Supplementary Plots

S1 Chapter 4: Cross-validation and MI for the continuous outcome

S1.1 Continuous outcome

In this section, all graphical output produced from the simulation study comparing methods for combining MI and CV is available. This includes graphs comparing the MSE estimates from the imputation methods to the fully-observed MSE estimates and, also, to the estimates from a larger validation set. In addition, graphs will be available for increased number of imputation datasets (M=25) and for increased proportion of missingness (40%).

S1.1.1 MSE from methods compared to the fully-observed MSE (MSE$_{imp}$-MSE$_{obs}$)

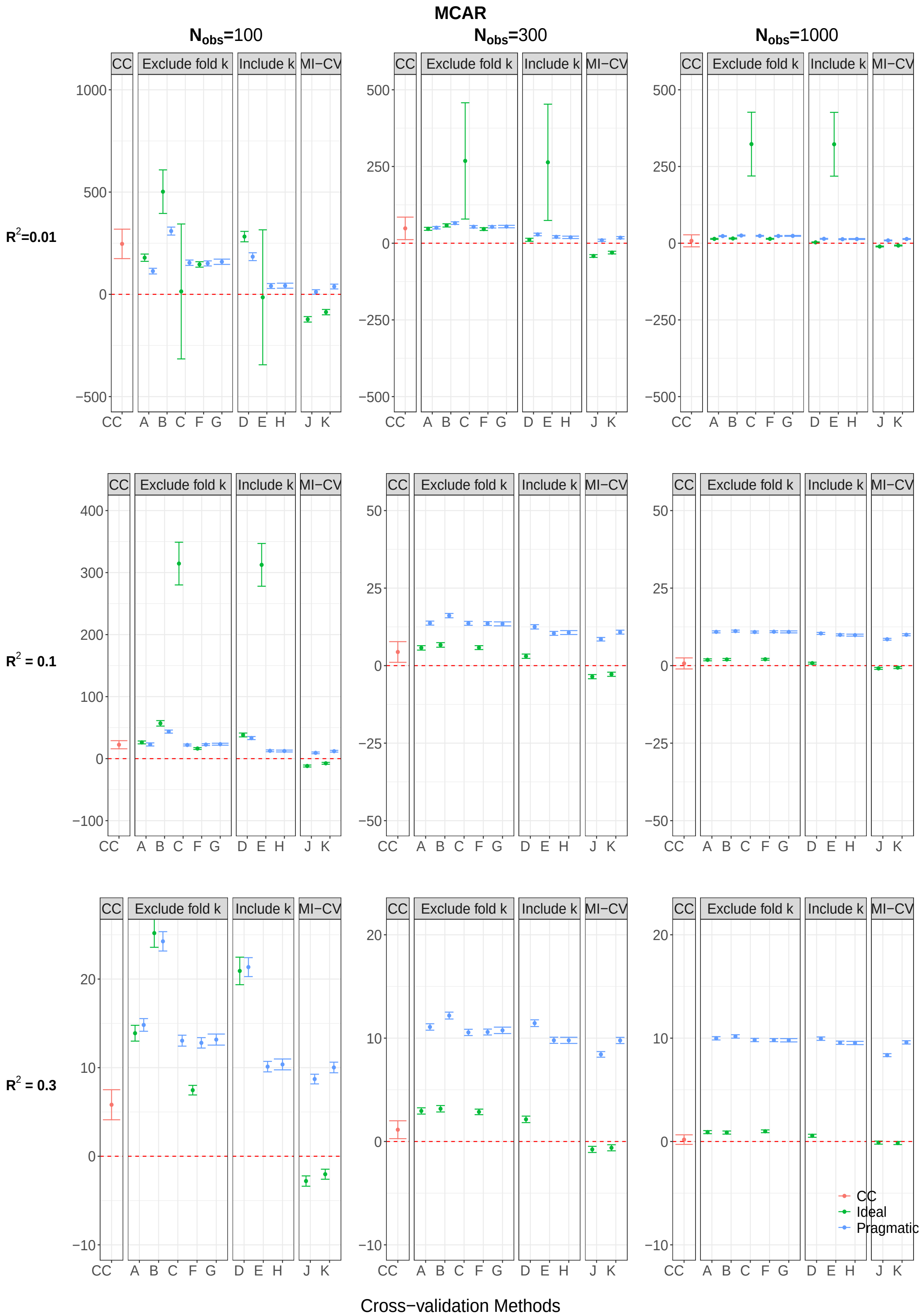

Figure S1: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

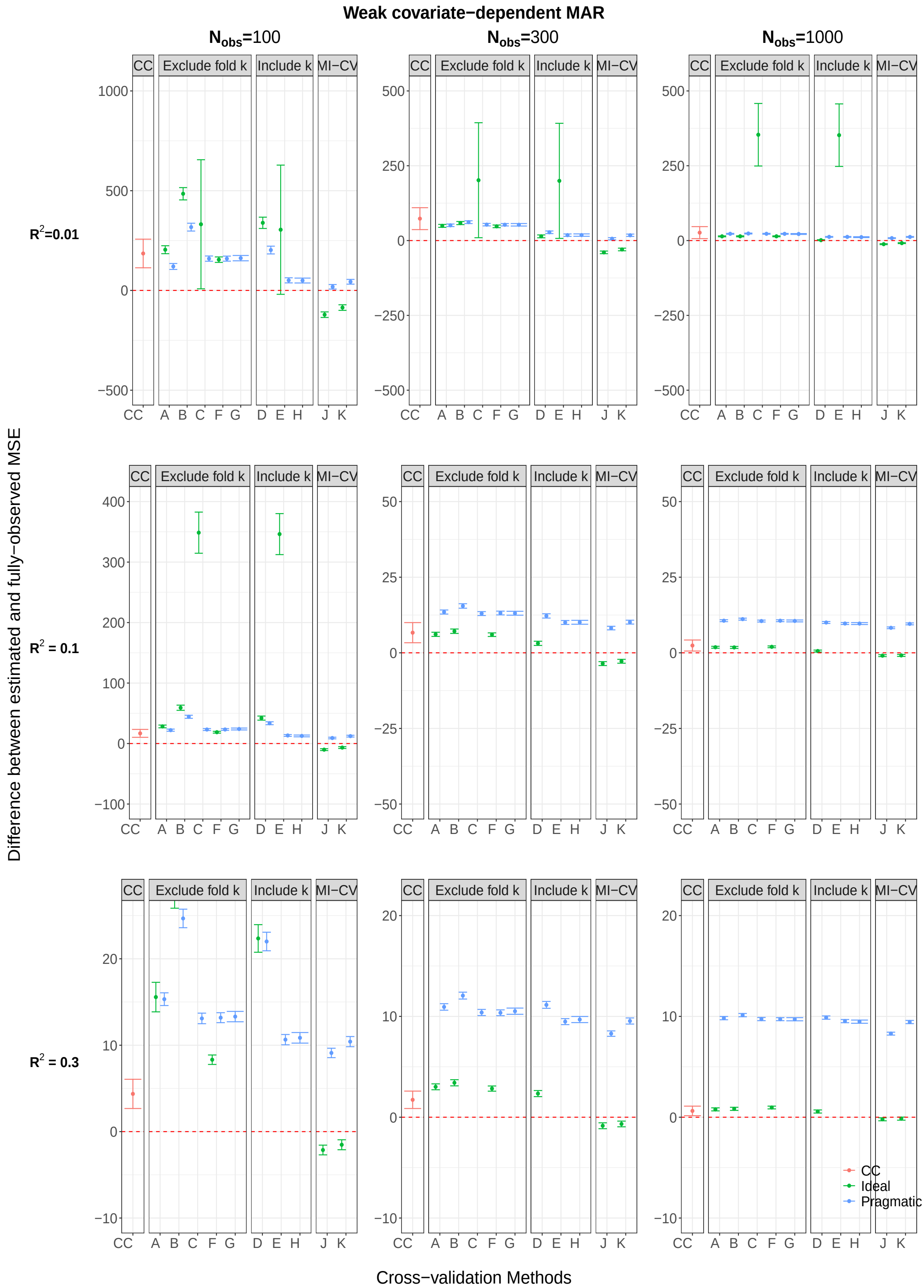

Figure S2: The difference MSE_{imp} - MSE_{obs} when data are weak covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

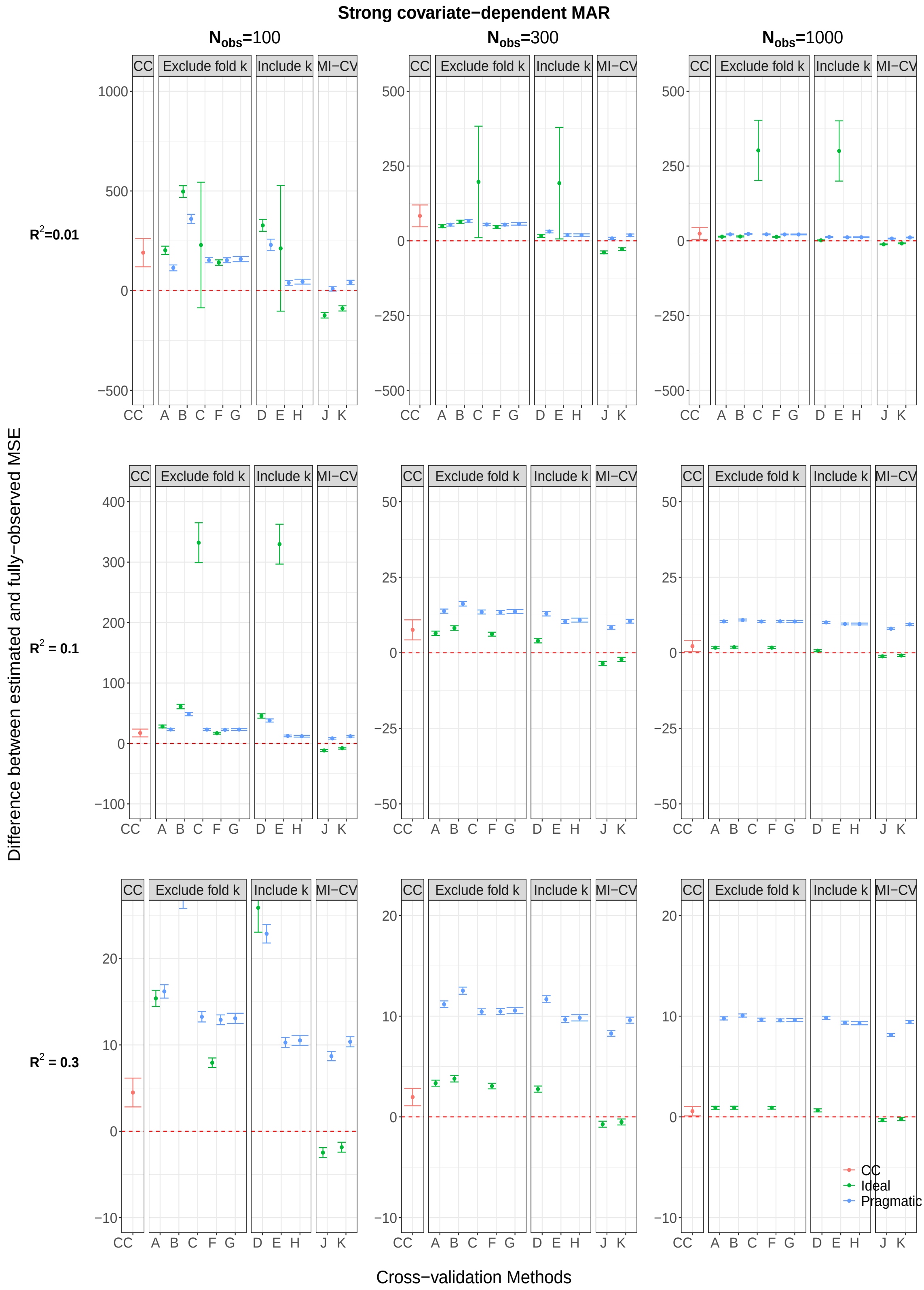

Figure S3: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

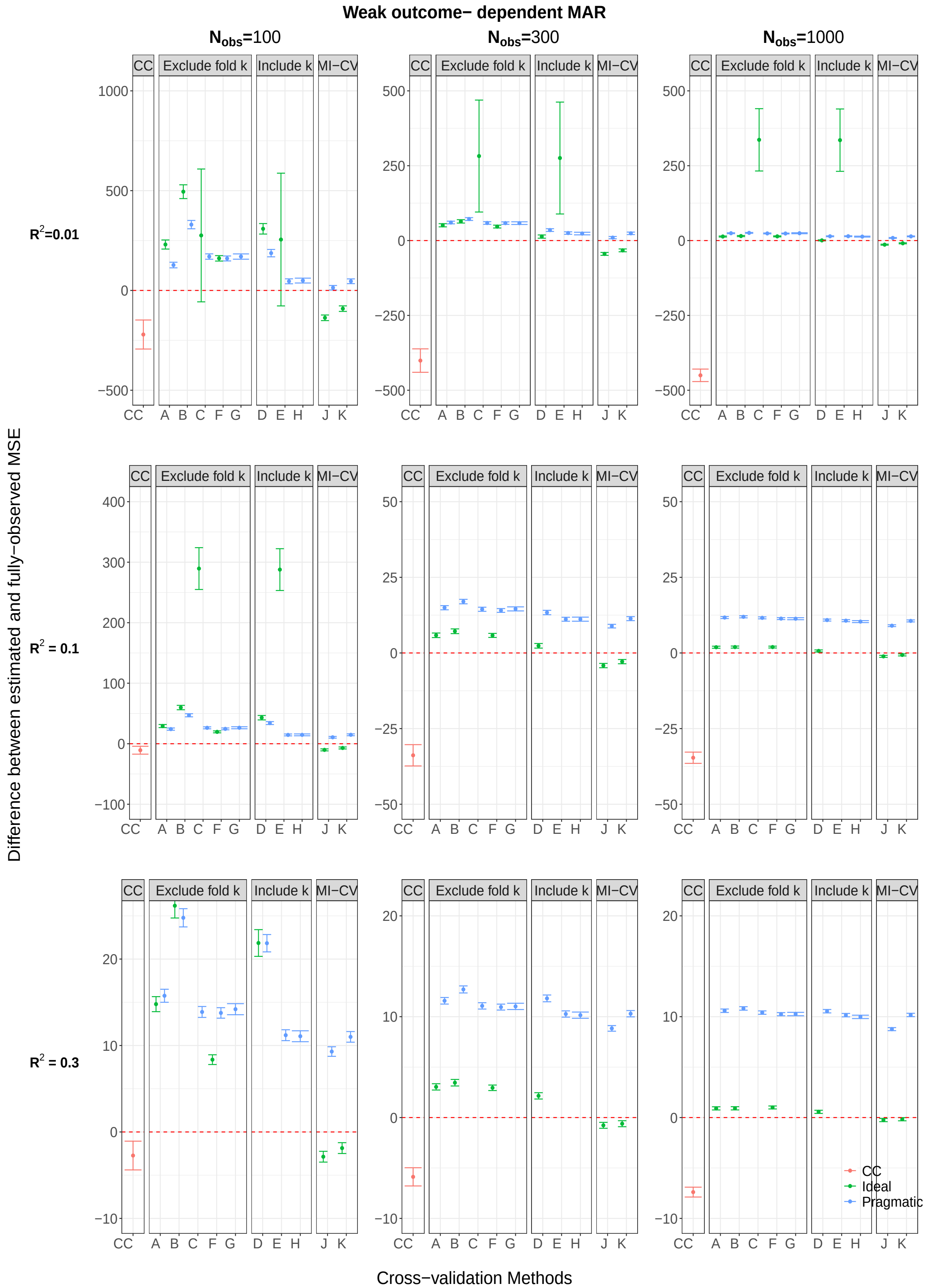

Figure S4: The difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

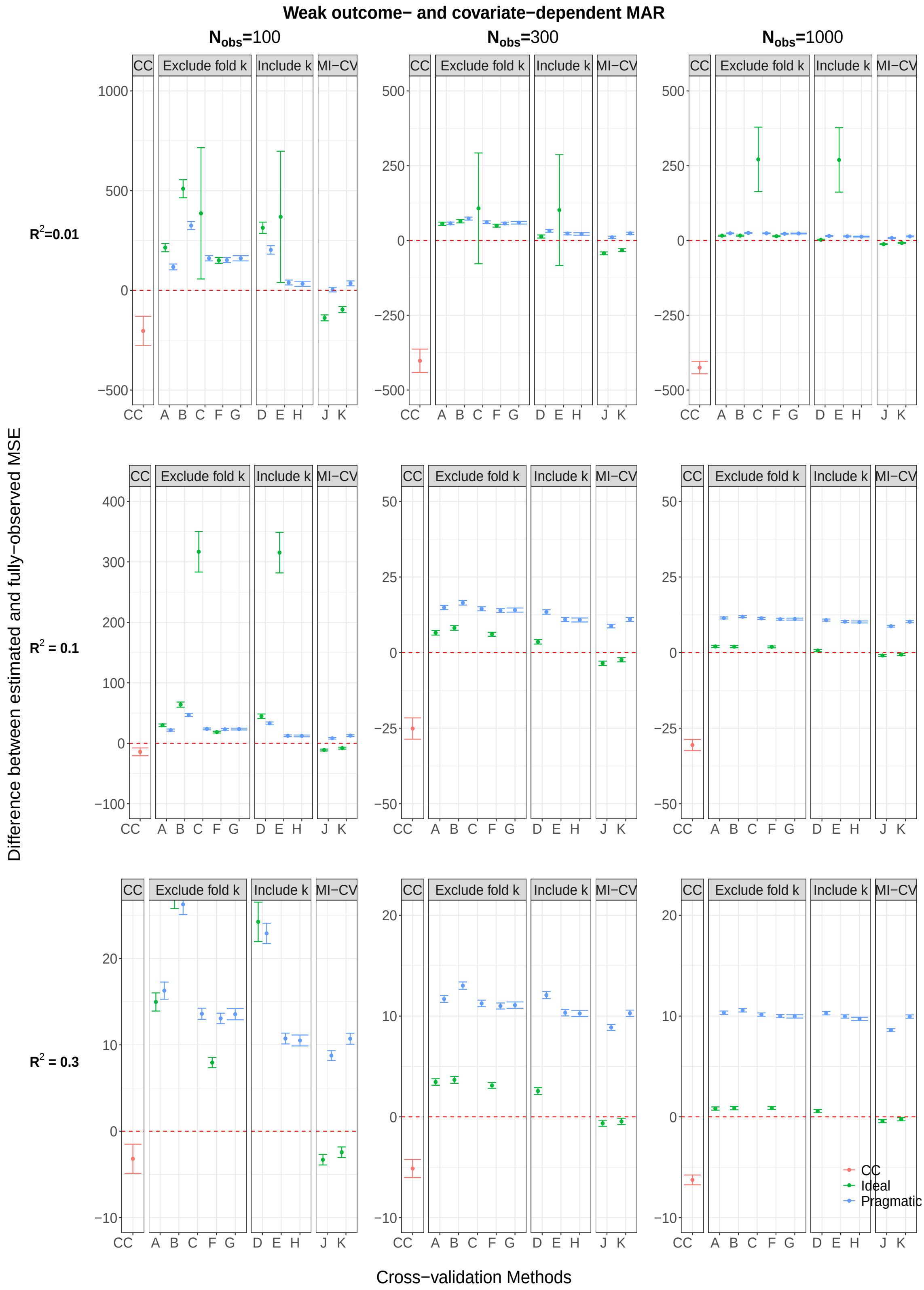

Figure S5: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

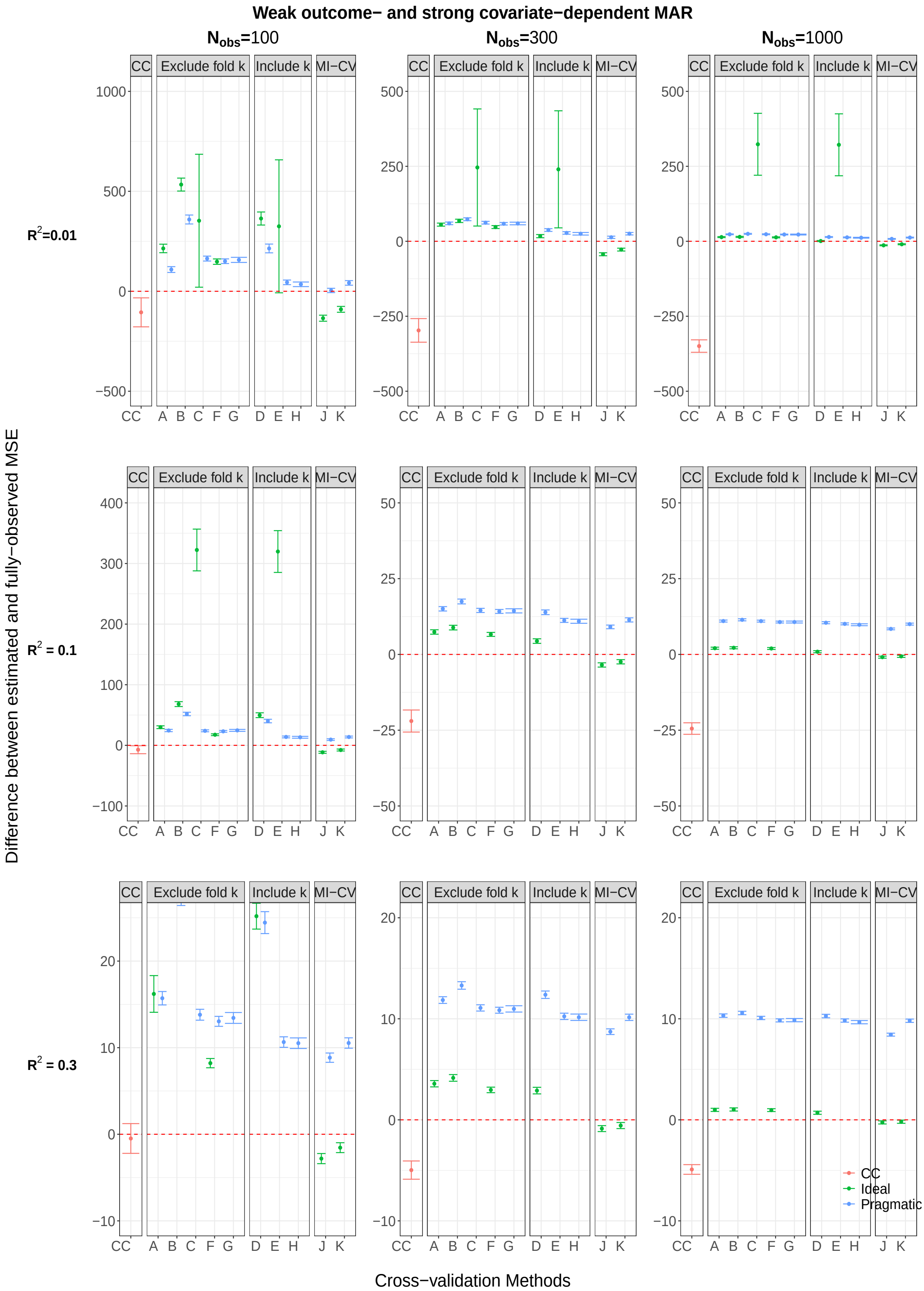

Figure S6: The difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S1.1.2 Comparing data leakage (MSE$_{imp}$ - MSE$_{obs}$)

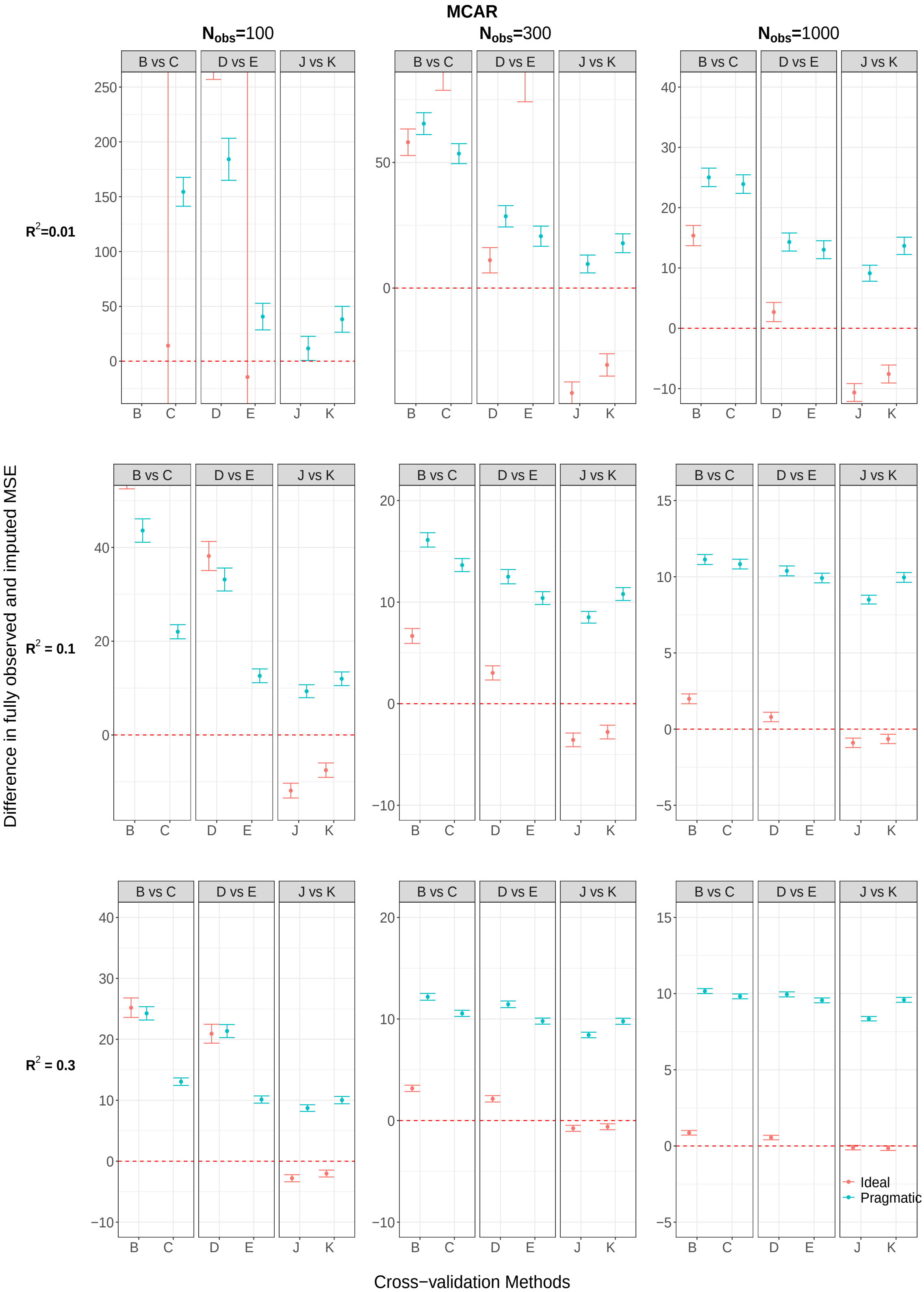

Figure S7: Assessing data leakage within the imputation process for cross-validation. The difference MSE$_{imp}$ - MSE$_{obs}$ is compared when data are MCAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

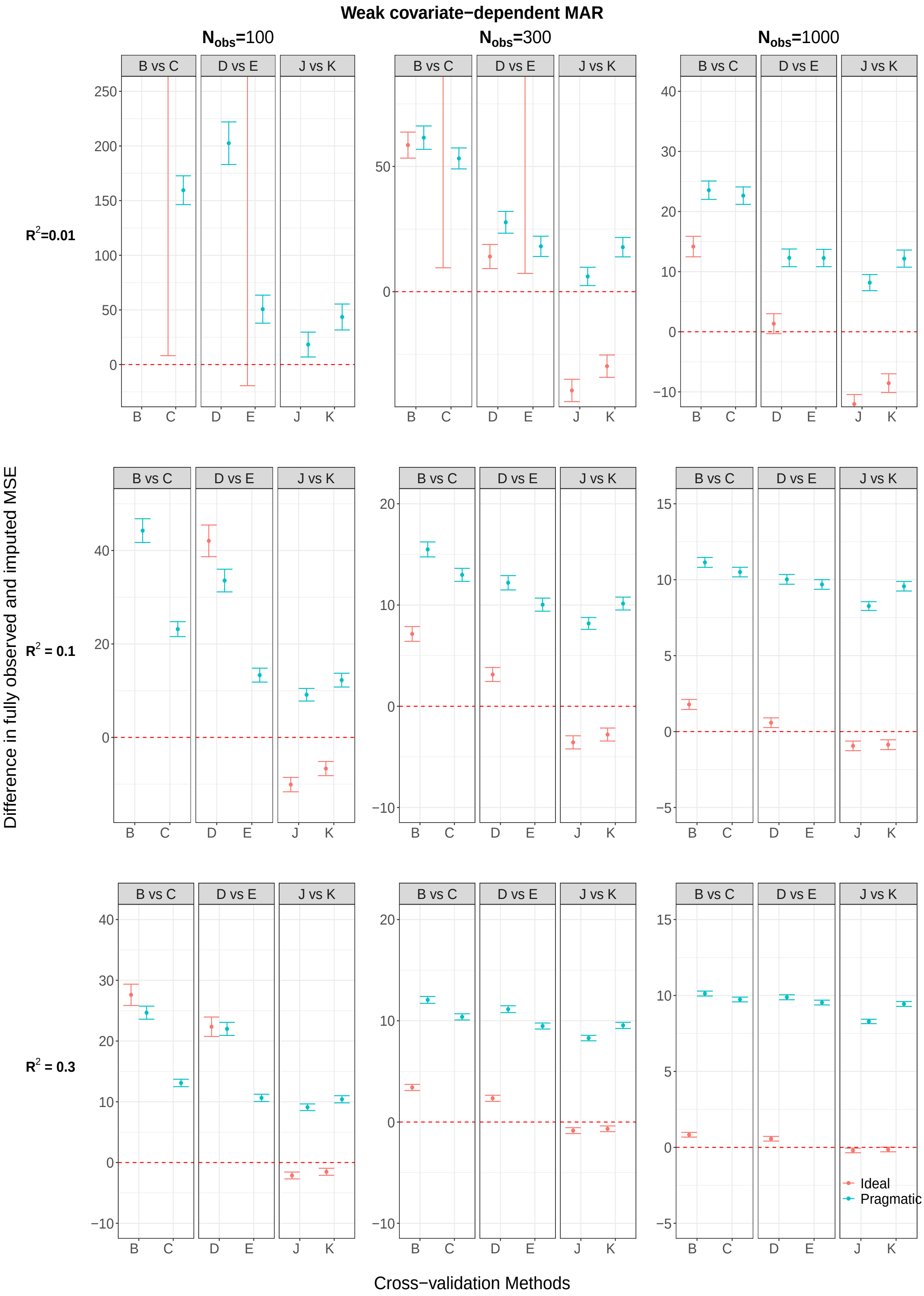

Figure S8: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{obs} is compared when data are weak covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

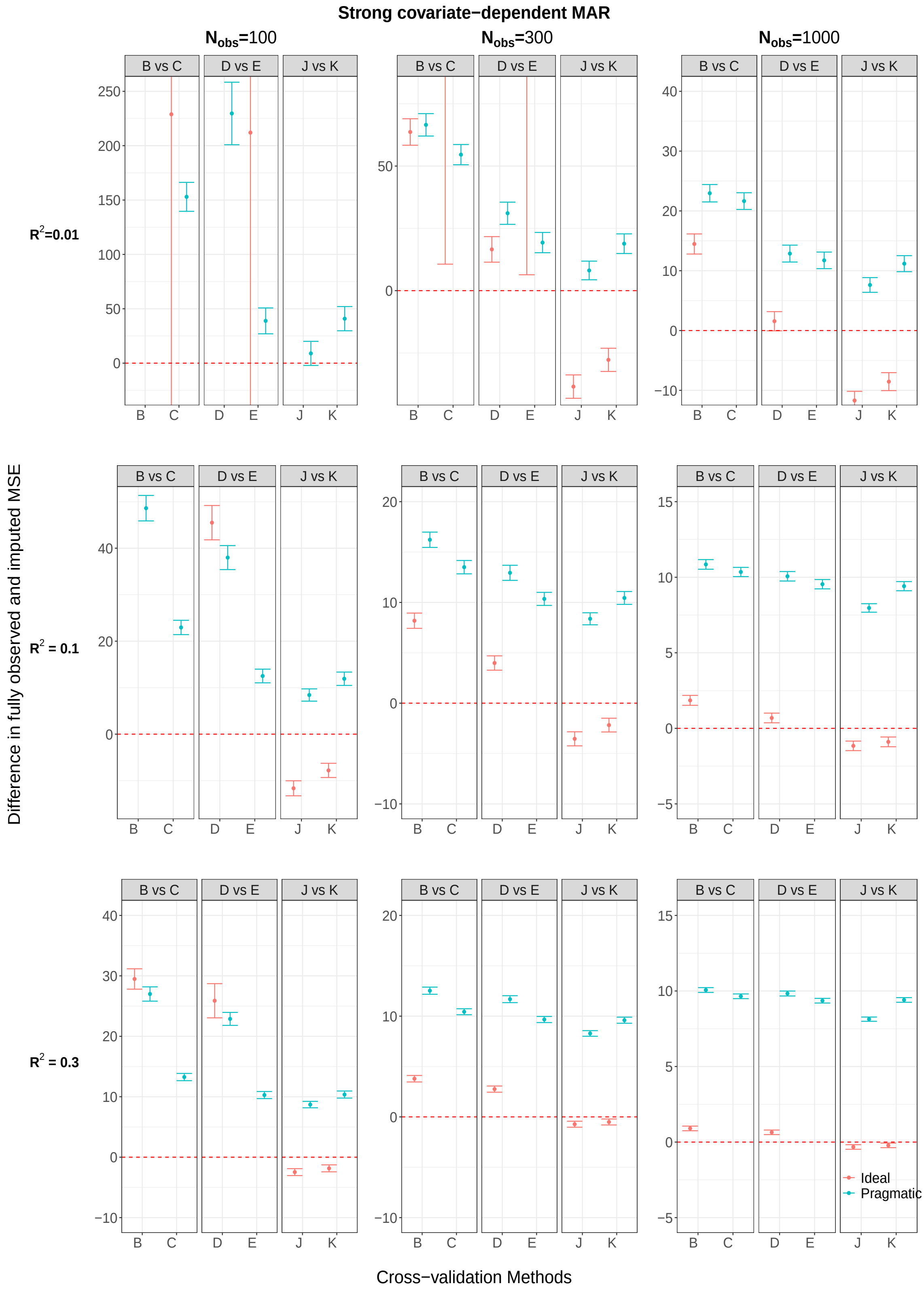

Figure S9: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{obs} is compared when data are strong covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

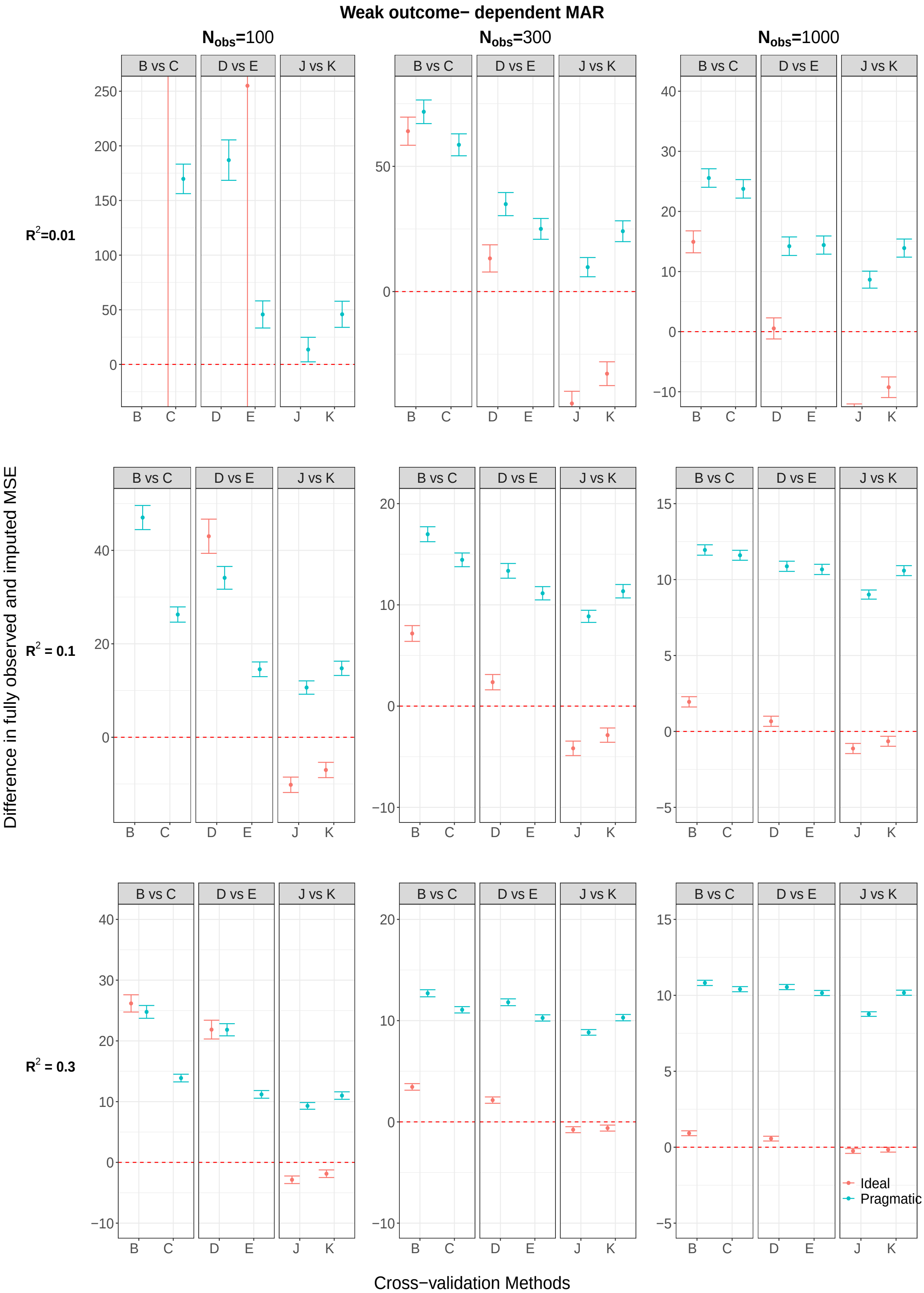

Figure S10: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{obs} is compared when data are weak outcome-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

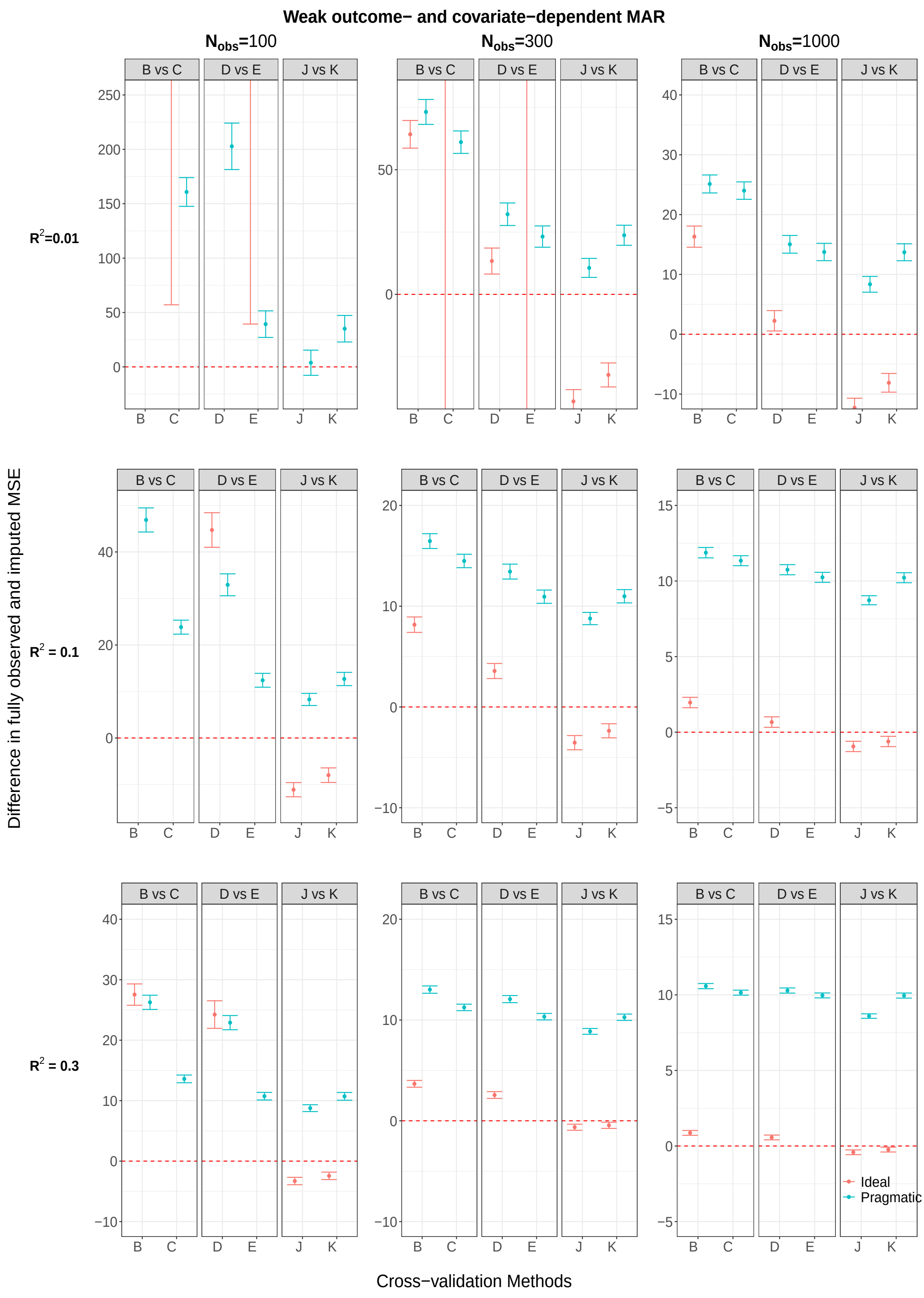

Figure S11: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{obs} is compared when data are weak outcome- and covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

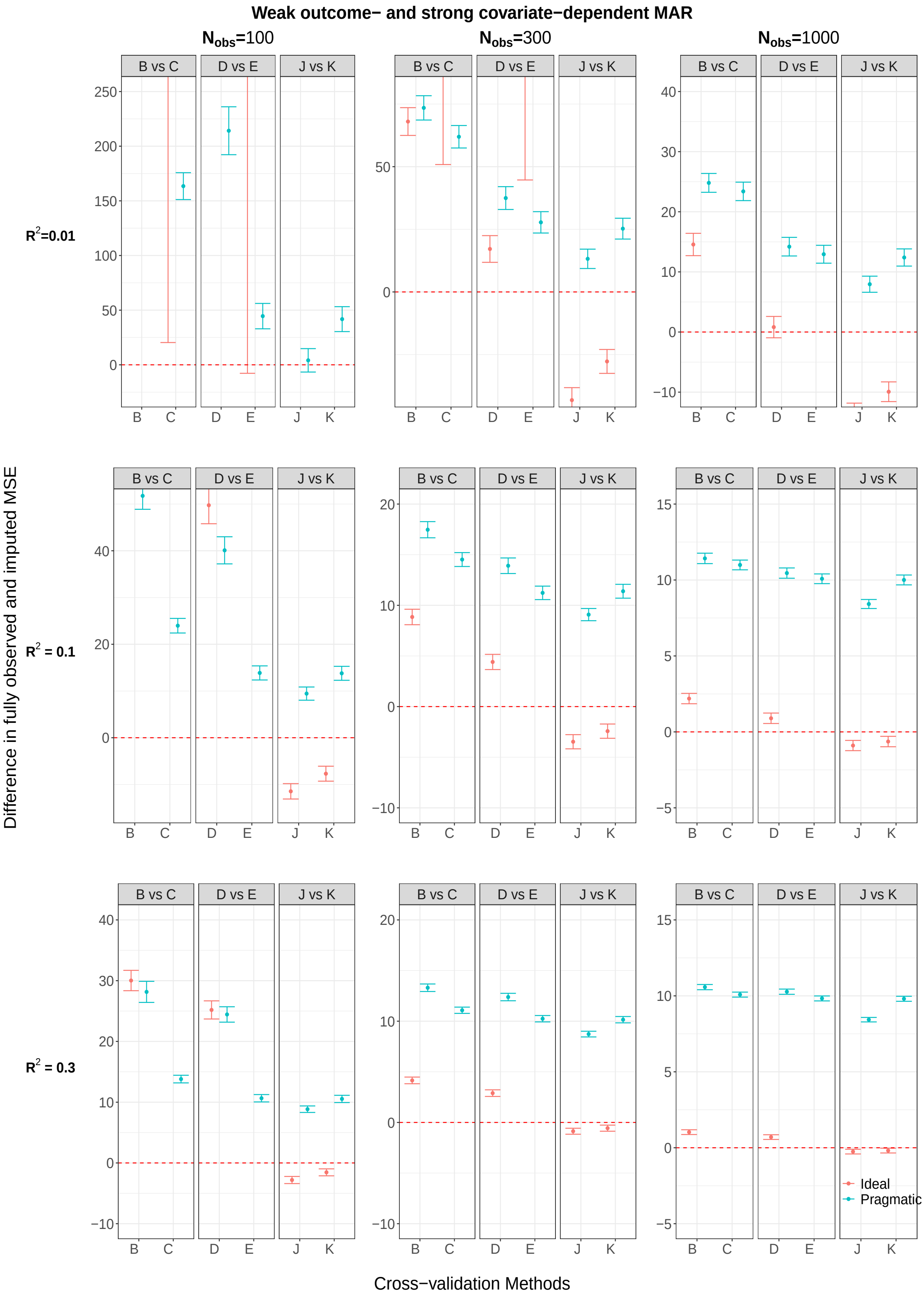

Figure S12: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{obs} is compared when data are weak outcome- and strong covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S1.1.3 The proportion of missingness is 40% (MSE_{imp}-MSE_{obs})

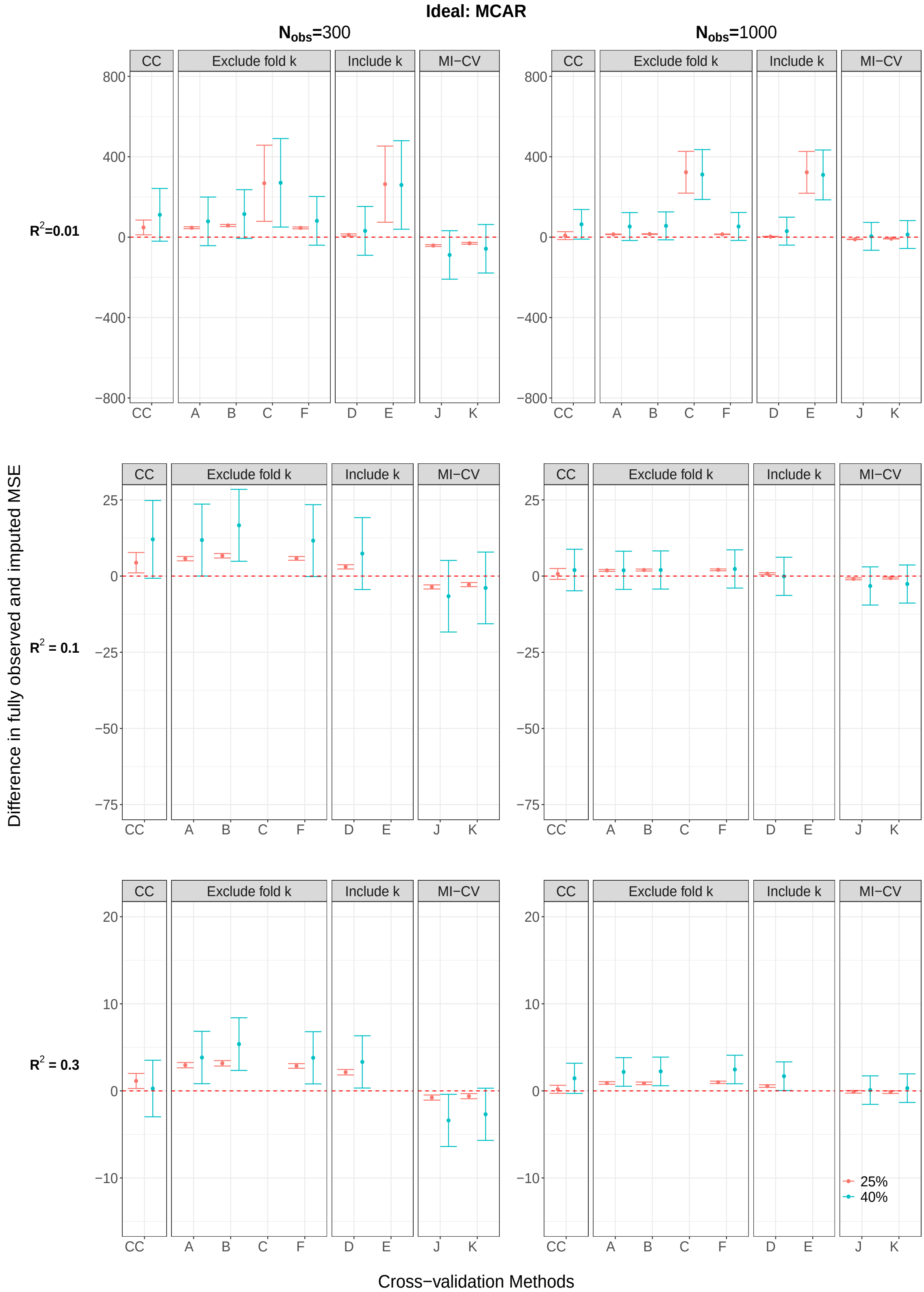

Figure S13: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are MCAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

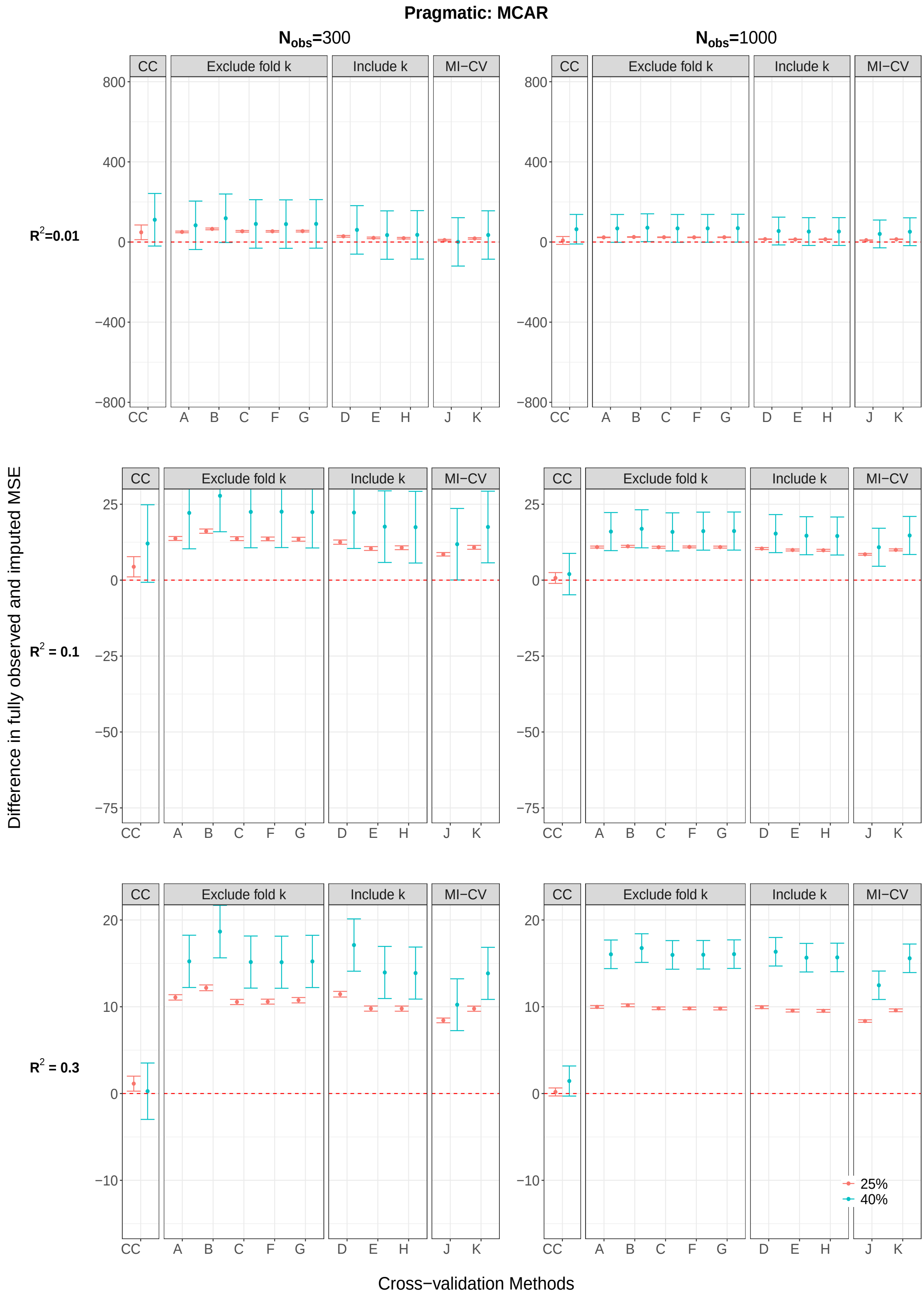

Figure S14: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are MCAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

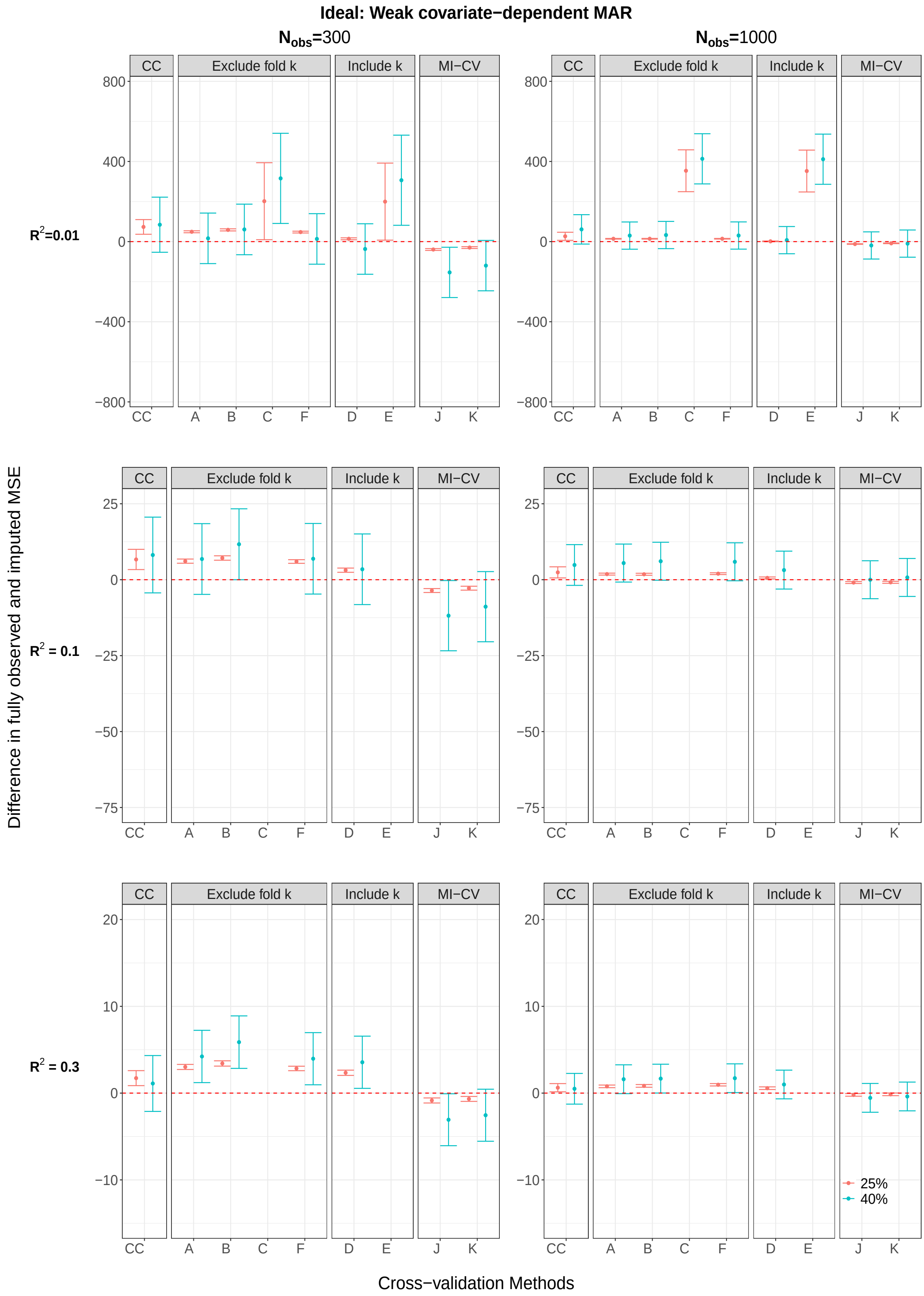

Figure S15: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

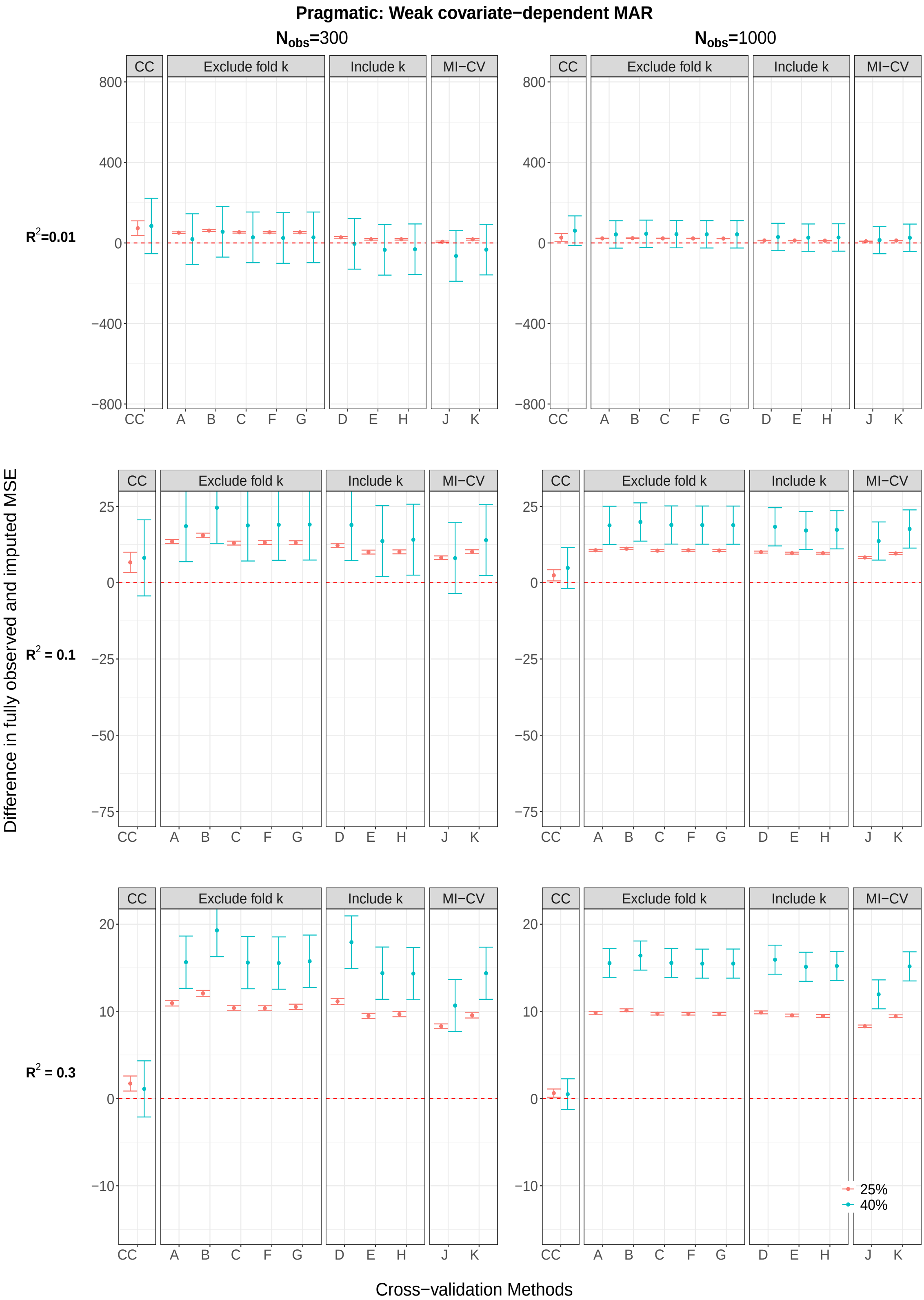

Figure S16: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

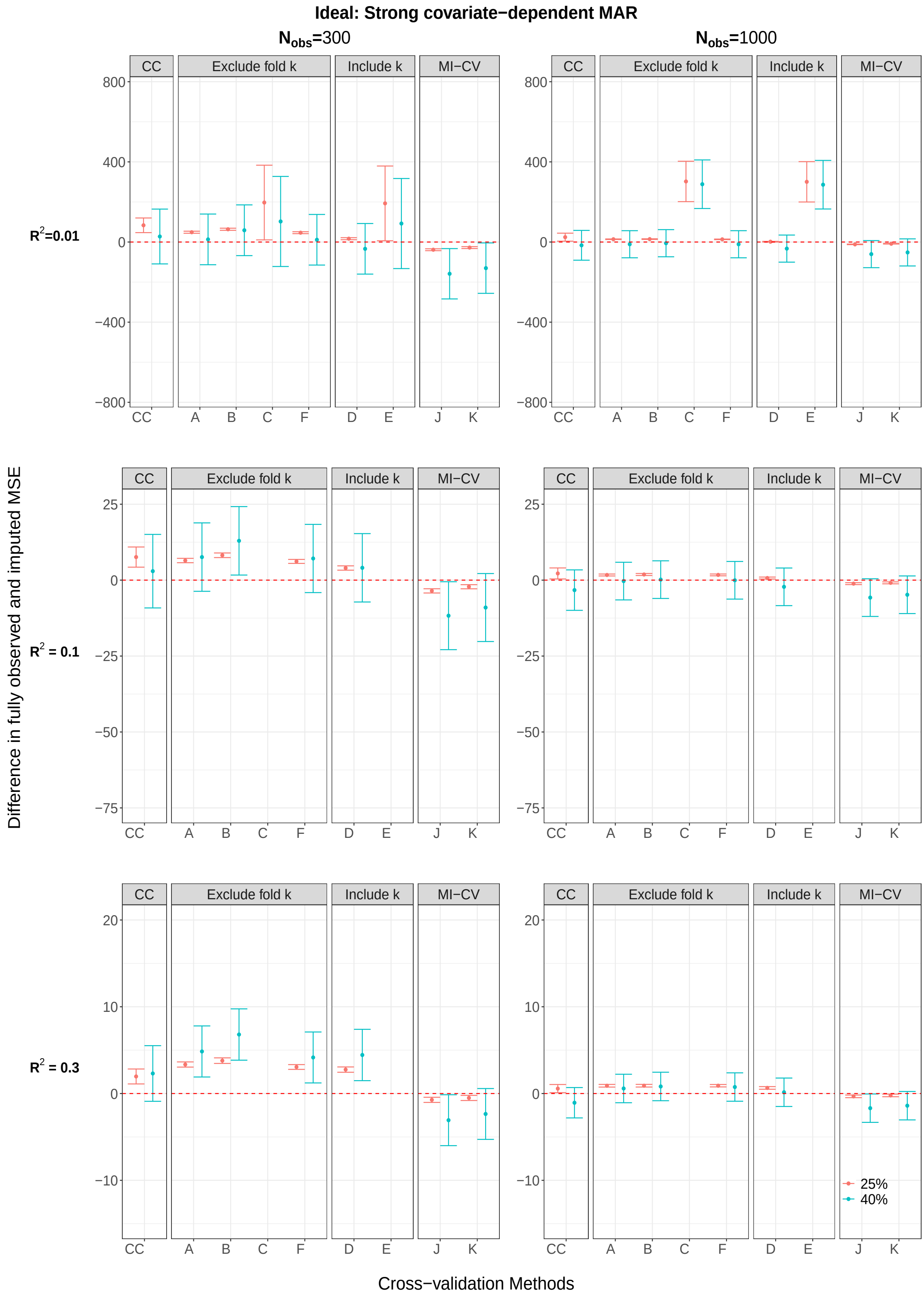

Figure S17: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

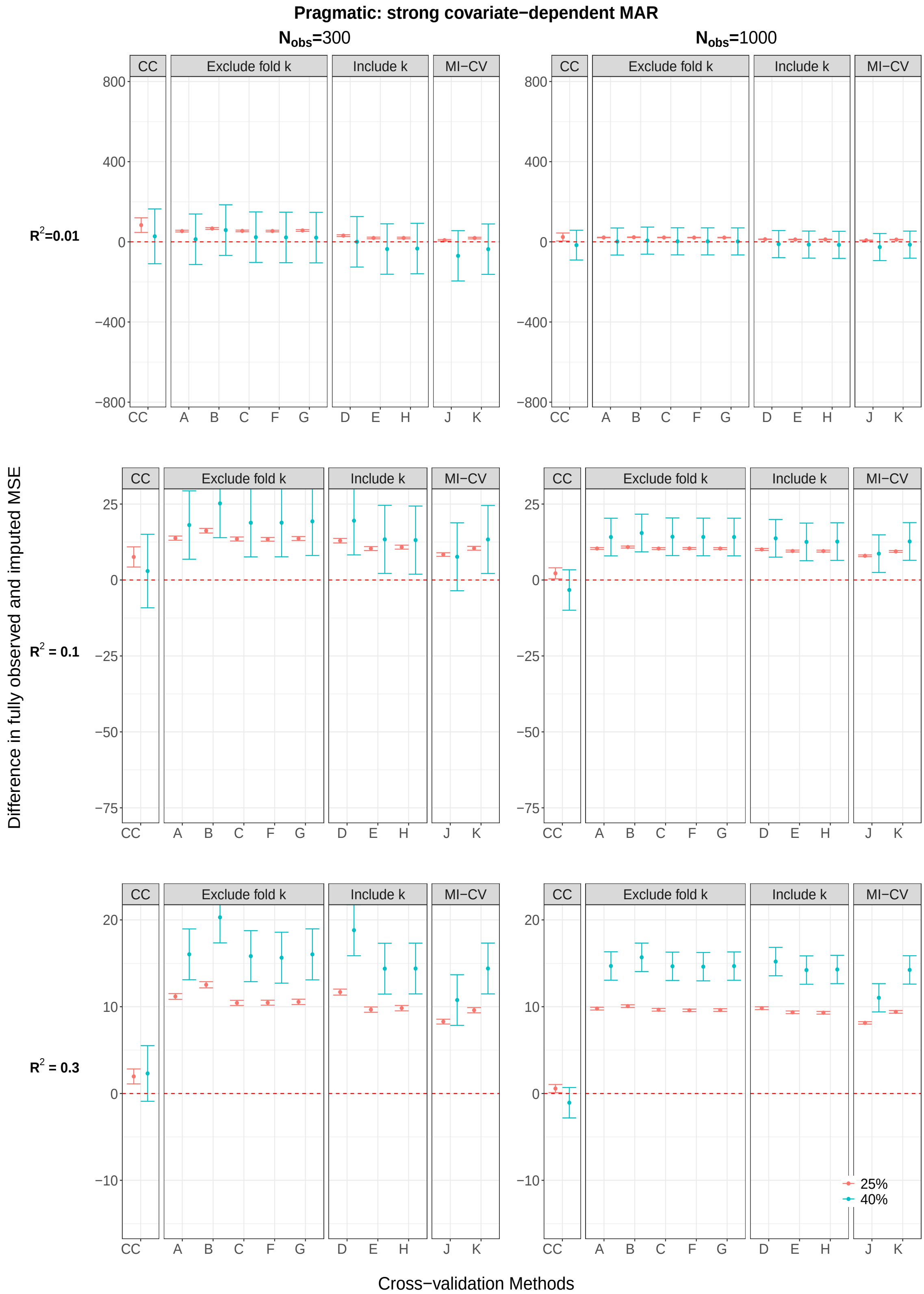

Figure S18: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

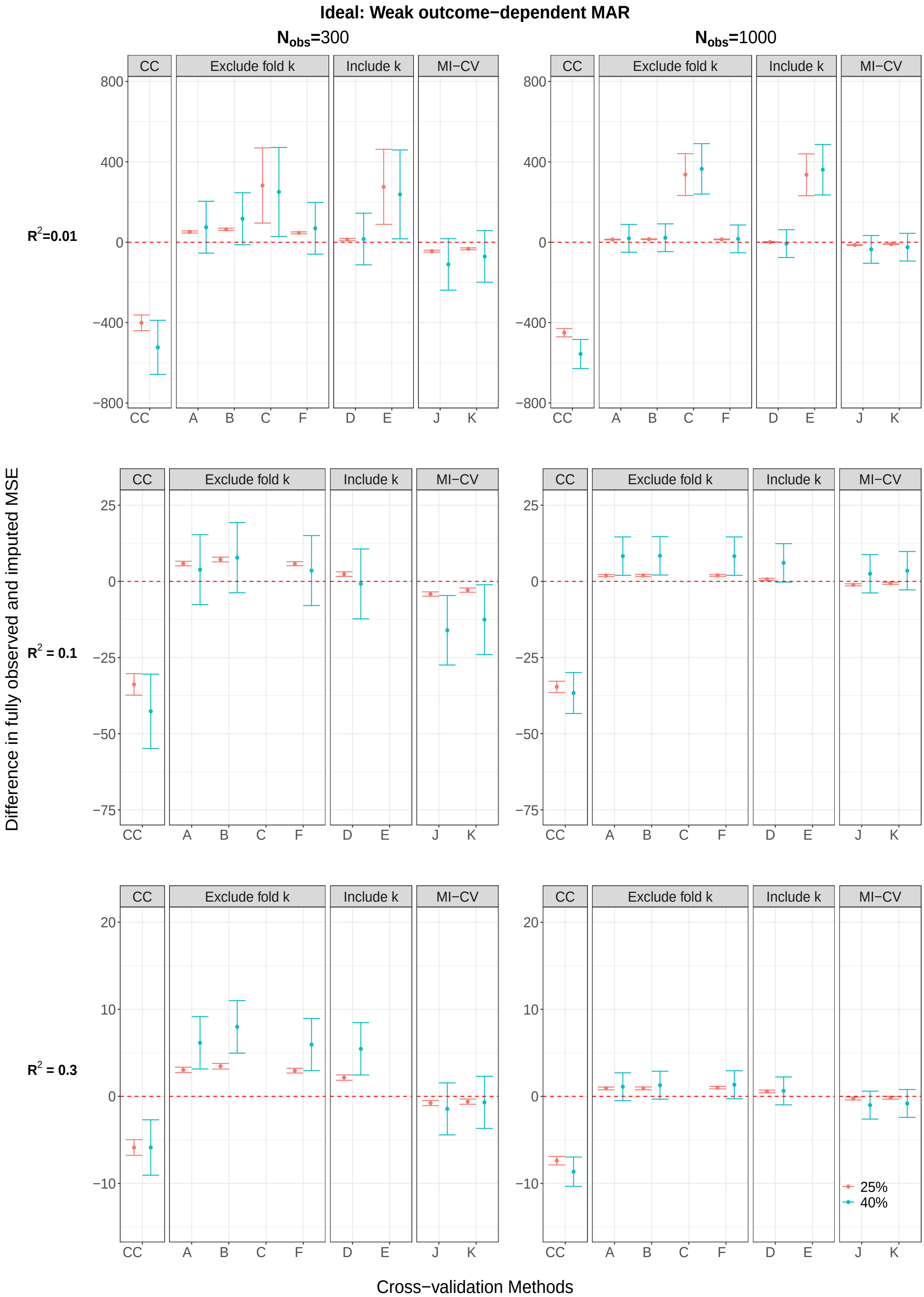

Figure S19: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

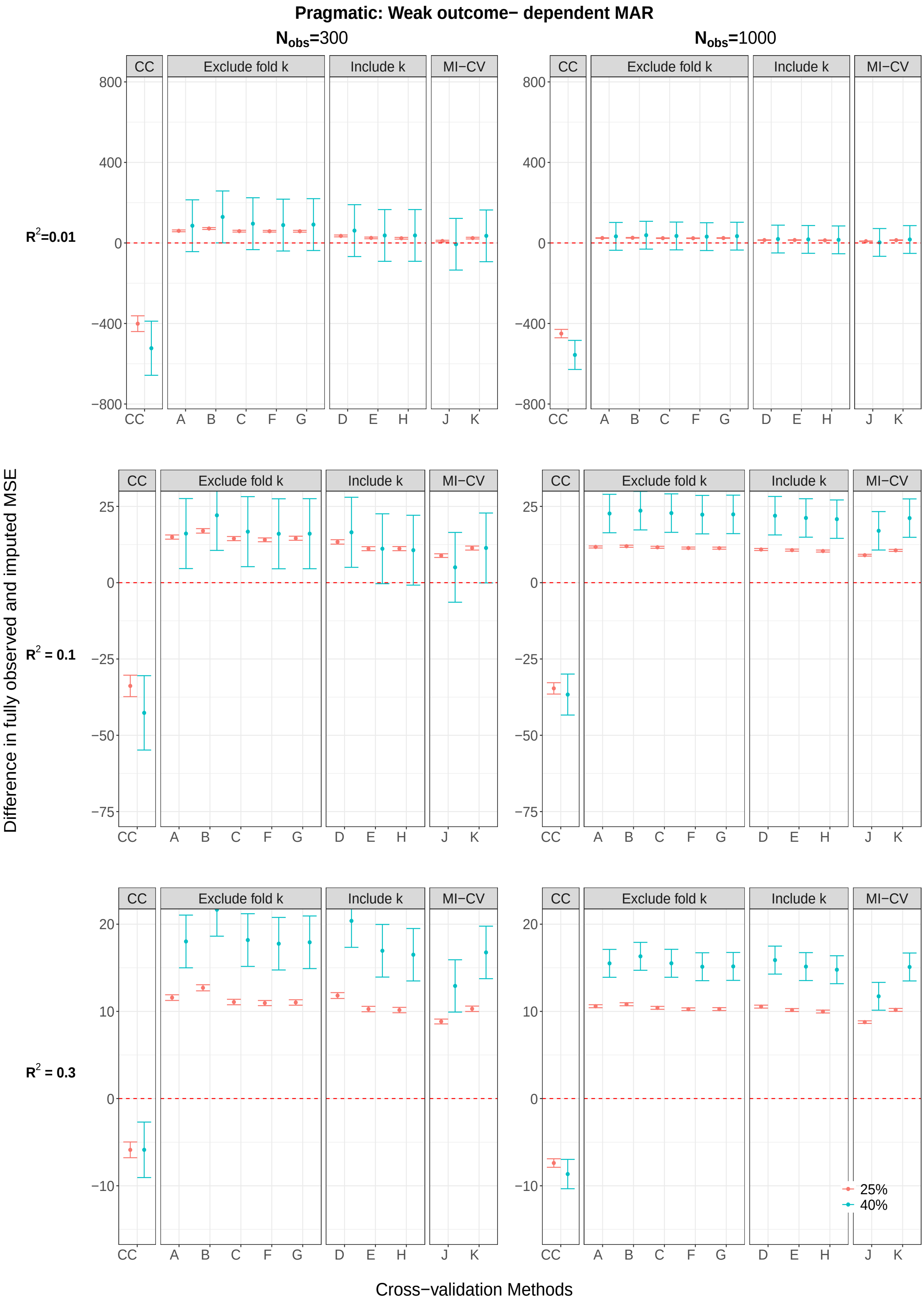

Figure S20: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

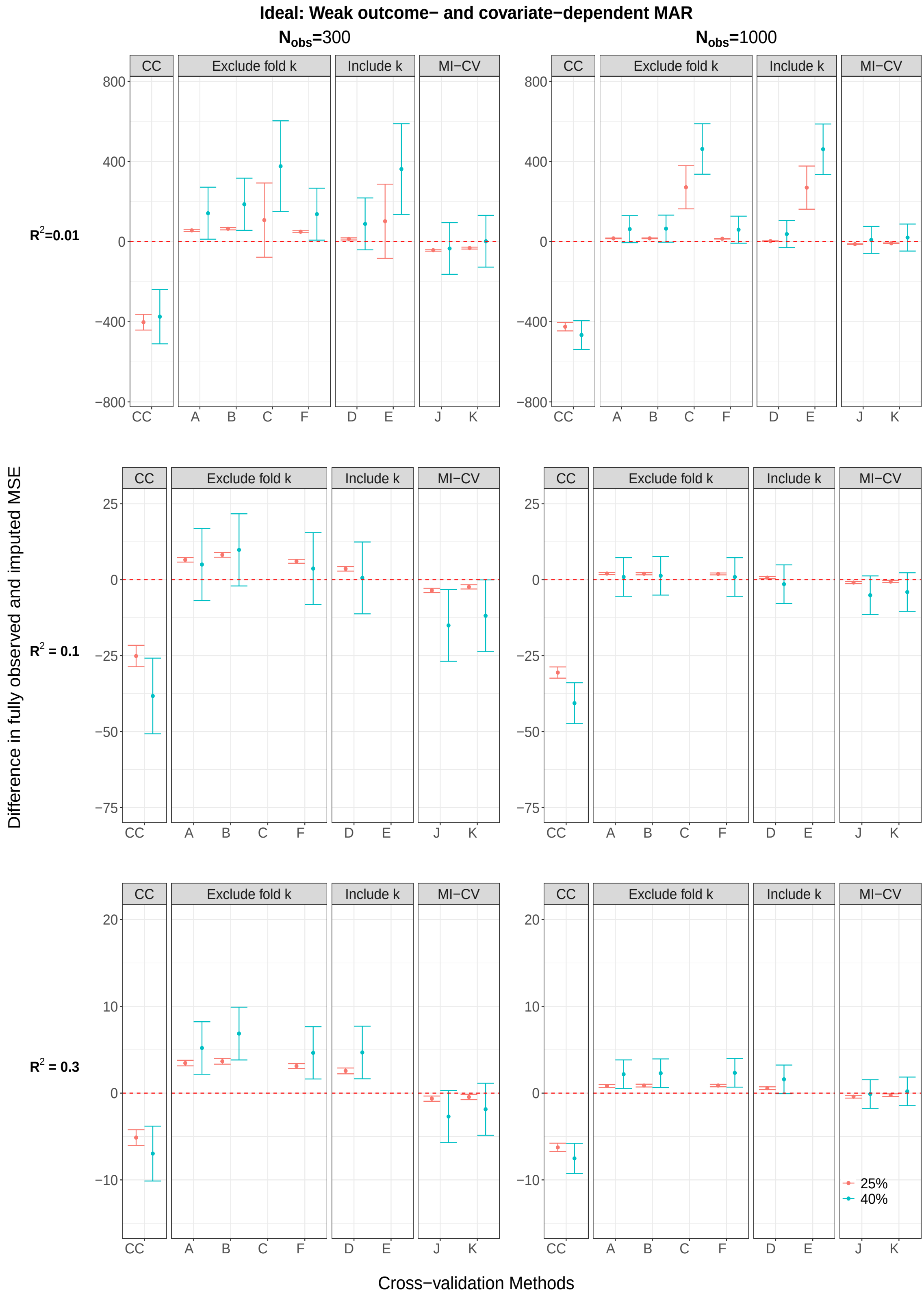

Figure S21: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome- and covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

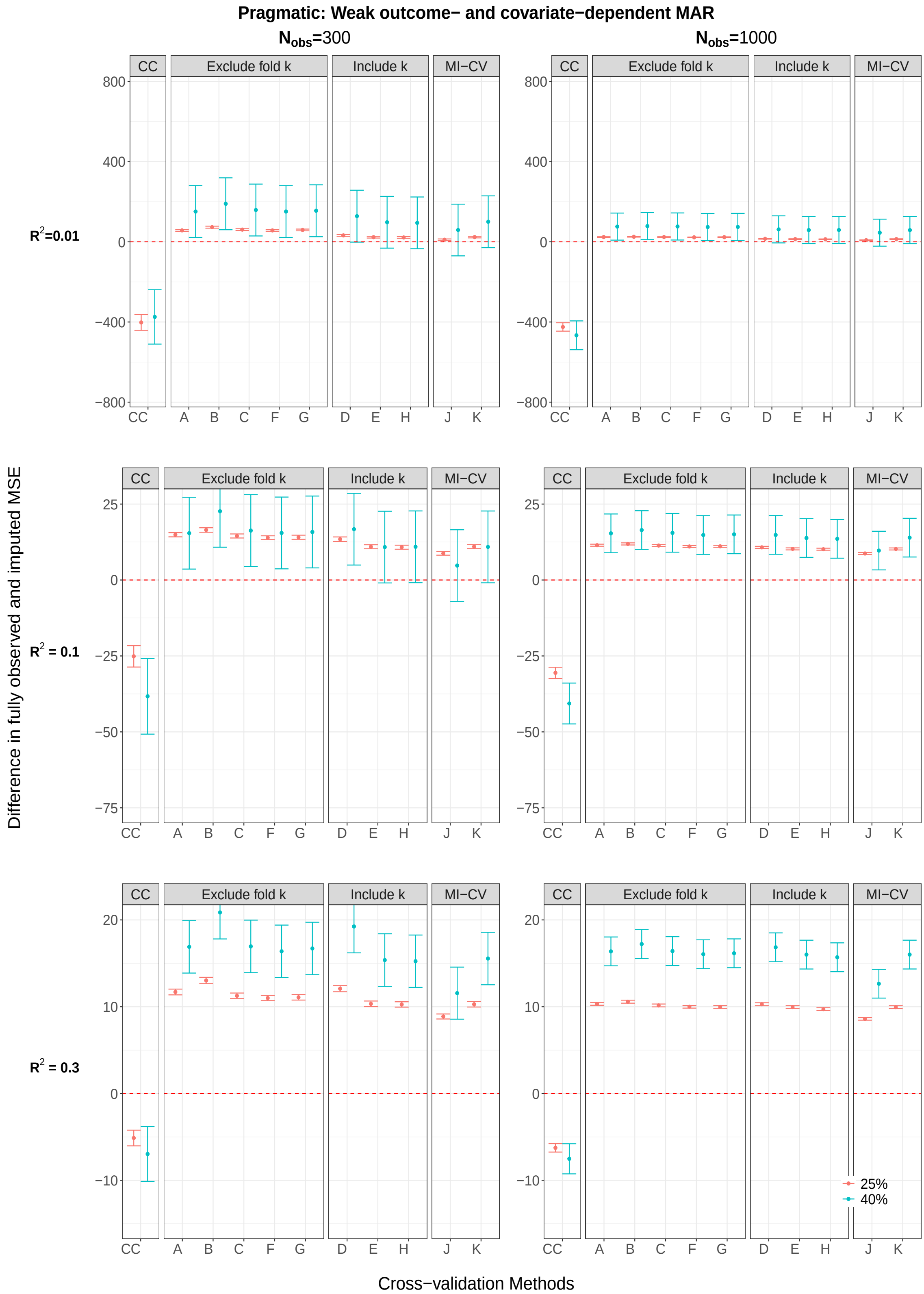

Figure S22: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

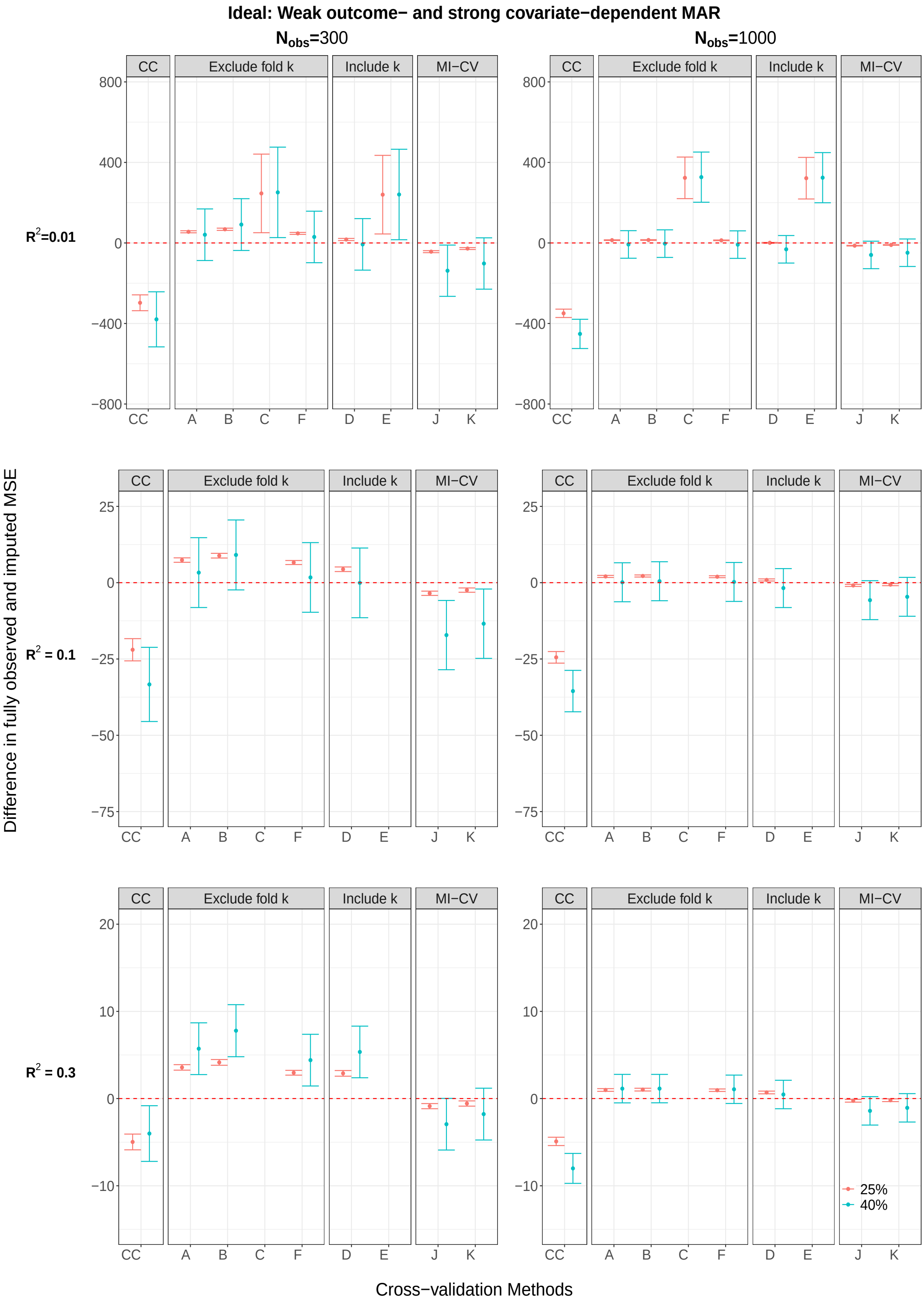

Figure S23: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

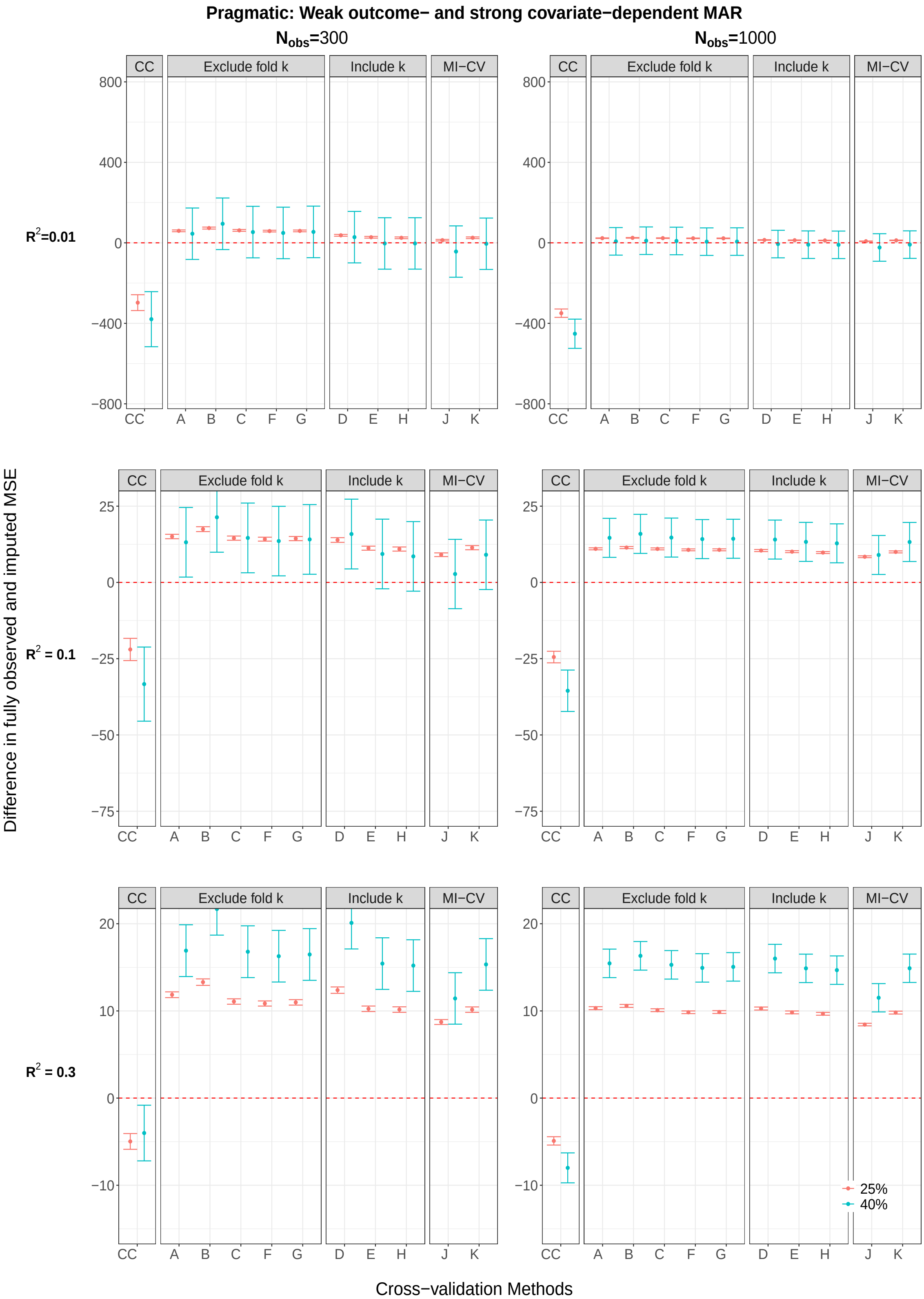

Figure S24: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR and $R^2 = 0.1$ for cross-validation when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S1.1.4 Comparing M=5 versus M=25 (MSE$_{imp}$-MSE$_{obs}$)

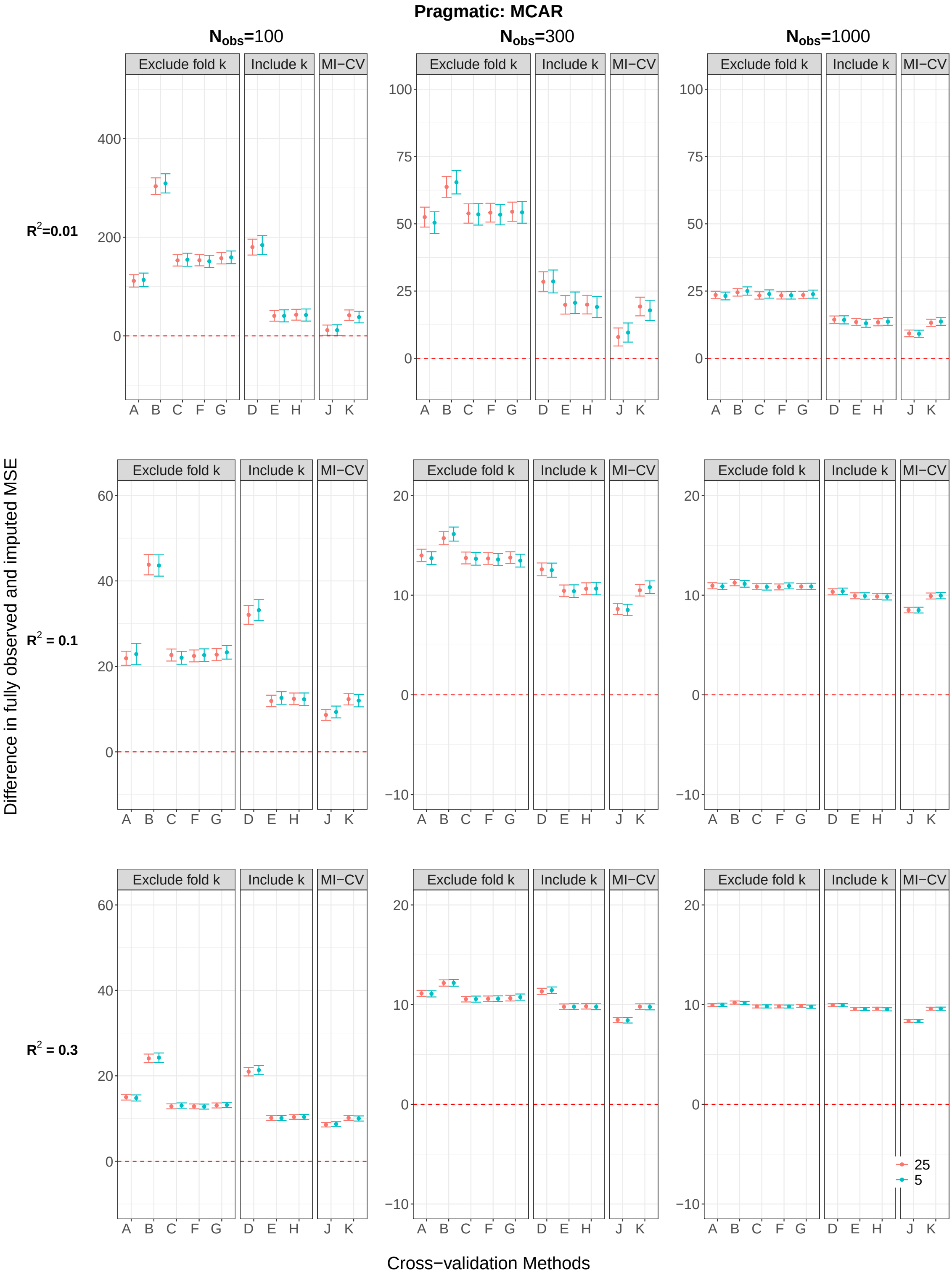

Figure S25: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

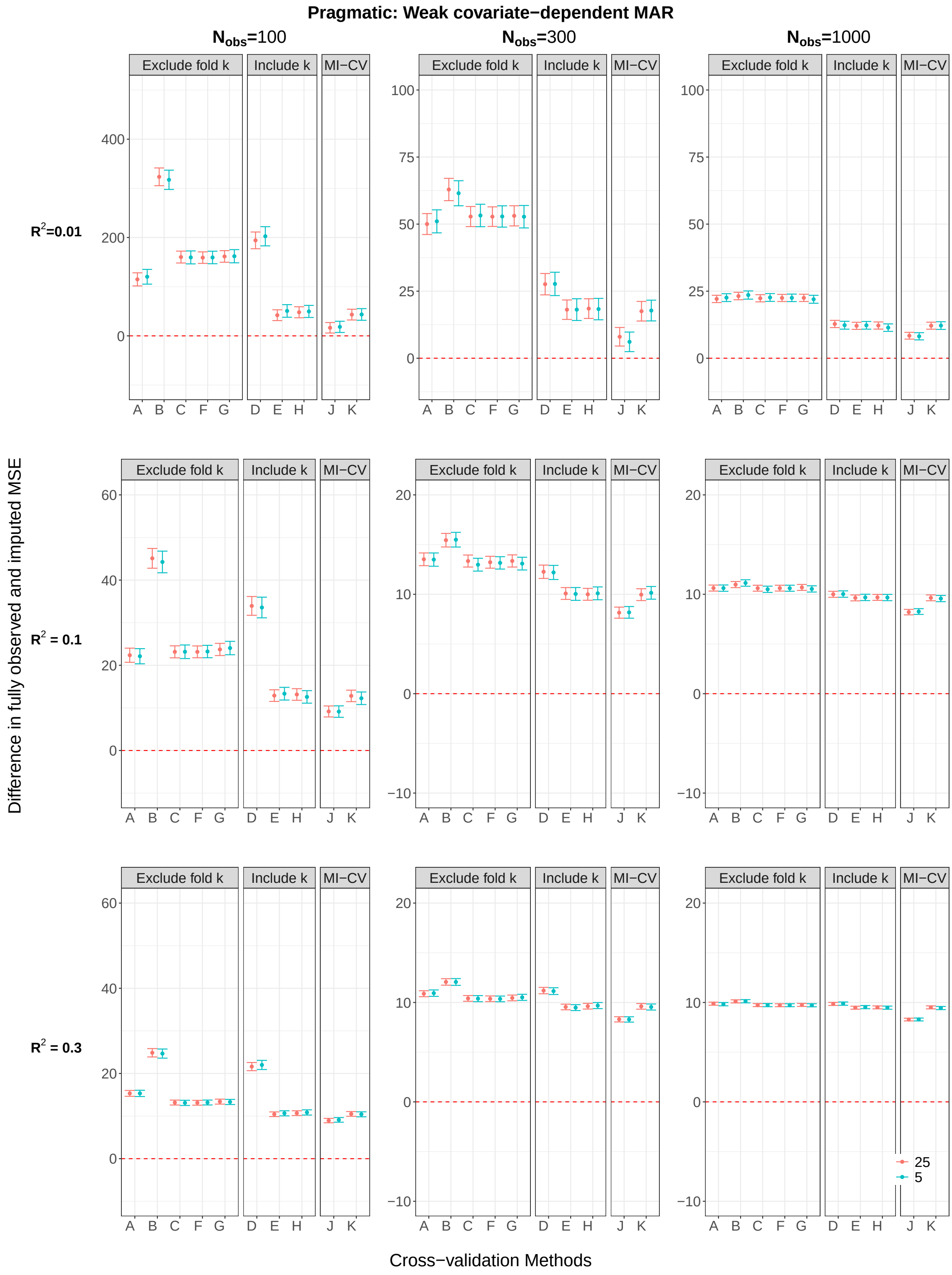

Figure S26: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

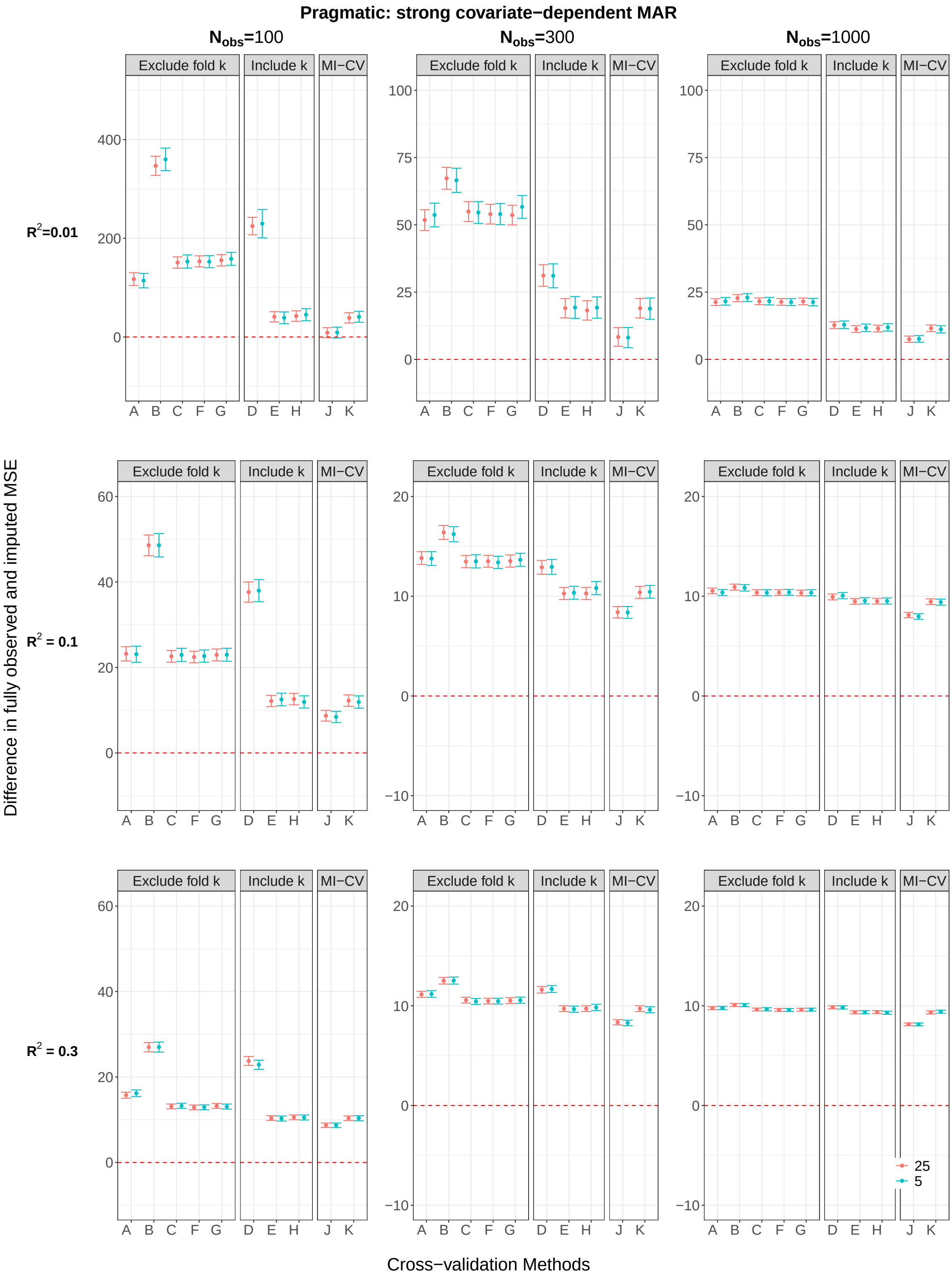

Figure S27: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

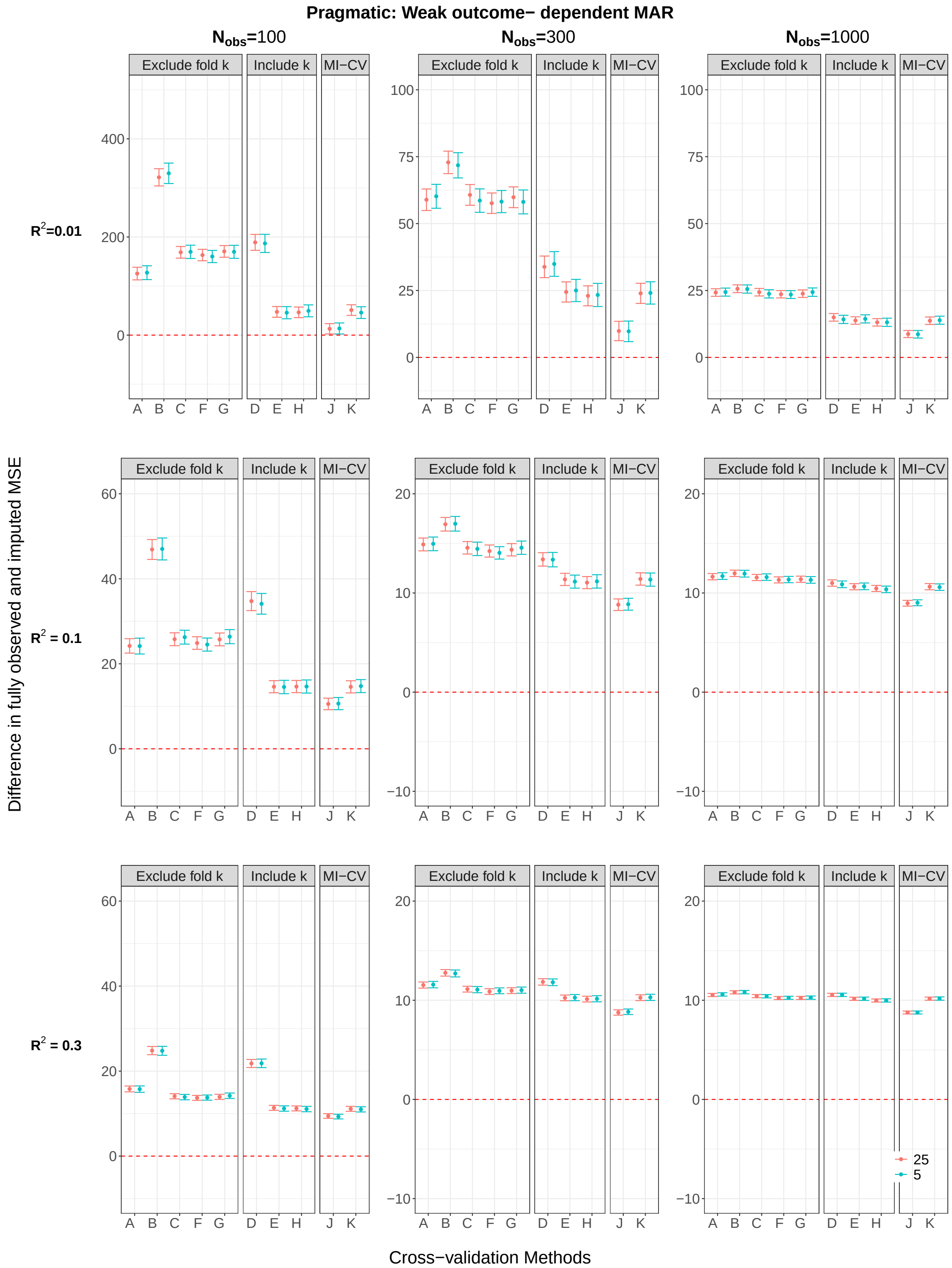

Figure S28: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

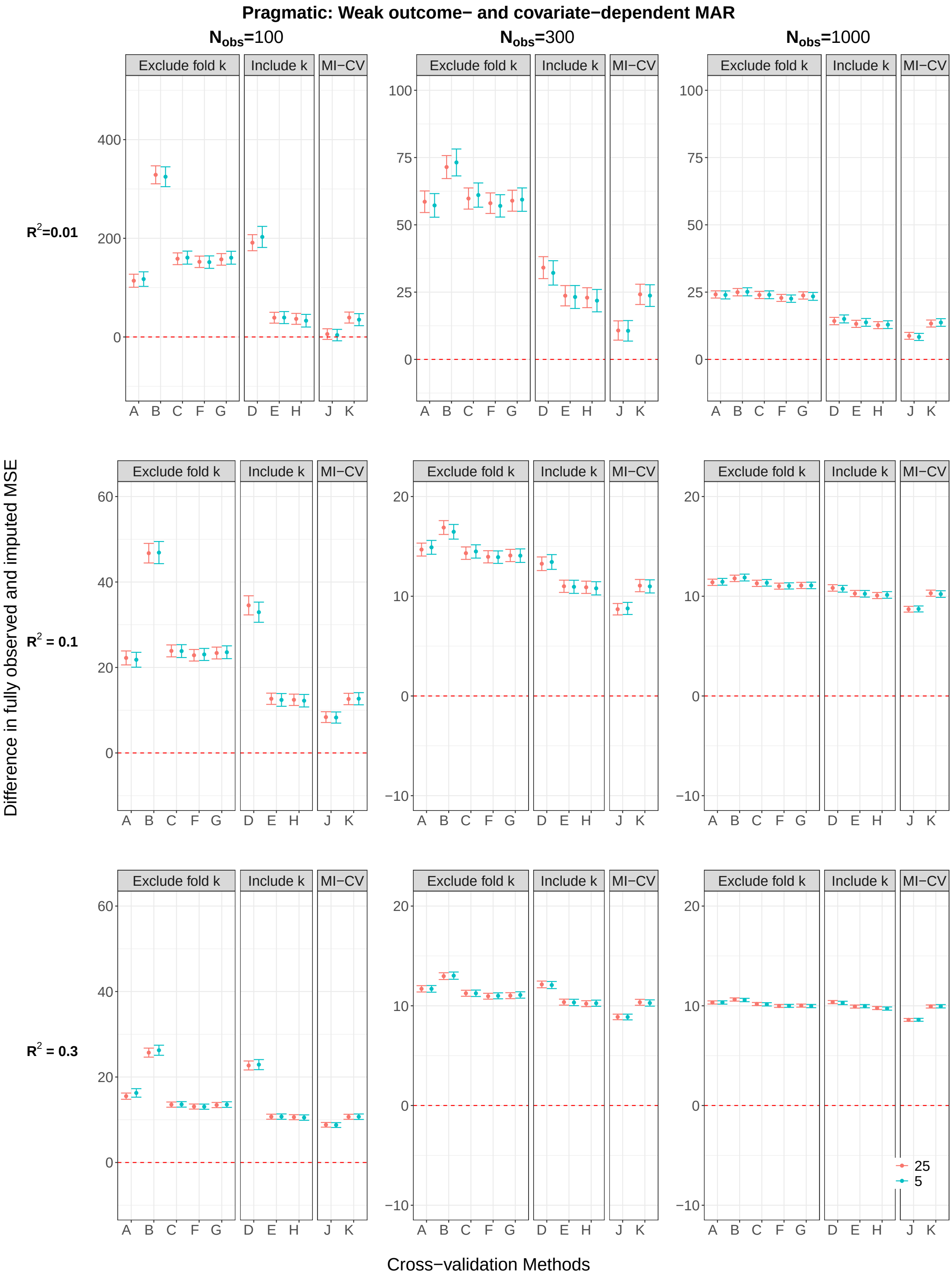

Figure S29: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

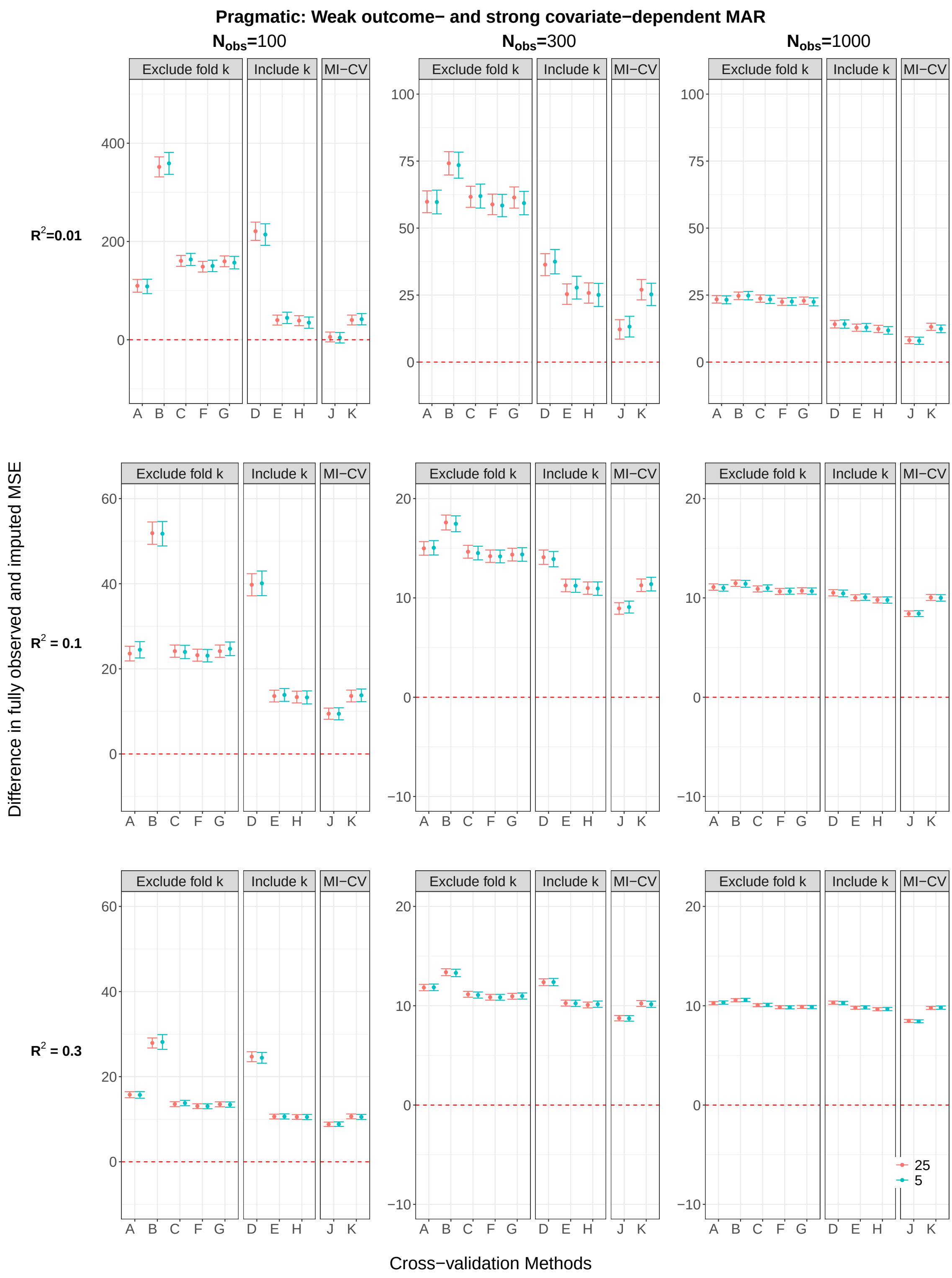

Figure S30: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and strong covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

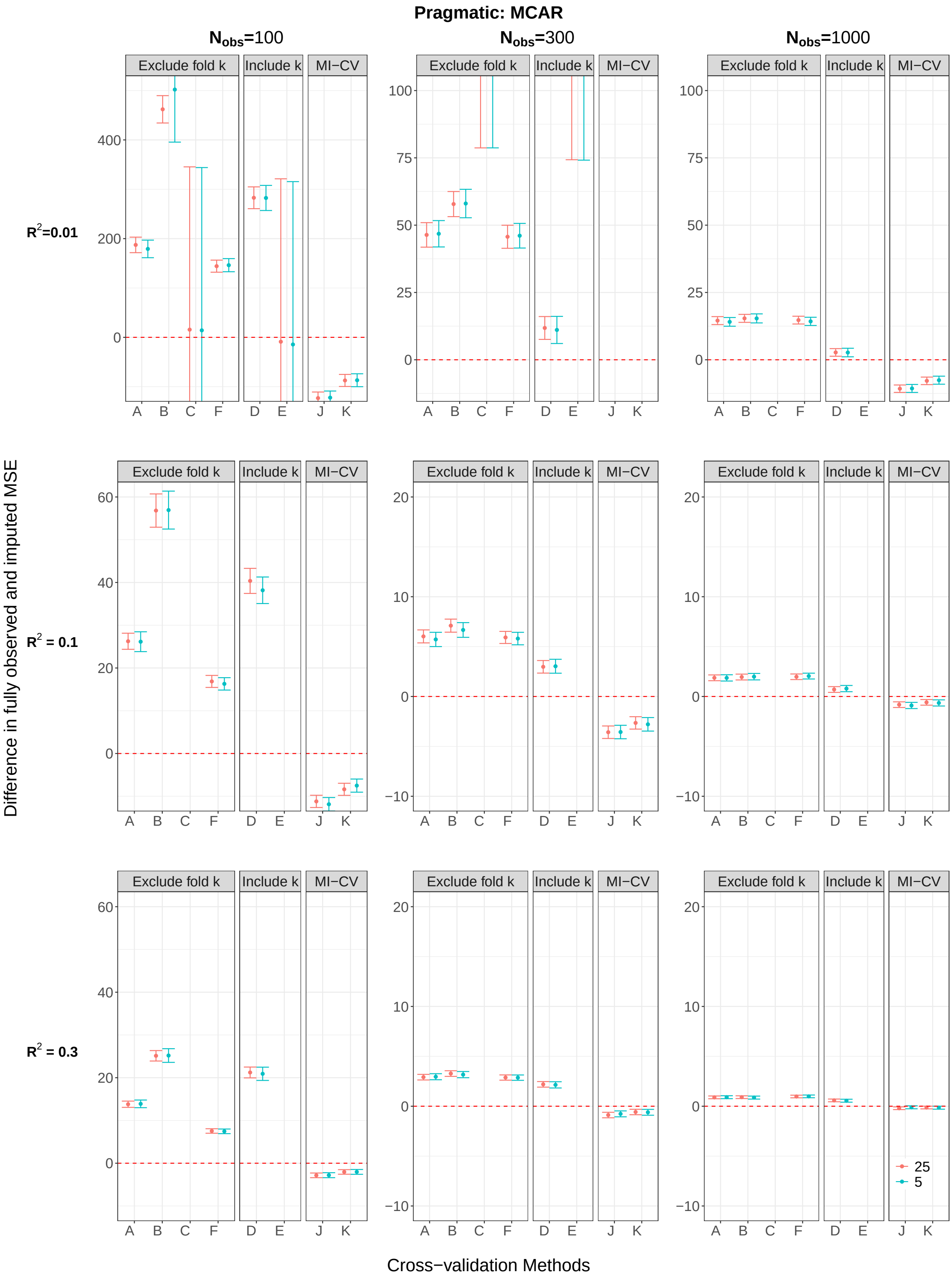

Figure S31: The difference MSE_{imp} - MSE_{obs} when data are MCAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

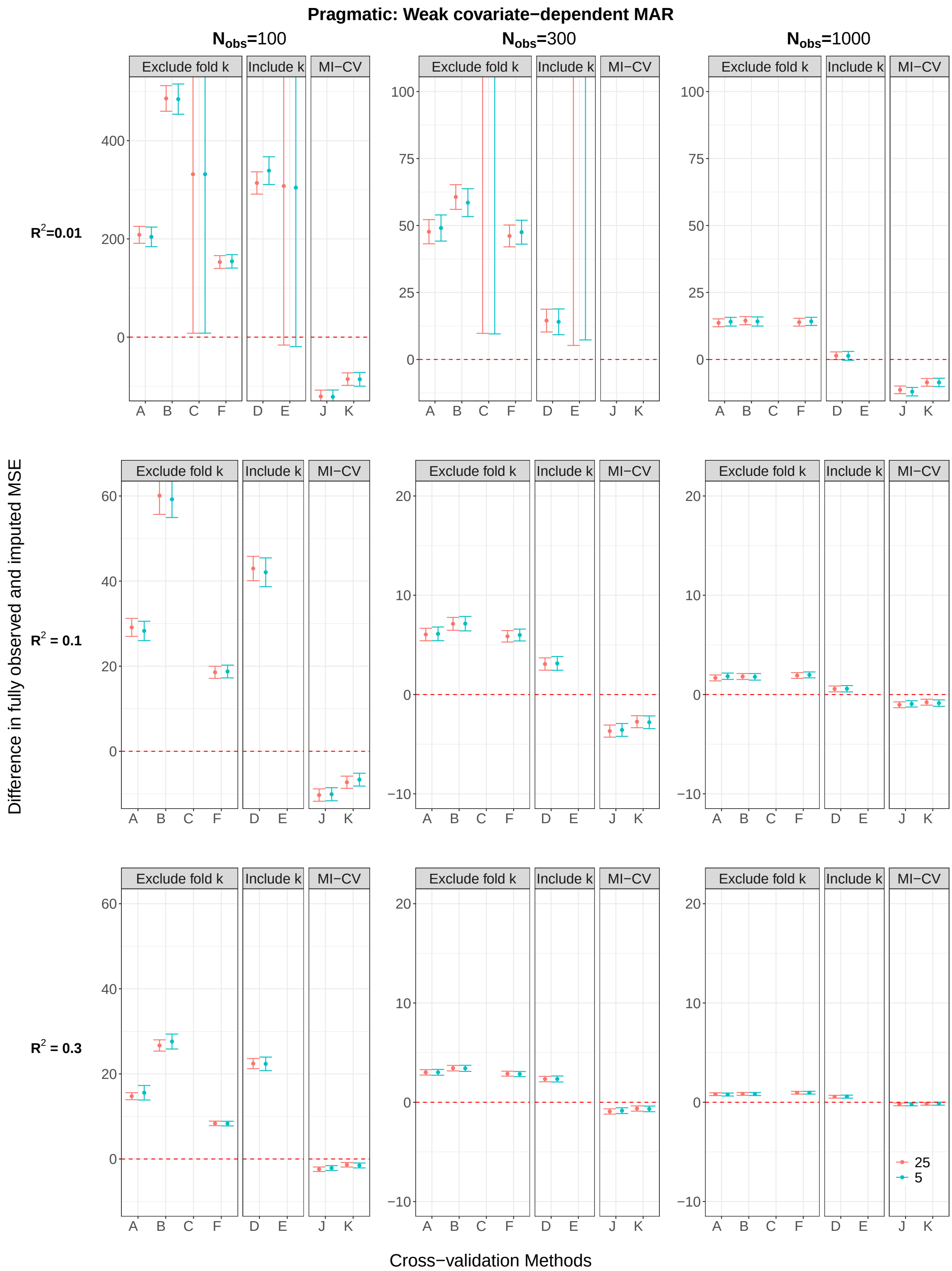

Figure S32: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

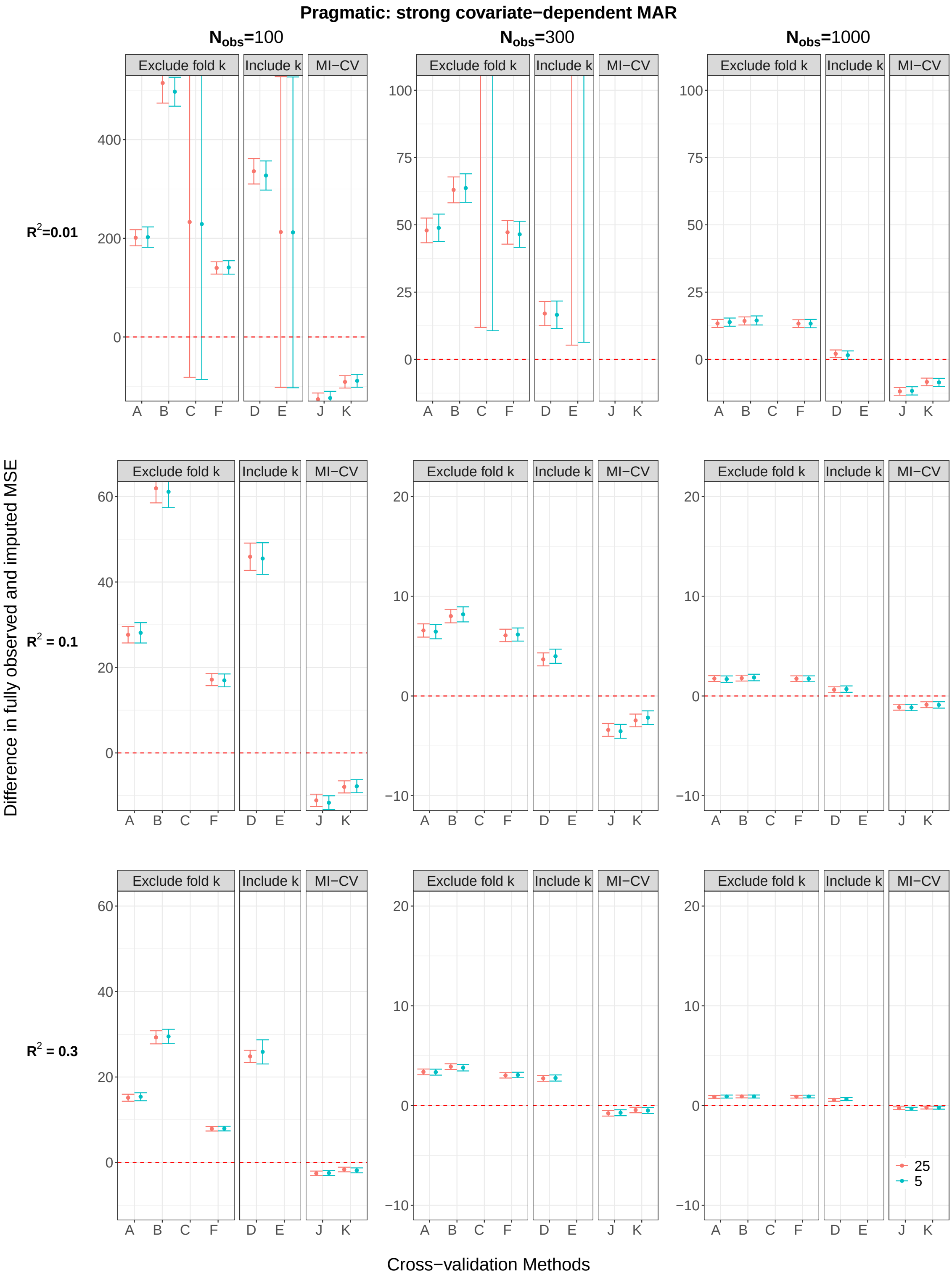

Figure S33: The difference MSE_{imp} - MSE_{obs} when data are strong covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

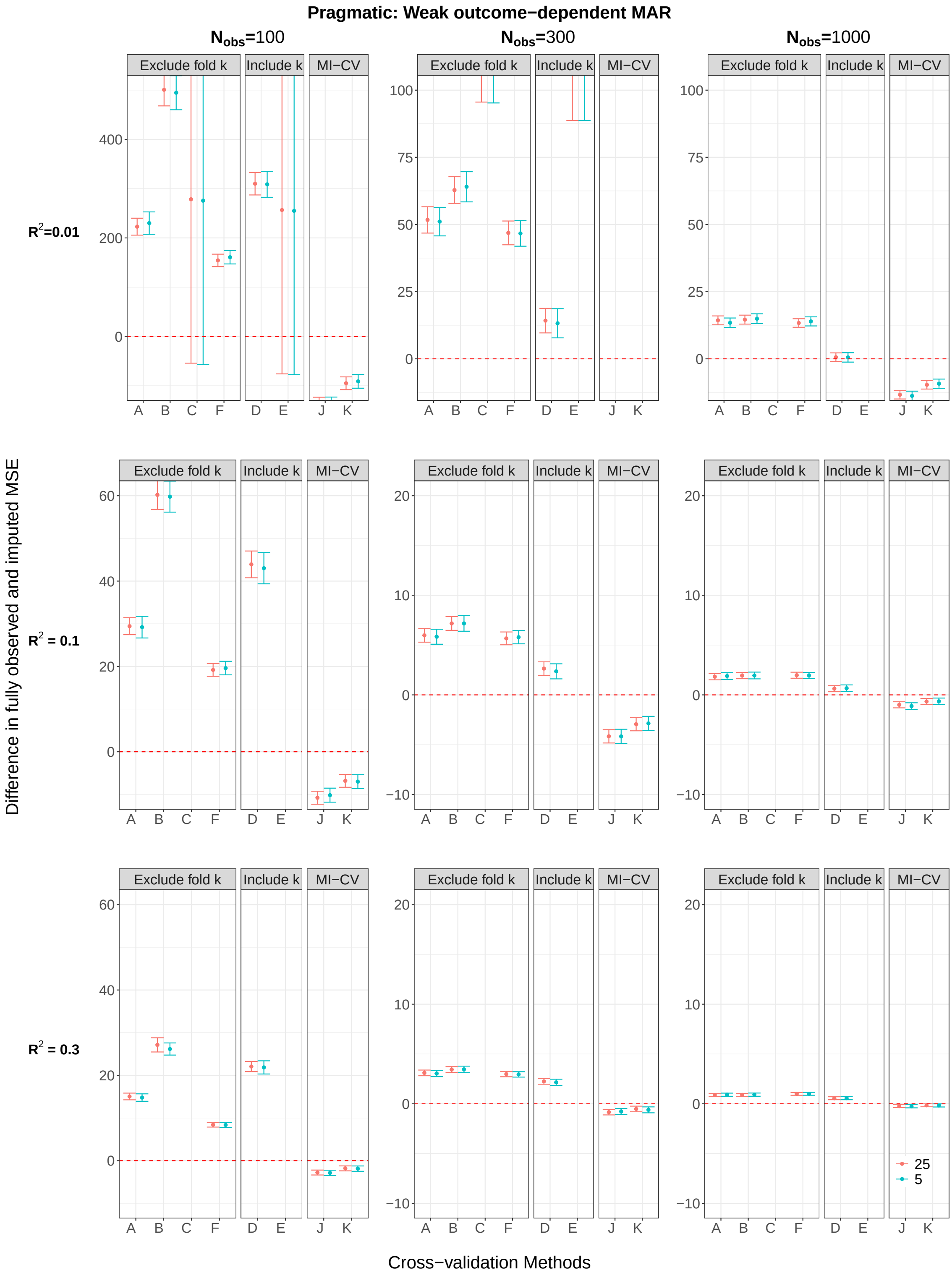

Figure S34: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

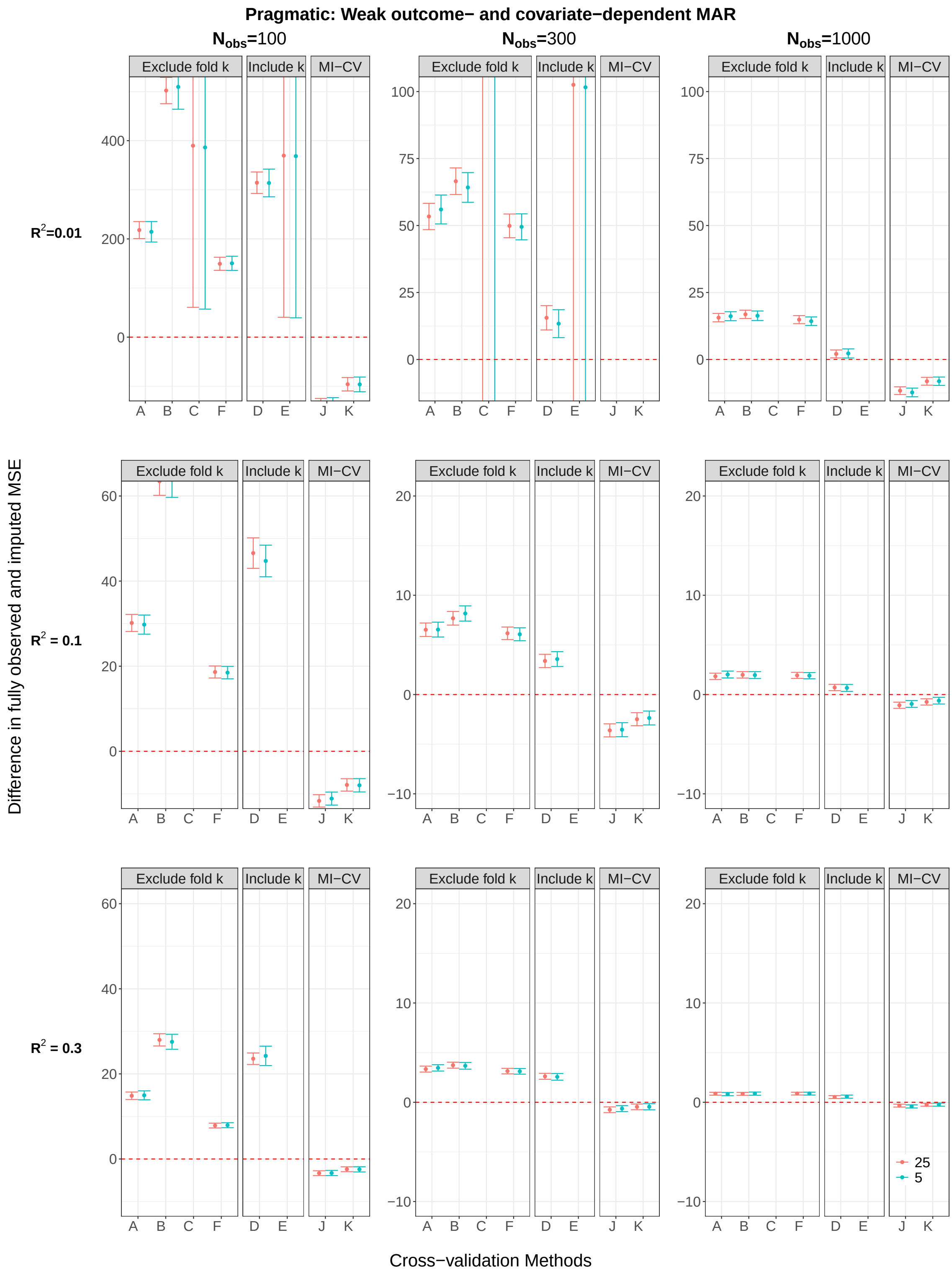

Figure S35: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

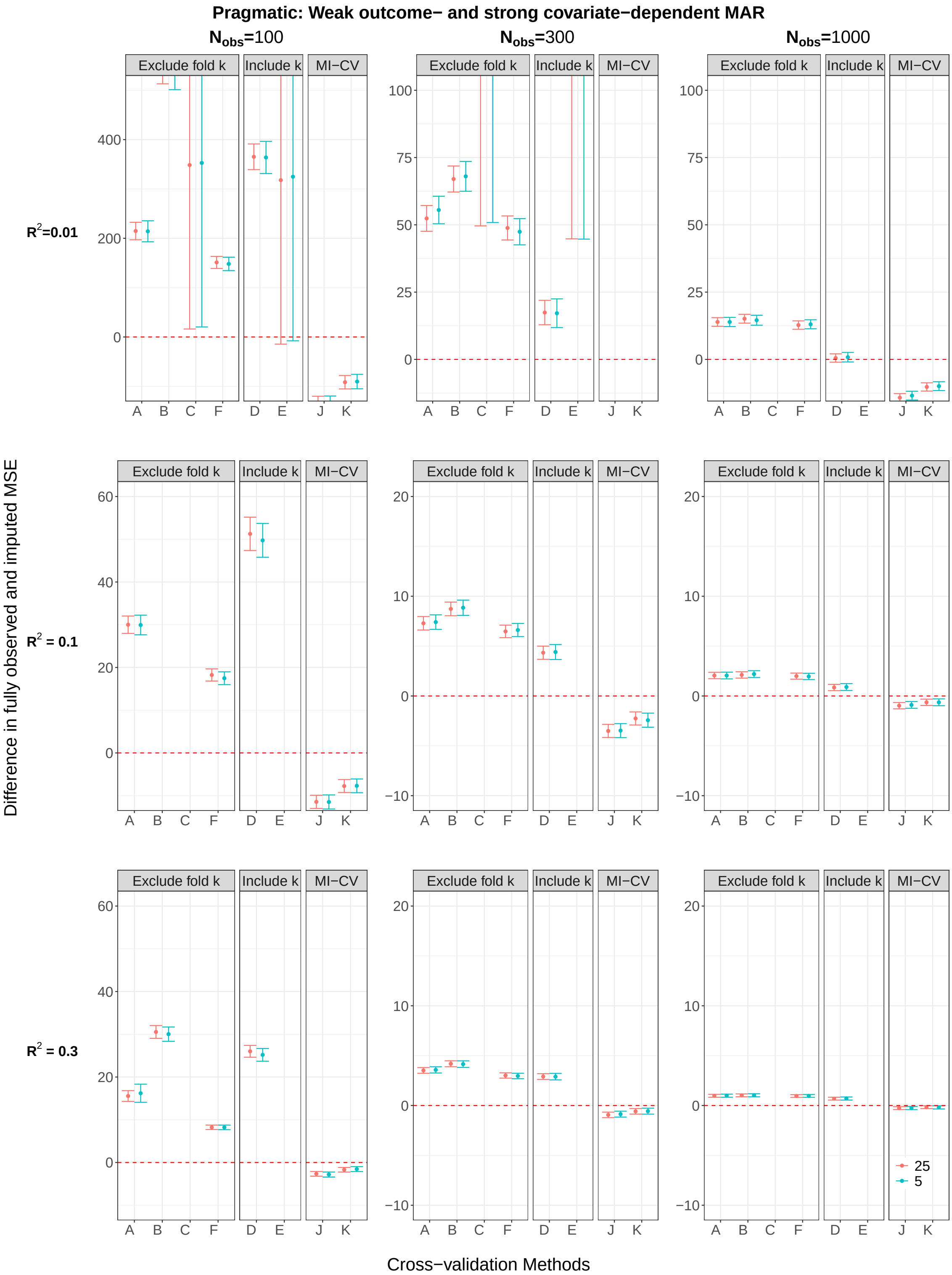

Figure S36: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and strong covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S1.1.5 MSE from imputation methods compared to the target MSE (MSE_{target}) using a larger validation set

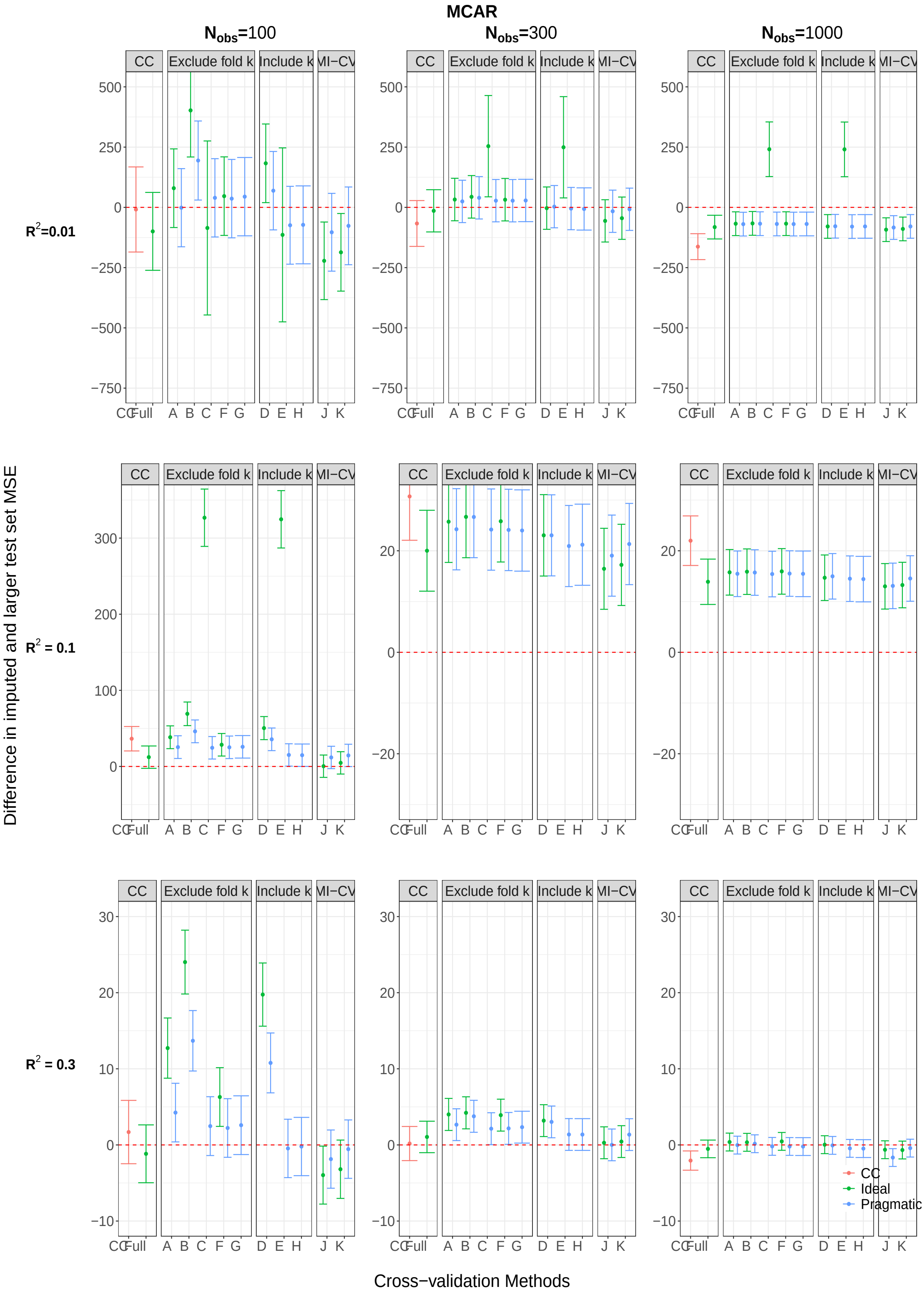

Figure S37: The difference MSE_{imp} - MSE_{target} when data are MCAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

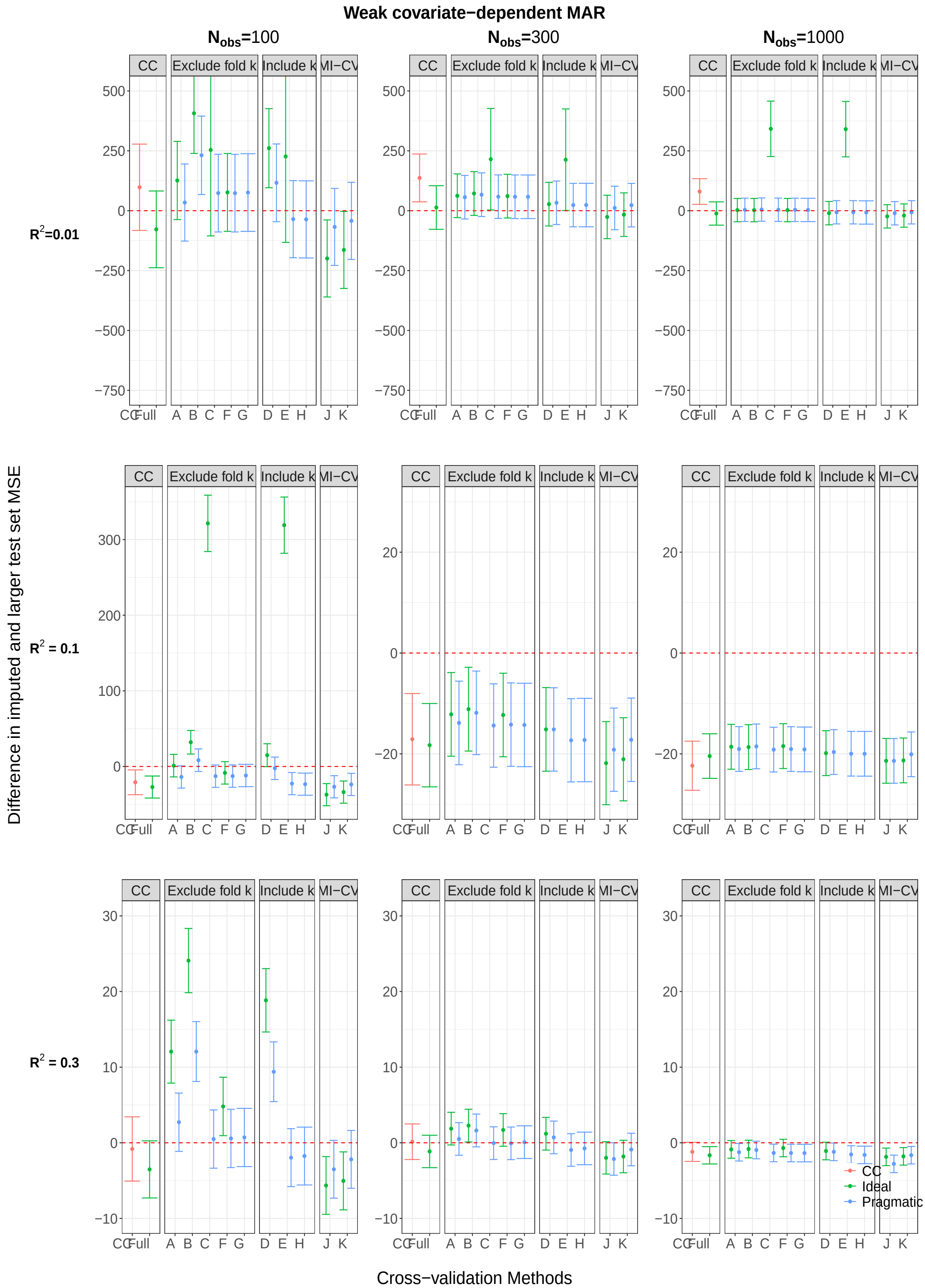

Figure S38: The difference MSE$_{imp}$ - MSE$_{target}$ when data are weak covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

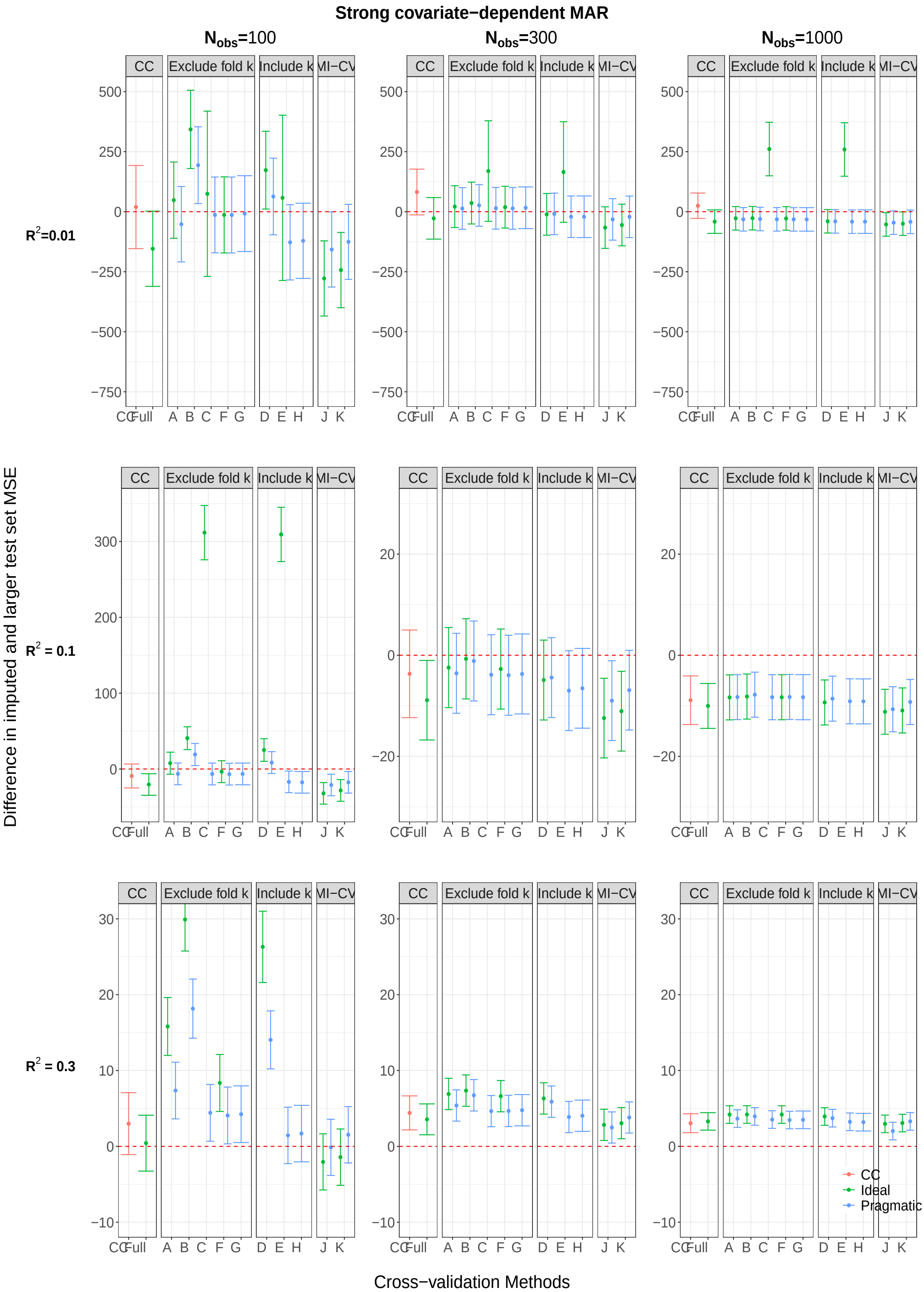

Figure S39: The difference MSE_{imp} - MSE_{target} when data are strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

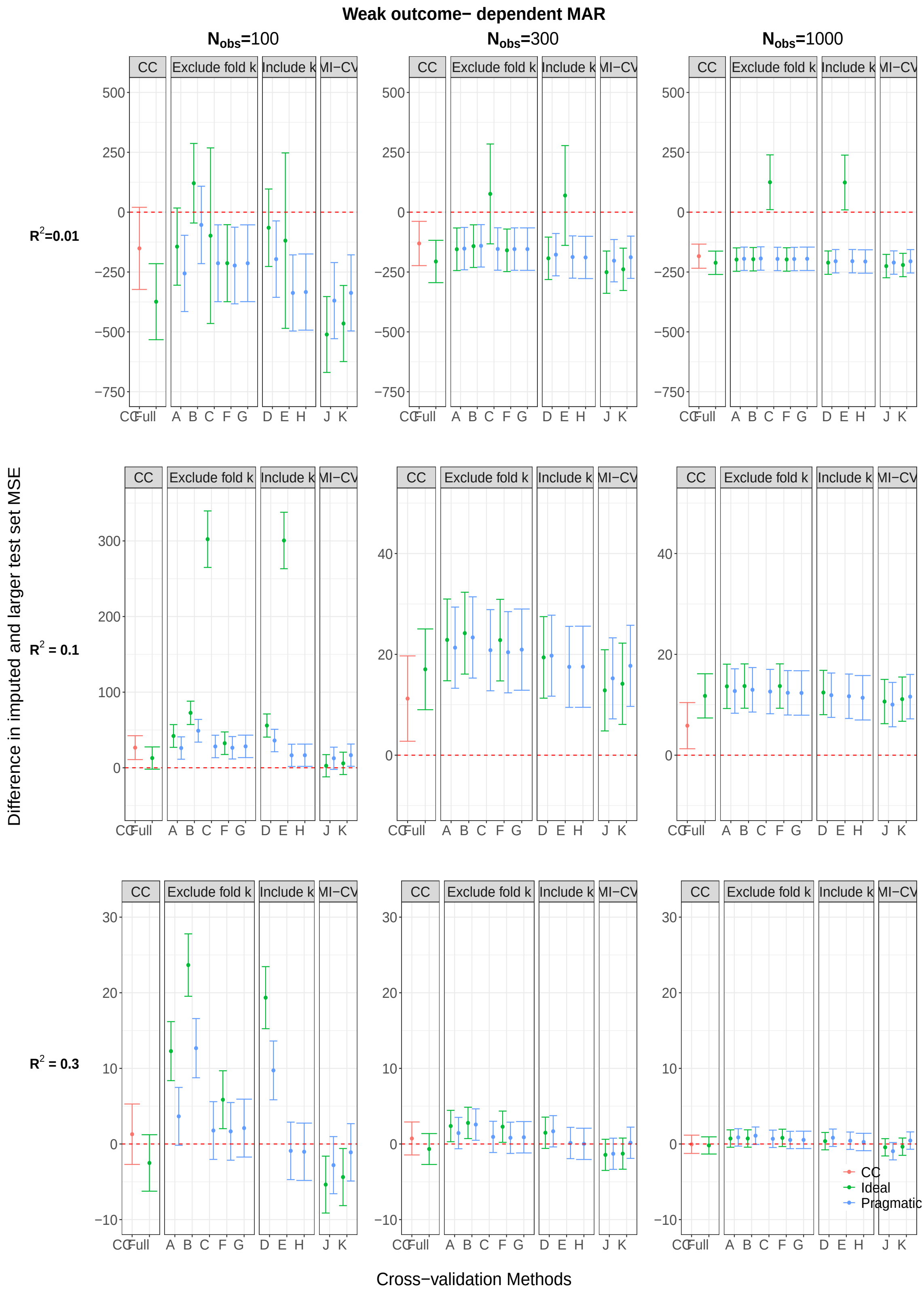

Figure S40: The difference MSE$_{imp}$ - MSE$_{target}$ when data are weak outcome-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

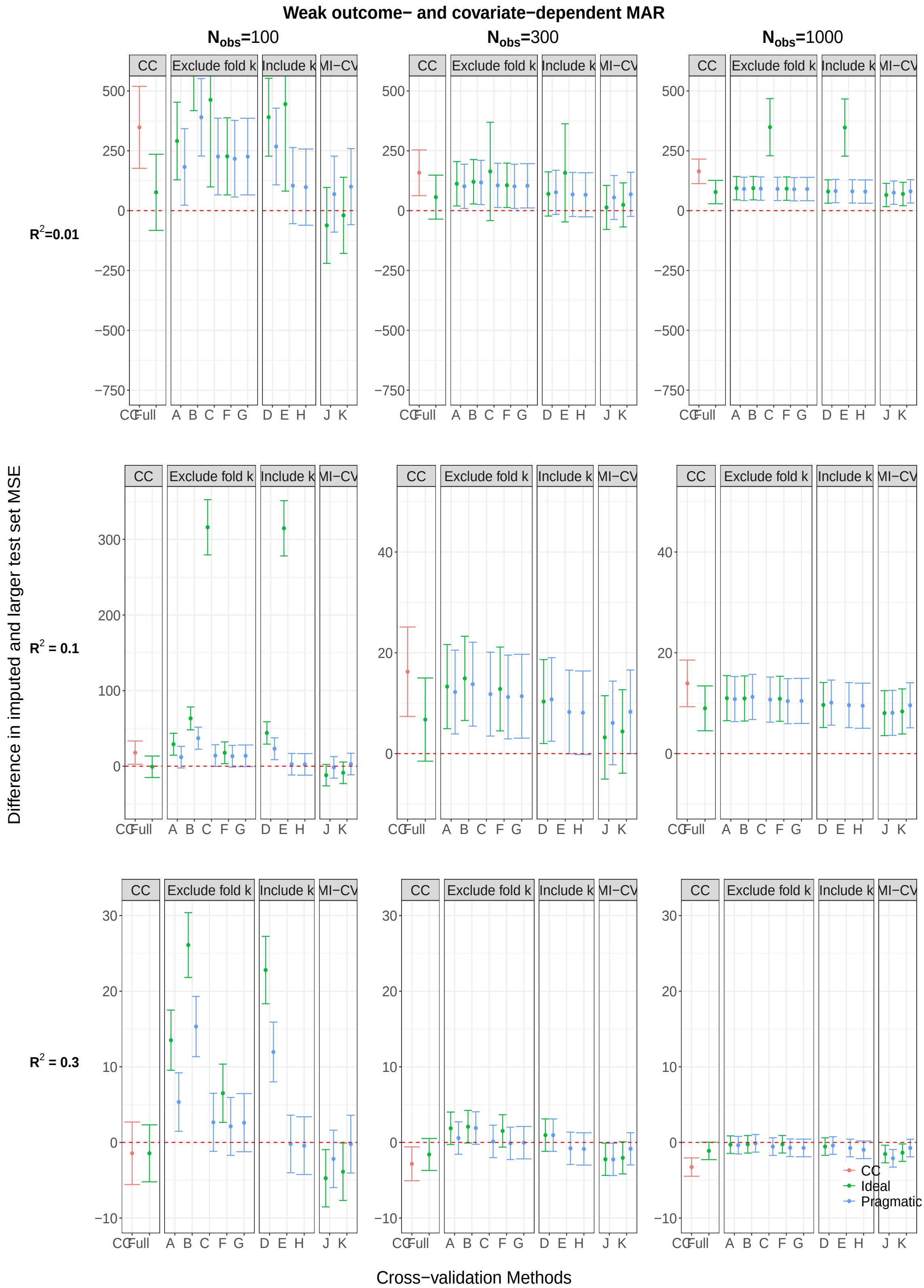

Figure S41: The difference MSE$_{imp}$ - MSE$_{target}$ when data are weak outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

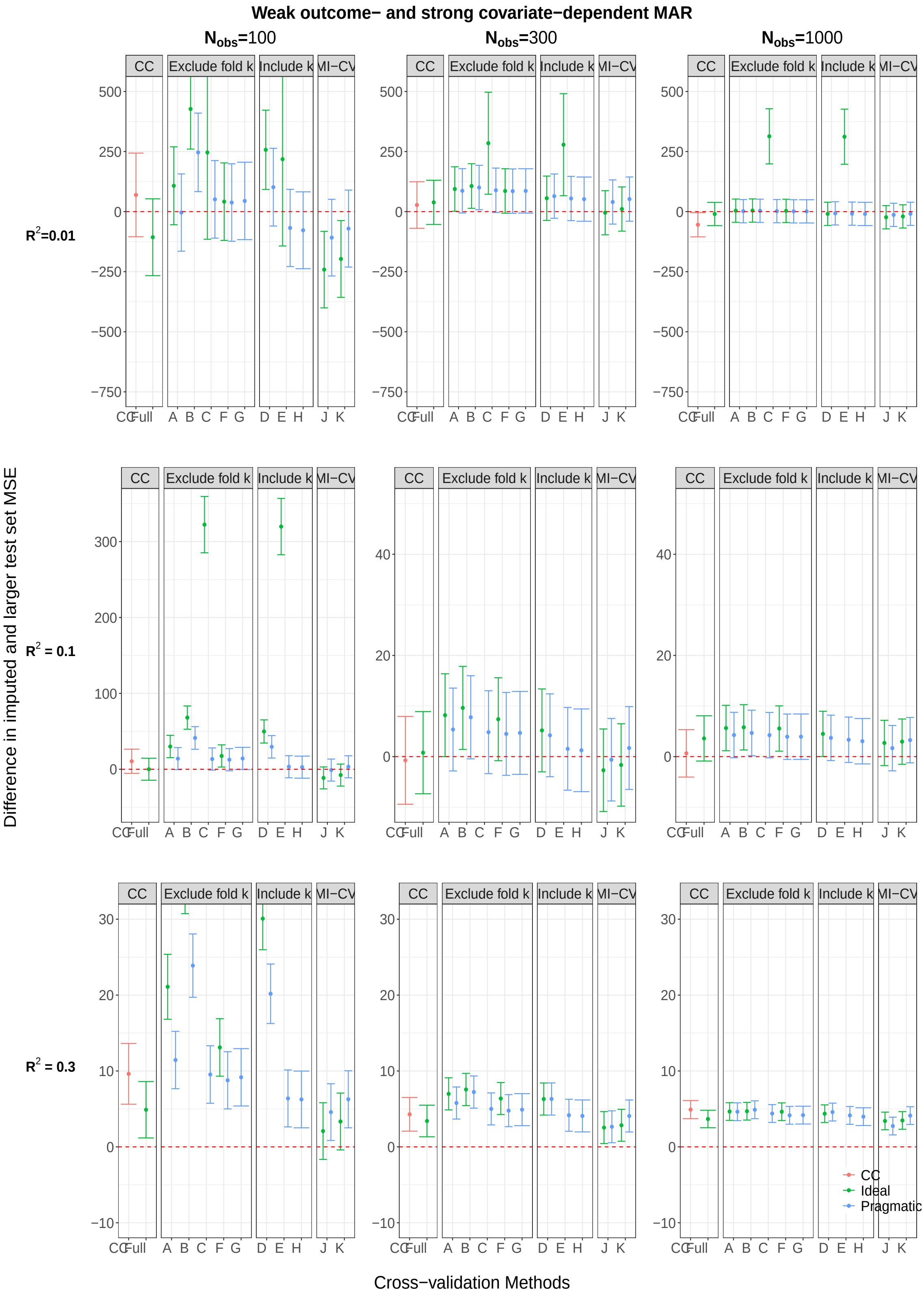

Figure S42: The difference MSE_{imp} - MSE_{target} when data are weak outcome- and strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S1.1.6 Comparing data leakage (MSE$_{imp}$ - MSE$_{target}$)

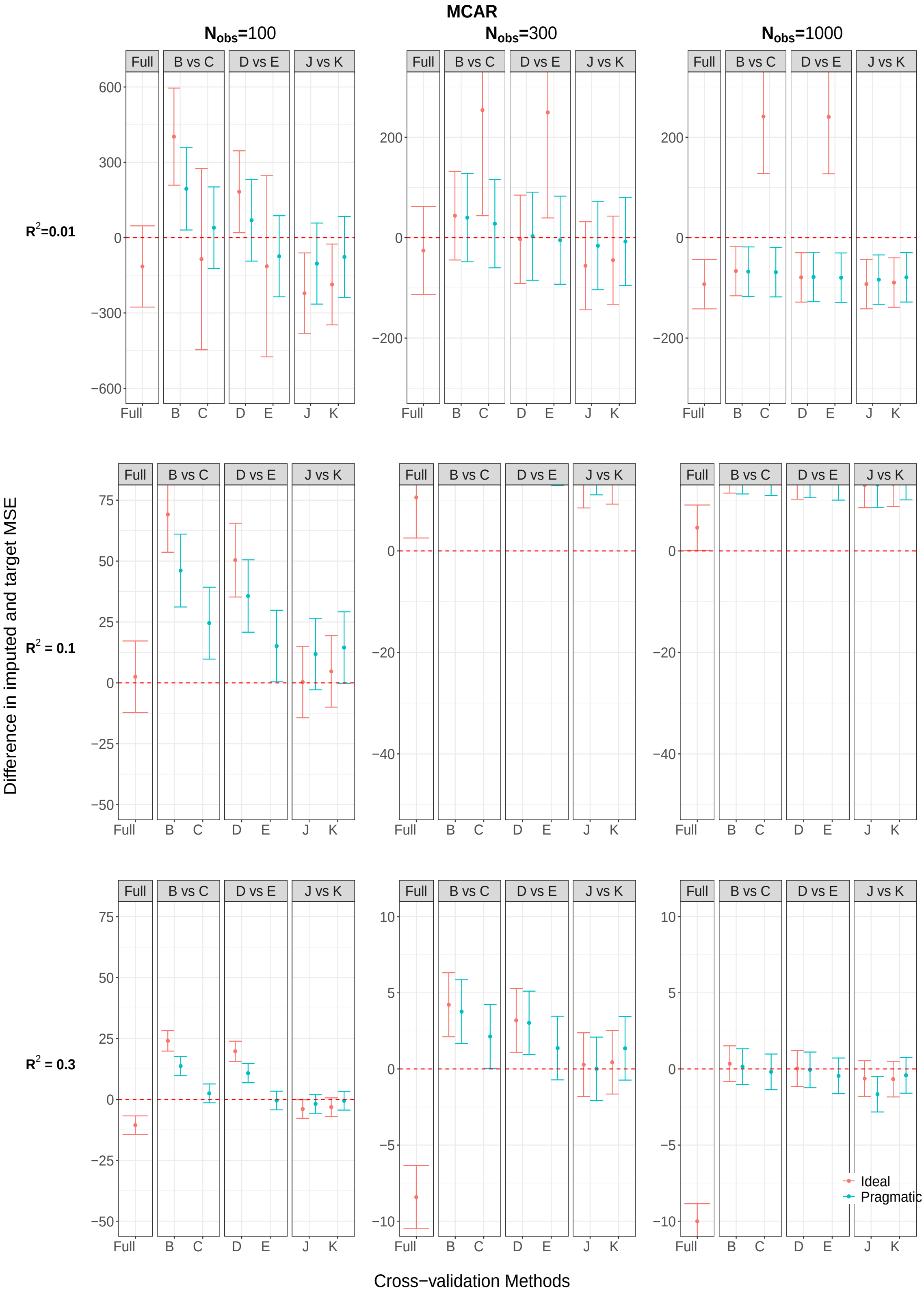

Figure S43: Assessing data leakage within the imputation process for cross-validation. The difference MSE$_{imp}$ - MSE$_{target}$ is compared when data are MCAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

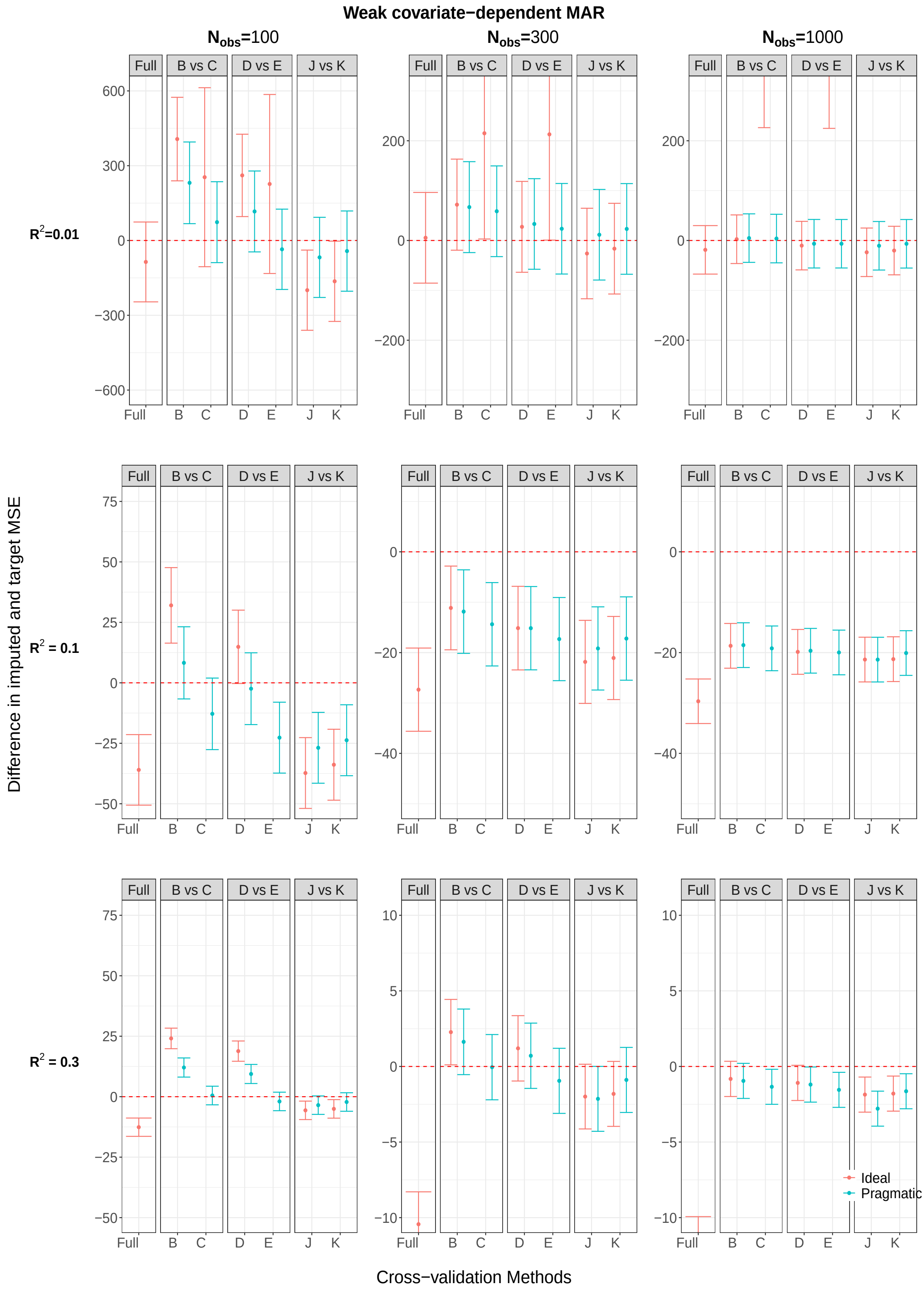

Figure S44: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{target} is compared when data are weak covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

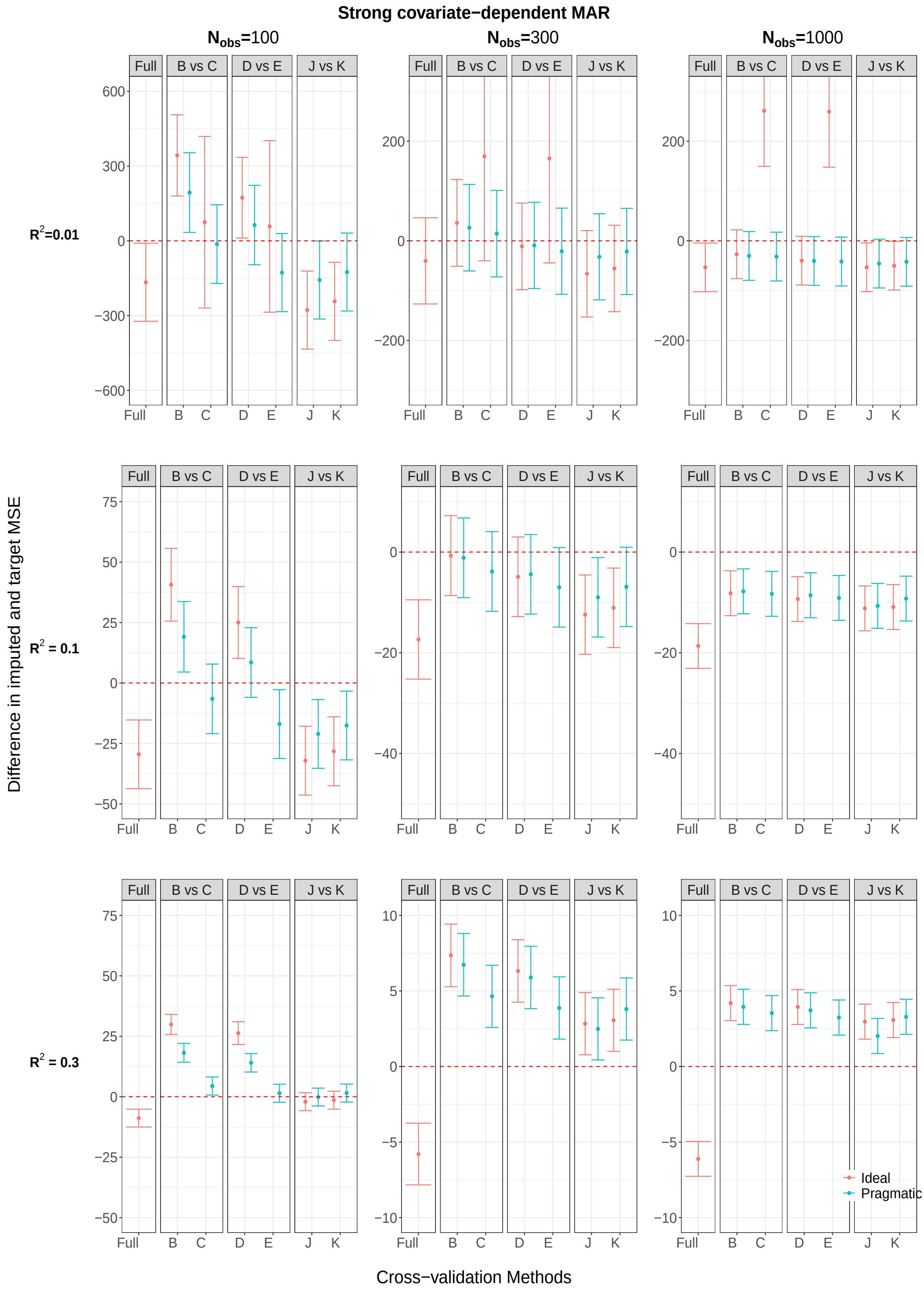

Figure S45: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{target} is compared when data are strong covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

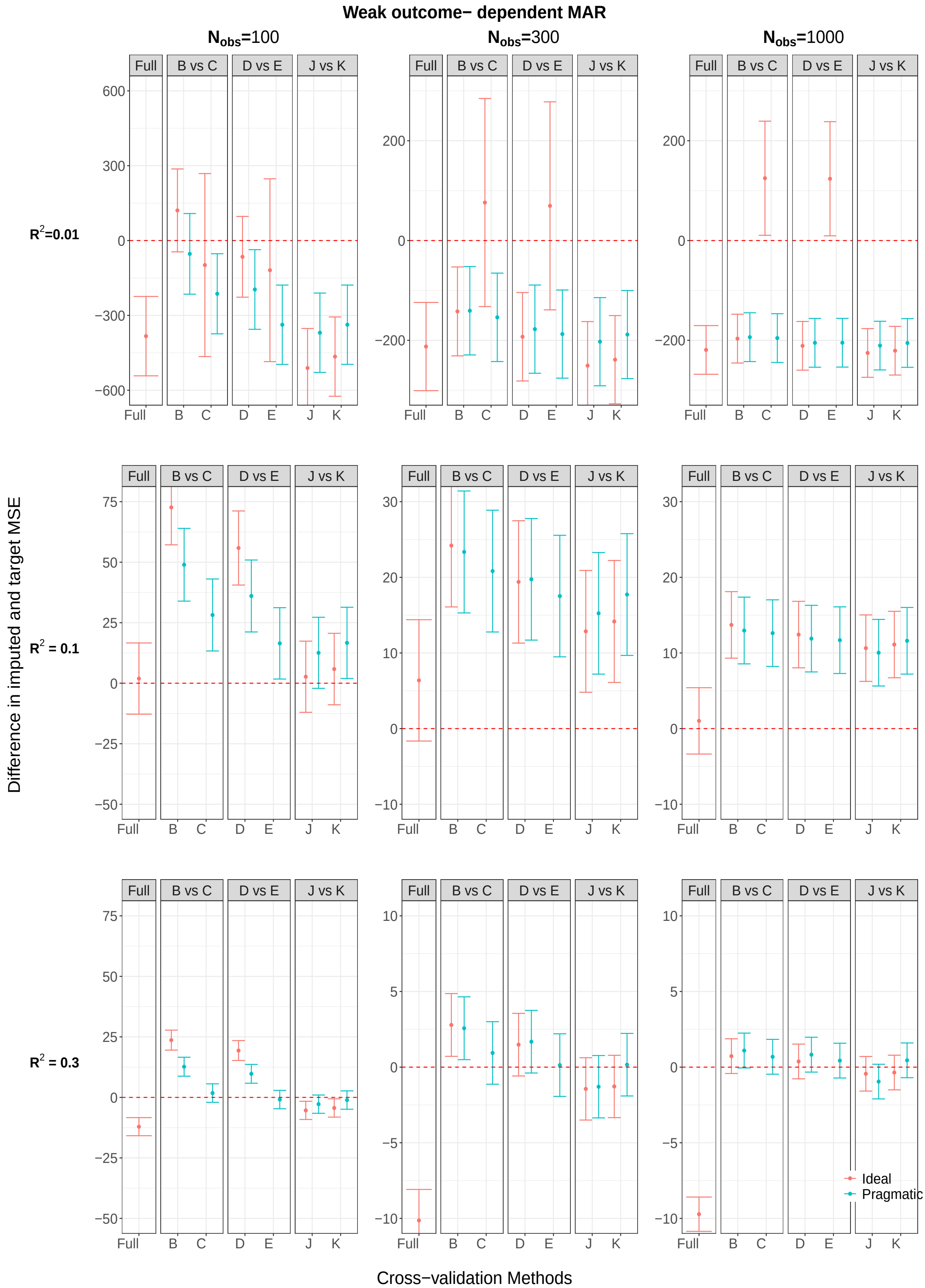

Figure S46: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{target} is compared when data are weak outcome-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

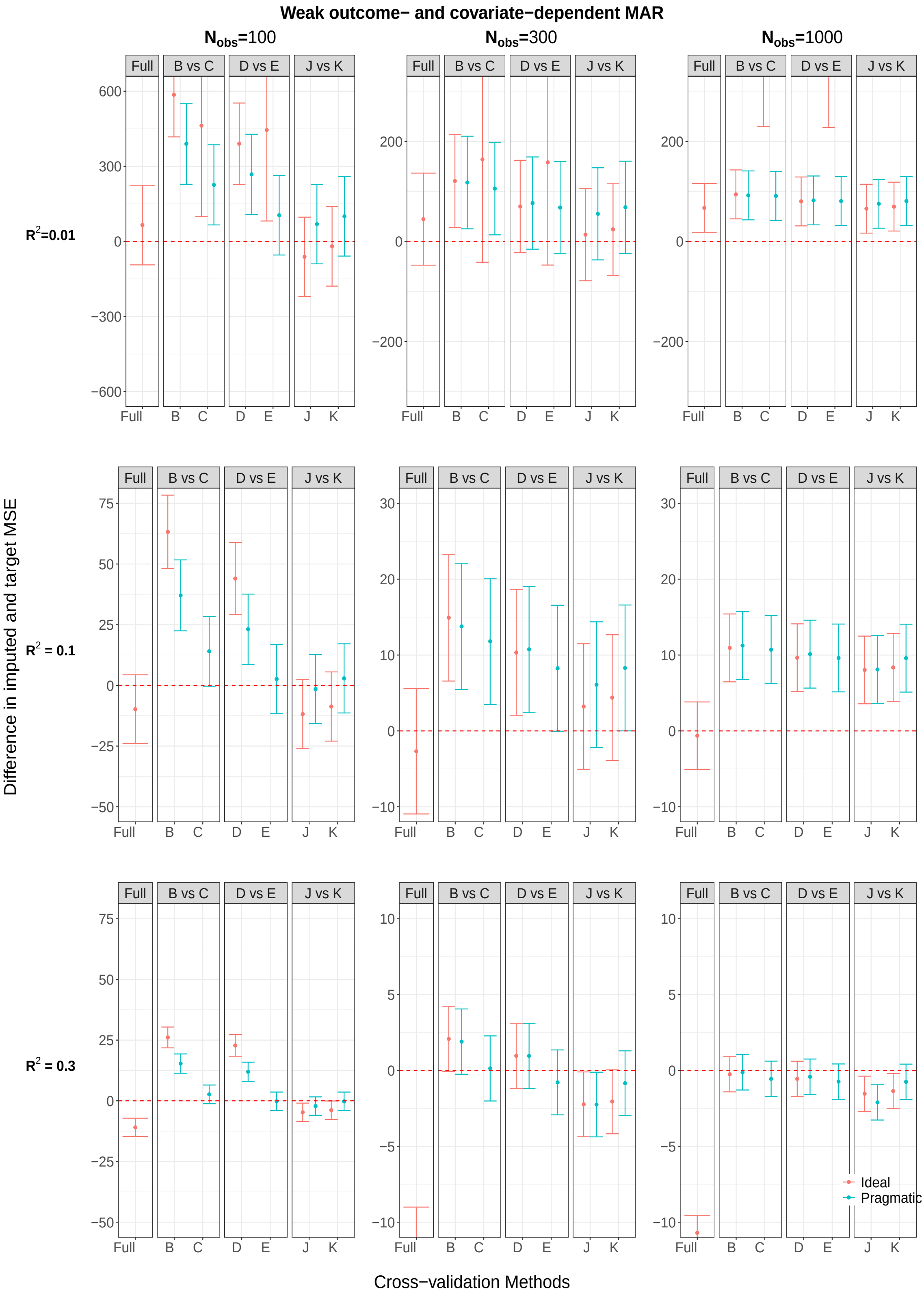

Figure S47: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{target} is compared when data are weak outcome- and covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

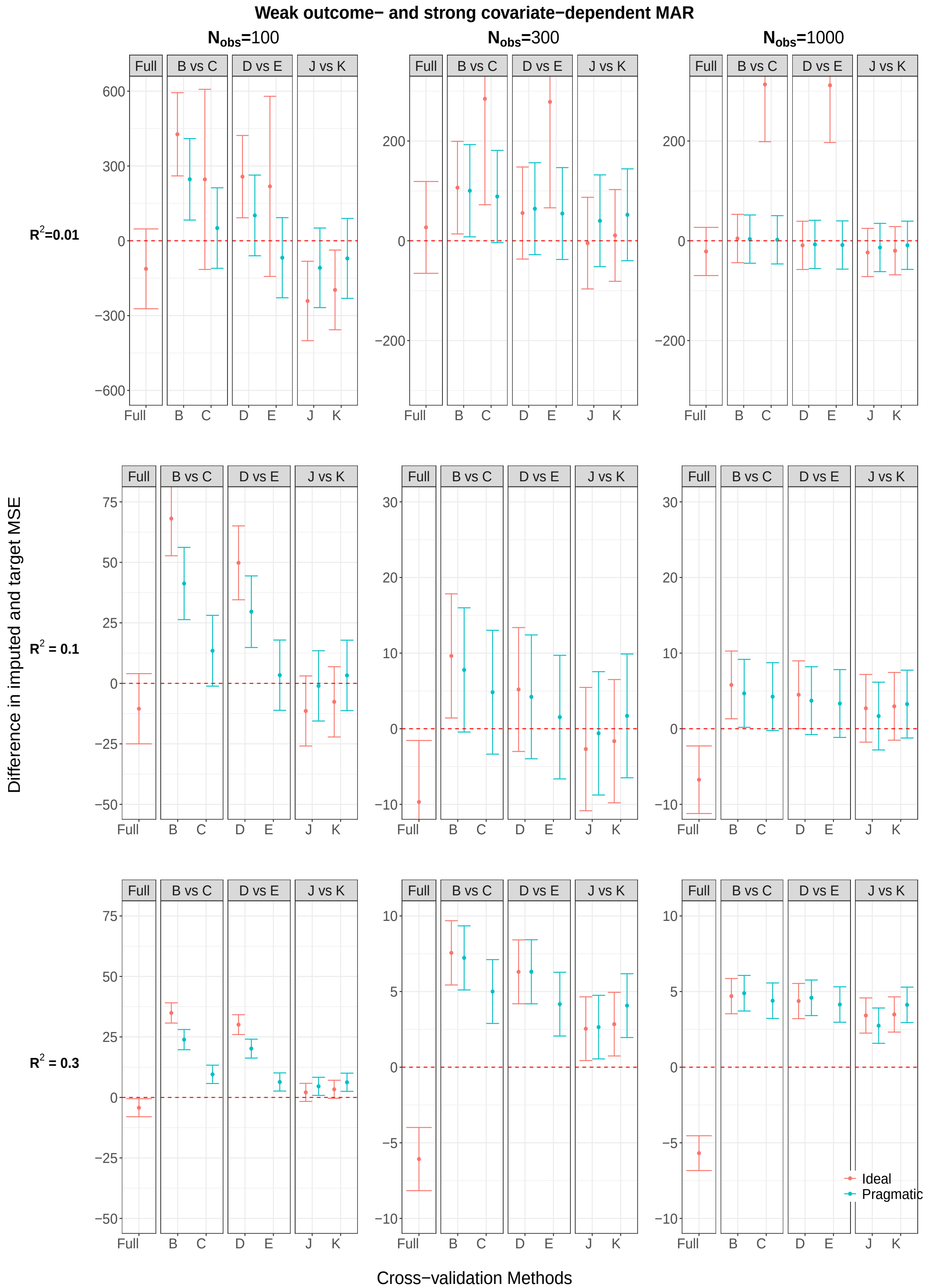

Figure S48: Assessing data leakage within the imputation process for cross-validation. The difference MSE_{imp} - MSE_{target} is compared when data are weak outcome- and strong covariate-dependent MAR. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2 Chapter 4: Cross-validation and MI for the binary outcome

S2.1 AUC

S2.1.1 AUC from imputation methods compared to the fully-observed AUC (AUC_{imp}-AUC_{obs})

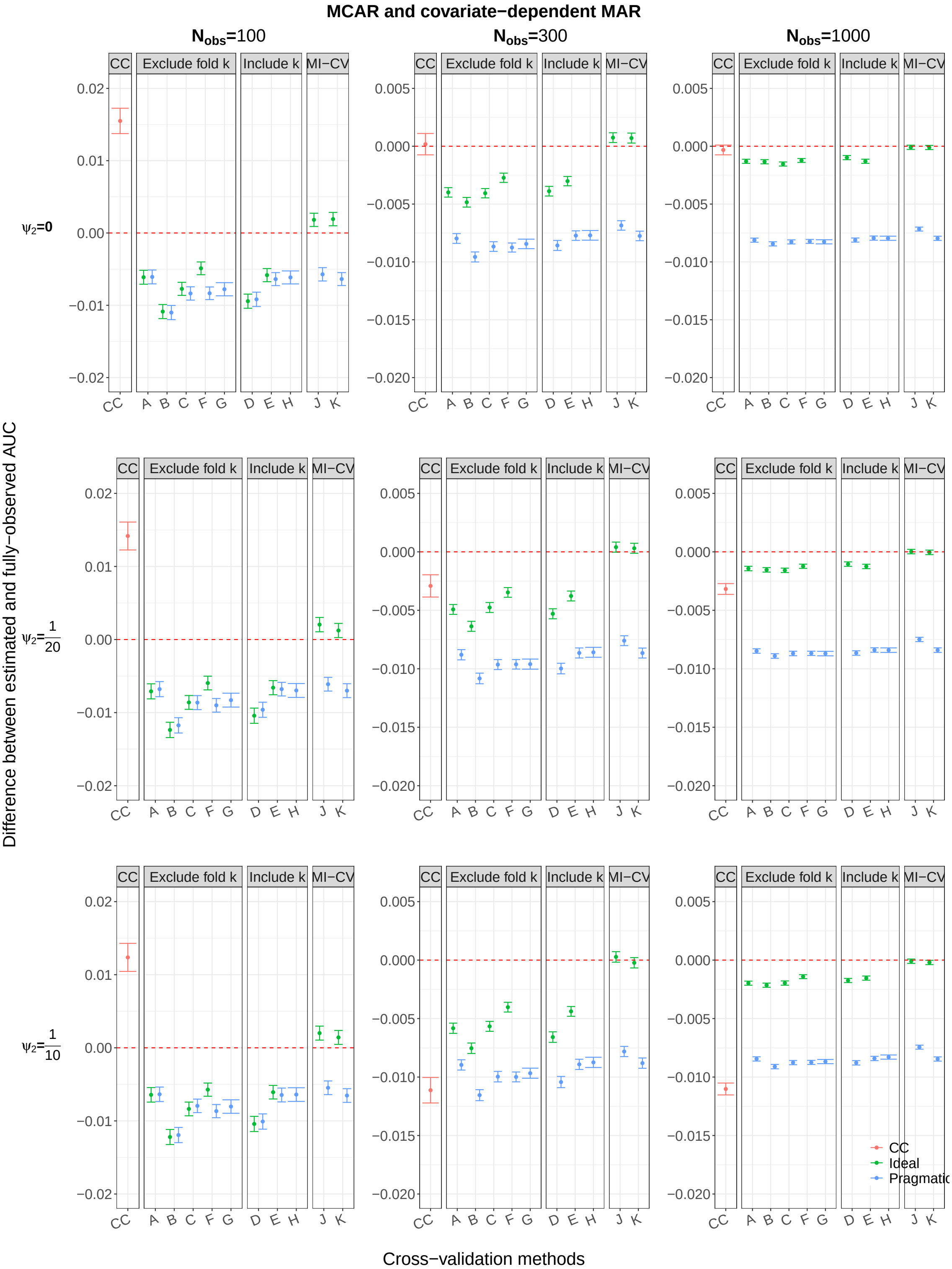

Figure S1: The difference AUC_{imp} - AUC_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. The average AUC when data are fully observed is 0.78. CC (complete case), methods A-K are

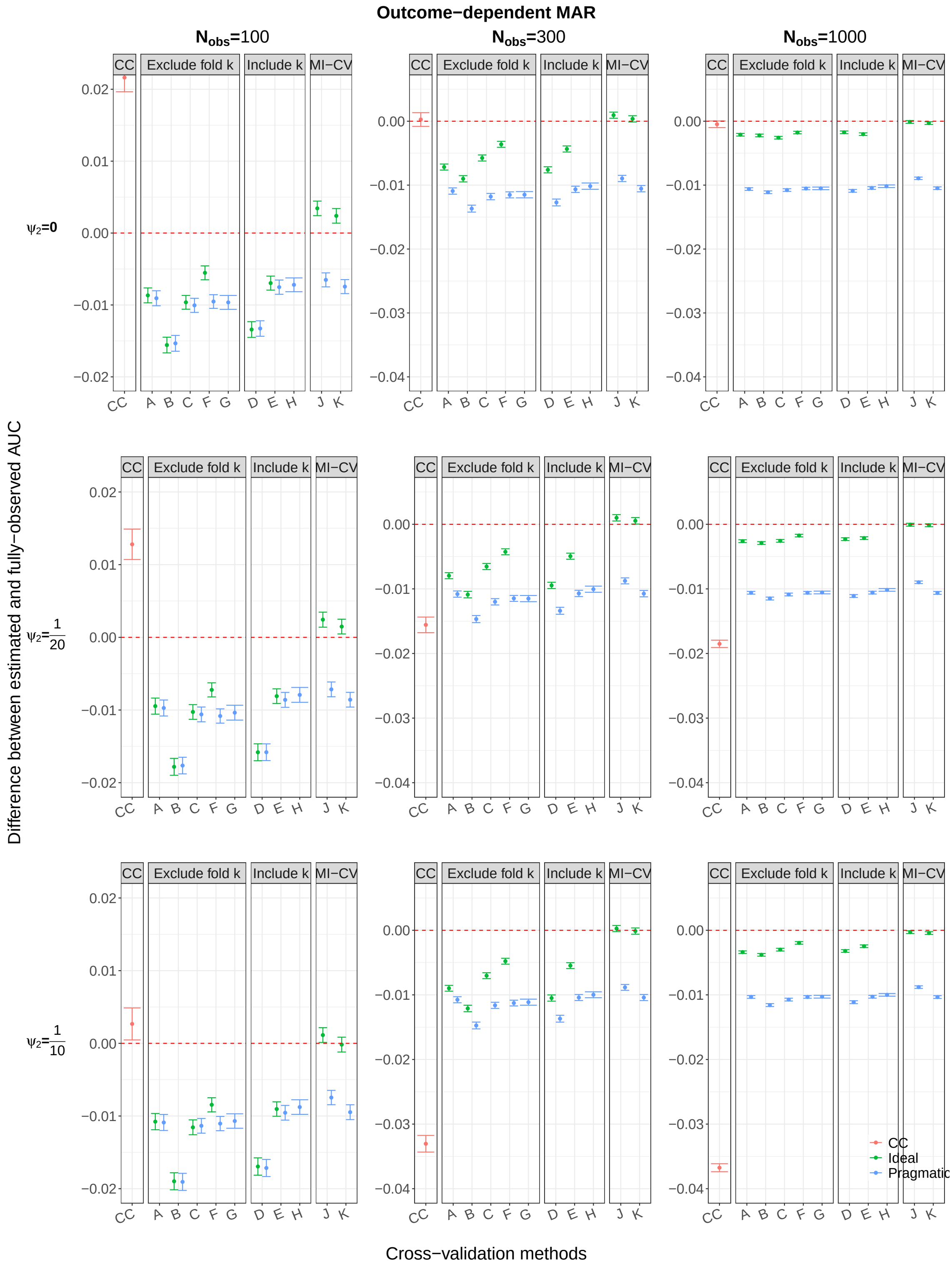

Figure S2: The difference AUC$_{imp}$ - AUC$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{obs}$. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.1.2 The proportion of missingness is 40% (AUC_{imp}-AUC_{obs})

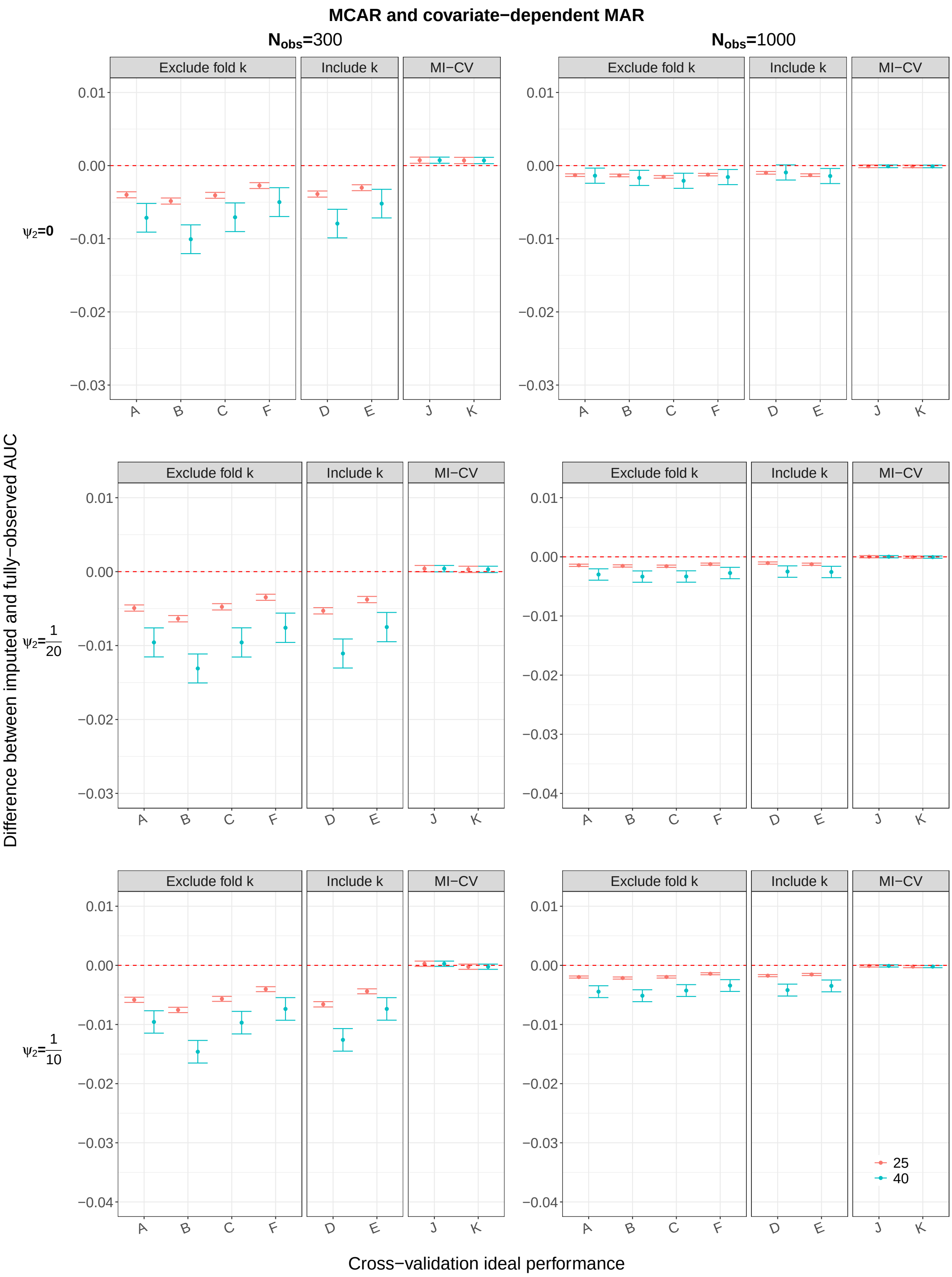

Figure S3: Comparing the impact of increasing the percentage of missingness on the difference AUC_{imp} - AUC_{obs} when data are MCAR and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. Red denotes AUC_{imp} - AUC_{obs} when 25% of X_1 values are missing and blue denotes AUC_{imp} - AUC_{obs} when 40% of X_1 values are missing. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3.

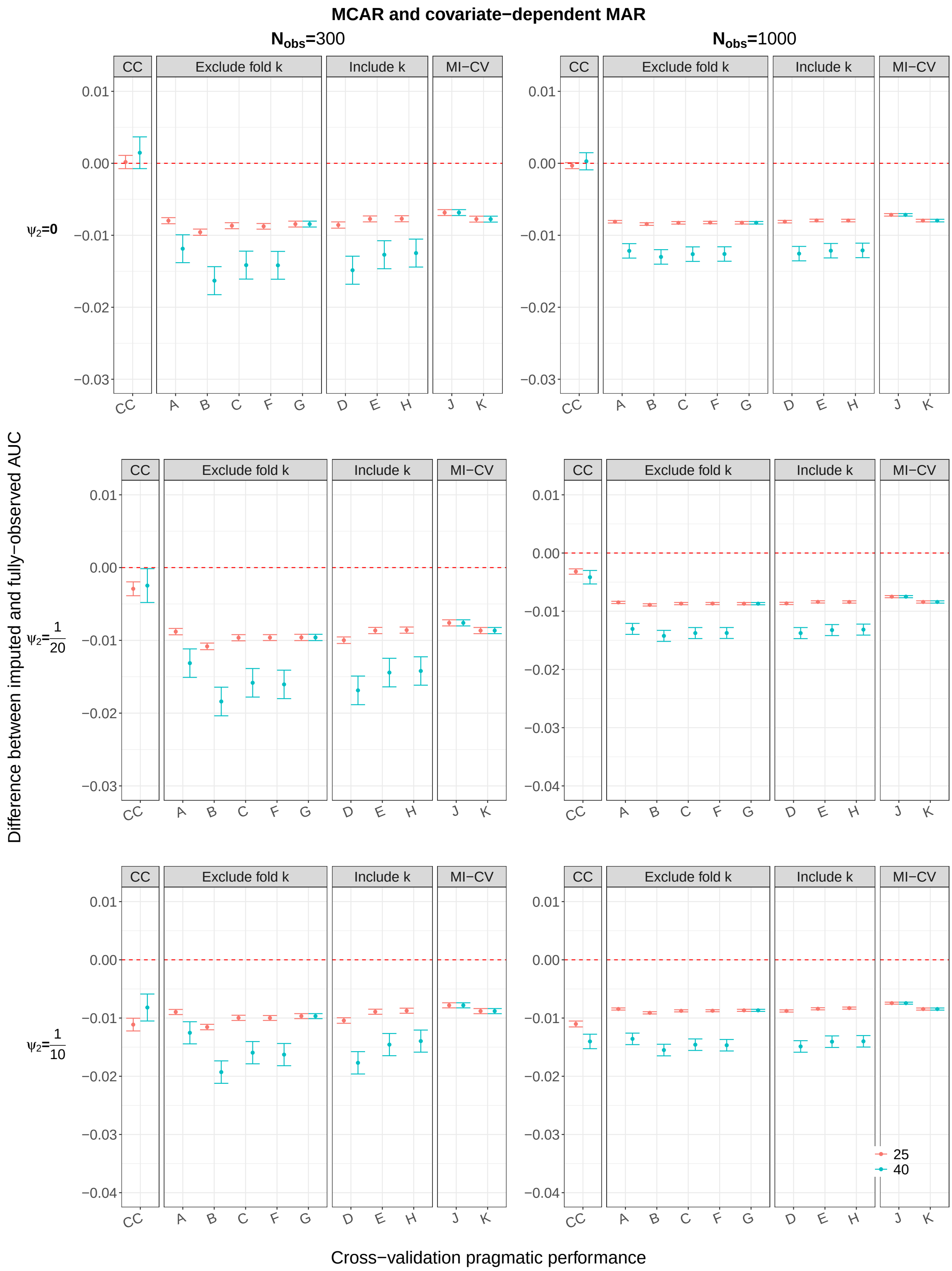

Figure S4: Comparing the impact of increasing the percentage of missingness on the difference AUC_{imp} - AUC_{obs} when data are MCAR and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. Red denotes AUC_{imp} - AUC_{obs} when 25% of X_1 values are missing and blue denotes AUC_{imp} - AUC_{obs} when 40% of X_1 values are missing. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3.

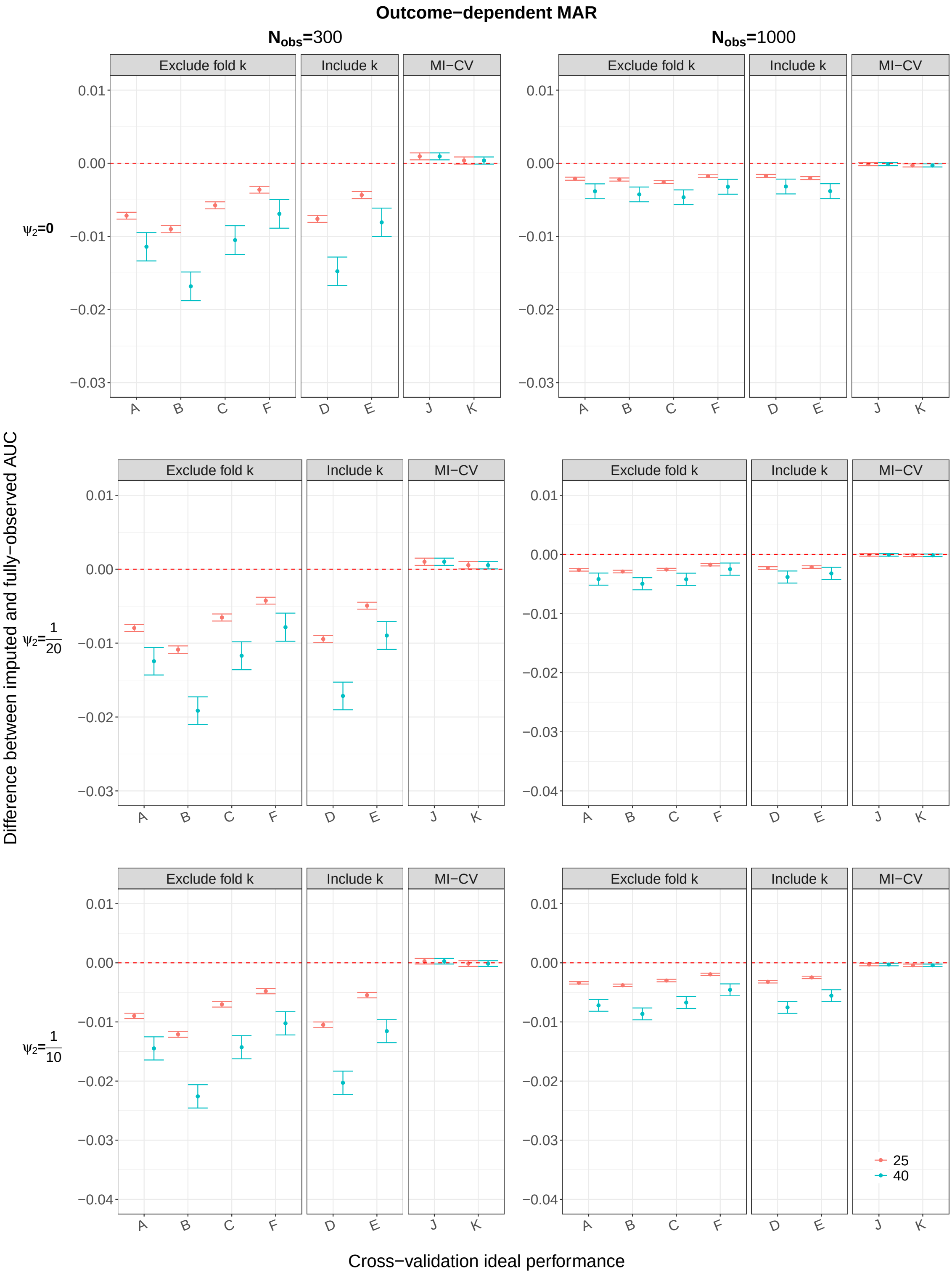

Figure S5: Comparing the impact of increasing the percentage of missingness on the difference AUC$_{imp}$ - AUC$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{obs}$. Red denotes AUC$_{imp}$ - AUC$_{obs}$ when 25% of X_1 values are missing and blue denotes AUC$_{imp}$ - AUC$_{obs}$ when 40% of X_1 values are missing. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3.

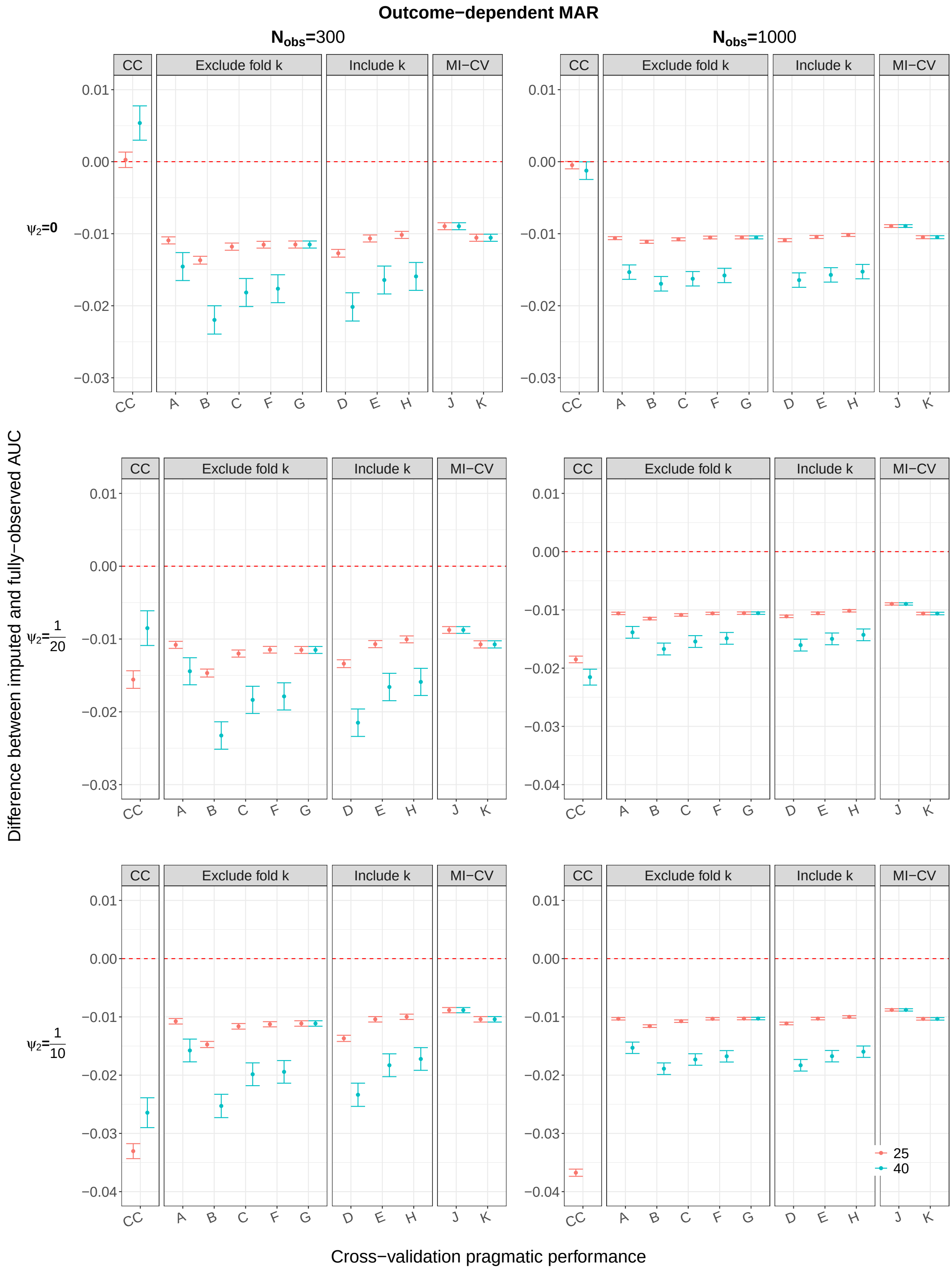

Figure S6: Comparing the impact of increasing the percentage of missingness on the difference AUC$_{imp}$ - AUC$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{obs}$. Red denotes AUC$_{imp}$ - AUC$_{obs}$ when 25% of X_1 values are missing and blue denotes AUC$_{imp}$ - AUC$_{obs}$ when 40% of X_1 values are missing. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3.

S2.1.3 Comparing M=5 versus M=25 (AUC_{imp}-AUC_{obs})

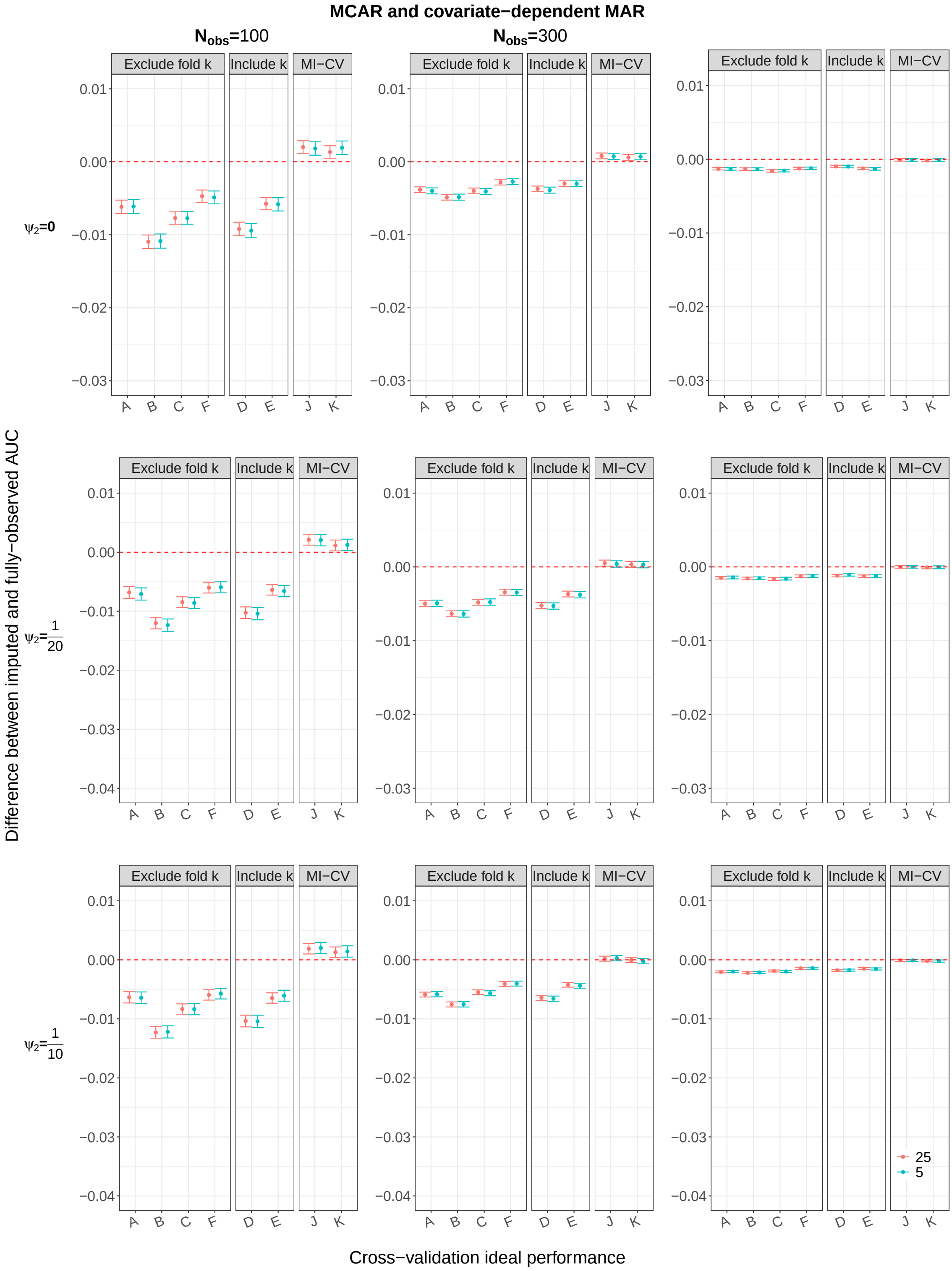

Figure S7: The difference AUC_{imp} - AUC_{obs} when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

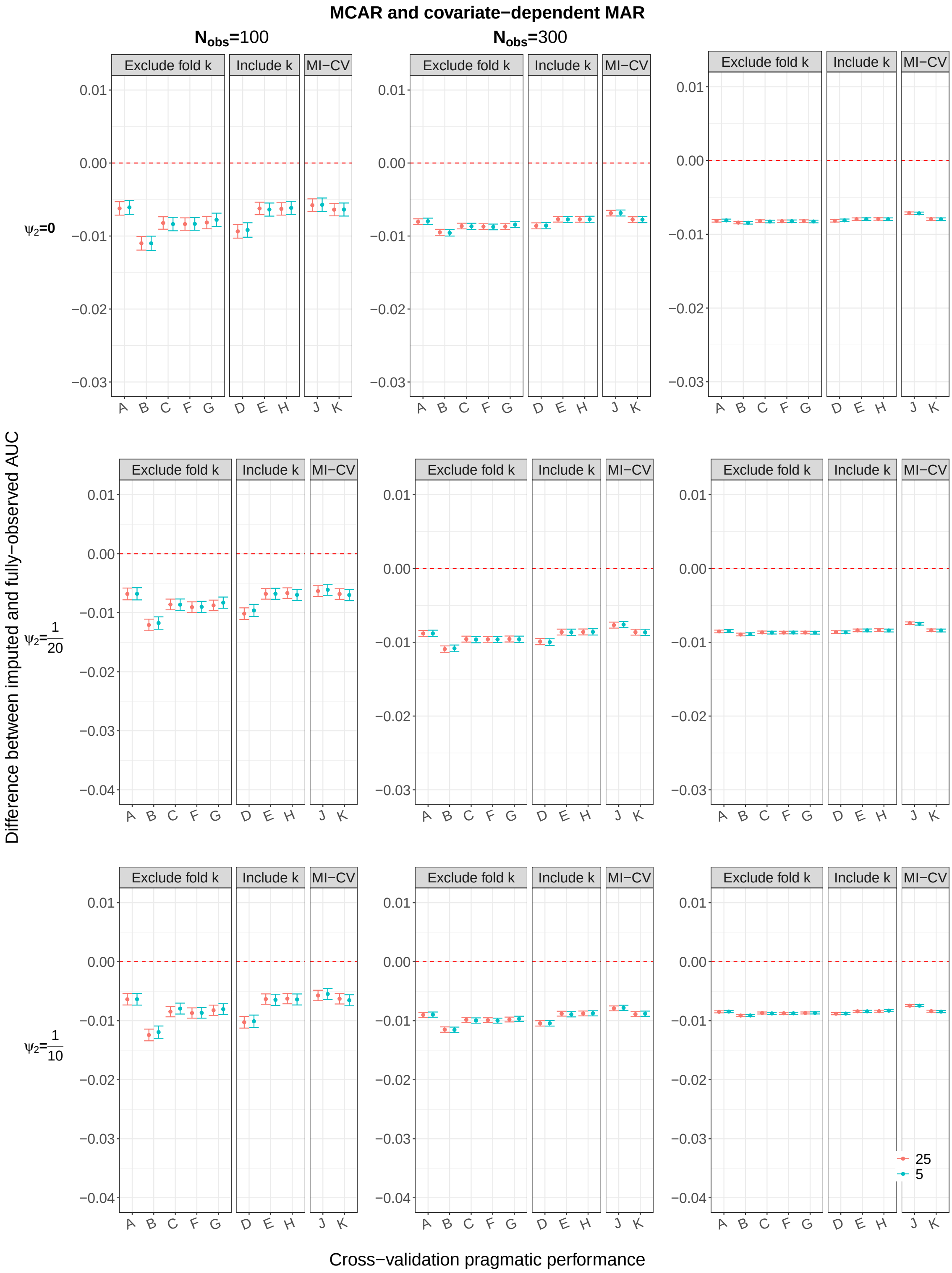

Figure S8: The difference AUC$_{imp}$ - AUC$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{obs}$. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

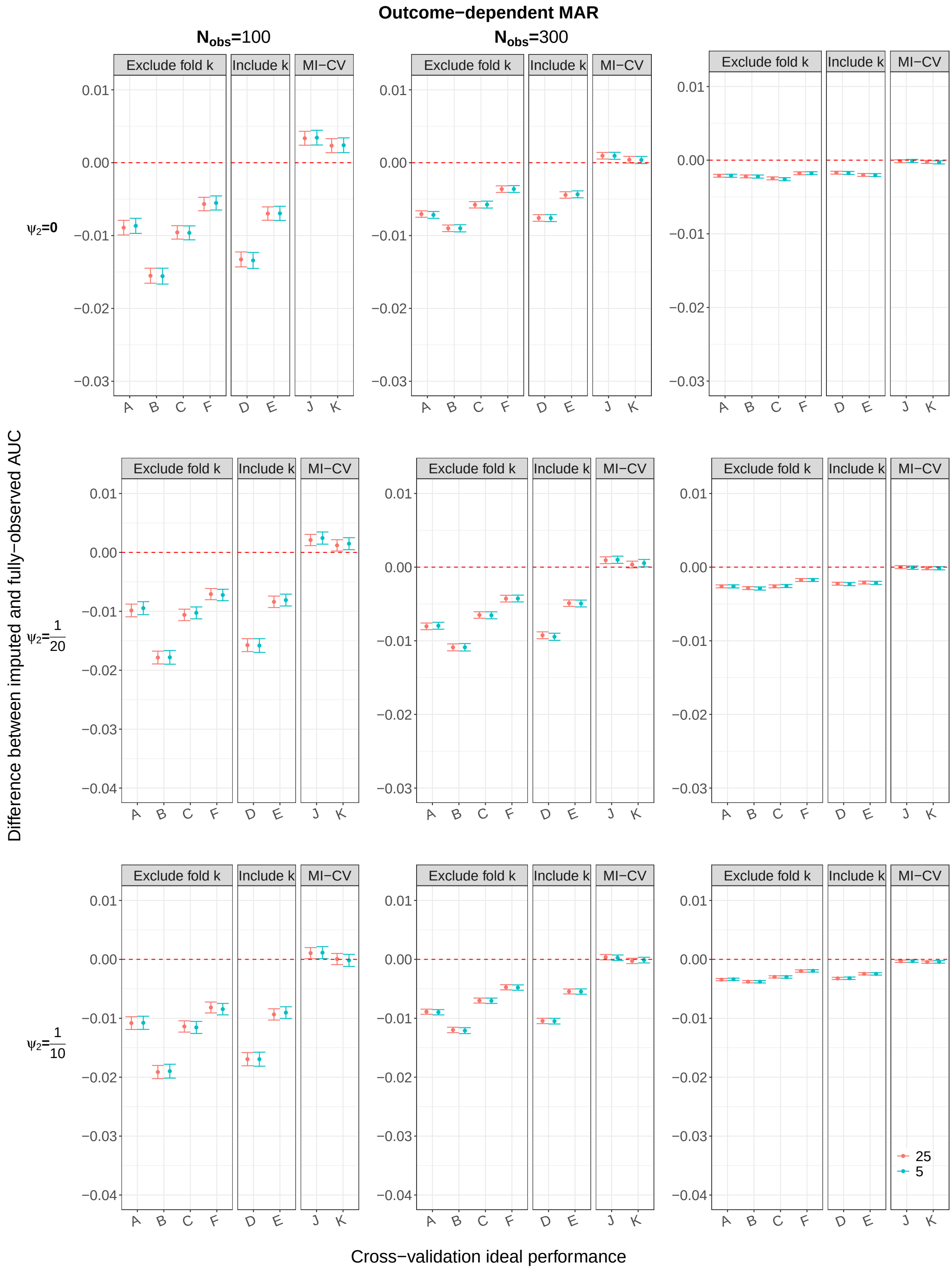

Figure S9: The difference AUC_{imp} - AUC_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

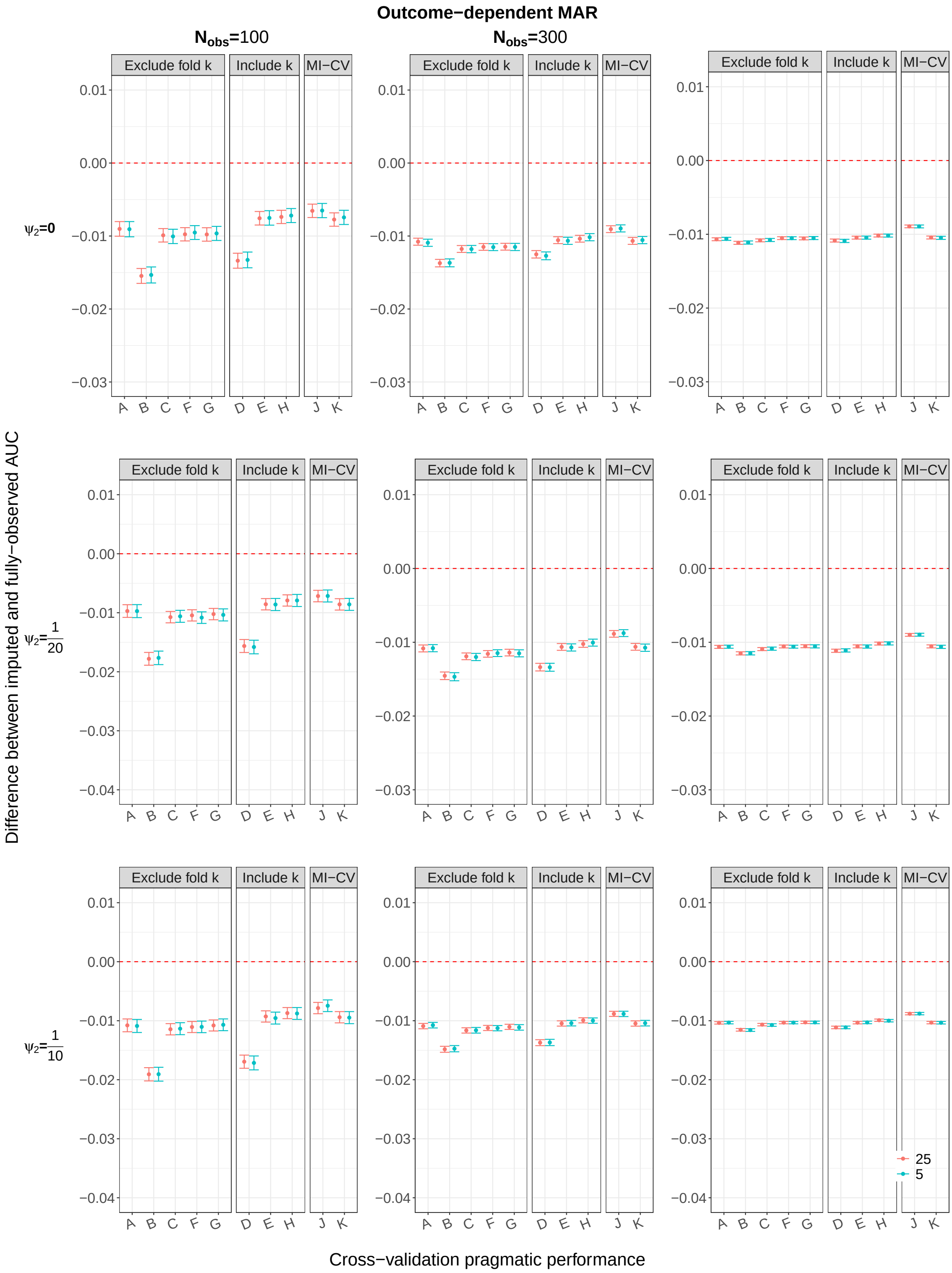

Figure S10: The difference AUC$_{imp}$ - AUC$_{obs}$ when data are outcome-dependent or outcome-and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{obs}$. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.1.4 AUC from imputation methods compared to the target AUC (AUC_{target}) using a larger validation set

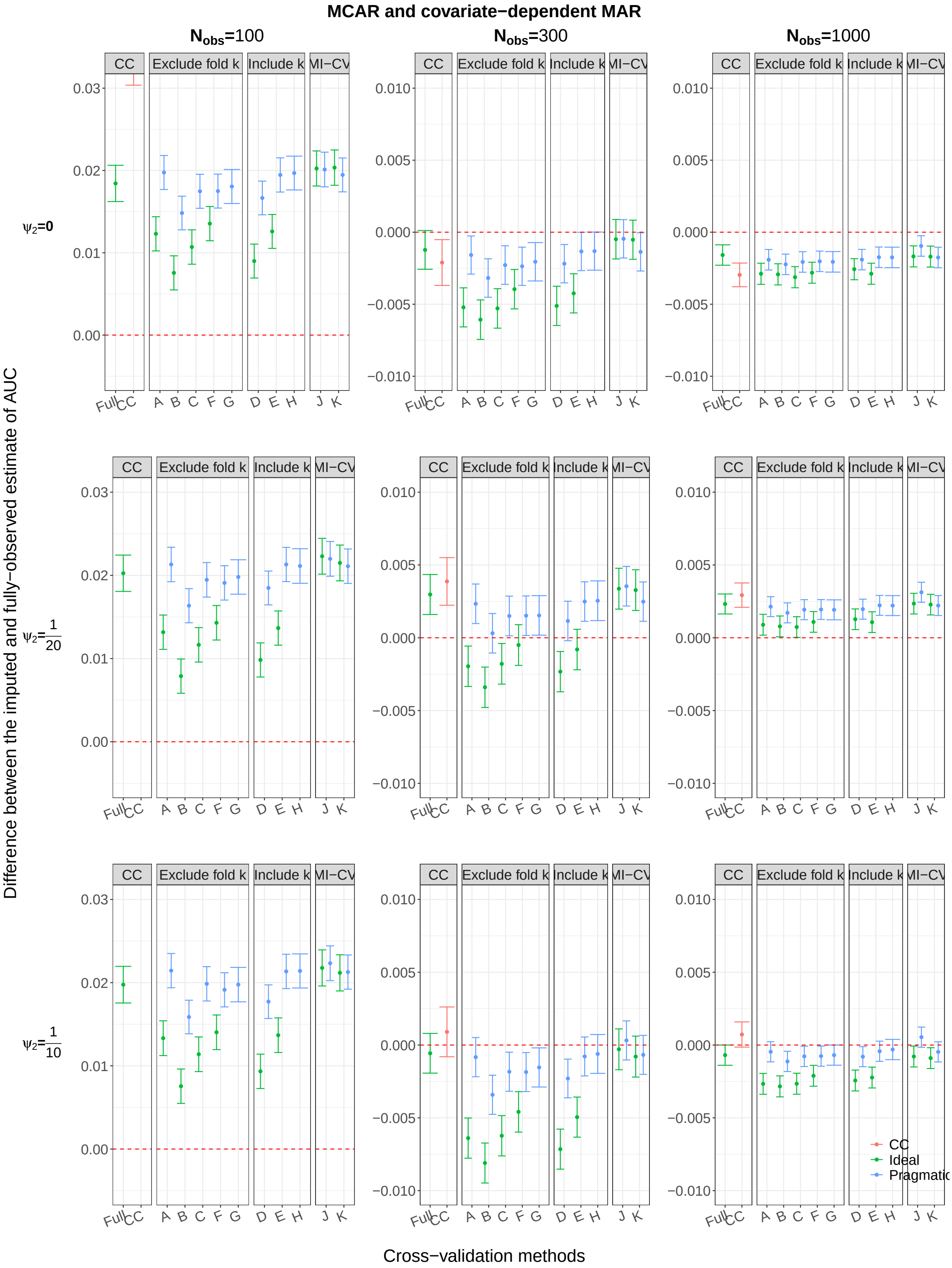

Figure S11: The difference AUC_{imp} - AUC_{target} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{target}. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3.

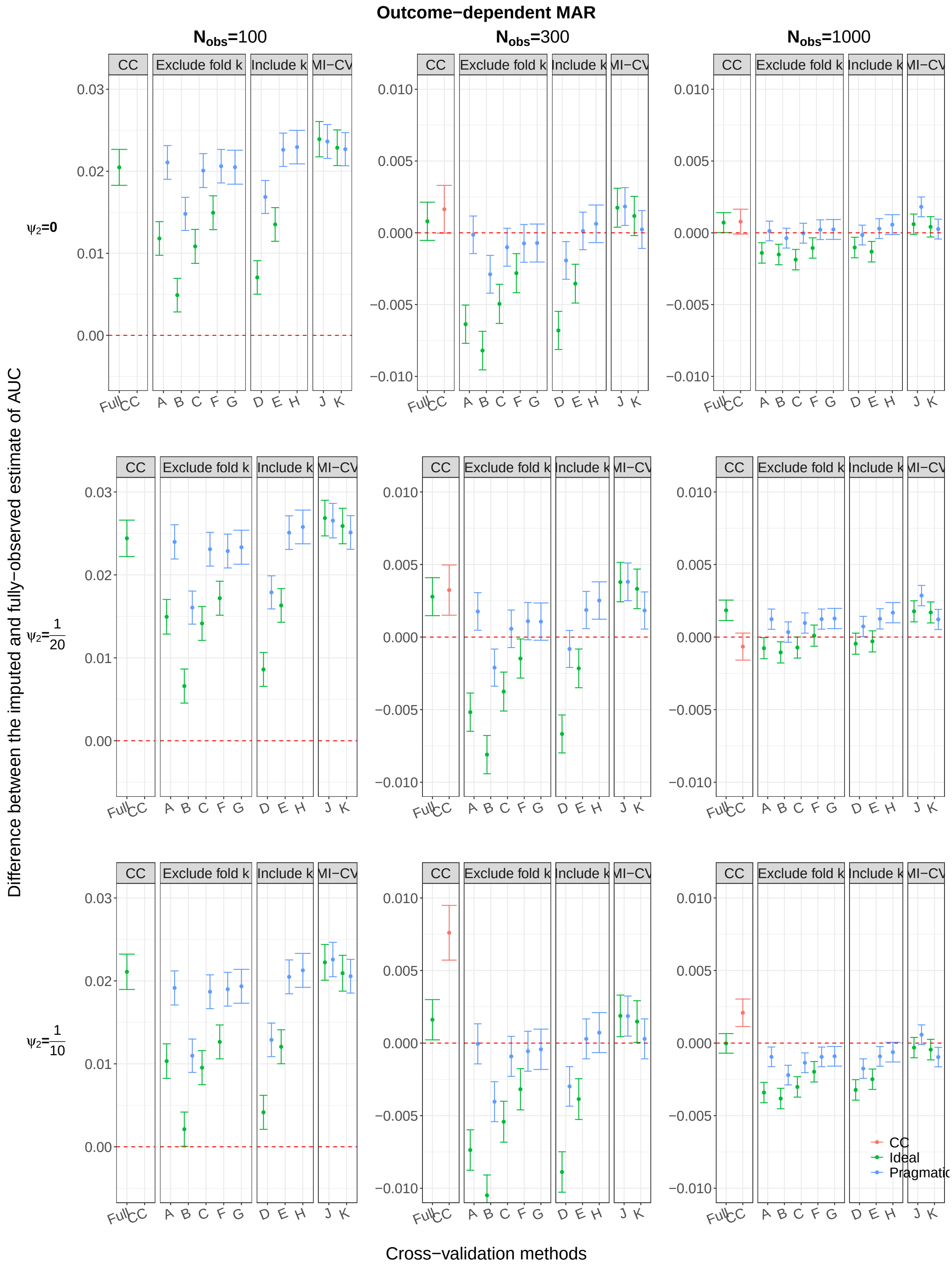

Figure S12: The difference AUC$_{imp}$ - AUC$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{target}$. The average AUC when data are fully-observed is 0.78. CC (complete-case); methods A-K are described in Table 2.3.

S2.2 Brier Score

S2.2.1 Brier score from imputation methods compared to the fully-observed Brier score (Brier$_{imp}$-Brier$_{obs}$)

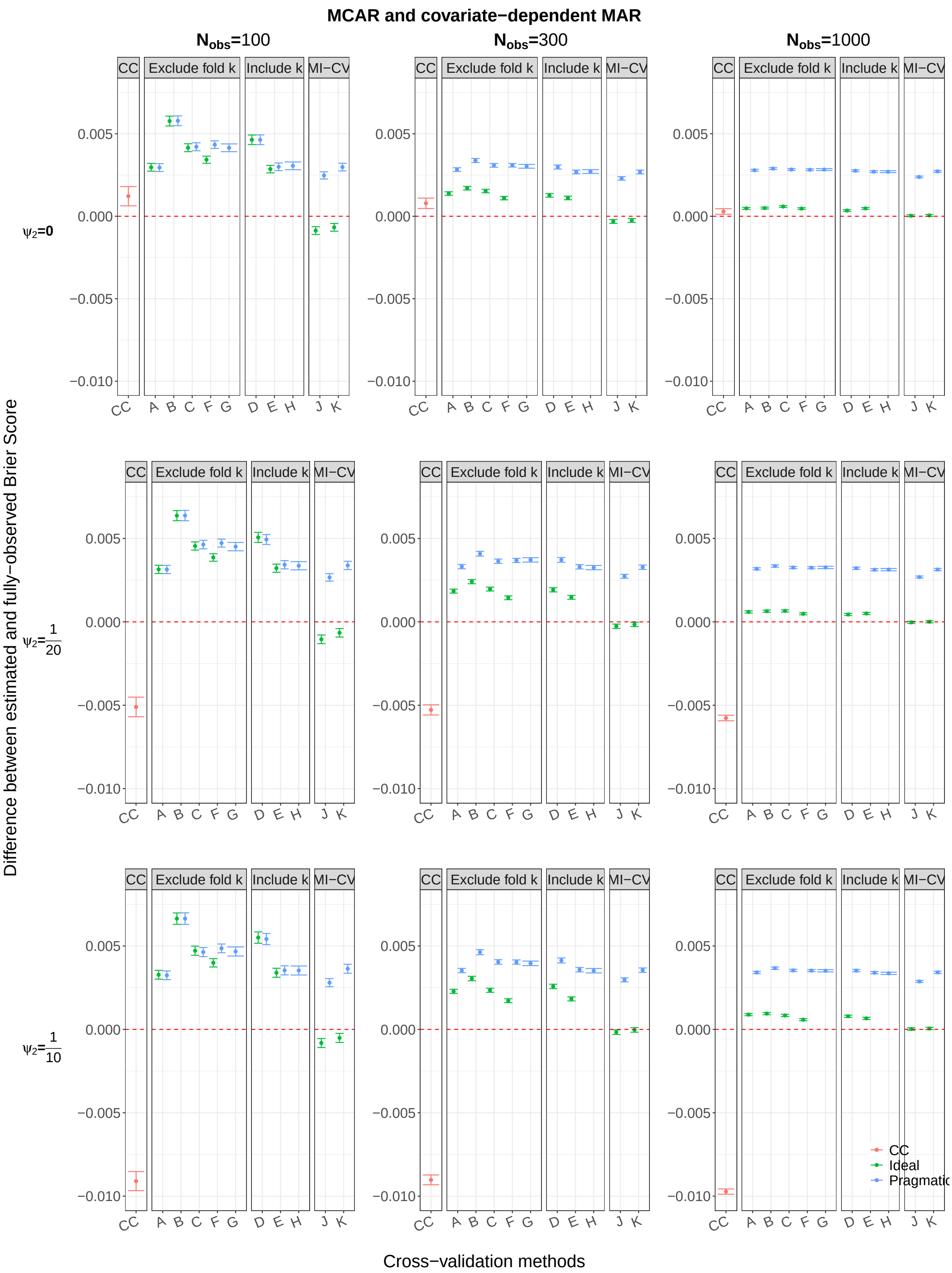

Figure S13: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

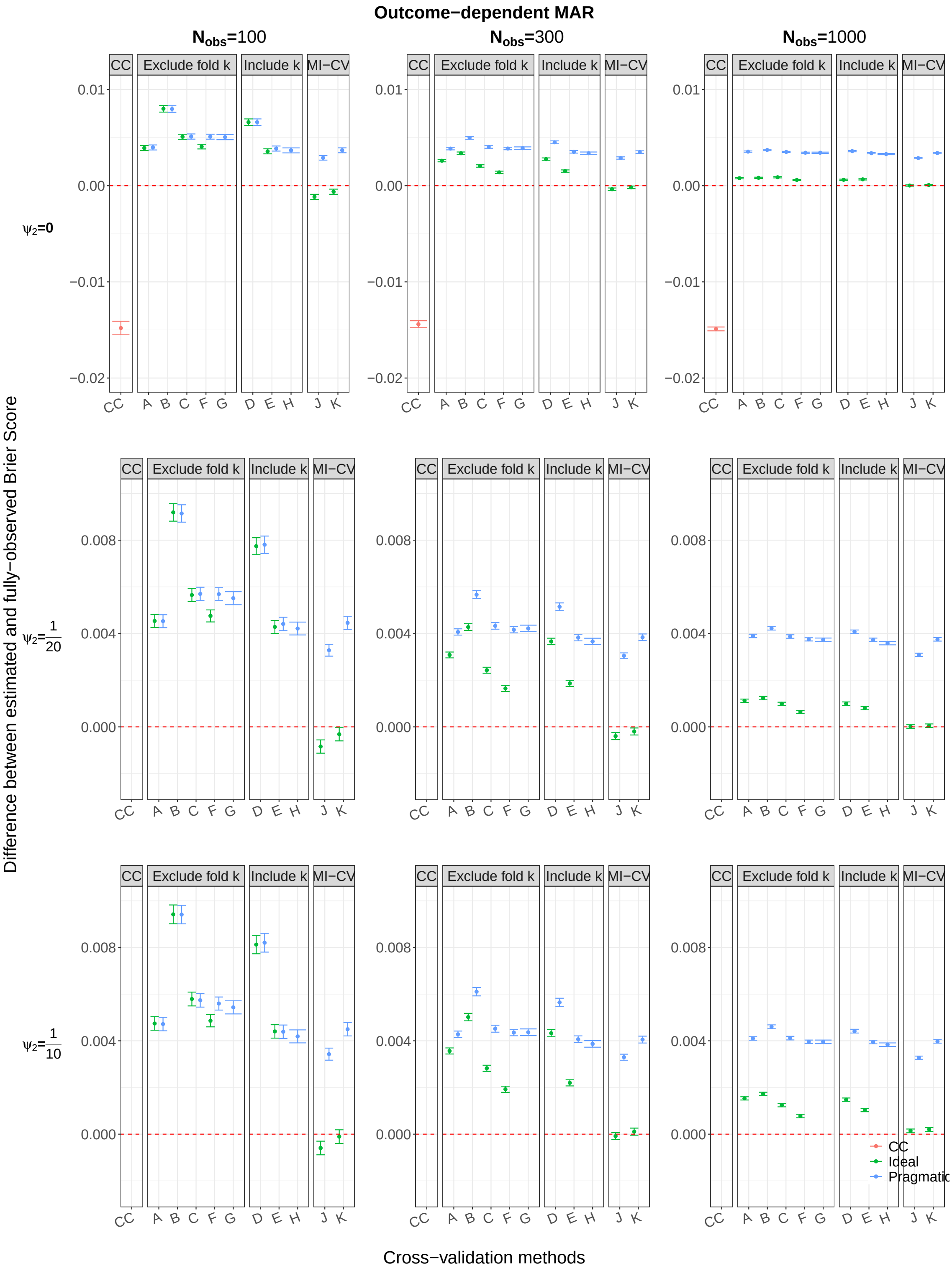

Figure S14: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.2.2 The proportion of missingness is 40% (Brier$_{imp}$-Brier$_{obs}$)

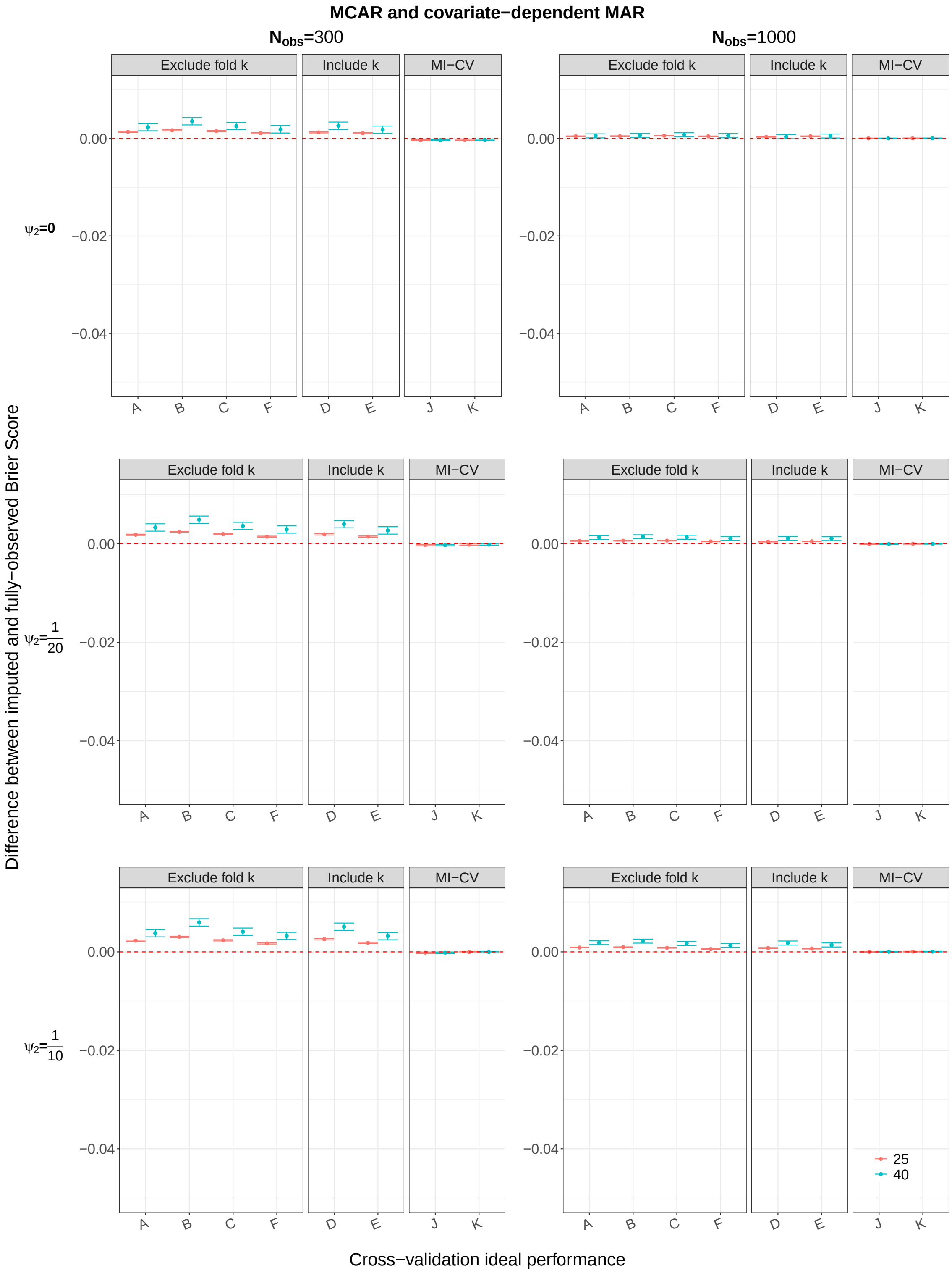

Figure S15: Comparing the impact of increasing the percentage of missingness on the difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of BrierC$_{imp}$ - Brier$_{obs}$. Red denotes Brier$_{imp}$ - Brier$_{obs}$ when 25% of X_1 values are missing and blue denotes Brier$_{imp}$ - Brier$_{obs}$ when 40% of X_1 values are missing. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3.

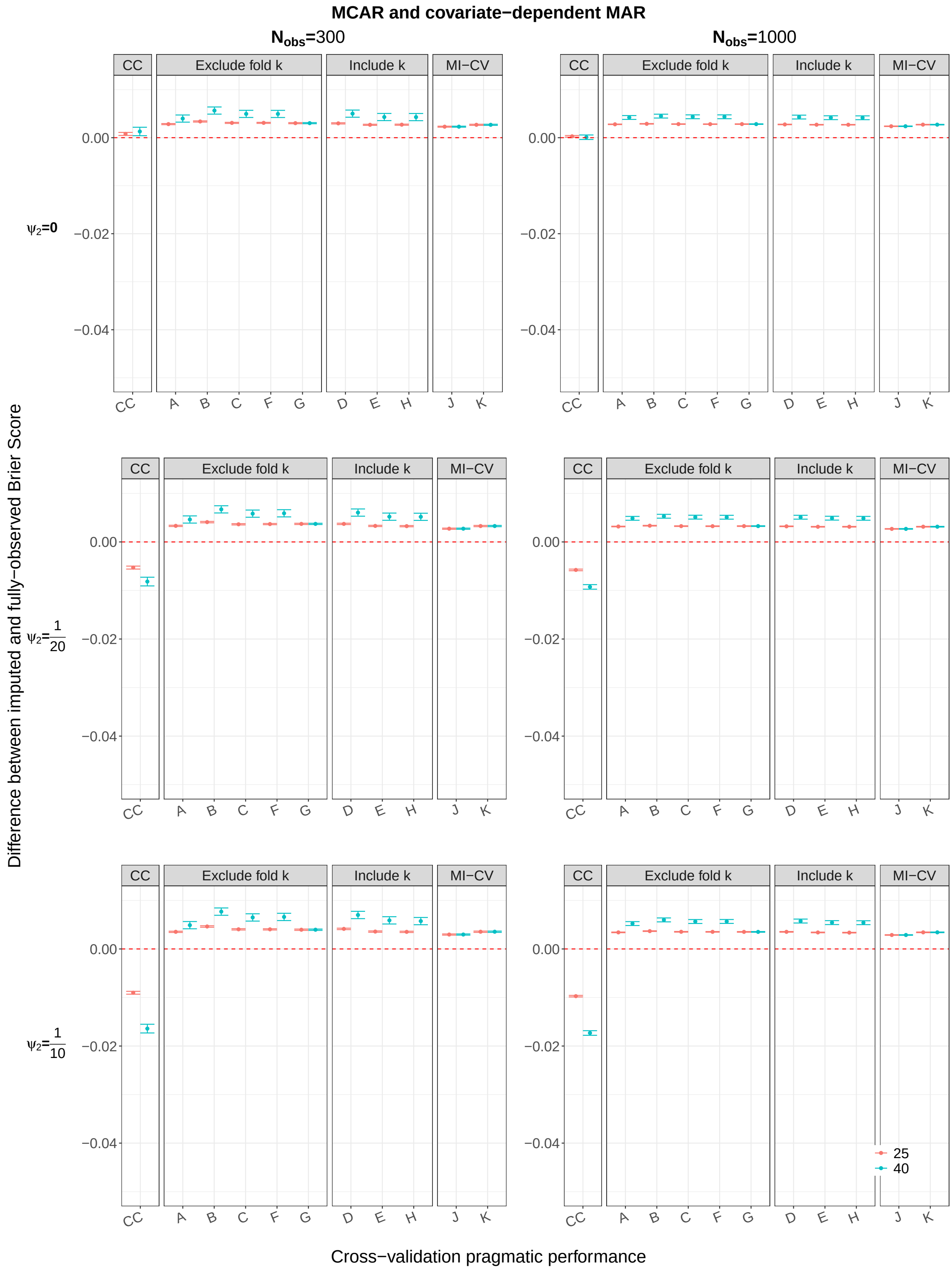

Figure S16: Cross-validation: Comparing the impact of increased missingness on the pragmatic performance of the methods when compared to the fully-observed data's estimate of the Brier score when data are MCAR or covariate-dependent MAR

Figure S17: Comparing the impact of increasing the percentage of missingness on the difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of BrierC$_{imp}$ - Brier$_{obs}$. Red denotes Brier$_{imp}$ - Brier$_{obs}$ when 25% of X_1 values are missing and blue denotes Brier$_{imp}$ - Brier$_{obs}$ when 40% of X_1 values are missing. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3.

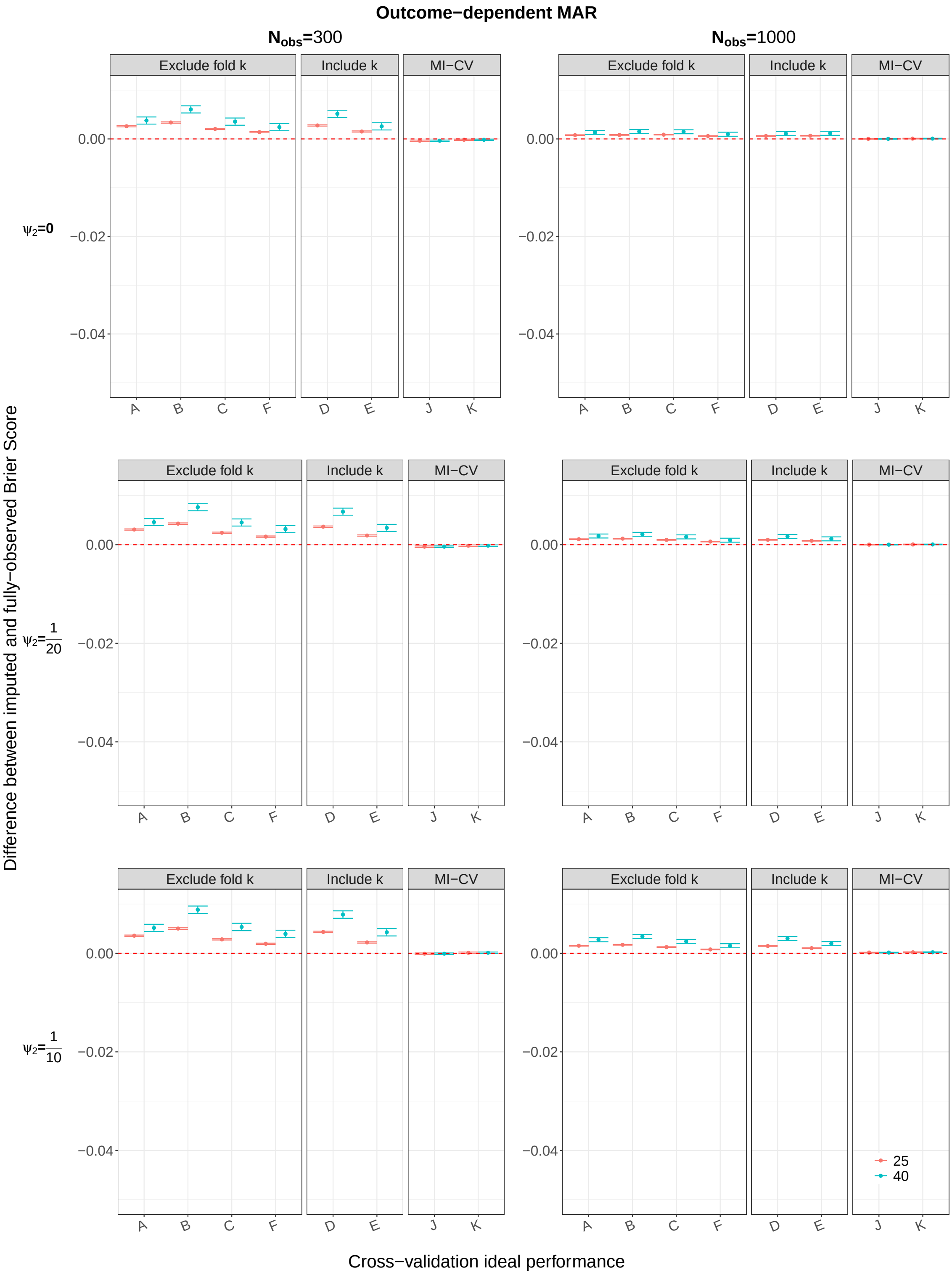

Figure S18: Comparing the impact of increasing the percentage of missingness on the difference Brier_{imp} - Brier_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Brier_{imp} - Brier_{obs}. Red denotes Brier_{imp} - Brier_{obs} when 25% of X_1 values are missing and blue denotes Brier_{imp} - Brier_{obs} when 40% of X_1 values are missing. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3.

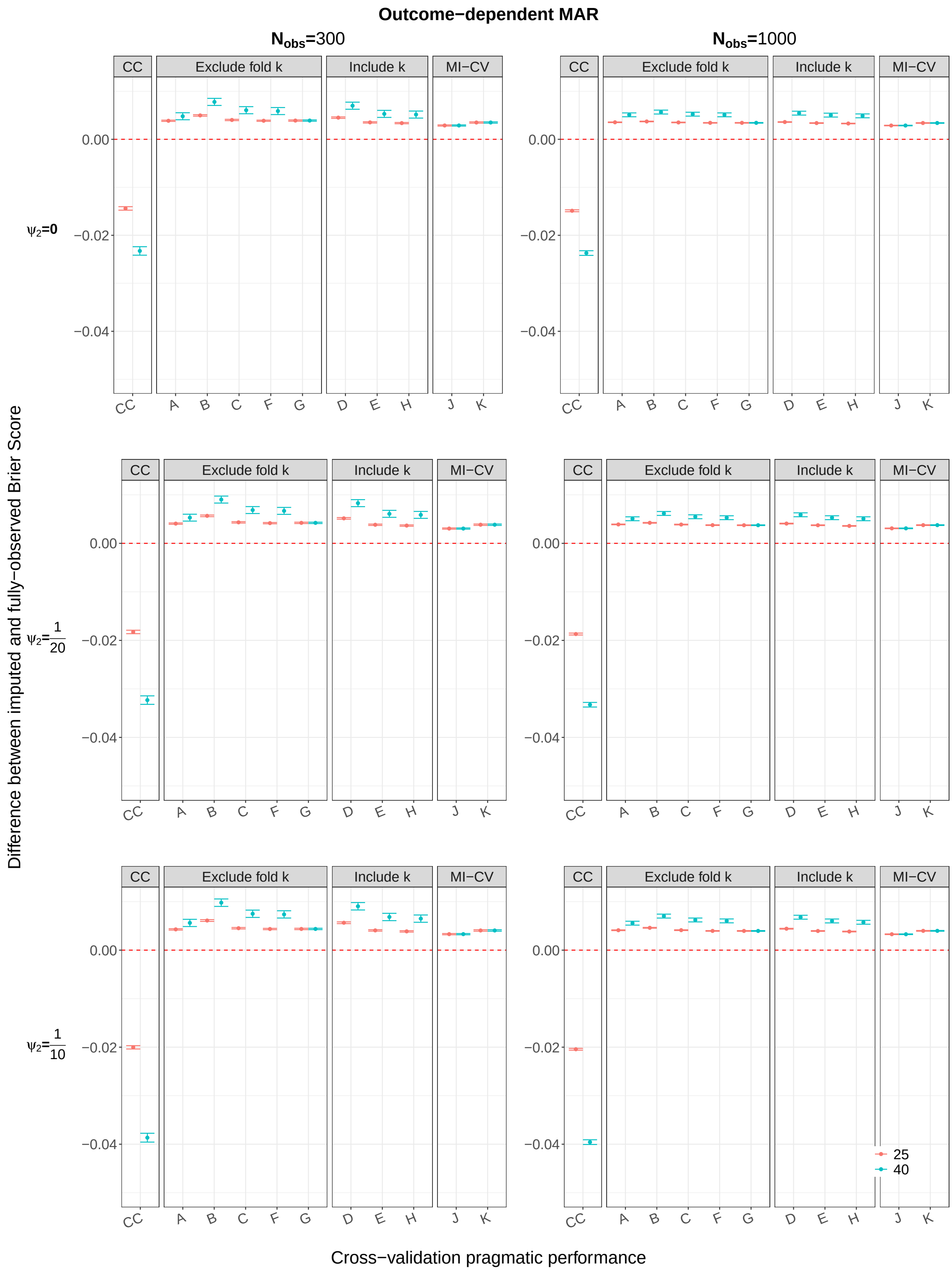

Figure S19: Comparing the impact of increasing the percentage of missingness on the difference Brier_{imp} - Brier_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Brier_{imp} - Brier_{obs}. Red denotes Brier_{imp} - Brier_{obs} when 25% of X_1 values are missing and blue denotes Brier_{imp} - Brier_{obs} when 40% of X_1 values are missing. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3.

S2.2.3 Comparing M=5 versus M=25 (Brier$_{imp}$-Brier$_{obs}$)

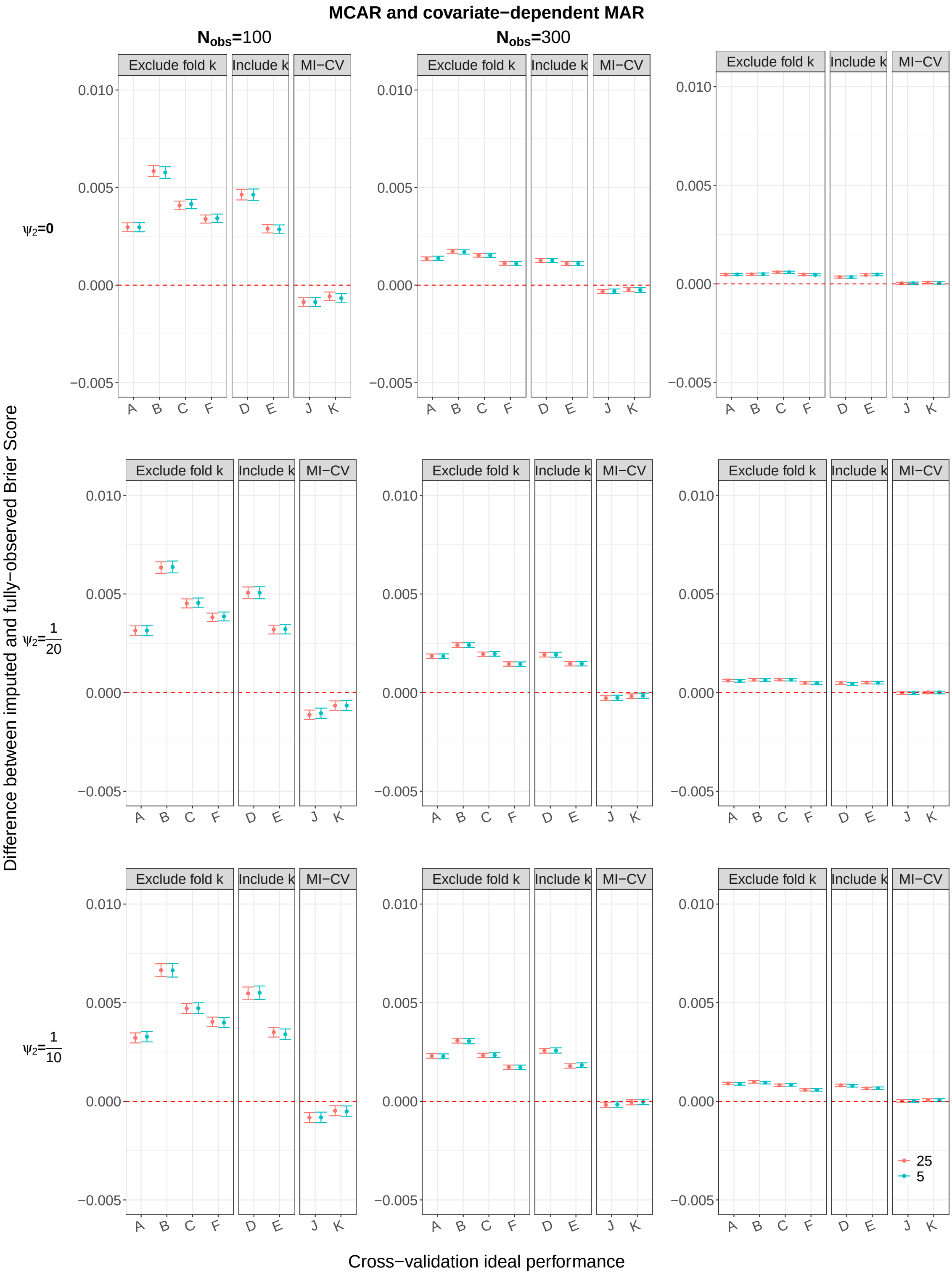

Figure S20: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

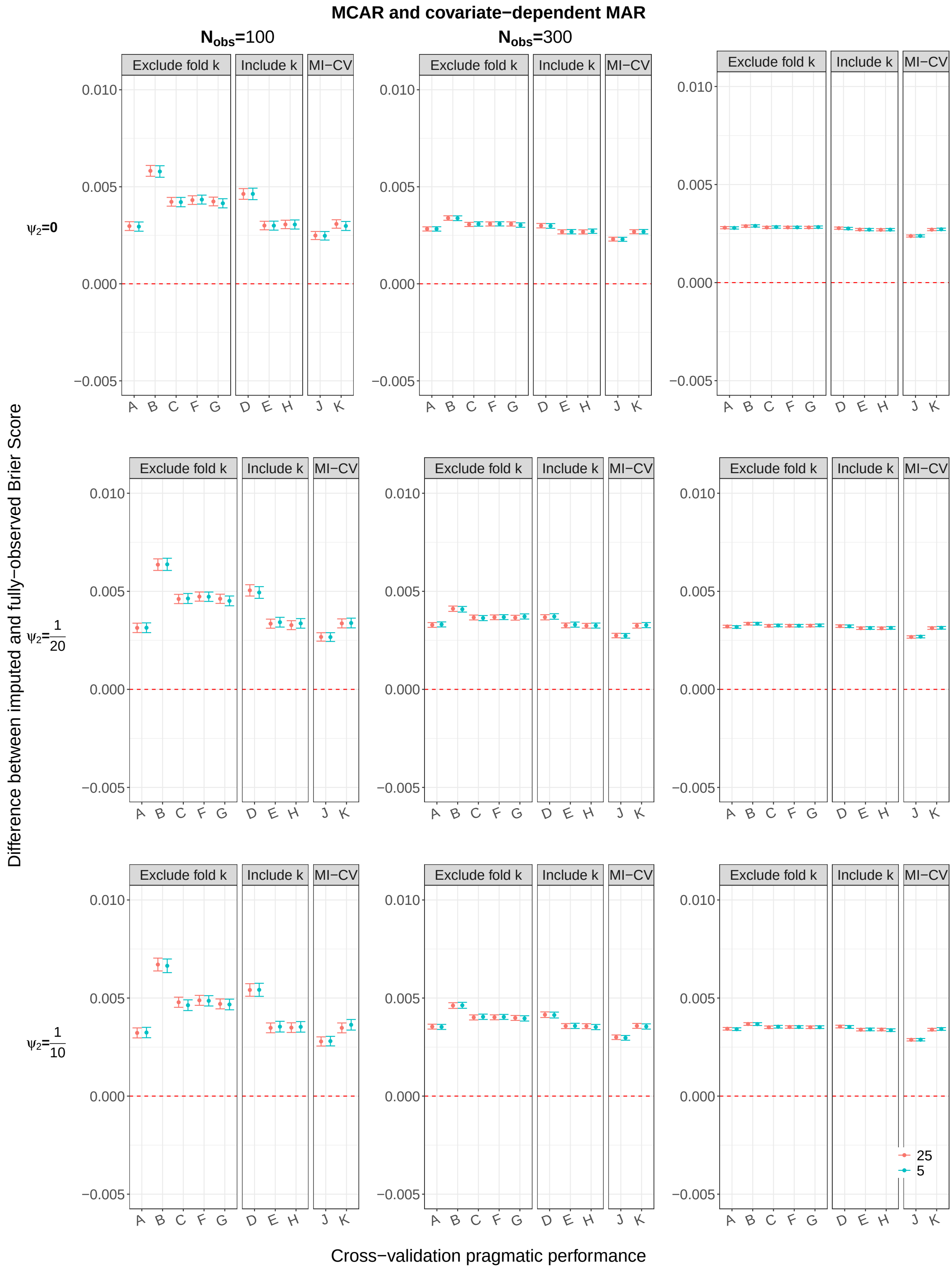

Figure S21: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

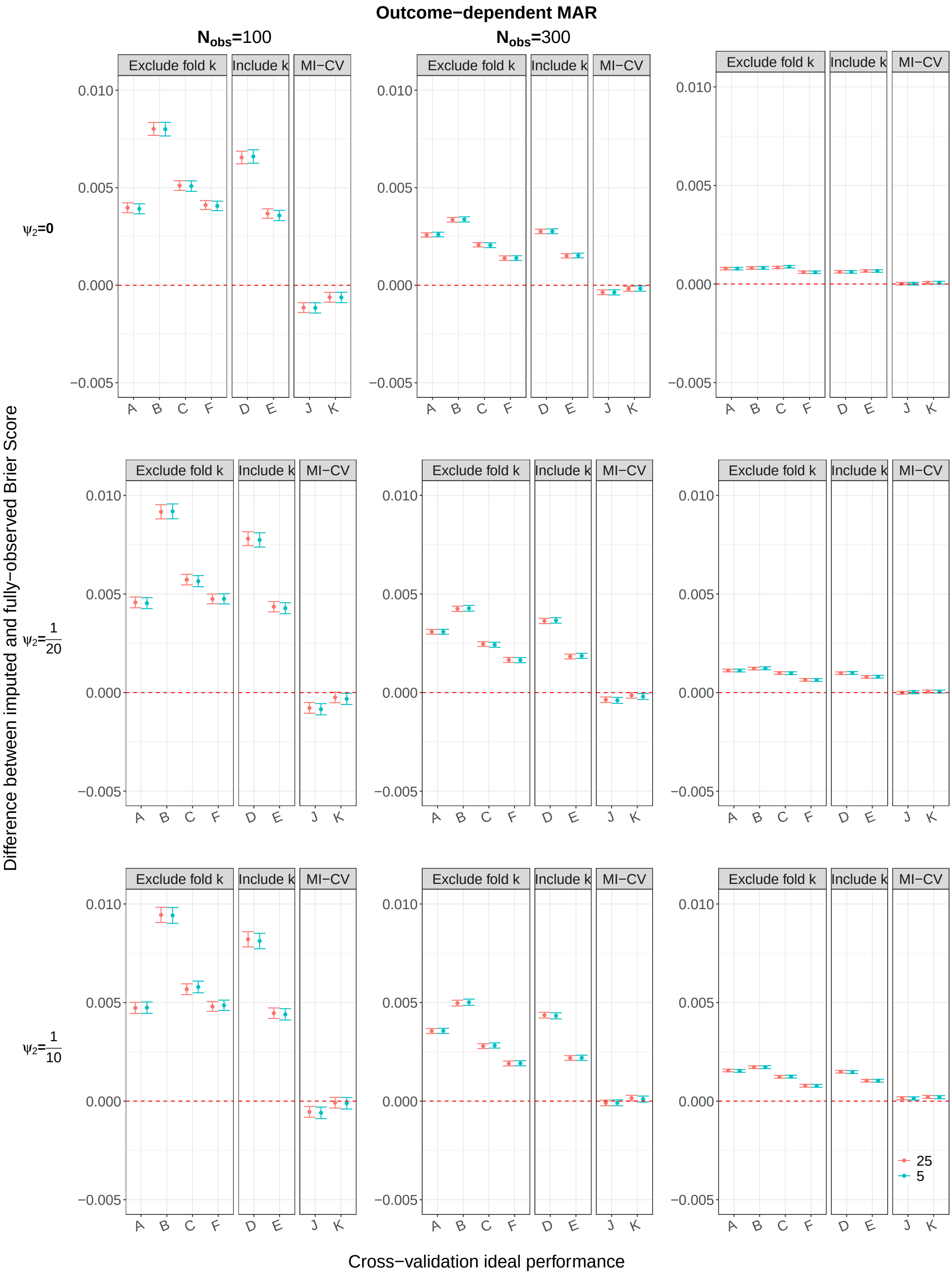

Figure S22: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

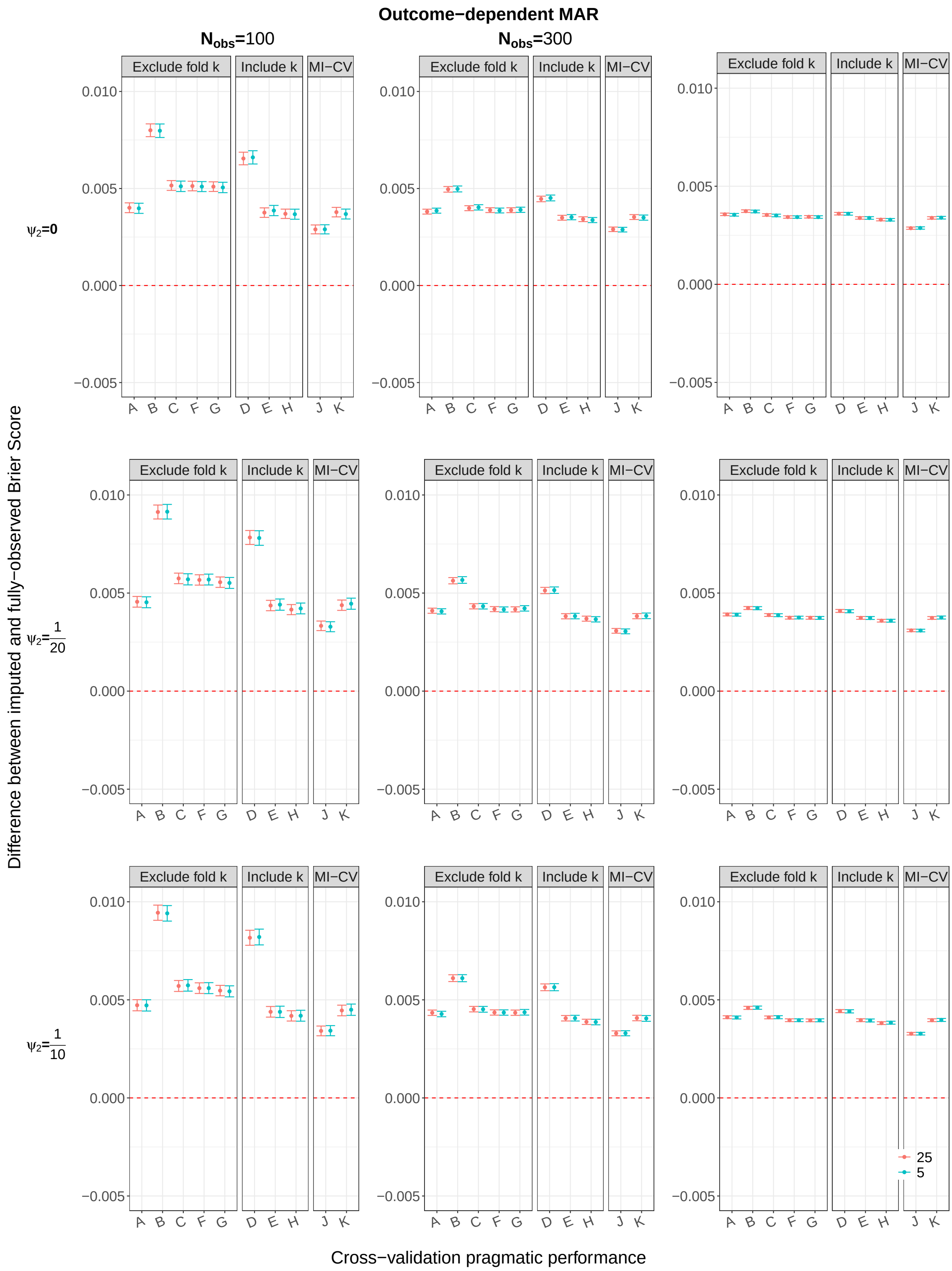

Figure S23: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. The average Brier score when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.2.4 Brier score from imputation methods compared to the target Brier score (Brier$_{target}$) using a larger validation set

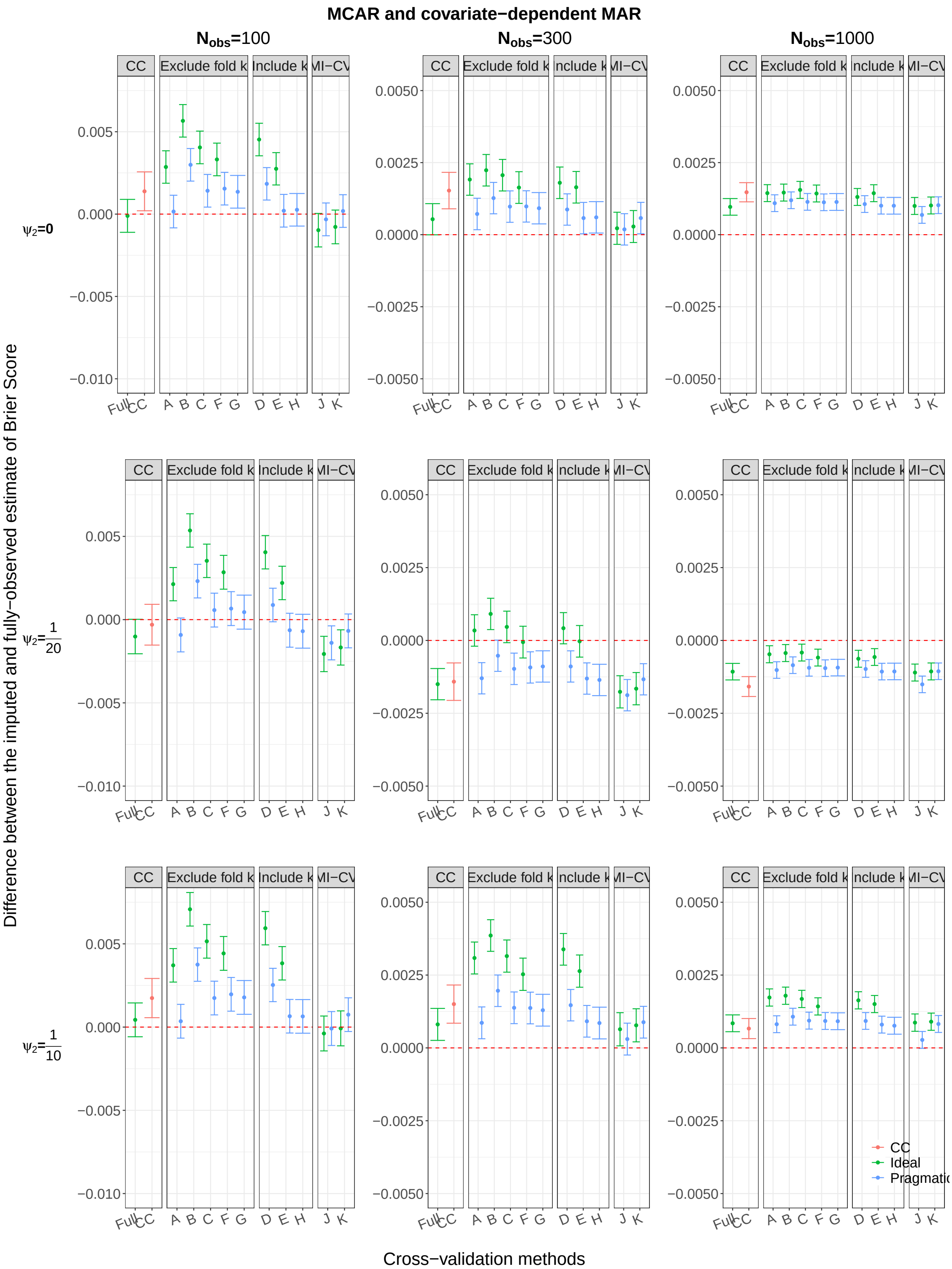

Figure S24: The difference Brier$_{imp}$ - Brier$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{target}$. The average Brier when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3.

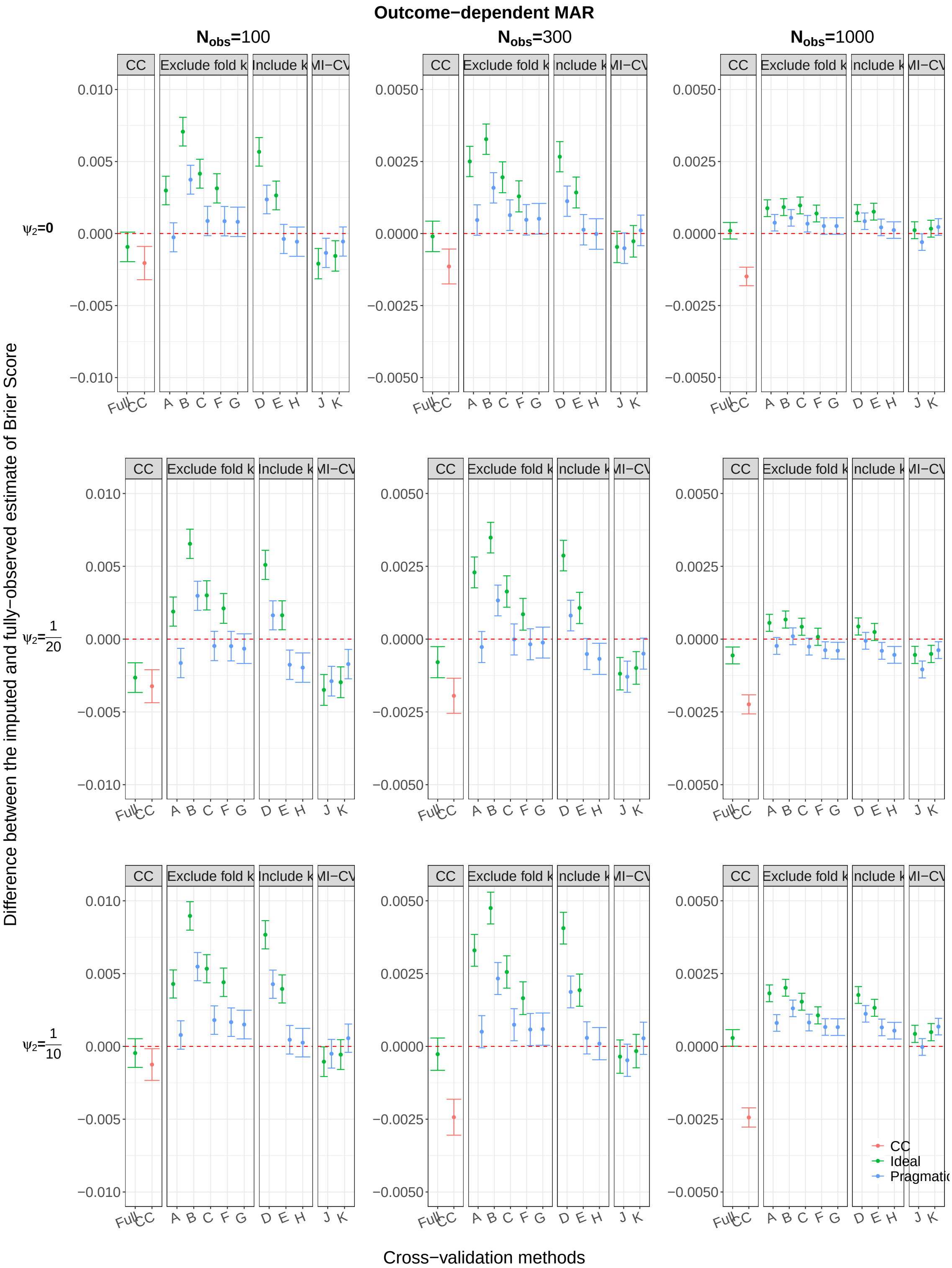

Figure S25: The difference Brier$_{imp}$ - Brier$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{target}$. The average Brier when data are fully-observed is 0.17. CC (complete-case); methods A-K are described in Table 2.3.

S2.3 Calibration intercept

S2.3.1 Calibration intercept from imputation methods compared to the fully-observed calibration intercept (intercept_{imp}-intercept_{obs})

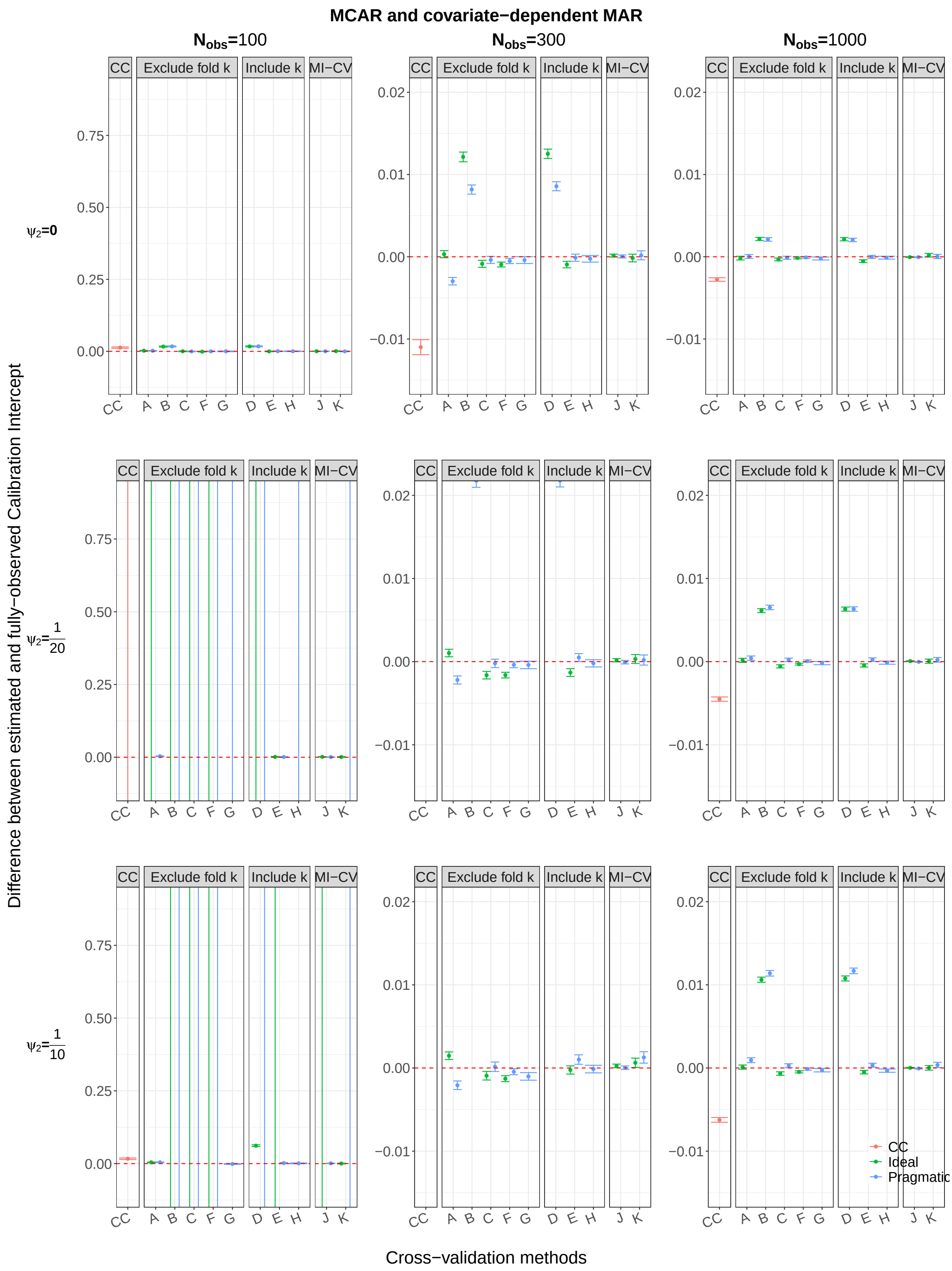

Figure S26: The difference Intercept_{imp} - Intercept_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept_{imp} - Intercept_{obs}. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

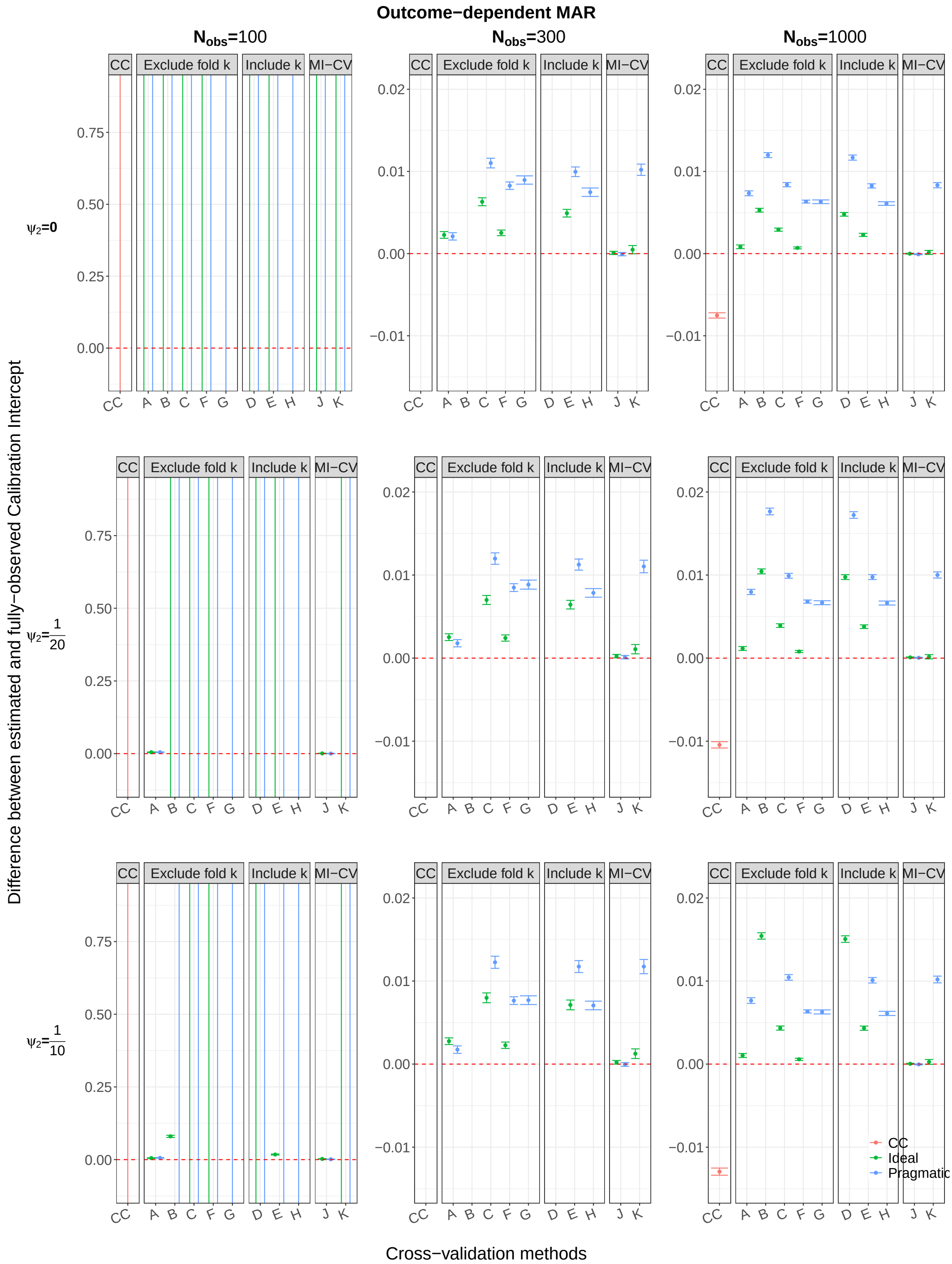

Figure S27: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.3.2 The proportion of missingness is 40% (intercept$_{imp}$-intercept$_{obs}$)

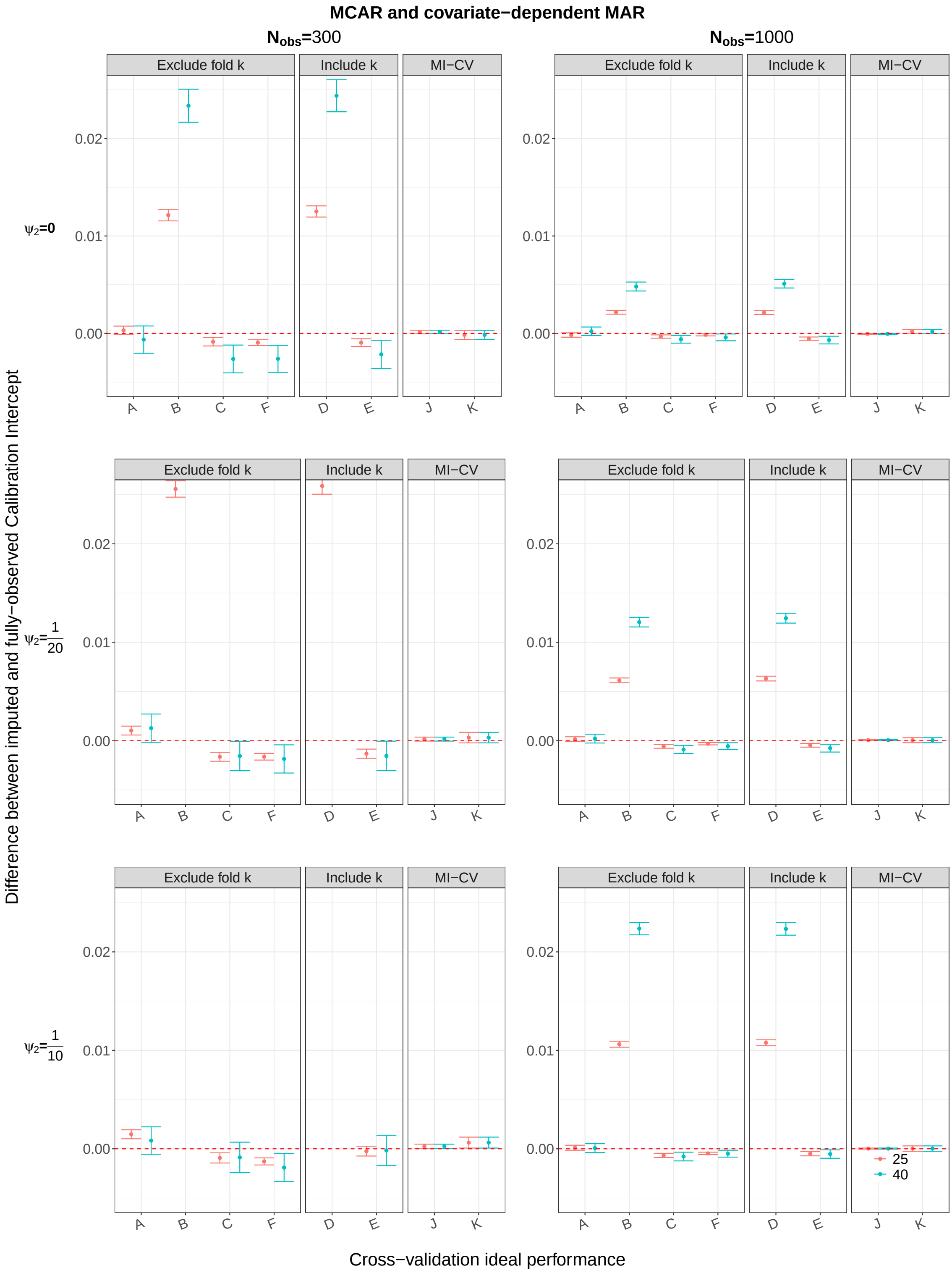

Figure S28: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

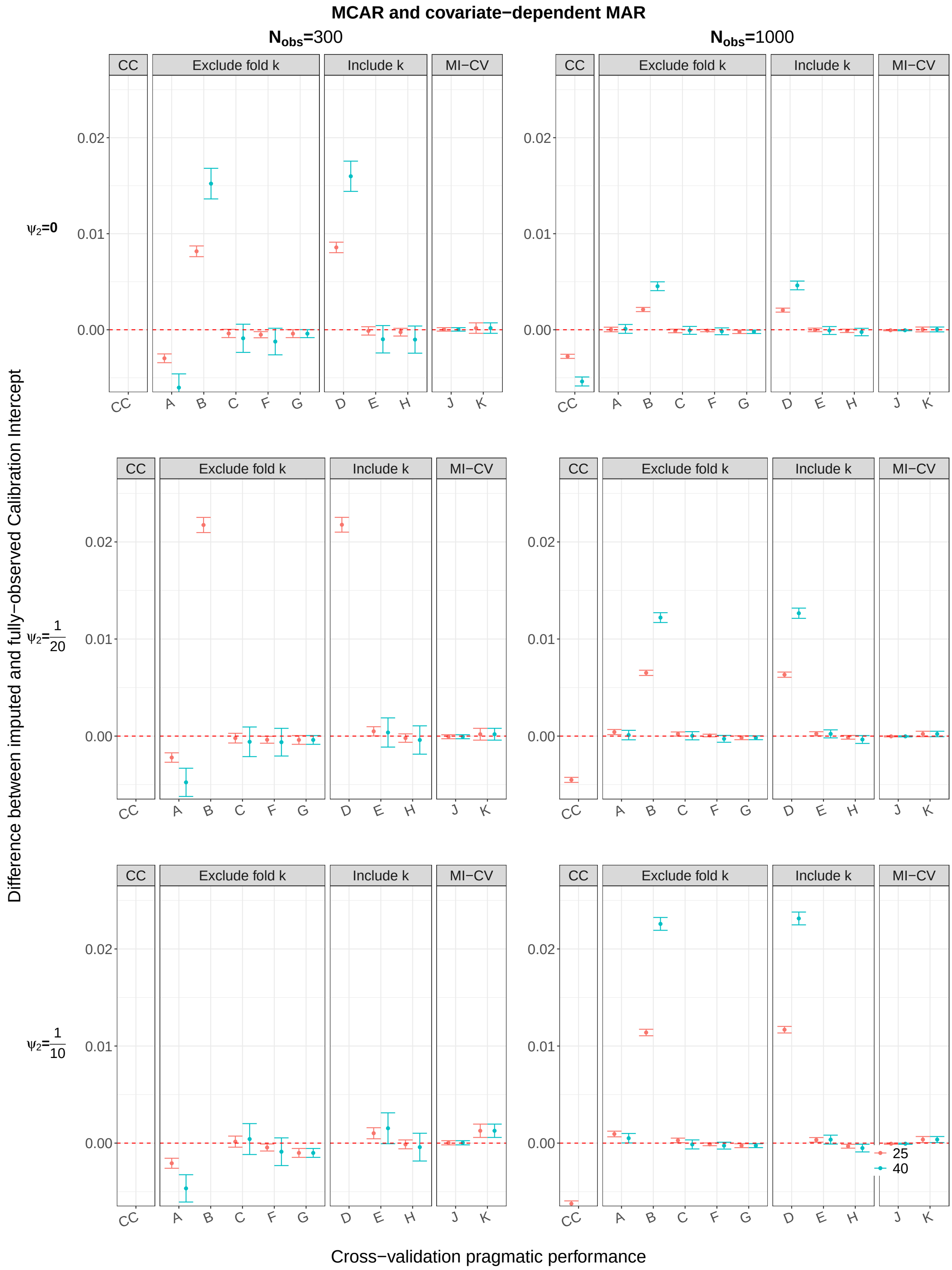

Figure S29: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

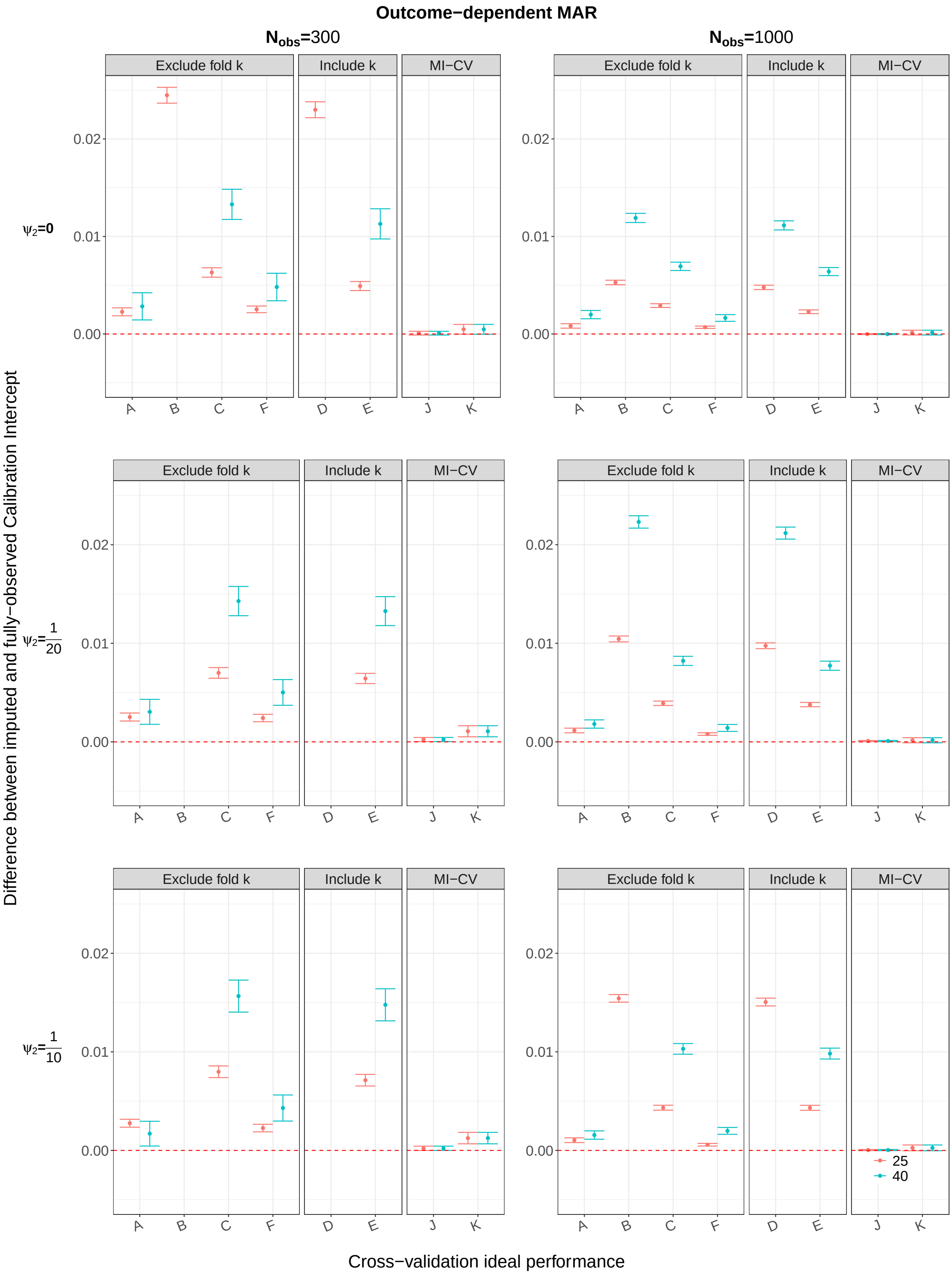

Figure S30: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

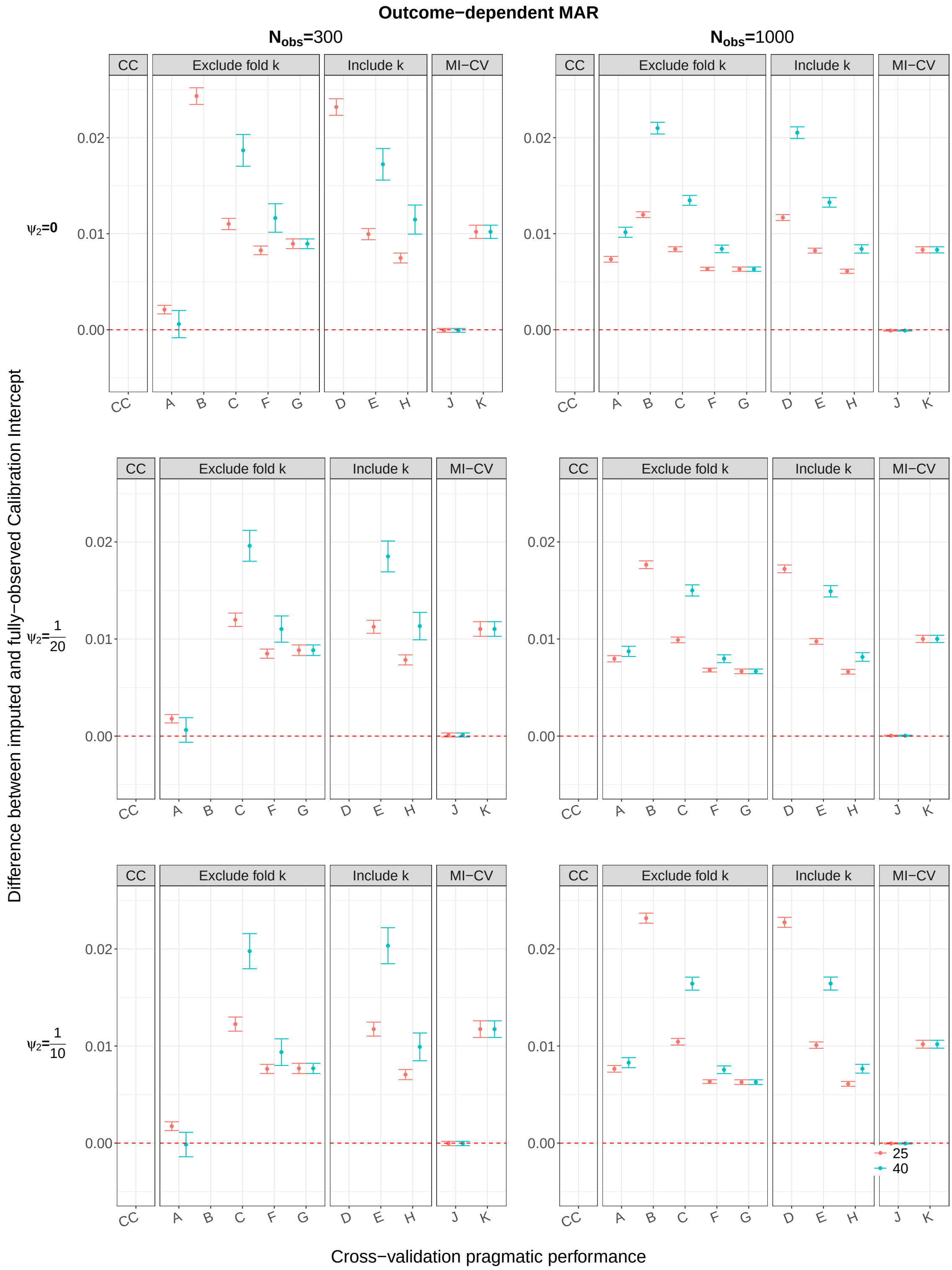

Figure S31: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

S2.3.3 Comparing M=5 versus M=25 (intercept$_{imp}$-intercept$_{obs}$)

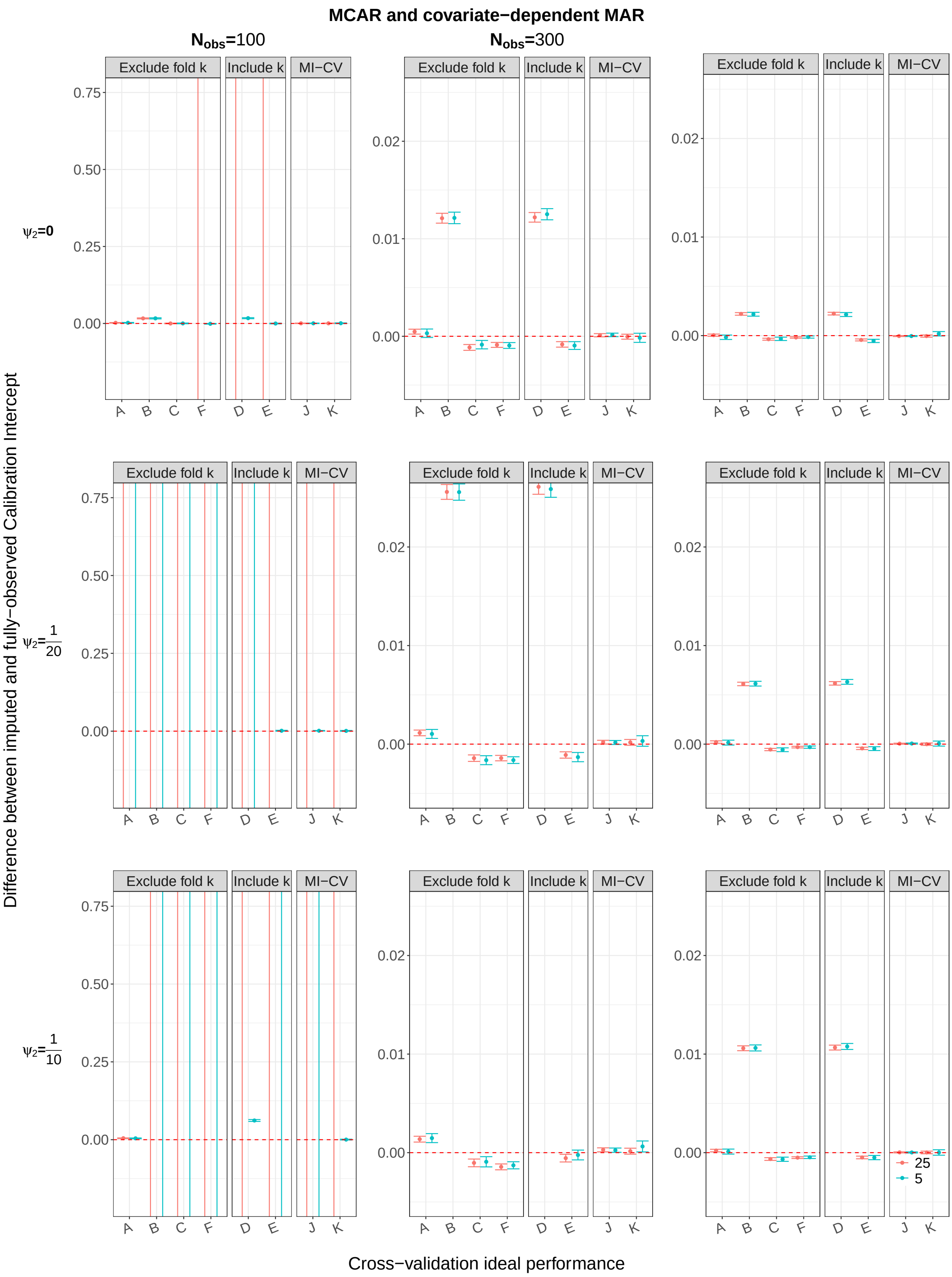

Figure S32: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

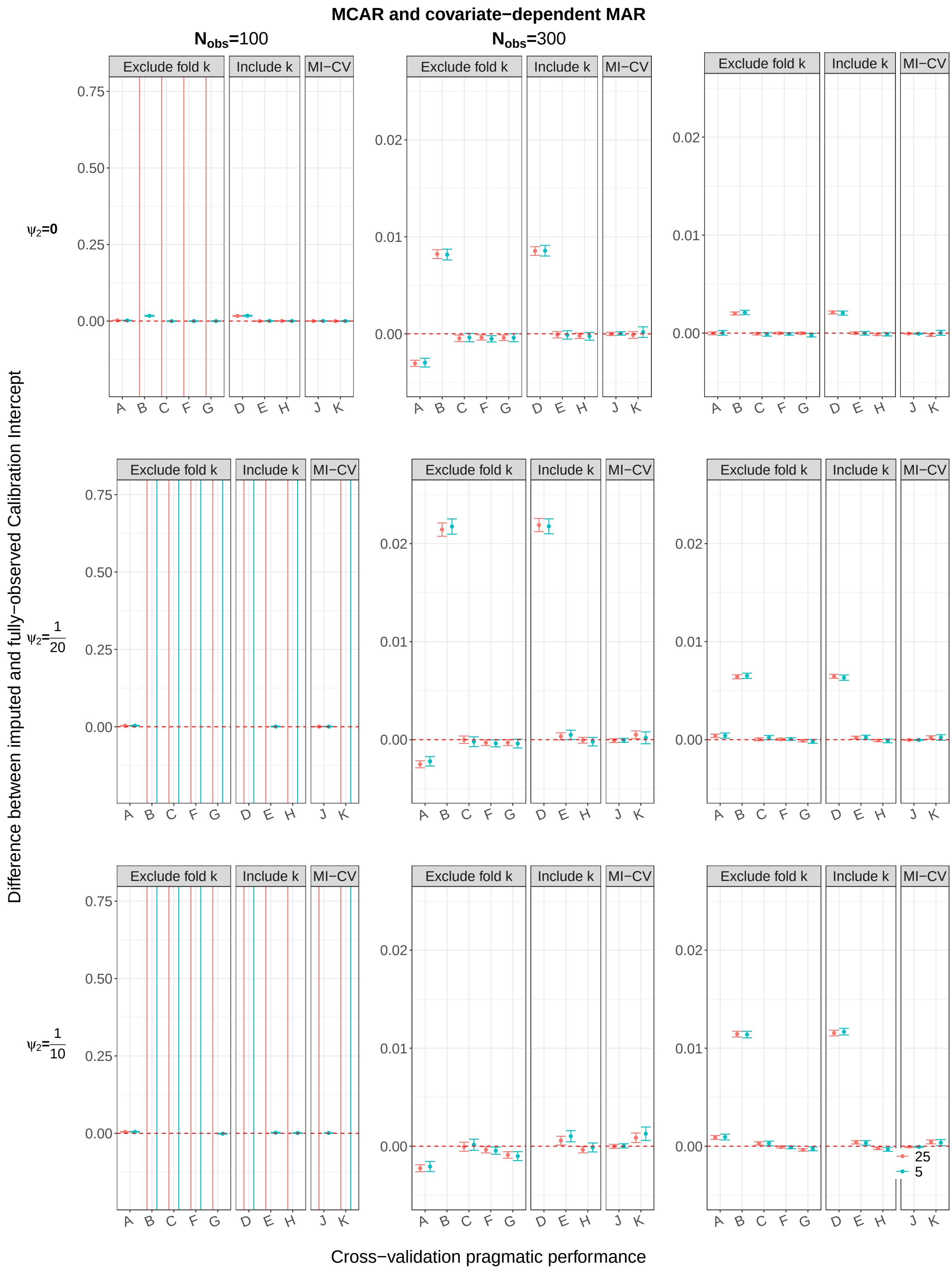

Figure S33: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

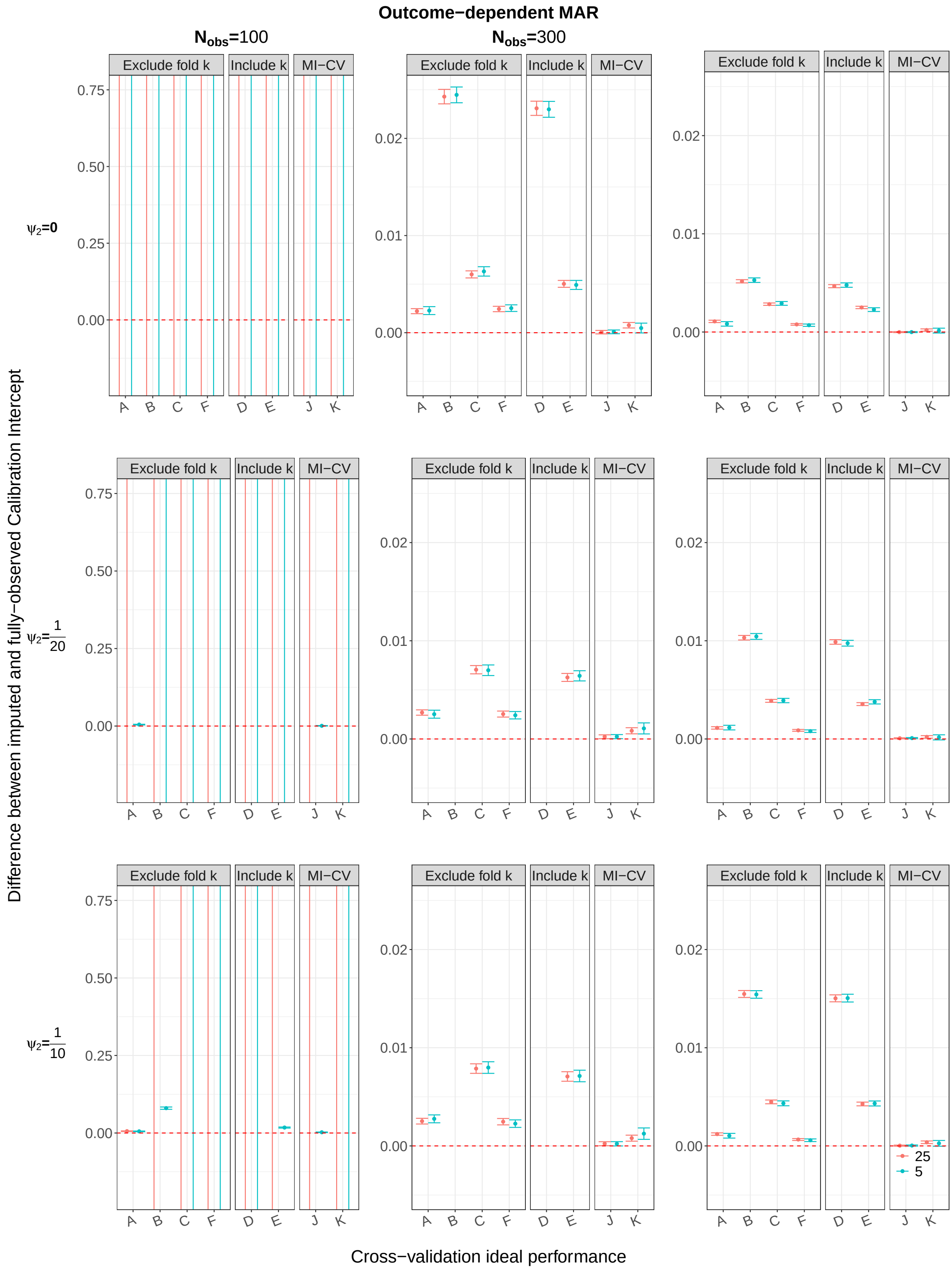

Figure S34: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when outcome-dependent or outcome-and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

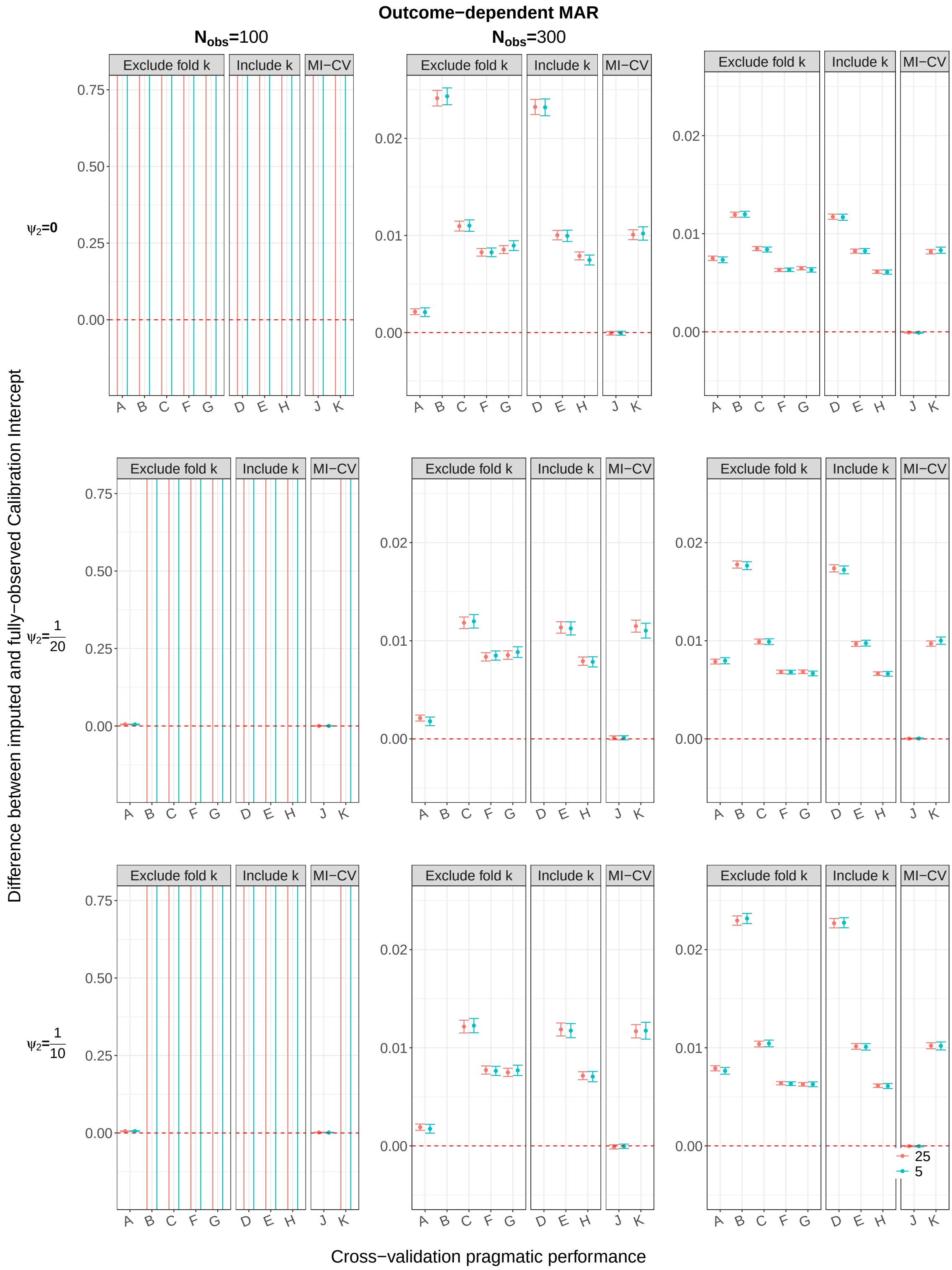

Figure S35: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.3.4 Calibration intercept from imputation methods compared to the target calibration intercept (intercept$_{target}$) using a larger validation set

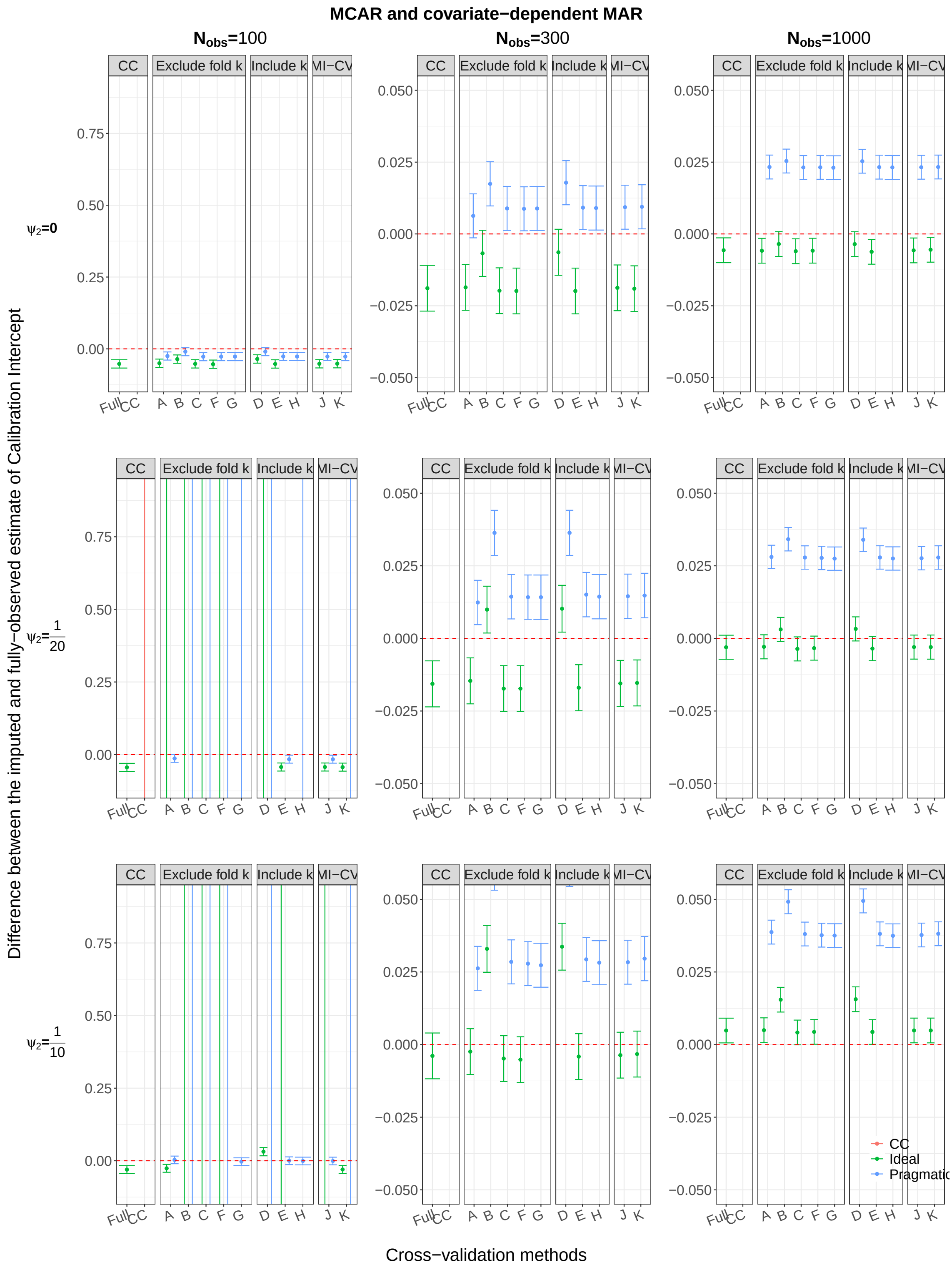

Figure S36: The difference Intercept$_{imp}$ - Intercept$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{target}$. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

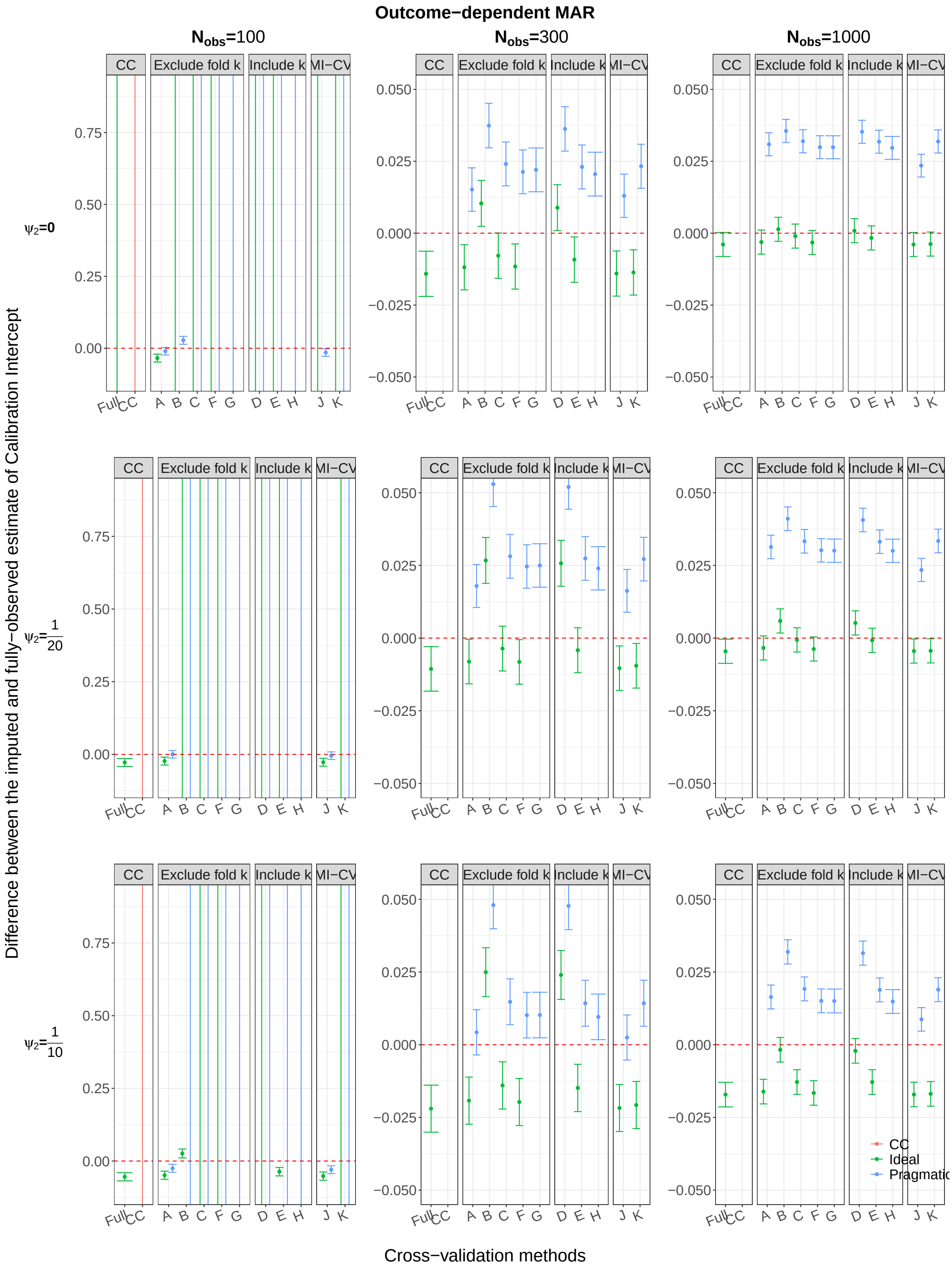

Figure S37: The difference Intercept$_{imp}$ - Intercept$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{target}$. The average Calibration intercept when data are fully-observed is 0.02 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

S2.4 Calibration slope

S2.4.1 Calibration slope from imputation methods compared to the fully-observed calibration slope (slope$_{imp}$-slope$_{obs}$)

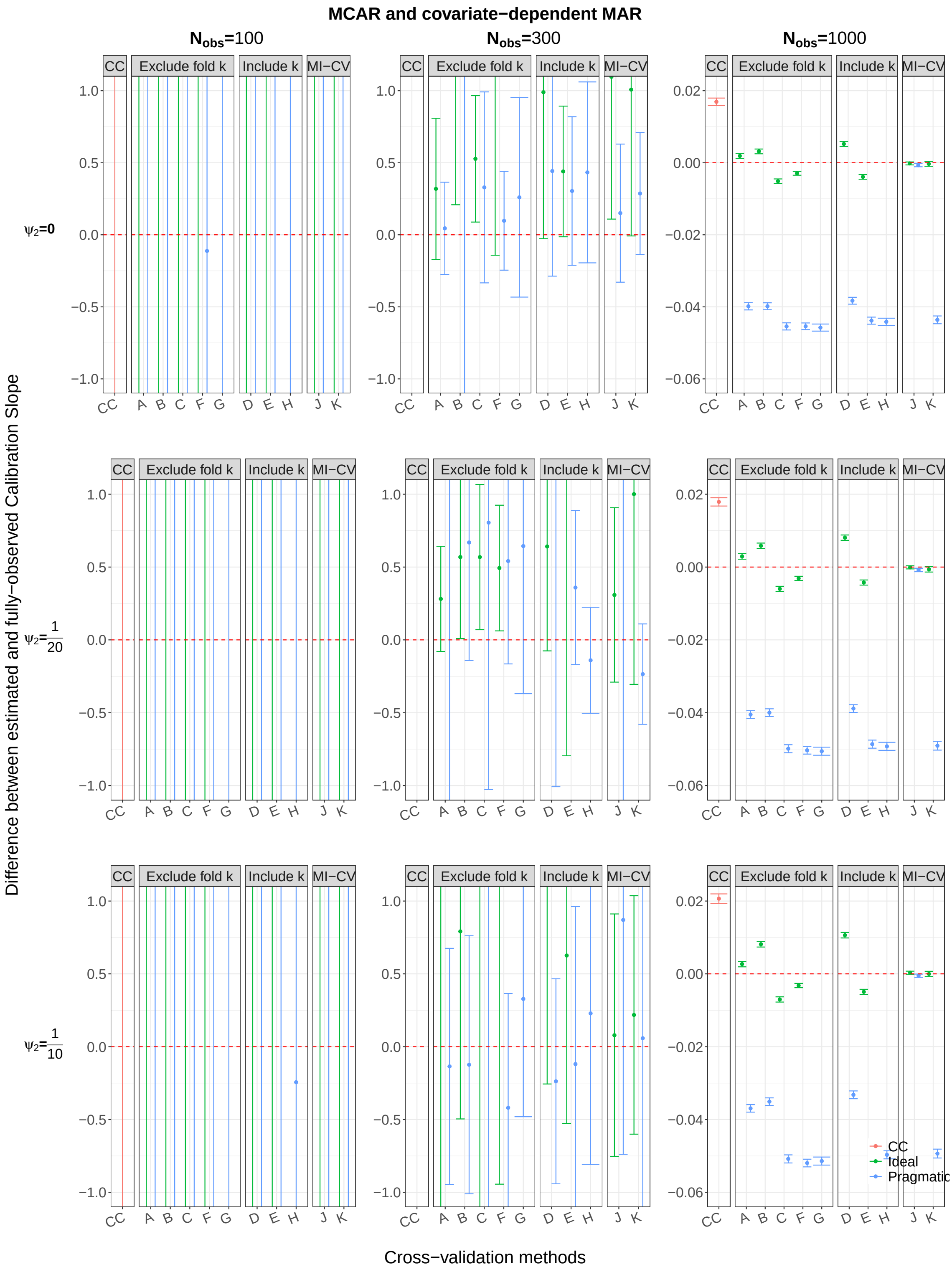

Figure S38: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

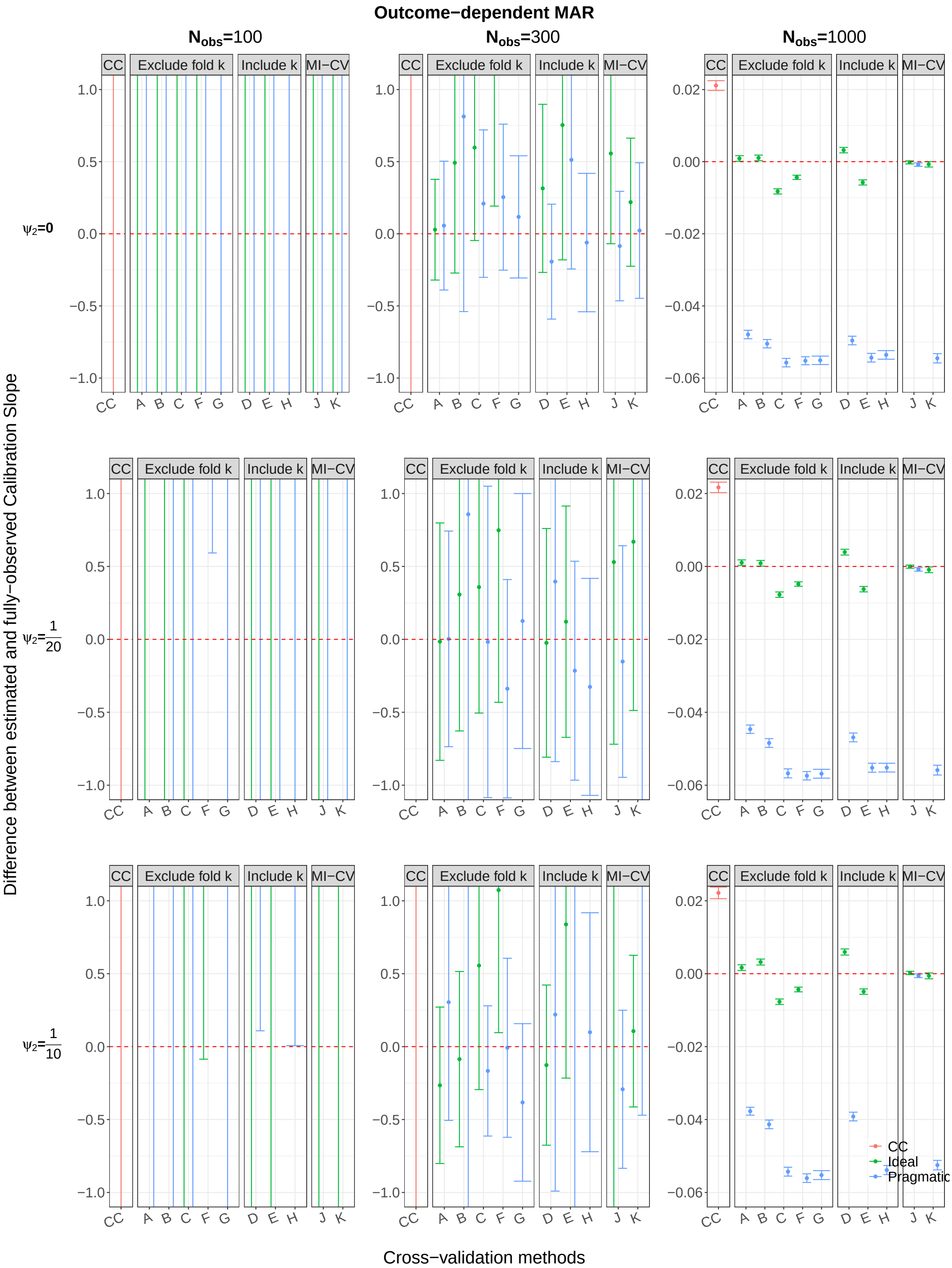

Figure S39: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.4.2 The proportion of missingness is 40% (slope$_{imp}$-slope$_{obs}$)

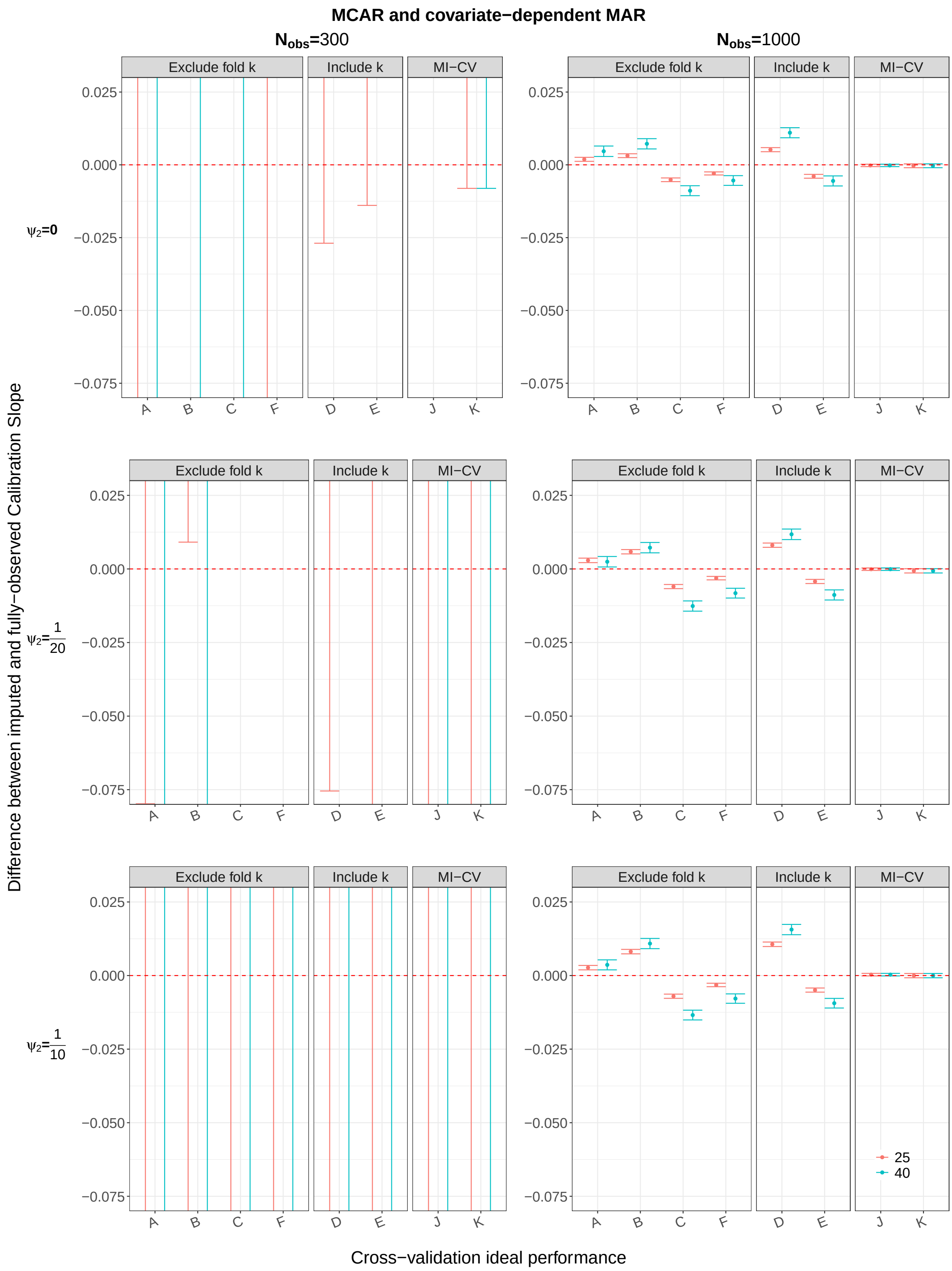

Figure S40: Comparing the impact of increasing the percentage of missingness on the difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. Red denotes Slope$_{imp}$ - Slope$_{obs}$ when 25% of X_1 values are missing and blue denotes Slope$_{imp}$ - Slope$_{obs}$ when 40% of X_1 values are missing. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

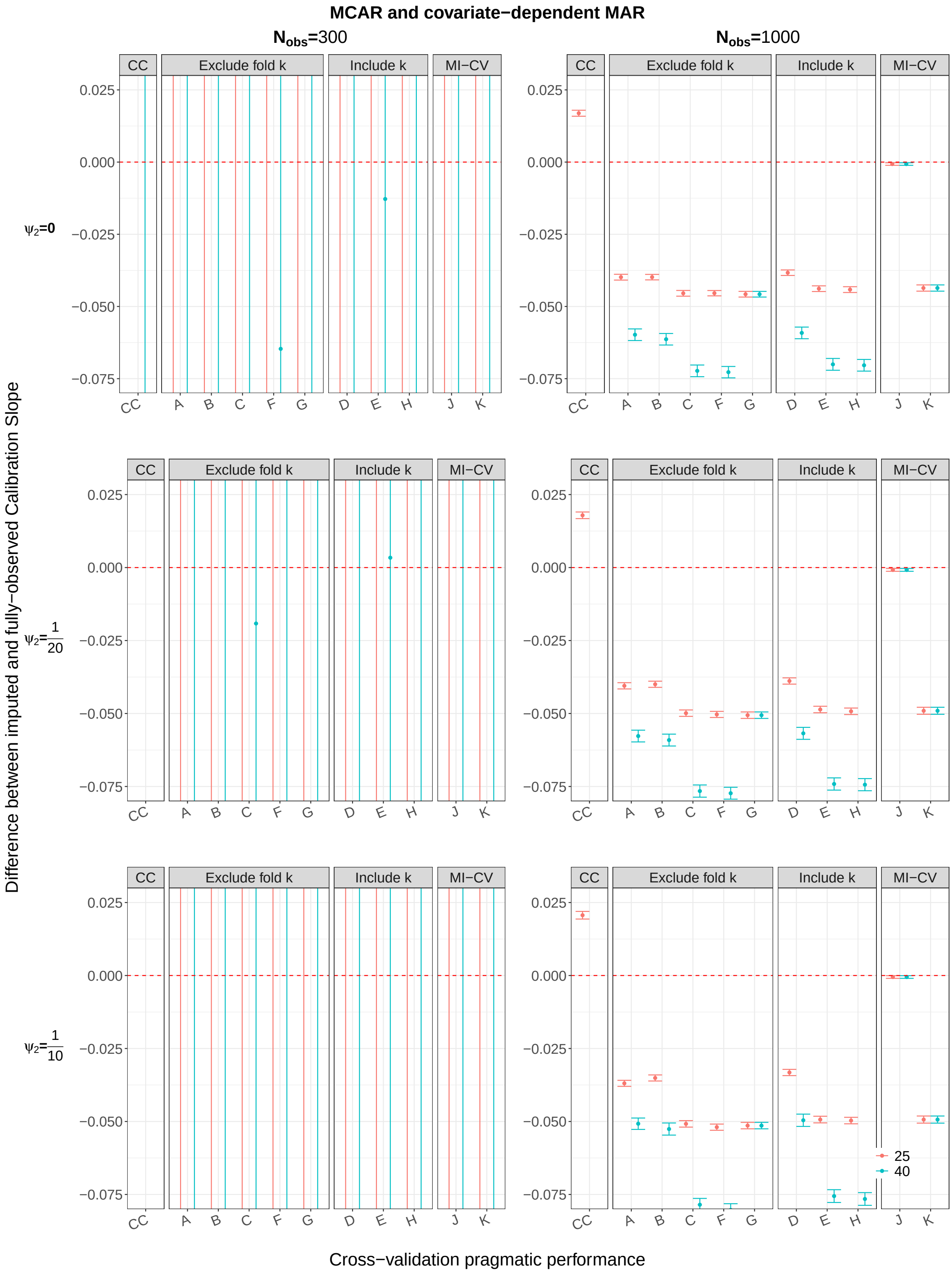

Figure S41: Comparing the impact of increasing the percentage of missingness on the difference Slope_{imp} - Slope_{obs} when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. Red denotes Slope_{imp} - Slope_{obs} when 25% of X_1 values are missing and blue denotes Slope_{imp} - Slope_{obs} when 40% of X_1 values are missing. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

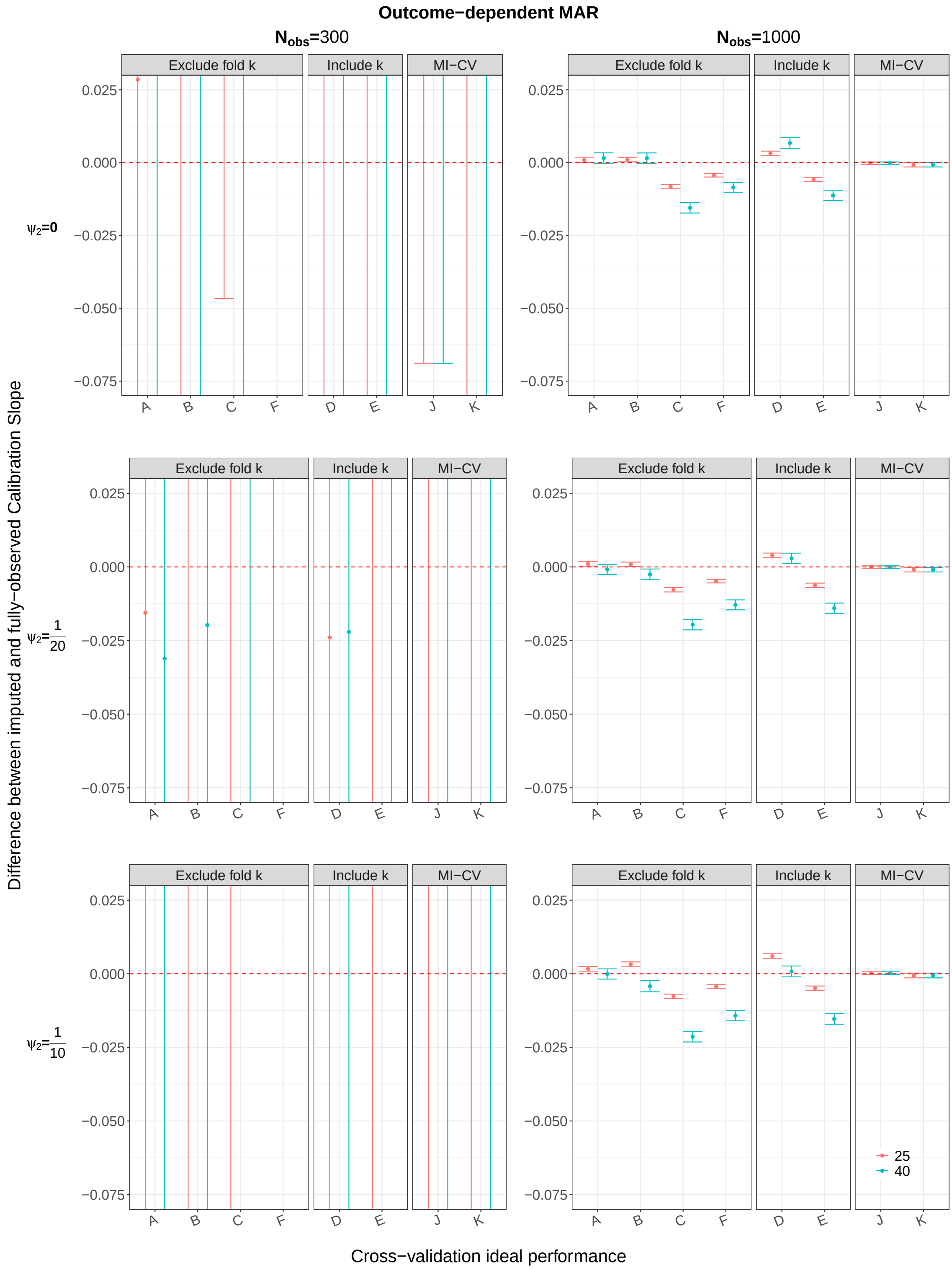

Figure S42: Comparing the impact of increasing the percentage of missingness on the difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. Red denotes Slope$_{imp}$ - Slope$_{obs}$ when 25% of X_1 values are missing and blue denotes Slope$_{imp}$ - Slope$_{obs}$ when 40% of X_1 values are missing. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

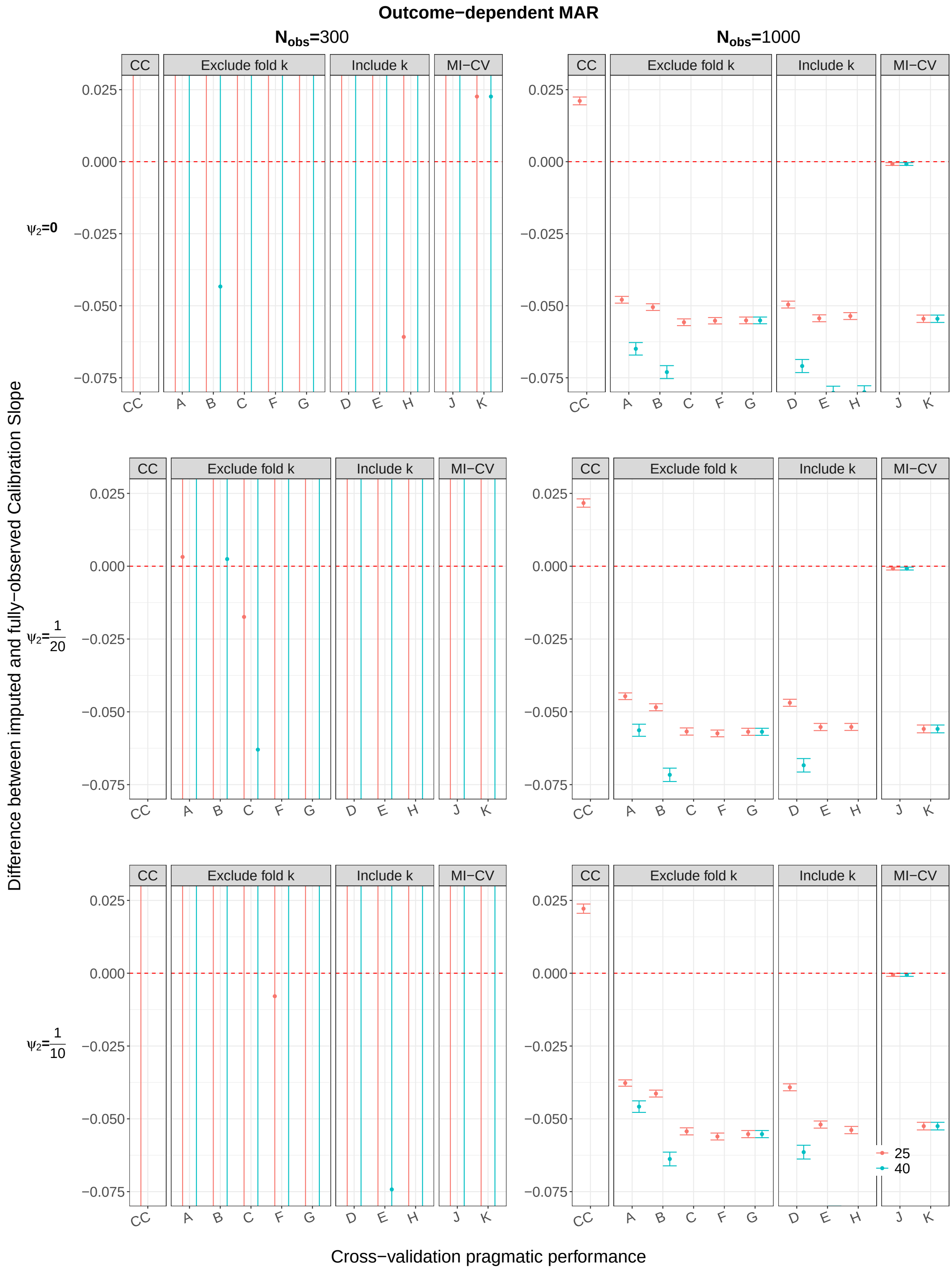

Figure S43: Comparing the impact of increasing the percentage of missingness on the difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. Red denotes Slope$_{imp}$ - Slope$_{obs}$ when 25% of X_1 values are missing and blue denotes Slope$_{imp}$ - Slope$_{obs}$ when 40% of X_1 values are missing. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3.

S2.4.3 Comparing M=5 versus M=25 (slope$_{imp}$-slope$_{obs}$)

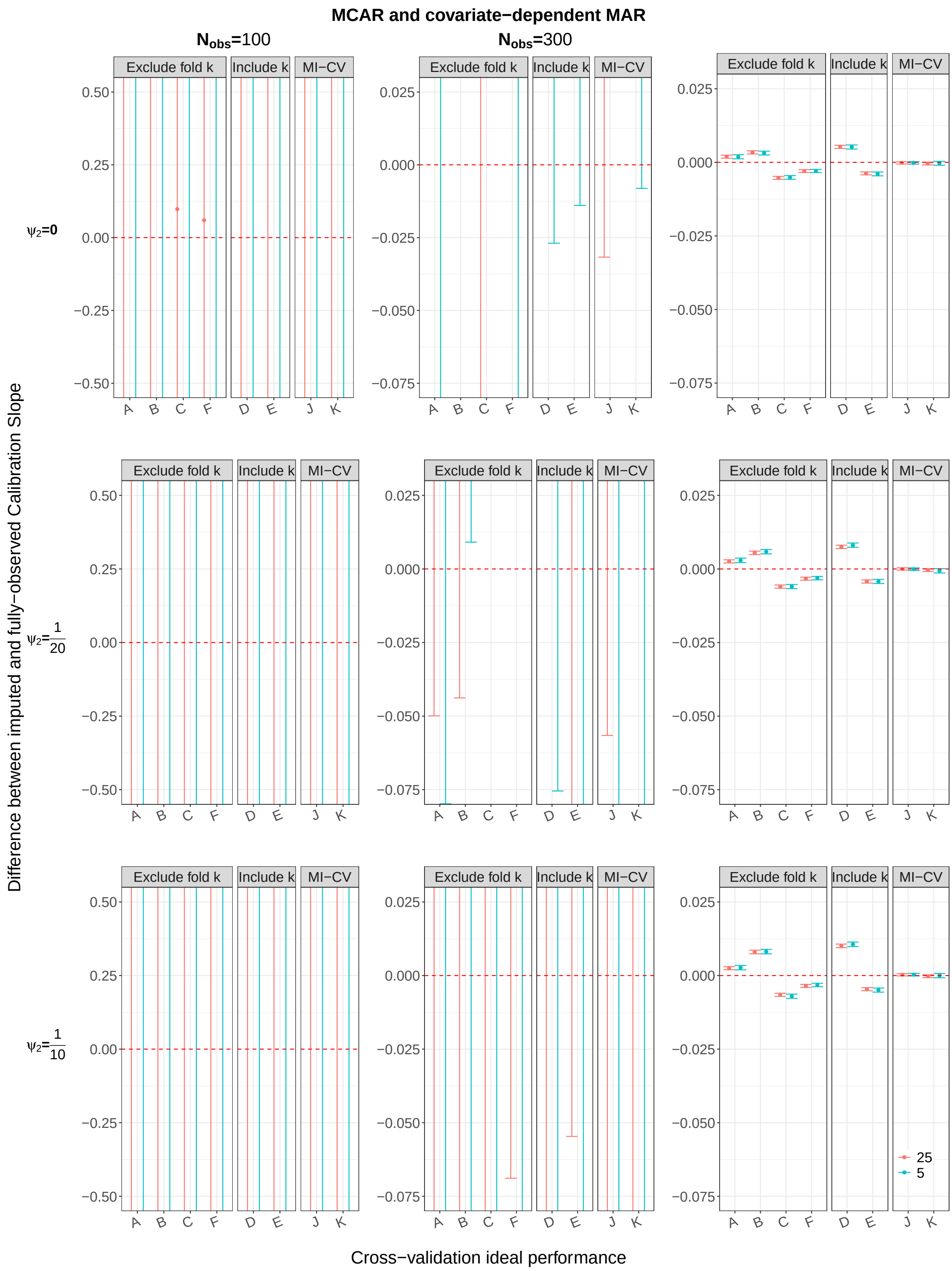

Figure S44: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

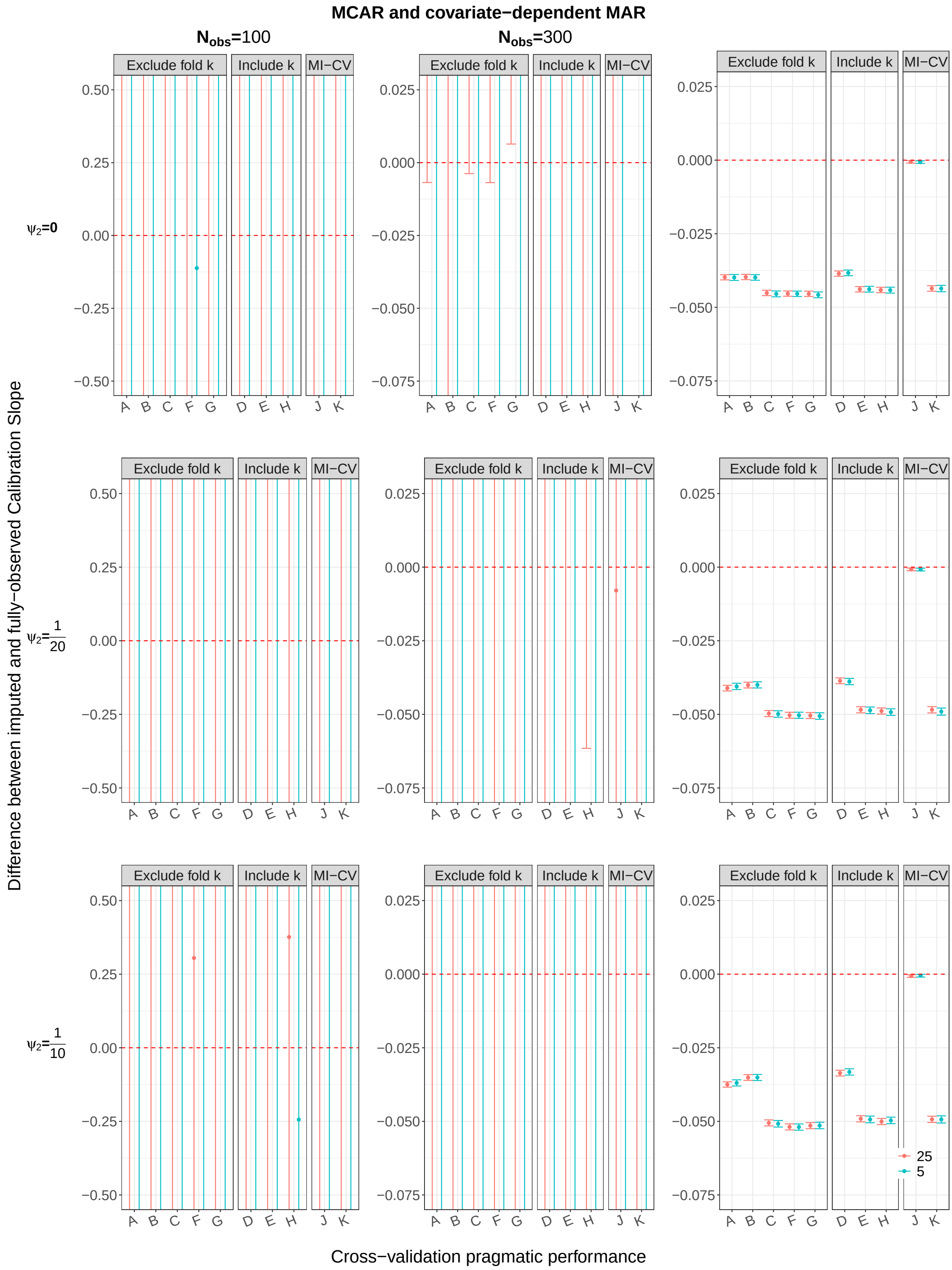

Figure S45: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

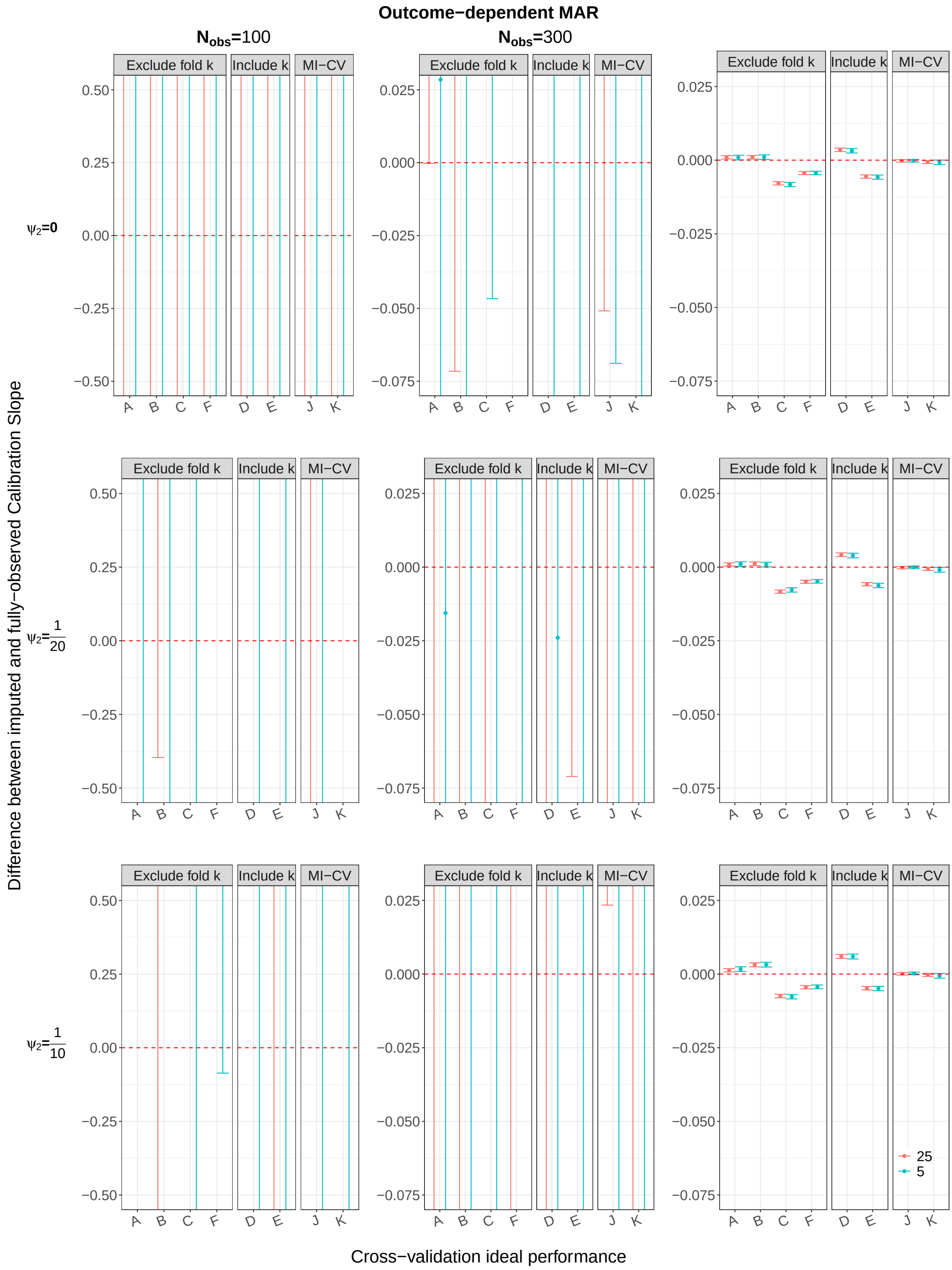

Figure S46: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome-and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

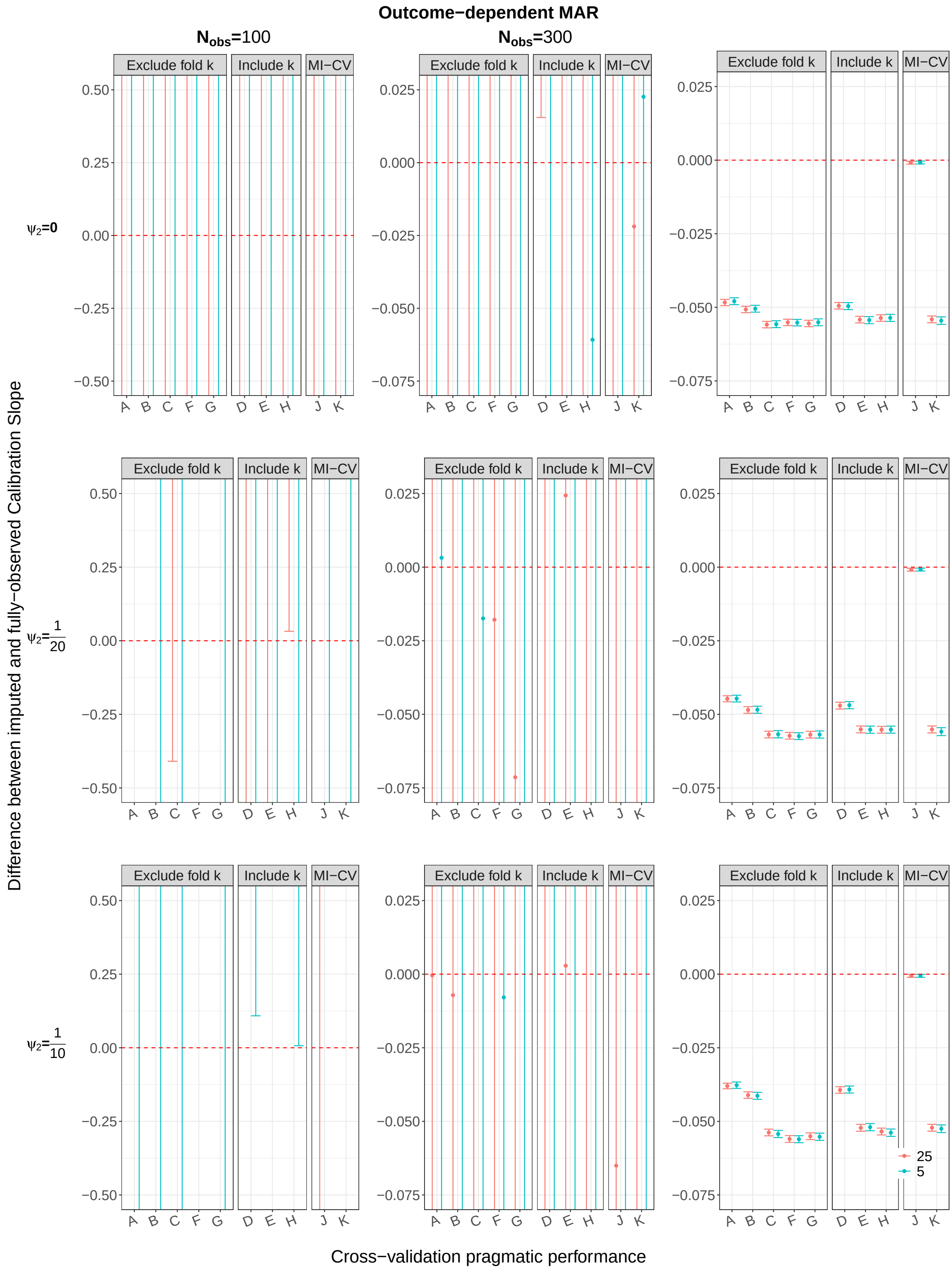

Figure S47: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. The average Calibration slope when data are fully-observed is 1.04 for larger sample sizes. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.4.4 Calibration slope from imputation methods compared to the target calibration slope (slope$_{target}$) using a larger validation set

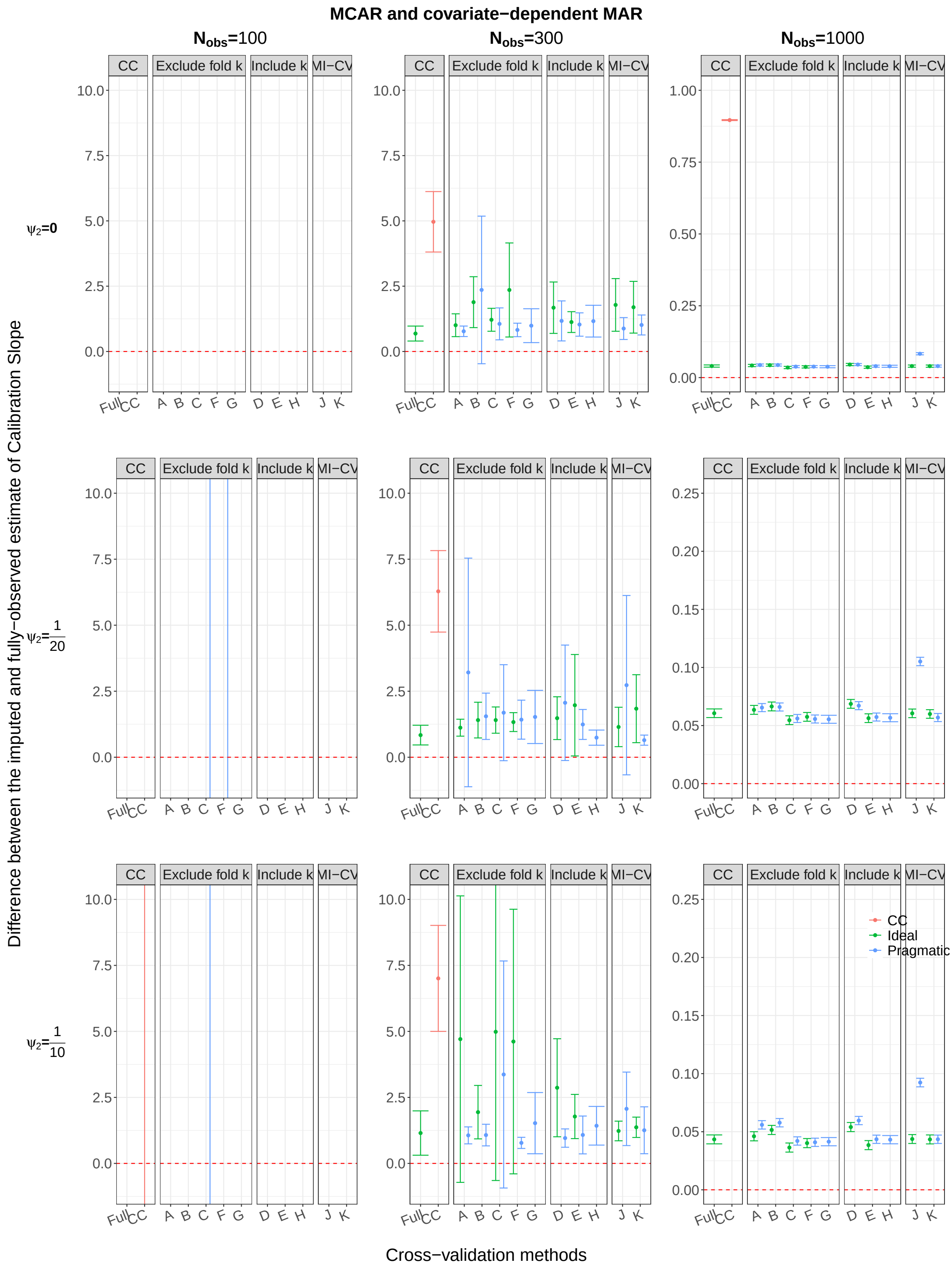

Figure S48: The difference Slope$_{imp}$ - Slope$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{target}$. The average Calibration slope when data are fully-observed is 1.8 for a sample size of 300 and 1.04 for a sample size 1000. CC (complete-case); methods A-K are described in Table 2.3.

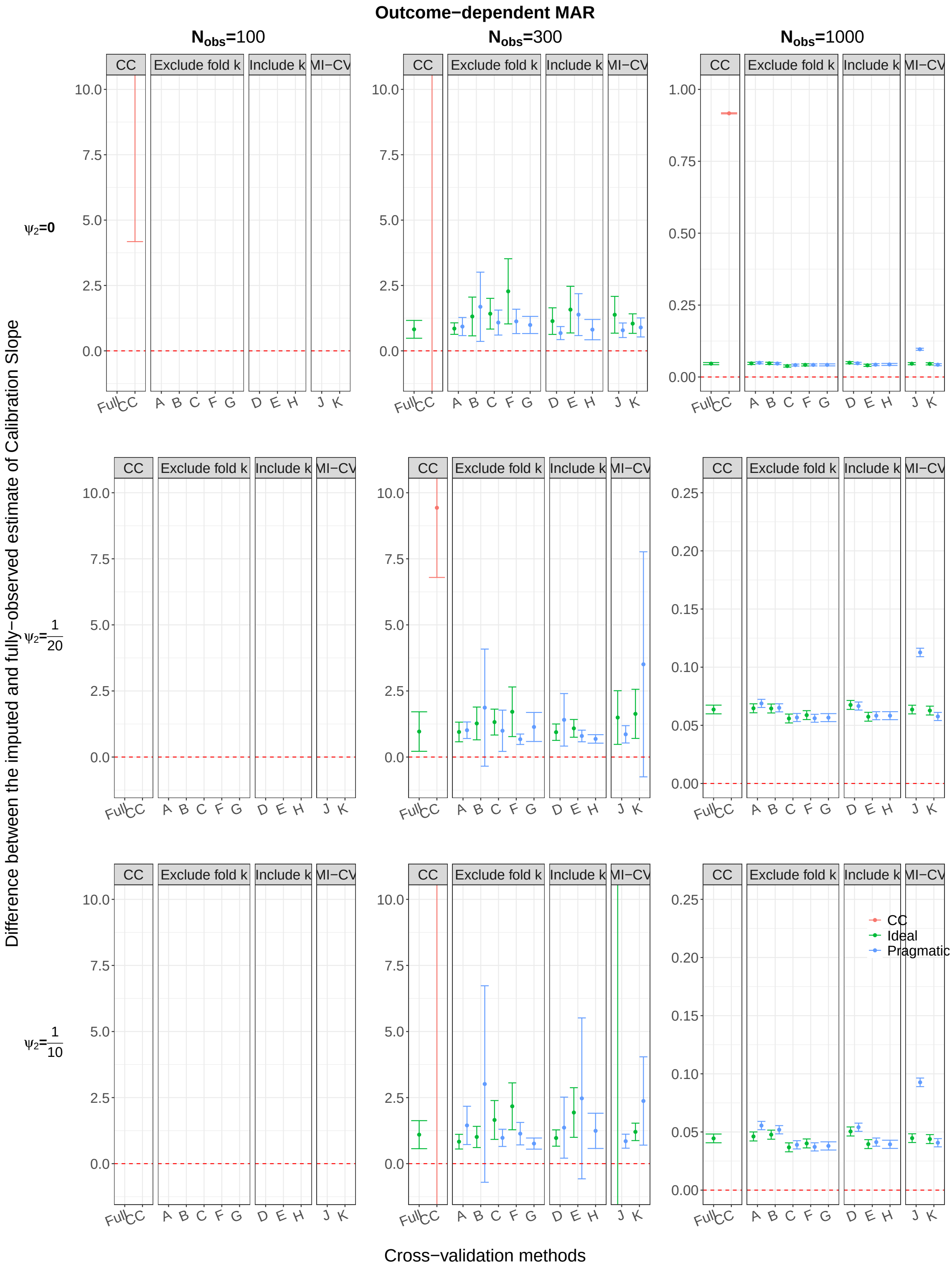

Figure S49: The difference Slope$_{imp}$ - Slope$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{target}$. The average Calibration slope when data are fully-observed is 1.8 for a sample size of 300 and 1.04 for a sample size 1000. CC (complete-case); methods A-K are described in Table 2.3.

S2.5 Is data leakage an issue?

S2.5.1 AUC

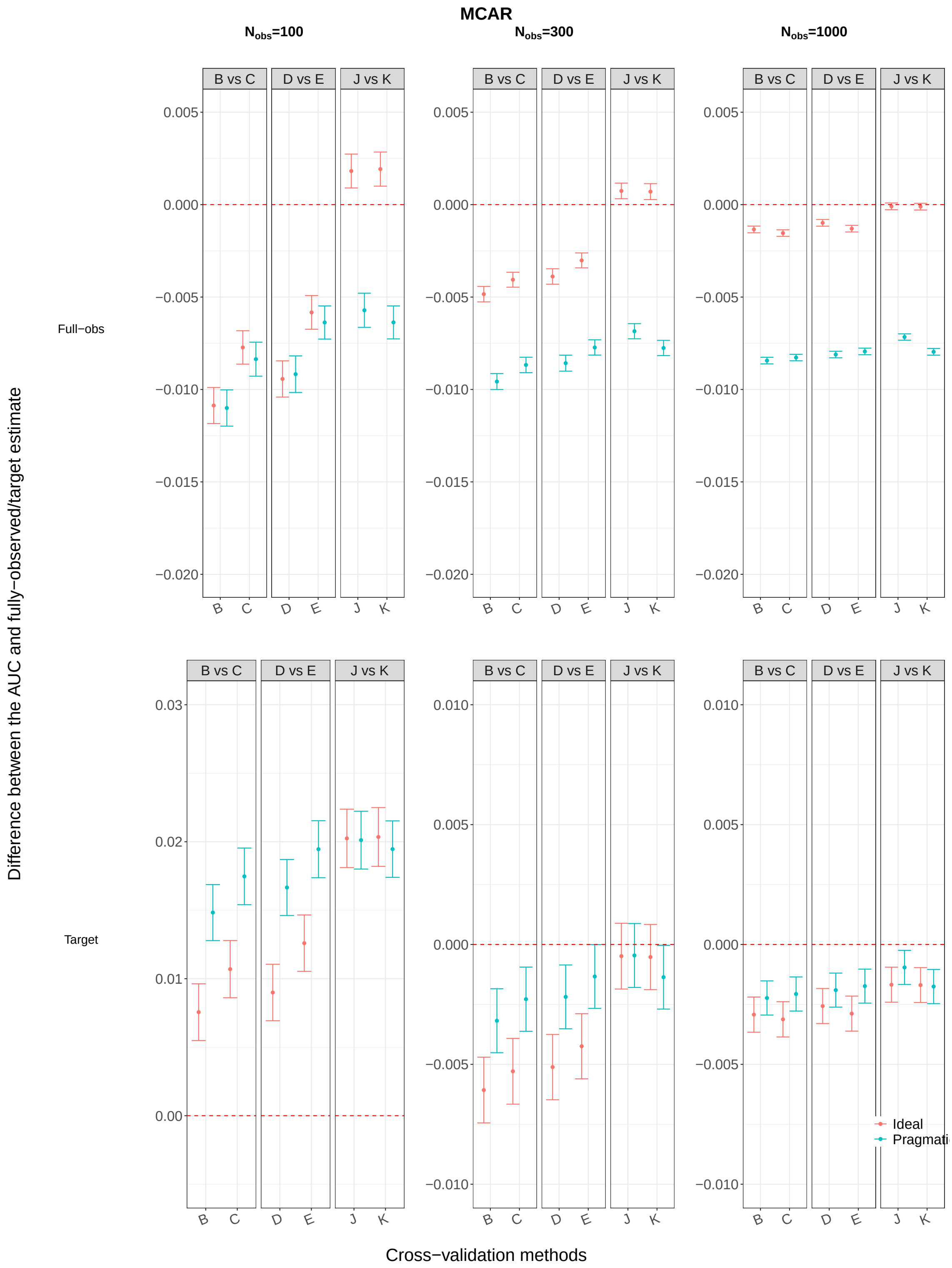

Figure S50: Assessing data leakage within the imputation process for cross-validation. The differences AUC$_{imp}$ - AUC$_{obs}$ and AUC$_{imp}$ - AUC$_{target}$ are compared when data are MCAR. Methods are compared to both the AUC estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

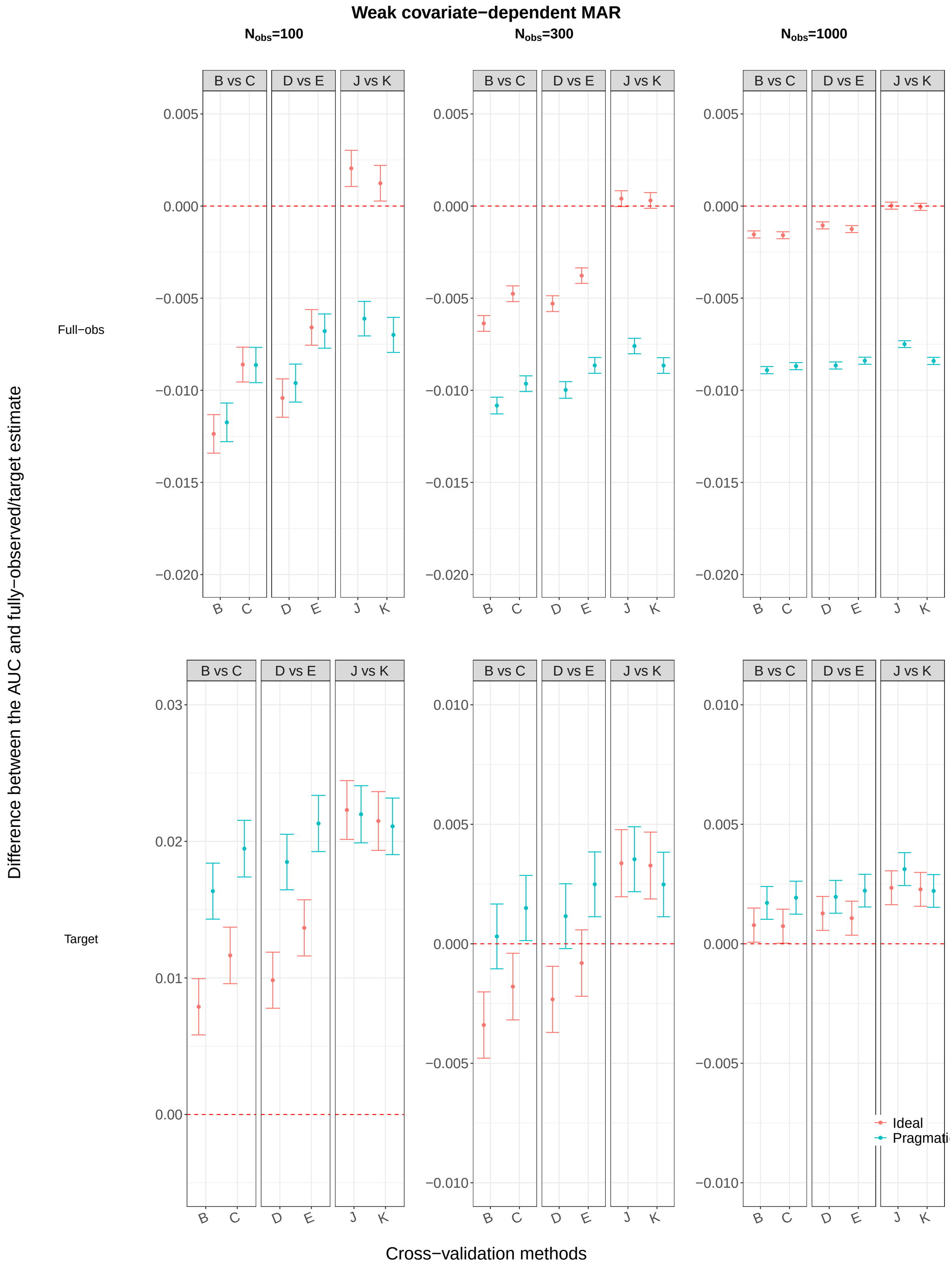

Figure S51: Assessing data leakage within the imputation process for cross-validation. The differences AUC$_{imp}$ - AUC$_{obs}$ and AUC$_{imp}$ - AUC$_{target}$ are compared when data are weak covariate-dependent MAR. Methods are compared to both the AUC estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

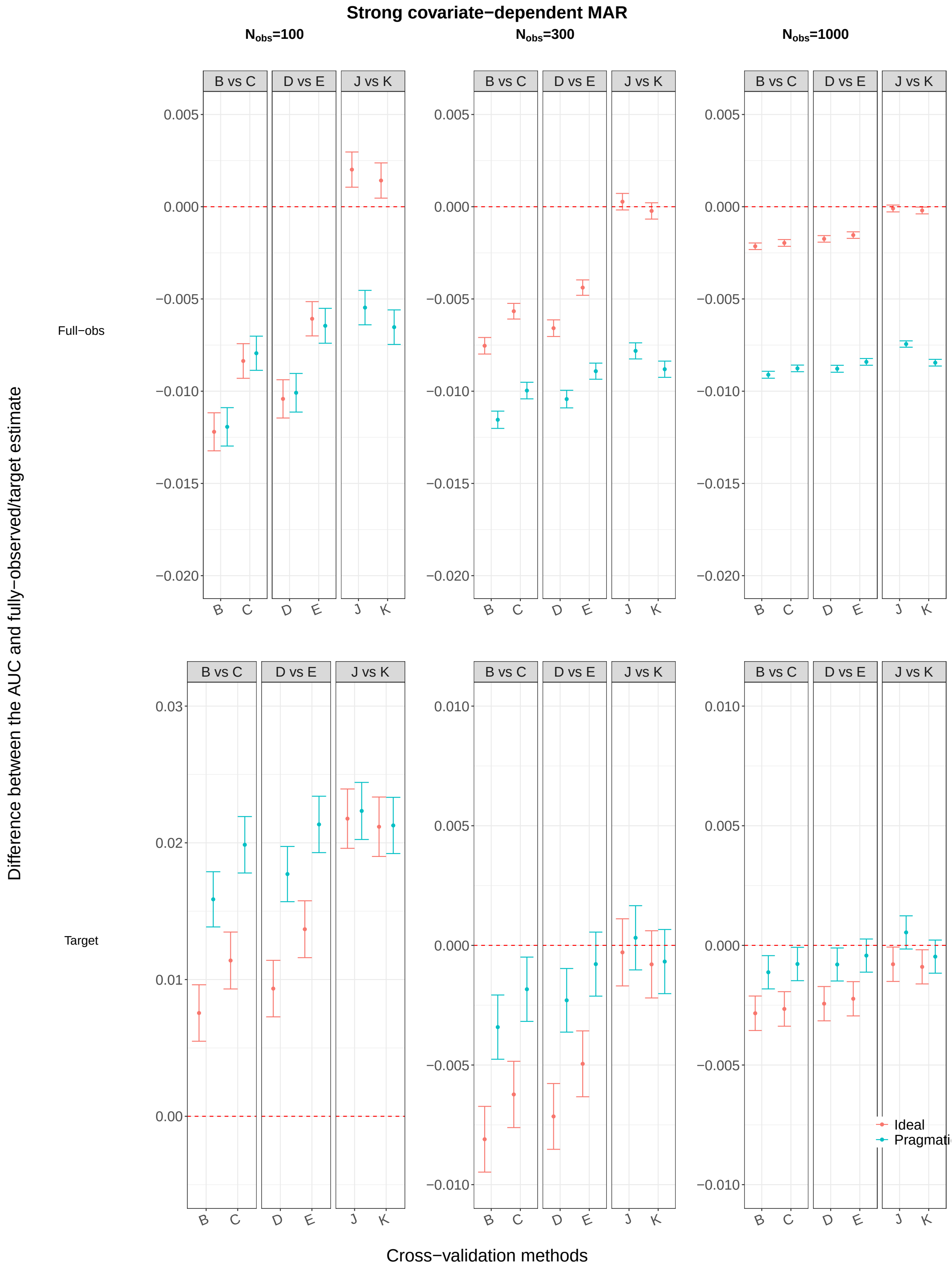

Figure S52: Assessing data leakage within the imputation process for cross-validation. The differences AUC$_{imp}$ - AUC$_{obs}$ and AUC$_{imp}$ - AUC$_{target}$ are compared when data are strong covariate-dependent MAR. Methods are compared to both the AUC estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

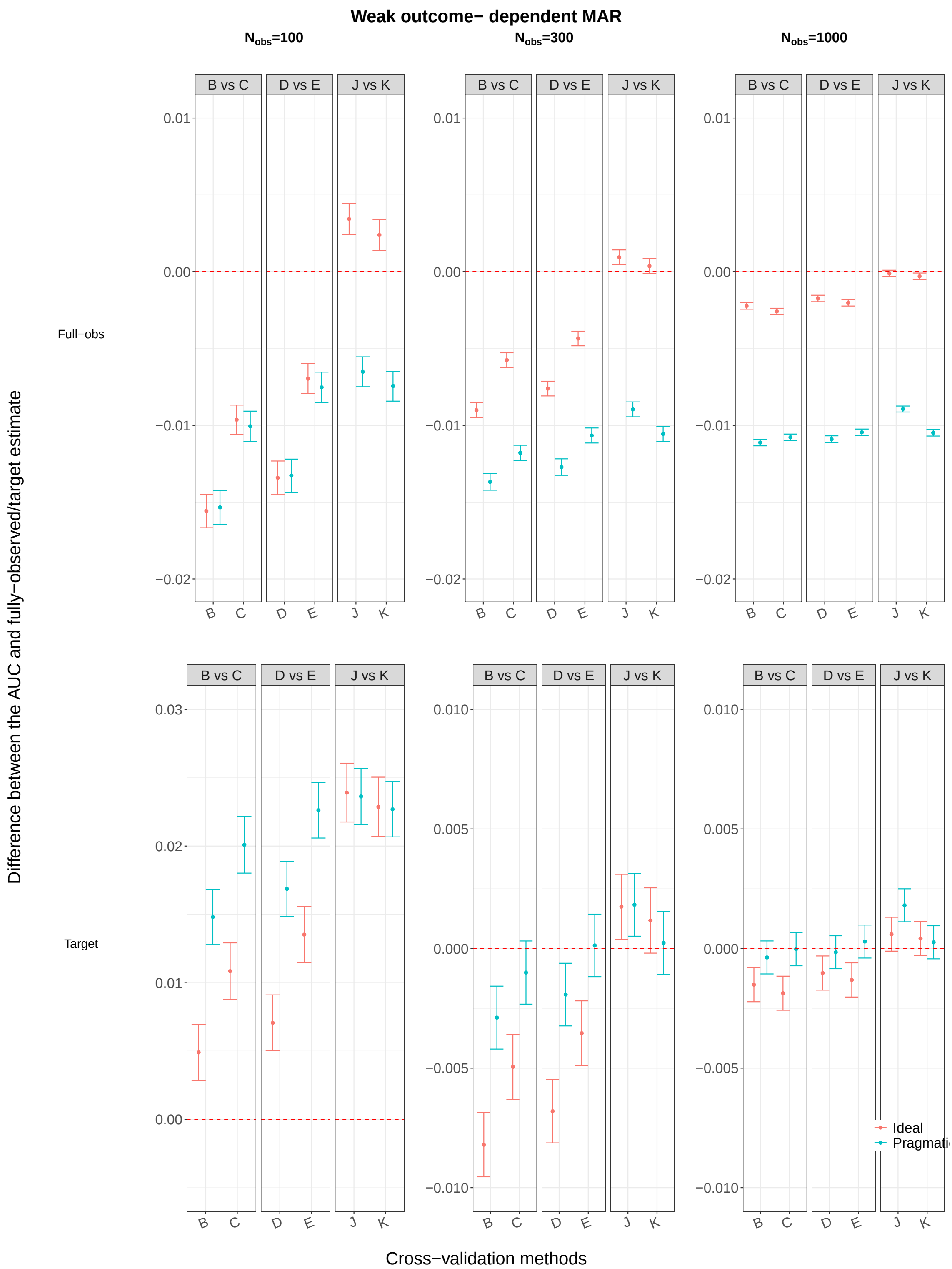

Figure S53: Assessing data leakage within the imputation process for cross-validation. The differences AUC$_{imp}$ - AUC$_{obs}$ and AUC$_{imp}$ - AUC$_{target}$ are compared when data are weak outcome-dependent MAR. Methods are compared to both the AUC estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

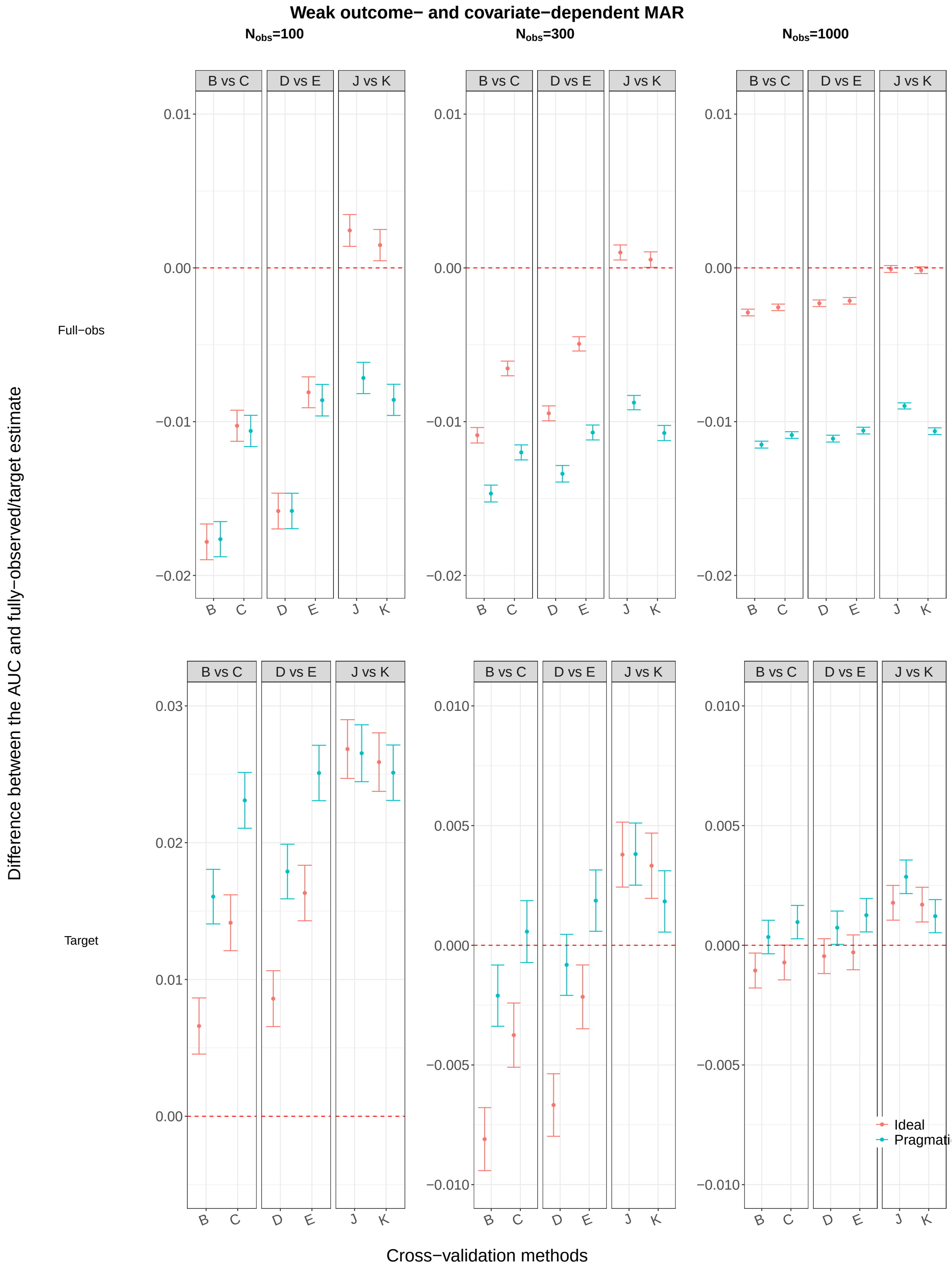

Figure S54: Assessing data leakage within the imputation process for cross-validation. The differences AUC$_{imp}$ - AUC$_{obs}$ and AUC$_{imp}$ - AUC$_{target}$ are compared when data are weak outcome- and covariate-dependent MAR. Methods are compared to both the AUC estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

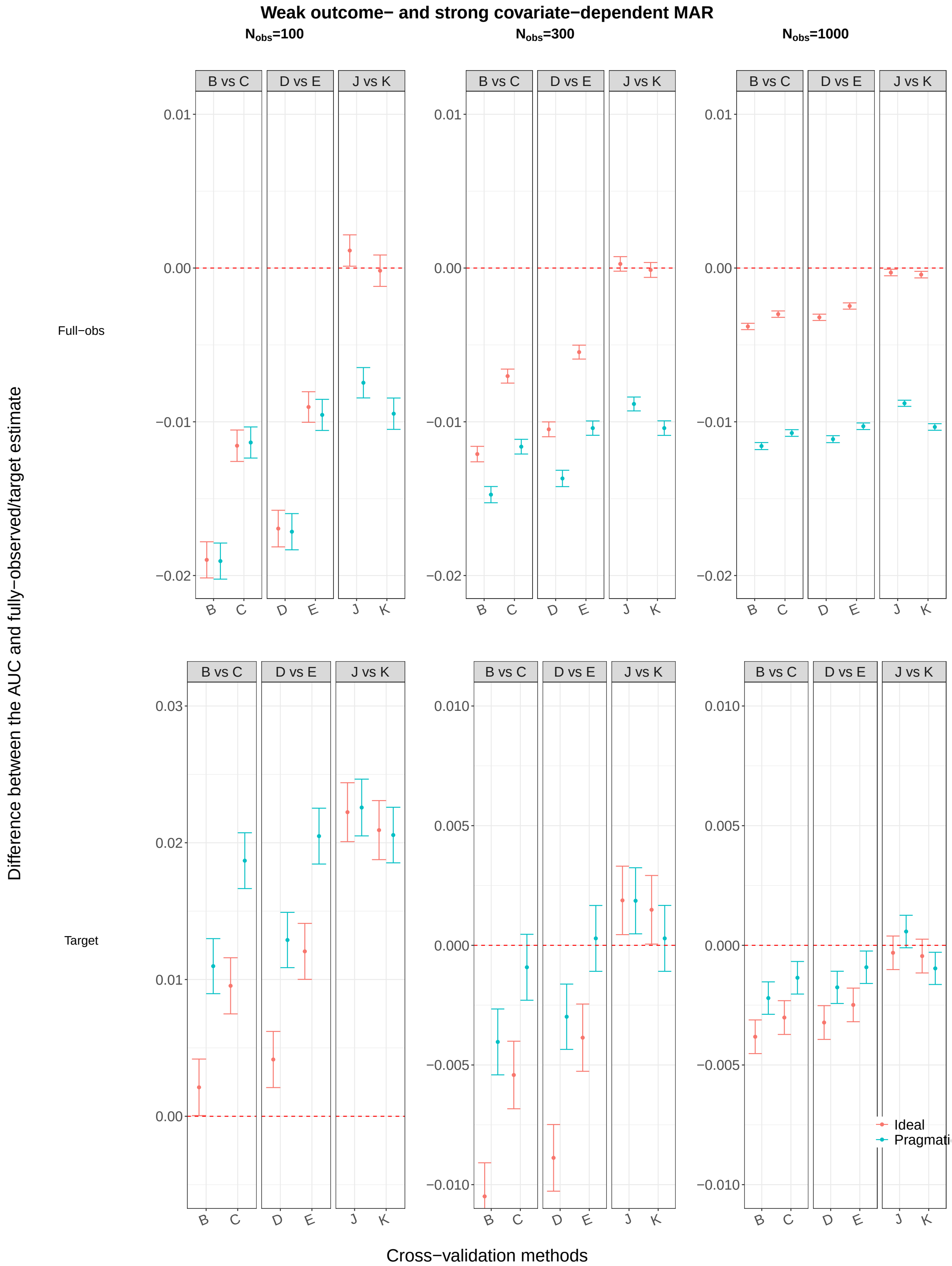

Figure S55: Assessing data leakage within the imputation process for cross-validation. The differences AUC$_{imp}$ - AUC$_{obs}$ and AUC$_{imp}$ - AUC$_{target}$ are compared when data are weak outcome- and strong covariate-dependent MAR. Methods are compared to both the AUC estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.5.2 Brier Score

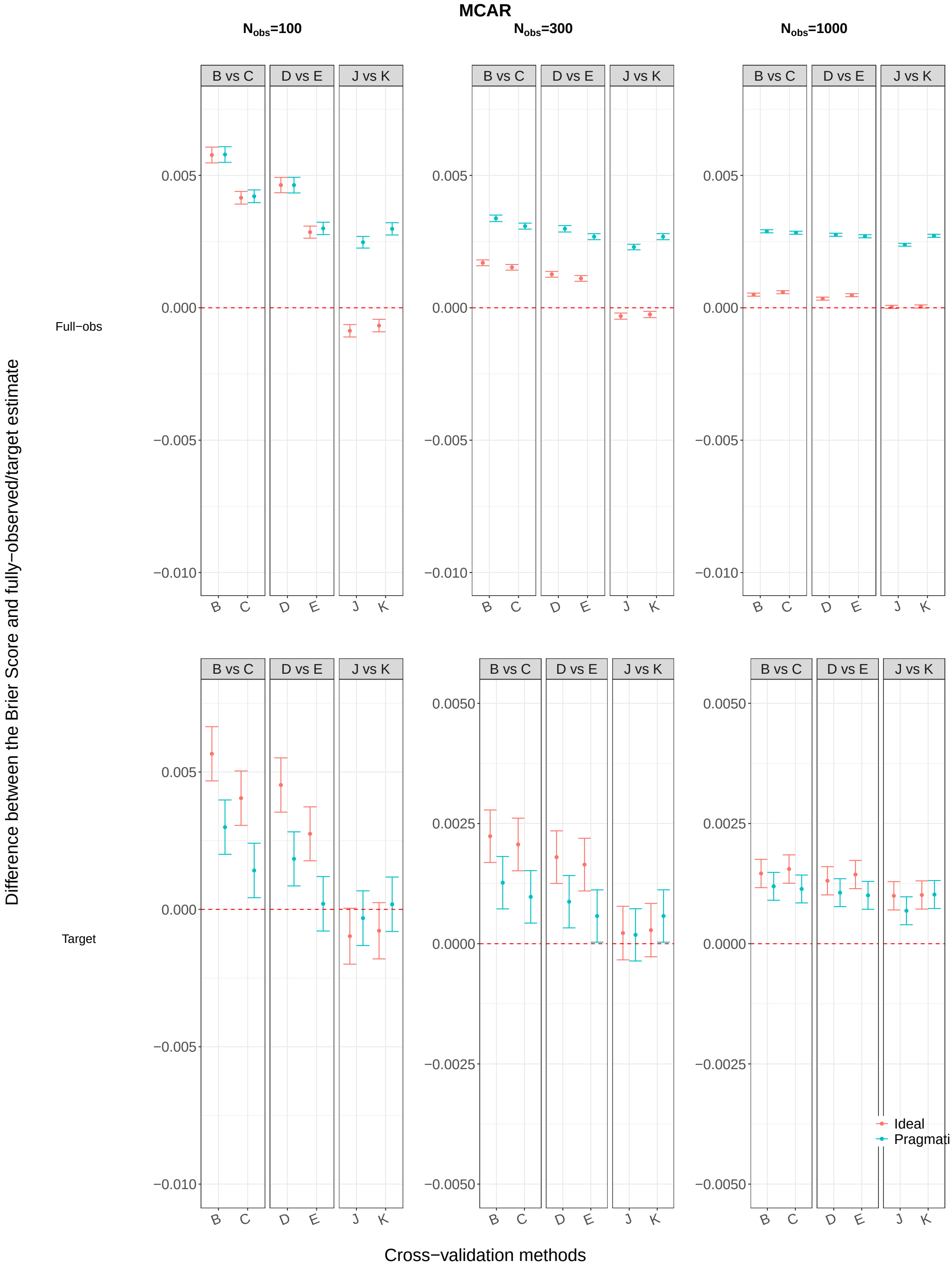

Figure S56: Assessing data leakage within the imputation process for cross-validation. The differences Brier_{imp} - Brier_{obs} and Brier_{imp} - Brier_{target} are compared when data are MCAR. Methods are compared to both the Brier score estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

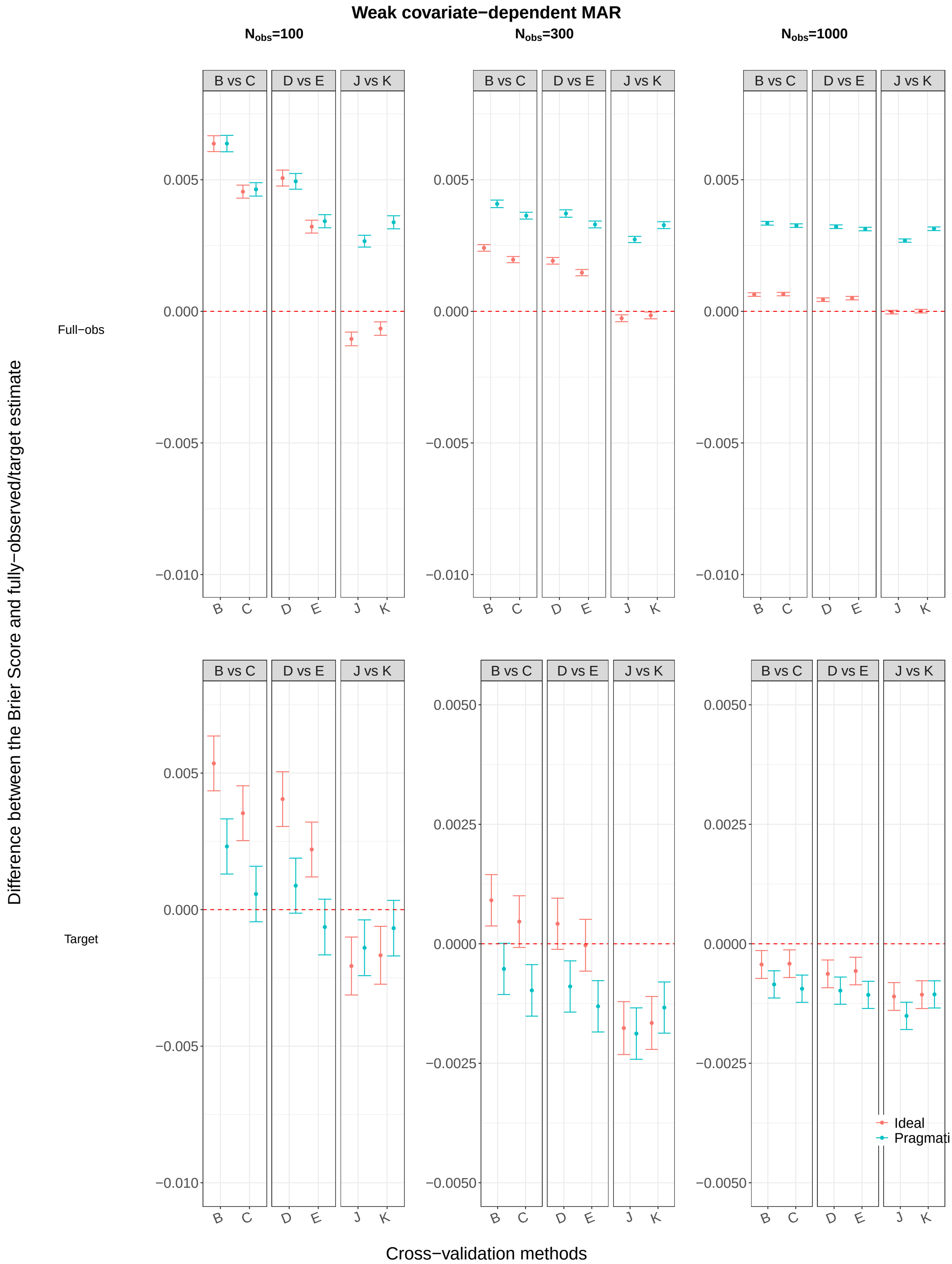

Figure S57: Assessing data leakage within the imputation process for cross-validation. The differences Brier$_{imp}$ - Brier$_{obs}$ and Brier$_{imp}$ - Brier$_{target}$ are compared when data are weak covariate-dependent MAR. Methods are compared to both the Brier score estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

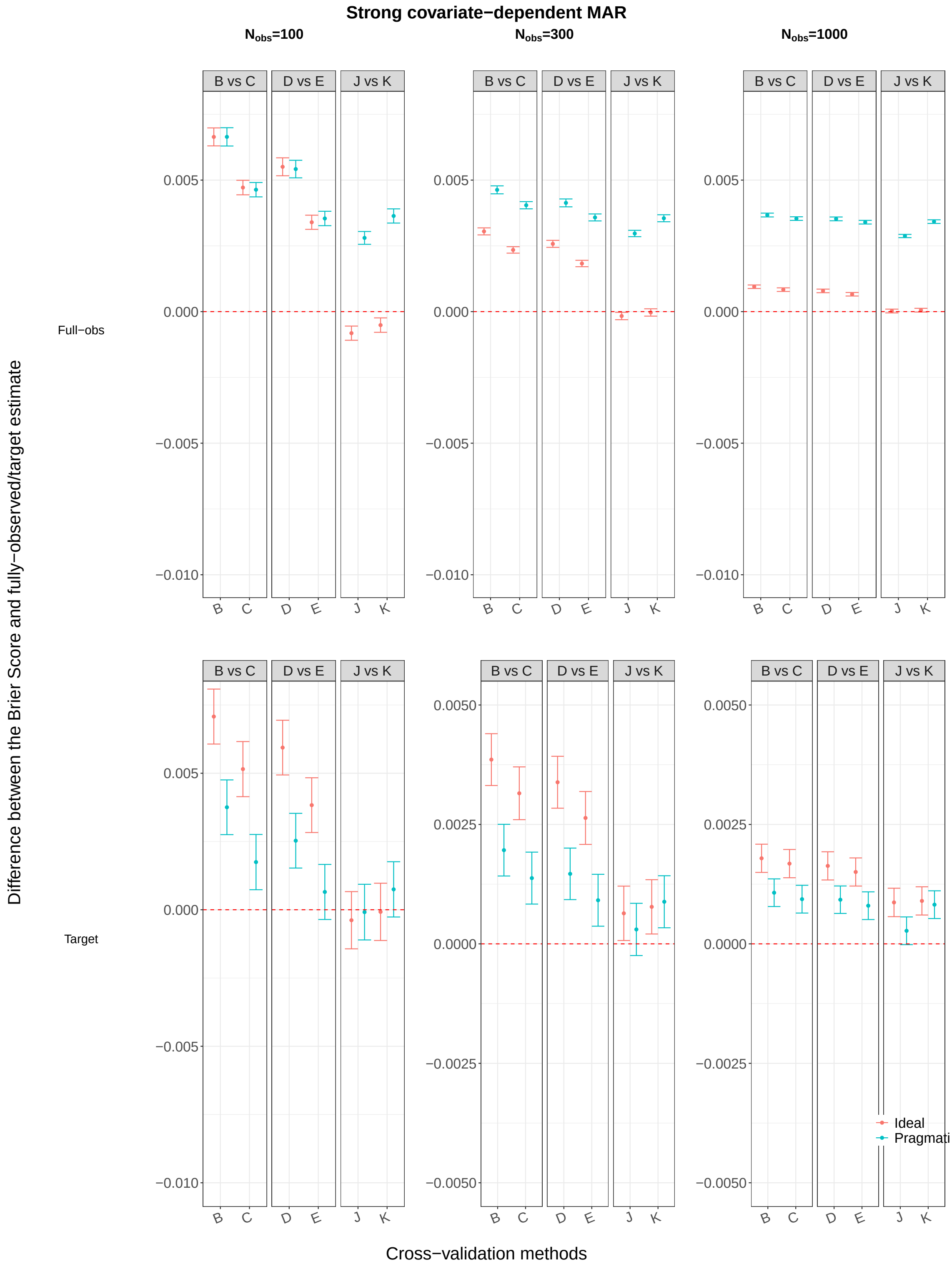

Figure S58: Assessing data leakage within the imputation process for cross-validation. The differences Brier$_{imp}$ - Brier$_{obs}$ and Brier$_{imp}$ - Brier$_{target}$ are compared when data are strong covariate-dependent MAR. Methods are compared to both the Brier score estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

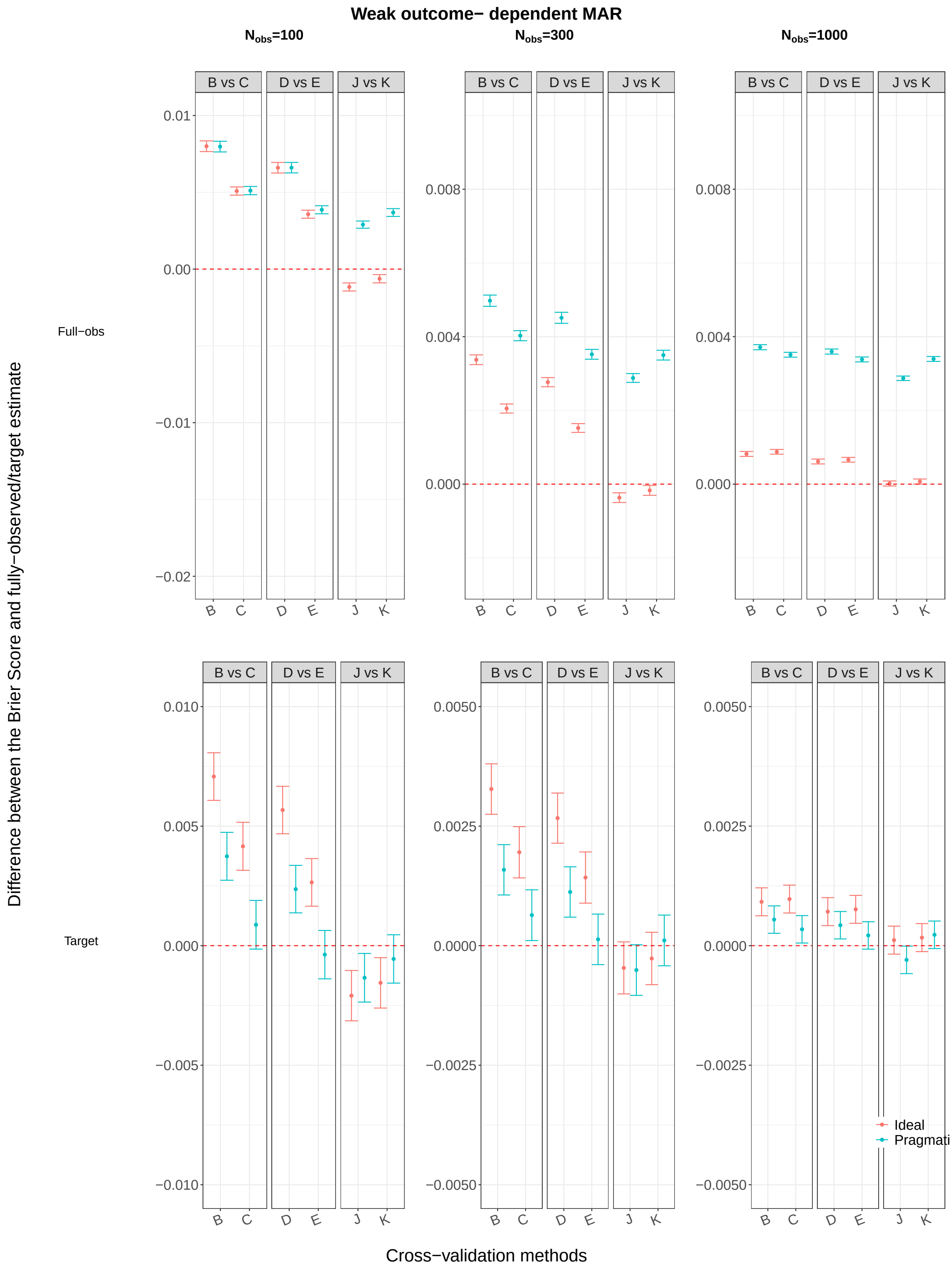

Figure S59: Assessing data leakage within the imputation process for cross-validation. The differences Brier$_{imp}$ - Brier$_{obs}$ and Brier$_{imp}$ - Brier$_{target}$ are compared when data are weak outcome-dependent MAR. Methods are compared to both the Brier score estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

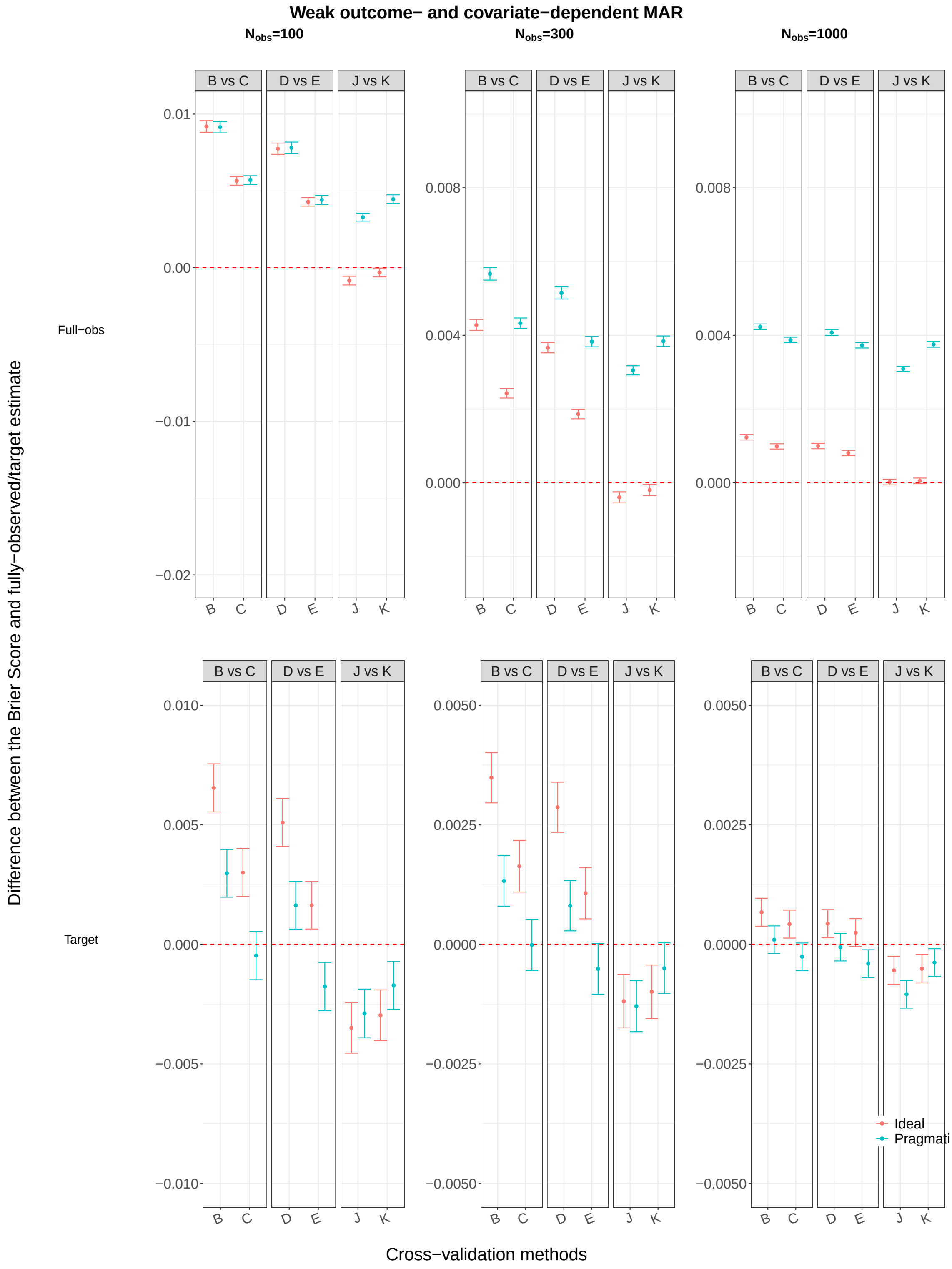

Figure S60: Assessing data leakage within the imputation process for cross-validation. The differences Brier$_{imp}$ - Brier$_{obs}$ and Brier$_{imp}$ - Brier$_{target}$ are compared when data are weak outcome- and covariate-dependent MAR. Methods are compared to both the Brier score estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

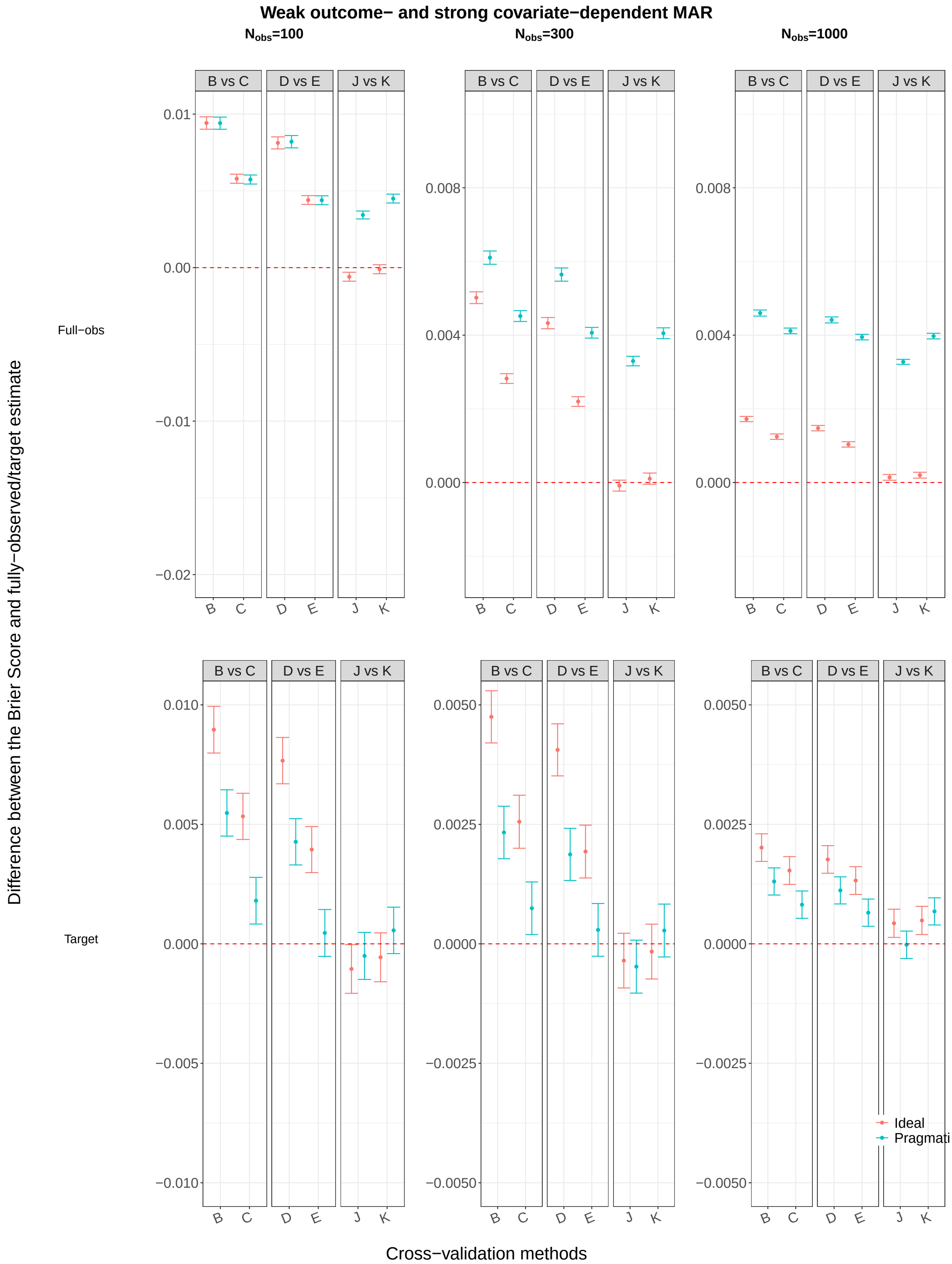

Figure S61: Assessing data leakage within the imputation process for cross-validation. The differences Brier$_{imp}$ - Brier$_{obs}$ and Brier$_{imp}$ - Brier$_{target}$ are compared when data are weak outcome- and strong covariate-dependent MAR. Methods are compared to both the Brier score estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.5.3 Calibration intercept

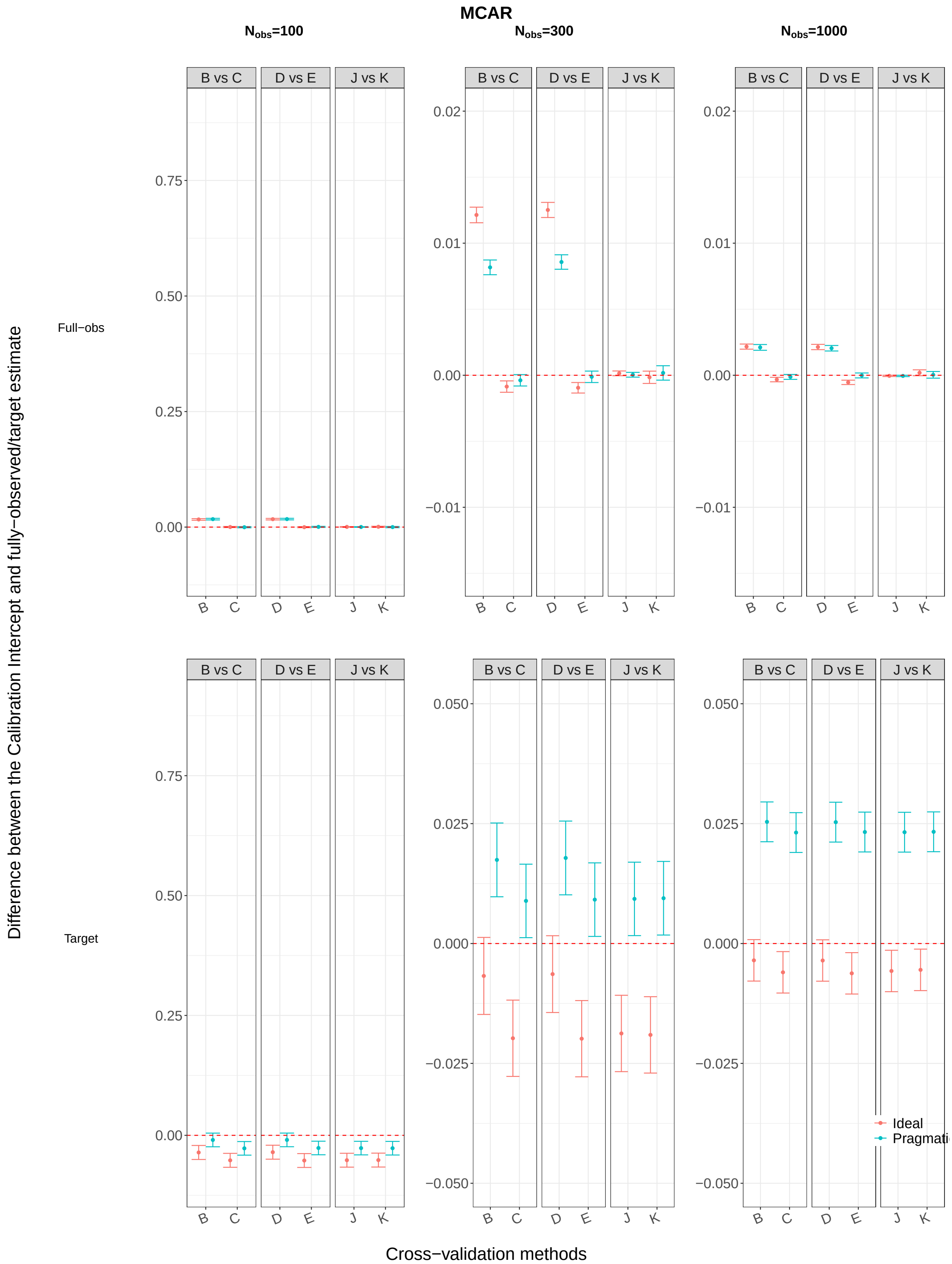

Figure S62: Assessing data leakage within the imputation process for cross-validation. The differences Intercept_{imp} - Intercept_{obs} and Intercept_{imp} - $\text{Intercept}_{target}$ are compared when data are MCAR. Methods are compared to both the calibration intercept estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

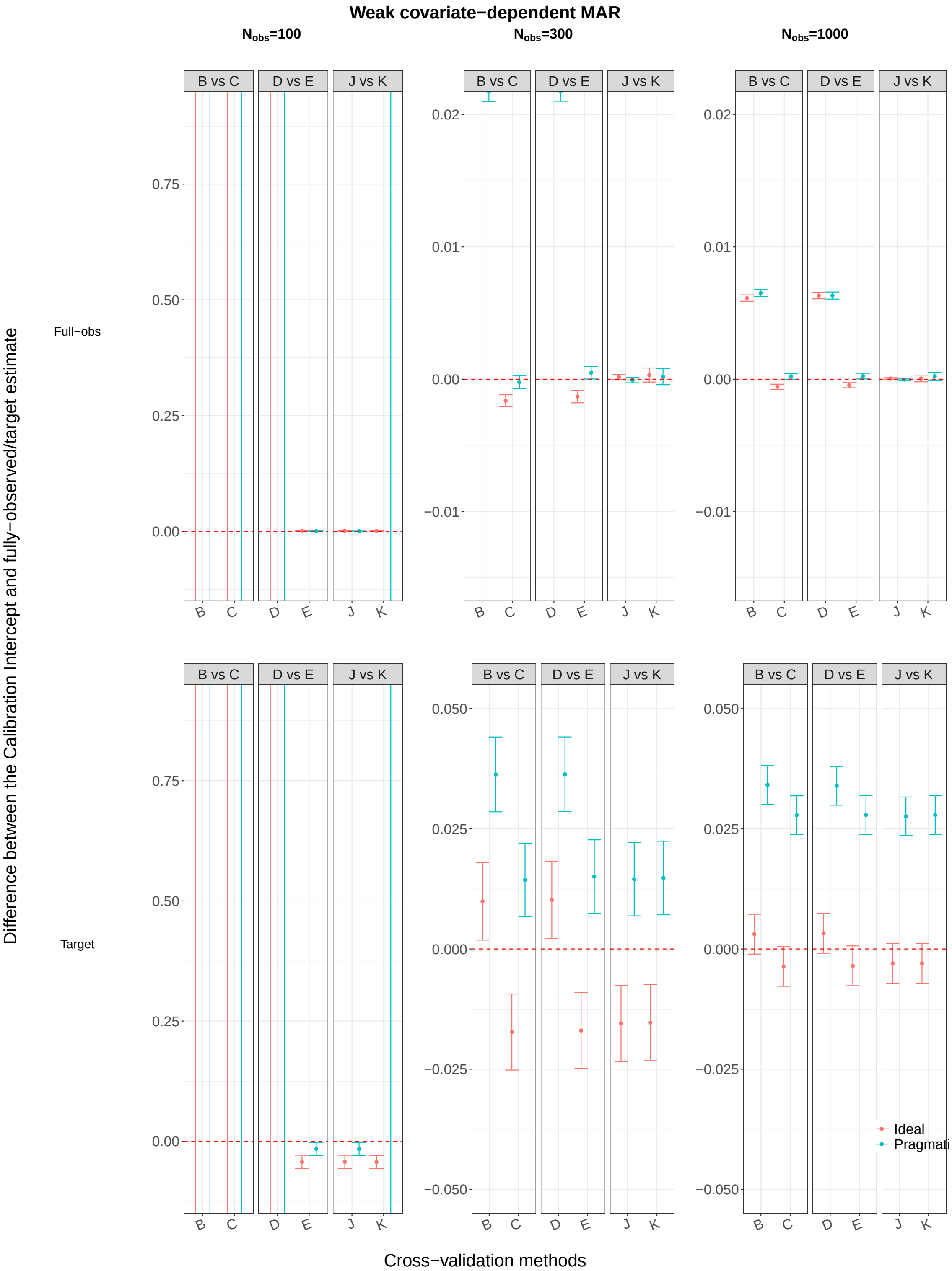

Figure S63: Assessing data leakage within the imputation process for cross-validation. The differences Intercept$_{imp}$ - Intercept$_{obs}$ and Intercept$_{imp}$ - Intercept$_{target}$ are compared when data are weak covariate-dependent MAR. Methods are compared to both the calibration intercept estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

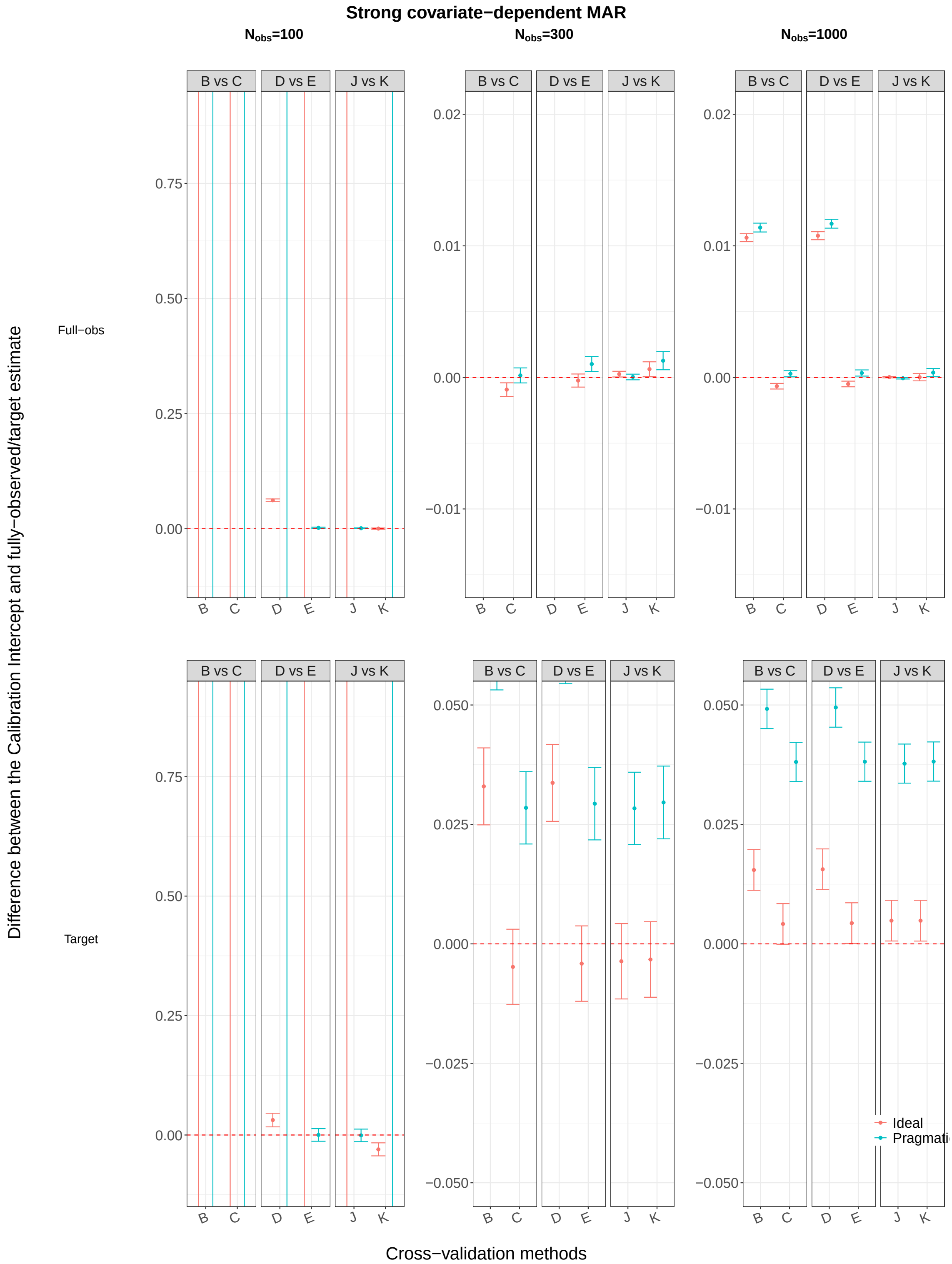

Figure S64: Assessing data leakage within the imputation process for cross-validation. The differences Intercept$_{imp}$ - Intercept$_{obs}$ and Intercept$_{imp}$ - Intercept$_{target}$ are compared when data are strong covariate-dependent MAR. Methods are compared to both the calibration intercept estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

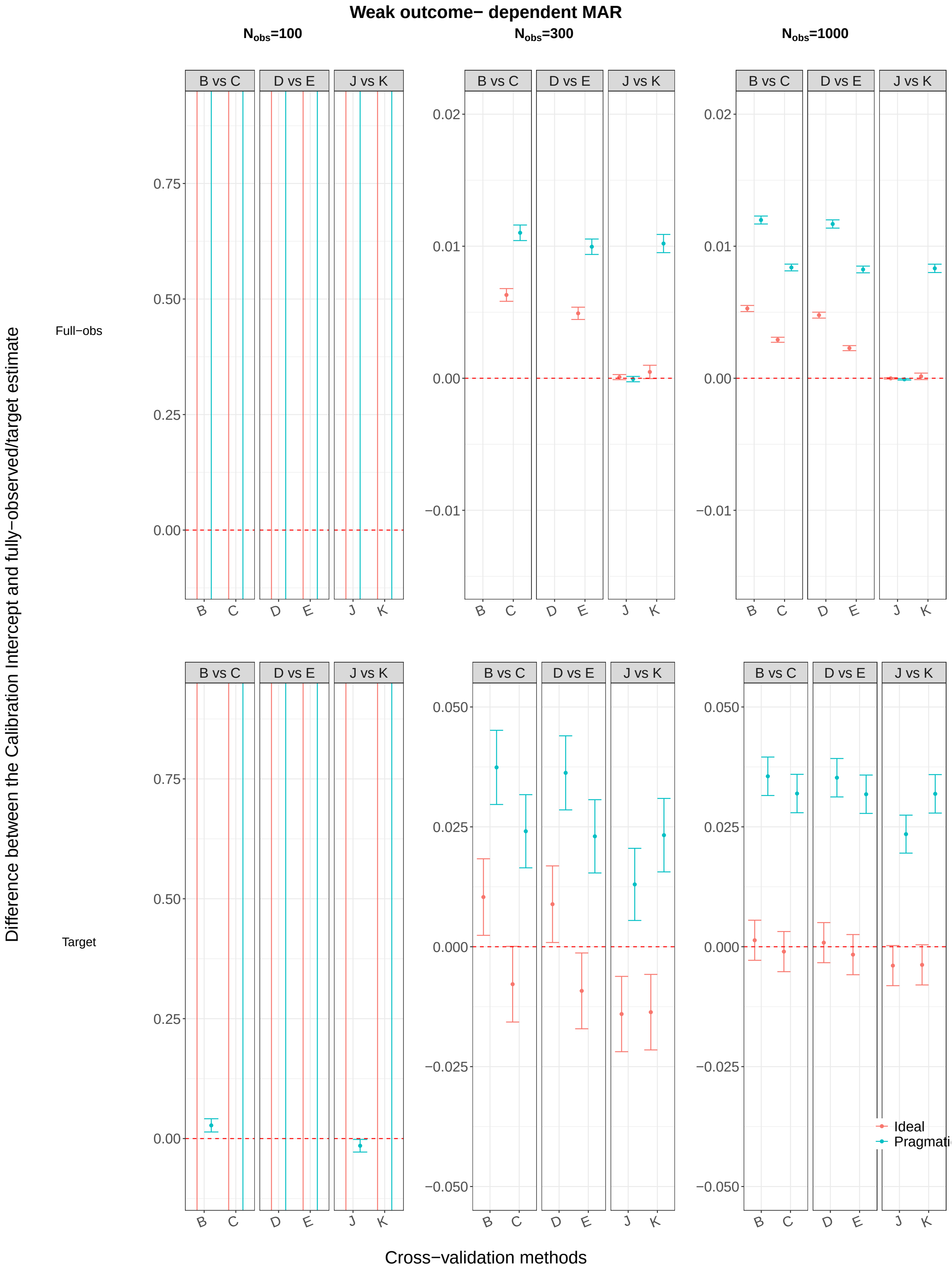

Figure S65: Assessing data leakage within the imputation process for cross-validation. The differences Intercept_{imp} - Intercept_{obs} and Intercept_{imp} - $\text{Intercept}_{target}$ are compared when data are weak outcome-dependent MAR. Methods are compared to both the calibration intercept estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

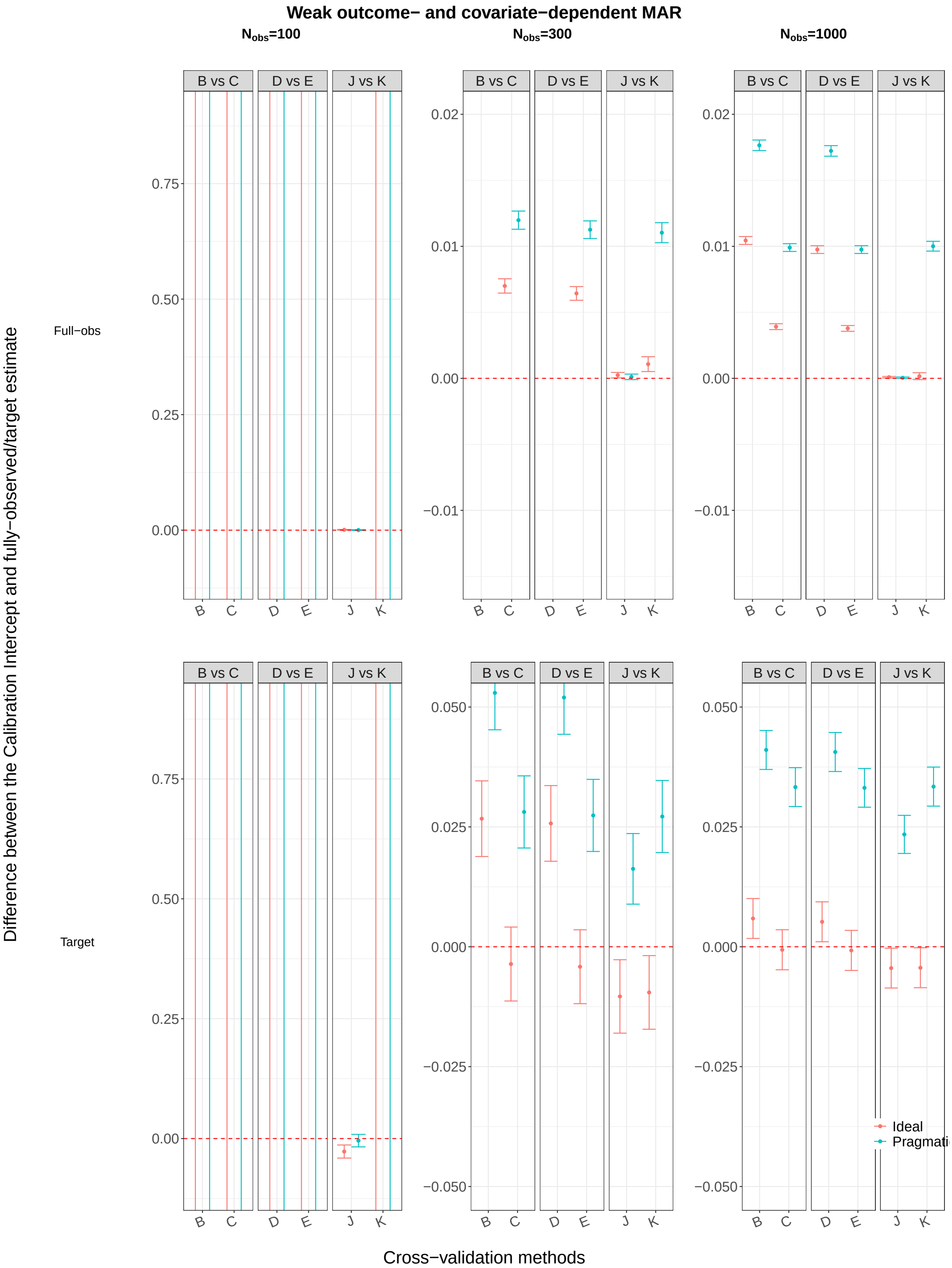

Figure S66: Assessing data leakage within the imputation process for cross-validation. The differences Intercept_{imp} - Intercept_{obs} and Intercept_{imp} - $\text{Intercept}_{target}$ are compared when data are weak outcome- and covariate-dependent MAR. Methods are compared to both the calibration intercept estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

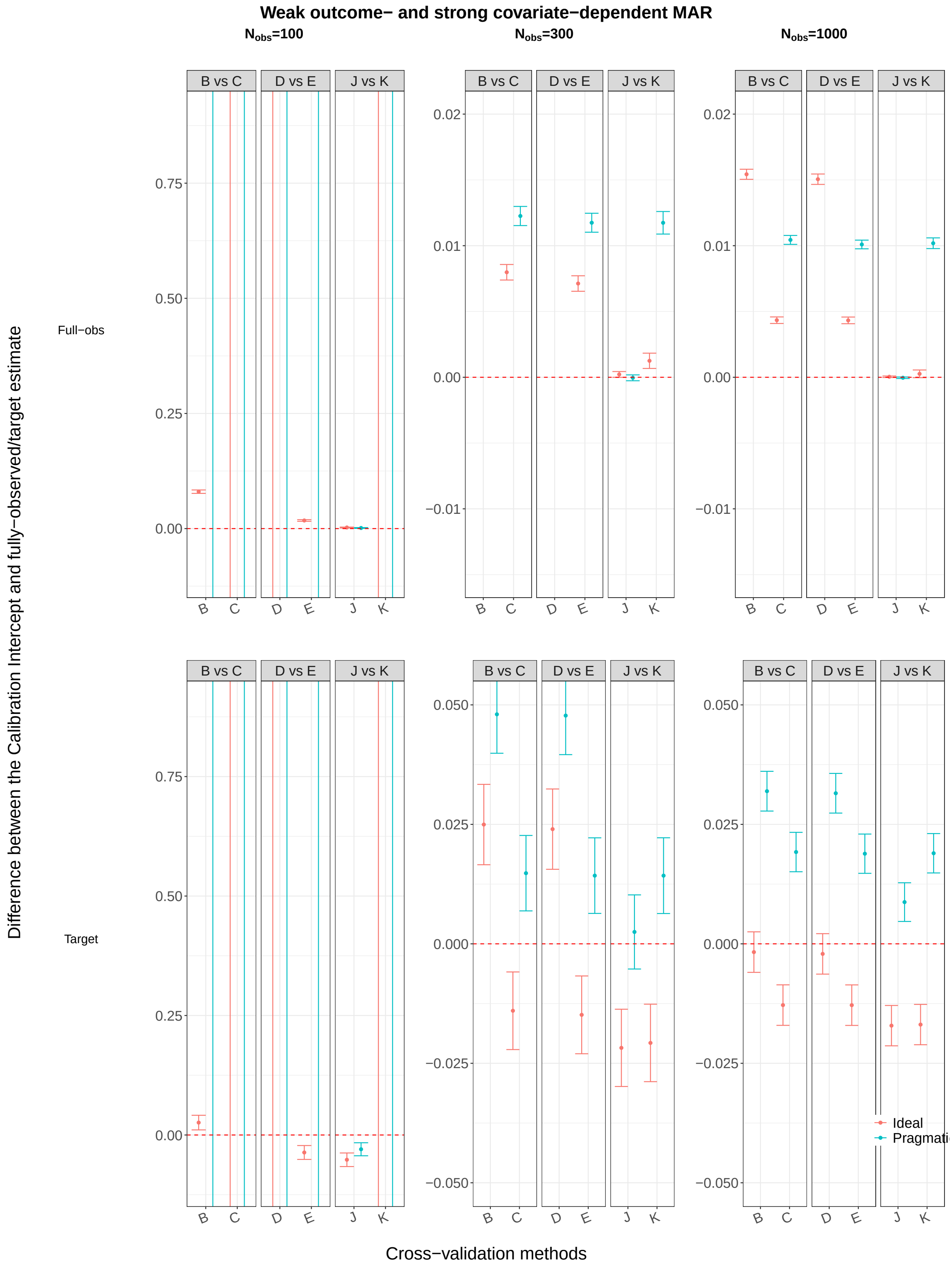

Figure S67: Assessing data leakage within the imputation process for cross-validation. The differences Intercept$_{imp}$ - Intercept$_{obs}$ and Intercept$_{imp}$ - Intercept$_{target}$ are compared when data are weak outcome- and strong covariate-dependent MAR. Methods are compared to both the calibration intercept estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S2.5.4 Calibration slope

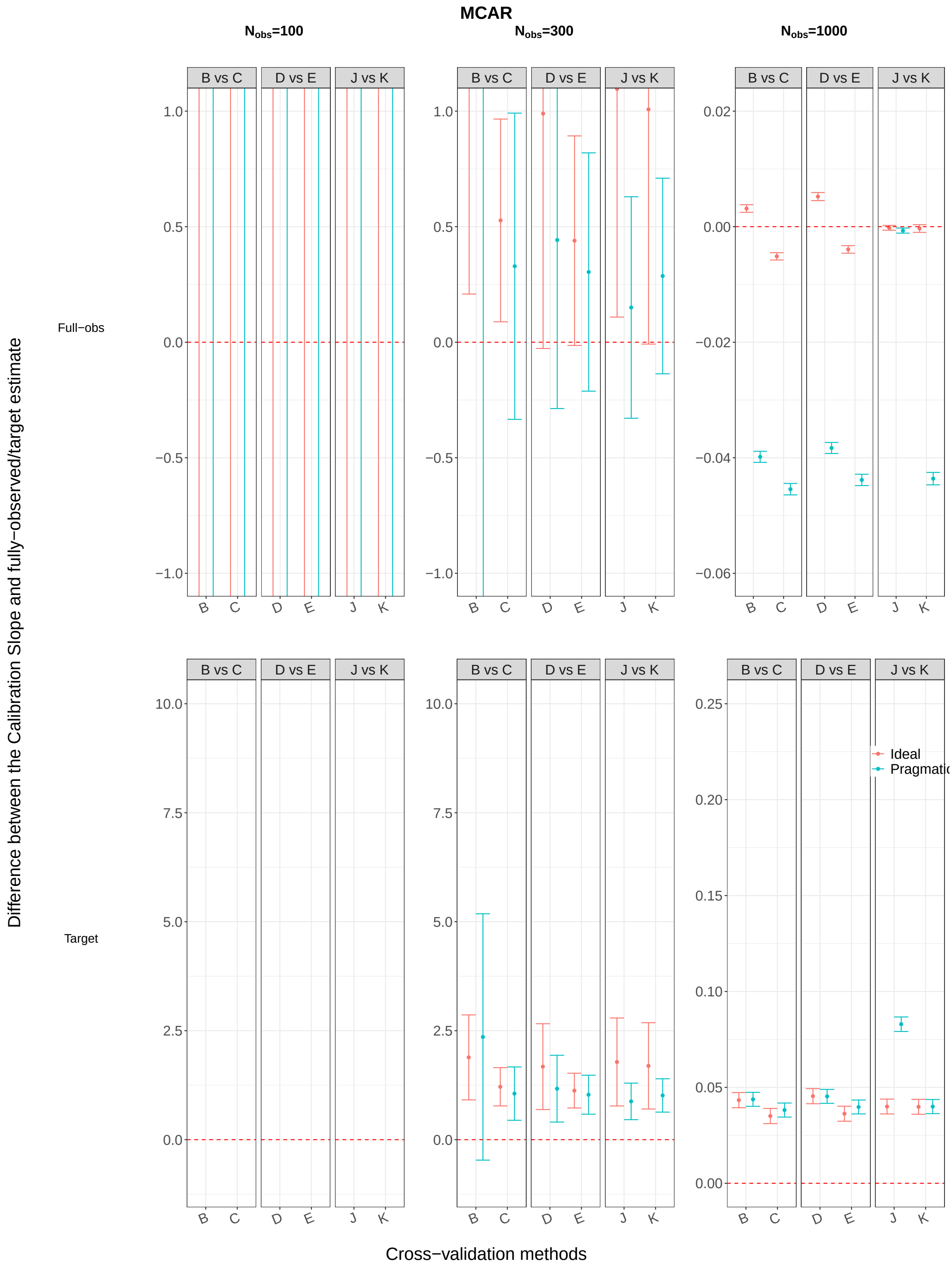

Figure S68: Assessing data leakage within the imputation process for cross-validation. The differences Slope$_{imp}$ - Slope$_{obs}$ and Slope$_{imp}$ - Slope$_{target}$ are compared when data are MCAR. Methods are compared to both the calibration slope estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

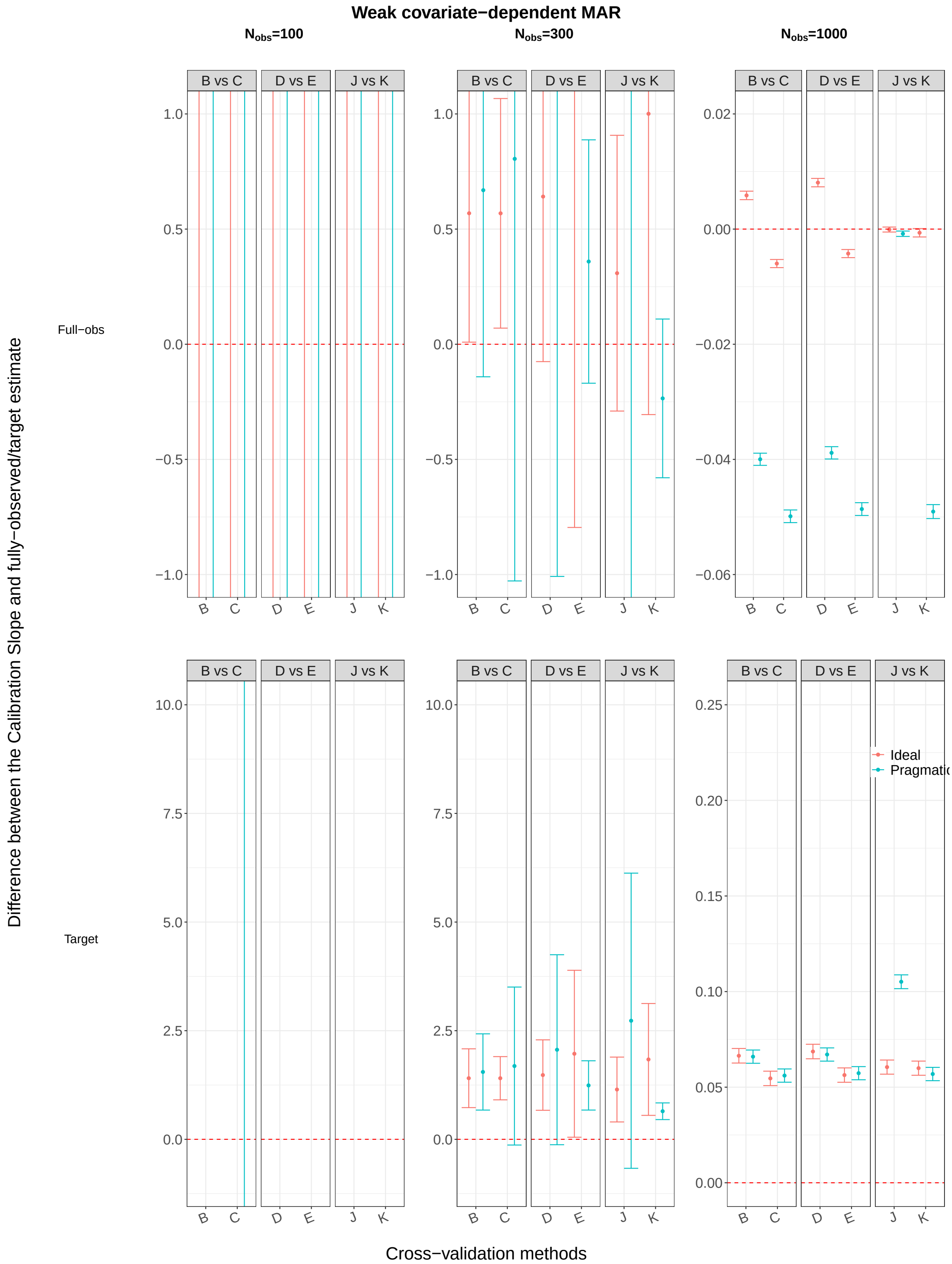

Figure S69: Assessing data leakage within the imputation process for cross-validation. The differences Slope$_{imp}$ - Slope$_{obs}$ and Slope$_{imp}$ - Slope$_{target}$ are compared when data are weak covariate-dependent MAR. Methods are compared to both the calibration slope estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

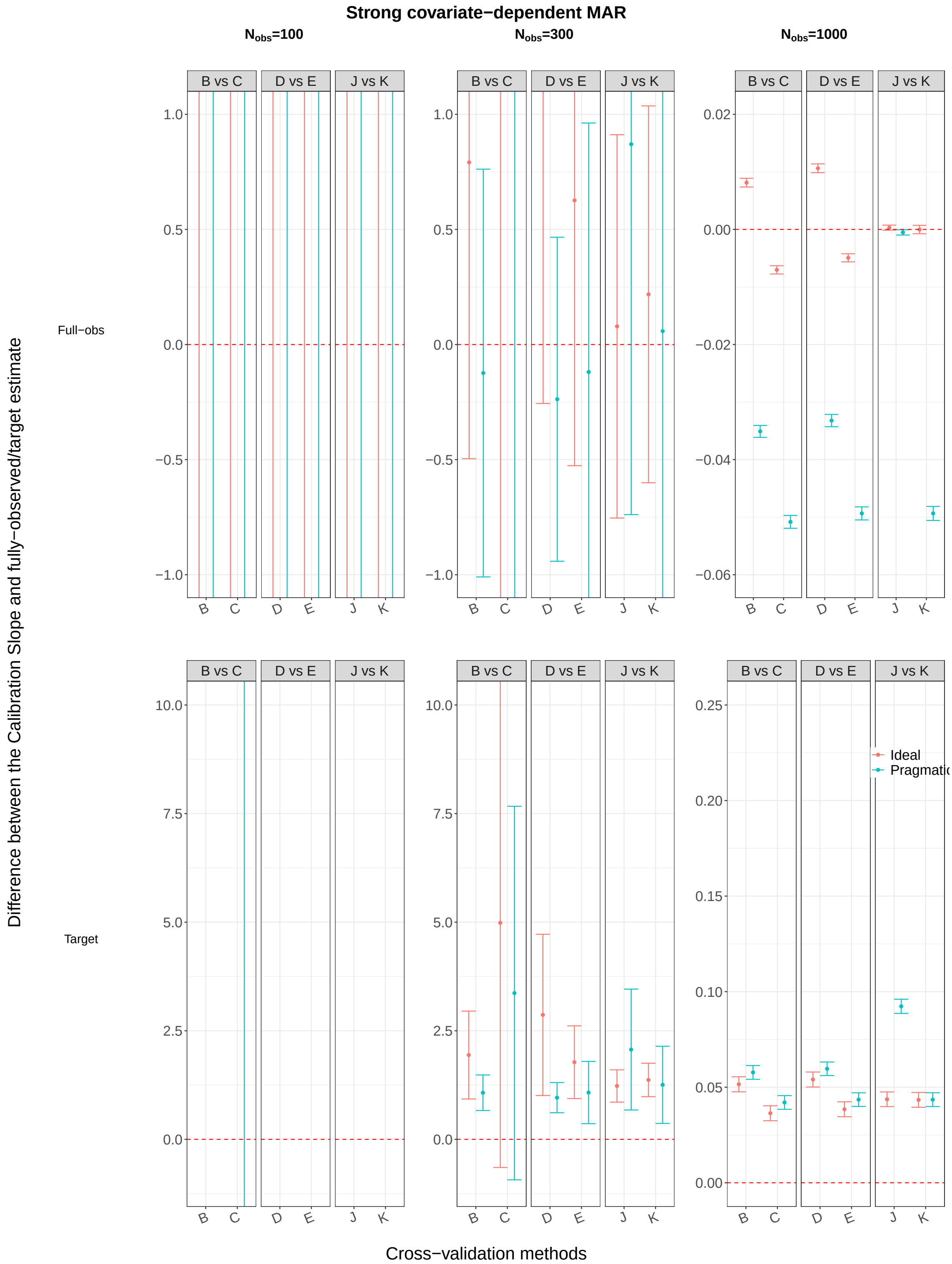

Figure S70: Assessing data leakage within the imputation process for cross-validation. The differences Slope$_{imp}$ - Slope$_{obs}$ and Slope$_{imp}$ - Slope$_{target}$ are compared when data are strong covariate-dependent MAR. Methods are compared to both the calibration slope estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

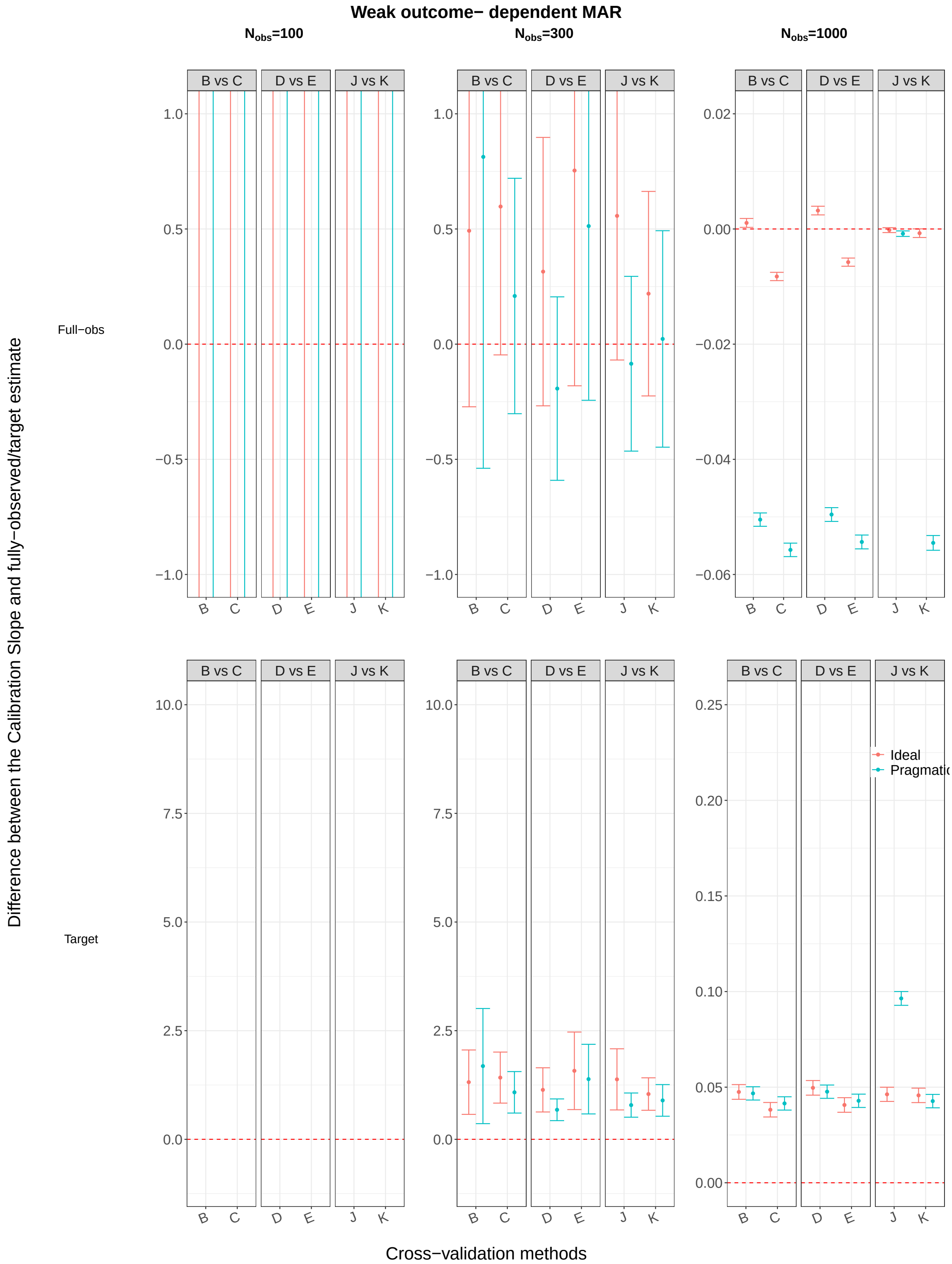

Figure S71: Assessing data leakage within the imputation process for cross-validation. The differences Slope$_{imp}$ - Slope$_{obs}$ and Slope$_{imp}$ - Slope$_{target}$ are compared when data are weak outcome-dependent MAR. Methods are compared to both the calibration slope estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

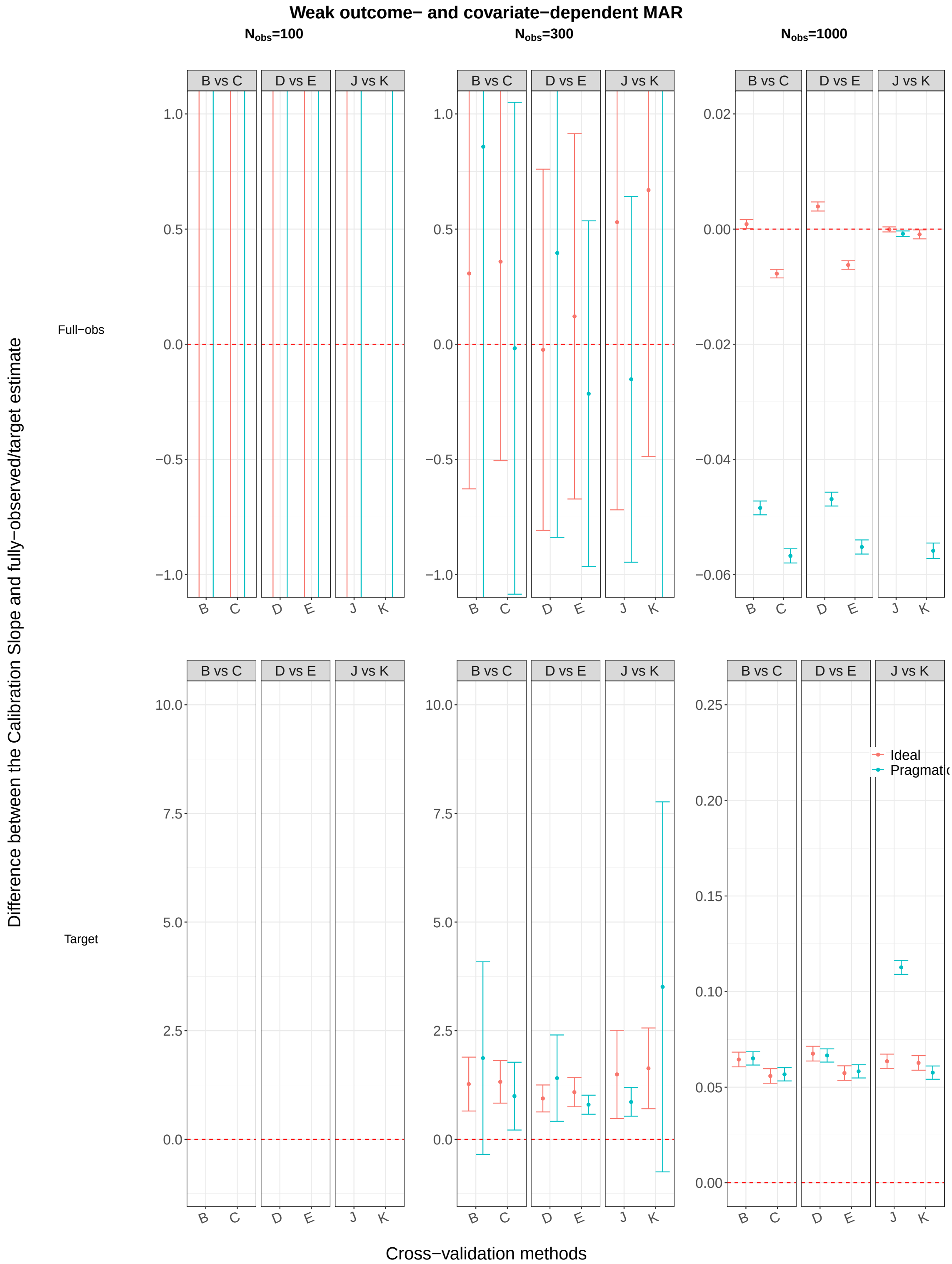

Figure S72: Assessing data leakage within the imputation process for cross-validation. The differences Slope$_{imp}$ - Slope$_{obs}$ and Slope$_{imp}$ - Slope$_{target}$ are compared when data are weak outcome- and covariate-dependent MAR. Methods are compared to both the calibration slope estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

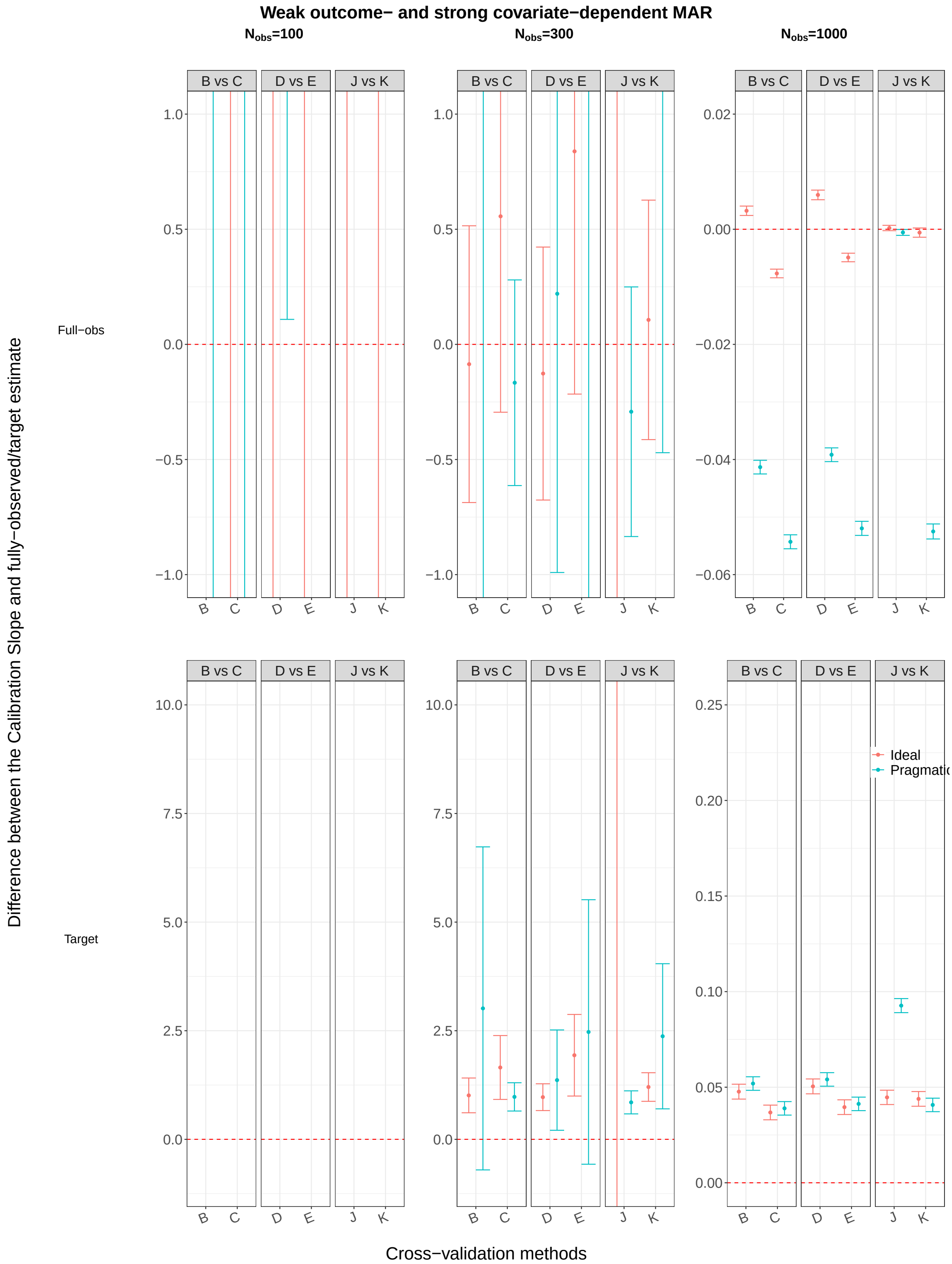

Figure S73: Assessing data leakage within the imputation process for cross-validation. The differences Slope$_{imp}$ - Slope$_{obs}$ and Slope$_{imp}$ - Slope$_{target}$ are compared when data are weak outcome- and strong covariate-dependent MAR. Methods are compared to both the calibration slope estimate when data are fully-observed (Full-obs, row 1) and the target estimate (Target, row 2) from a larger validation set. CC (complete-case); methods A-K are described in Table 2.3 and summarised in Table 4.4.

S3 Chapter 6: Bootstrap and MI (continuous outcome)

S3.1 Reusing versus re-imputing for test performance of the standard algorithm

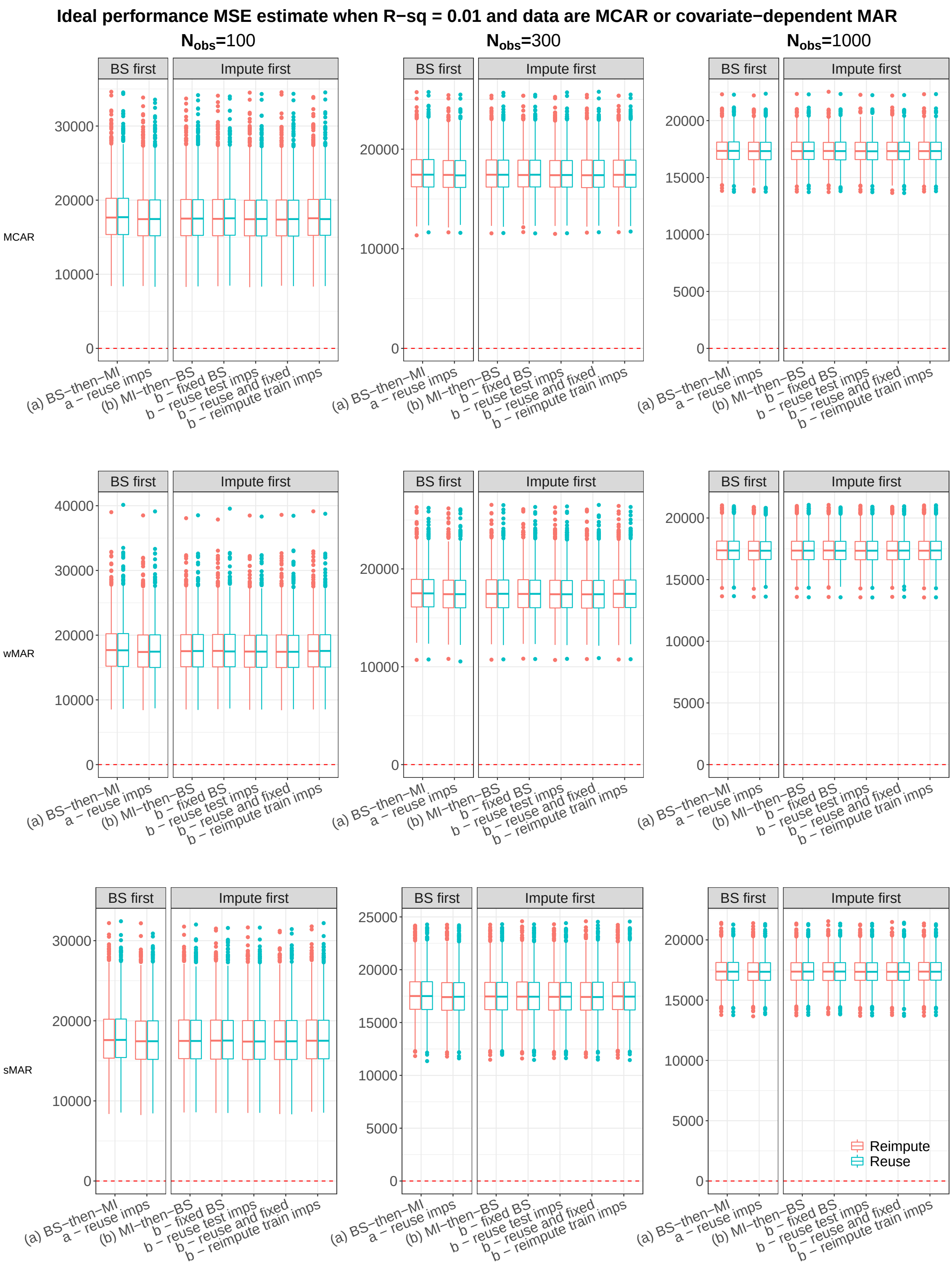

Figure S1: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR and $R^2 = 0.01$. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE$_{imp}$-MSE$_{obs}$). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

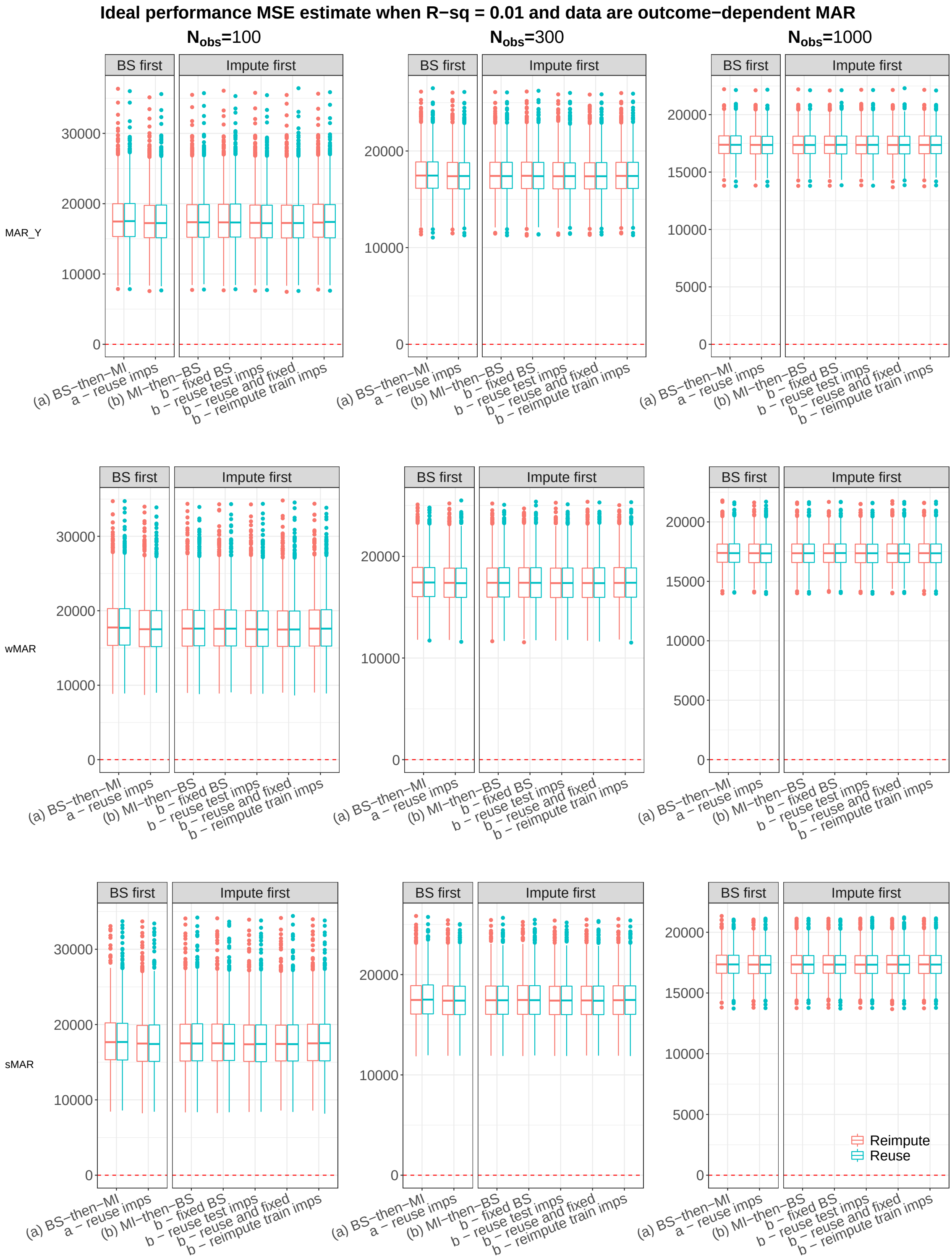

Figure S2: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR and $R^2 = 0.01$. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}- MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

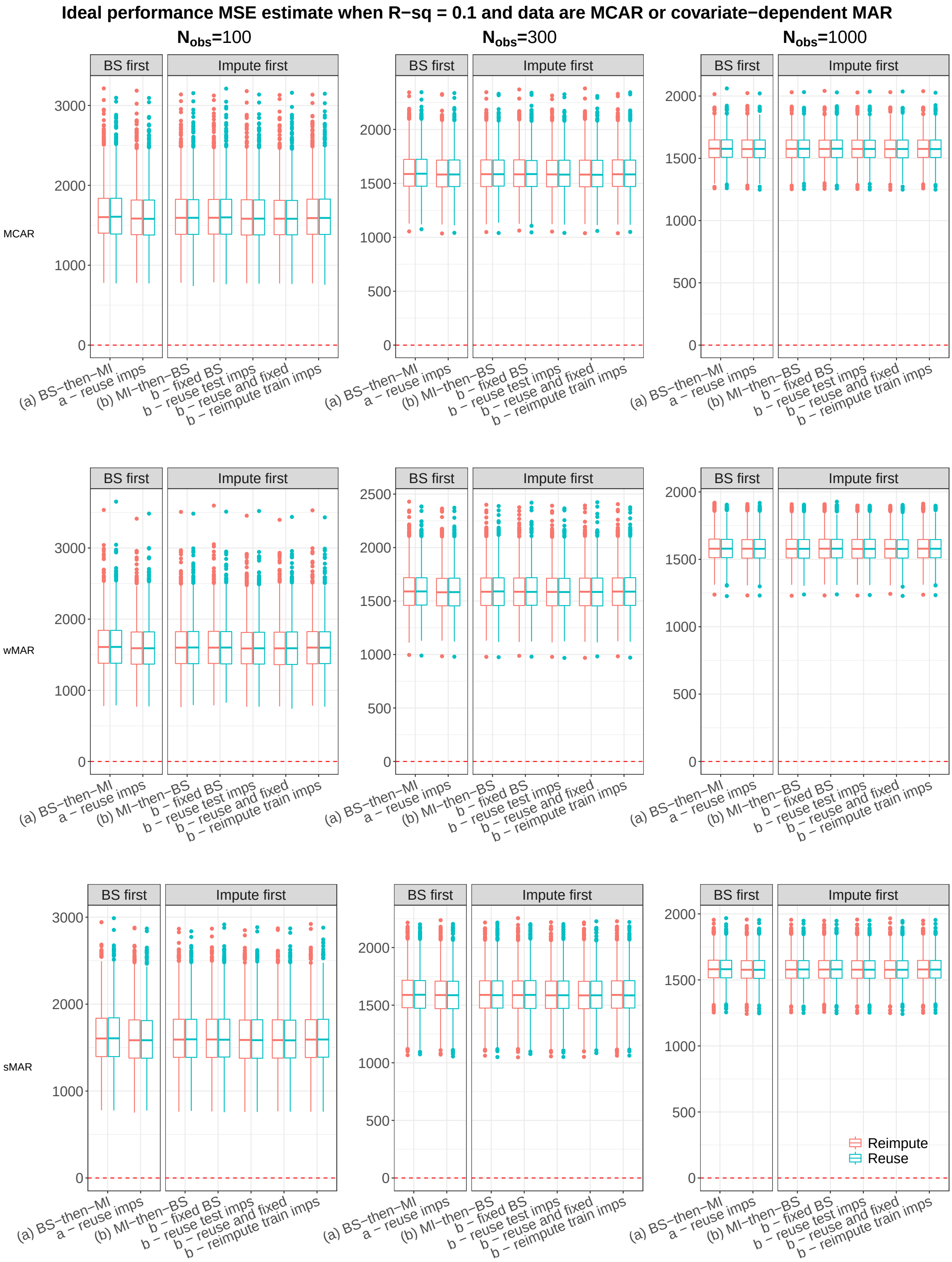

Figure S3: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR and $R^2 = 0.1$. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}-MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

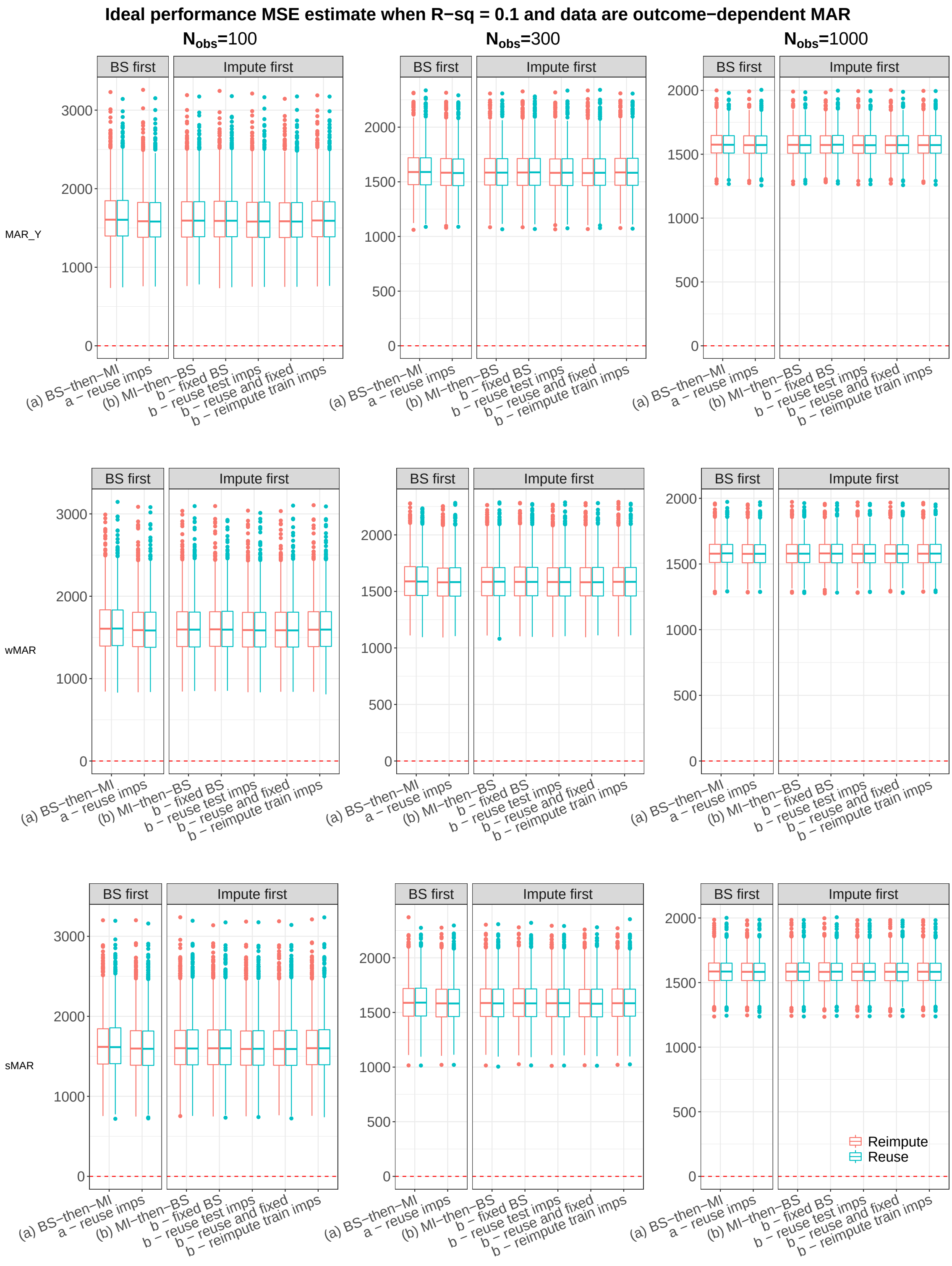

Figure S4: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR and $R^2 = 0.1$. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}- MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

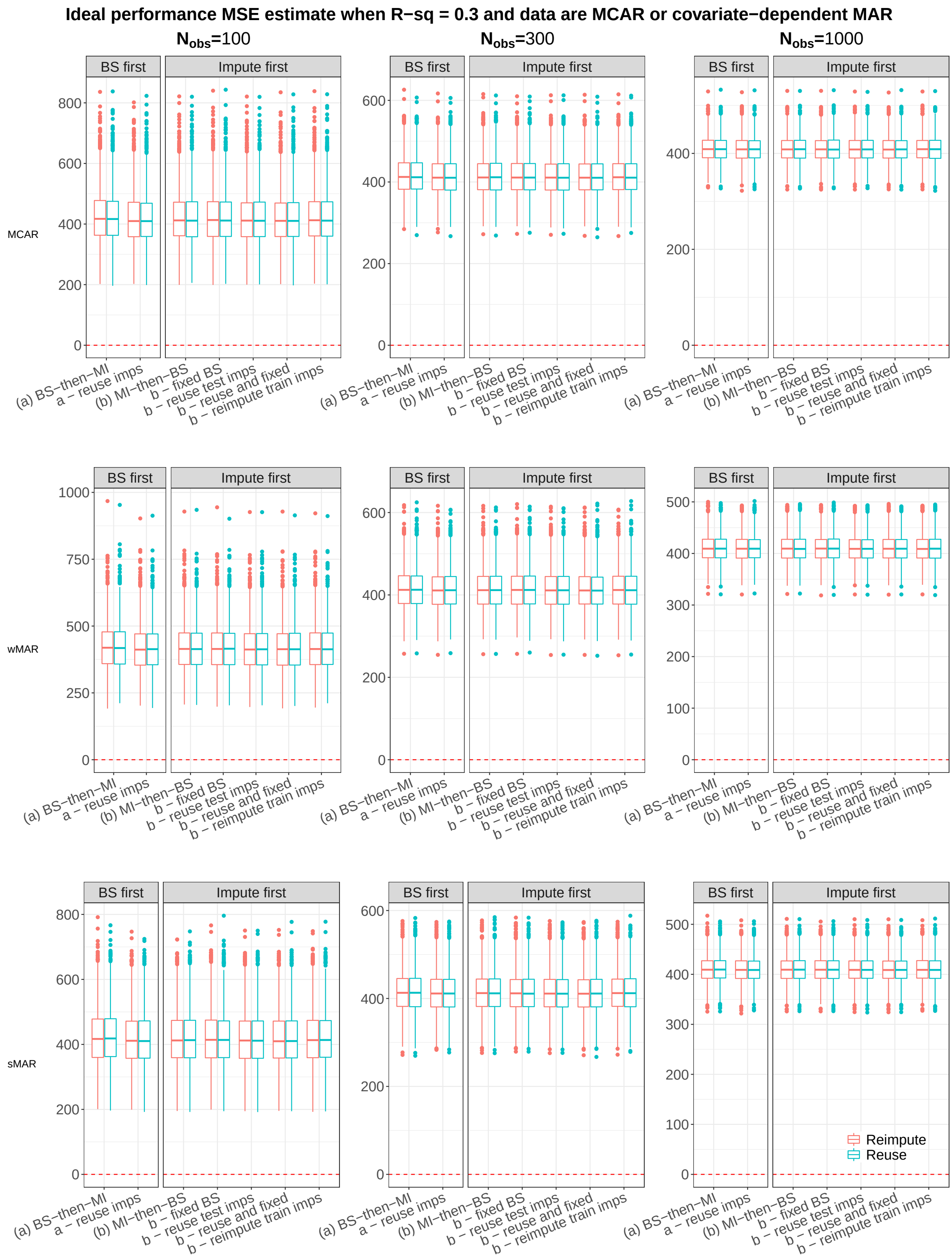

Figure S5: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR and $R^2 = 0.3$. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}-MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

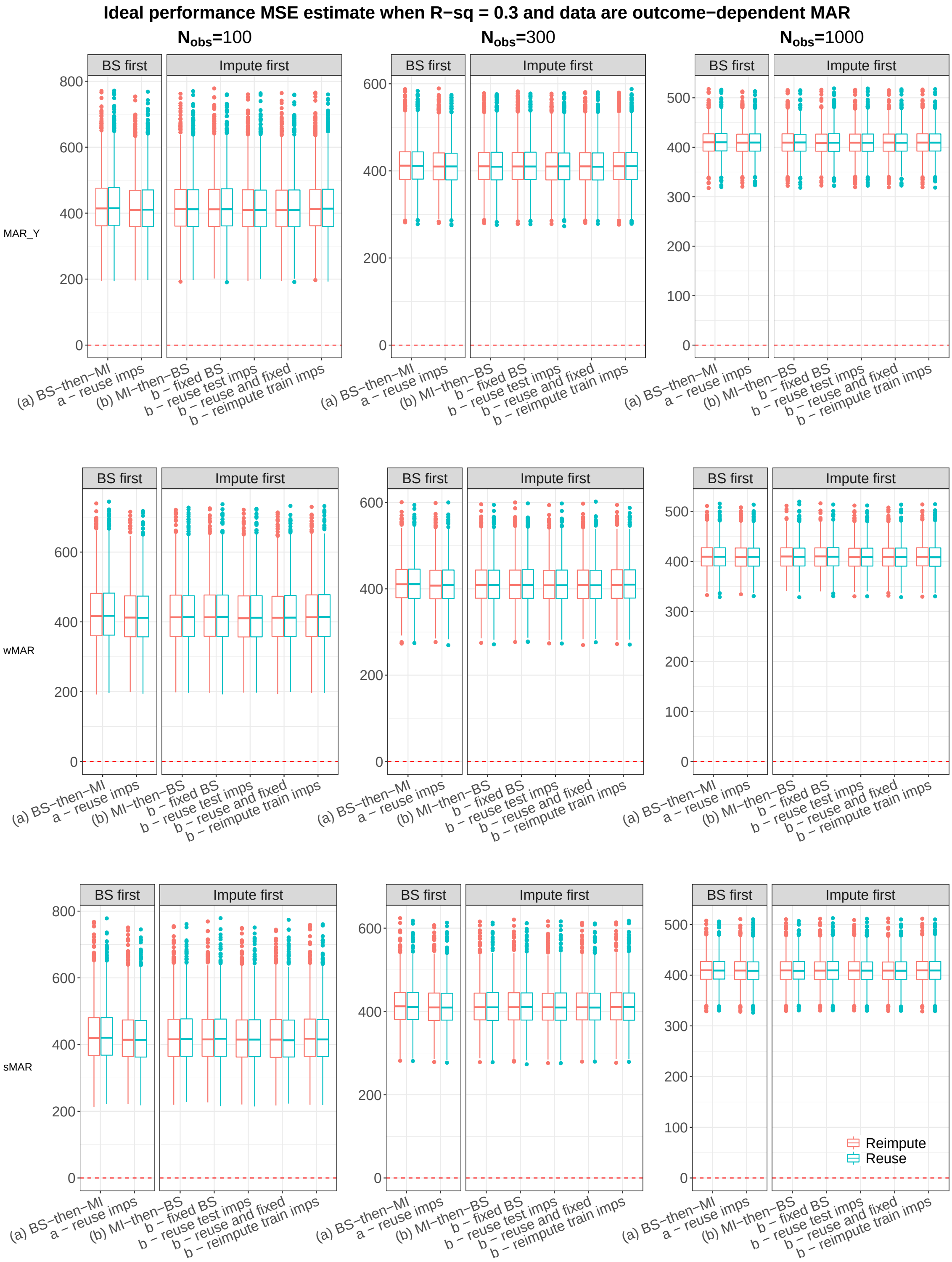

Figure S6: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR and $R^2 = 0.3$. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}- MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

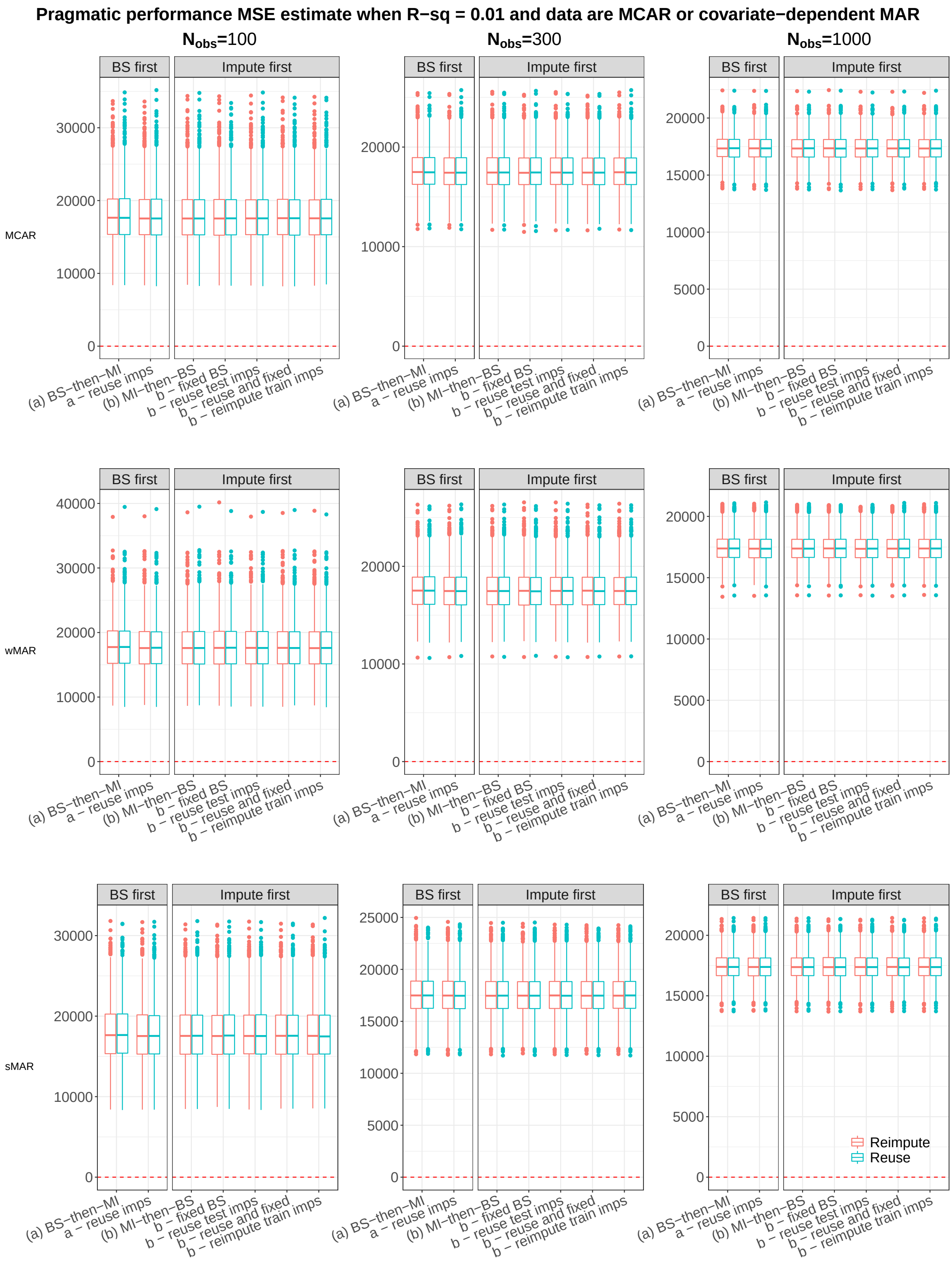

Figure S7: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR and $R^2 = 0.01$. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}-MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

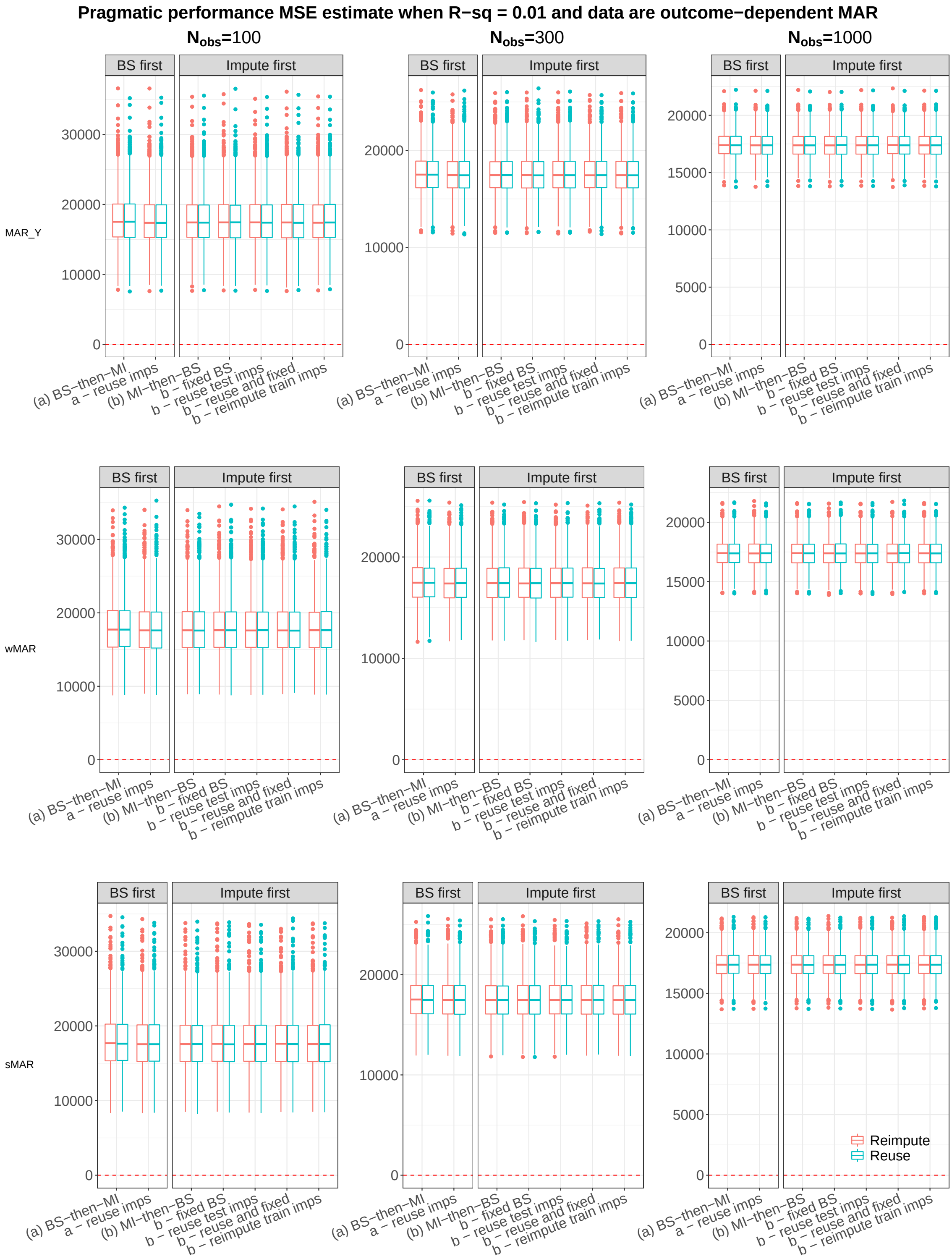

Figure S8: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR and $R^2 = 0.01$. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE$_{imp}$- MSE$_{obs}$). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

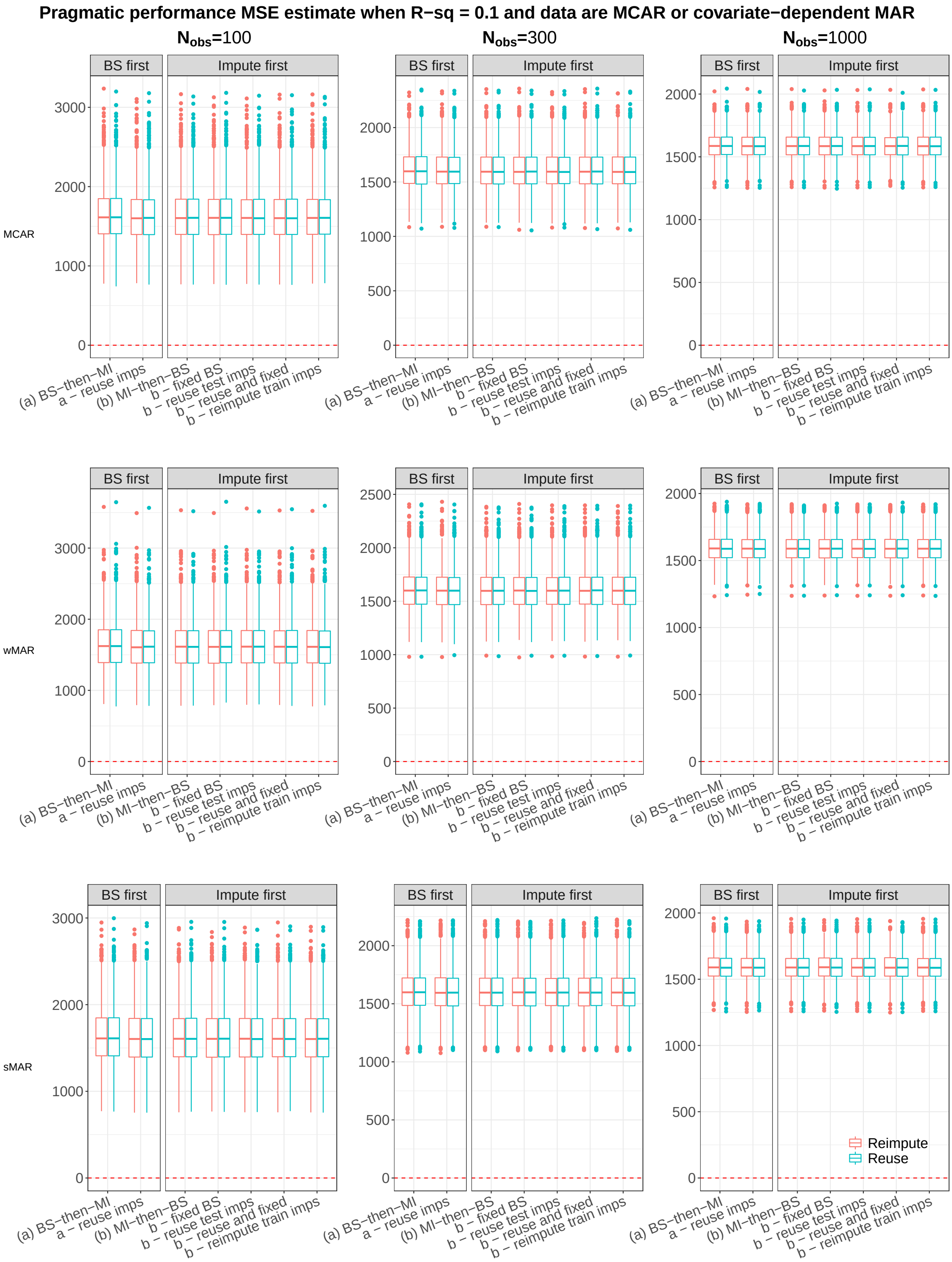

Figure S9: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR and $R^2 = 0.1$. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}-MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

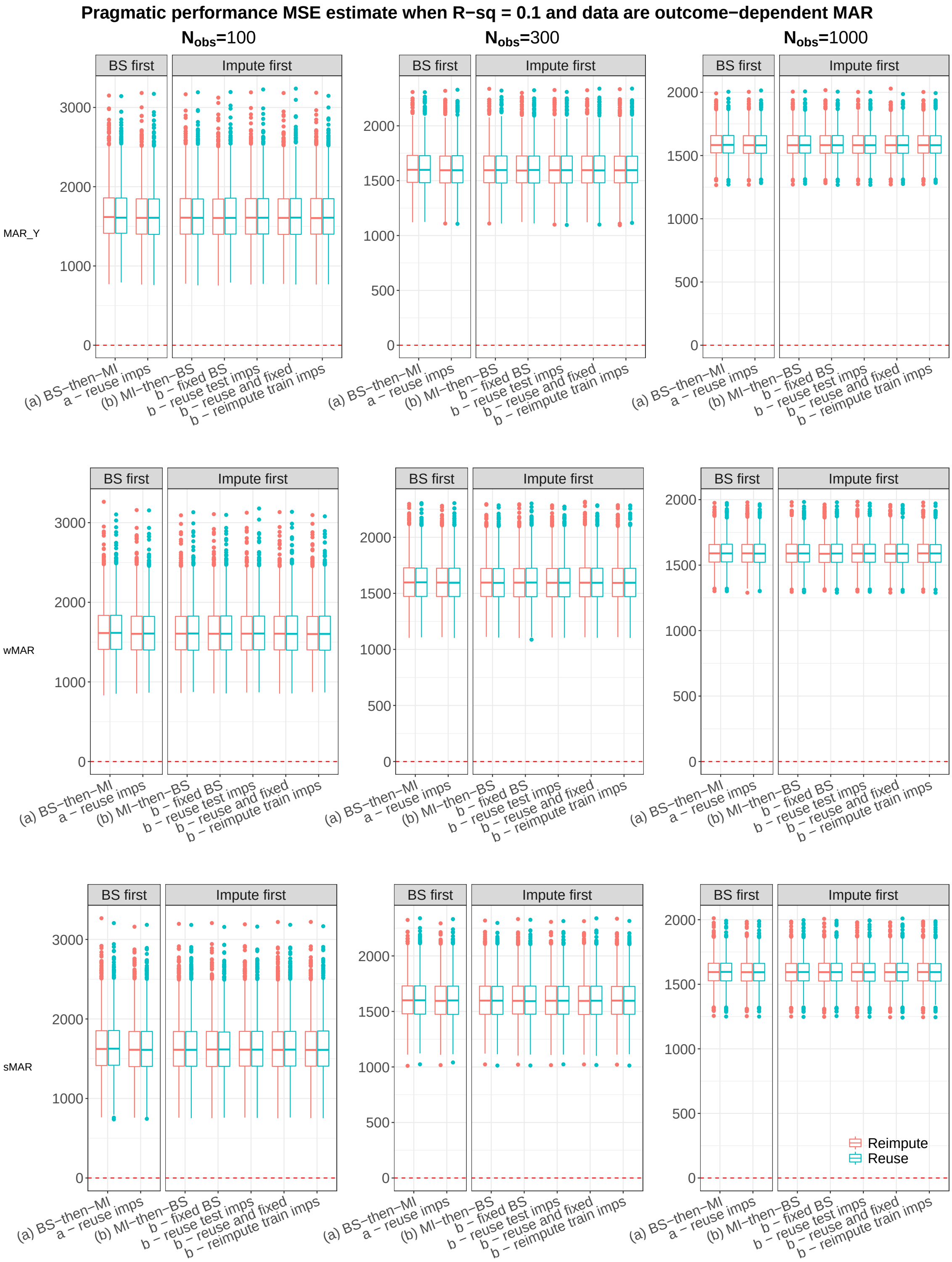

Figure S10: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR and $R^2 = 0.1$. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}- MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

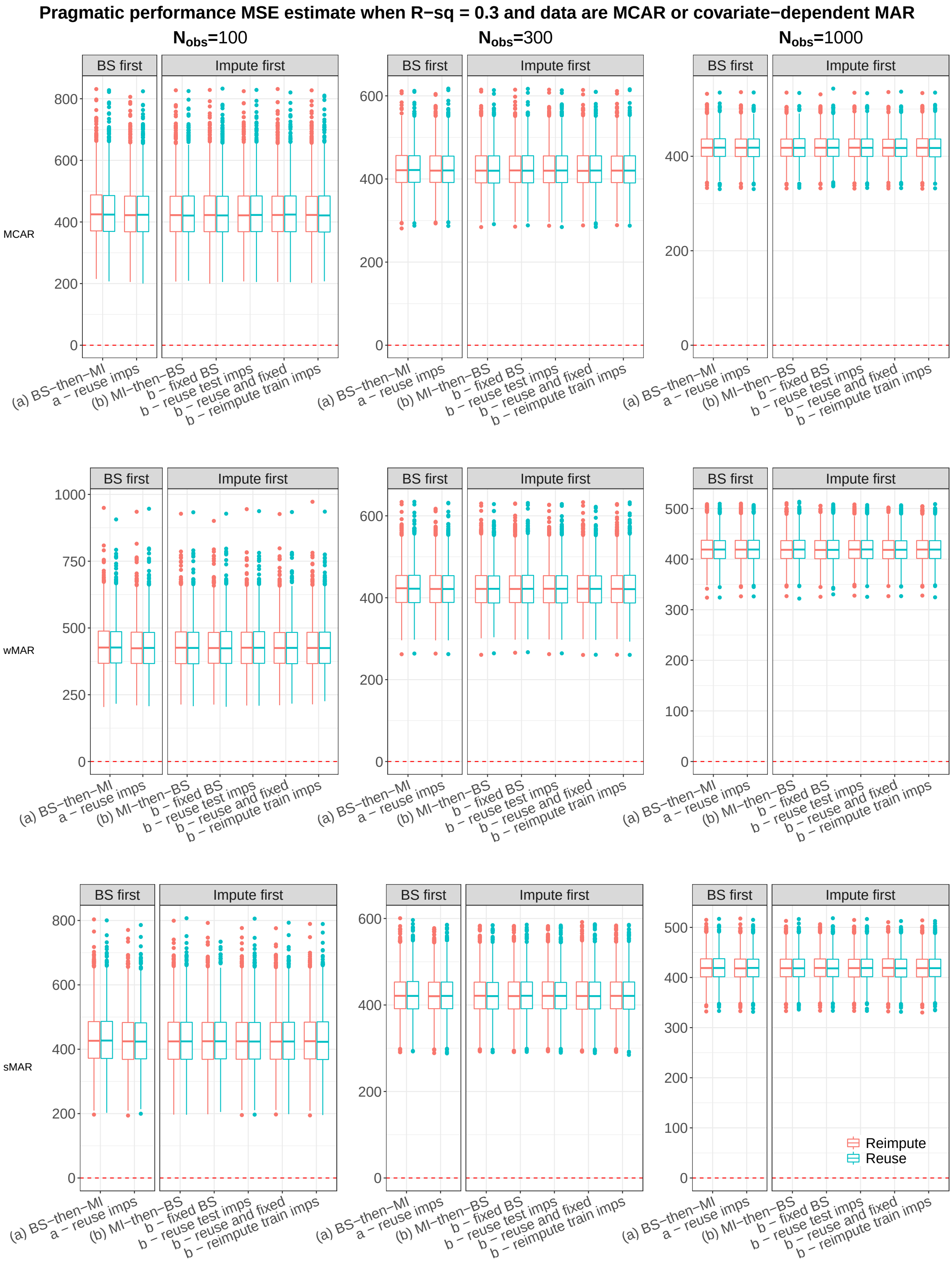

Figure S11: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR and $R^2 = 0.3$. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}- MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

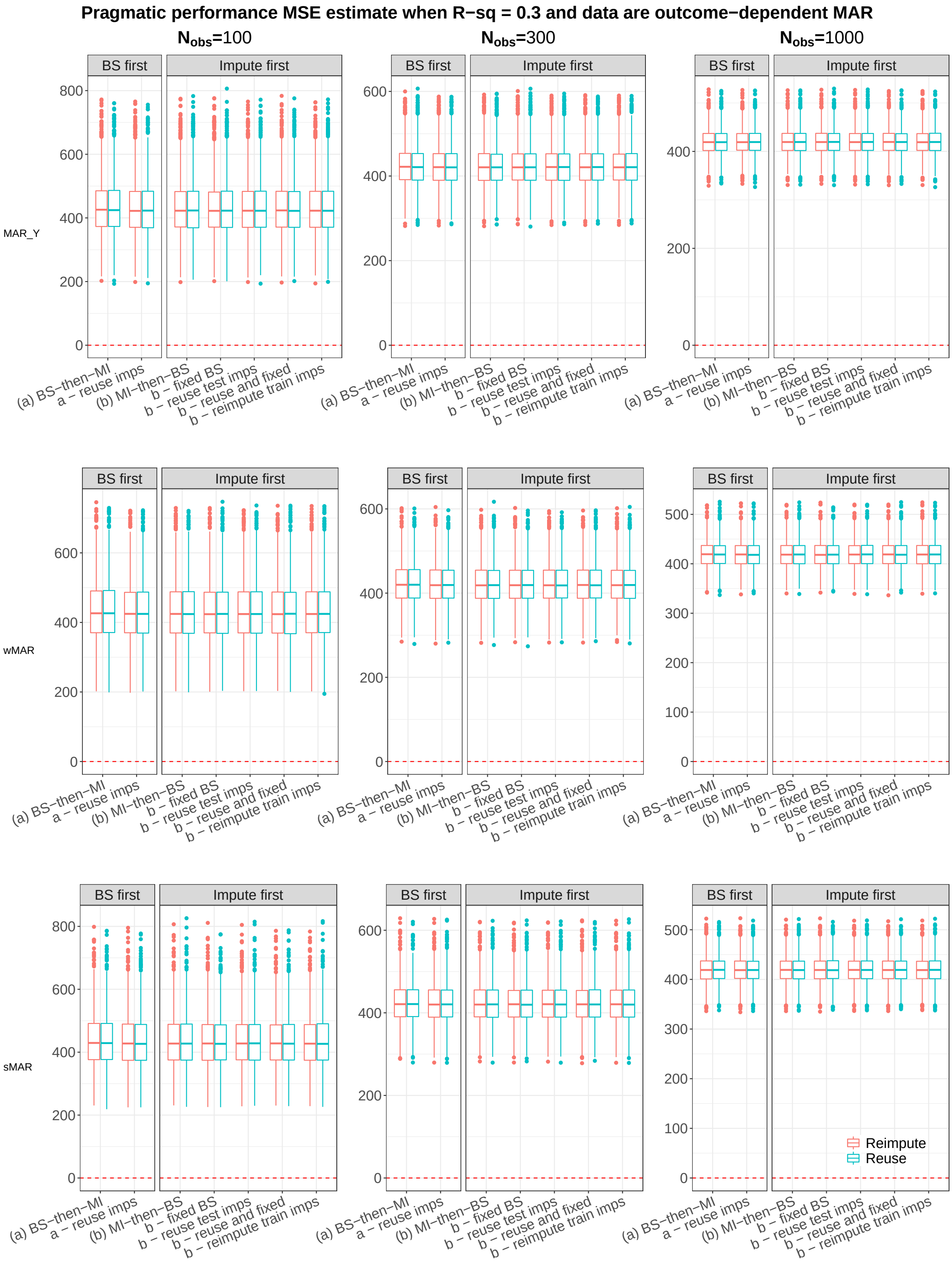

Figure S12: A comparison of reusing versus re-imputing test datasets on the MSE estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR and $R^2 = 0.3$. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the MSE estimated when data are fully-observed (MSE_{imp}- MSE_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.2 The standard bootstrap

S3.2.1 MSE from imputation methods compared to the fully-observed MSE (MSE_{imp}-MSE_{obs})

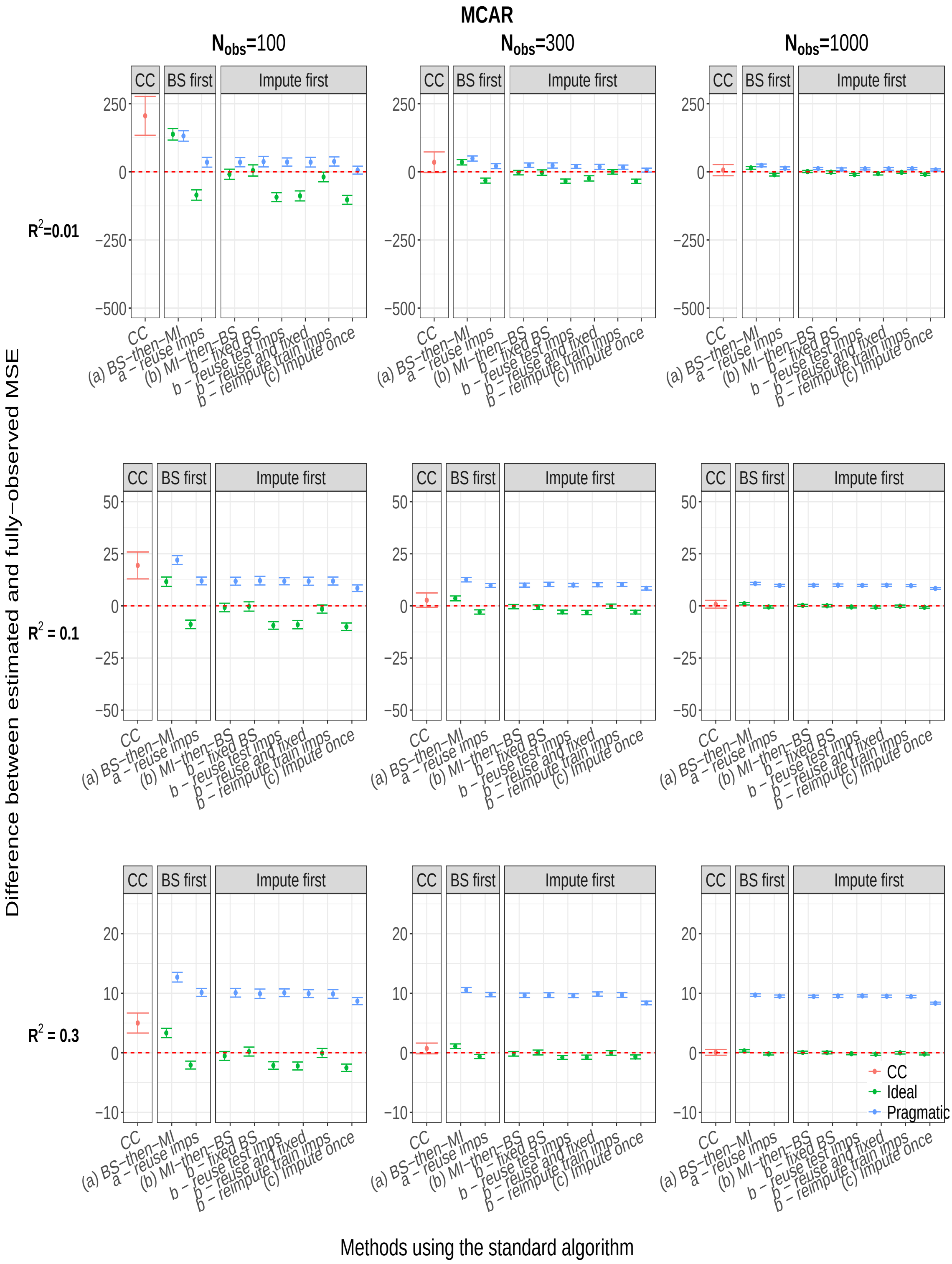

Figure S13: The difference MSE_{imp} - MSE_{obs} when data are MCAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

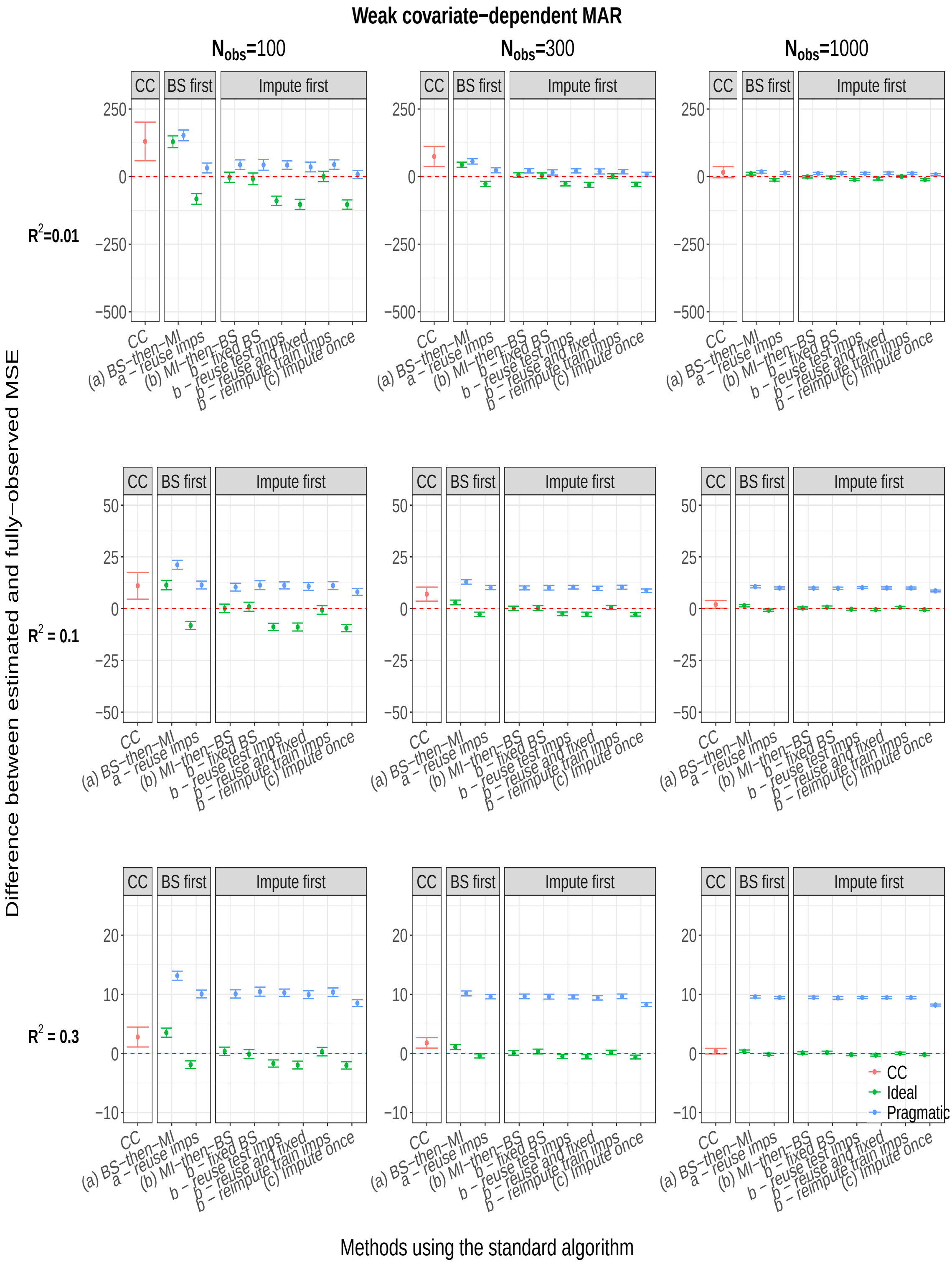

Figure S14: The difference MSE_{imp} - MSE_{obs} when data are weak covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

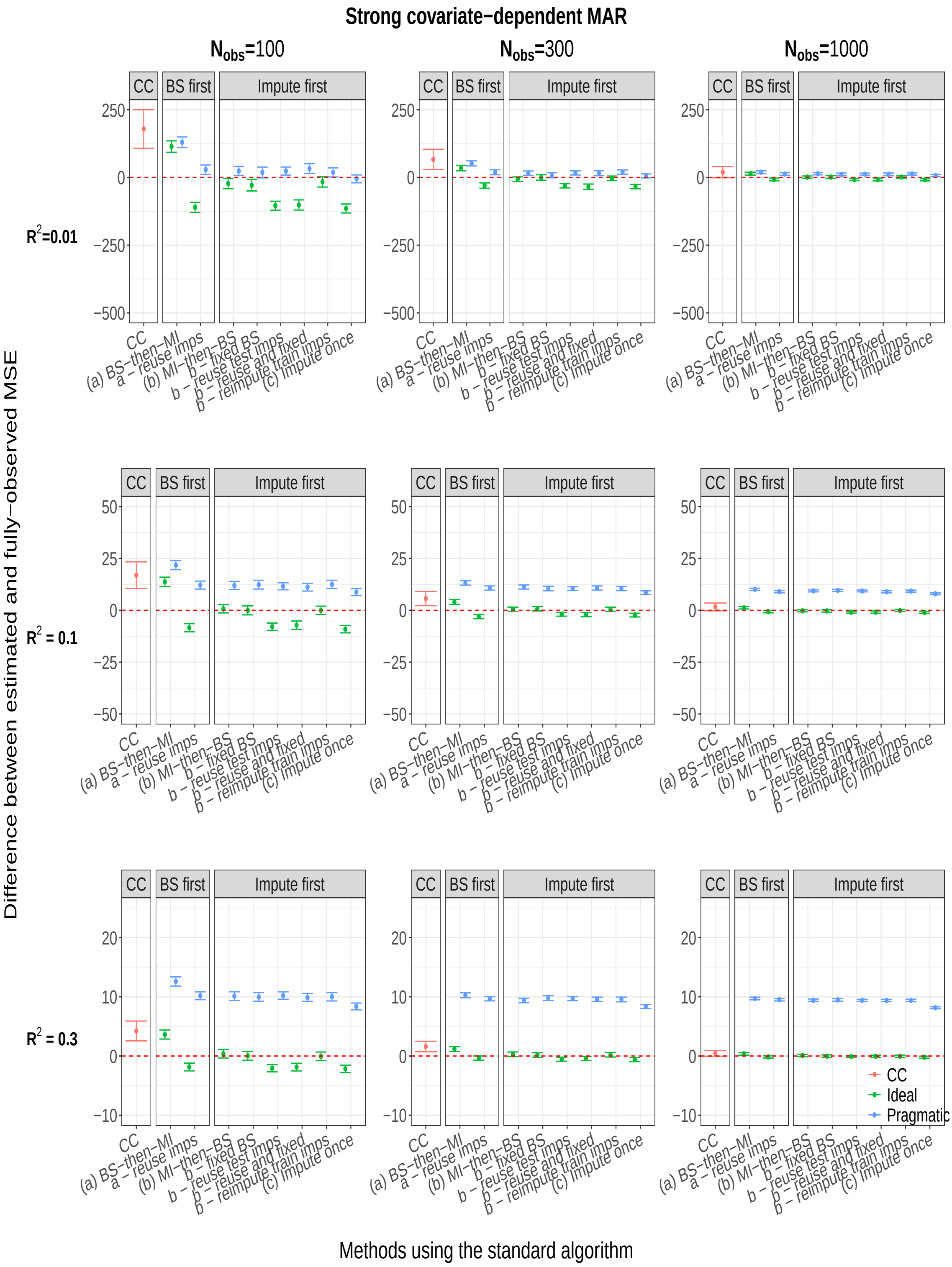

Figure S15: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

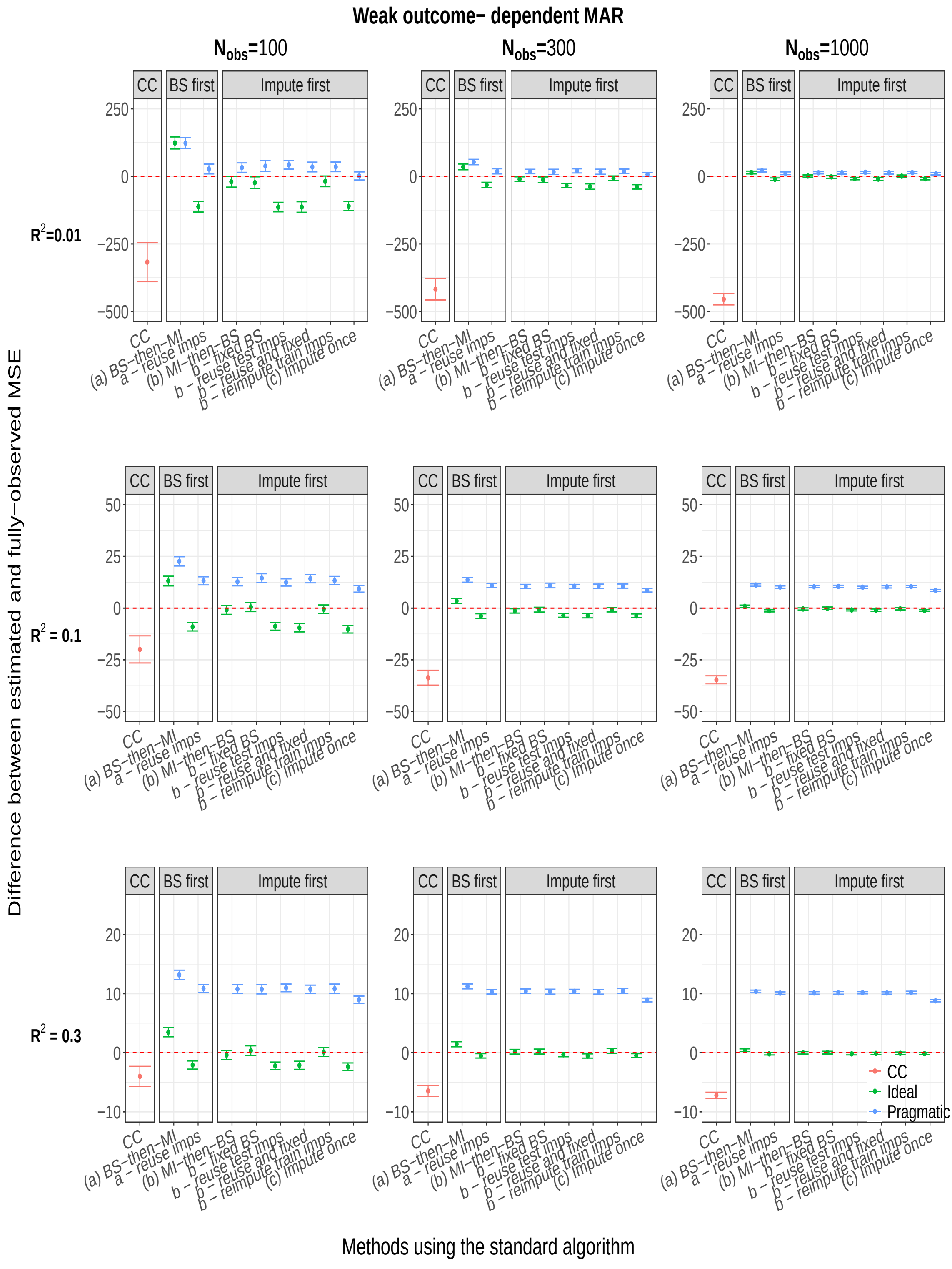

Figure S16: The difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

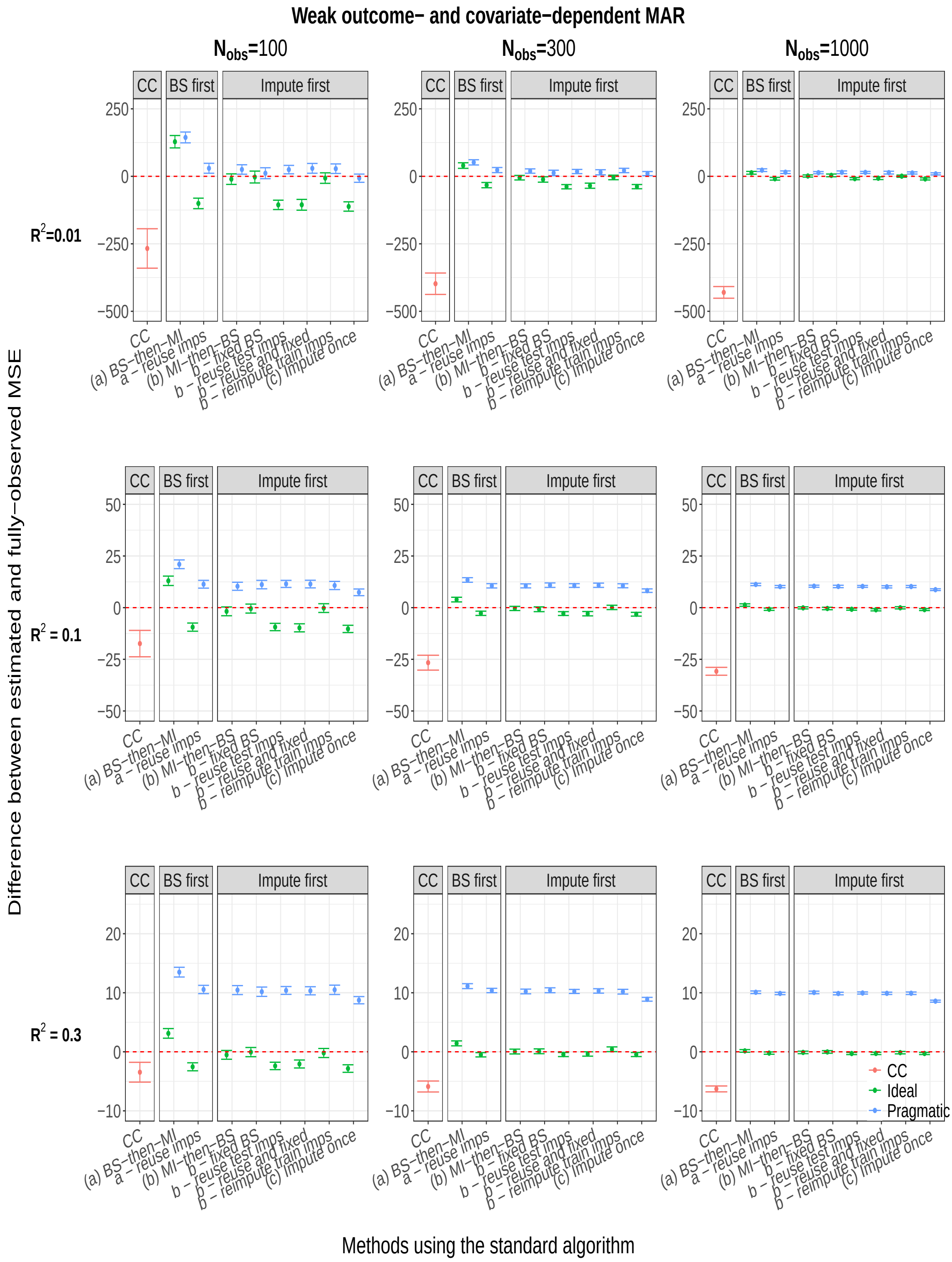

Figure S17: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

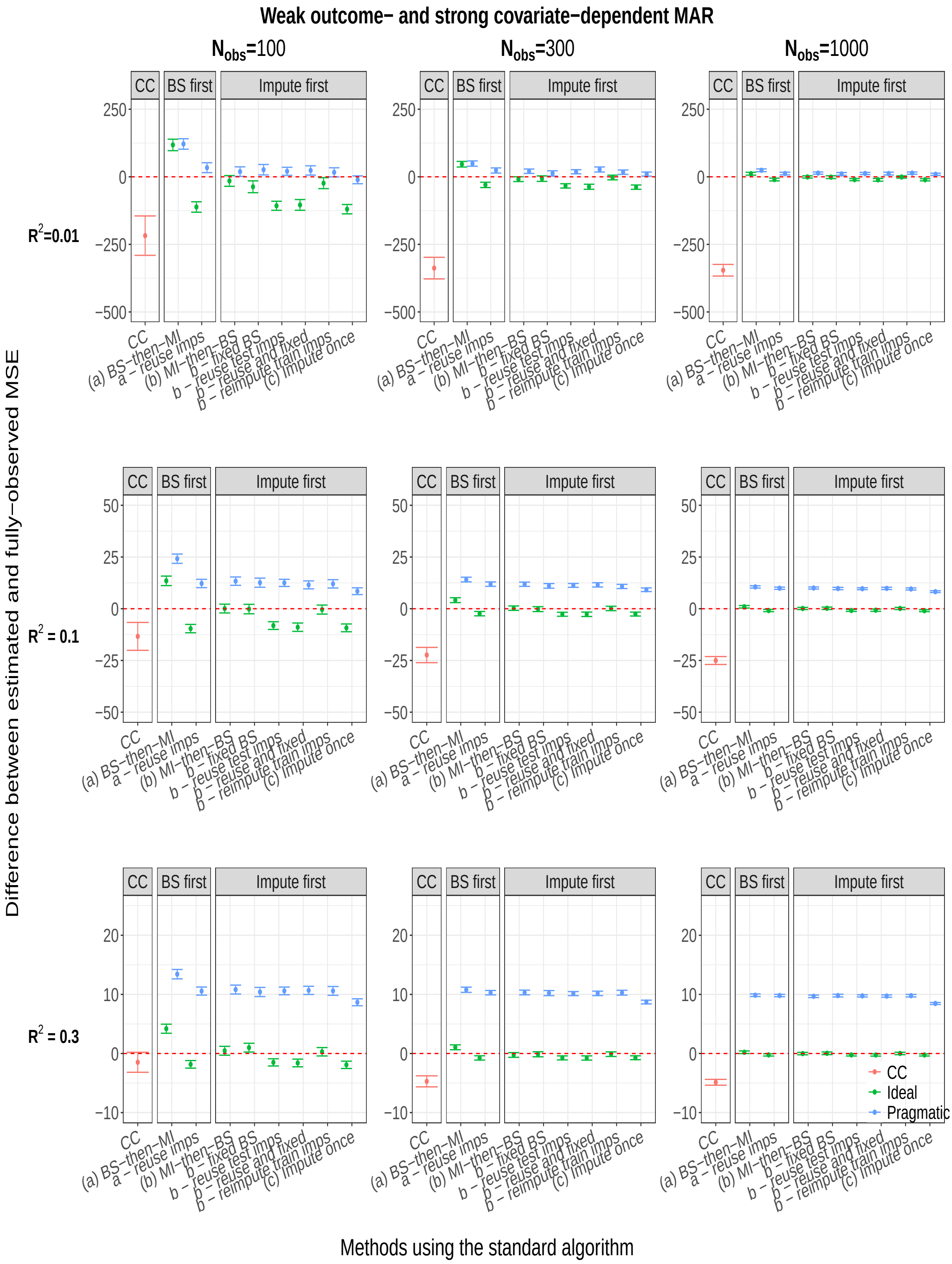

Figure S18: The difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.2.2 The proportion of missingness is 40% (MSE_{imp}-MSE_{obs})

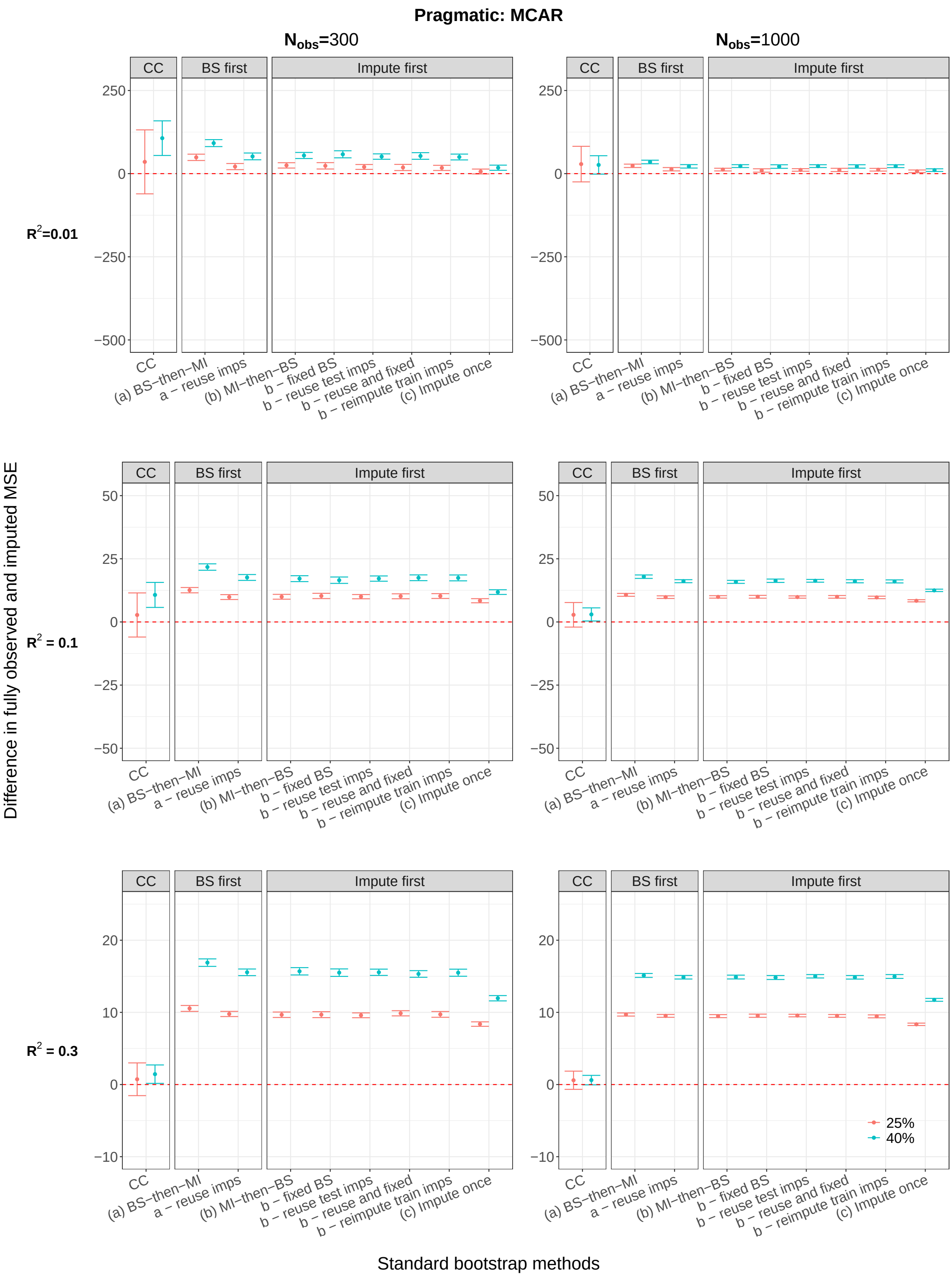

Figure S19: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are MCAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

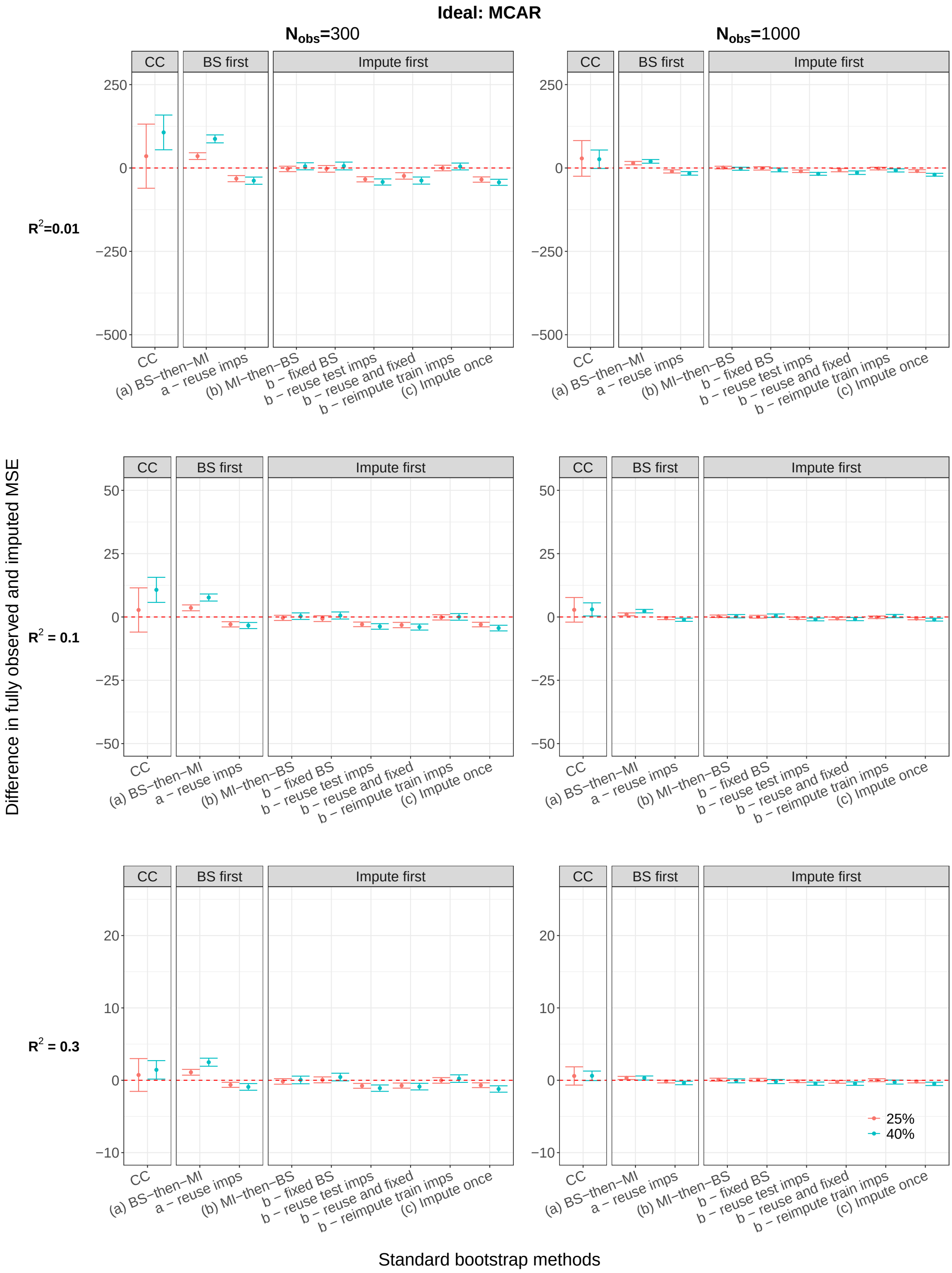

Figure S20: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are MCAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

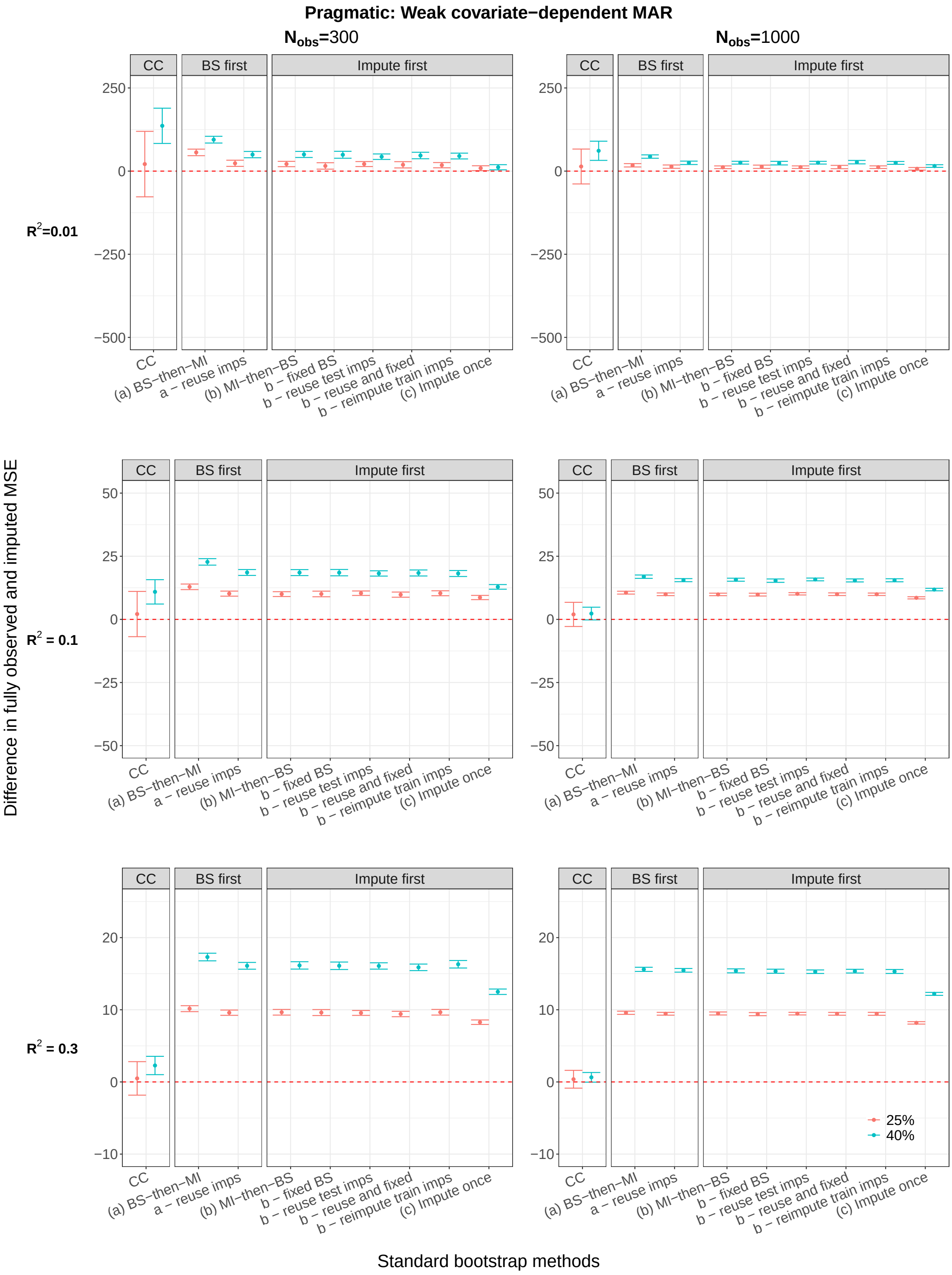

Figure S21: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak covariate-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

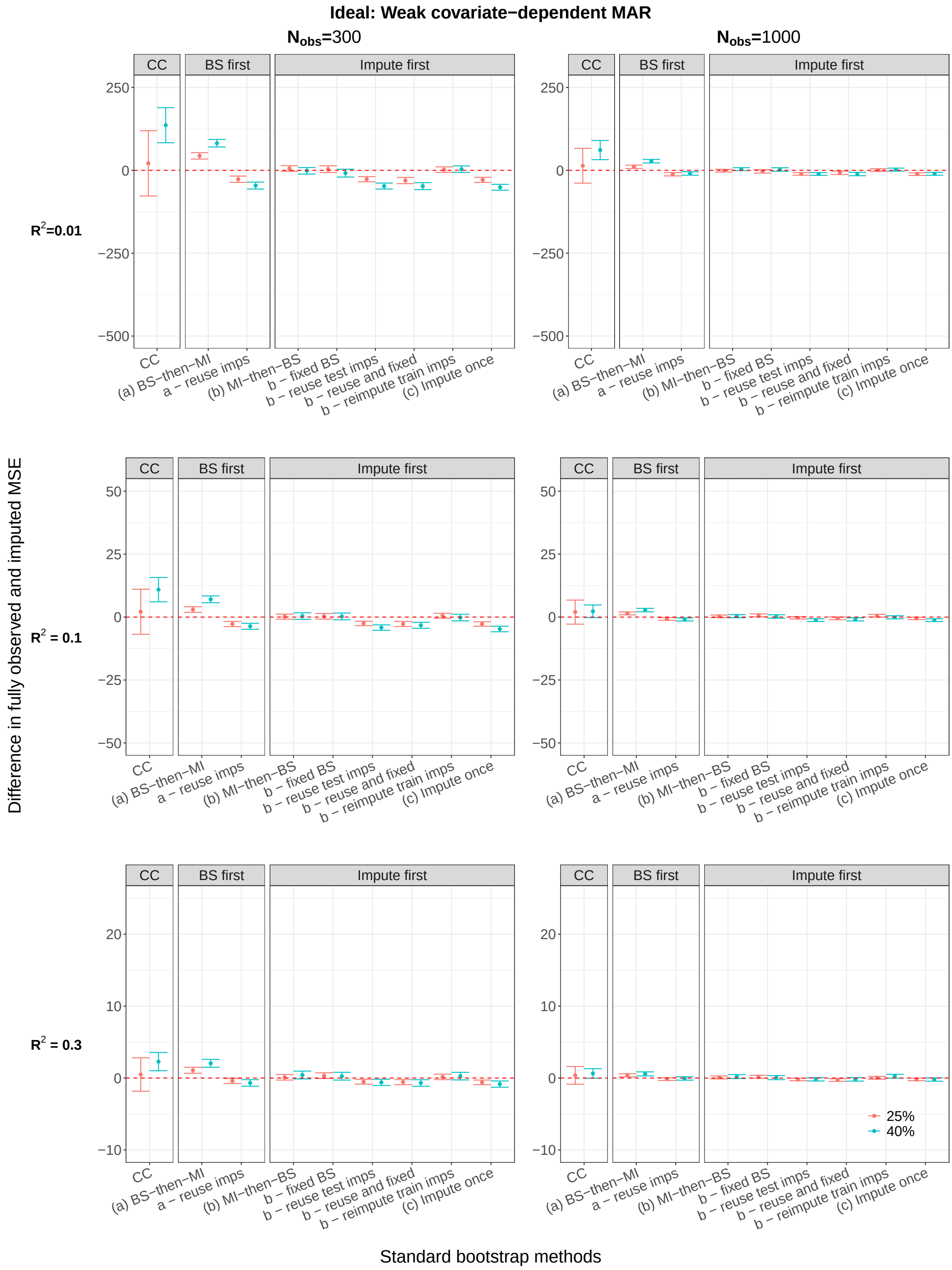

Figure S22: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak covariate-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

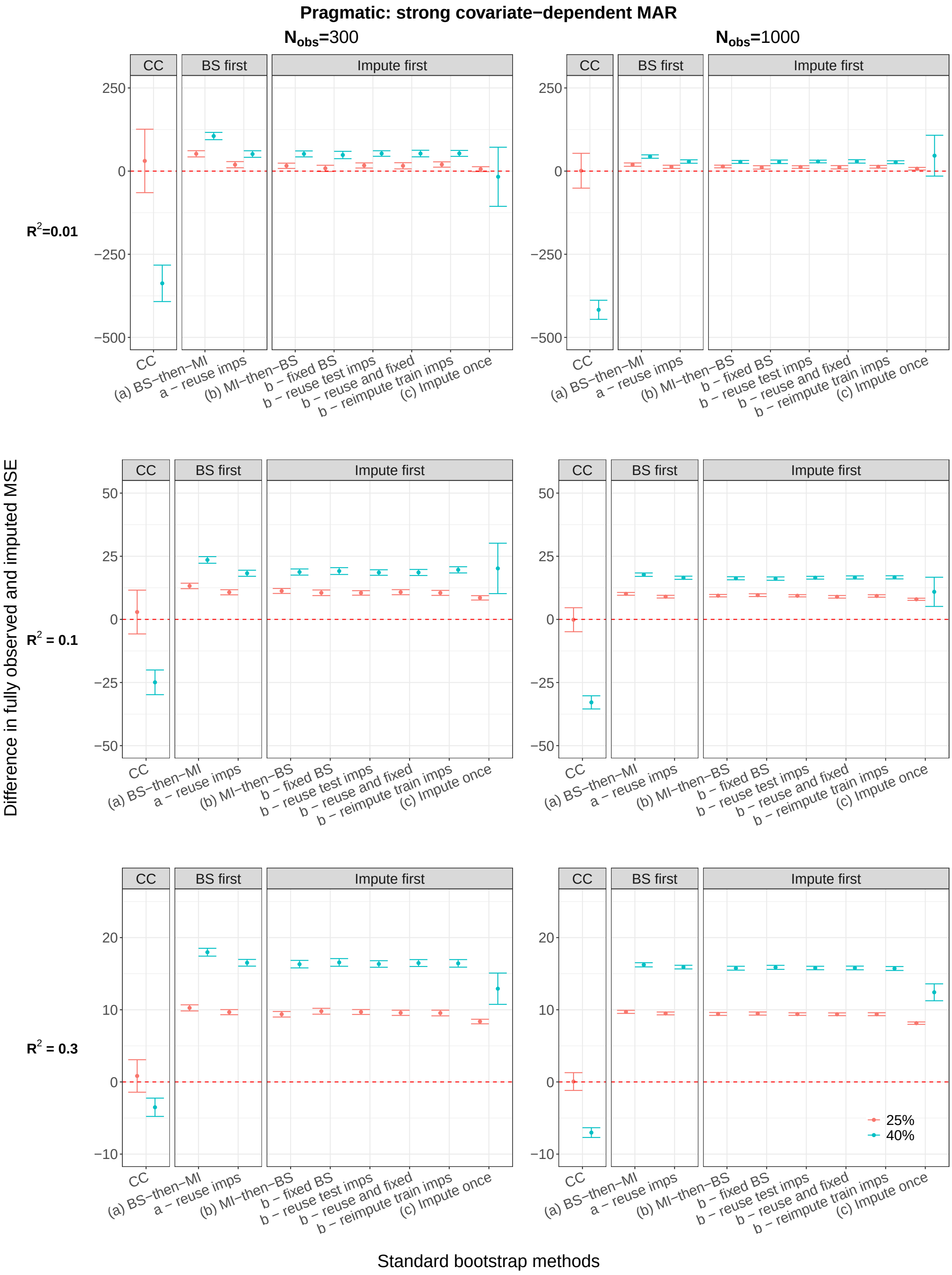

Figure S23: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

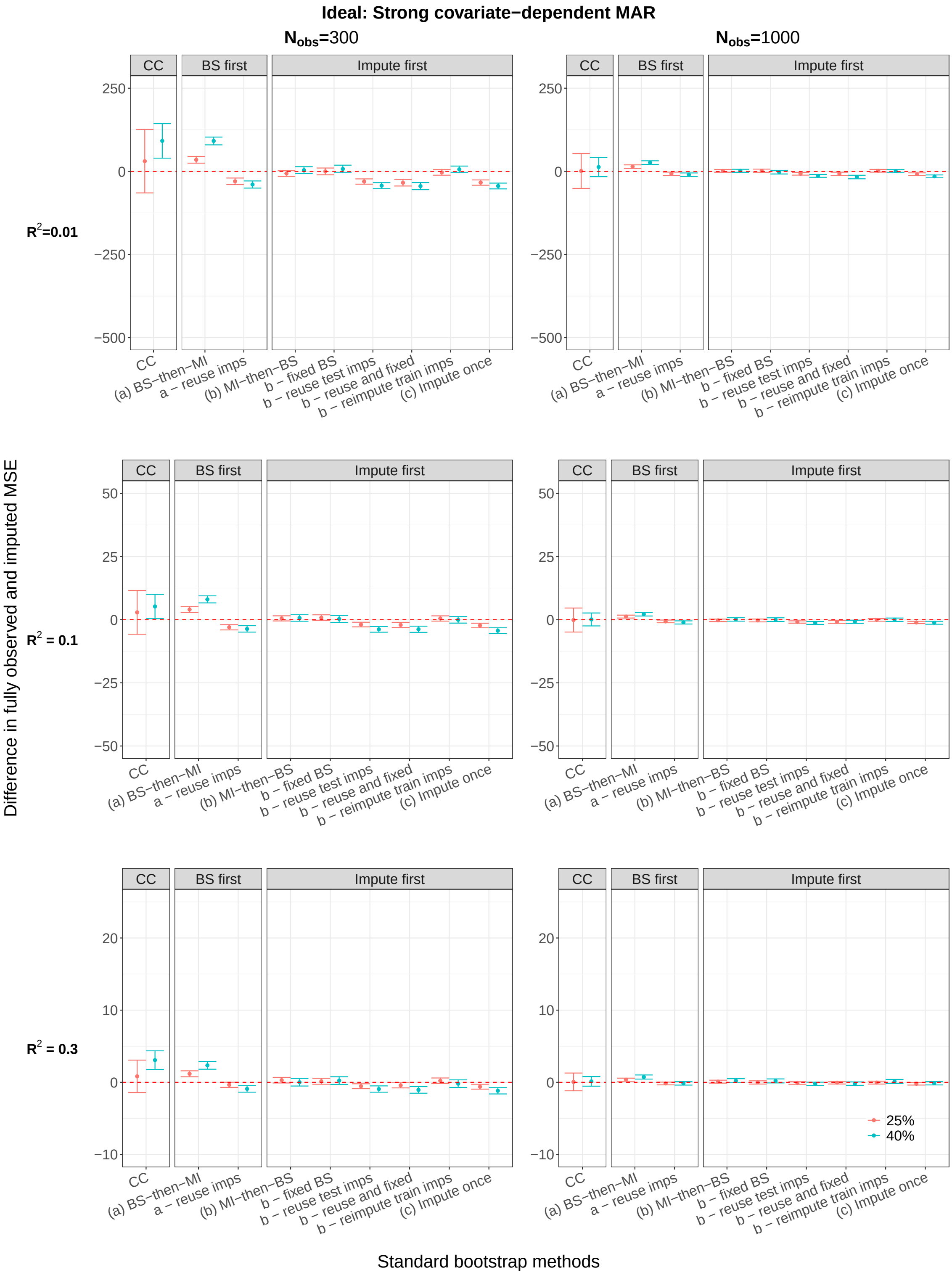

Figure S24: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

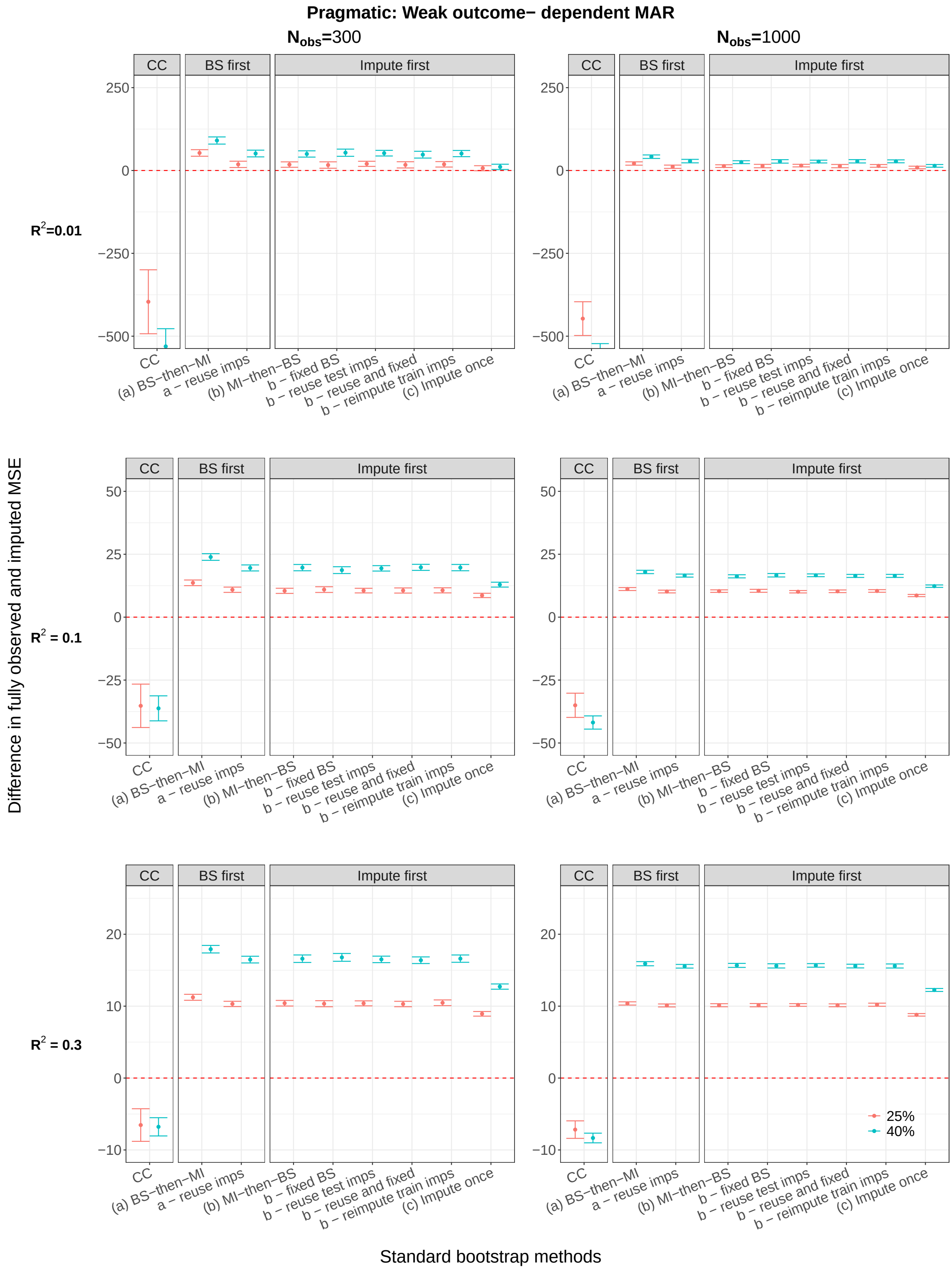

Figure S25: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

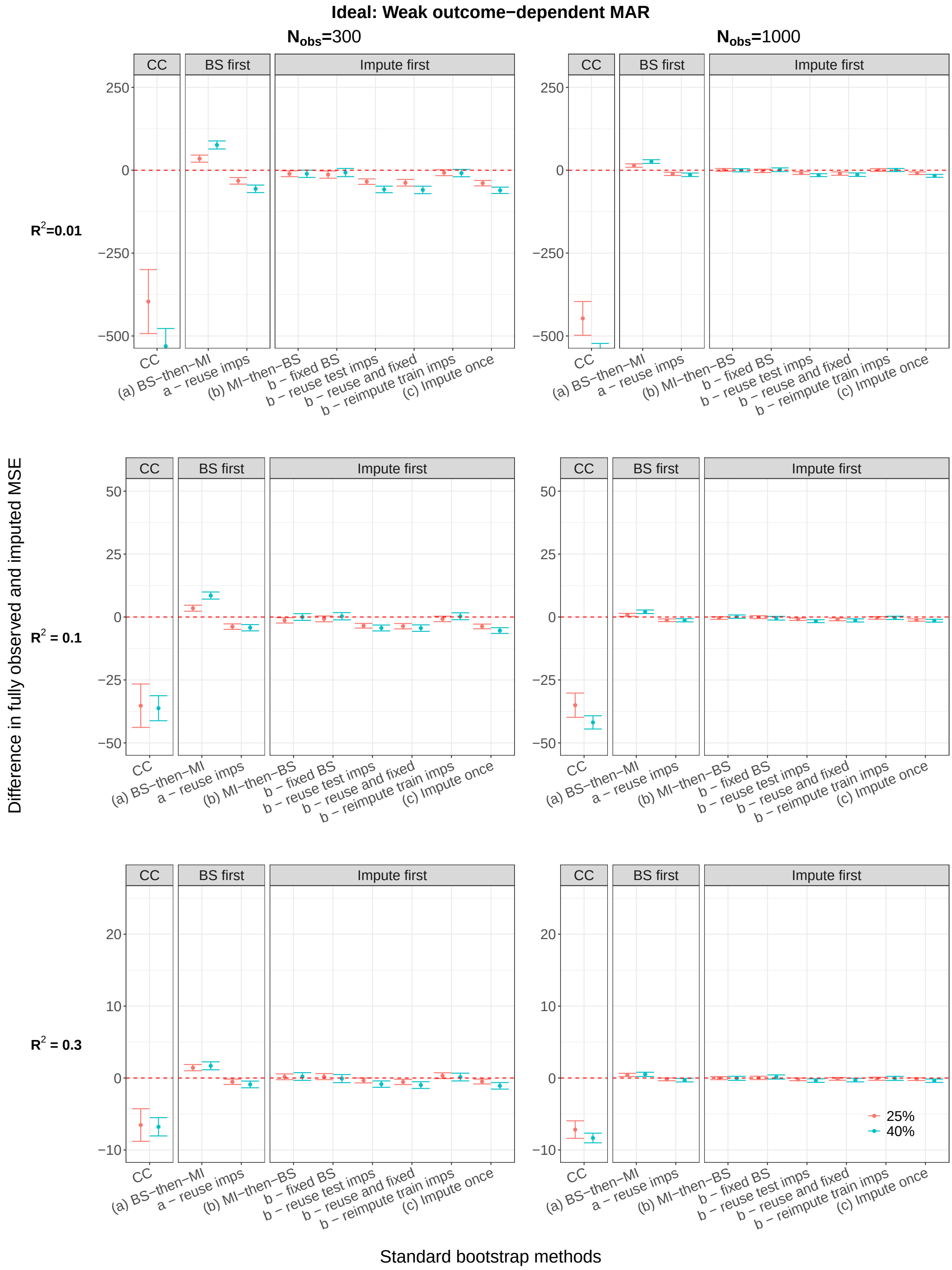

Figure S26: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

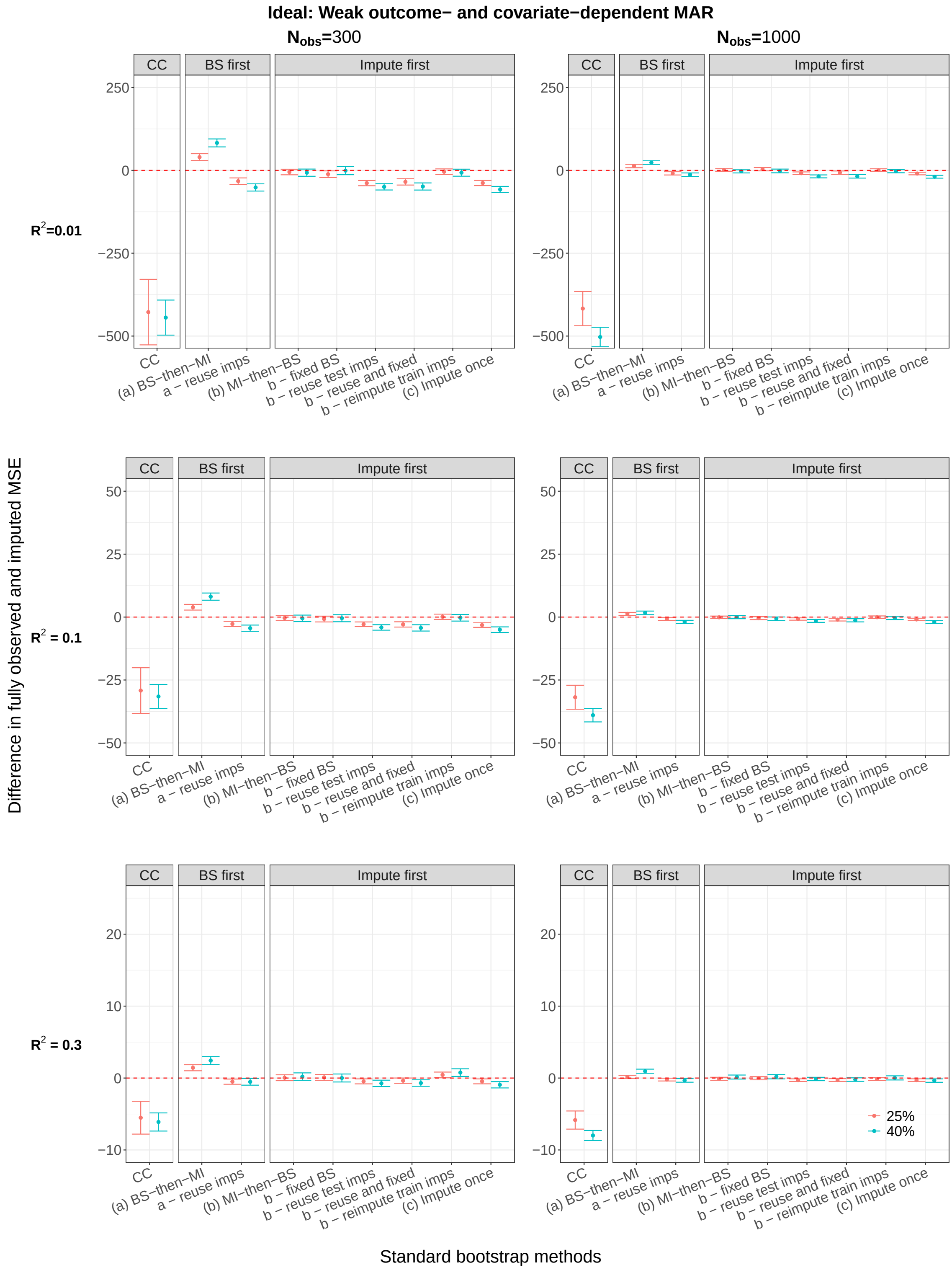

Figure S27: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

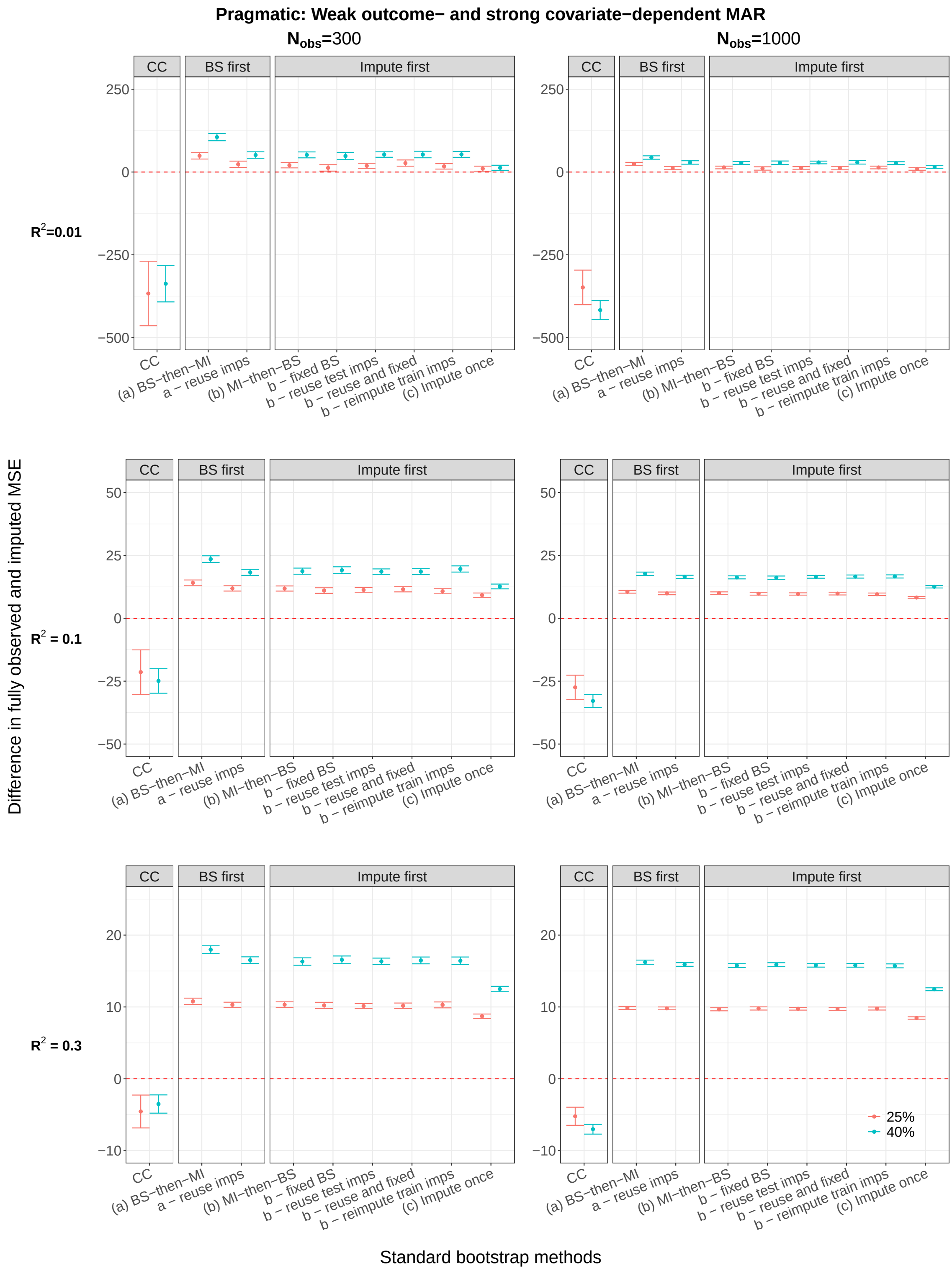

Figure S28: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR for the *standard* bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.2.3 Comparing M=5 versus M=25 (MSE_{imp}-MSE_{obs})

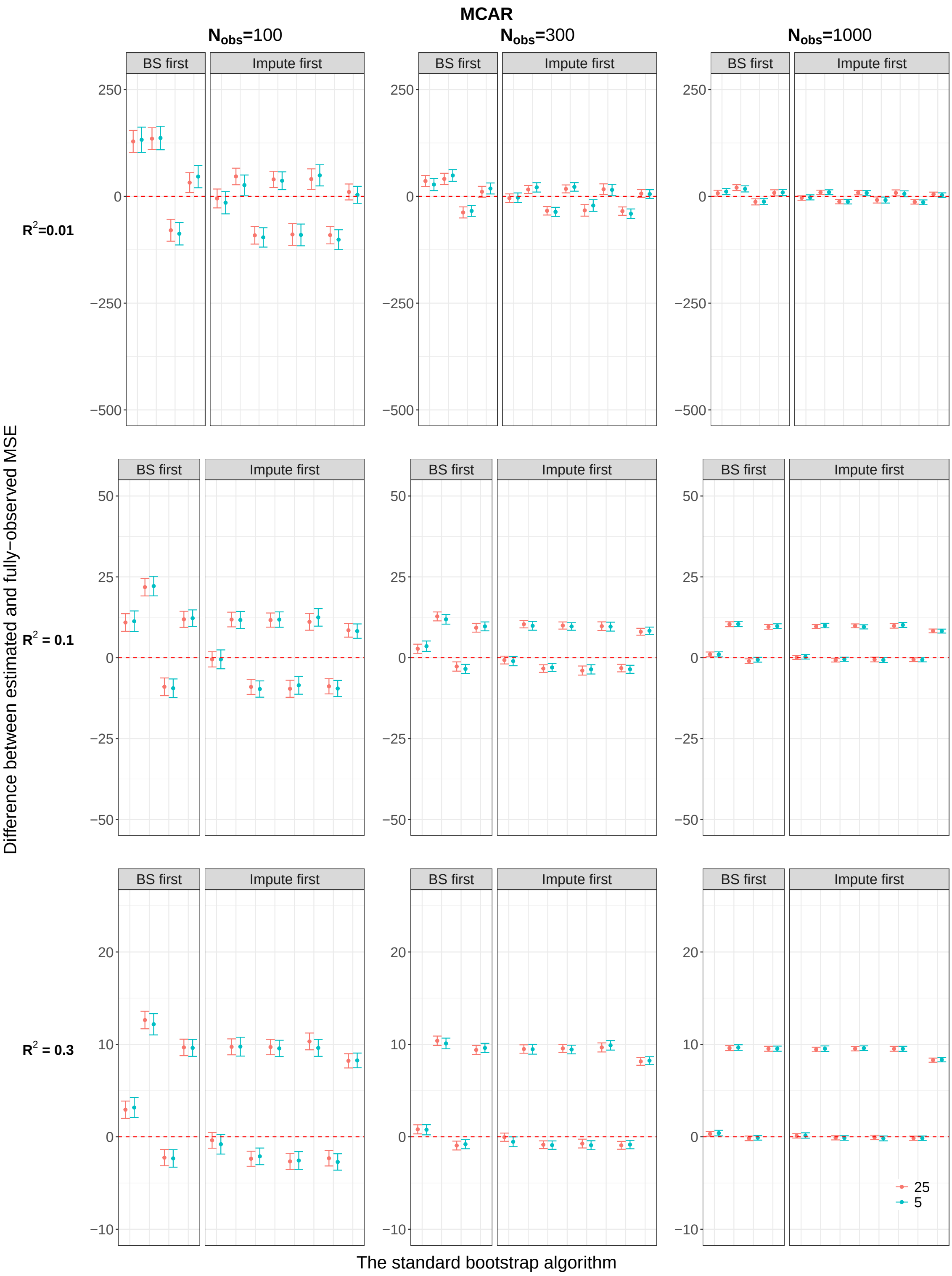

Figure S29: The difference MSE_{imp} - MSE_{obs} when data are MCAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

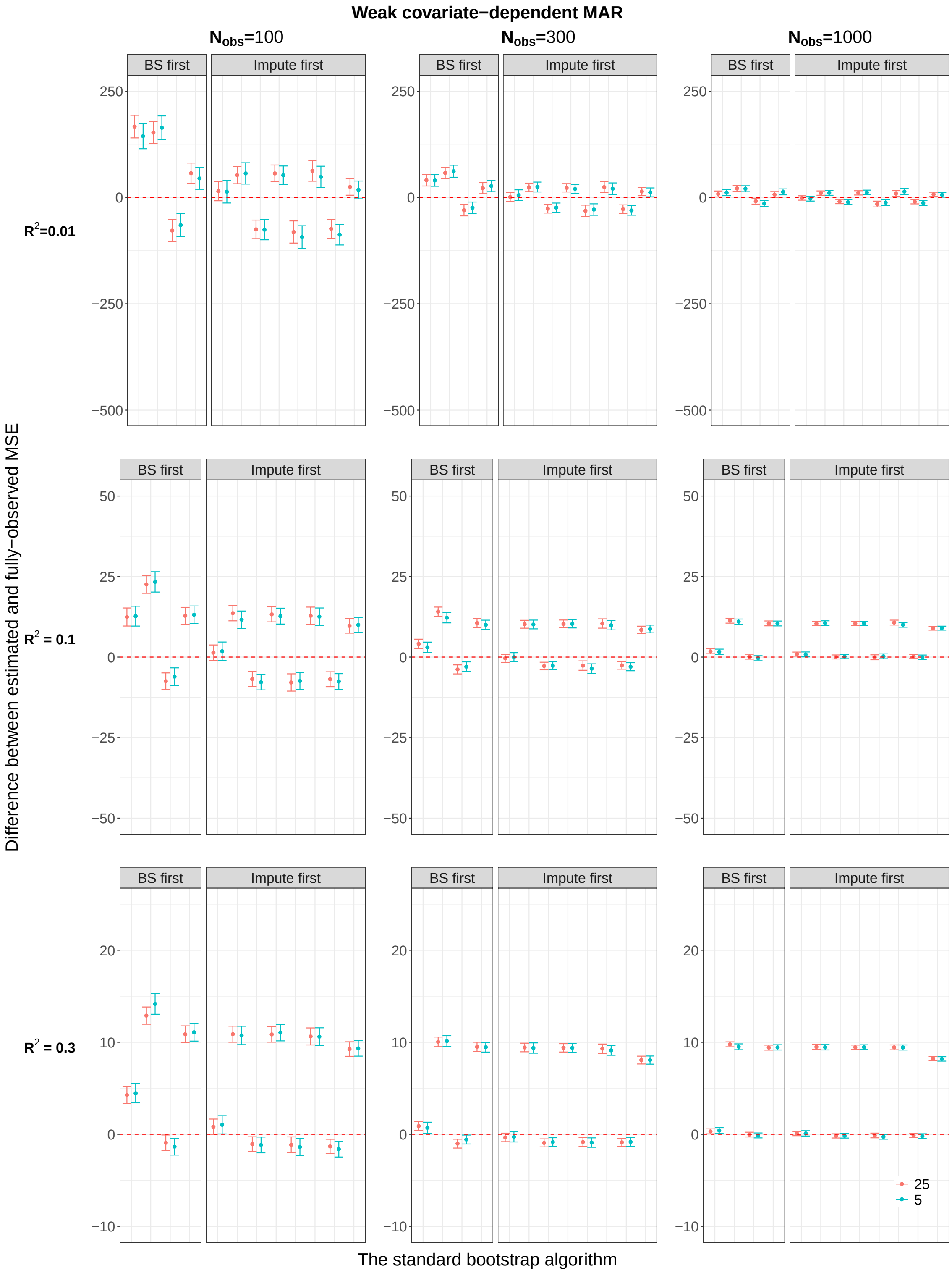

Figure S30: The difference MSE_{imp} - MSE_{obs} when data are weak covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

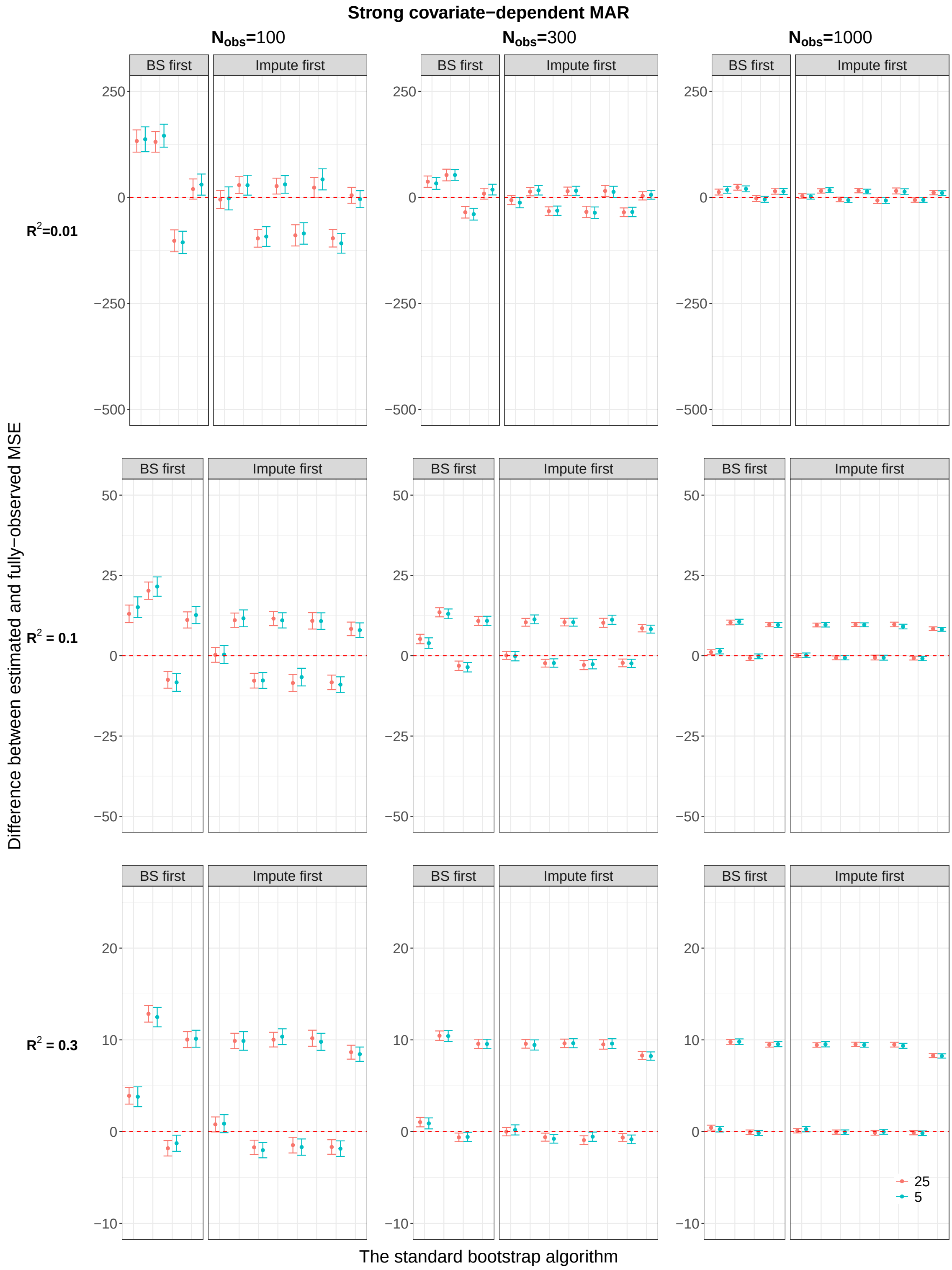

Figure S31: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

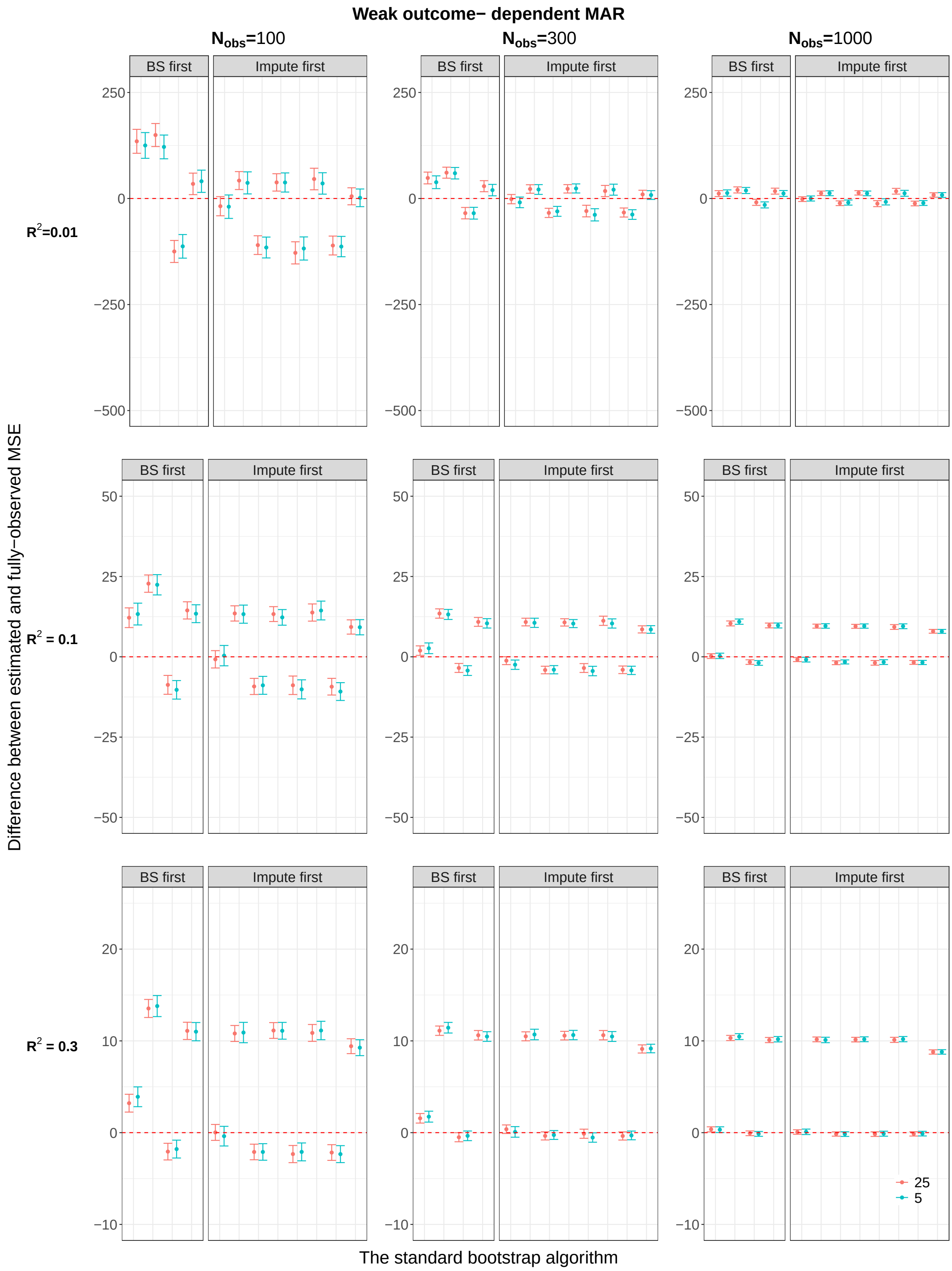

Figure S32: The difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

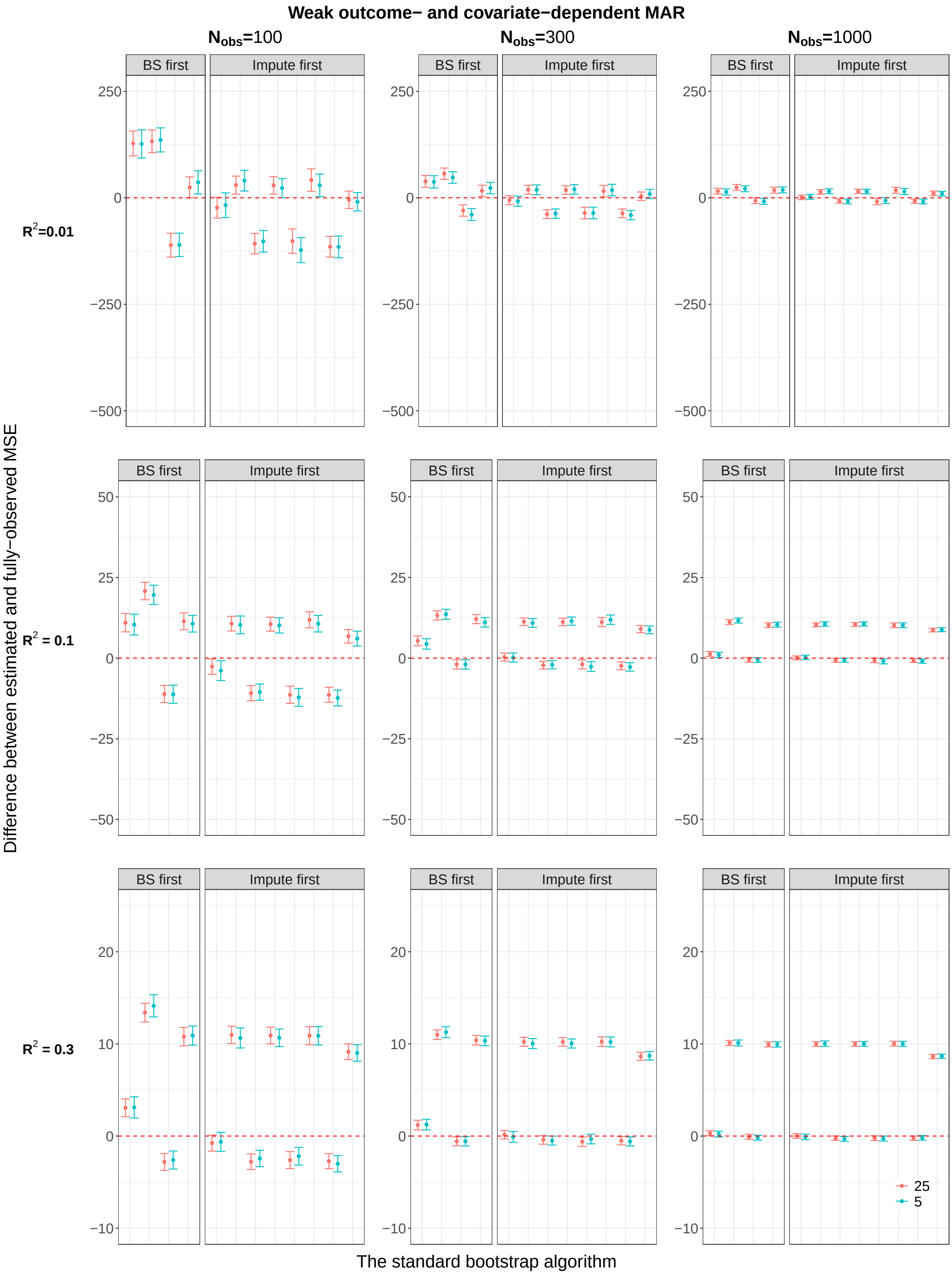

Figure S33: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

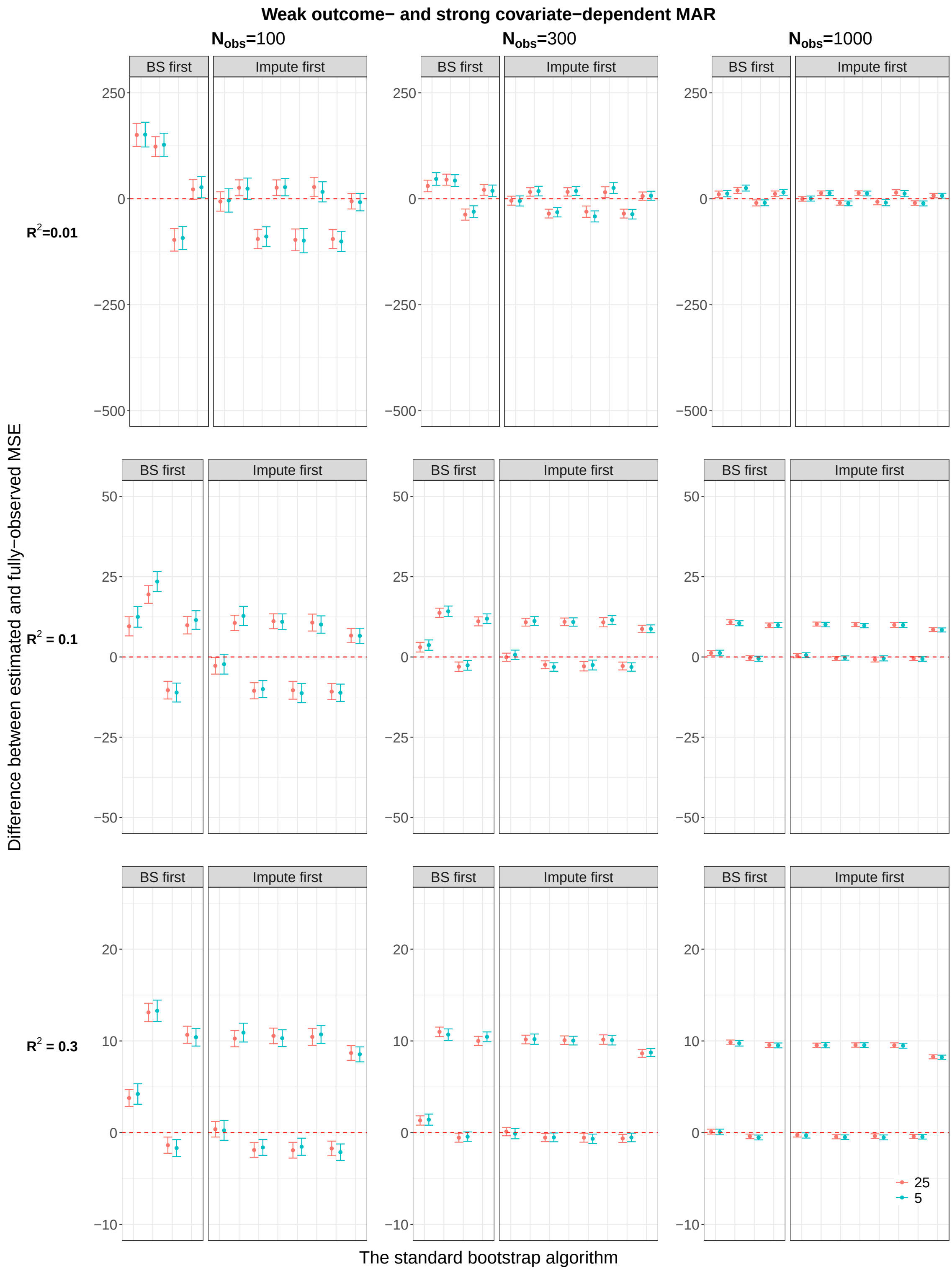

Figure S34: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and strong covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.2.4 MSE from imputation methods compared to the target MSE (MSE_{target}) using a larger validation set

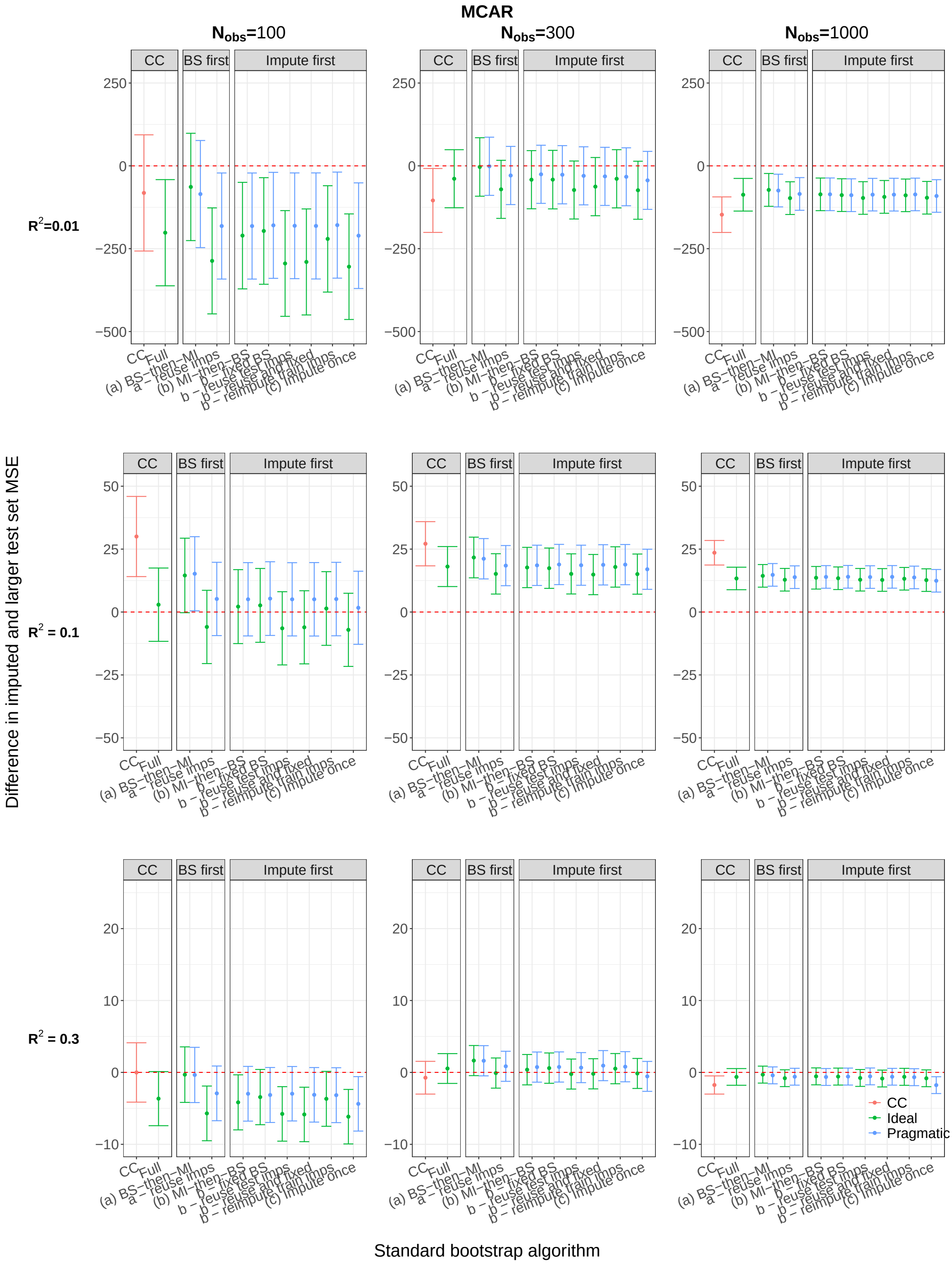

Figure S35: The difference MSE_{imp} - MSE_{target} when data are MCAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

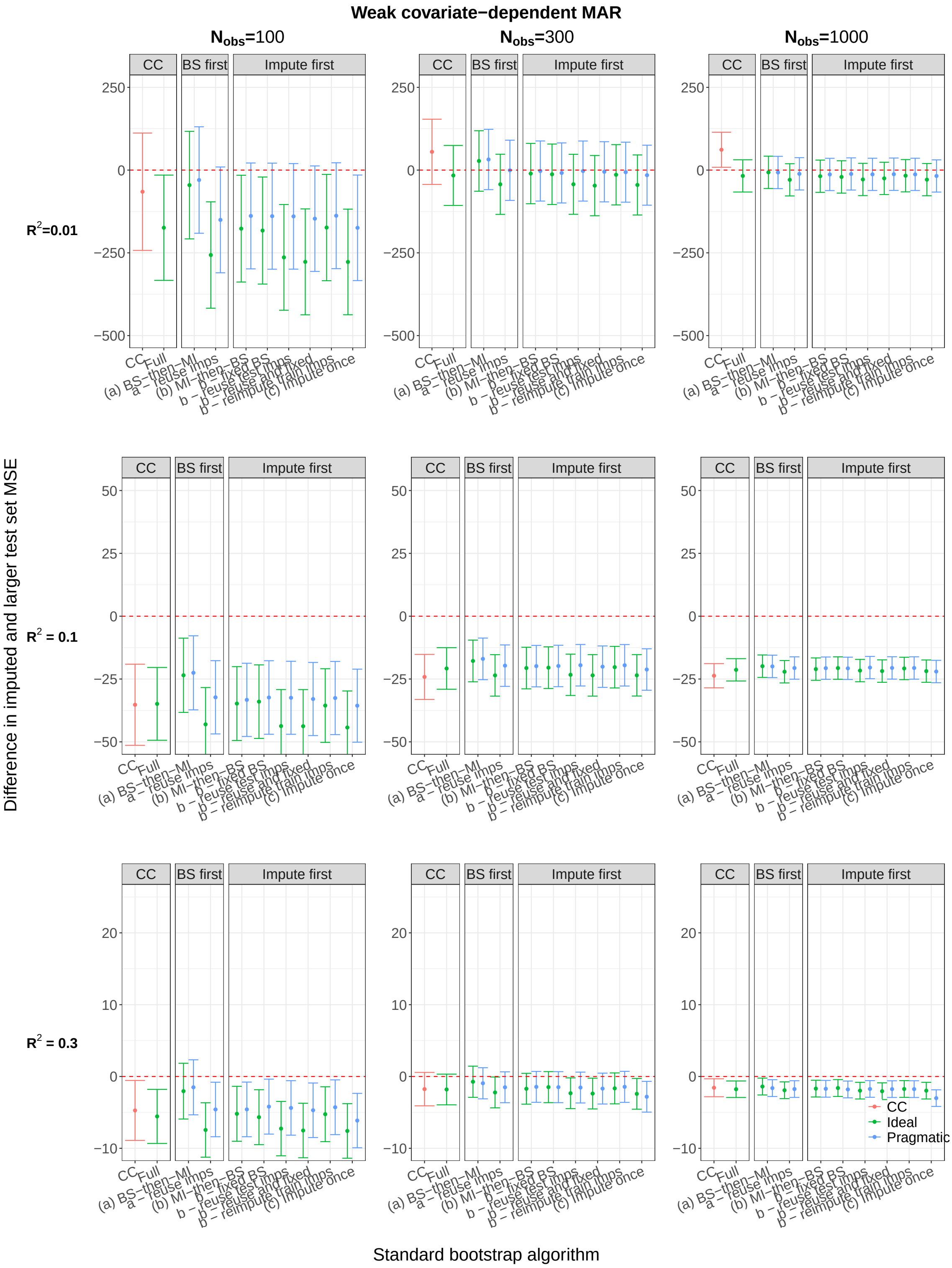

Figure S36: The difference MSE_{imp} - MSE_{target} when data are weak covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

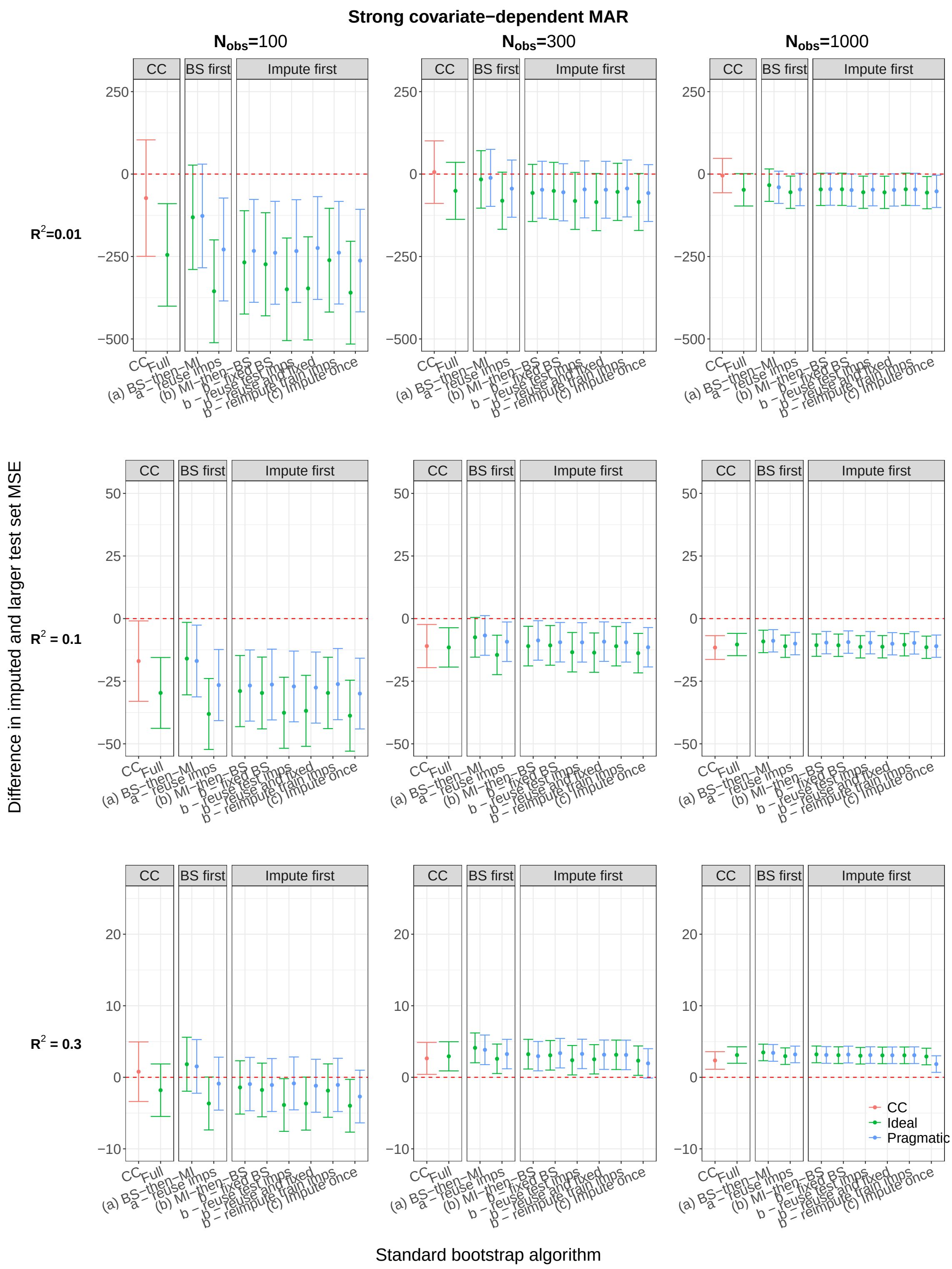

Figure S37: The difference MSE$_{imp}$ - MSE$_{target}$ when data are strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

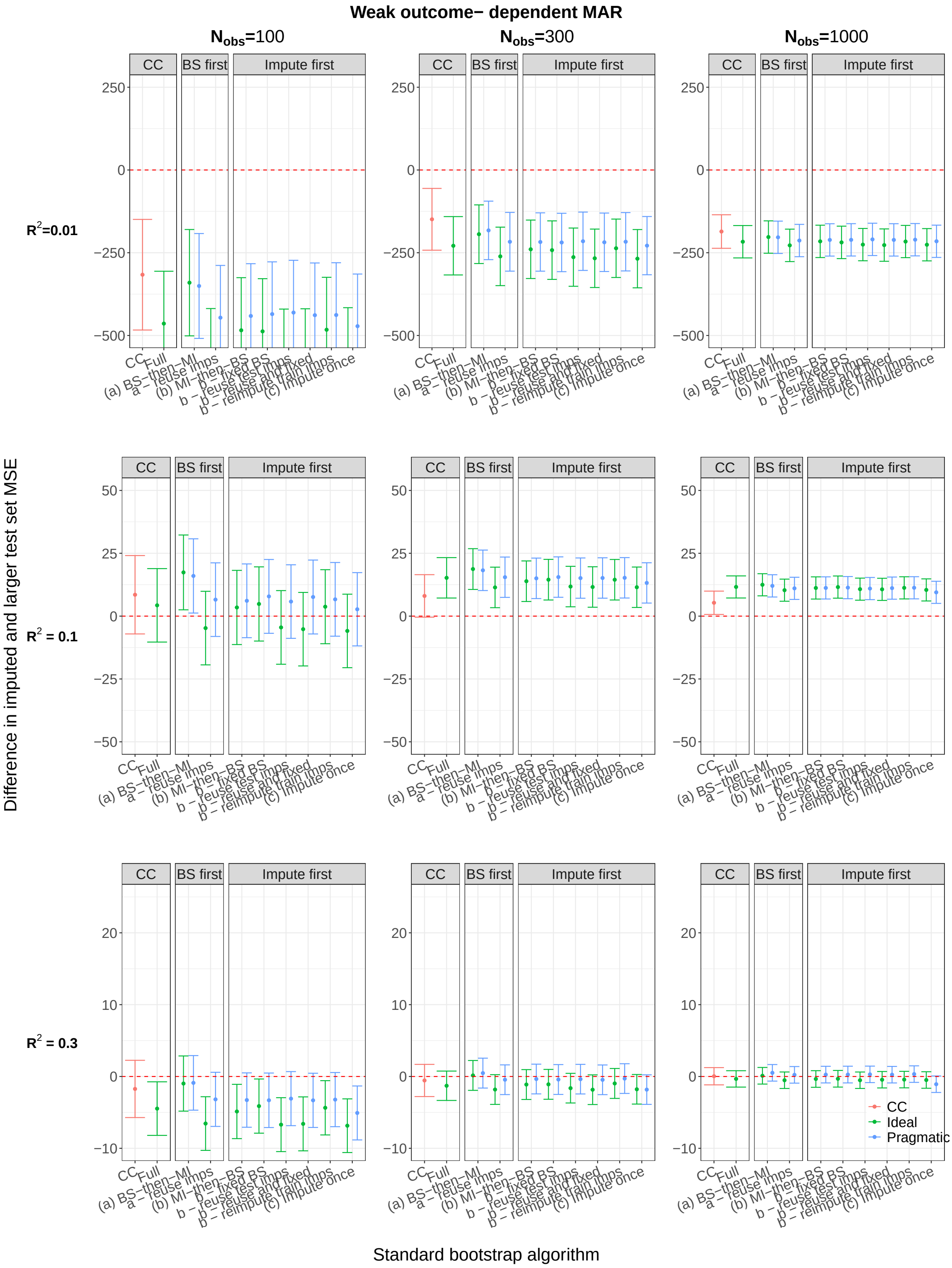

Figure S38: The difference MSE_{imp} - MSE_{target} when data are weak outcome-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

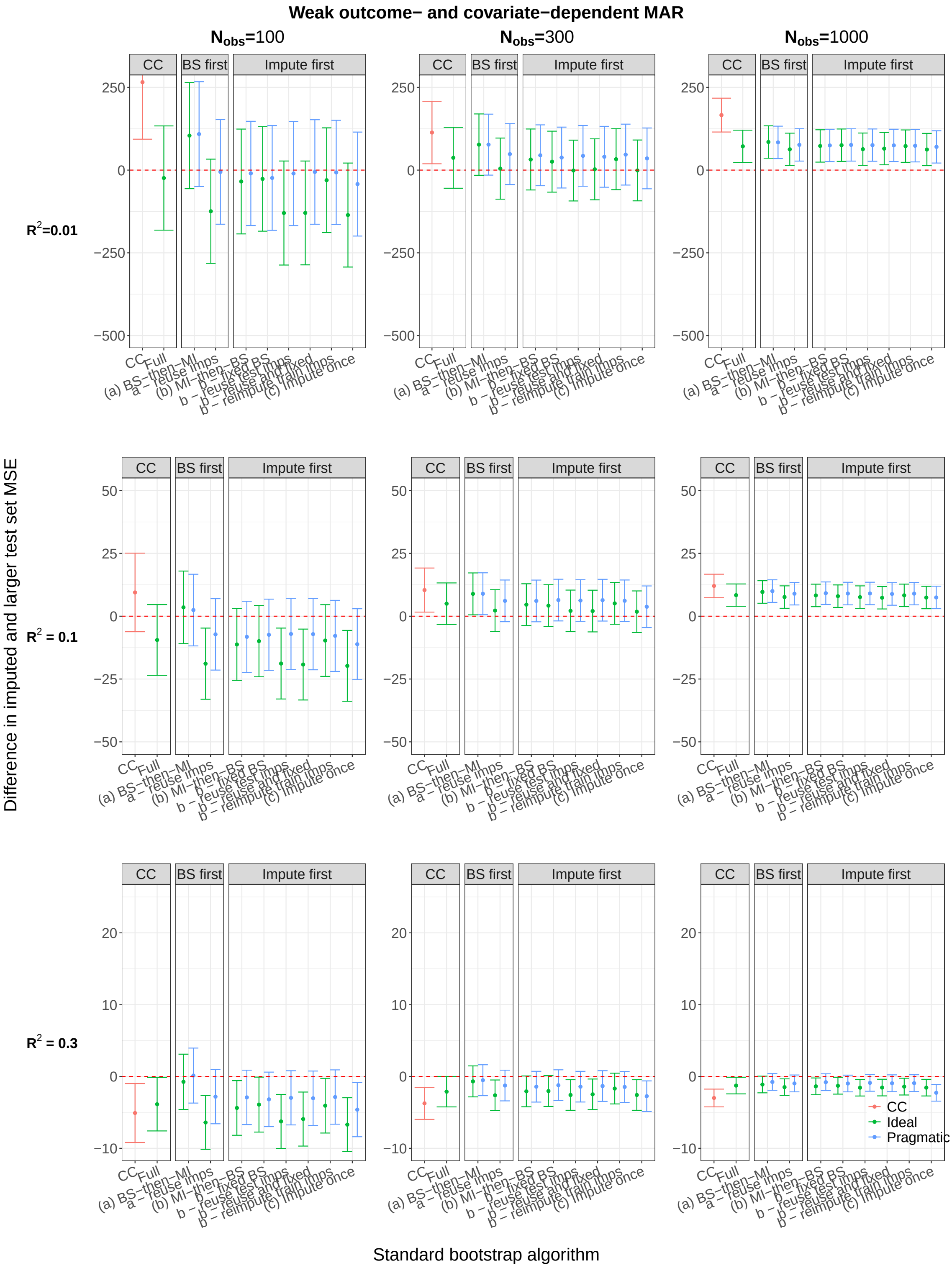

Figure S39: The difference MSE$_{imp}$ - MSE$_{target}$ when data are weak outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

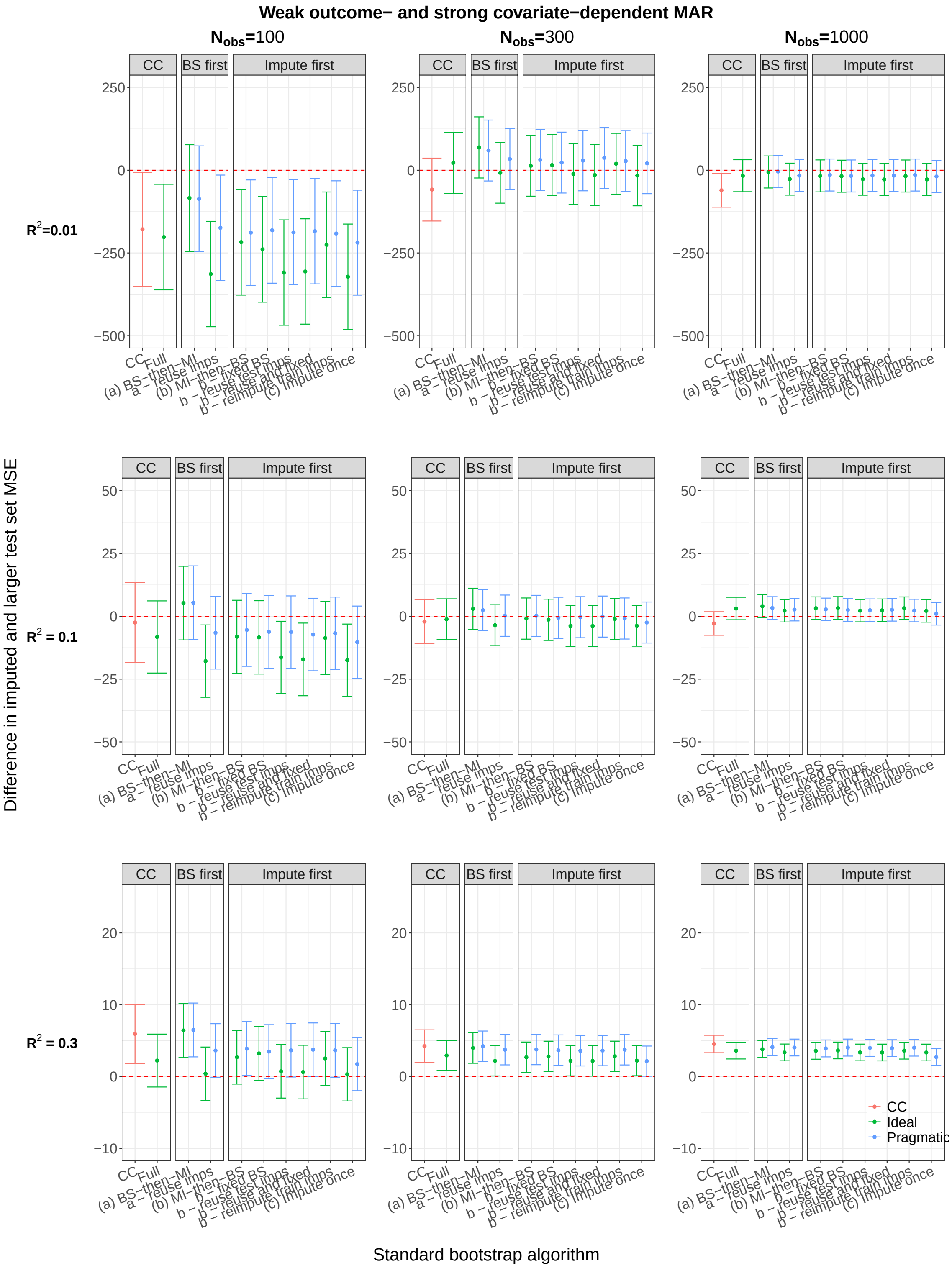

Figure S40: The difference MSE_{imp} - MSE_{target} when data are weak outcome- and strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.3 The 0.632 bootstrap

S3.3.1 MSE from imputation methods compared to the fully-observed MSE (MSE_{imp}-MSE_{obs})

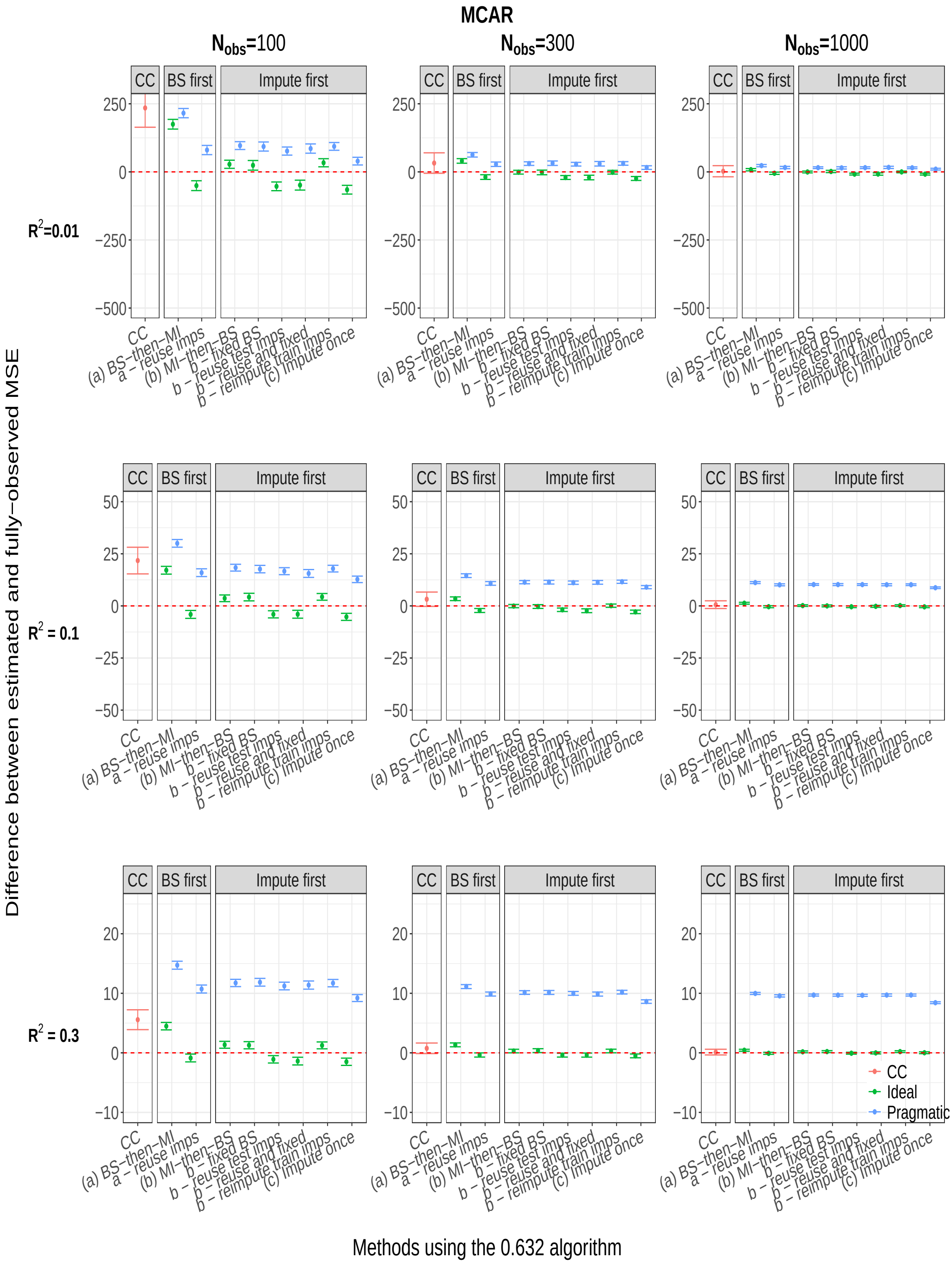

Figure S41: The difference MSE_{imp} - MSE_{obs} when data are MCAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

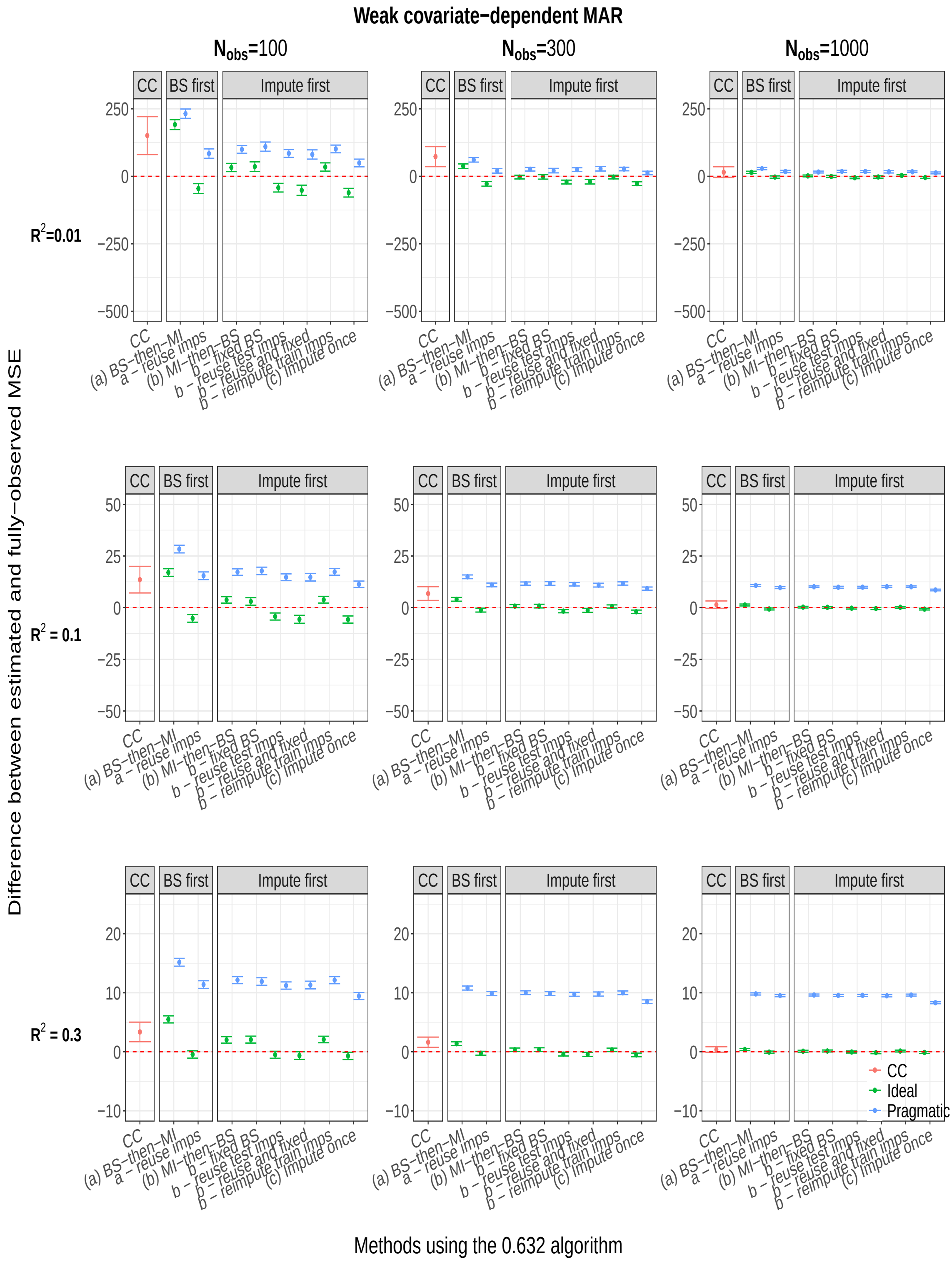

Figure S42: The difference MSE_{imp} - MSE_{obs} when data are weak covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

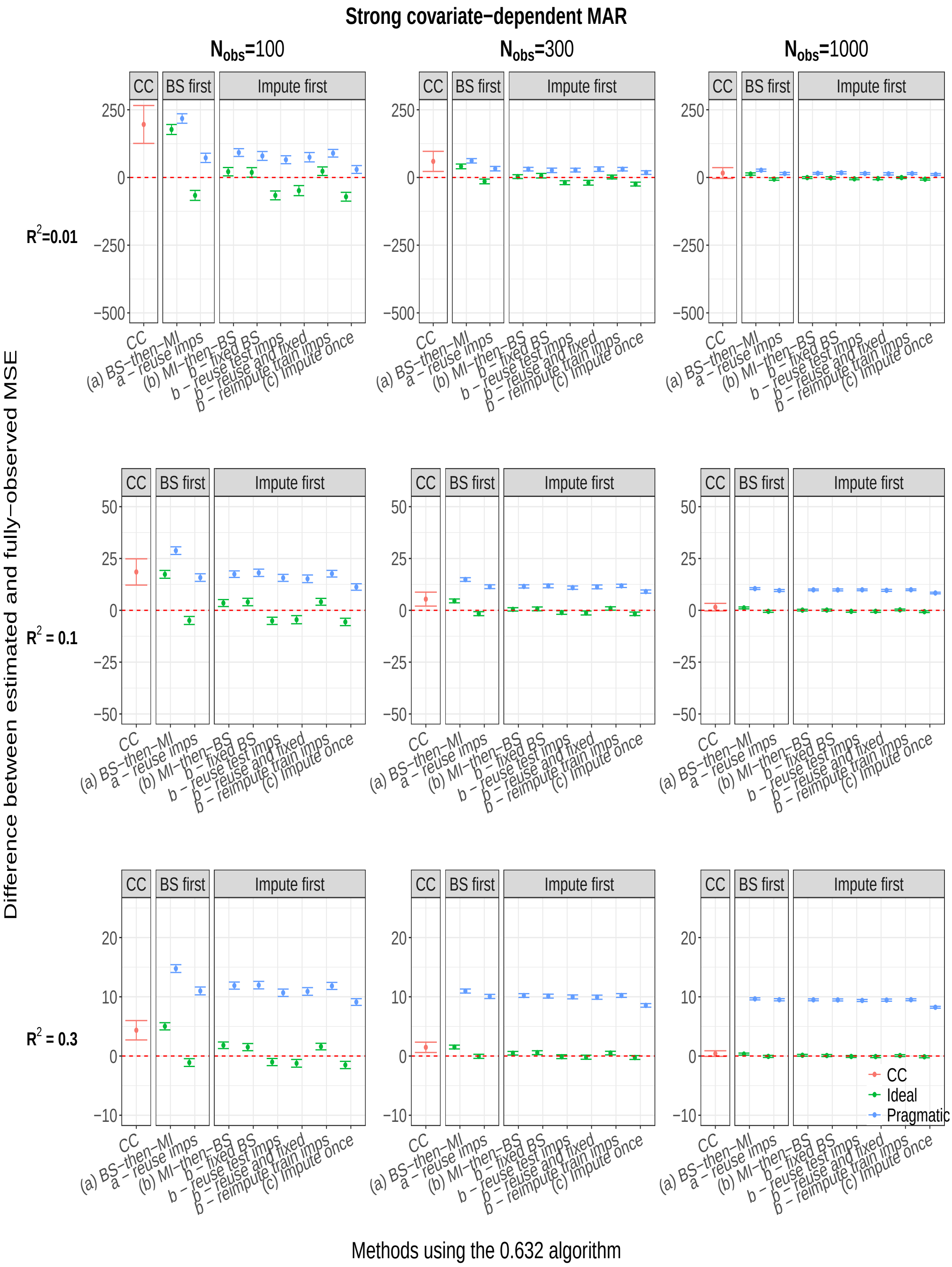

Figure S43: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

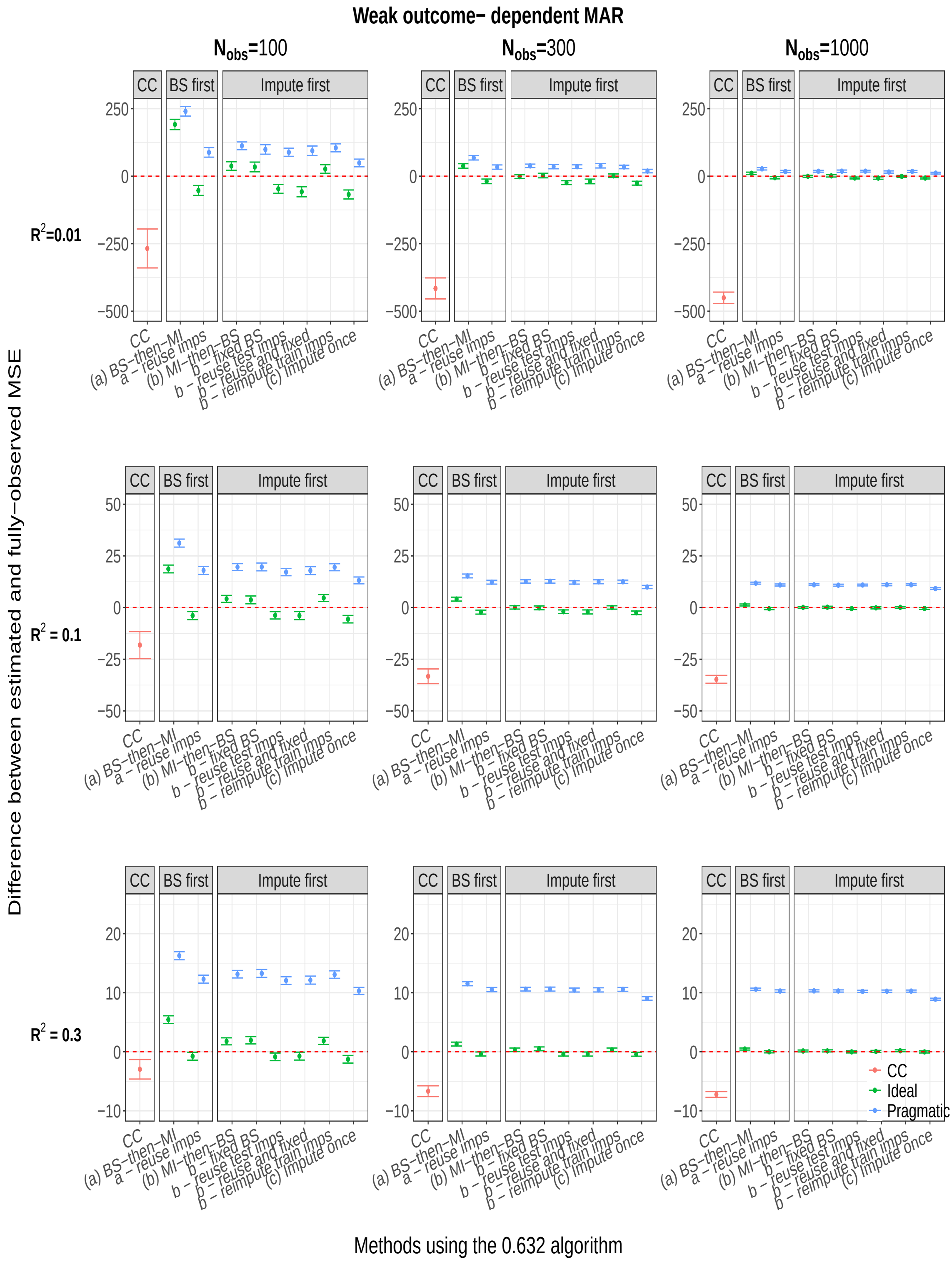

Figure S44: The difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

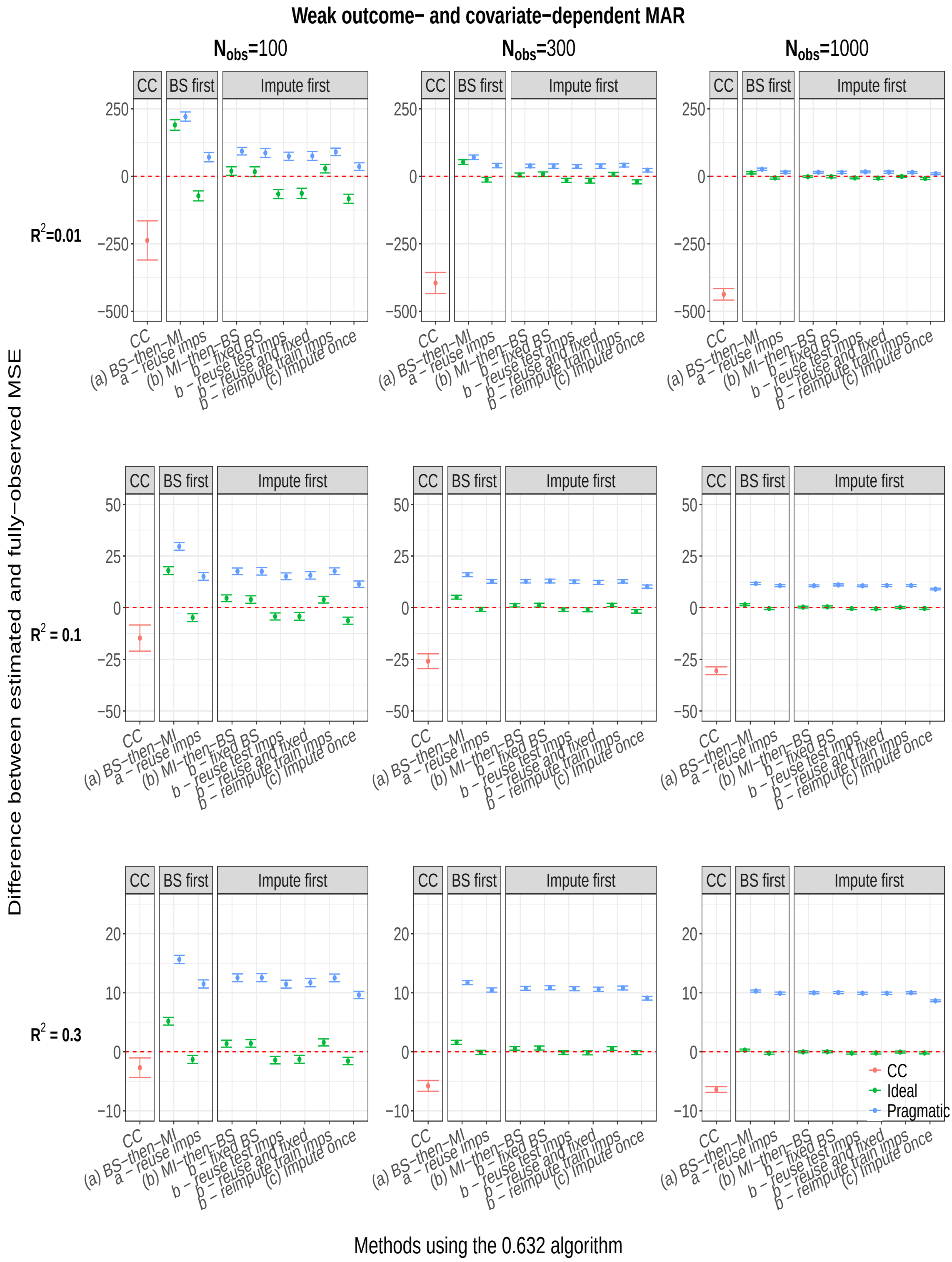

Figure S45: The difference MSE_{imp} - MSE_{obs} when data are weak outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

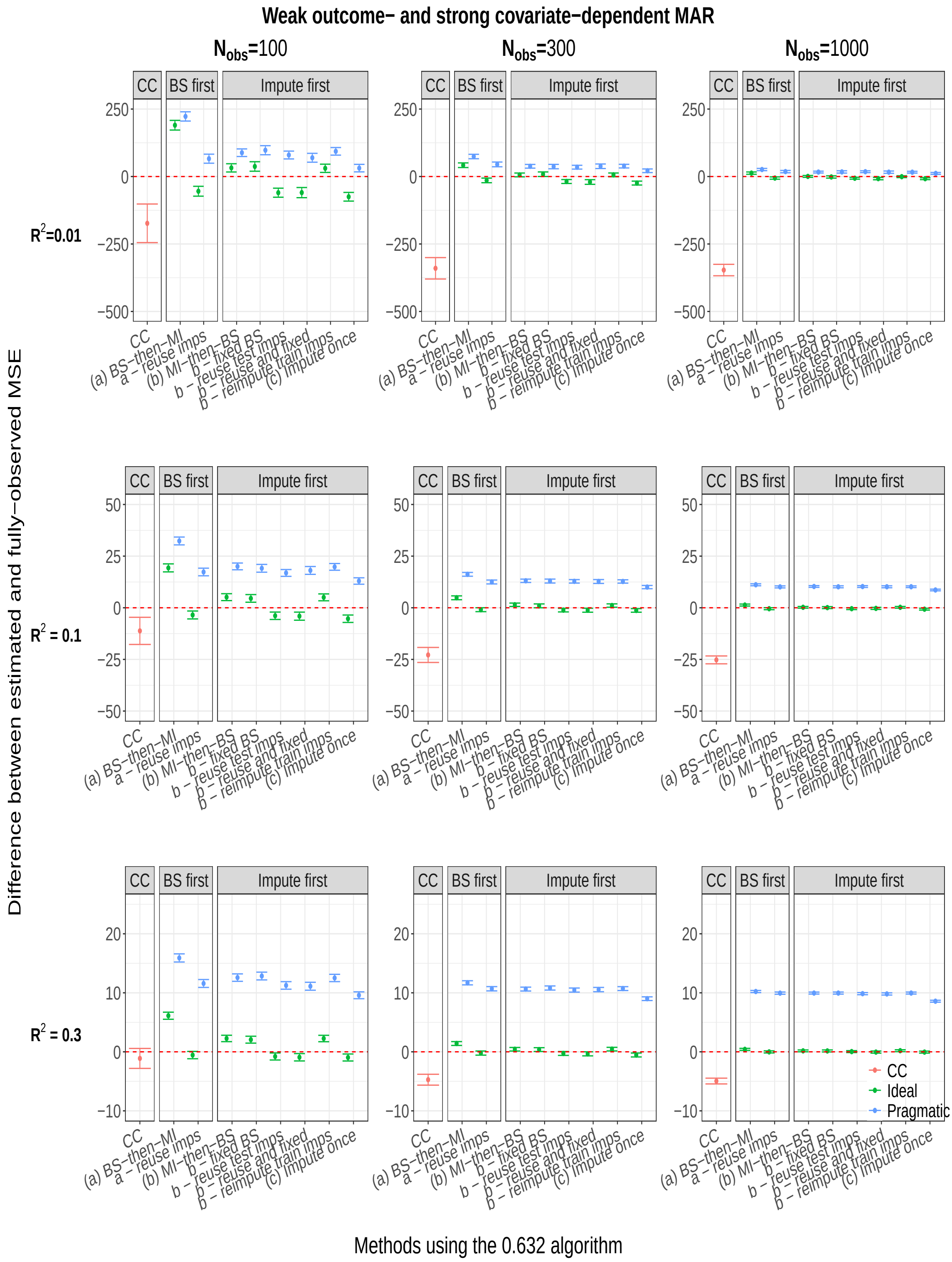

Figure S46: The difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.3.2 The proportion of missingness is 40% (MSE_{imp}-MSE_{obs})

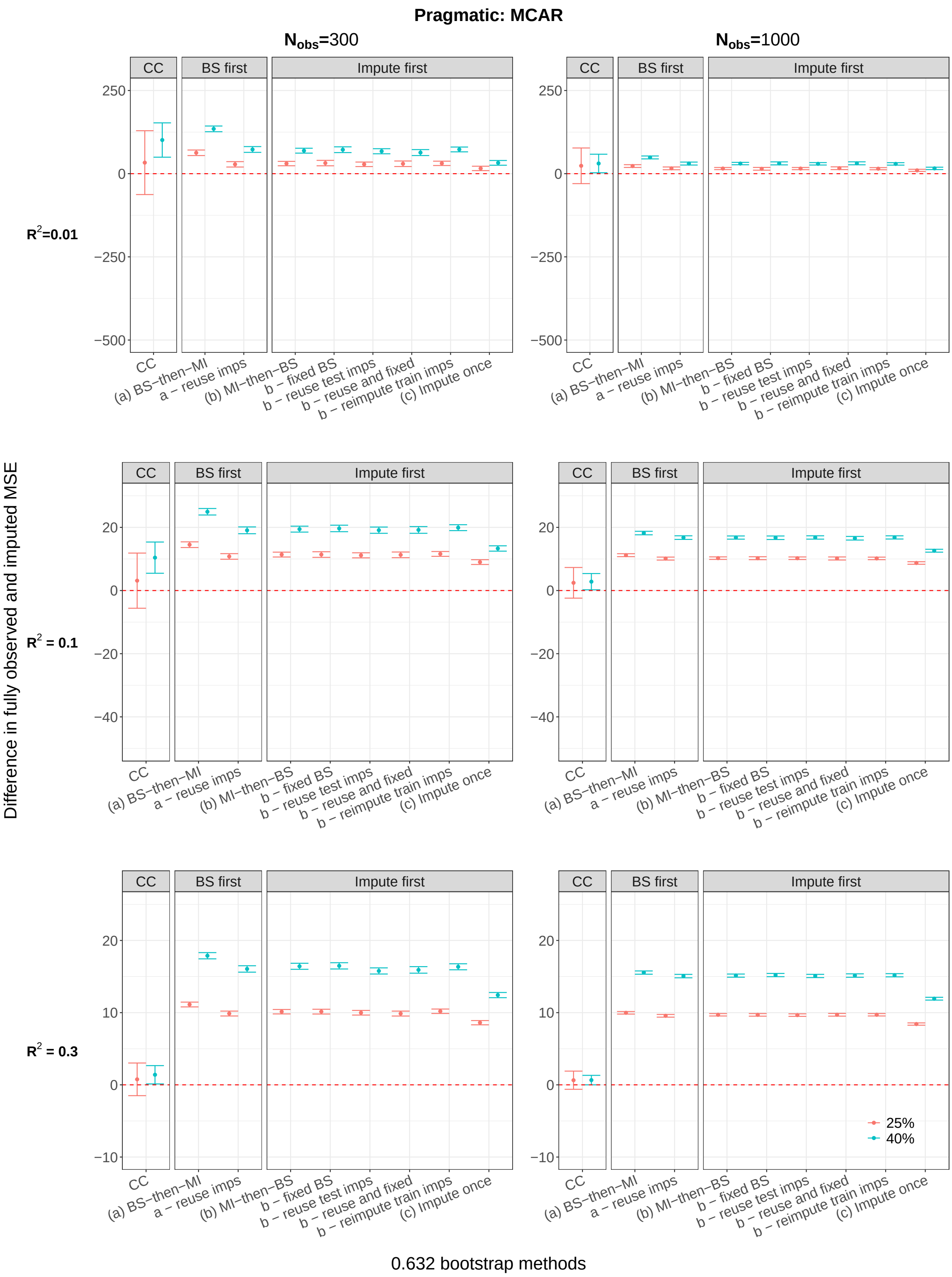

Figure S47: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are MCAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

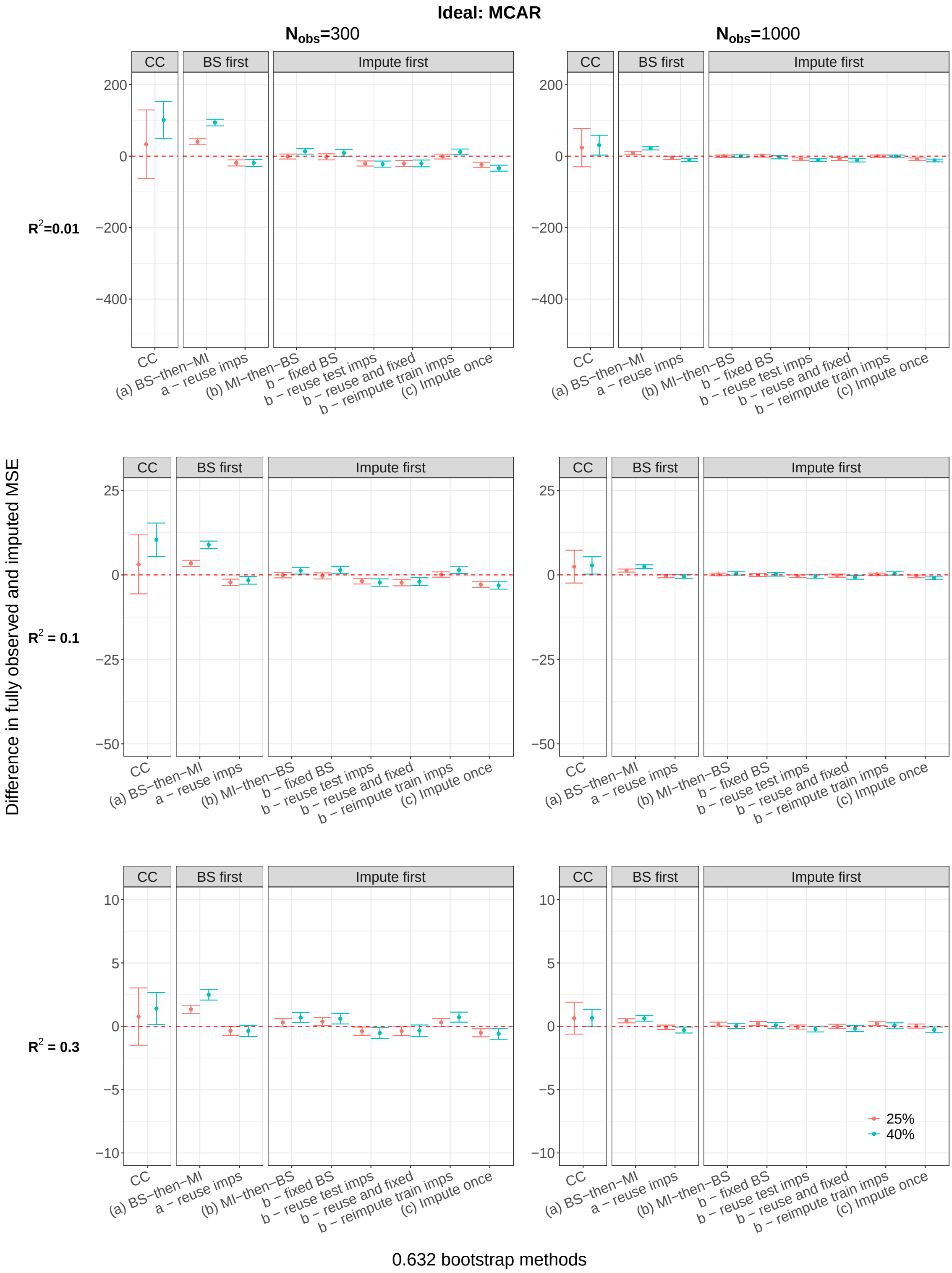

Figure S48: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are MCAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

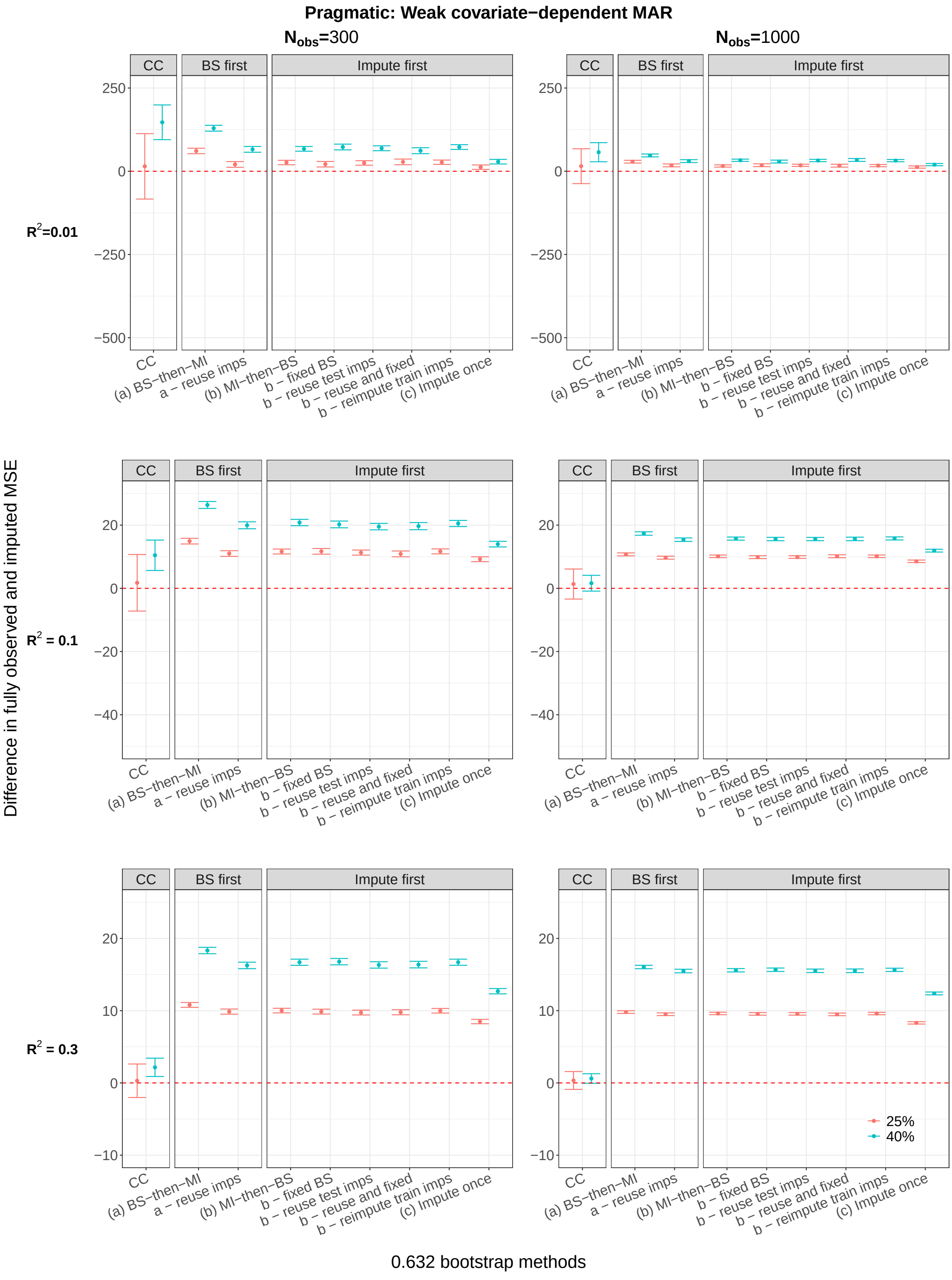

Figure S49: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

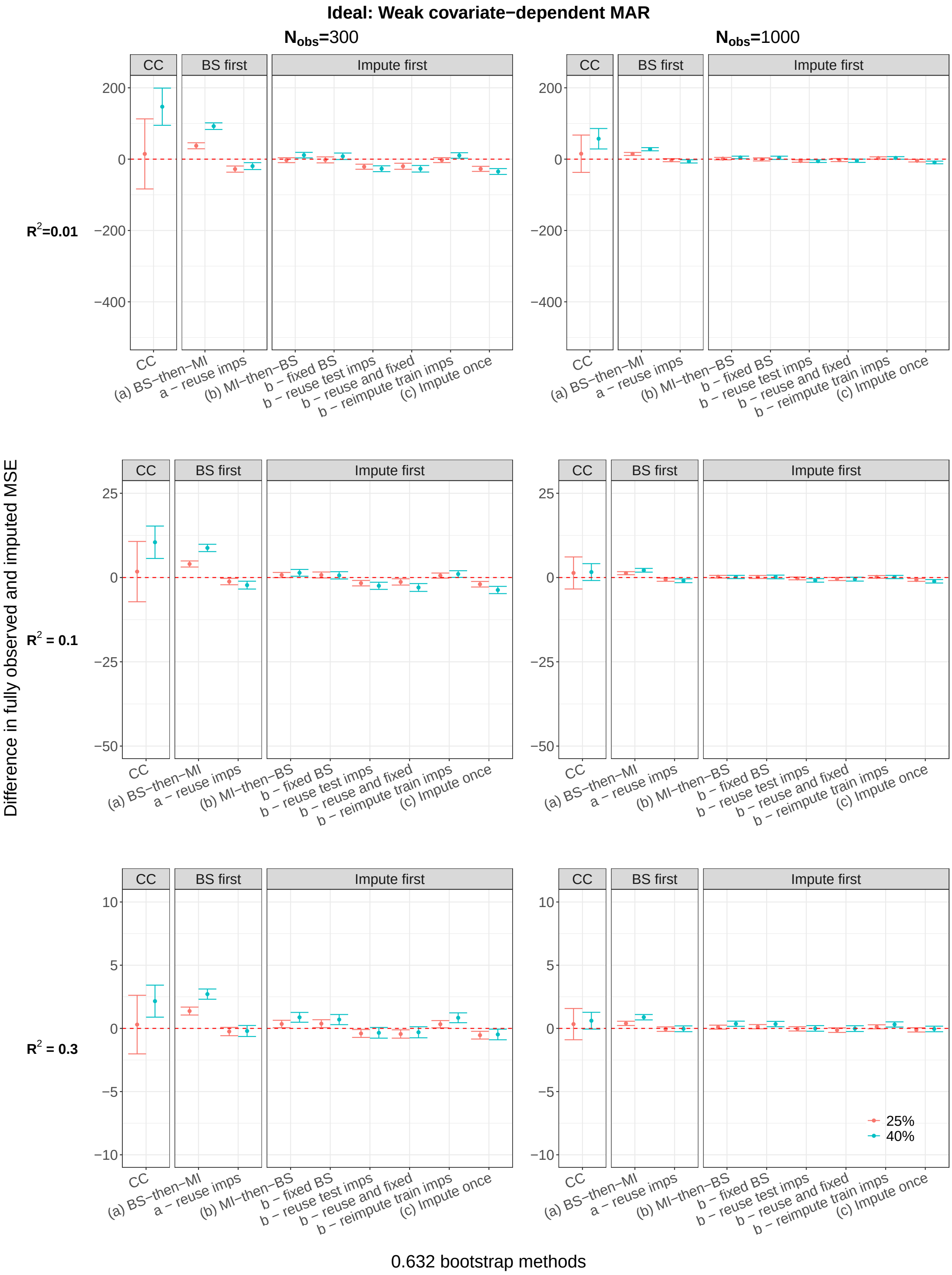

Figure S50: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

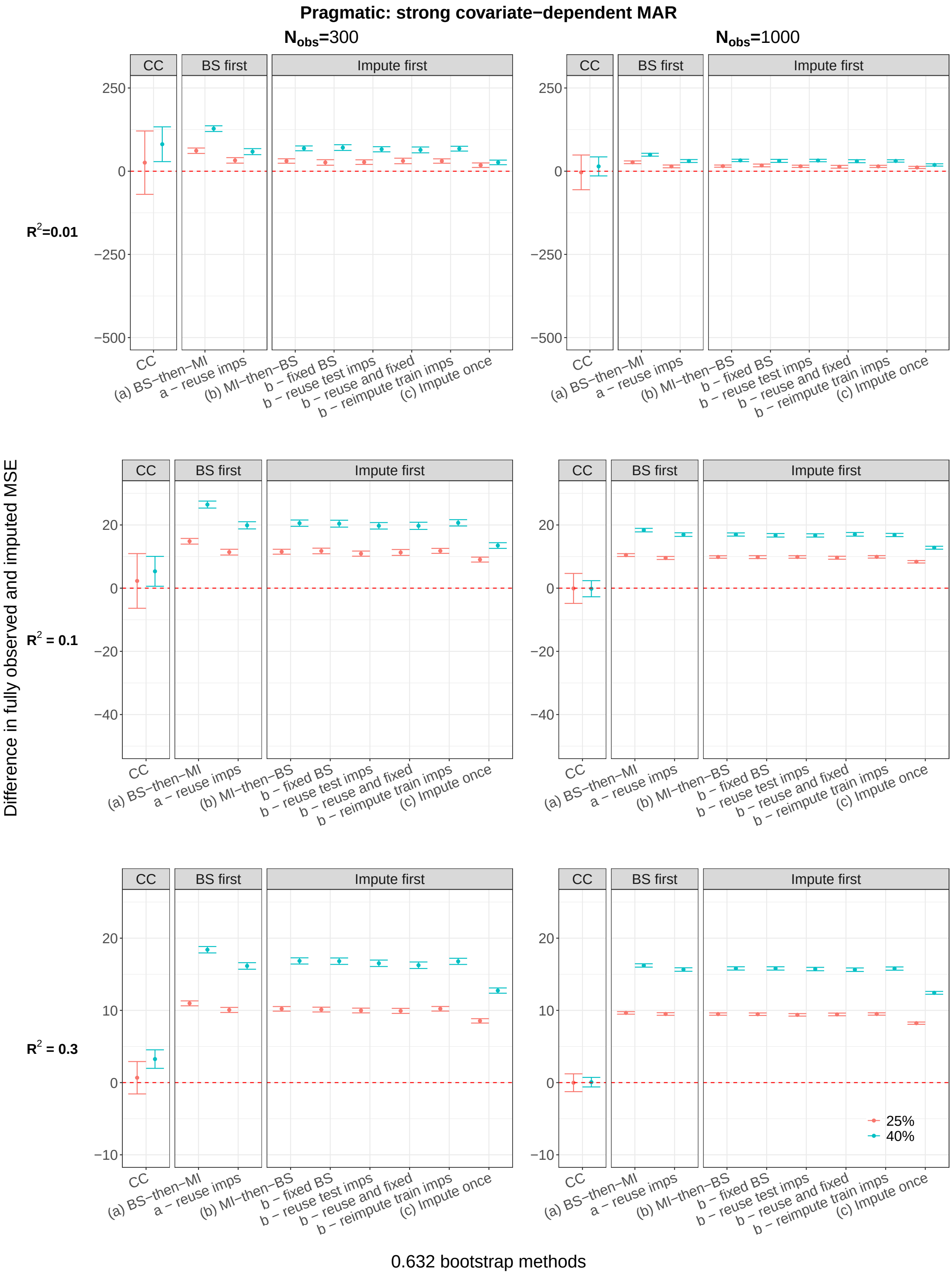

Figure S51: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

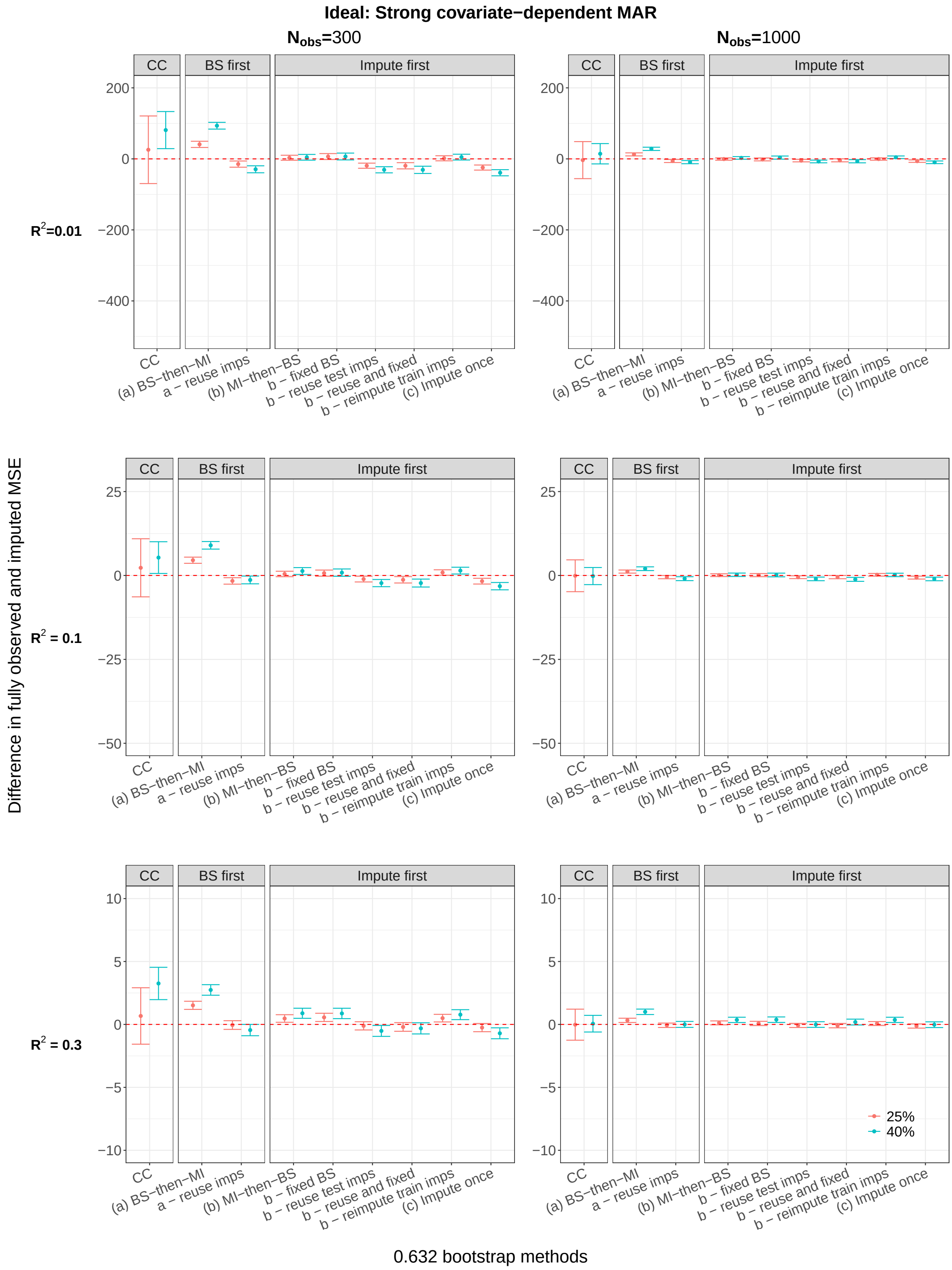

Figure S52: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are strong covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

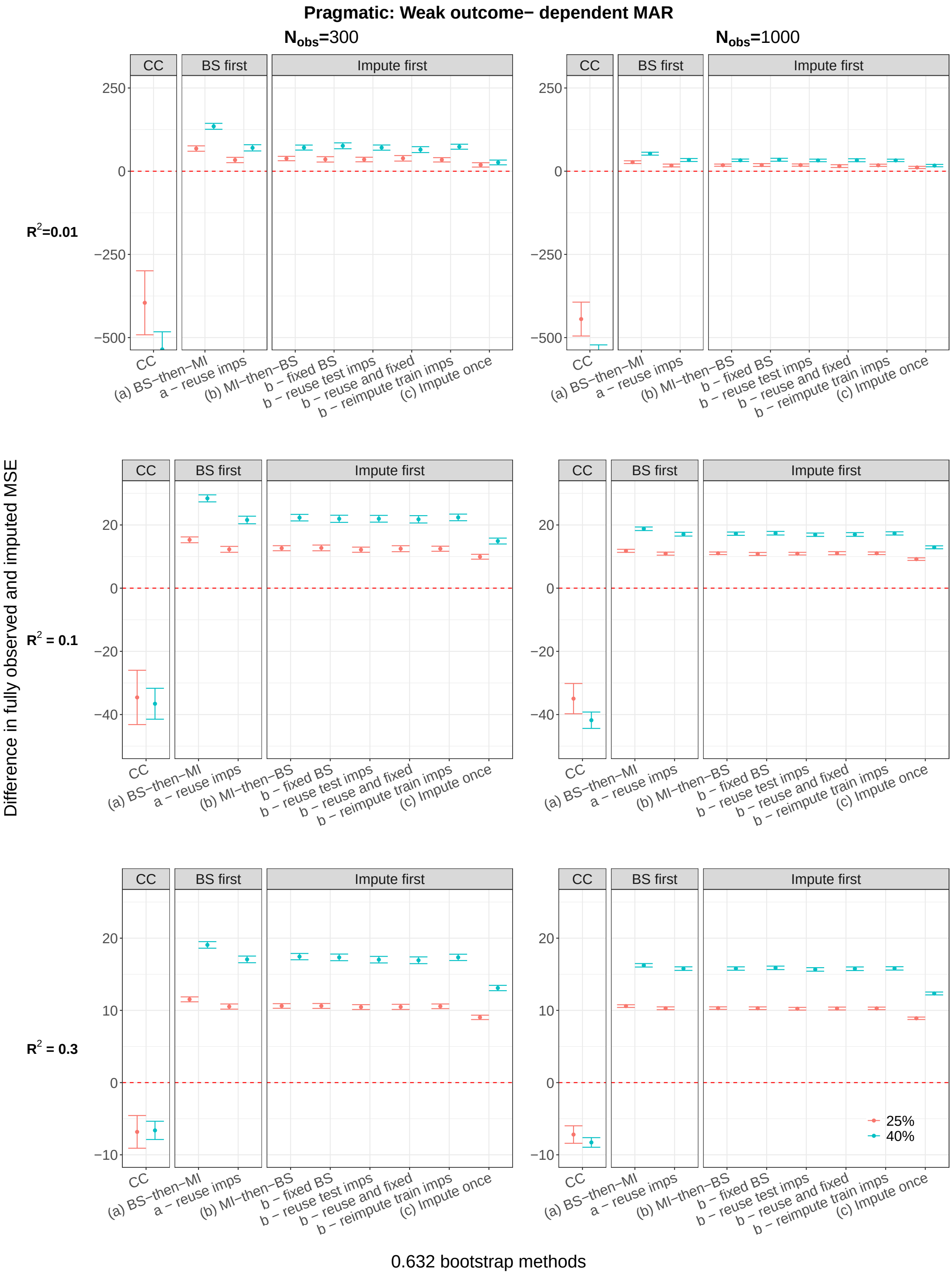

Figure S53: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

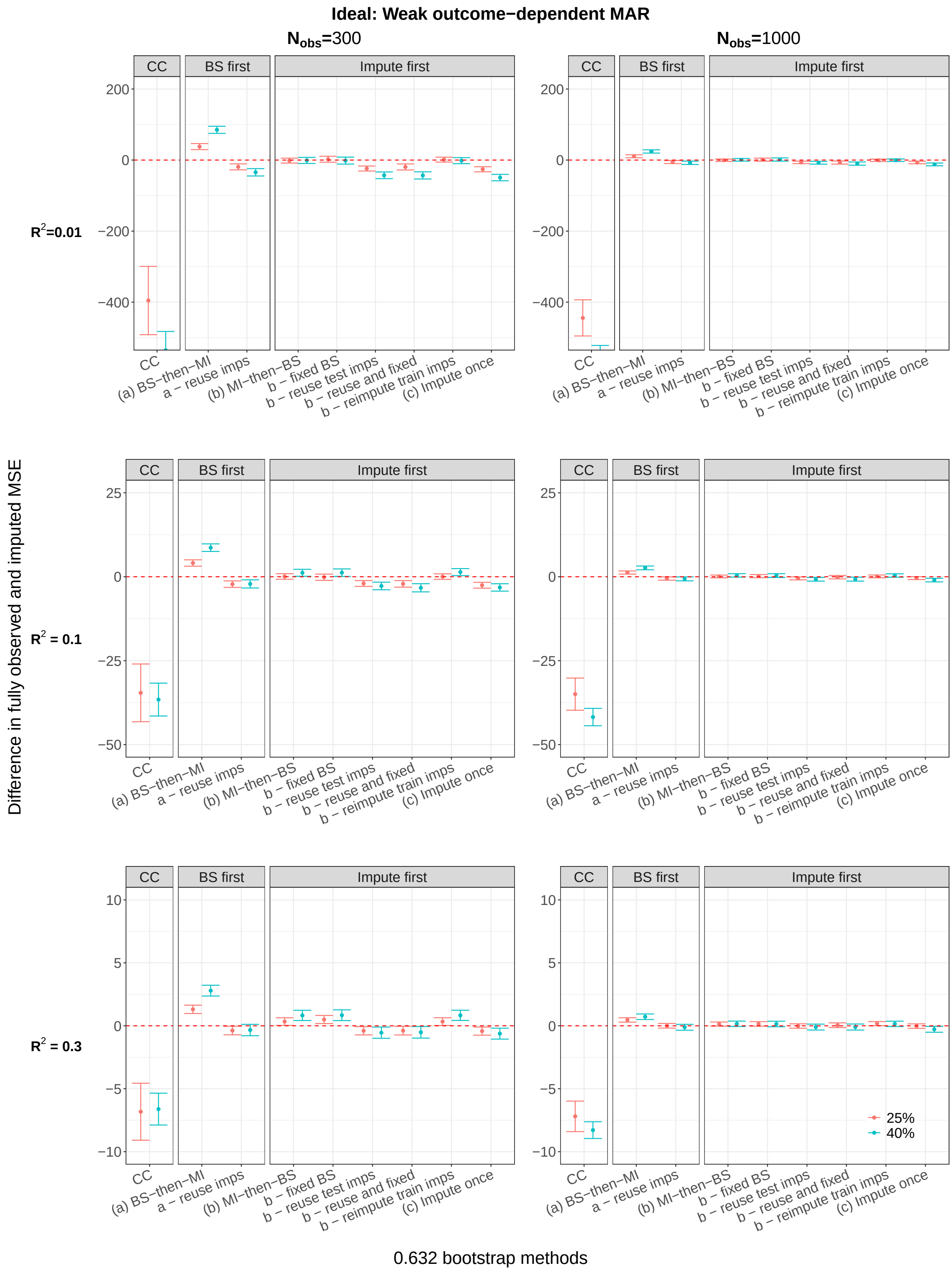

Figure S54: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

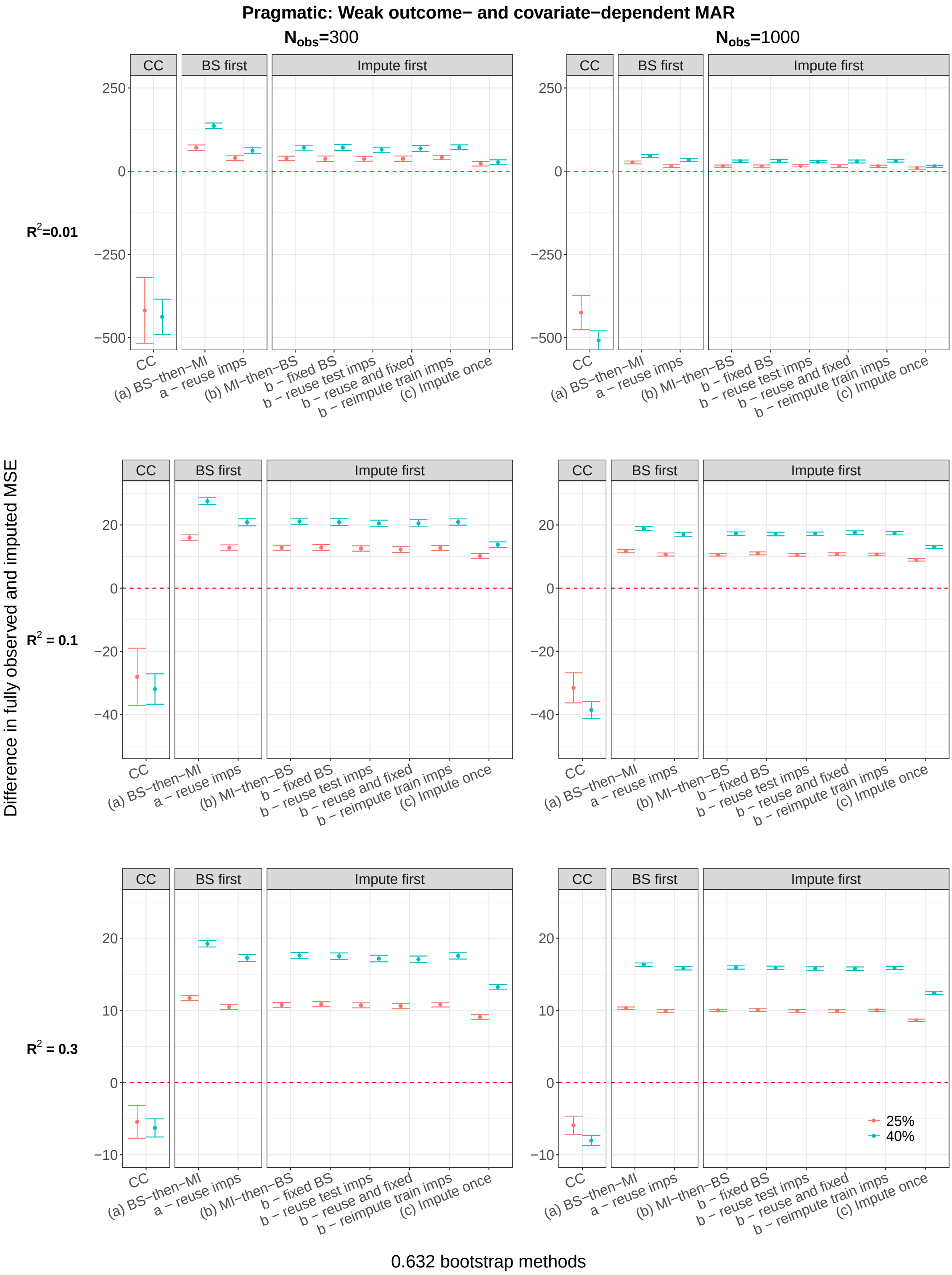

Figure S55: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

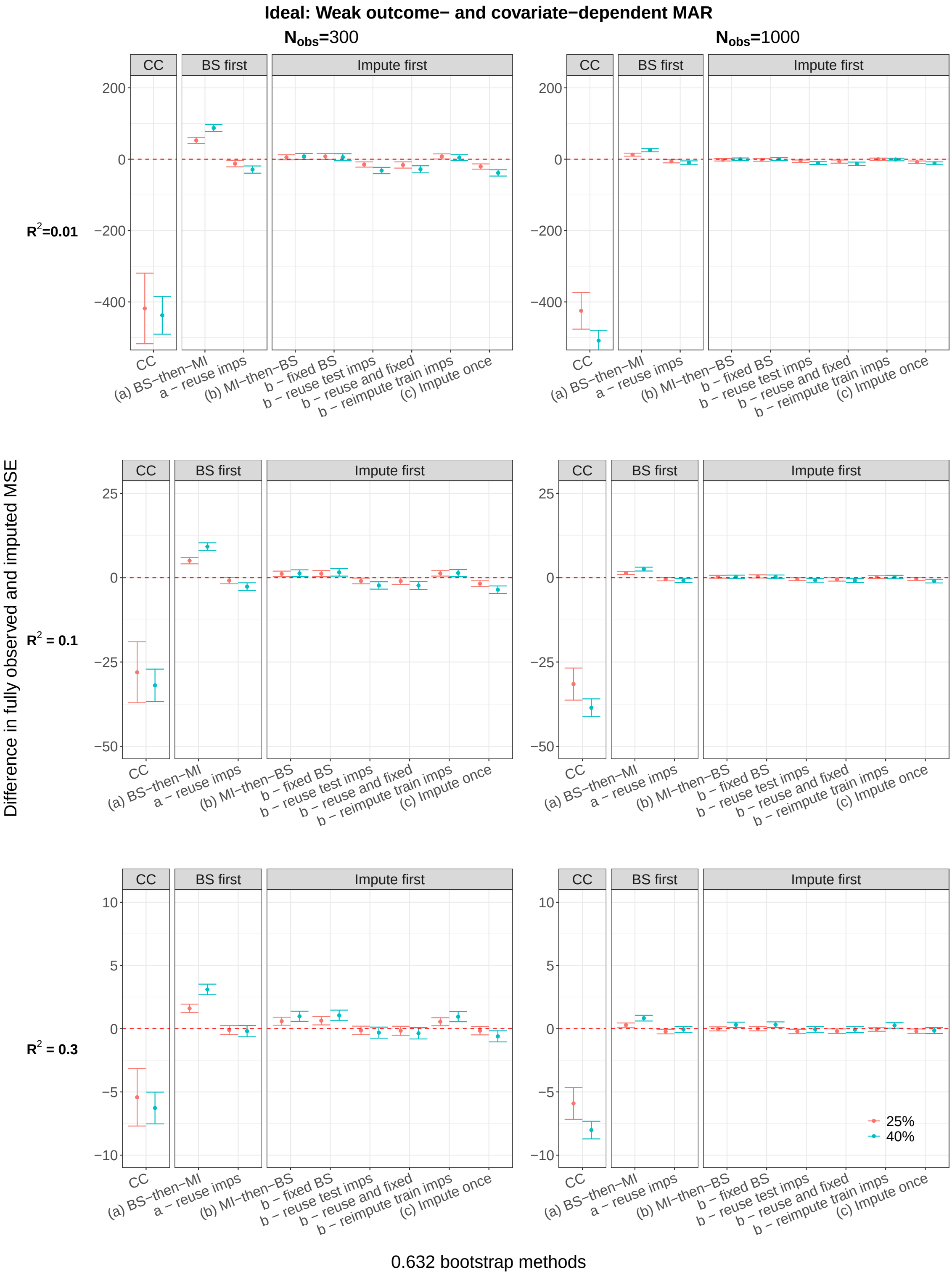

Figure S56: Comparing the impact of increasing the percentage of missingness on the difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak outcome- and covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. Red denotes MSE$_{imp}$ - MSE$_{obs}$ when 25% of X_1 values are missing and blue denotes MSE$_{imp}$ - MSE$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

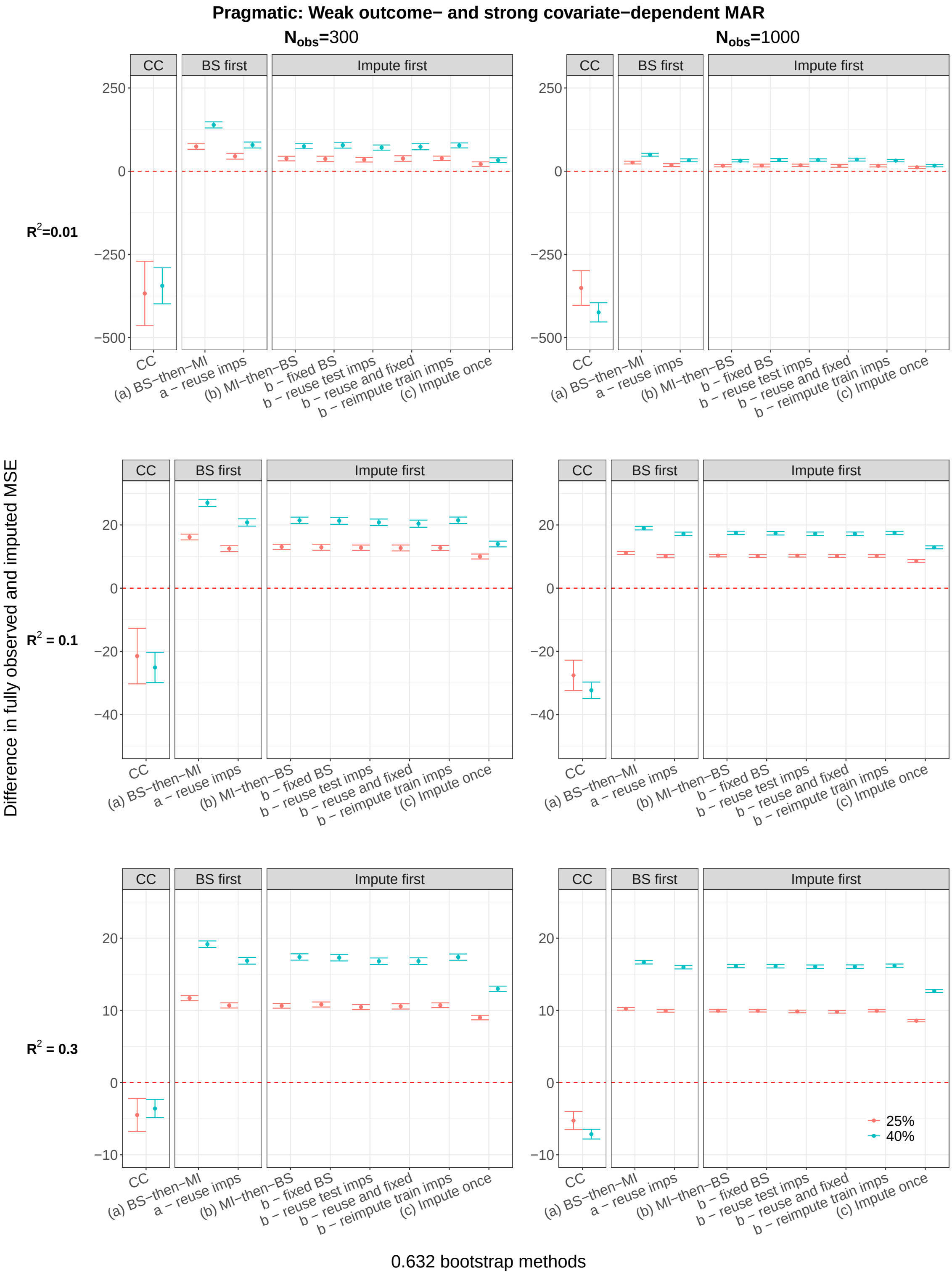

Figure S57: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

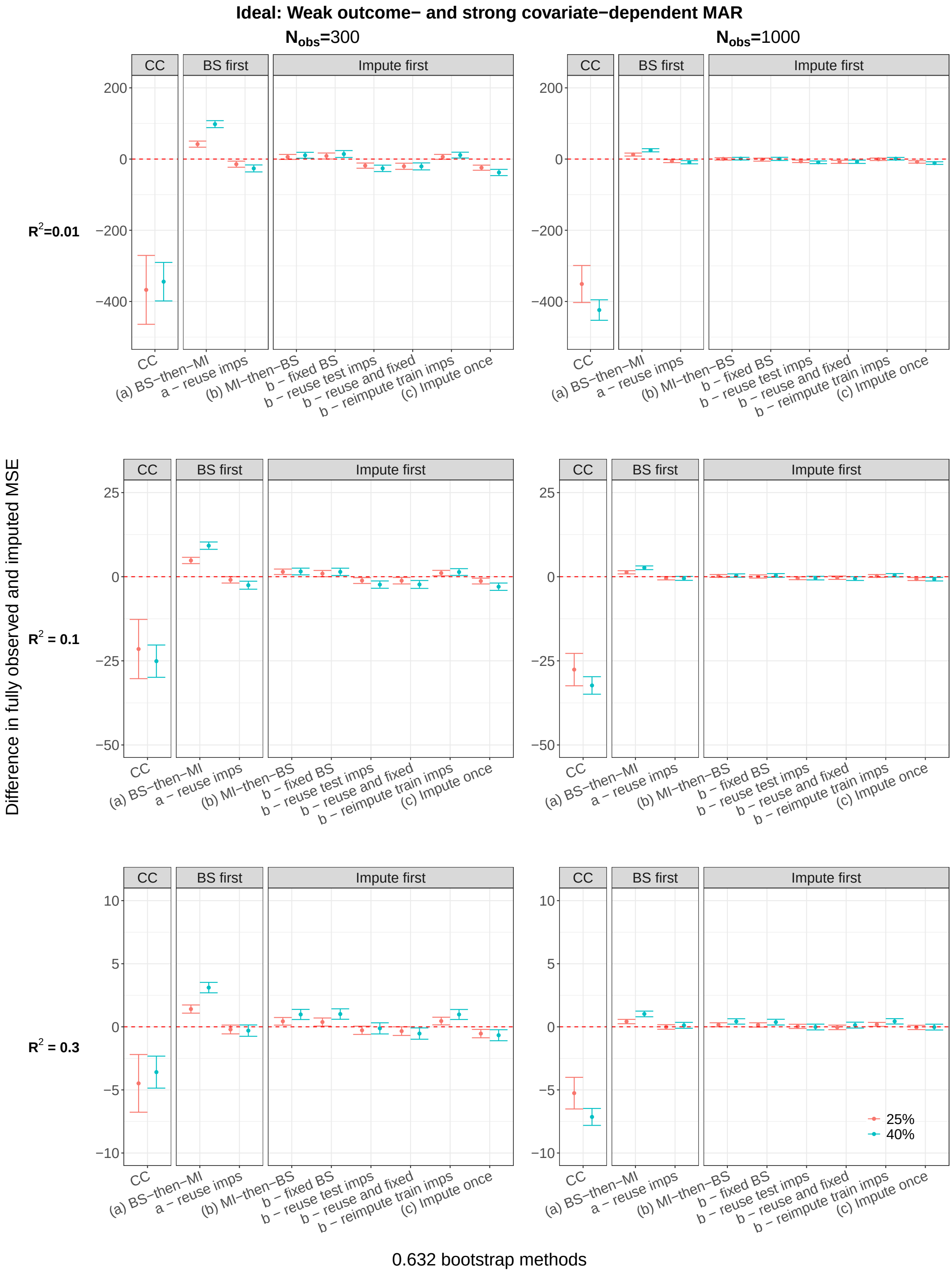

Figure S58: Comparing the impact of increasing the percentage of missingness on the difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR for the 0.632 bootstrap algorithm when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. Red denotes MSE_{imp} - MSE_{obs} when 25% of X_1 values are missing and blue denotes MSE_{imp} - MSE_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.3.3 Comparing M=5 versus M=25 (MSE_{imp}-MSE_{obs})

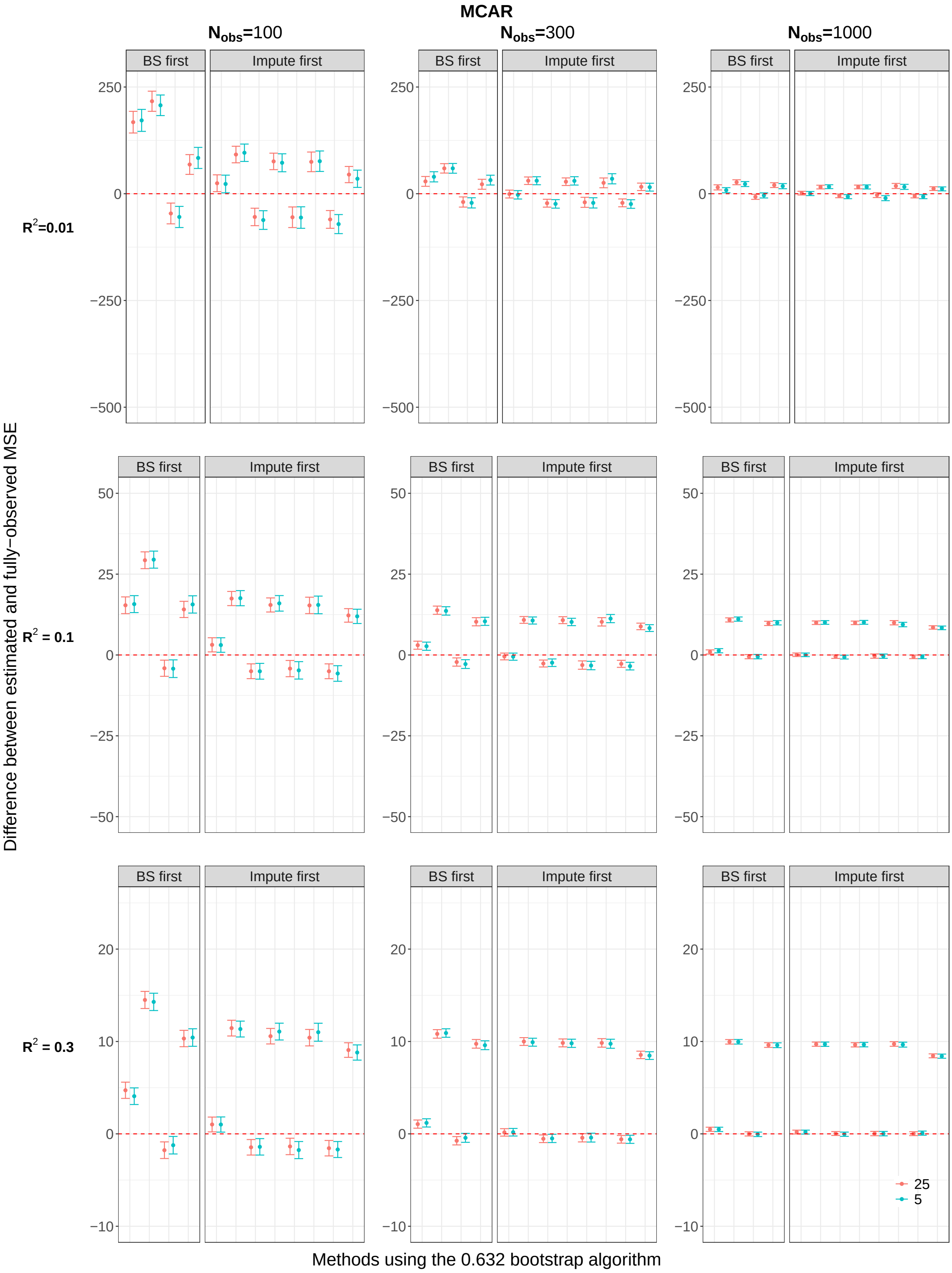

Figure S59: The difference MSE_{imp} - MSE_{obs} when data are MCAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

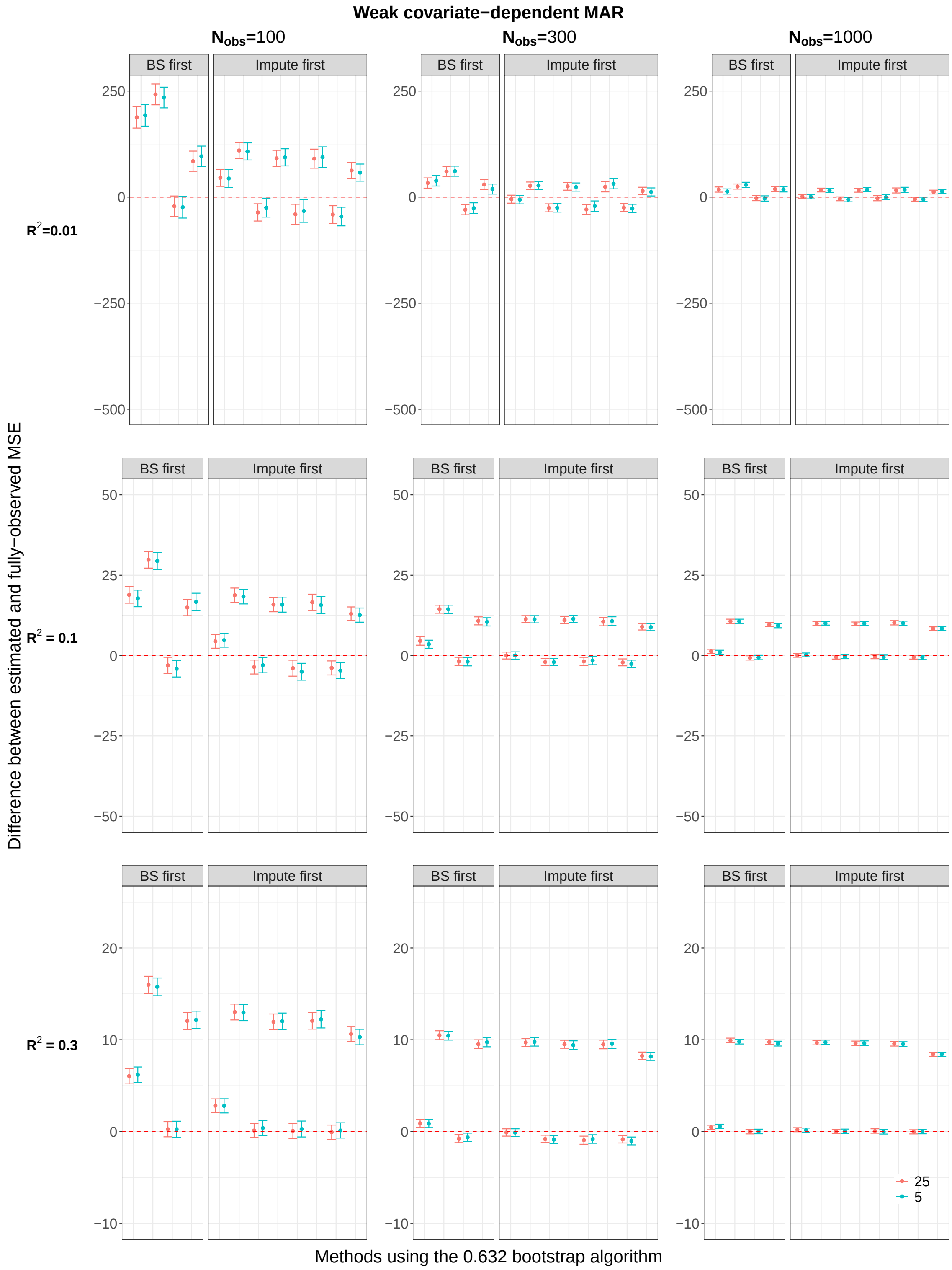

Figure S60: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are weak covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

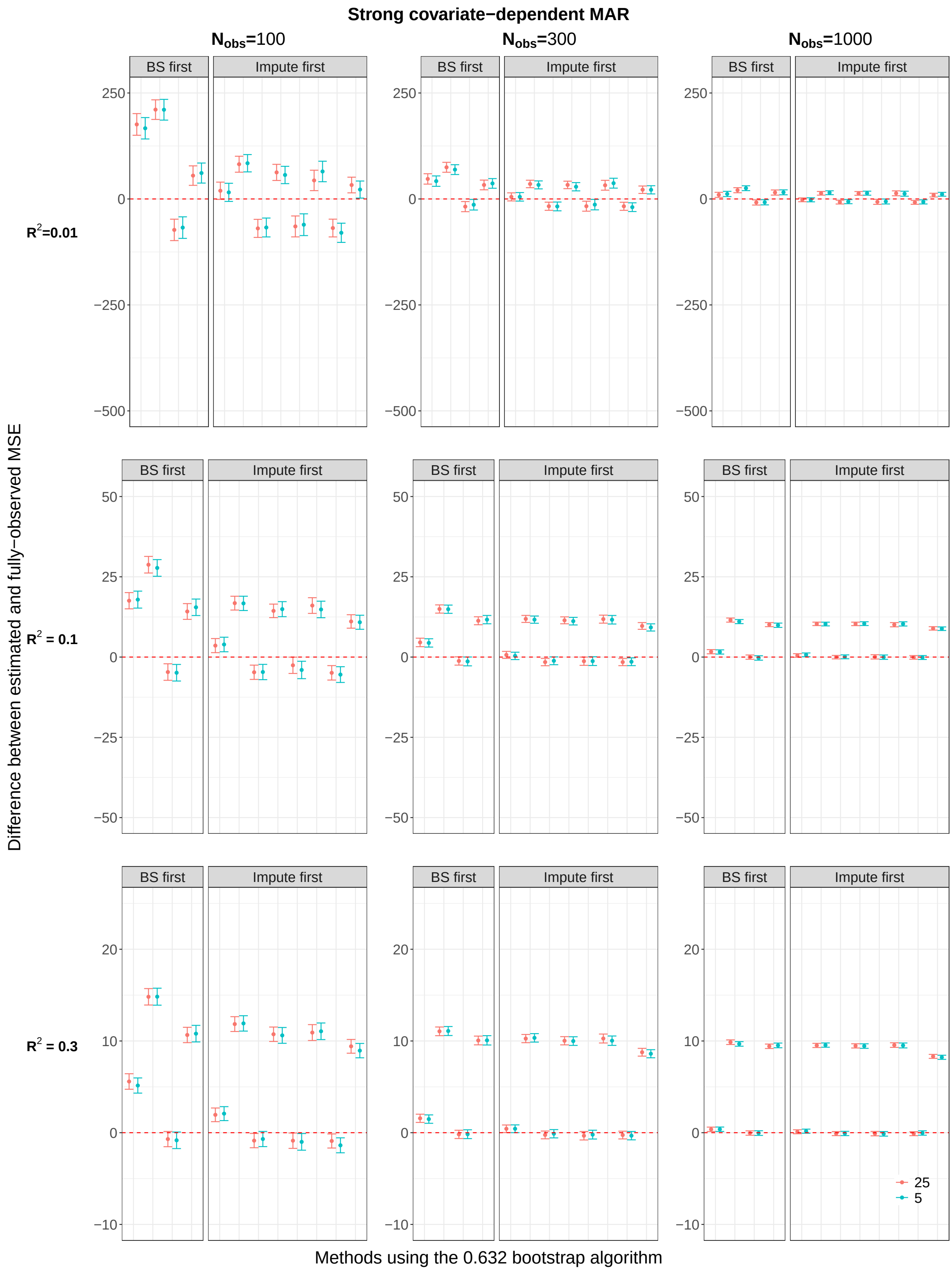

Figure S61: The difference MSE_{imp} - MSE_{obs} when data are strong covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

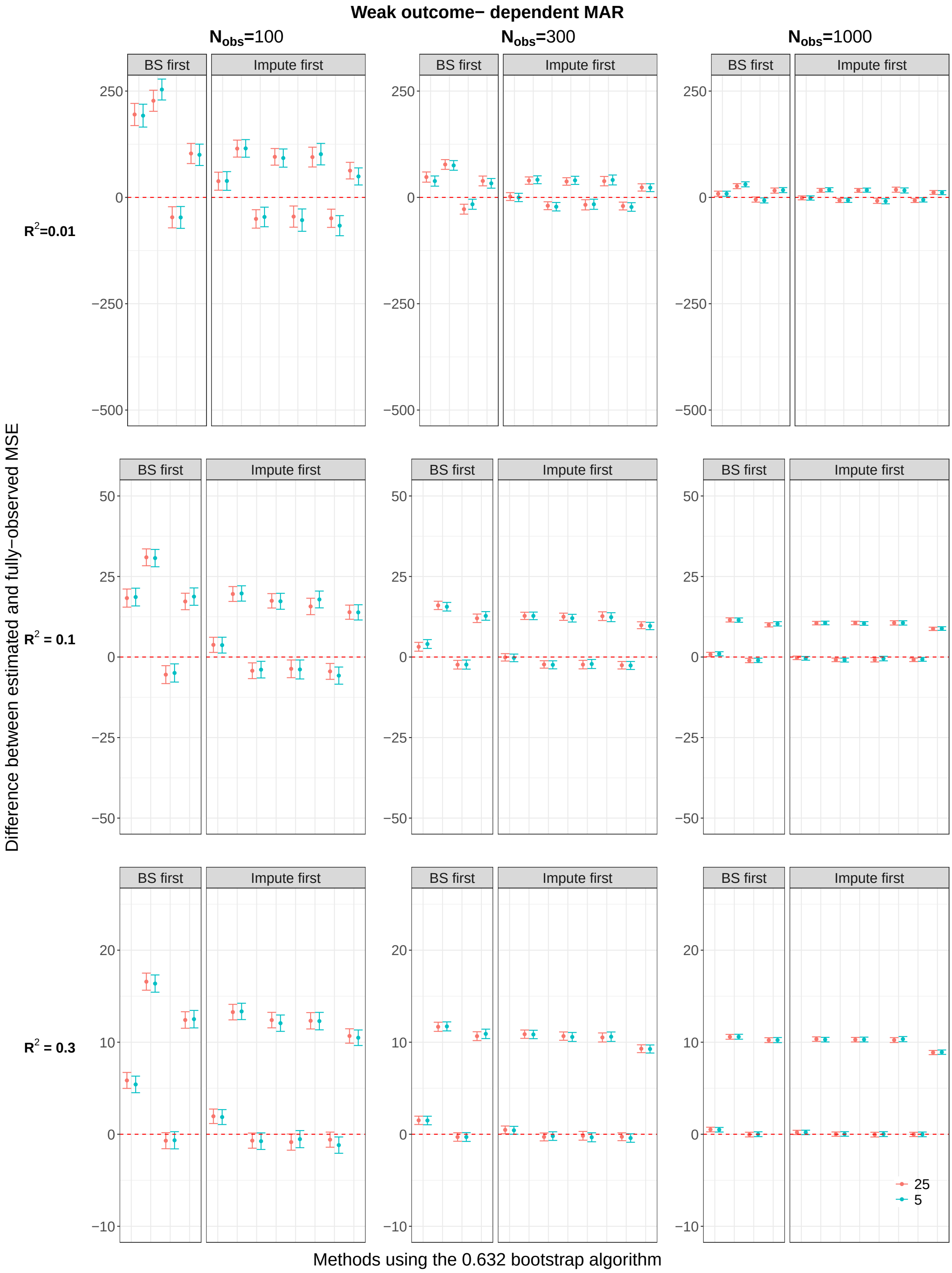

Figure S62: The difference MSE_{imp} - MSE_{obs} when data are weak outcome-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

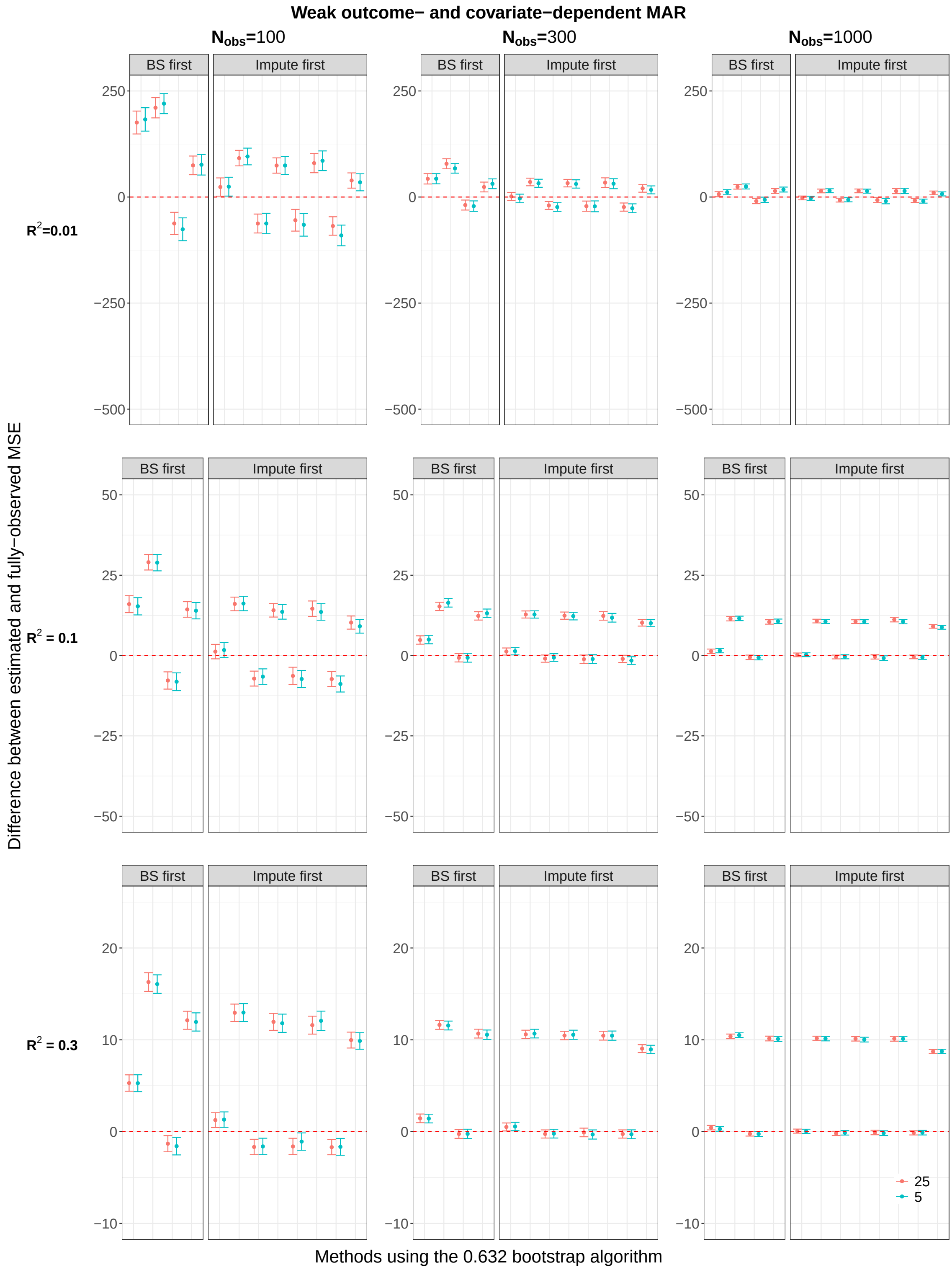

Figure S63: The difference MSE_{imp} - MSE_{obs} when data are weak outcome- and covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

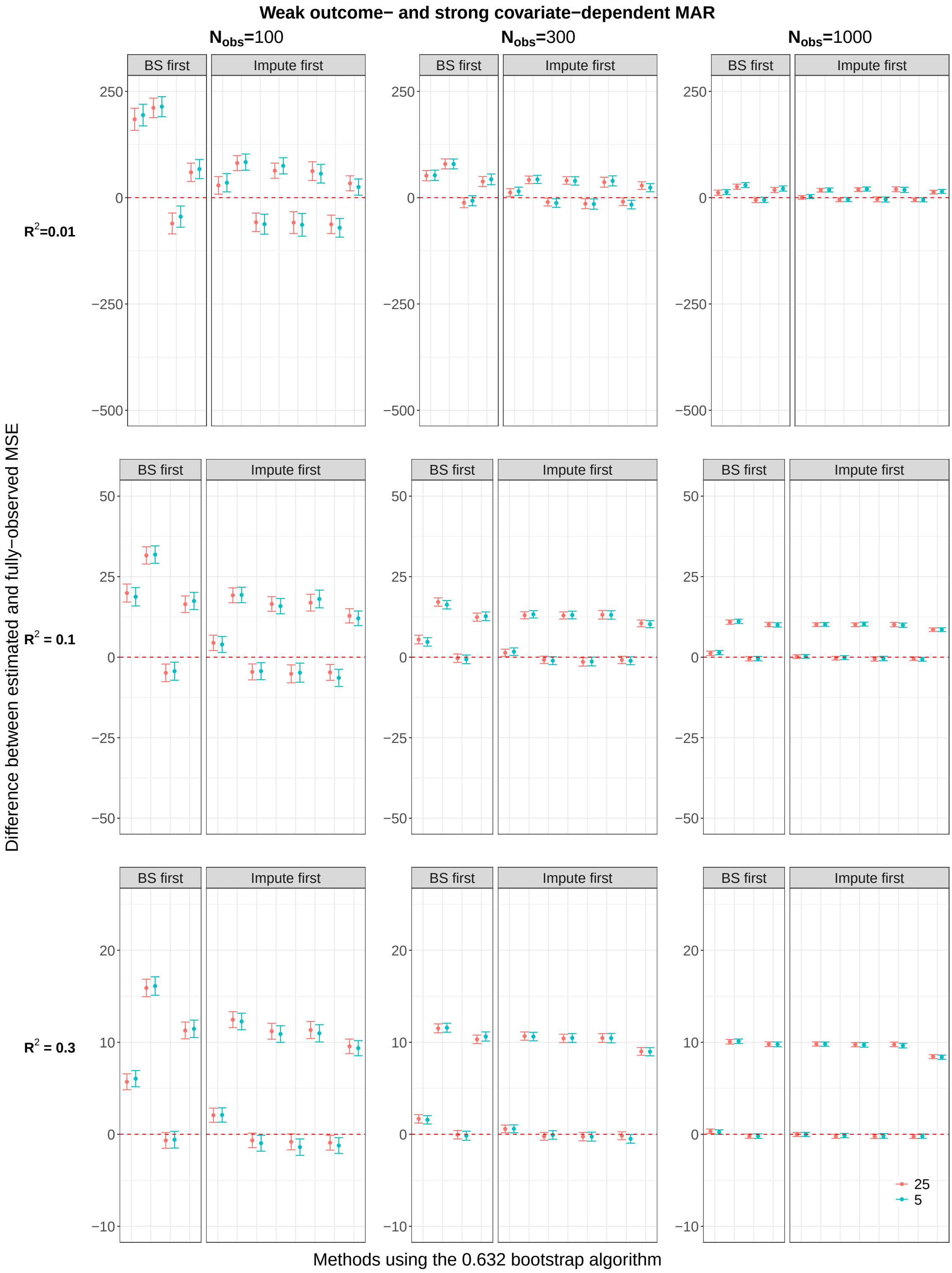

Figure S64: The difference MSE_{imp} - MSE_{obs} when data are weak outcome- and strong covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.3.4 MSE from imputation methods compared to the target MSE (MSE_{target}) using a larger validation set

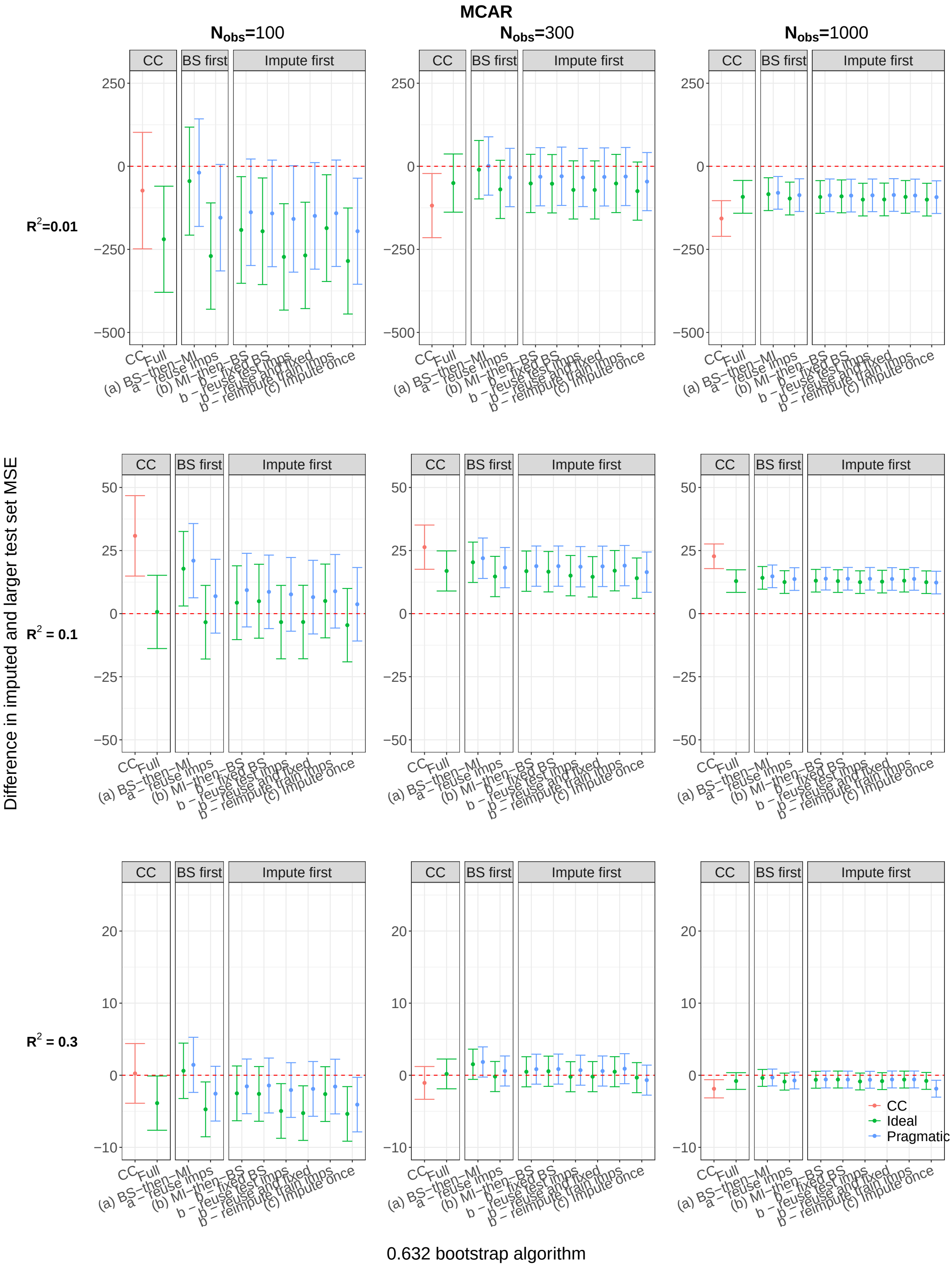

Figure S65: The difference MSE_{imp} - MSE_{target} when data are MCAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

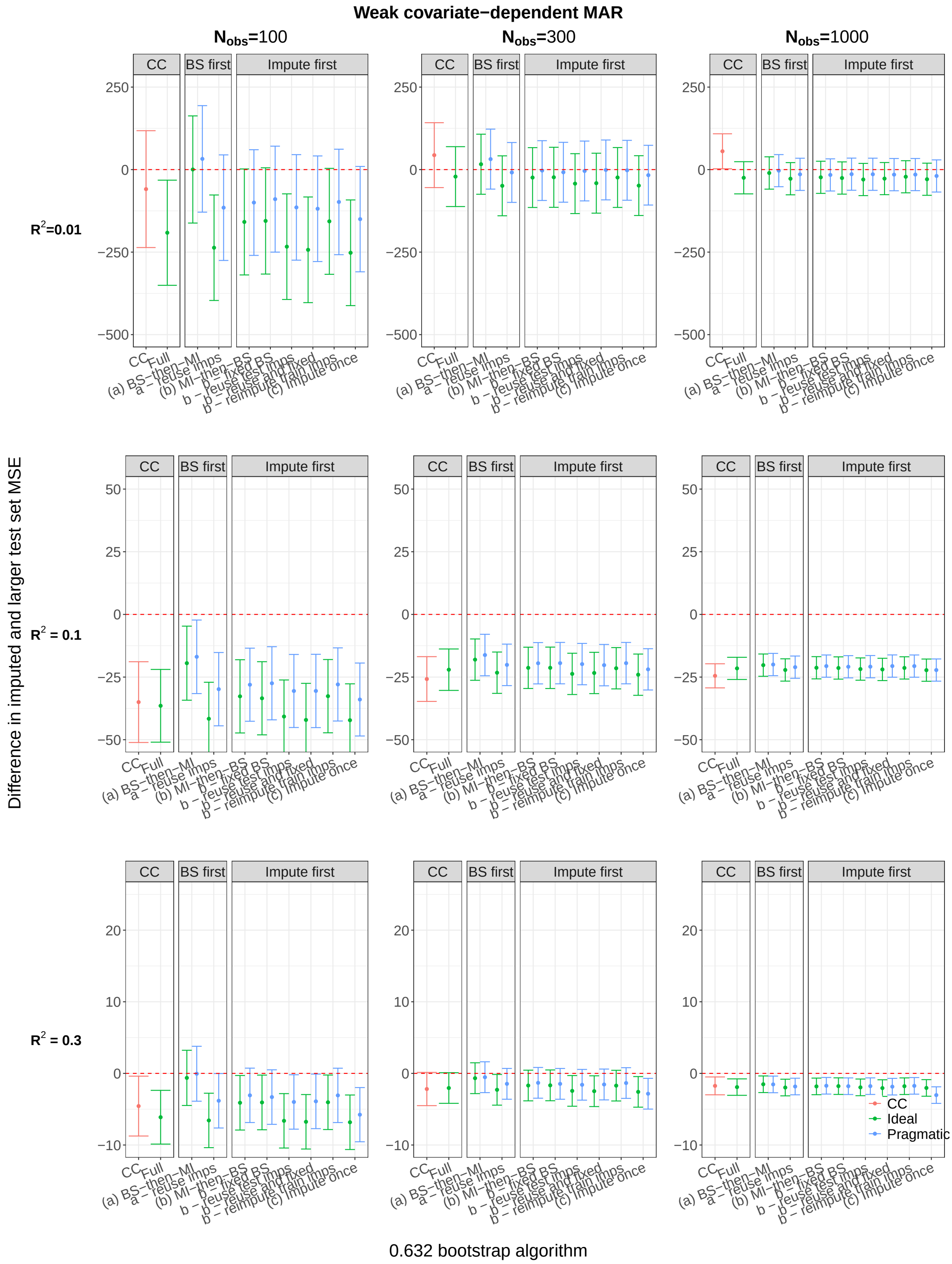

Figure S66: The difference MSE_{imp} - MSE_{target} when data are weak covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

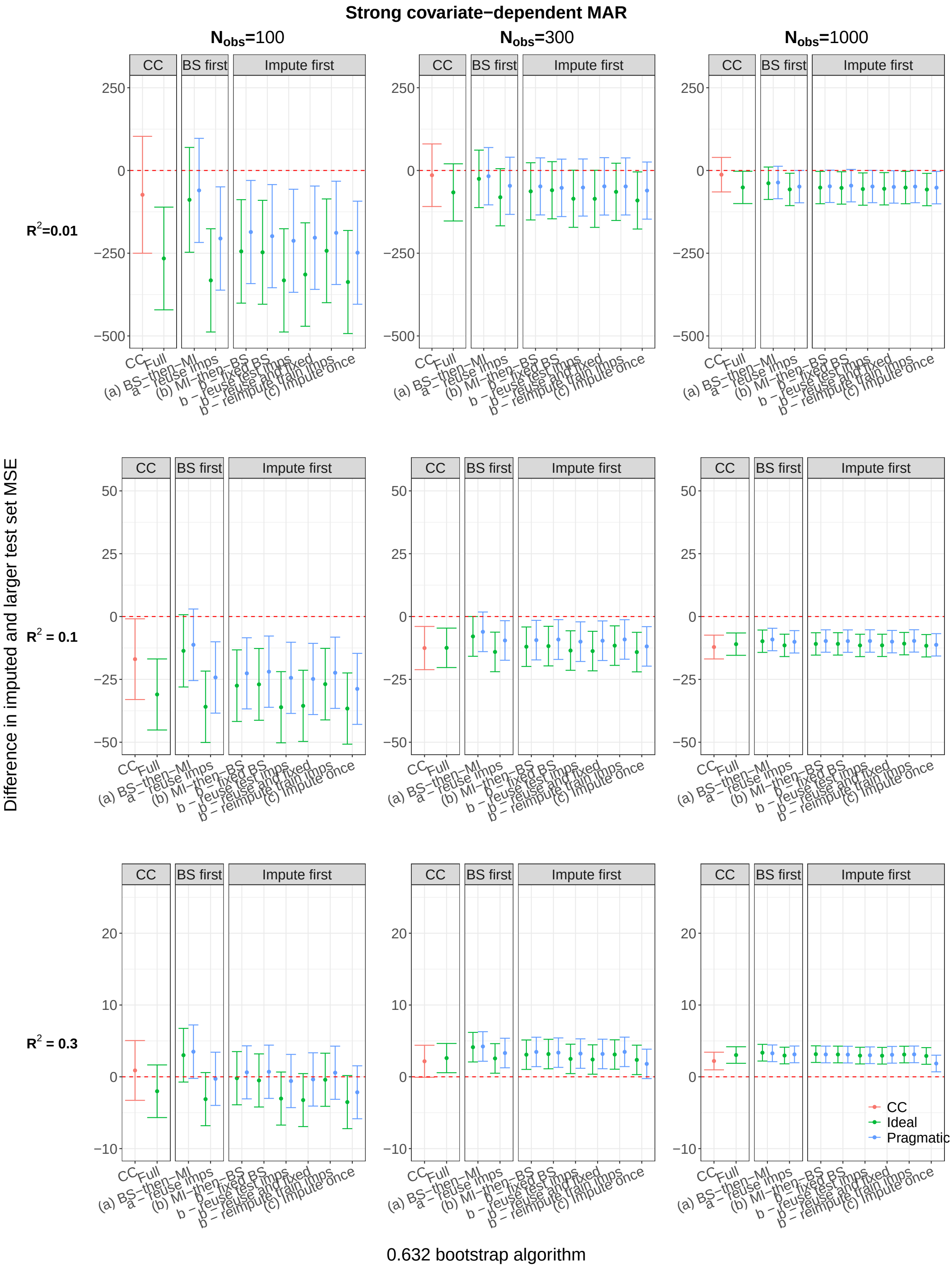

Figure S67: The difference MSE$_{imp}$ - MSE$_{target}$ when data are strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

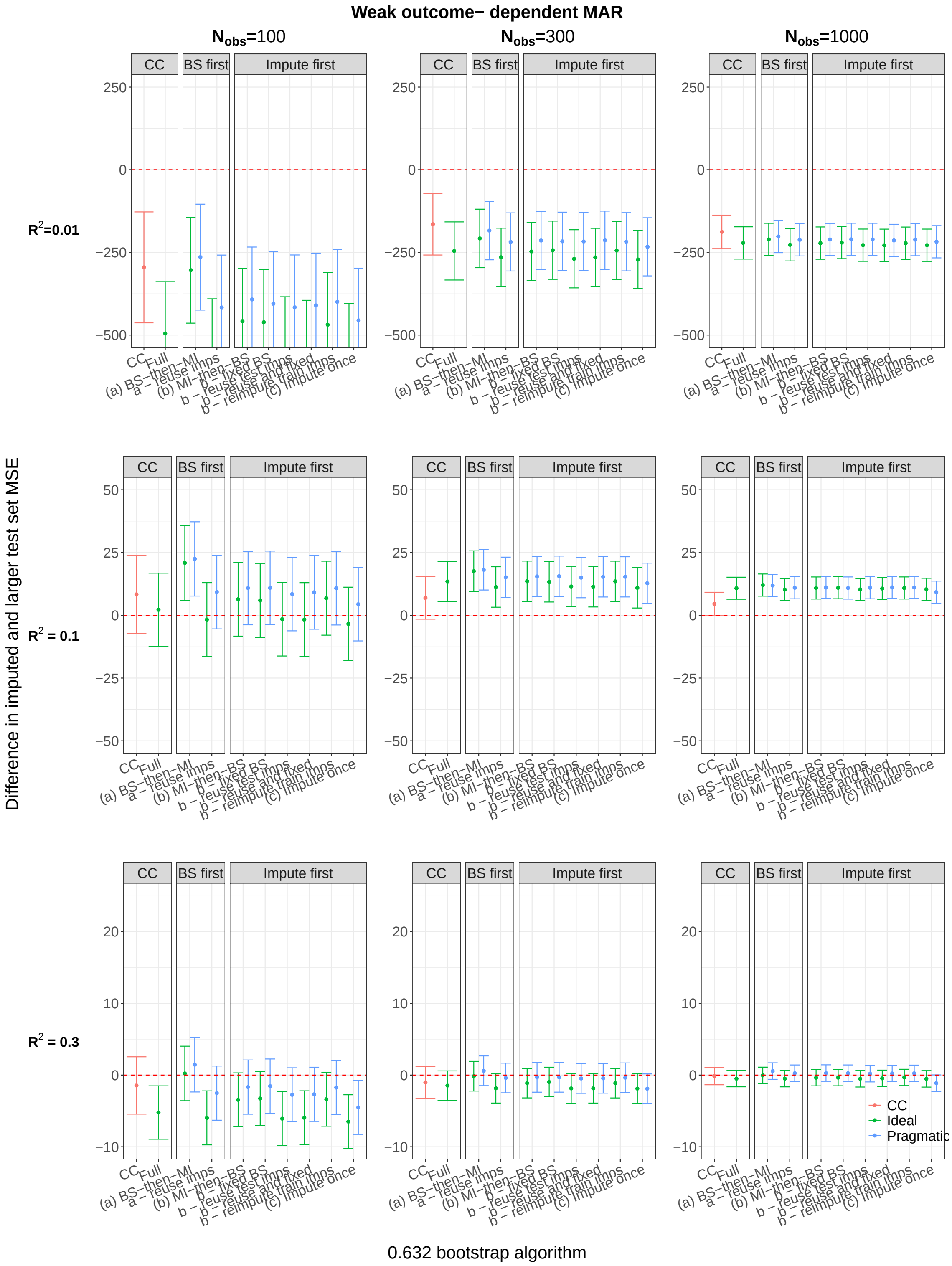

Figure S68: The difference MSE$_{imp}$ - MSE$_{target}$ when data are weak outcome-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

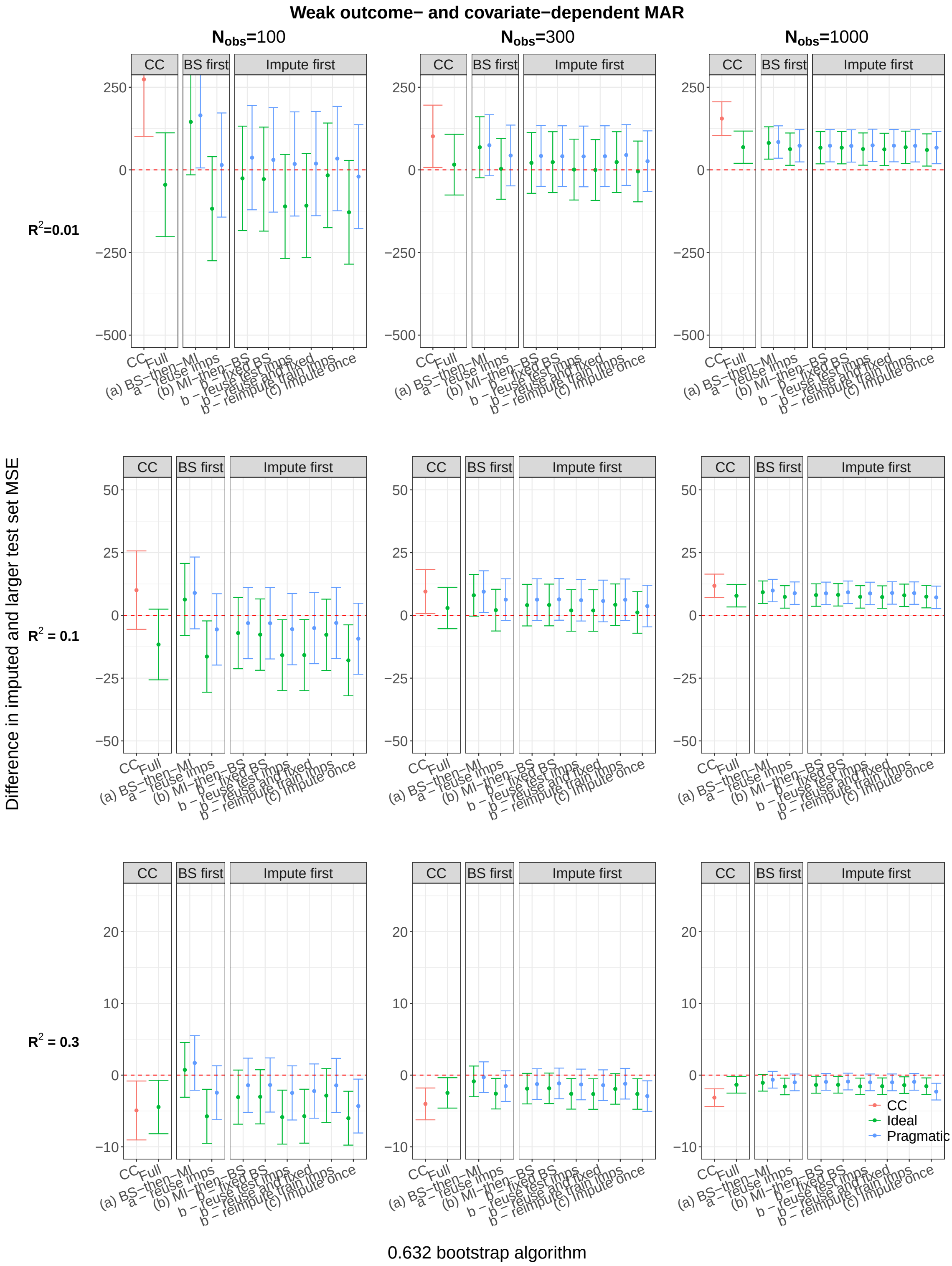

Figure S69: The difference MSE$_{imp}$ - MSE$_{target}$ when data are weak outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

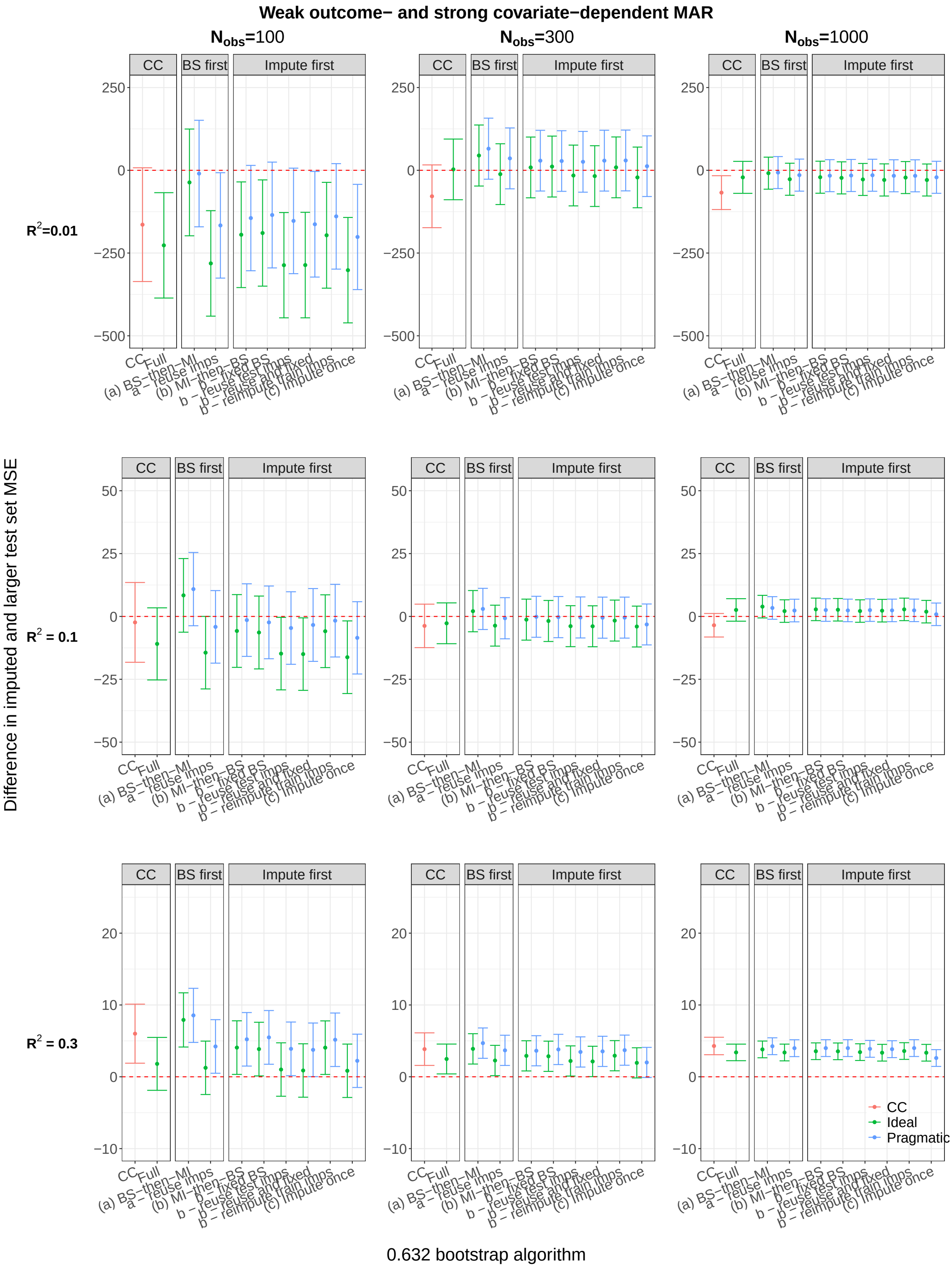

Figure S70: The difference MSE$_{imp}$ - MSE$_{target}$ when data are weak outcome- and strong covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S3.4 Comparing internal validation methods

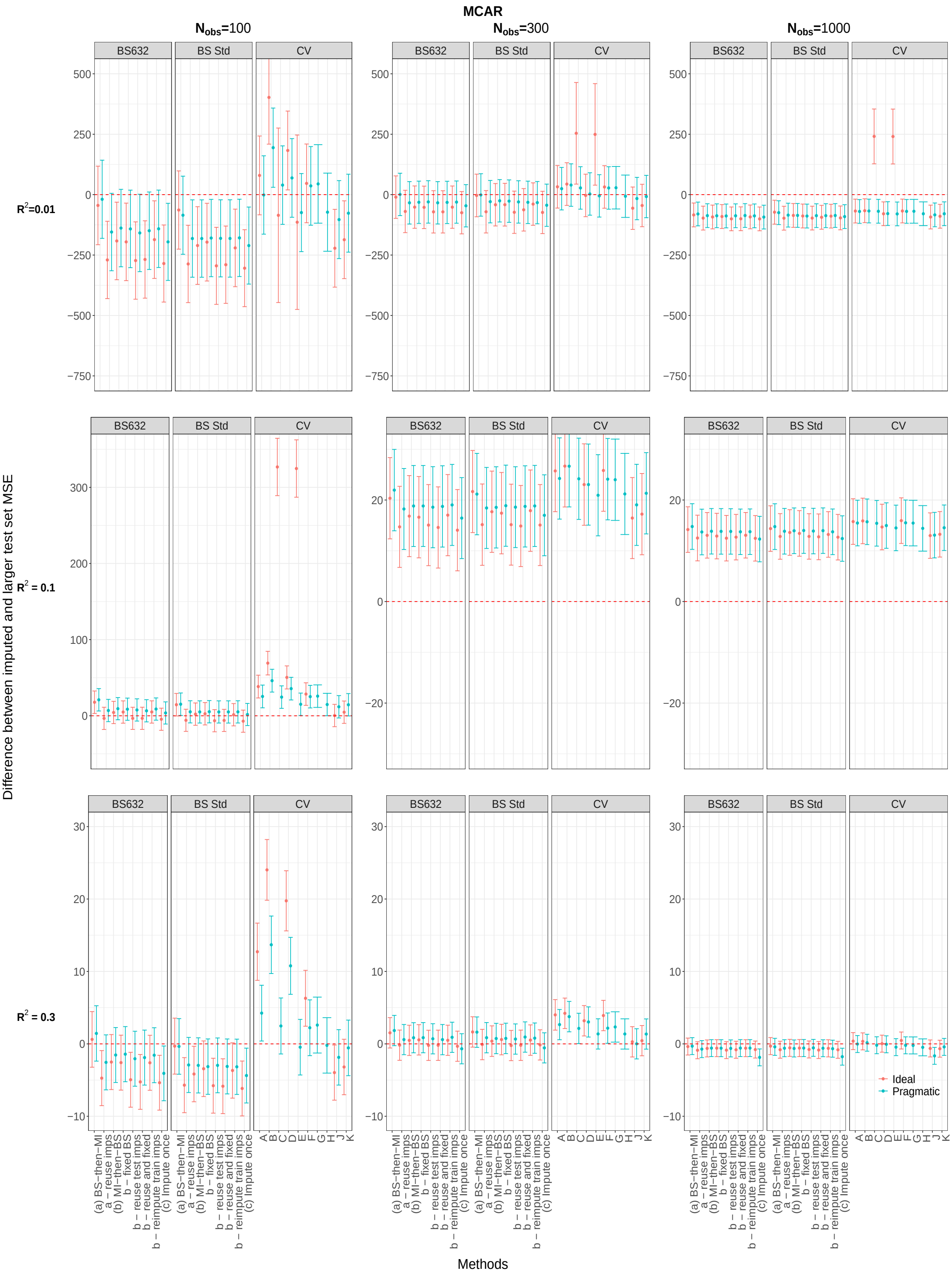

Figure S71: Comparing cross-validation and the 0.632 and *standard* Bootstrap using the target MSE. Error bars of the difference in the imputed MSE and the MSE estimate from a larger validation set are presented for the MCAR scenario. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

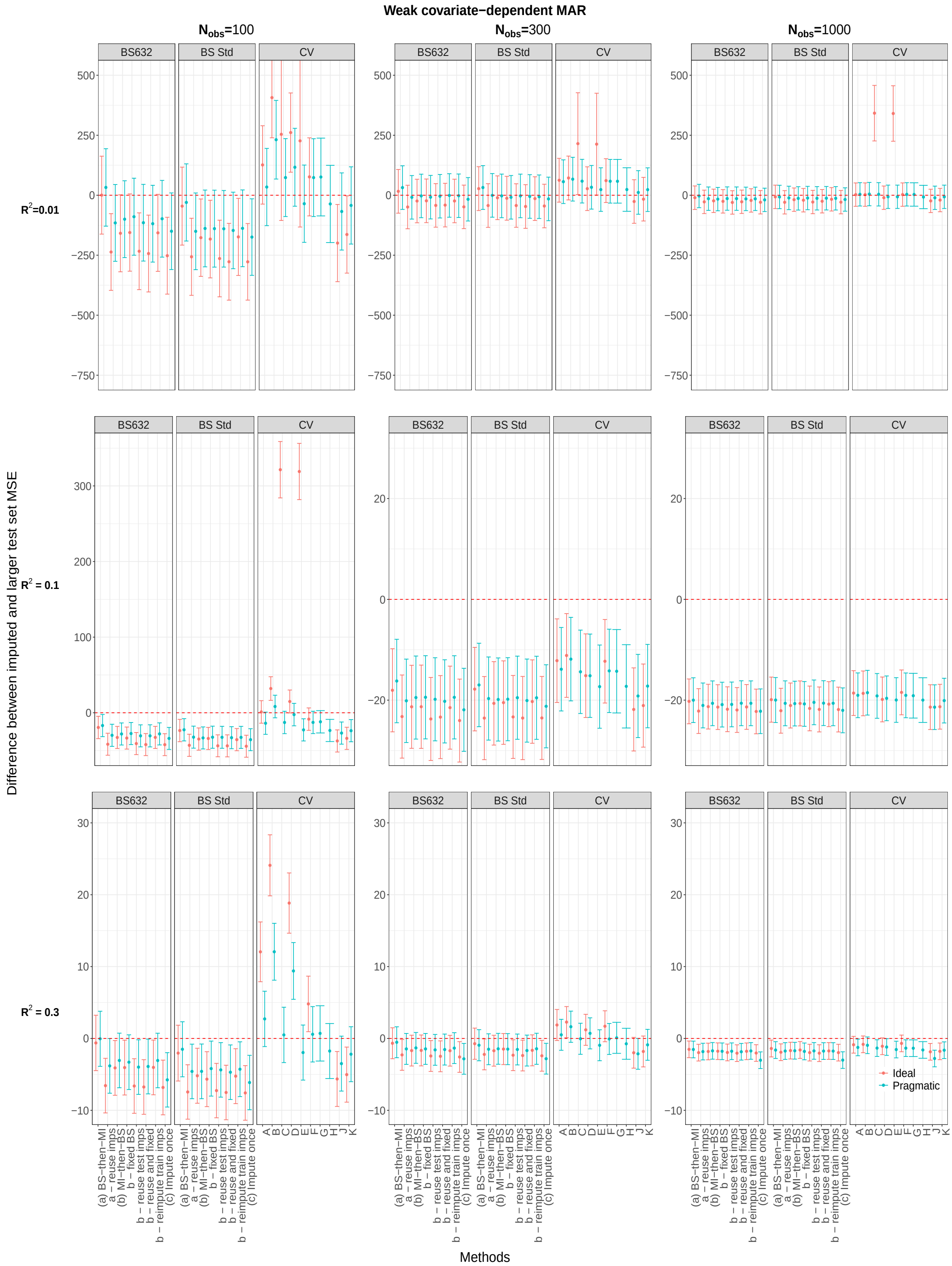

Figure S72: Comparing cross-validation and the 0.632 and *standard* Bootstrap using the target MSE. Error bars of the difference in the imputed MSE and the MSE estimate from a larger validation set are presented for the weak covariate-dependent MAR scenario. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

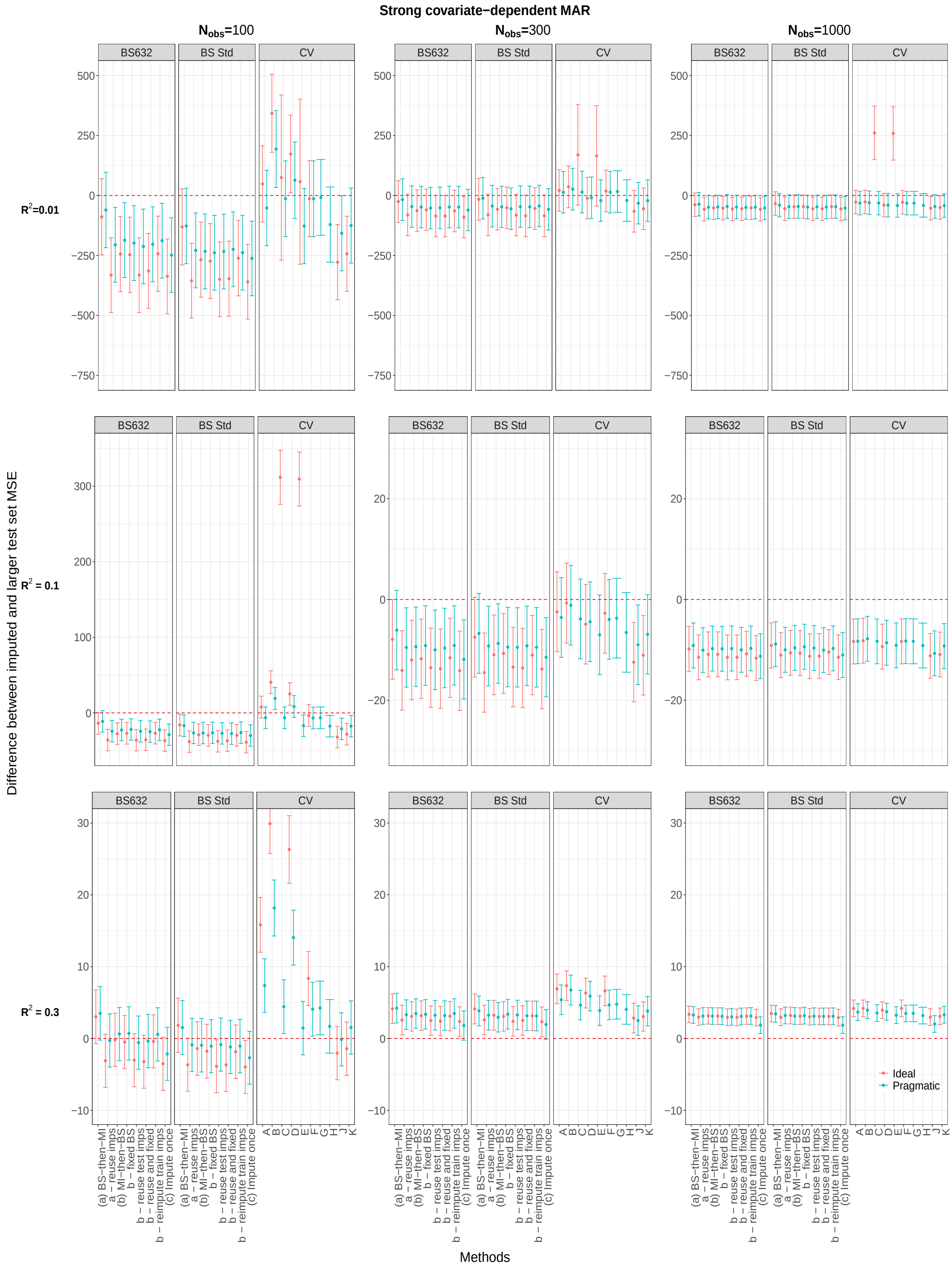

Figure S73: Comparing cross-validation and the 0.632 and *standard* Bootstrap using the target MSE. Error bars of the difference in the imputed MSE and the MSE estimate from a larger validation set are presented for the strong covariate-dependent MAR scenario. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

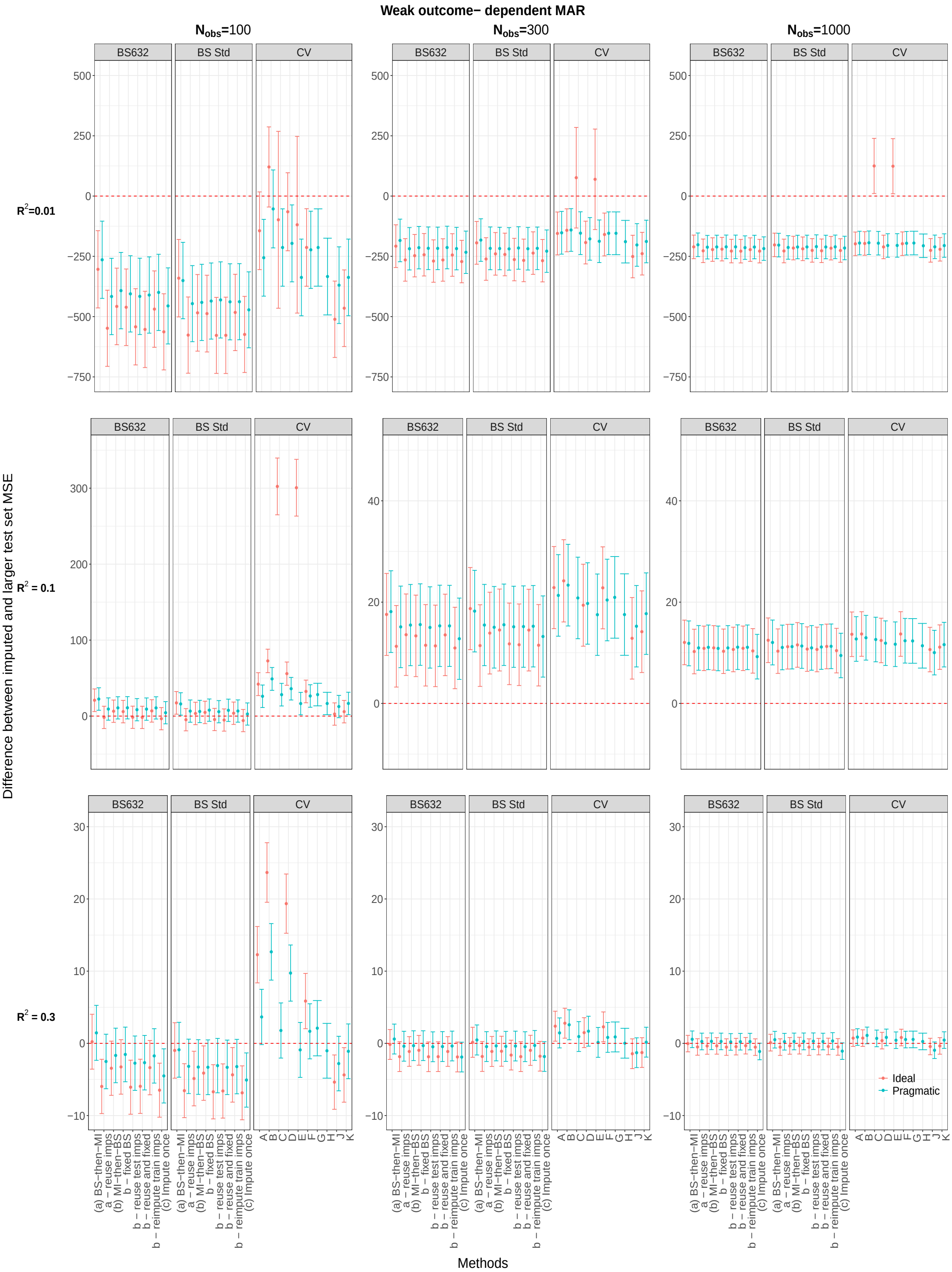

Figure S74: Comparing cross-validation and the 0.632 and *standard* Bootstrap using the target MSE. Error bars of the difference in the imputed MSE and the MSE estimate from a larger validation set are presented for the weak outcome-dependent MAR scenario. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

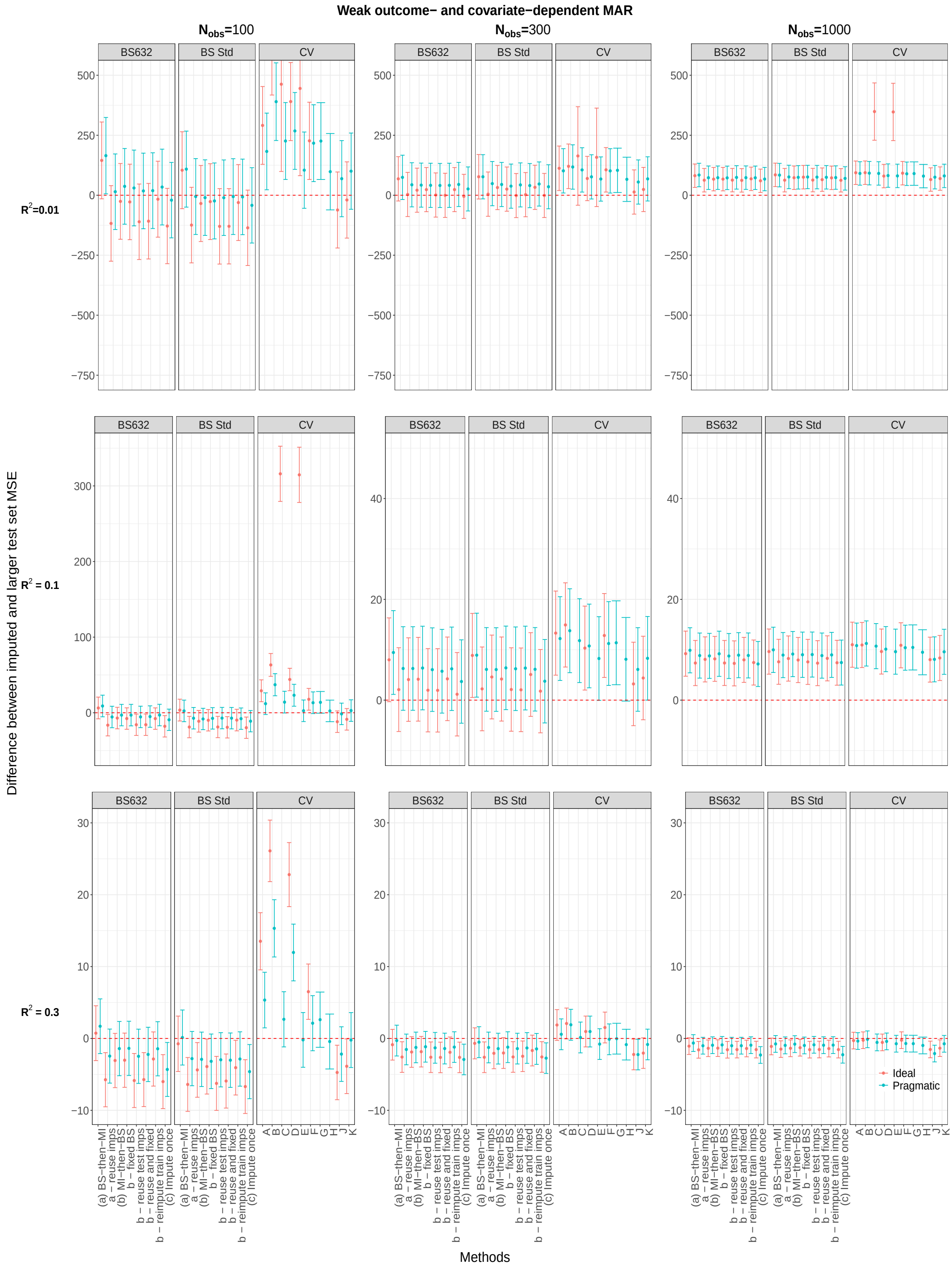

Figure S75: Comparing cross-validation and the 0.632 and *standard* Bootstrap using the target MSE. Error bars of the difference in the imputed MSE and the MSE estimate from a larger validation set are presented for the weak outcome- and covariate-dependent MAR scenario. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

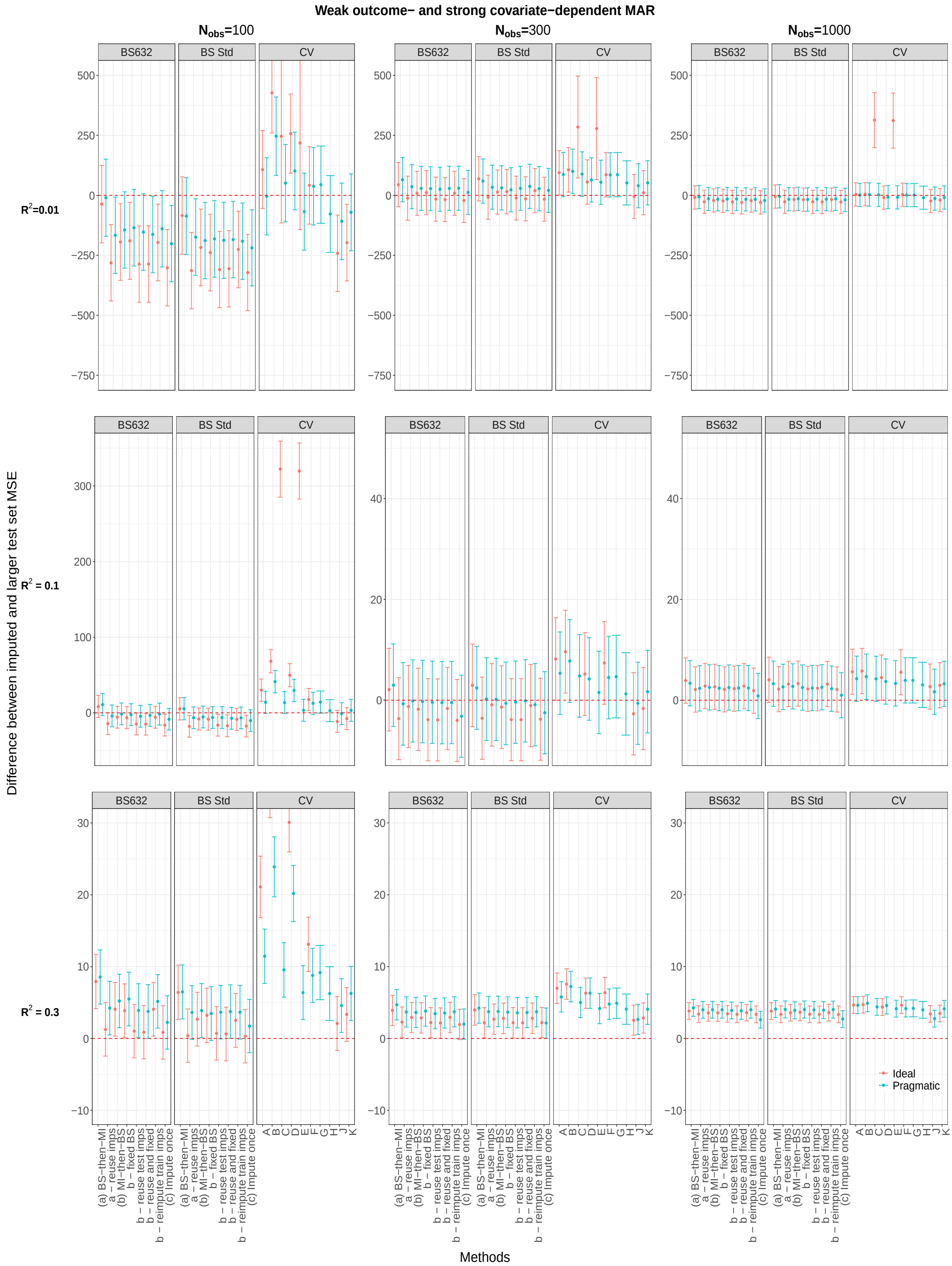

Figure S76: Comparing cross-validation and the 0.632 and *standard* Bootstrap using the target MSE. Error bars of the difference in the imputed MSE and the MSE estimate from a larger validation set are presented for the weak outcome- and strong covariate-dependent MAR scenario. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

S4 Chapter C: Bootstrap and MI (binary outcome)

S4.1 The *standard* bootstrap: AUC

S4.1.1 Reusing versus re-imputing for test performance of the *standard* algorithm

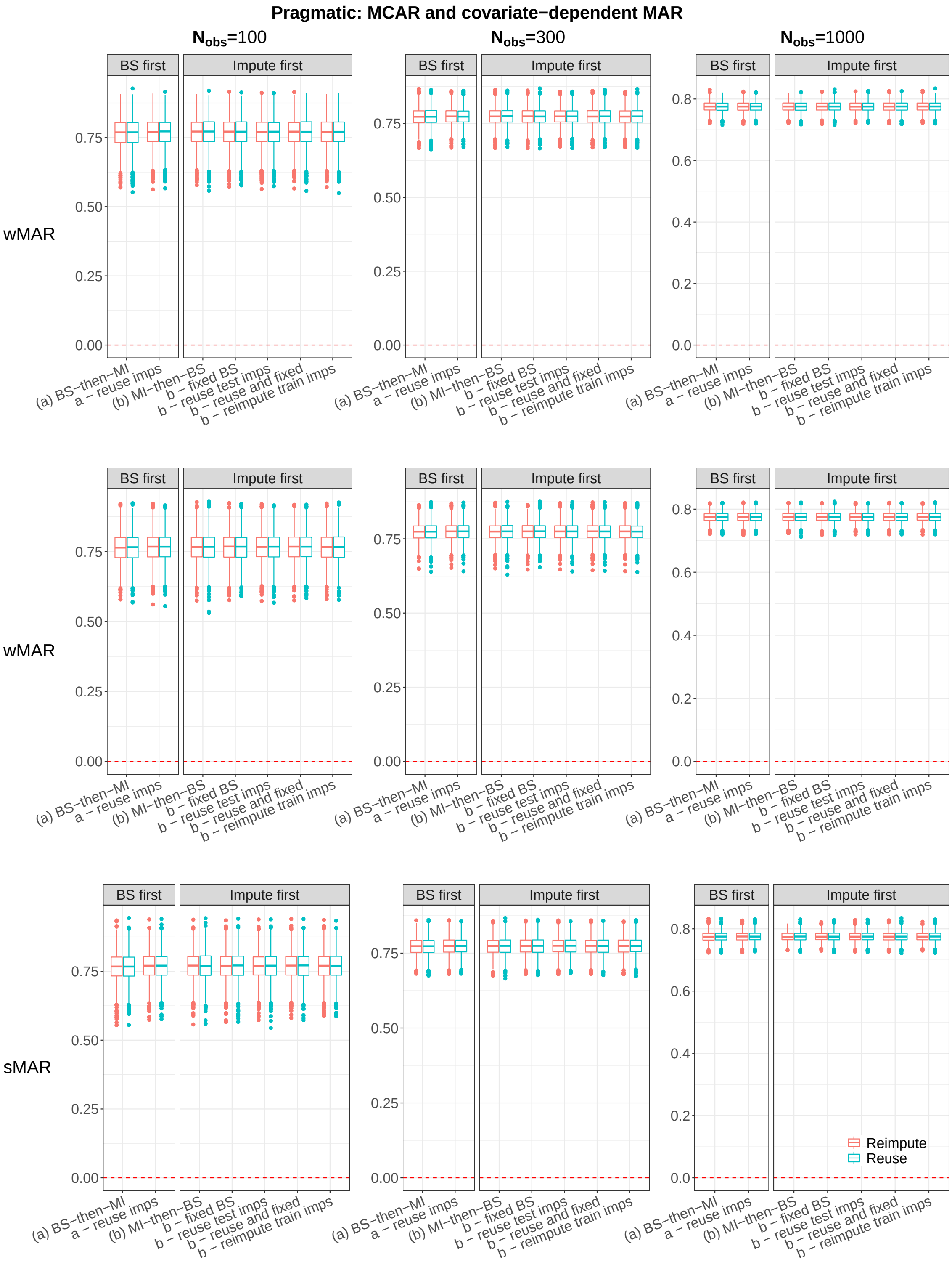

Figure S1: A comparison of reusing versus re-imputing test datasets on the AUC estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the AUC estimated when data are fully-observed (AUC_{imp}- AUC_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

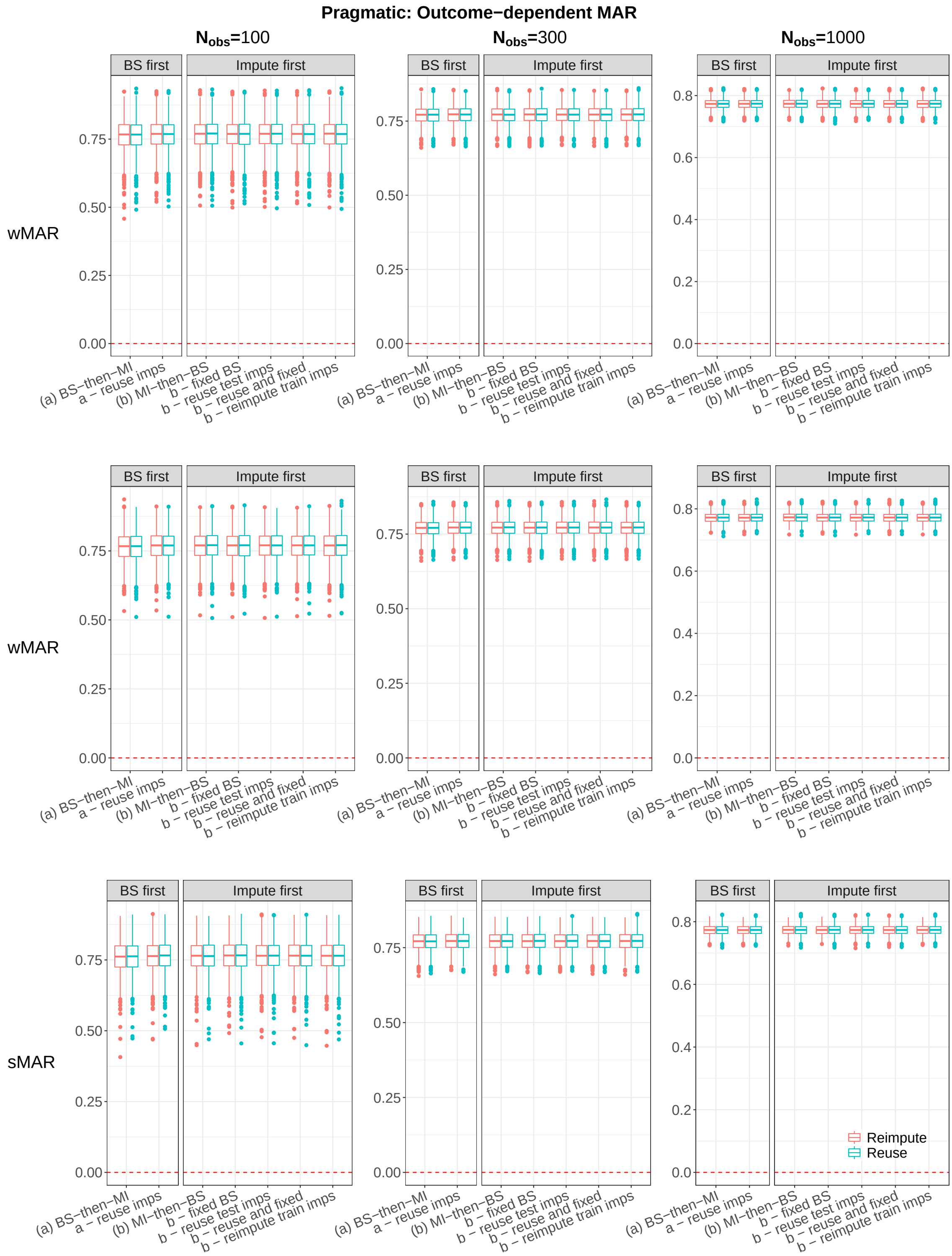

Figure S2: A comparison of reusing versus re-imputing test datasets on the AUC estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the AUC estimated when data are fully-observed (AUC_{imp}- AUC_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

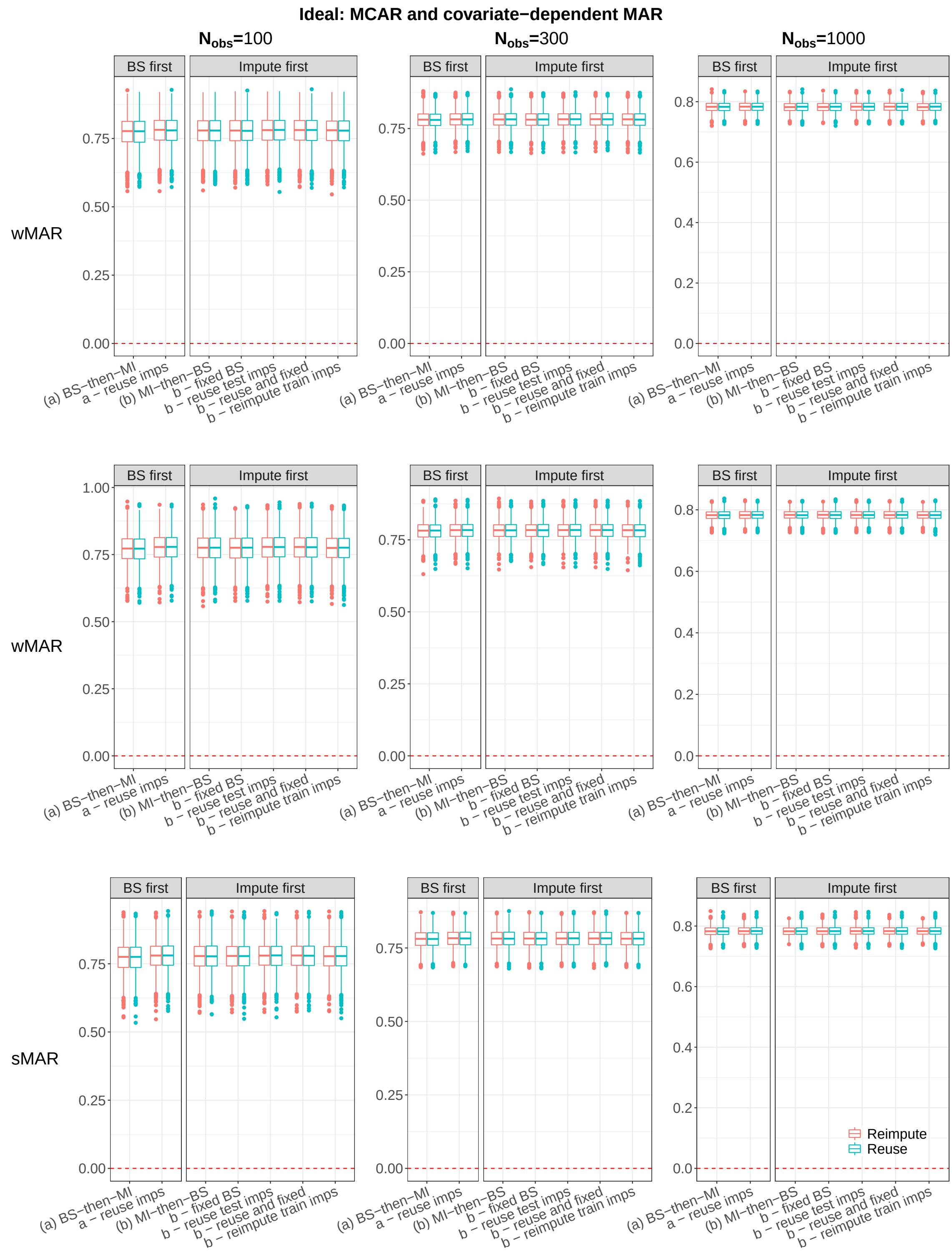

Figure S3: A comparison of reusing versus re-imputing test datasets on the AUC estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the AUC estimated when data are fully-observed (AUC_{imp}- AUC_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

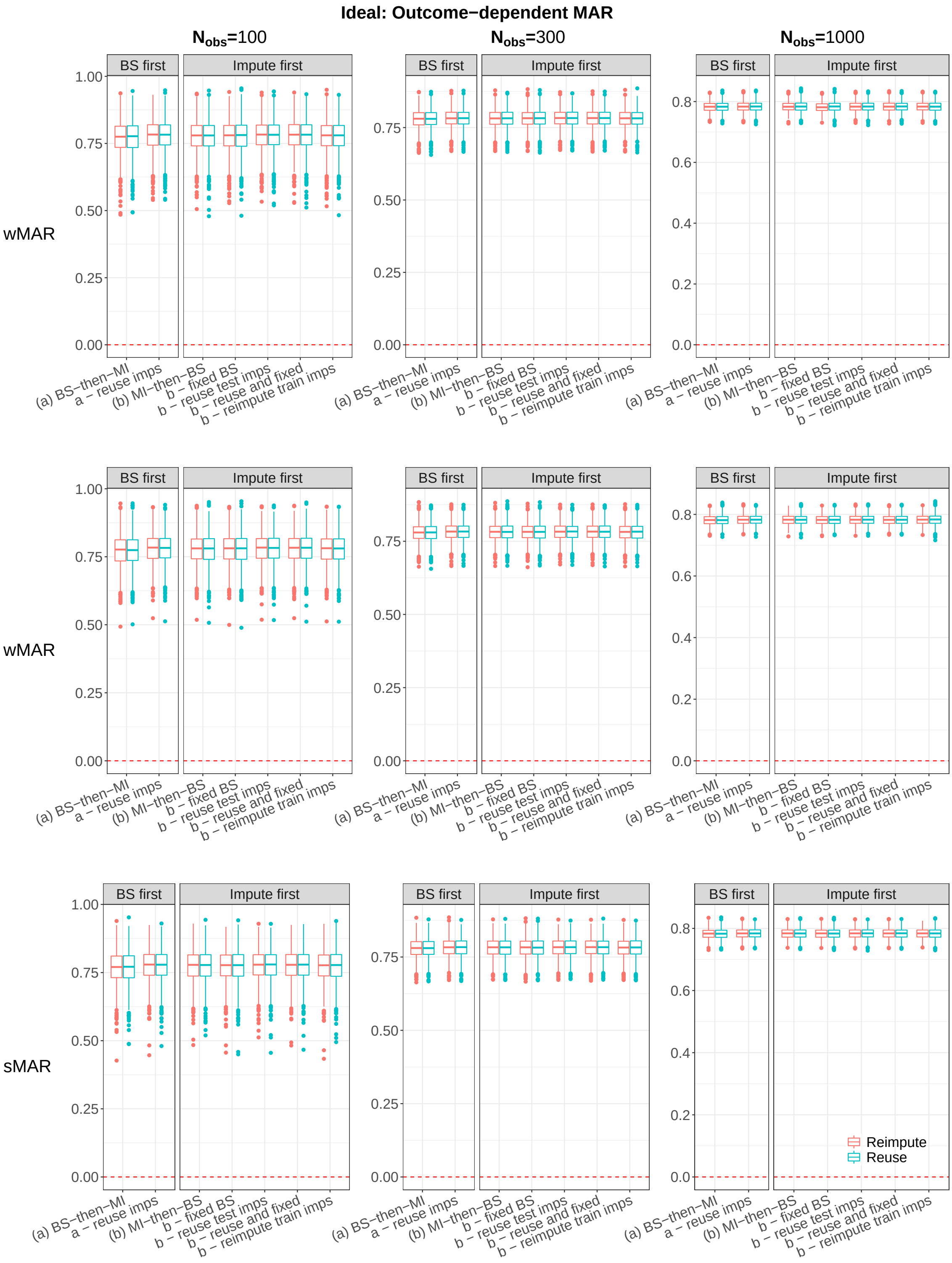

Figure S4: A comparison of reusing versus re-imputing test datasets on the AUC estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the AUC estimated when data are fully-observed (AUC_{imp}- AUC_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.1.2 AUC from imputation methods compared to the fully-observed AUC (AUC_{imp}-AUC_{obs})

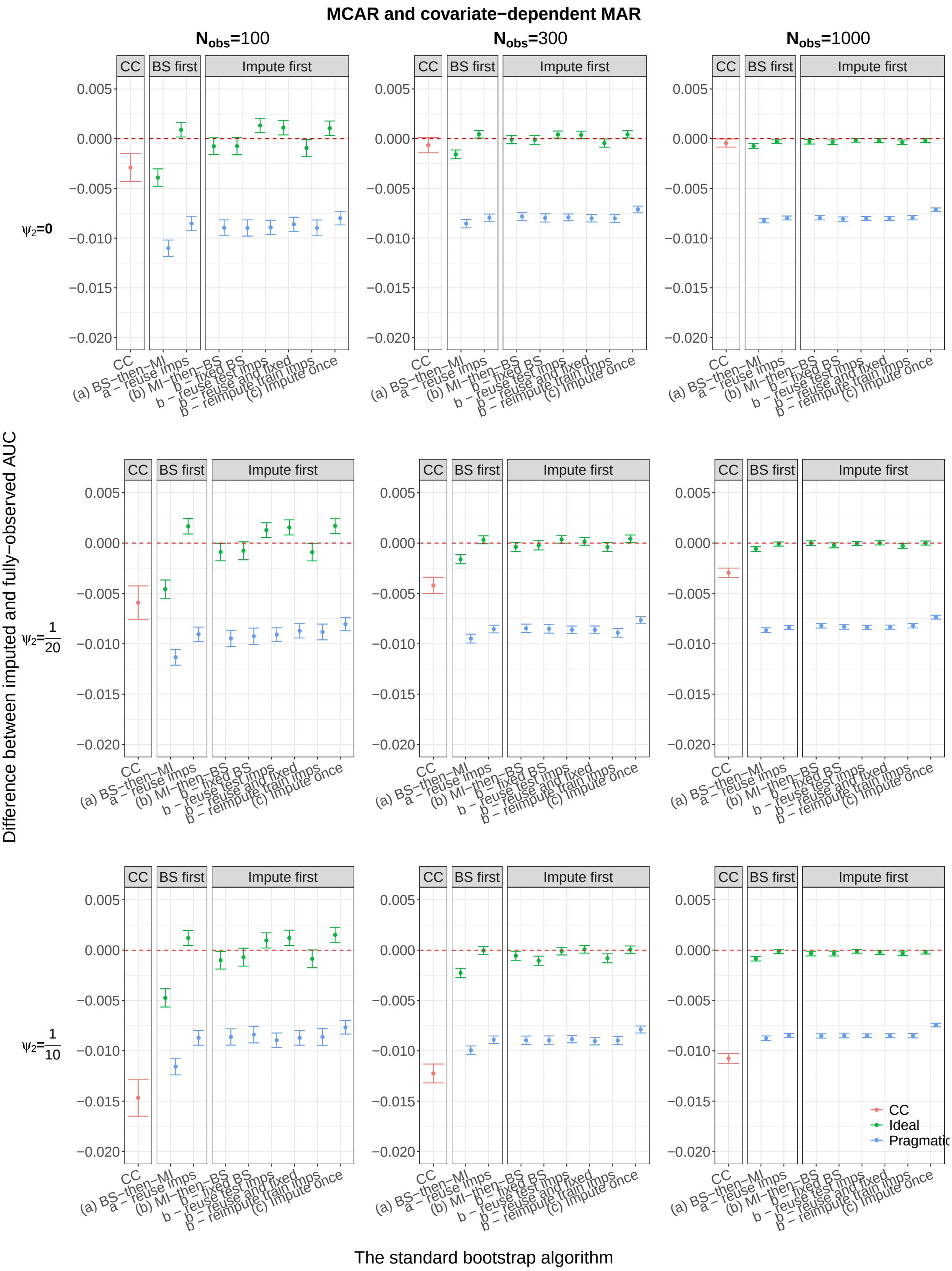

Figure S5: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are MCAR or covariate-dependent MAR. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

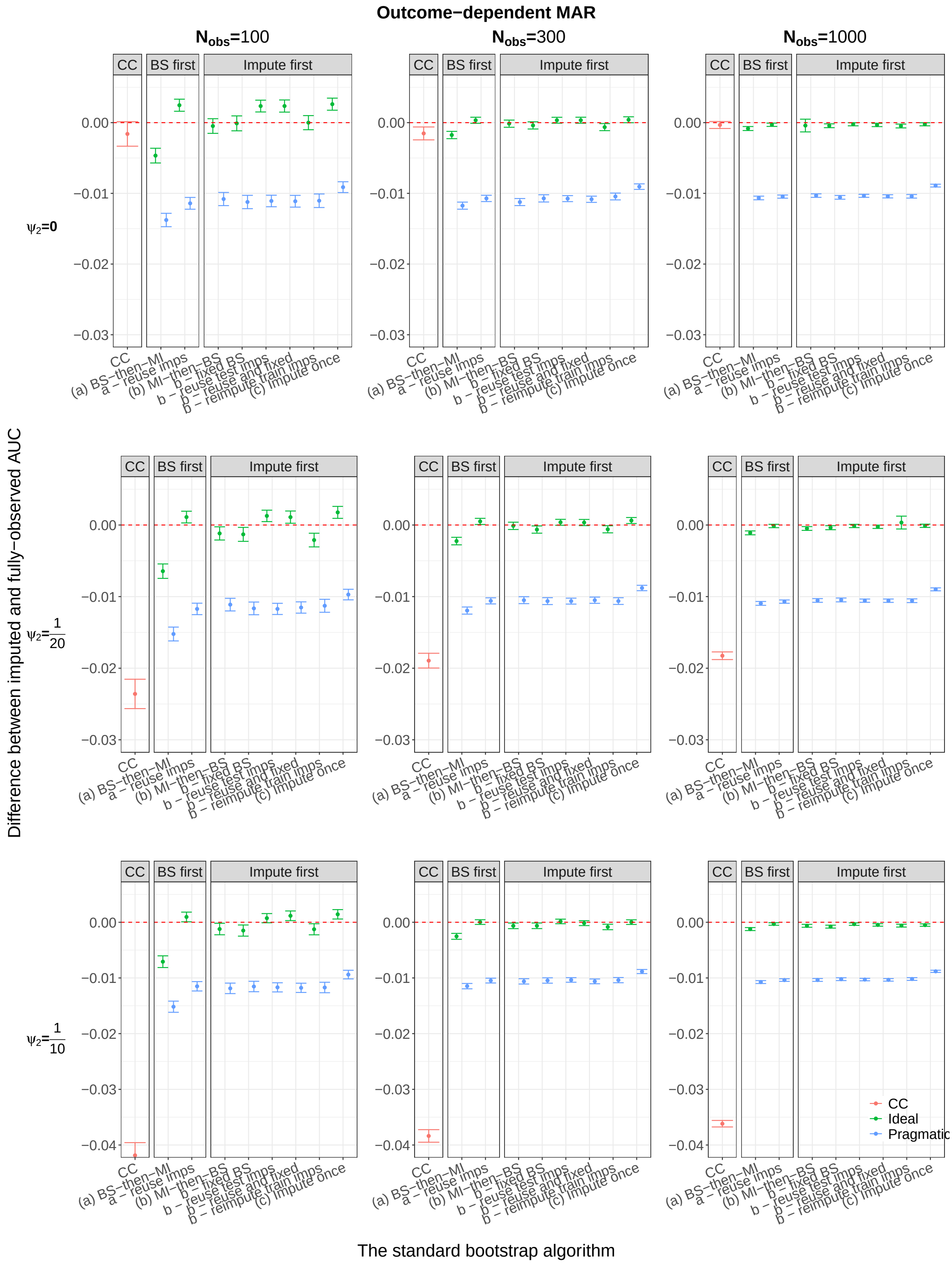

Figure S6: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are outcome-dependent or outcome- and covariate-dependent MAR. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.1.3 The proportion of missingness is 40% (AUC_{imp}-AUC_{obs})

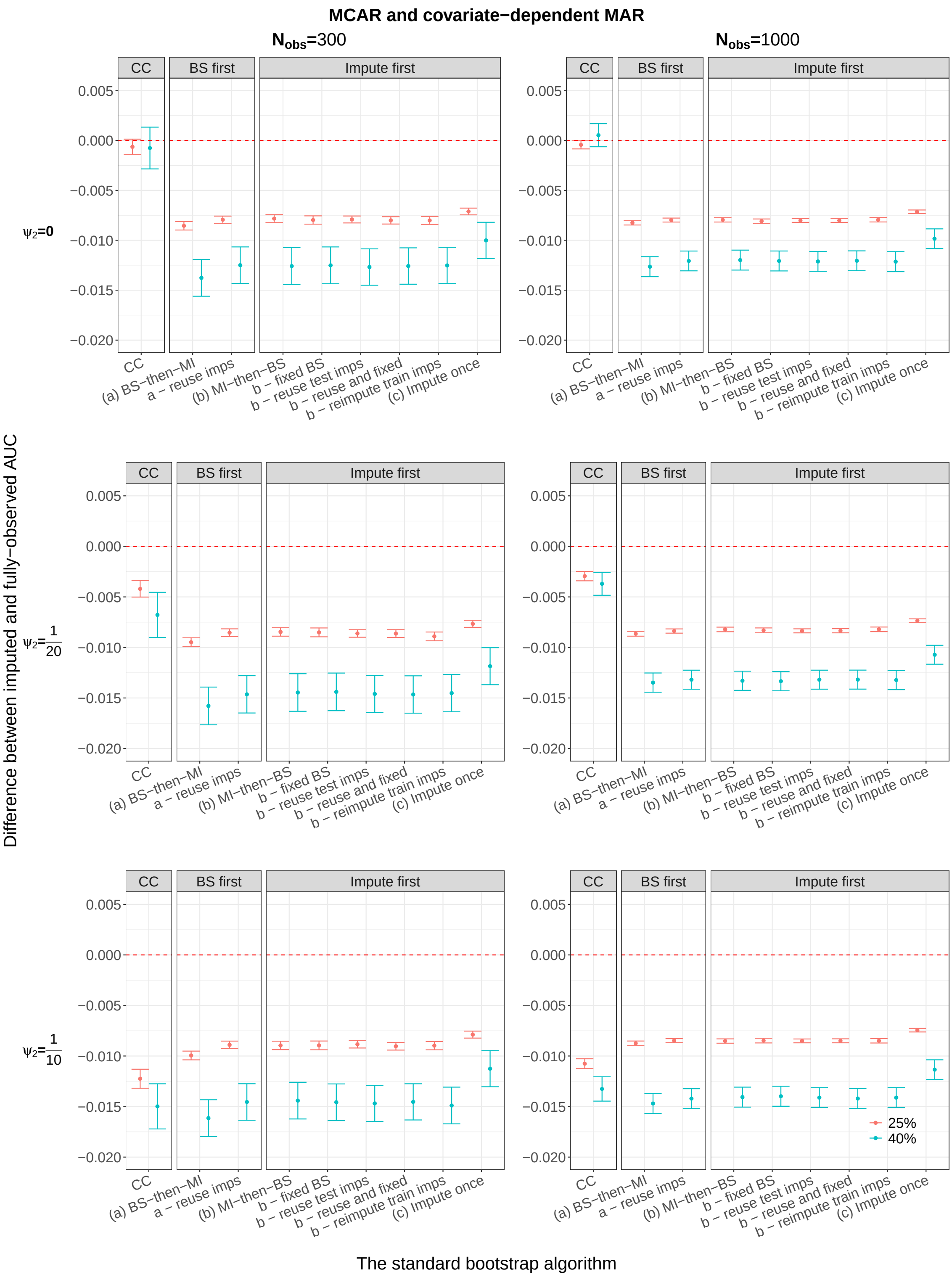

Figure S7: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are MCAR or covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for pragmatic performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

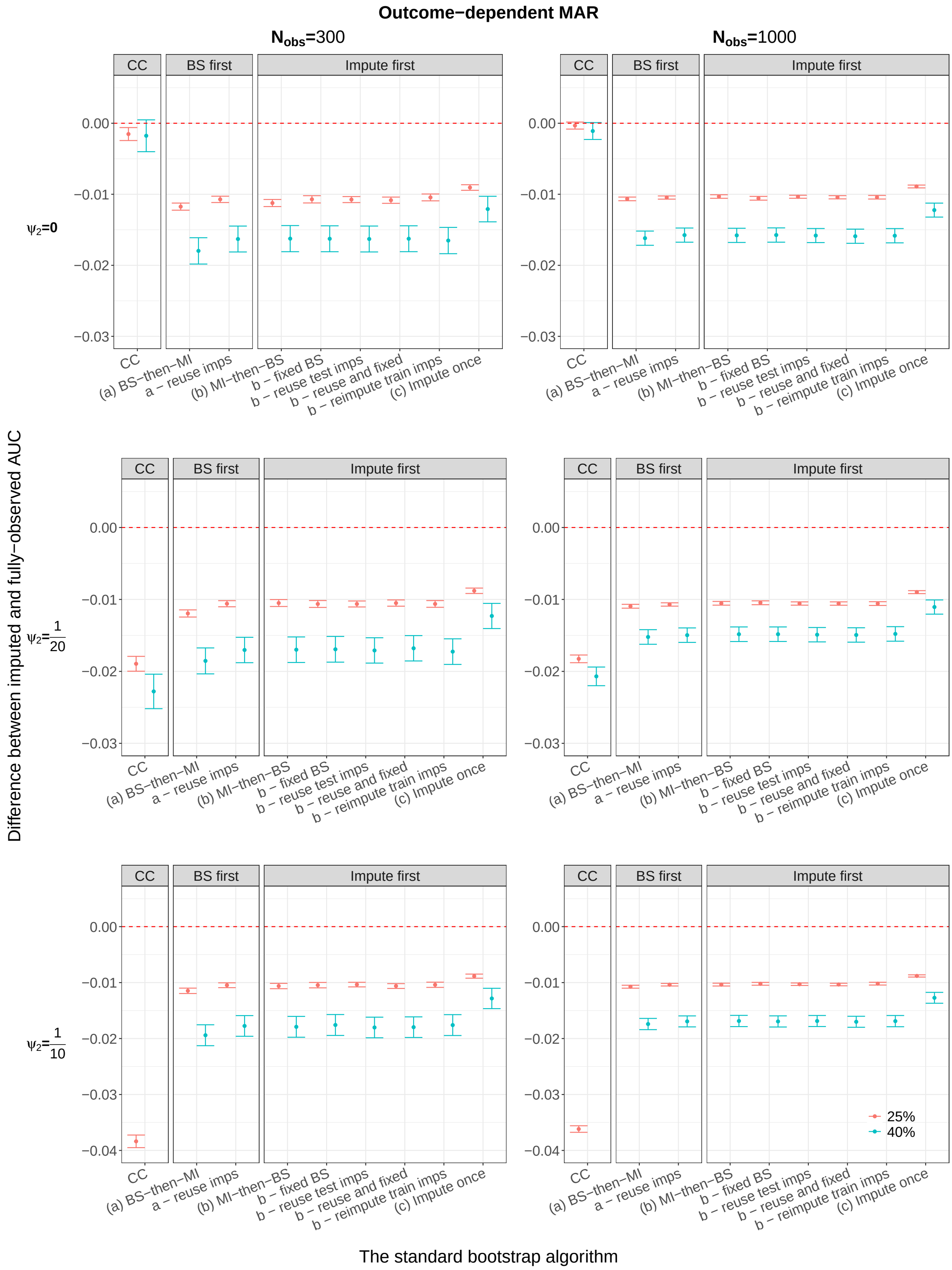

Figure S8: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are outcome-dependent or outcome- and covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for pragmatic performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

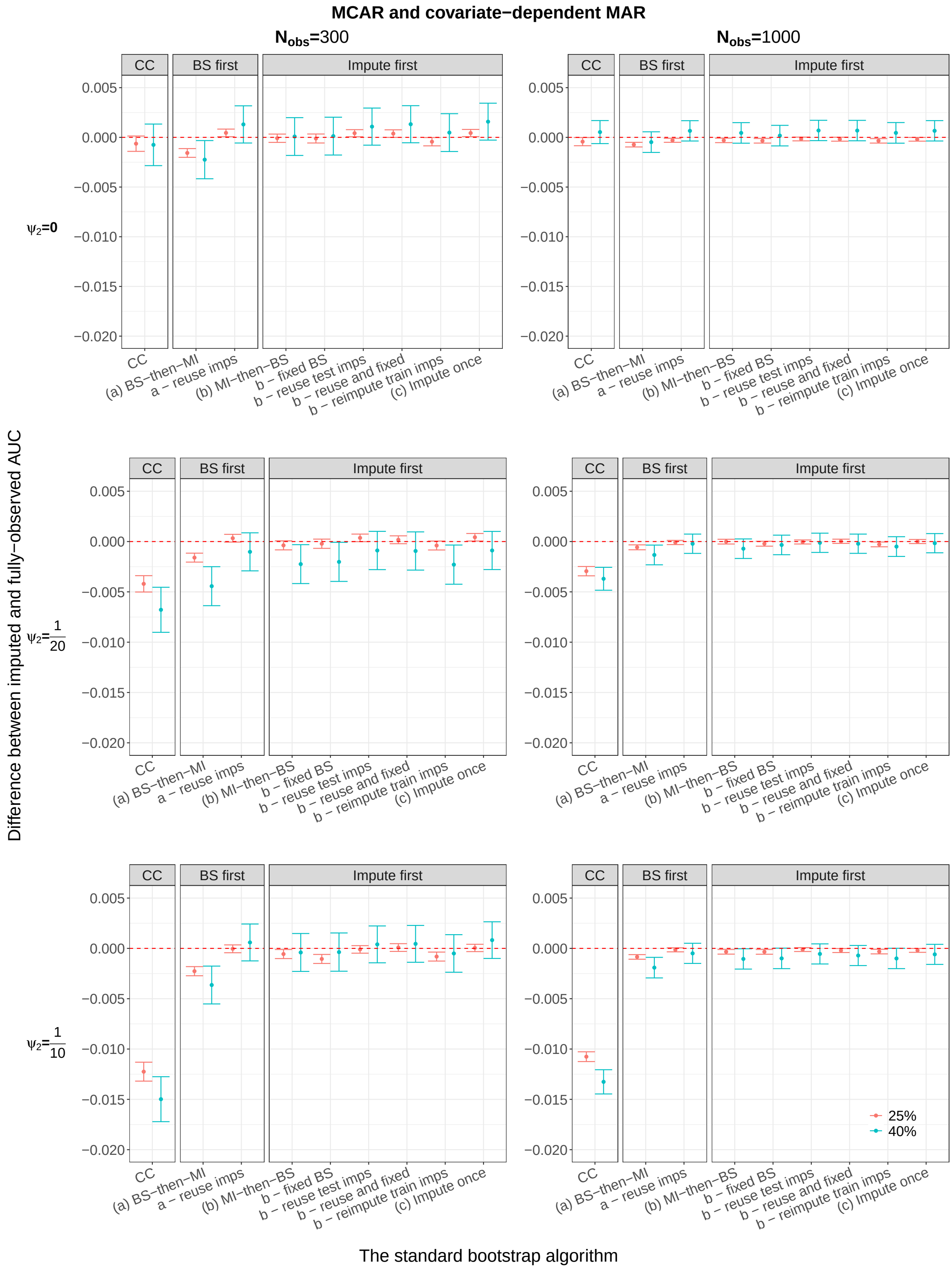

Figure S9: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are MCAR or covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for ideal performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

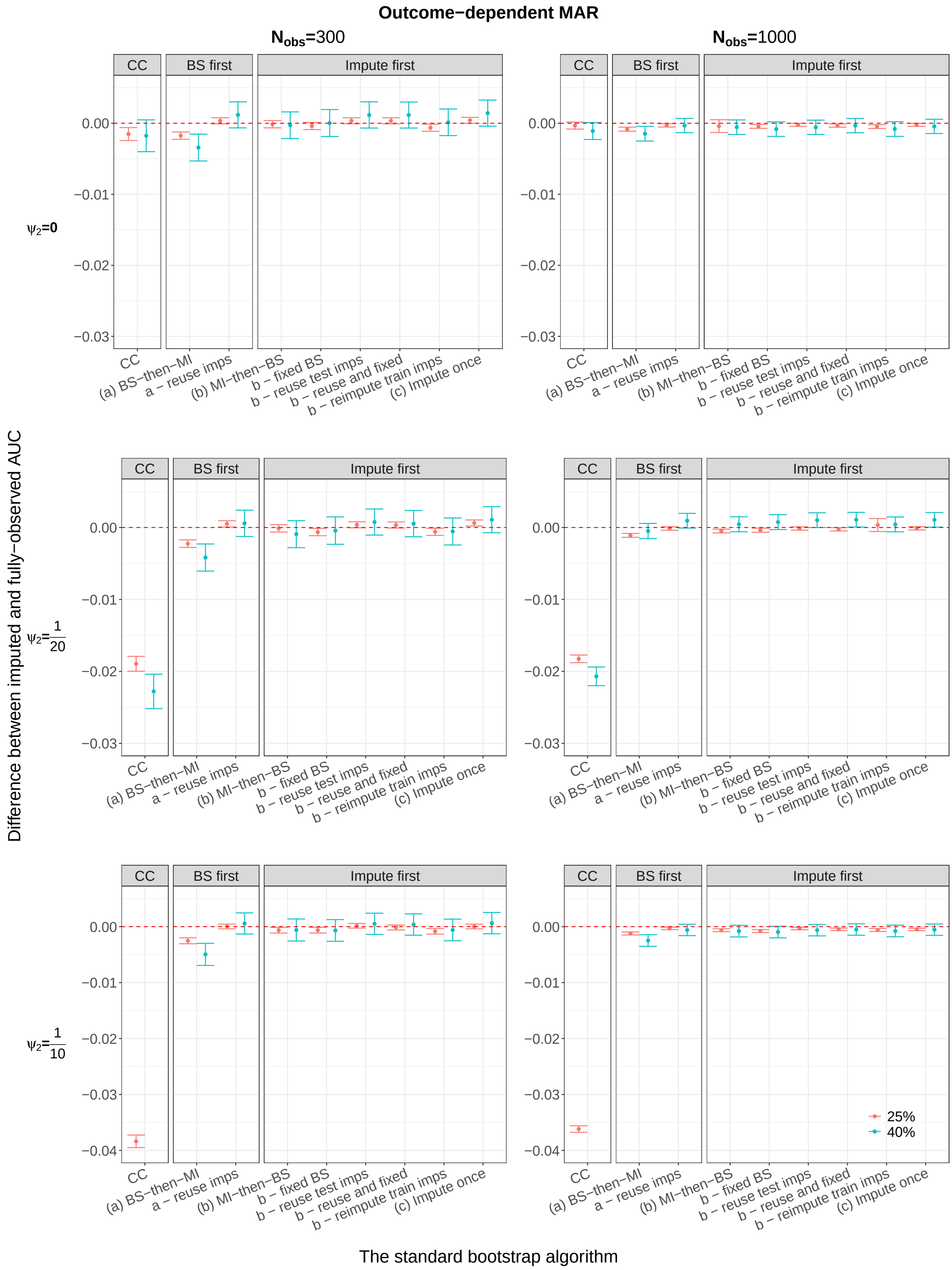

Figure S10: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are outcome-dependent or outcome- and covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for ideal performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.1.4 Comparing M=5 versus M=25 (AUC_{imp}-AUC_{obs})

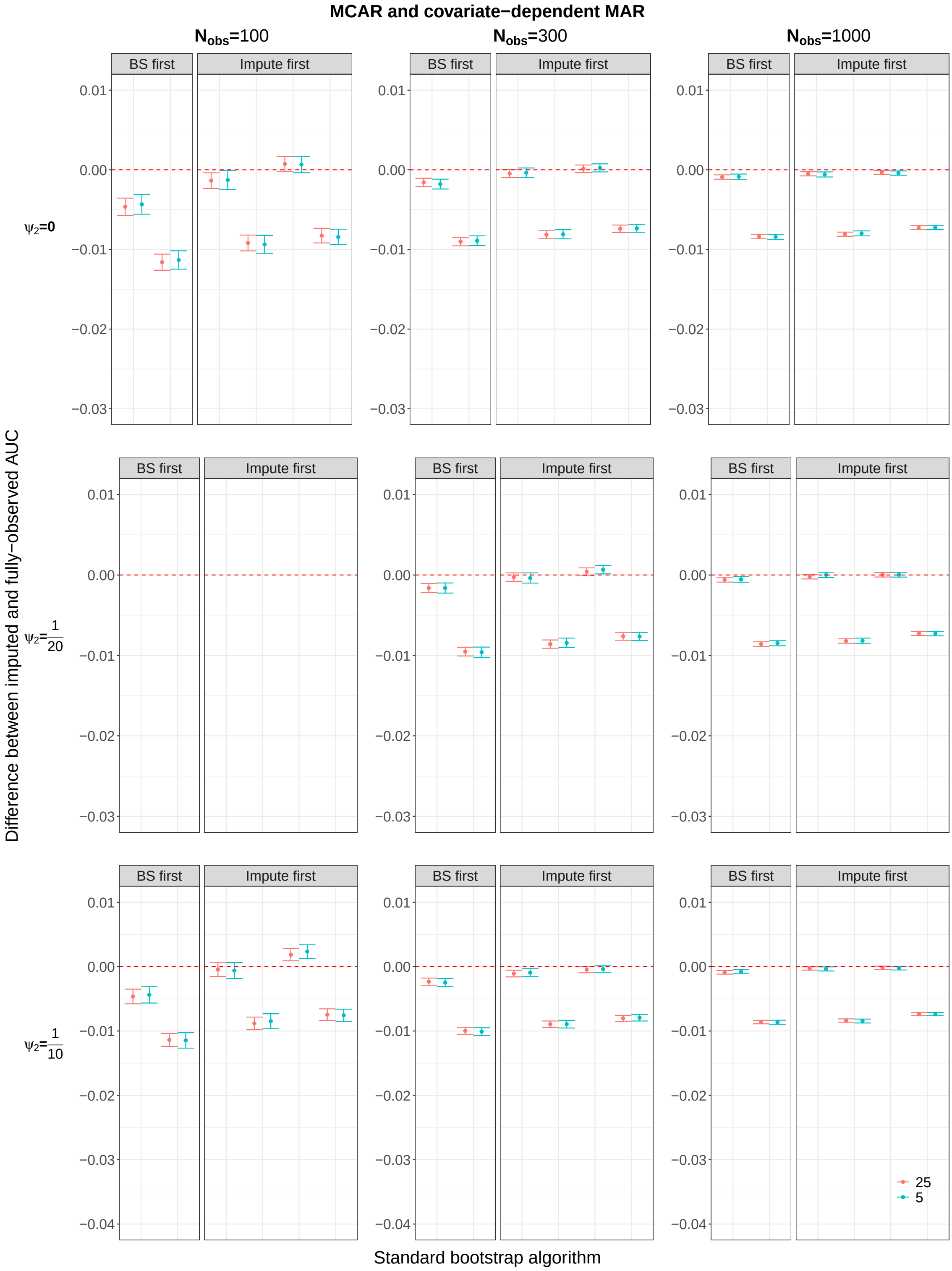

Figure S11: The difference AUC_{imp} - AUC_{obs} when data are MCAR or covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

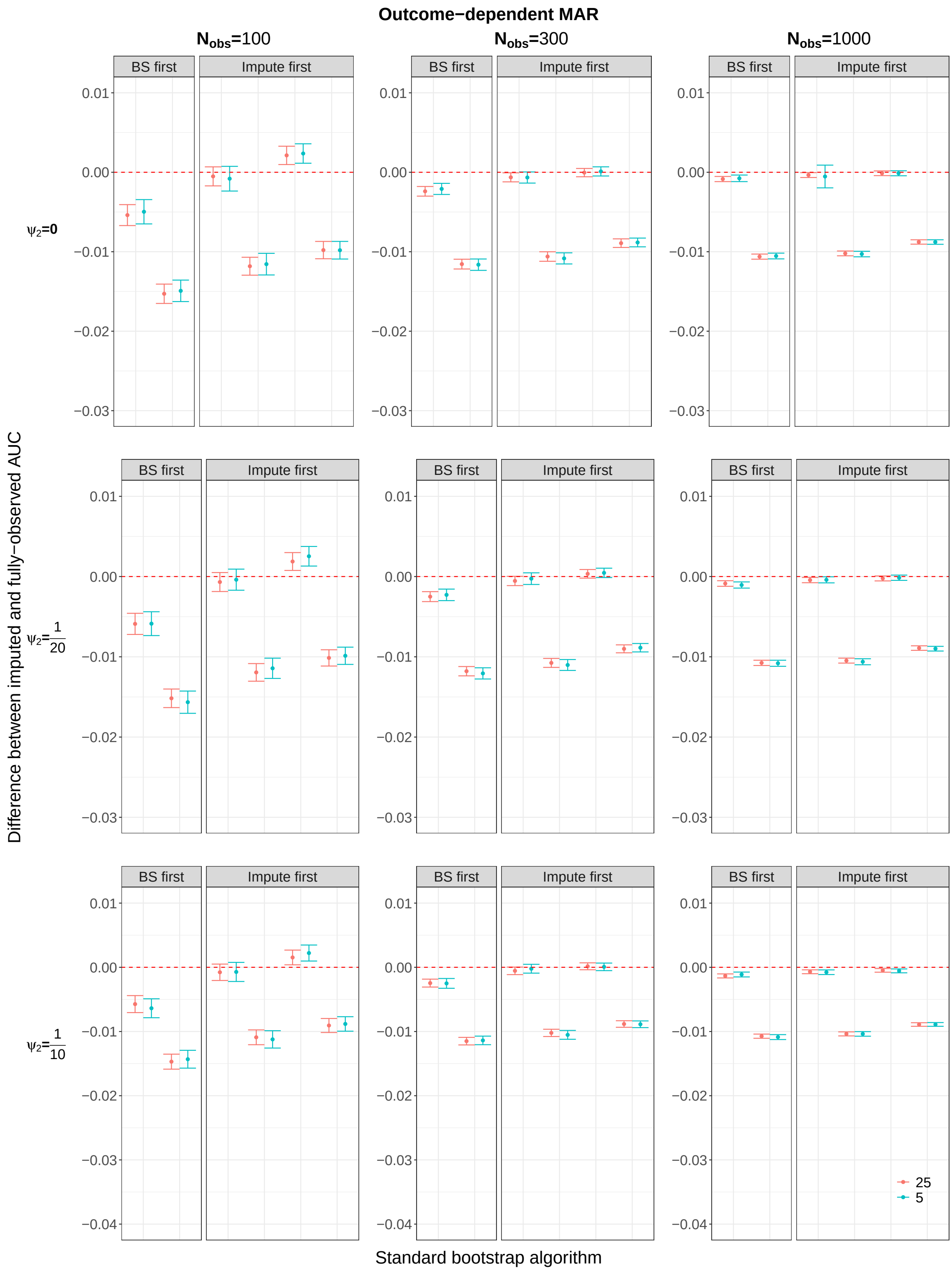

Figure S12: The difference AUC_{imp} - AUC_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.1.5 AUC from imputation methods compared to the target AUC (AUC_{target}) using a larger validation set

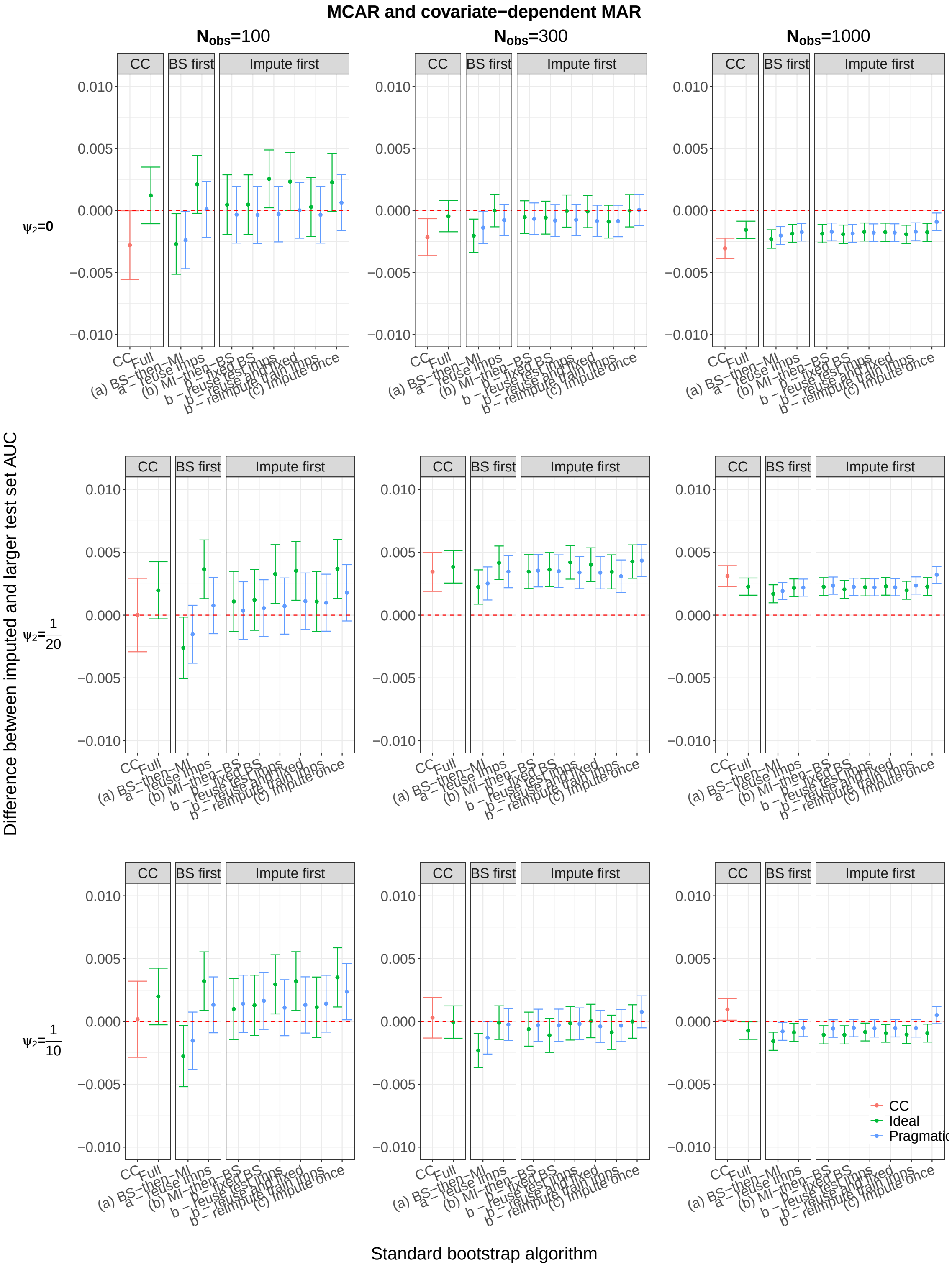

Figure S13: The difference AUC_{imp} - AUC_{target} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

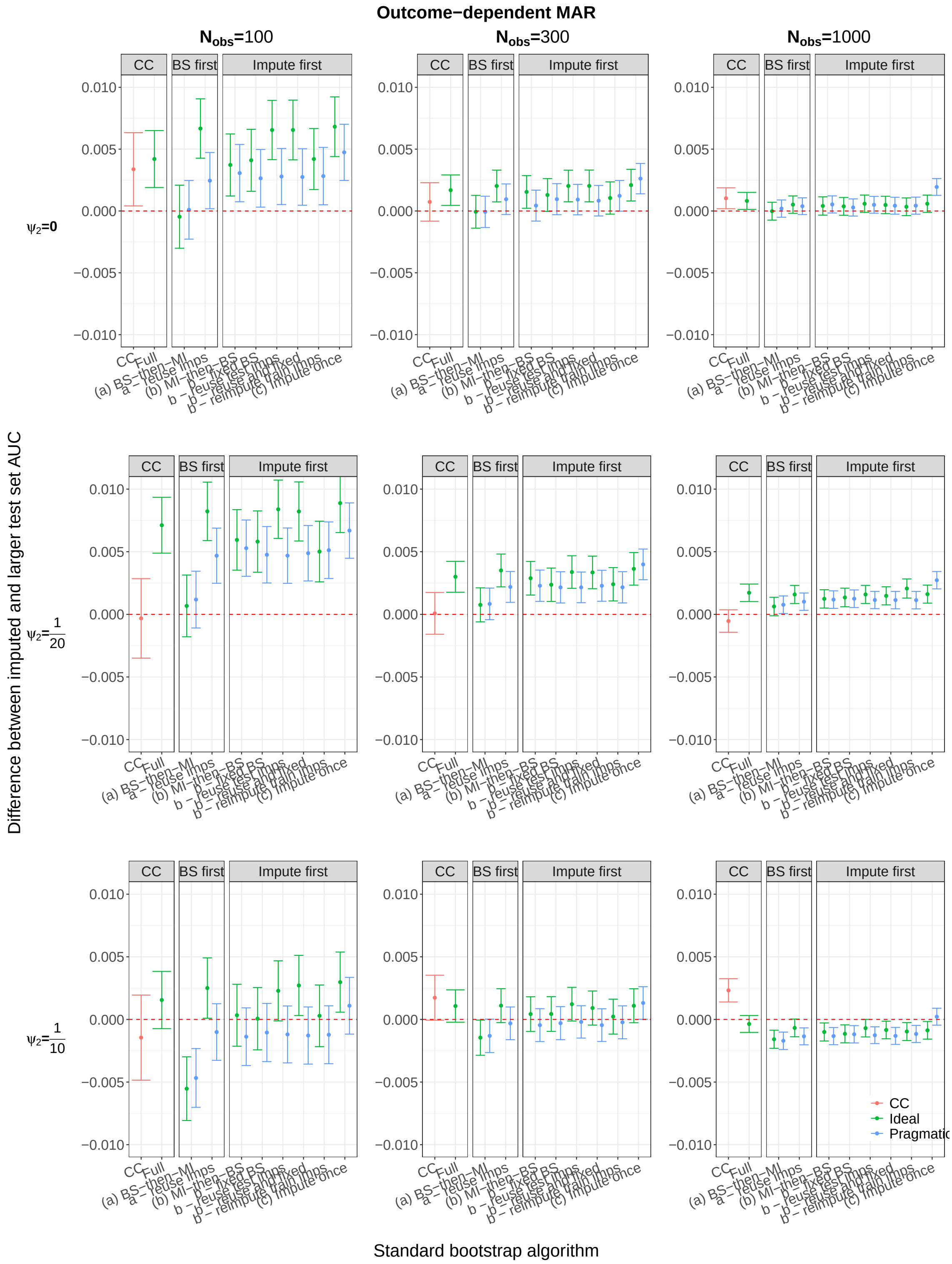

Figure S14: The difference AUC$_{imp}$ - AUC$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.2 The standard bootstrap: Brier Score

S4.2.1 Reusing versus re-imputing for test performance of the standard algorithm

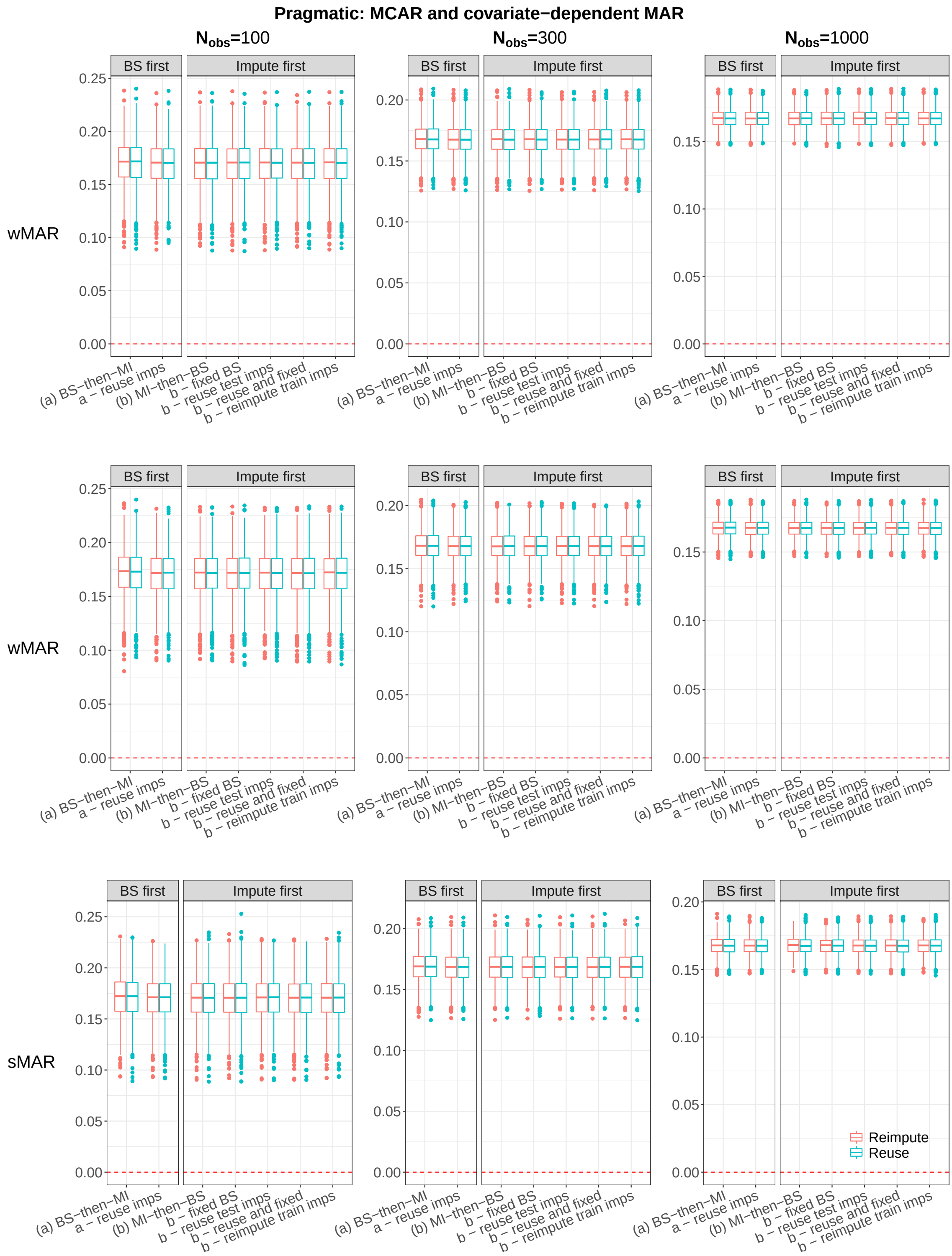

Figure S15: A comparison of reusing versus re-imputing test datasets on the Brier score estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the Brier score estimated when data are fully-observed (Brier_{imp}-Brier_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

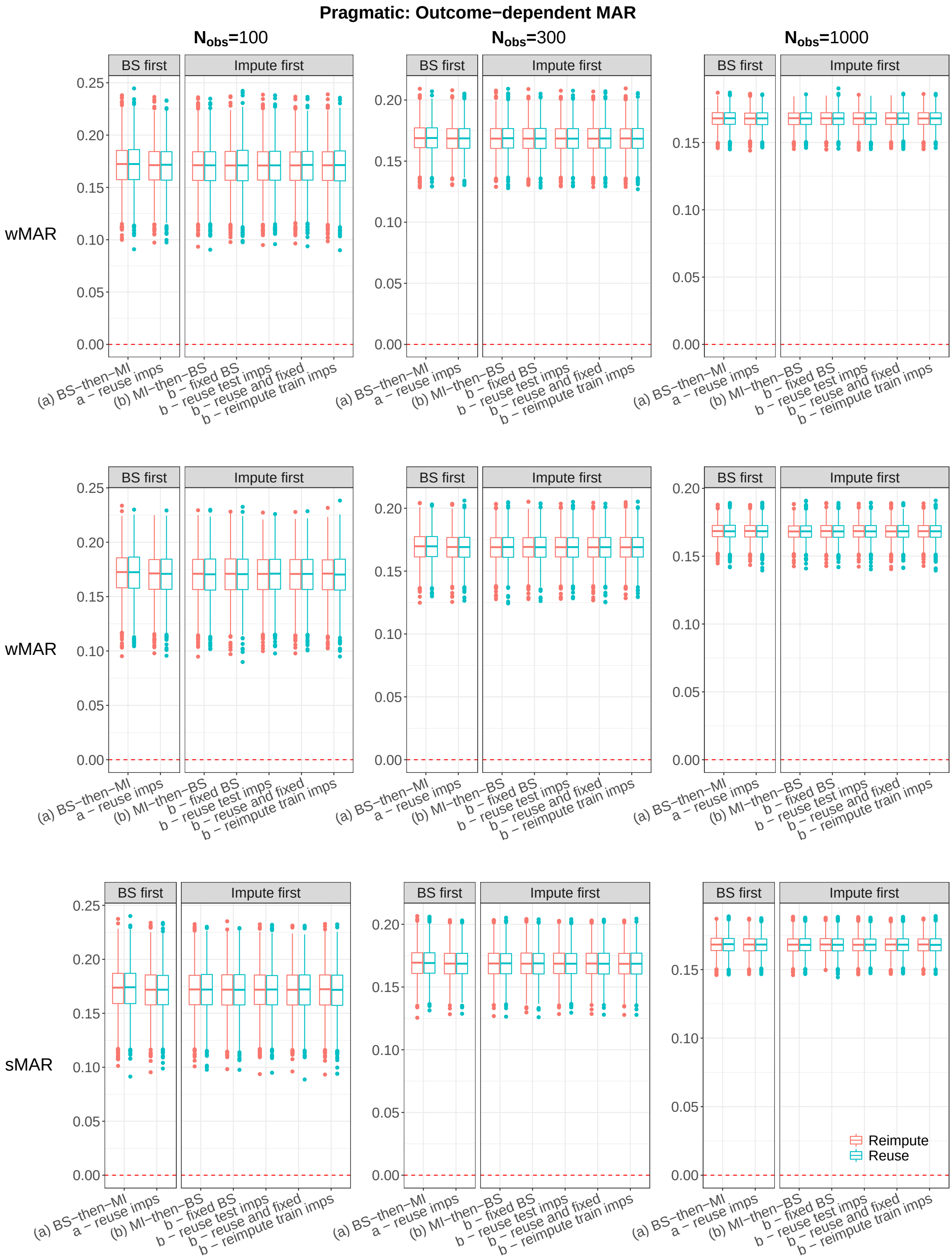

Figure S16: A comparison of reusing versus re-imputing test datasets on the Brier score estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for each method which are compared to the Brier score estimated when data are fully-observed (Brier$_{imp}$- Brier$_{obs}$). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

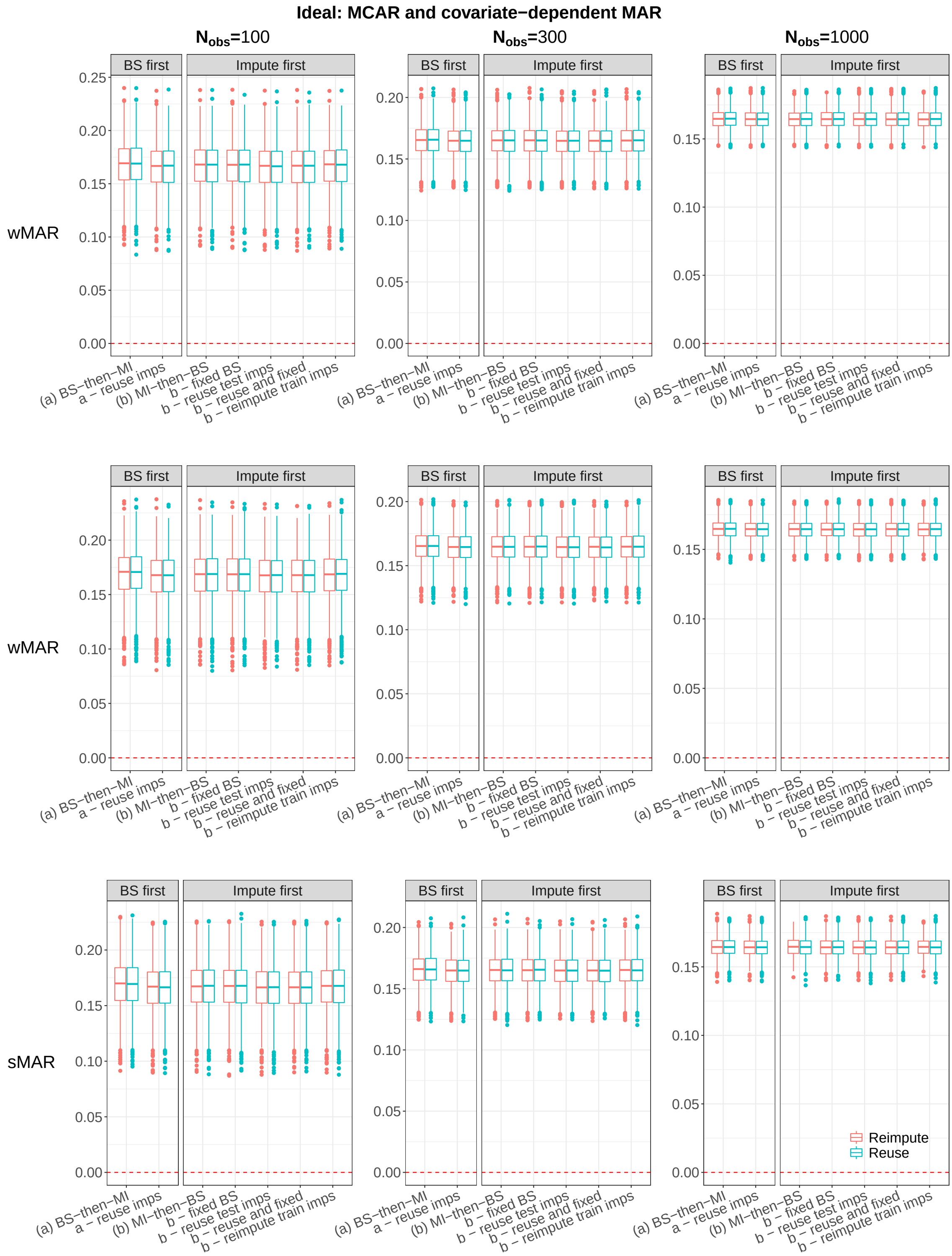

Figure S17: A comparison of reusing versus re-imputing test datasets on the Brier score estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the Brier score estimated when data are fully-observed (Brier_{imp}- Brier_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

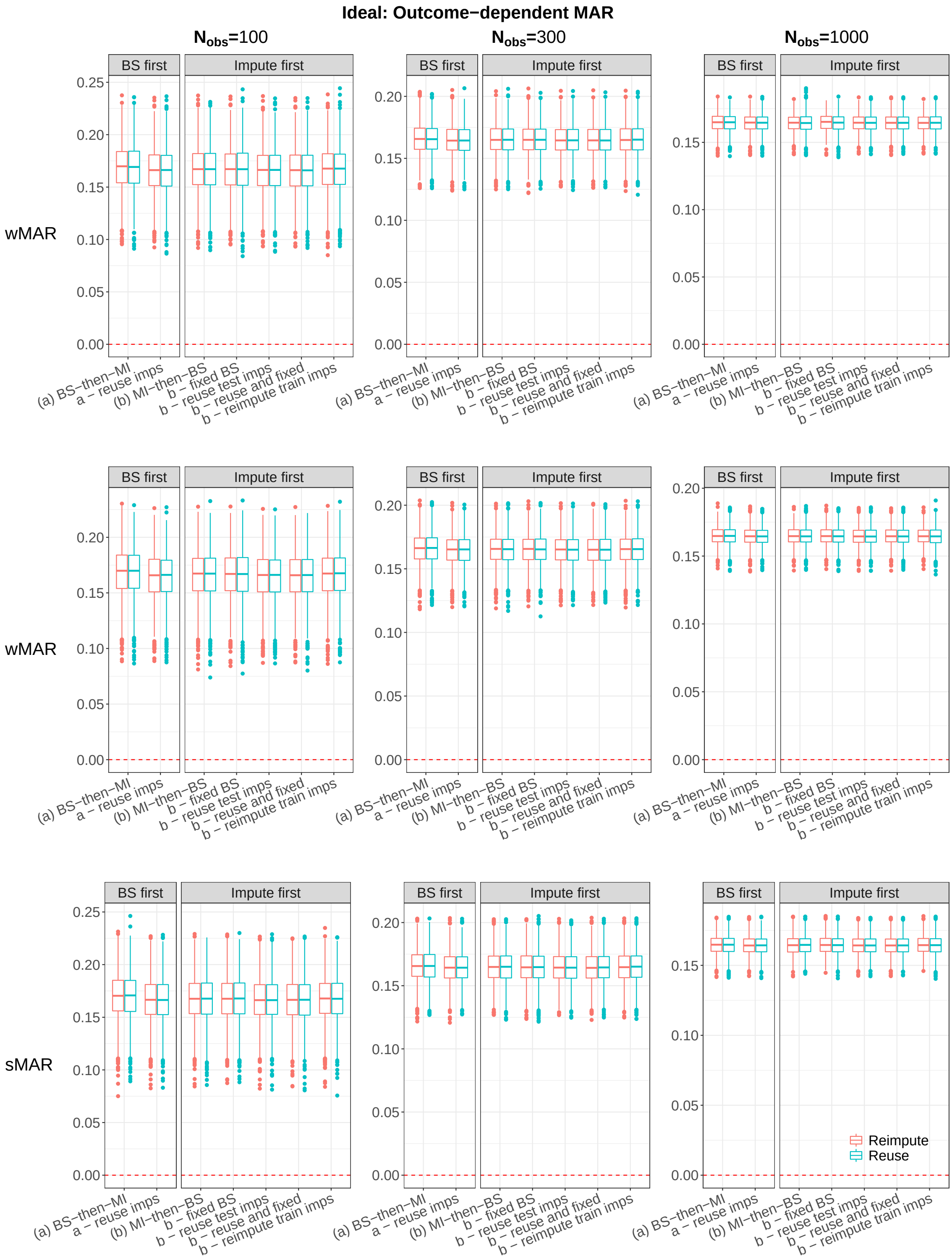

Figure S18: A comparison of reusing versus re-imputing test datasets on the Brier score estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for each method which are compared to the Brier score estimated when data are fully-observed (Brier_{imp}- Brier_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.2.2 Brier Score from imputation methods compared to the fully-observed Brier Score (Brier_{imp}-Brier_{obs})

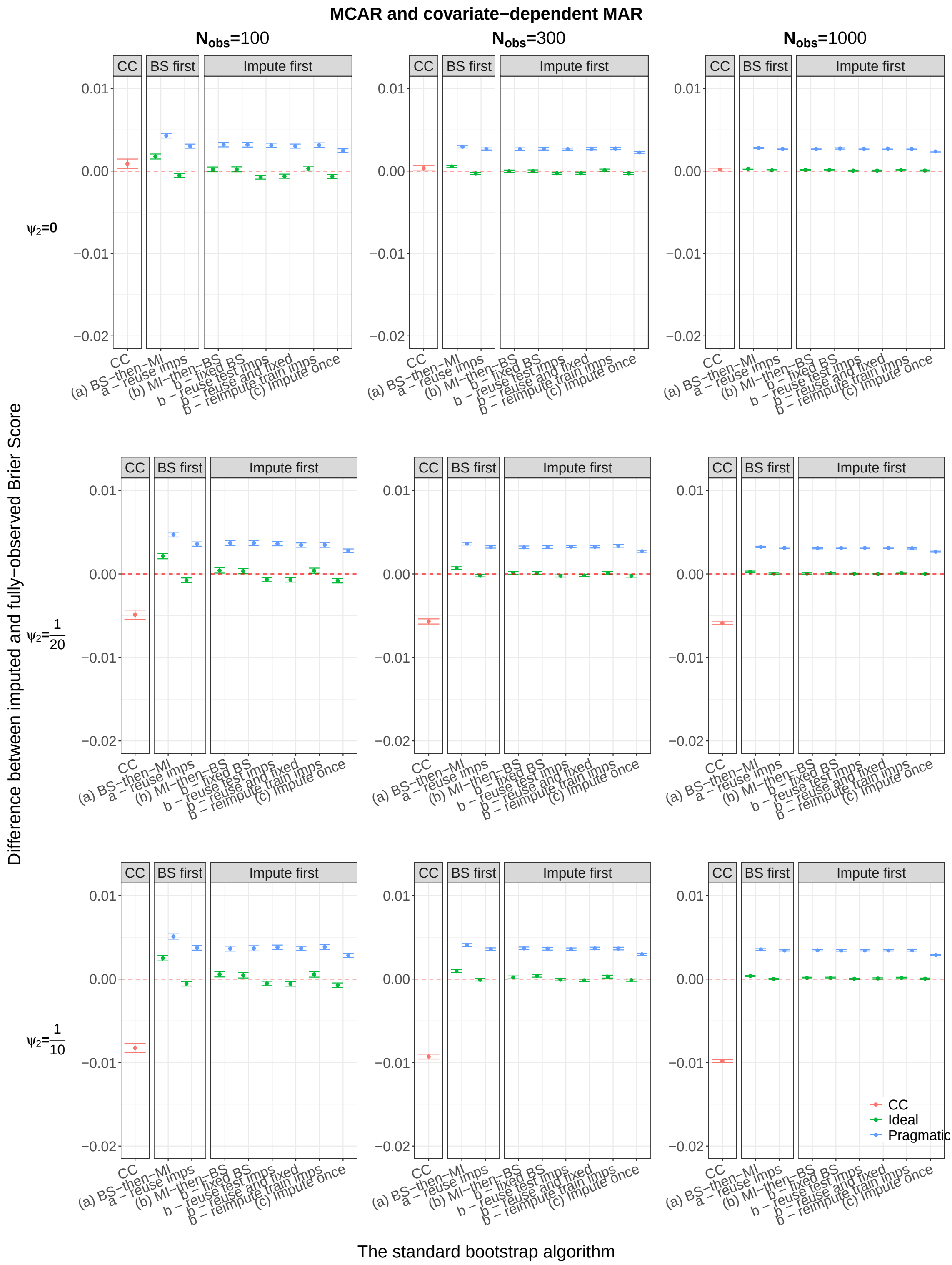

Figure S19: The difference Brier_{imp} - Brier_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier_{imp} - Brier_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

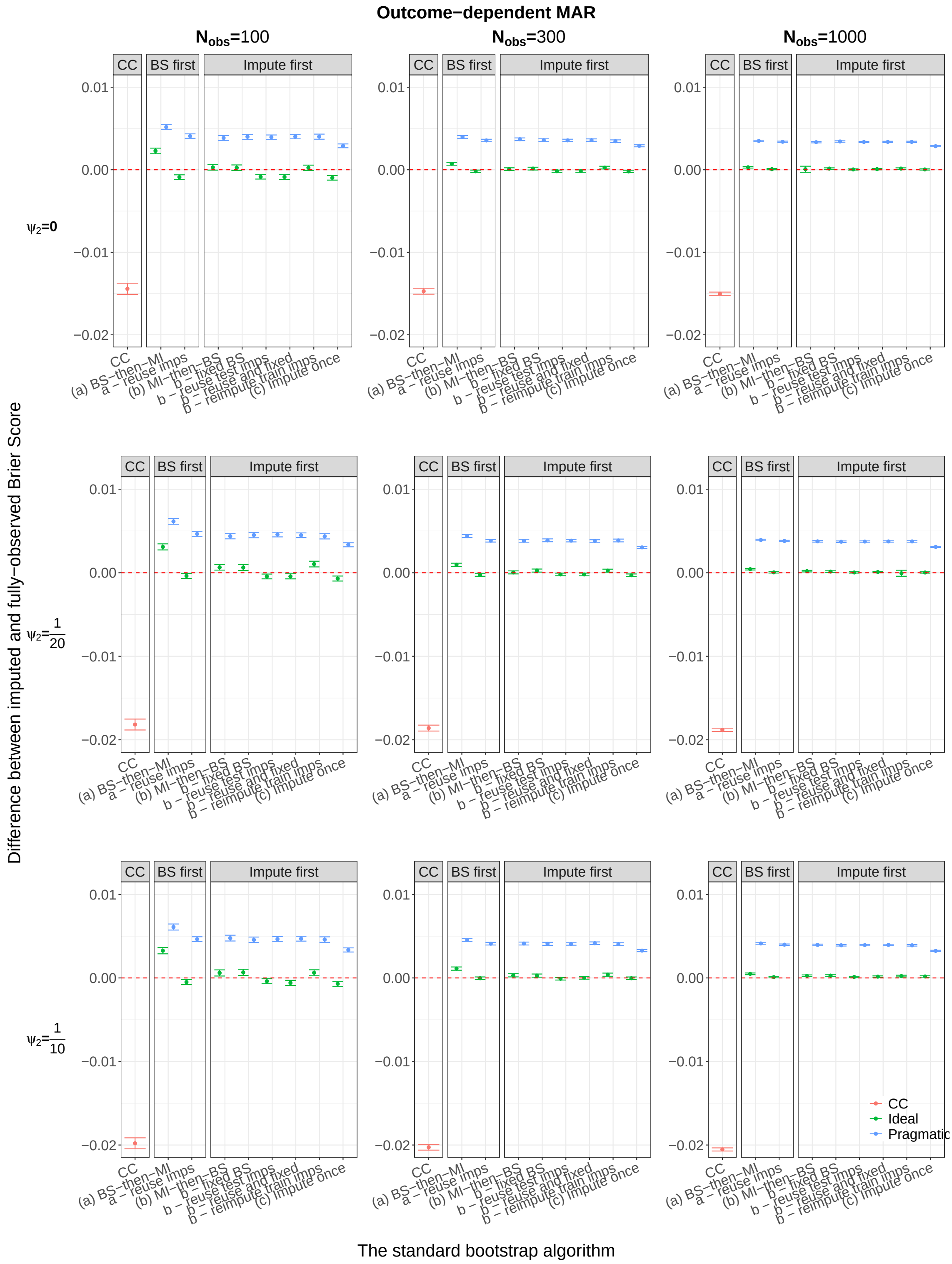

Figure S20: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.2.3 The proportion of missingness is 40% (Brier_{imp}-Brier_{obs})

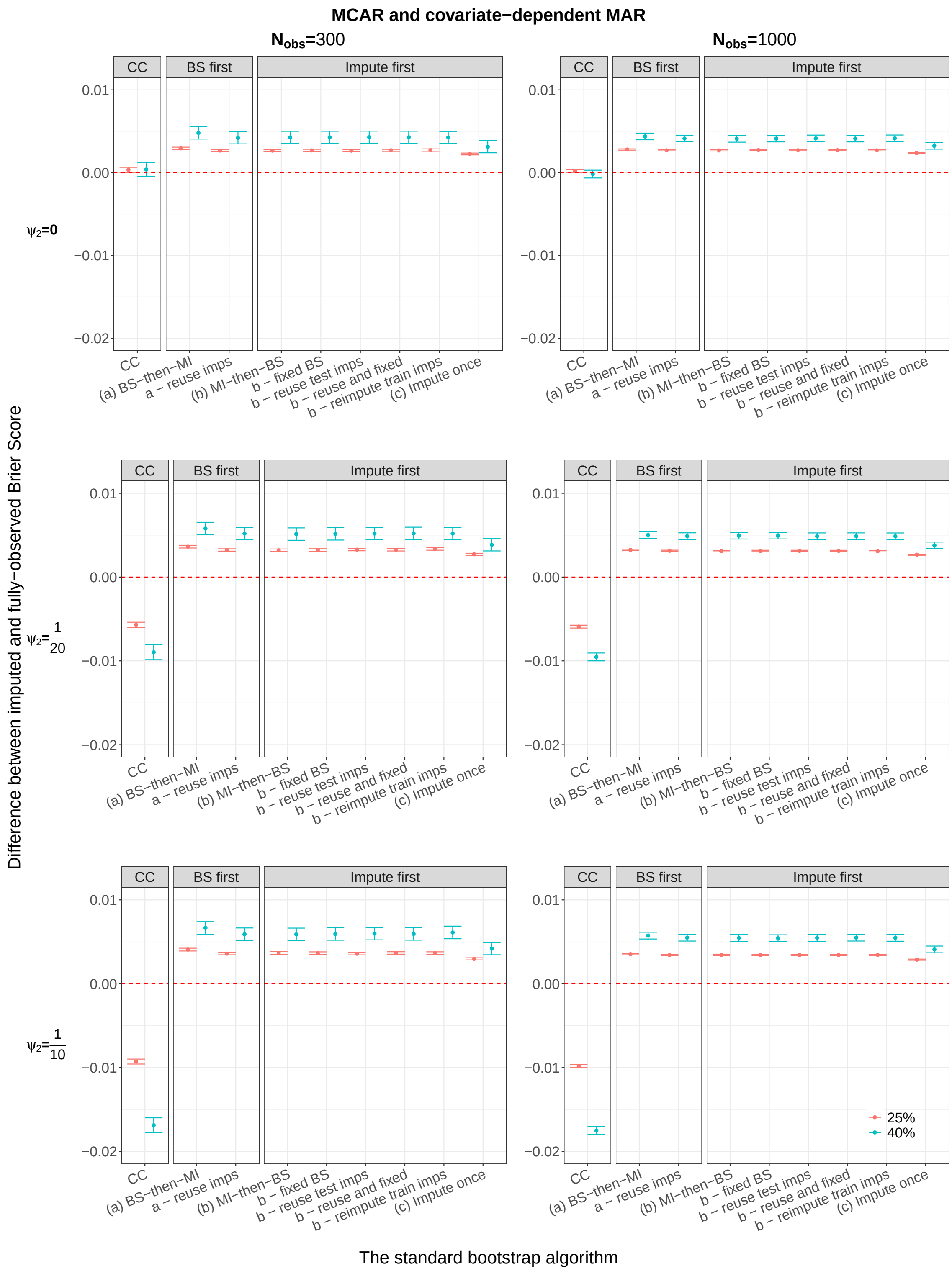

Figure S21: Comparing the impact of increasing the percentage of missingness on the difference Brier_{imp} - Brier_{obs} when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of BrierC_{imp} - Brier_{obs}. Red denotes Brier_{imp} - Brier_{obs} when 25% of X_1 values are missing and blue denotes Brier_{imp} - Brier_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

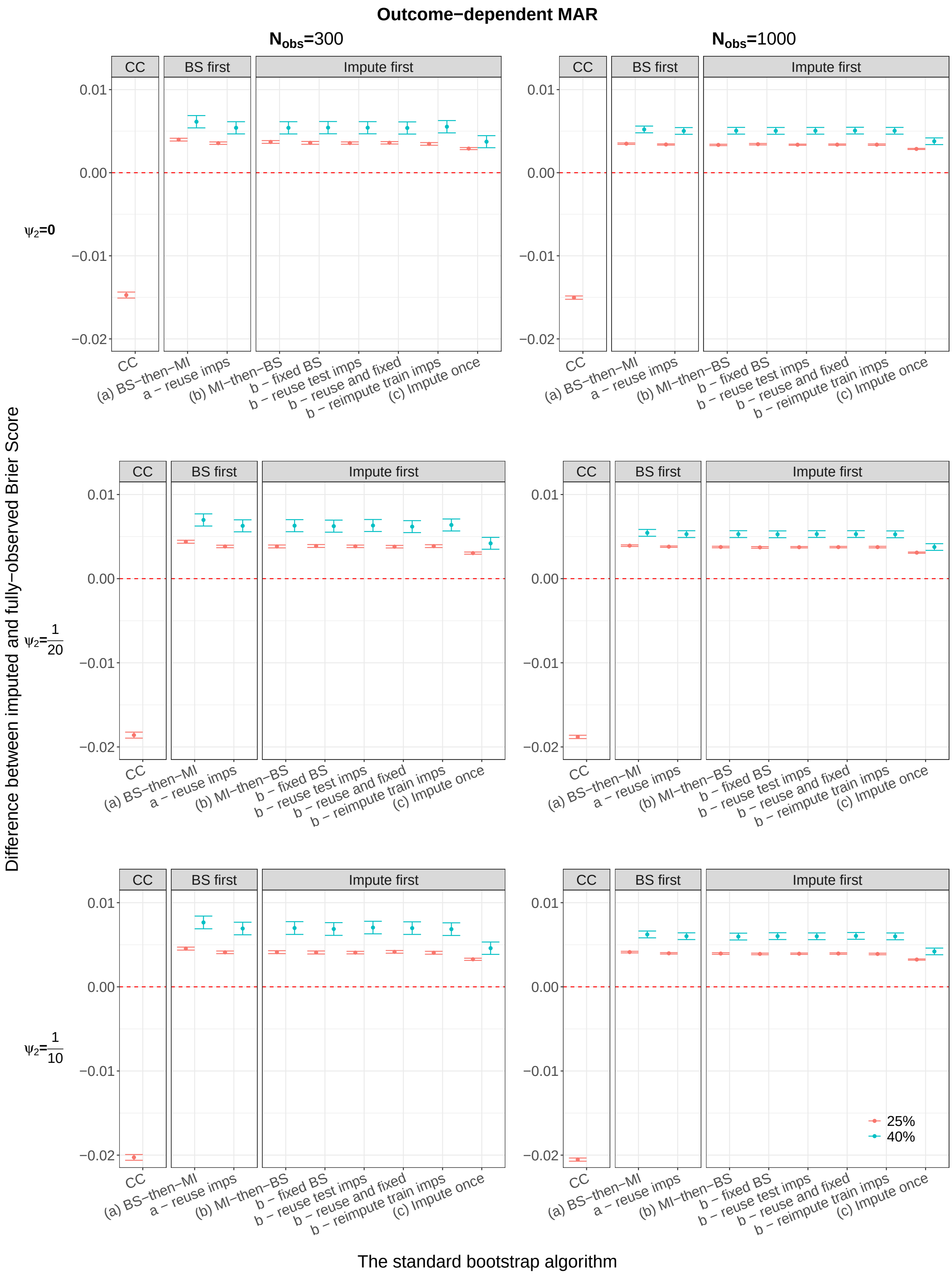

Figure S22: Comparing the impact of increasing the percentage of missingness on the difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. Red denotes Brier$_{imp}$ - Brier$_{obs}$ when 25% of X_1 values are missing and blue denotes Brier$_{imp}$ - Brier$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

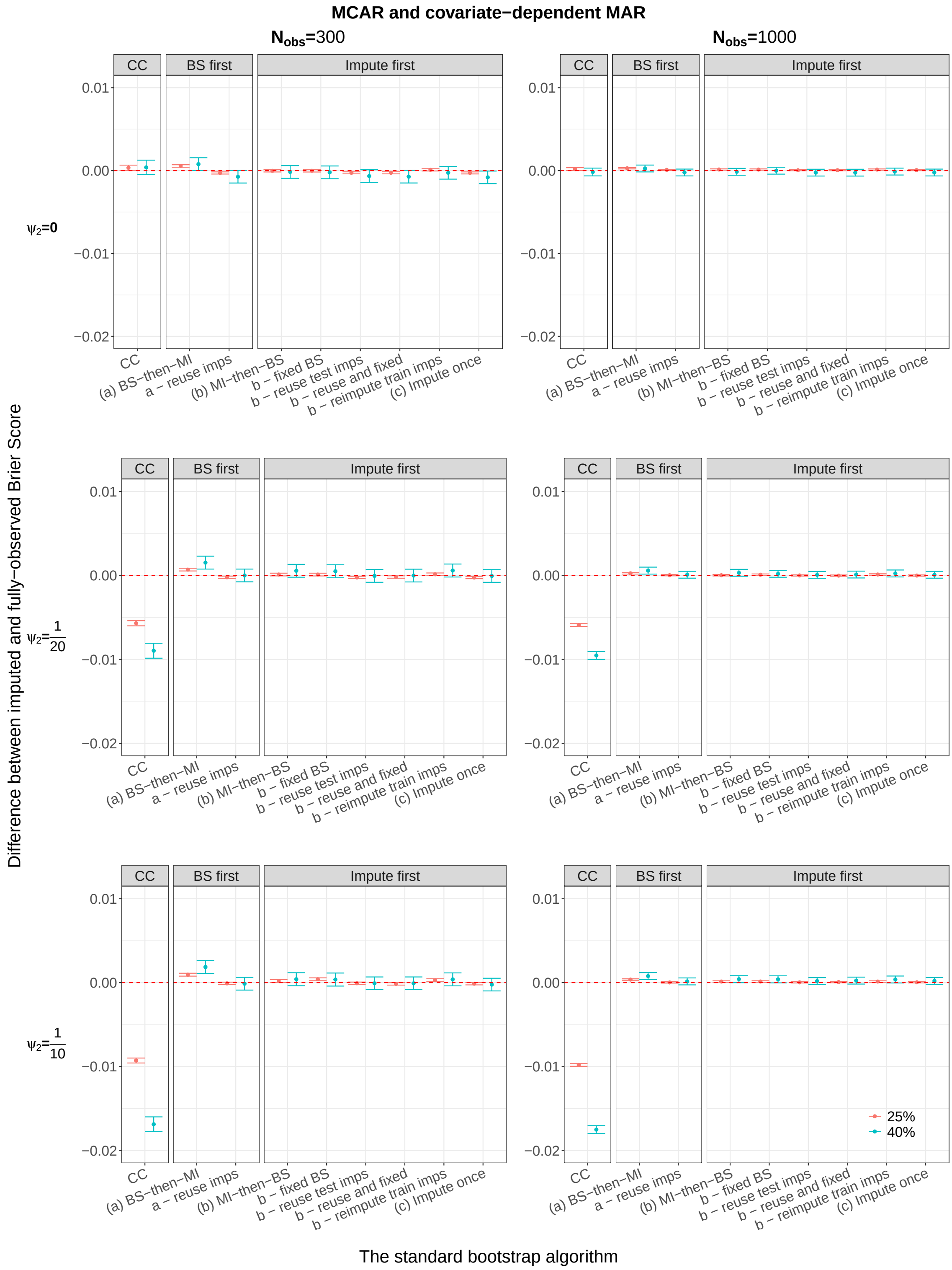

Figure S23: Comparing the impact of increasing the percentage of missingness on the difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of BrierC$_{imp}$ - Brier$_{obs}$. Red denotes Brier$_{imp}$ - Brier$_{obs}$ when 25% of X_1 values are missing and blue denotes Brier$_{imp}$ - Brier$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

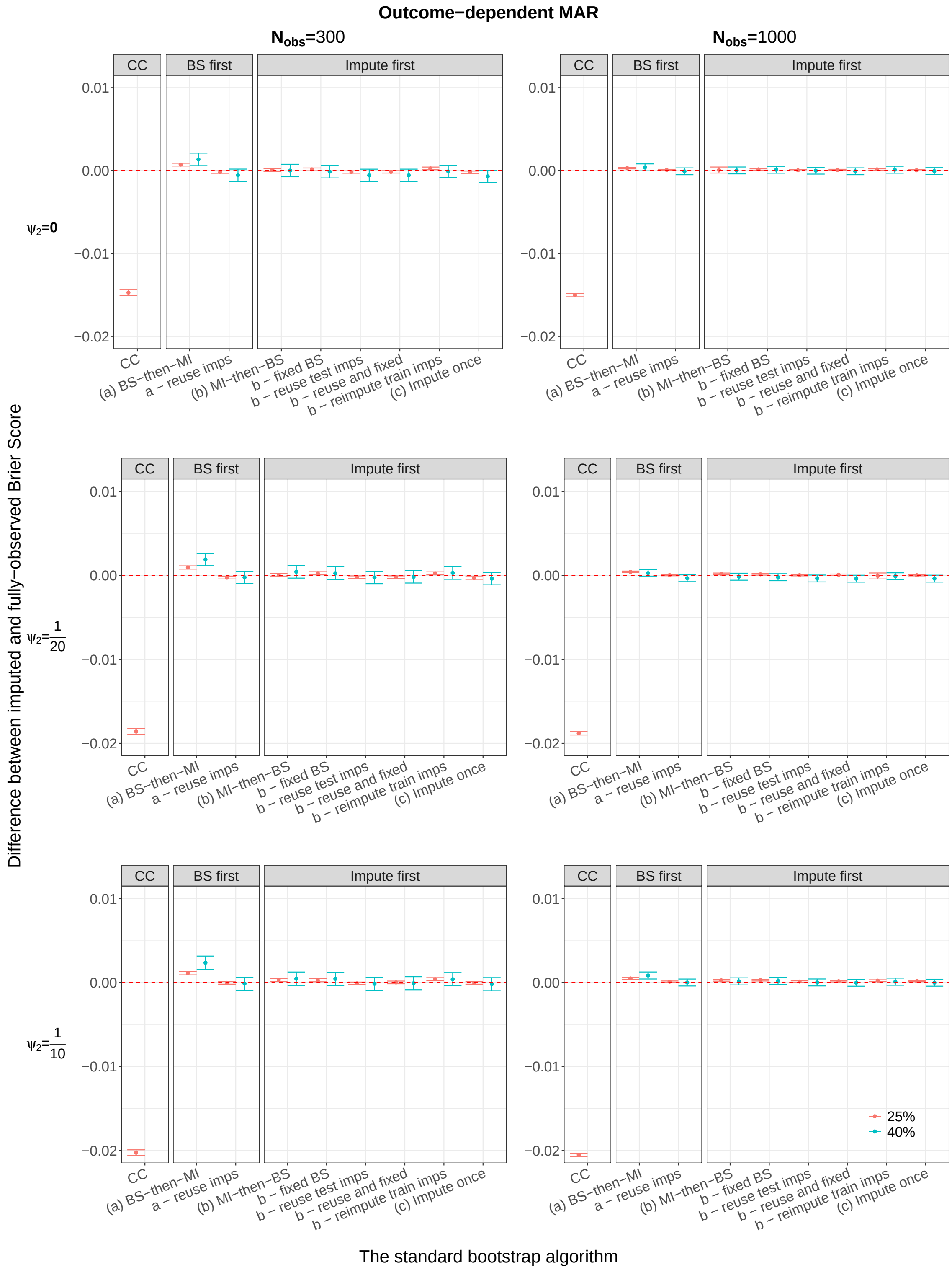

Figure S24: Comparing the impact of increasing the percentage of missingness on the difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. Red denotes Brier$_{imp}$ - Brier$_{obs}$ when 25% of X_1 values are missing and blue denotes Brier$_{imp}$ - Brier$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.2.4 Comparing M=5 versus M=25 (Brier$_{imp}$-Brier$_{obs}$)

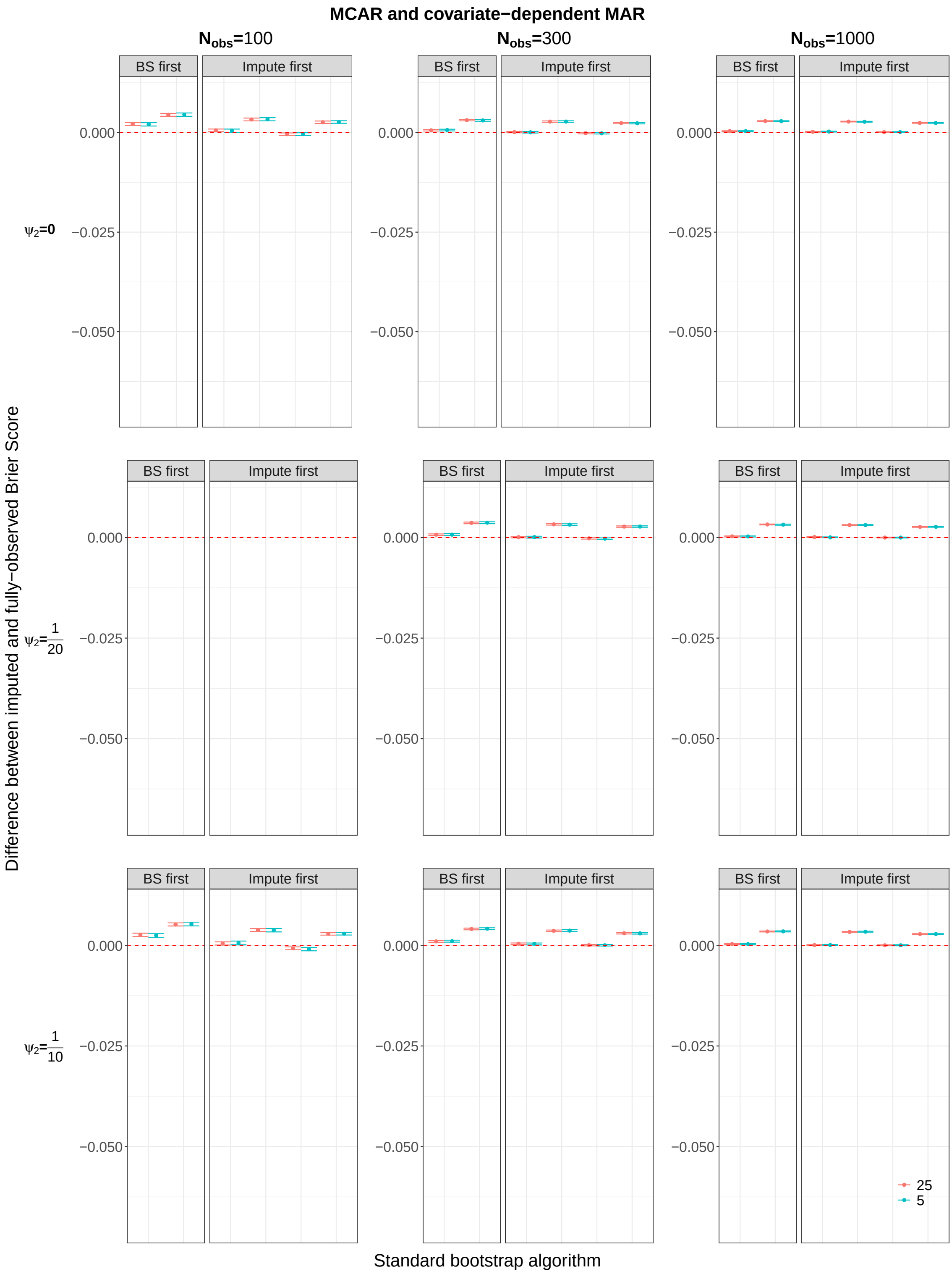

Figure S25: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

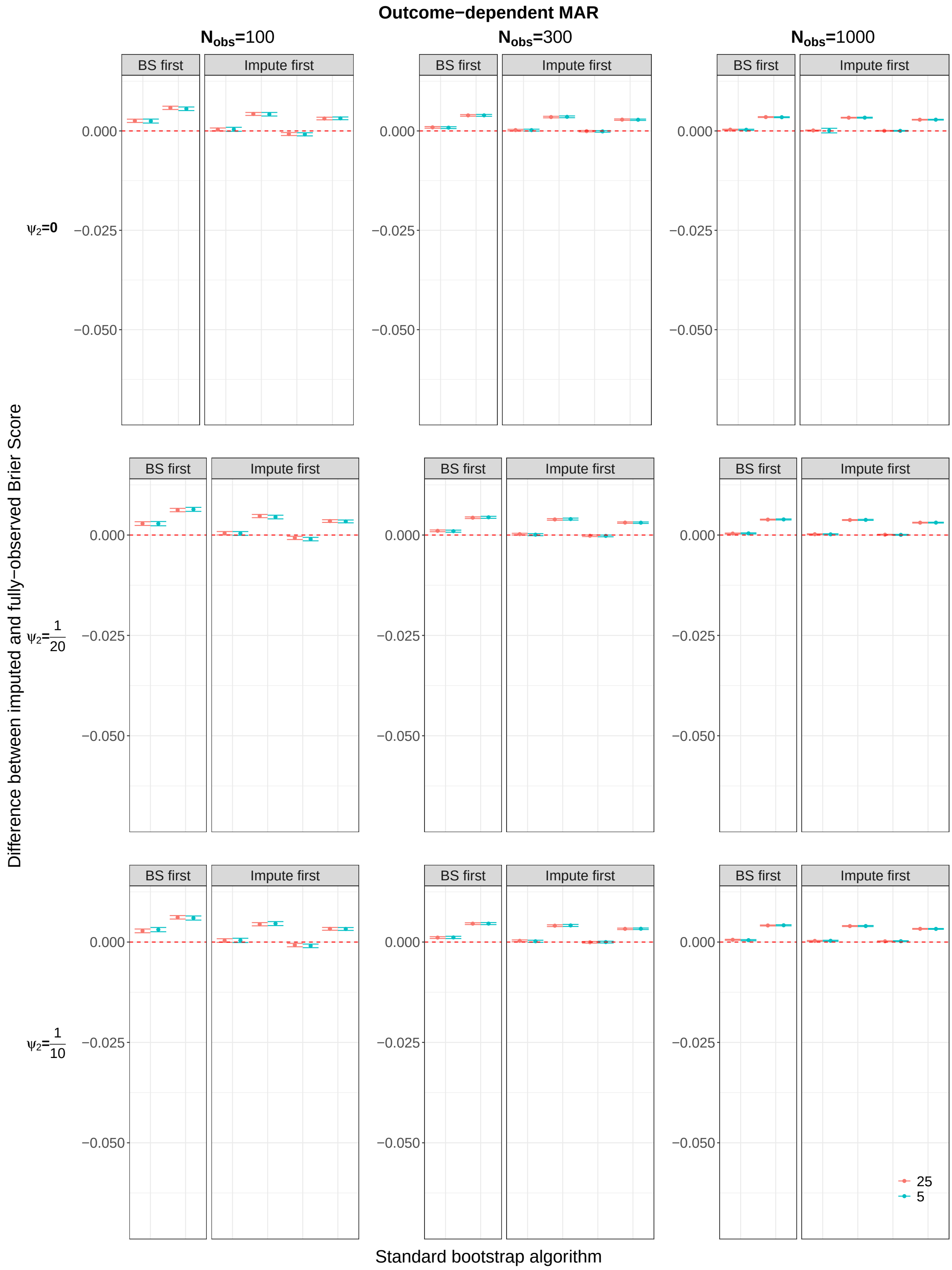

Figure S26: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.2.5 Brier Score from imputation methods compared to the target Brier Score (Brier_{target}) using a larger validation set

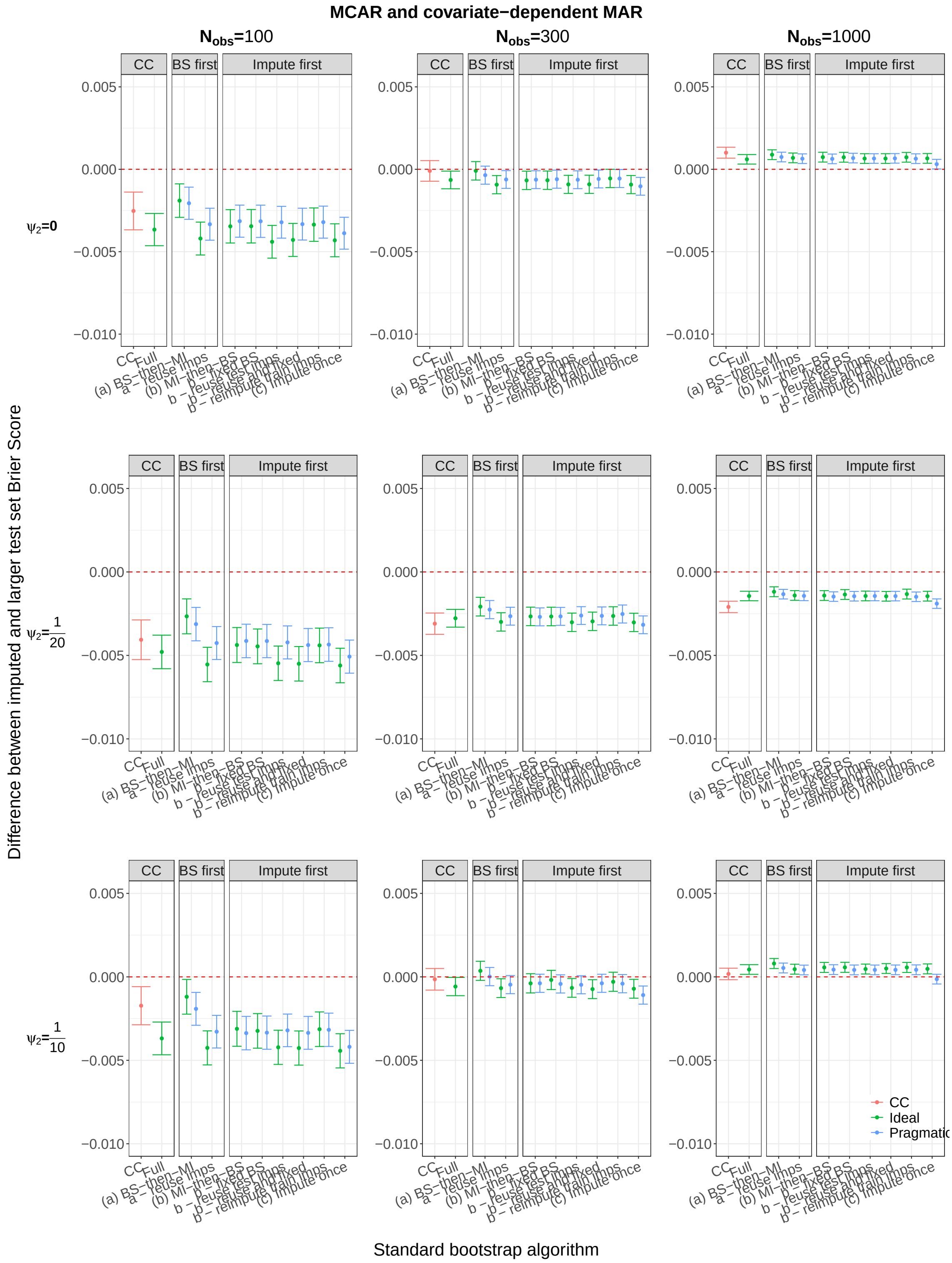

Figure S27: The difference Brier_{imp} - Brier_{target} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier_{imp} - Brier_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

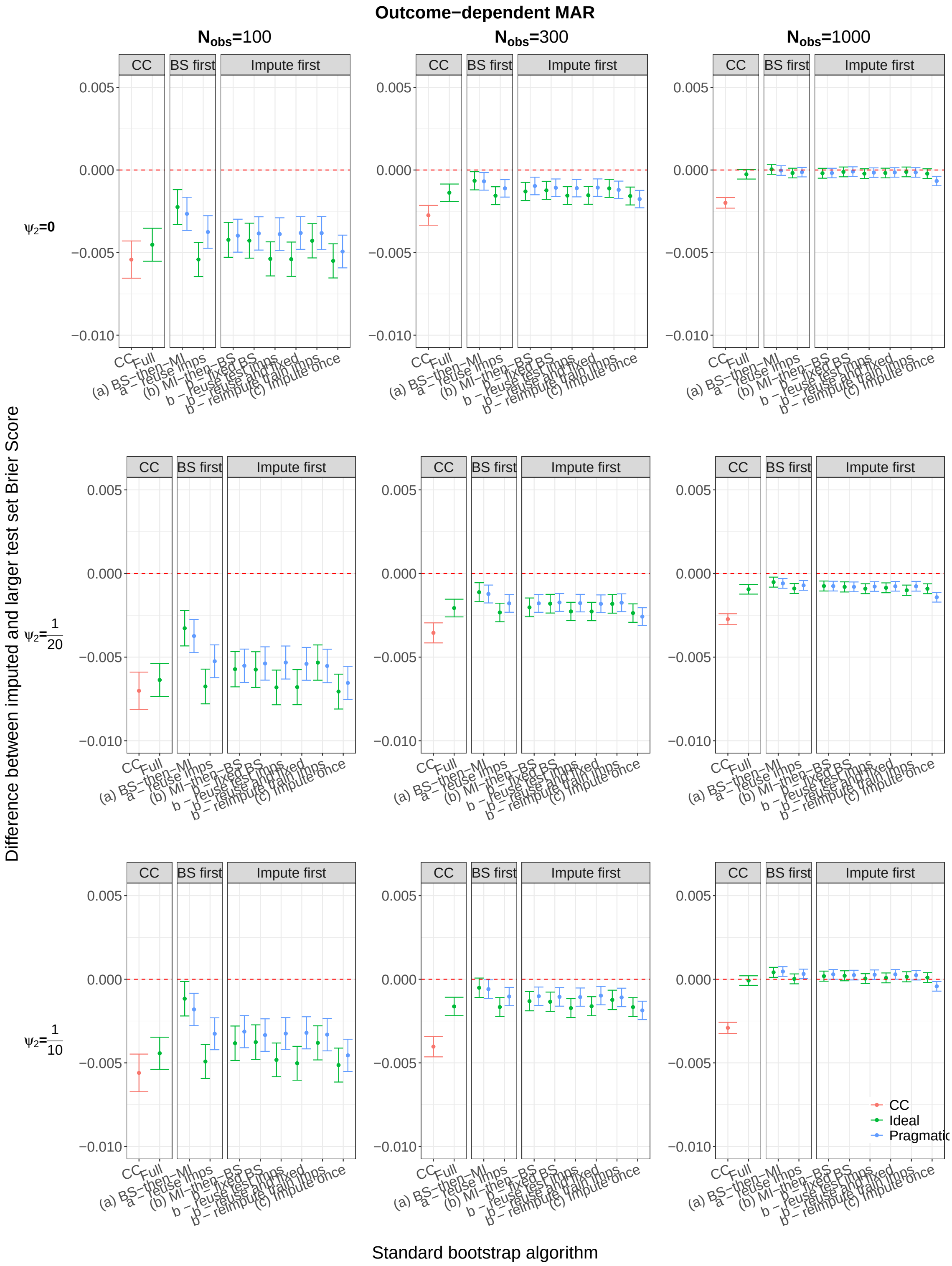

Figure S28: The difference Brier$_{imp}$ - Brier$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.3 The standard bootstrap: Calibration intercept and slope

S4.3.1 Reusing versus re-imputing for test performance of the standard algorithm

Intercept

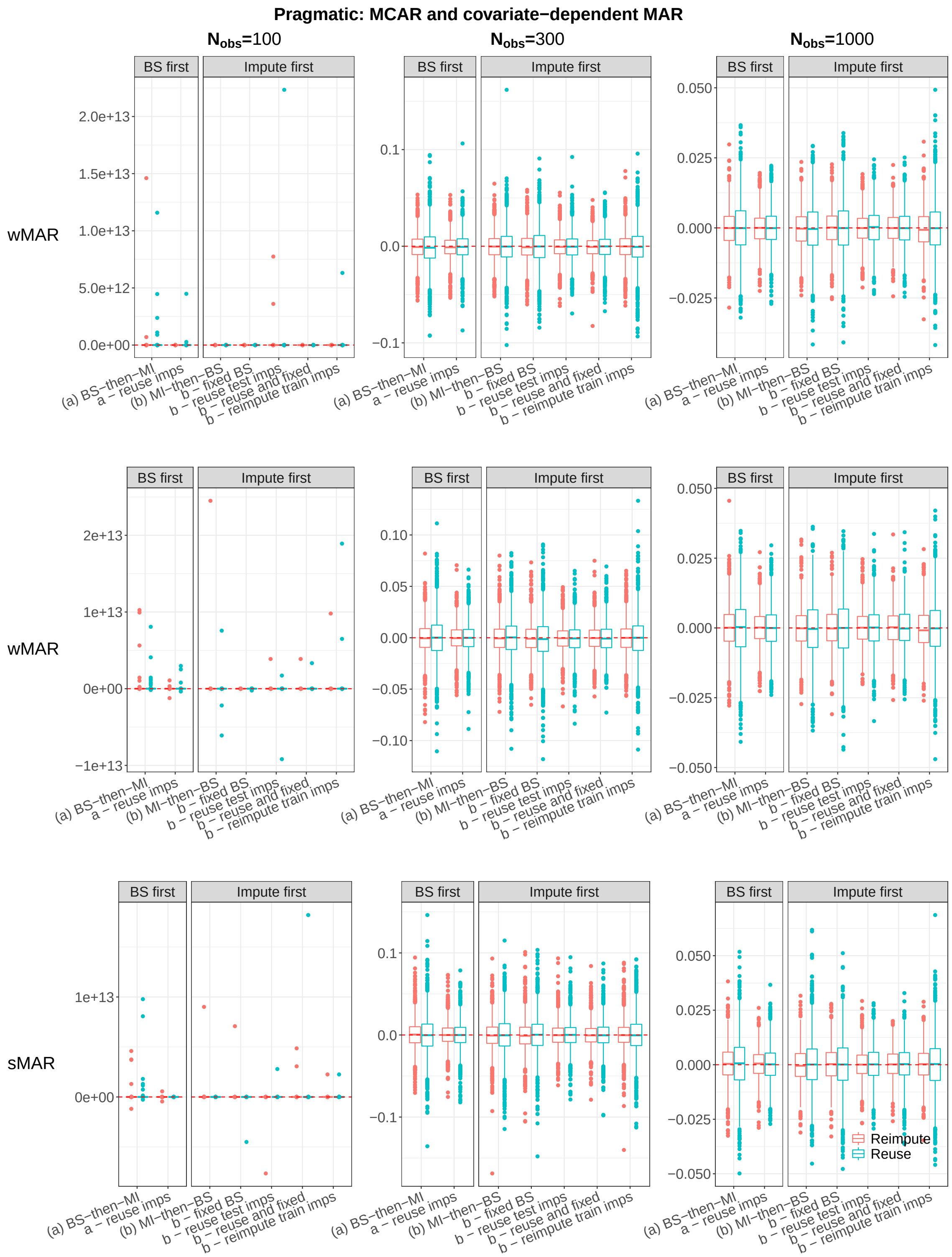

Figure S29: A comparison of reusing versus re-imputing test datasets on the calibration intercept estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for the pragmatic performance each method which are compared to the Brier score estimated when data are fully-observed (Intercept_{imp}- Intercept_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

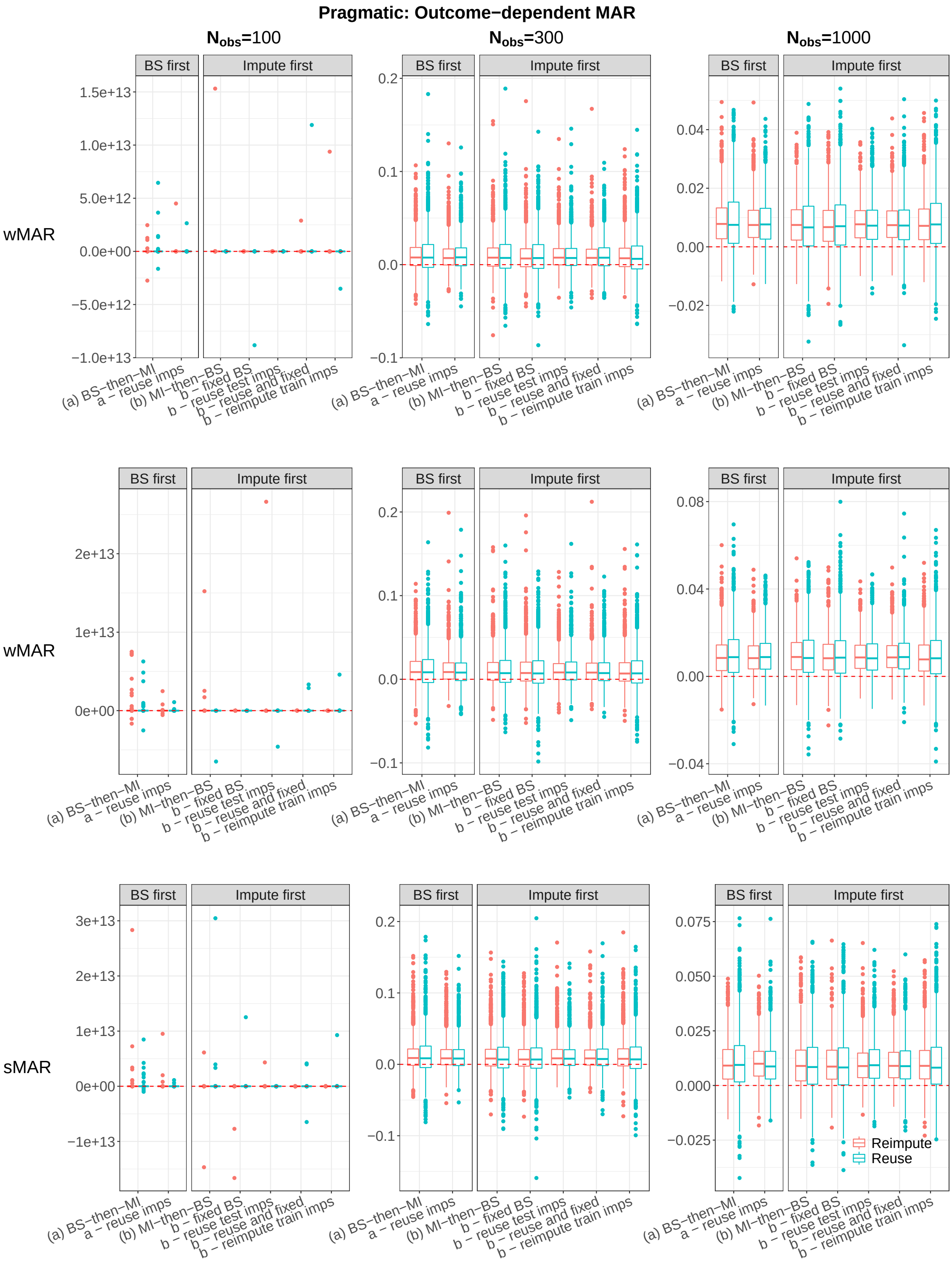

Figure S30: A comparison of reusing versus re-imputing test datasets on the calibration intercept estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for the pragmatic performance of each method which are compared to the Brier score estimated when data are fully-observed (Intercept$_{imp}$- Intercept$_{obs}$). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

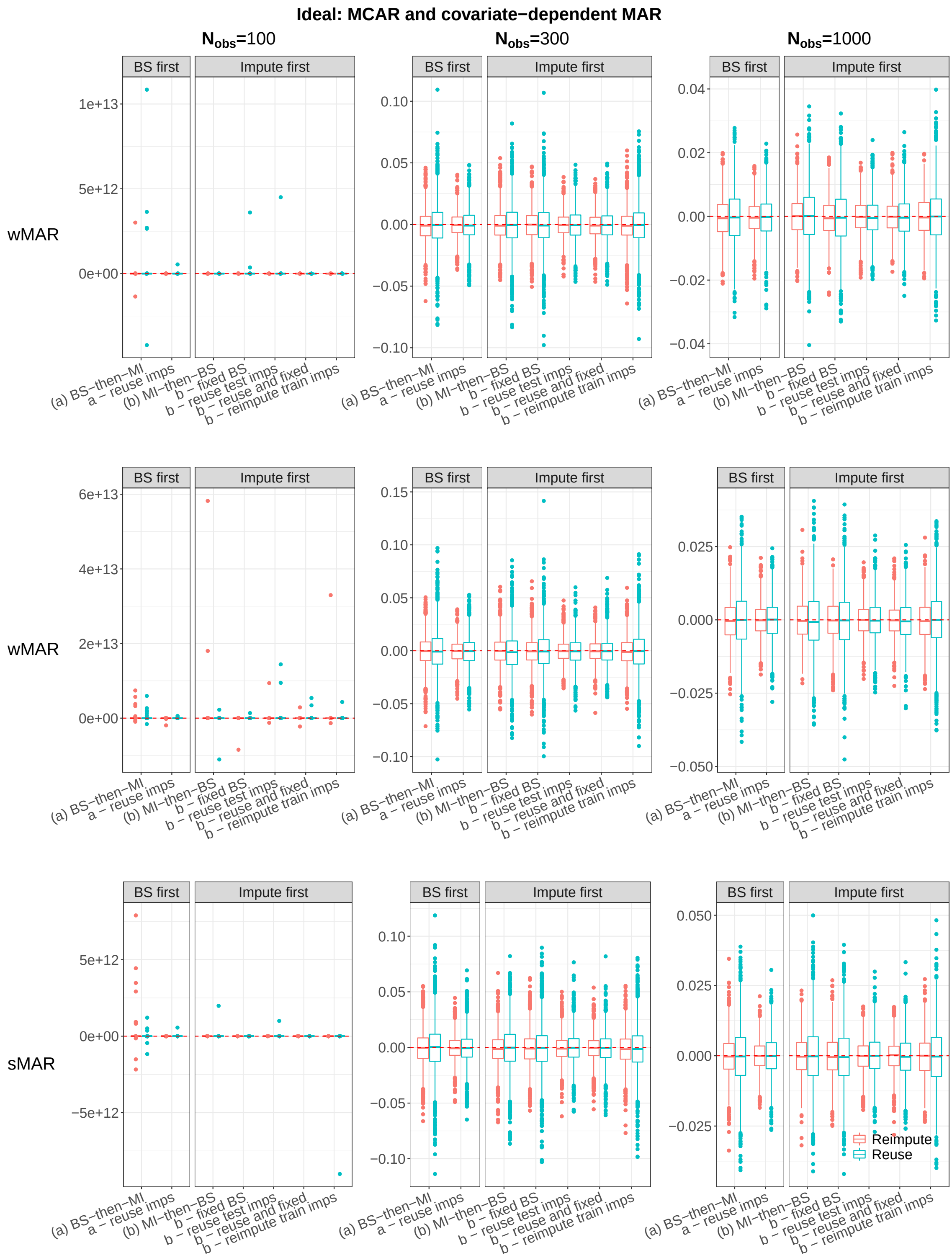

Figure S31: A comparison of reusing versus re-imputing test datasets on the calibration intercept estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for the ideal performance each method which are compared to the Brier score estimated when data are fully-observed (Intercept$_{imp}$- Intercept$_{obs}$). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

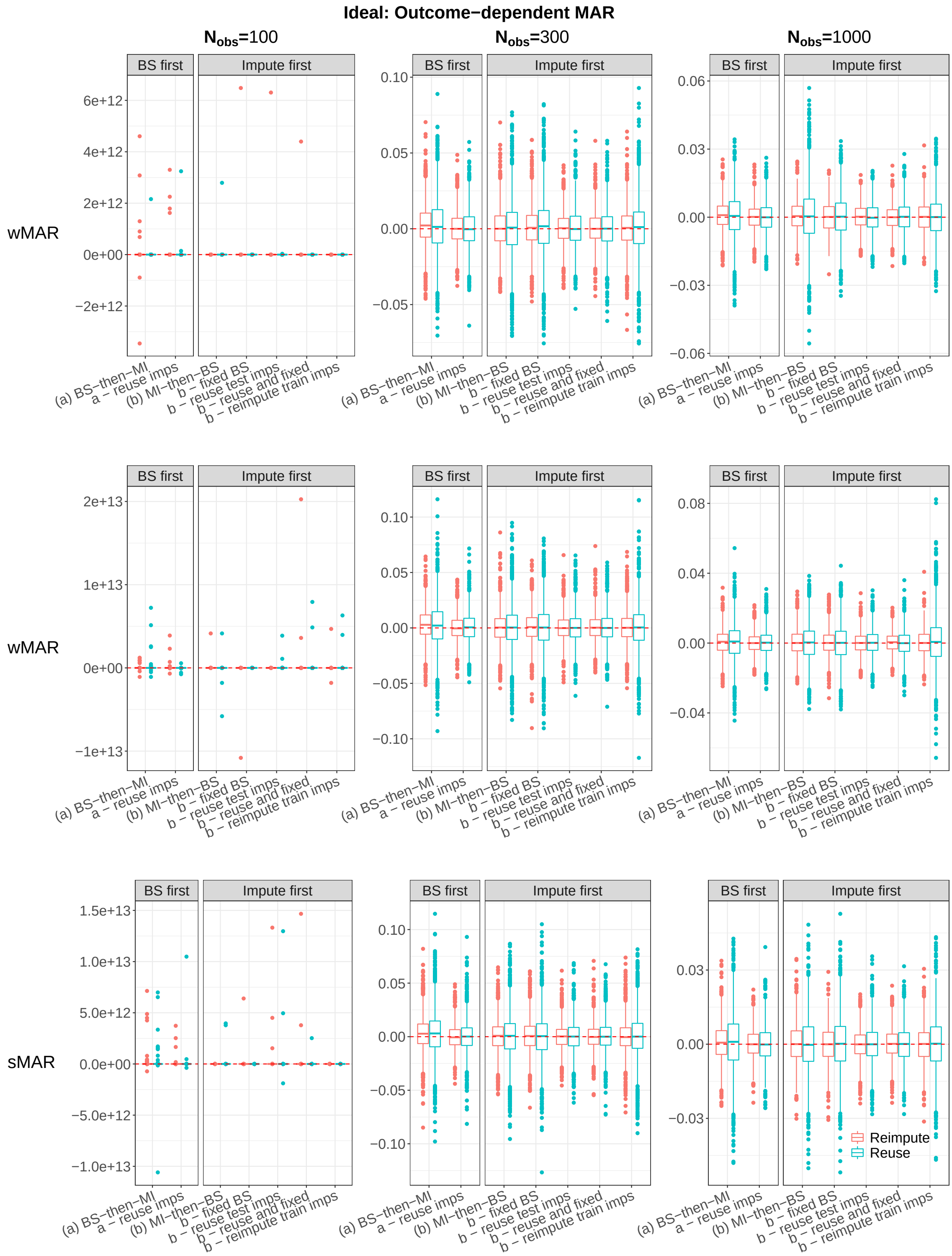

Figure S32: A comparison of reusing versus re-imputing test datasets on the calibration intercept estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for the pragmatic performance of each method which are compared to the Brier score estimated when data are fully-observed (Intercept$_{imp}$- Intercept$_{obs}$). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

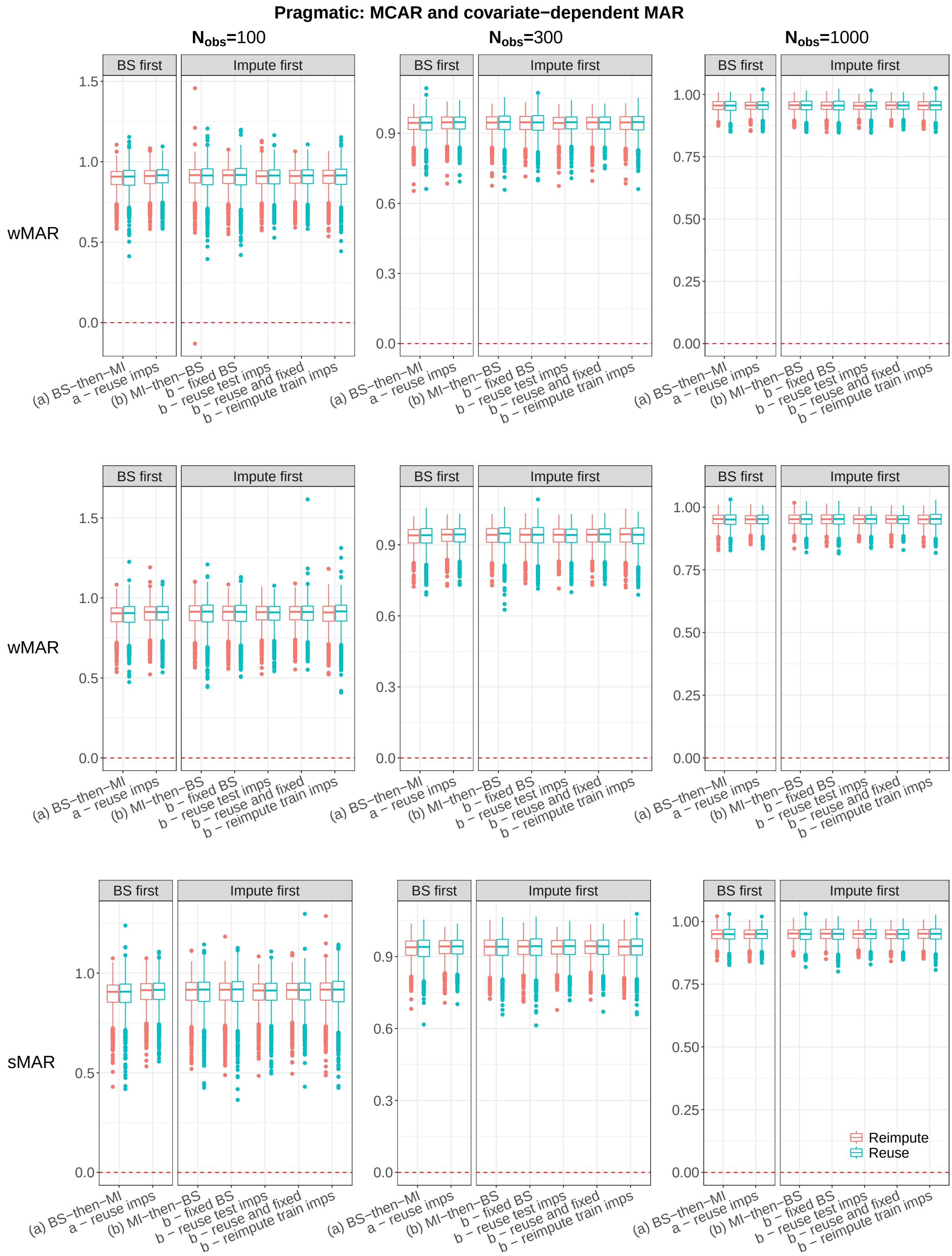

Figure S33: A comparison of reusing versus re-imputing test datasets on the calibration slope estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for the pragmatic performance each method which are compared to the Brier score estimated when data are fully-observed (Slope_{imp}- Slope_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

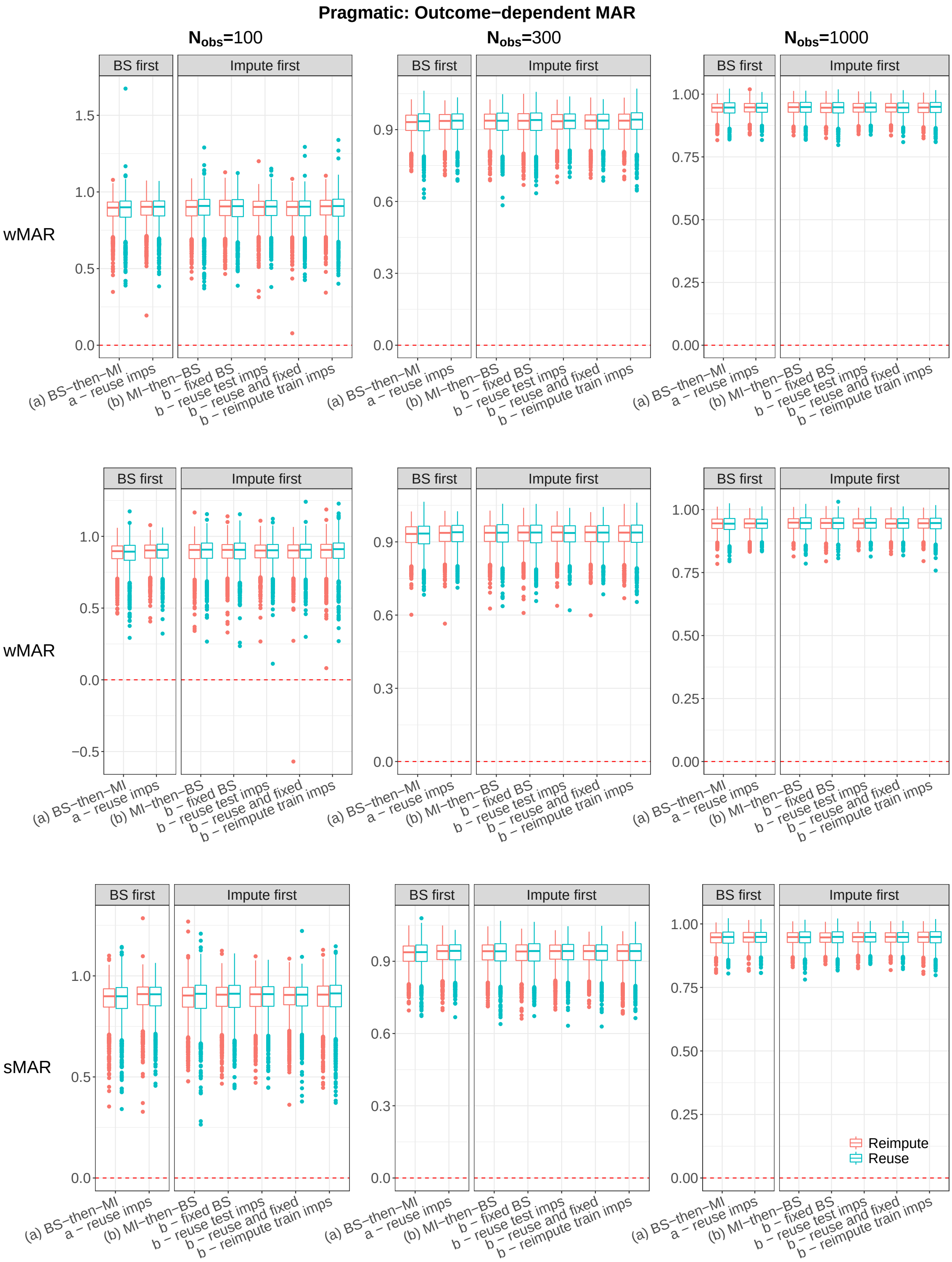

Figure S34: A comparison of reusing versus re-imputing test datasets on the calibration slope estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for the pragmatic performance of each method which are compared to the Brier score estimated when data are fully-observed (Slope_{imp}- Slope_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

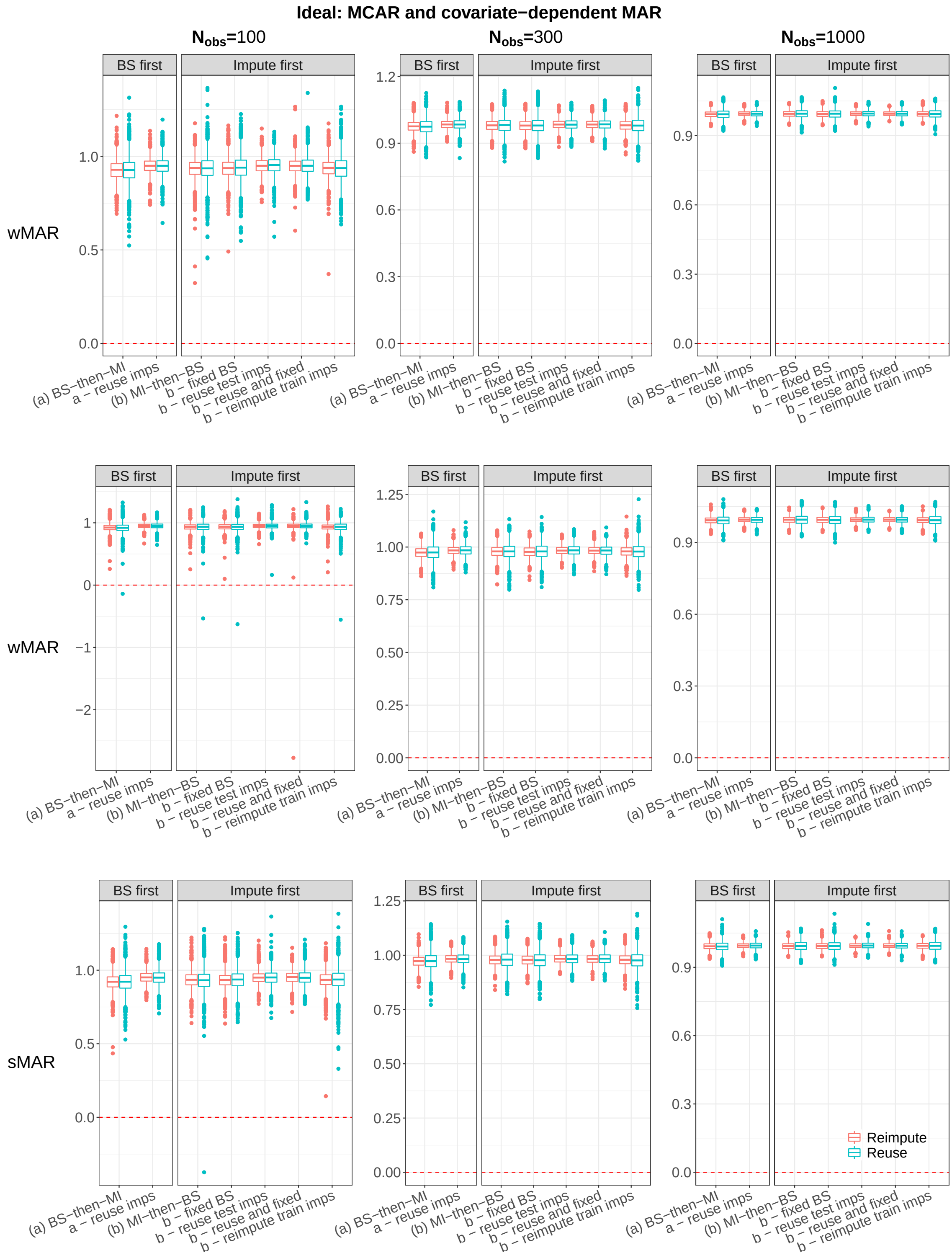

Figure S35: A comparison of reusing versus re-imputing test datasets on the calibration slope estimates for the *standard* bootstrap algorithm when data are MCAR or covariate-dependent MAR. The boxplots display the pragmatic performance estimates of the MSE from 2000 repetitions for the ideal performance each method which are compared to the Brier score estimated when data are fully-observed (Slope$_{imp}$- Slope$_{obs}$). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

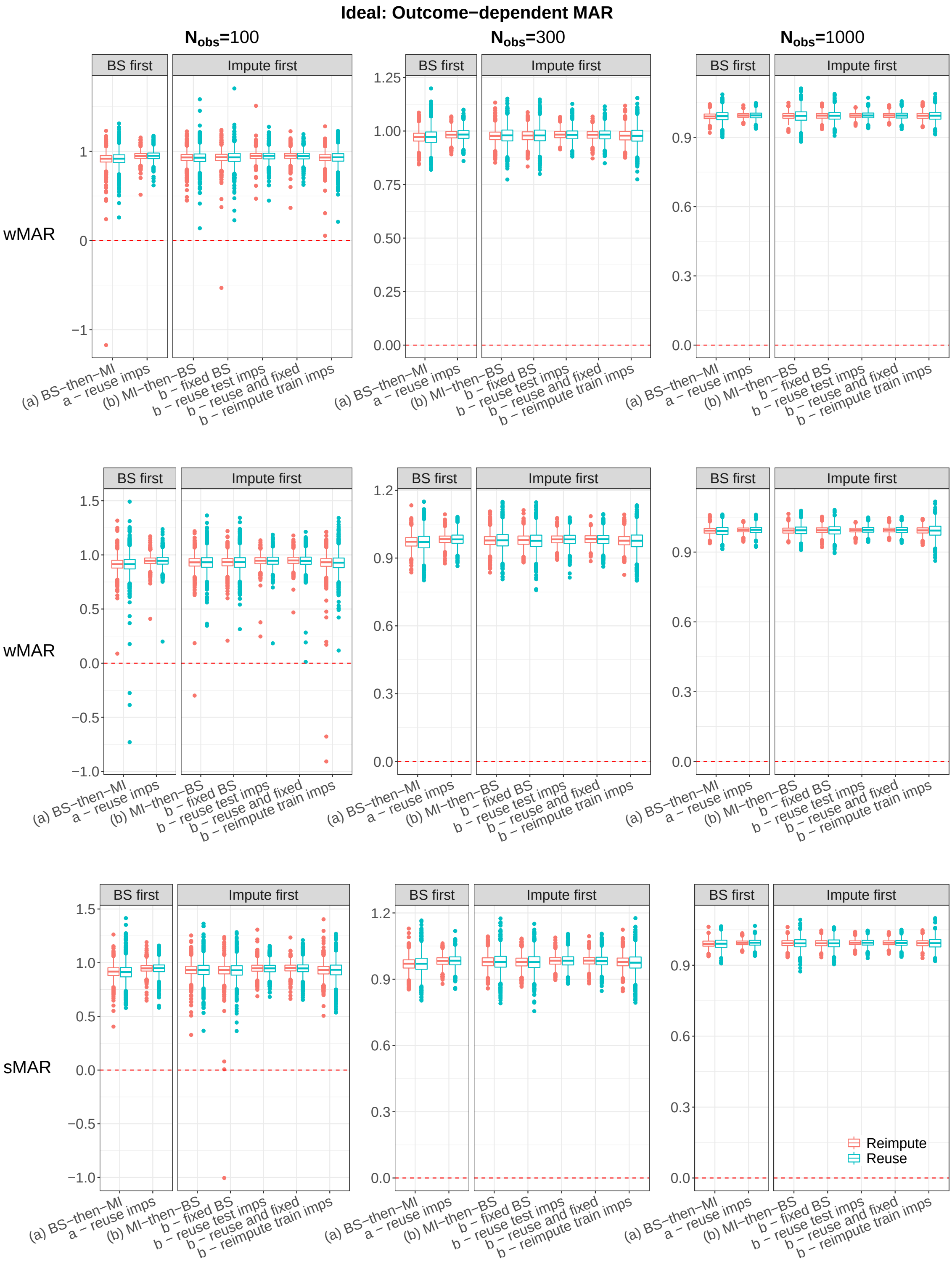

Figure S36: A comparison of reusing versus re-imputing test datasets on the calibration slope estimates for the *standard* bootstrap algorithm when data are outcome-dependent or outcome- and covariate-dependent MAR. The boxplots display the ideal performance estimates of the MSE from 2000 repetitions for the pragmatic performance of each method which are compared to the Brier score estimated when data are fully-observed (Slope_{imp}- Slope_{obs}). CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.3.2 Calibration intercept and slope from imputation methods compared to the fully-observed Calibration intercept and slope (Cal$_{imp}$-Cal$_{obs}$)

Intercept

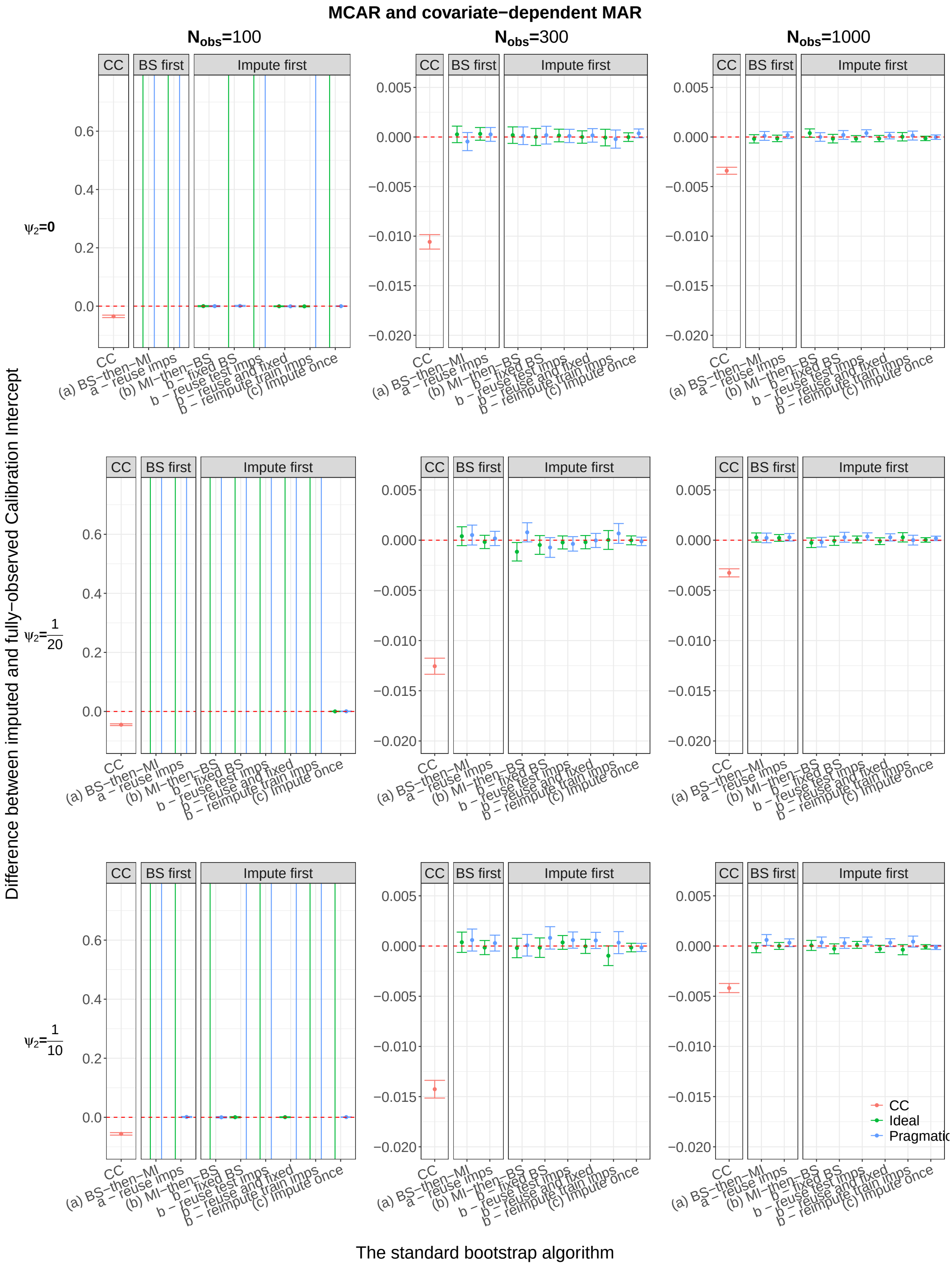

Figure S37: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

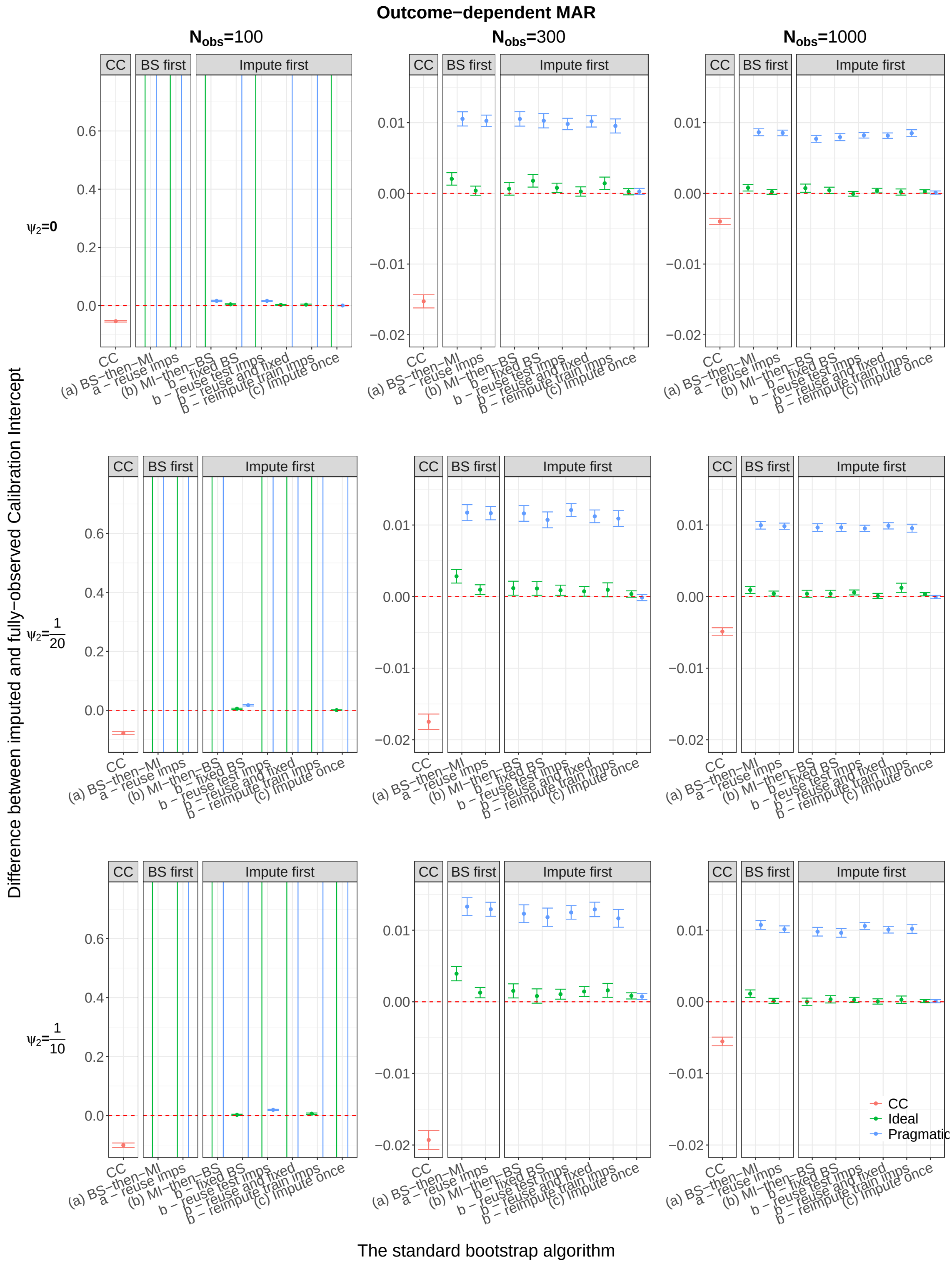

Figure S38: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

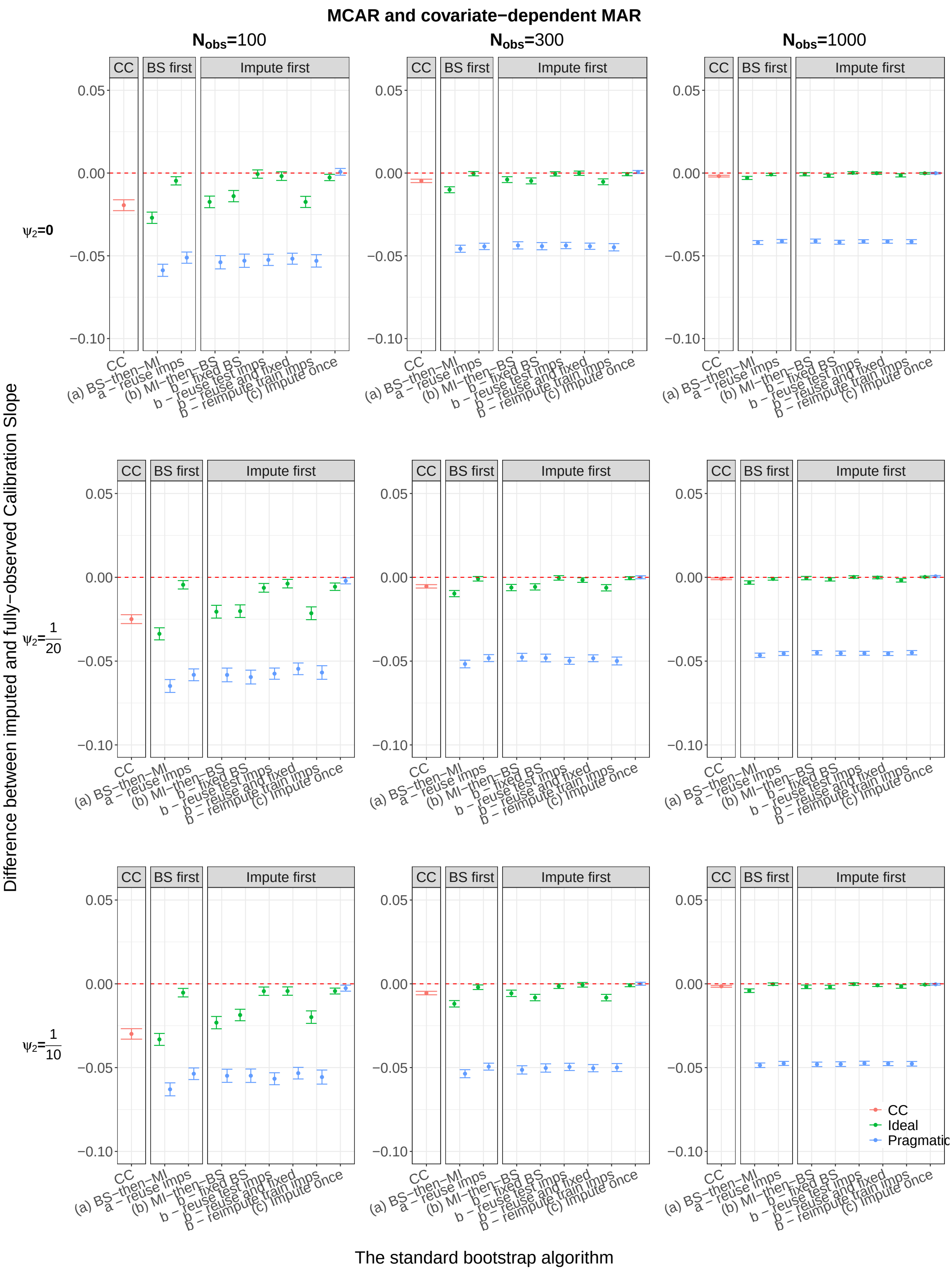

Figure S39: The difference Slope_{imp} - Slope_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

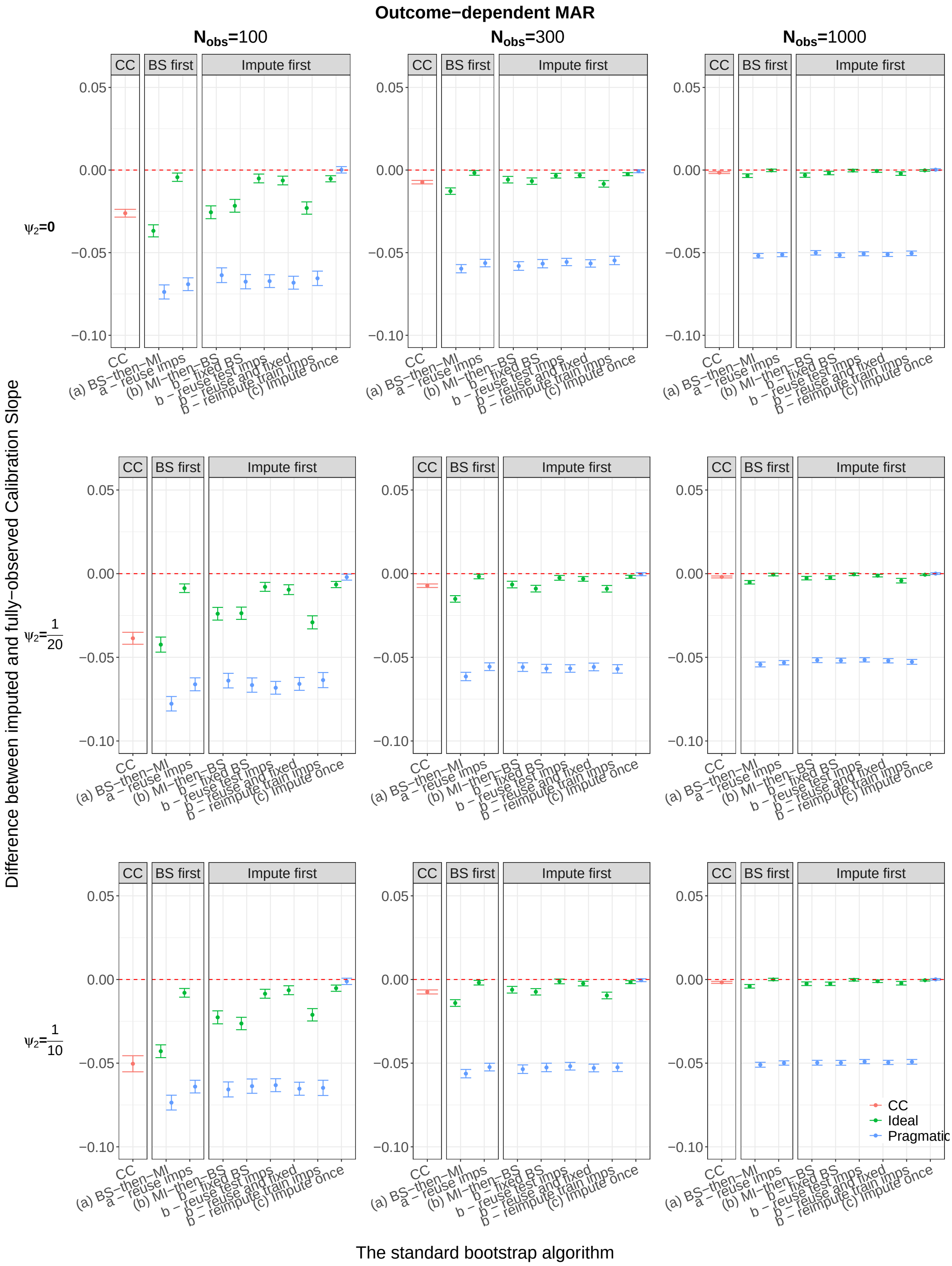

Figure S40: The difference Slope_{imp} - Slope_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.3.3 The proportion of missingness is 40% ($\mathbf{Cal}_{imp}$-$\mathbf{Cal}_{obs}$)

Intercept

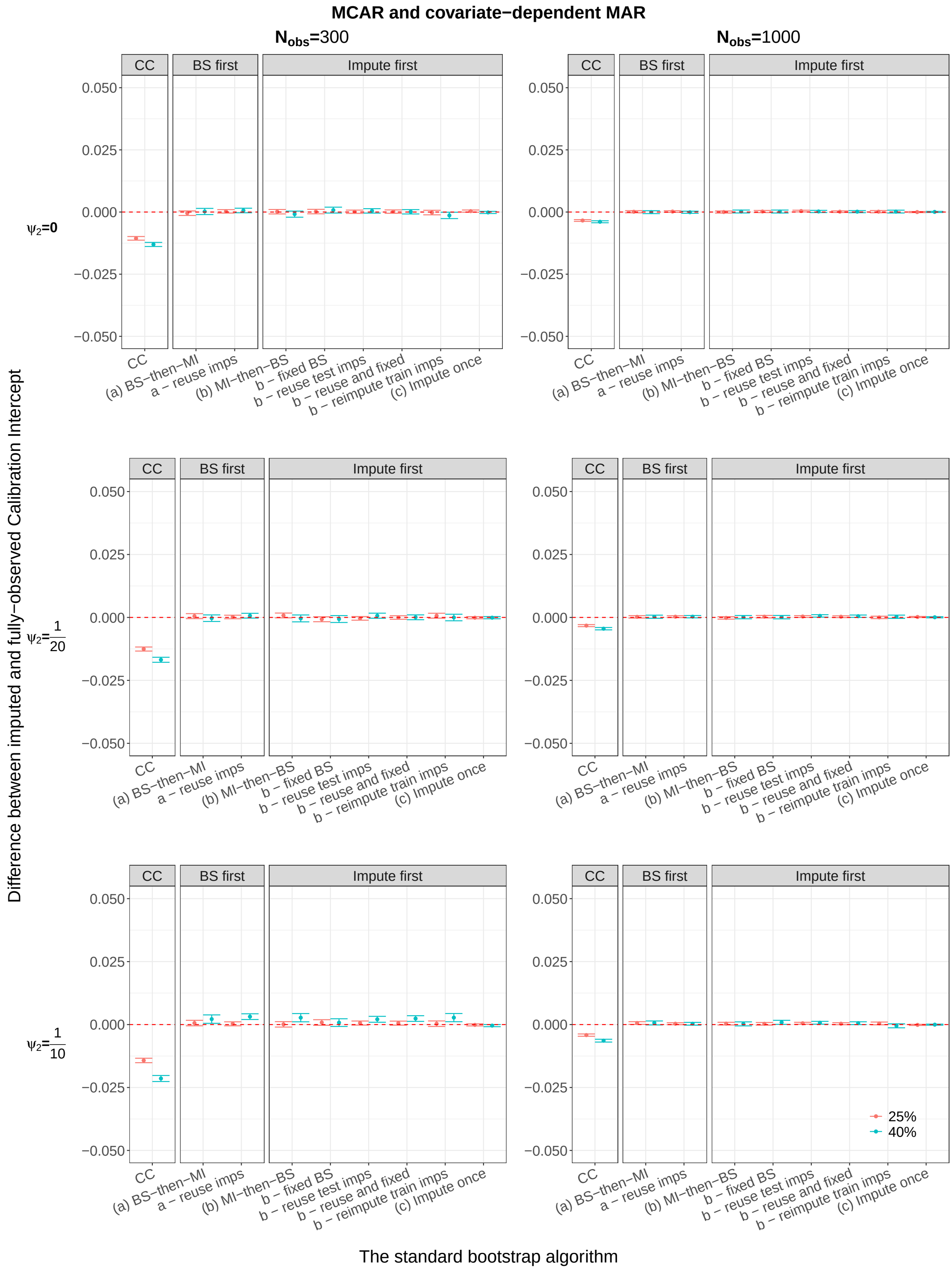

Figure S41: Comparing the impact of increasing the percentage of missingness on the difference Intercept_{imp} - Intercept_{obs} when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept_{imp} - Intercept_{obs}. Red denotes Intercept_{imp} - Intercept_{obs} when 25% of X_1 values are missing and blue denotes Intercept_{imp} - Intercept_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

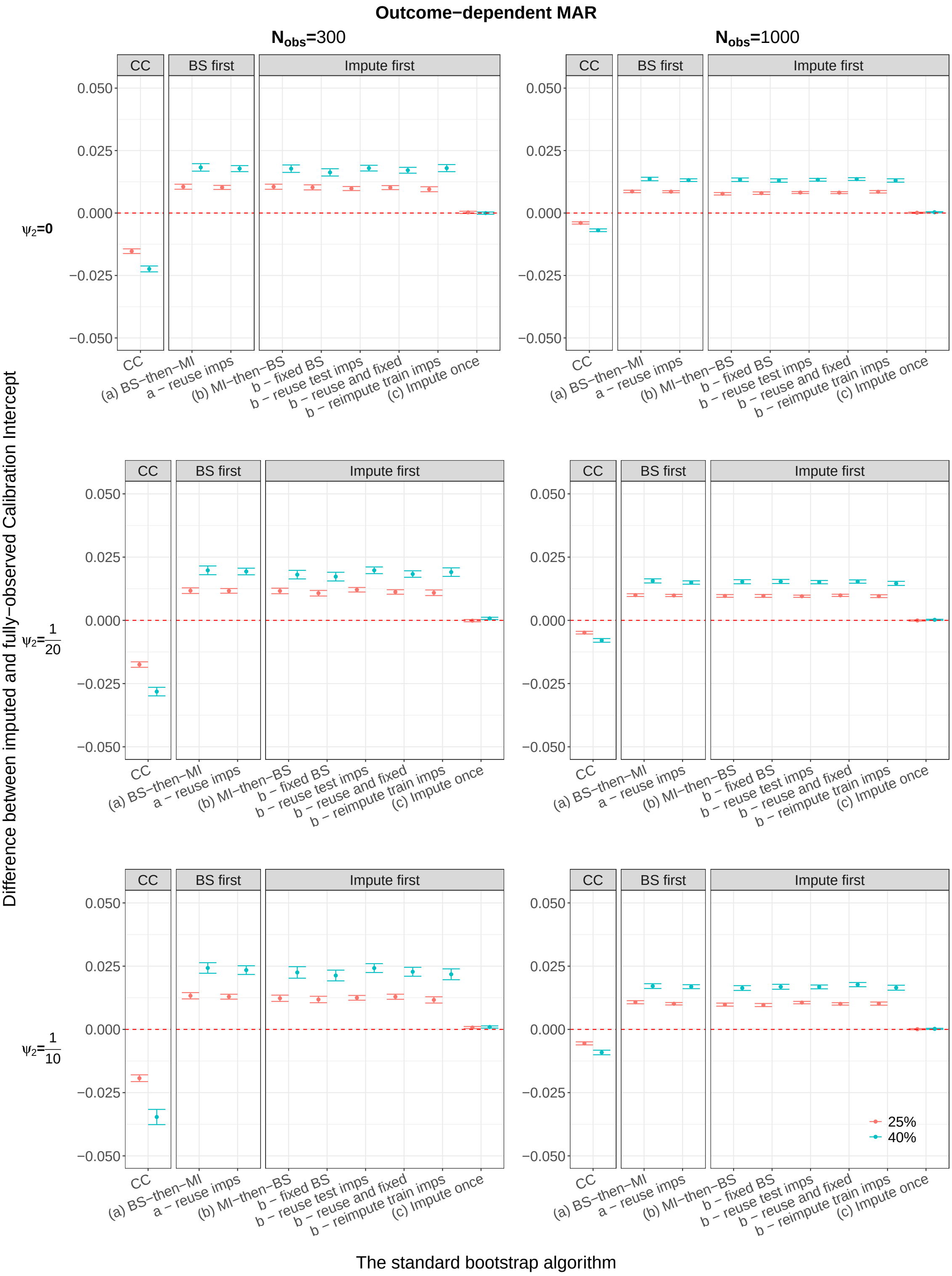

Figure S42: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

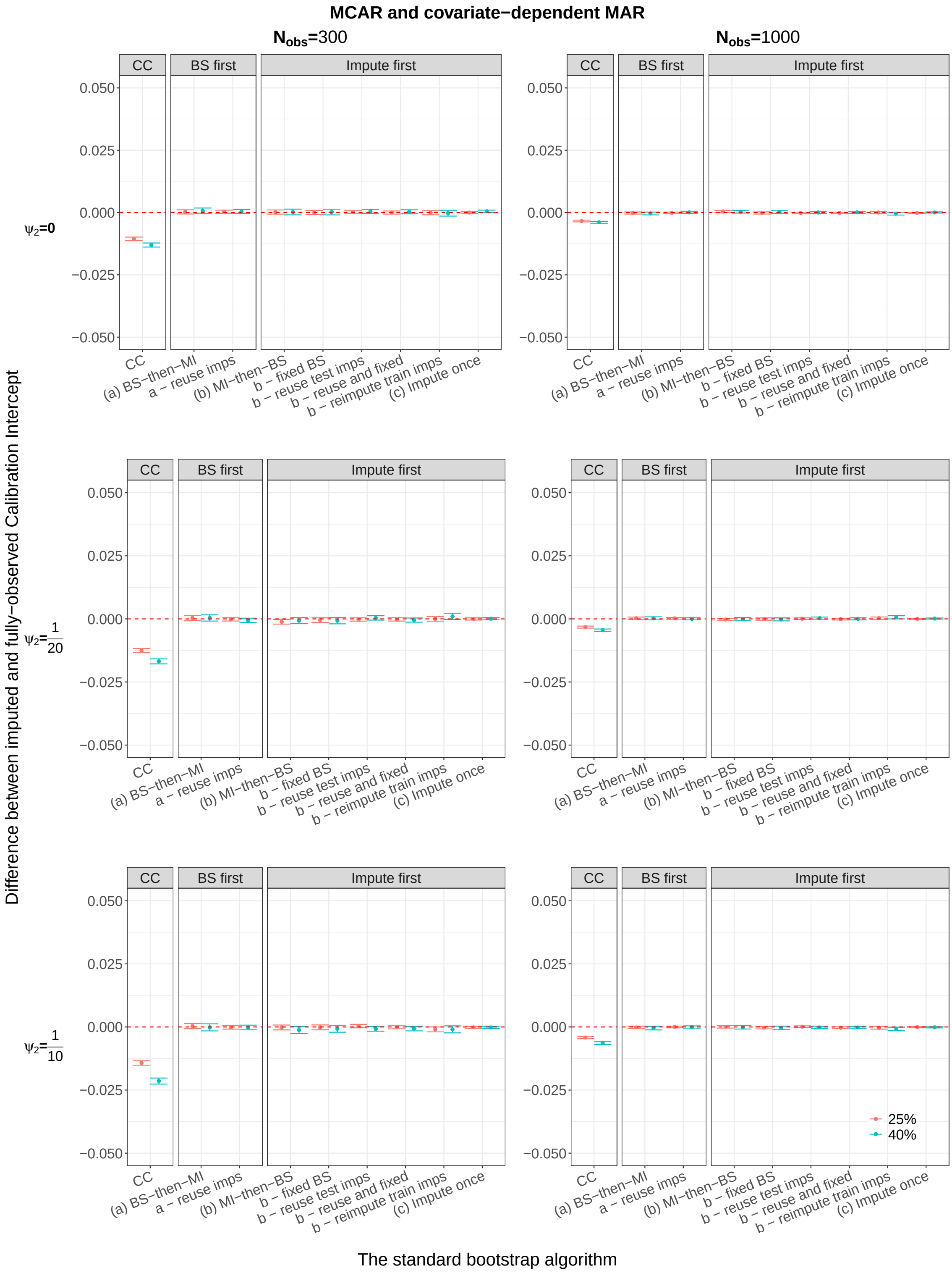

Figure S43: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

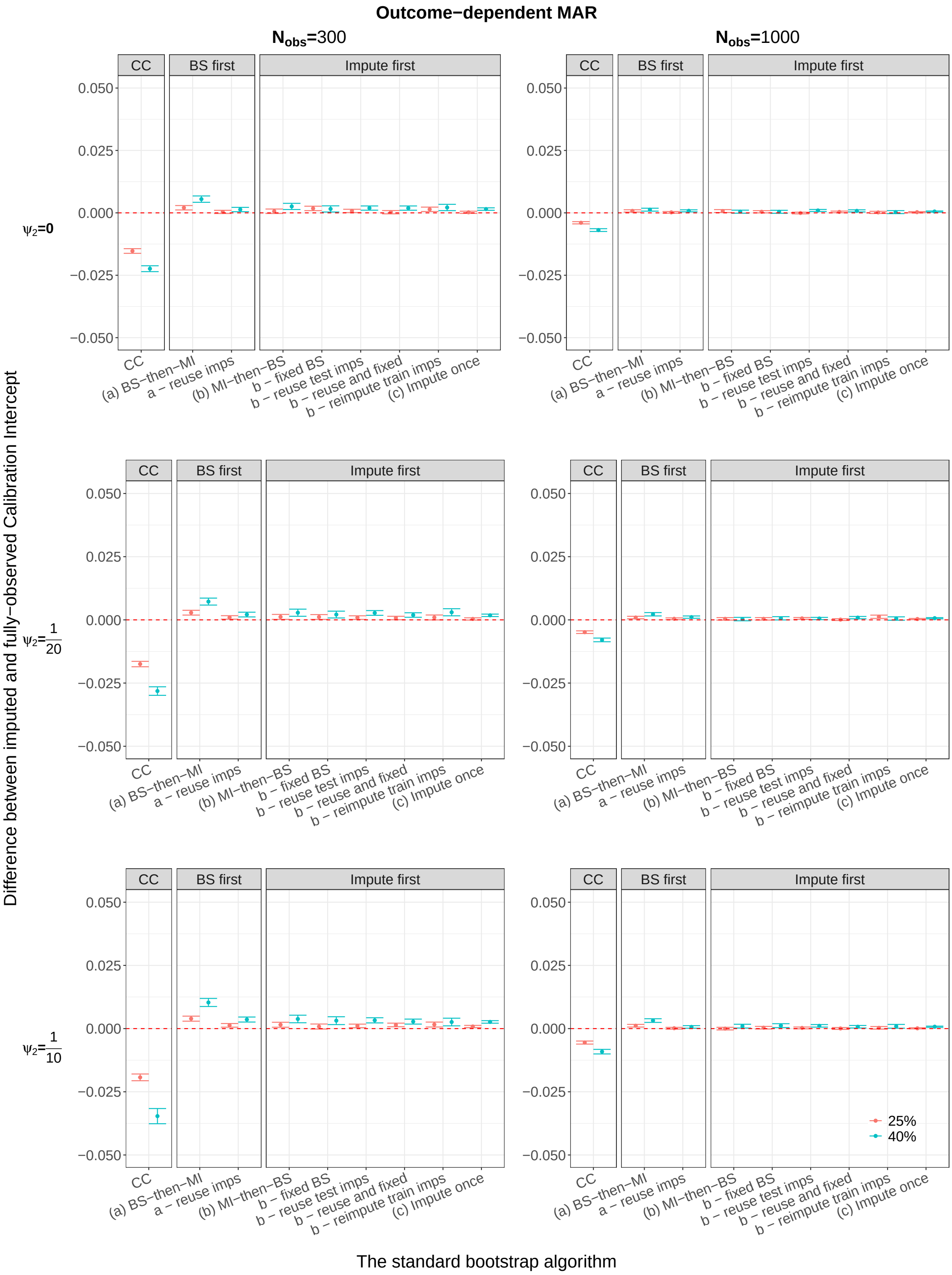

Figure S44: Comparing the impact of increasing the percentage of missingness on the difference Intercept_{imp} - Intercept_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept_{imp} - Intercept_{obs}. Red denotes Intercept_{imp} - Intercept_{obs} when 25% of X_1 values are missing and blue denotes Intercept_{imp} - Intercept_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

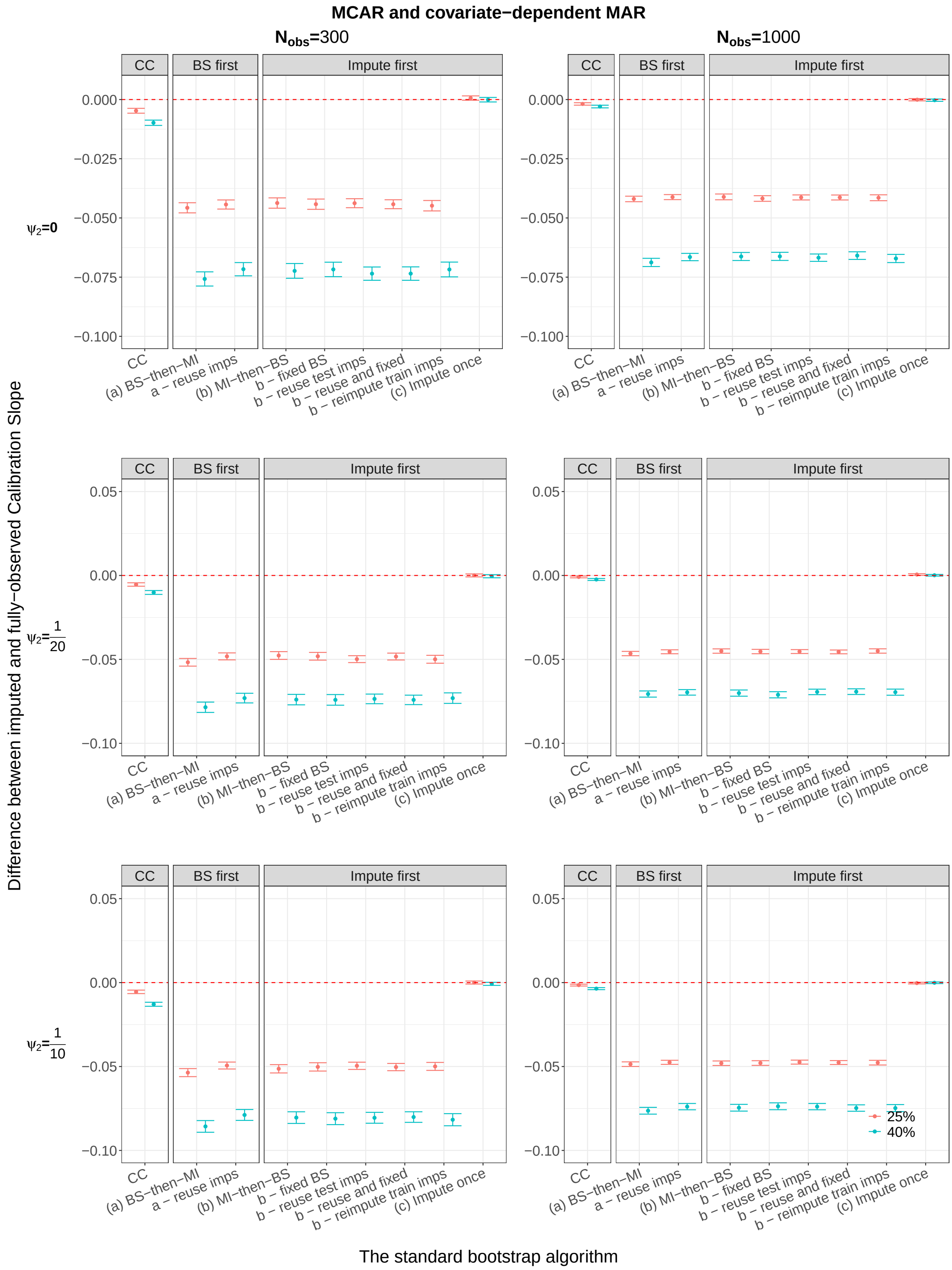

Figure S45: Comparing the impact of increasing the percentage of missingness on the difference Slope_{imp} - Slope_{obs} when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. Red denotes Slope_{imp} - Slope_{obs} when 25% of X_1 values are missing and blue denotes Slope_{imp} - Slope_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

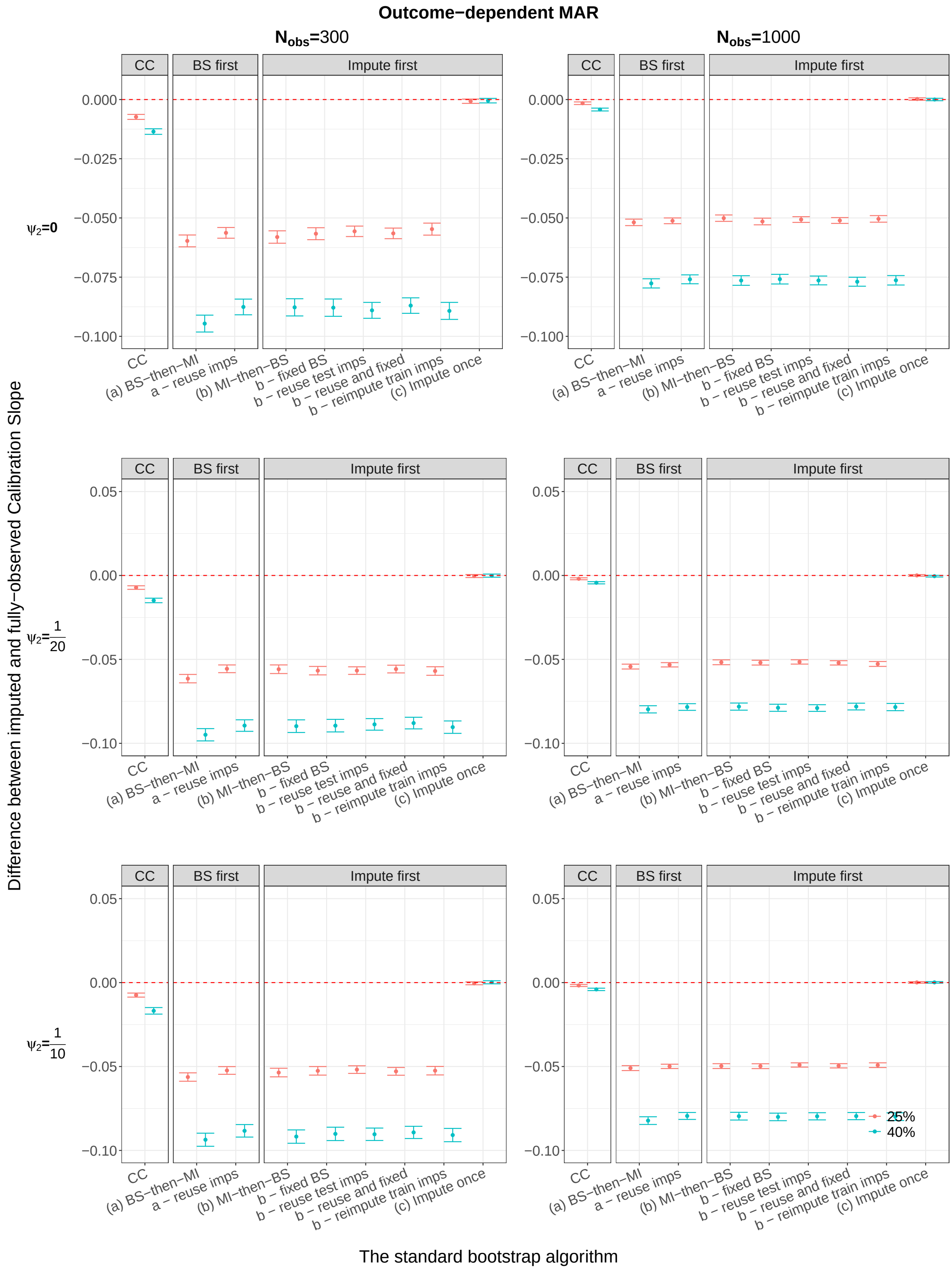

Figure S46: Comparing the impact of increasing the percentage of missingness on the difference Slope_{imp} - Slope_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. Red denotes Slope_{imp} - Slope_{obs} when 25% of X_1 values are missing and blue denotes Slope_{imp} - Slope_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

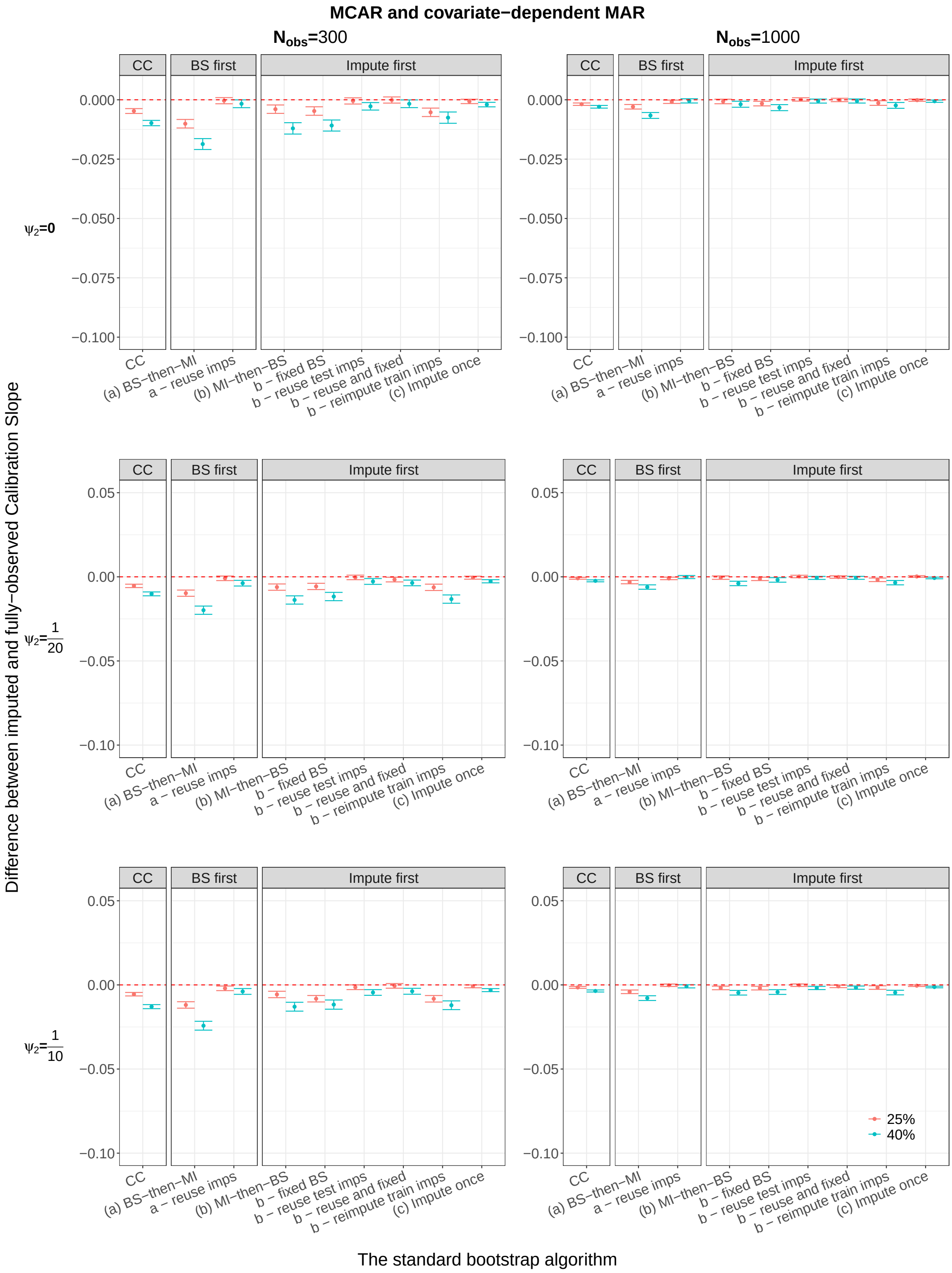

Figure S47: Comparing the impact of increasing the percentage of missingness on the difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. Red denotes Slope$_{imp}$ - Slope$_{obs}$ when 25% of X_1 values are missing and blue denotes Slope$_{imp}$ - Slope$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

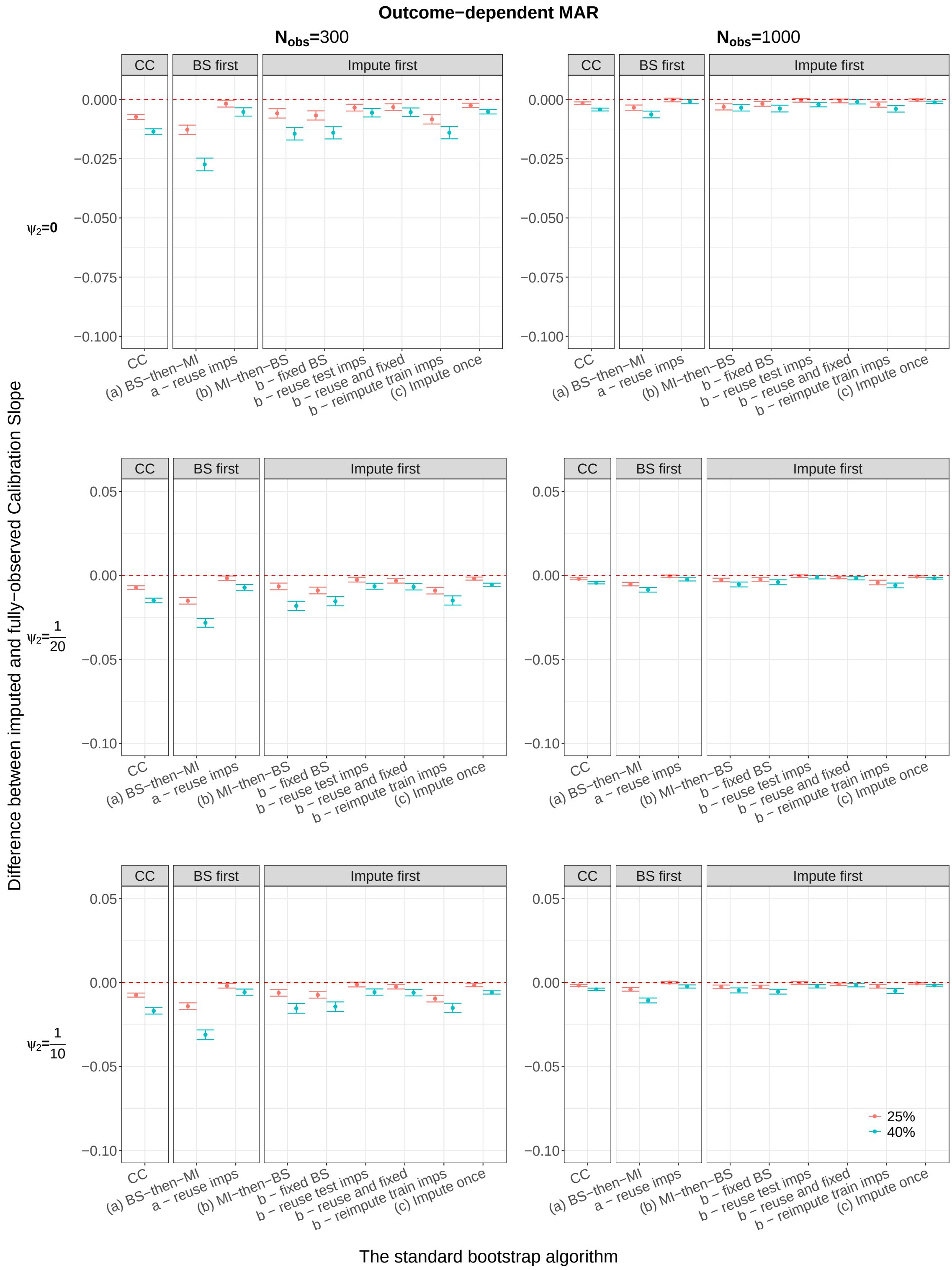

Figure S48: Comparing the impact of increasing the percentage of missingness on the difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. Red denotes Slope$_{imp}$ - Slope$_{obs}$ when 25% of X_1 values are missing and blue denotes Slope$_{imp}$ - Slope$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.3.4 Comparing M=5 versus M=25 (Cal_{imp}-Cal_{obs})

Intercept

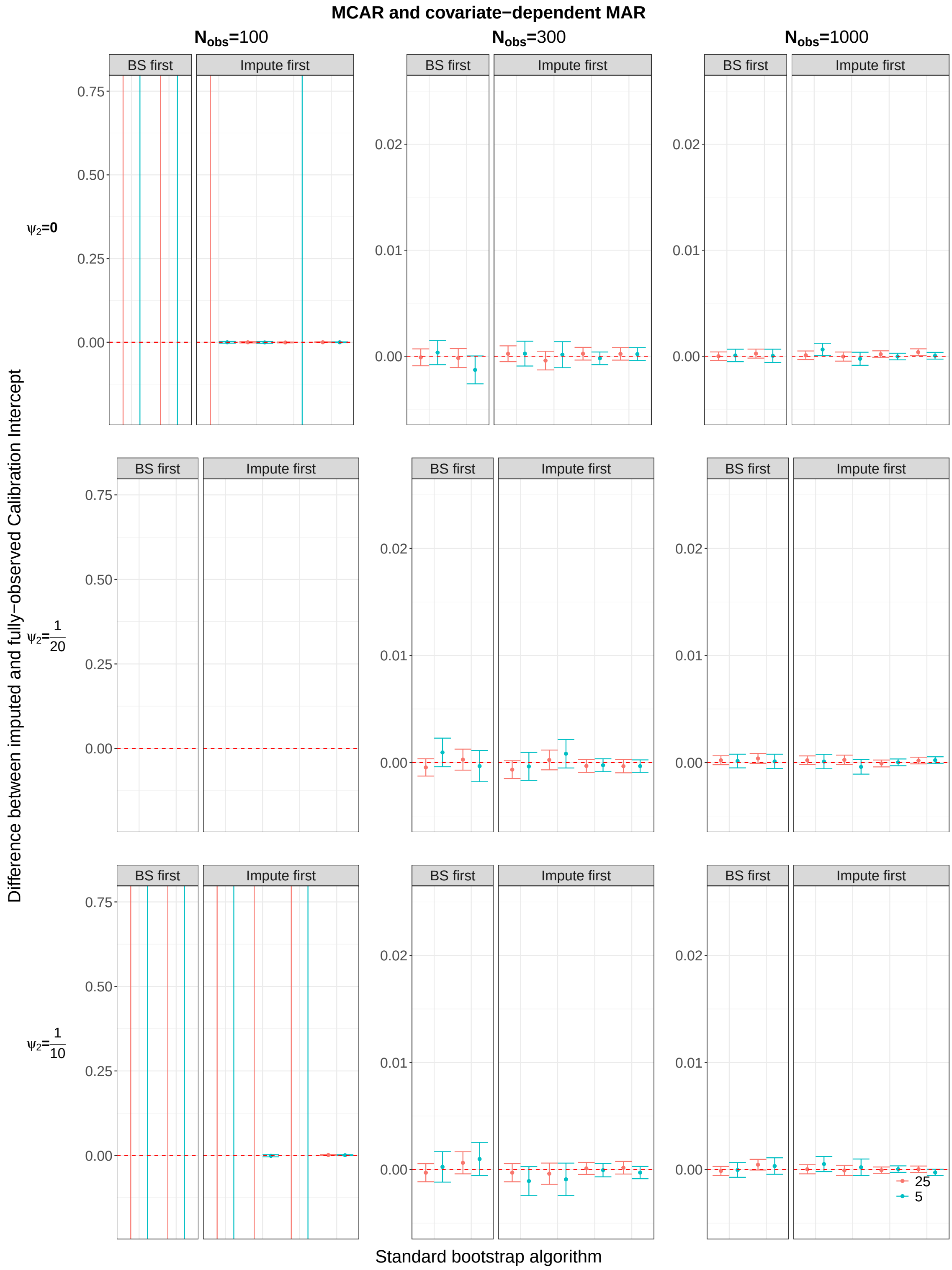

Figure S49: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

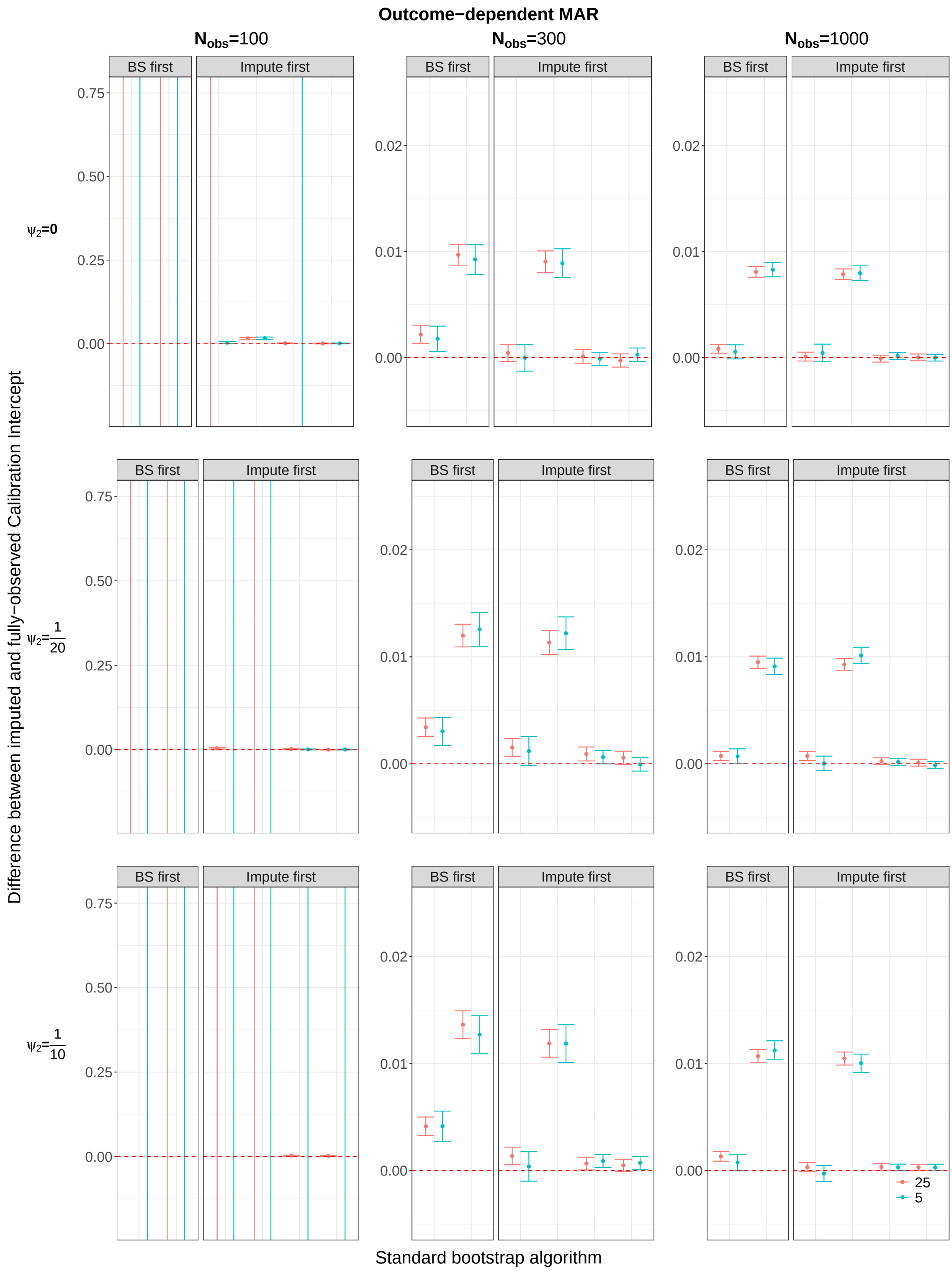

Figure S50: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

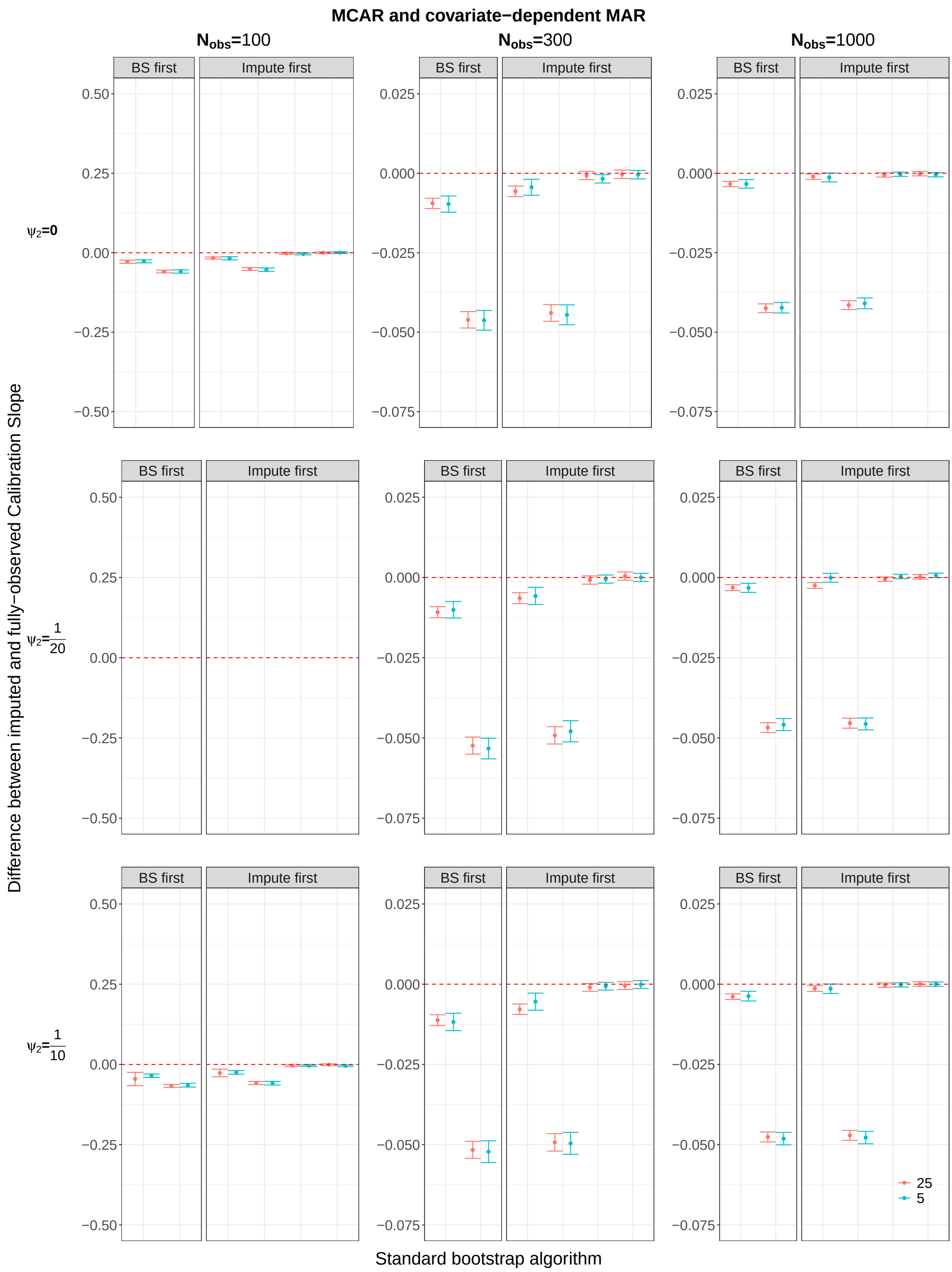

Figure S51: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

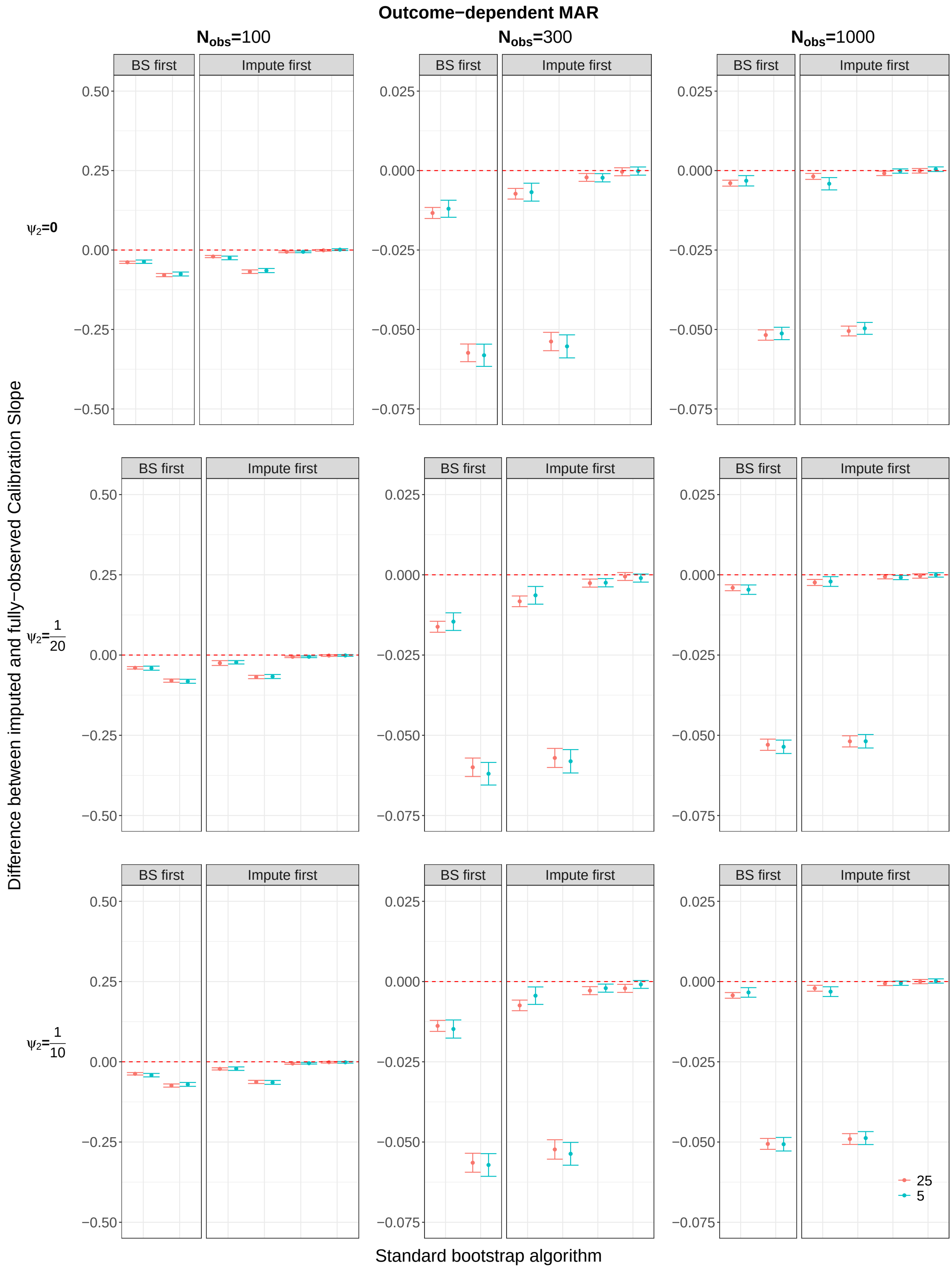

Figure S52: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.3.5 Calibration intercept and slope from imputation methods compared to the target Calibration intercept and slope (Cal$_{target}$) using a larger validation set

Intercept

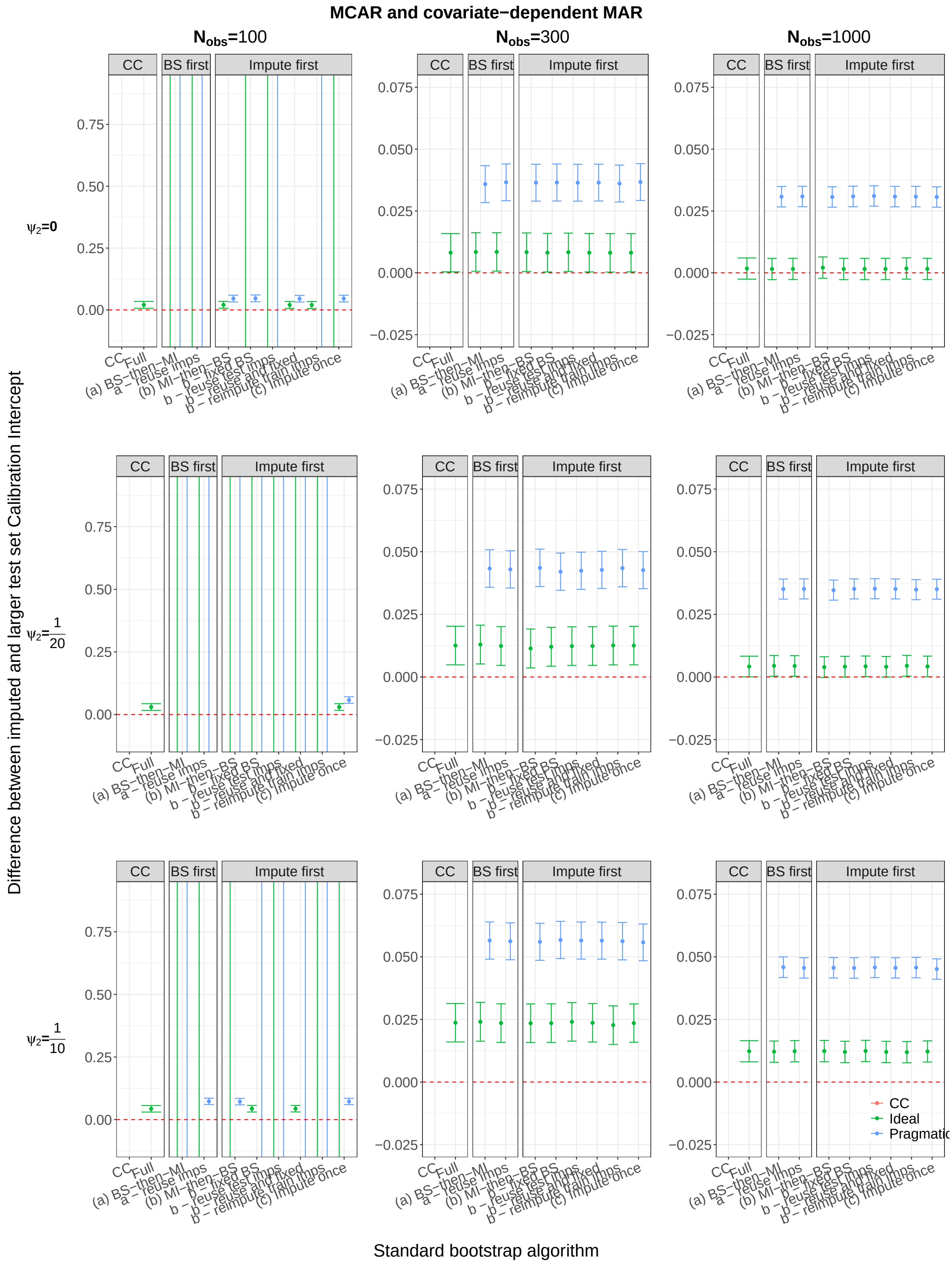

Figure S53: The difference Intercept$_{imp}$ - Intercept$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

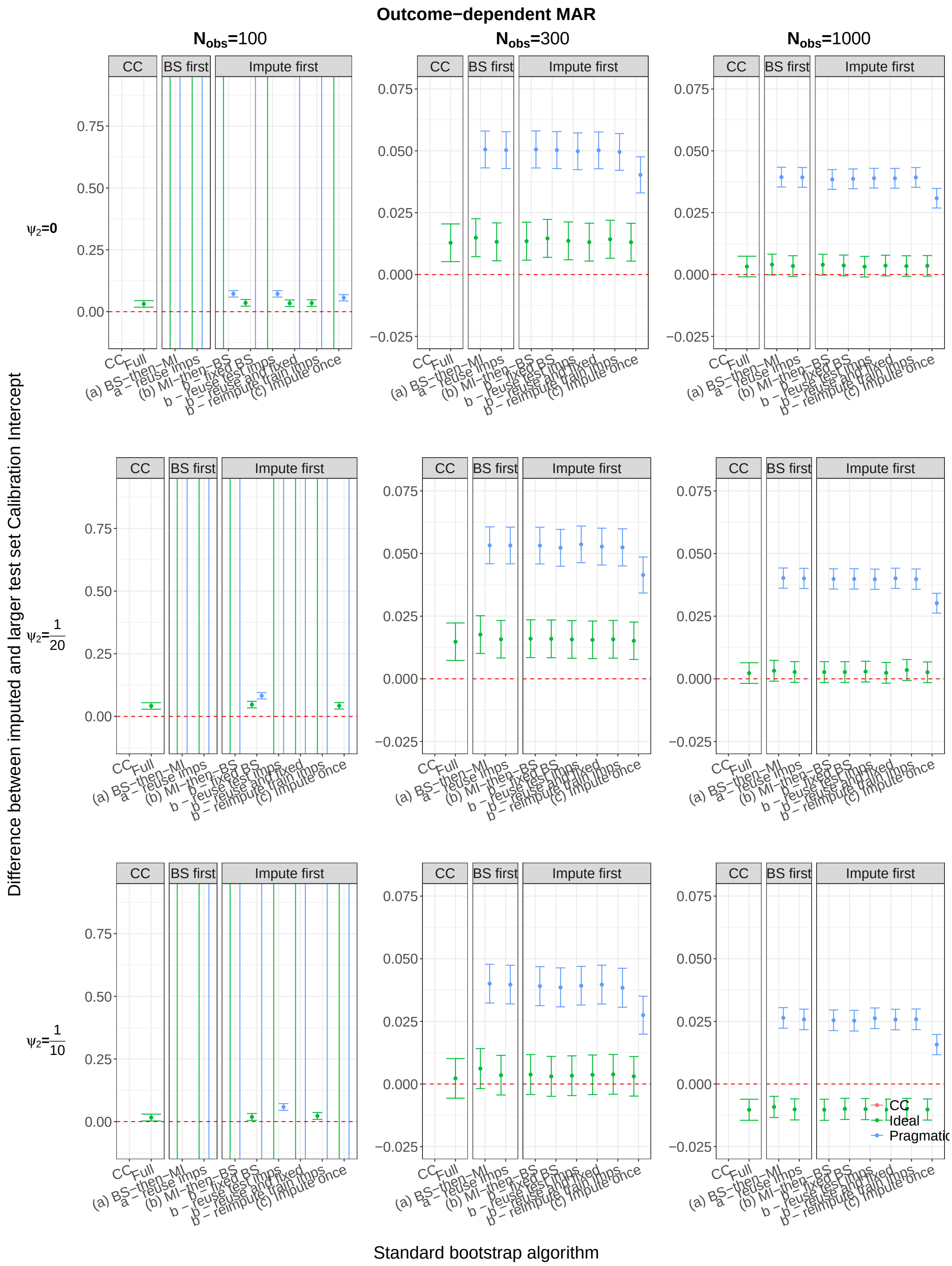

Figure S54: The difference Intercept$_{imp}$ - Intercept$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

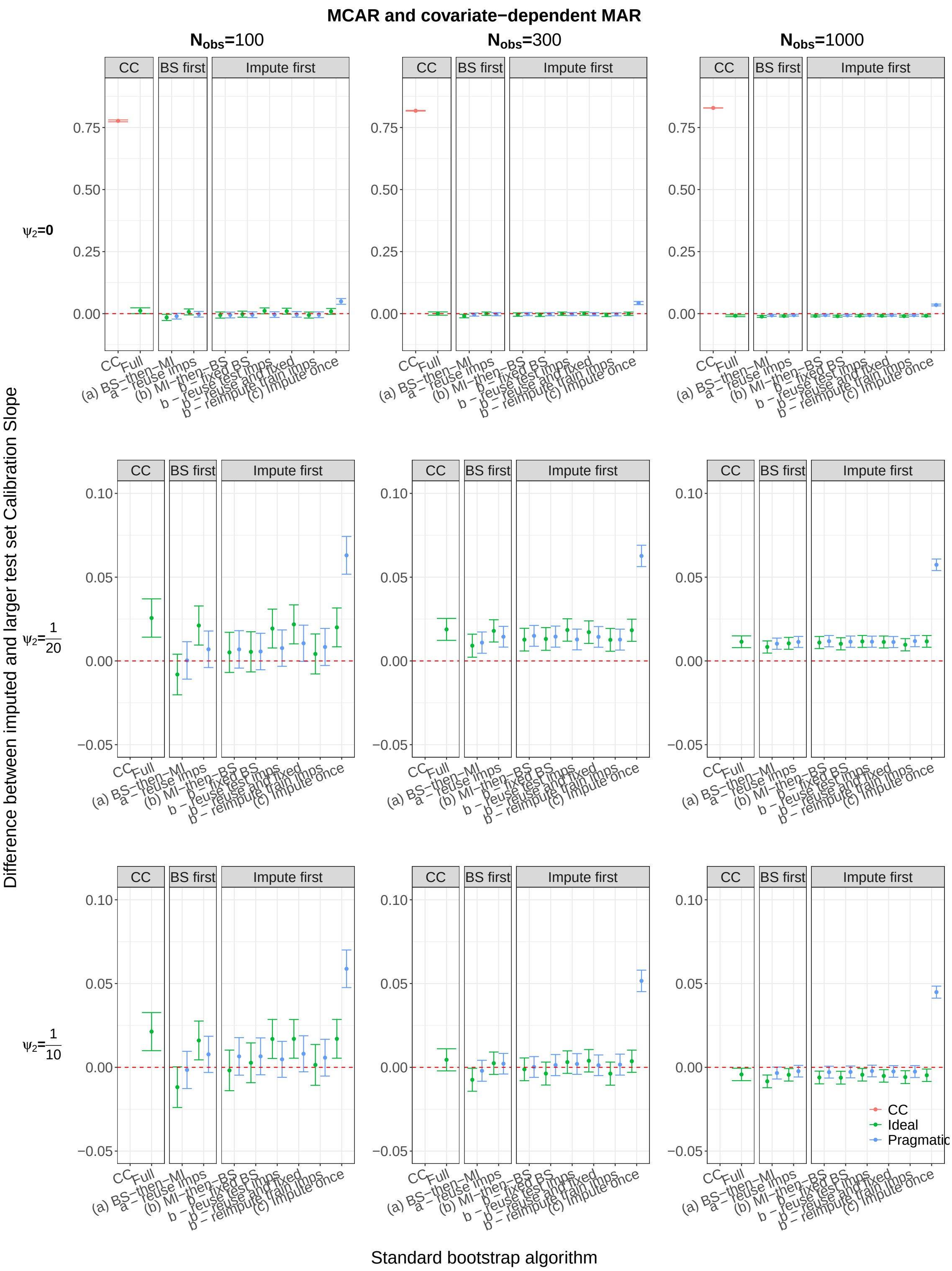

Figure S55: The difference Slope$_{imp}$ - Slope$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

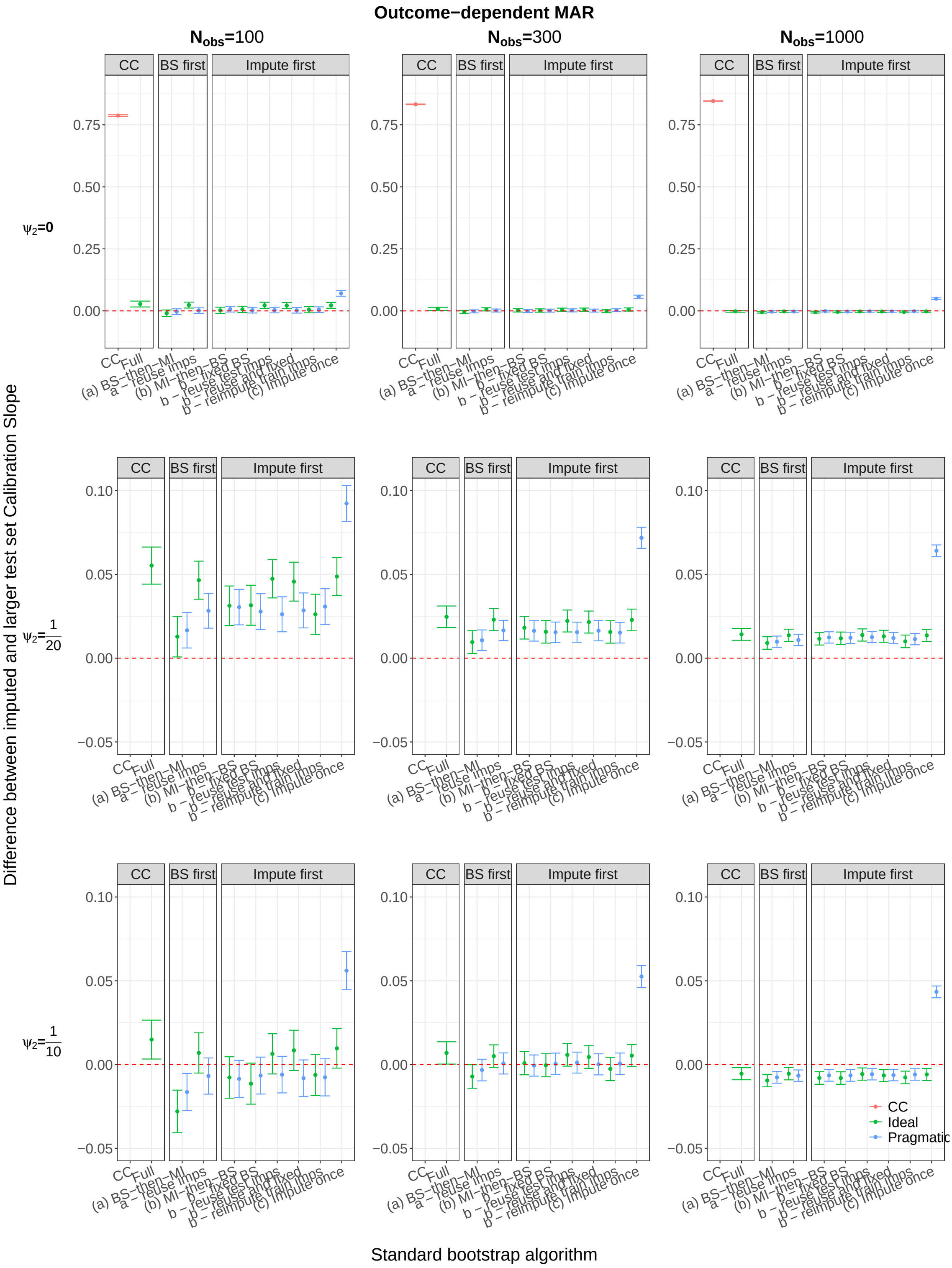

Figure S56: The difference Slope$_{imp}$ - Slope$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.4 The 0.632 bootstrap: AUC

S4.4.1 AUC from imputation methods compared to the fully-observed AUC (AUC_{imp}-AUC_{obs})

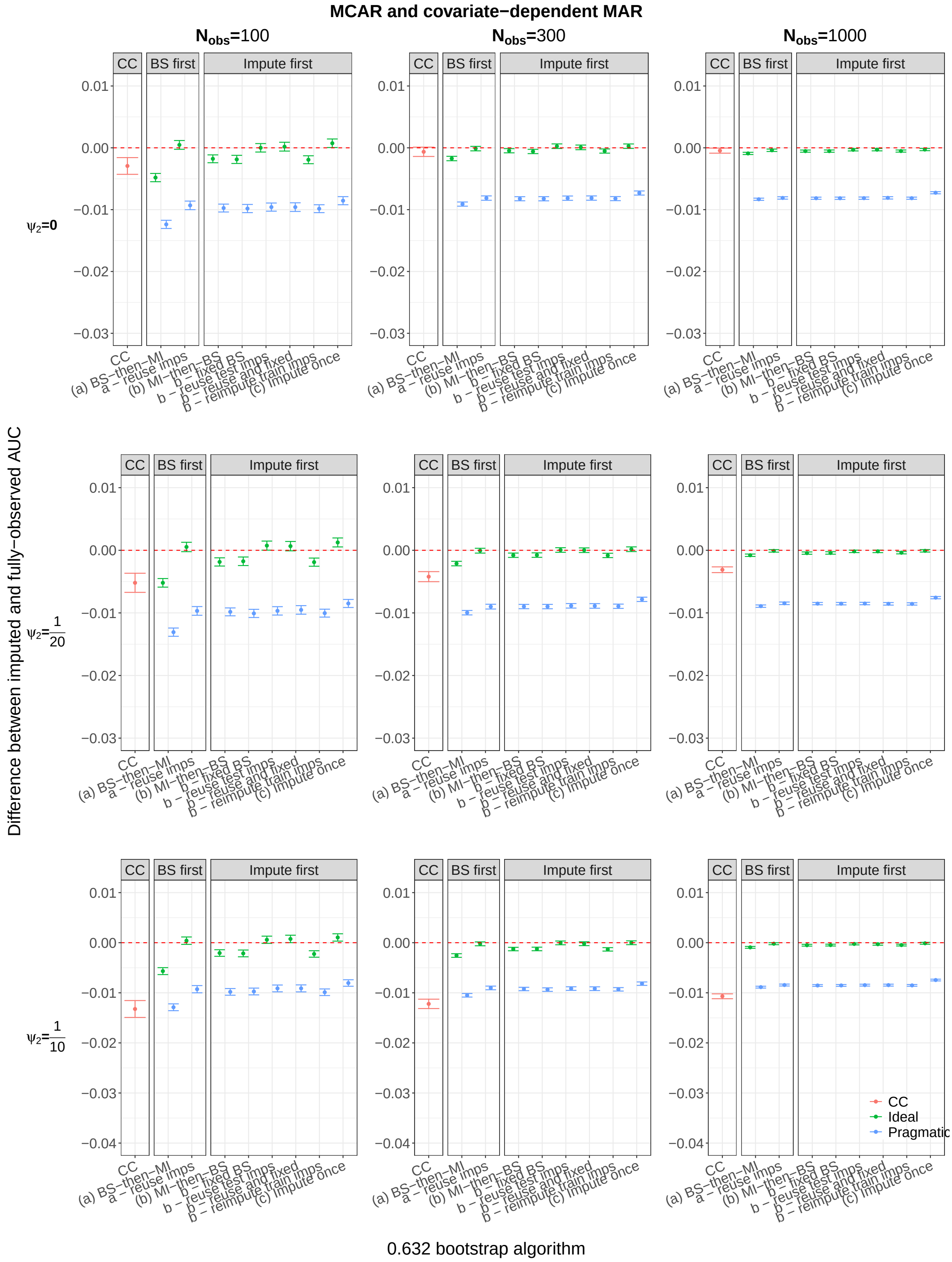

Figure S57: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are MCAR or covariate-dependent MAR. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

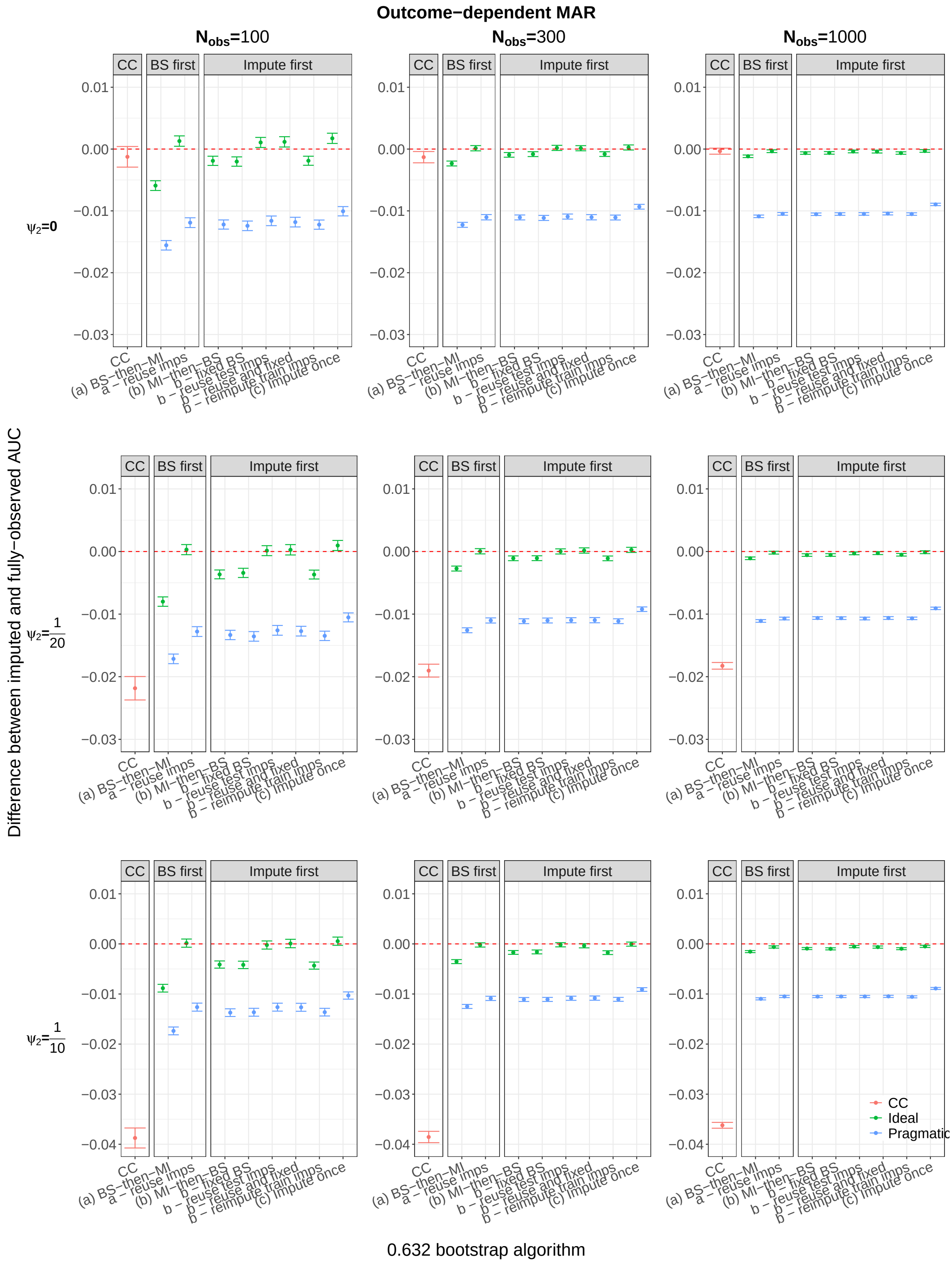

Figure S58: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are outcome-dependent or outcome- and covariate-dependent MAR. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.4.2 The proportion of missingness is 40% (AUC_{imp}-AUC_{obs})

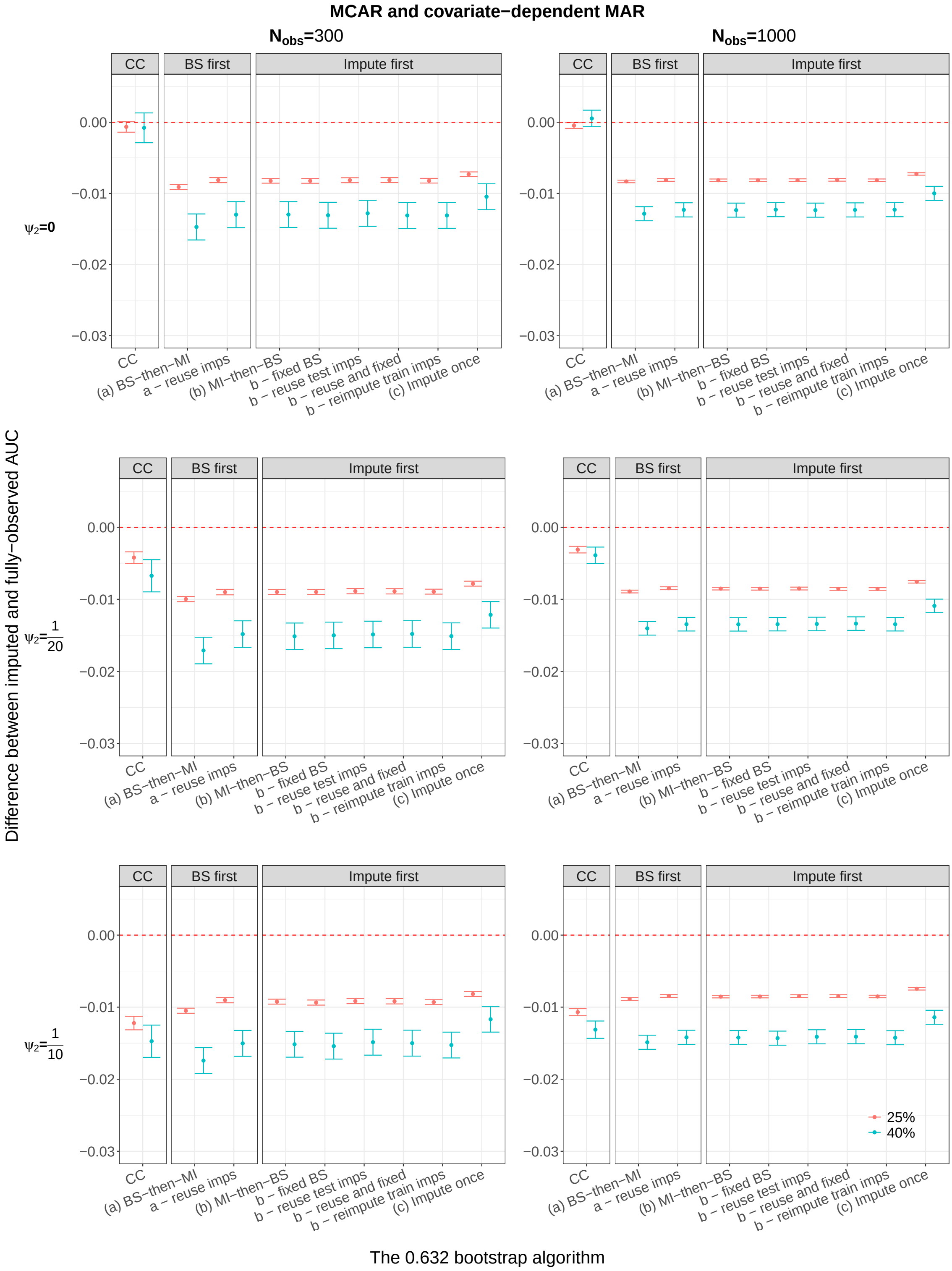

Figure S59: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are MCAR or covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for pragmatic performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

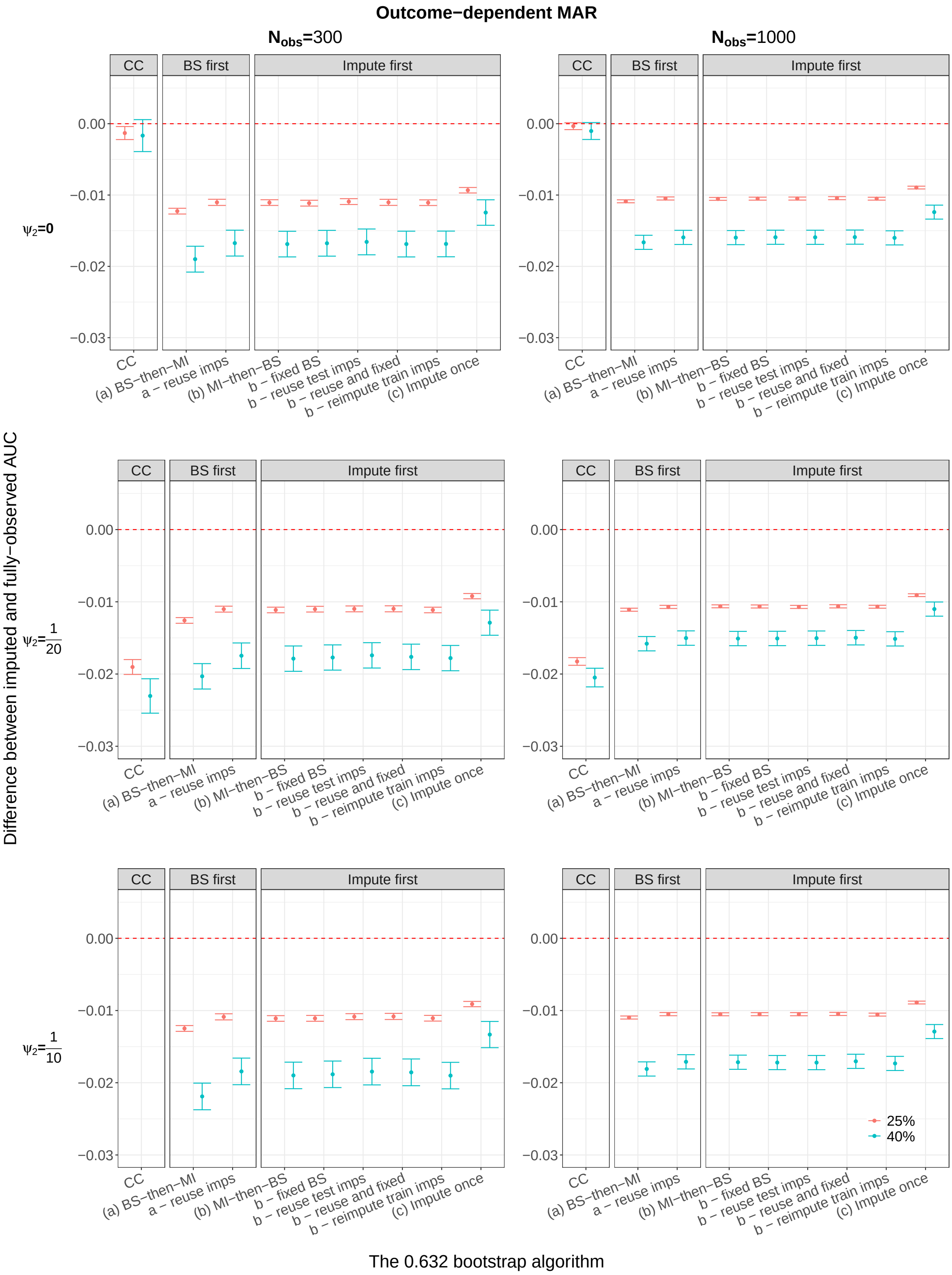

Figure S60: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are outcome-dependent or outcome- and covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for pragmatic performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

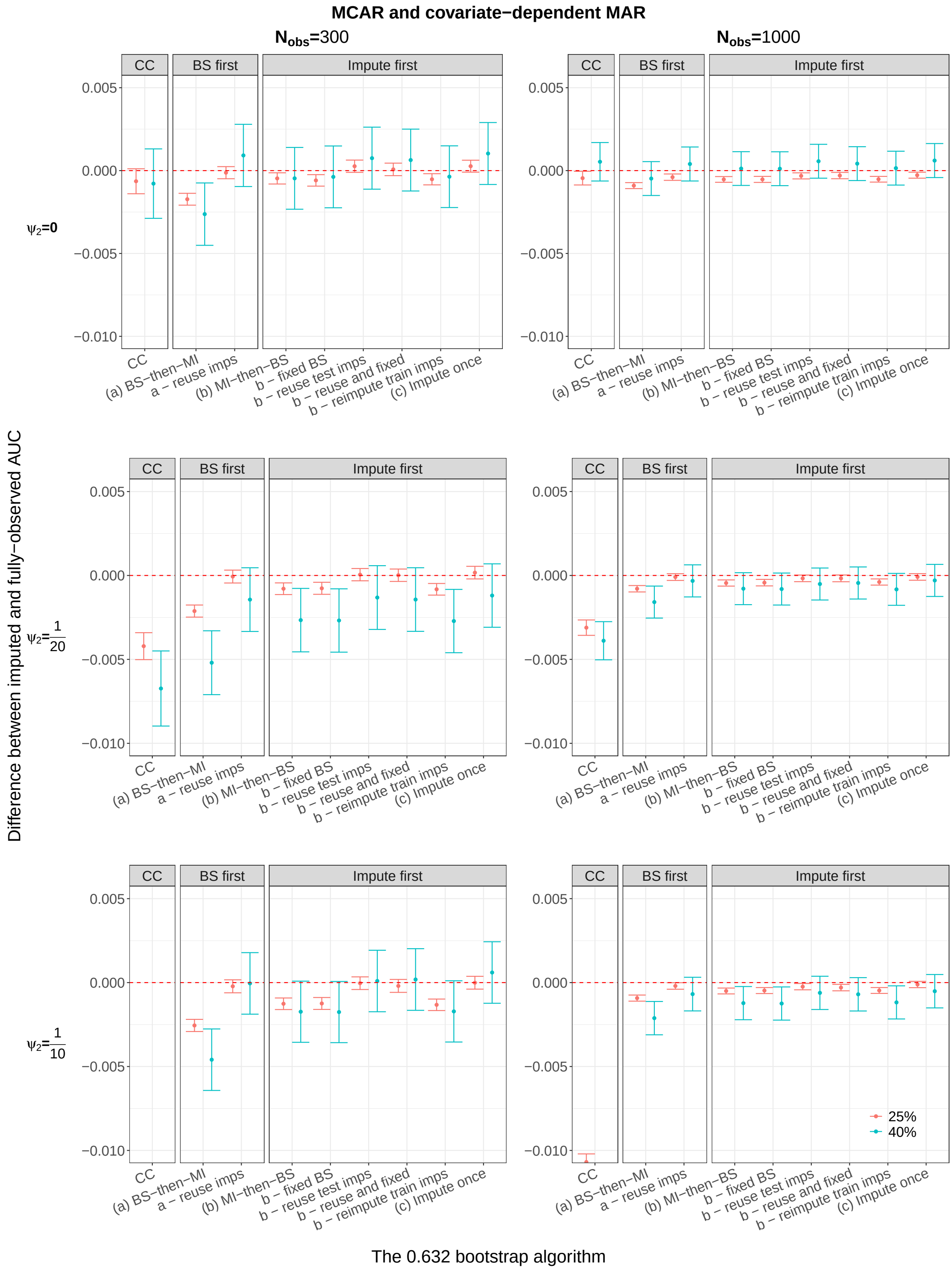

Figure S61: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are MCAR or covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for ideal performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

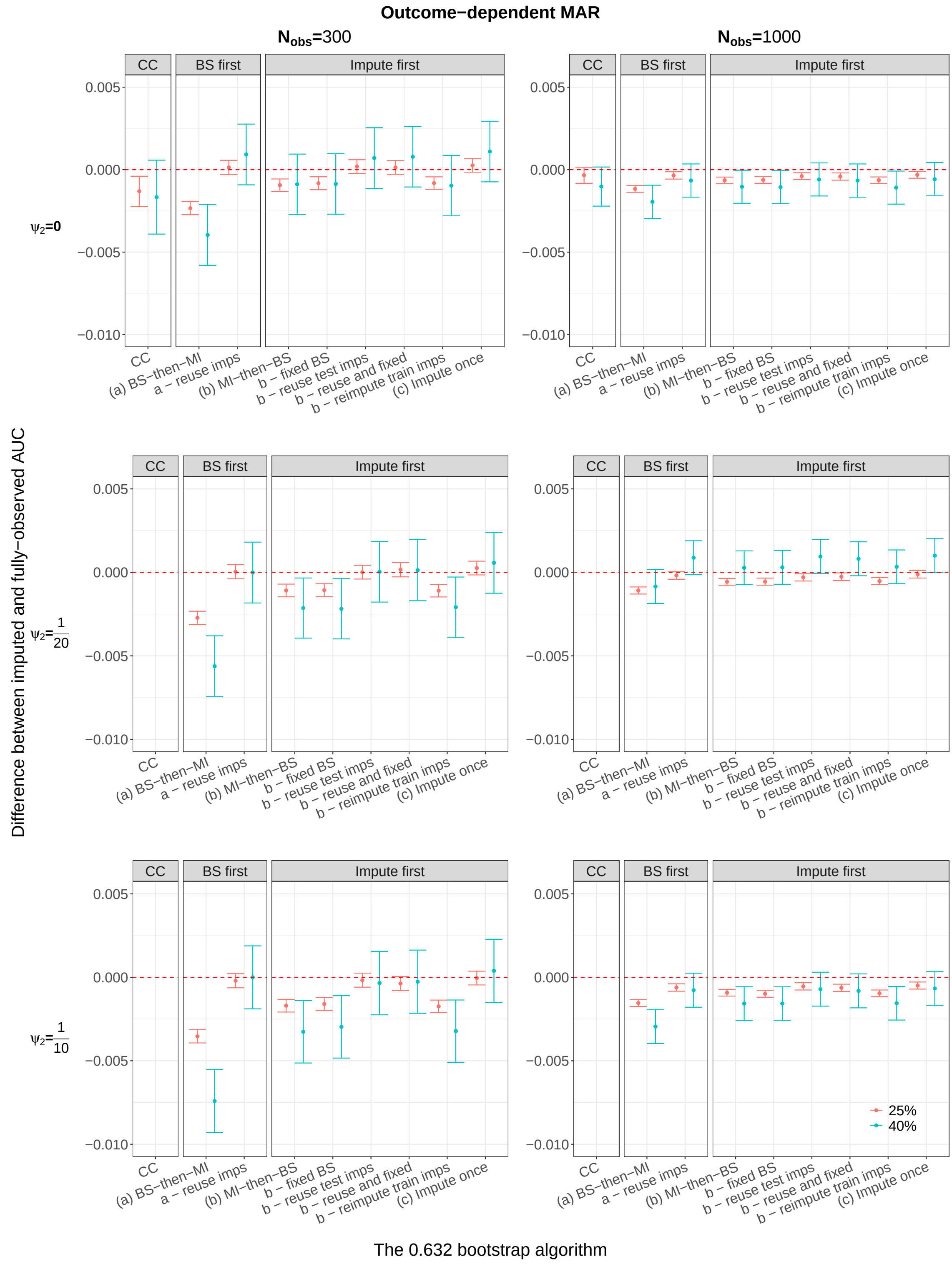

Figure S62: Error bars of the difference in the AUC from the imputation methods and the AUC estimate when data are fully-observed, with Monte Carlo 95% confidence intervals, when data are outcome-dependent or outcome- and covariate-dependent MAR. The graph compares the AUC estimates when 25% of X_1 values are missing versus 40% missing for ideal performance. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.4.3 Comparing M=5 versus M=25 (AUC_{imp}-AUC_{obs})

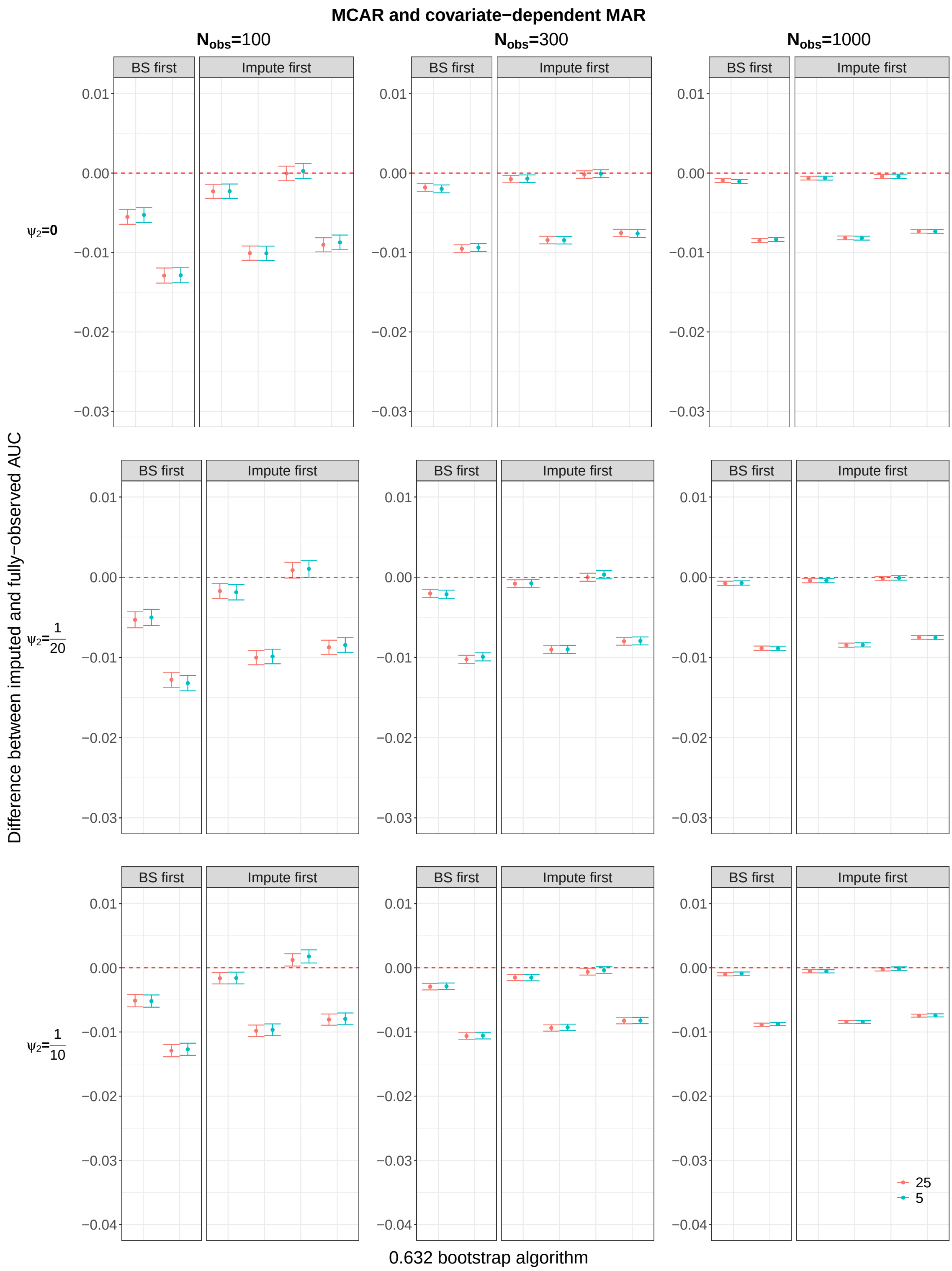

Figure S63: The difference AUC_{imp} - AUC_{obs} when data are MCAR or covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

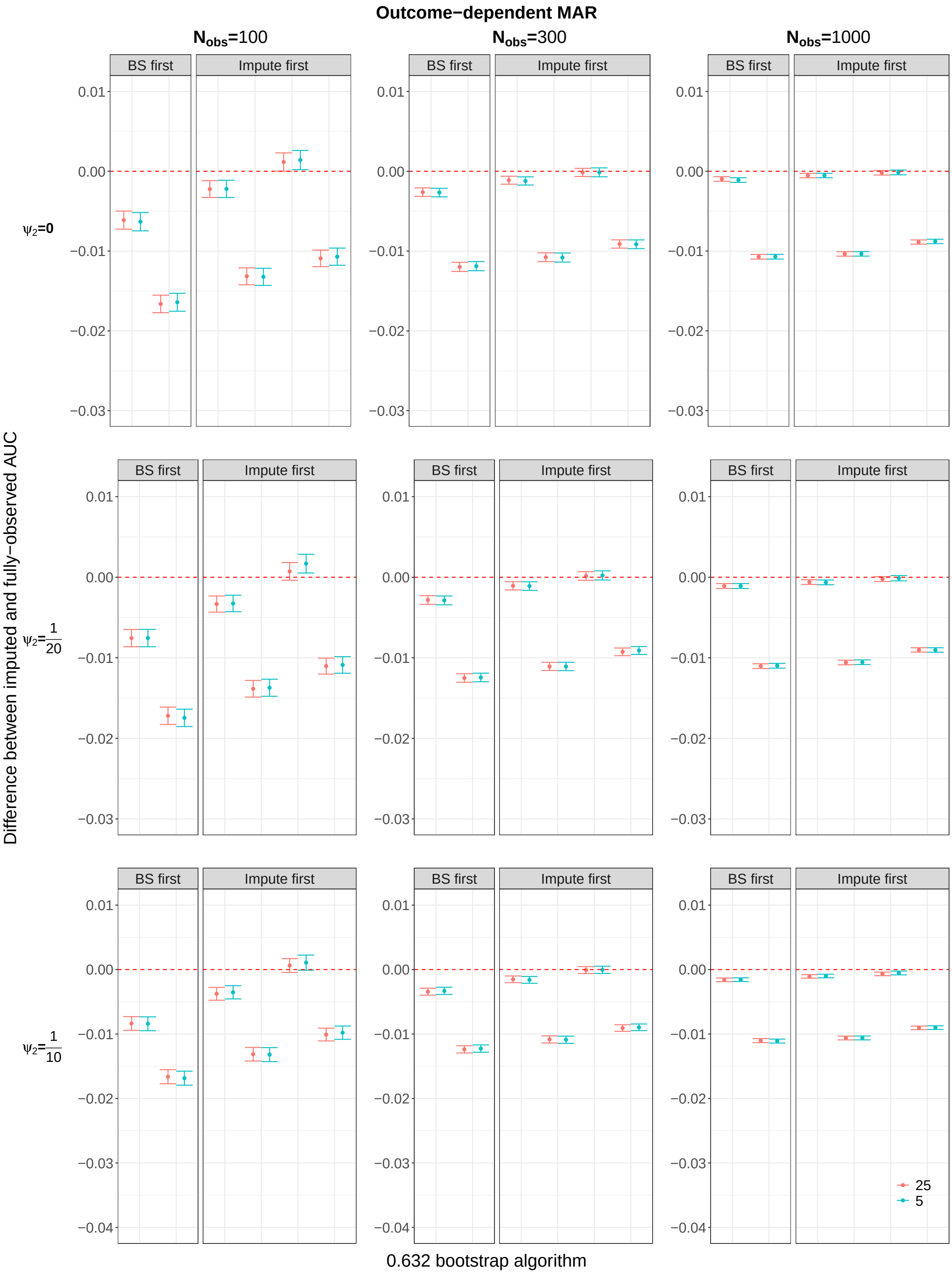

Figure S64: The difference AUC_{imp} - AUC_{obs} when data are outcome-dependent or outcome-and covariate-dependent MAR for $M = 25$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC_{imp} - AUC_{obs}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.4.4 AUC from imputation methods compared to the target AUC (AUC$_{target}$) using a larger validation set

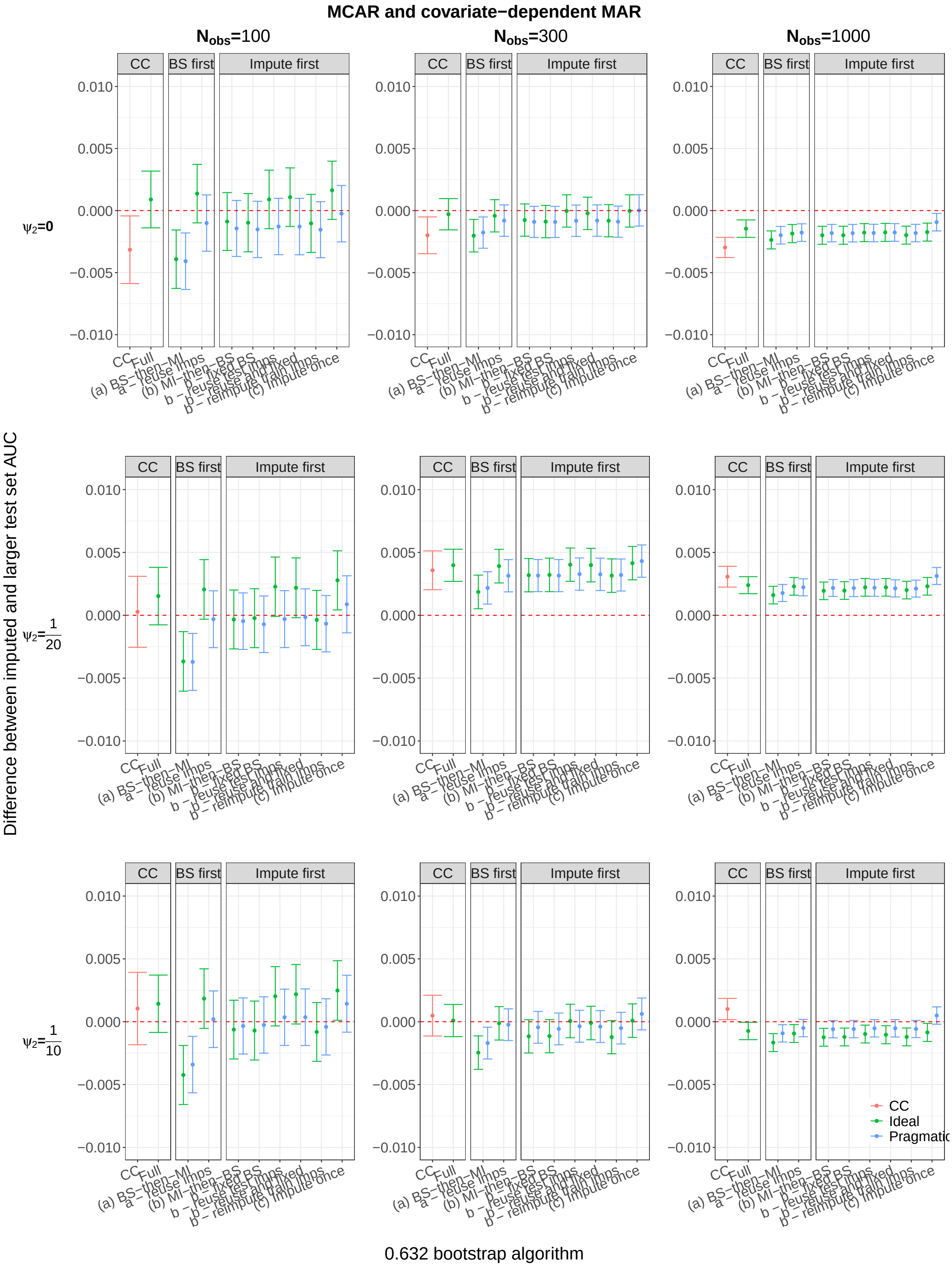

Figure S65: The difference AUC$_{imp}$ - AUC$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

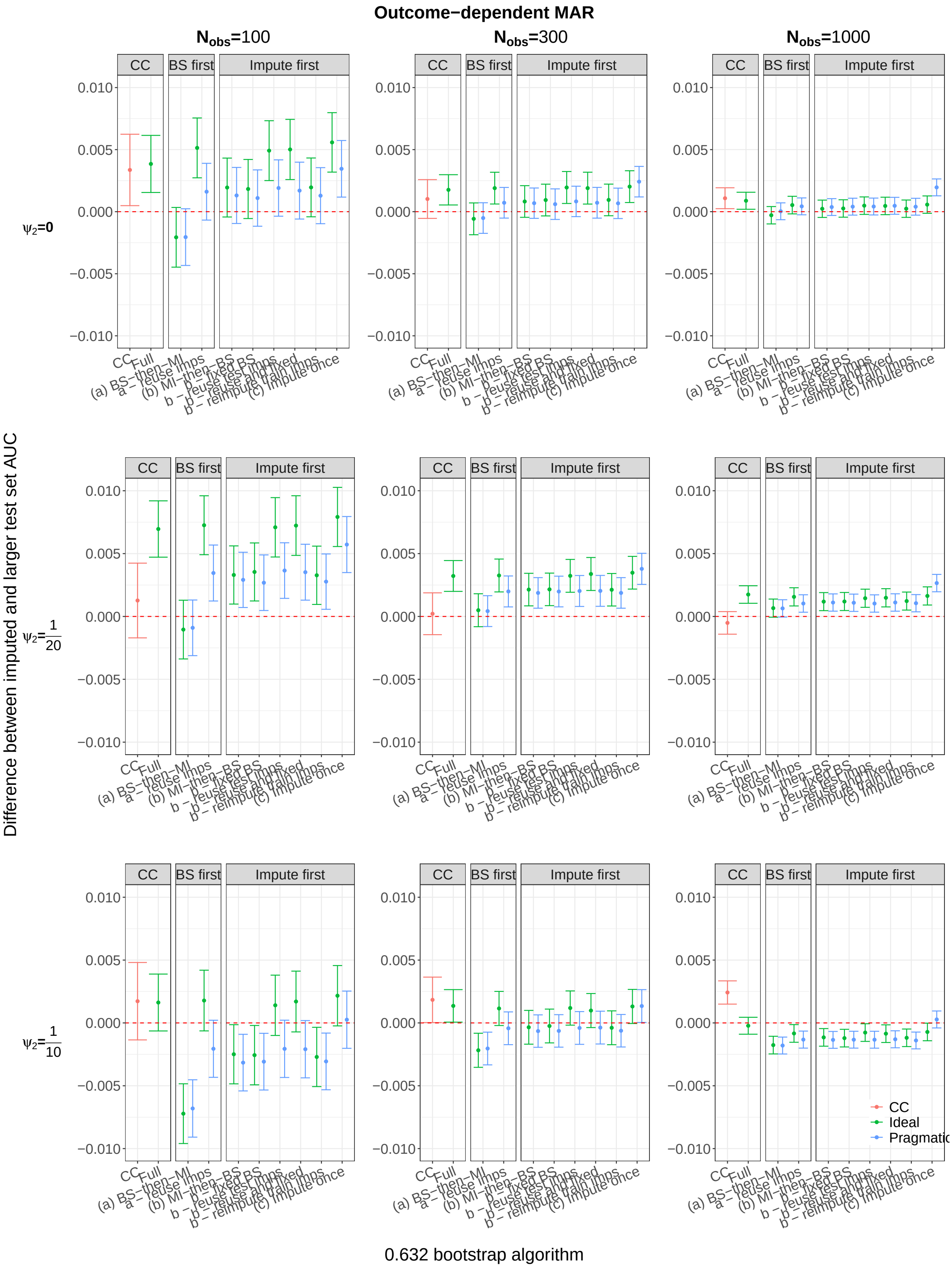

Figure S66: The difference AUC$_{imp}$ - AUC$_{target}$ when data are outcome-dependent or outcome-and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of AUC$_{imp}$ - AUC$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.5 The 0.632 bootstrap: Brier Score

S4.5.1 Brier Score from imputation methods compared to the fully-observed Brier Score (Brier$_{imp}$-Brier$_{obs}$)

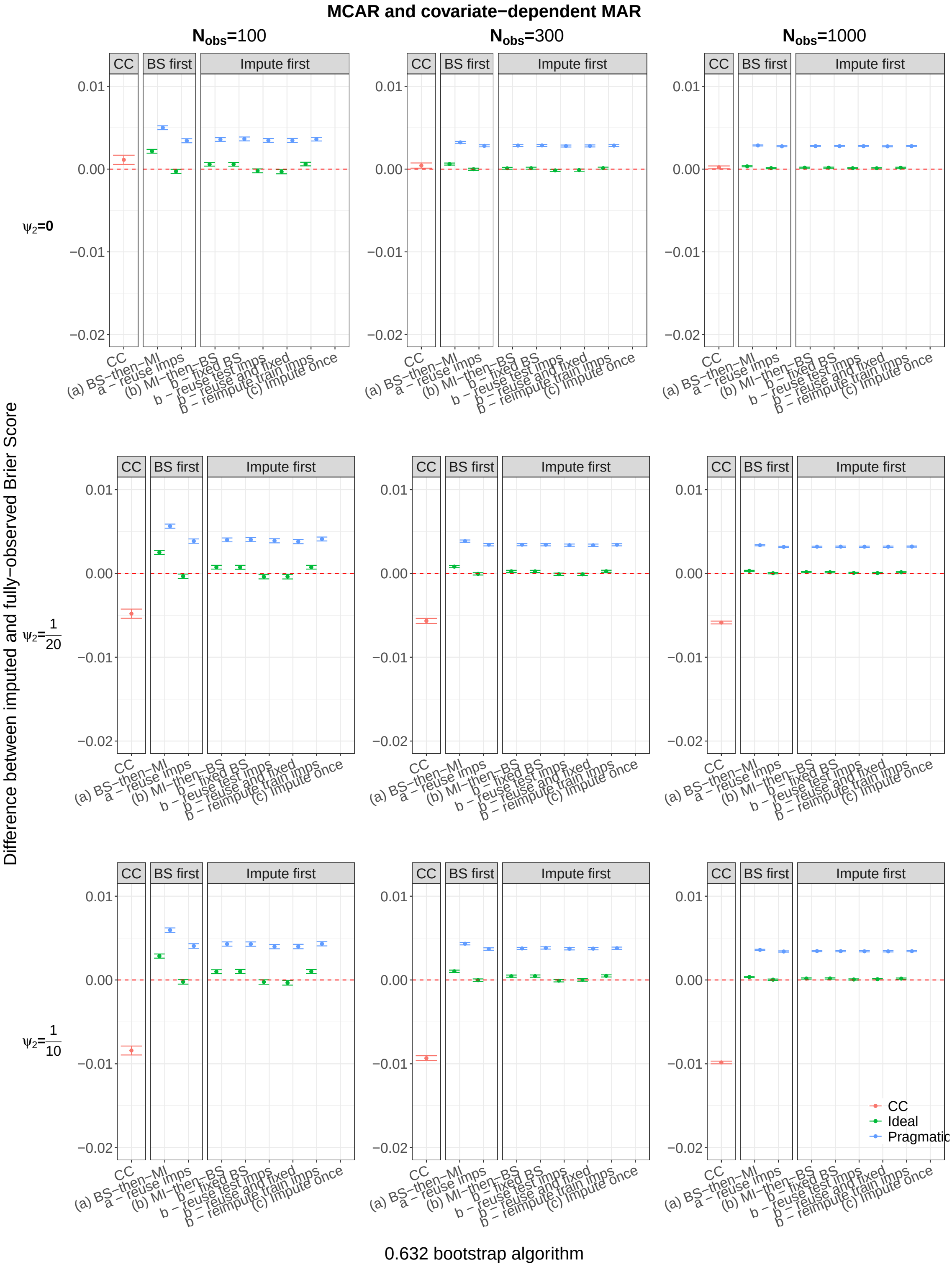

Figure S67: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

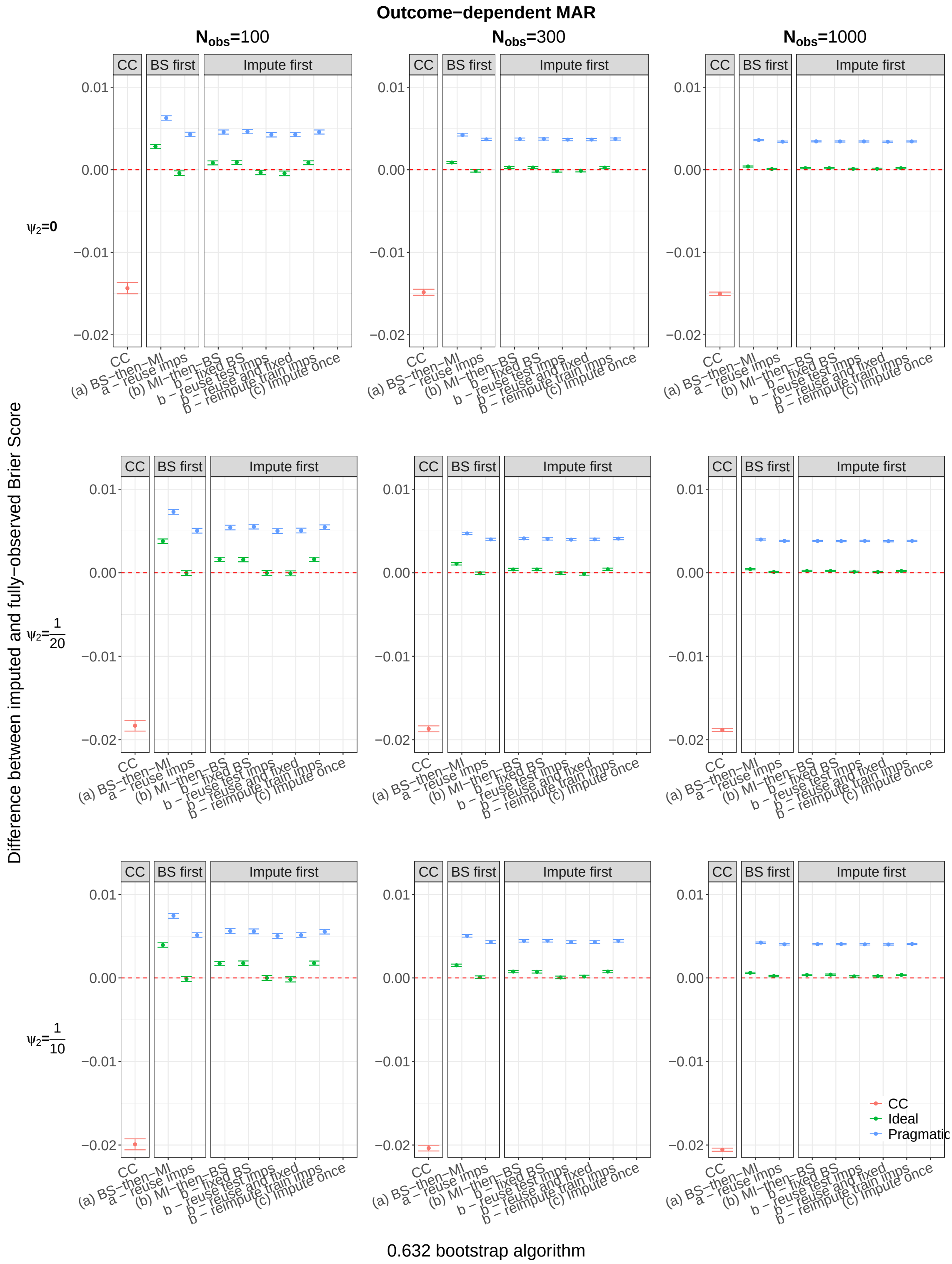

Figure S68: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.5.2 The proportion of missingness is 40% (Brier_{imp}-Brier_{obs})

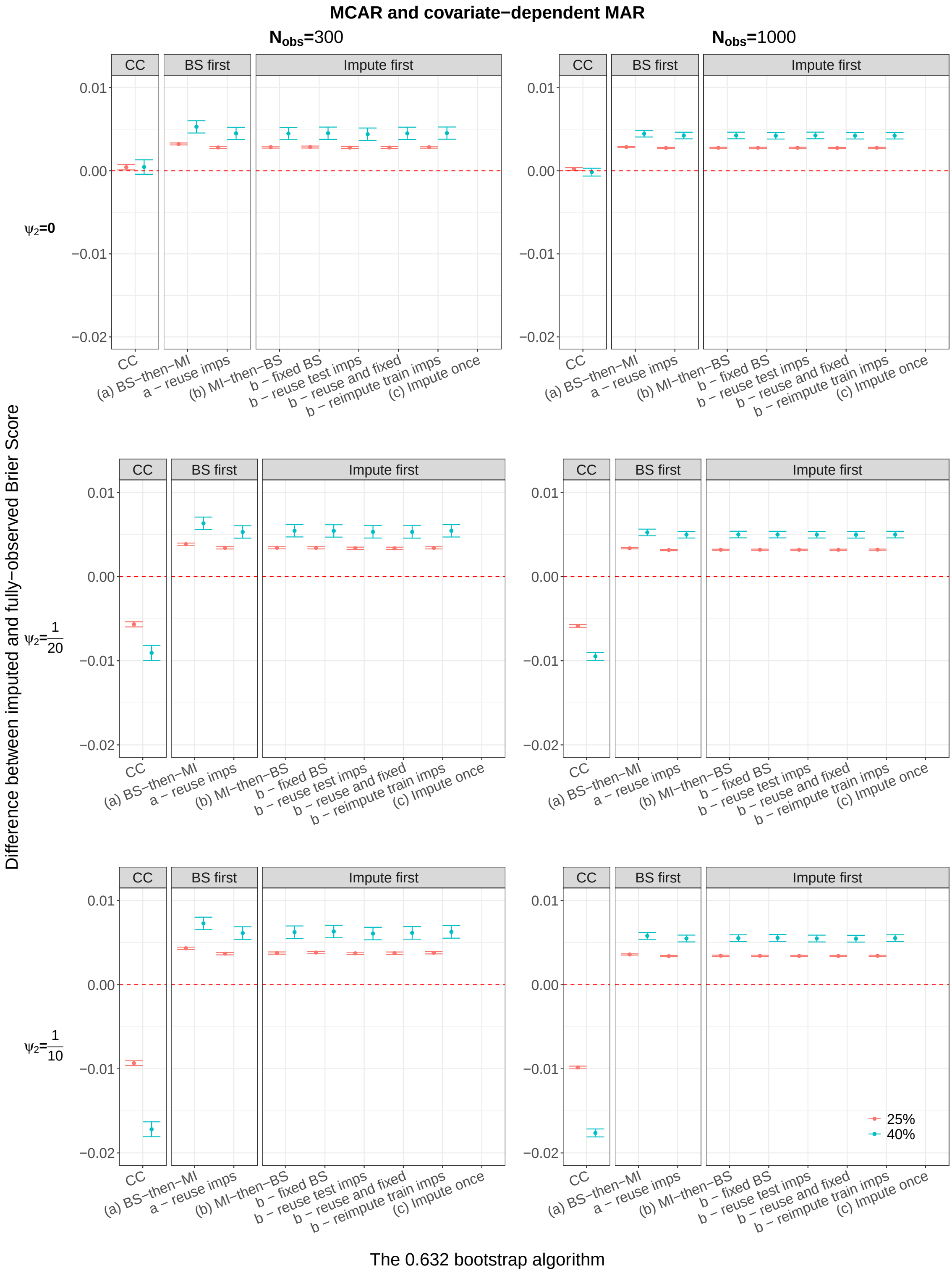

Figure S69: Comparing the impact of increasing the percentage of missingness on the difference Brier_{imp} - Brier_{obs} when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of BrierC_{imp} - Brier_{obs}. Red denotes Brier_{imp} - Brier_{obs} when 25% of X_1 values are missing and blue denotes Brier_{imp} - Brier_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

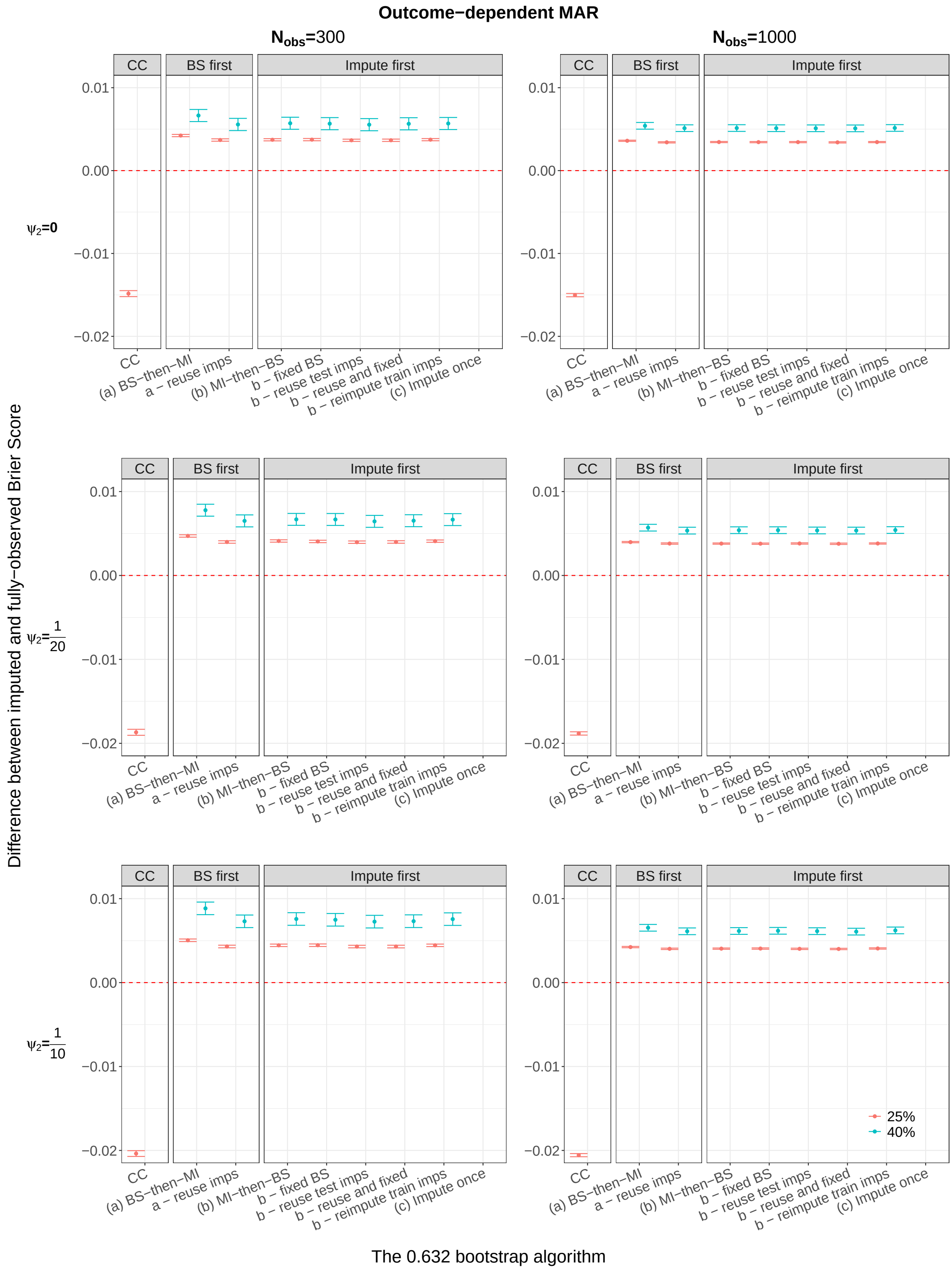

Figure S70: Comparing the impact of increasing the percentage of missingness on the difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. Red denotes Brier$_{imp}$ - Brier$_{obs}$ when 25% of X_1 values are missing and blue denotes Brier$_{imp}$ - Brier$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

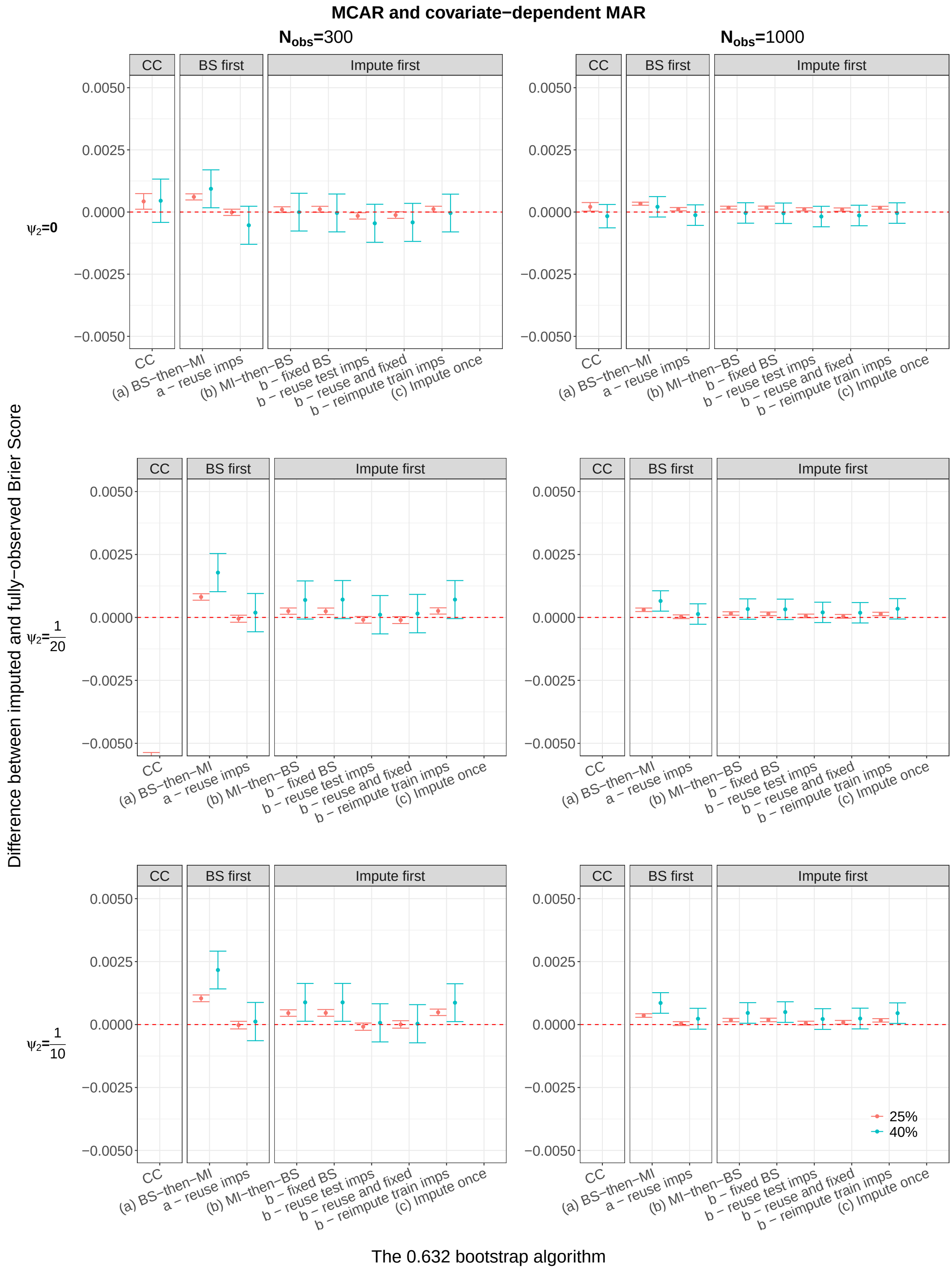

Figure S71: Comparing the impact of increasing the percentage of missingness on the difference Brier_{imp} - Brier_{obs} when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of BrierC_{imp} - Brier_{obs}. Red denotes Brier_{imp} - Brier_{obs} when 25% of X_1 values are missing and blue denotes Brier_{imp} - Brier_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

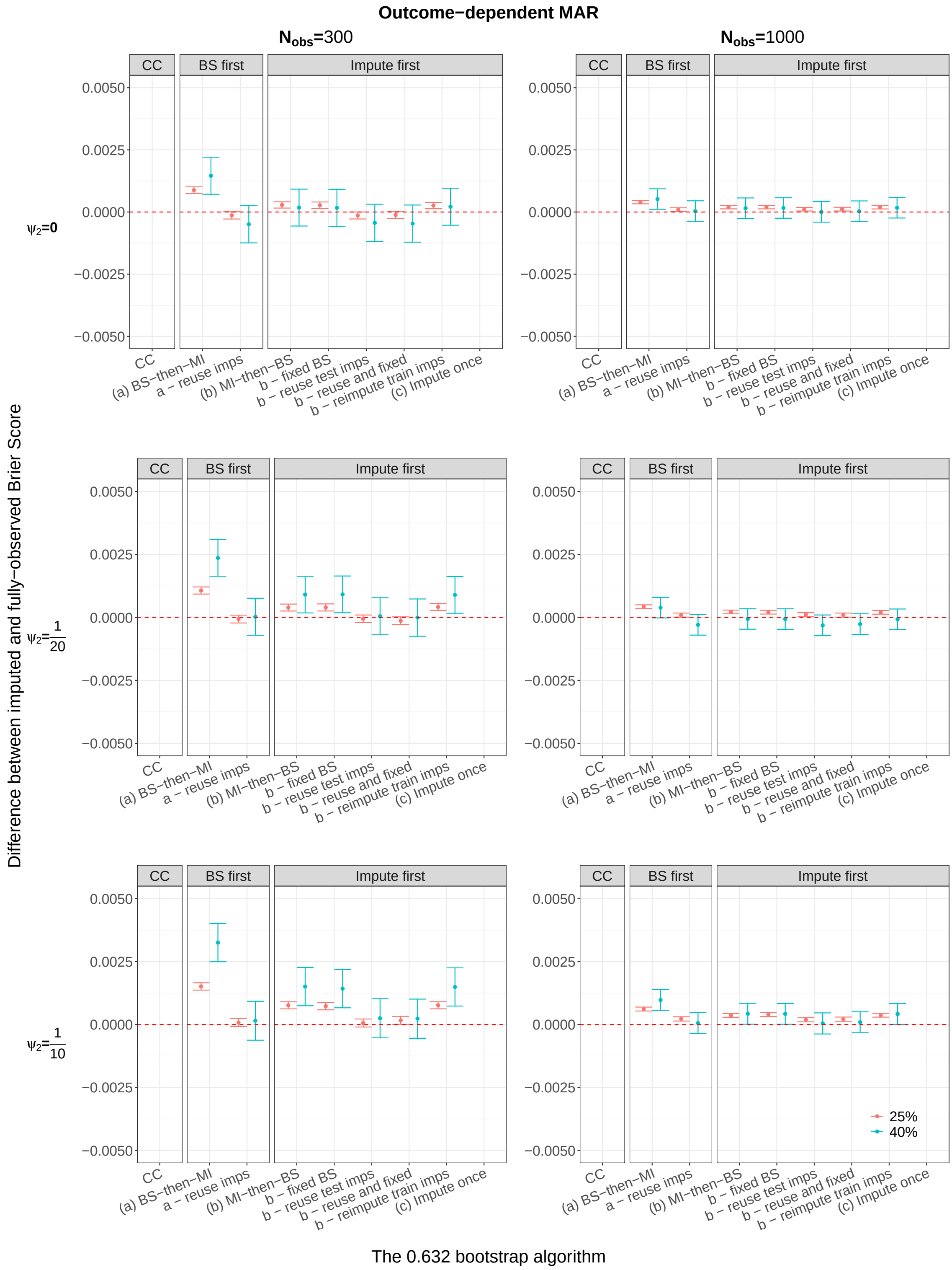

Figure S72: Comparing the impact of increasing the percentage of missingness on the difference Brier_{imp} - Brier_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Brier_{imp} - Brier_{obs}. Red denotes Brier_{imp} - Brier_{obs} when 25% of X_1 values are missing and blue denotes Brier_{imp} - Brier_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.5.3 Comparing M=5 versus M=25 (Brier$_{imp}$-Brier$_{obs}$)

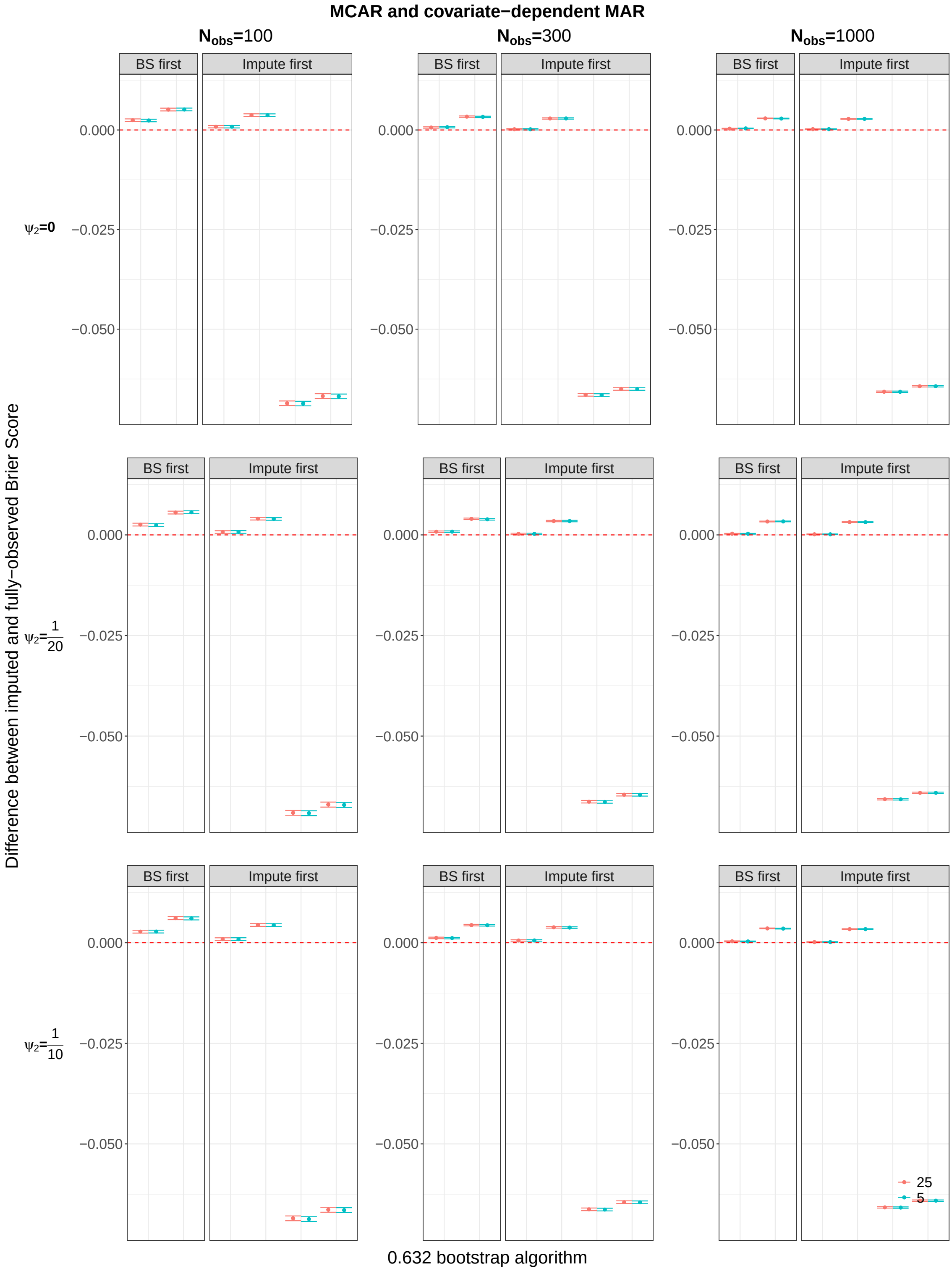

Figure S73: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

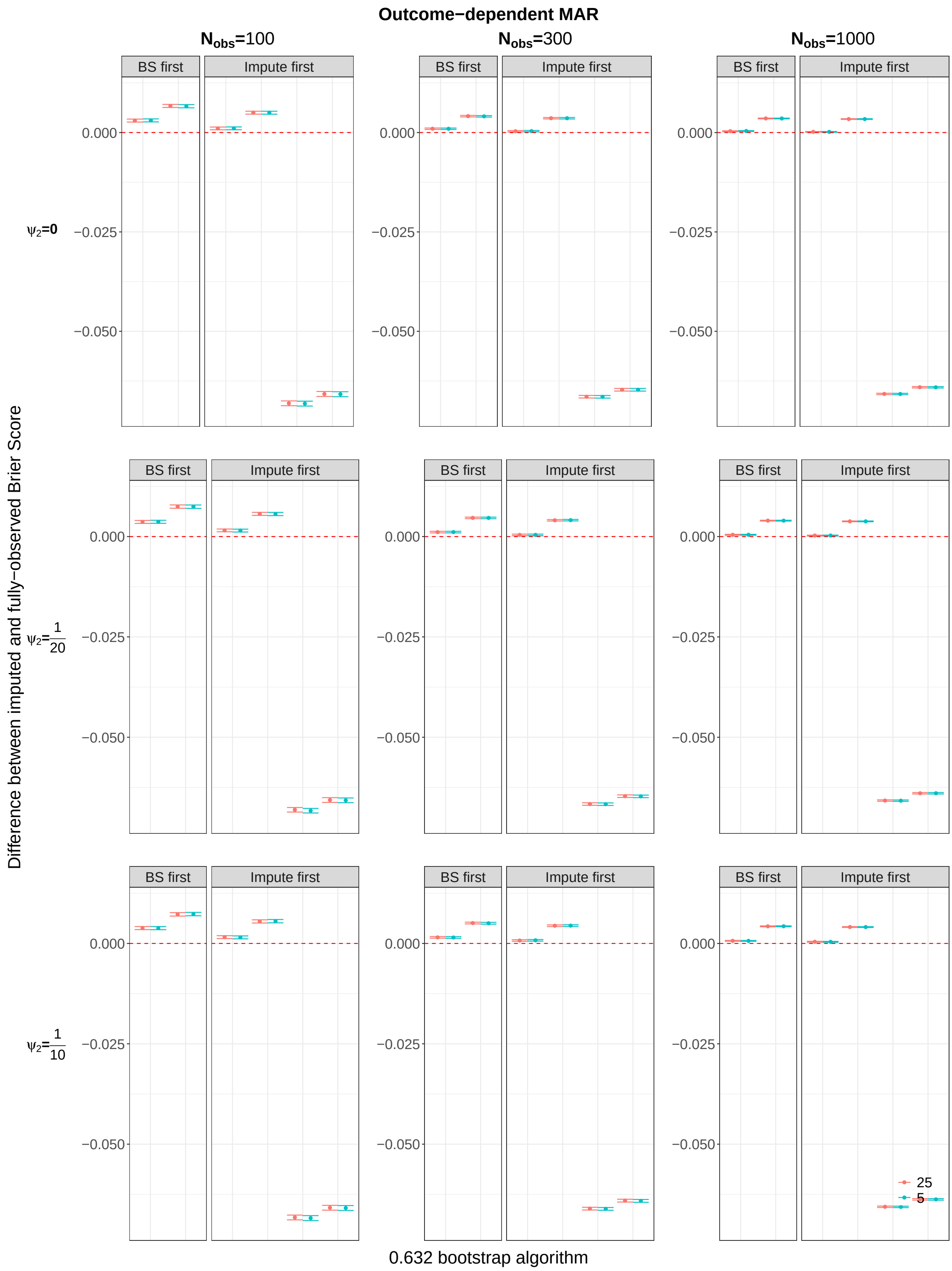

Figure S74: The difference Brier$_{imp}$ - Brier$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.5.4 Brier Score from imputation methods compared to the target Brier Score (Brier_{target}) using a larger validation set

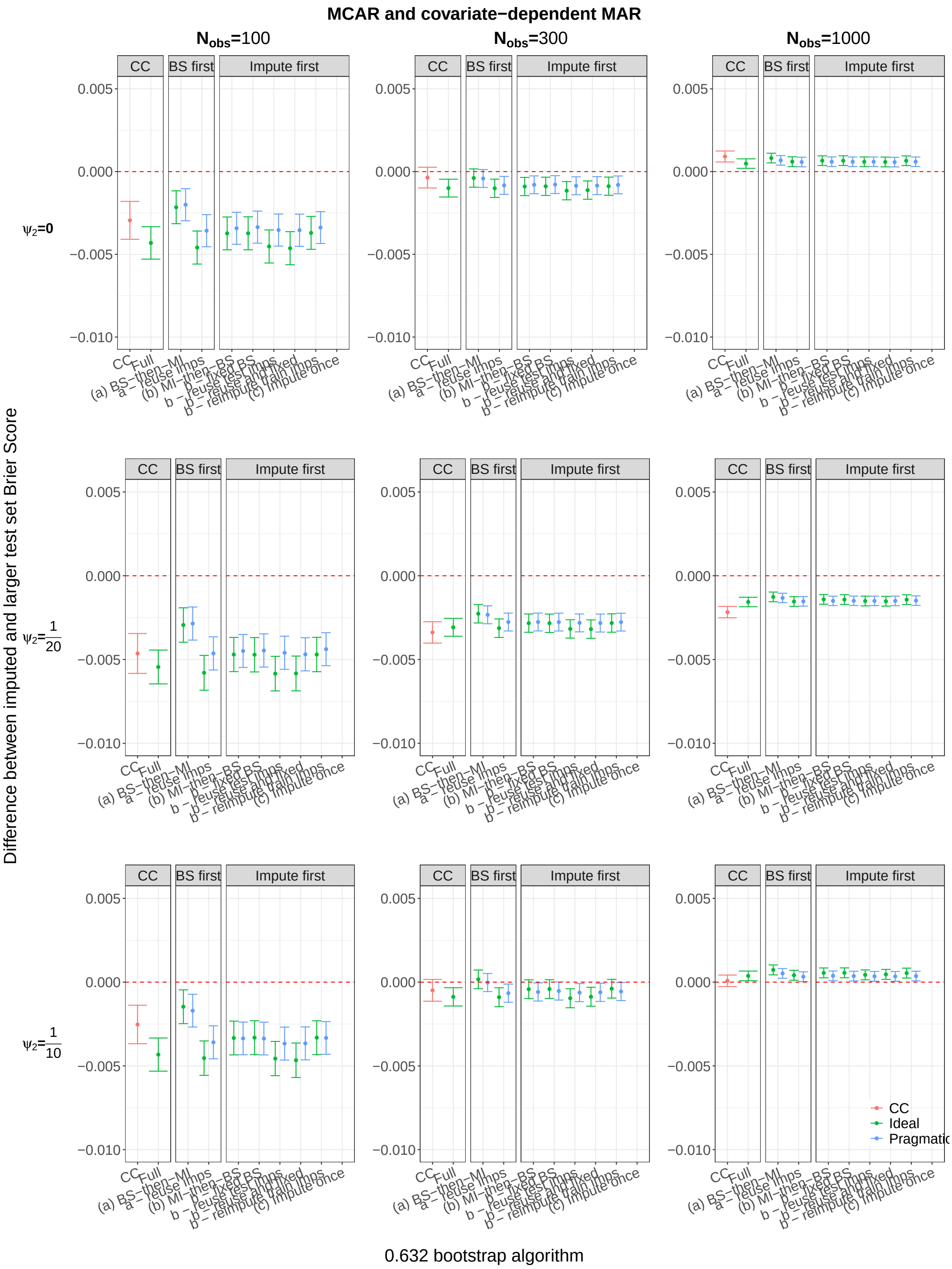

Figure S75: The difference Brier_{imp} - Brier_{target} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier_{imp} - Brier_{target}. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

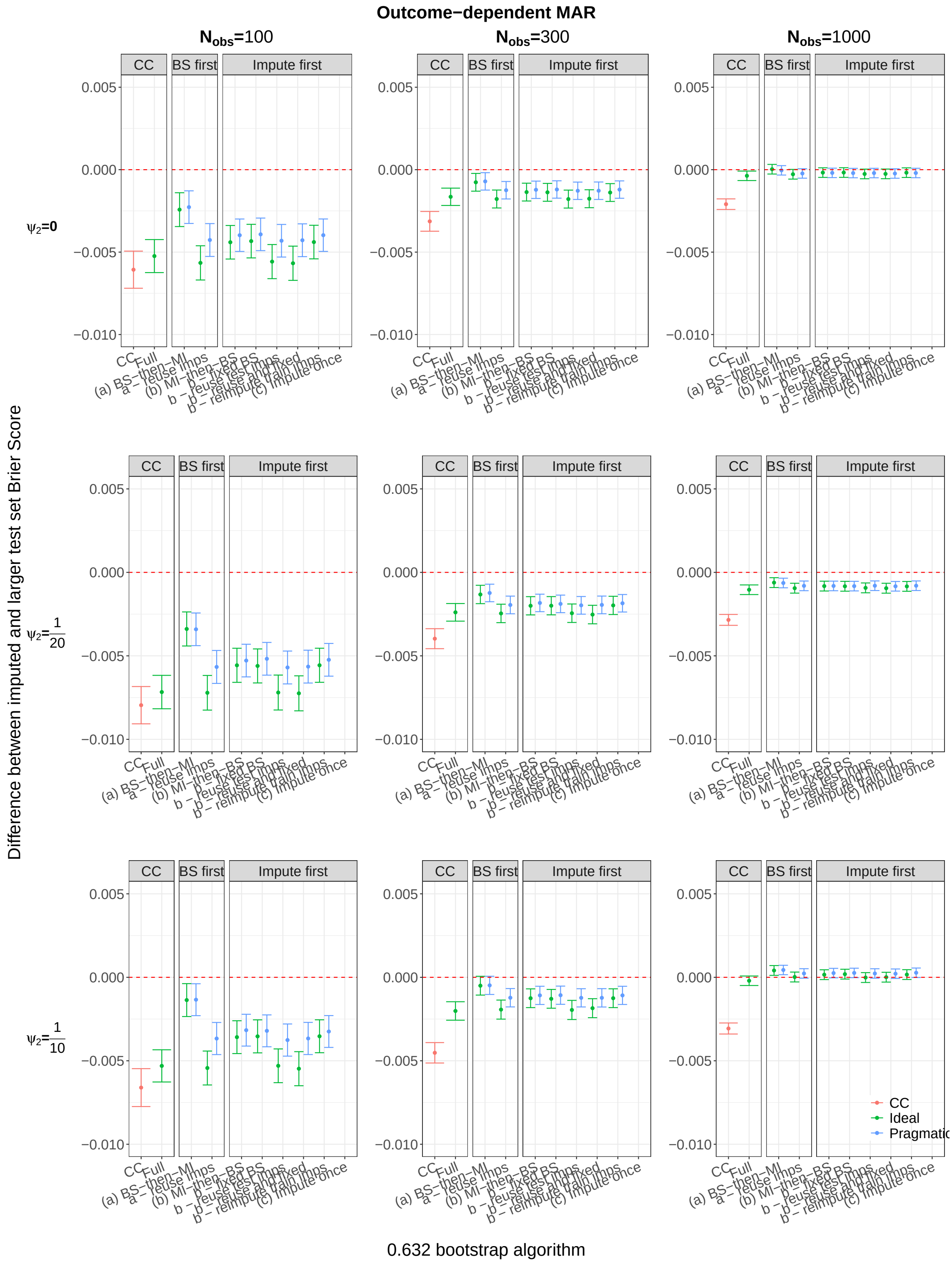

Figure S76: The difference Brier$_{imp}$ - Brier$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Brier$_{imp}$ - Brier$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.6 The 0.632 bootstrap: Calibration intercept and slope

S4.6.1 Calibration intercept and slope from imputation methods compared to the fully-observed Calibration intercept and slope (Cal$_{imp}$-Cal$_{obs}$)

Intercept

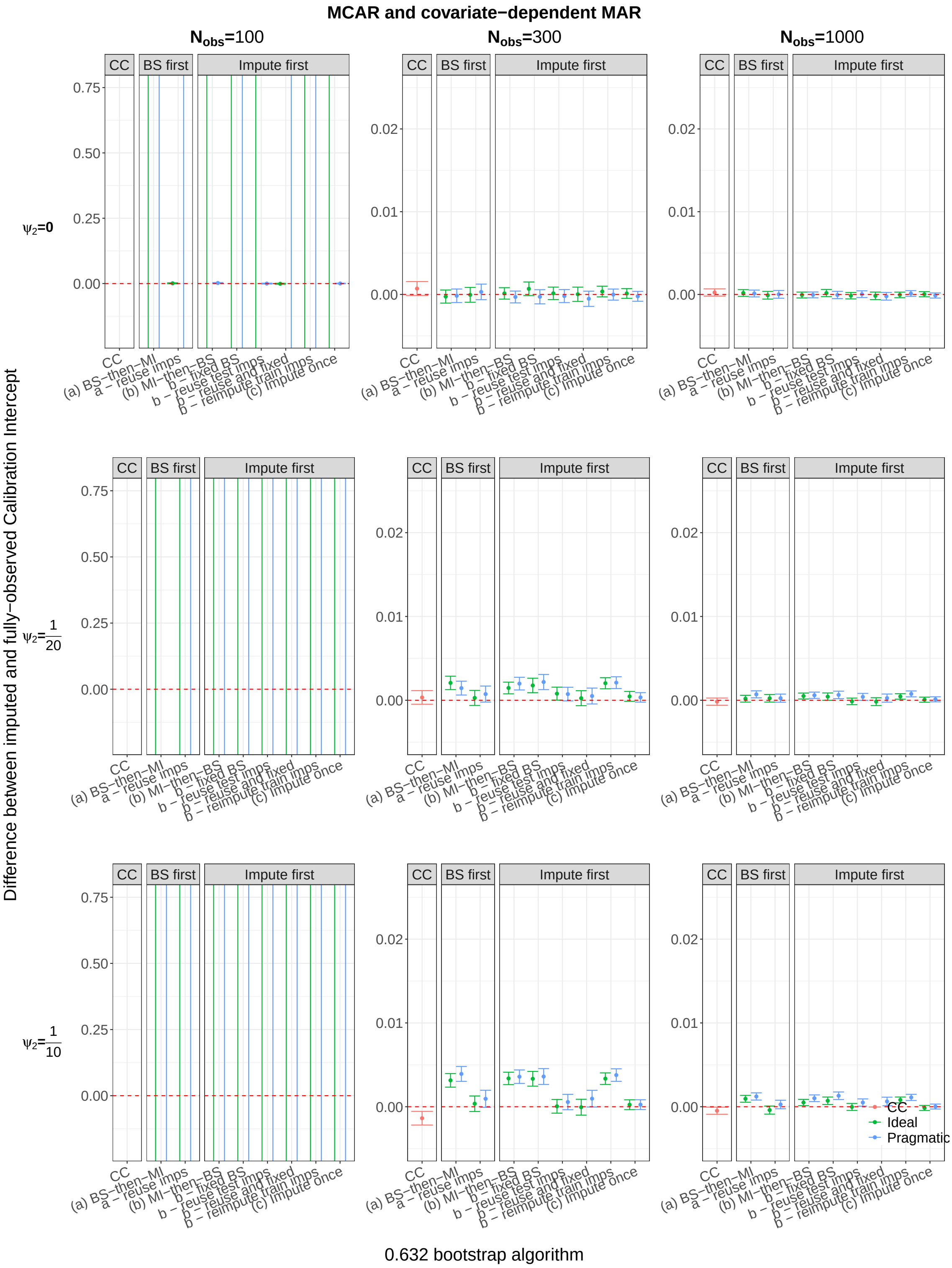

Figure S77: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

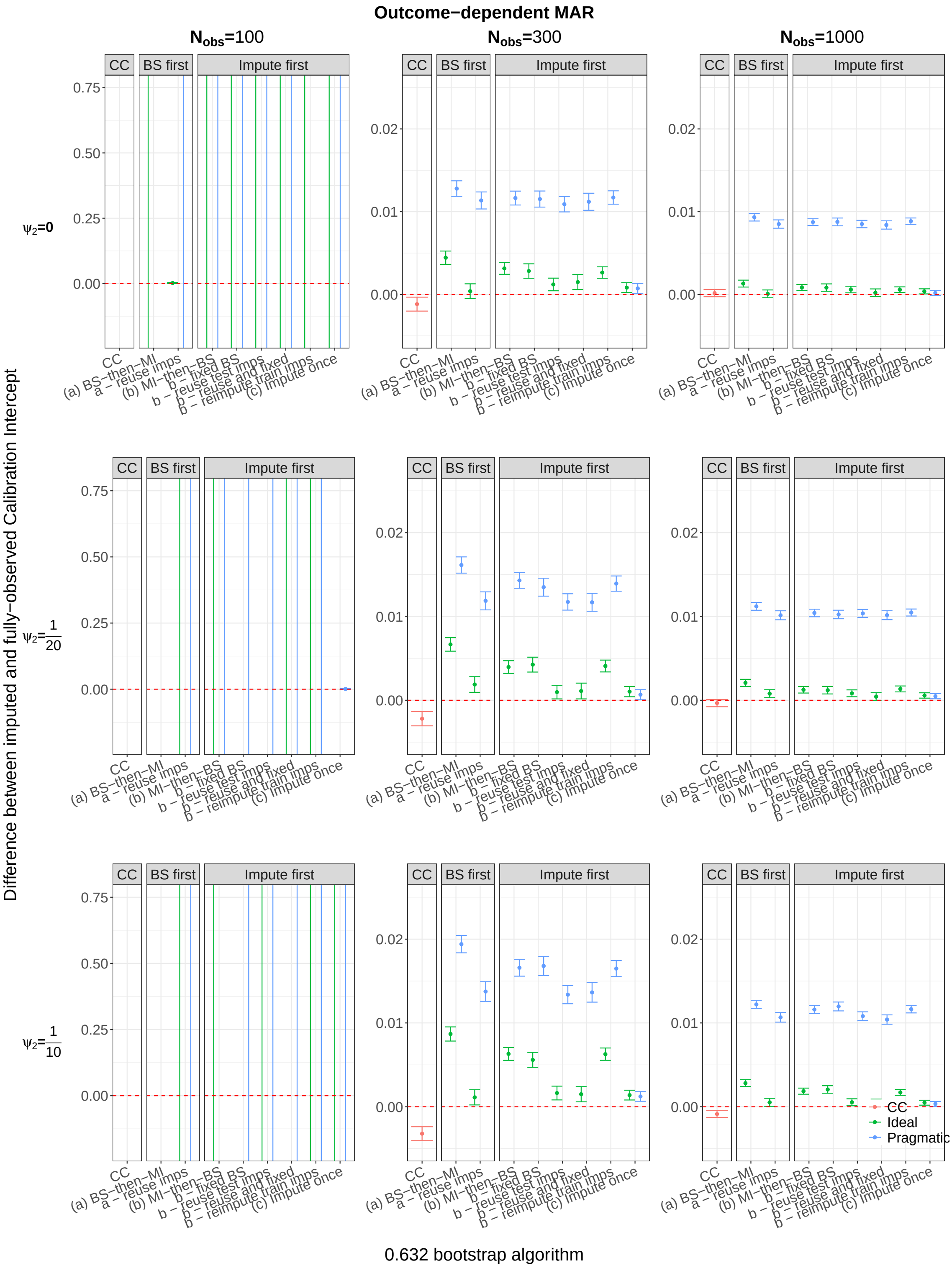

Figure S78: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

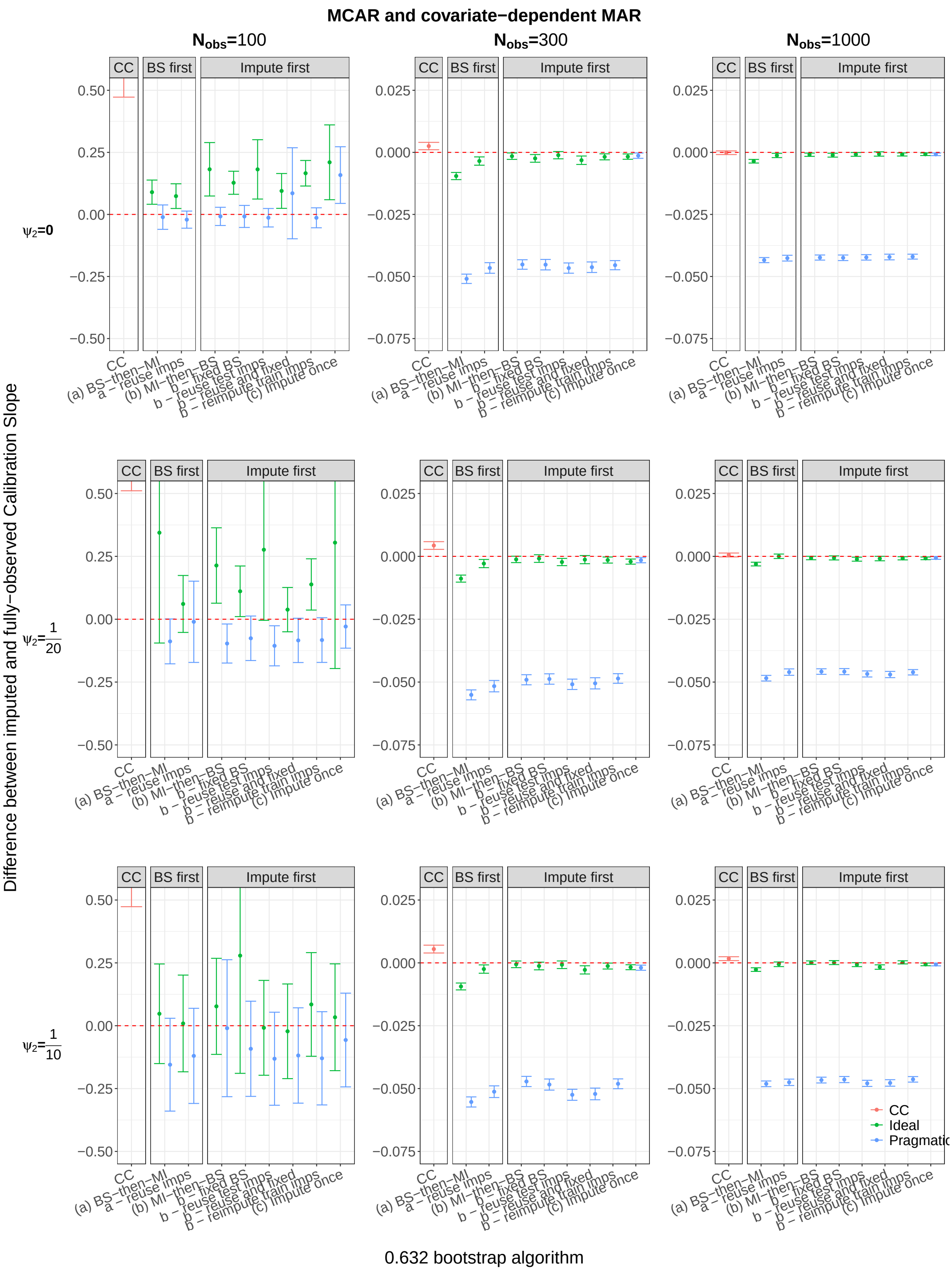

Figure S79: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

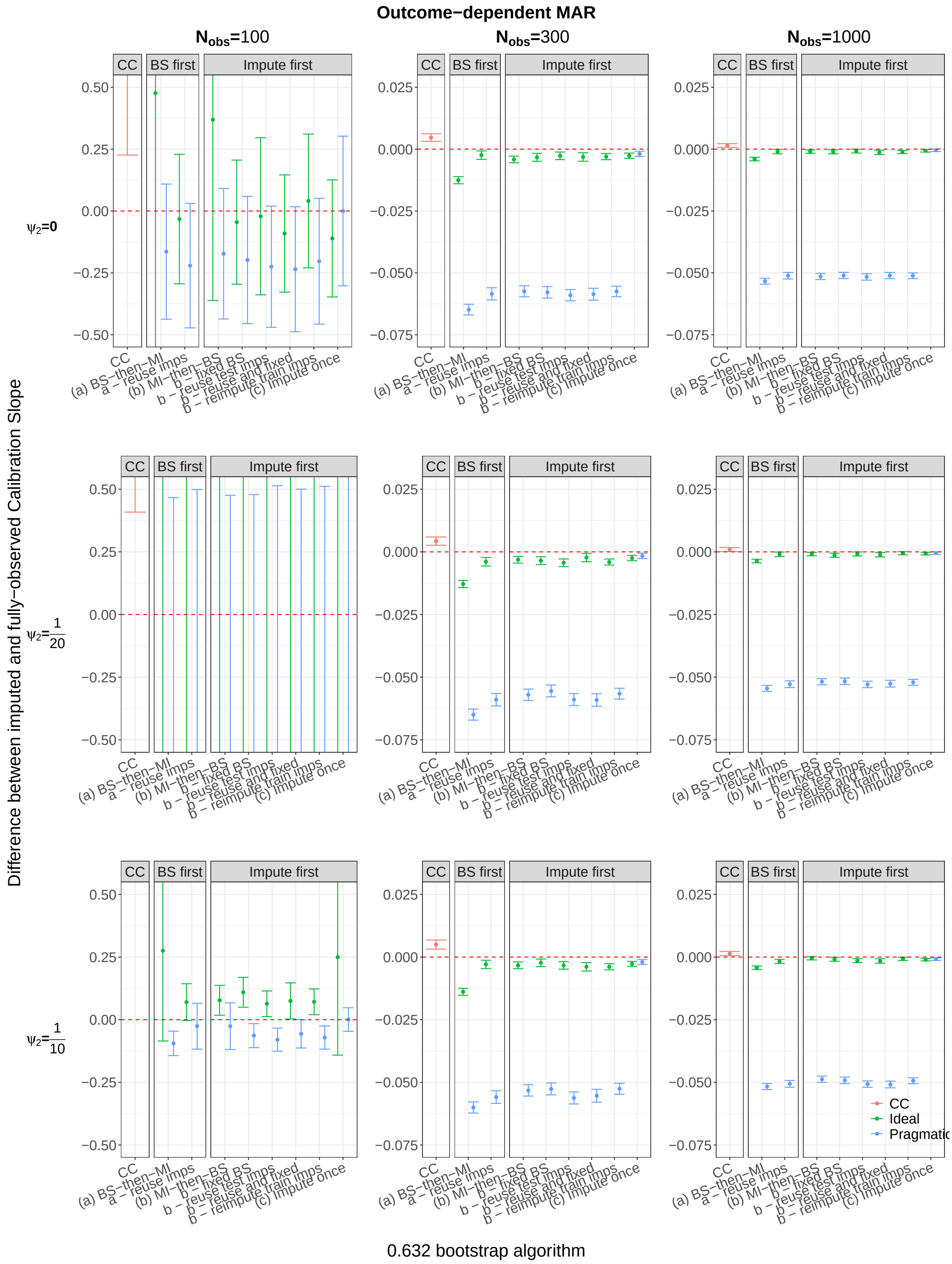

Figure S80: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.6.2 The proportion of missingness is 40% (Cal$_{imp}$-Cal$_{obs}$)

Intercept

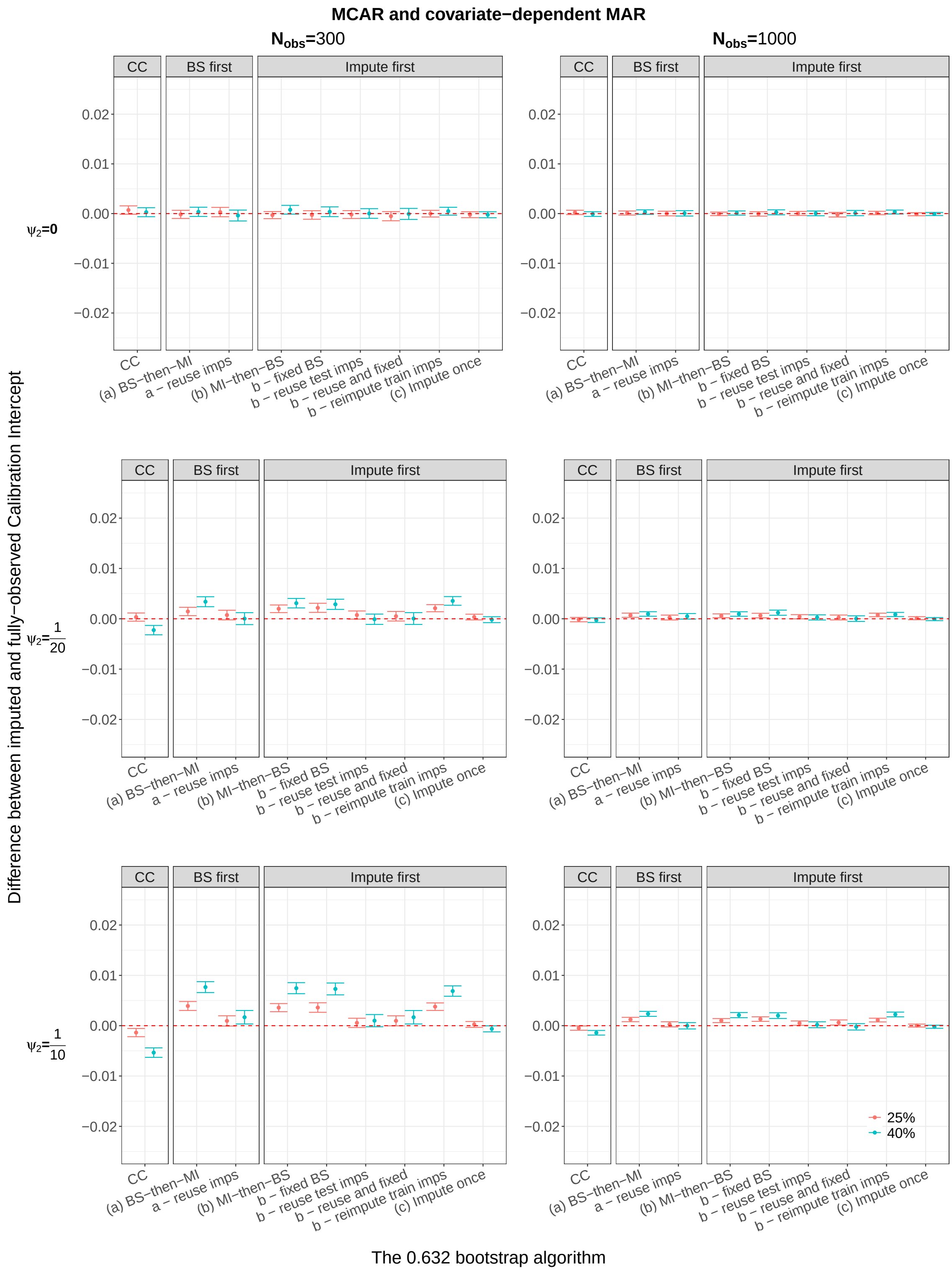

Figure S81: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

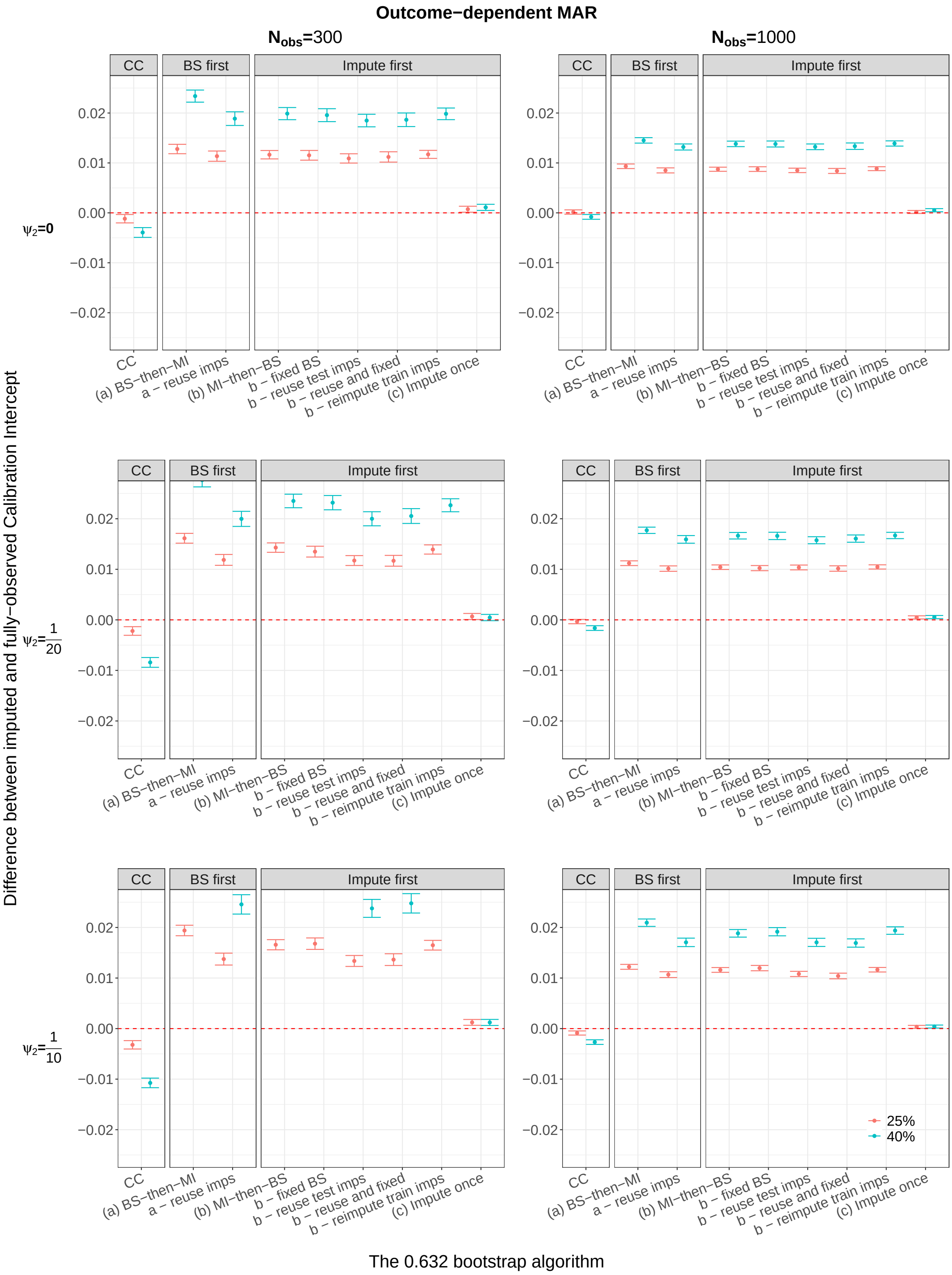

Figure S82: Comparing the impact of increasing the percentage of missingness on the difference Intercept_{imp} - Intercept_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Intercept_{imp} - Intercept_{obs}. Red denotes Intercept_{imp} - Intercept_{obs} when 25% of X_1 values are missing and blue denotes Intercept_{imp} - Intercept_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

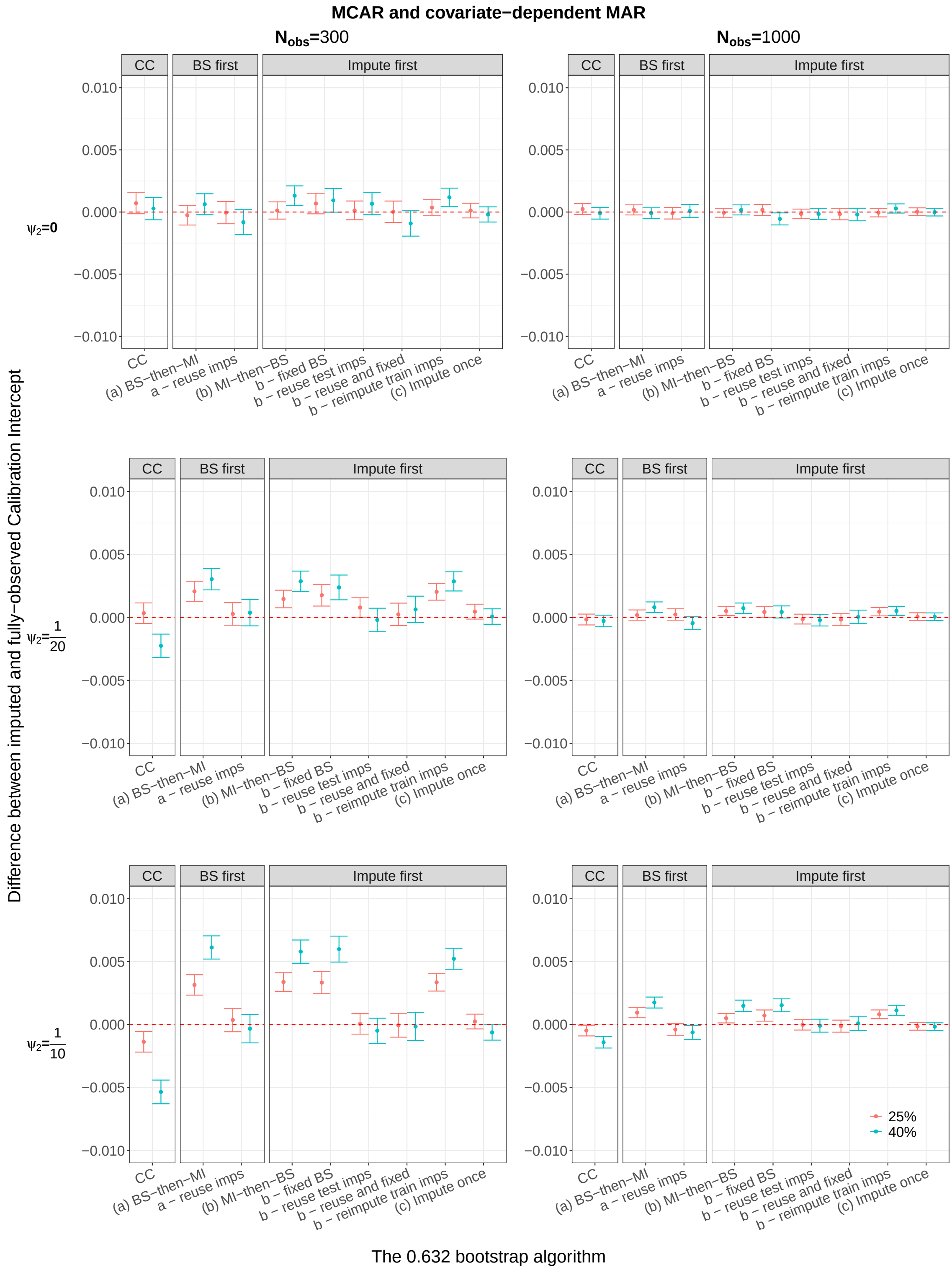

Figure S83: Comparing the impact of increasing the percentage of missingness on the difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. Red denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 25% of X_1 values are missing and blue denotes Intercept$_{imp}$ - Intercept$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

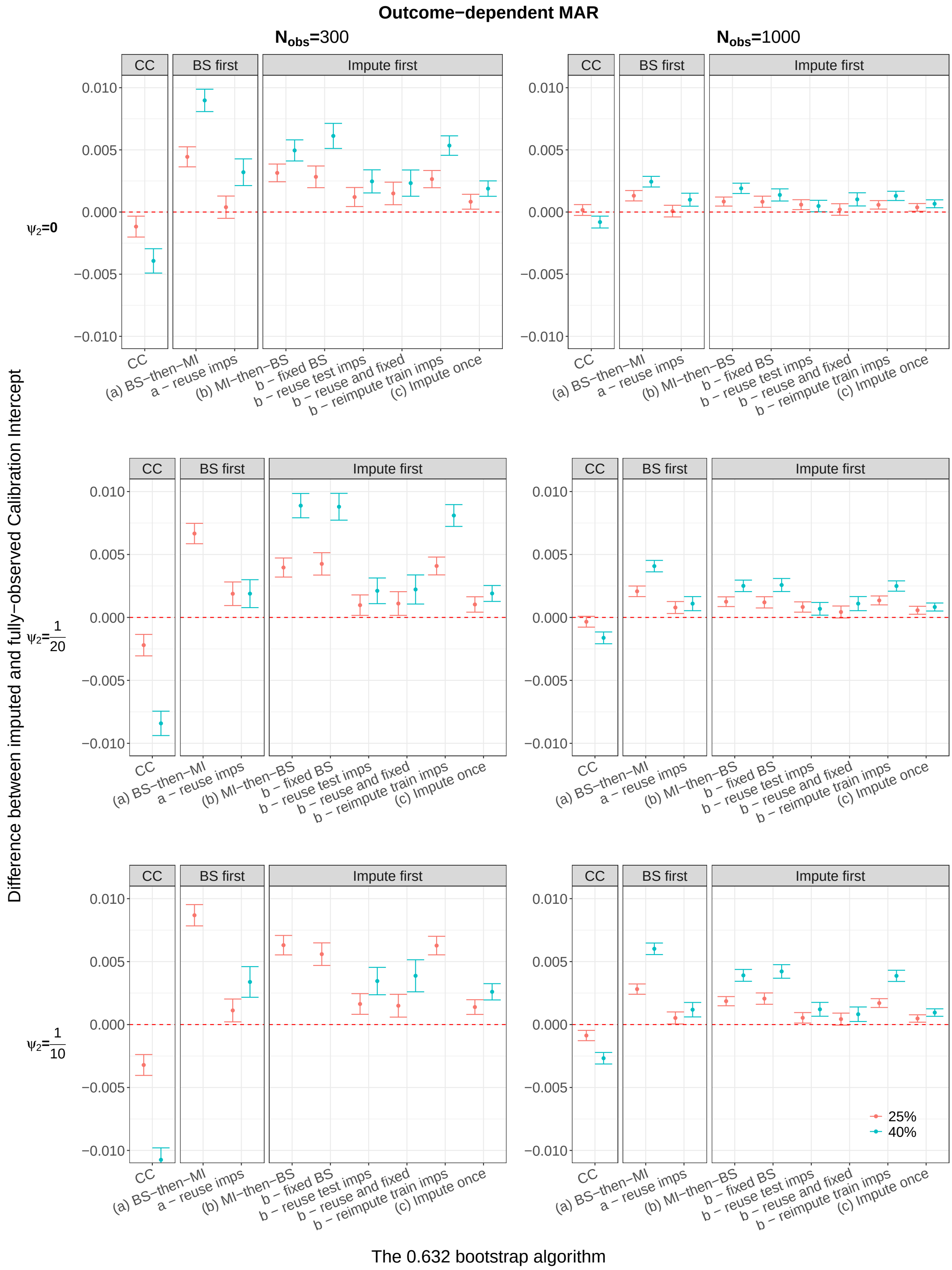

Figure S84: Comparing the impact of increasing the percentage of missingness on the difference Intercept_{imp} - Intercept_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Intercept_{imp} - Intercept_{obs}. Red denotes Intercept_{imp} - Intercept_{obs} when 25% of X_1 values are missing and blue denotes Intercept_{imp} - Intercept_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

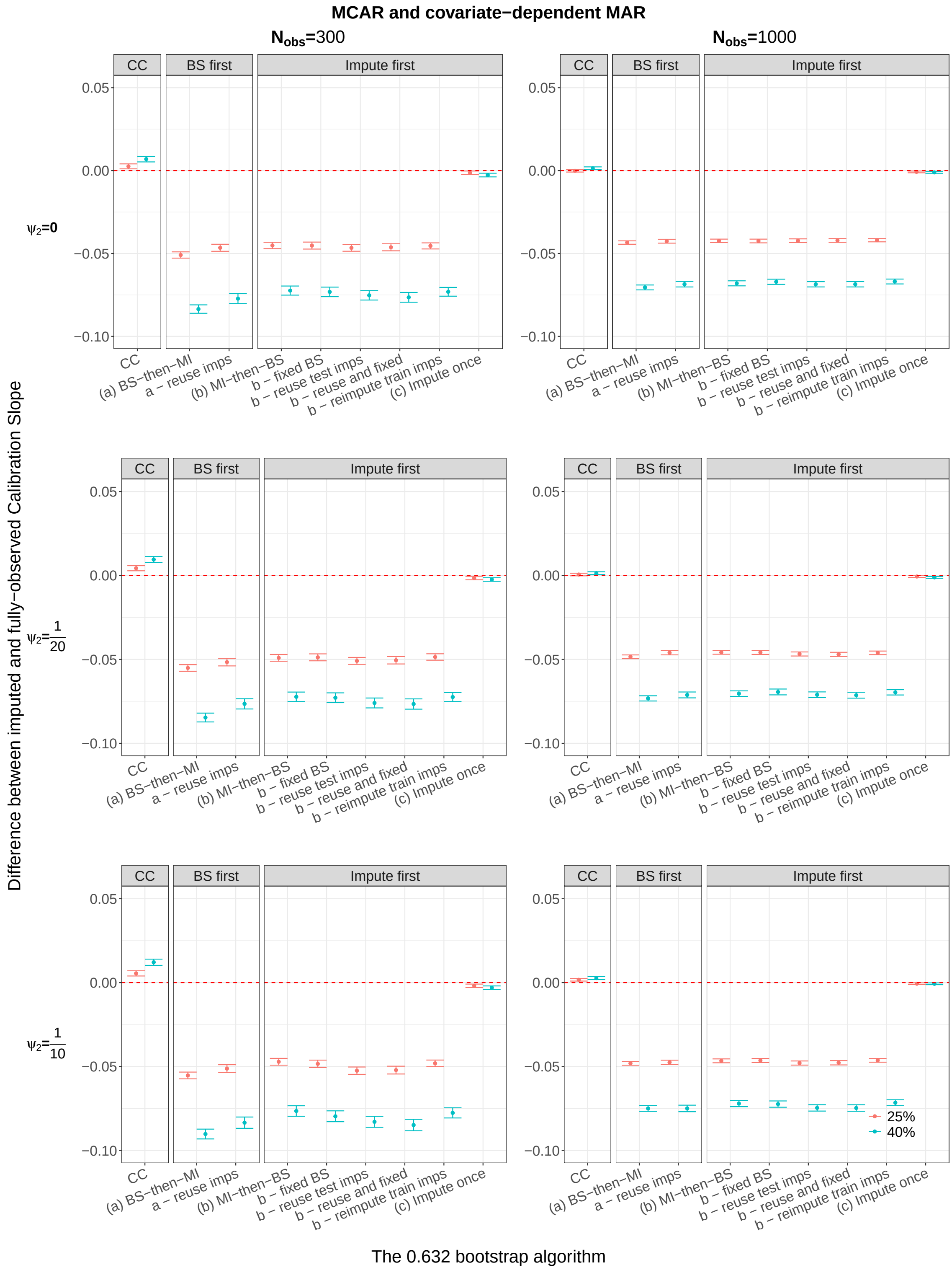

Figure S85: Comparing the impact of increasing the percentage of missingness on the difference Slope_{imp} - Slope_{obs} when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. Red denotes Slope_{imp} - Slope_{obs} when 25% of X_1 values are missing and blue denotes Slope_{imp} - Slope_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

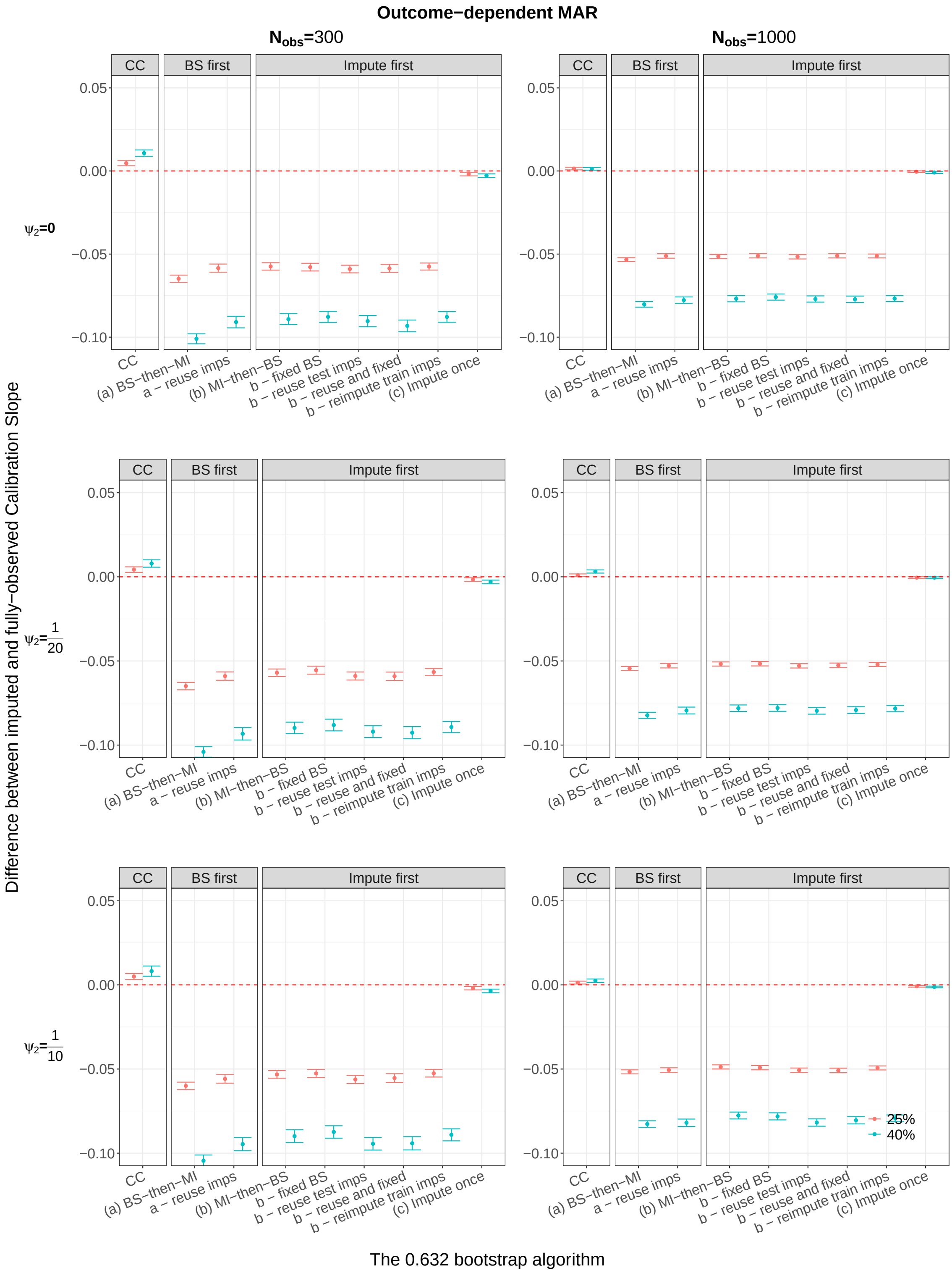

Figure S86: Comparing the impact of increasing the percentage of missingness on the difference Slope_{imp} - Slope_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for pragmatic performance and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. Red denotes Slope_{imp} - Slope_{obs} when 25% of X_1 values are missing and blue denotes Slope_{imp} - Slope_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

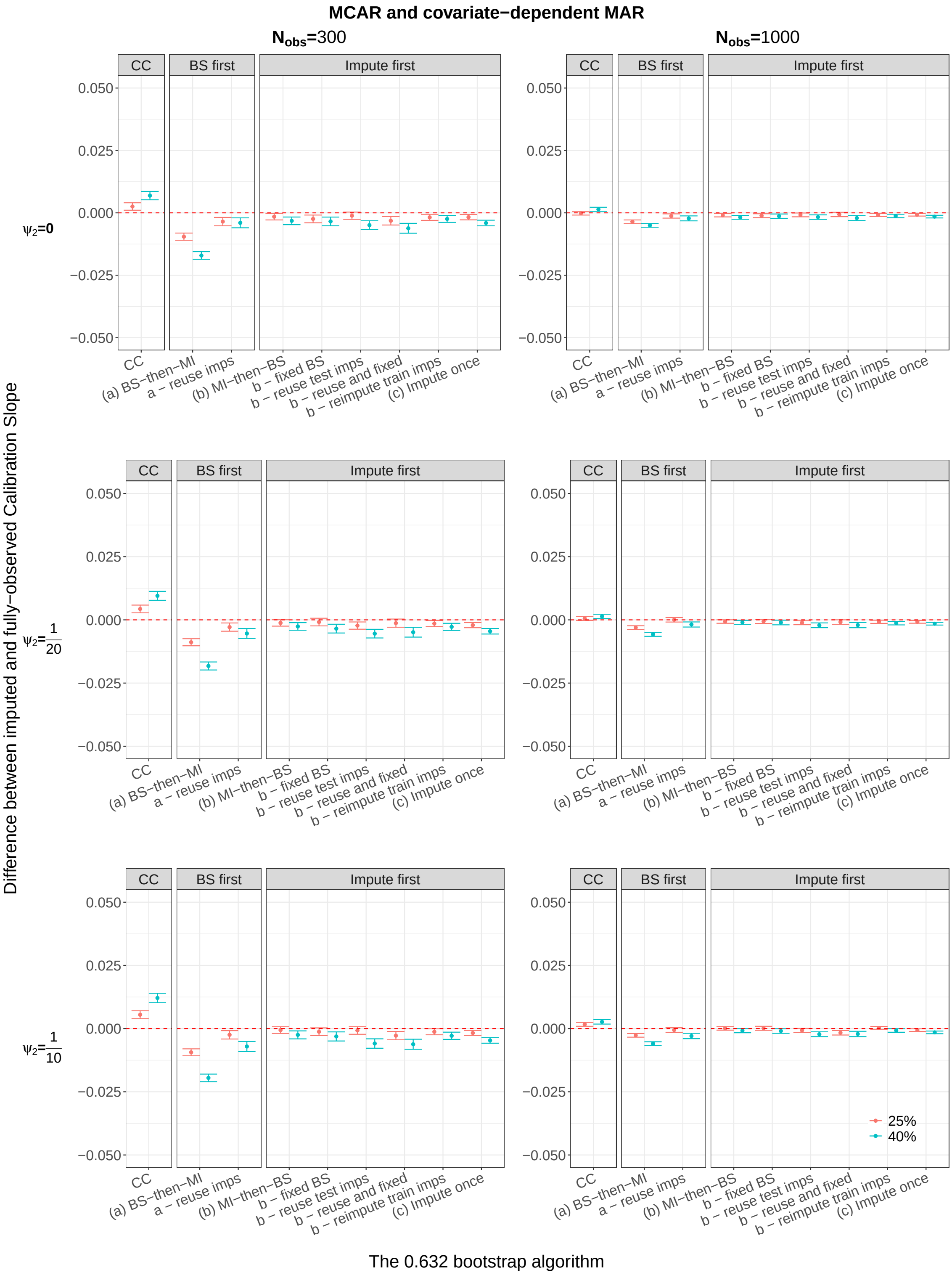

Figure S87: Comparing the impact of increasing the percentage of missingness on the difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. Red denotes Slope$_{imp}$ - Slope$_{obs}$ when 25% of X_1 values are missing and blue denotes Slope$_{imp}$ - Slope$_{obs}$ when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

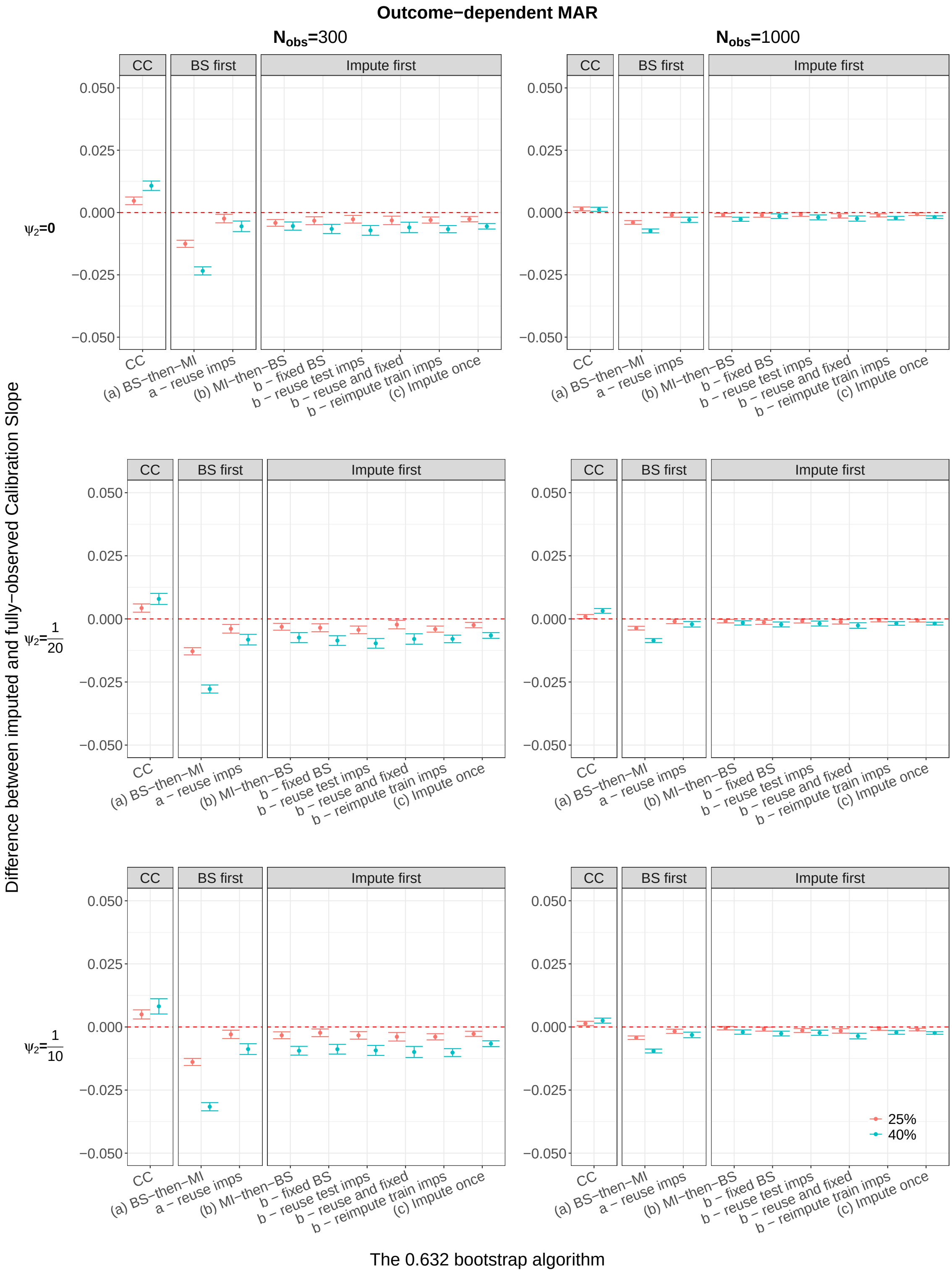

Figure S88: Comparing the impact of increasing the percentage of missingness on the difference Slope_{imp} - Slope_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR when $M = 5$. The error bars summarise results from the 2000 repetitions for ideal performance and the limits represent the Monte Carlo 95% confidence interval of Slope_{imp} - Slope_{obs}. Red denotes Slope_{imp} - Slope_{obs} when 25% of X_1 values are missing and blue denotes Slope_{imp} - Slope_{obs} when 40% of X_1 values are missing. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.6.3 Comparing M=5 versus M=25 (Cal_{imp}-Cal_{obs})

Intercept

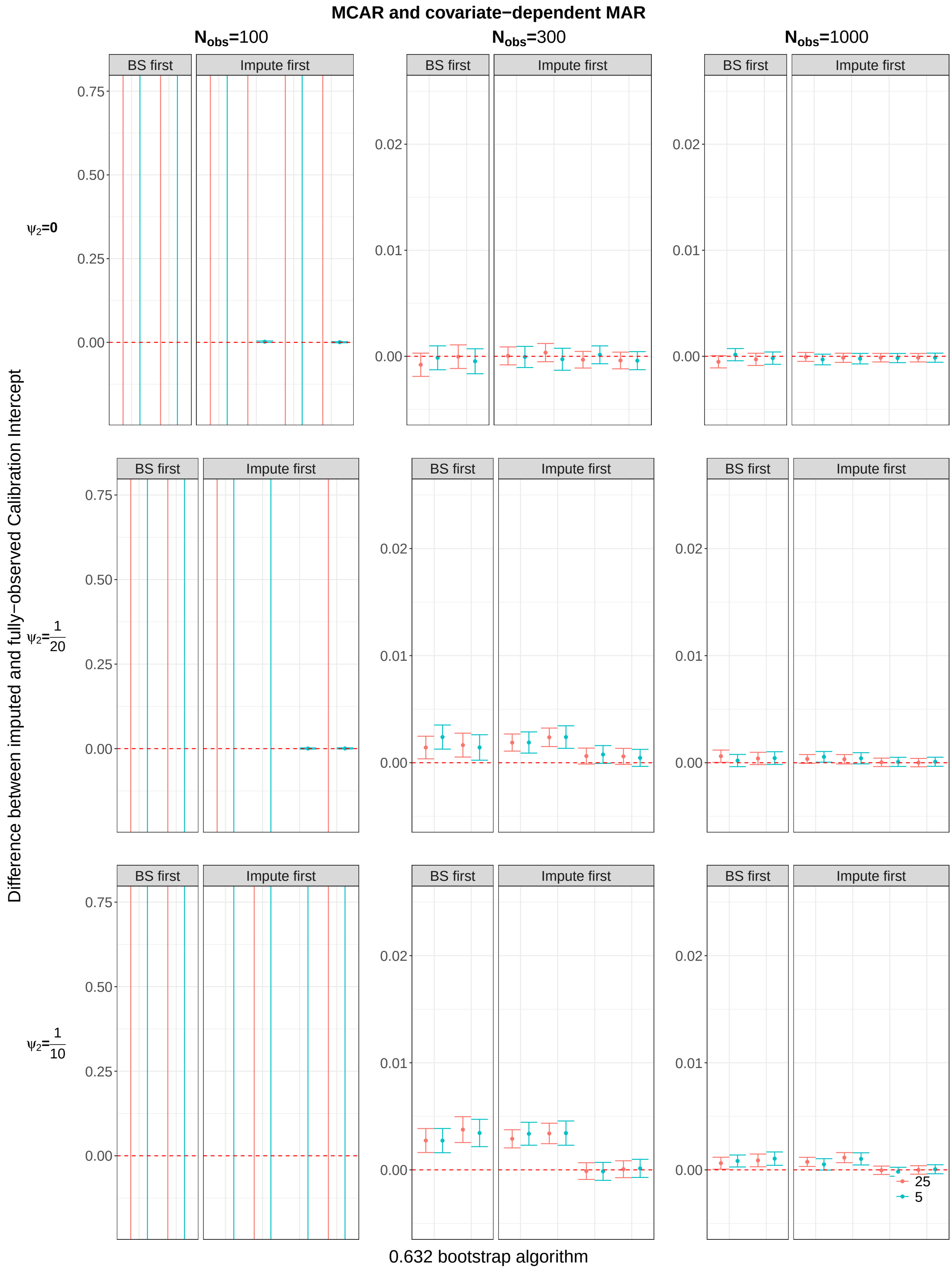

Figure S89: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

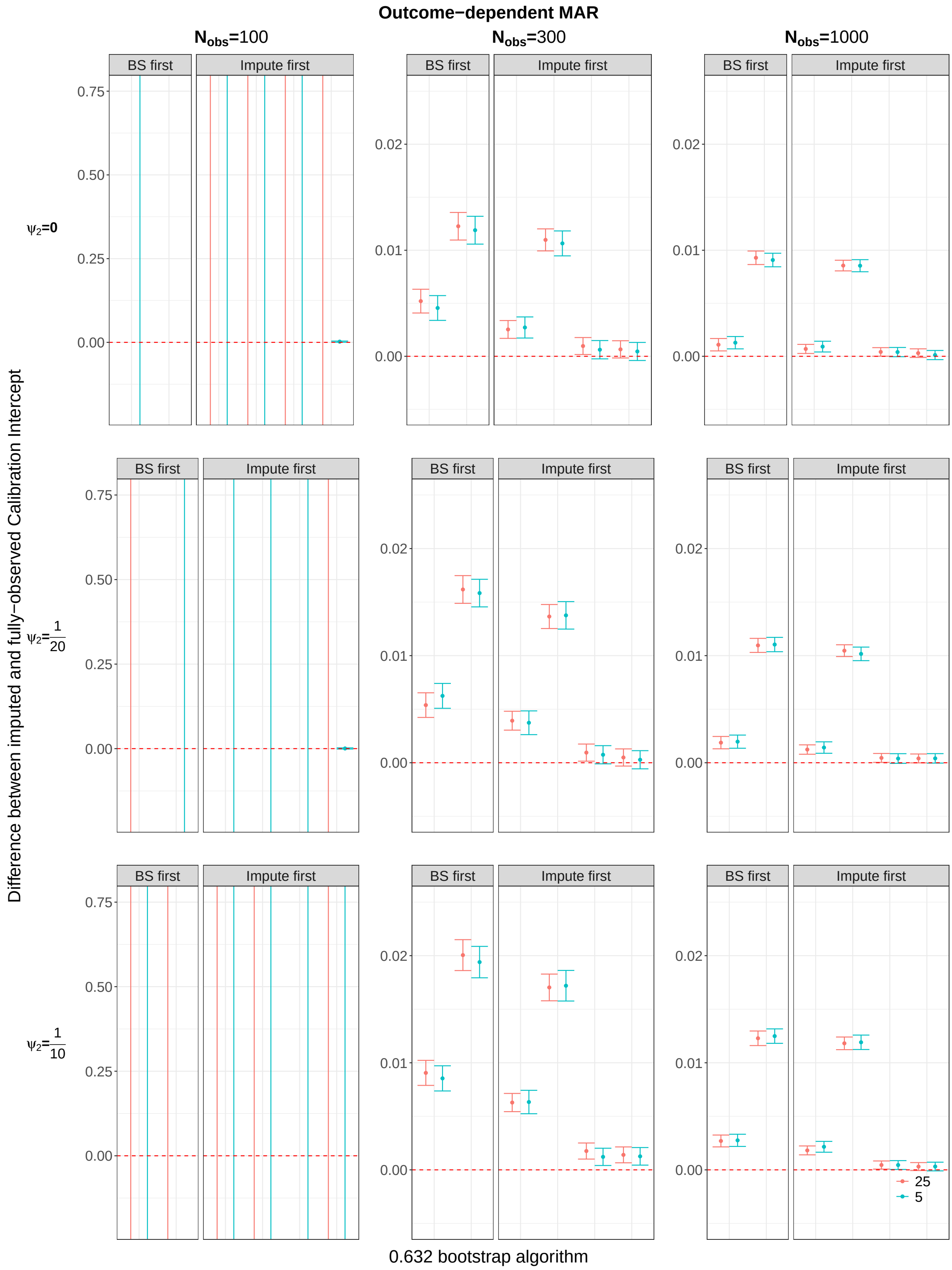

Figure S90: The difference Intercept$_{imp}$ - Intercept$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

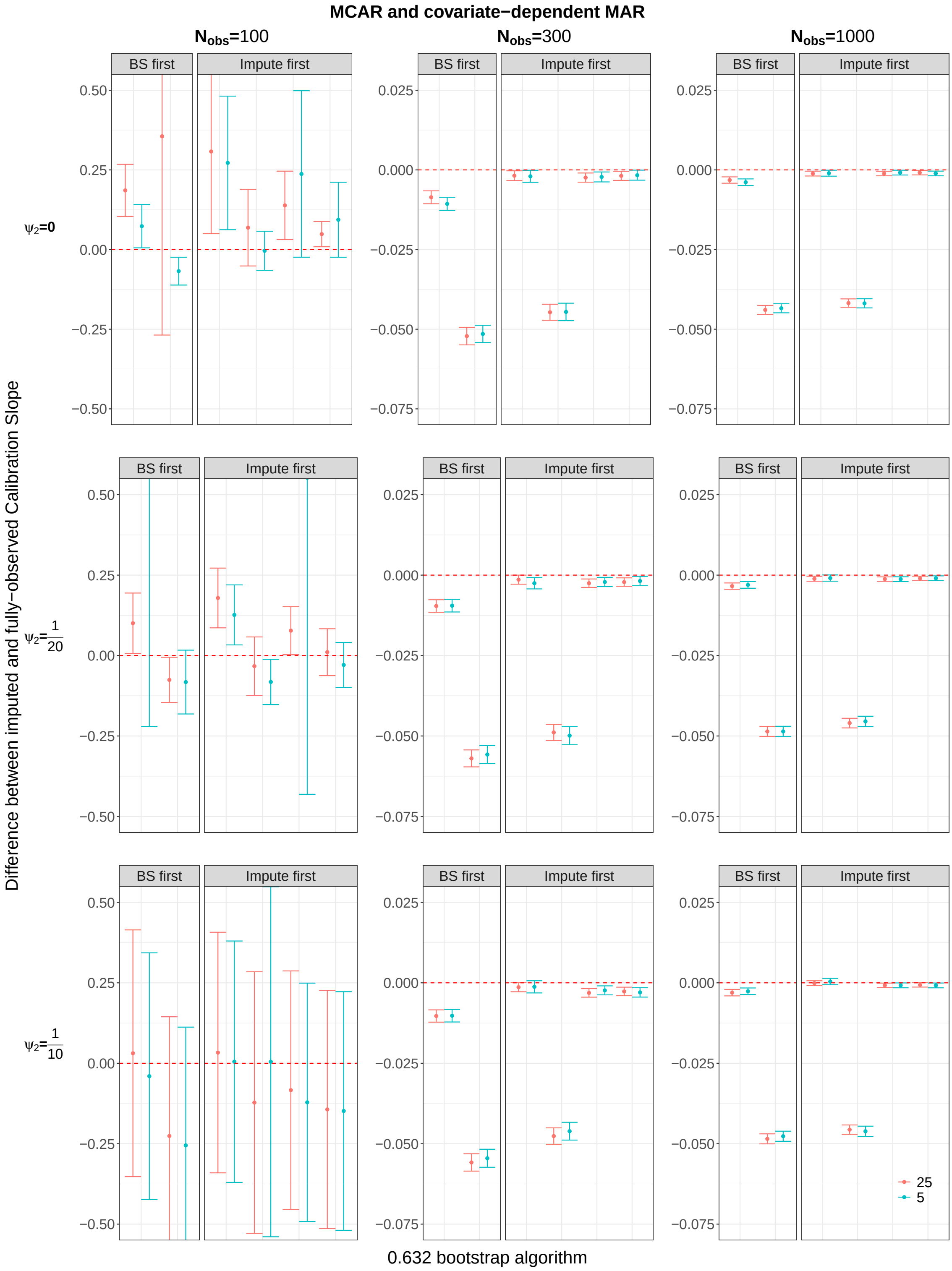

Figure S91: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

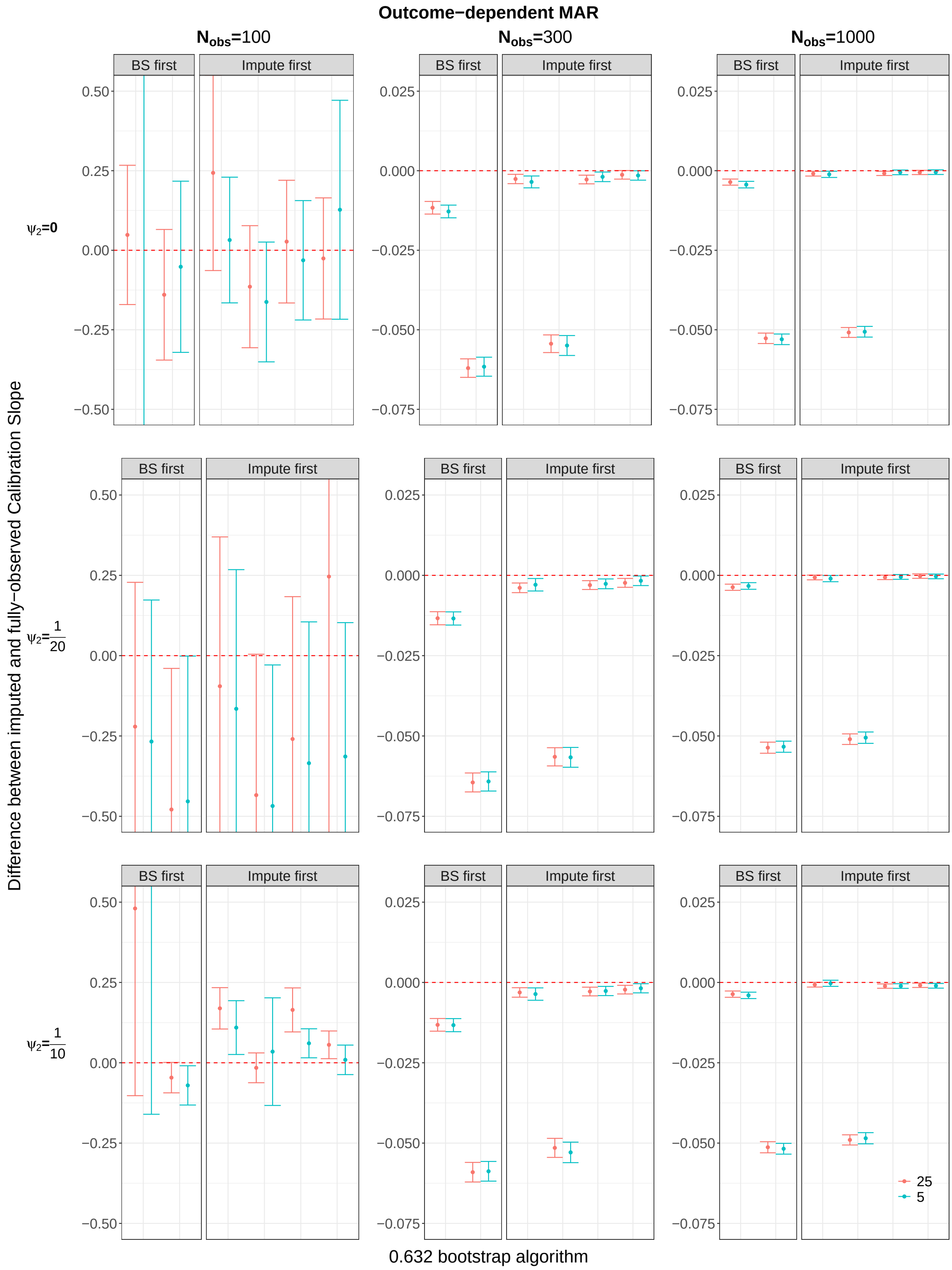

Figure S92: The difference Slope$_{imp}$ - Slope$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 25$ versus $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions for and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{obs}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.6.4 Calibration intercept and slope from imputation methods compared to the target Calibration intercept and slope ($\mathbf{Cal}_{target}$) using a larger validation set

Intercept

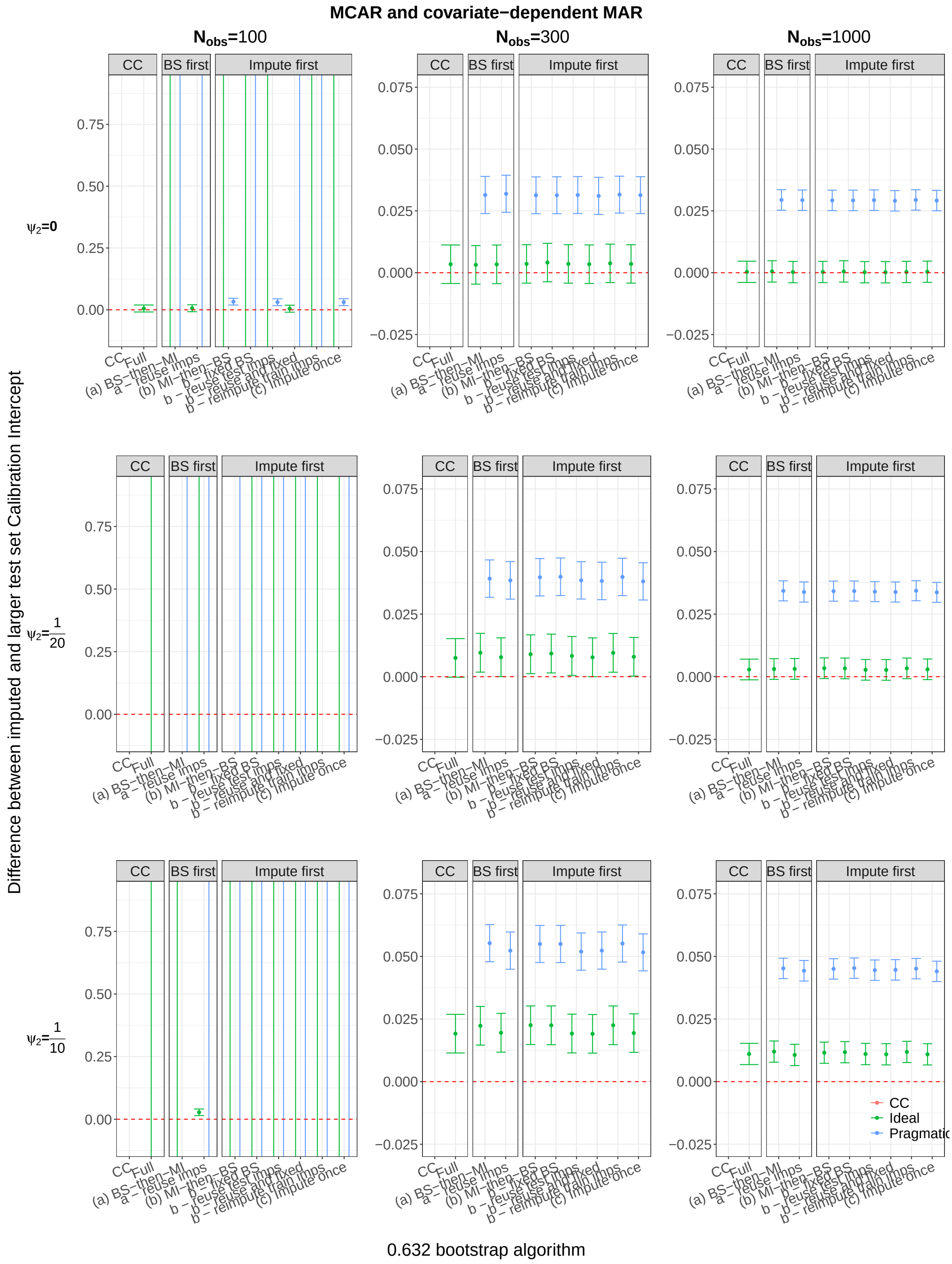

Figure S93: The difference Intercept$_{imp}$ - Intercept$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

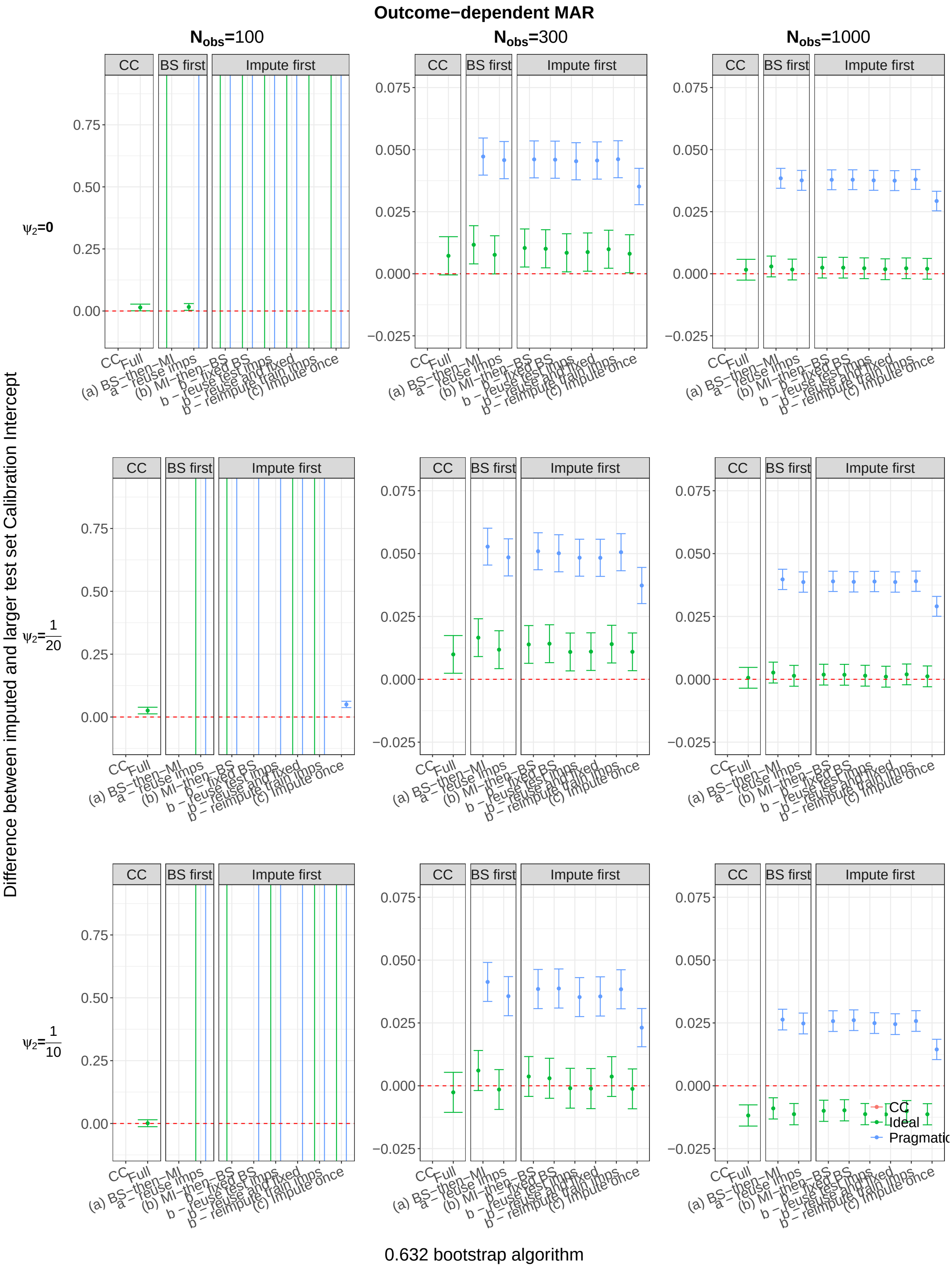

Figure S94: The difference Intercept$_{imp}$ - Intercept$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Intercept$_{imp}$ - Intercept$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

Slope

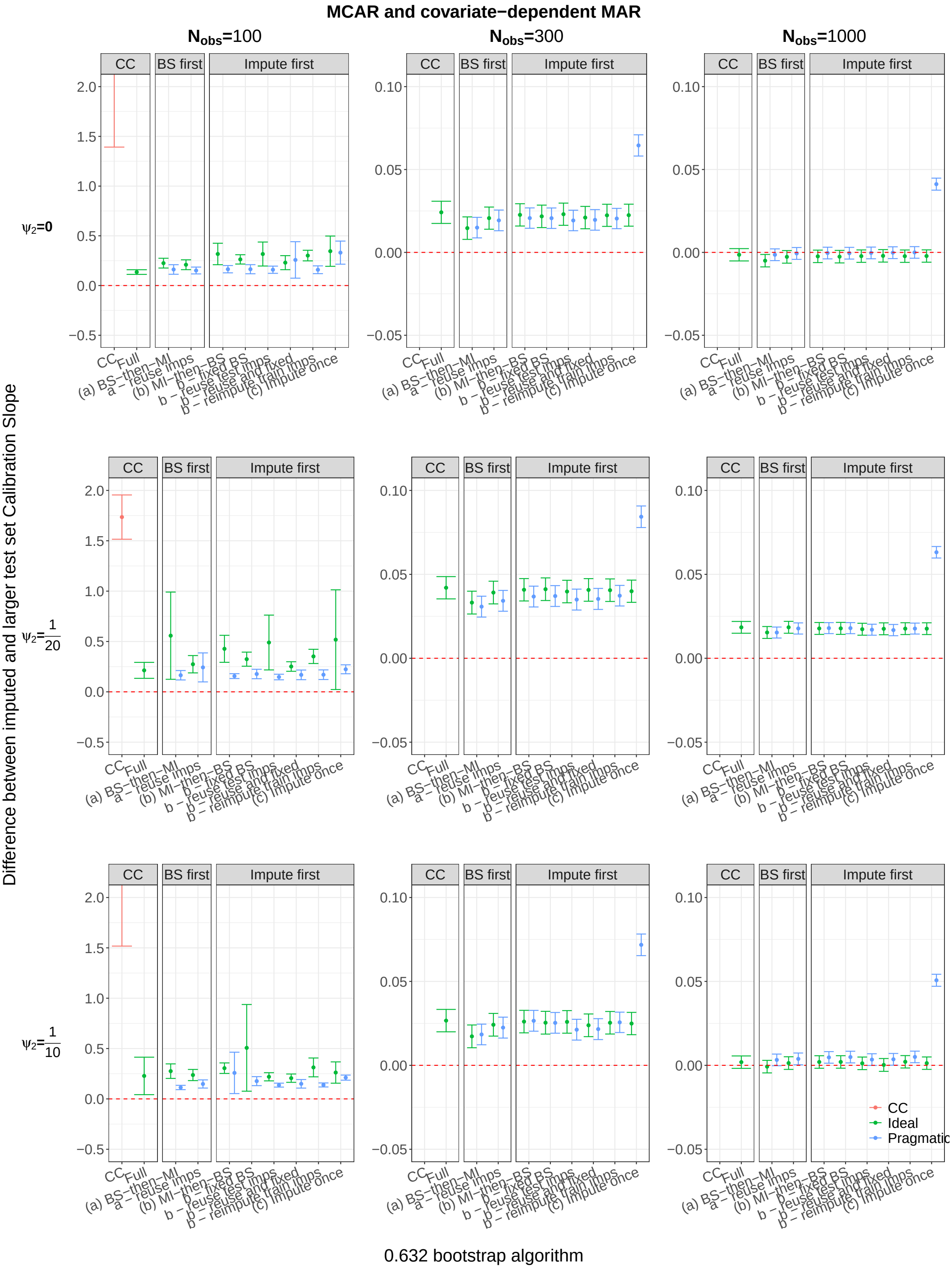

Figure S95: The difference Slope$_{imp}$ - Slope$_{target}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

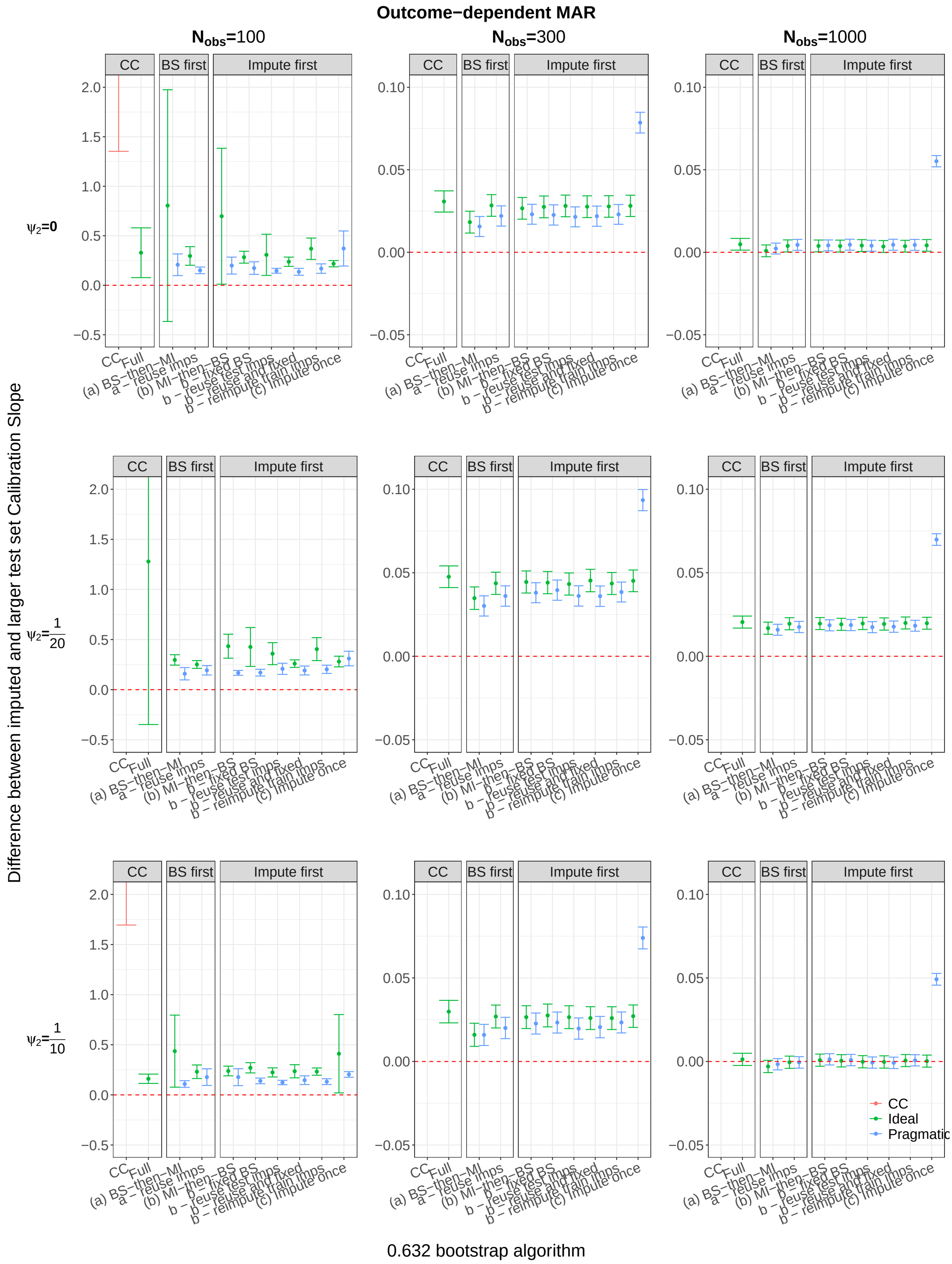

Figure S96: The difference Slope$_{imp}$ - Slope$_{target}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of Slope$_{imp}$ - Slope$_{target}$. CC (complete-case); methods are described in Section 2.7 or Table 6.1.

S4.7 Comparing internal validation methods

S4.7.1 AUC

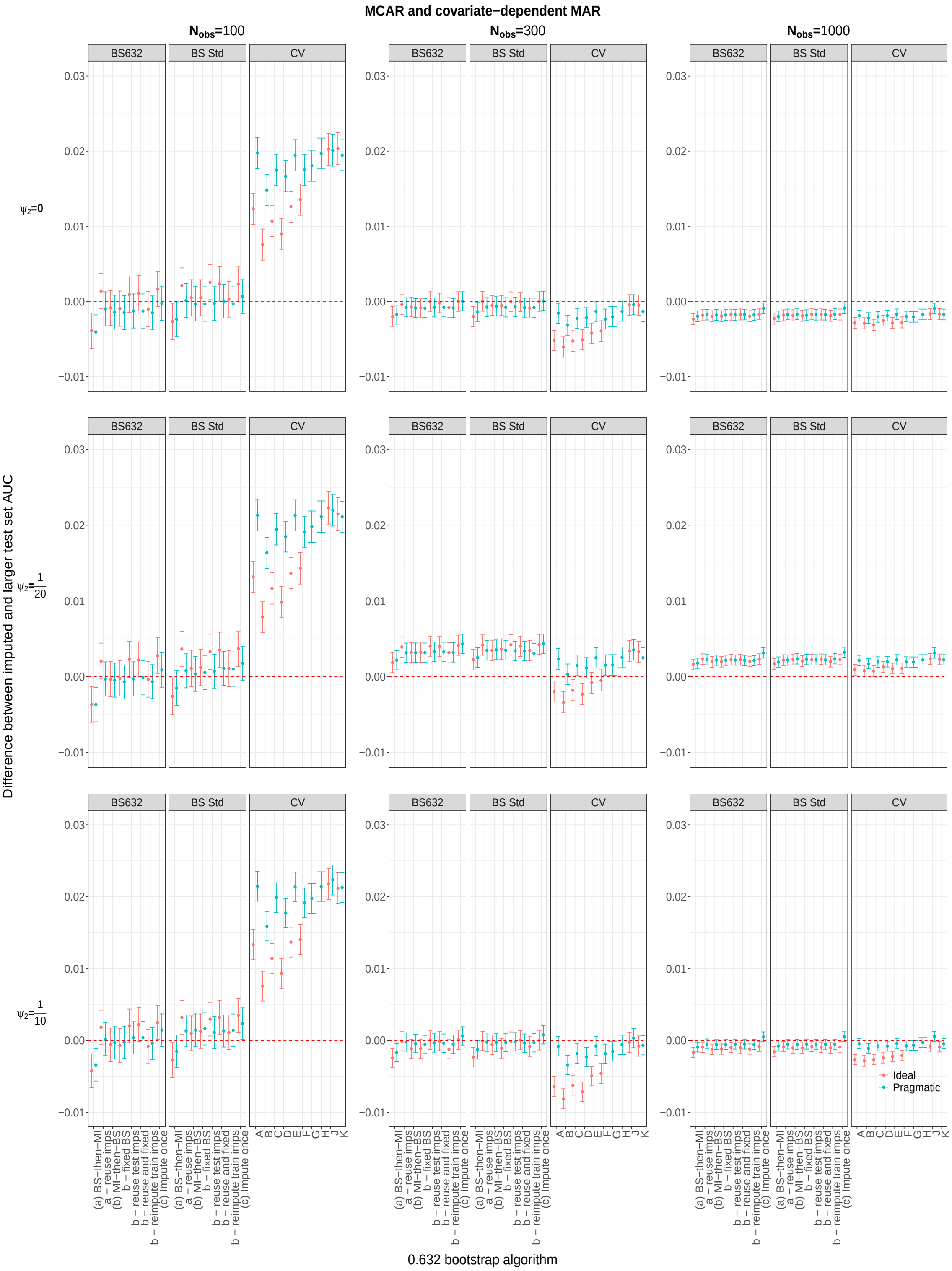

Figure S97: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the AUC. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are MCAR or covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

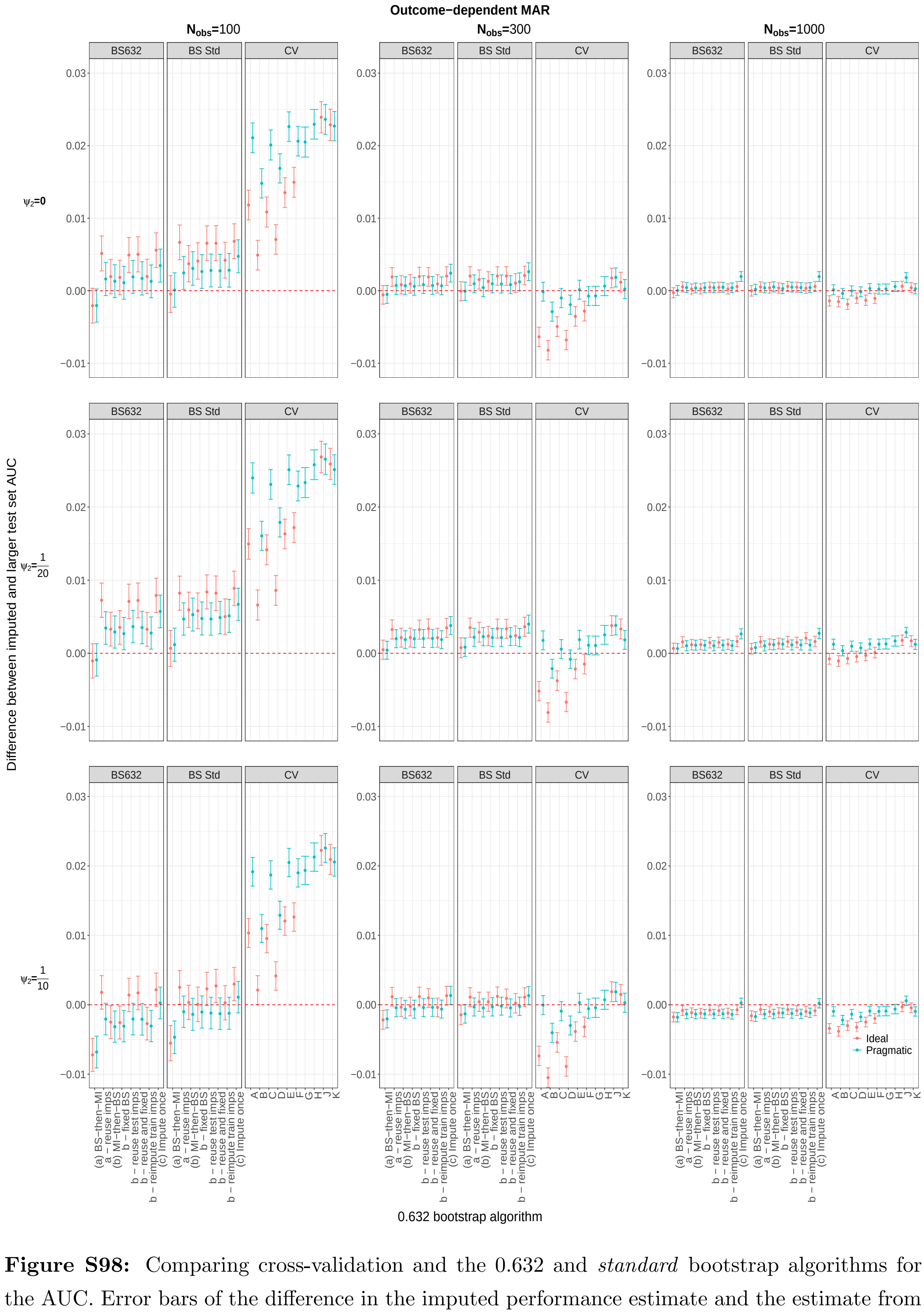

Figure S98: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the AUC. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are outcome-dependent or outcome- and covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

S4.7.2 Brier score

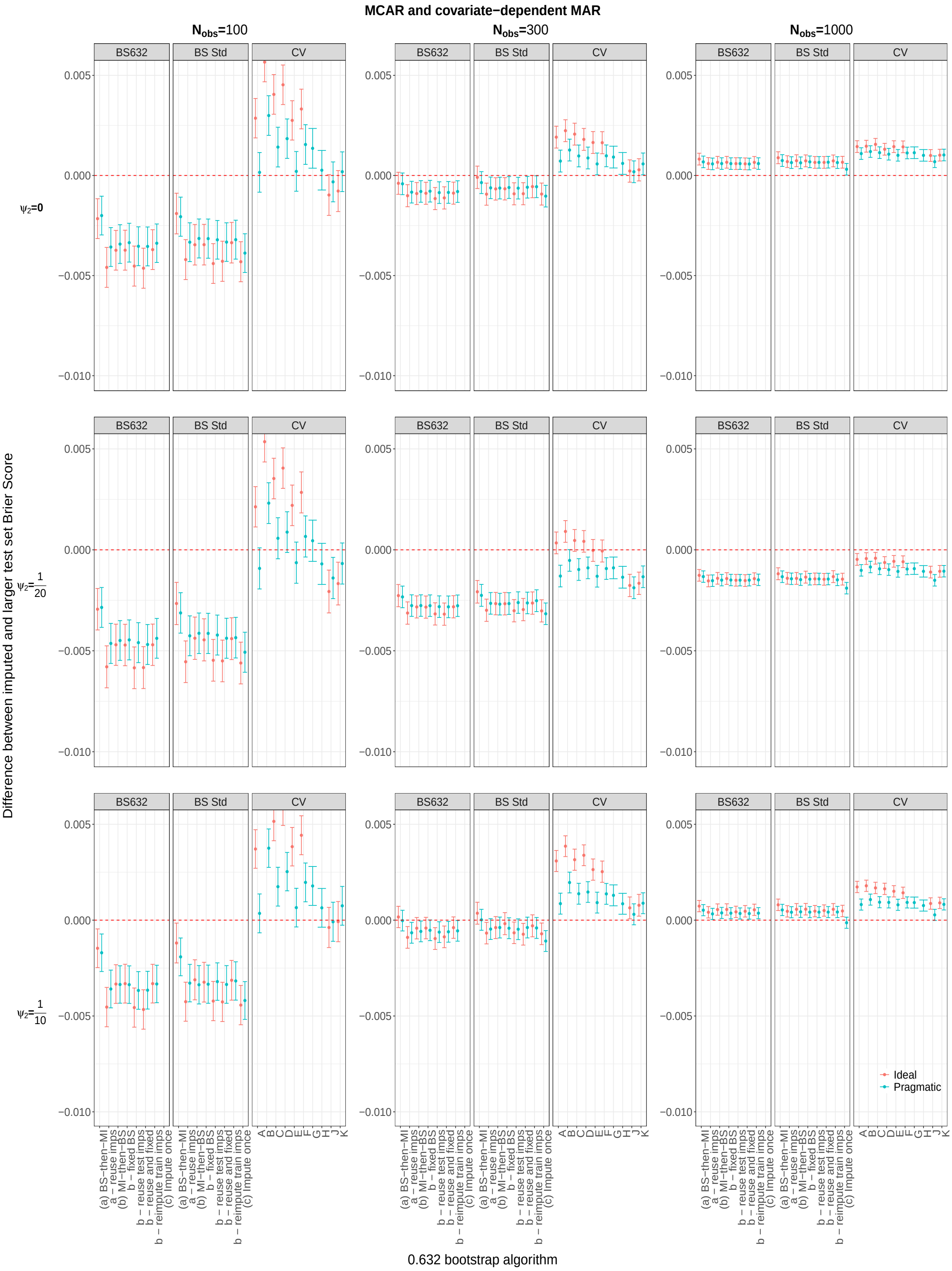

Figure S99: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the Brier score. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are MCAR or covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

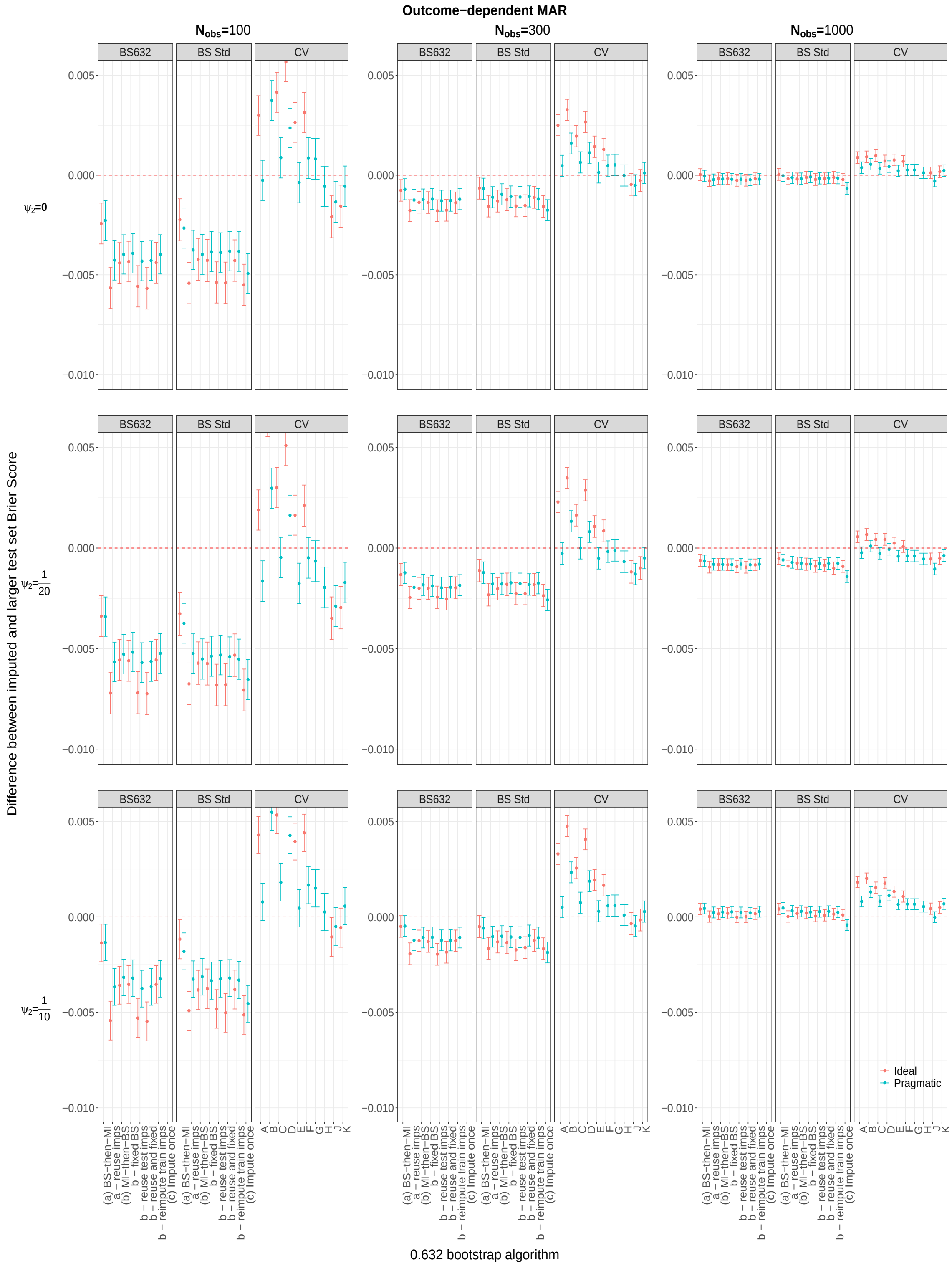

Figure S100: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the Brier score. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are outcome-dependent or outcome- and covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

S4.7.3 Calibration intercept

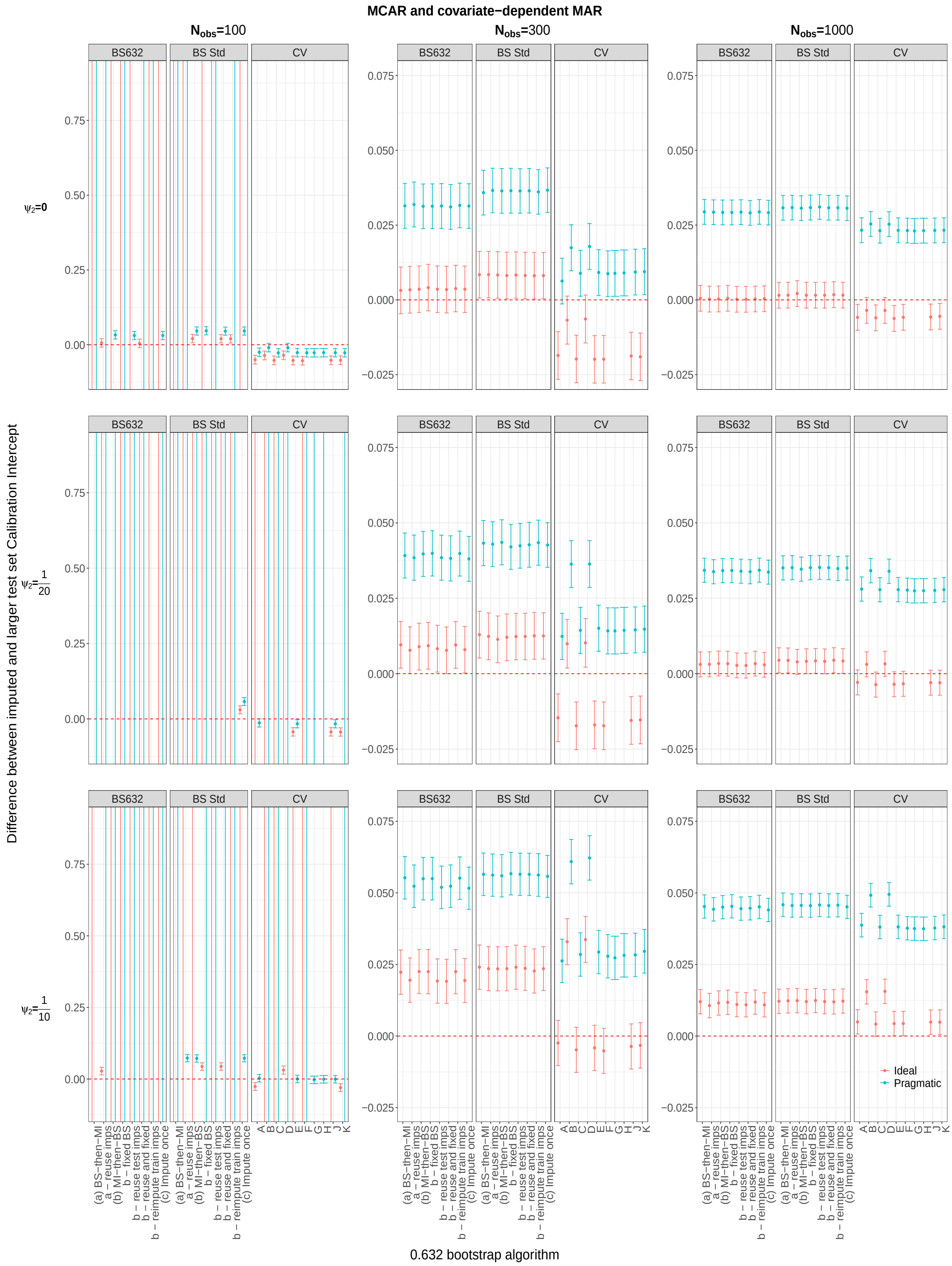

Figure S101: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the calibration intercept. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are MCAR or covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

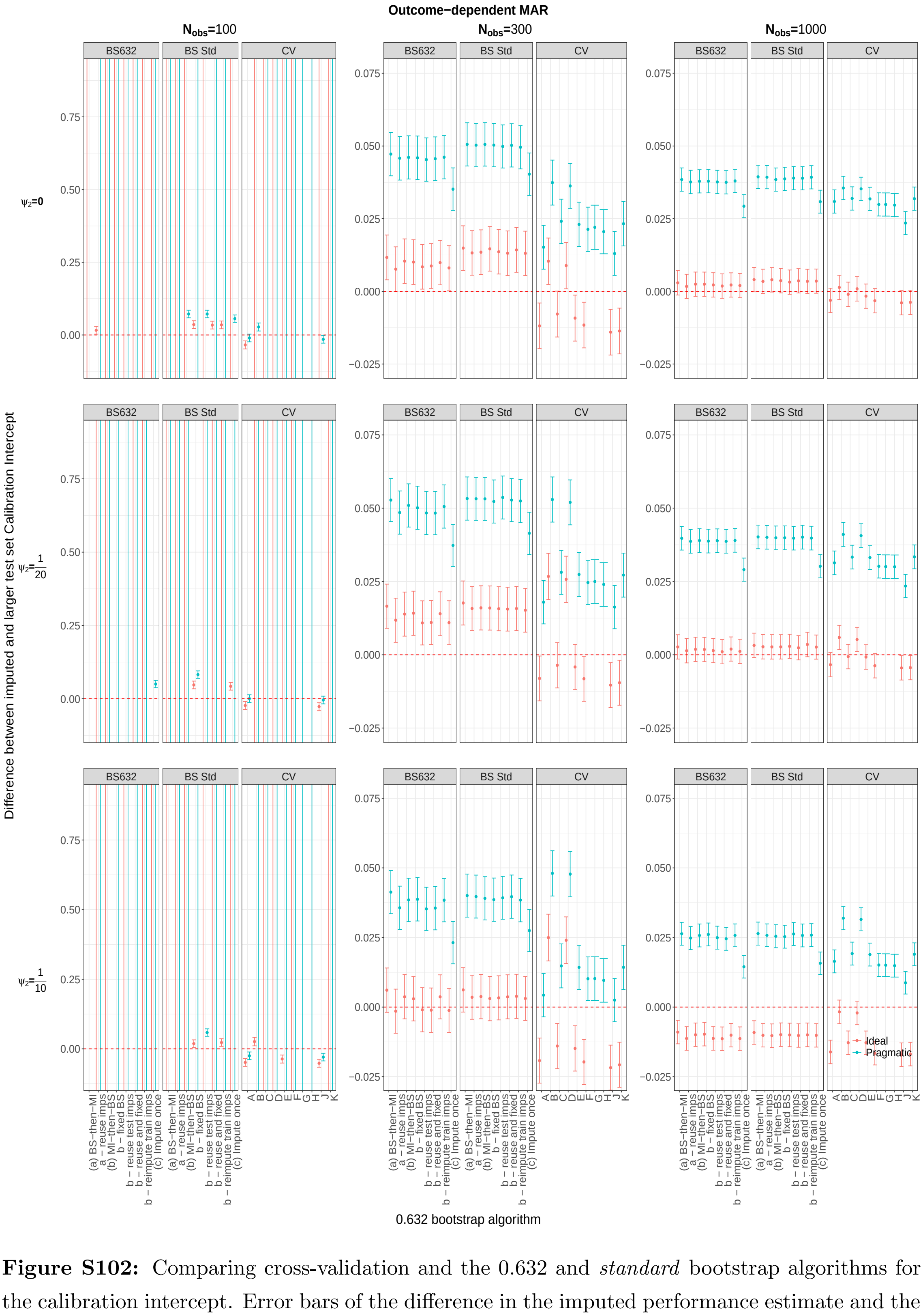

Figure S102: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the calibration intercept. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are outcome-dependent or outcome- and covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

S4.7.4 Calibration slope

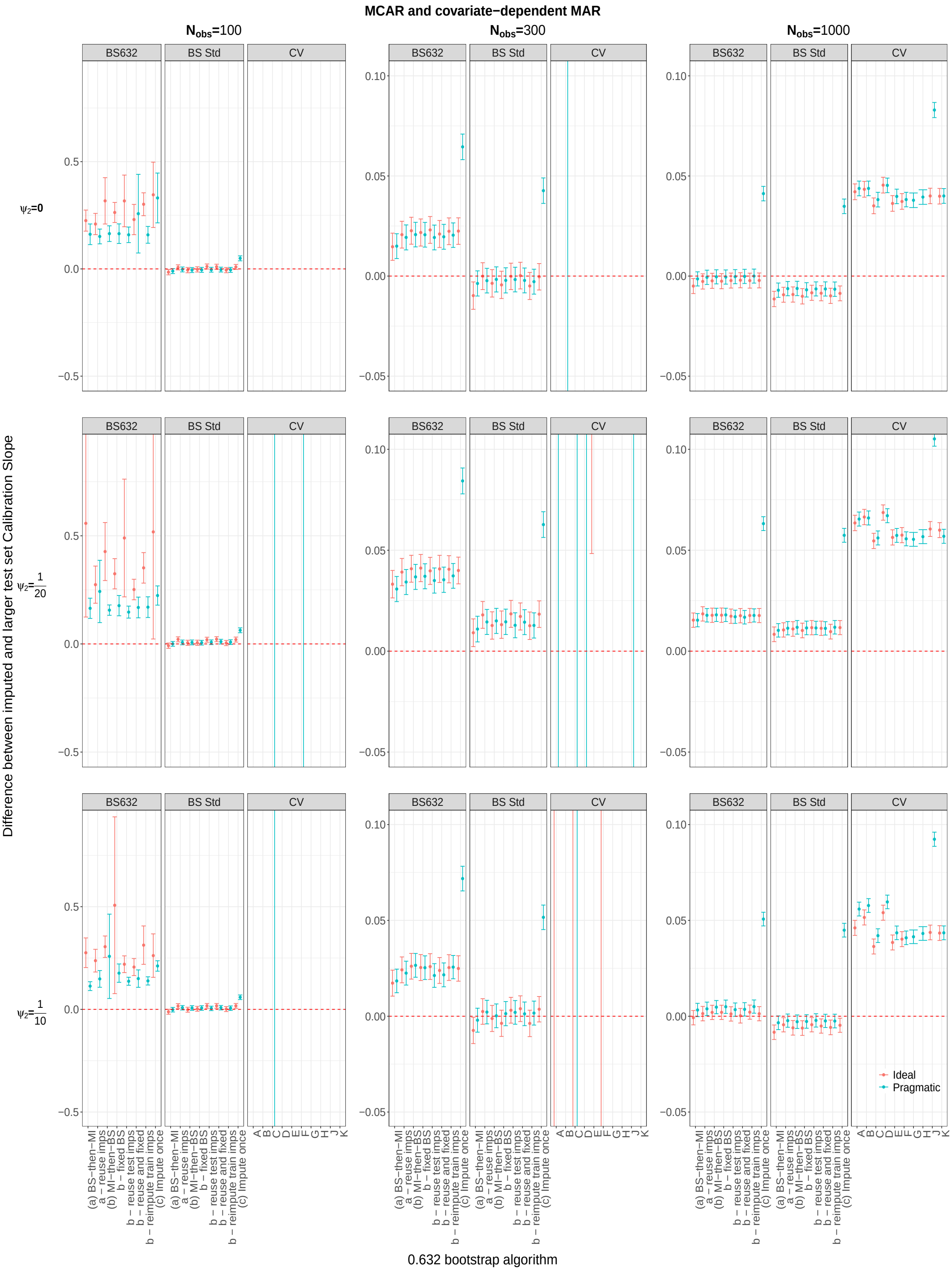

Figure S103: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the calibration slope. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are MCAR or covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

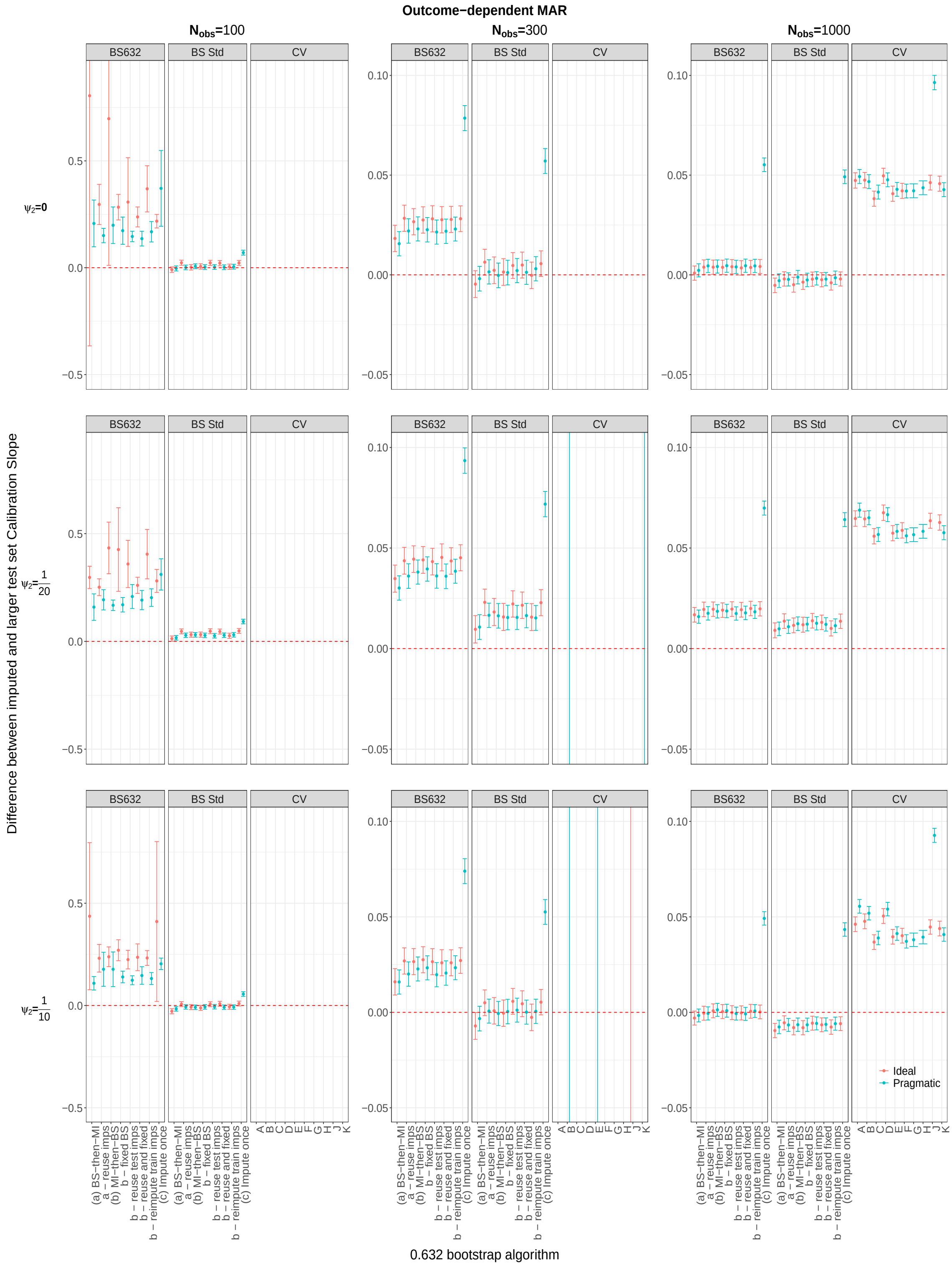

Figure S104: Comparing cross-validation and the 0.632 and *standard* bootstrap algorithms for the calibration slope. Error bars of the difference in the imputed performance estimate and the estimate from a larger test set are presented for the scenario when data are outcome-dependent or outcome- and covariate-dependent MAR. CC (complete-case); CV methods A-K are described in Table 2.3; bootstrap methods are described in Section 2.7 or Table 6.1.

S5 Chapter 9: Simulation study results for FPS, comparison of MSE (Section 9.3)

S5.1 Cross-validation

S5.1.1 No origin shift transformation applied

True exponent is 0

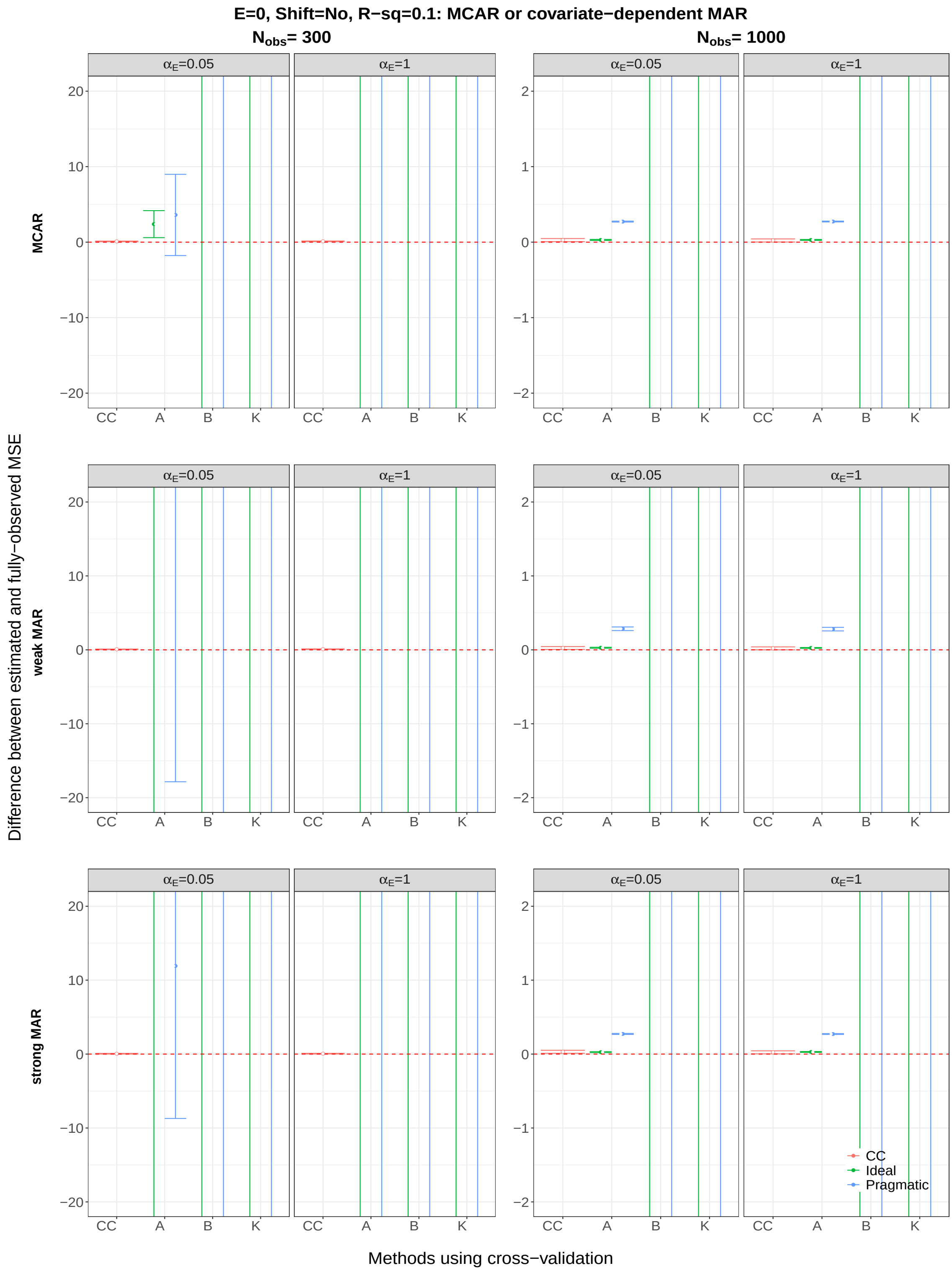

Figure S1: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

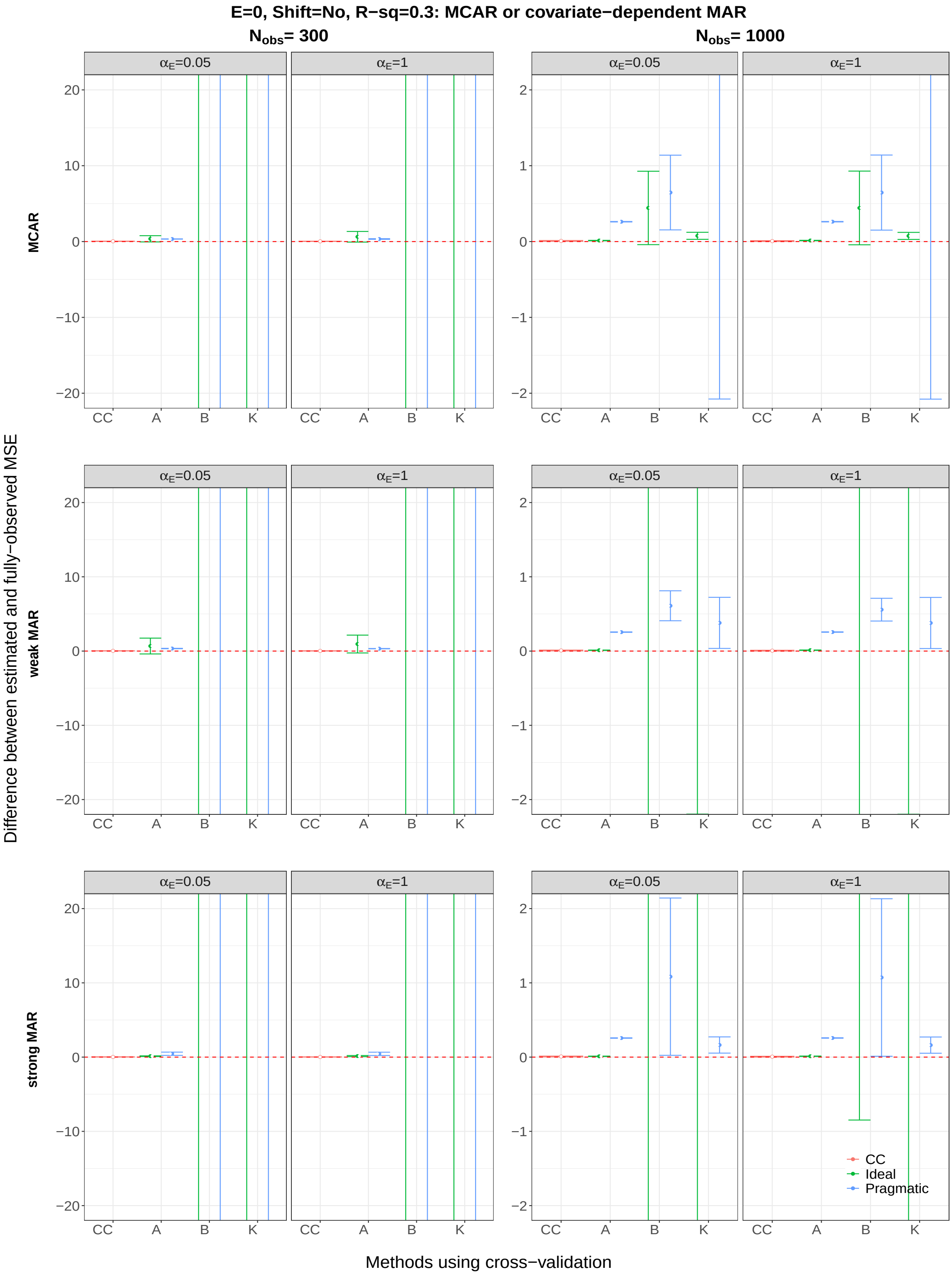

Figure S2: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

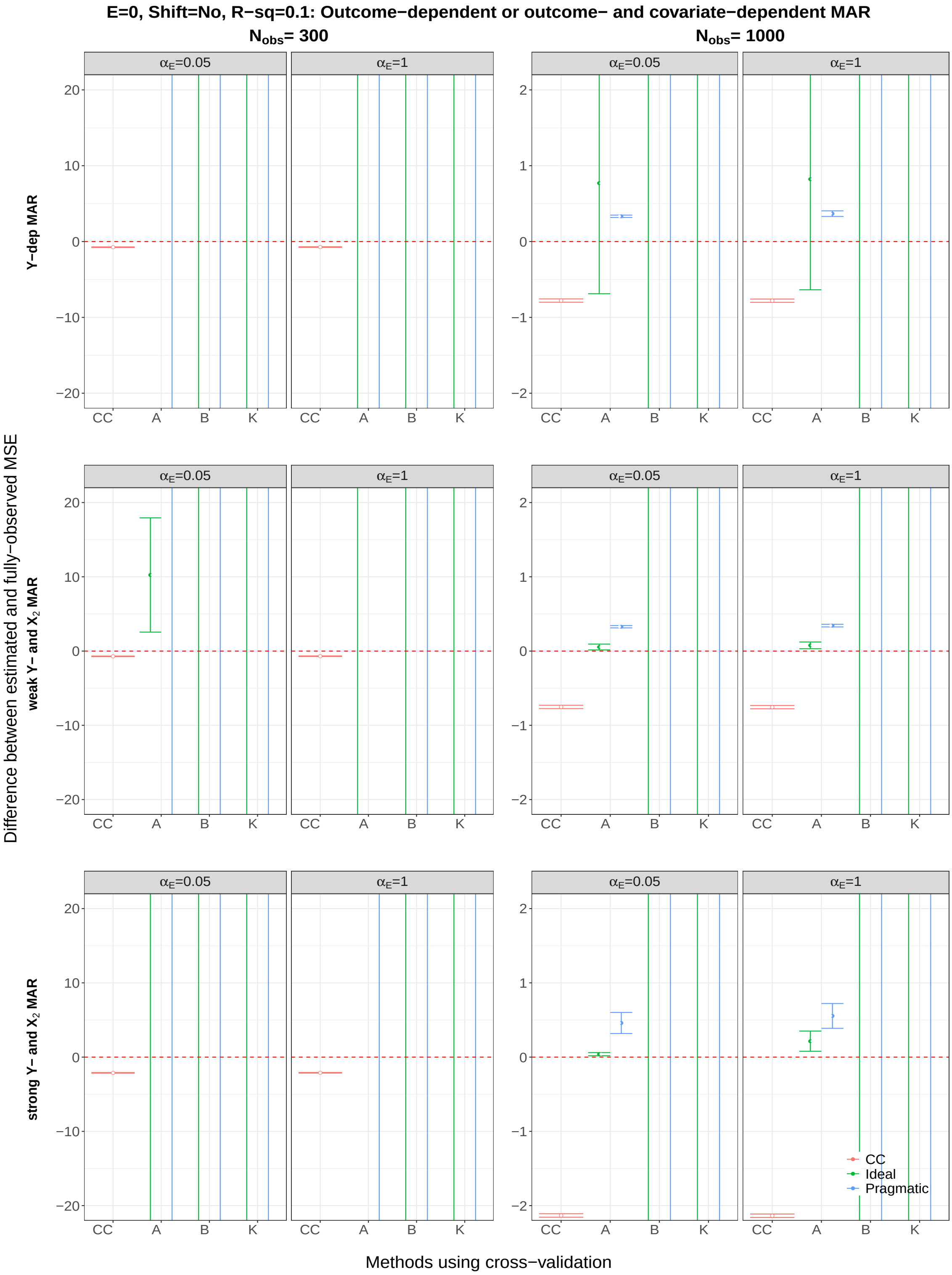

Figure S3: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

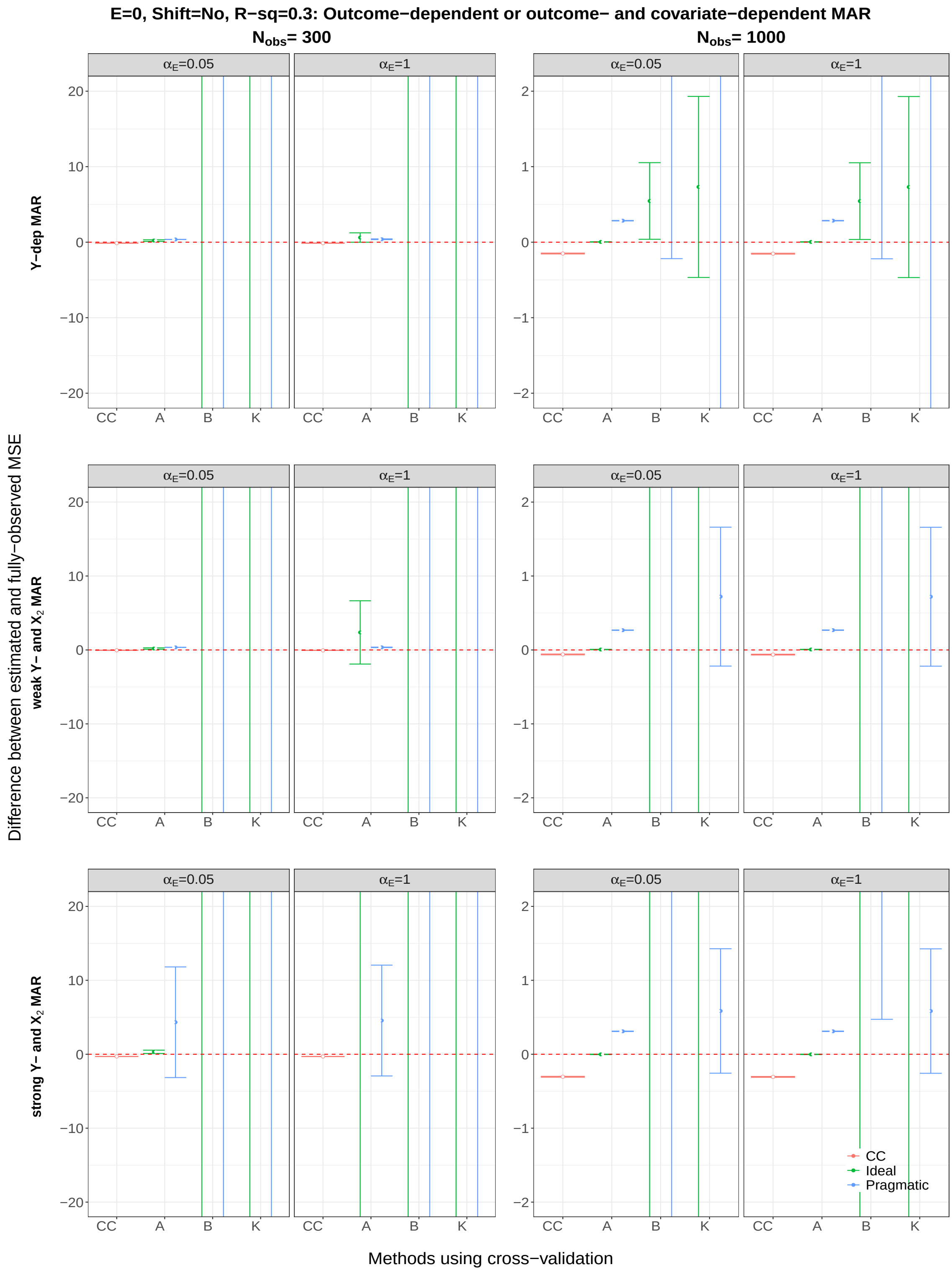

Figure S4: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

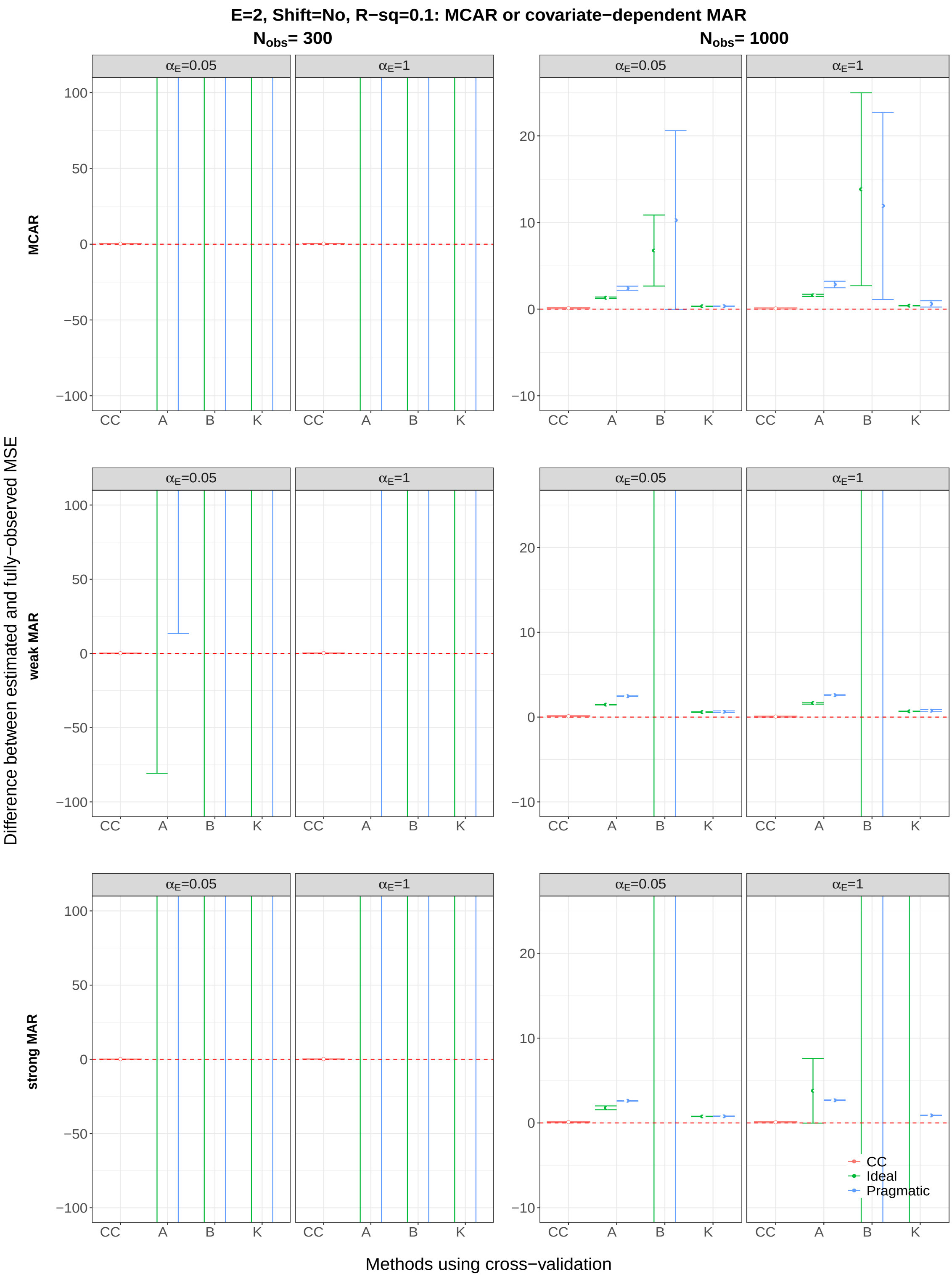

Figure S5: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

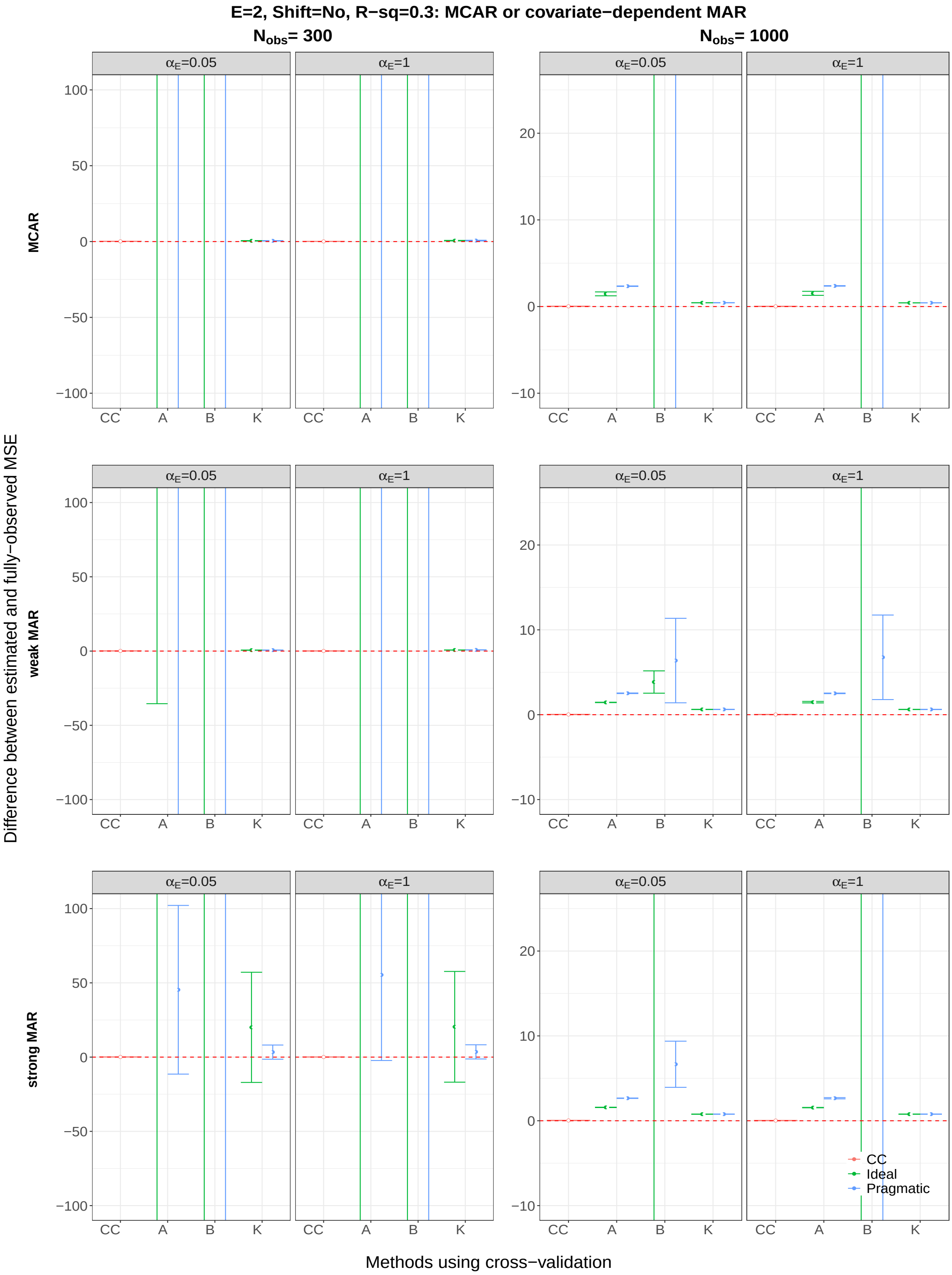

Figure S6: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

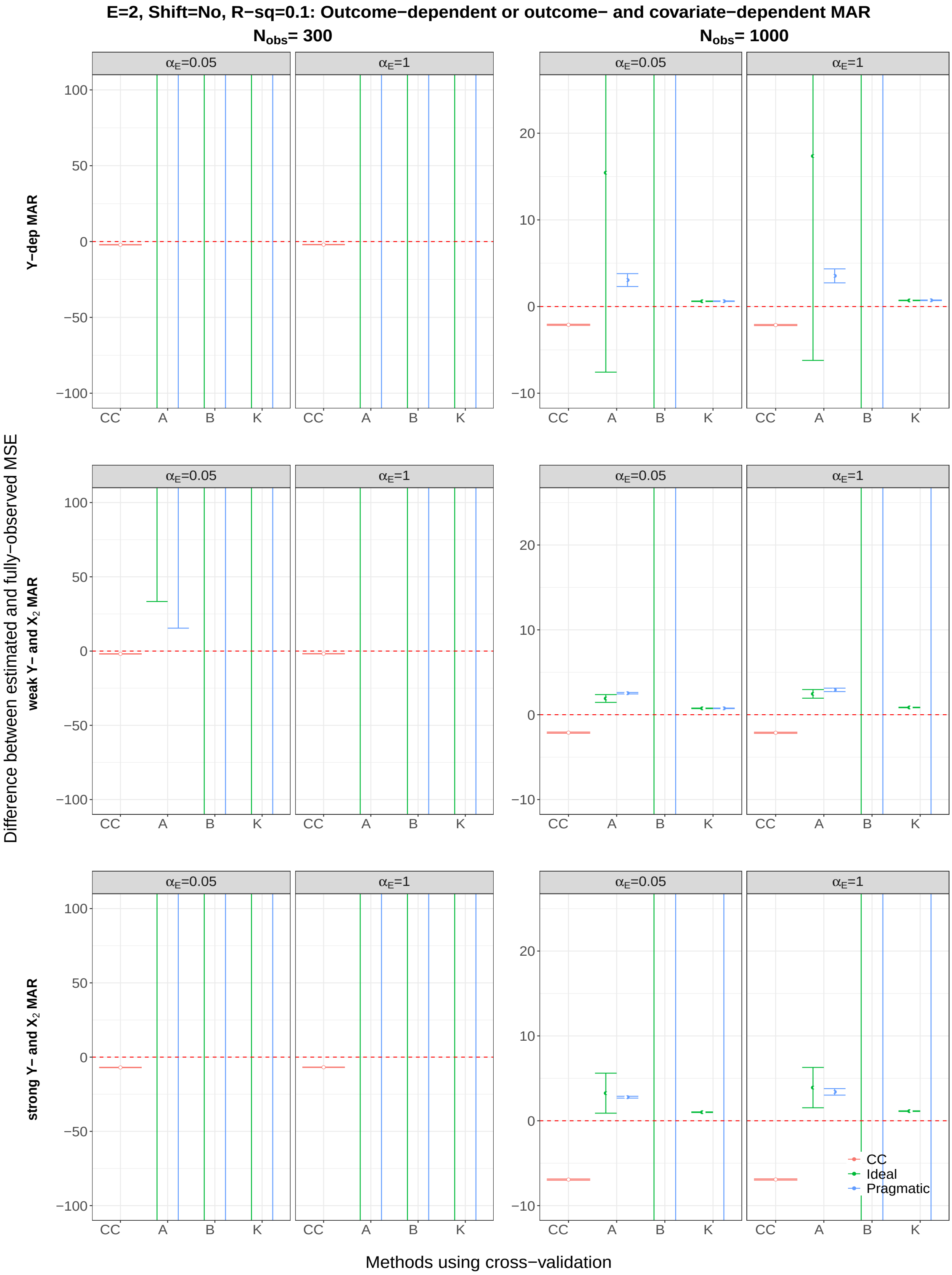

Figure S7: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

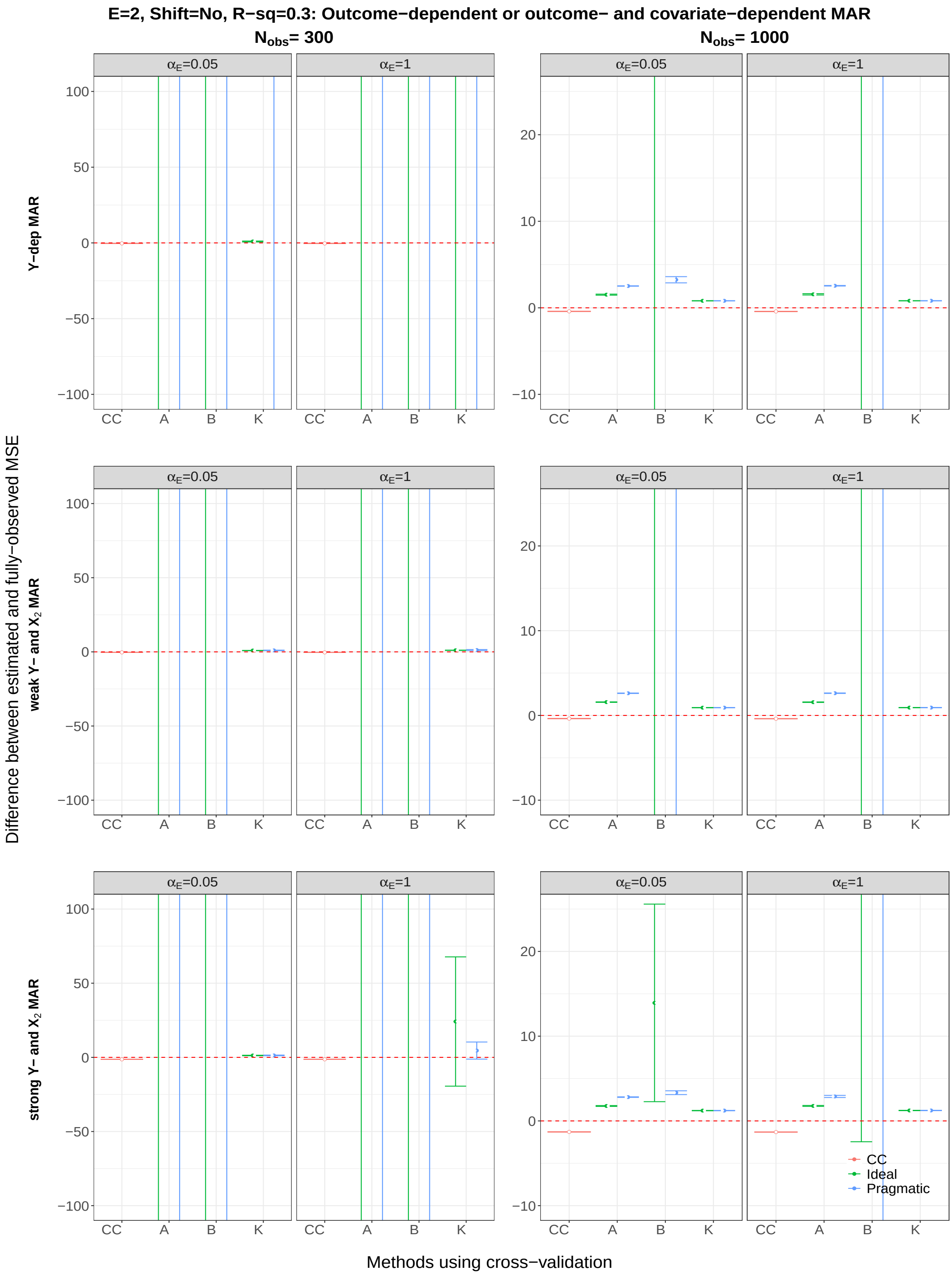

Figure S8: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

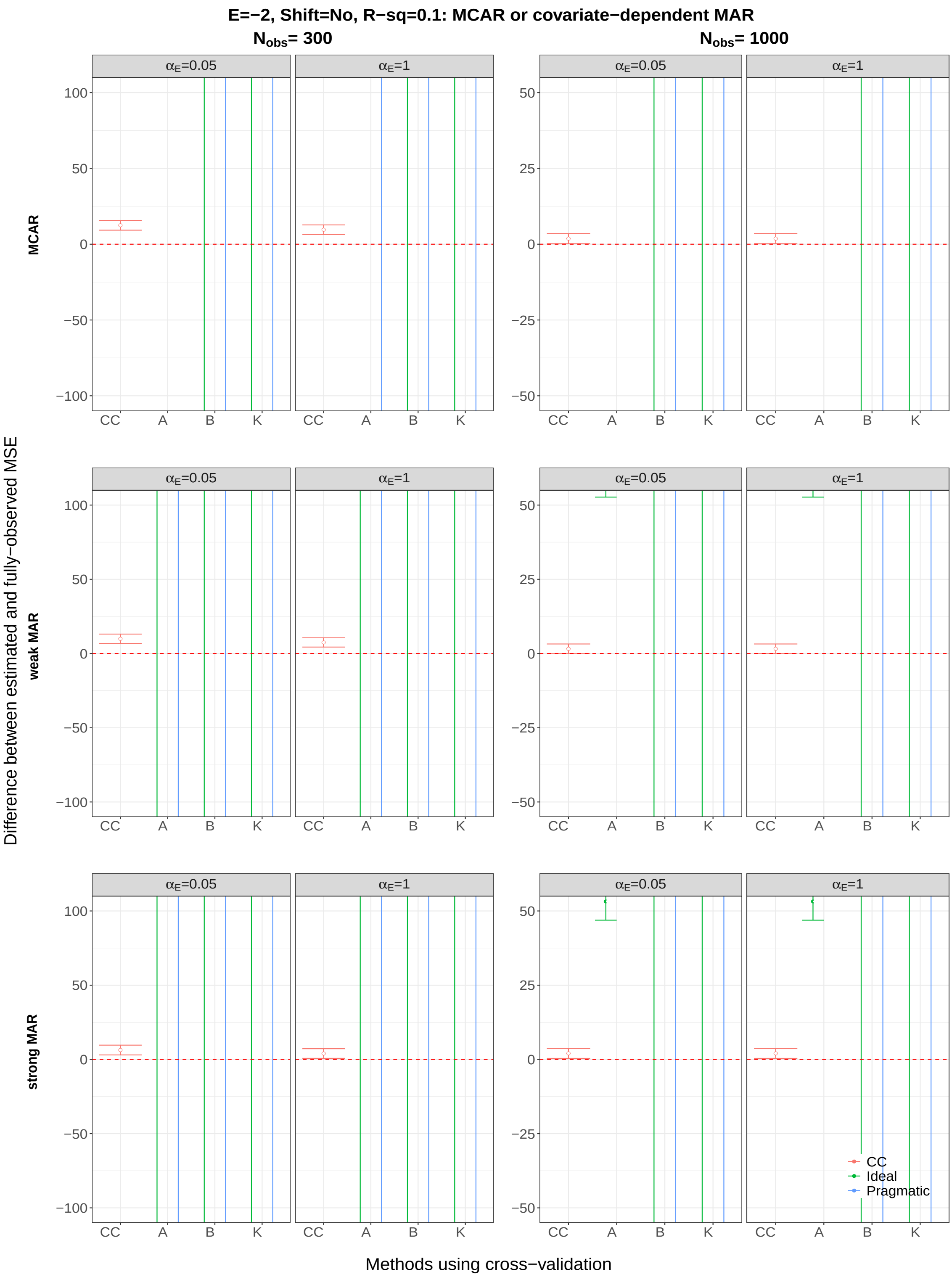

Figure S9: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

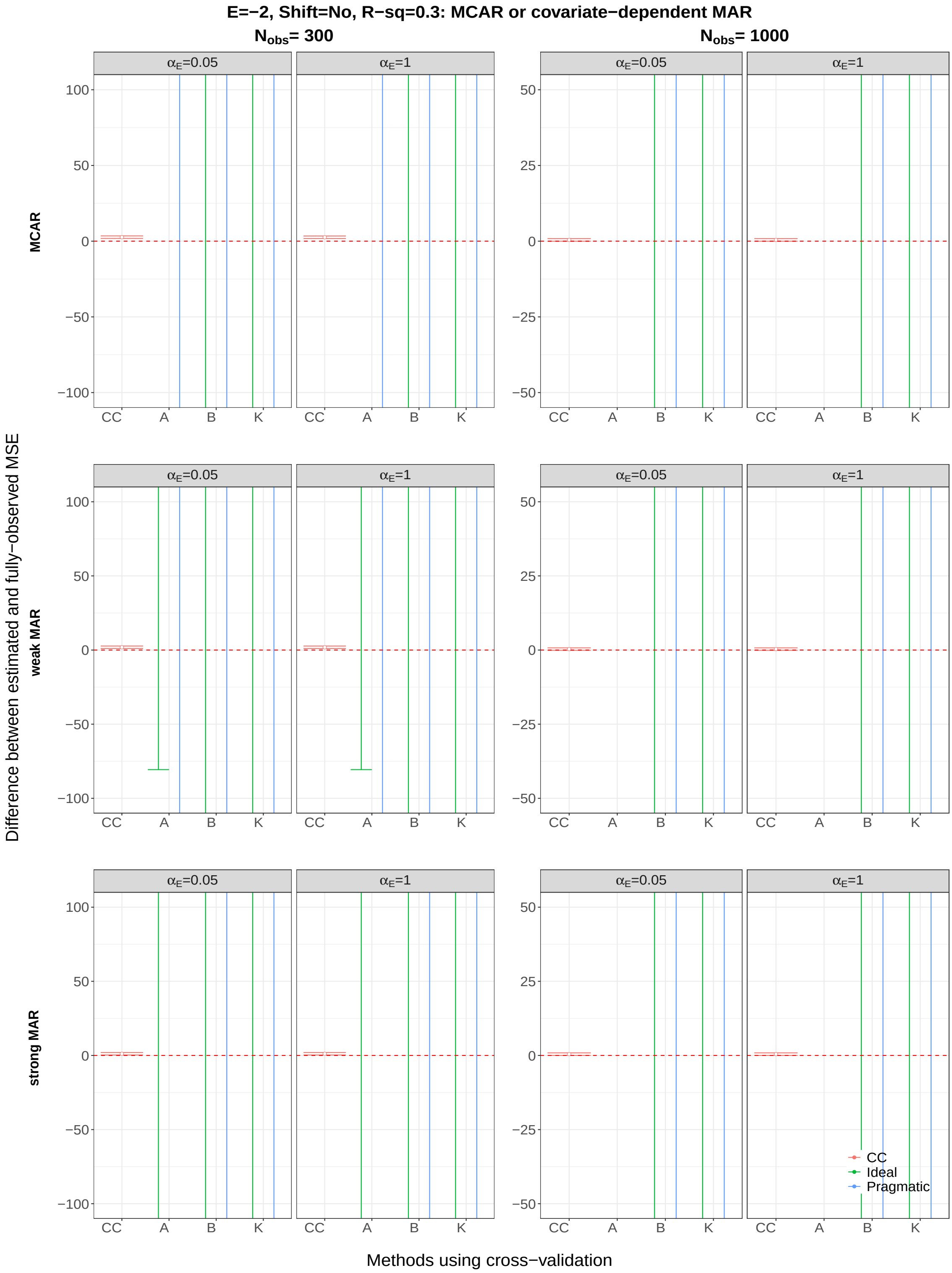

Figure S10: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

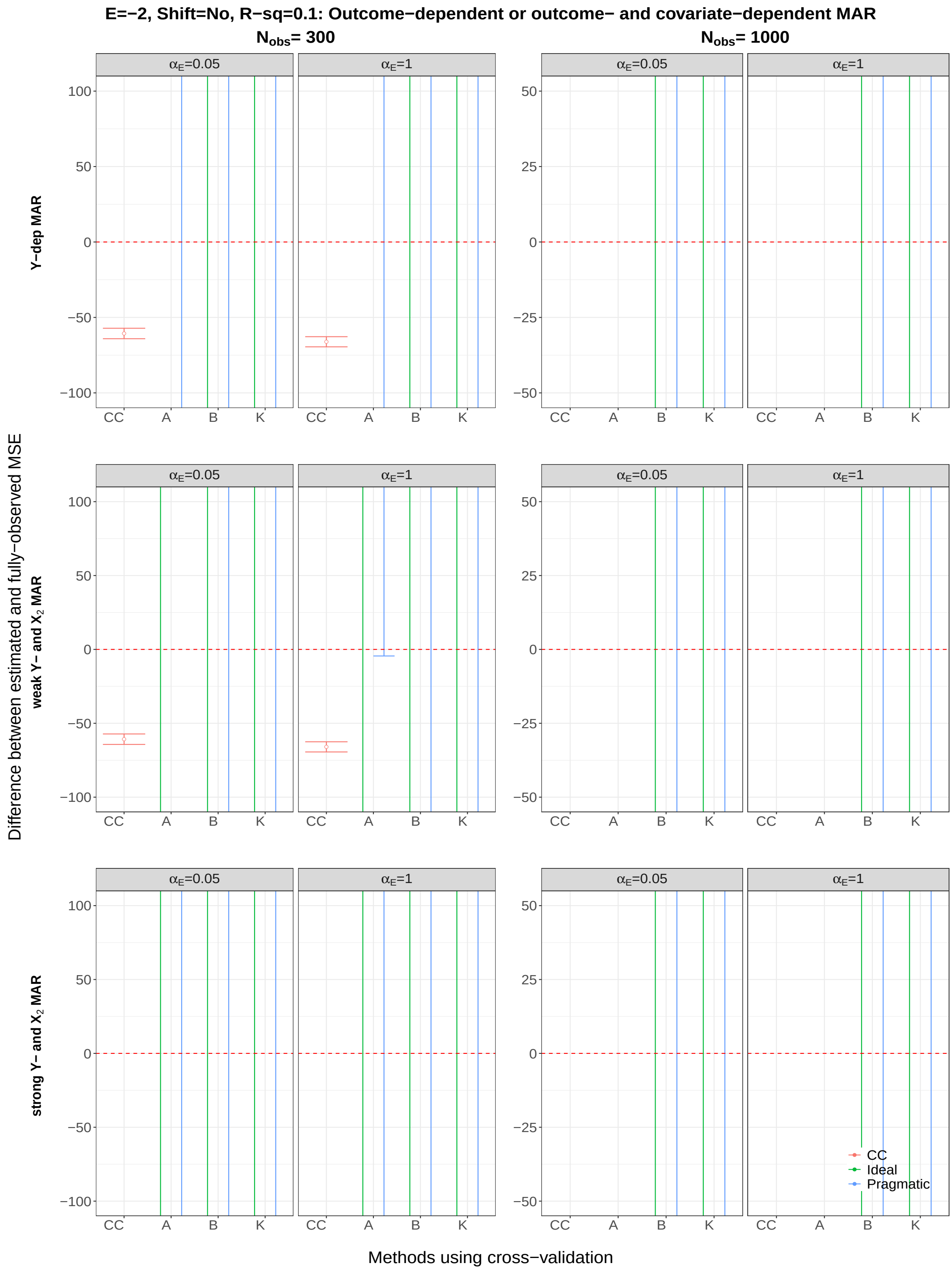

Figure S11: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

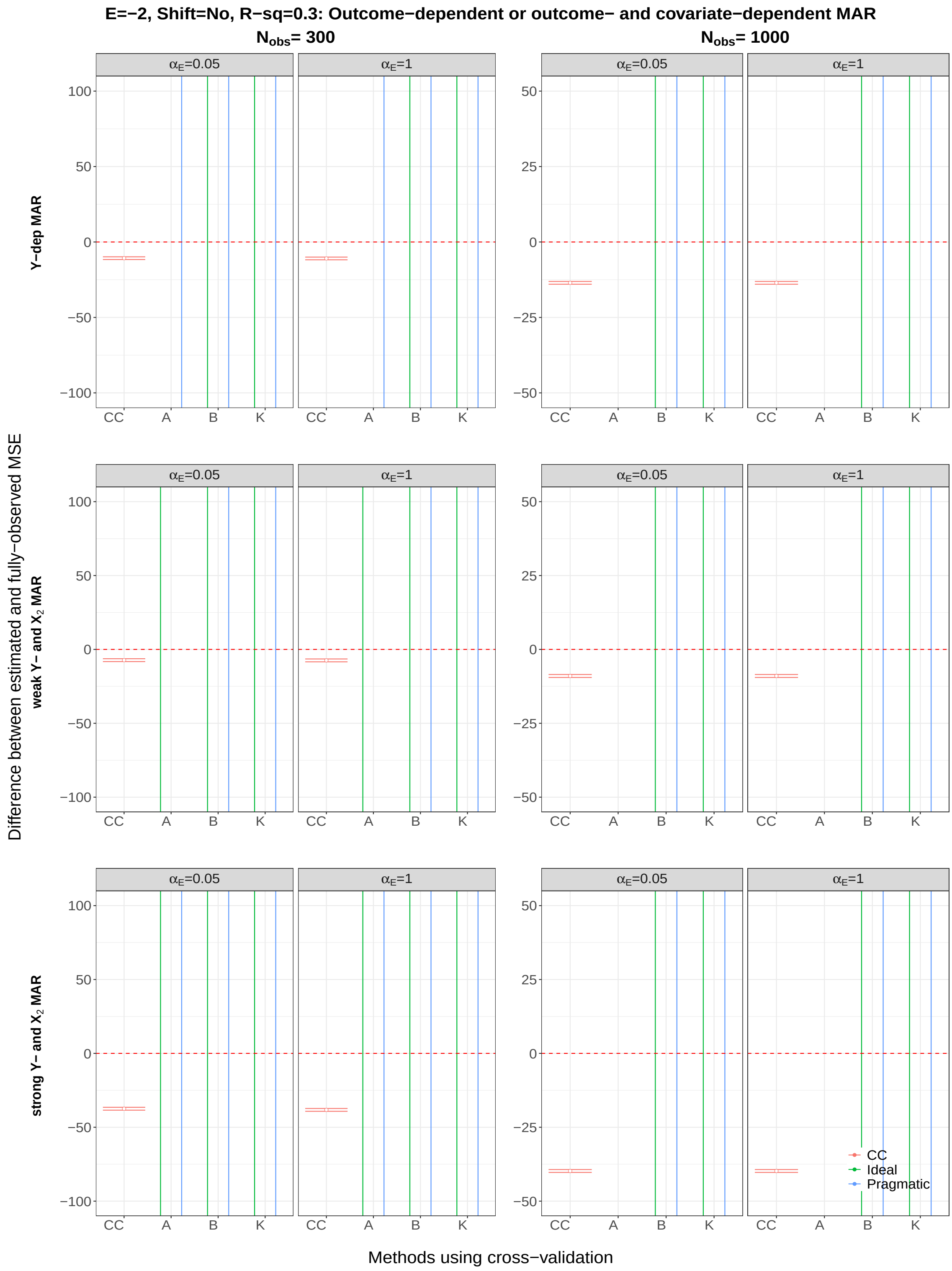

Figure S12: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S5.1.2 An origin shift transformation applied

True exponent is 0

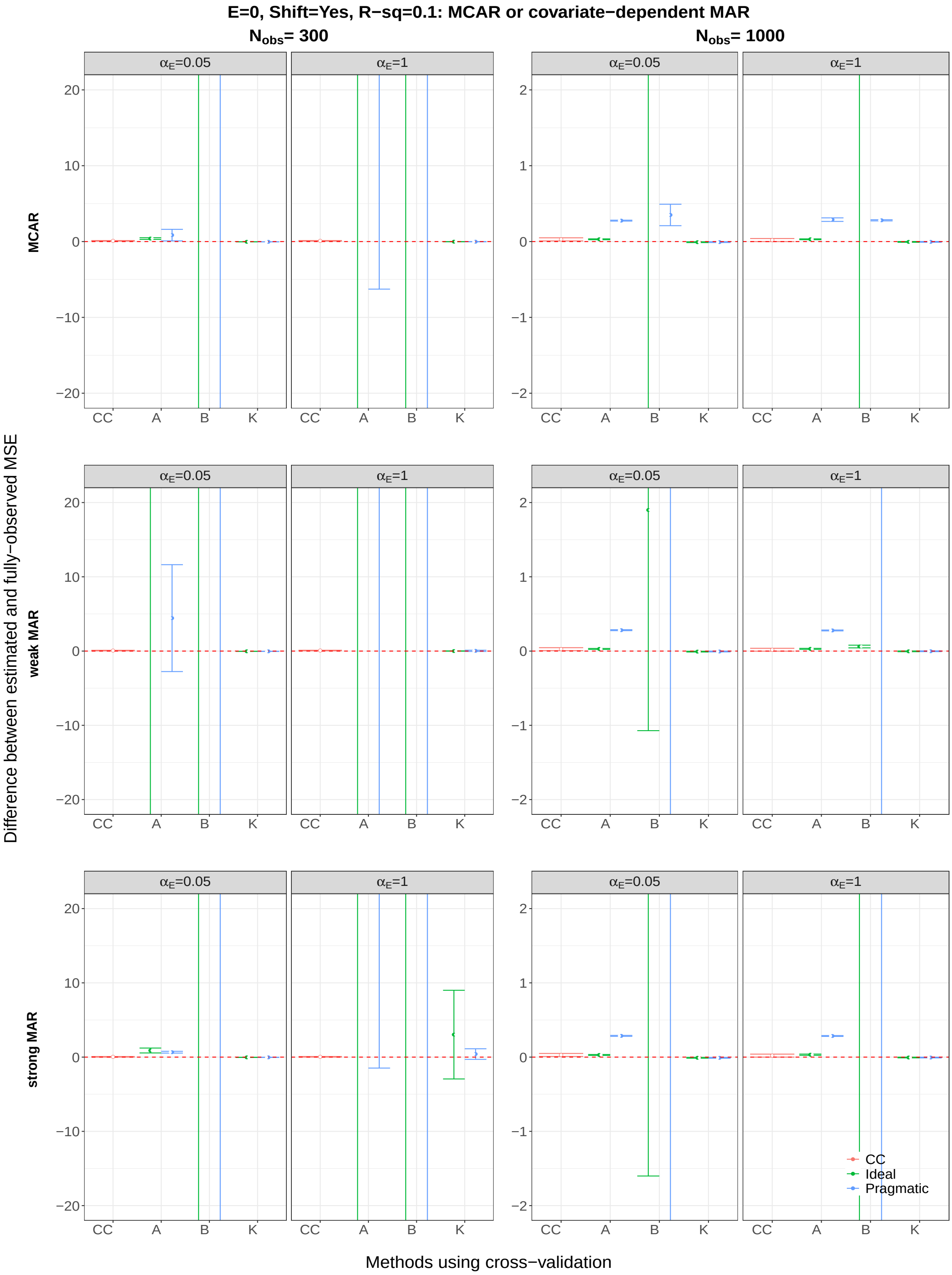

Figure S13: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

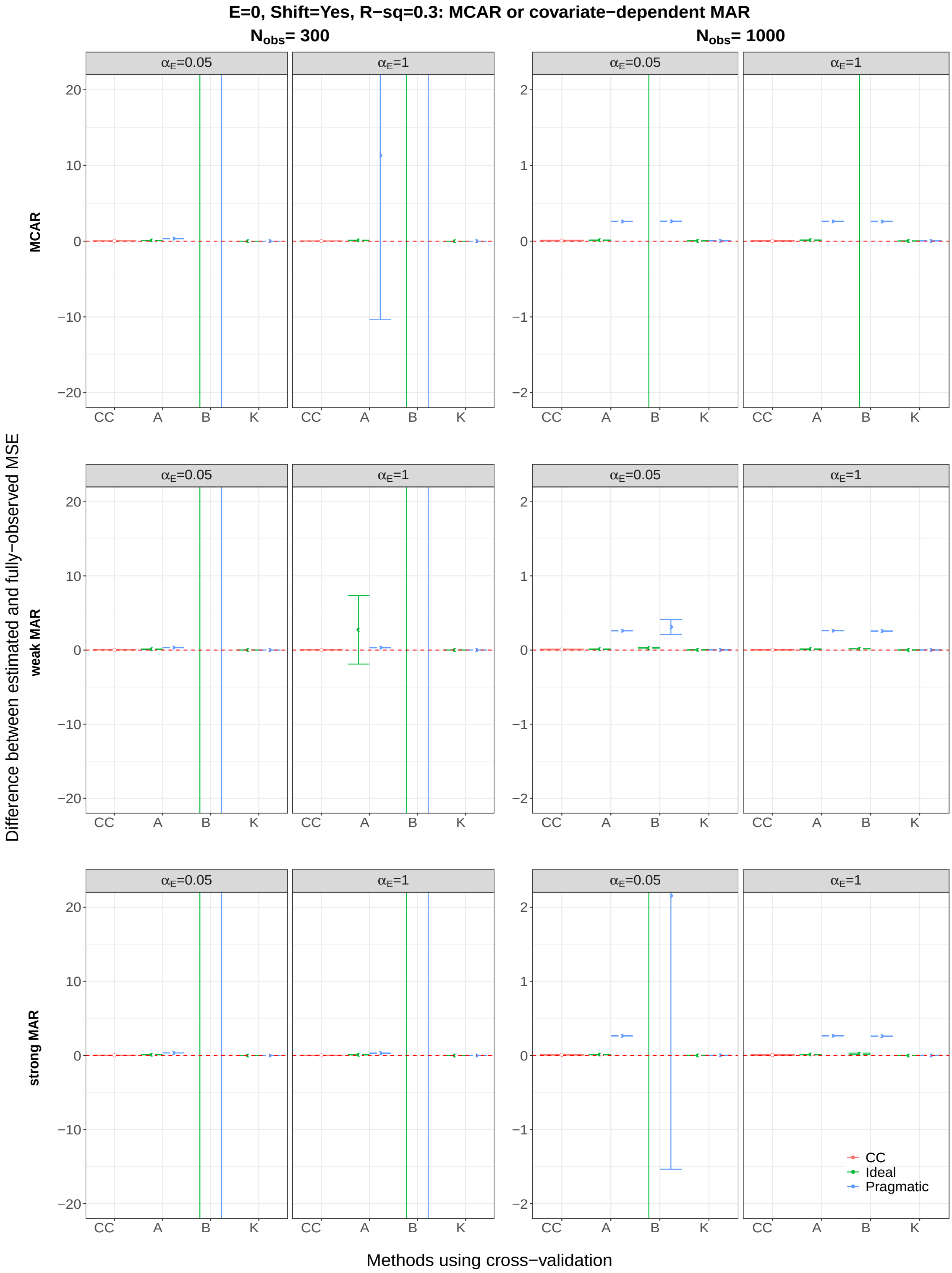

Figure S14: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

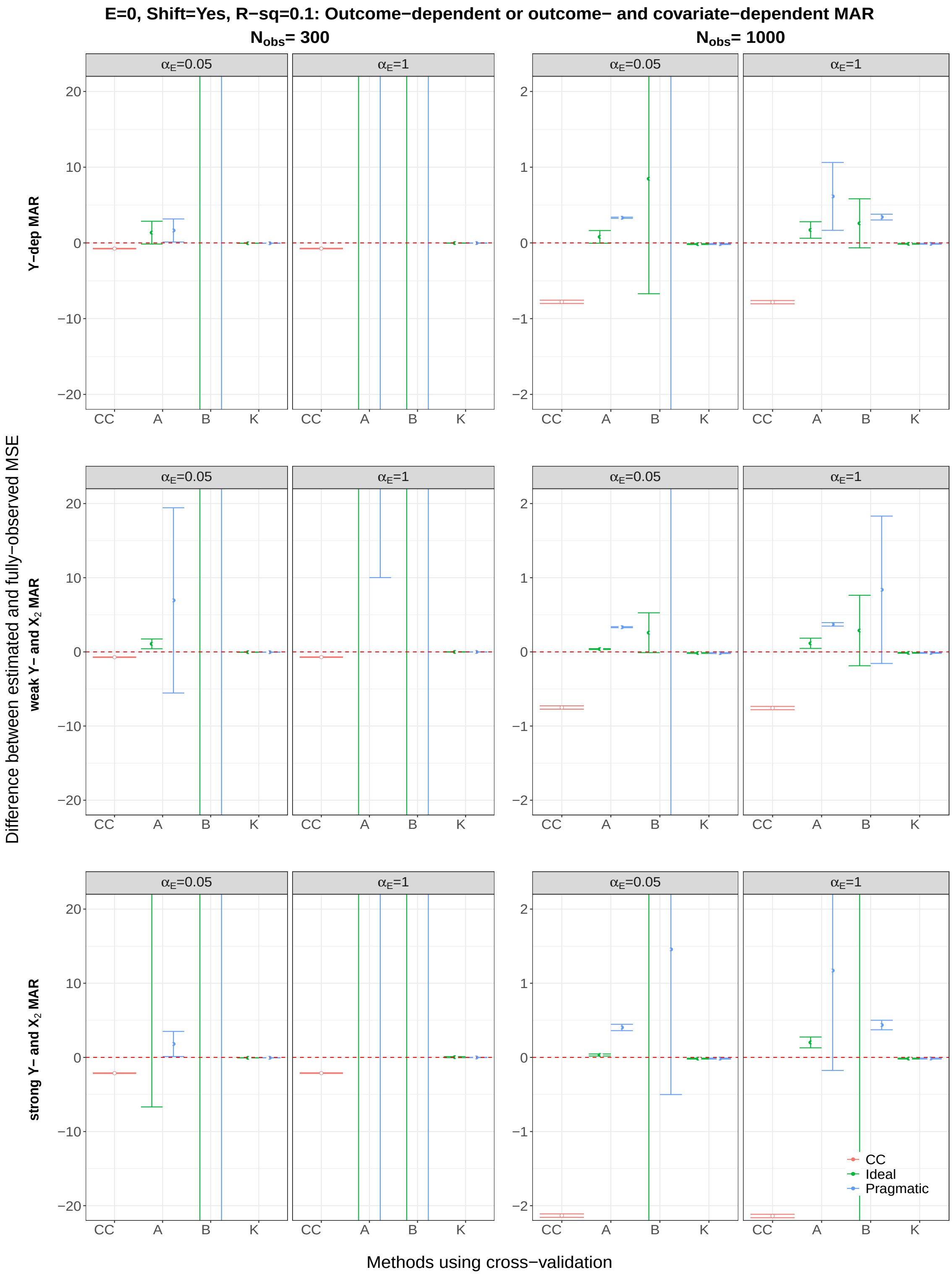

Figure S15: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

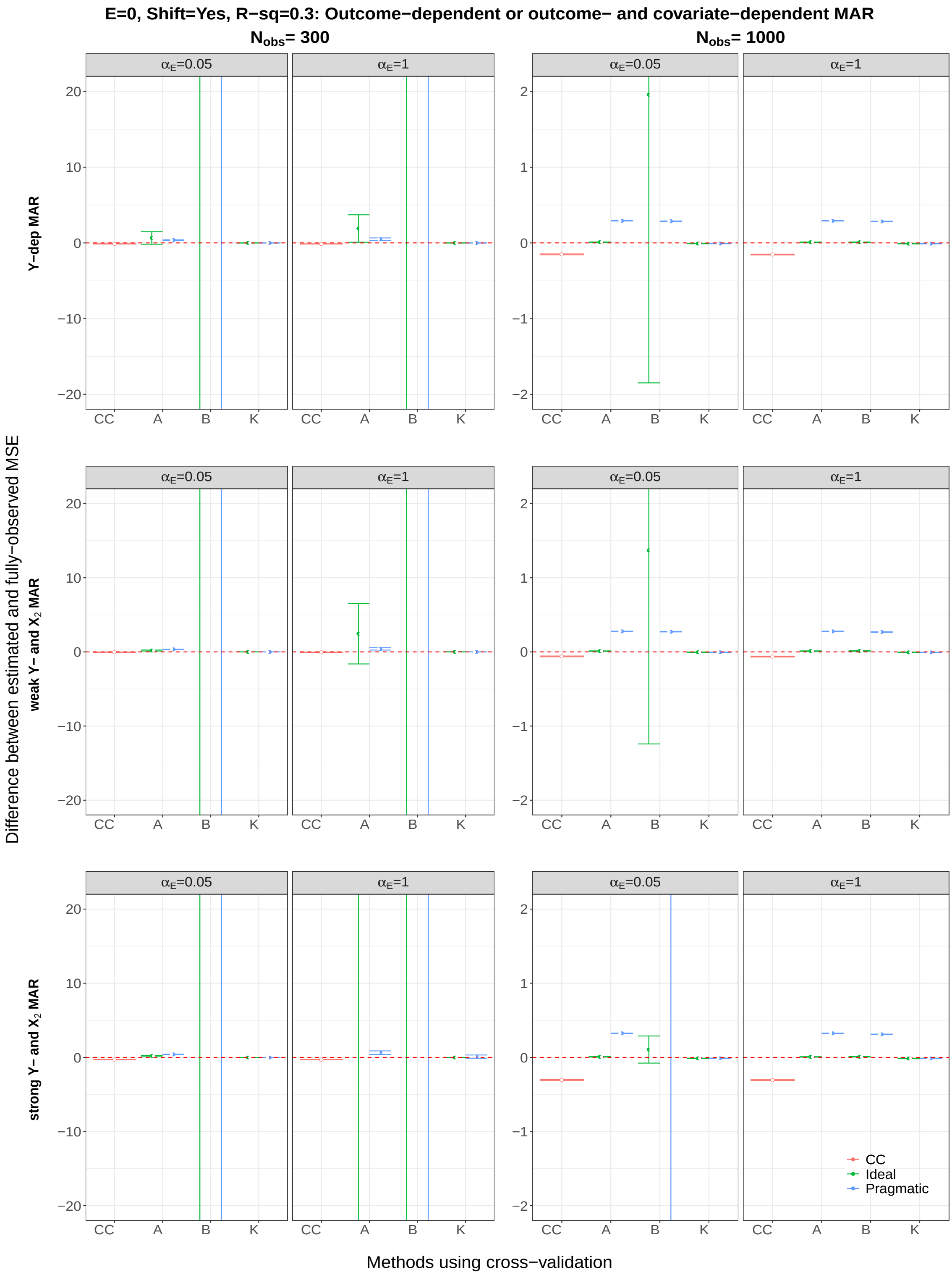

Figure S16: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

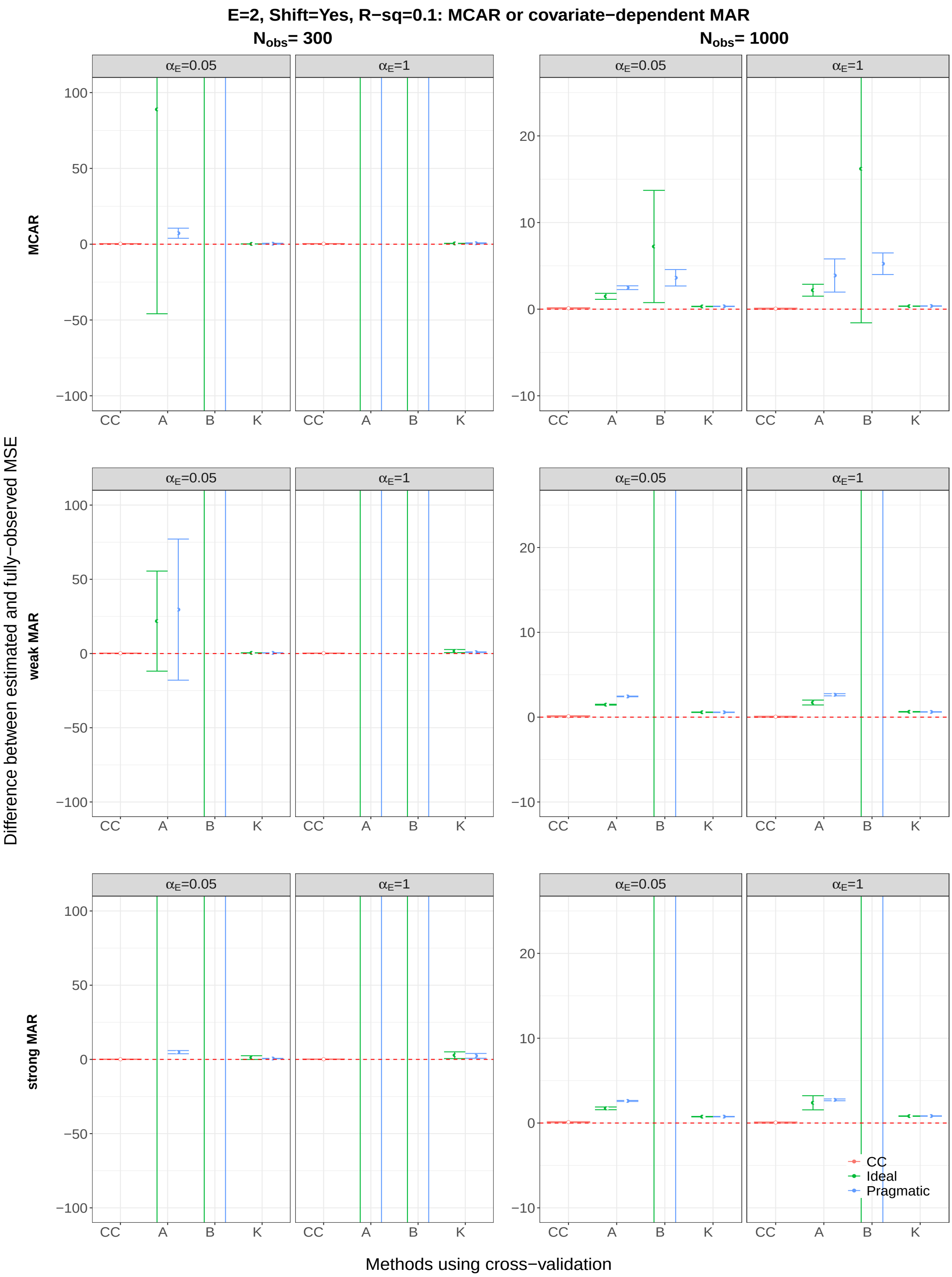

Figure S17: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

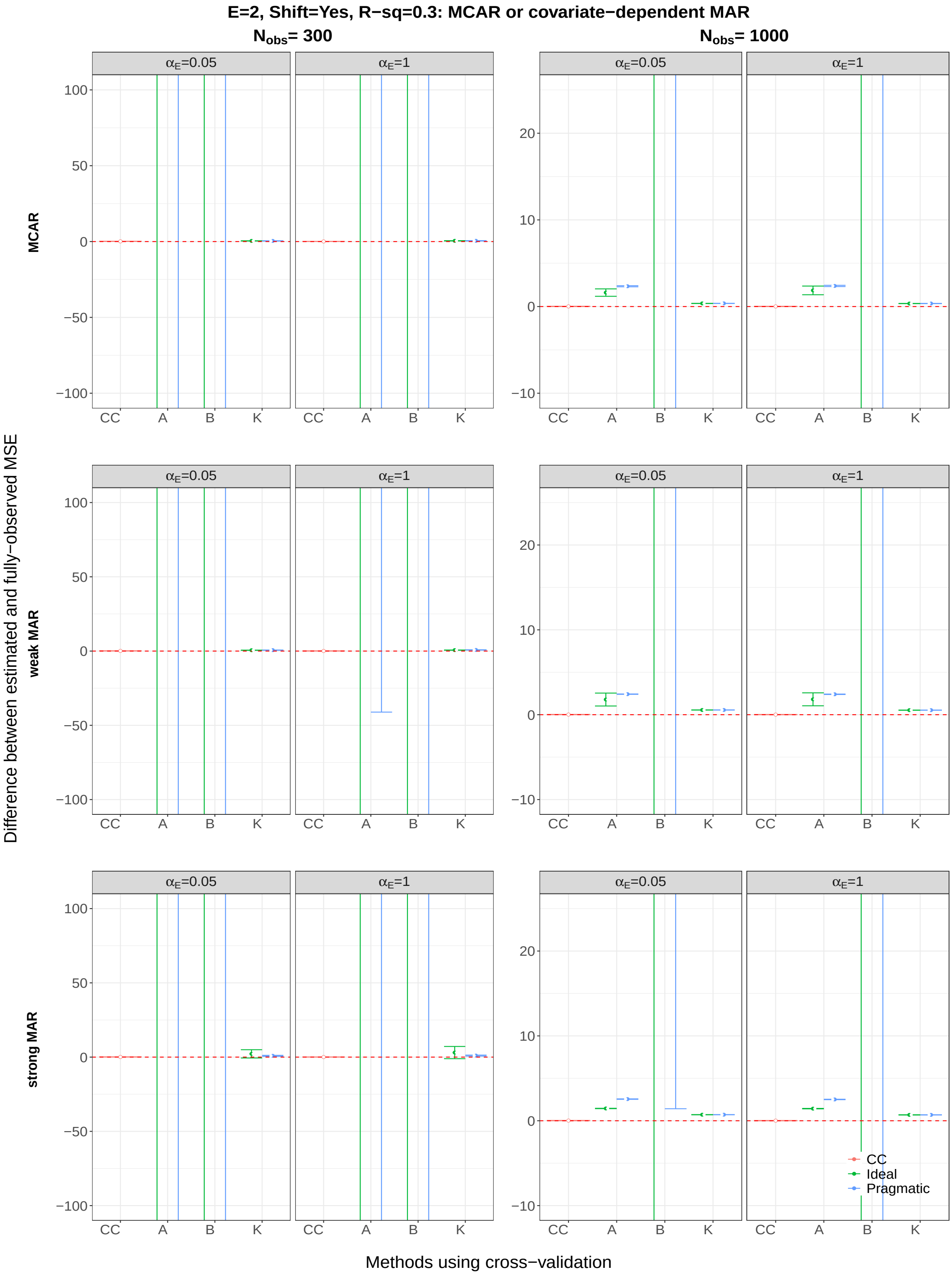

Figure S18: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

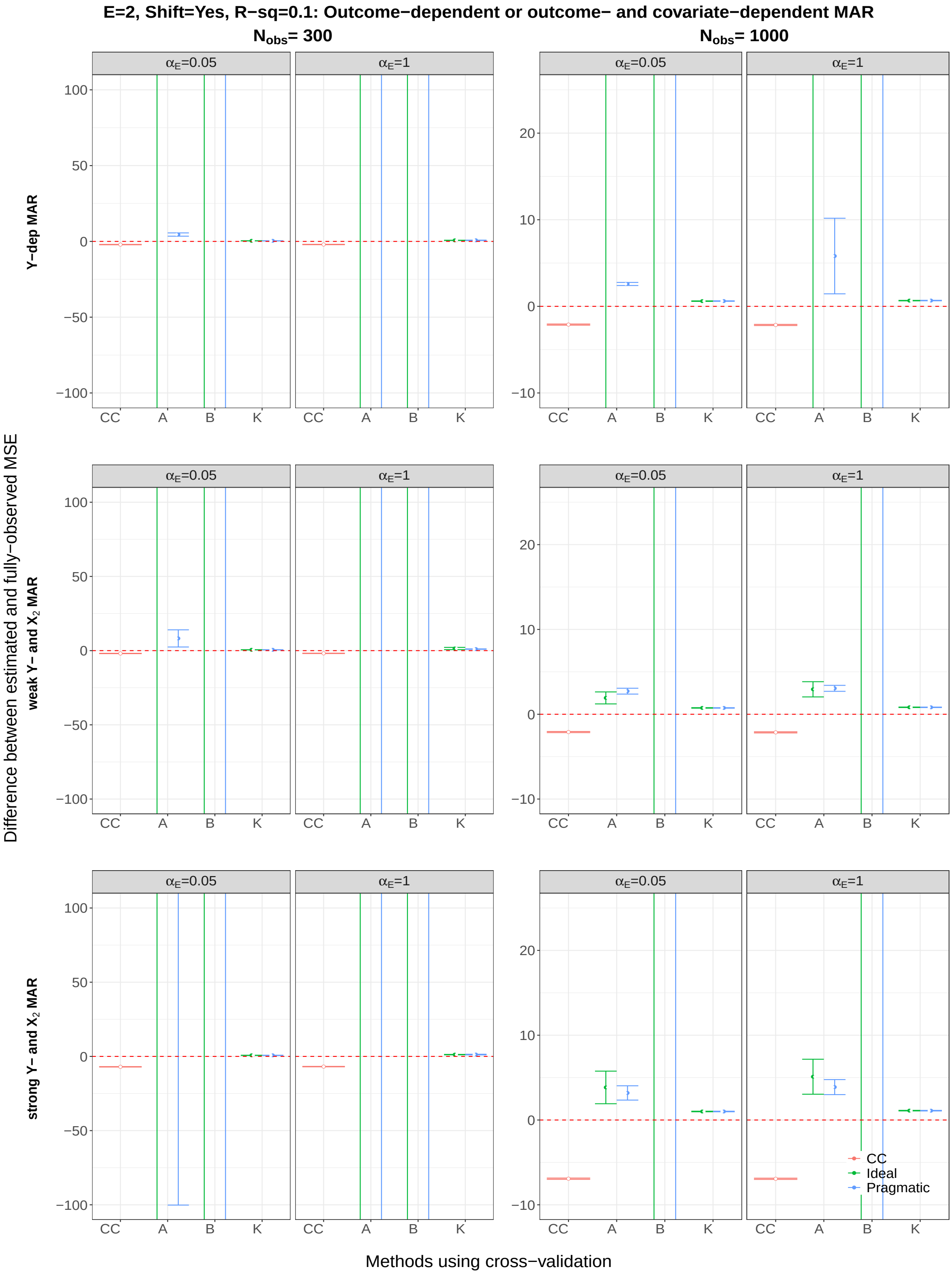

Figure S19: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

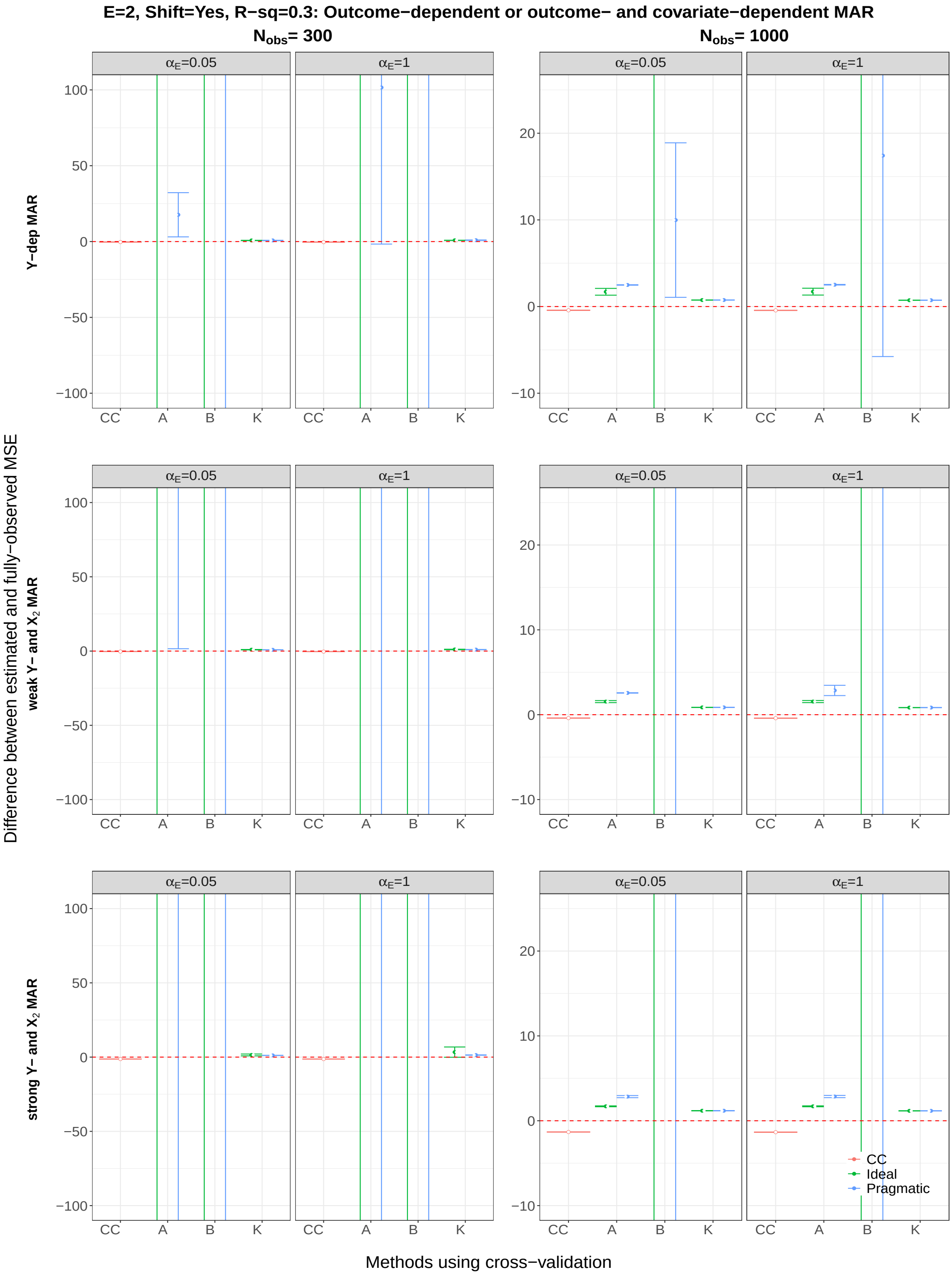

Figure S20: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

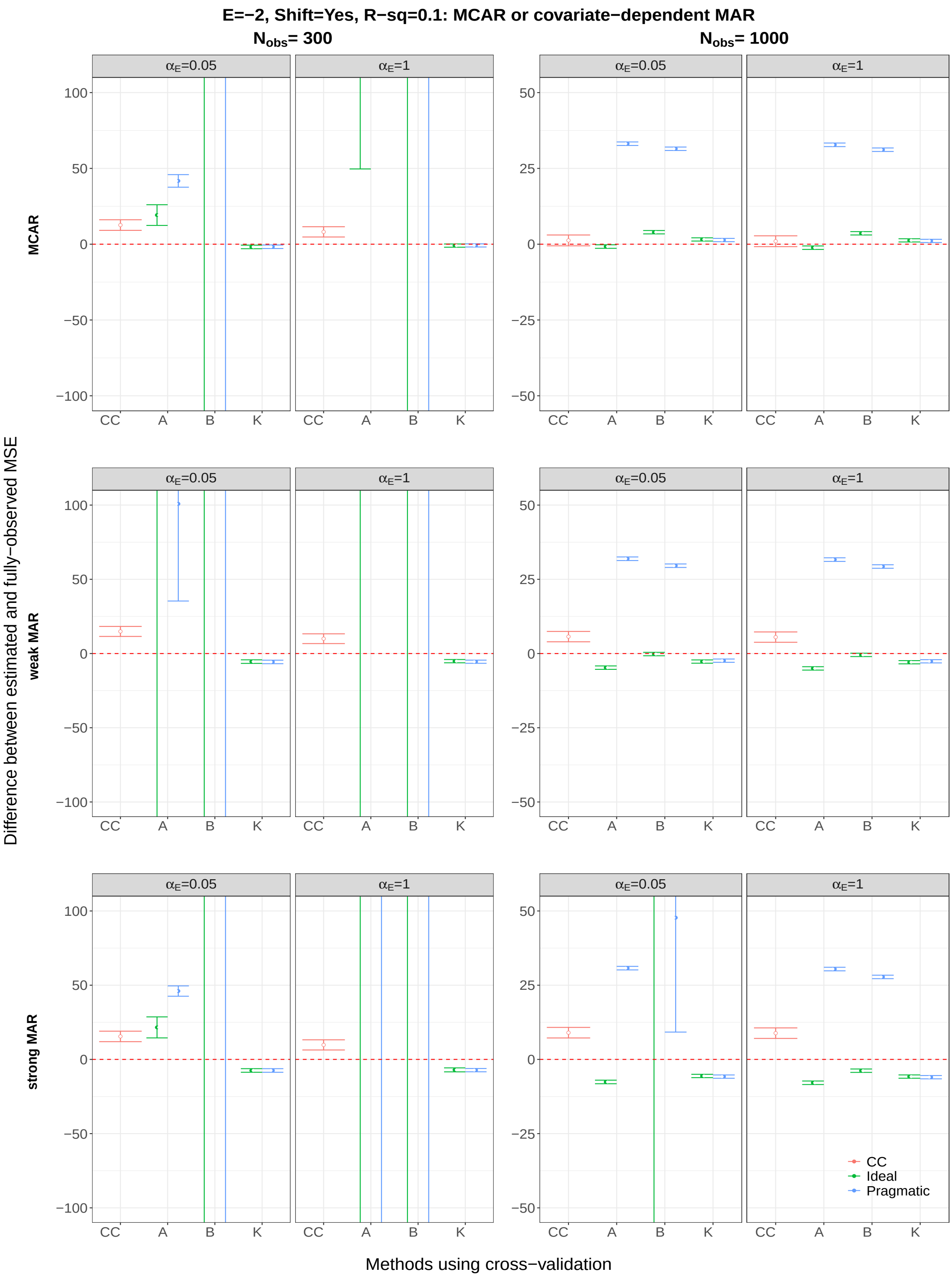

Figure S21: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

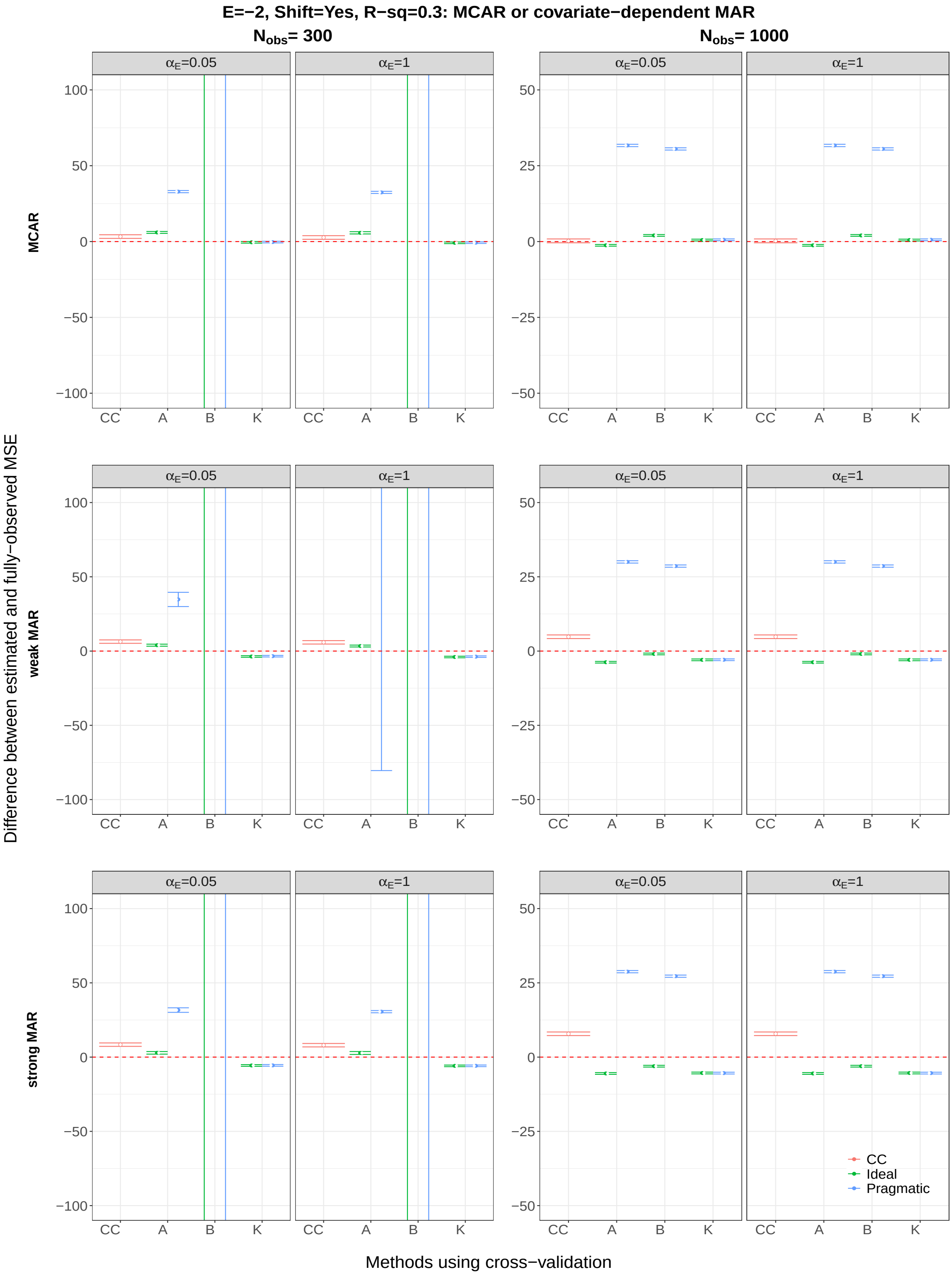

Figure S22: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

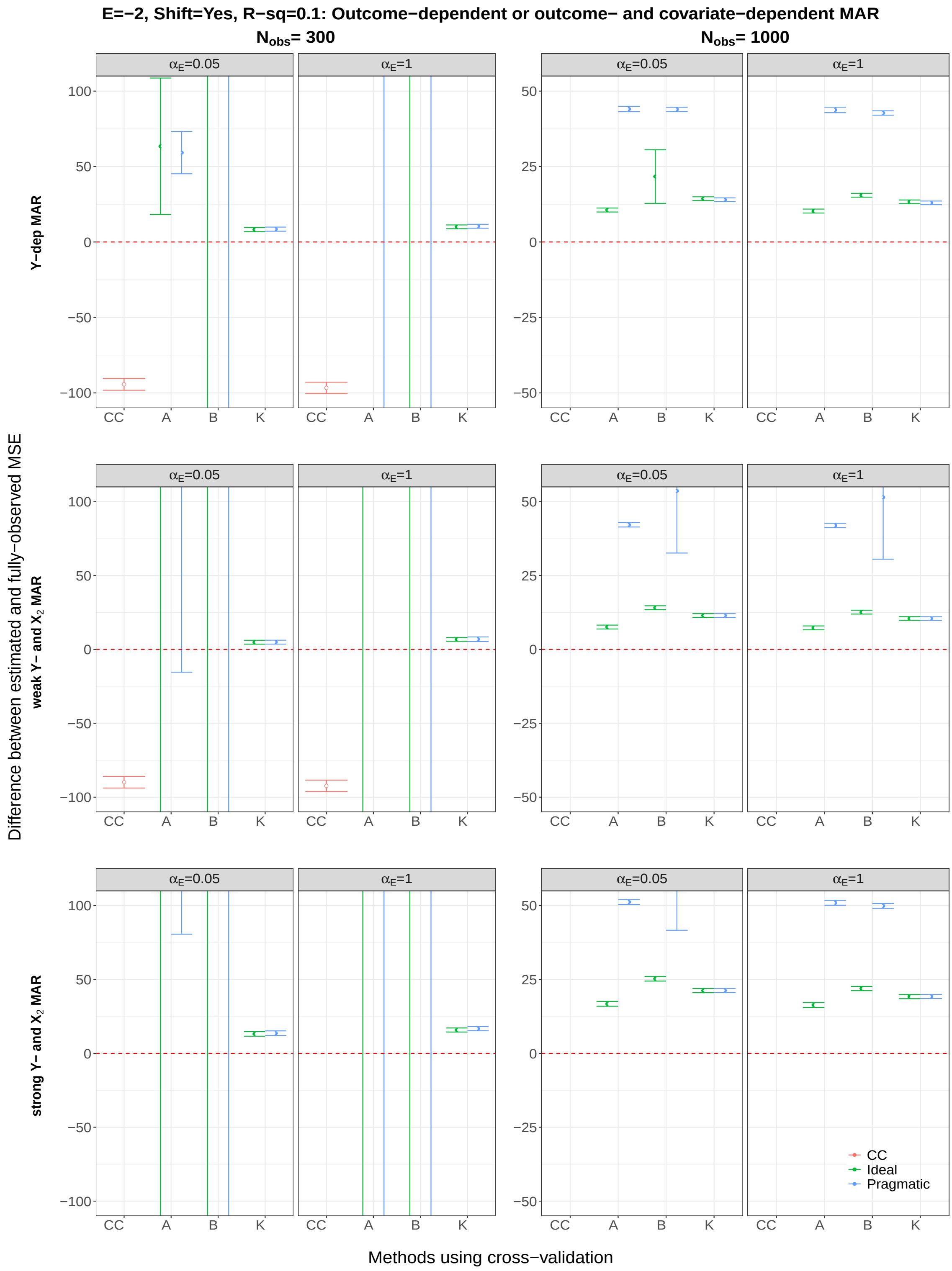

Figure S23: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

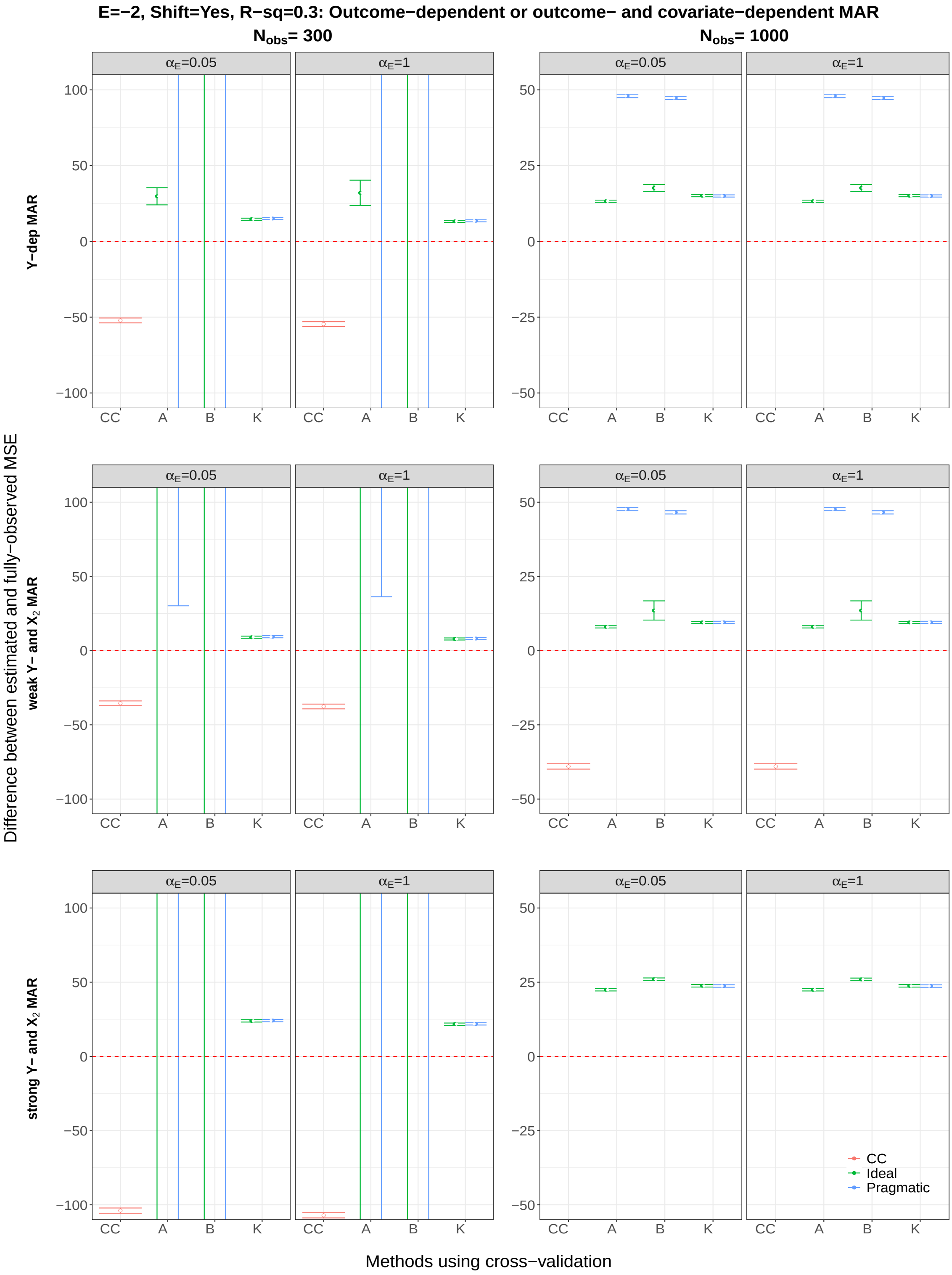

Figure S24: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S5.2 The 0.632 bootstrap

S5.2.1 No origin shift transformation applied

True exponent is 0

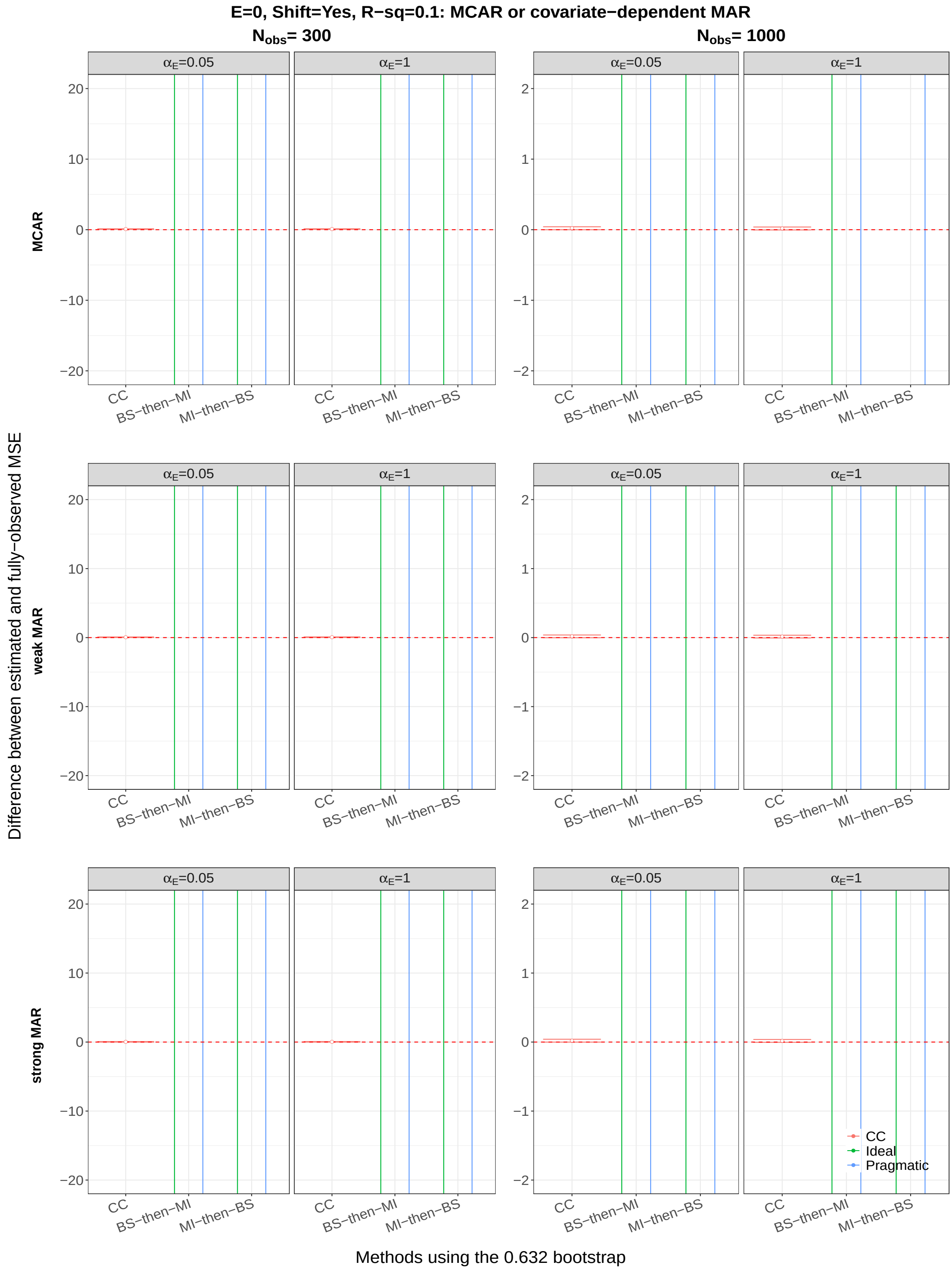

Figure S25: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

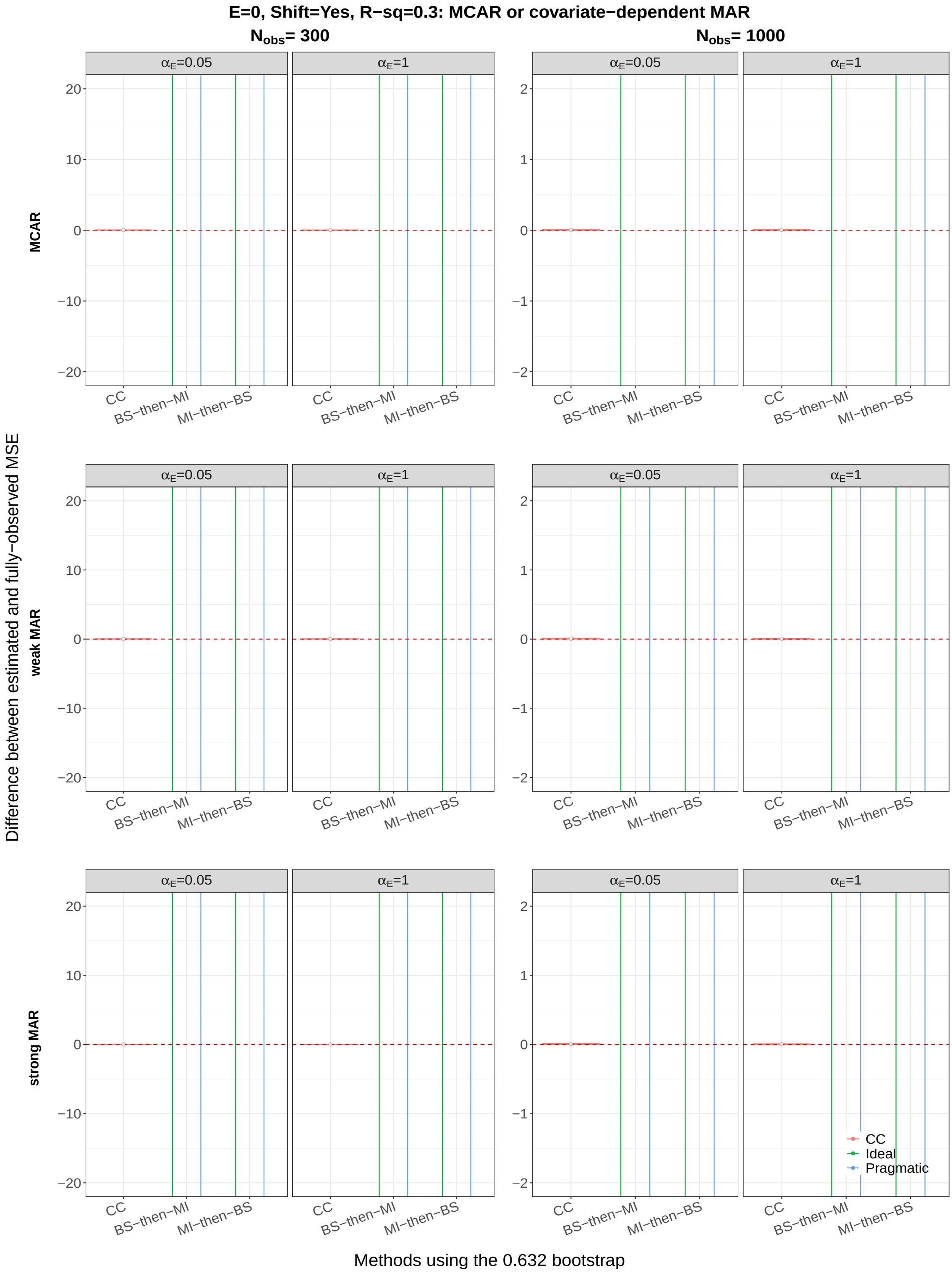

Figure S26: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

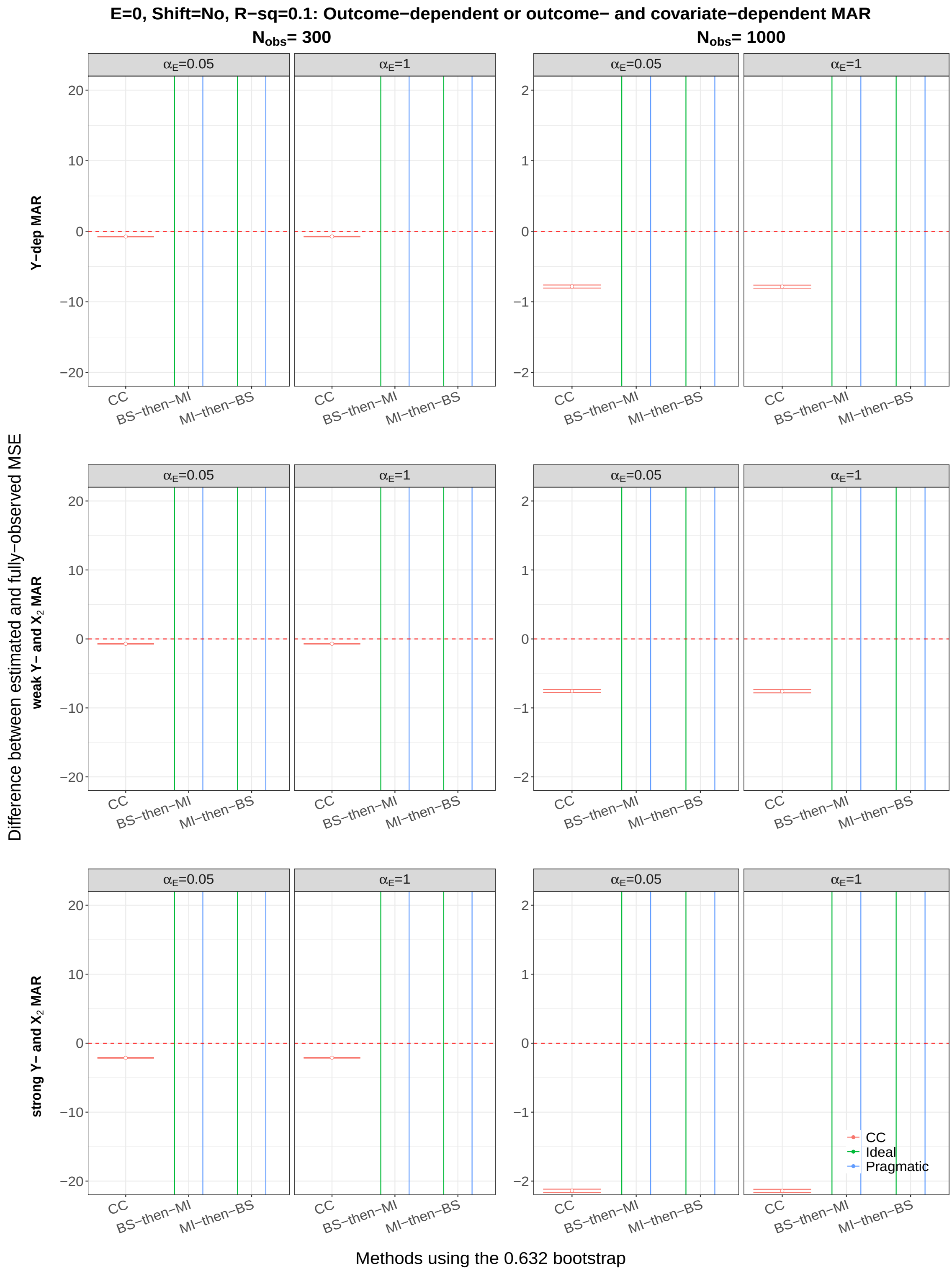

Figure S27: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

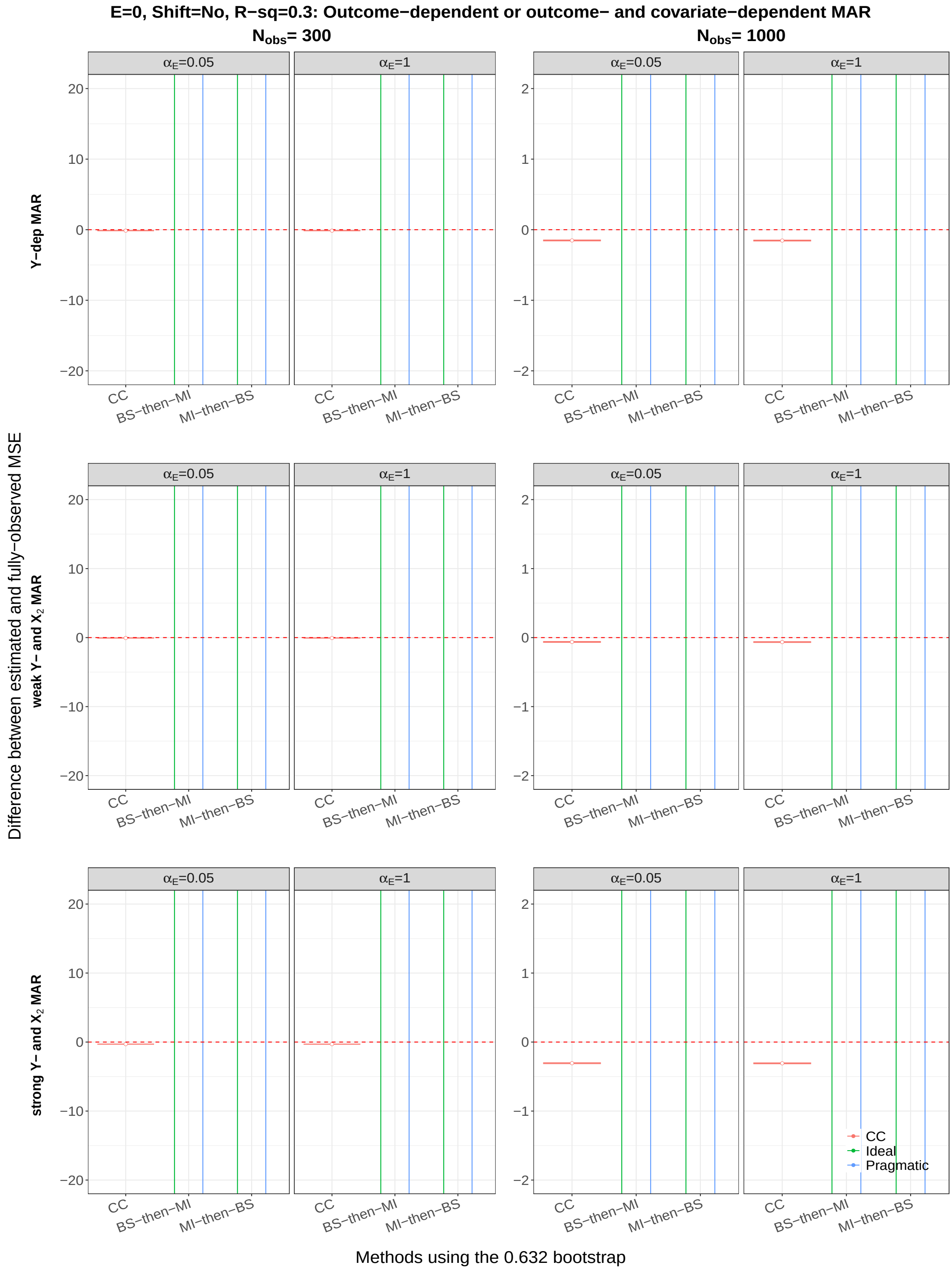

Figure S28: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

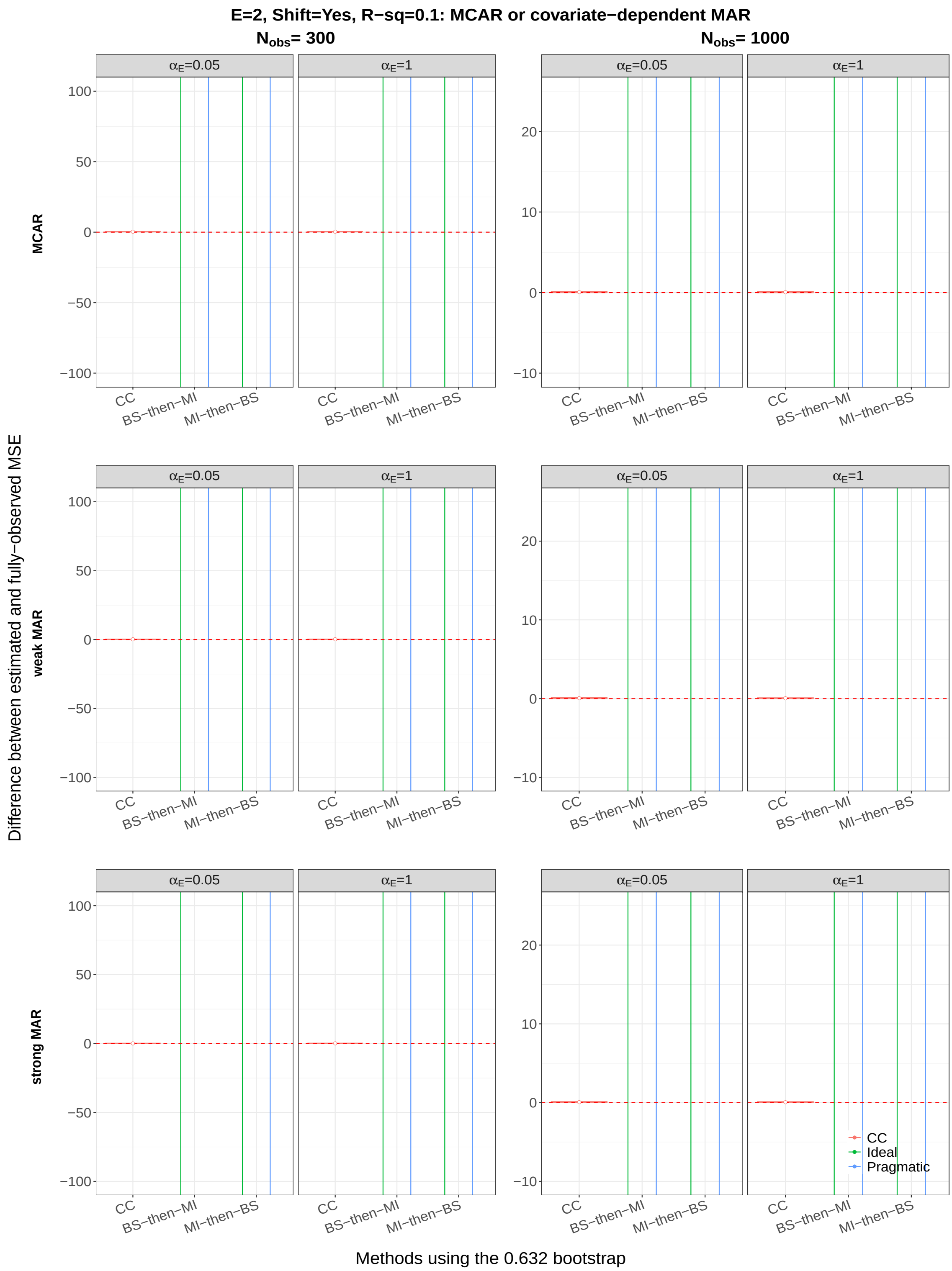

Figure S29: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

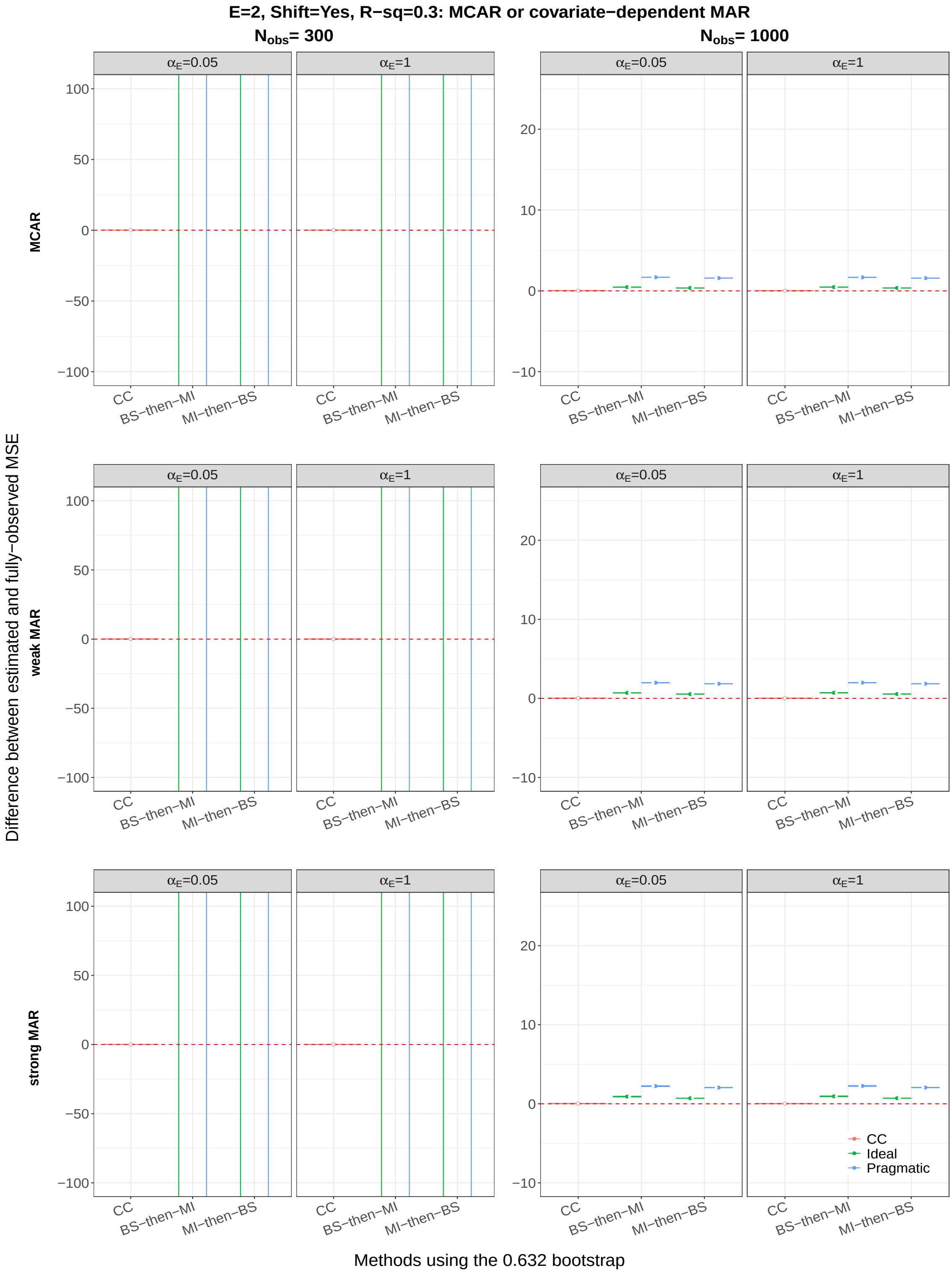

Figure S30: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

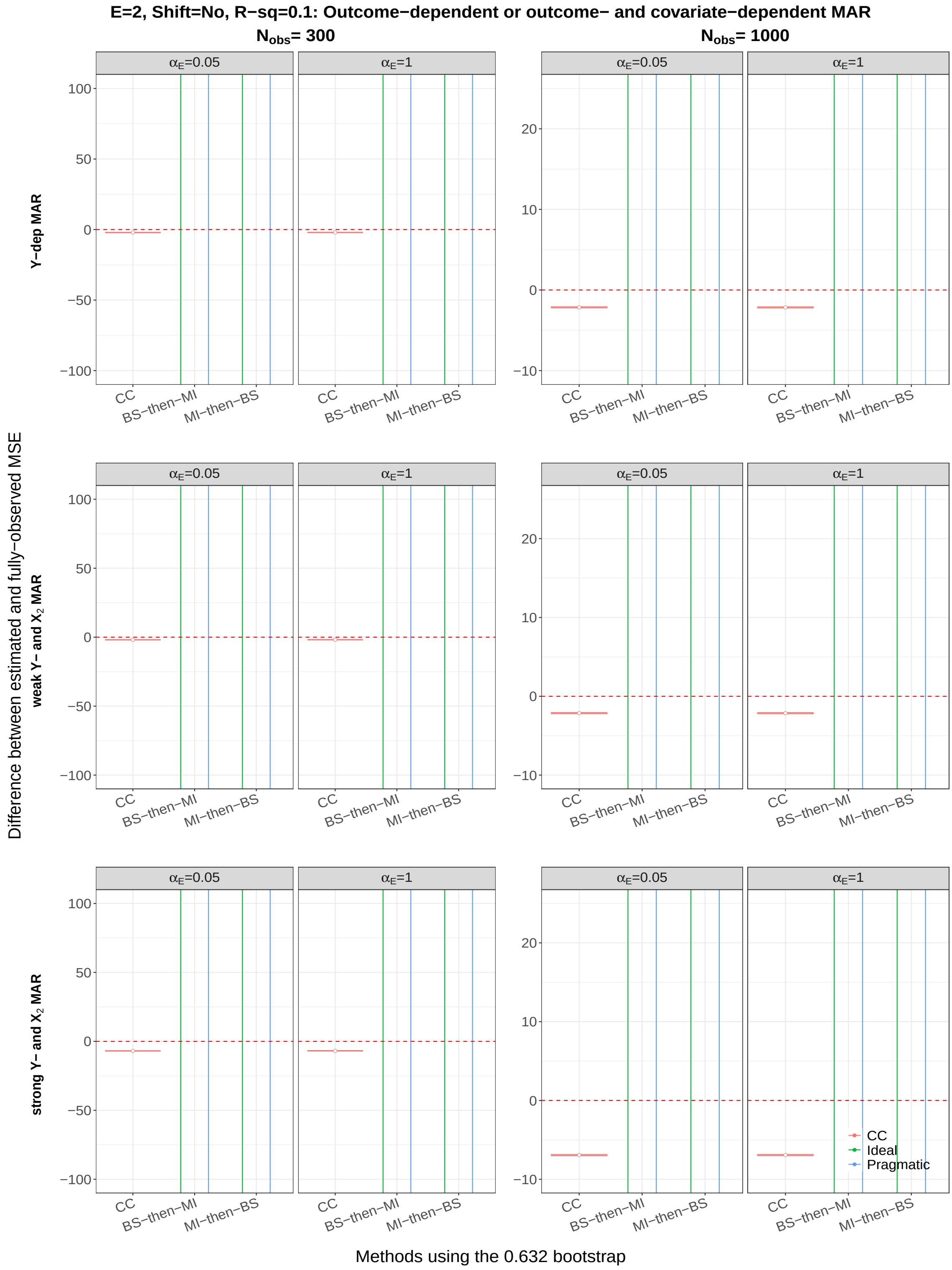

Figure S31: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

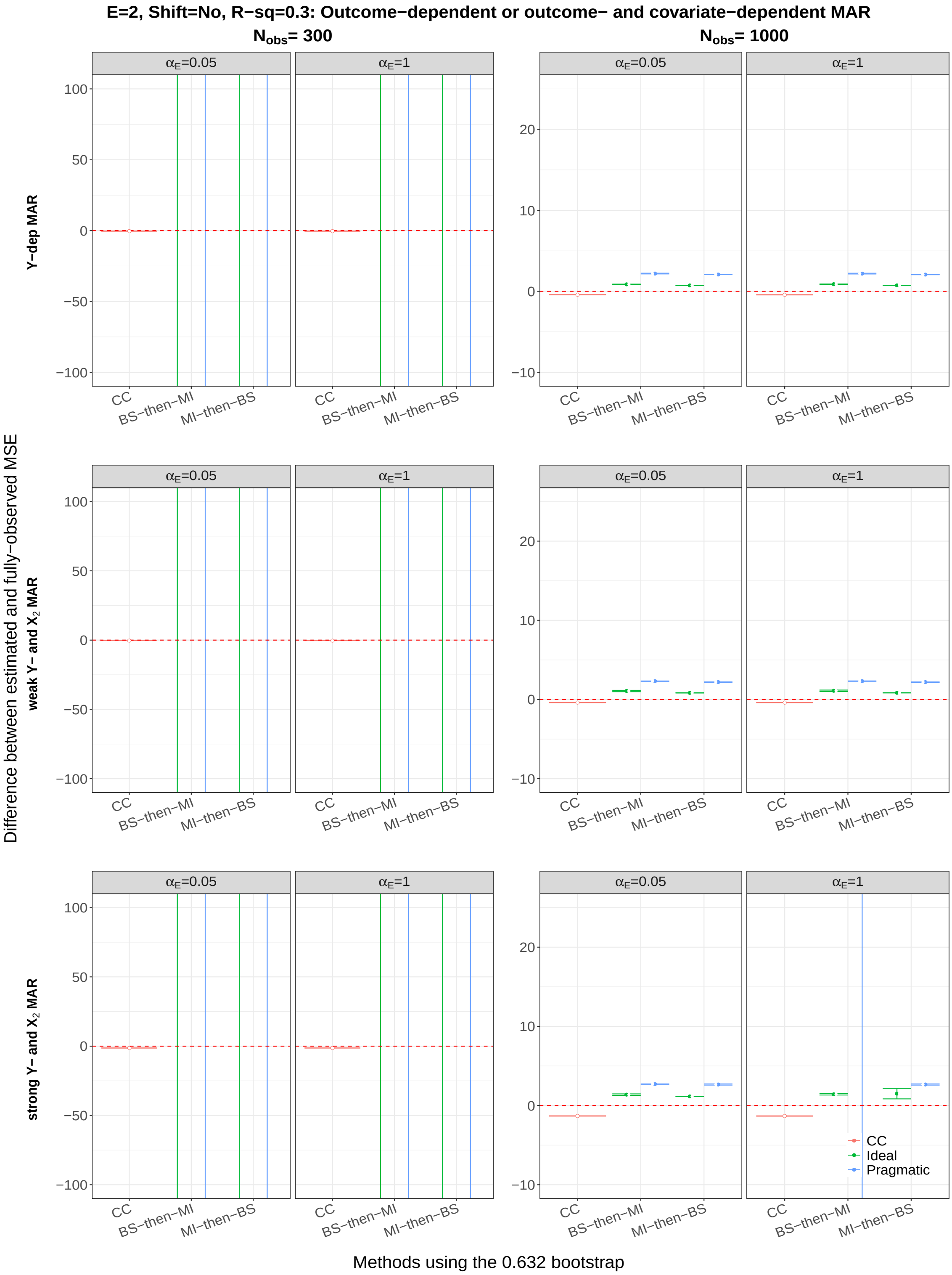

Figure S32: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

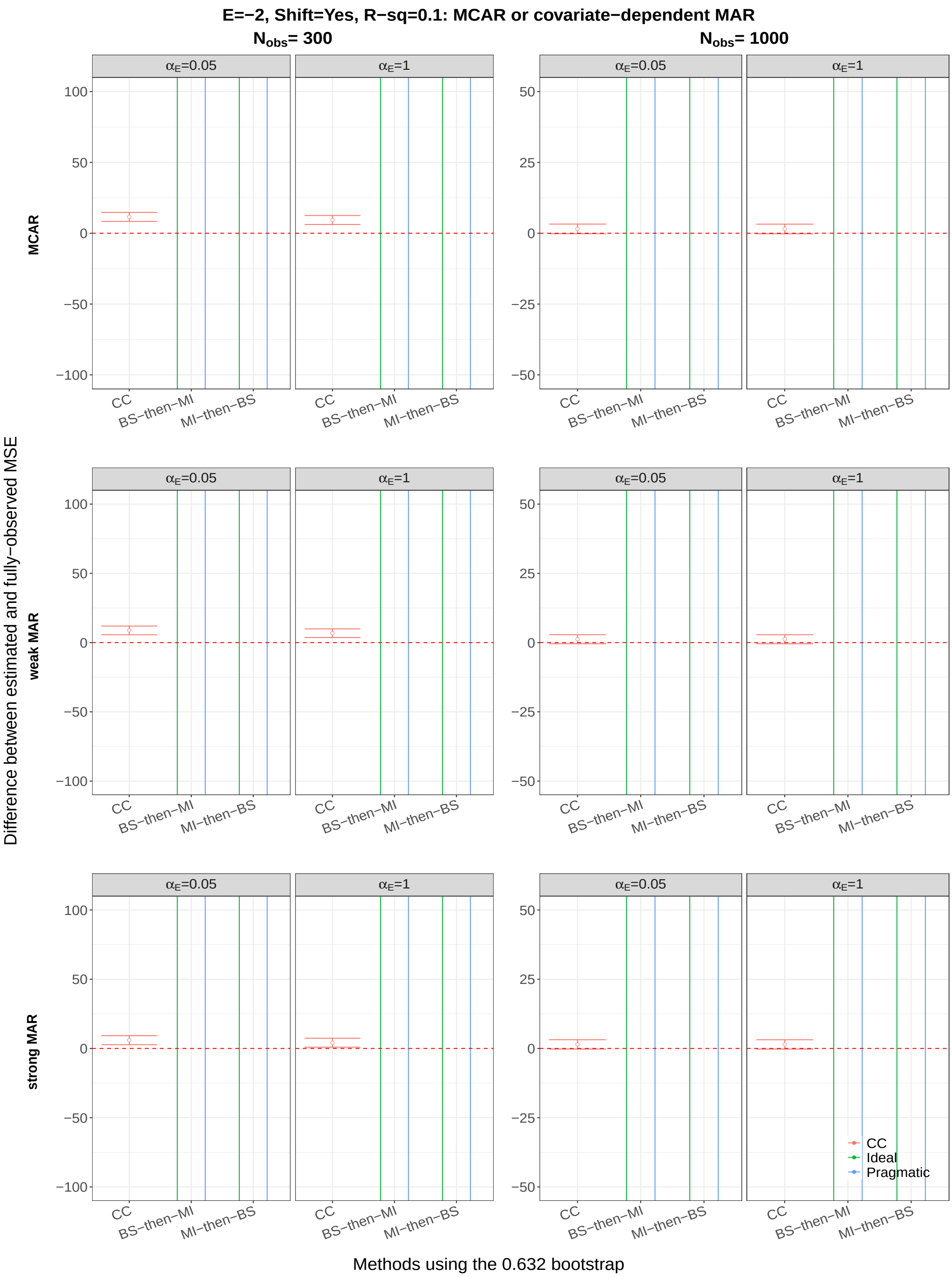

Figure S33: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

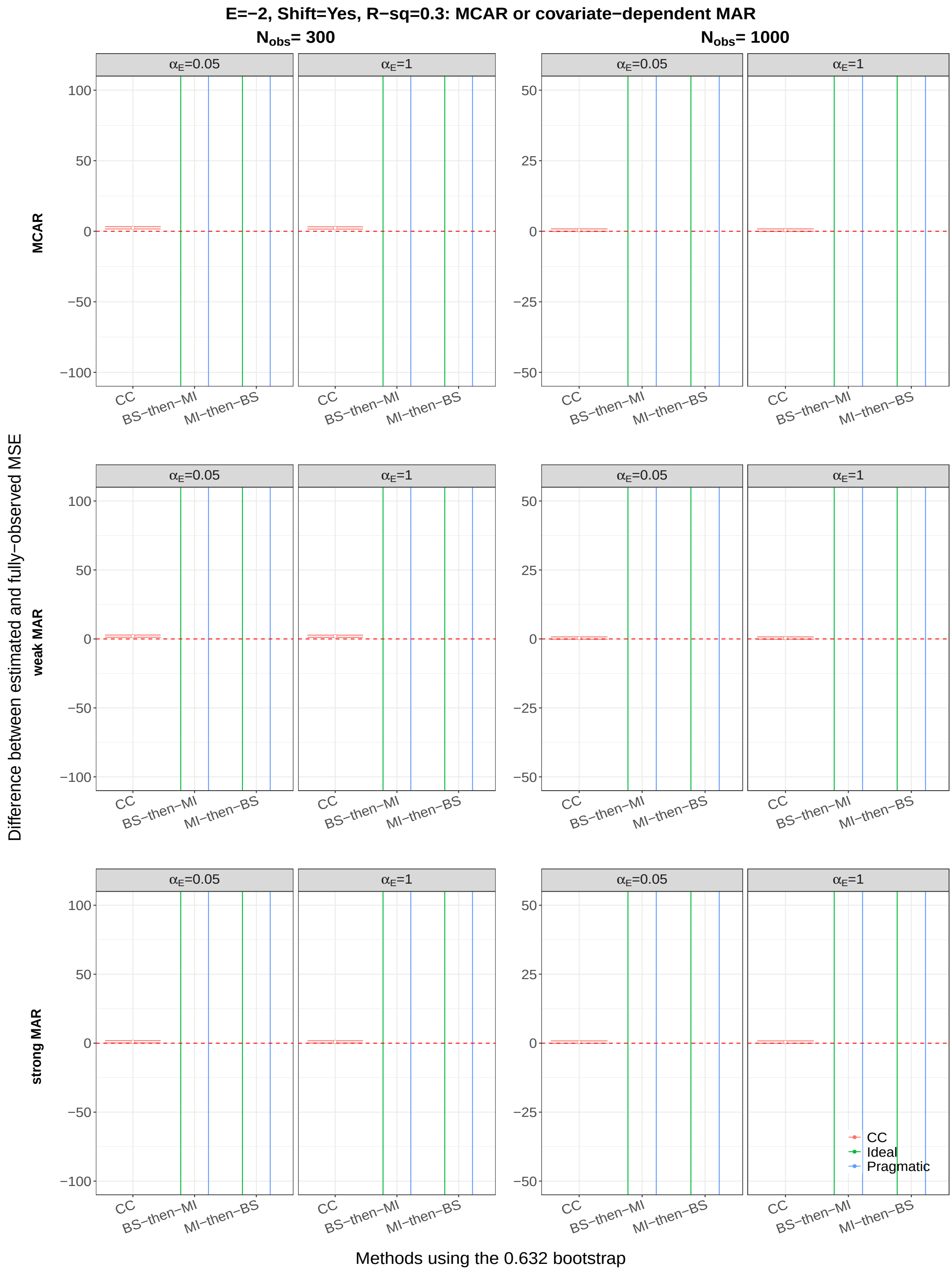

Figure S34: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

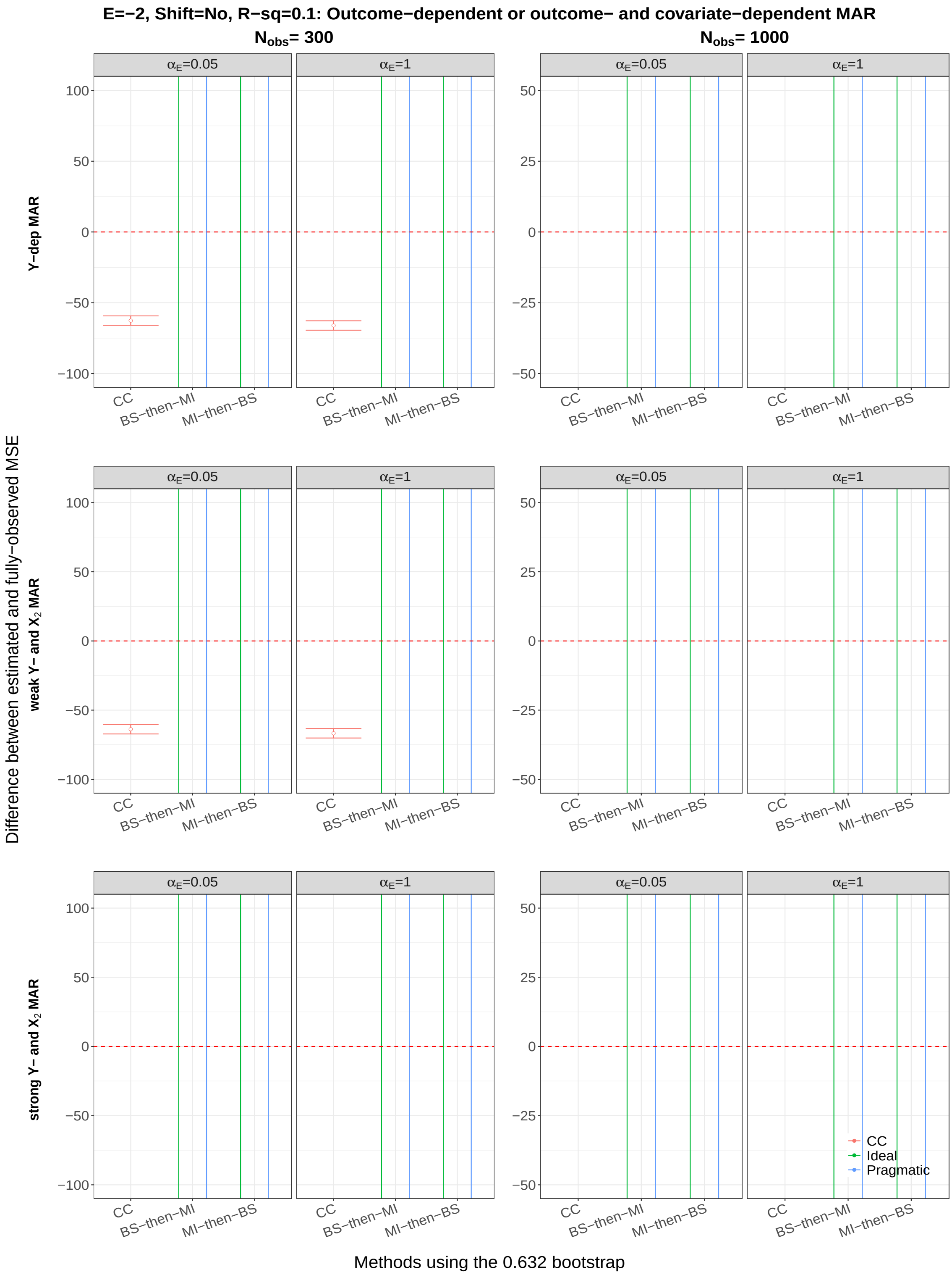

Figure S35: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

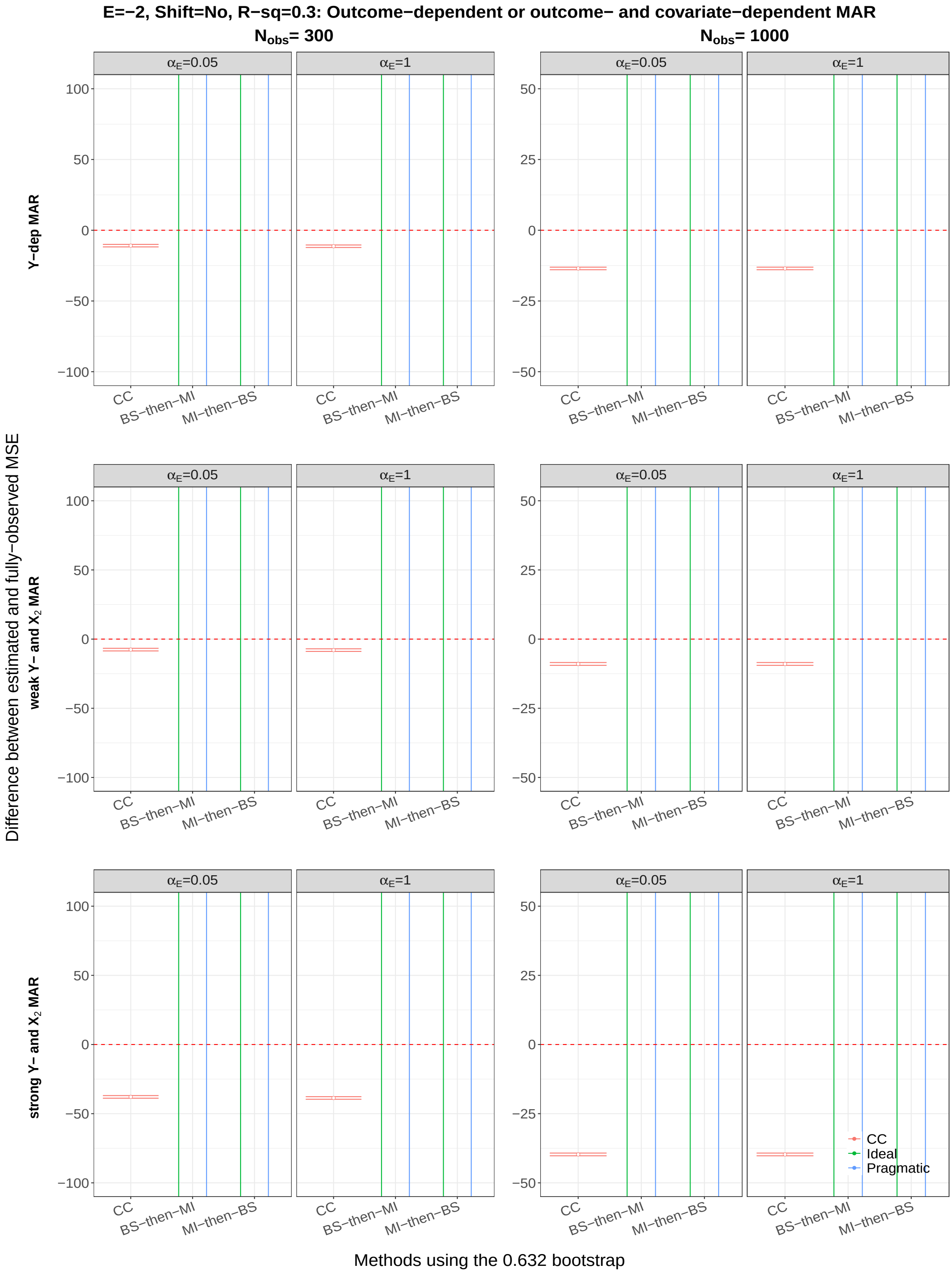

Figure S36: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S5.2.2 An origin shift transformation applied

True exponent is 0

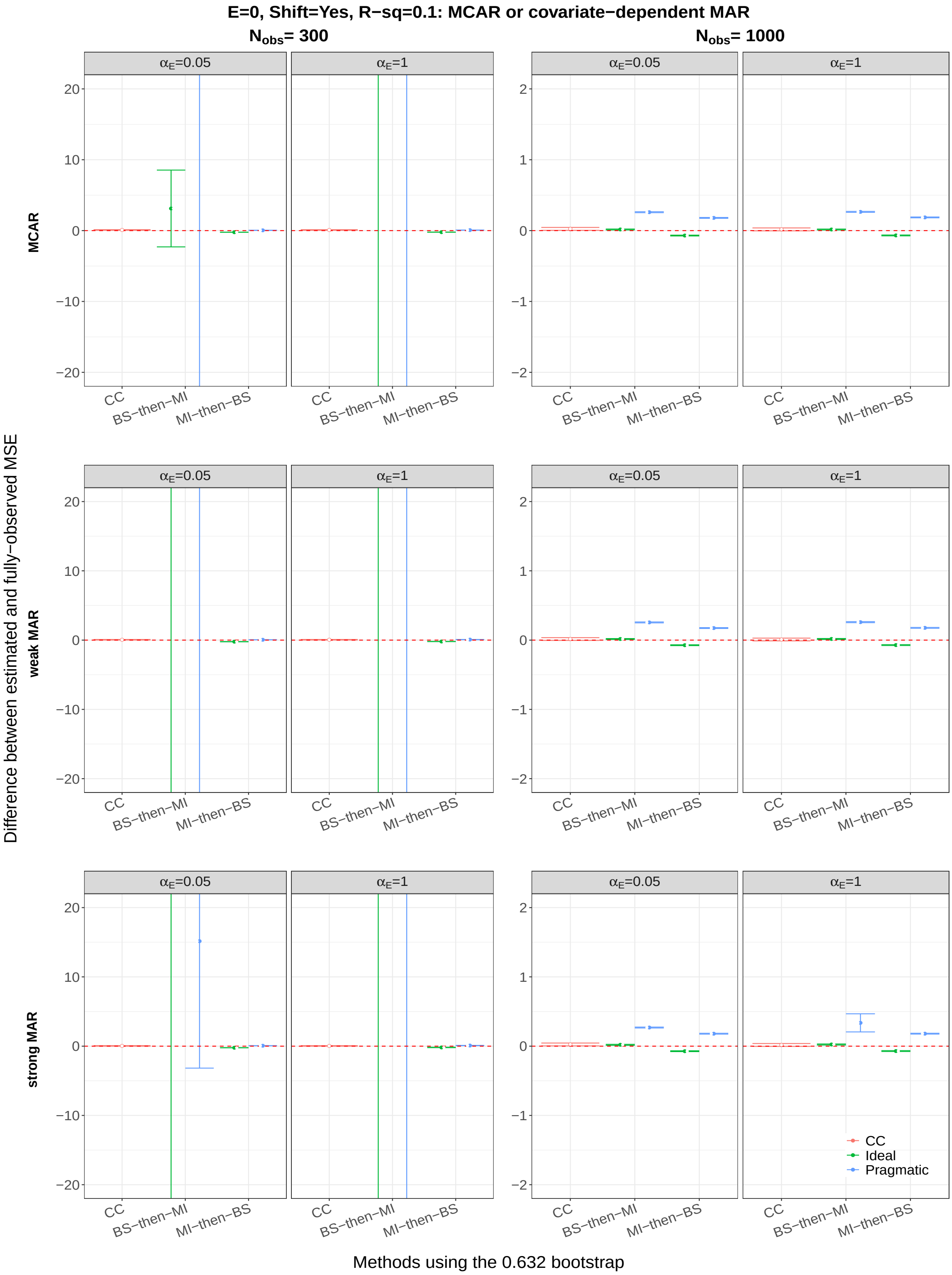

Figure S37: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

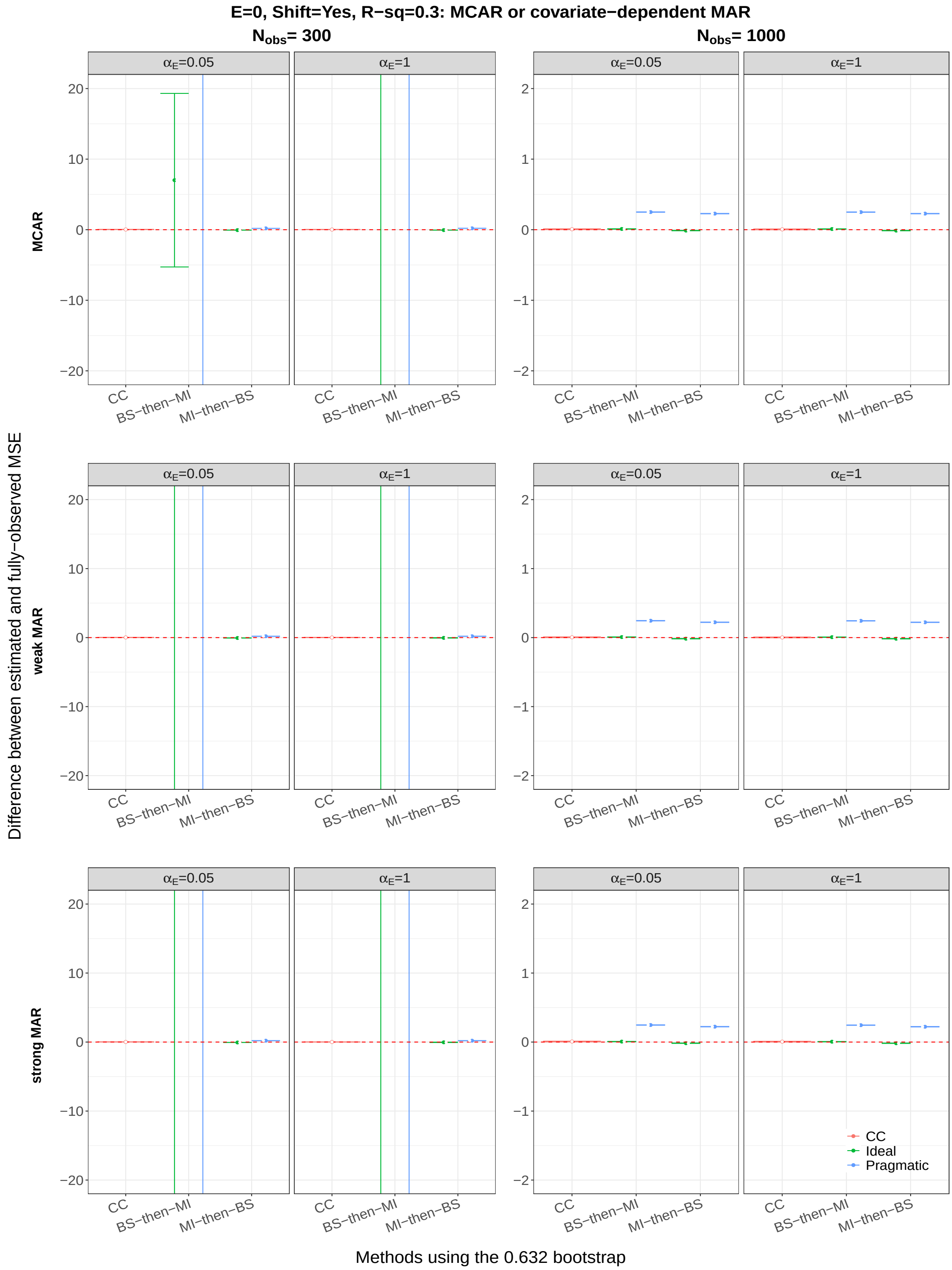

Figure S38: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

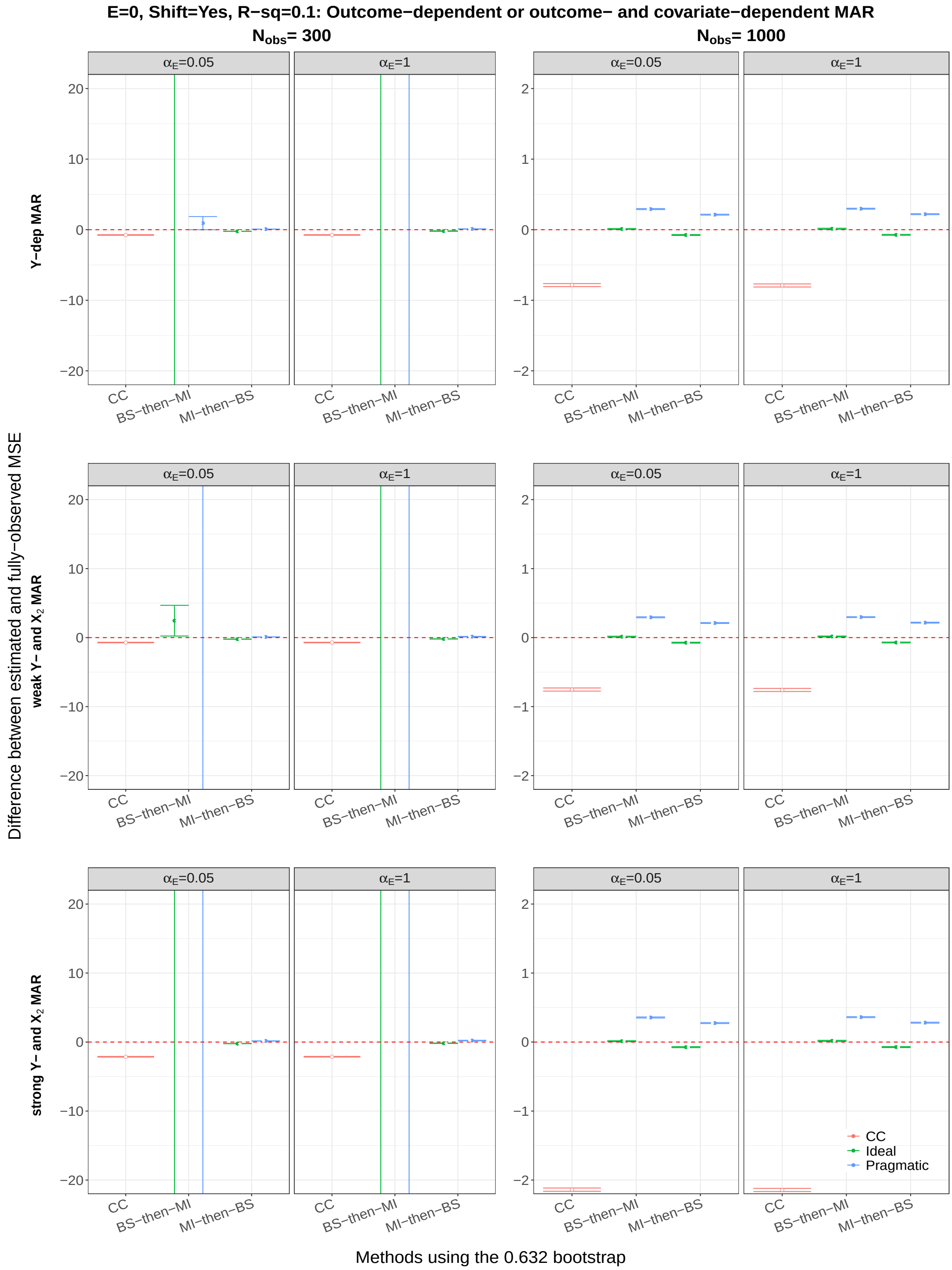

Figure S39: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

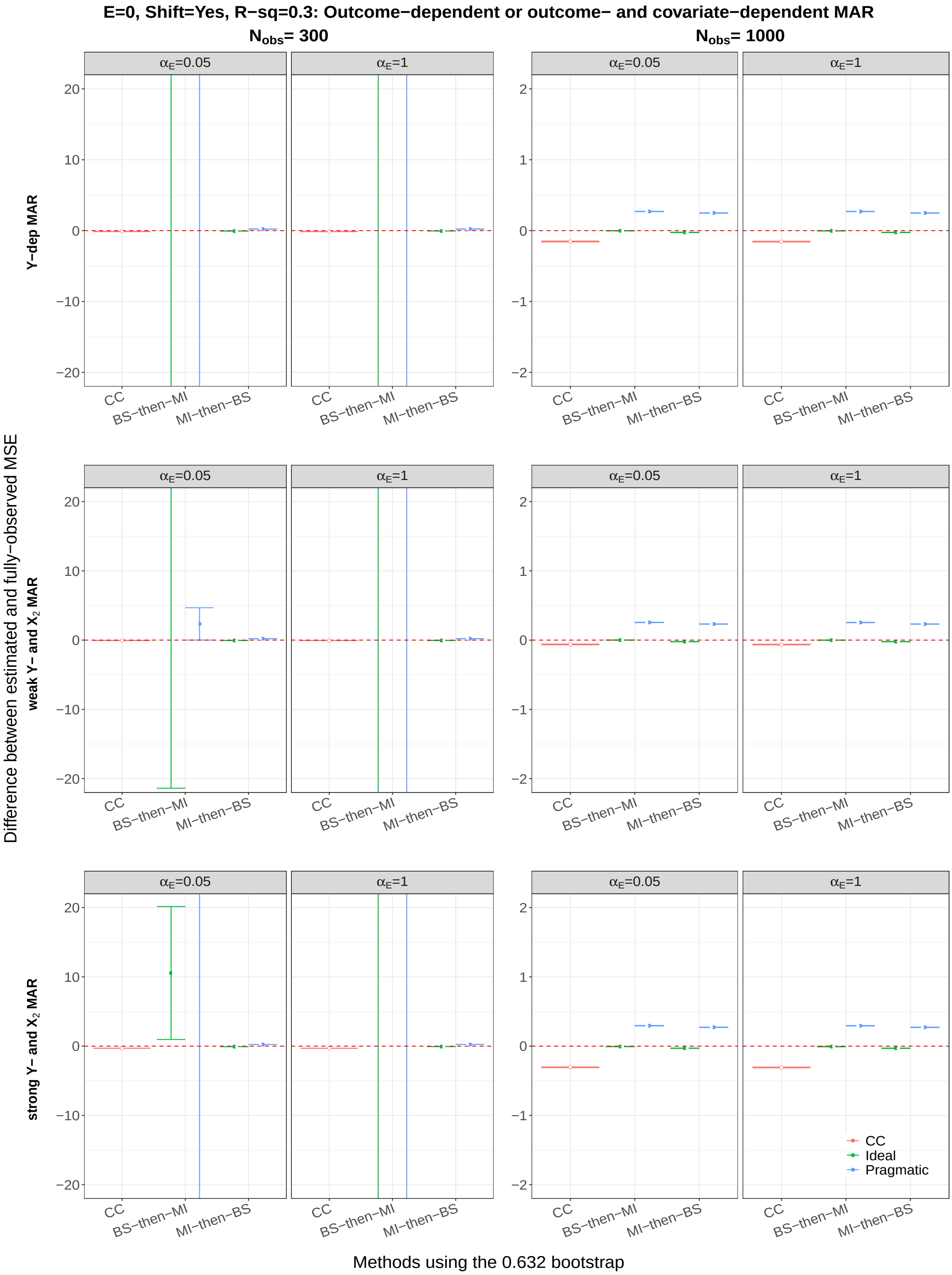

Figure S40: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

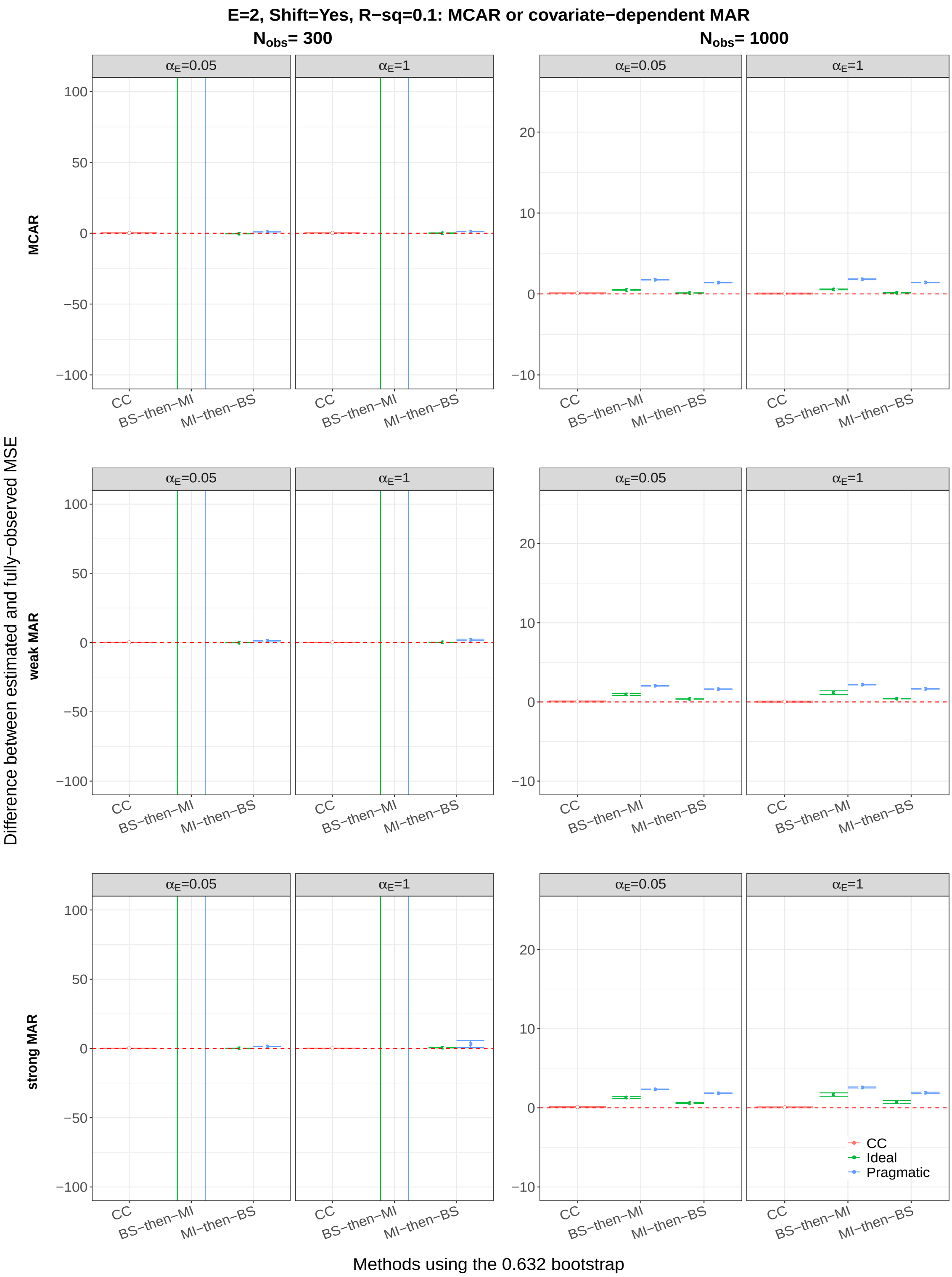

Figure S41: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

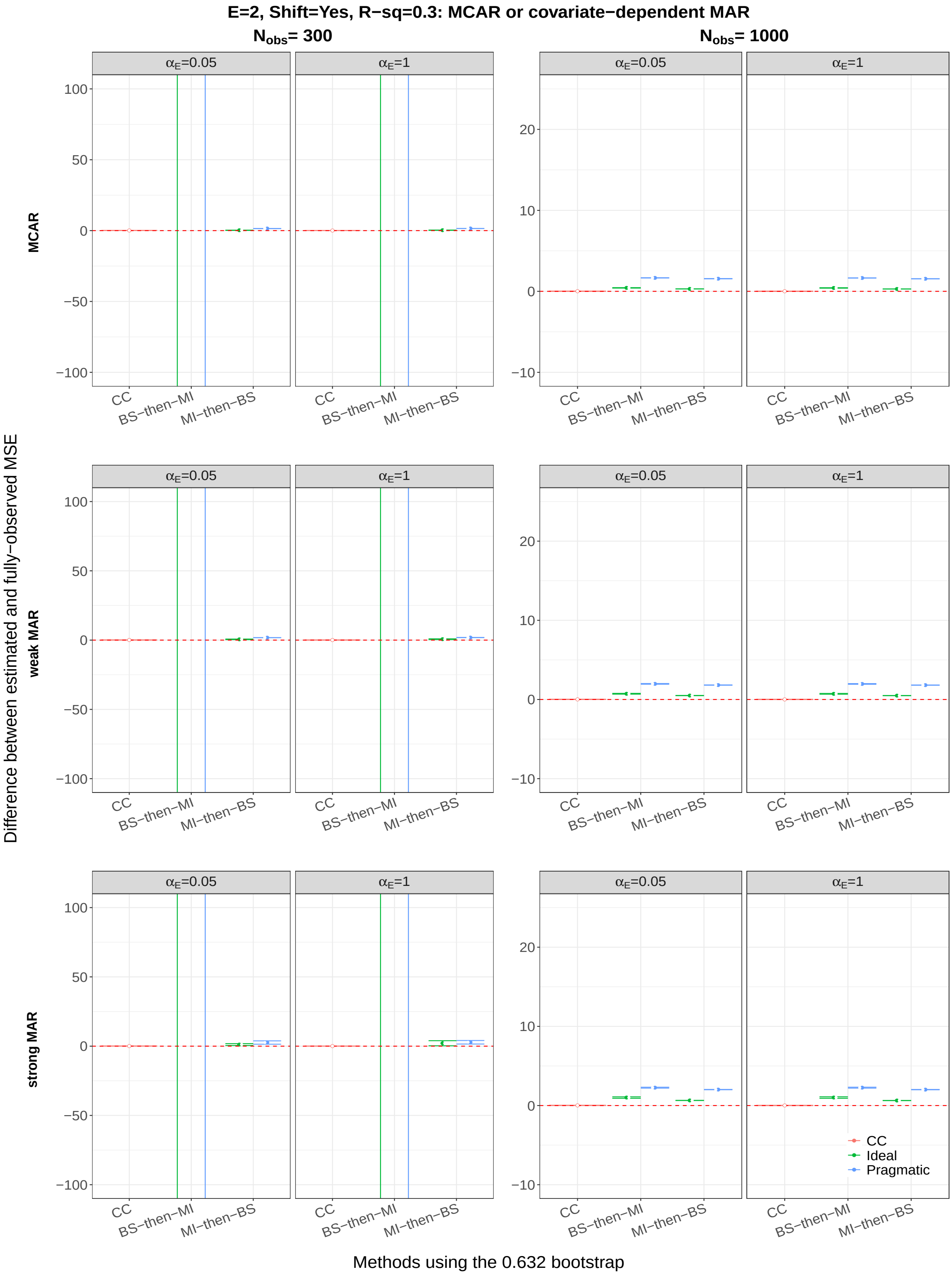

Figure S42: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

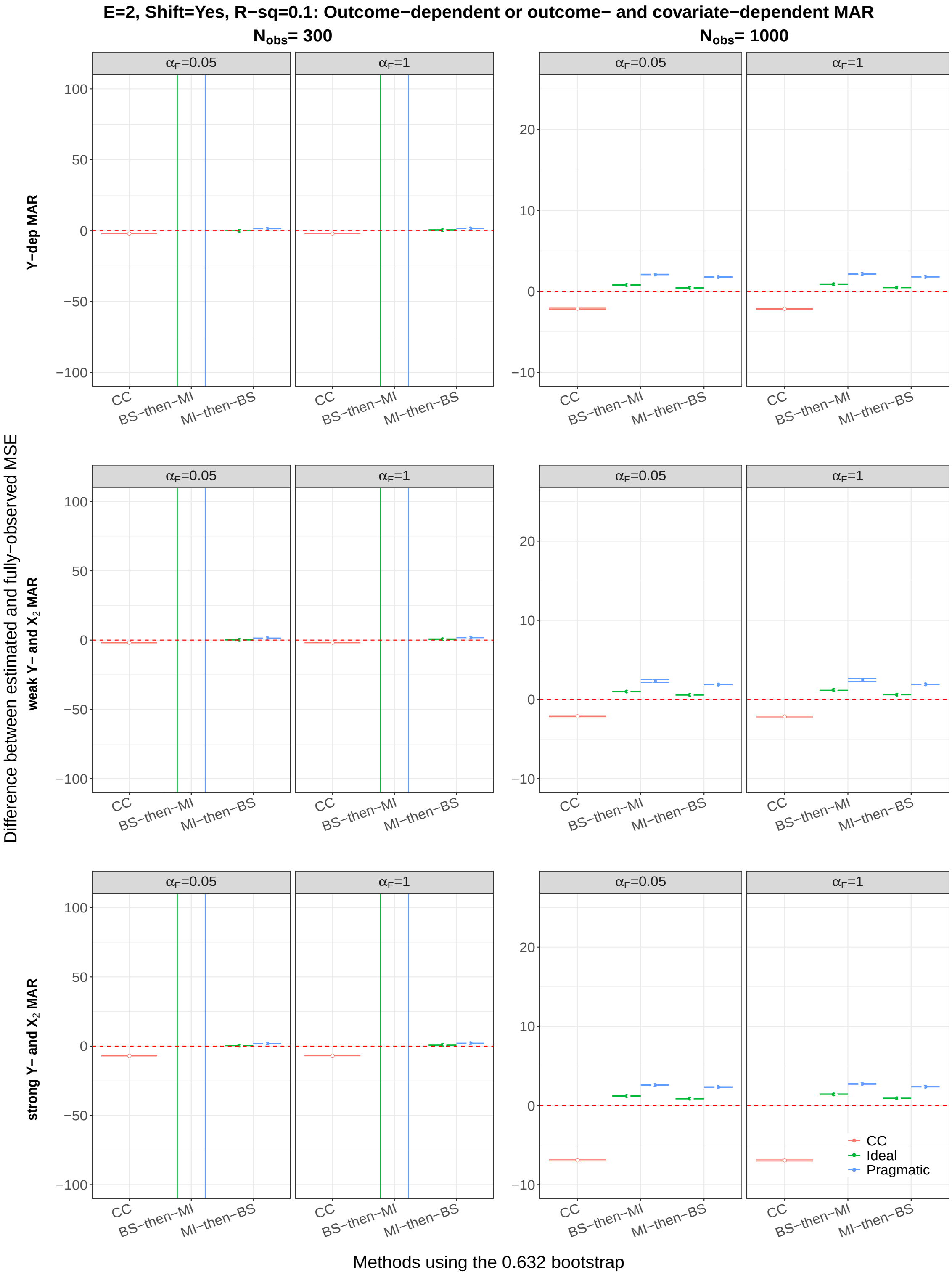

Figure S43: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

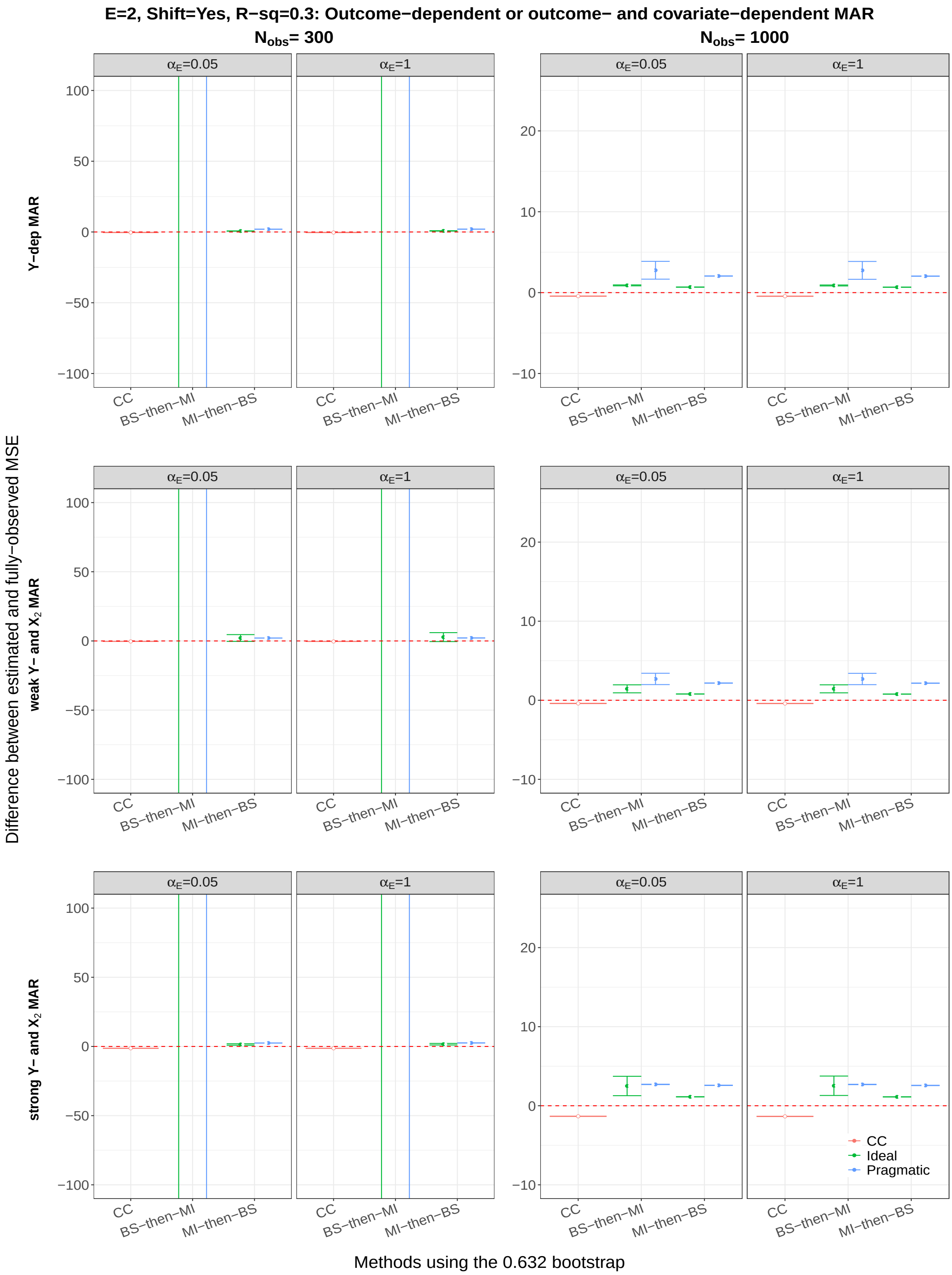

Figure S44: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

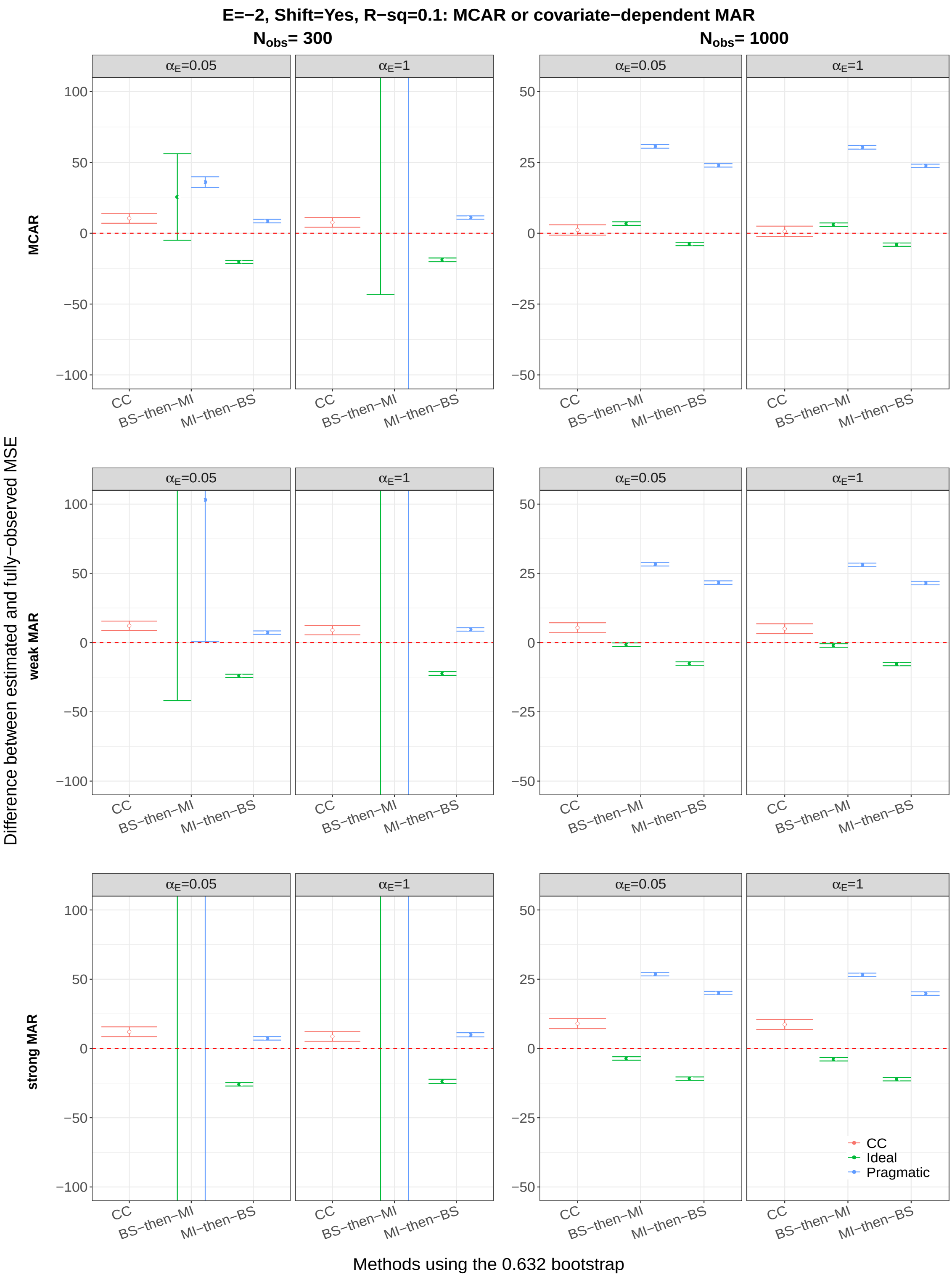

Figure S45: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

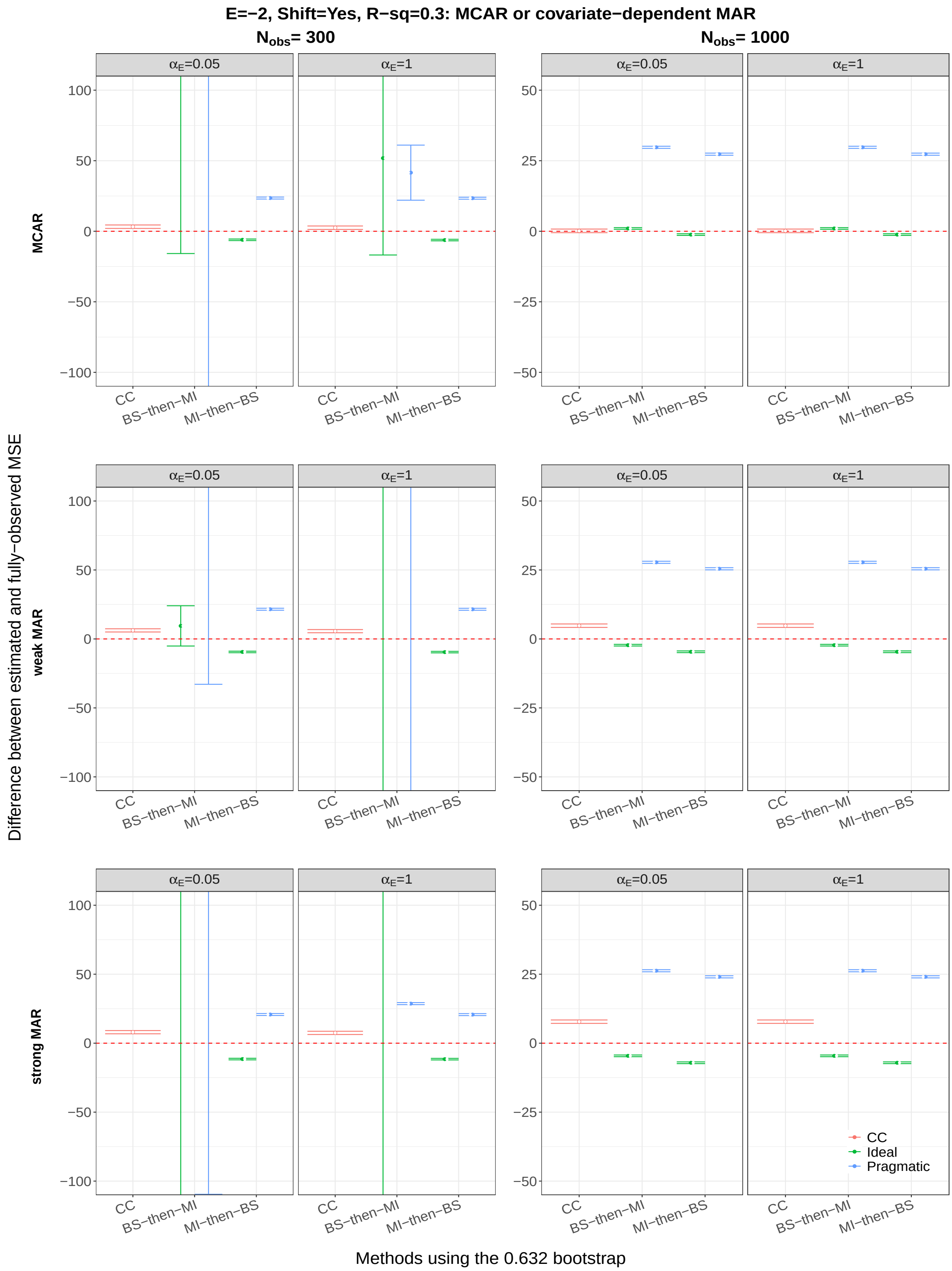

Figure S46: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

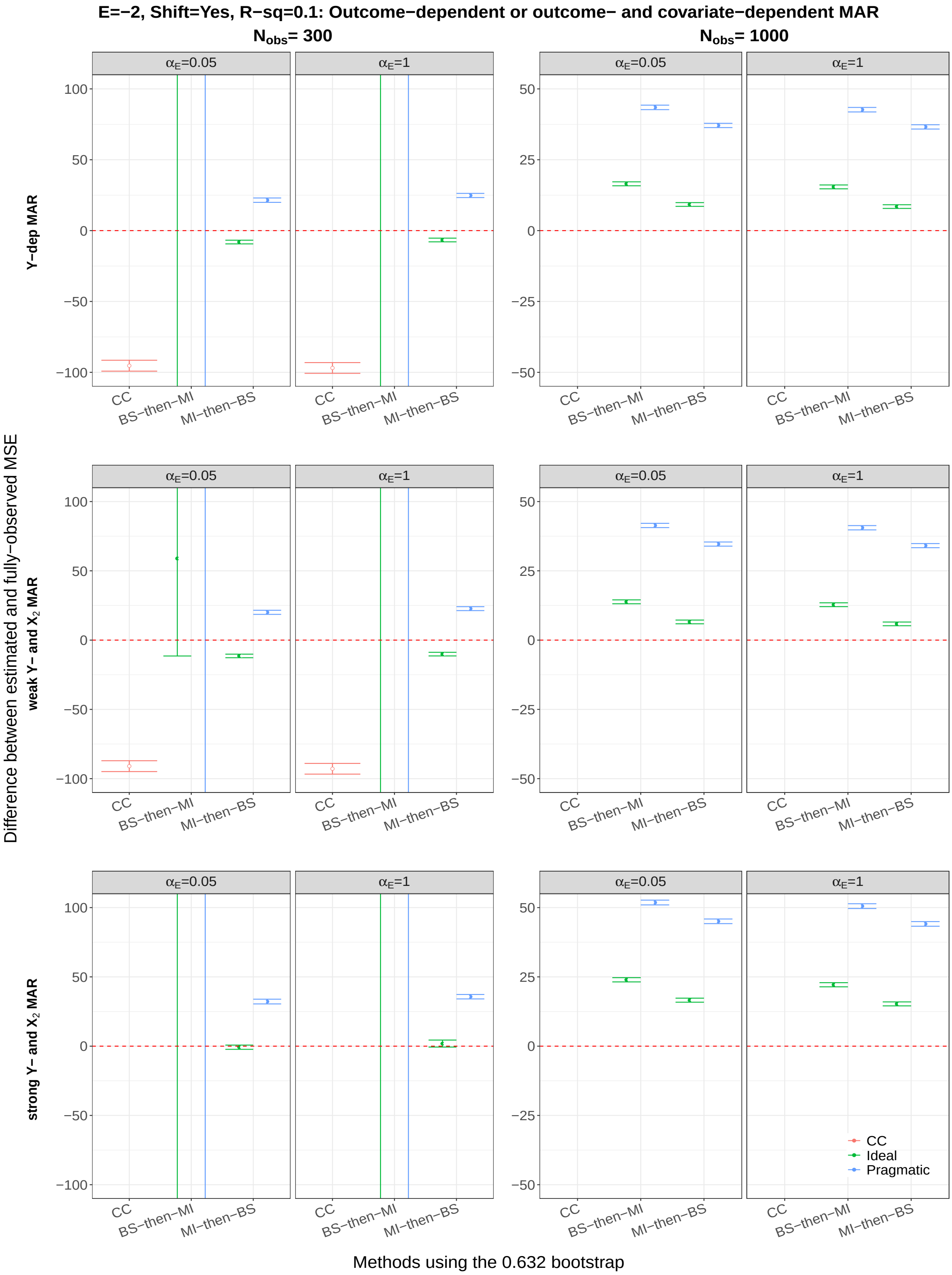

Figure S47: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

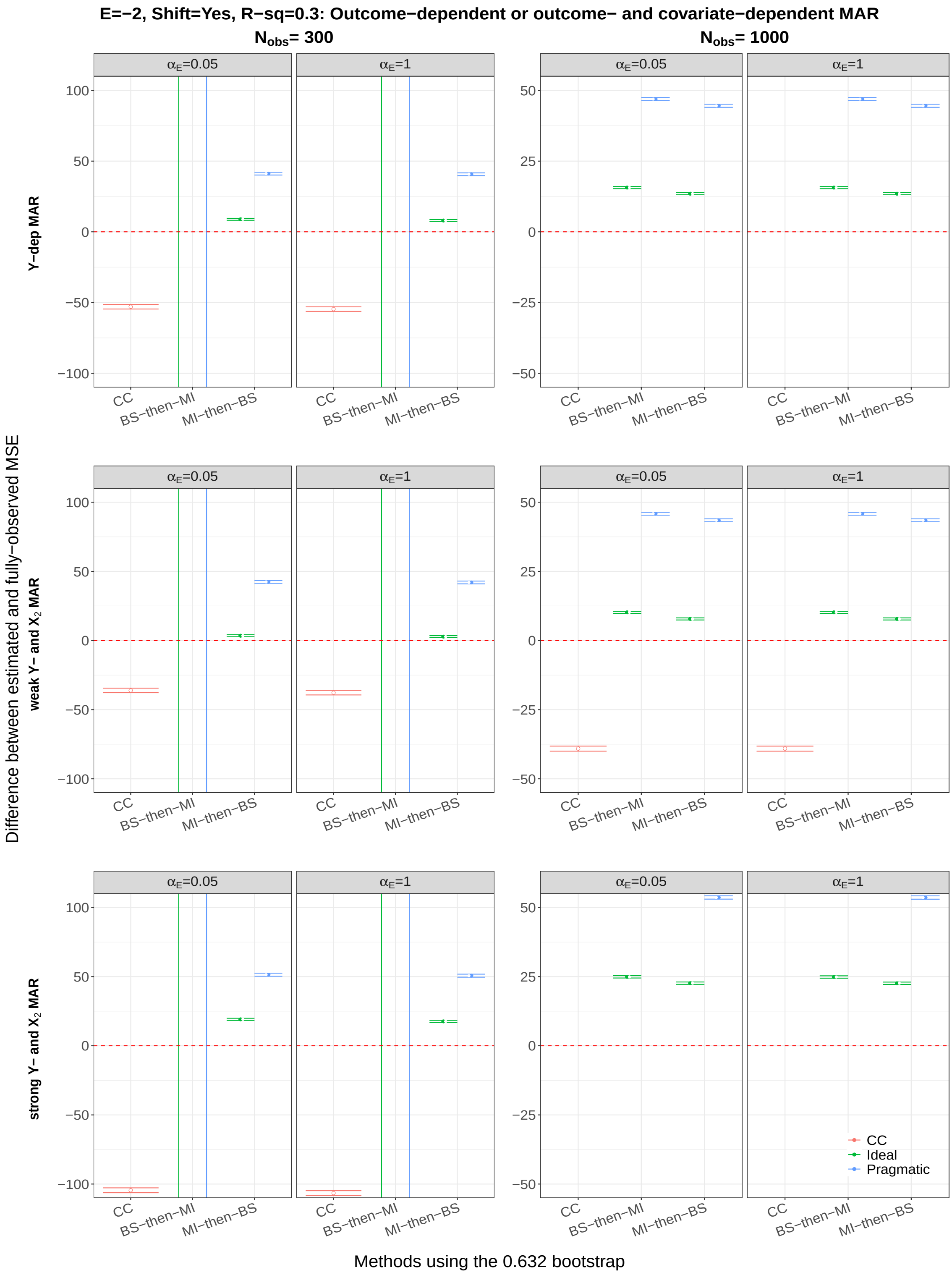

Figure S48: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6 Chapter 9: Simulation study results for FPS, exponent selection (Section 9.3)

S6.1 ABB exponent selection

S6.1.1 Cross-validation

True exponent is 0

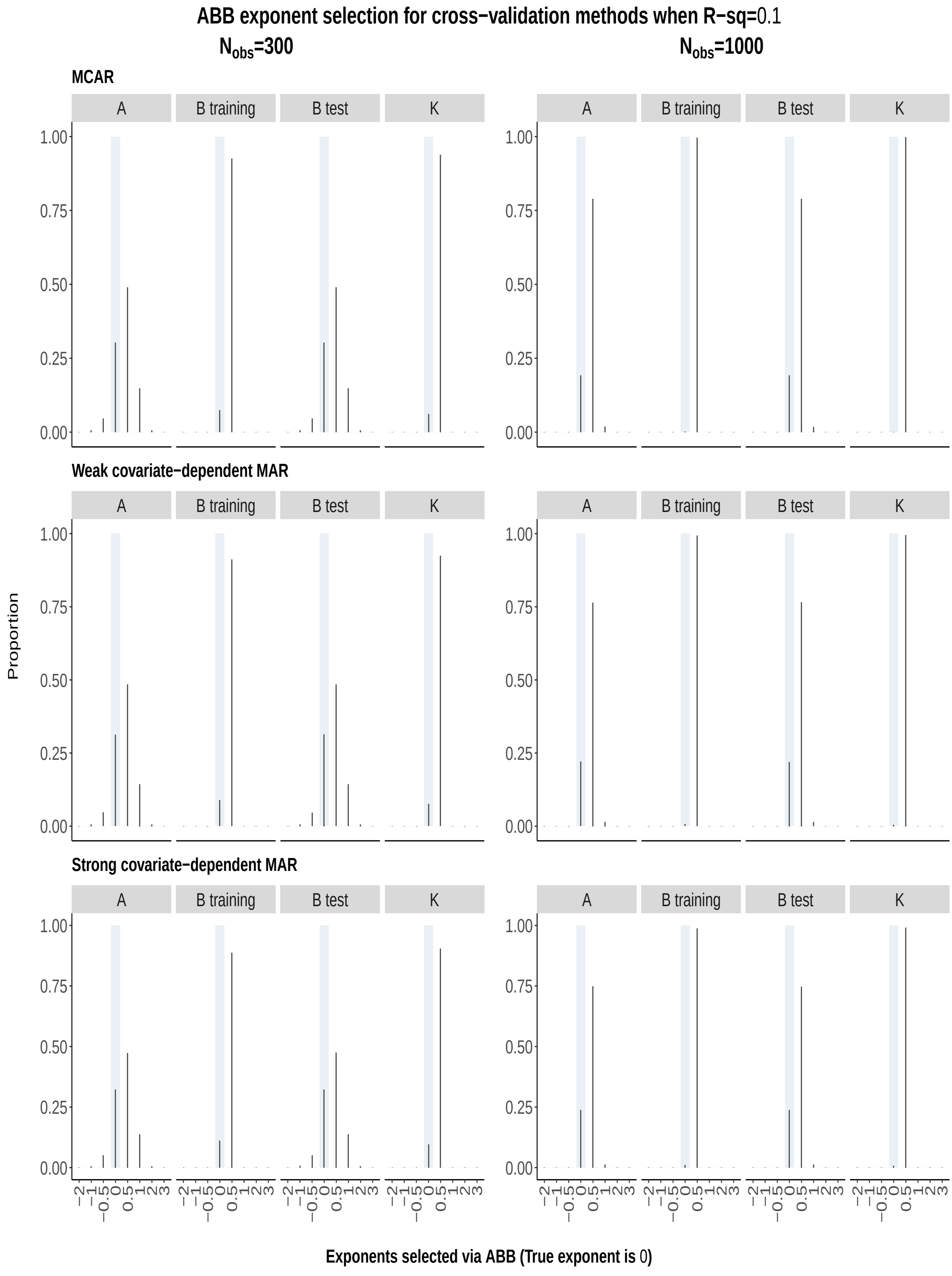

Figure S1: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

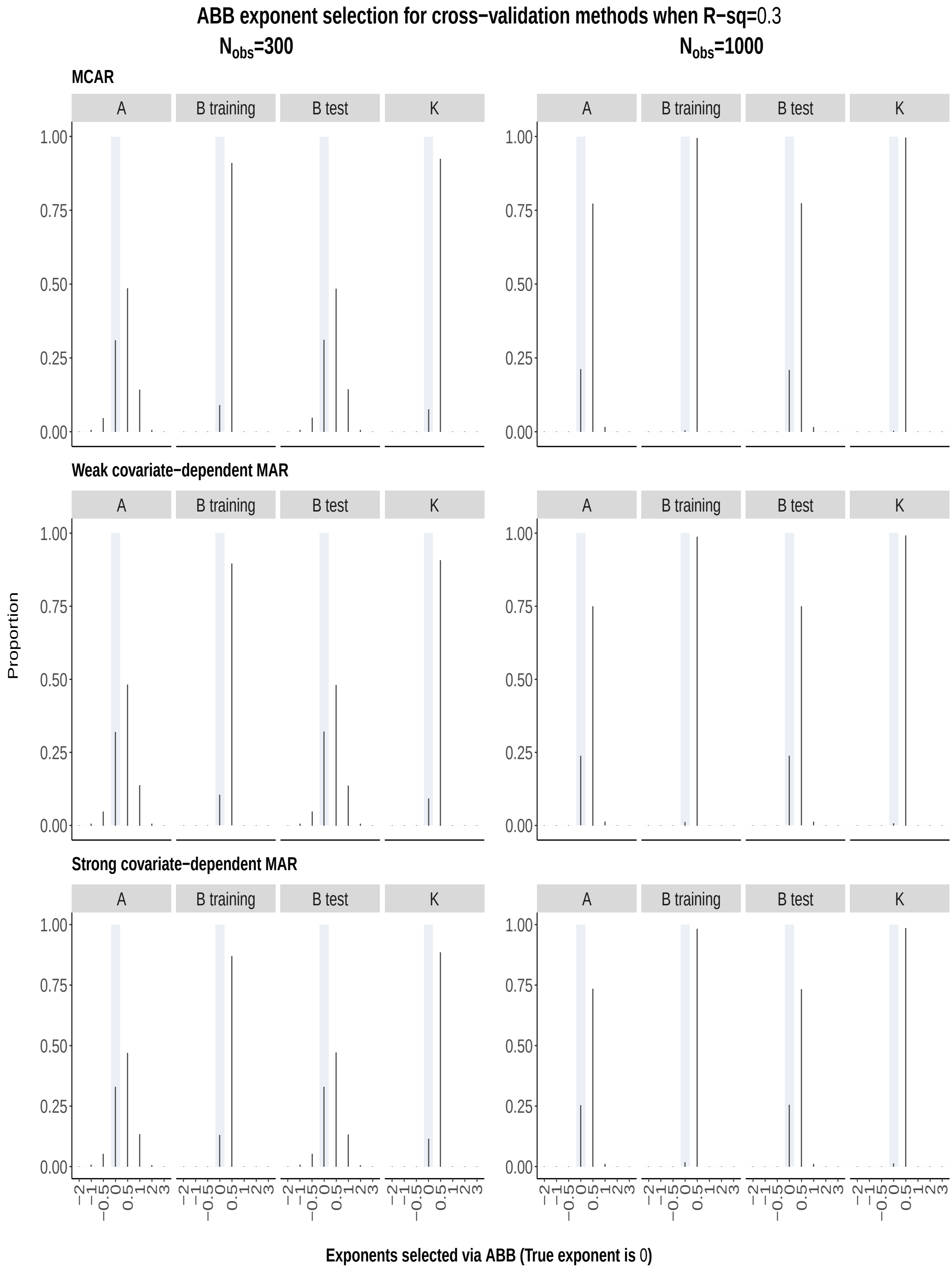

Figure S2: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

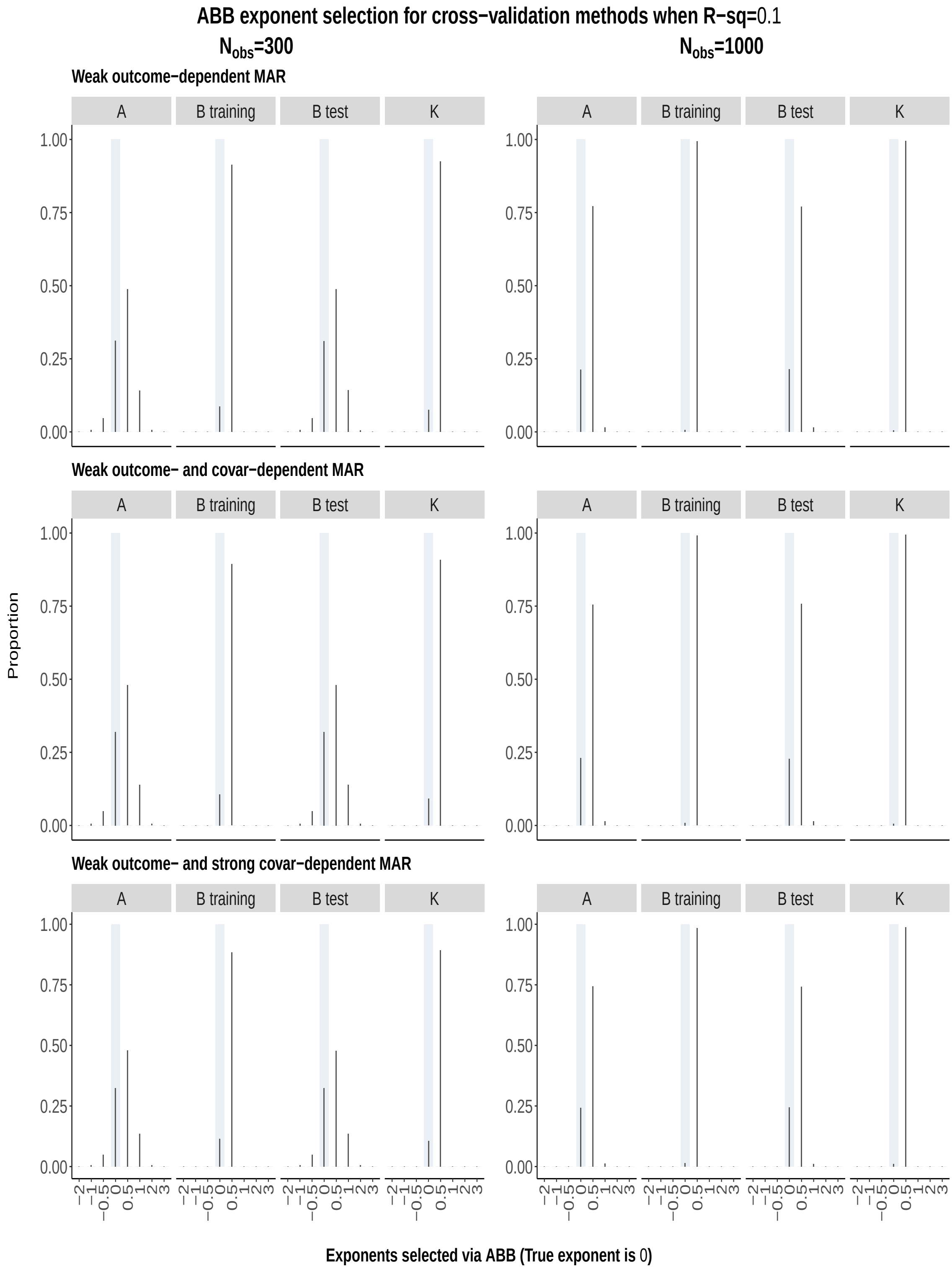

Figure S3: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

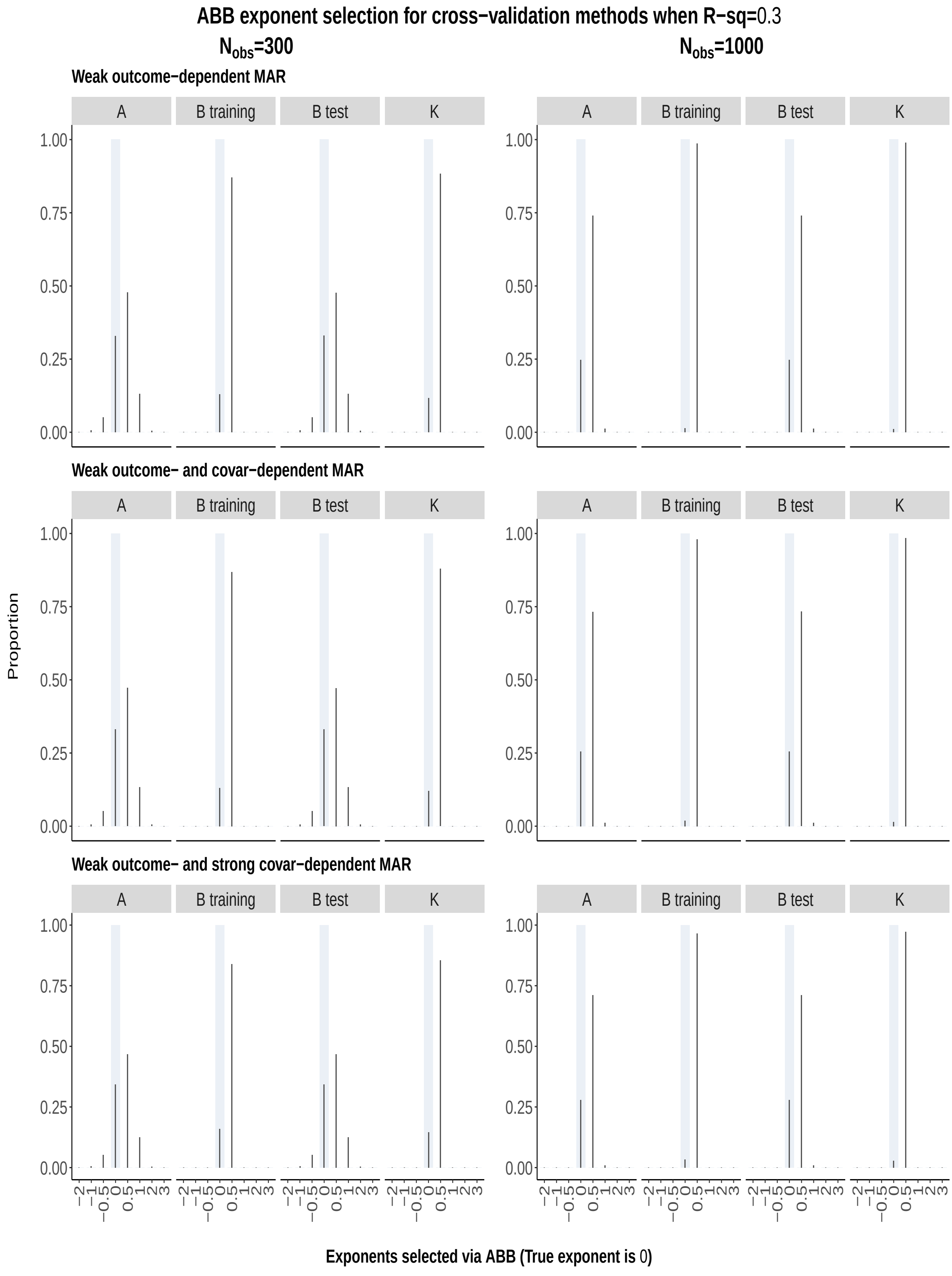

Figure S4: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

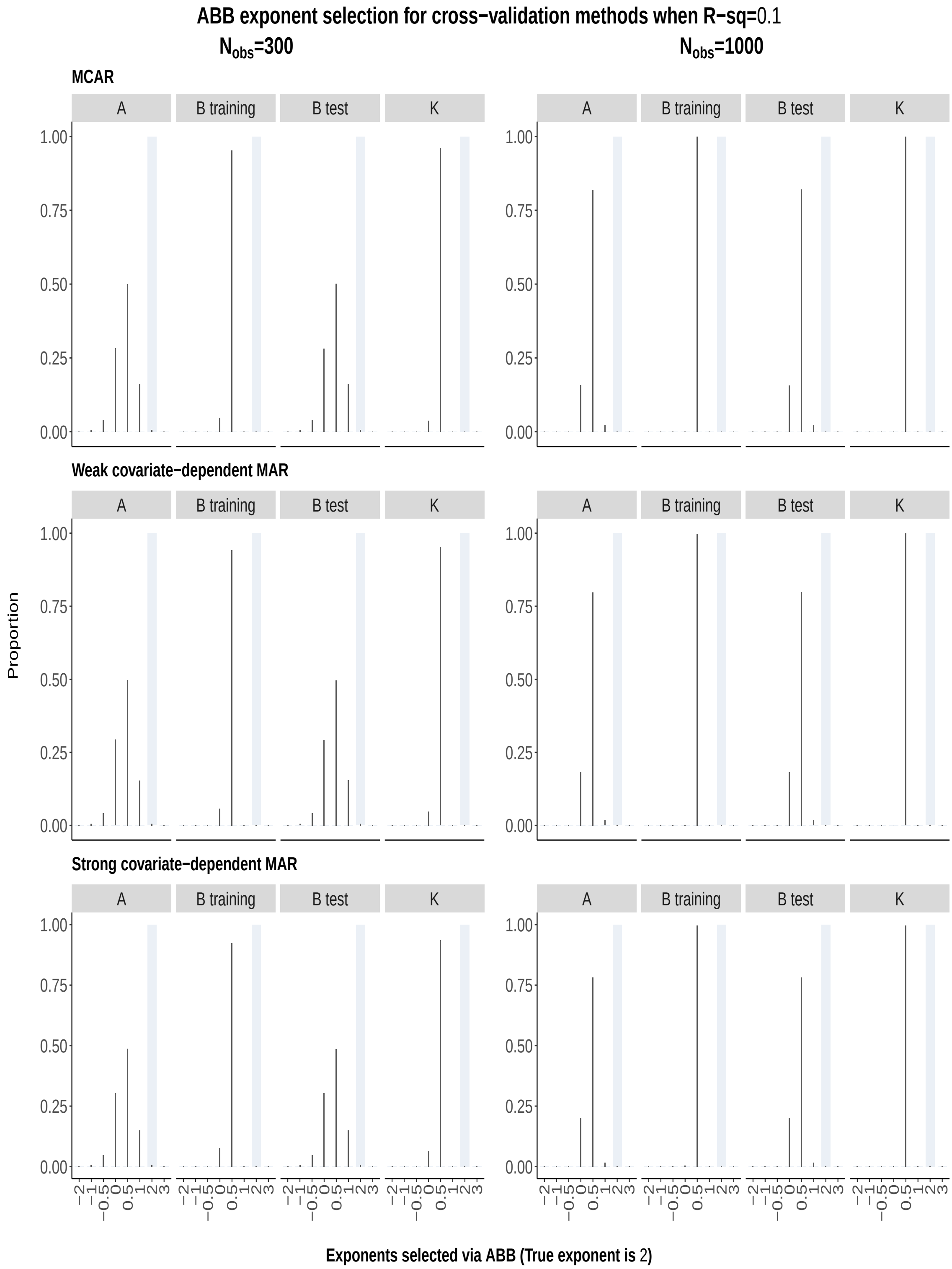

Figure S5: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

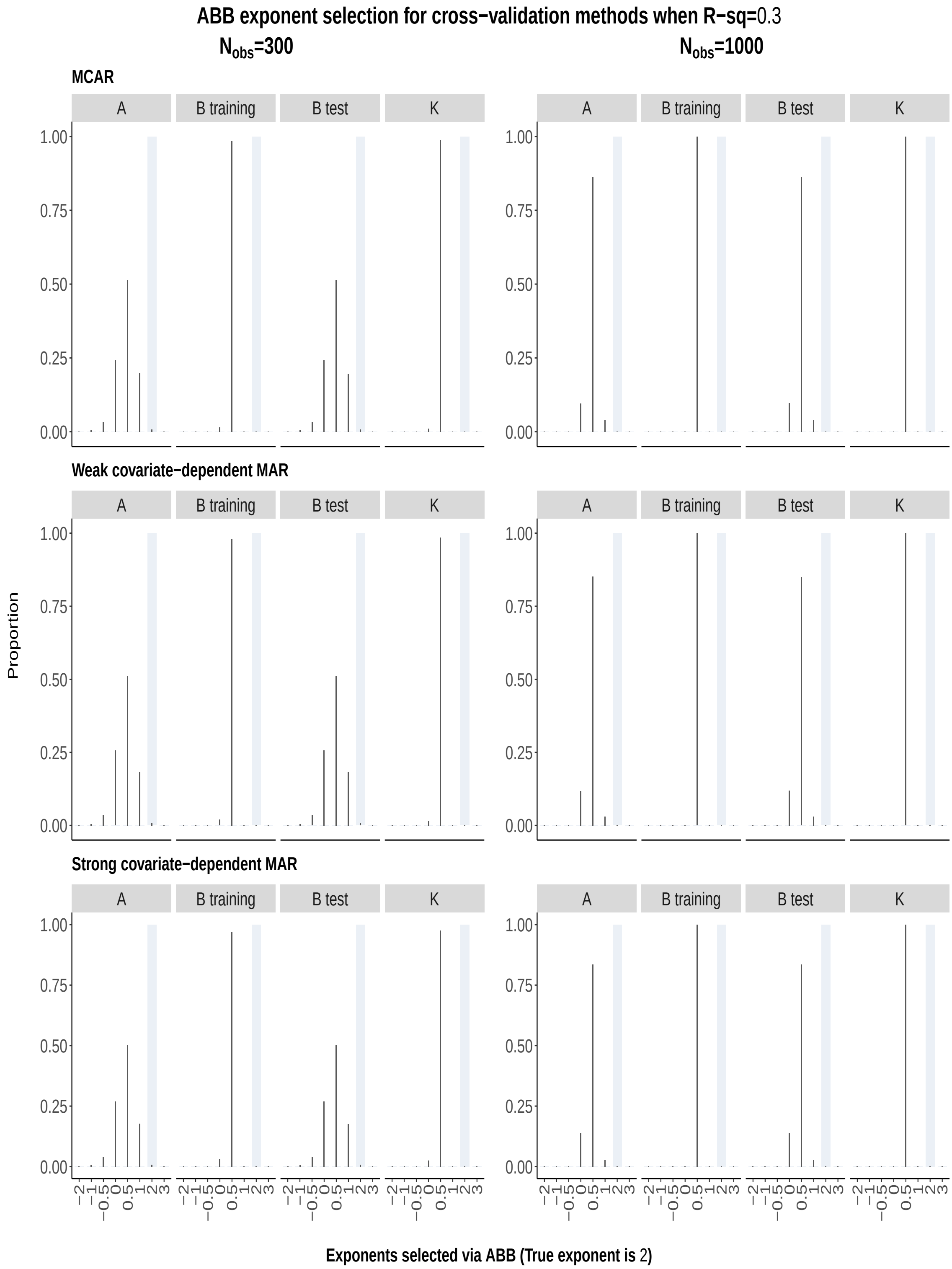

Figure S6: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

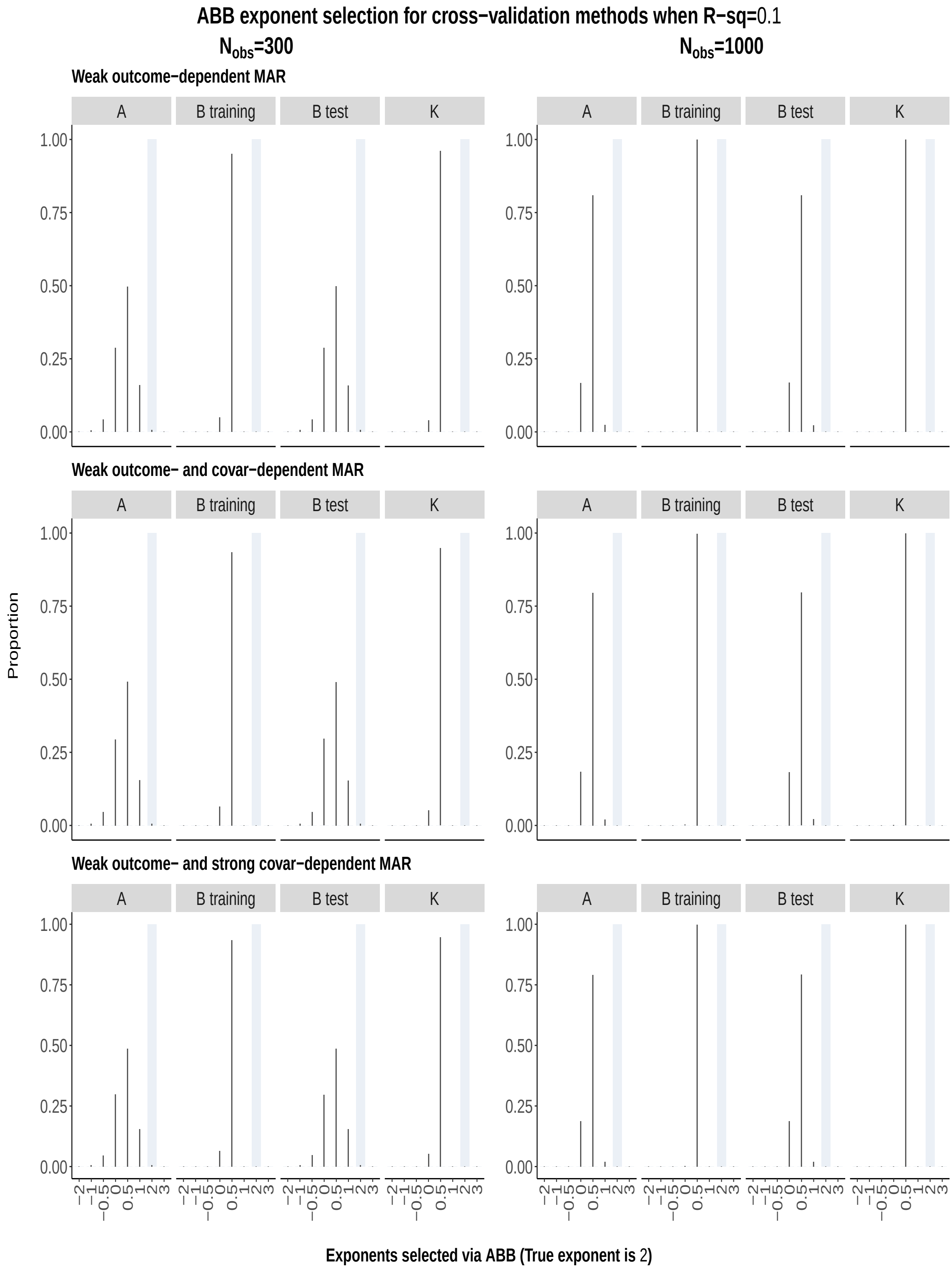

Figure S7: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

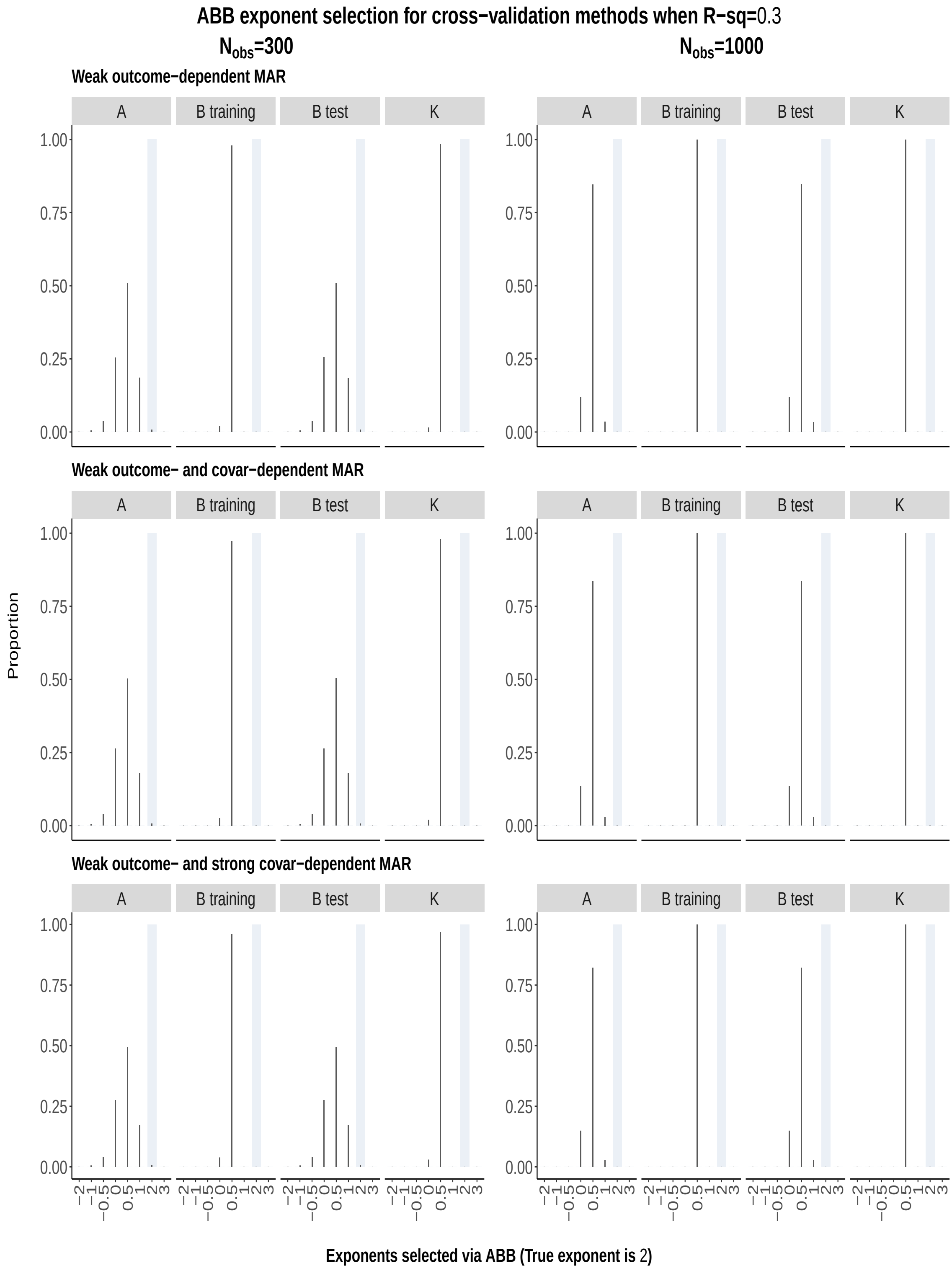

Figure S8: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

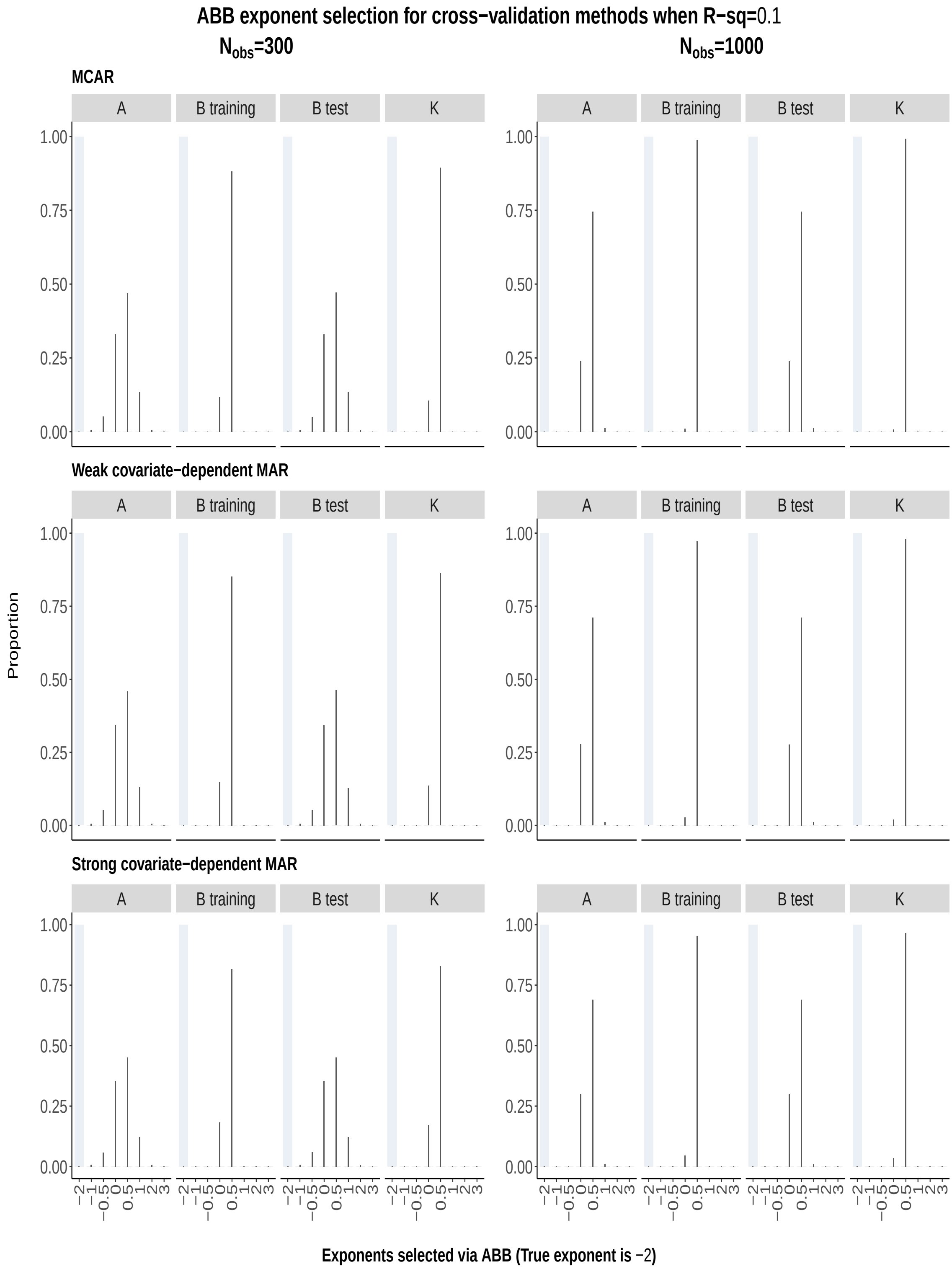

Figure S9: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

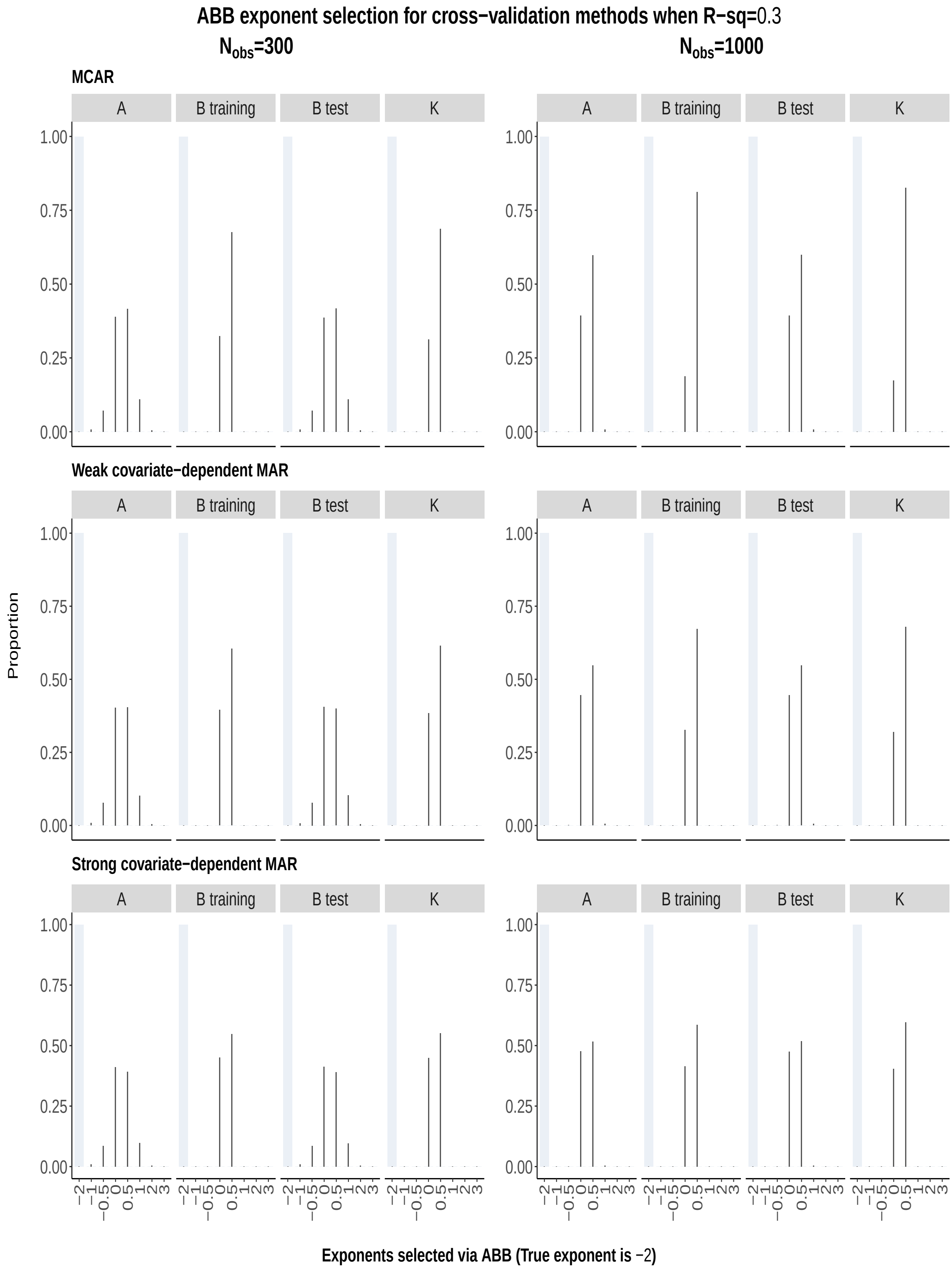

Figure S10: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

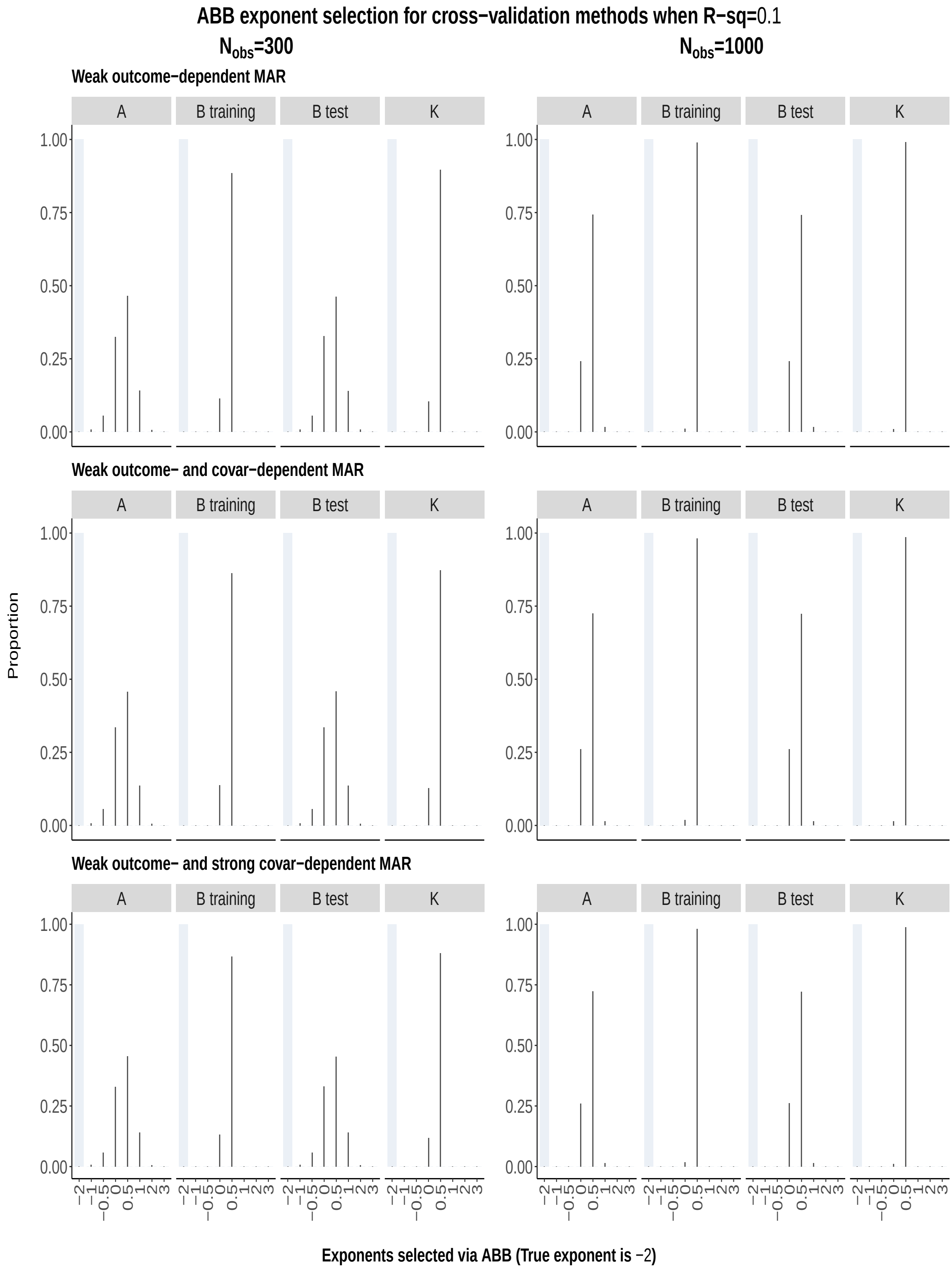

Figure S11: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

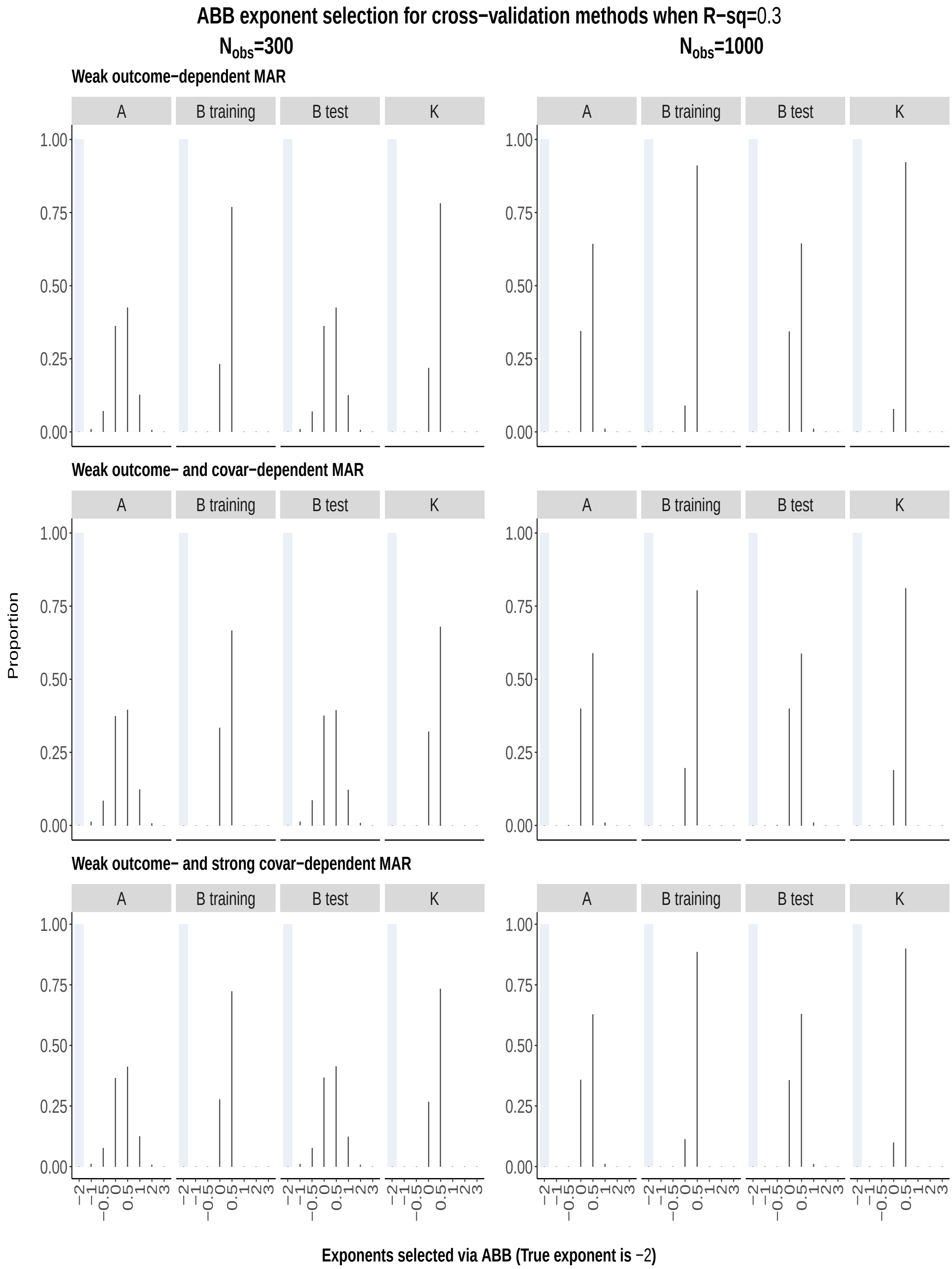

Figure S12: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.1.2 The 0.632 bootstrap

True exponent is 0

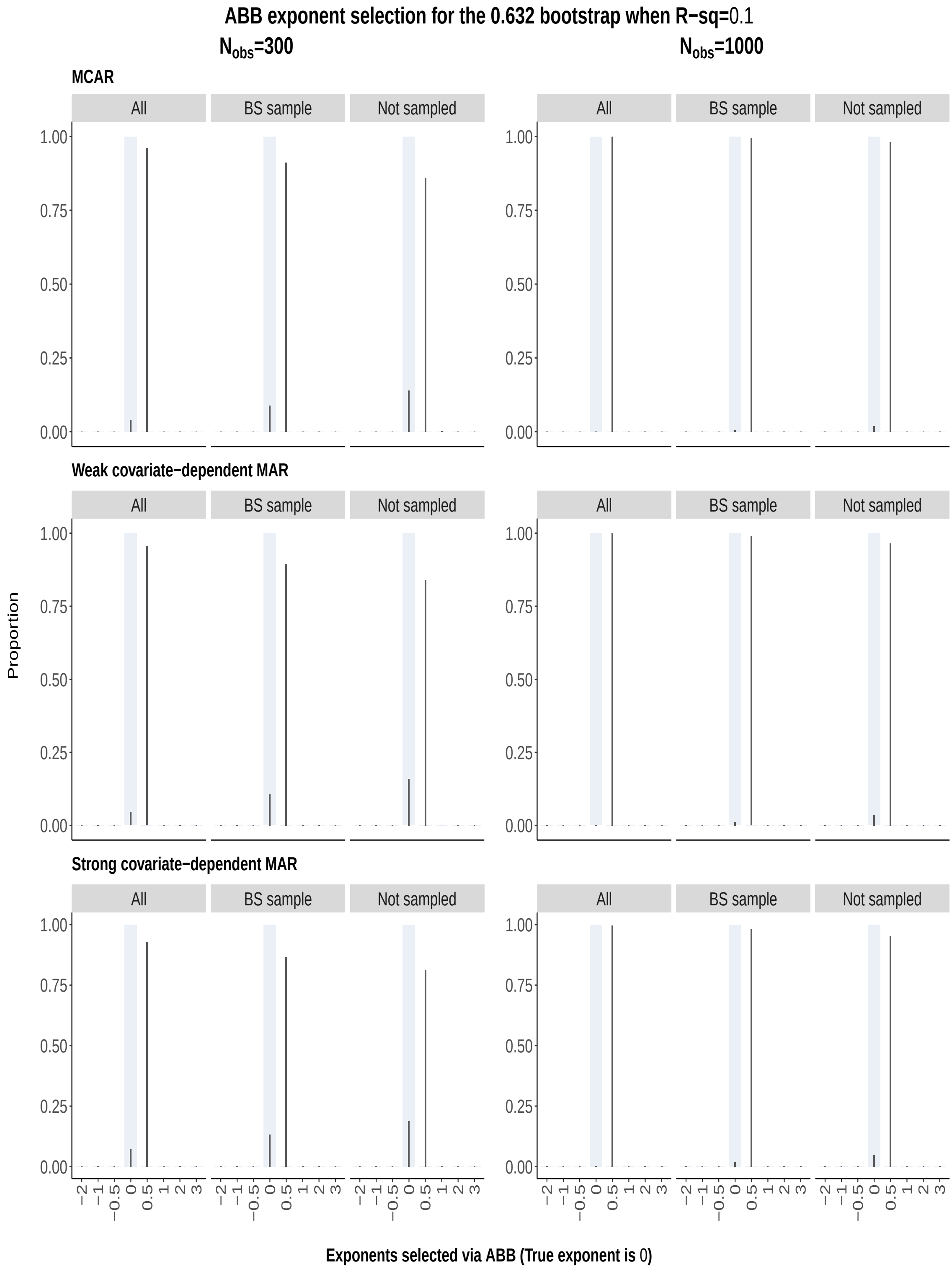

Figure S13: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

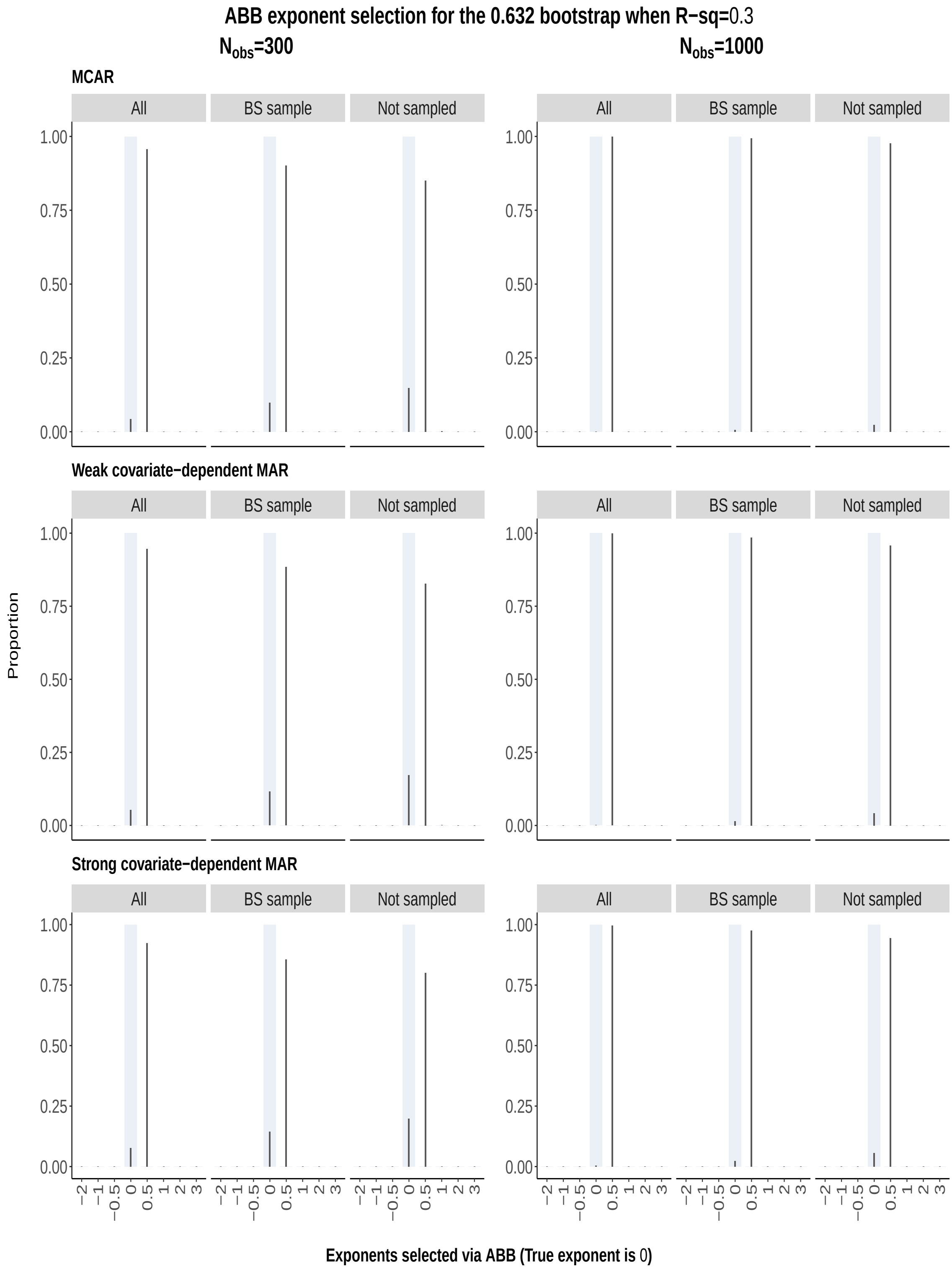

Figure S14: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

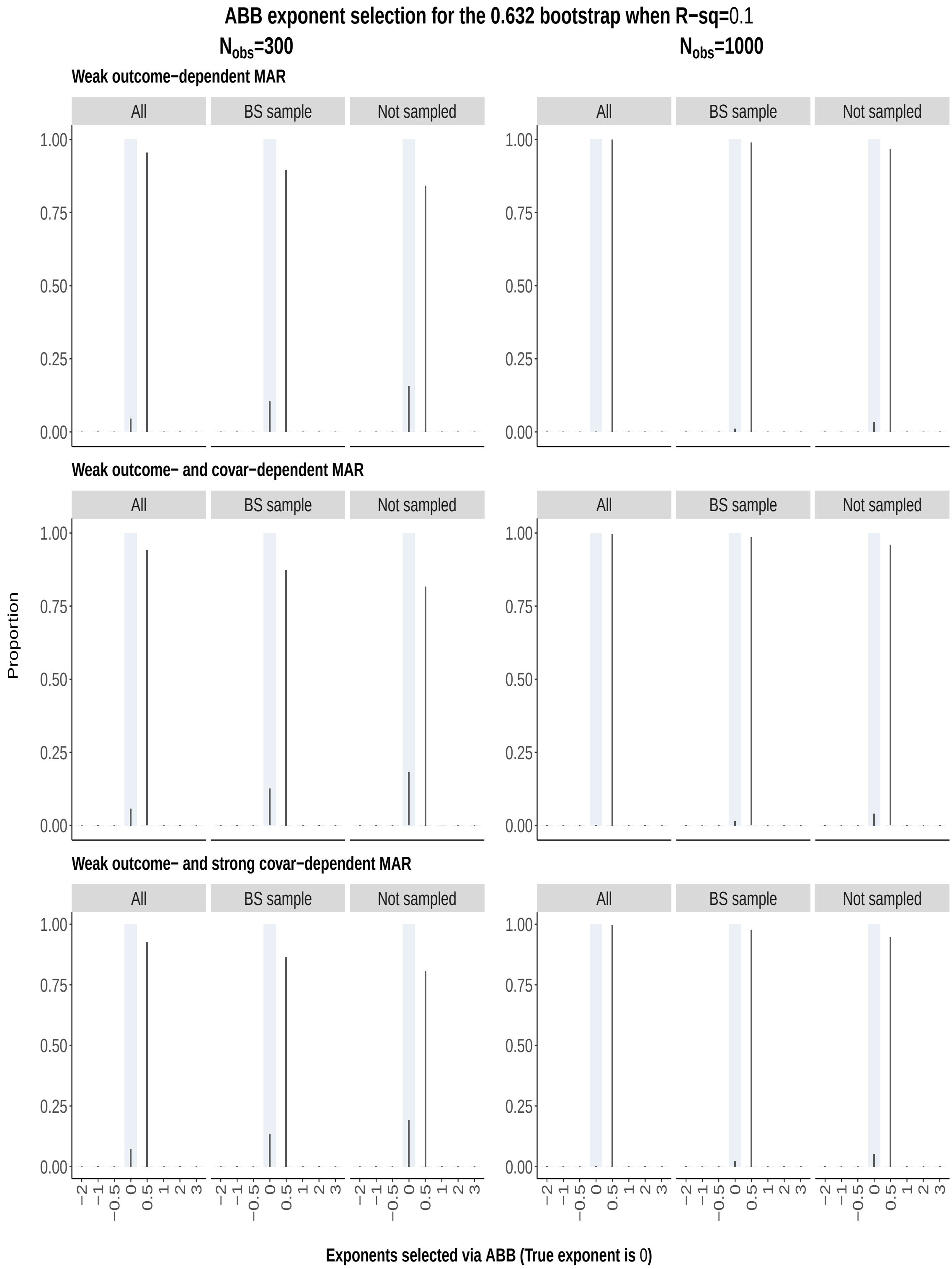

Figure S15: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

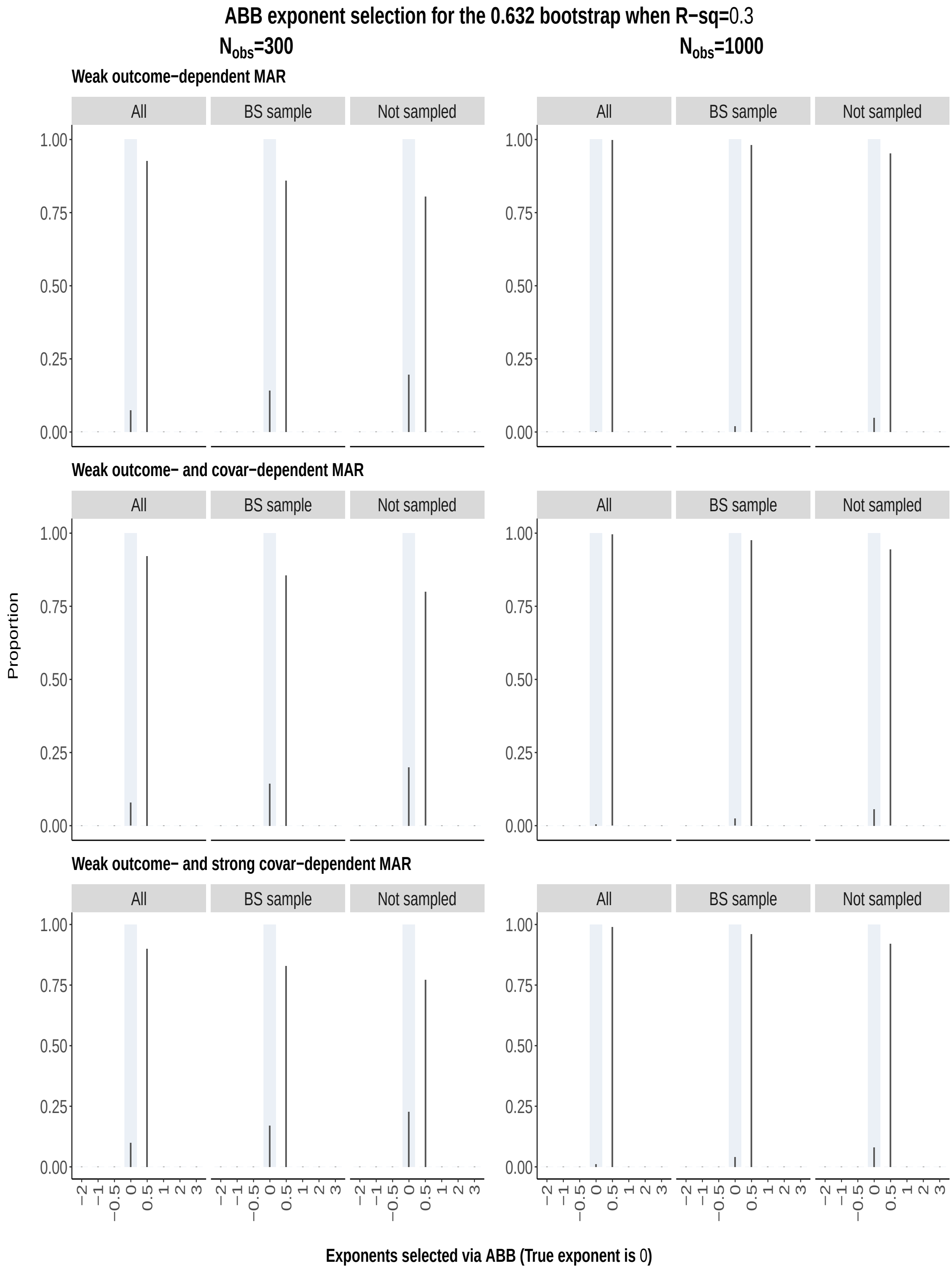

Figure S16: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

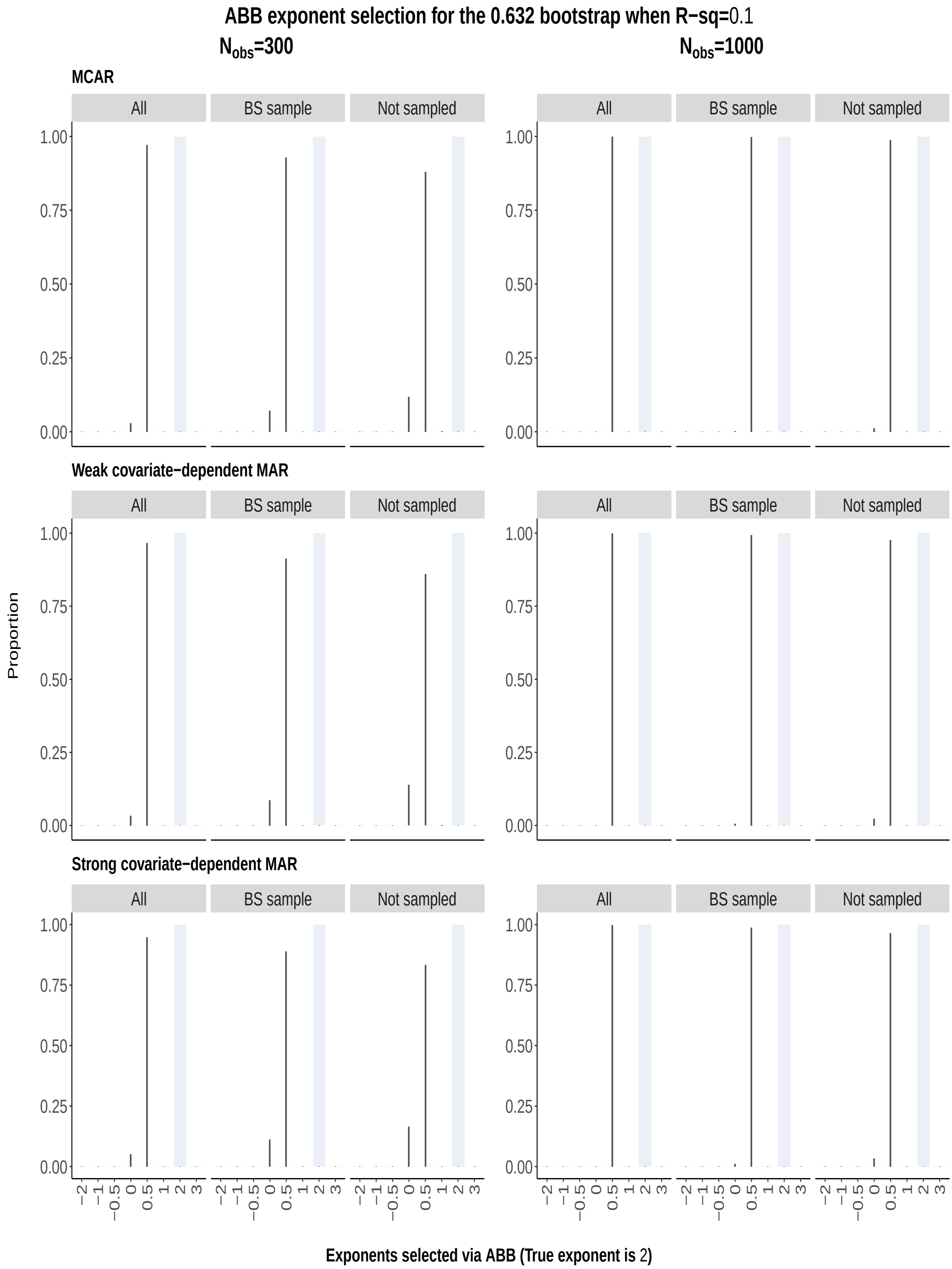

Figure S17: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

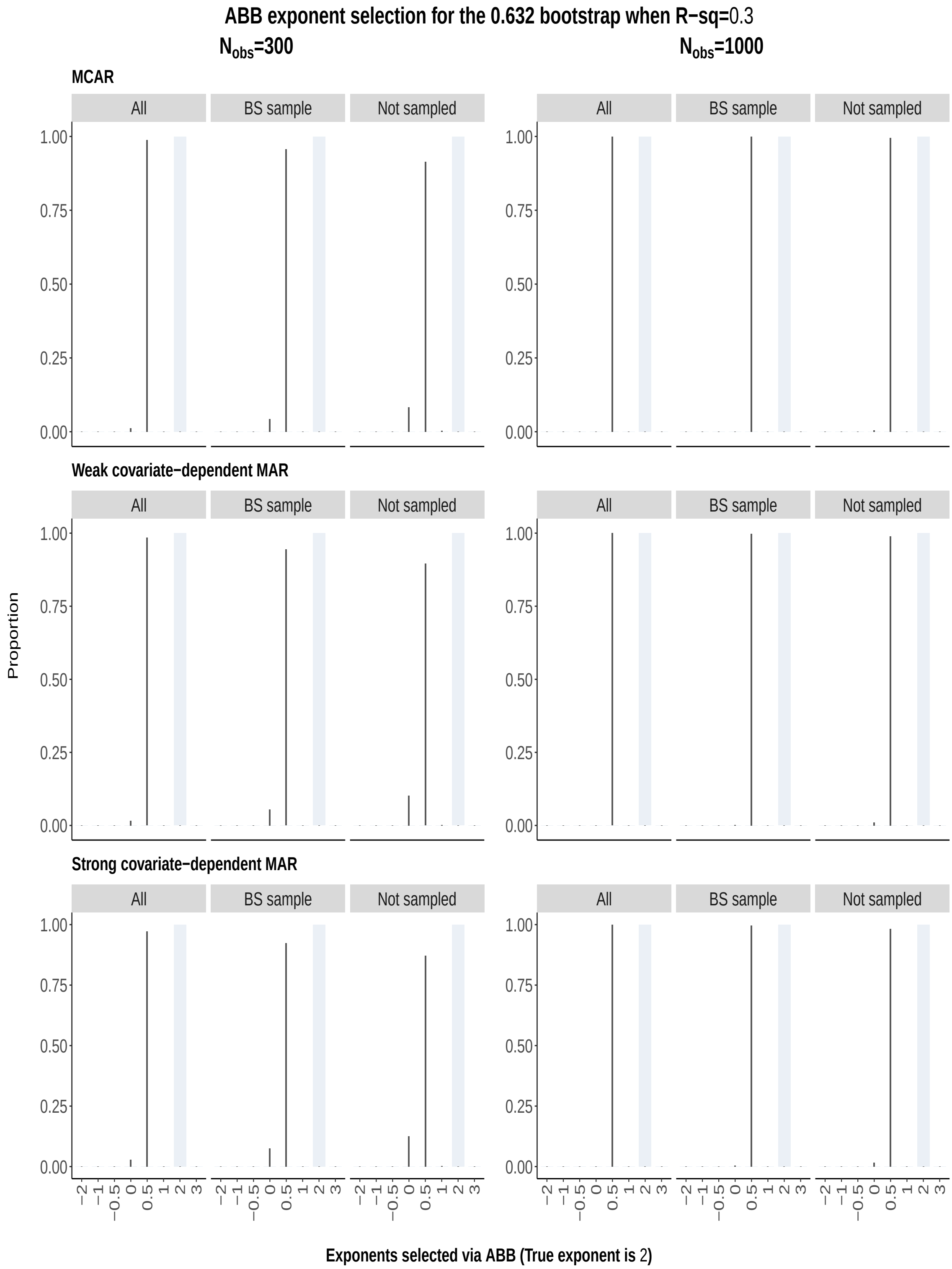

Figure S18: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

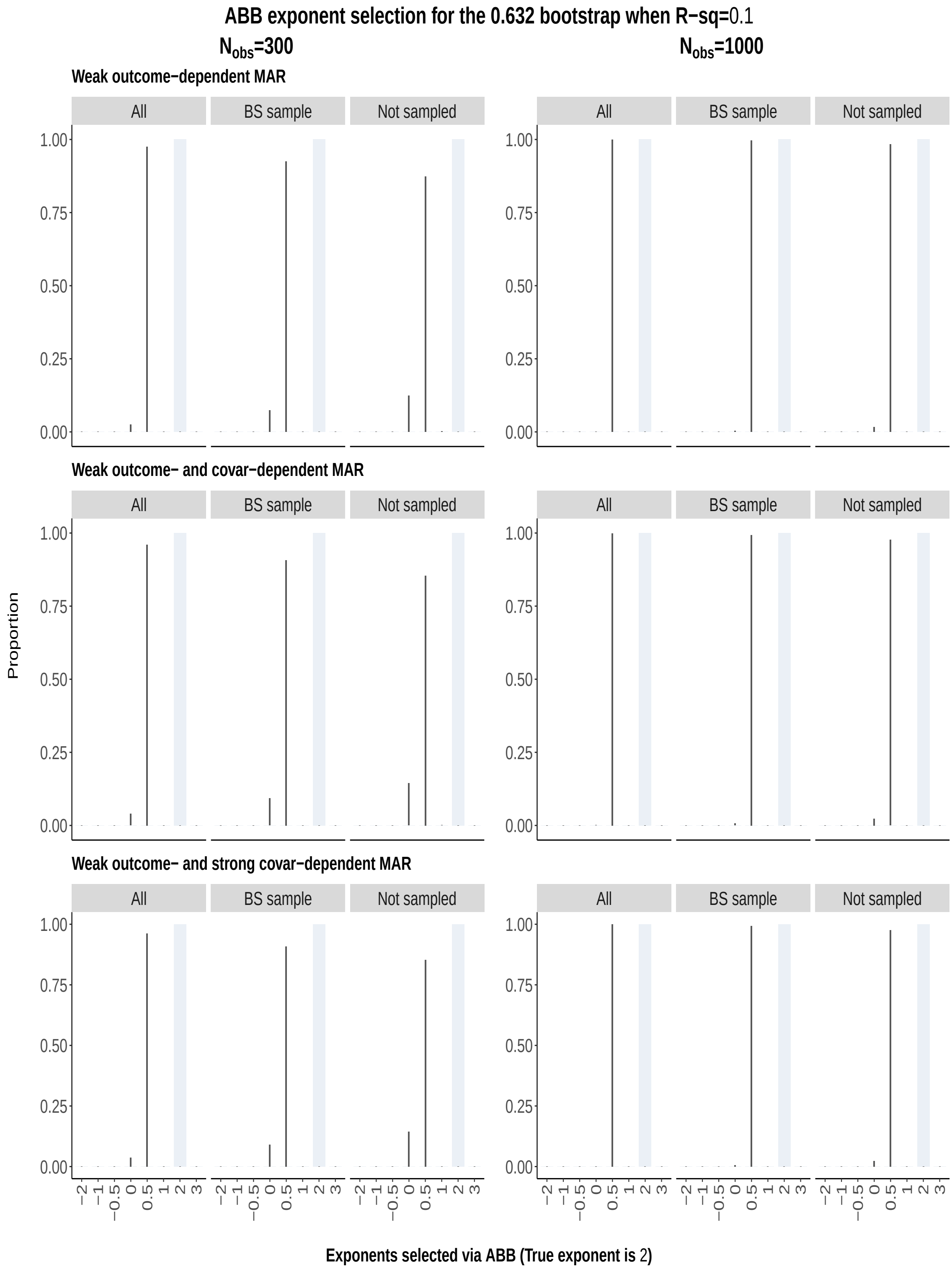

Figure S19: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

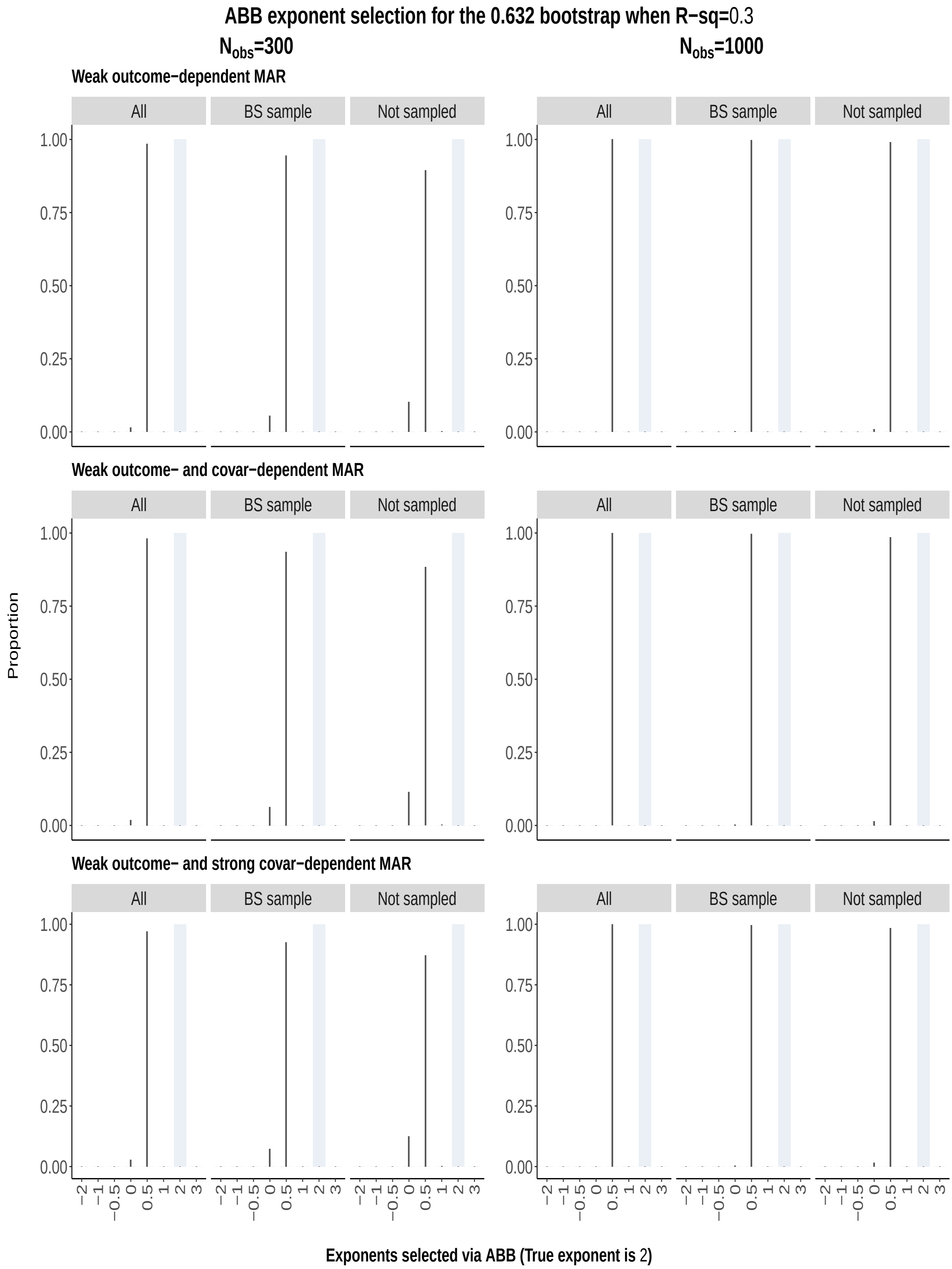

Figure S20: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

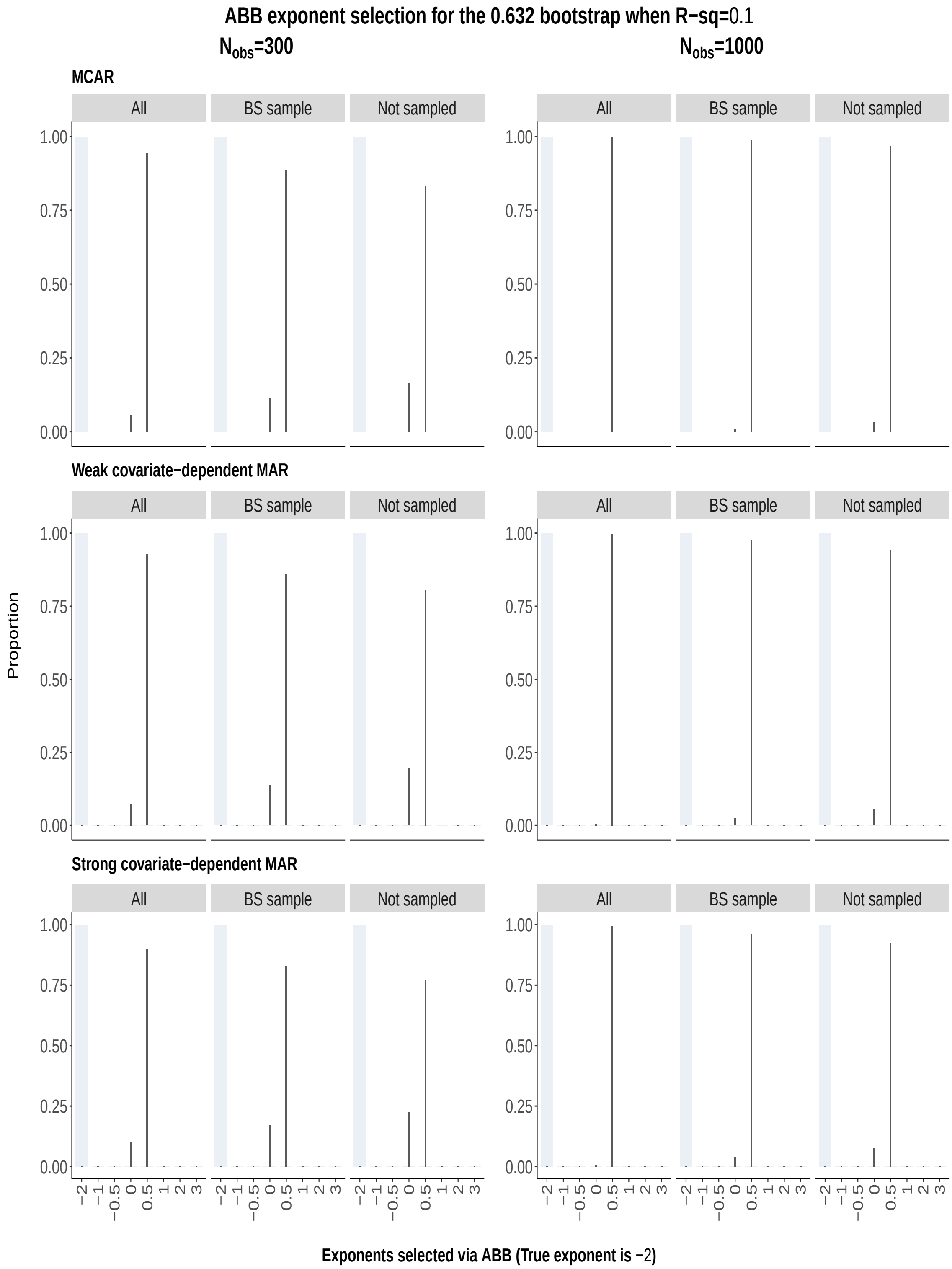

Figure S21: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

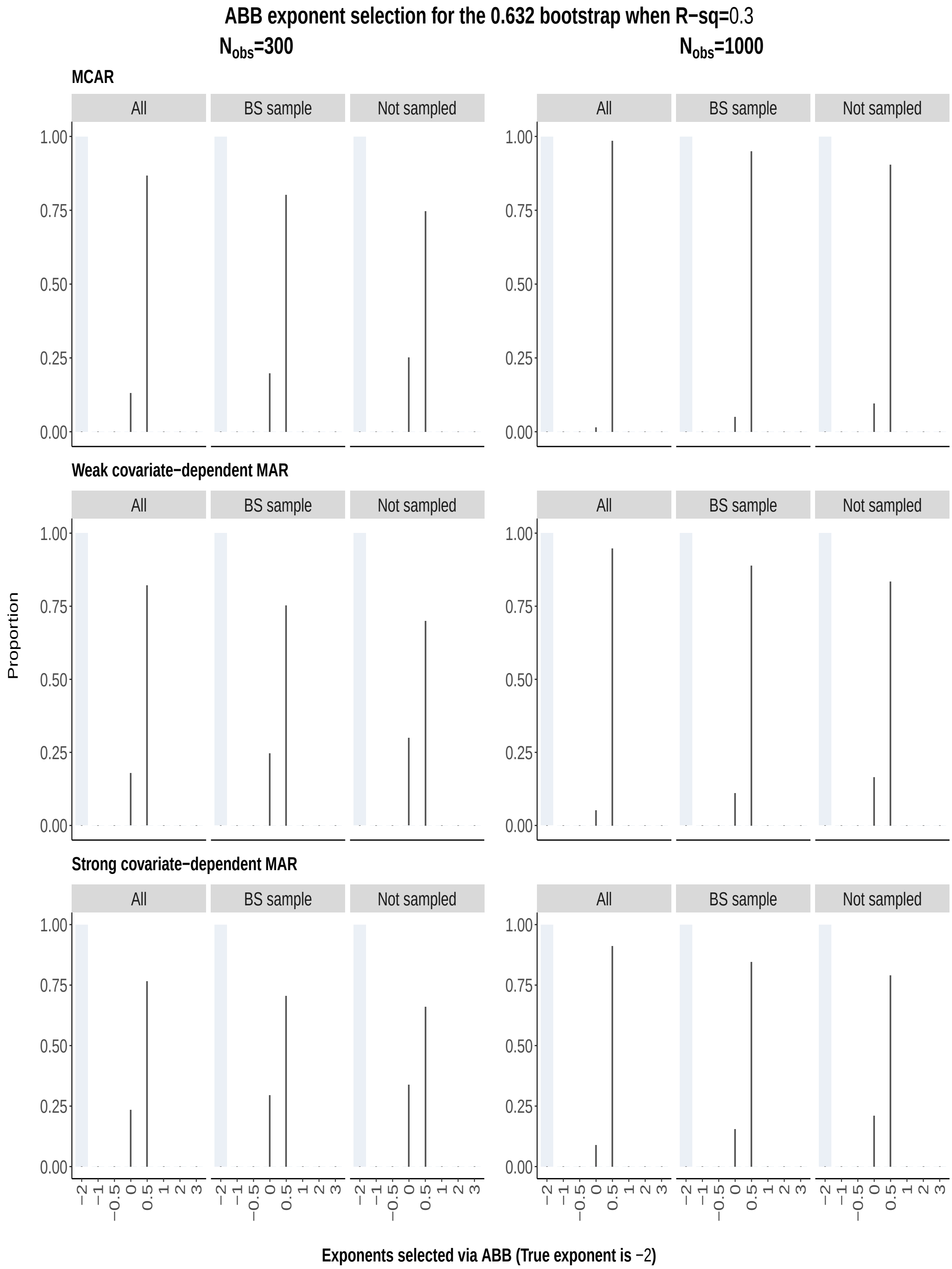

Figure S22: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

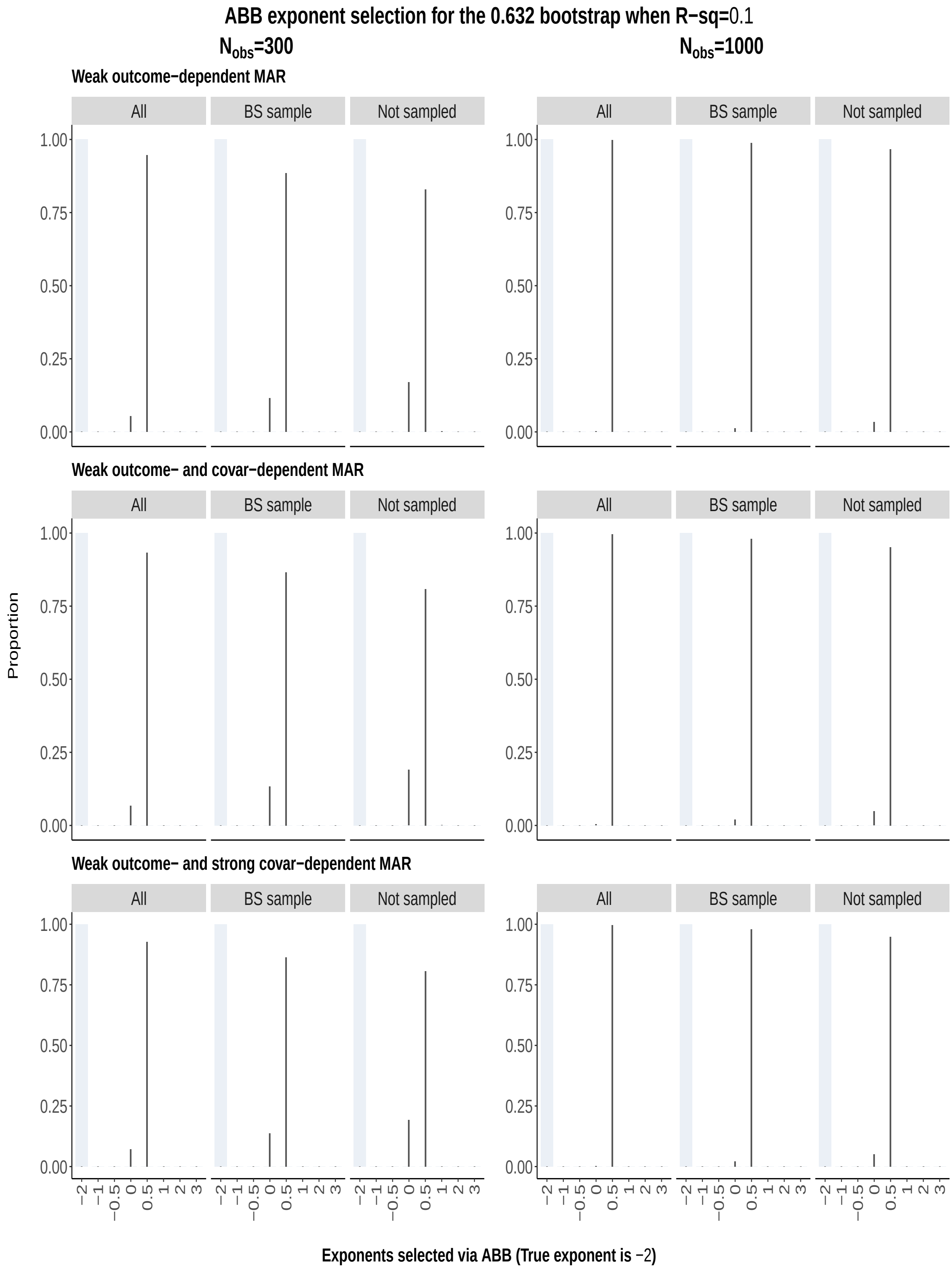

Figure S23: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

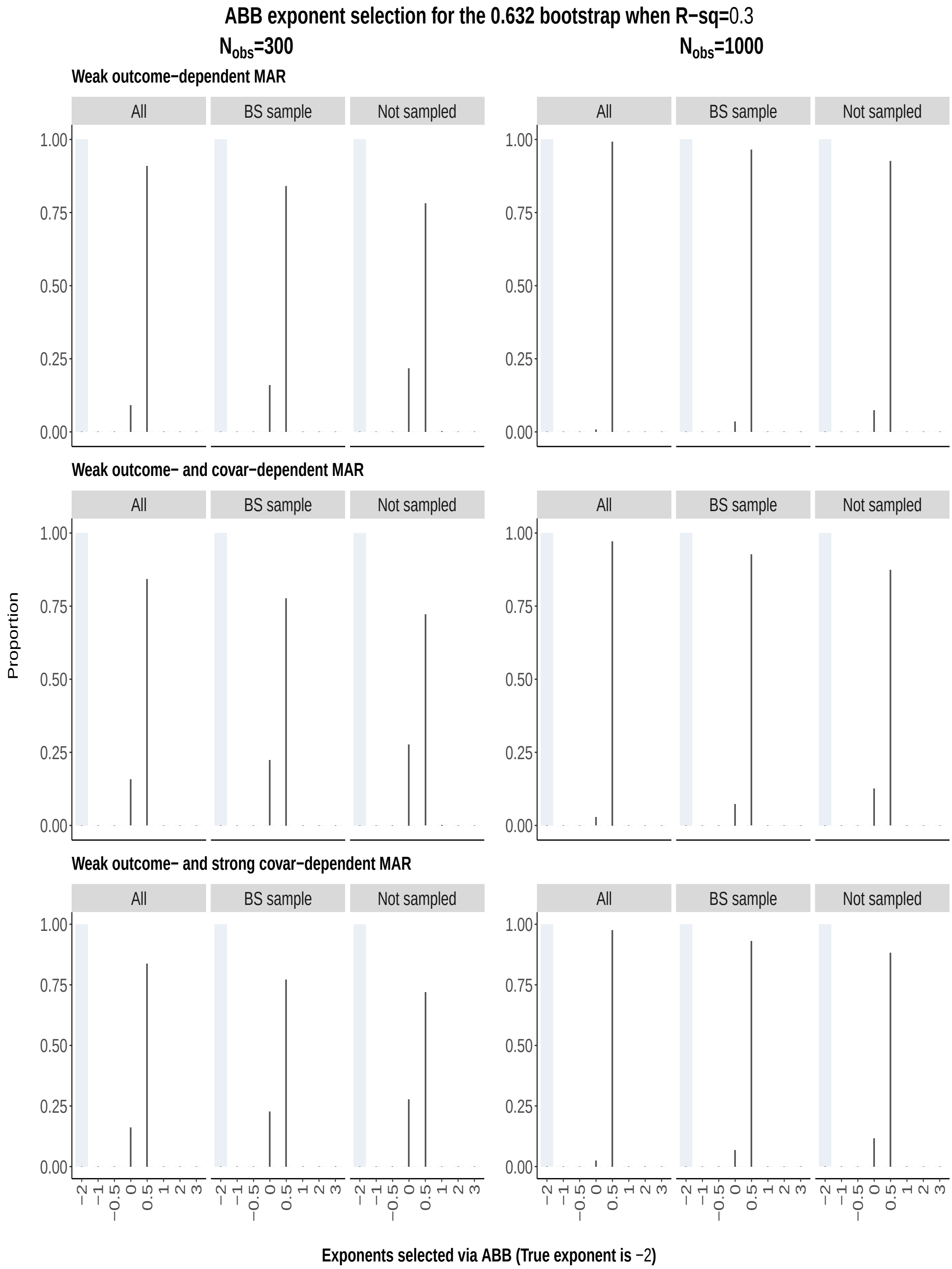

Figure S24: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2 Exponent selection from the FPS procedure

S6.2.1 Cross-validation, $\alpha_E = 1$ and no origin-shift

True exponent is 0

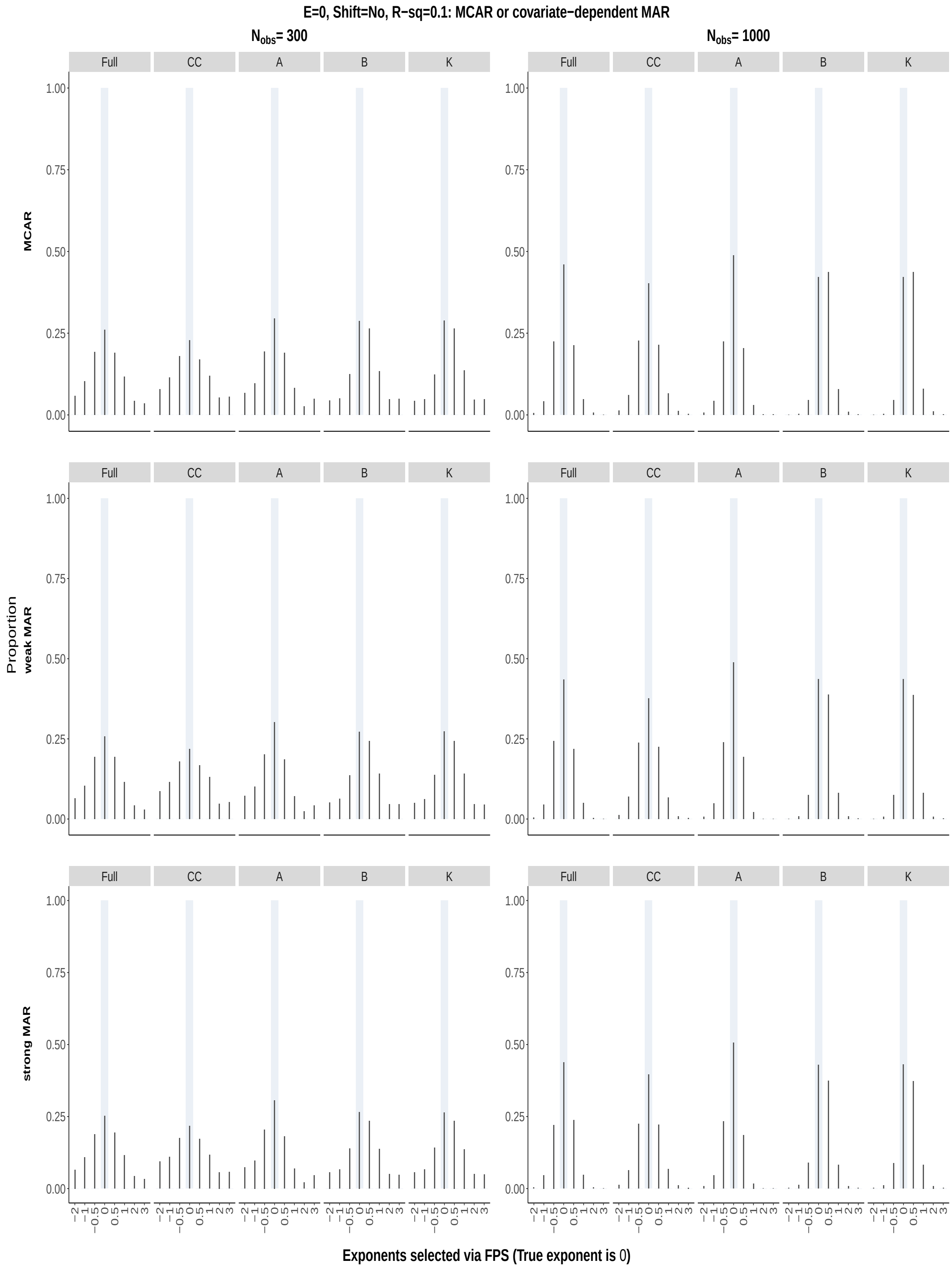

Figure S25: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

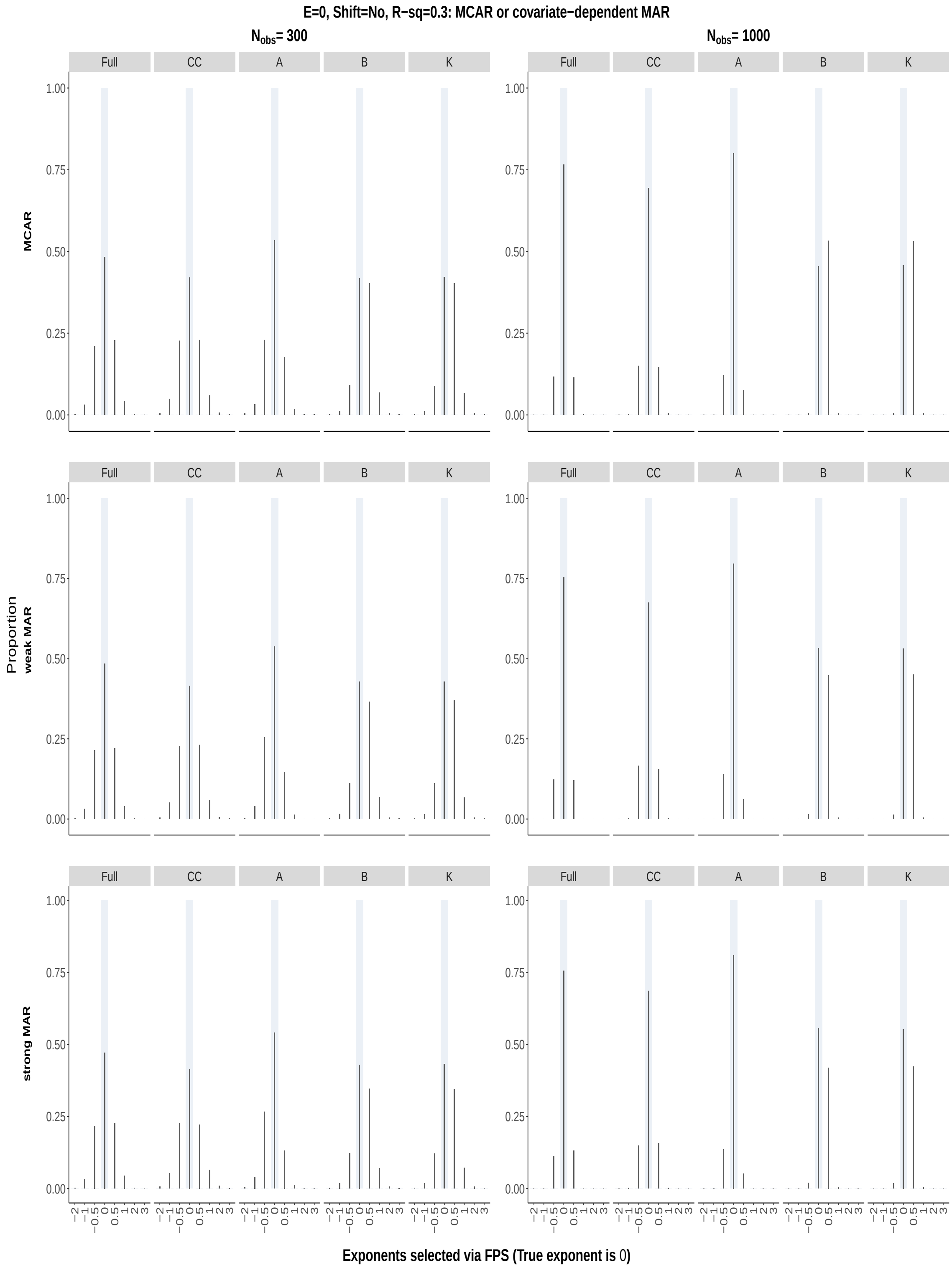

Figure S26: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

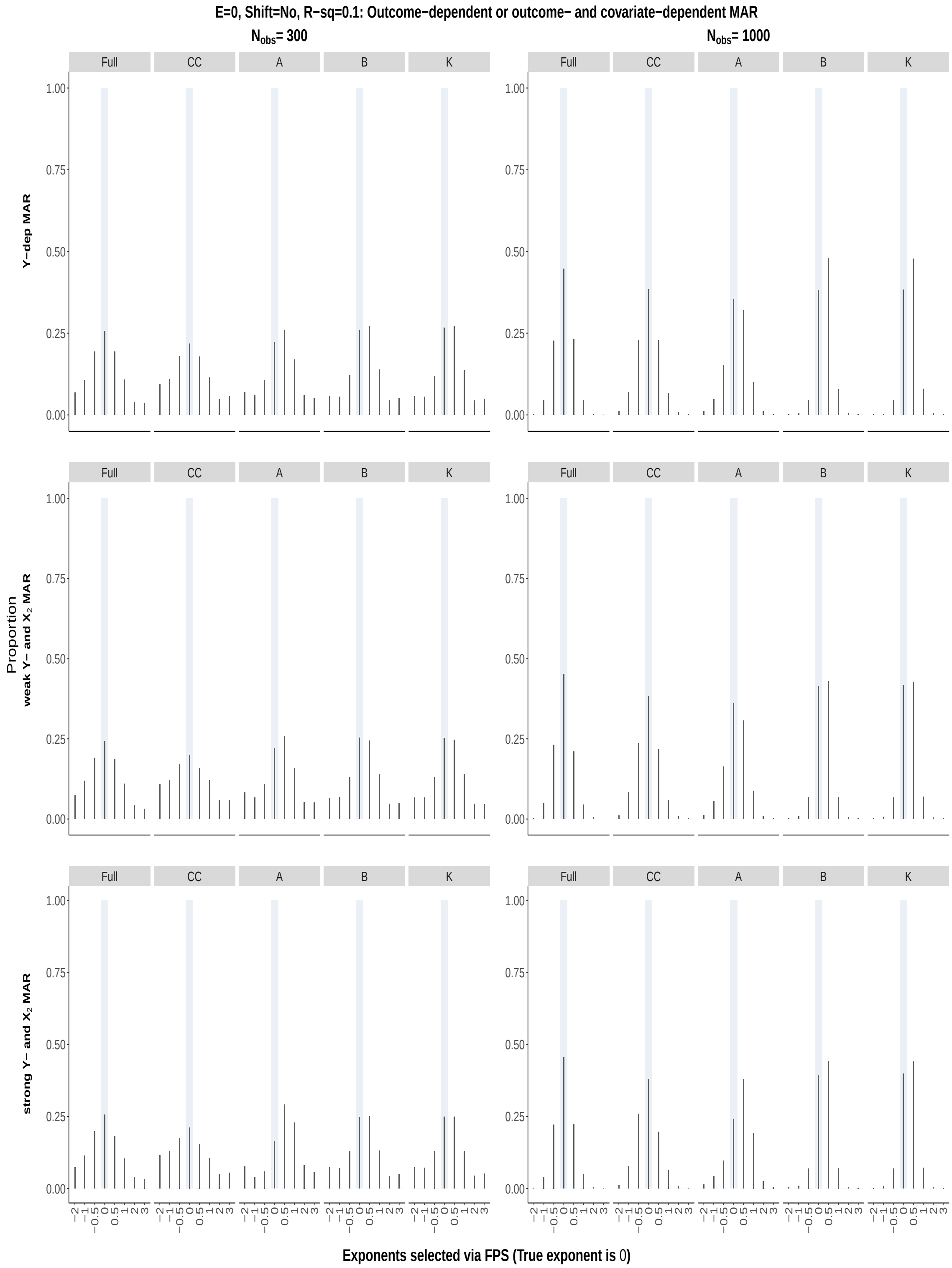

Figure S27: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

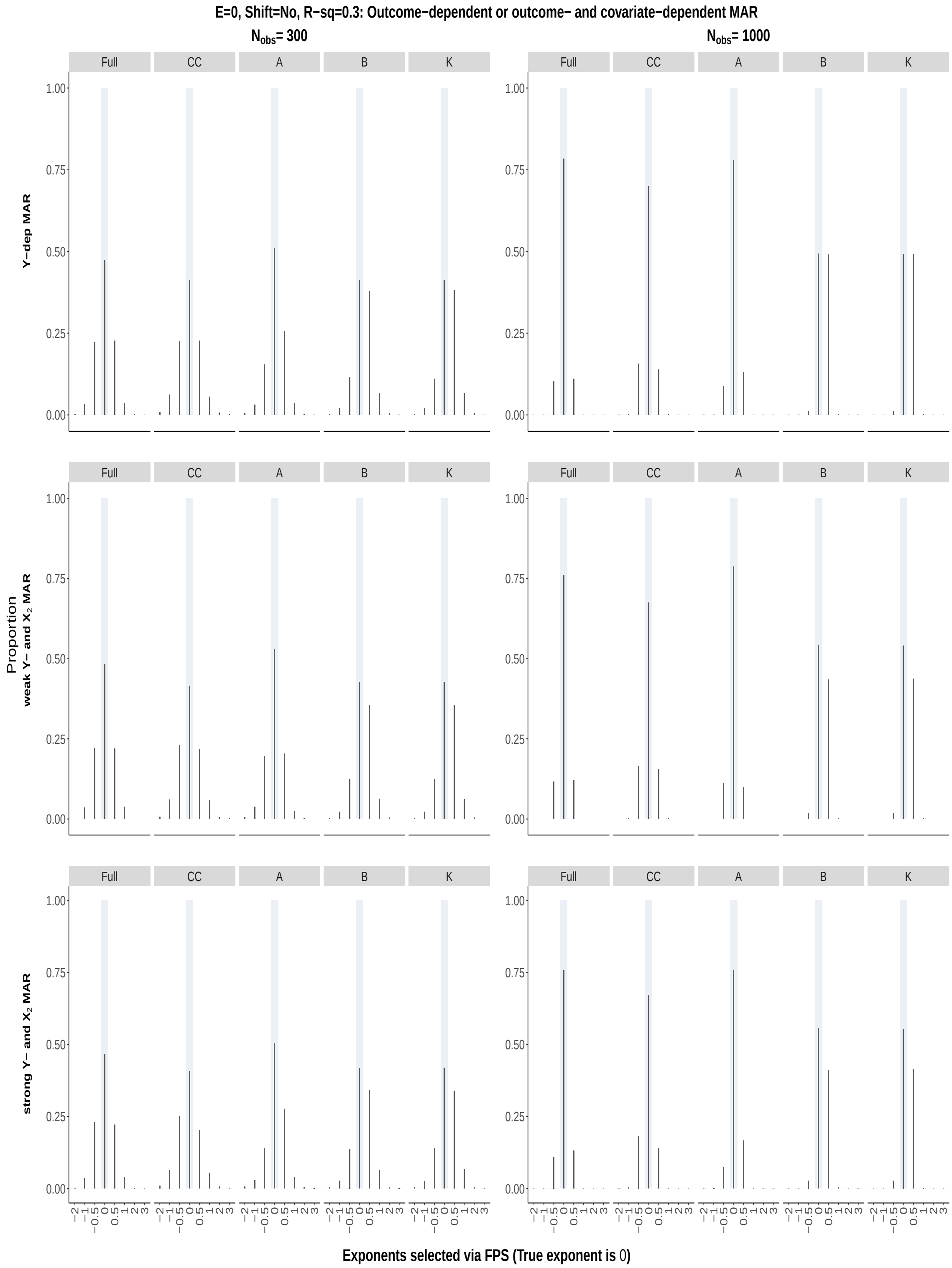

Figure S28: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

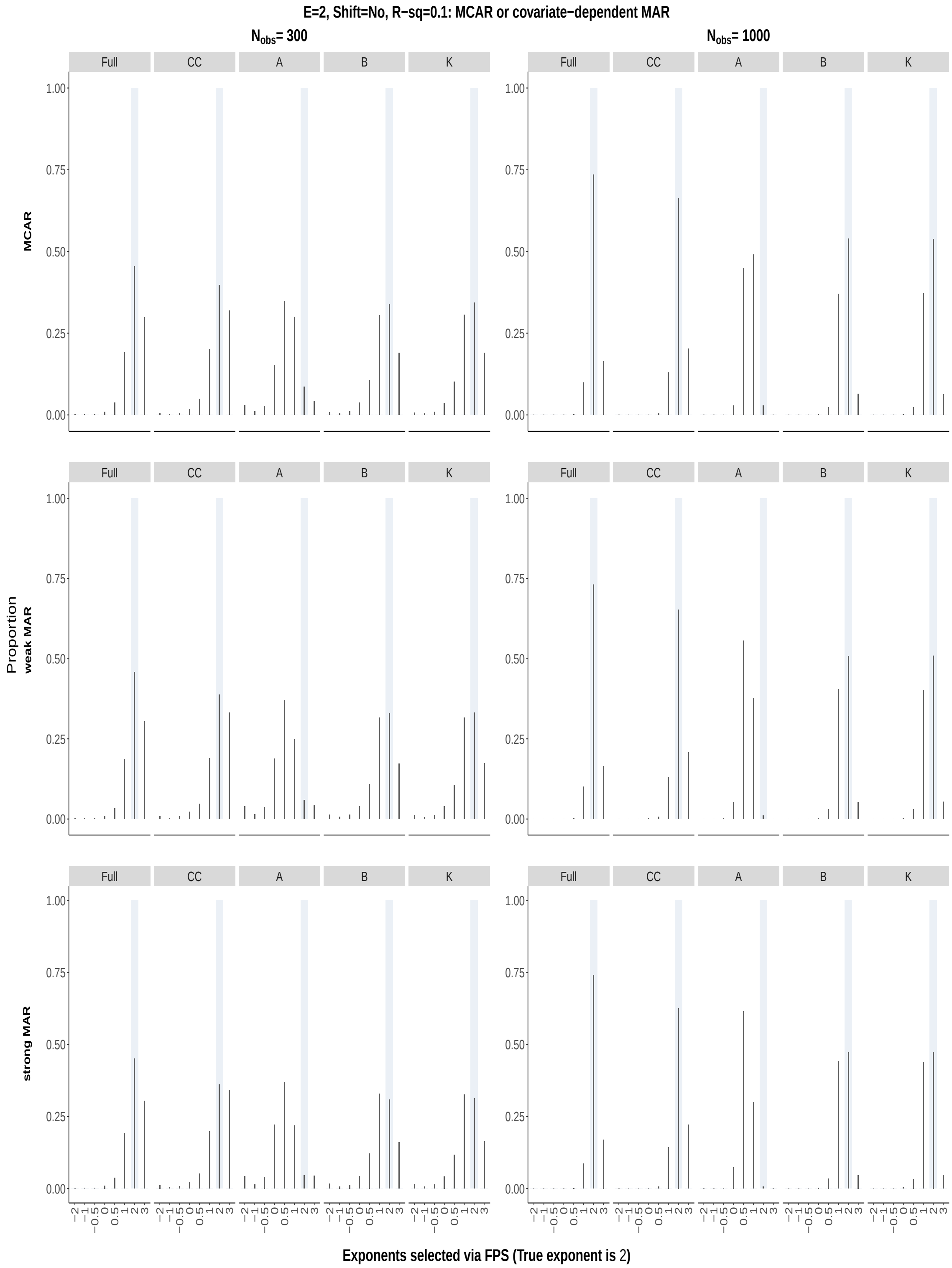

Figure S29: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

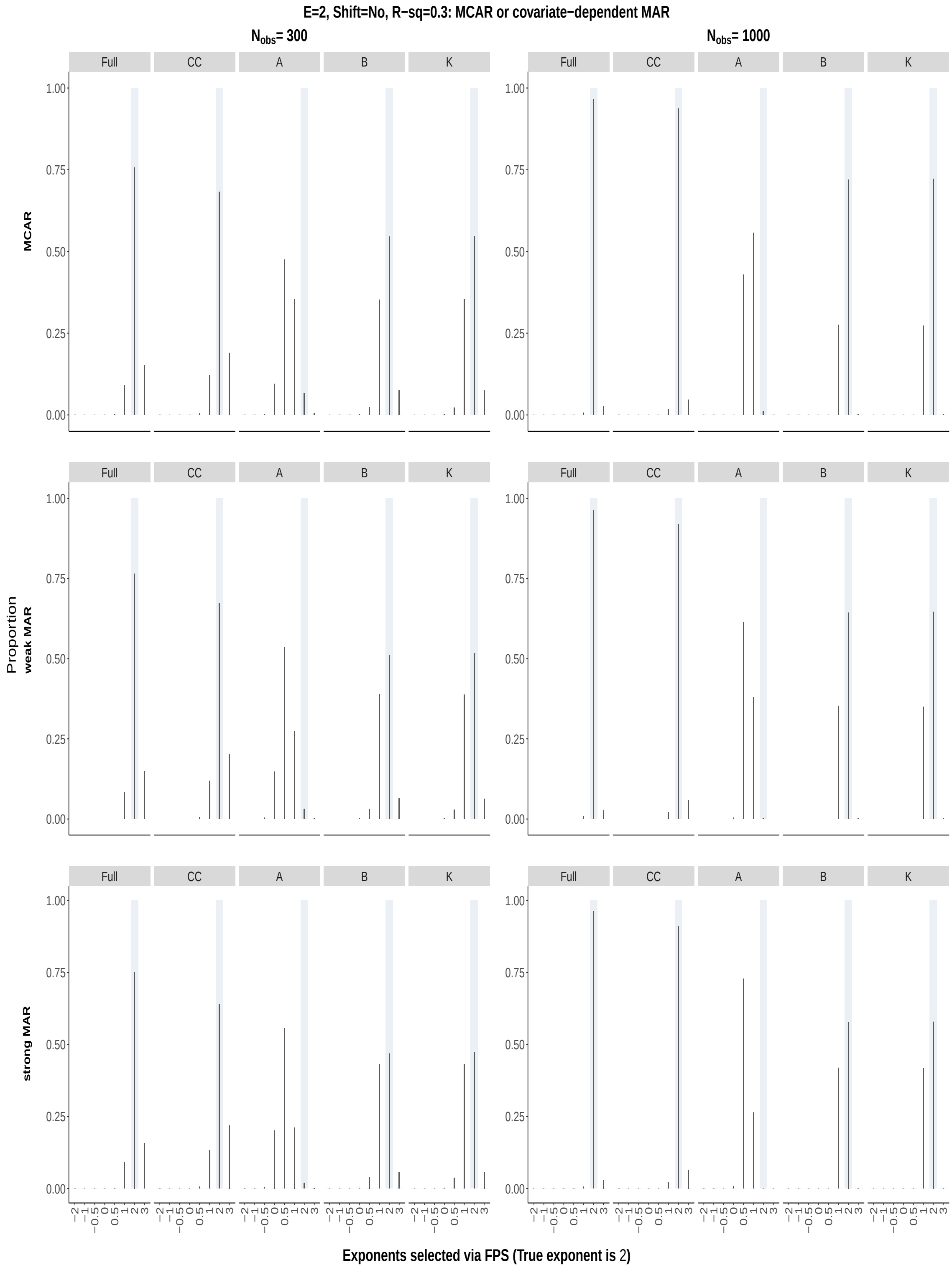

Figure S30: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

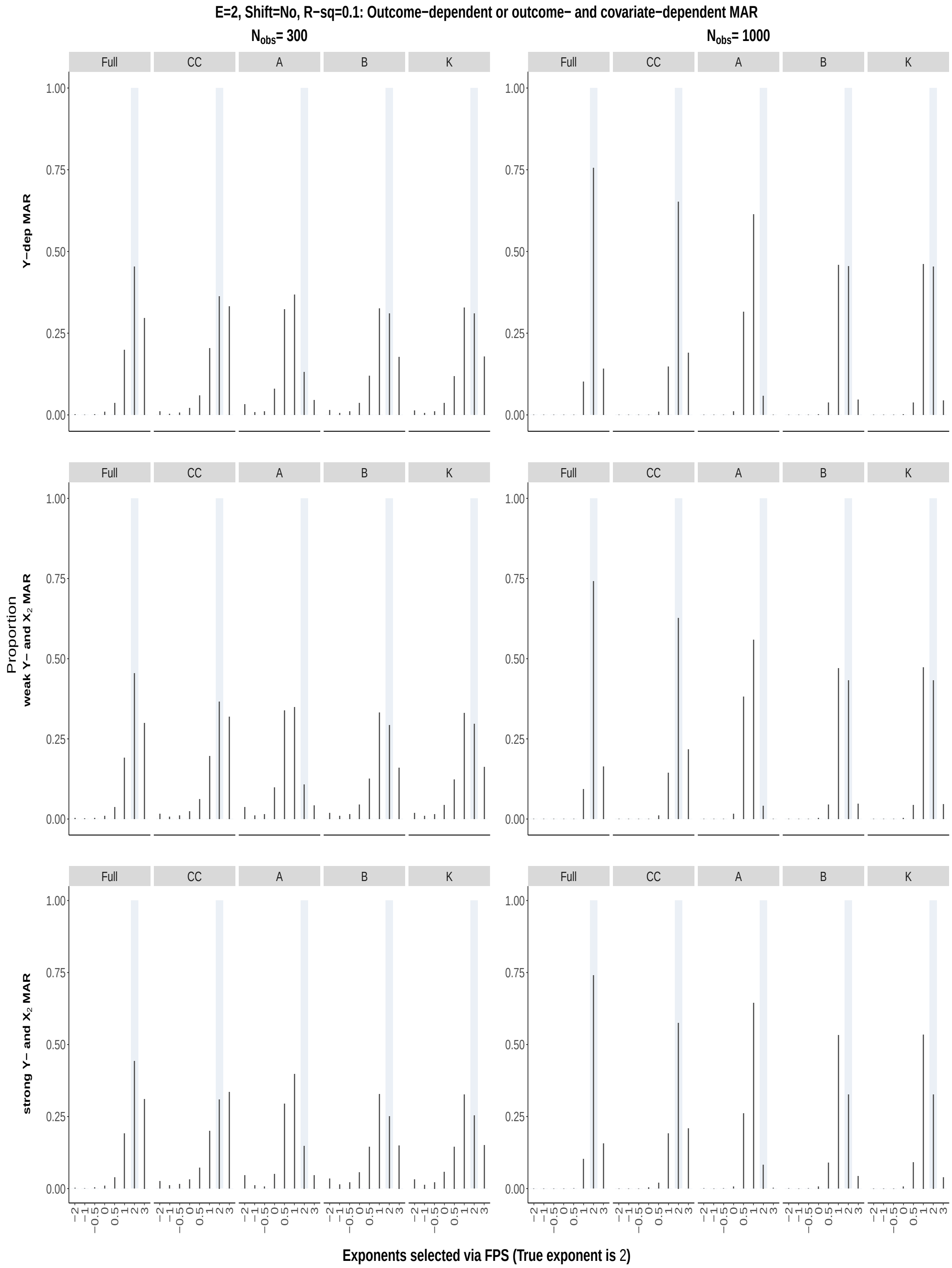

Figure S31: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

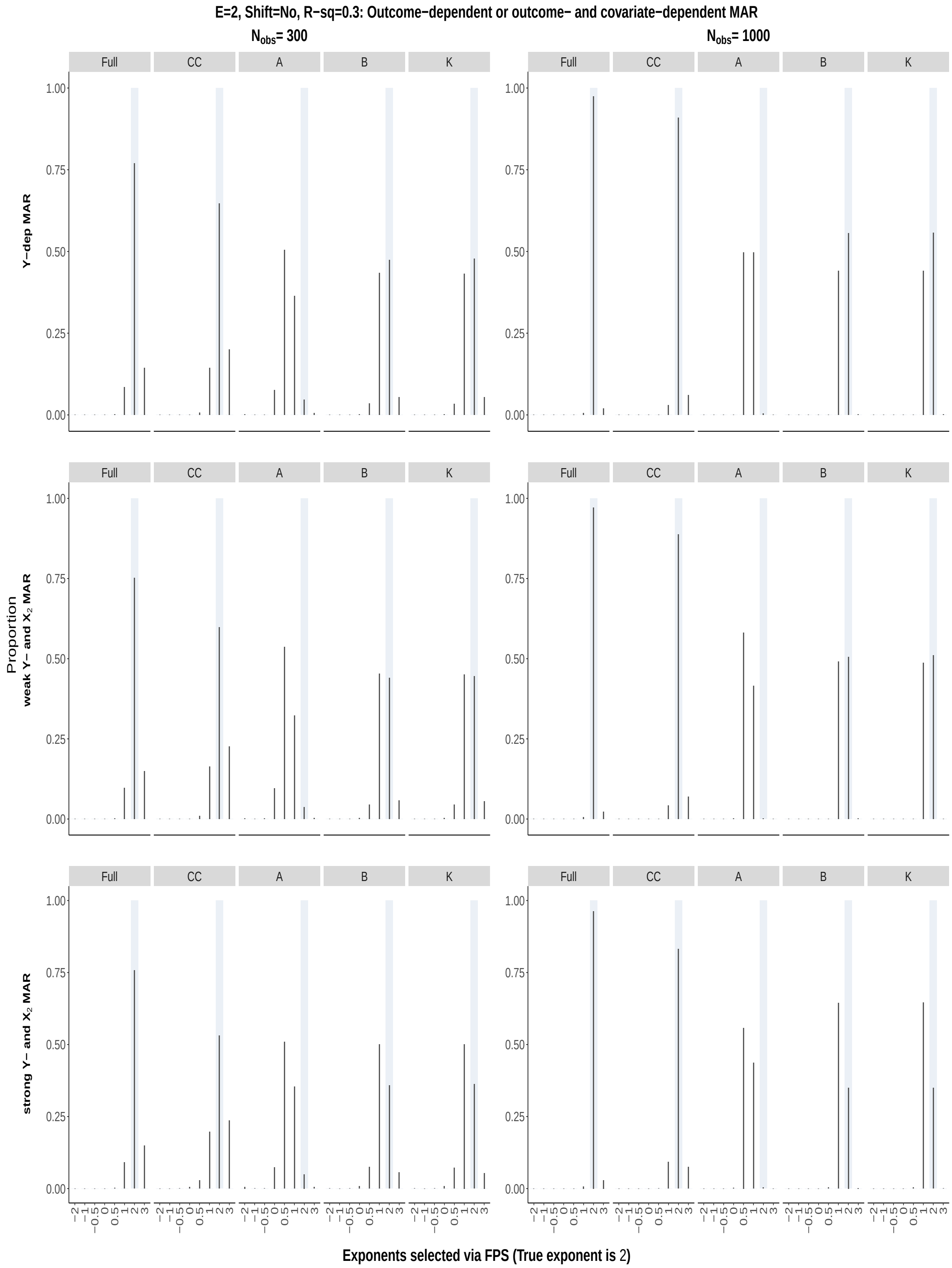

Figure S32: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

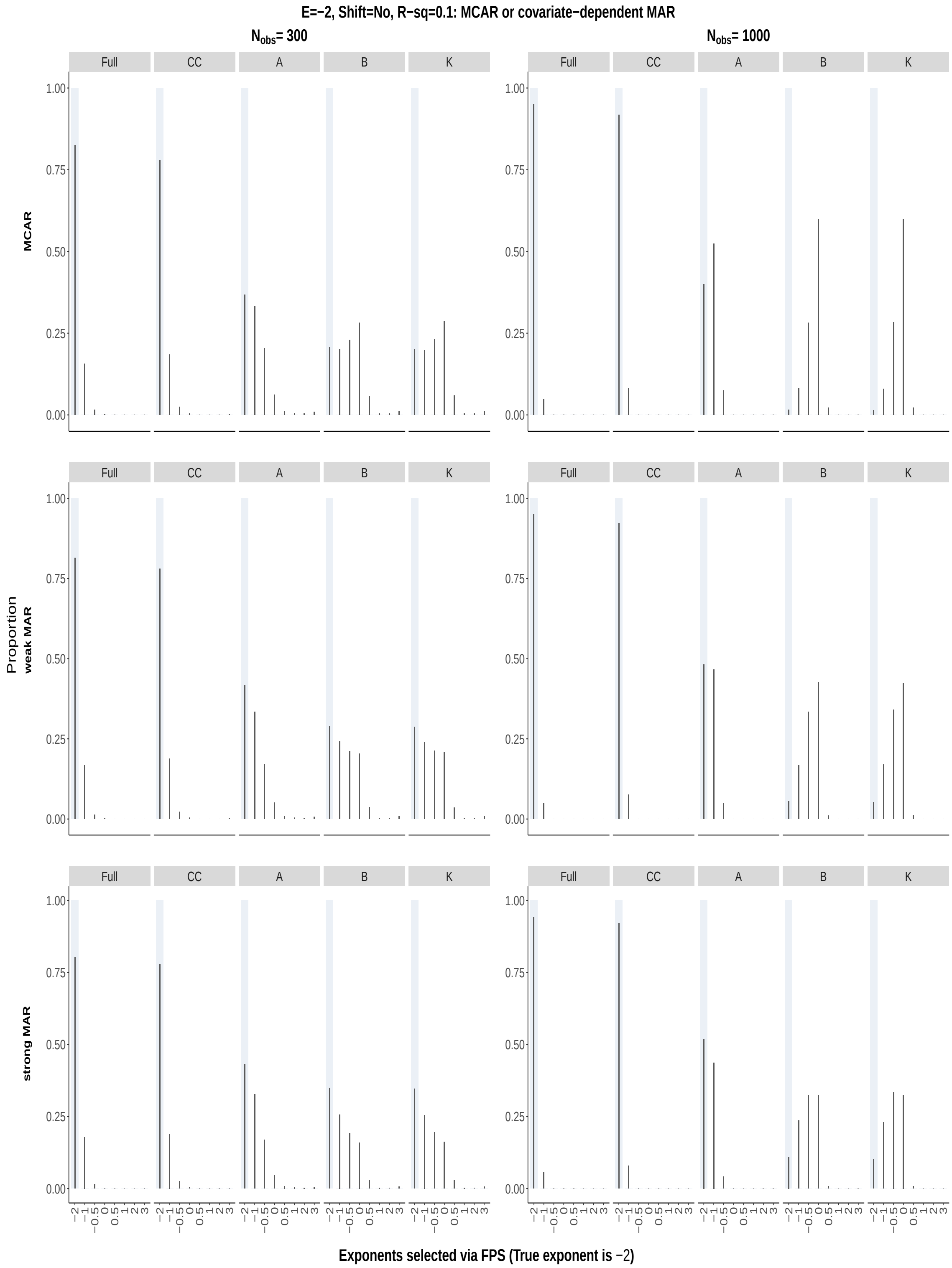

Figure S33: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

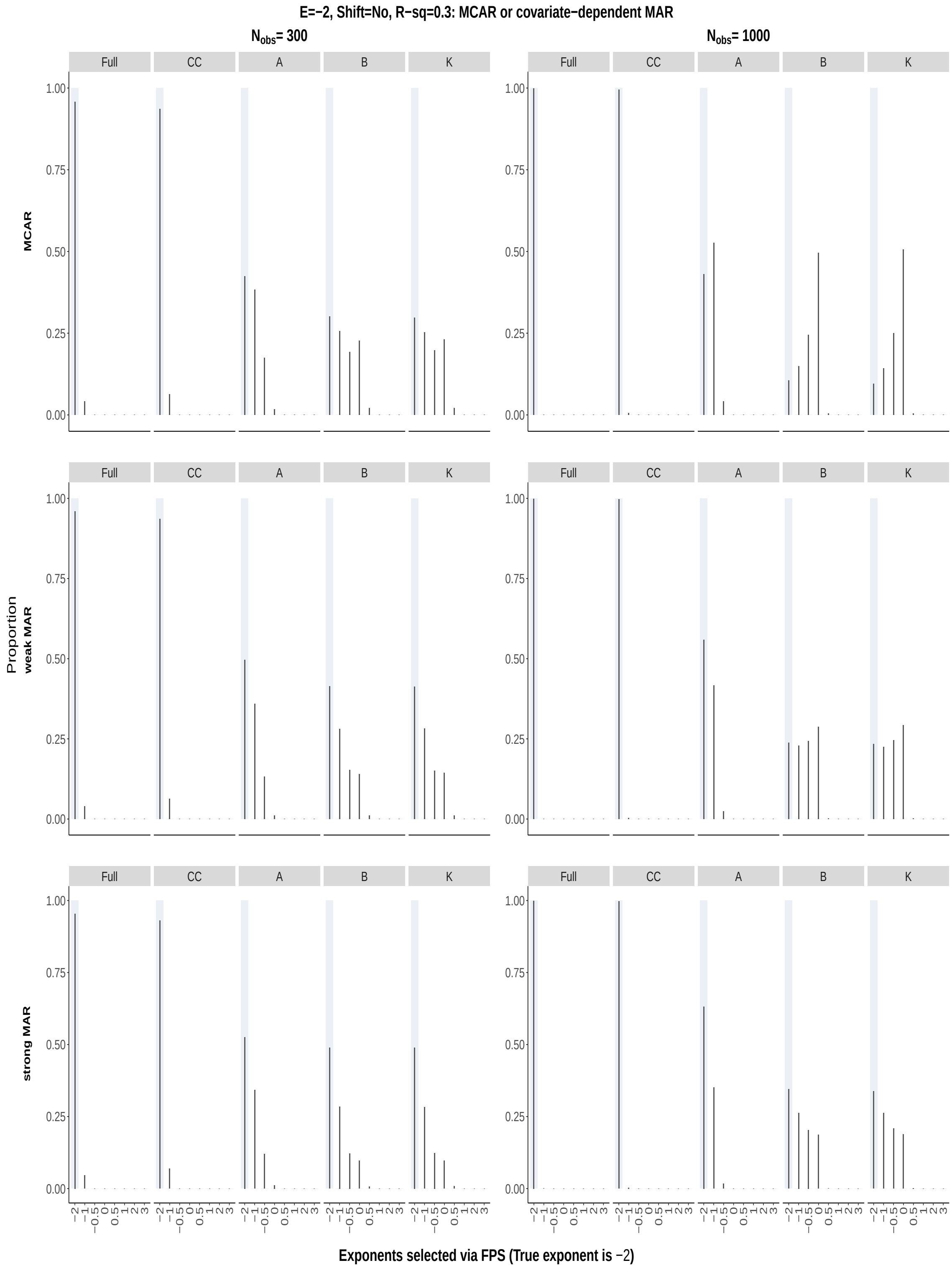

Figure S34: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

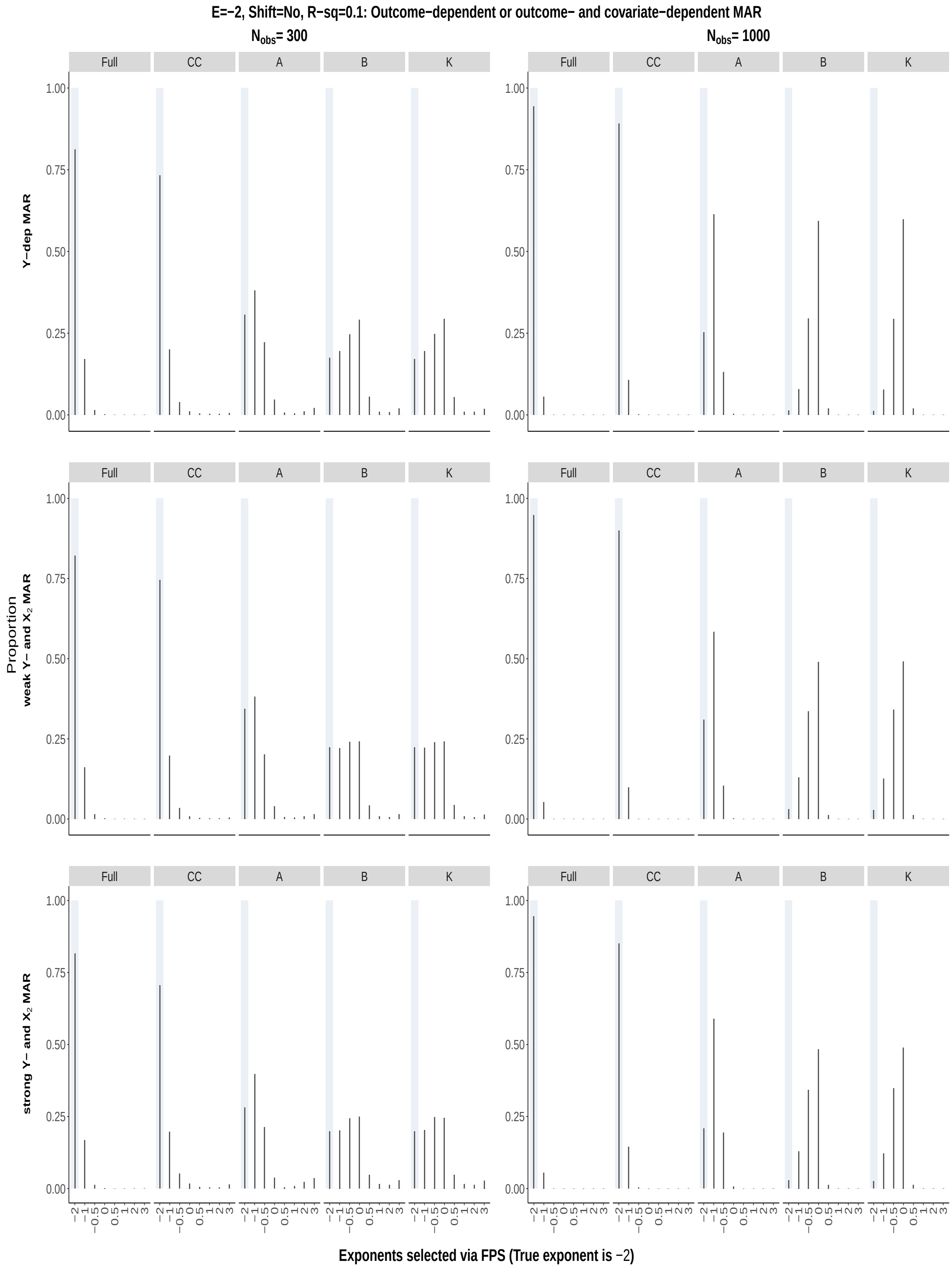

Figure S35: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

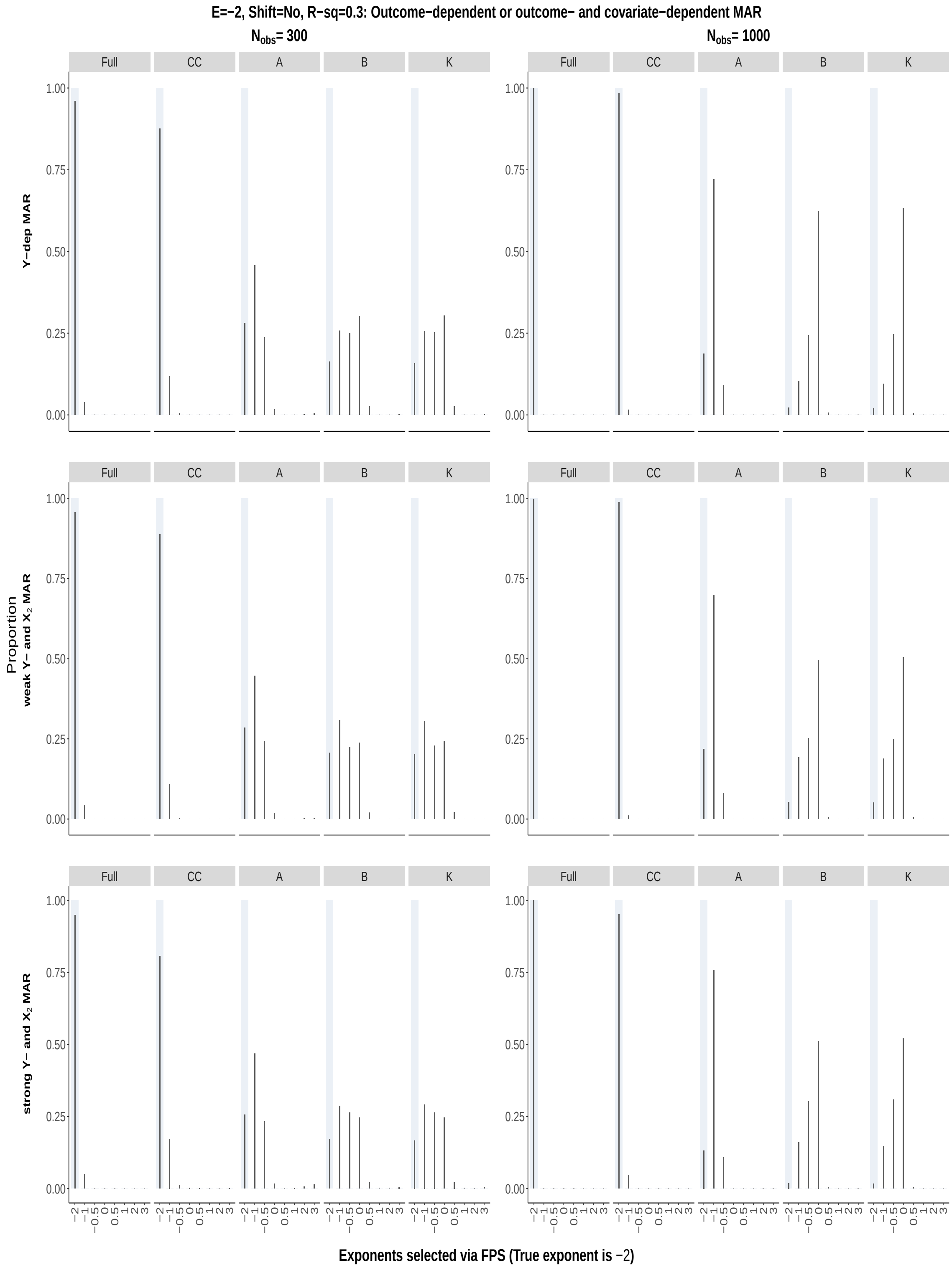

Figure S36: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.2 Cross-validation, $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

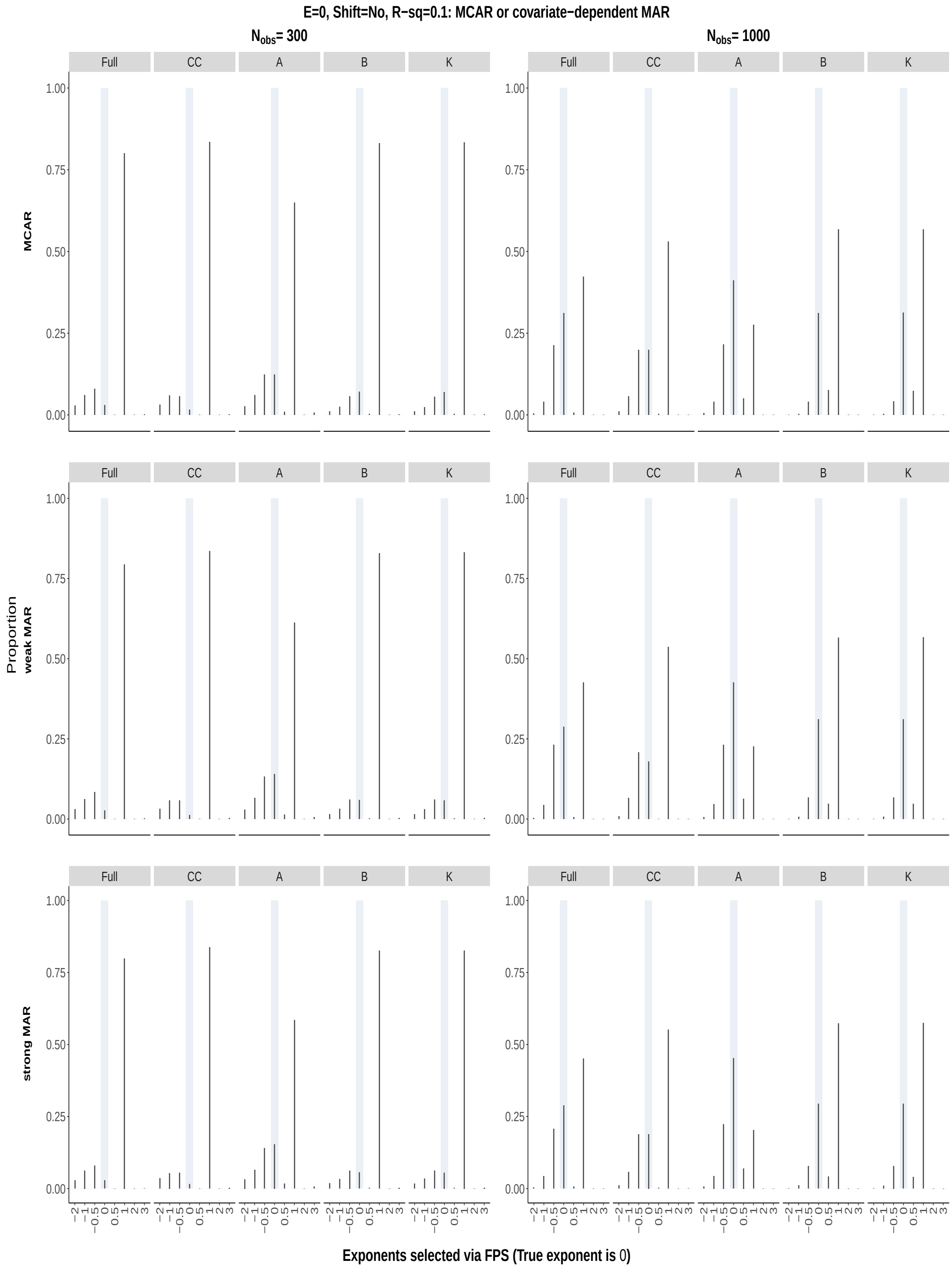

Figure S37: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

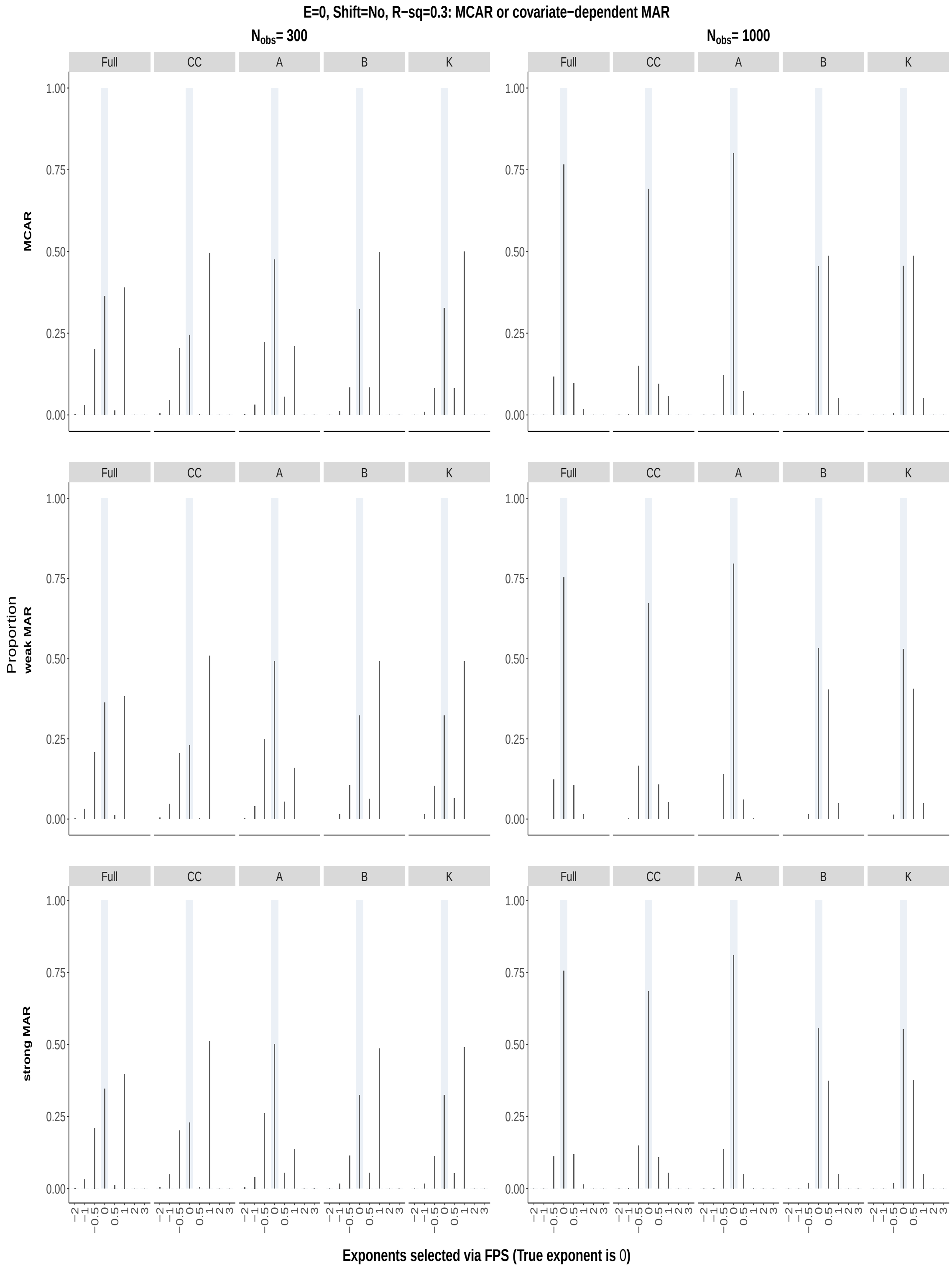

Figure S38: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

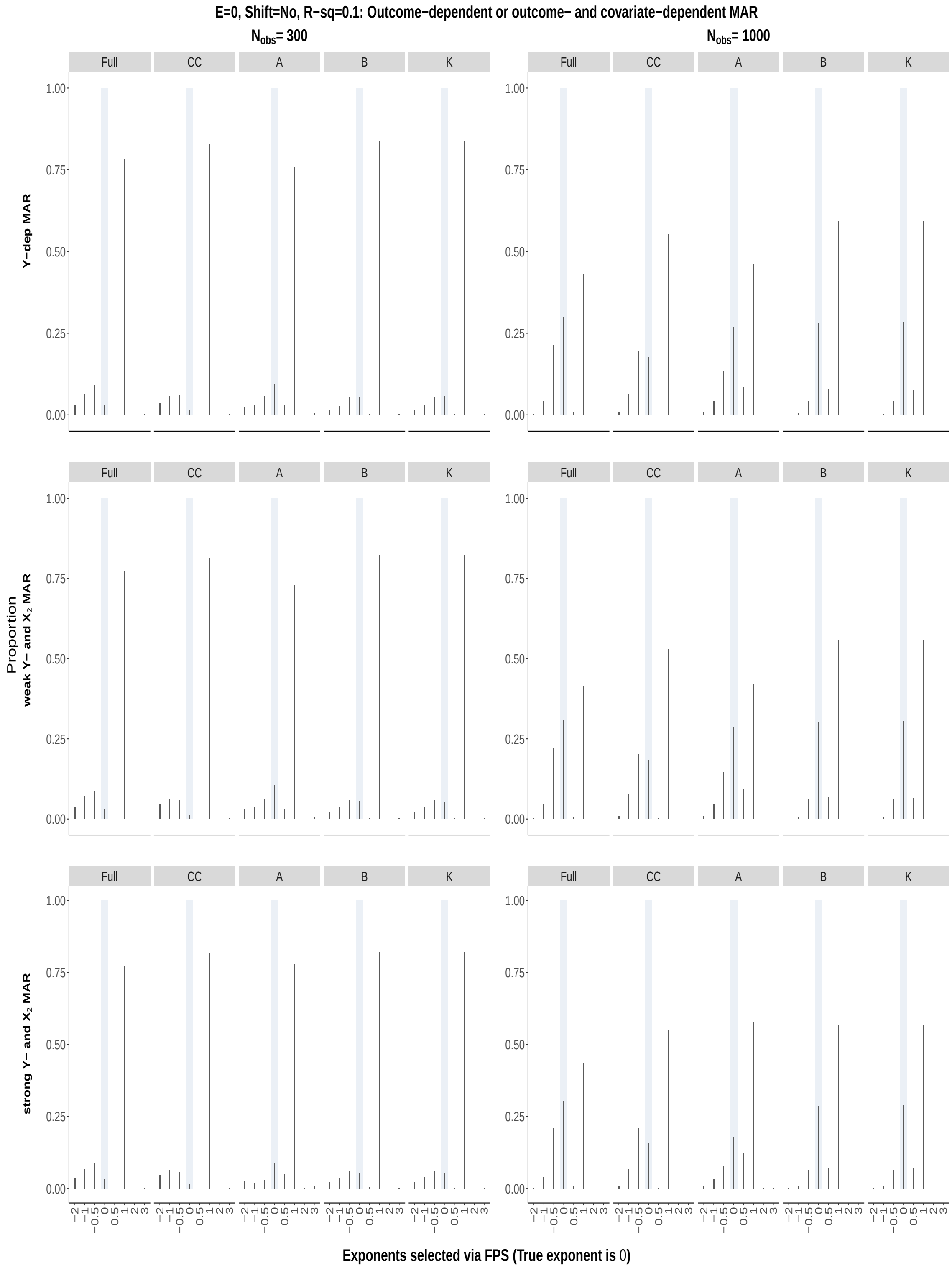

Figure S39: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

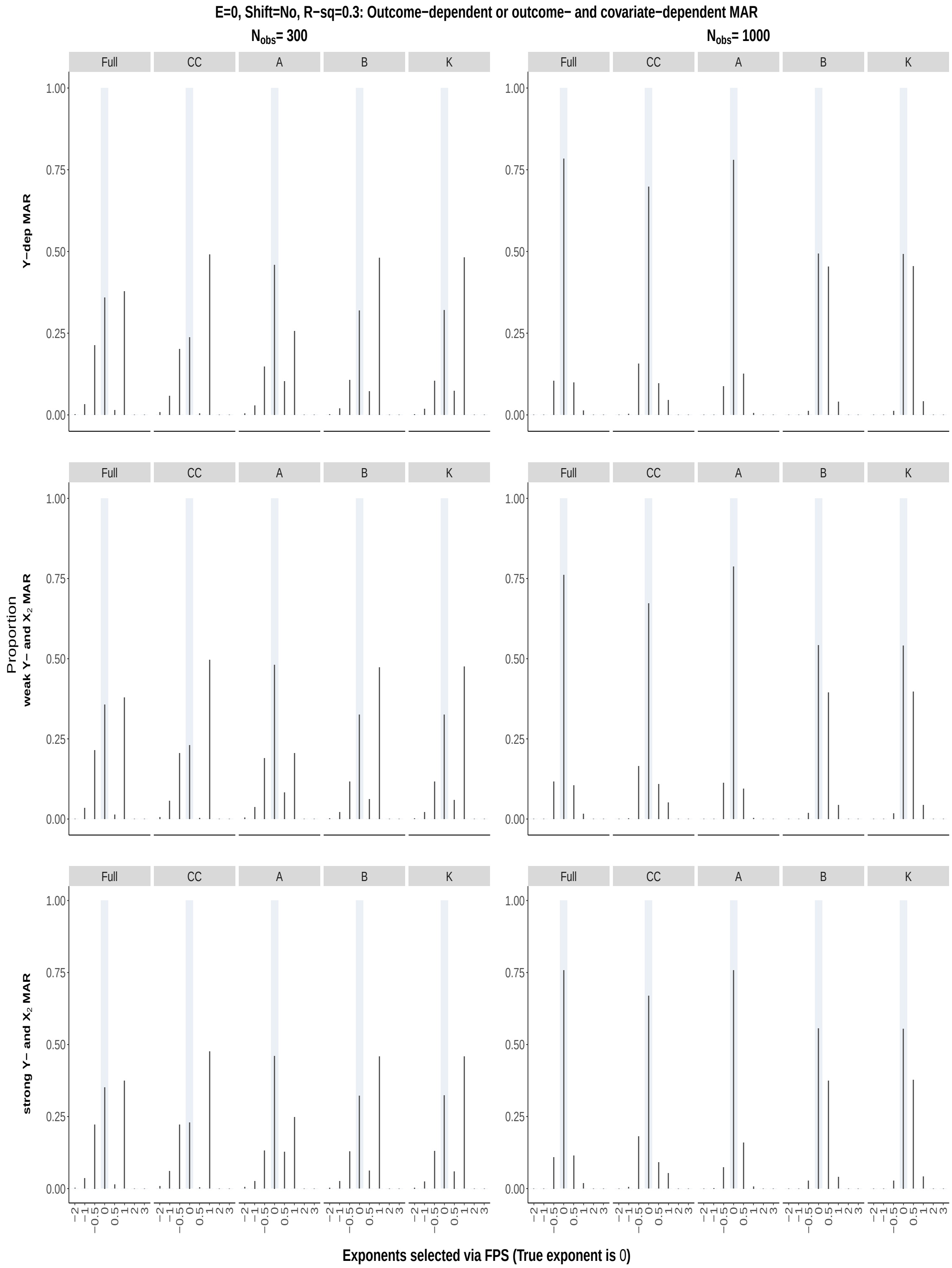

Figure S40: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

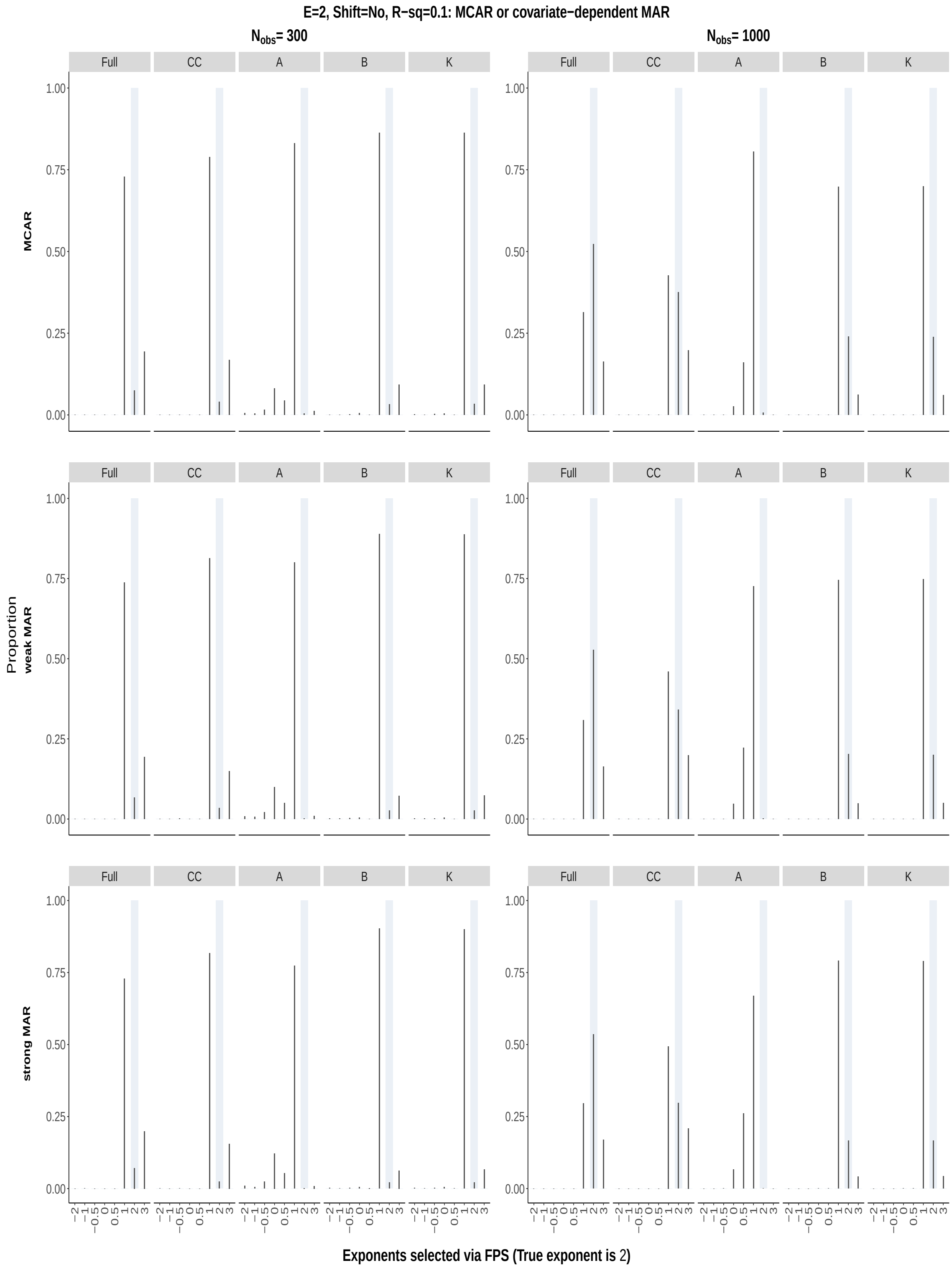

Figure S41: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

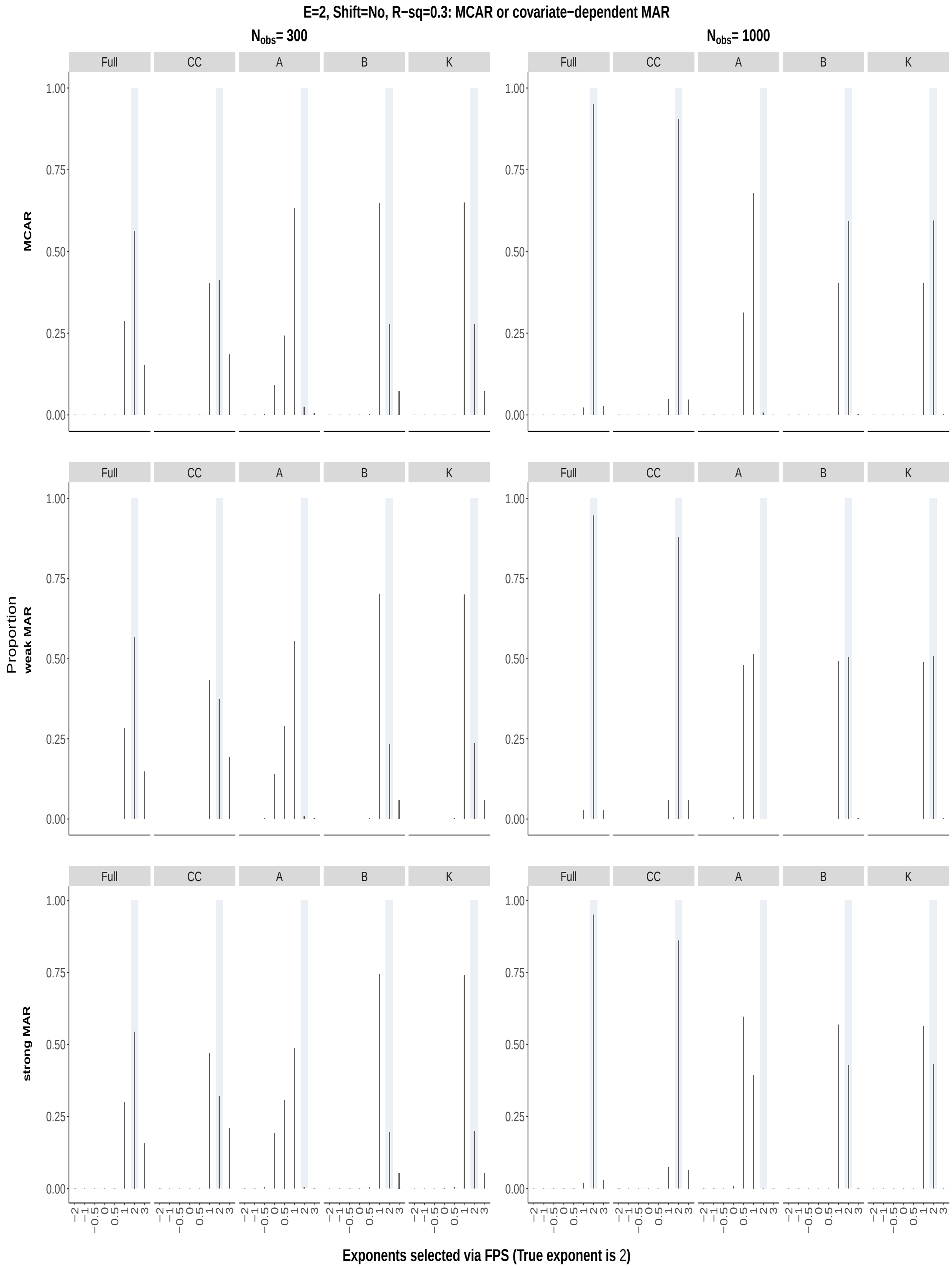

Figure S42: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

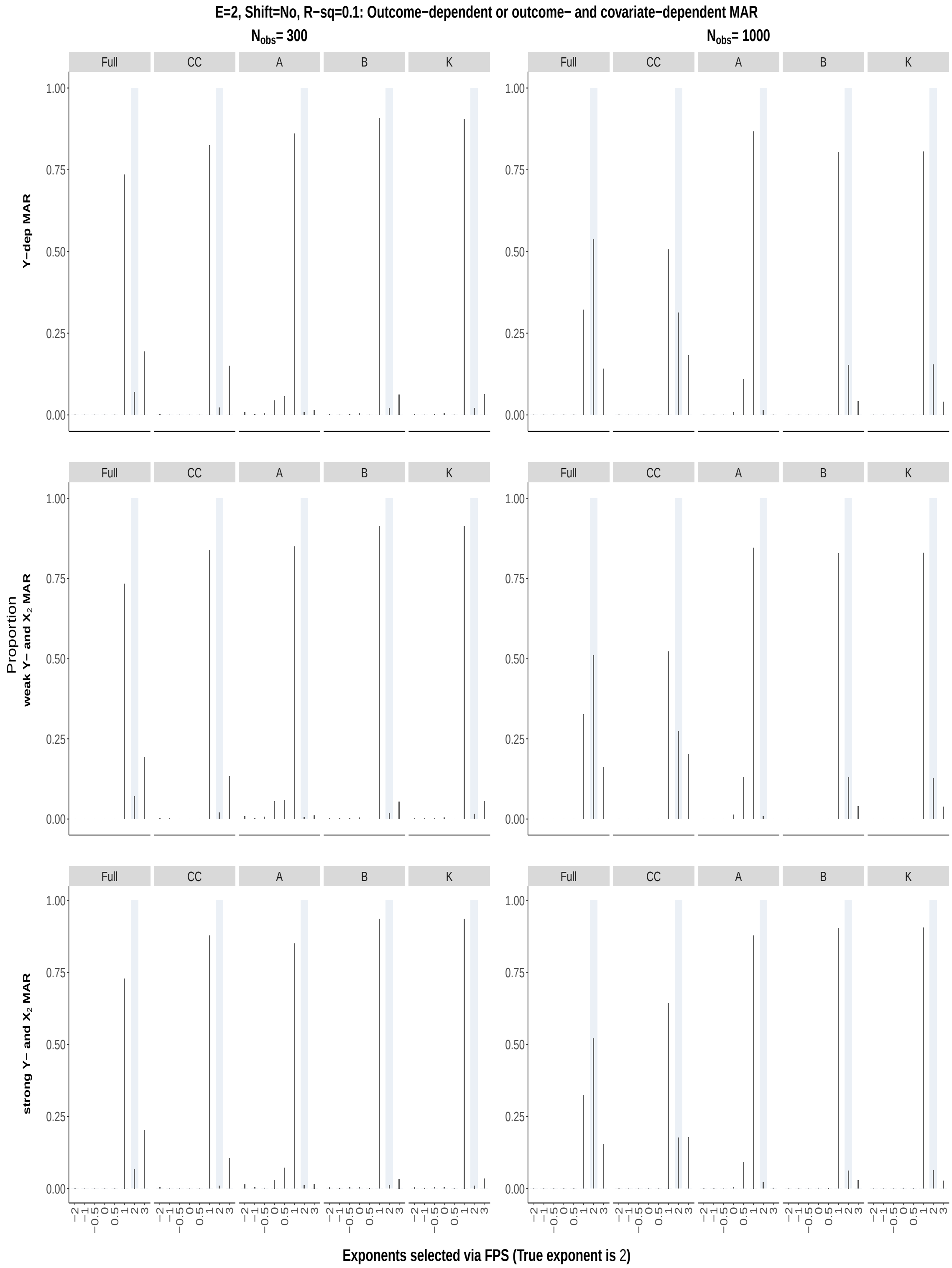

Figure S43: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

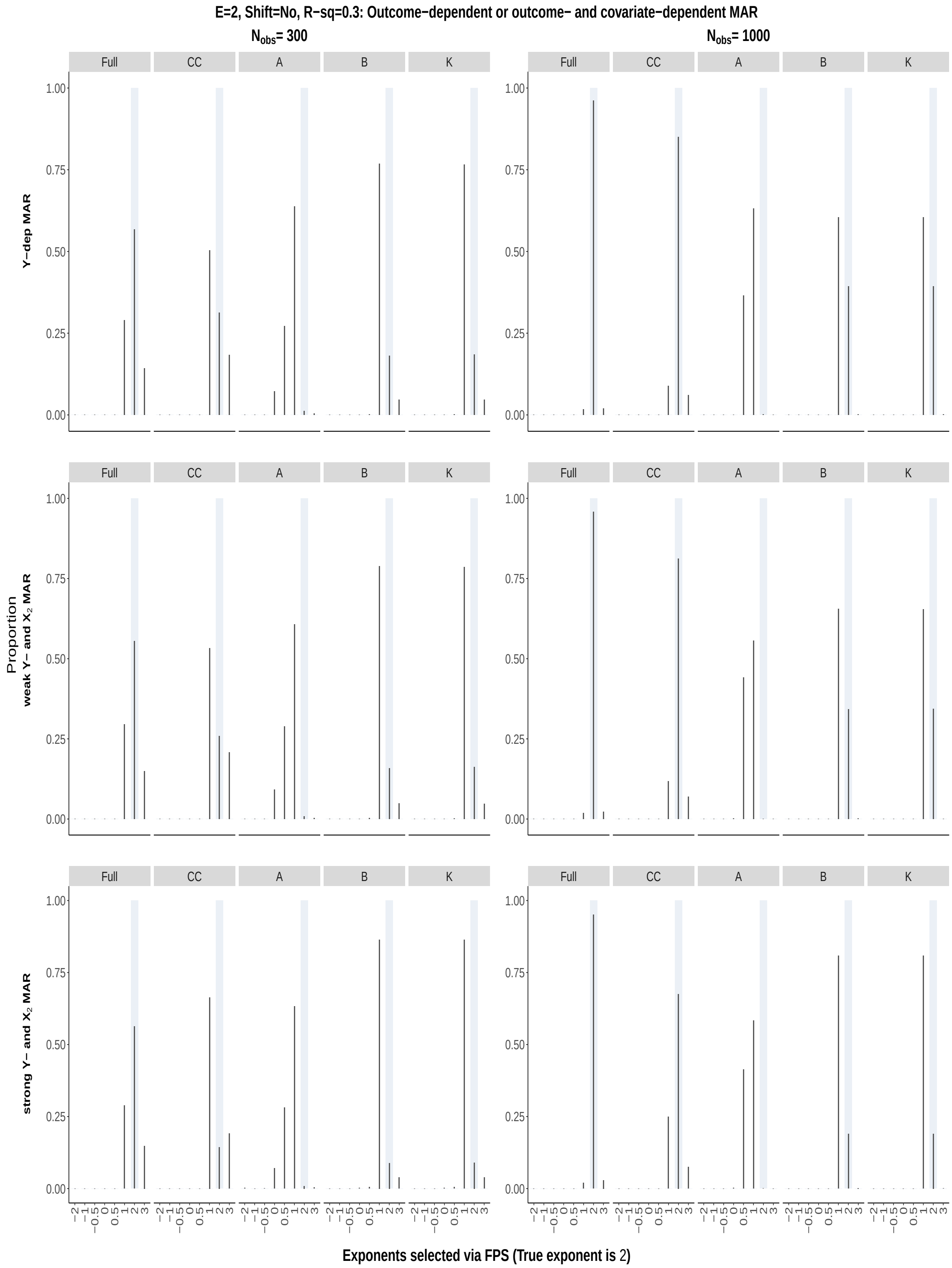

Figure S44: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

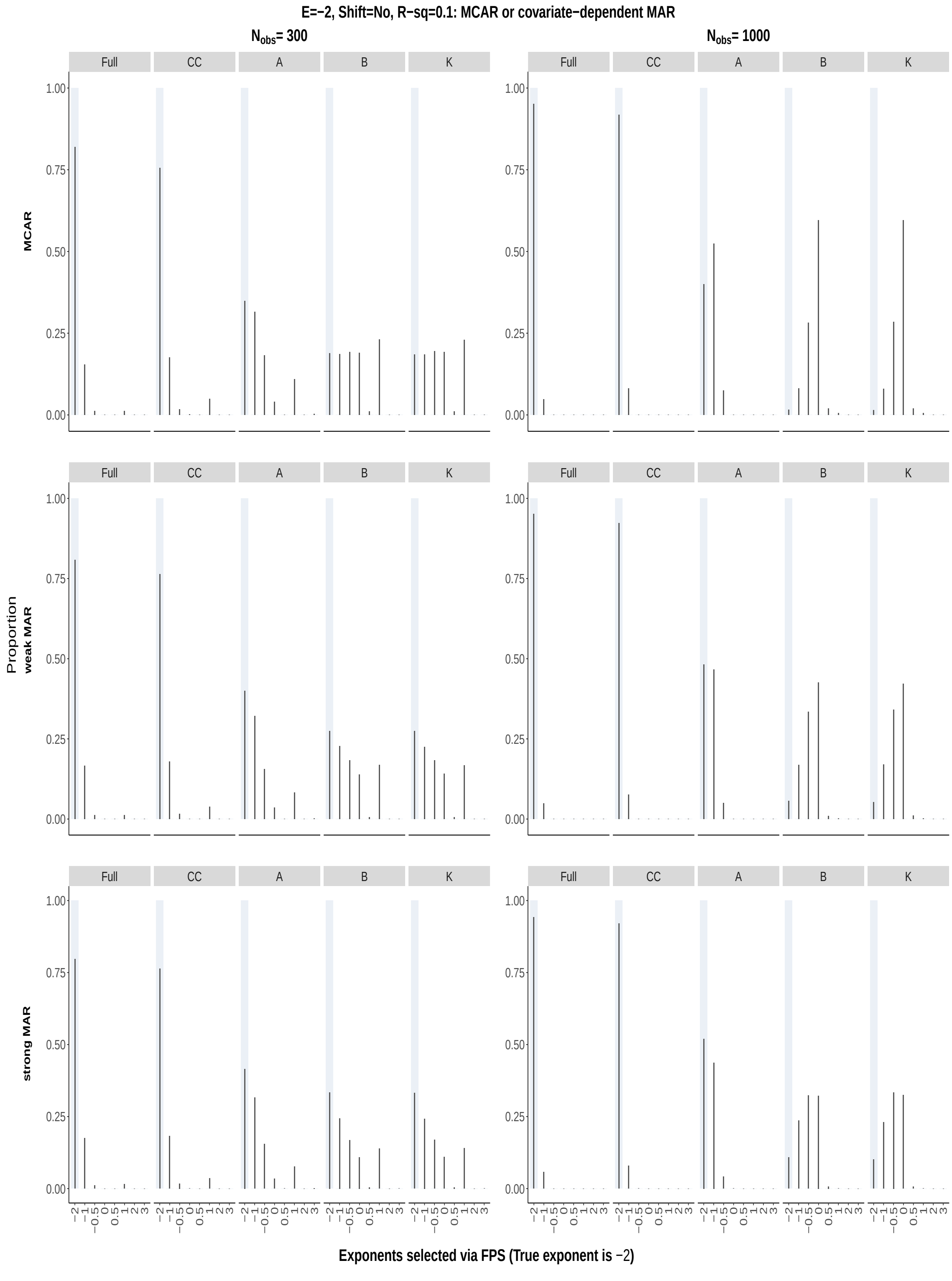

Figure S45: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

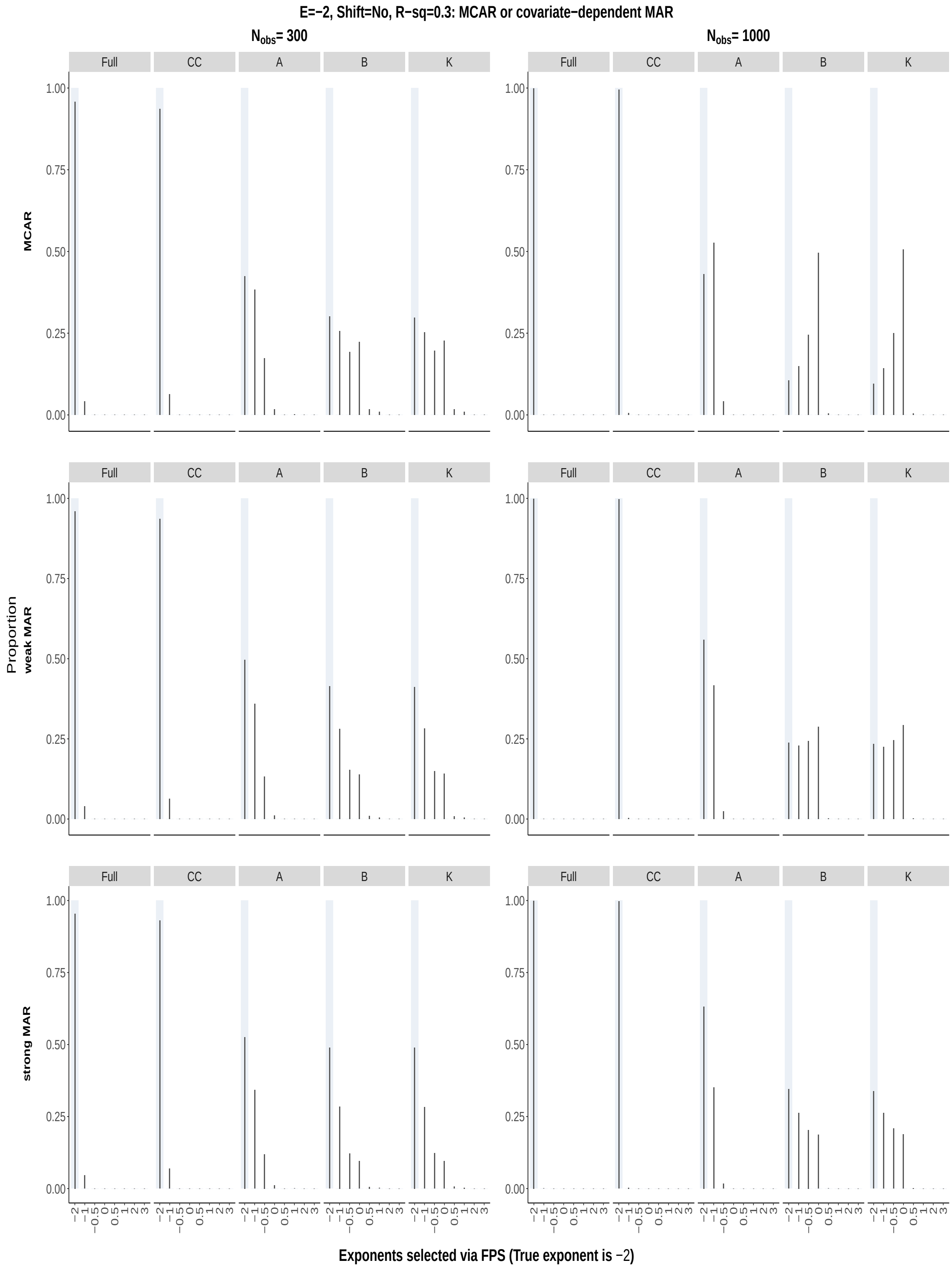

Figure S46: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

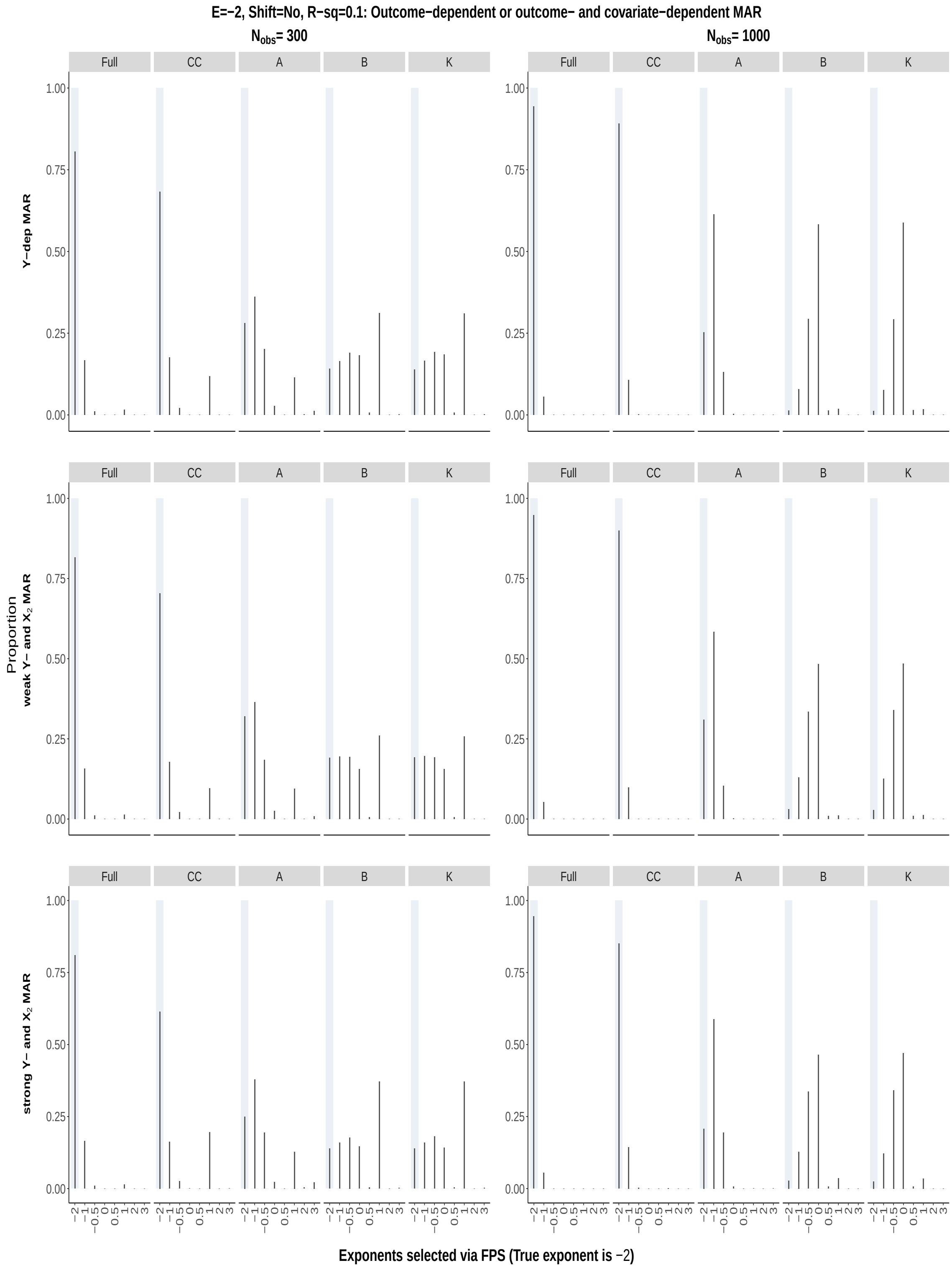

Figure S47: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

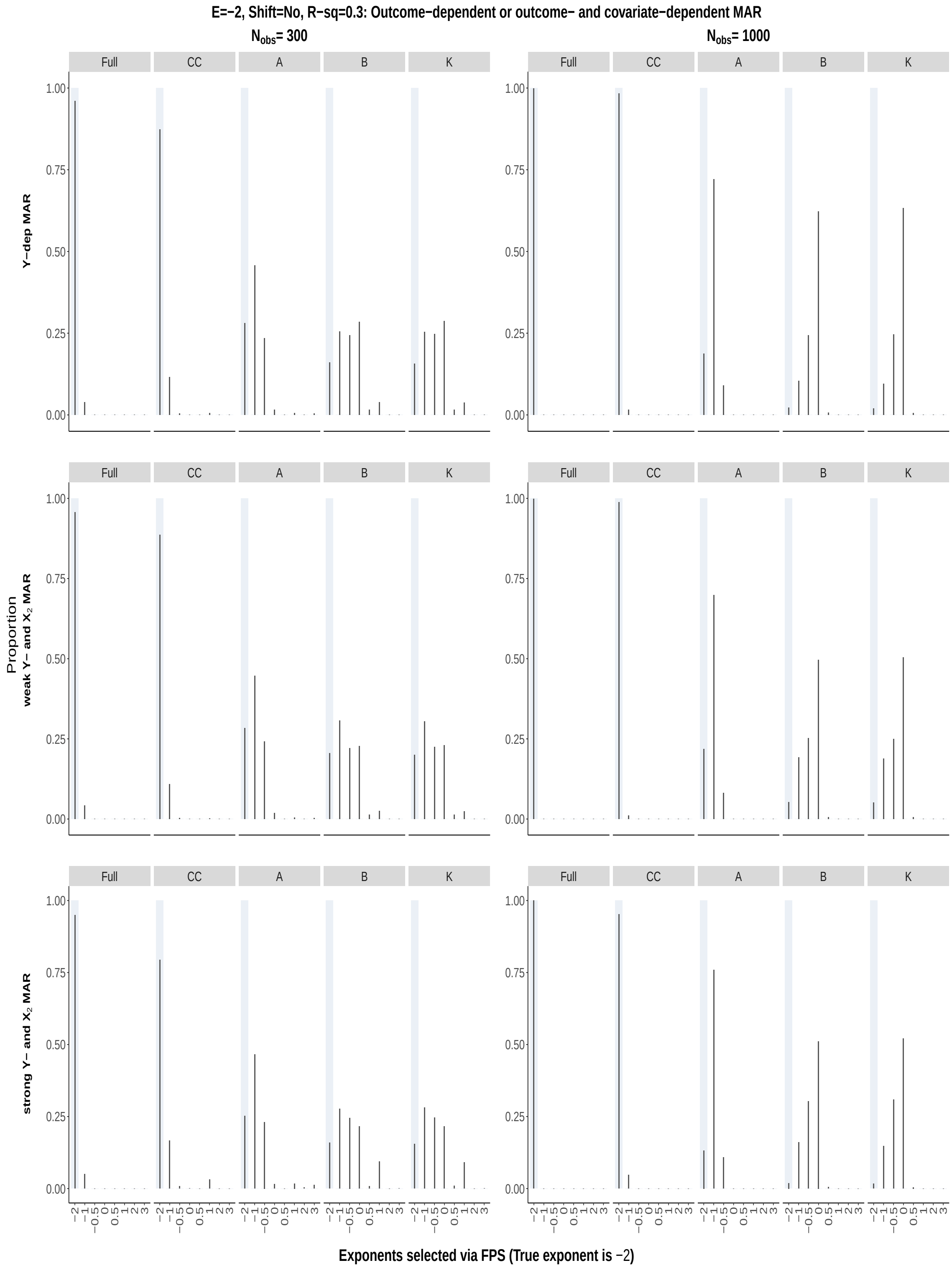

Figure S48: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.3 Cross-validation, $\alpha_E = 1$ and an origin-shift has been applied

True exponent is 0

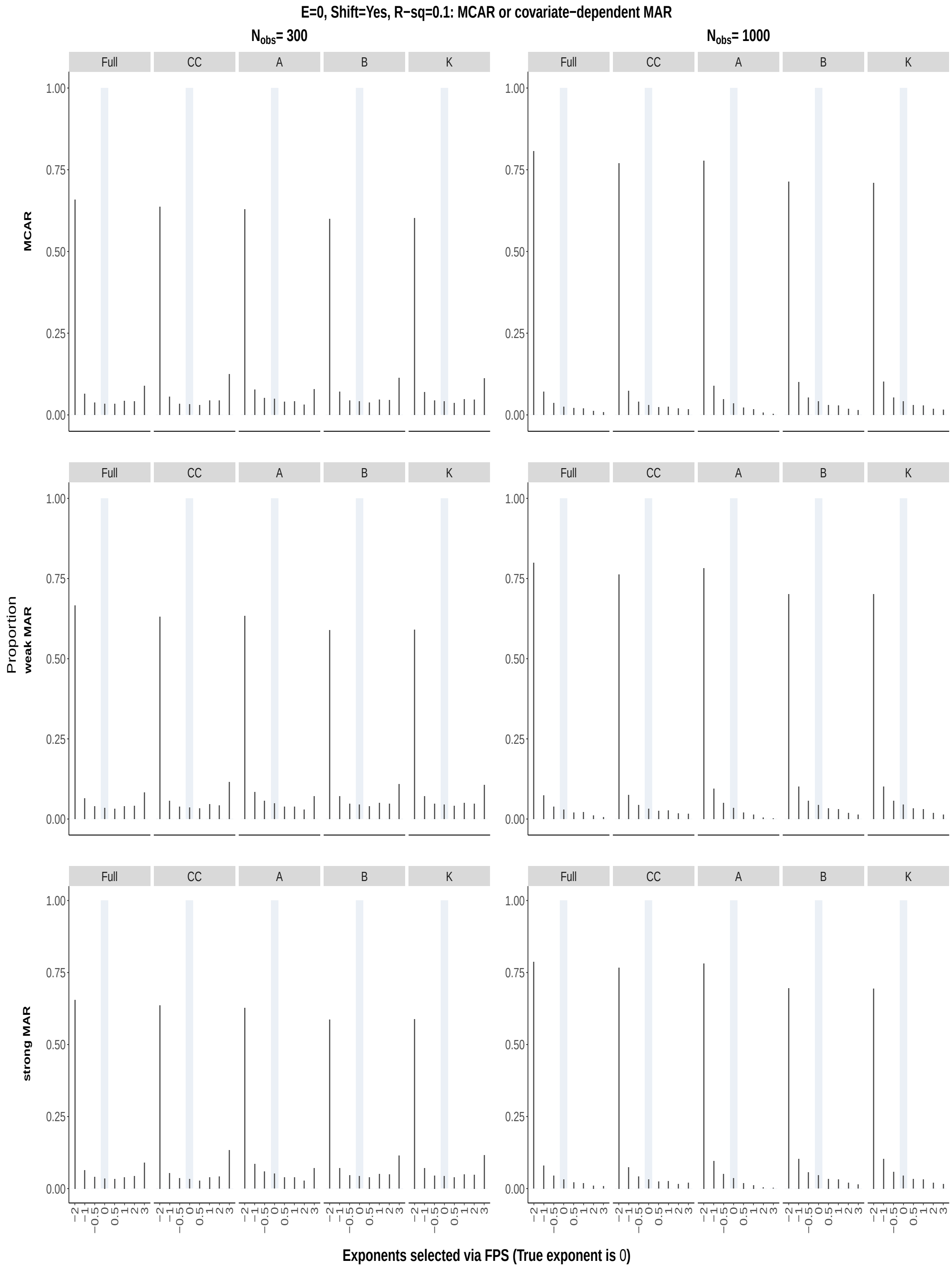

Figure S49: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

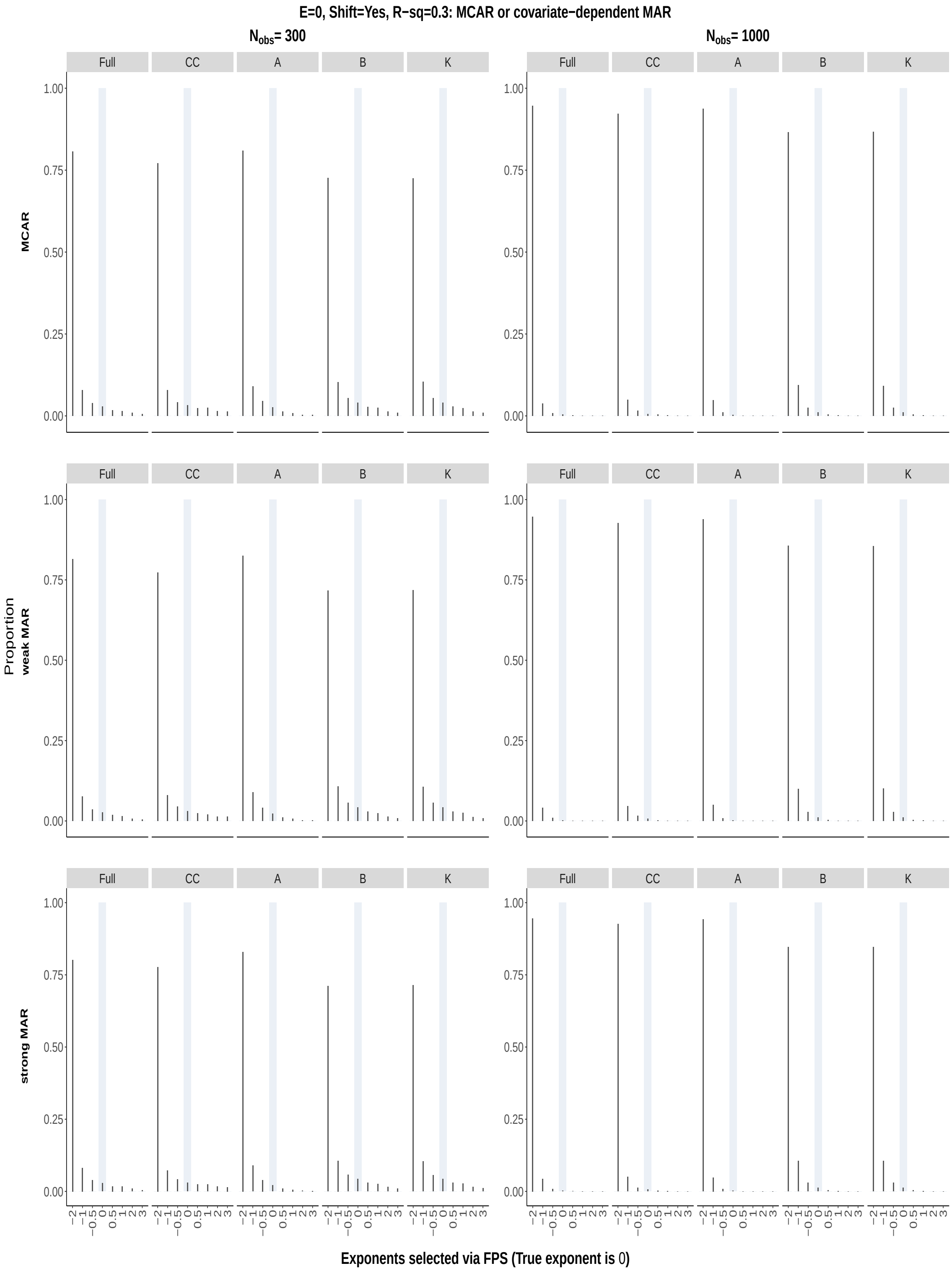

Figure S50: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

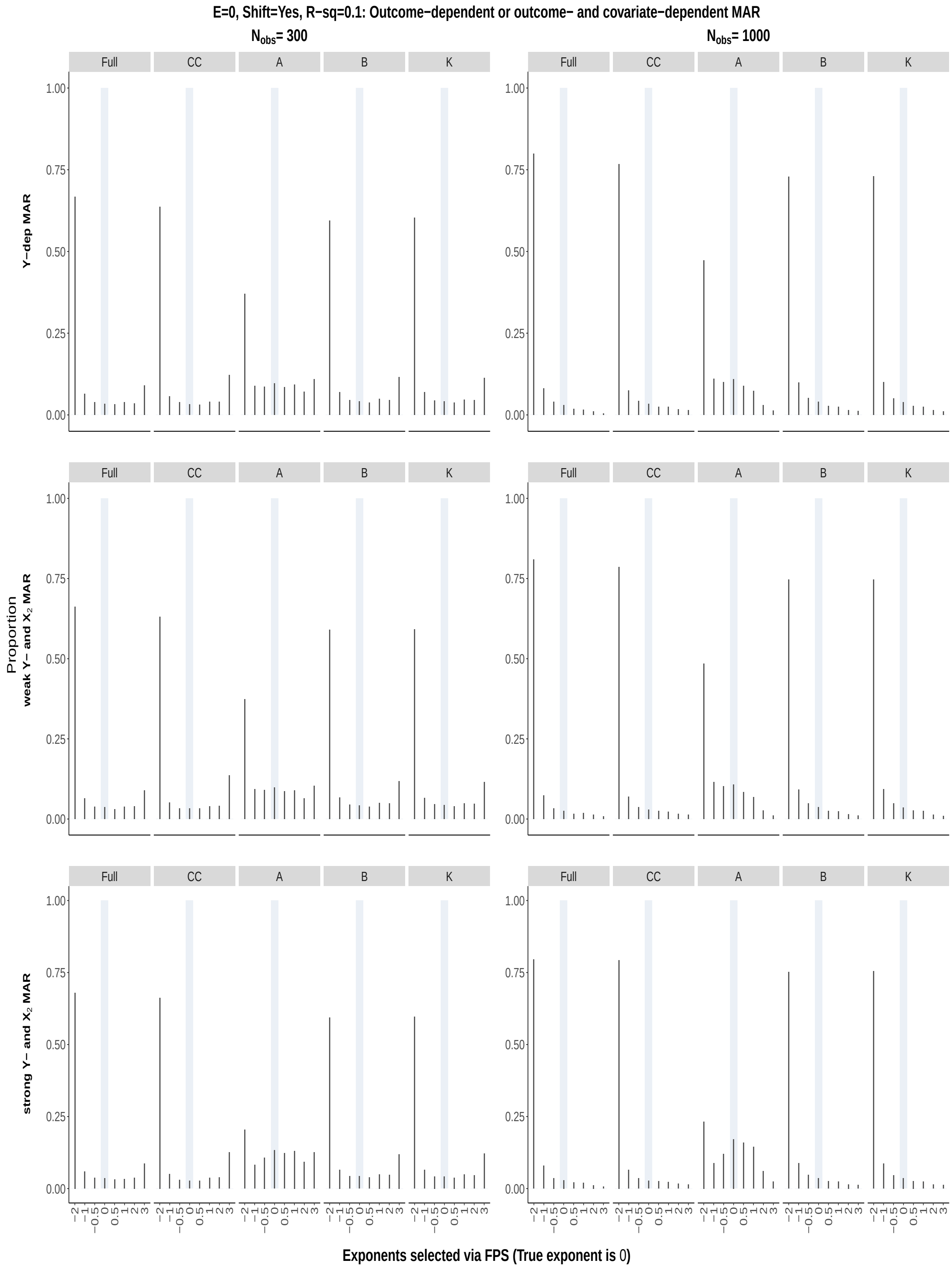

Figure S51: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

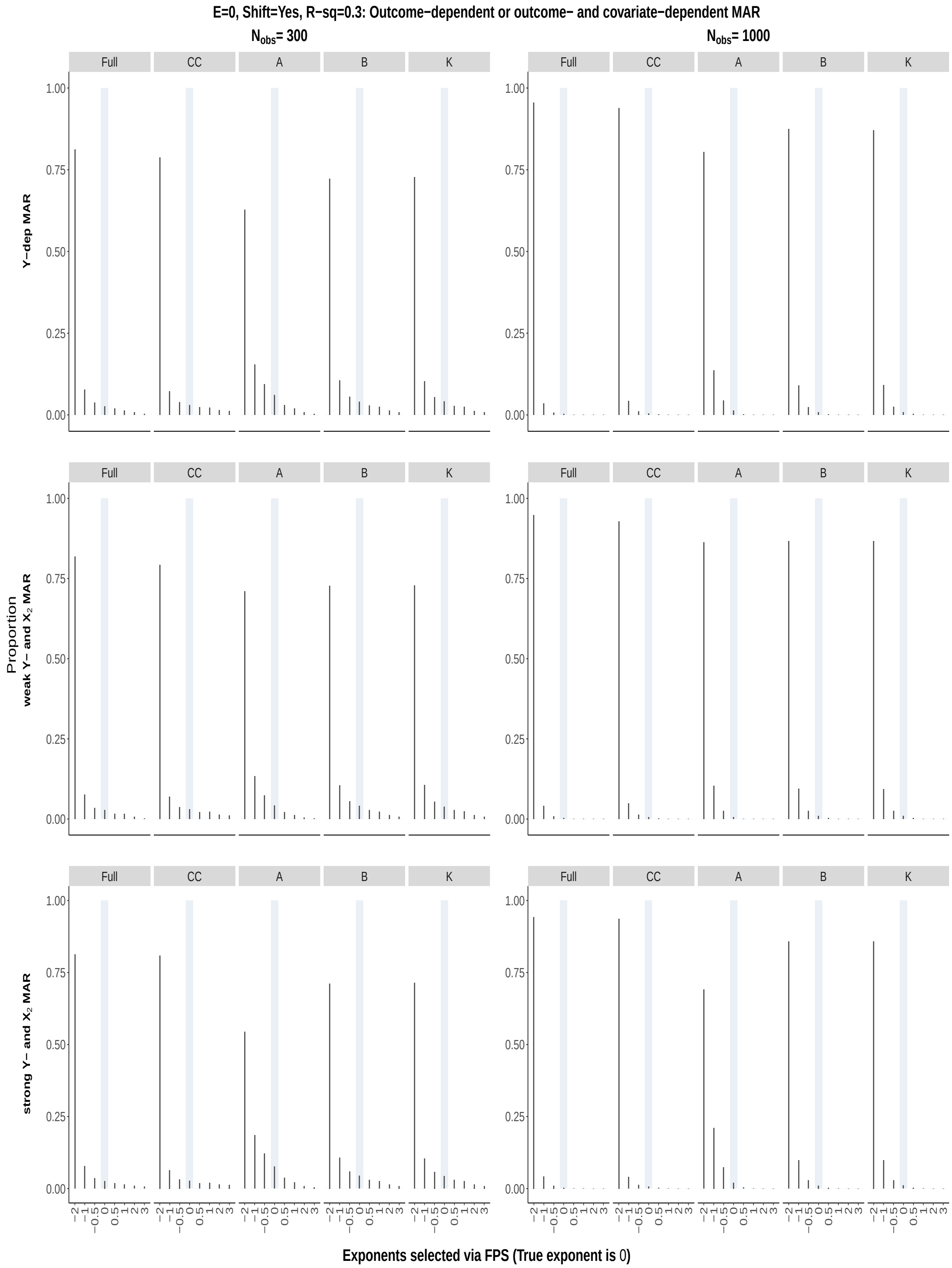

Figure S52: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

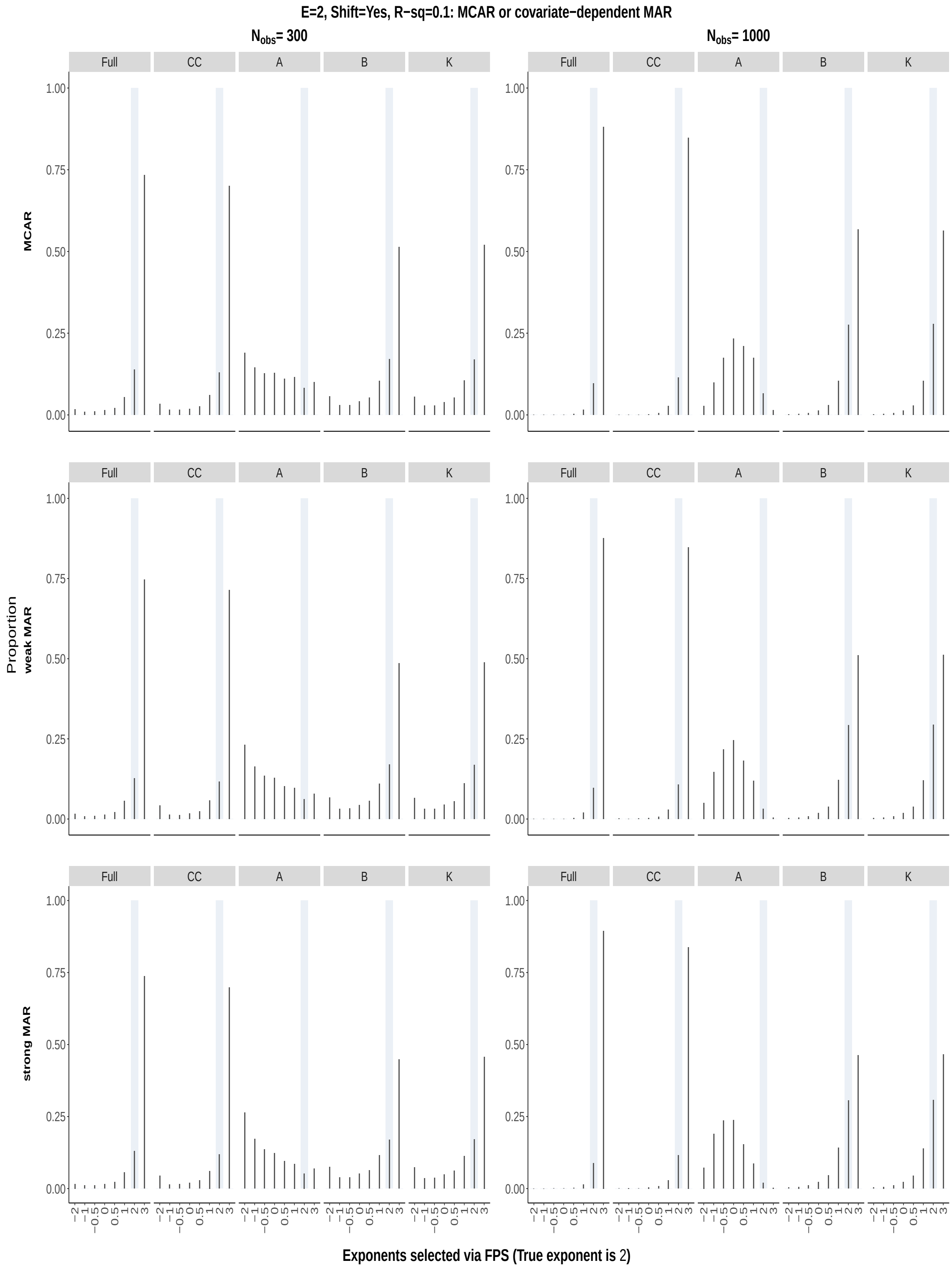

Figure S53: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

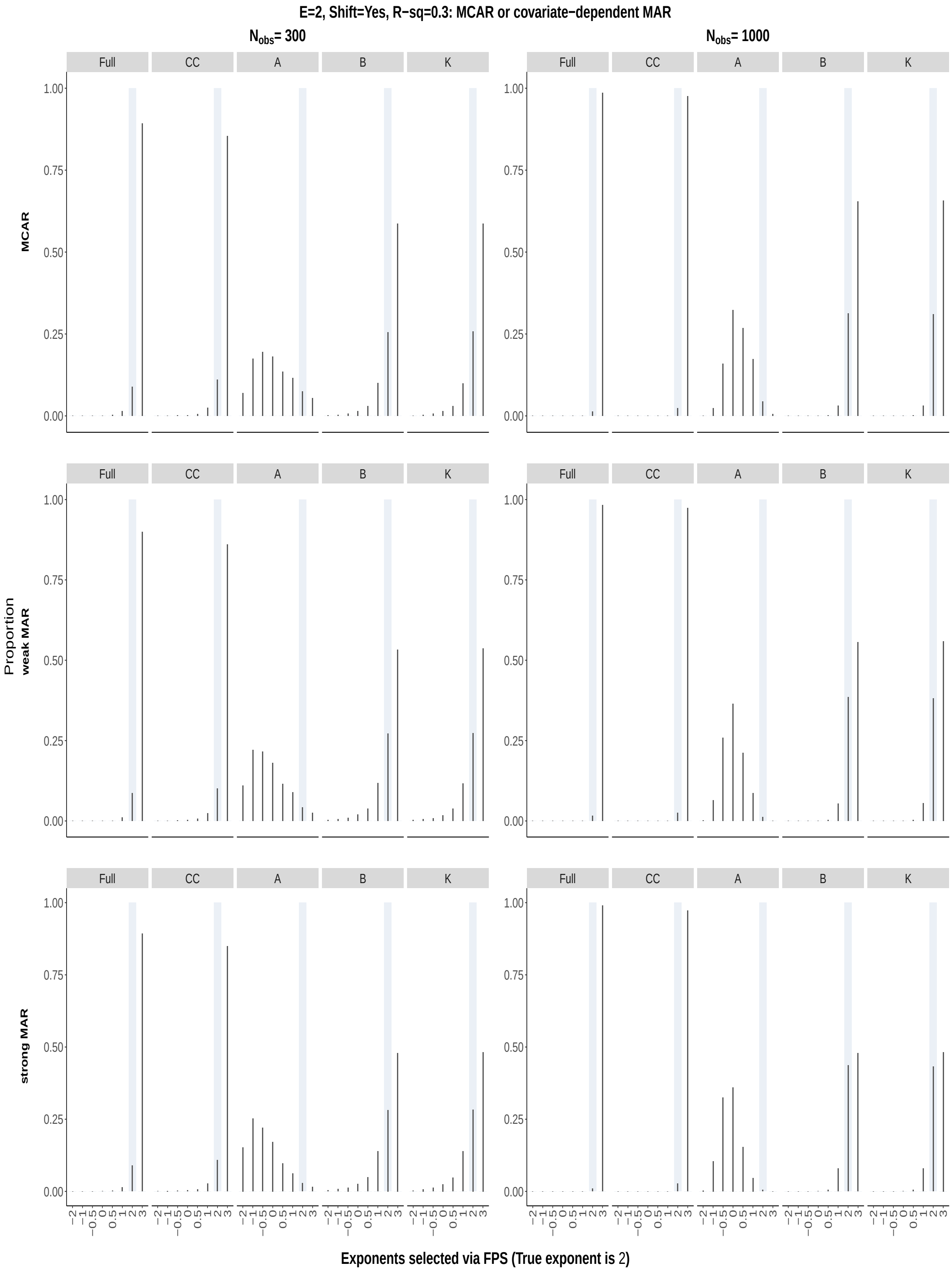

Figure S54: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

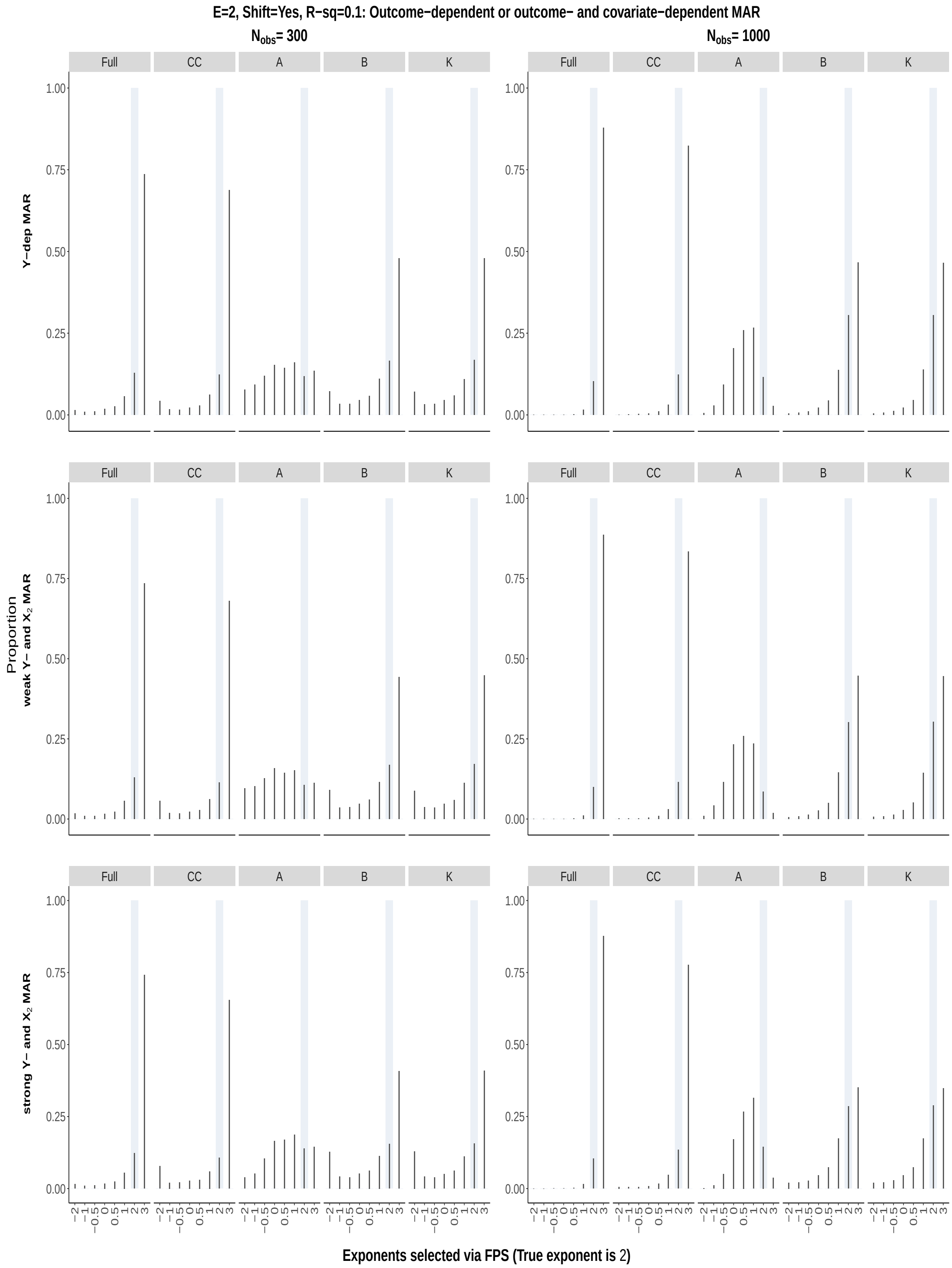

Figure S55: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

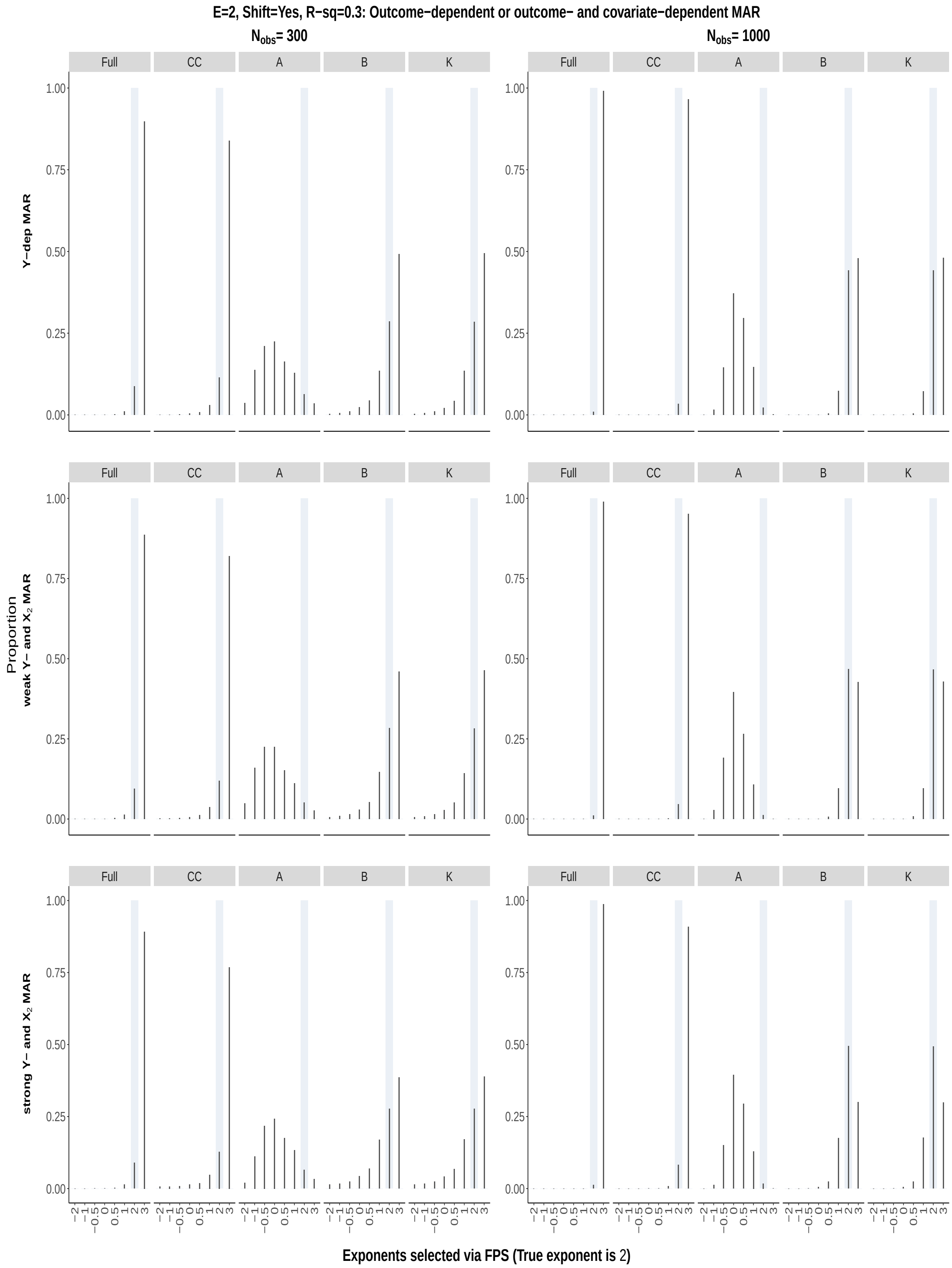

Figure S56: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

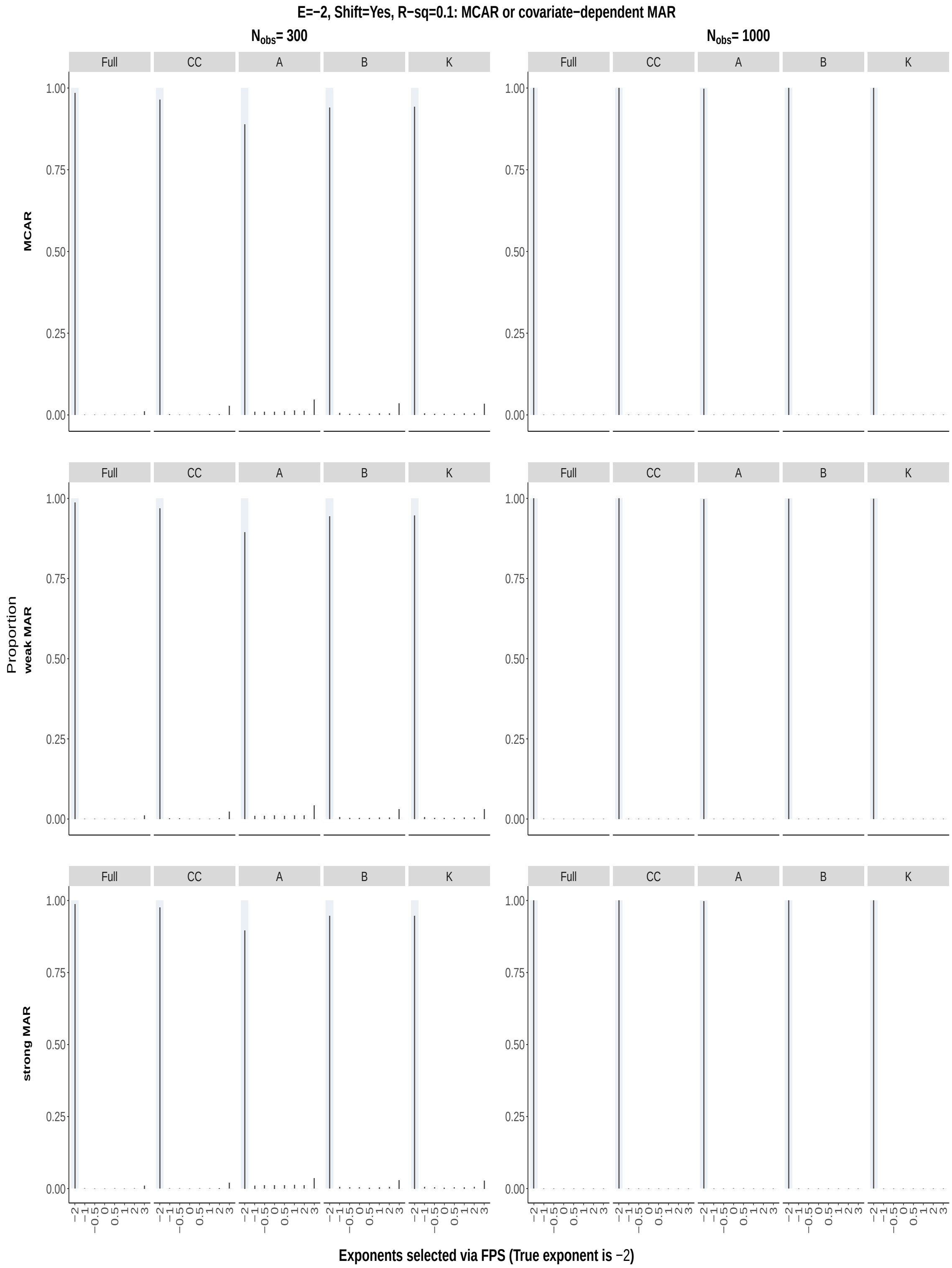

Figure S57: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

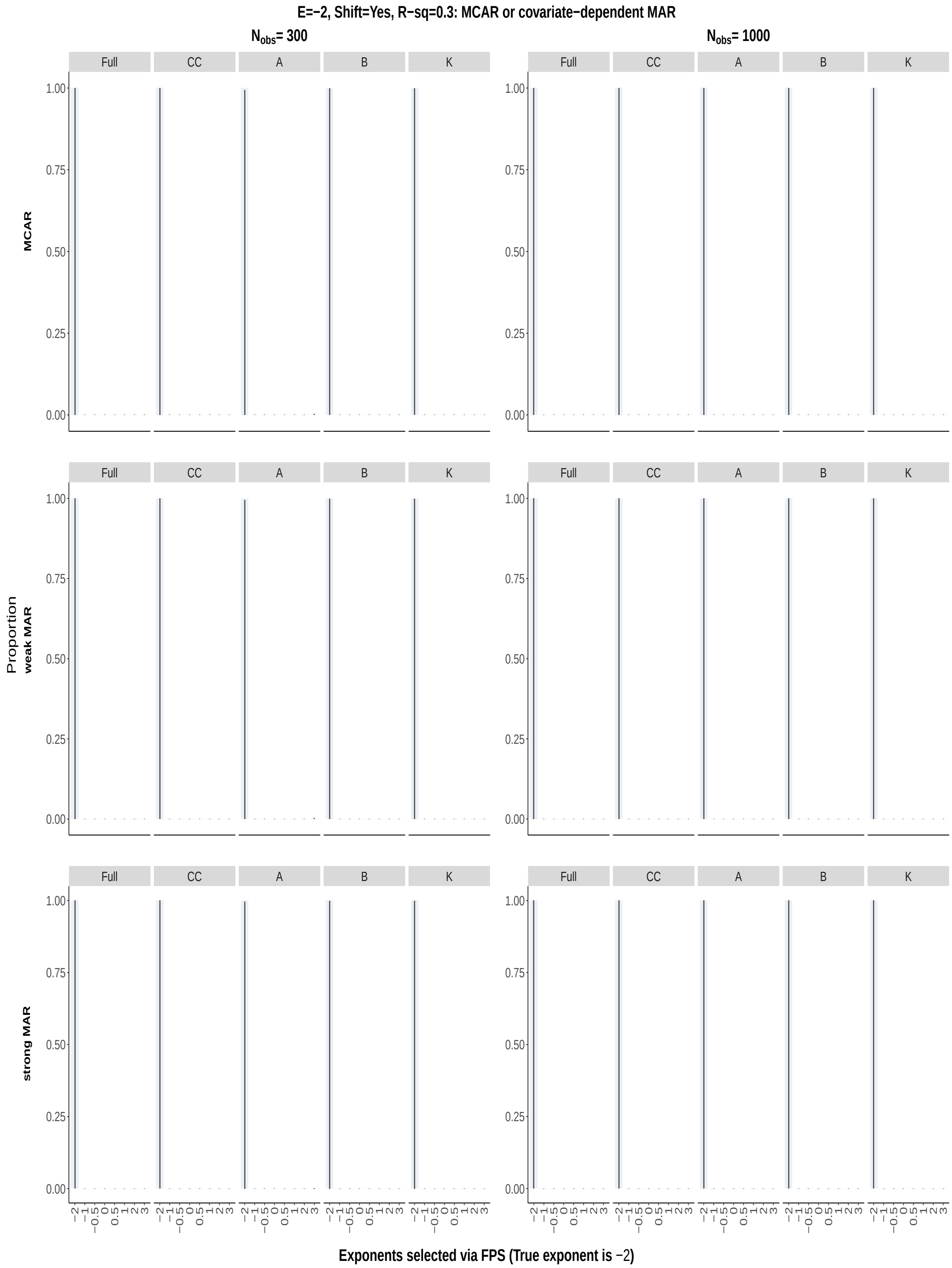

Figure S58: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

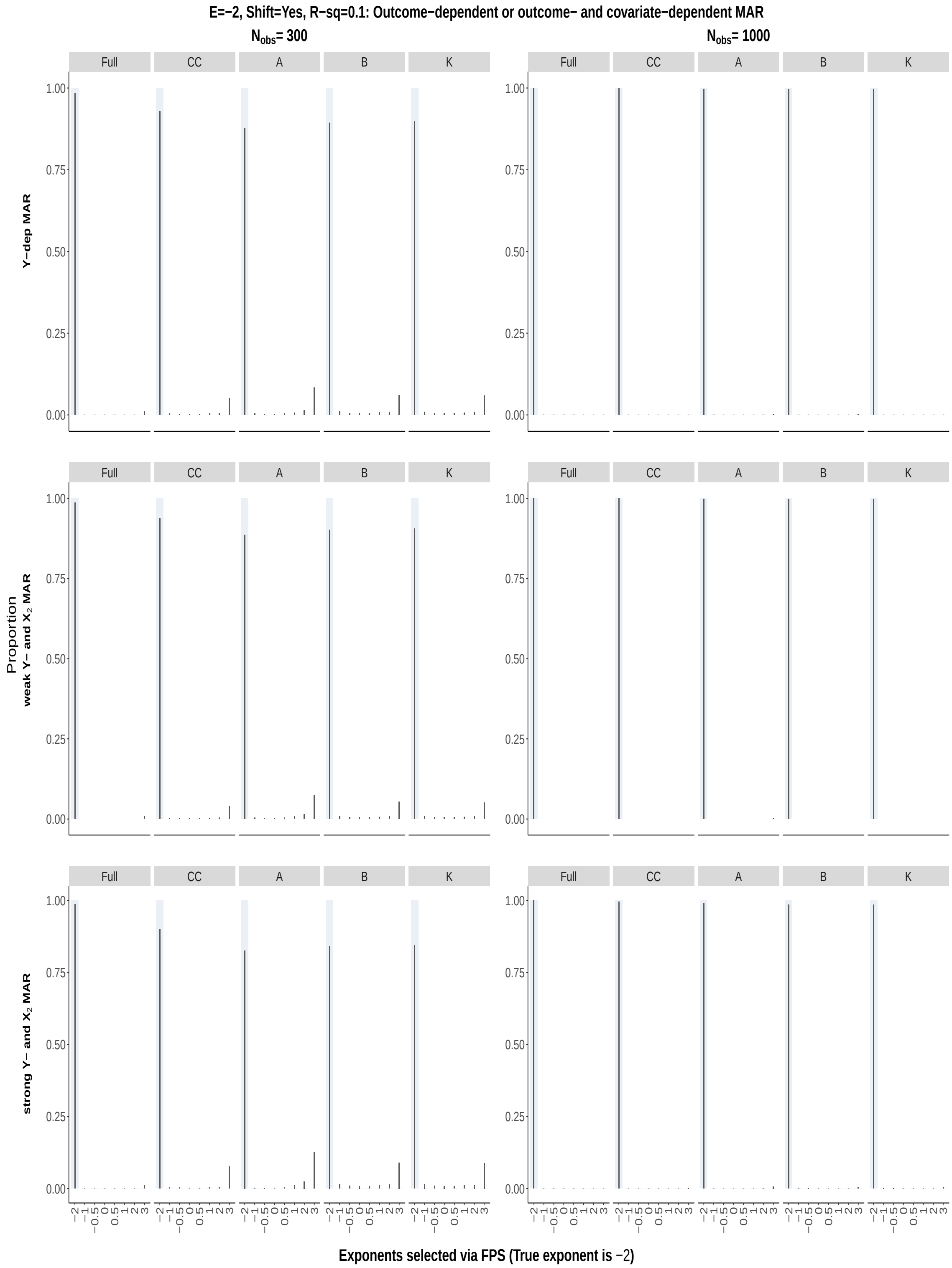

Figure S59: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

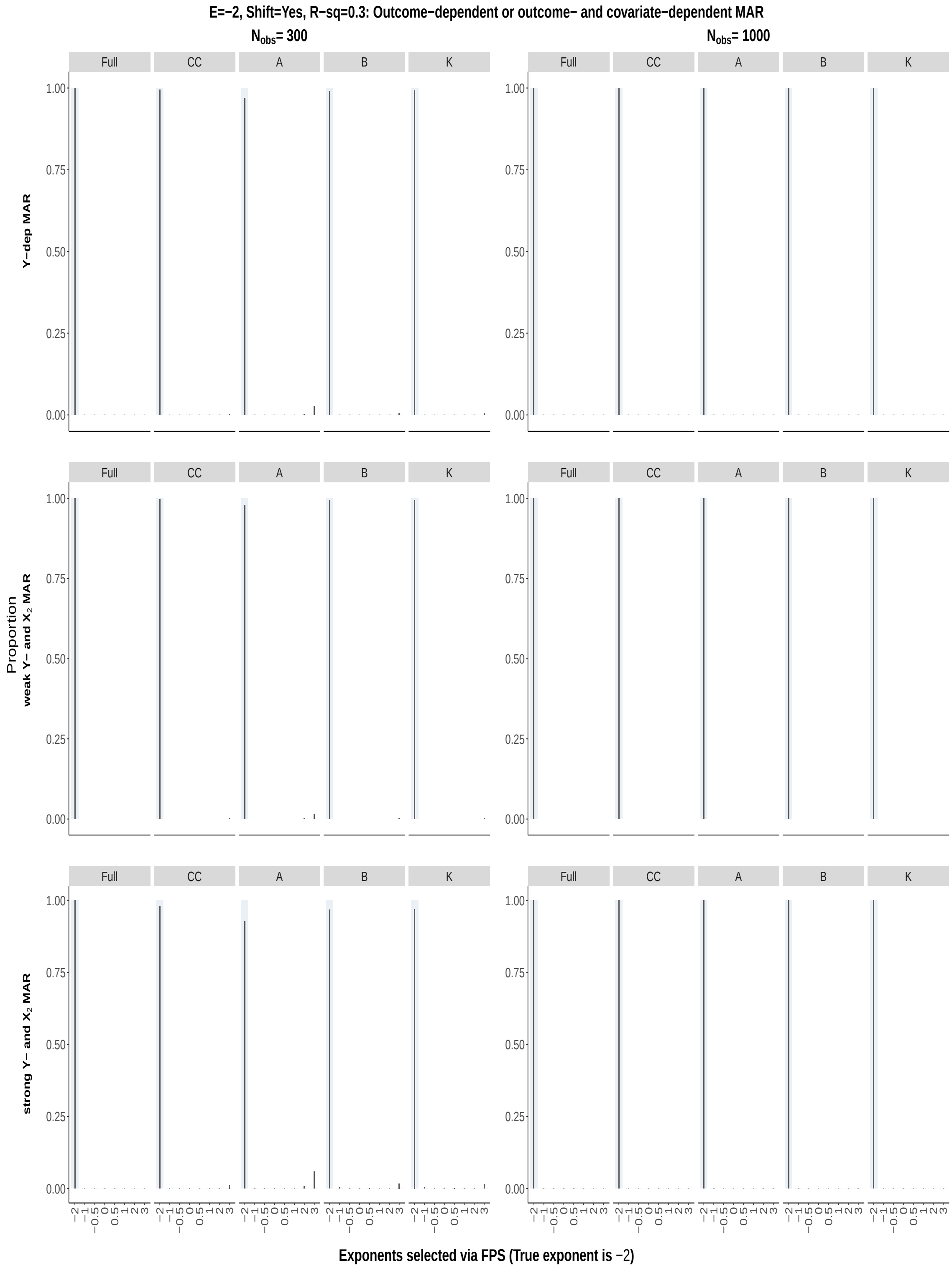

Figure S60: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.4 Cross-validation, $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

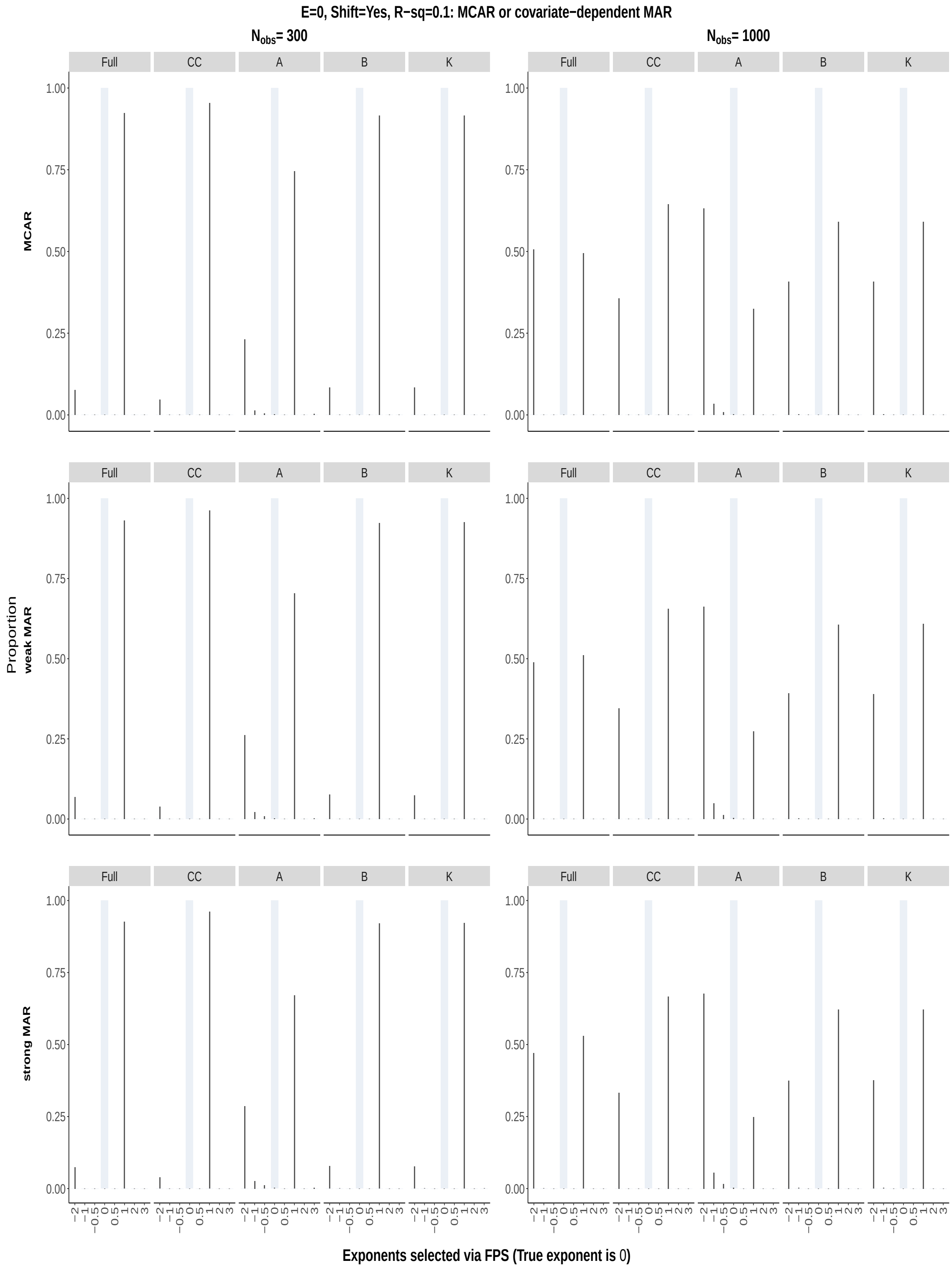

Figure S61: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

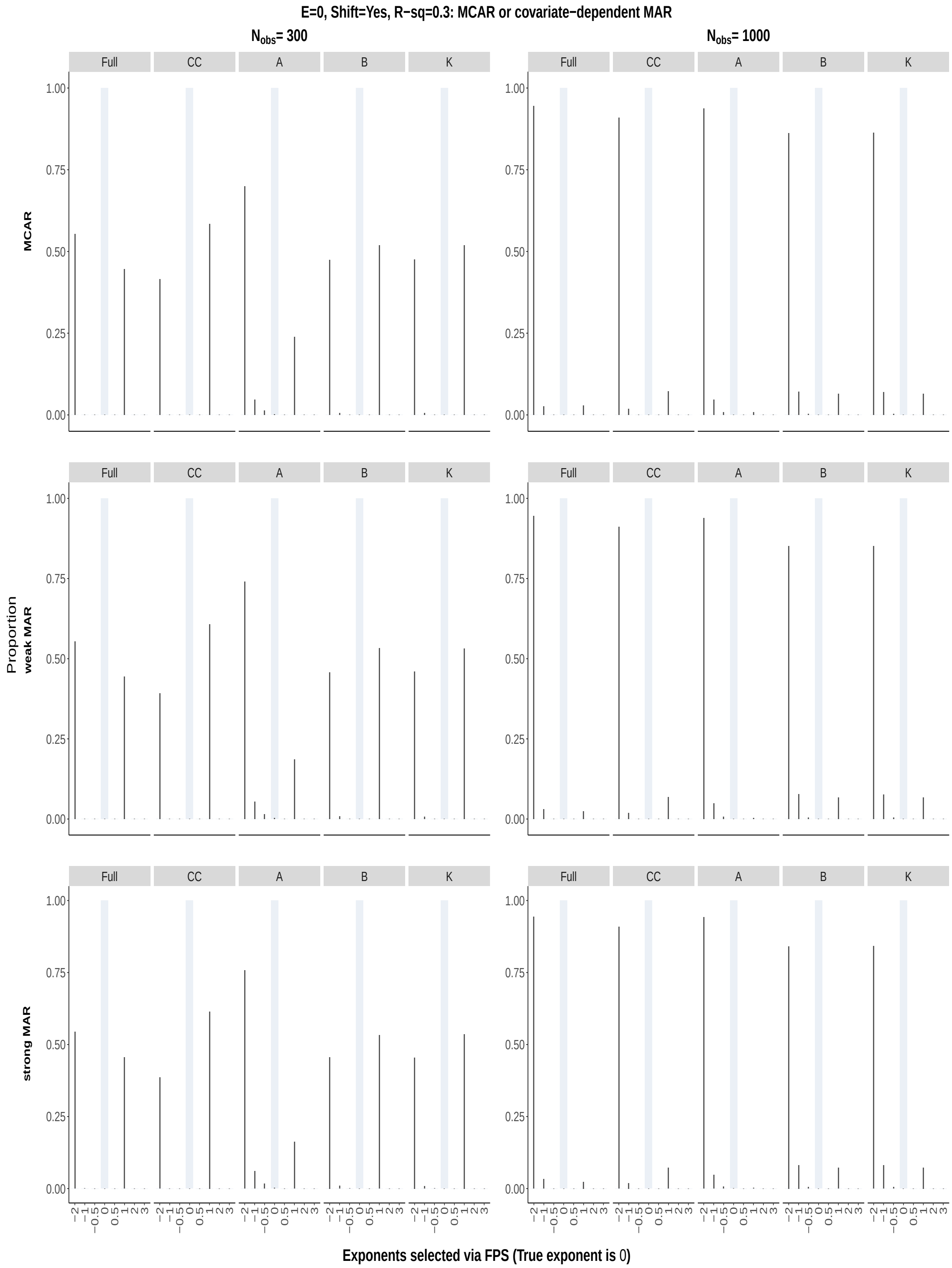

Figure S62: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

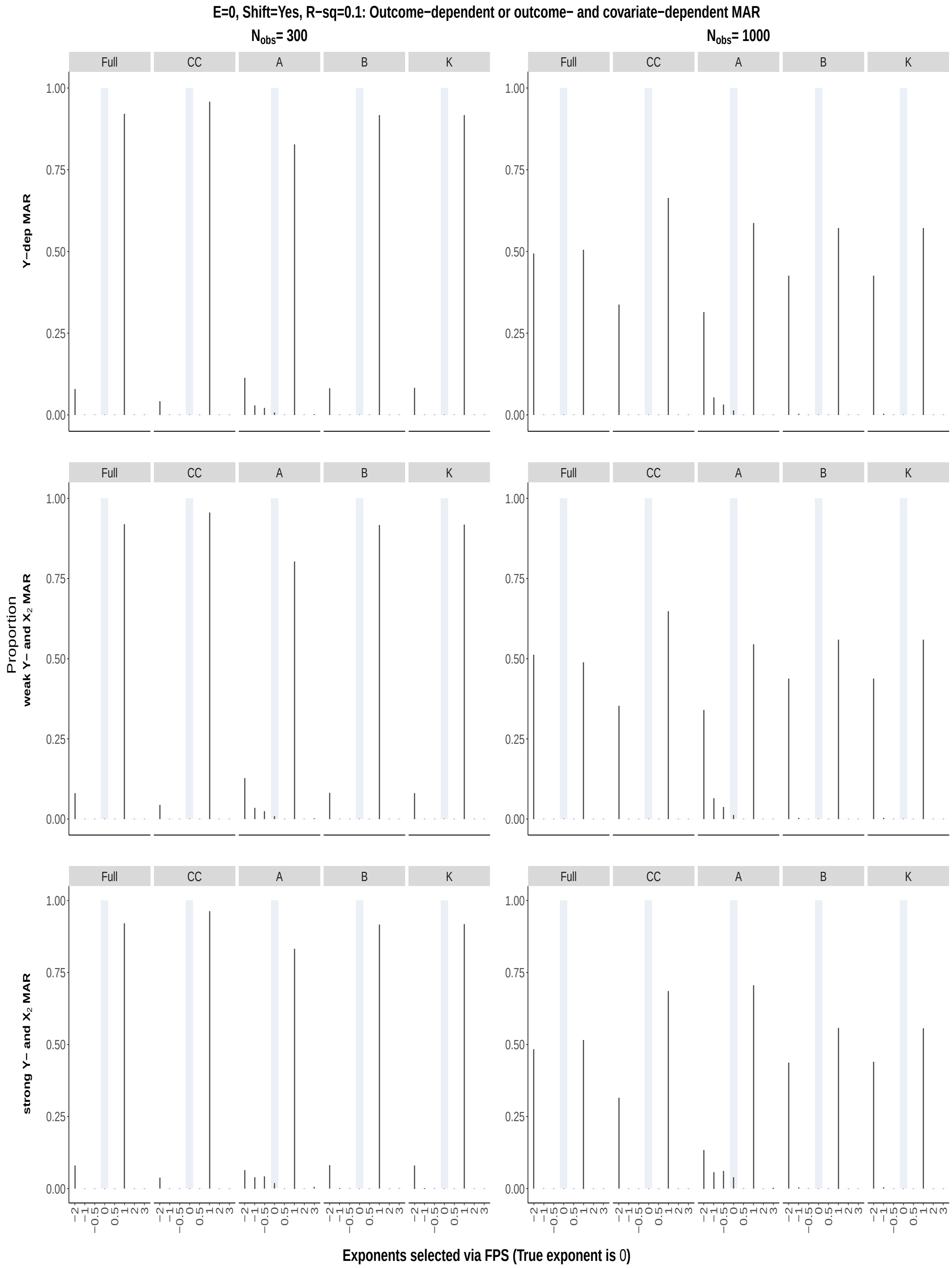

Figure S63: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

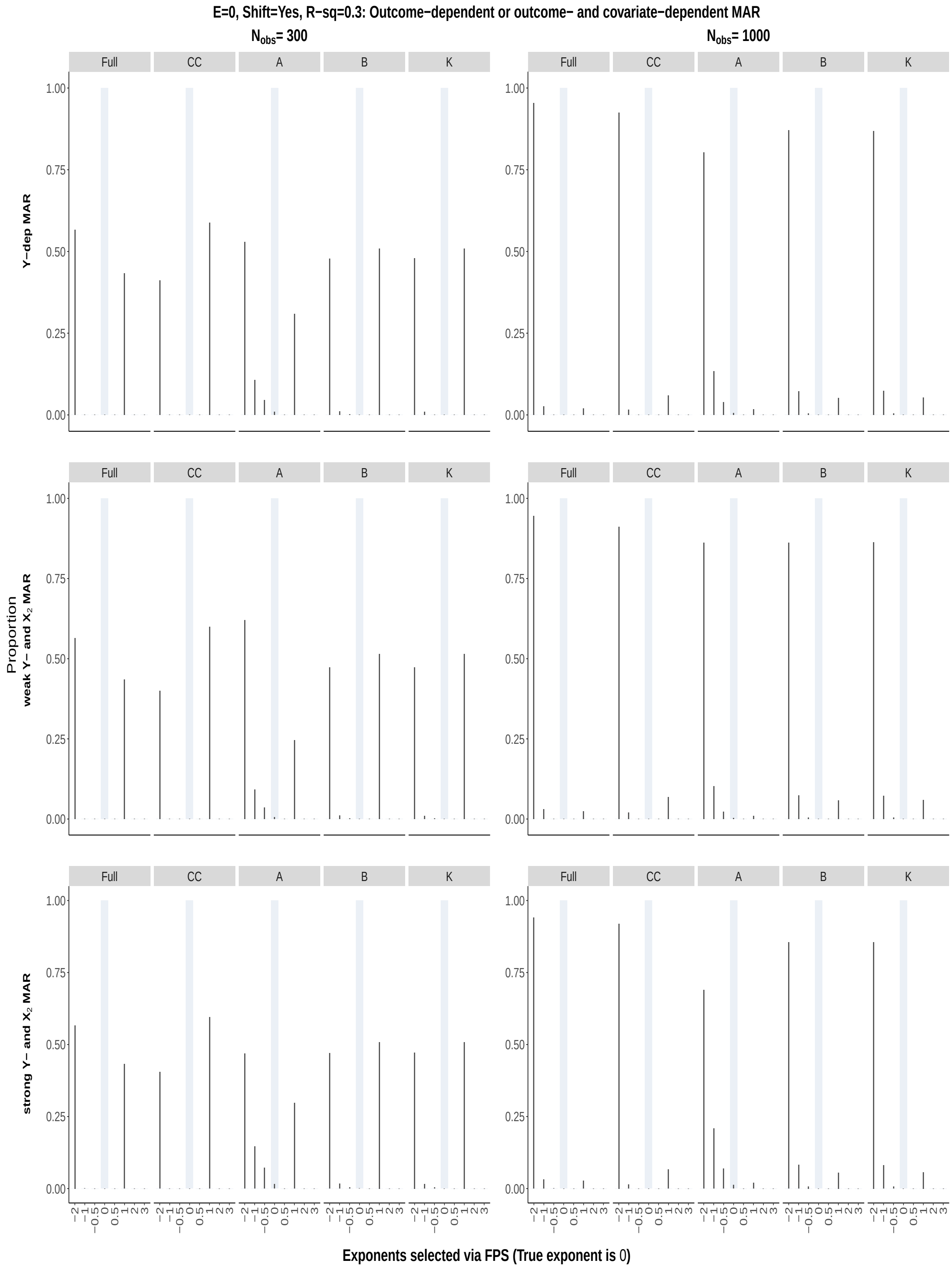

Figure S64: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

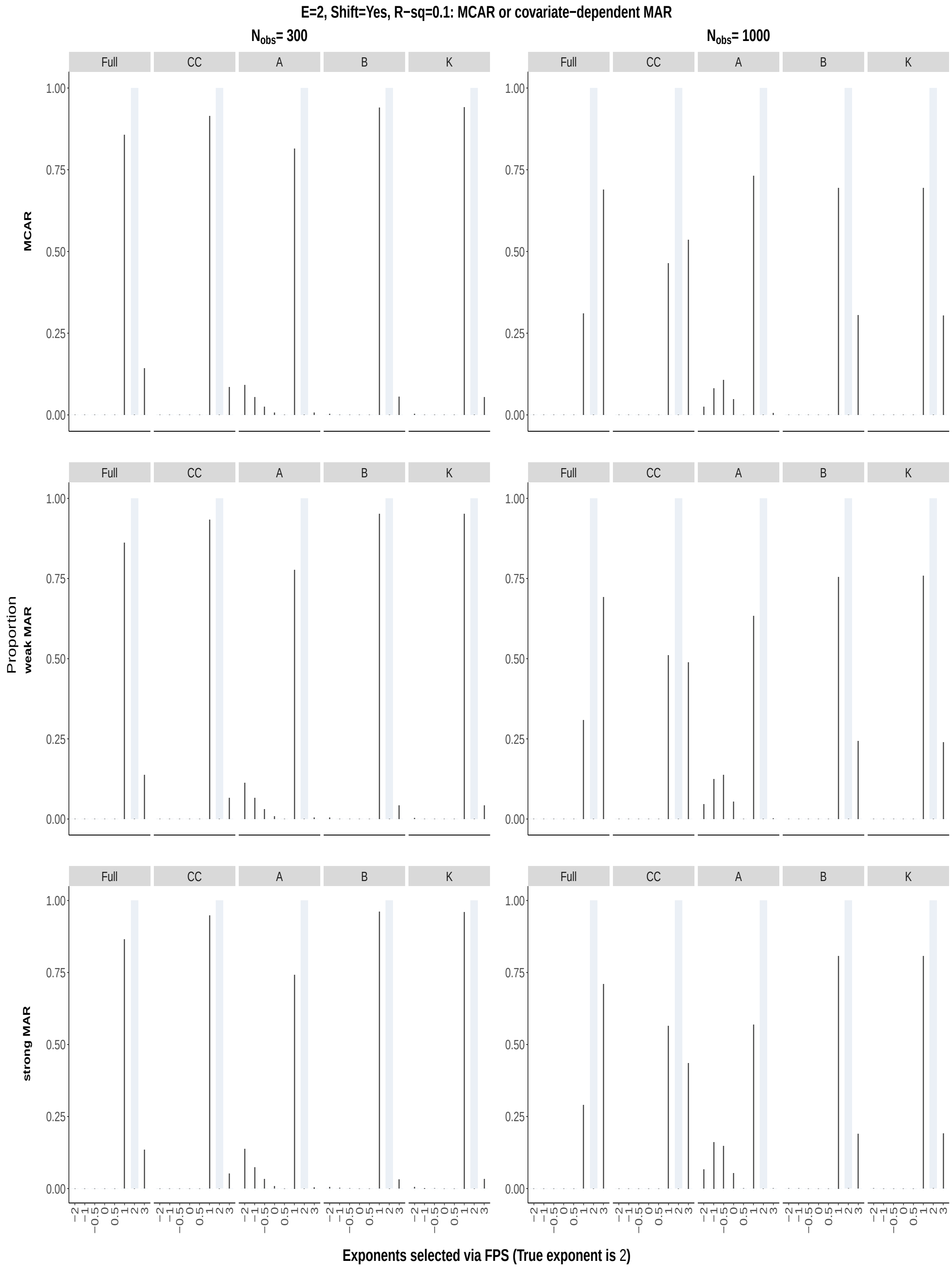

Figure S65: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

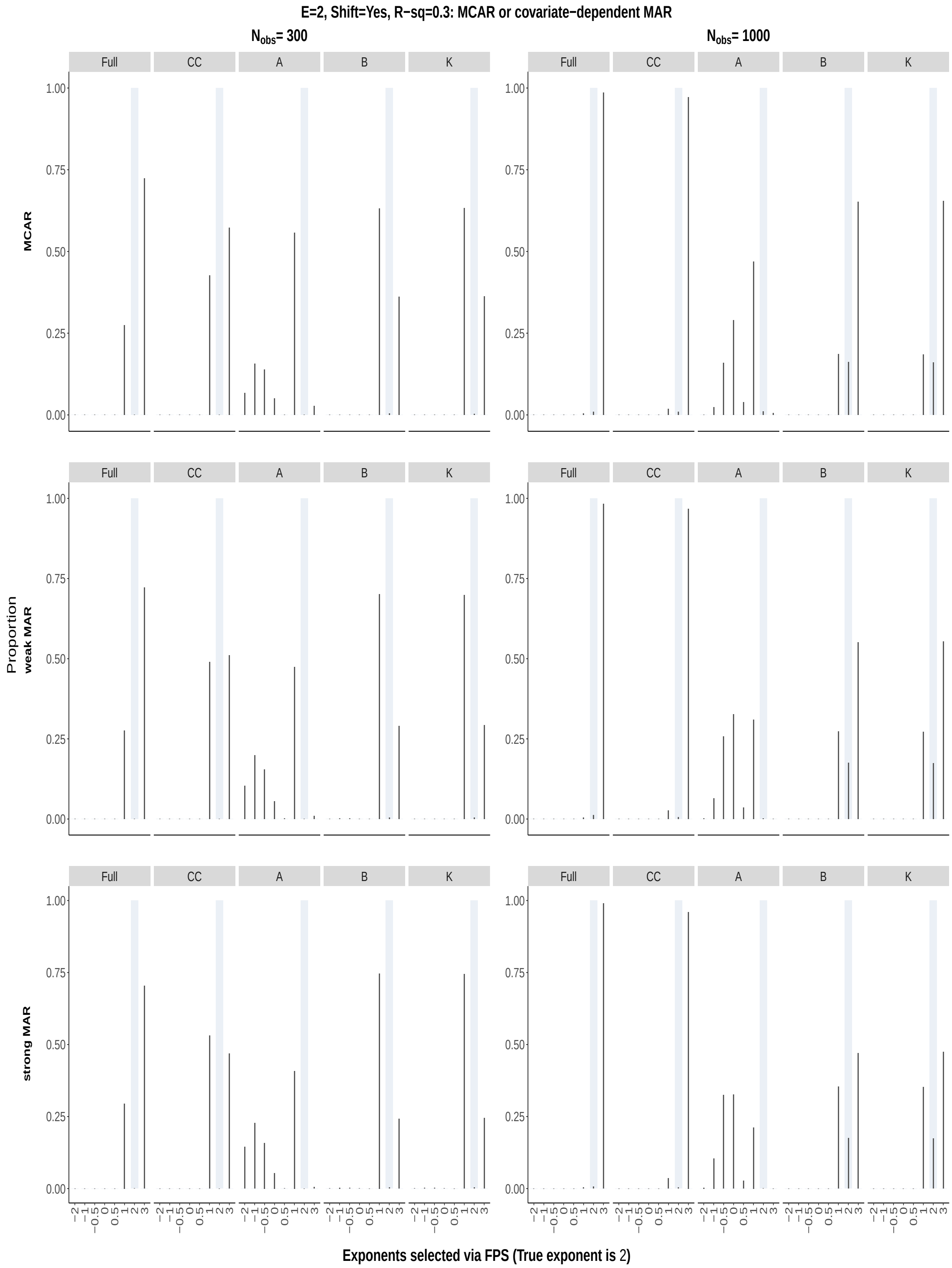

Figure S66: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

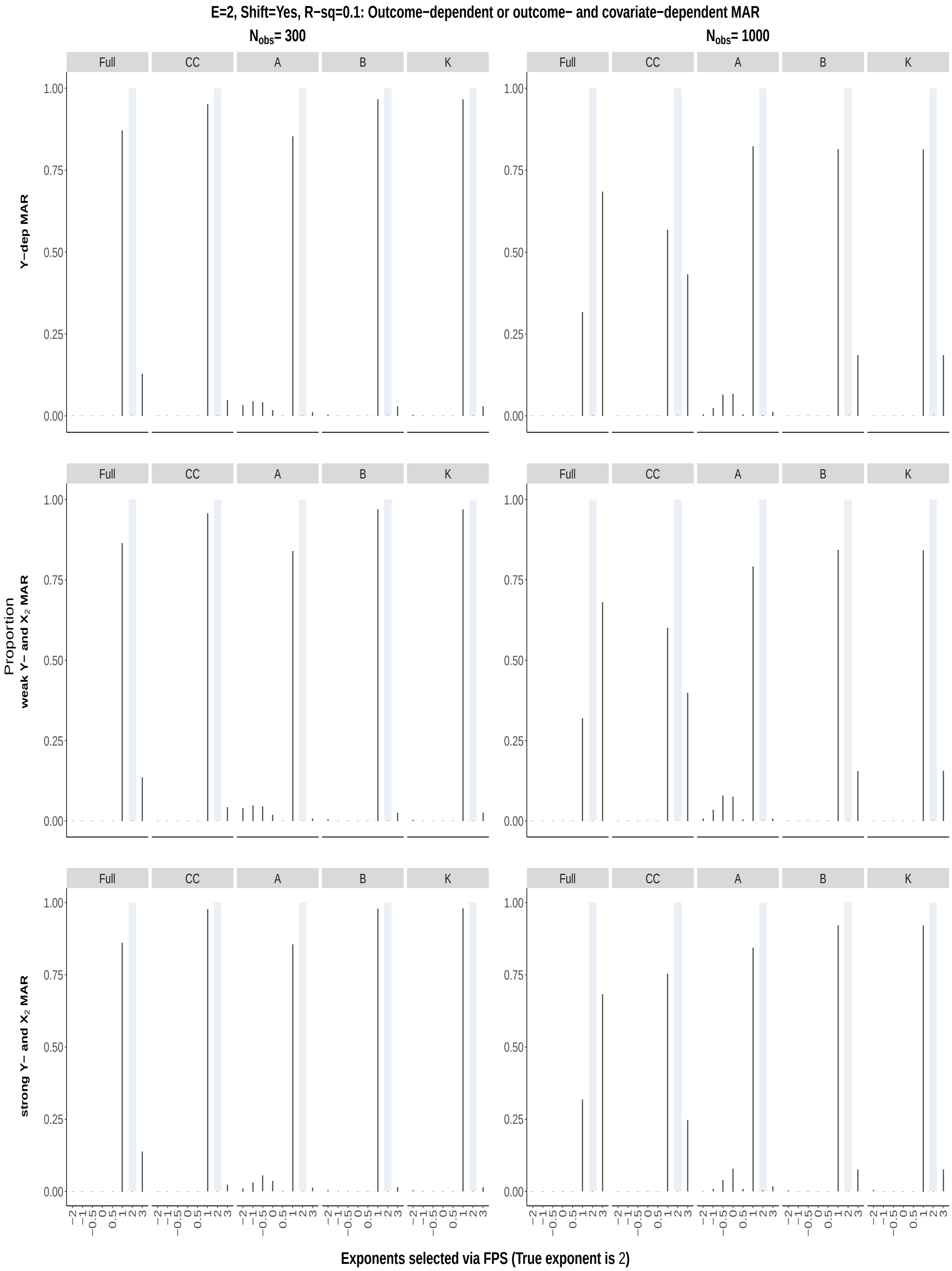

Figure S67: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

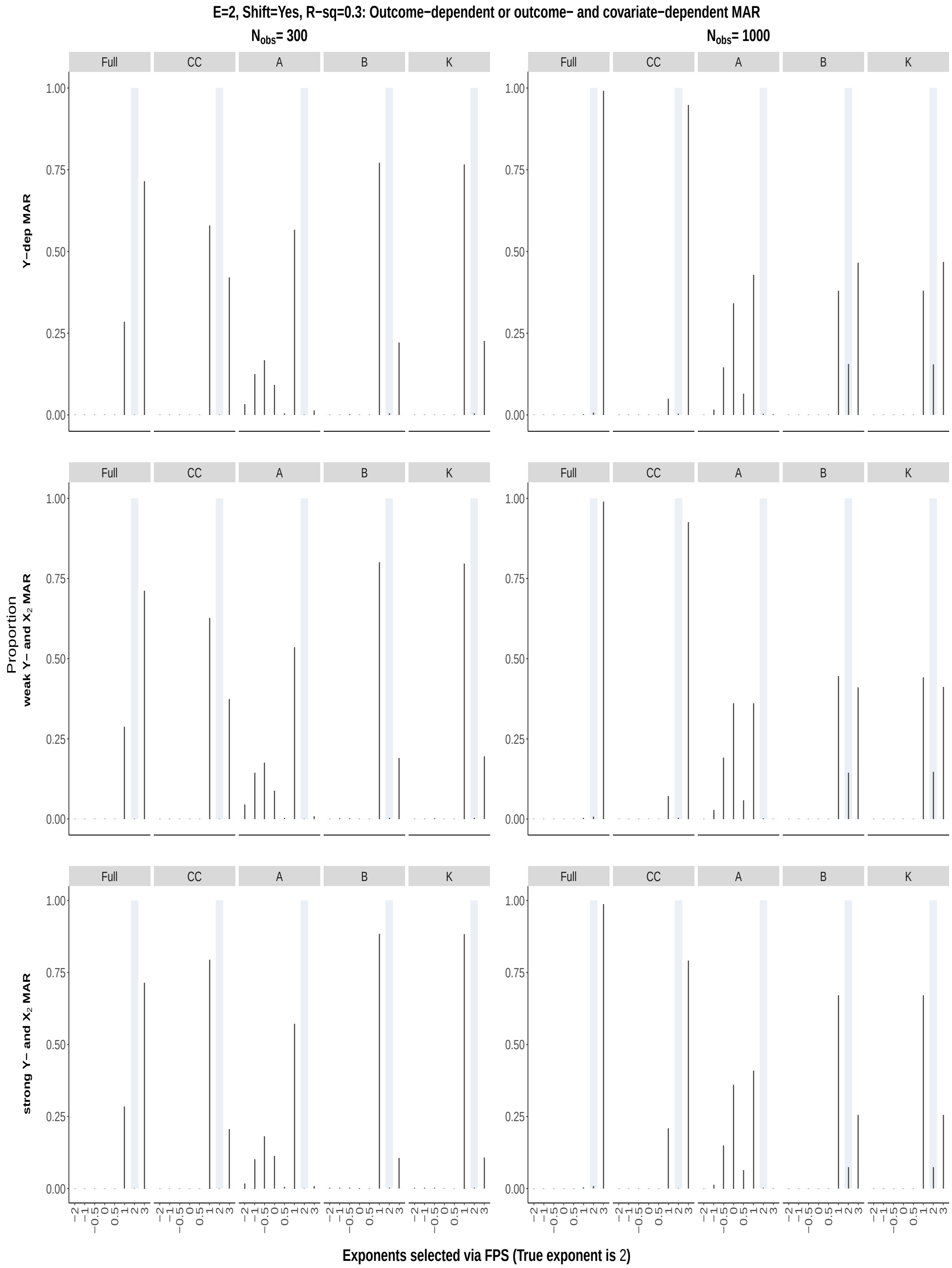

Figure S68: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

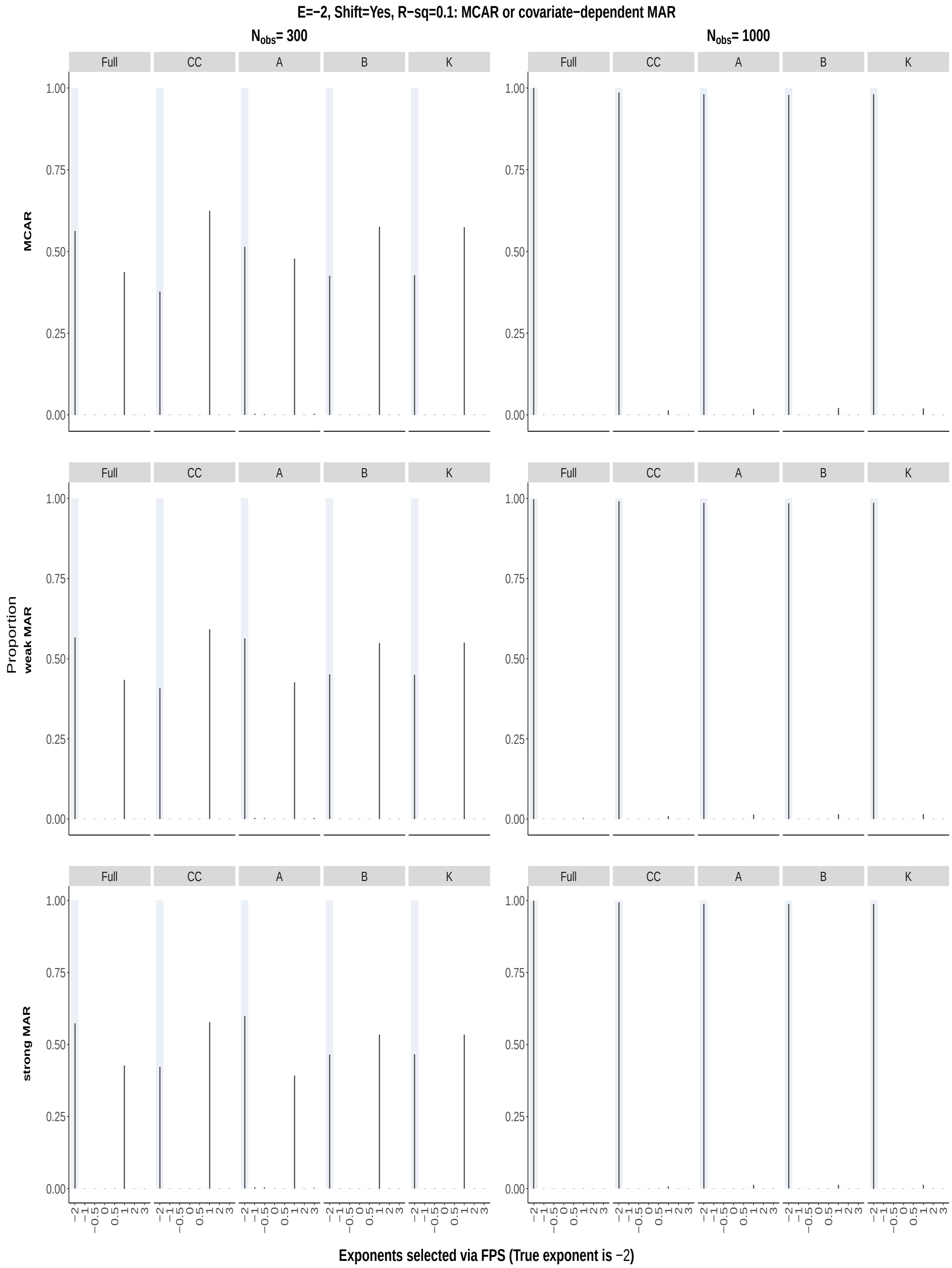

Figure S69: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

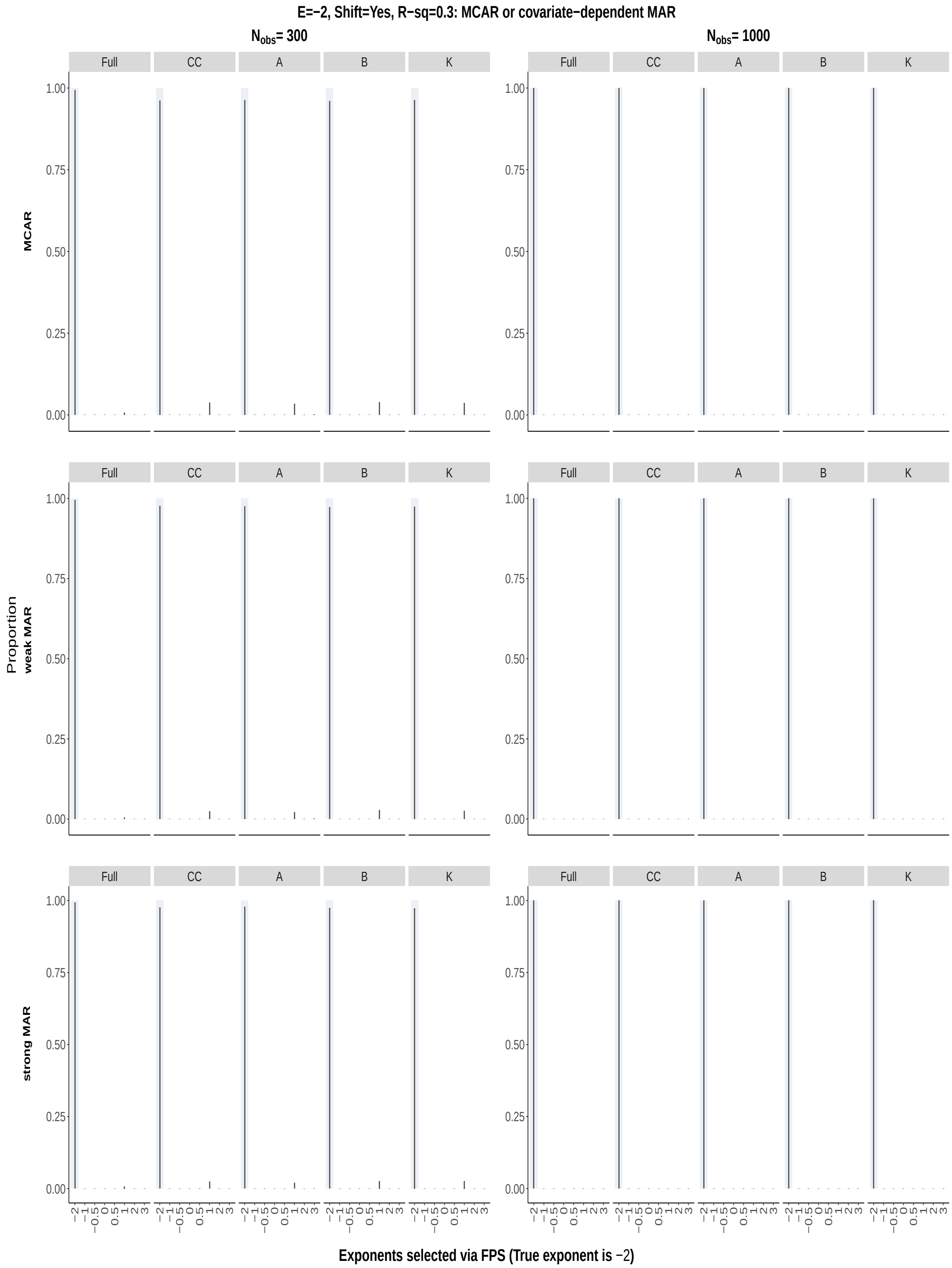

Figure S70: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

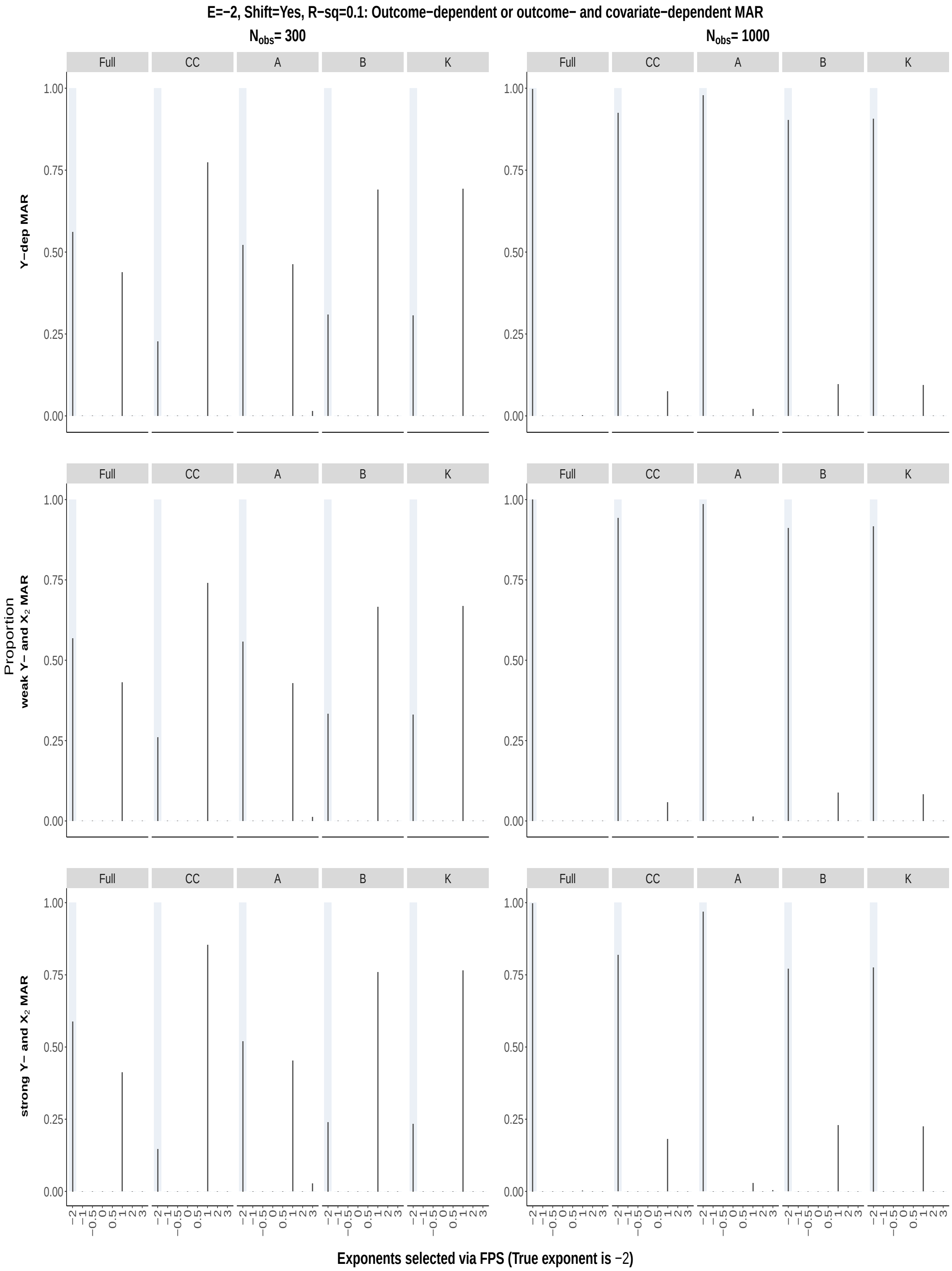

Figure S71: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

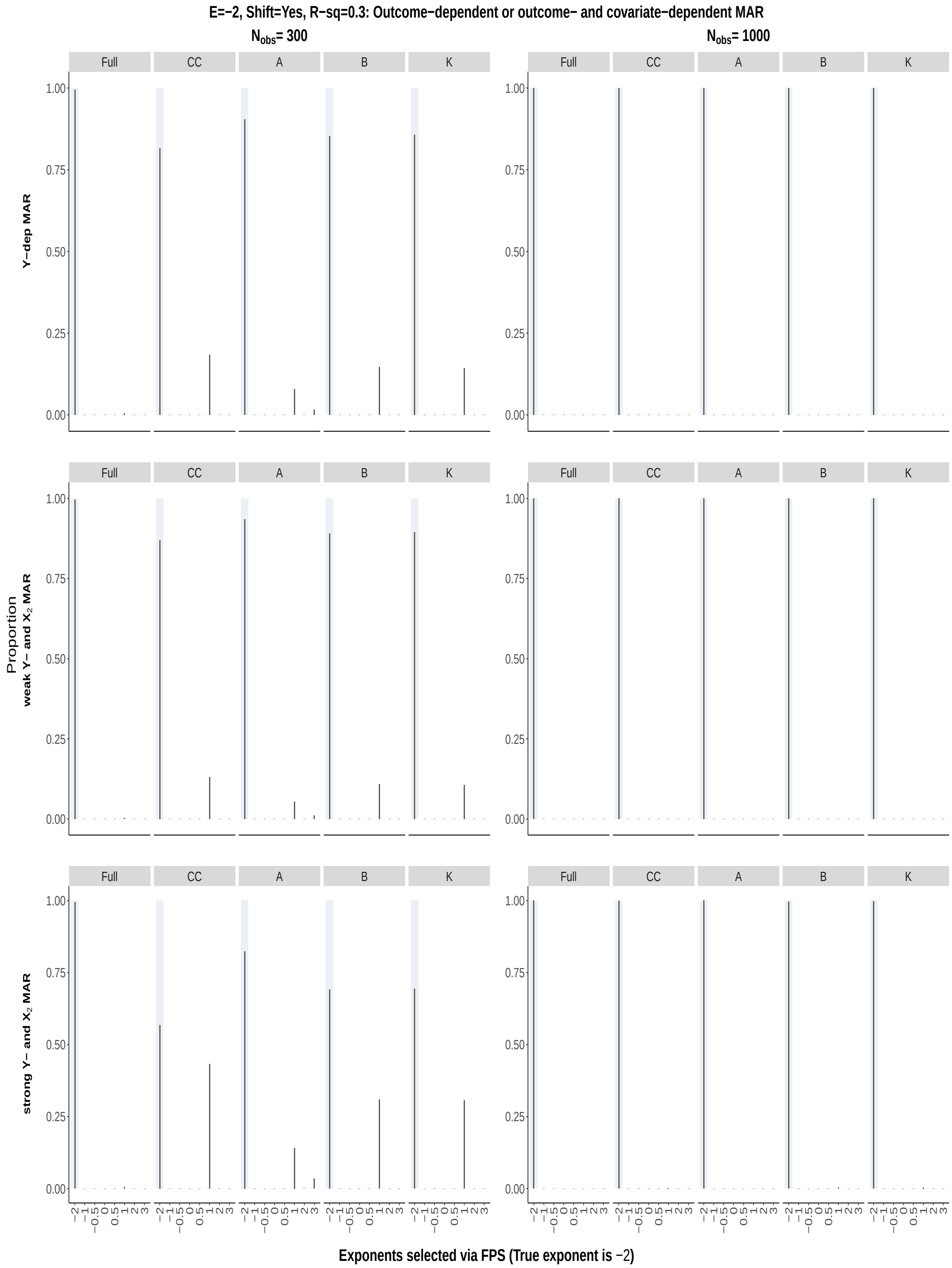

Figure S72: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.5 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\alpha_E = 1$ and no origin-shift

True exponent is 0

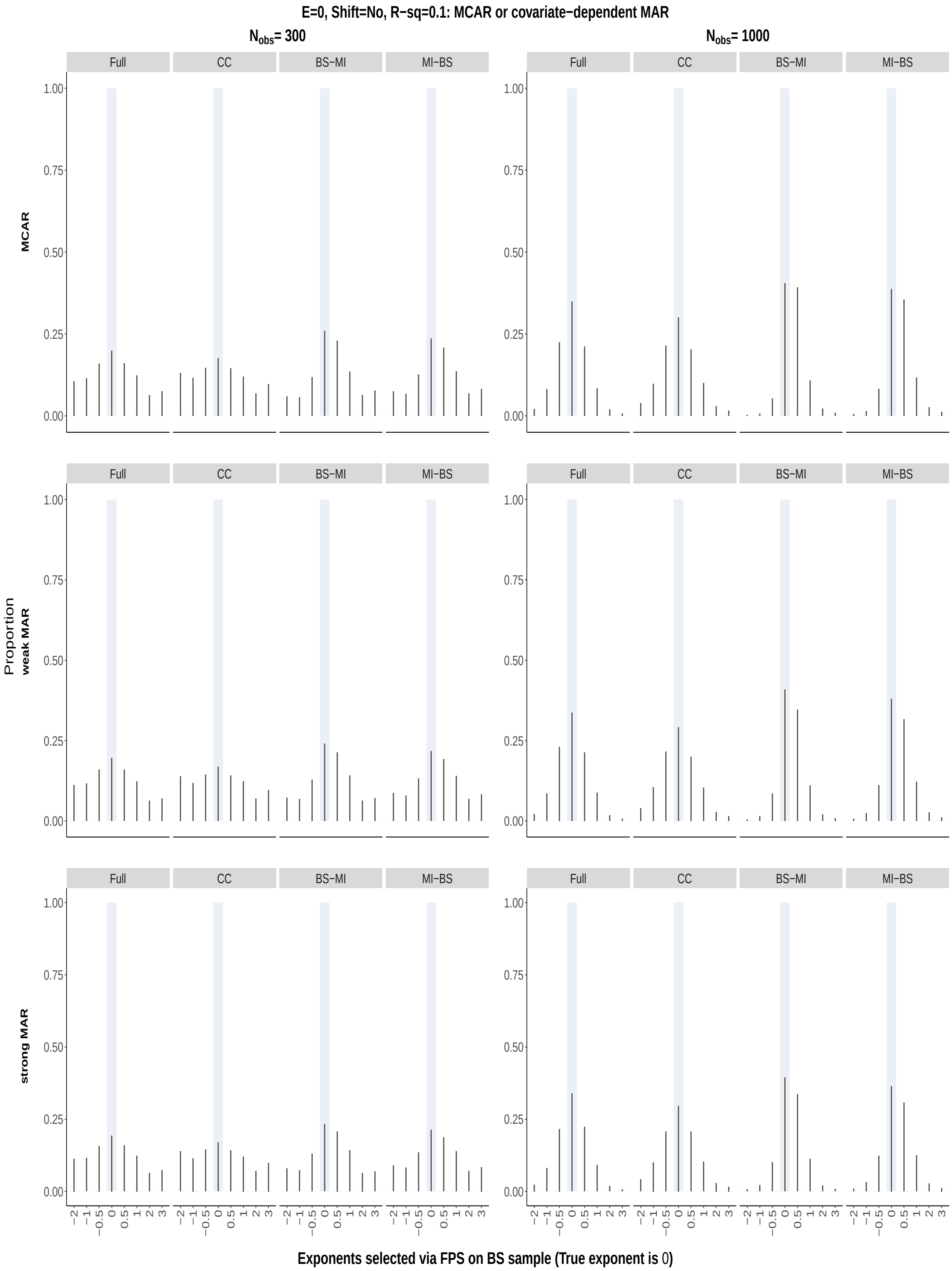

Figure S73: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

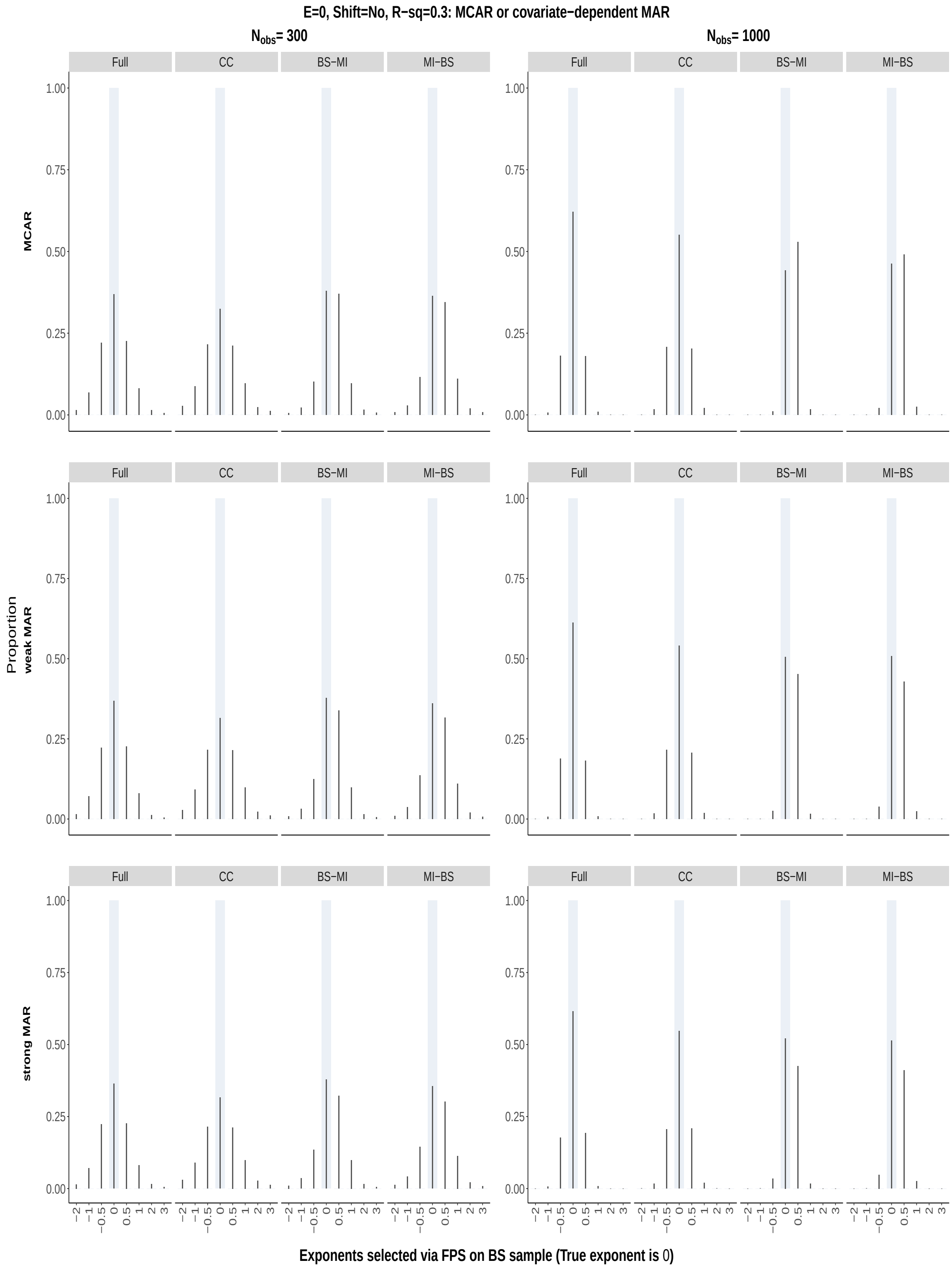

Figure S74: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

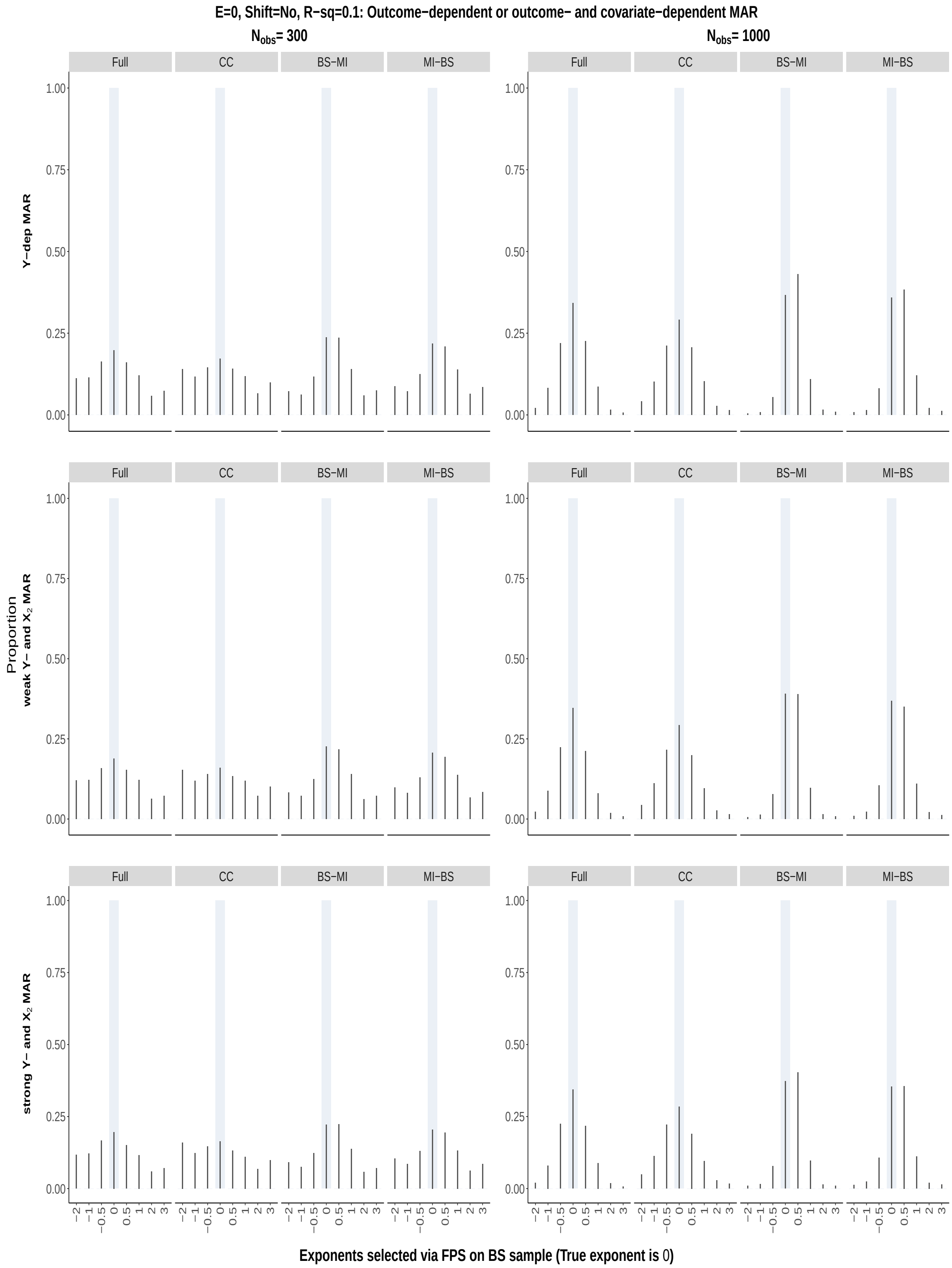

Figure S75: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

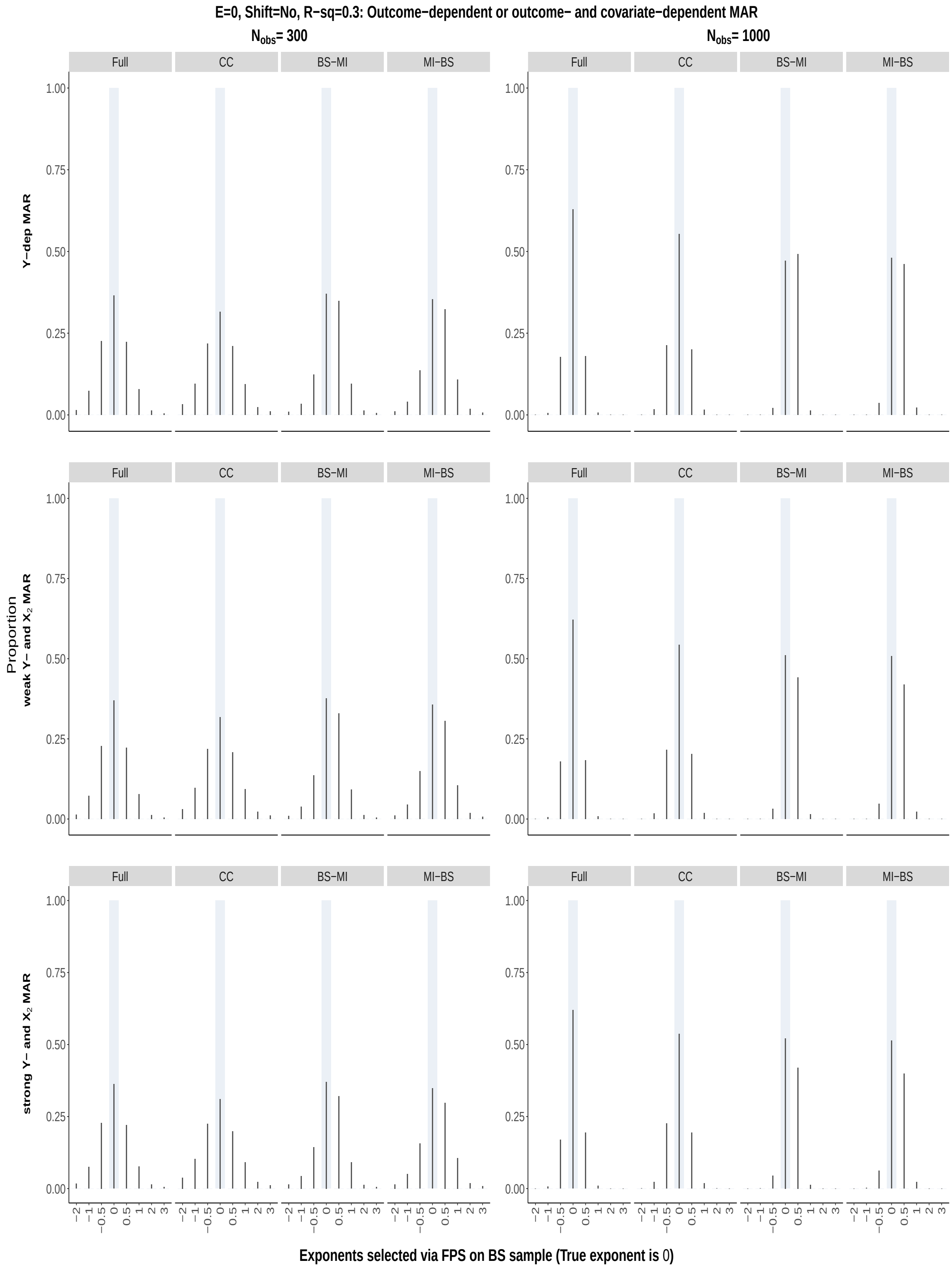

Figure S76: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

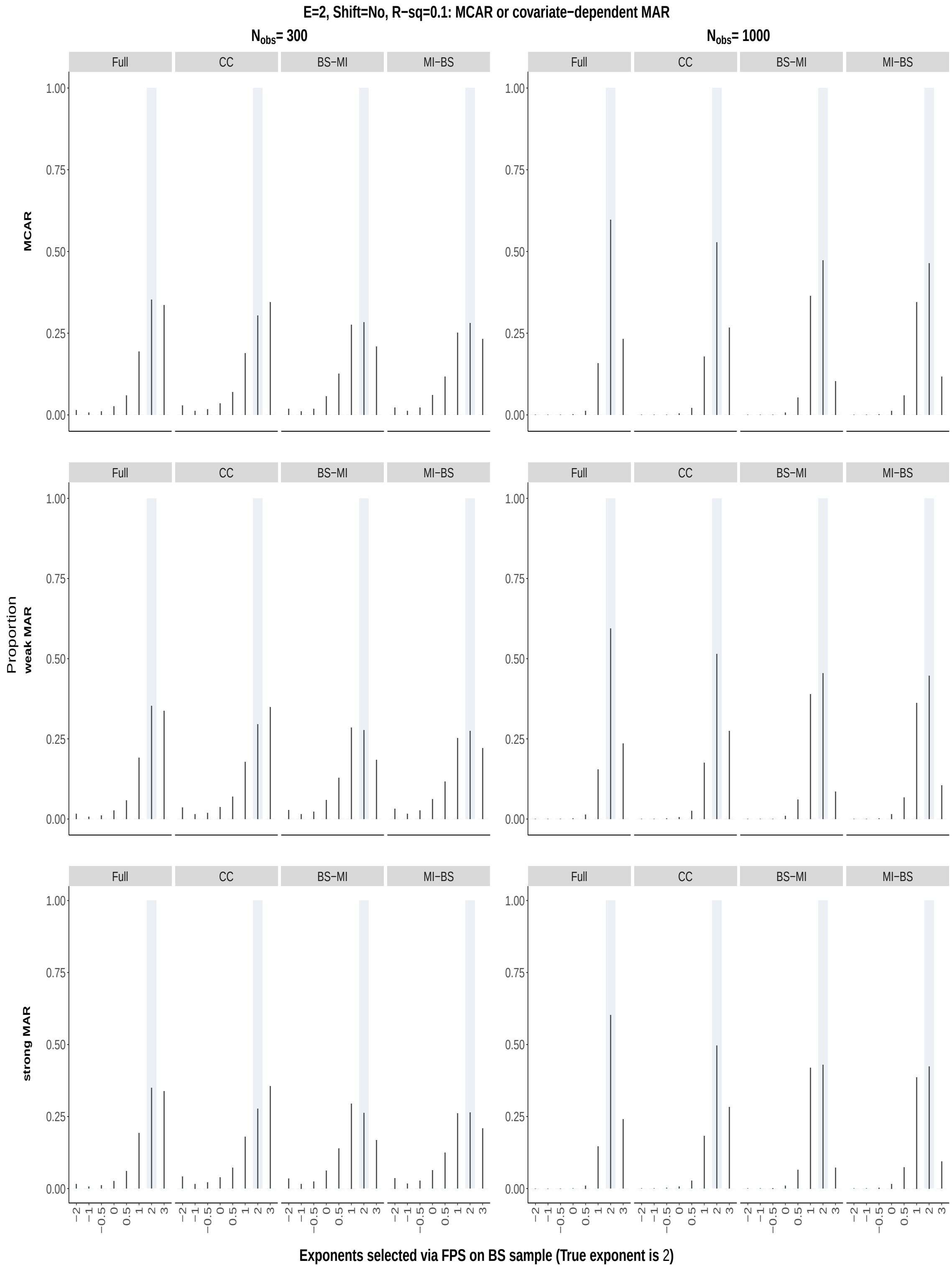

Figure S77: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

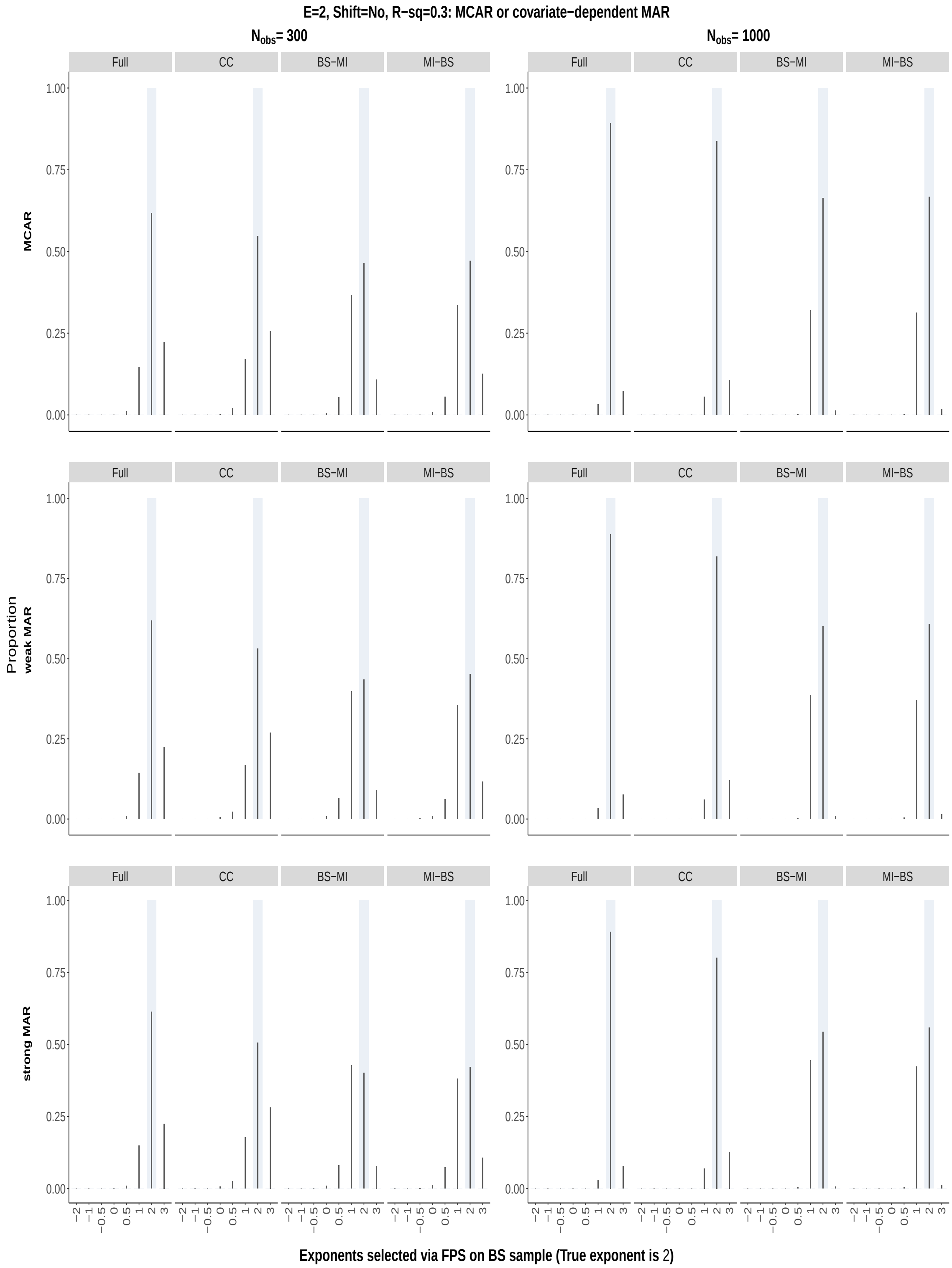

Figure S78: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

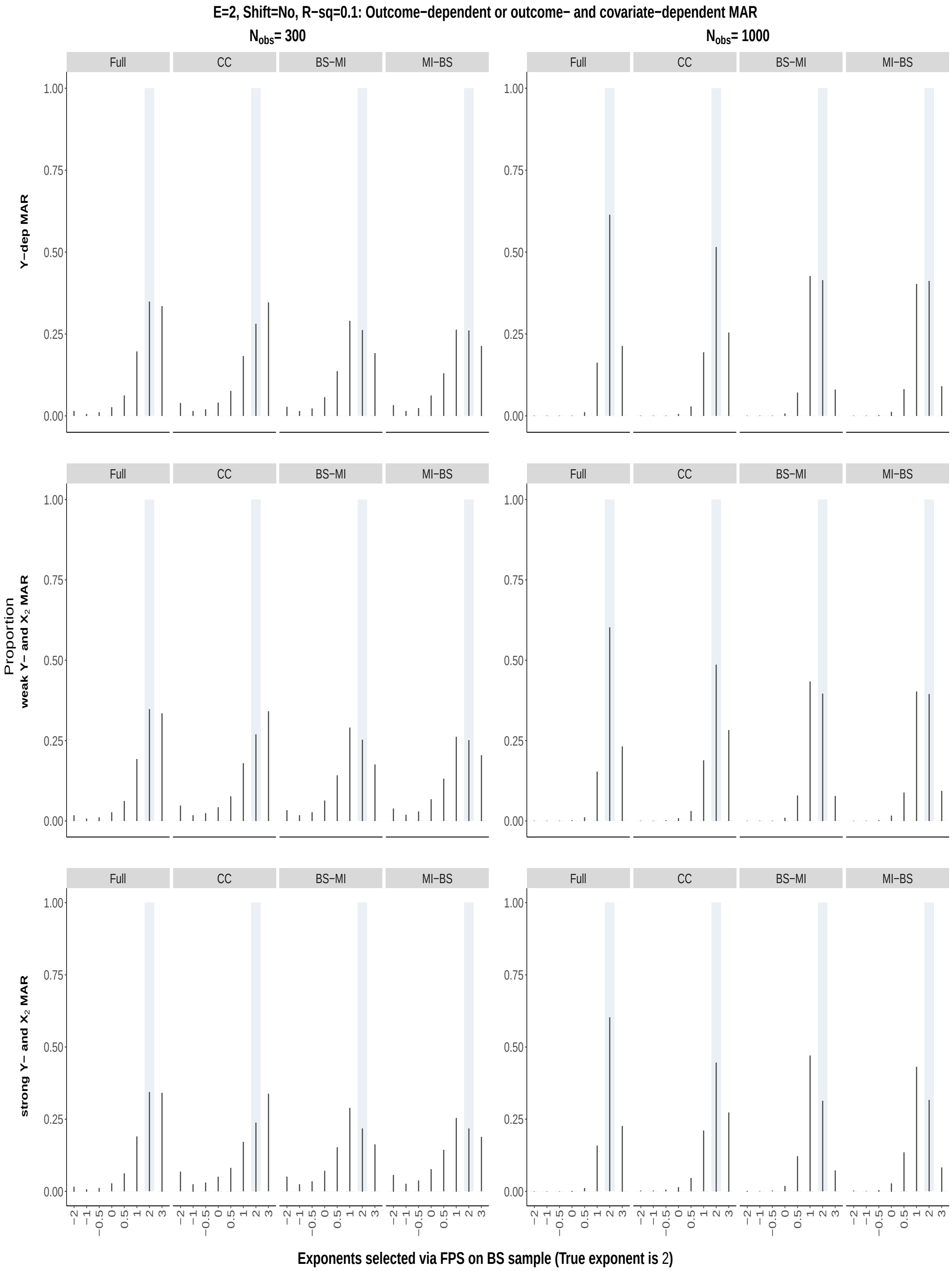

Figure S79: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

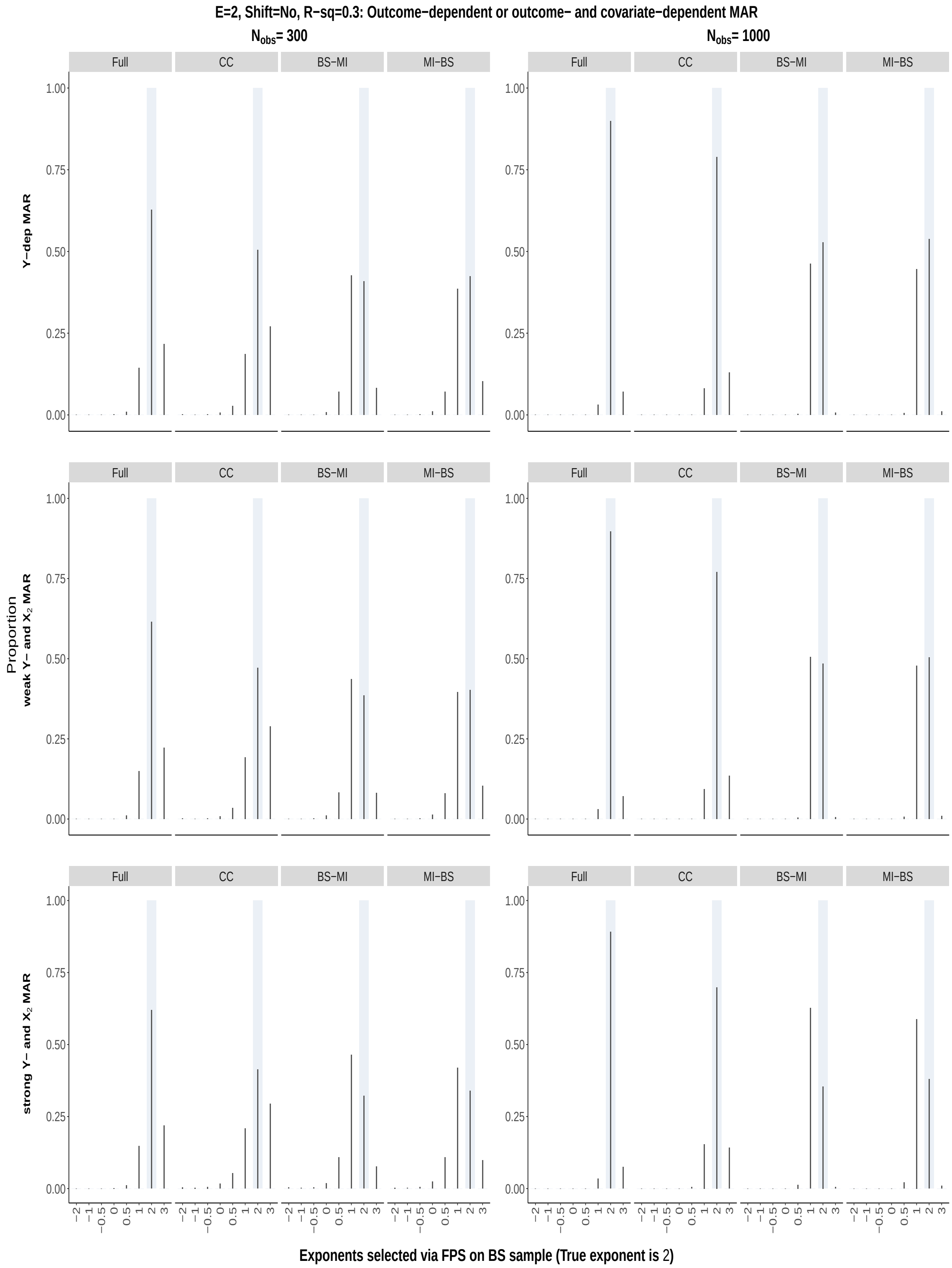

Figure S80: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

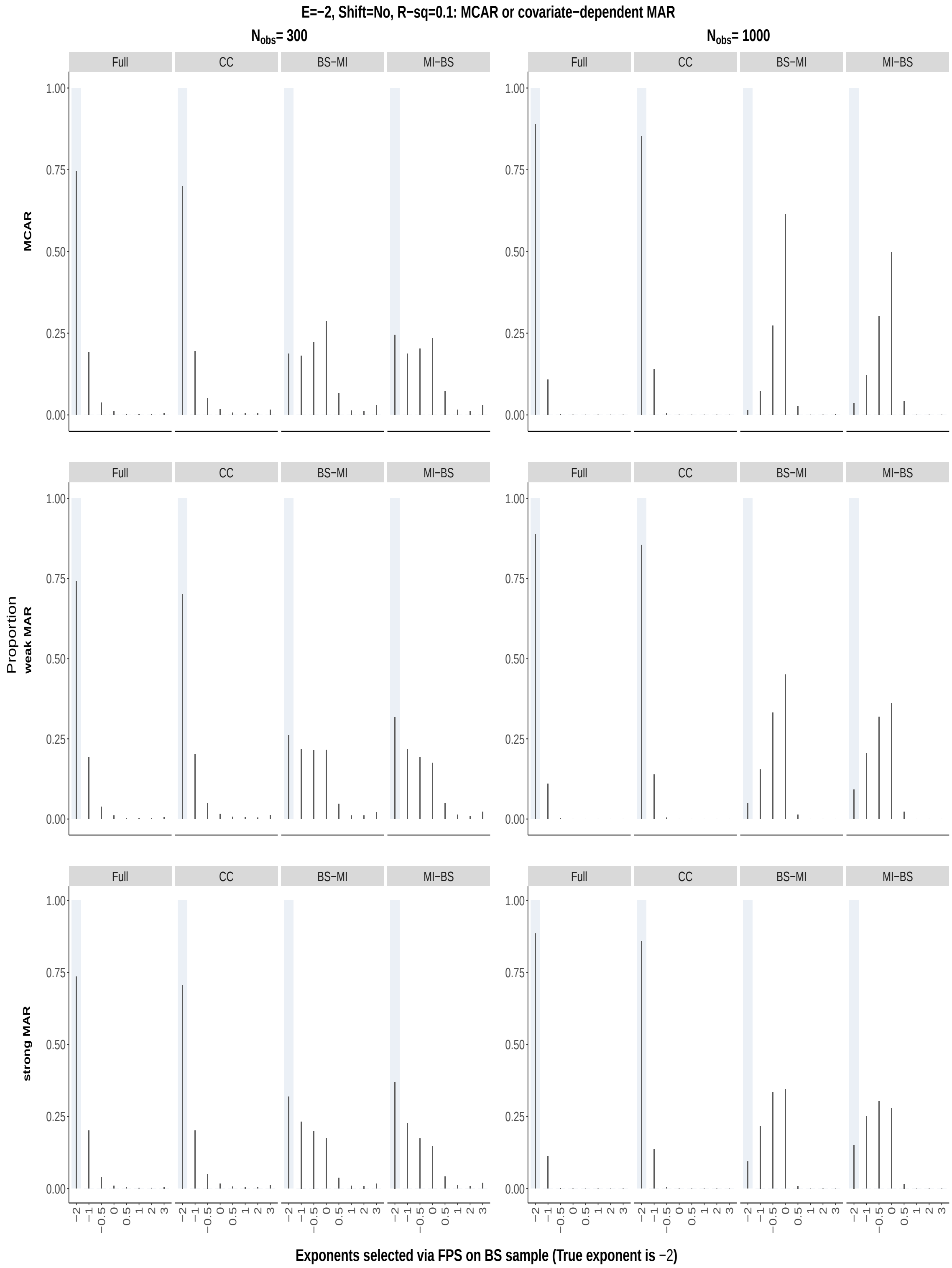

Figure S81: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

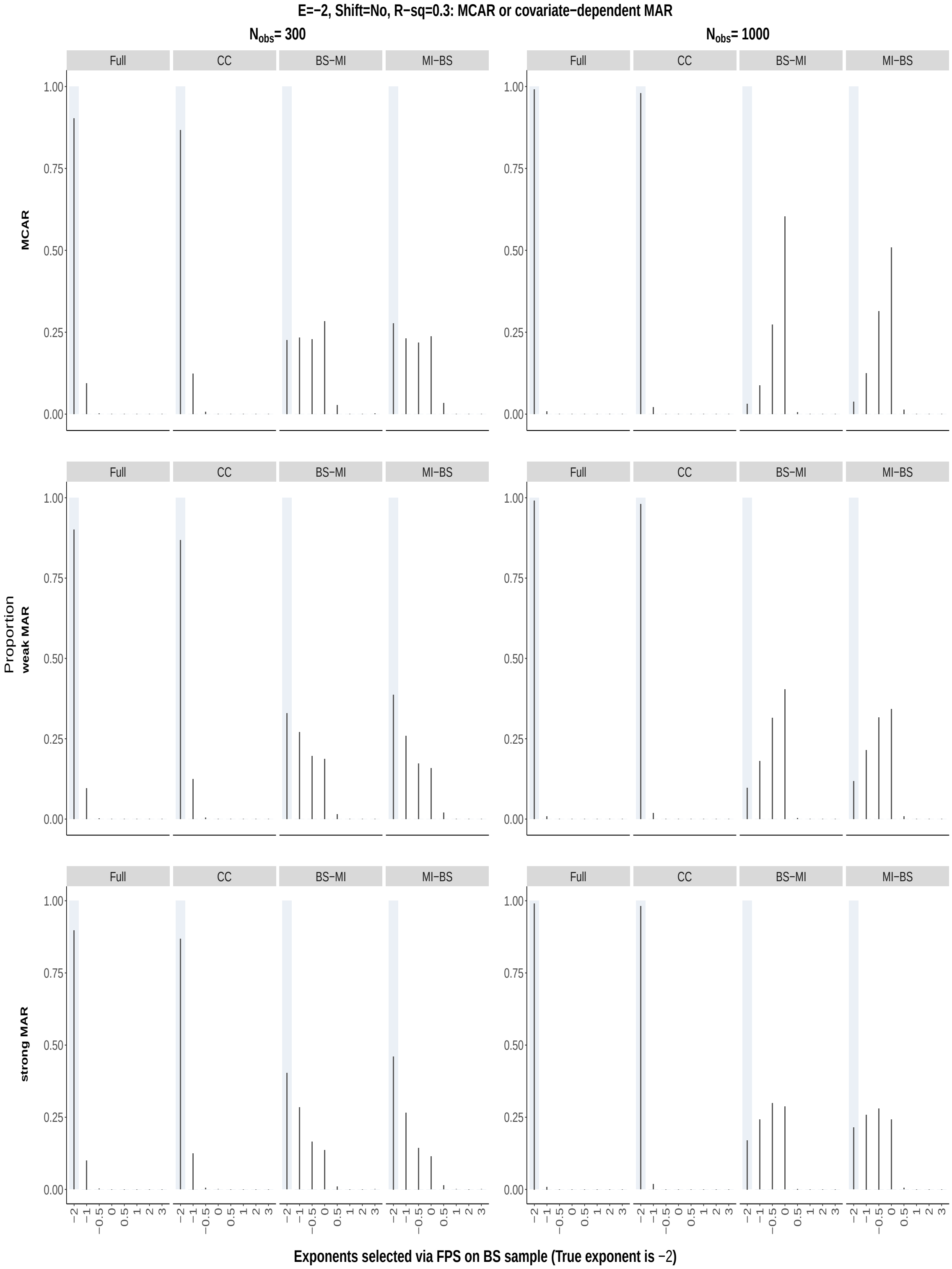

Figure S82: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

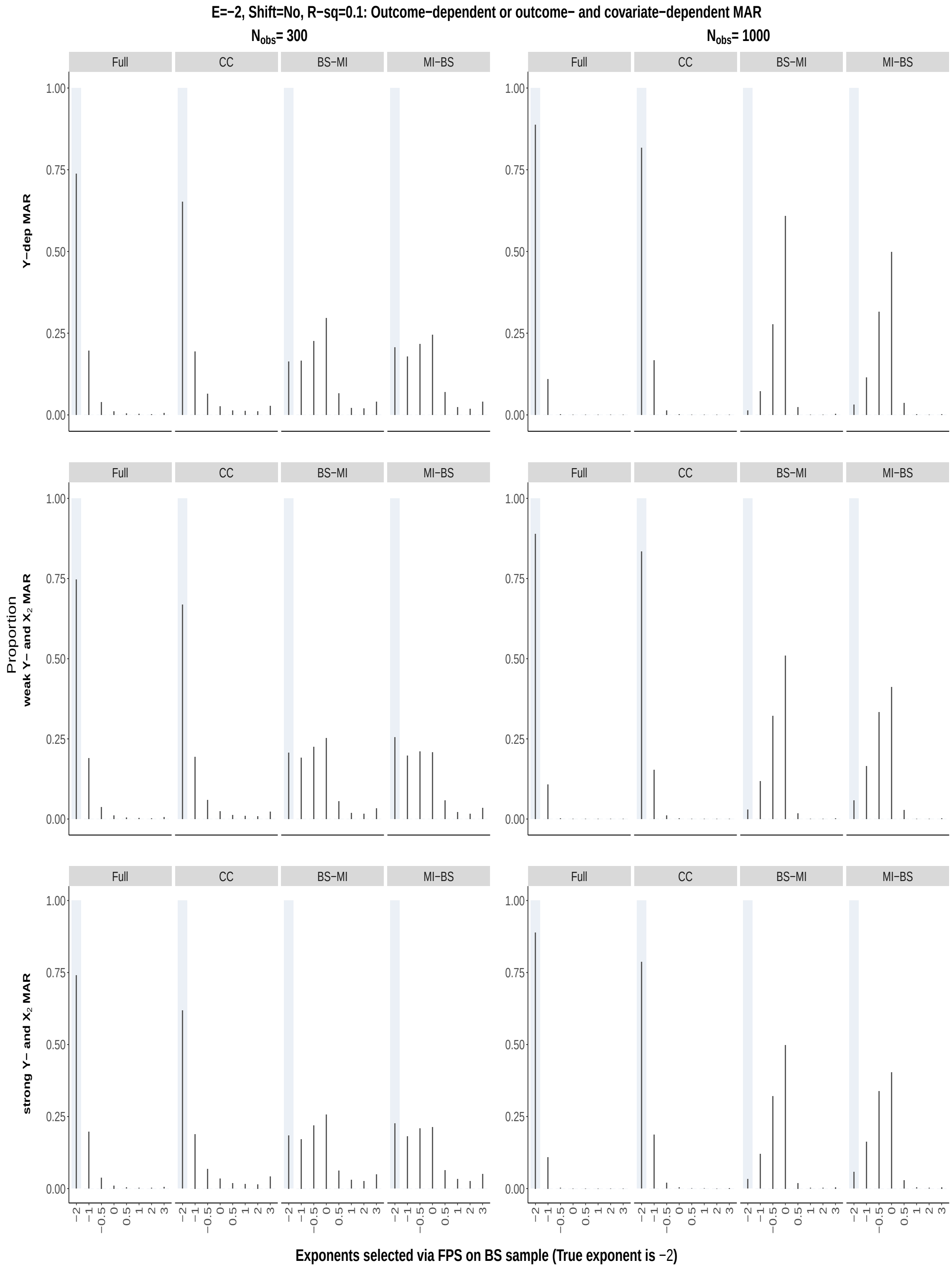

Figure S83: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

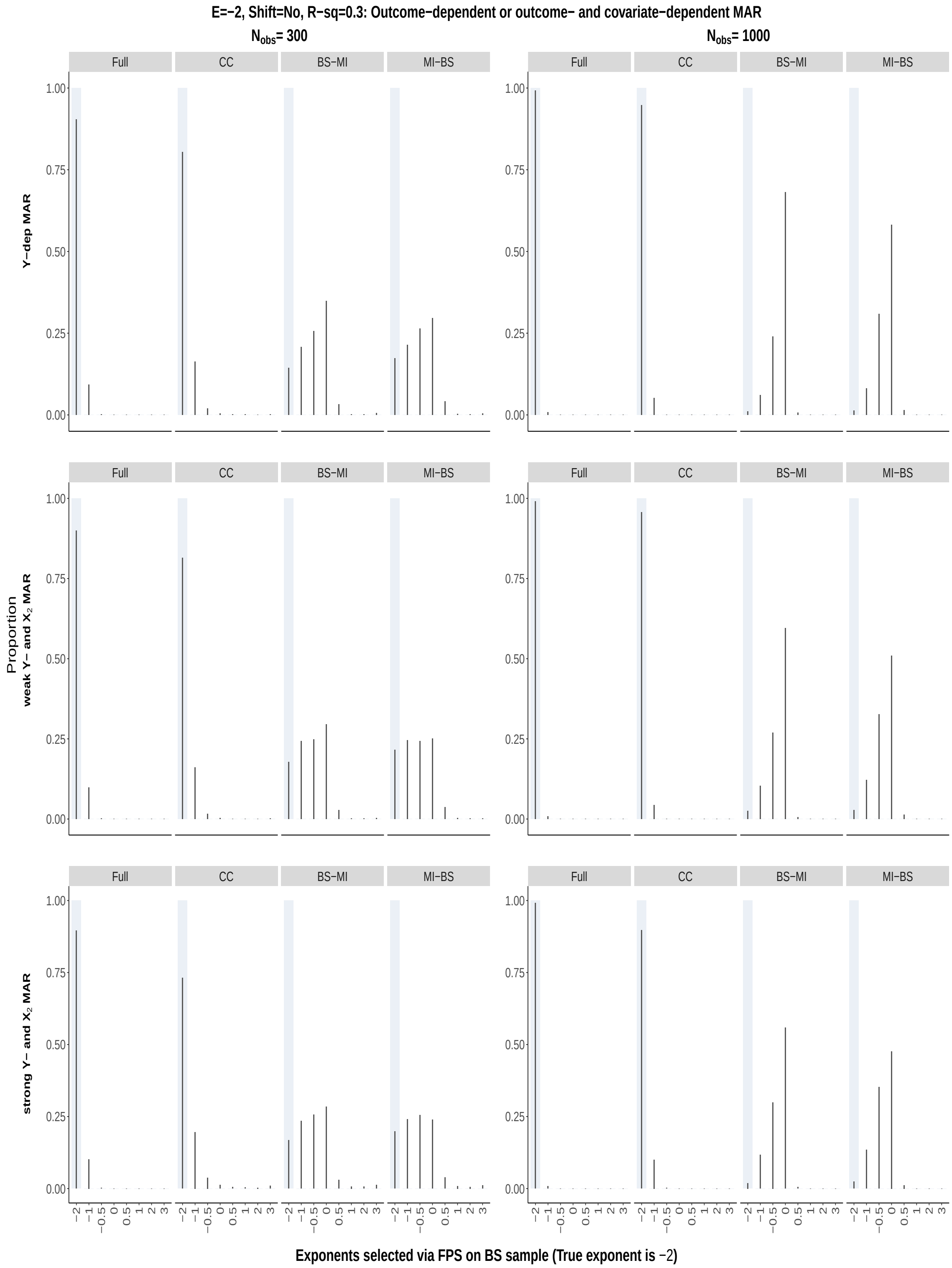

Figure S84: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.6 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

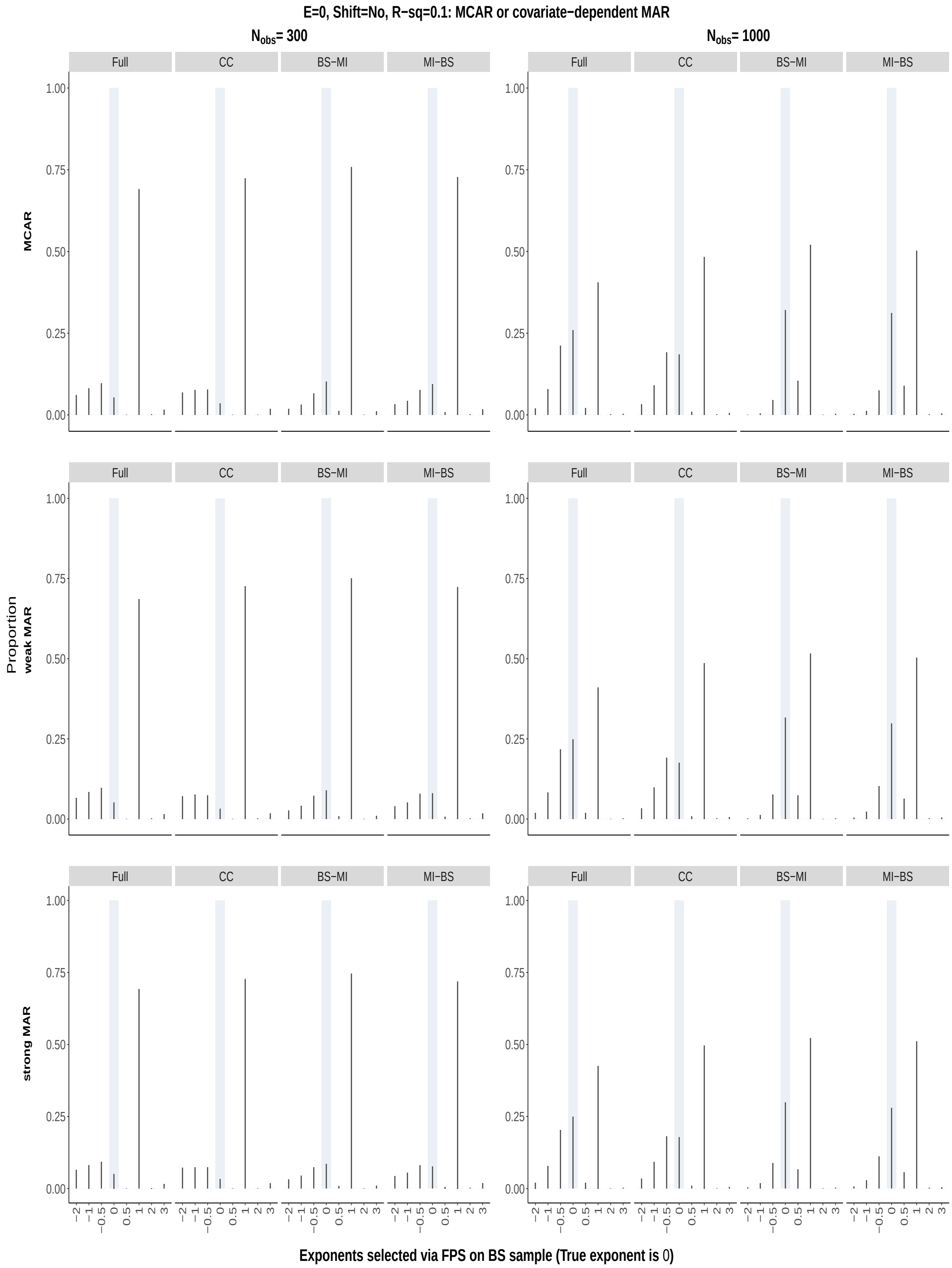

Figure S85: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

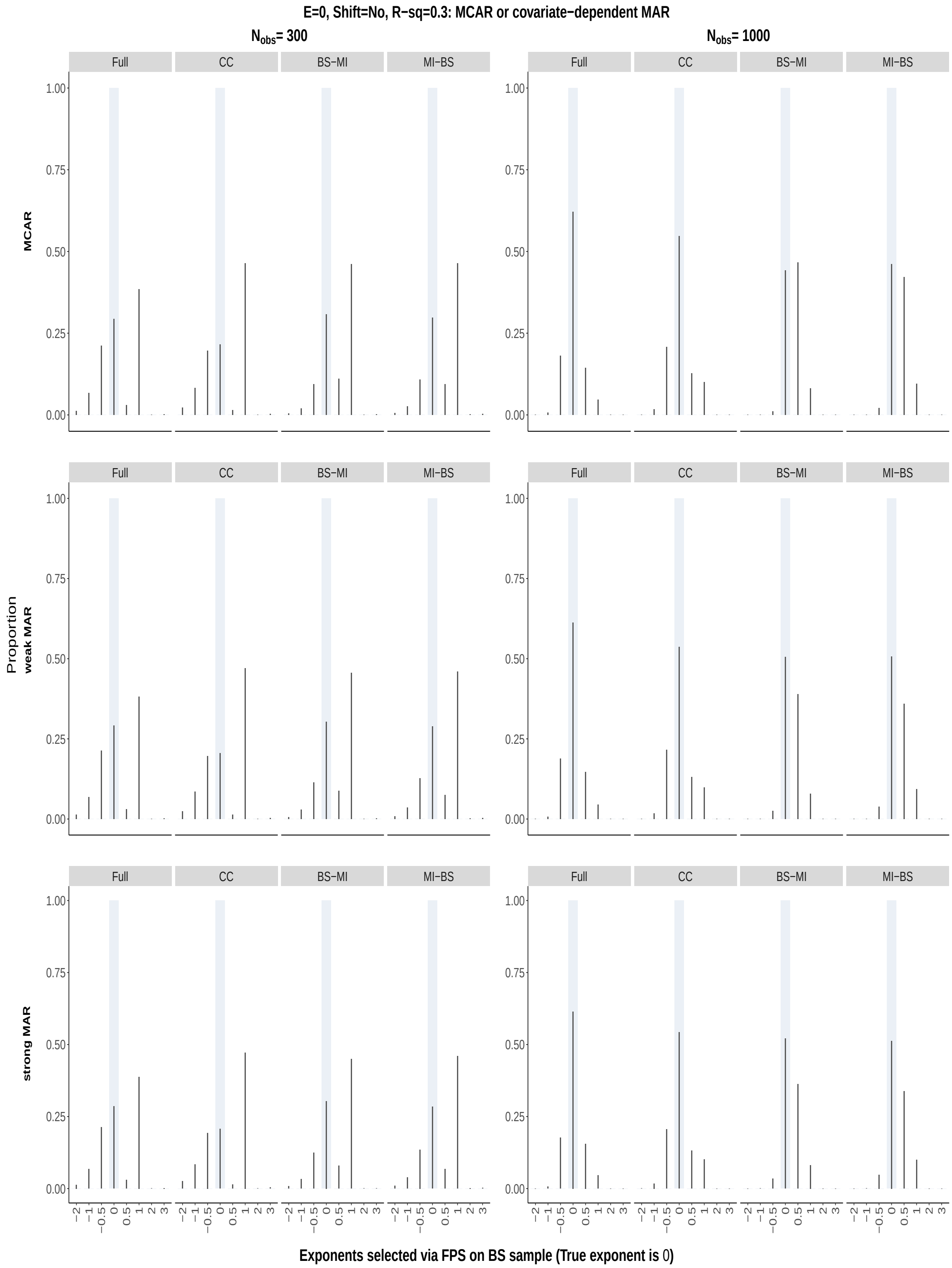

Figure S86: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

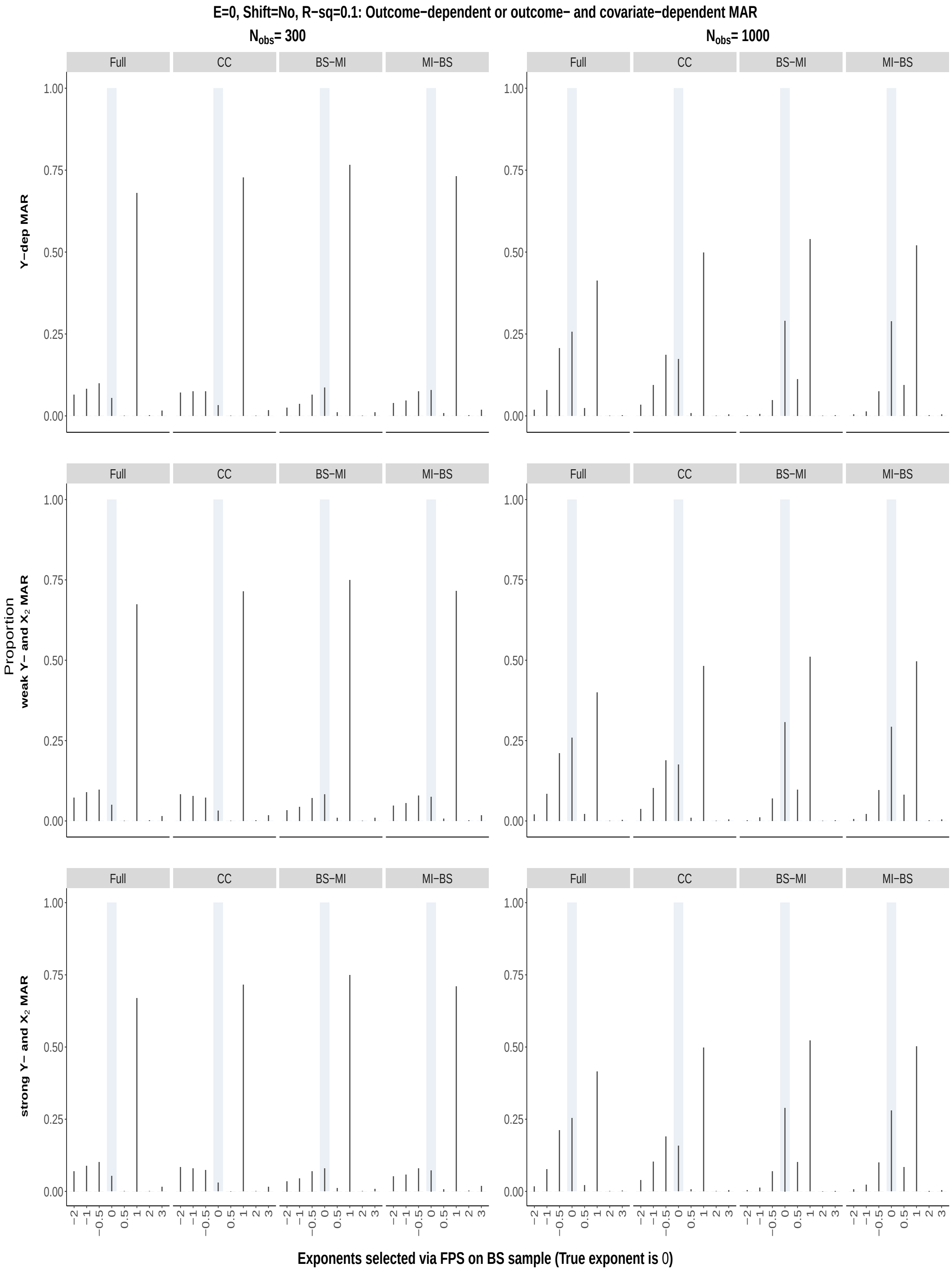

Figure S87: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

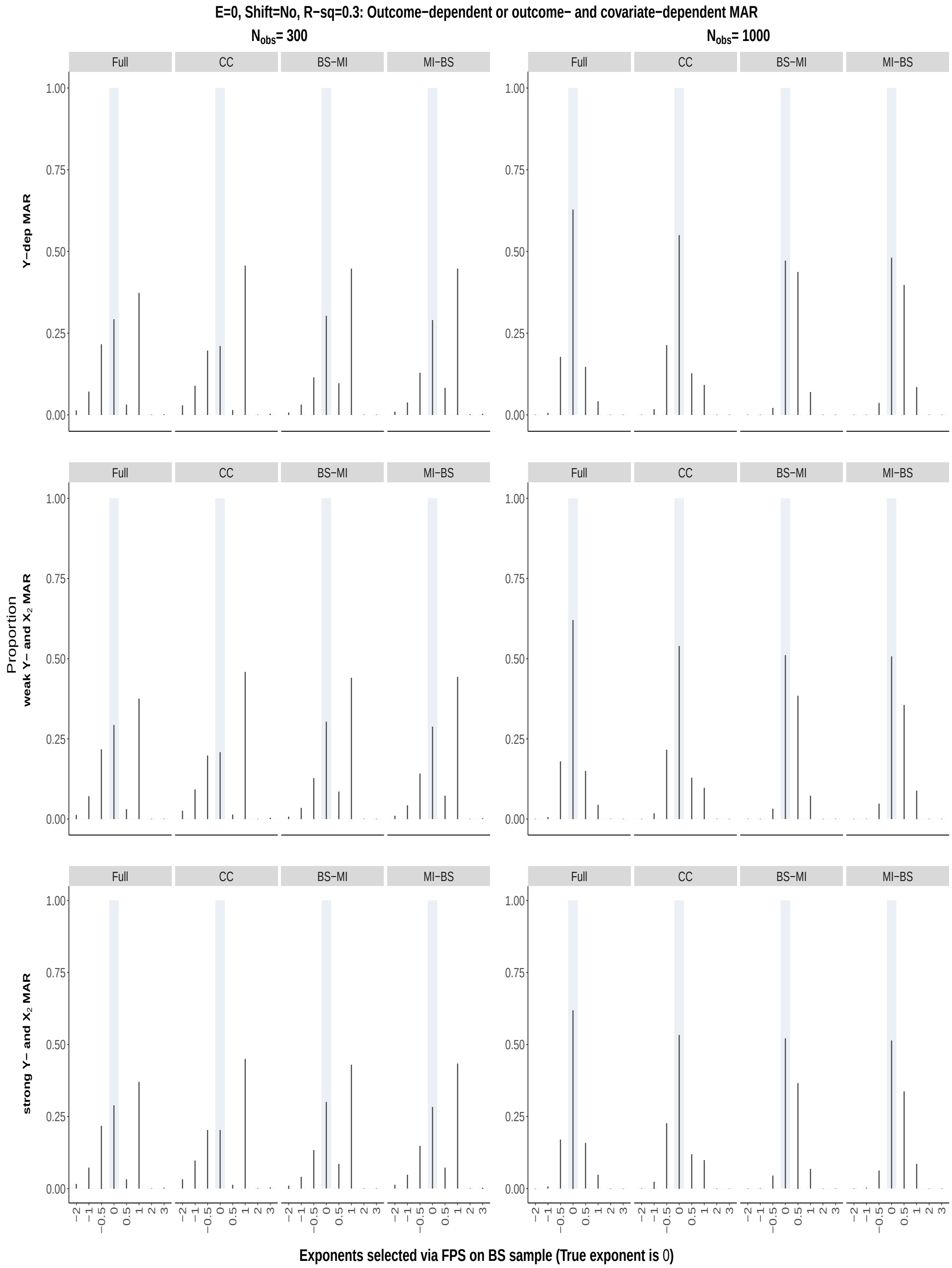

Figure S88: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

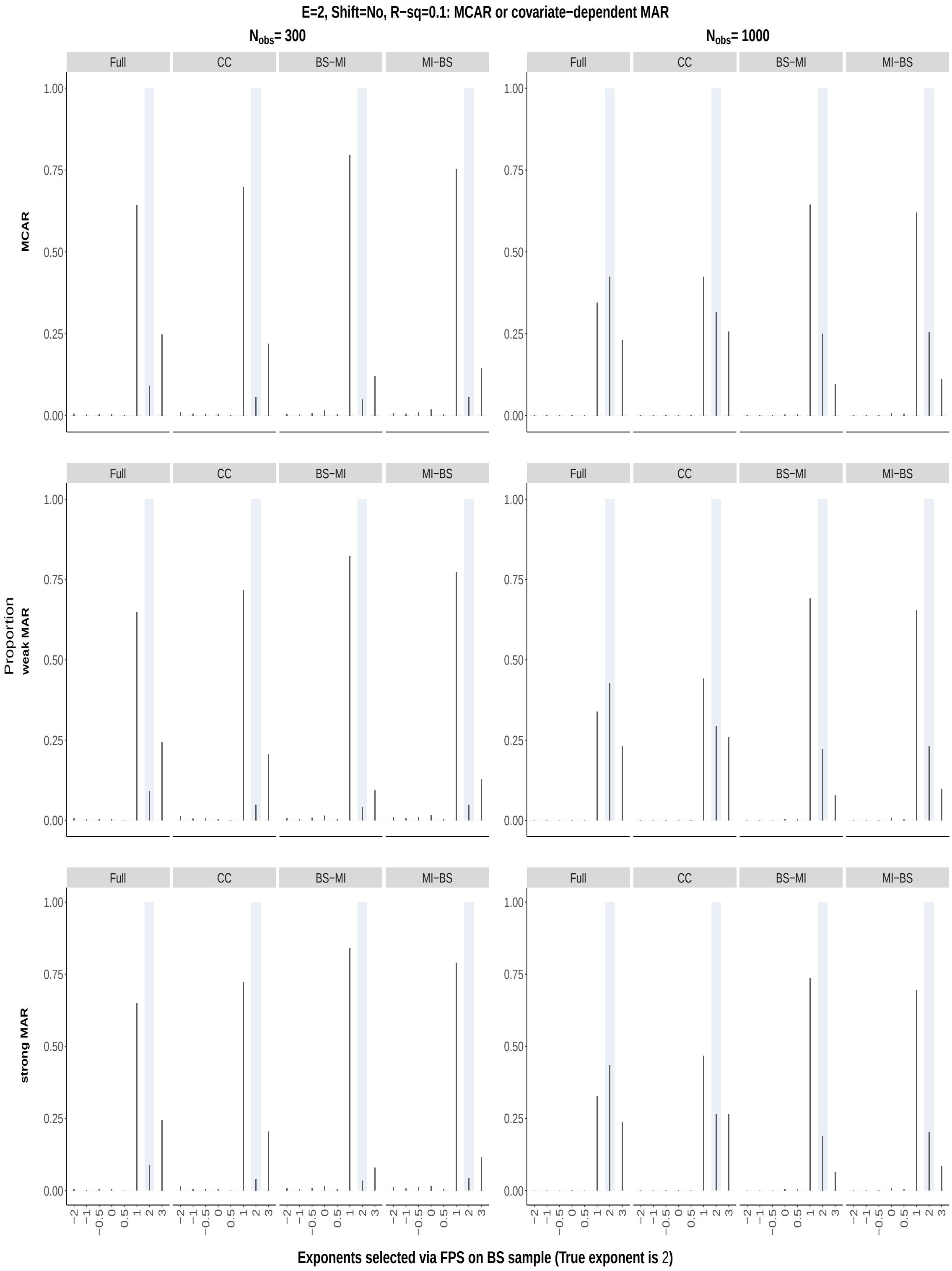

Figure S89: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

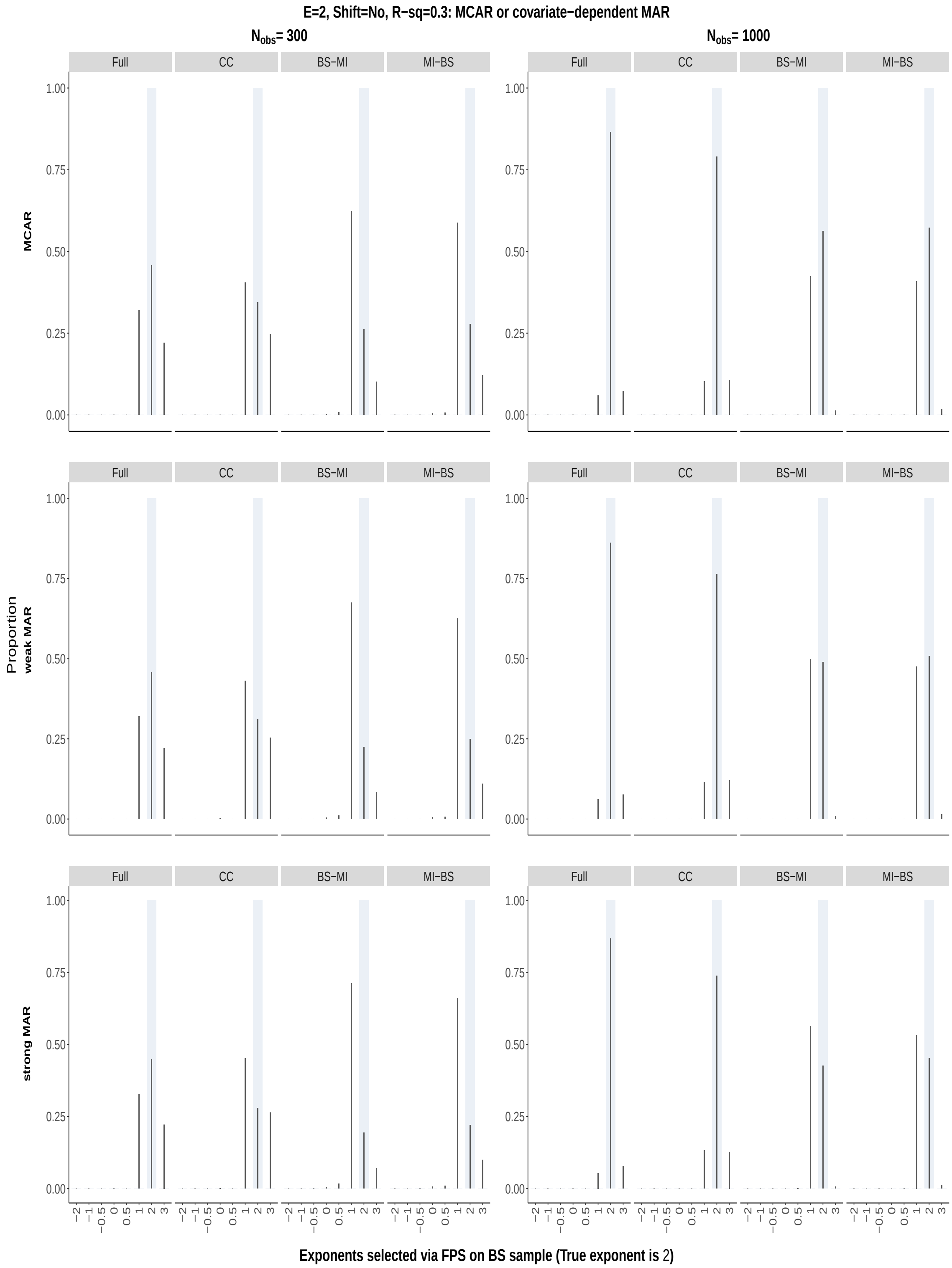

Figure S90: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

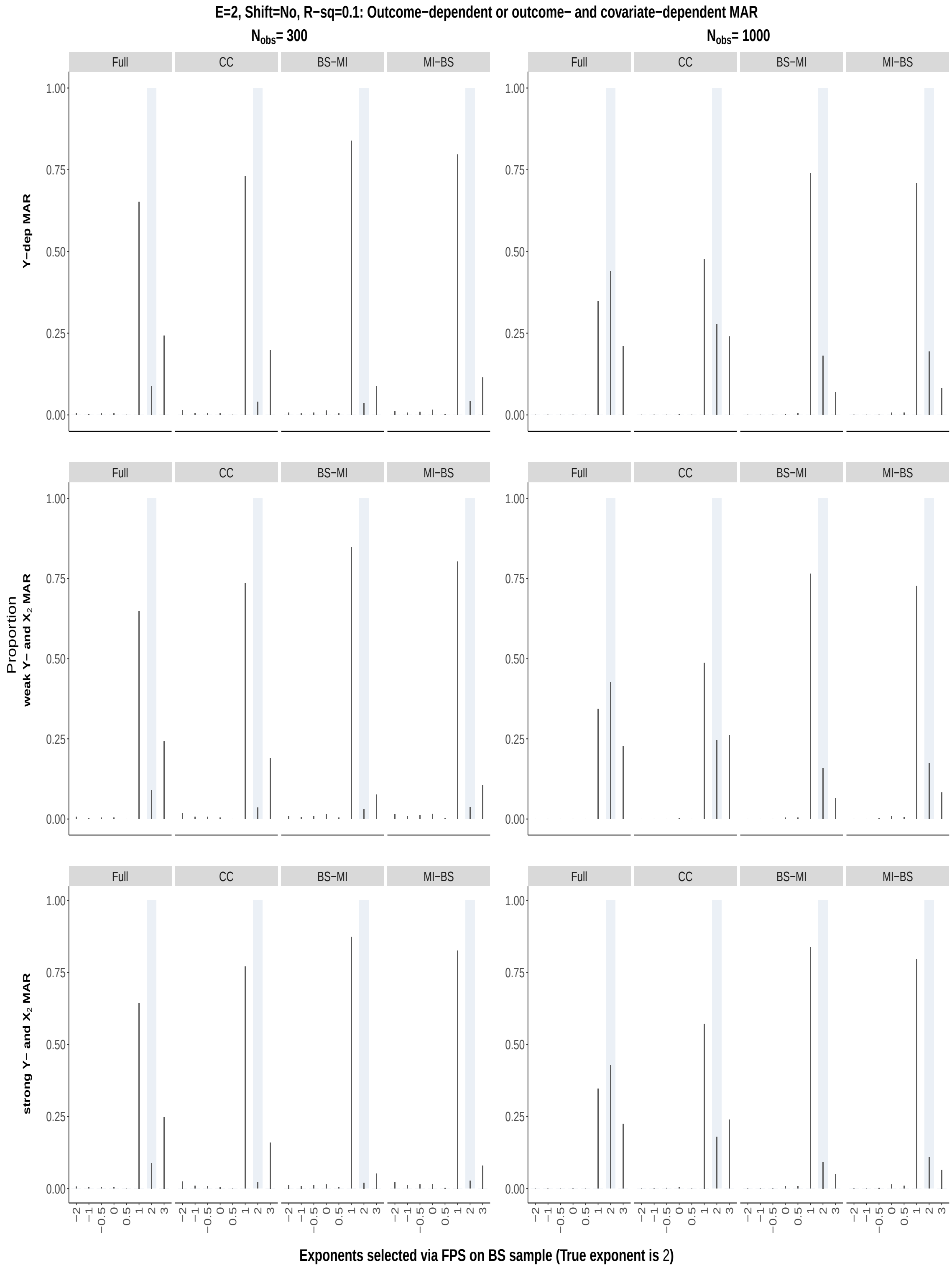

Figure S91: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

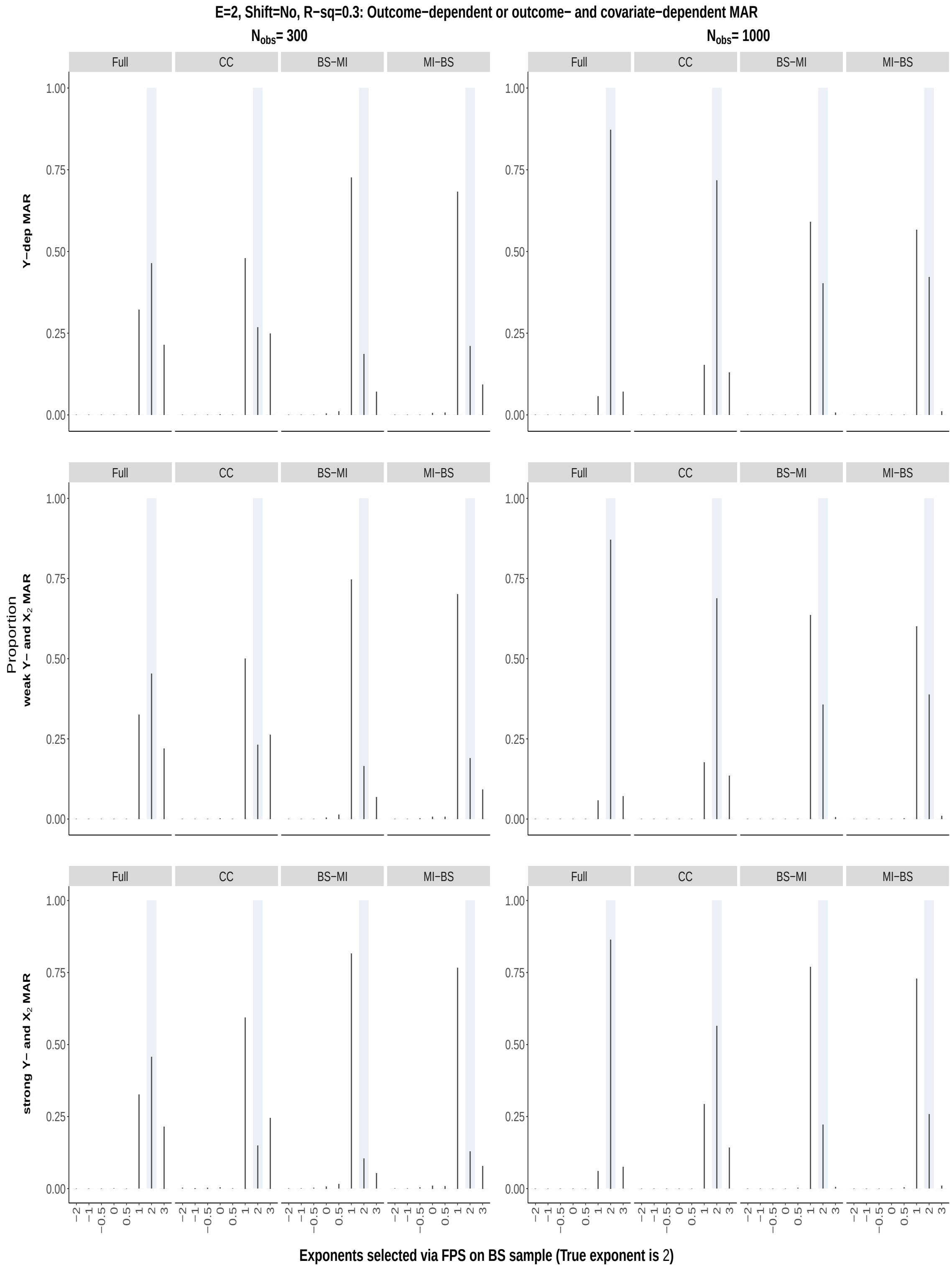

Figure S92: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

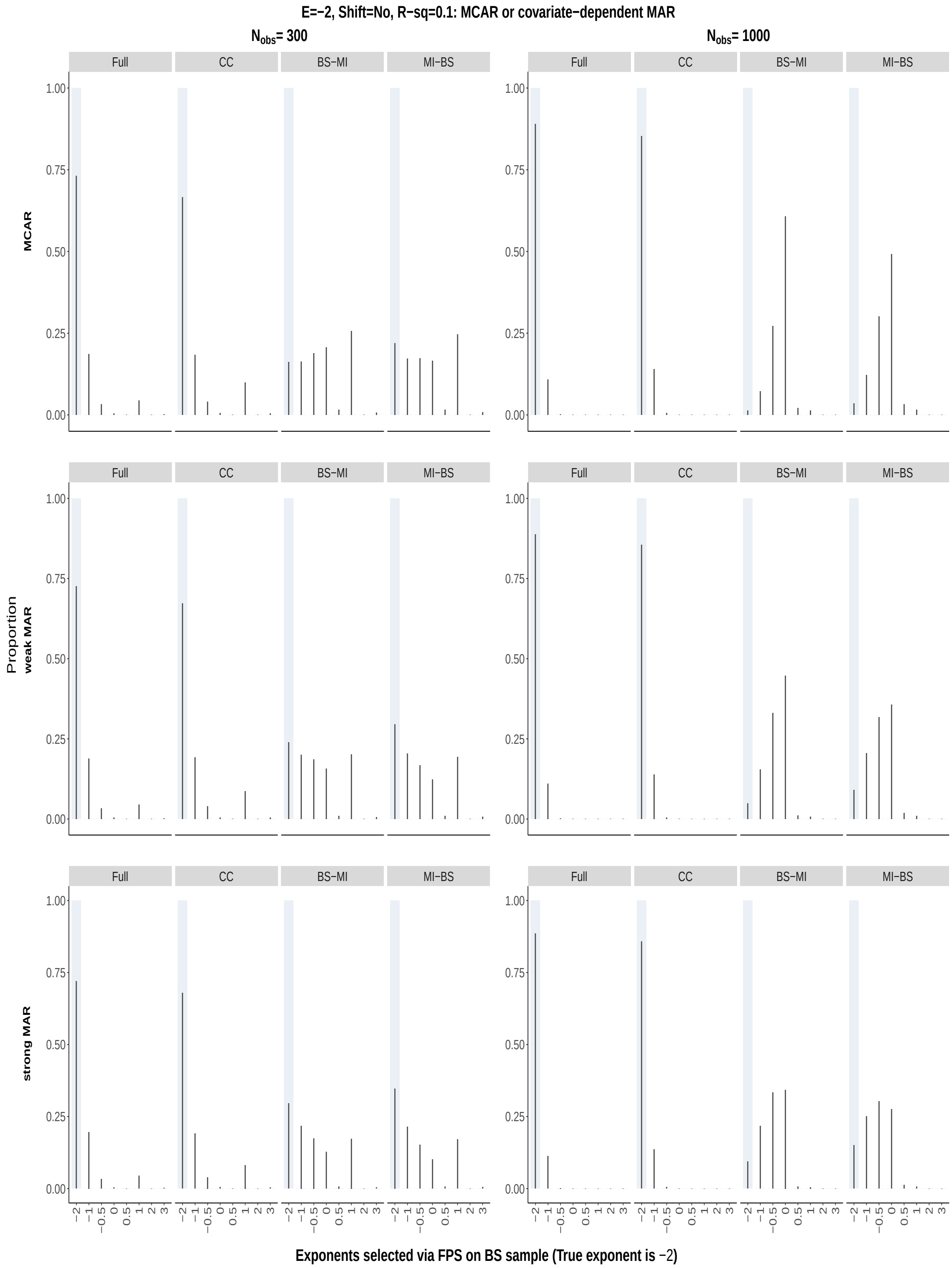

Figure S93: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

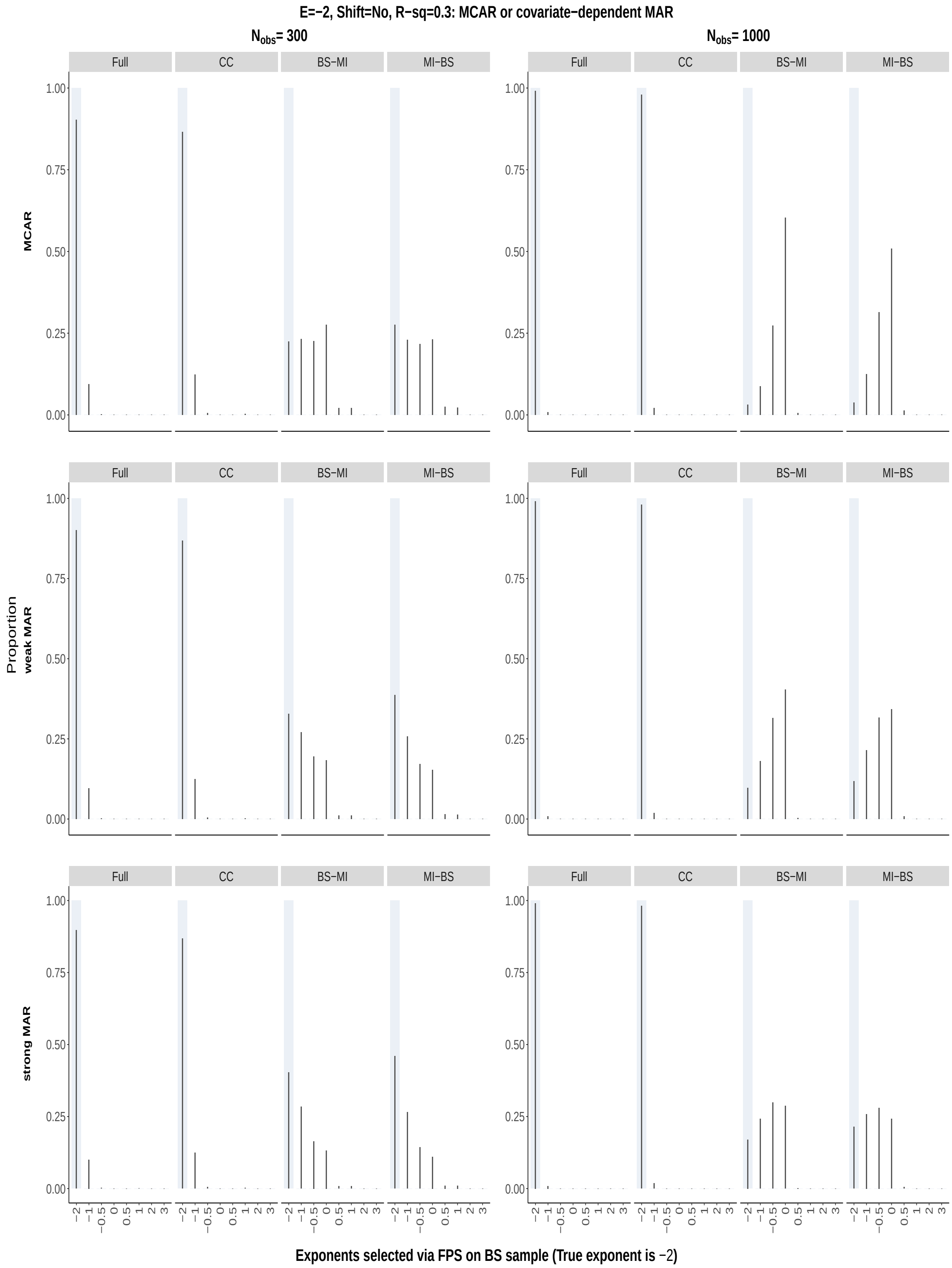

Figure S94: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

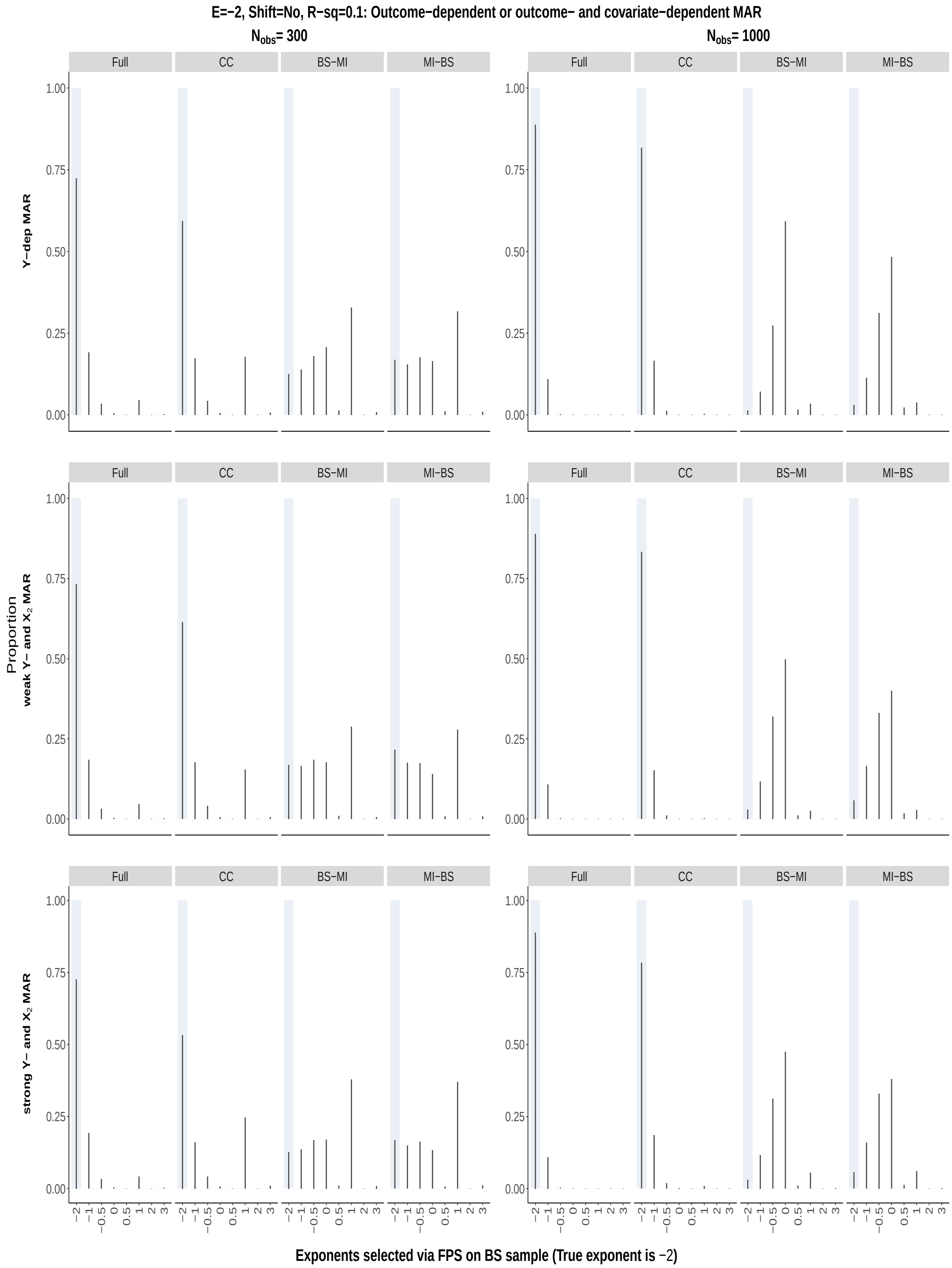

Figure S95: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

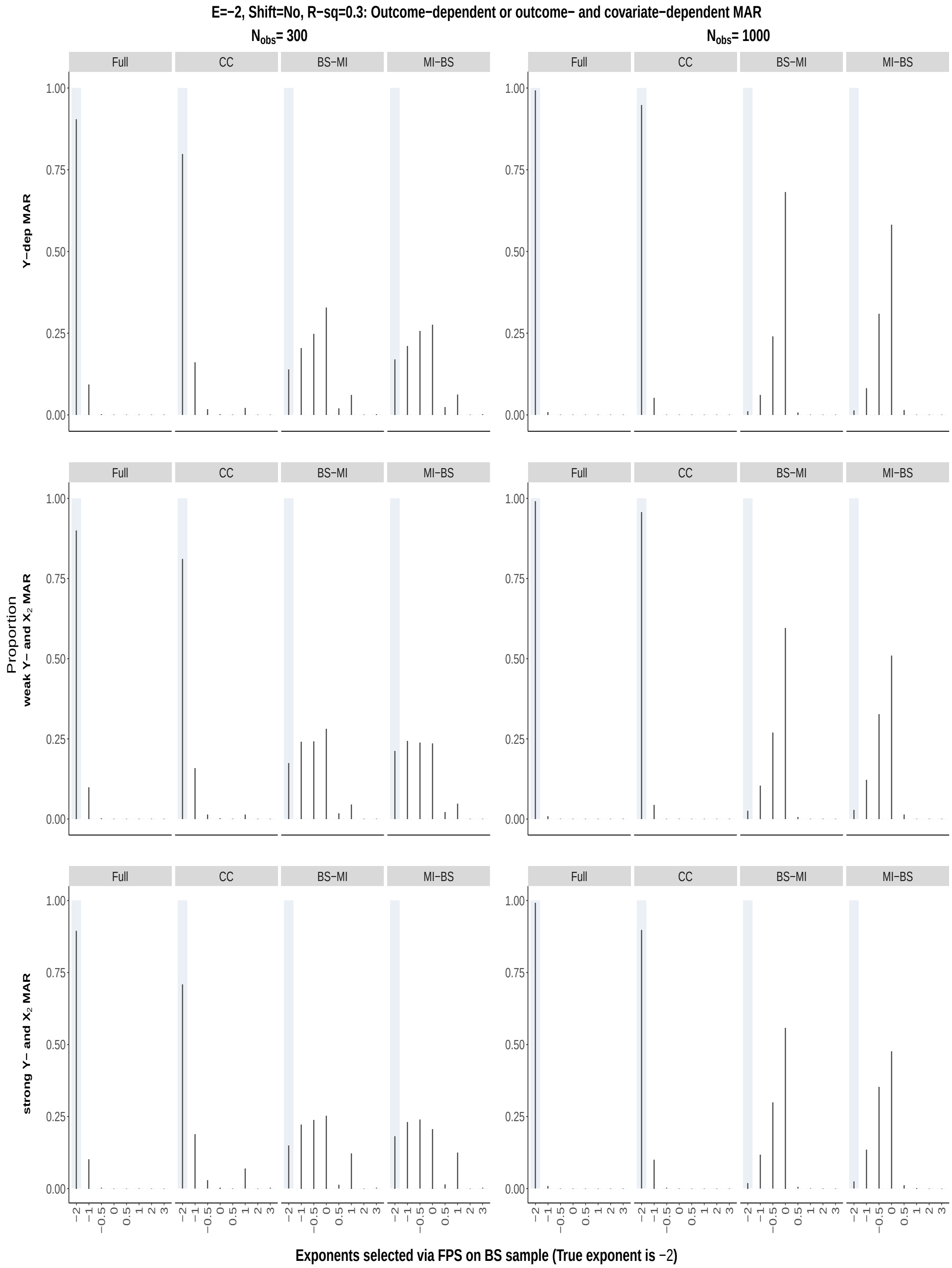

Figure S96: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.7 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\alpha_E = 1$ and an origin-shift has been applied

True exponent is 0

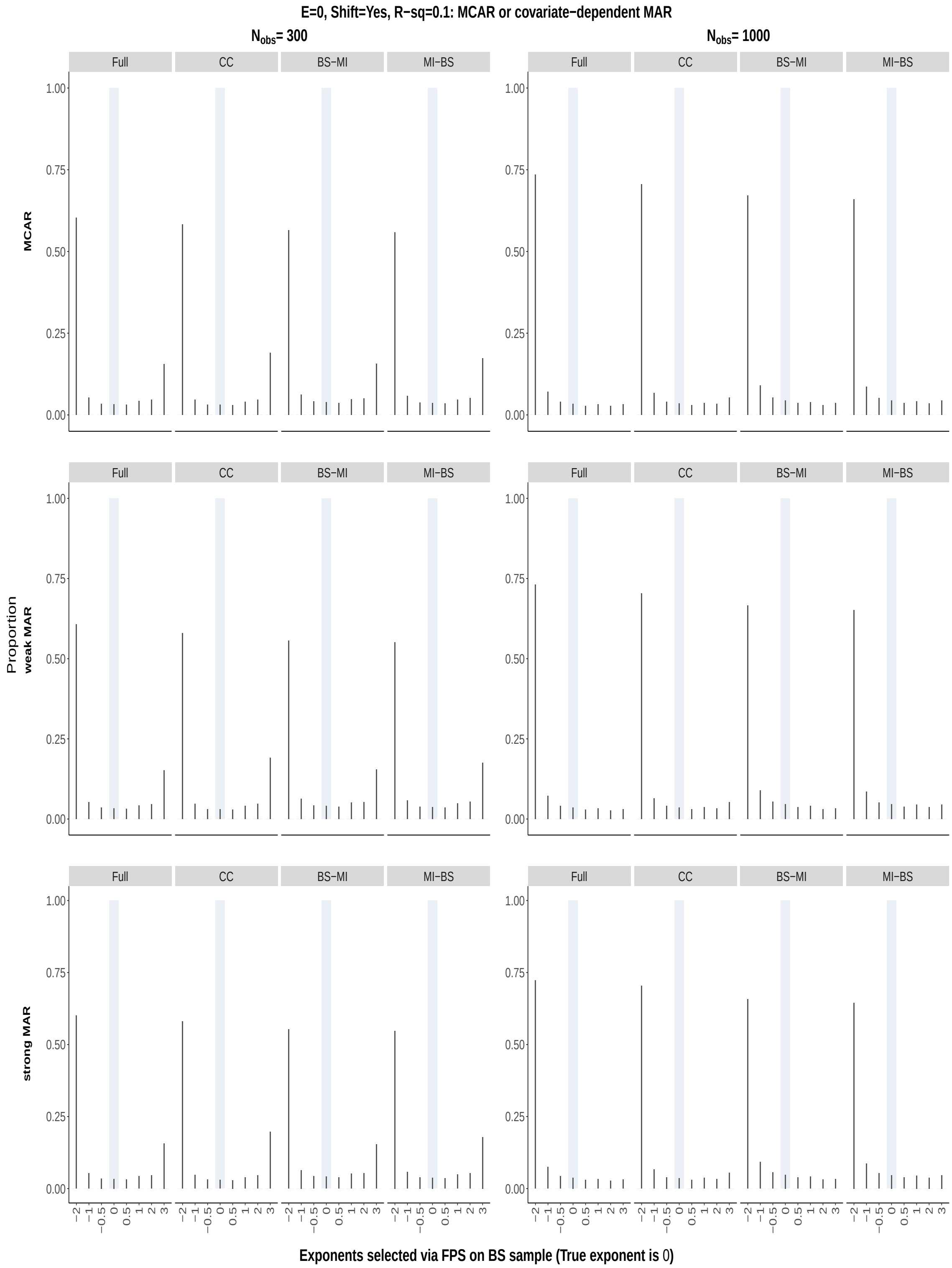

Figure S97: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

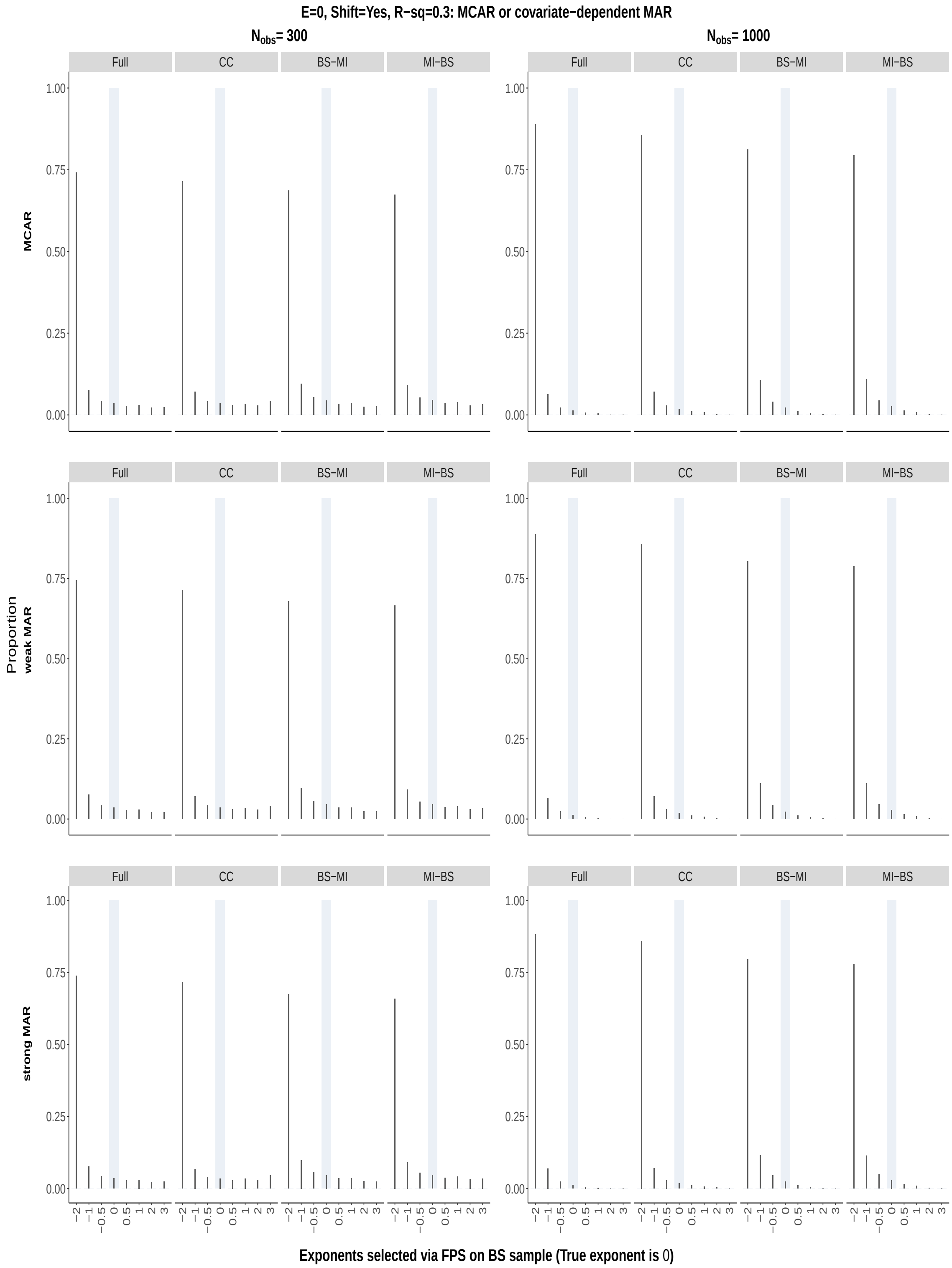

Figure S98: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

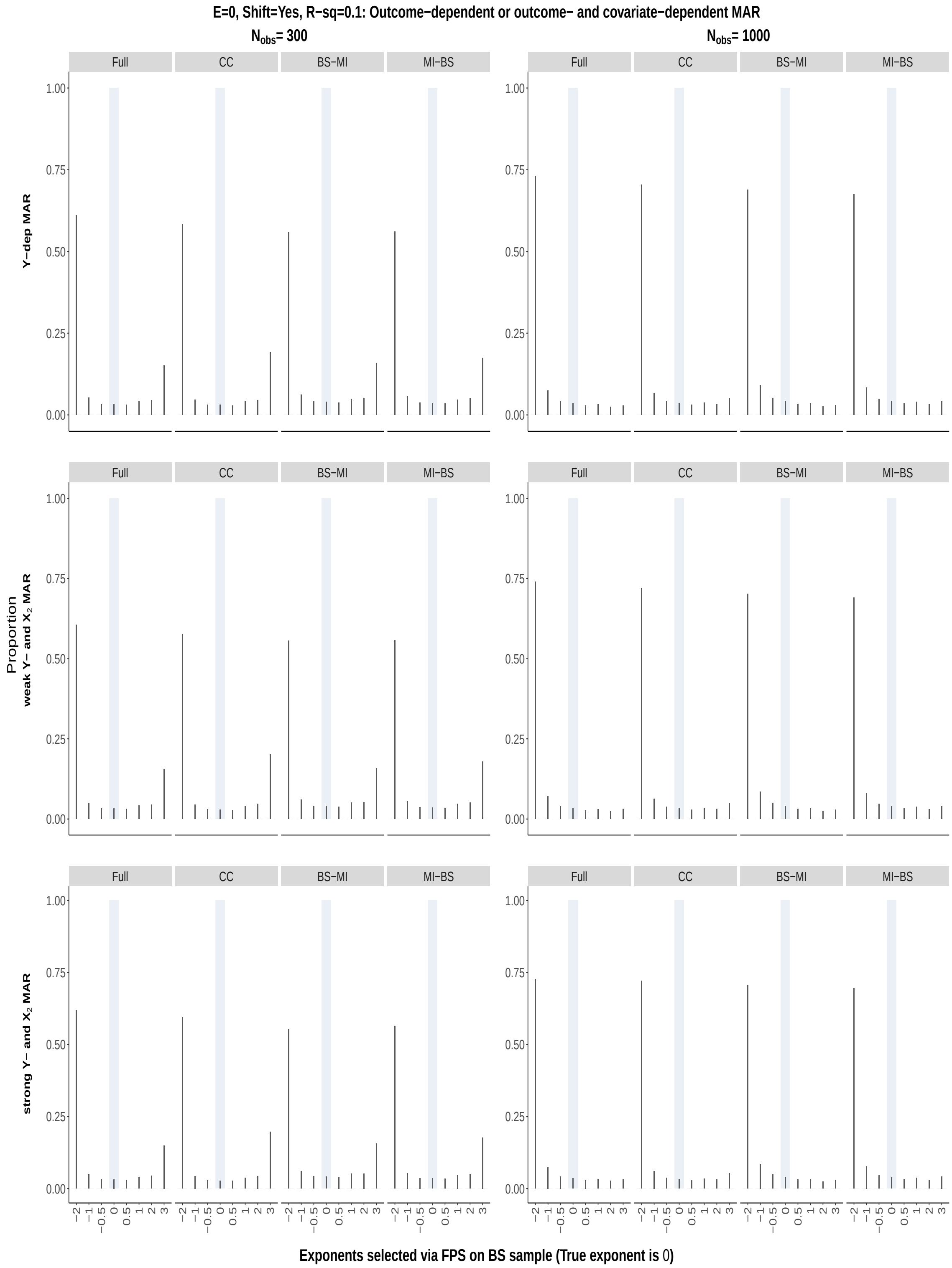

Figure S99: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

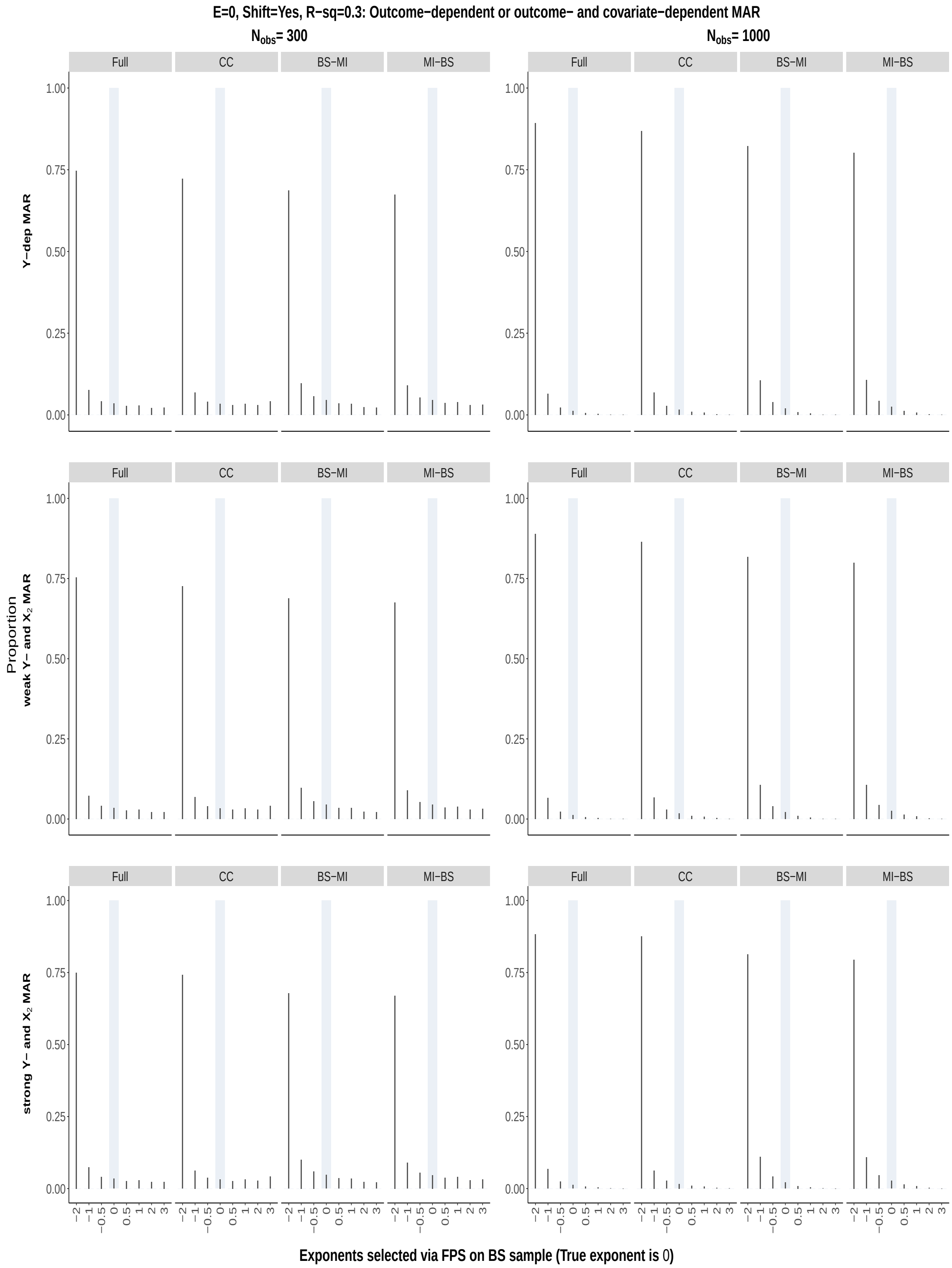

Figure S100: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

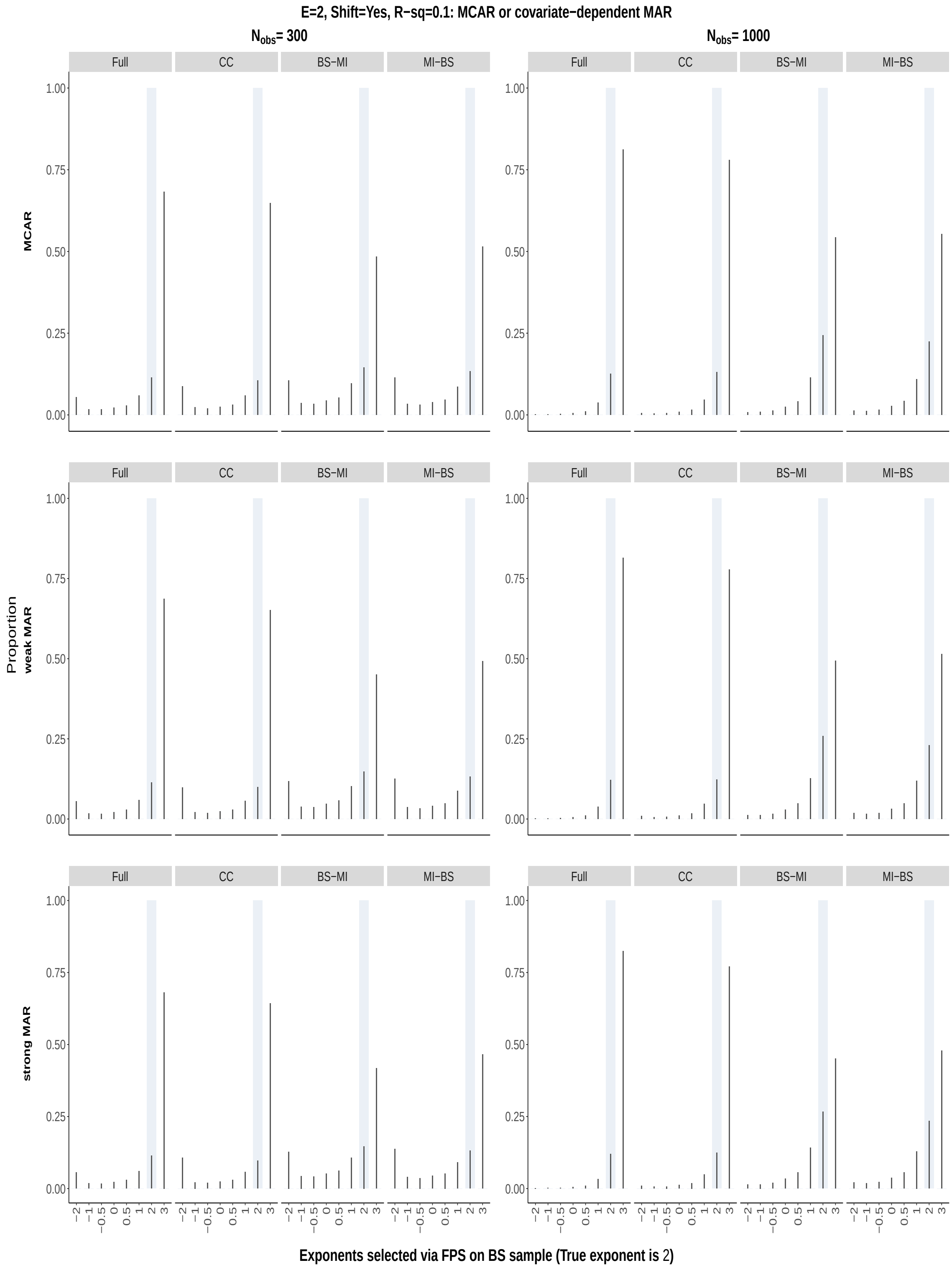

Figure S101: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

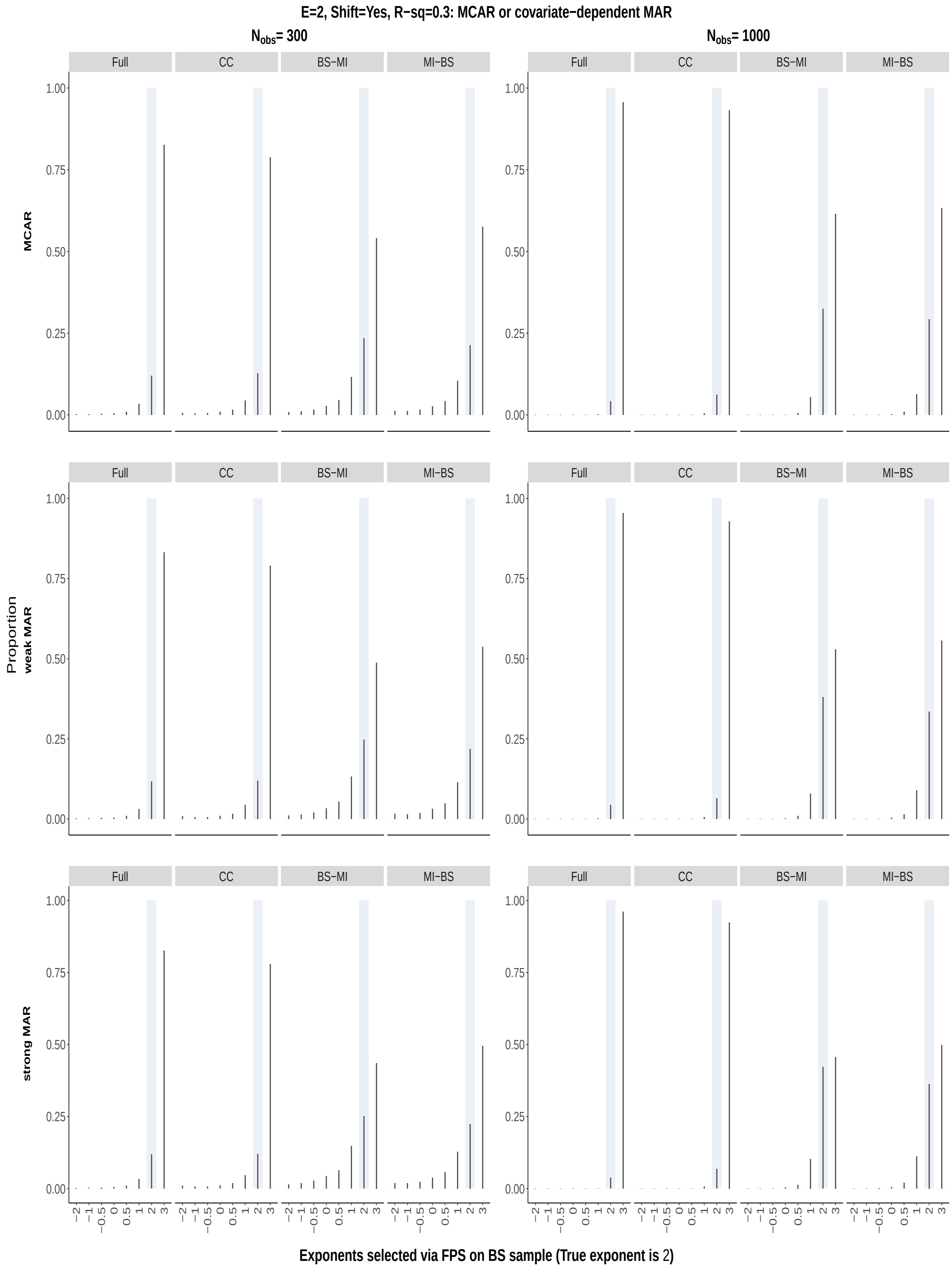

Figure S102: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

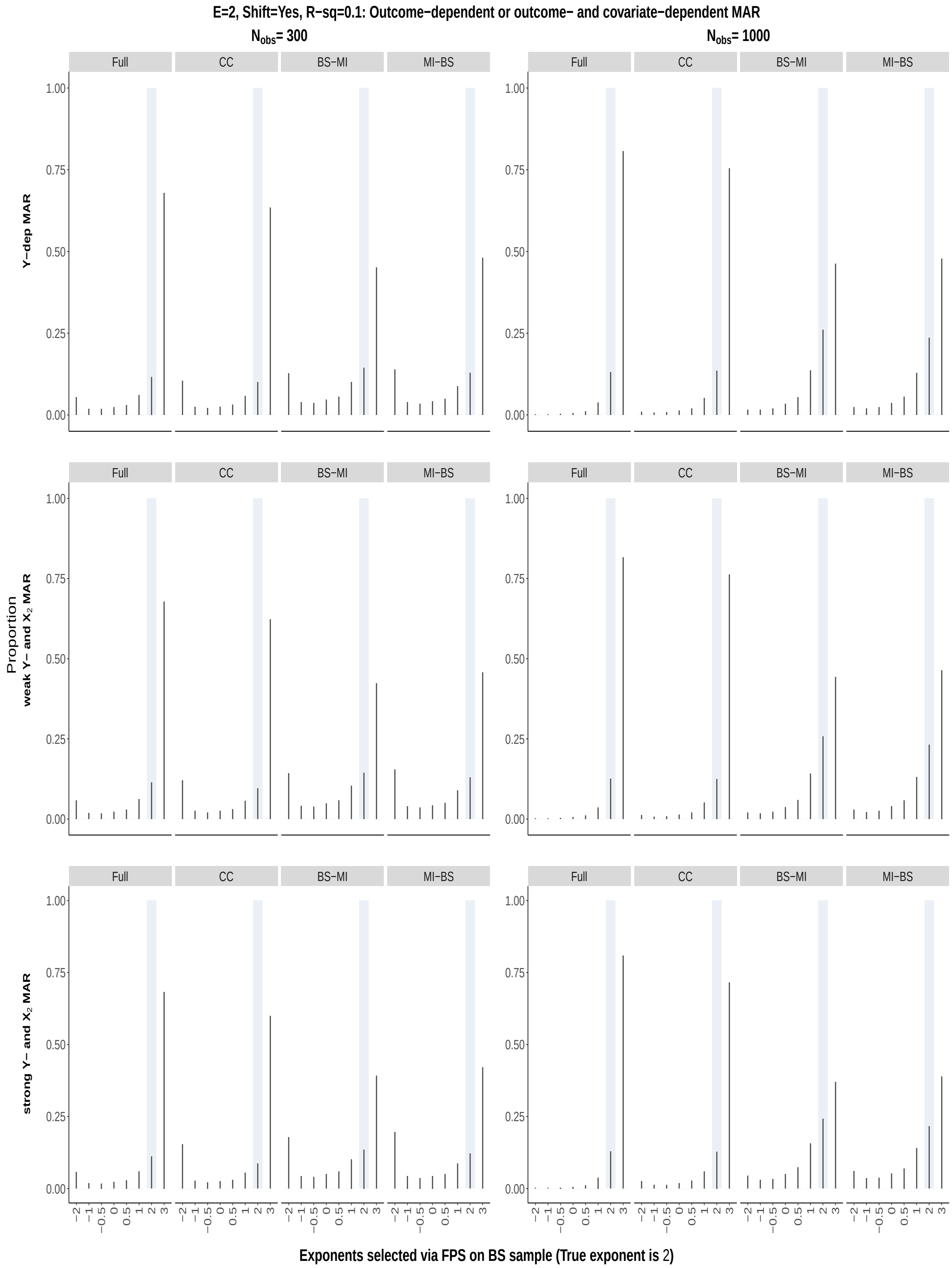

Figure S103: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

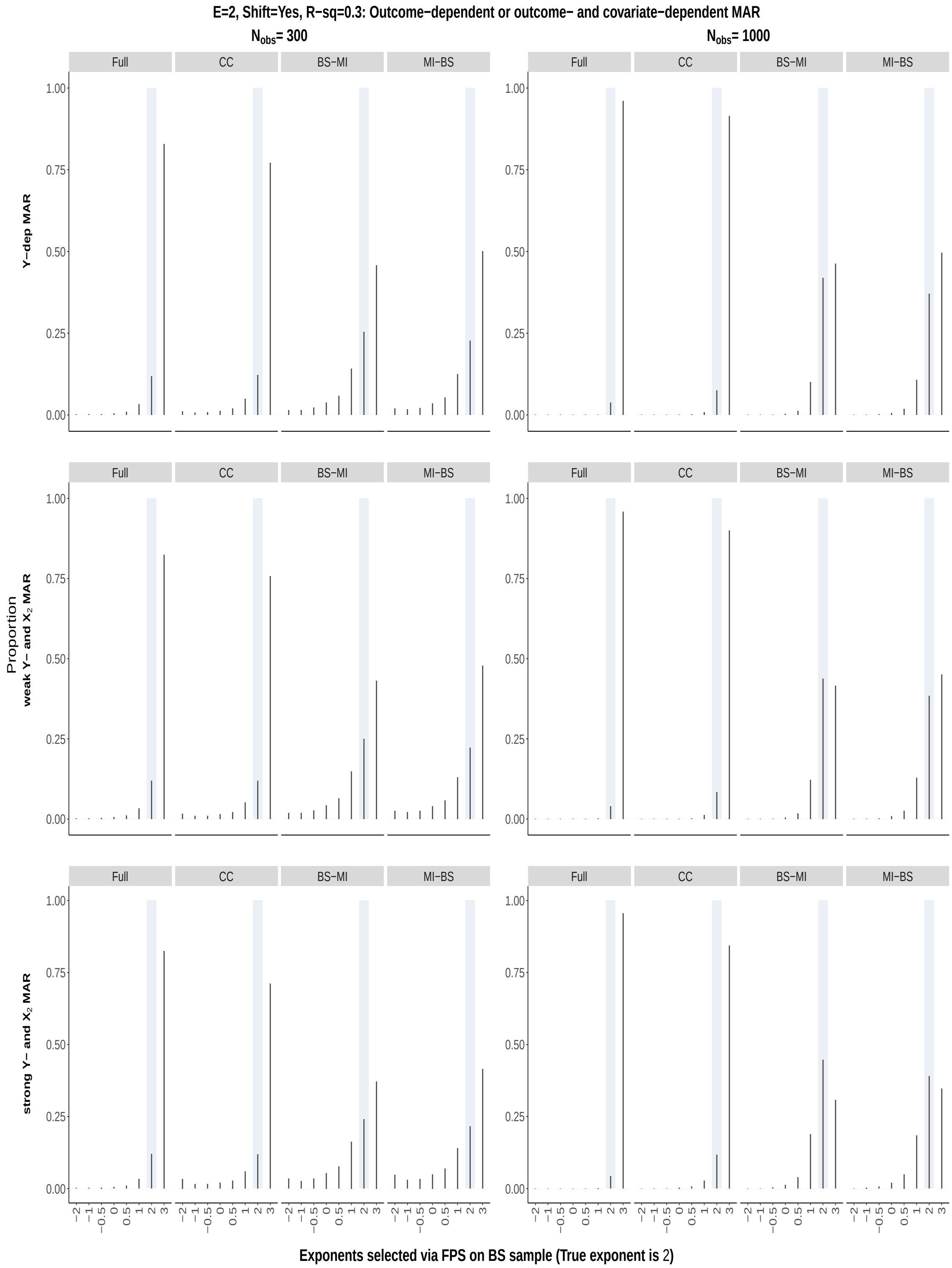

Figure S104: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

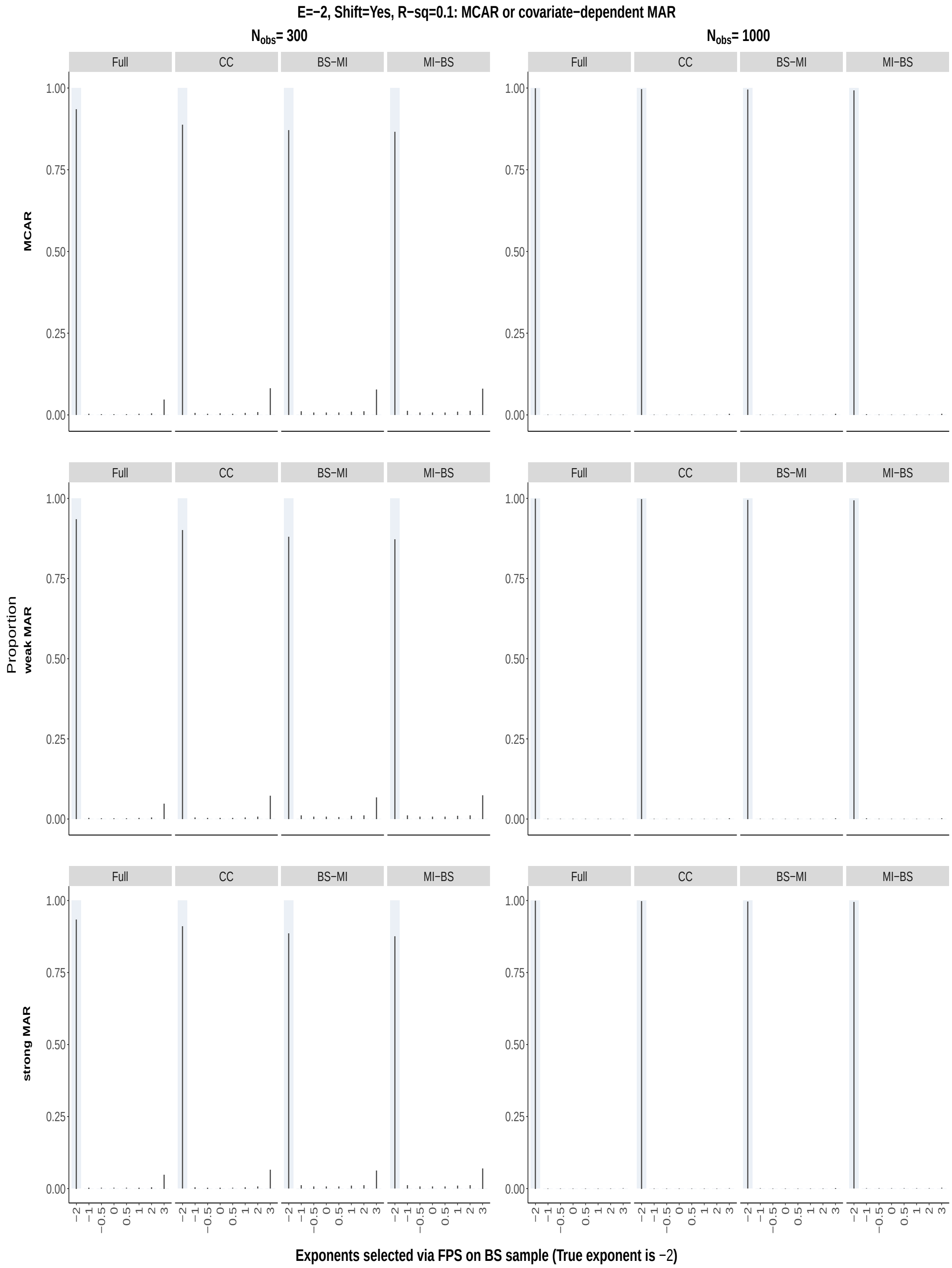

Figure S105: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

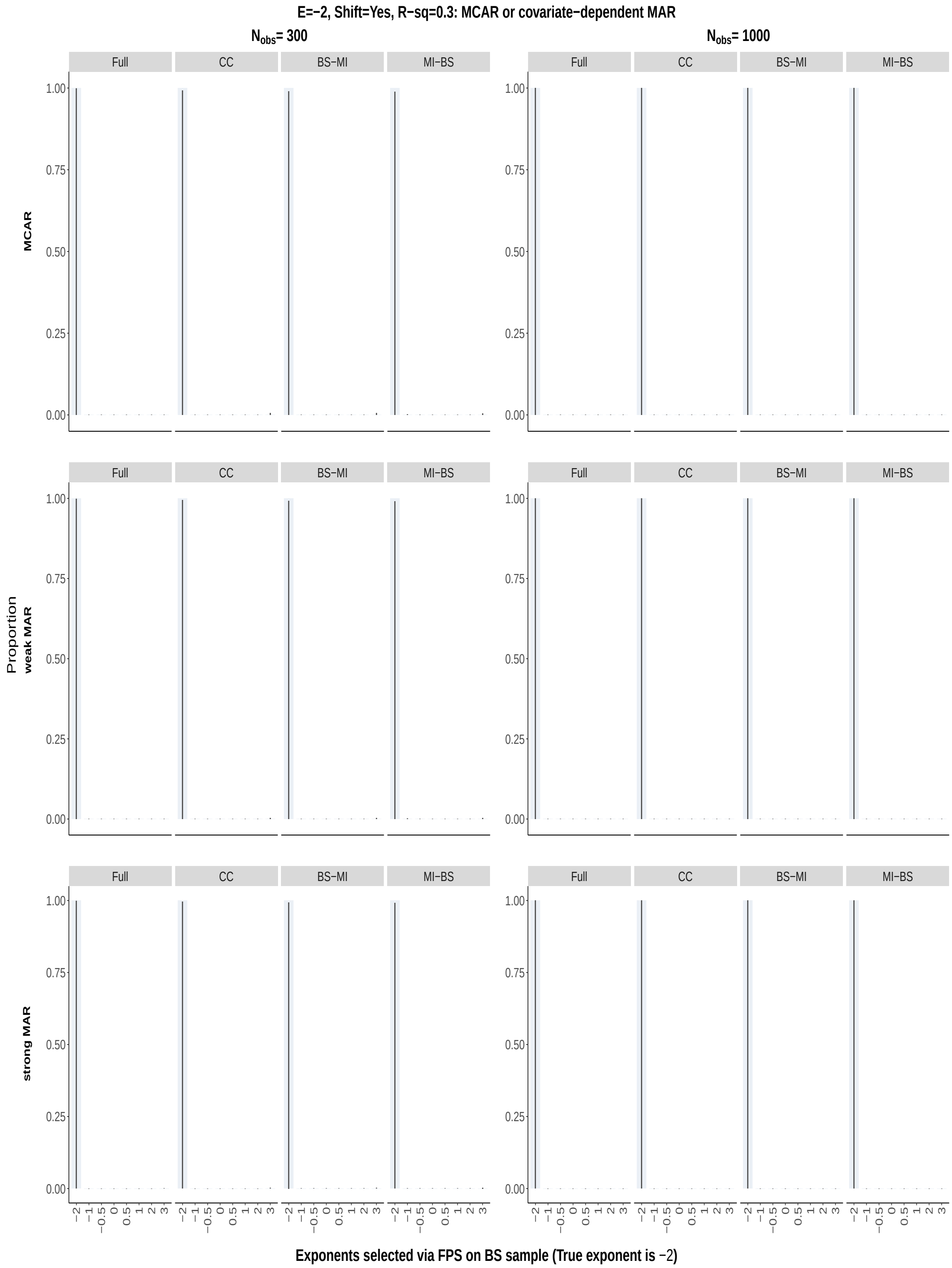

Figure S106: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

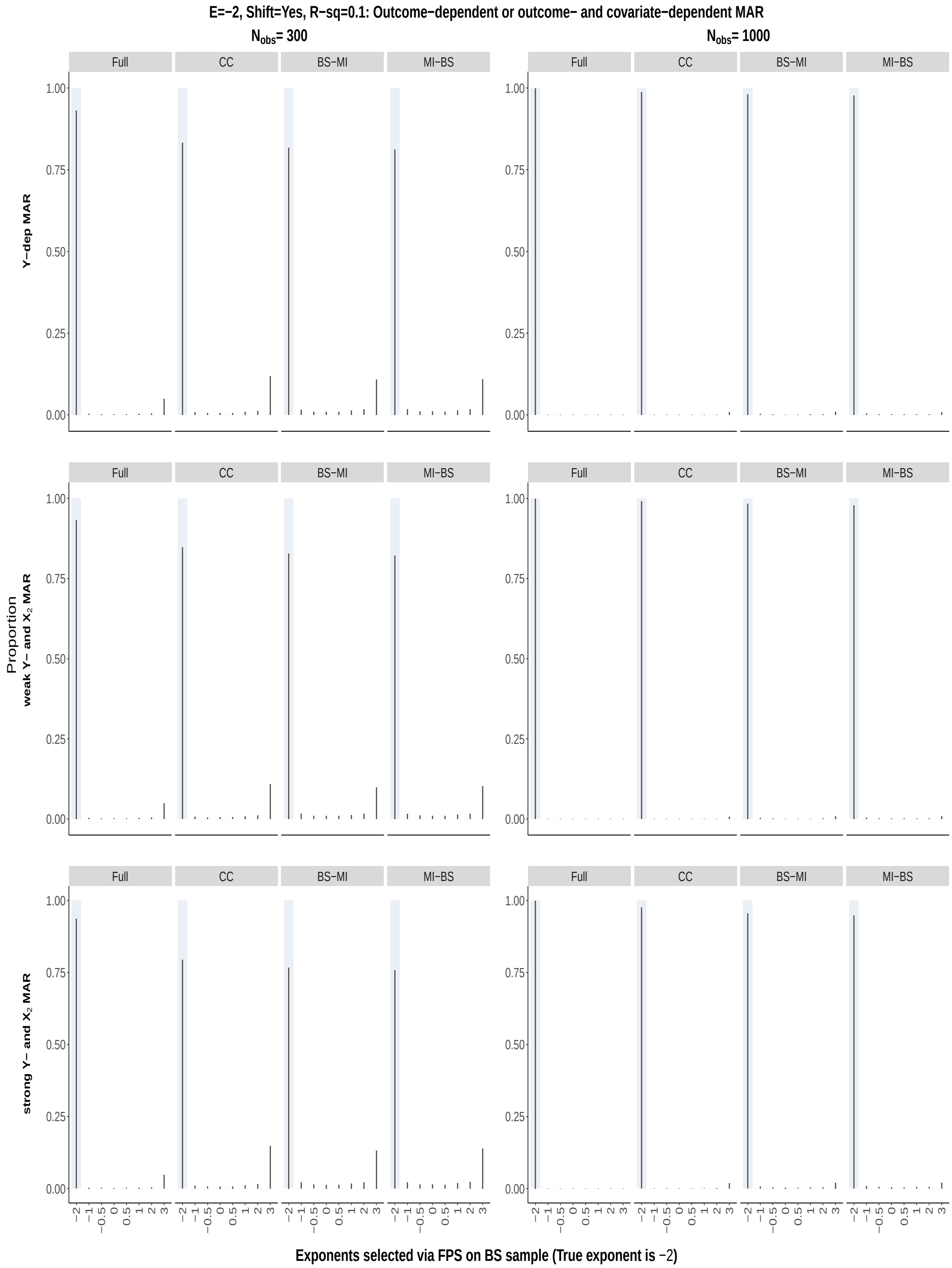

Figure S107: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

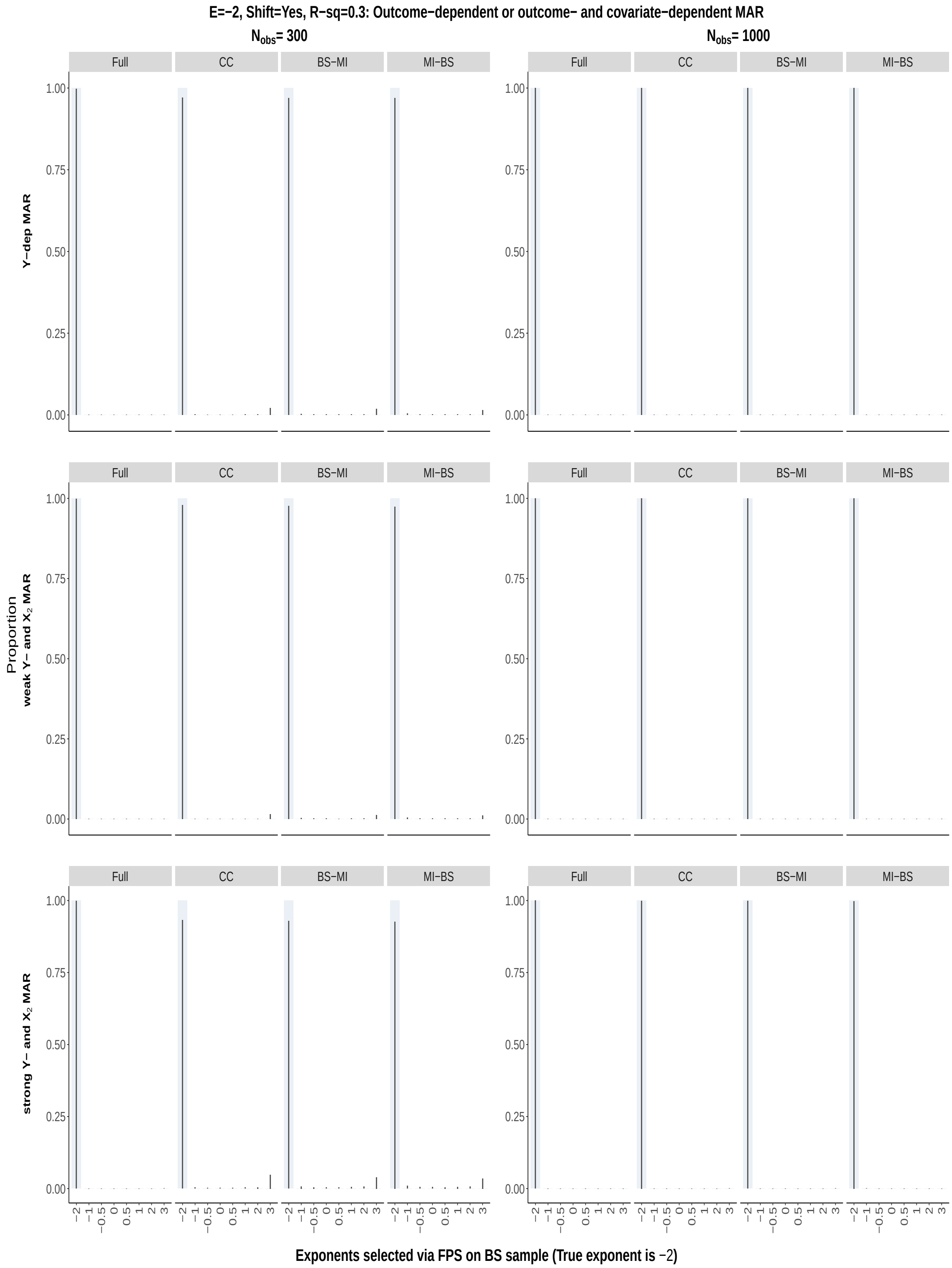

Figure S108: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.8 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

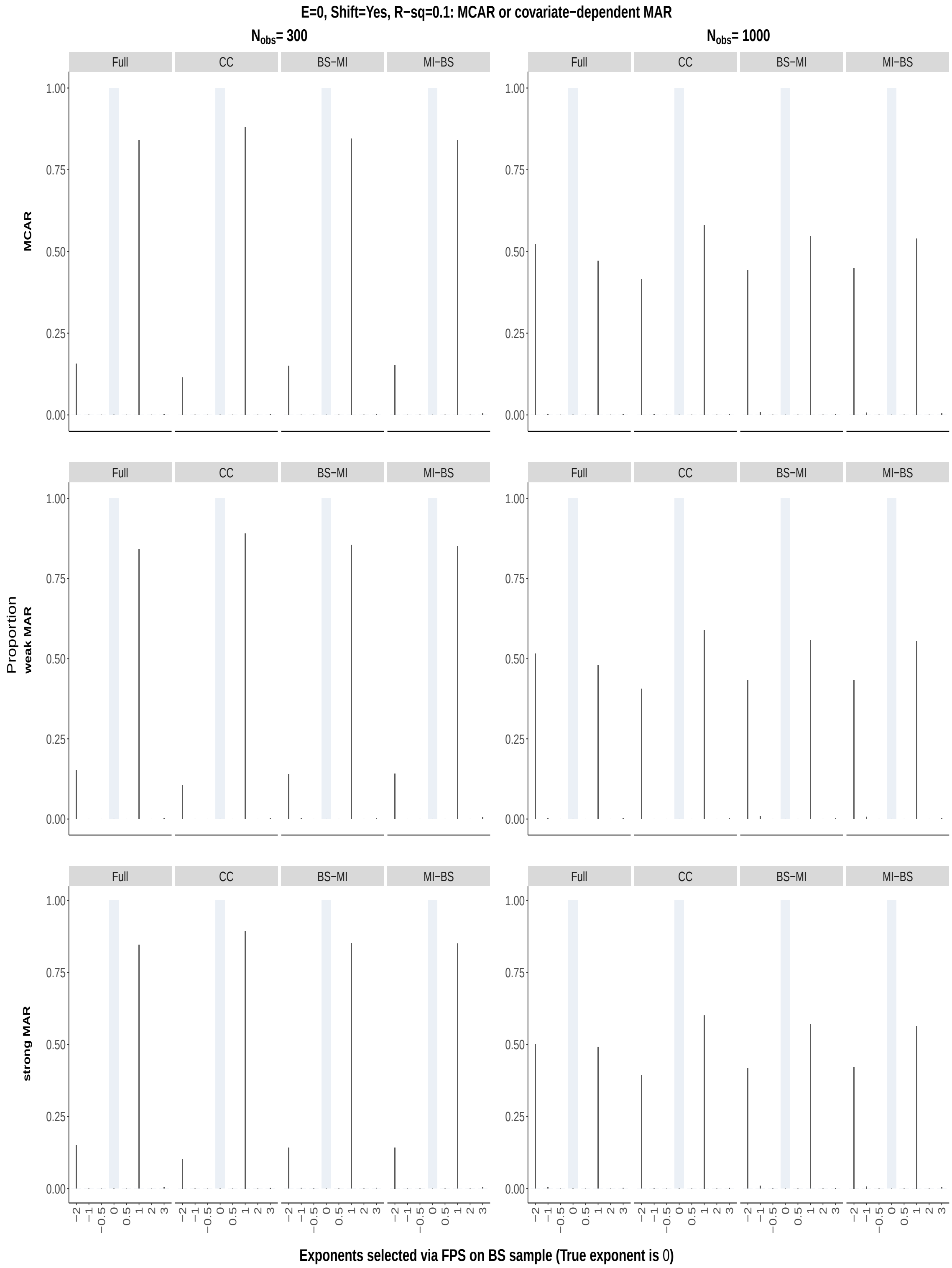

Figure S109: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

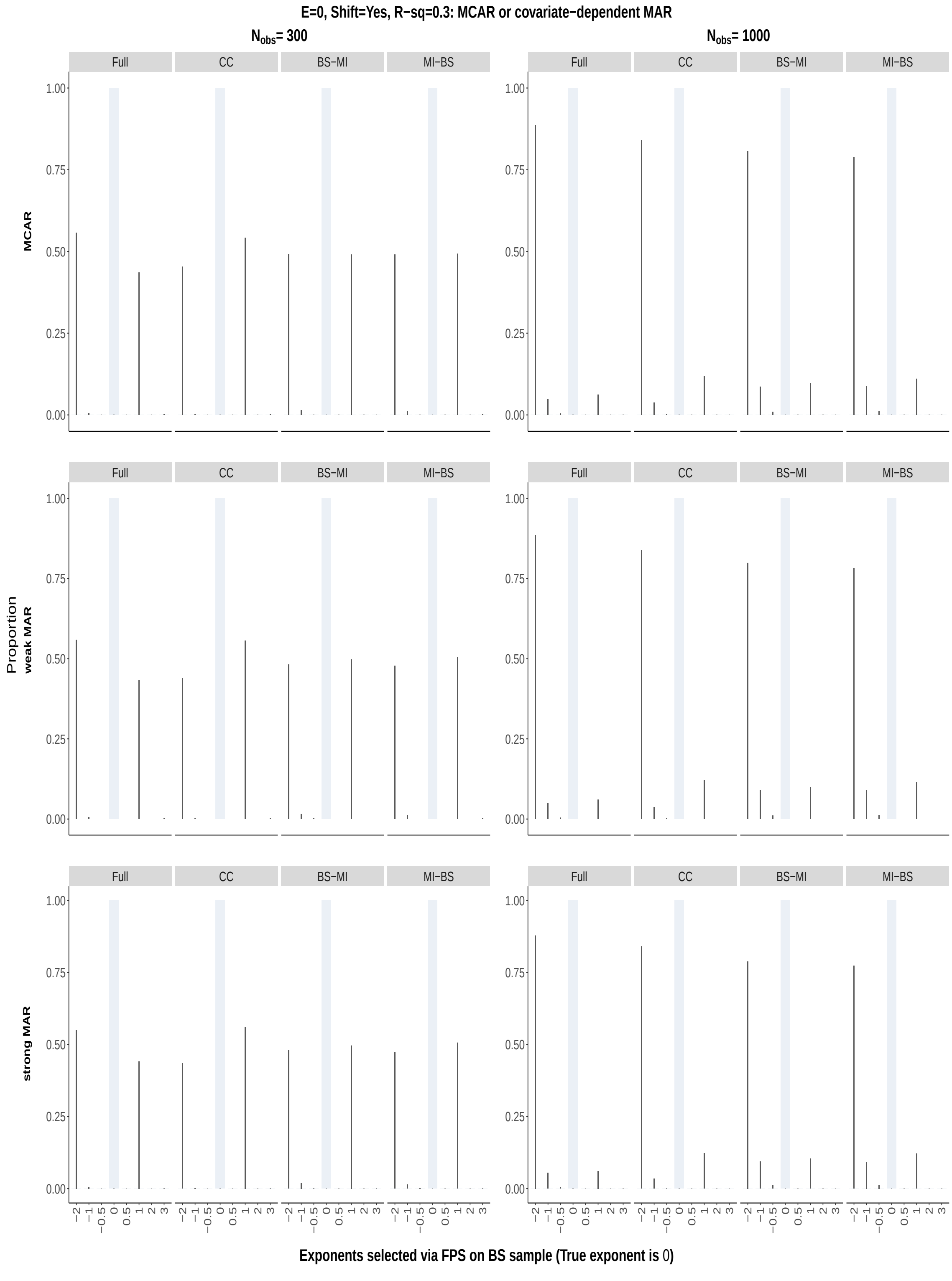

Figure S110: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

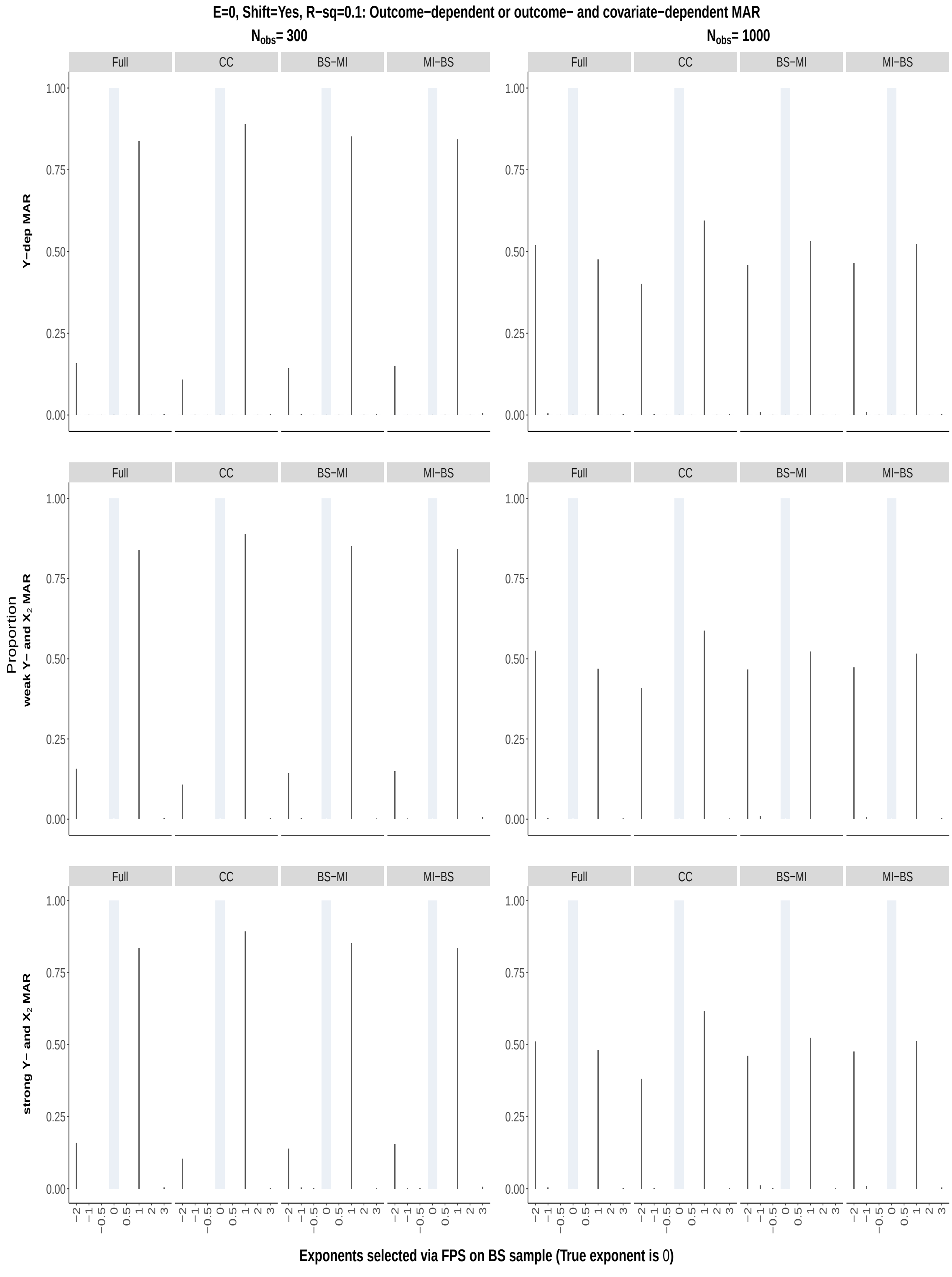

Figure S111: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

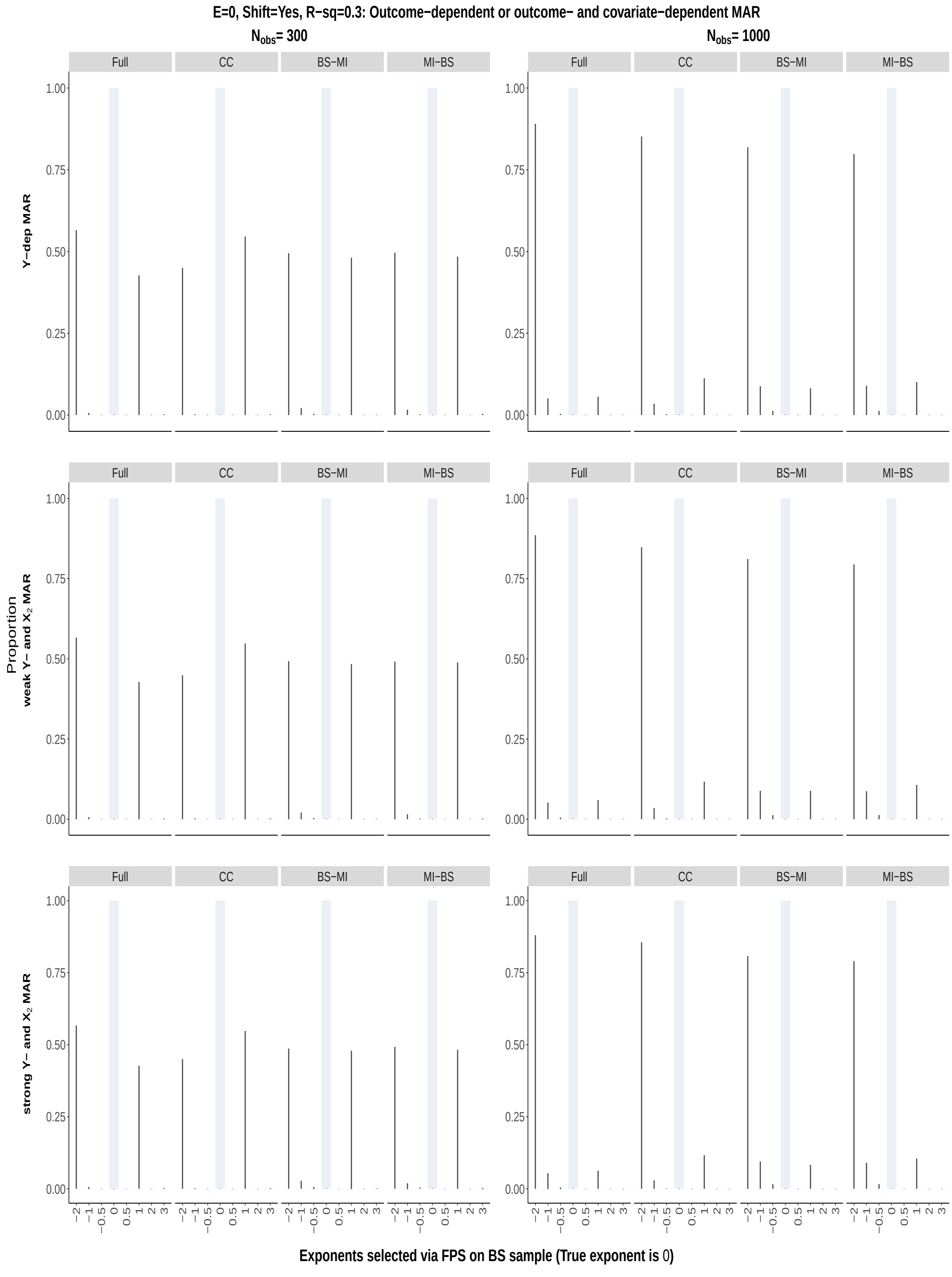

Figure S112: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

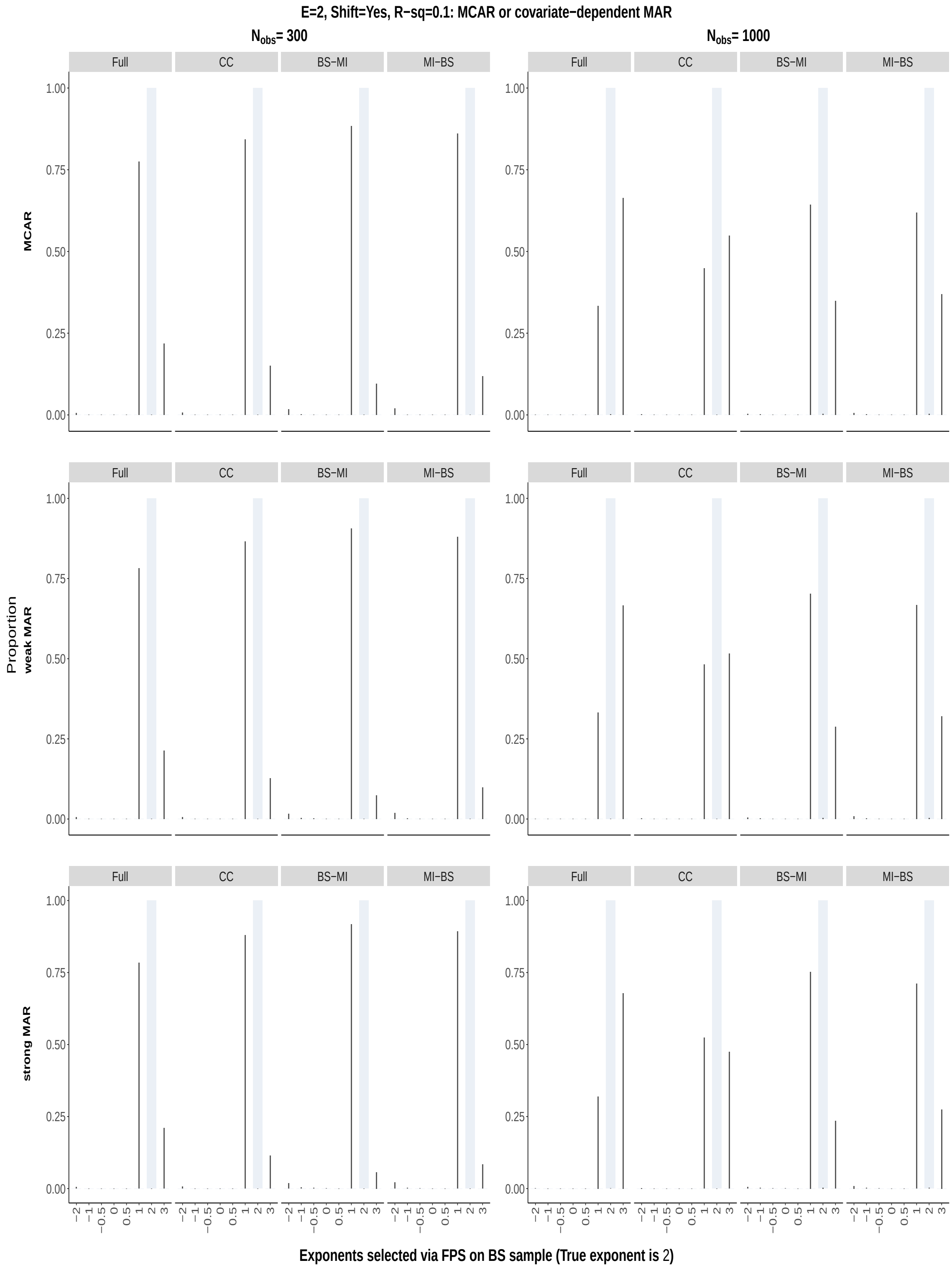

Figure S113: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

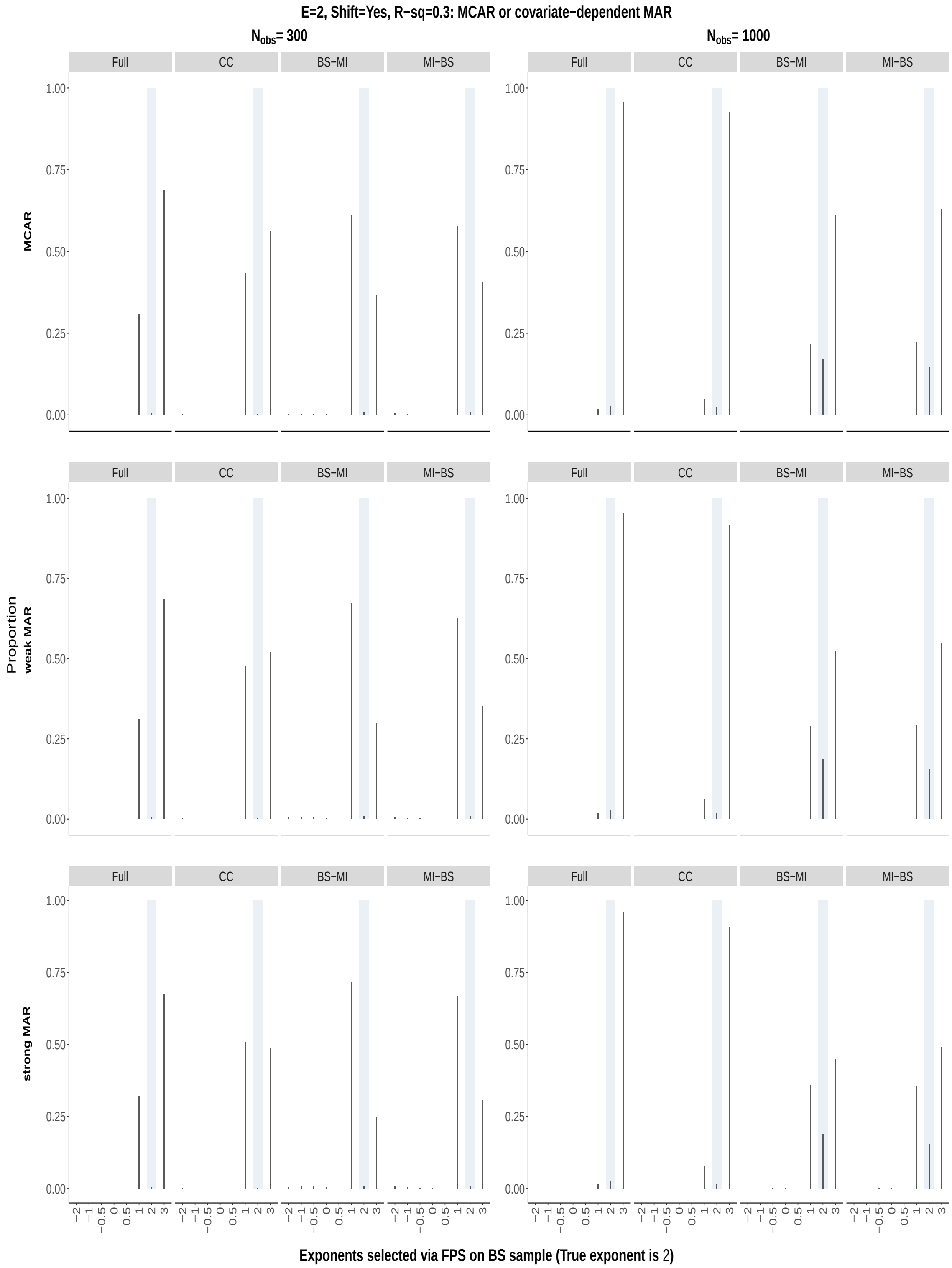

Figure S114: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

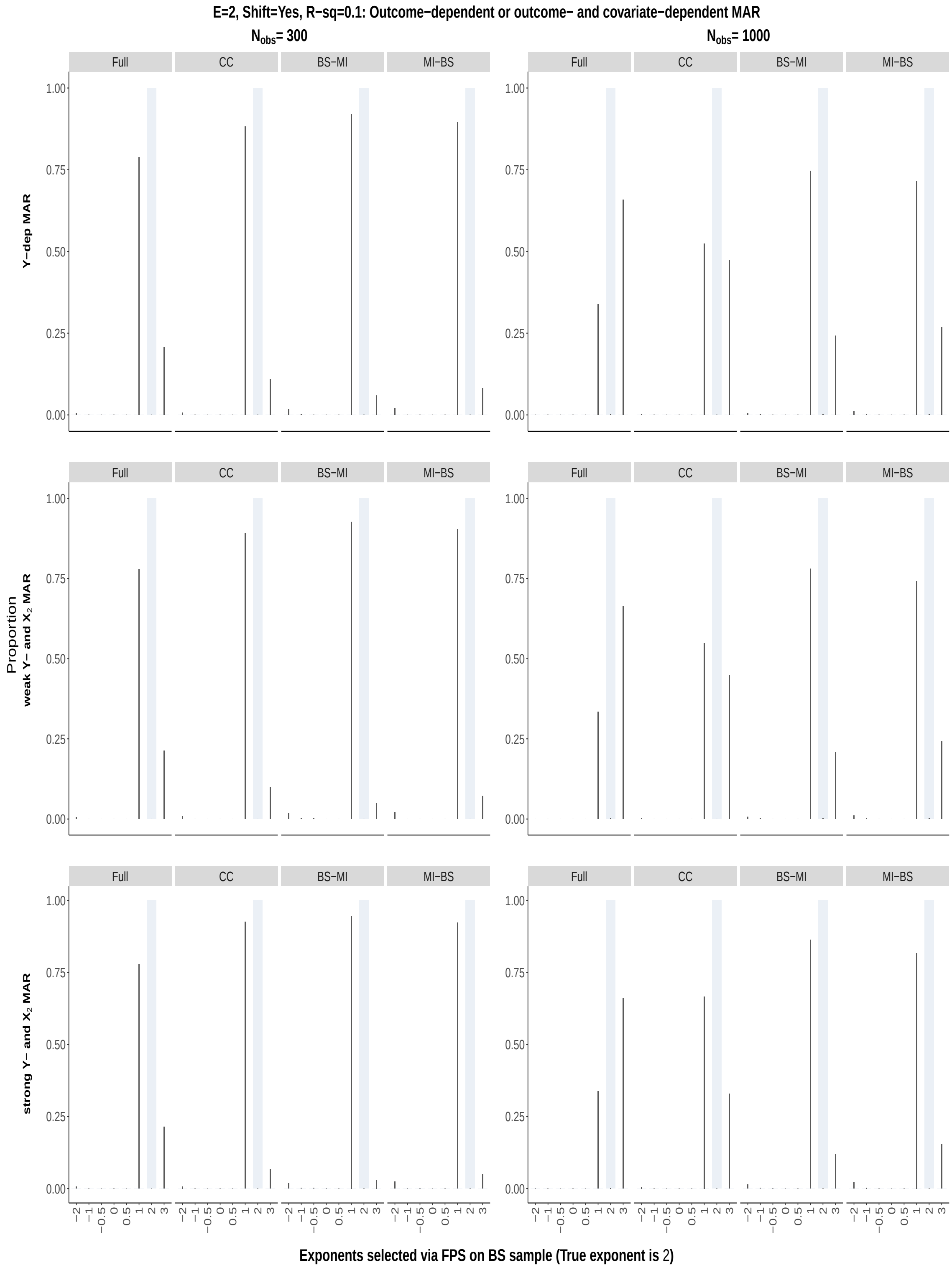

Figure S115: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

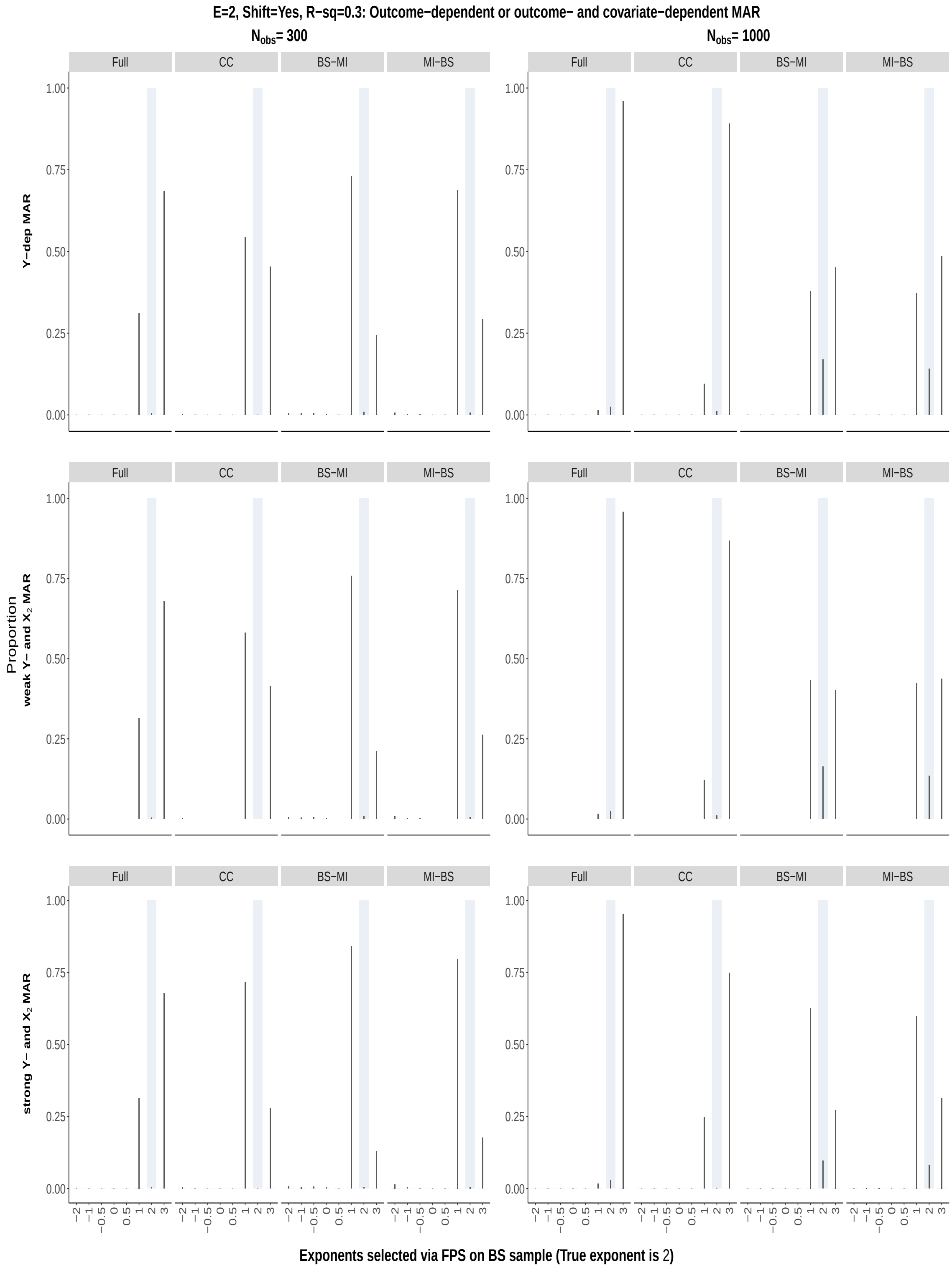

Figure S116: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

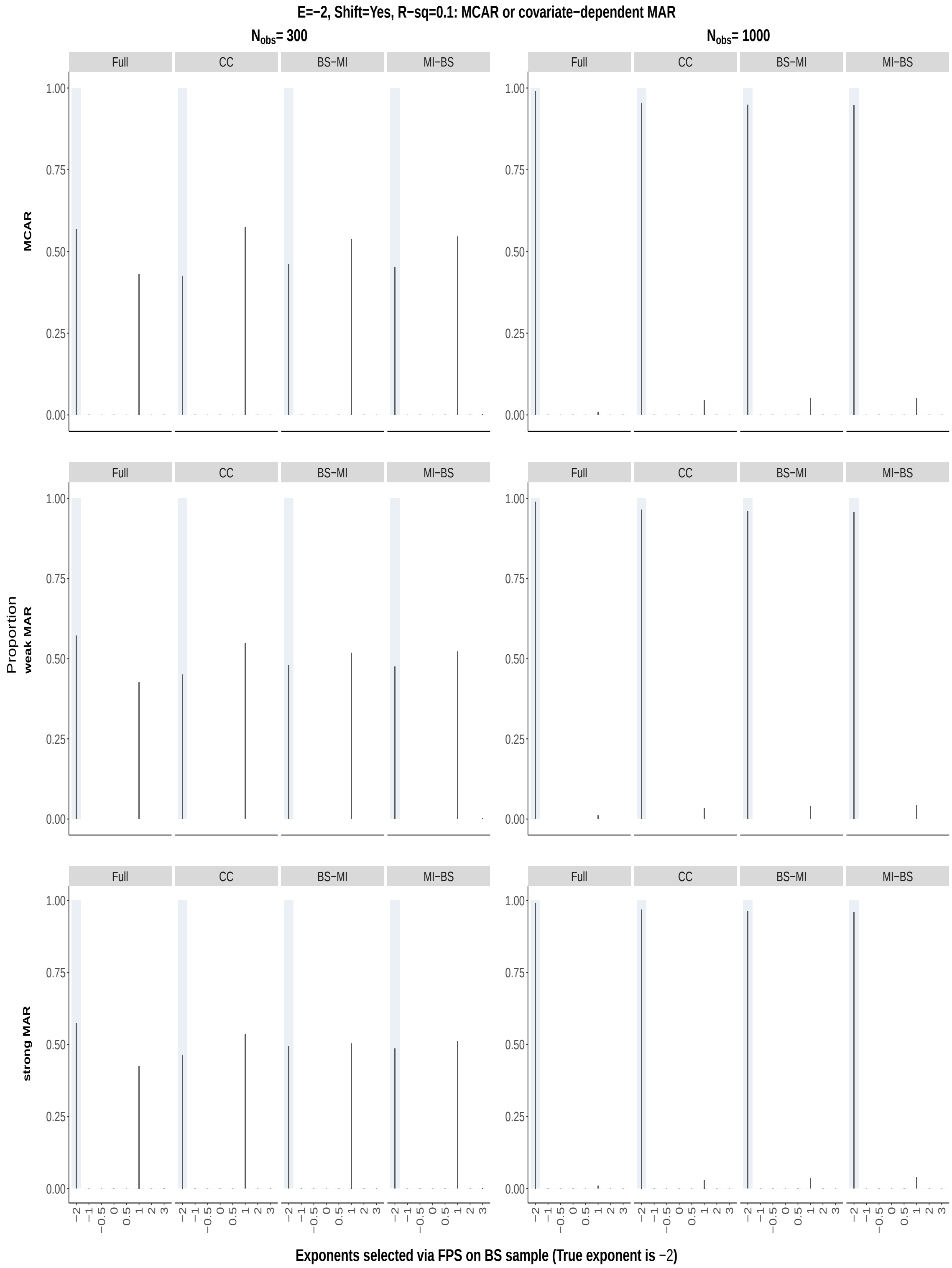

Figure S117: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

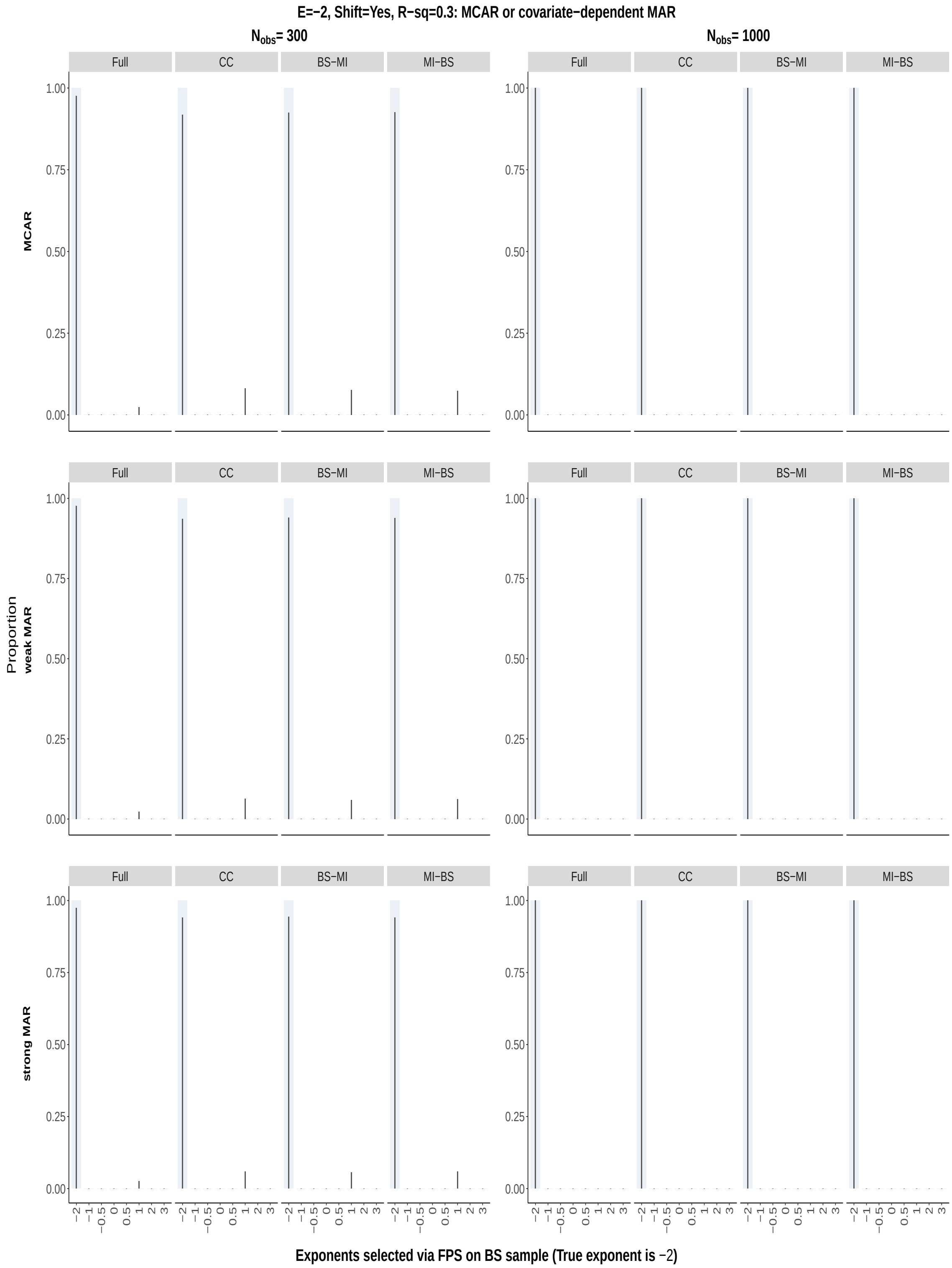

Figure S118: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

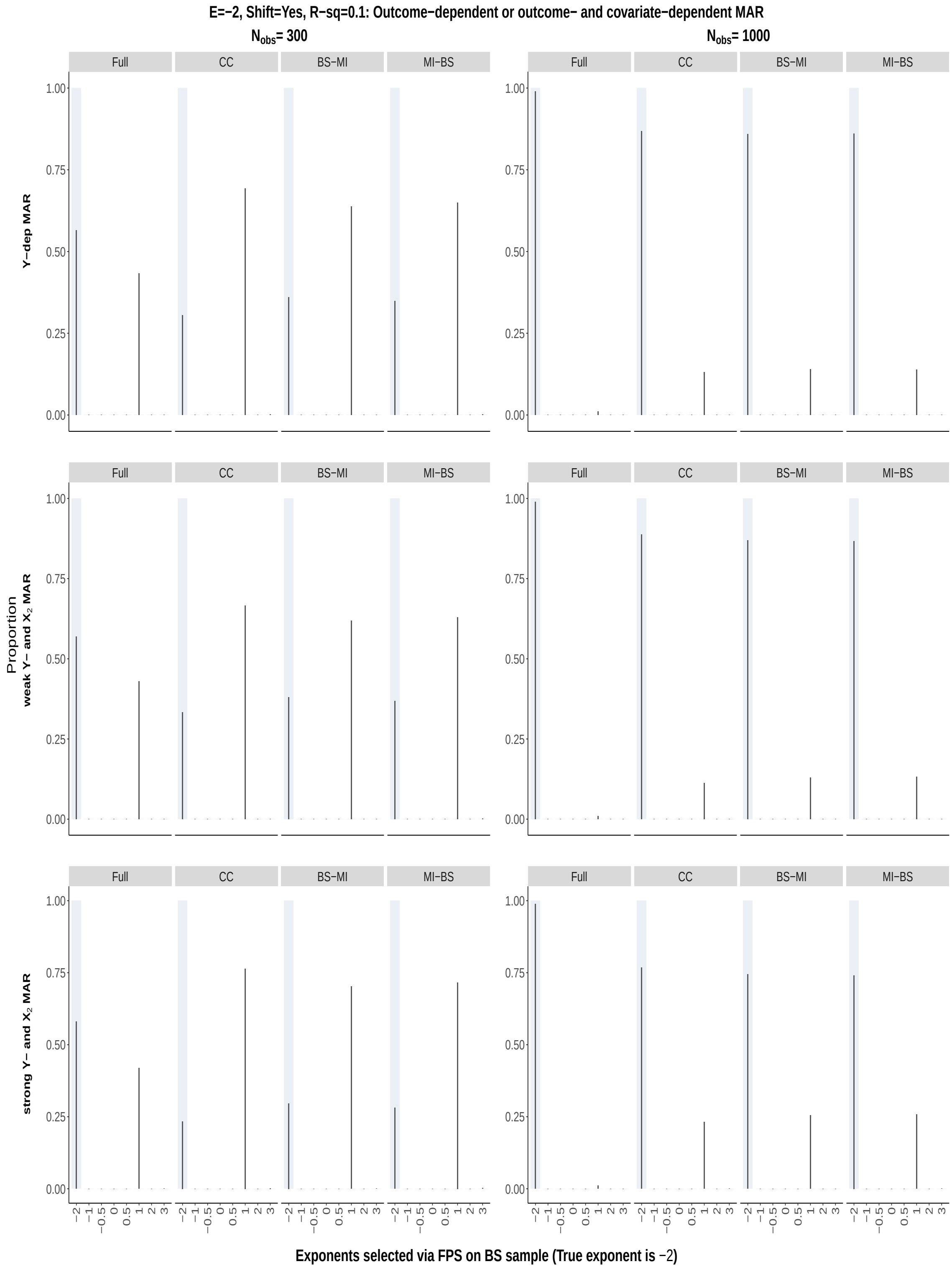

Figure S119: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

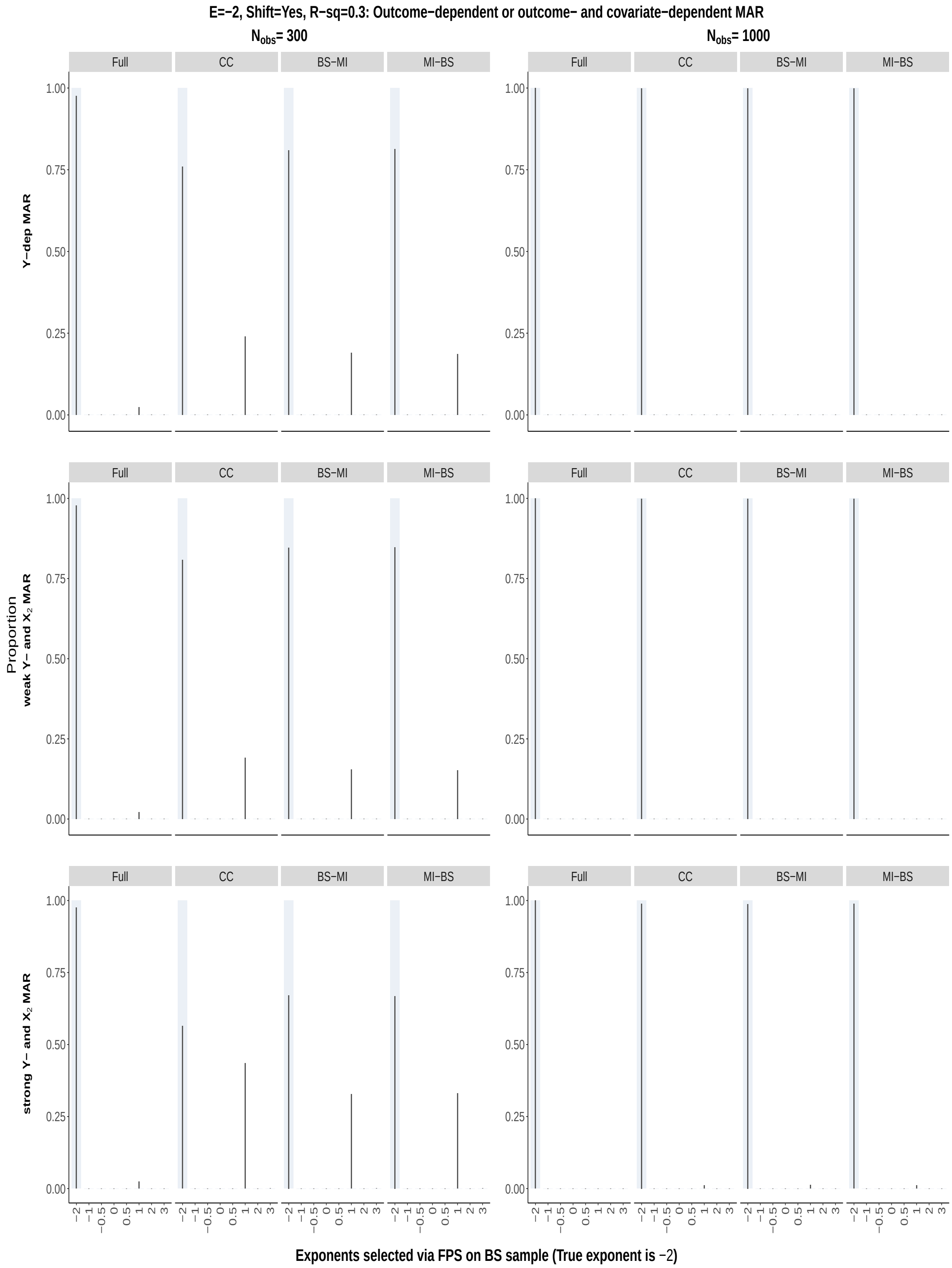

Figure S120: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.9 The 0.632 bootstrap, exponents selected using all the data: $\alpha_E = 1$ and no origin-shift

True exponent is 0

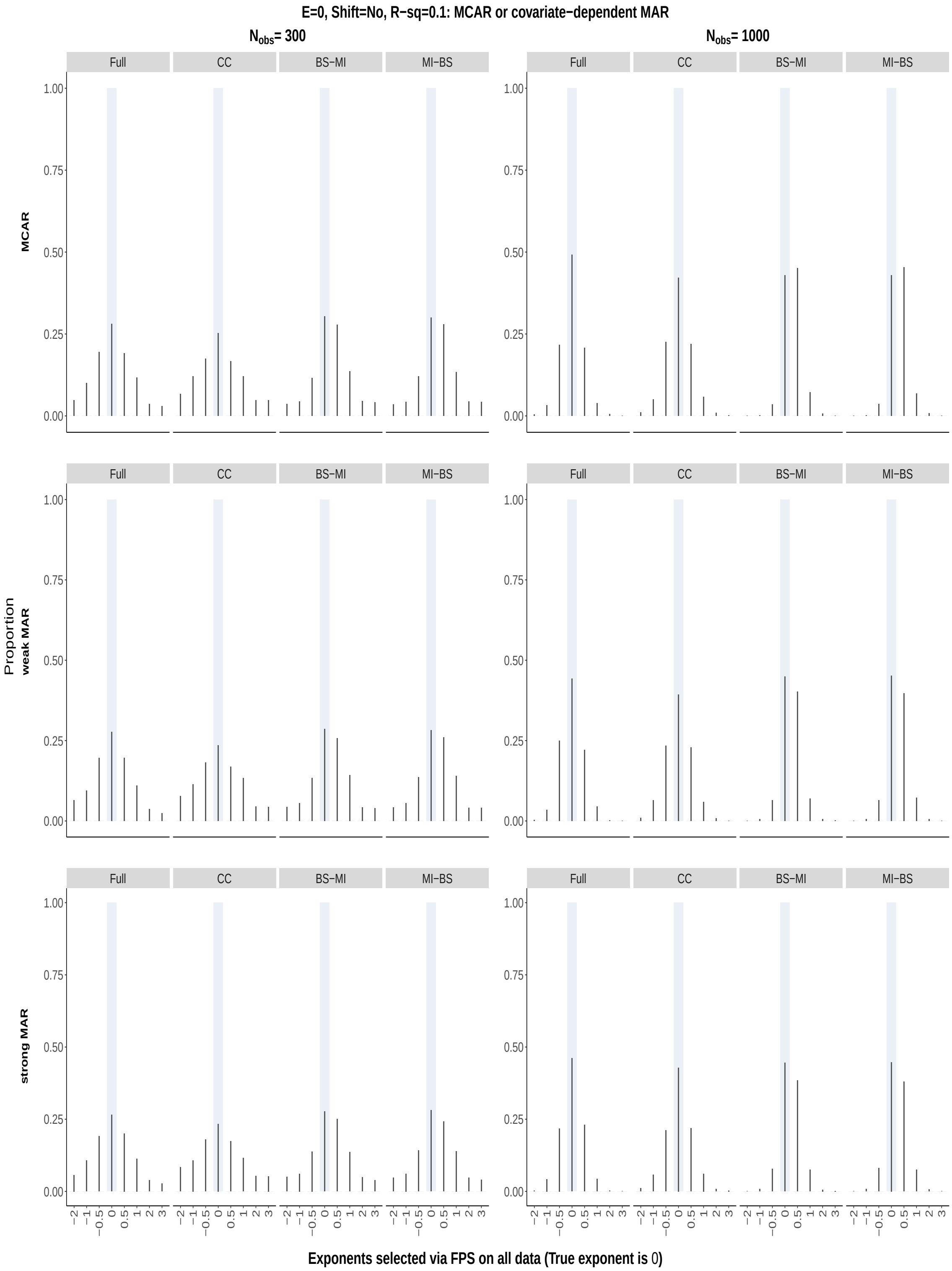

Figure S121: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

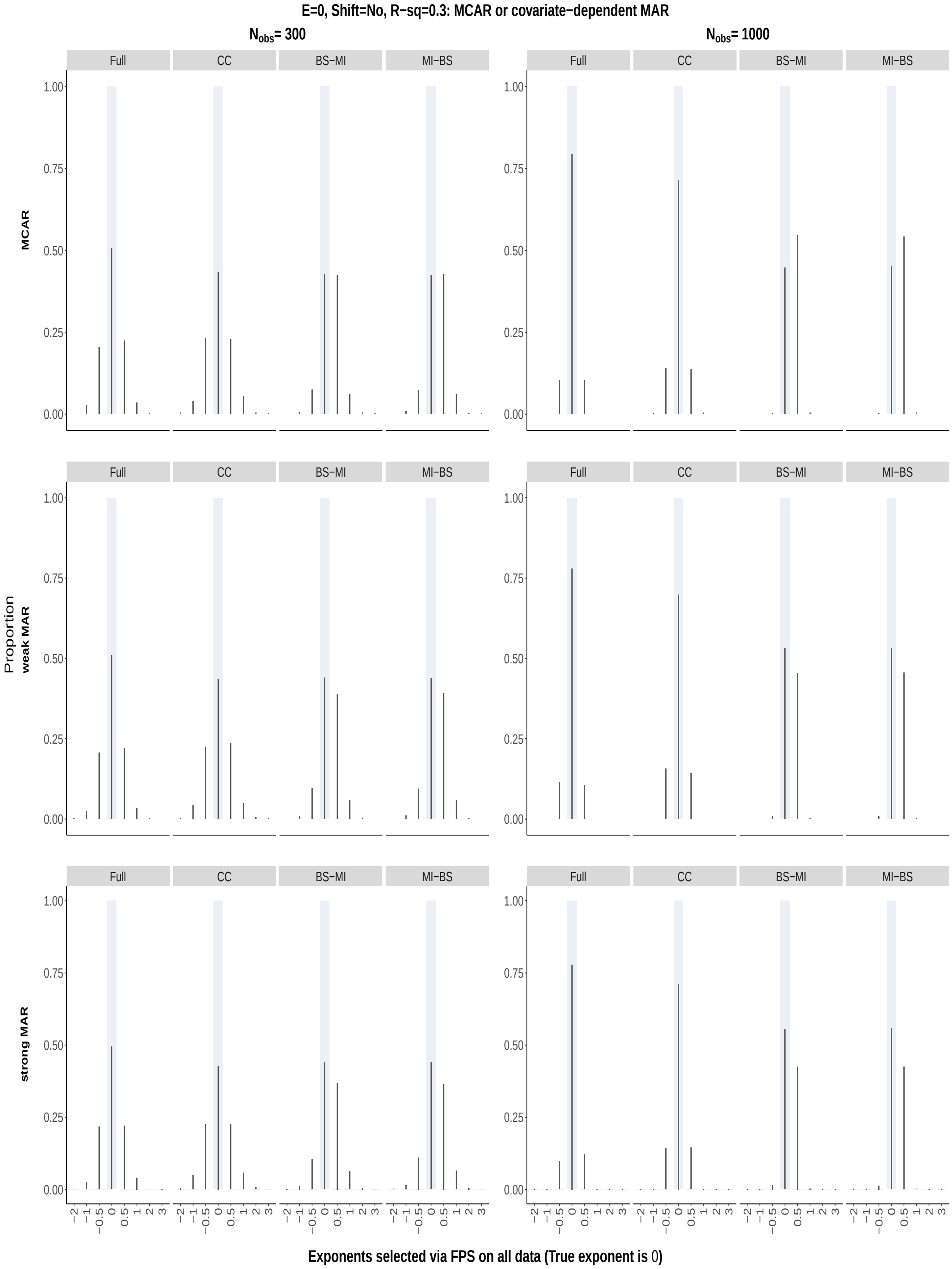

Figure S122: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

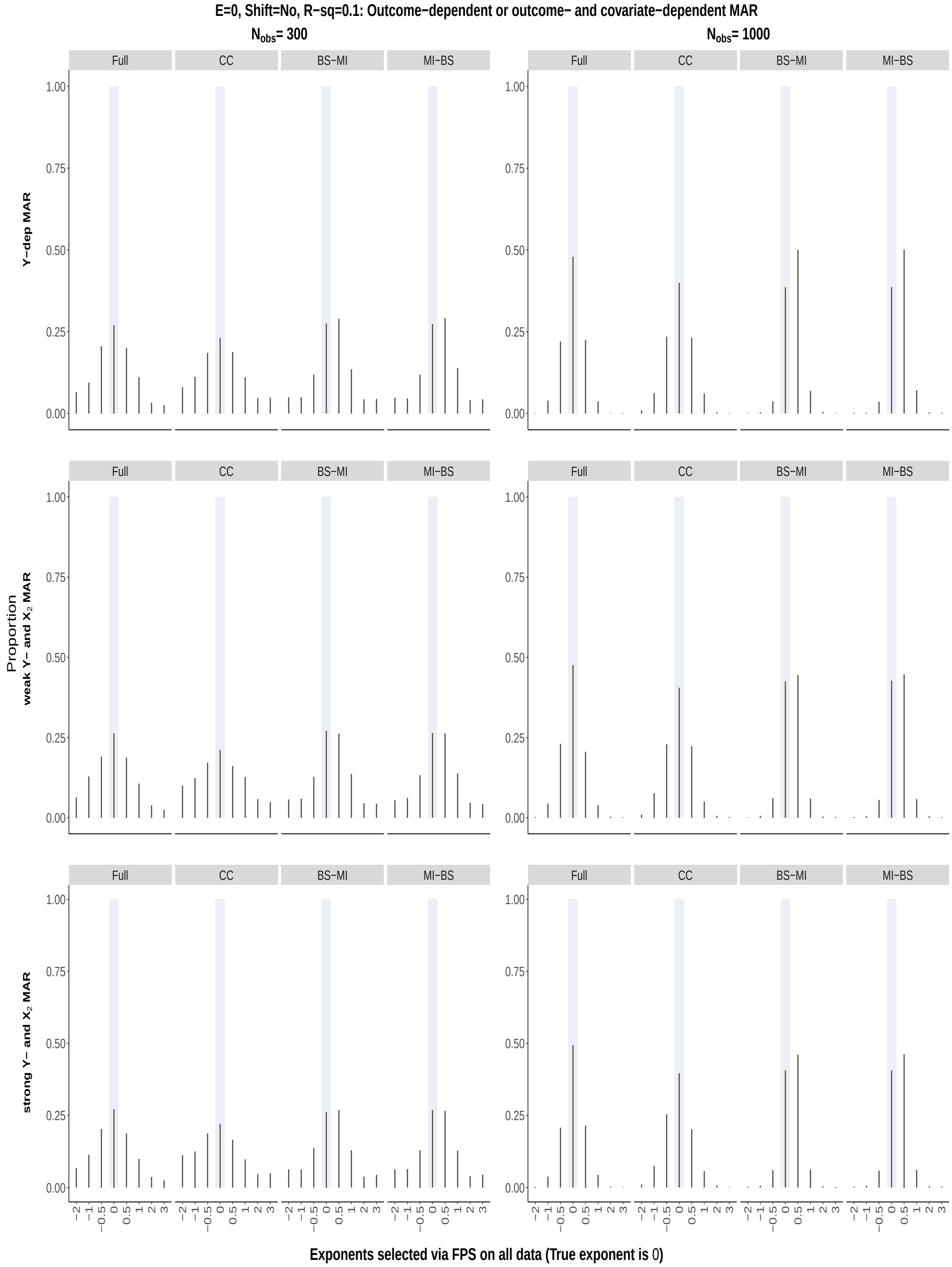

Figure S123: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

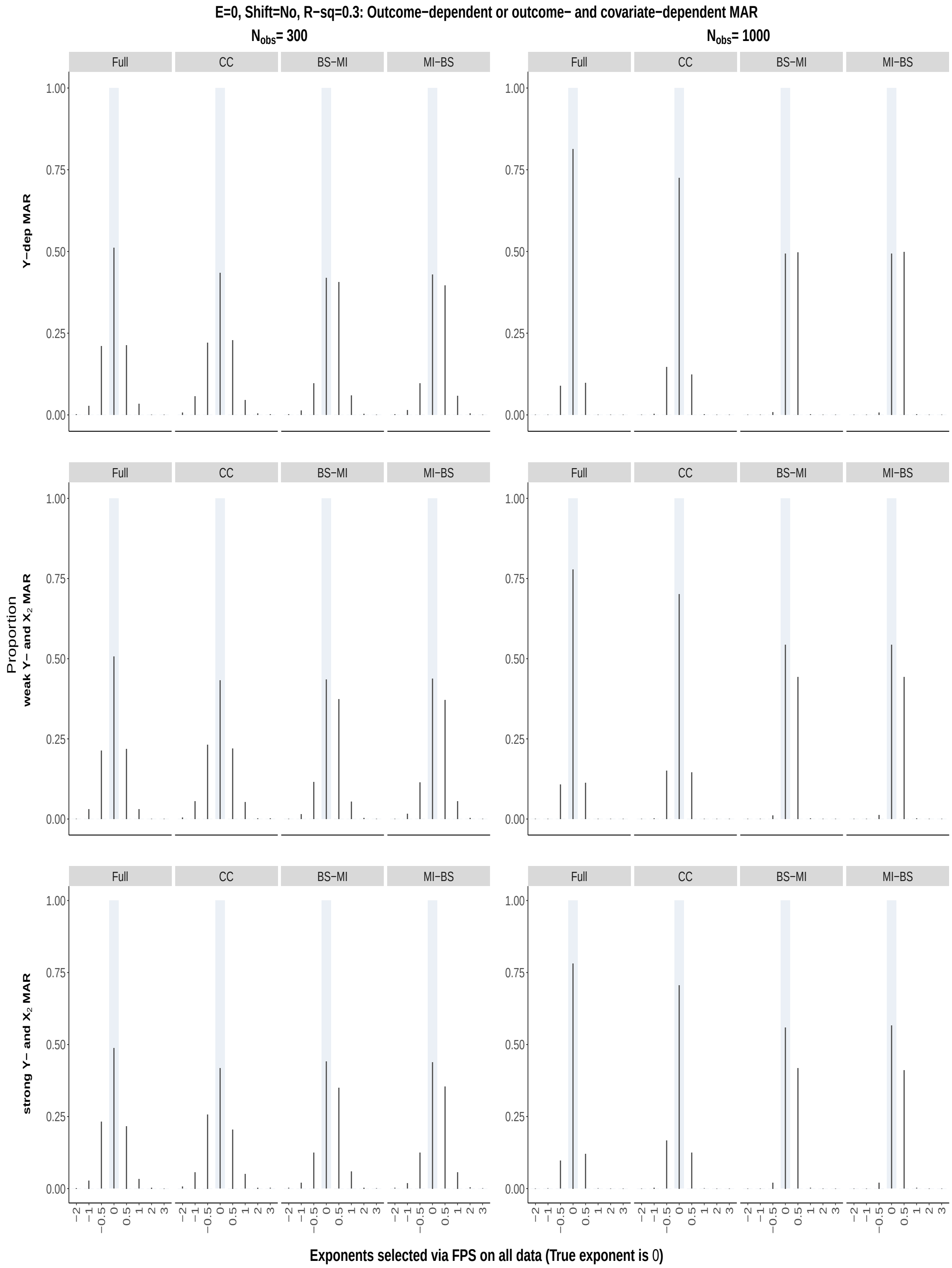

Figure S124: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

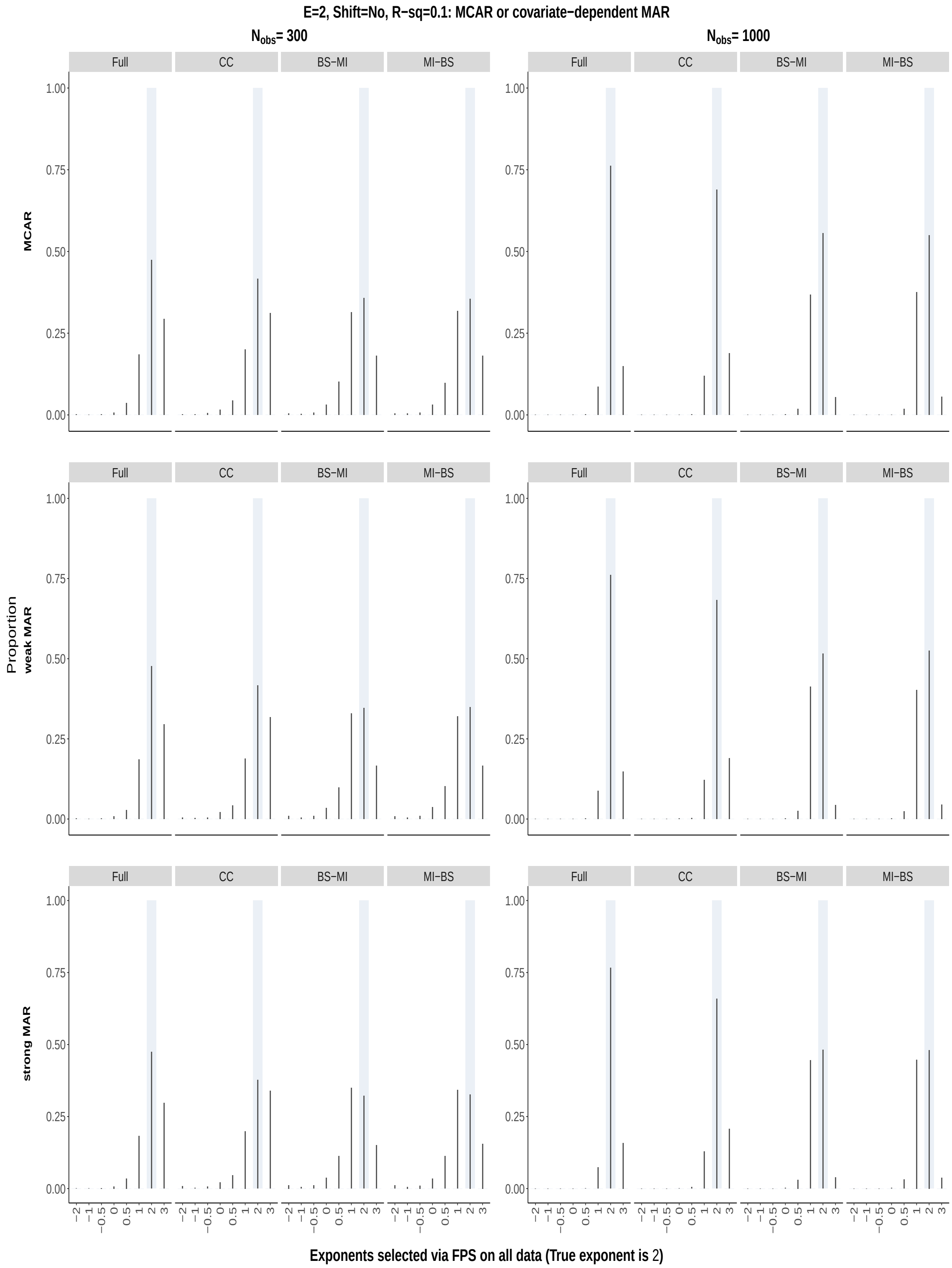

Figure S125: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

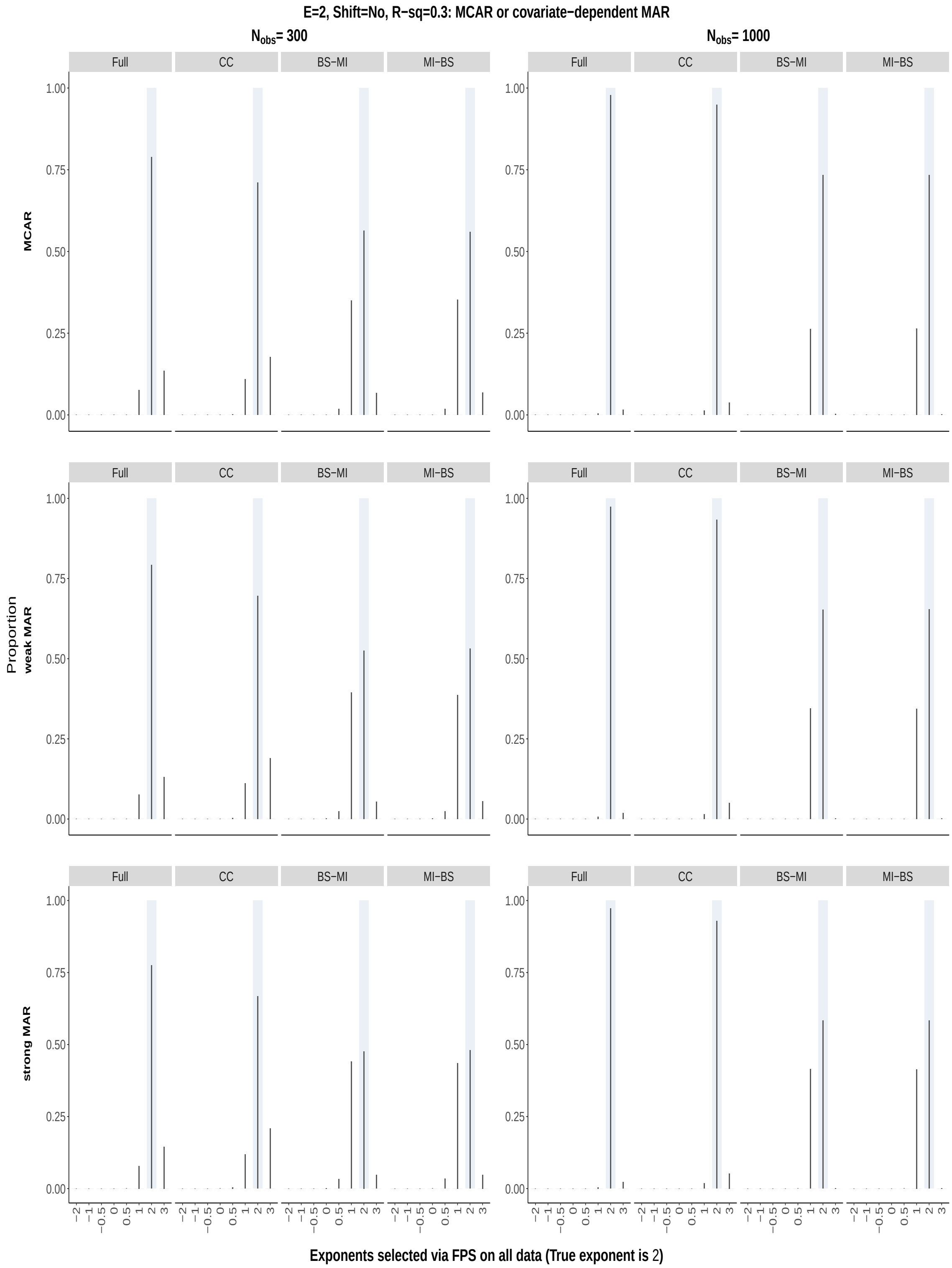

Figure S126: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

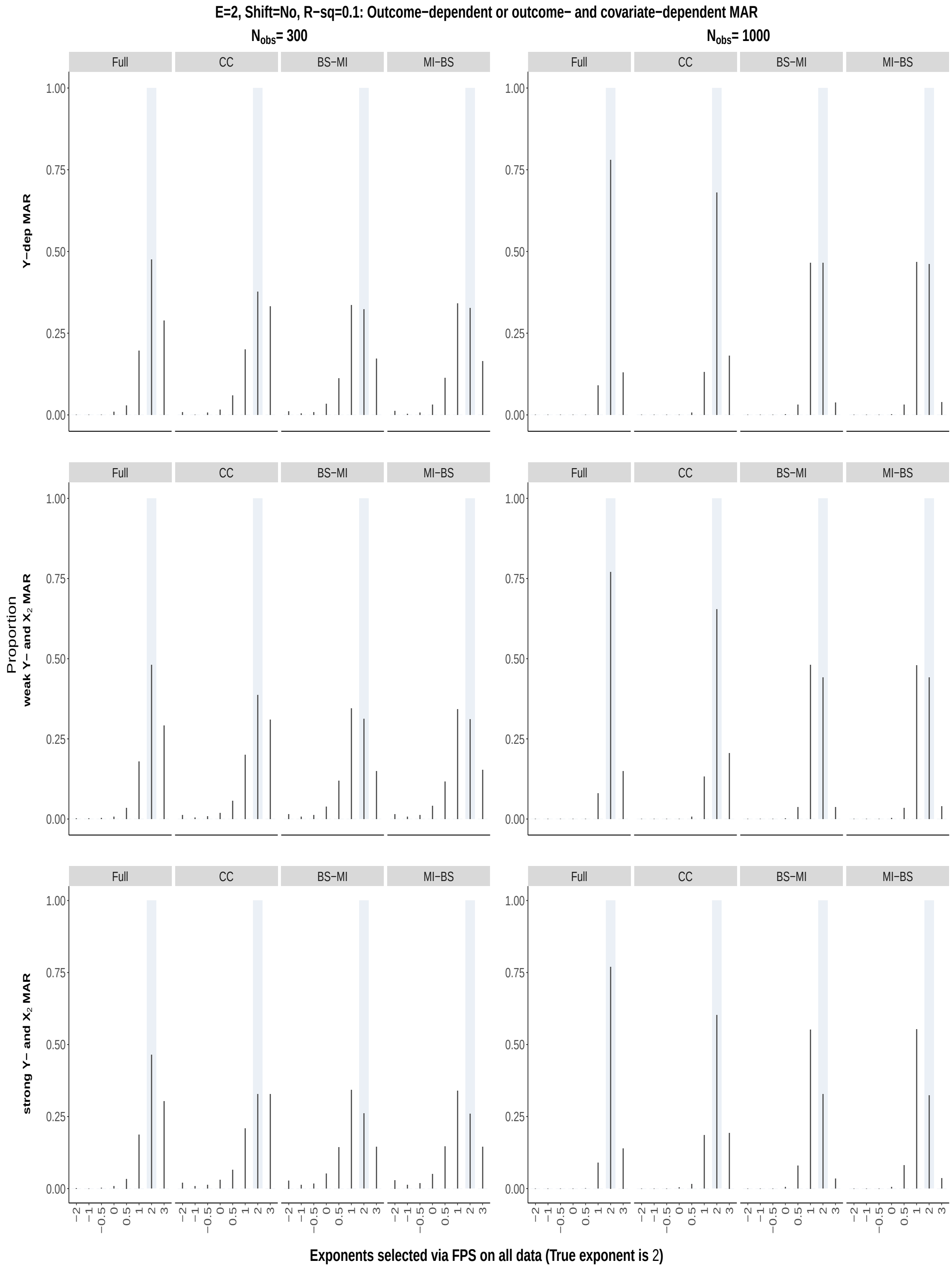

Figure S127: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

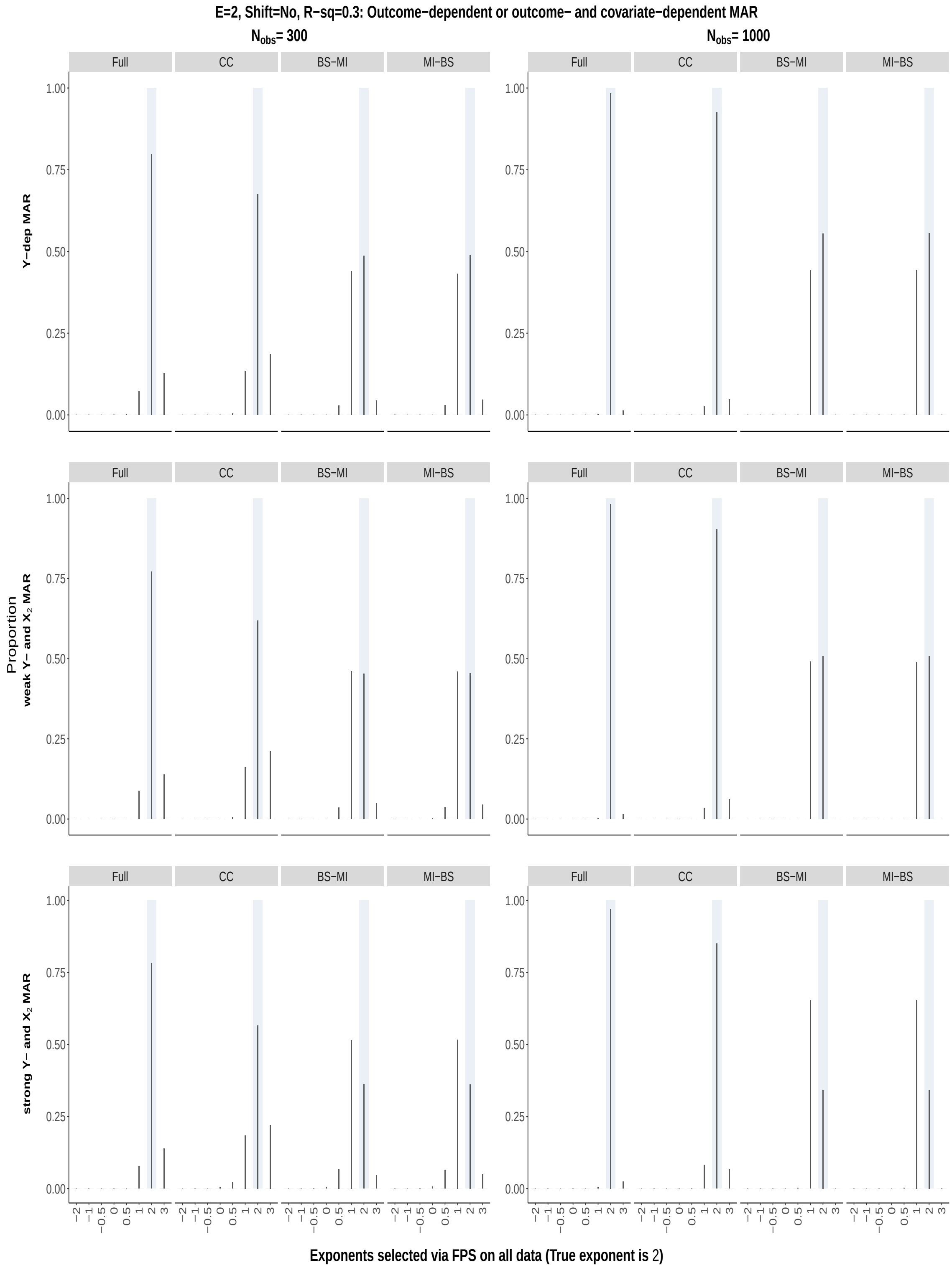

Figure S128: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

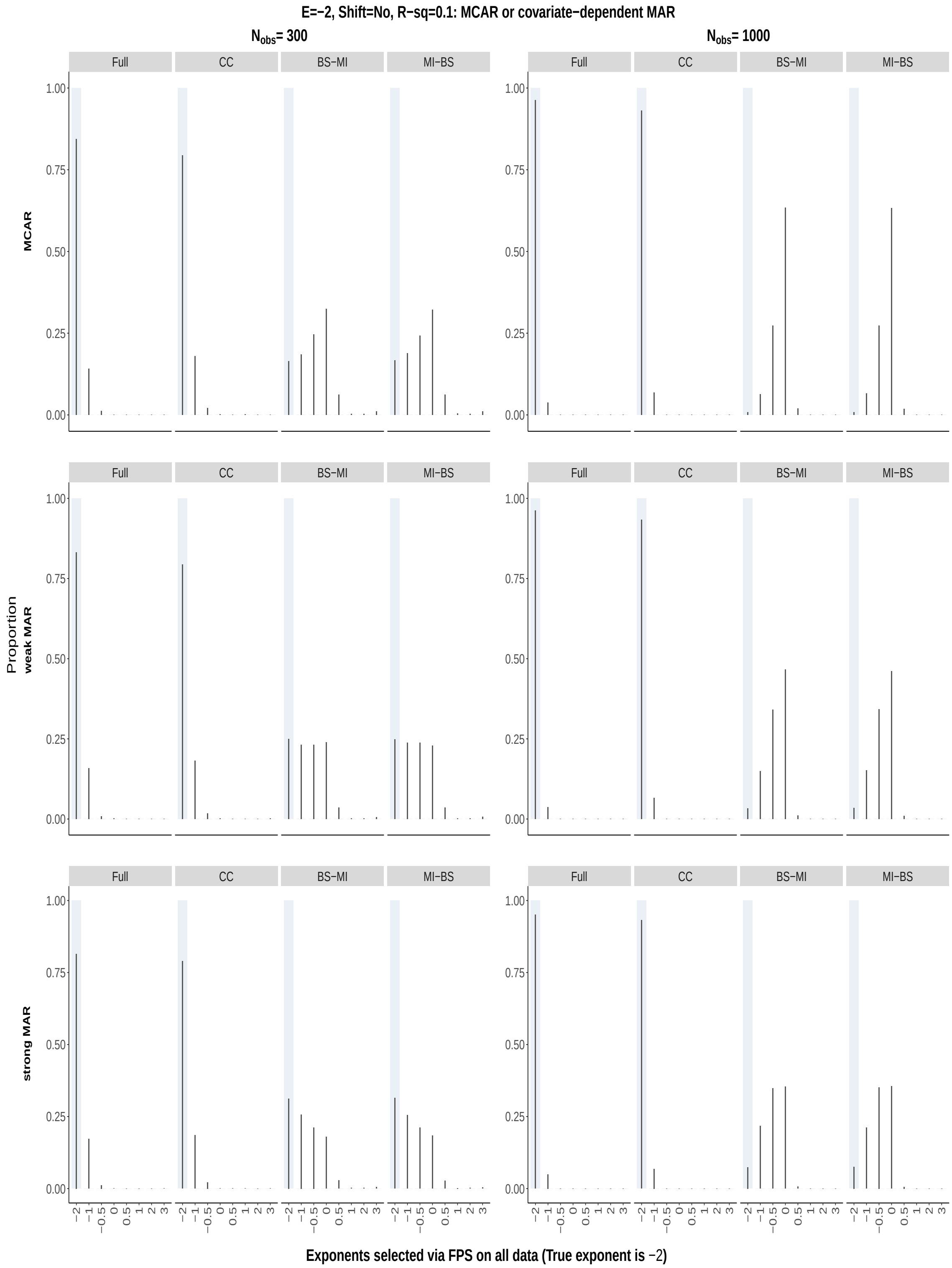

Figure S129: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

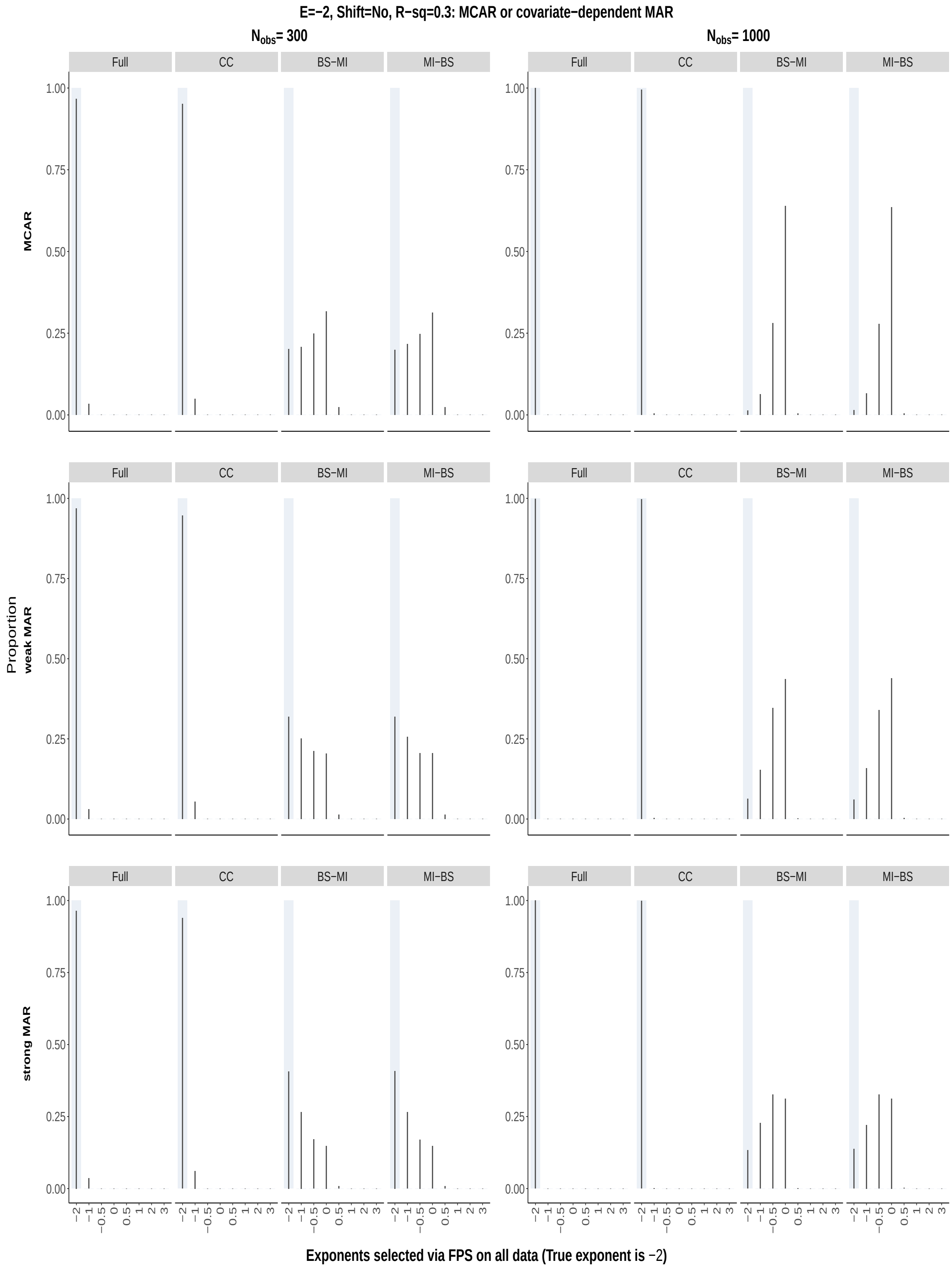

Figure S130: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

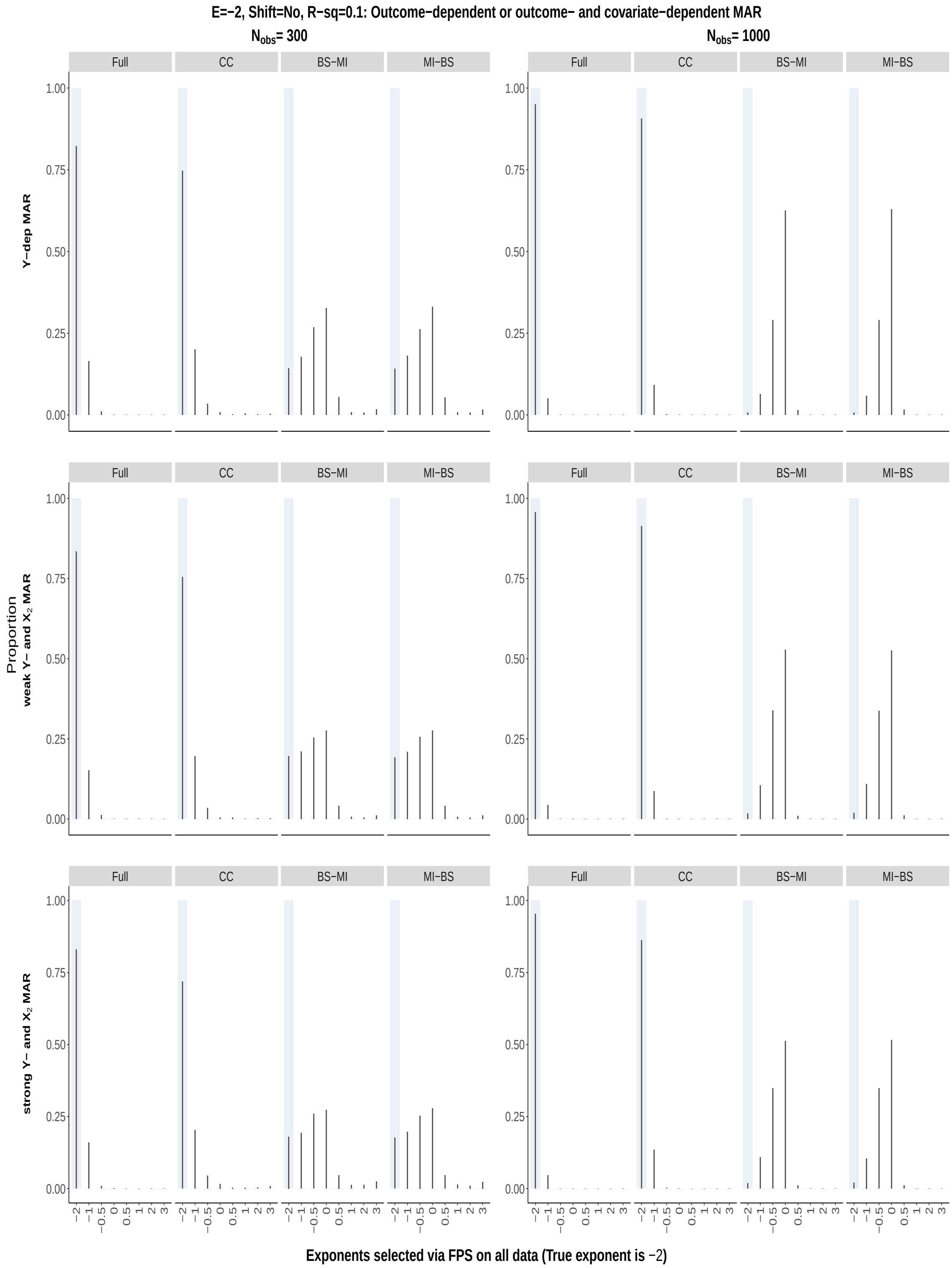

Figure S131: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

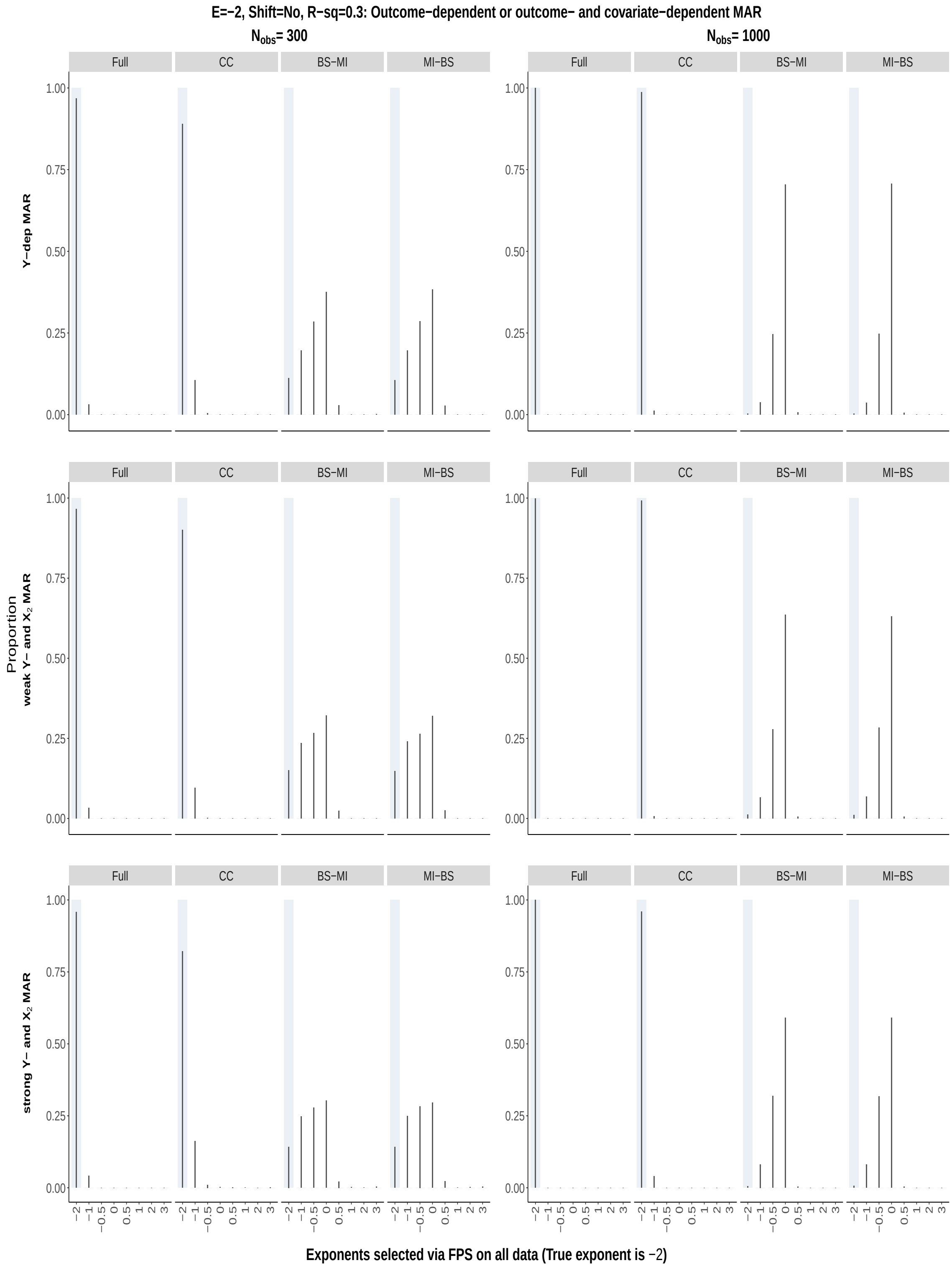

Figure S132: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.10 The 0.632 bootstrap, exponents selected using all the data: $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

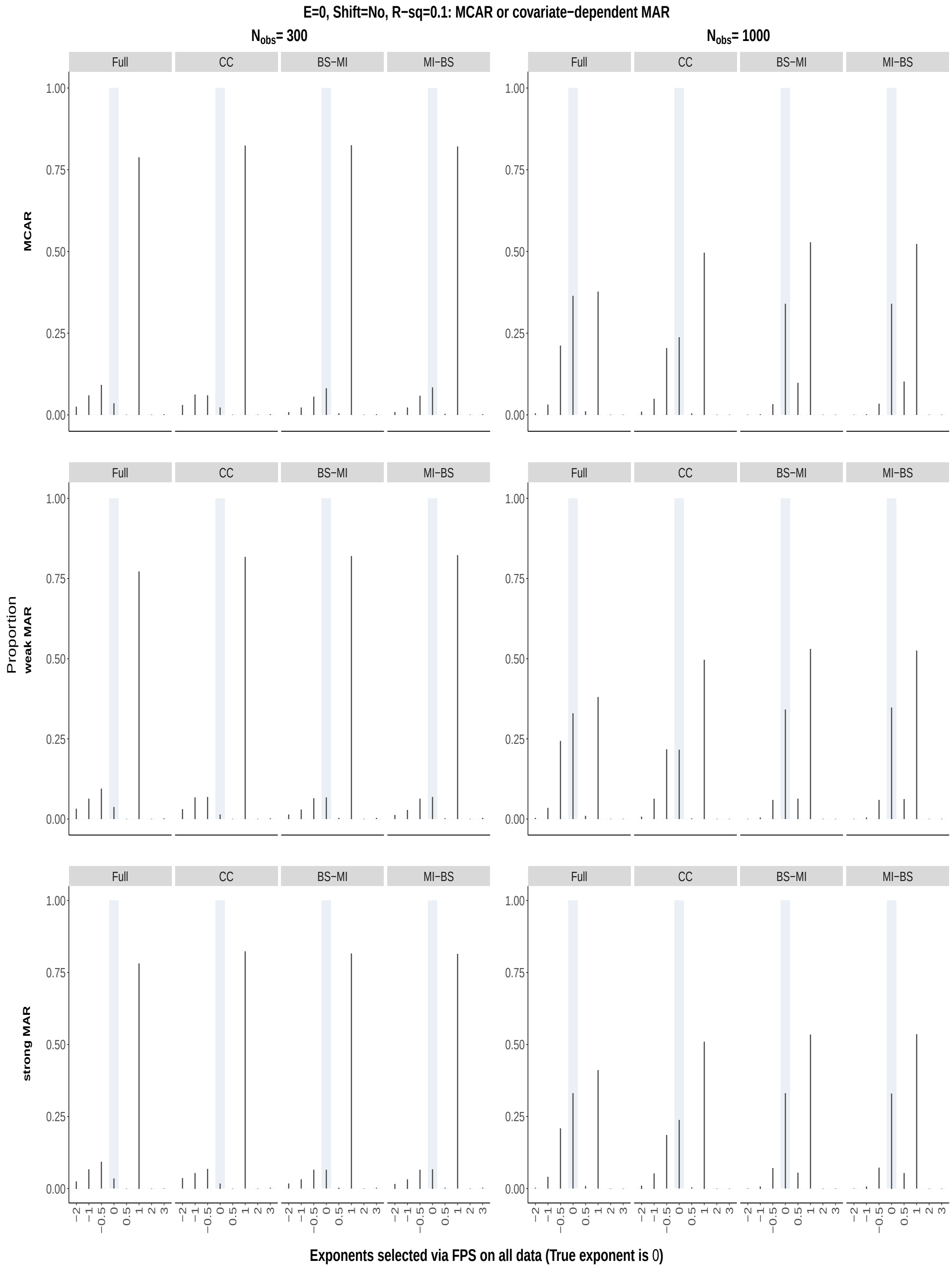

Figure S133: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

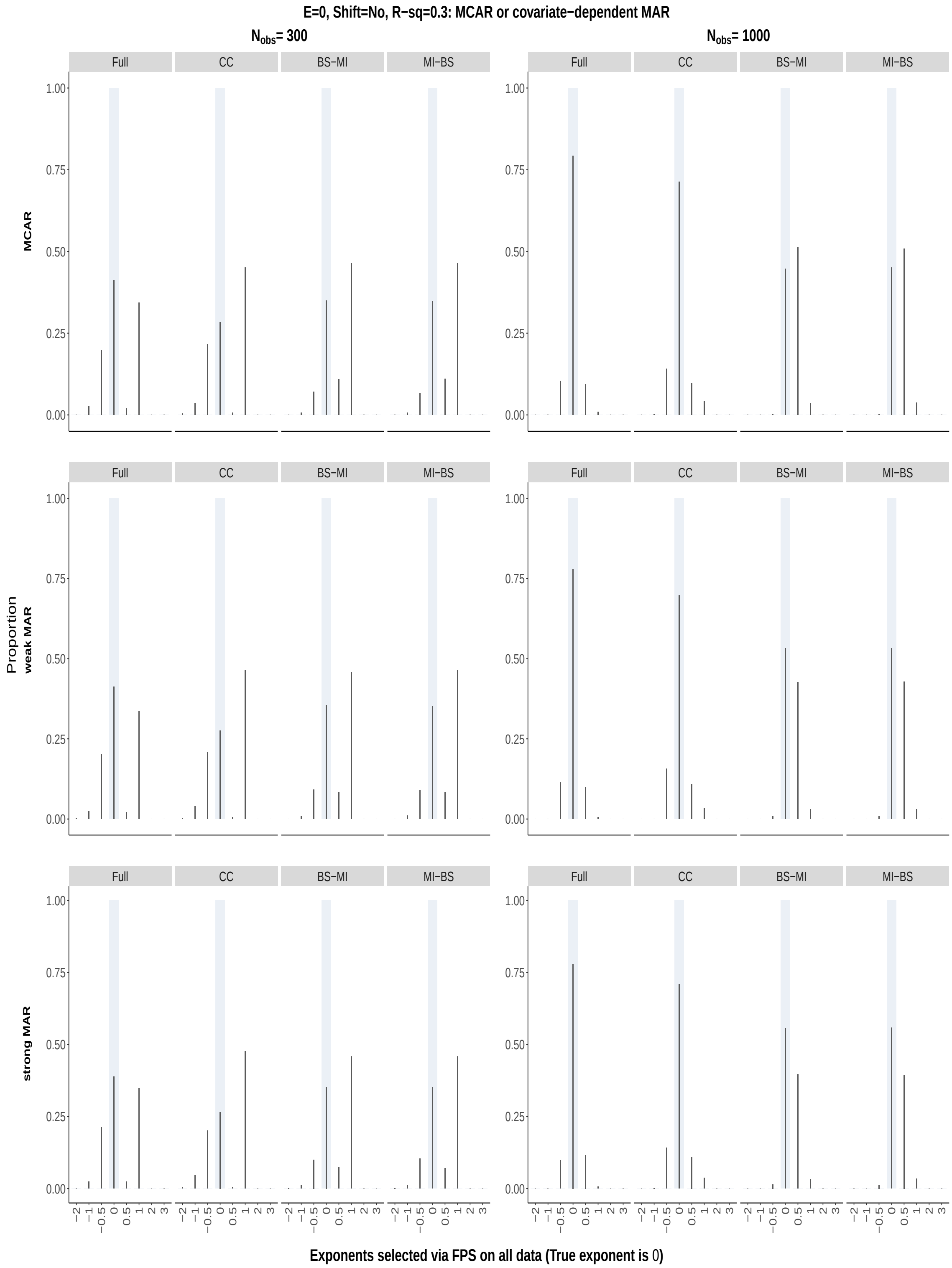

Figure S134: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

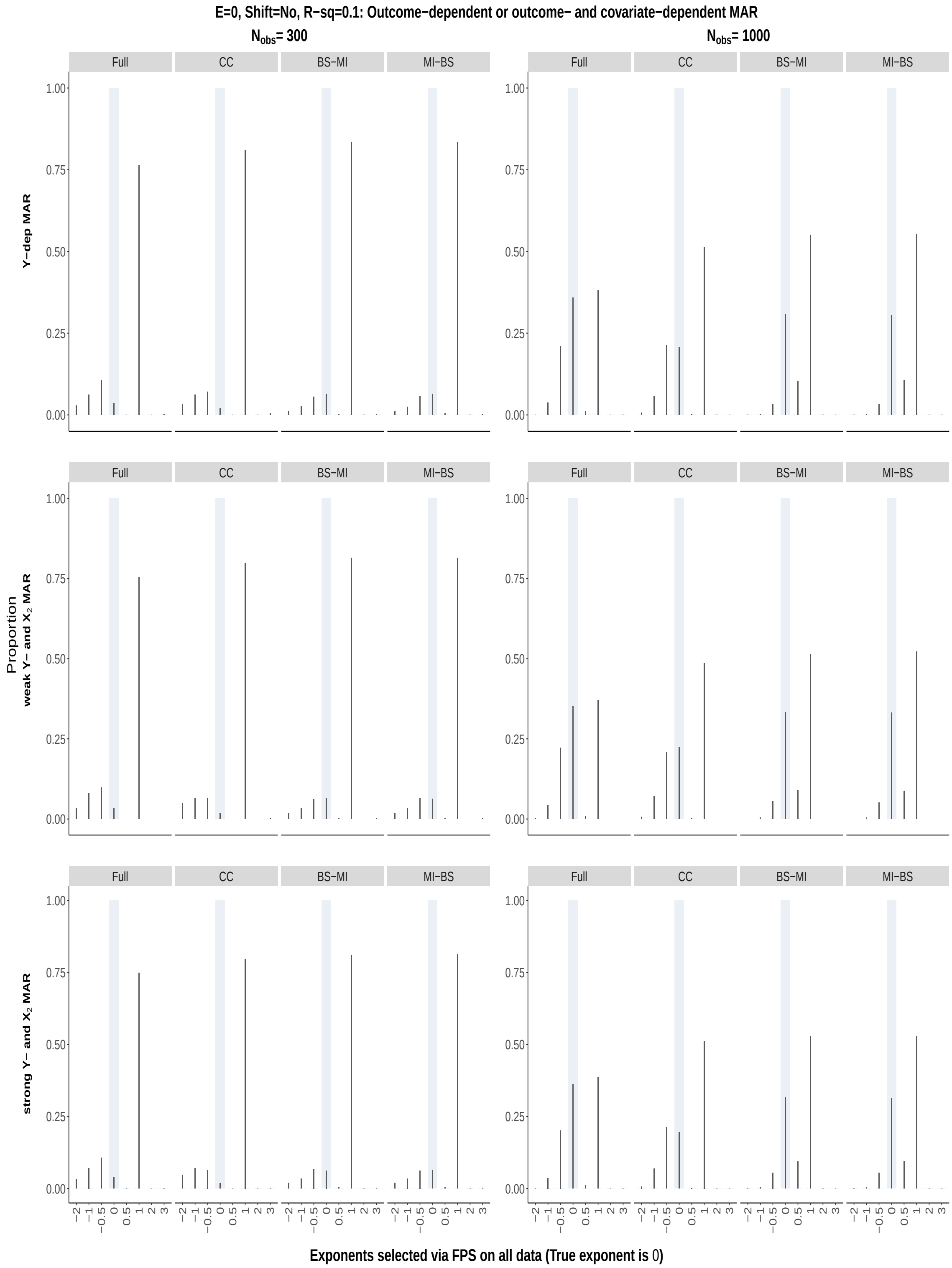

Figure S135: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

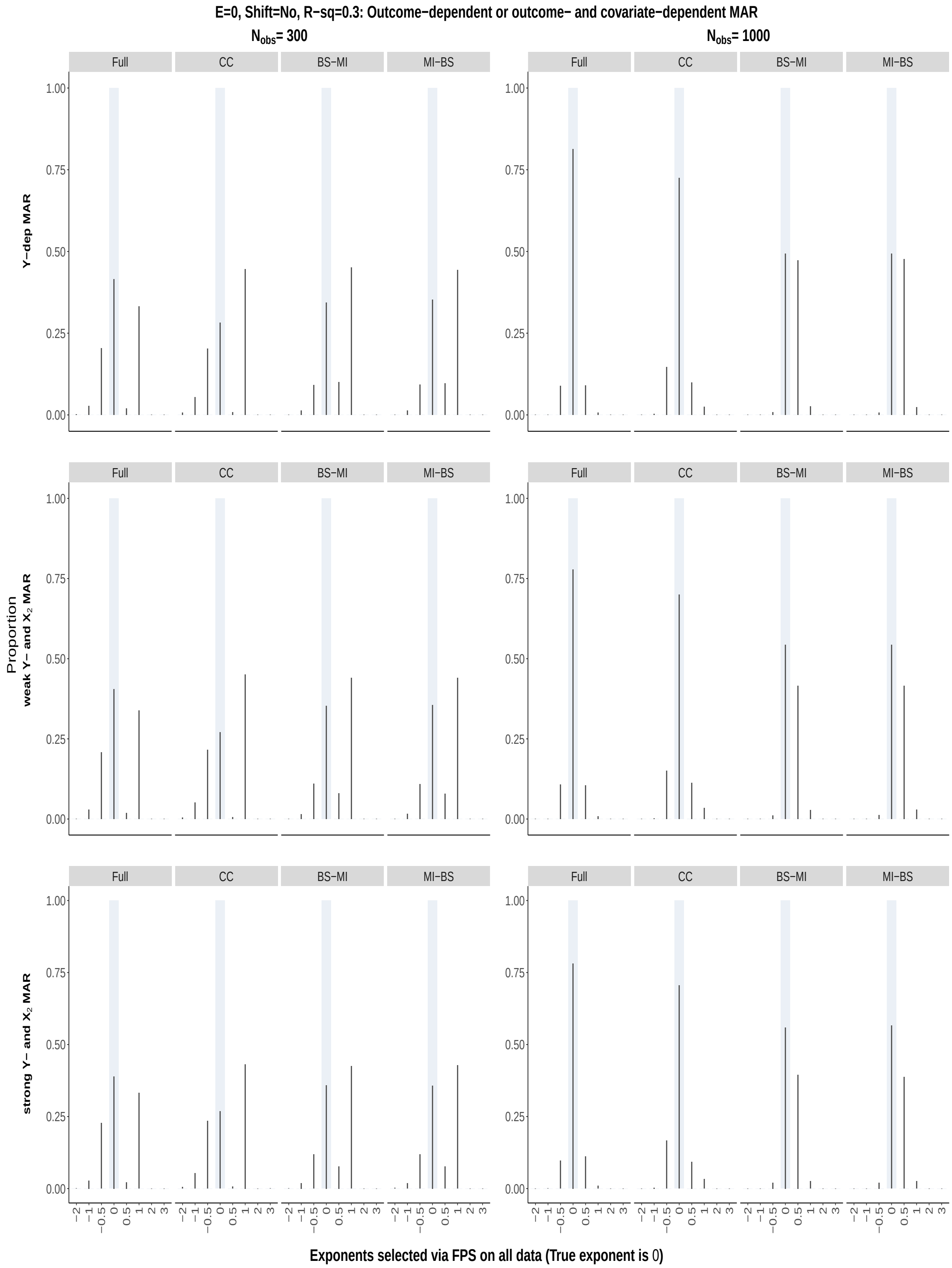

Figure S136: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

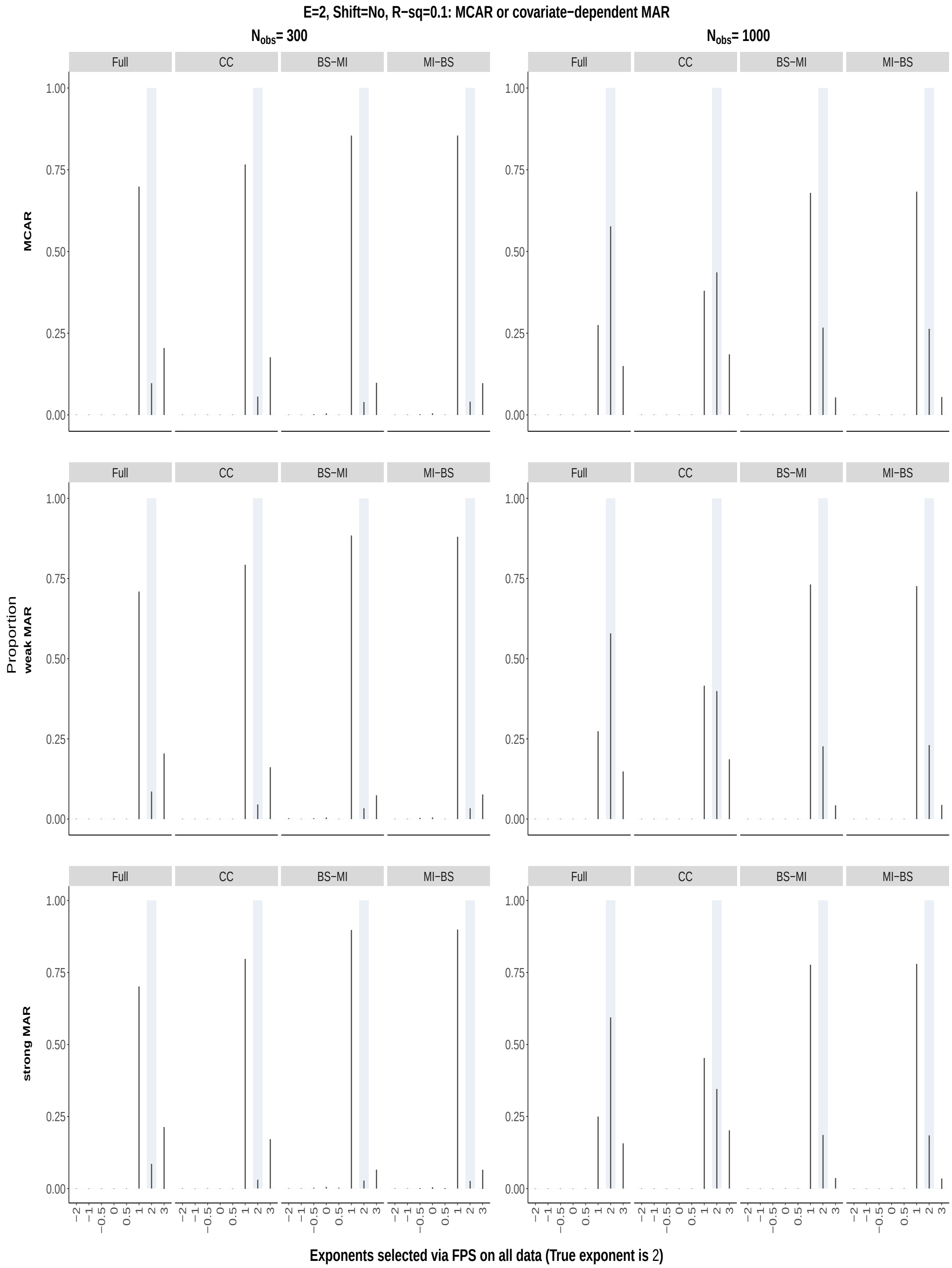

Figure S137: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

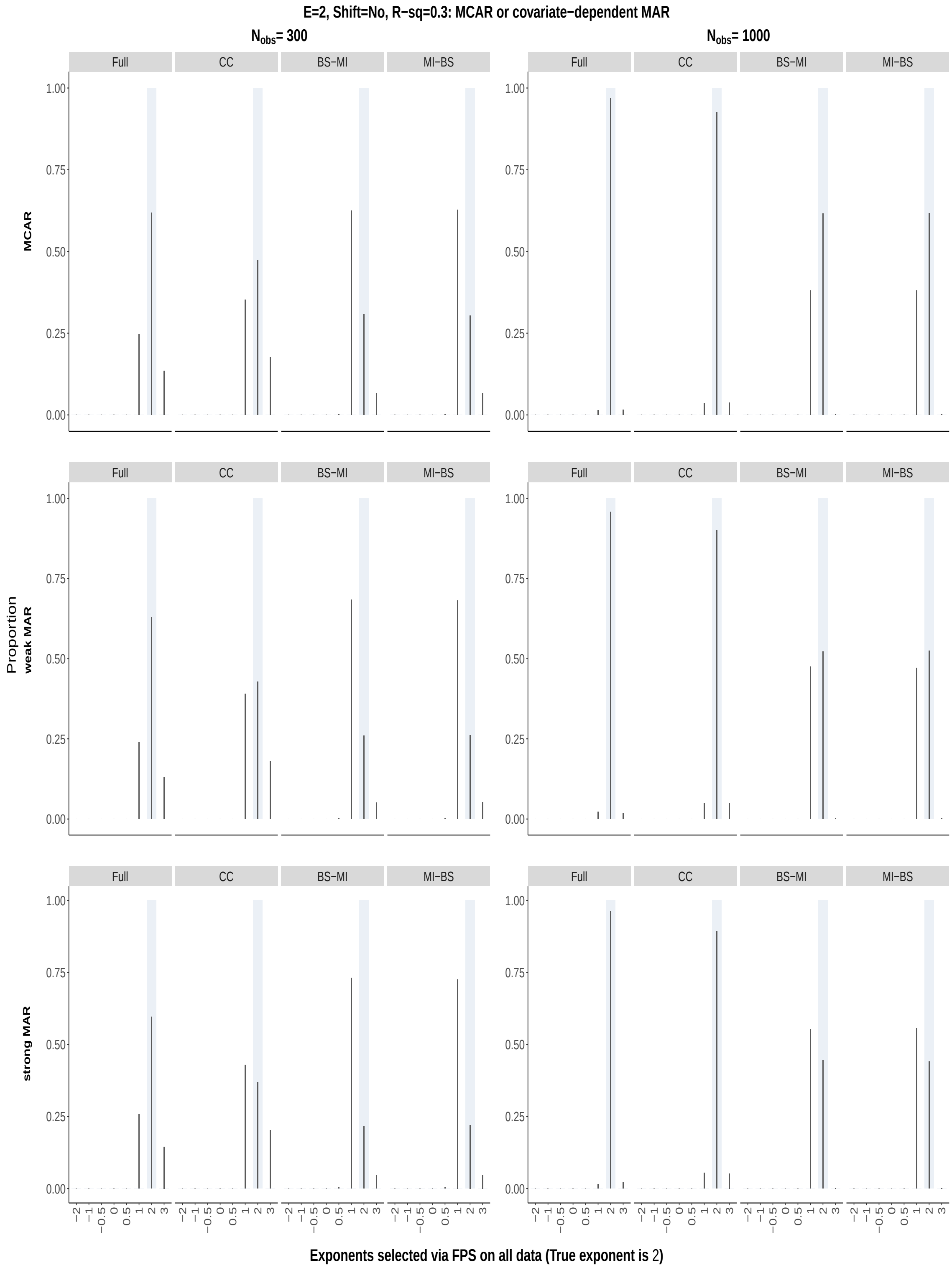

Figure S138: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

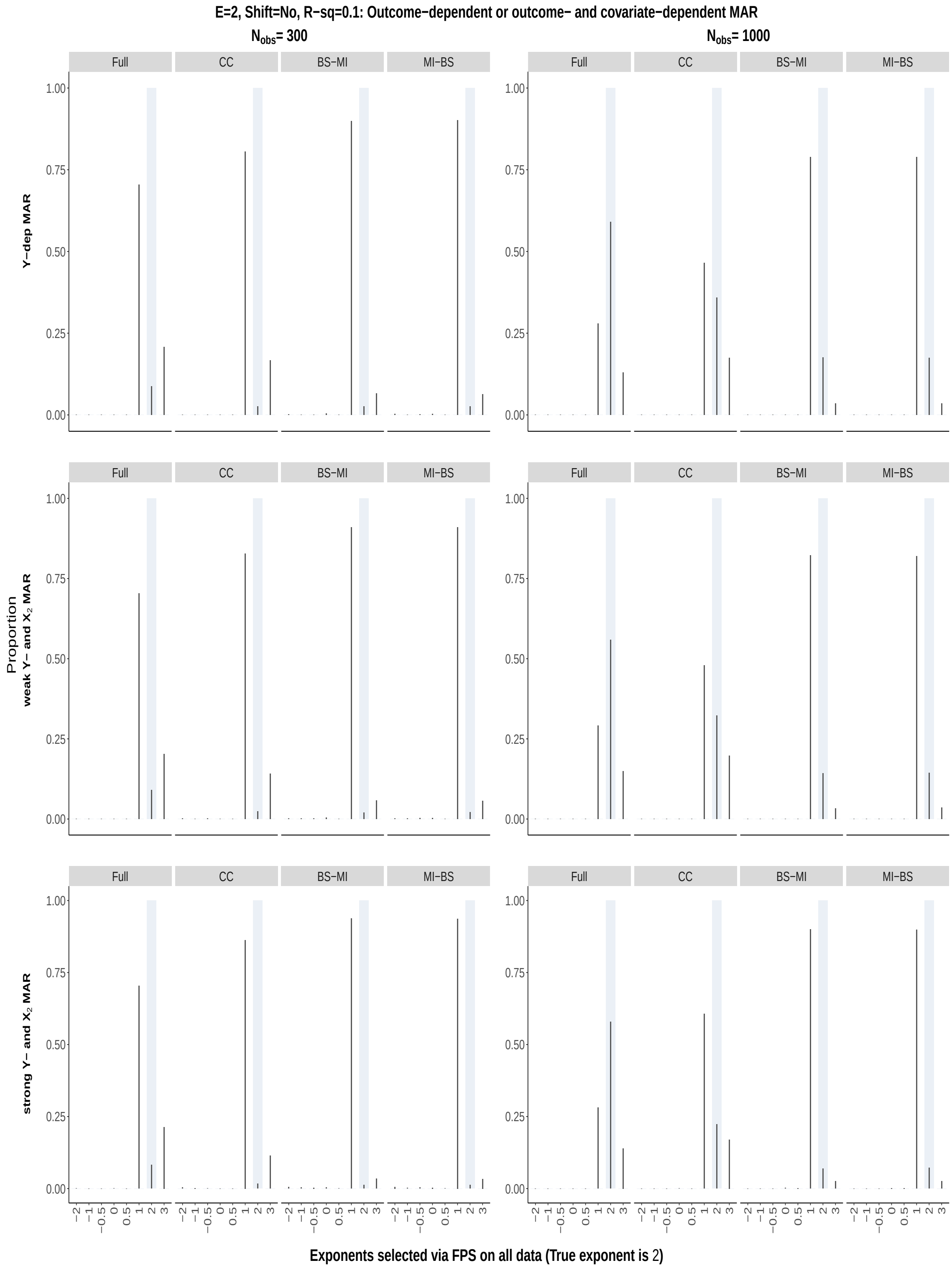

Figure S139: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

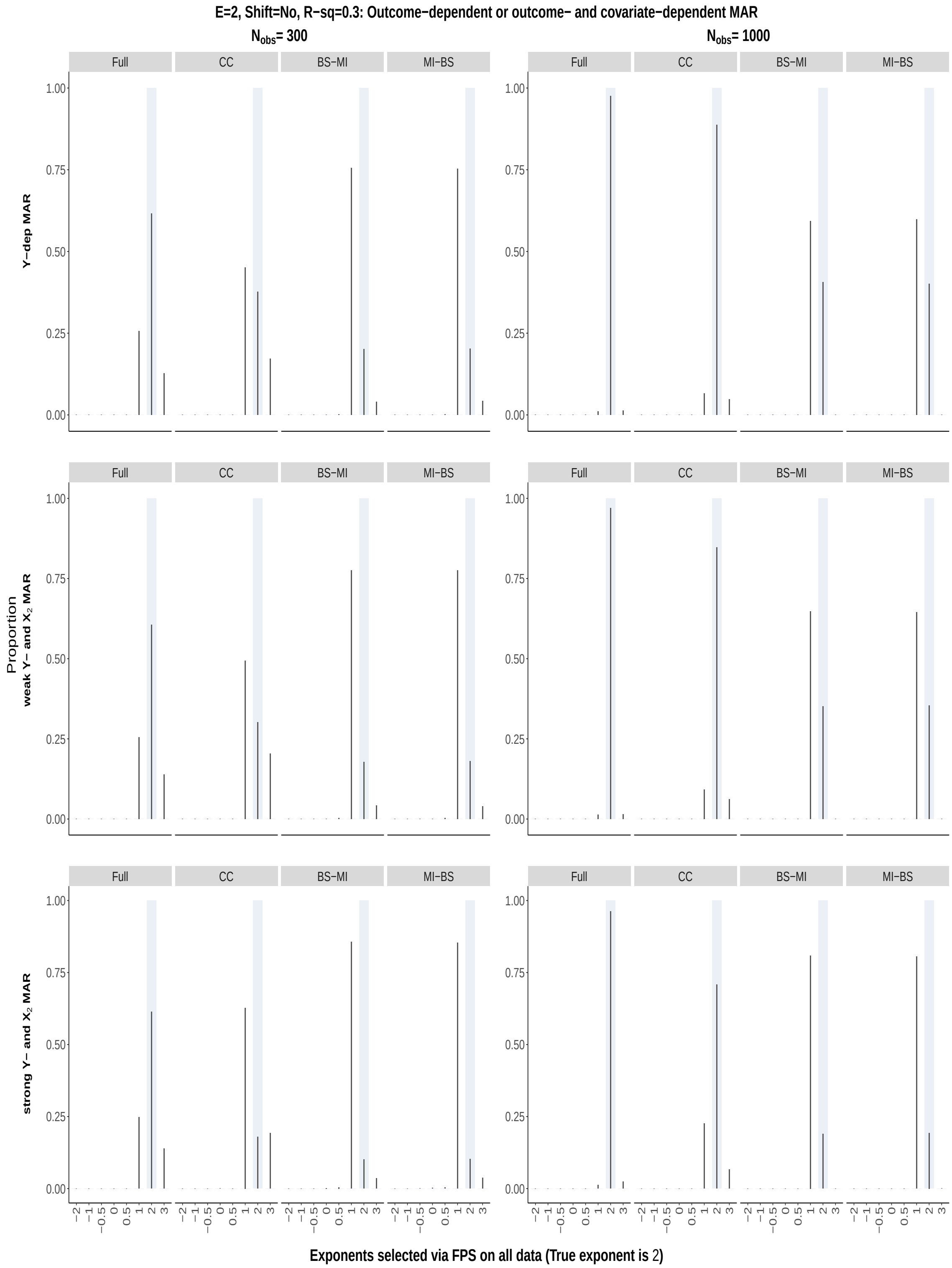

Figure S140: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

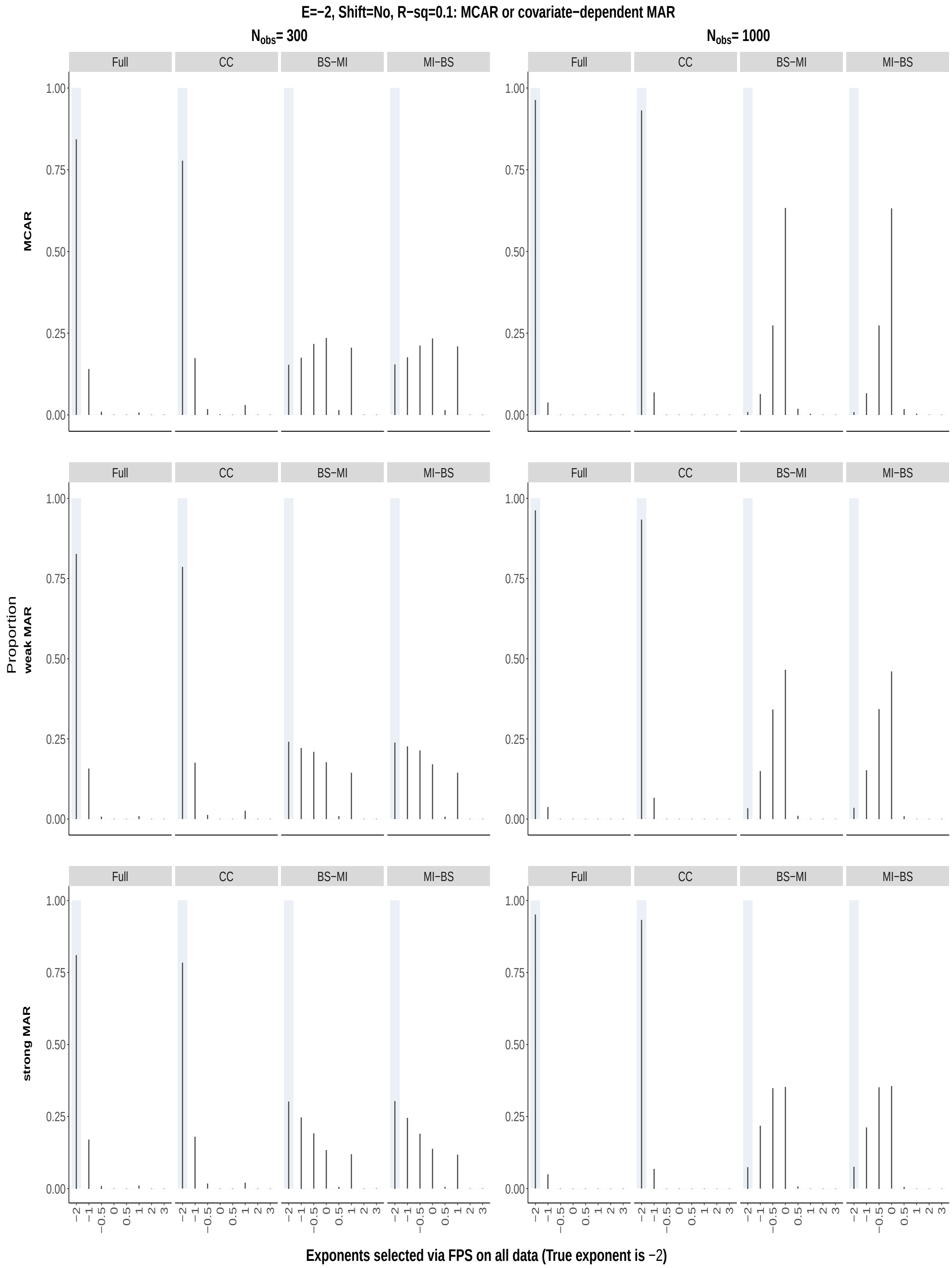

Figure S141: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

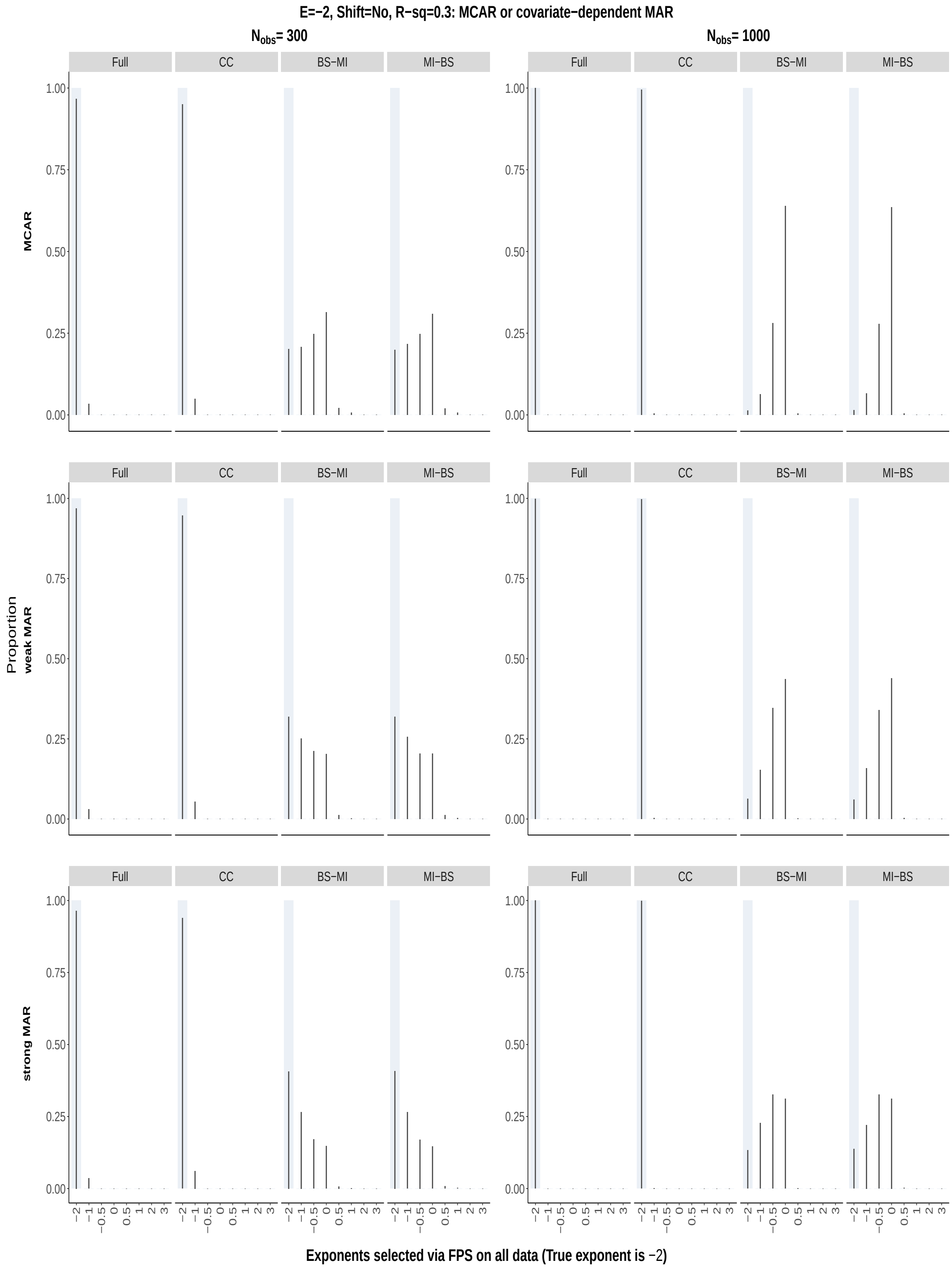

Figure S142: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

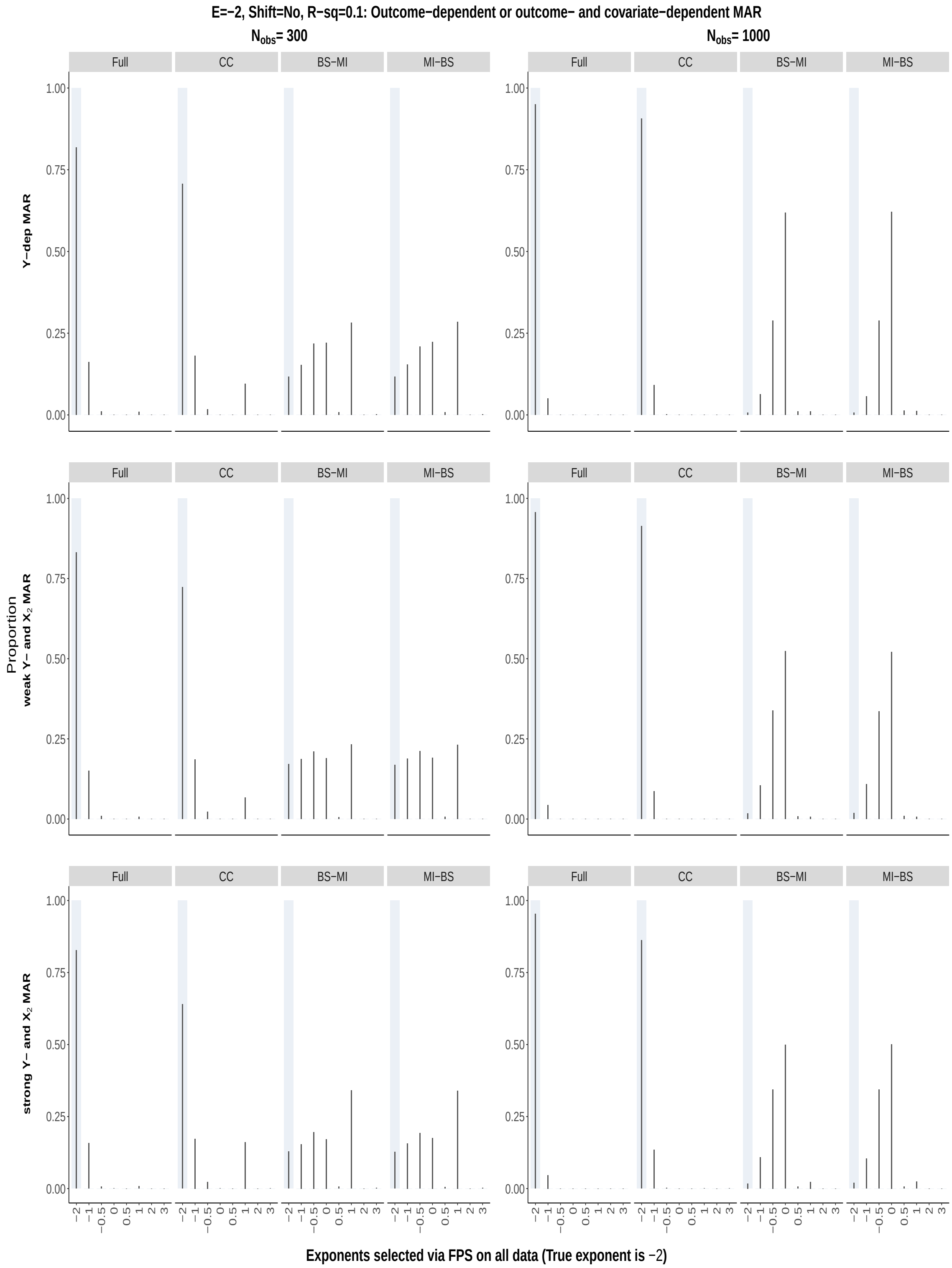

Figure S143: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

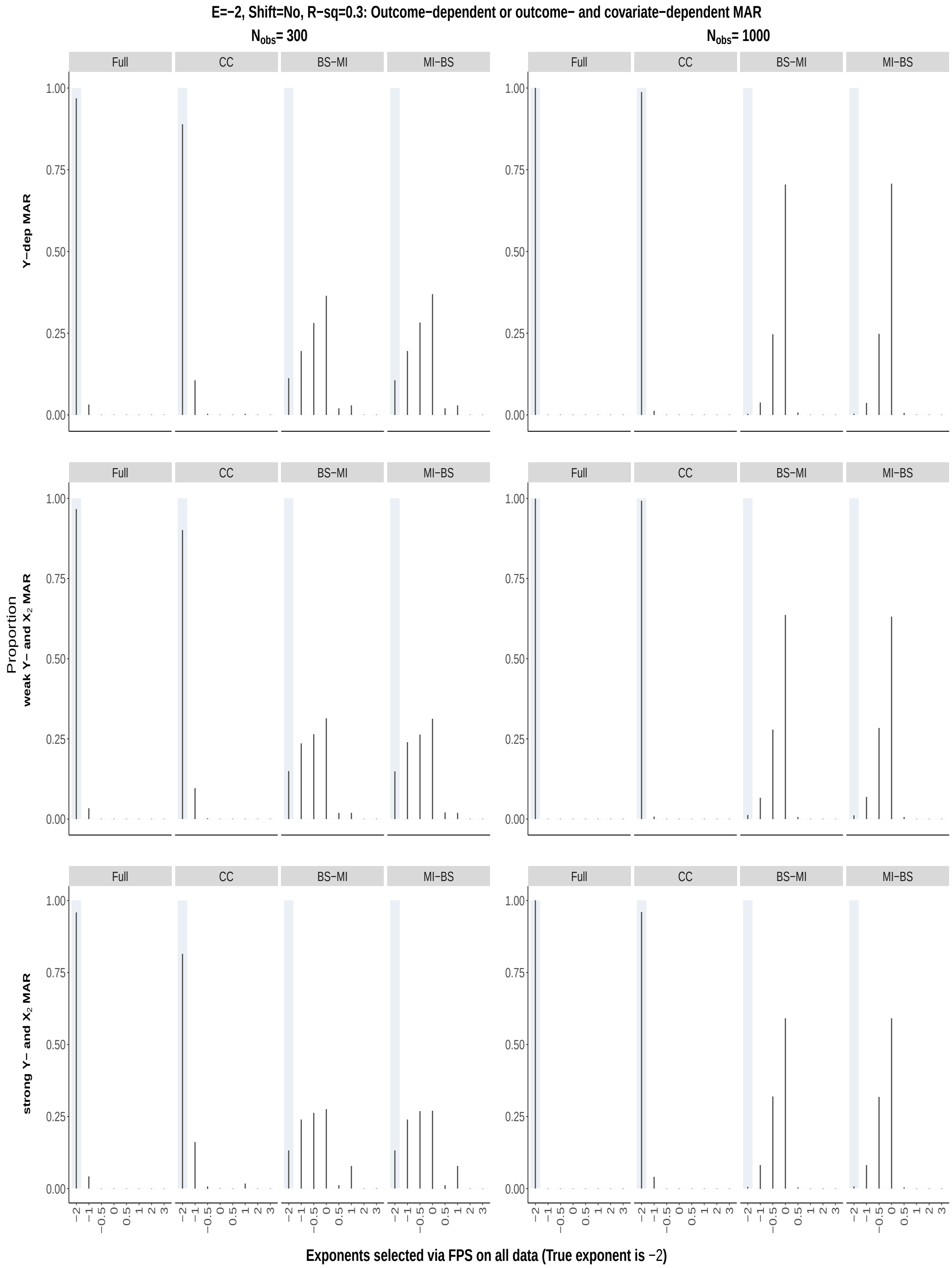

Figure S144: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.11 The 0.632 bootstrap, exponents selected using all the data: $\alpha_E = 1$ and an origin-shift has been applied

True exponent is 0

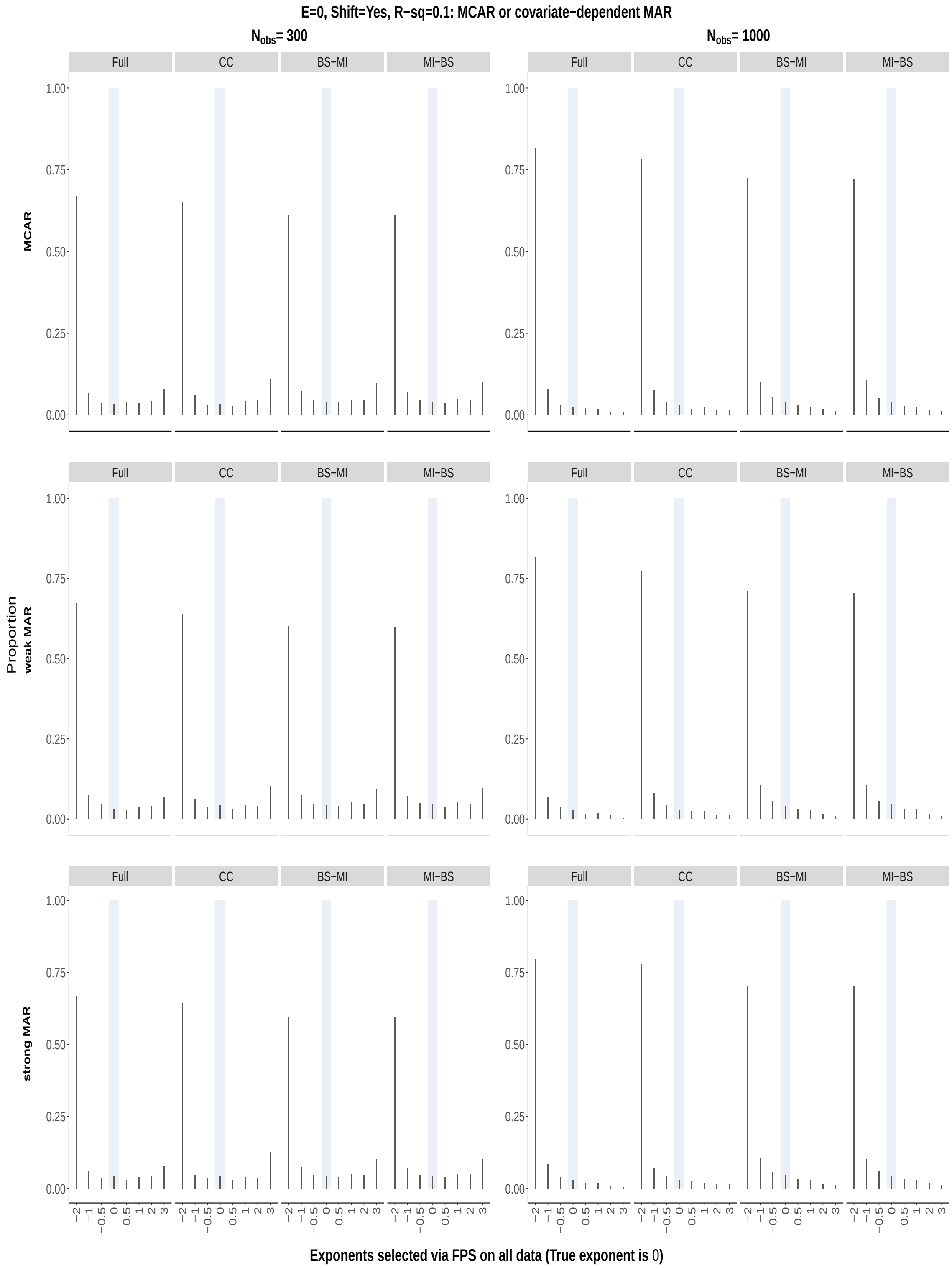

Figure S145: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

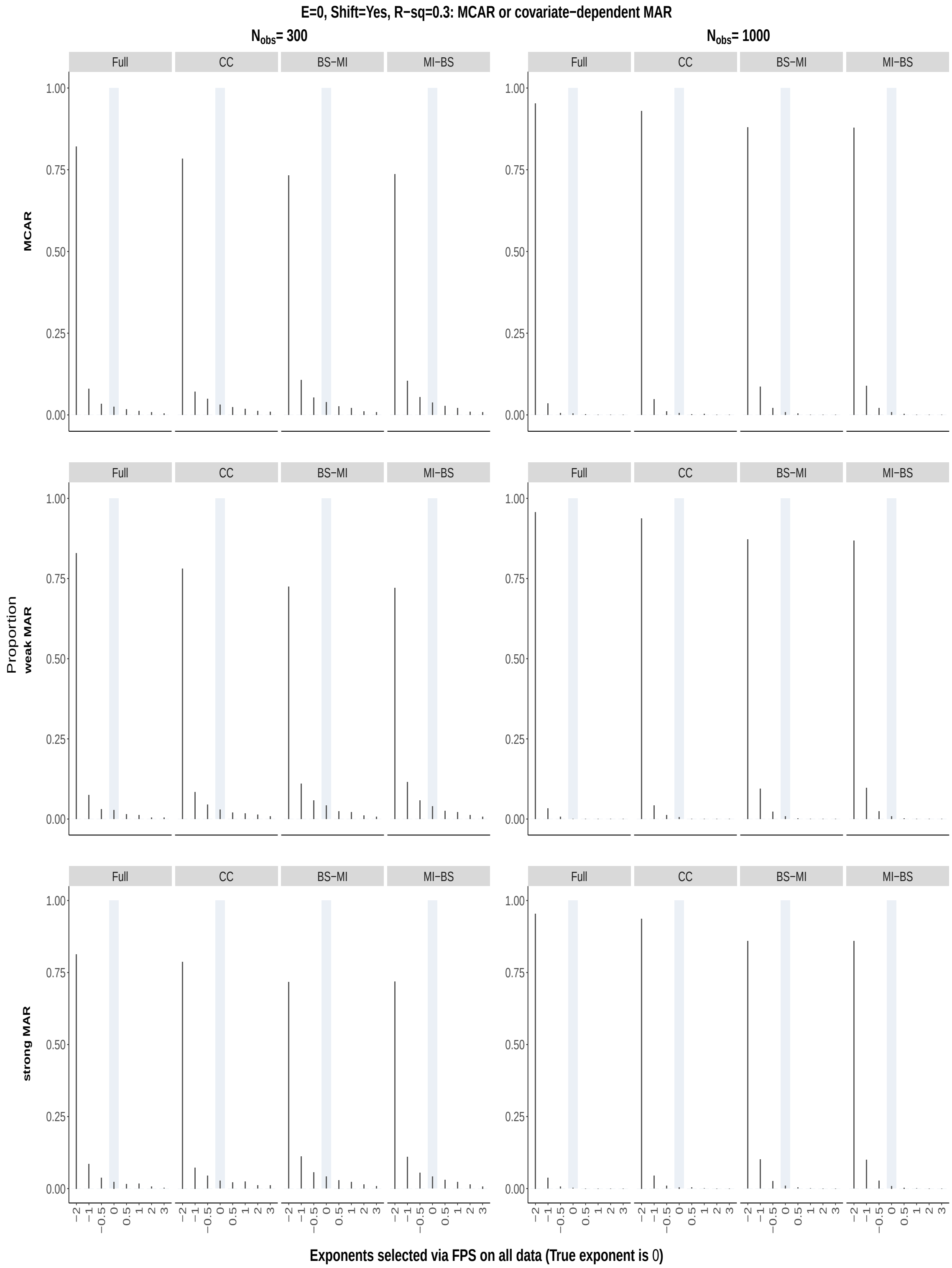

Figure S146: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

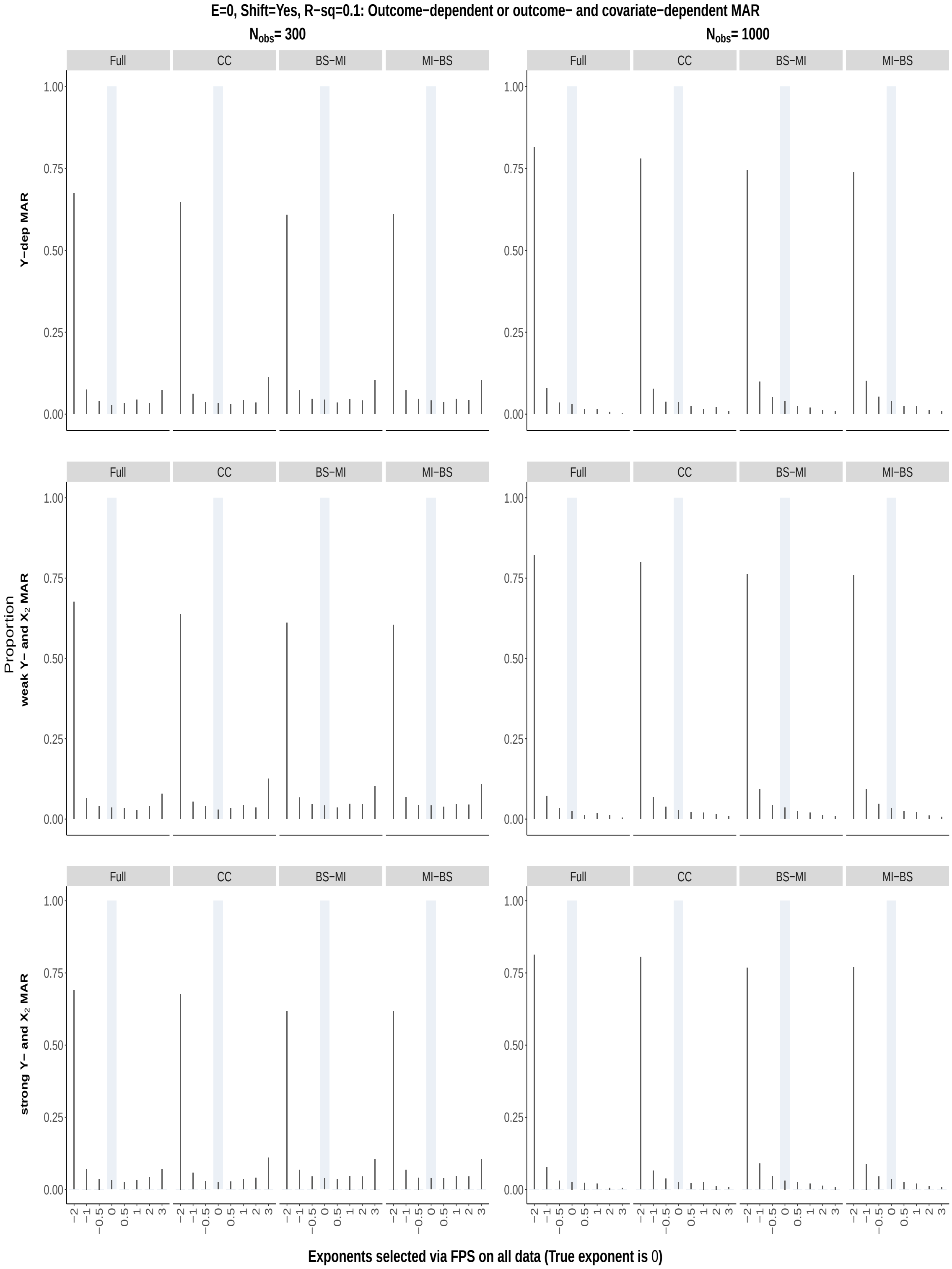

Figure S147: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

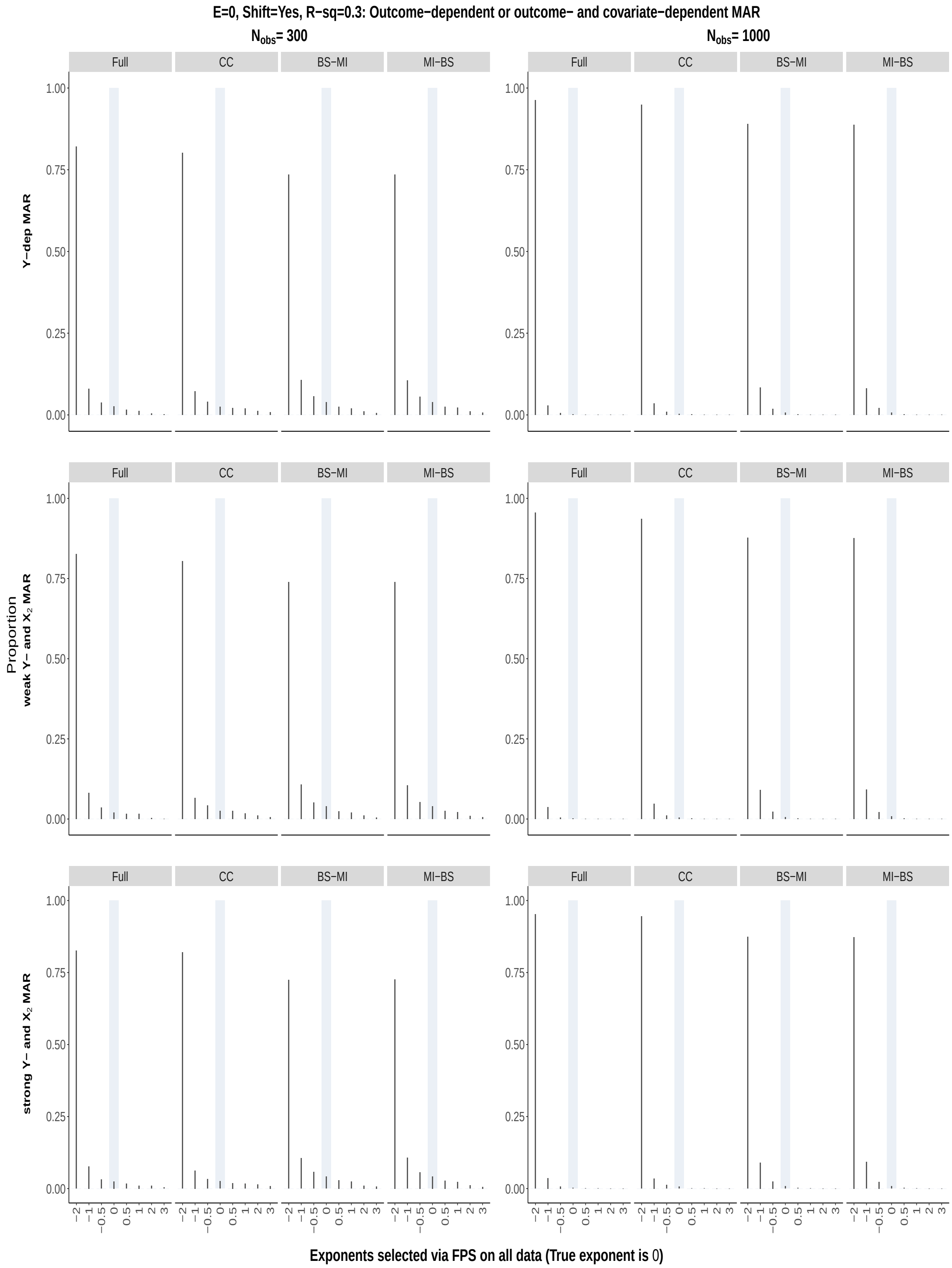

Figure S148: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

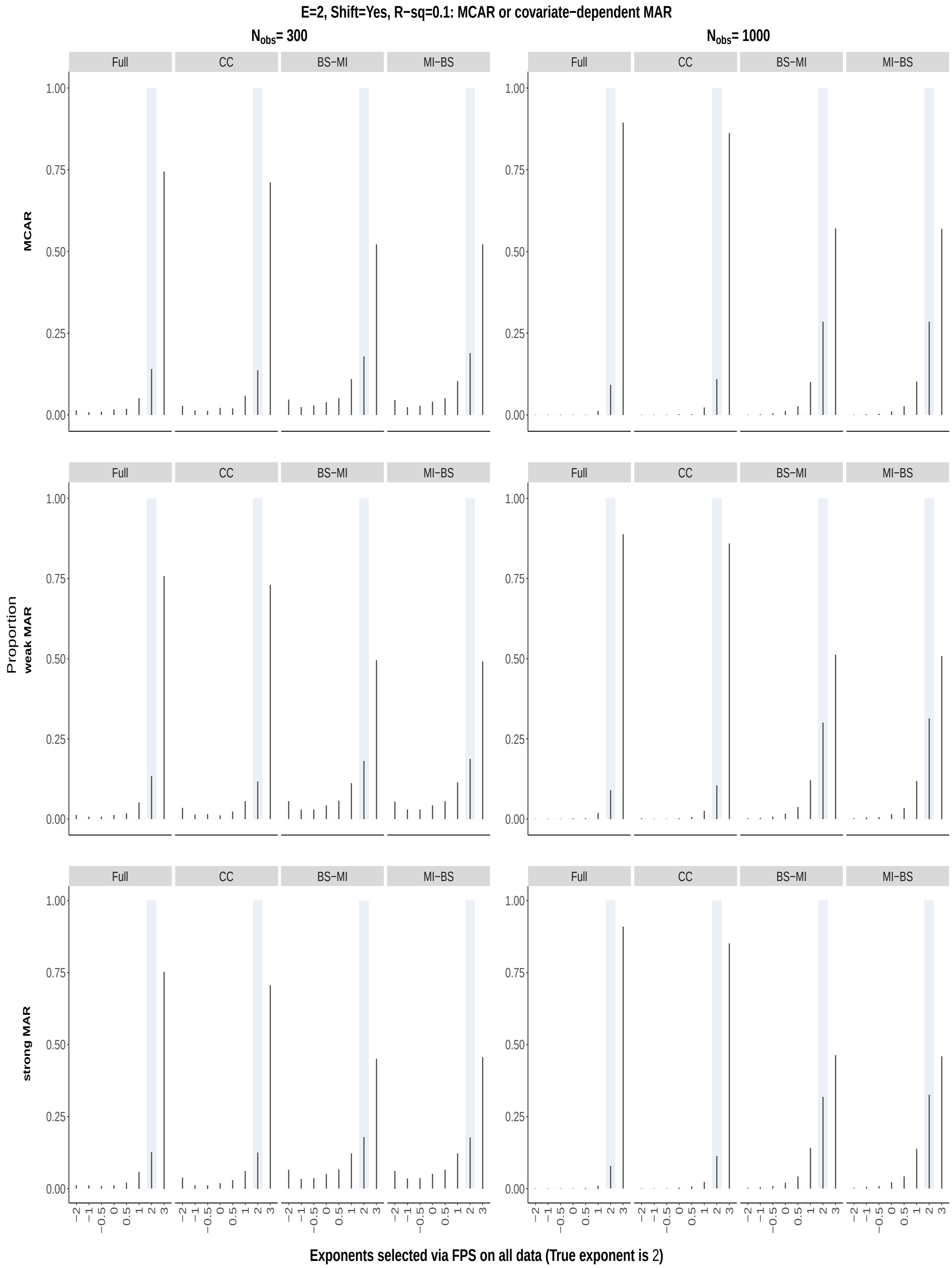

Figure S149: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

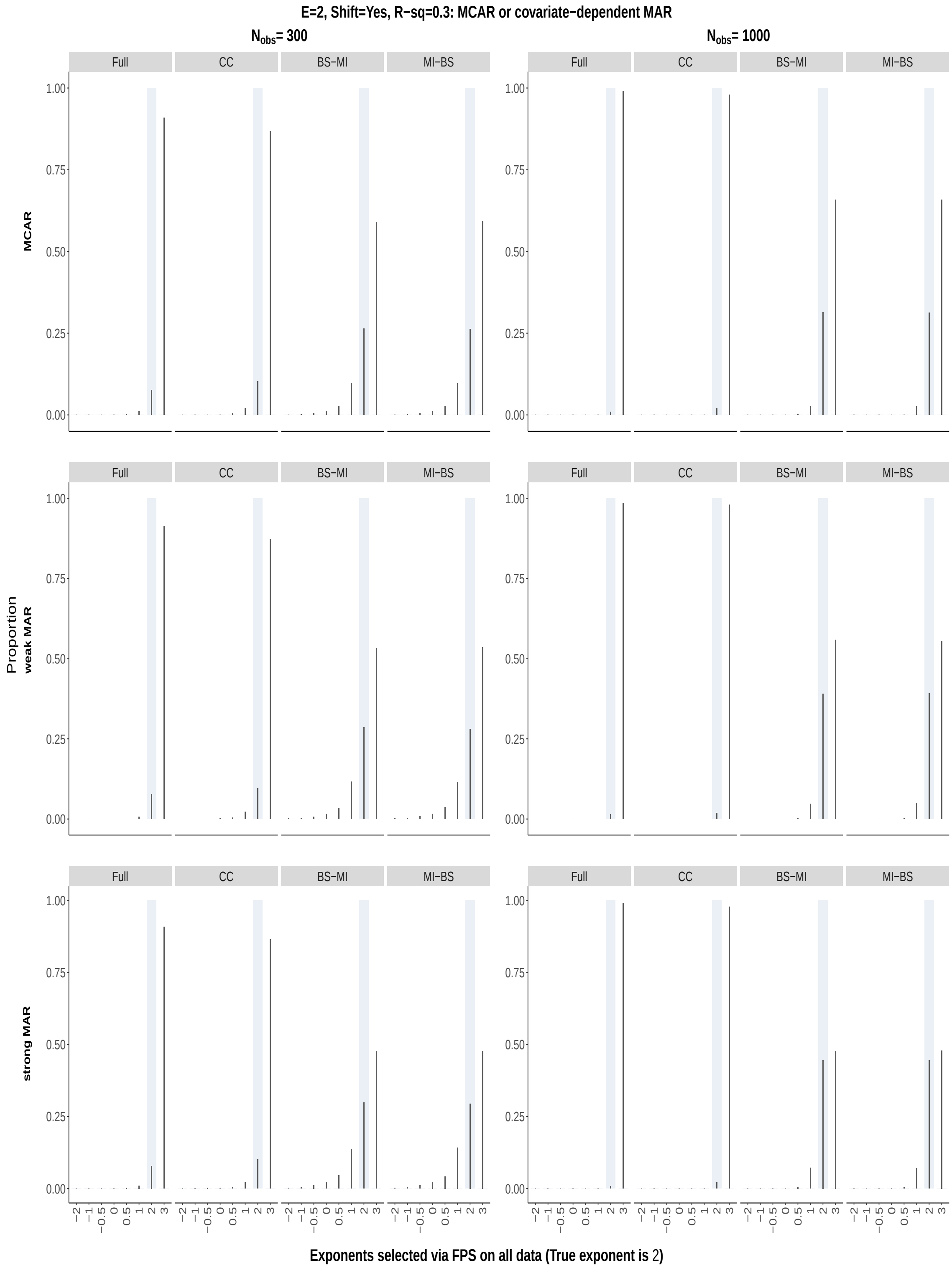

Figure S150: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

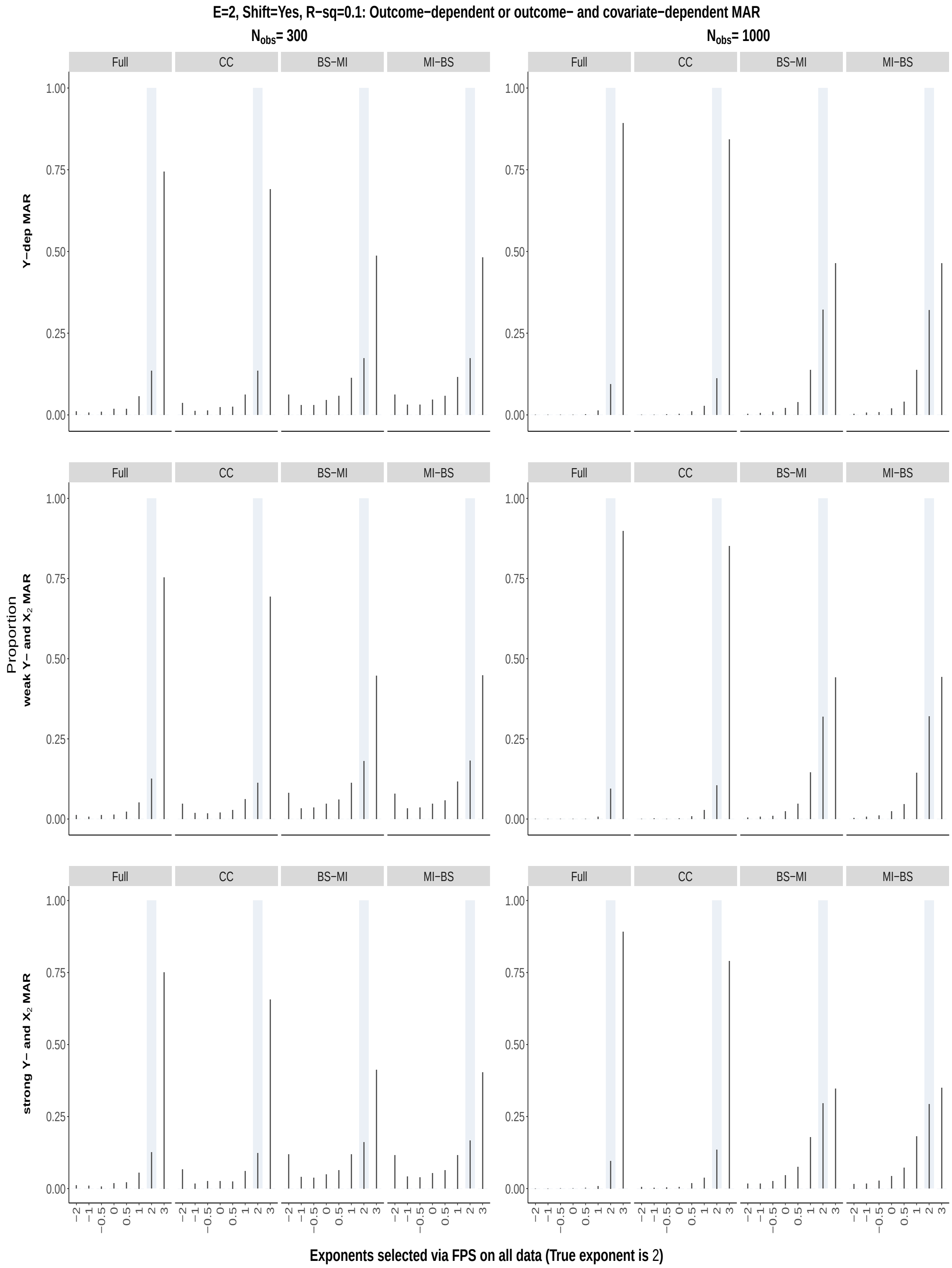

Figure S151: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

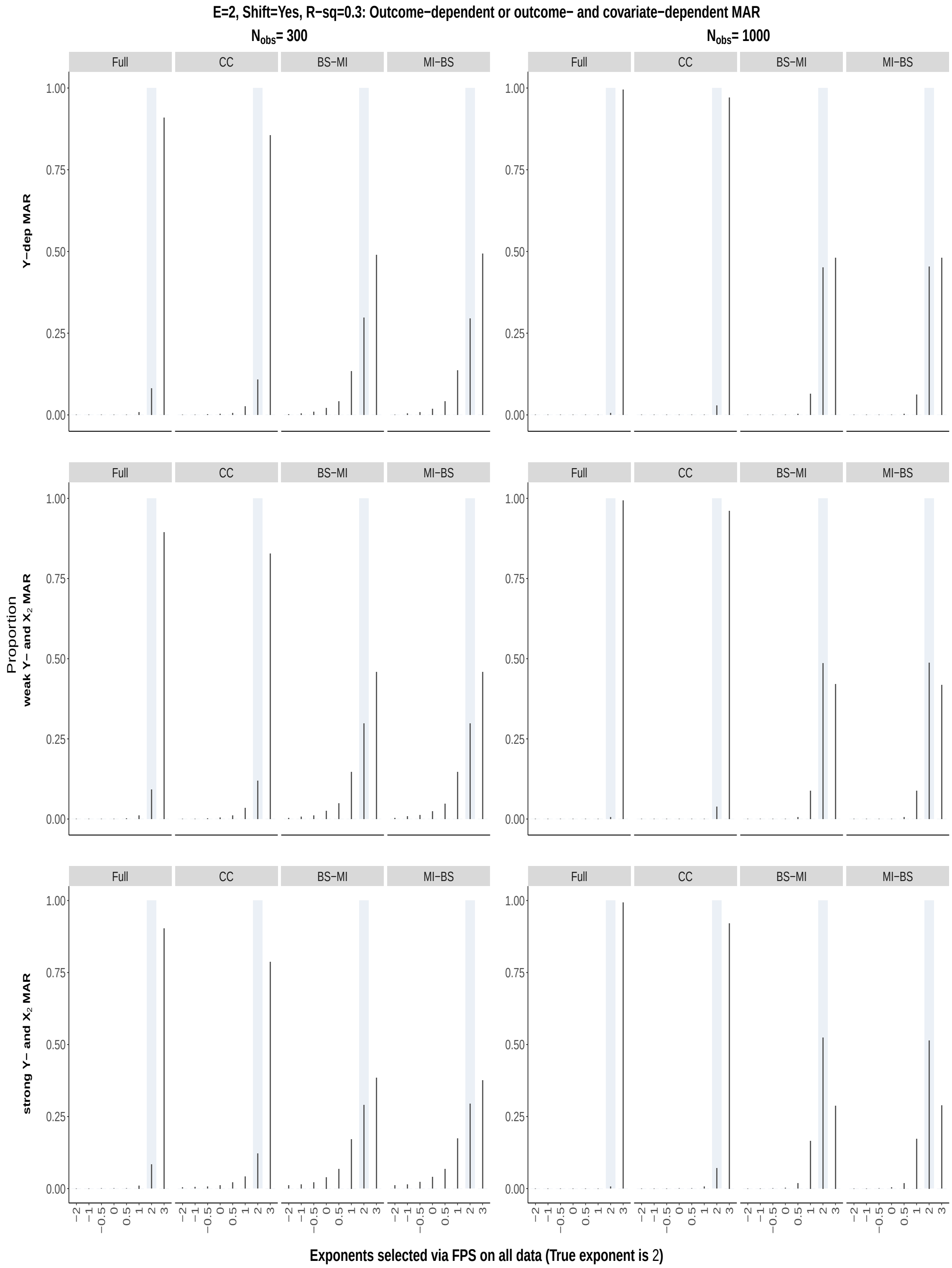

Figure S152: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

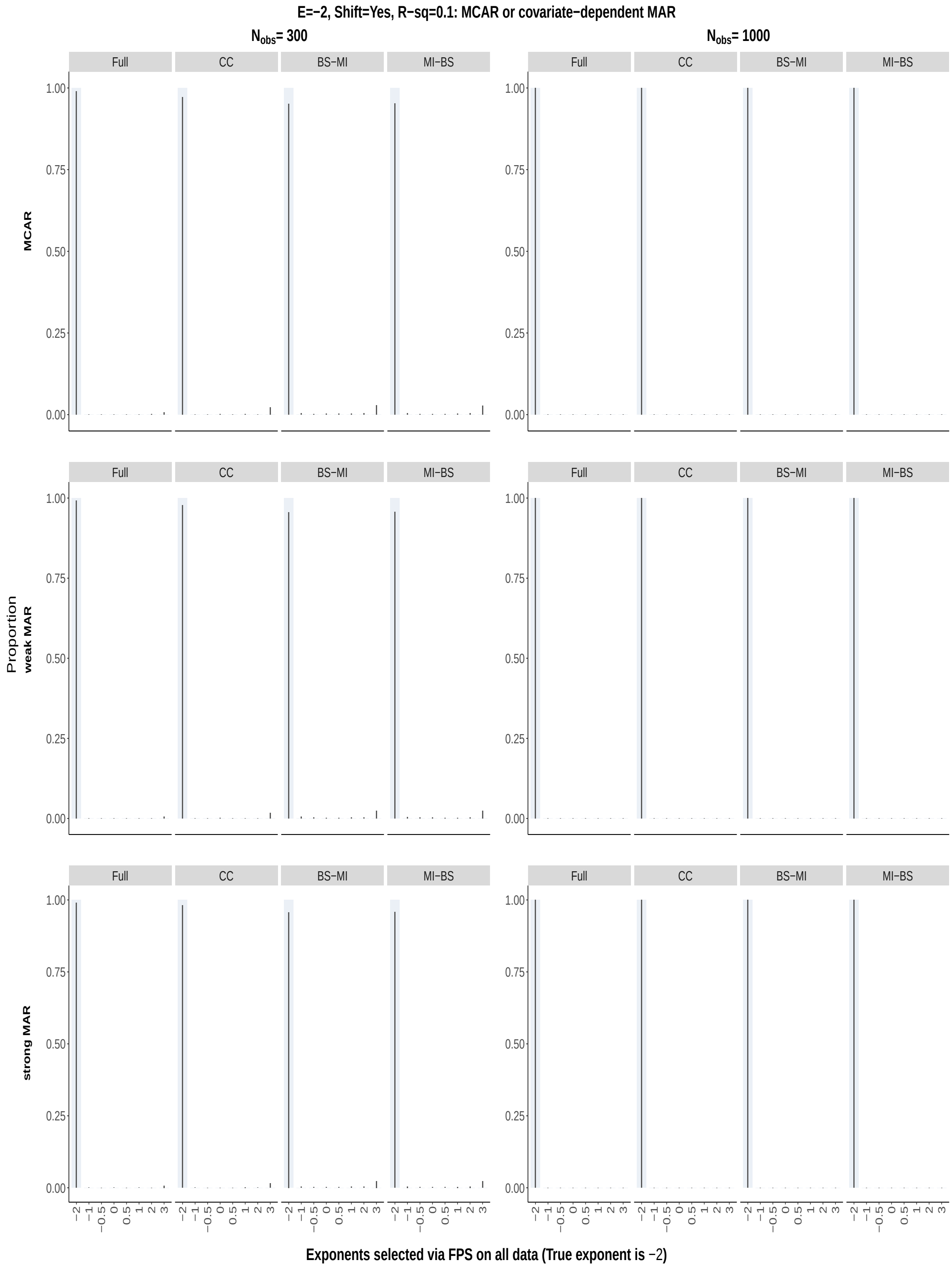

Figure S153: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

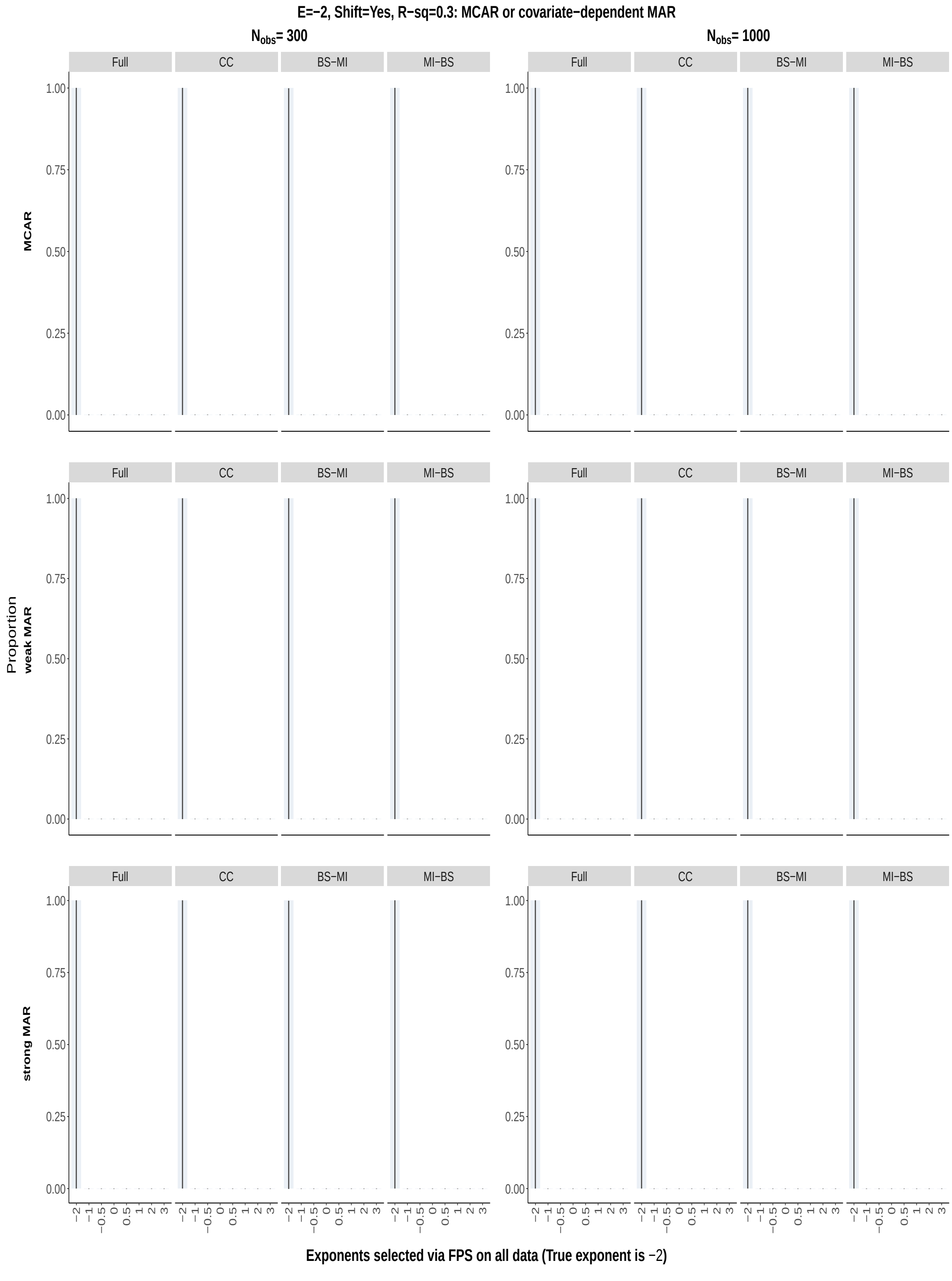

Figure S154: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

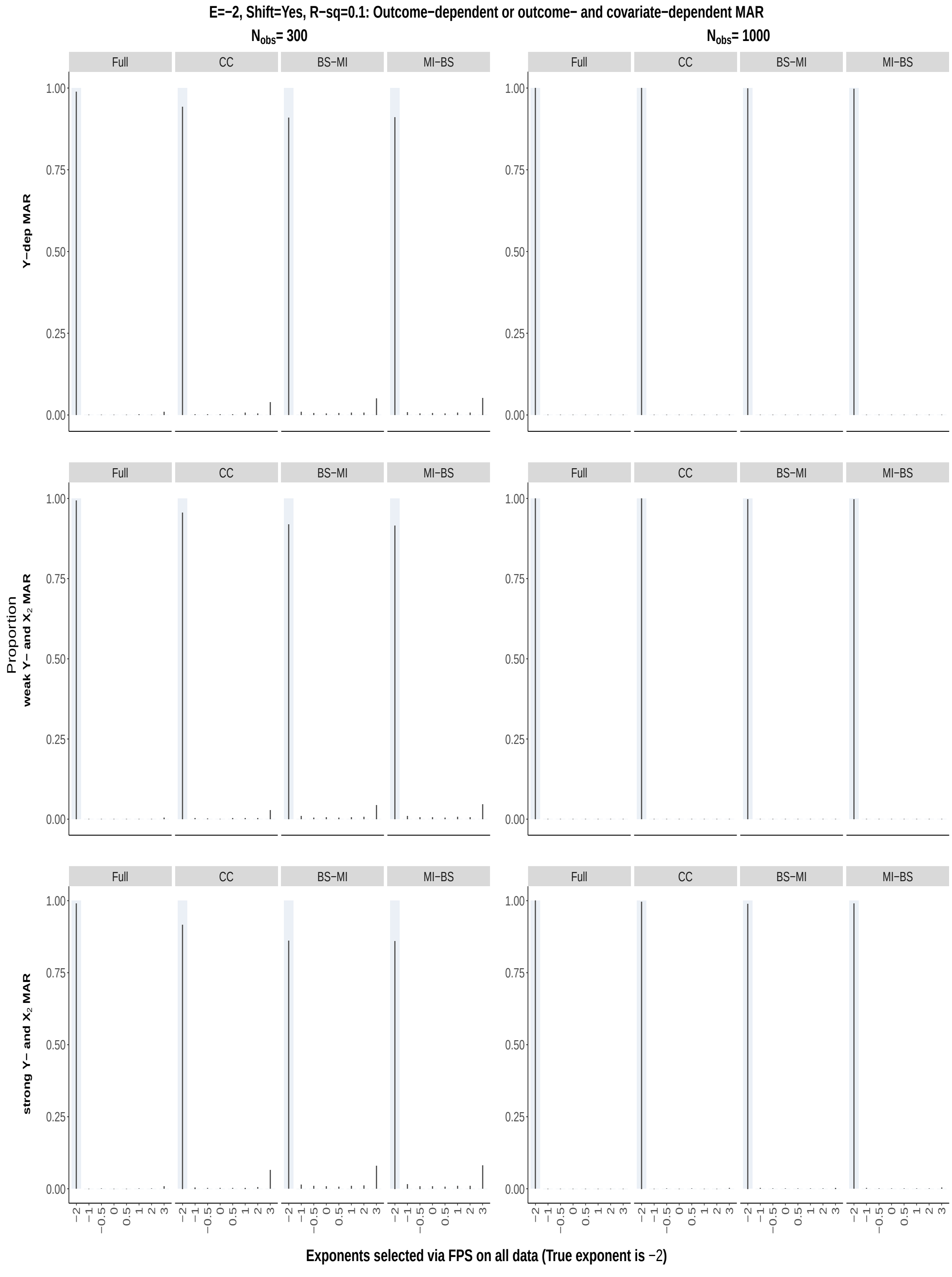

Figure S155: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

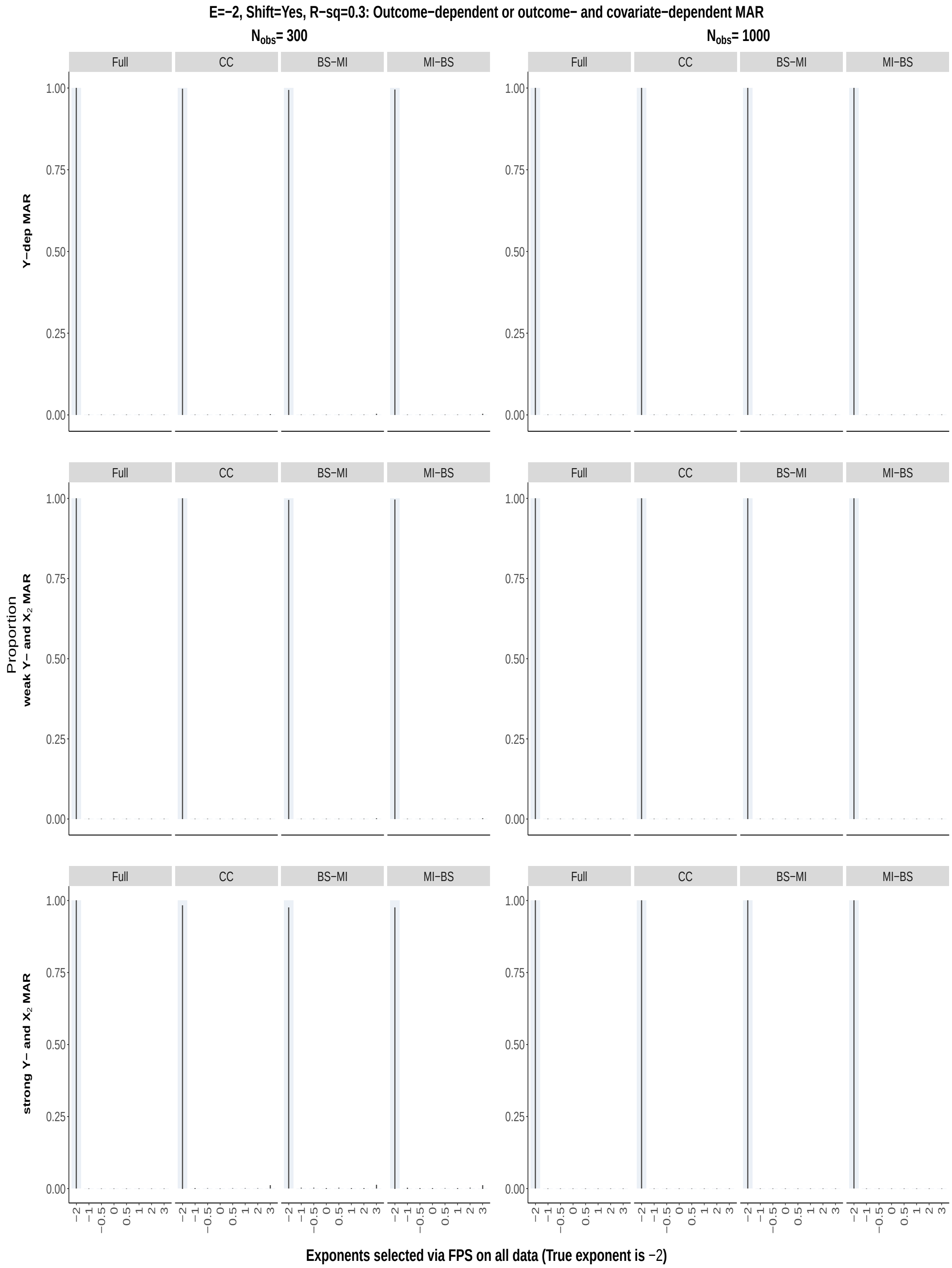

Figure S156: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S6.2.12 The 0.632 bootstrap, exponents selected using all the data: $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

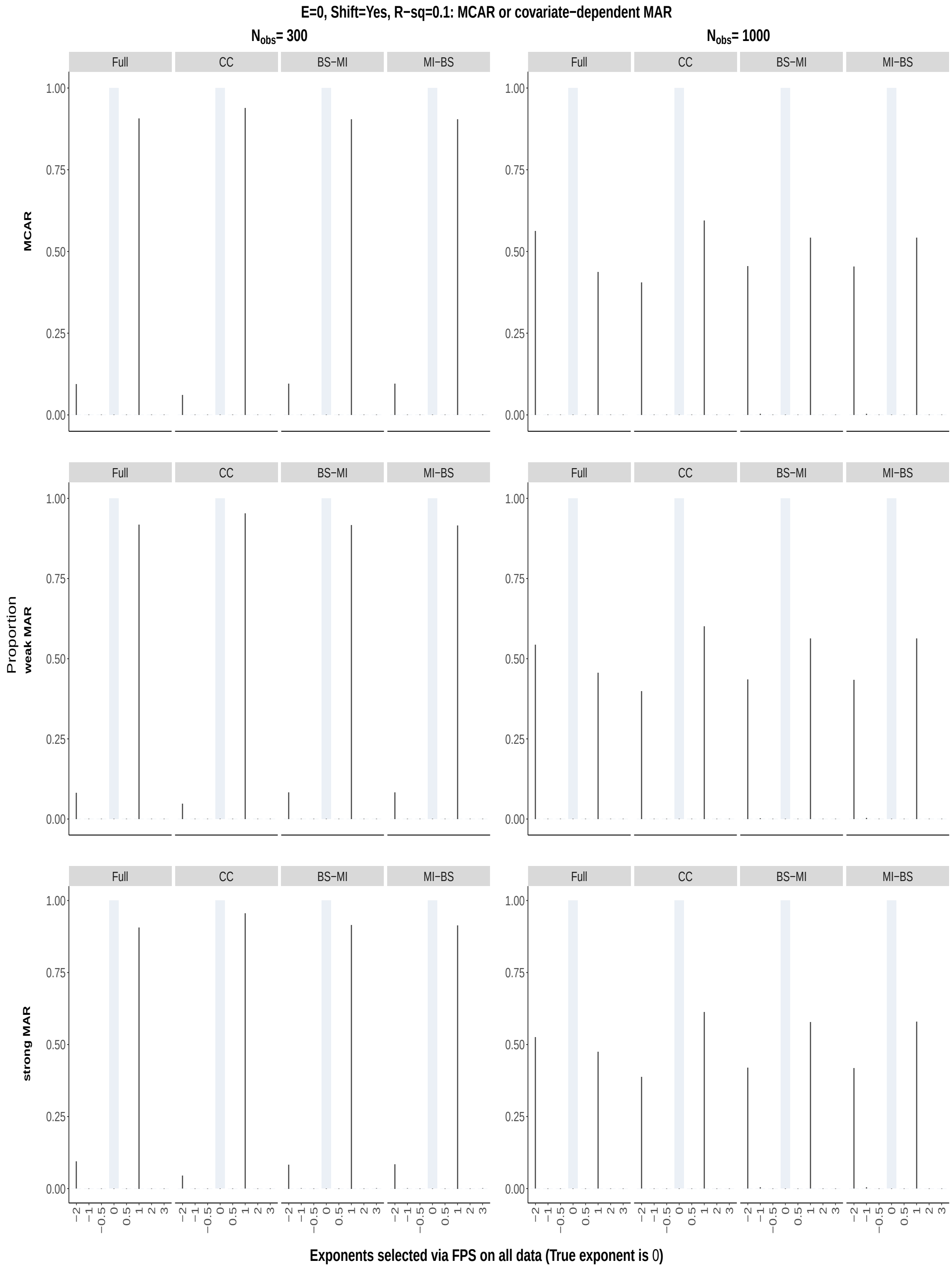

Figure S157: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

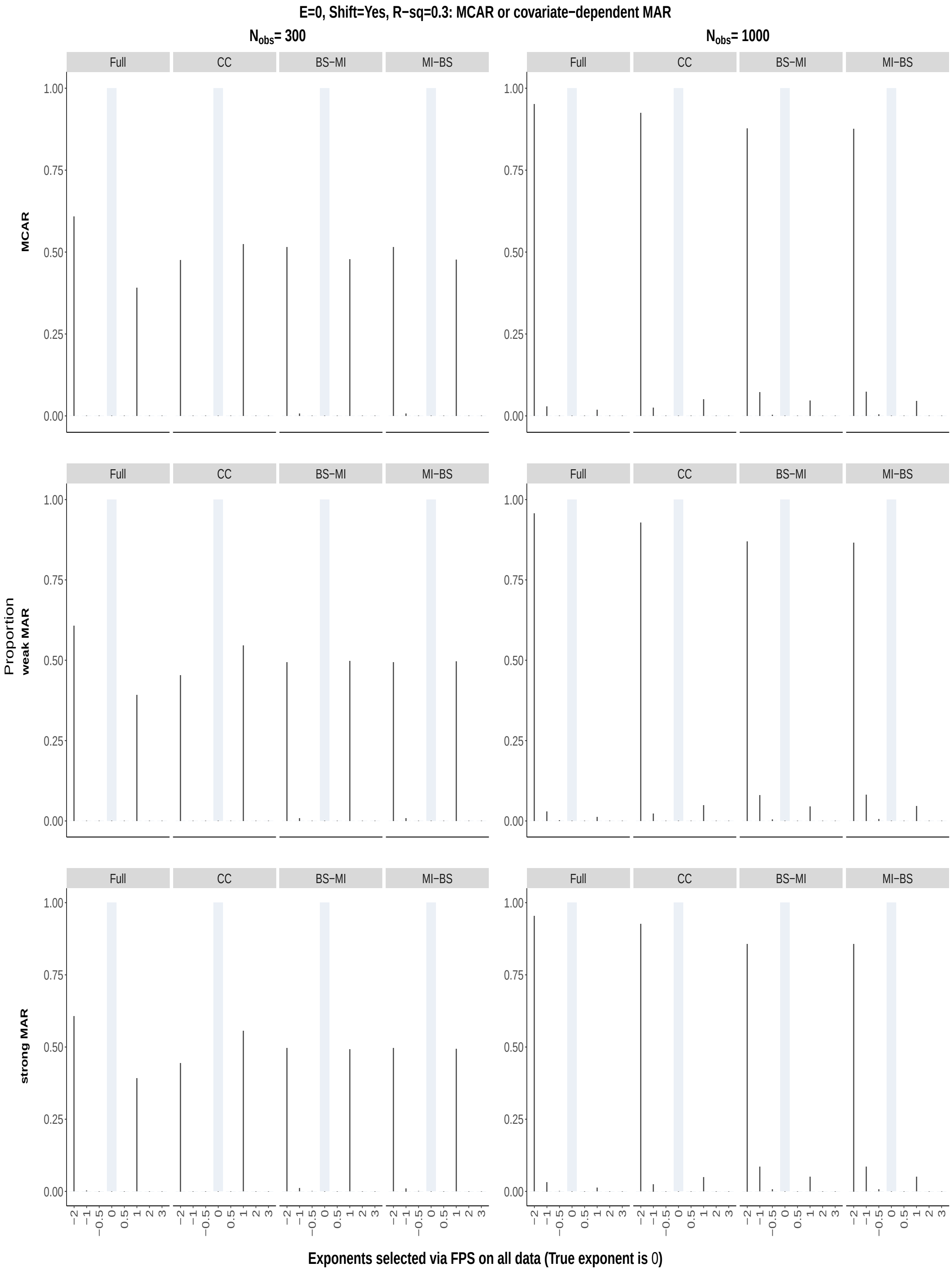

Figure S158: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

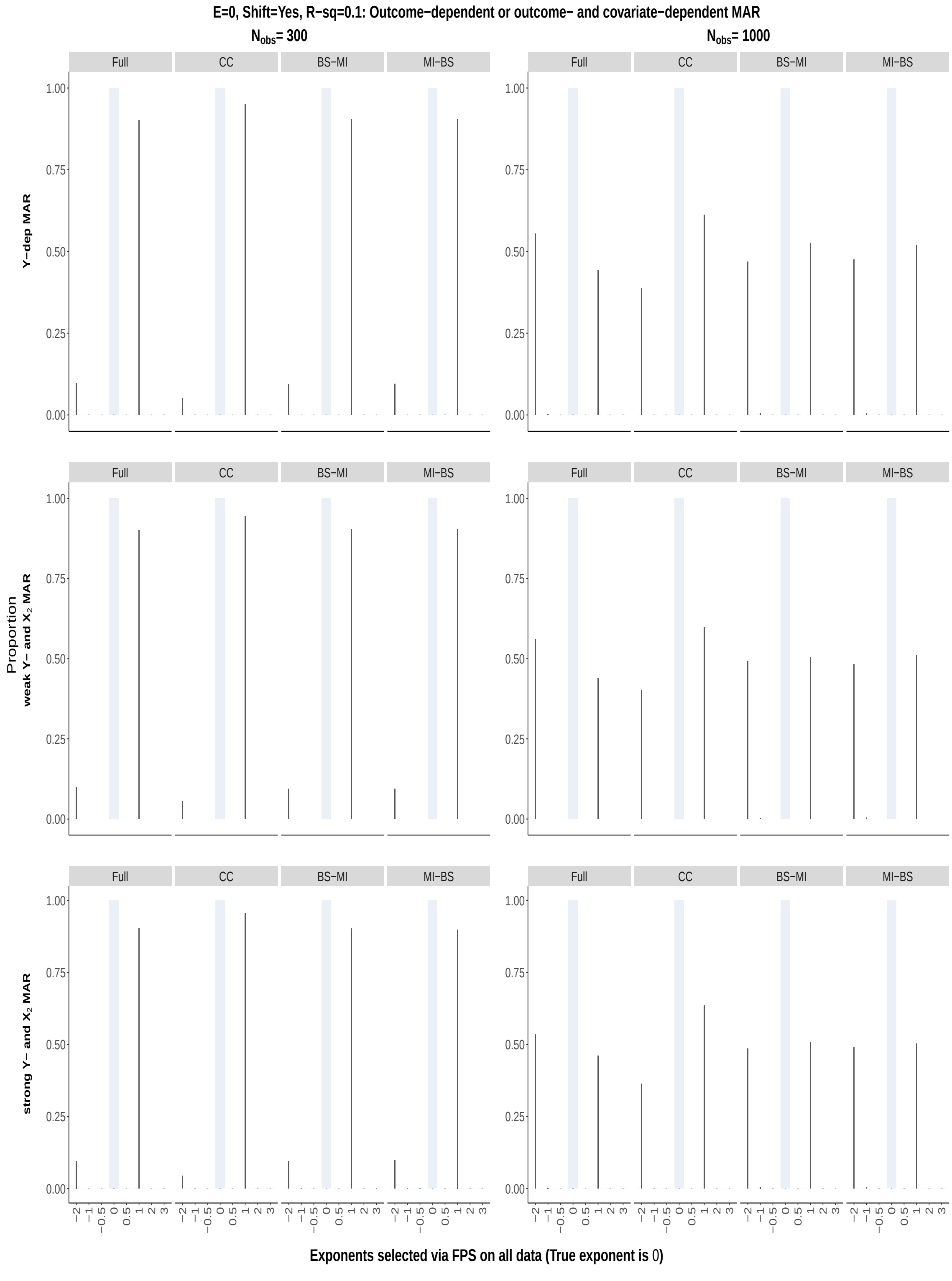

Figure S159: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

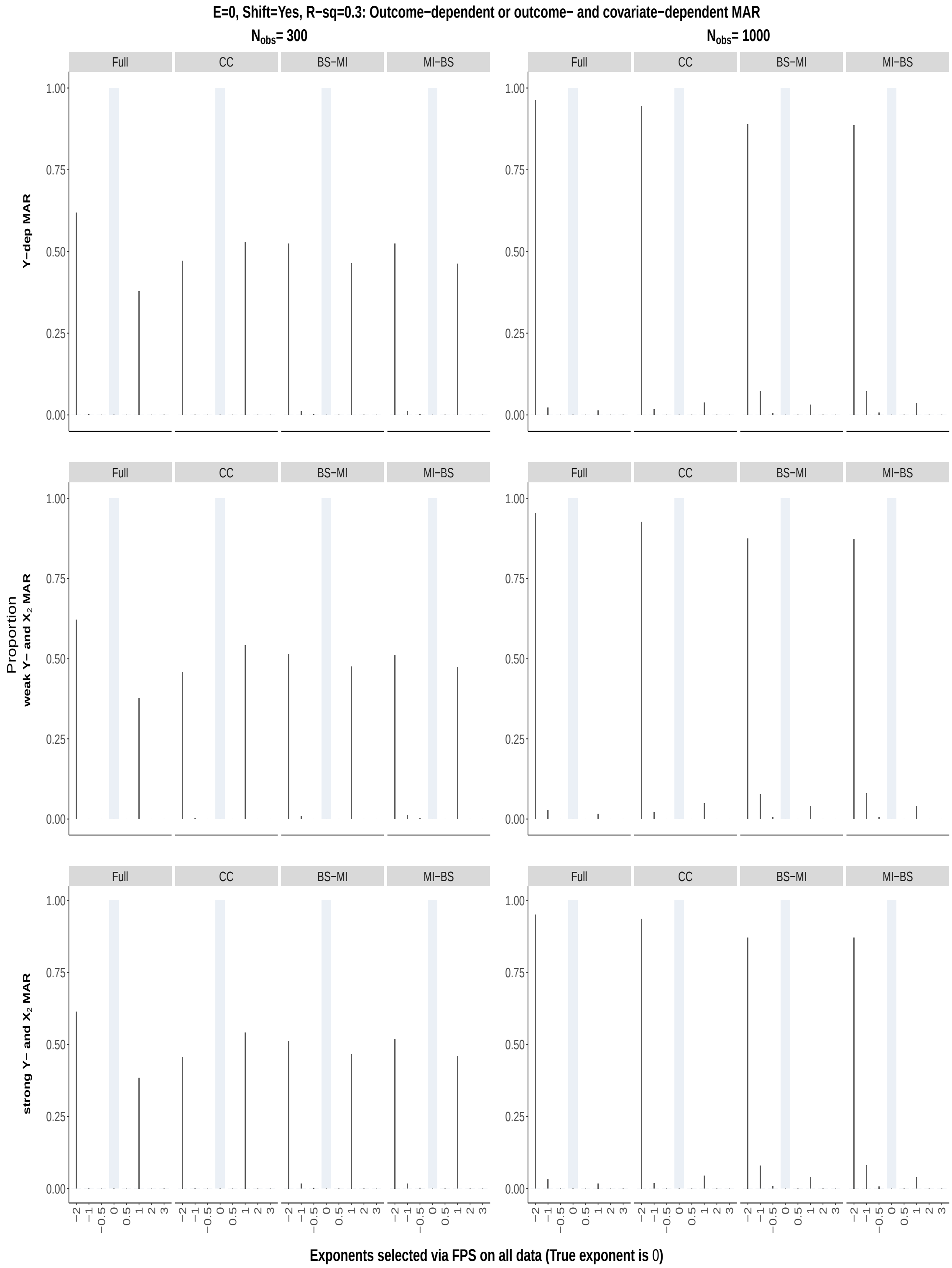

Figure S160: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

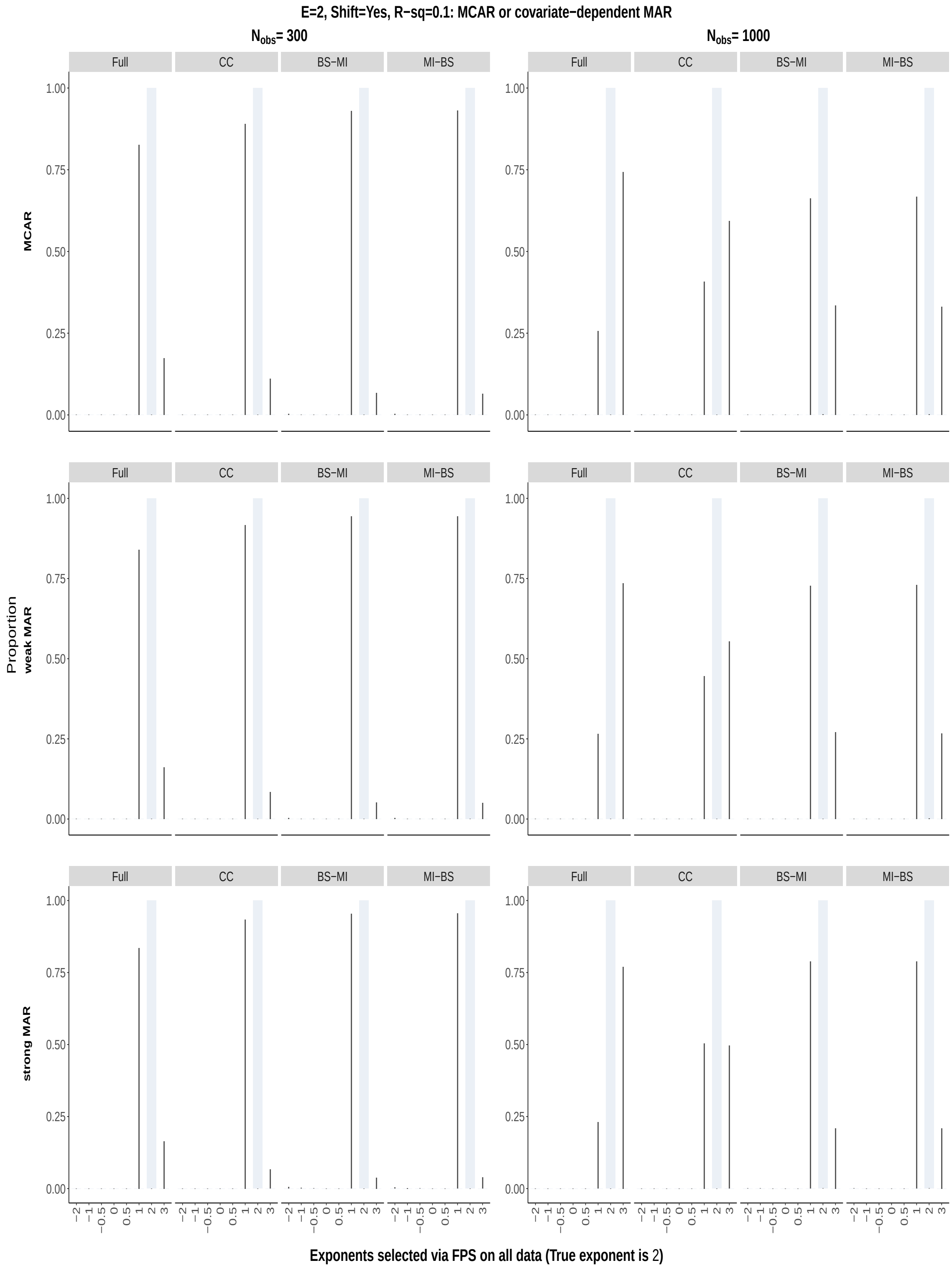

Figure S161: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

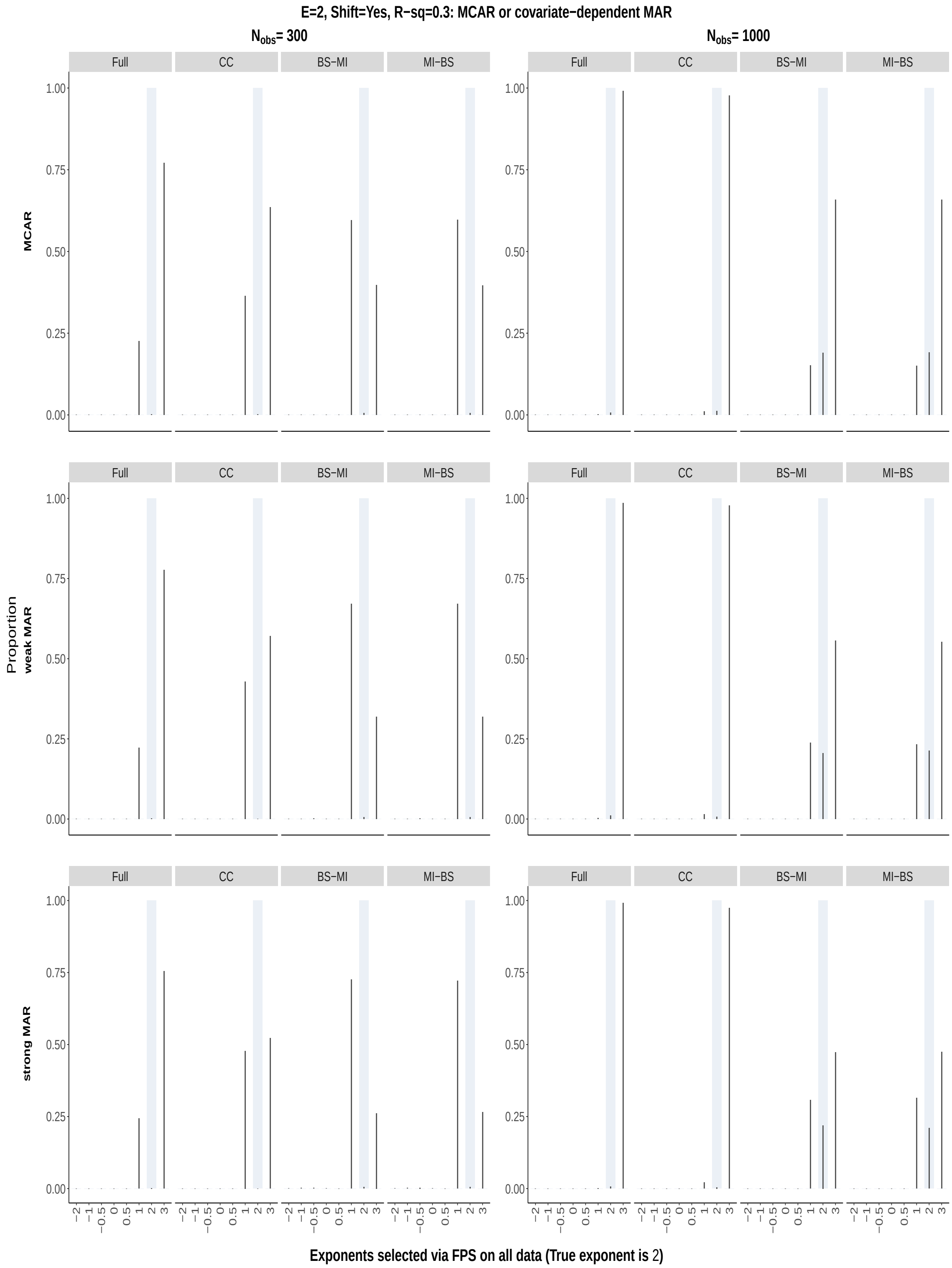

Figure S162: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

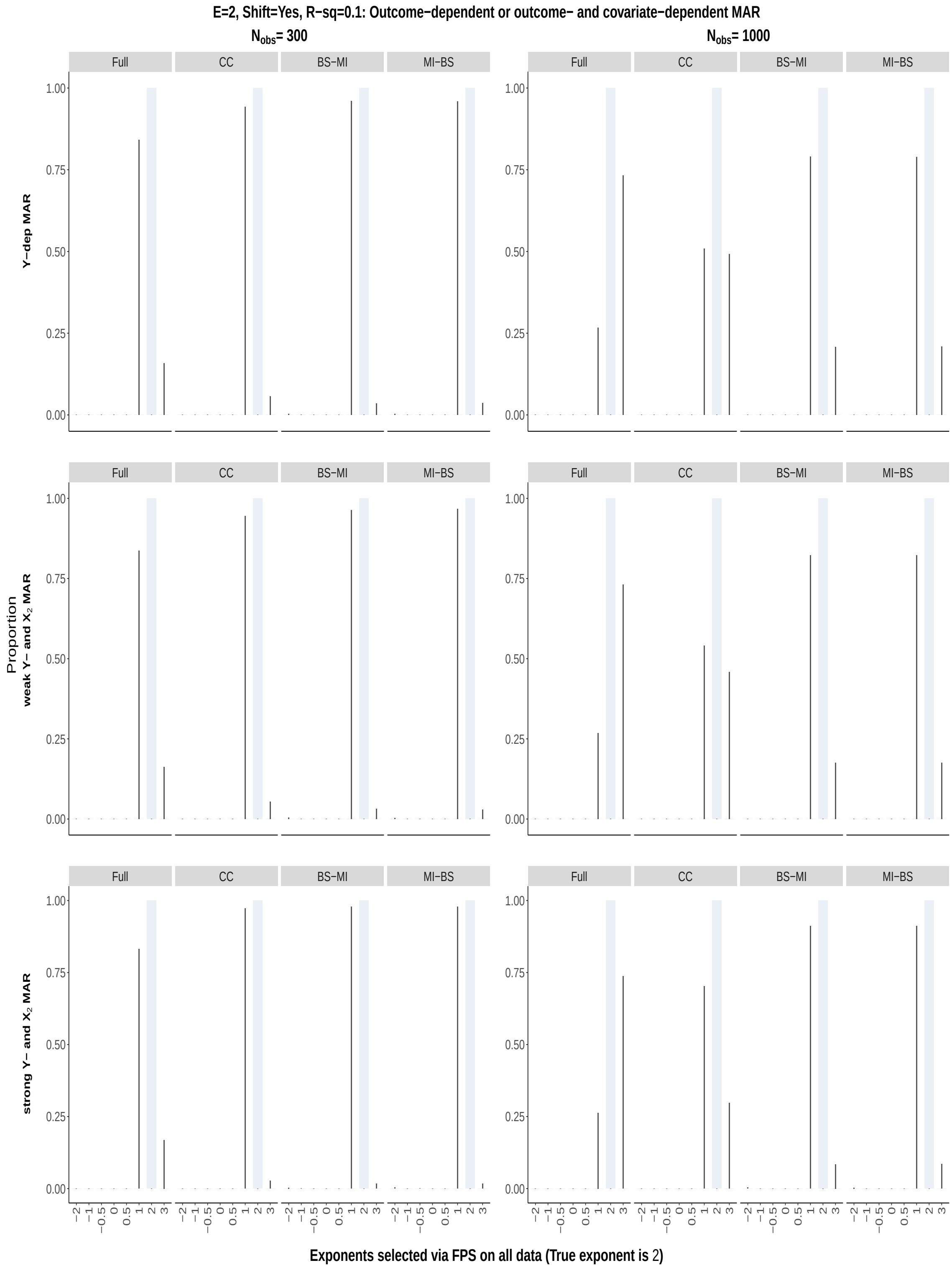

Figure S163: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

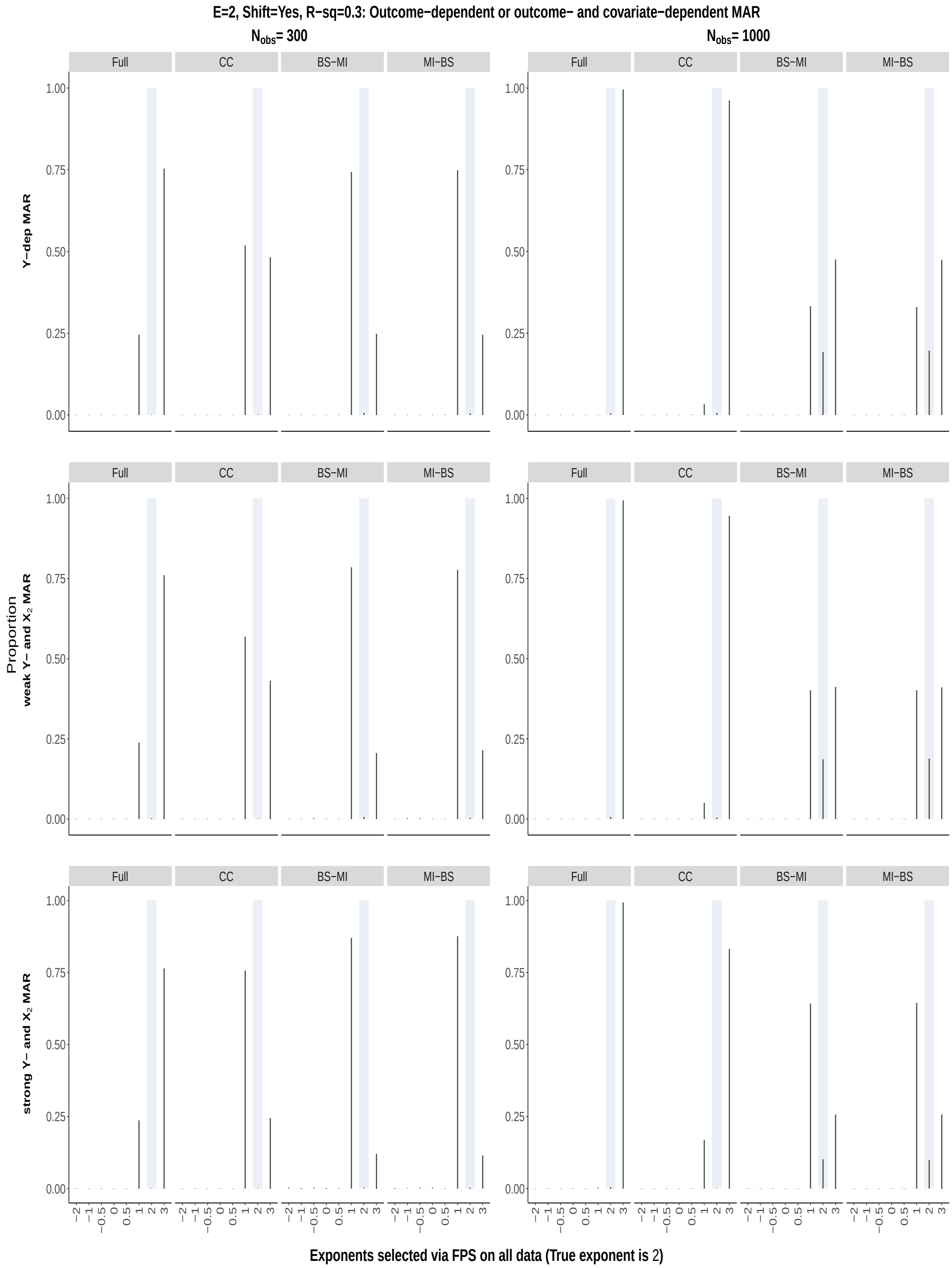

Figure S164: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

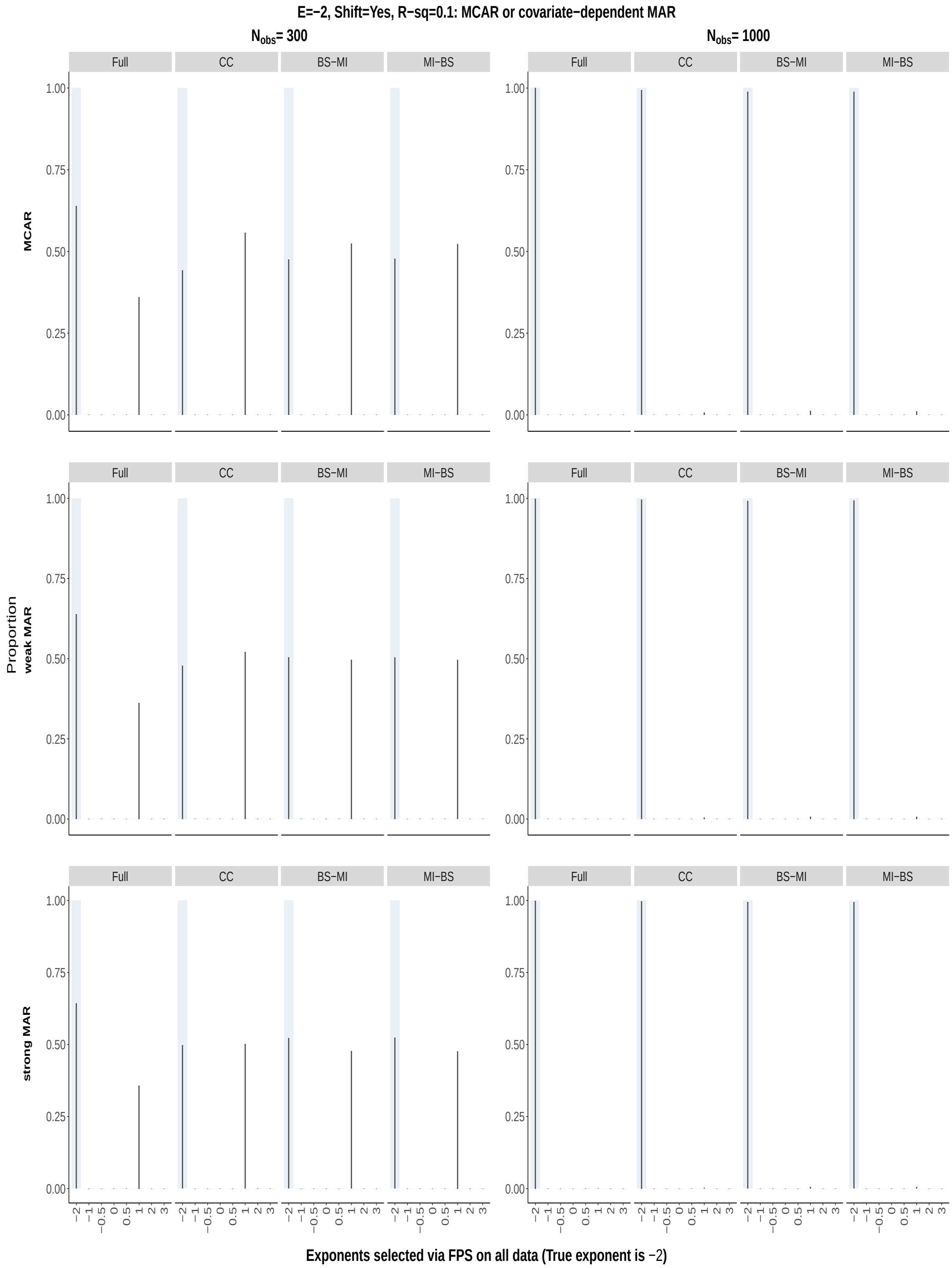

Figure S165: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

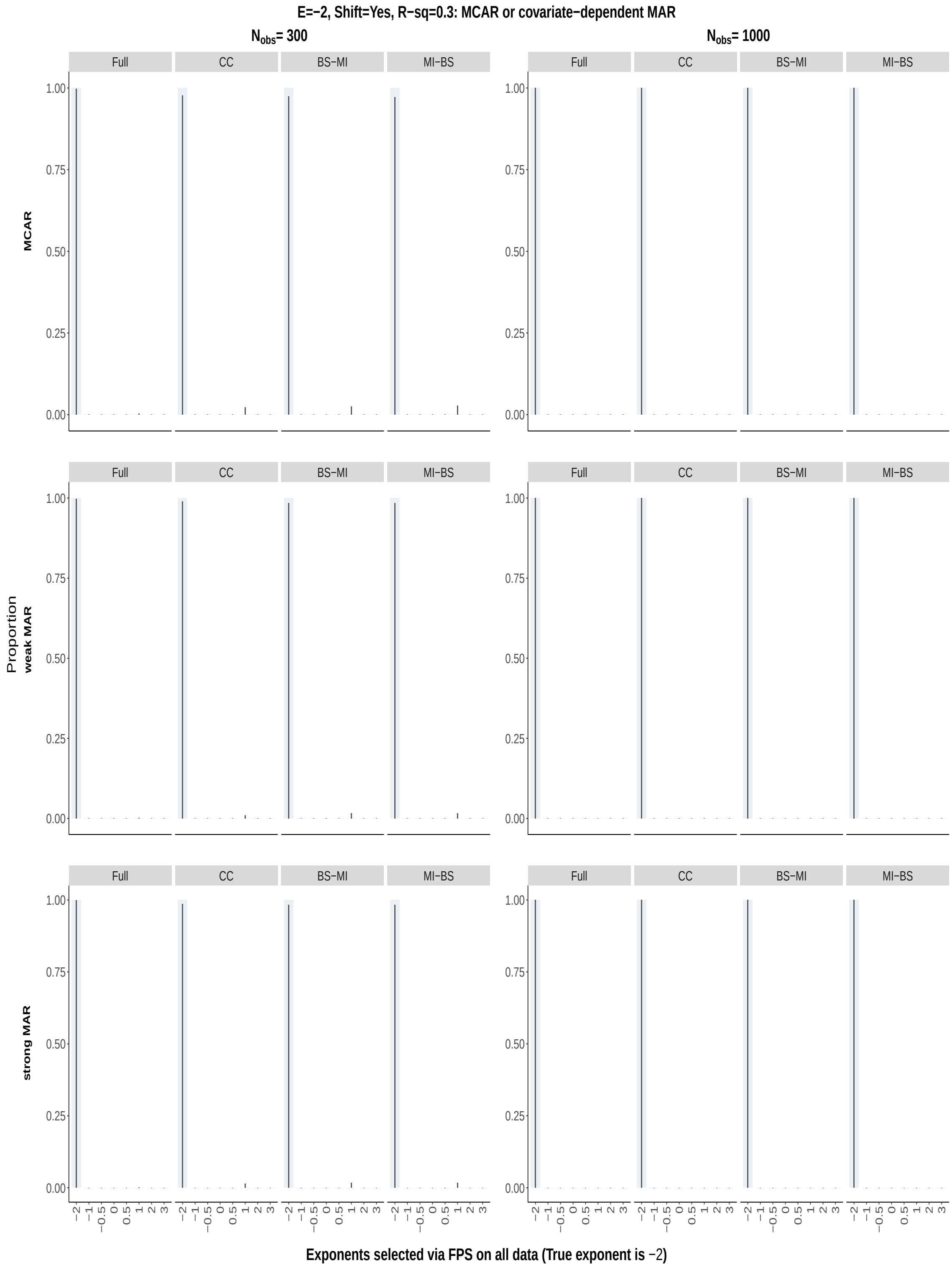

Figure S166: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

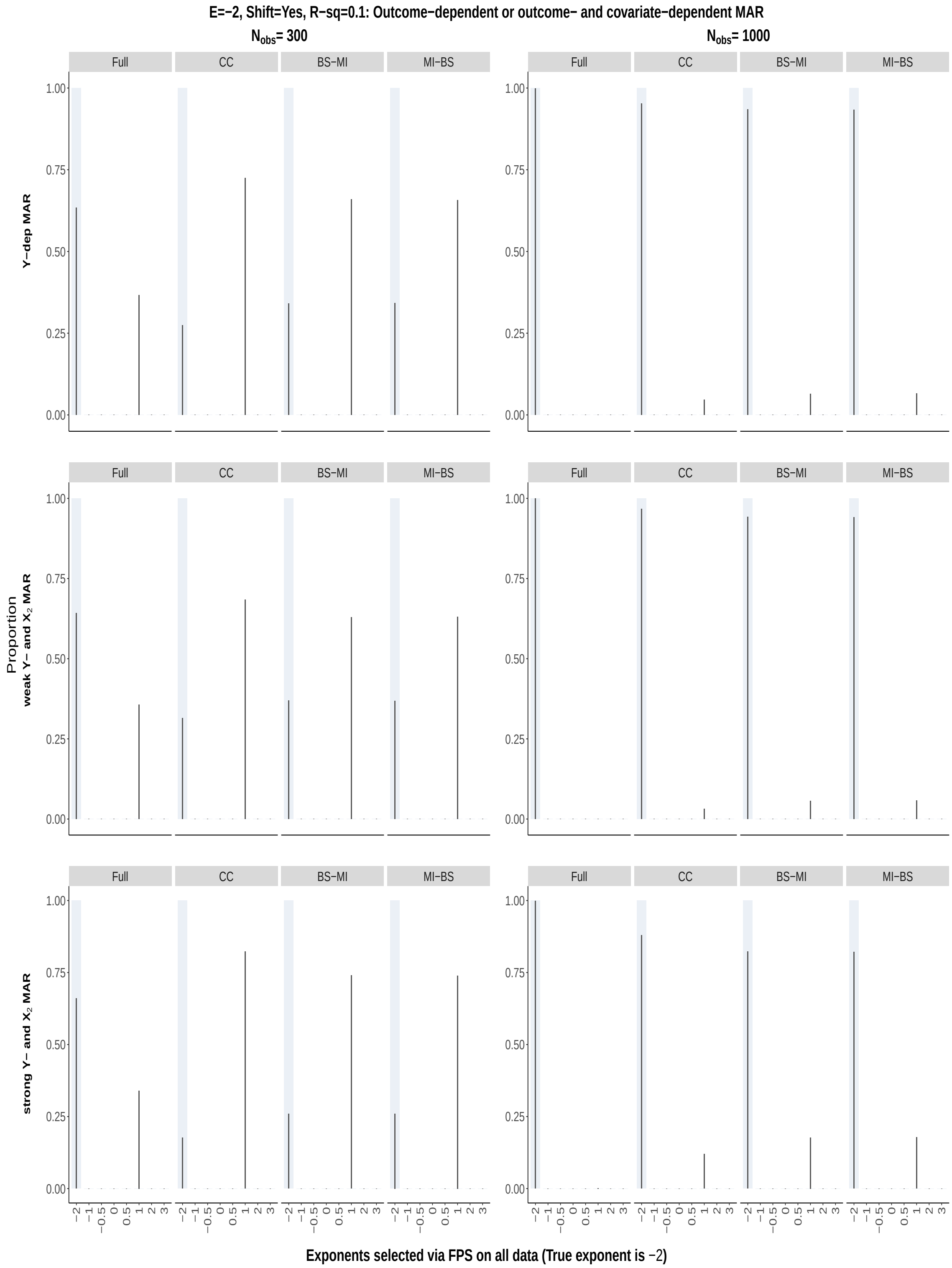

Figure S167: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

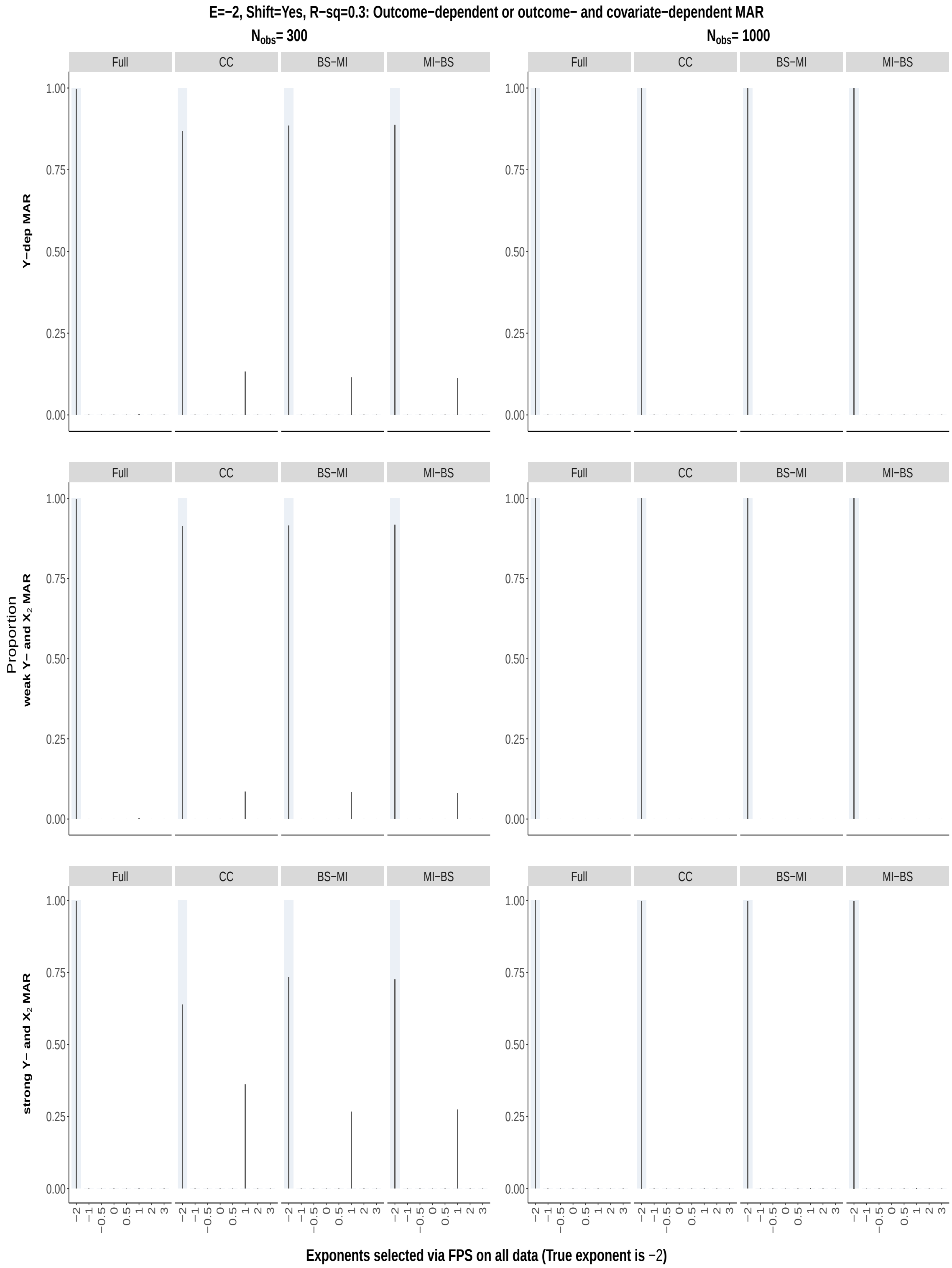

Figure S168: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7 Chapter 9: Simulation study results for MFP, comparison of MSE (Section 9.6)

S7.1 Cross-validation

S7.1.1 $\beta_2 = 1$ and an origin shift transformation has not been applied

True exponent is 0

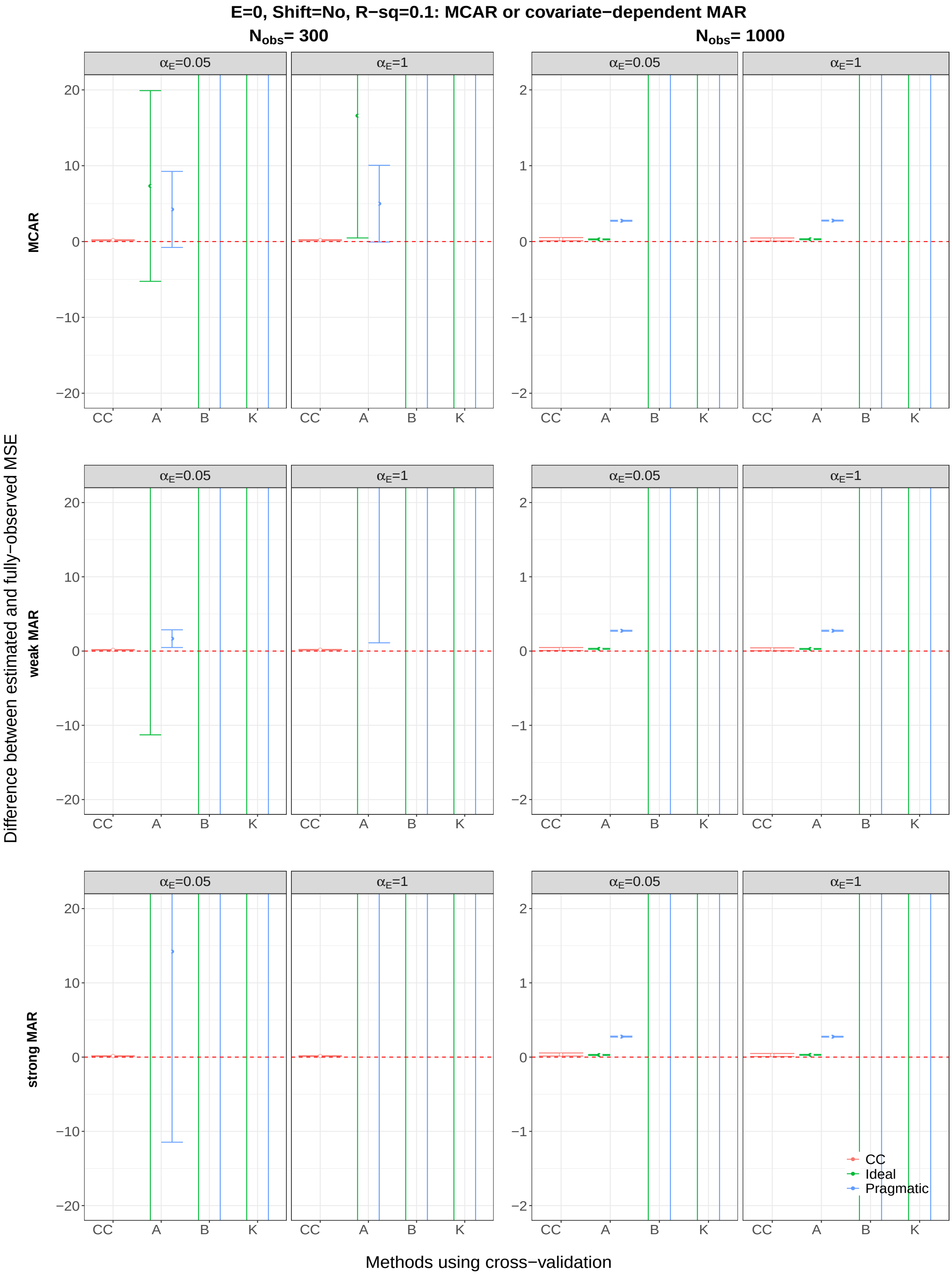

Figure S1: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

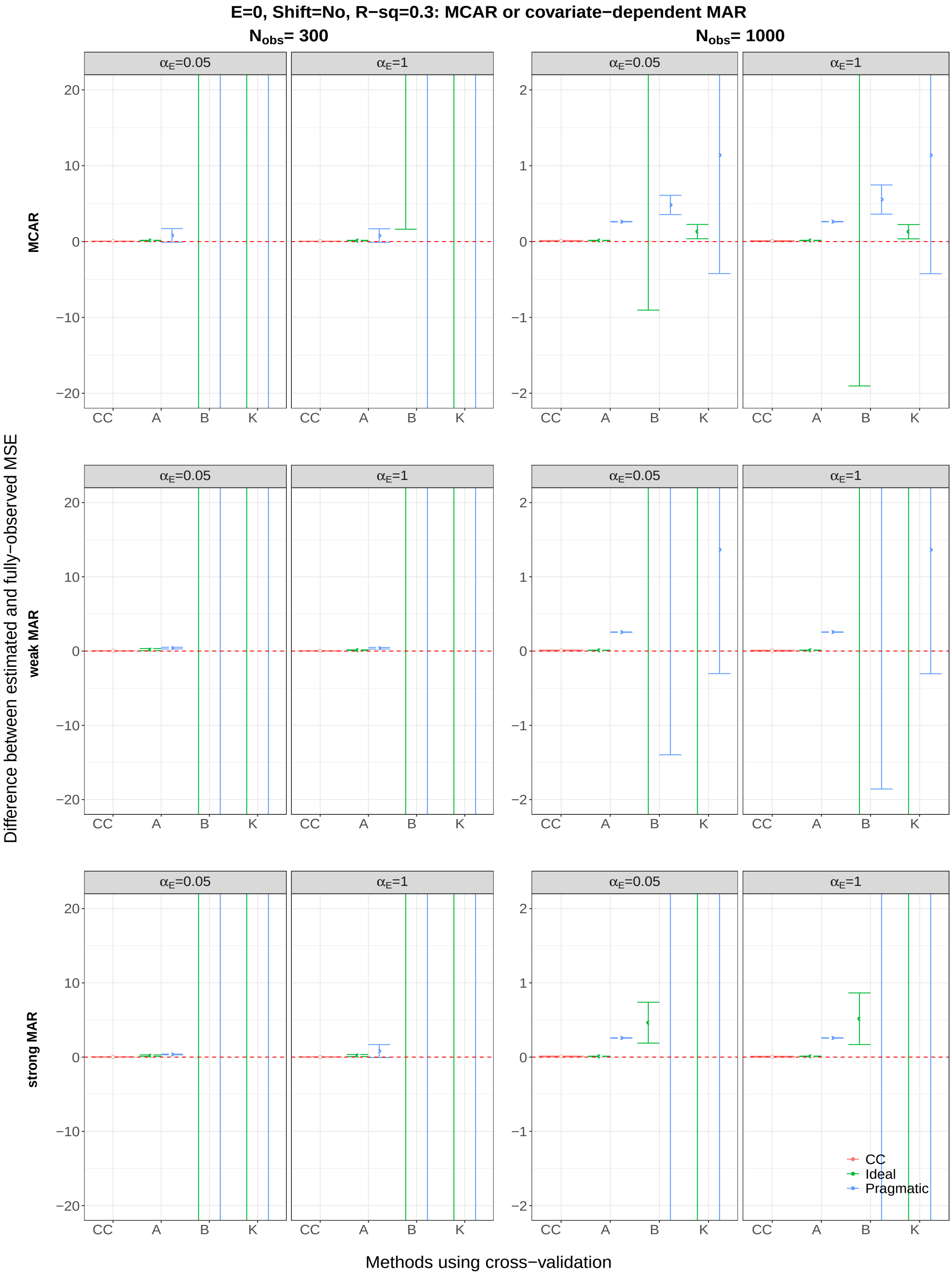

Figure S2: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

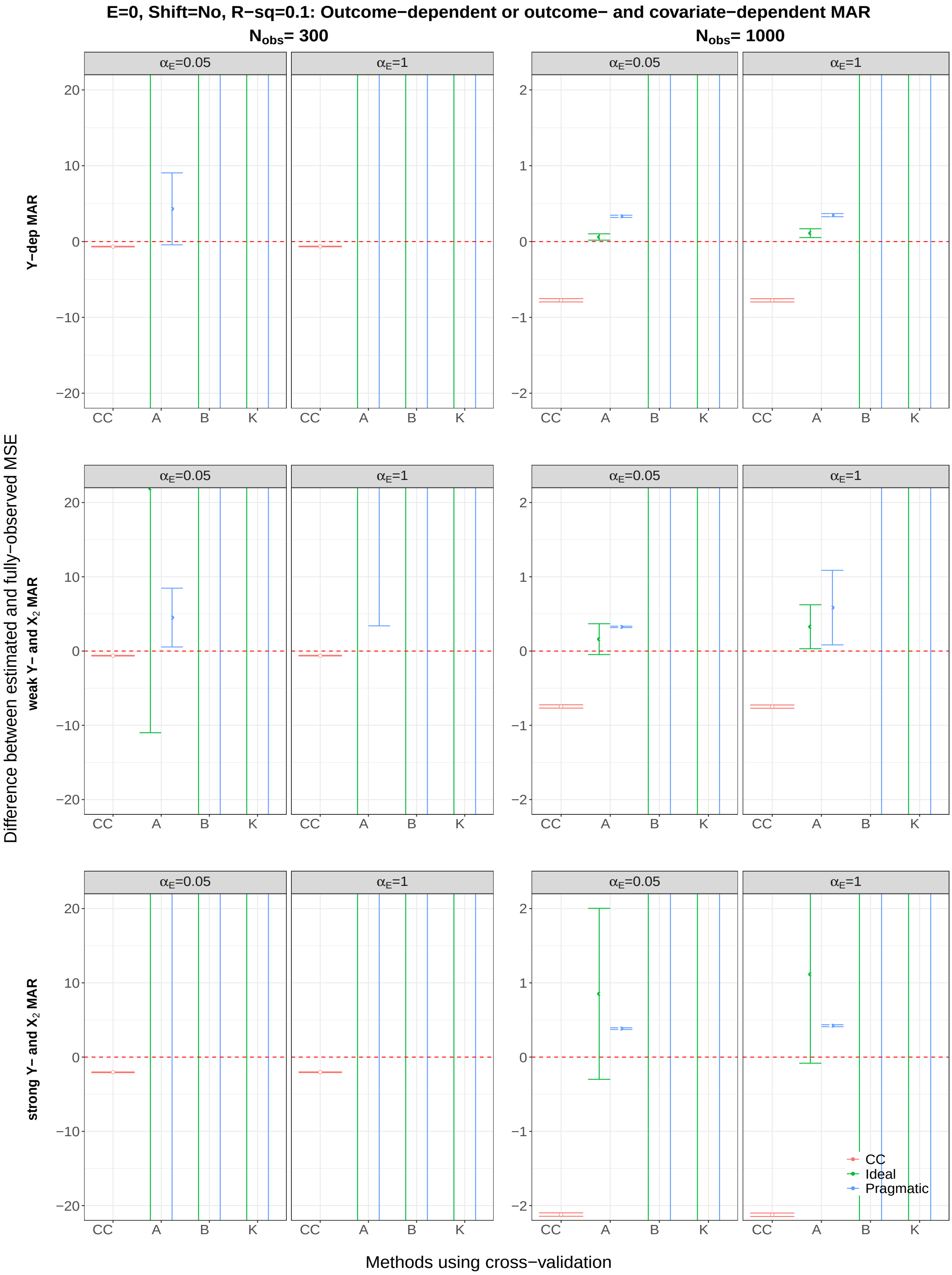

Figure S3: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

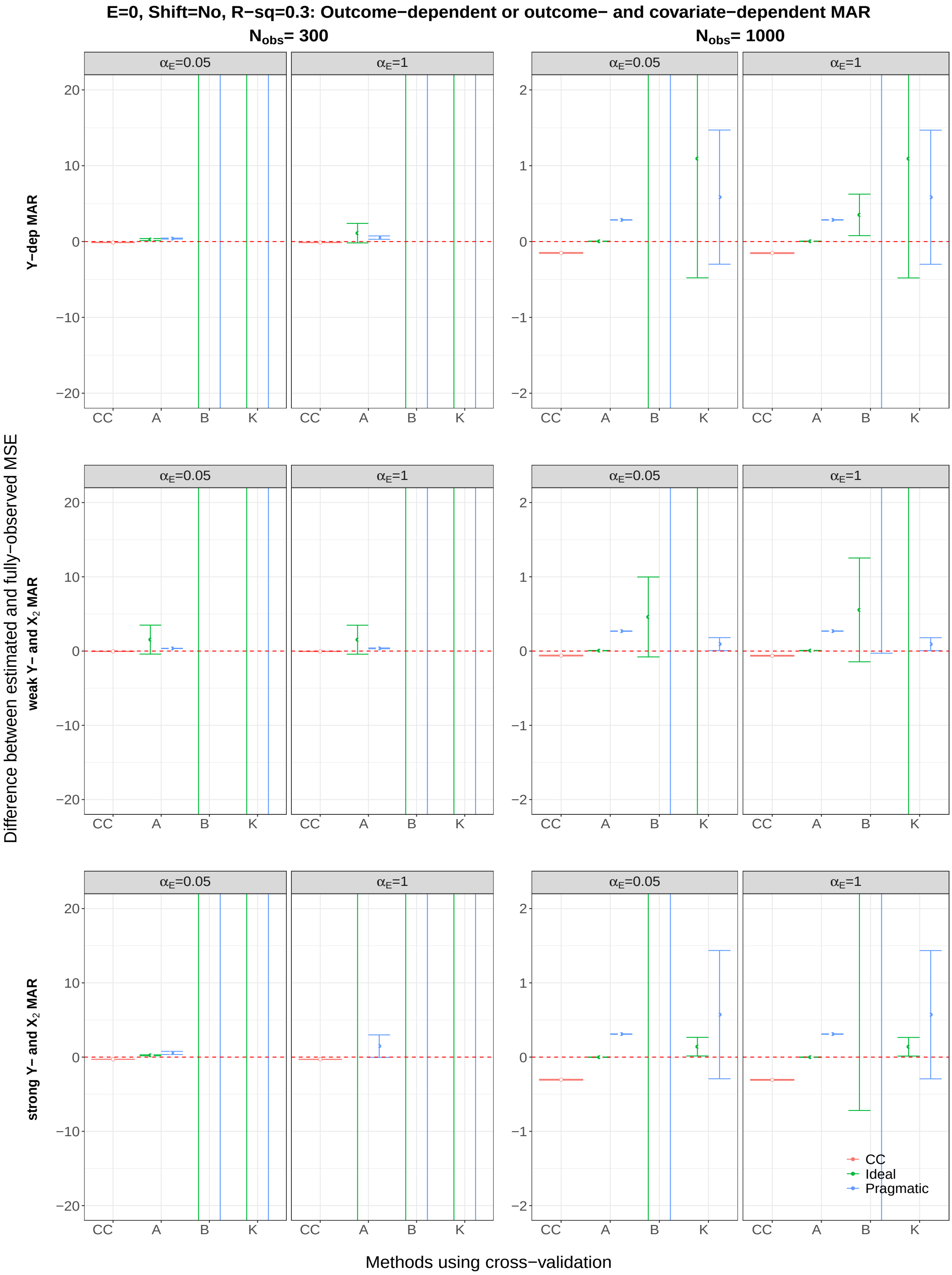

Figure S4: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

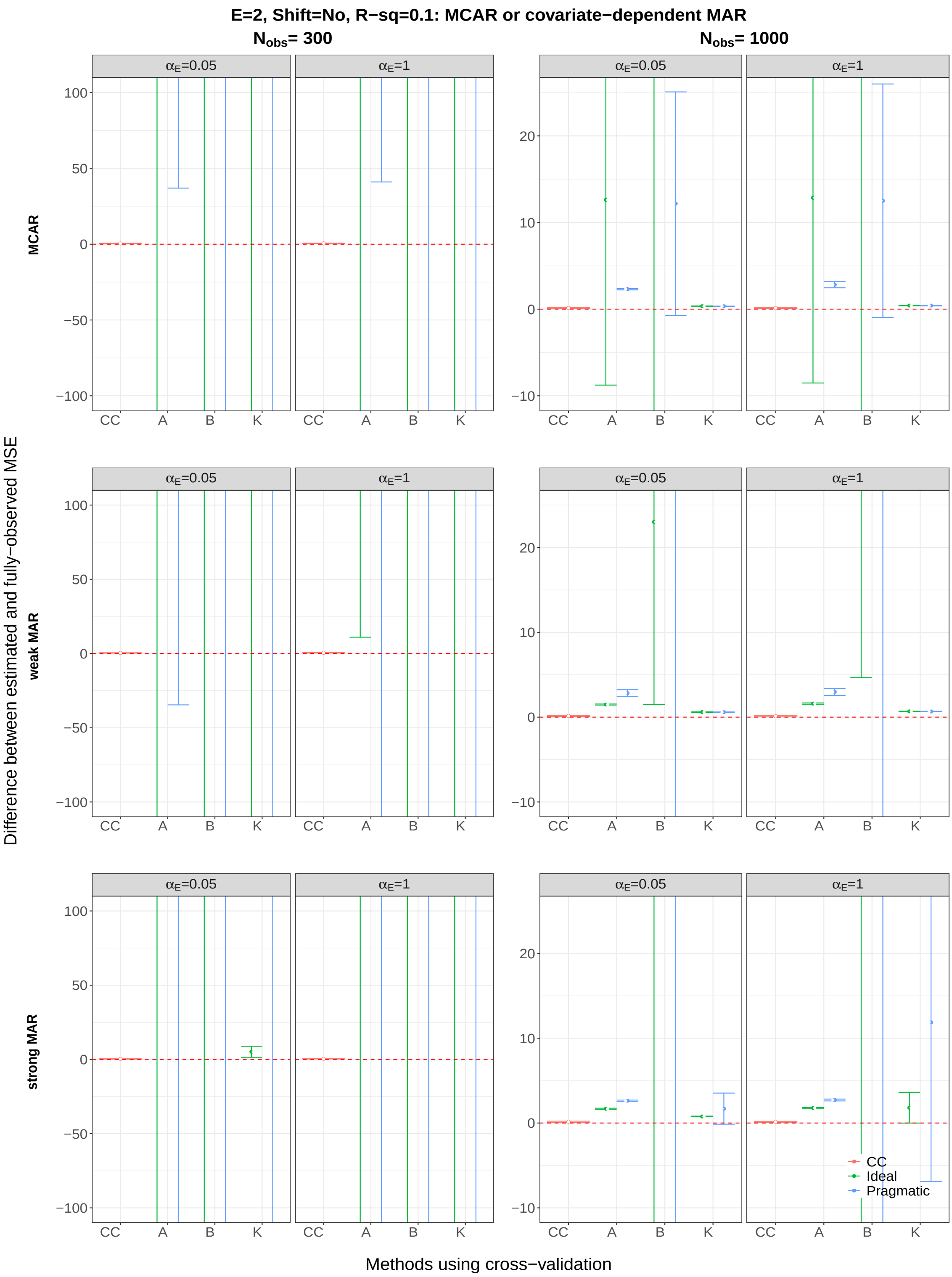

Figure S5: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

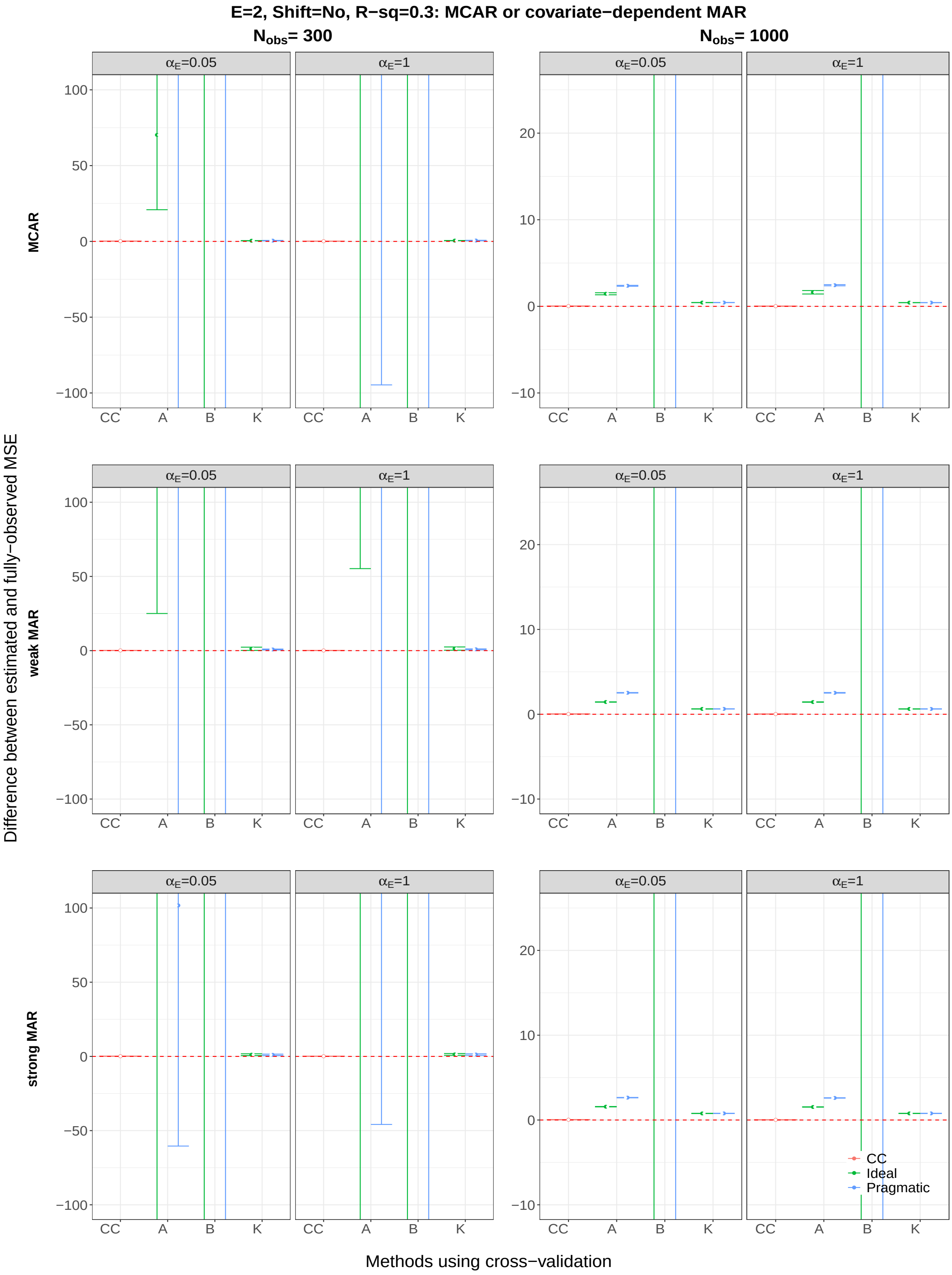

Figure S6: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

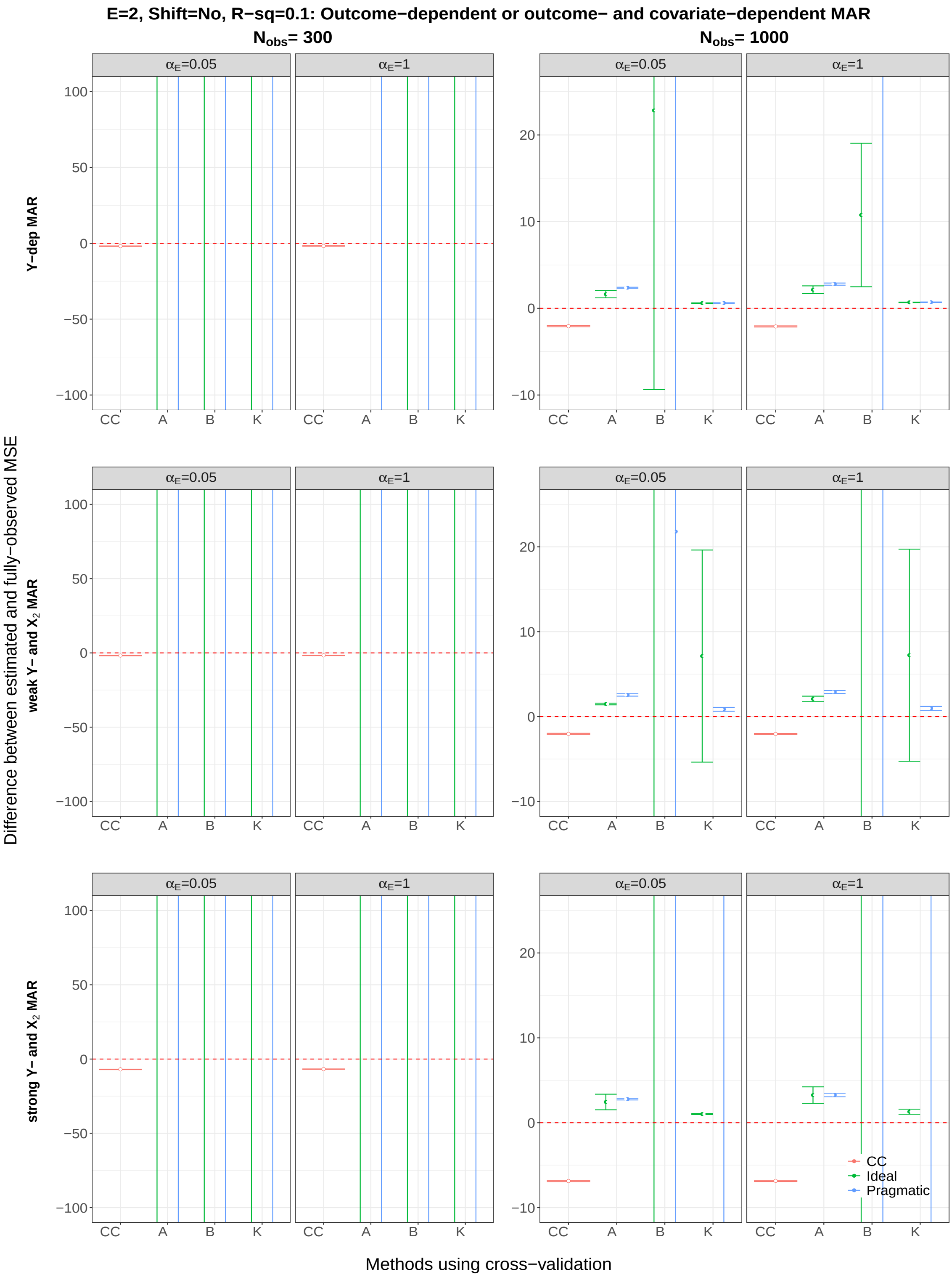

Figure S7: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

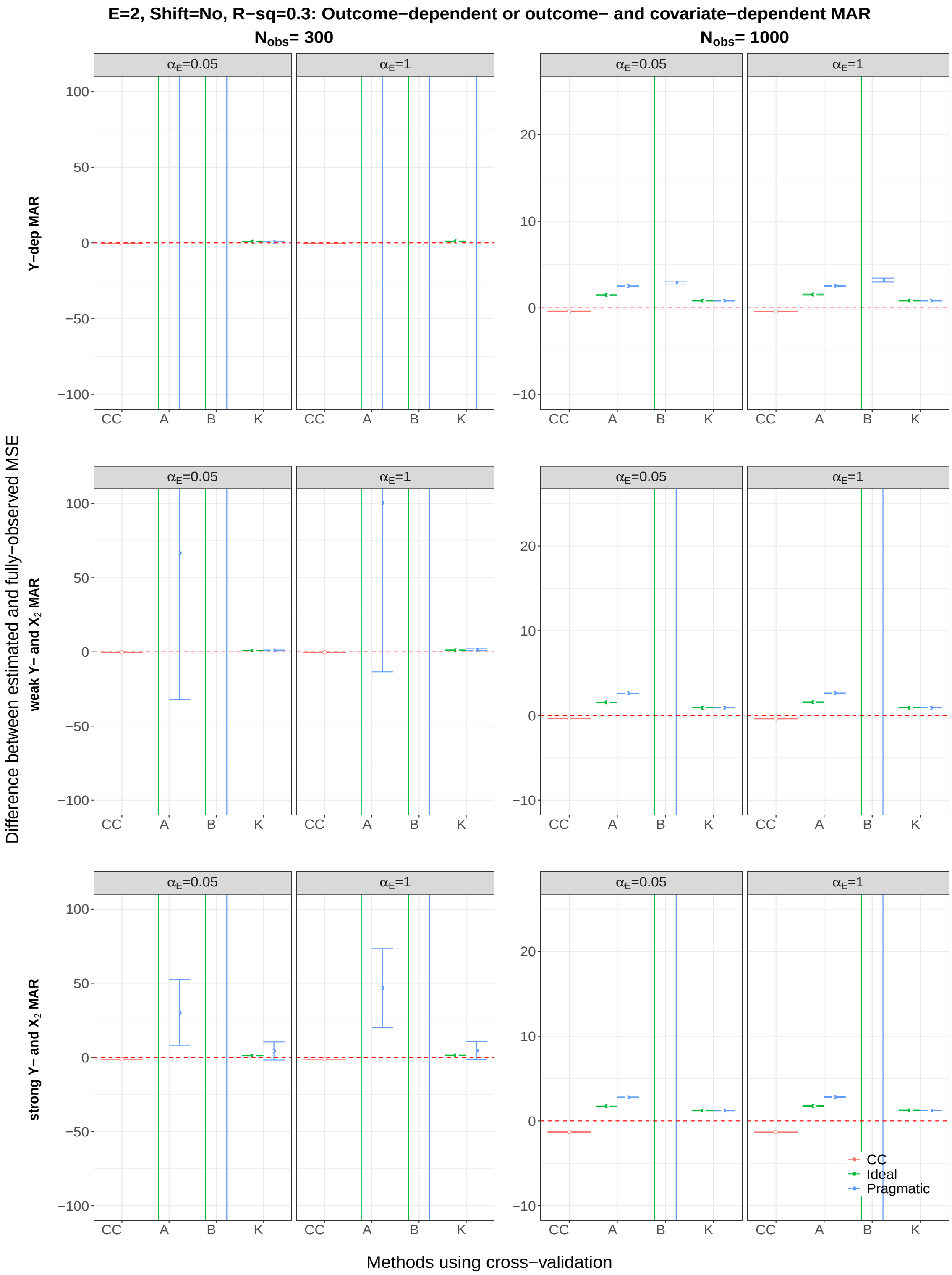

Figure S8: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

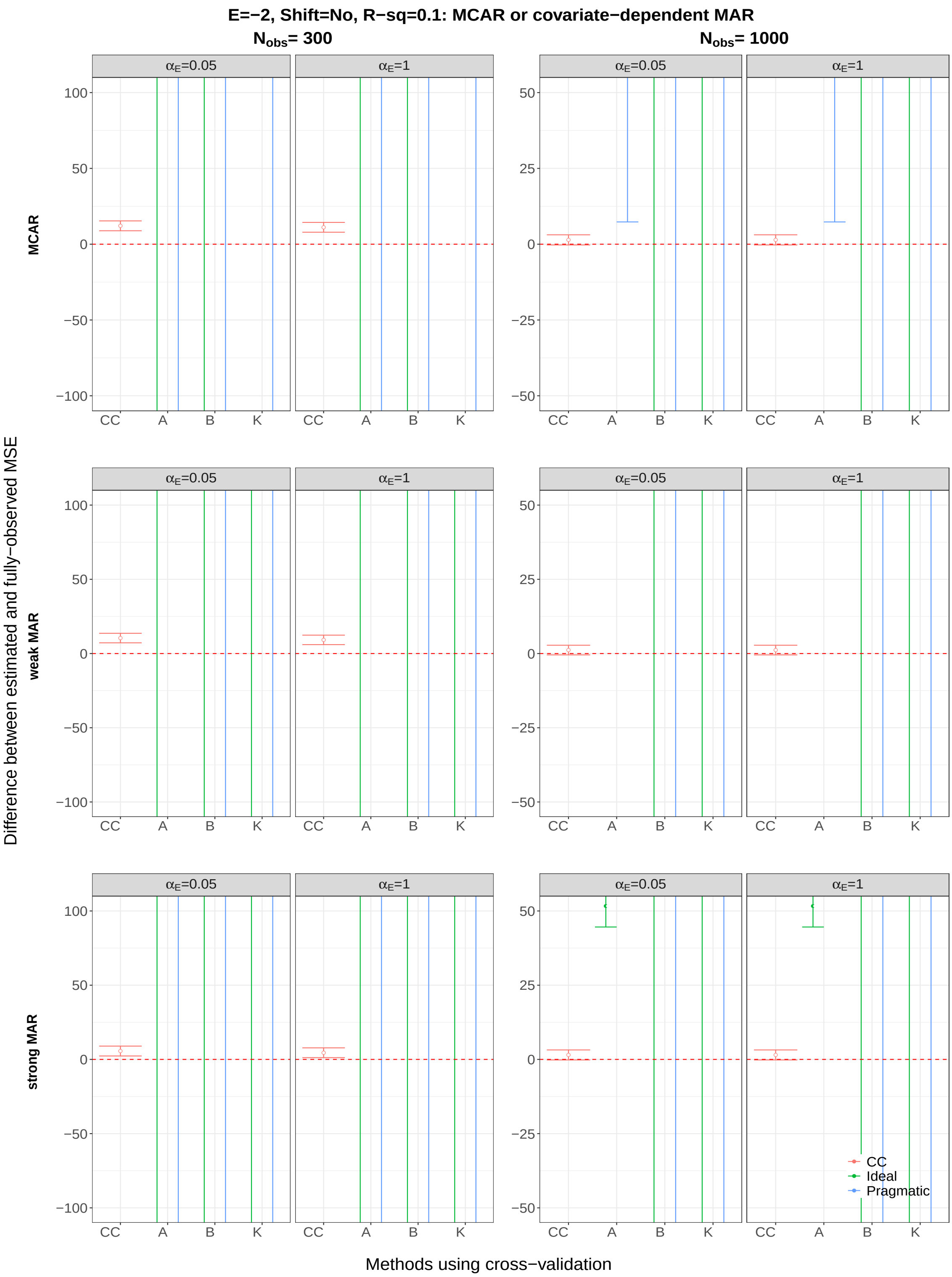

Figure S9: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

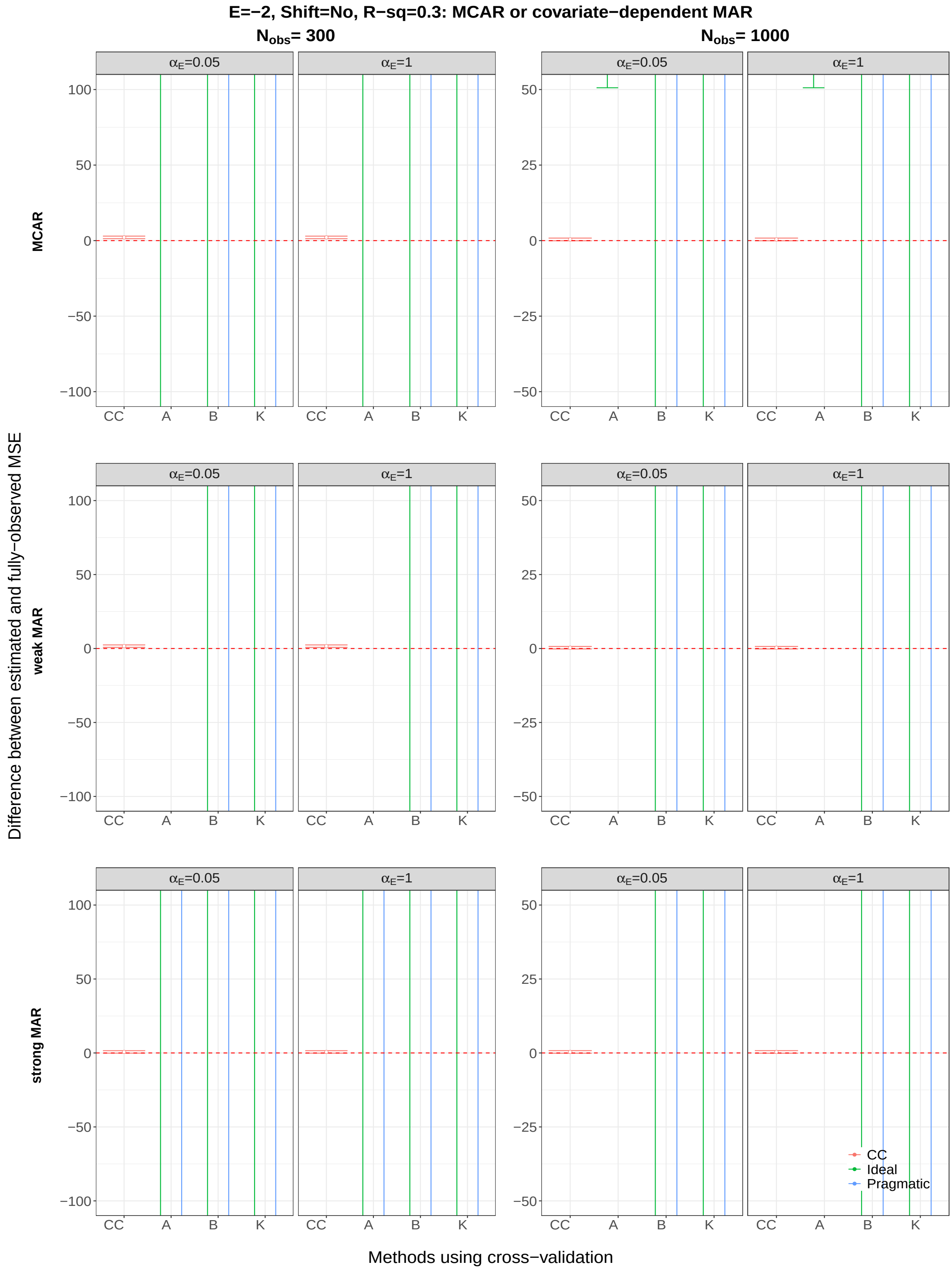

Figure S10: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

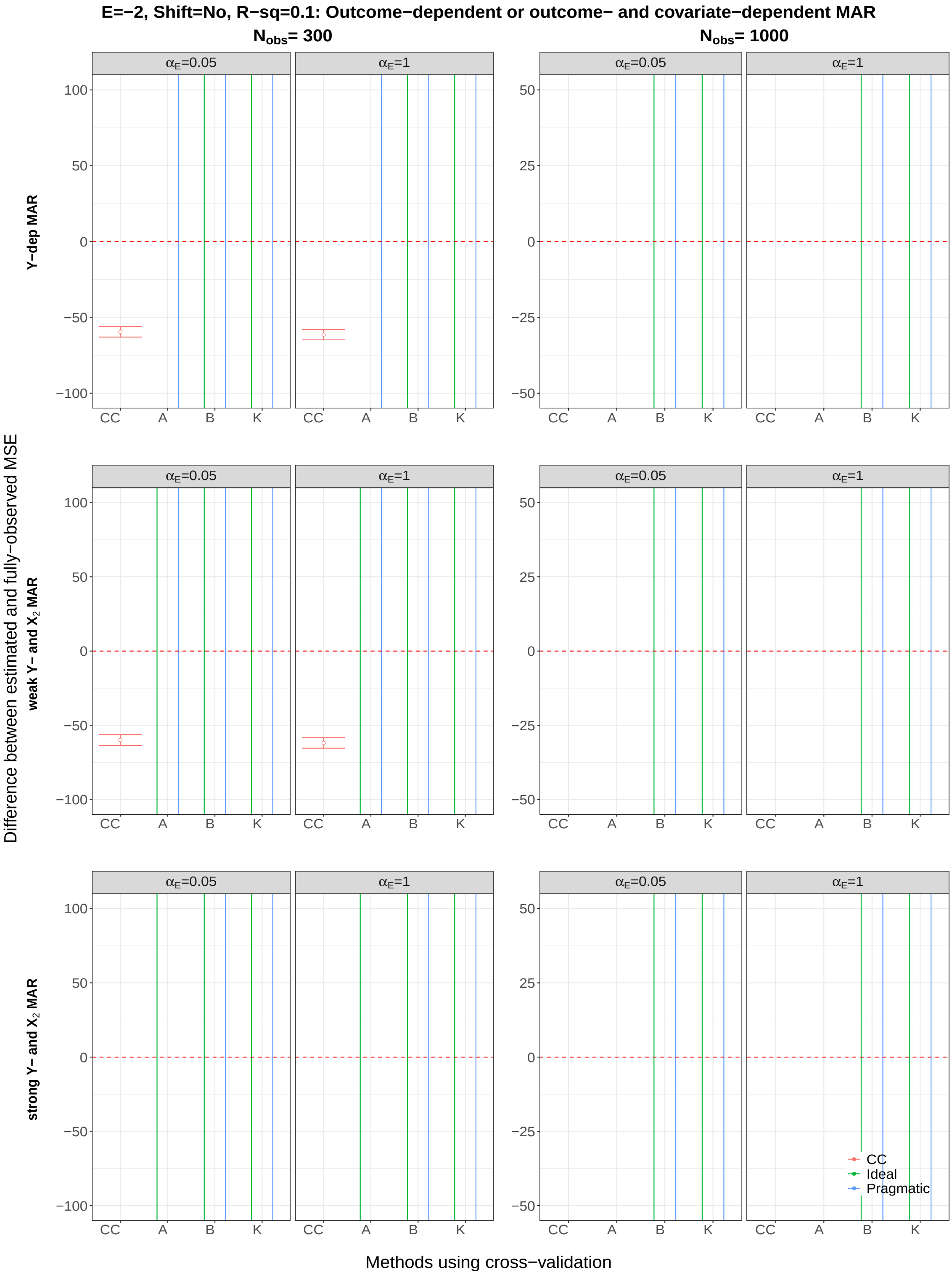

Figure S11: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

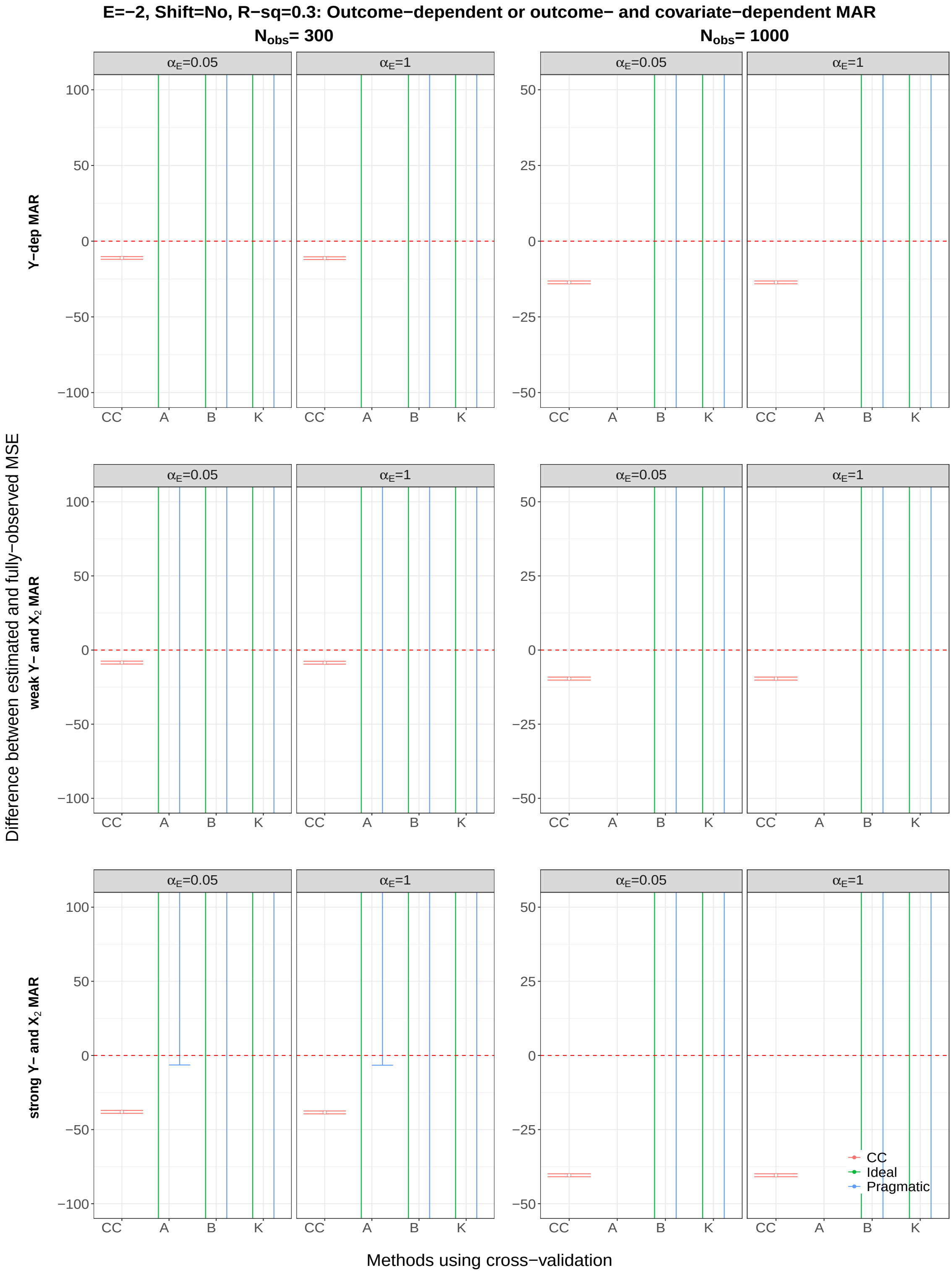

Figure S12: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7.1.2 $\beta_2 = 1$ and an origin shift transformation has been applied

True exponent is 0

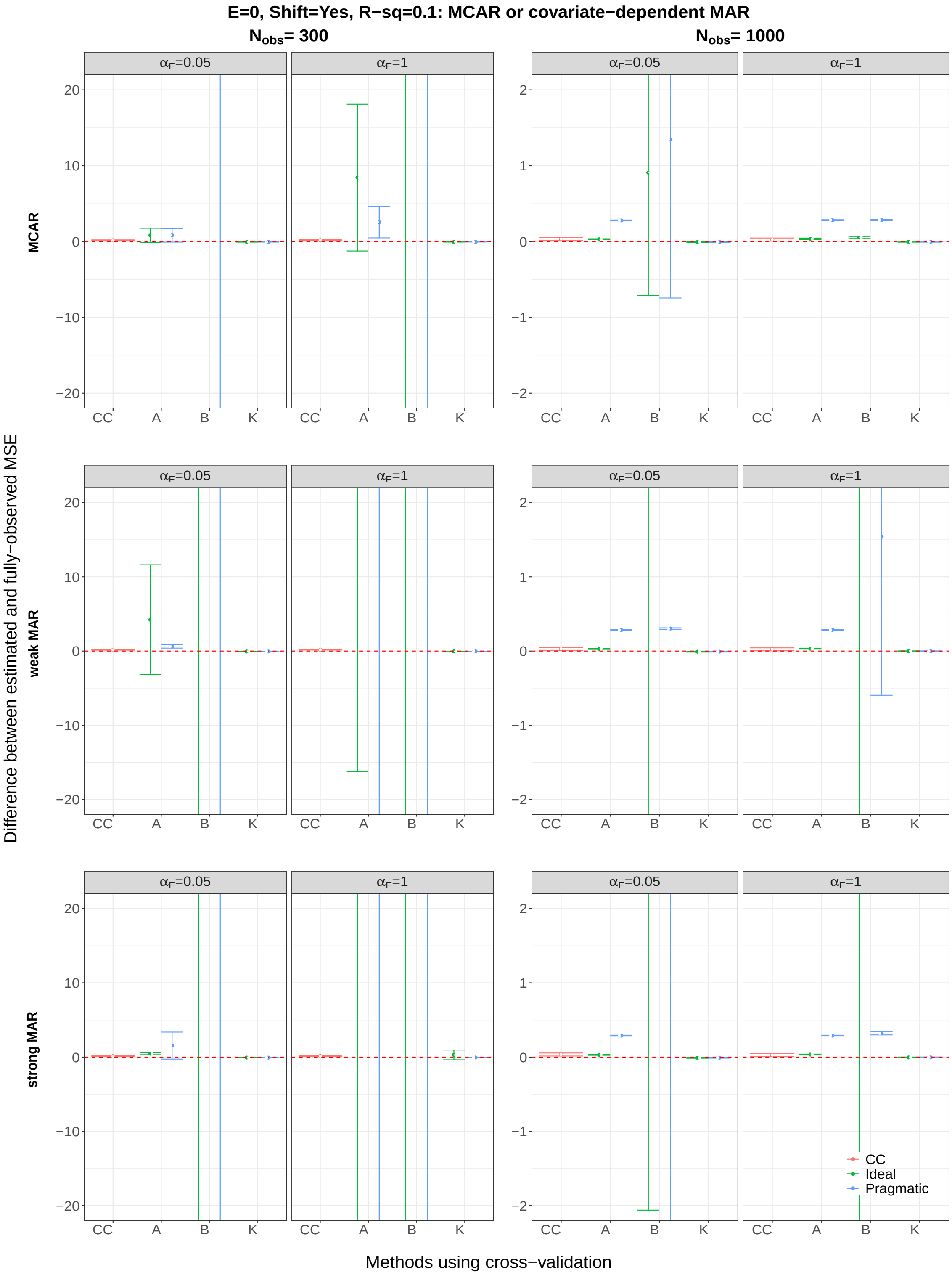

Figure S13: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

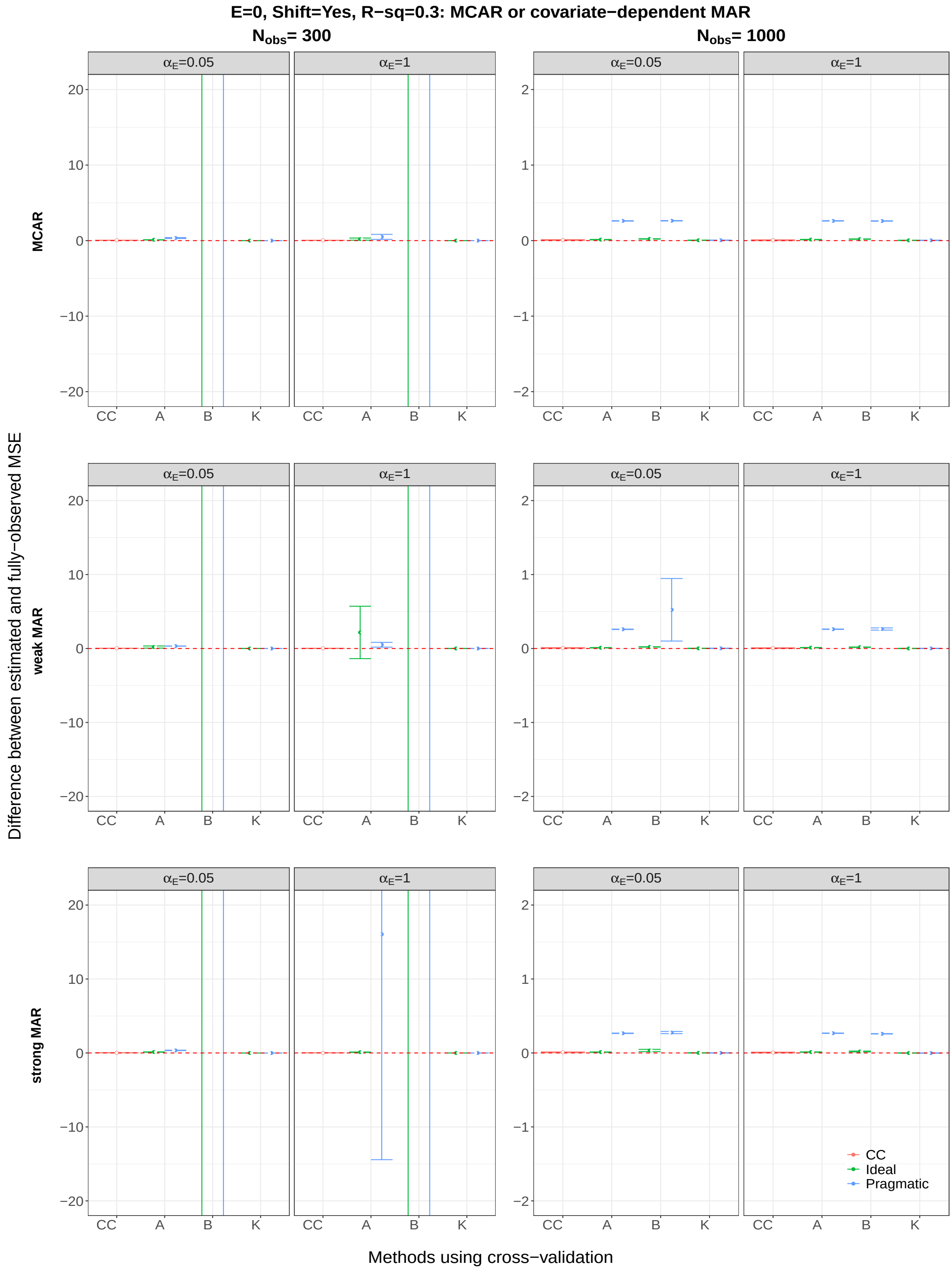

Figure S14: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

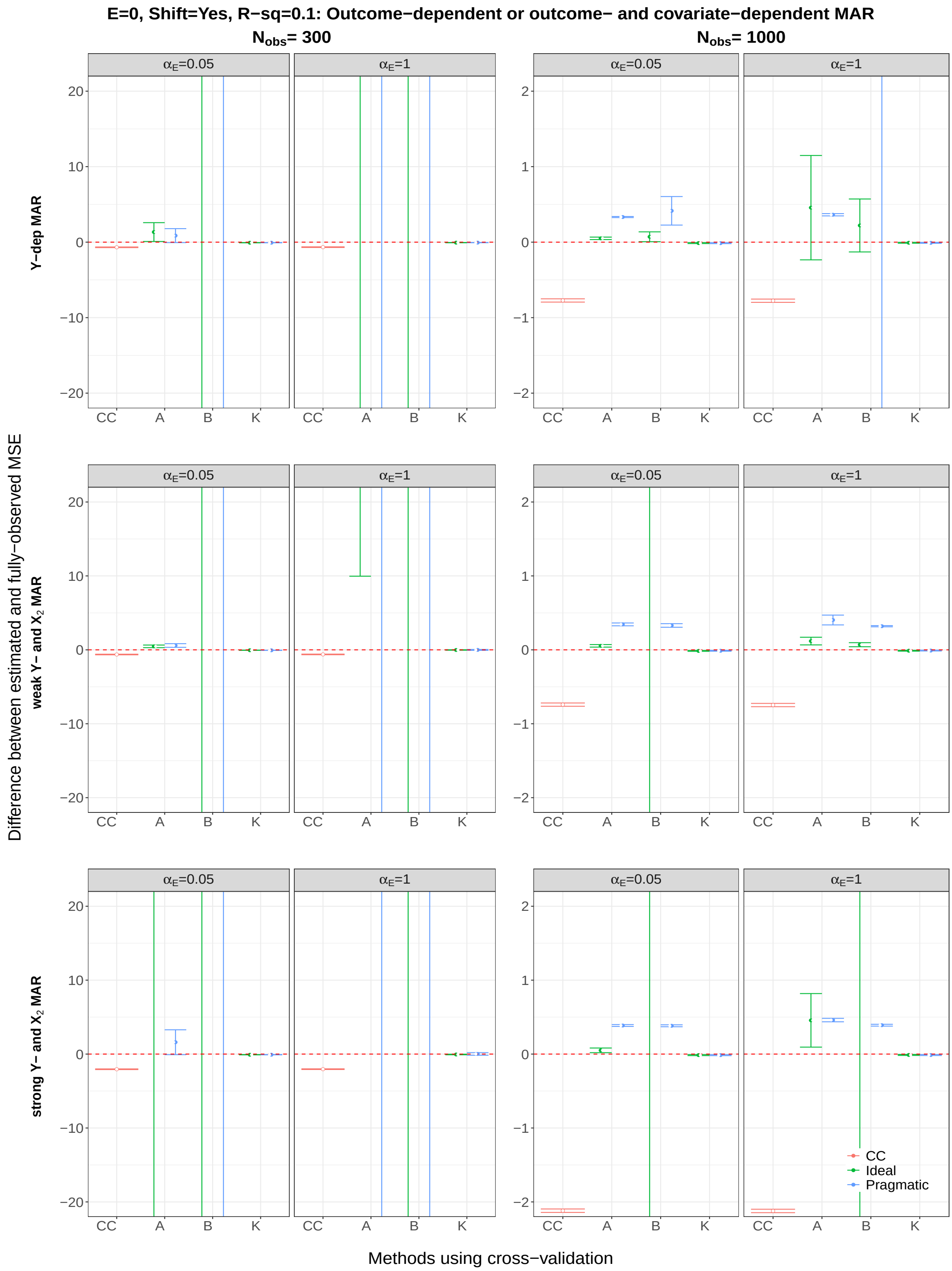

Figure S15: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

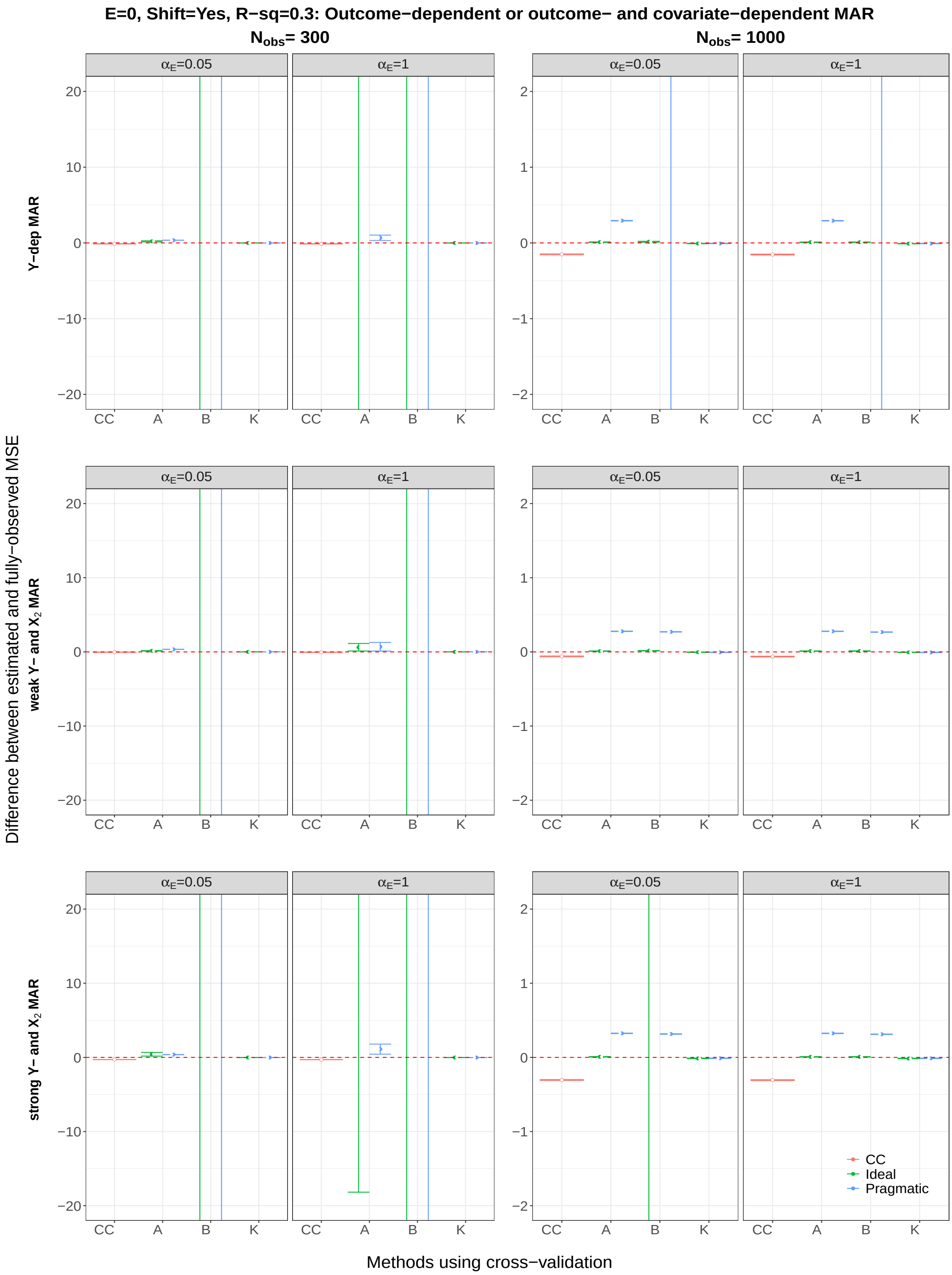

Figure S16: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

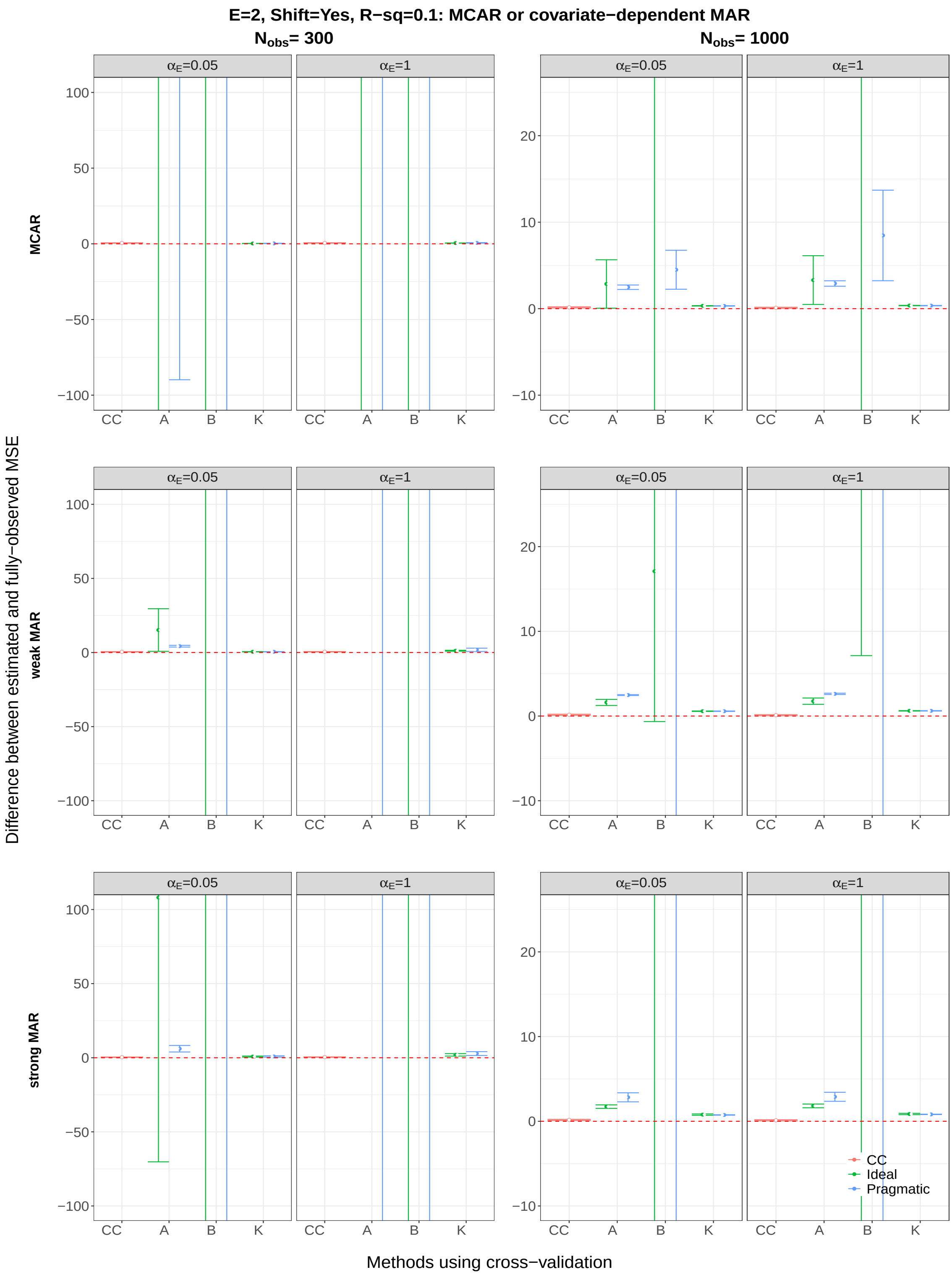

Figure S17: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

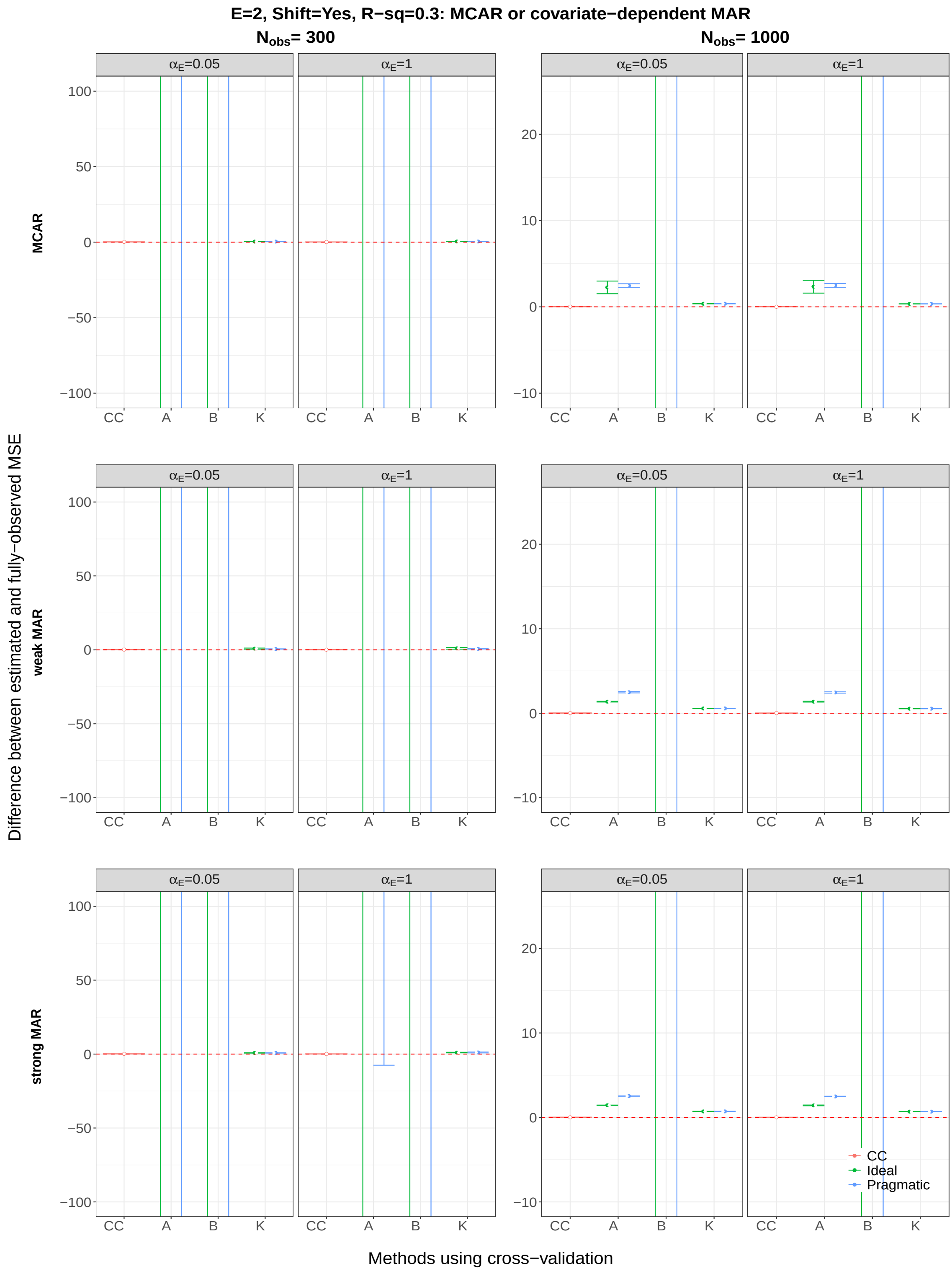

Figure S18: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

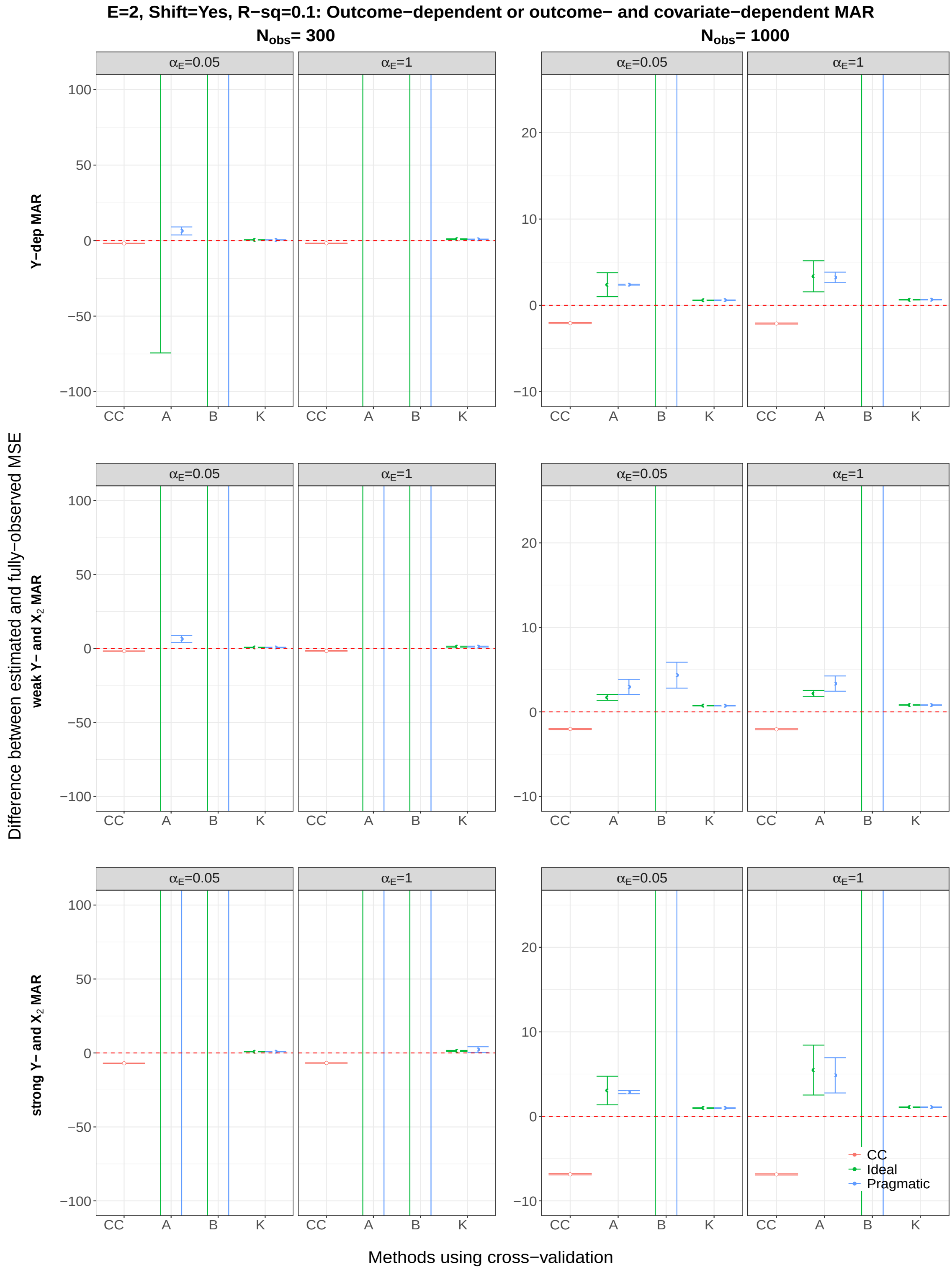

Figure S19: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

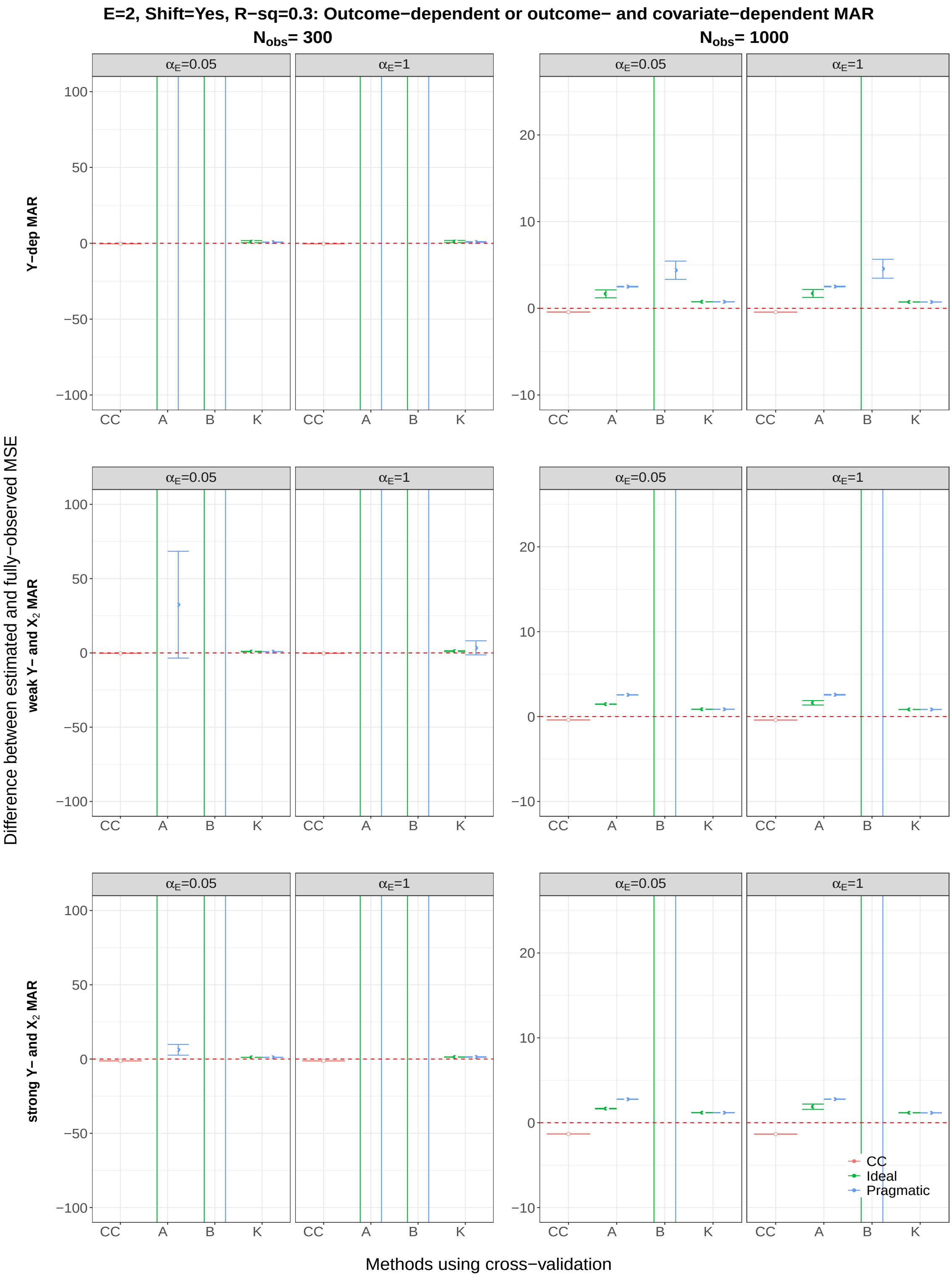

Figure S20: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

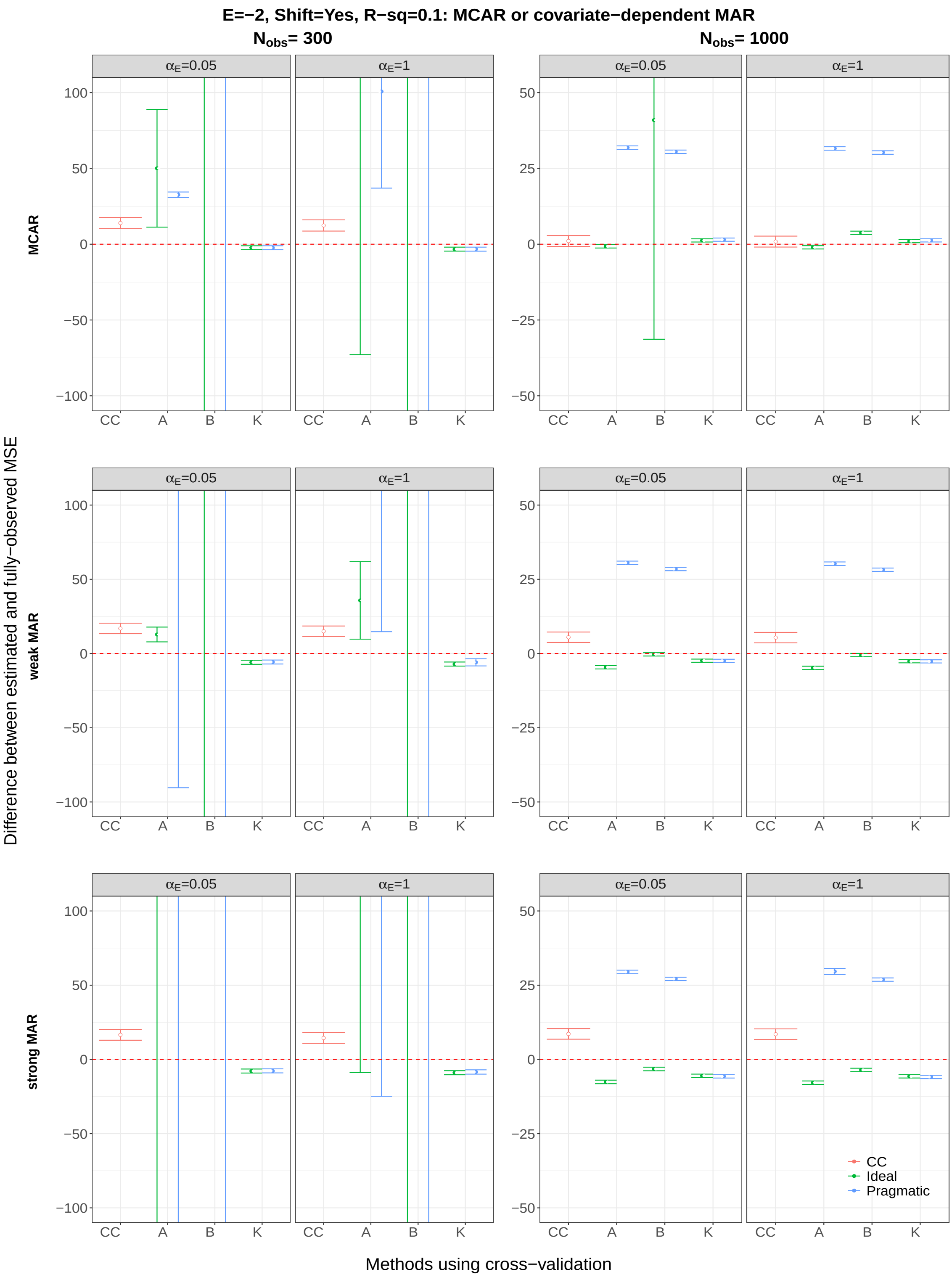

Figure S21: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

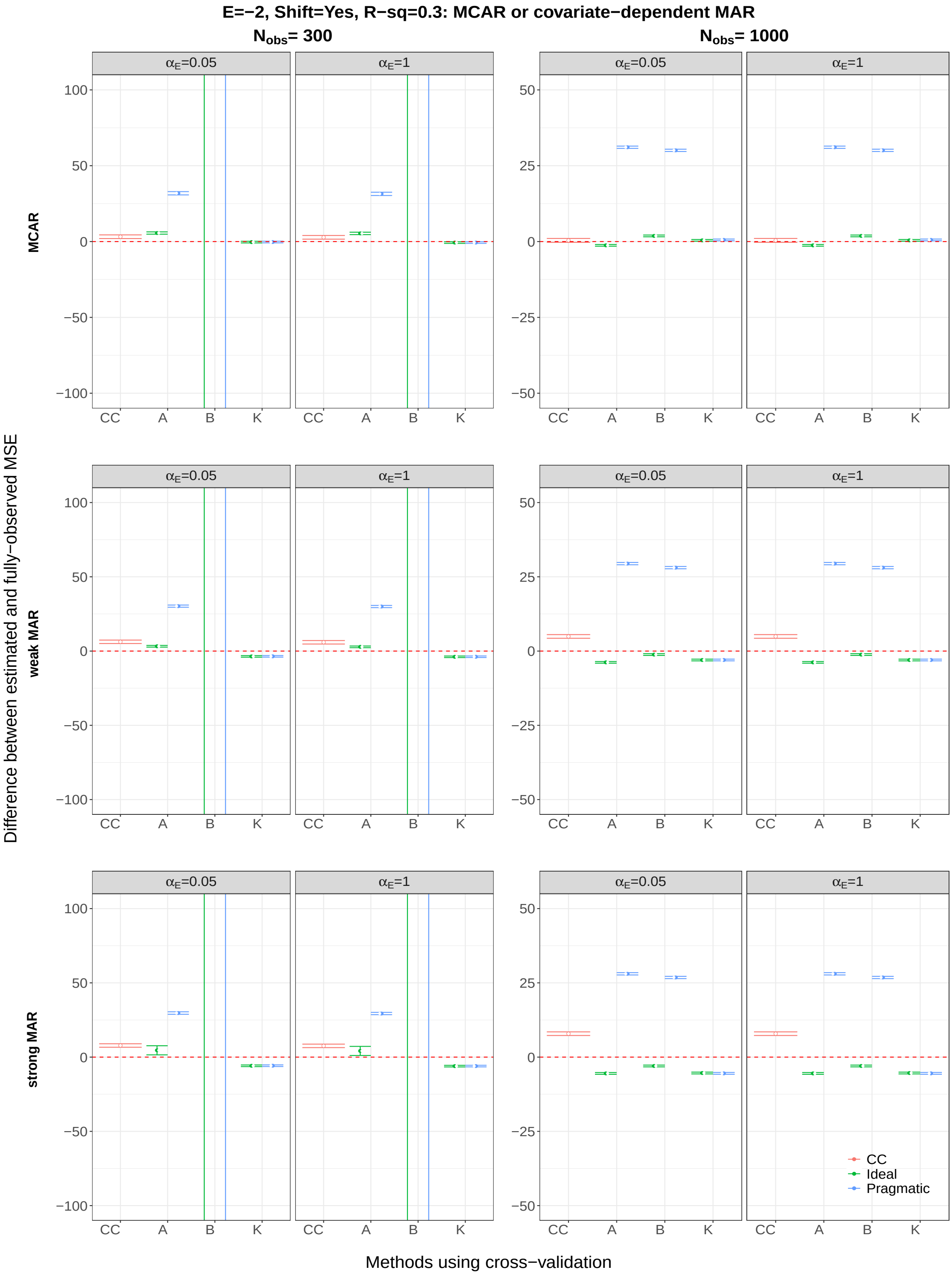

Figure S22: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

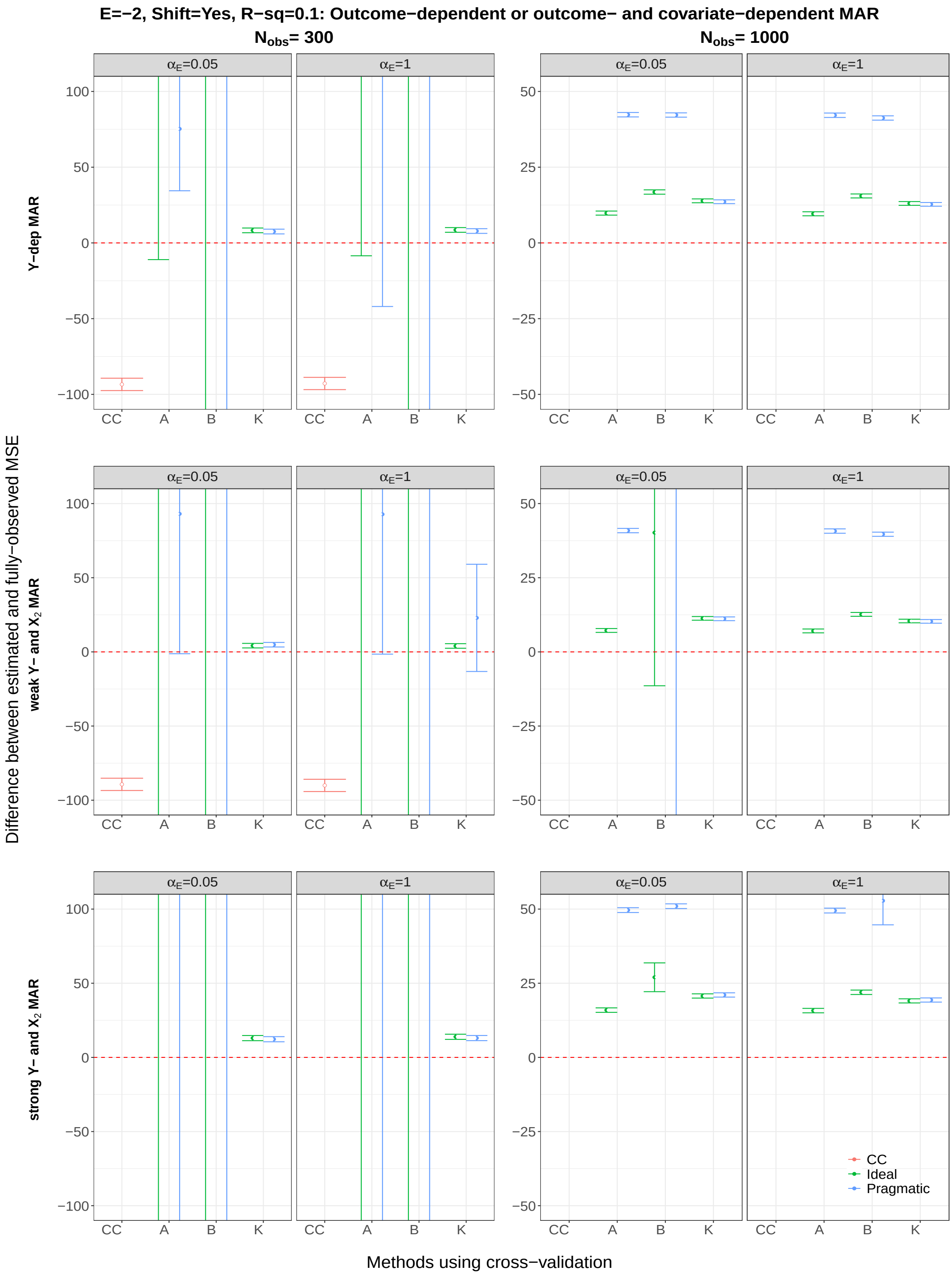

Figure S23: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

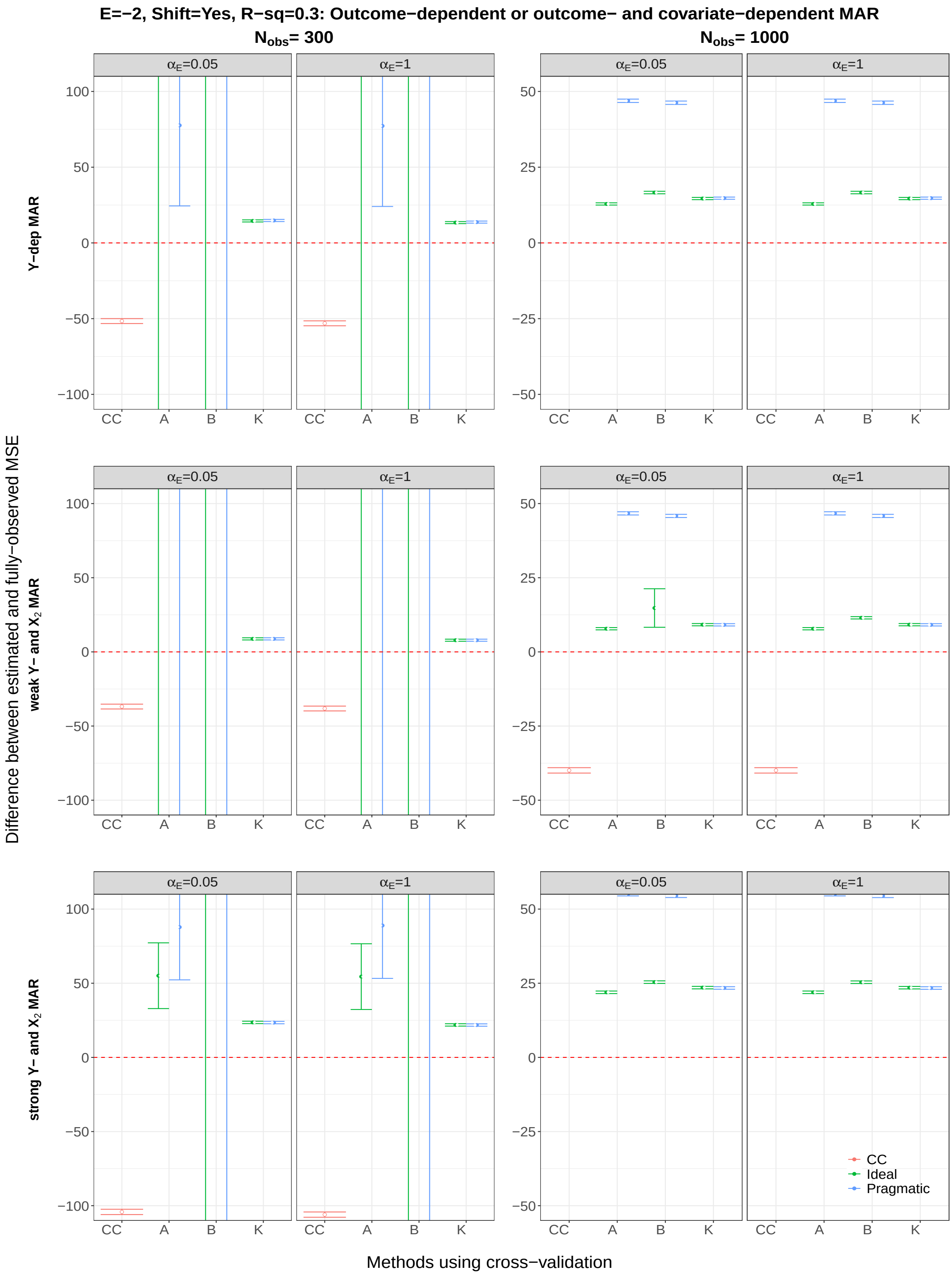

Figure S24: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7.1.3 $\beta_2 = 0$ and an origin shift transformation has not been applied

True exponent is 0

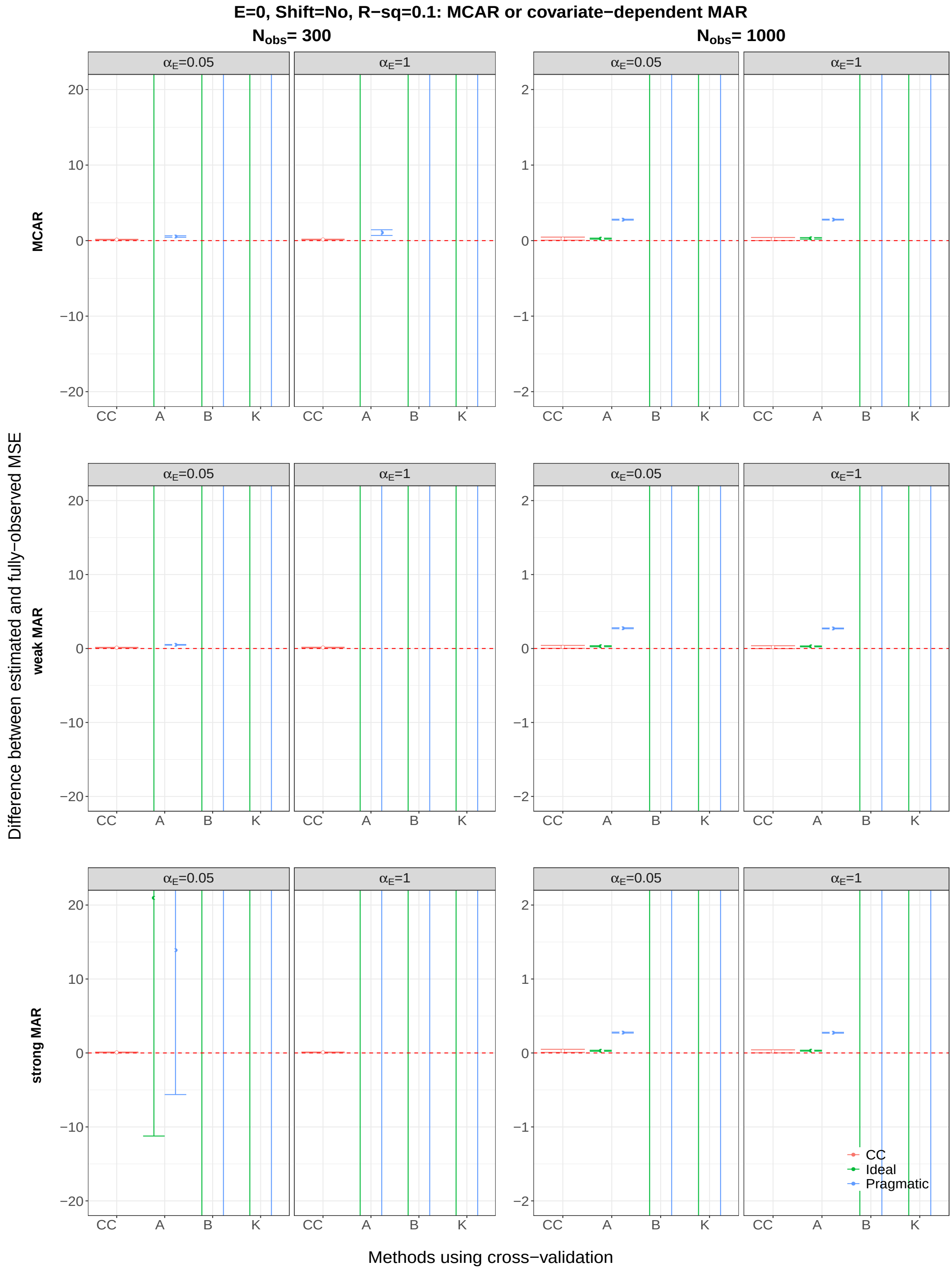

Figure S25: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

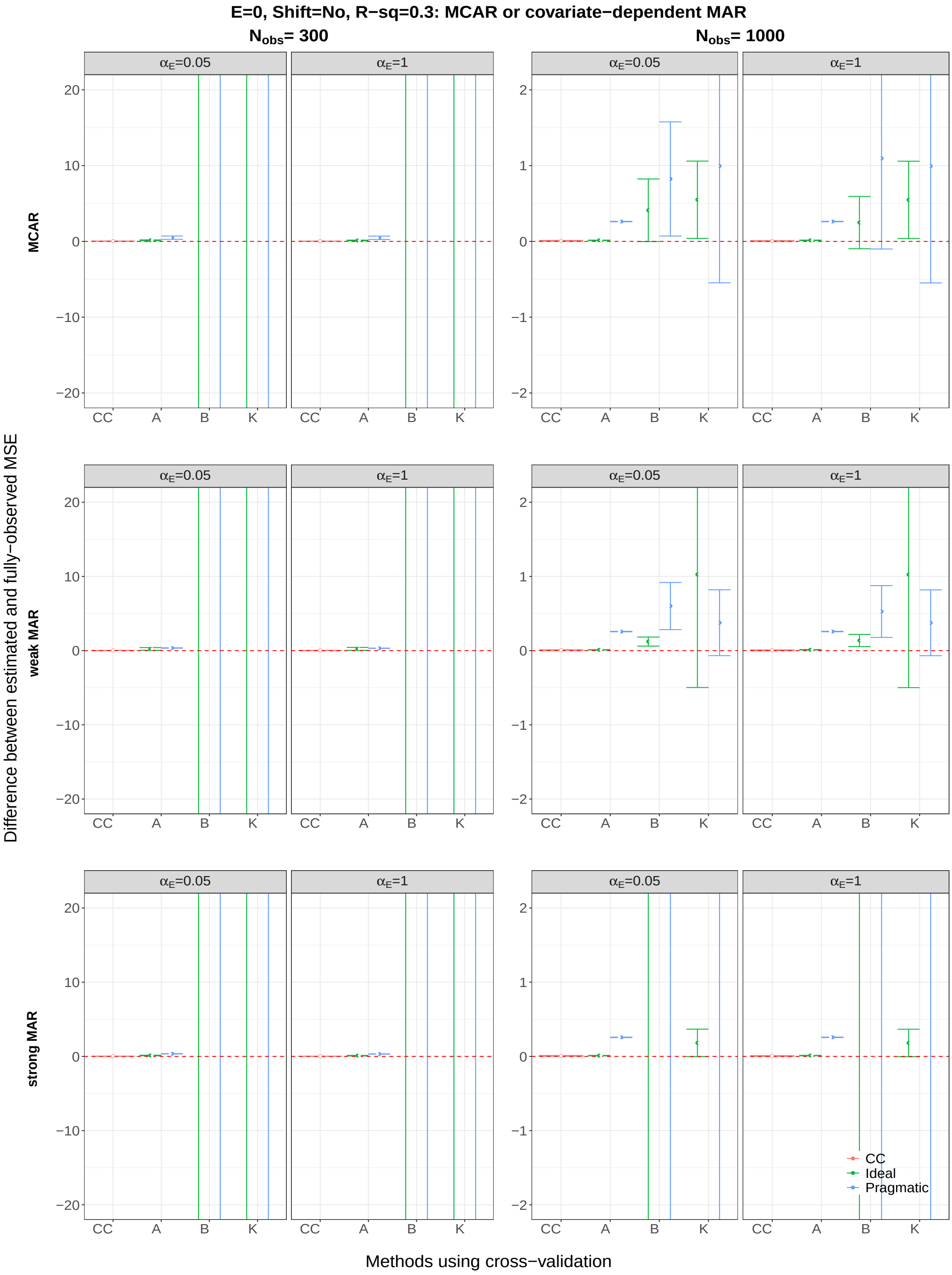

Figure S26: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

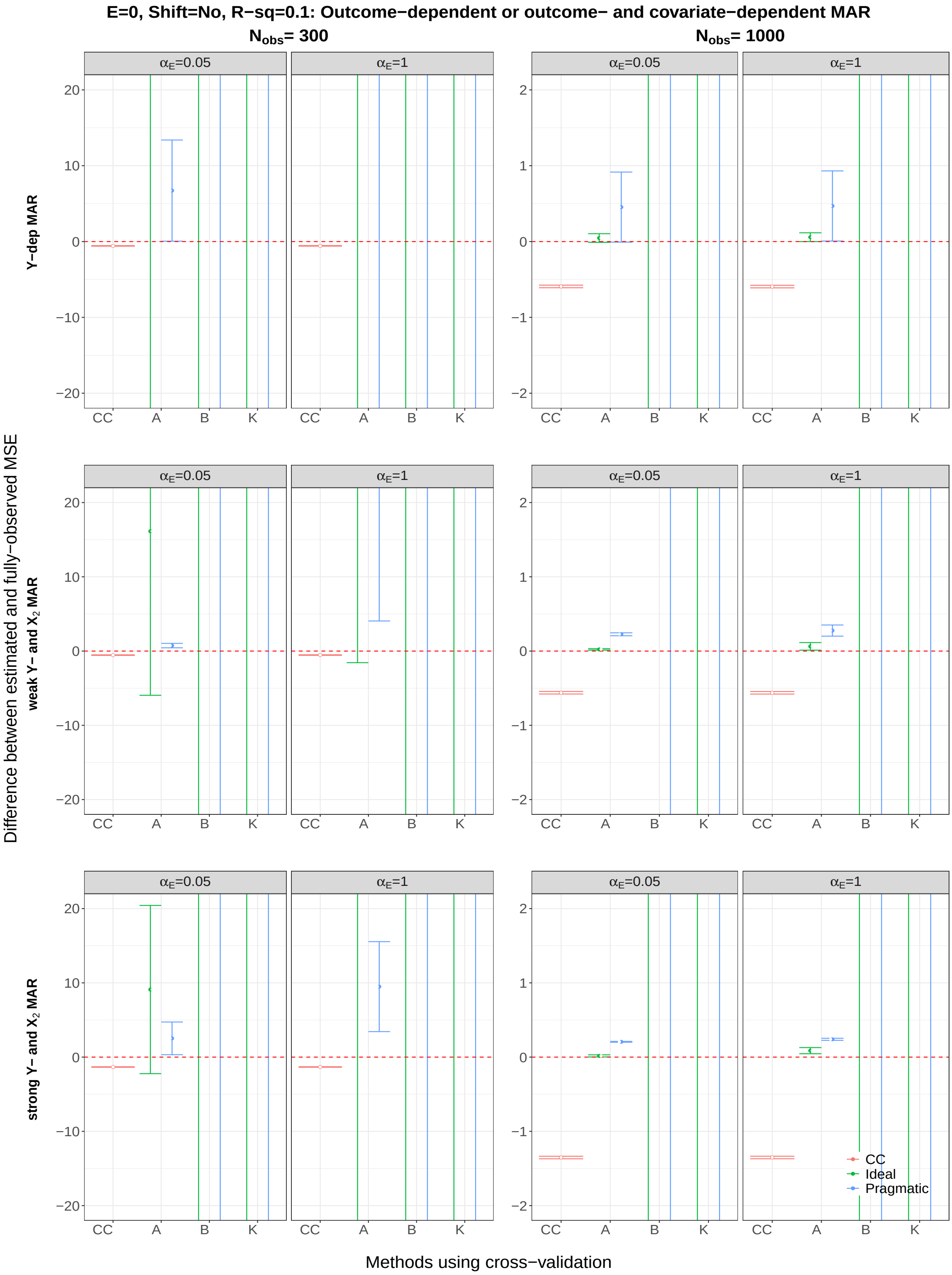

Figure S27: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

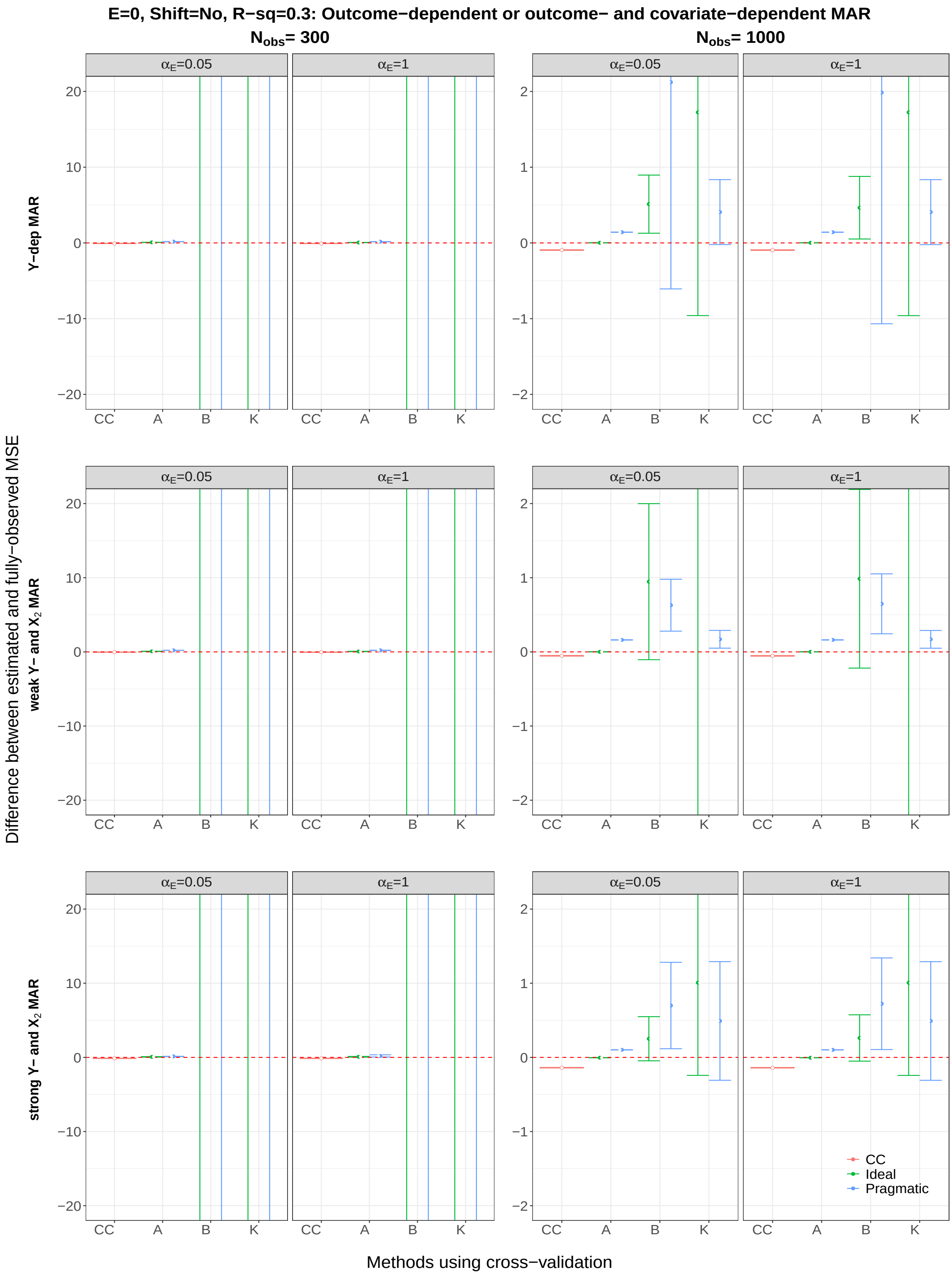

Figure S28: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

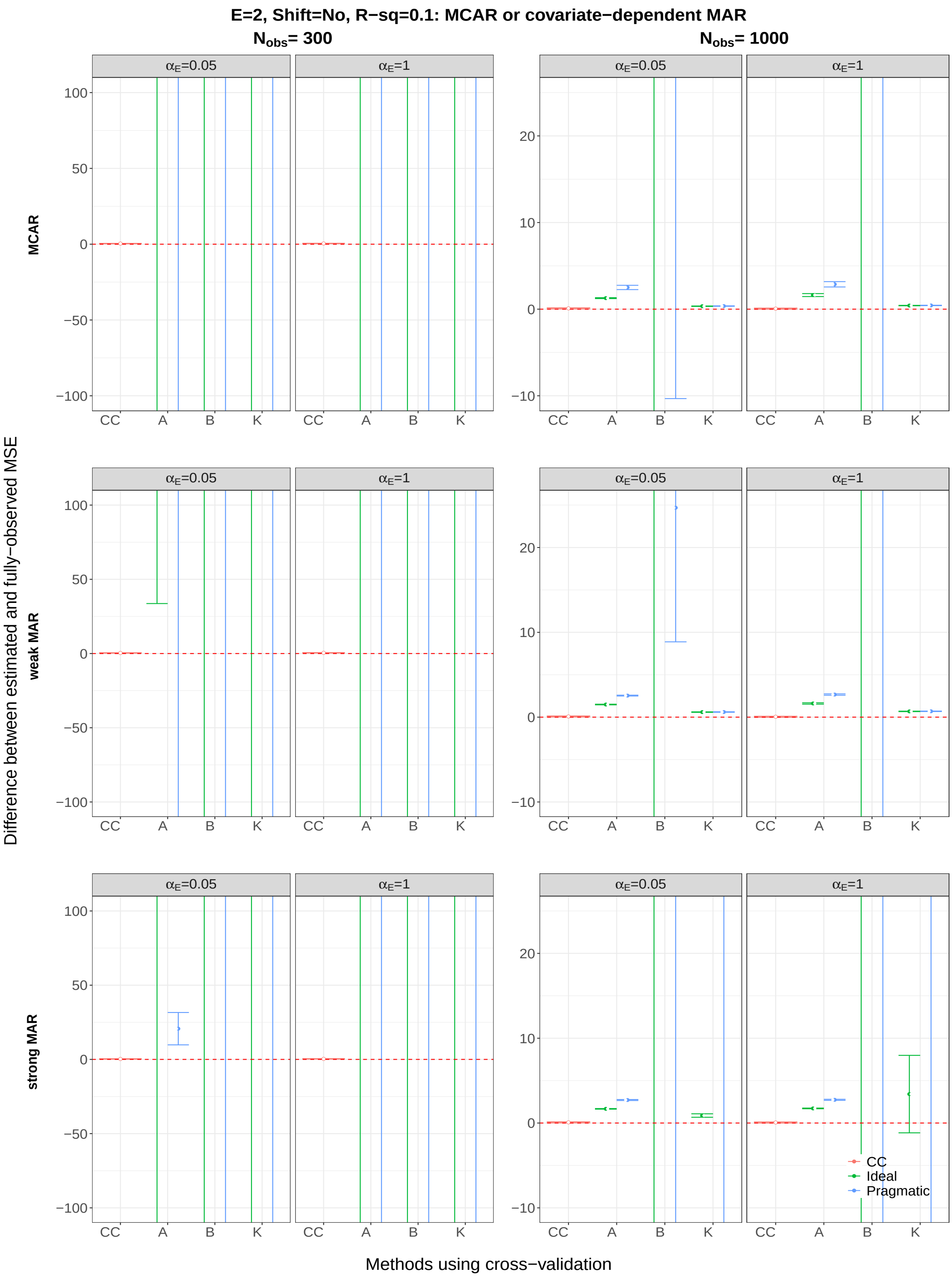

Figure S29: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

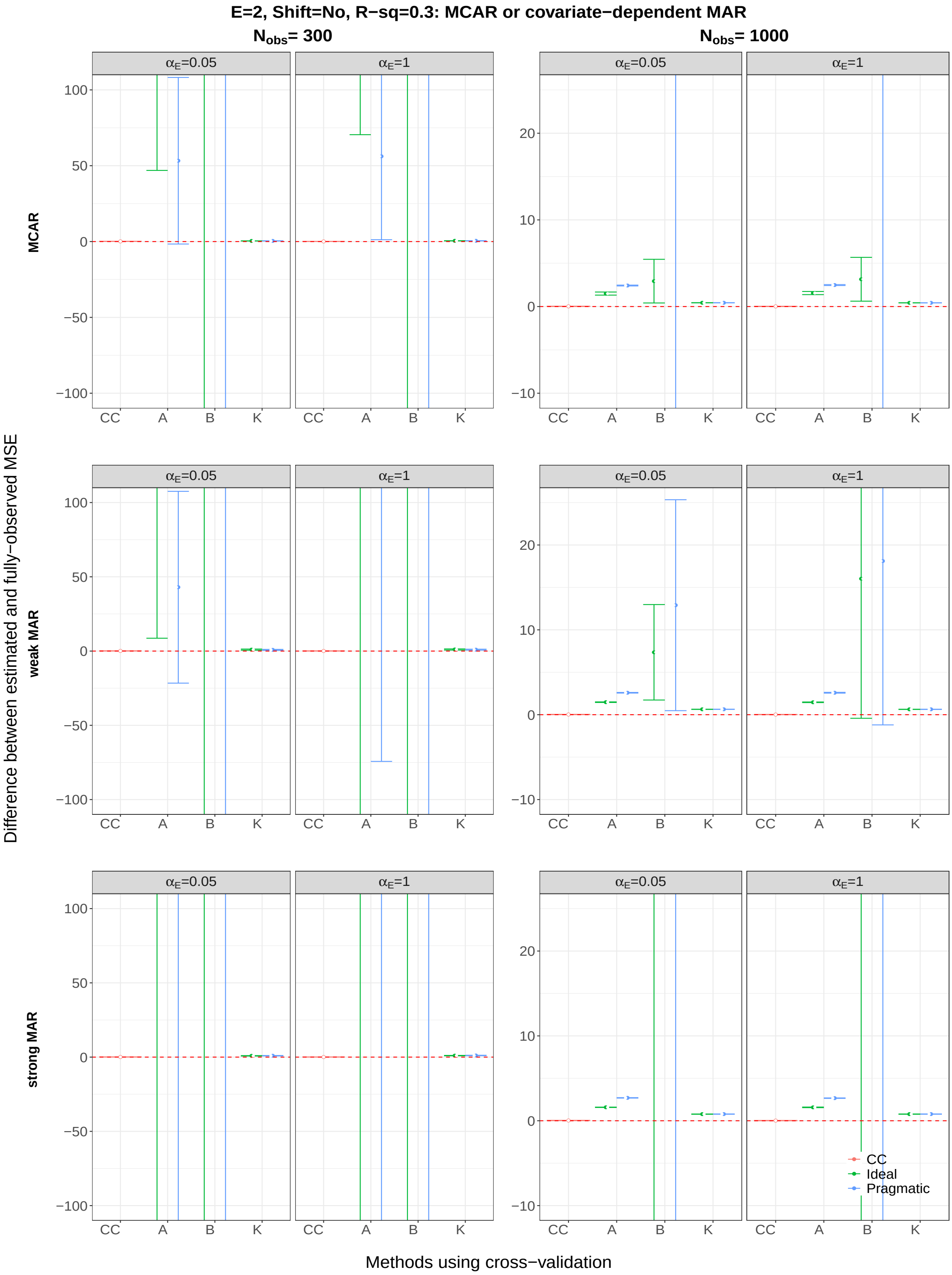

Figure S30: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

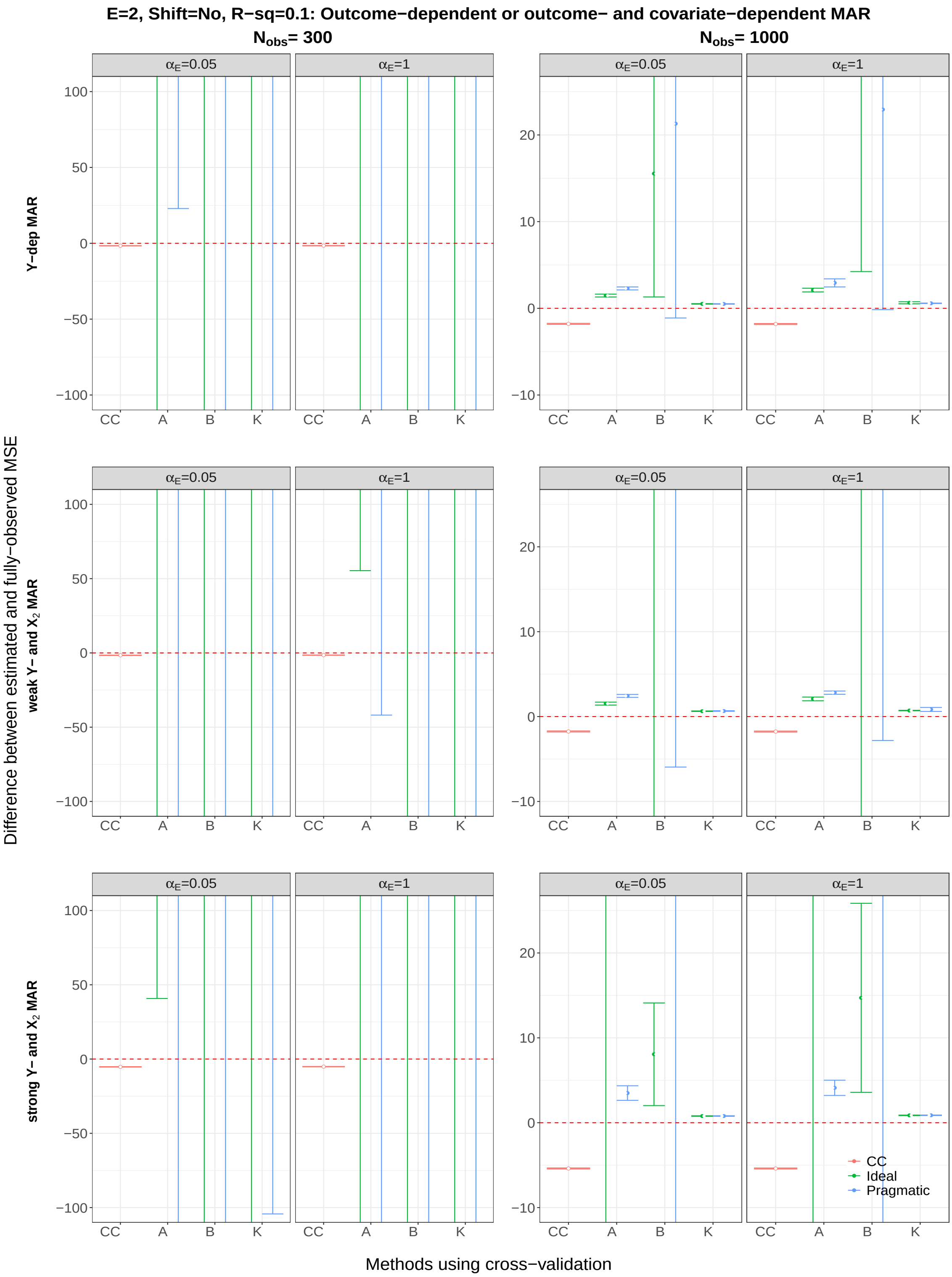

Figure S31: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

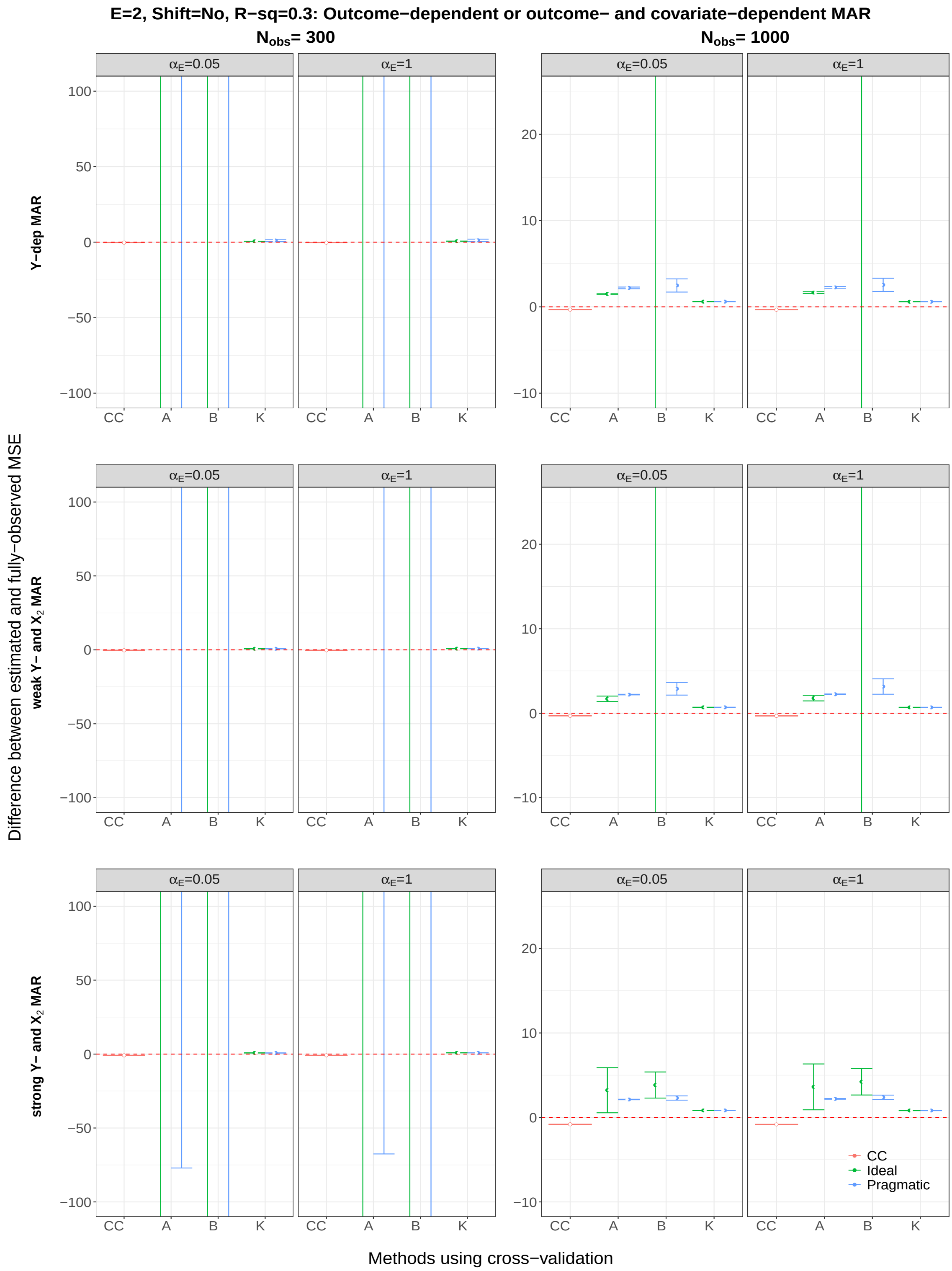

Figure S32: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

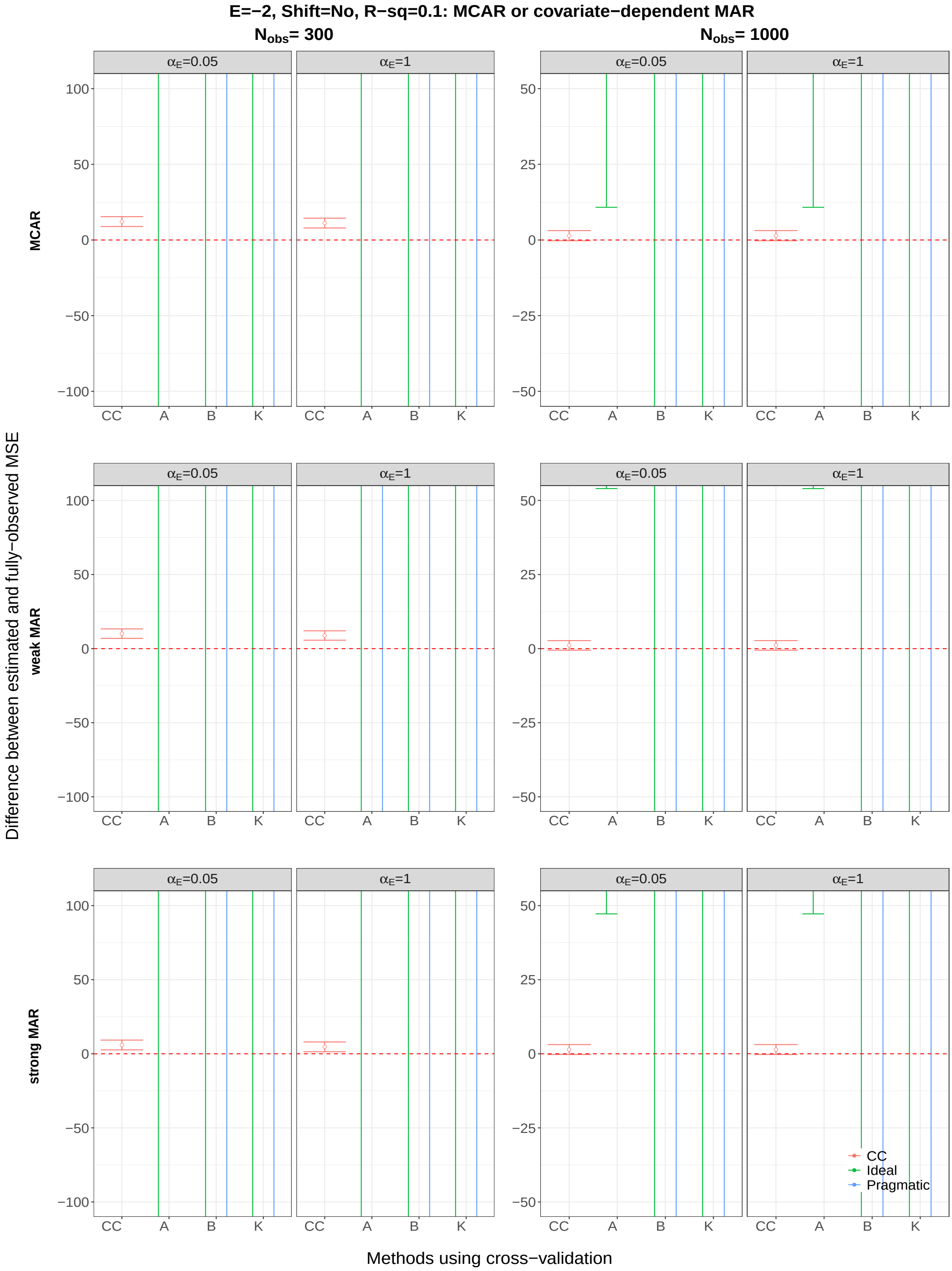

Figure S33: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

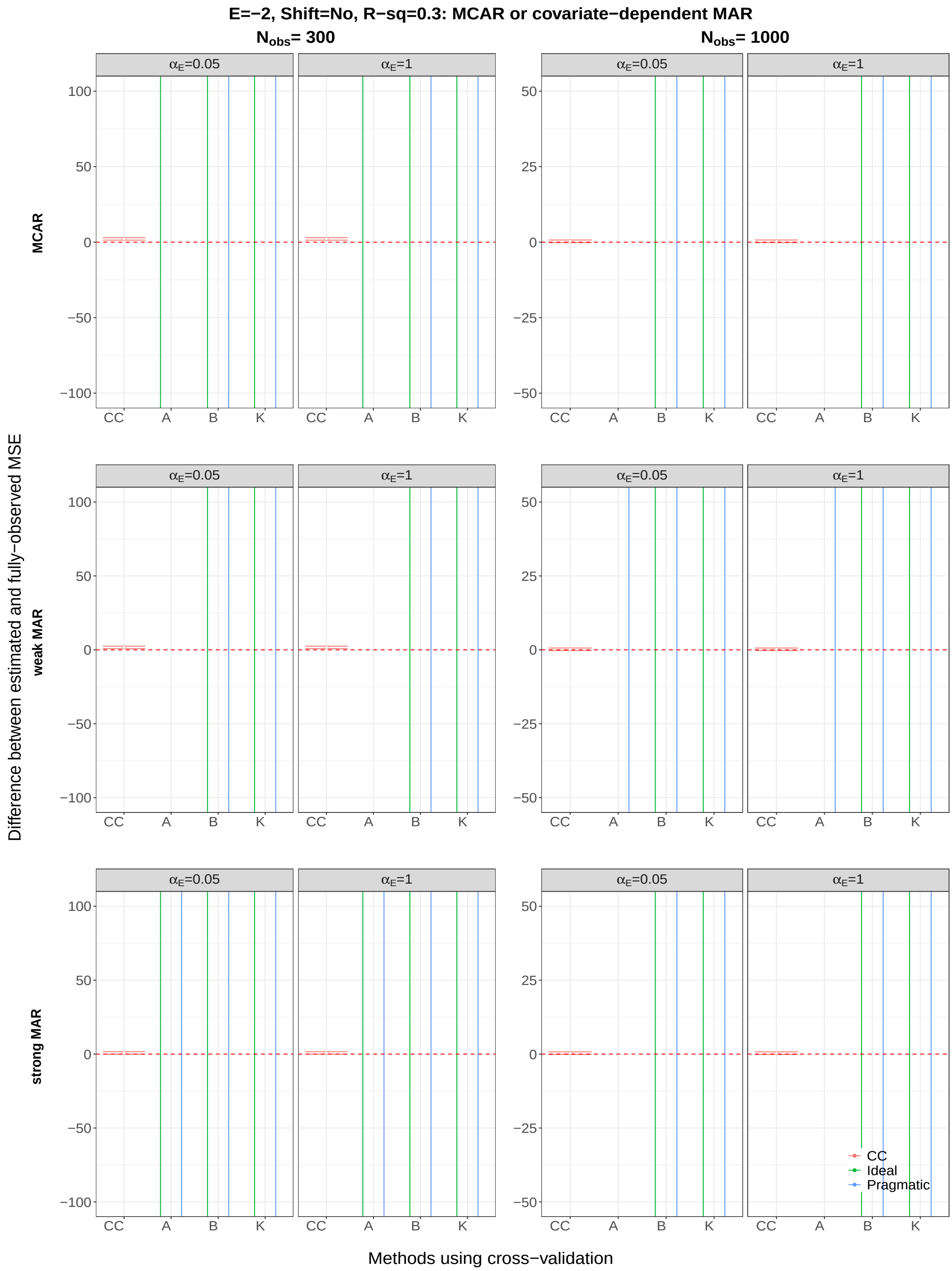

Figure S34: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

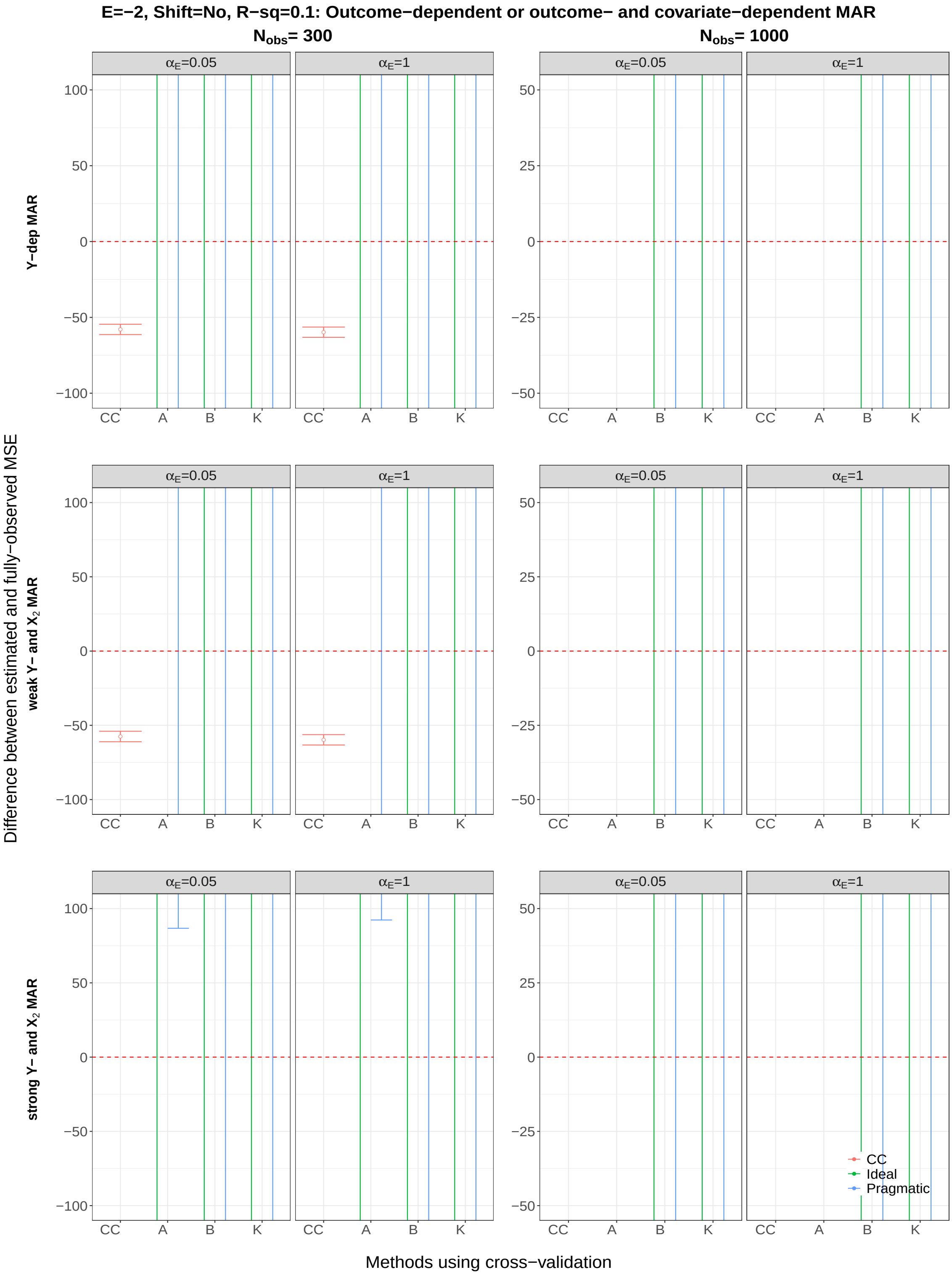

Figure S35: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

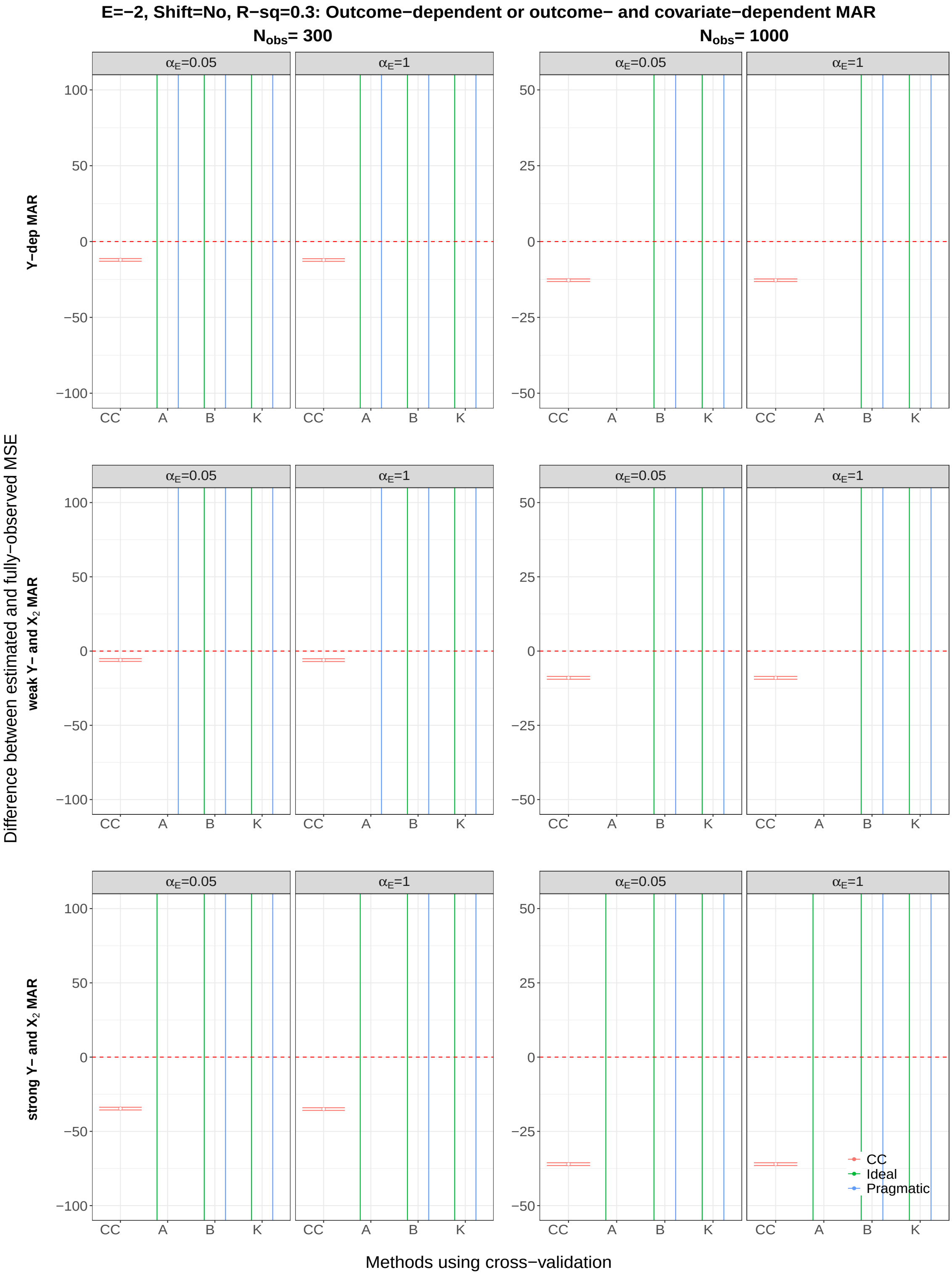

Figure S36: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7.1.4 $\beta_2 = 0$ and an origin shift transformation has been applied

True exponent is 0

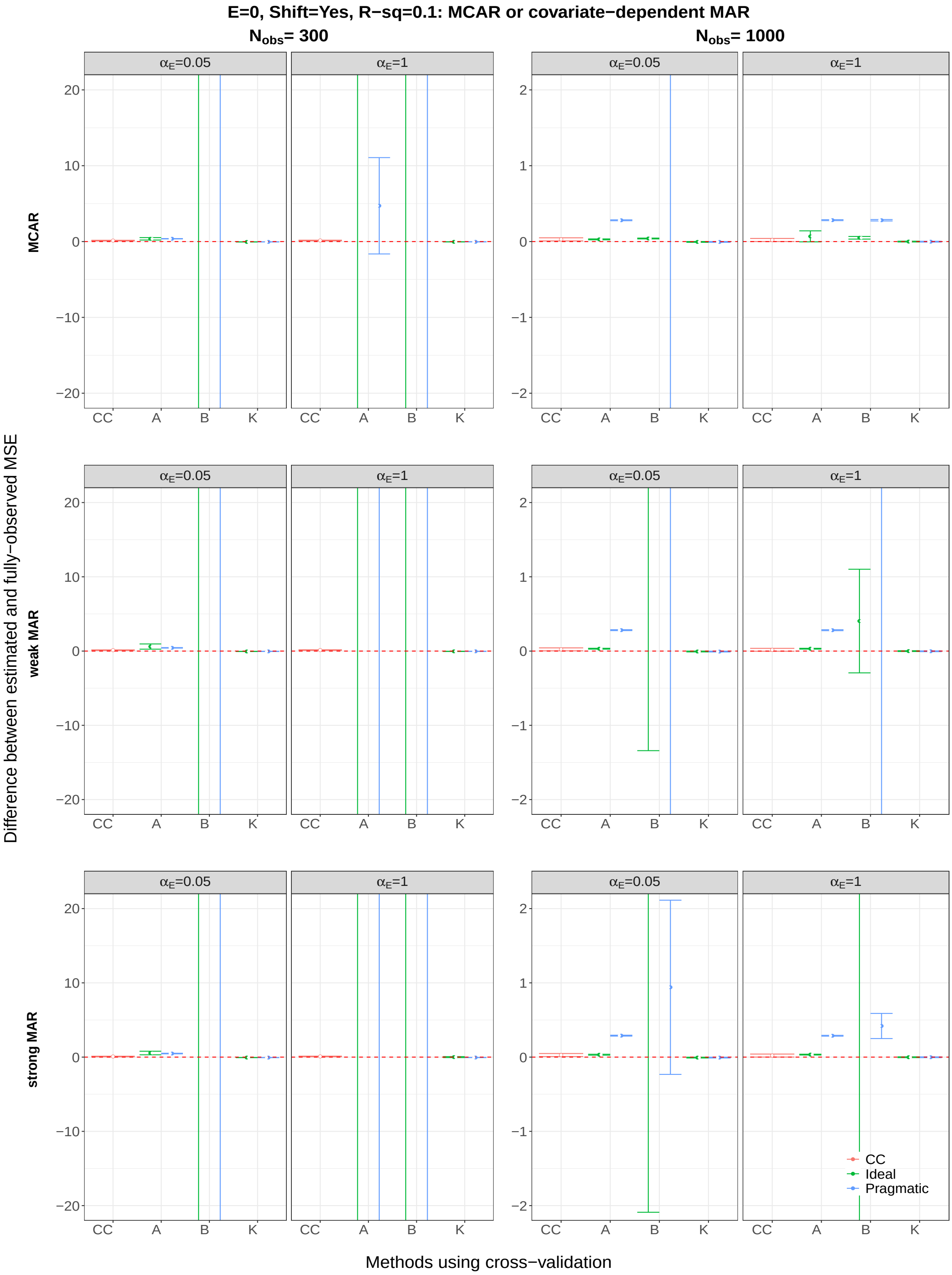

Figure S37: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

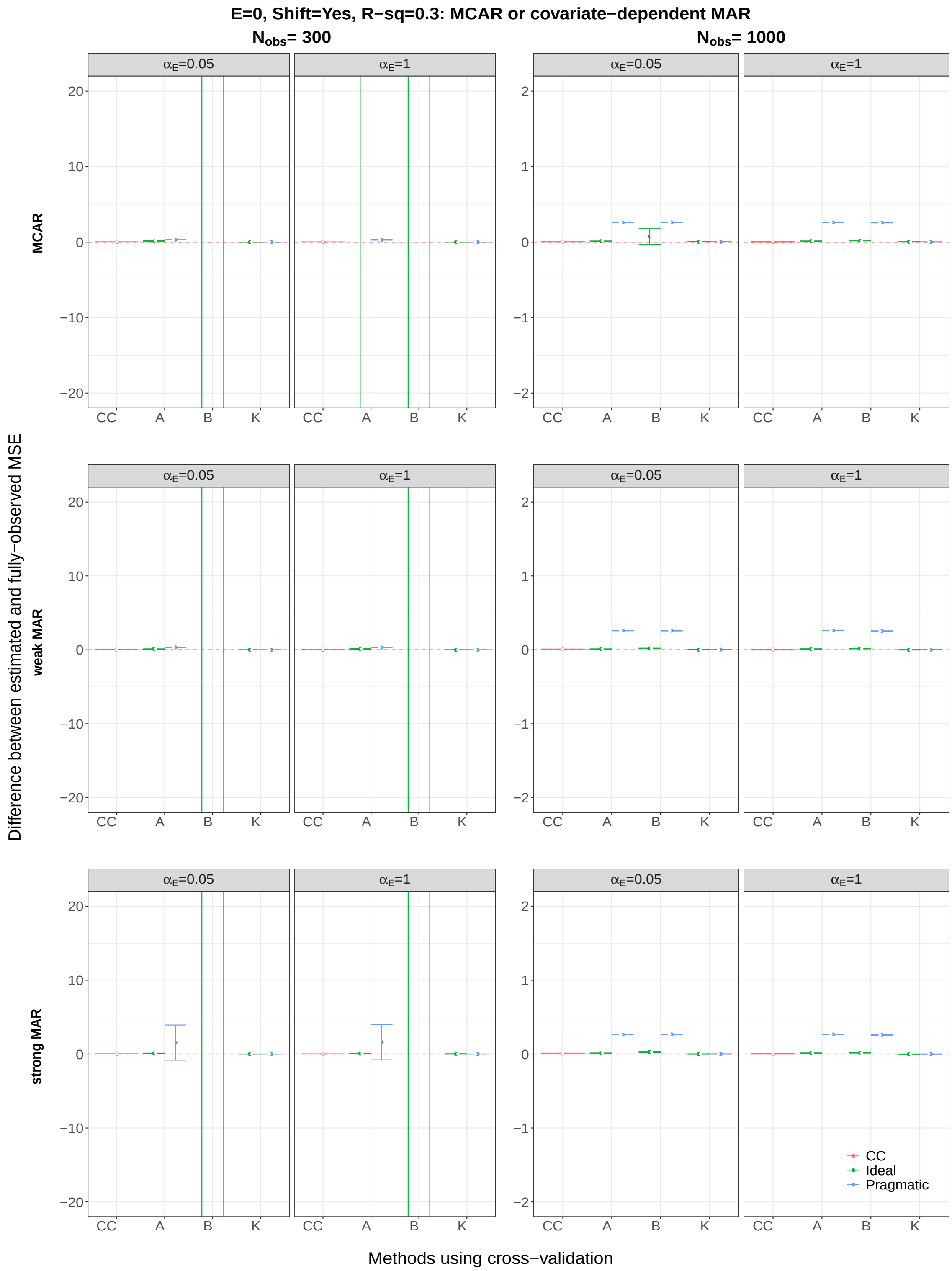

Figure S38: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

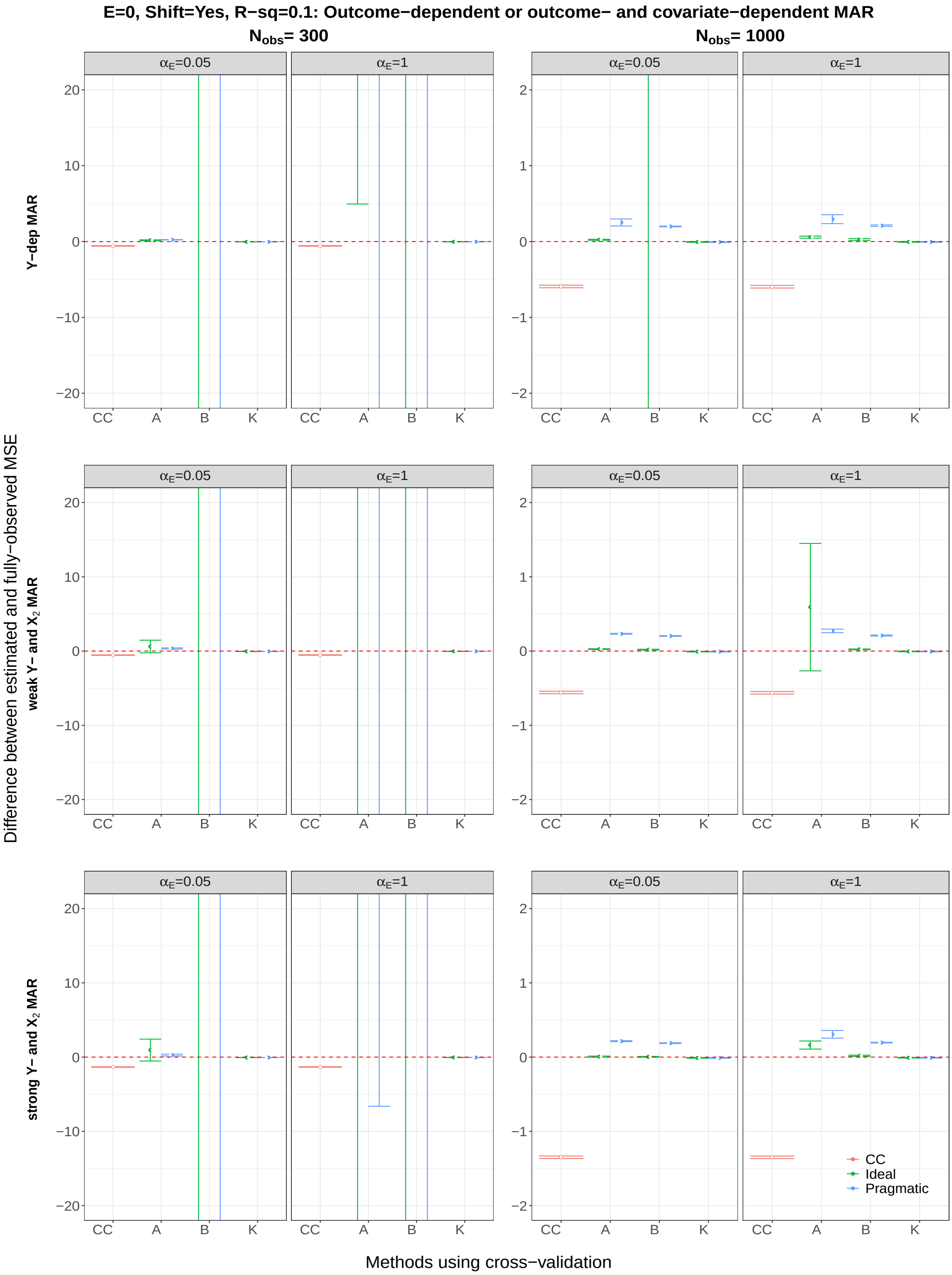

Figure S39: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

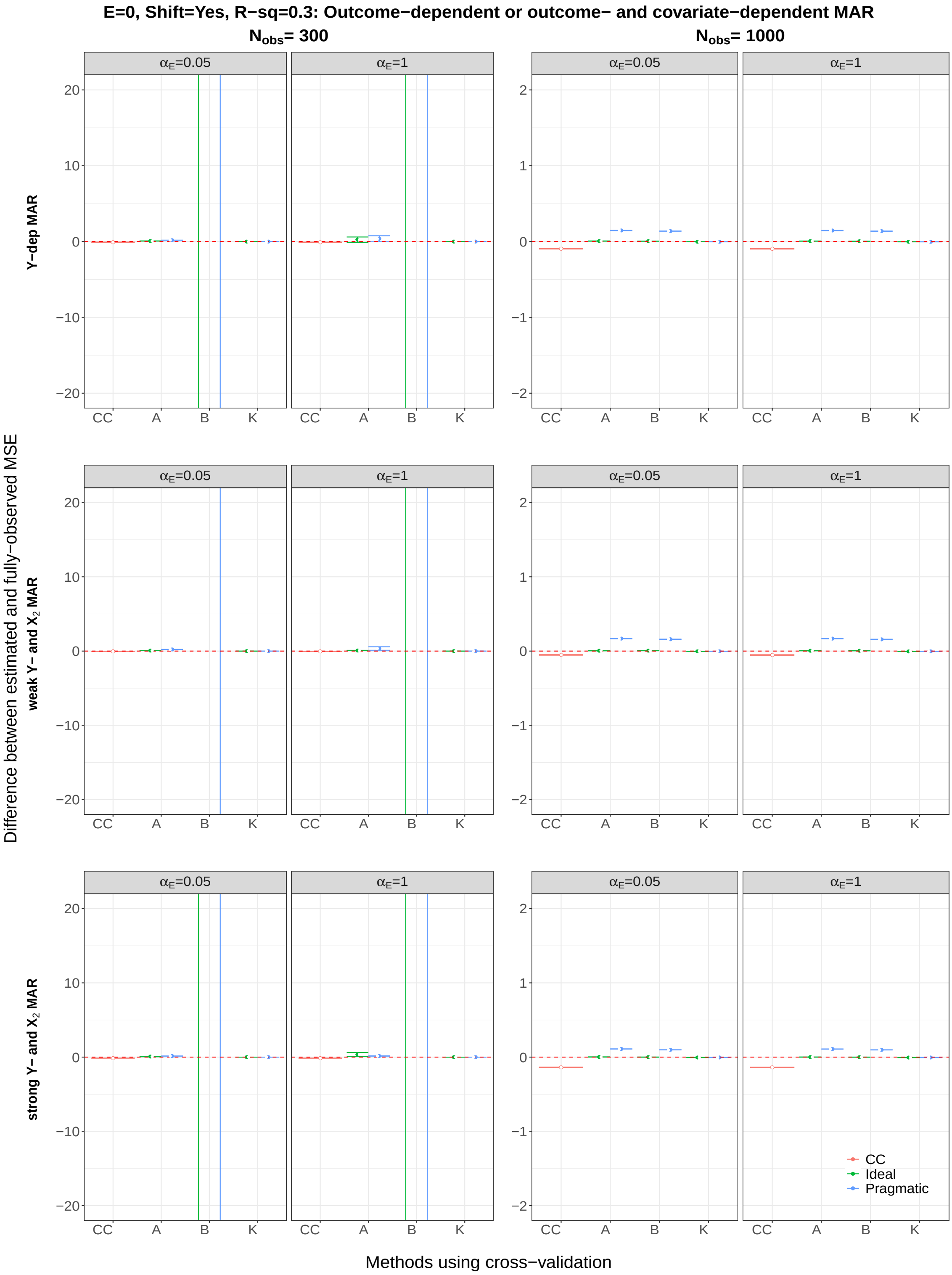

Figure S40: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

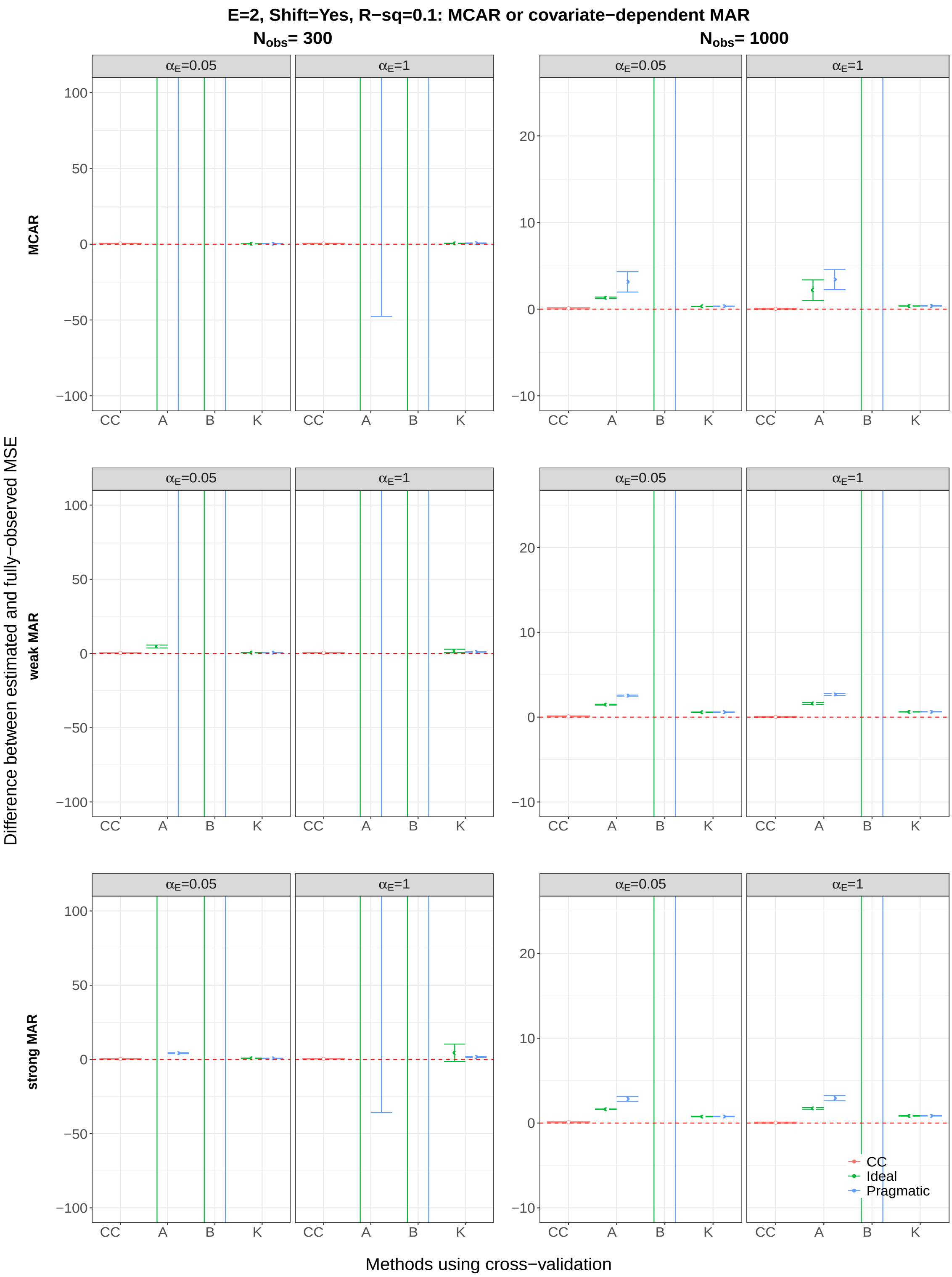

Figure S41: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

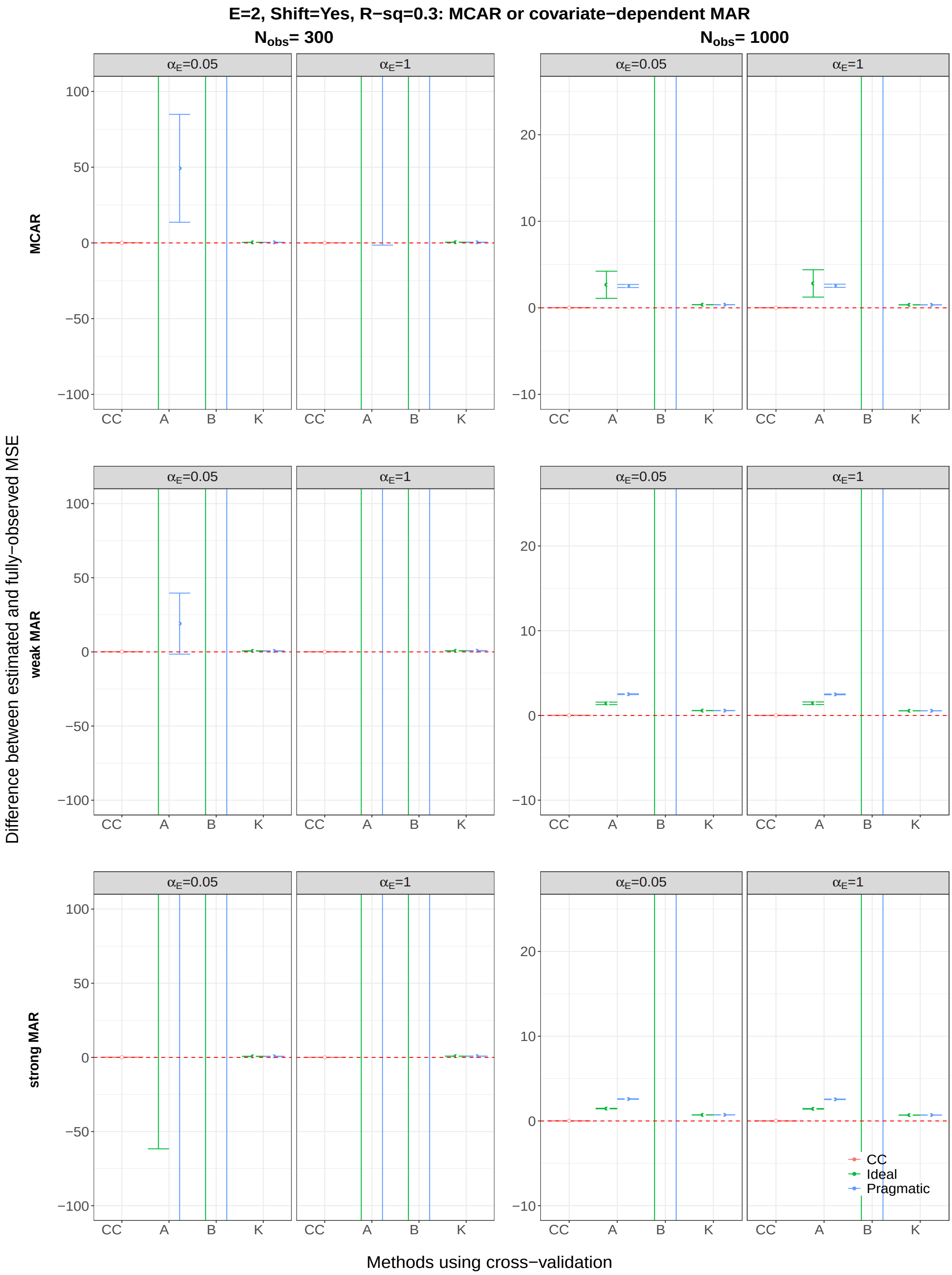

Figure S42: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

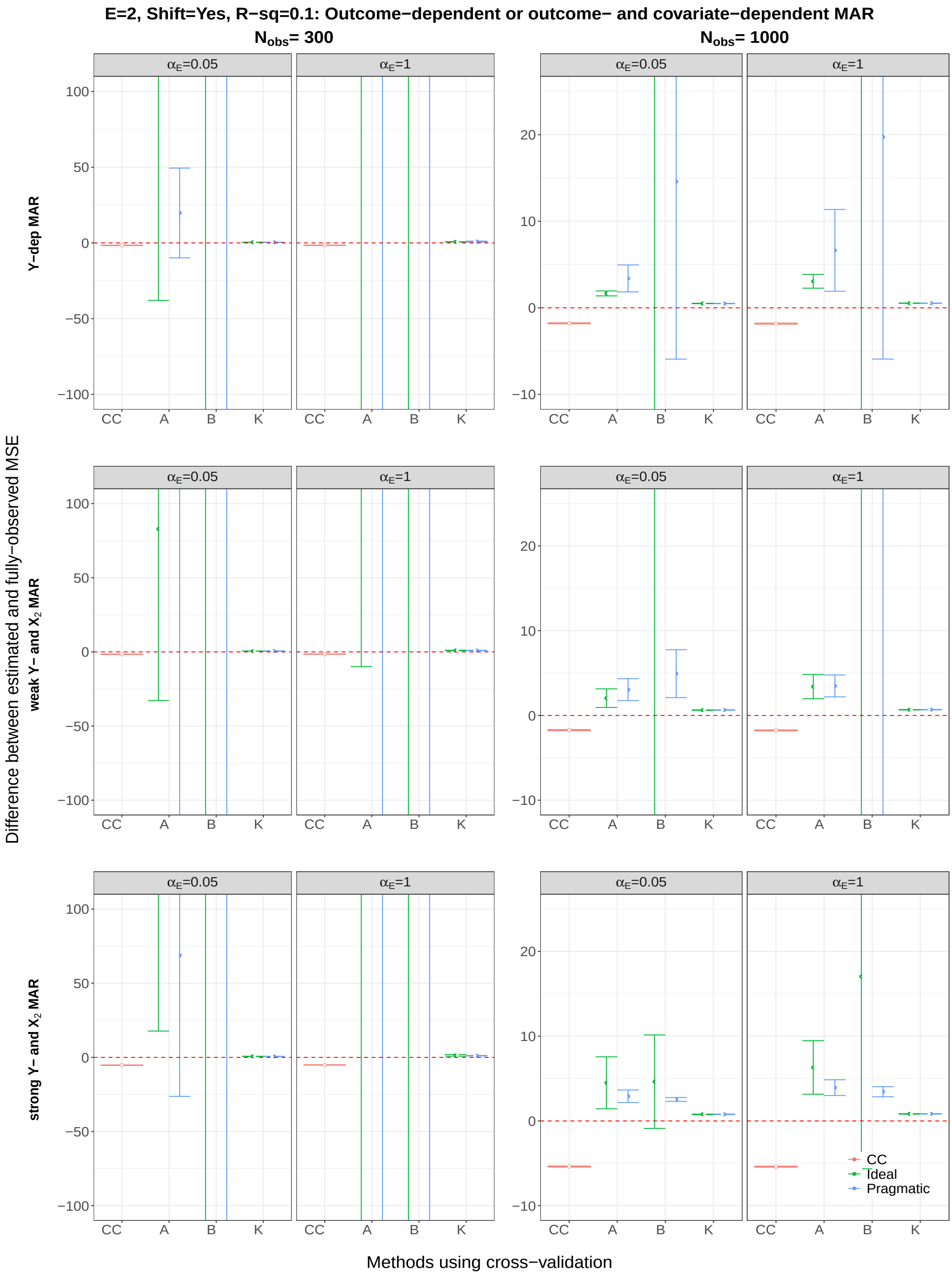

Figure S43: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

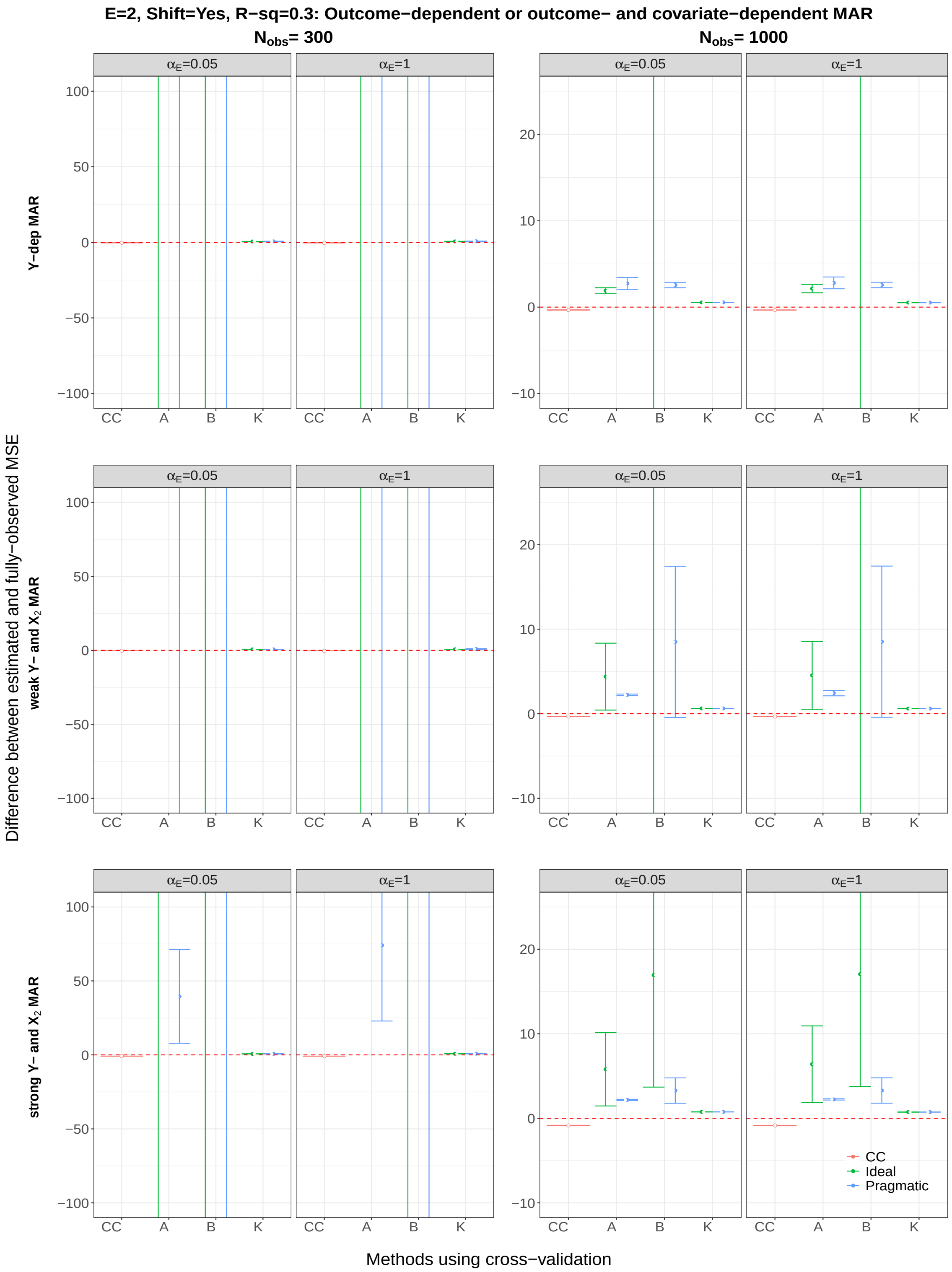

Figure S44: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

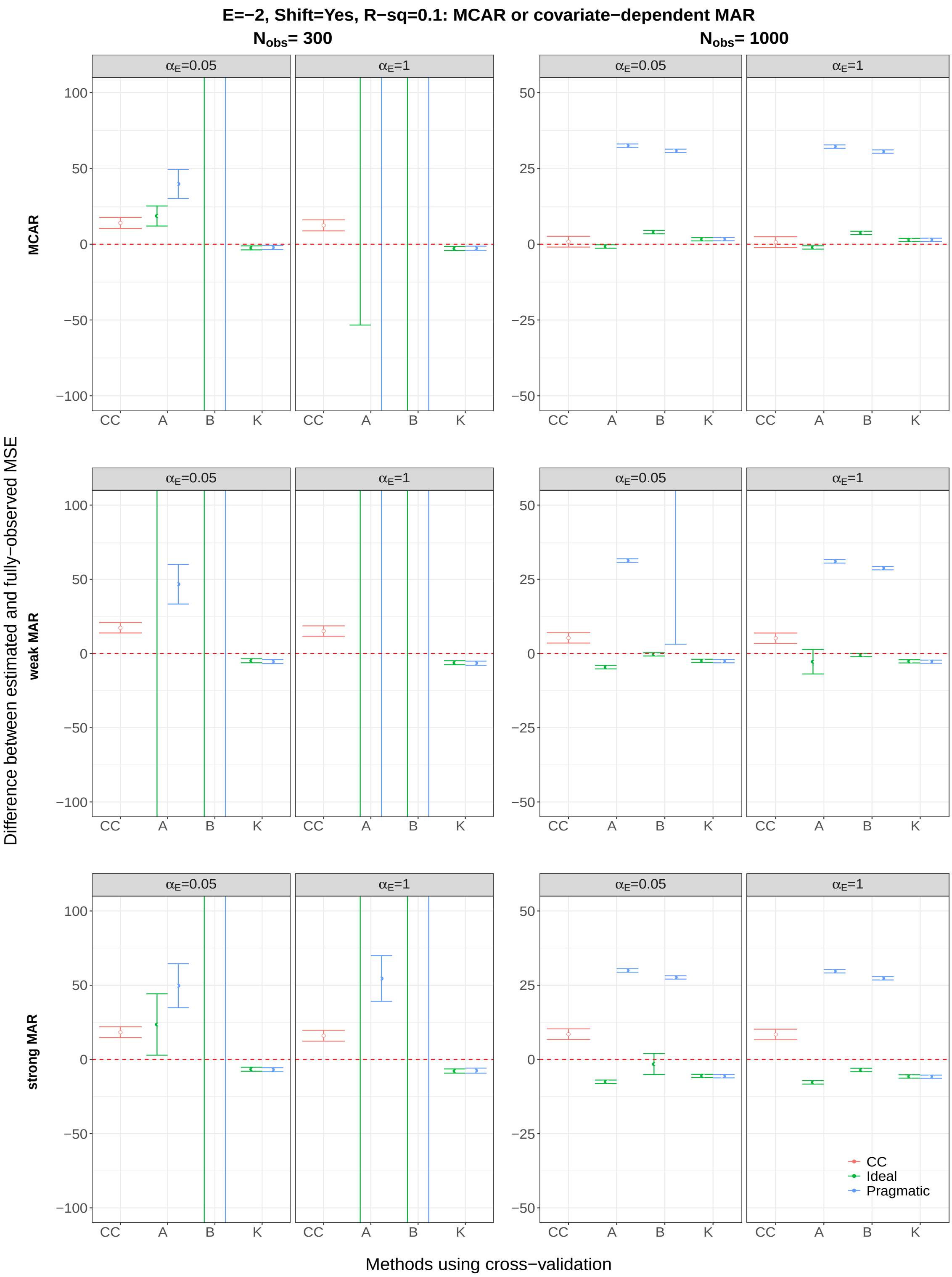

Figure S45: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

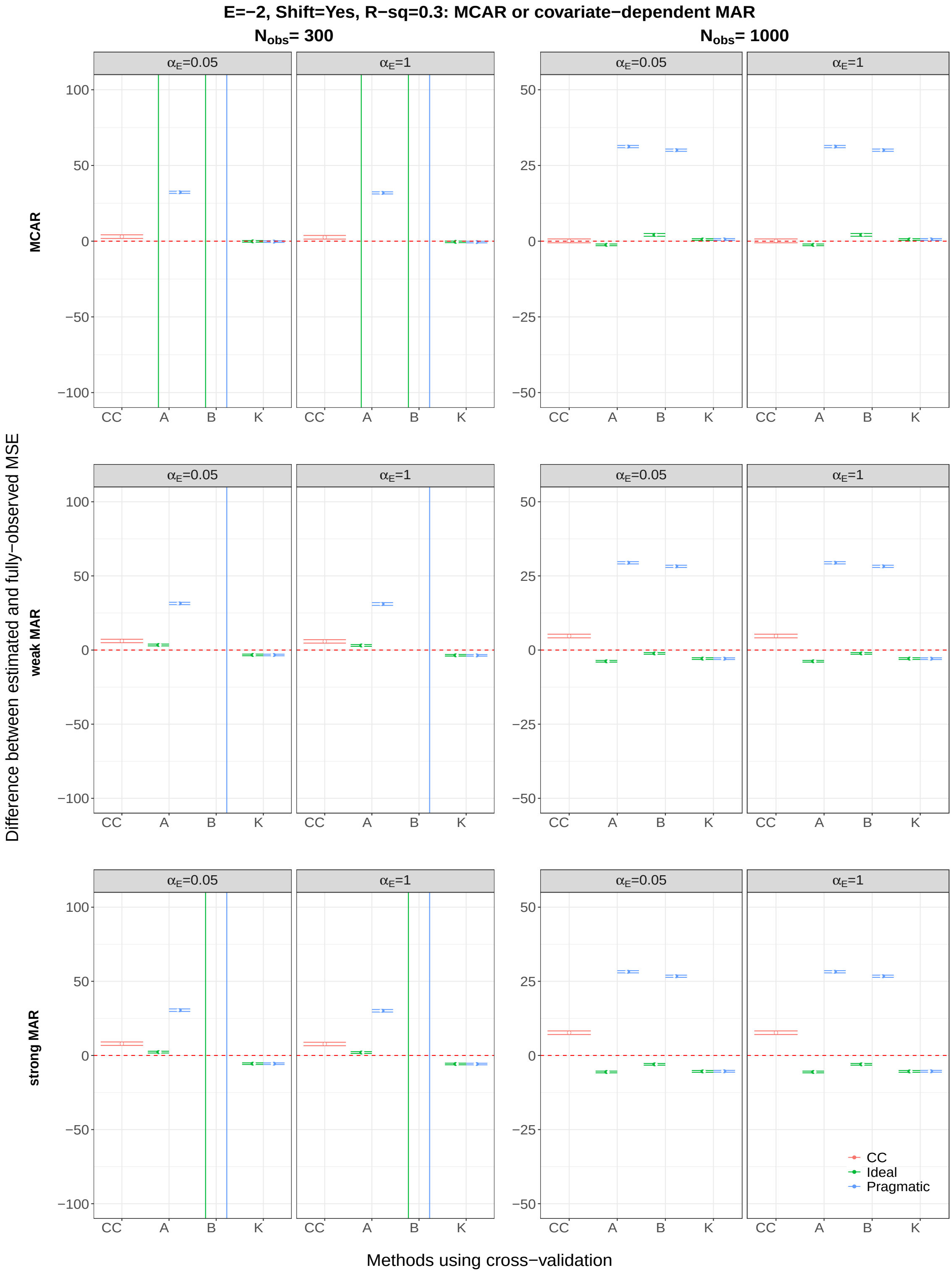

Figure S46: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

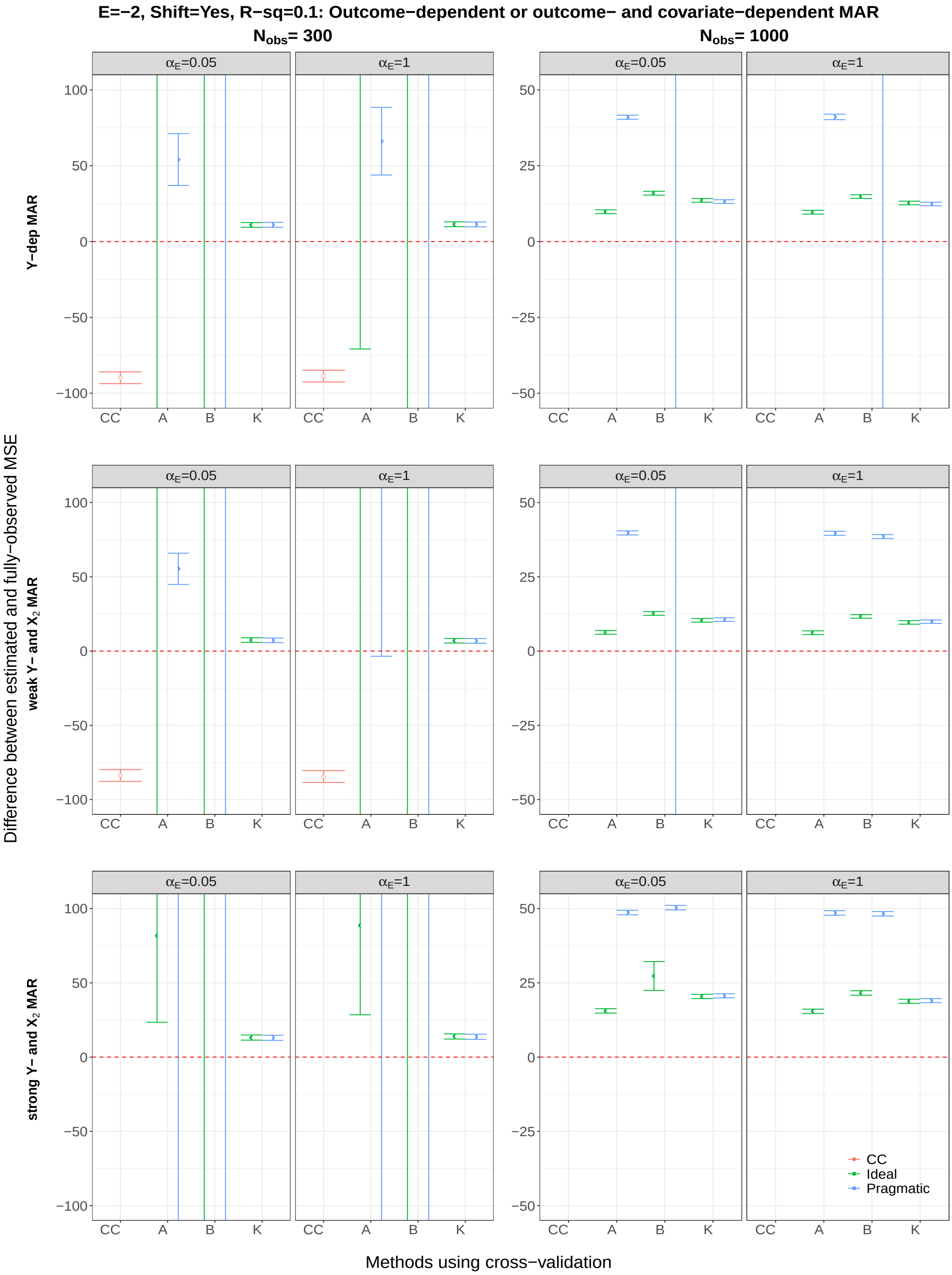

Figure S47: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

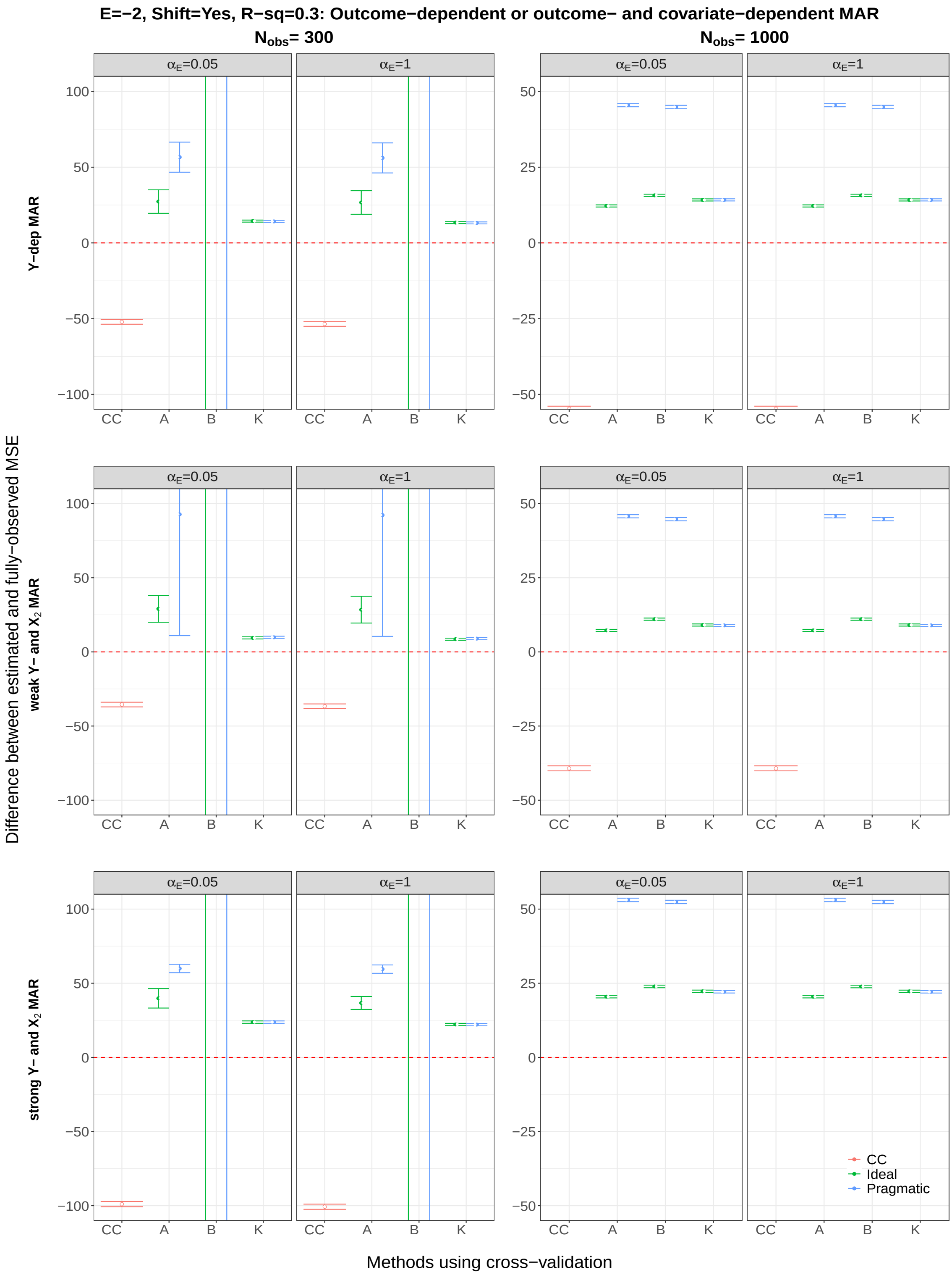

Figure S48: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7.2 The 0.632 bootstrap

S7.2.1 $\beta_2 = 1$ and an origin shift transformation has not been applied

True exponent is 0

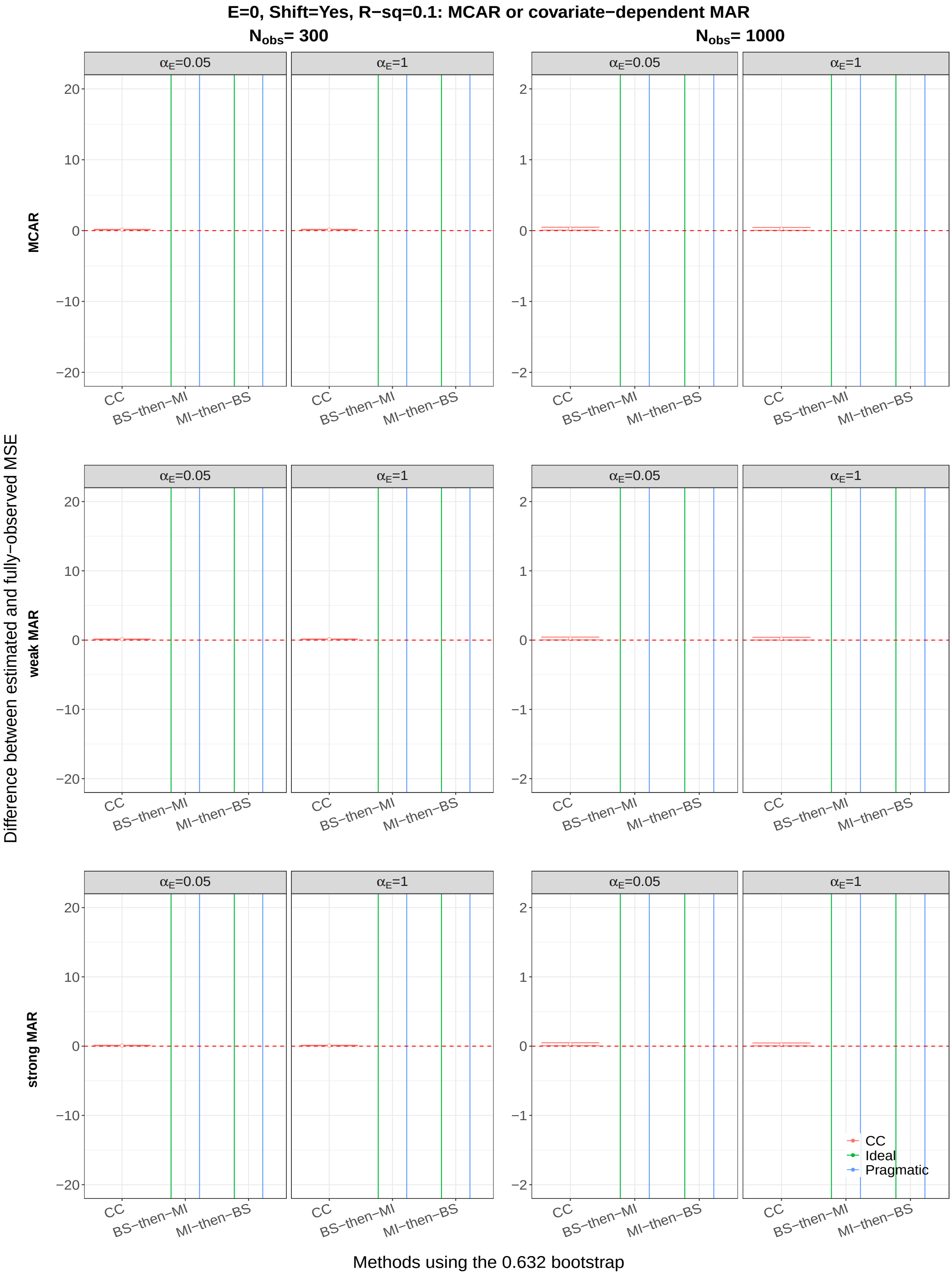

Figure S49: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

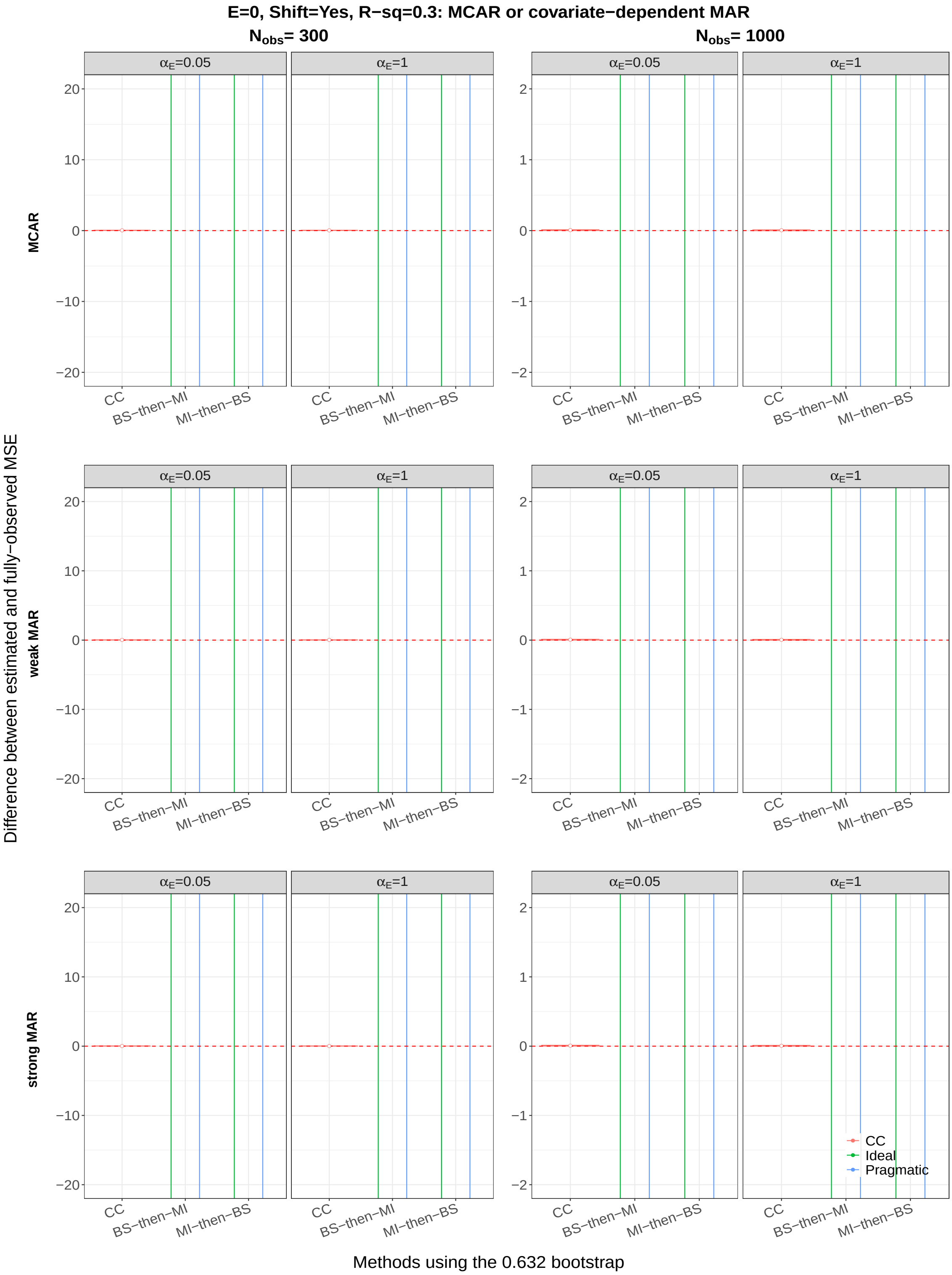

Figure S50: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

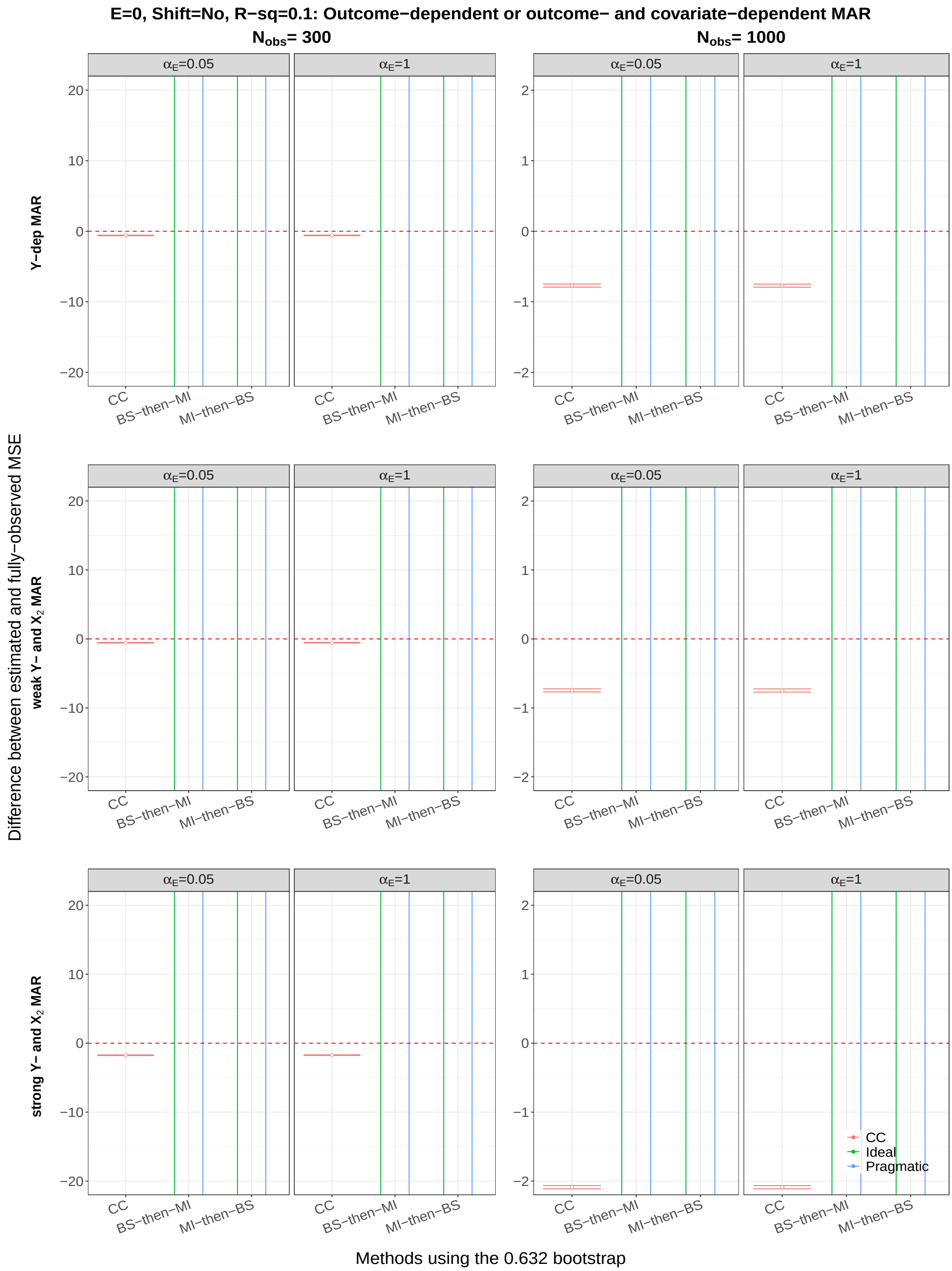

Figure S51: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

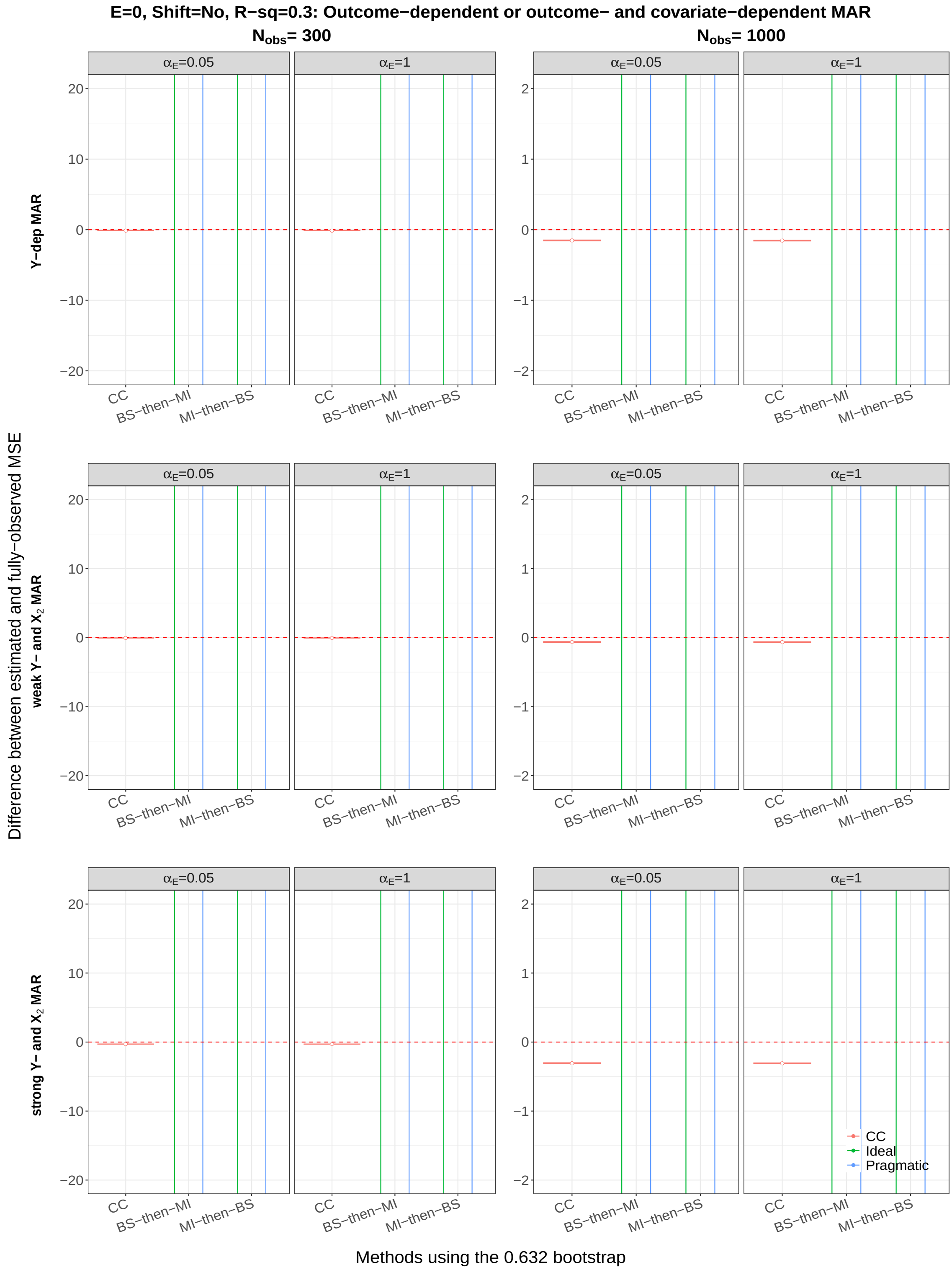

Figure S52: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

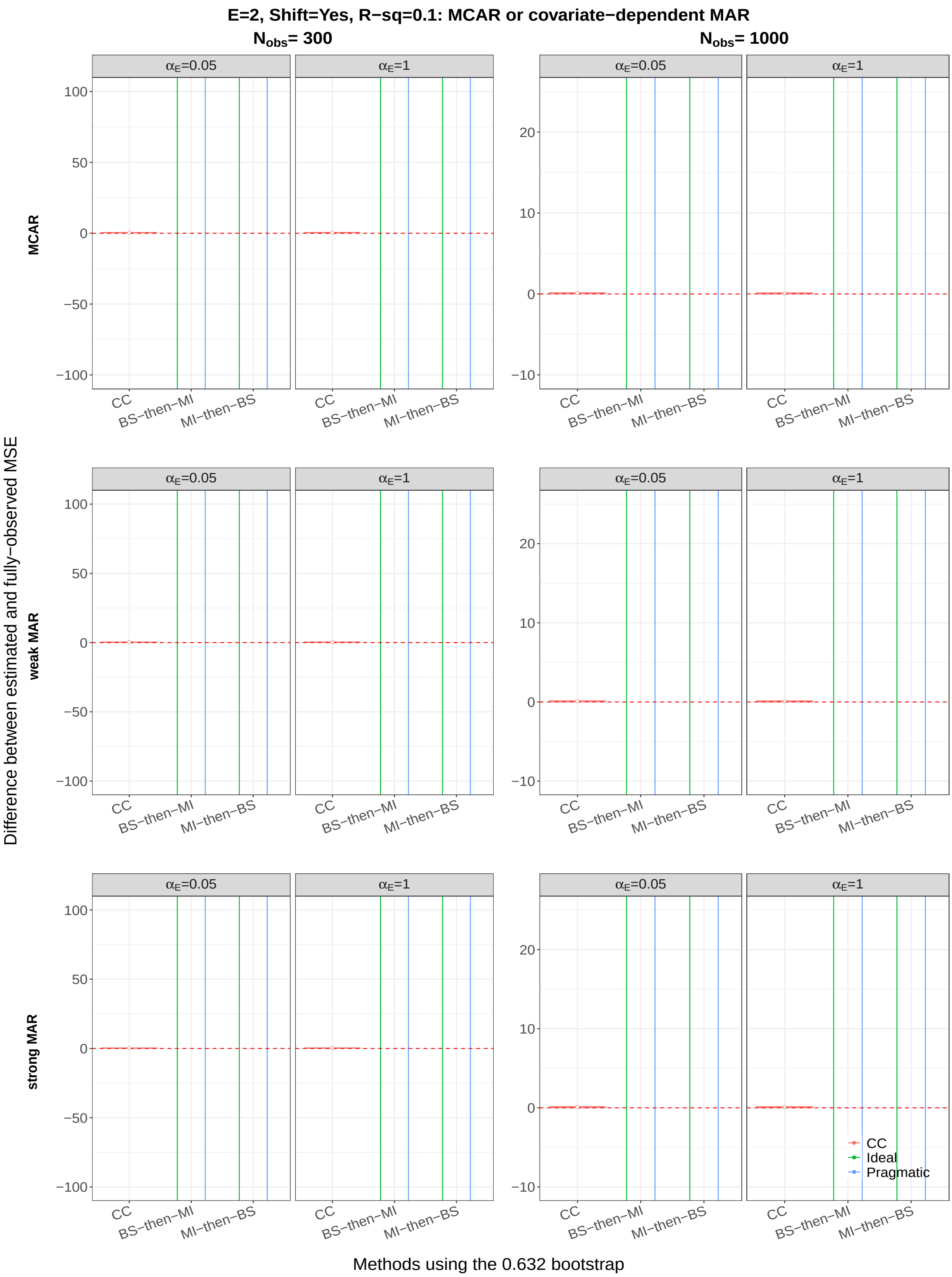

Figure S53: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

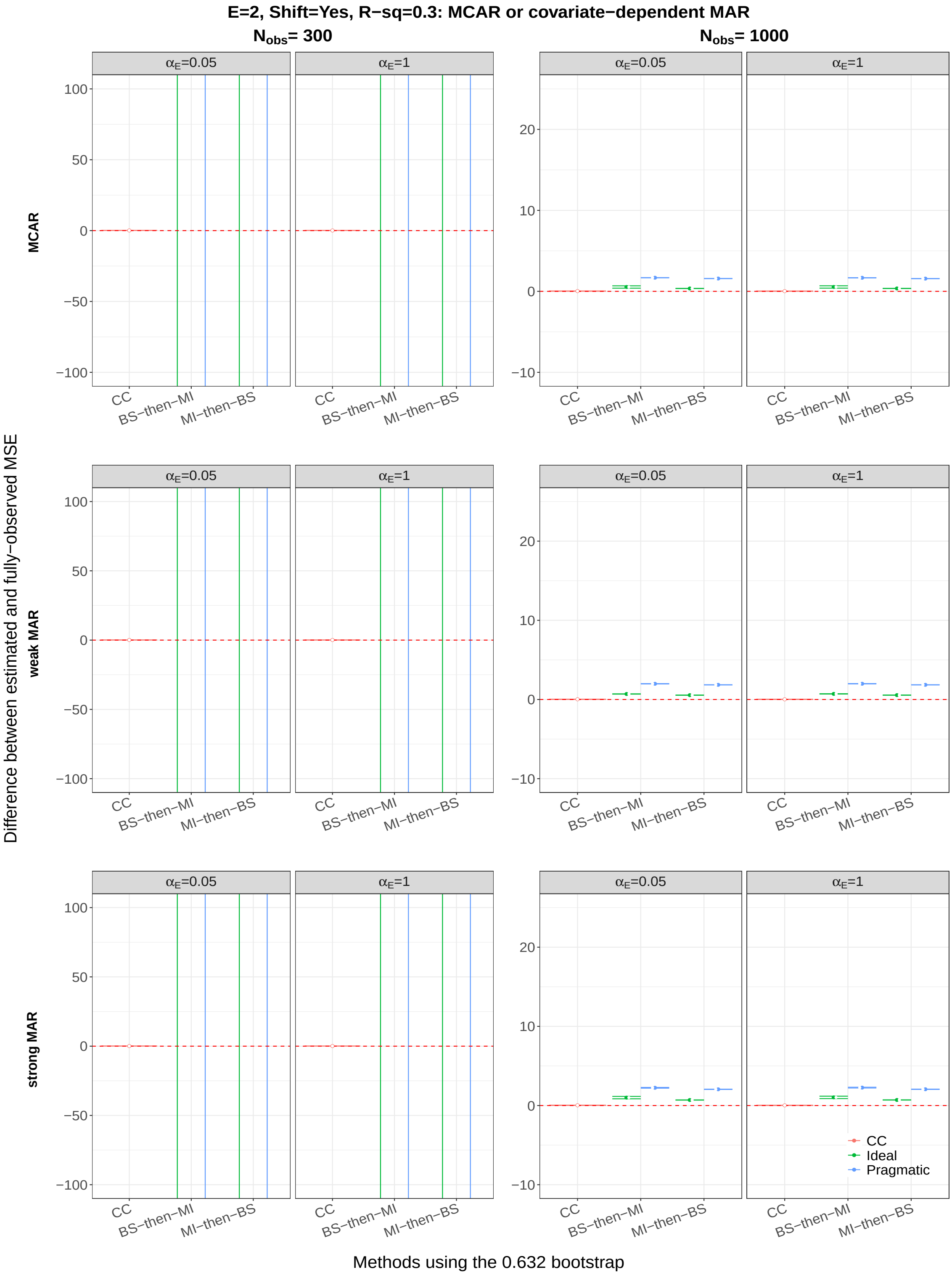

Figure S54: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

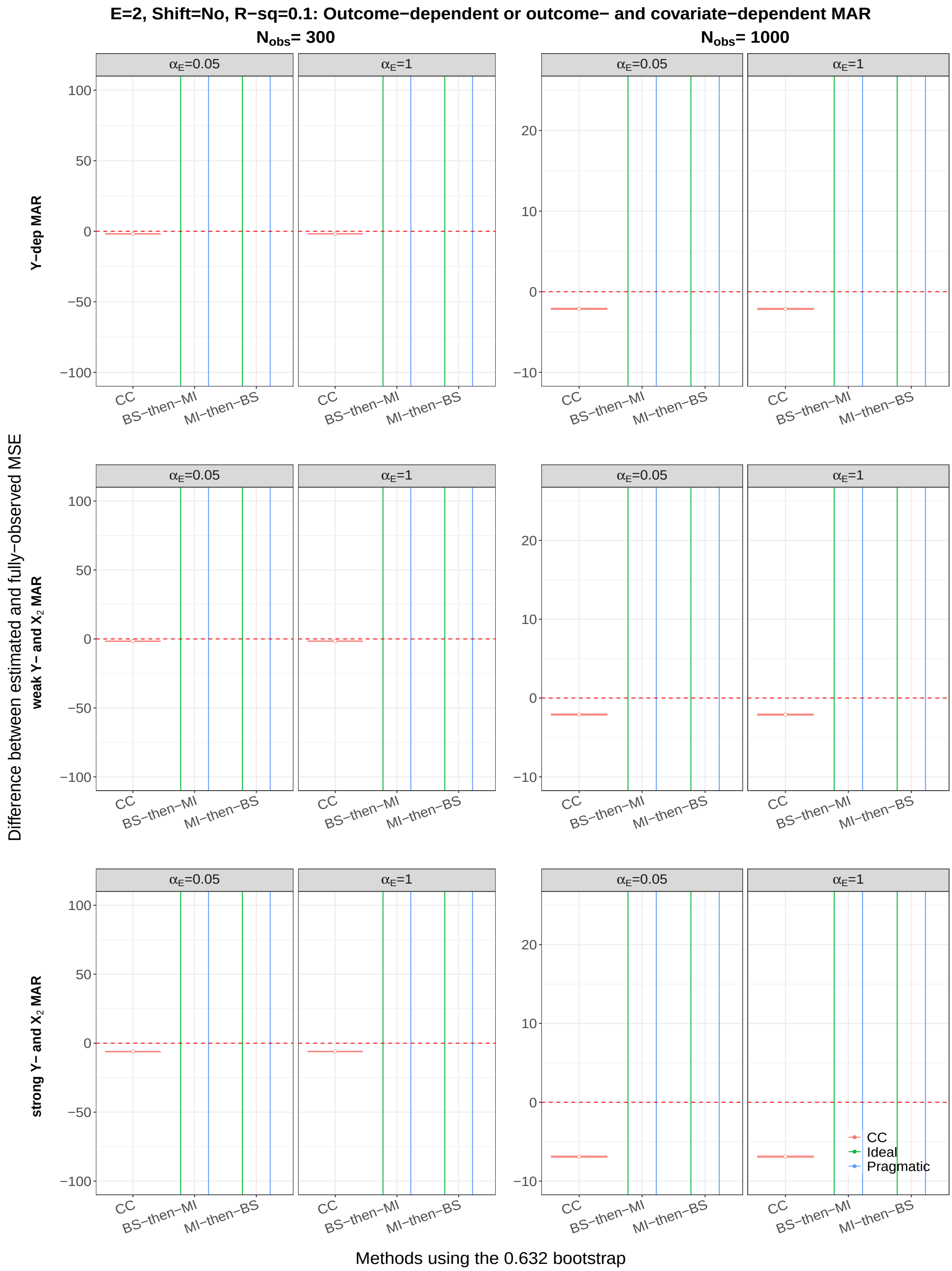

Figure S55: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

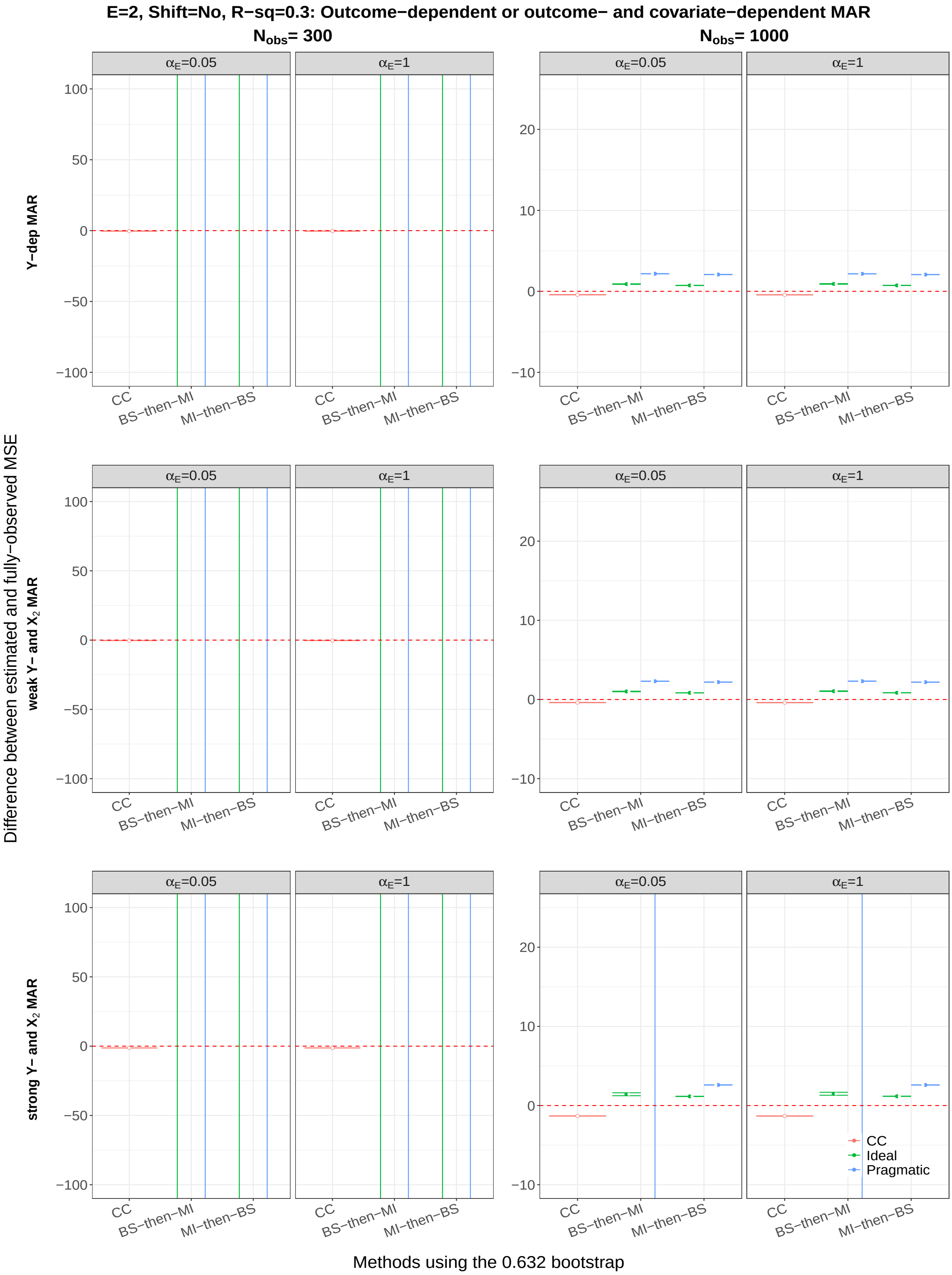

Figure S56: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

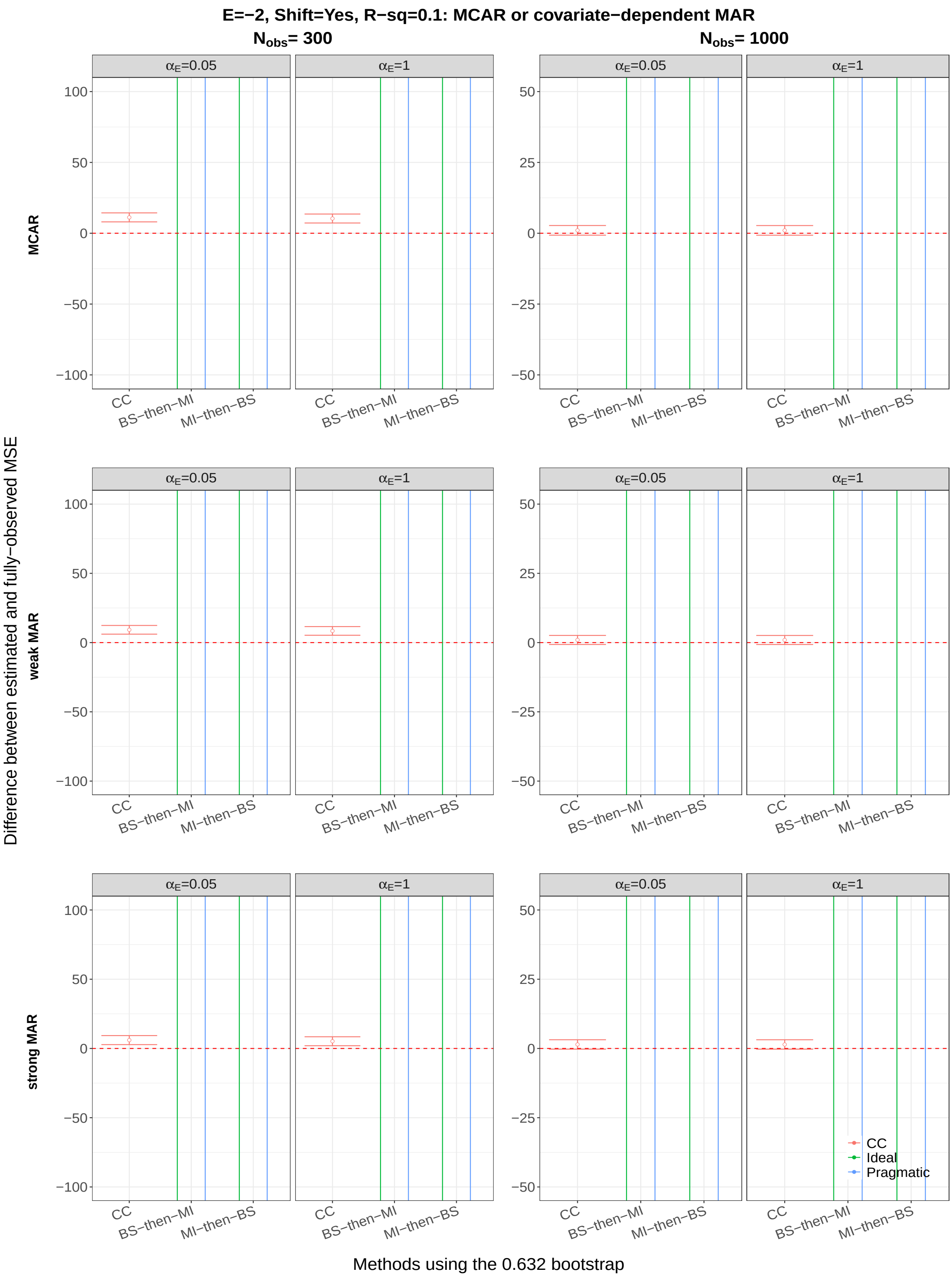

Figure S57: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

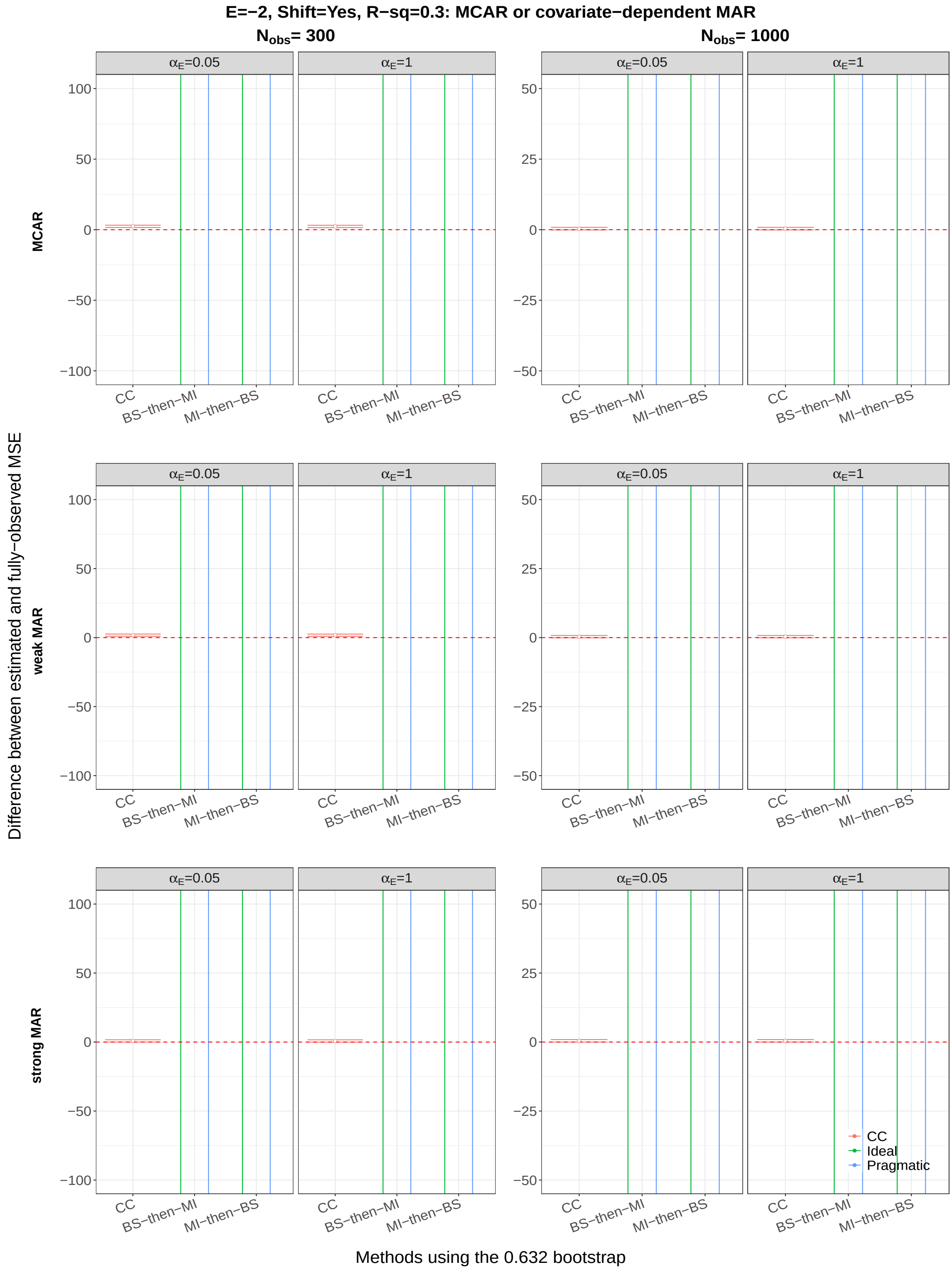

Figure S58: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

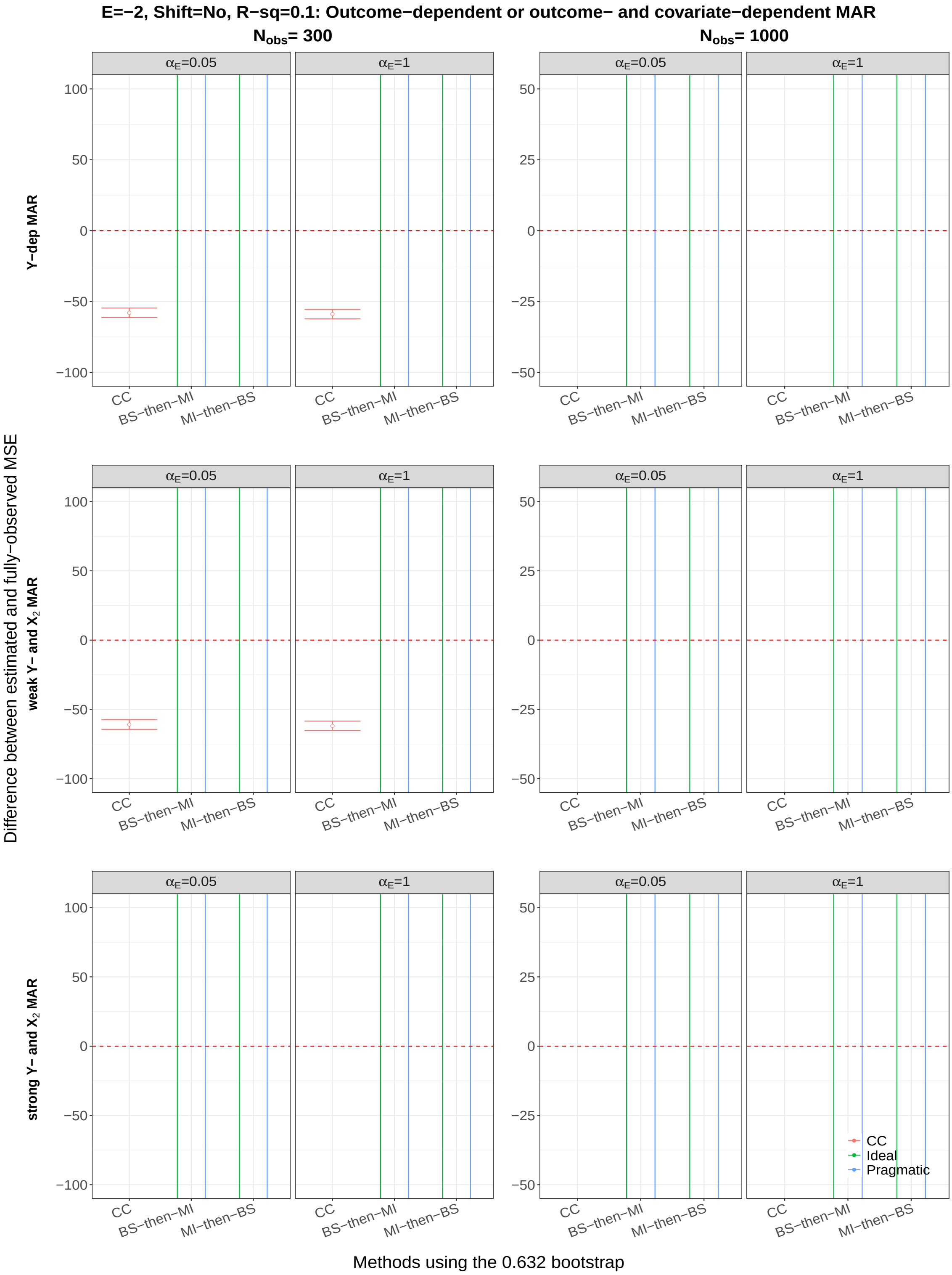

Figure S59: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

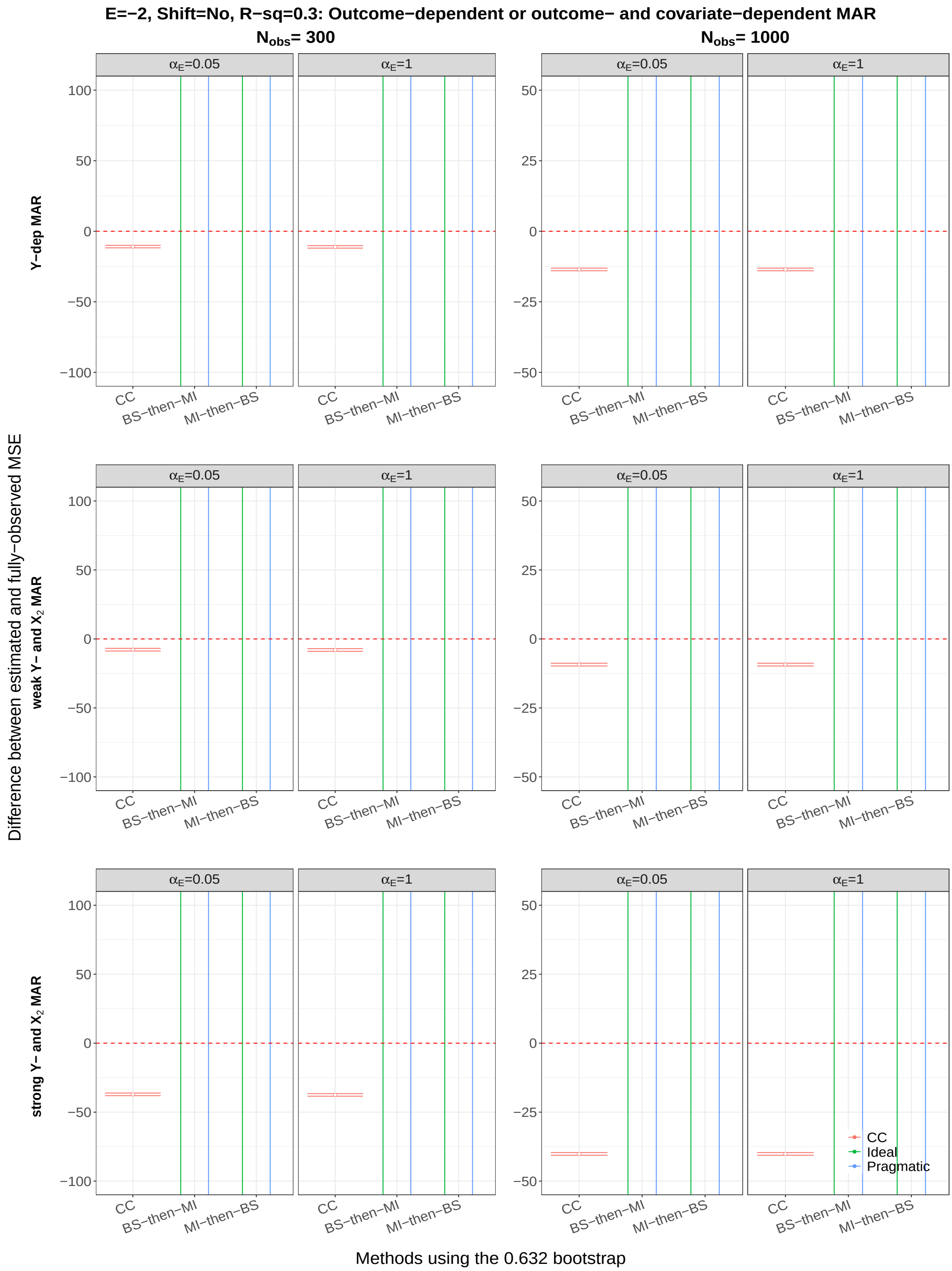

Figure S60: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7.2.2 $\beta_2 = 1$ and an origin shift transformation has been applied

True exponent is 0

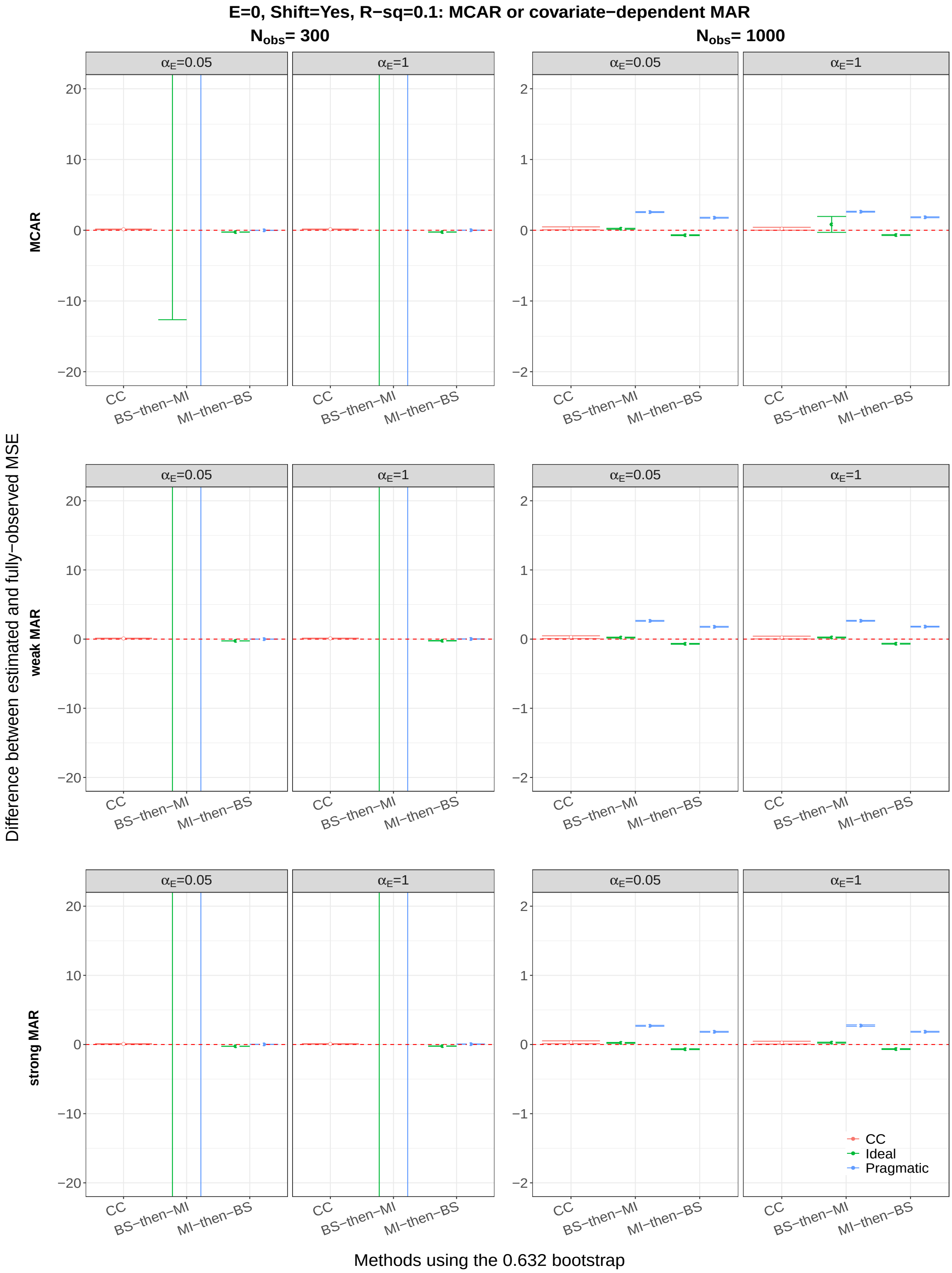

Figure S61: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

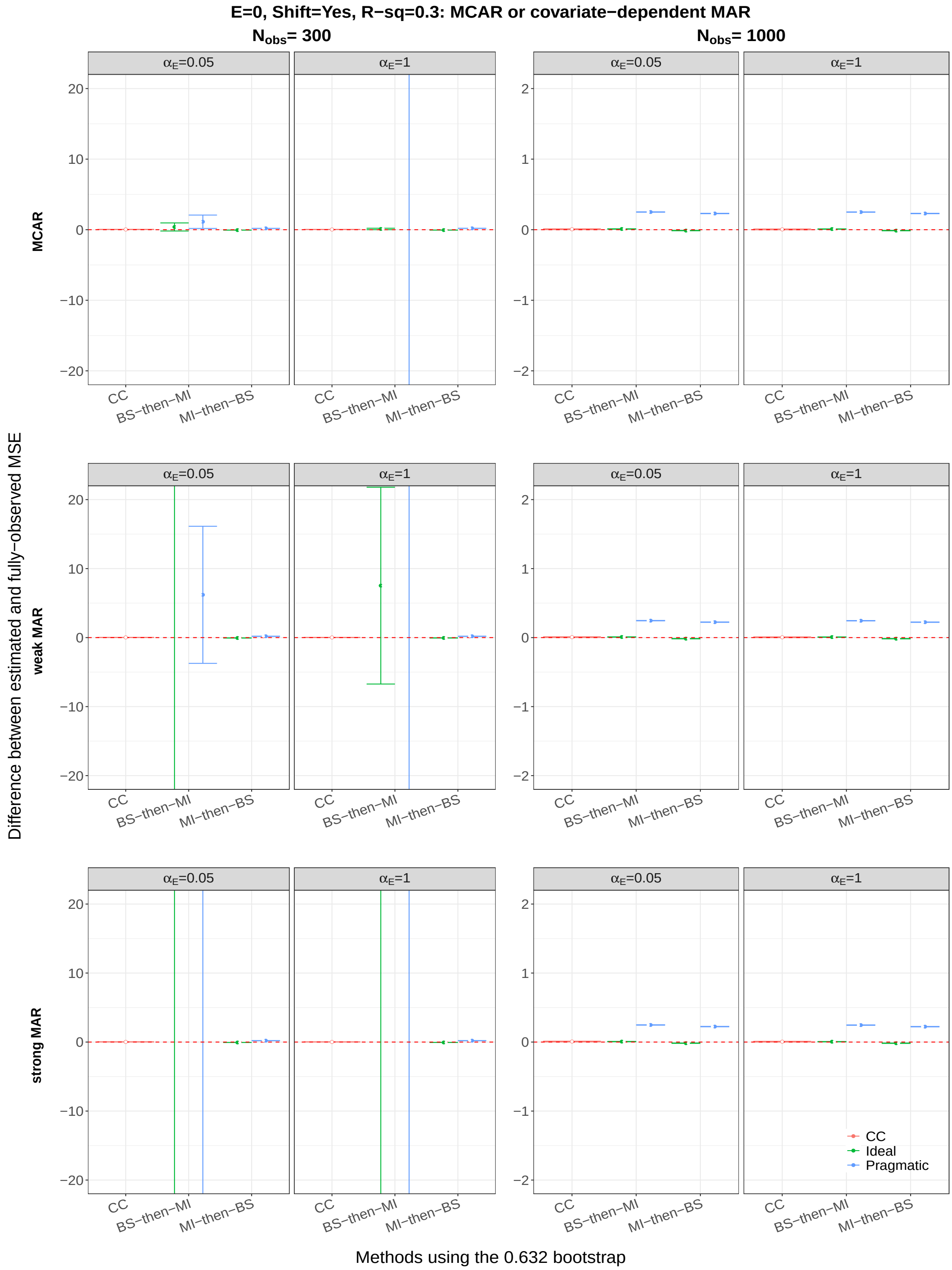

Figure S62: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

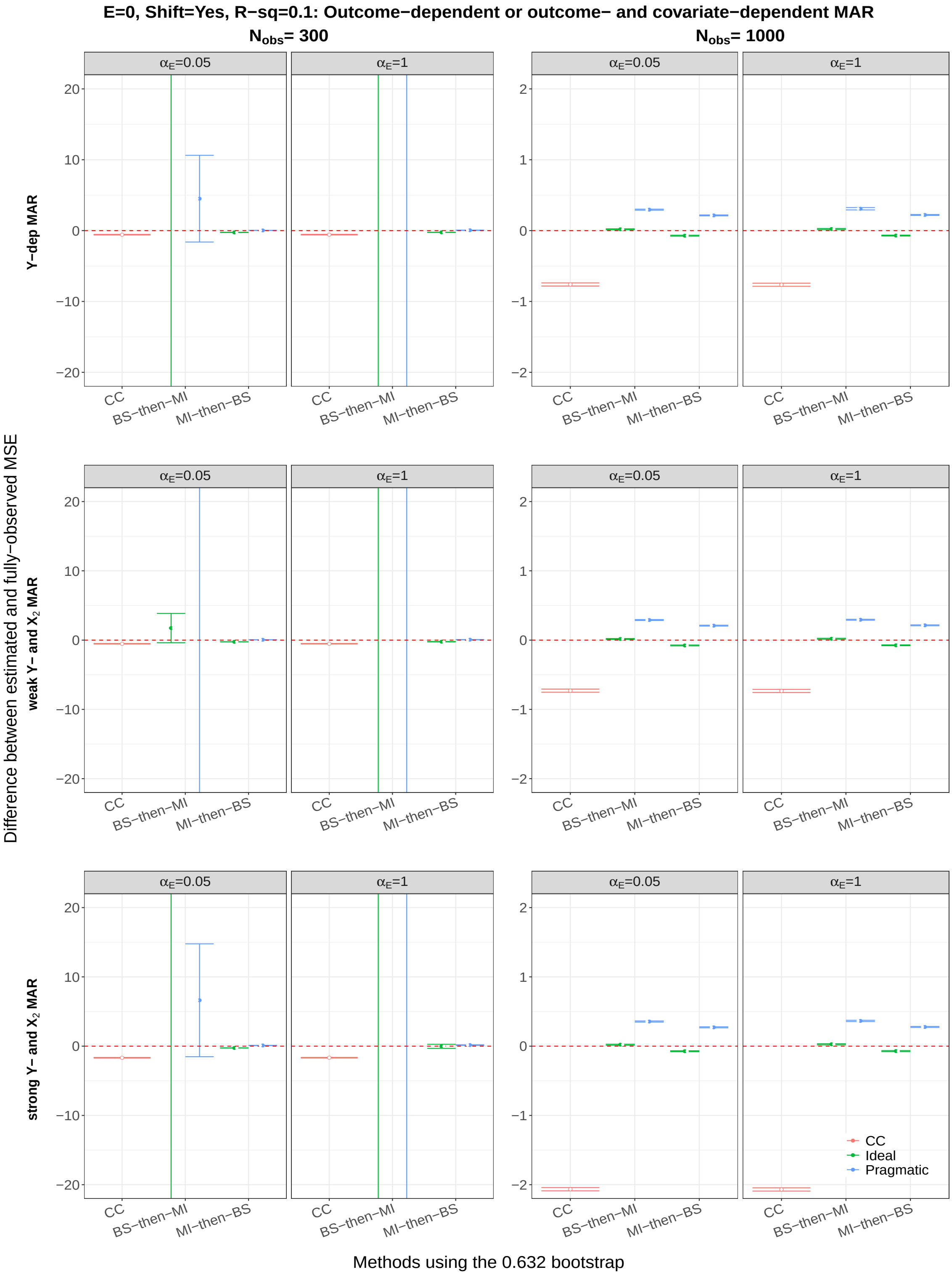

Figure S63: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

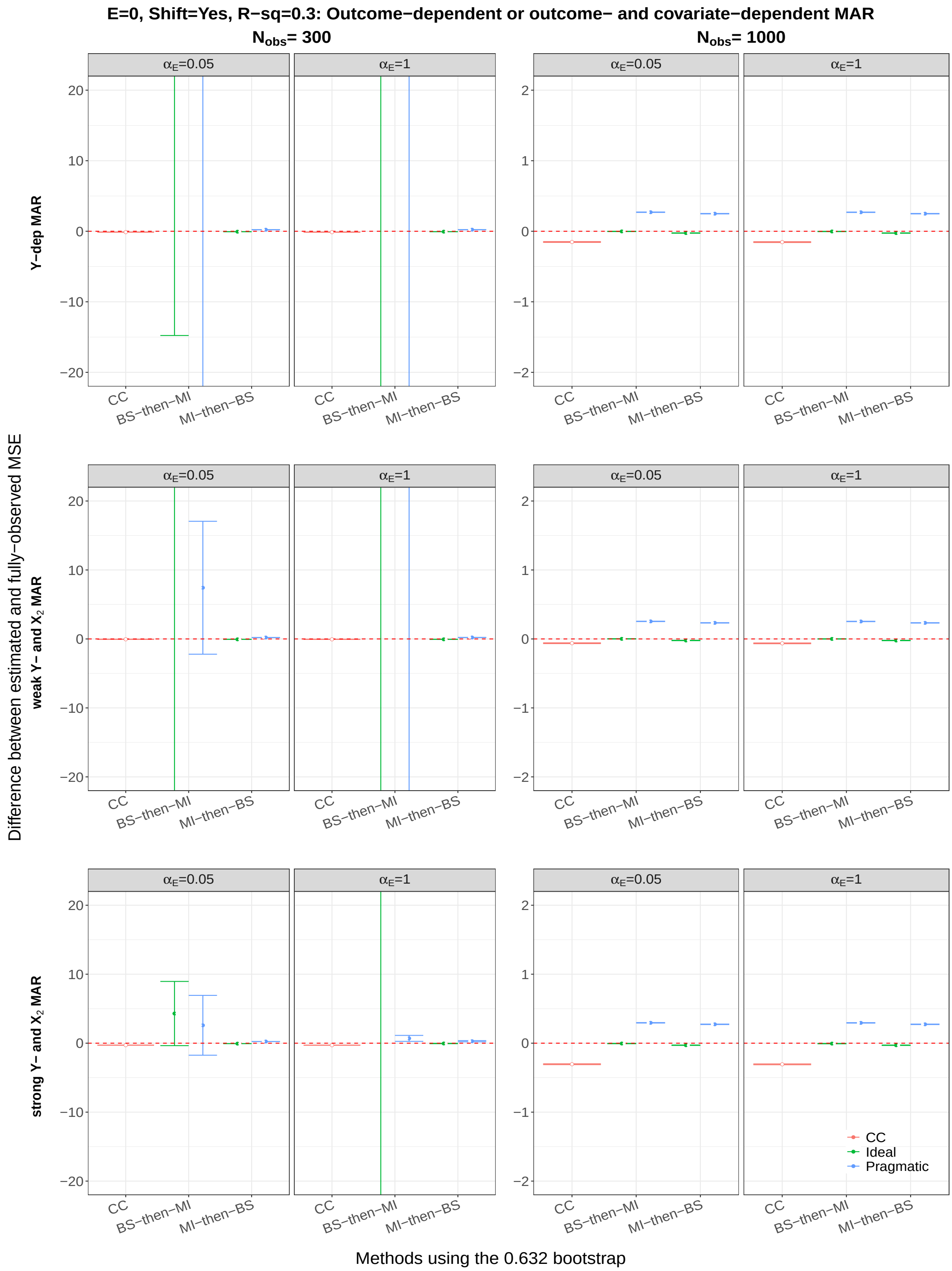

Figure S64: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

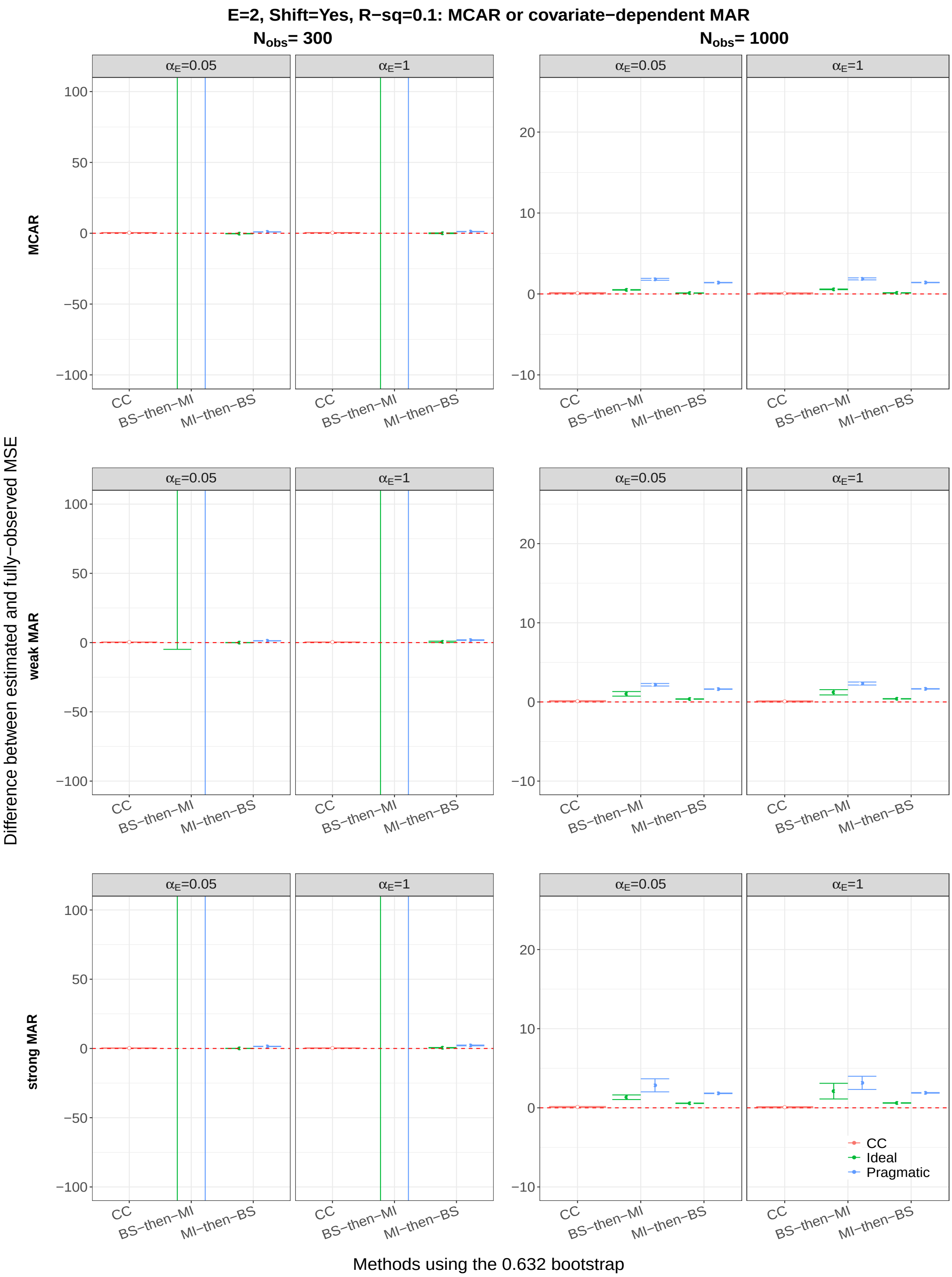

Figure S65: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

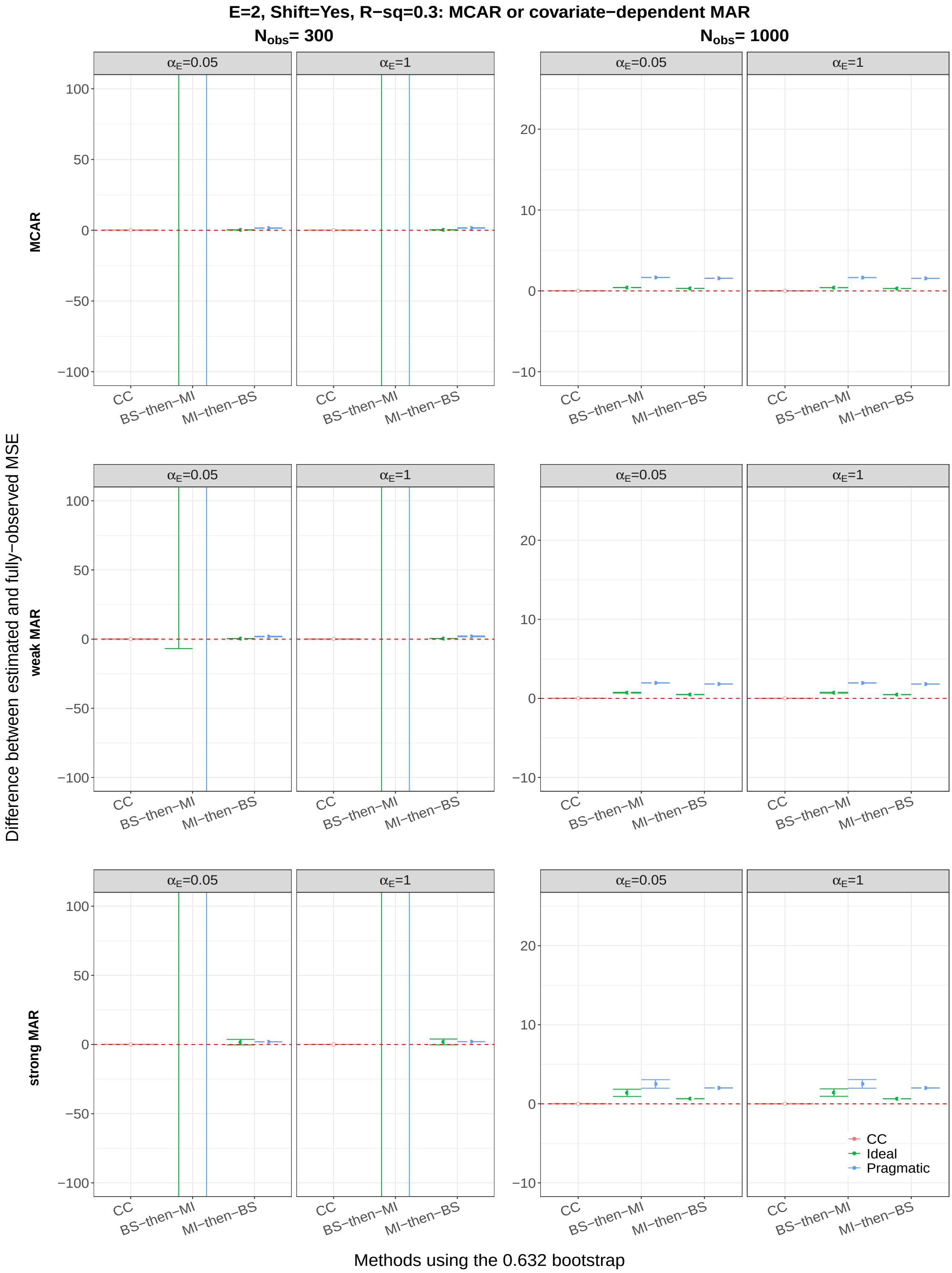

Figure S66: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

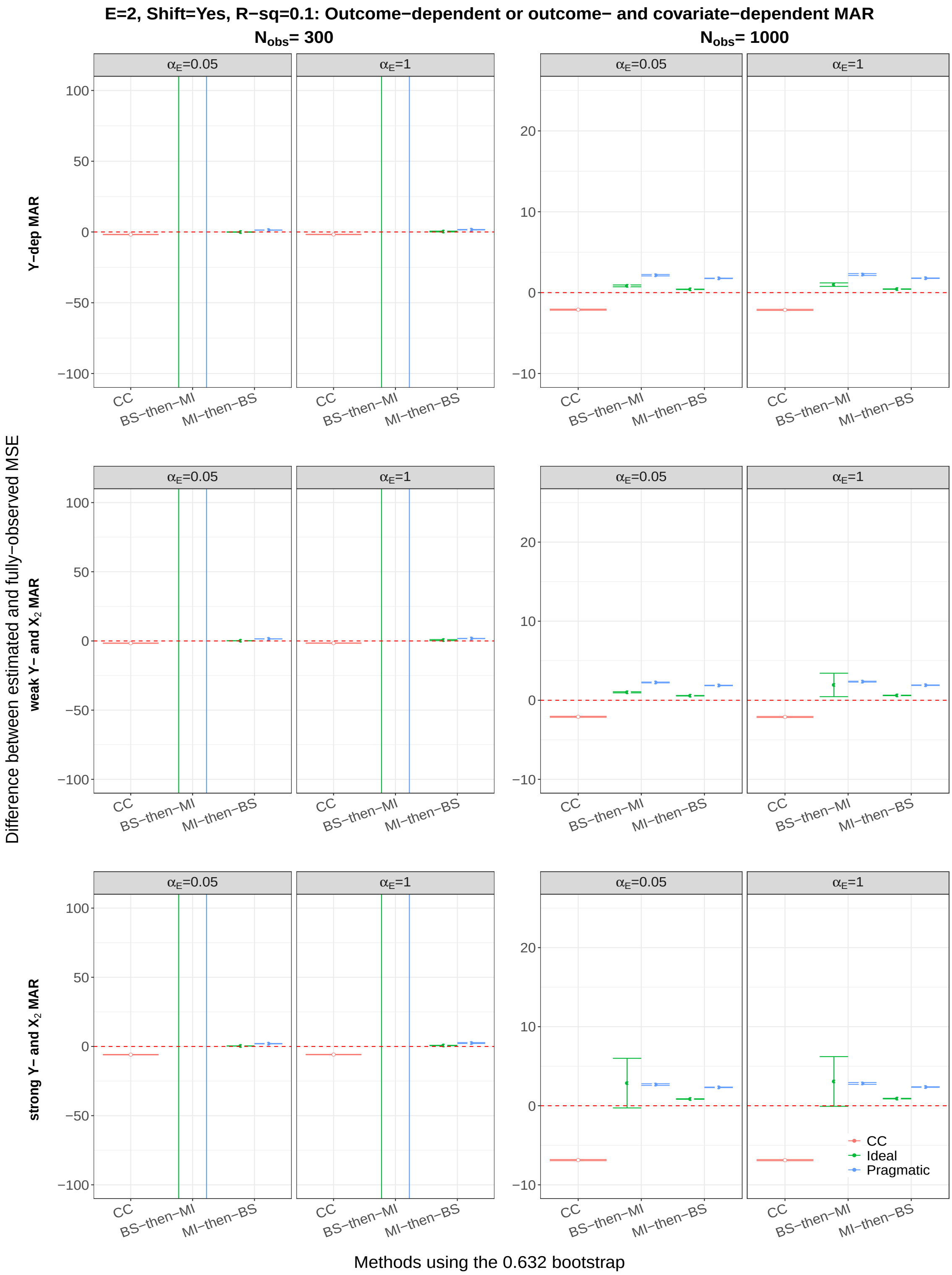

Figure S67: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

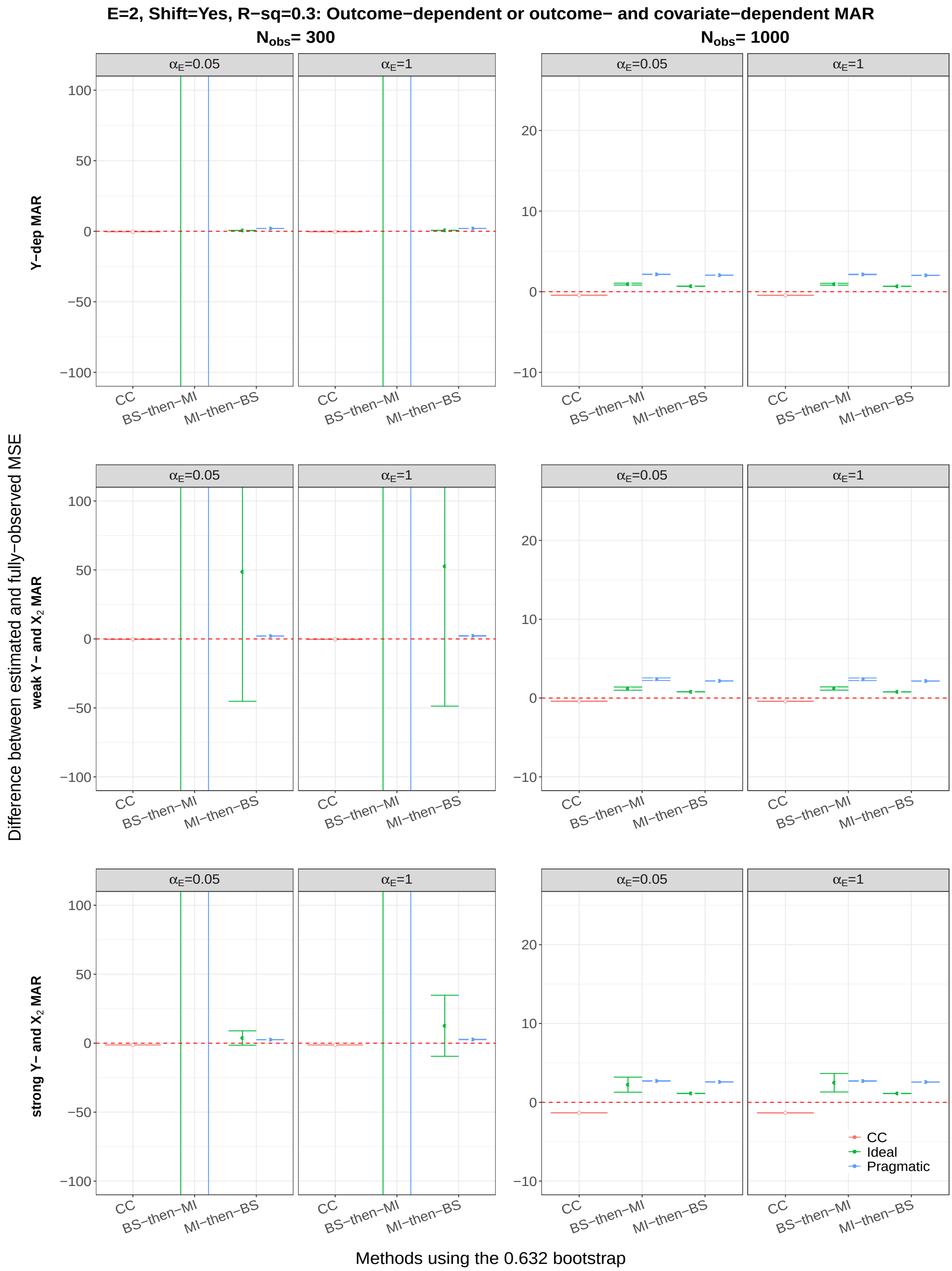

Figure S68: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

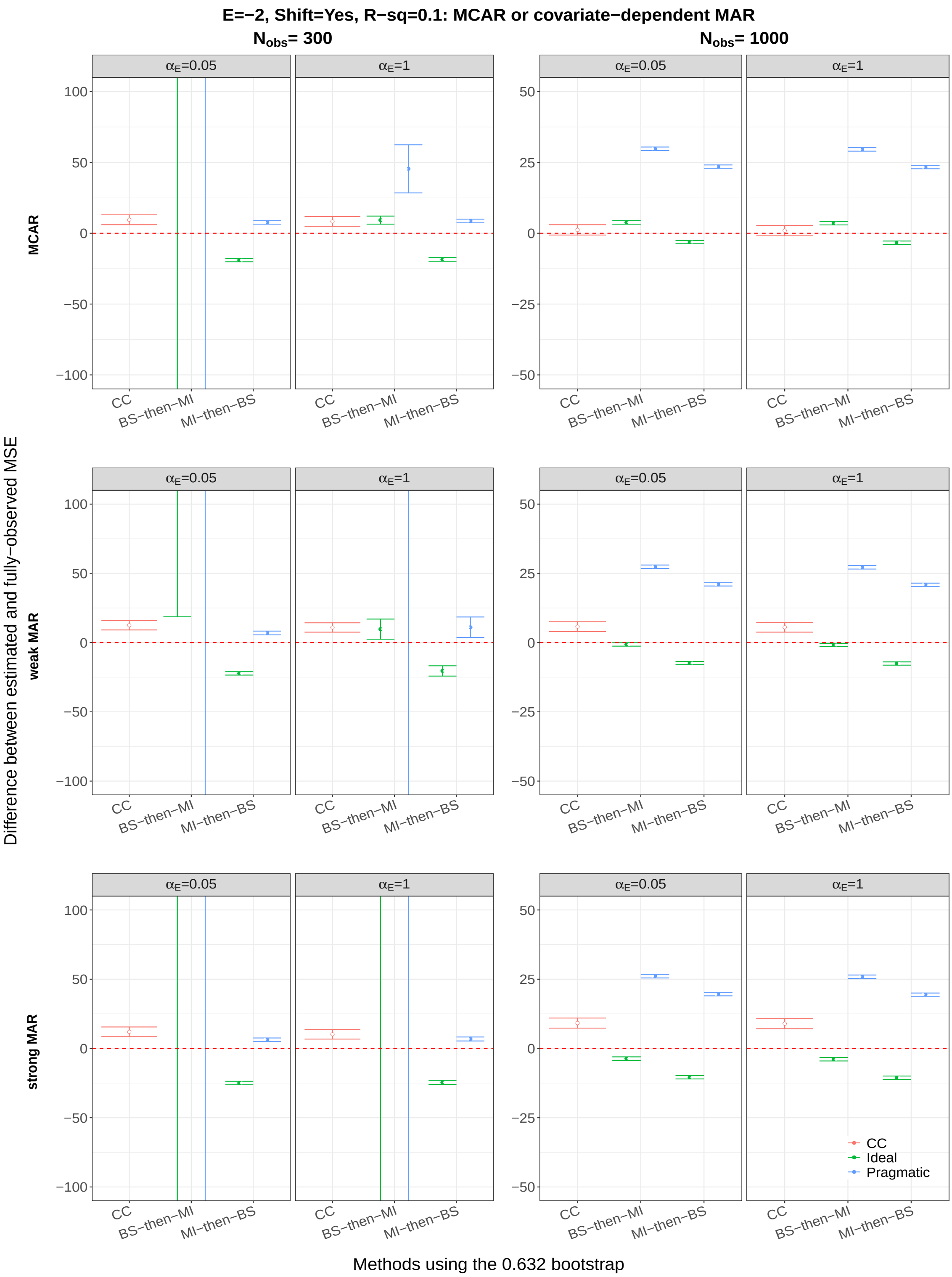

Figure S69: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

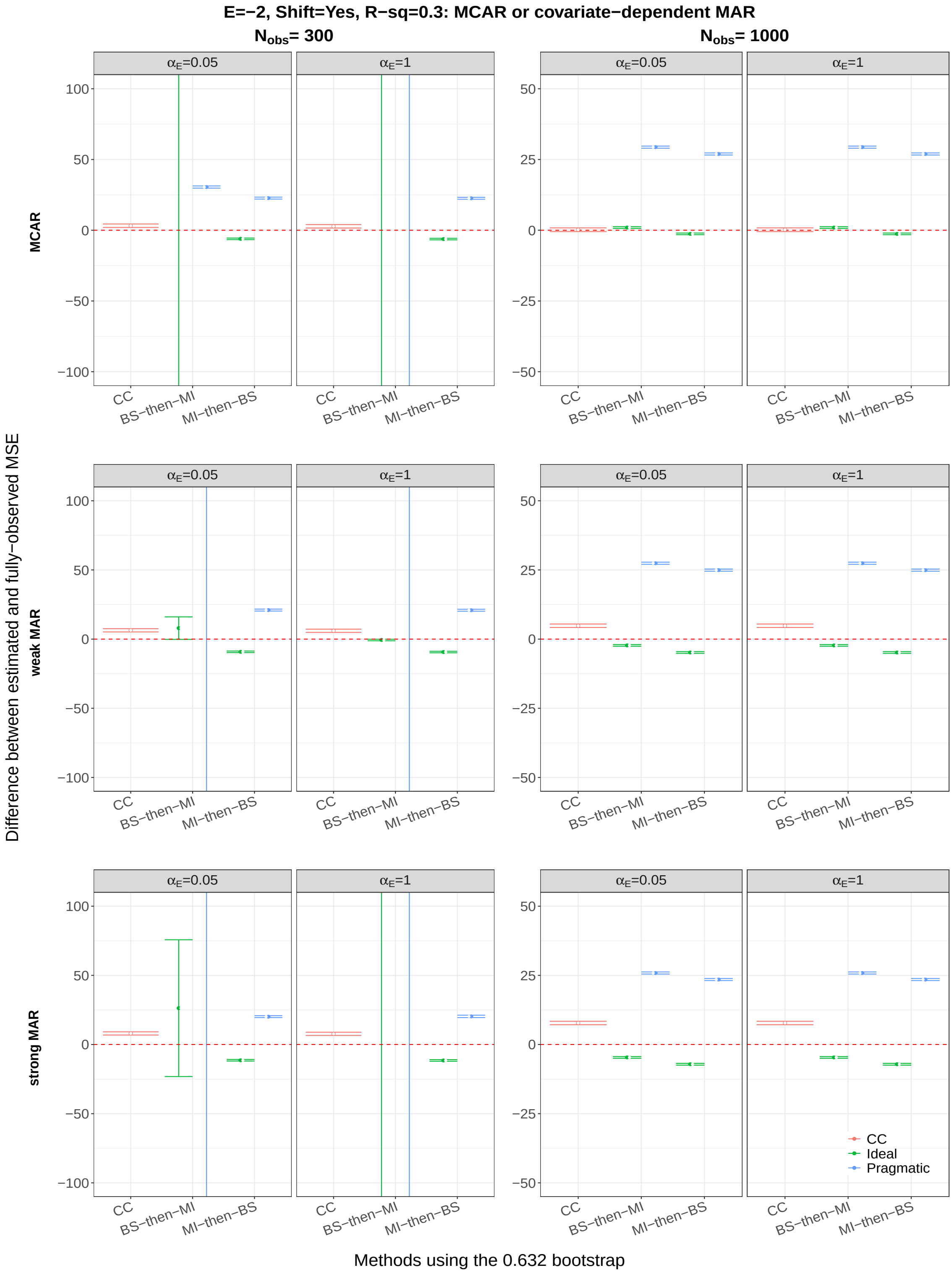

Figure S70: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

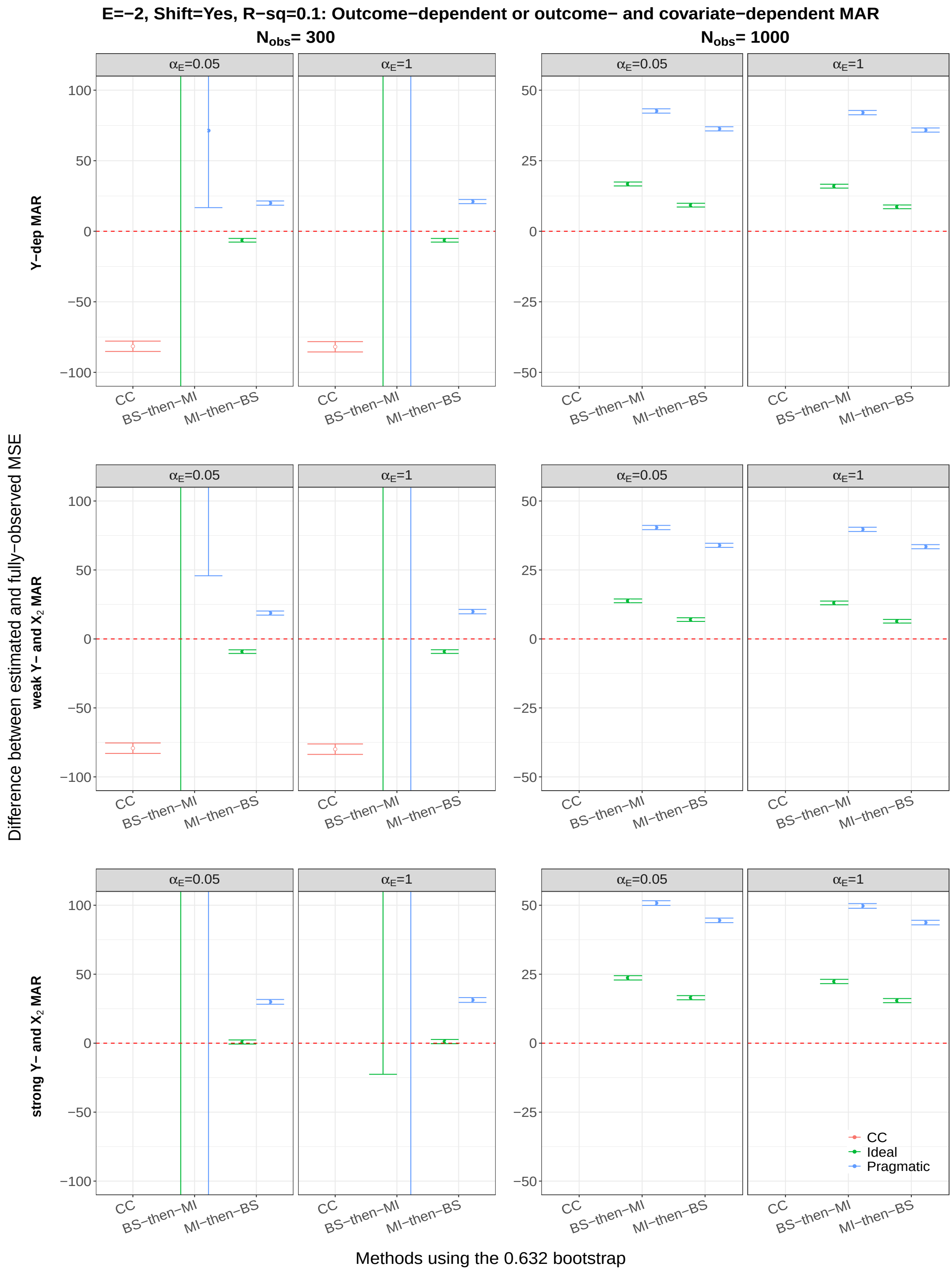

Figure S71: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

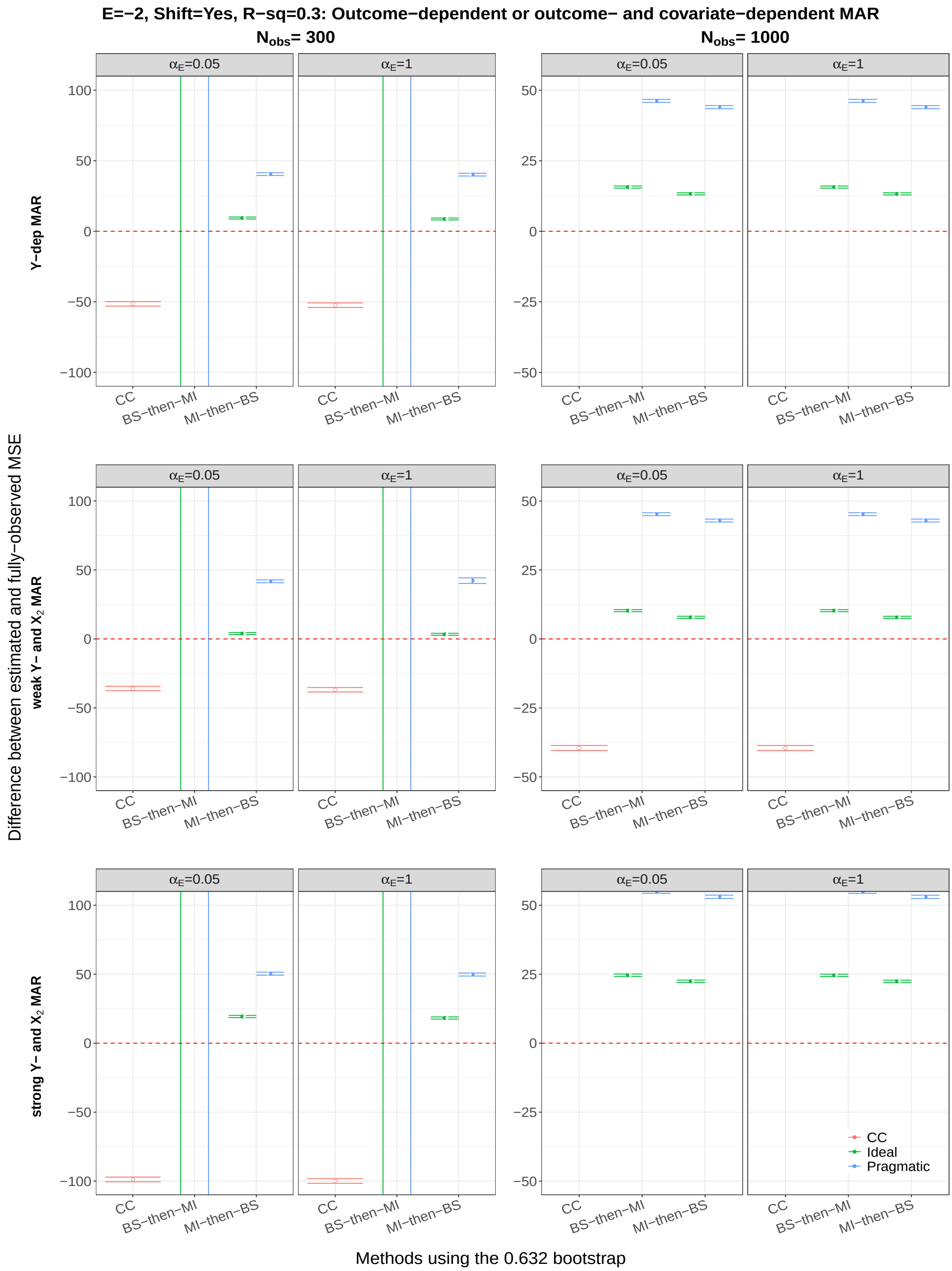

Figure S72: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7.2.3 $\beta_2 = 0$ and an origin shift transformation has not been applied

True exponent is 0

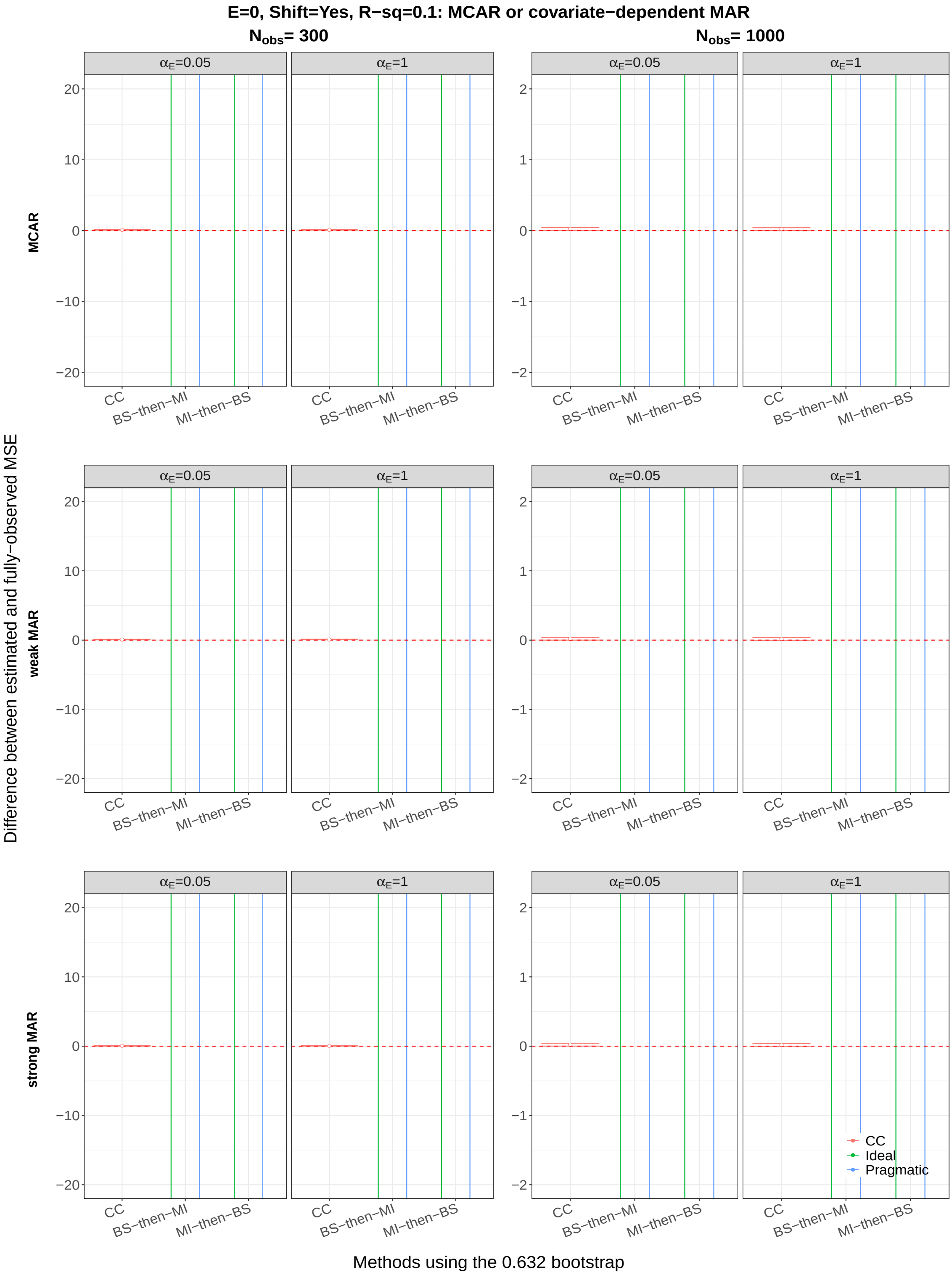

Figure S73: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

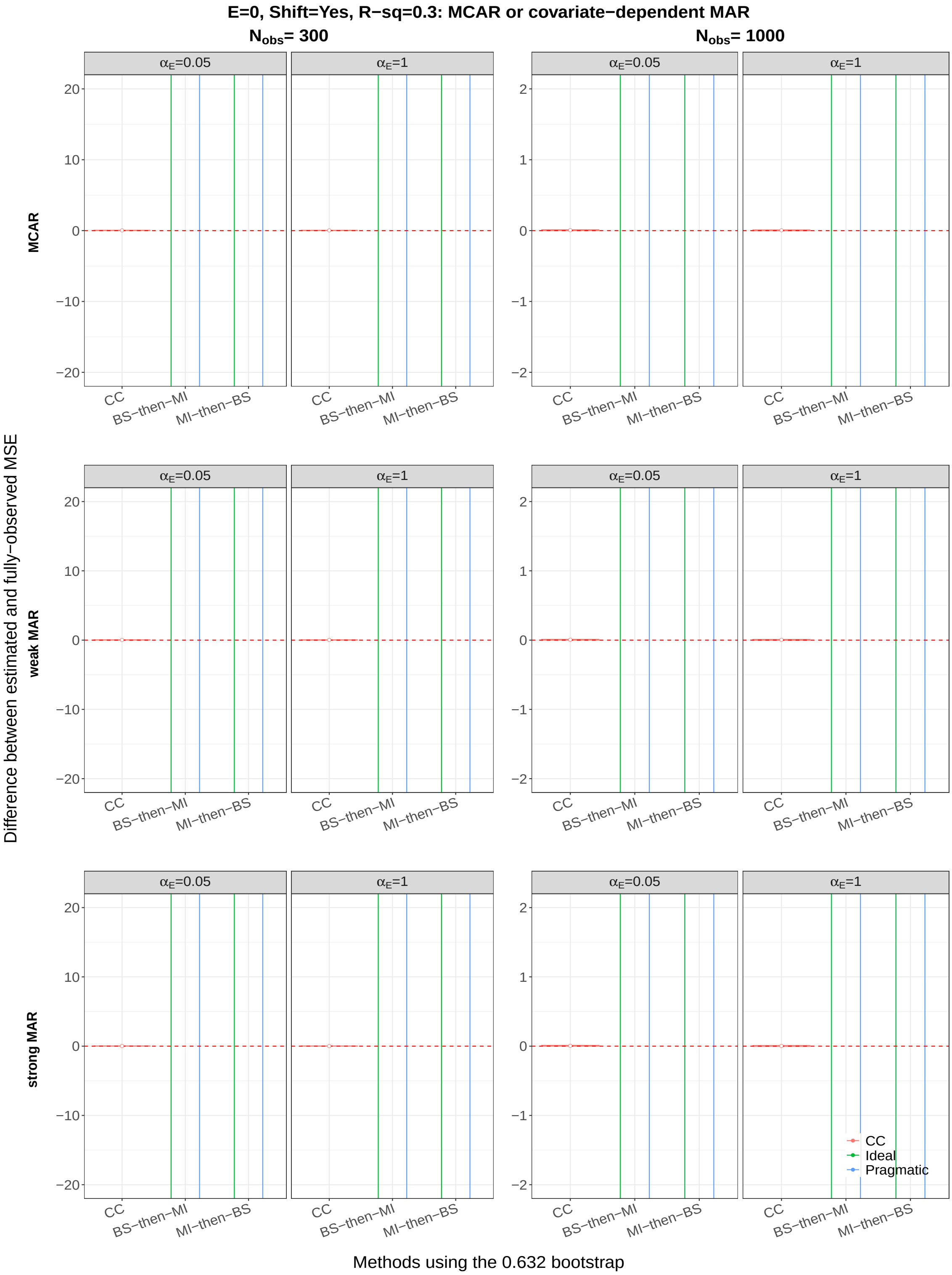

Figure S74: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

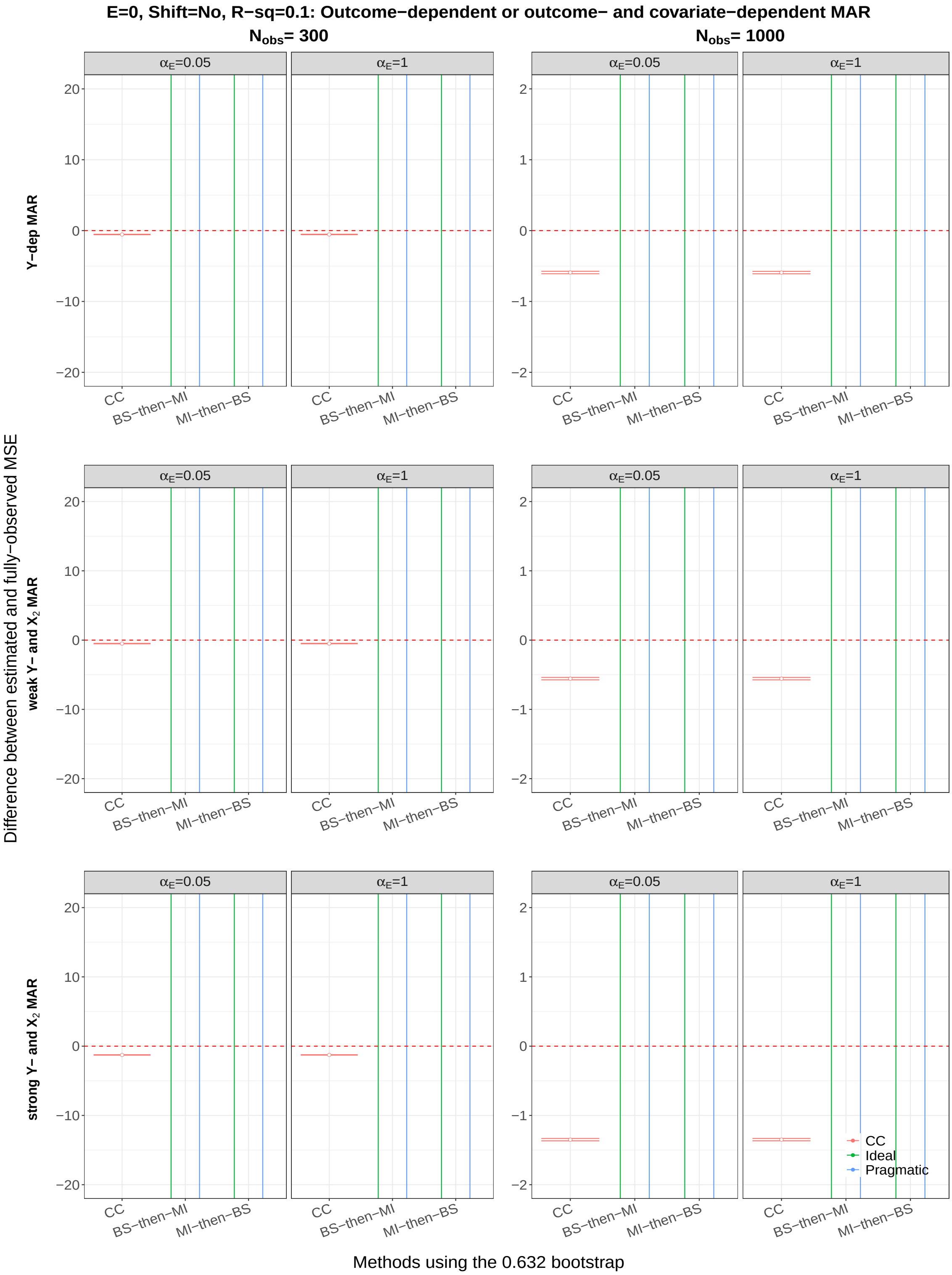

Figure S75: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

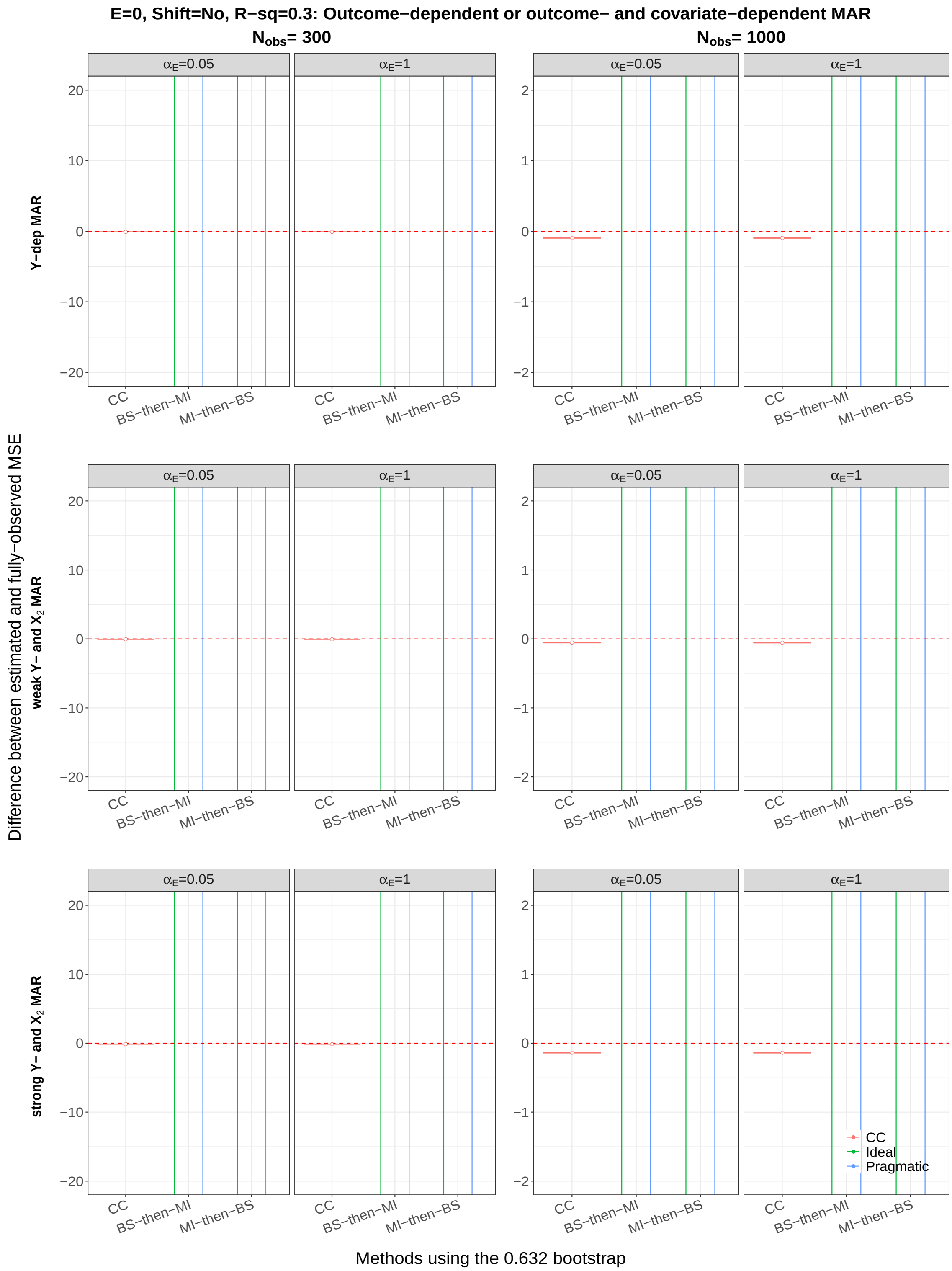

Figure S76: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

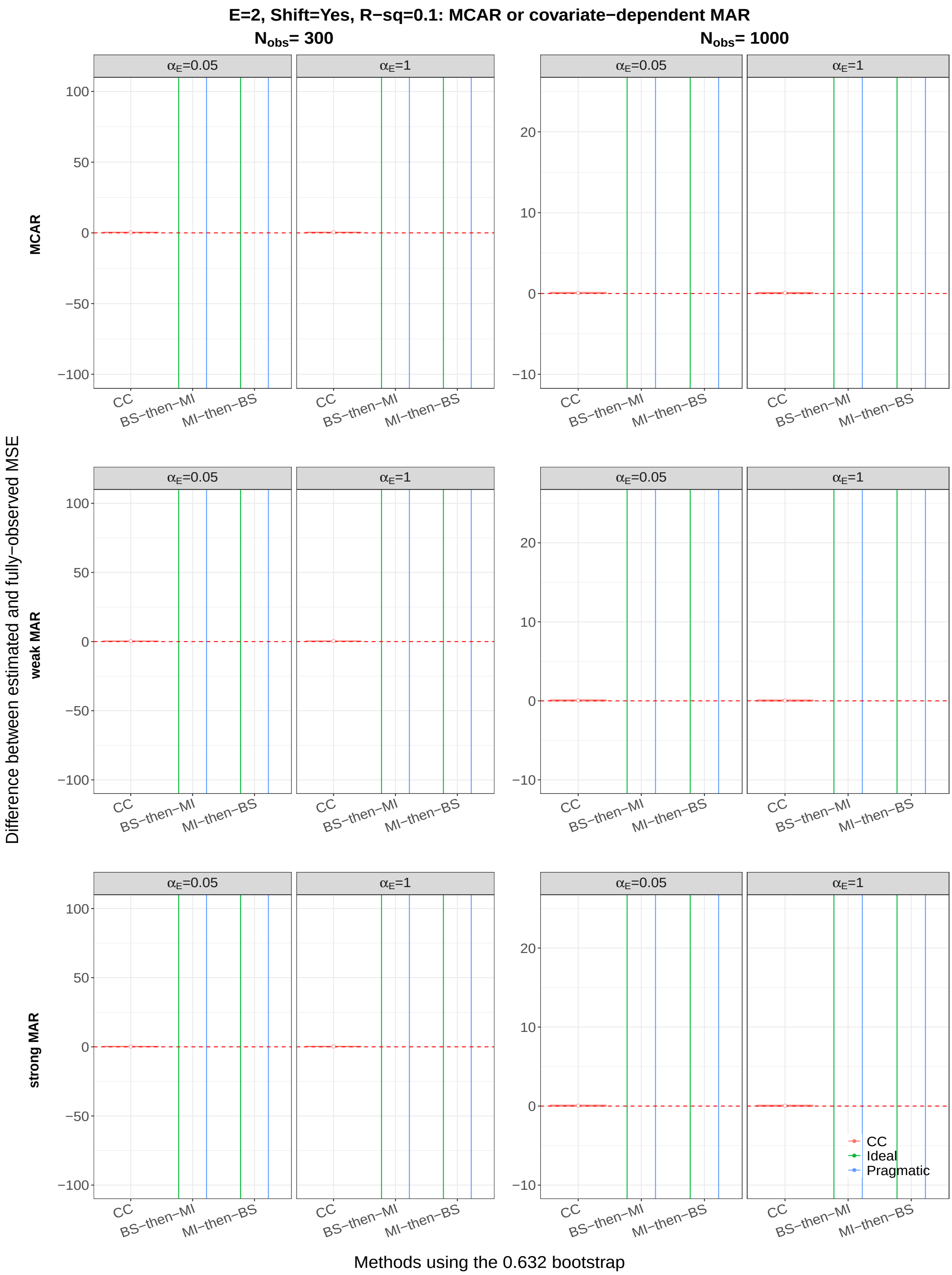

Figure S77: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

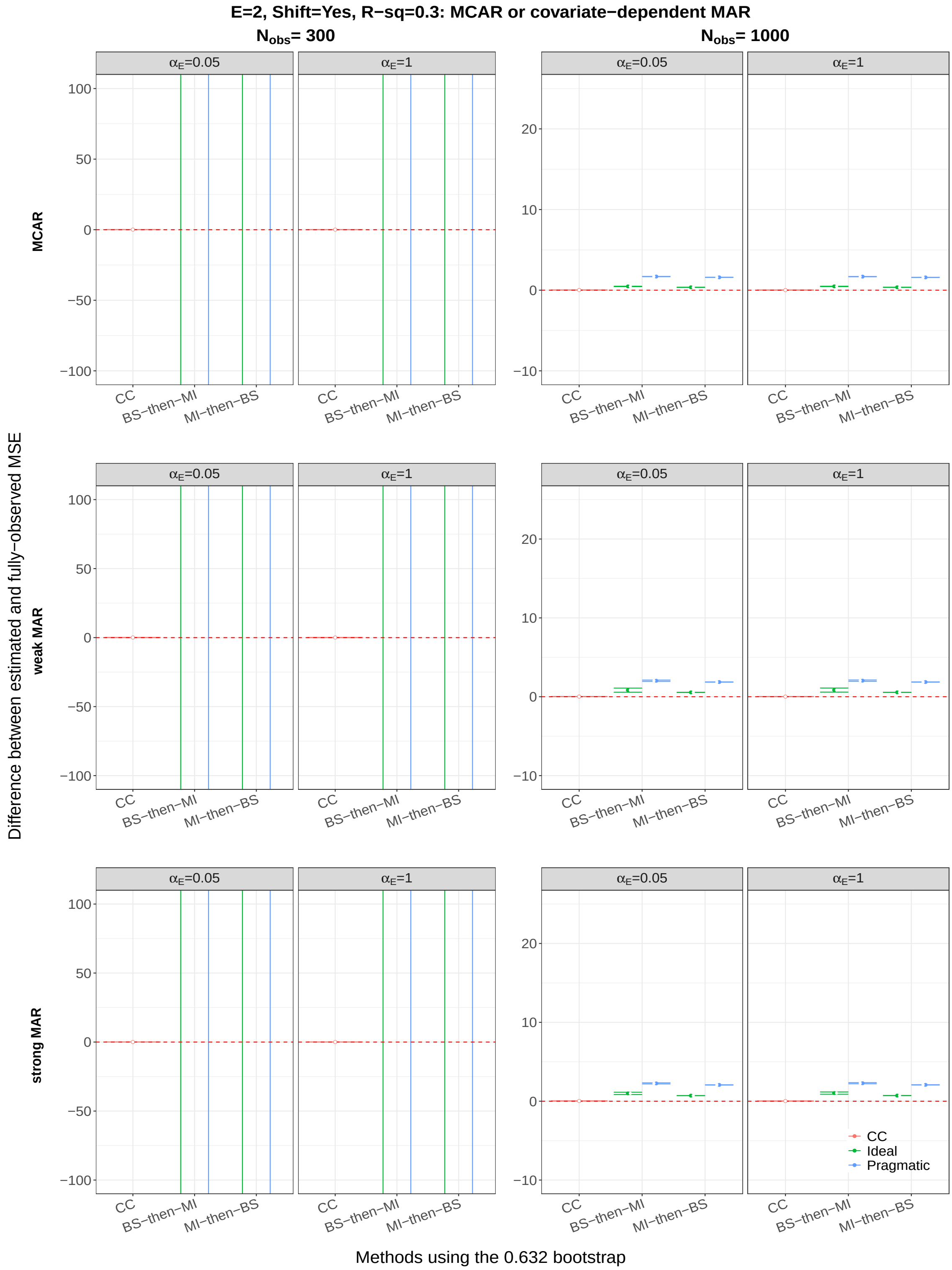

Figure S78: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

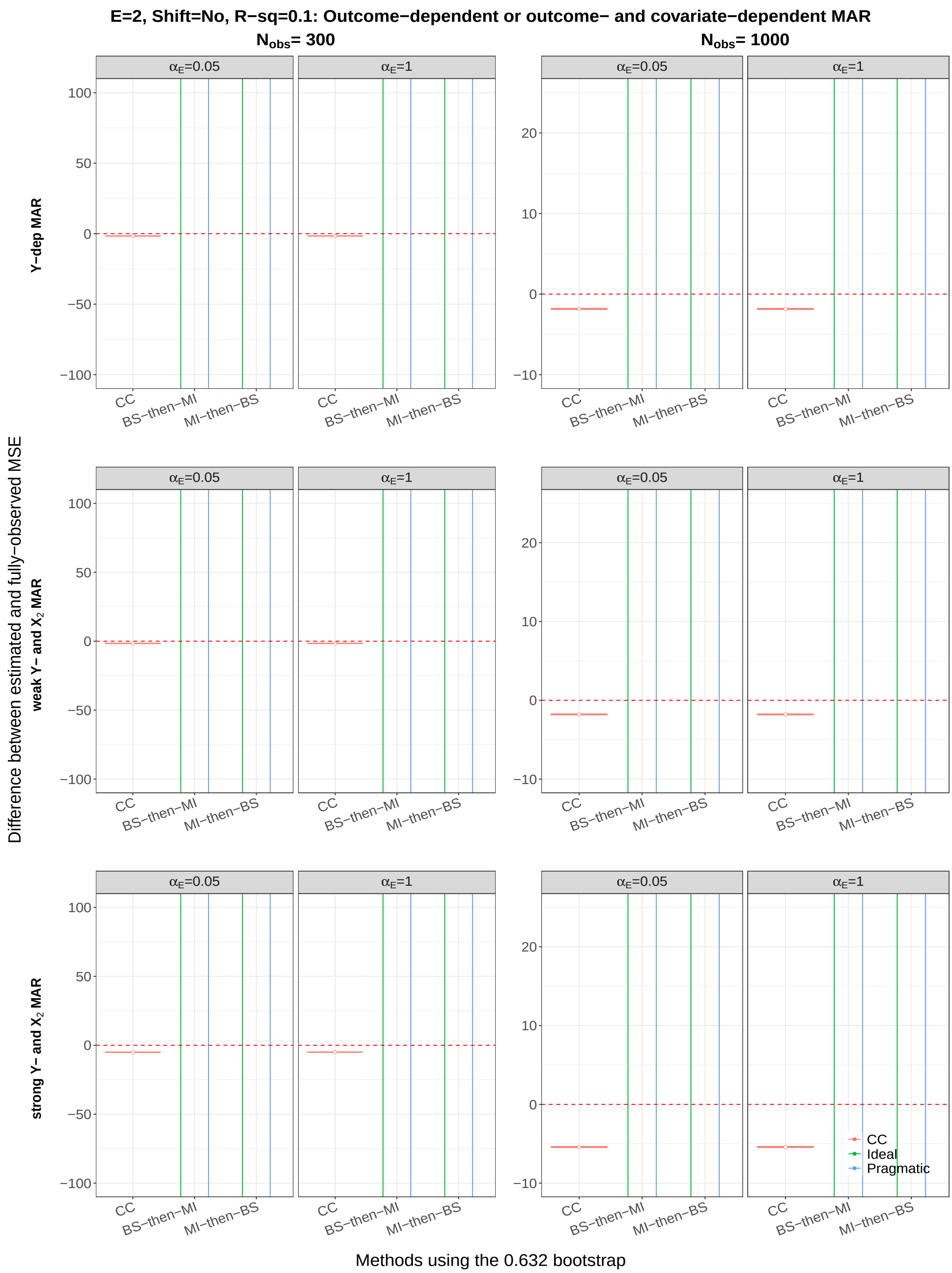

Figure S79: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

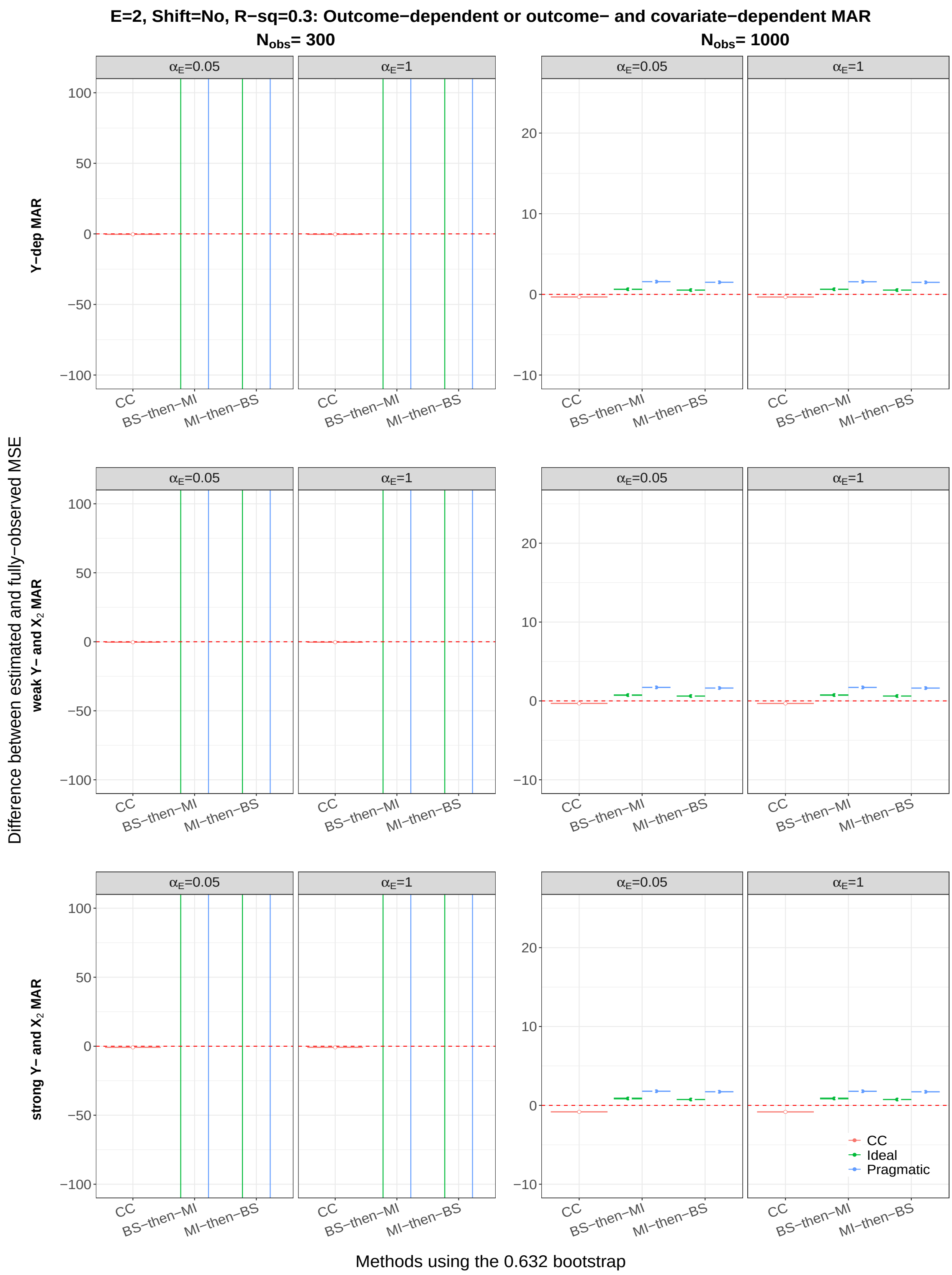

Figure S80: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

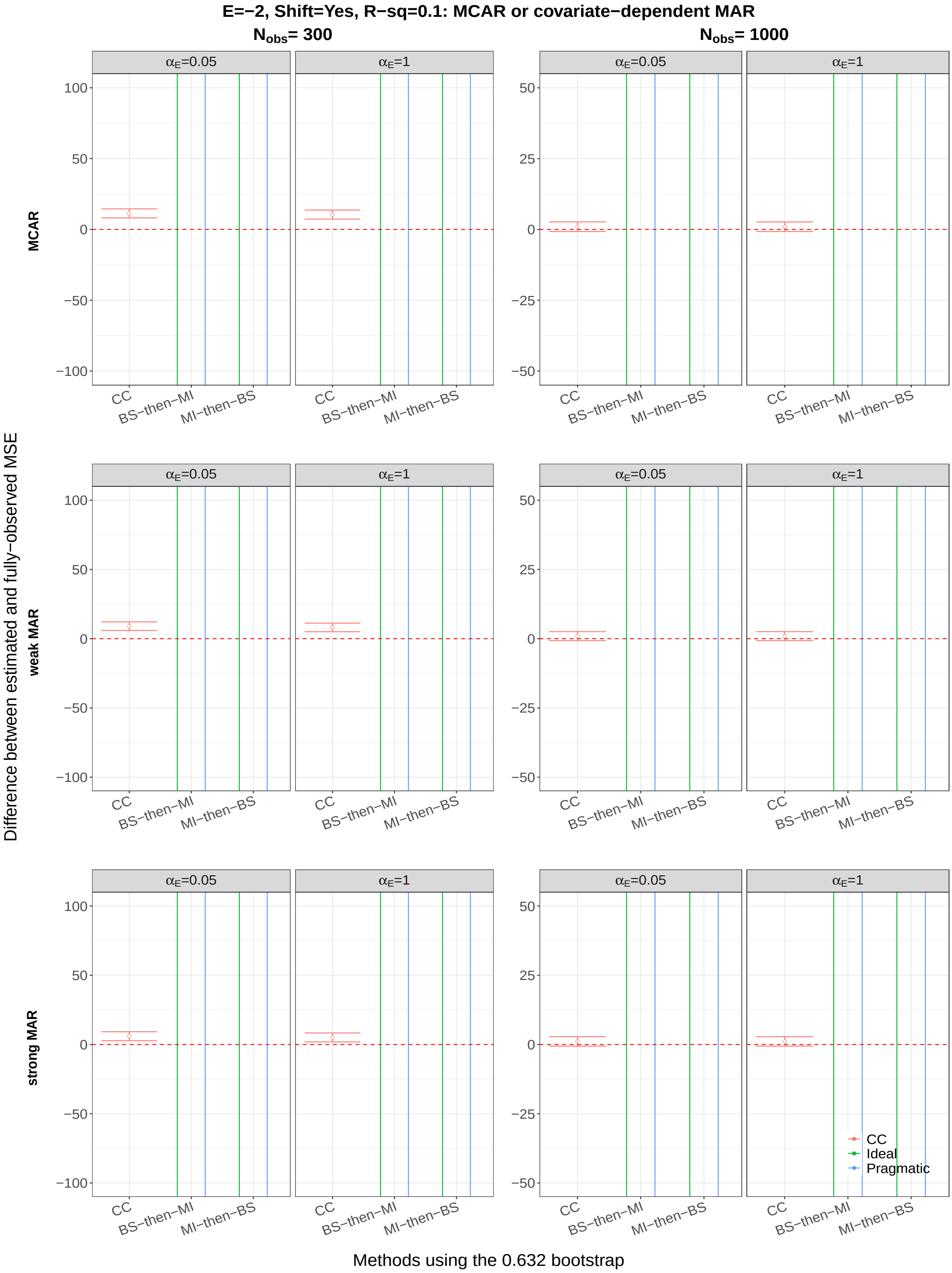

Figure S81: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

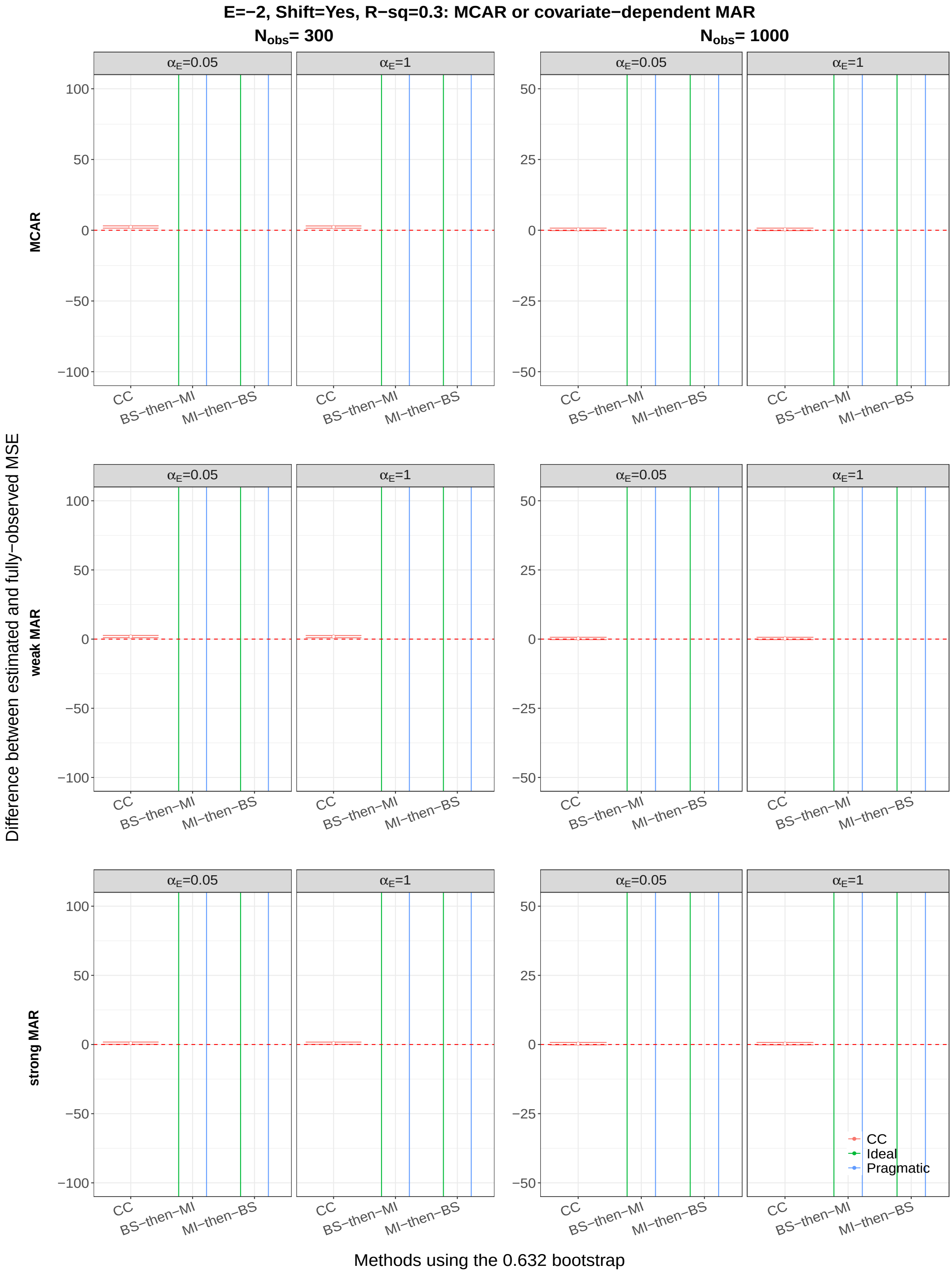

Figure S82: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

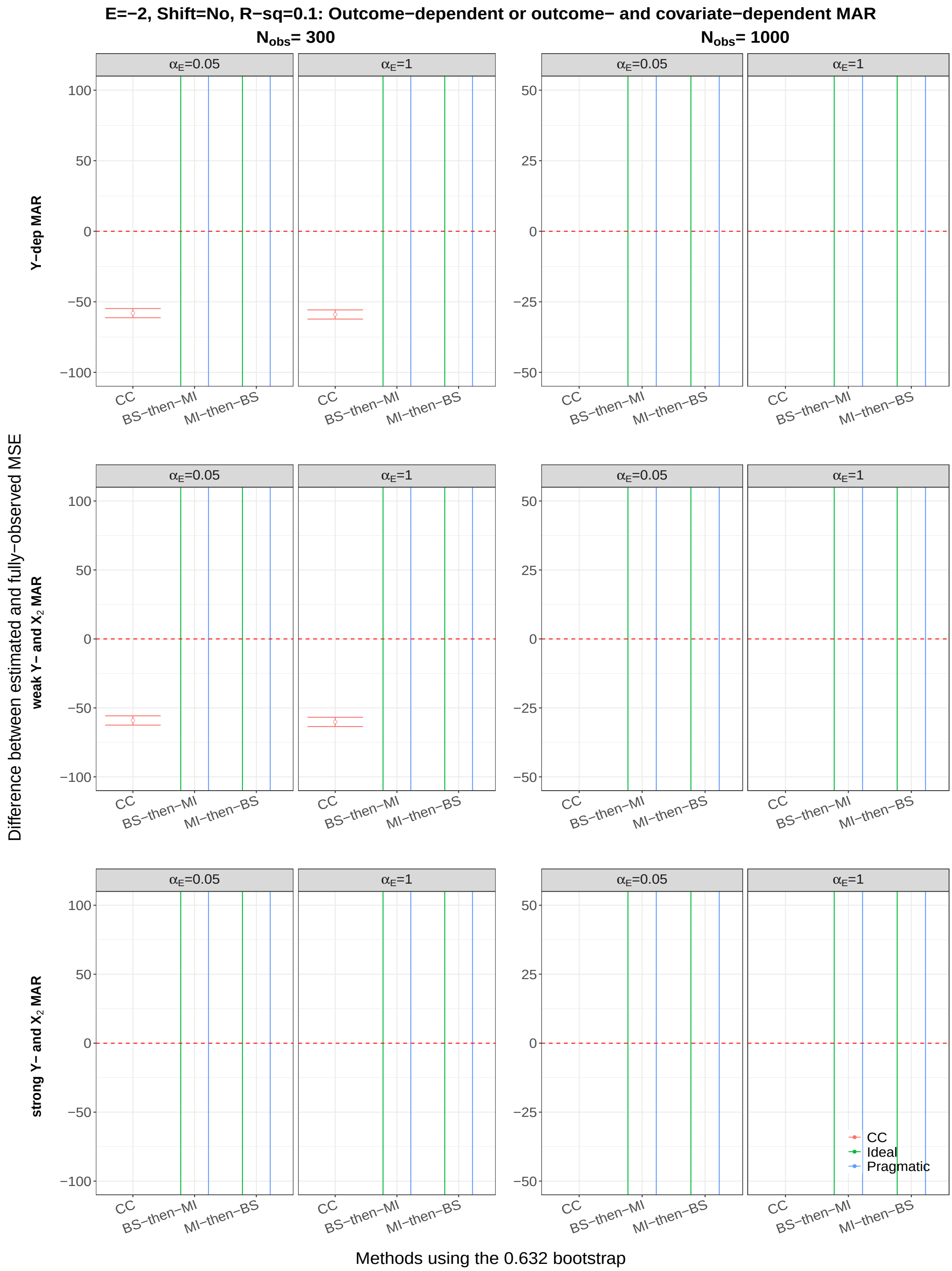

Figure S83: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

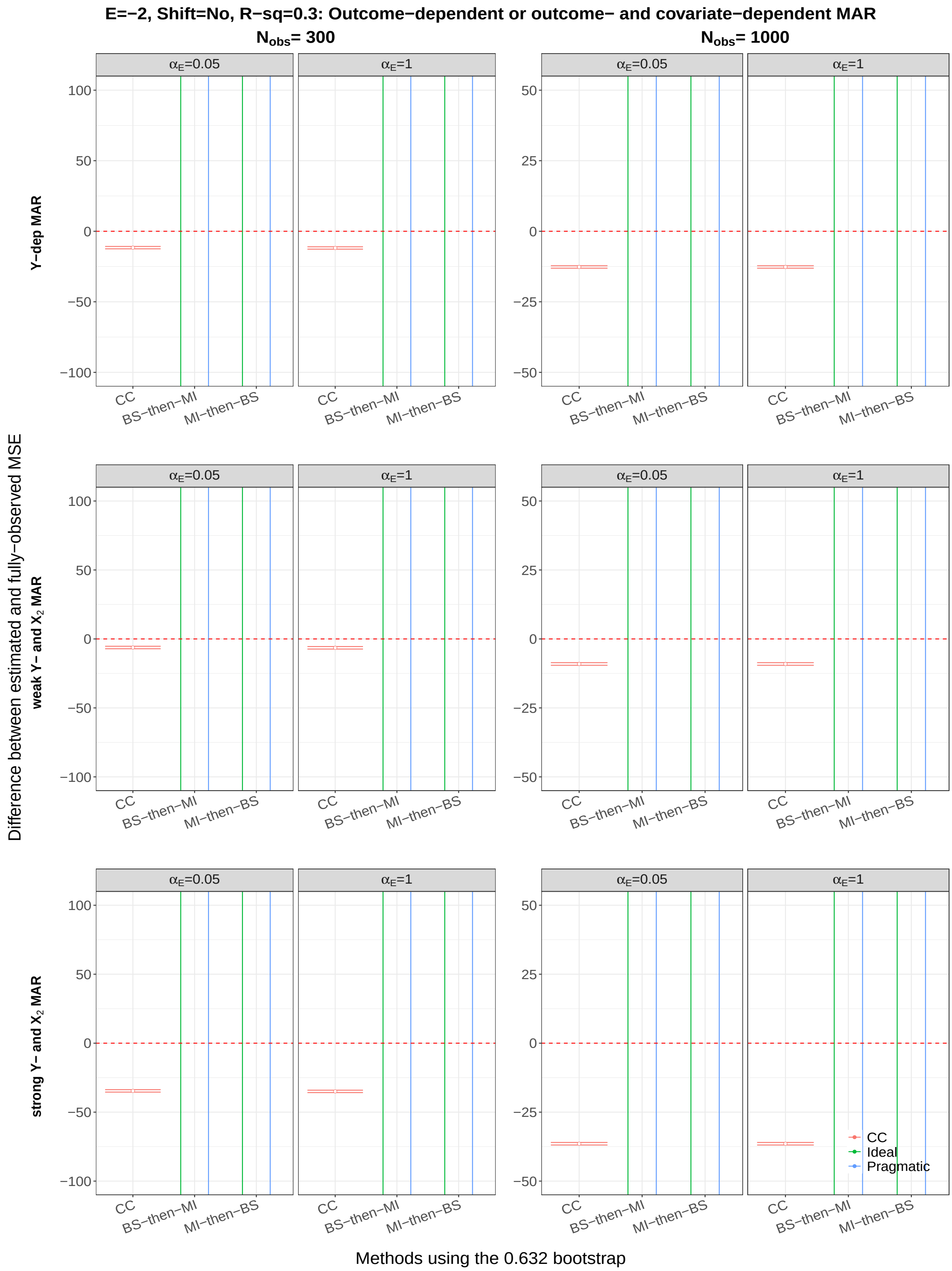

Figure S84: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has not been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S7.2.4 $\beta_2 = 0$ and an origin shift transformation has been applied

True exponent is 0

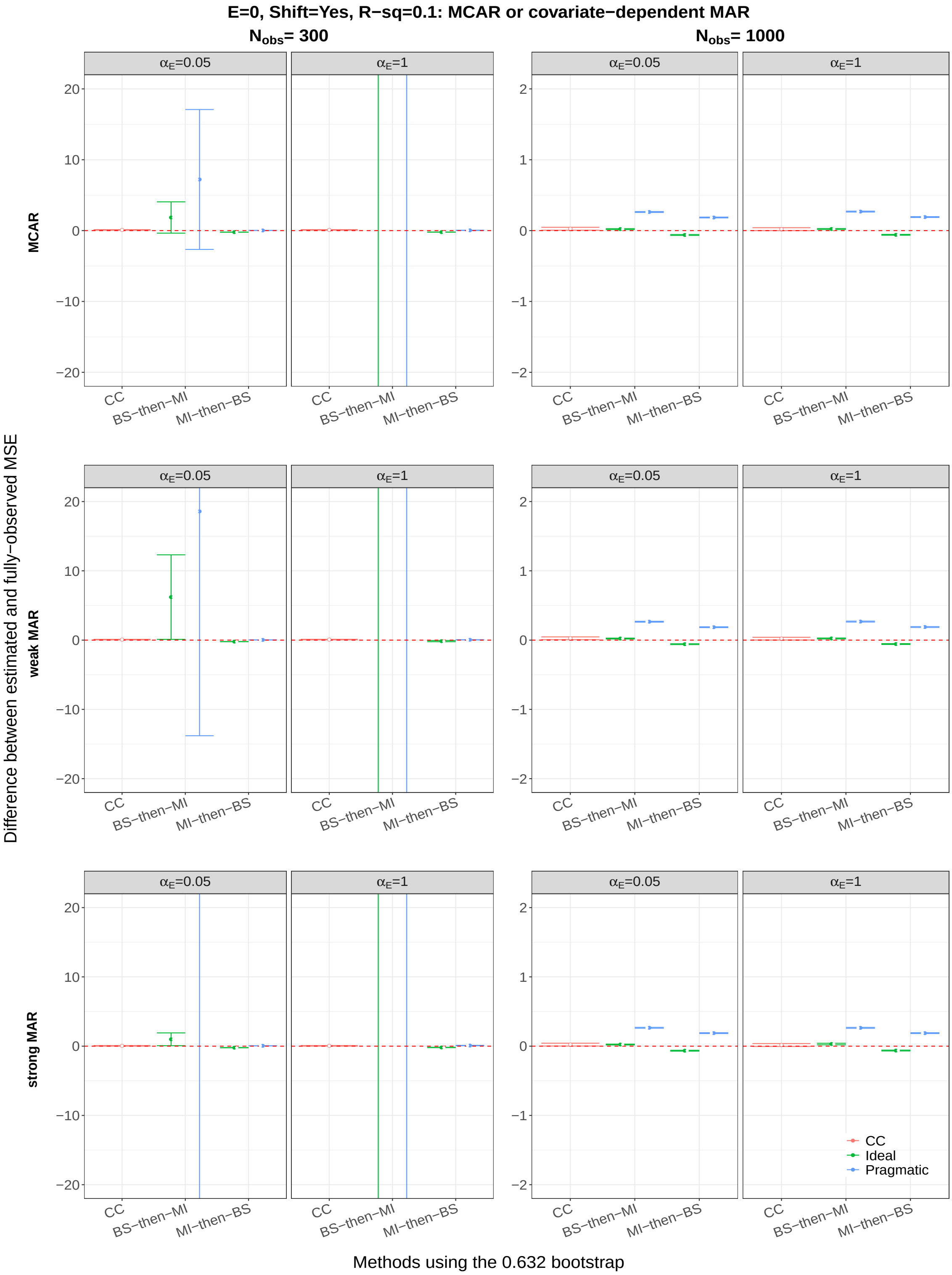

Figure S85: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

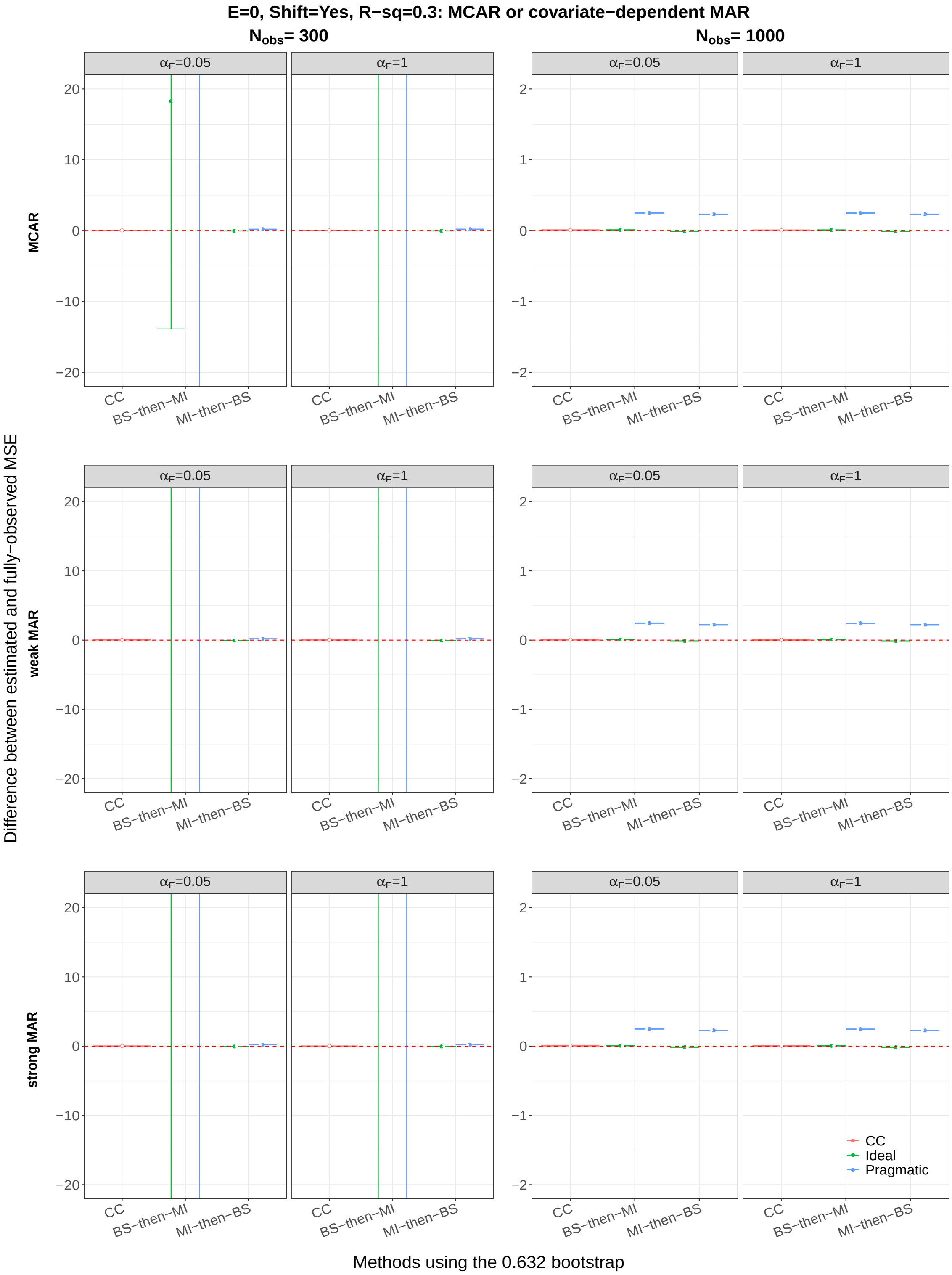

Figure S86: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

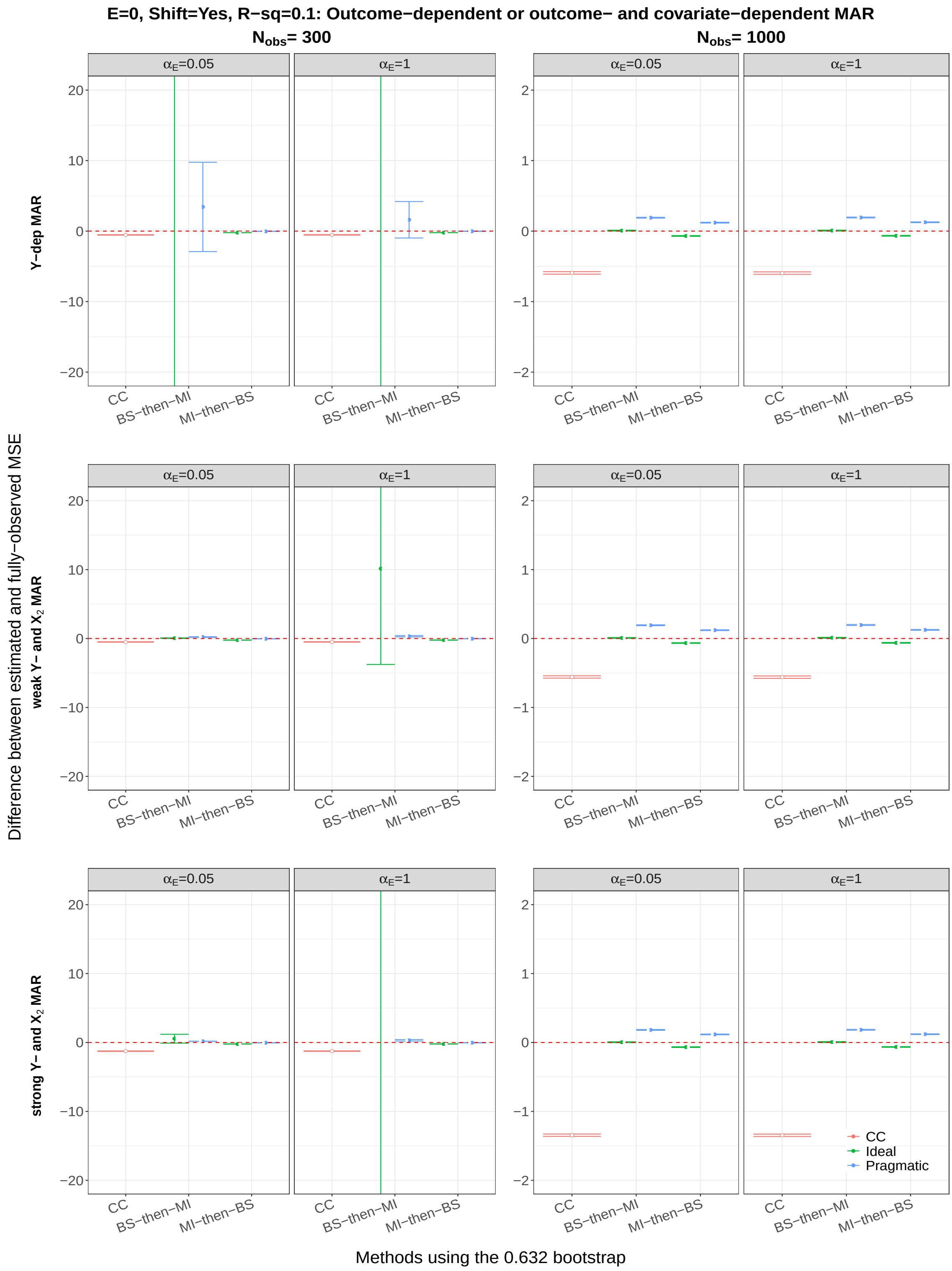

Figure S87: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

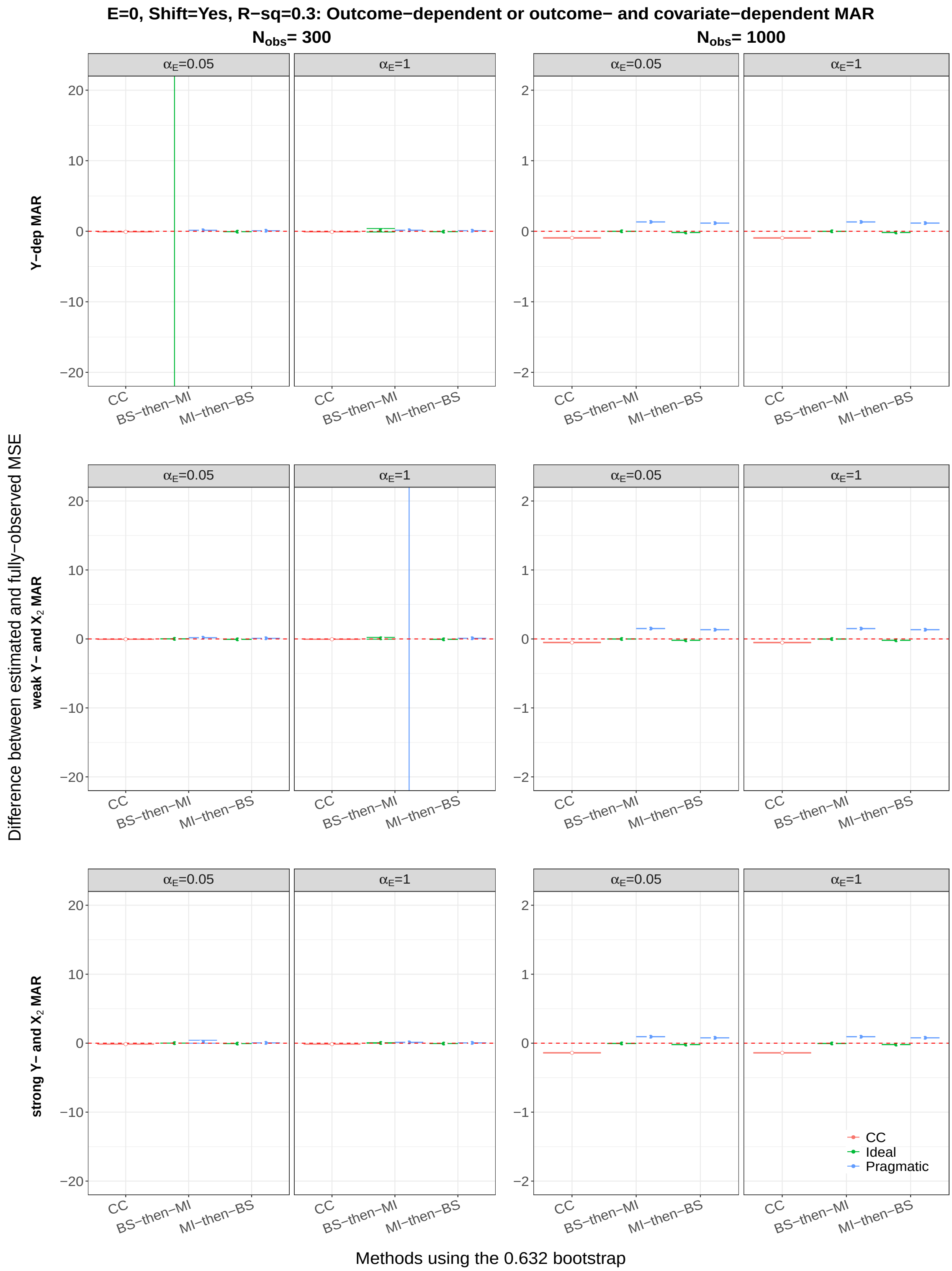

Figure S88: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 0, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

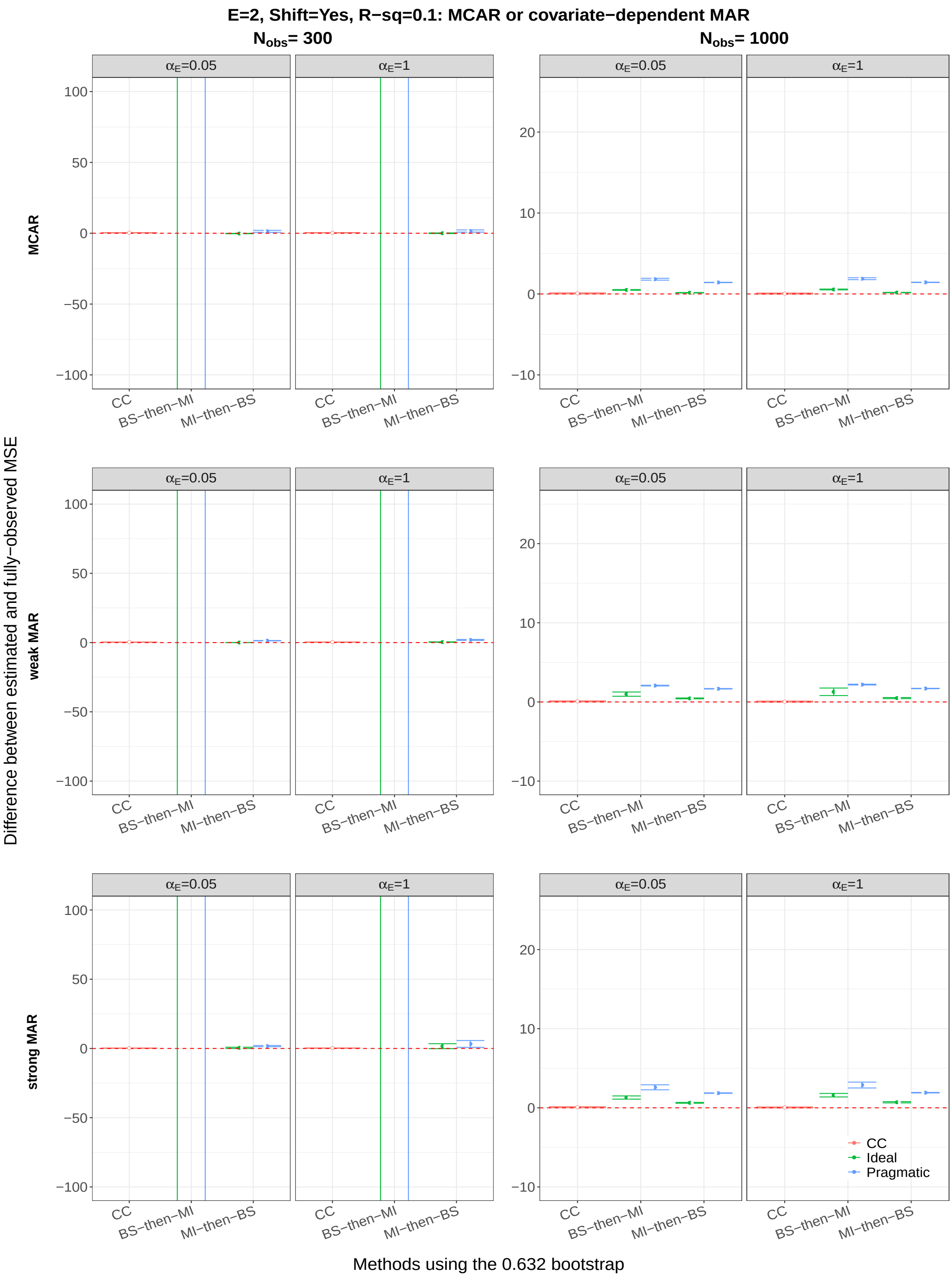

Figure S89: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

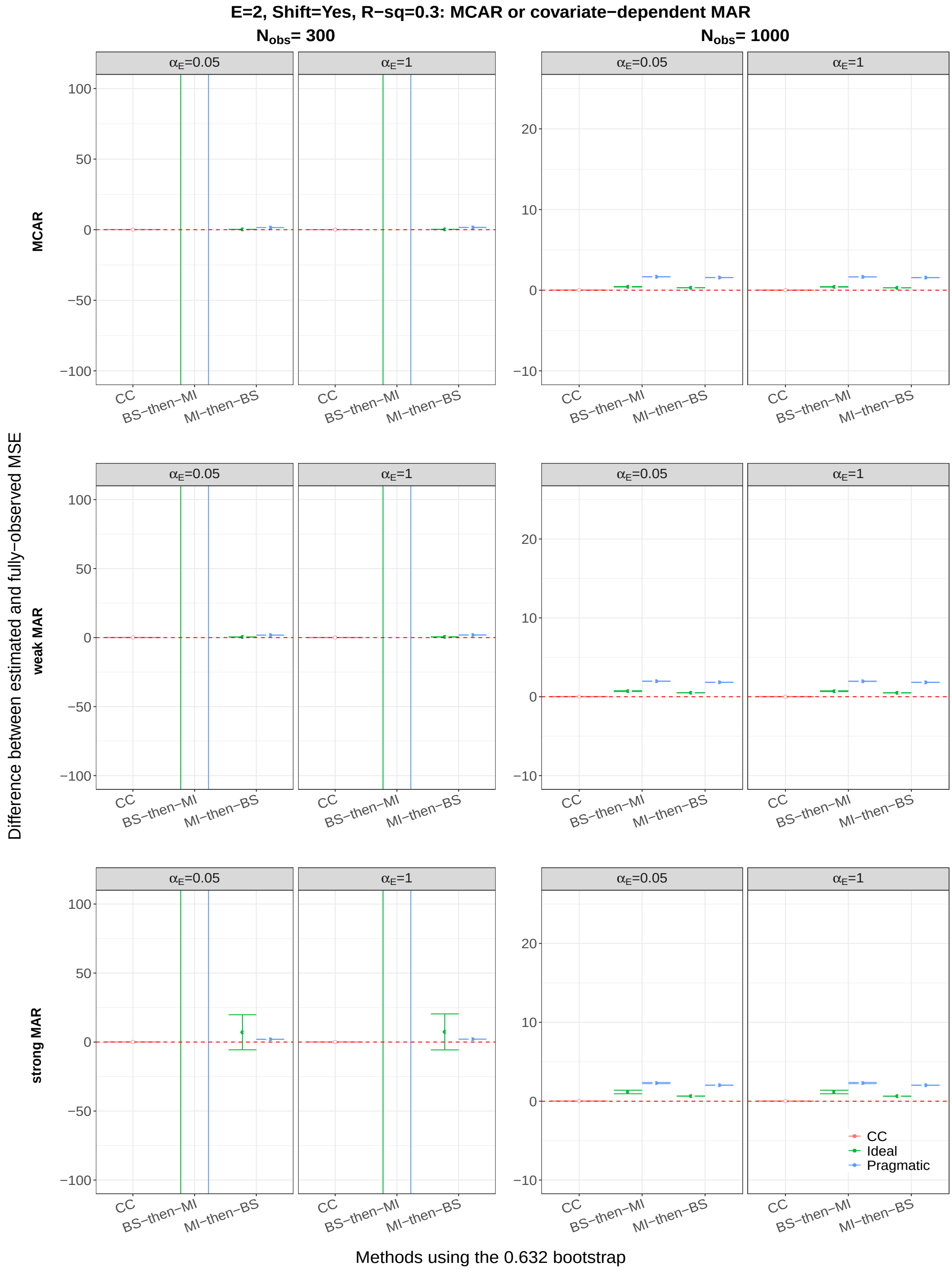

Figure S90: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

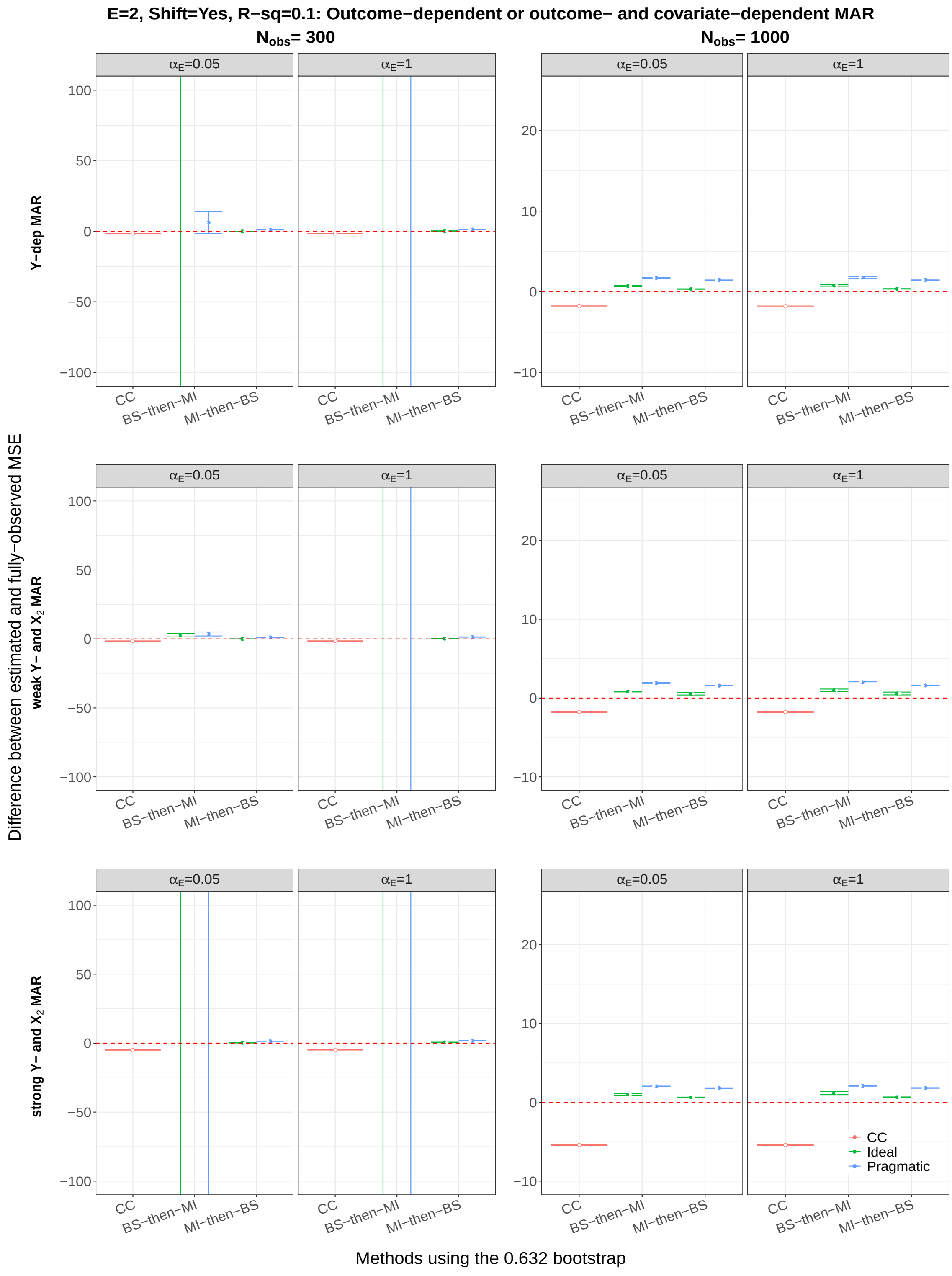

Figure S91: The difference MSE_{imp} - MSE_{obs} when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

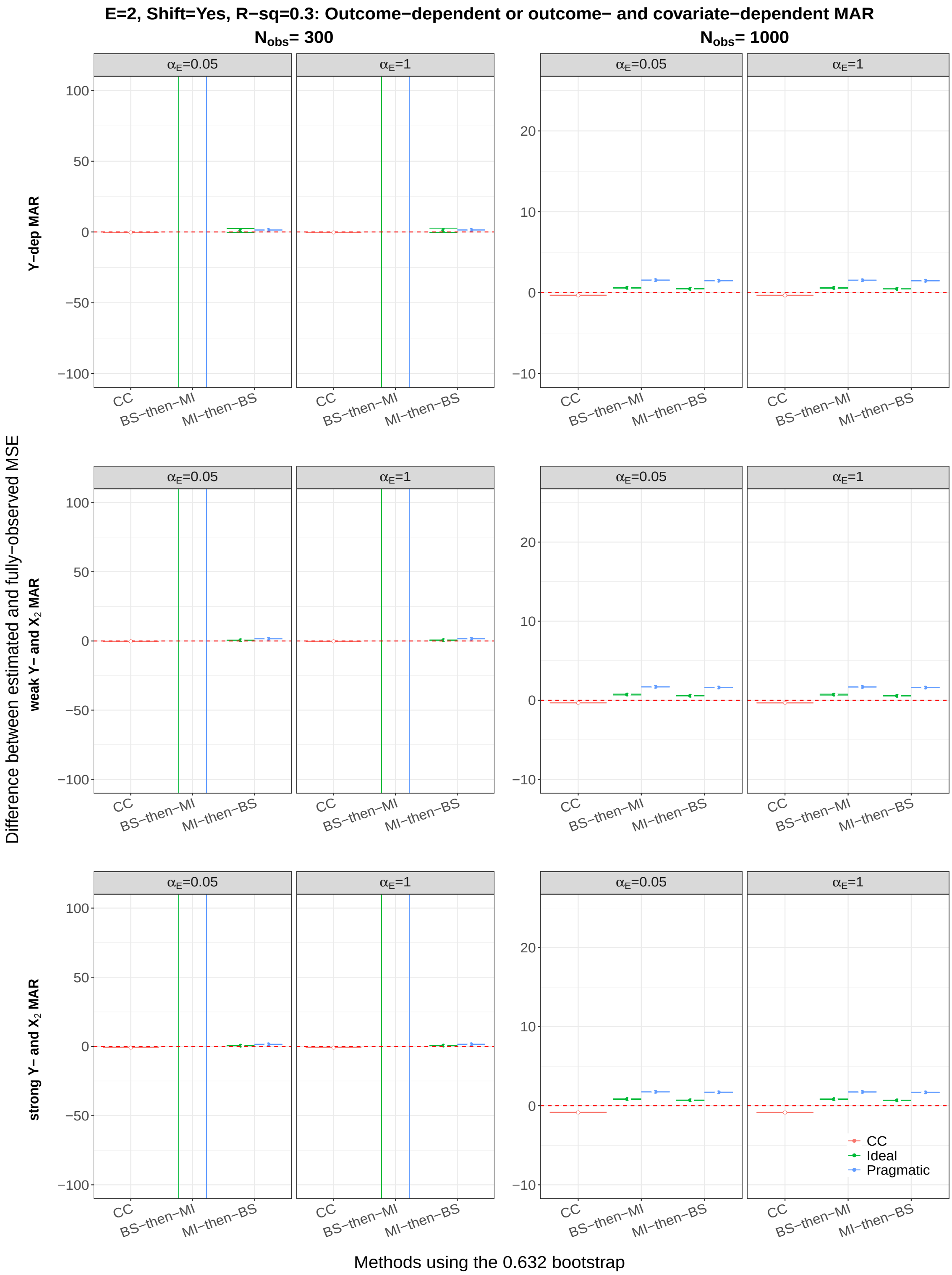

Figure S92: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is 2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

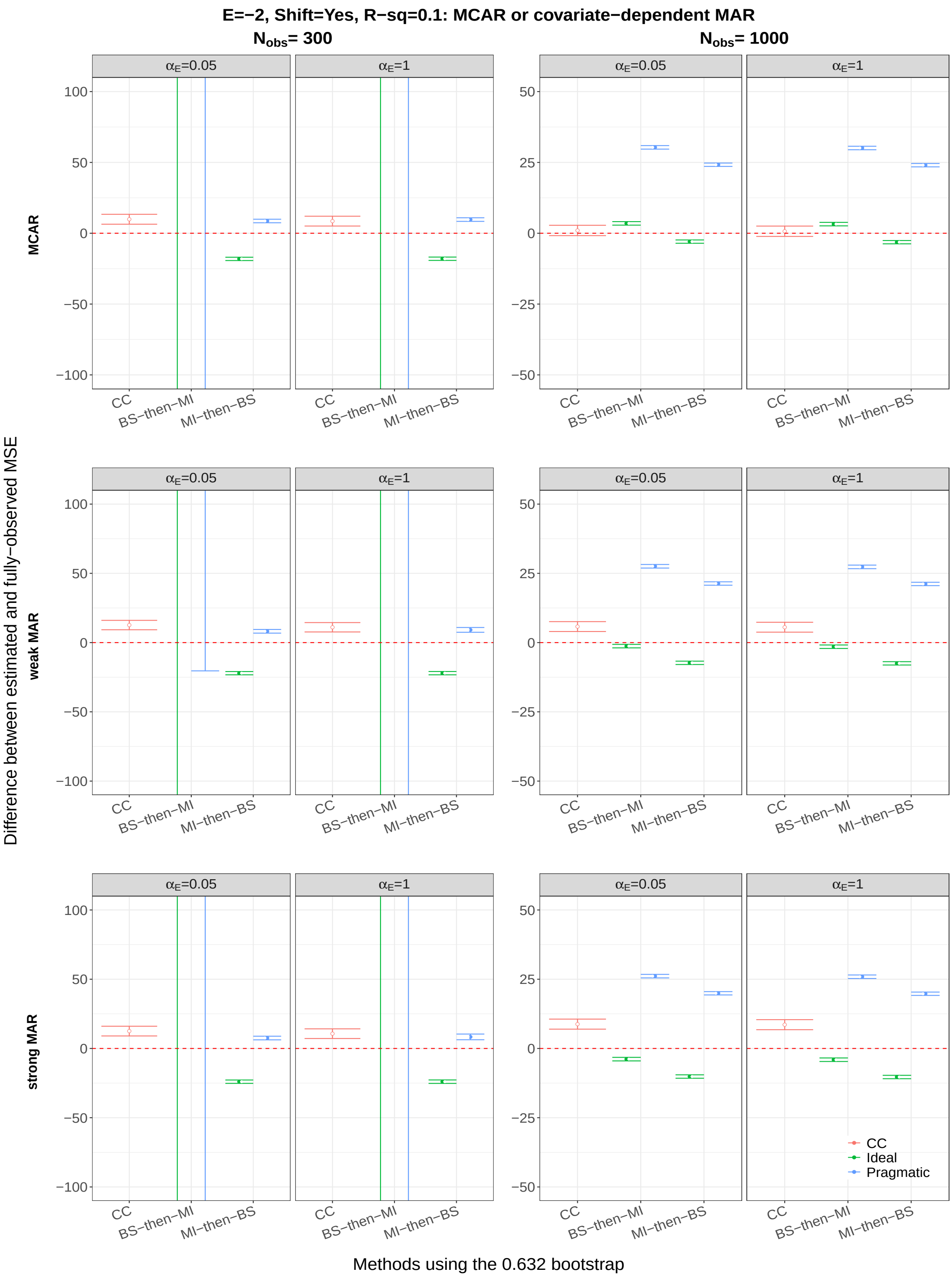

Figure S93: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

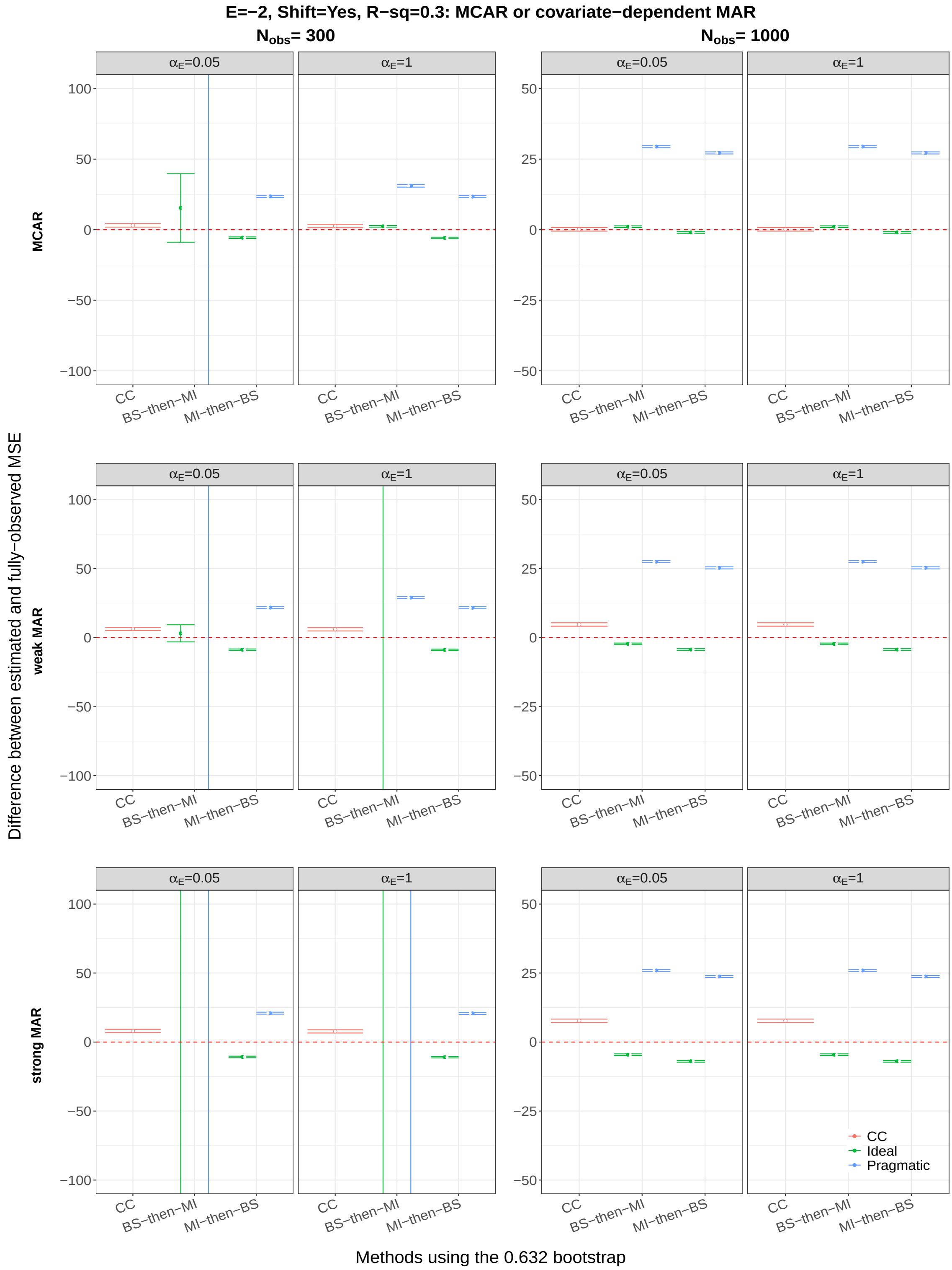

Figure S94: The difference MSE_{imp} - MSE_{obs} when data are MCAR or covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE_{imp} - MSE_{obs}. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

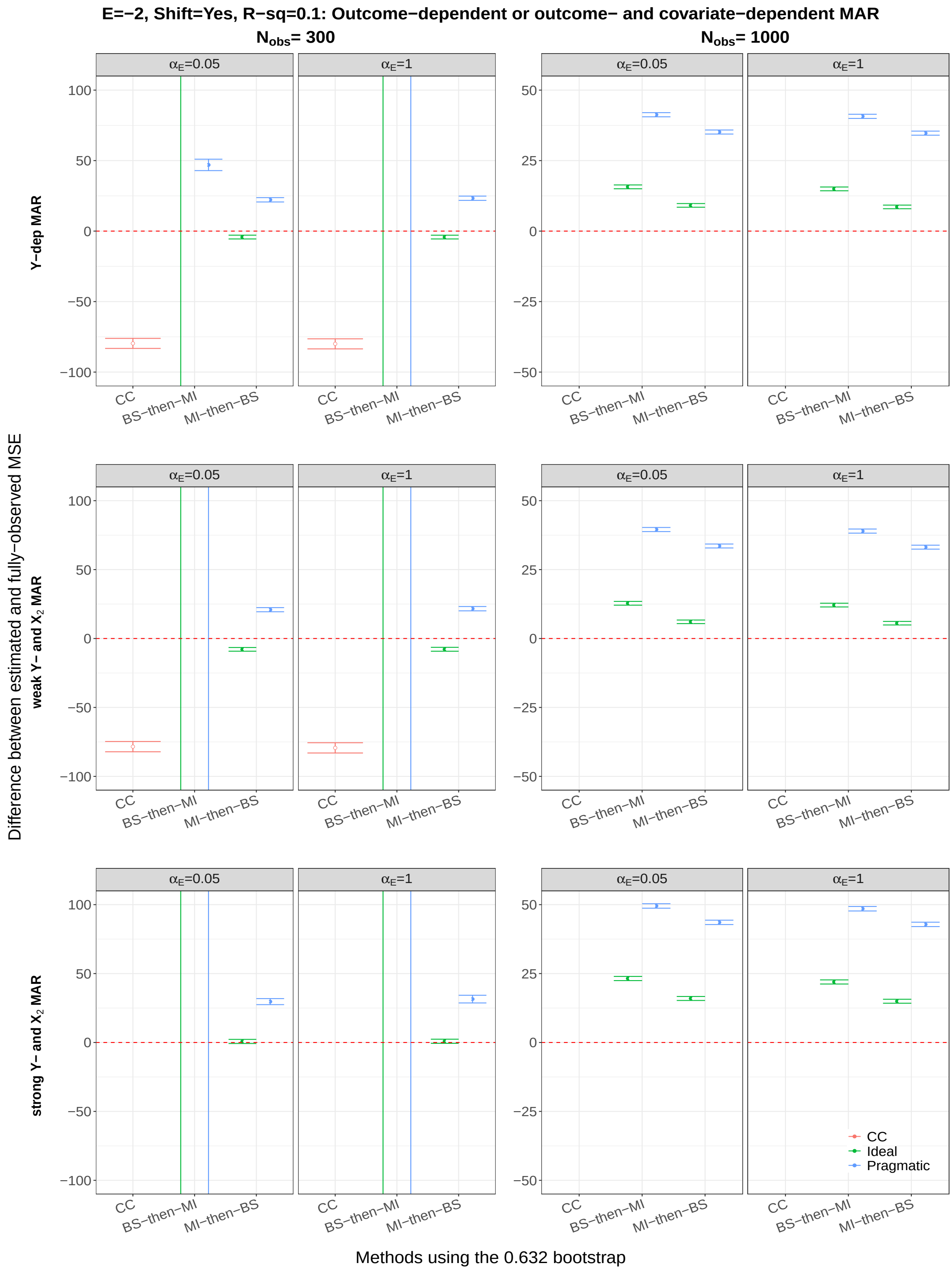

Figure S95: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.1$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

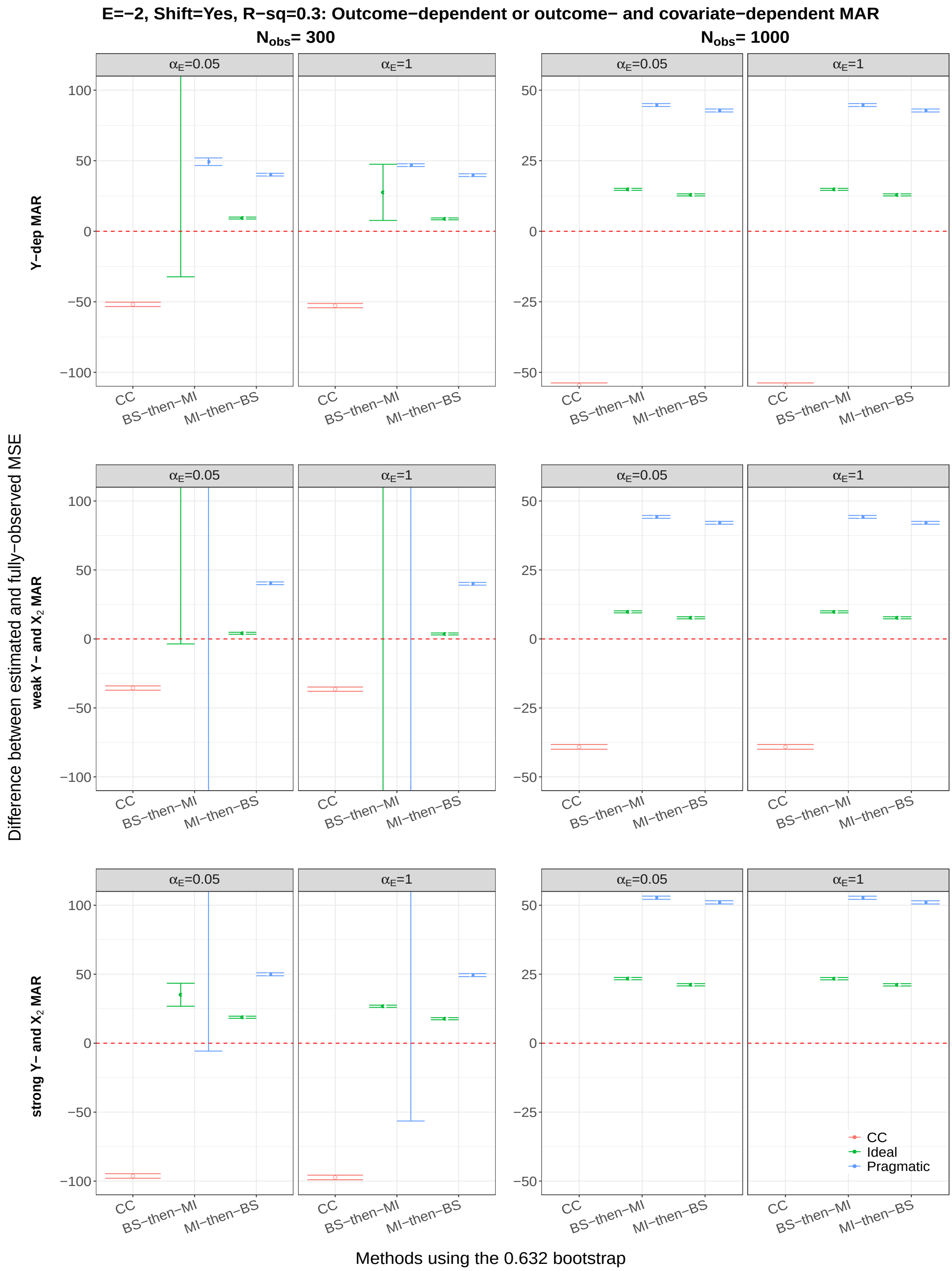

Figure S96: The difference MSE$_{imp}$ - MSE$_{obs}$ when data are outcome-dependent or outcome- and covariate-dependent MAR for $M = 5$ when 25% of values are missing in X_1. The true exponent, E, is -2, an origin-shift transformation has been applied and $R^2 = 0.3$. The error bars summarise results from the 2000 repetitions and the limits represent the Monte Carlo 95% confidence interval of MSE$_{imp}$ - MSE$_{obs}$. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8 Chapter 9: Simulation study results for MFP, exponent selection (Section 9.6)

S8.1 ABB exponent selection

S8.1.1 Cross-validation, $\beta_2 = 1$

True exponent is 0

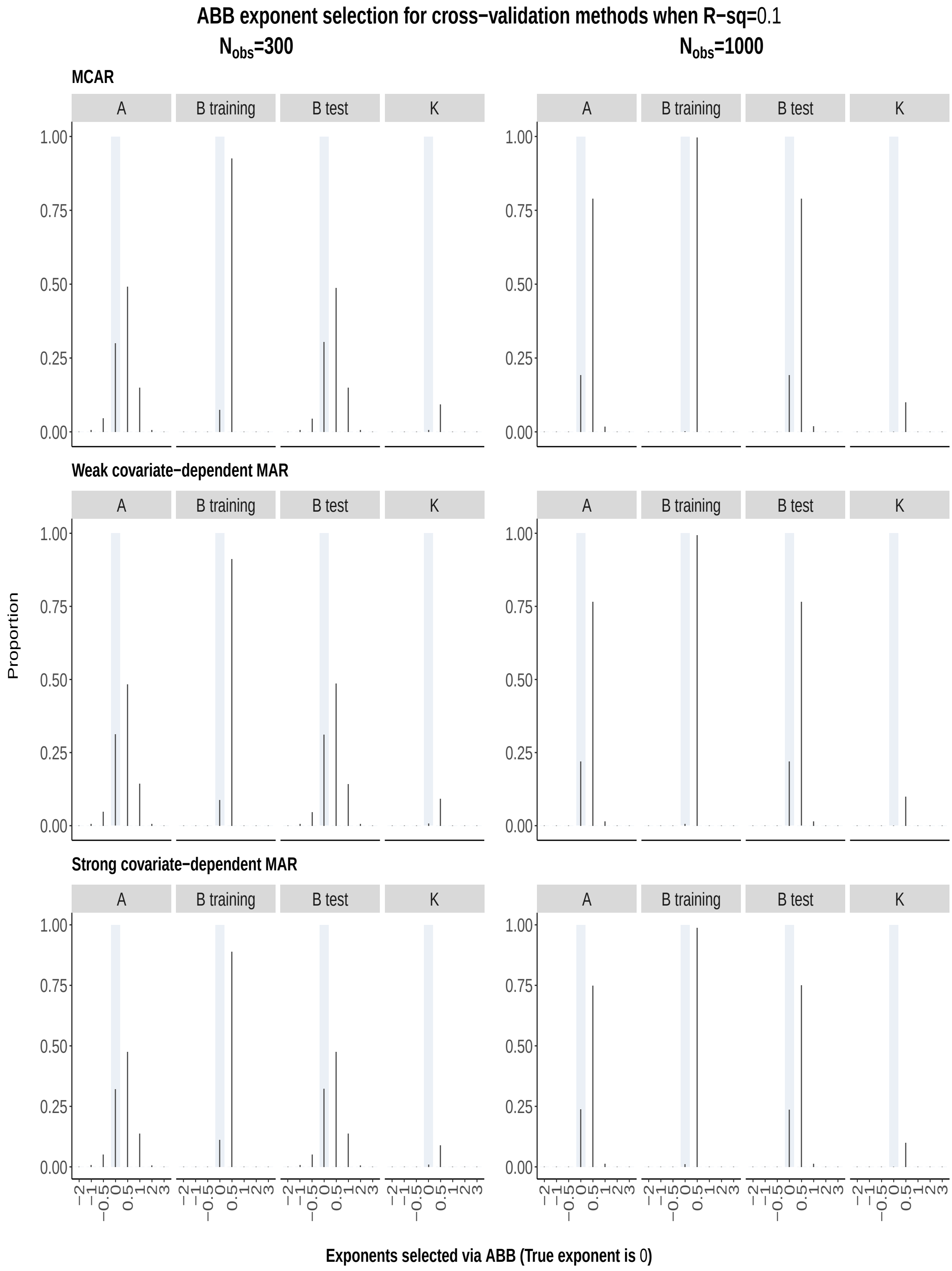

Figure S1: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

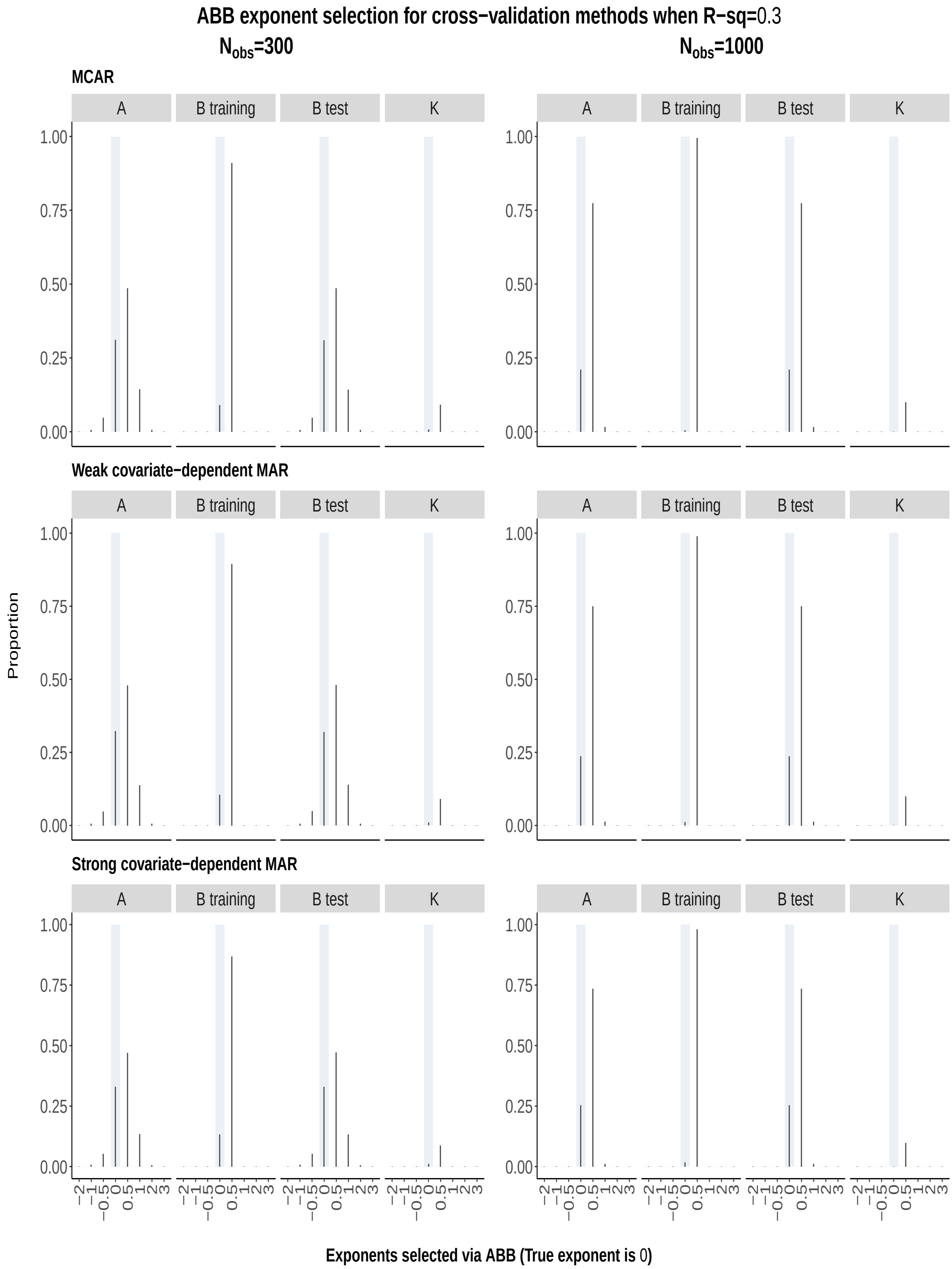

Figure S2: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

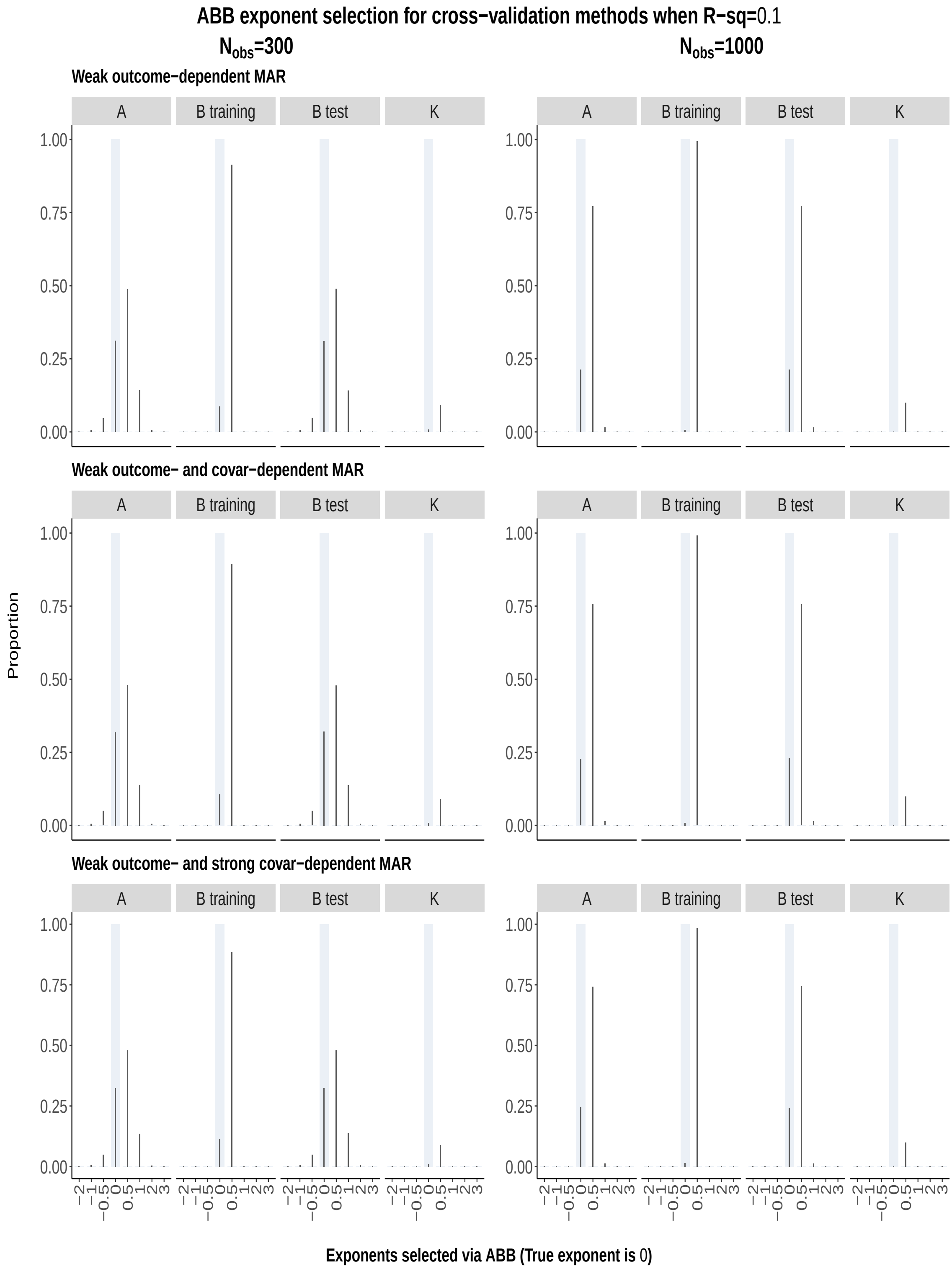

Figure S3: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

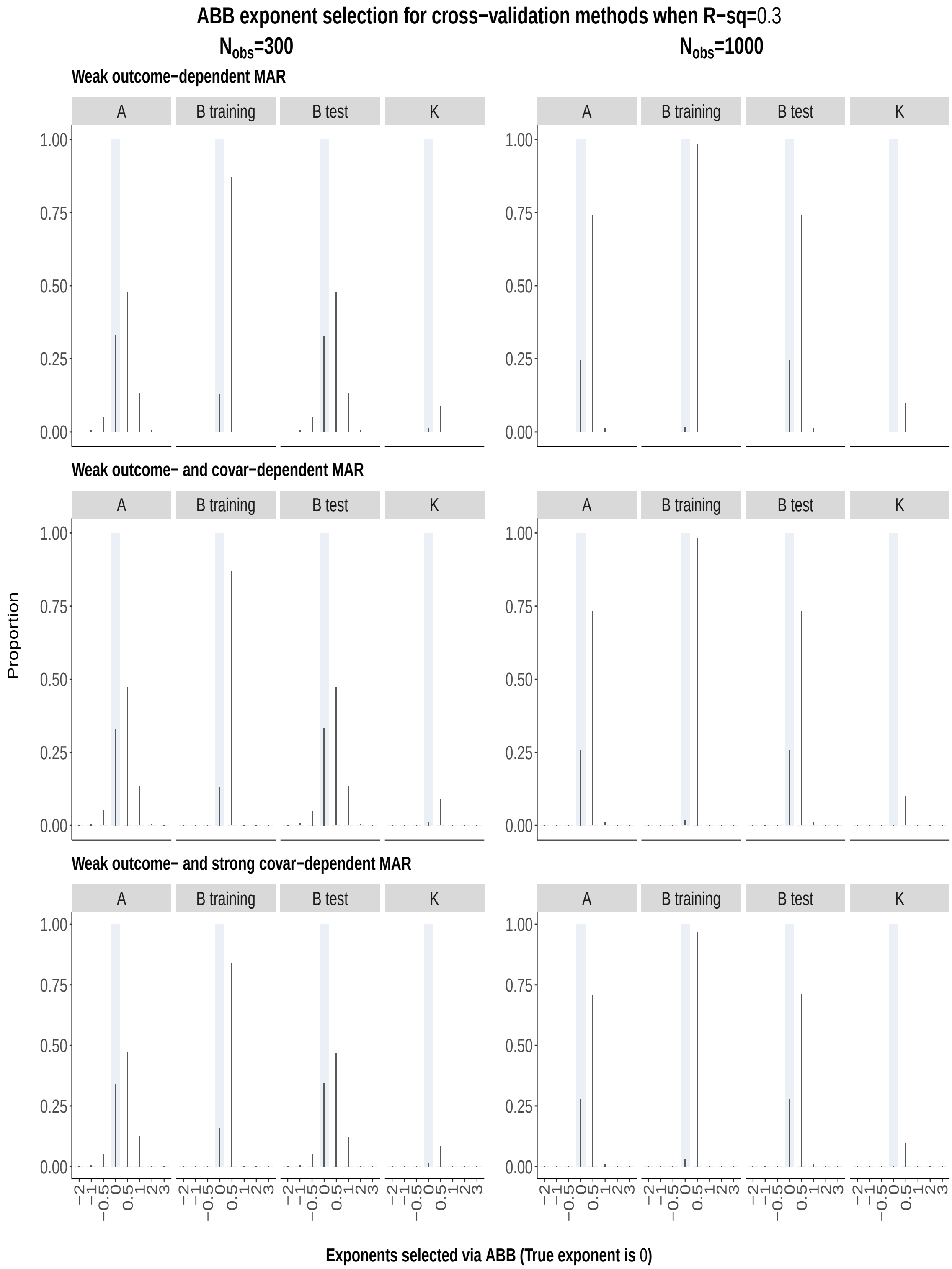

Figure S4: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

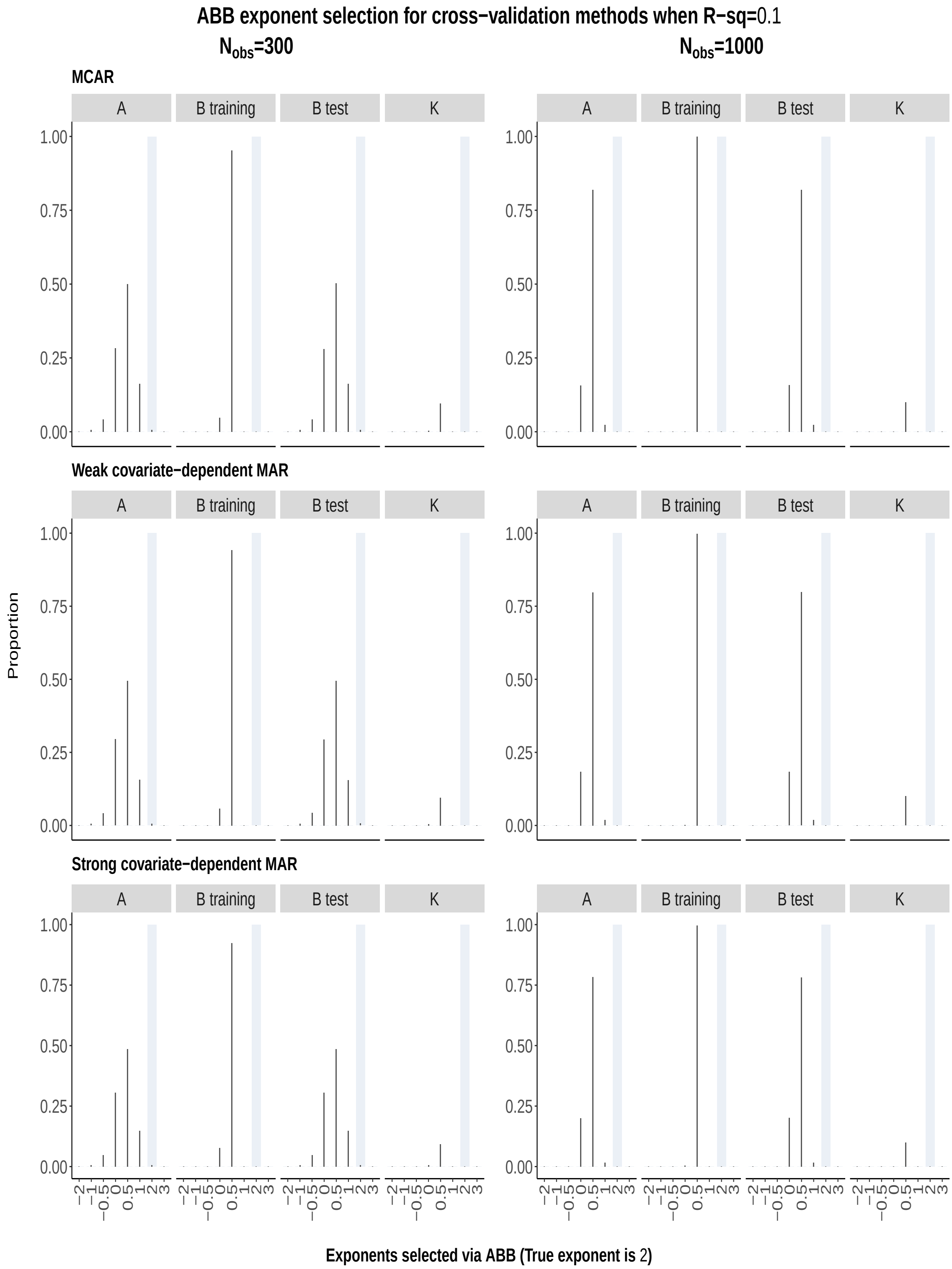

Figure S5: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

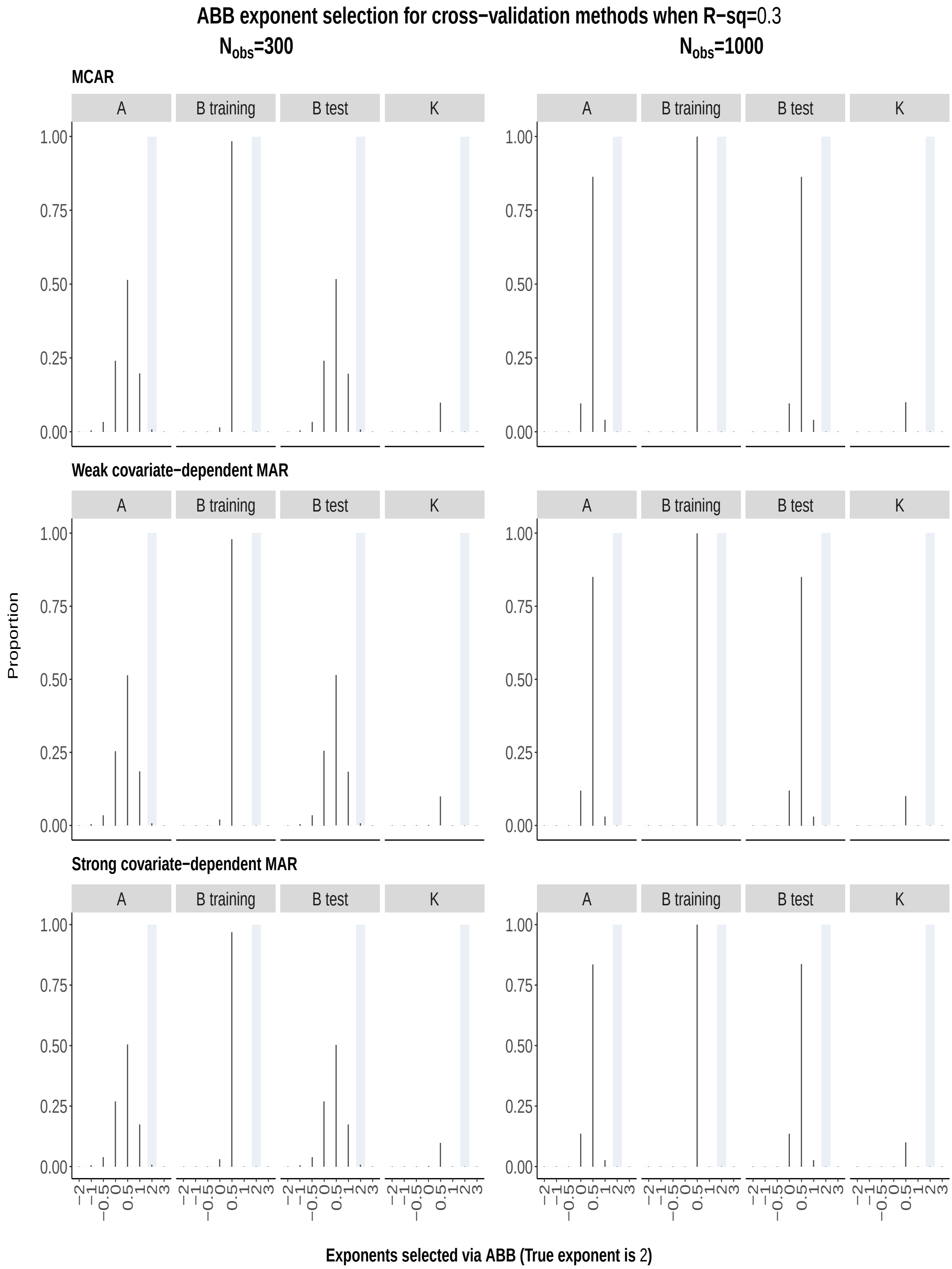

Figure S6: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

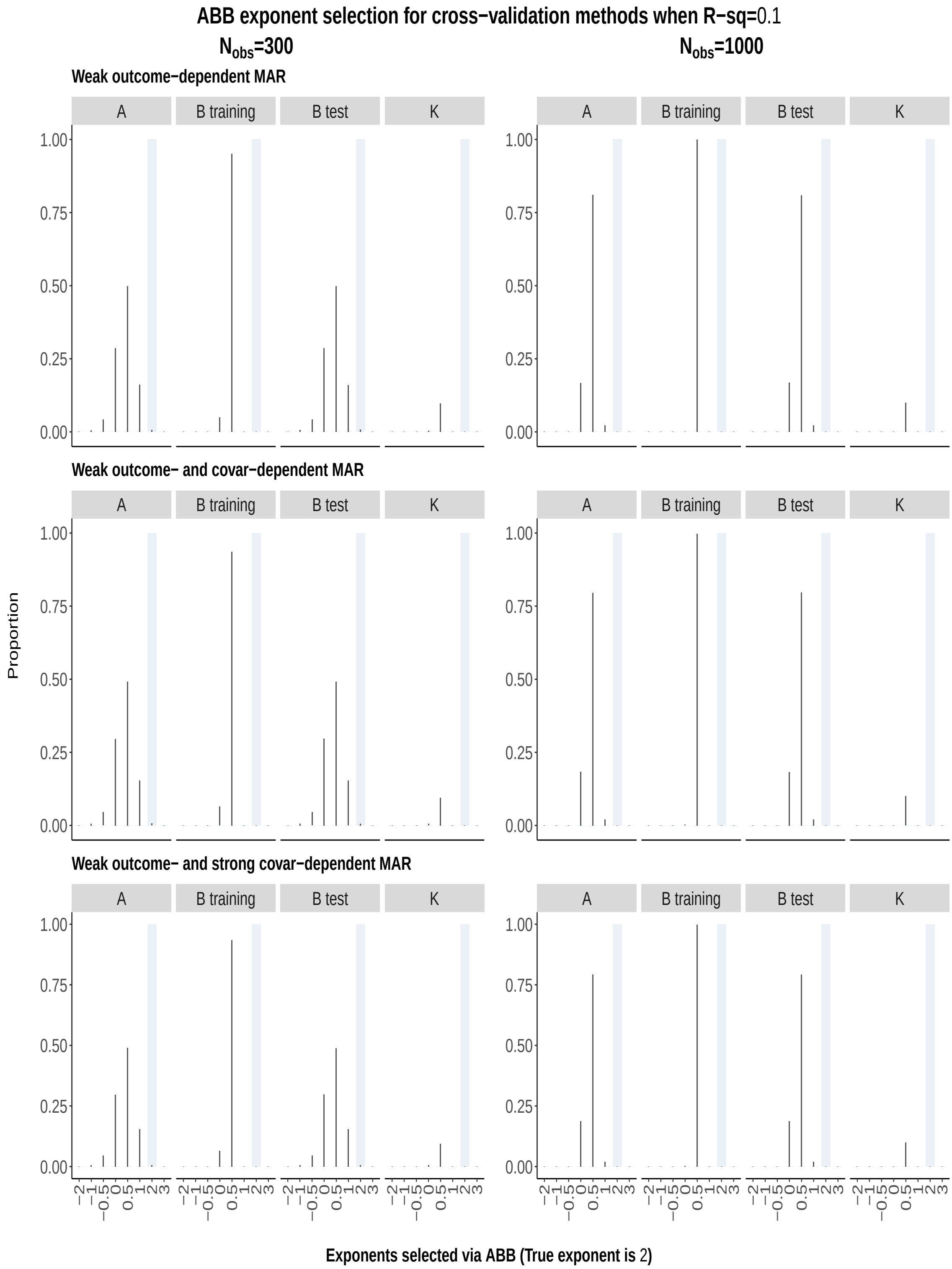

Figure S7: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

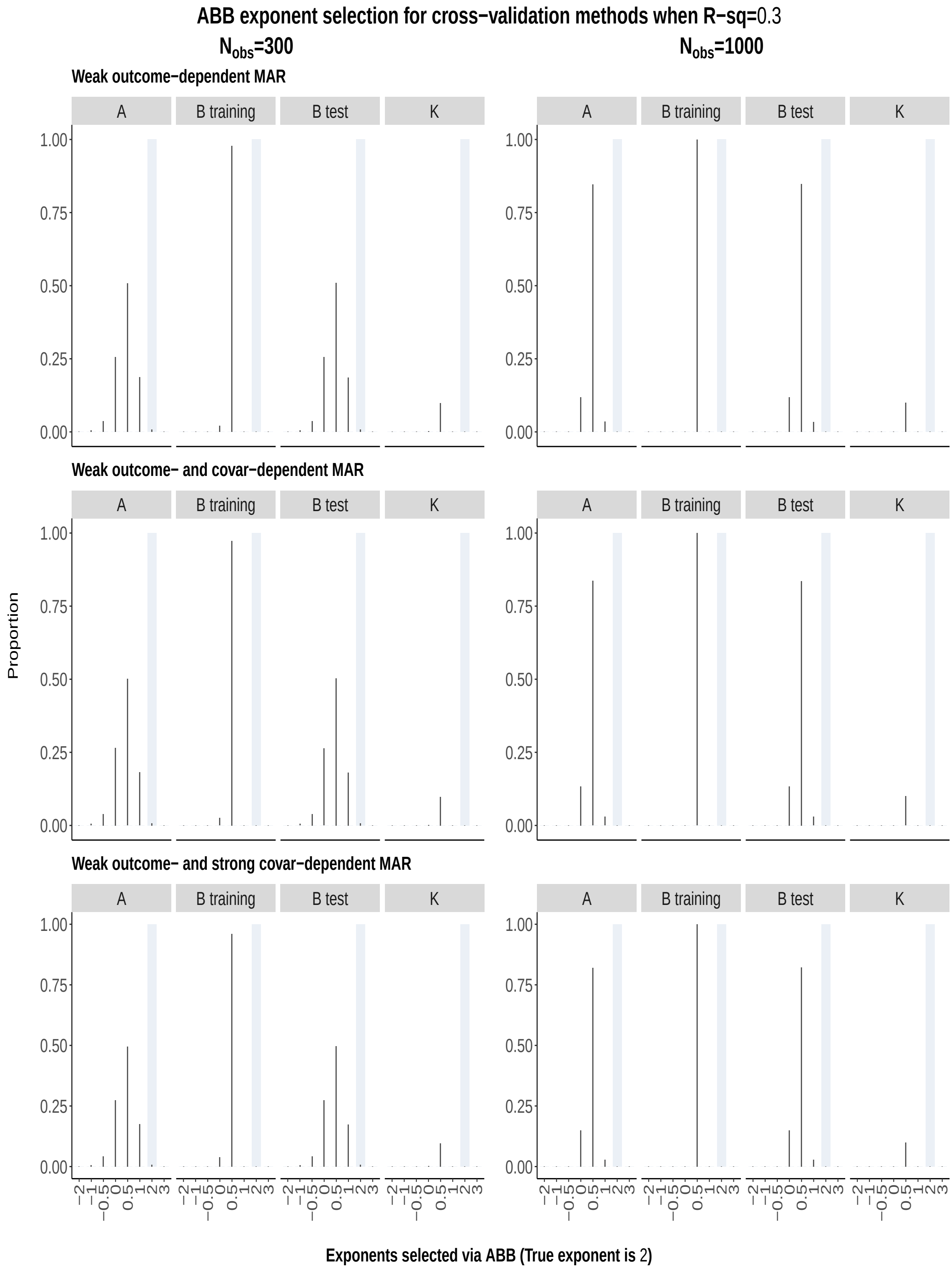

Figure S8: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

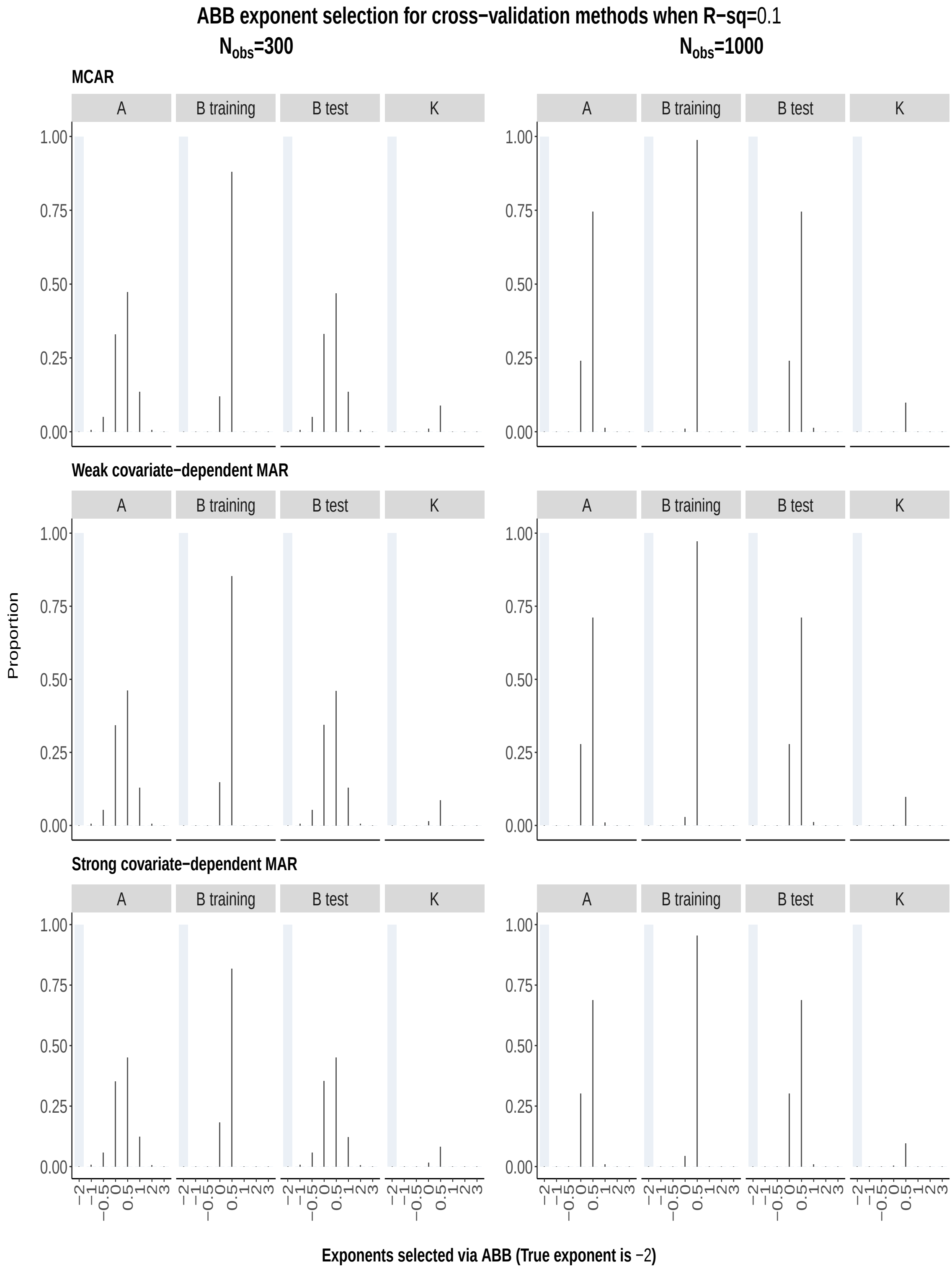

Figure S9: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

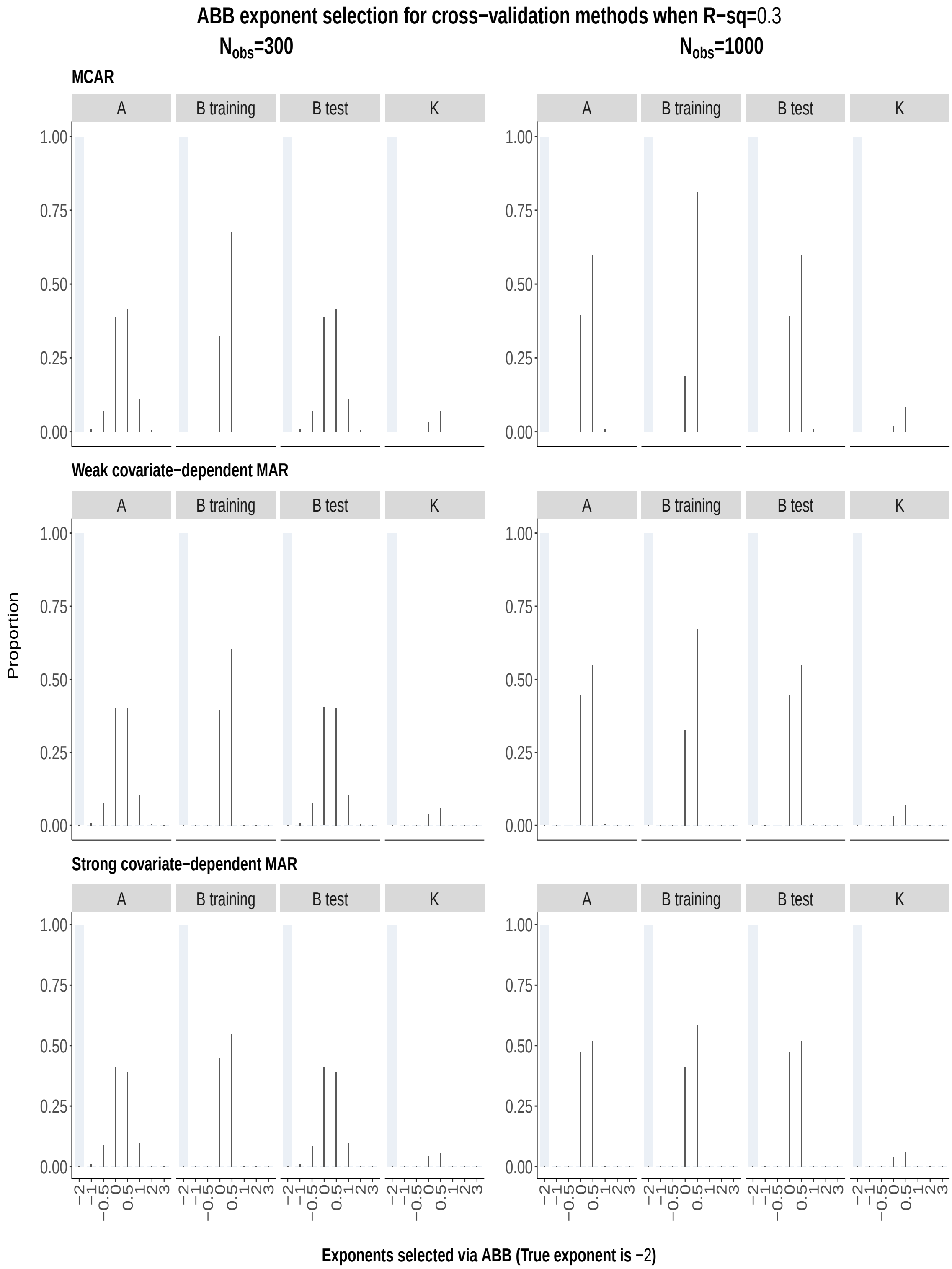

Figure S10: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

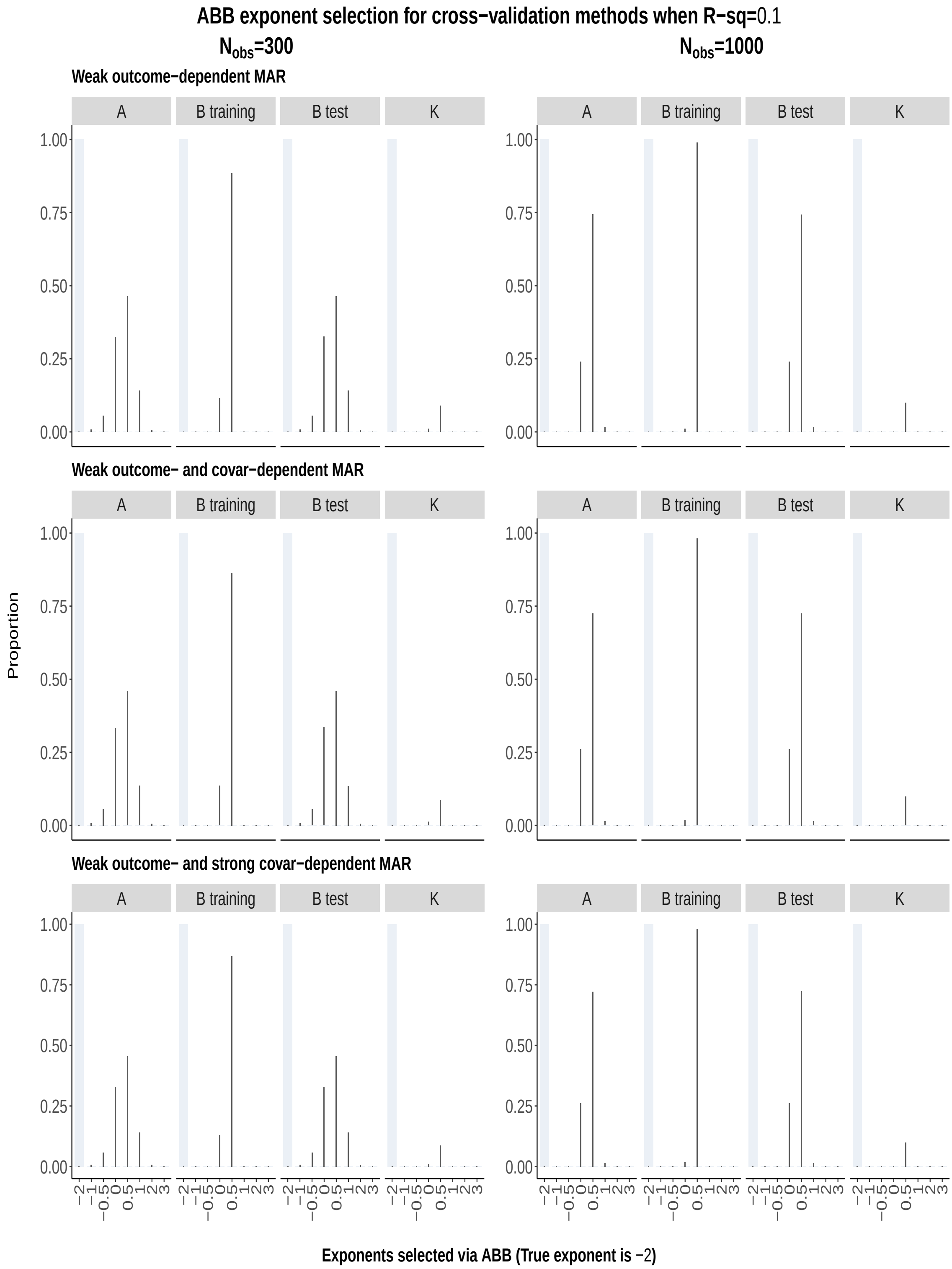

Figure S11: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

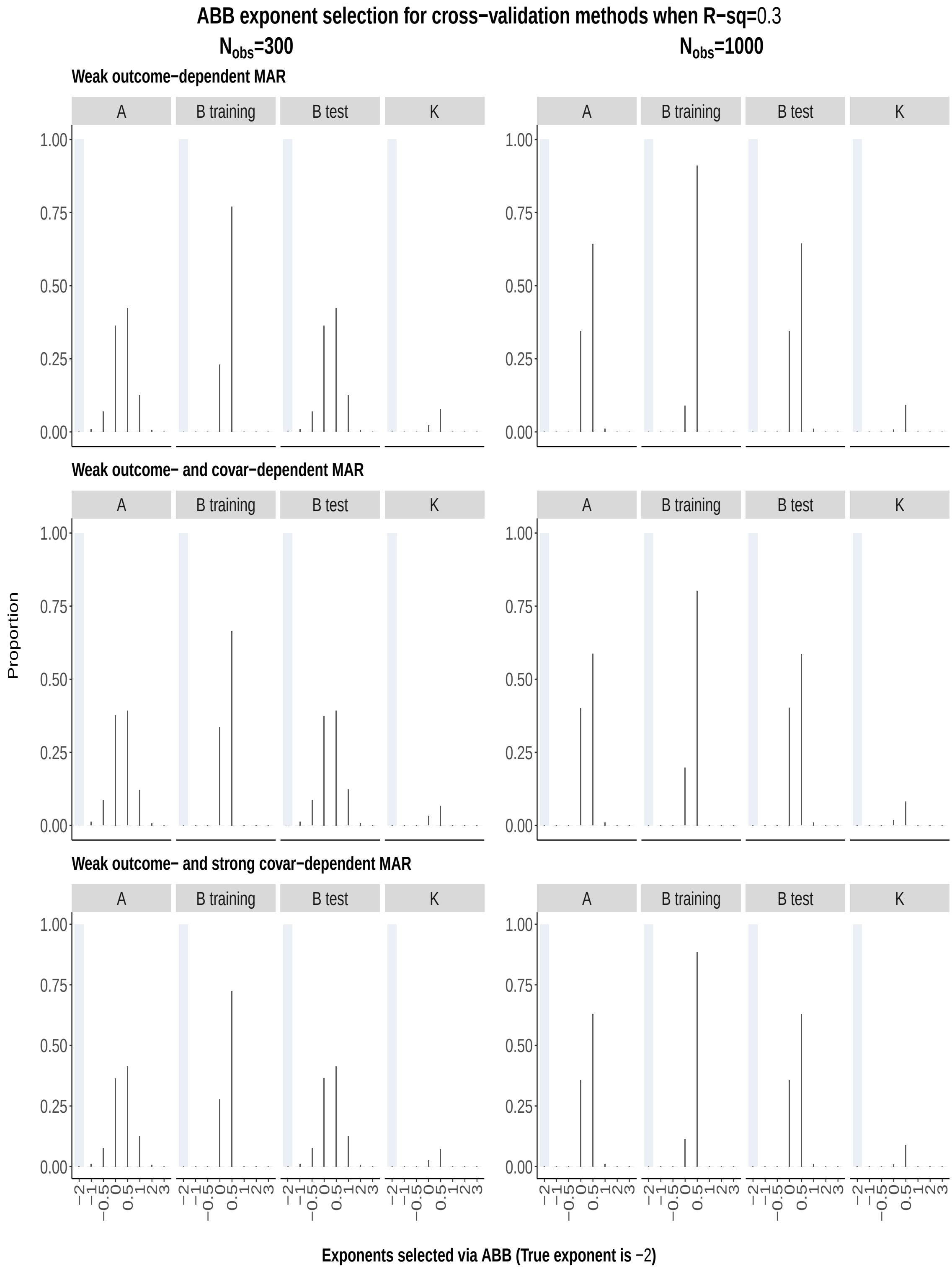

Figure S12: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.1.2 Cross-validation, $\beta_2 = 0$

True exponent is 0

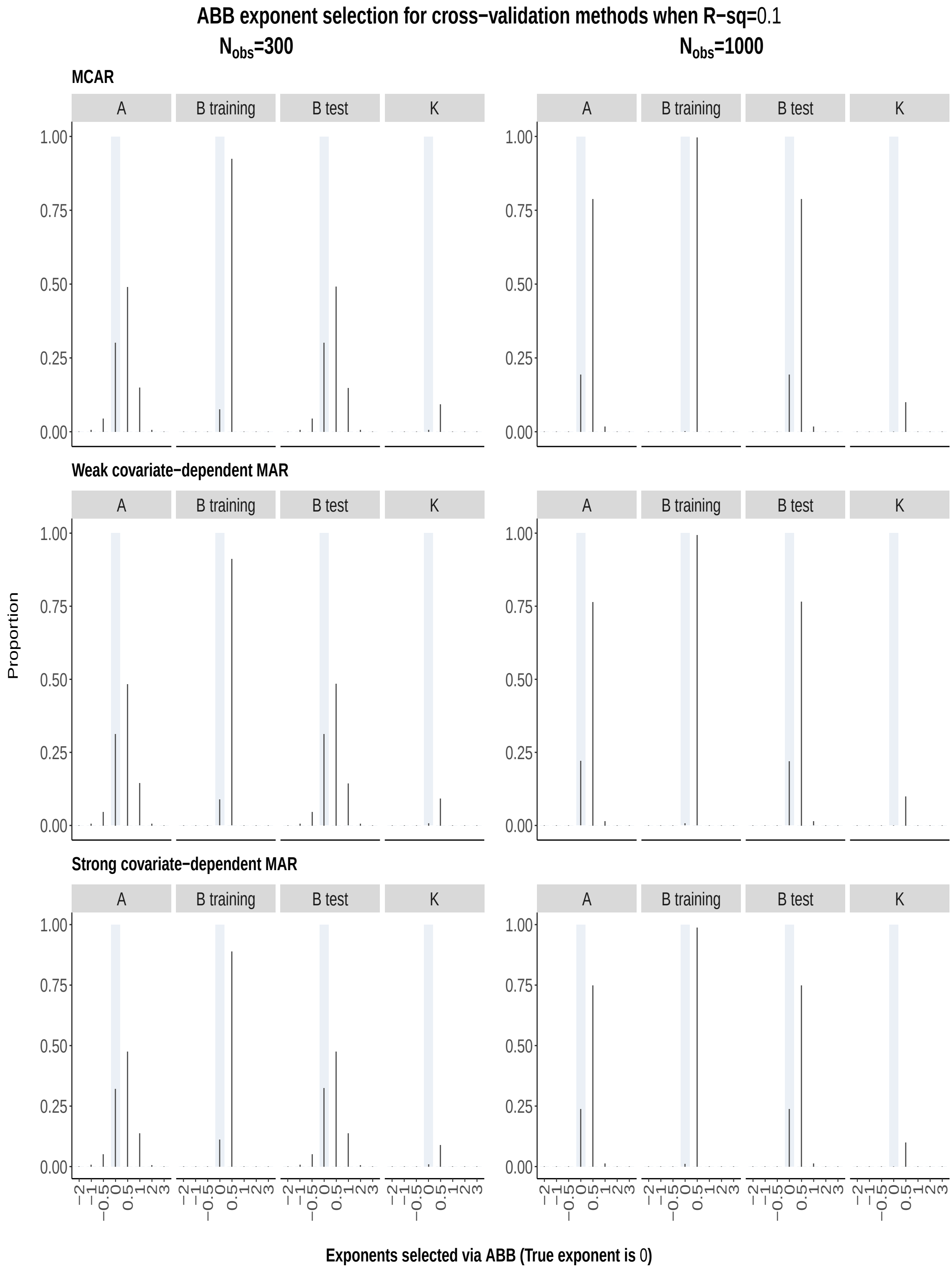

Figure S13: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

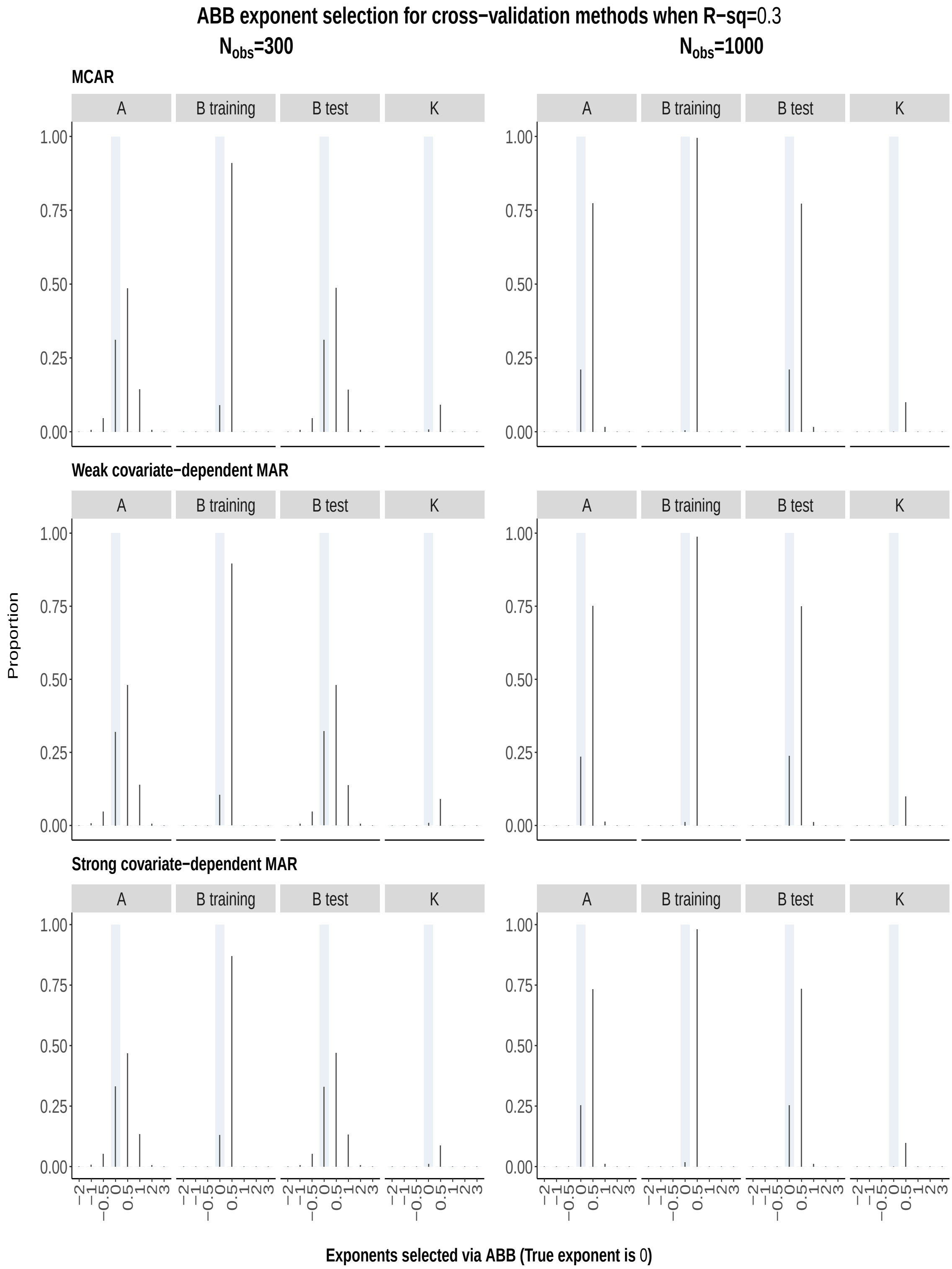

Figure S14: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

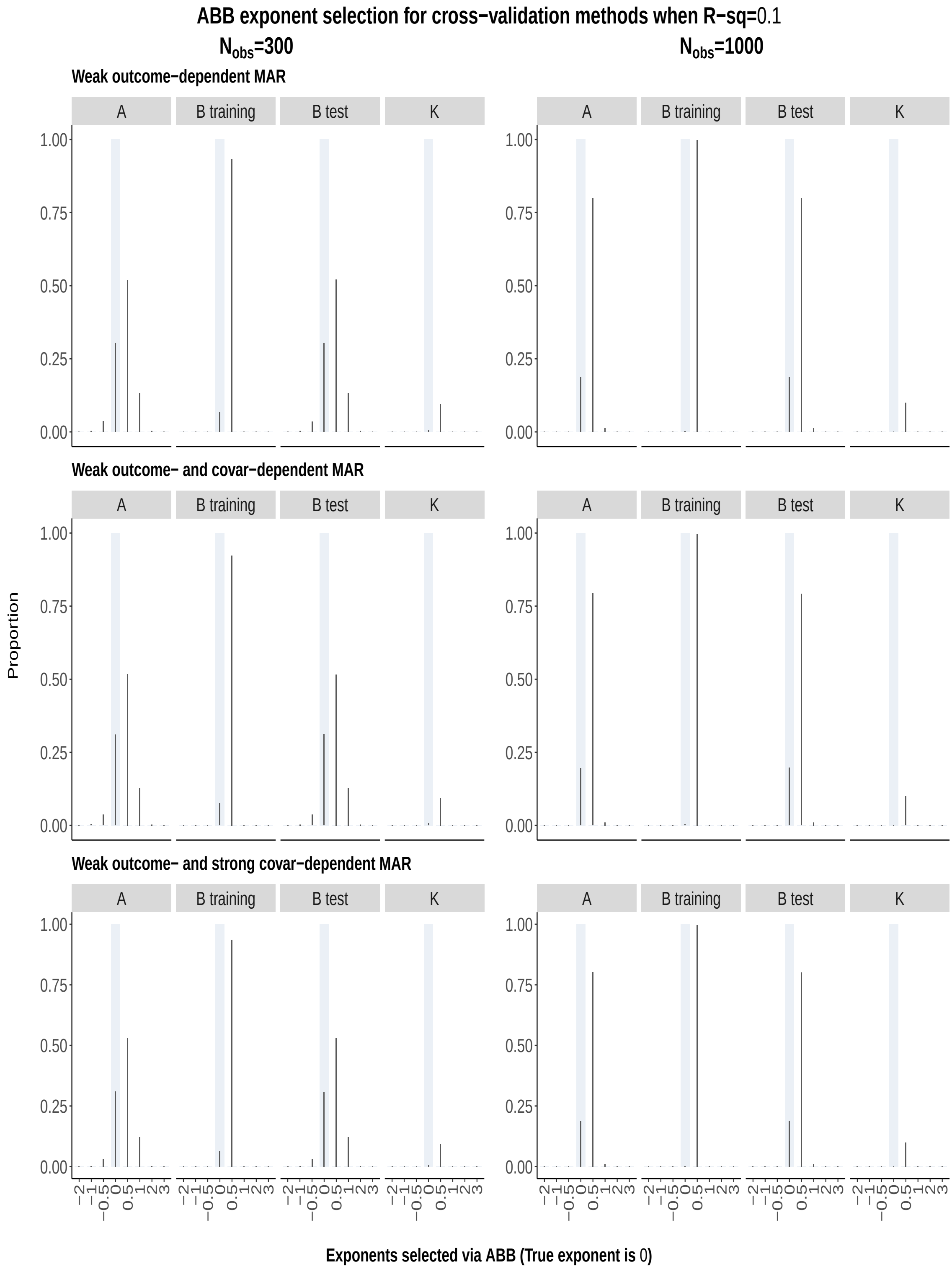

Figure S15: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

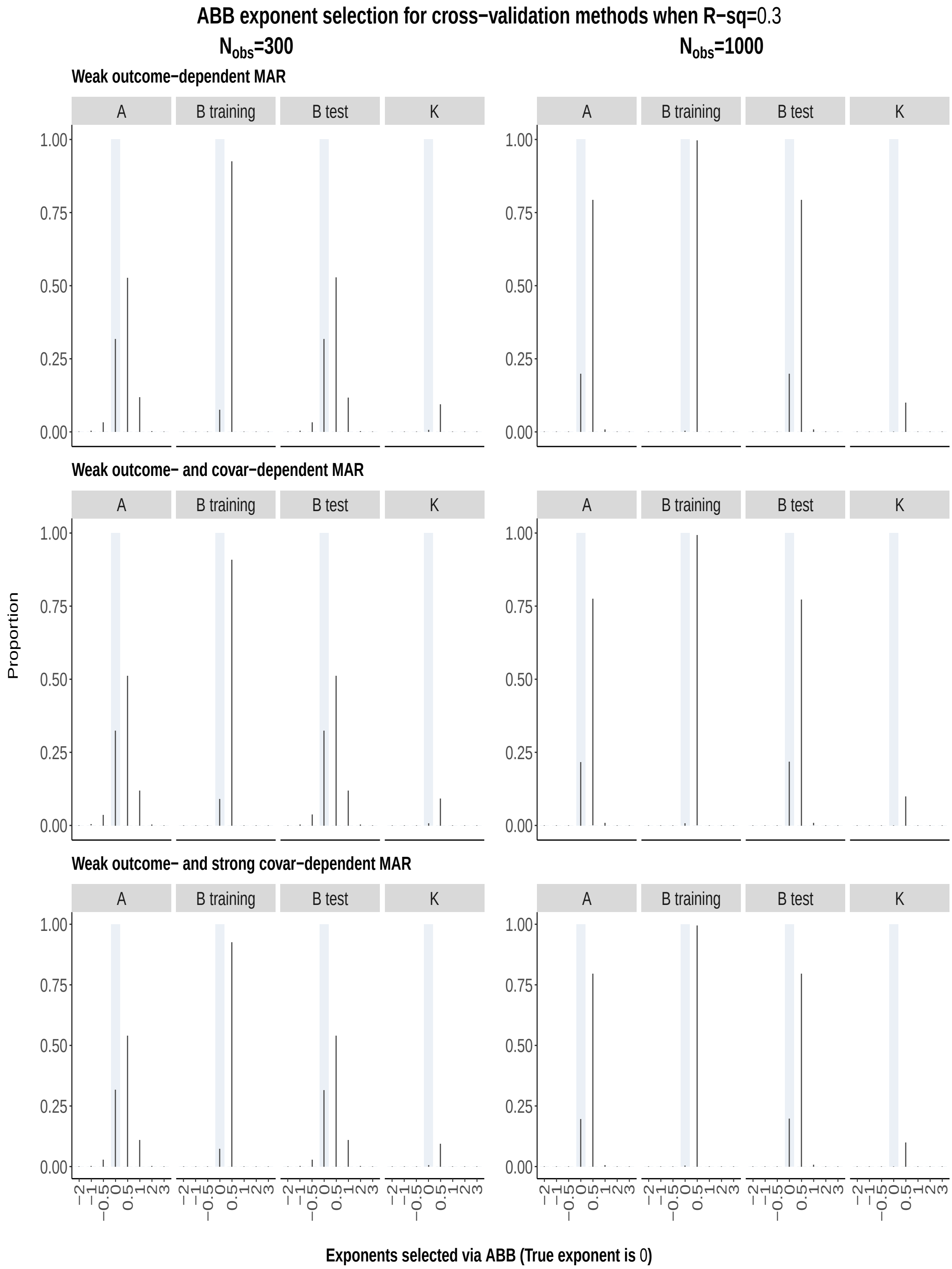

Figure S16: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

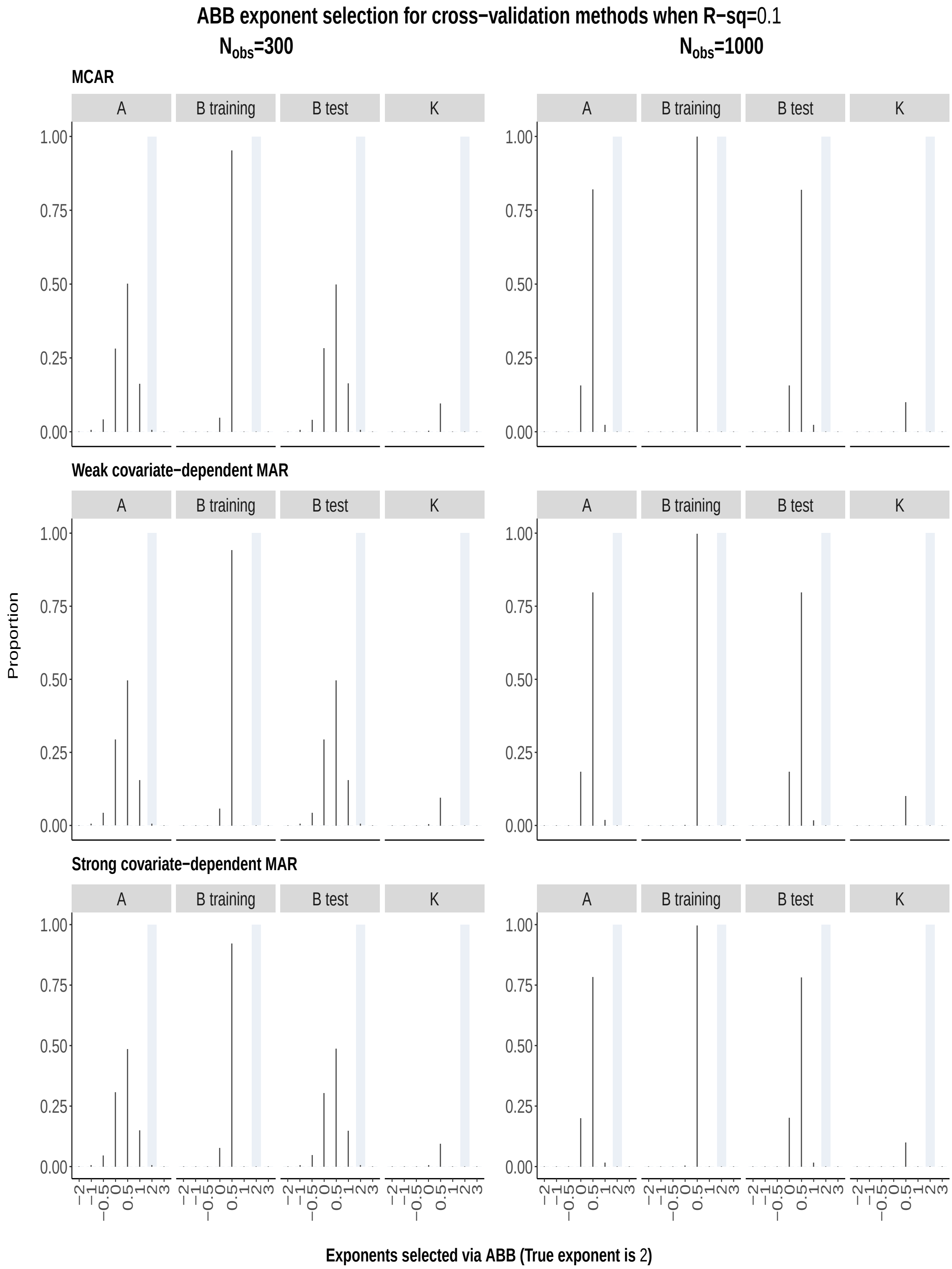

Figure S17: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

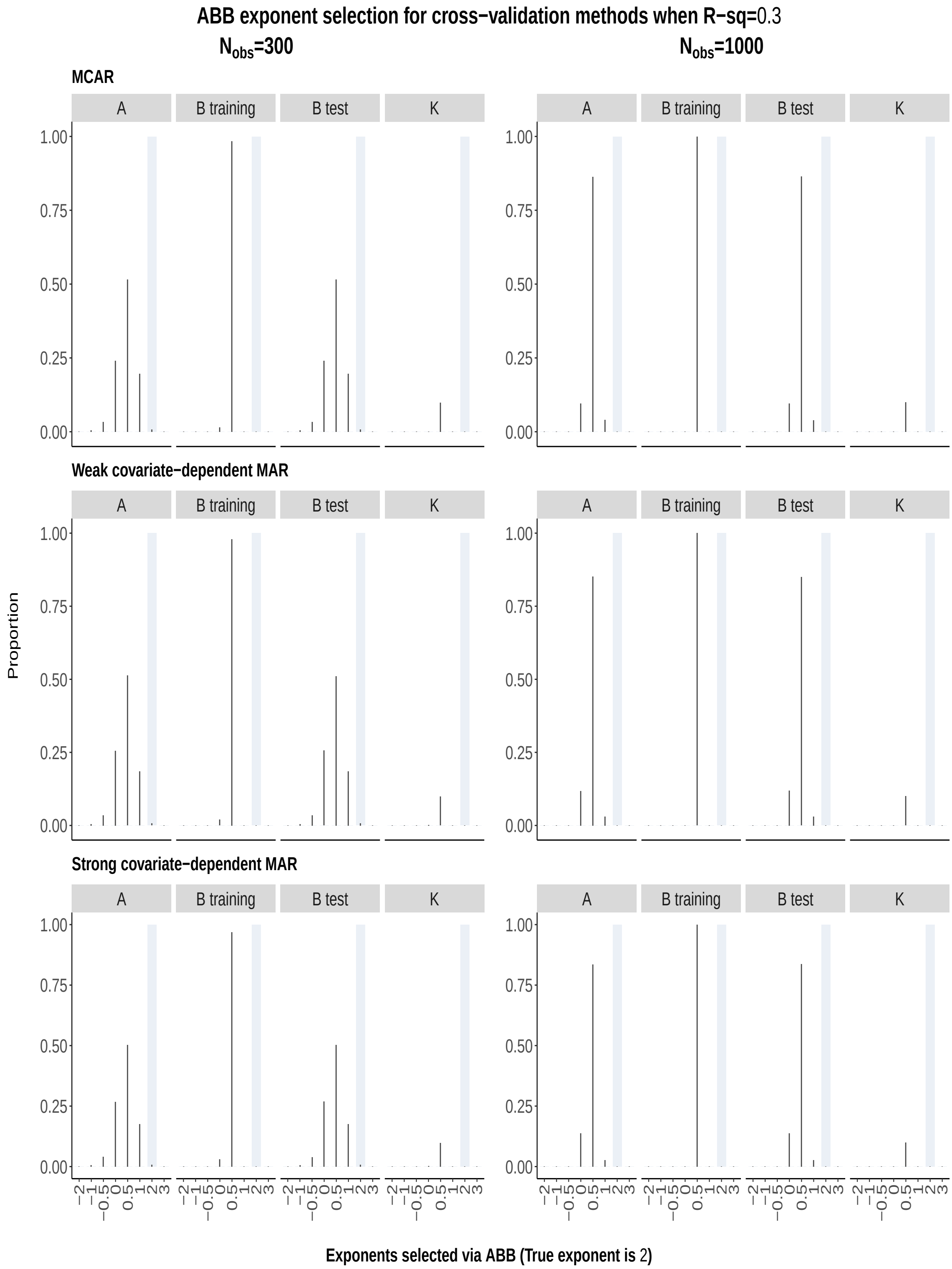

Figure S18: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

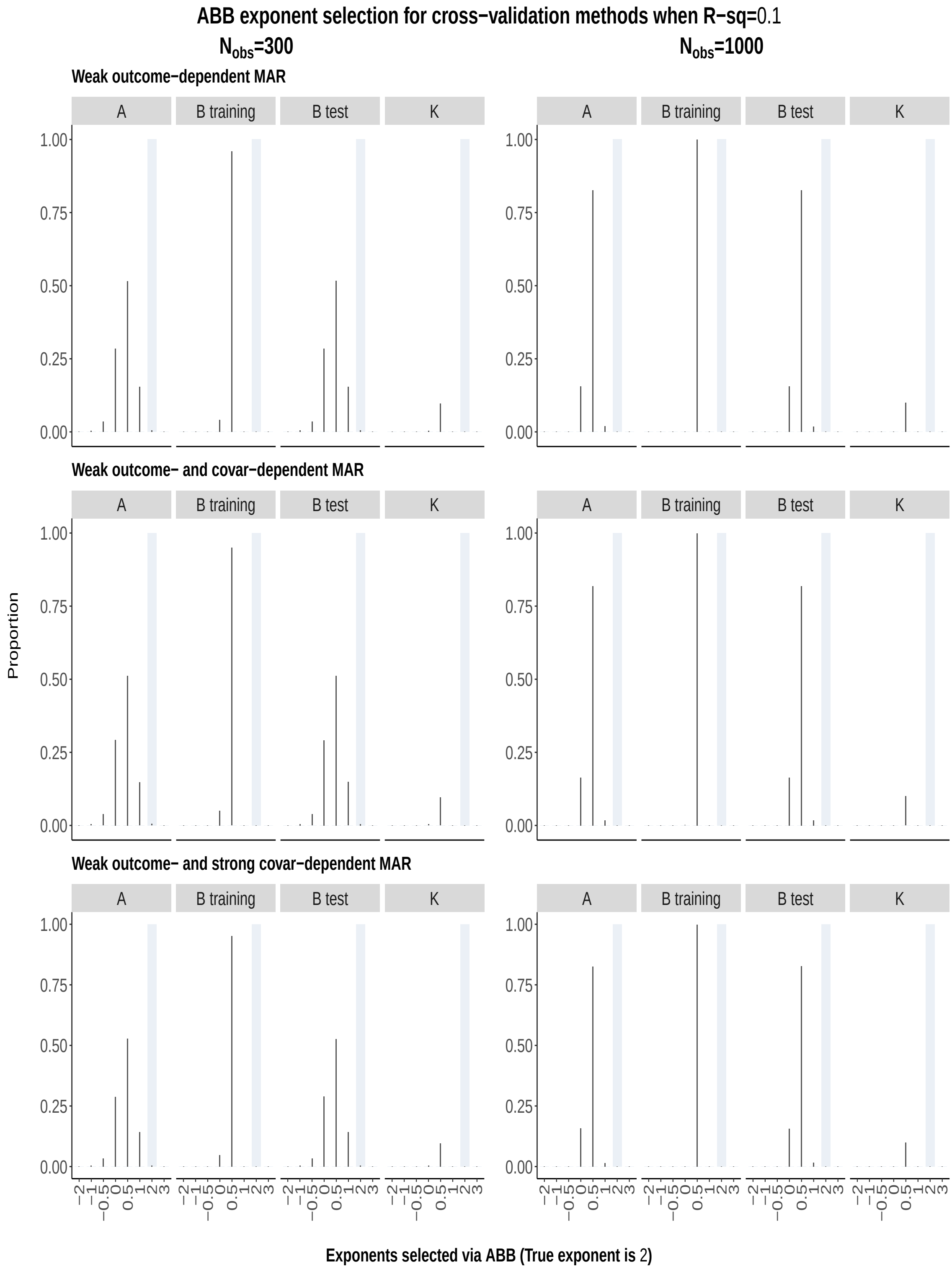

Figure S19: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

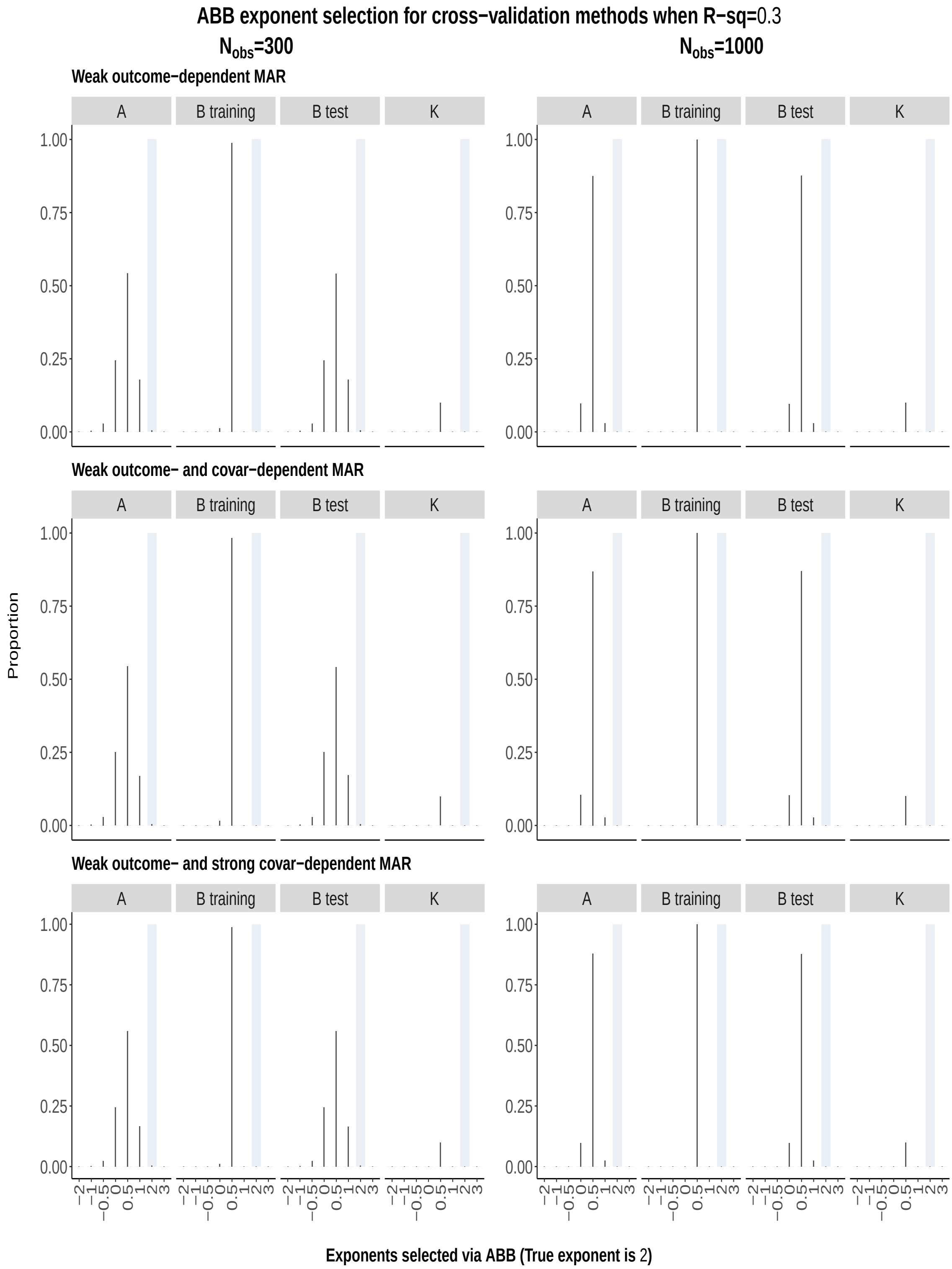

Figure S20: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

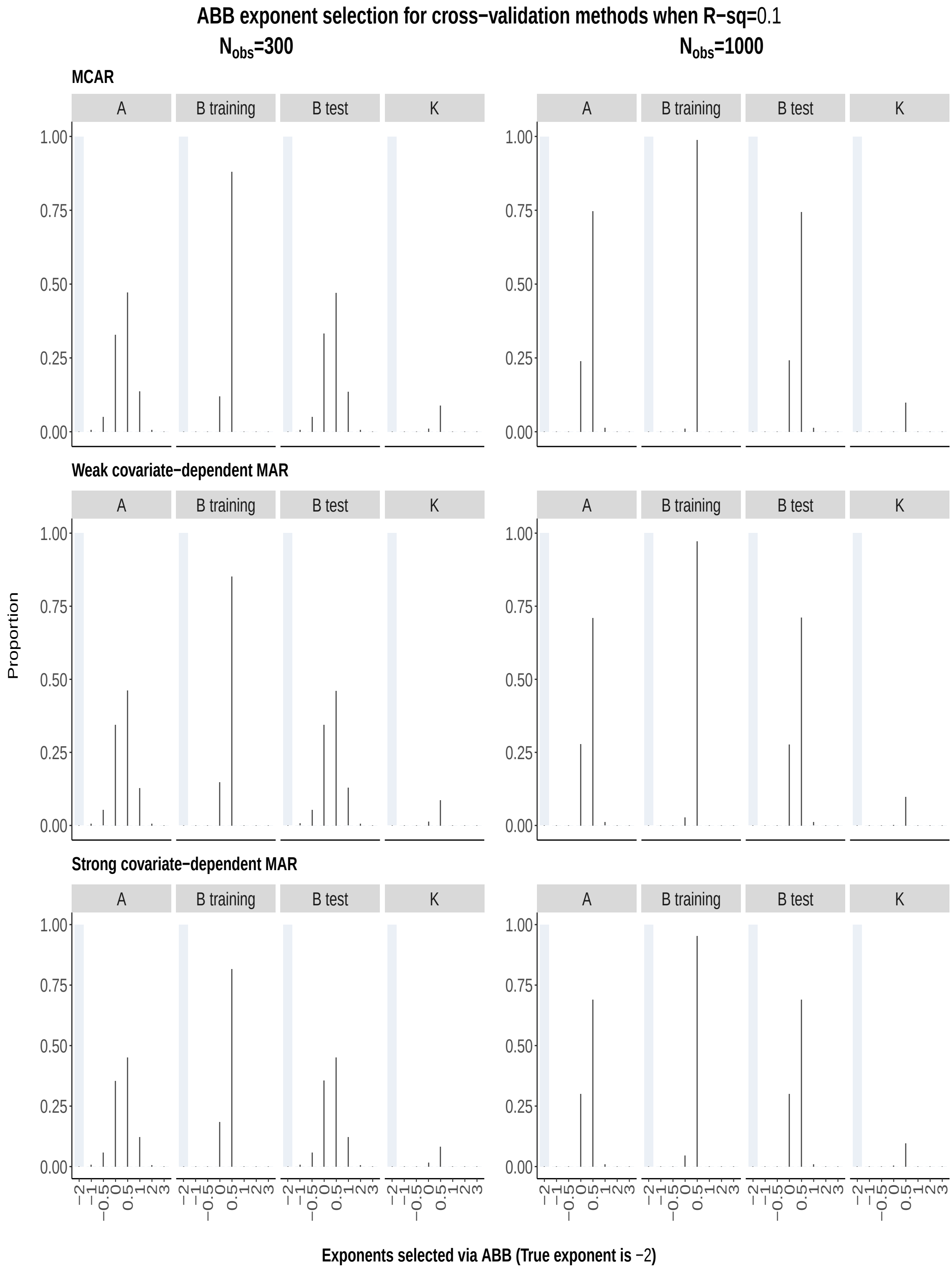

Figure S21: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

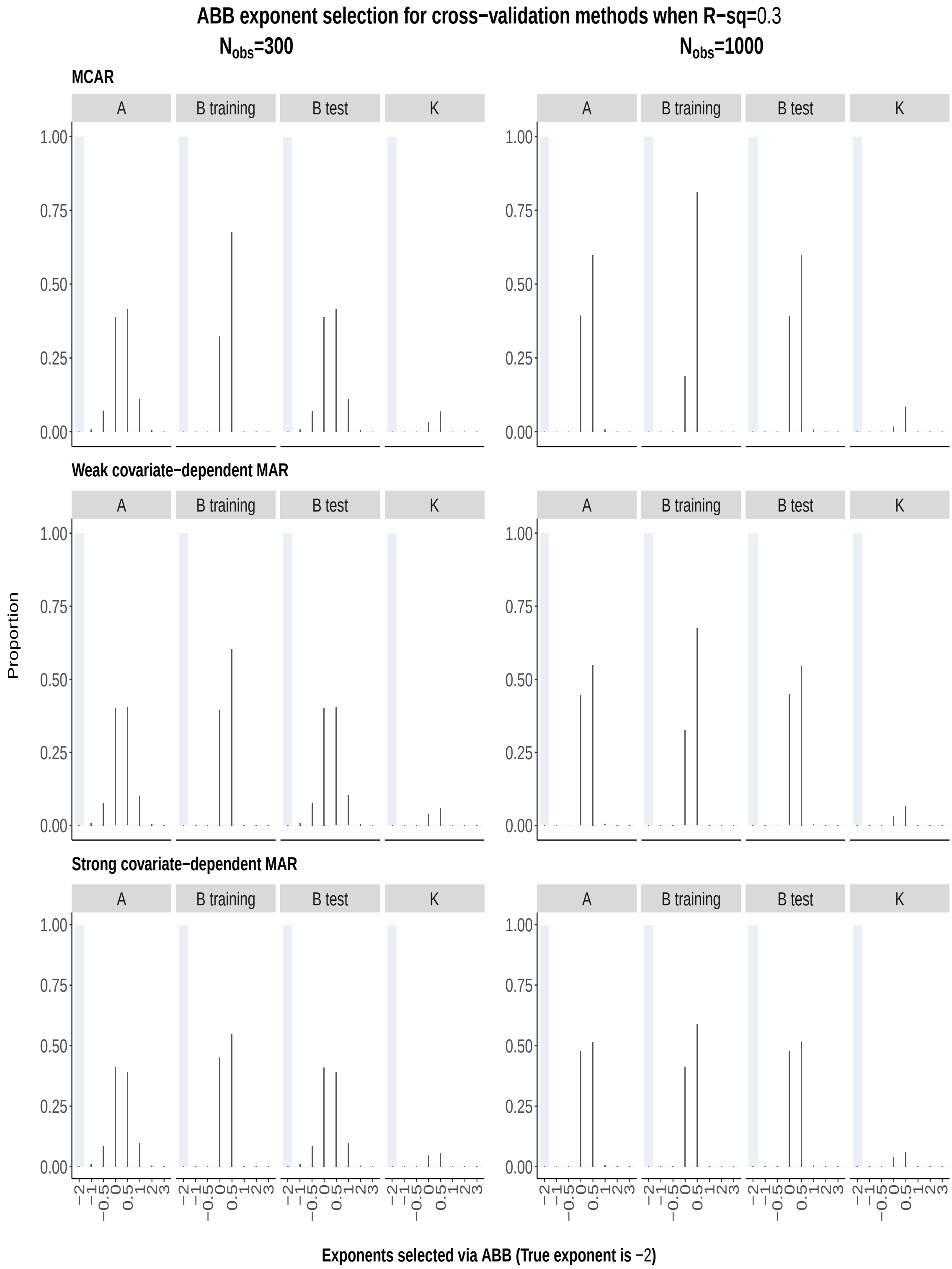

Figure S22: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

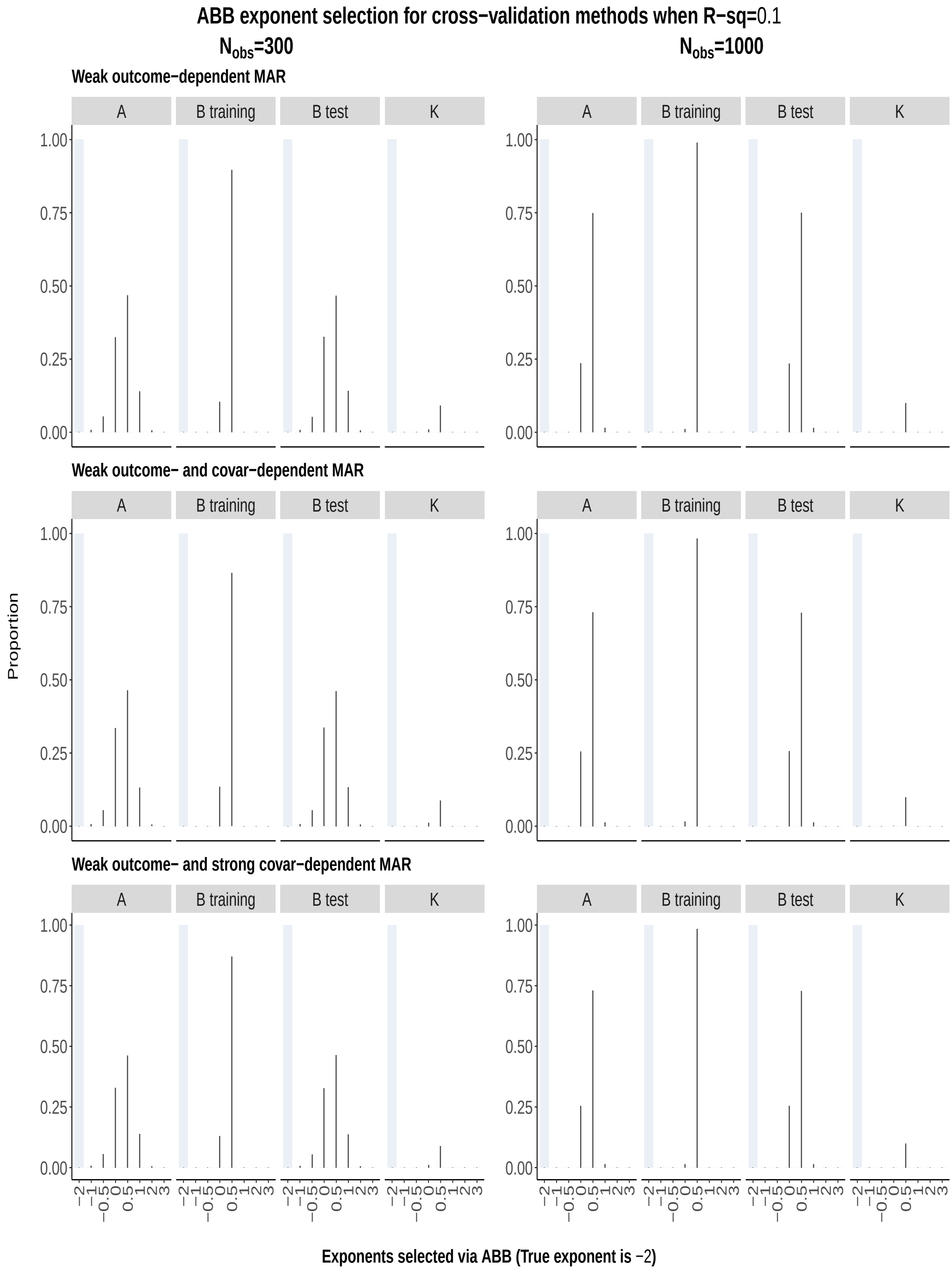

Figure S23: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

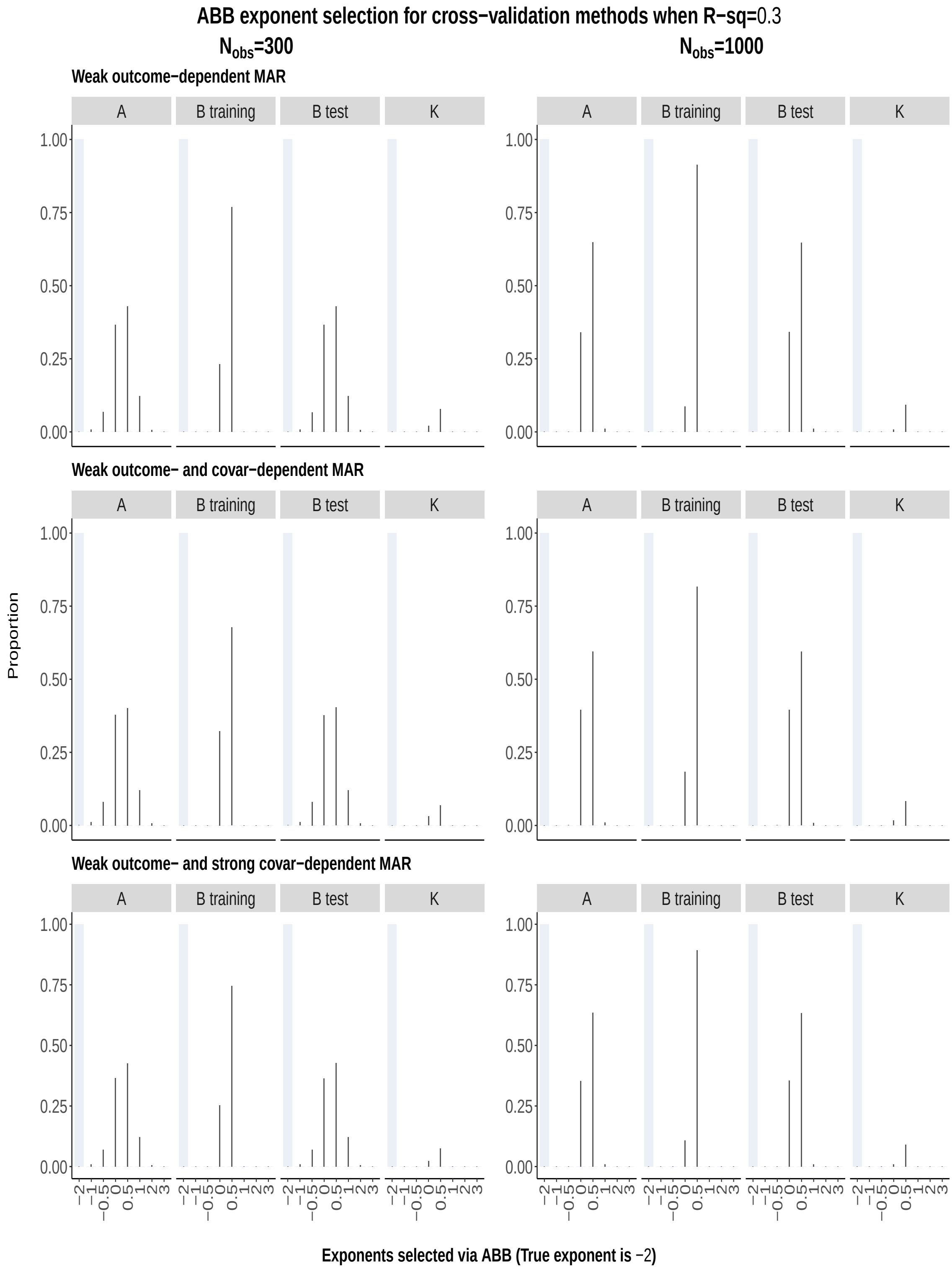

Figure S24: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.1.3 The 0.632 bootstrap, $\beta_2 = 1$

True exponent is 0

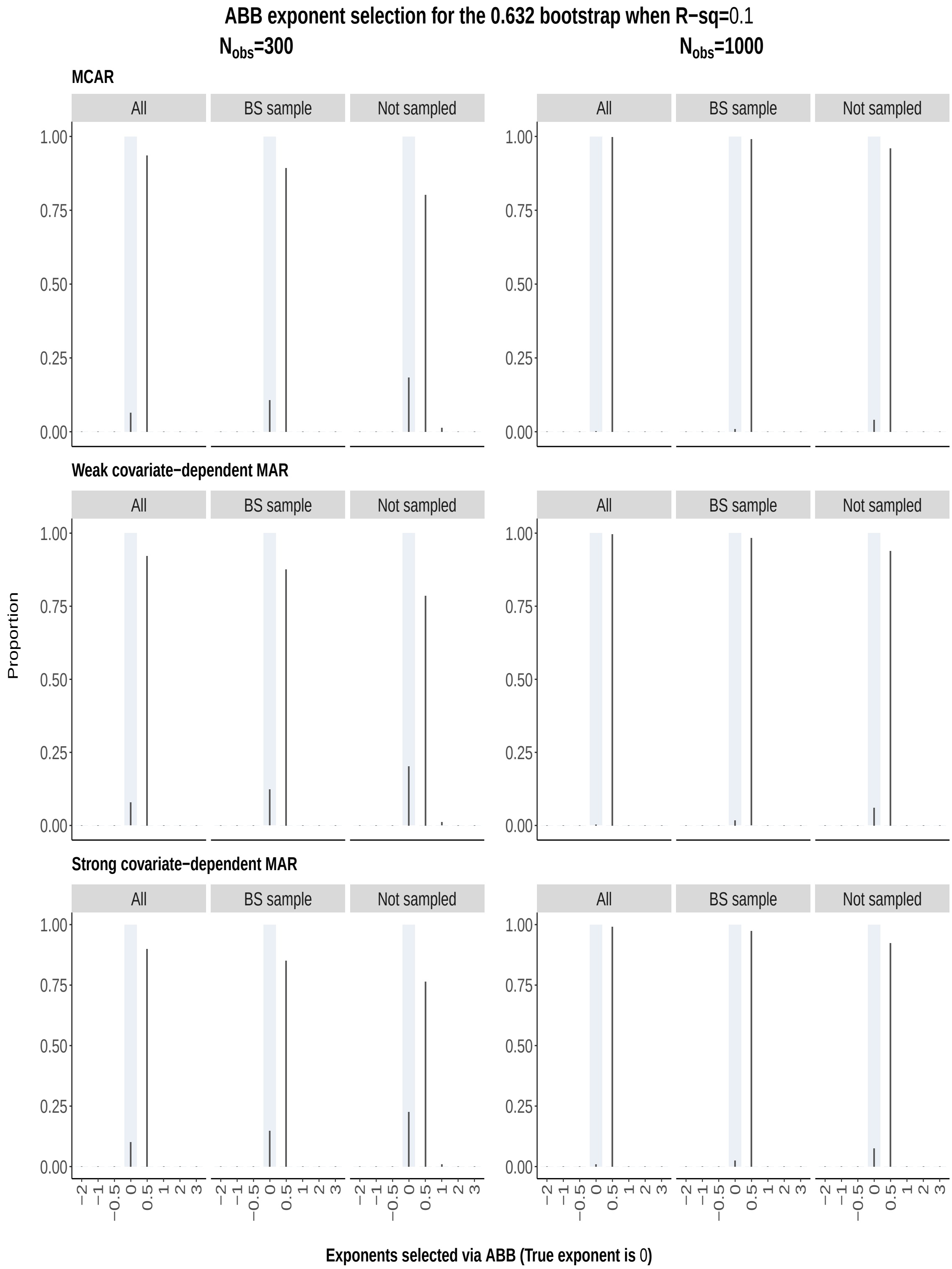

Figure S25: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

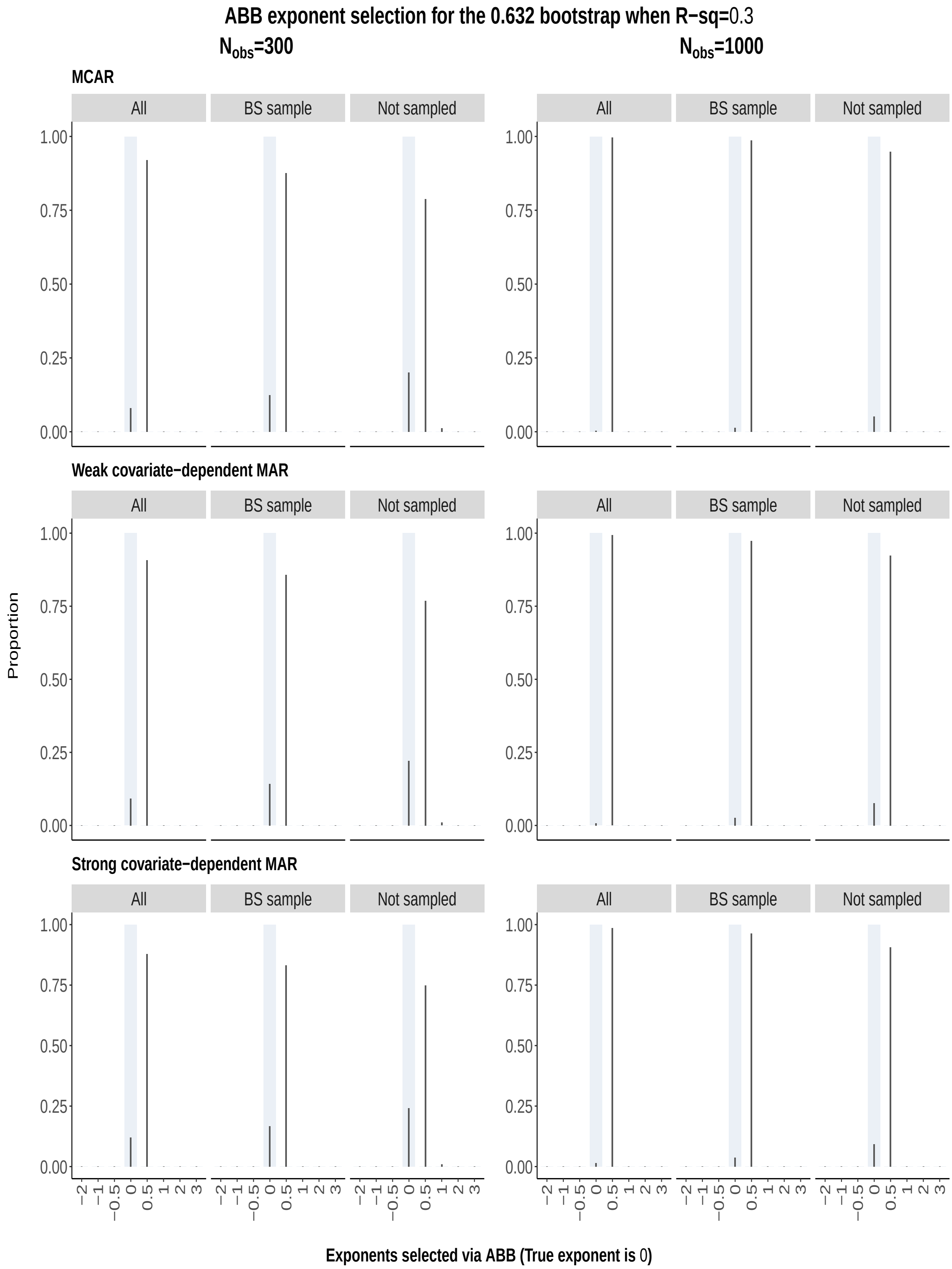

Figure S26: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

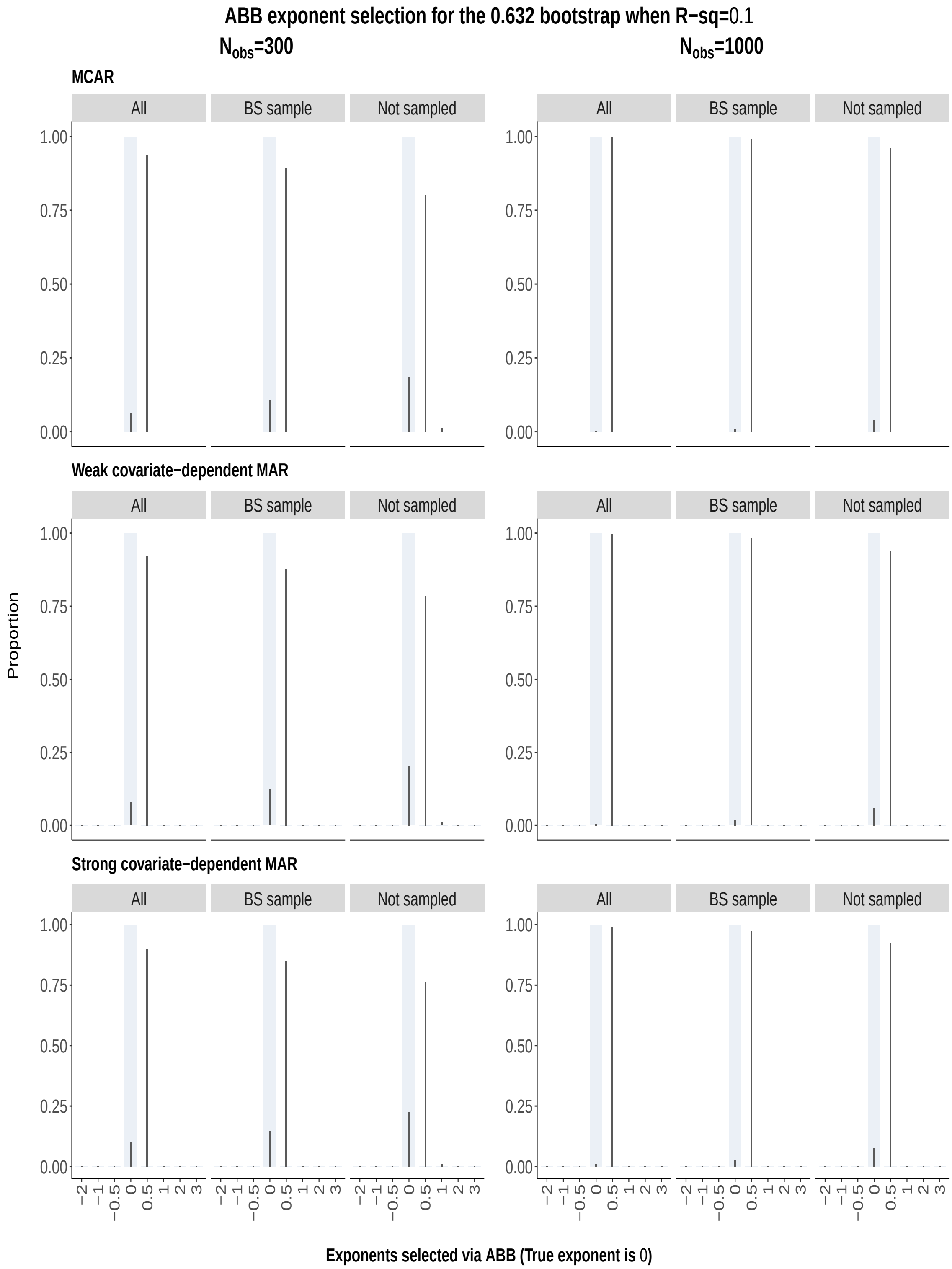

Figure S27: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

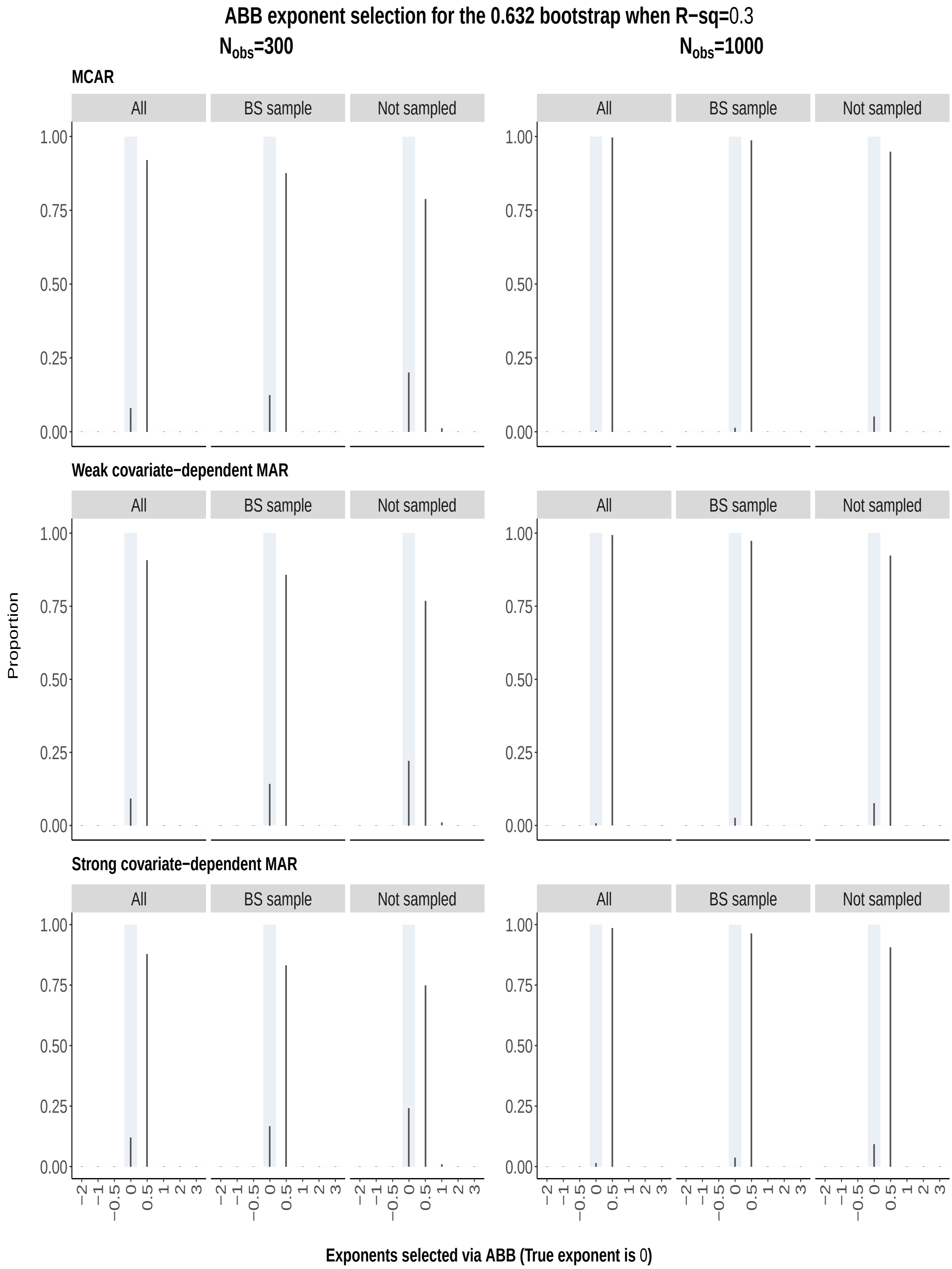

Figure S28: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

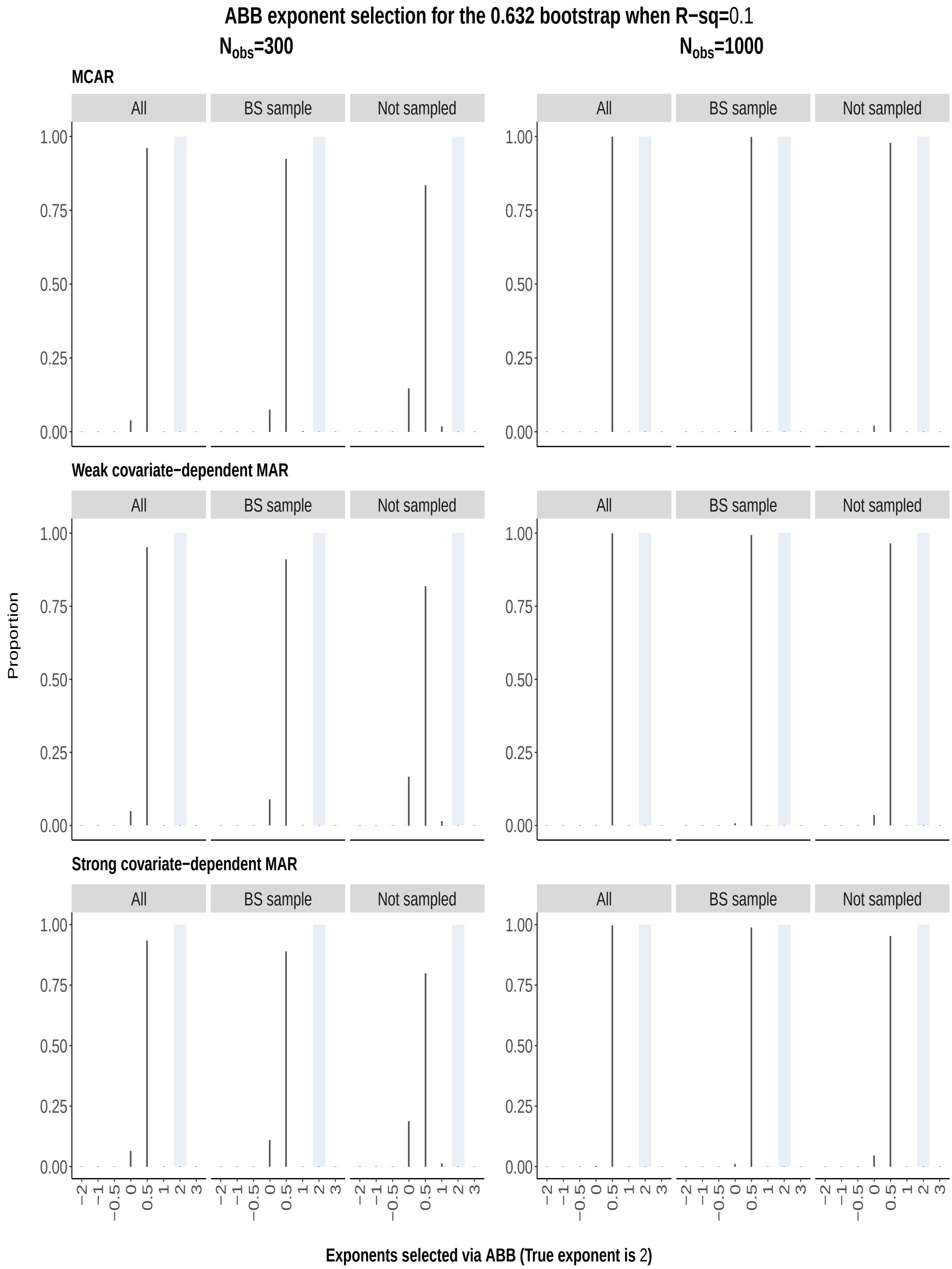

Figure S29: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

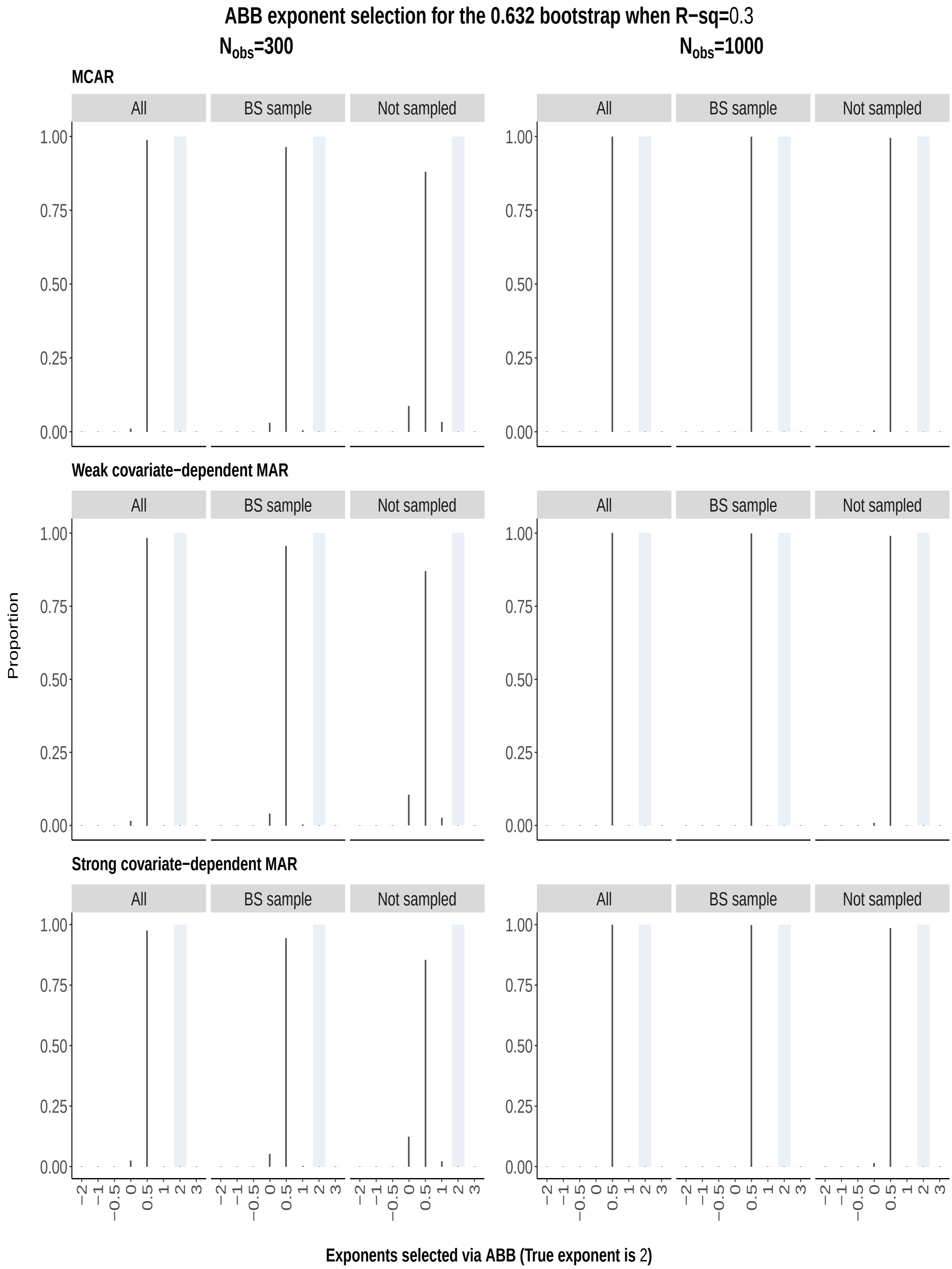

Figure S30: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

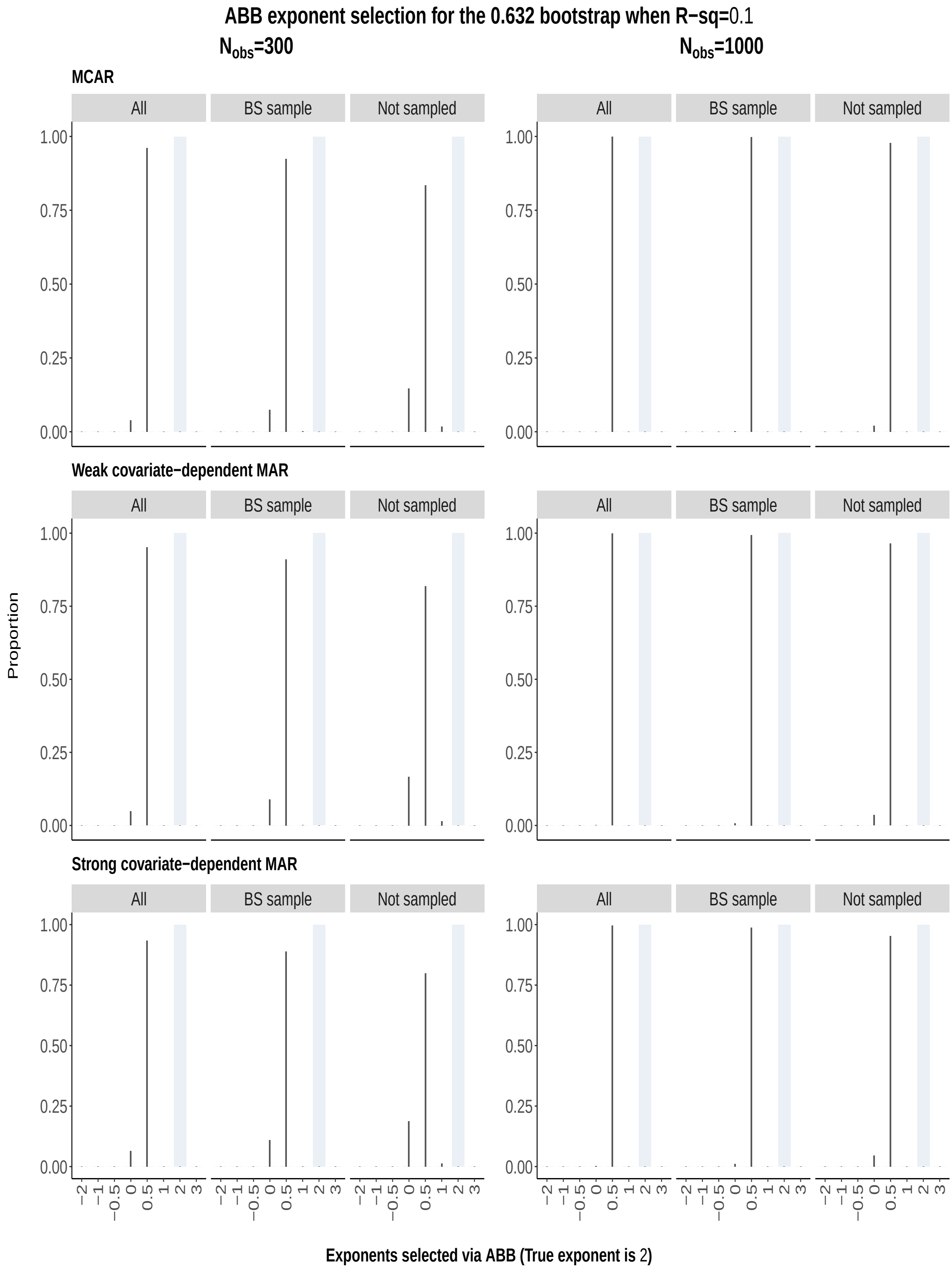

Figure S31: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

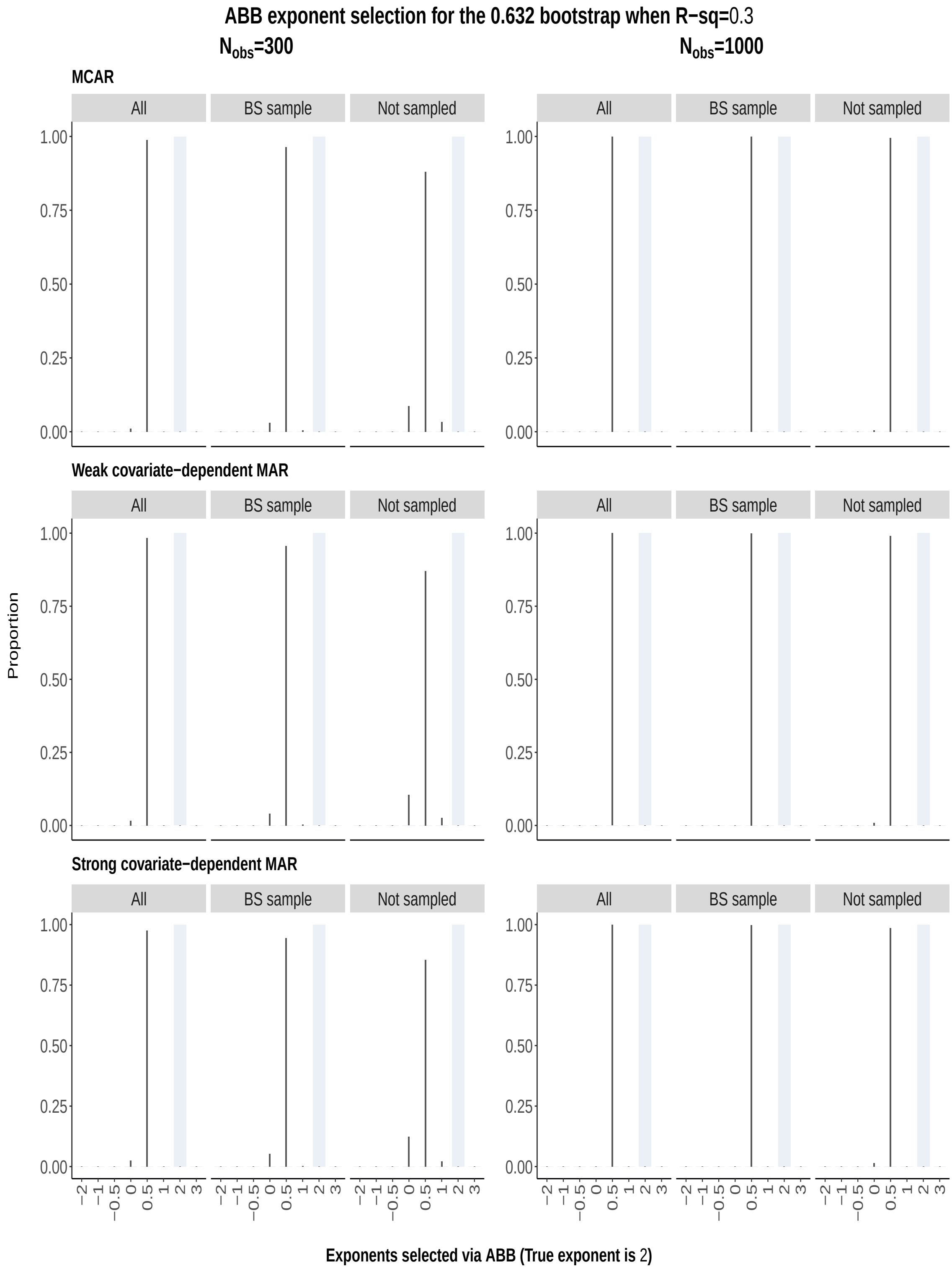

Figure S32: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

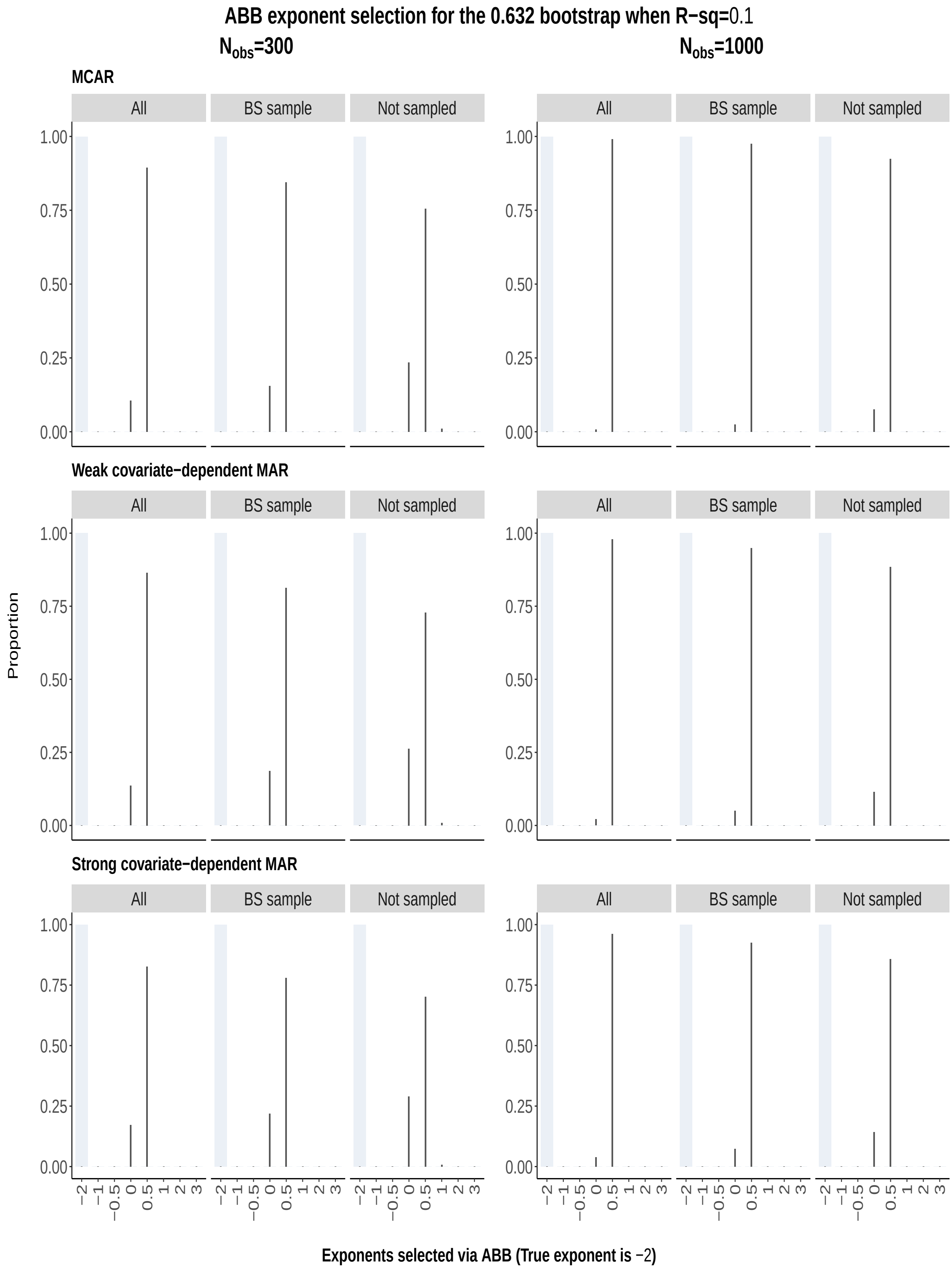

Figure S33: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

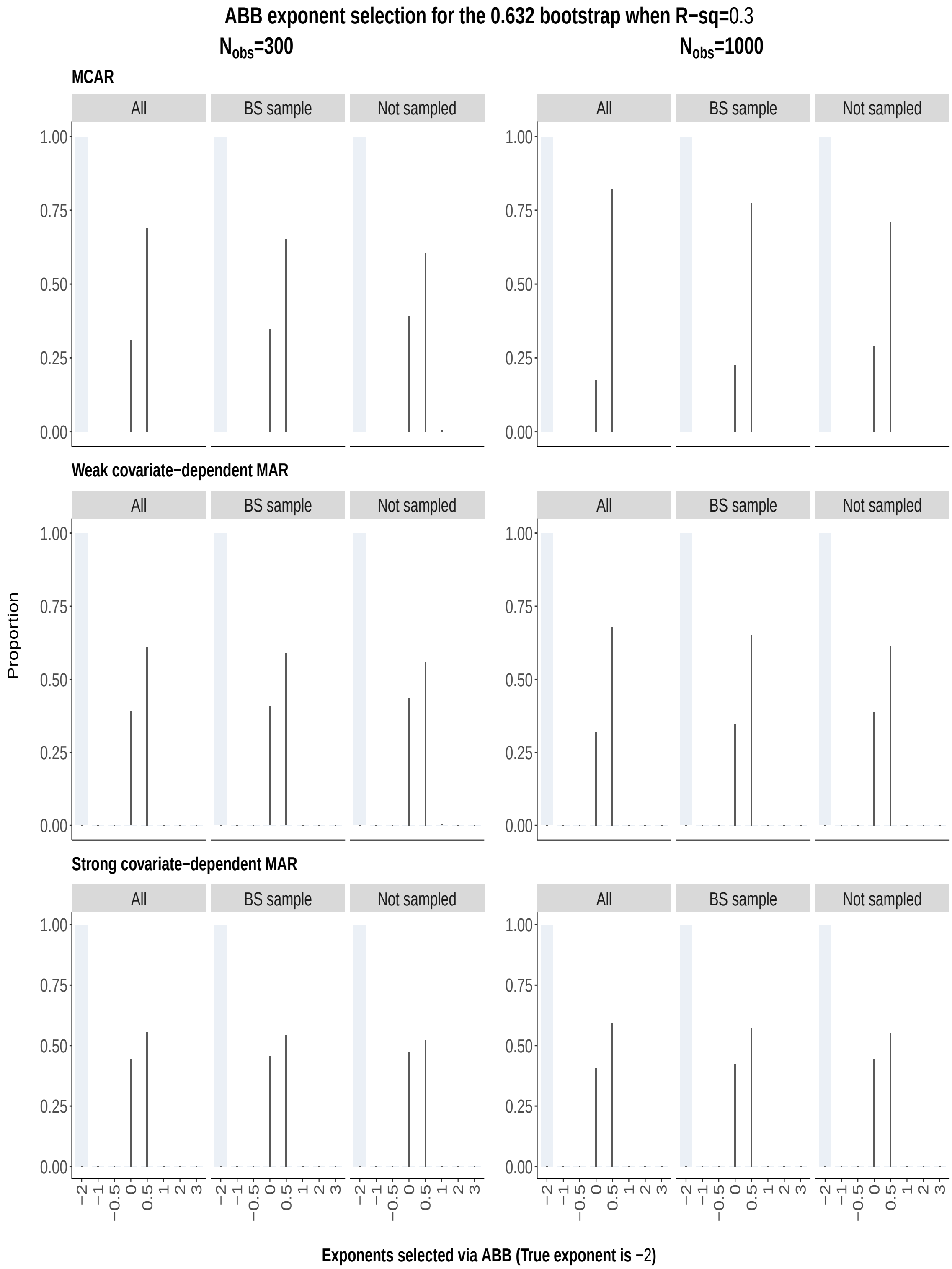

Figure S34: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

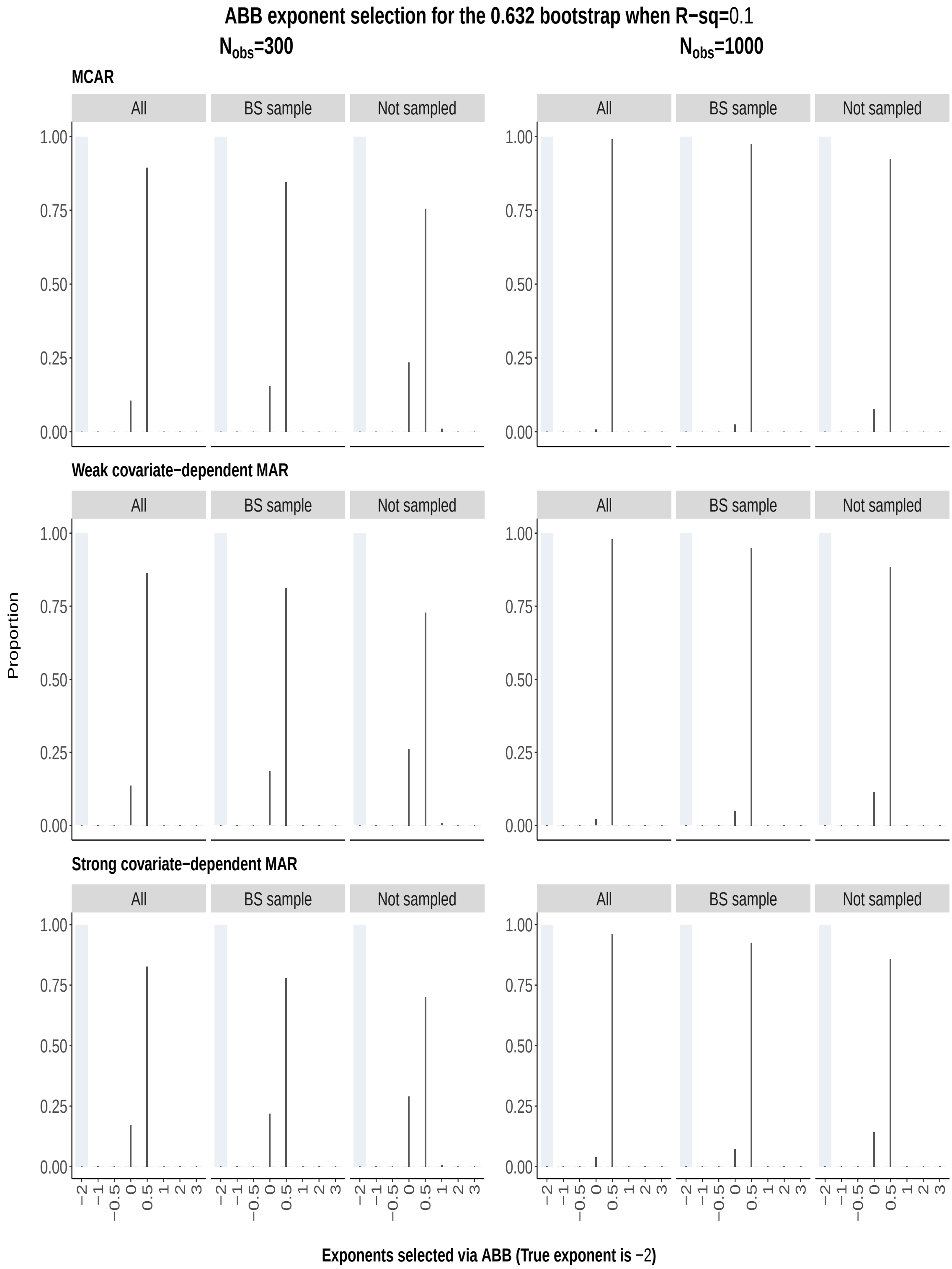

Figure S35: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

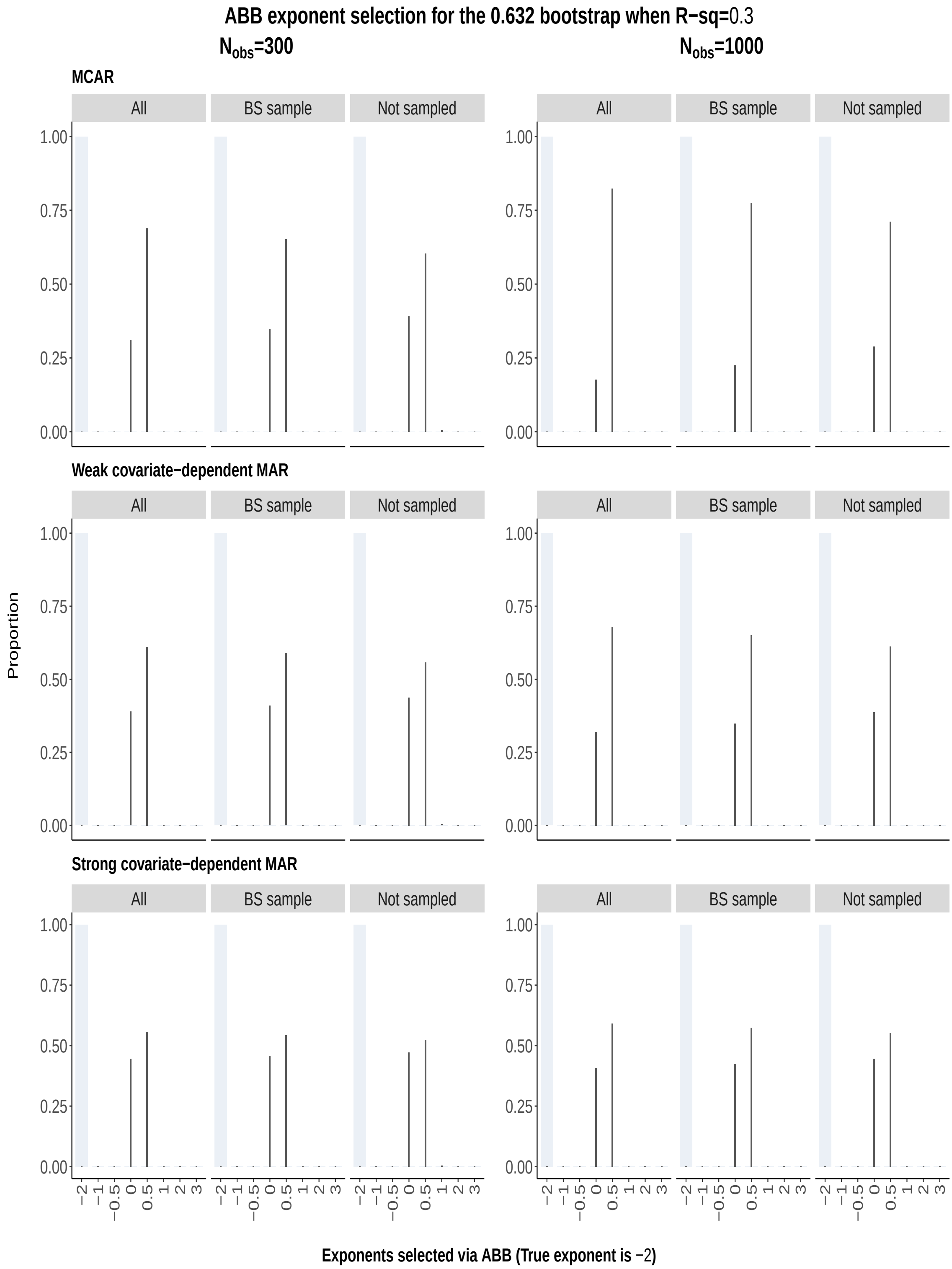

Figure S36: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.1.4 The 0.632 bootstrap, $\beta_2 = 0$

True exponent is 0

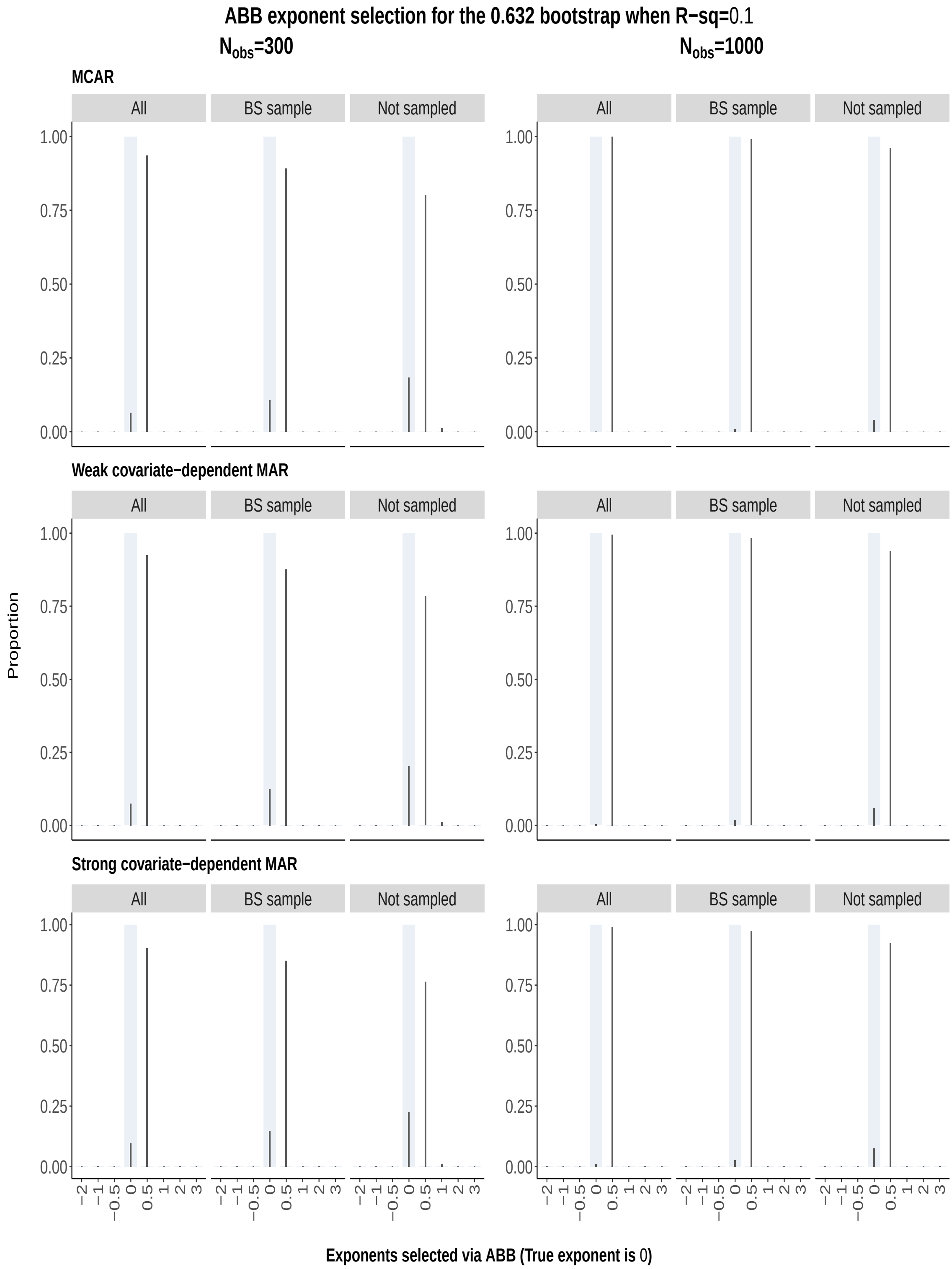

Figure S37: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

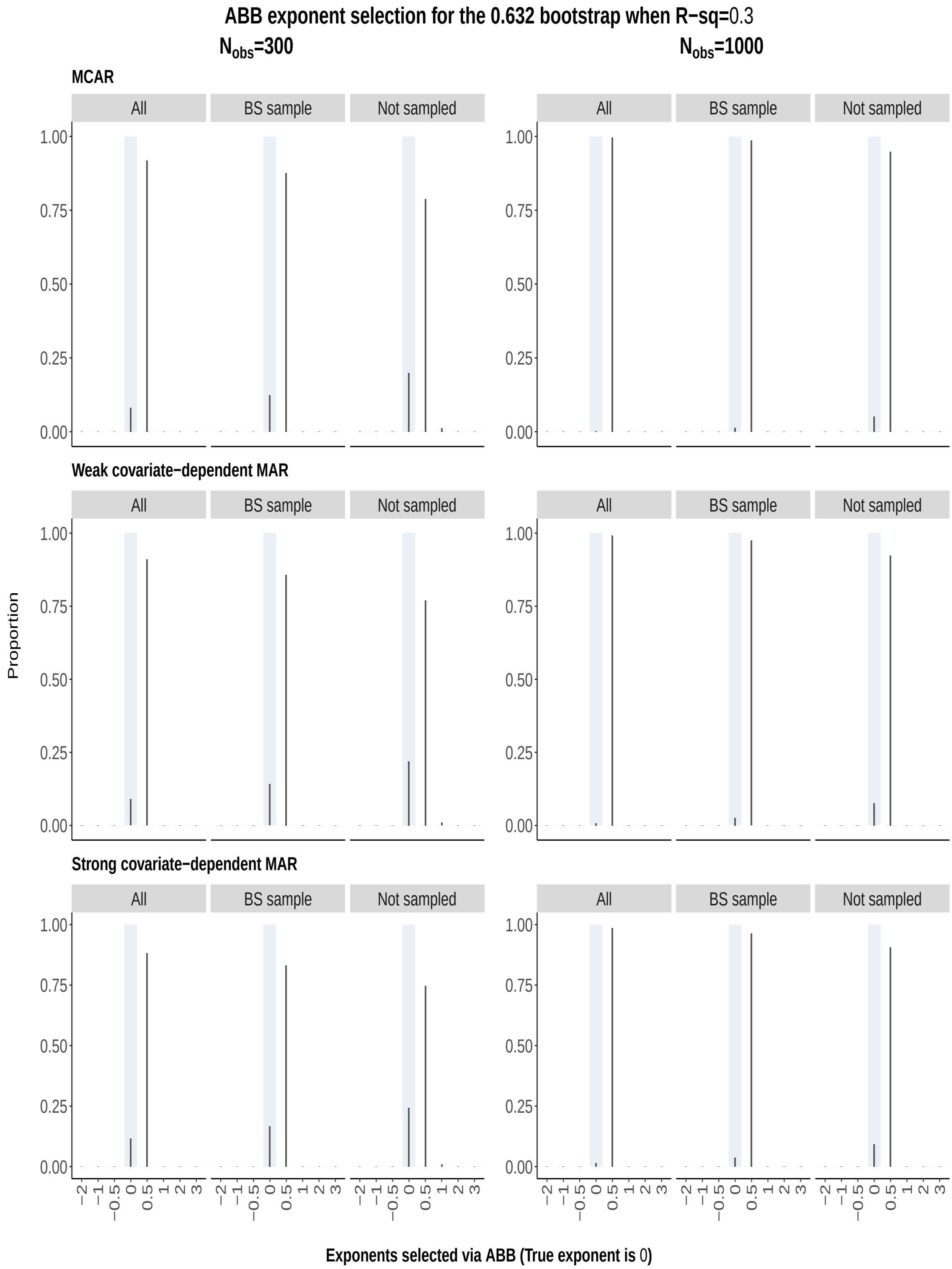

Figure S38: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

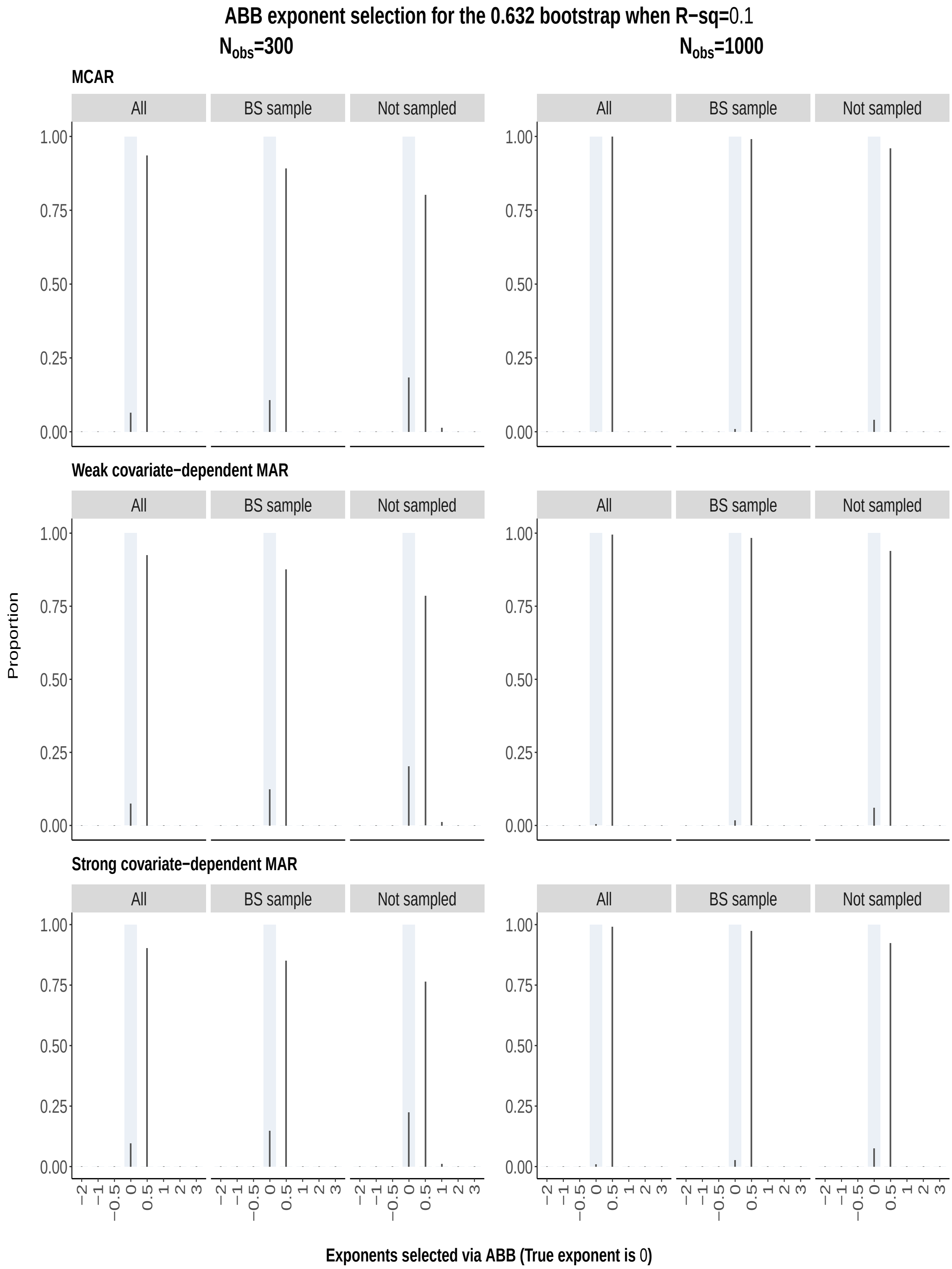

Figure S39: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

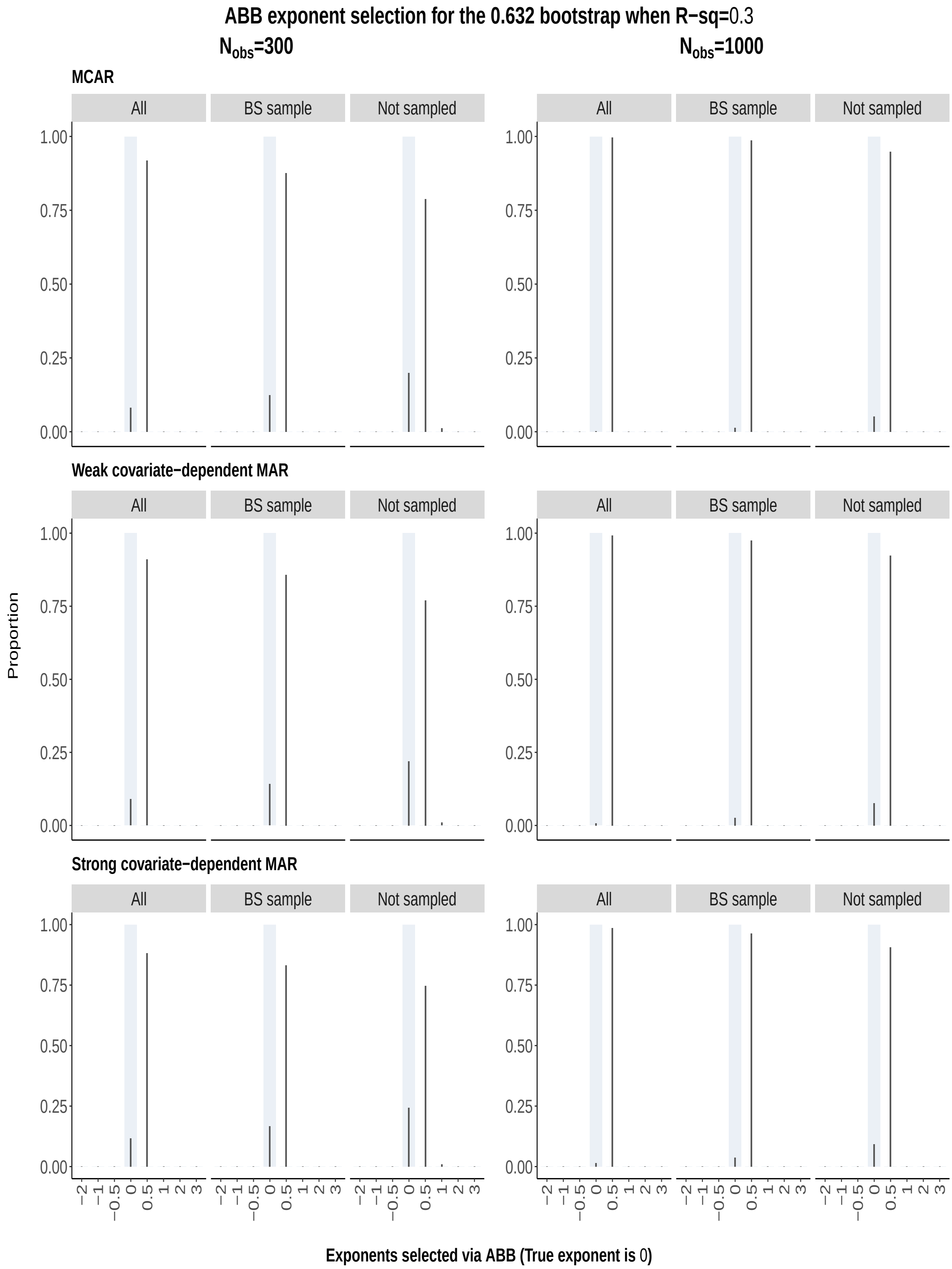

Figure S40: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

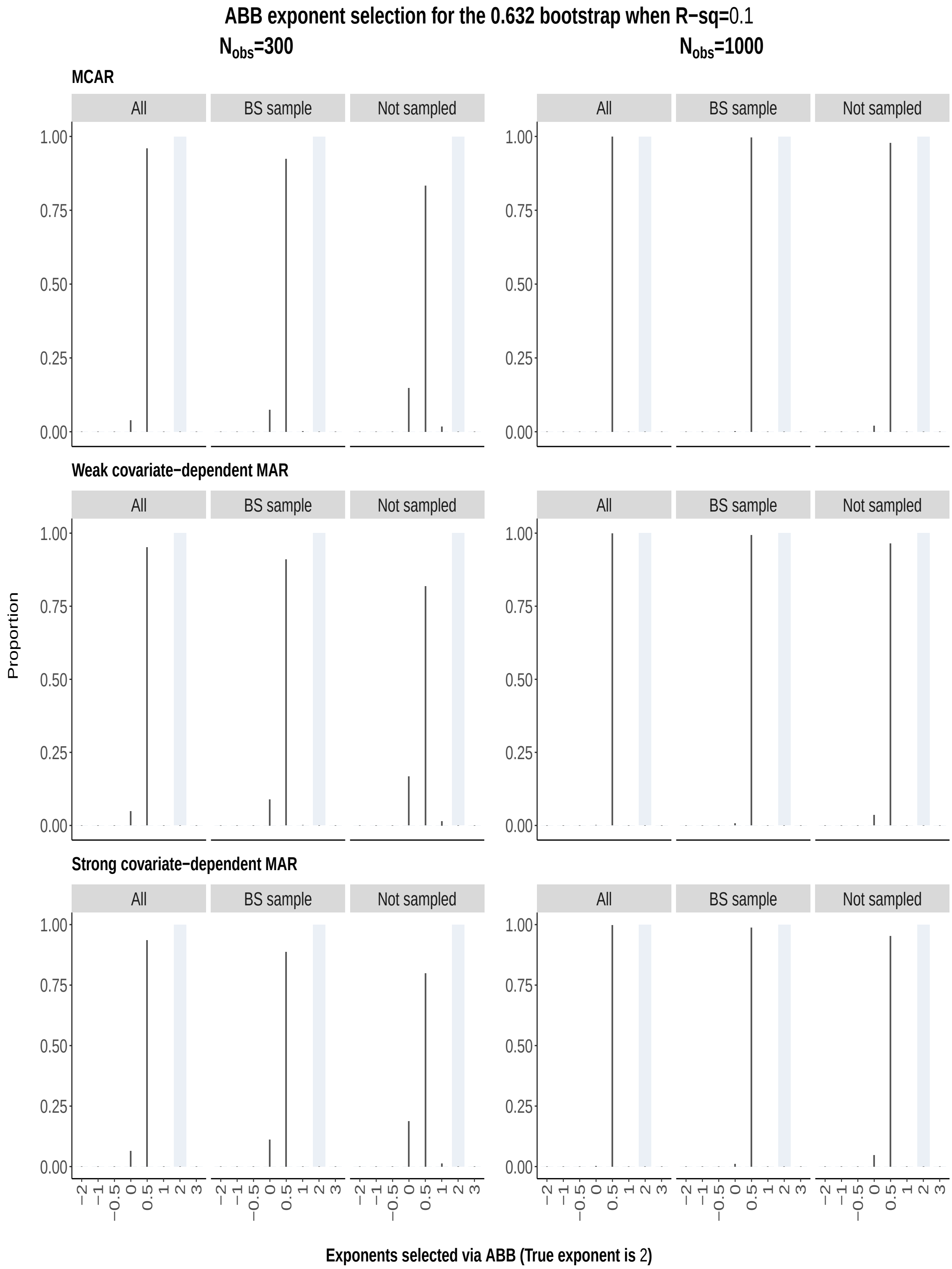

Figure S41: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

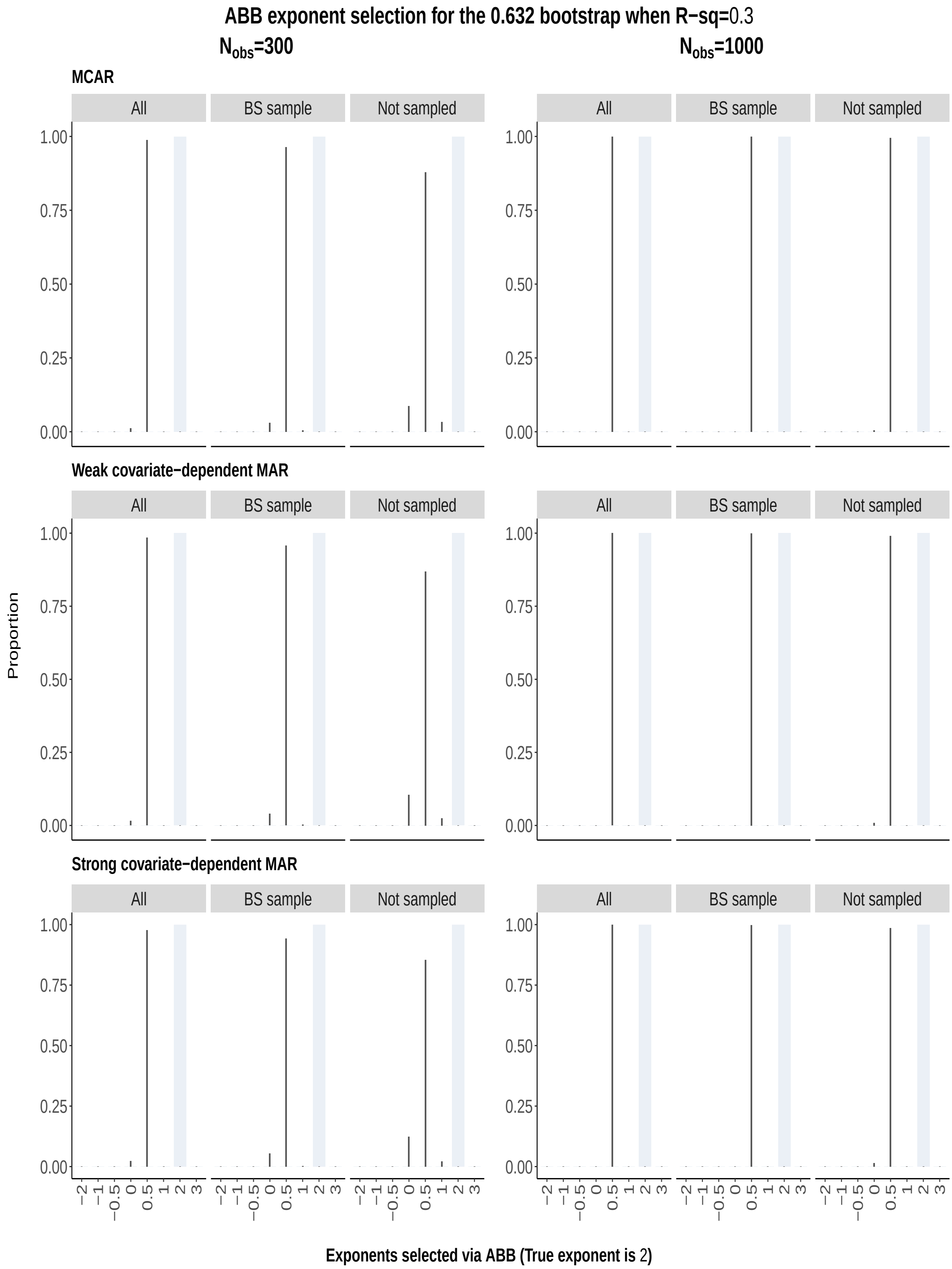

Figure S42: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

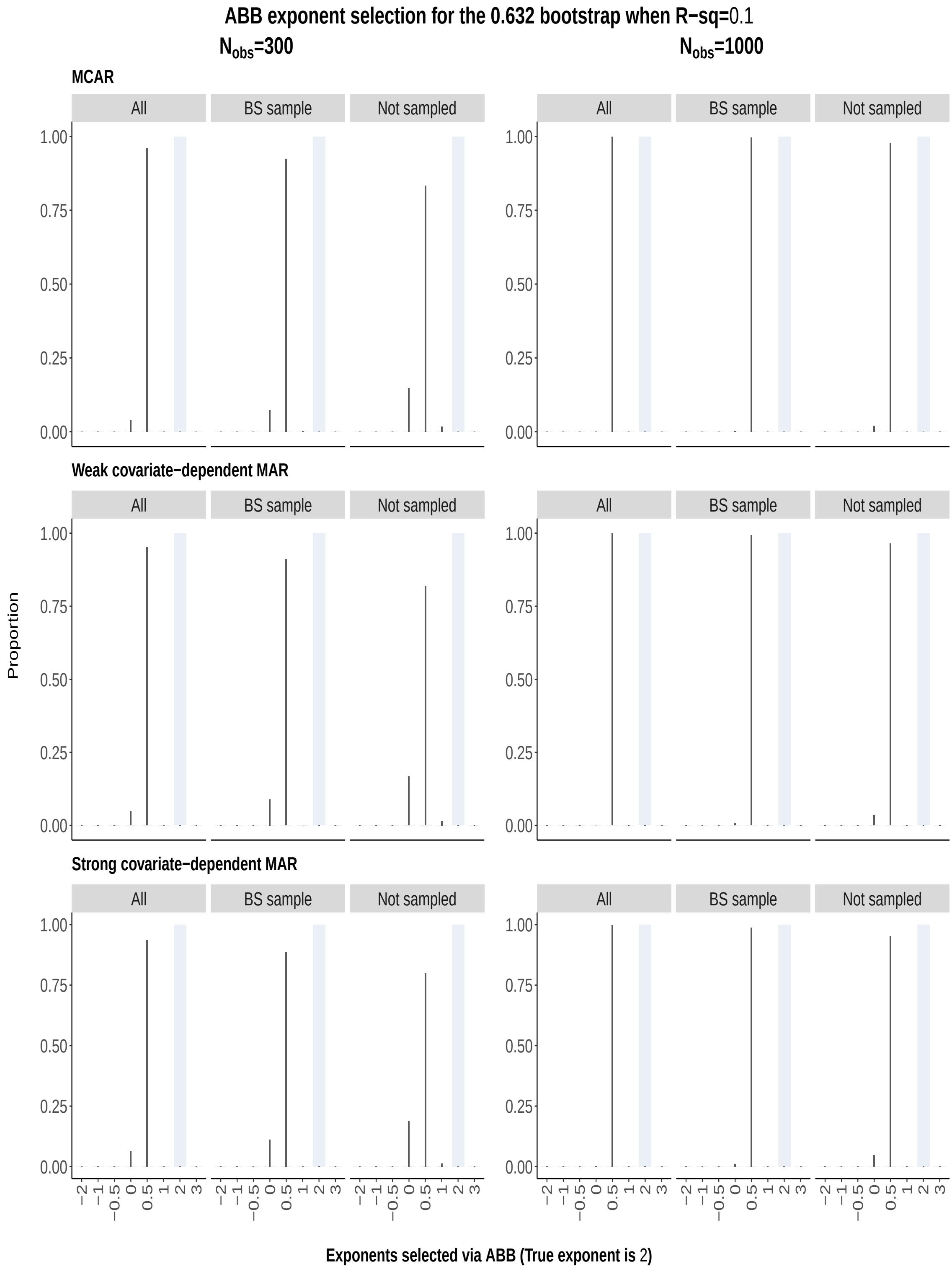

Figure S43: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

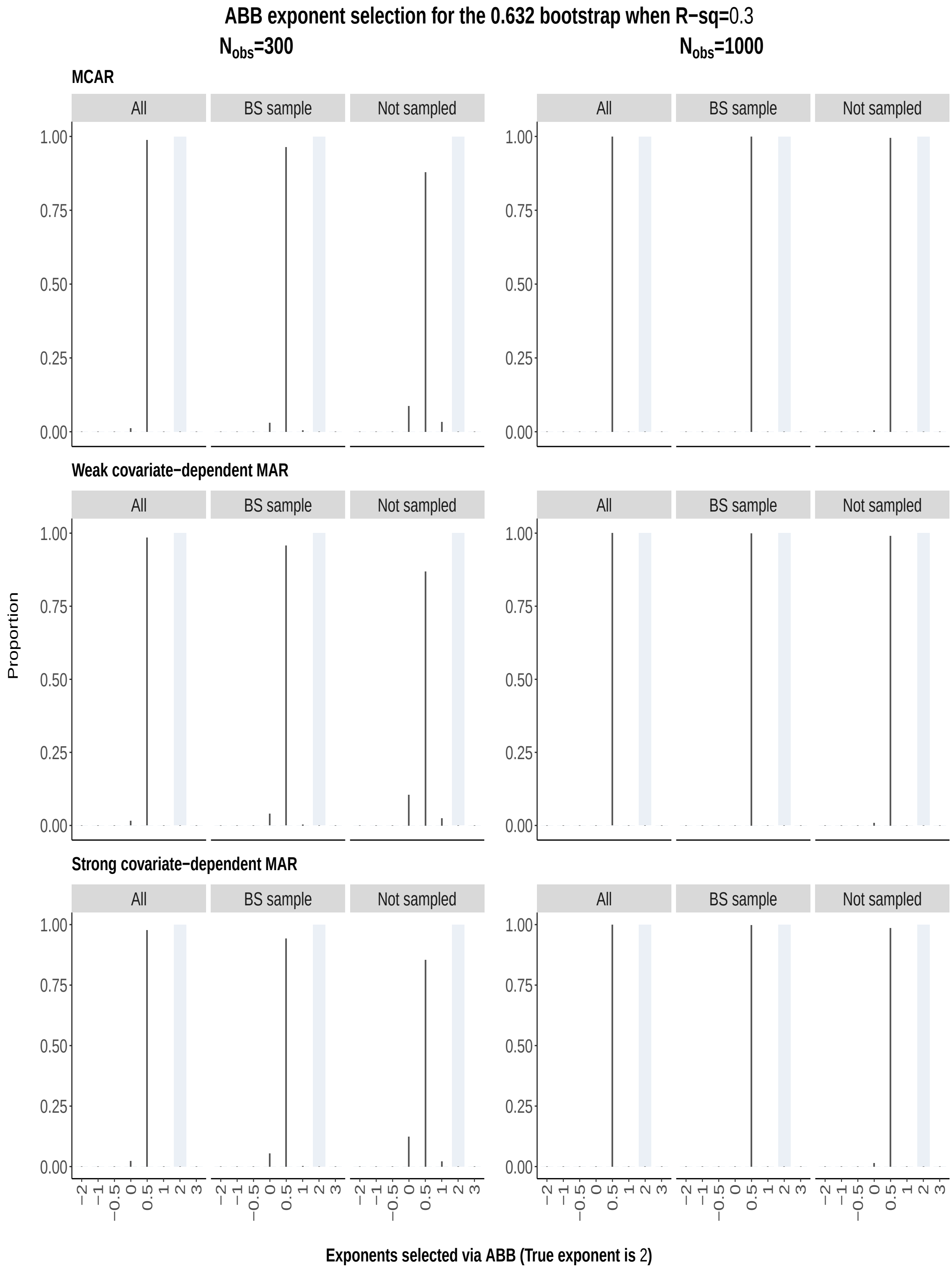

Figure S44: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

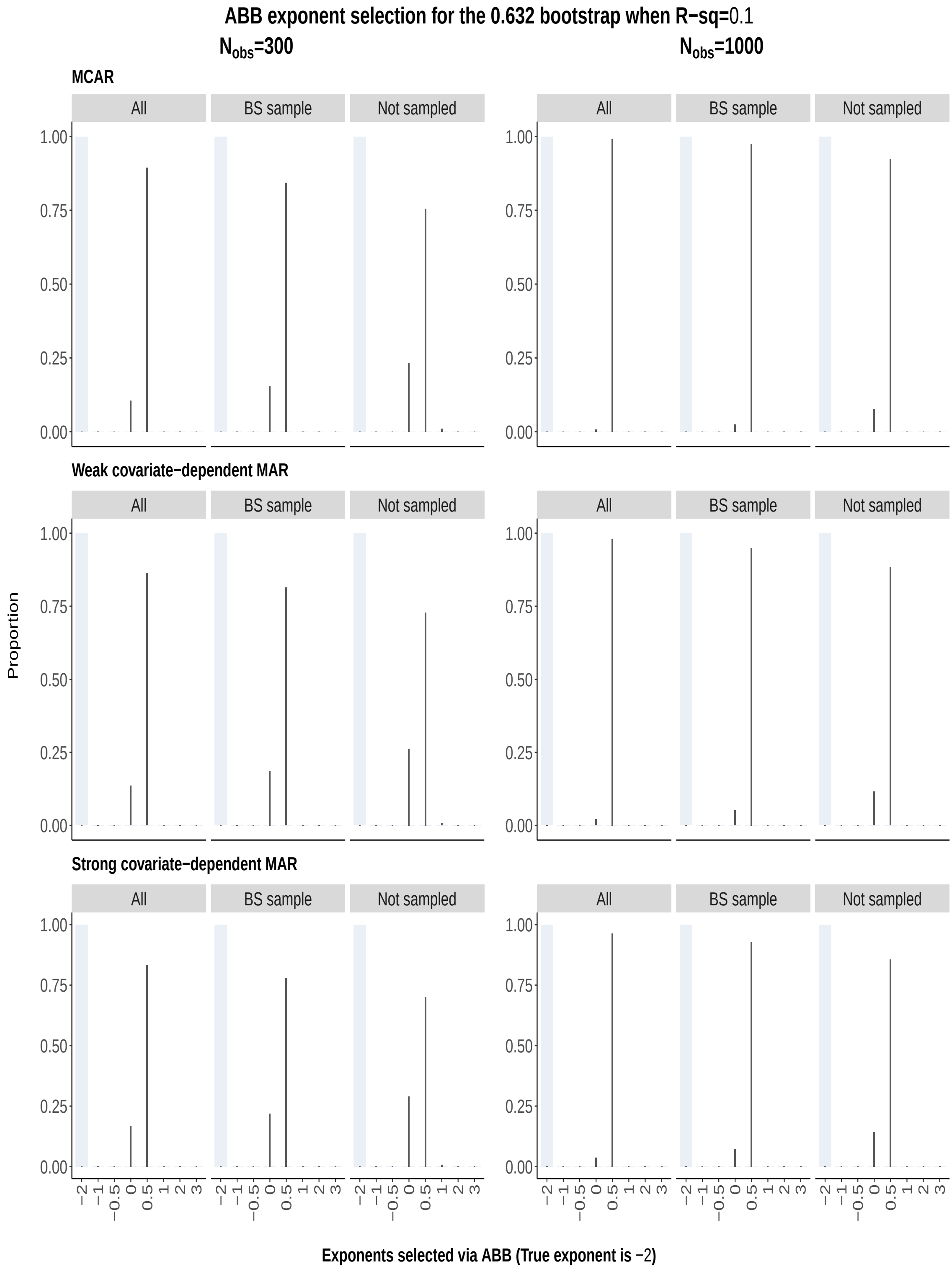

Figure S45: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

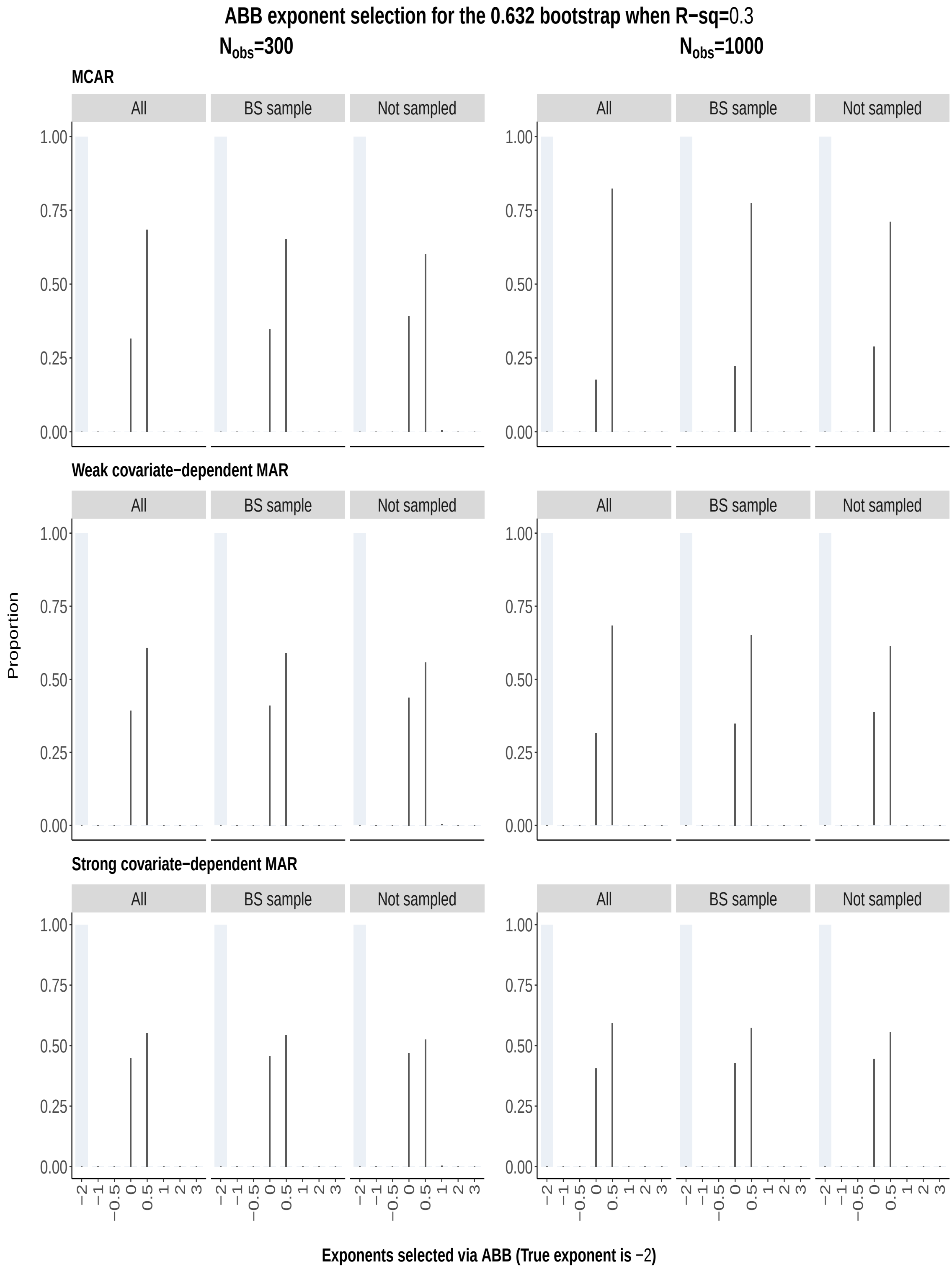

Figure S46: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

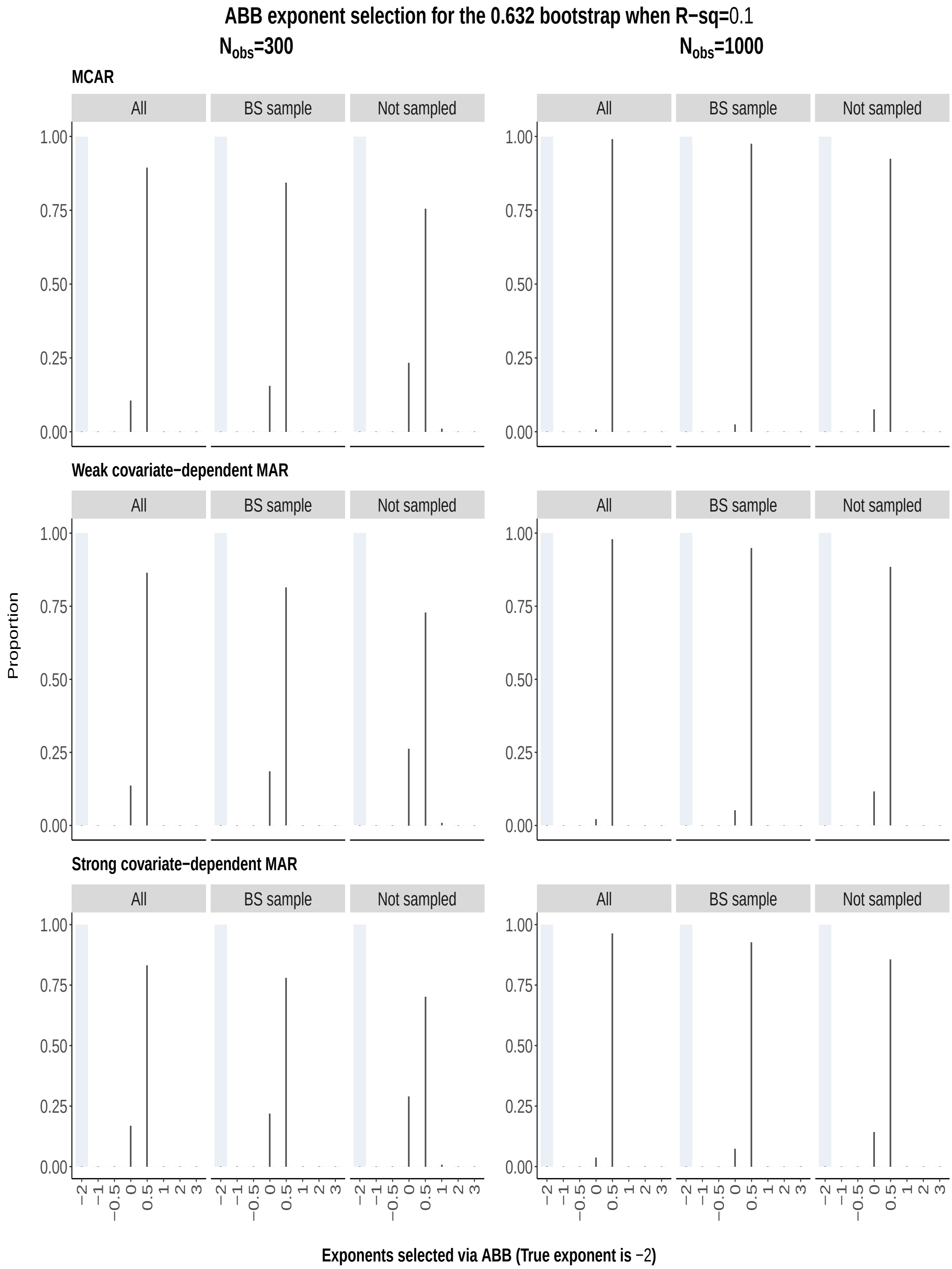

Figure S47: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

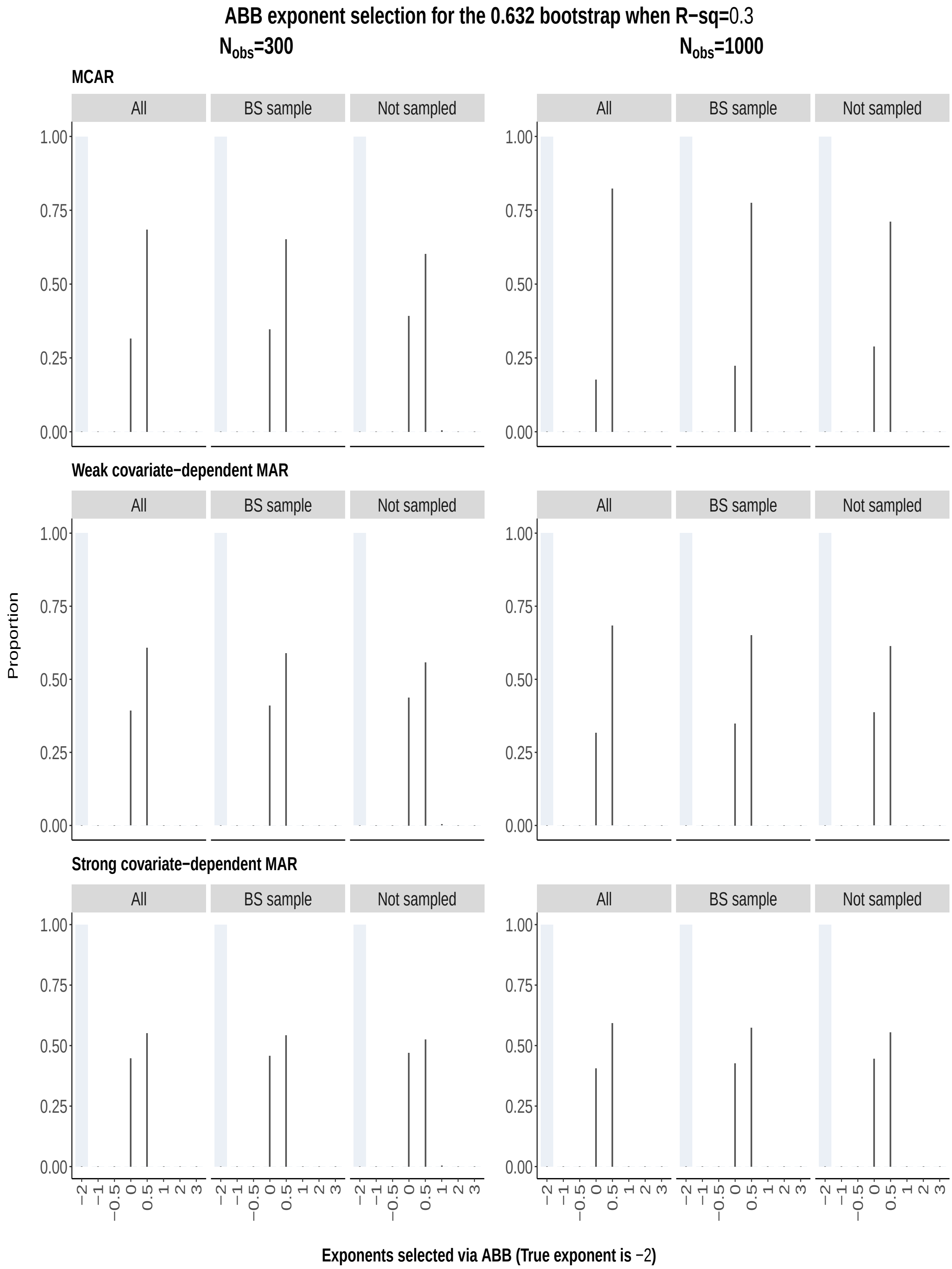

Figure S48: The proportion of times an exponent was selected to impute missing values ($M = 5$) using the ABB exponent selection when 25% of values are missing in X_1. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2, and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected in order to impute X_1^E. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2 Exponent selection from the FPS procedure

S8.2.1 Cross-validation, $\beta_2 = 1$, $\alpha_E = 1$ and no origin-shift

True exponent is 0

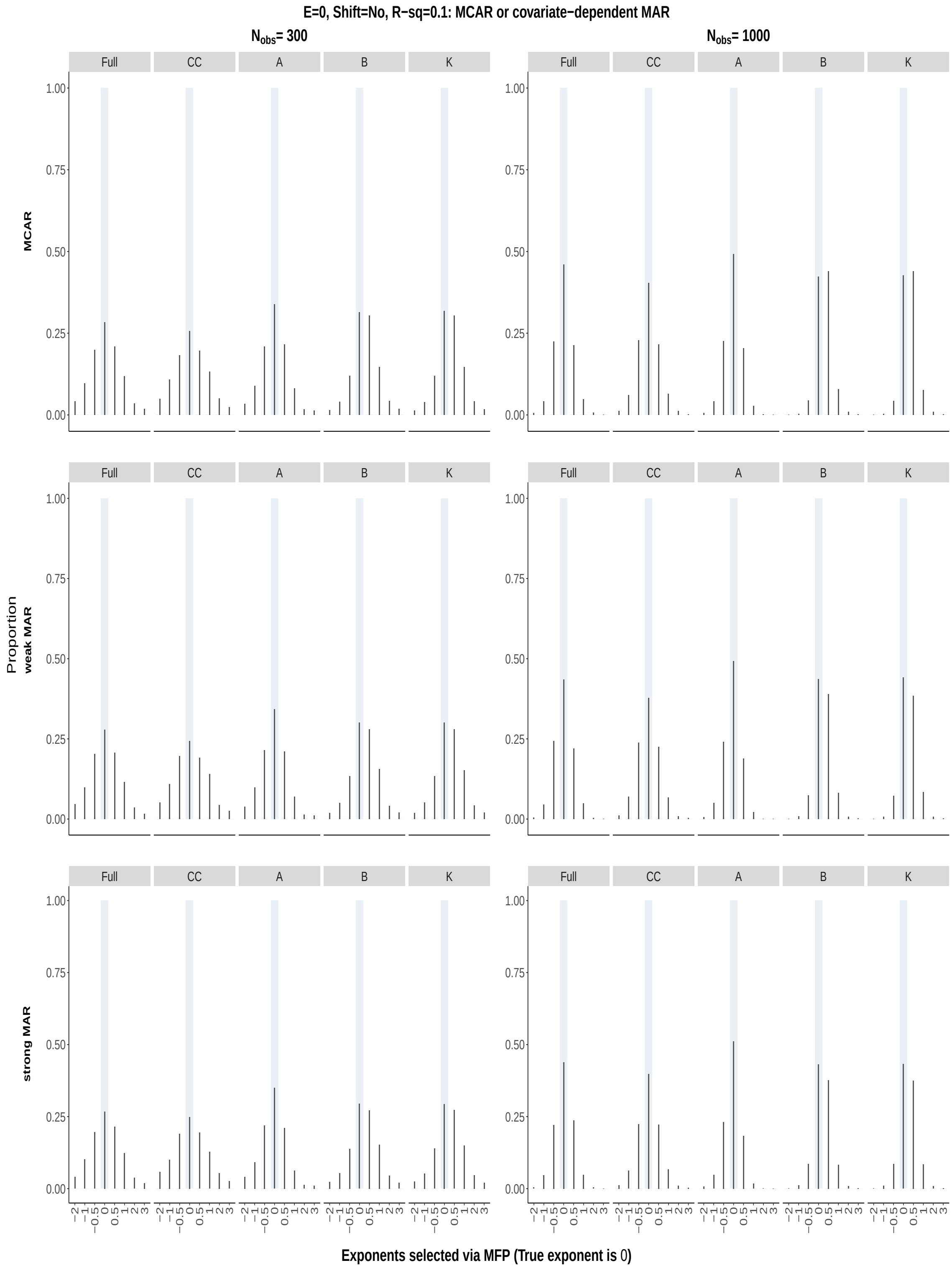

Figure S49: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

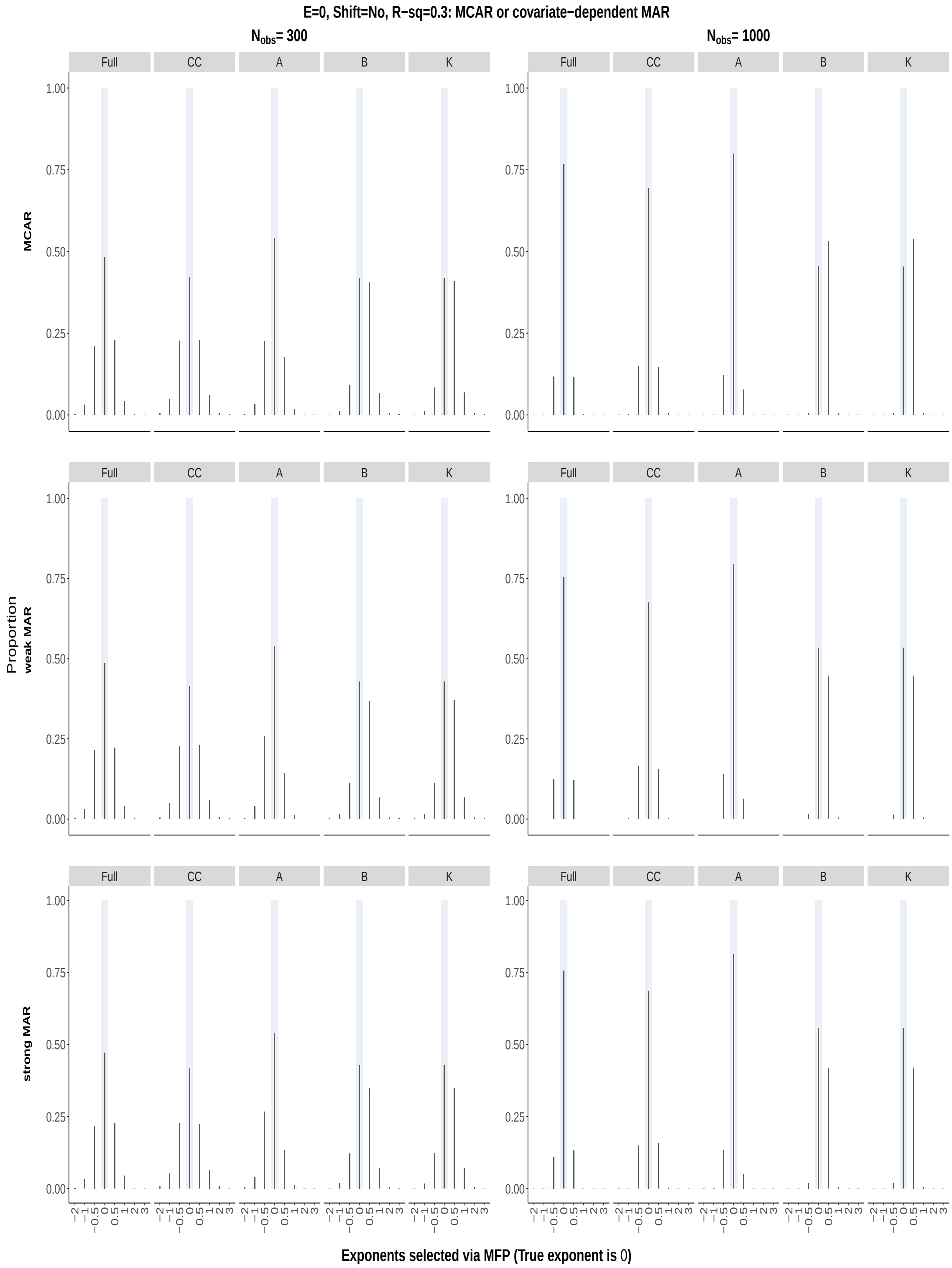

Figure S50: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

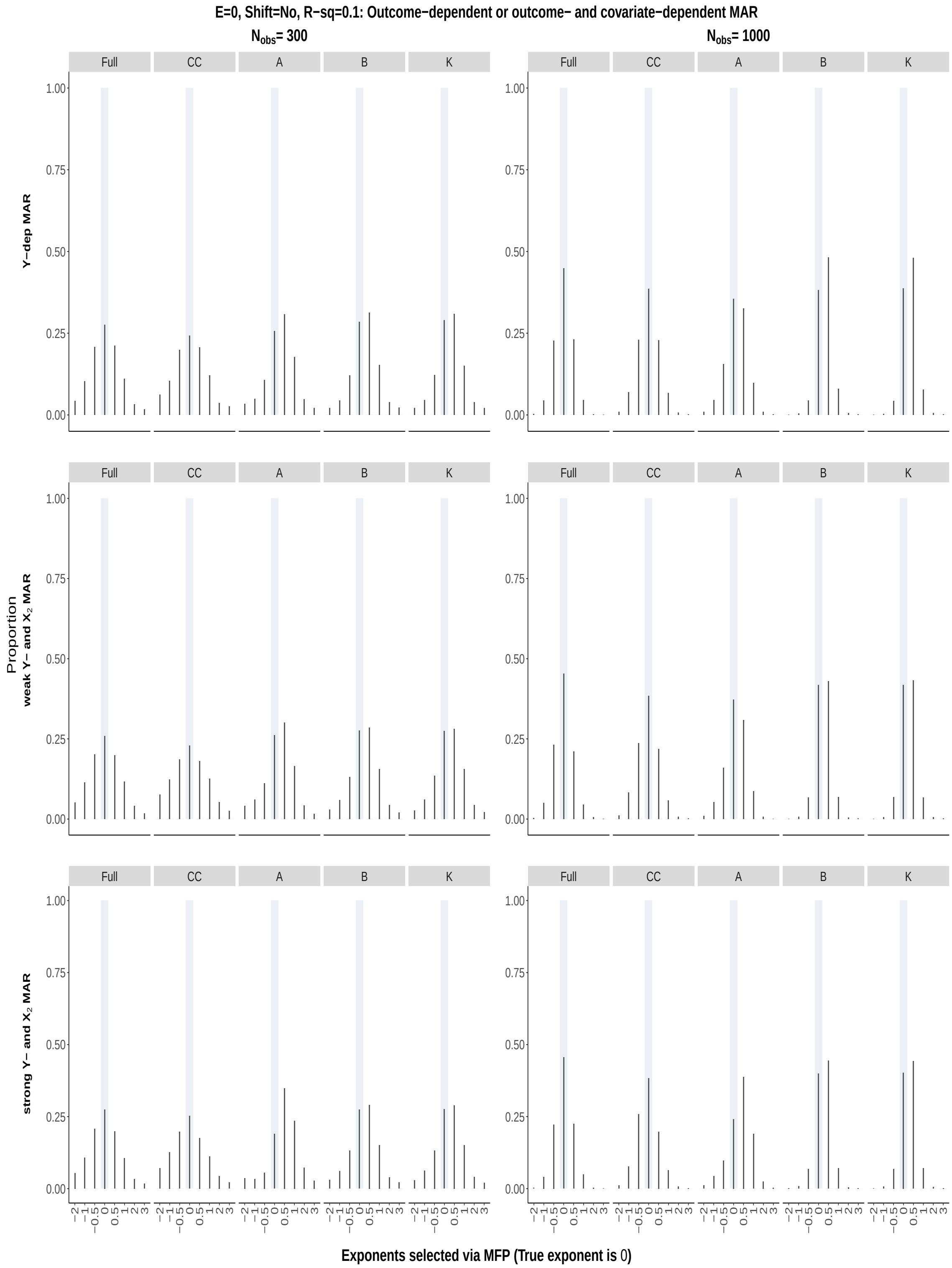

Figure S51: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

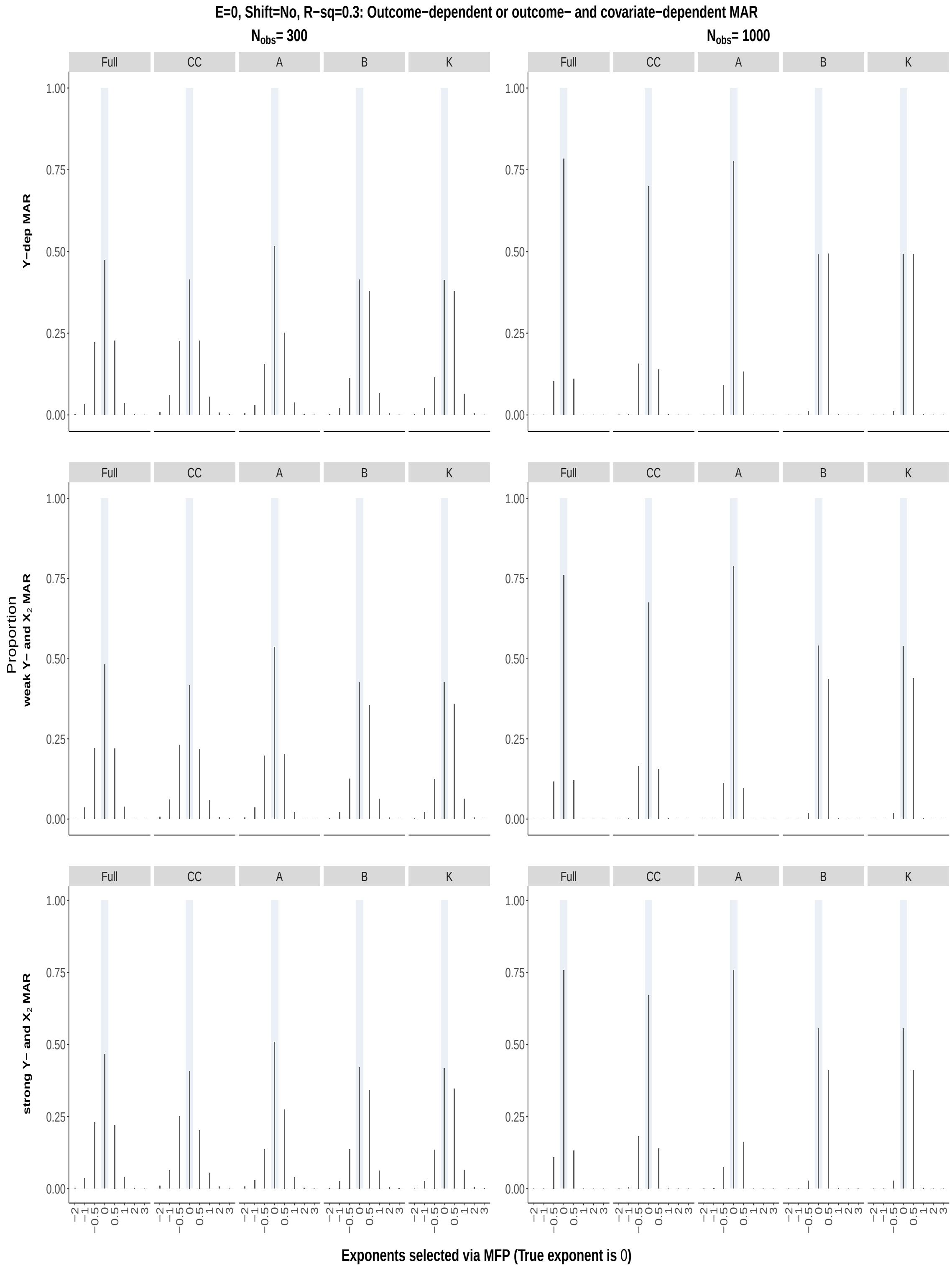

Figure S52: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

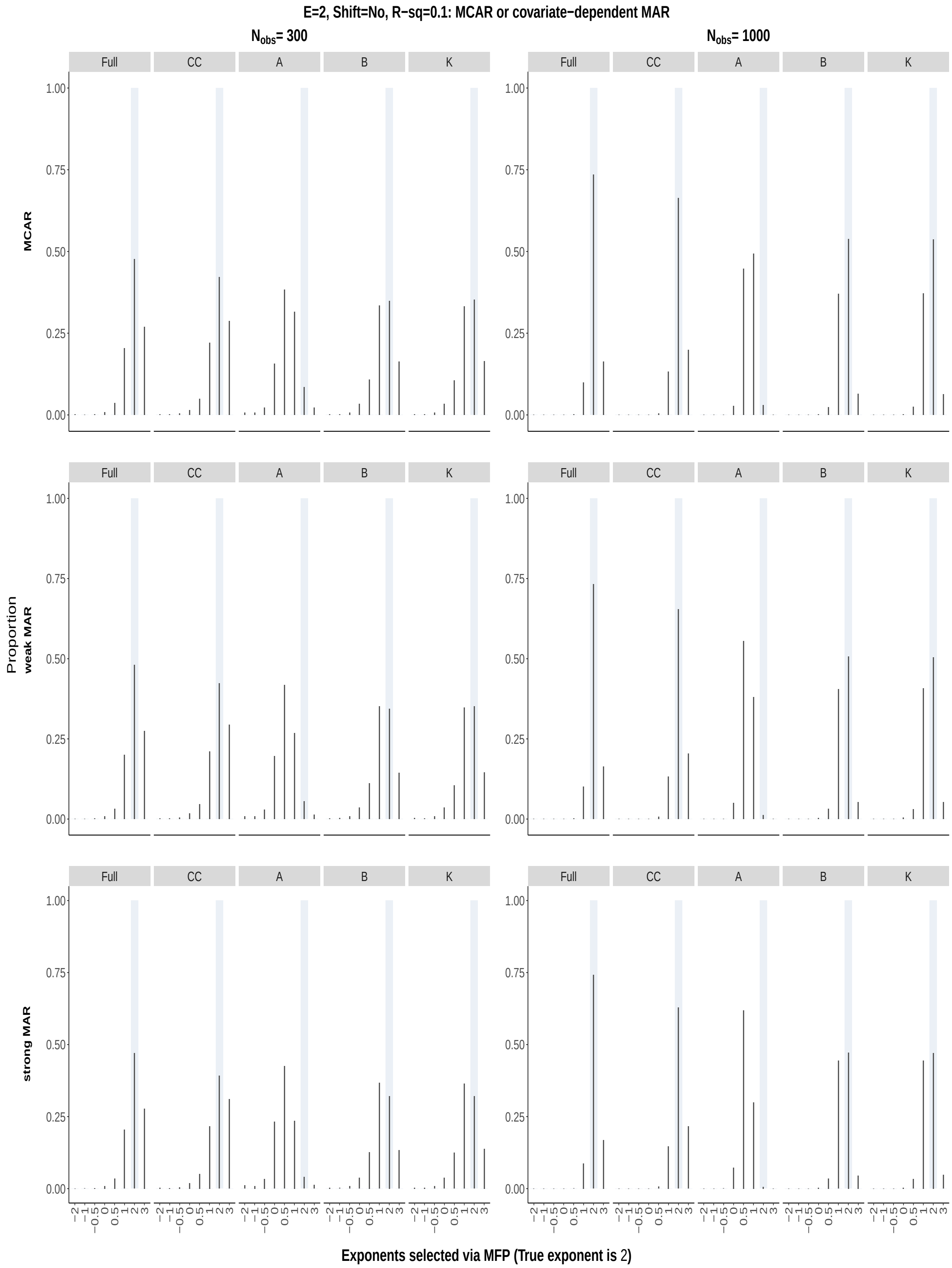

Figure S53: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

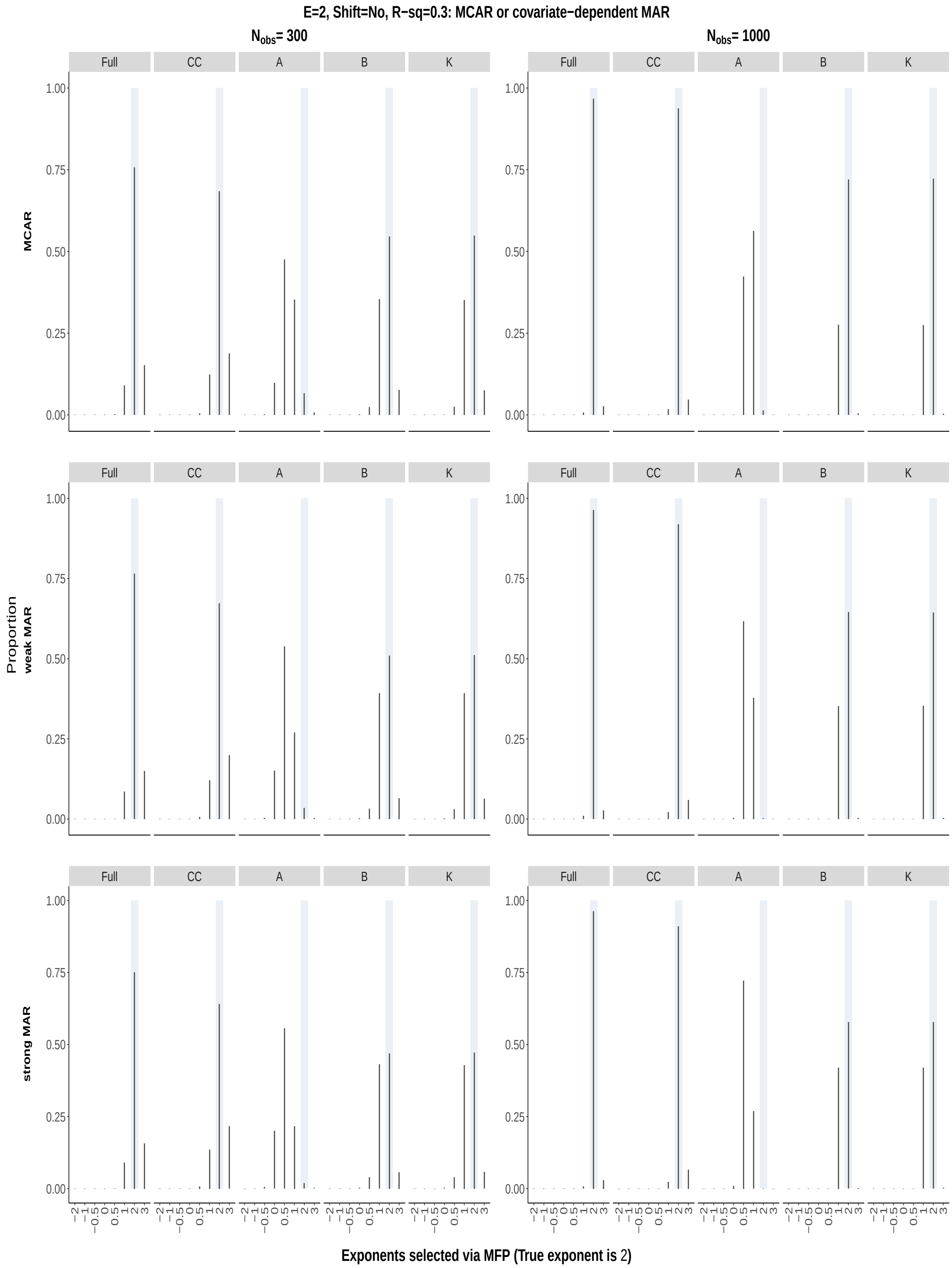

Figure S54: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

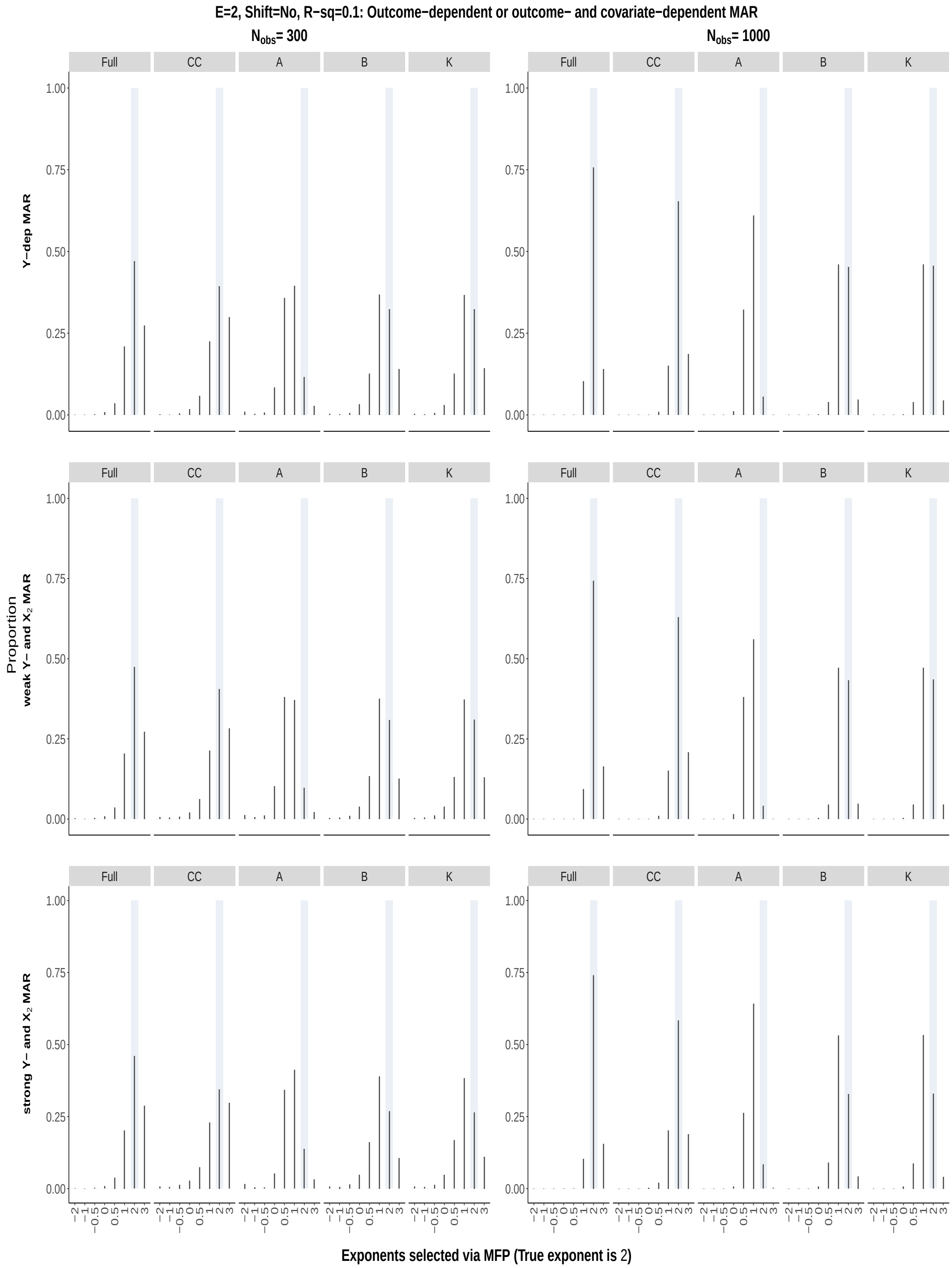

Figure S55: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

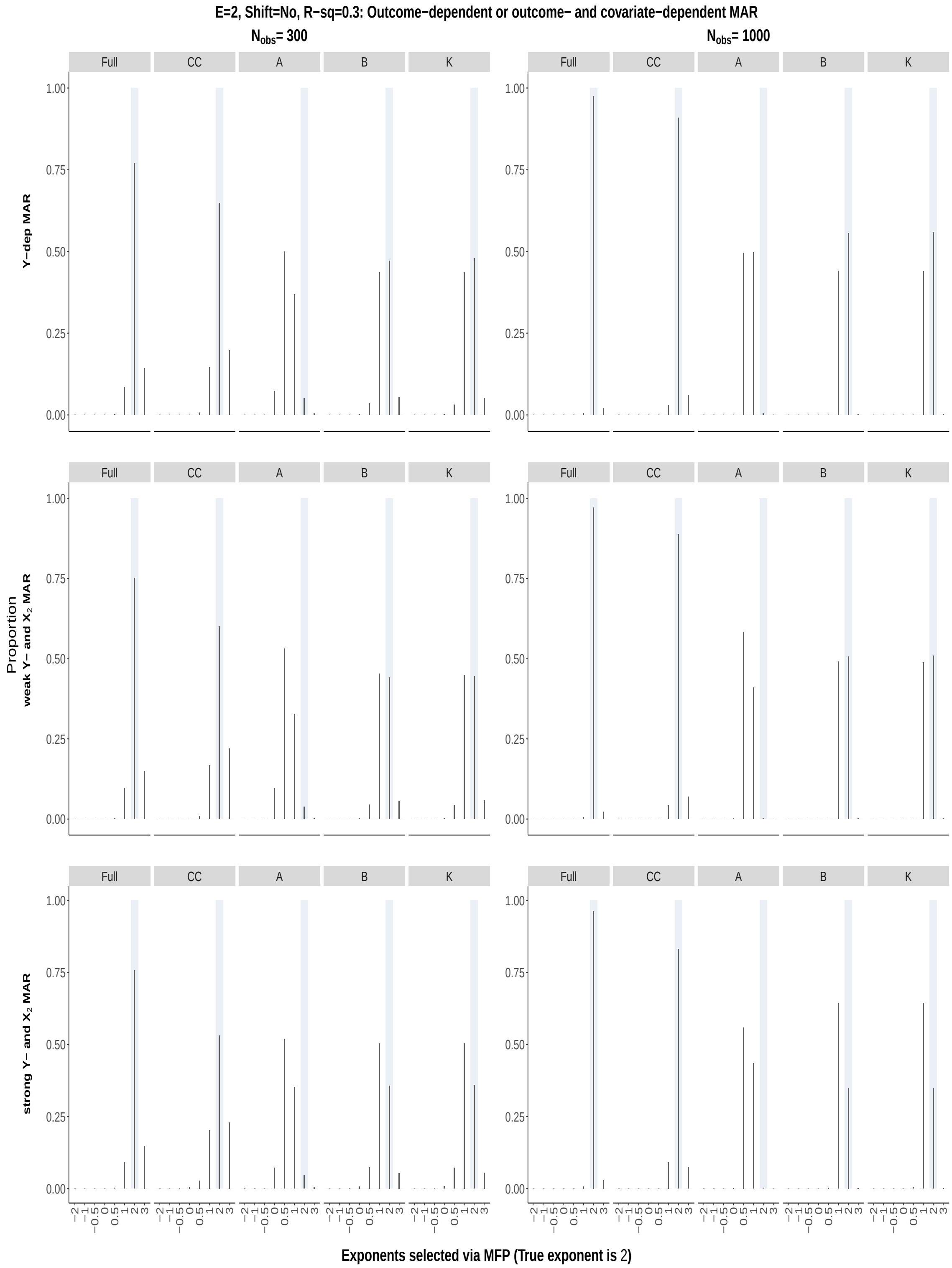

Figure S56: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

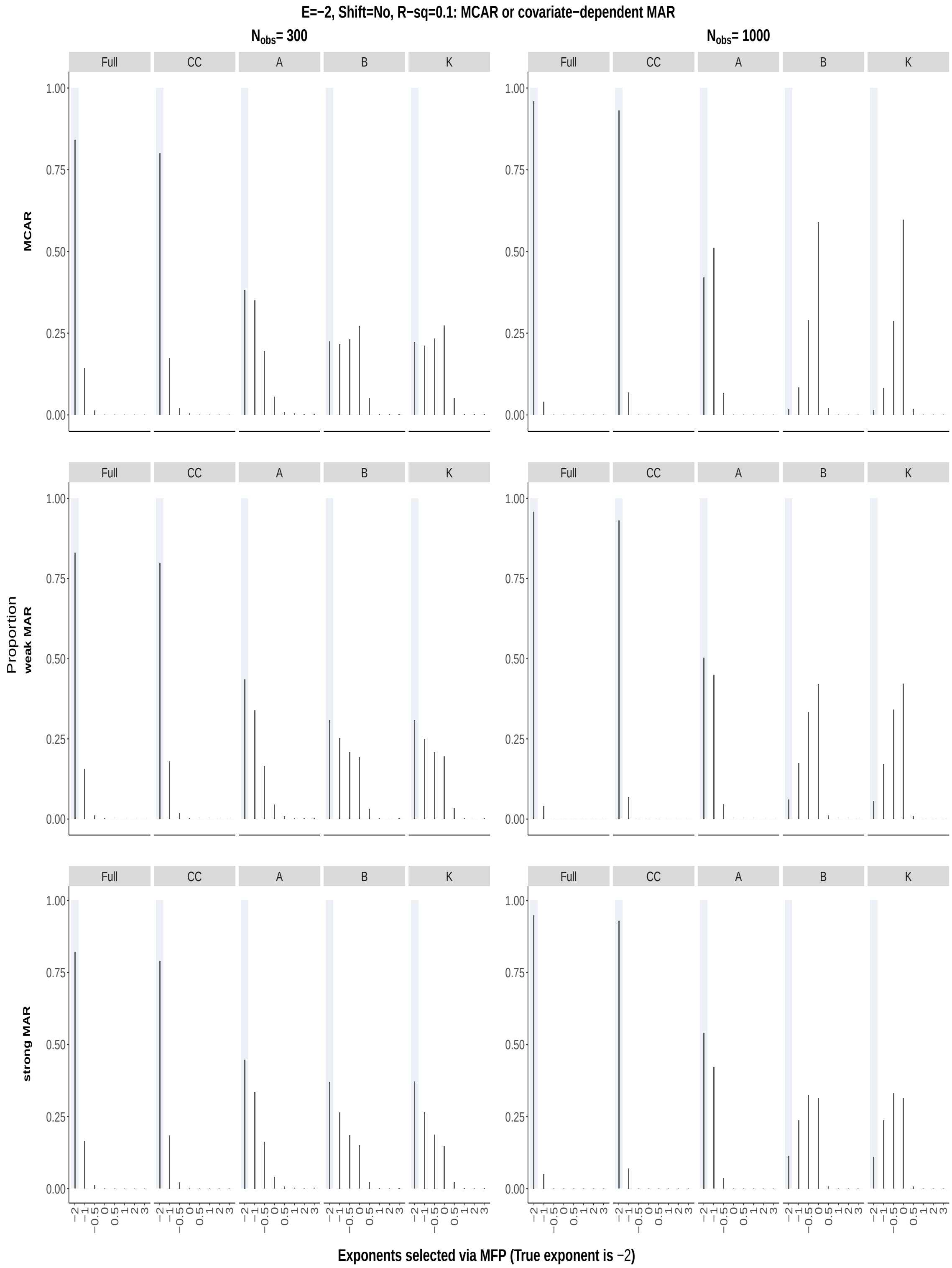

Figure S57: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

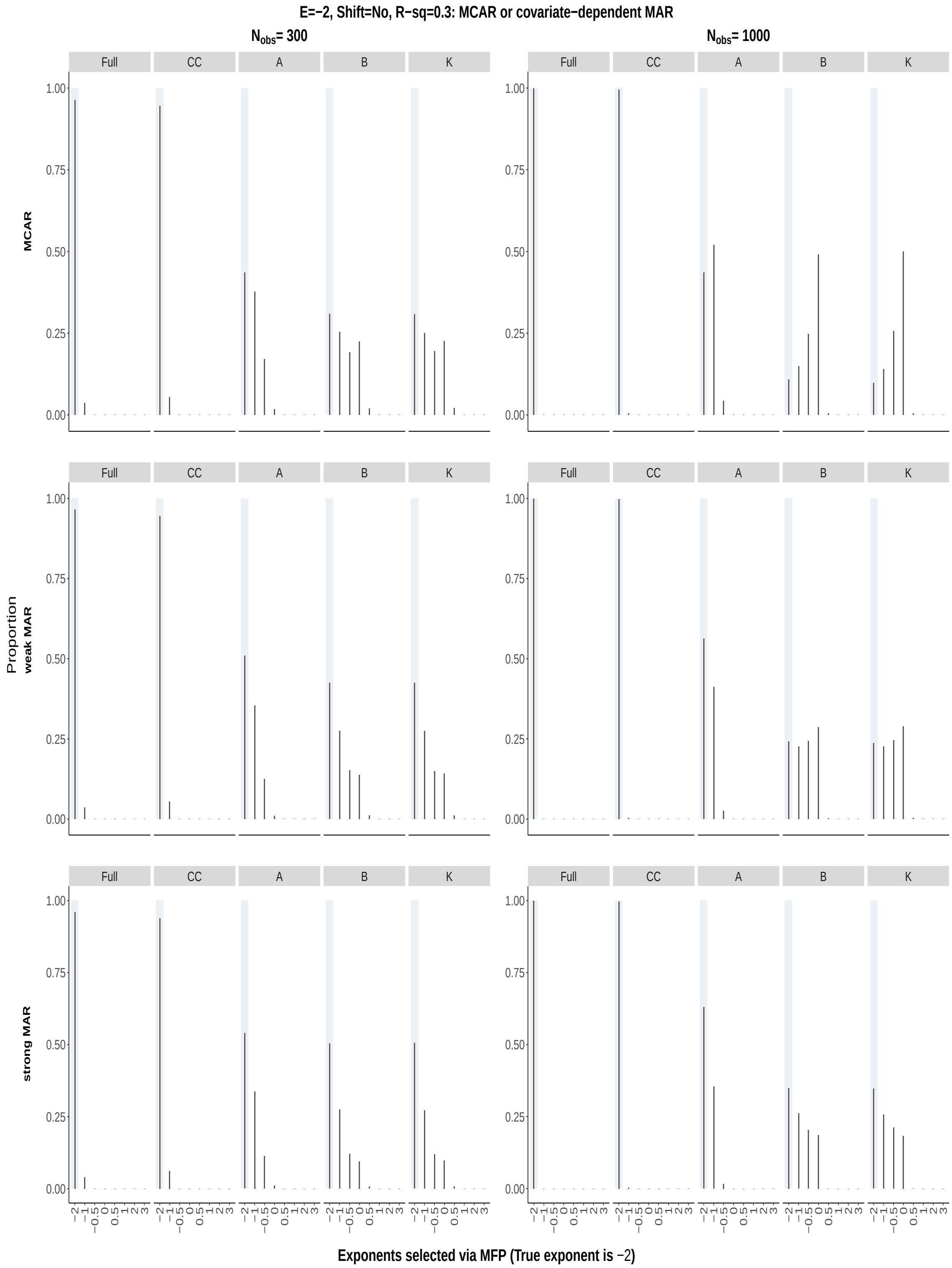

Figure S58: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

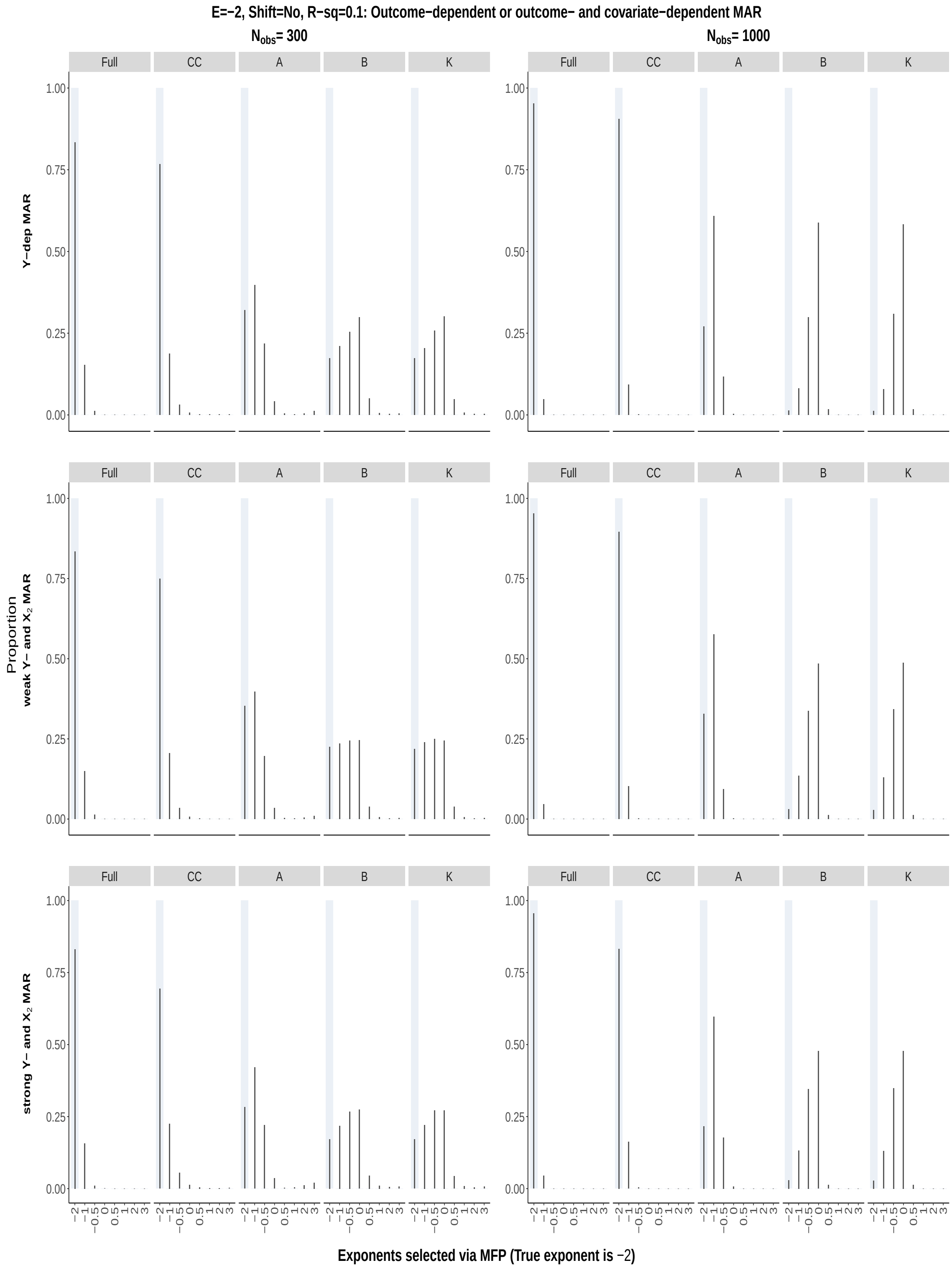

Figure S59: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

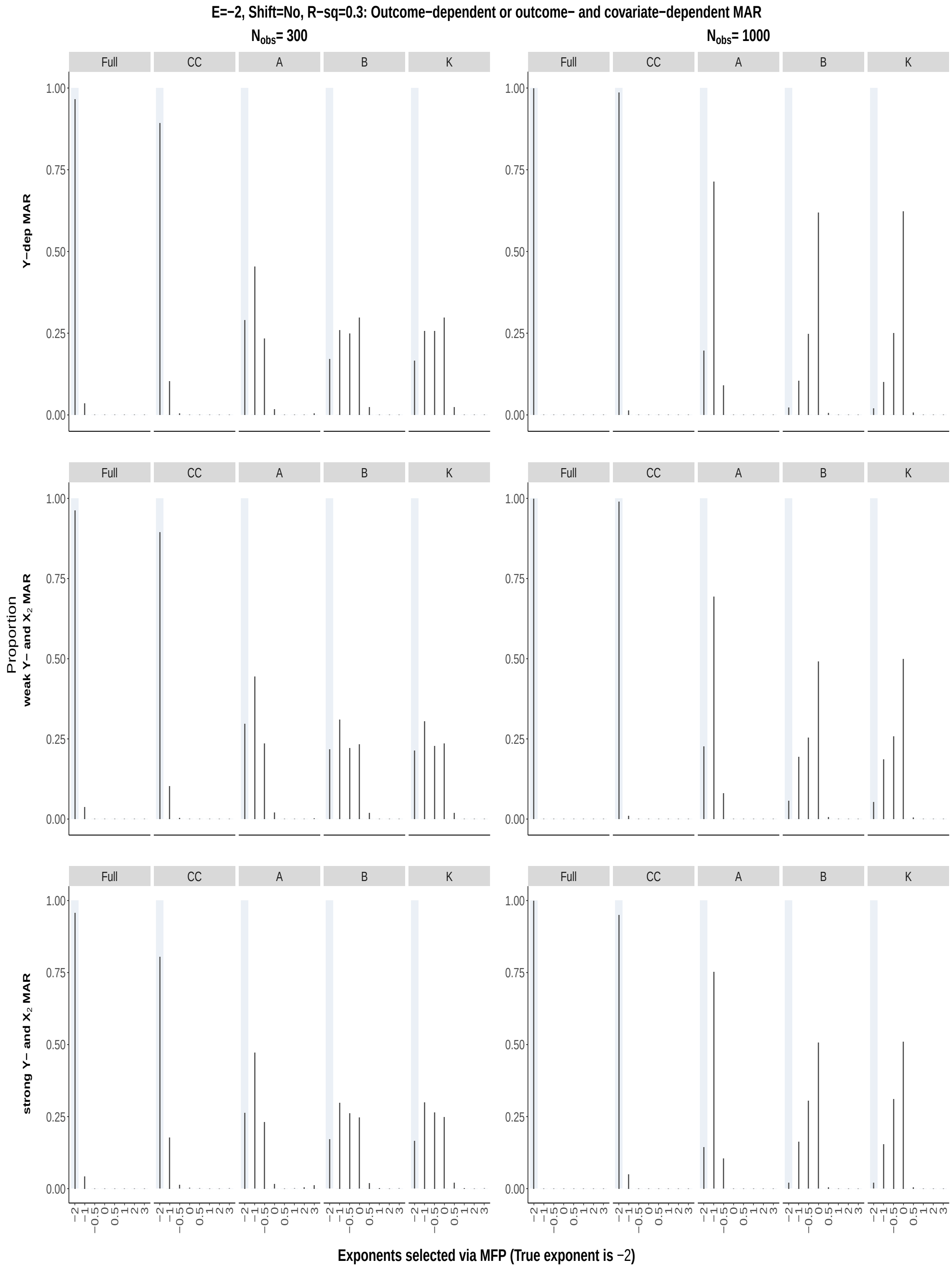

Figure S60: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.2 Cross-validation, $\beta_2 = 1$, $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

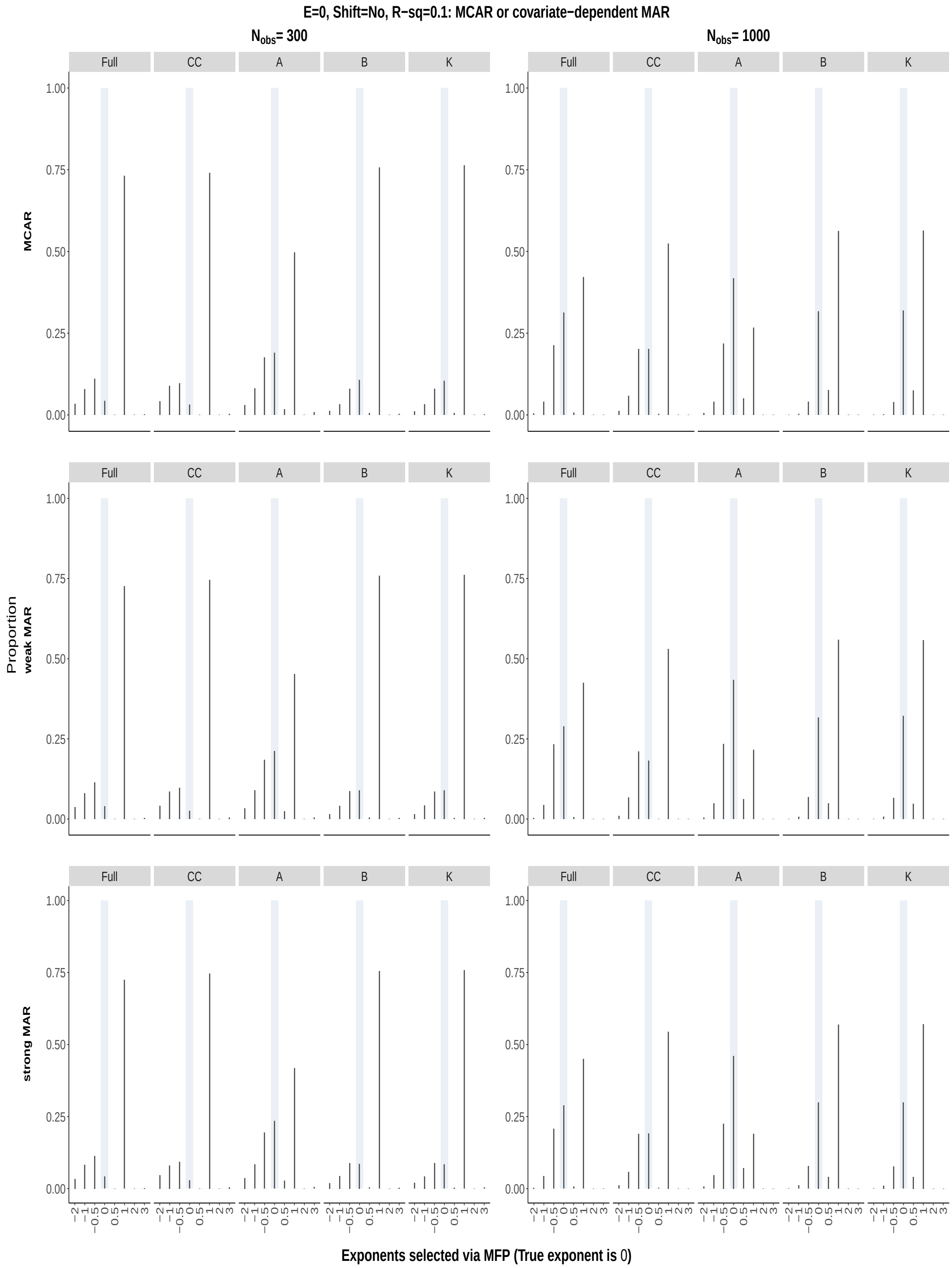

Figure S61: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

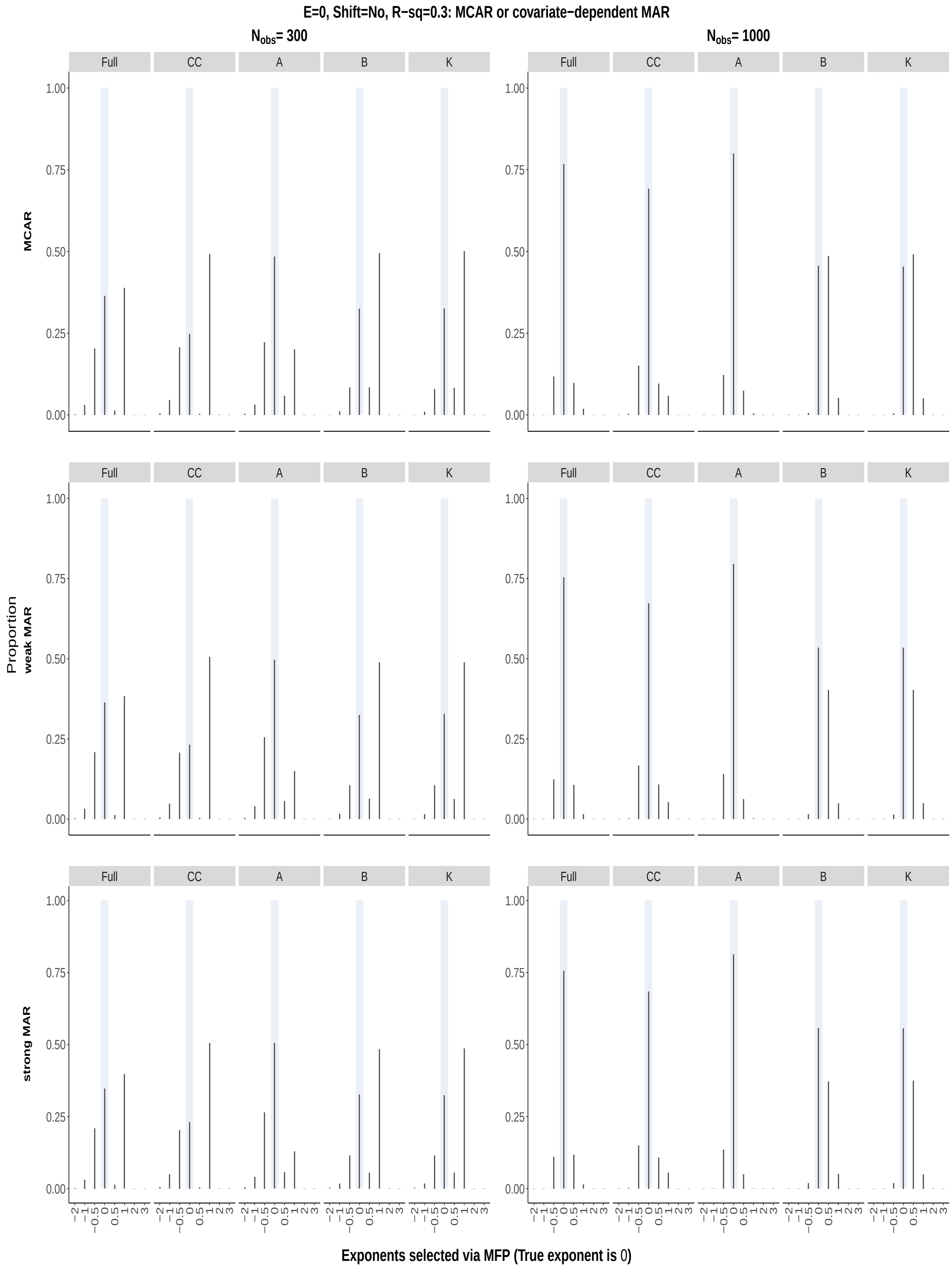

Figure S62: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

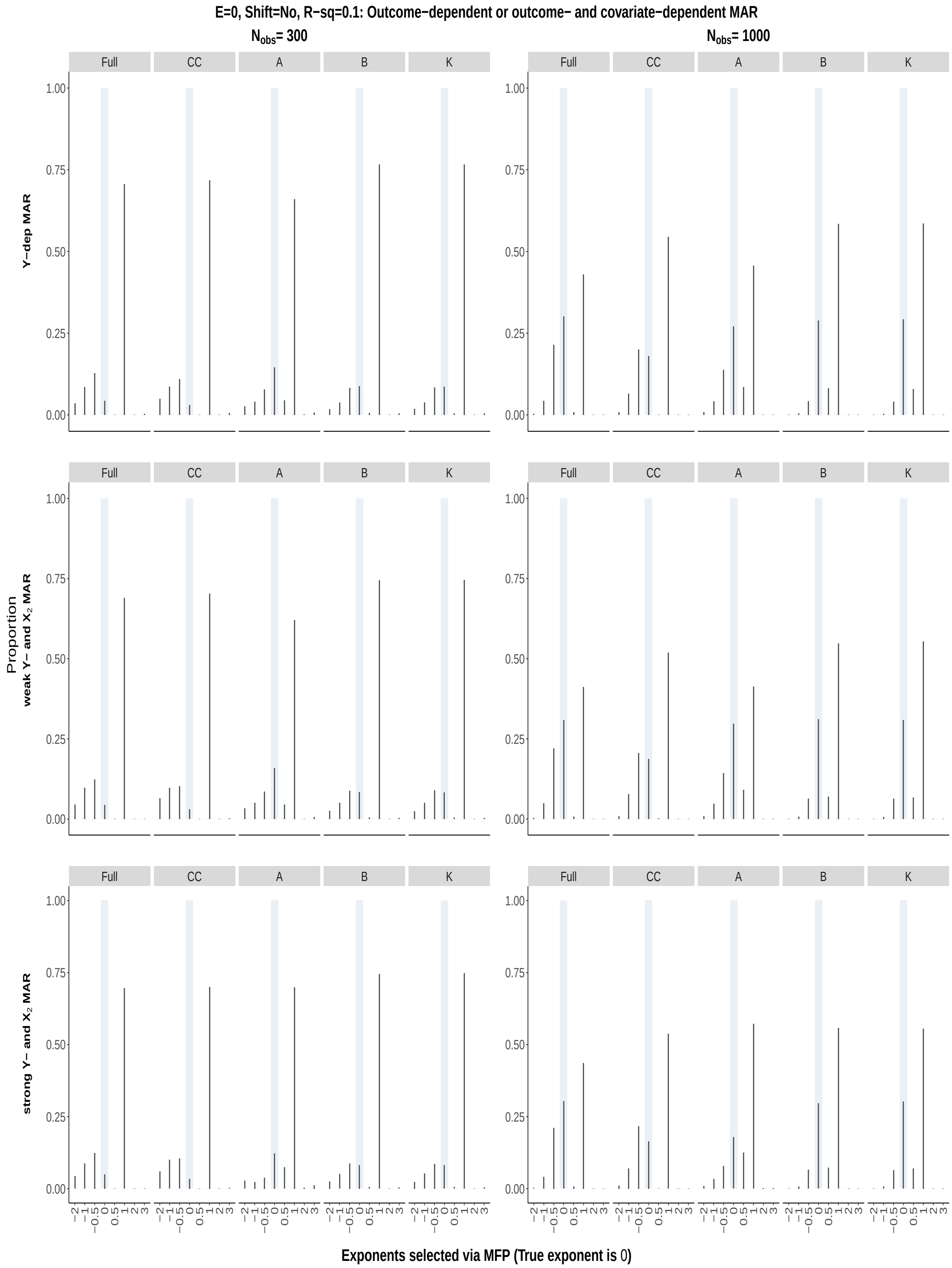

Figure S63: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

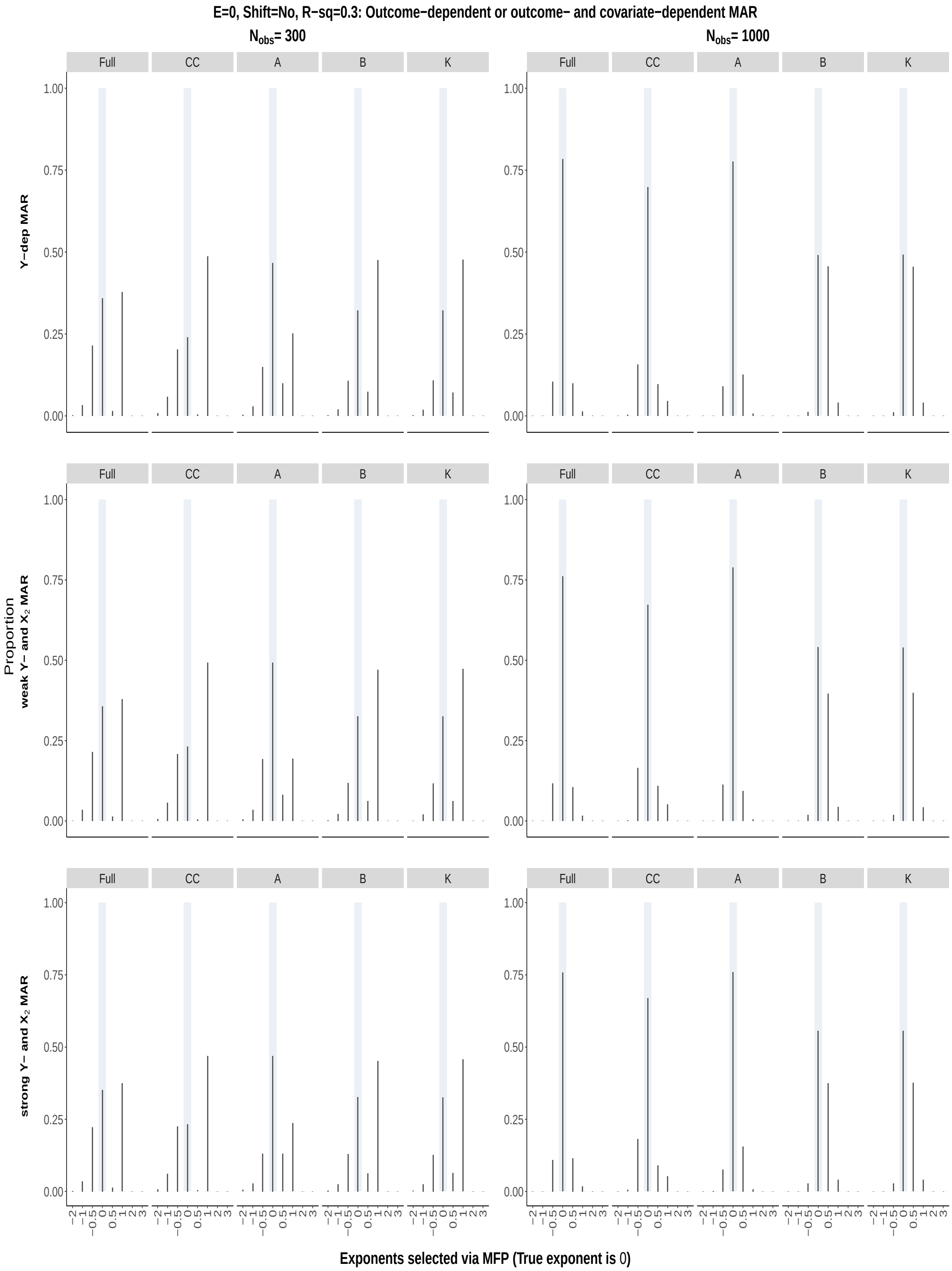

Figure S64: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

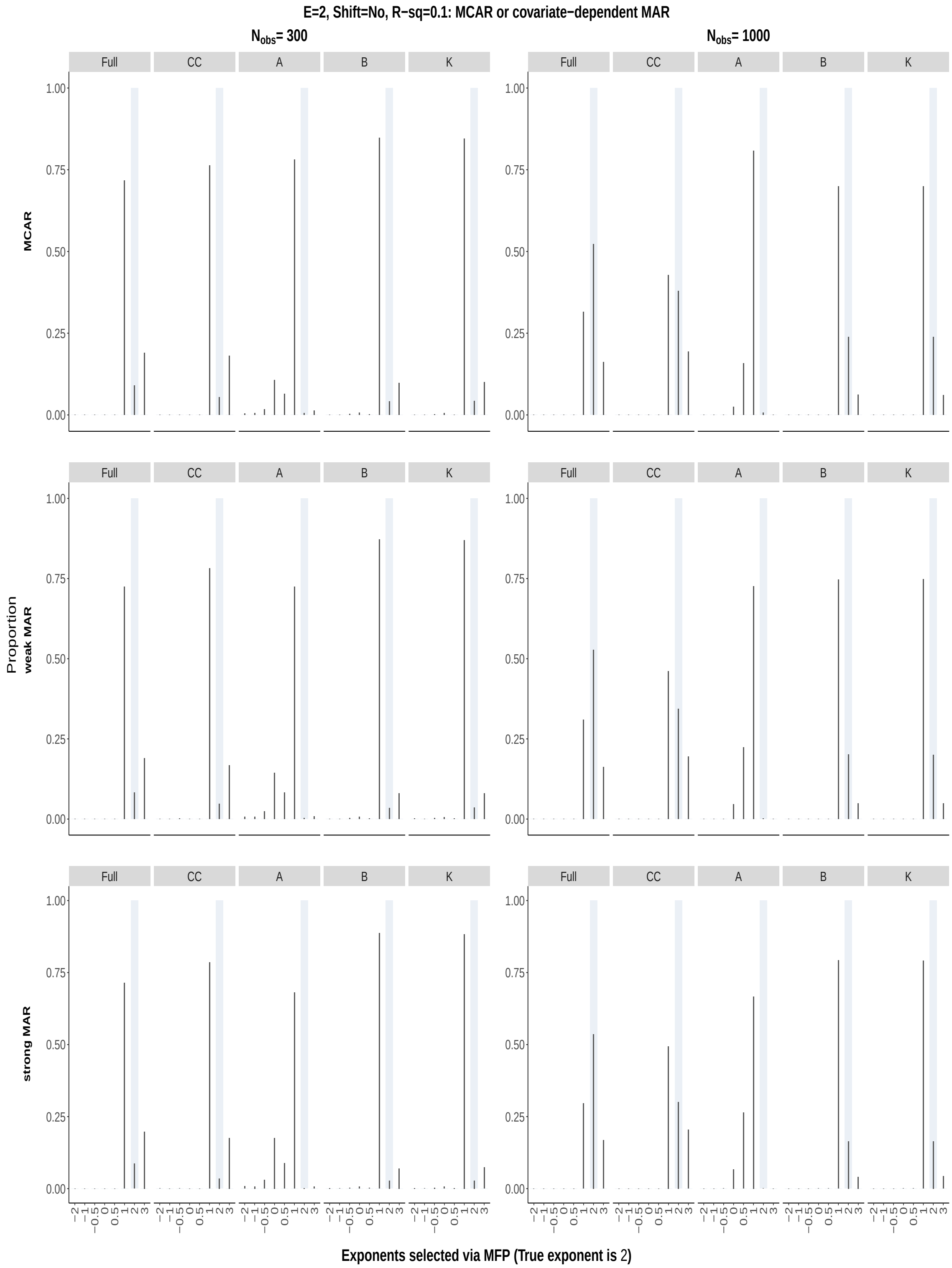

Figure S65: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

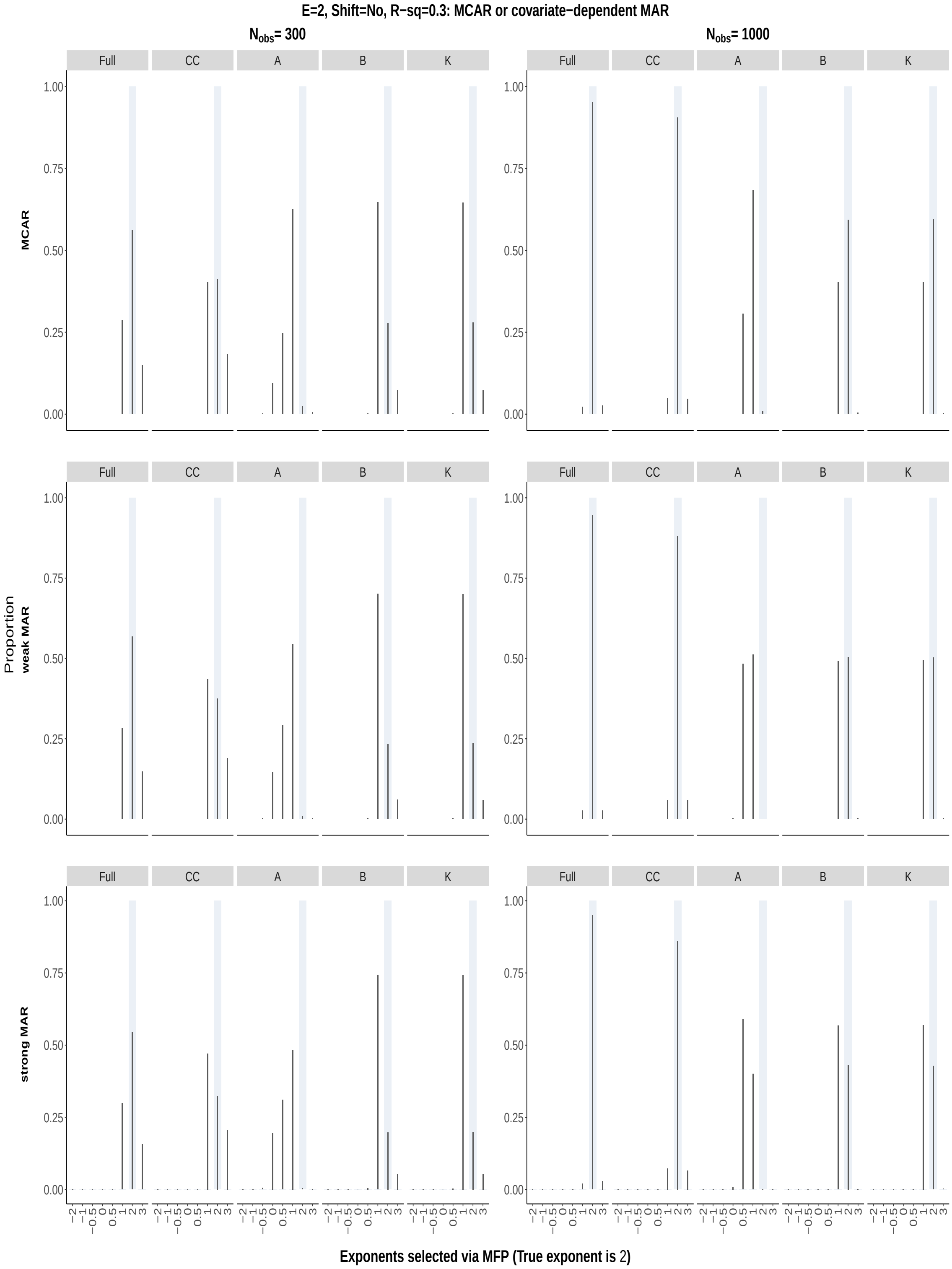

Figure S66: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

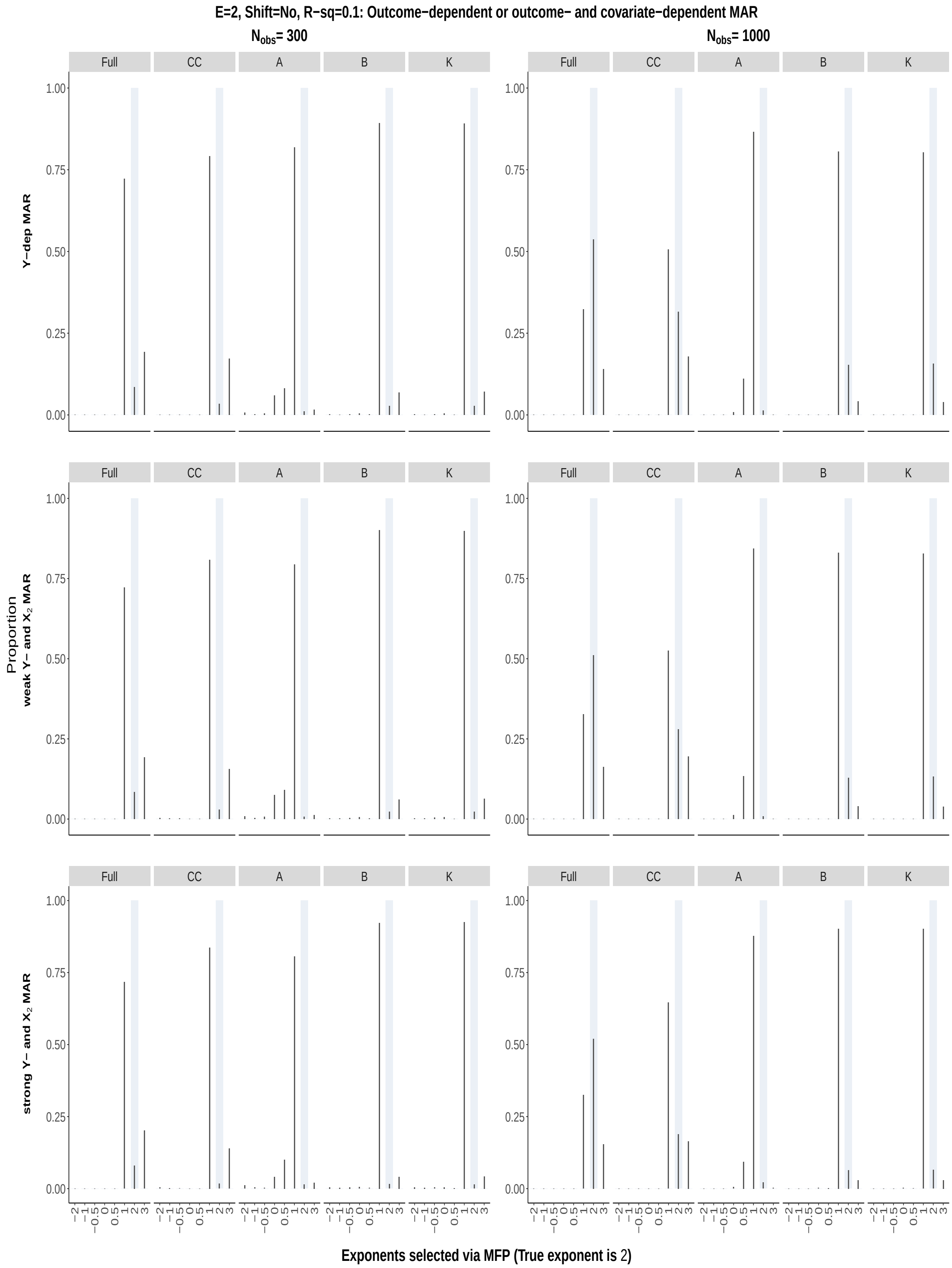

Figure S67: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

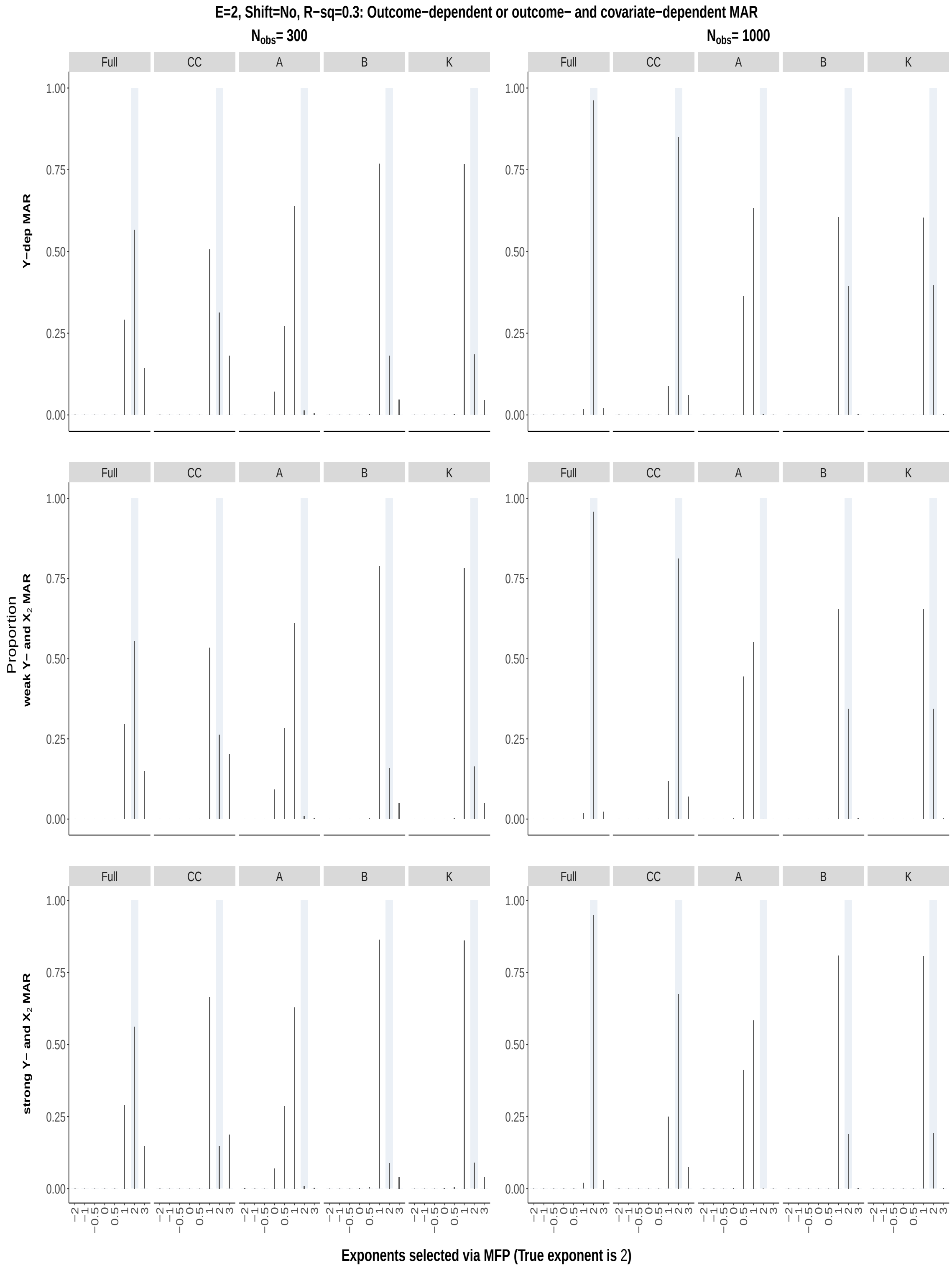

Figure S68: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

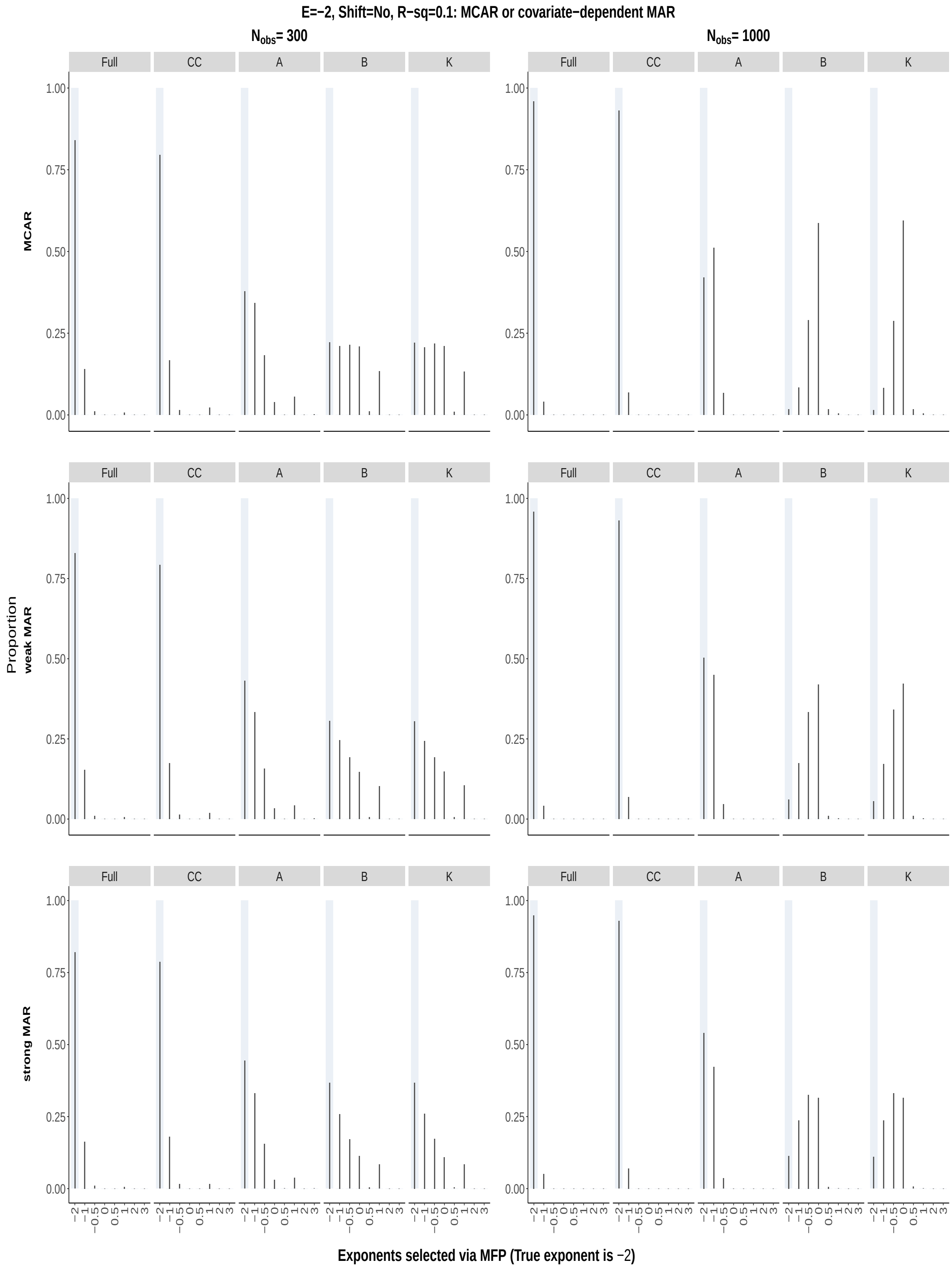

Figure S69: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

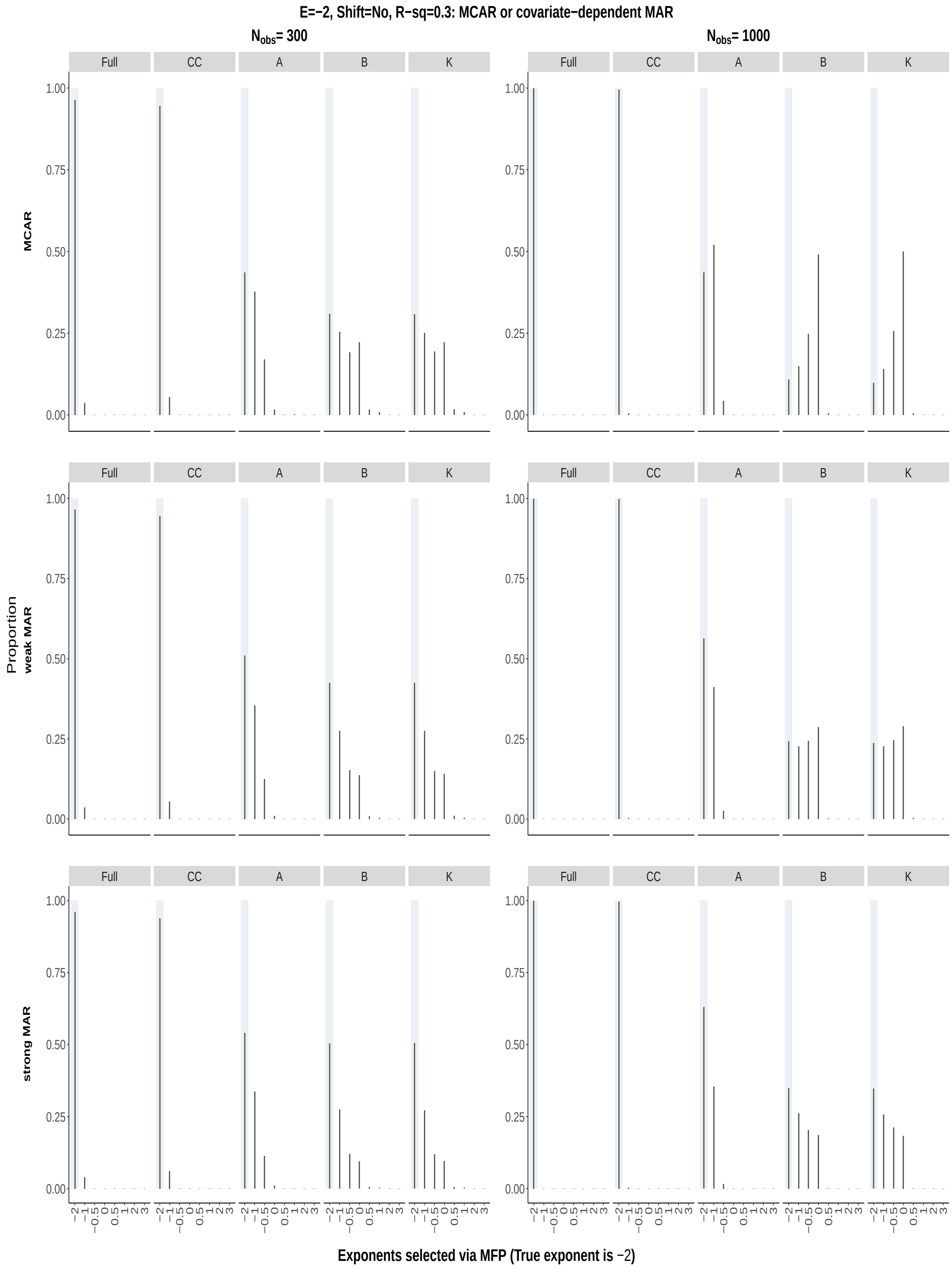

Figure S70: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

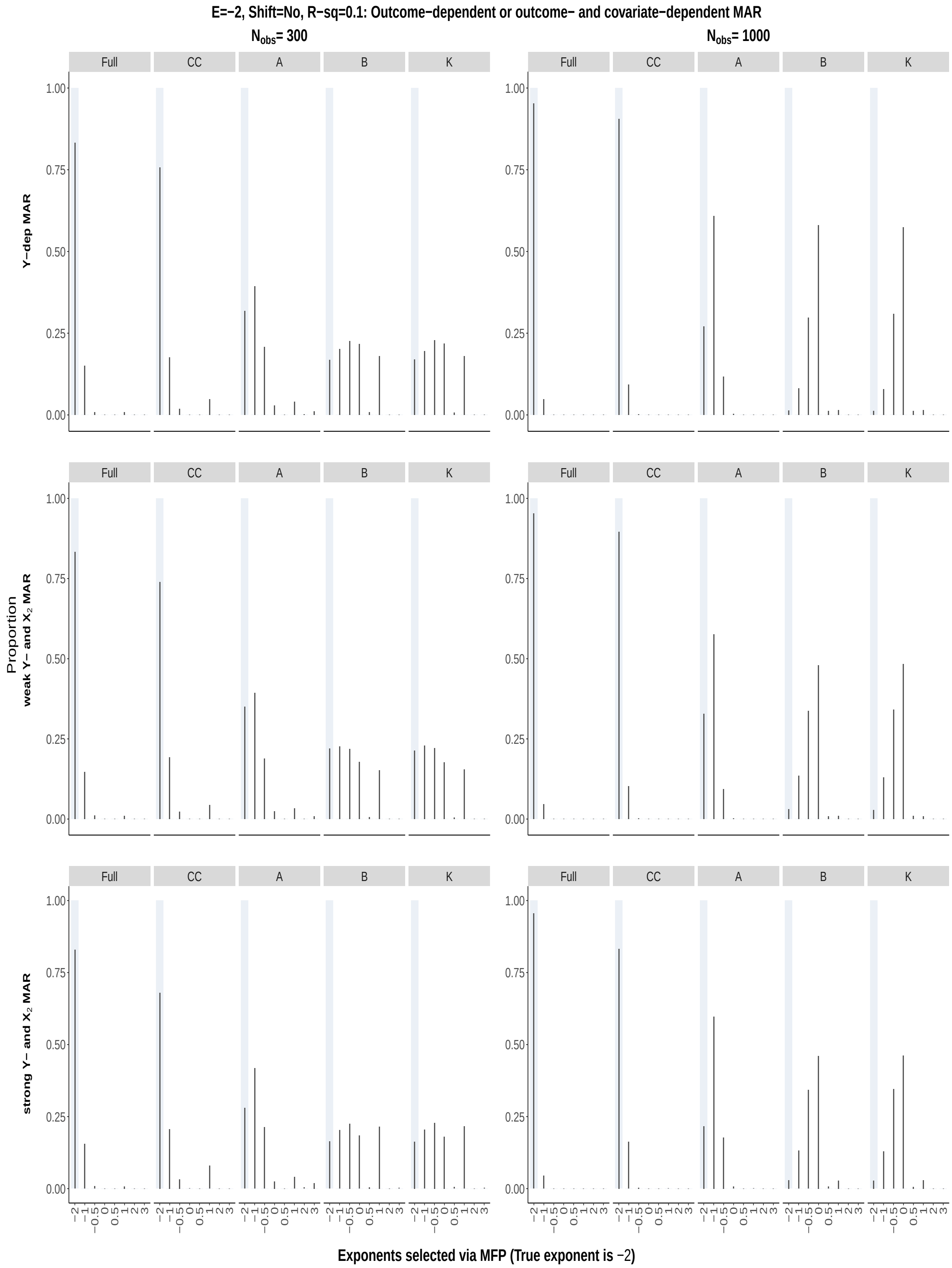

Figure S71: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

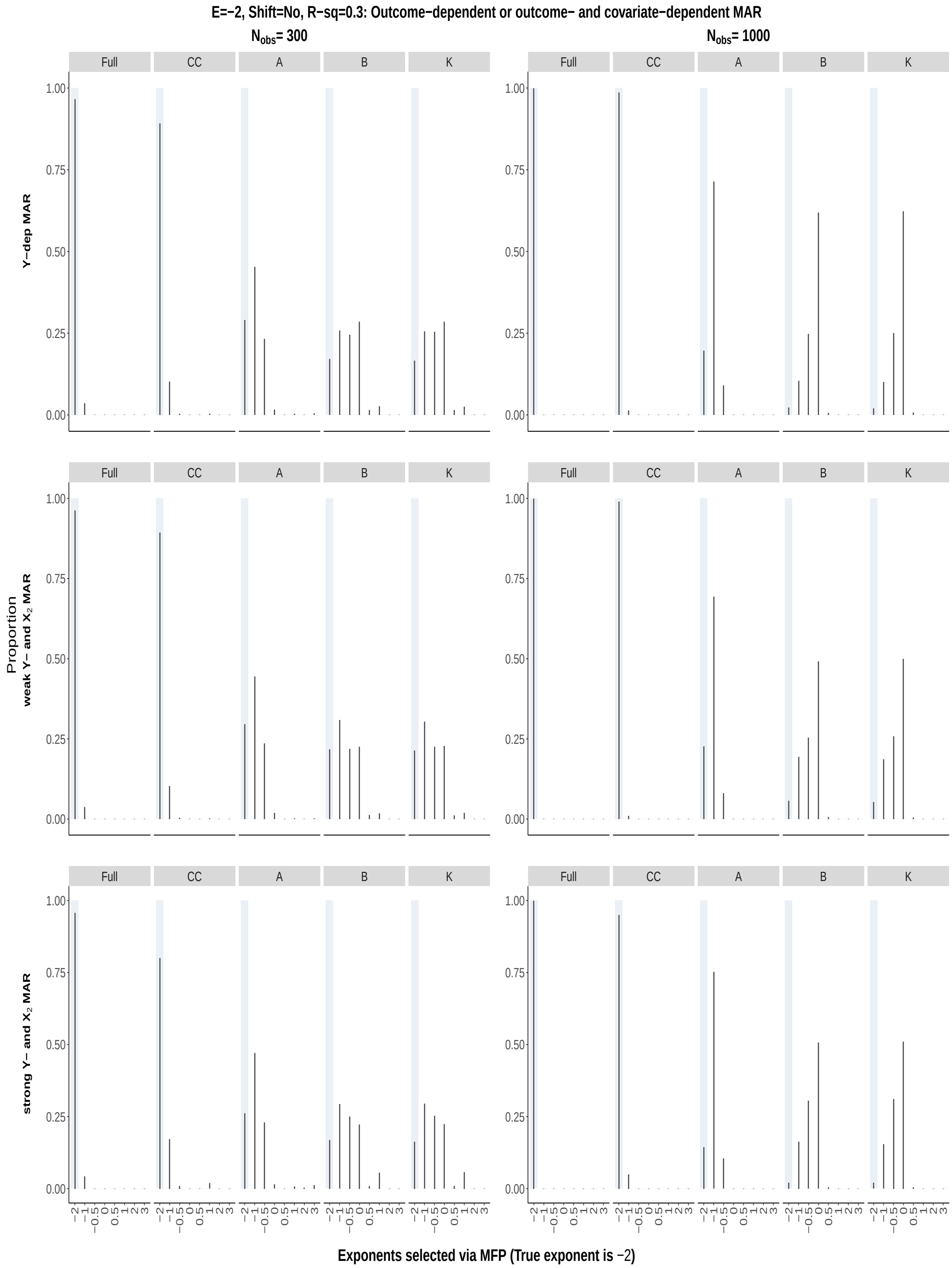

Figure S72: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.3 Cross-validation, $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been used

True exponent is 0

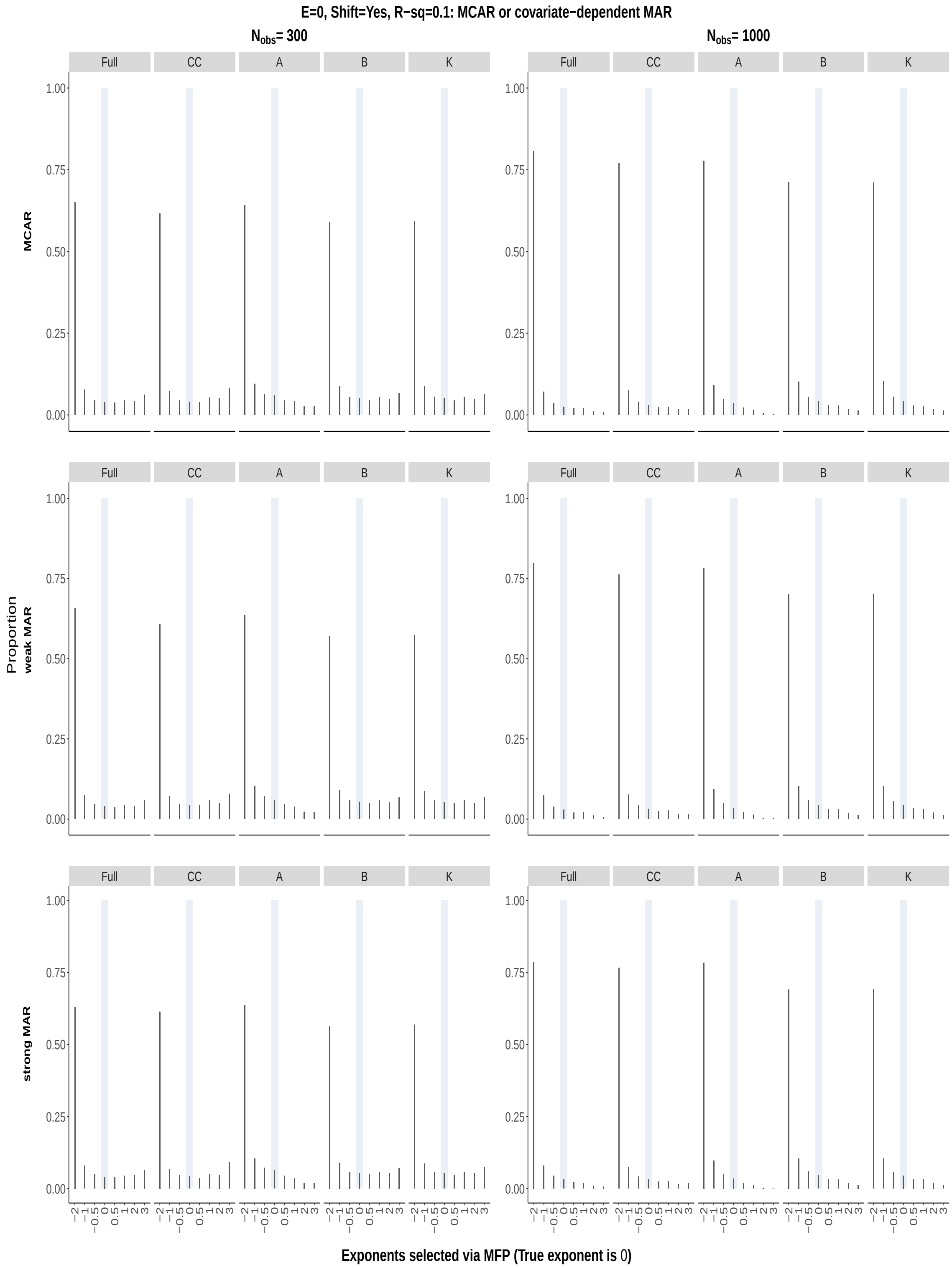

Figure S73: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

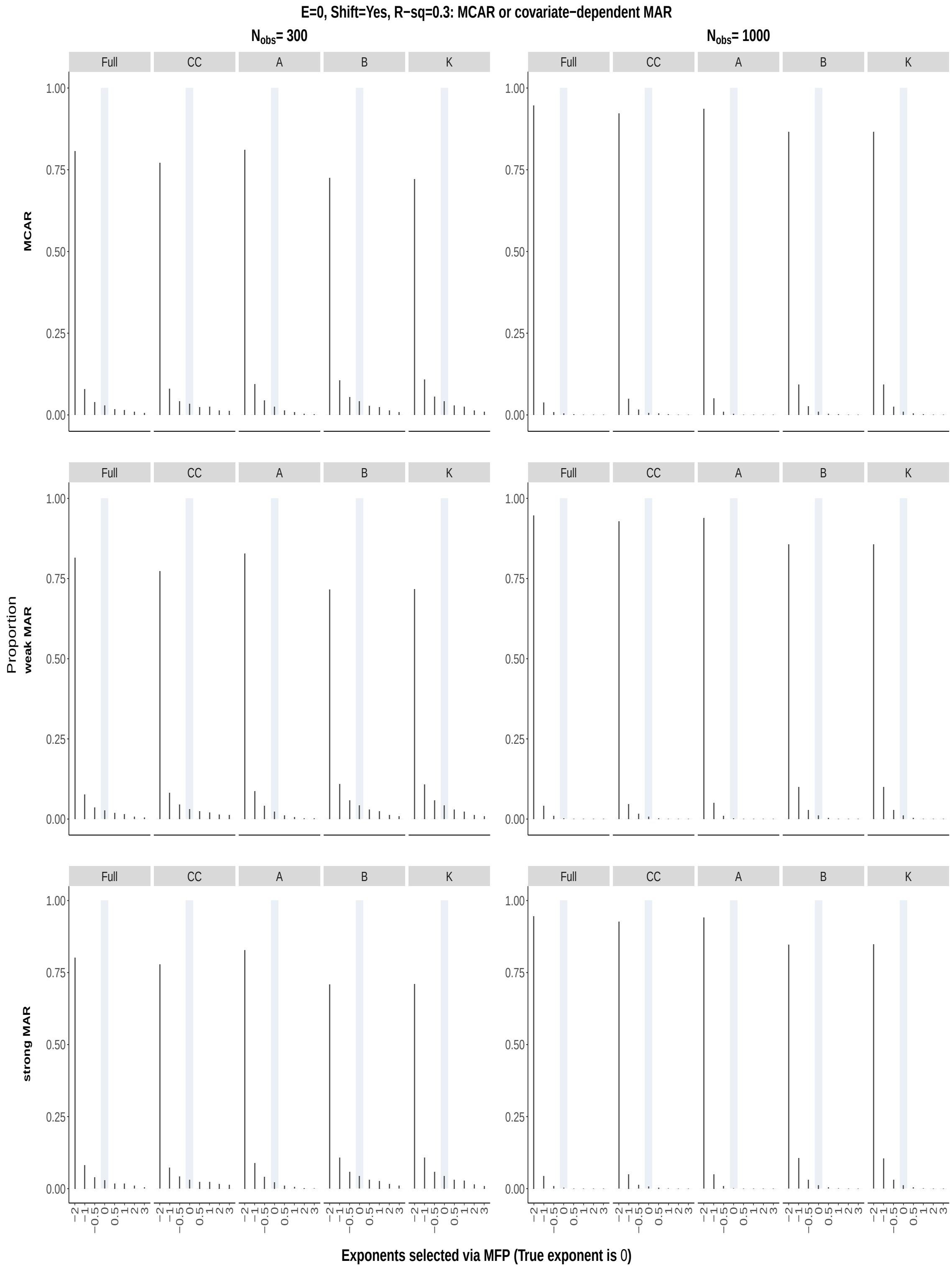

Figure S74: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

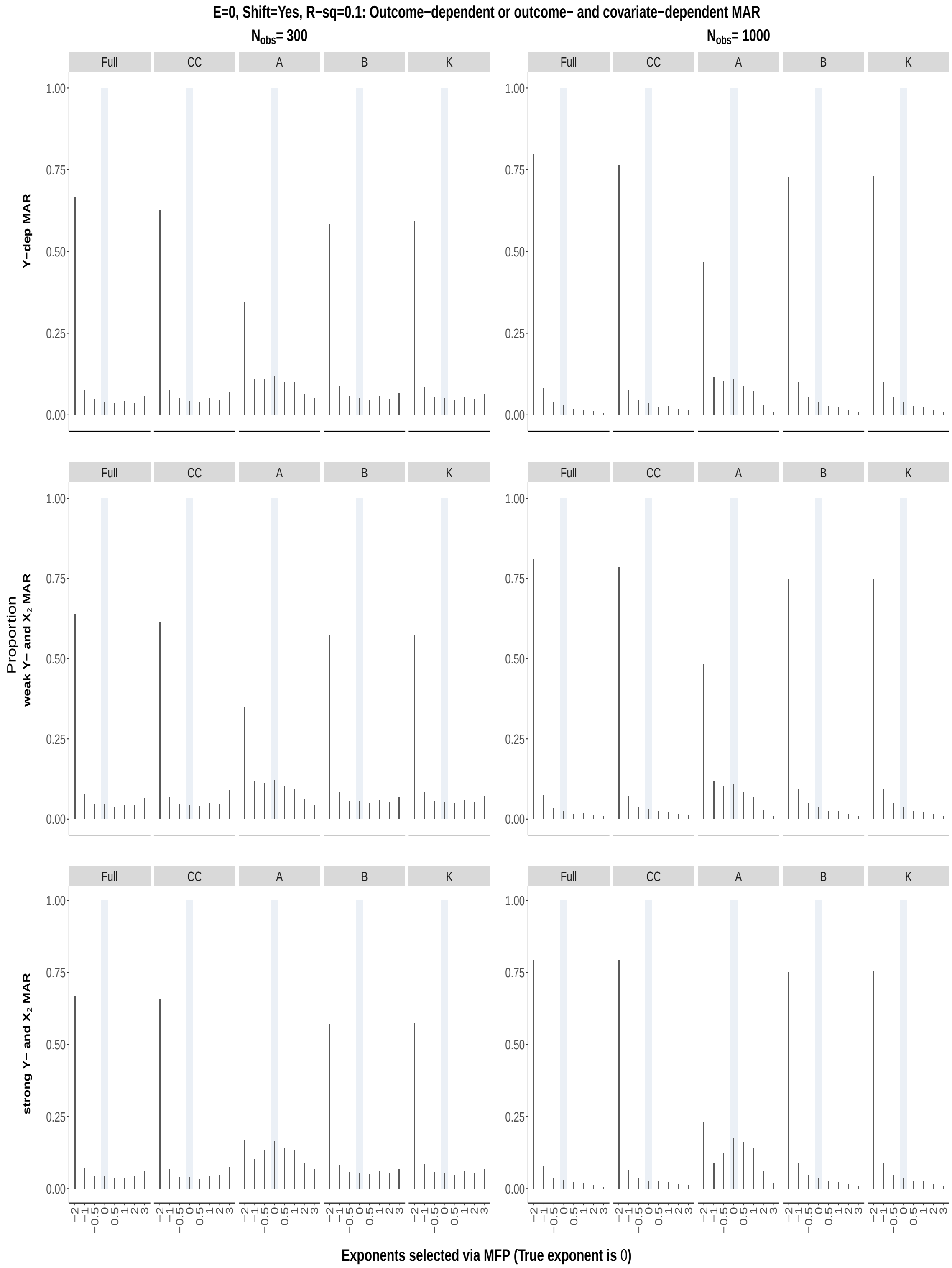

Figure S75: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

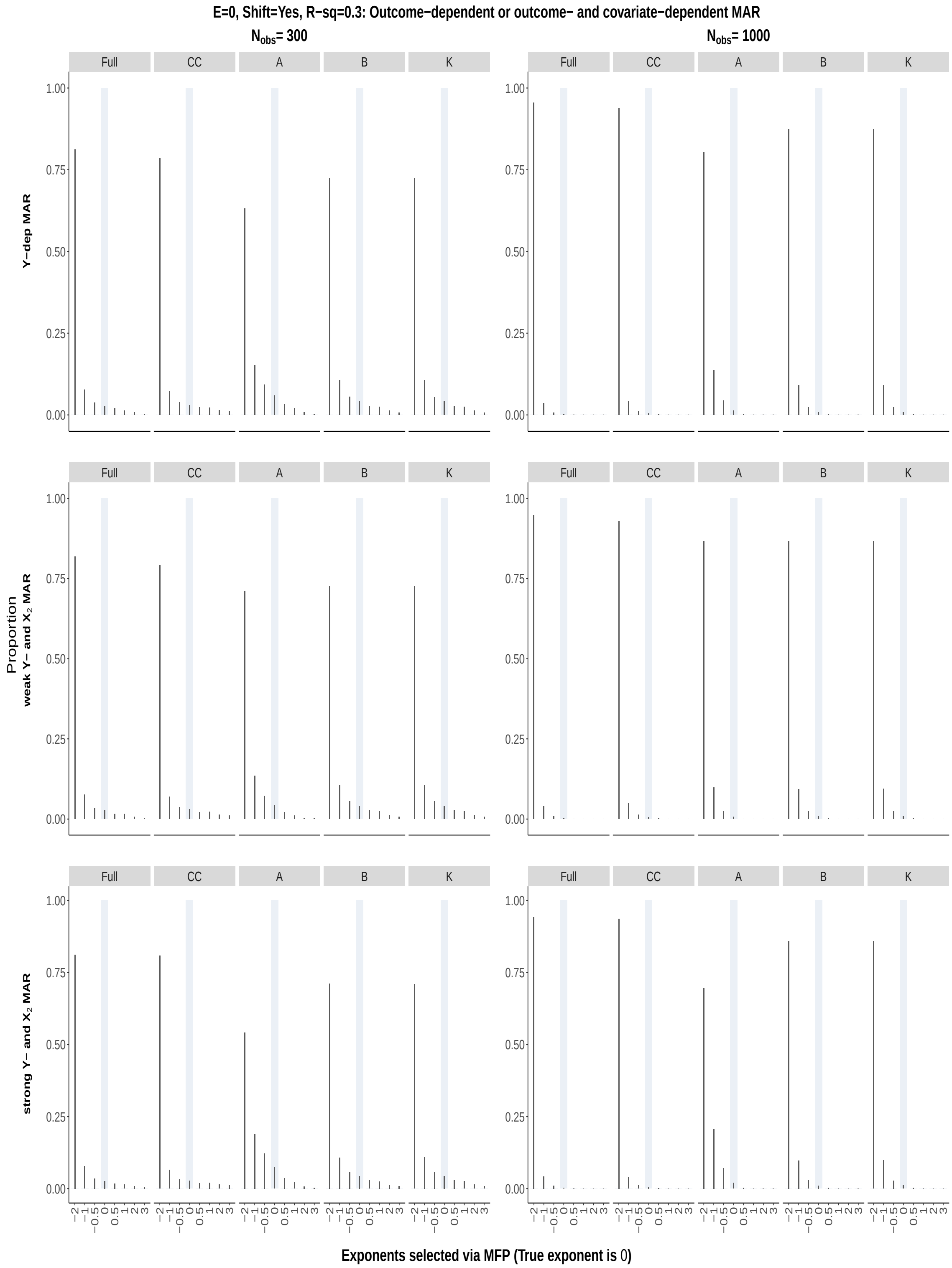

Figure S76: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

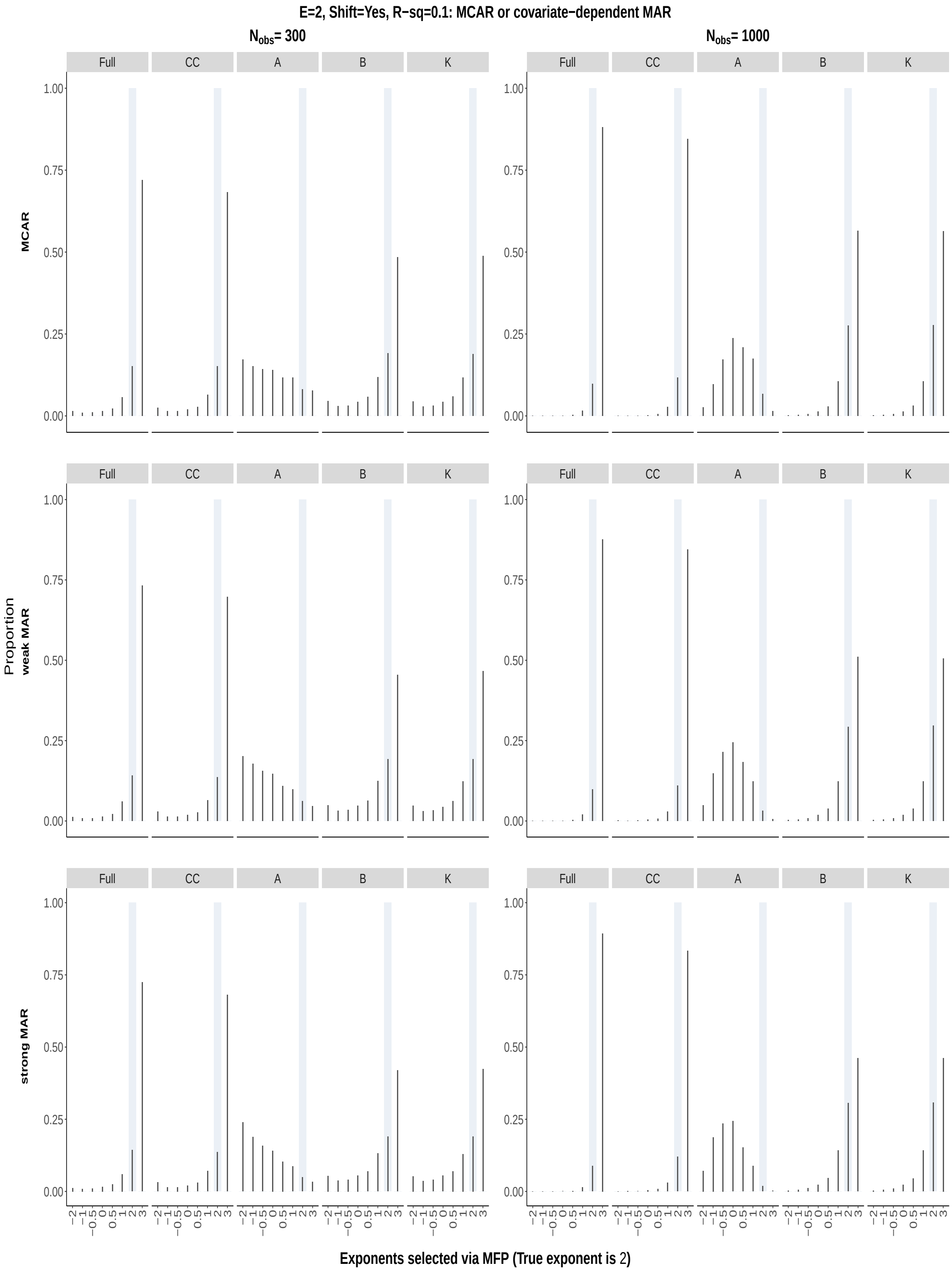

Figure S77: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

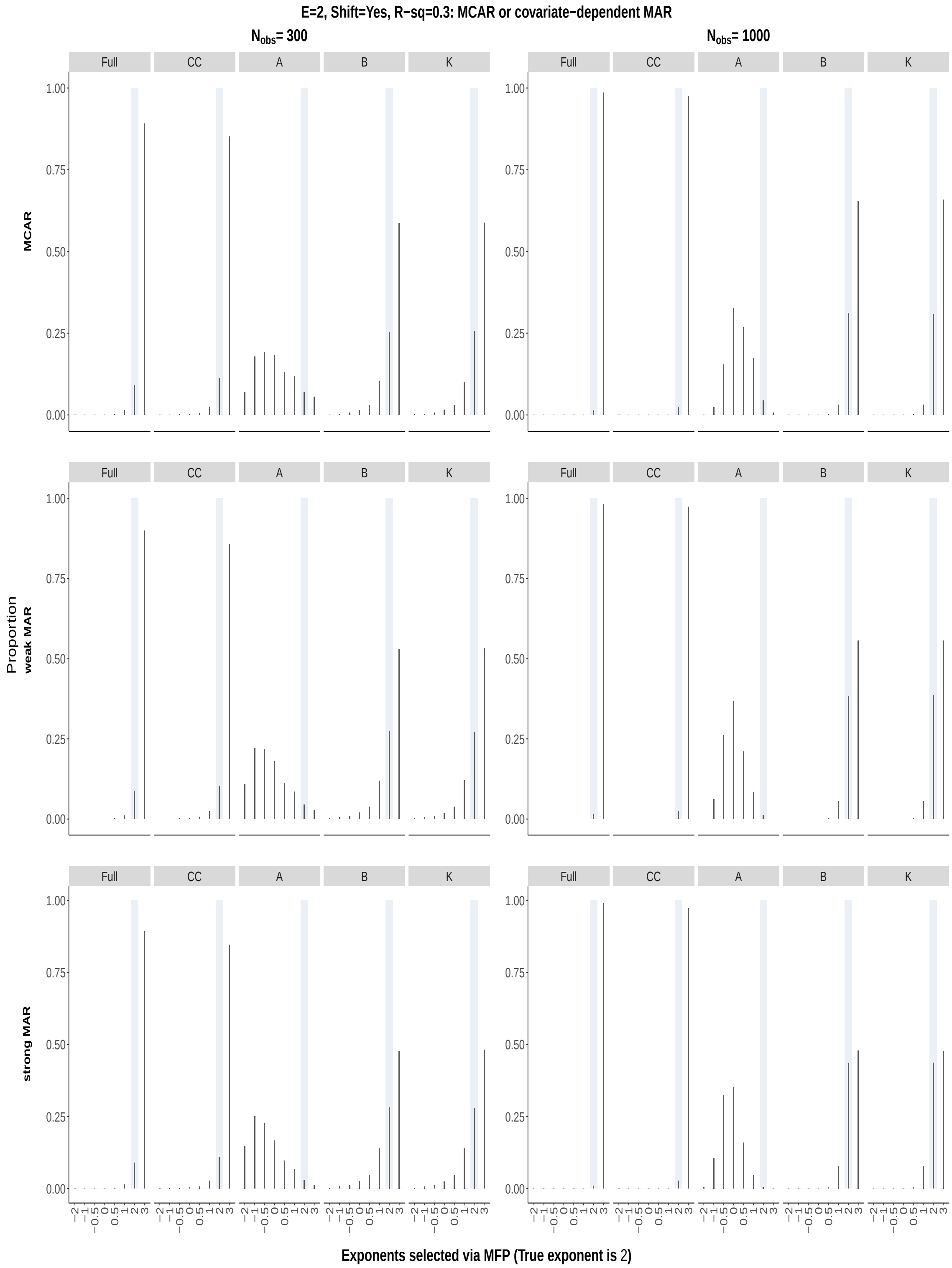

Figure S78: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

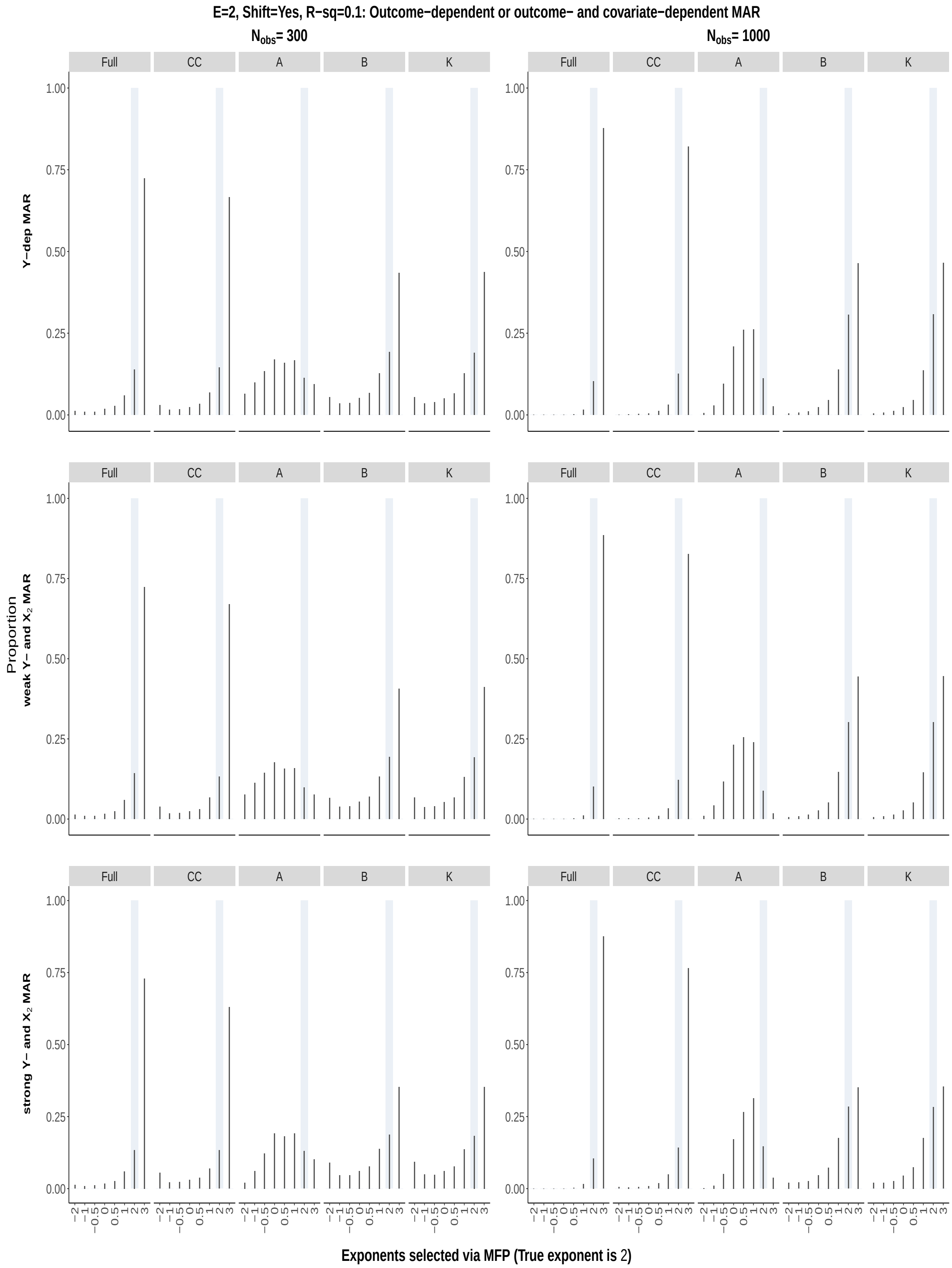

Figure S79: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

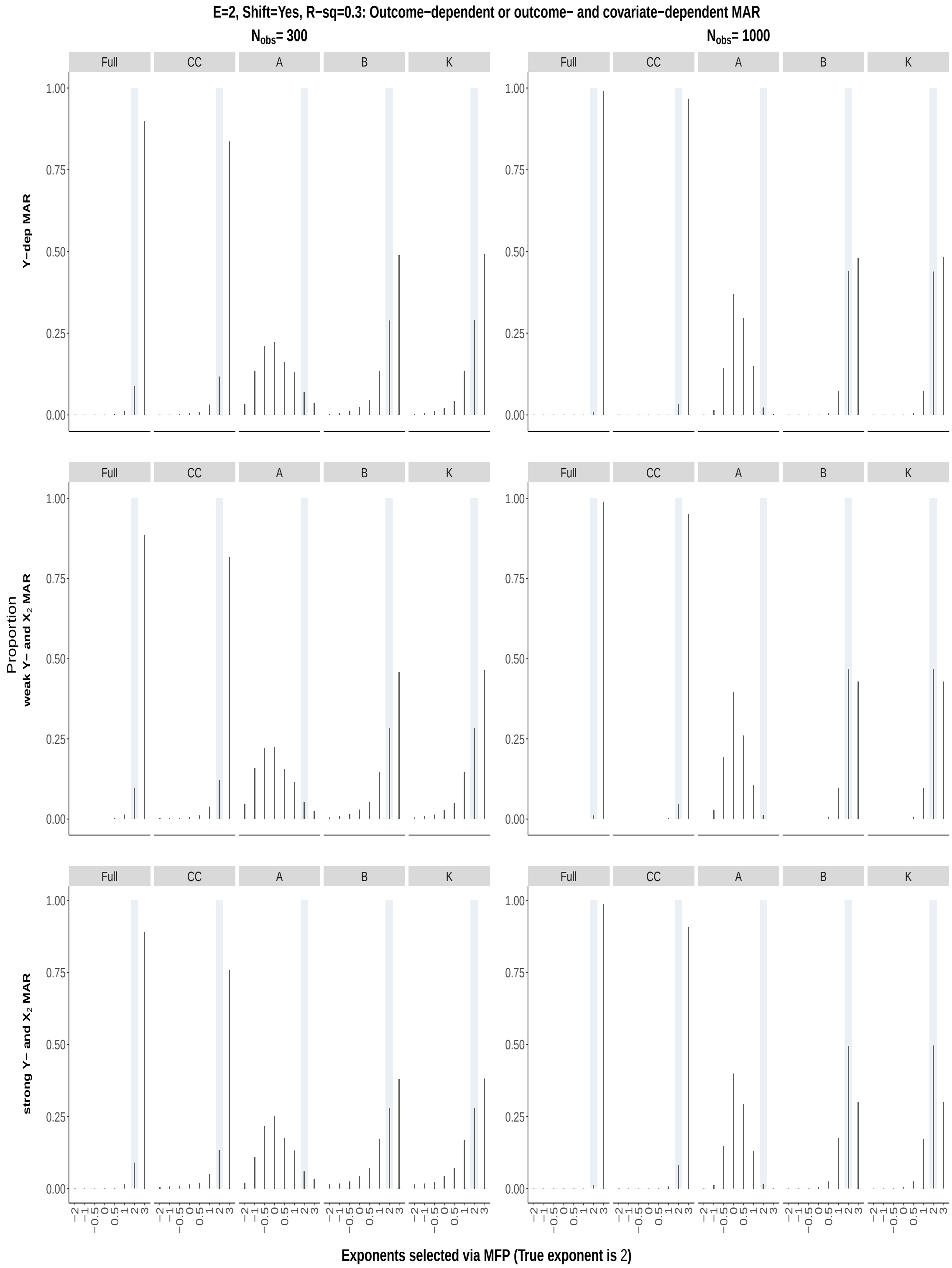

Figure S80: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

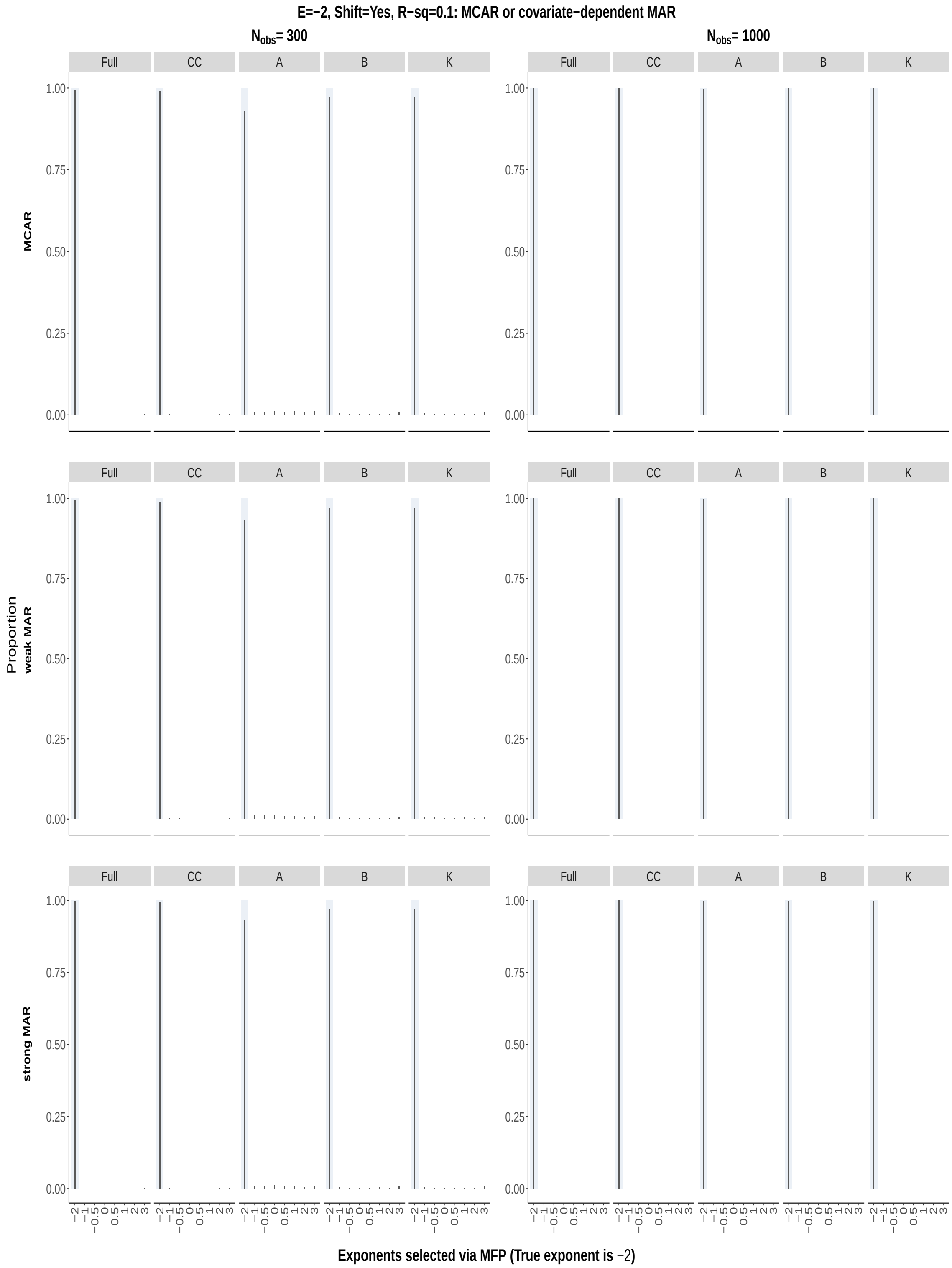

Figure S81: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

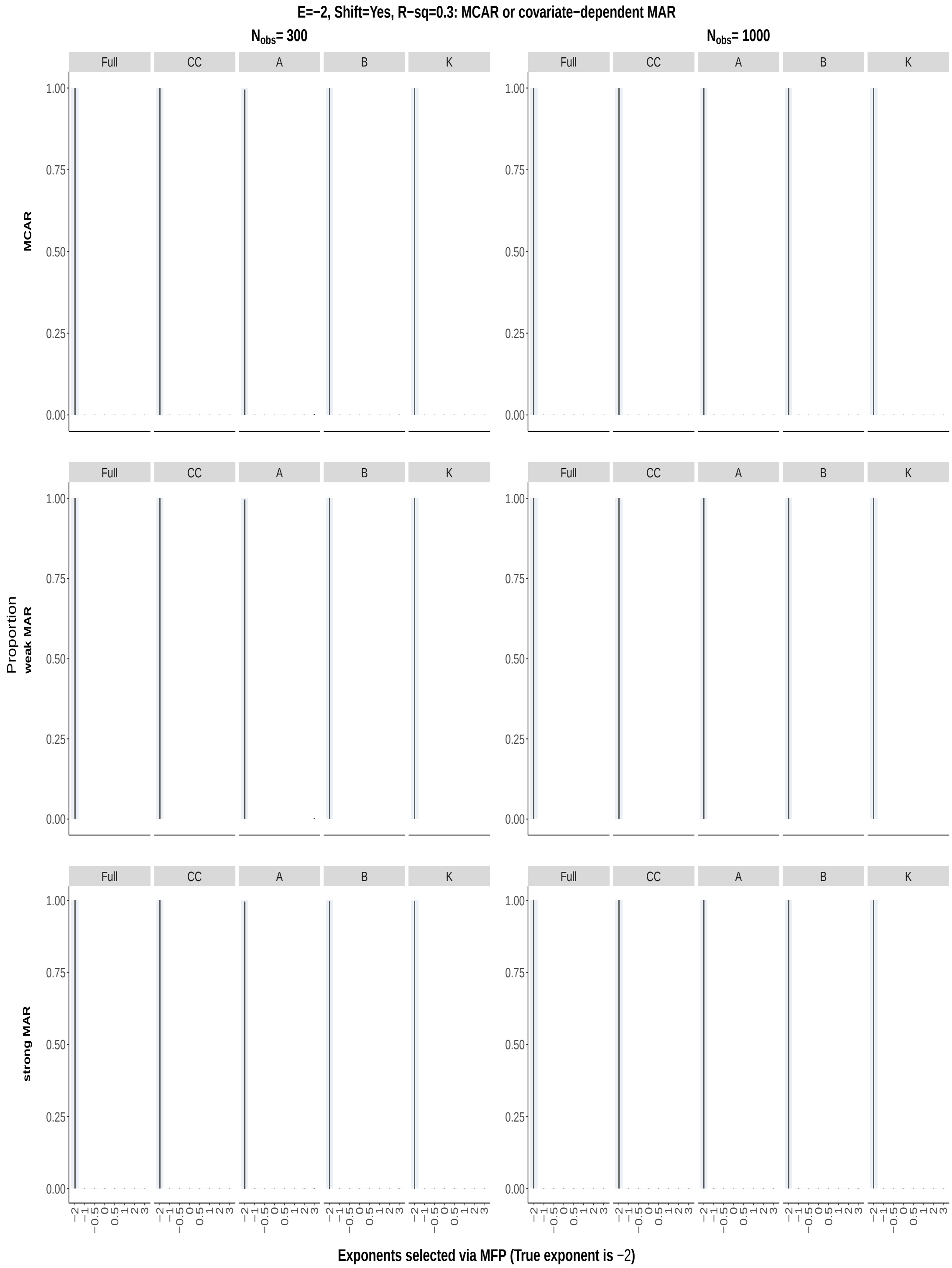

Figure S82: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

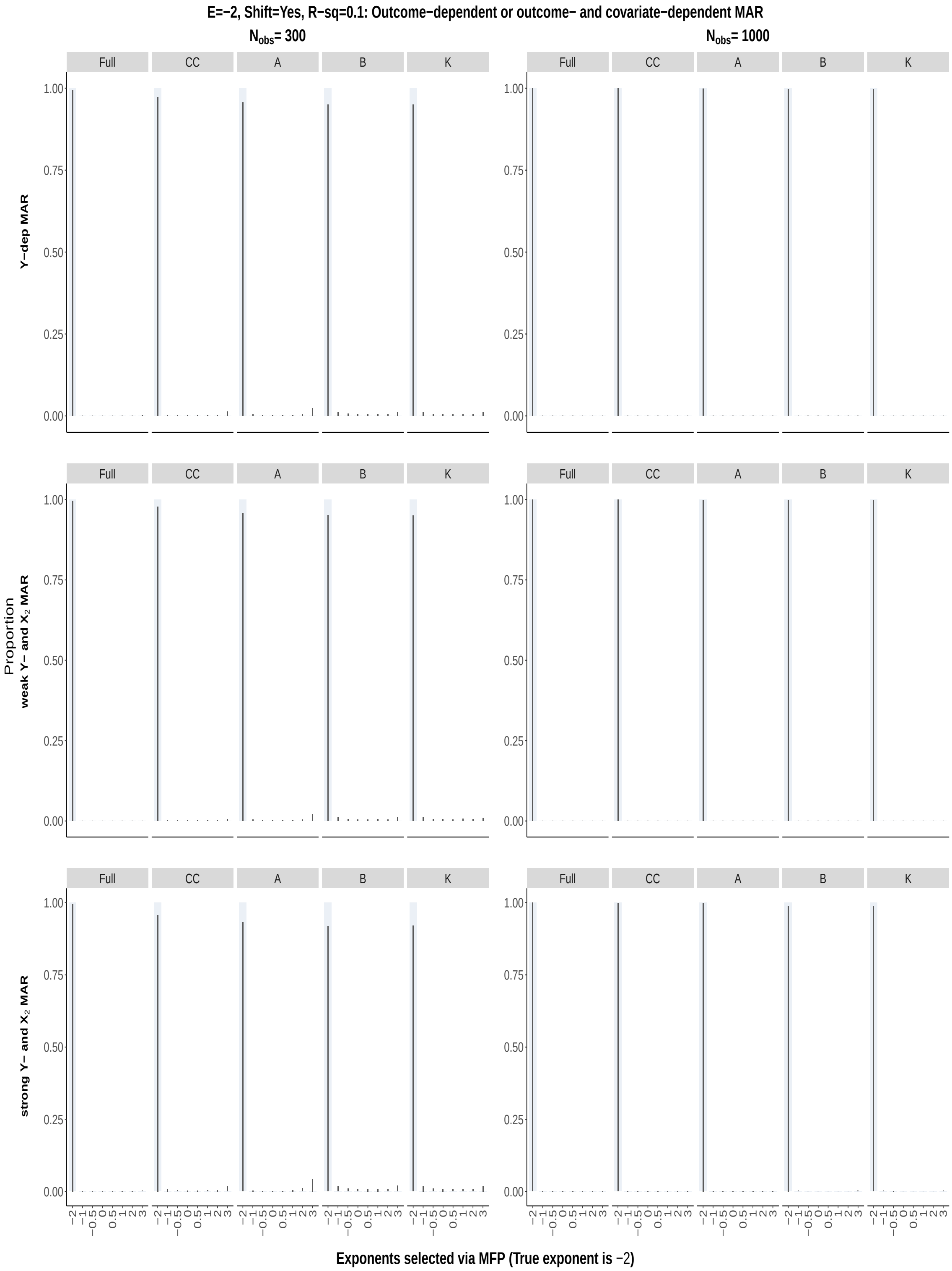

Figure S83: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

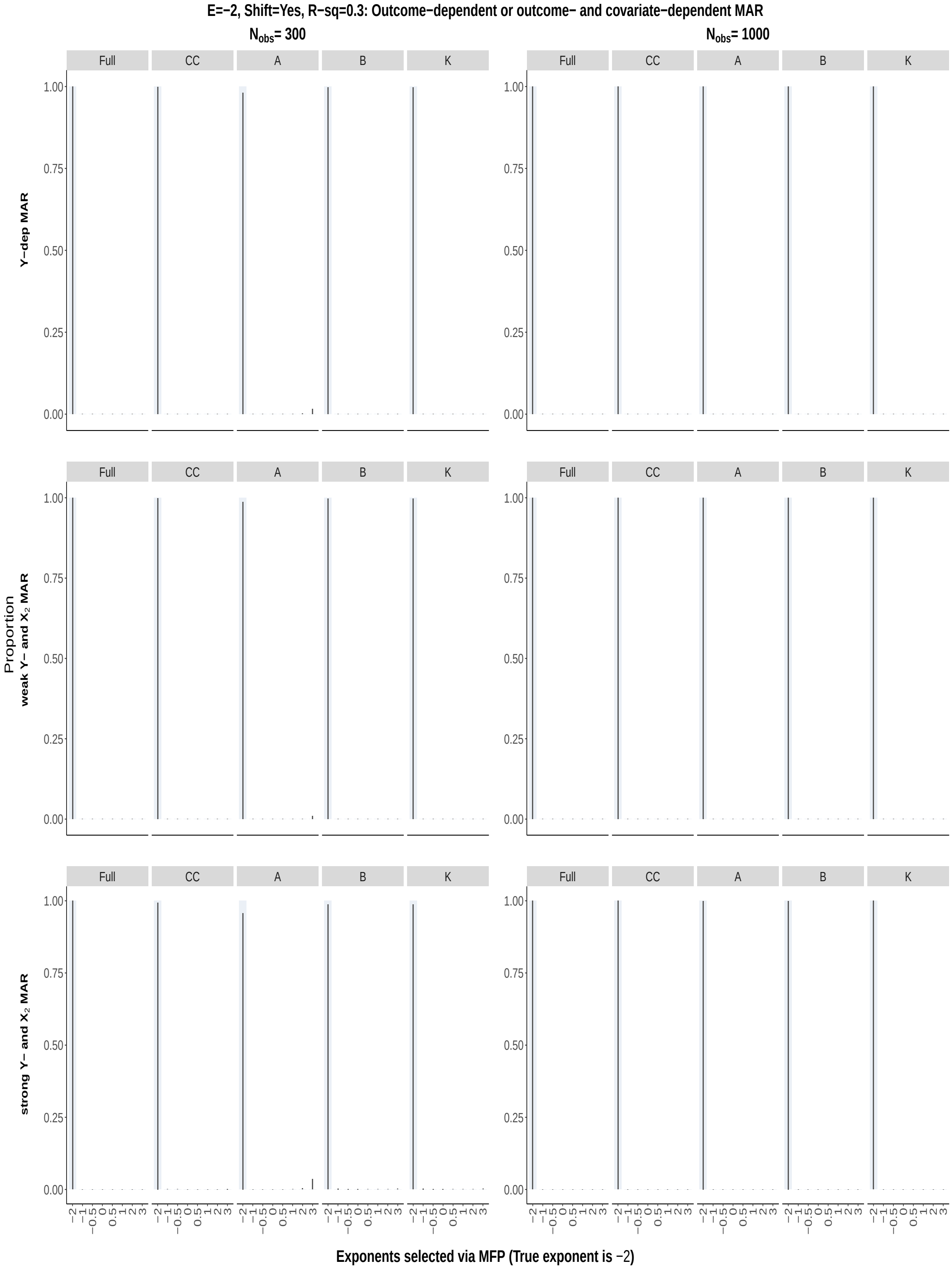

Figure S84: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.4 Cross-validation, $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been used

True exponent is 0

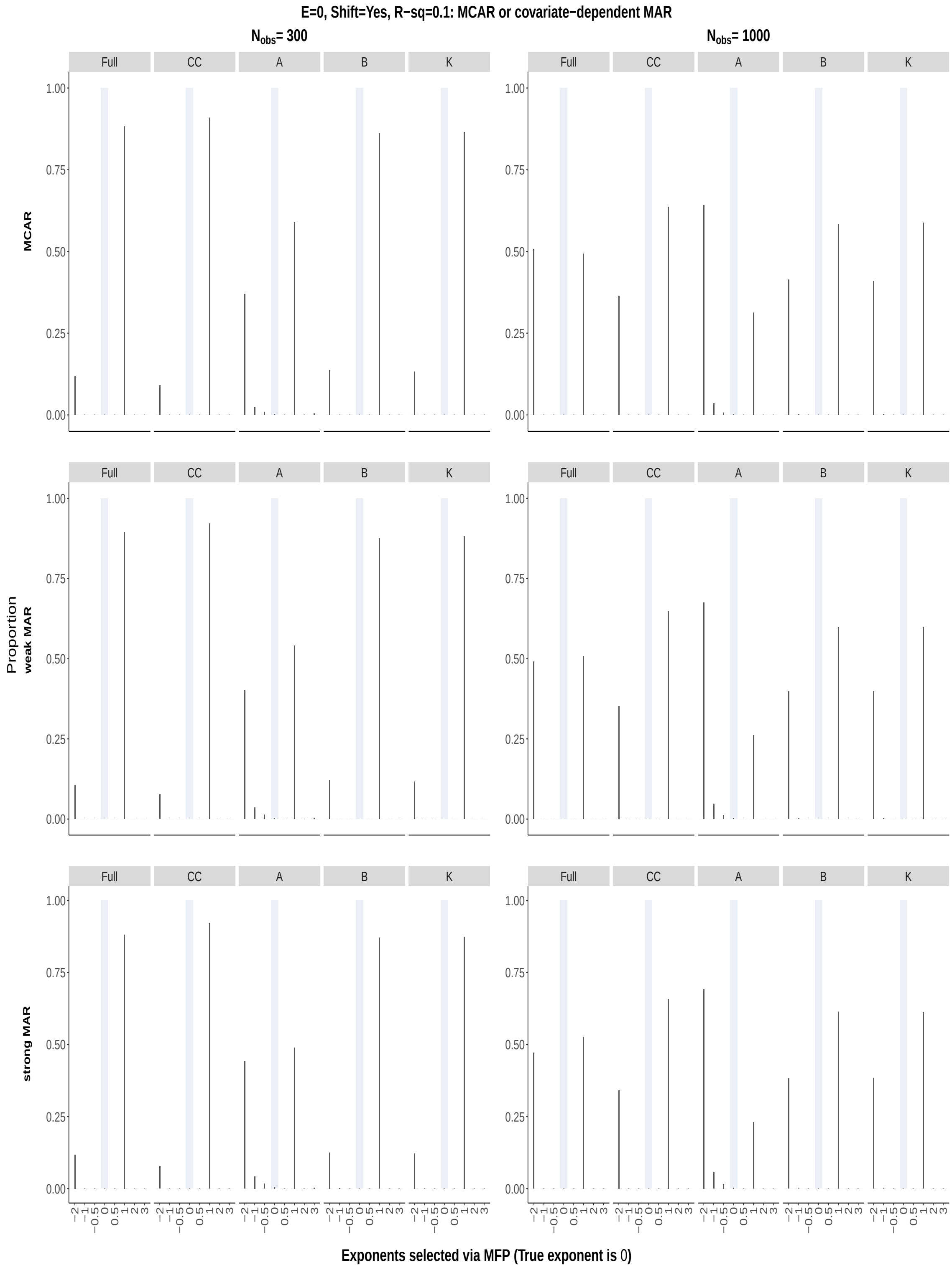

Figure S85: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

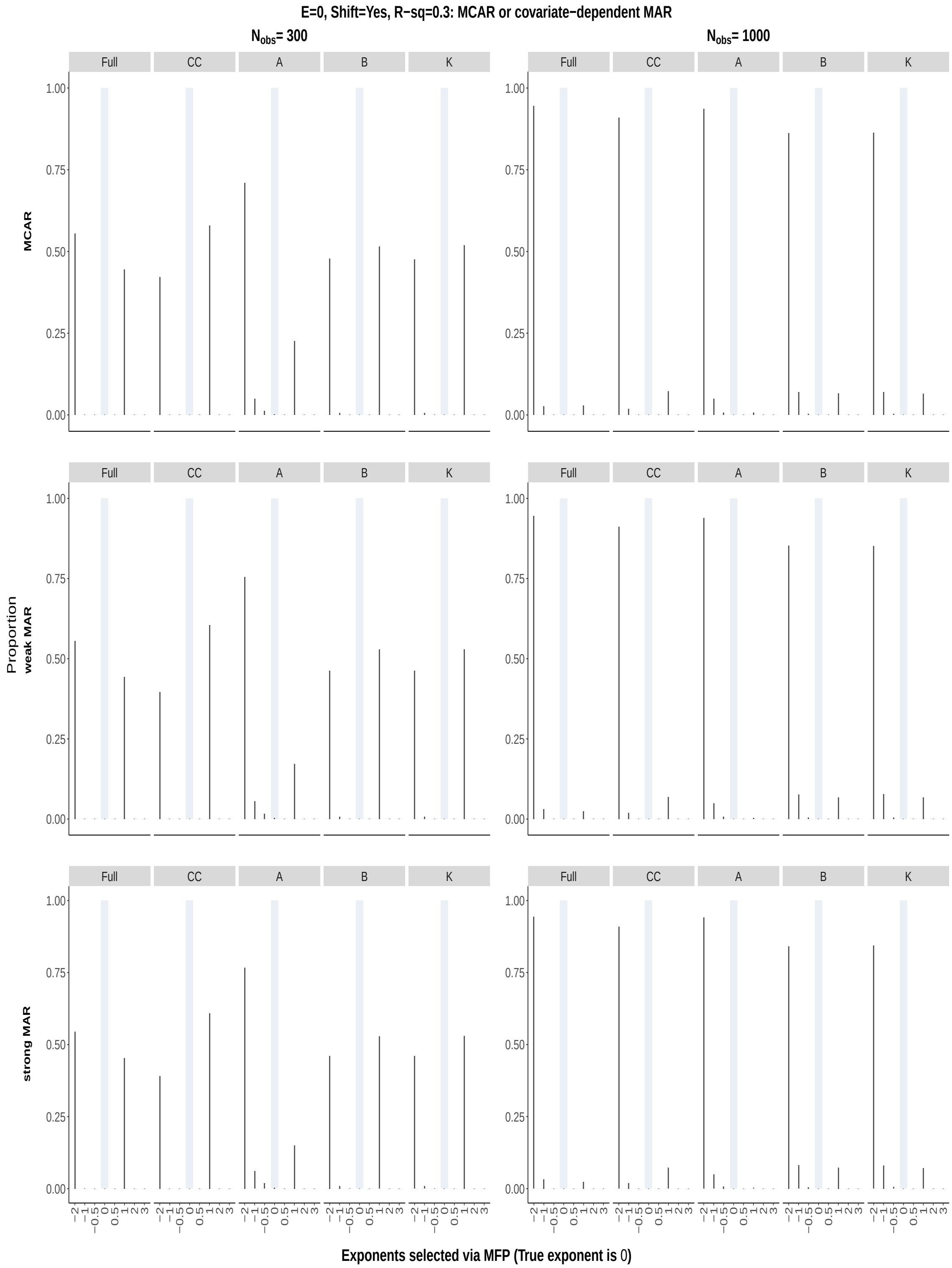

Figure S86: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

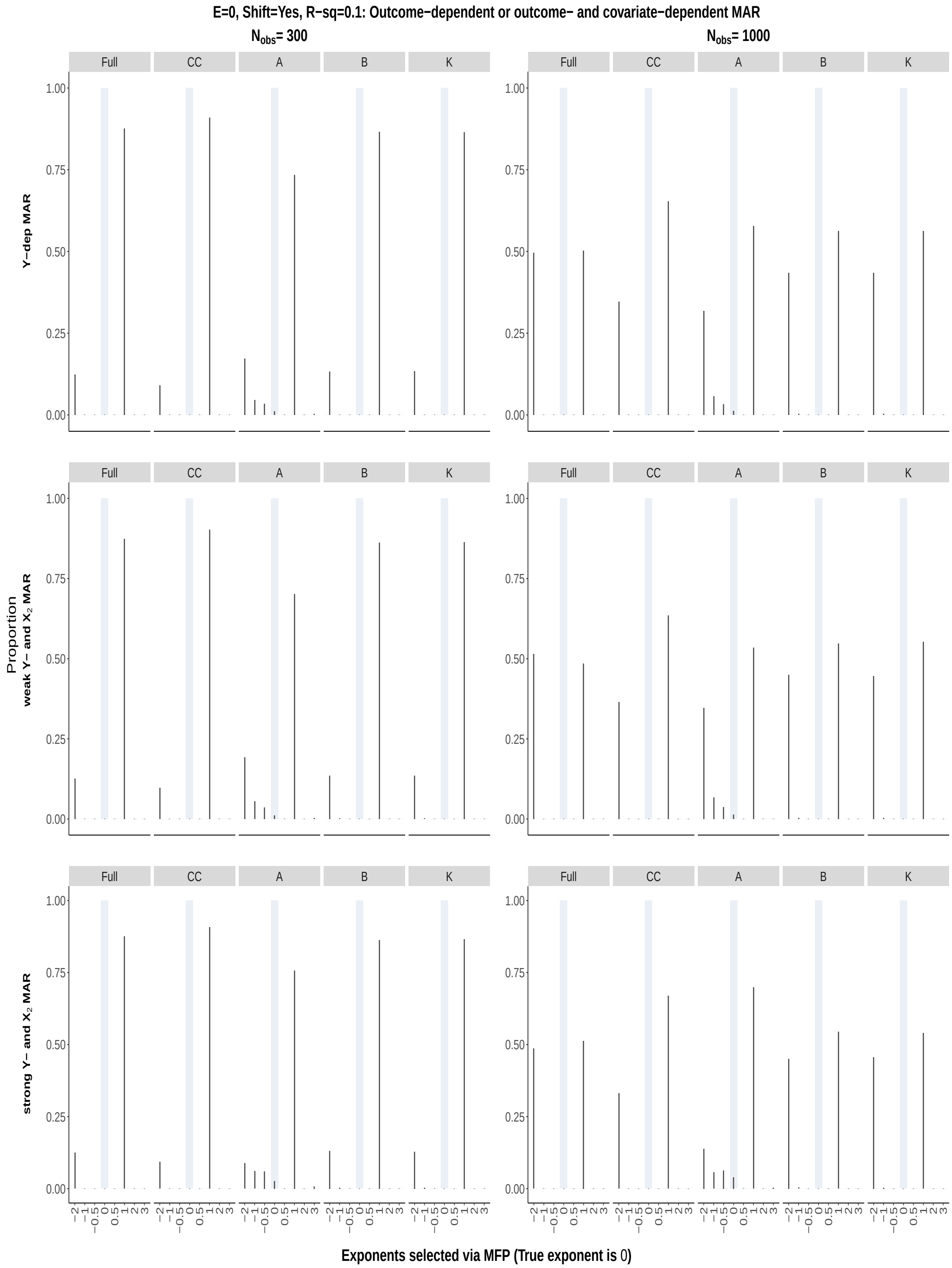

Figure S87: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

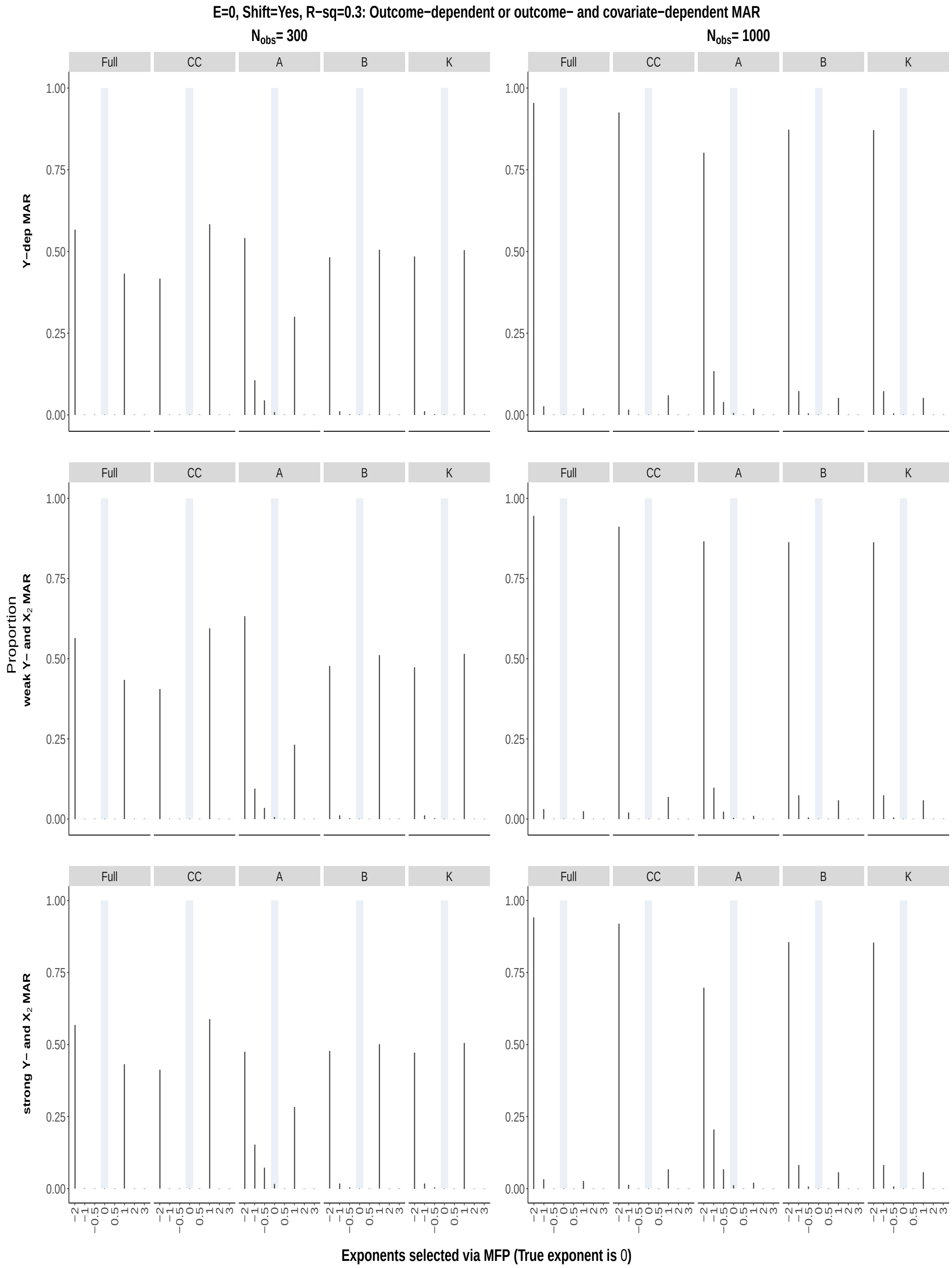

Figure S88: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

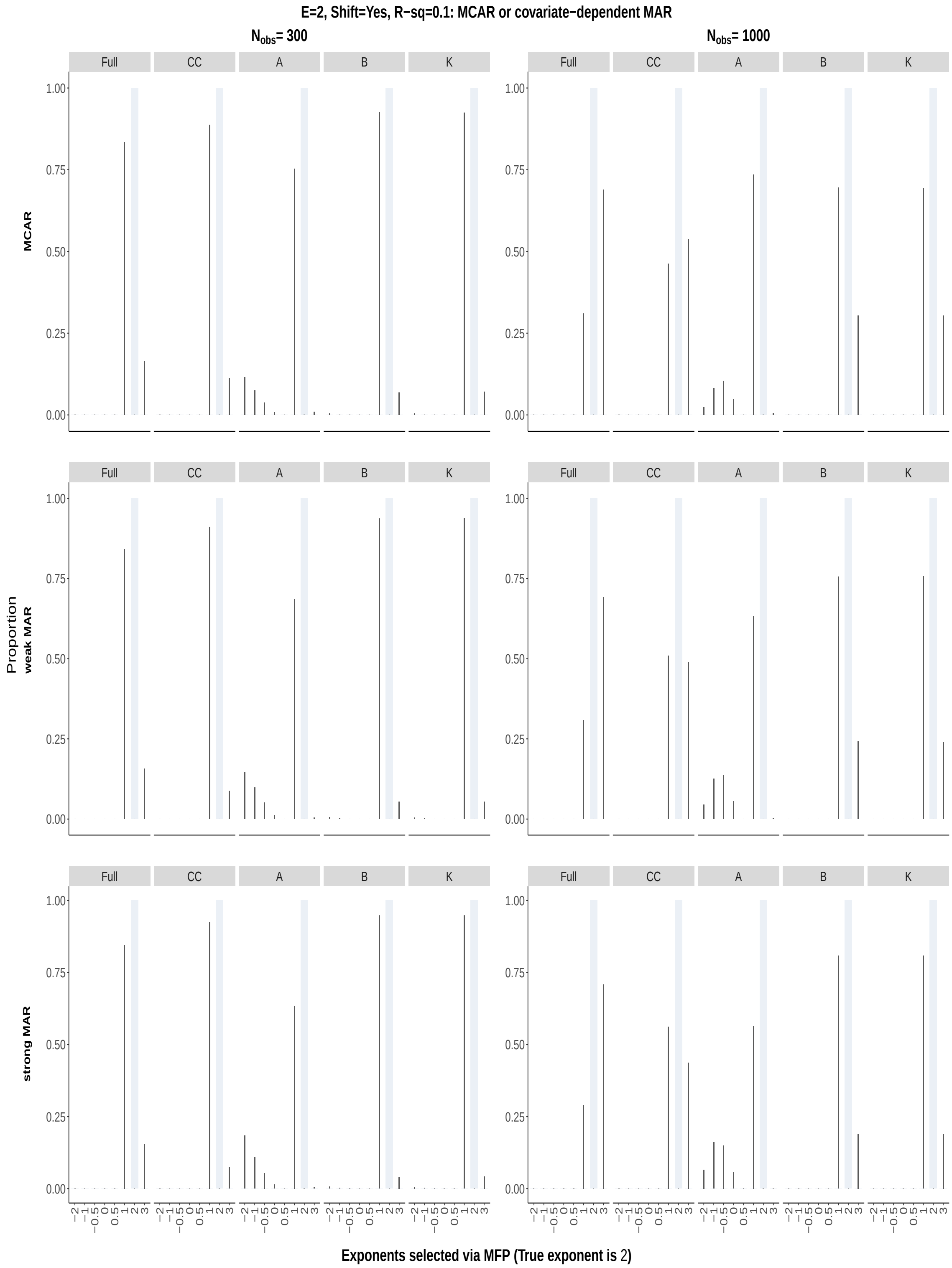

Figure S89: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

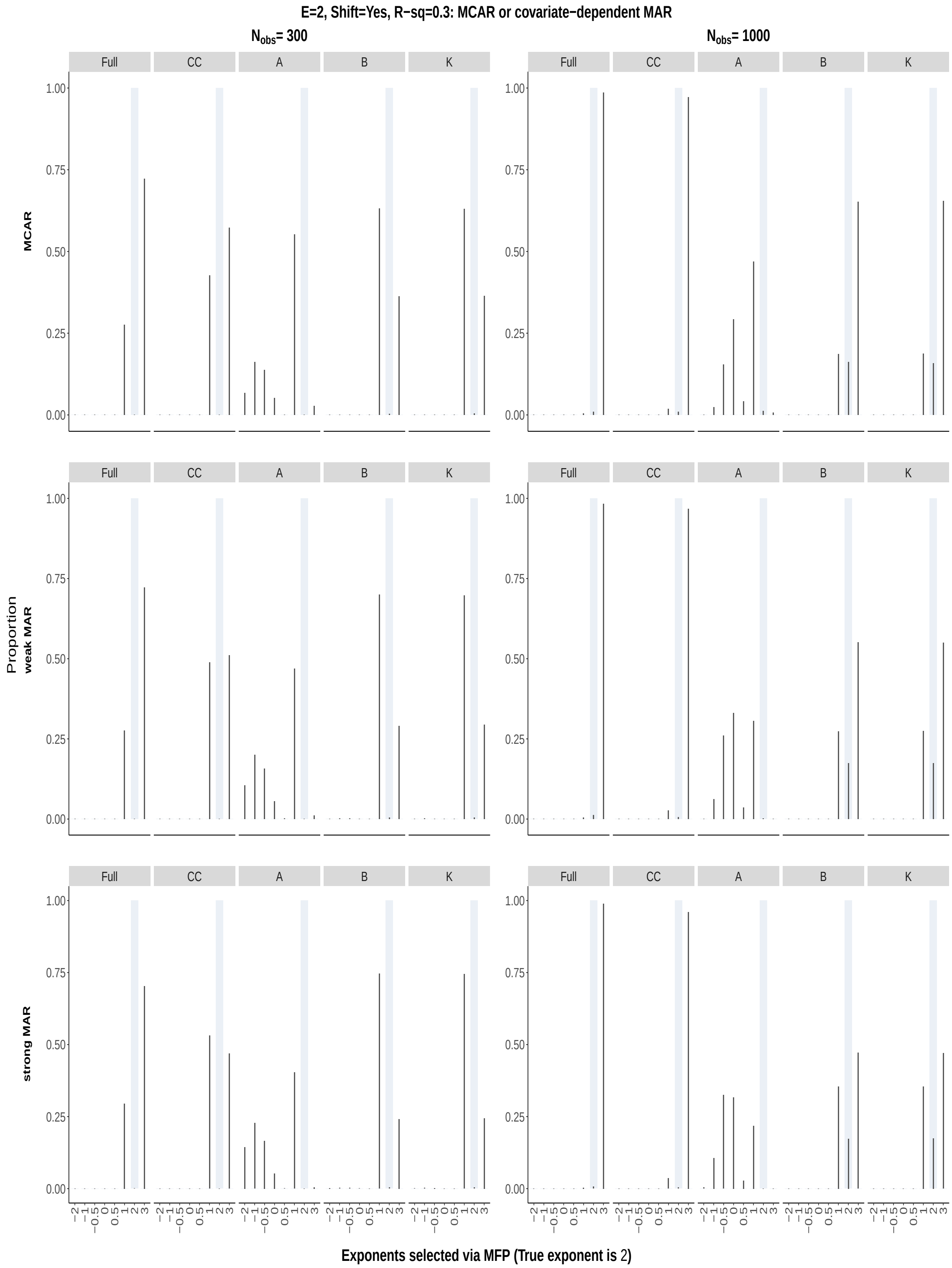

Figure S90: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

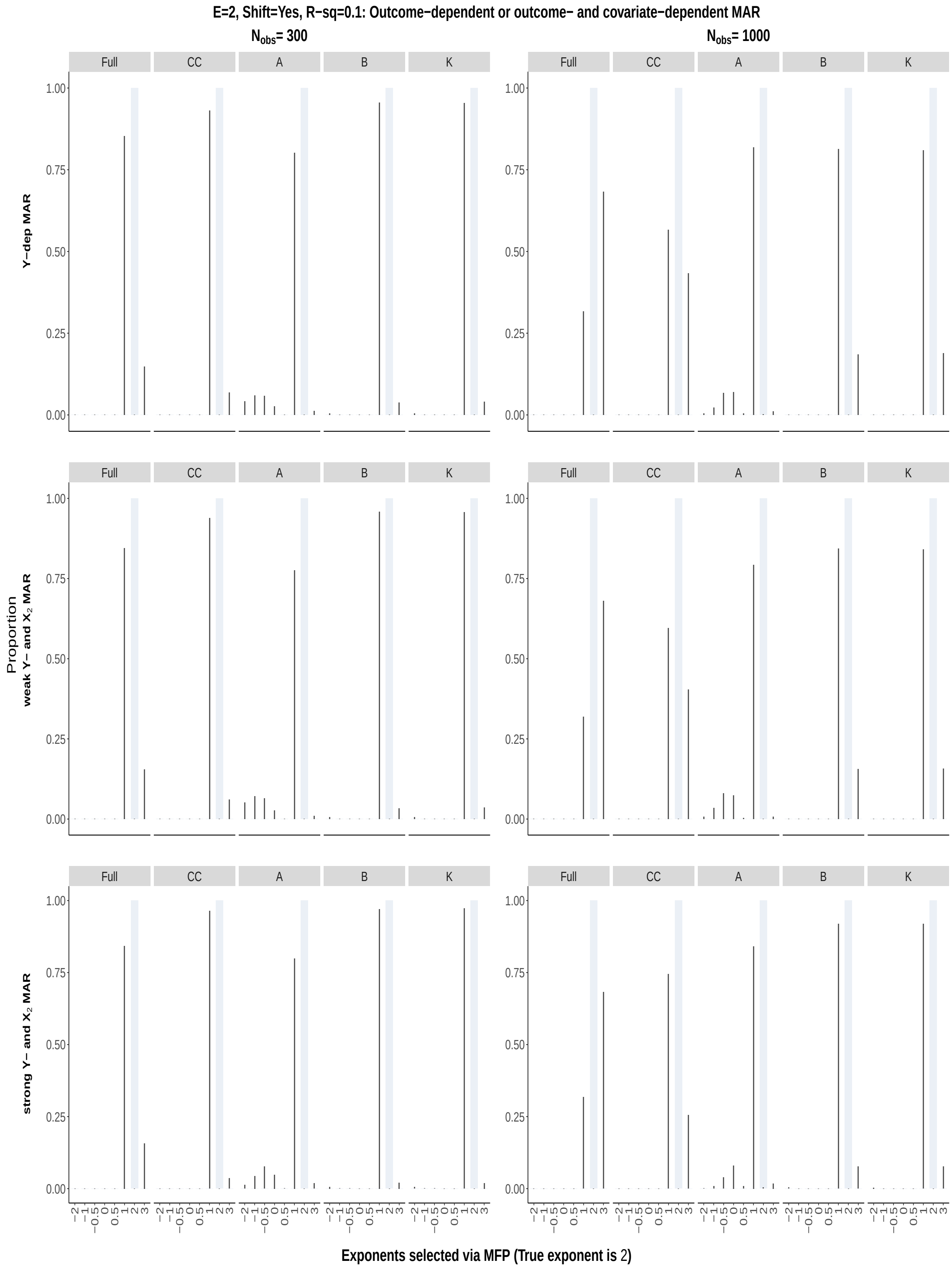

Figure S91: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

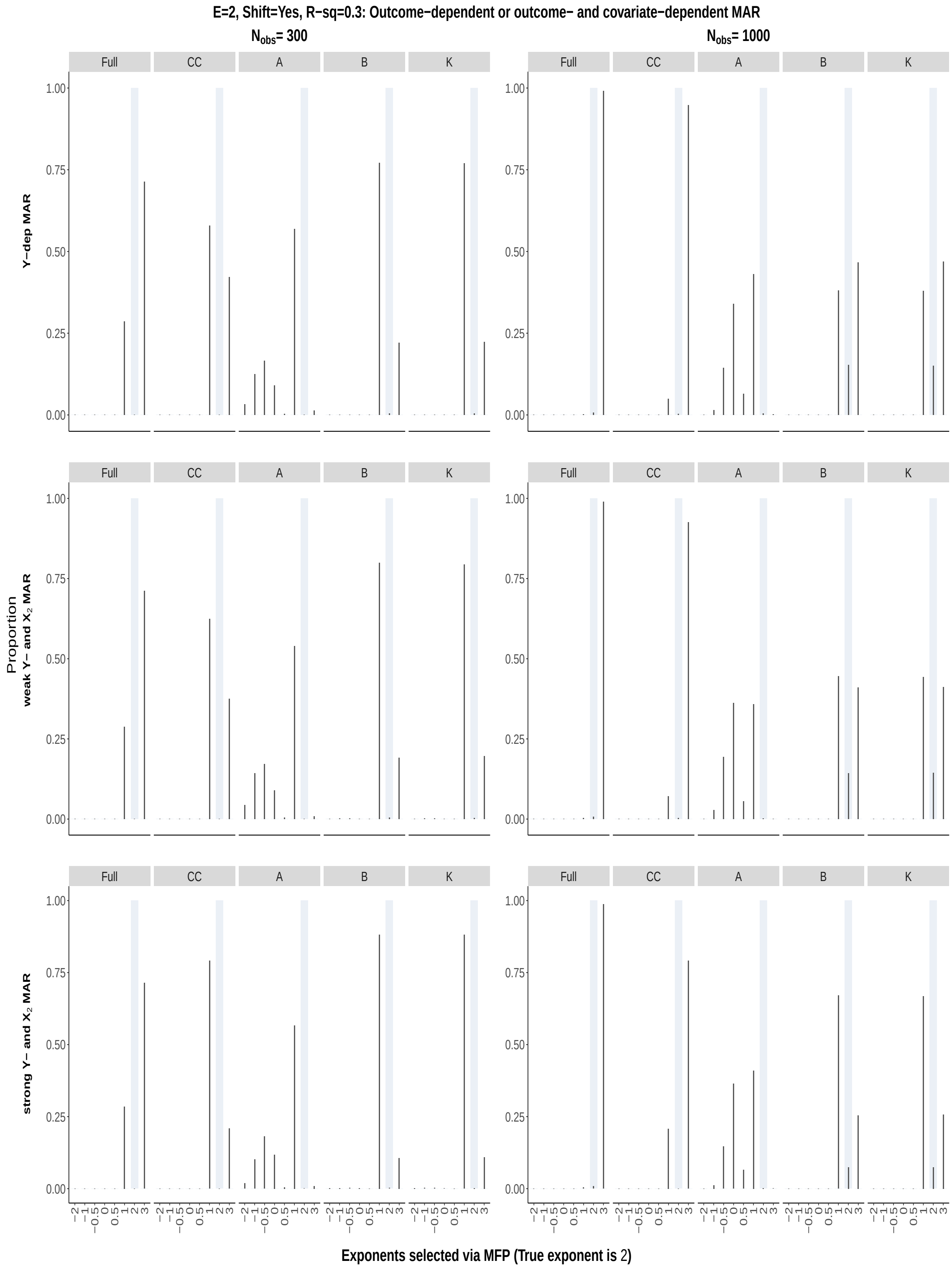

Figure S92: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

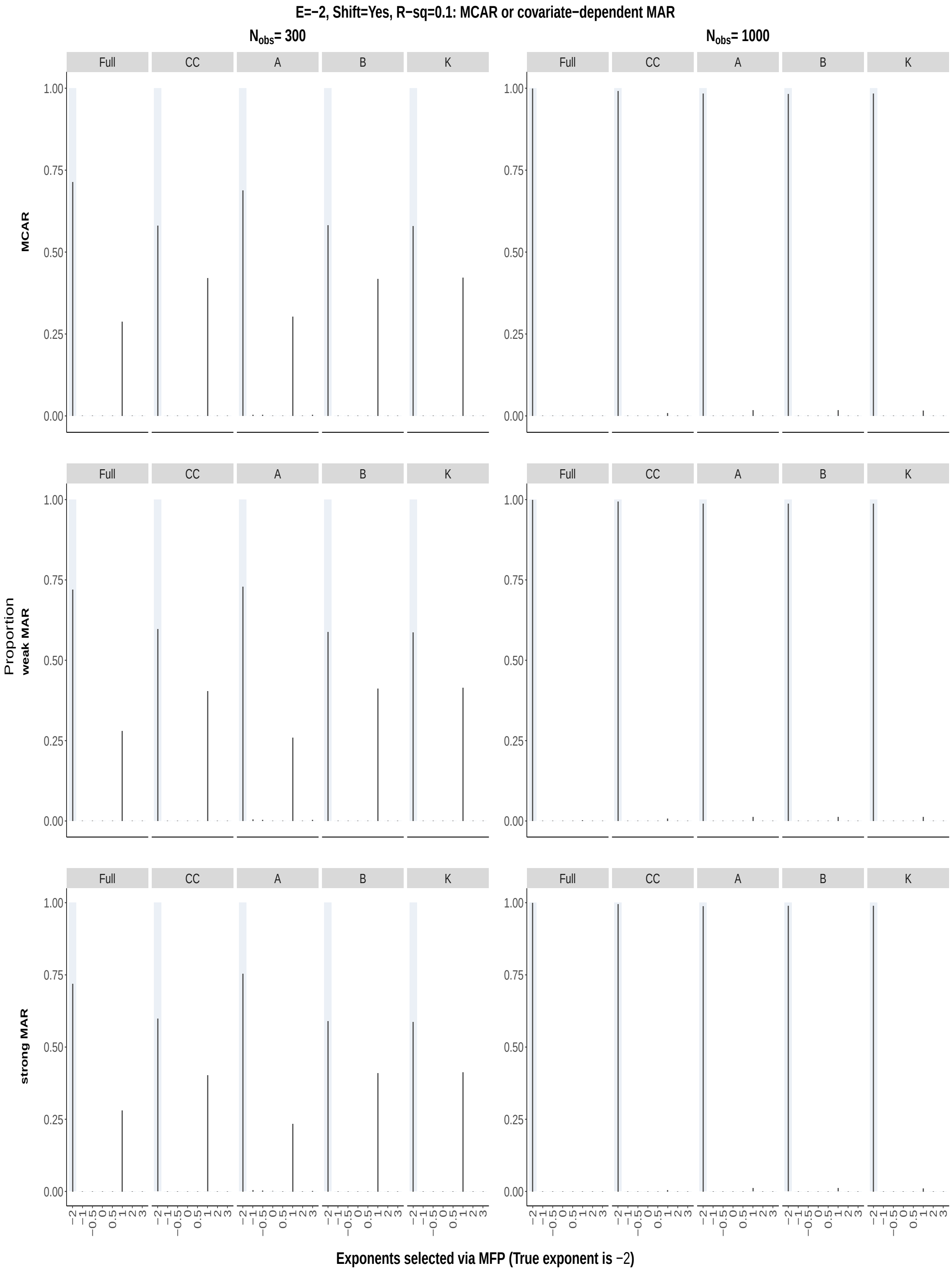

Figure S93: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

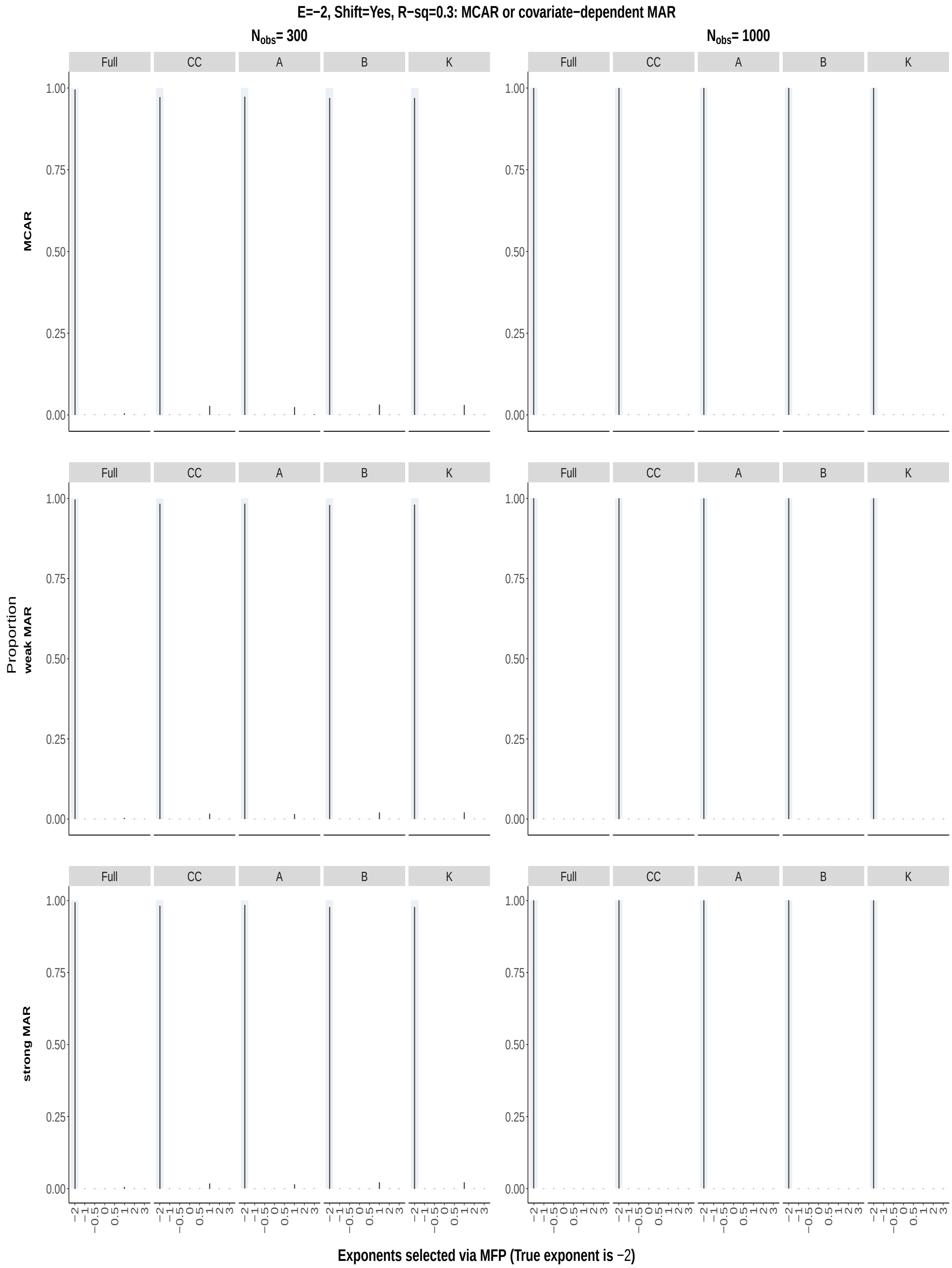

Figure S94: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

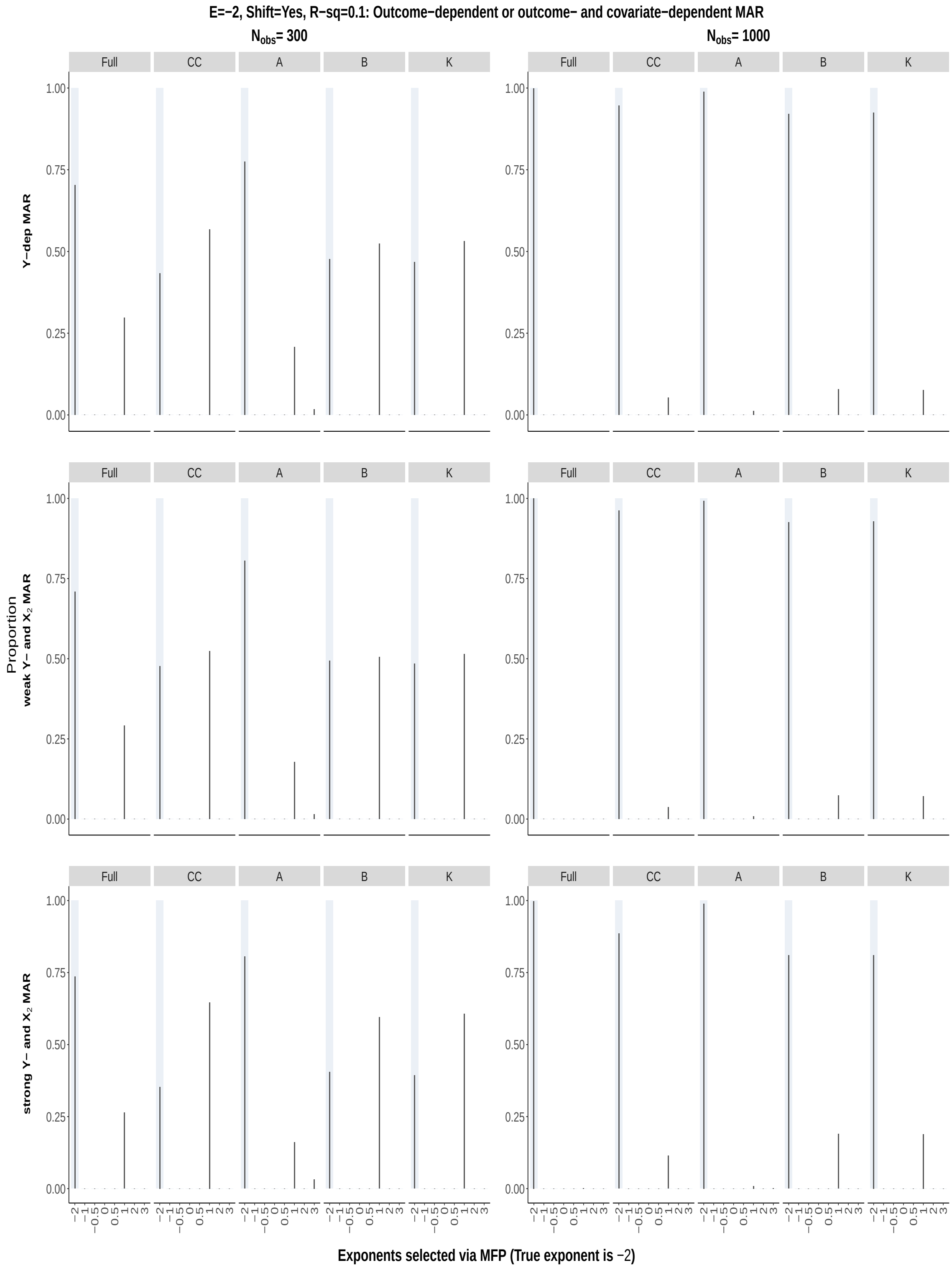

Figure S95: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

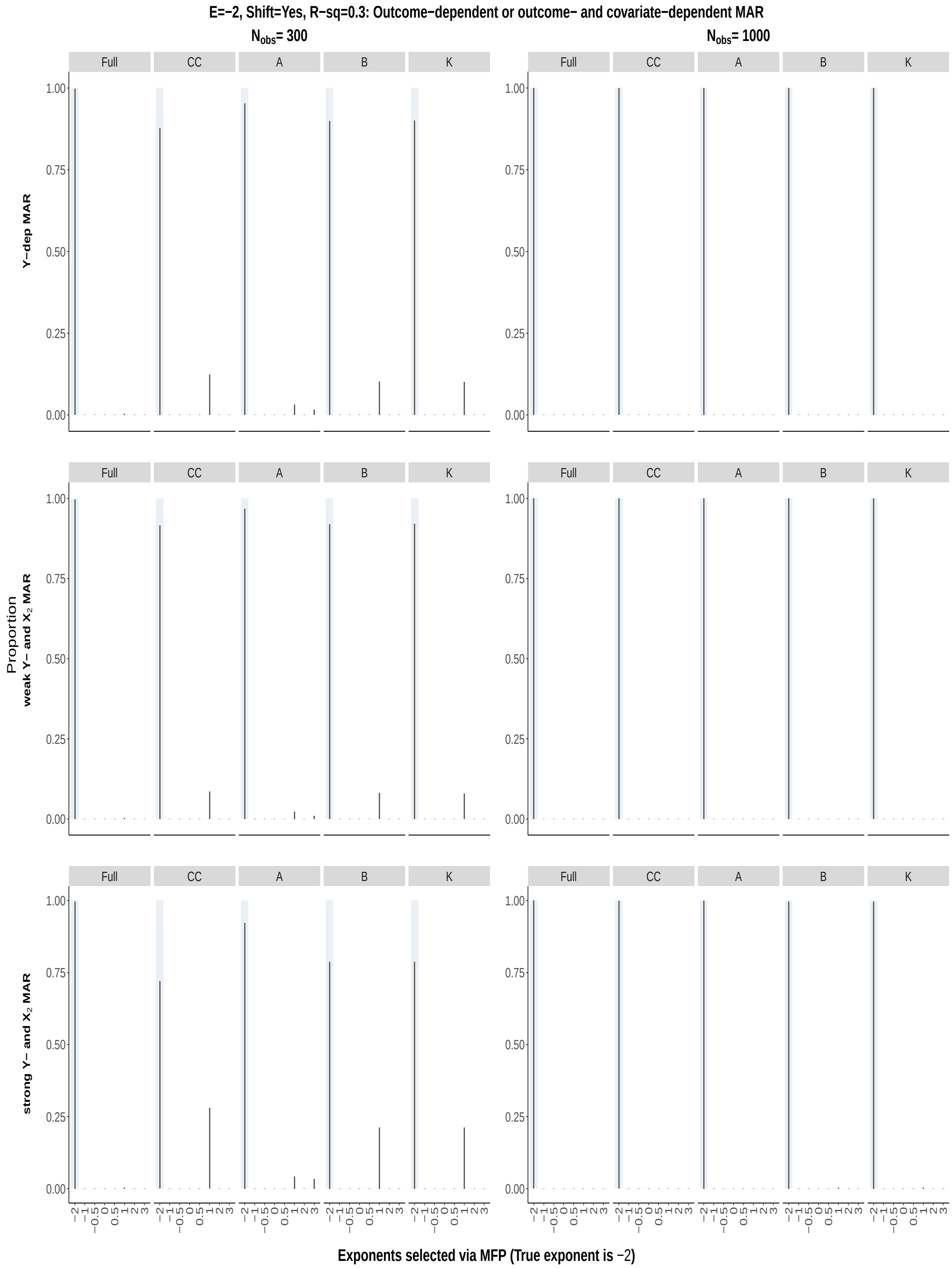

Figure S96: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.5 Cross-validation, $\beta_2 = 0$, $\alpha_E = 1$ and no origin-shift

True exponent is 0

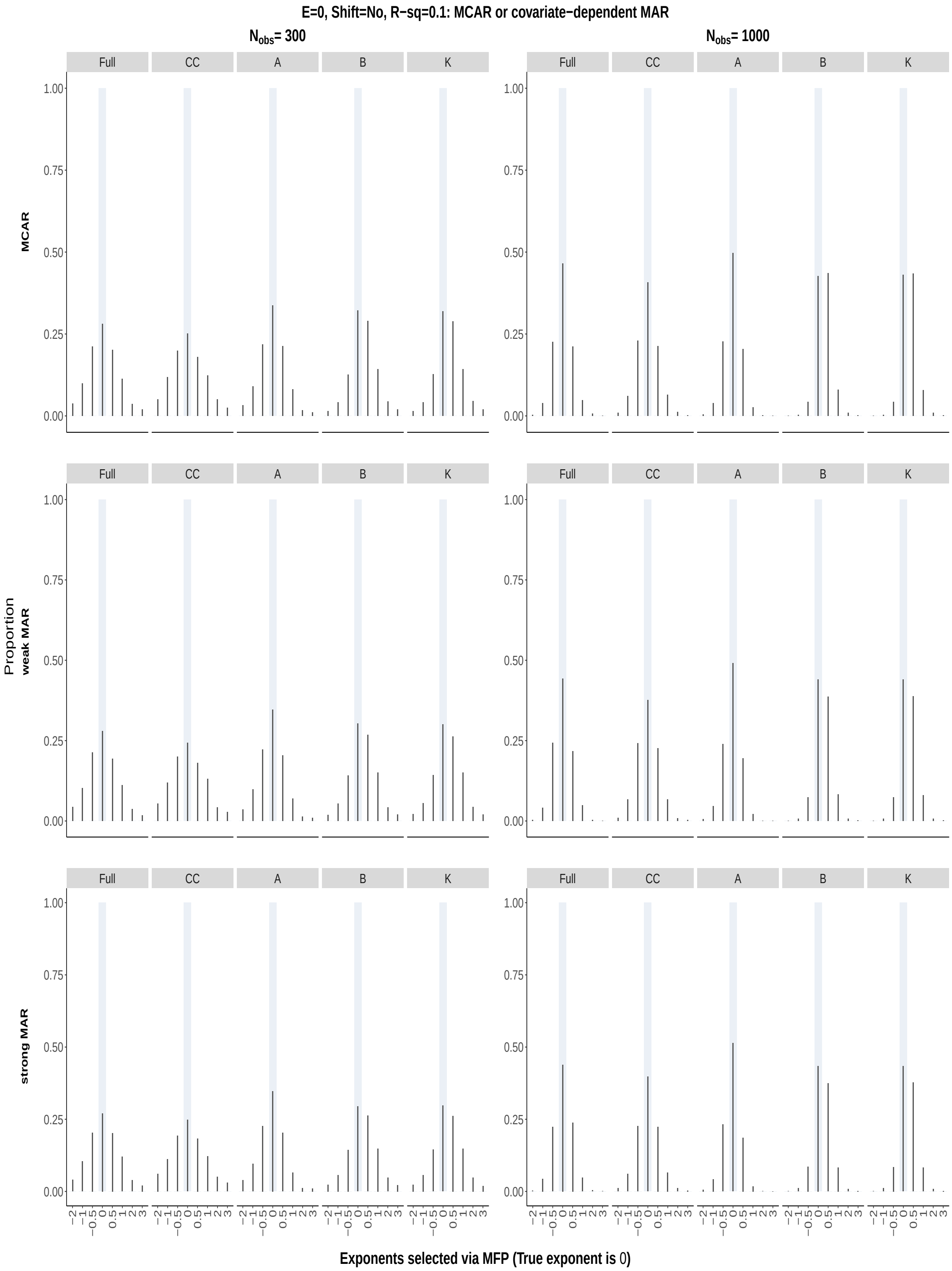

Figure S97: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

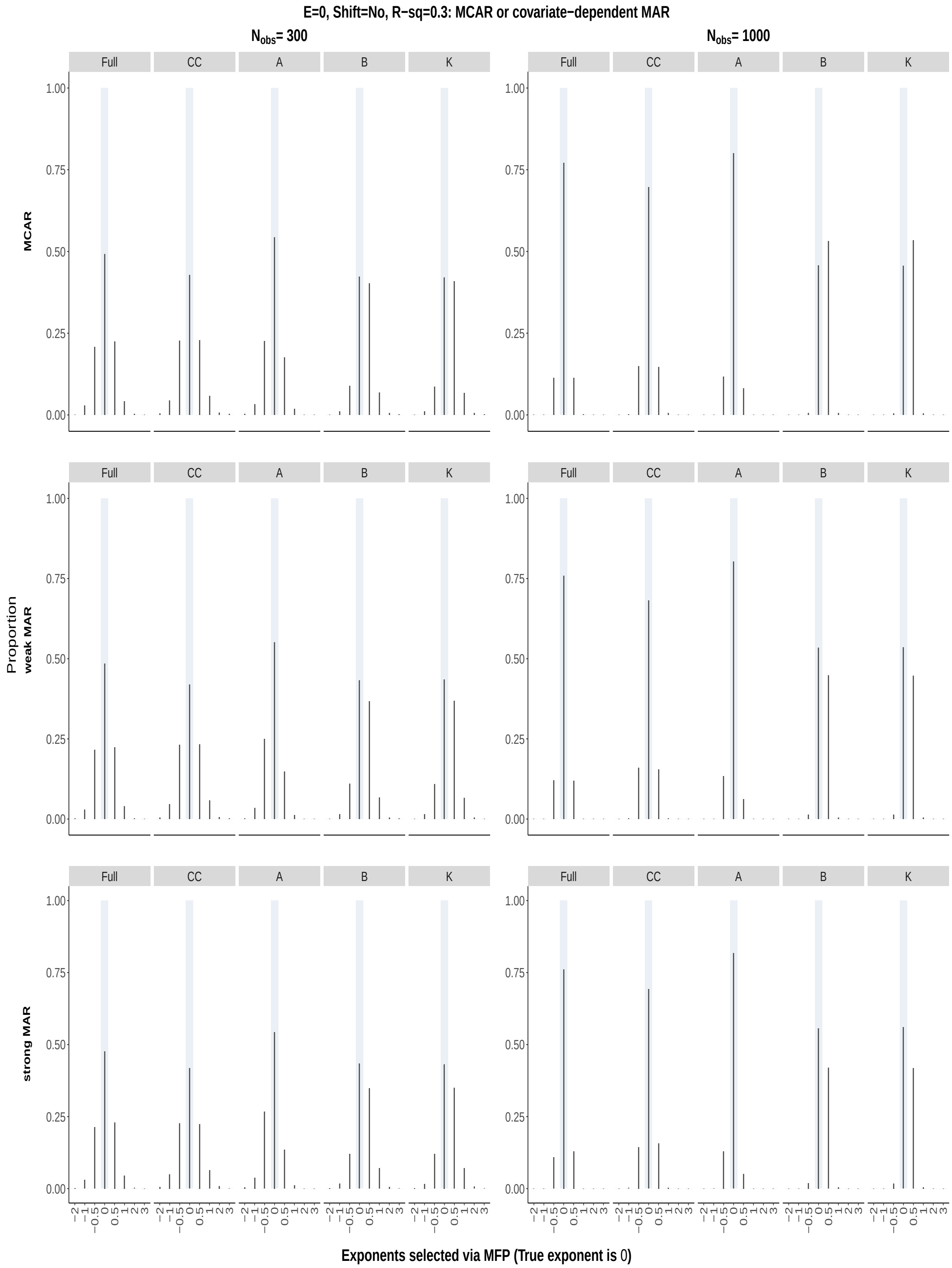

Figure S98: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

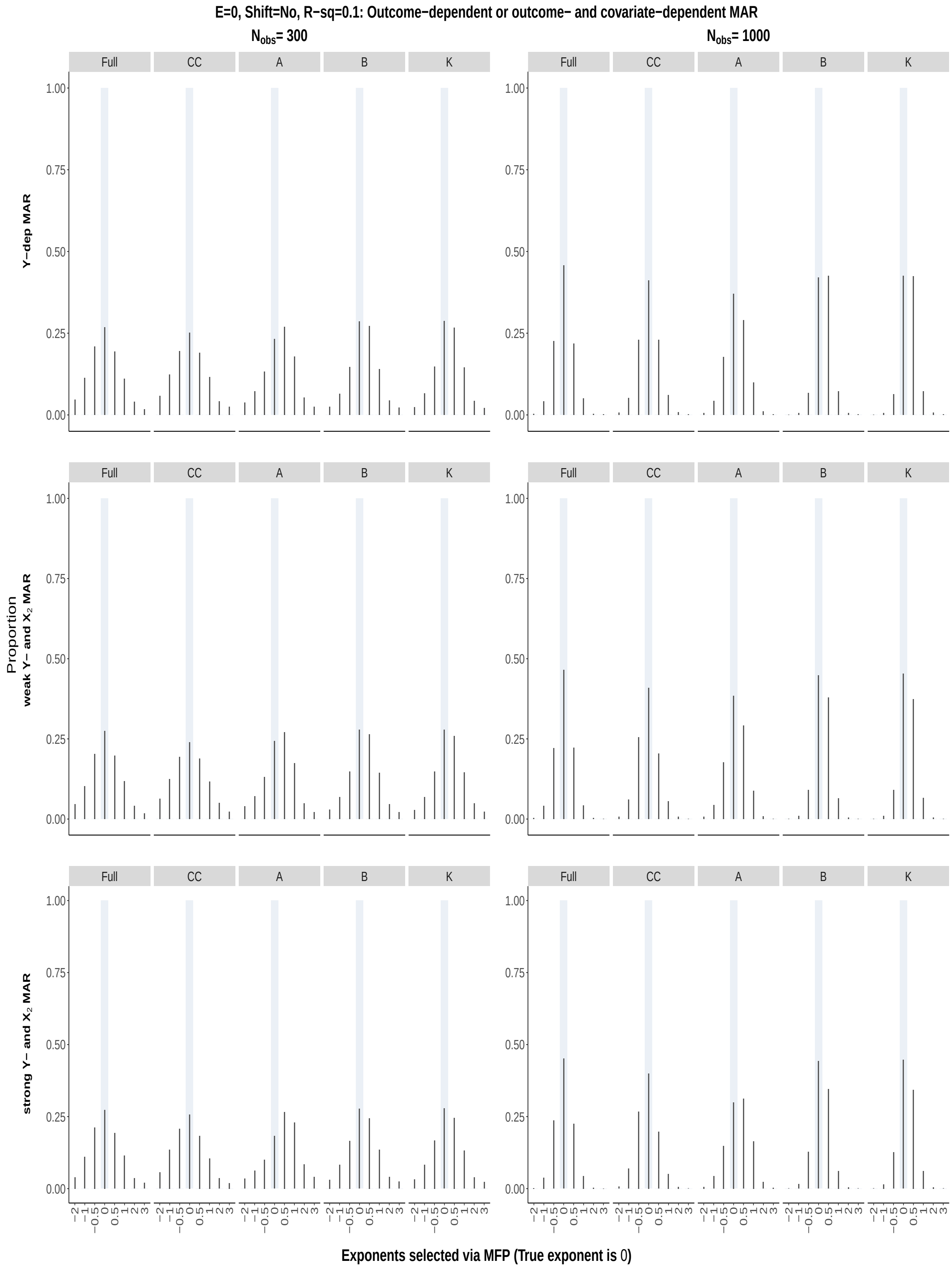

Figure S99: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

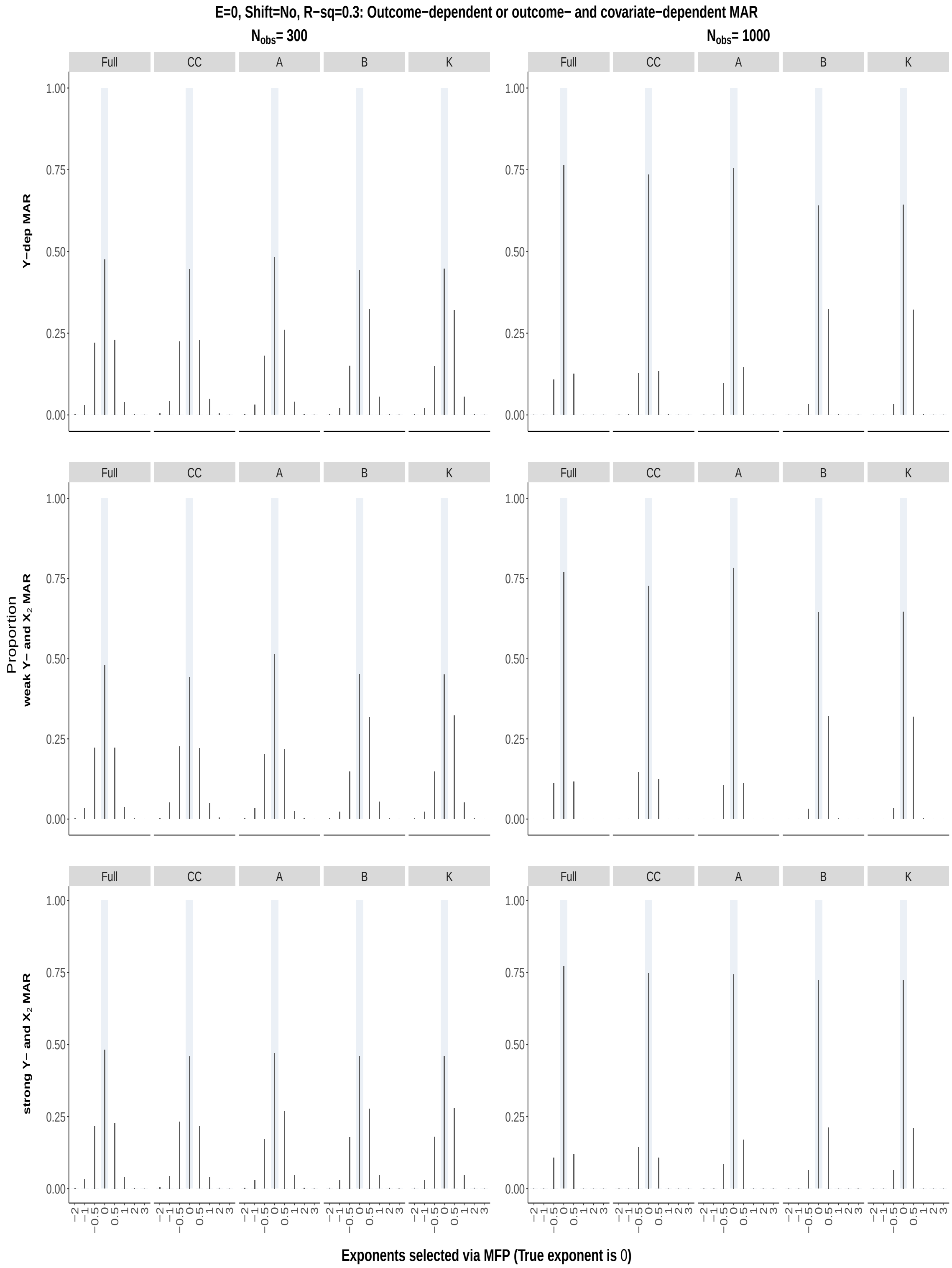

Figure S100: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

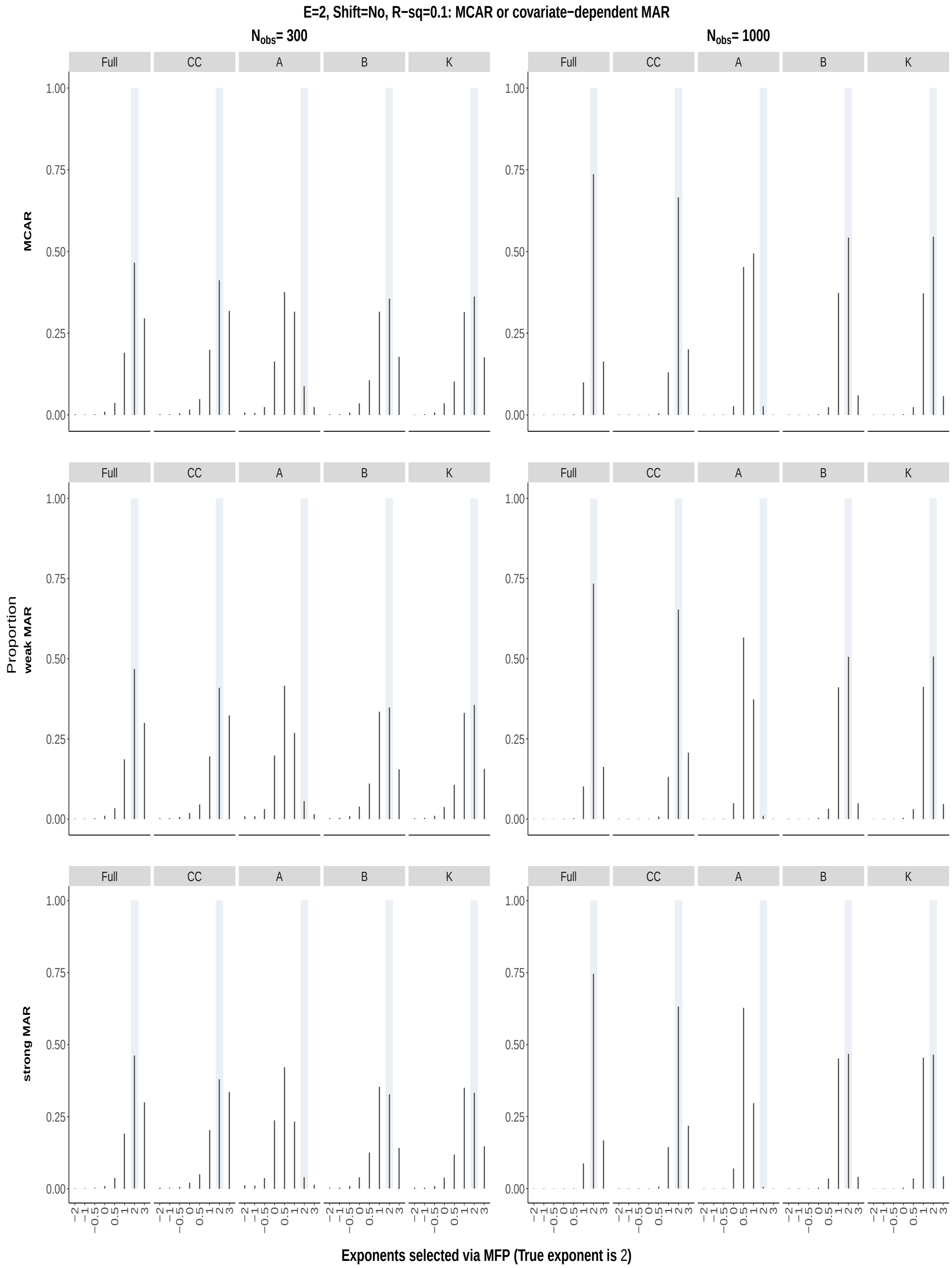

Figure S101: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

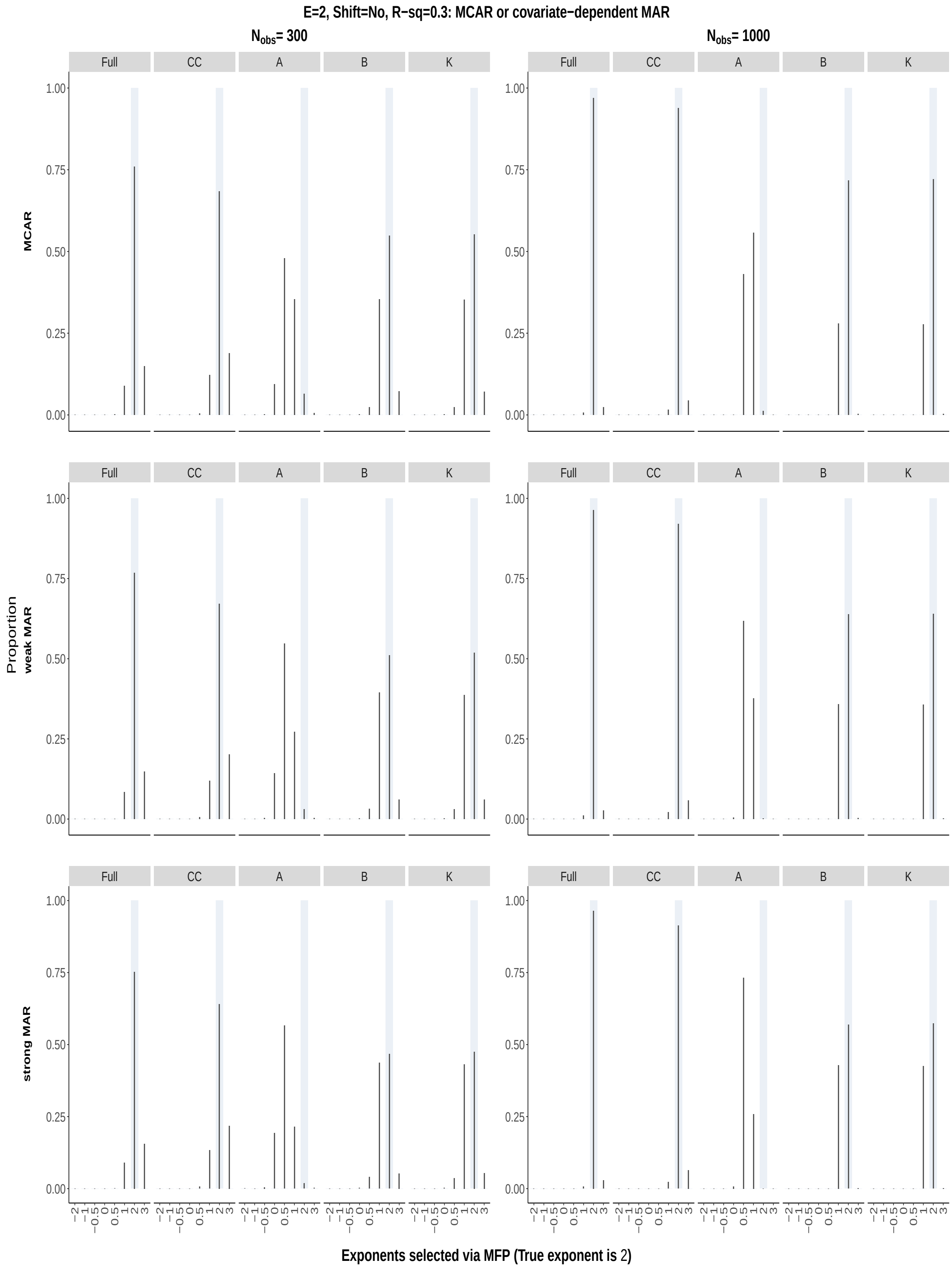

Figure S102: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

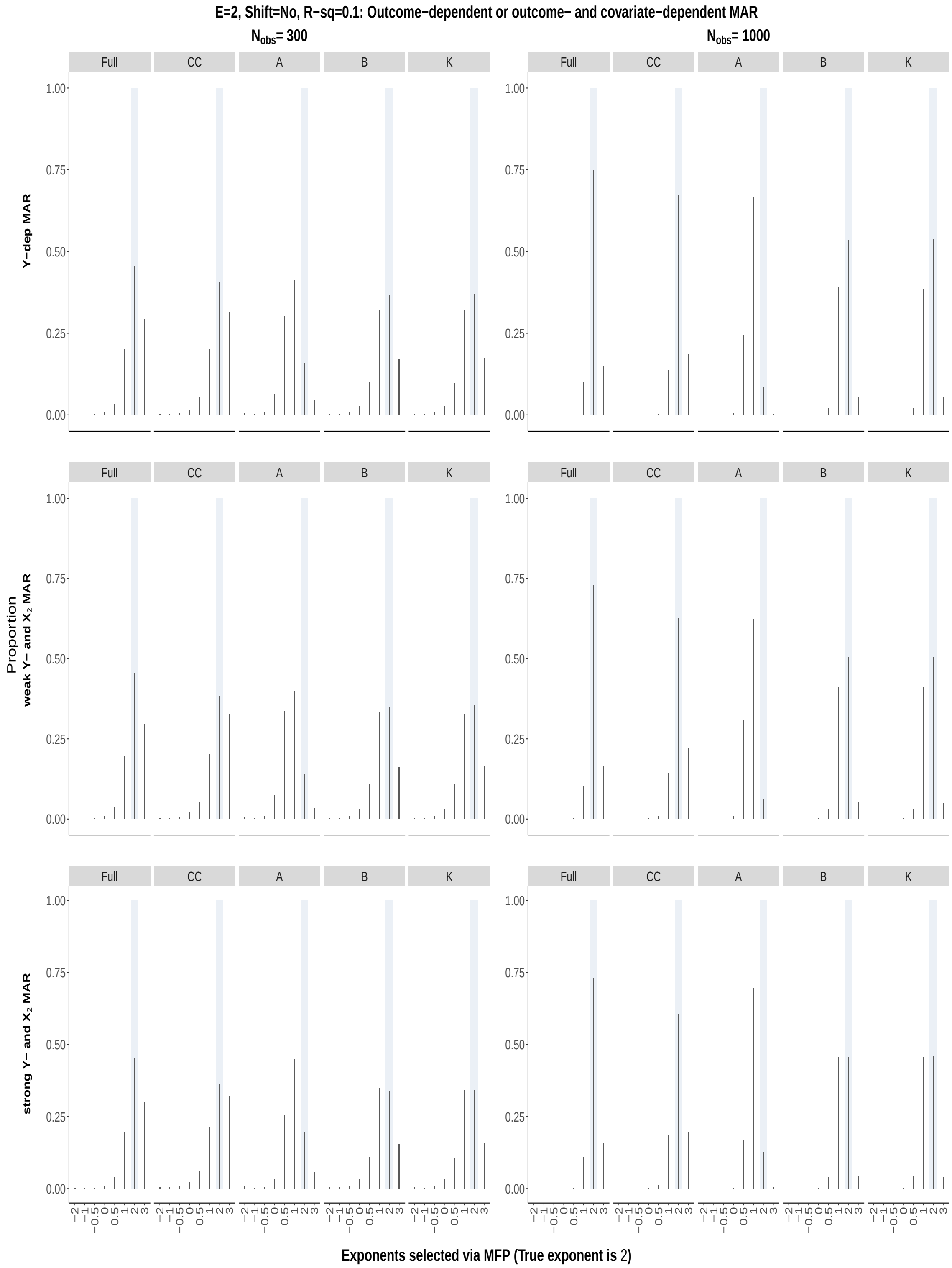

Figure S103: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

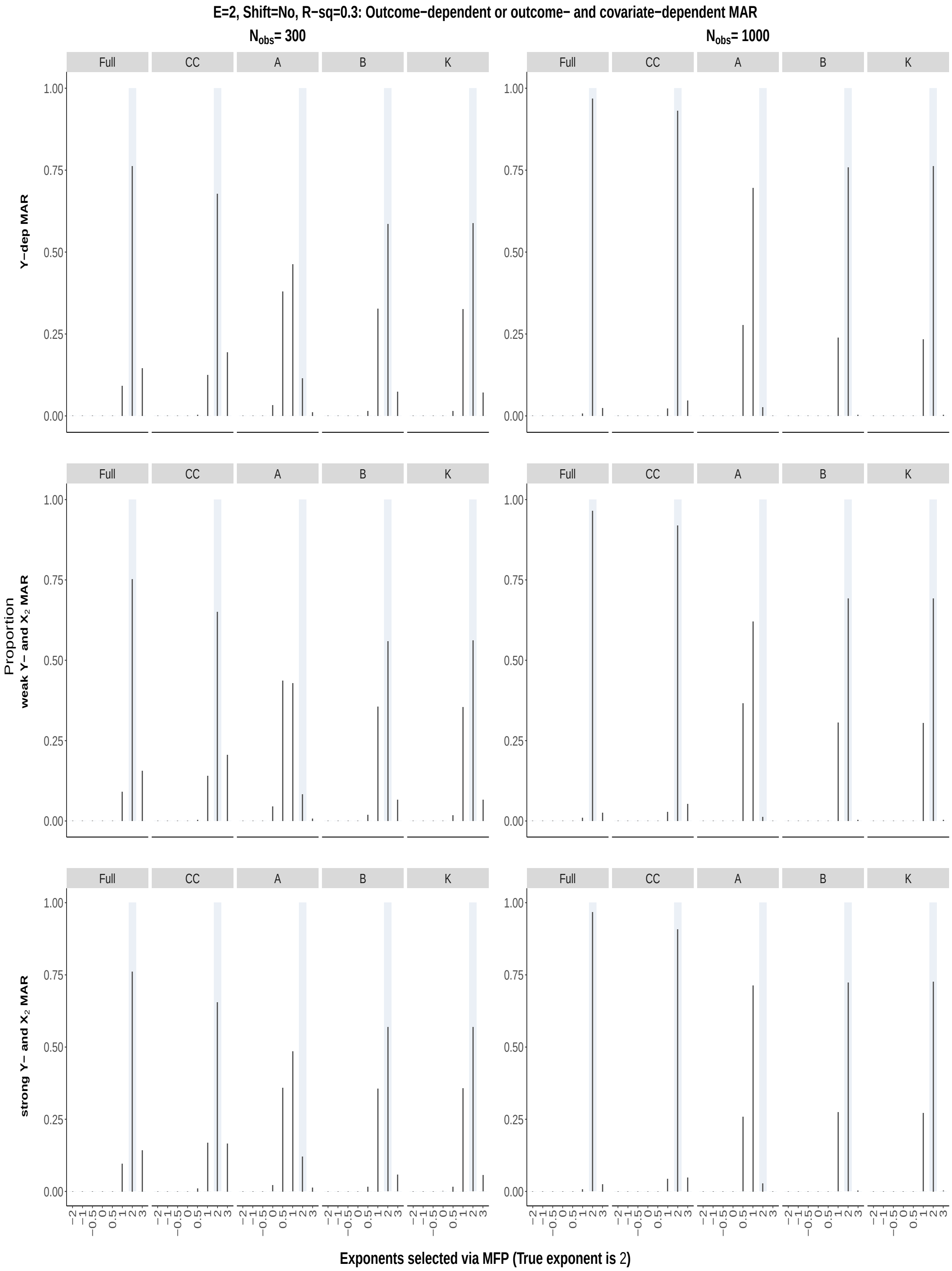

Figure S104: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

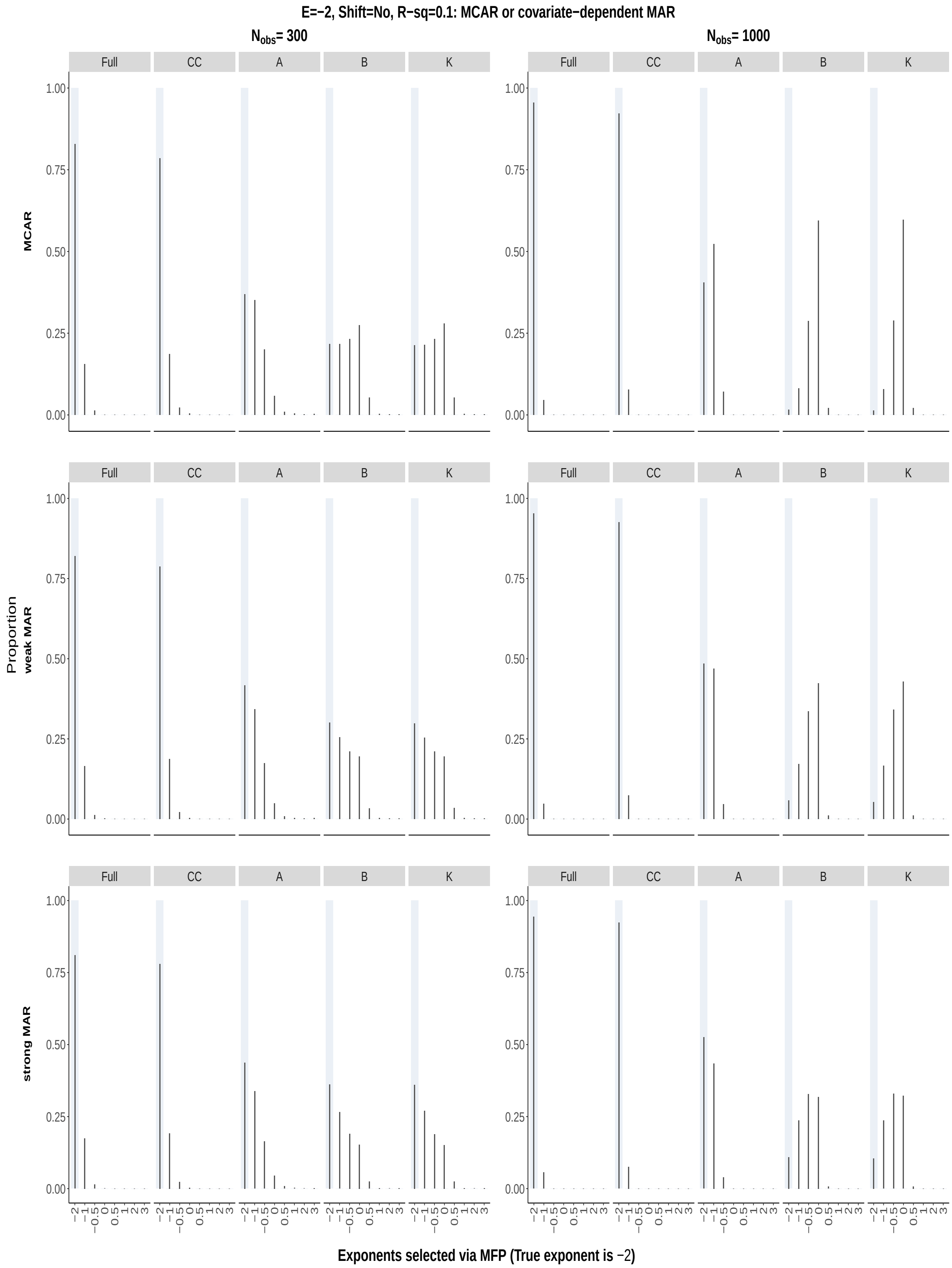

Figure S105: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

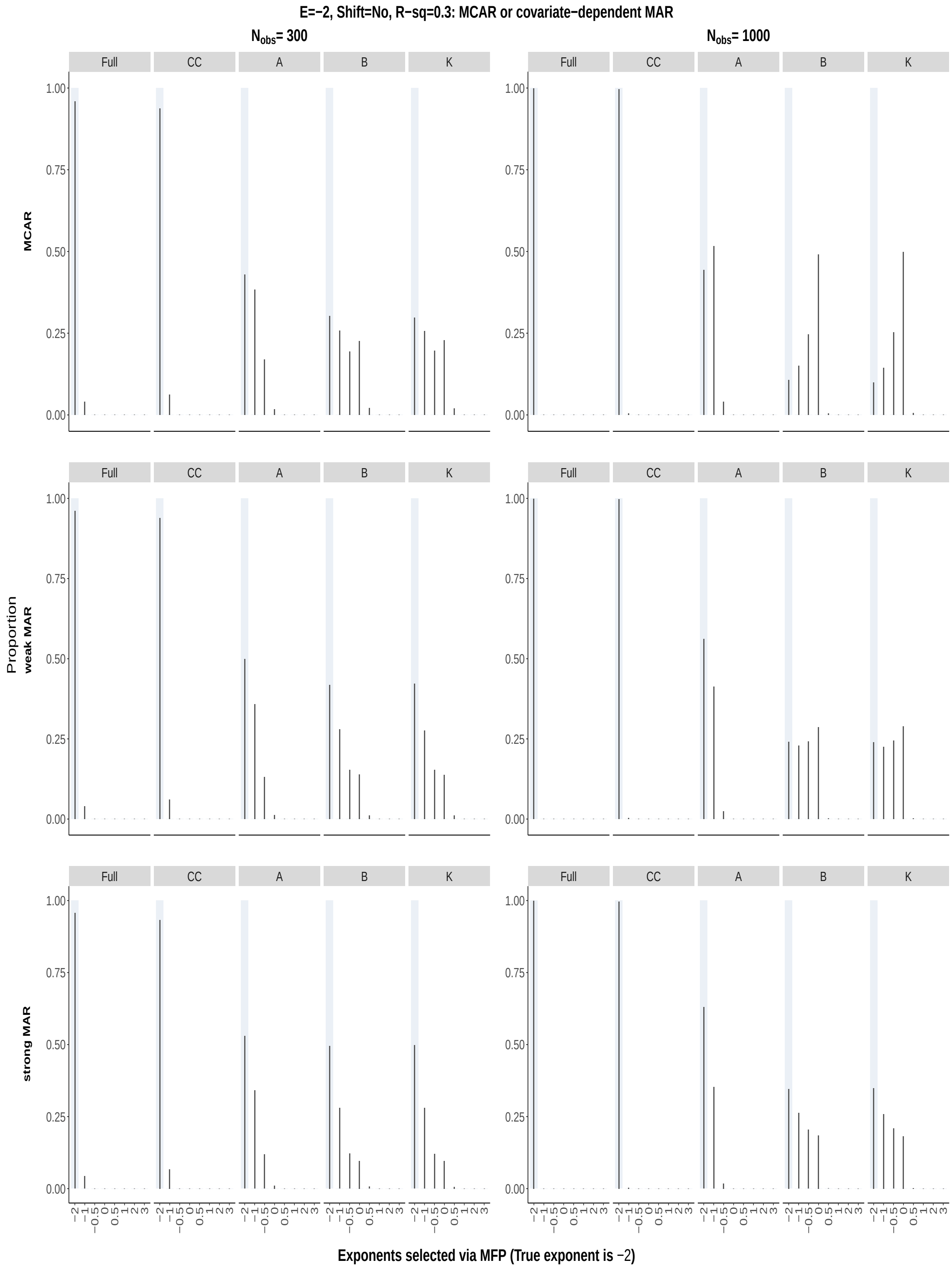

Figure S106: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

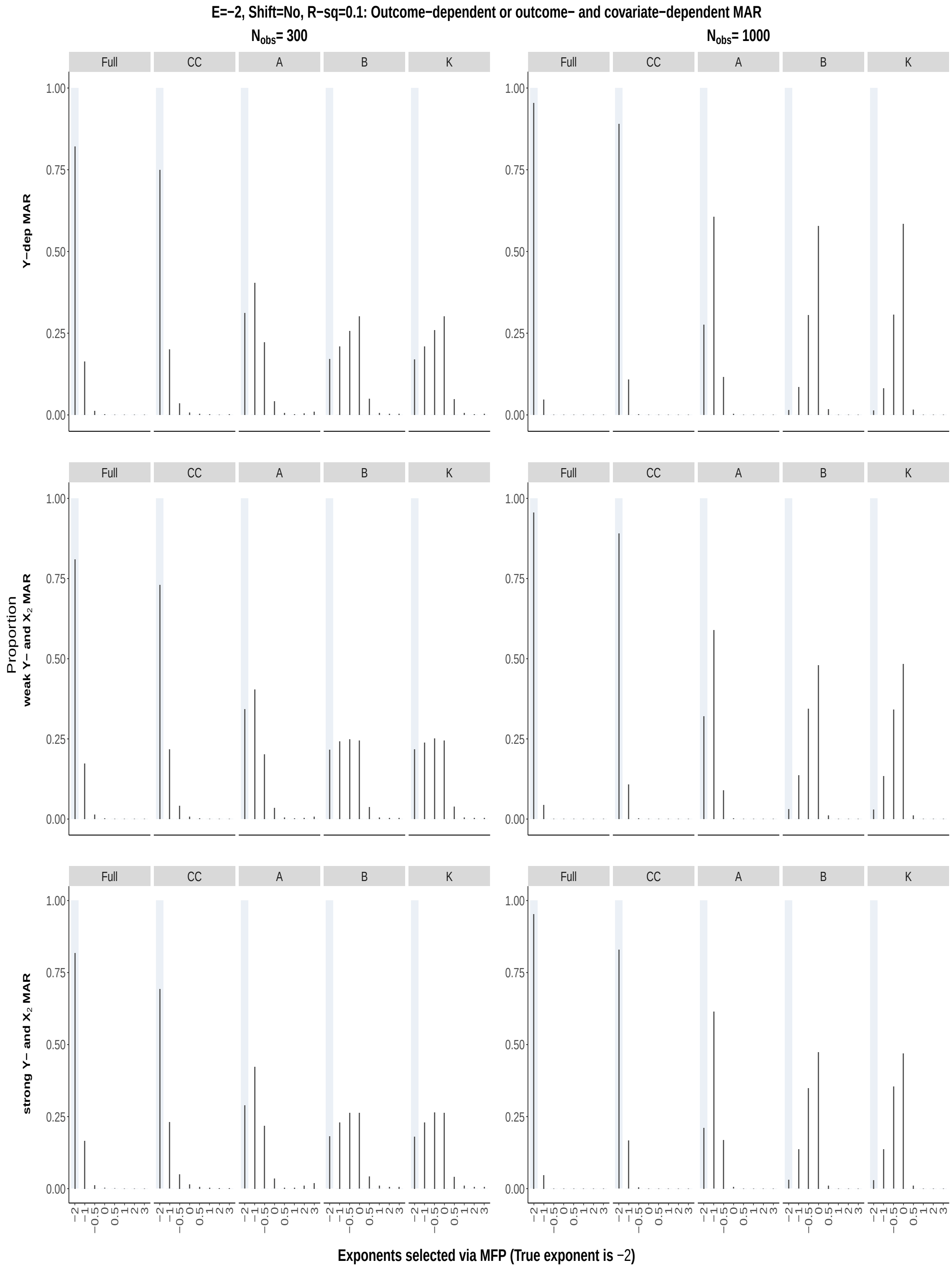

Figure S107: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

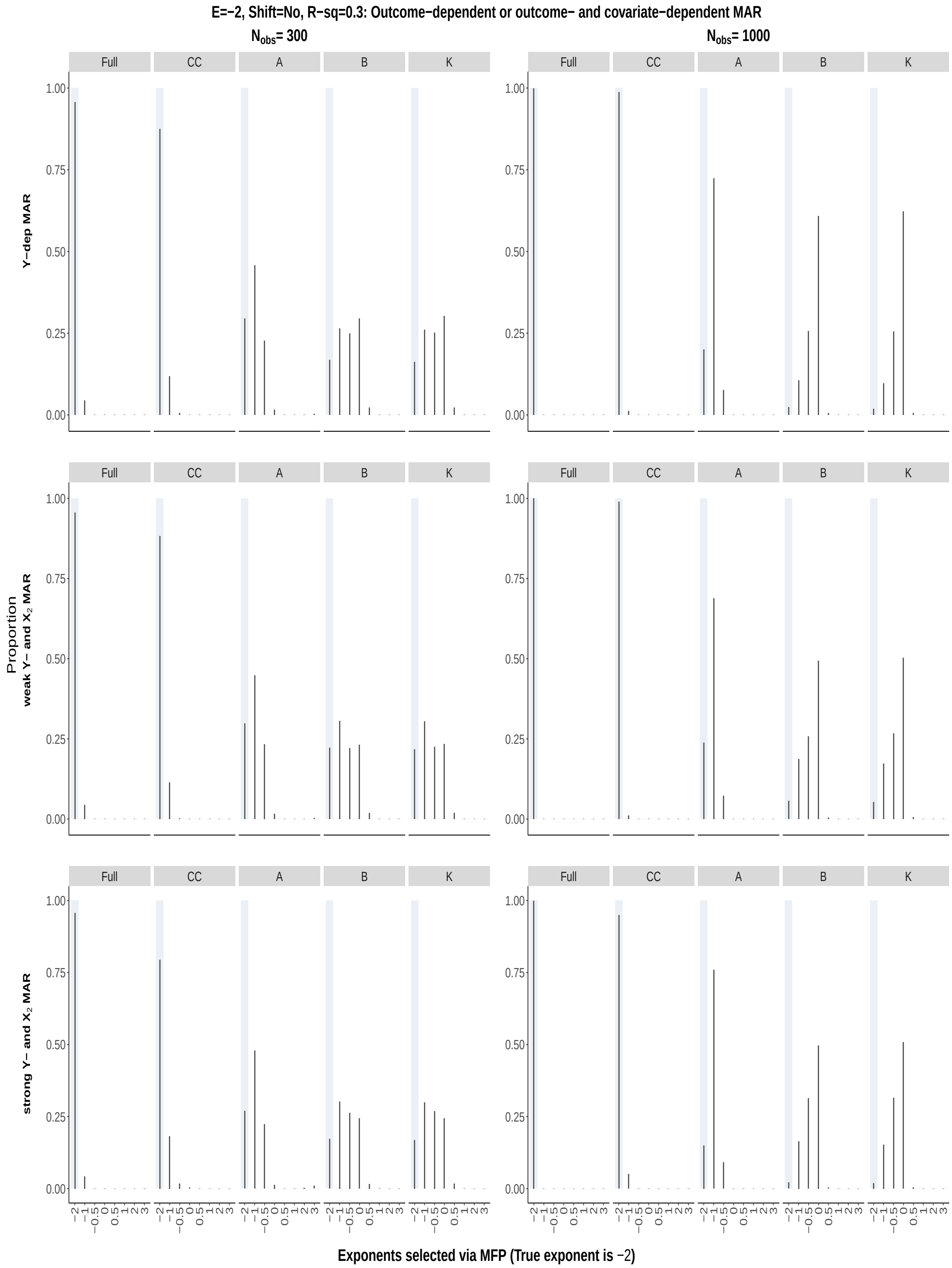

Figure S108: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.6 Cross-validation, $\beta_2 = 0$, $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

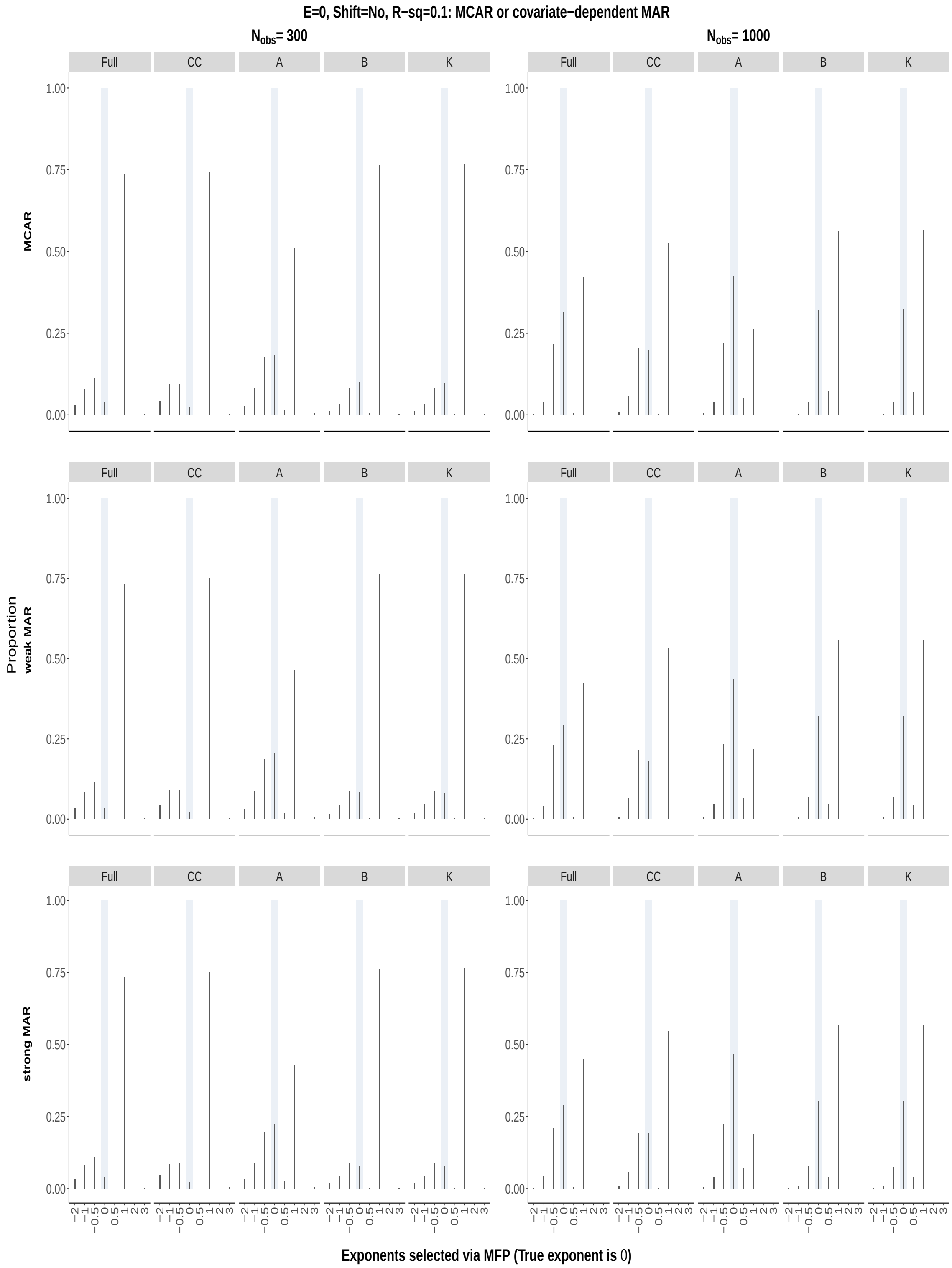

Figure S109: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

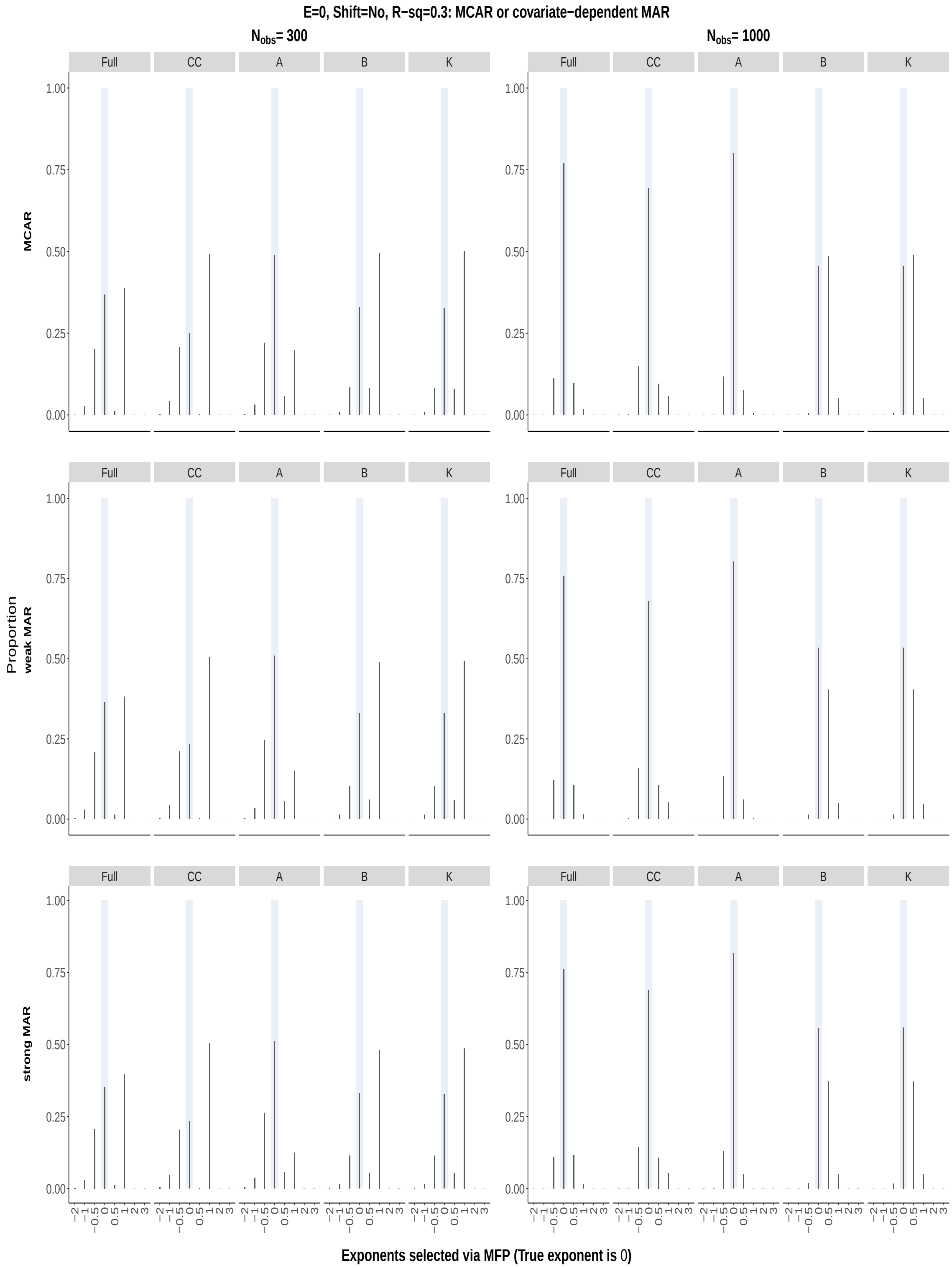

Figure S110: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

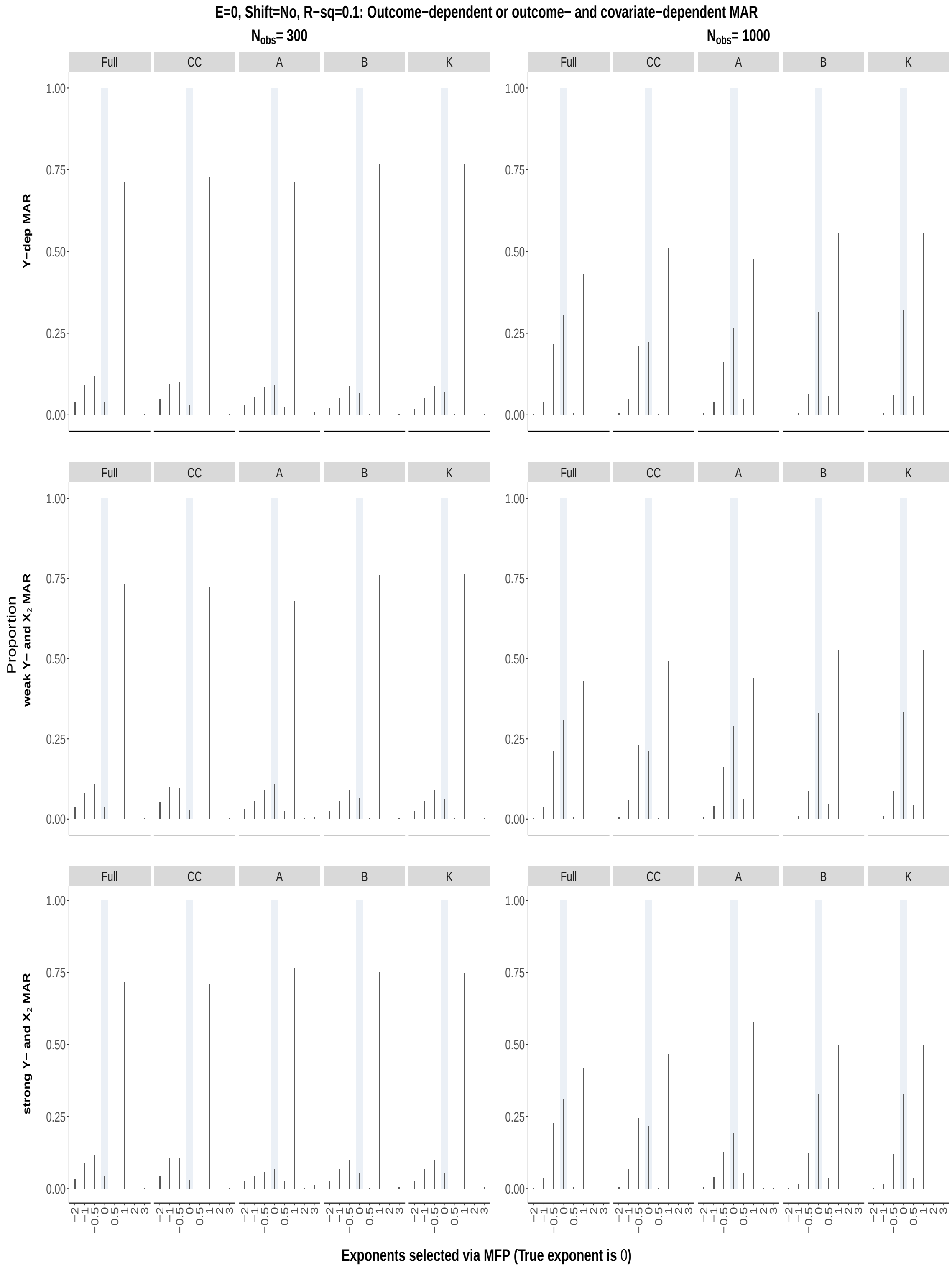

Figure S111: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

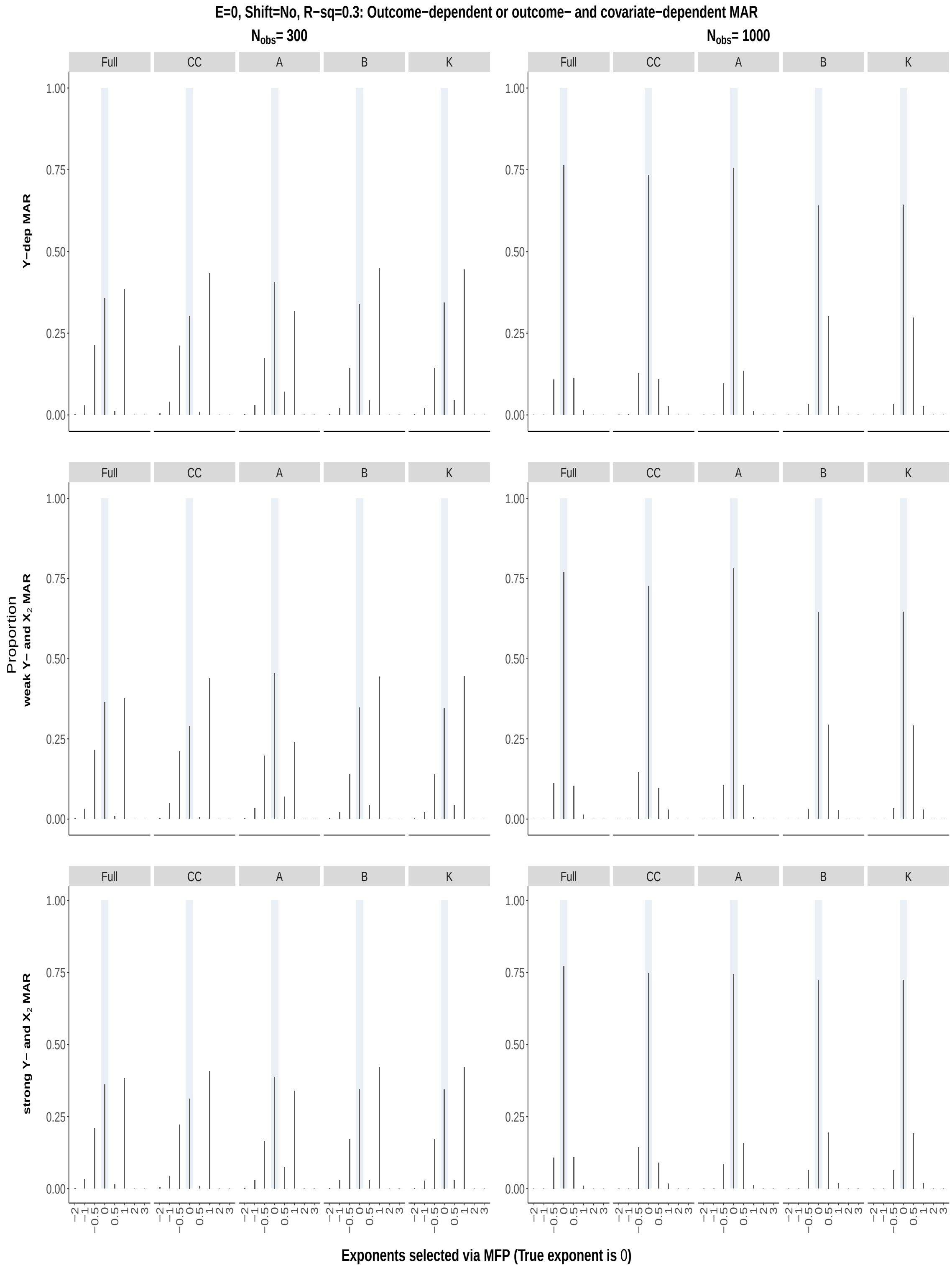

Figure S112: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

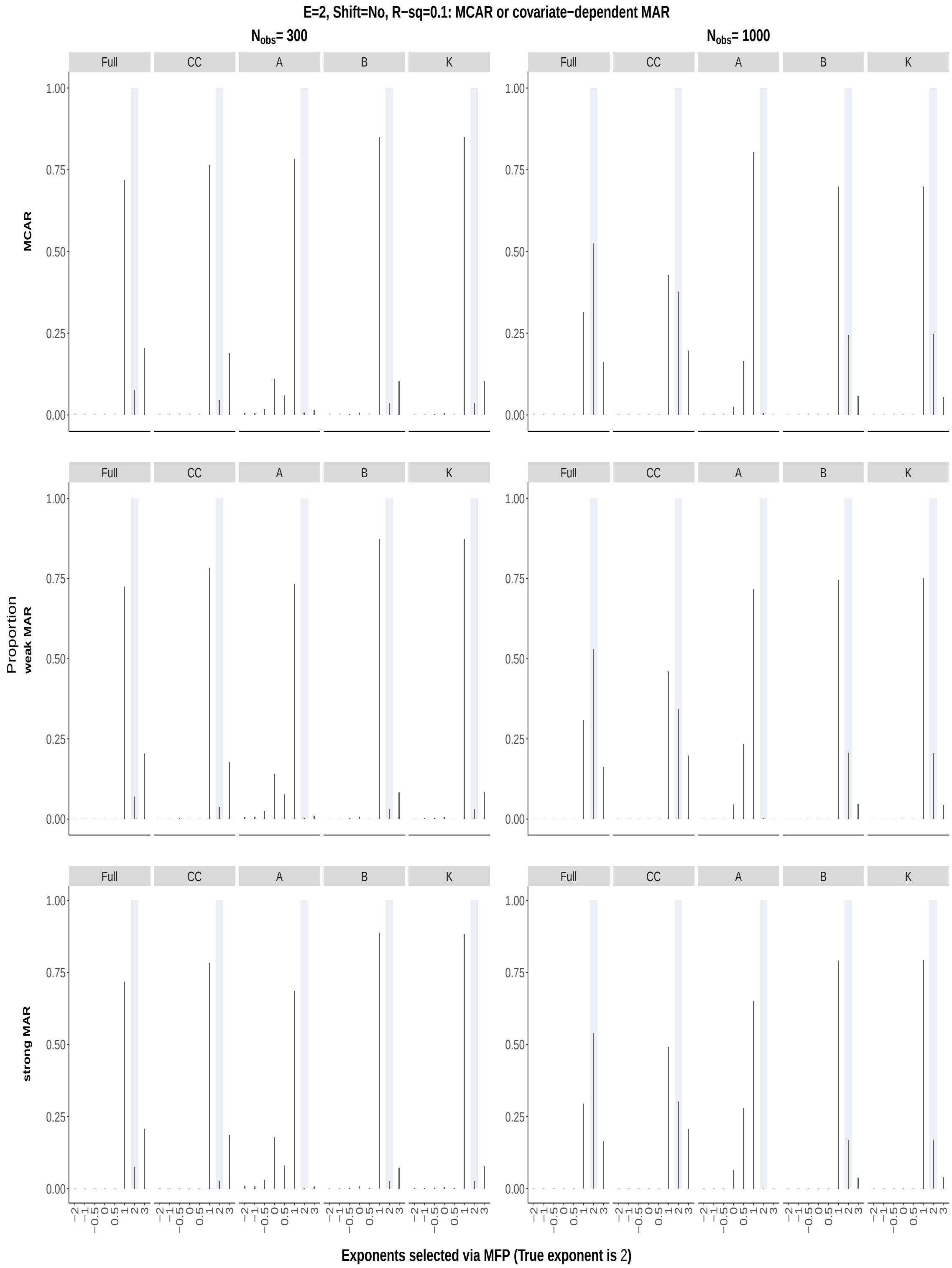

Figure S113: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

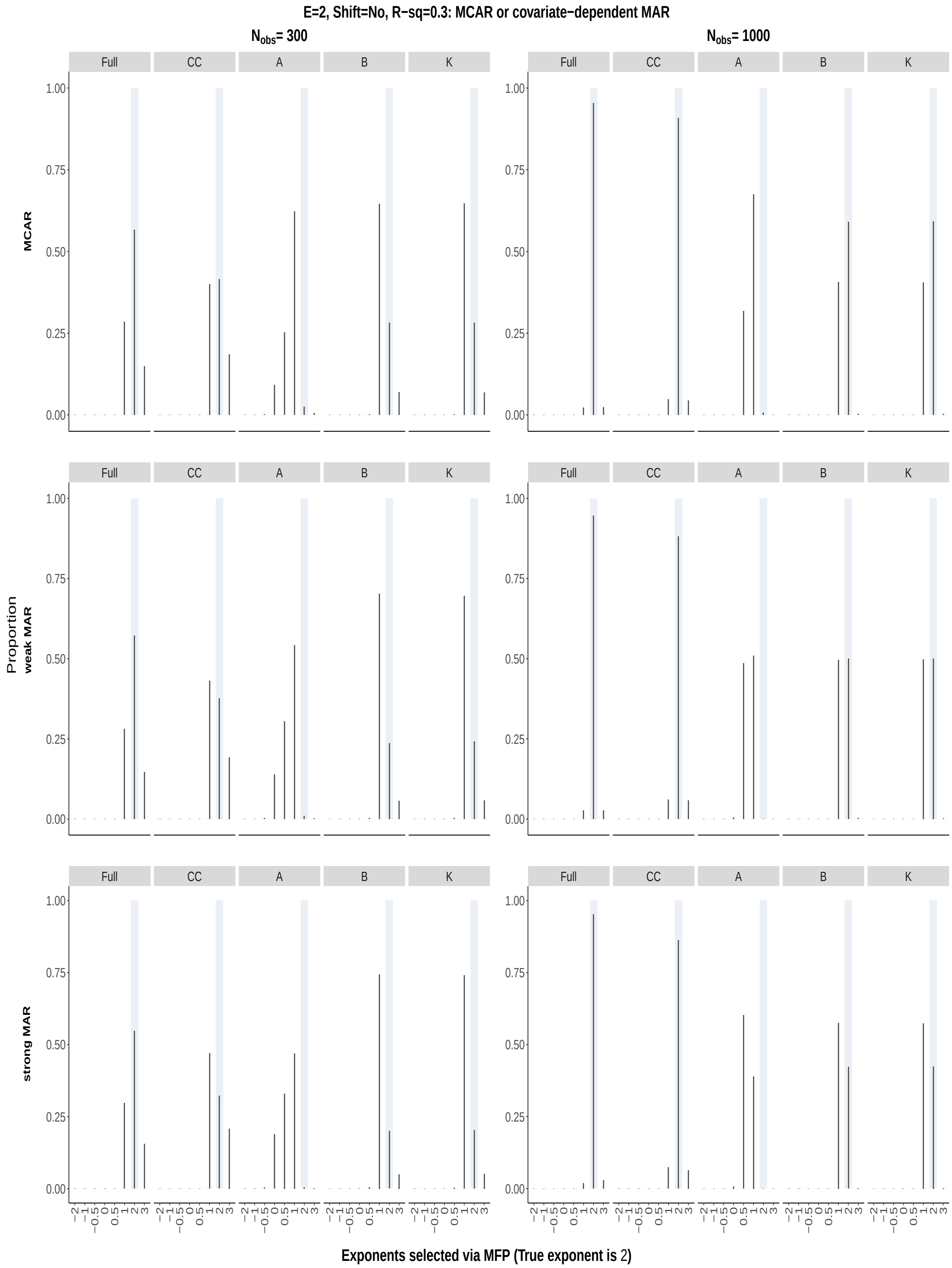

Figure S114: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

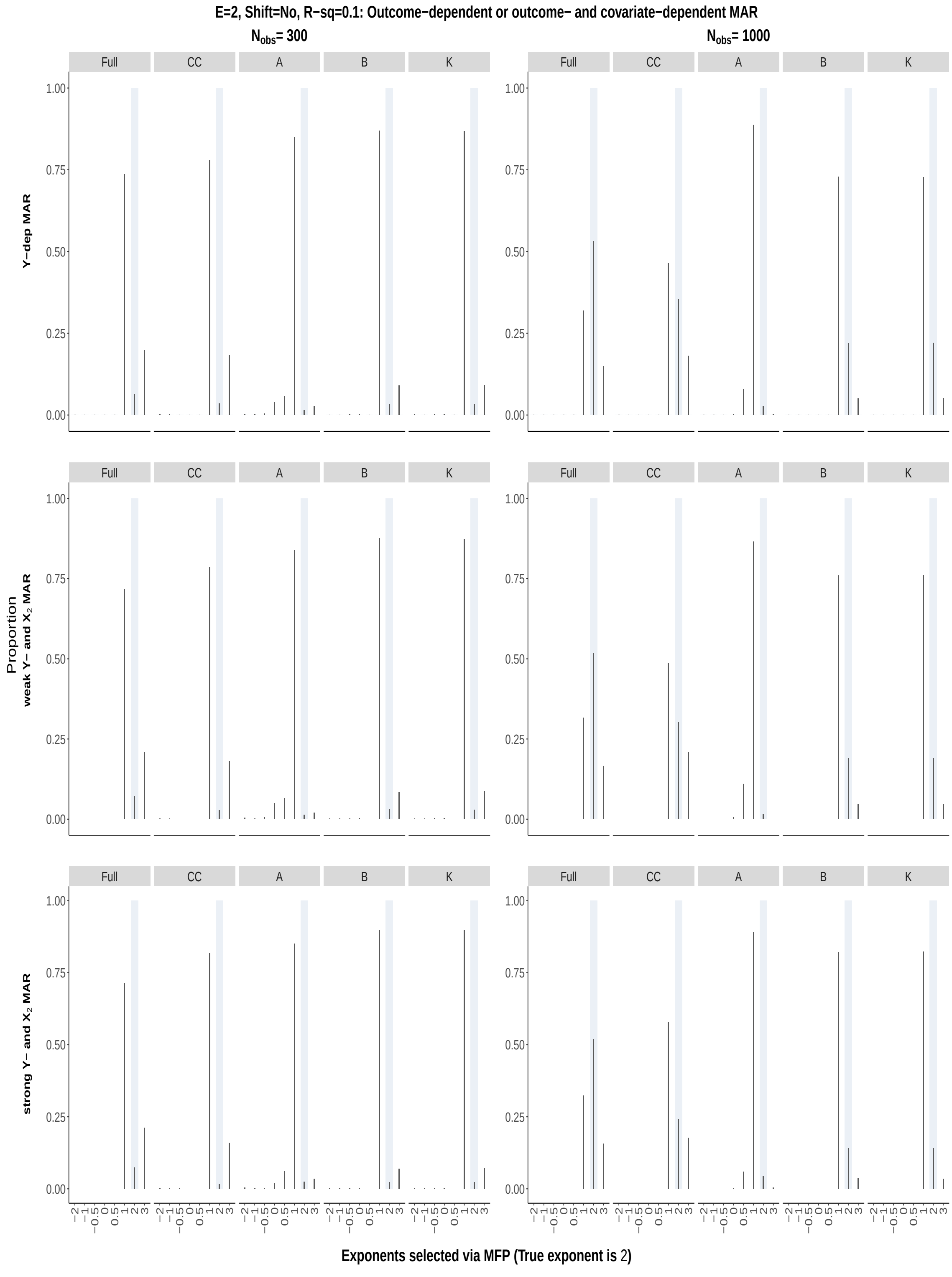

Figure S115: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

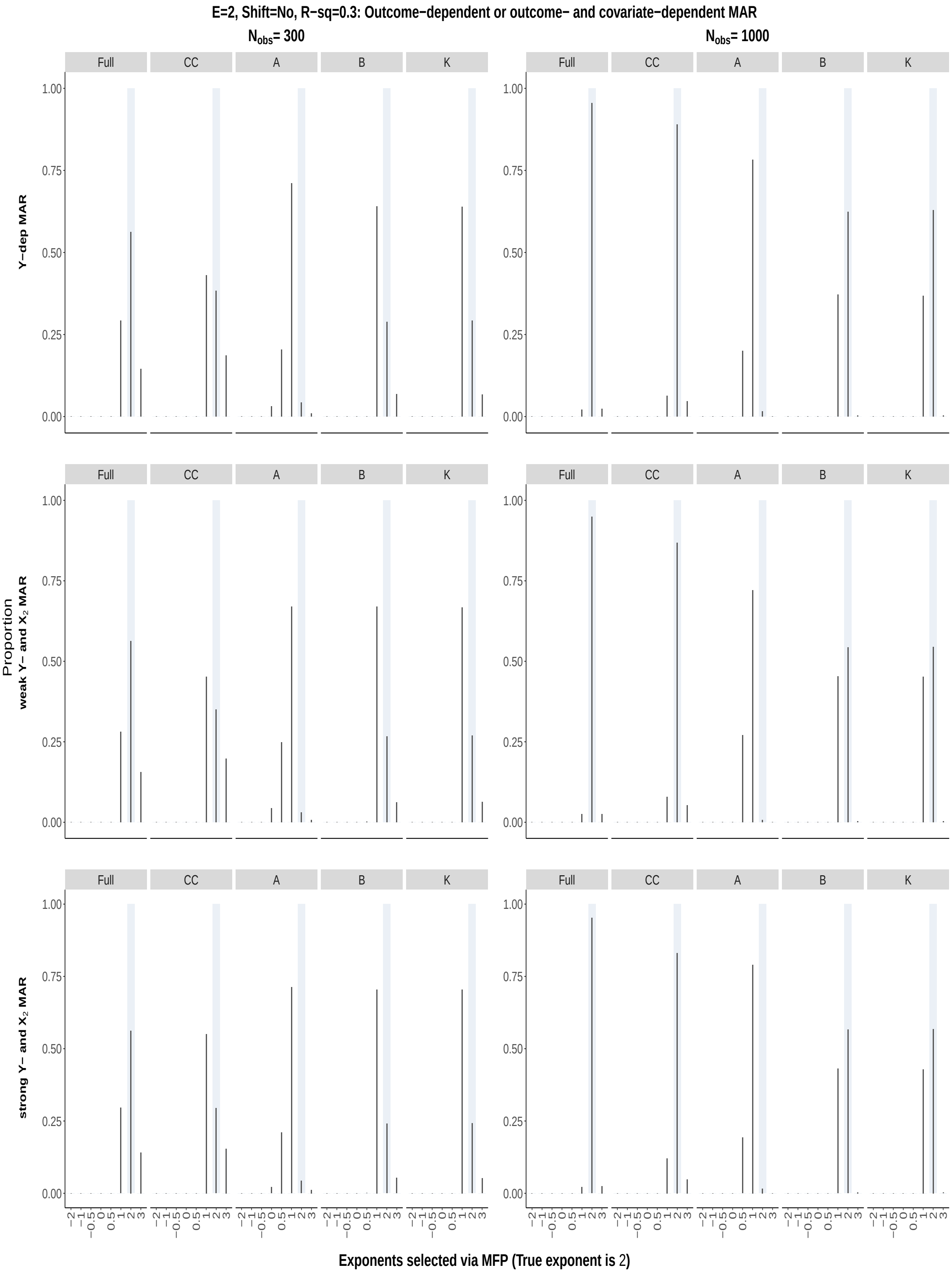

Figure S116: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

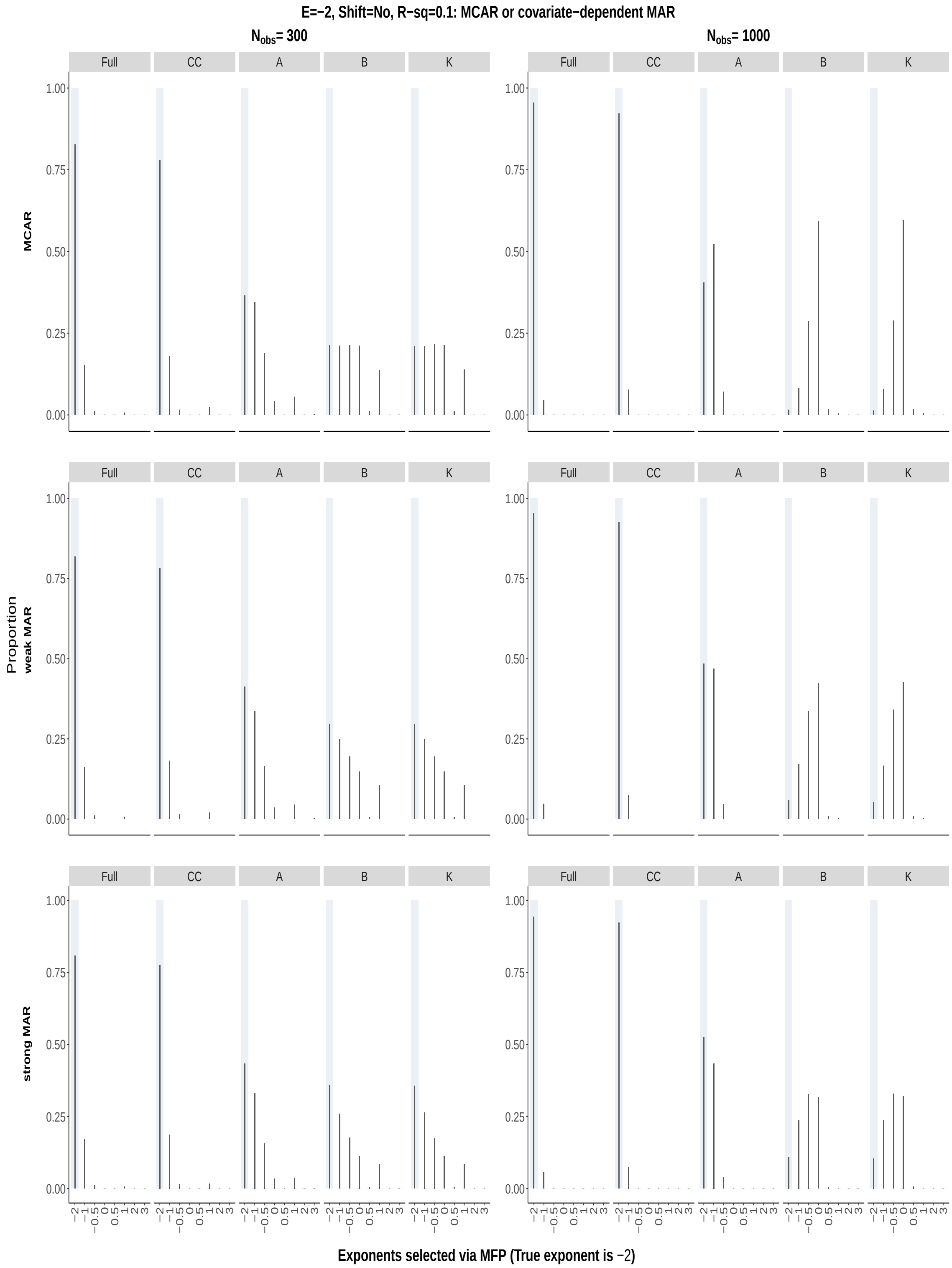

Figure S117: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

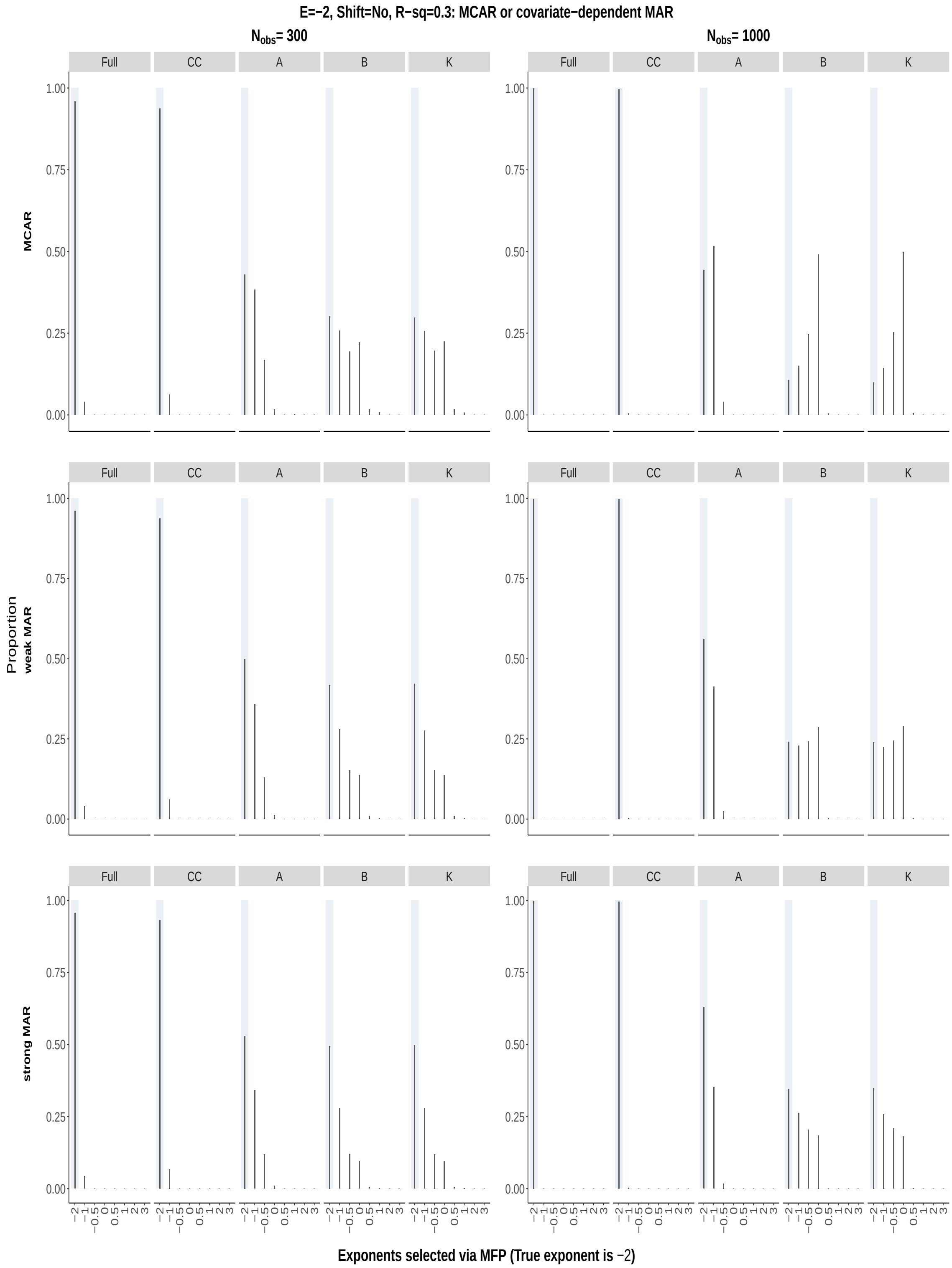

Figure S118: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

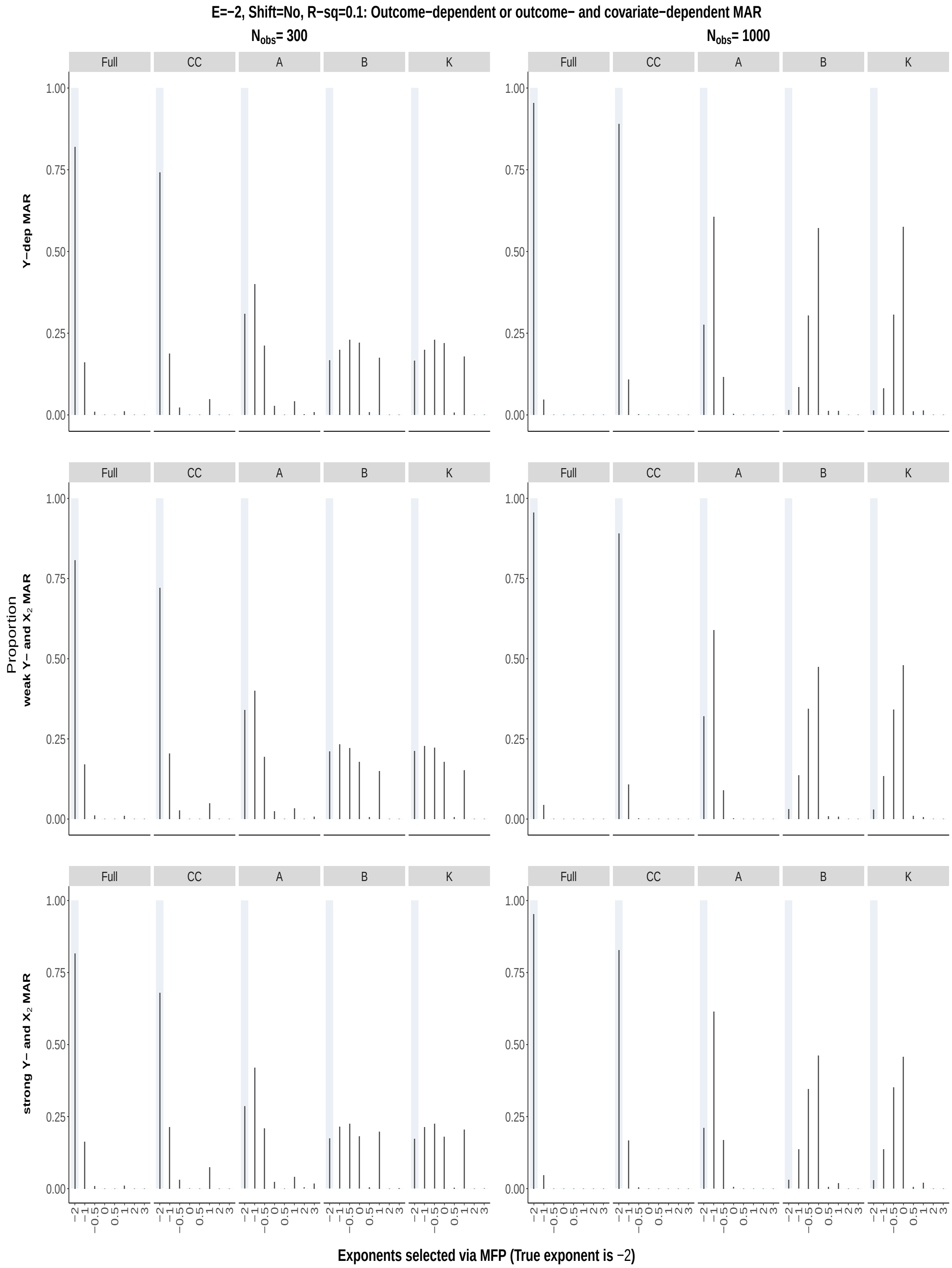

Figure S119: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

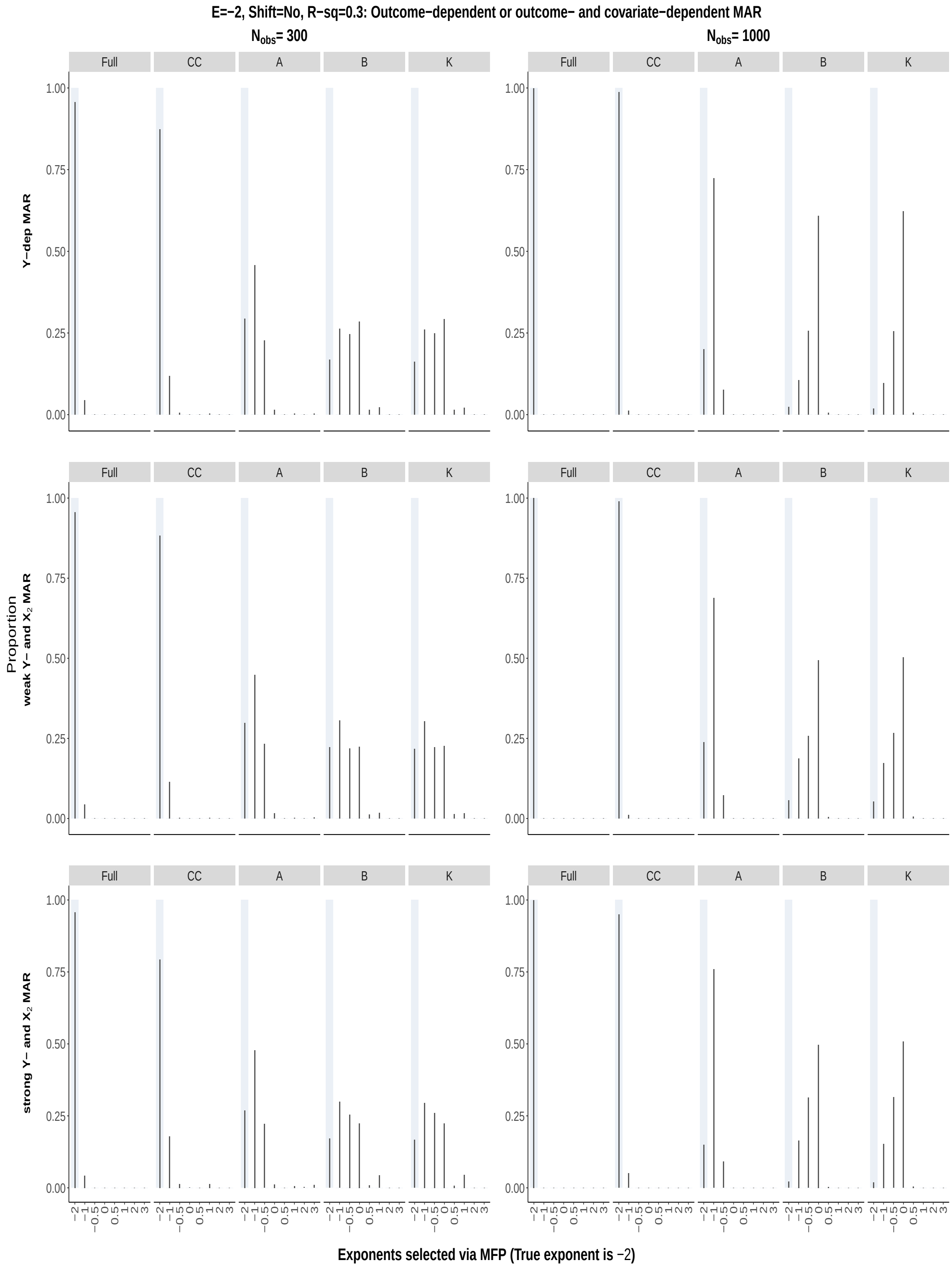

Figure S120: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.7 Cross-validation, $\beta_2 = 0$, $\alpha_E = 1$ and an origin-shift has been used

True exponent is 0

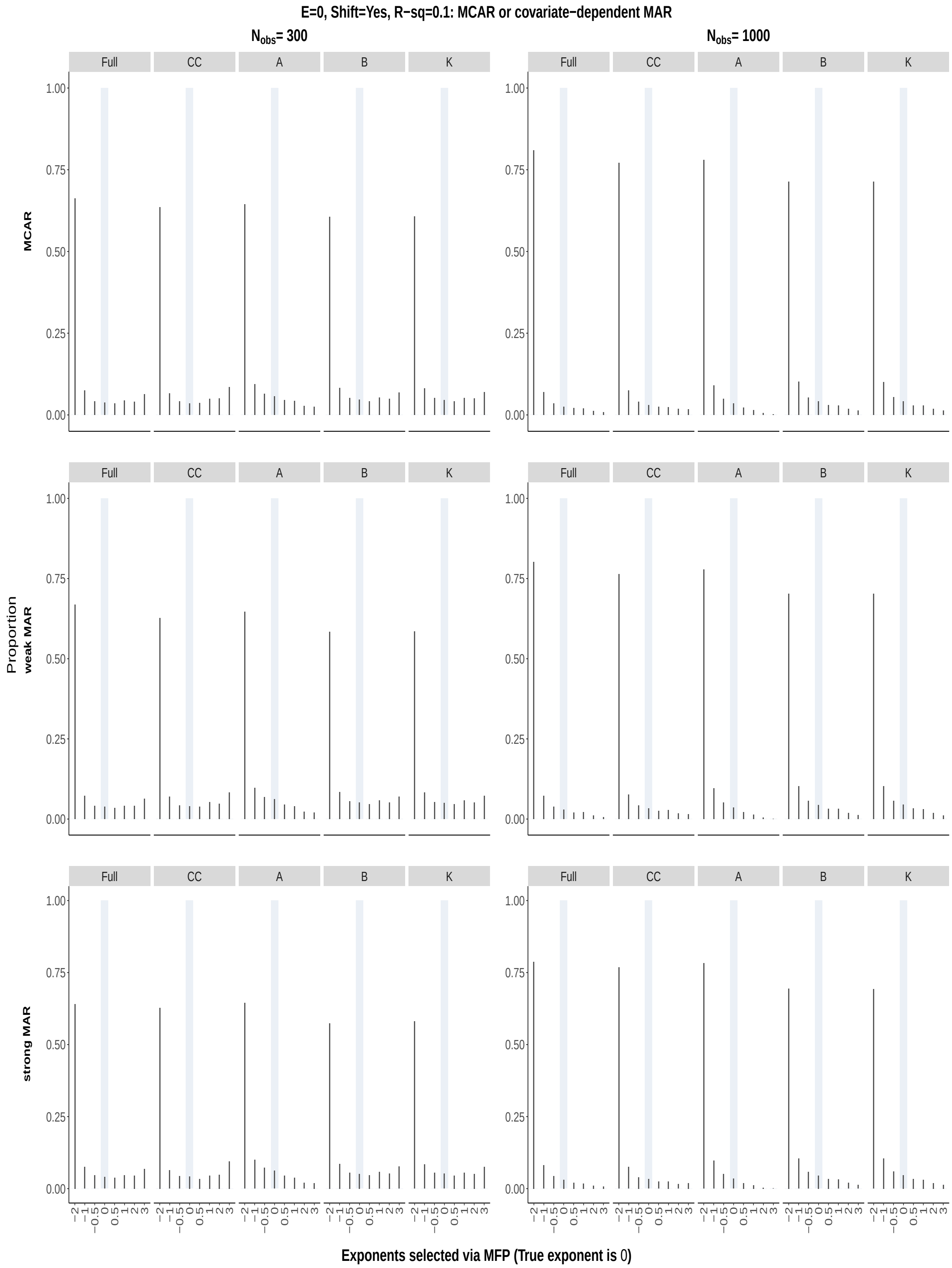

Figure S121: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

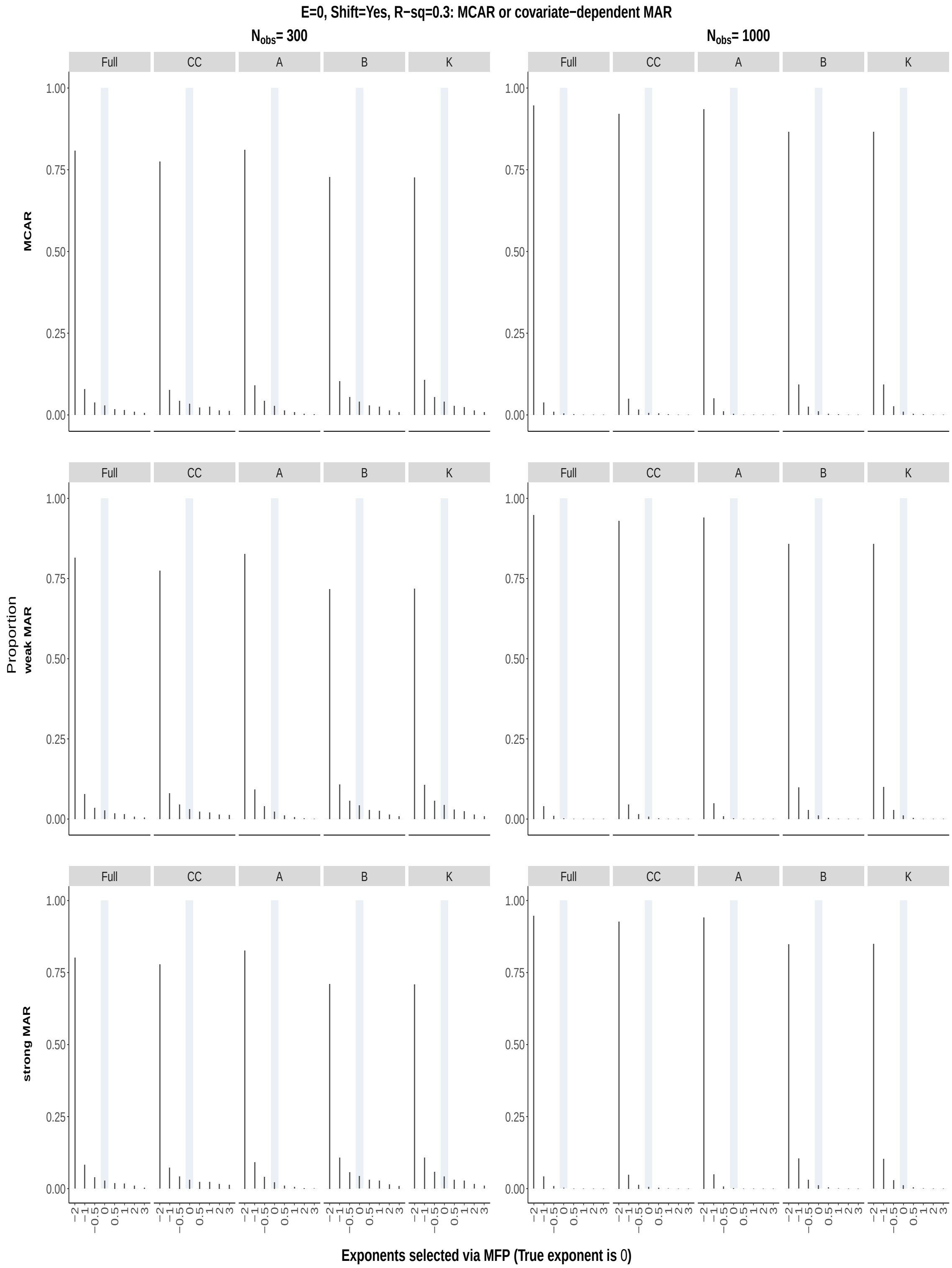

Figure S122: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

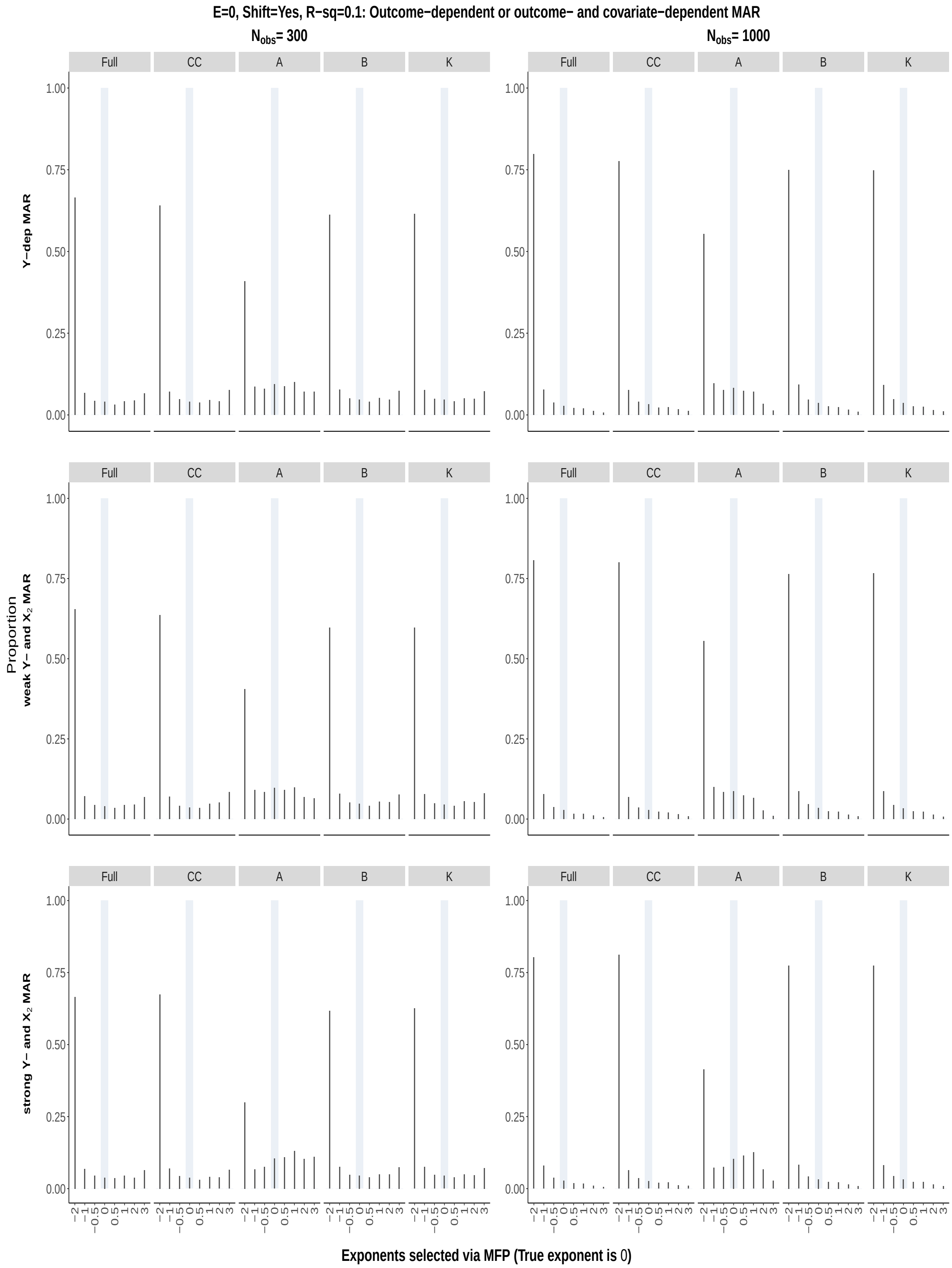

Figure S123: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

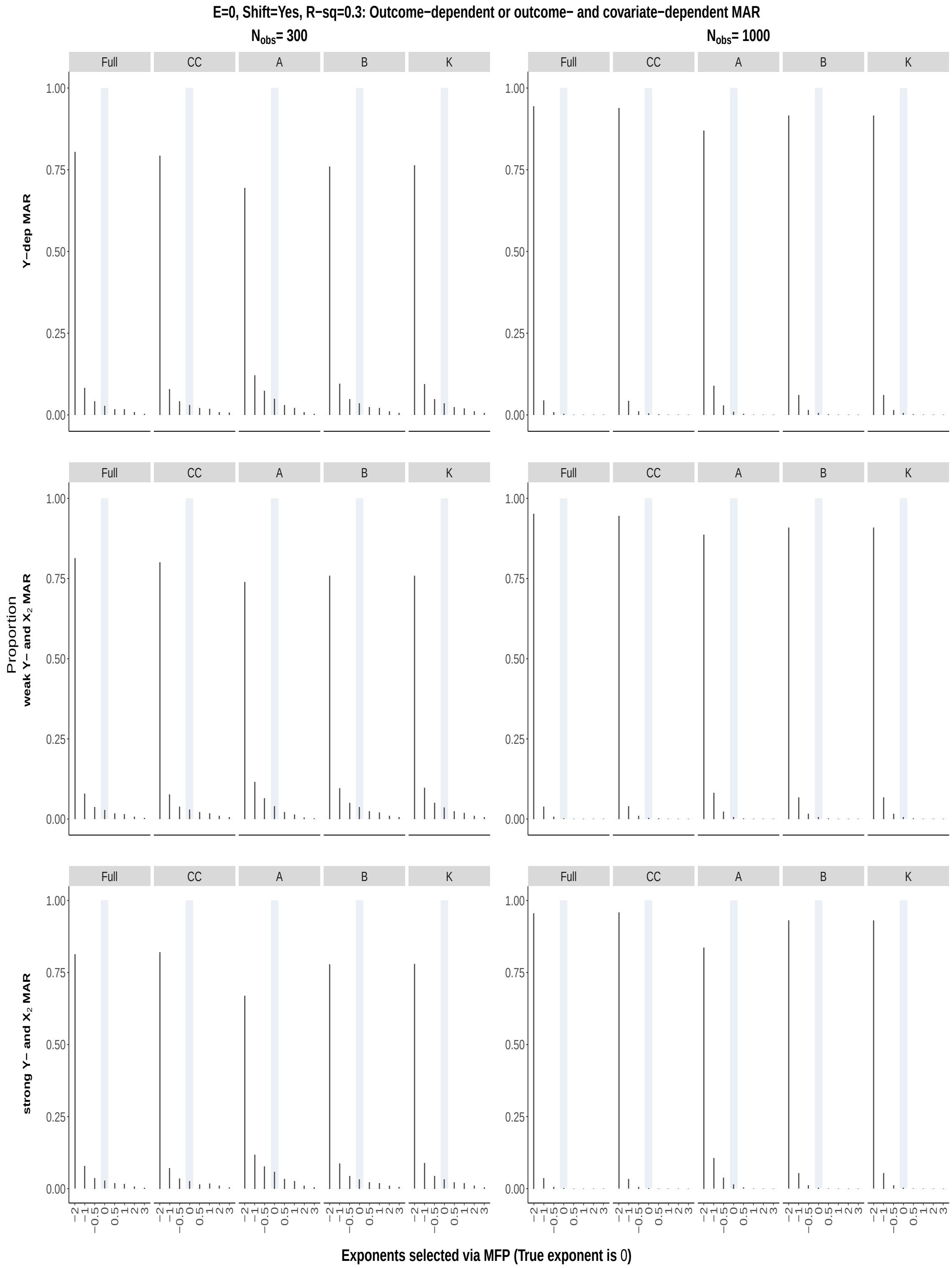

Figure S124: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

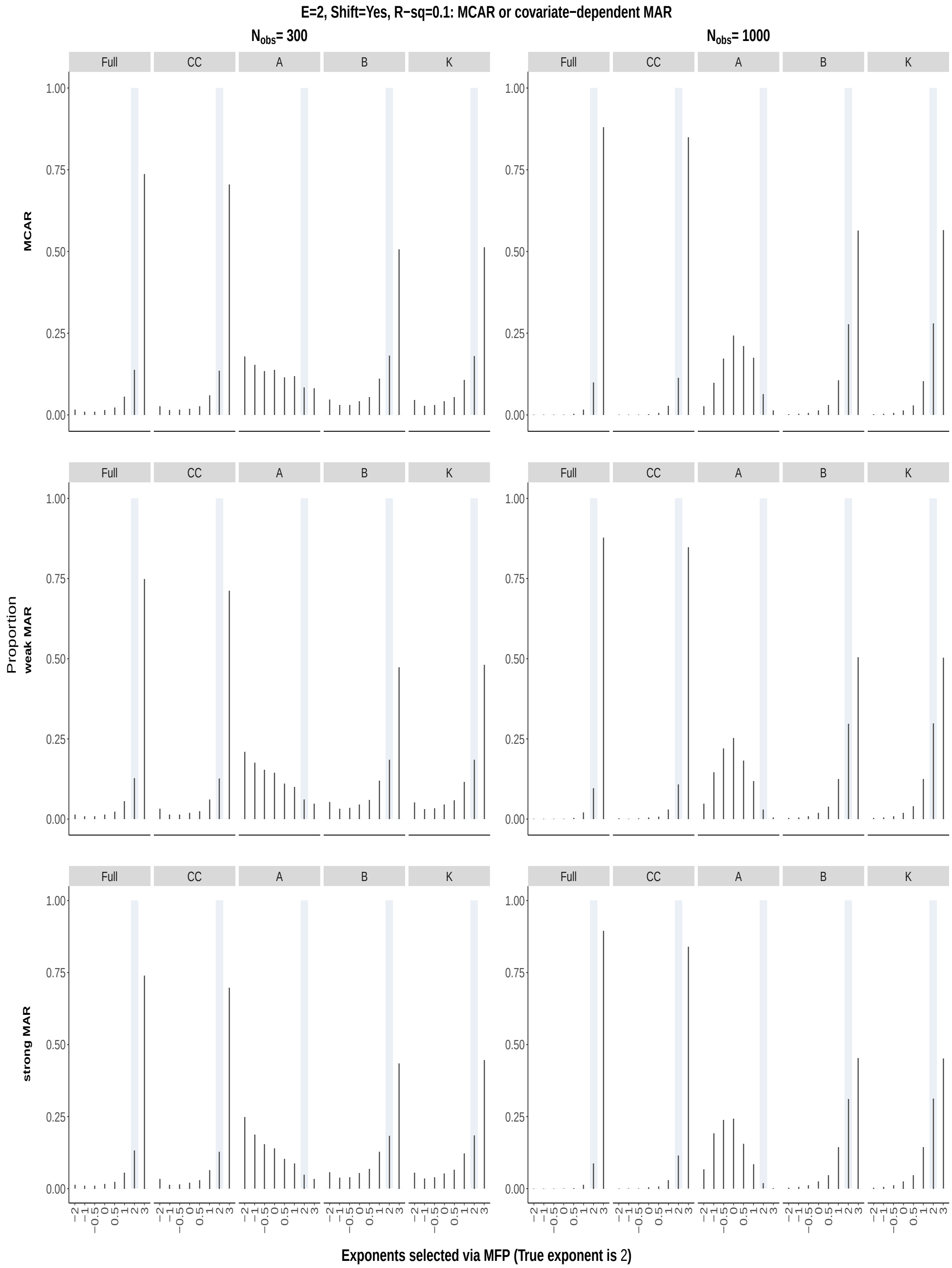

Figure S125: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

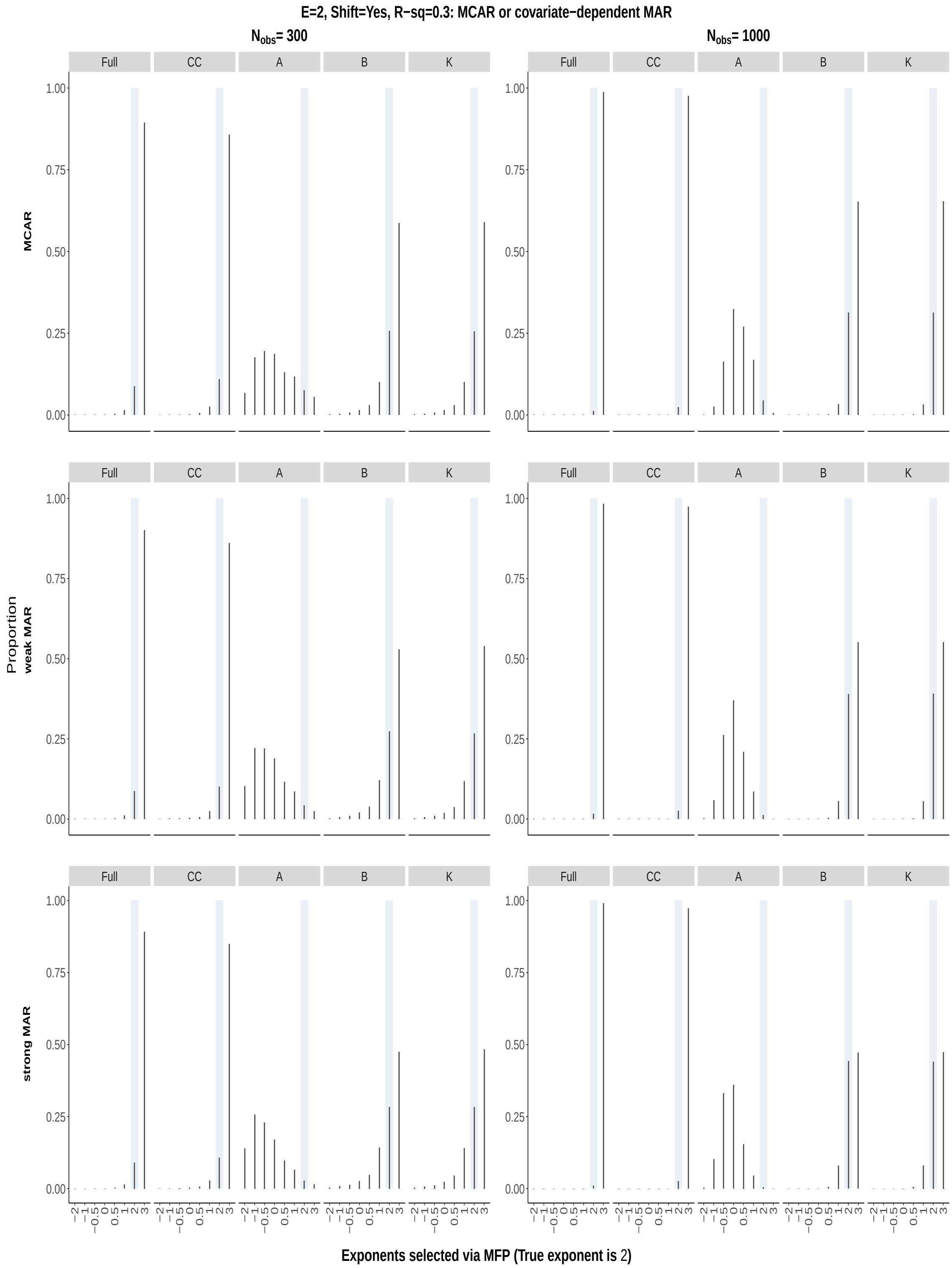

Figure S126: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

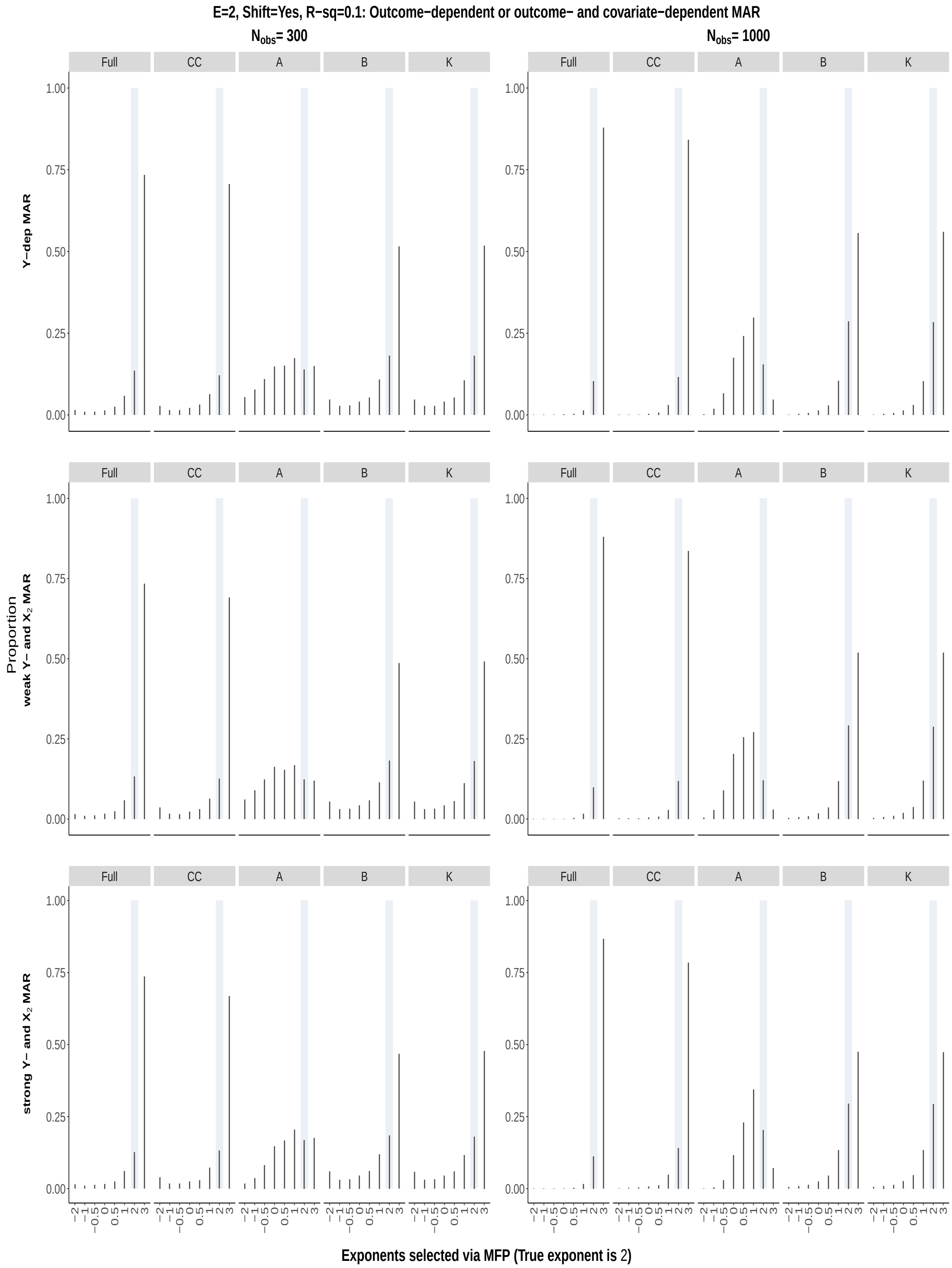

Figure S127: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

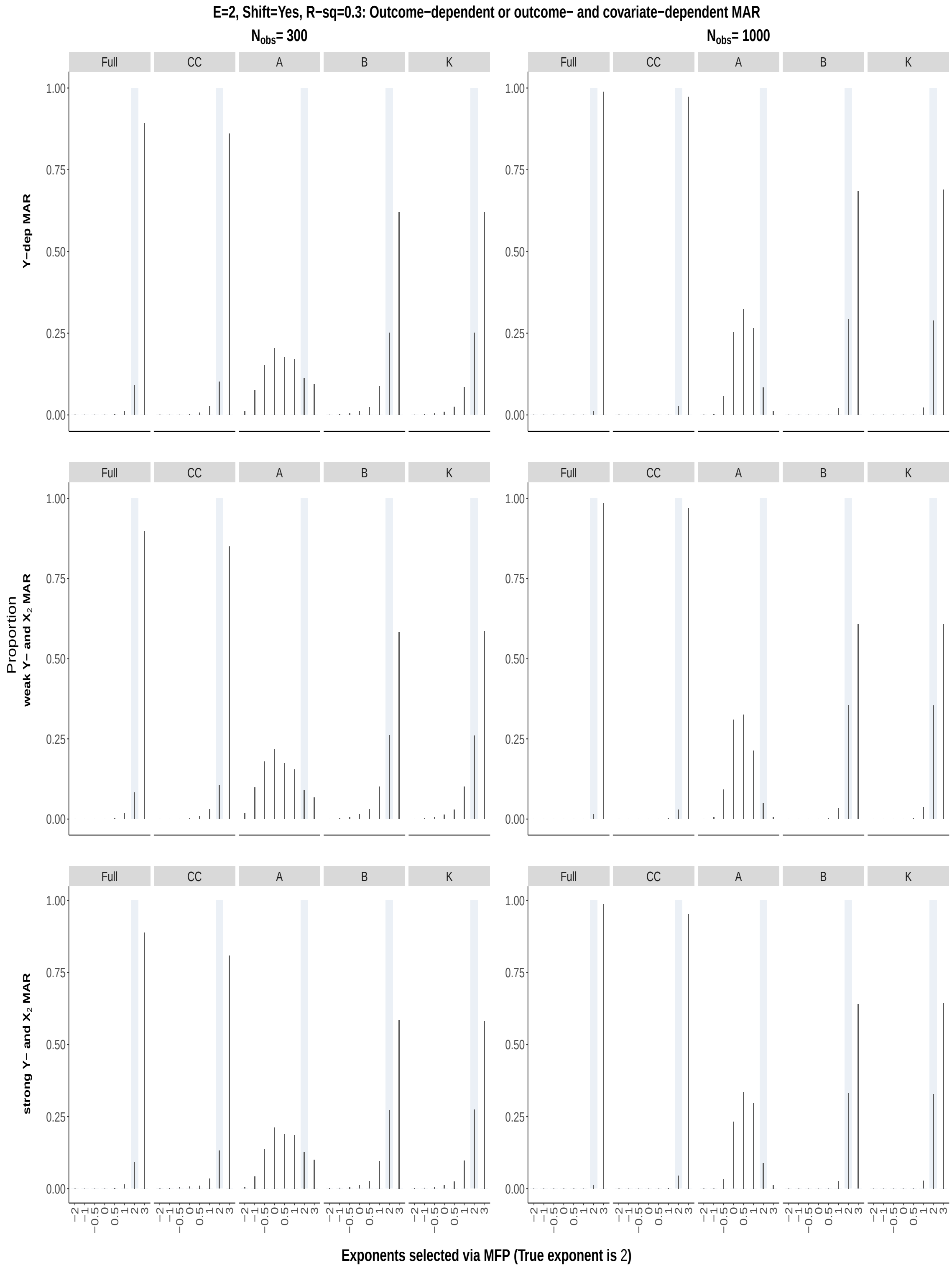

Figure S128: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

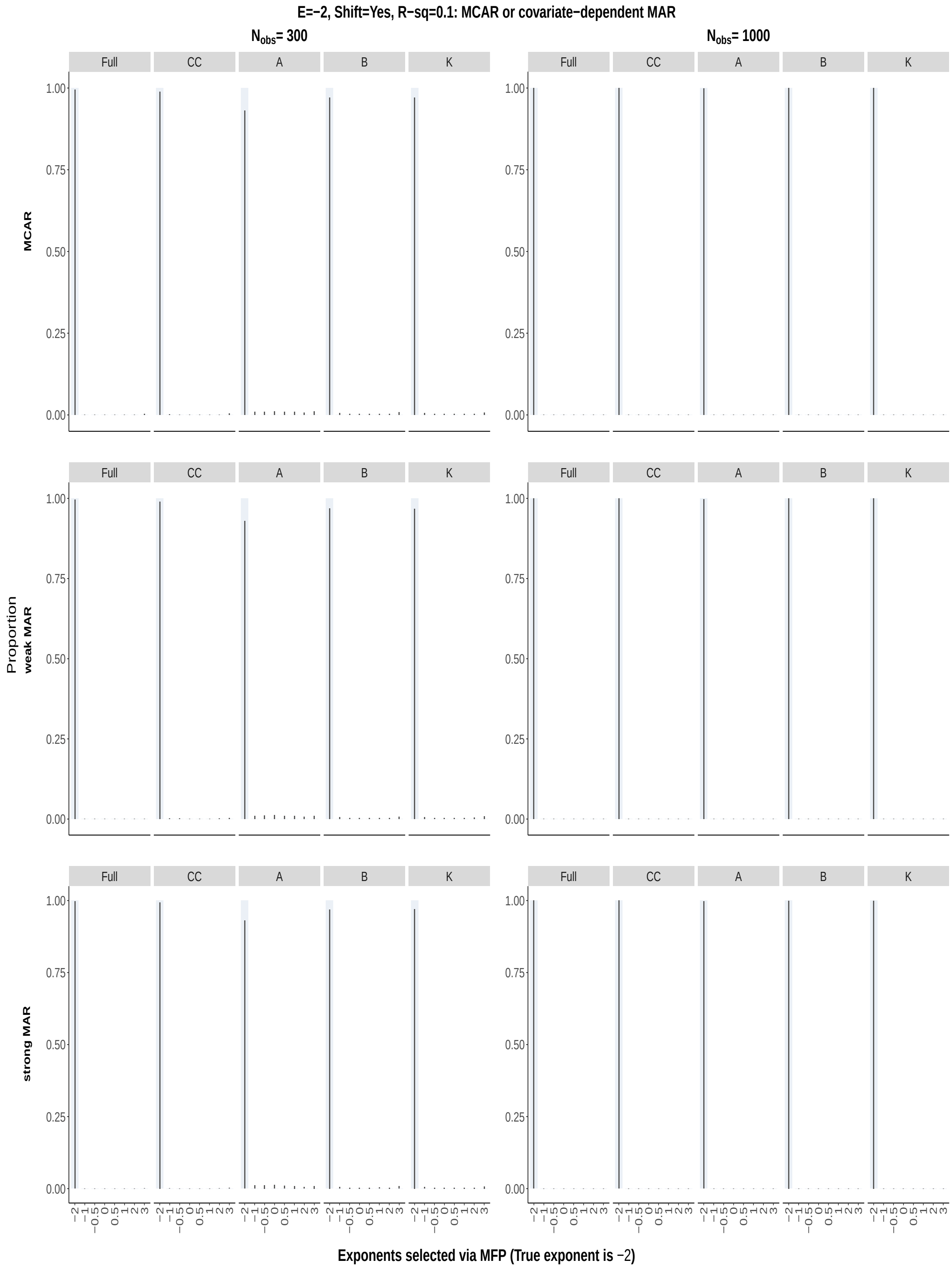

Figure S129: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

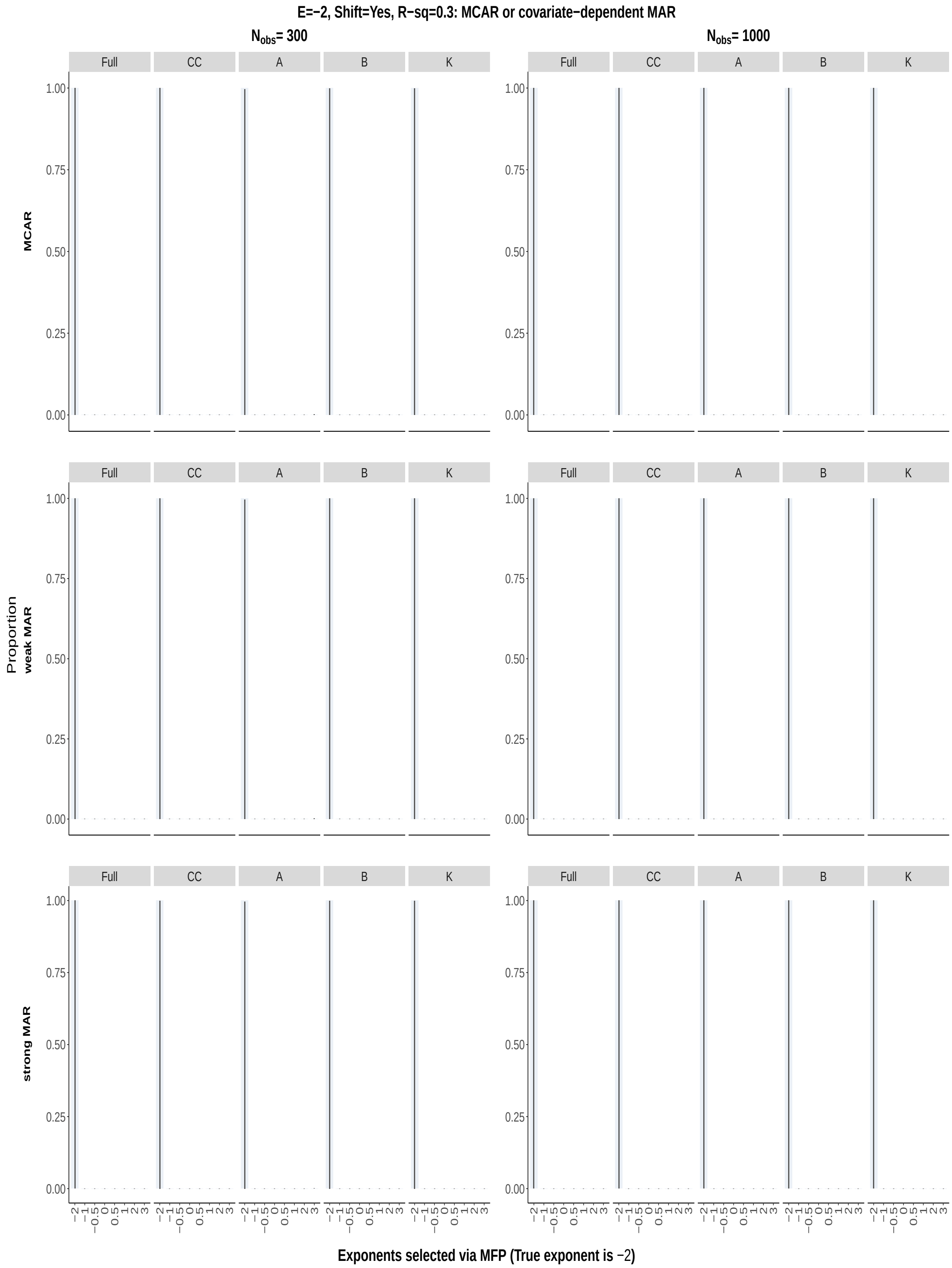

Figure S130: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

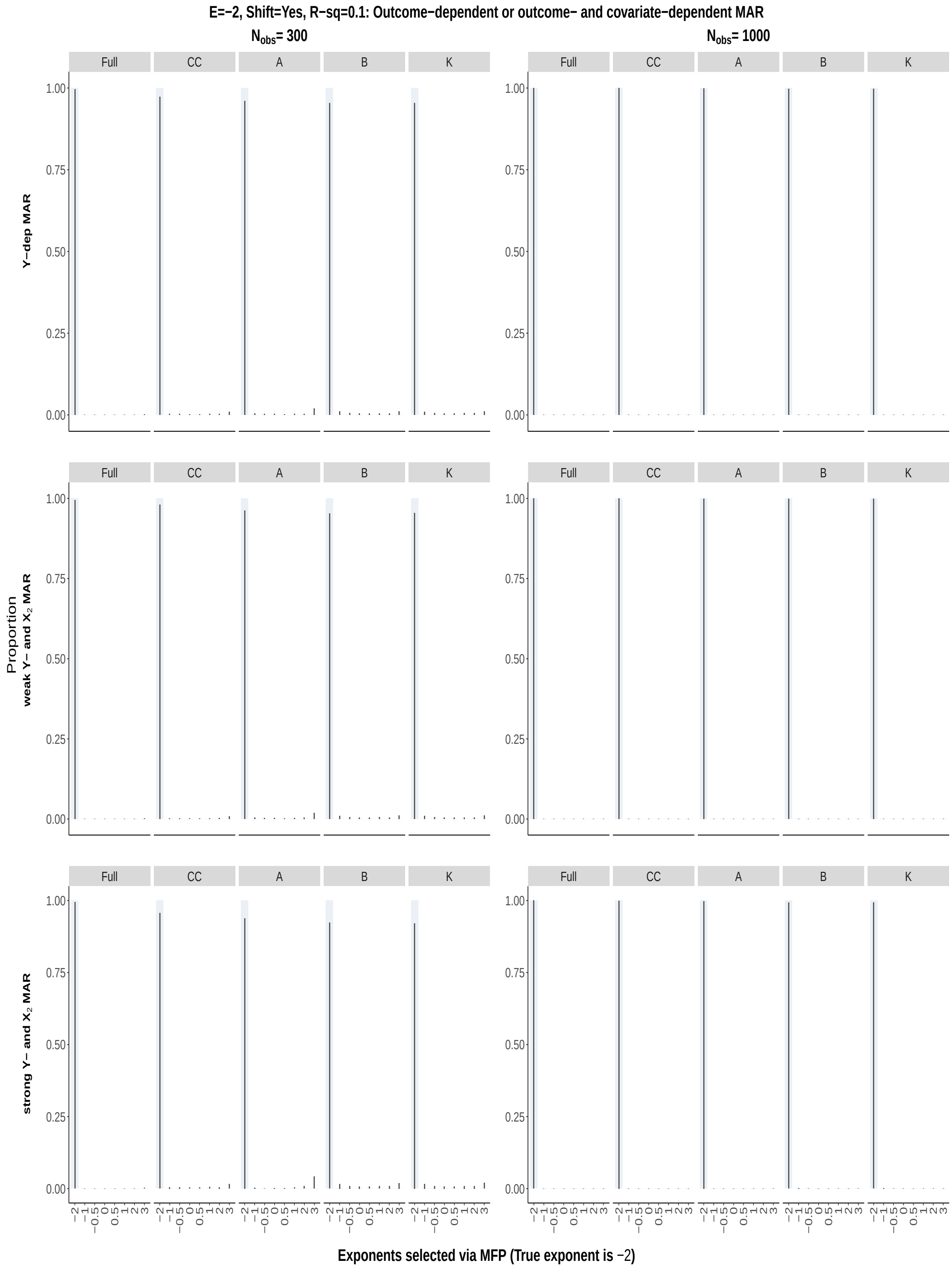

Figure S131: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

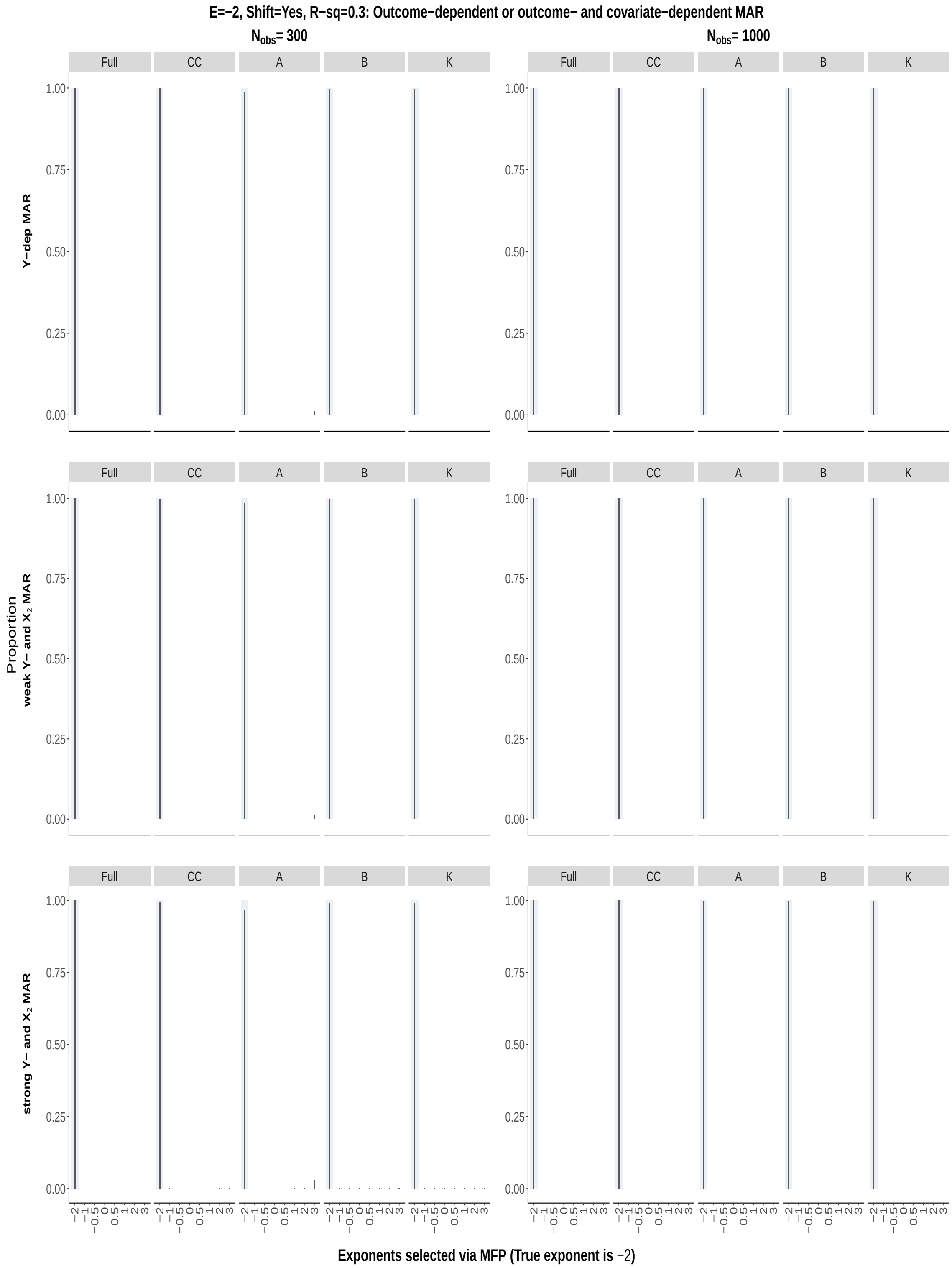

Figure S132: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.8 Cross-validation, $\beta_2 = 0$, $\alpha_E = 0.05$ and an origin-shift has been used

True exponent is 0

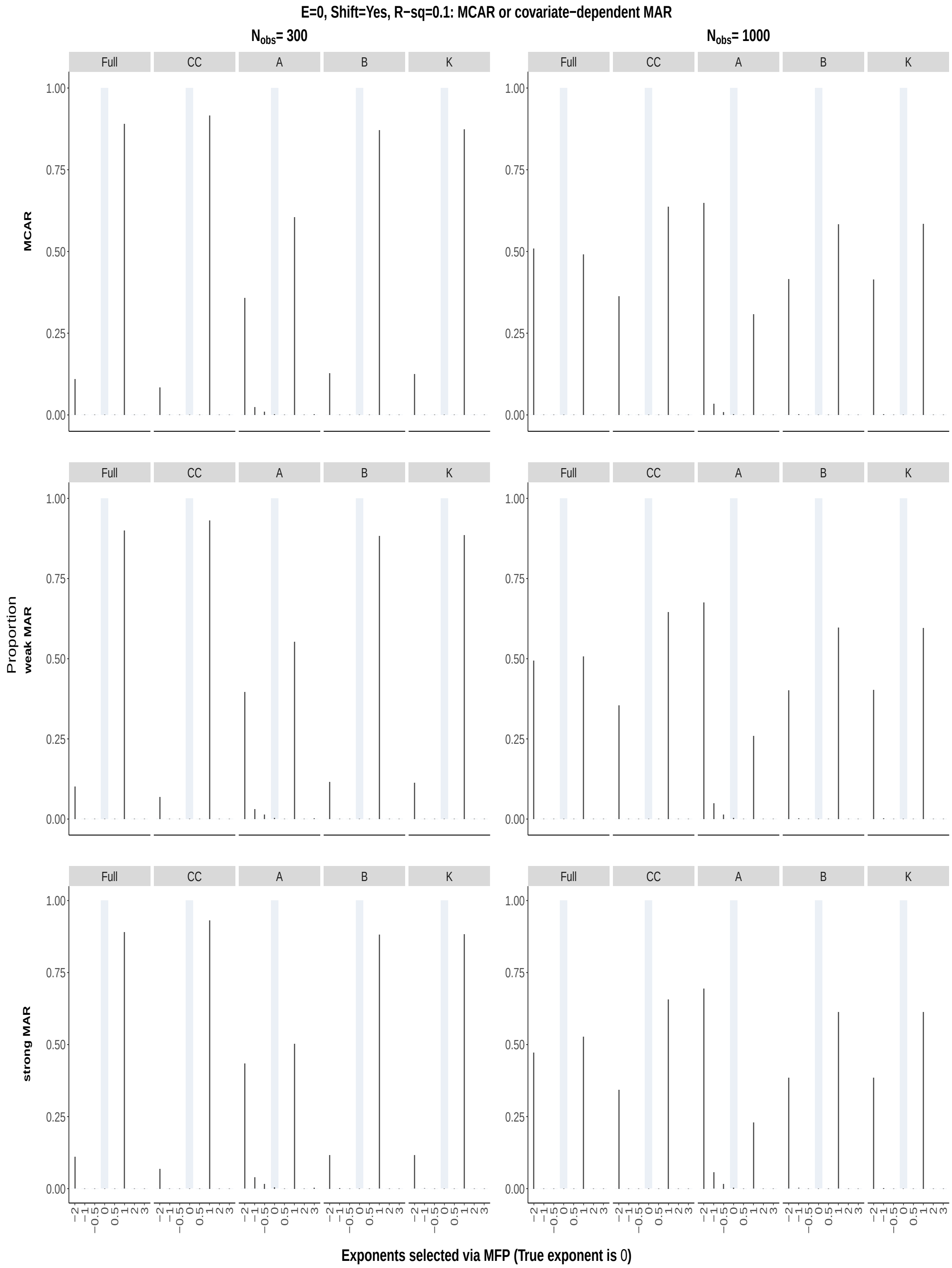

Figure S133: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

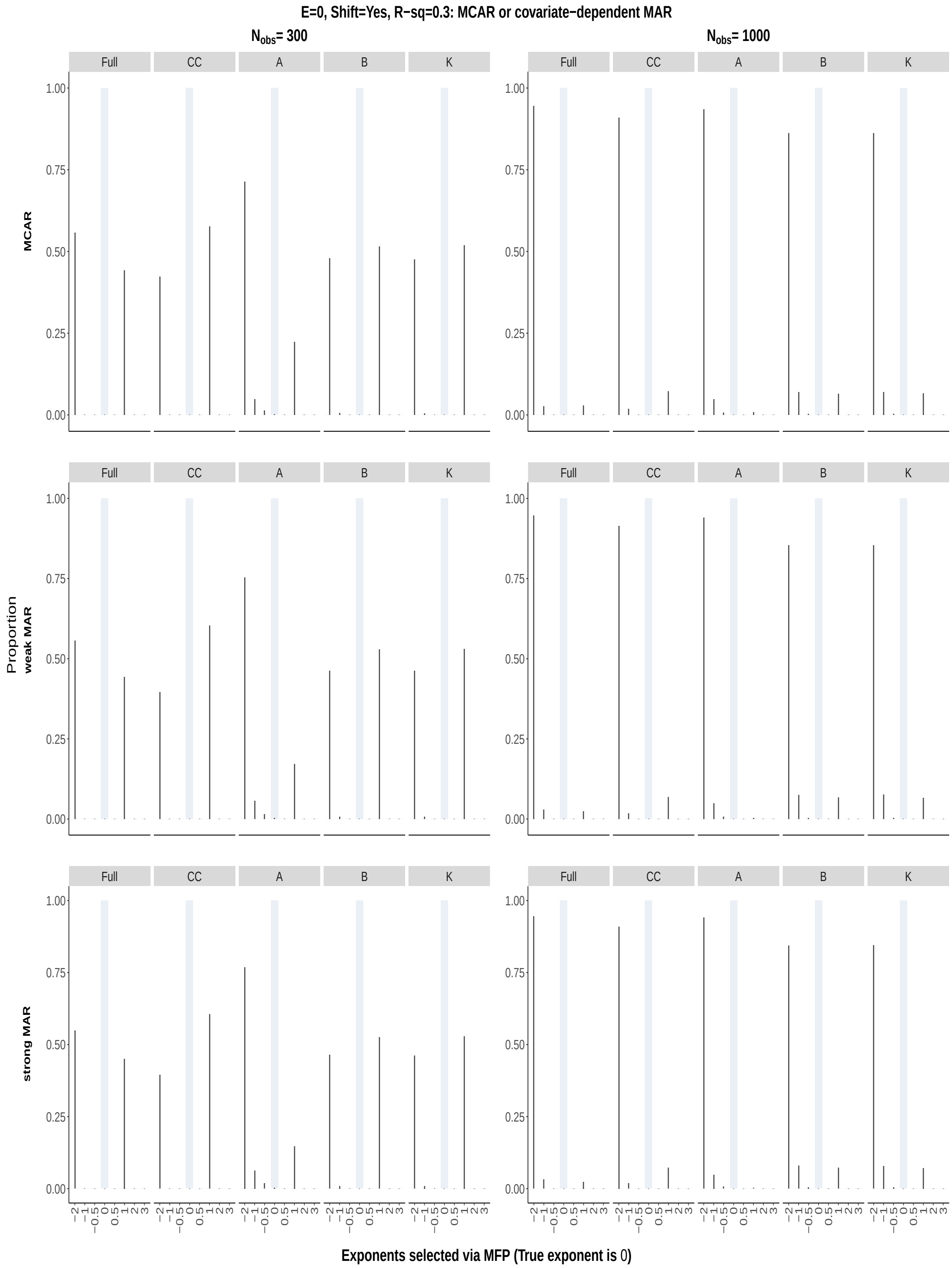

Figure S134: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

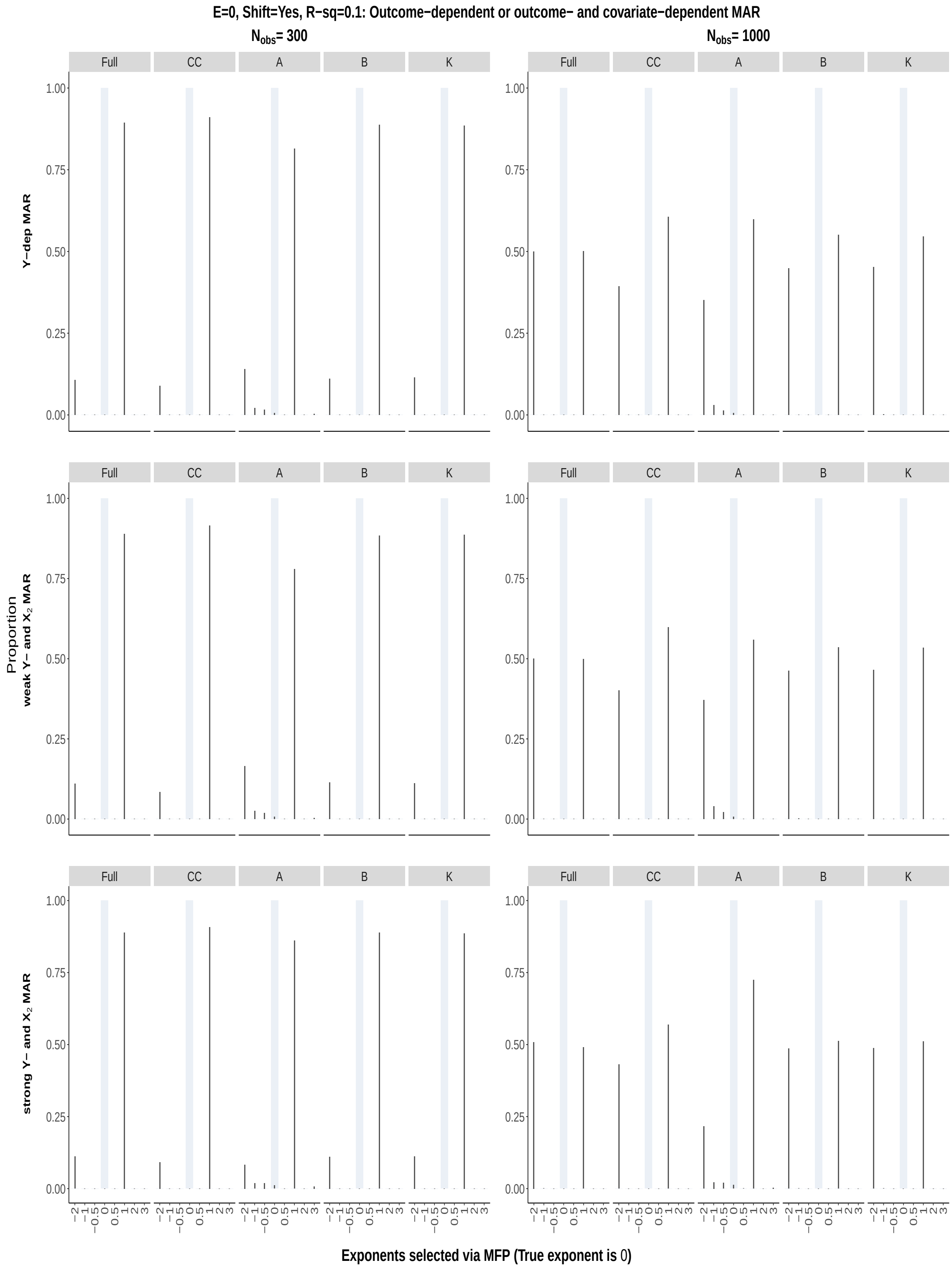

Figure S135: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

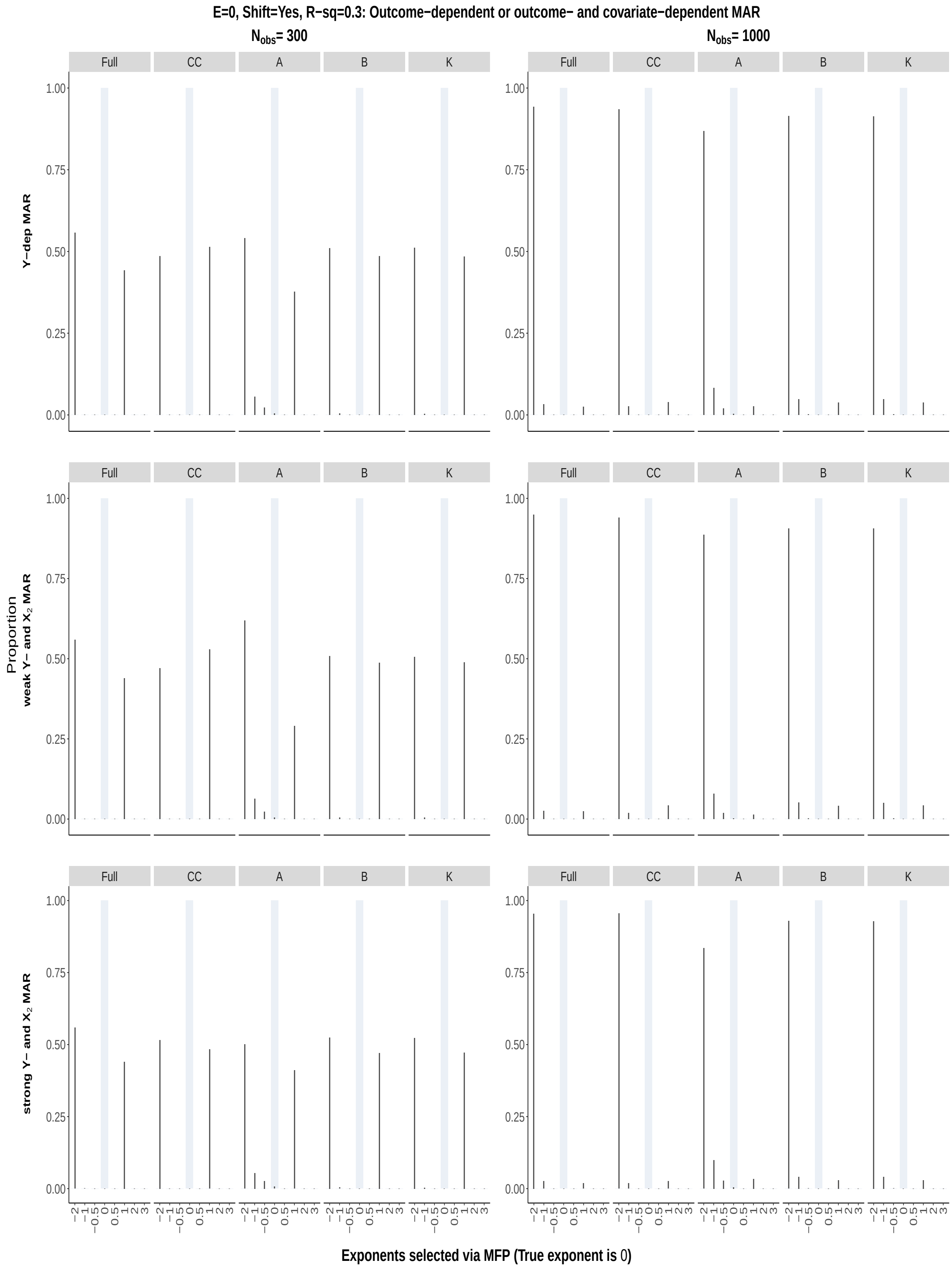

Figure S136: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

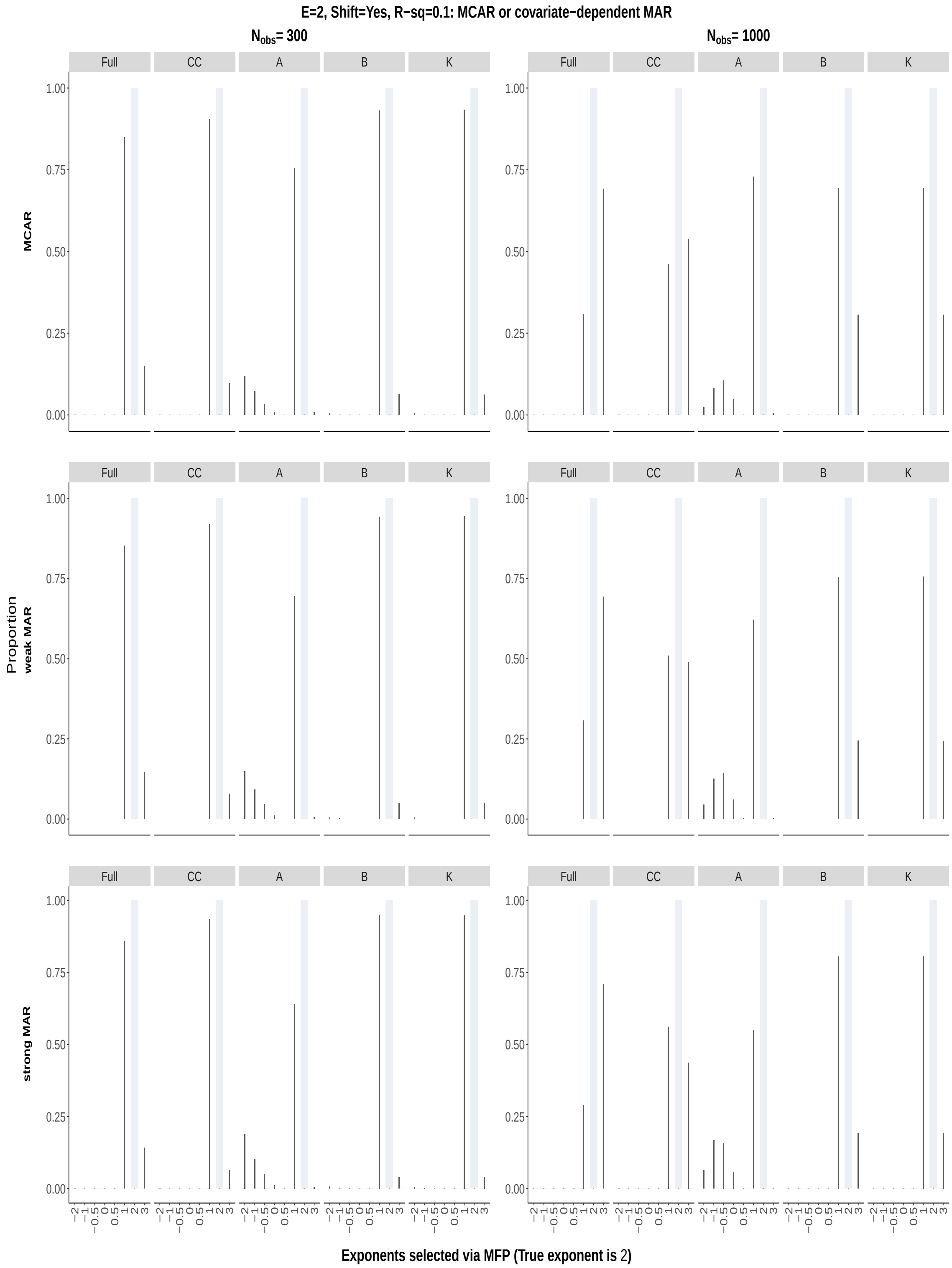

Figure S137: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

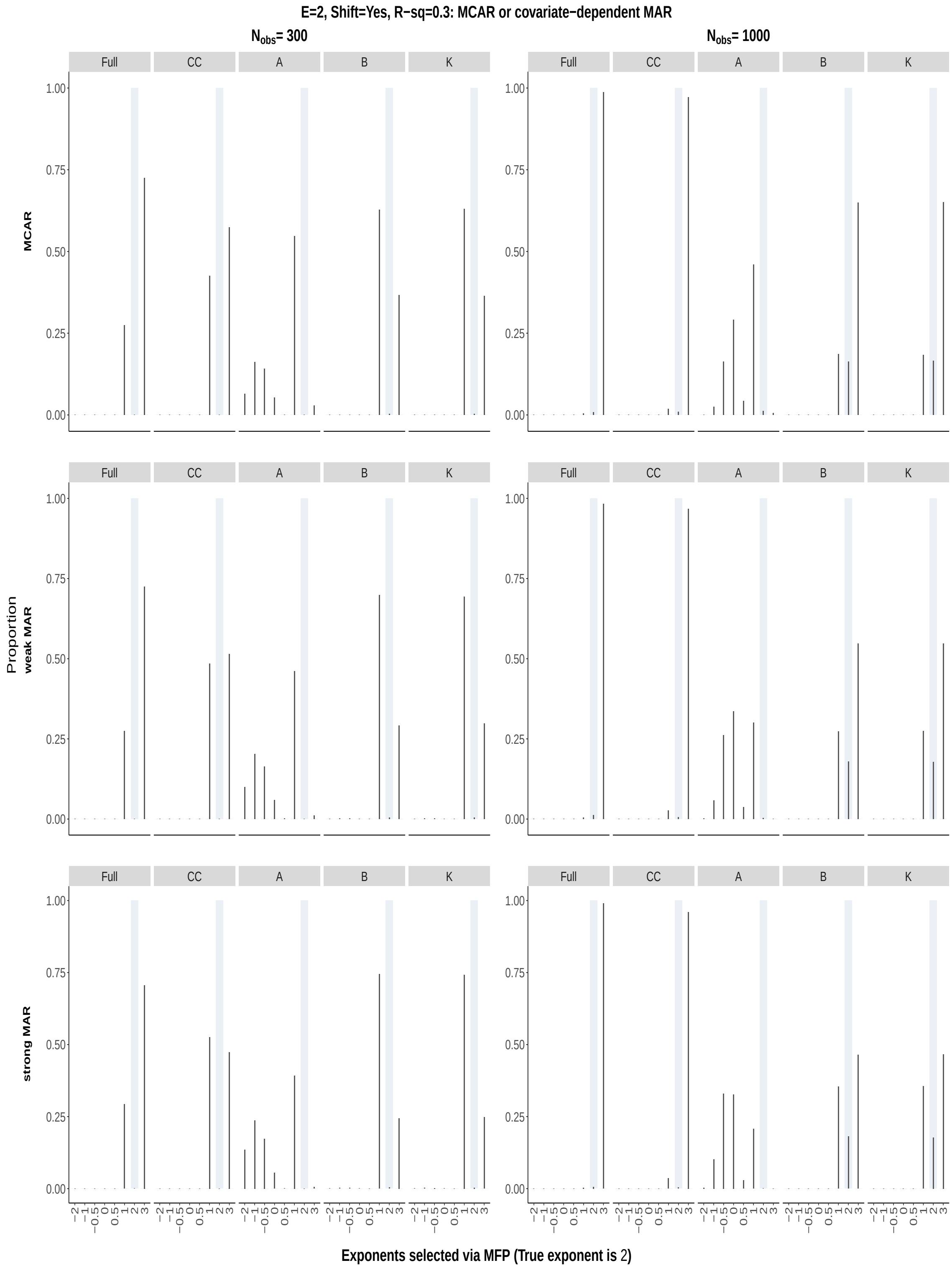

Figure S138: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

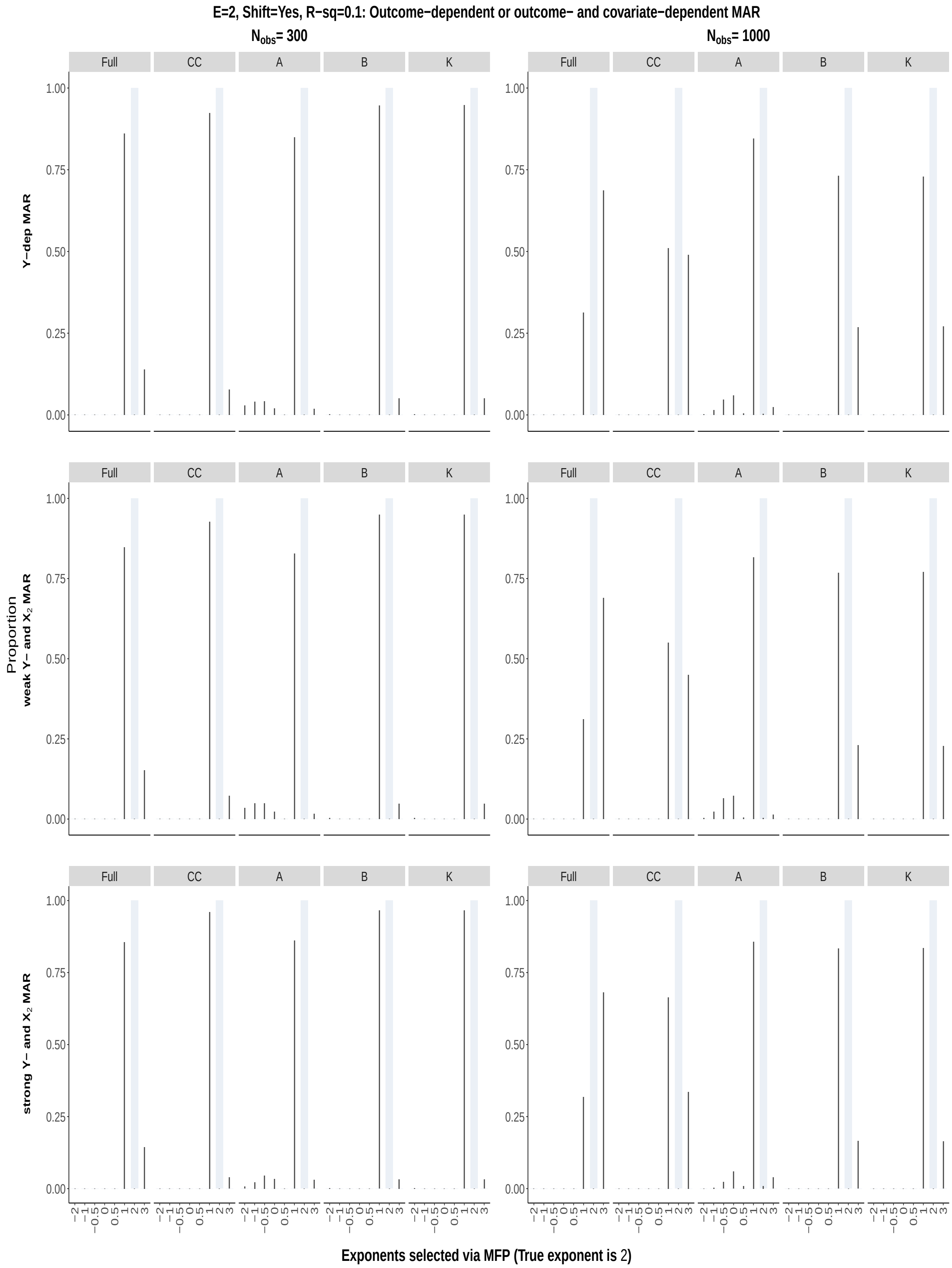

Figure S139: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

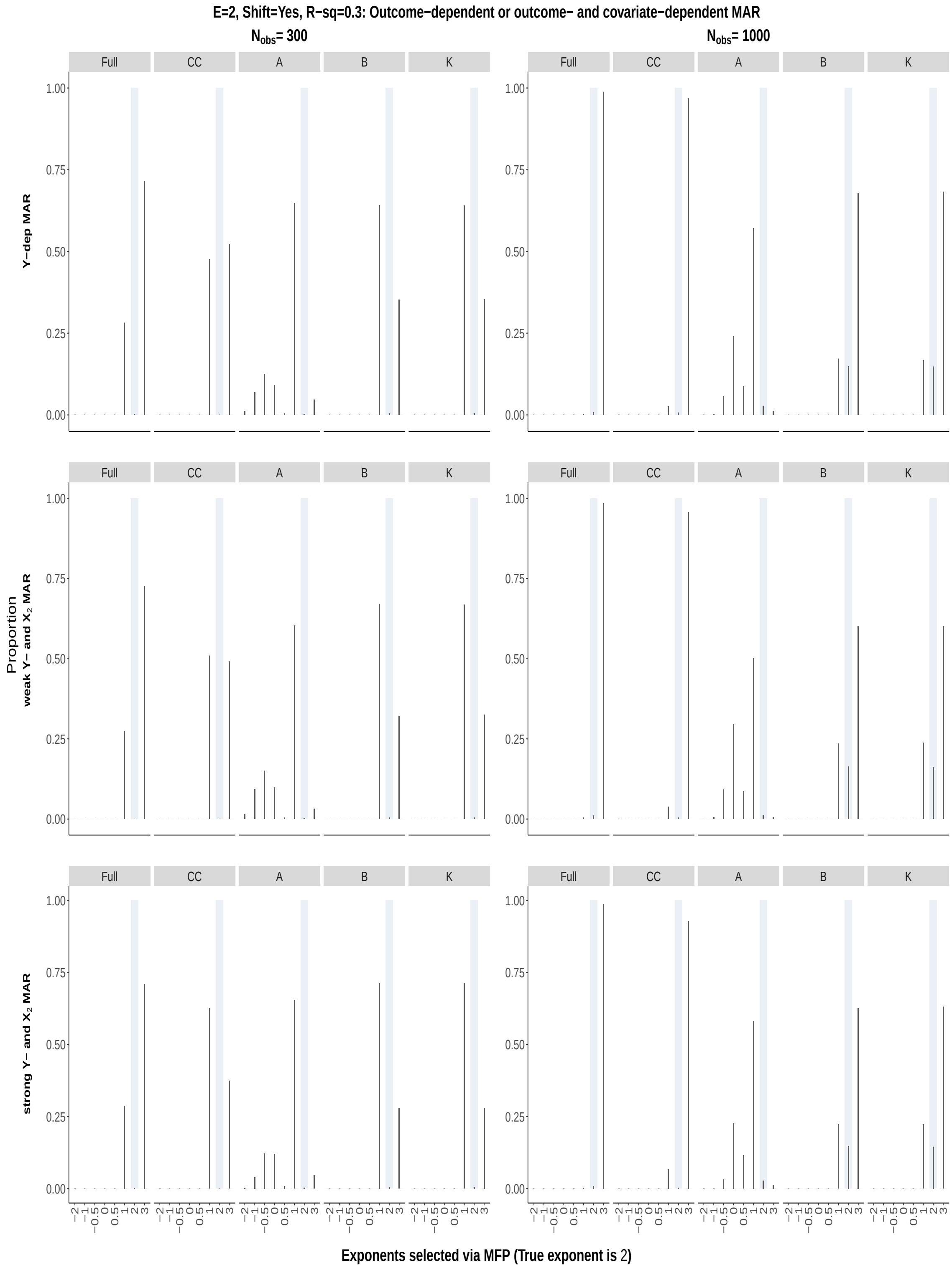

Figure S140: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

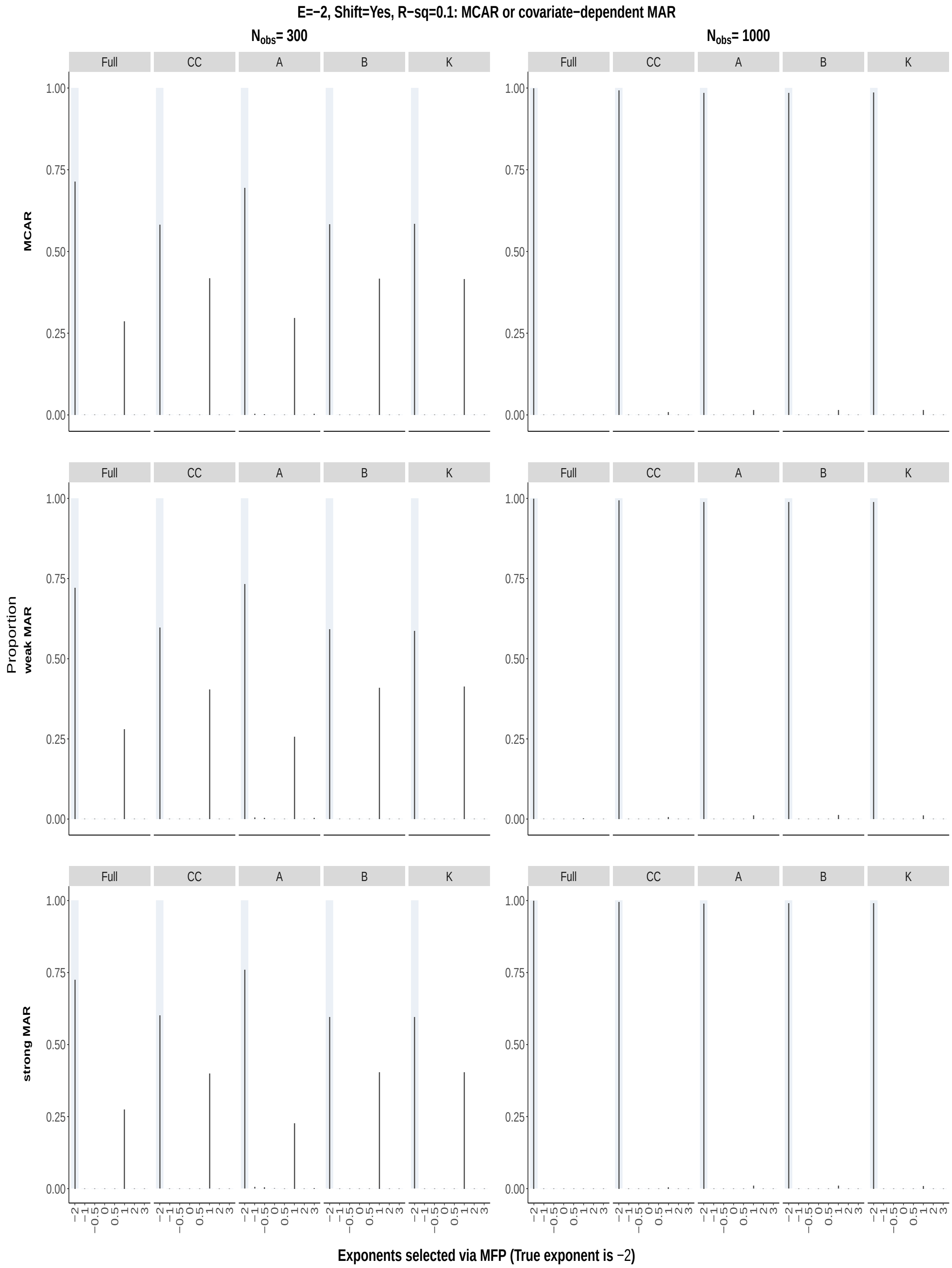

Figure S141: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

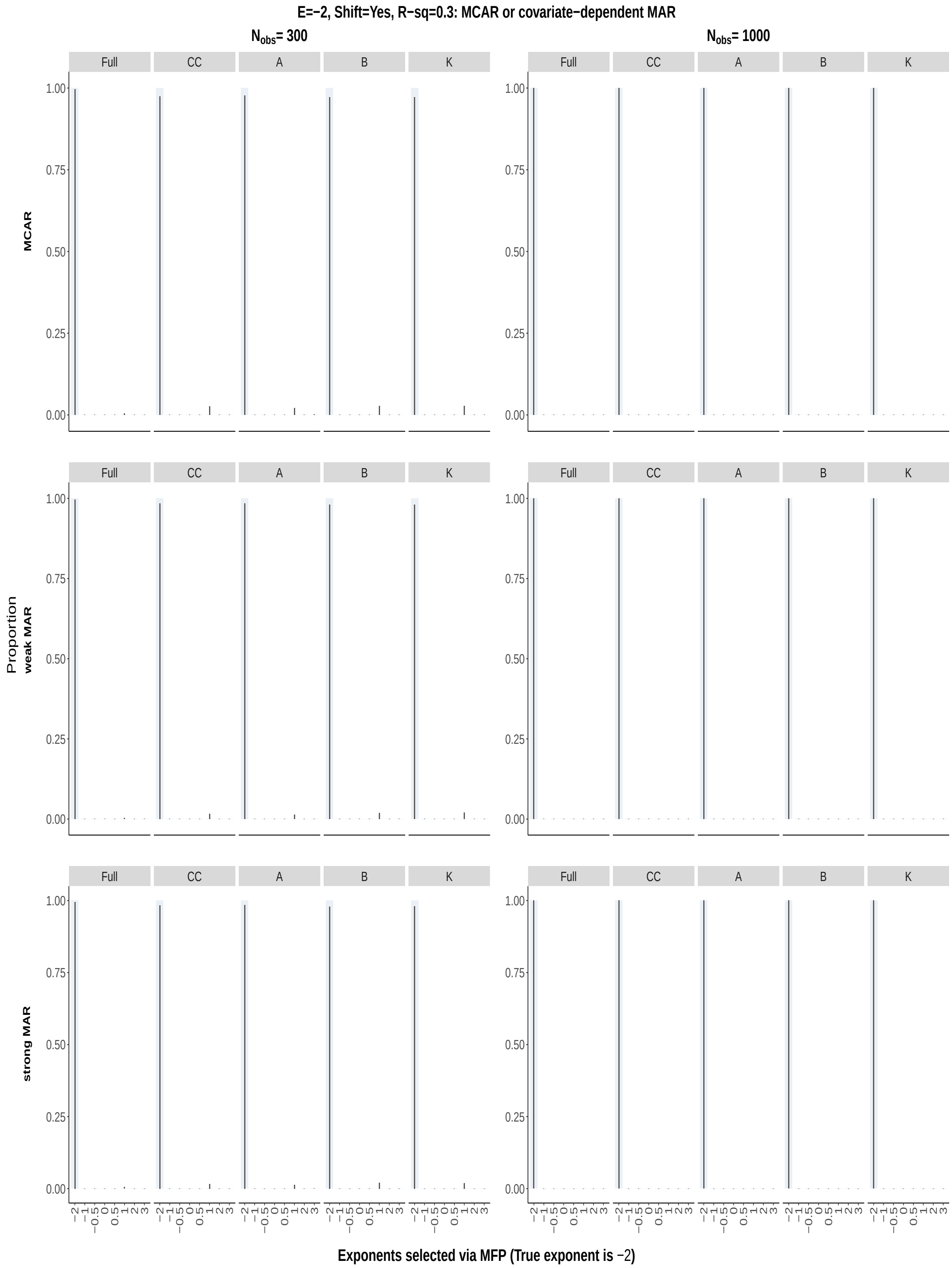

Figure S142: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

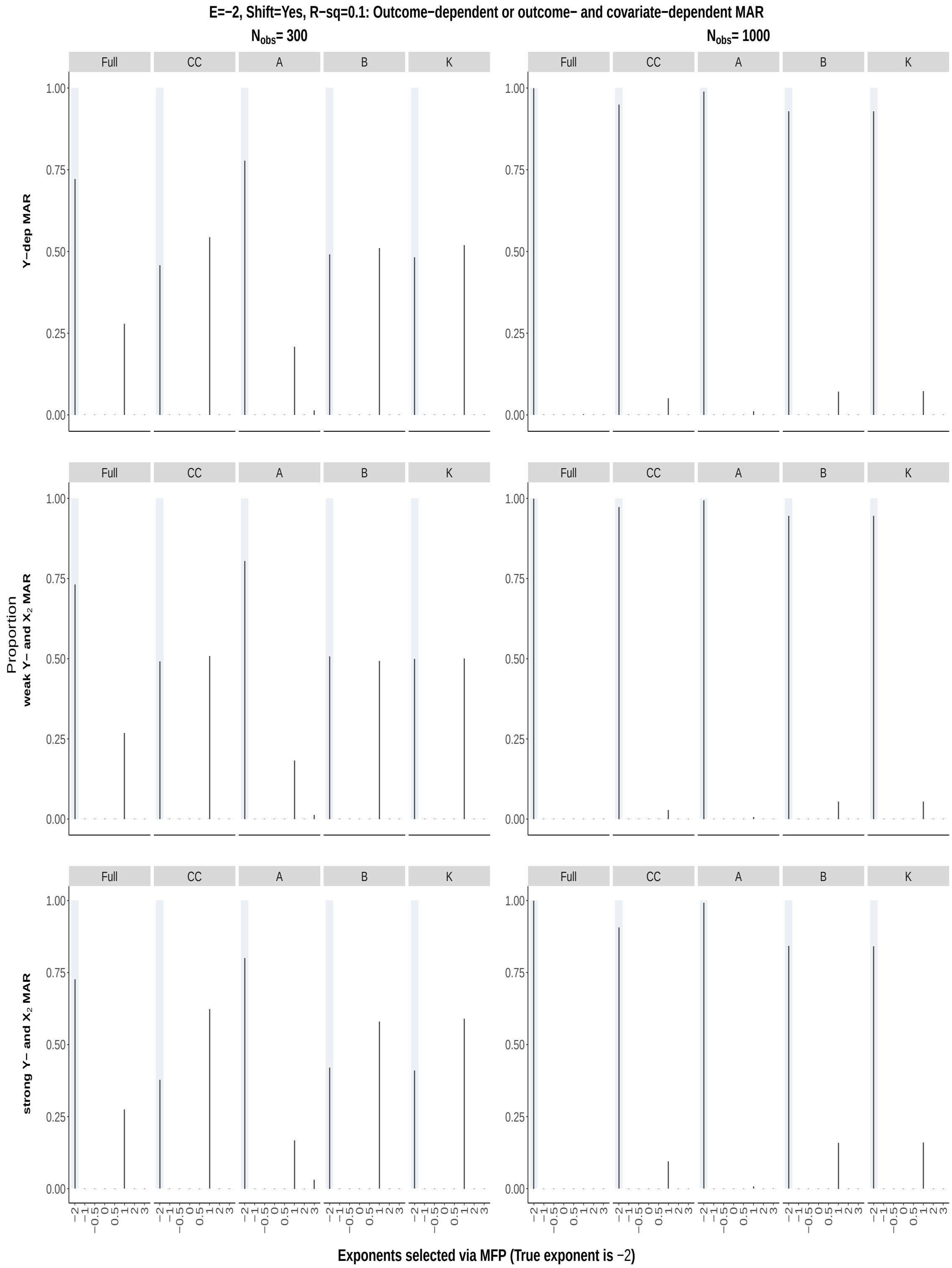

Figure S143: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

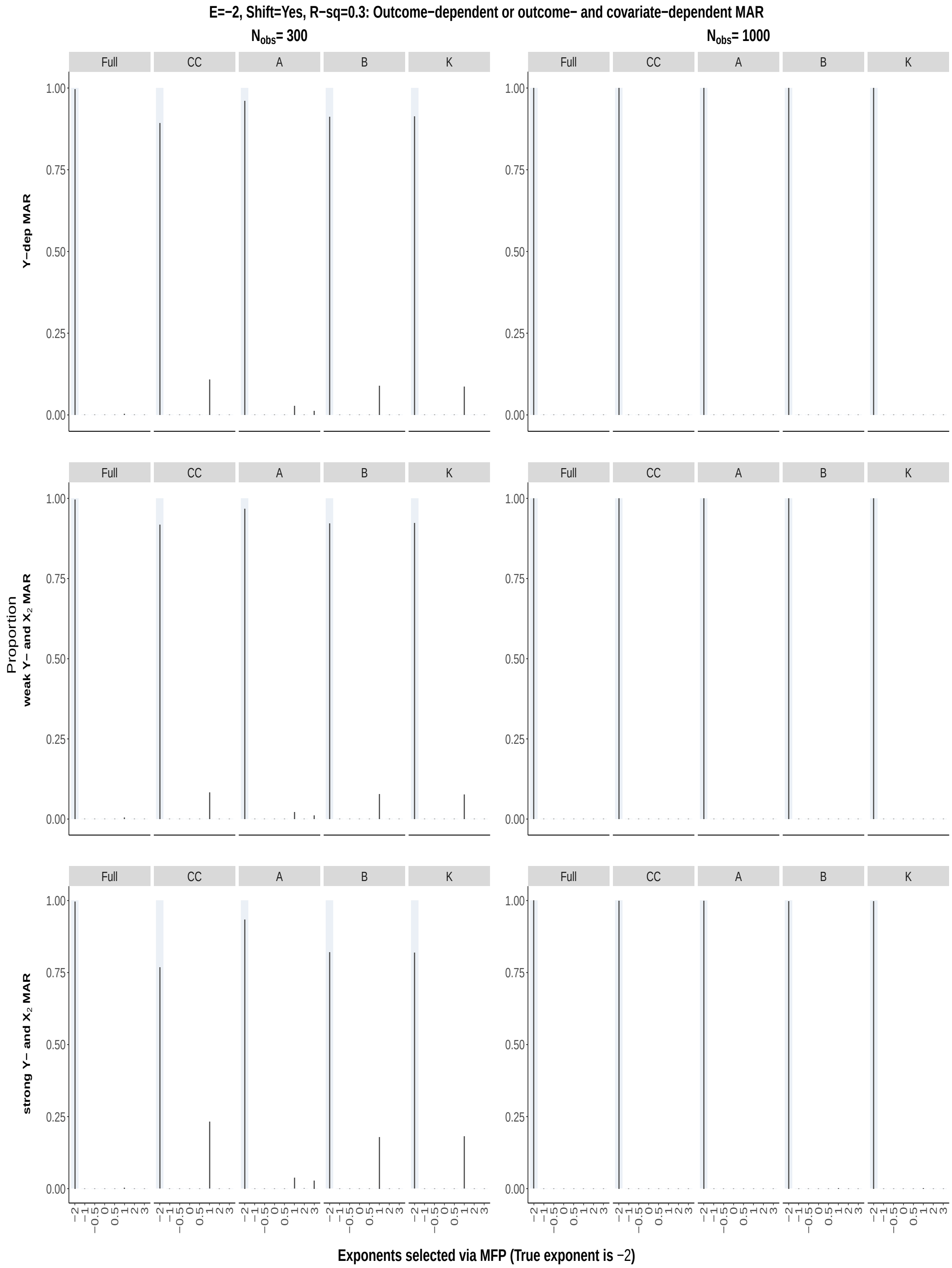

Figure S144: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.9 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 1$ and no origin-shift

True exponent is 0

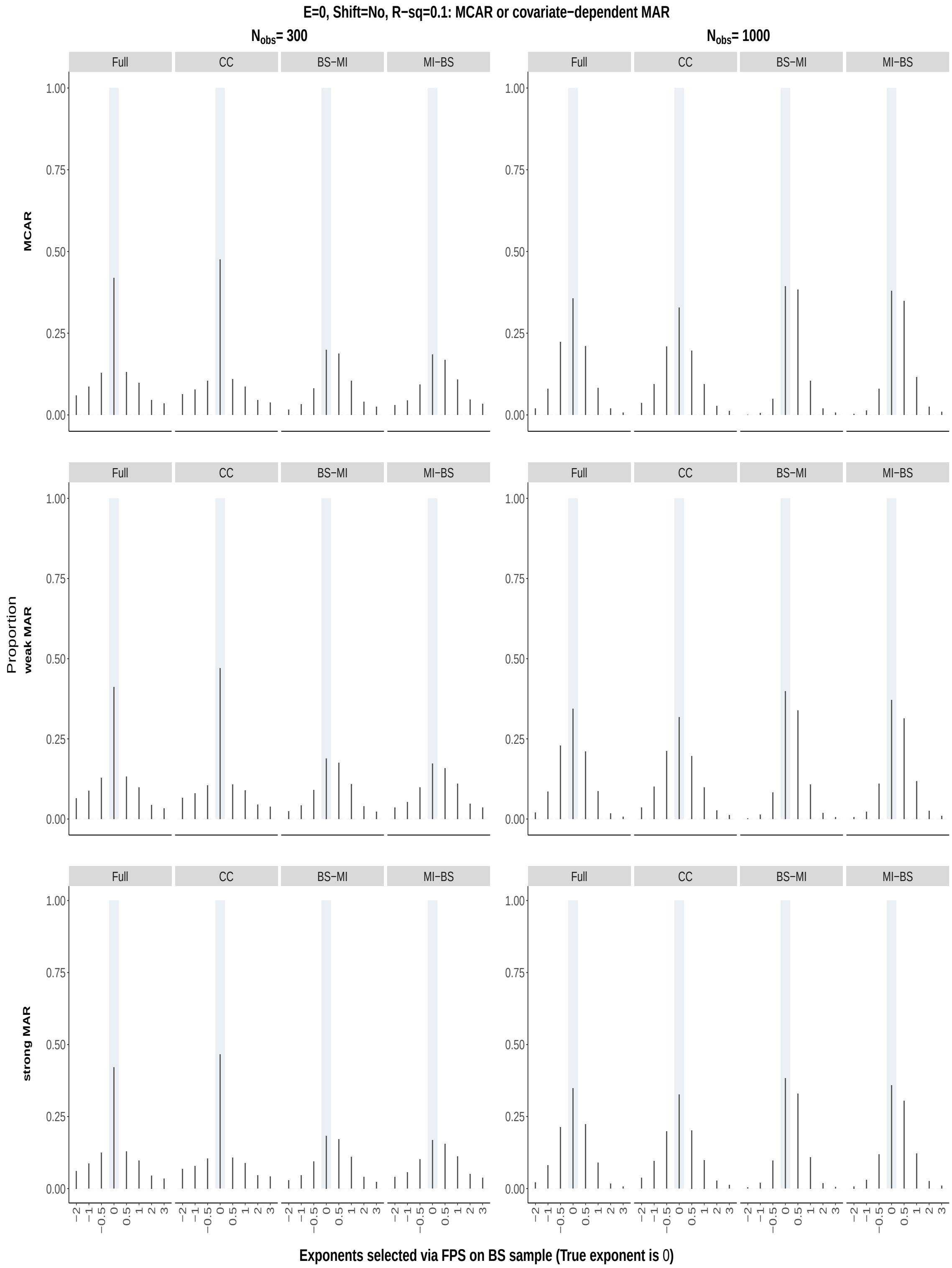

Figure S145: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

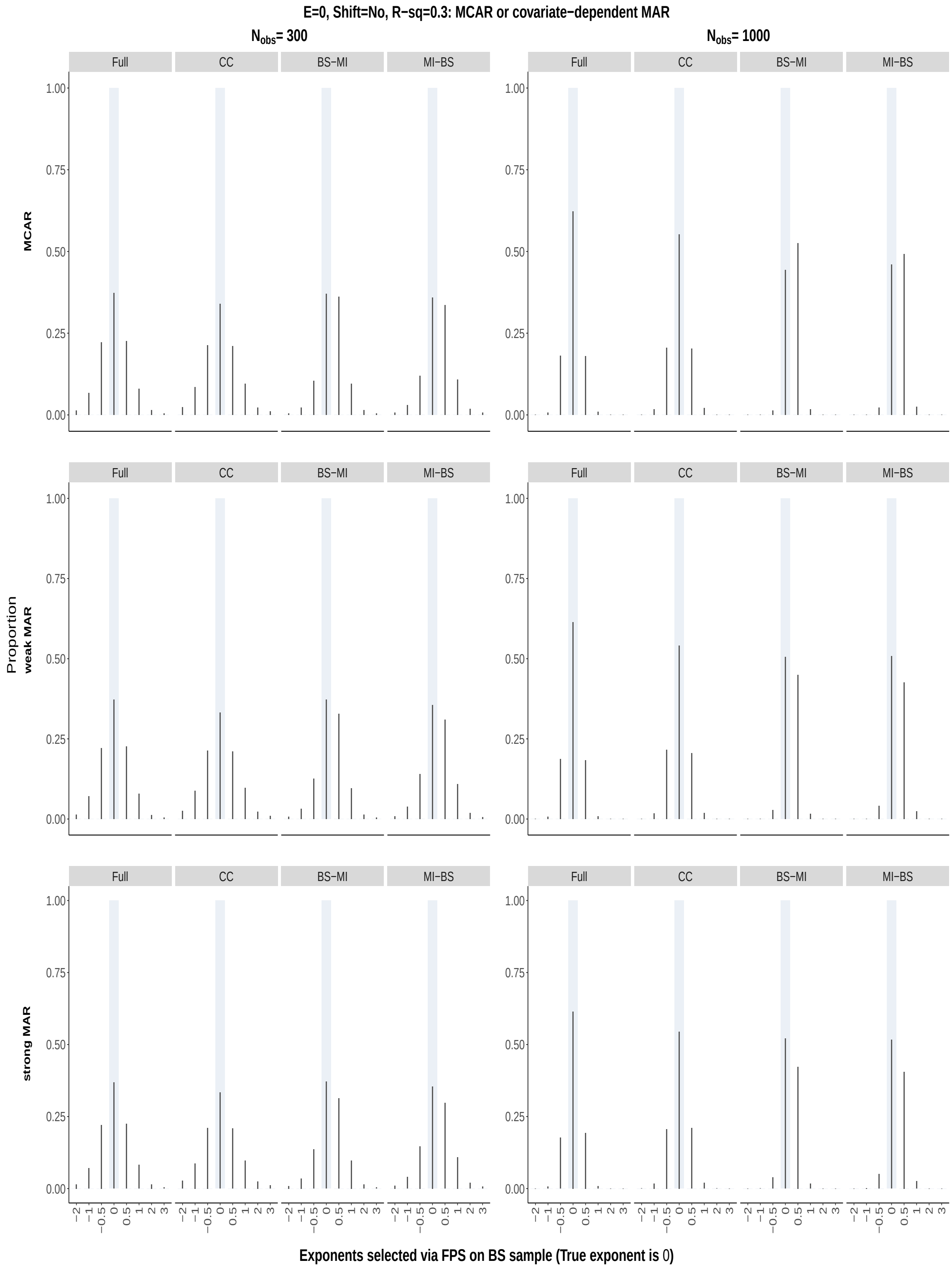

Figure S146: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

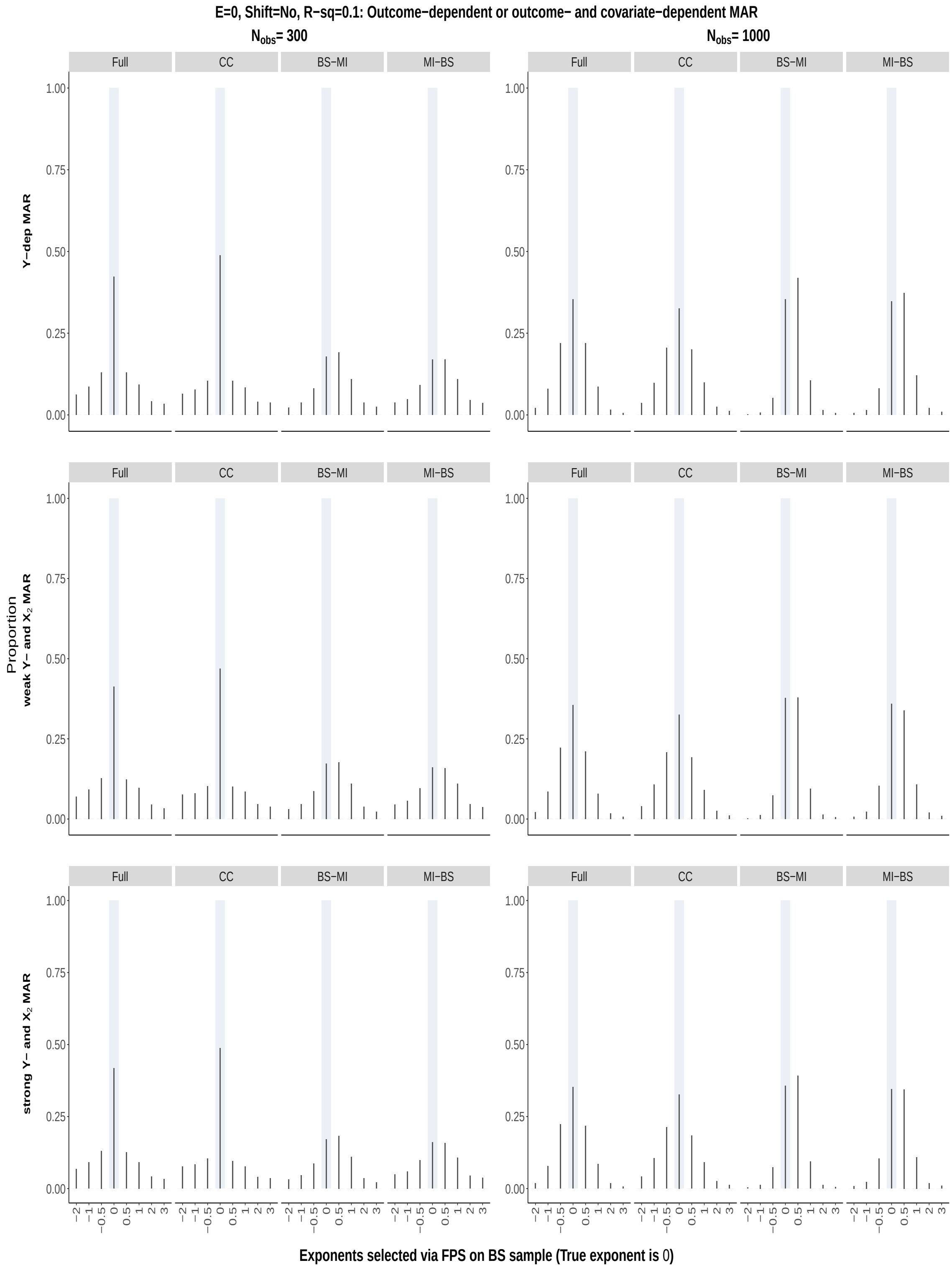

Figure S147: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

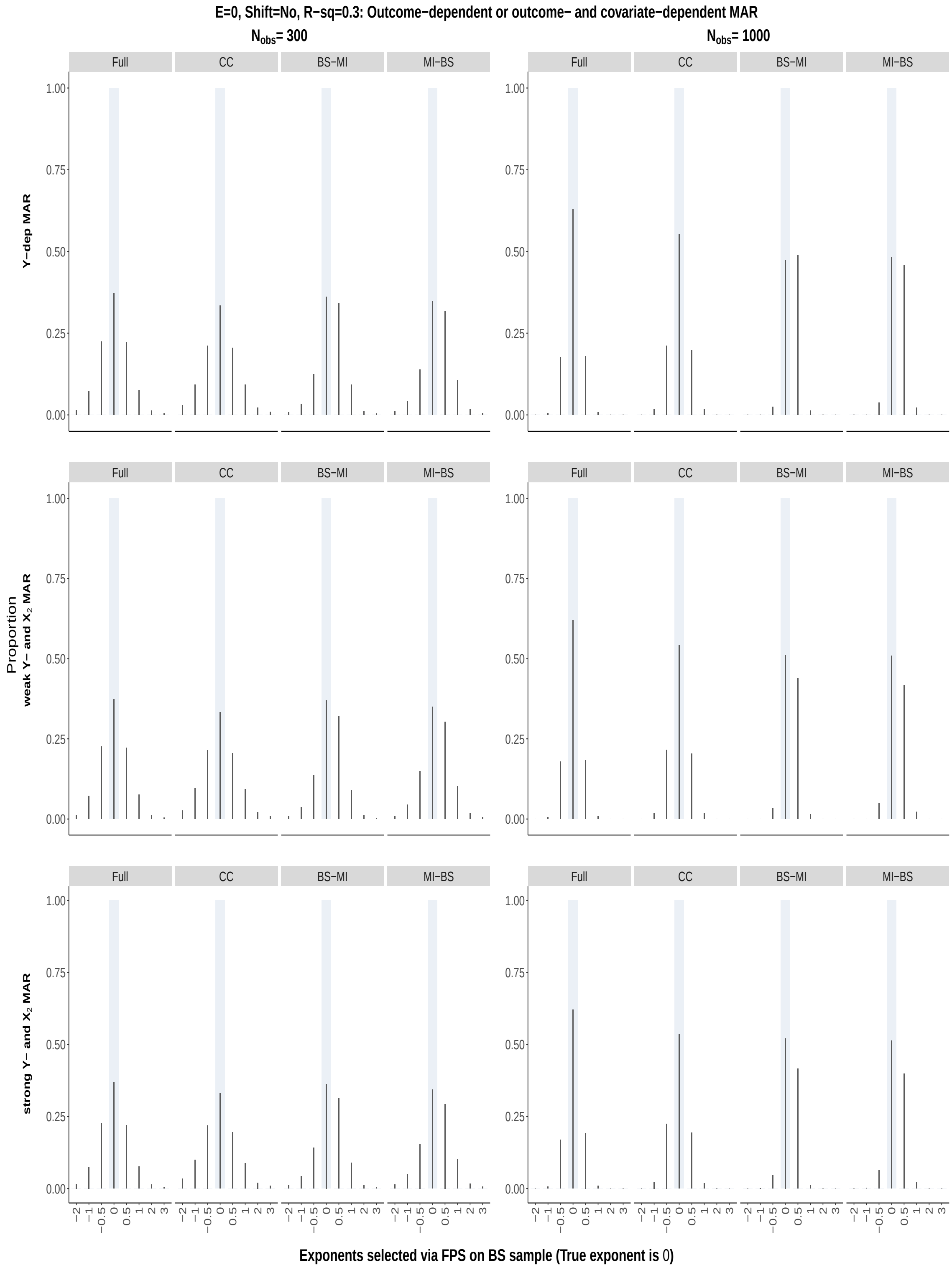

Figure S148: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

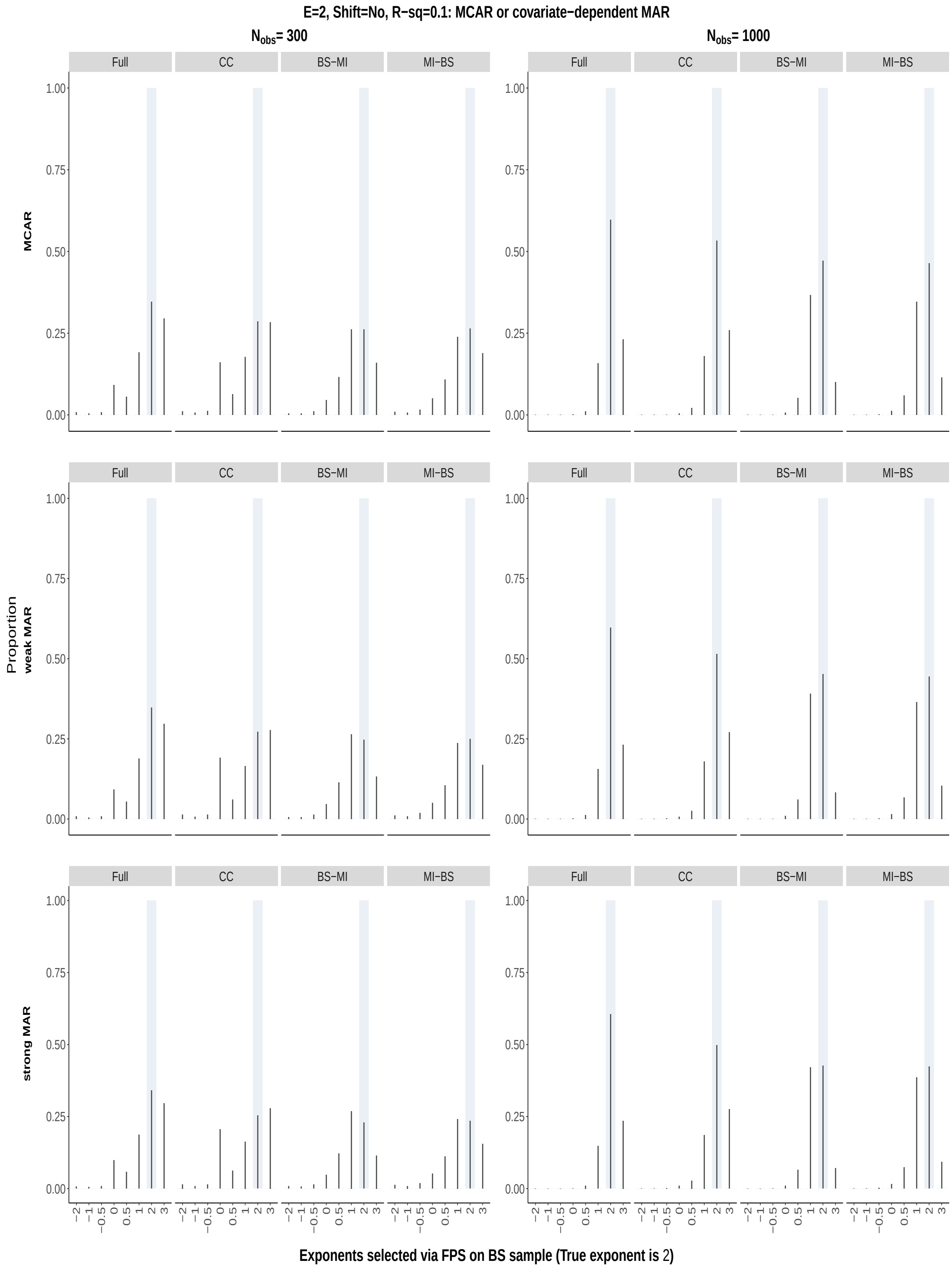

Figure S149: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

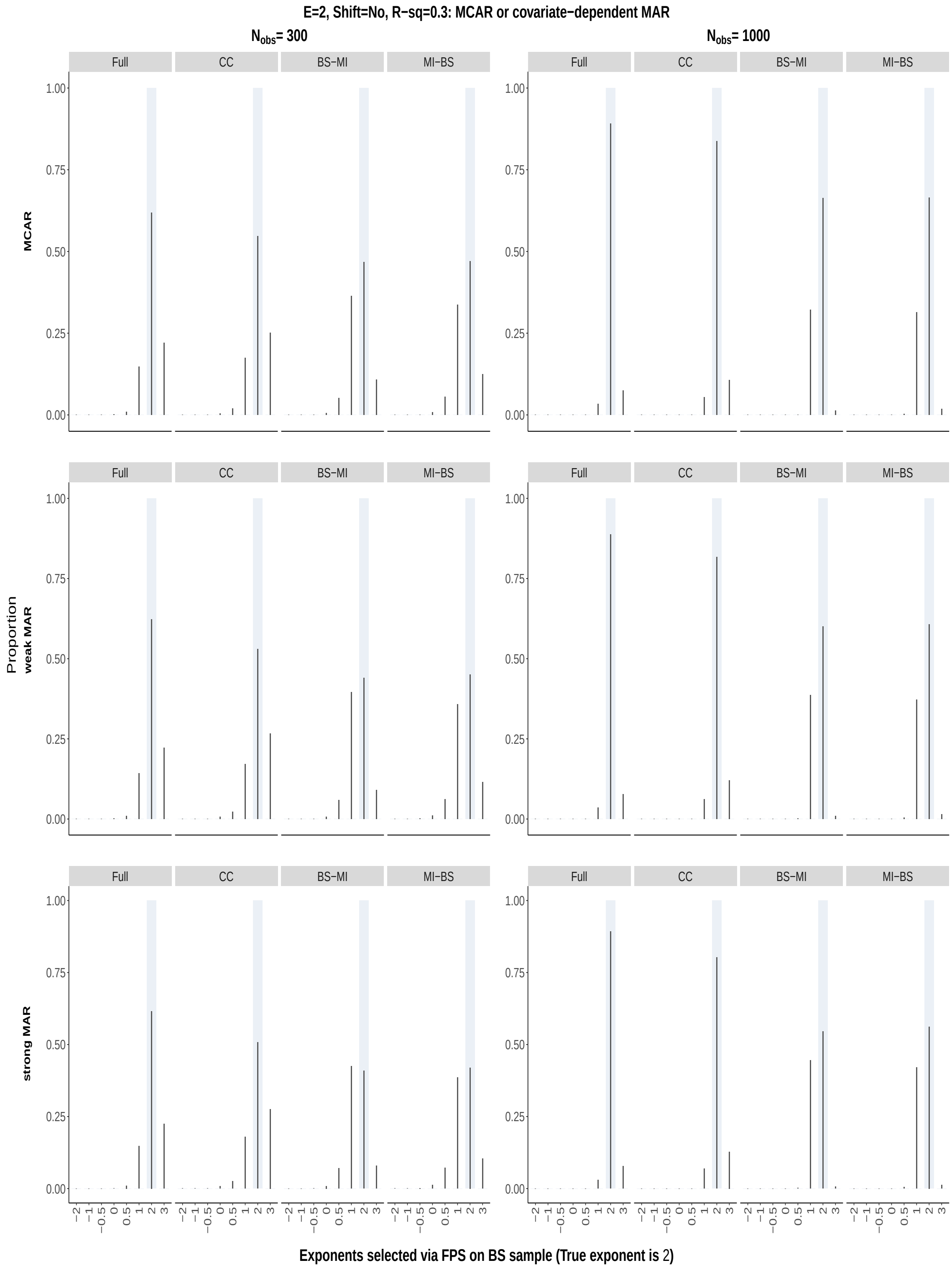

Figure S150: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

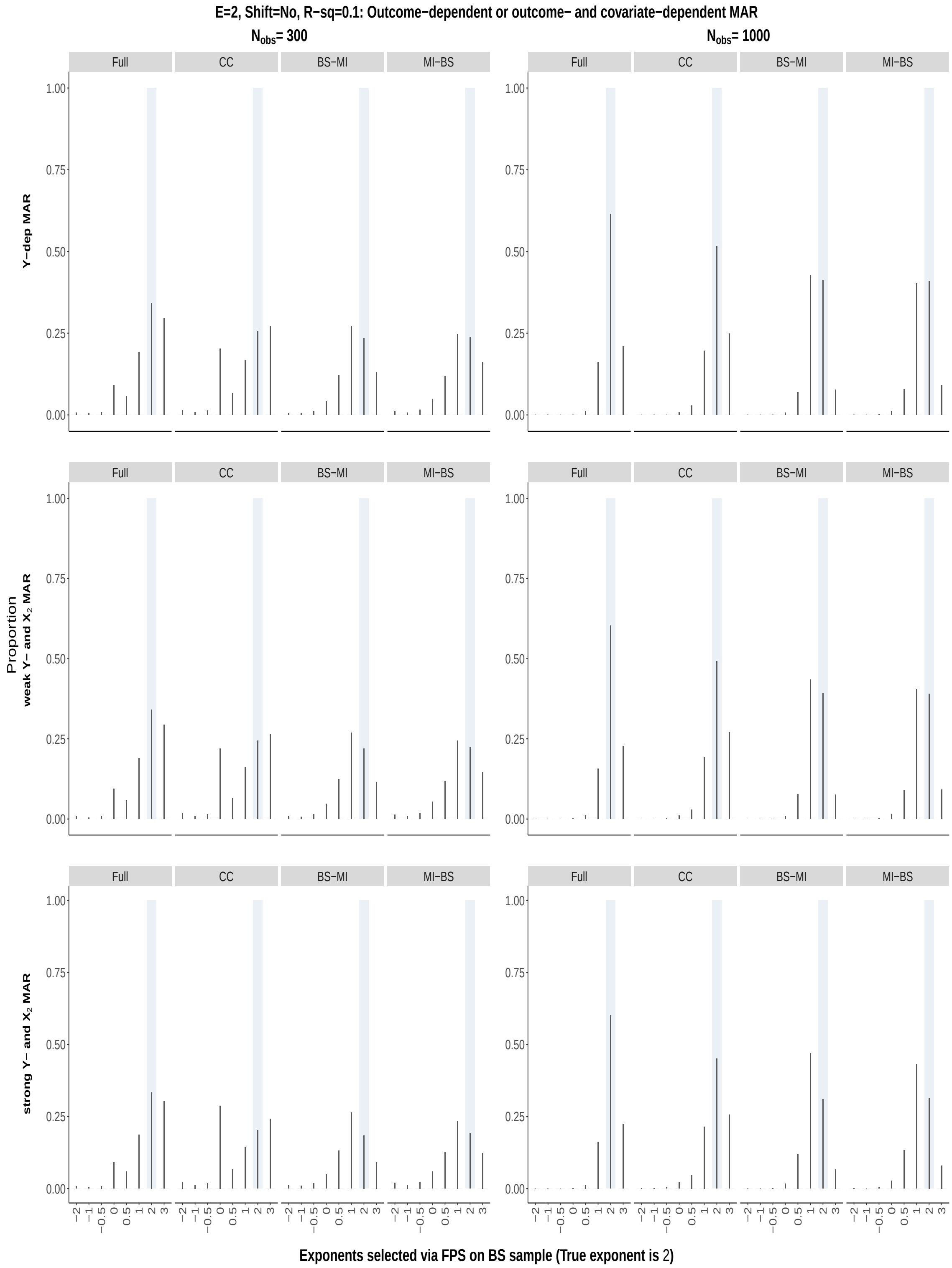

Figure S151: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

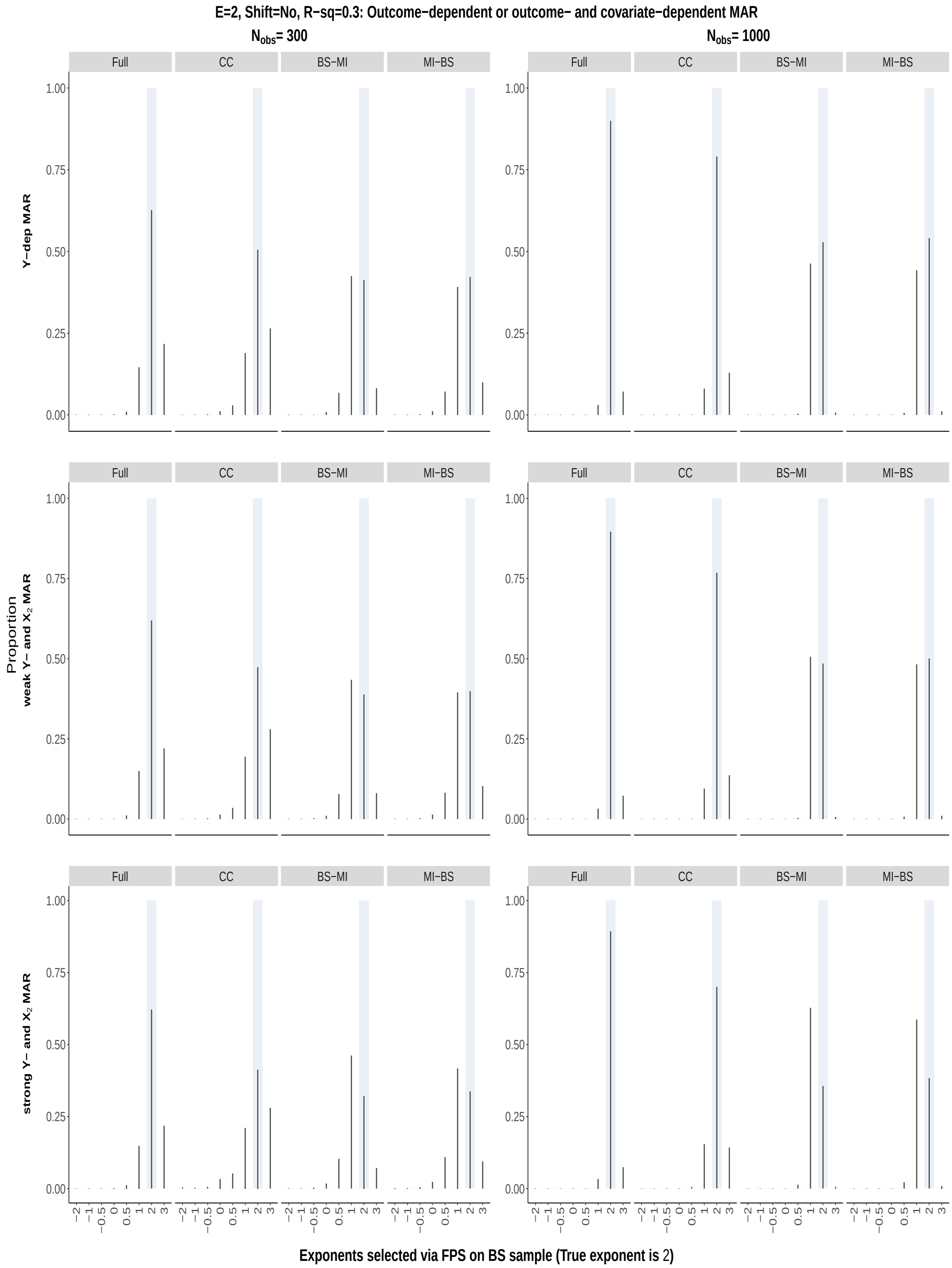

Figure S152: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

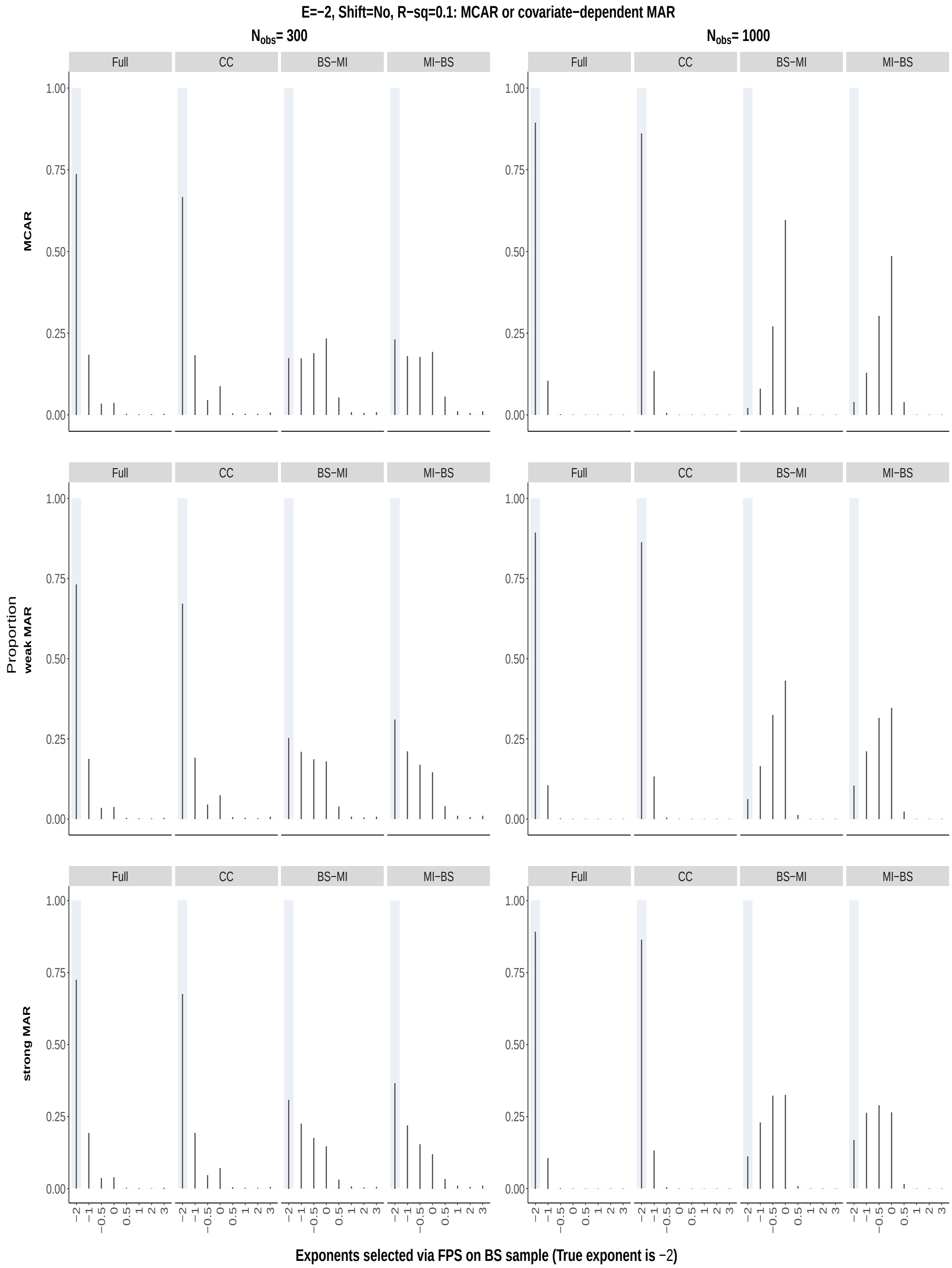

Figure S153: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

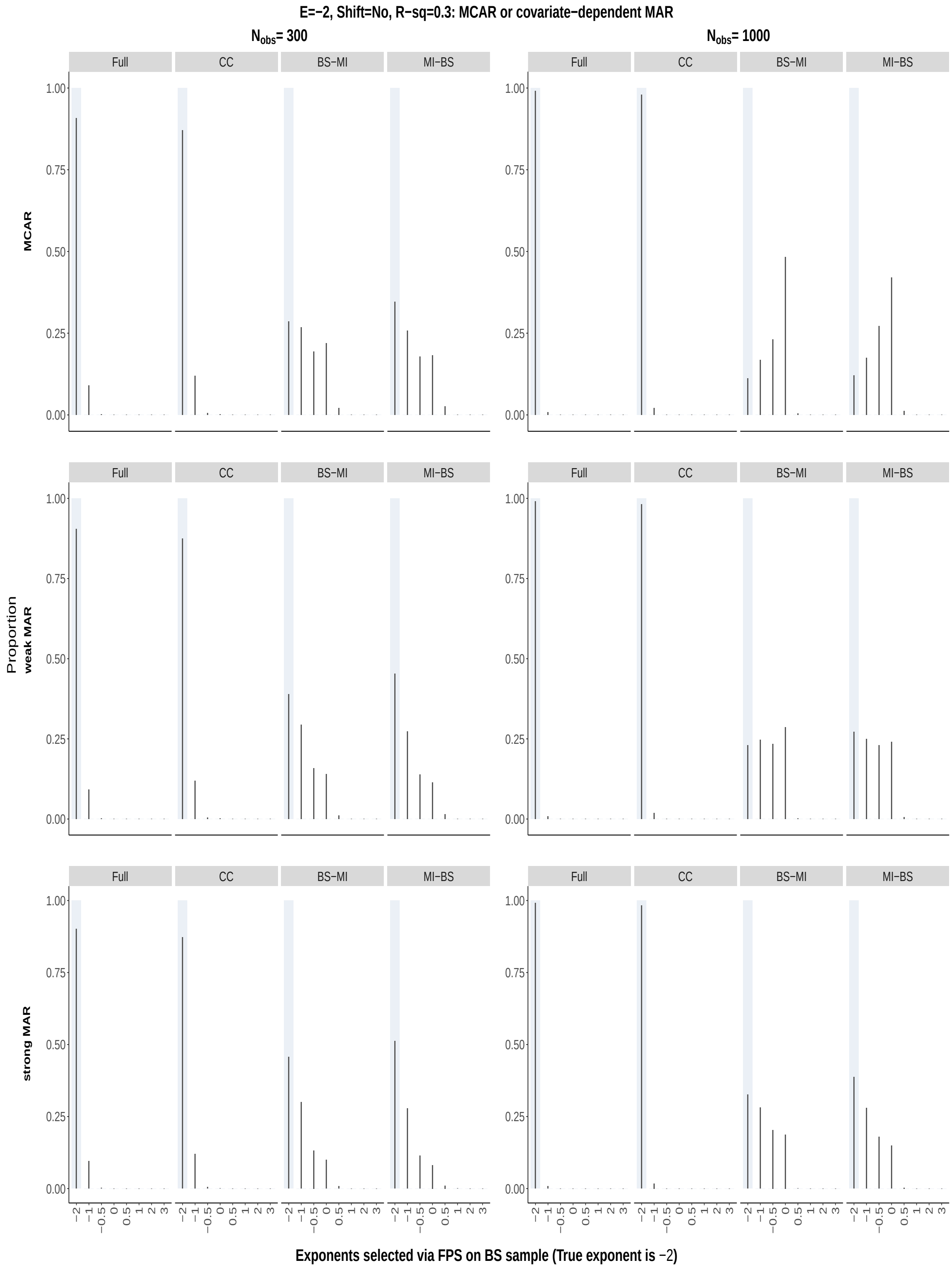

Figure S154: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

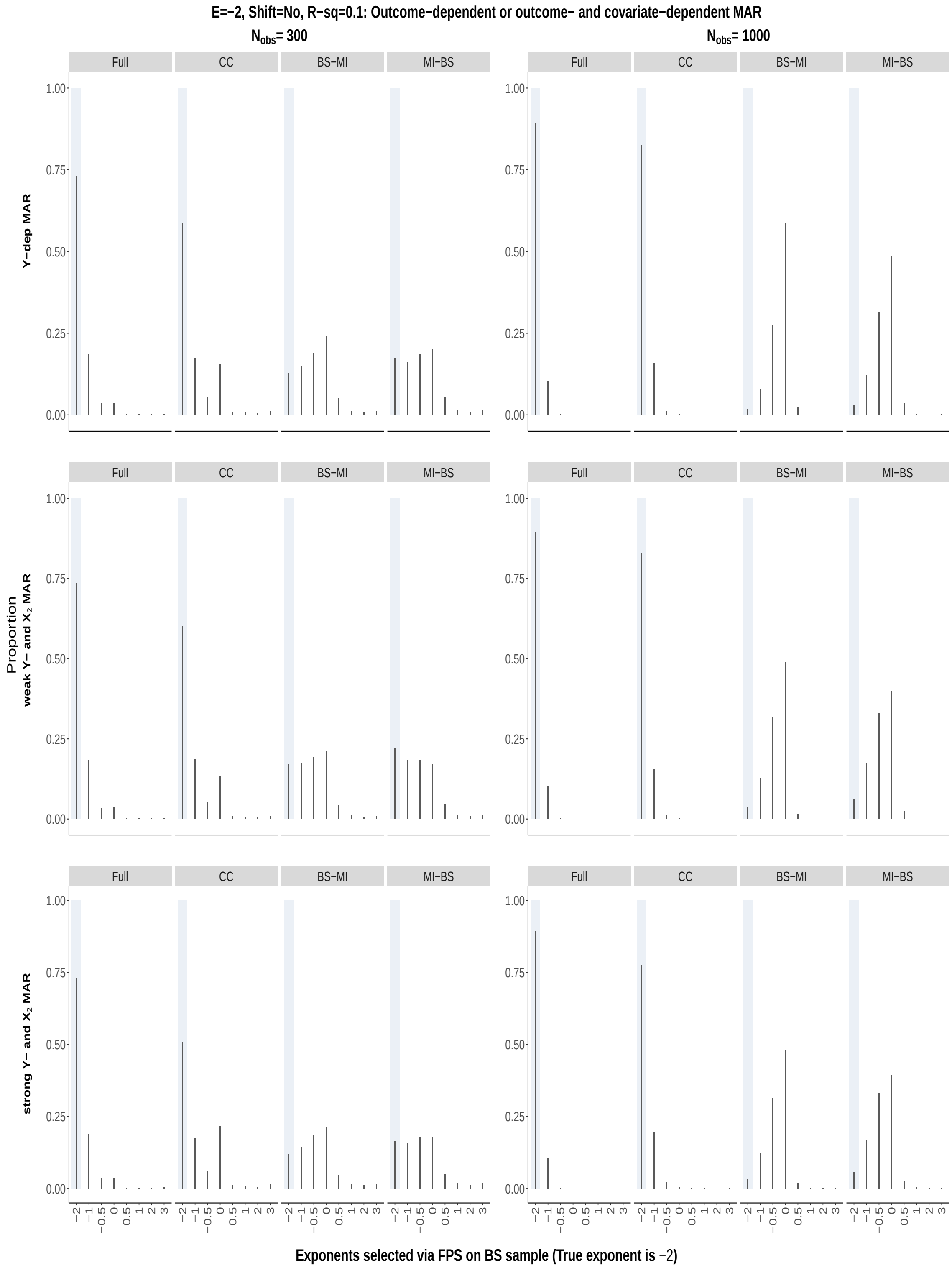

Figure S155: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

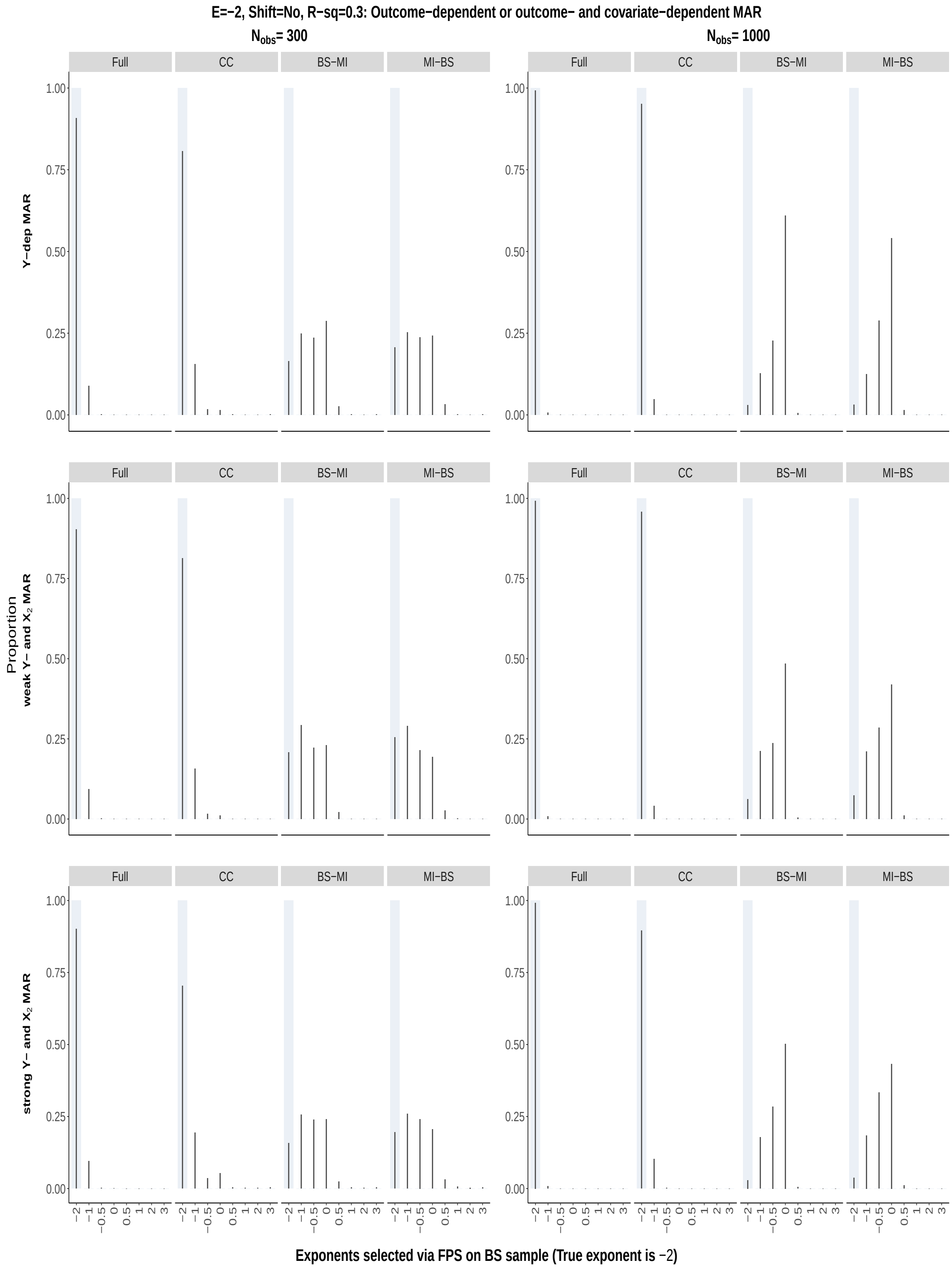

Figure S156: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.10 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

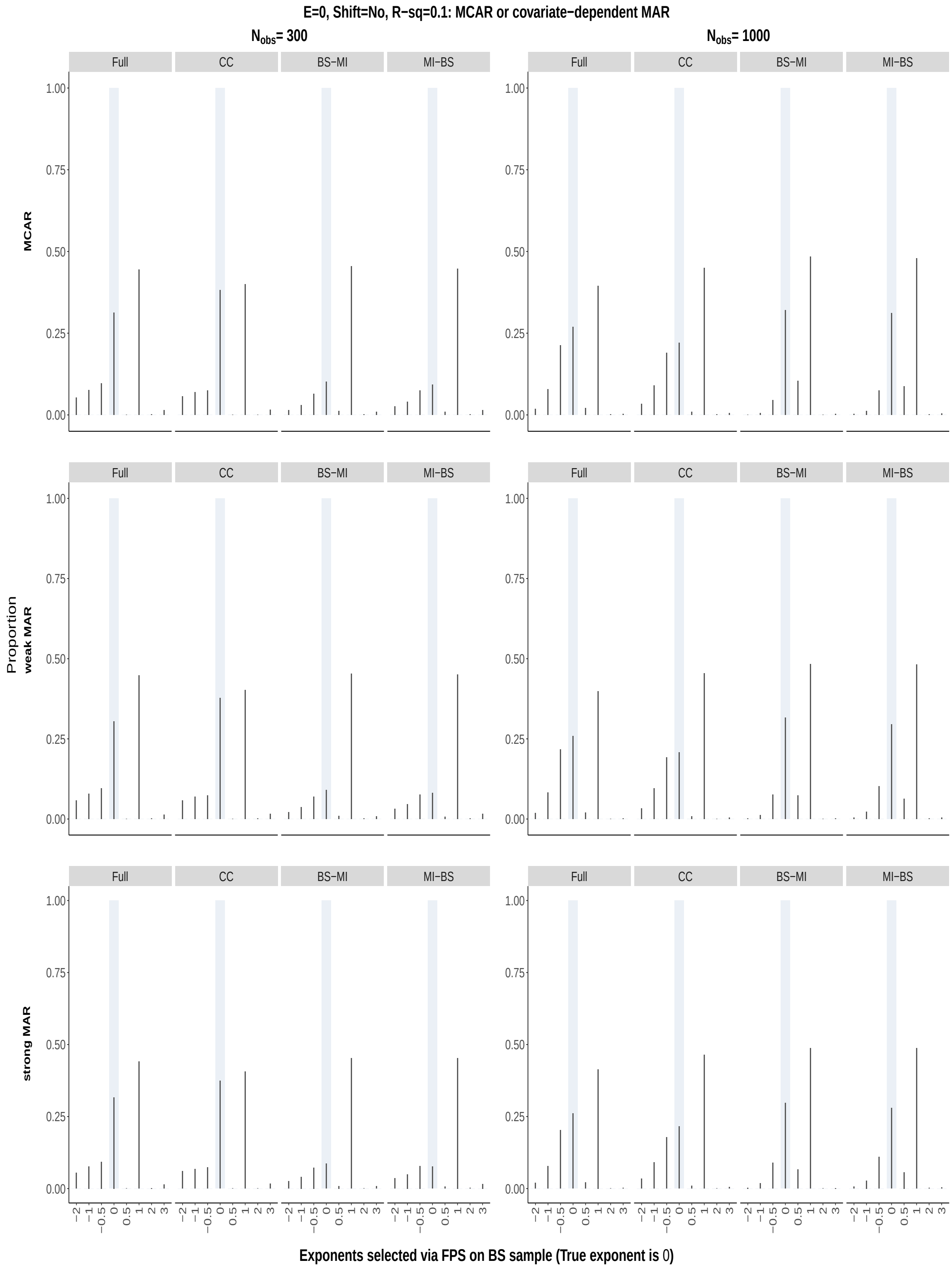

Figure S157: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

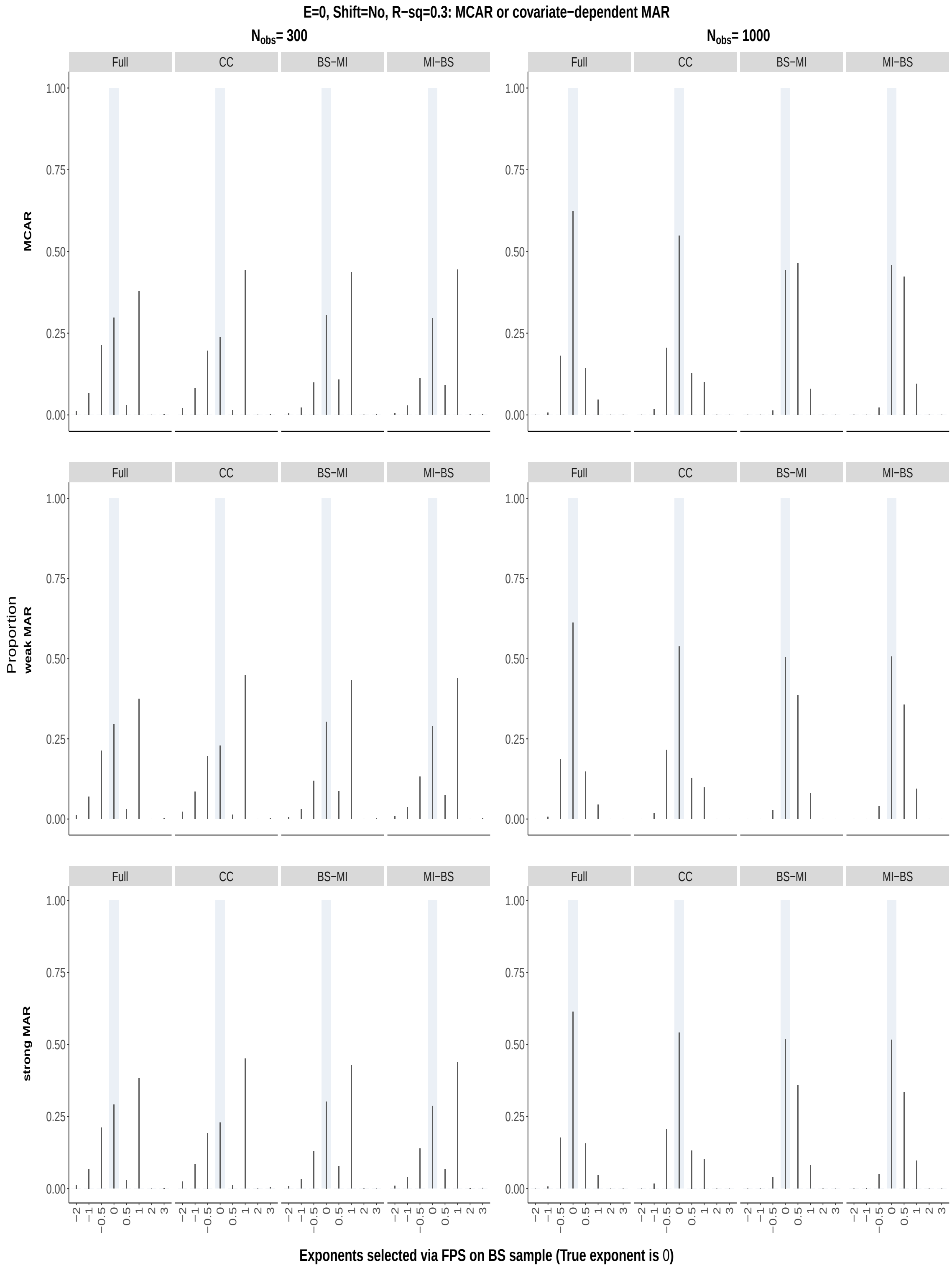

Figure S158: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

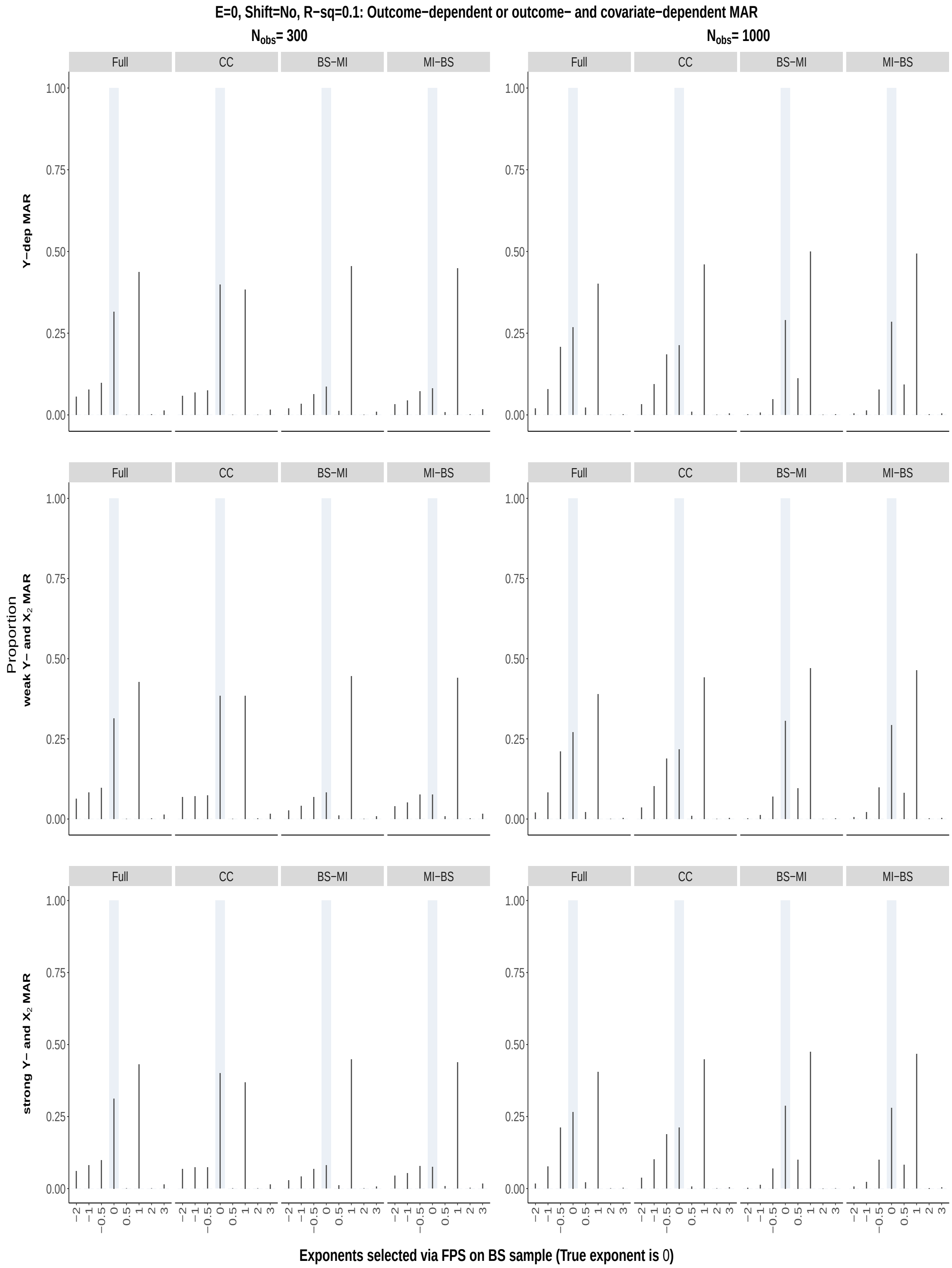

Figure S159: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

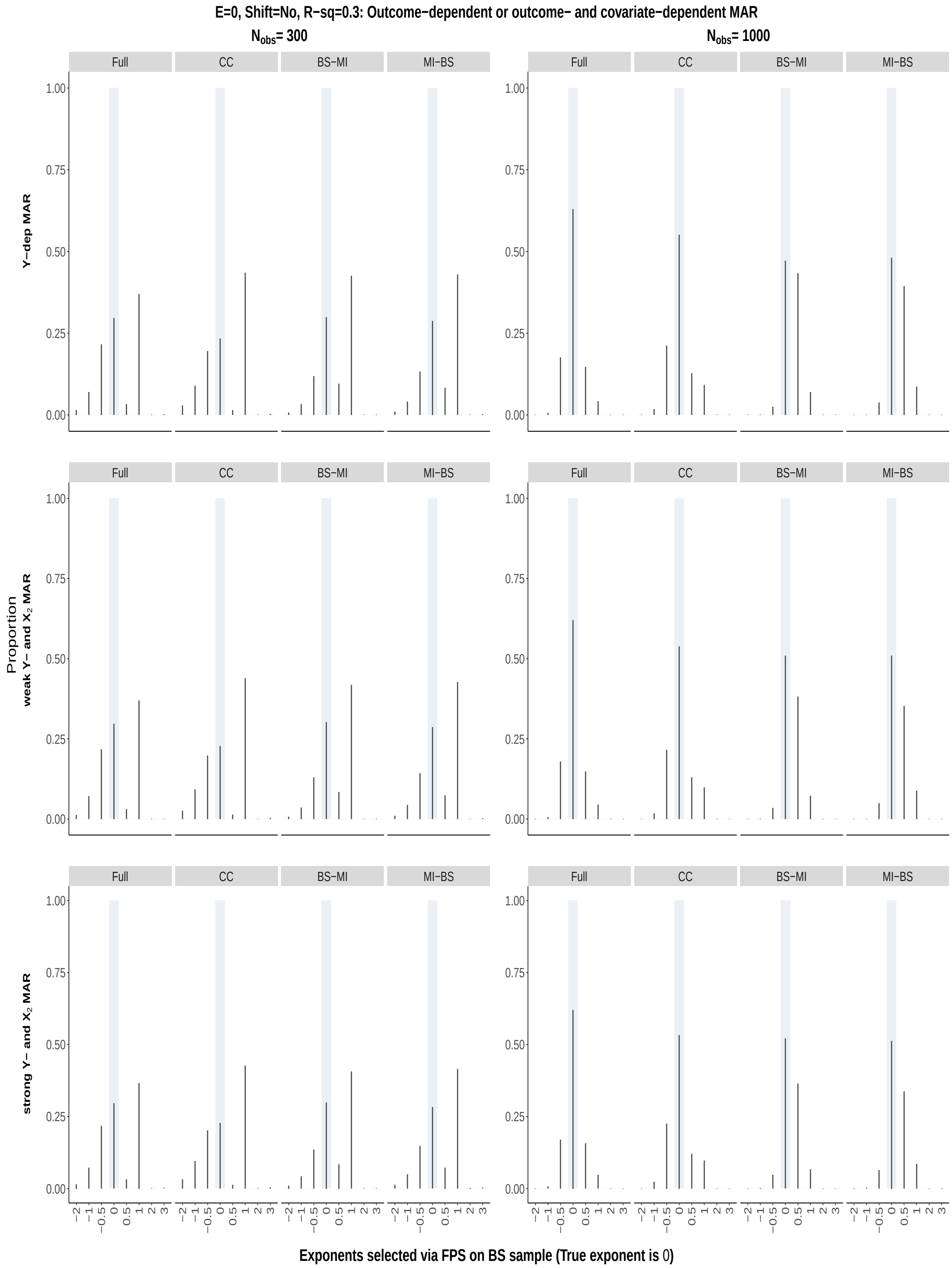

Figure S160: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

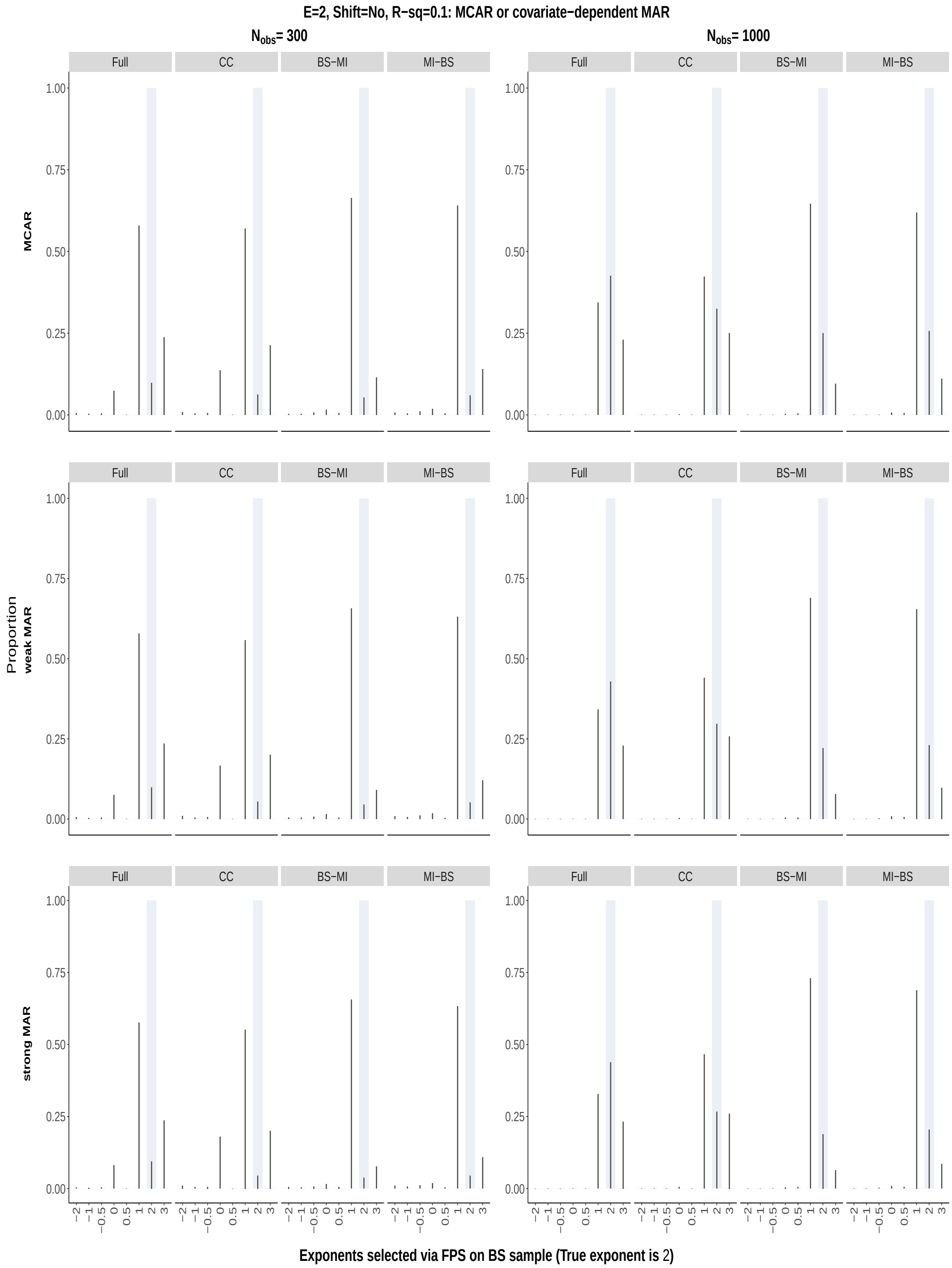

Figure S161: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

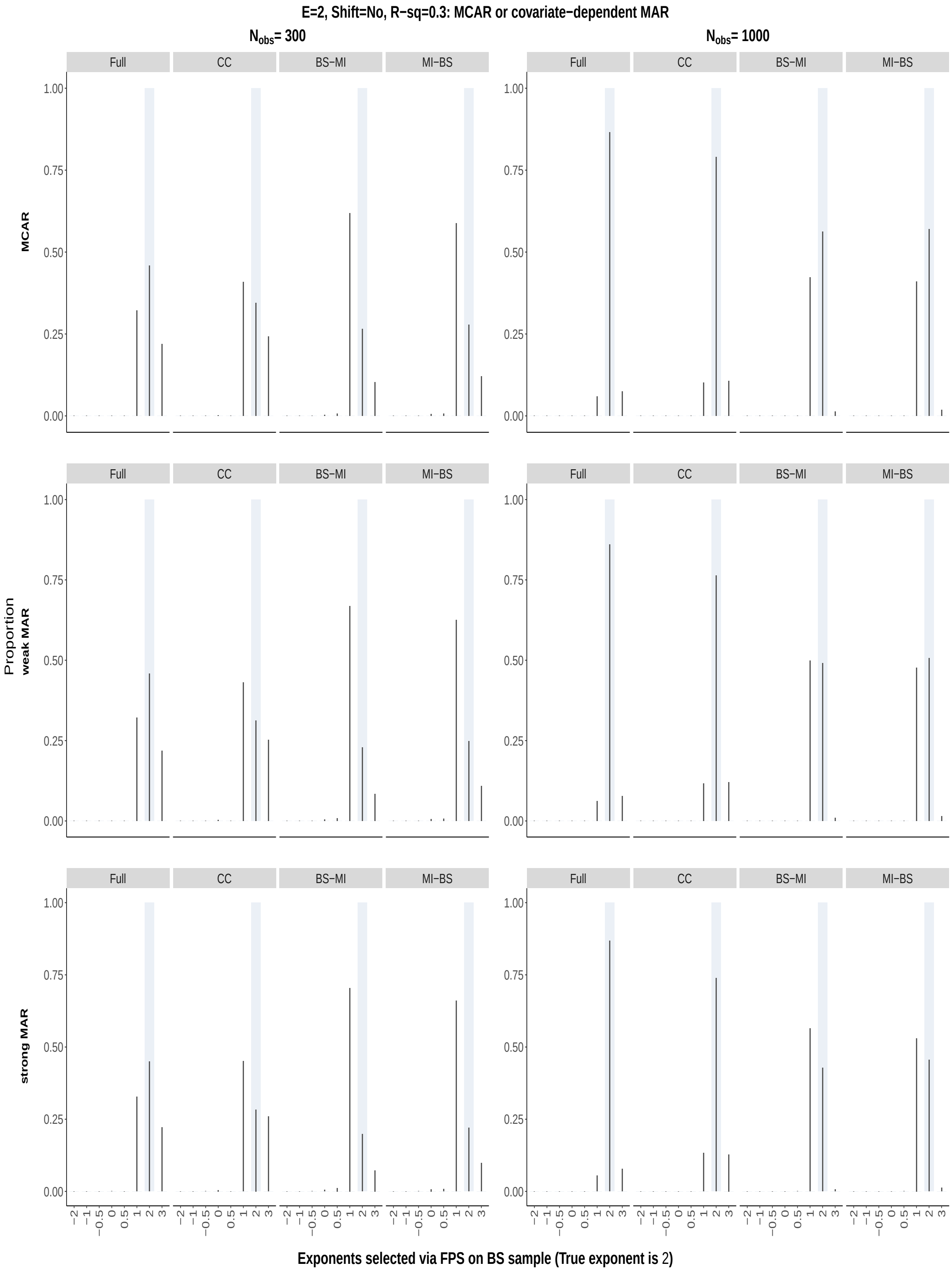

Figure S162: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

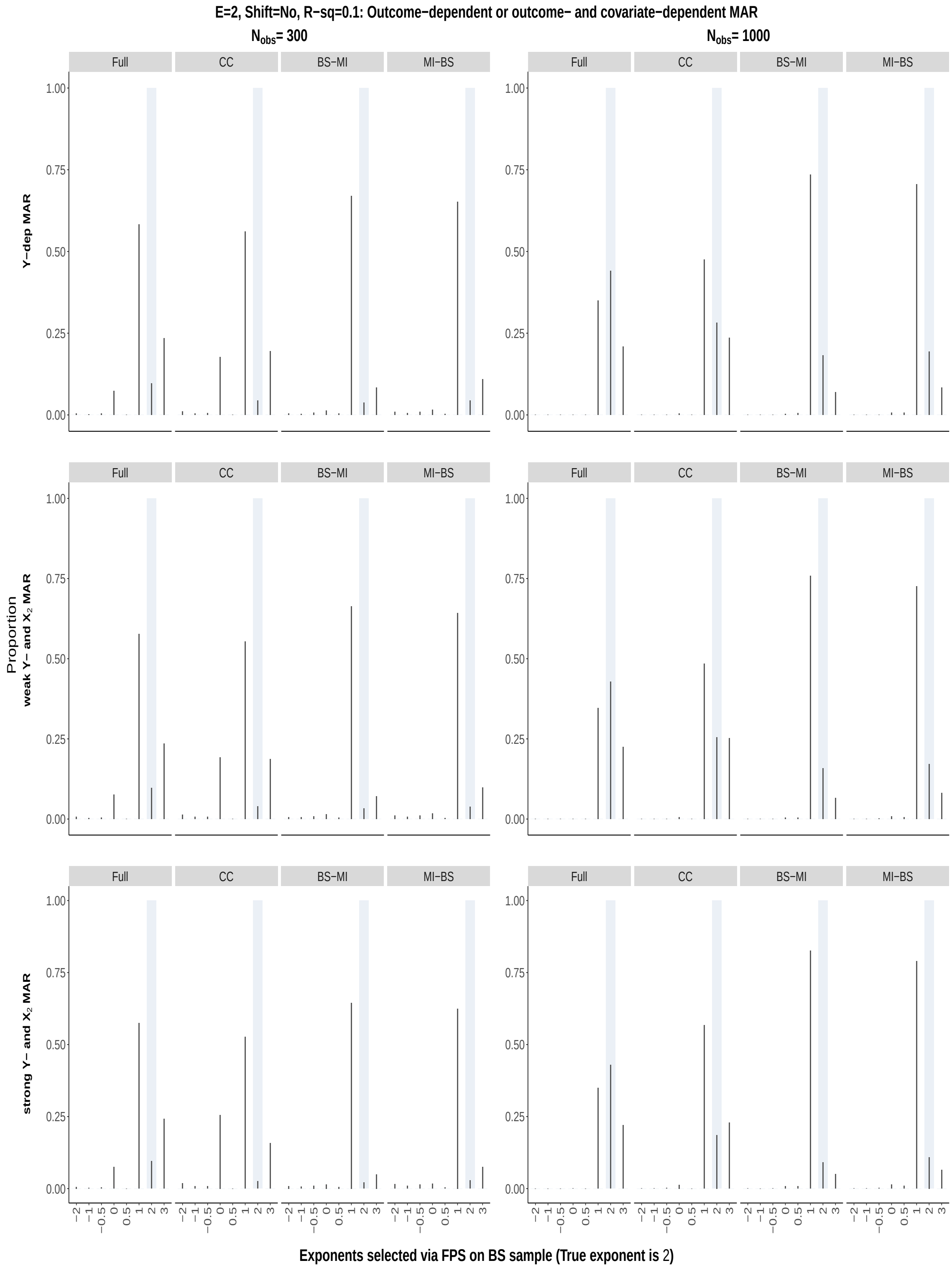

Figure S163: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

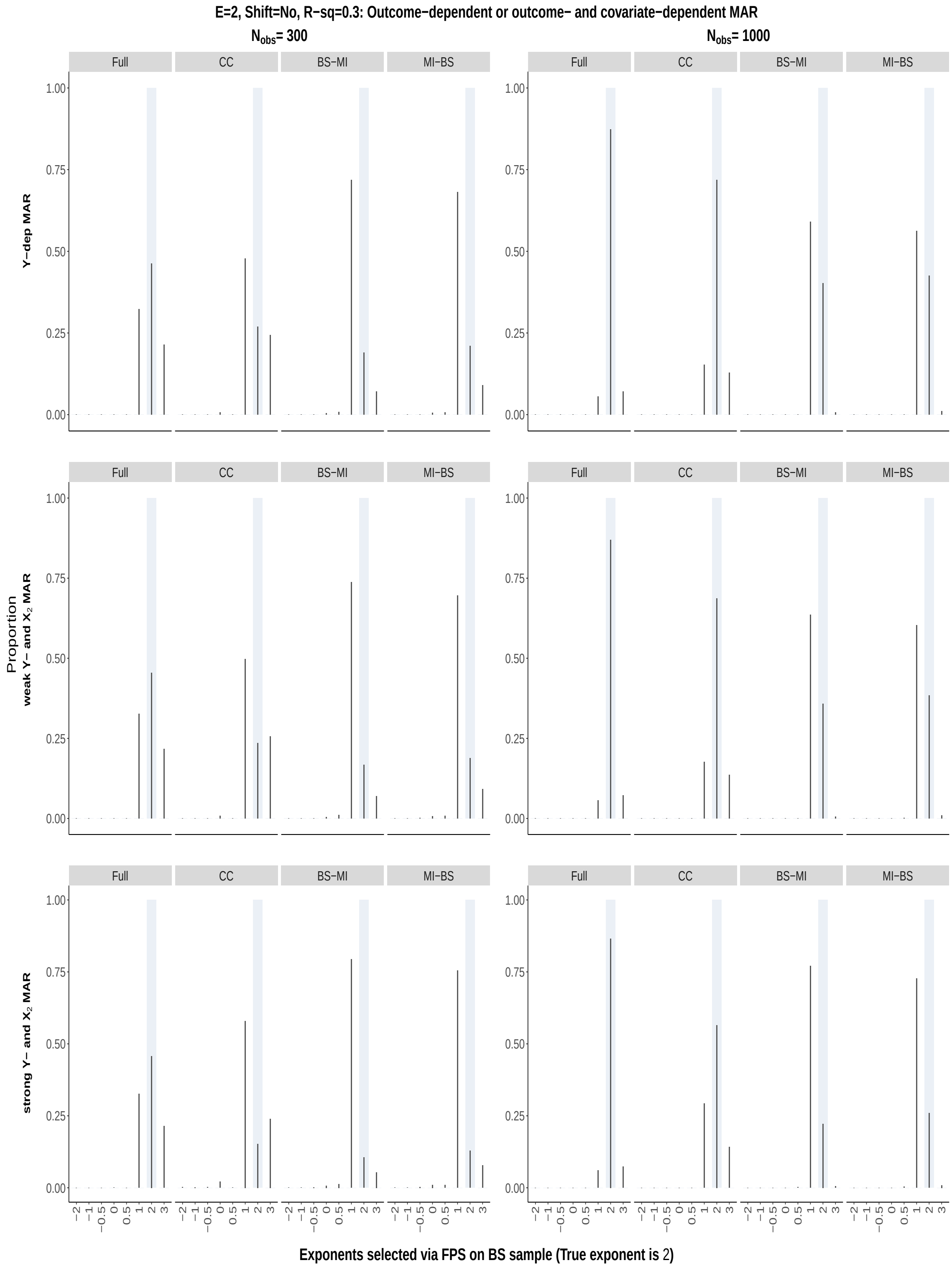

Figure S164: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

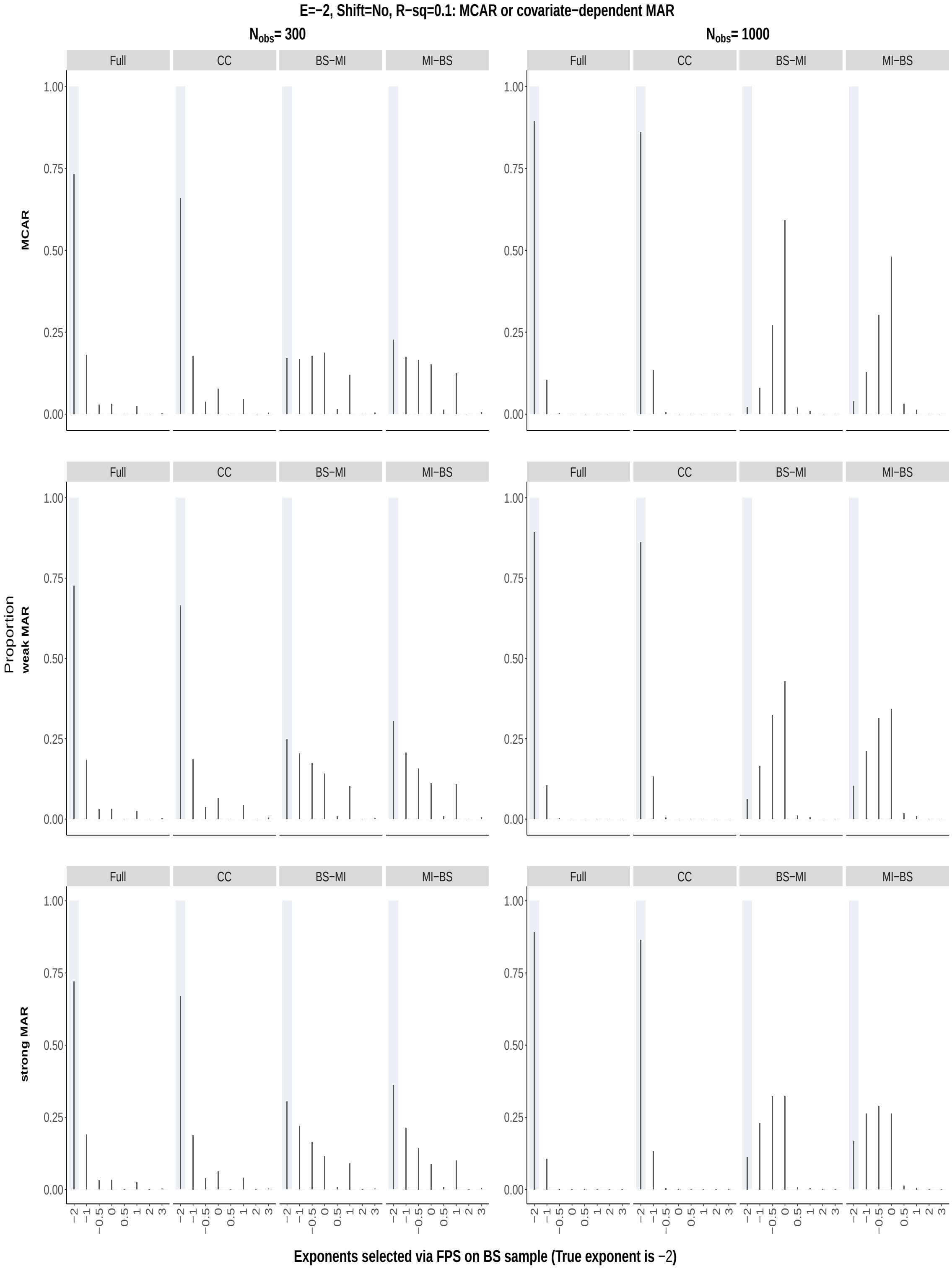

Figure S165: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

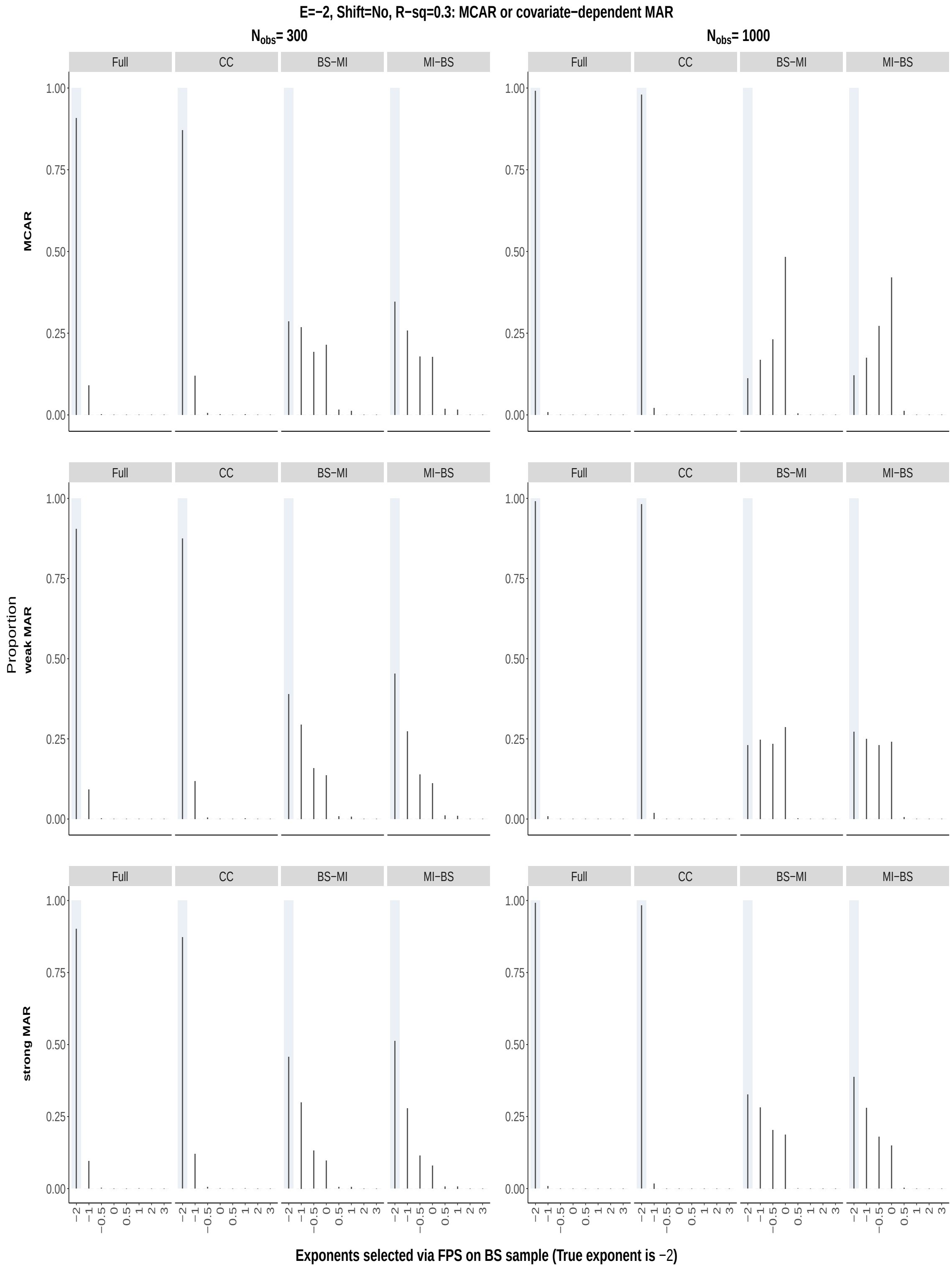

Figure S166: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

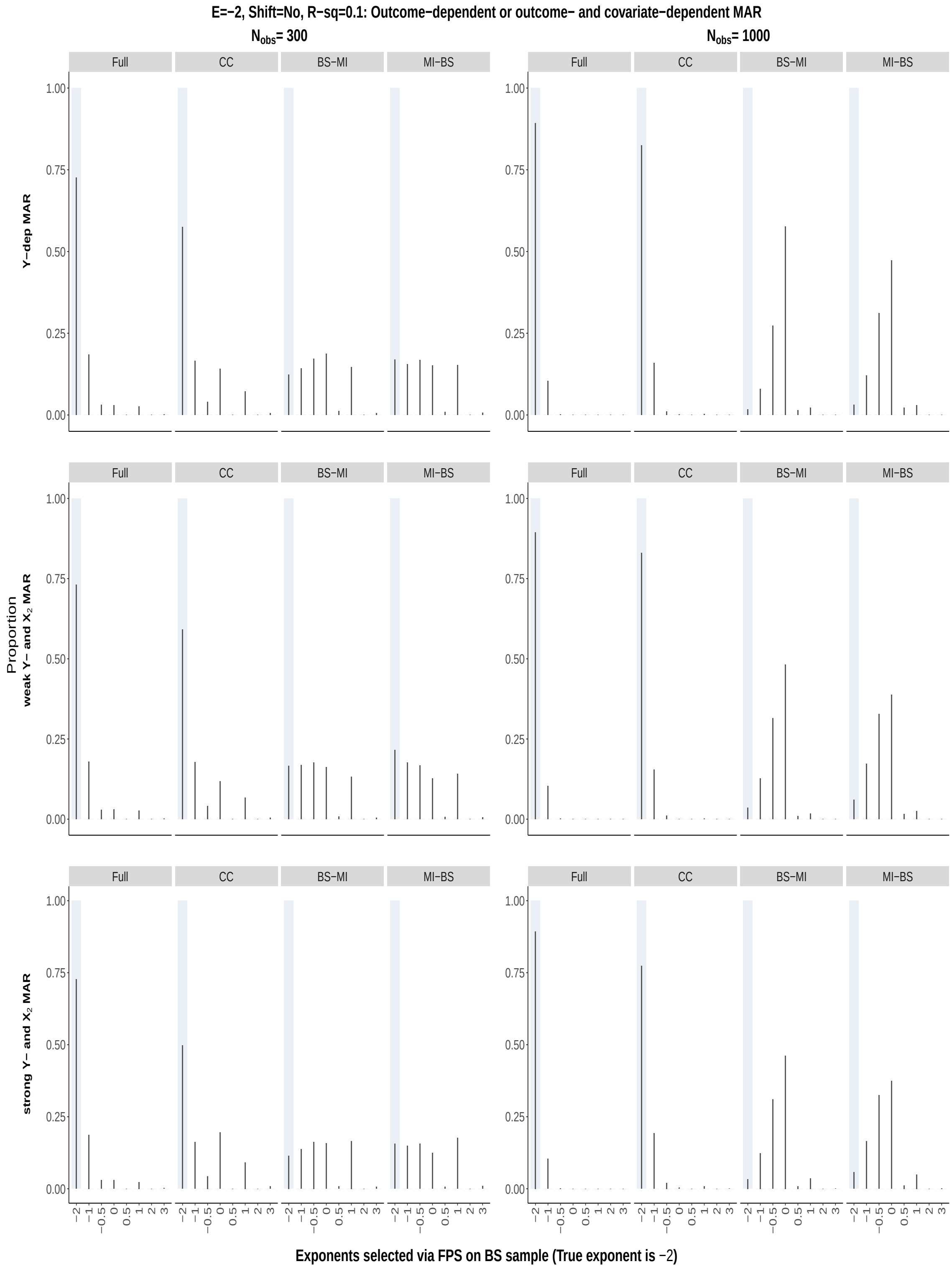

Figure S167: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

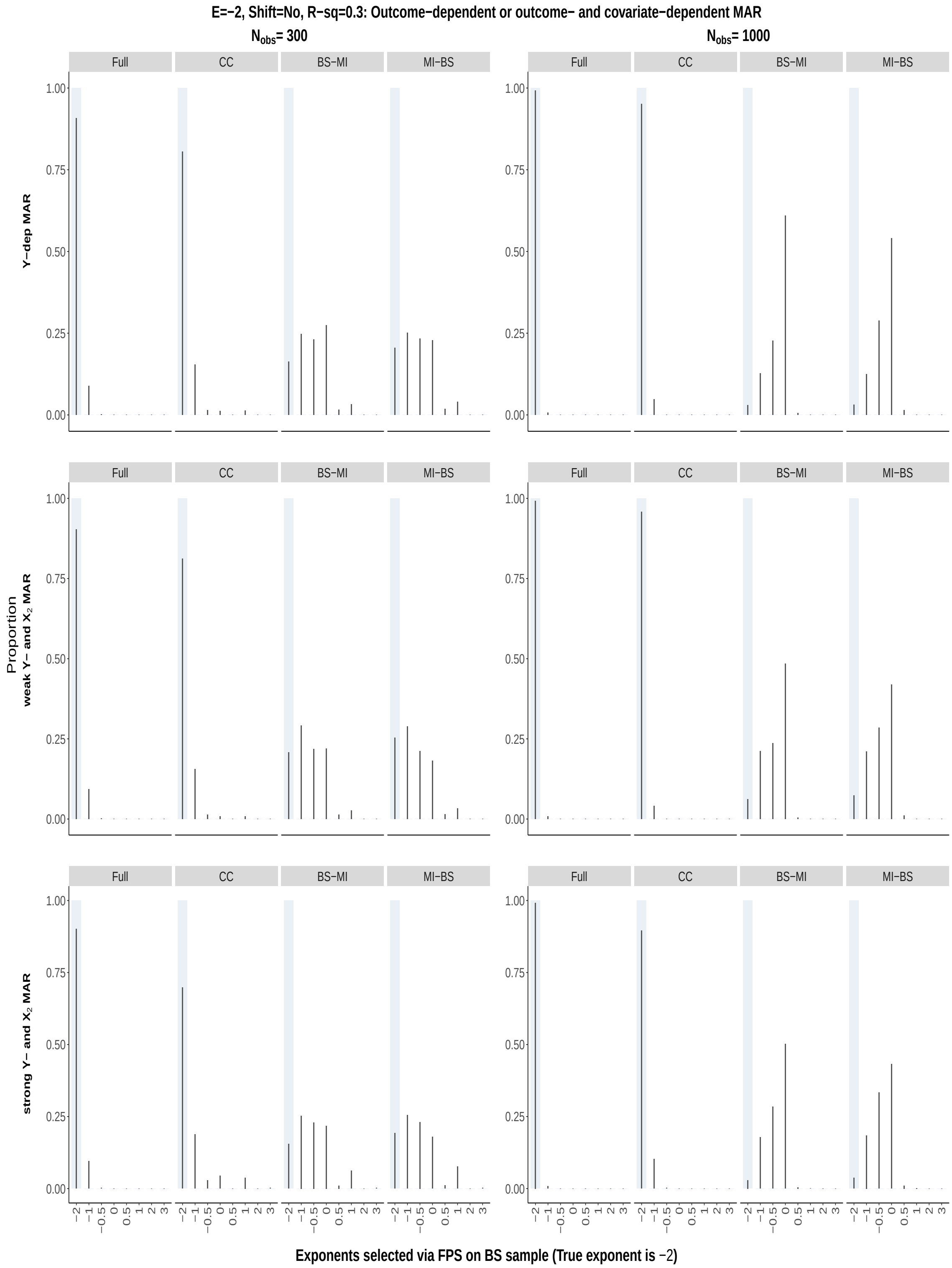

Figure S168: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.11 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

True exponent is 0

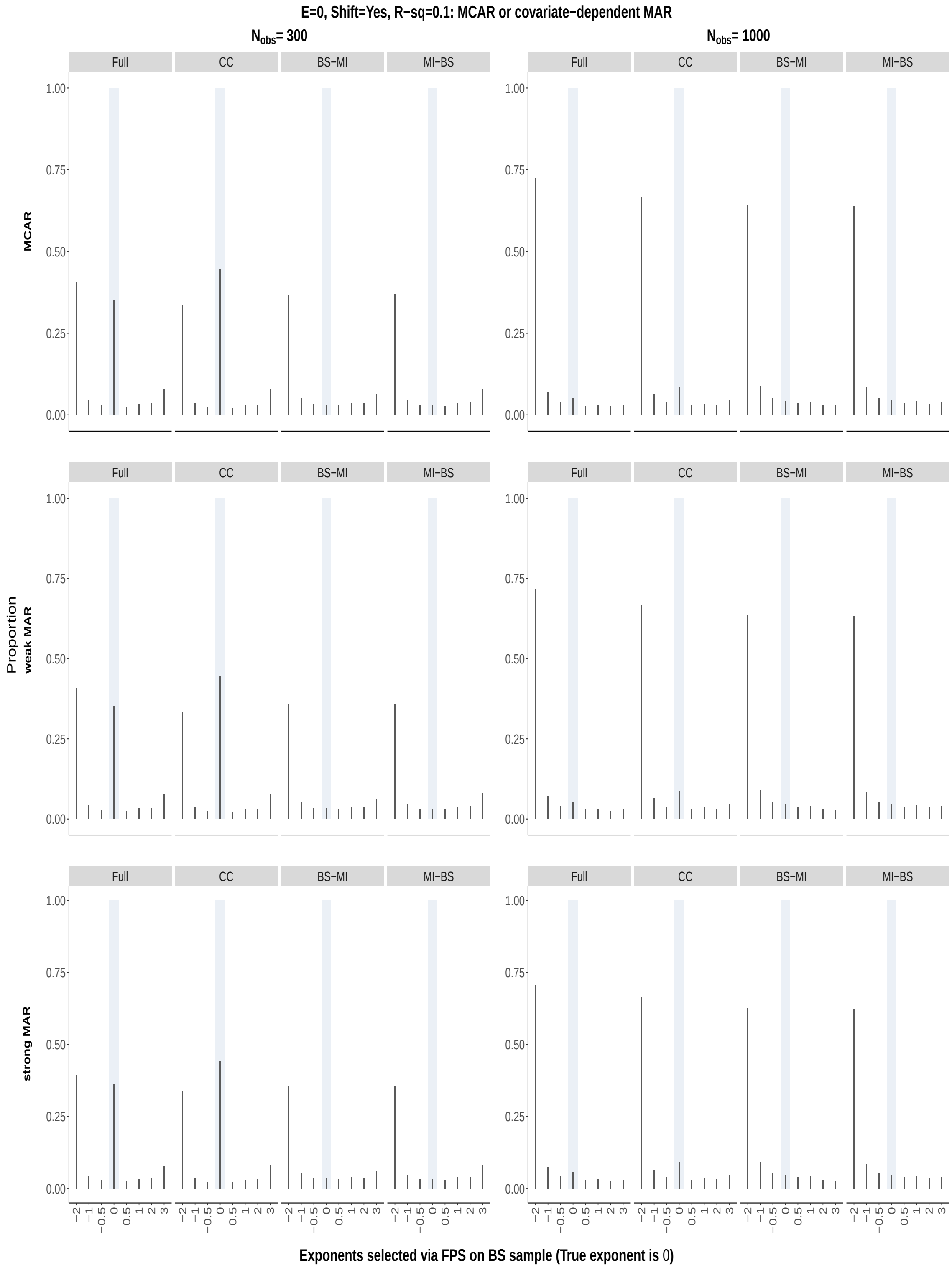

Figure S169: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

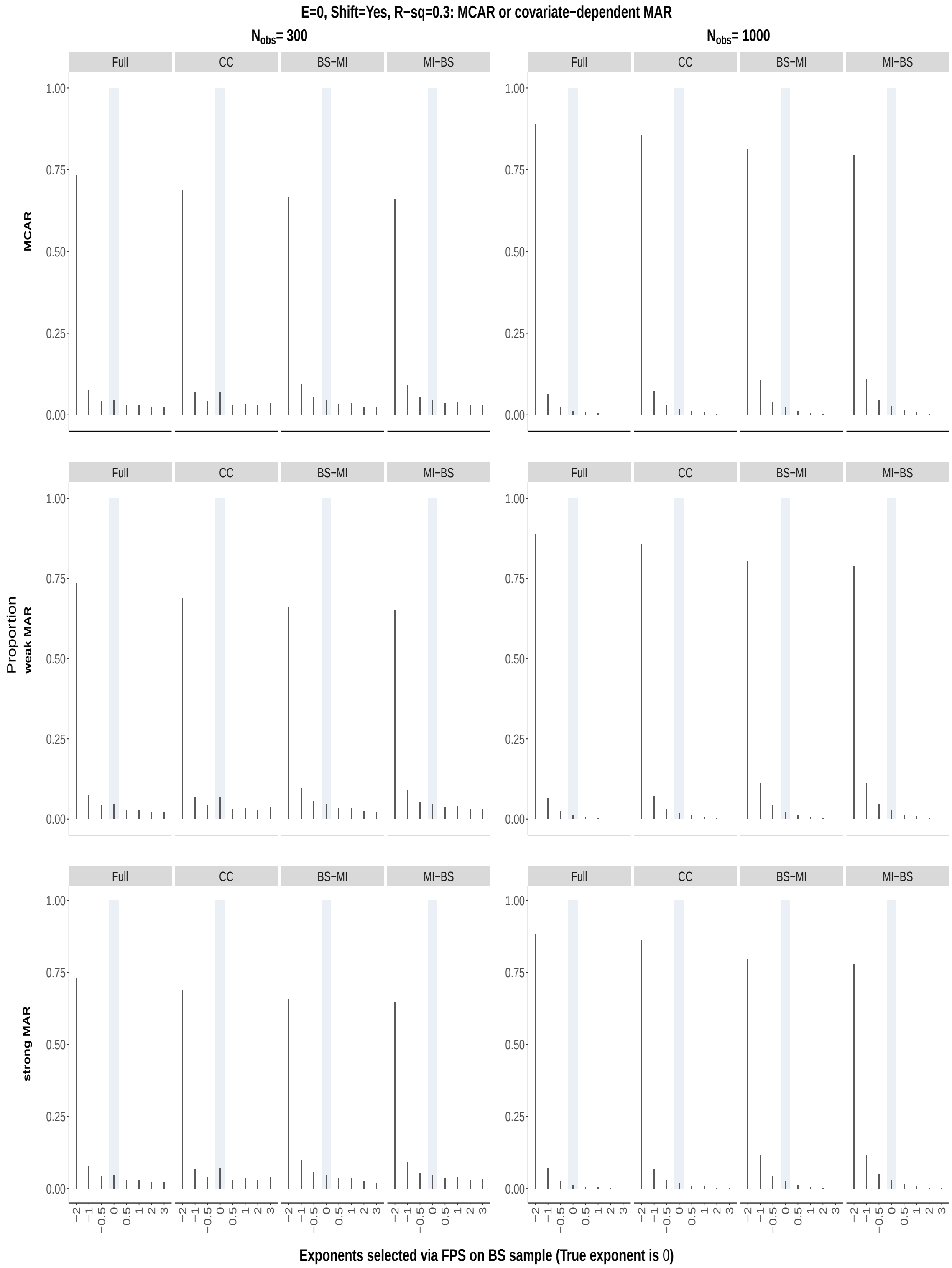

Figure S170: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

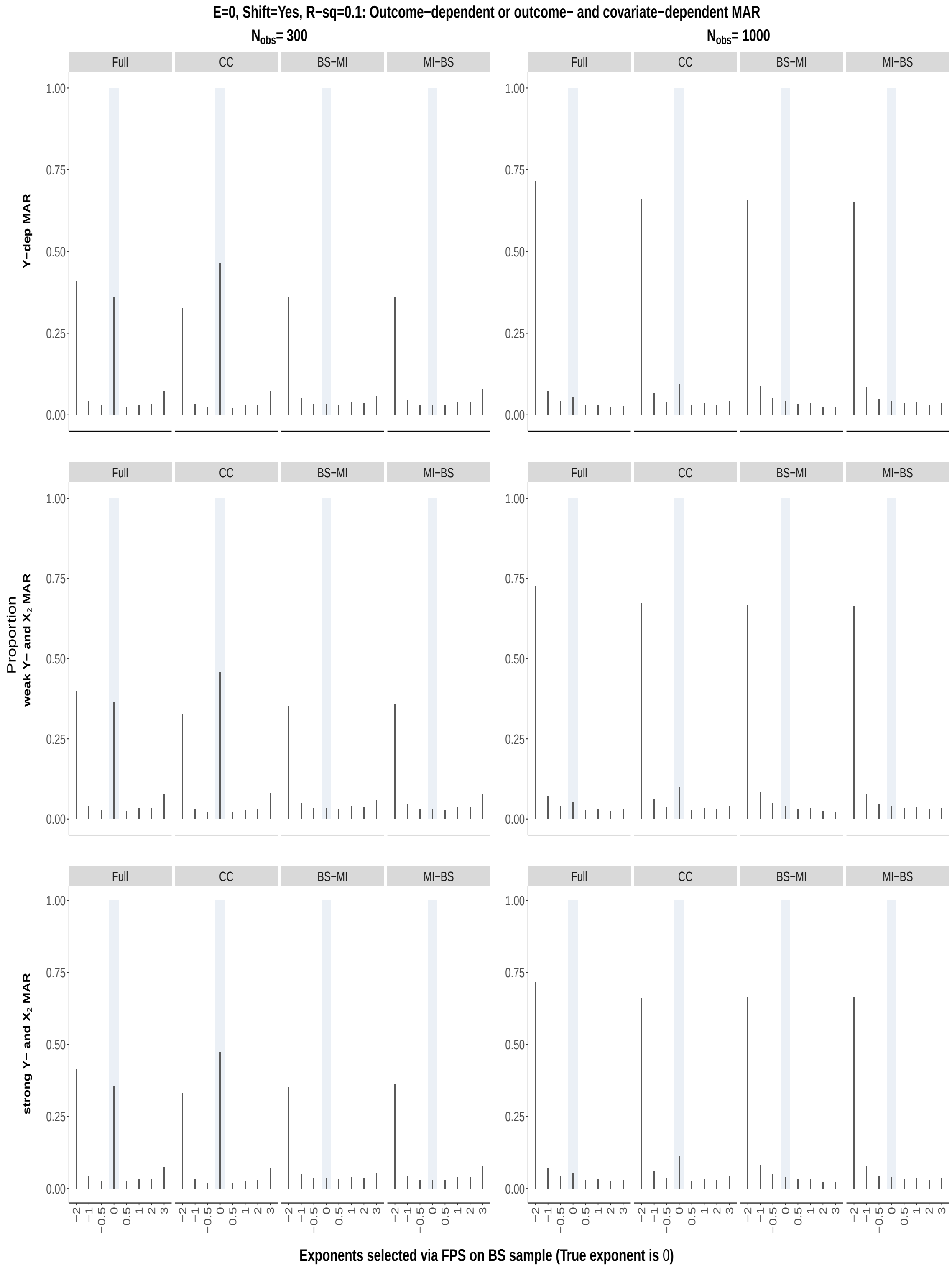

Figure S171: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

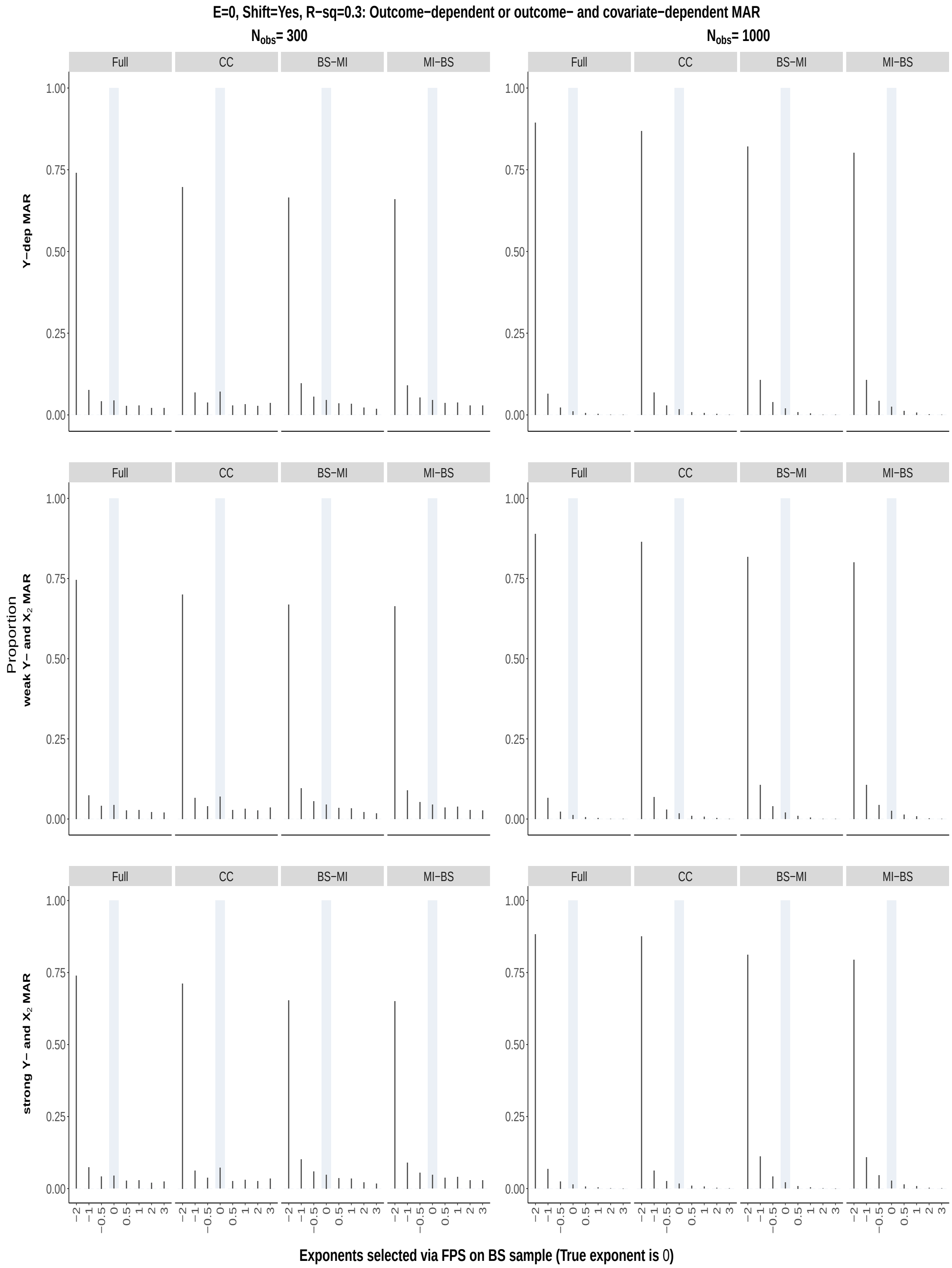

Figure S172: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

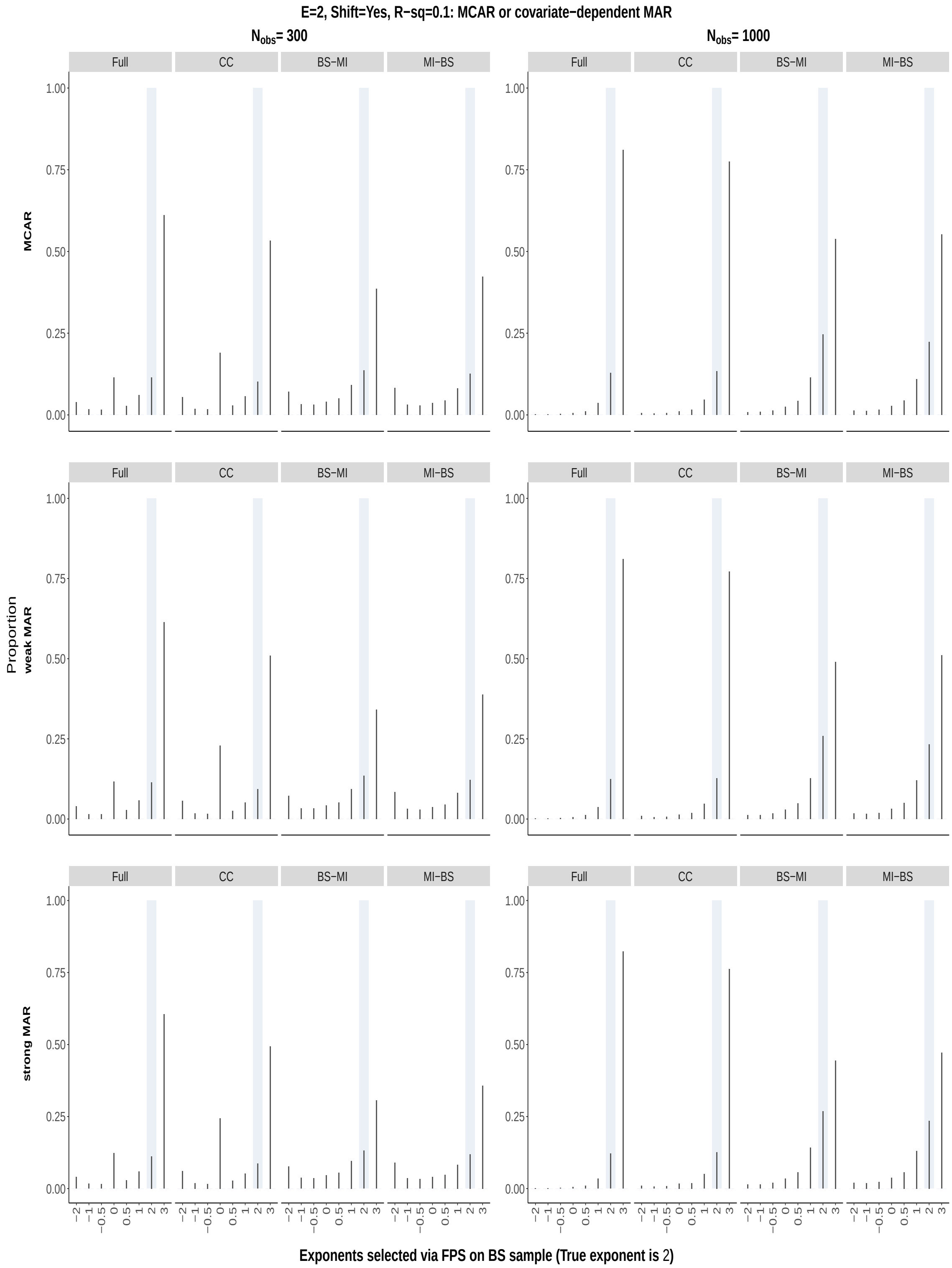

Figure S173: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

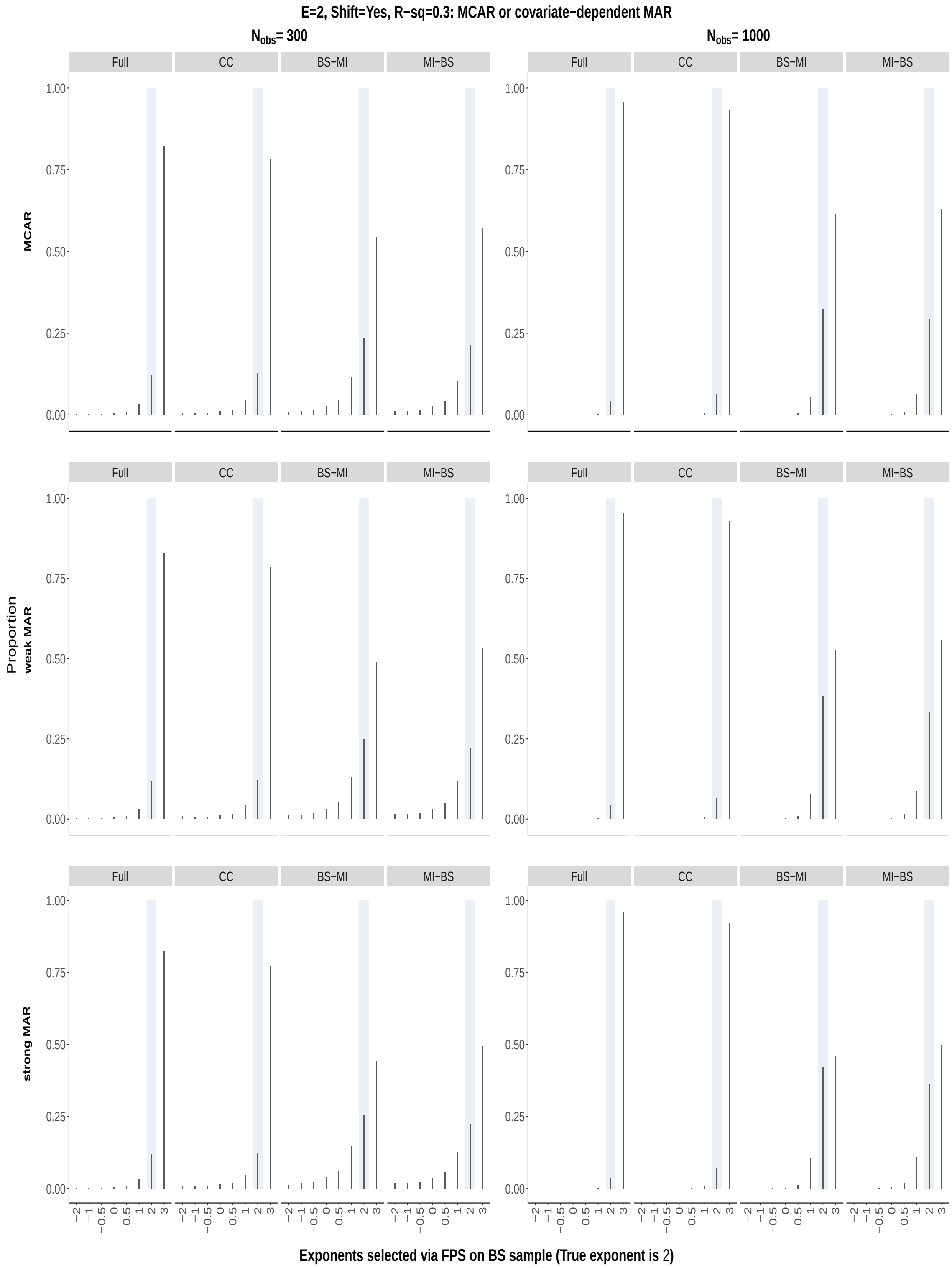

Figure S174: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

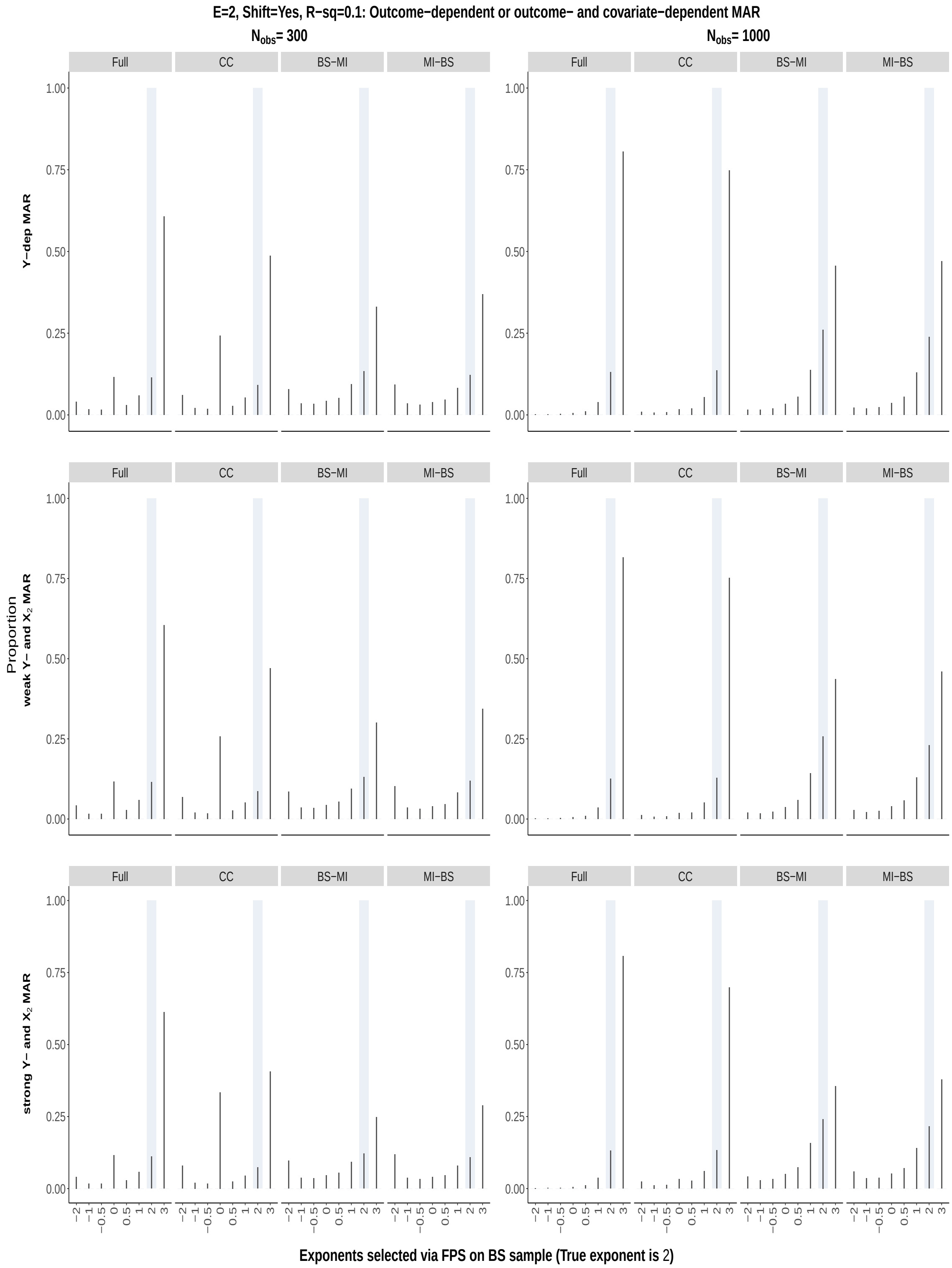

Figure S175: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

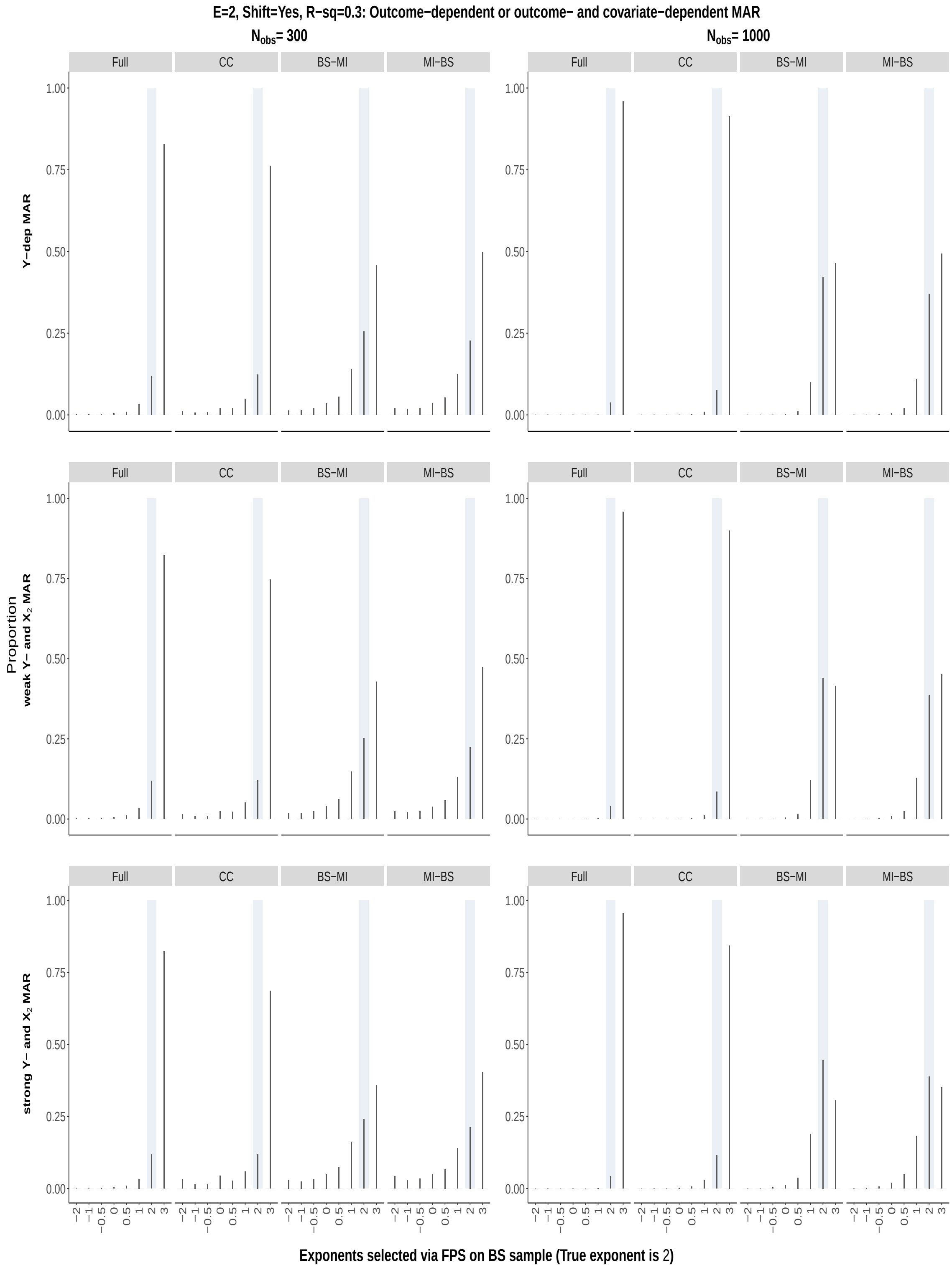

Figure S176: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

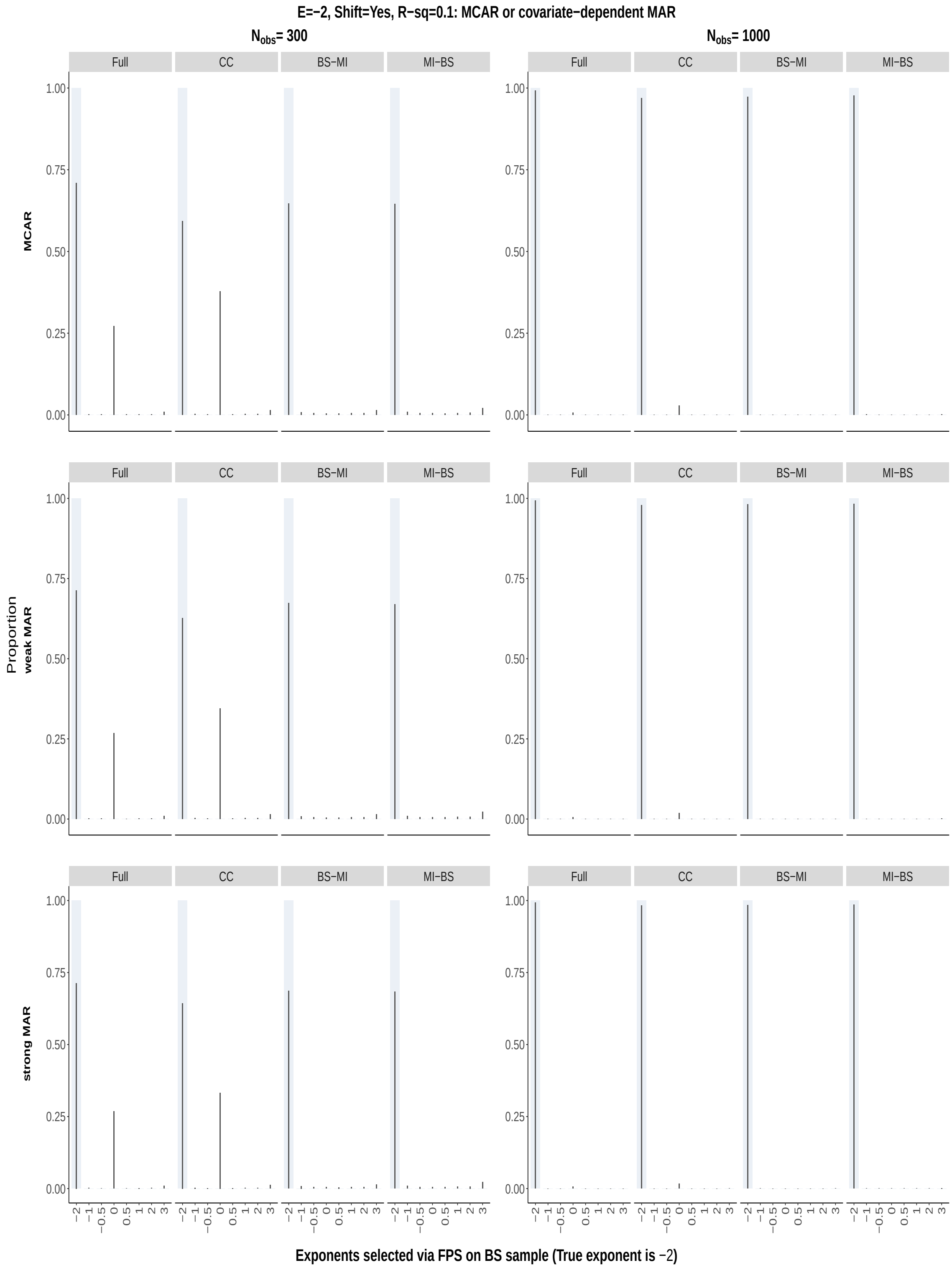

Figure S177: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

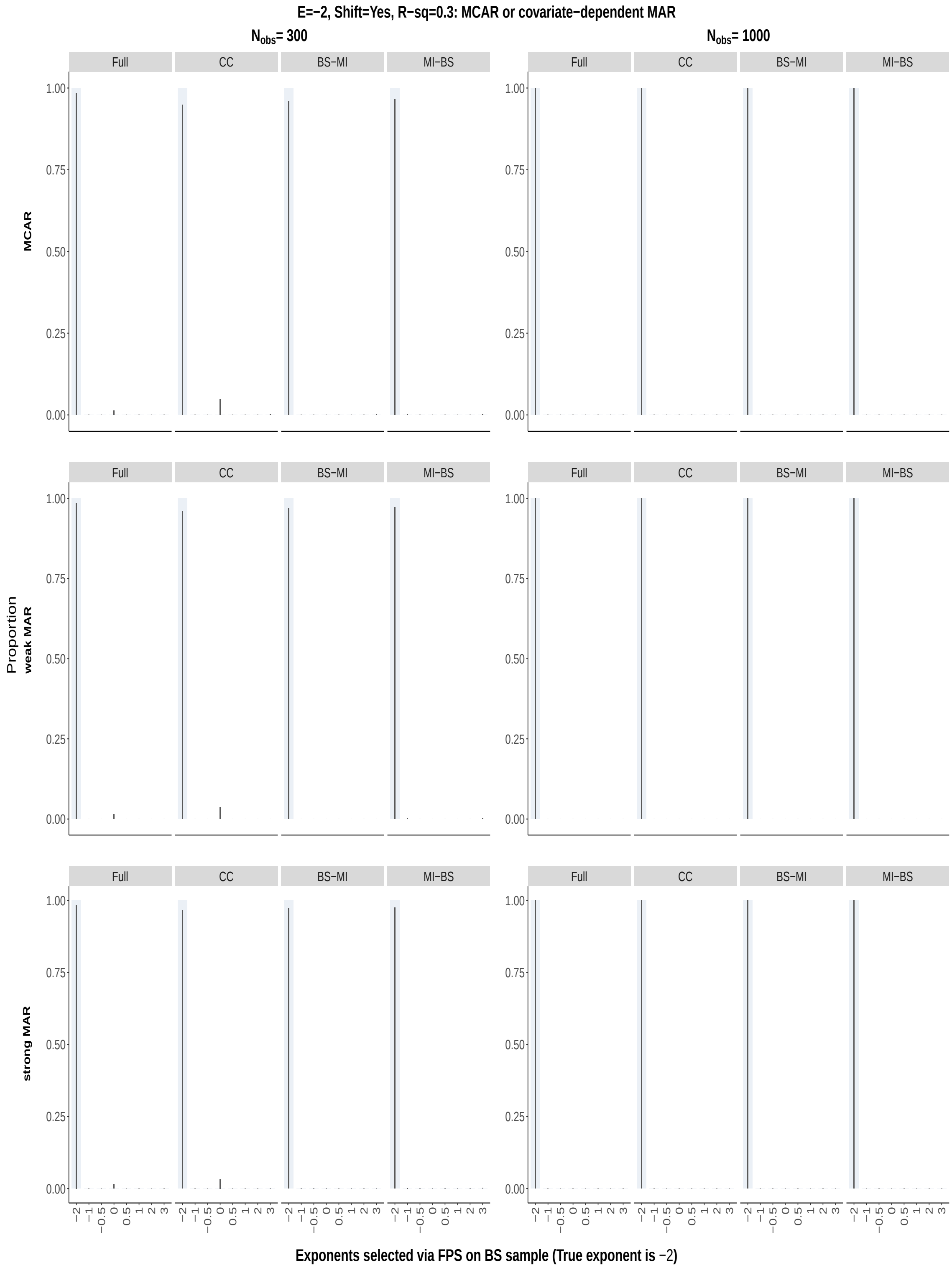

Figure S178: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

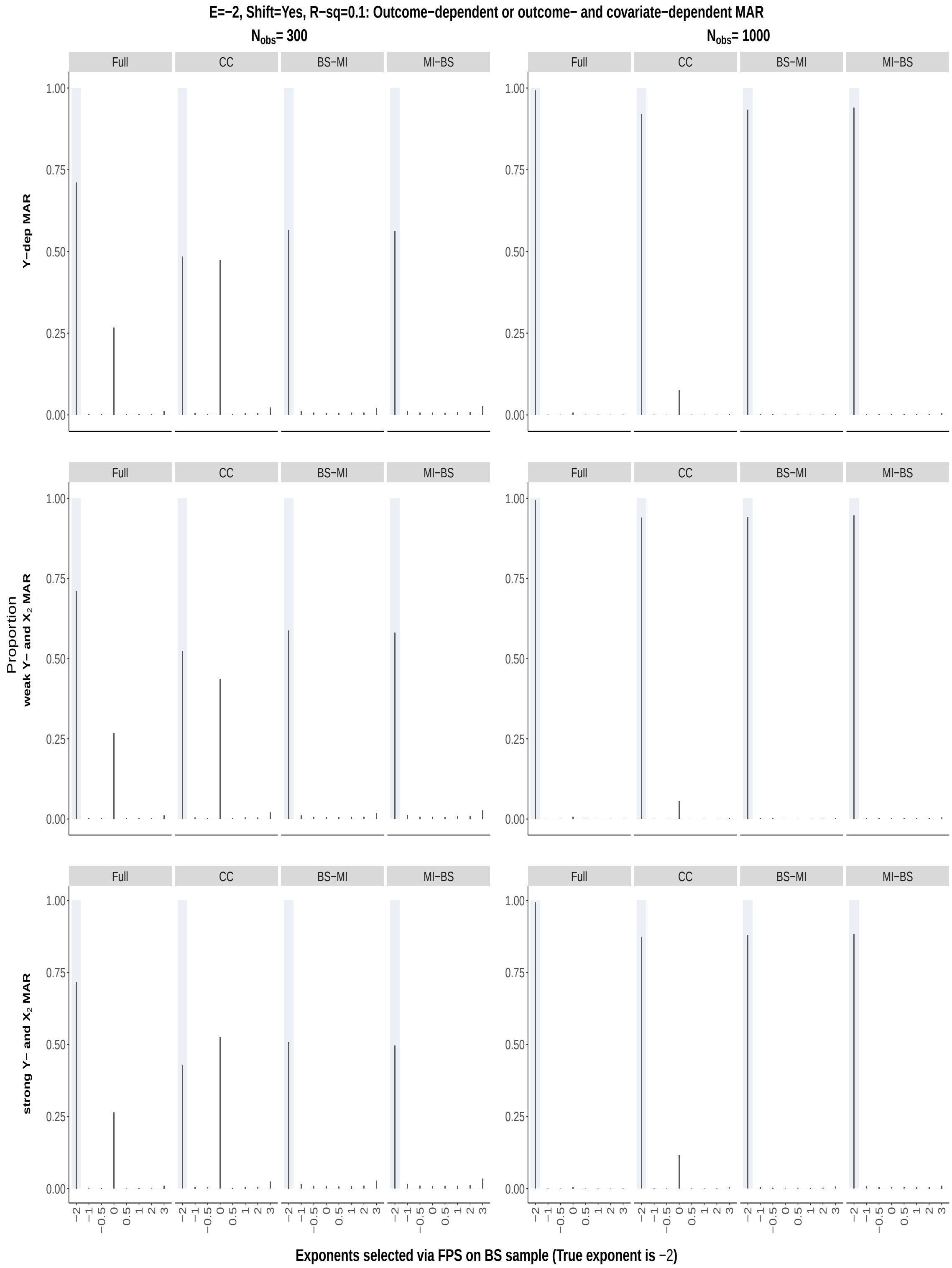

Figure S179: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

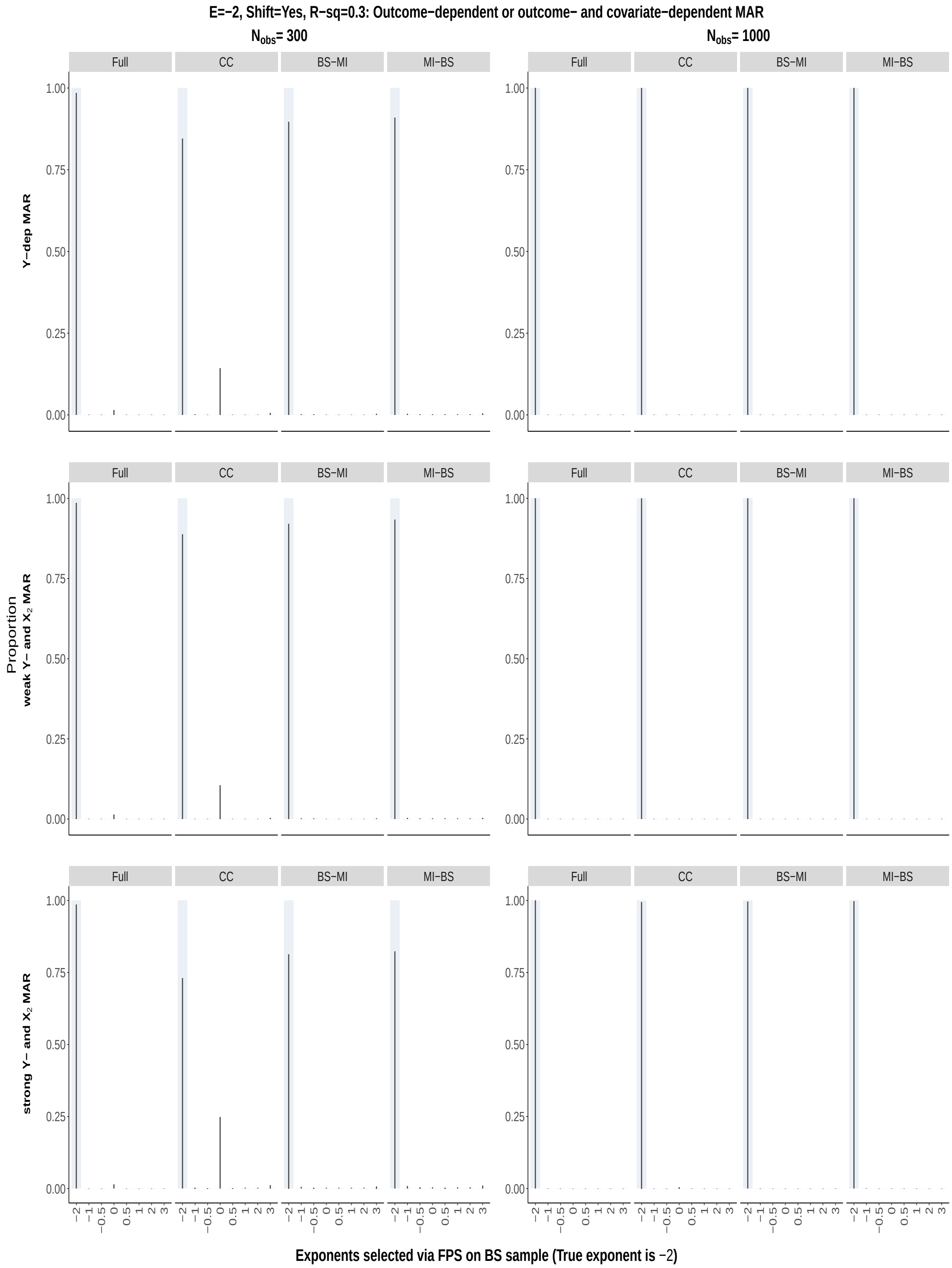

Figure S180: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.12 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

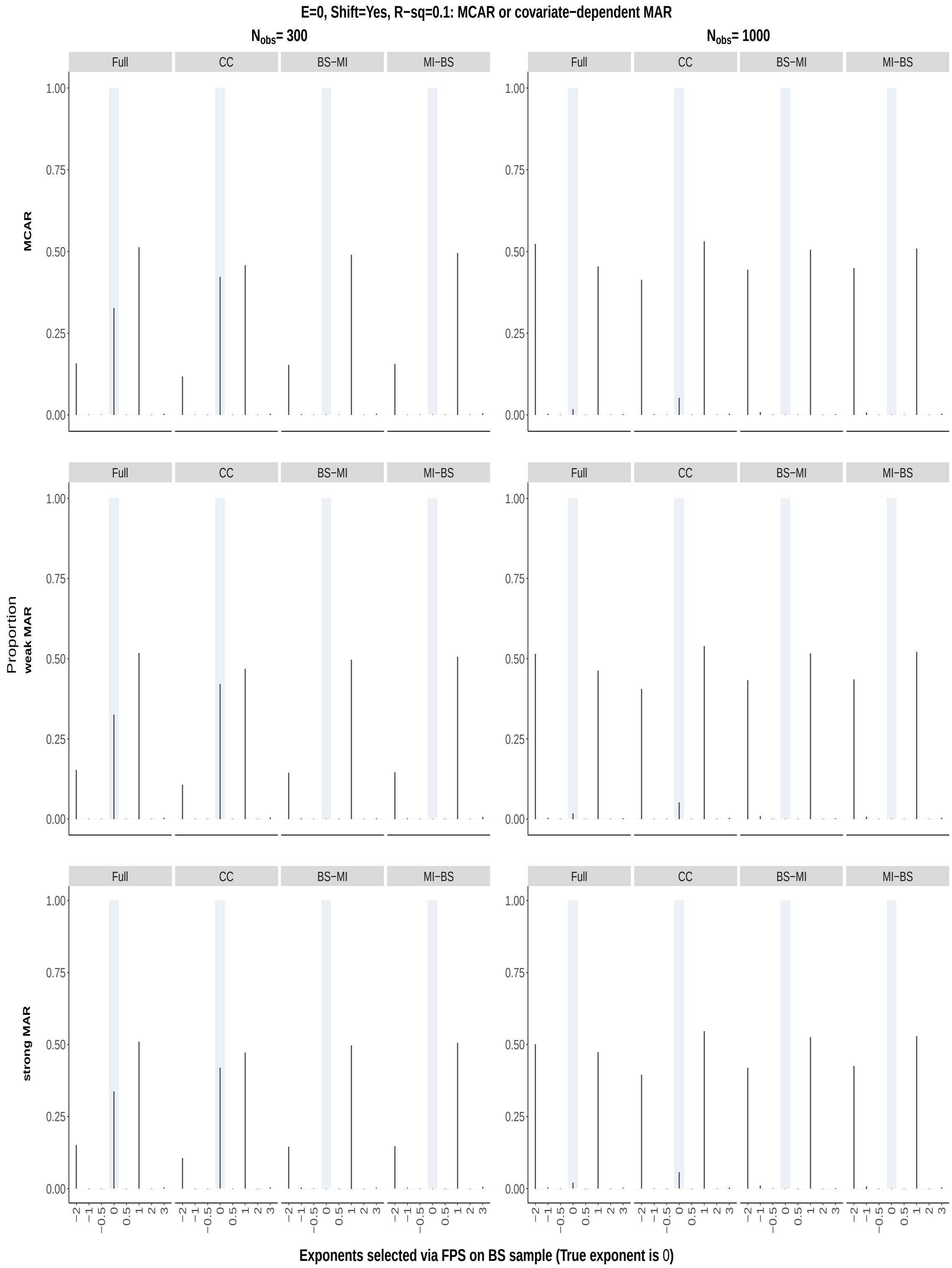

Figure S181: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

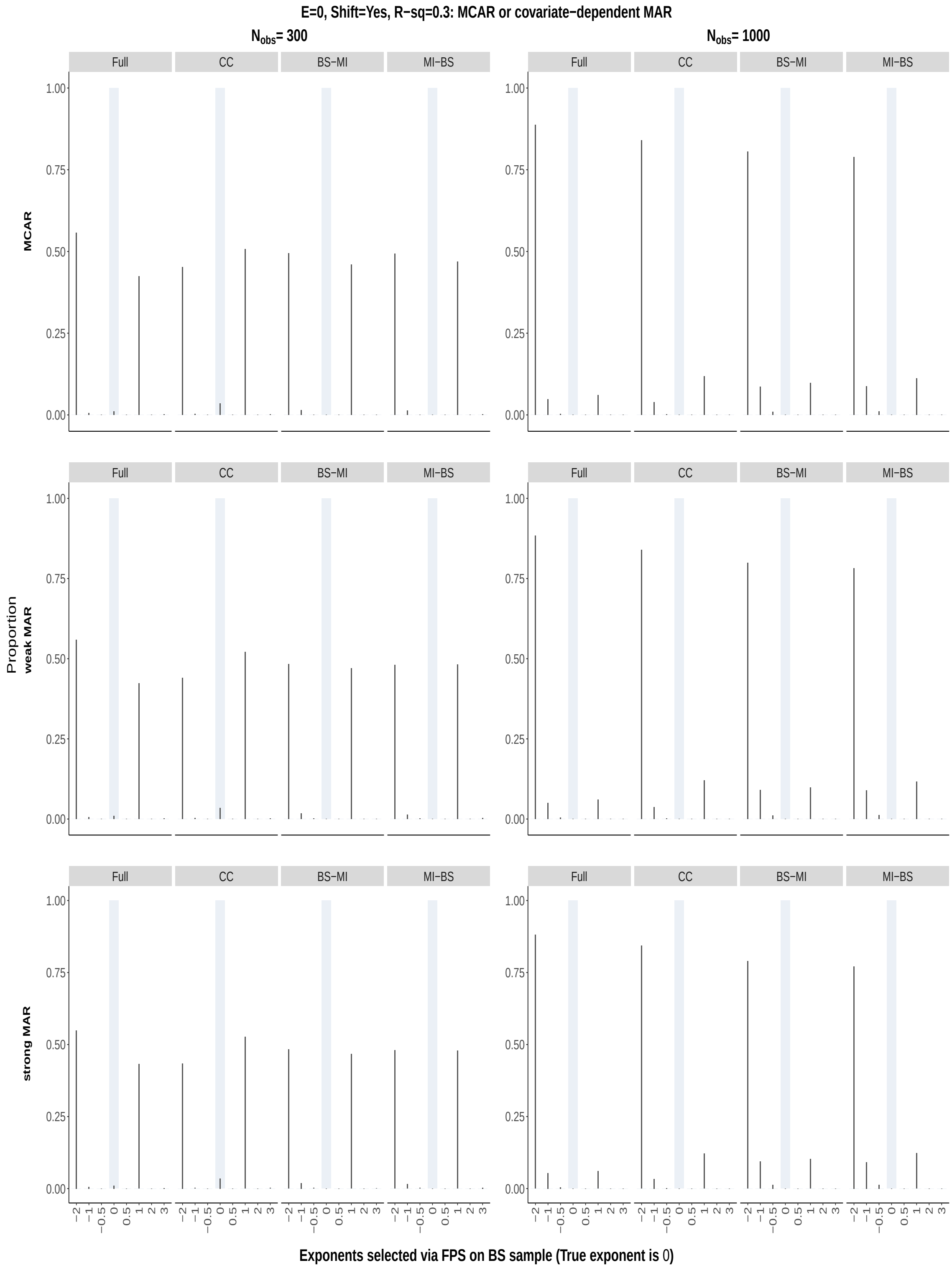

Figure S182: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

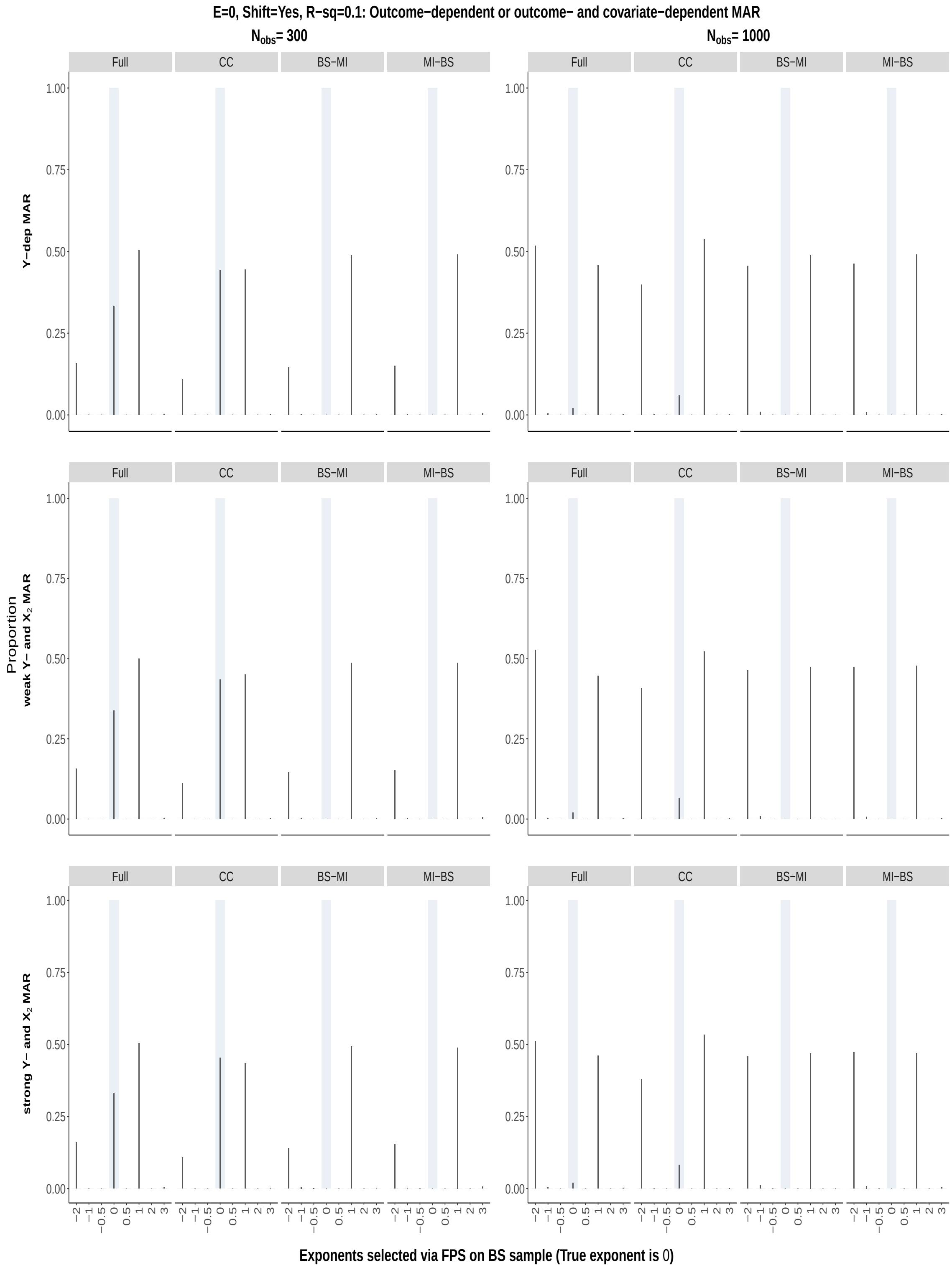

Figure S183: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

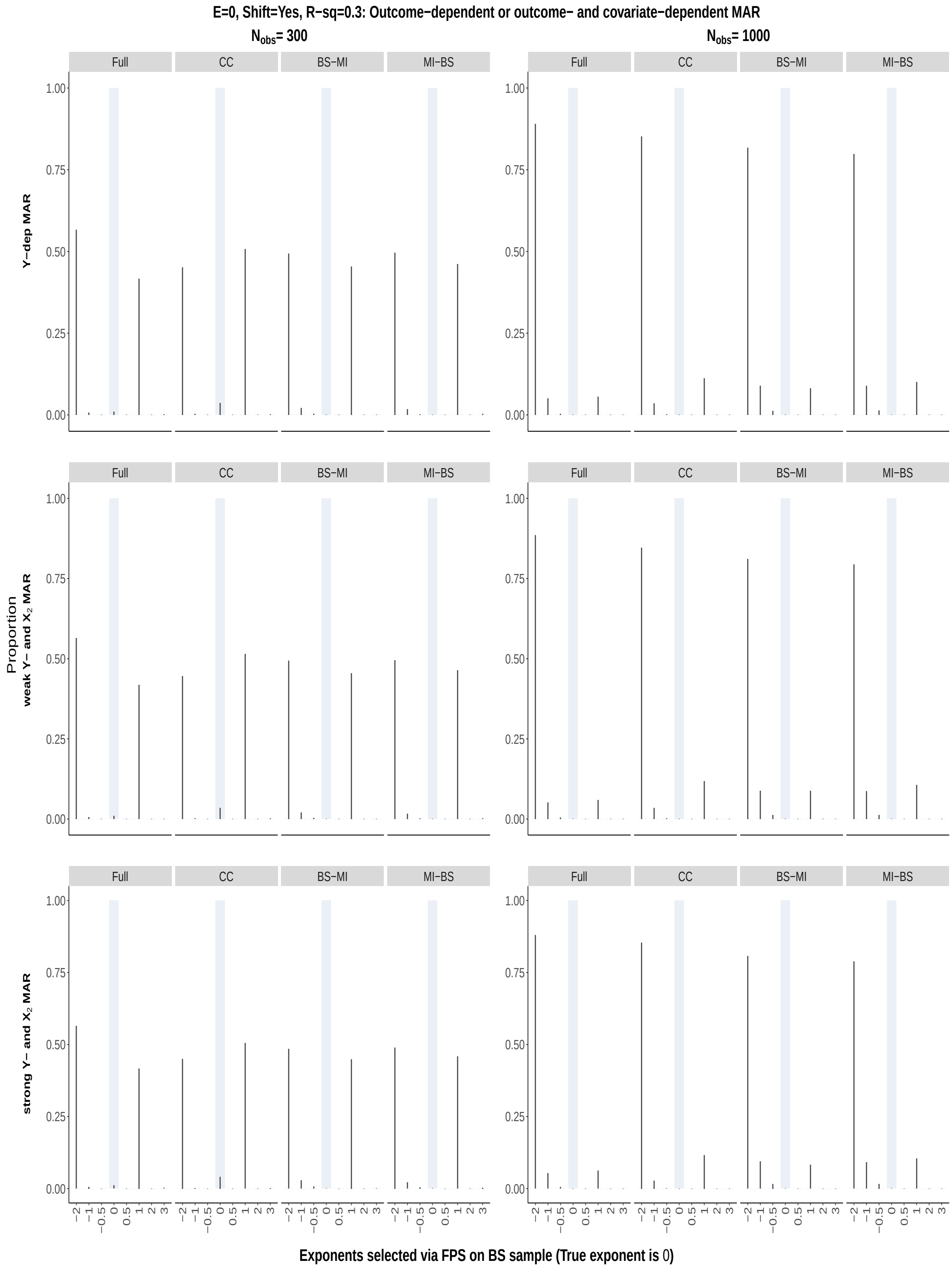

Figure S184: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

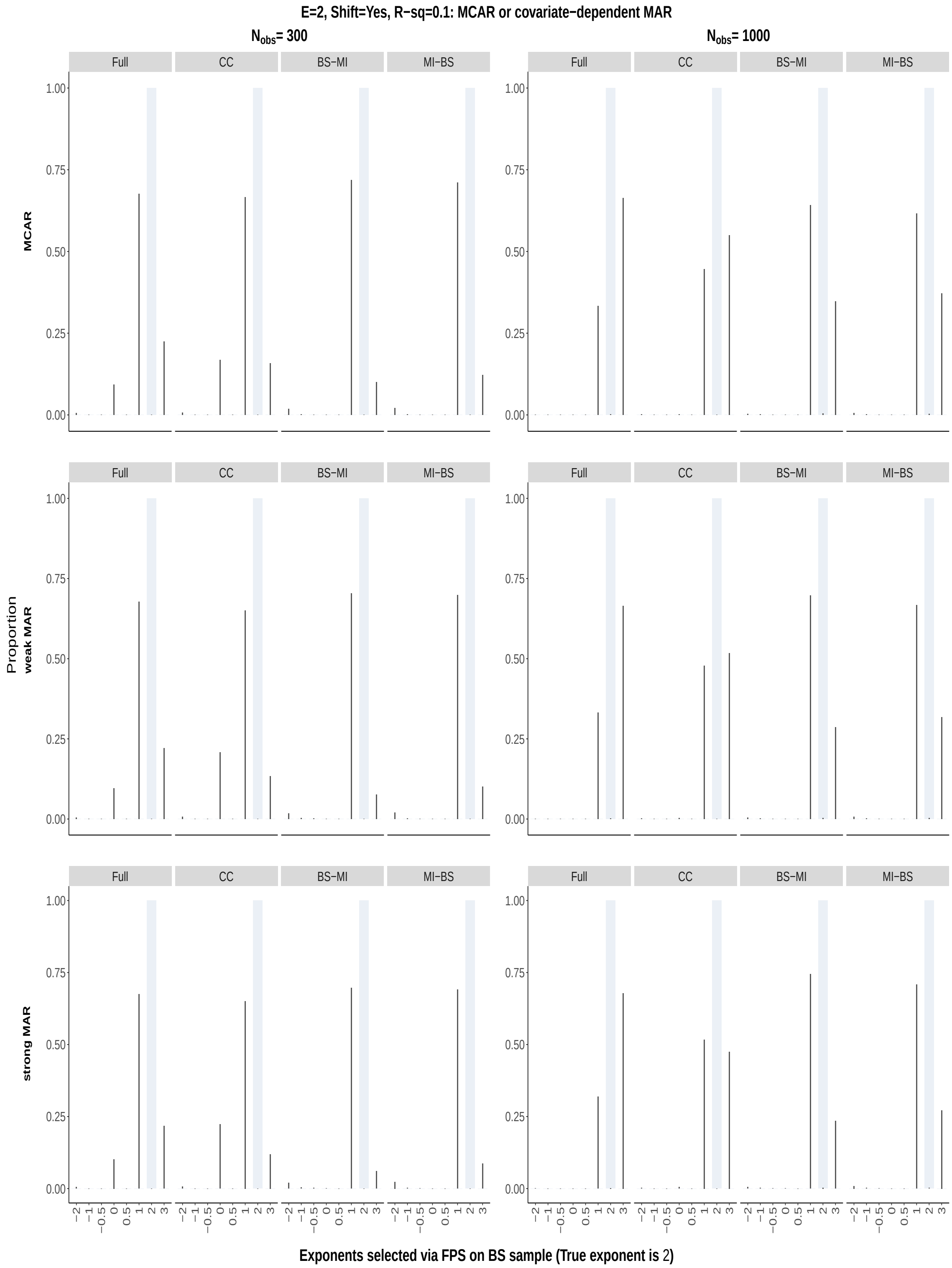

Figure S185: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

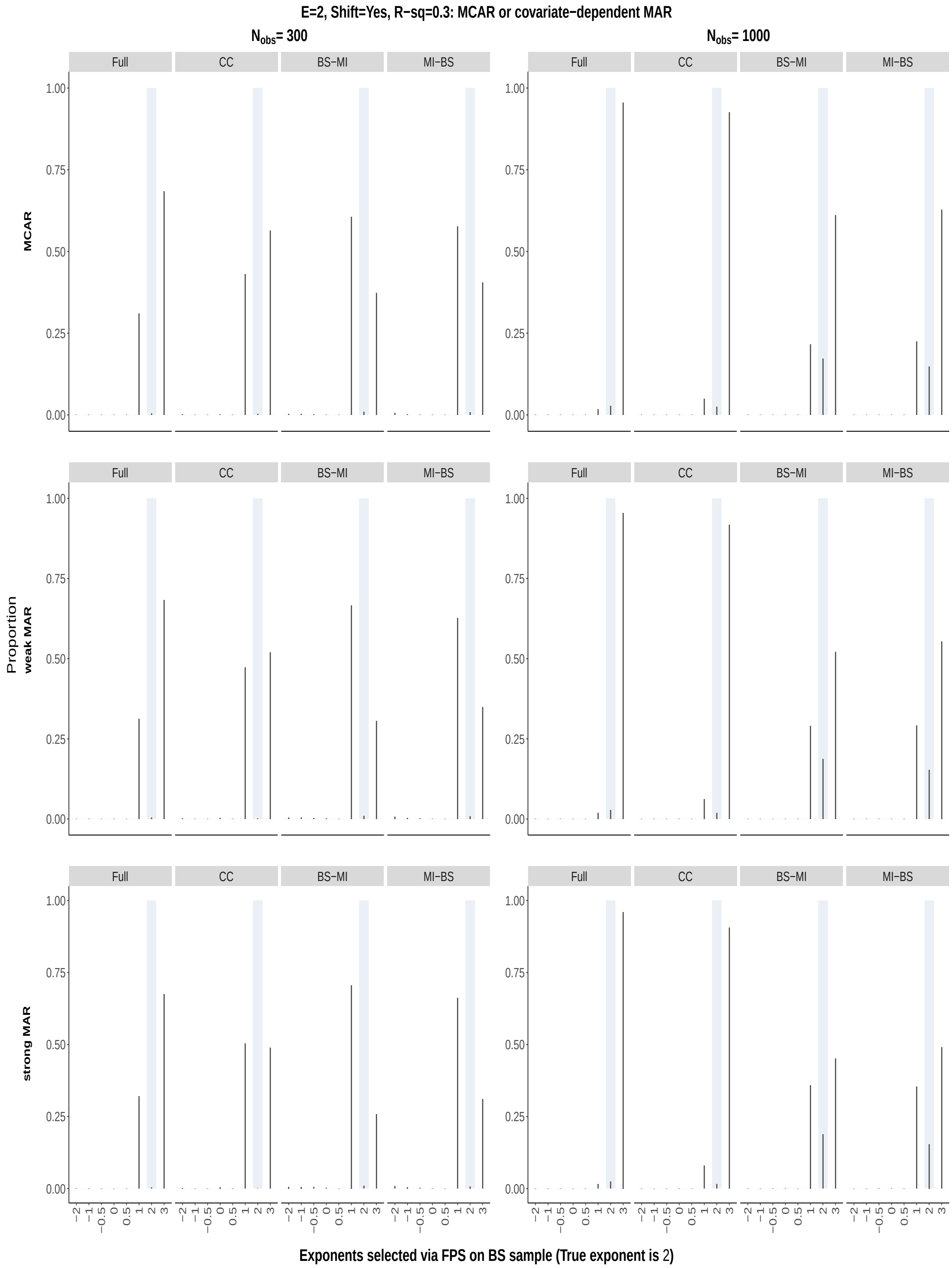

Figure S186: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

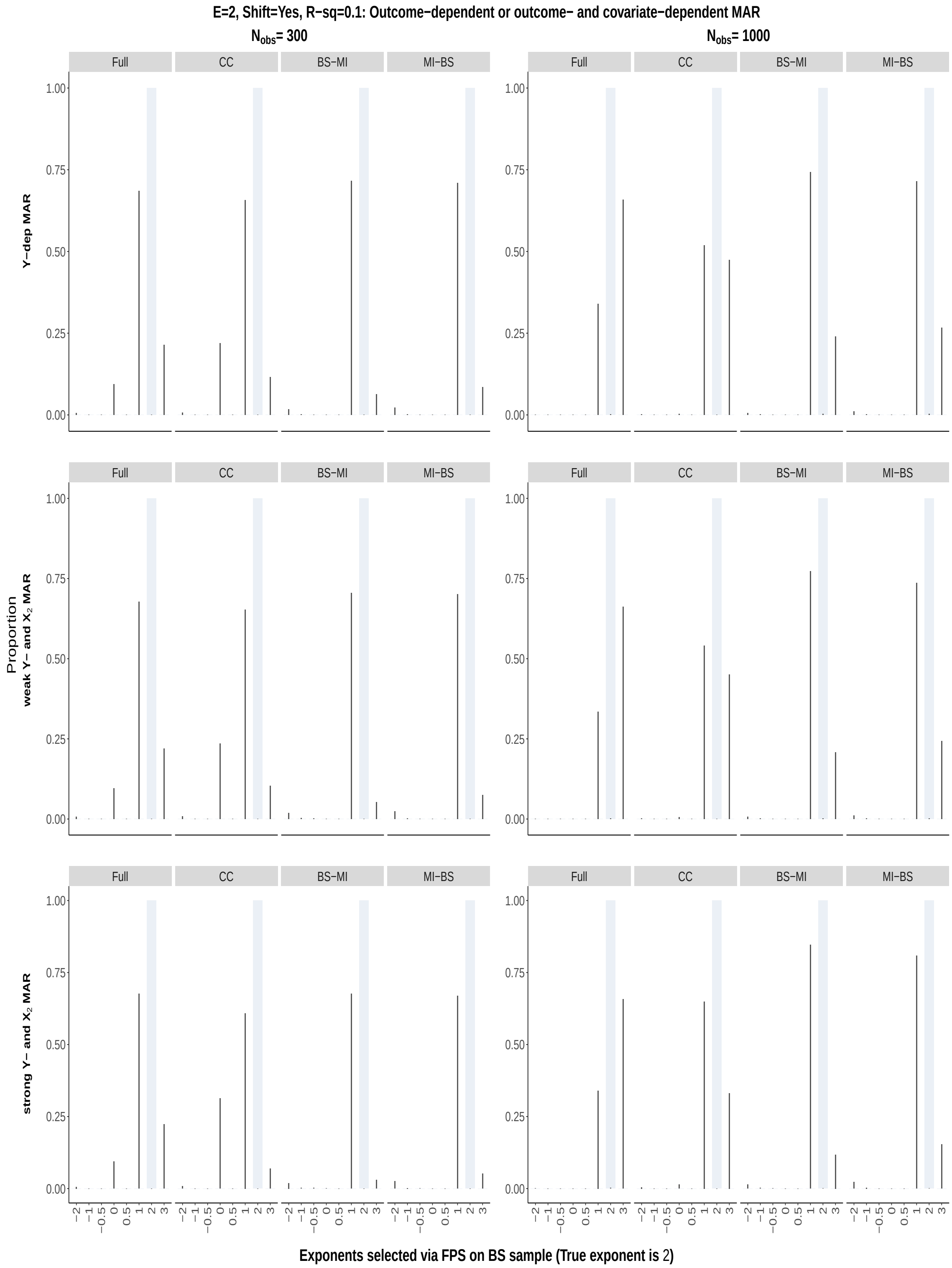

Figure S187: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

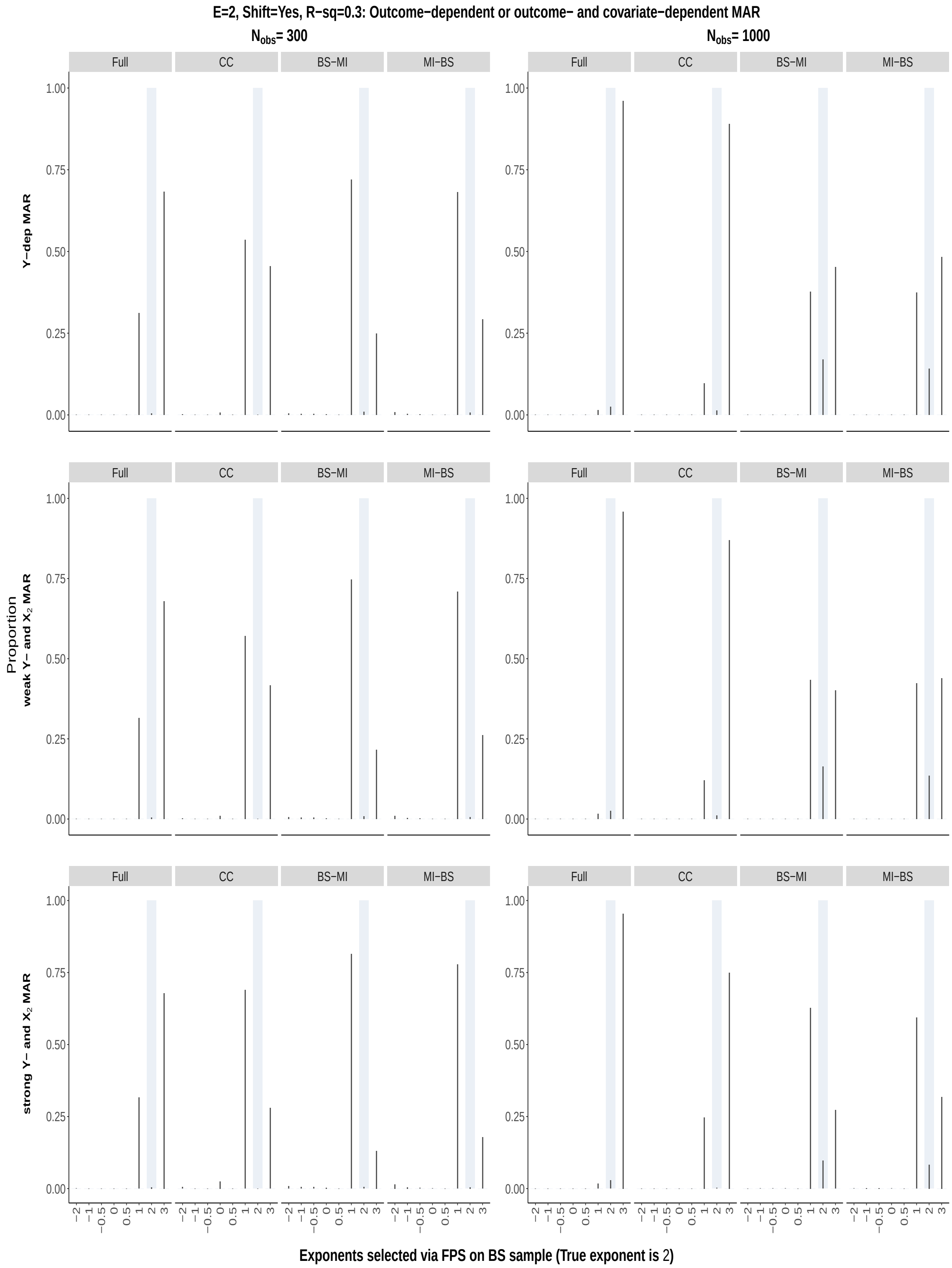

Figure S188: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

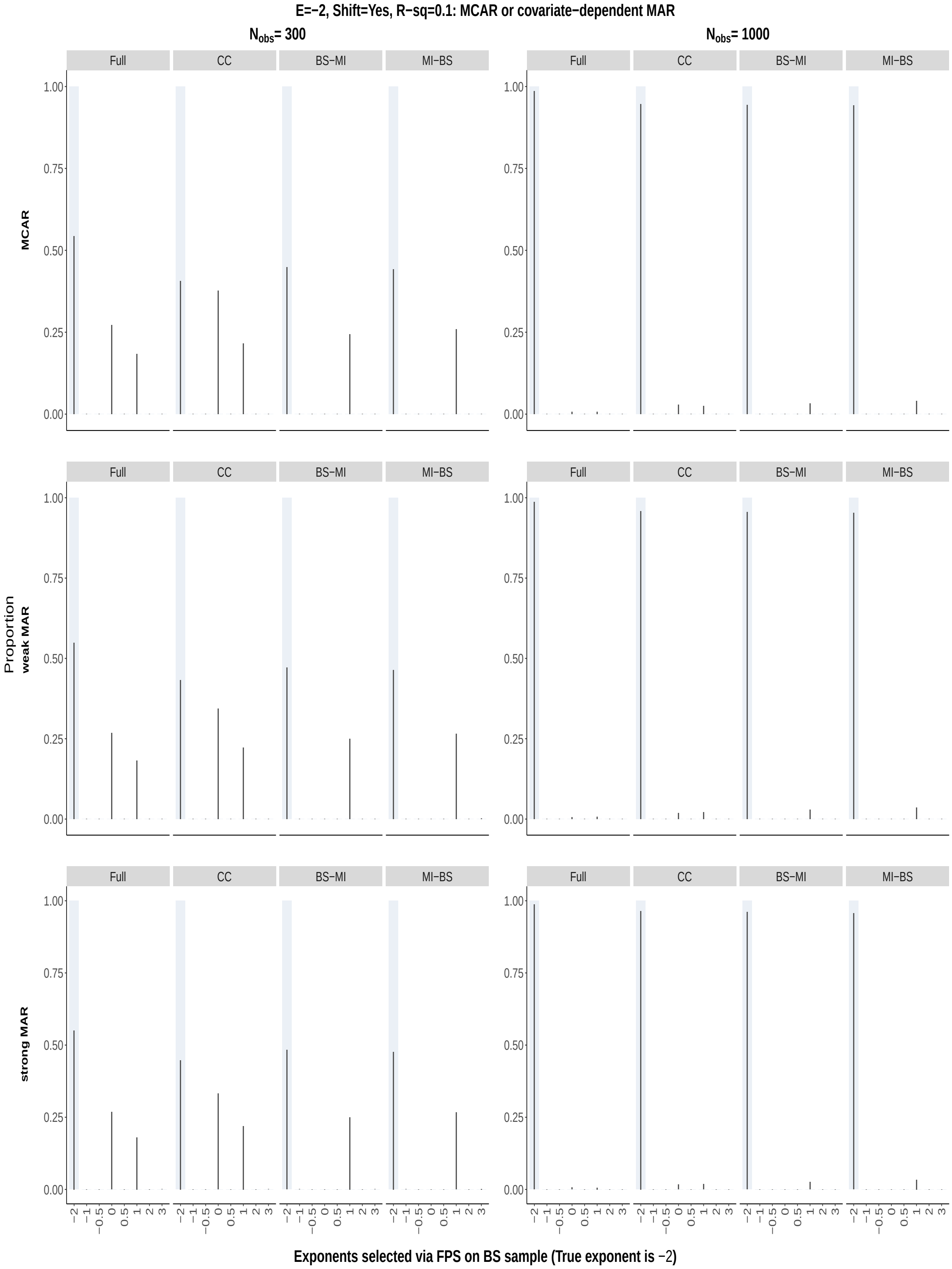

Figure S189: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

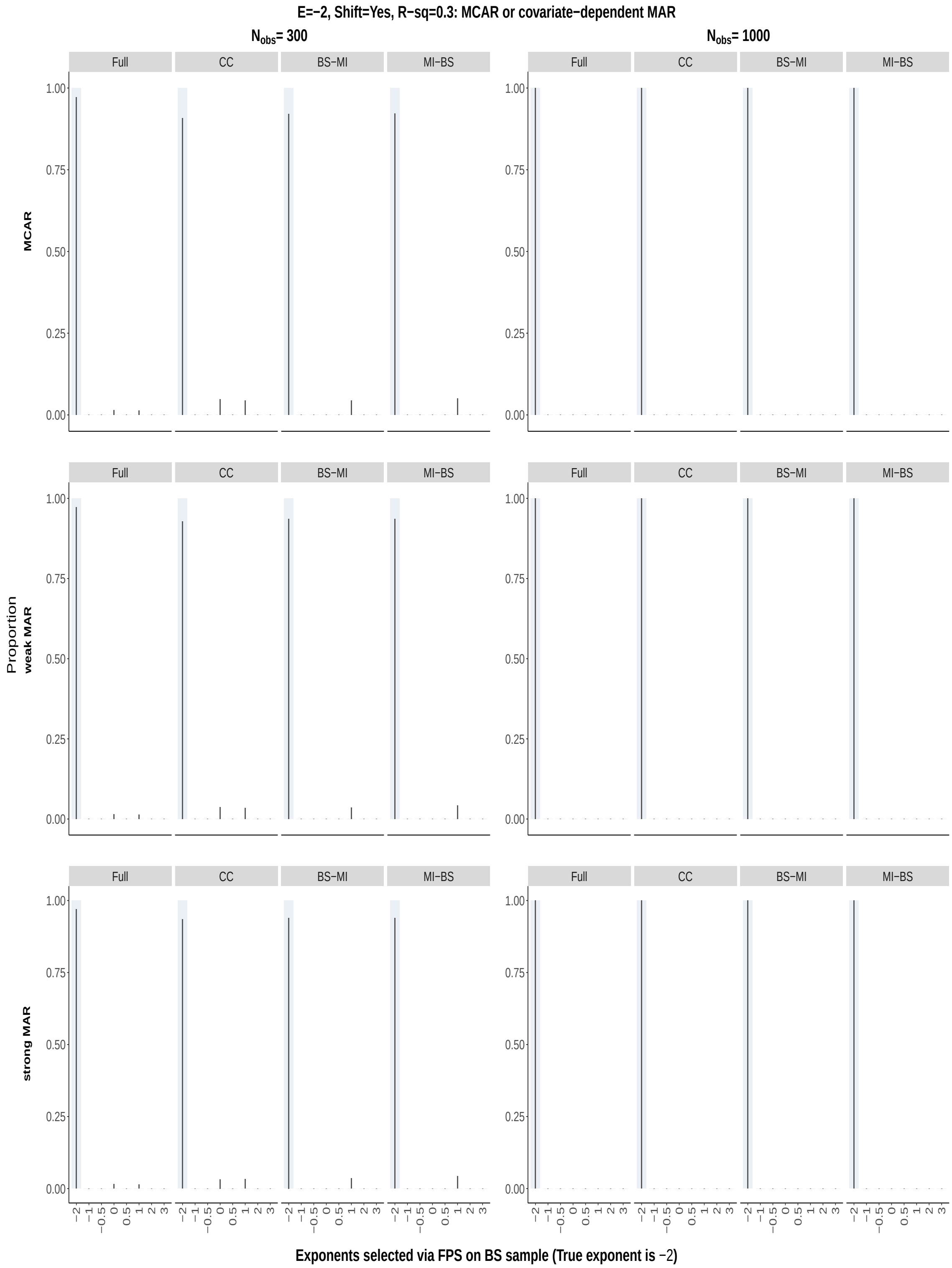

Figure S190: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

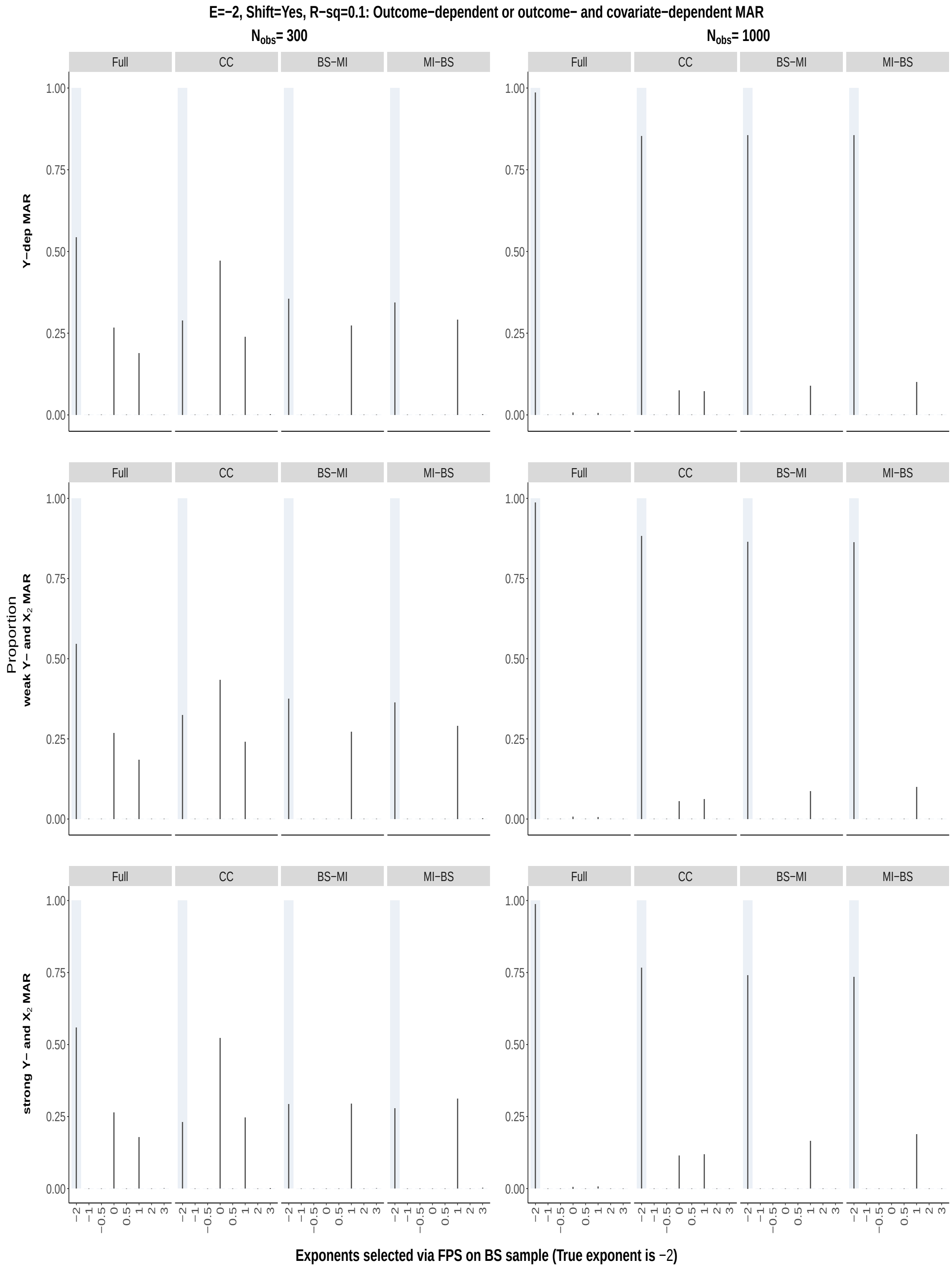

Figure S191: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

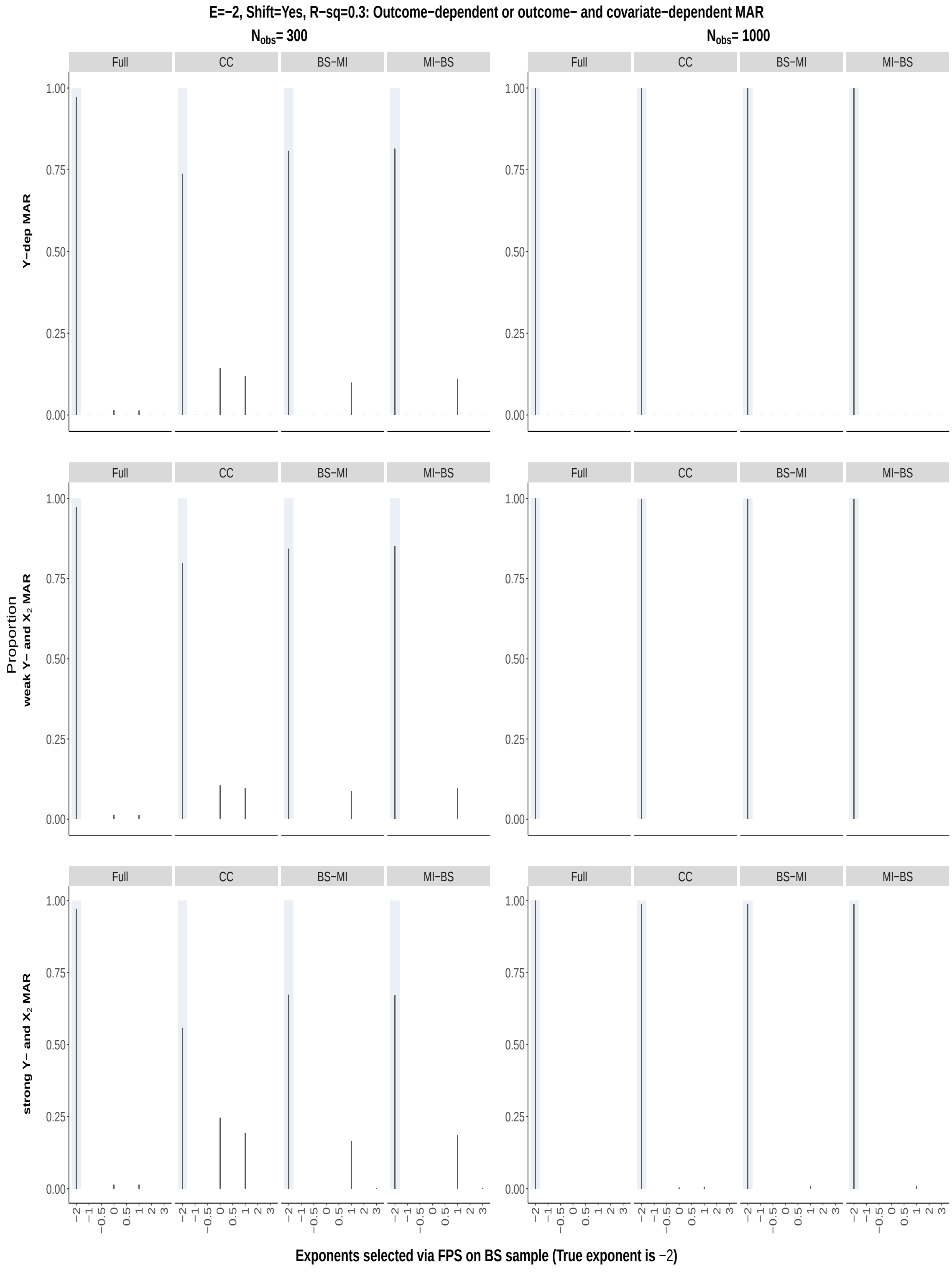

Figure S192: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.13 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 0$, $\alpha_E = 1$ and no origin-shift

True exponent is 0

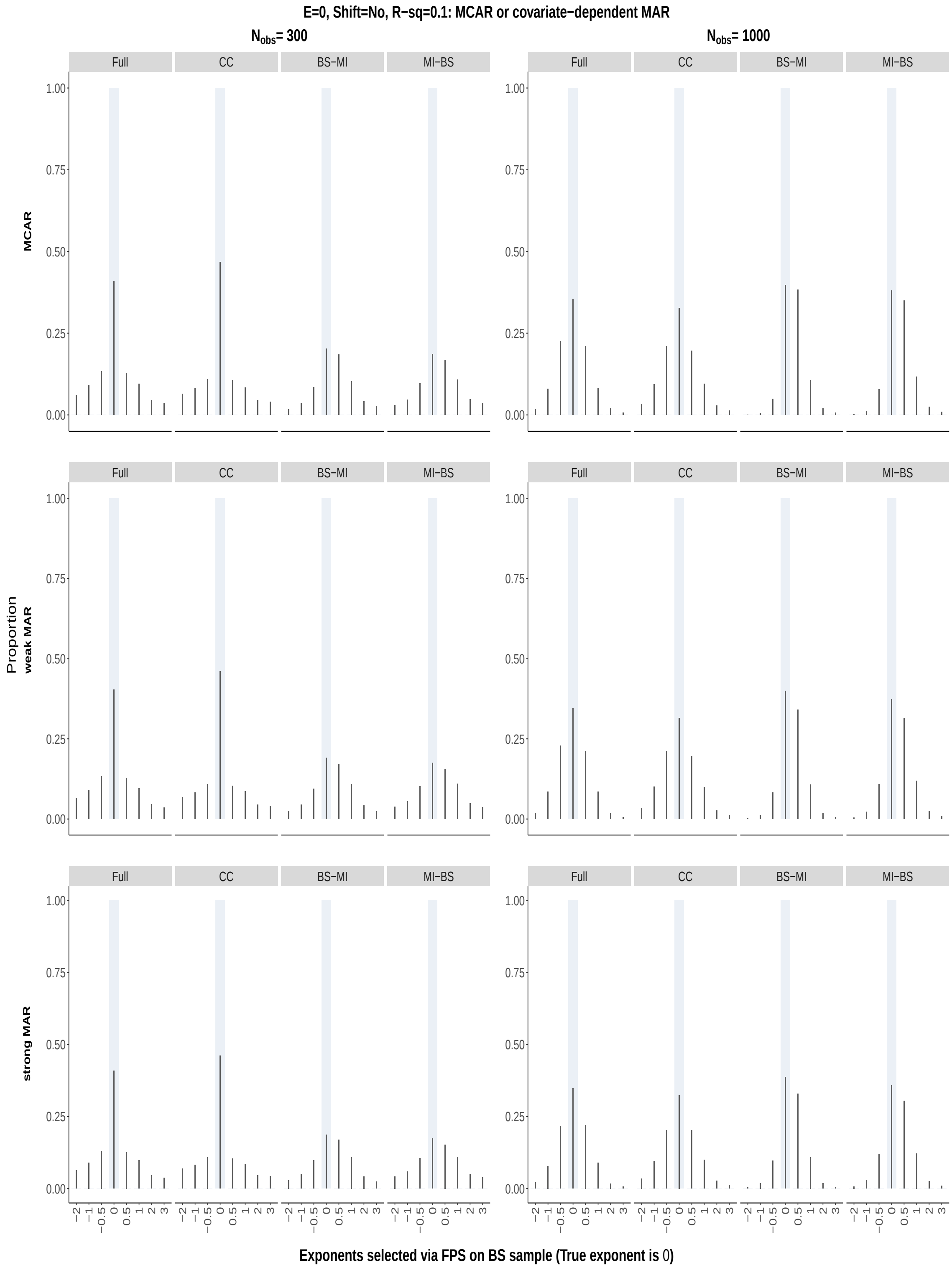

Figure S193: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

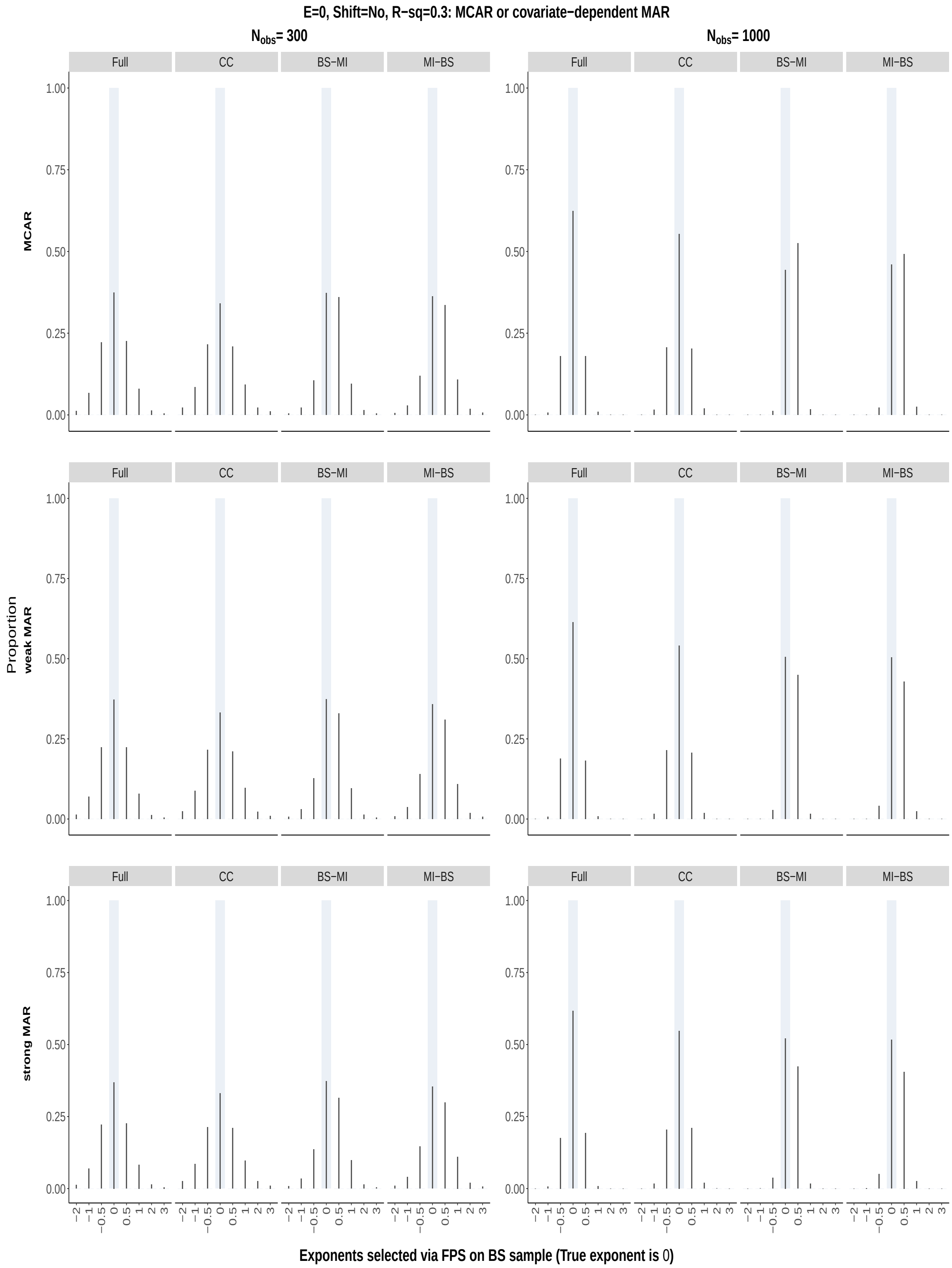

Figure S194: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

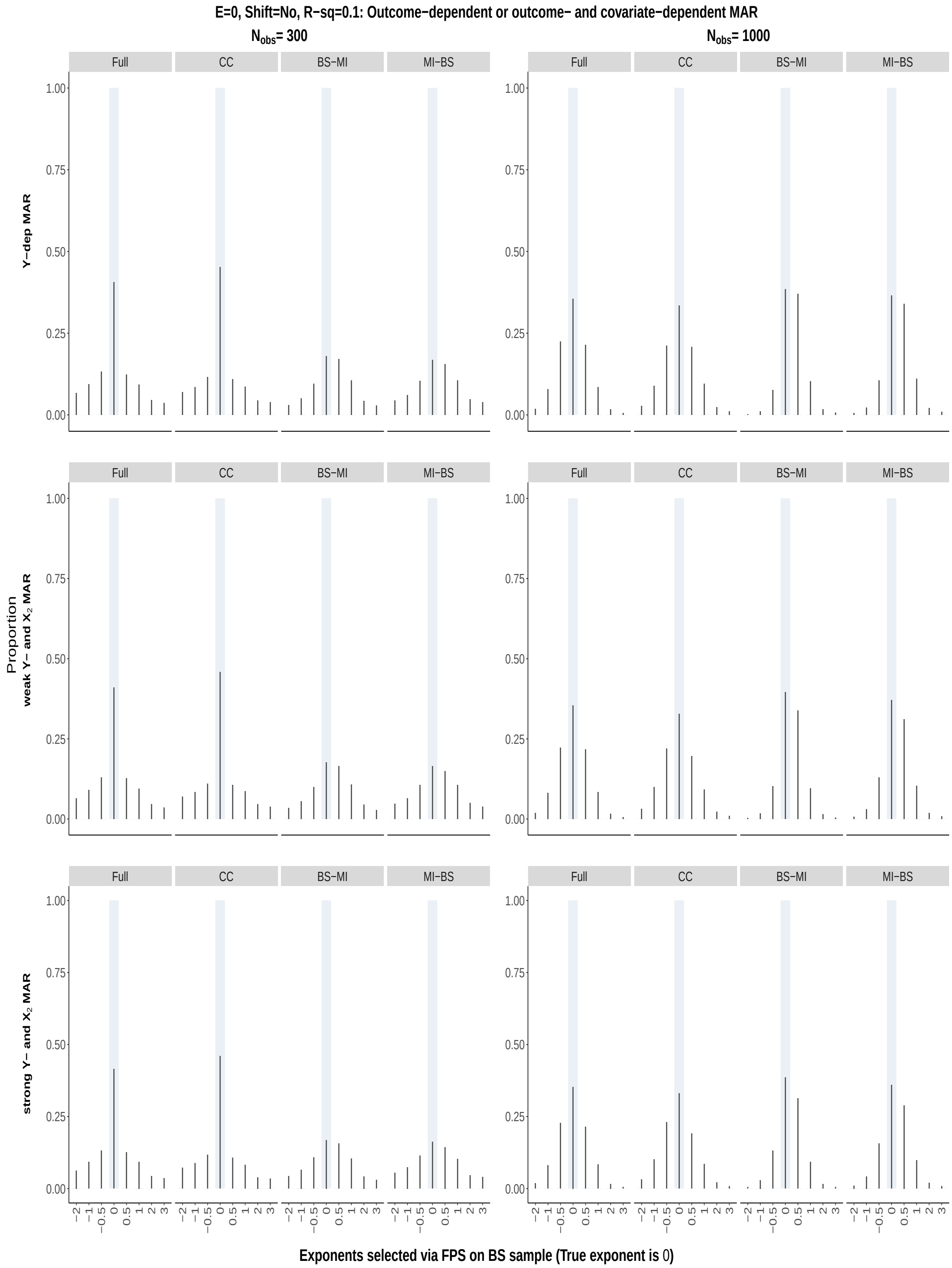

Figure S195: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

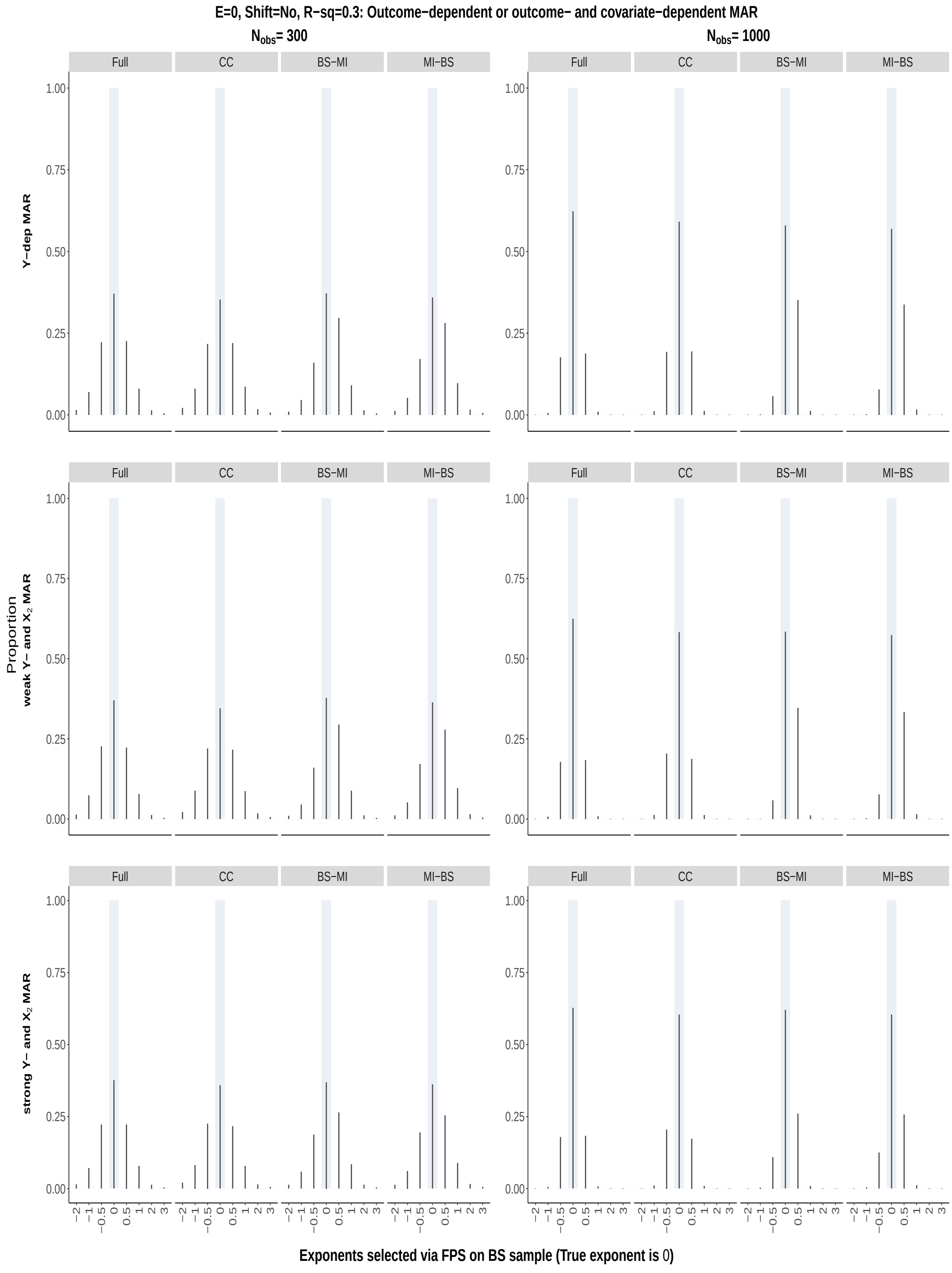

Figure S196: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

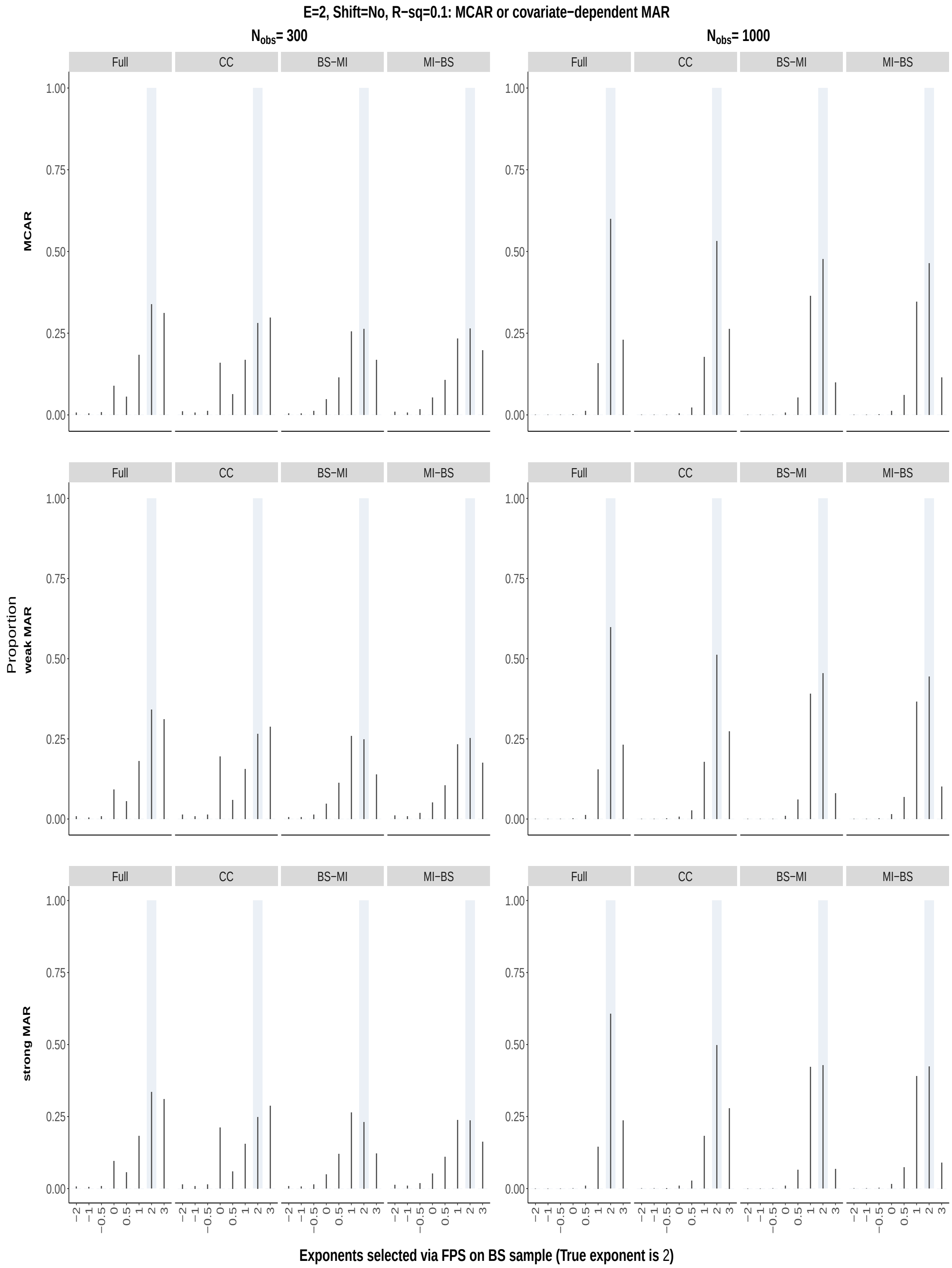

Figure S197: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

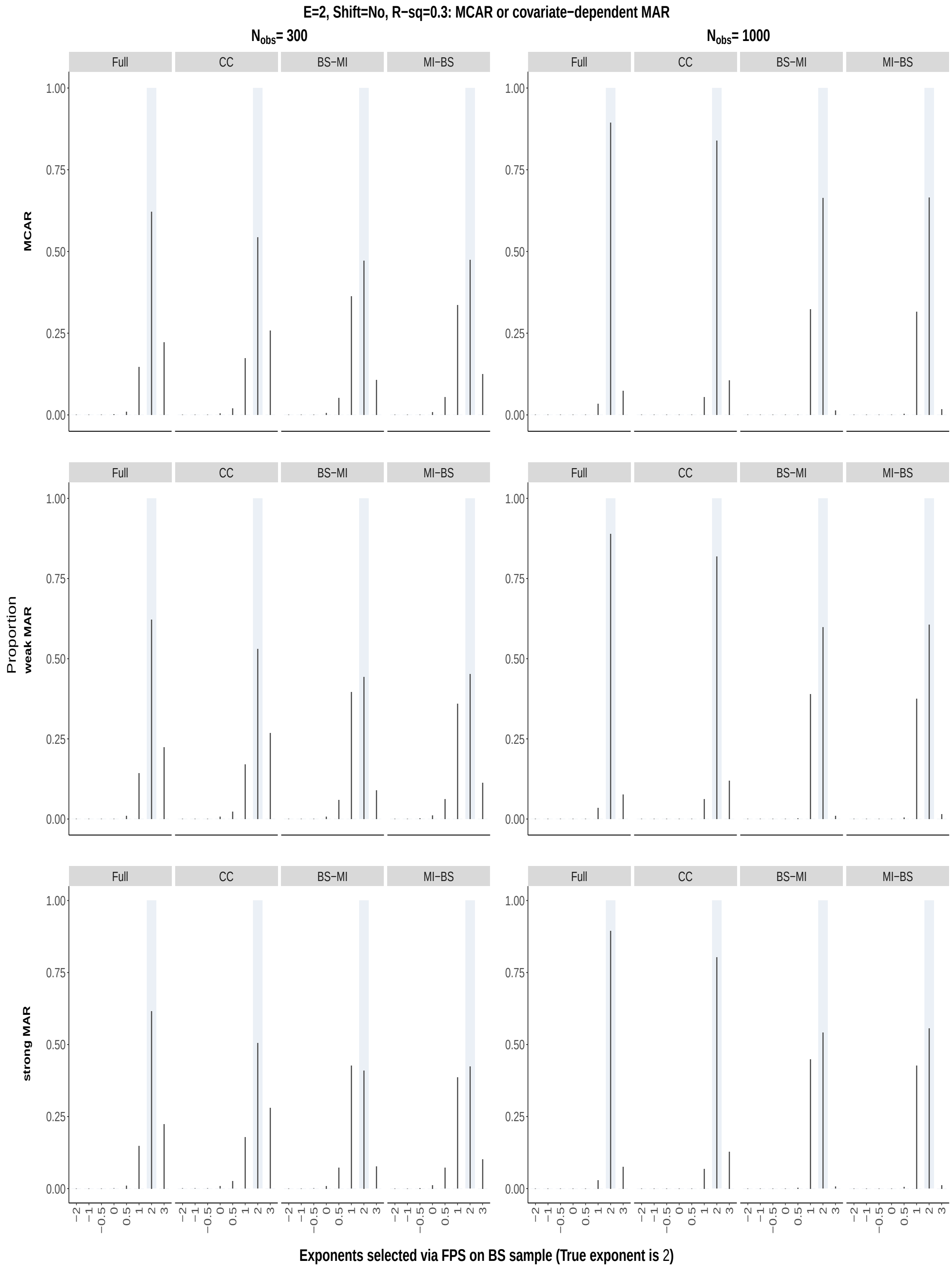

Figure S198: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

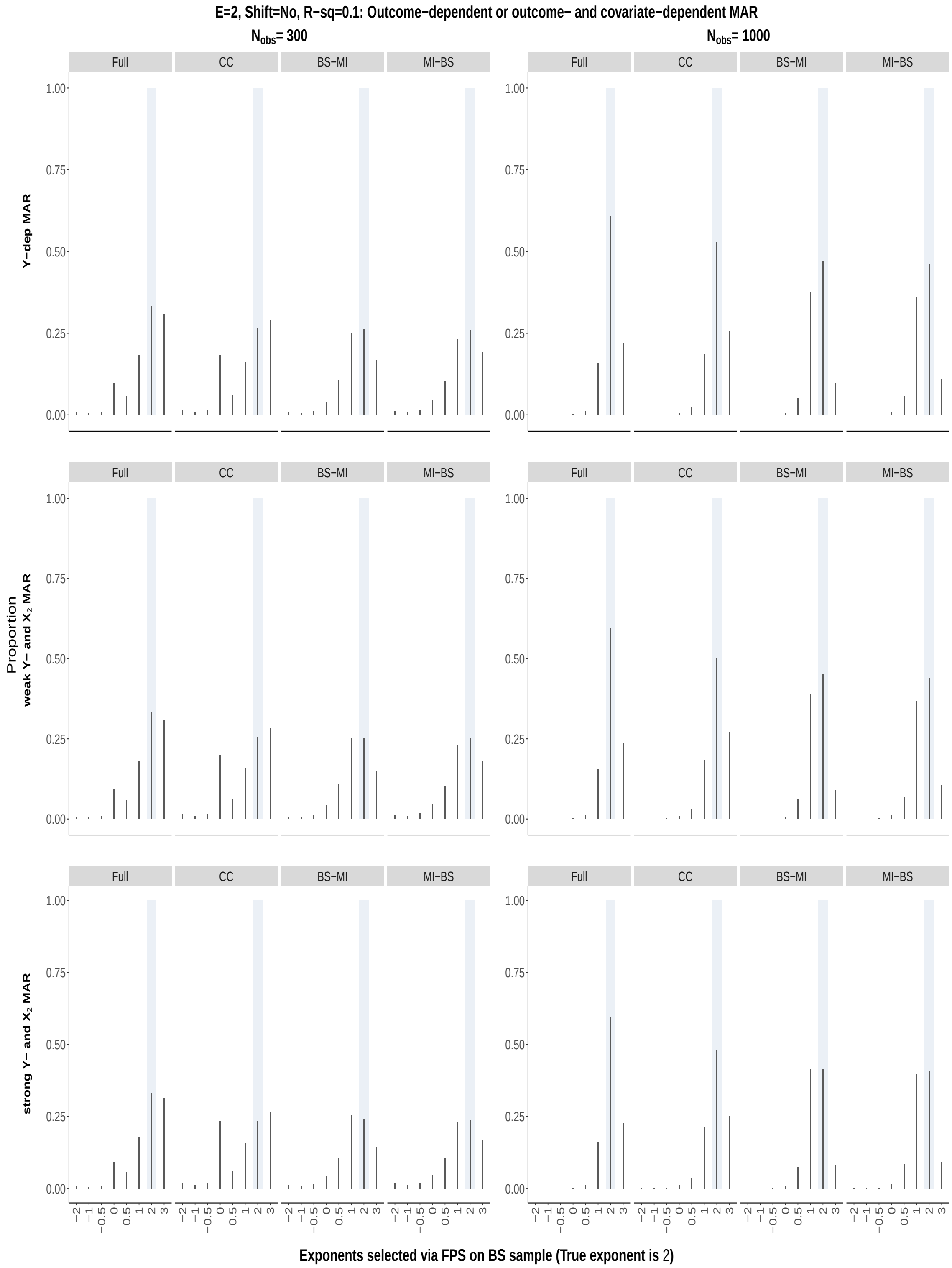

Figure S199: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

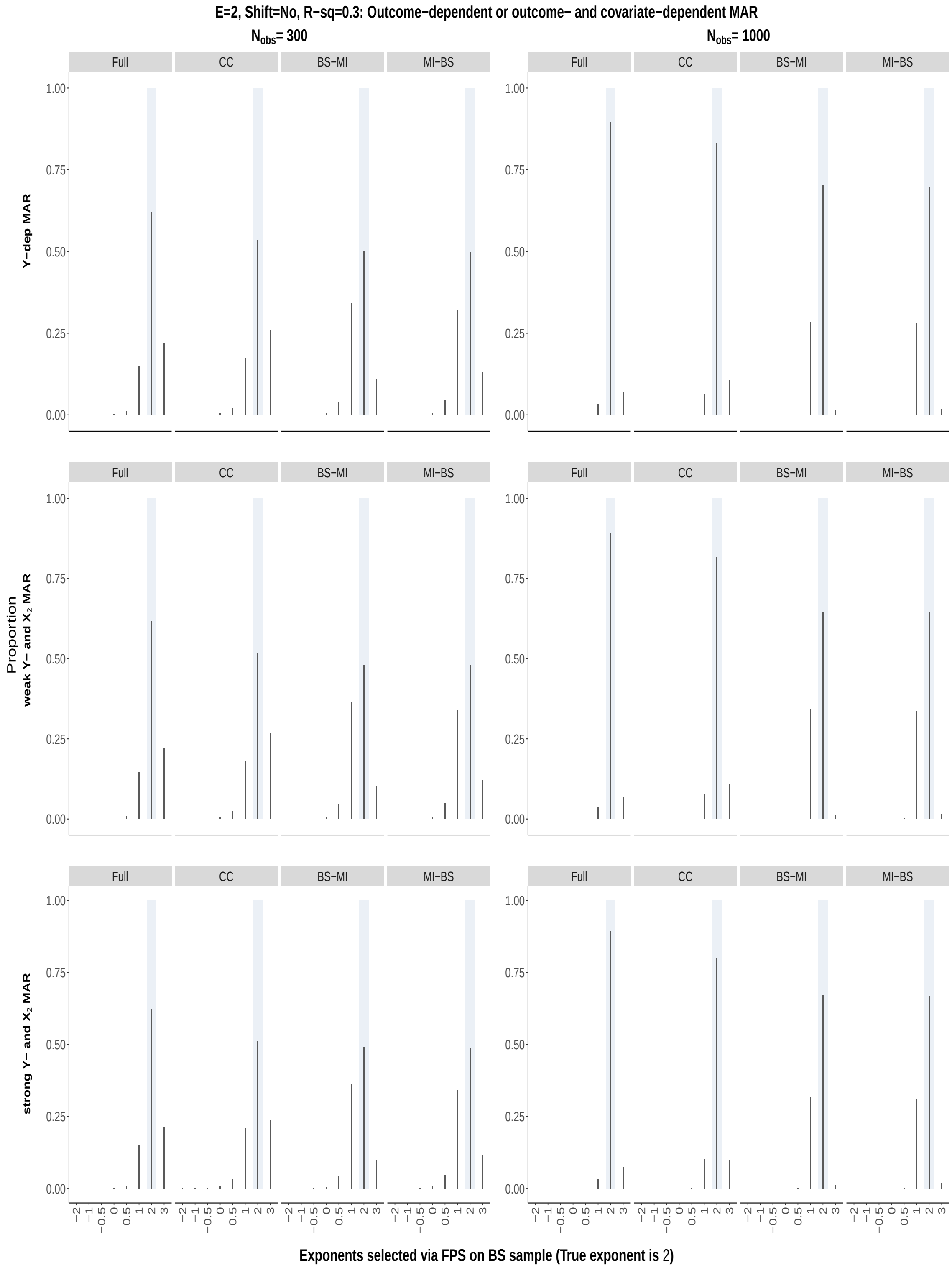

Figure S200: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

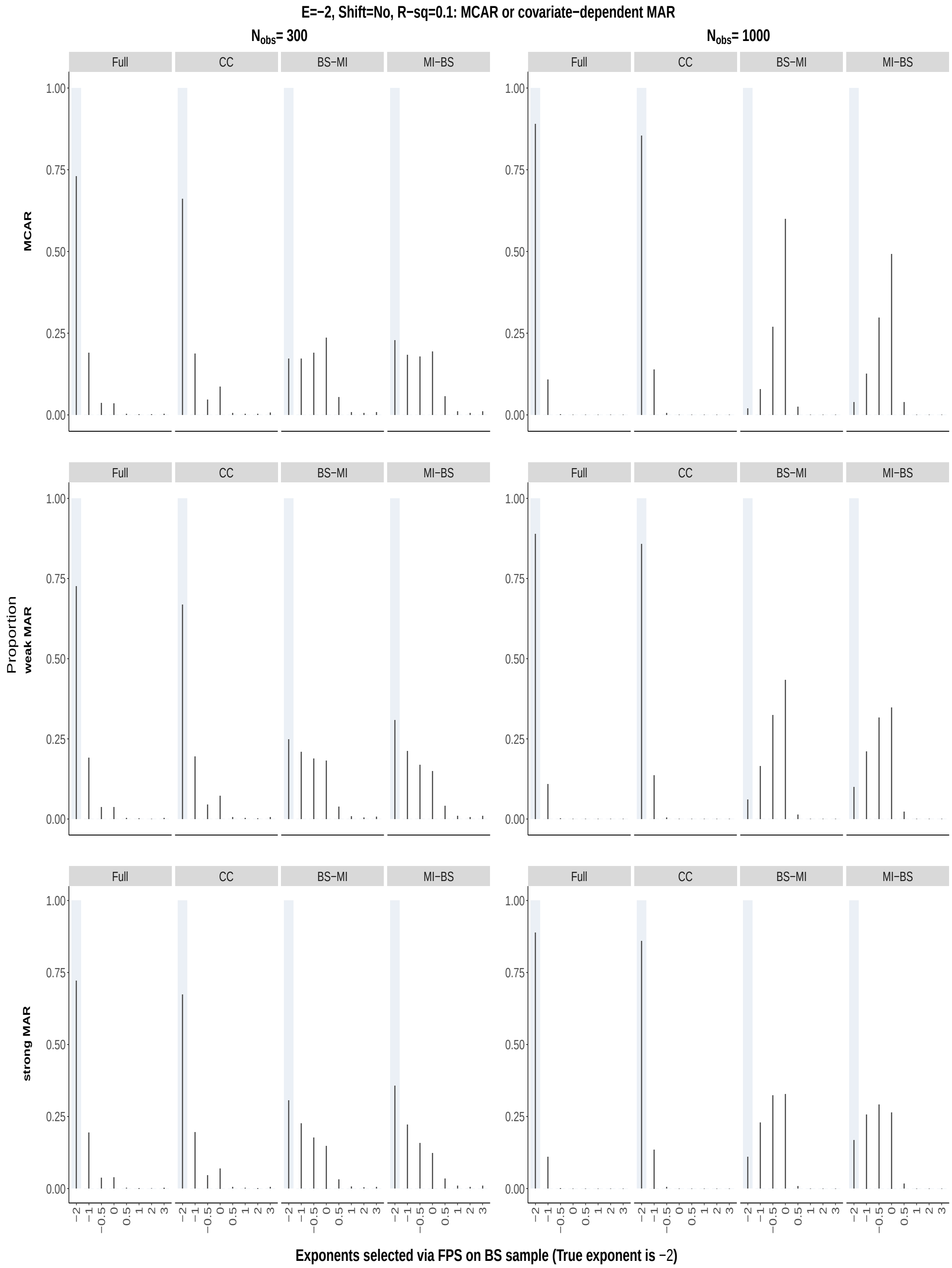

Figure S201: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

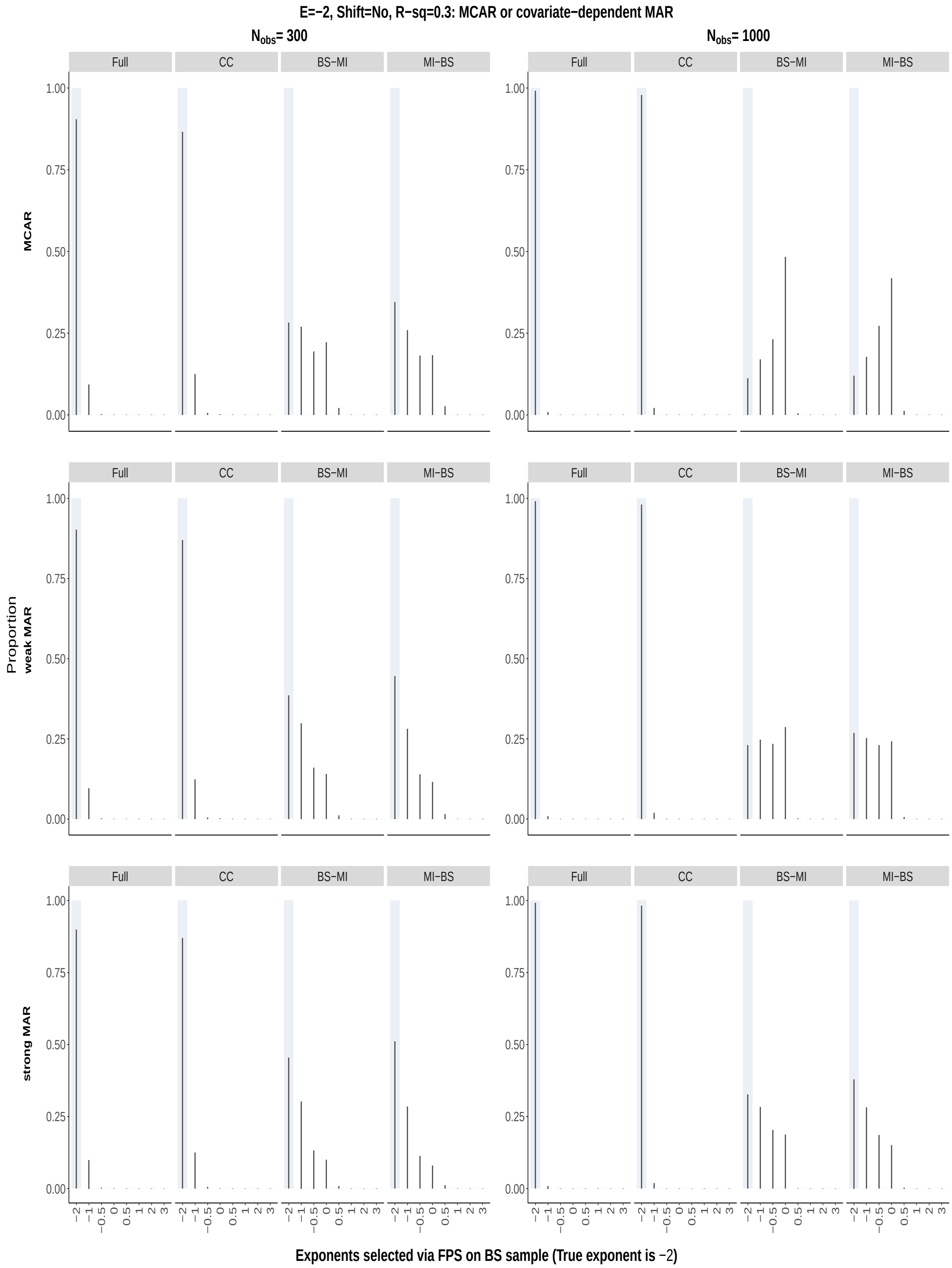

Figure S202: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

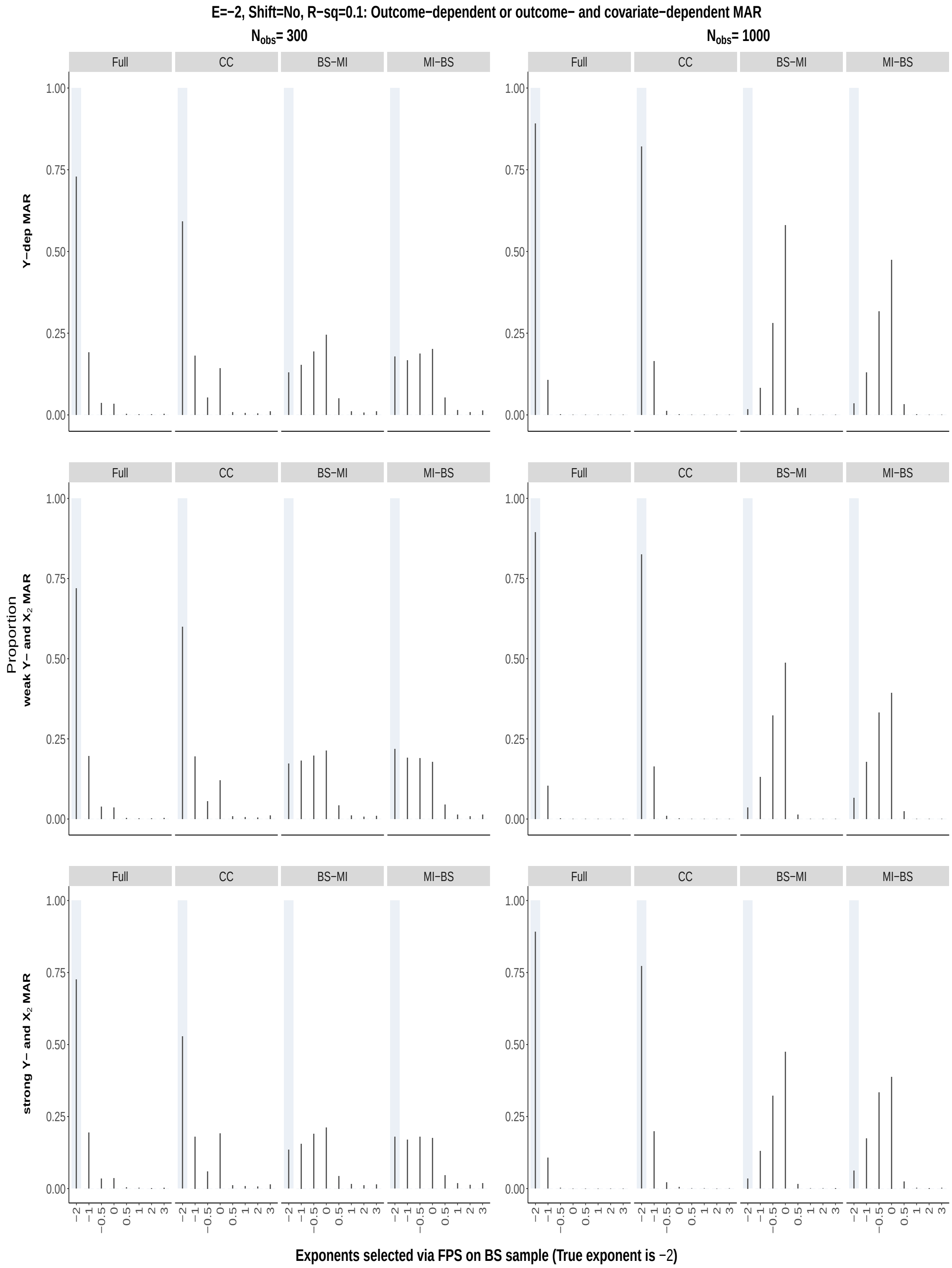

Figure S203: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

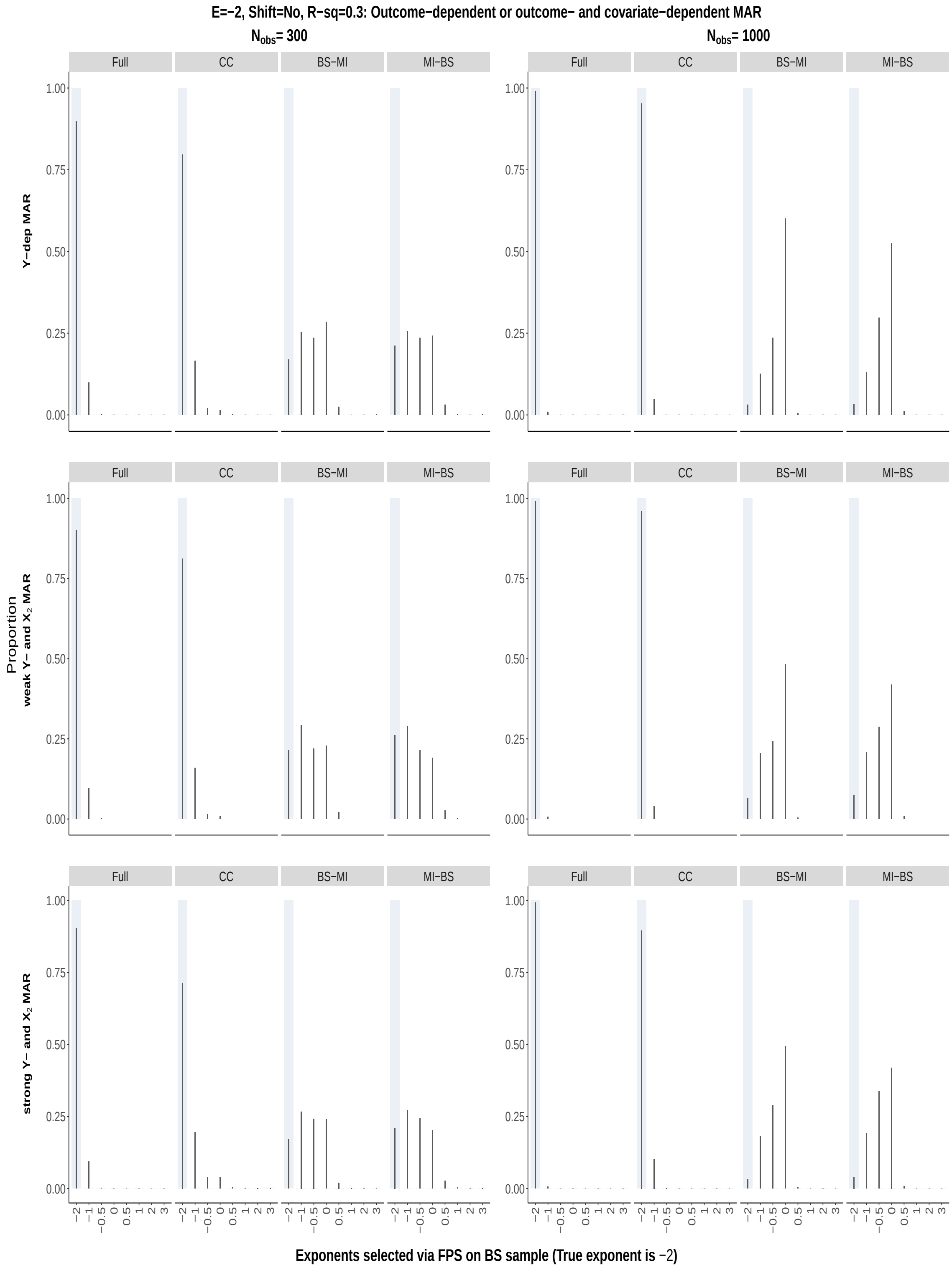

Figure S204: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.14 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 0$, $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

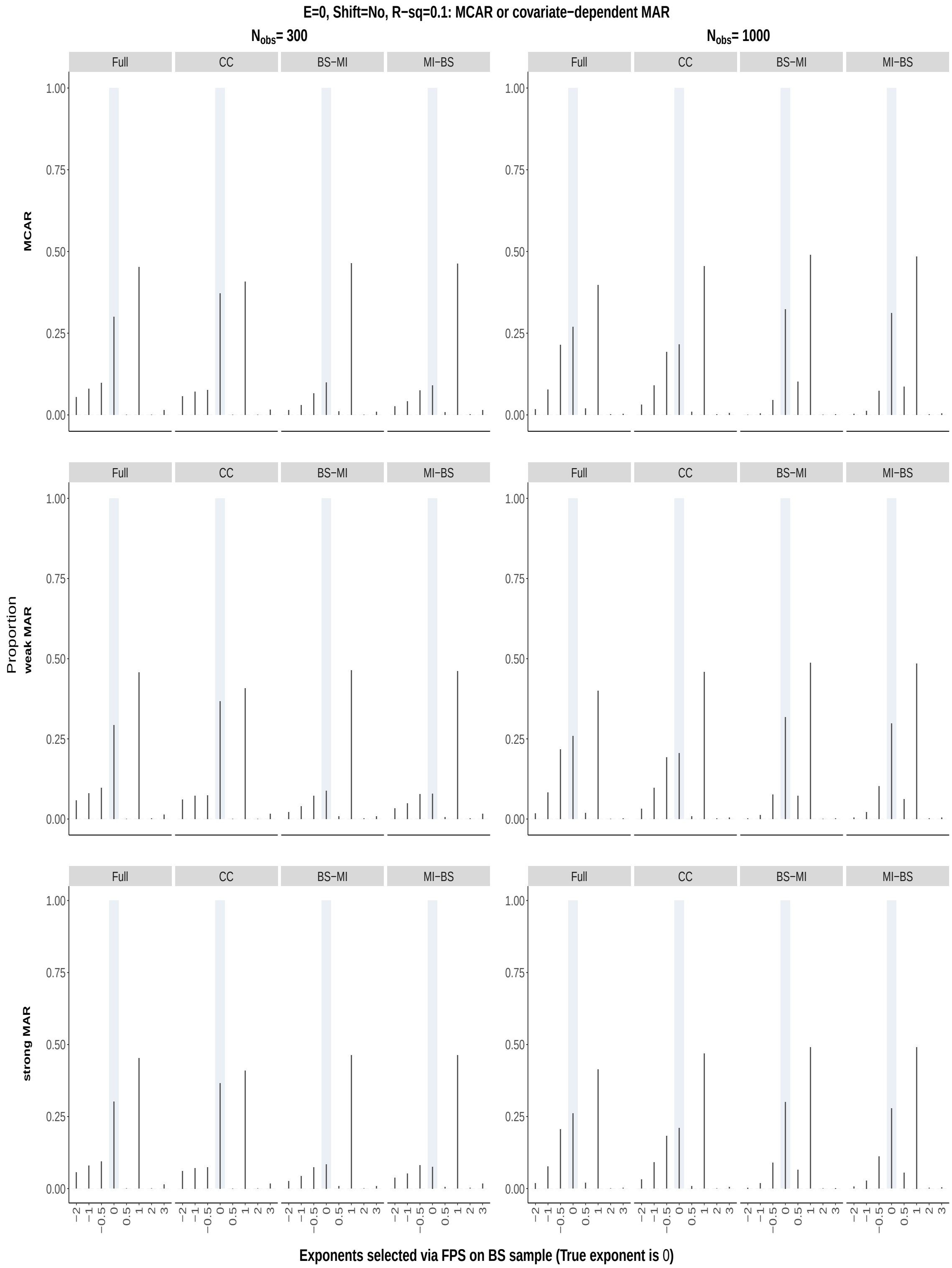

Figure S205: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

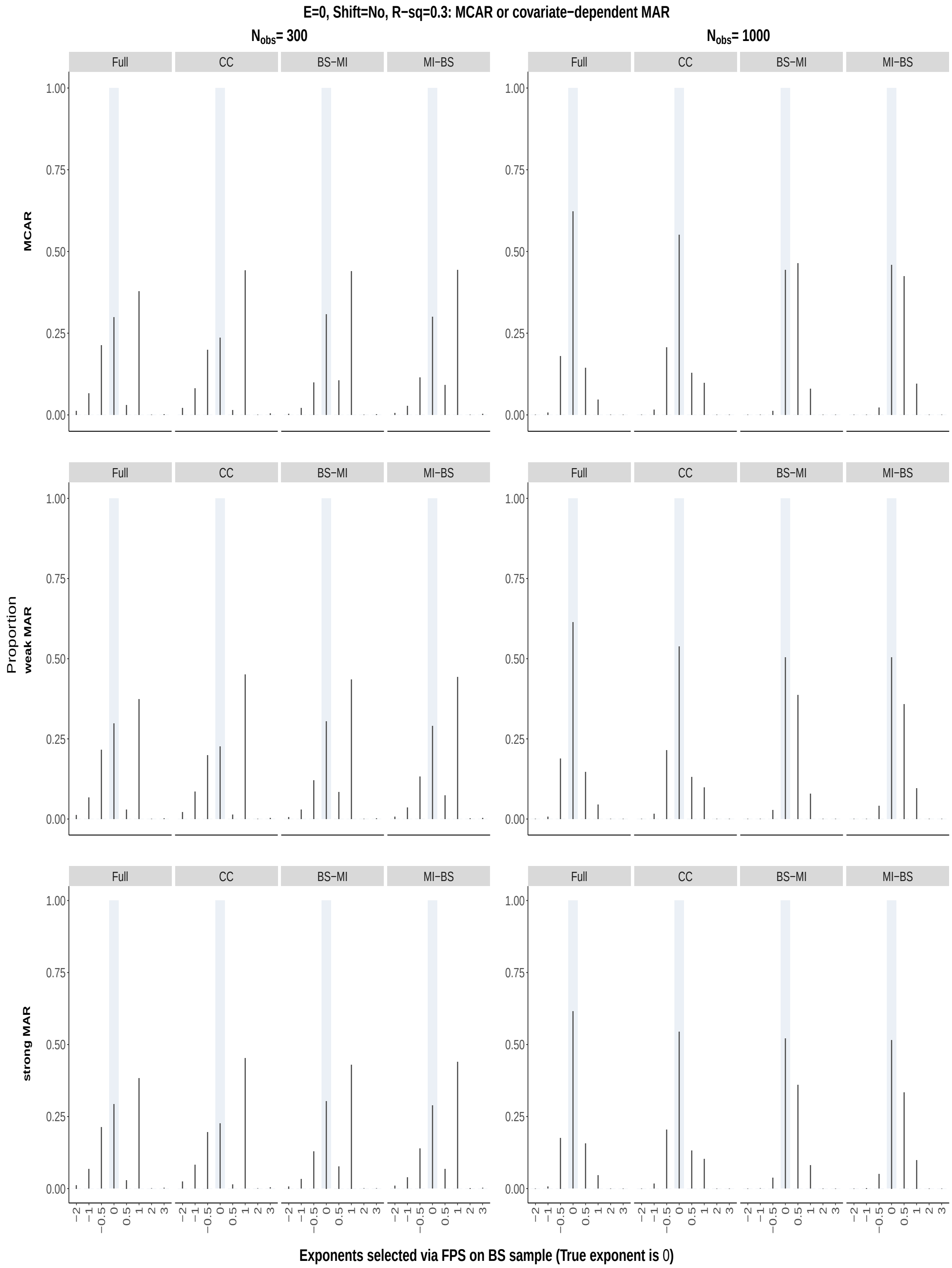

Figure S206: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

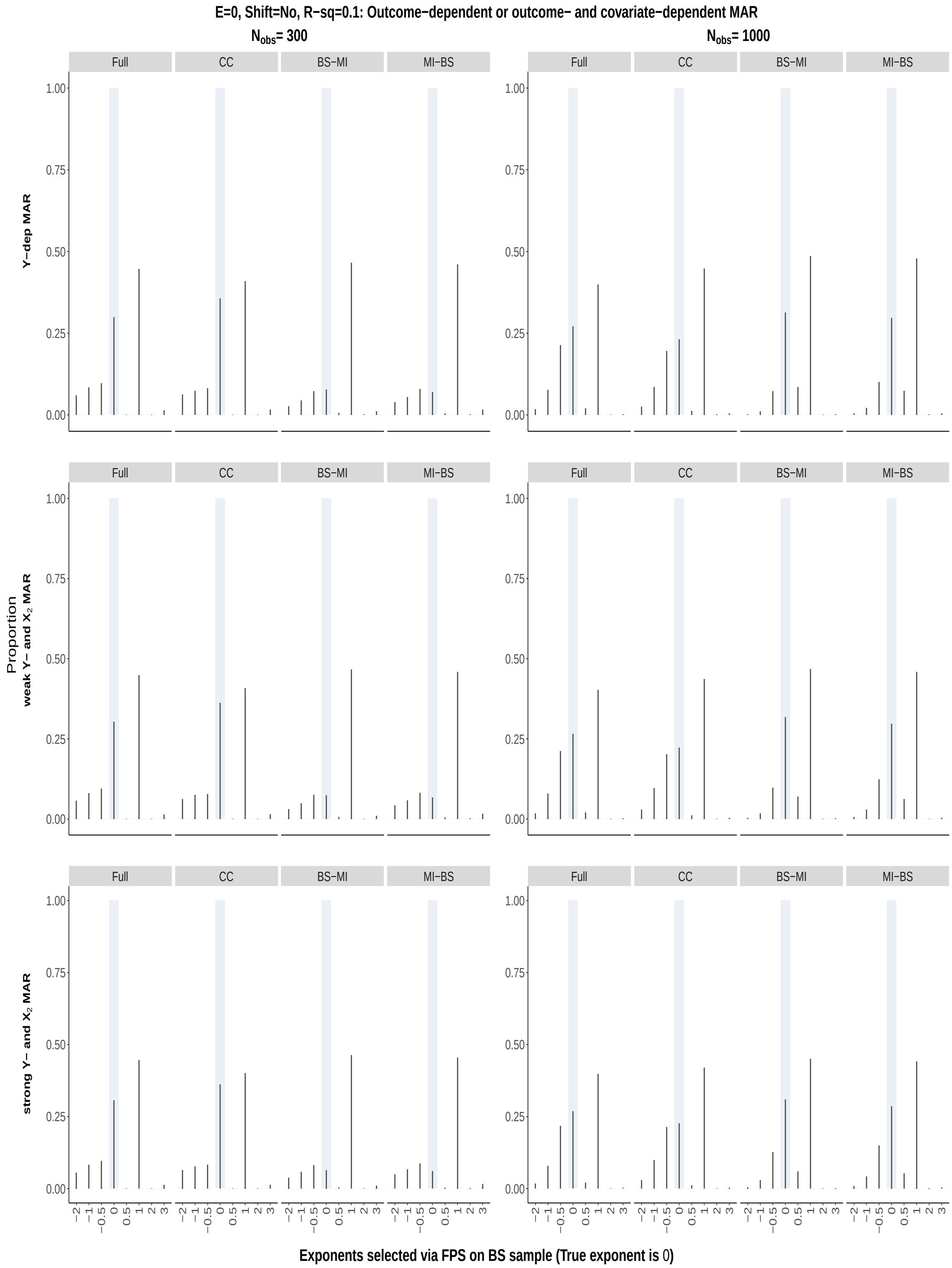

Figure S207: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

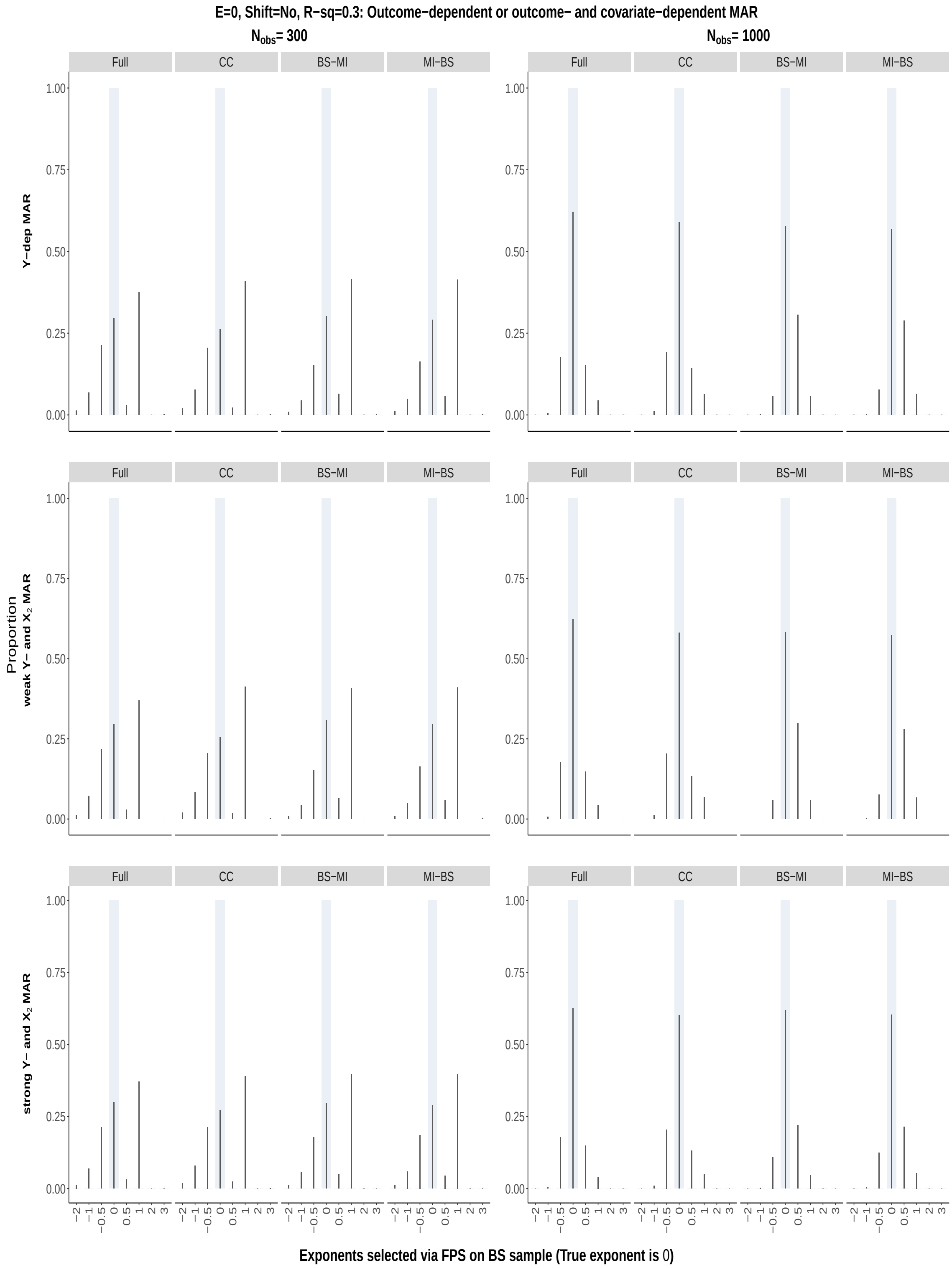

Figure S208: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

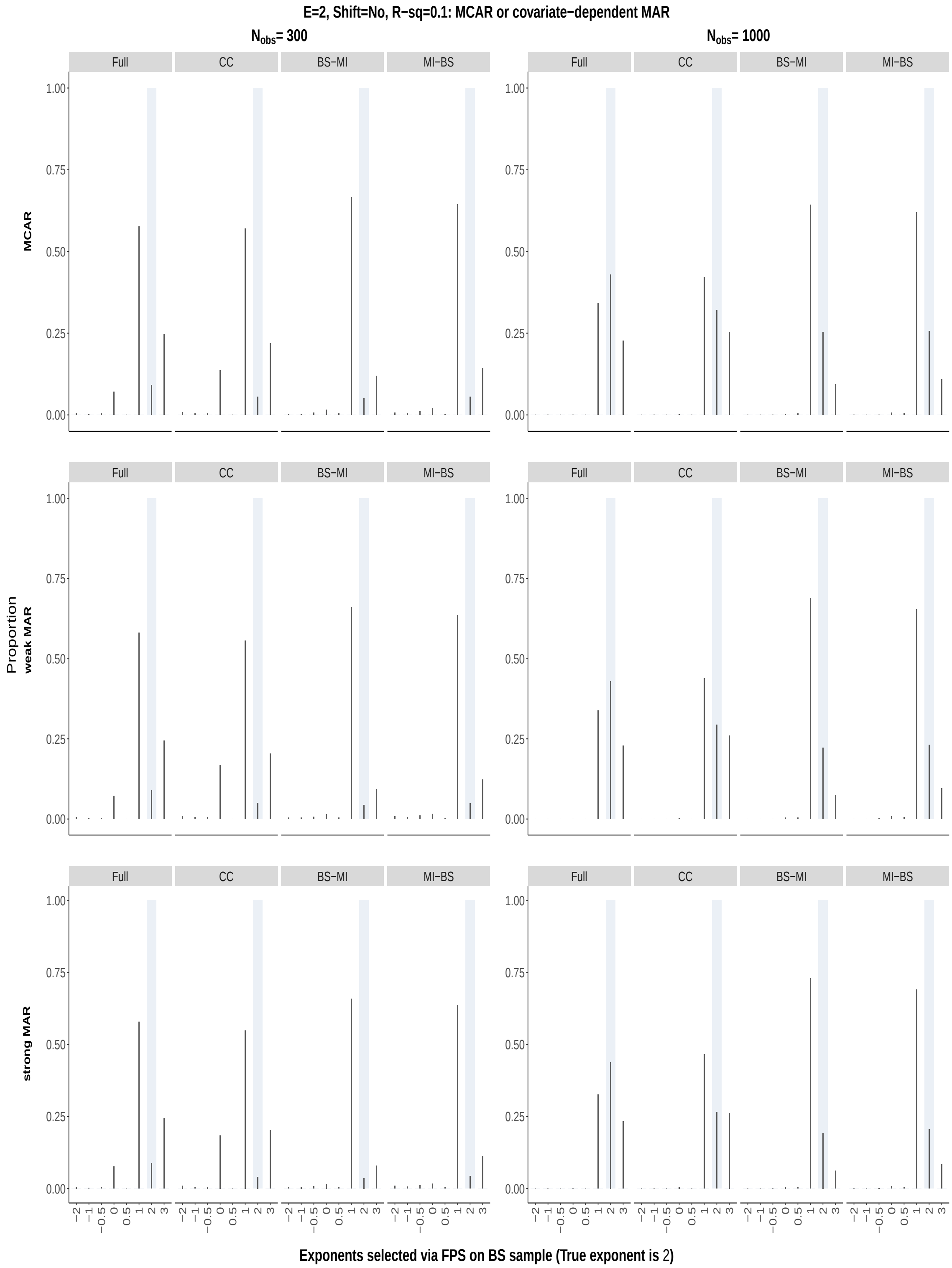

Figure S209: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

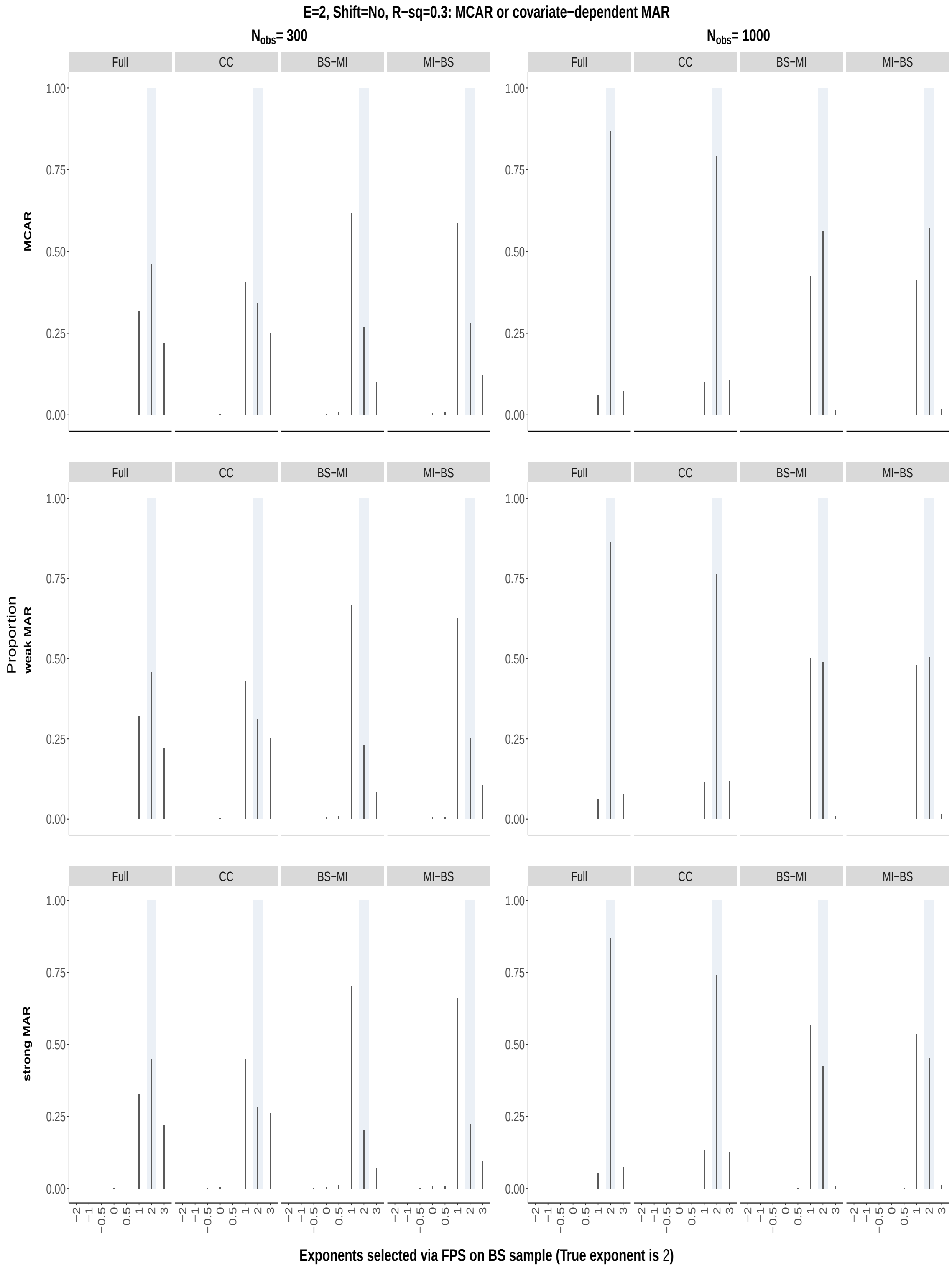

Figure S210: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

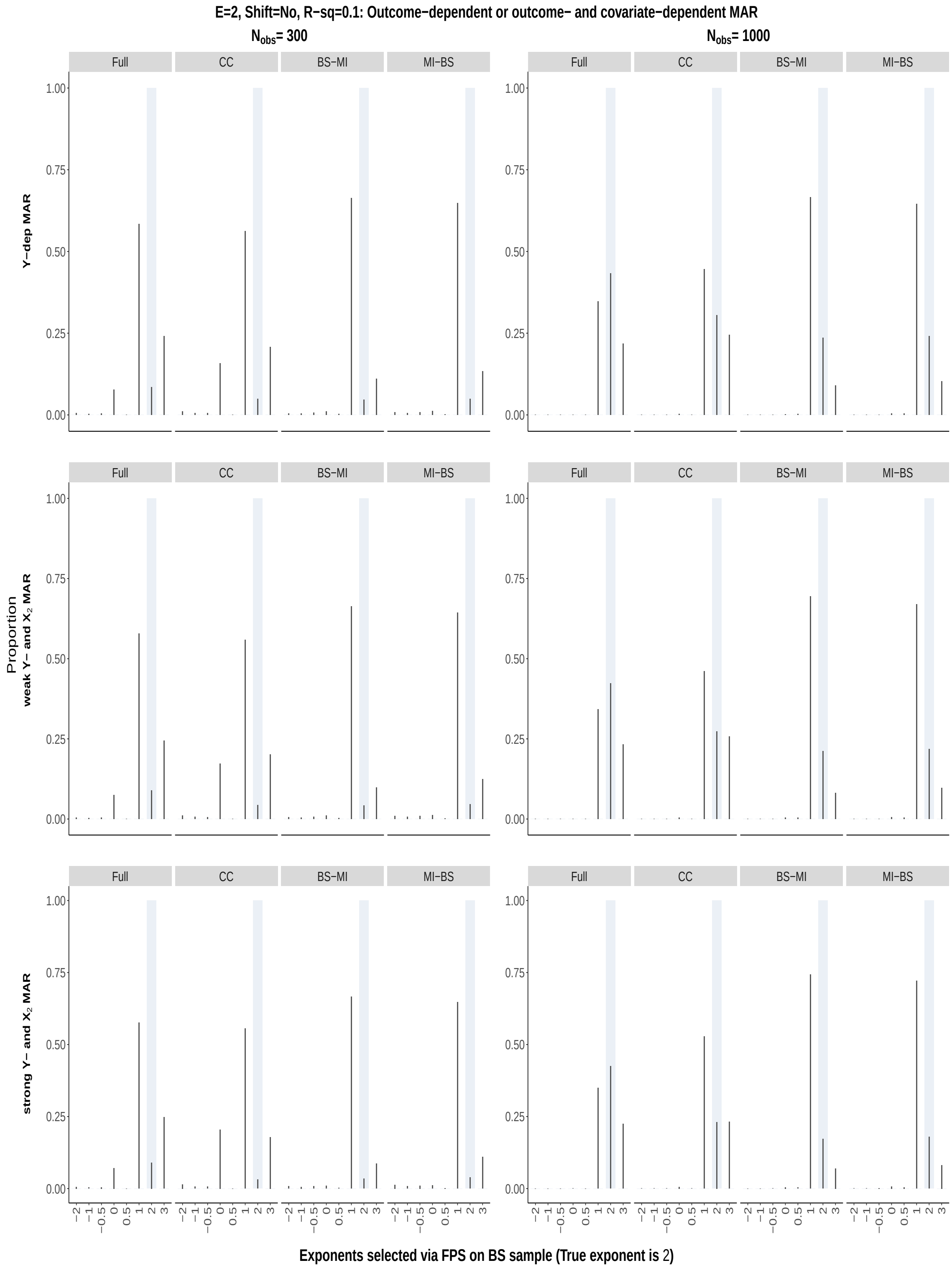

Figure S211: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

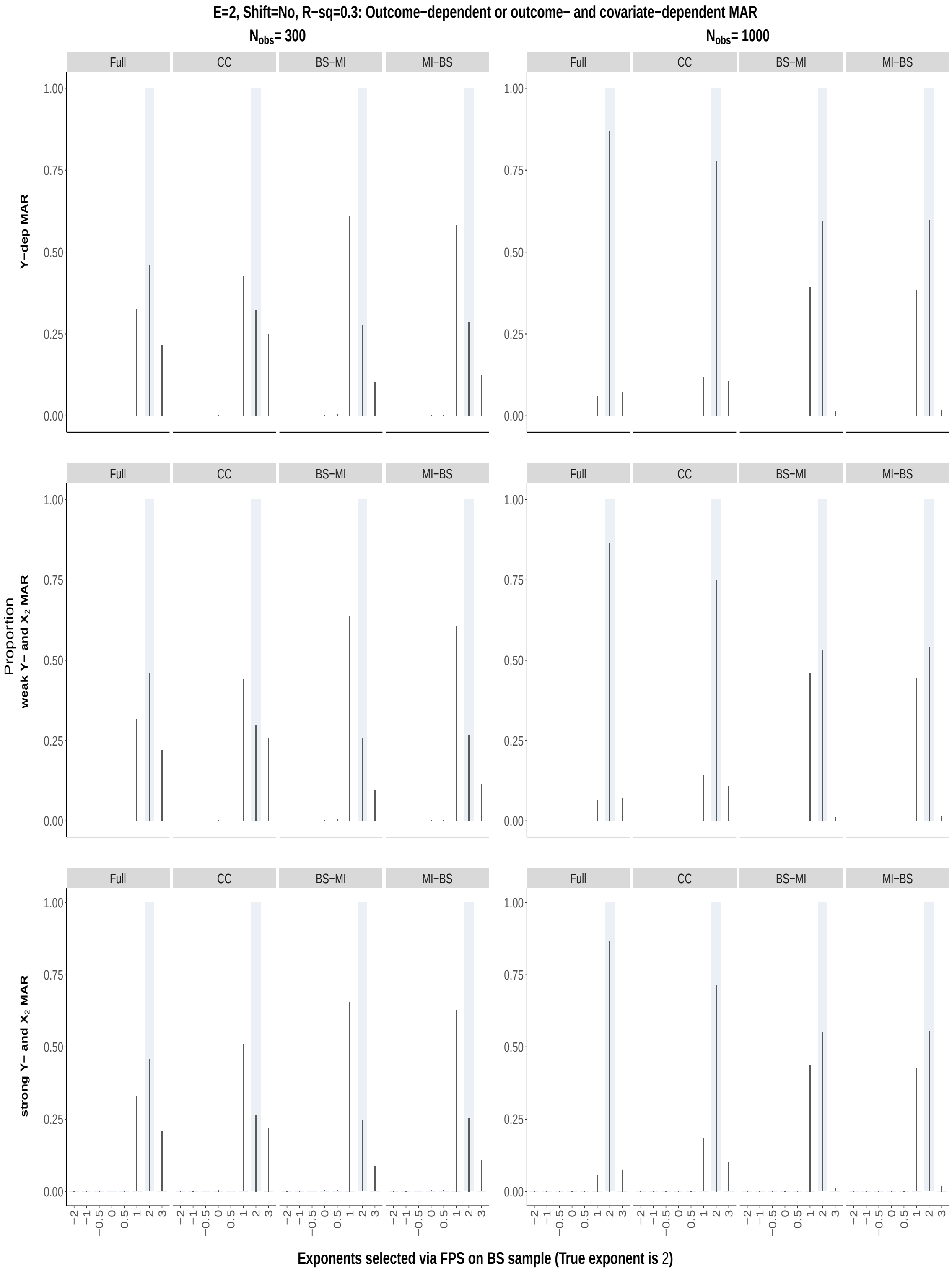

Figure S212: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

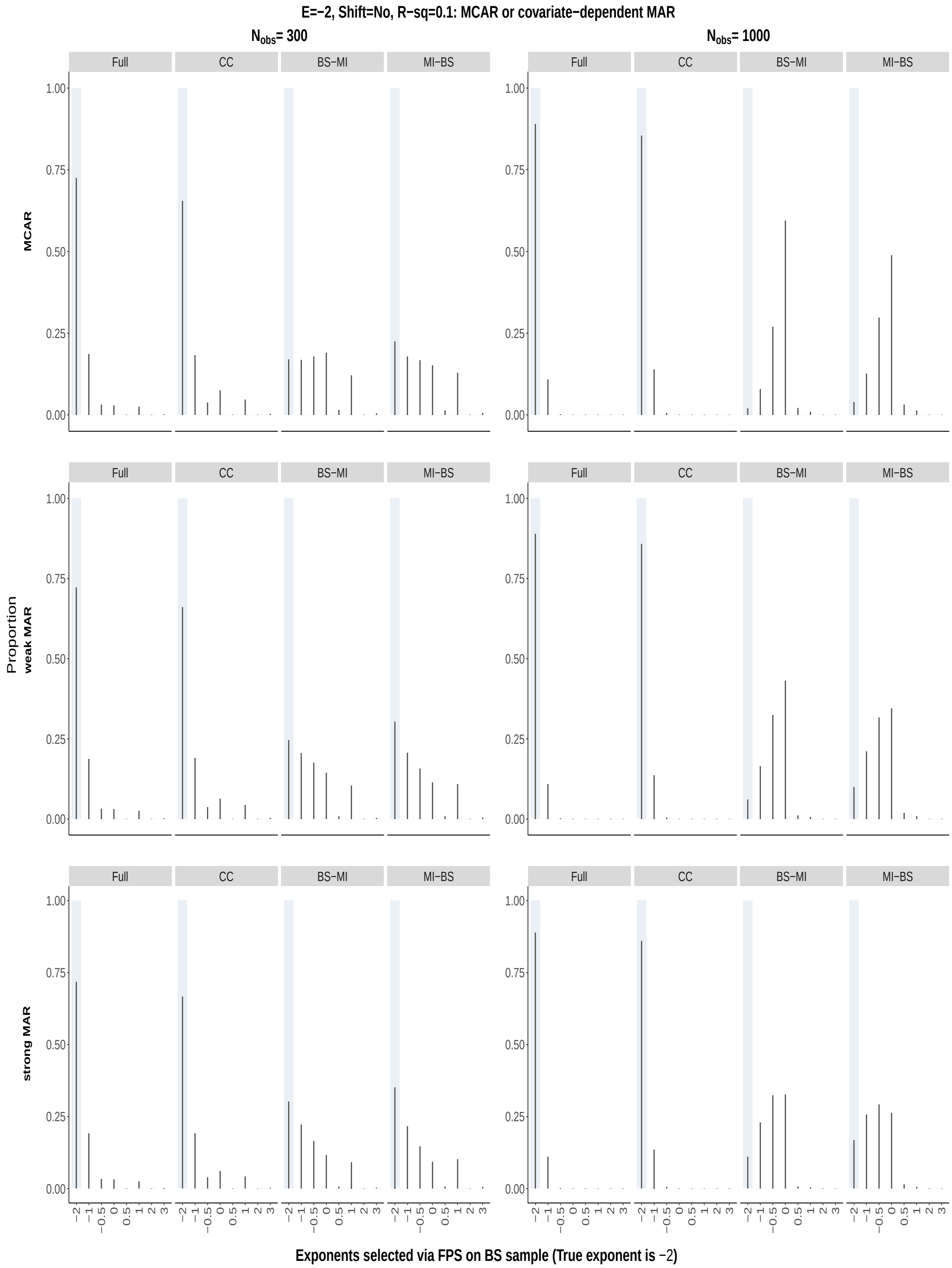

Figure S213: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

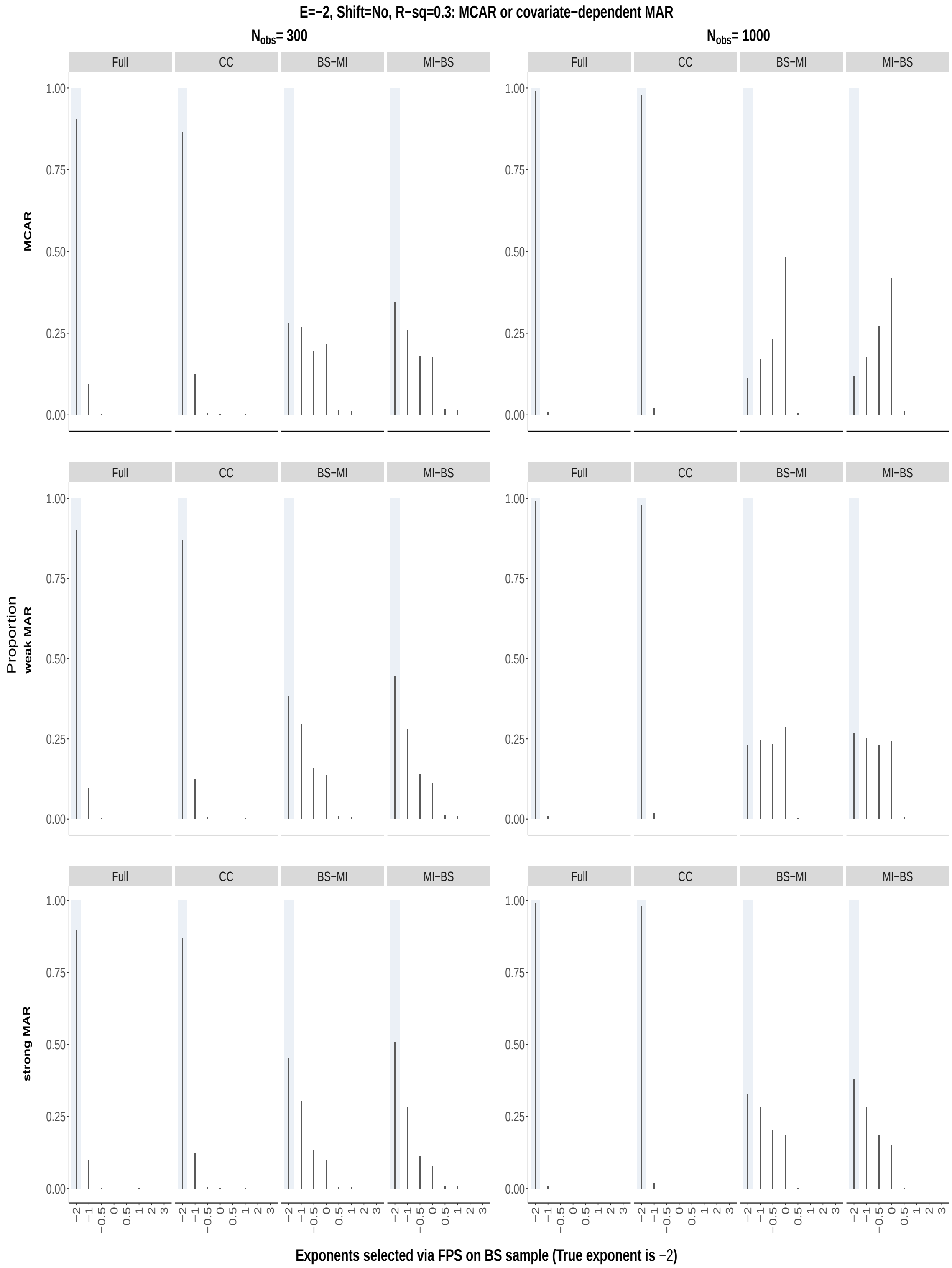

Figure S214: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

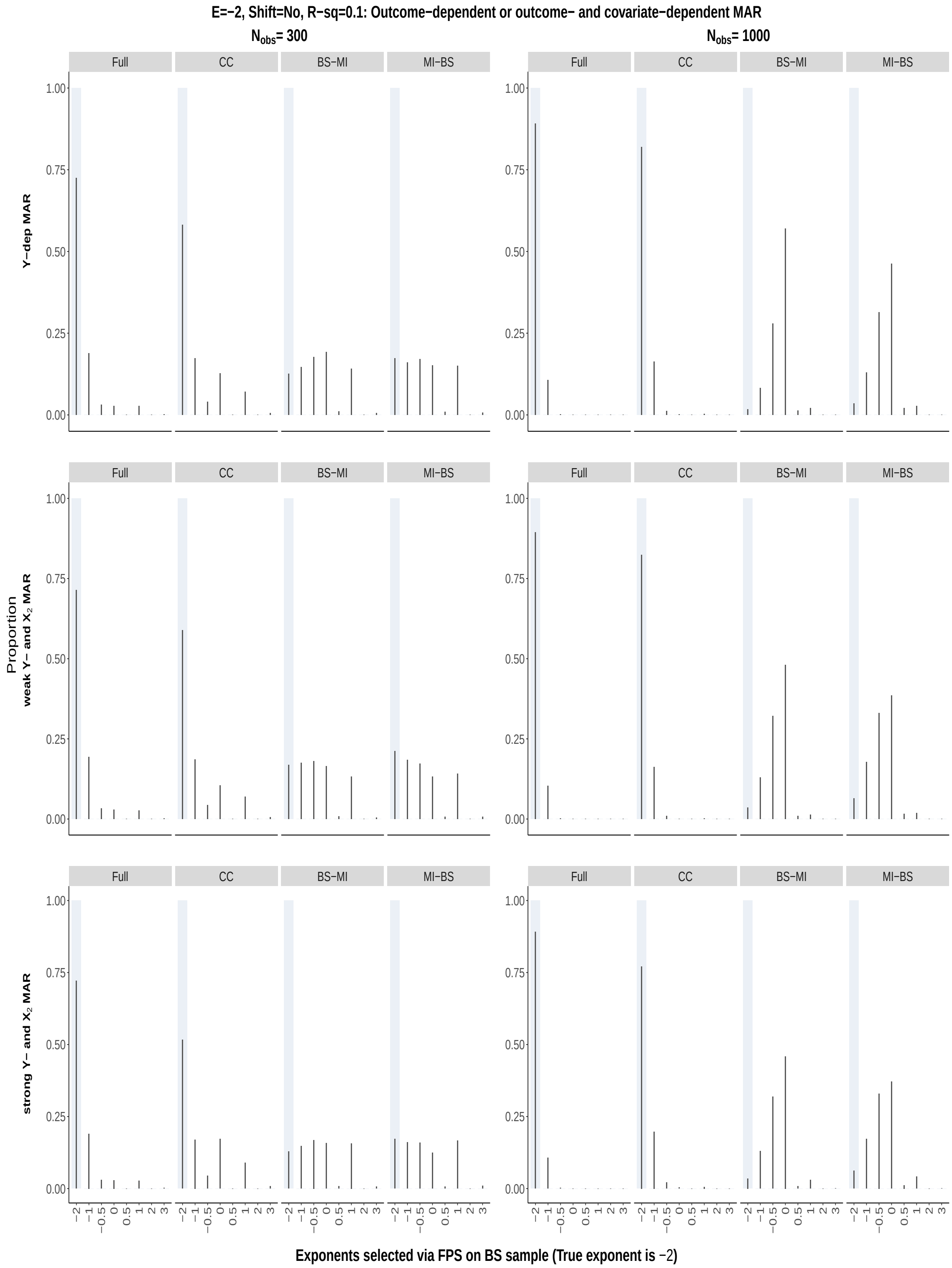

Figure S215: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

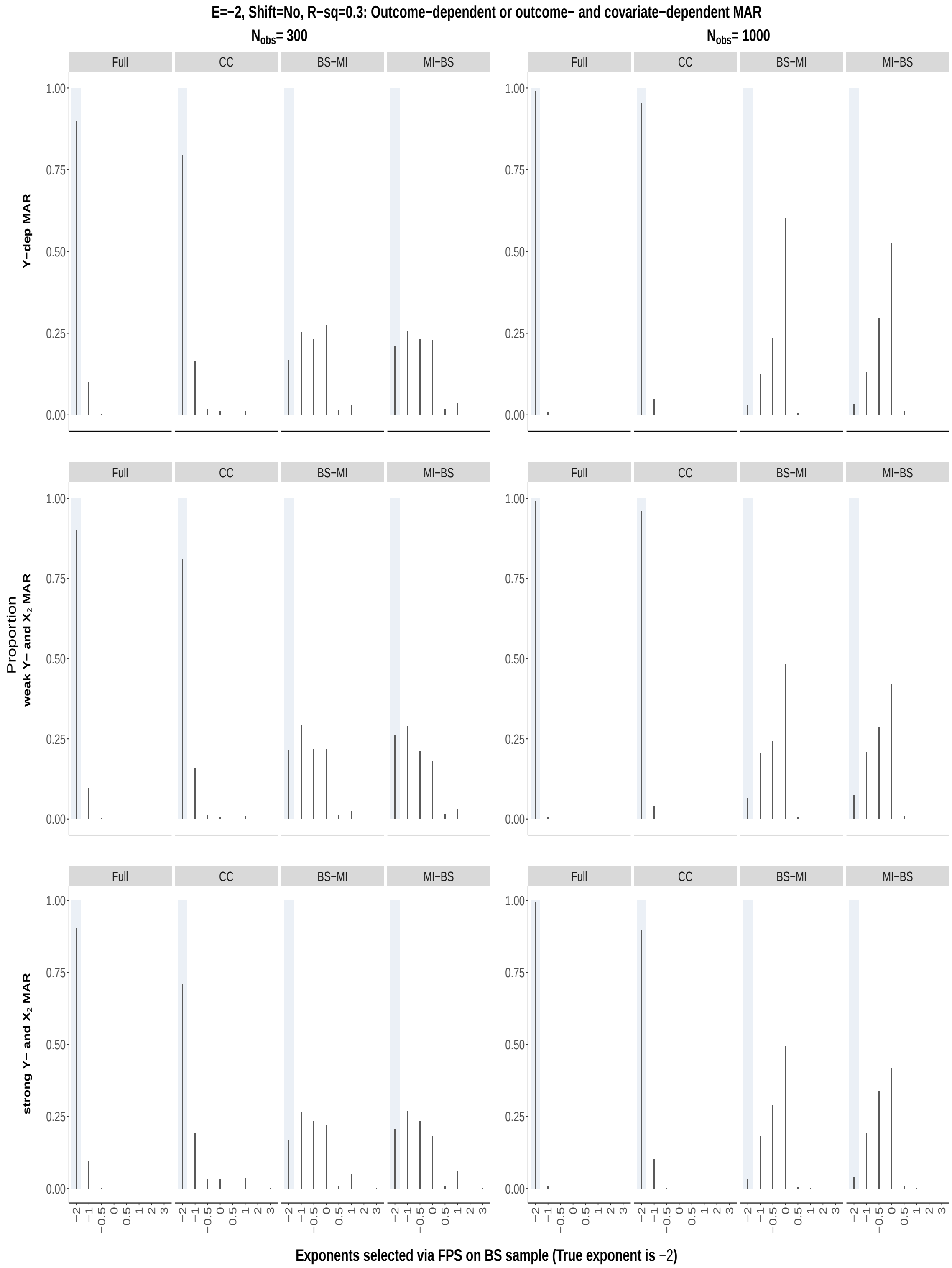

Figure S216: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.15 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 0$, $\alpha_E = 1$ and and an origin-shift has been applied

True exponent is 0

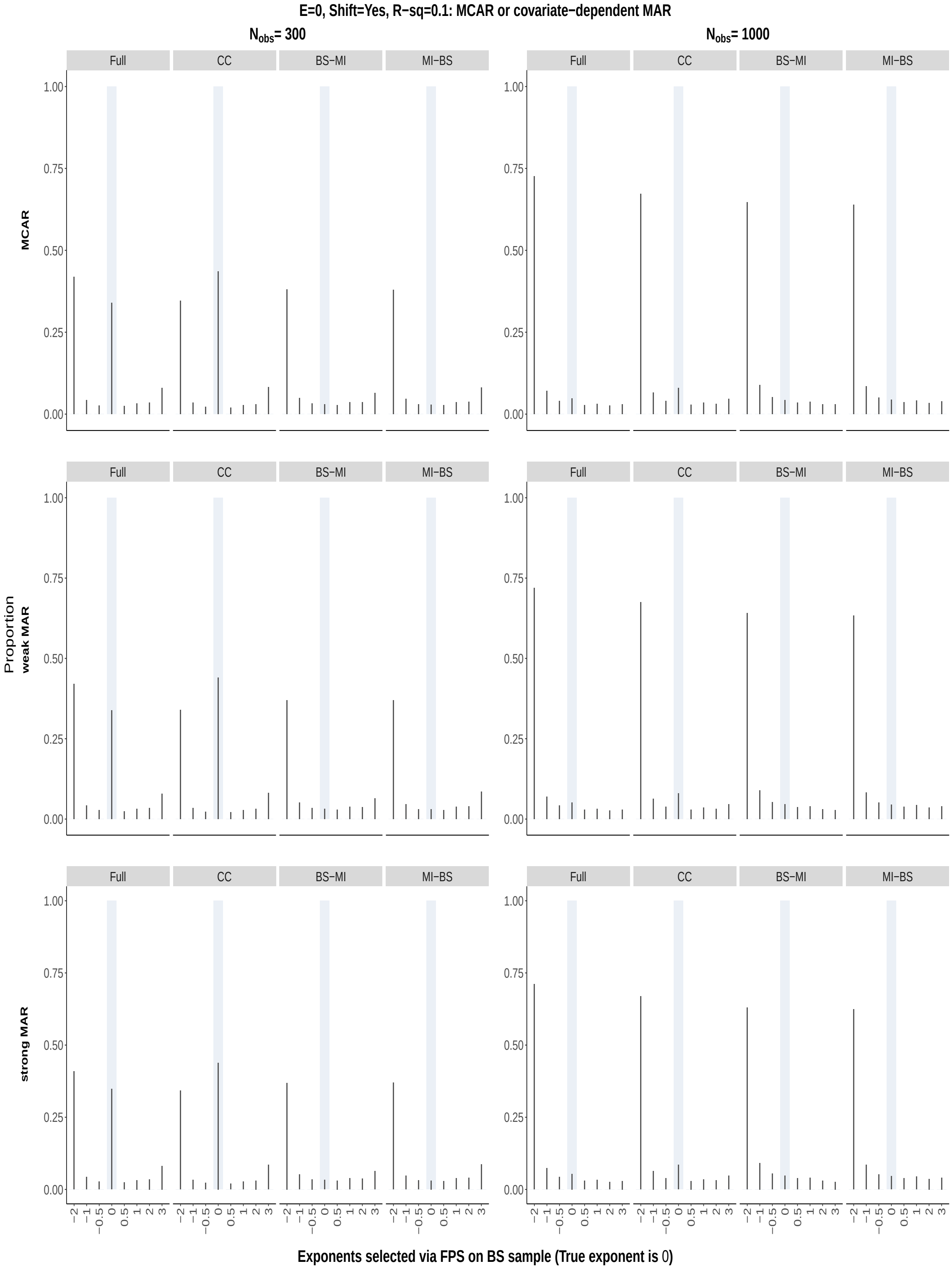

Figure S217: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

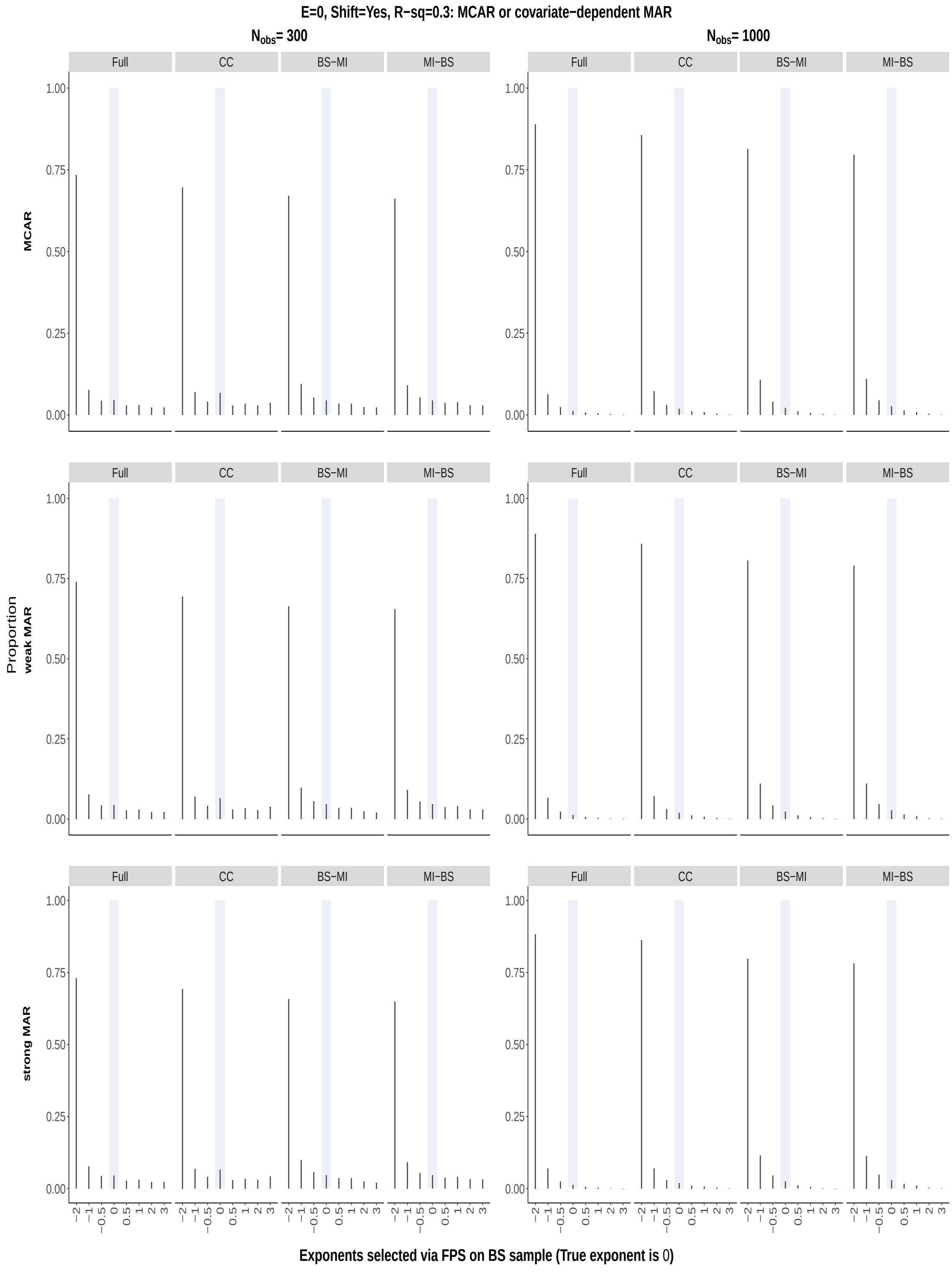

Figure S218: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

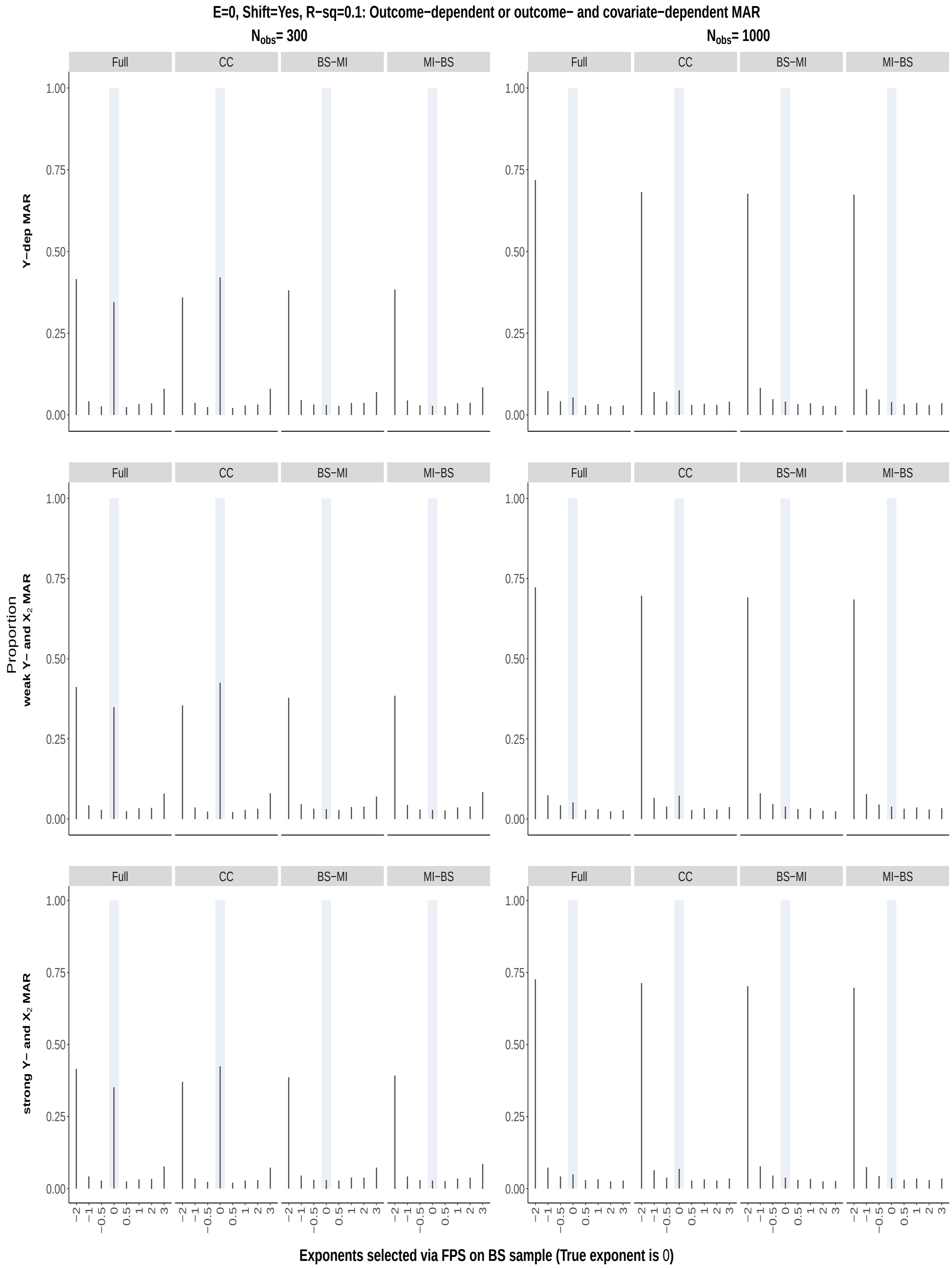

Figure S219: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

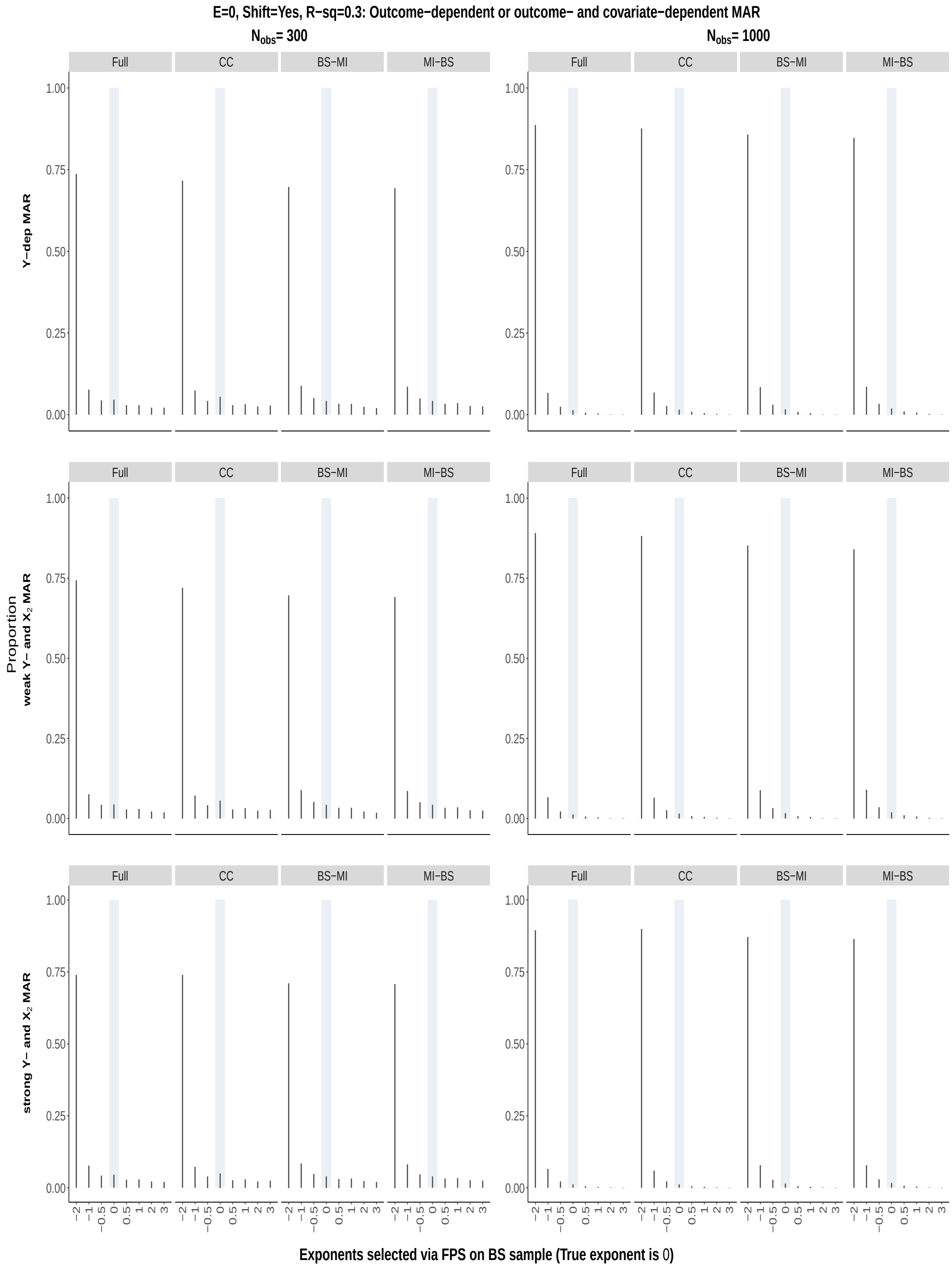

Figure S220: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

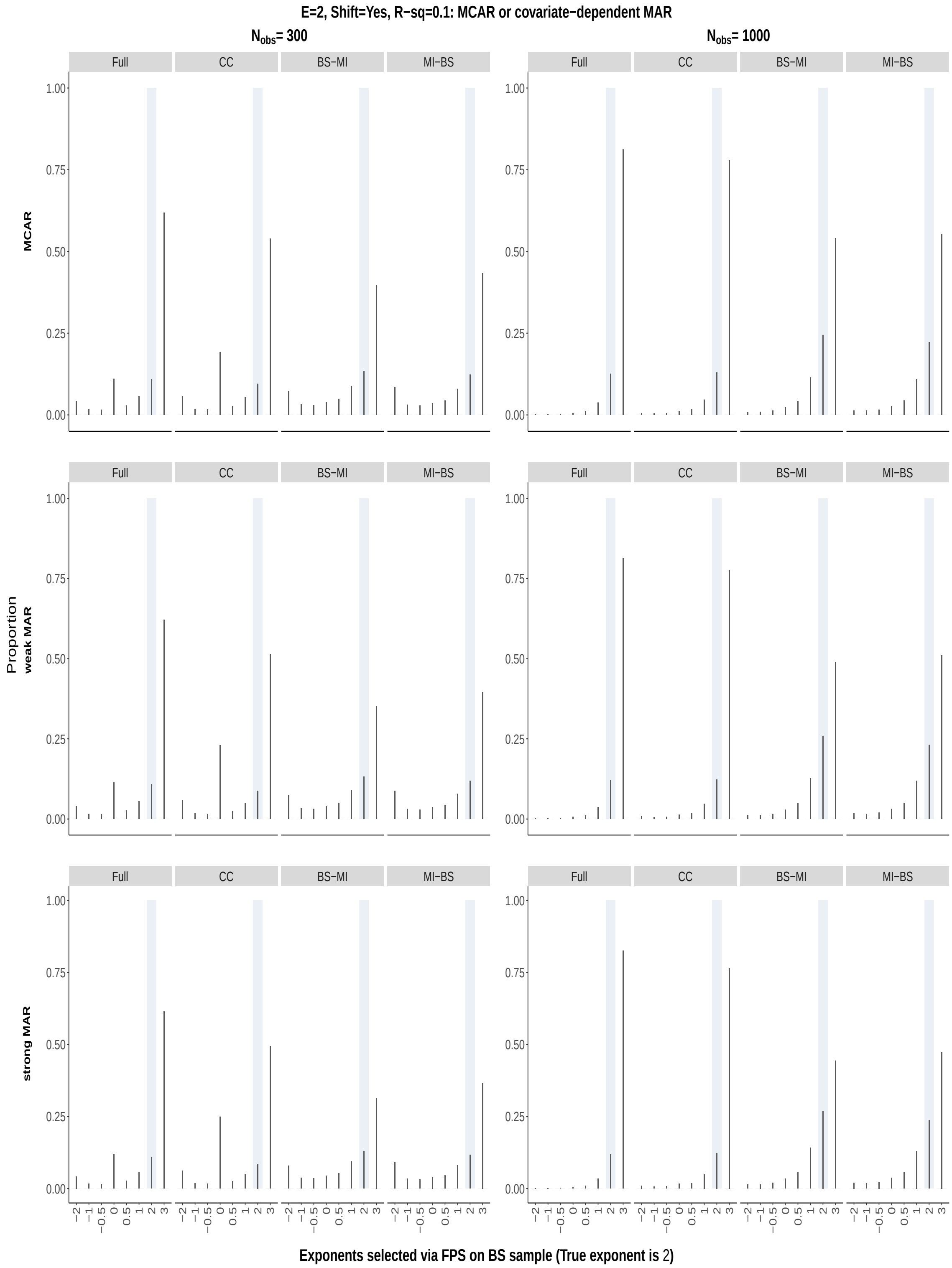

Figure S221: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

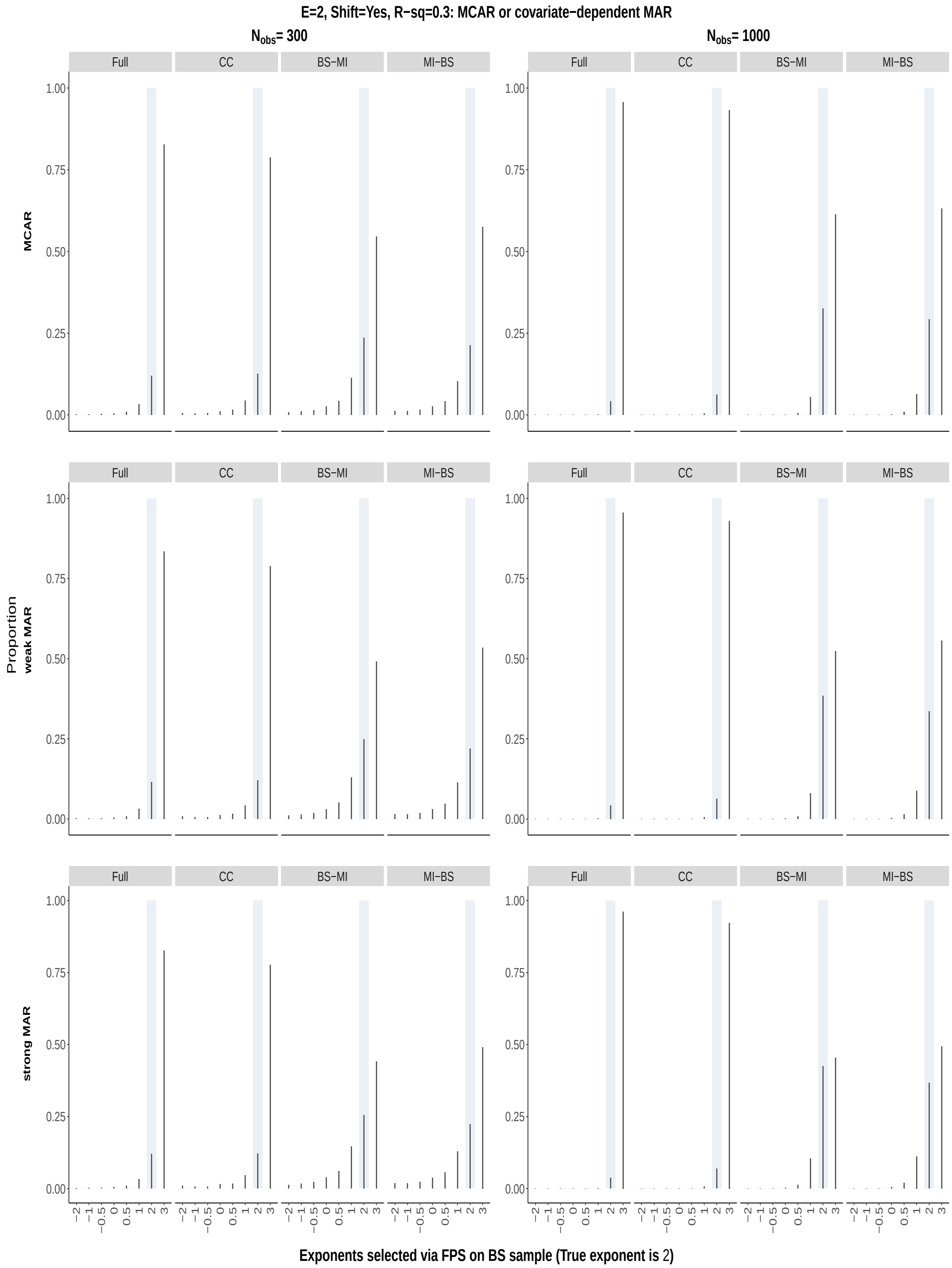

Figure S222: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

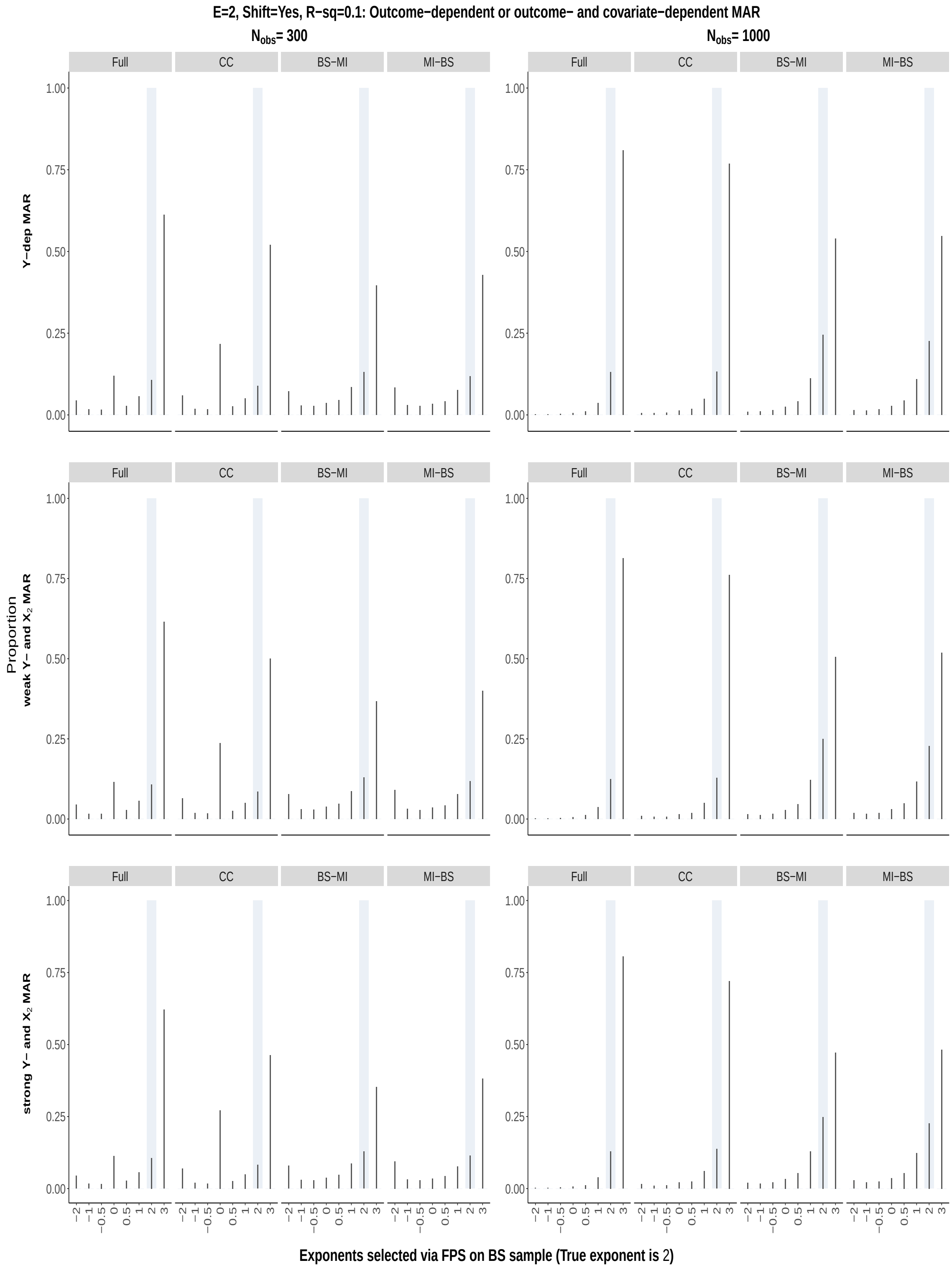

Figure S223: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

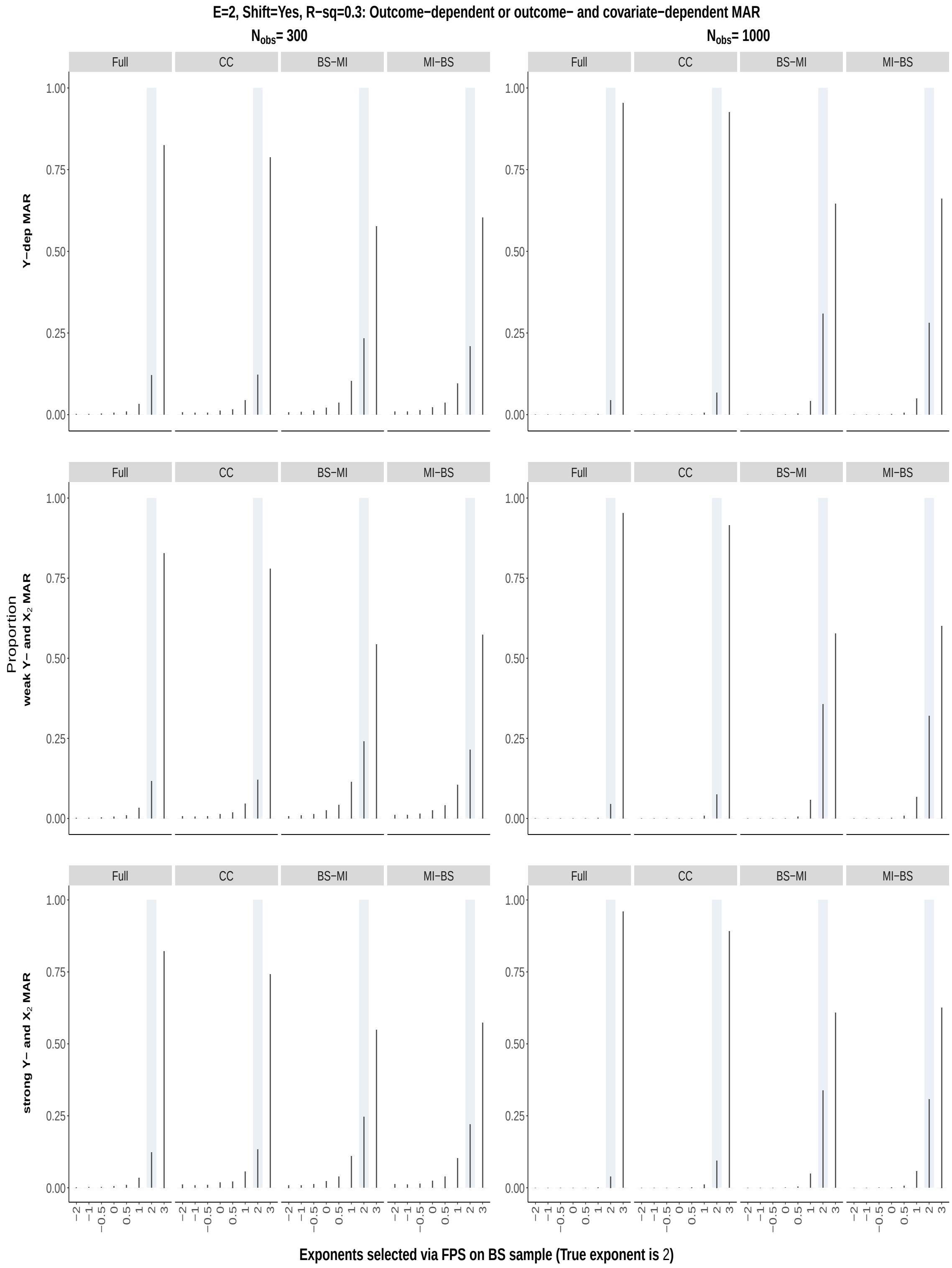

Figure S224: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

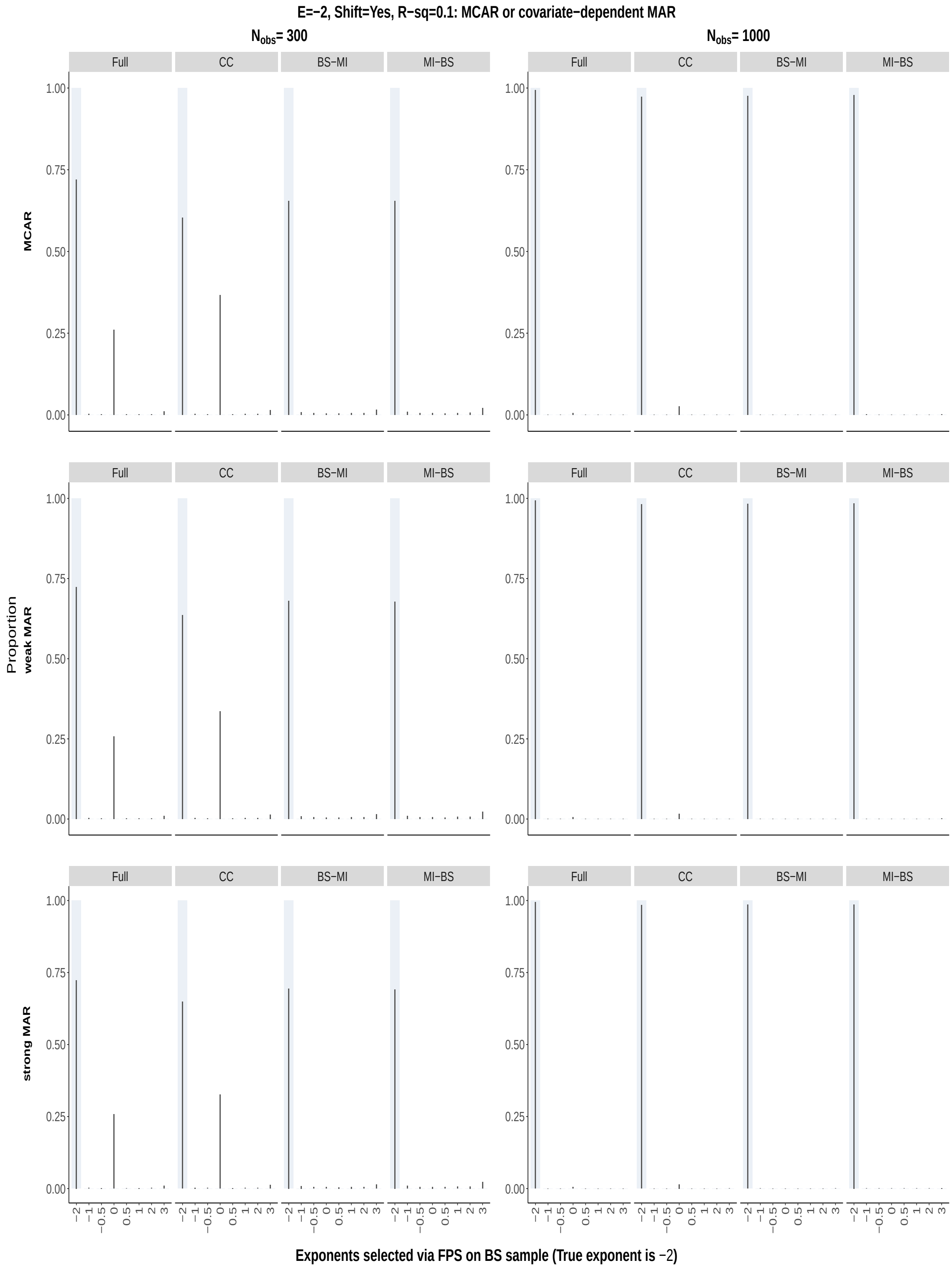

Figure S225: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

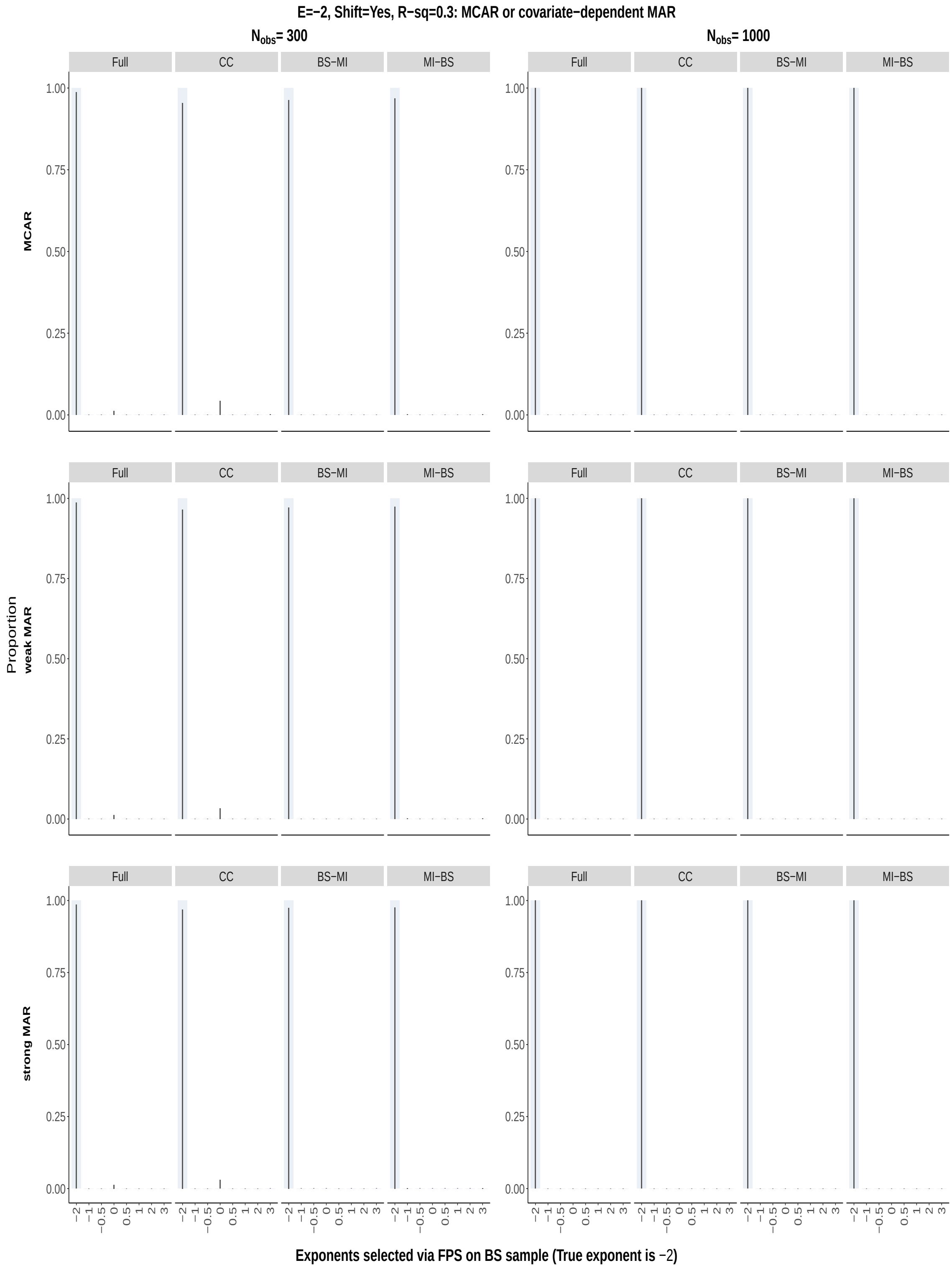

Figure S226: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

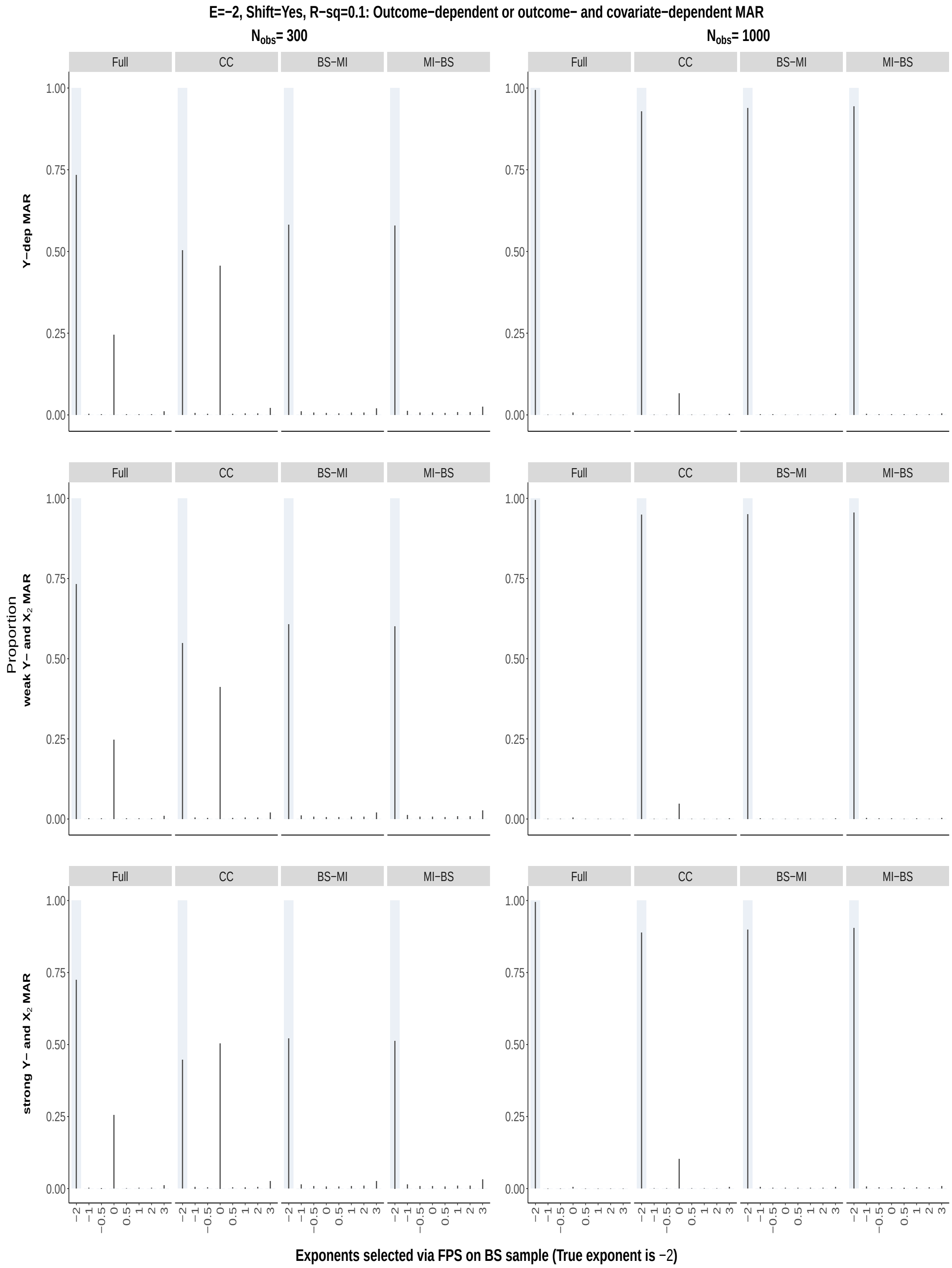

Figure S227: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

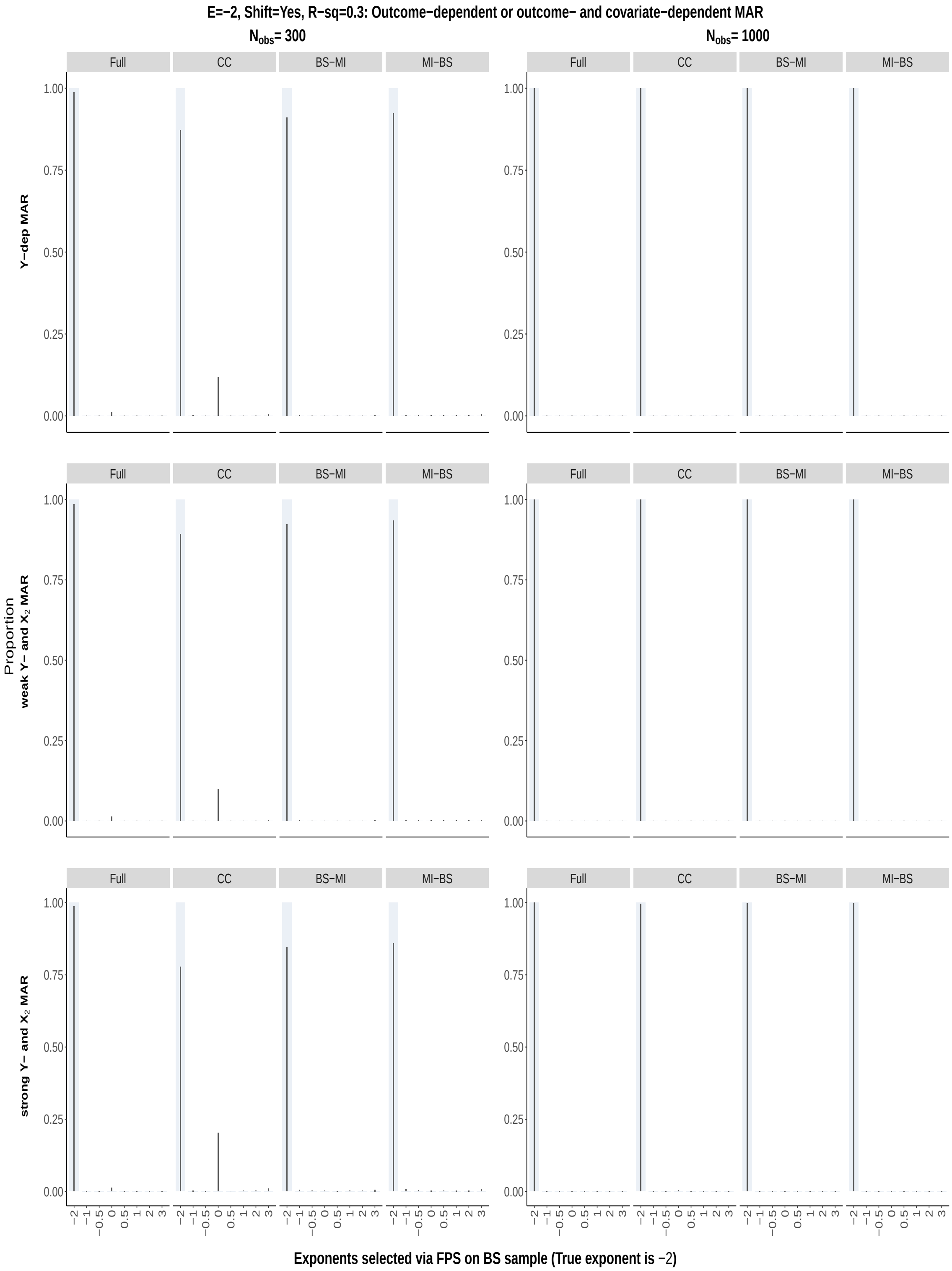

Figure S228: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.16 The 0.632 bootstrap, exponents selected in the bootstrap samples: $\beta_2 = 0$, $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

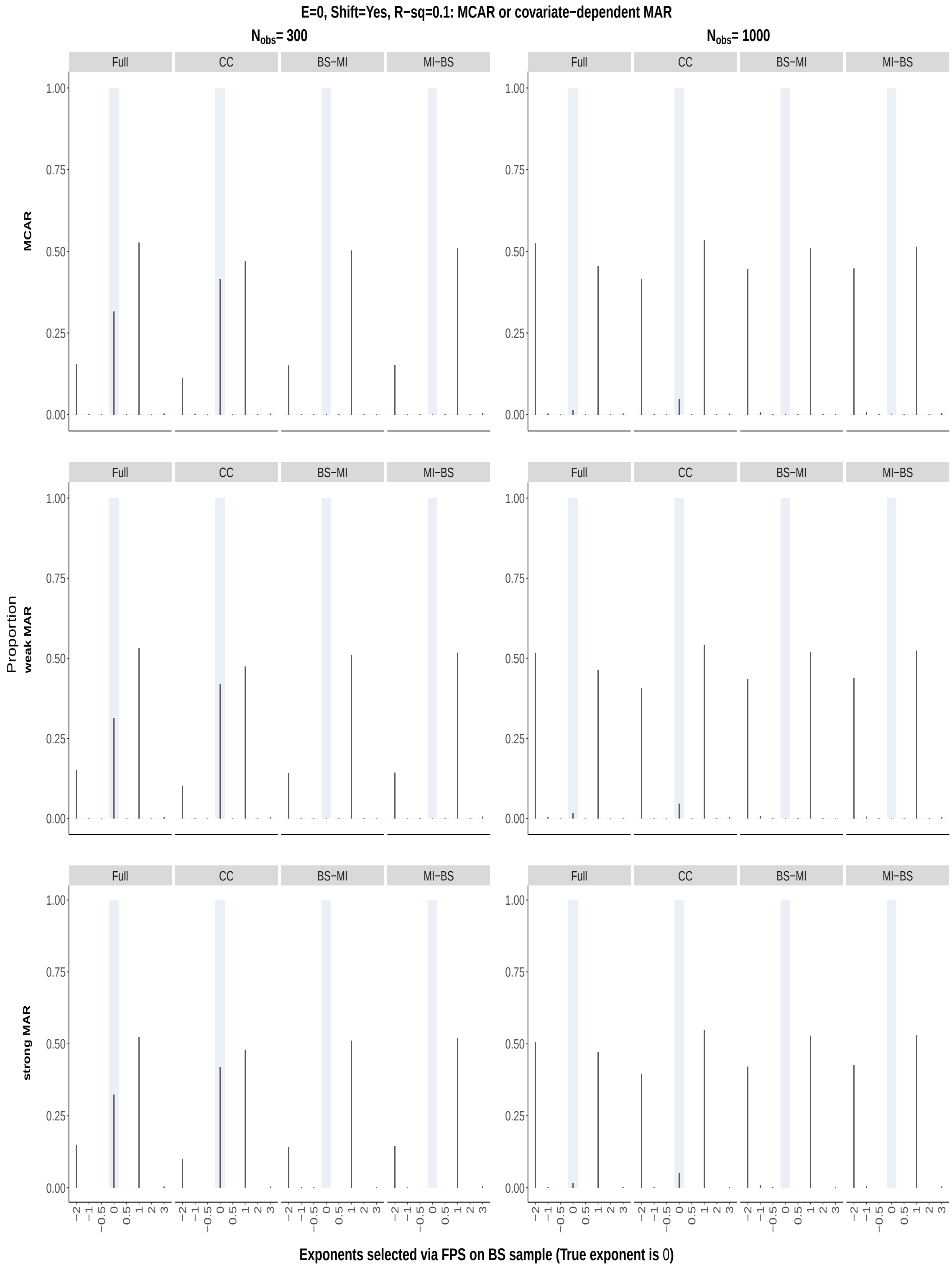

Figure S229: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

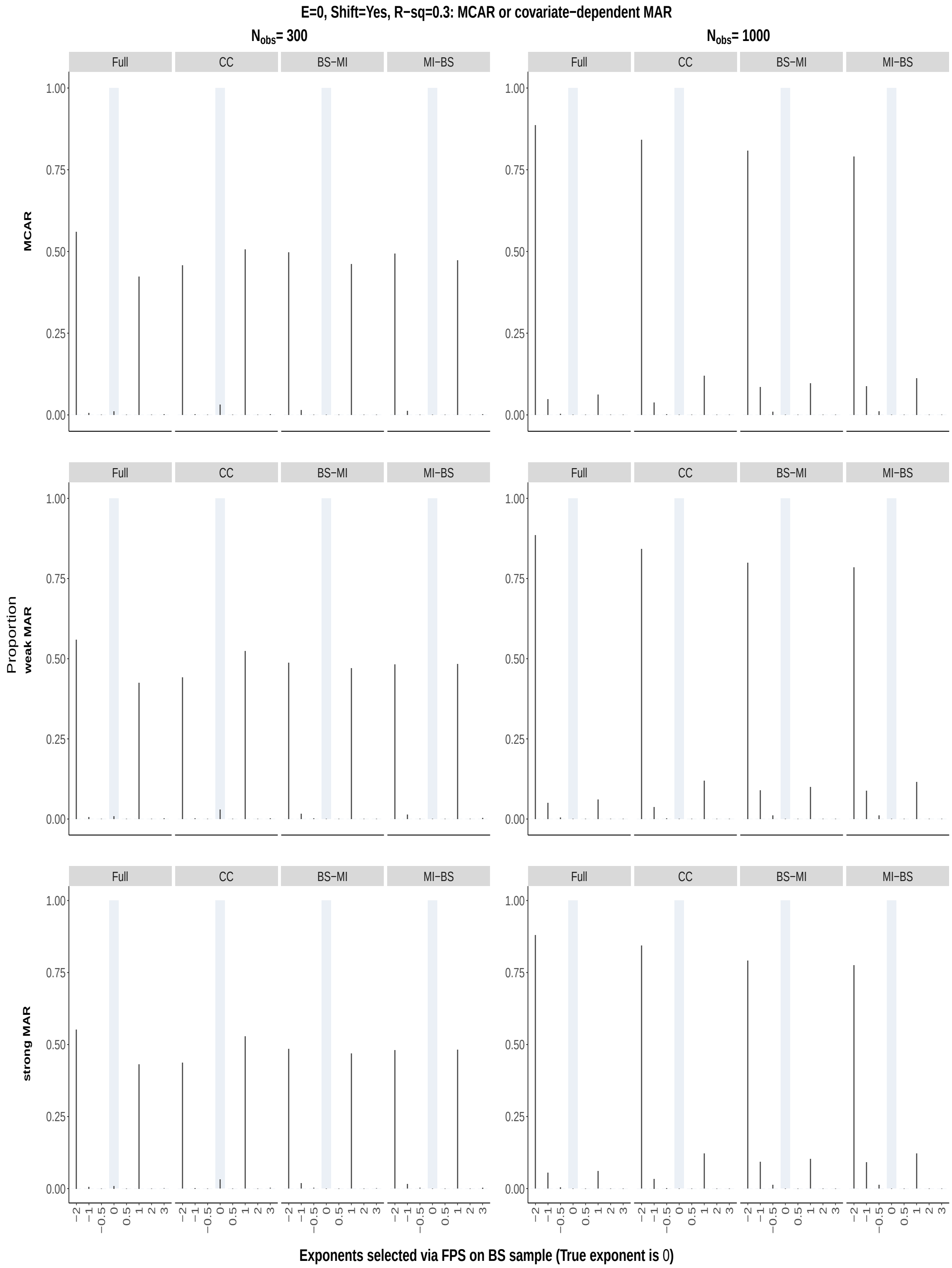

Figure S230: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

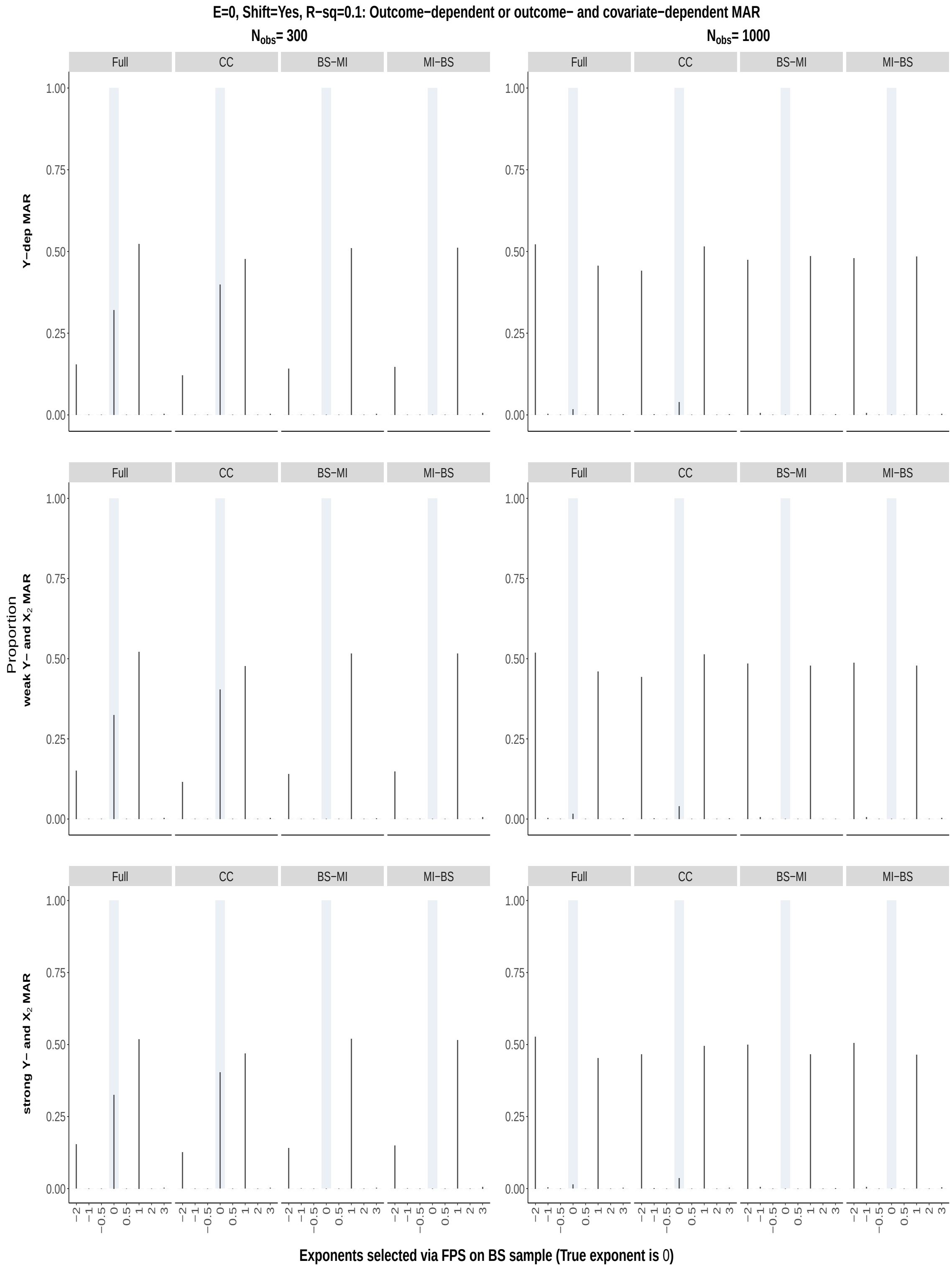

Figure S231: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

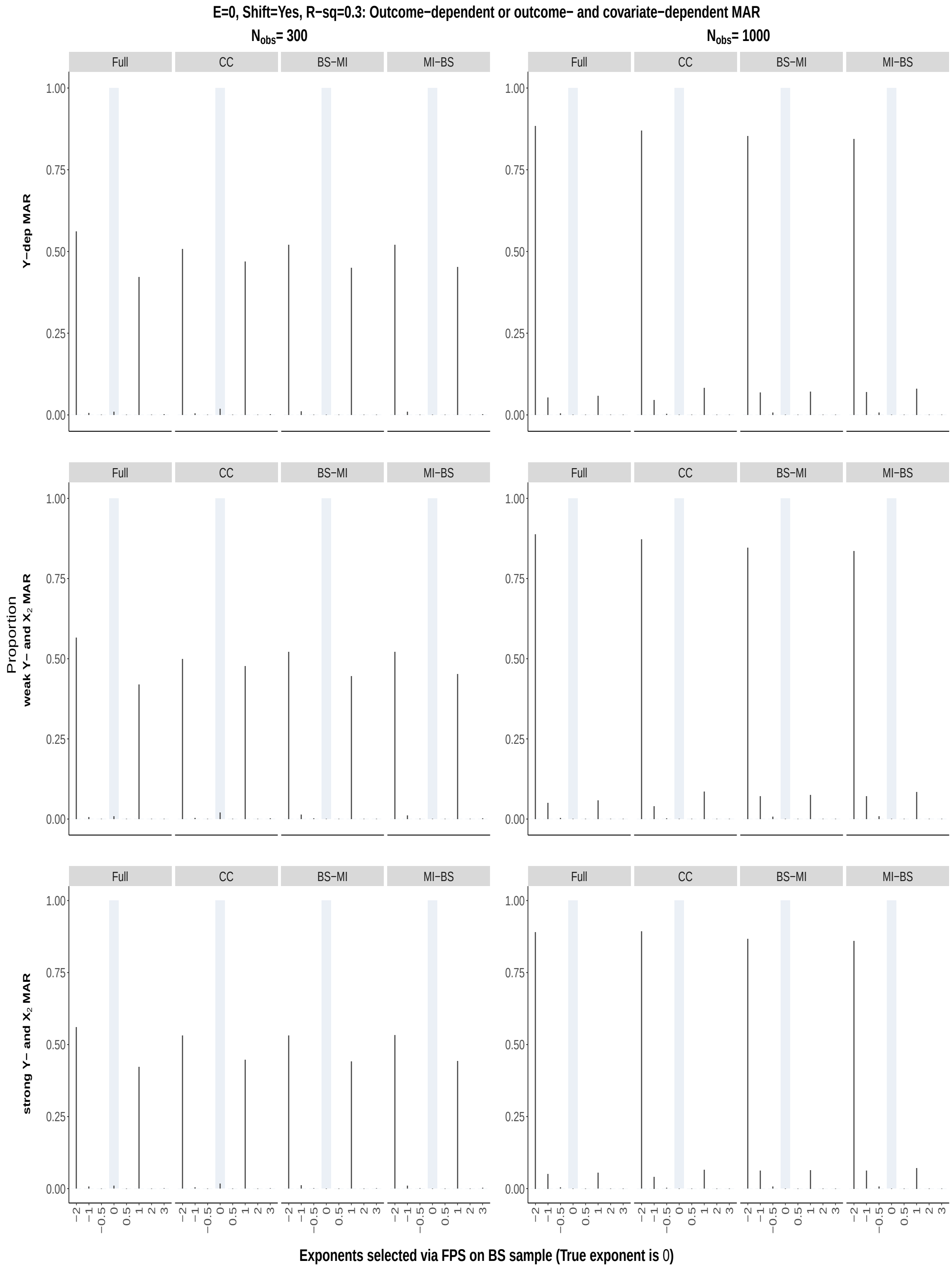

Figure S232: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

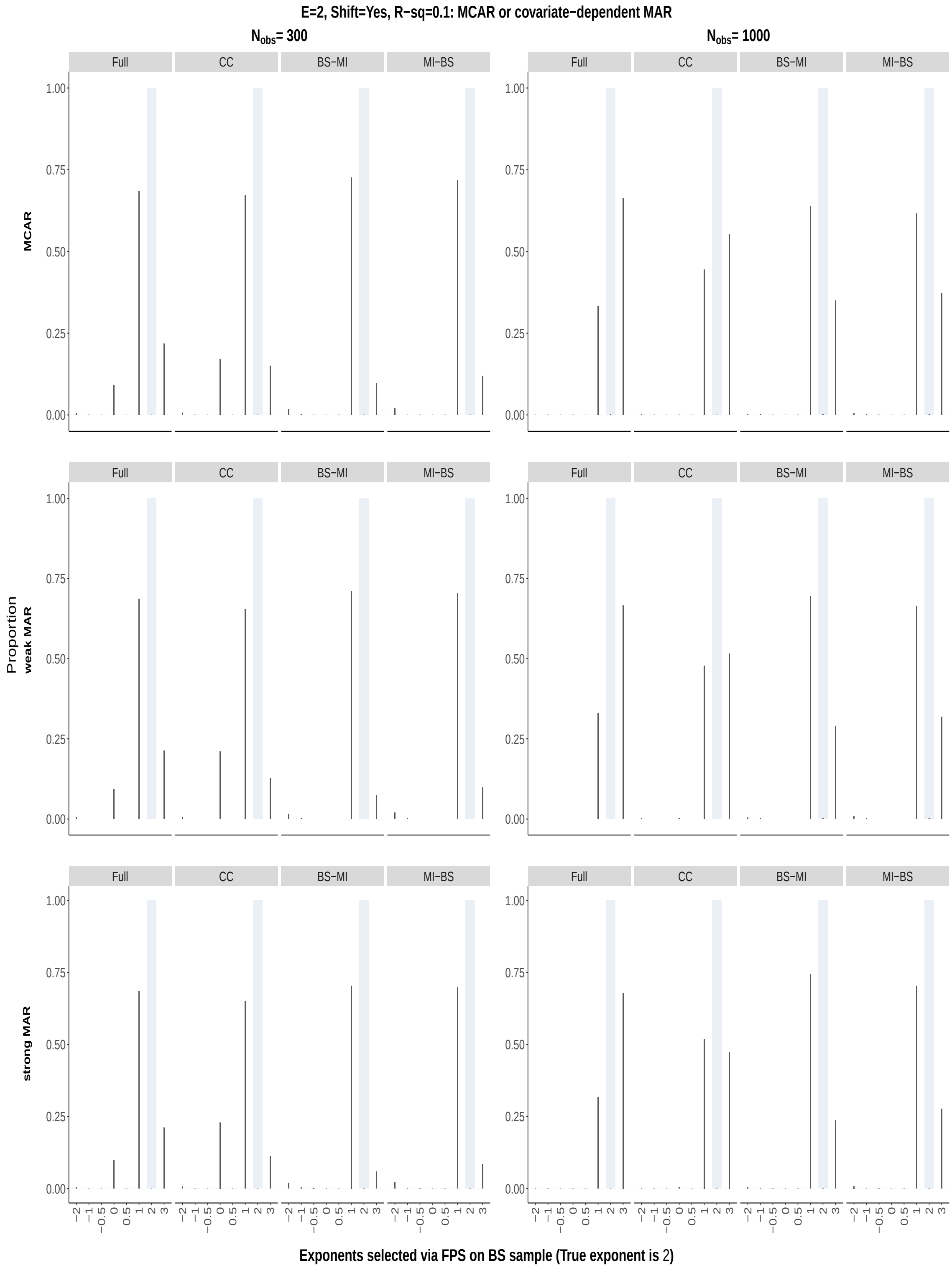

Figure S233: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

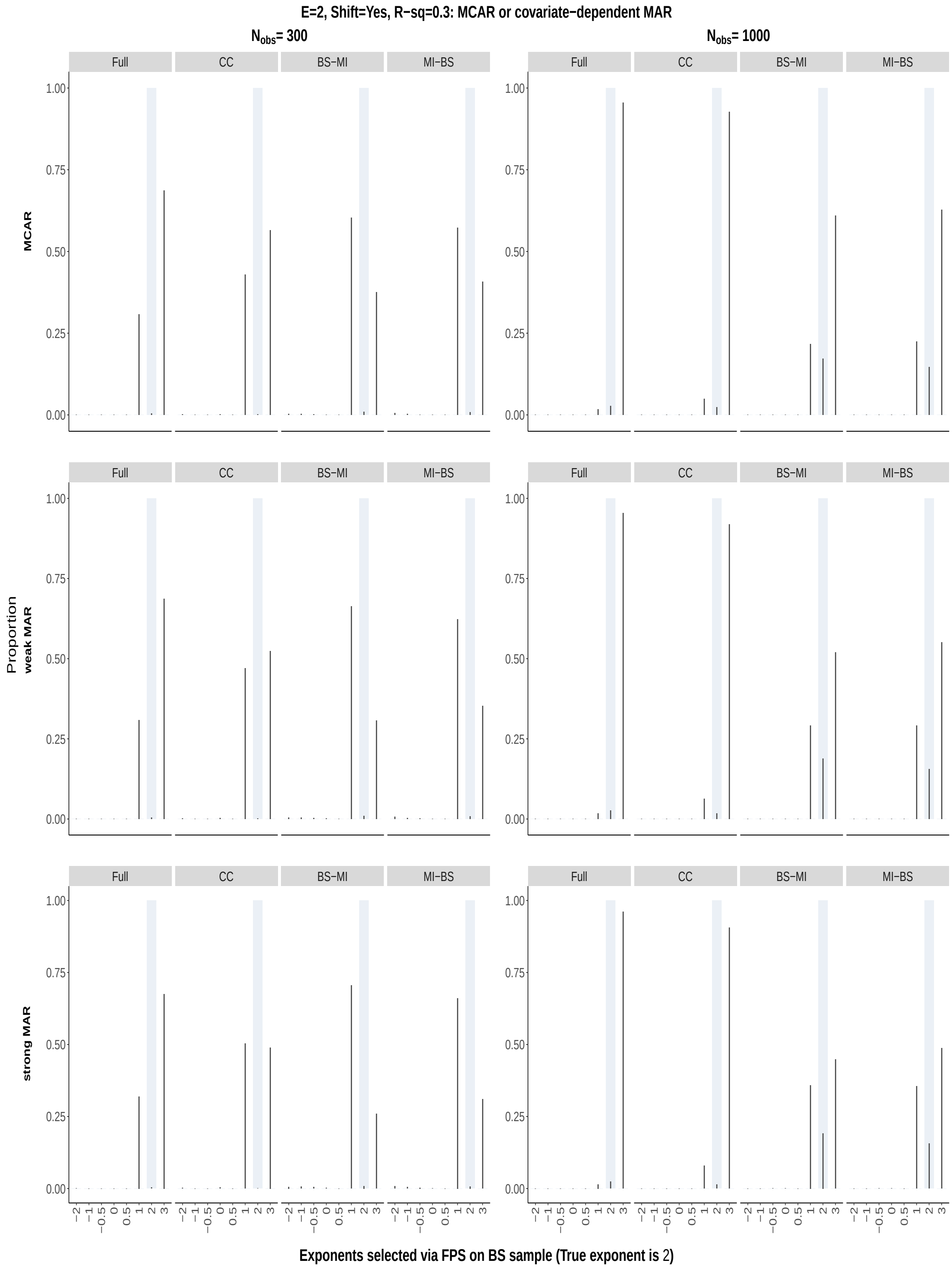

Figure S234: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

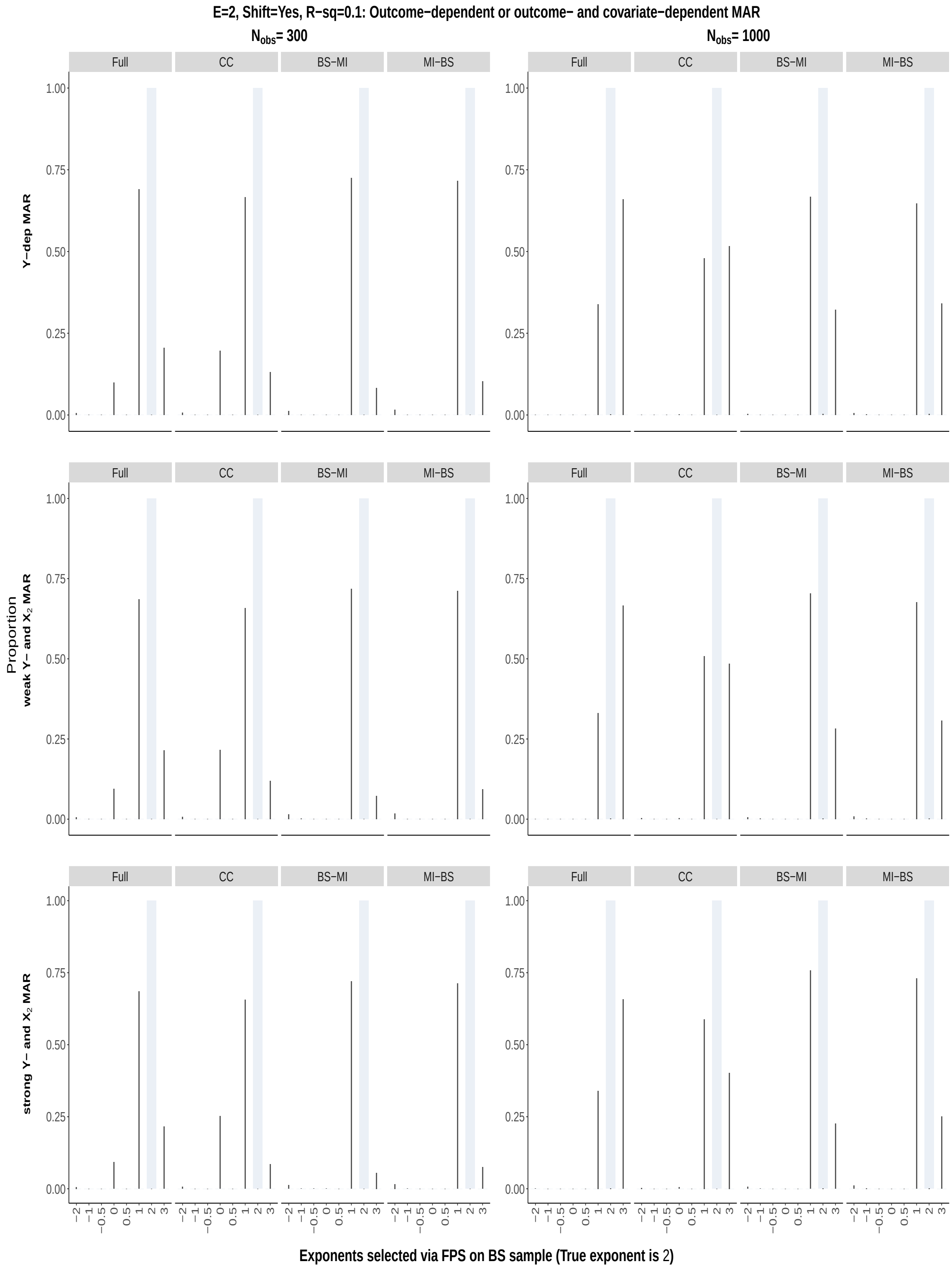

Figure S235: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

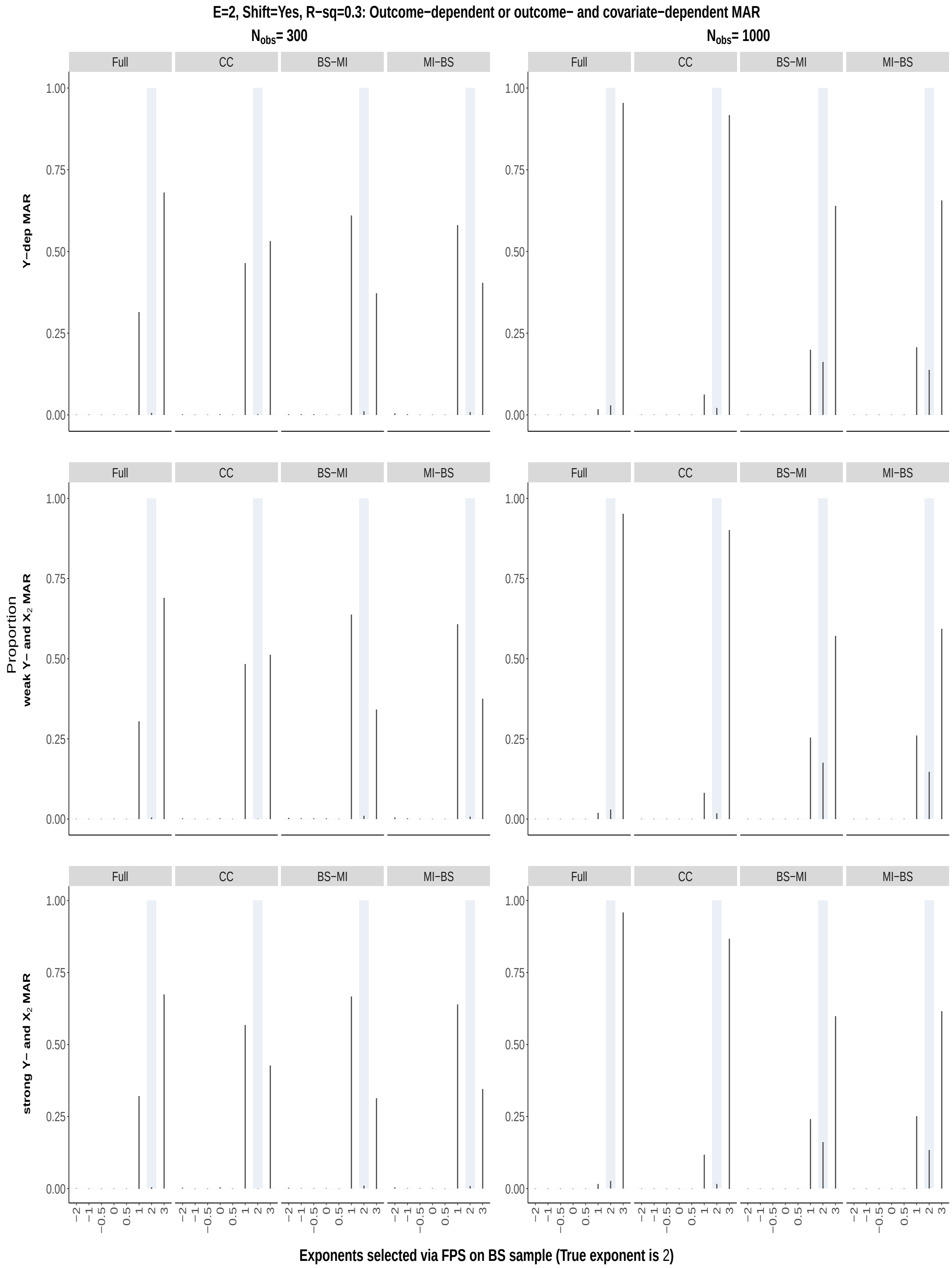

Figure S236: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

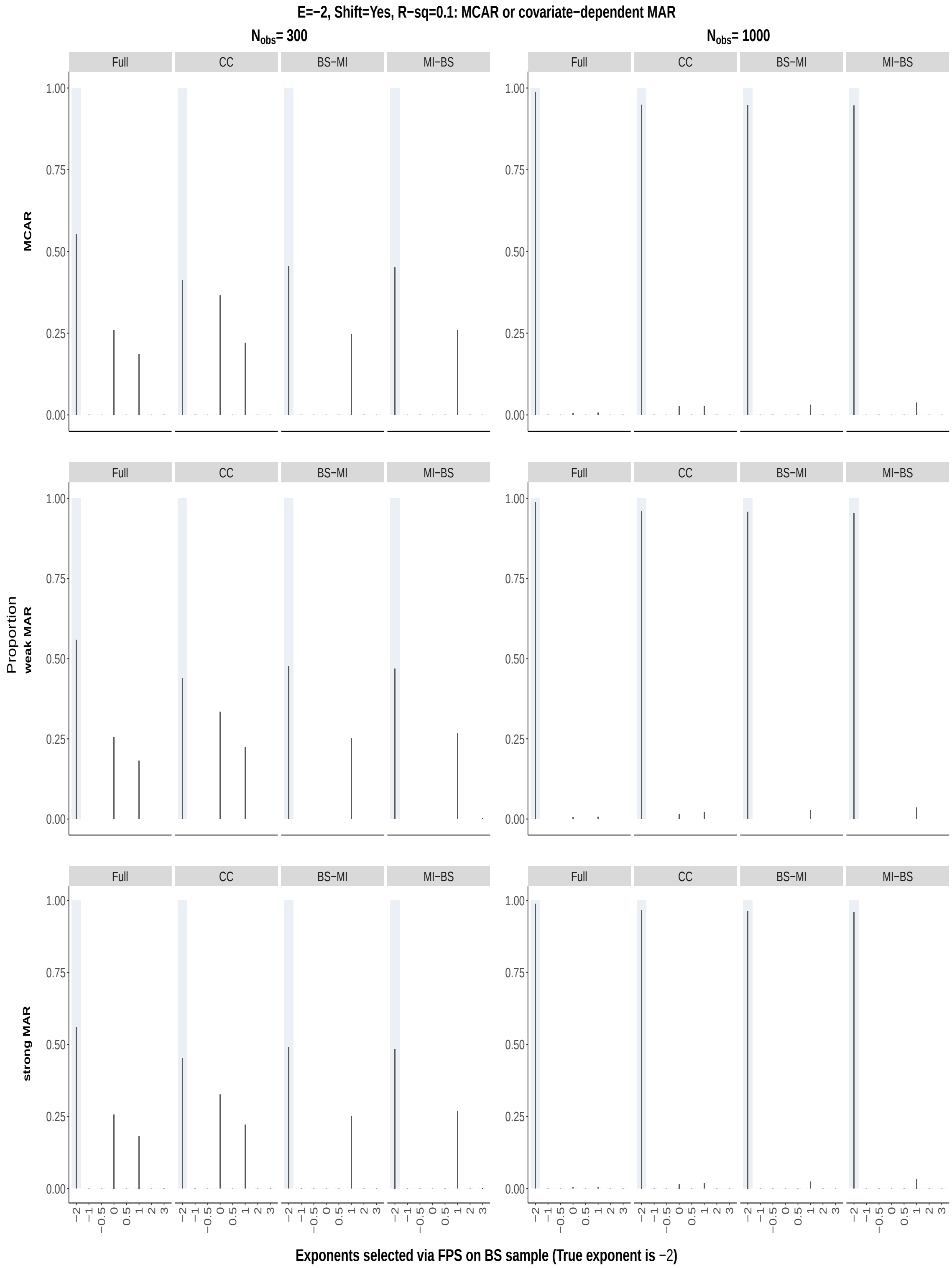

Figure S237: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

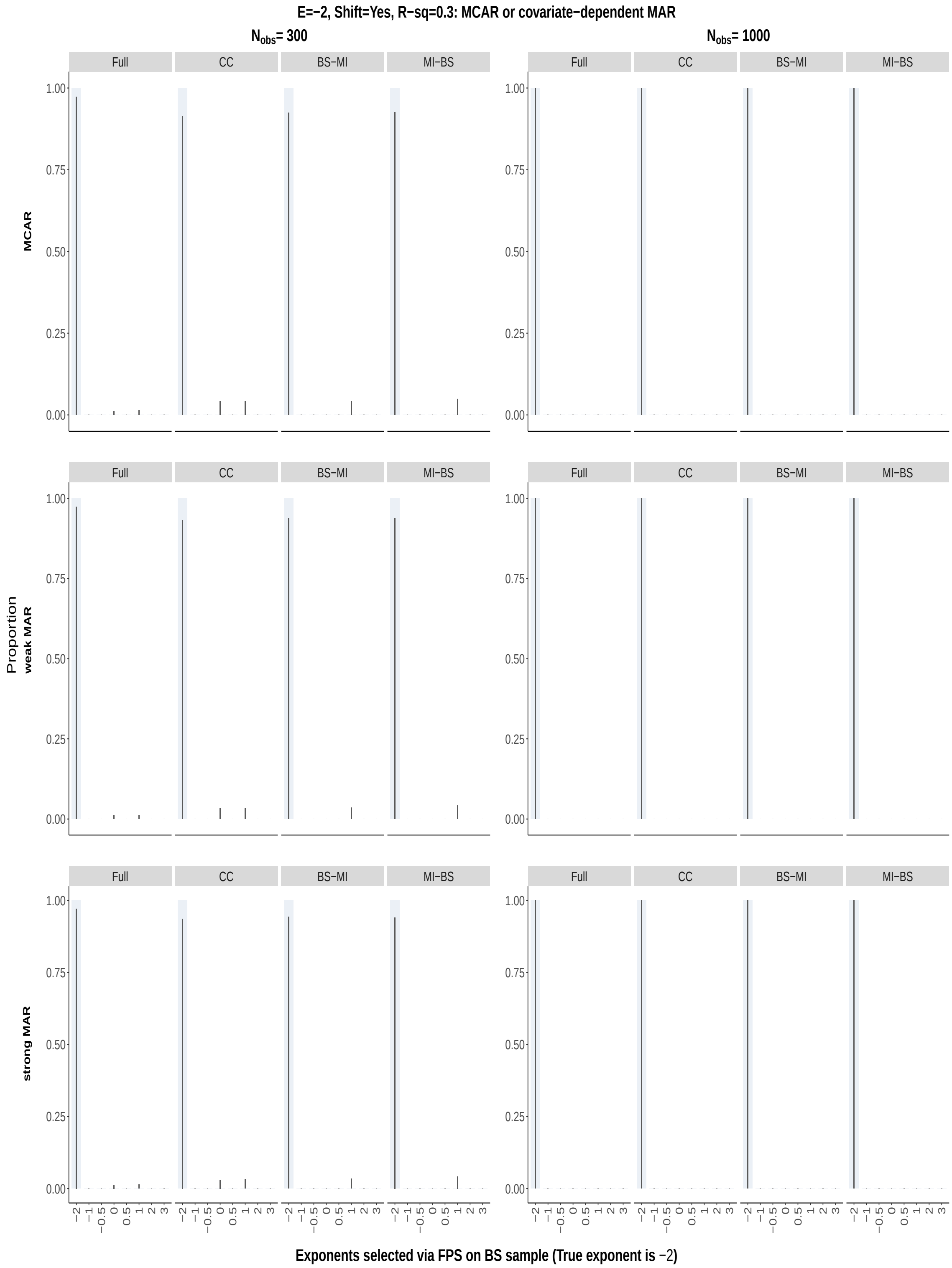

Figure S238: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

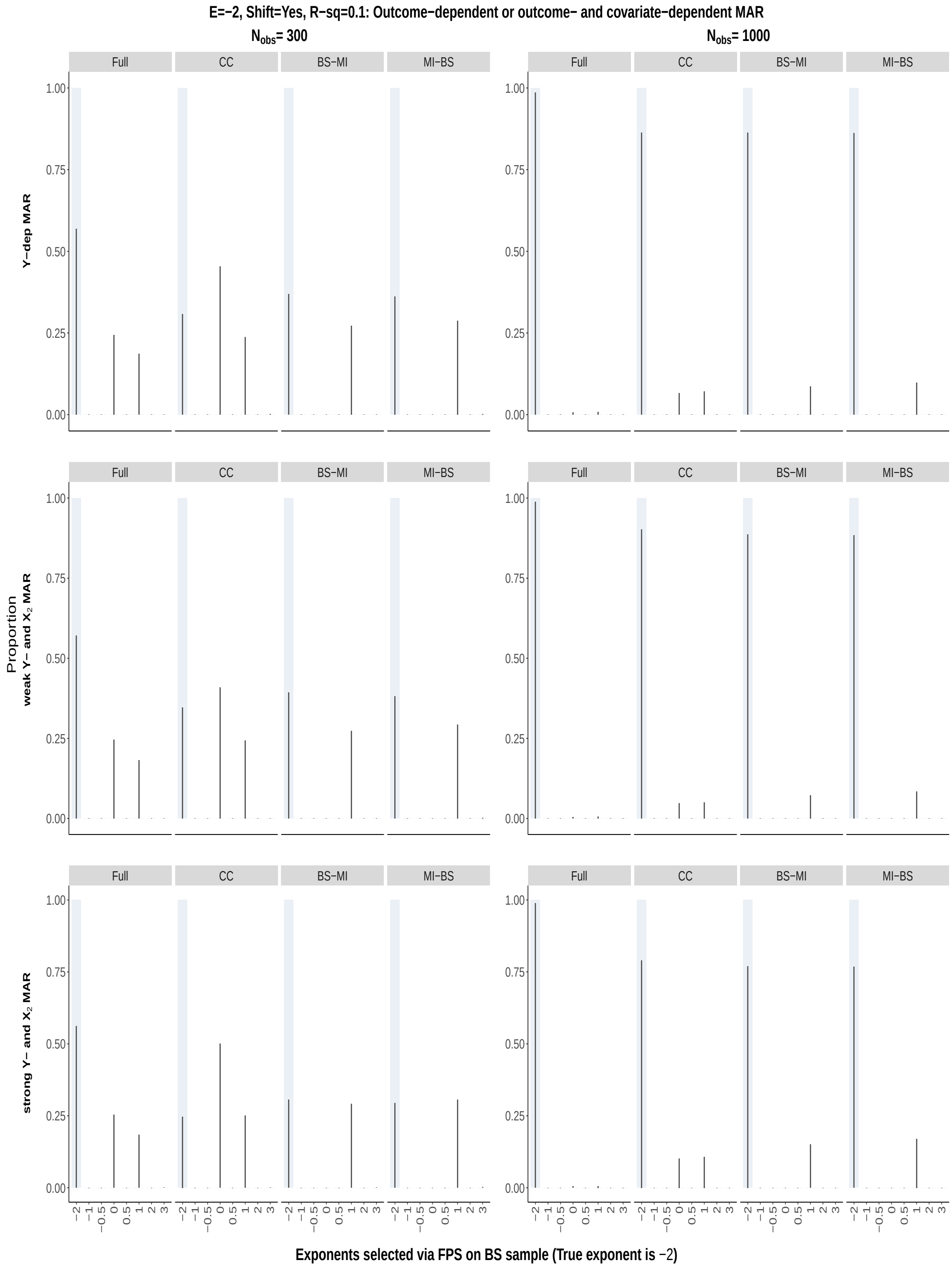

Figure S239: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

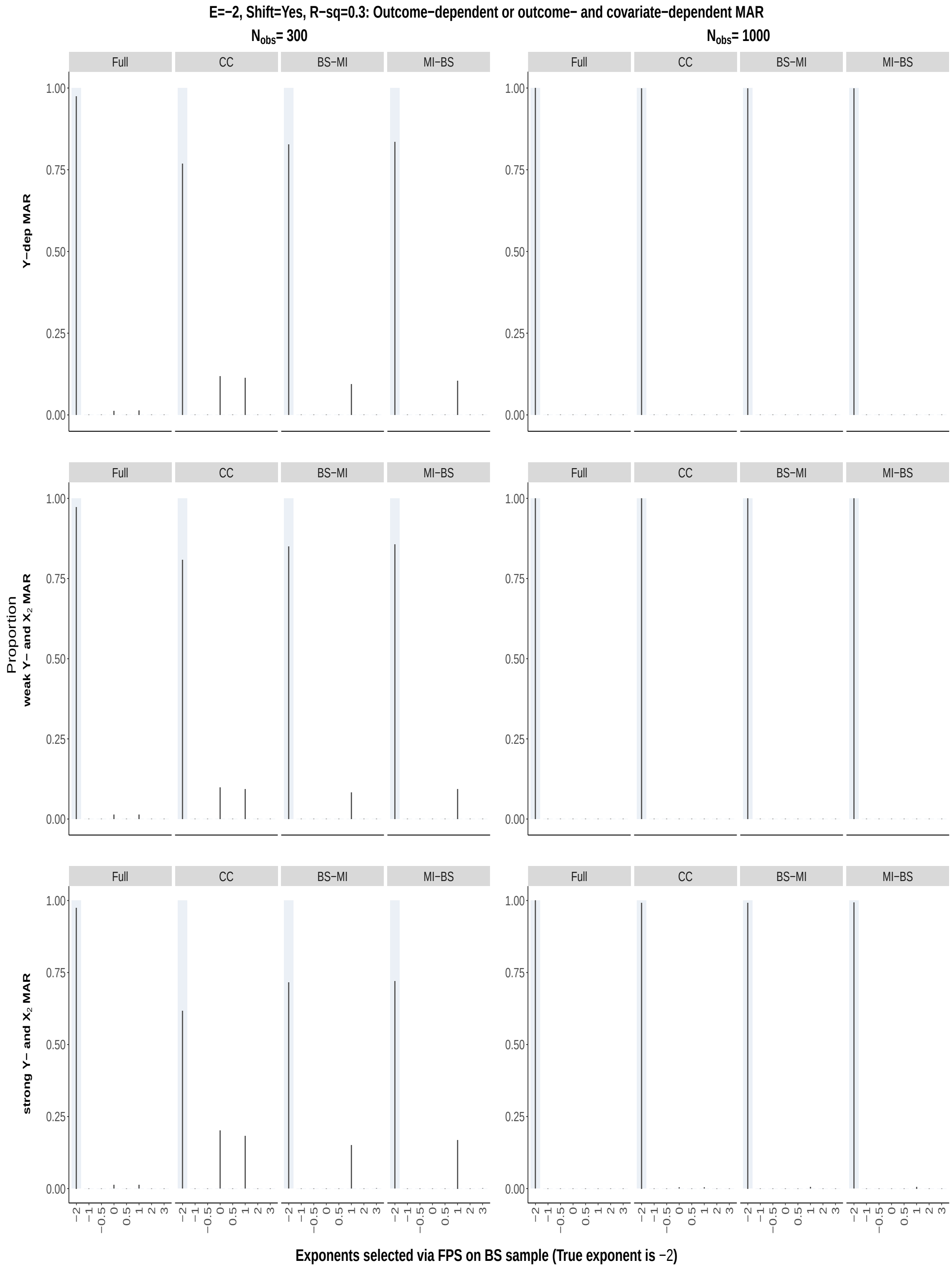

Figure S240: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.17 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 1$, $\alpha_E = 1$ and no origin-shift

True exponent is 0

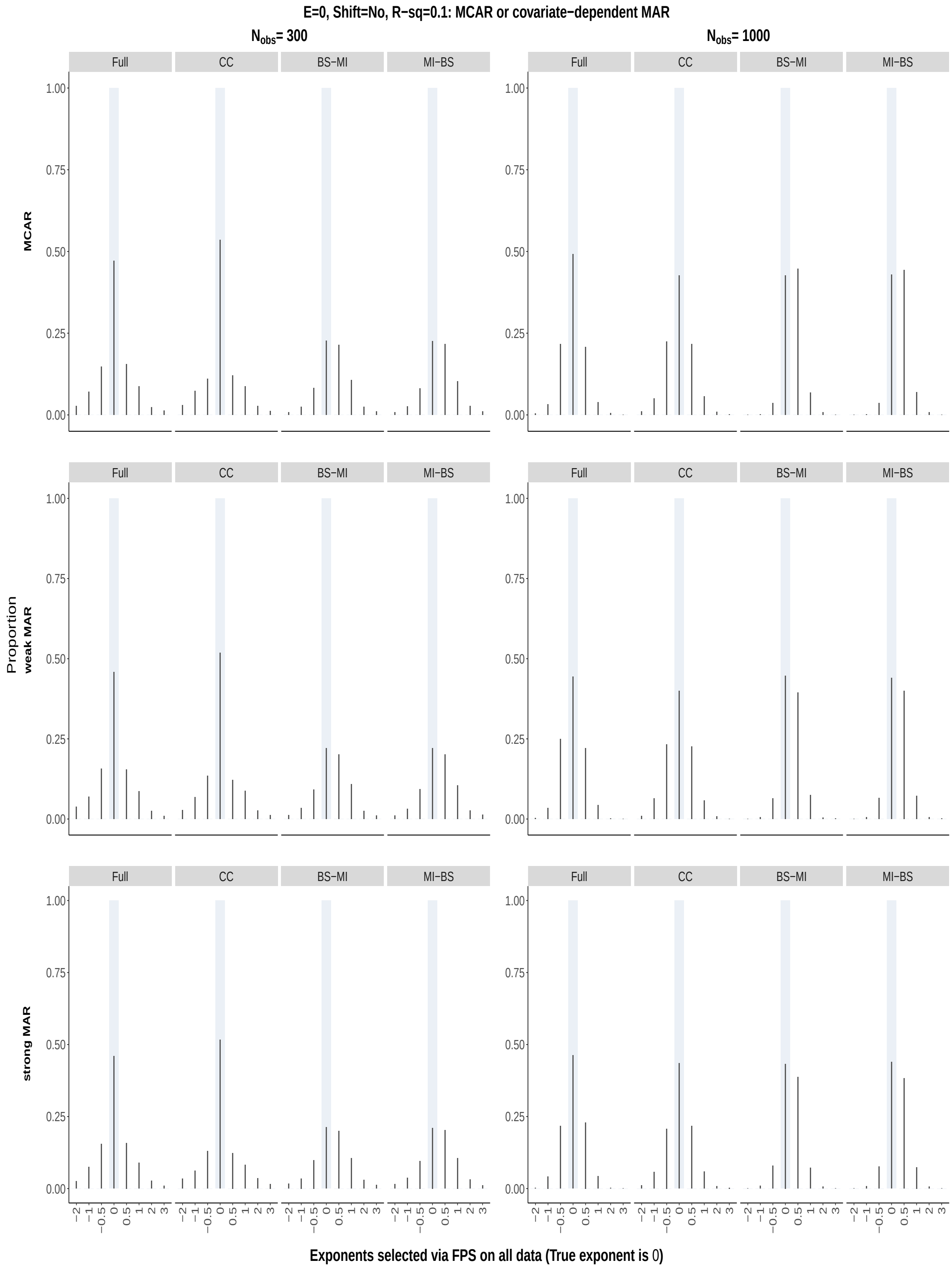

Figure S241: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

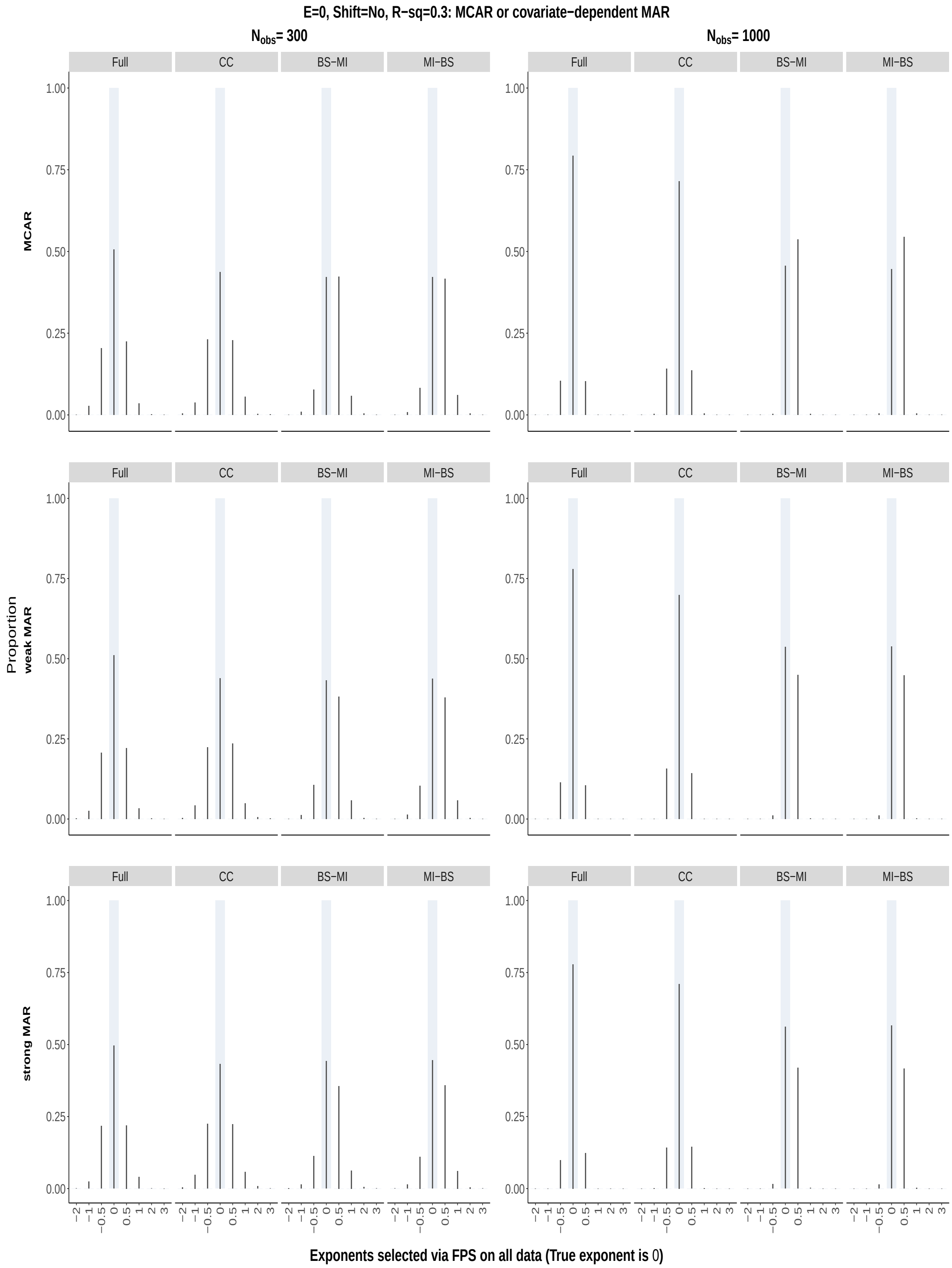

Figure S242: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

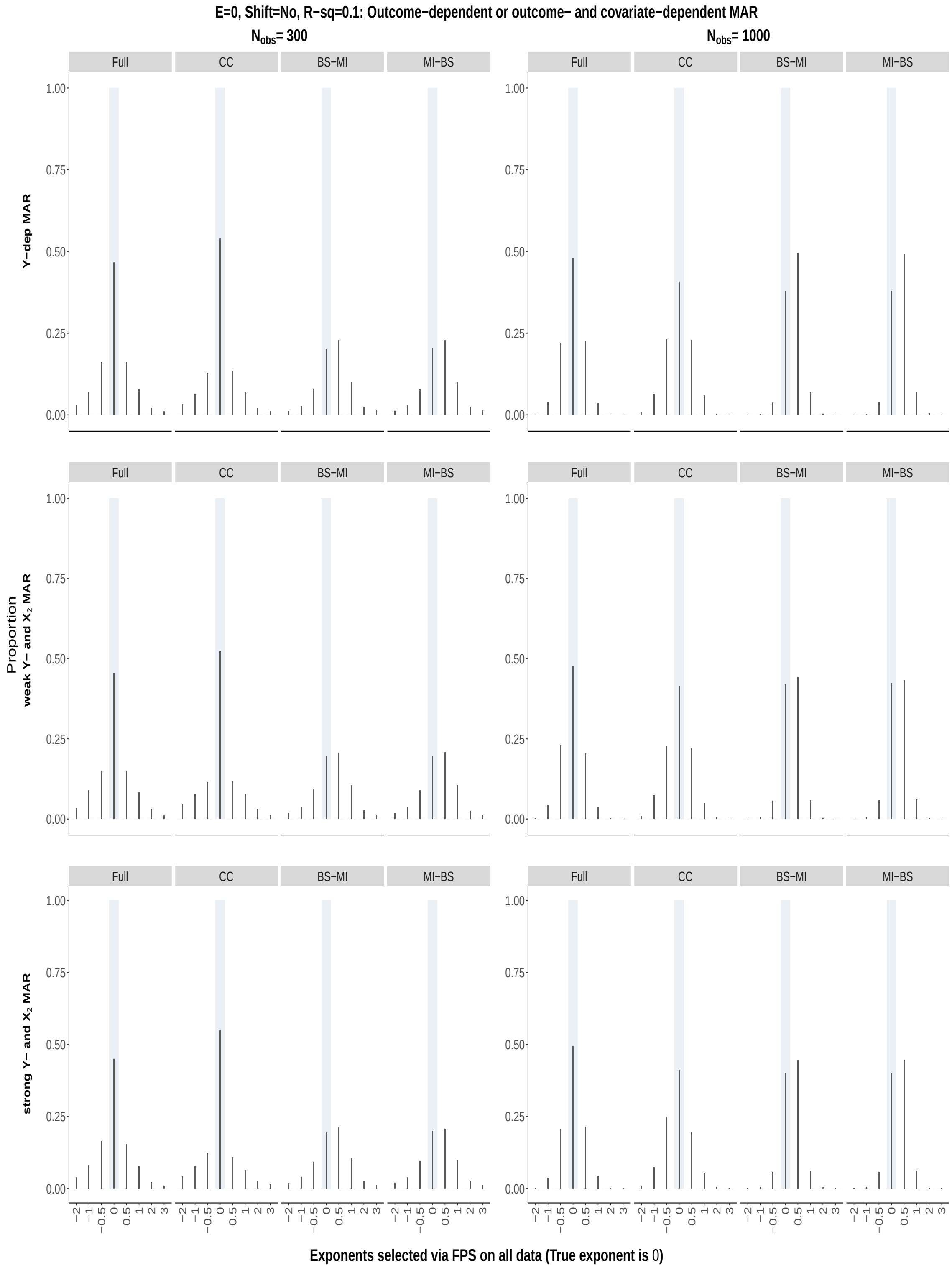

Figure S243: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

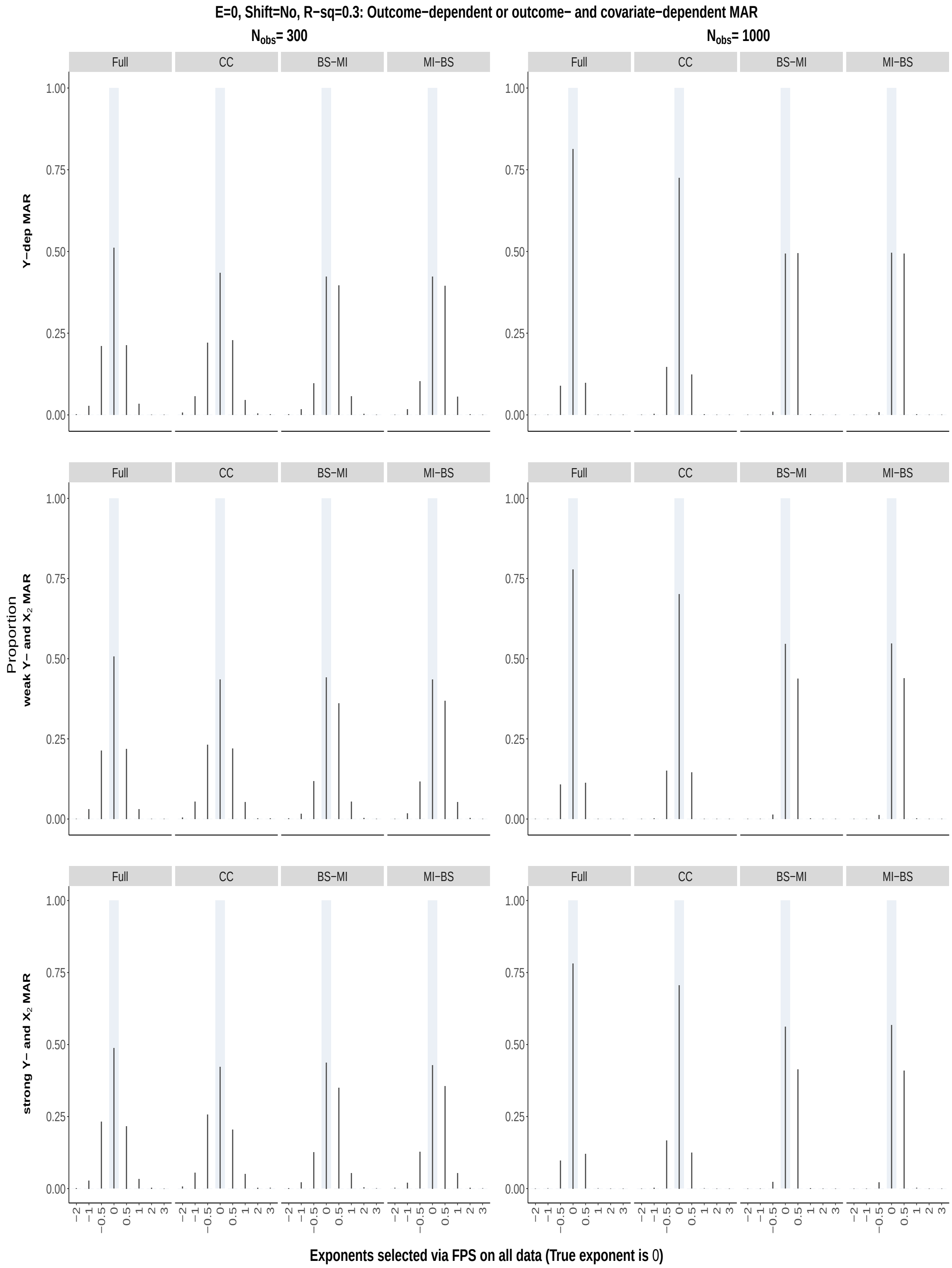

Figure S244: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

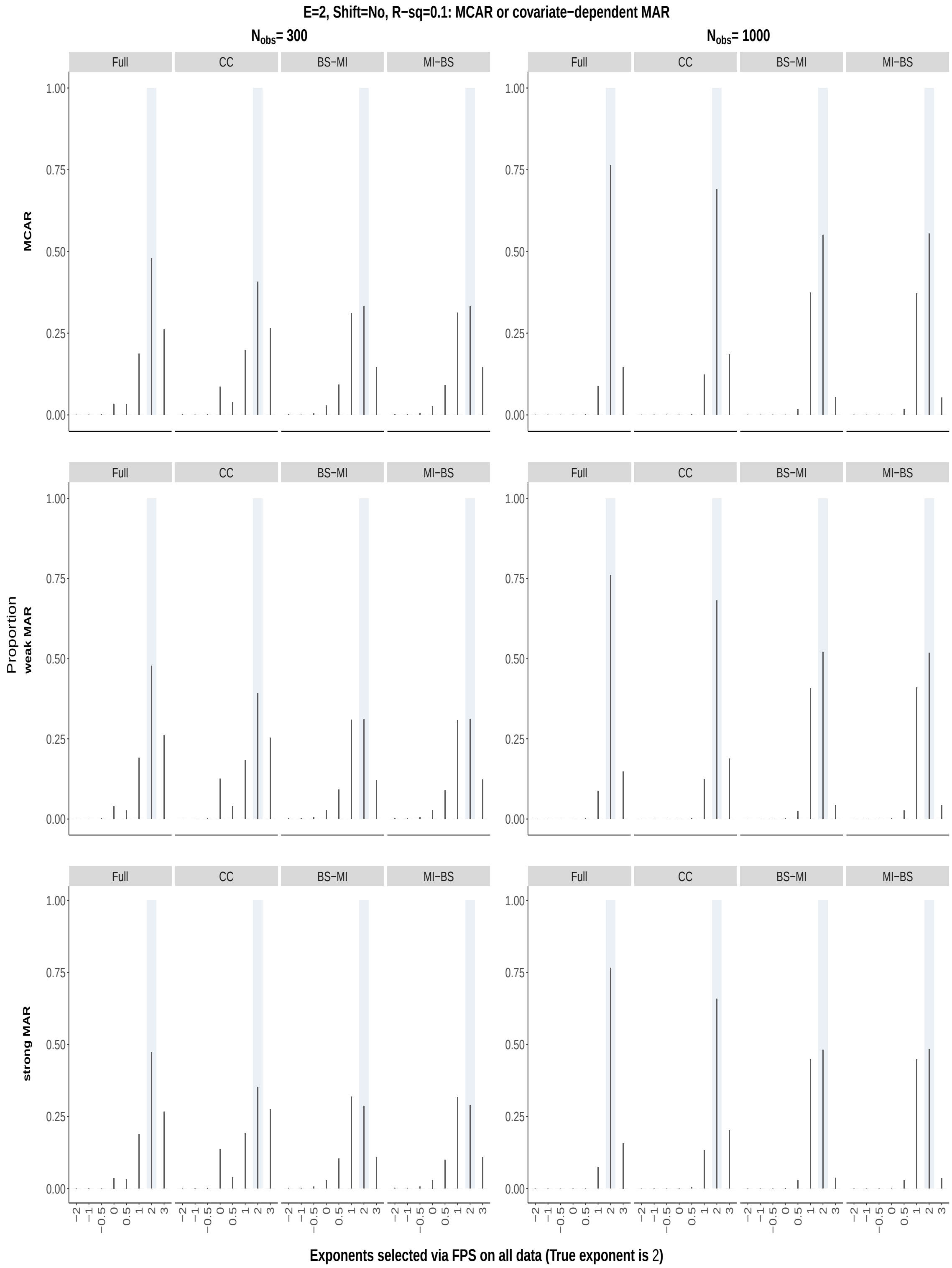

Figure S245: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

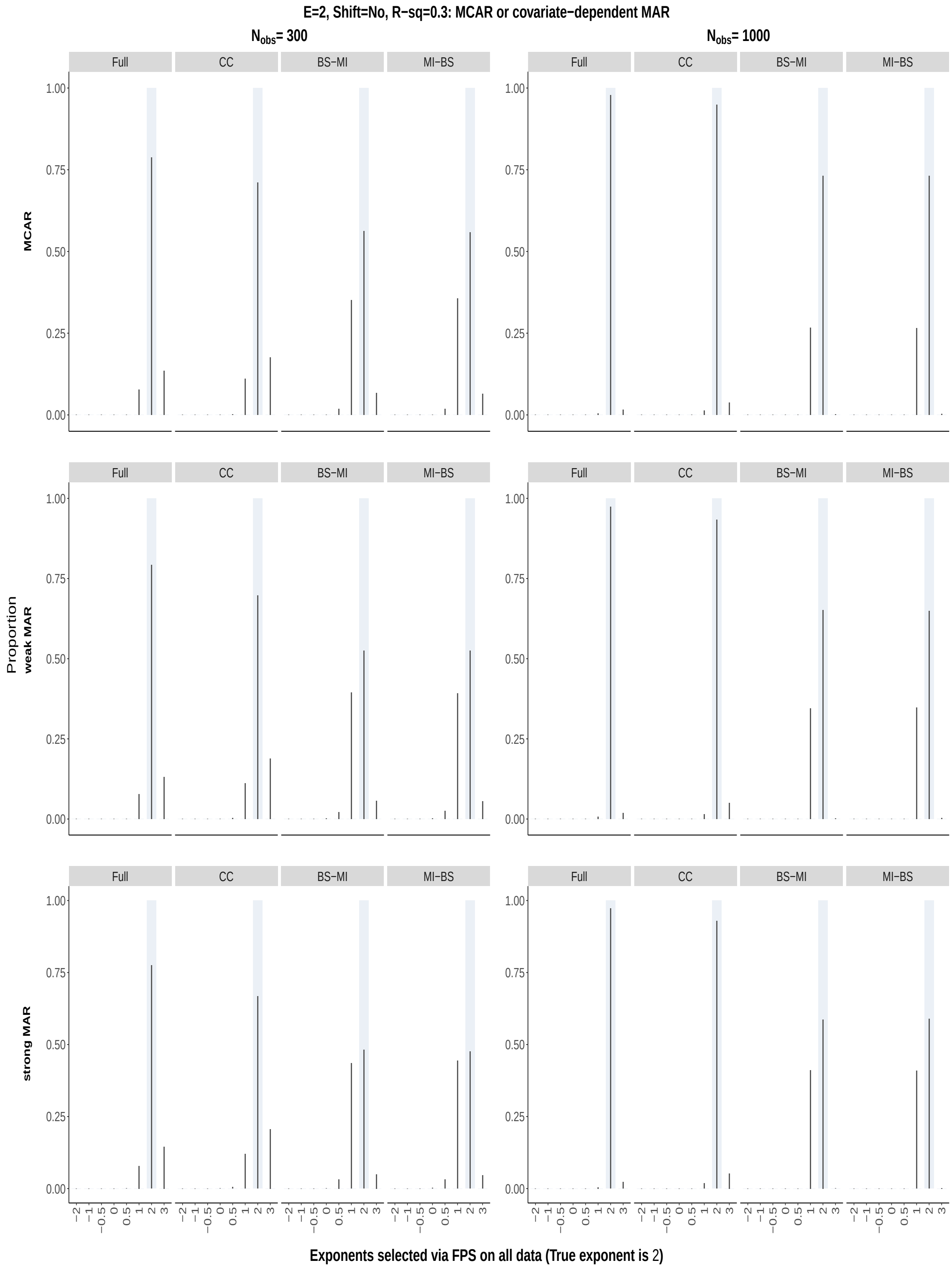

Figure S246: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

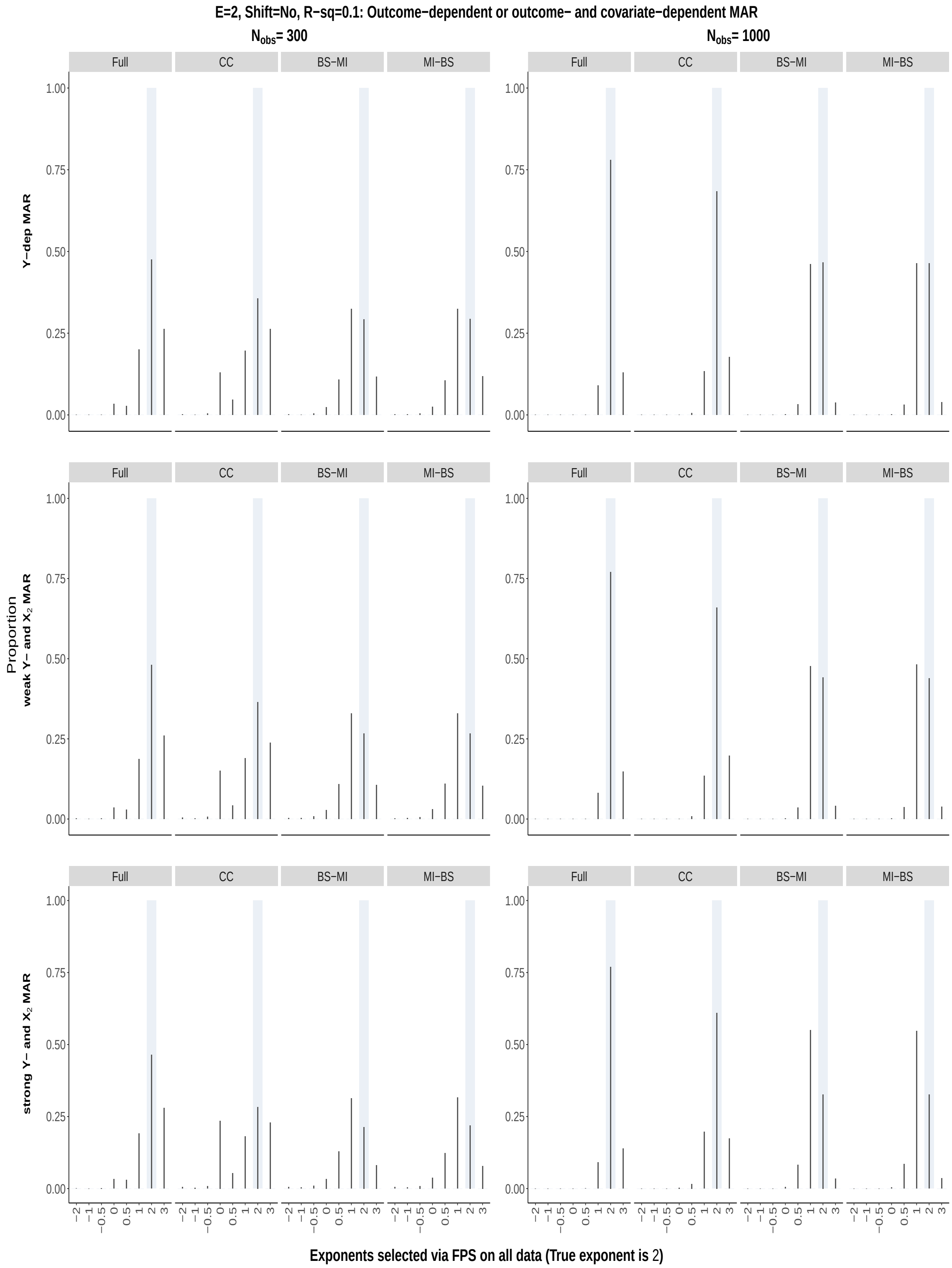

Figure S247: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

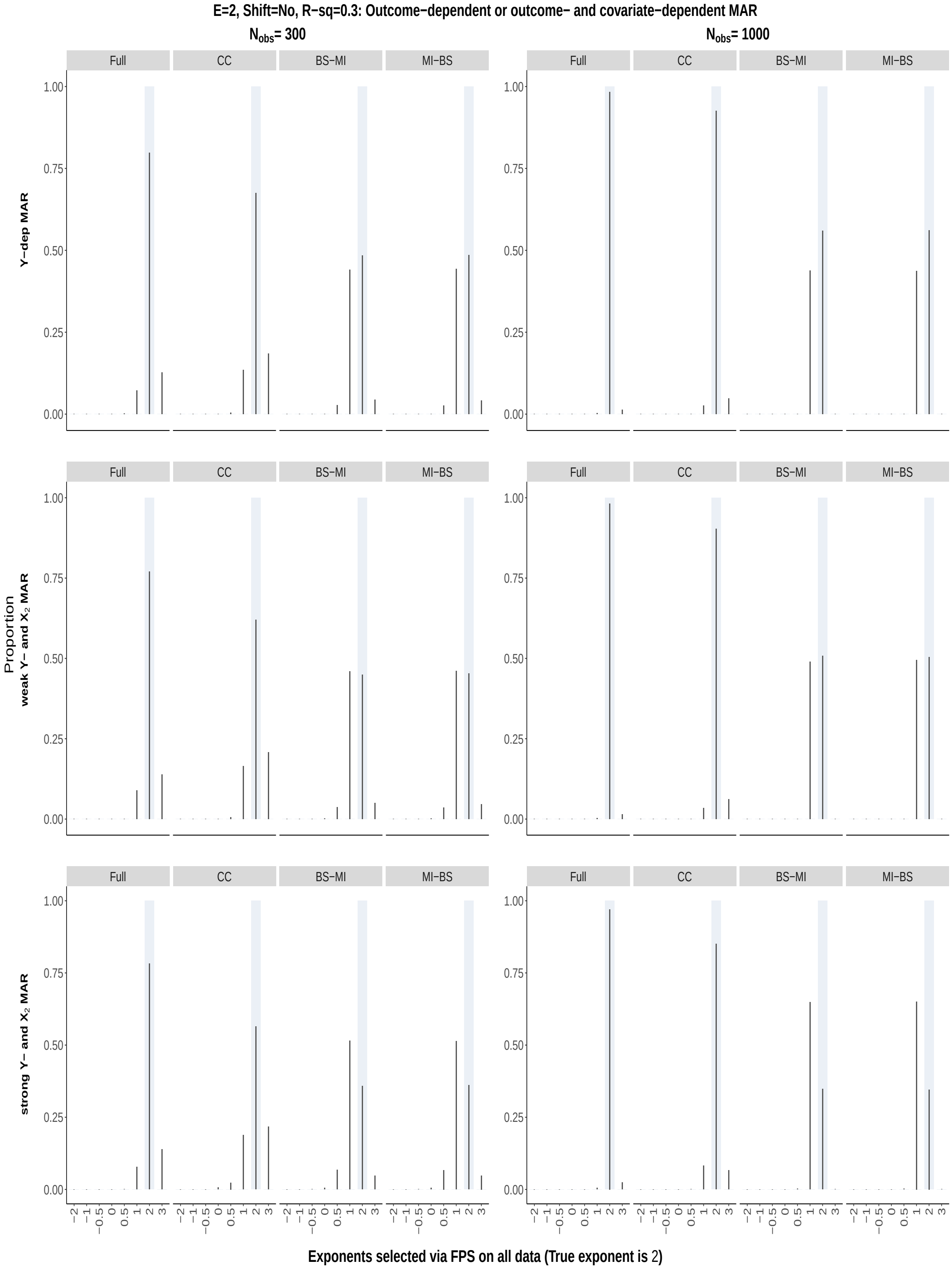

Figure S248: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

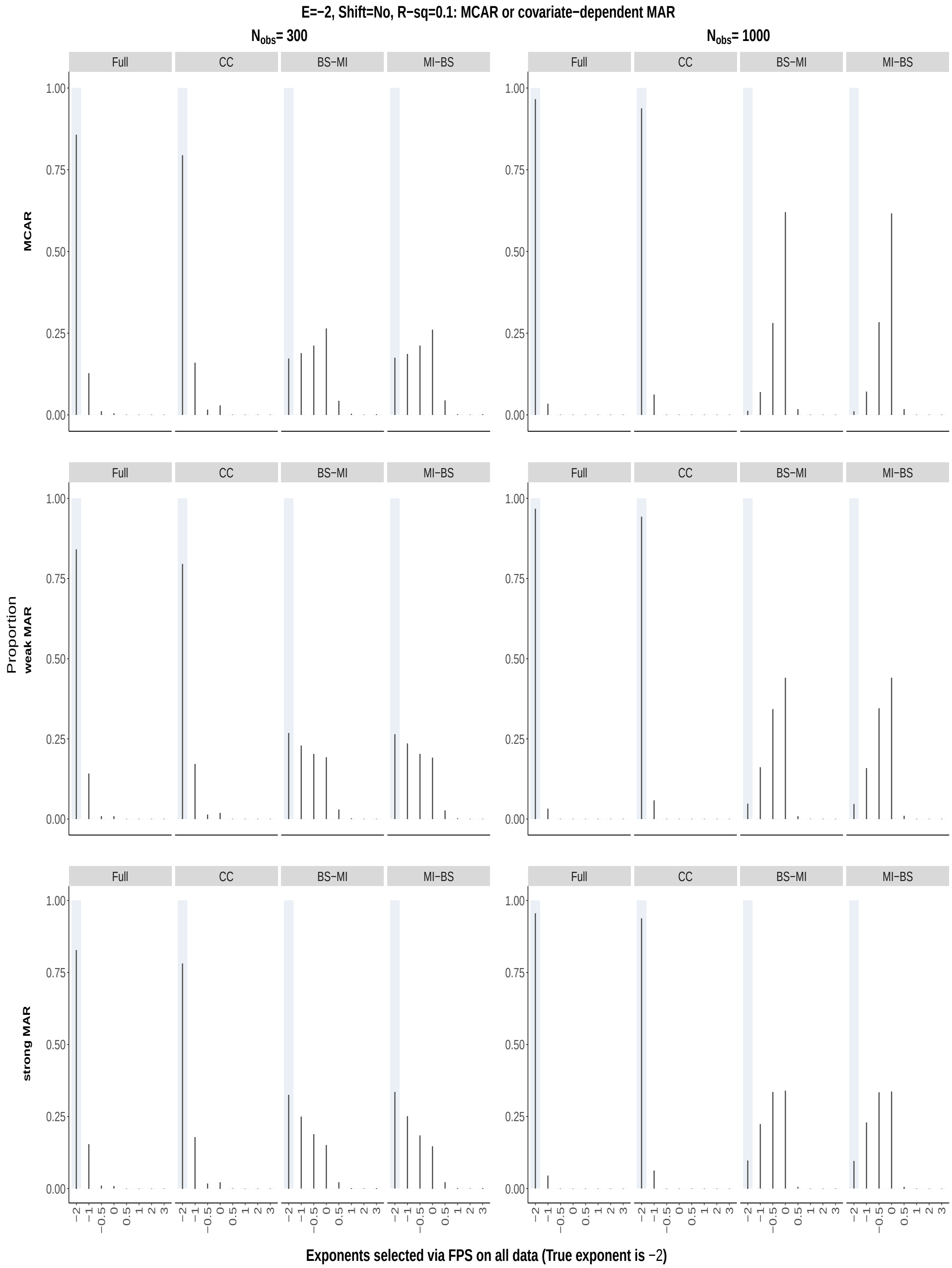

Figure S249: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

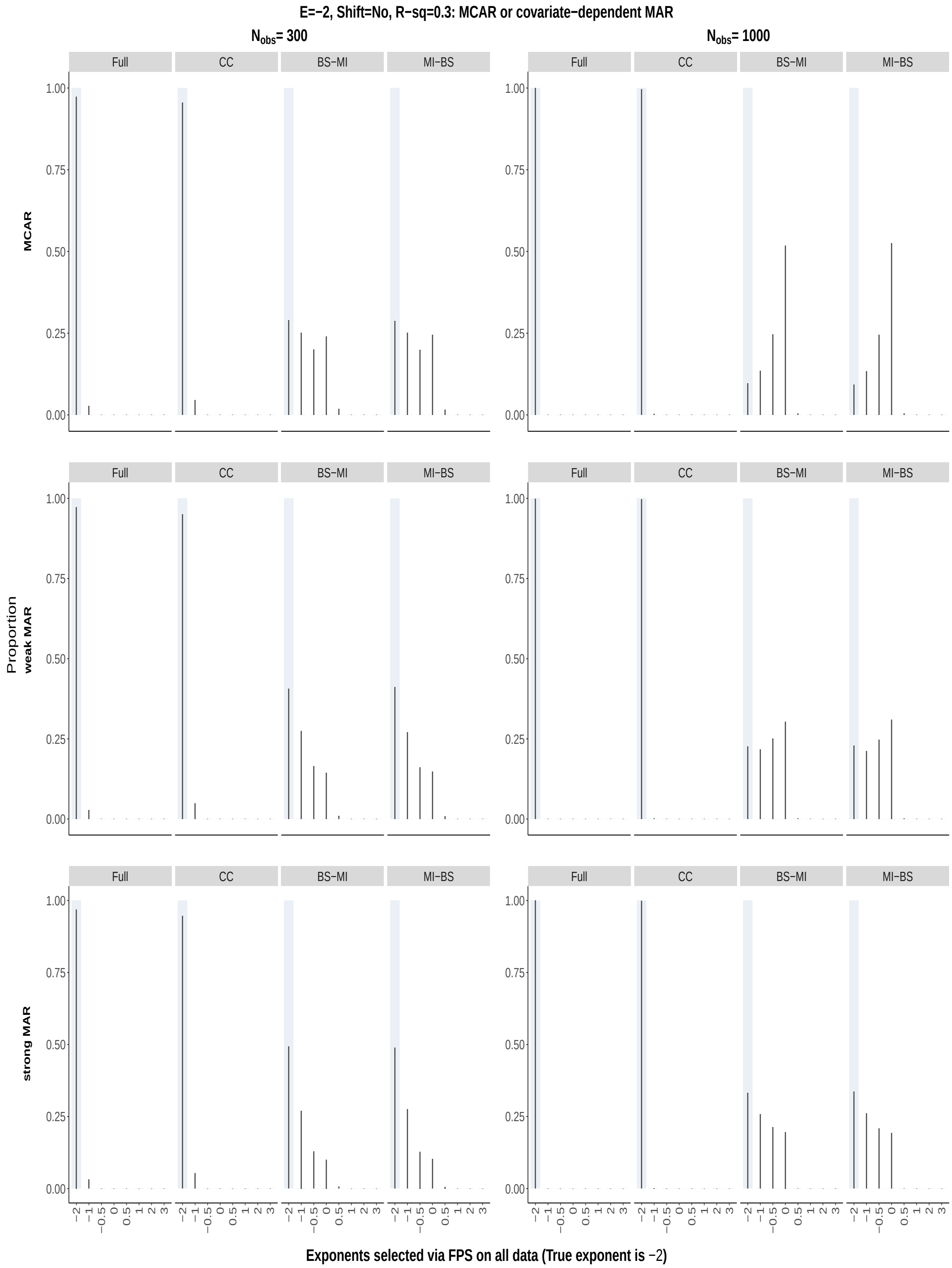

Figure S250: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

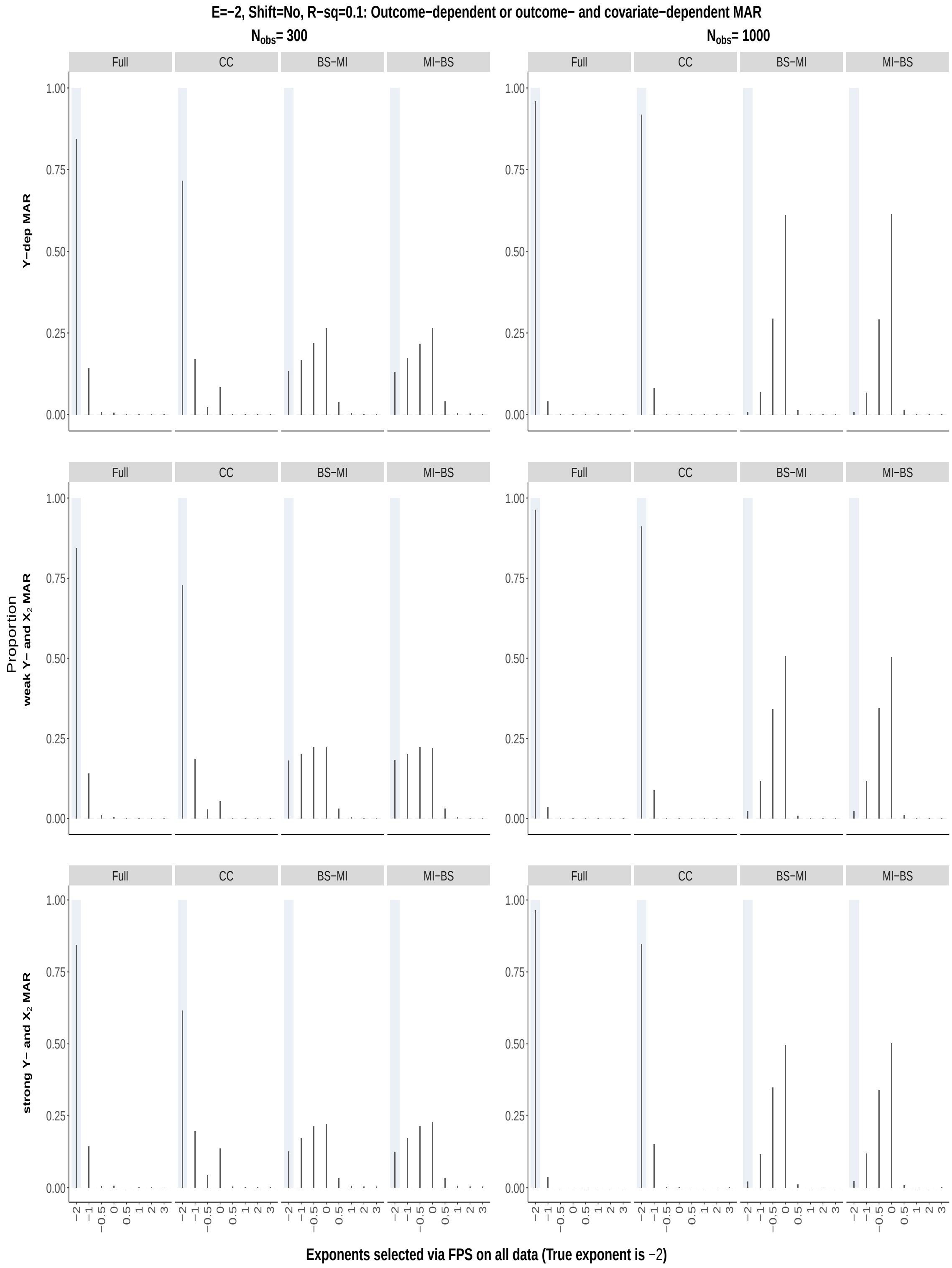

Figure S251: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

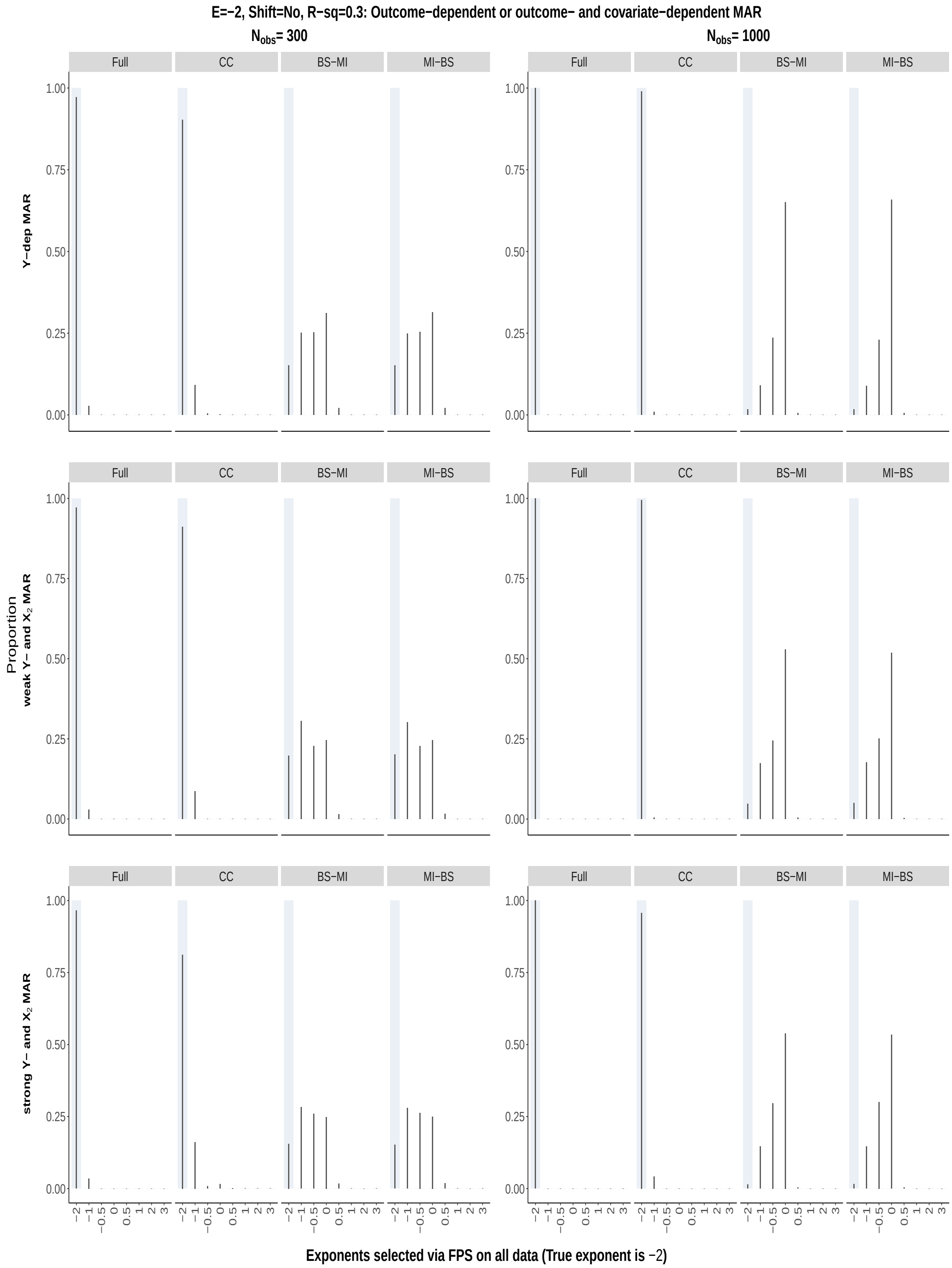

Figure S252: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.18 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 1$, $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

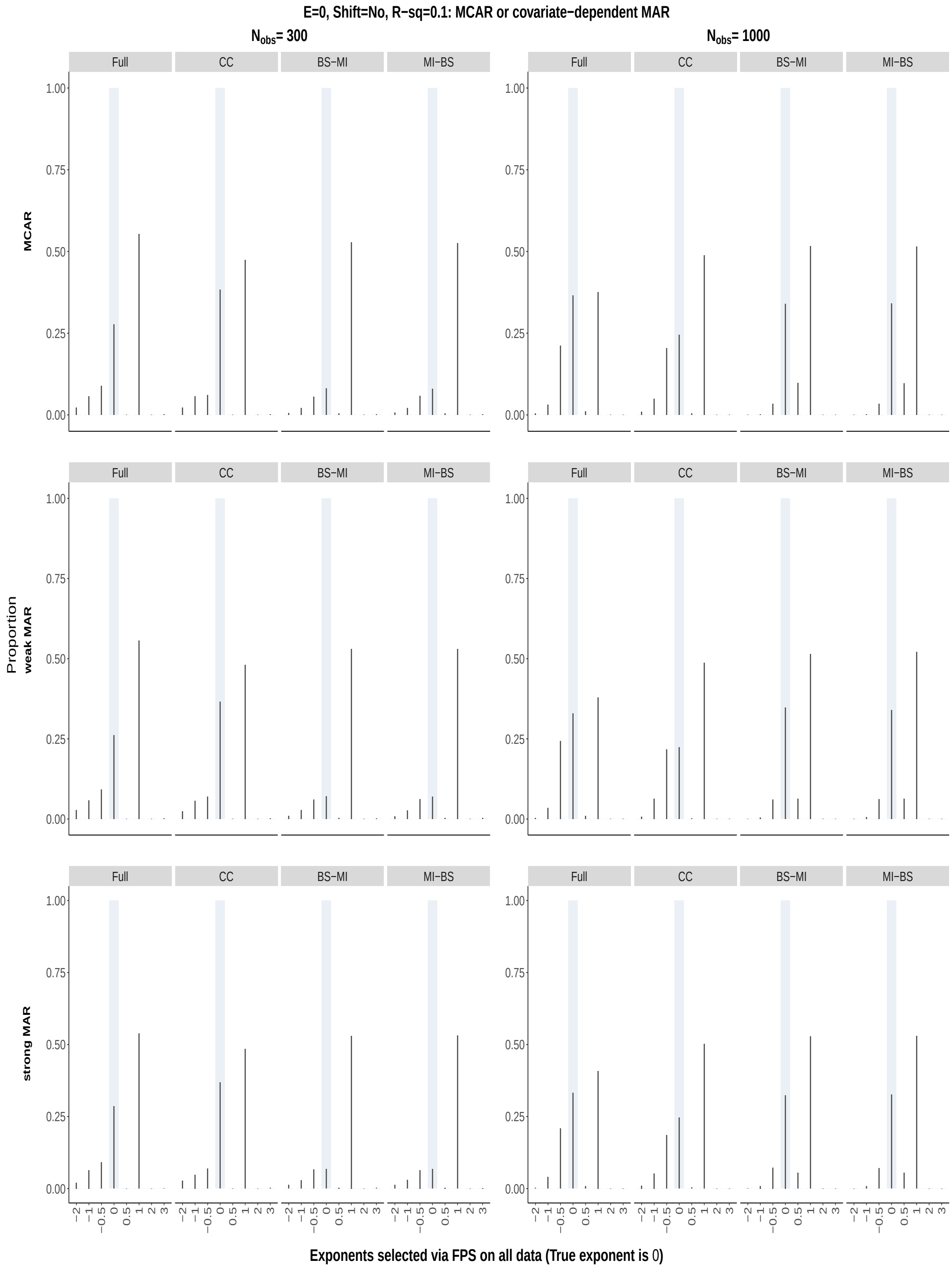

Figure S253: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

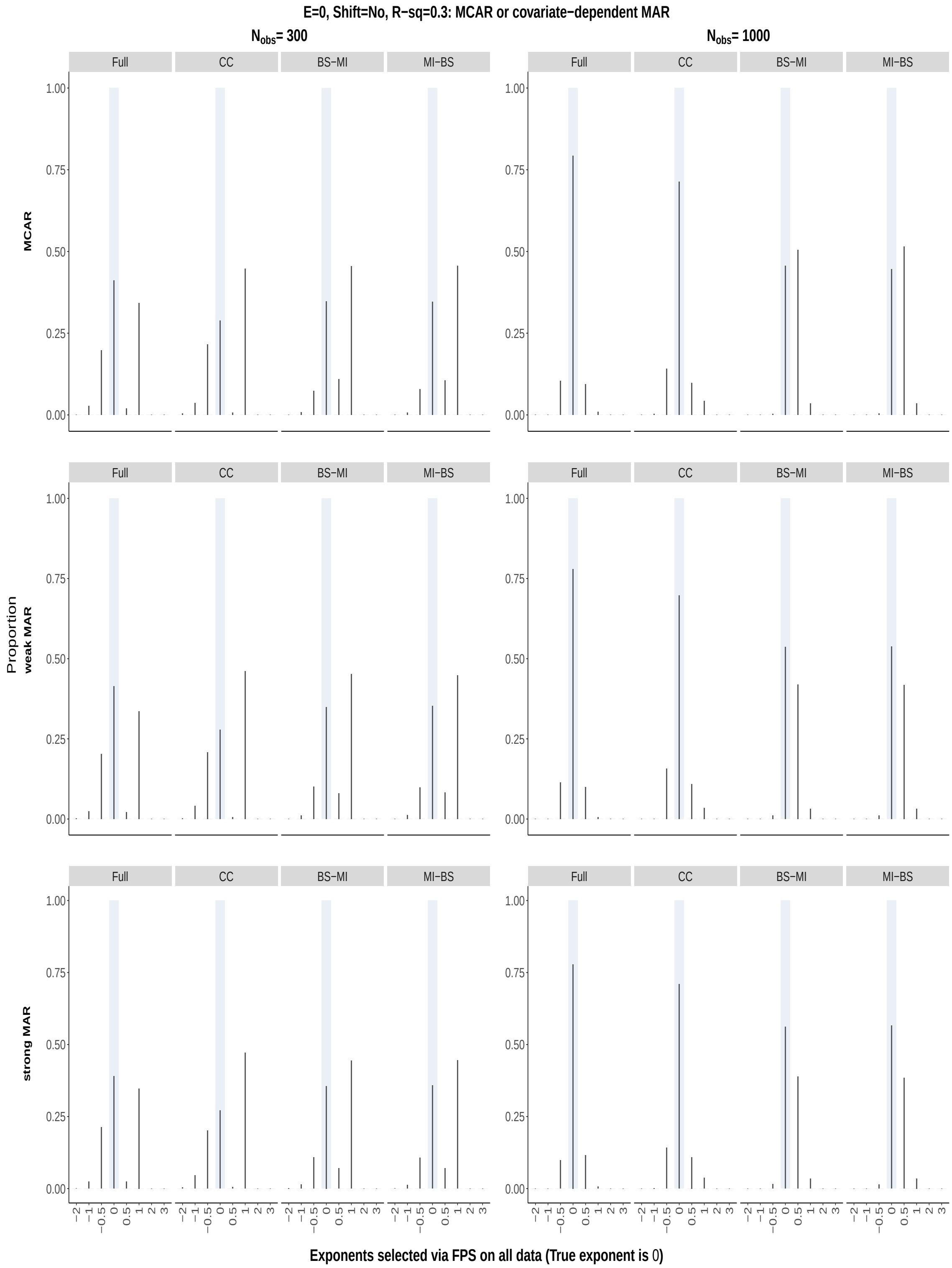

Figure S254: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

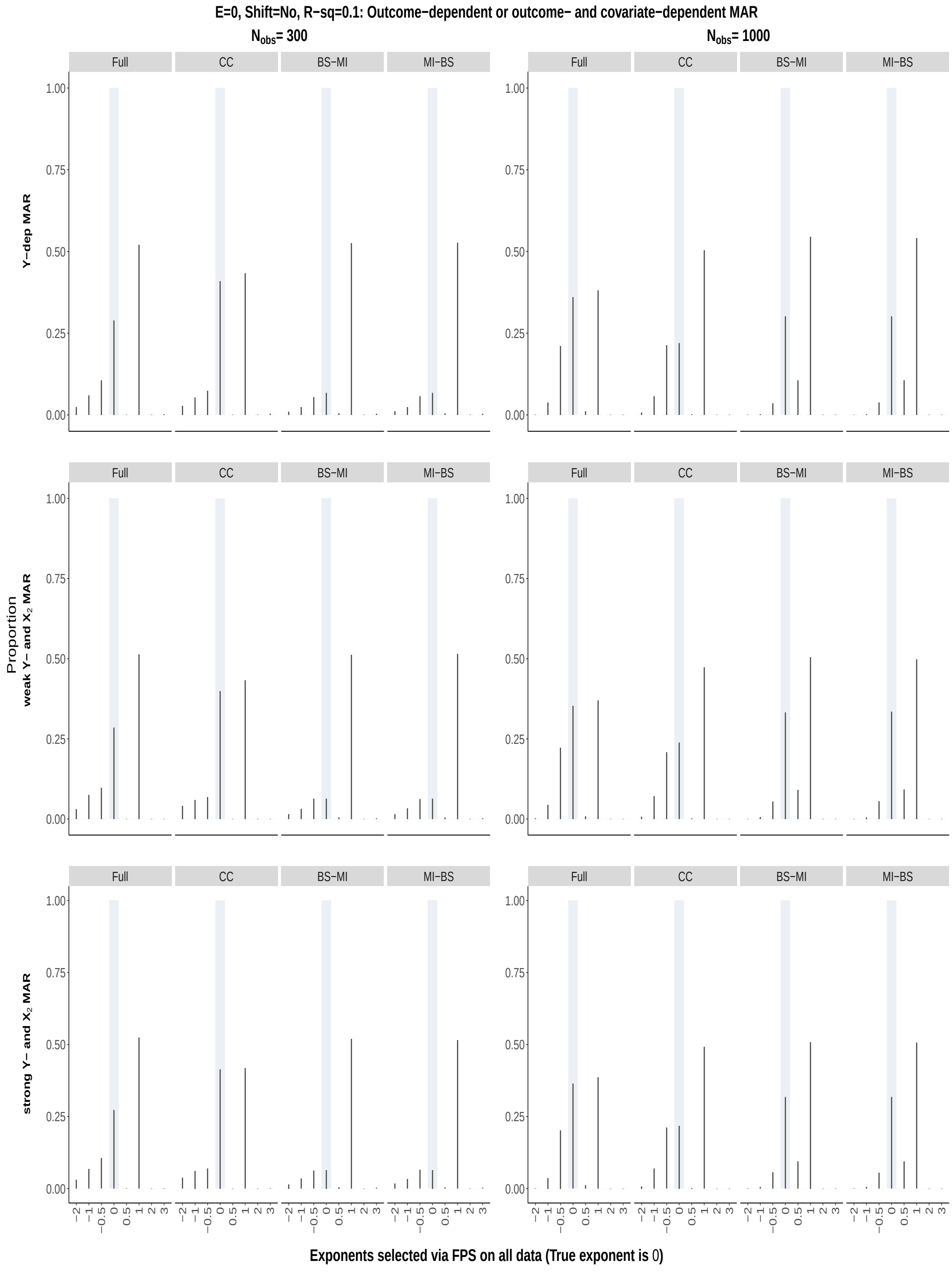

Figure S255: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

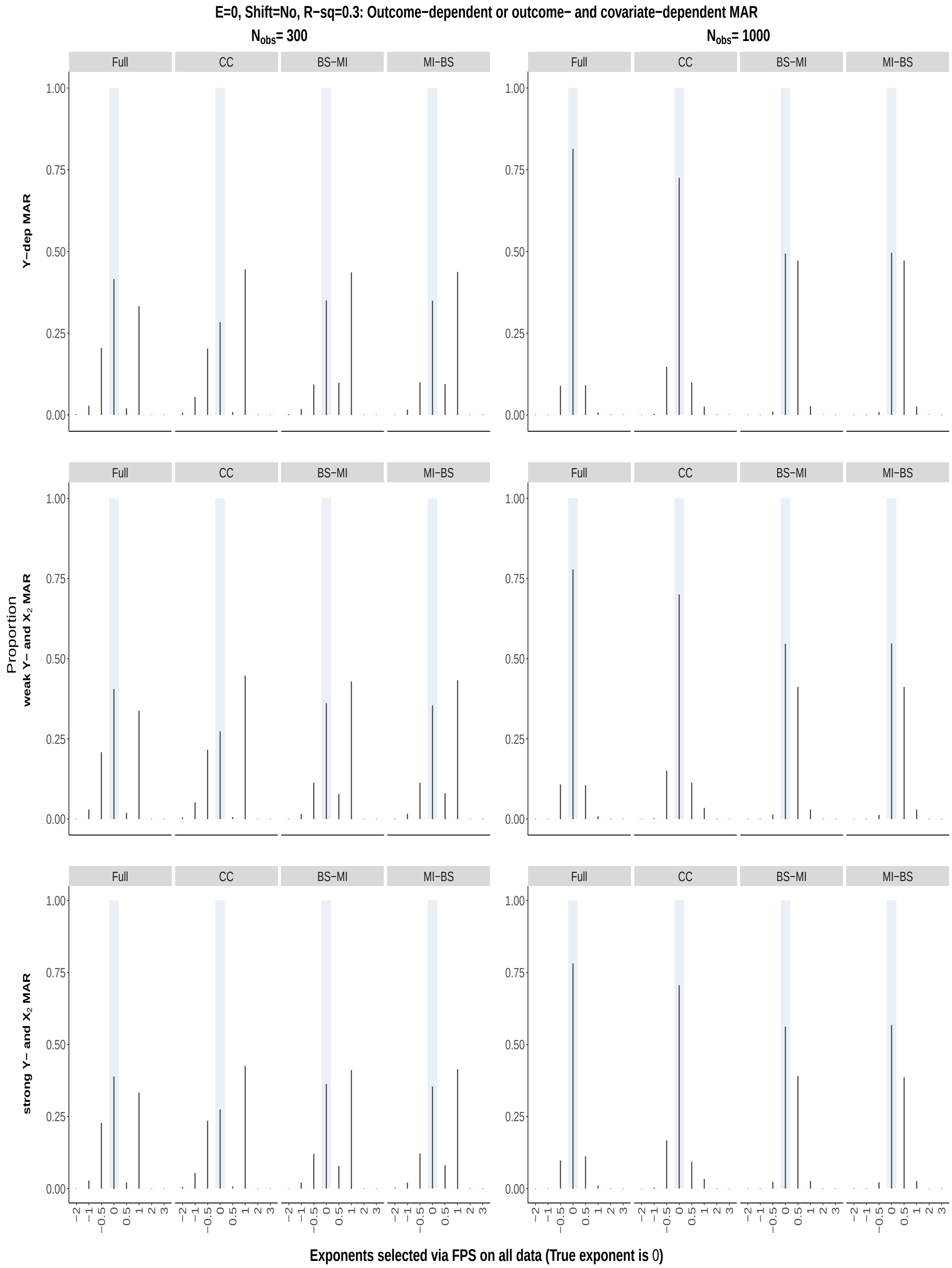

Figure S256: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

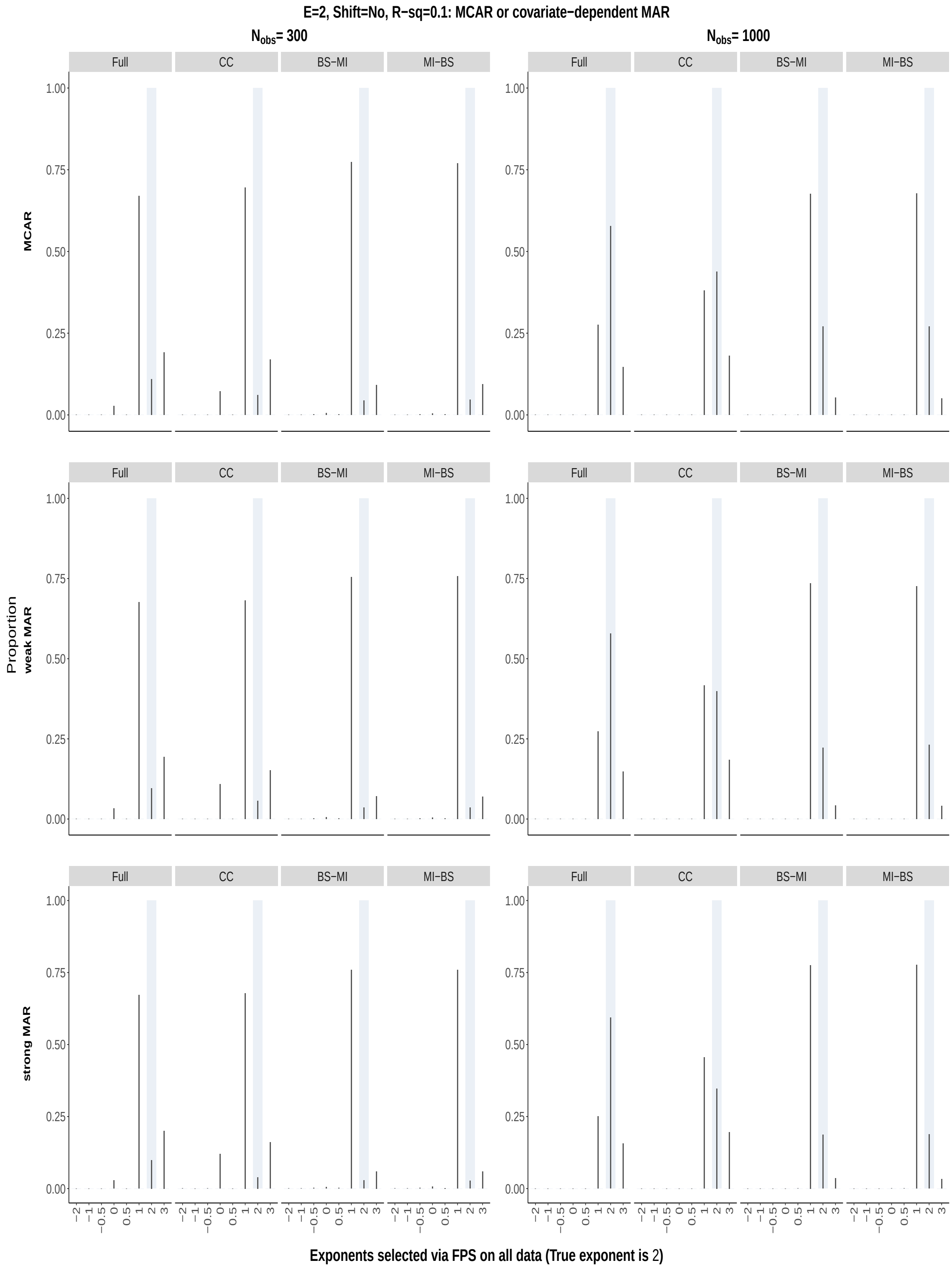

Figure S257: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

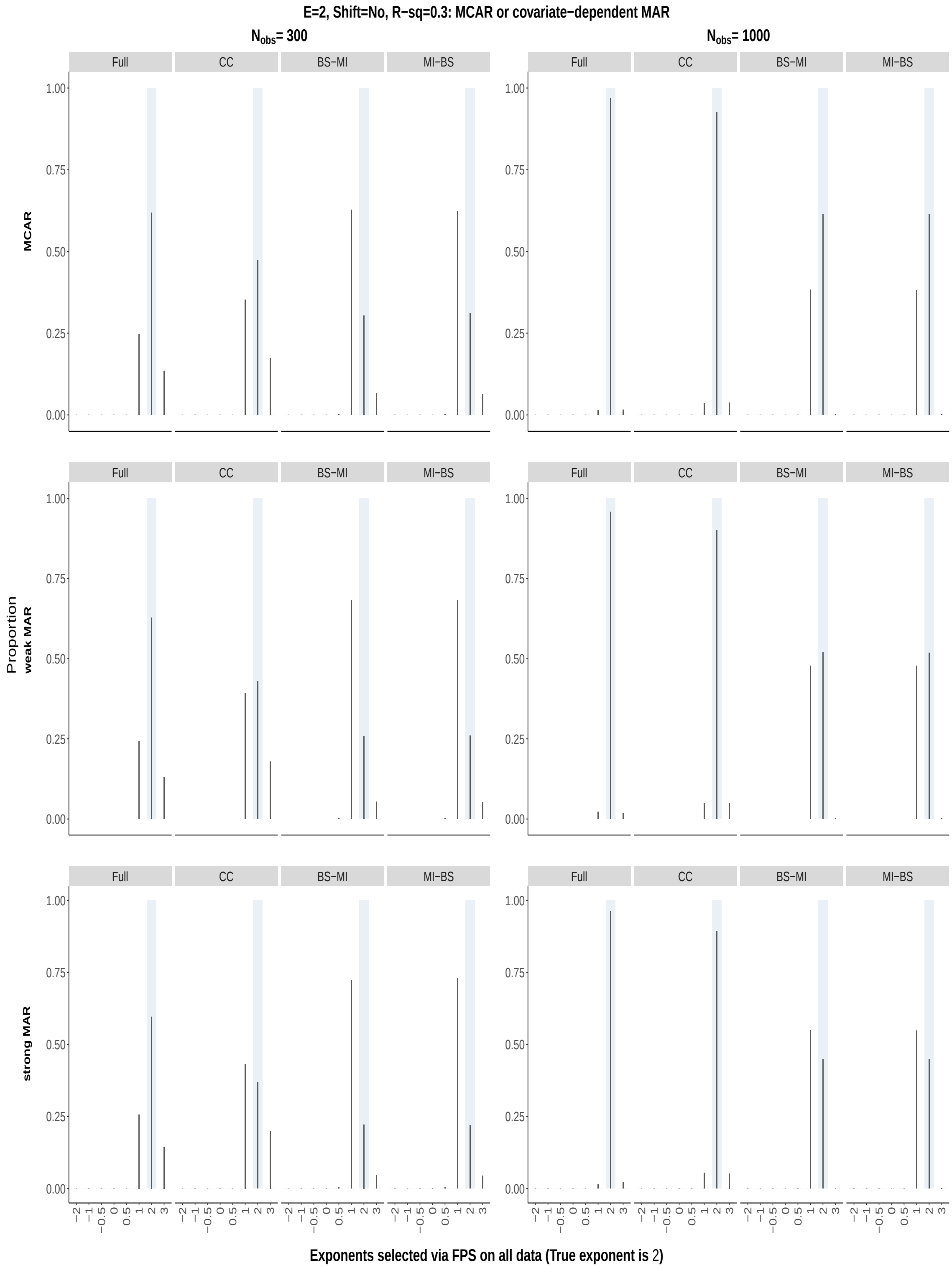

Figure S258: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

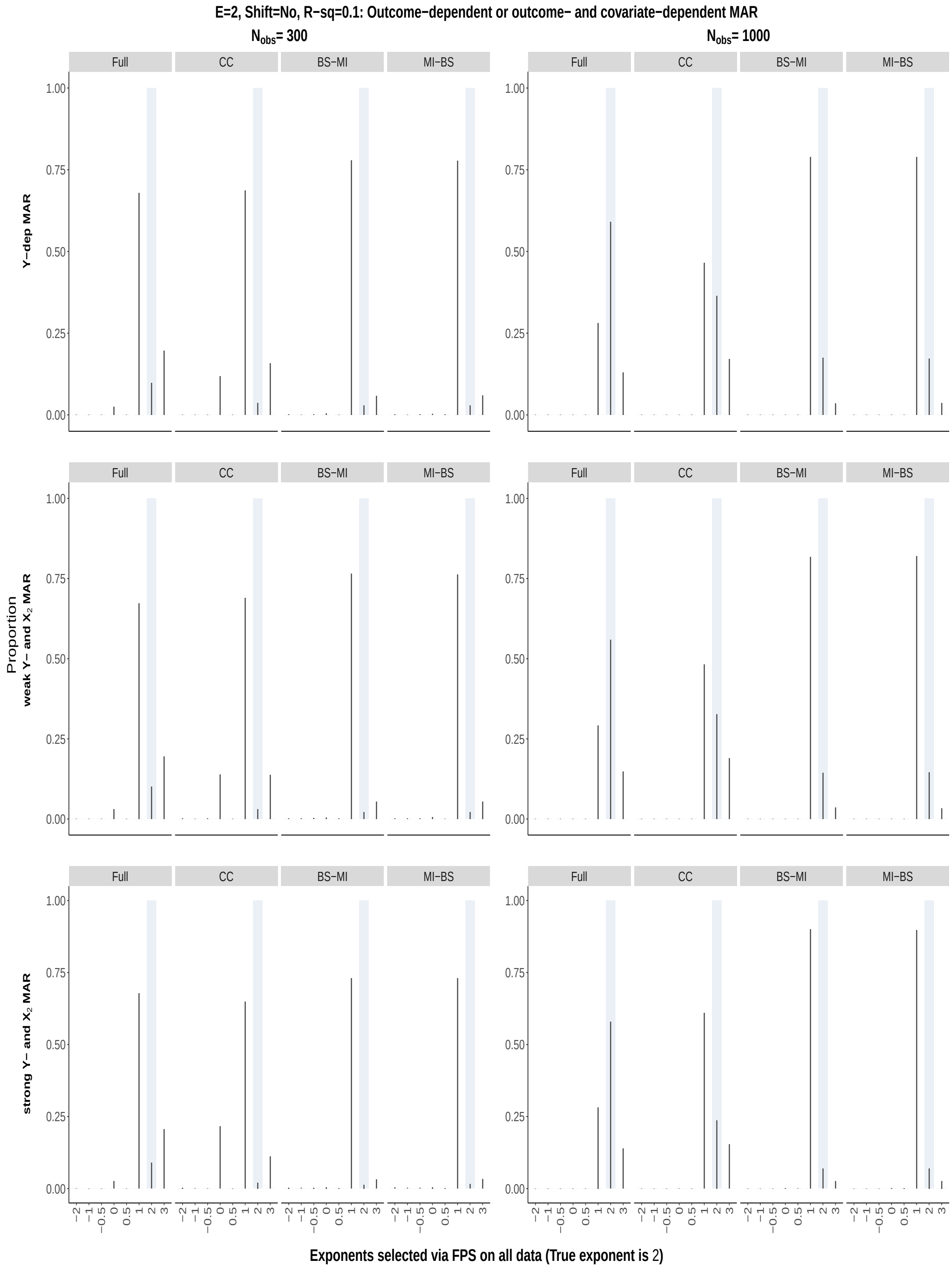

Figure S259: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

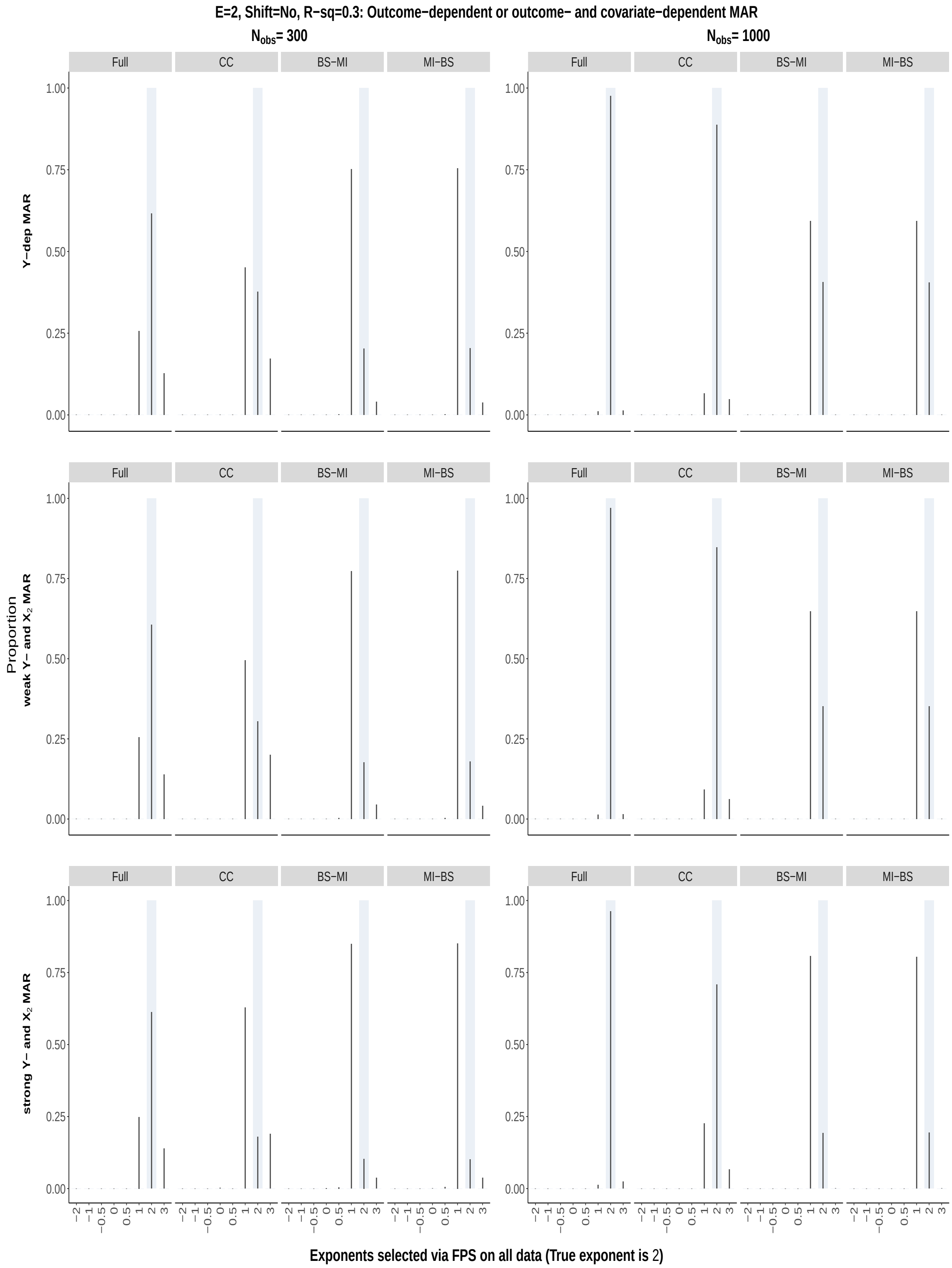

Figure S260: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

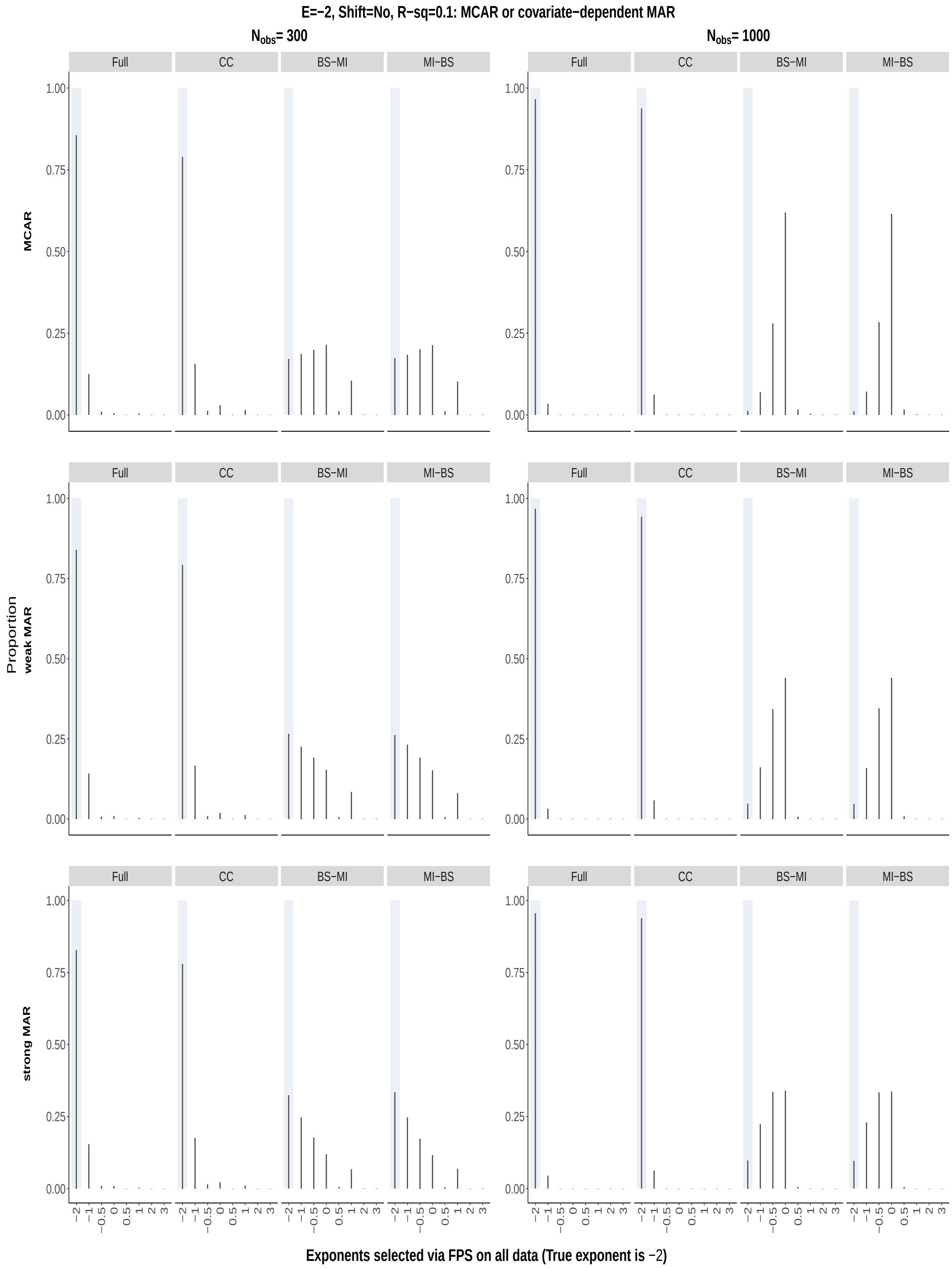

Figure S261: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

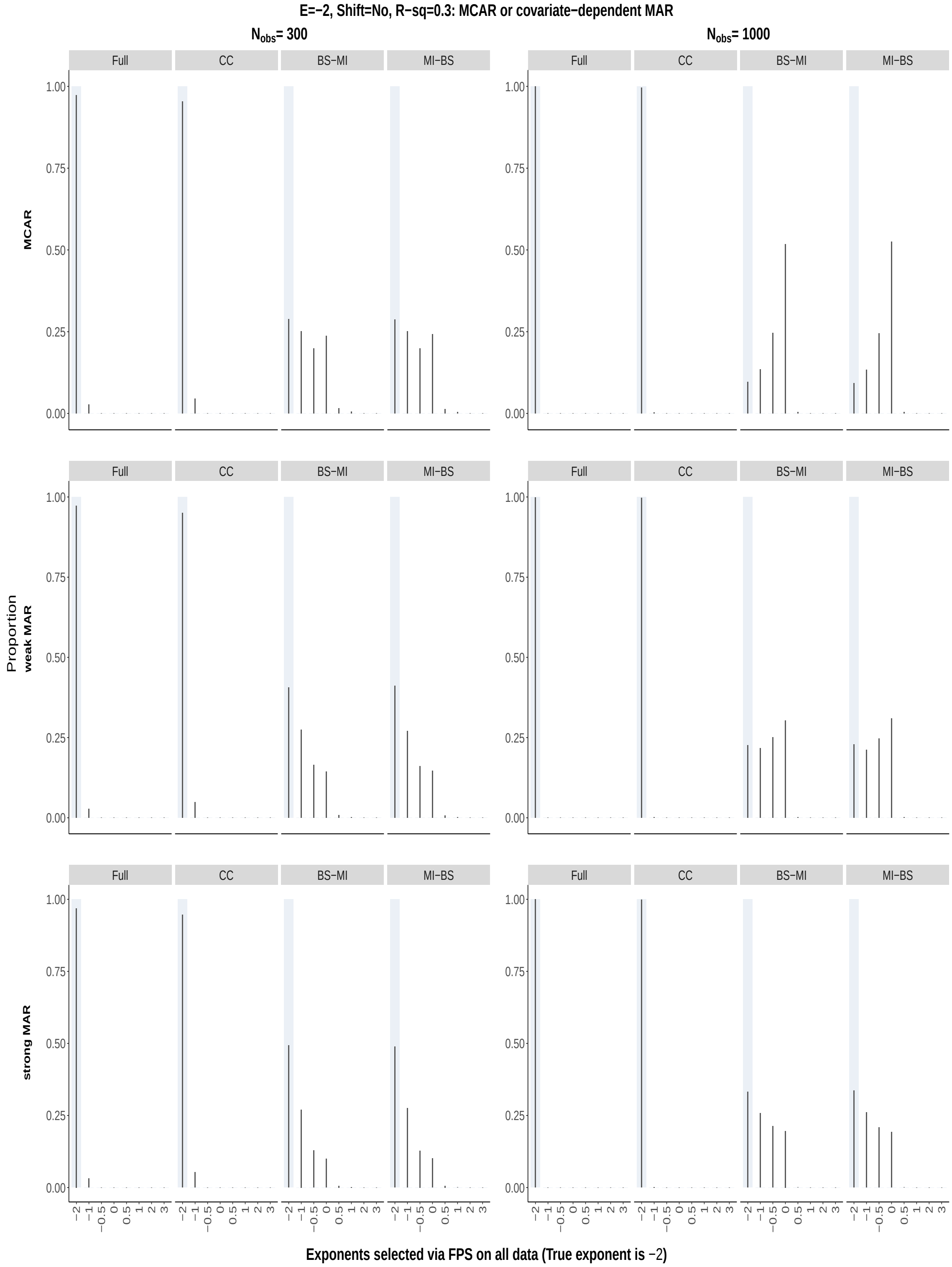

Figure S262: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

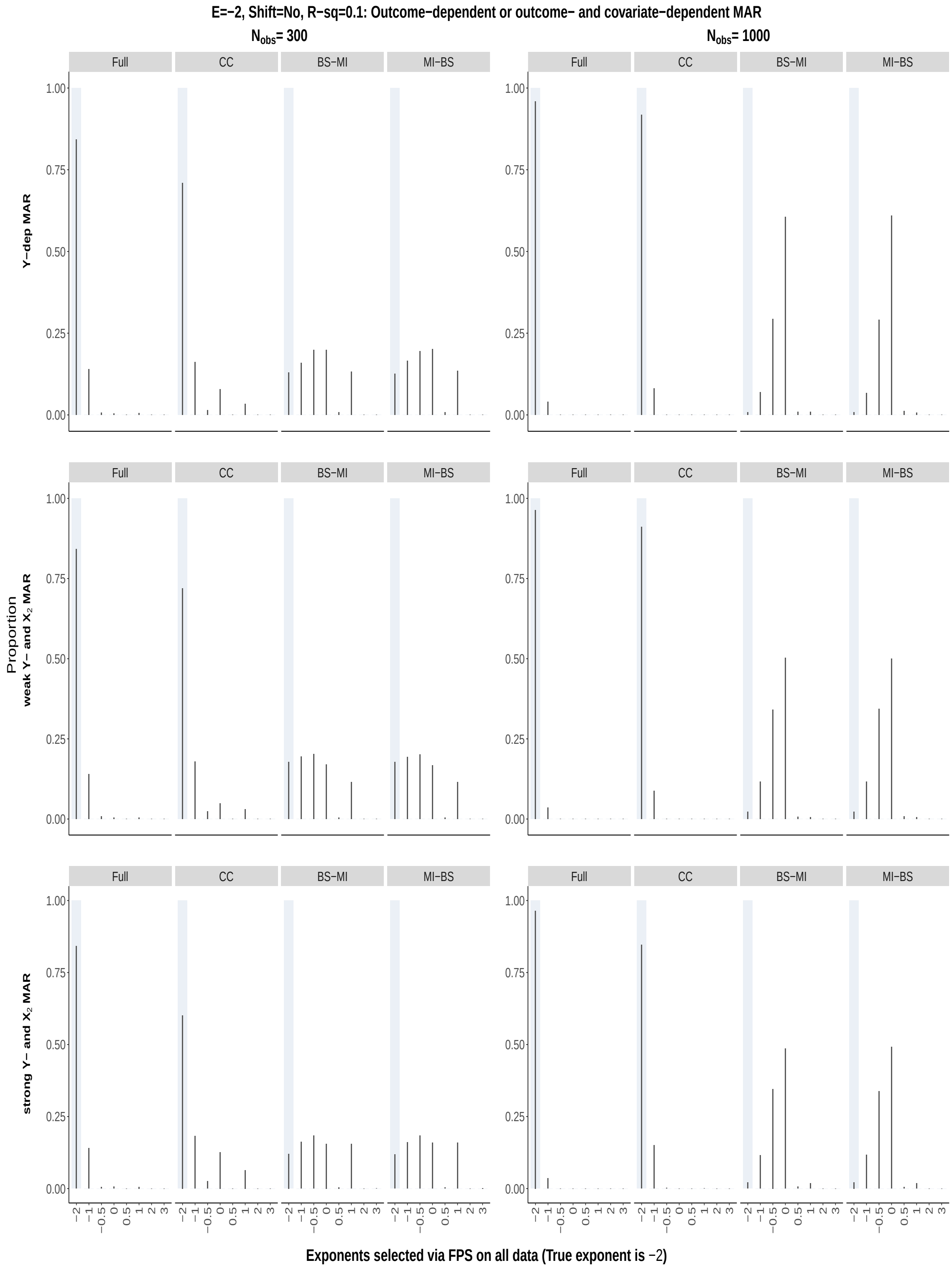

Figure S263: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

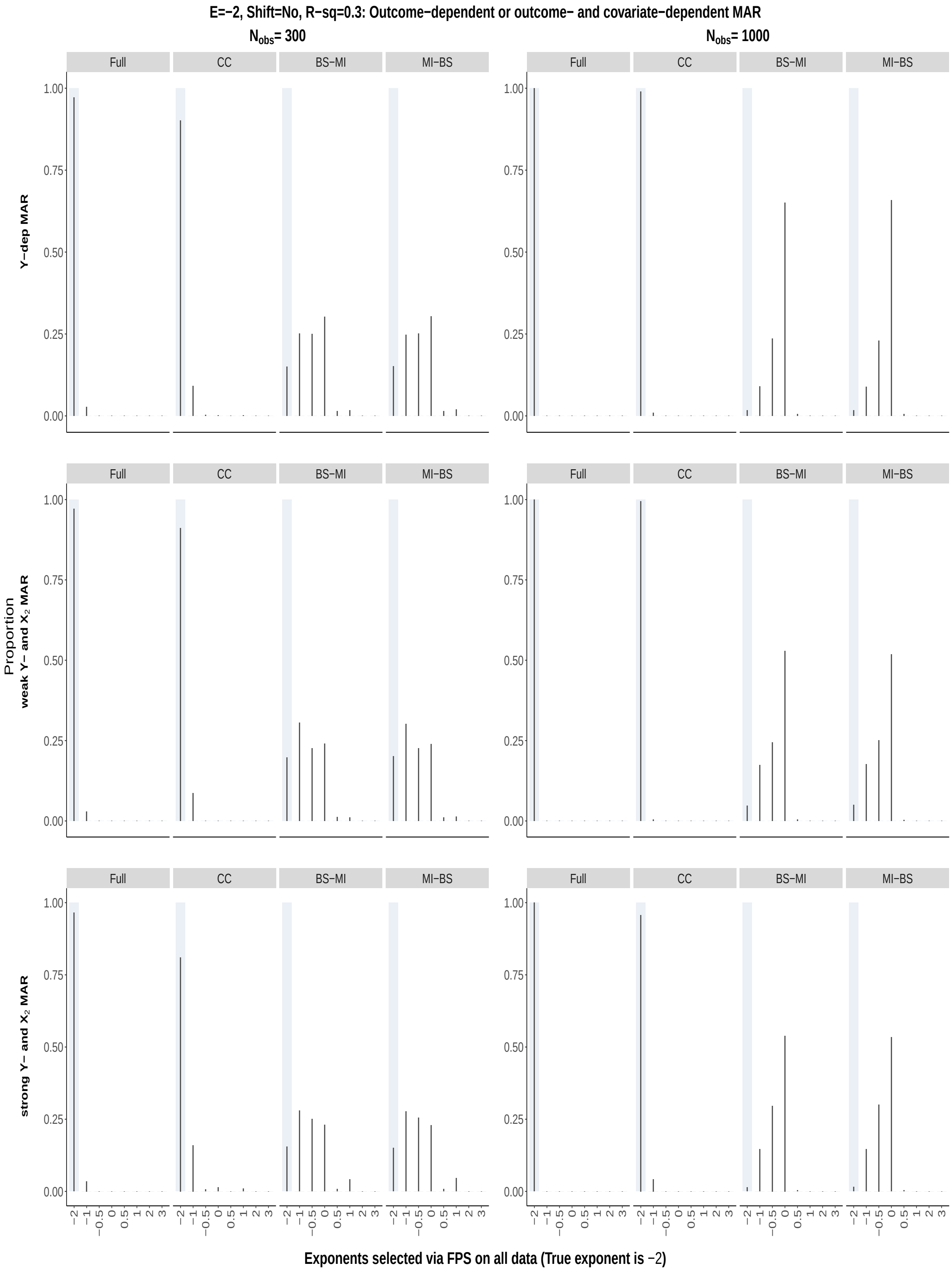

Figure S264: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.19 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

True exponent is 0

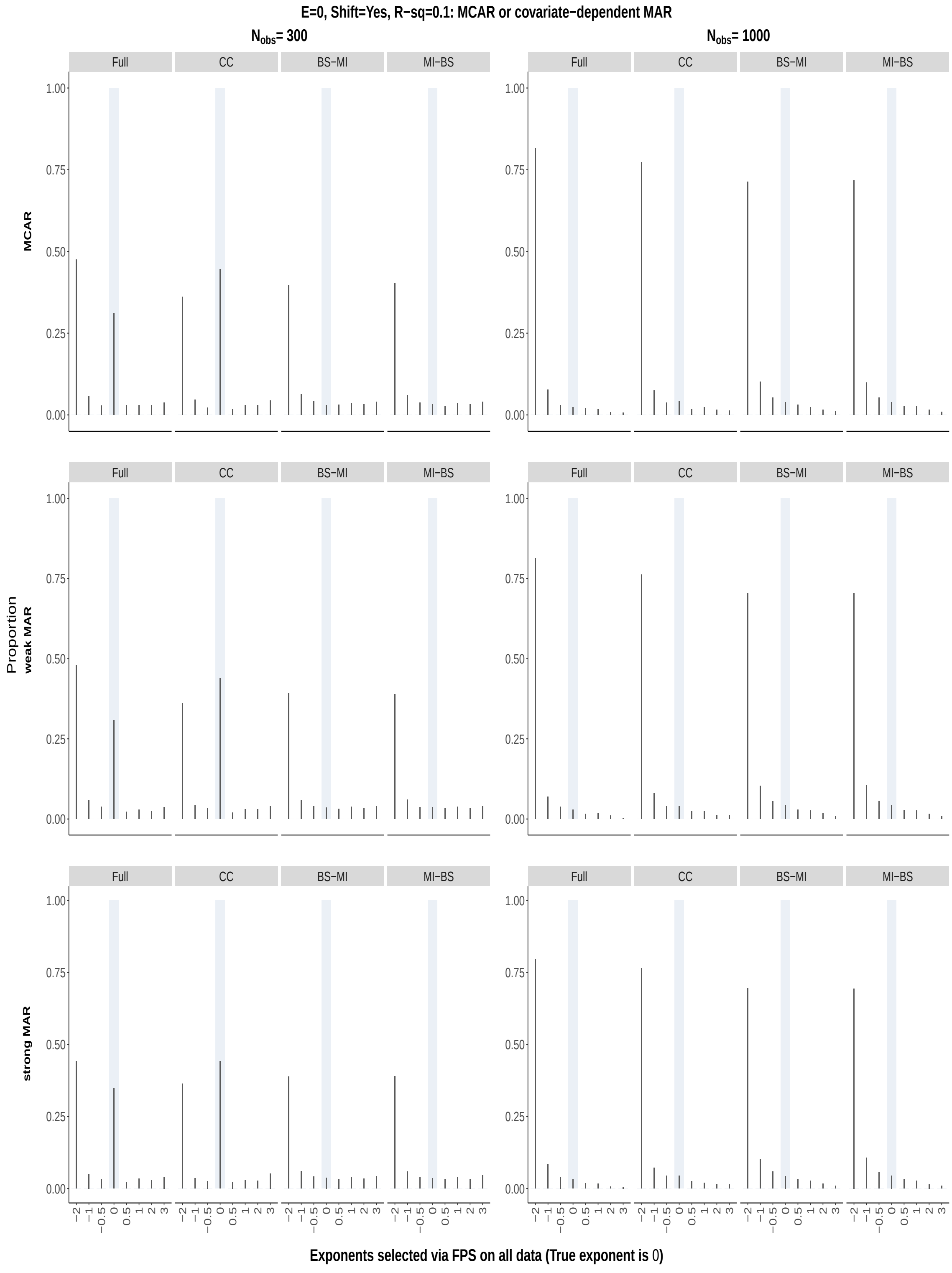

Figure S265: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

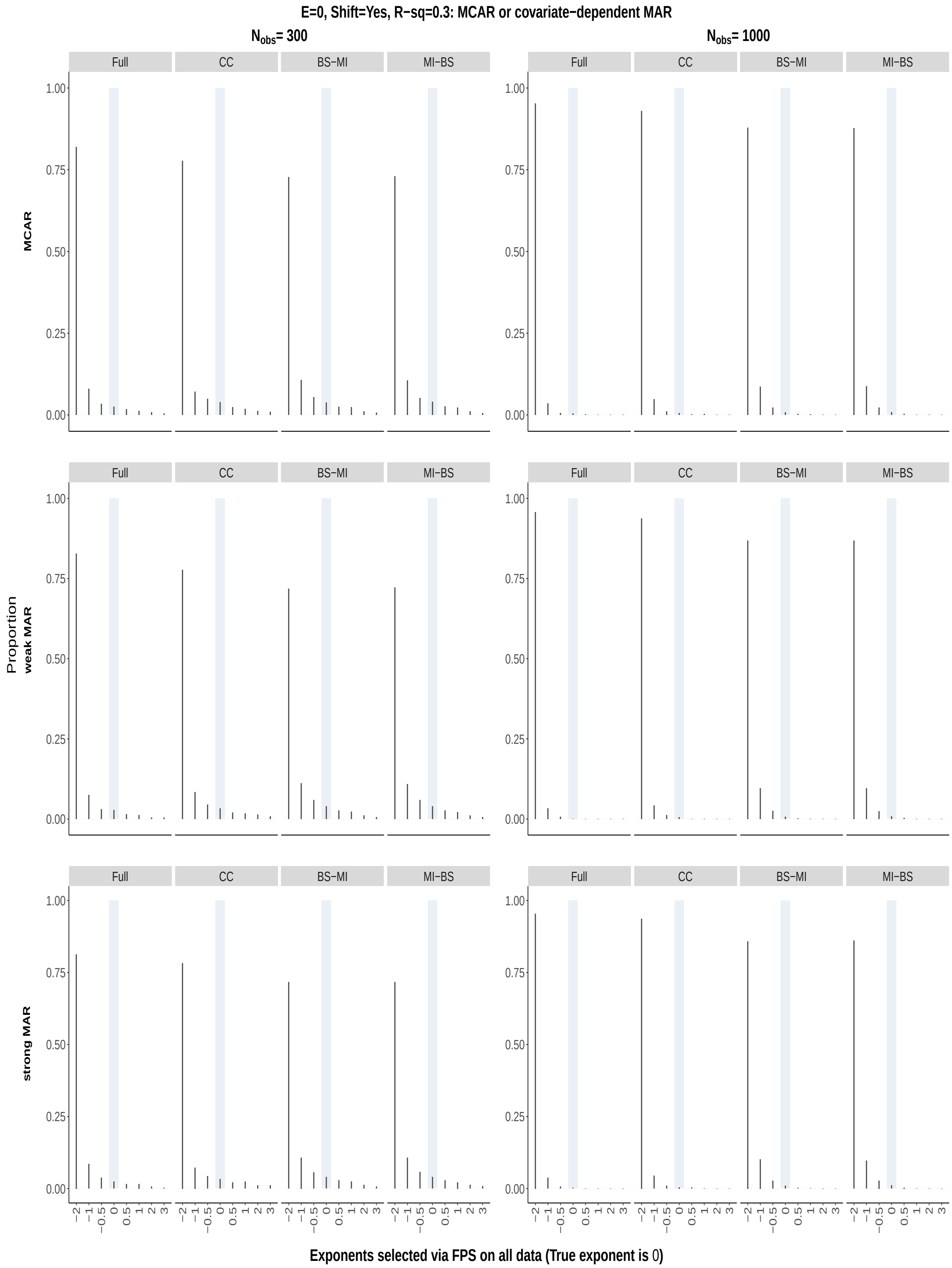

Figure S266: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

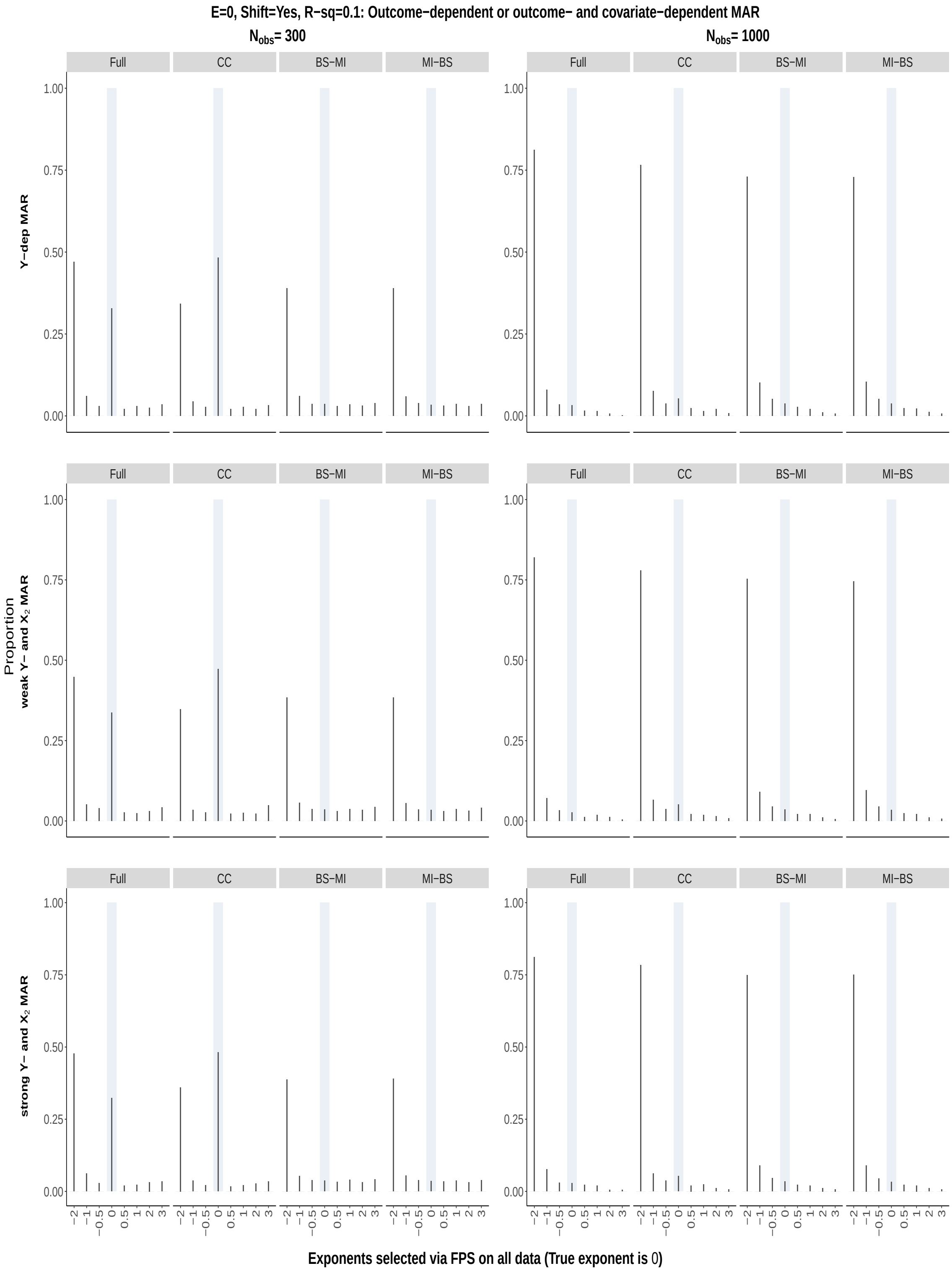

Figure S267: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

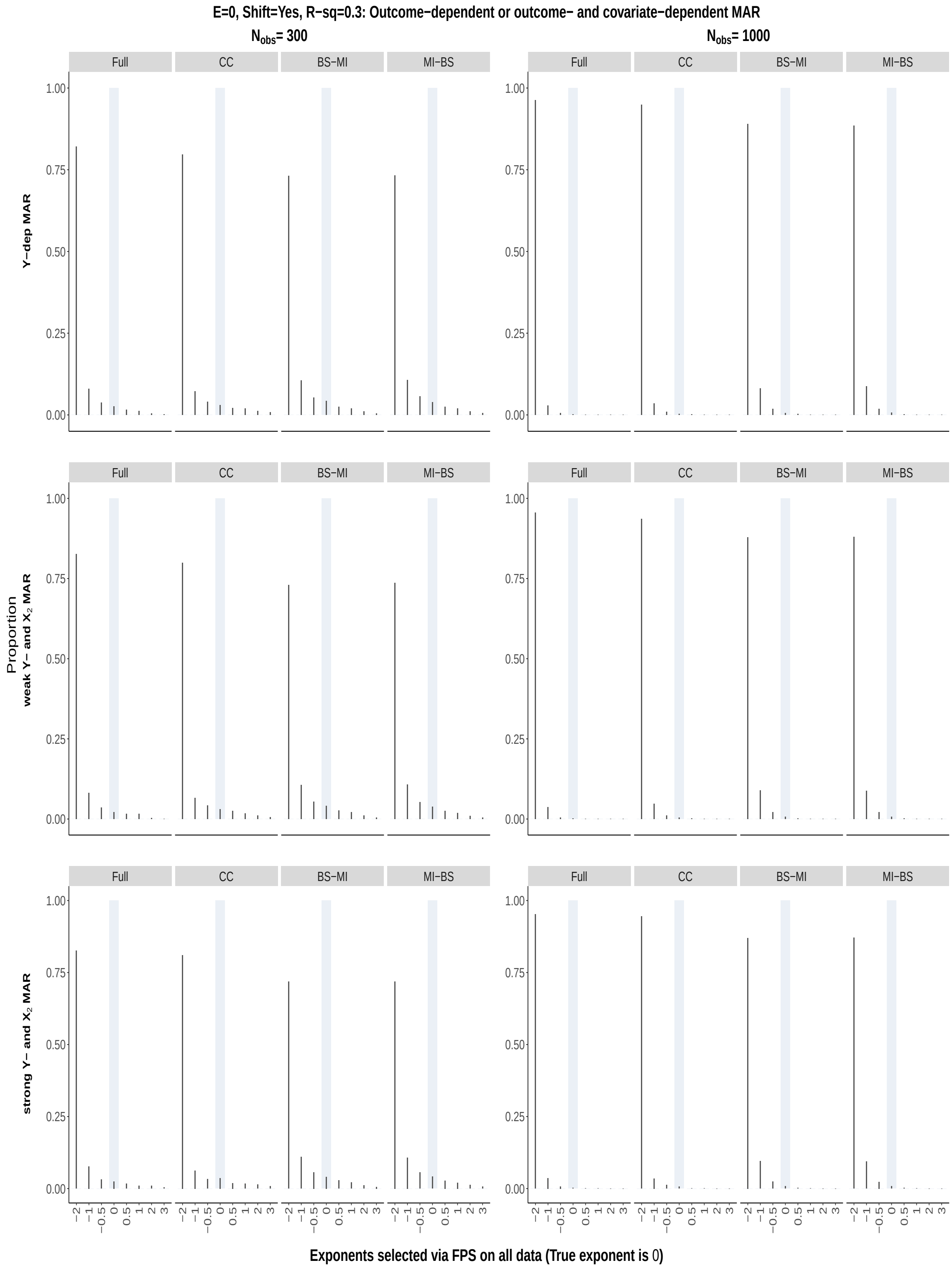

Figure S268: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

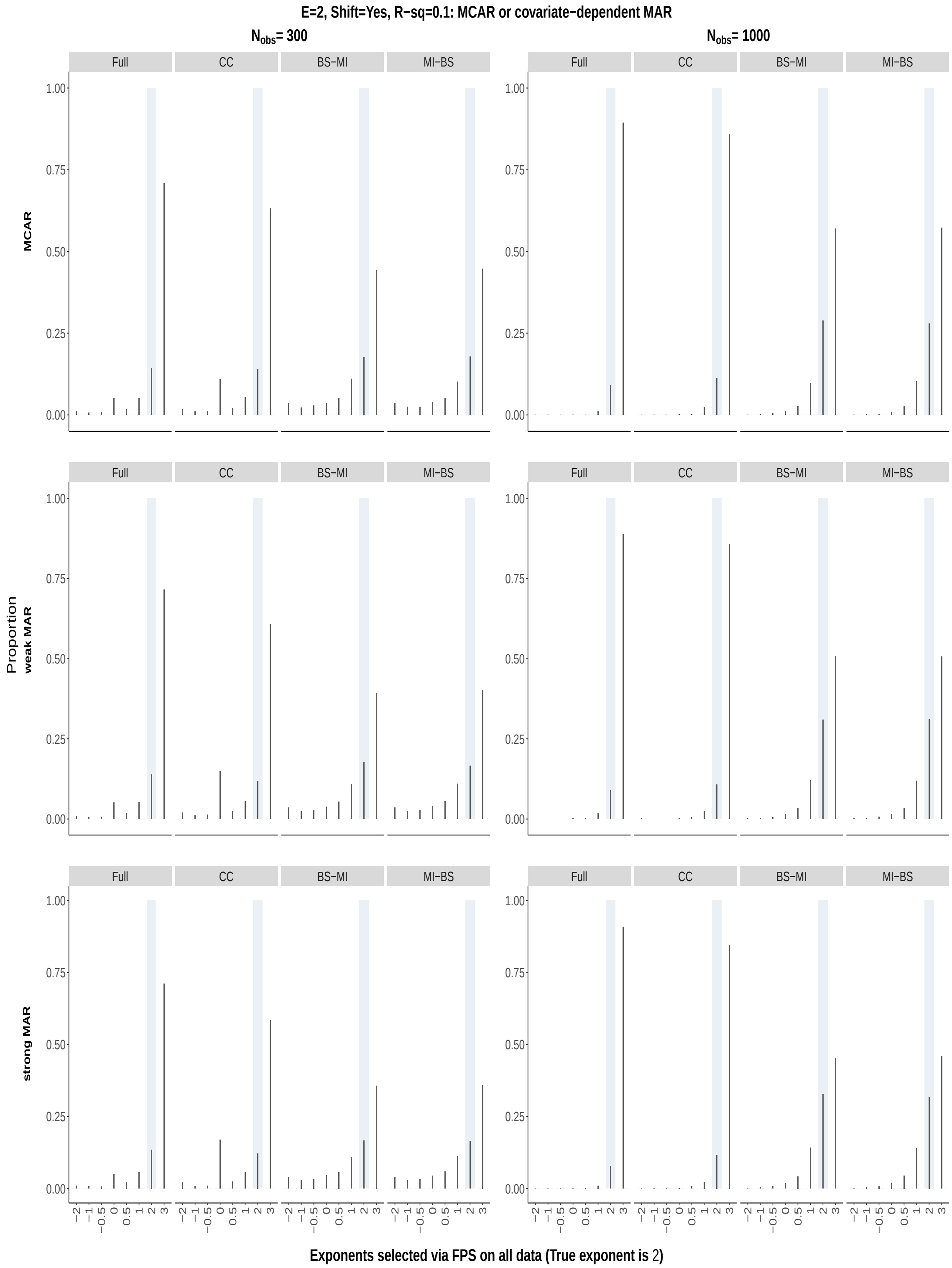

Figure S269: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

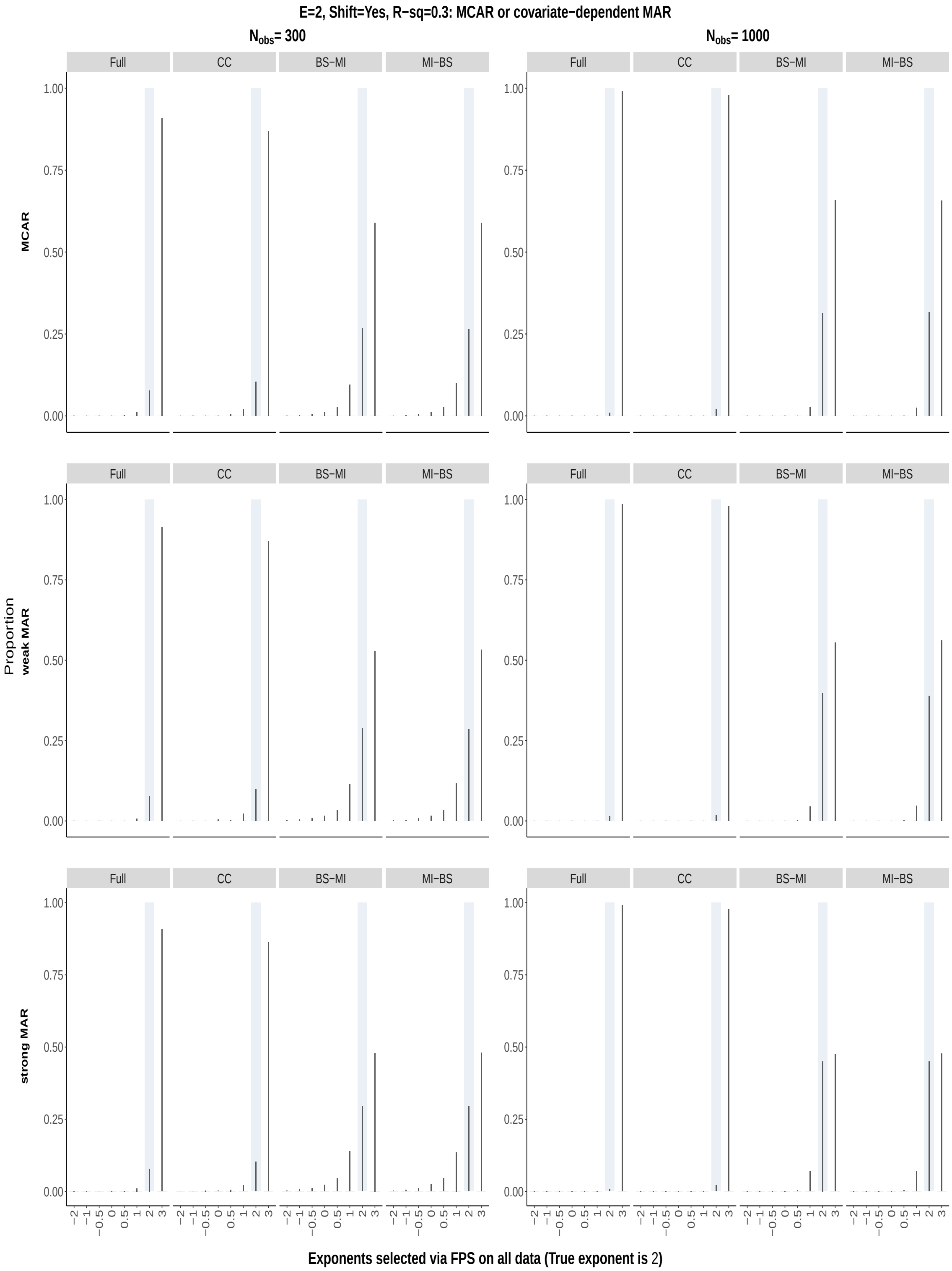

Figure S270: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

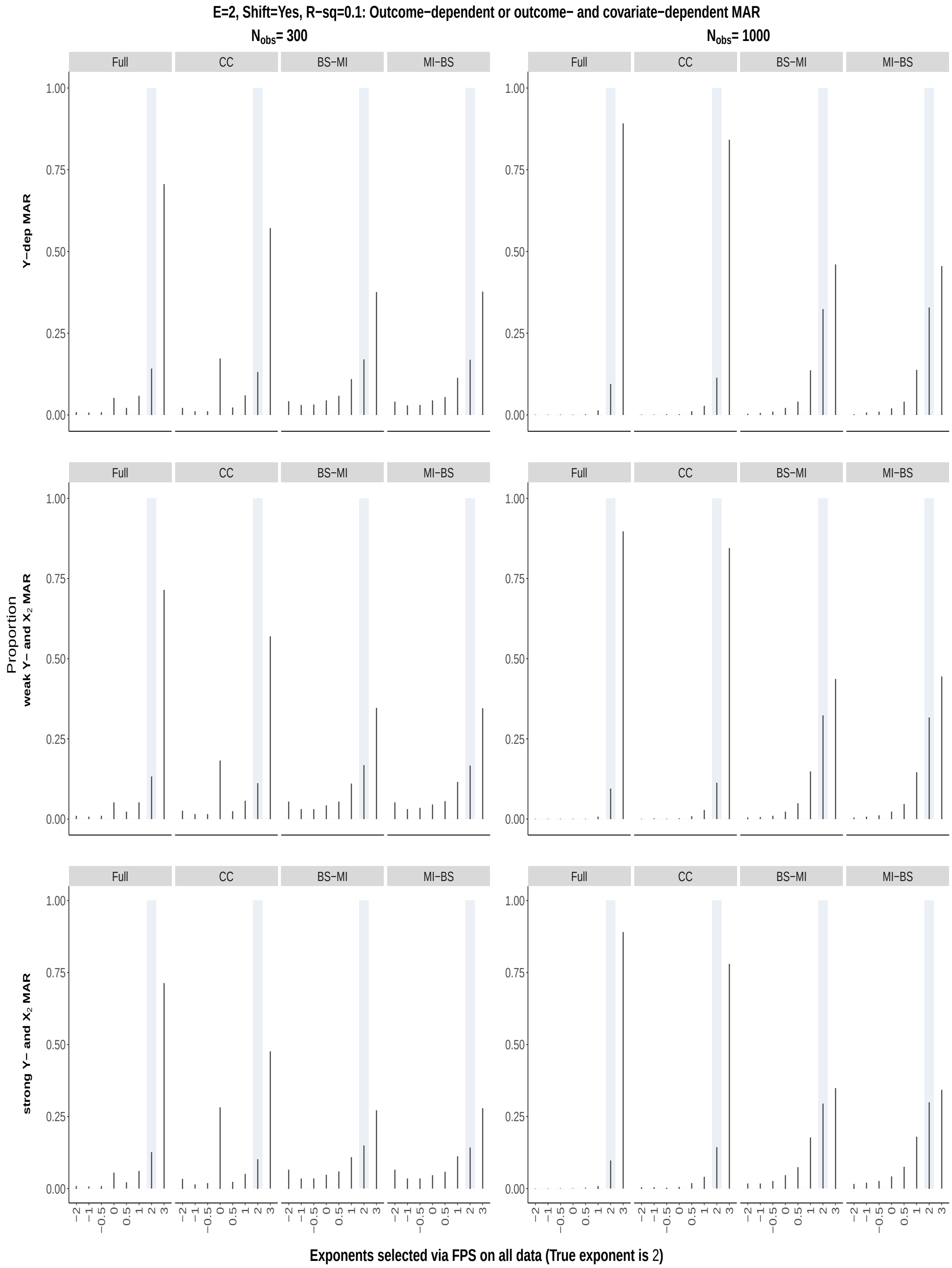

Figure S271: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

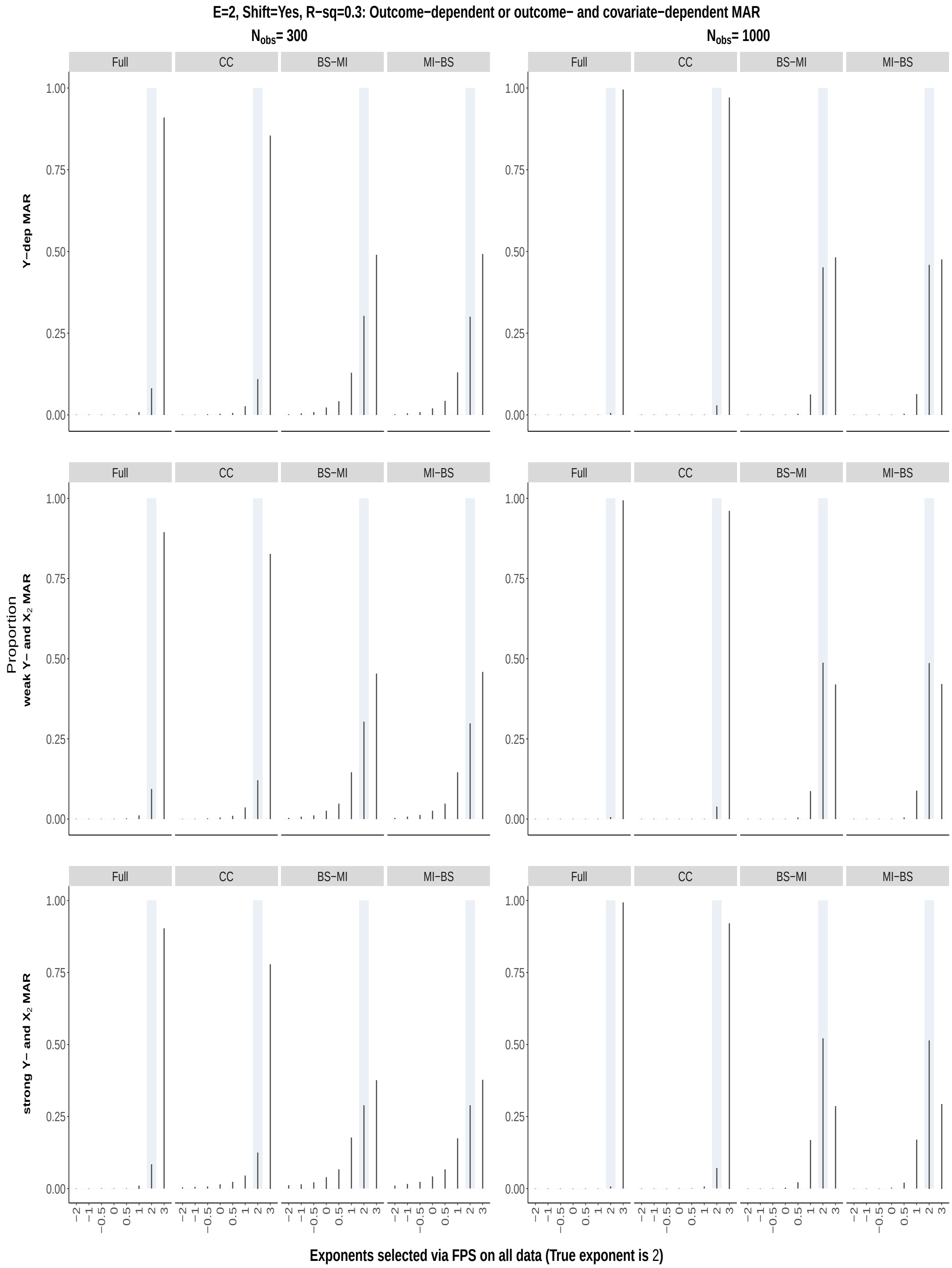

Figure S272: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

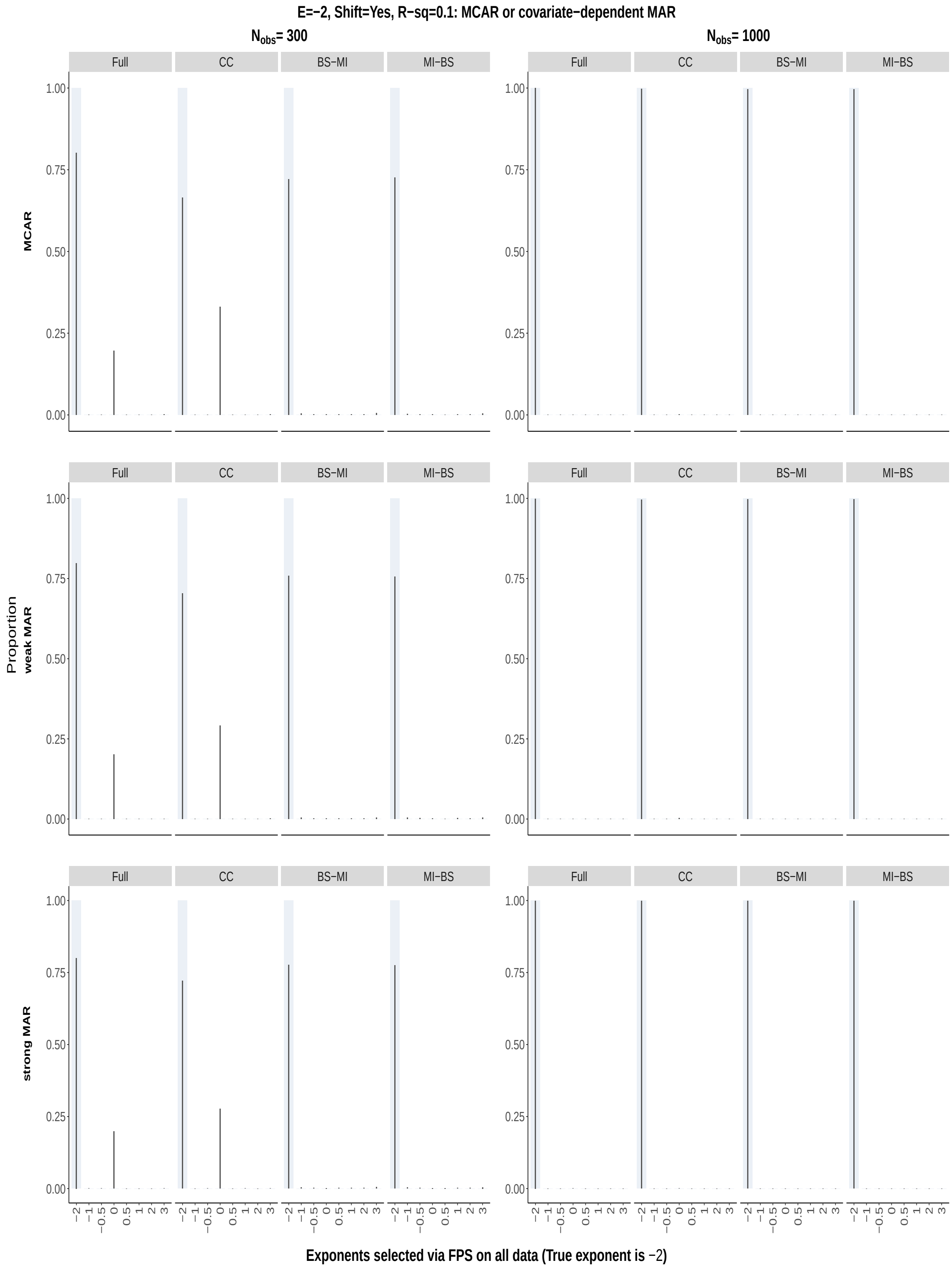

Figure S273: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

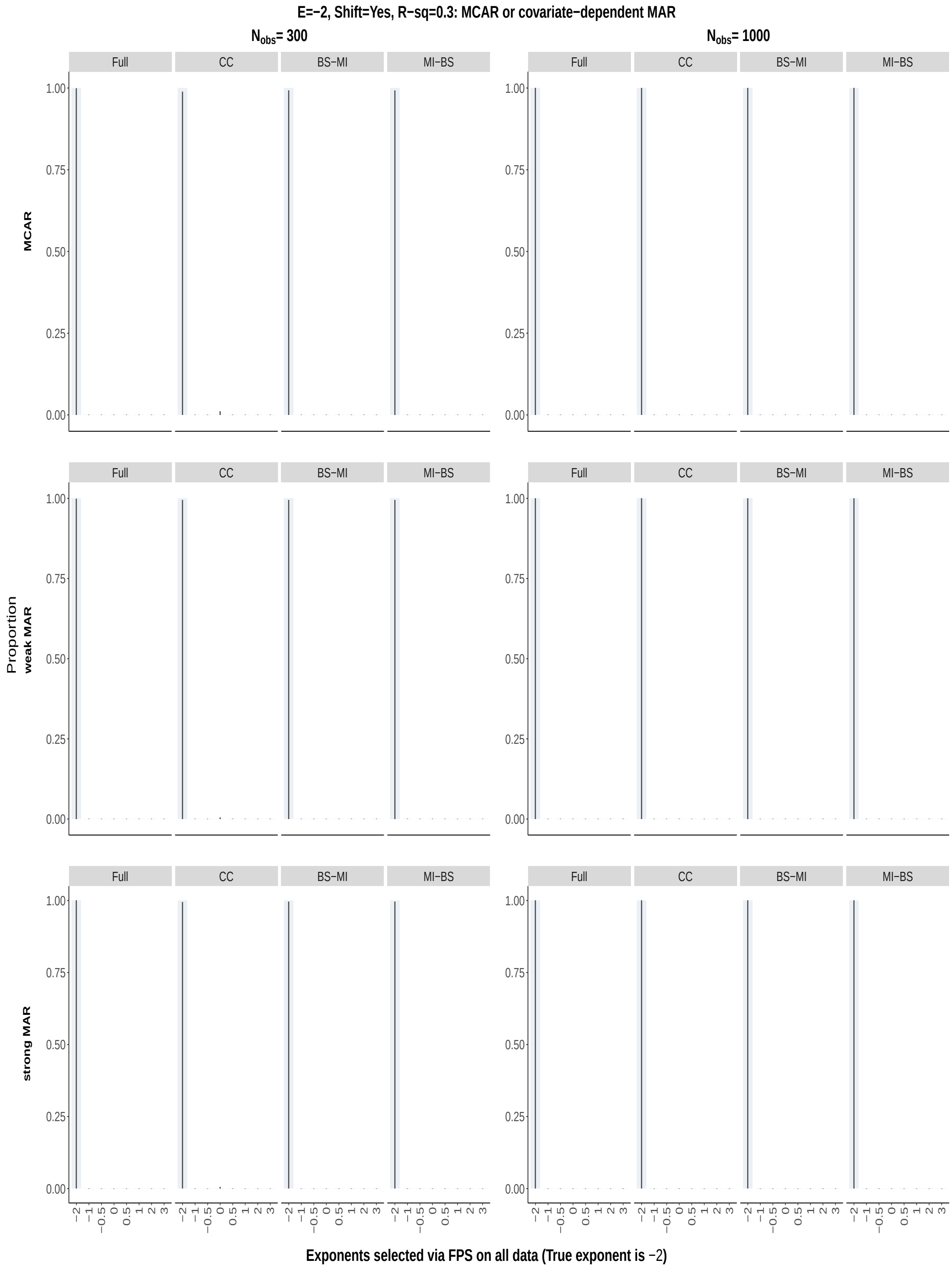

Figure S274: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

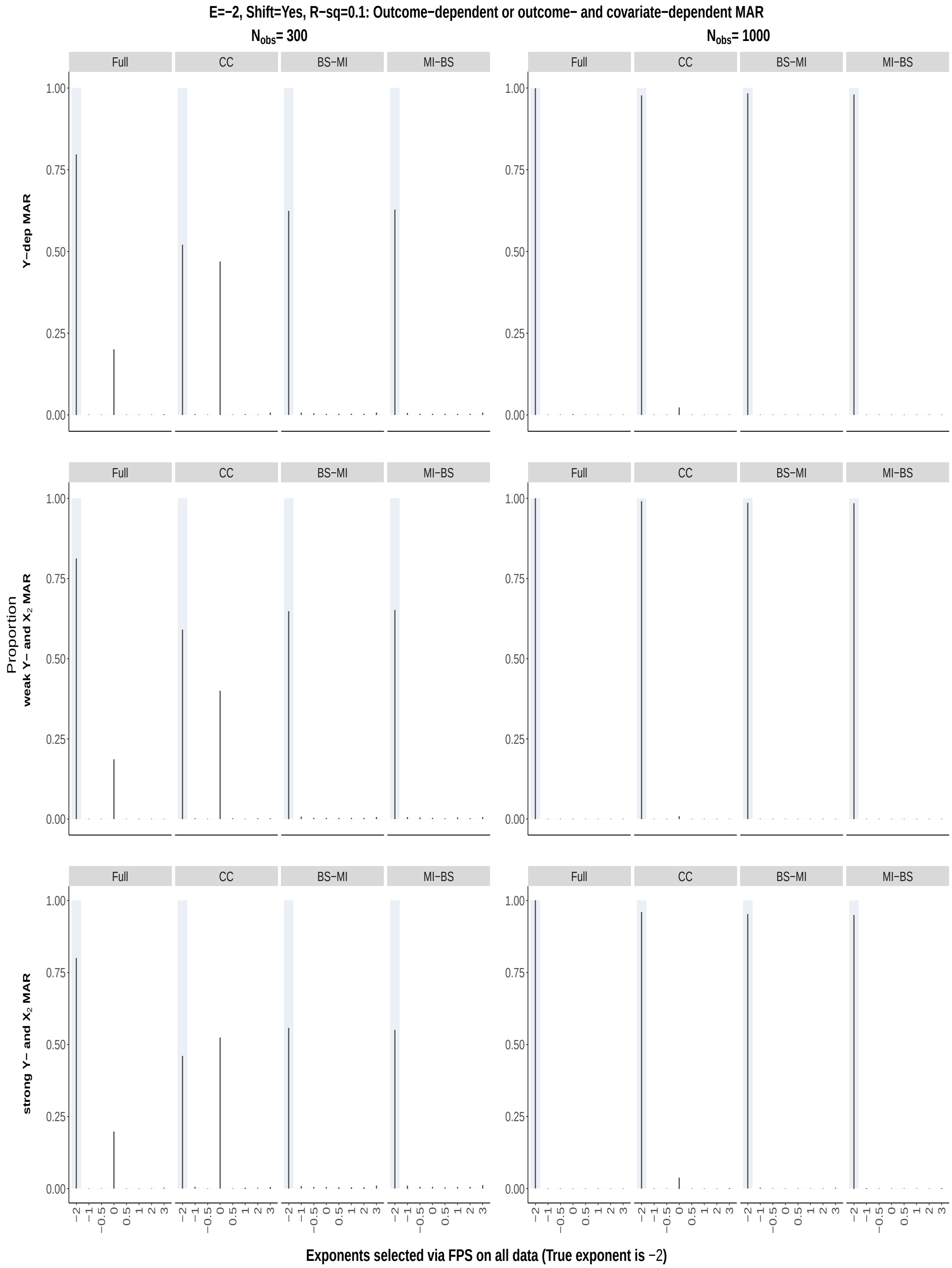

Figure S275: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

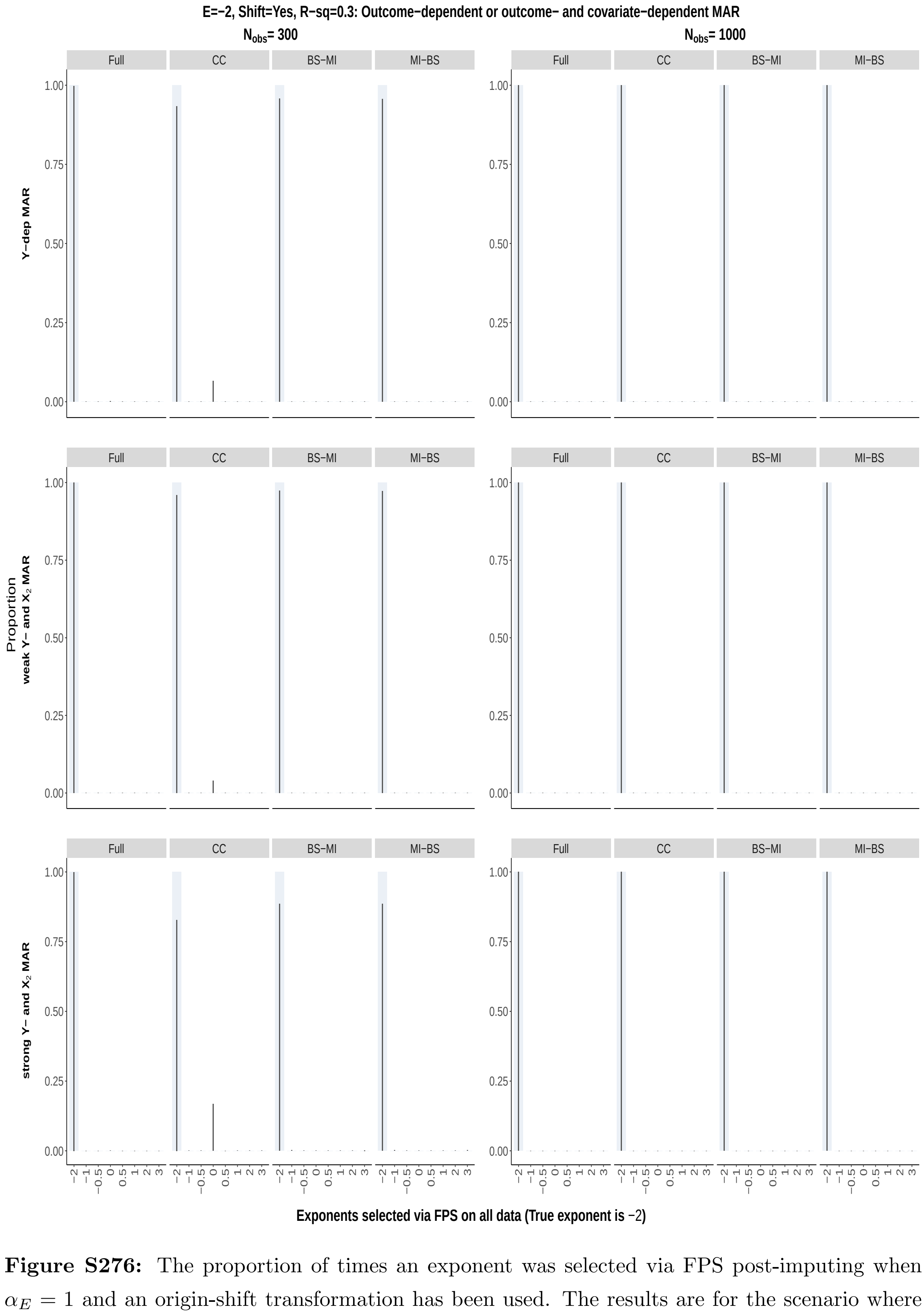

Figure S276: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.20 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

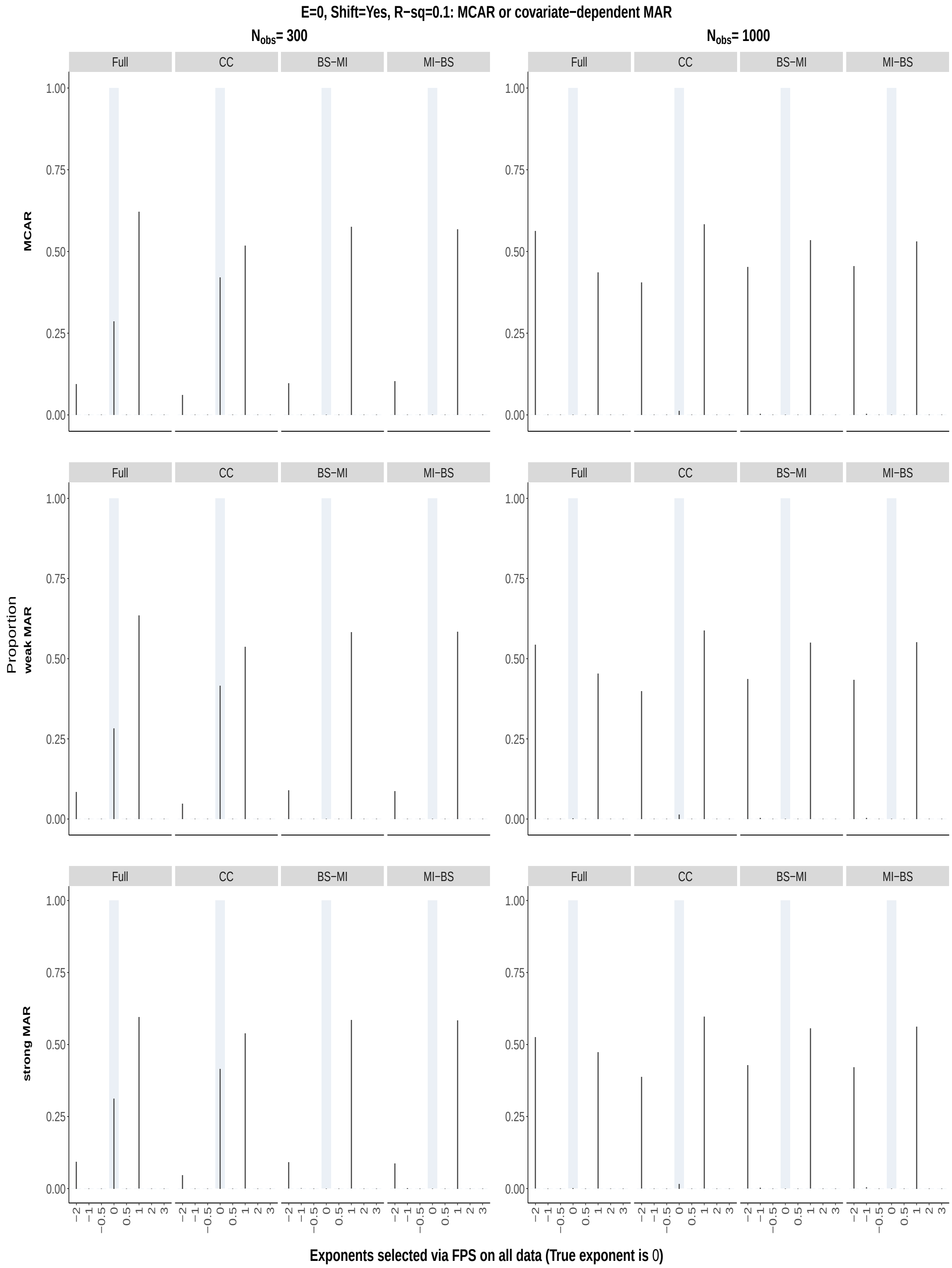

Figure S277: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

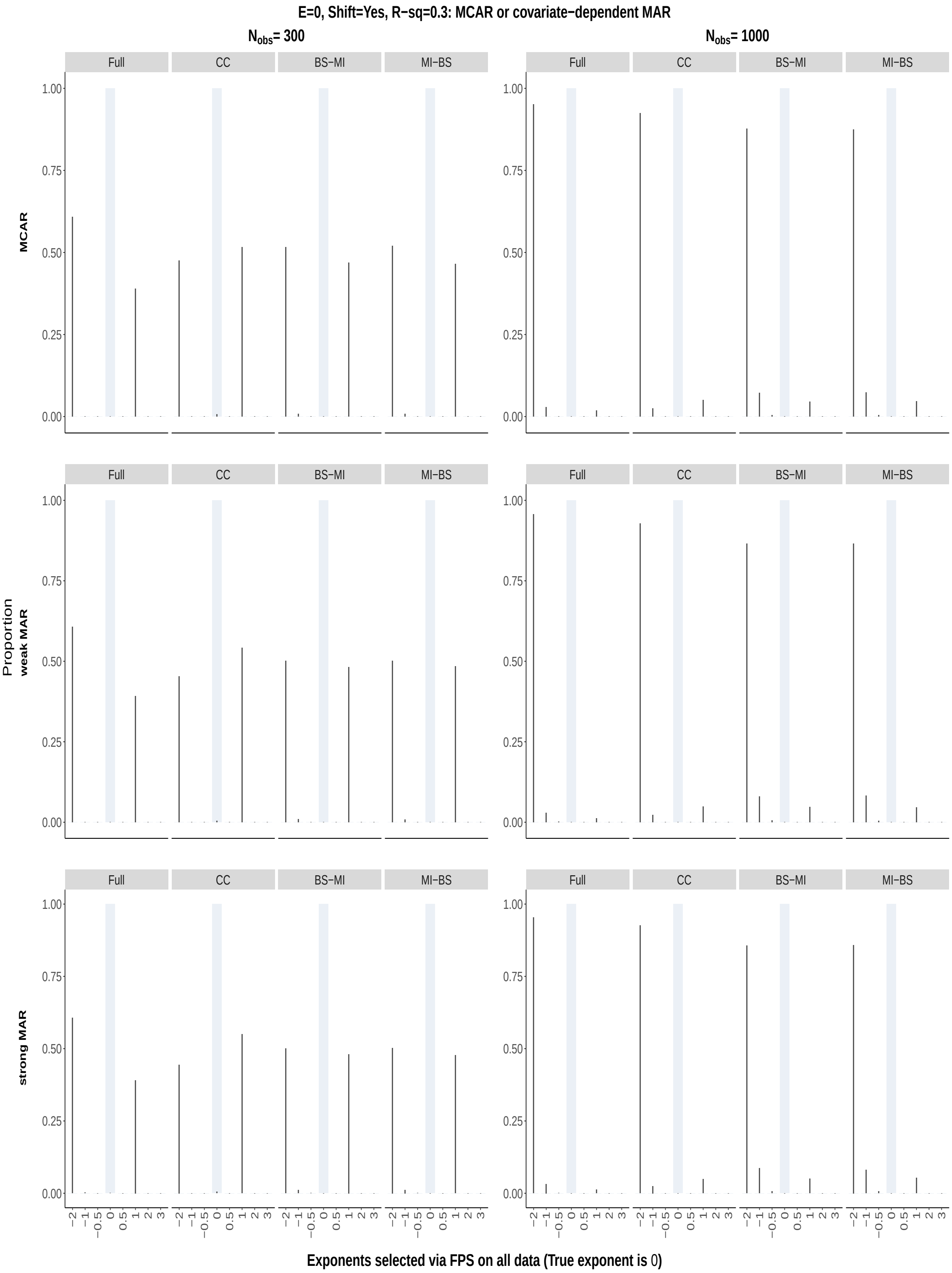

Figure S278: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

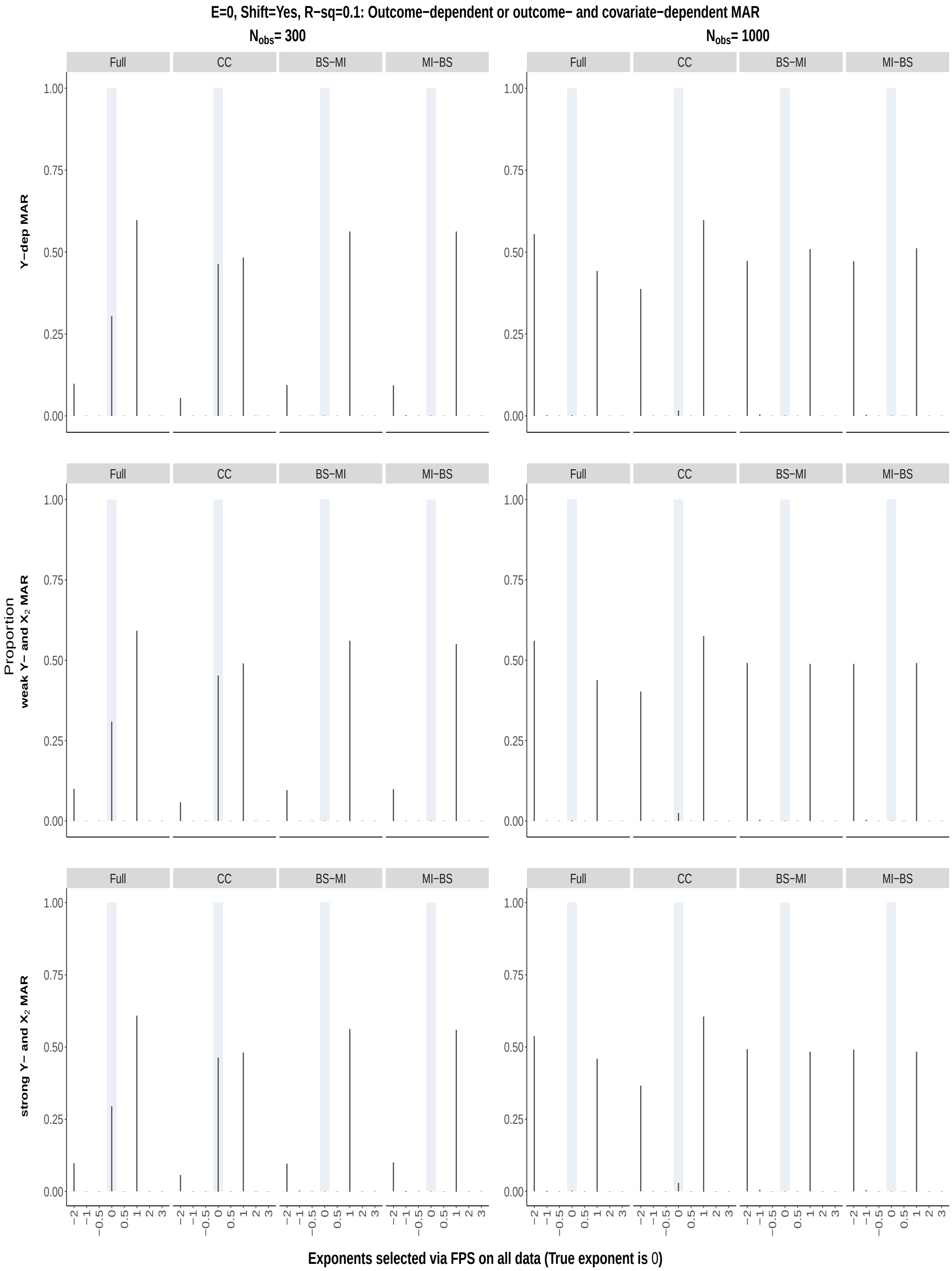

Figure S279: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

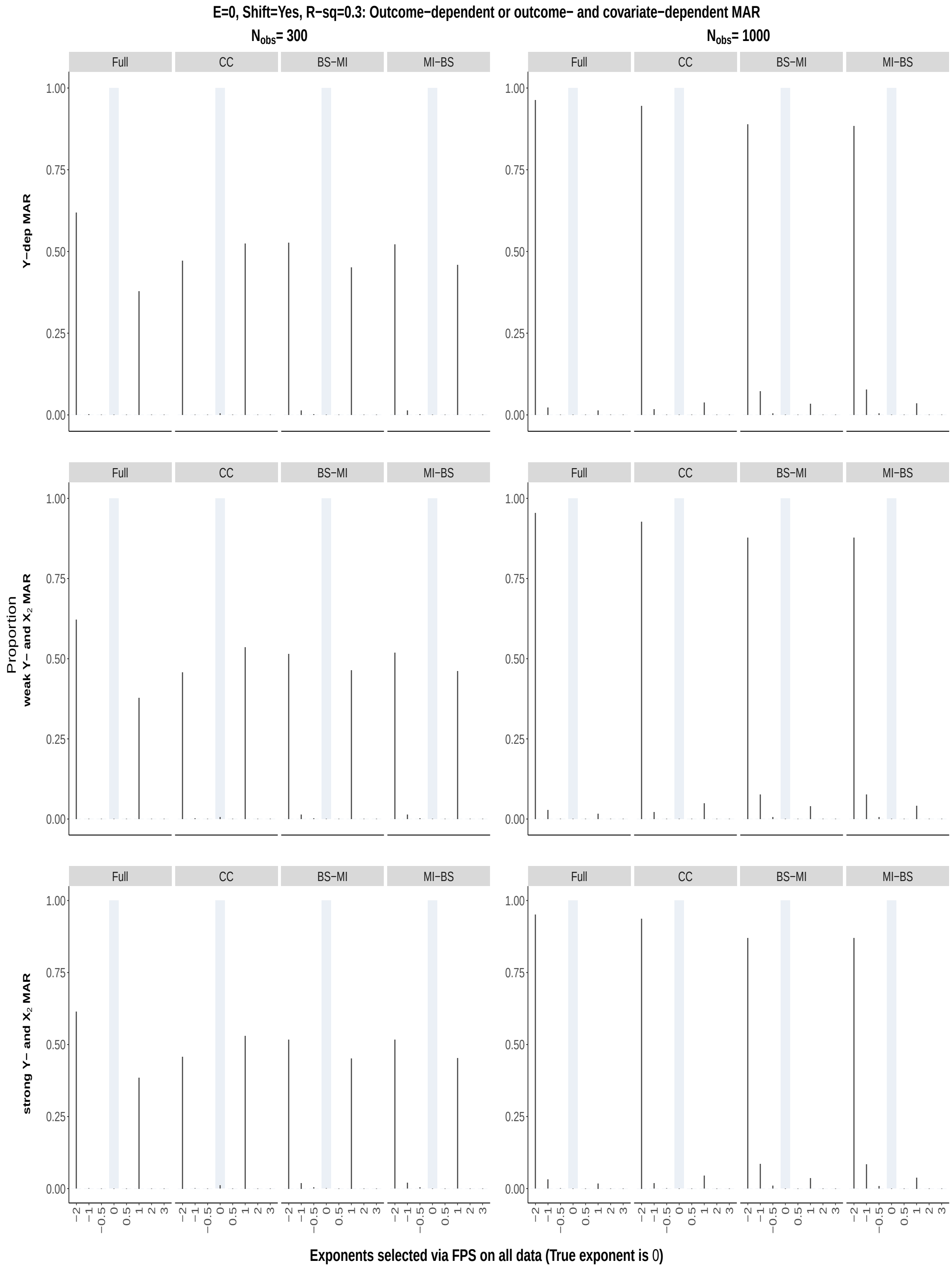

Figure S280: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

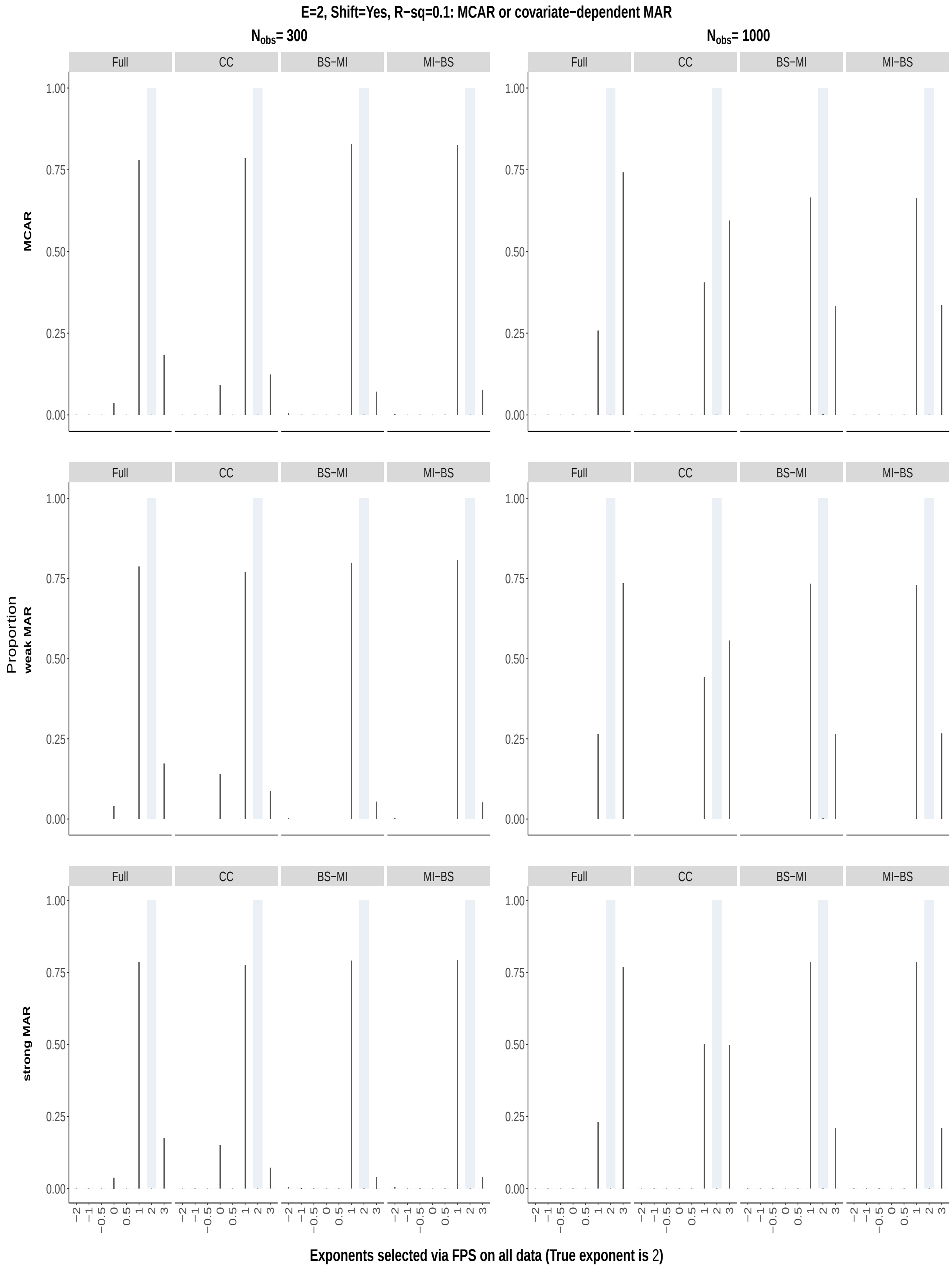

Figure S281: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

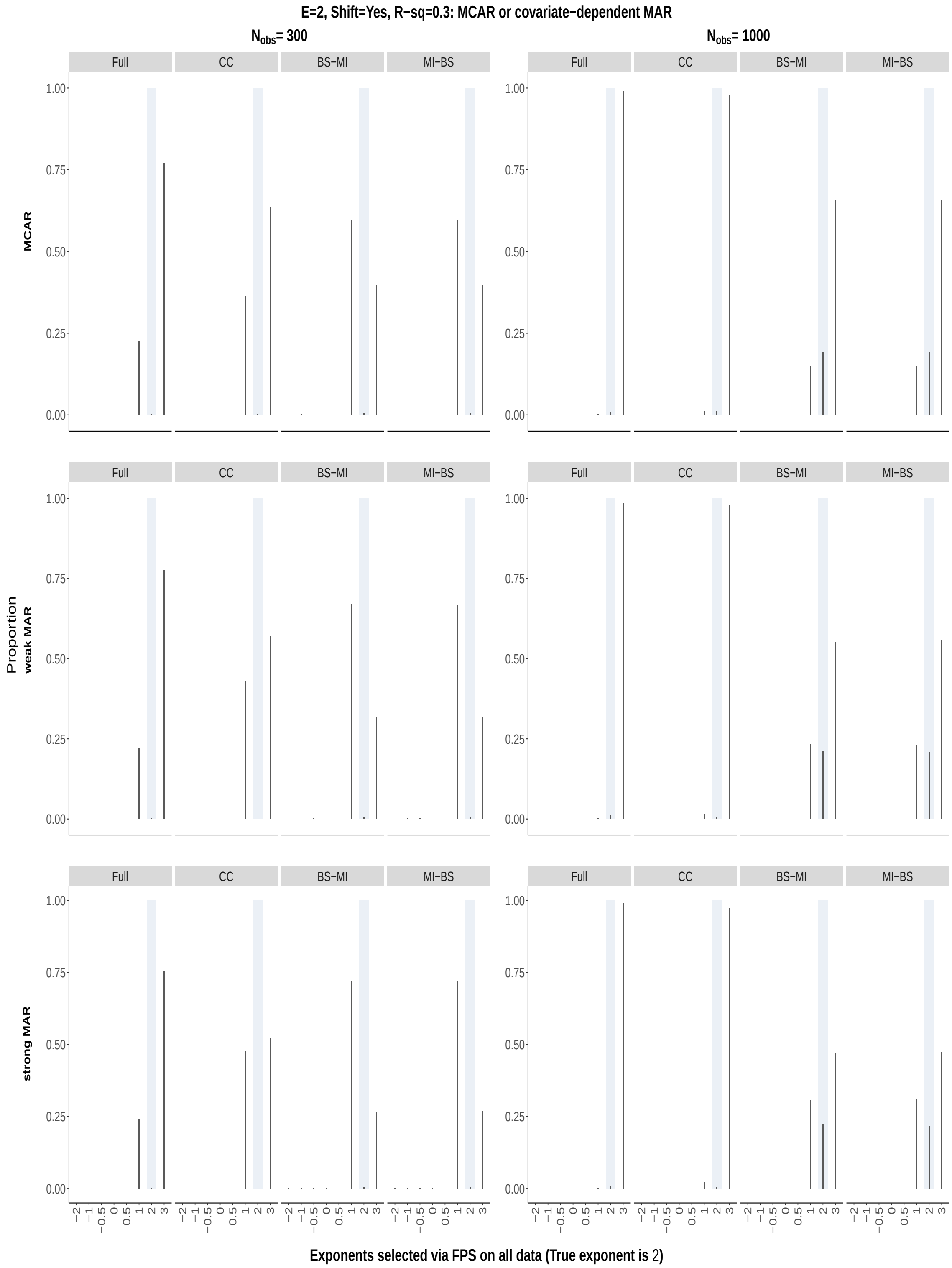

Figure S282: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

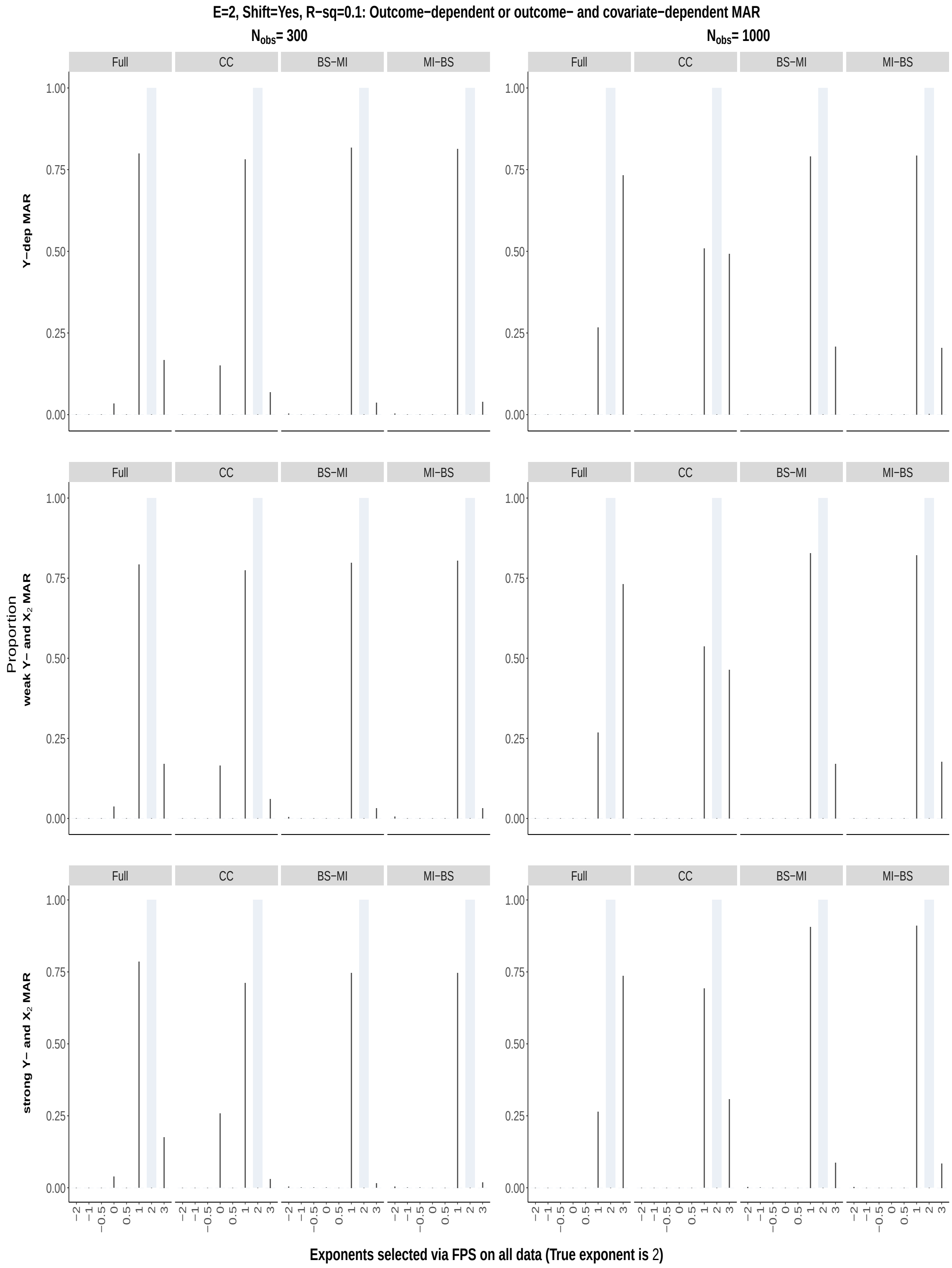

Figure S283: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

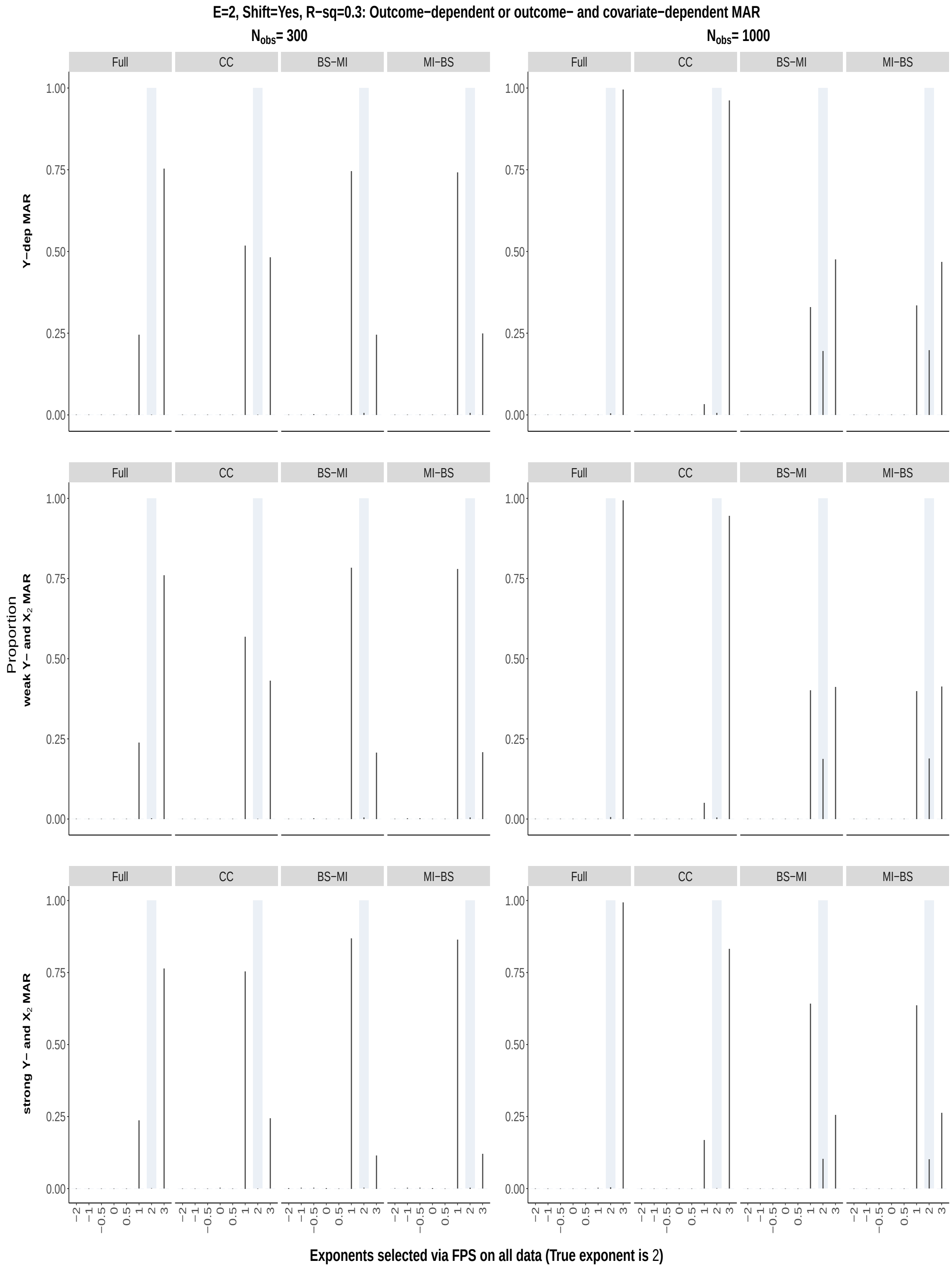

Figure S284: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

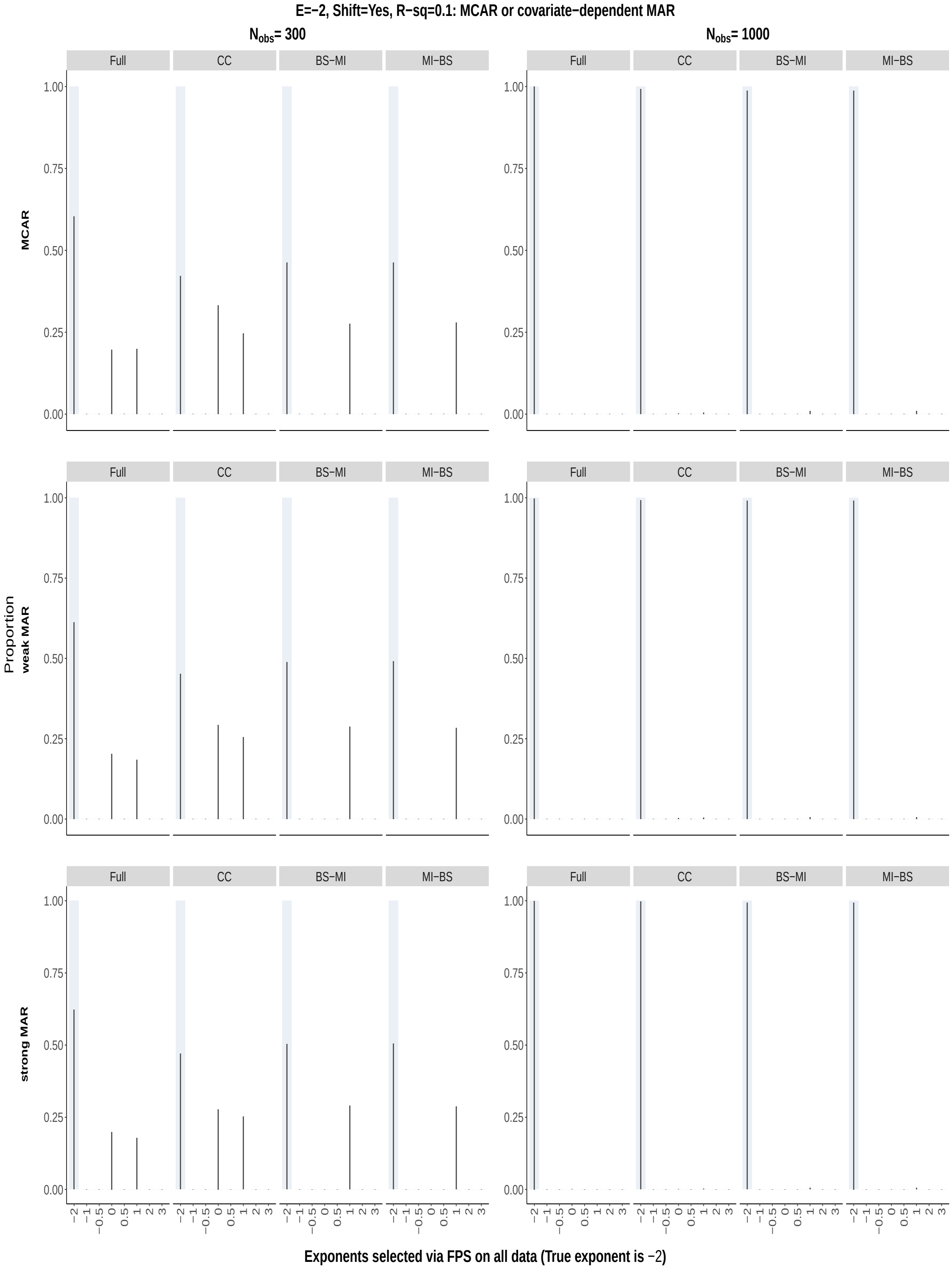

Figure S285: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

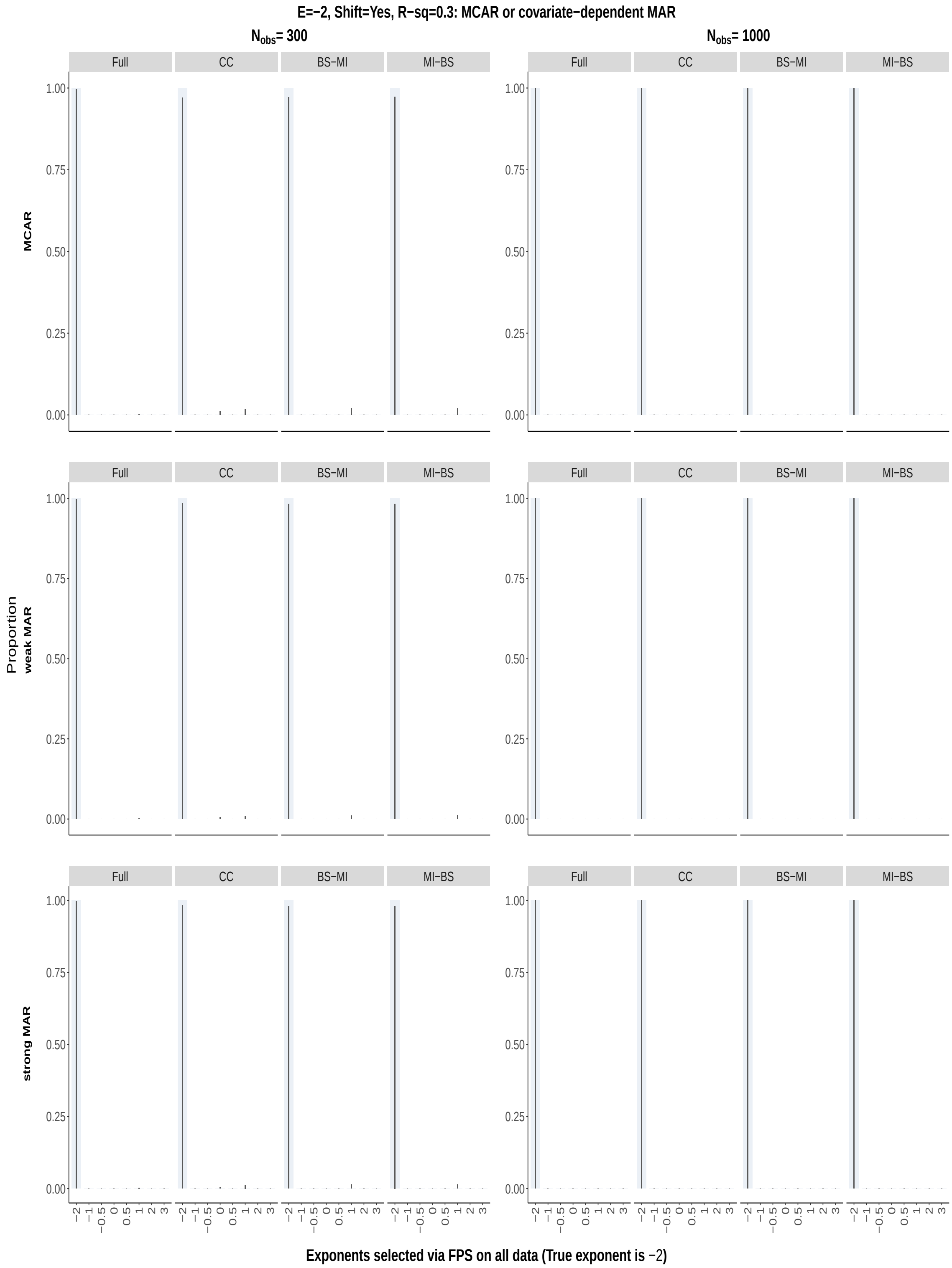

Figure S286: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

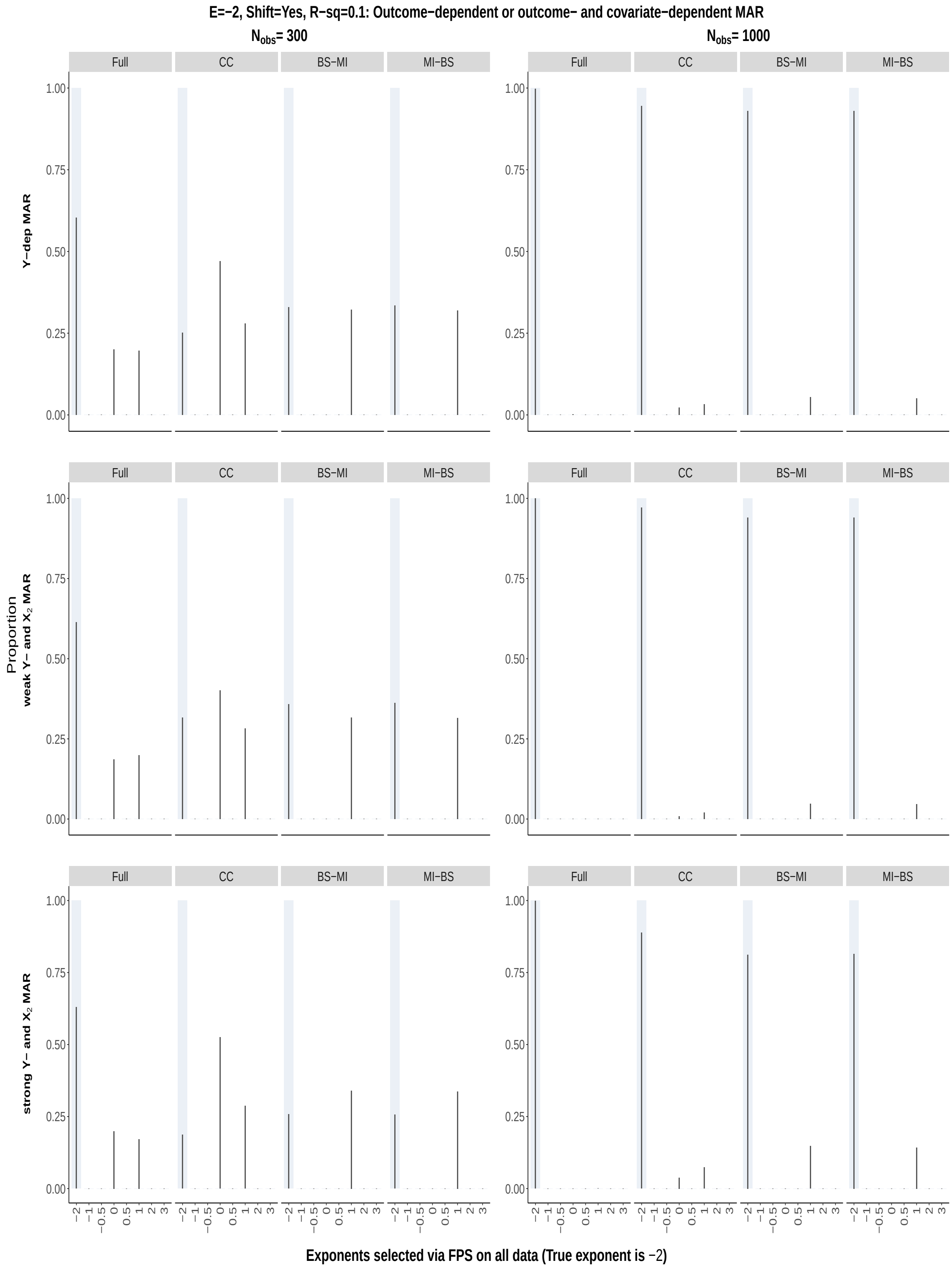

Figure S287: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

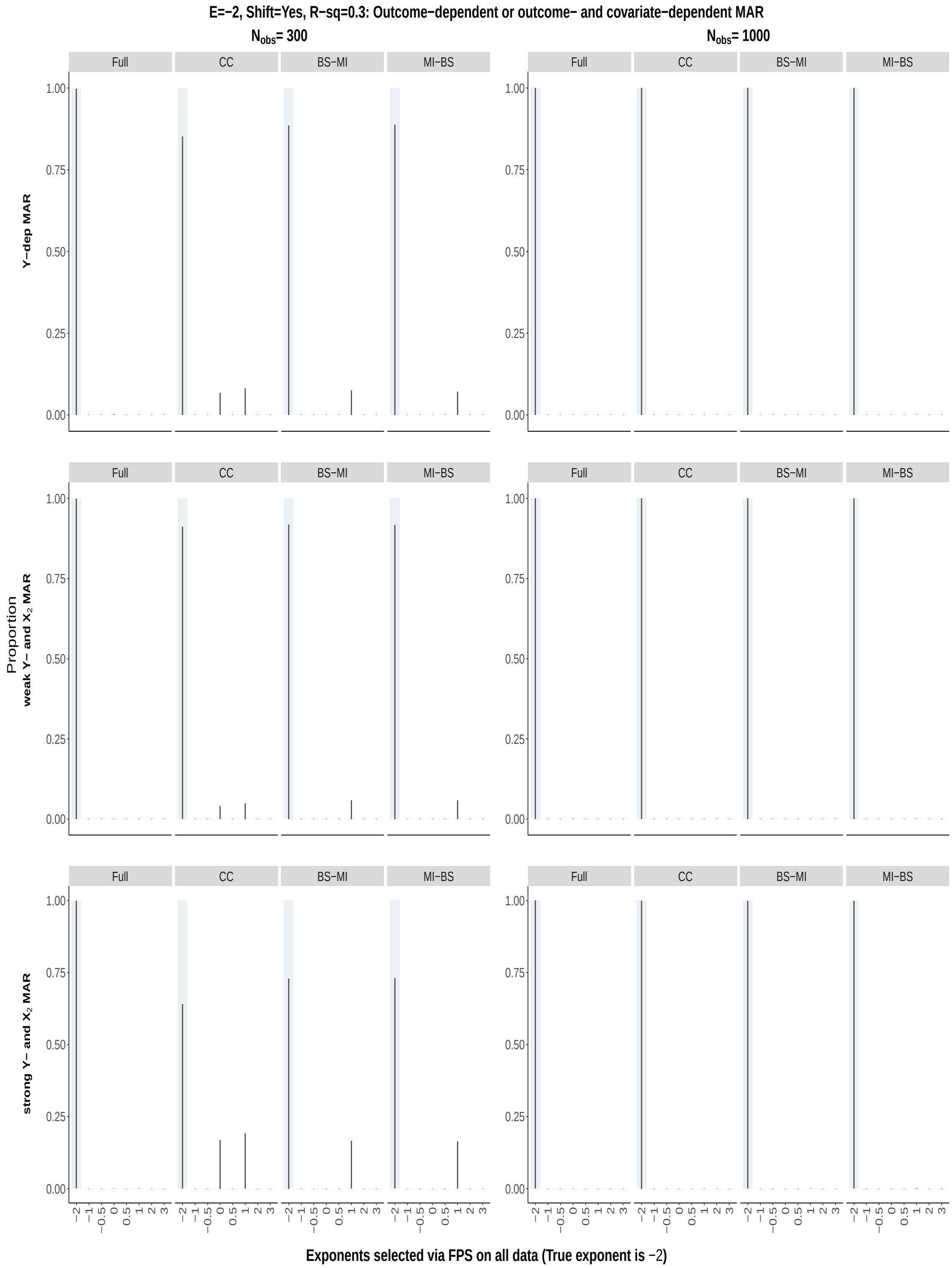

Figure S288: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.21 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 0$, $\alpha_E = 1$ and no origin-shift

True exponent is 0

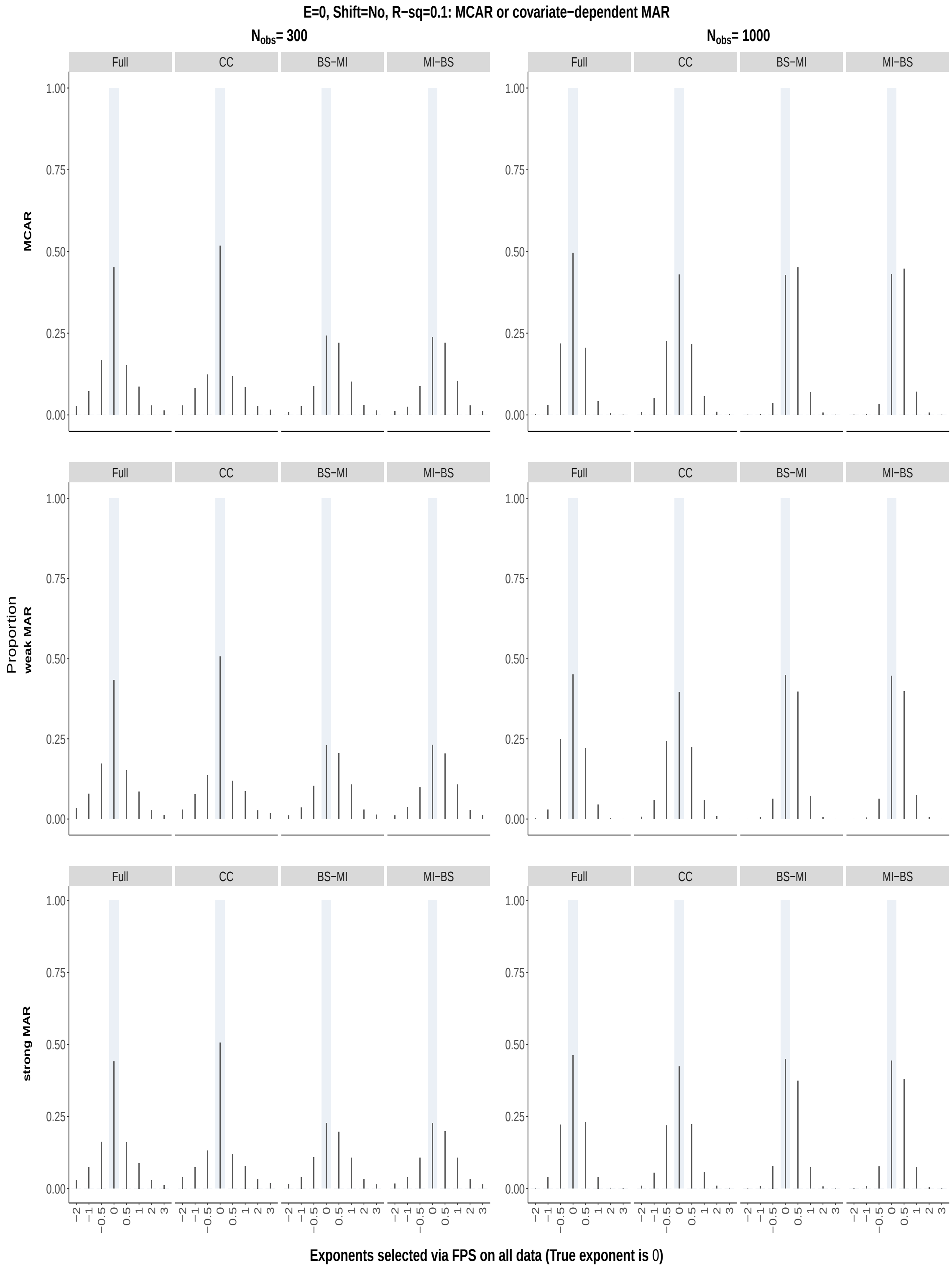

Figure S289: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

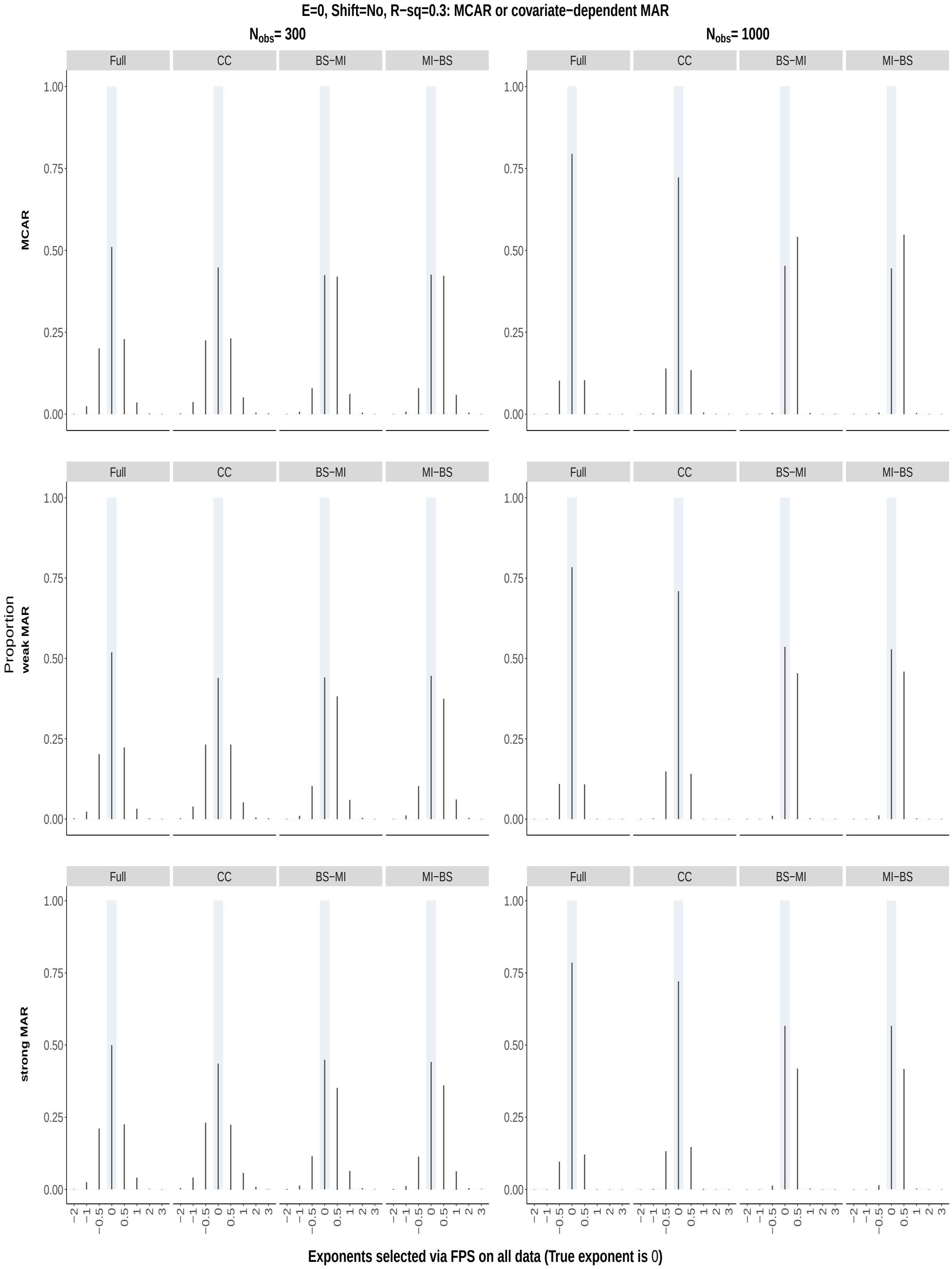

Figure S290: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

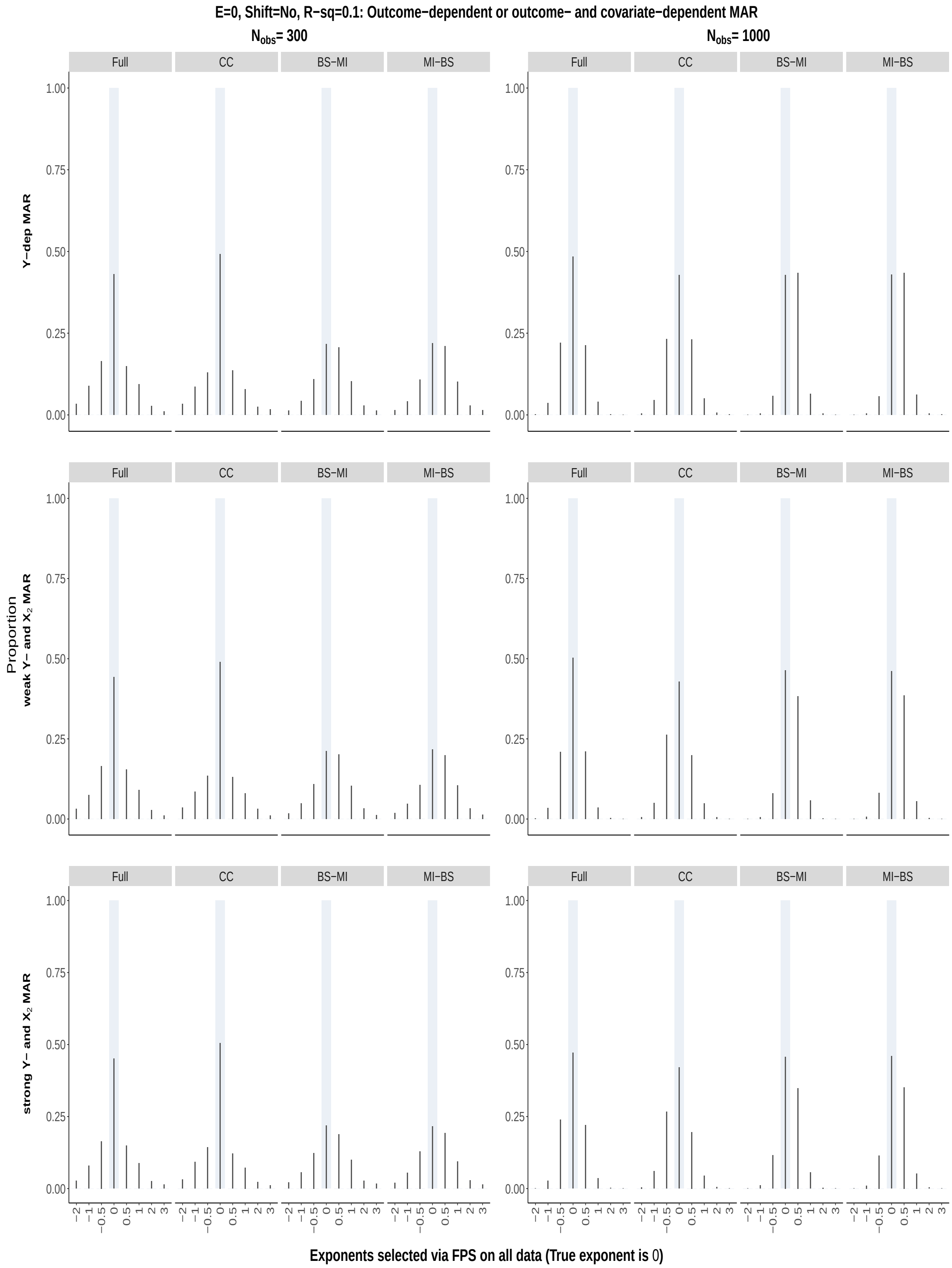

Figure S291: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

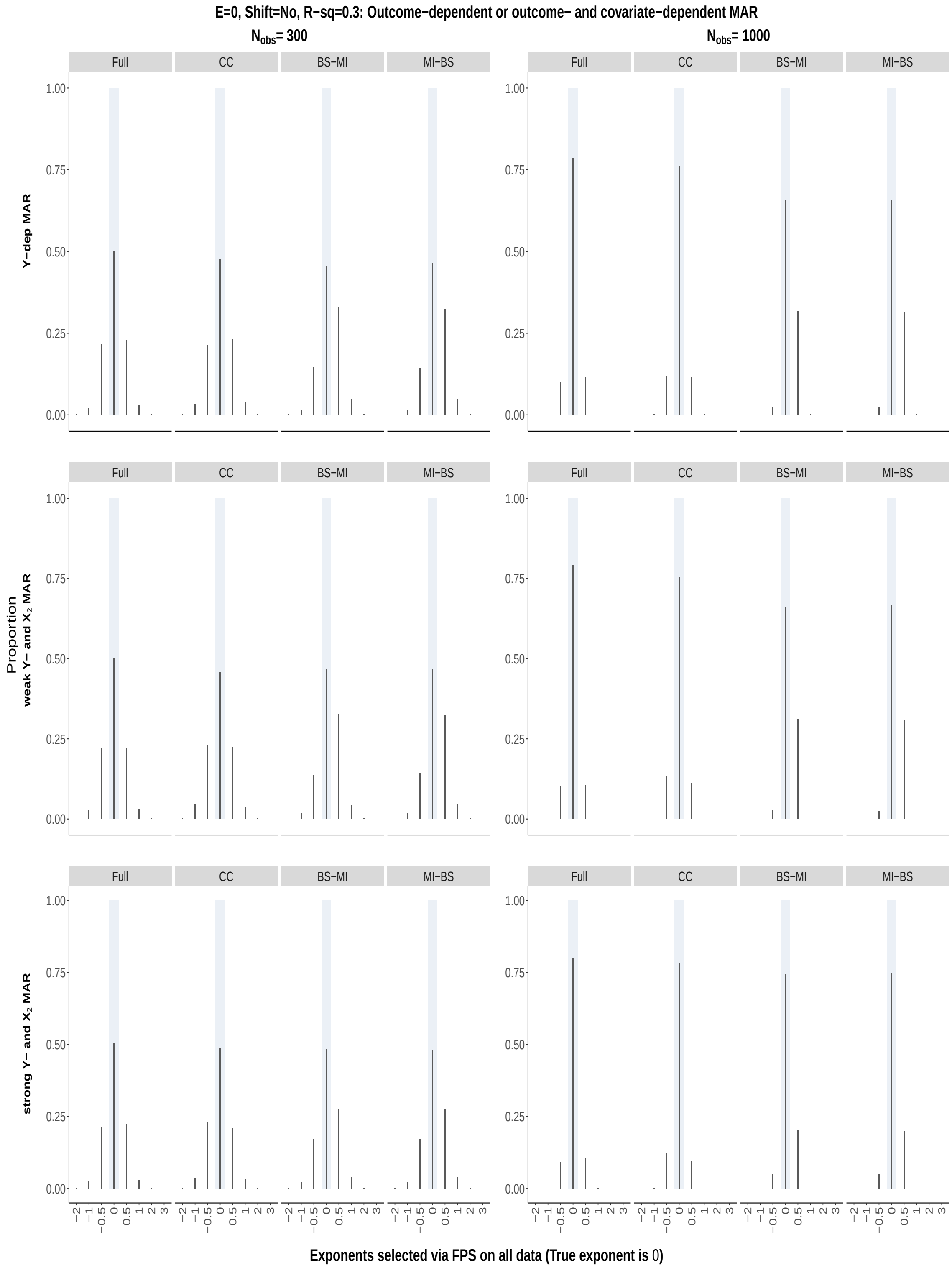

Figure S292: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

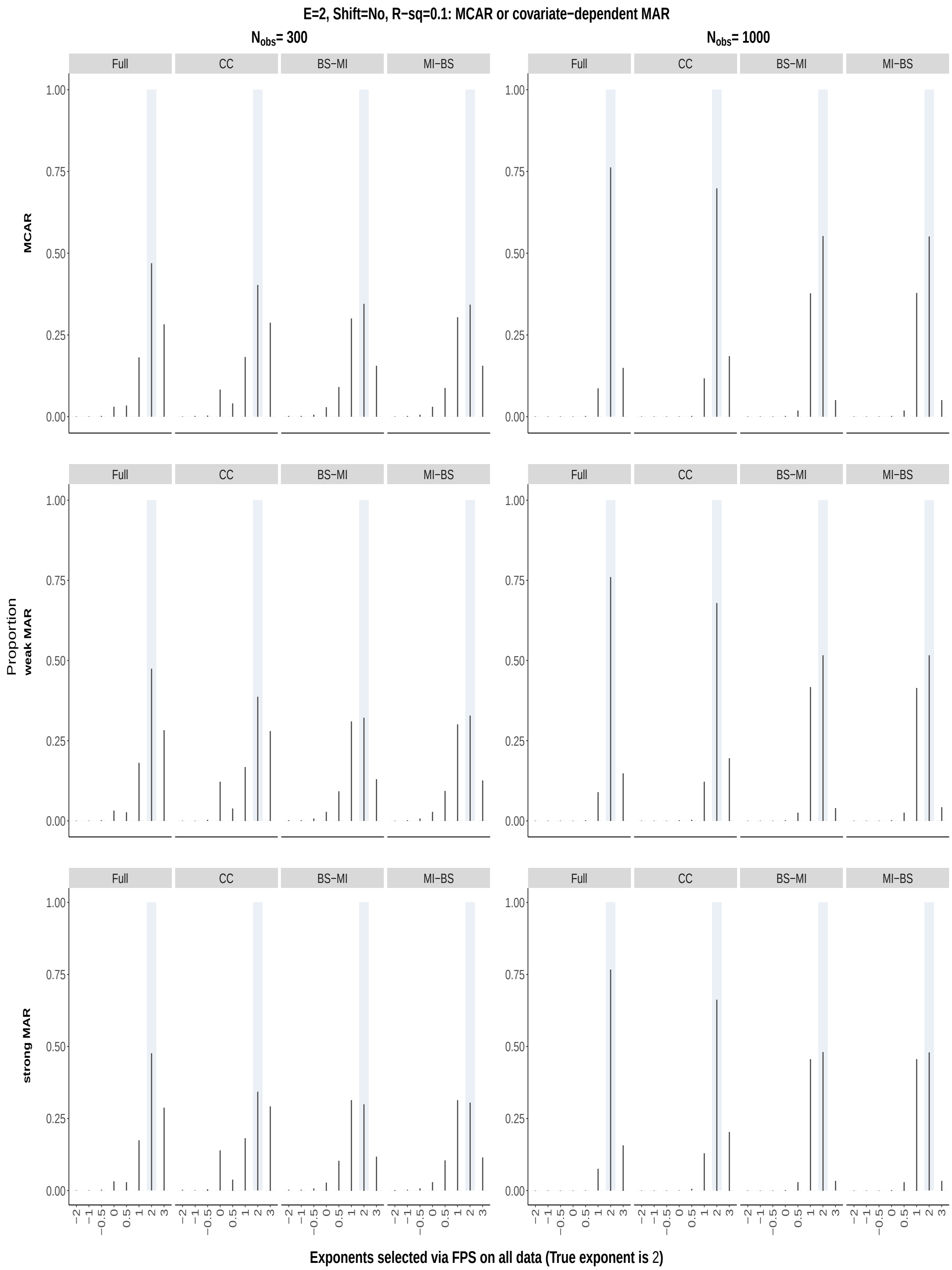

Figure S293: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

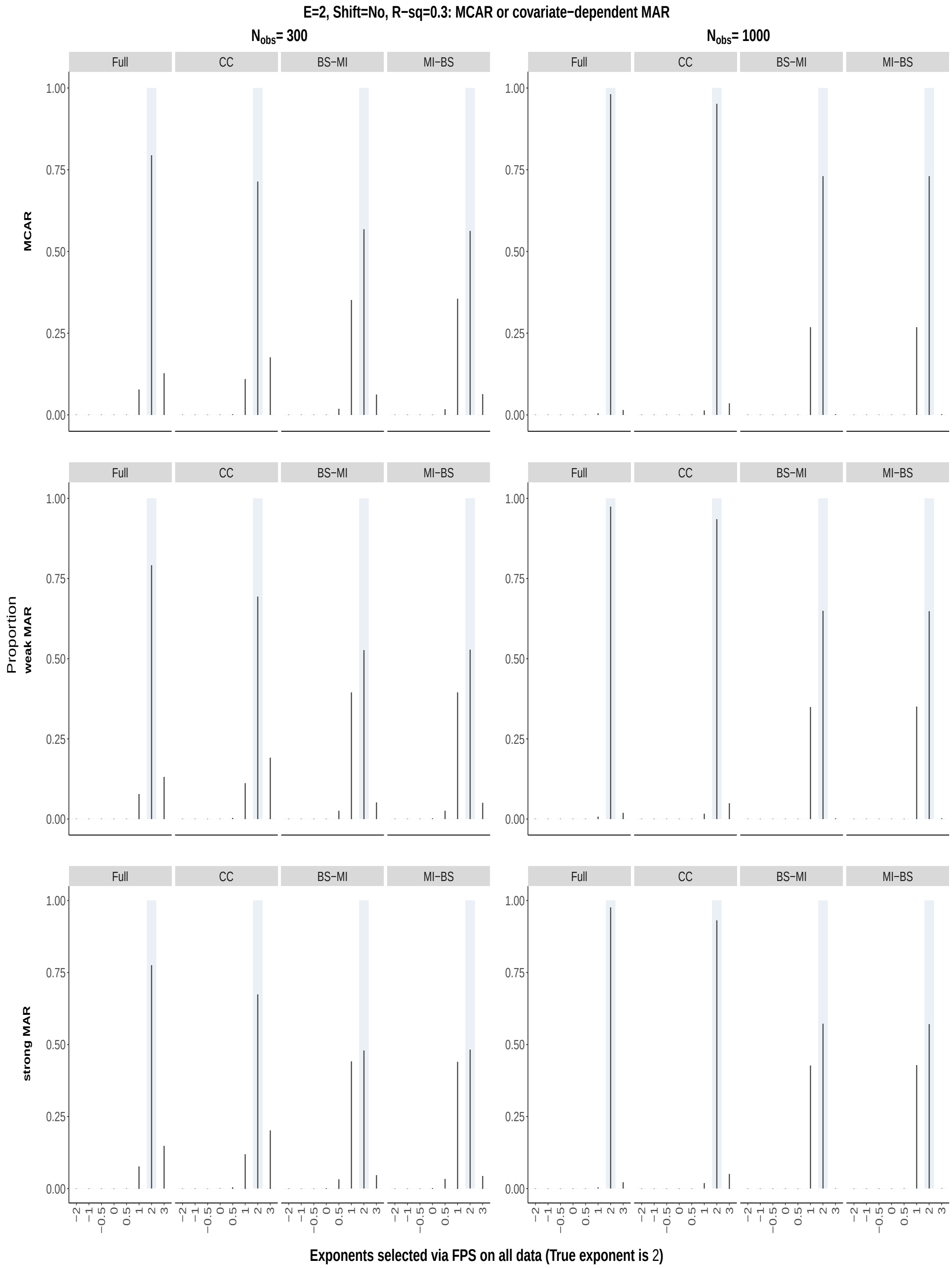

Figure S294: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

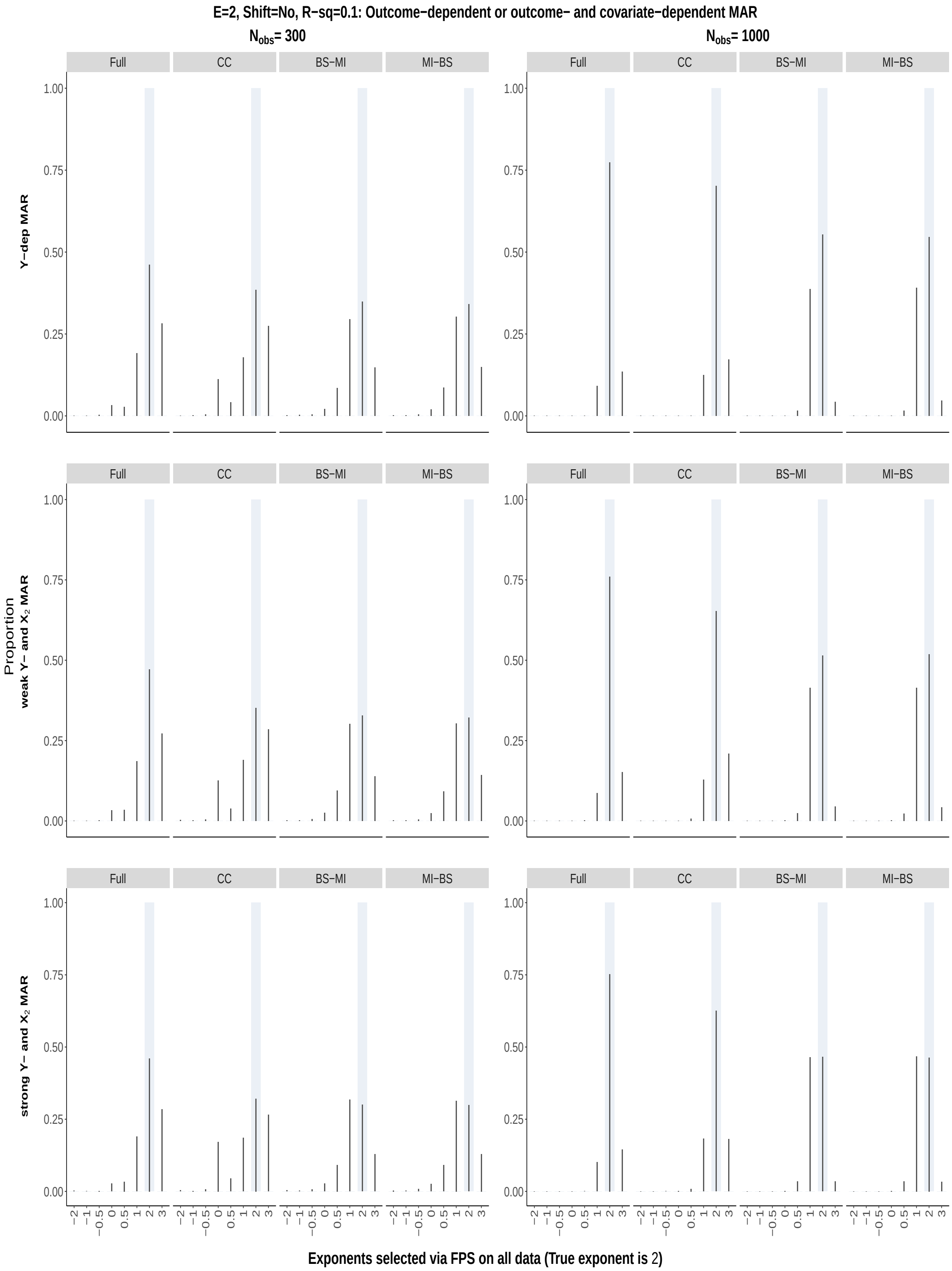

Figure S295: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

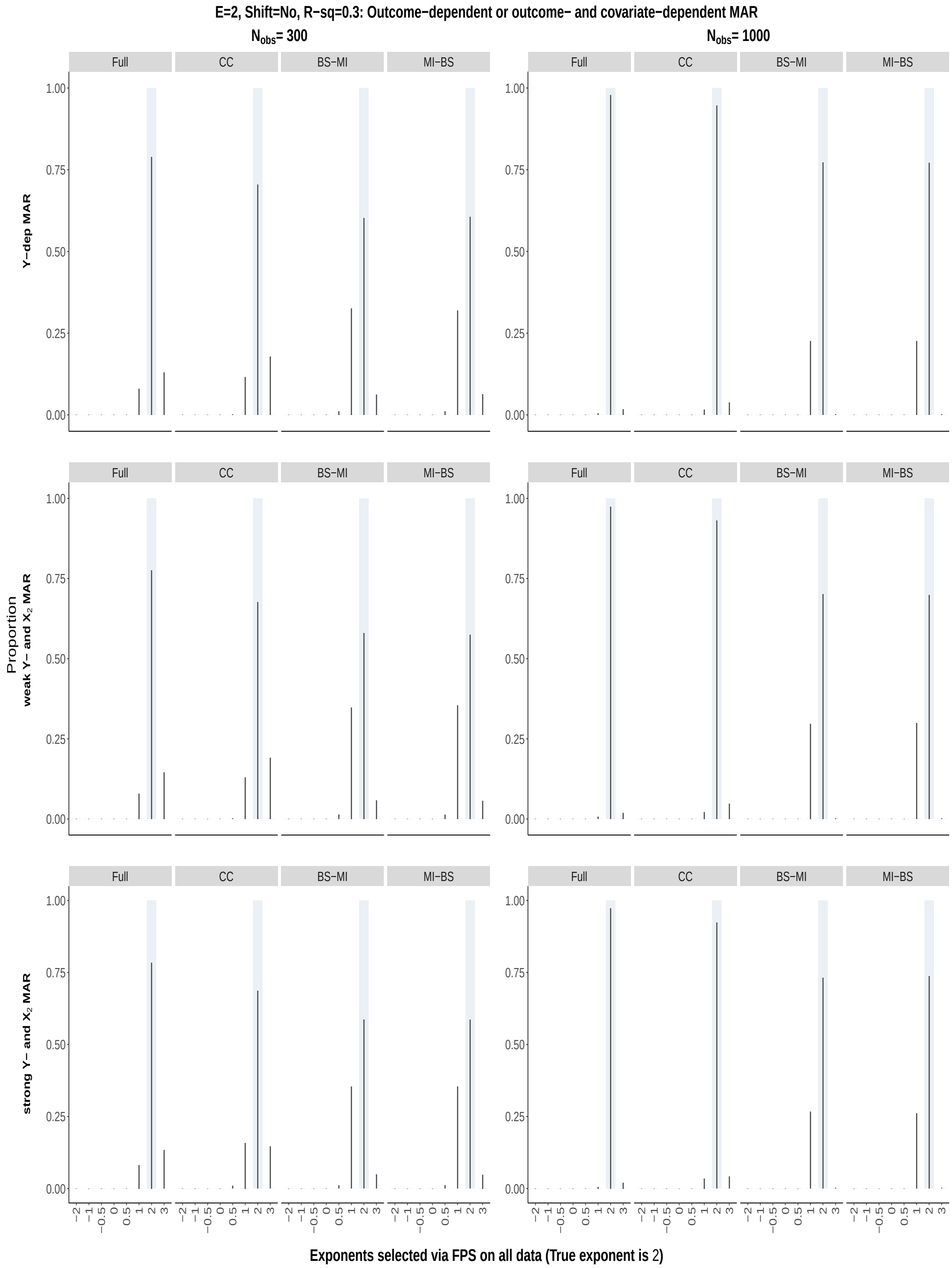

Figure S296: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

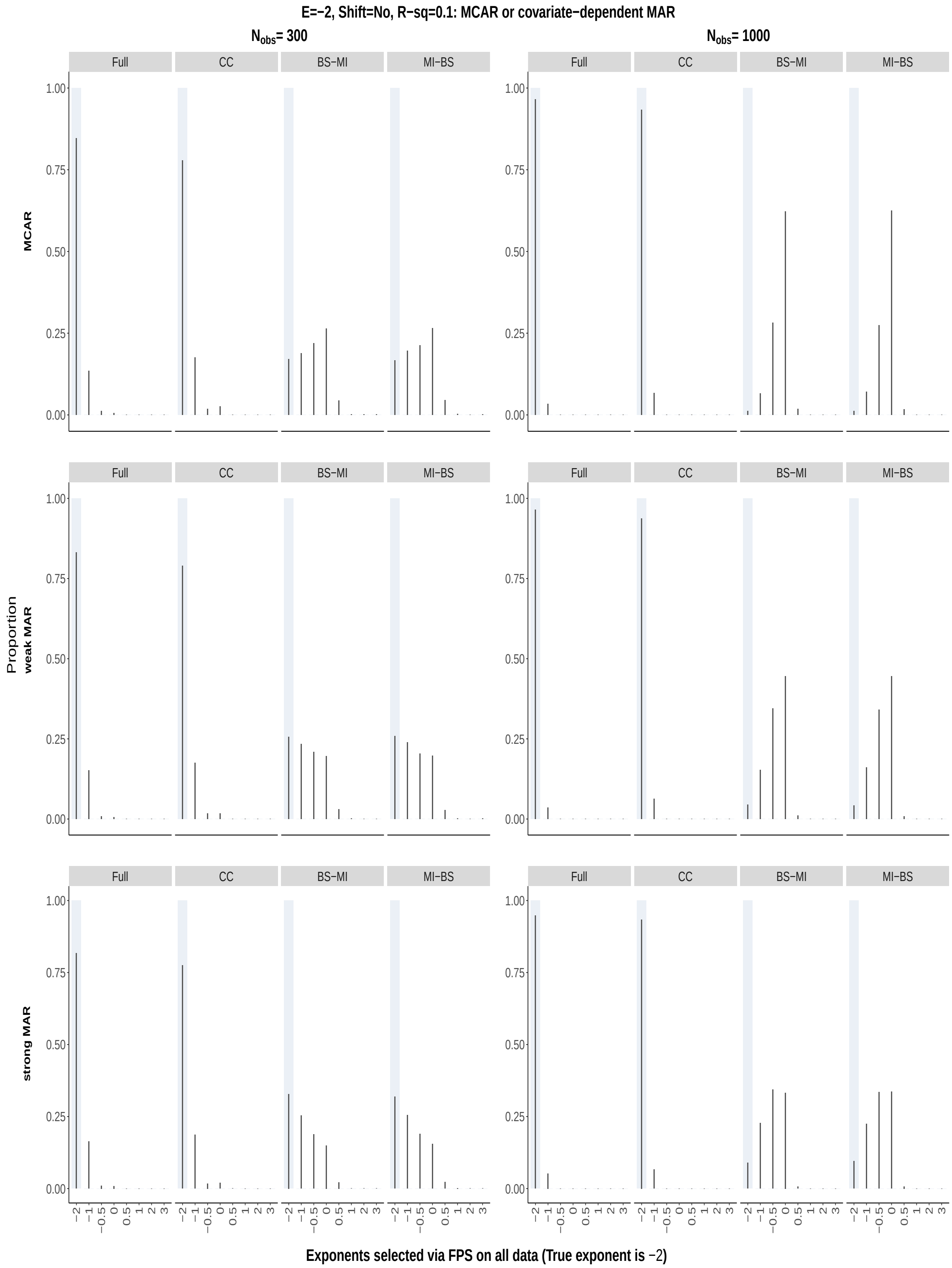

Figure S297: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

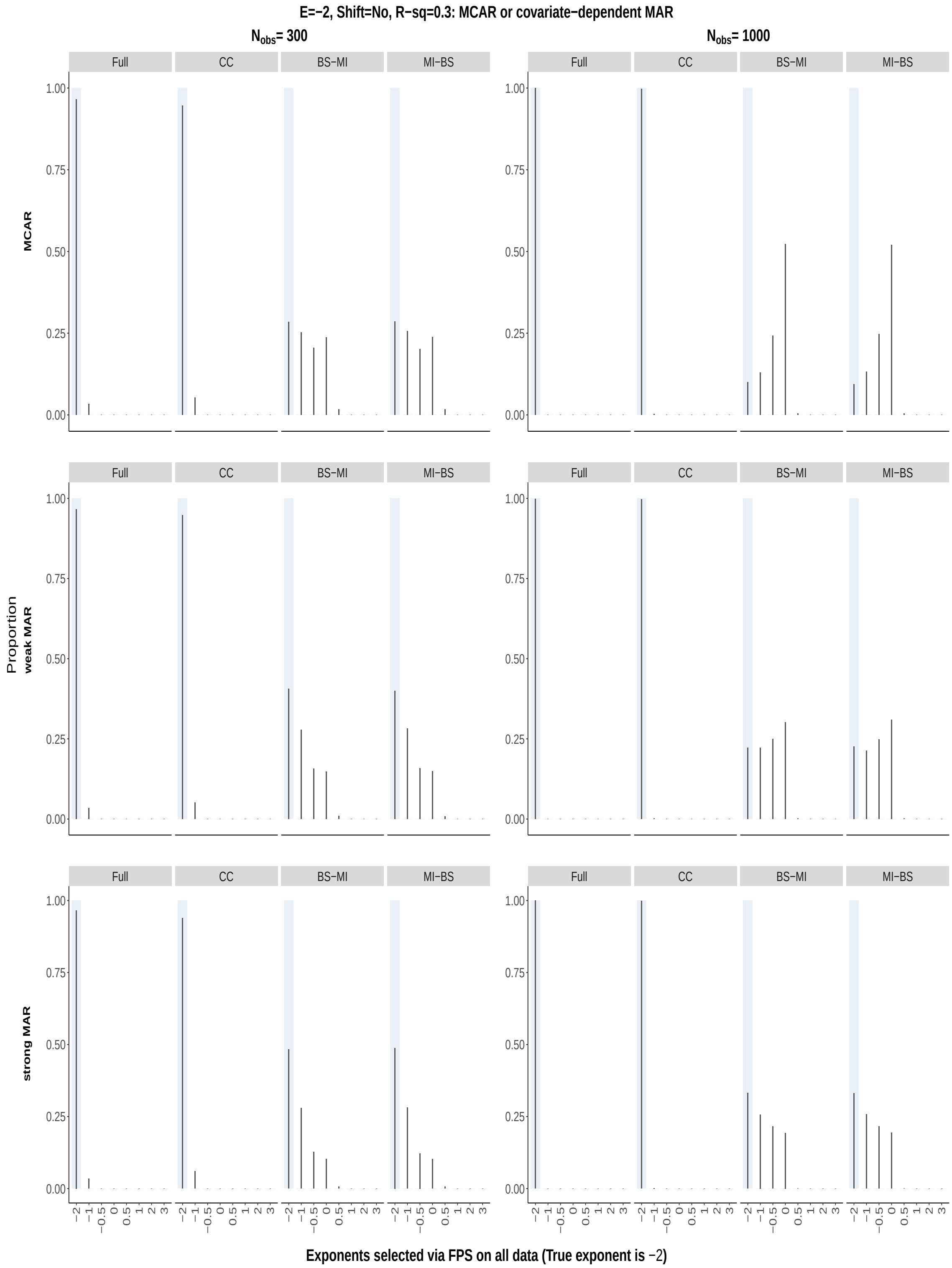

Figure S298: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

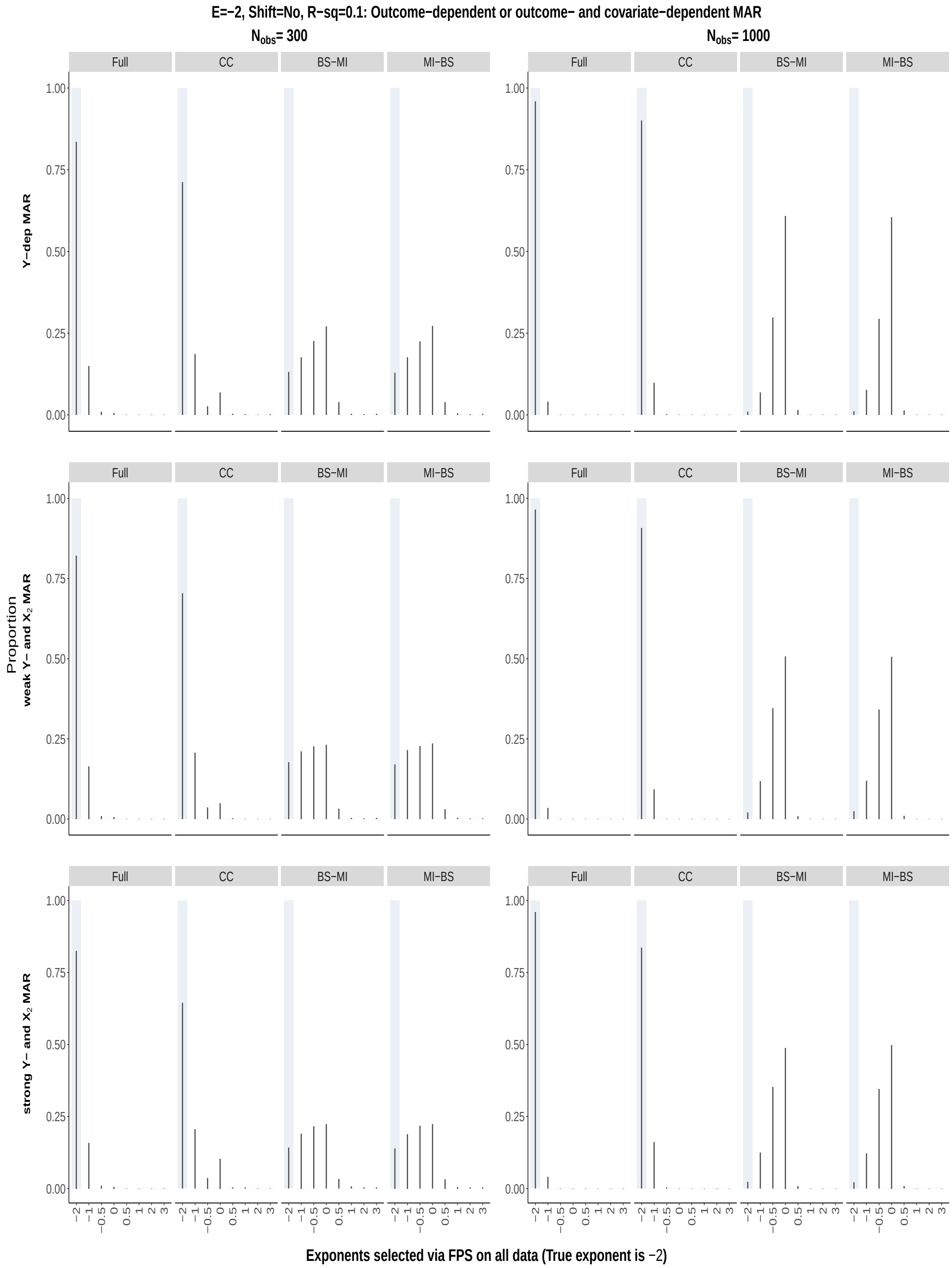

Figure S299: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

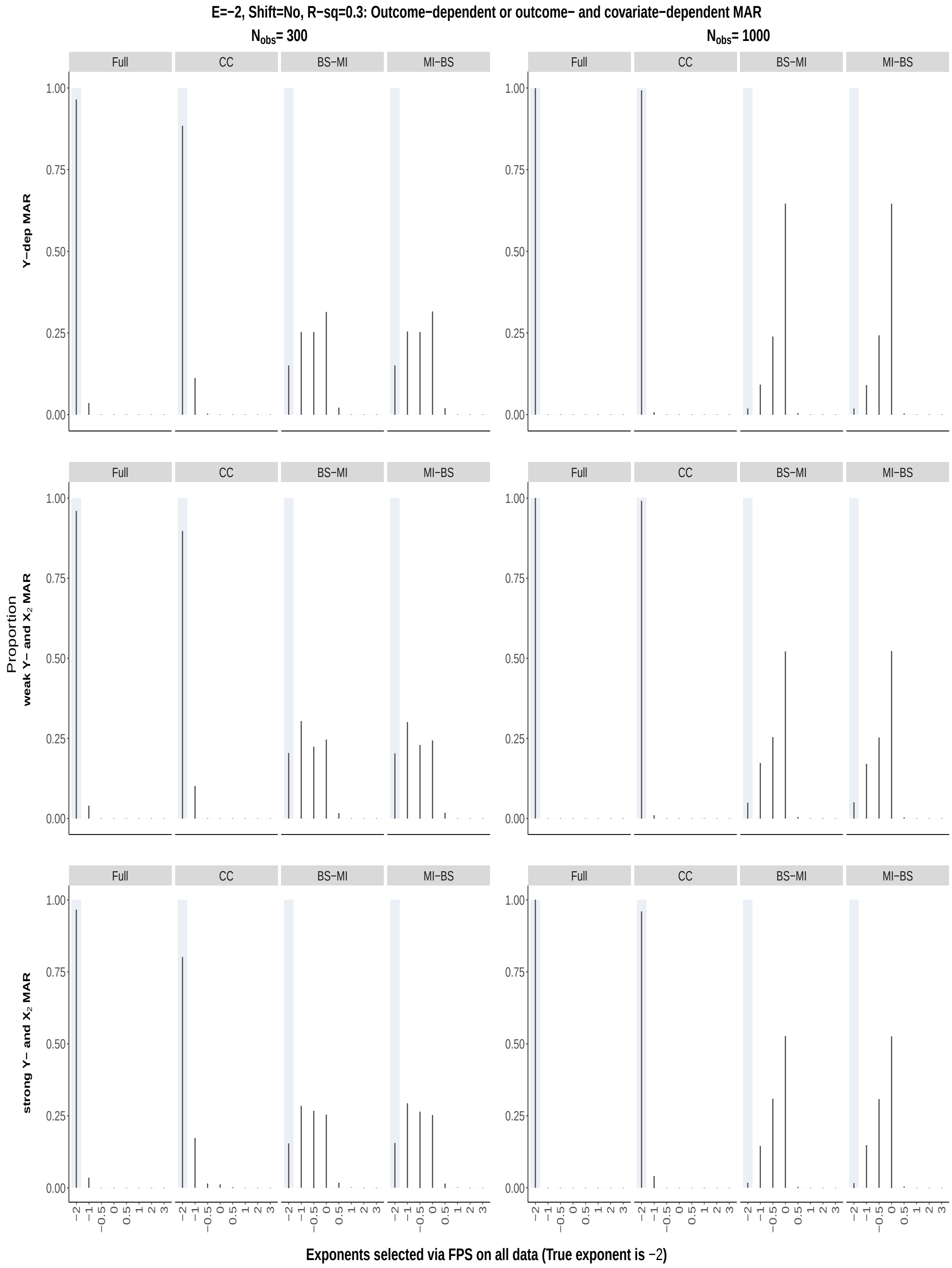

Figure S300: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.22 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 0$, $\alpha_E = 0.05$ and no origin-shift

True exponent is 0

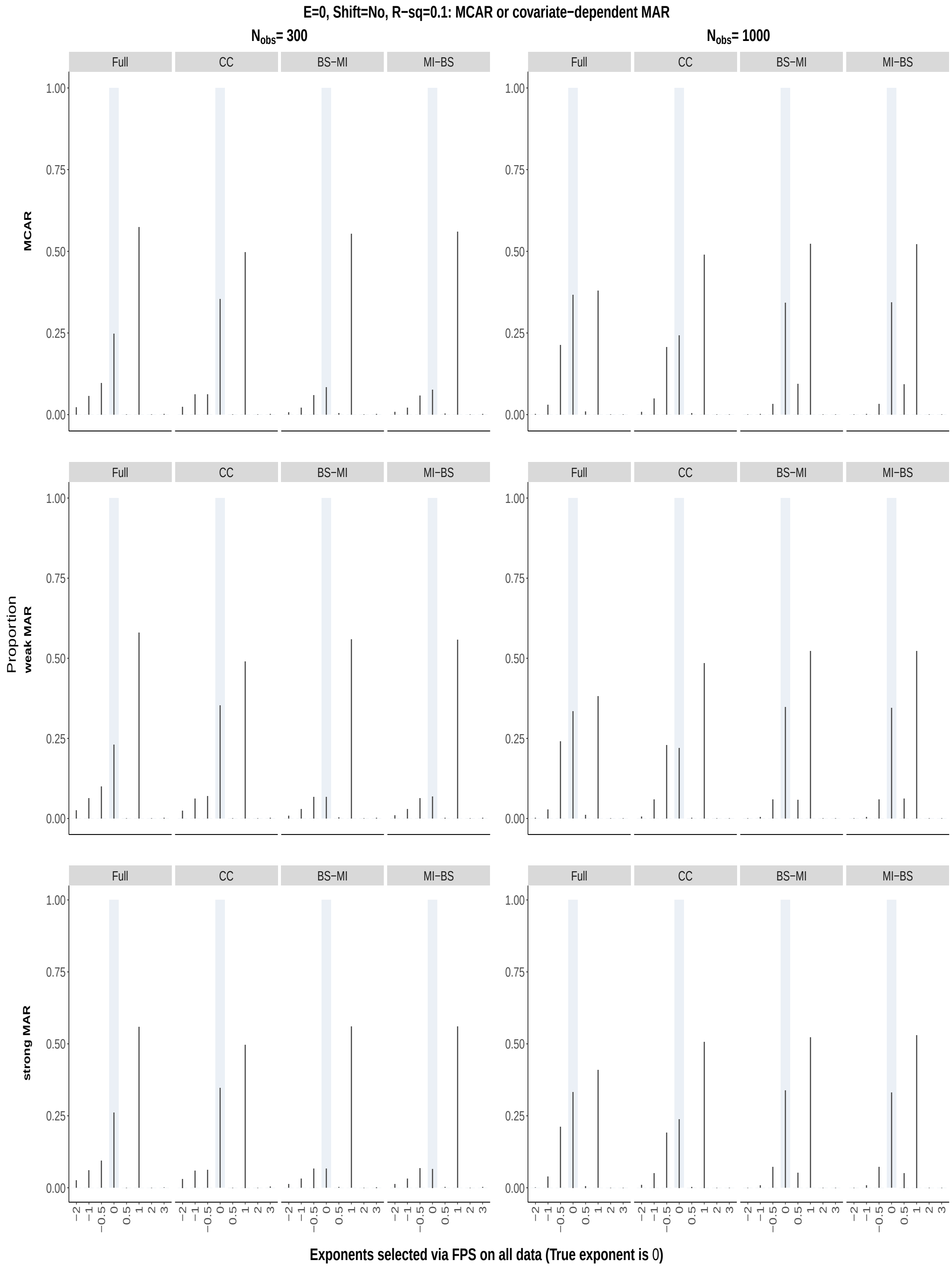

Figure S301: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

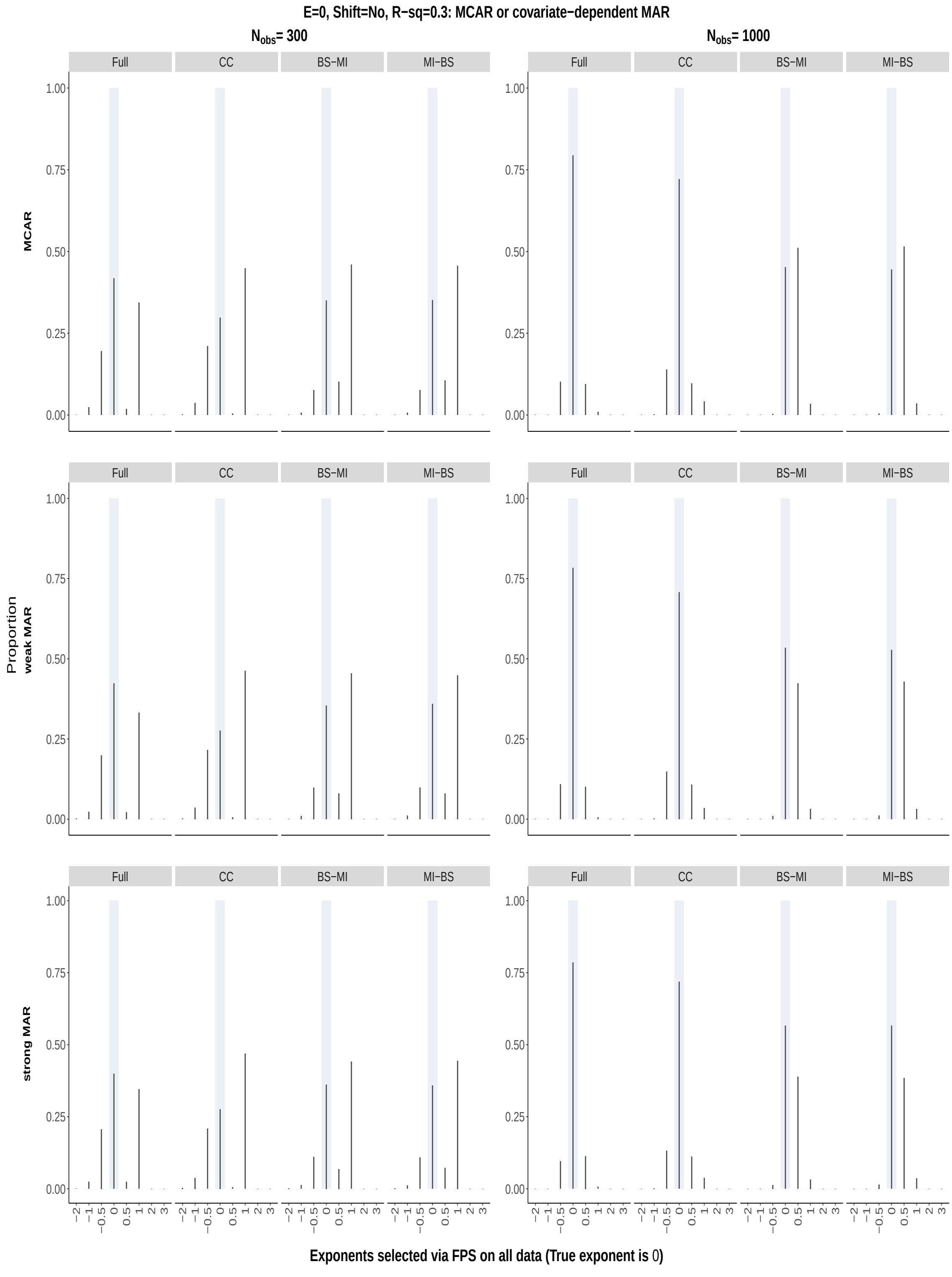

Figure S302: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

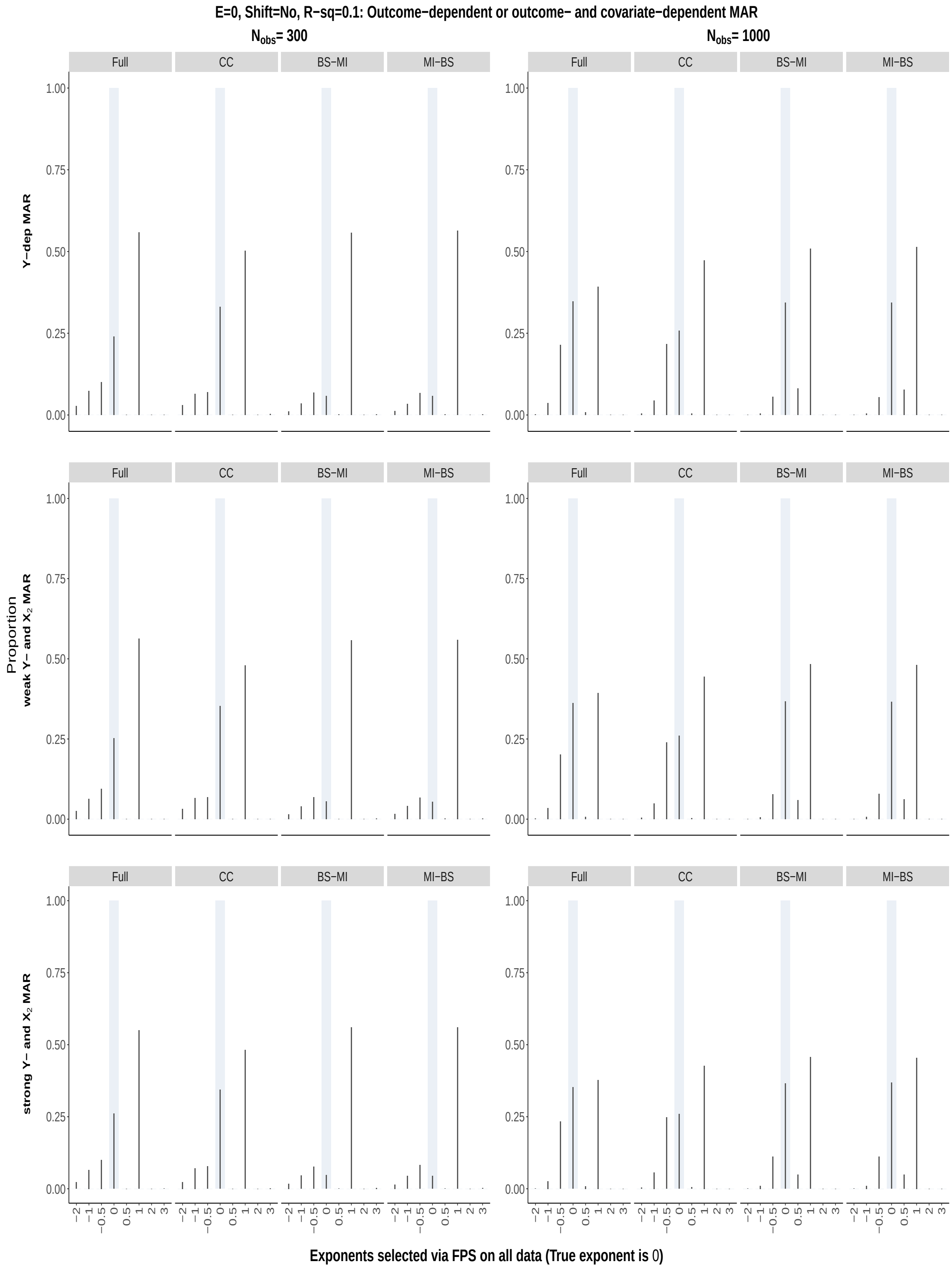

Figure S303: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

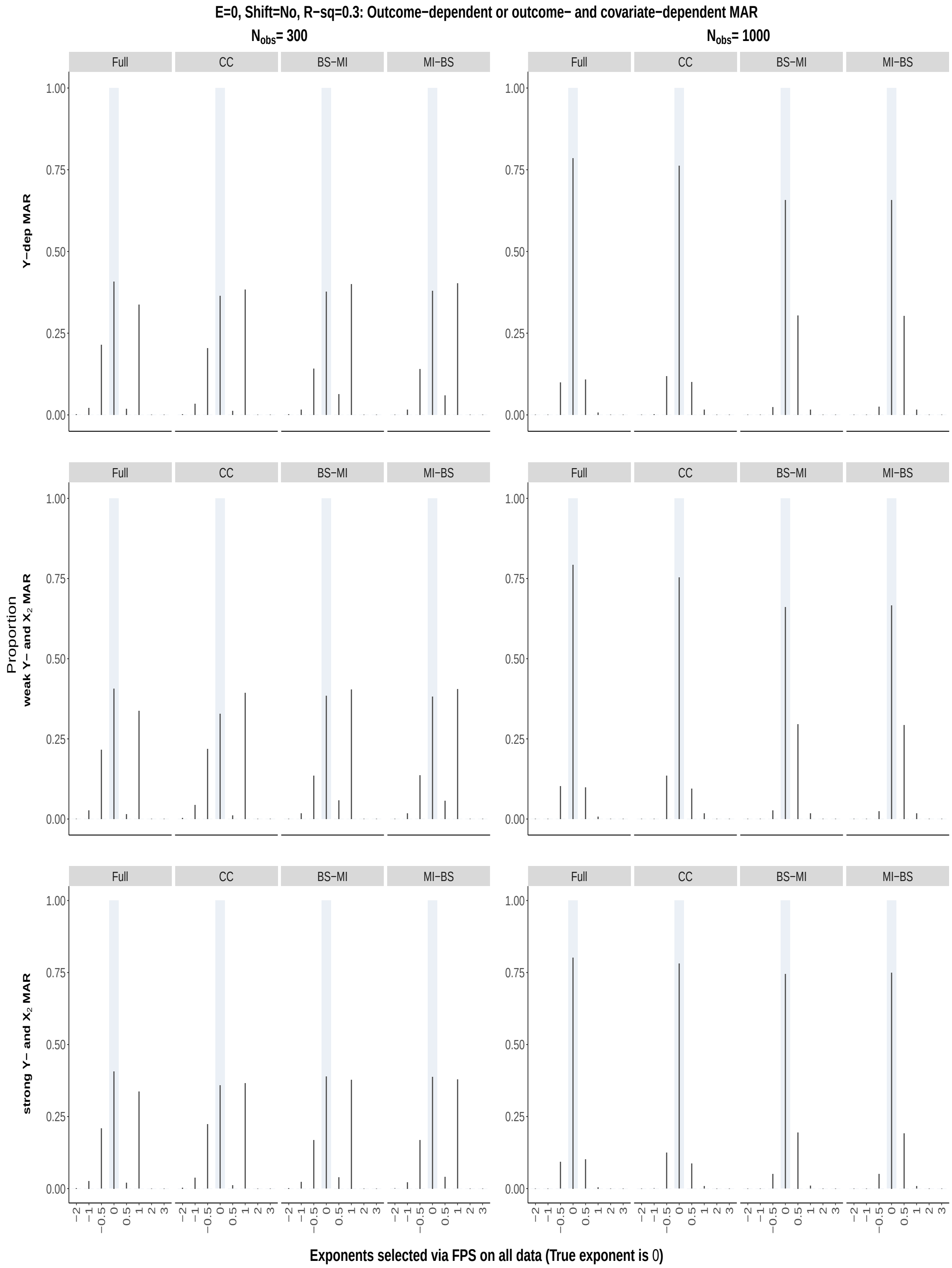

Figure S304: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

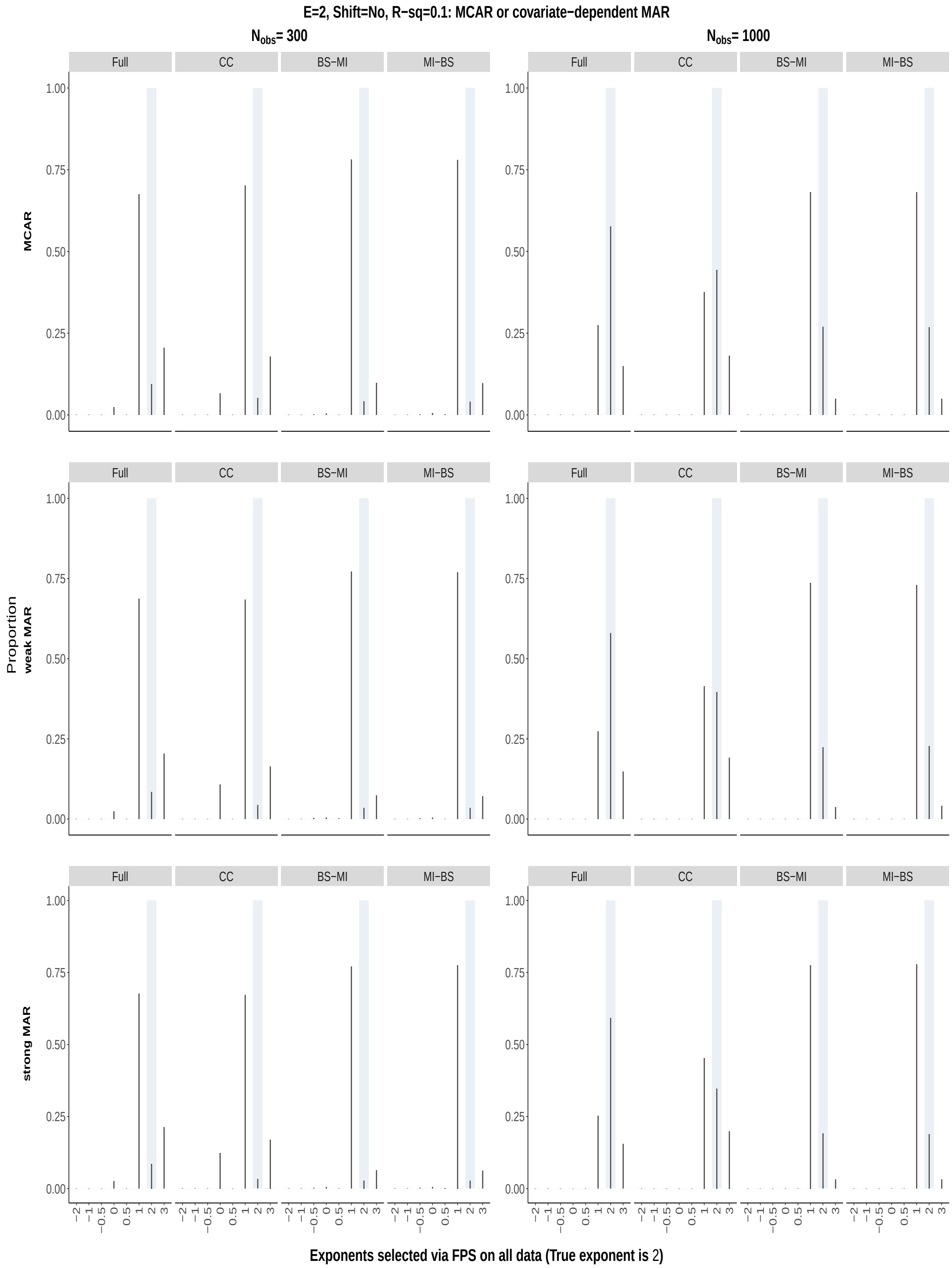

Figure S305: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

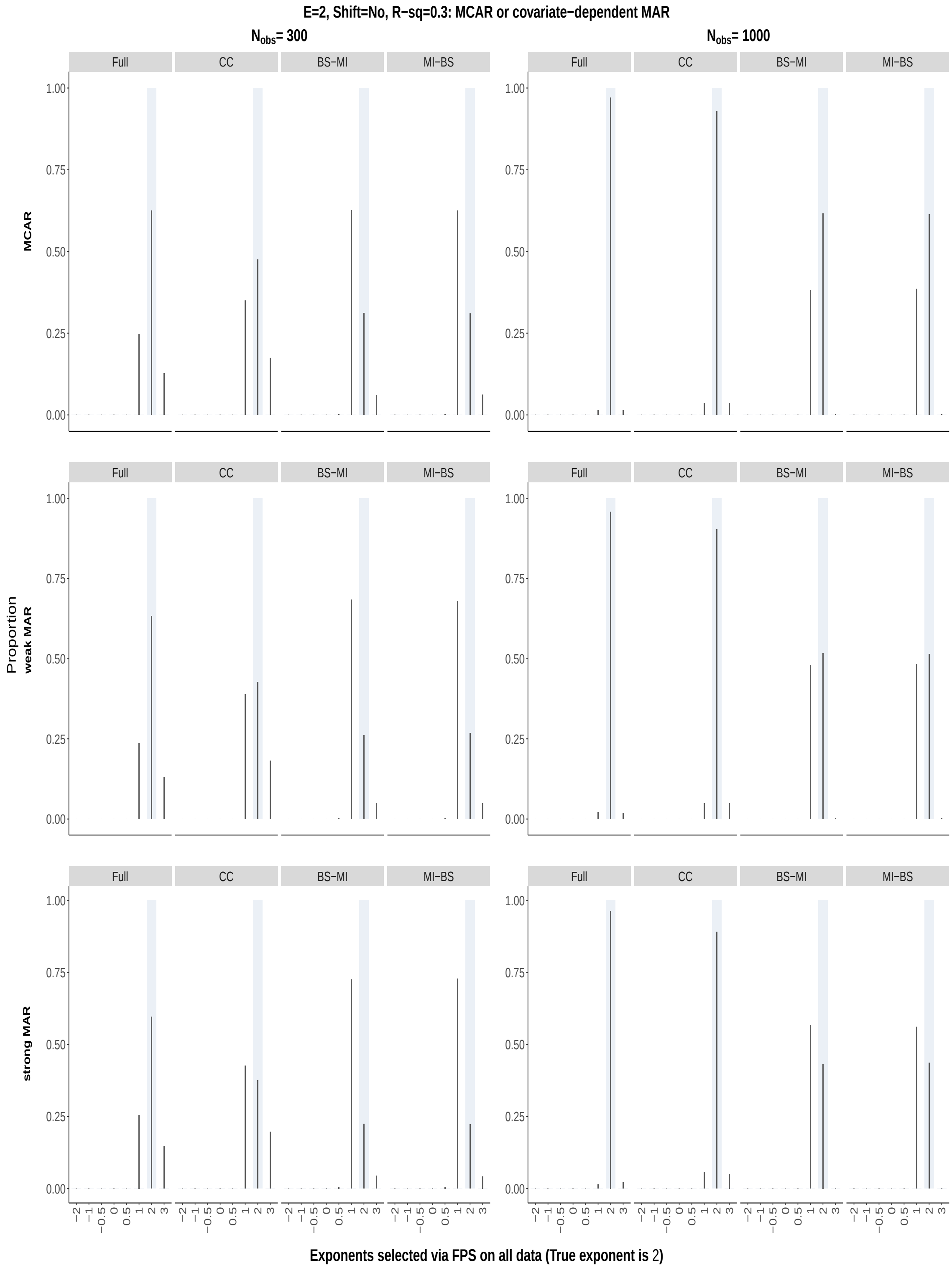

Figure S306: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

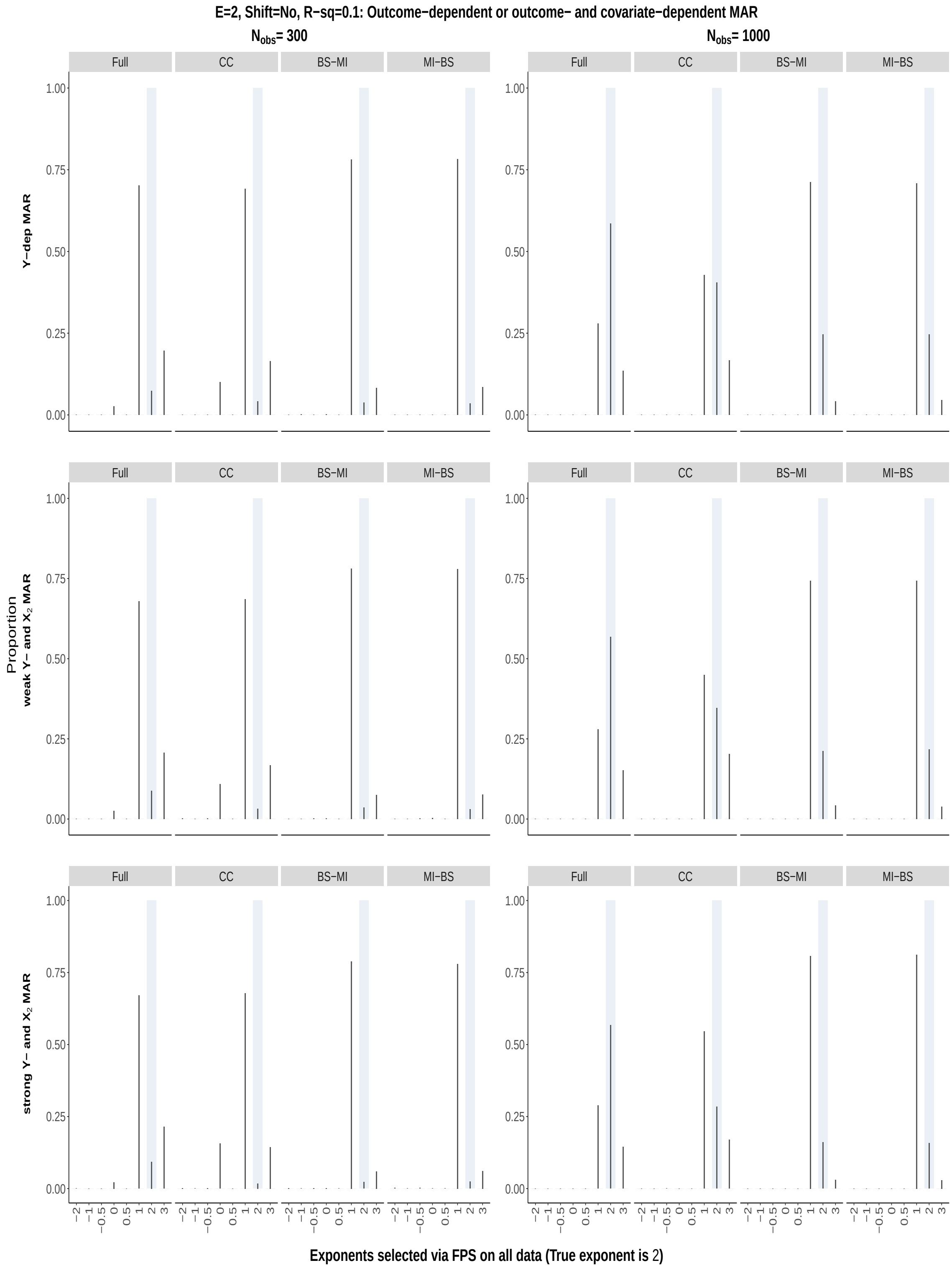

Figure S307: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

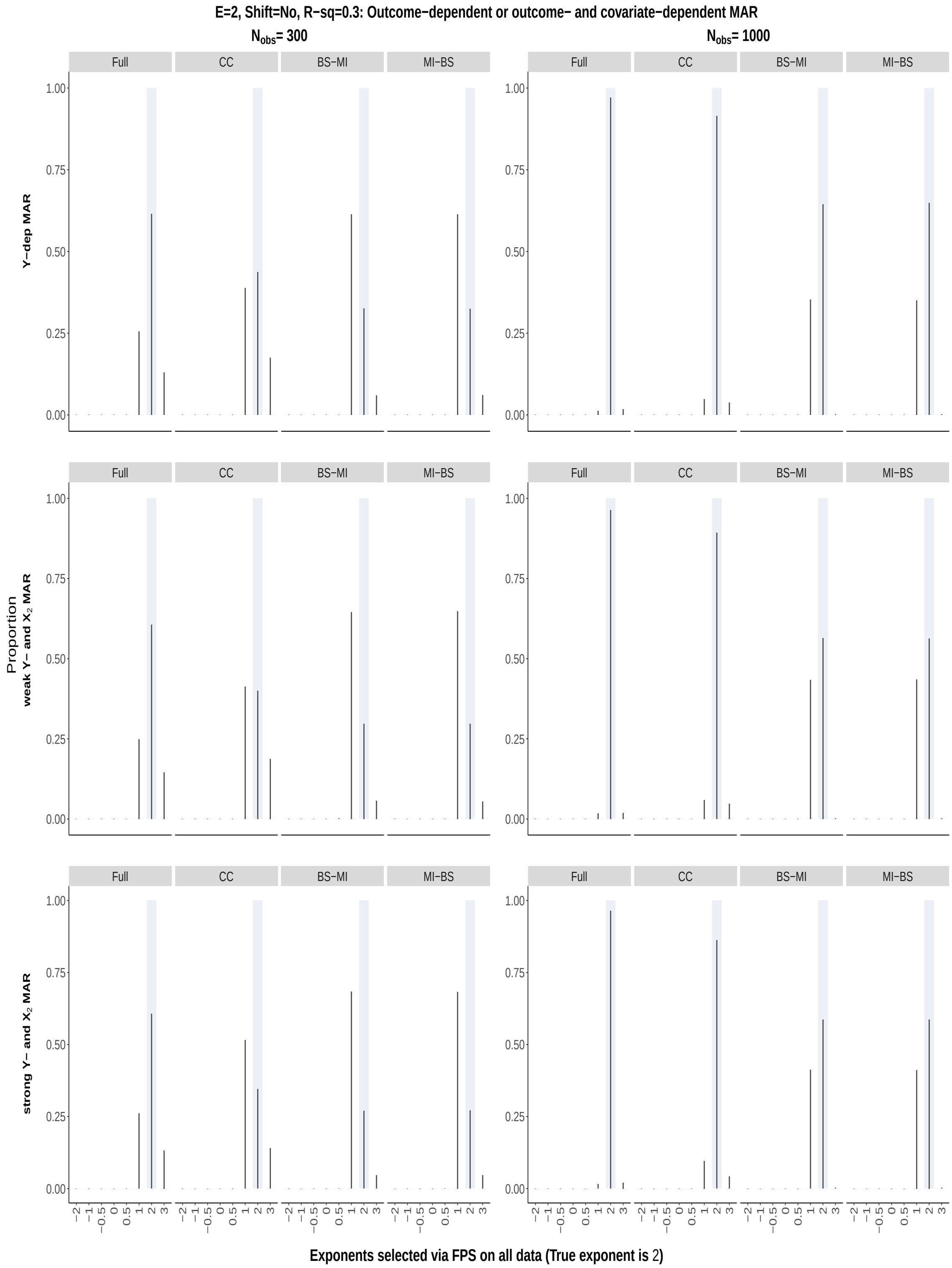

Figure S308: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

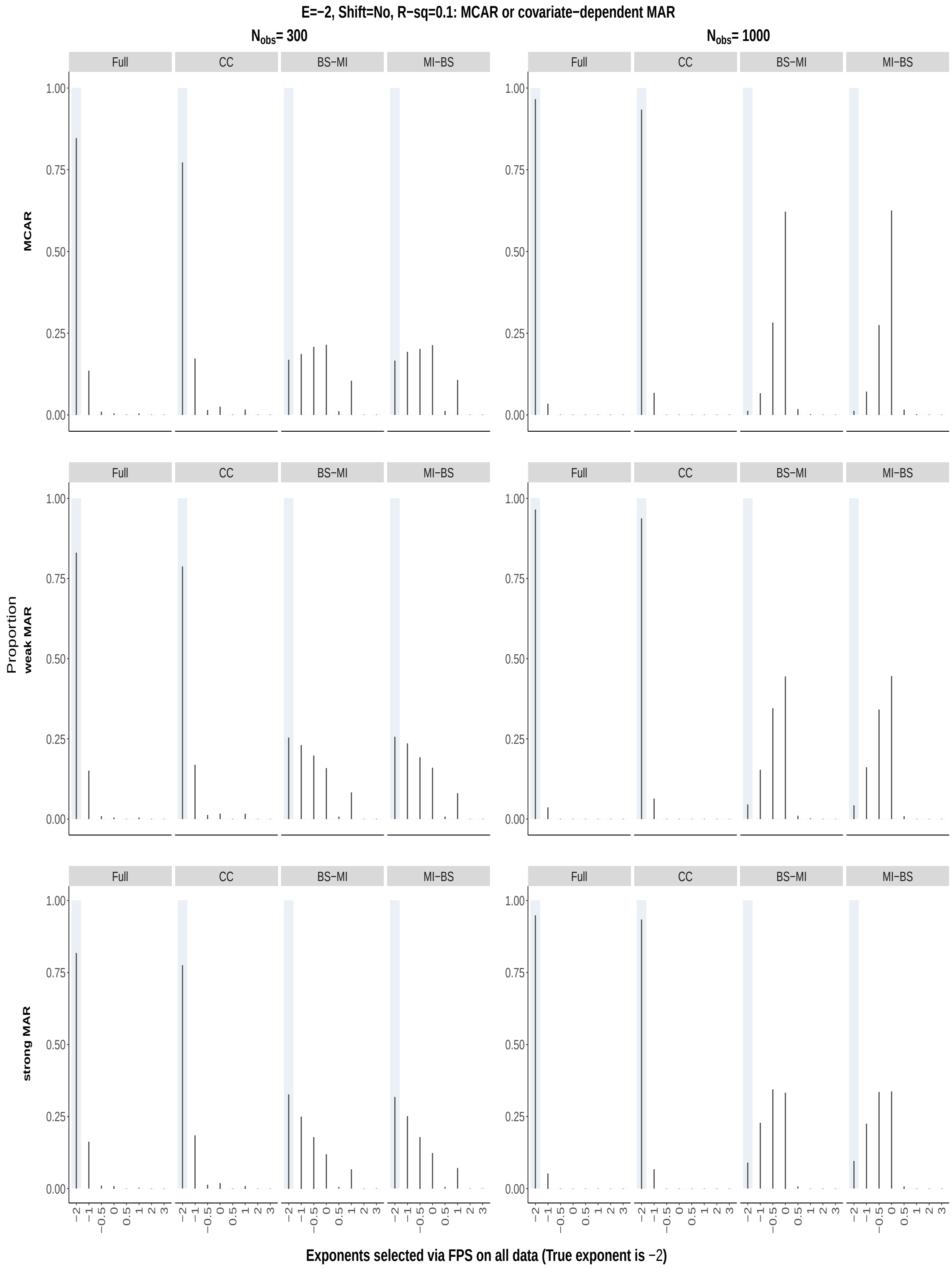

Figure S309: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

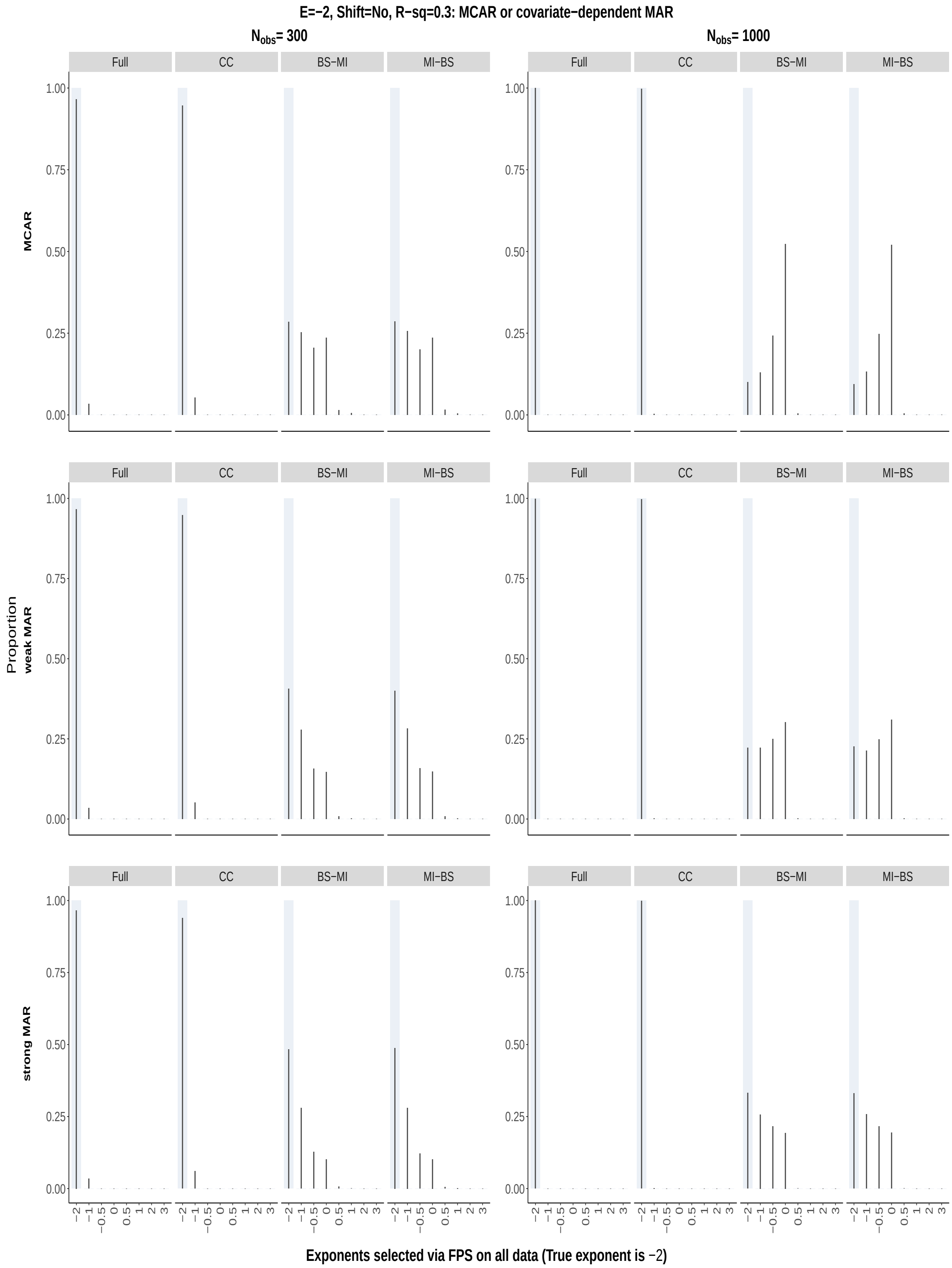

Figure S310: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

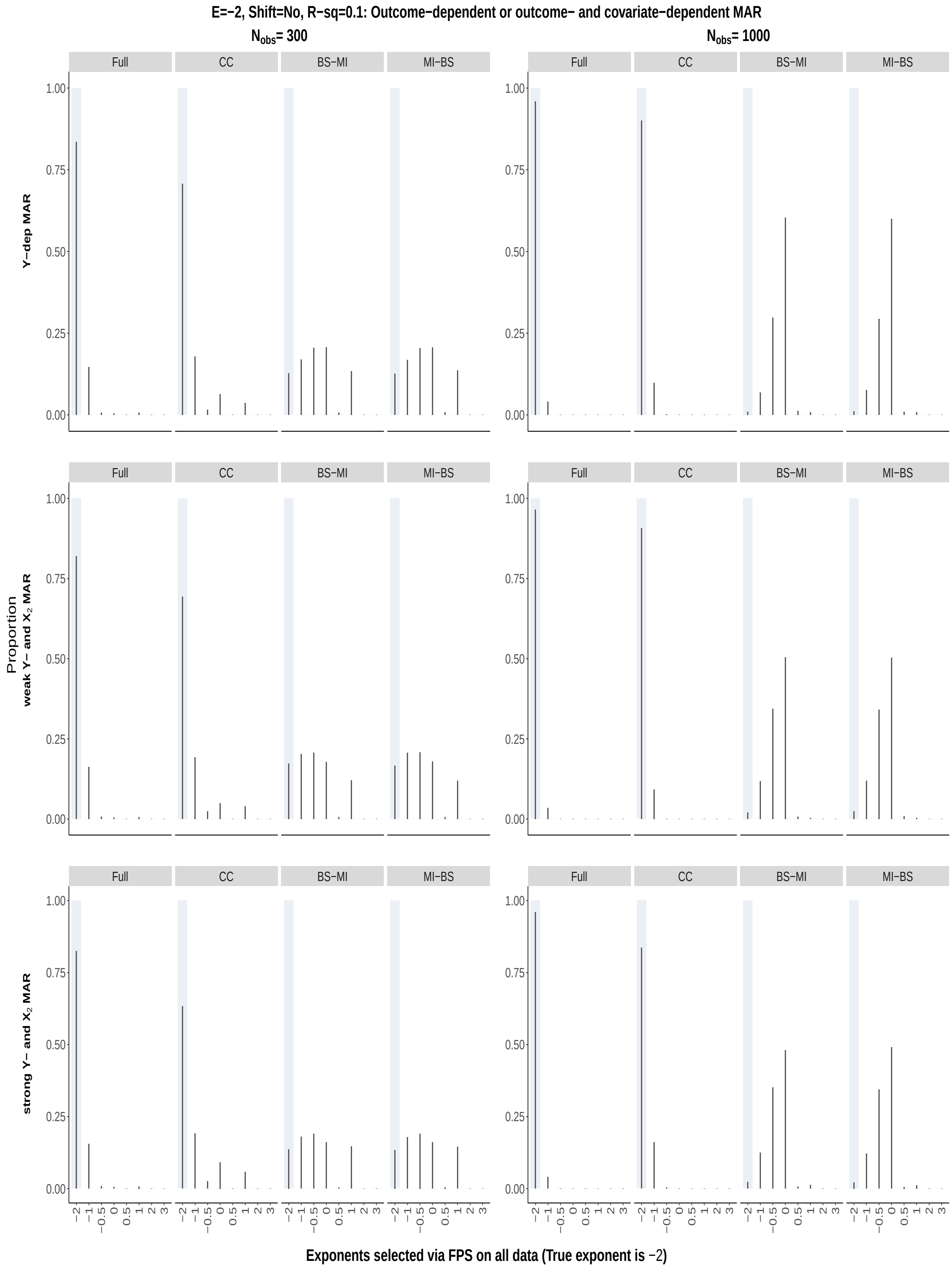

Figure S311: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

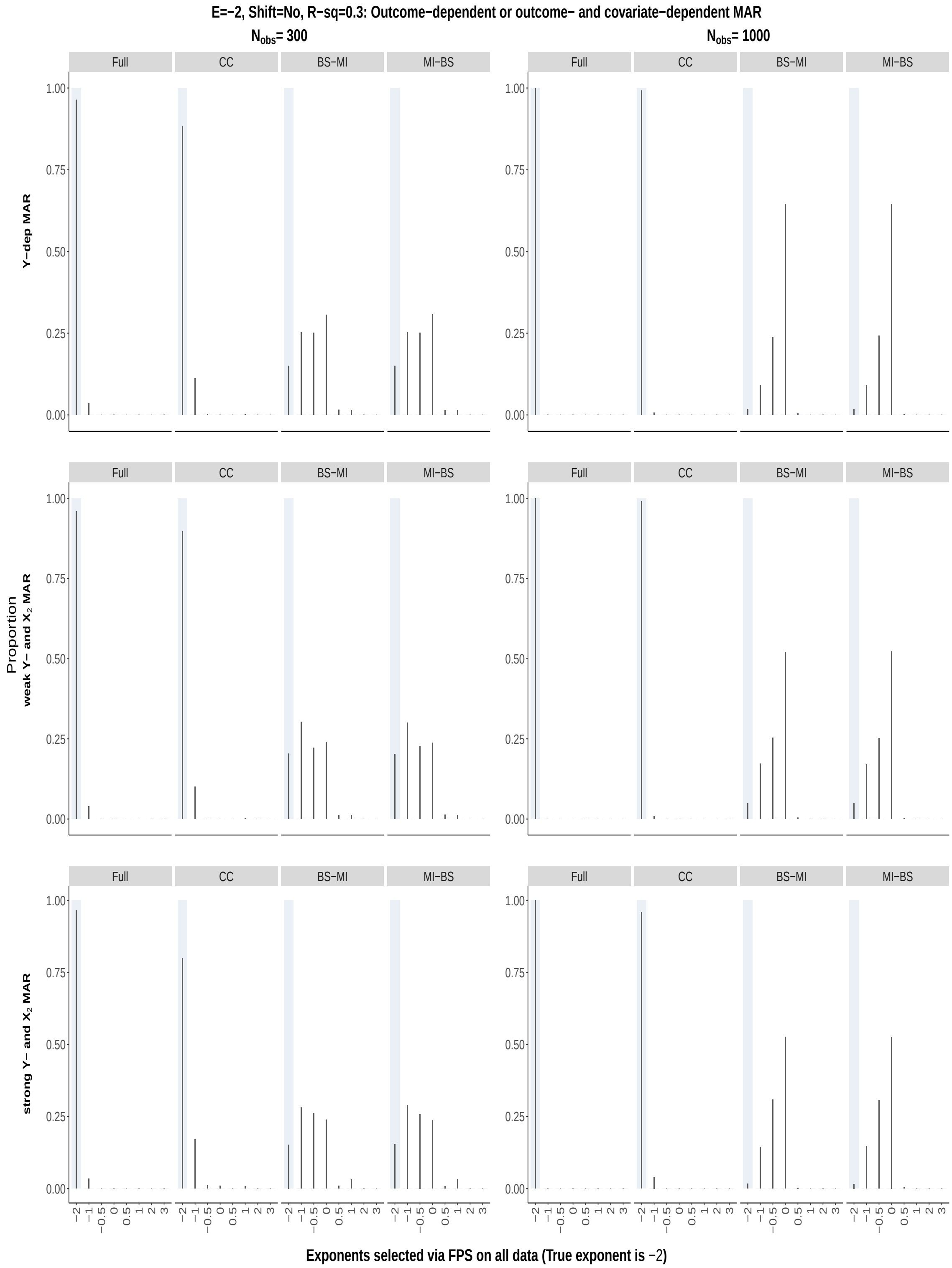

Figure S312: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.23 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 0$, $\alpha_E = 1$ and and an origin-shift has been applied

True exponent is 0

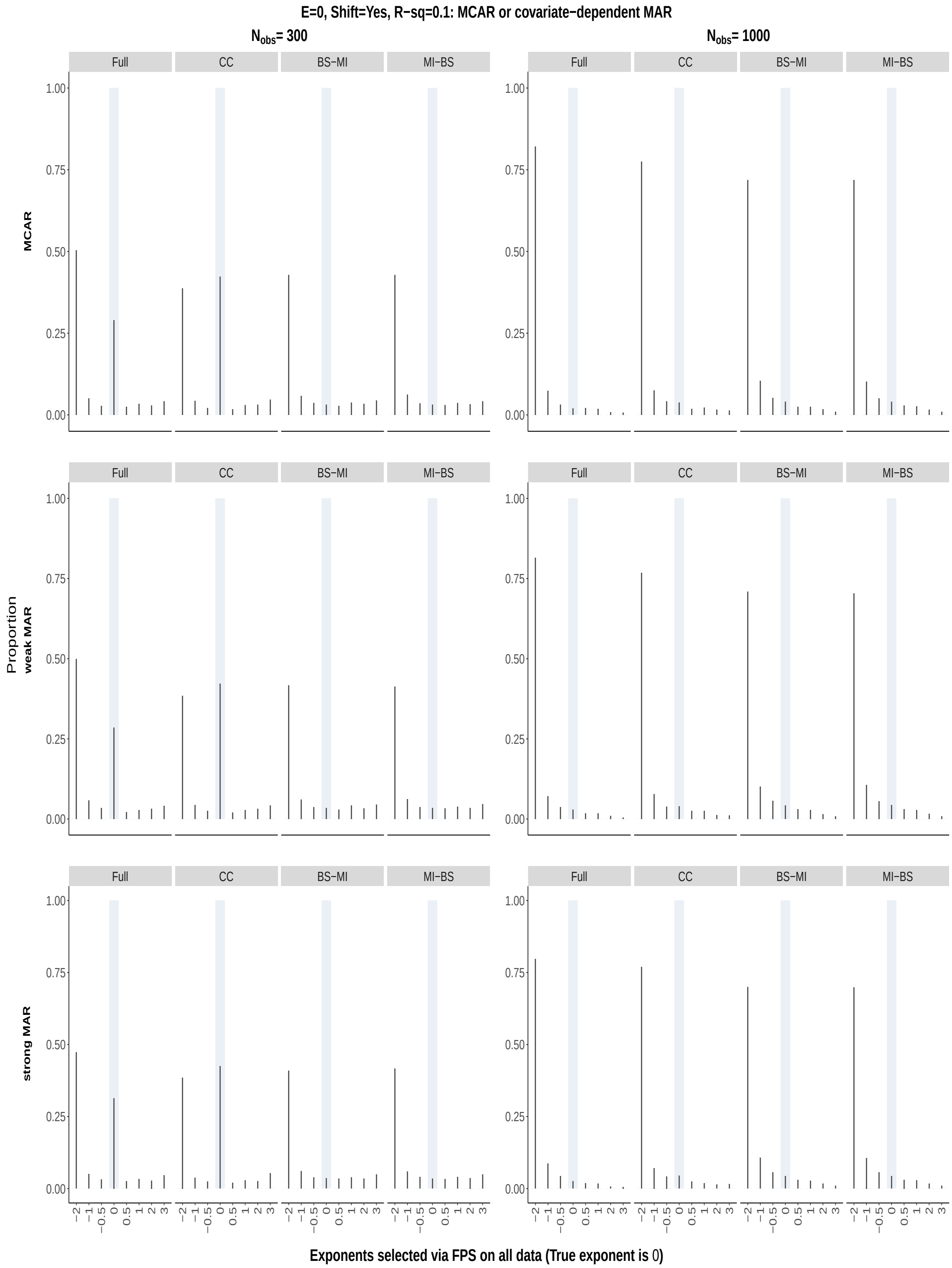

Figure S313: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

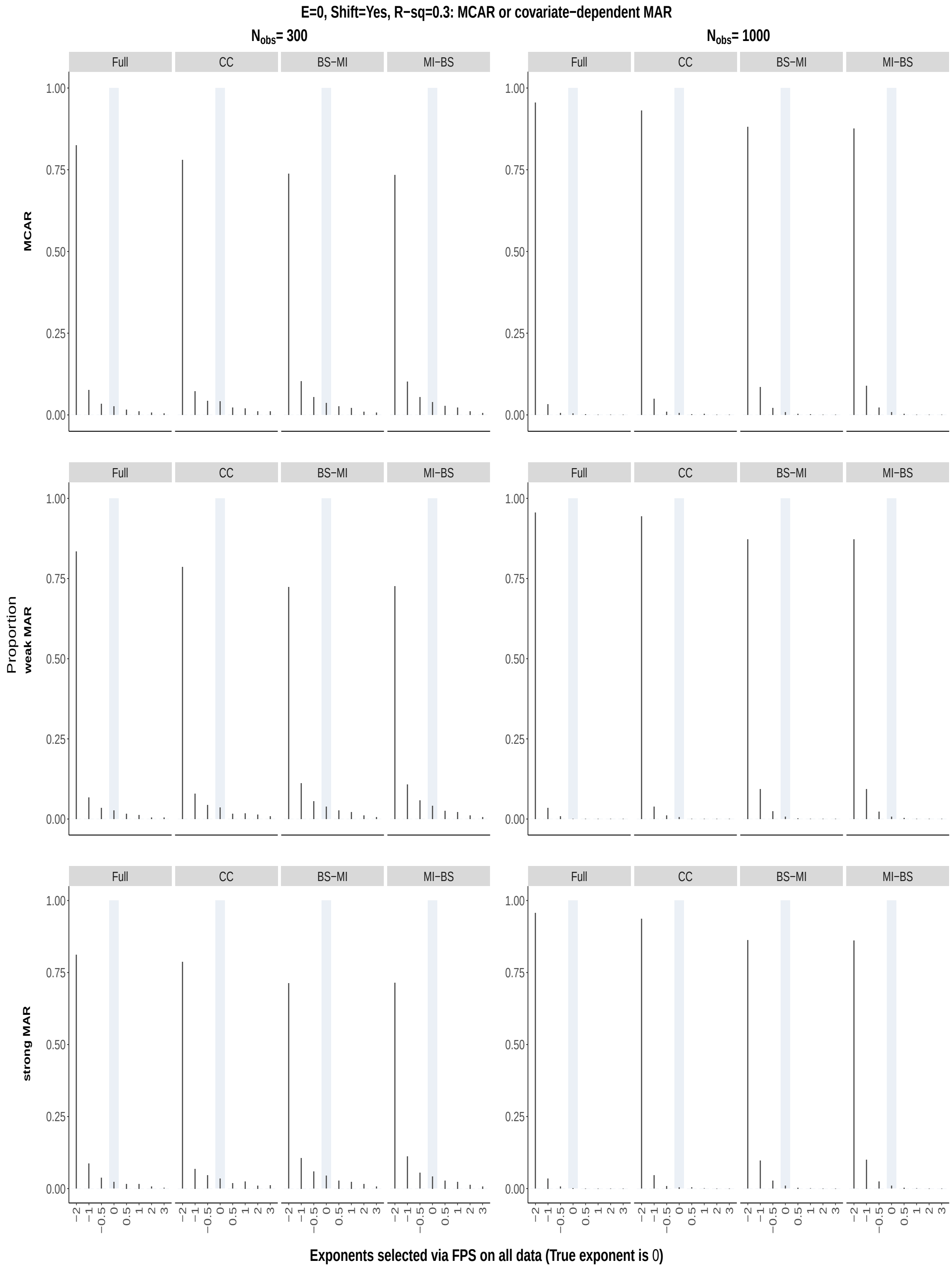

Figure S314: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

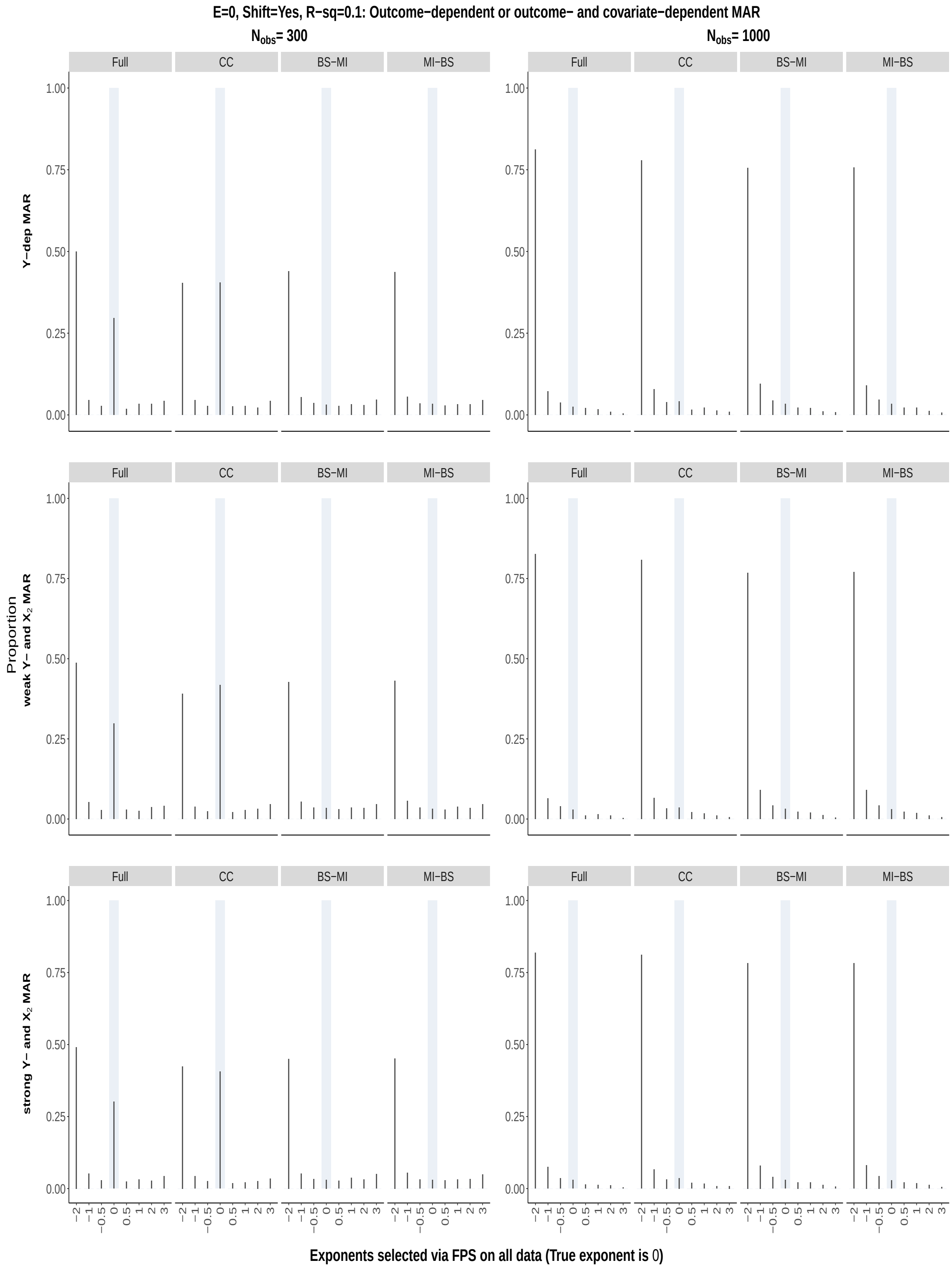

Figure S315: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

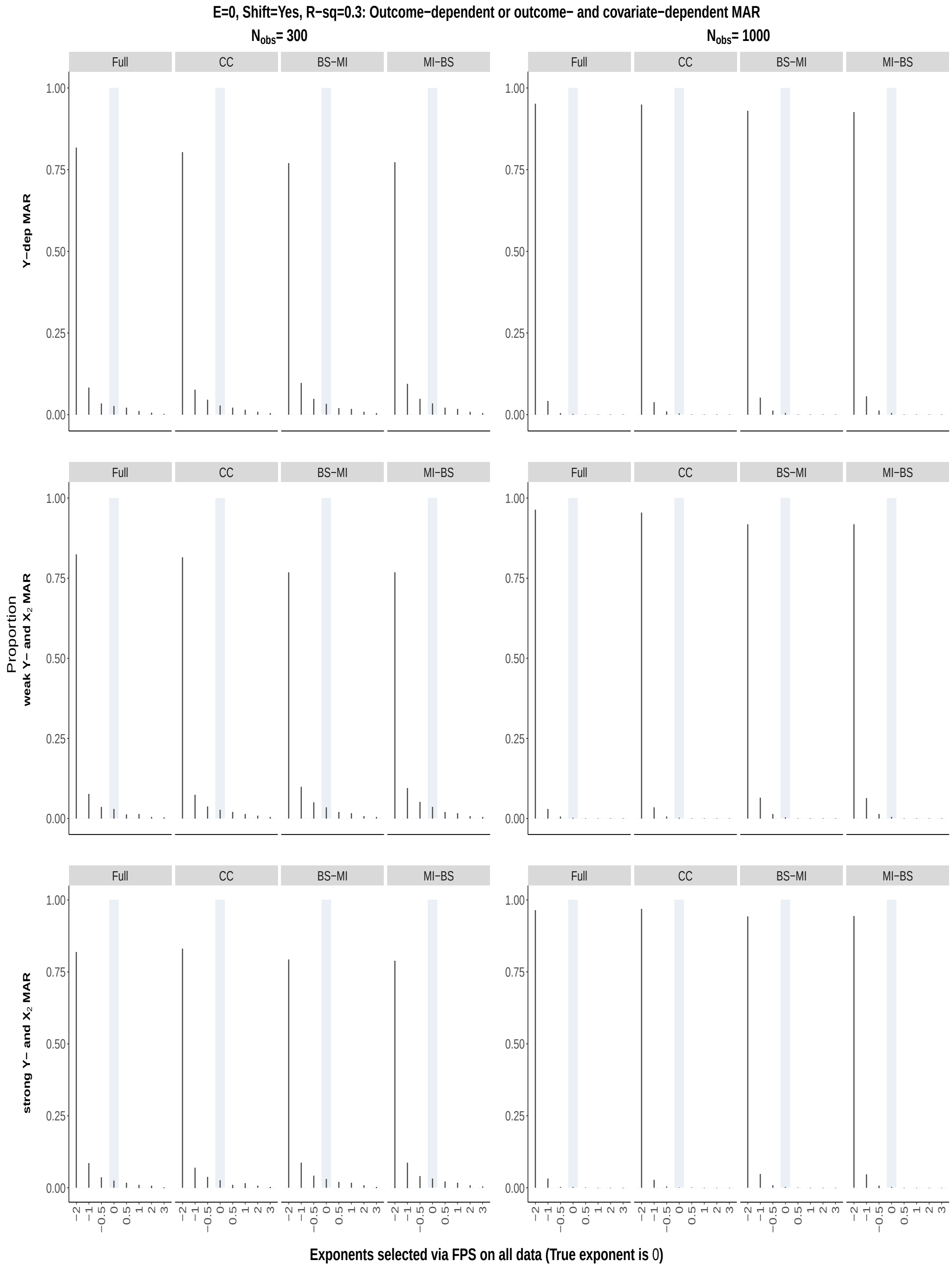

Figure S316: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

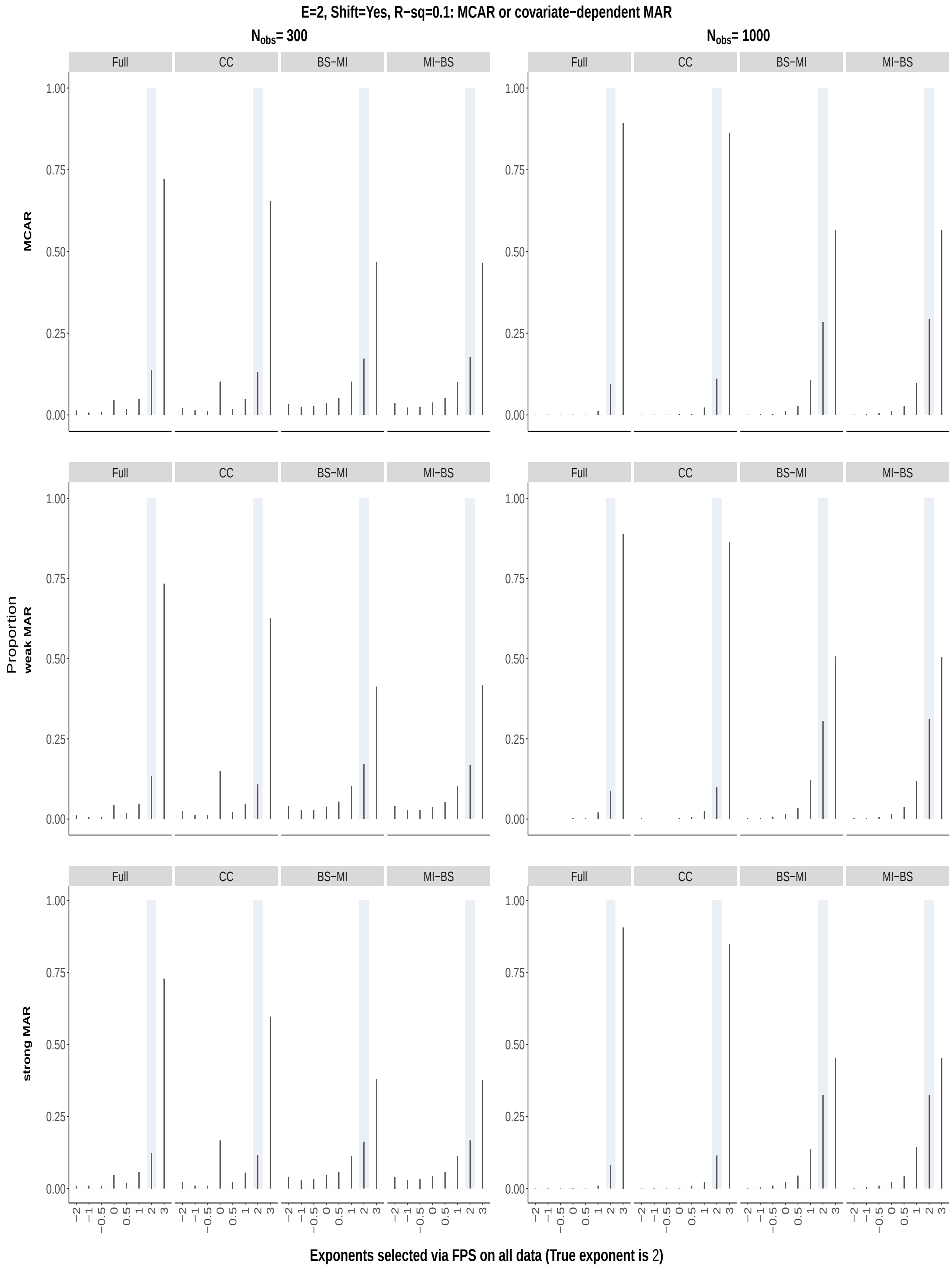

Figure S317: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

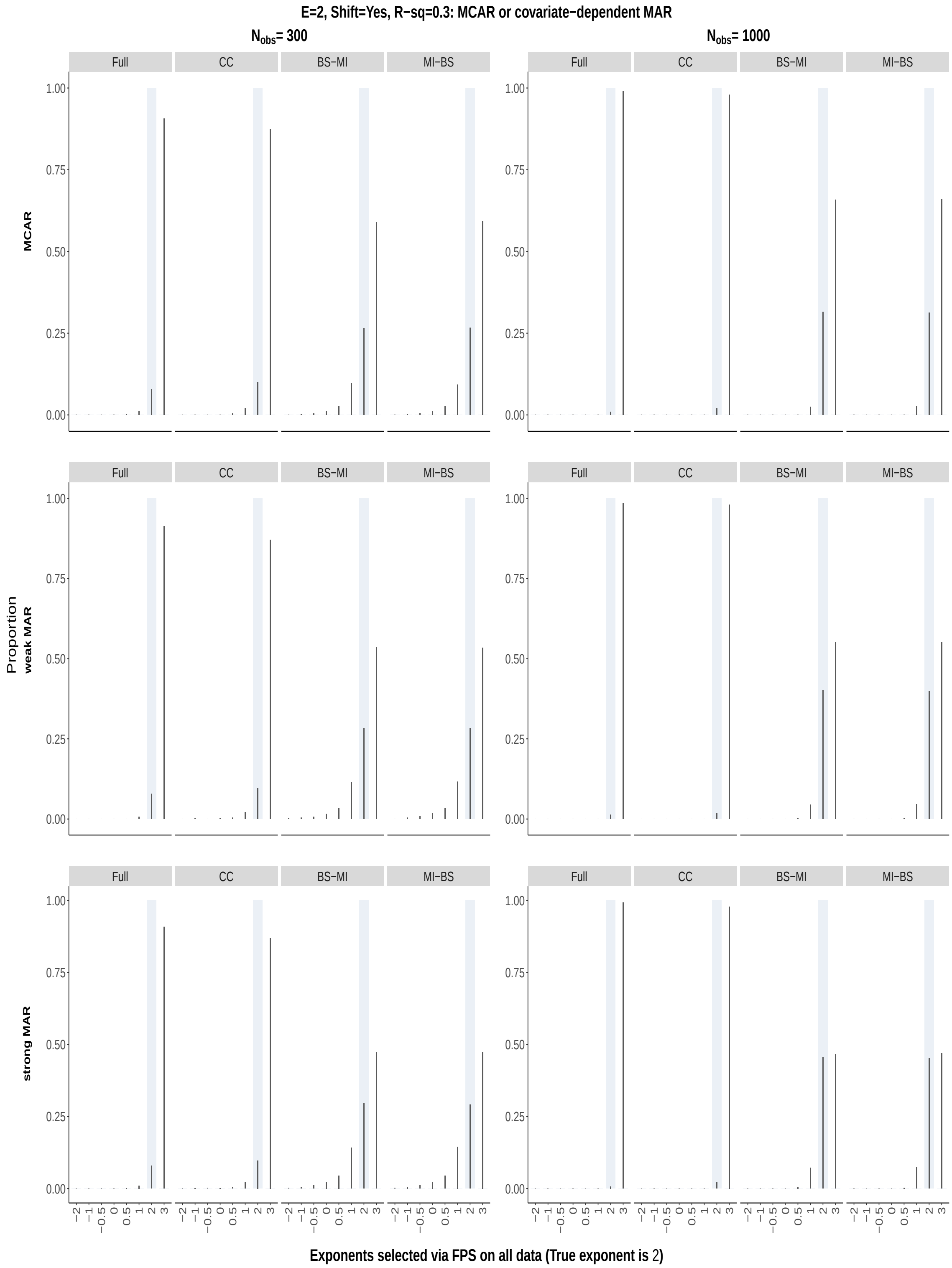

Figure S318: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

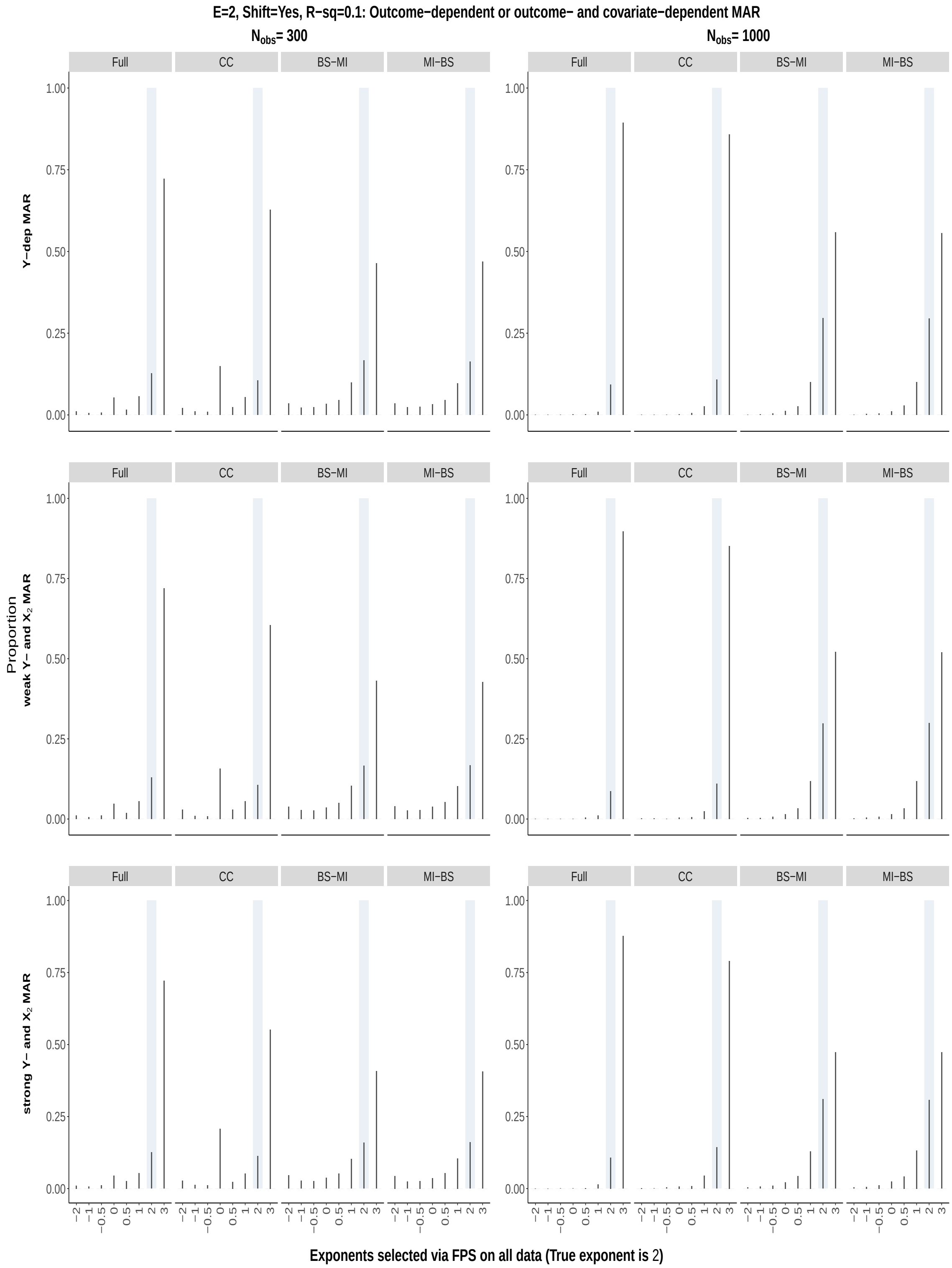

Figure S319: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

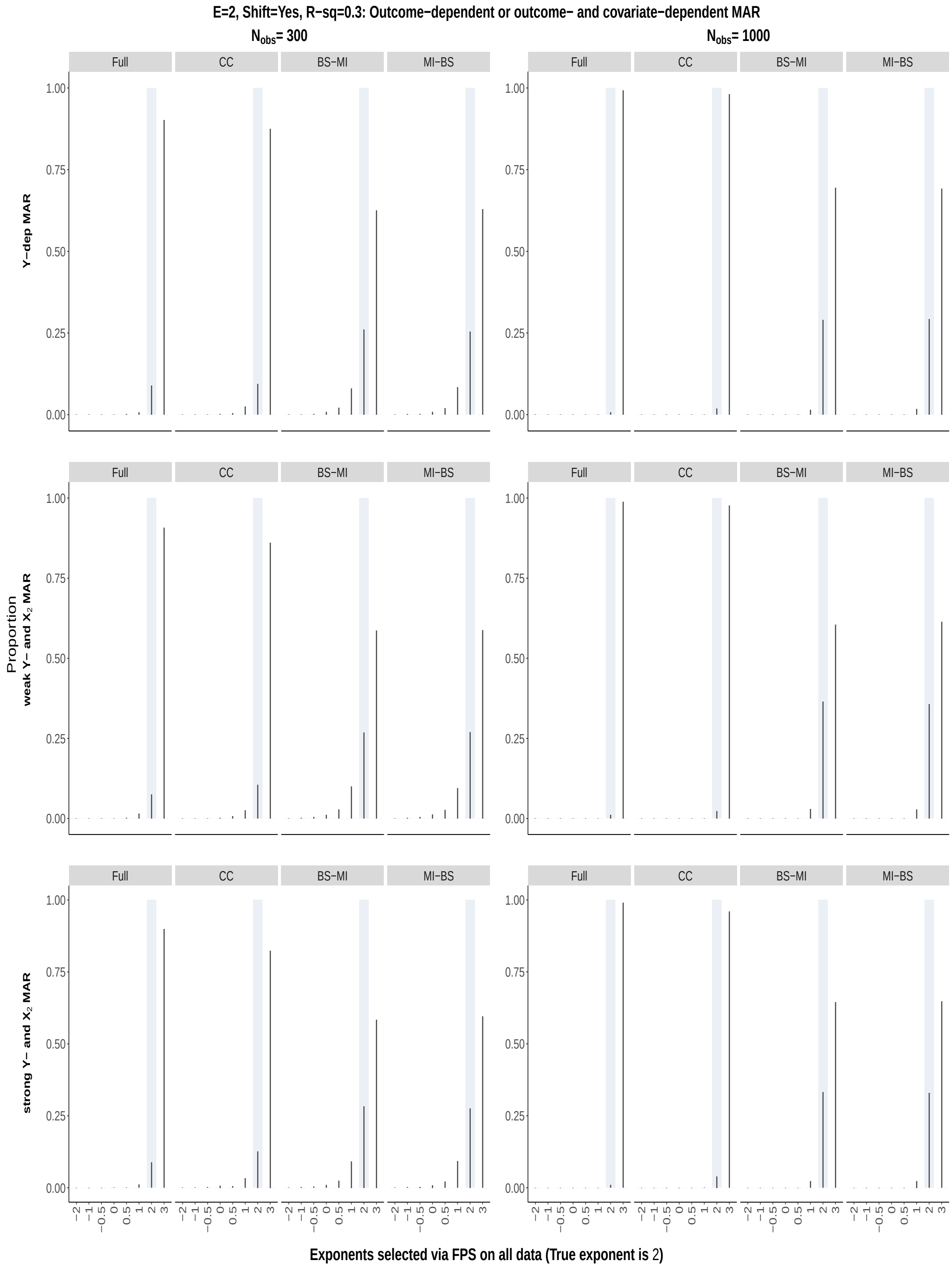

Figure S320: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

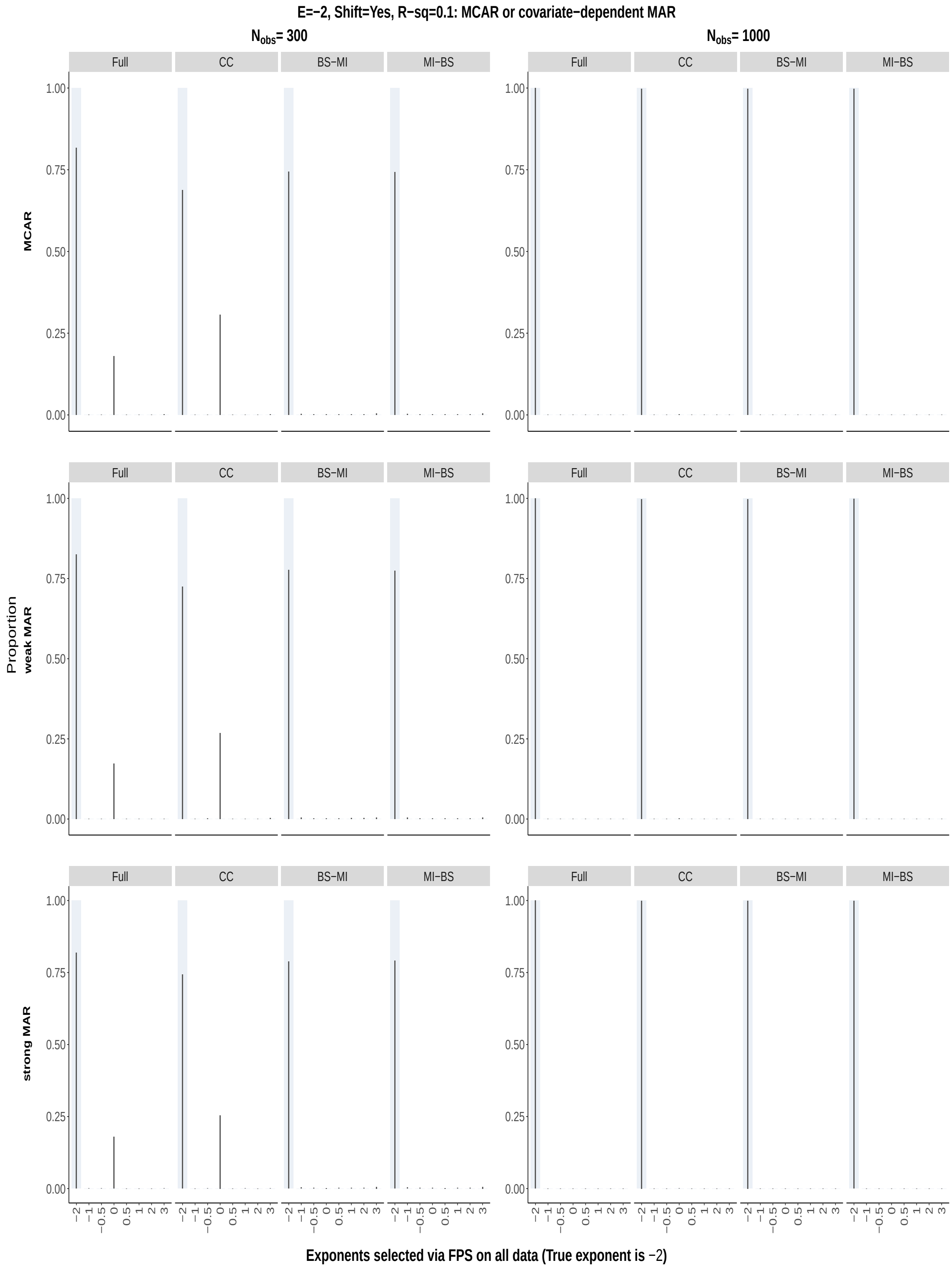

Figure S321: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

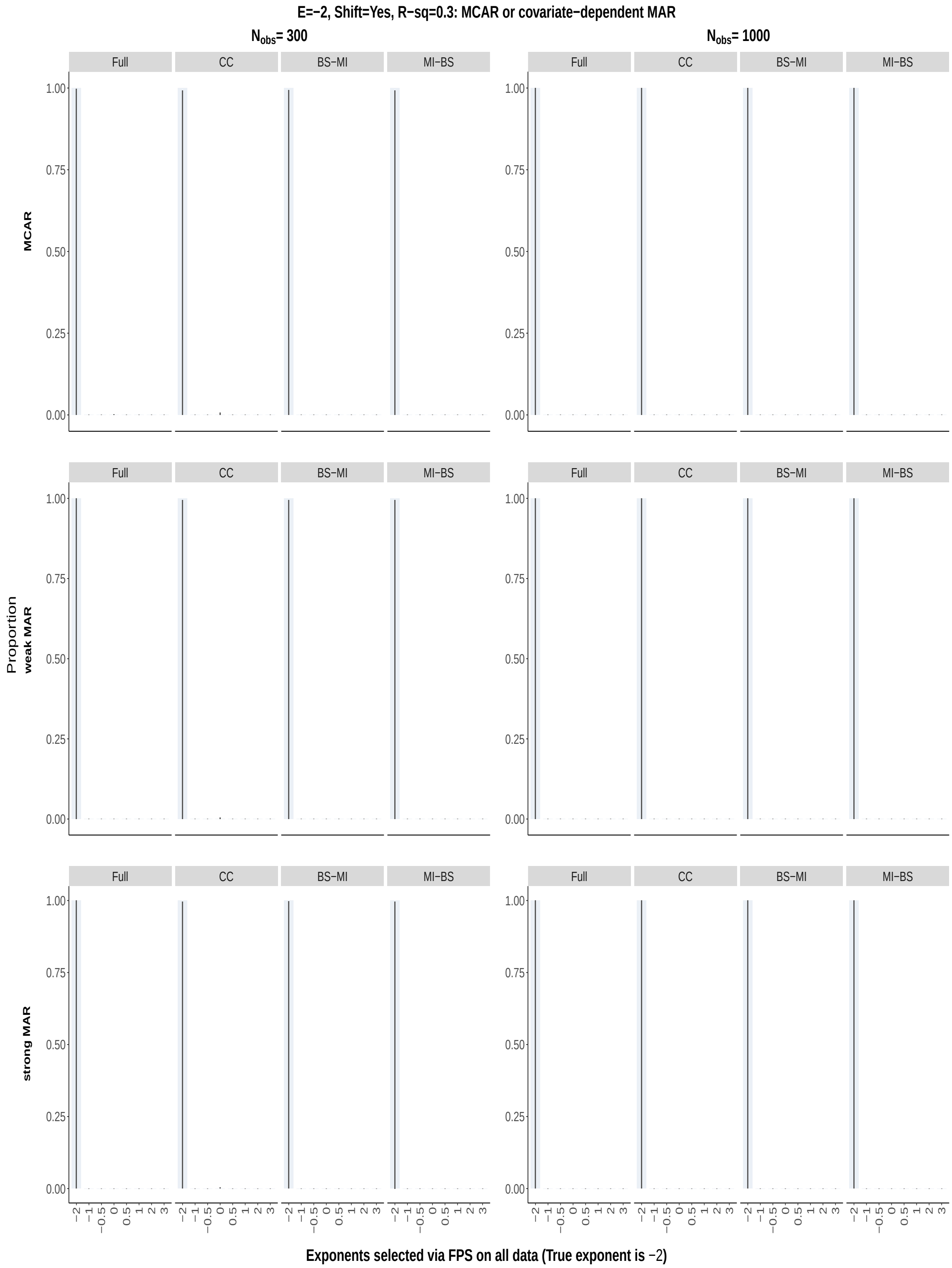

Figure S322: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

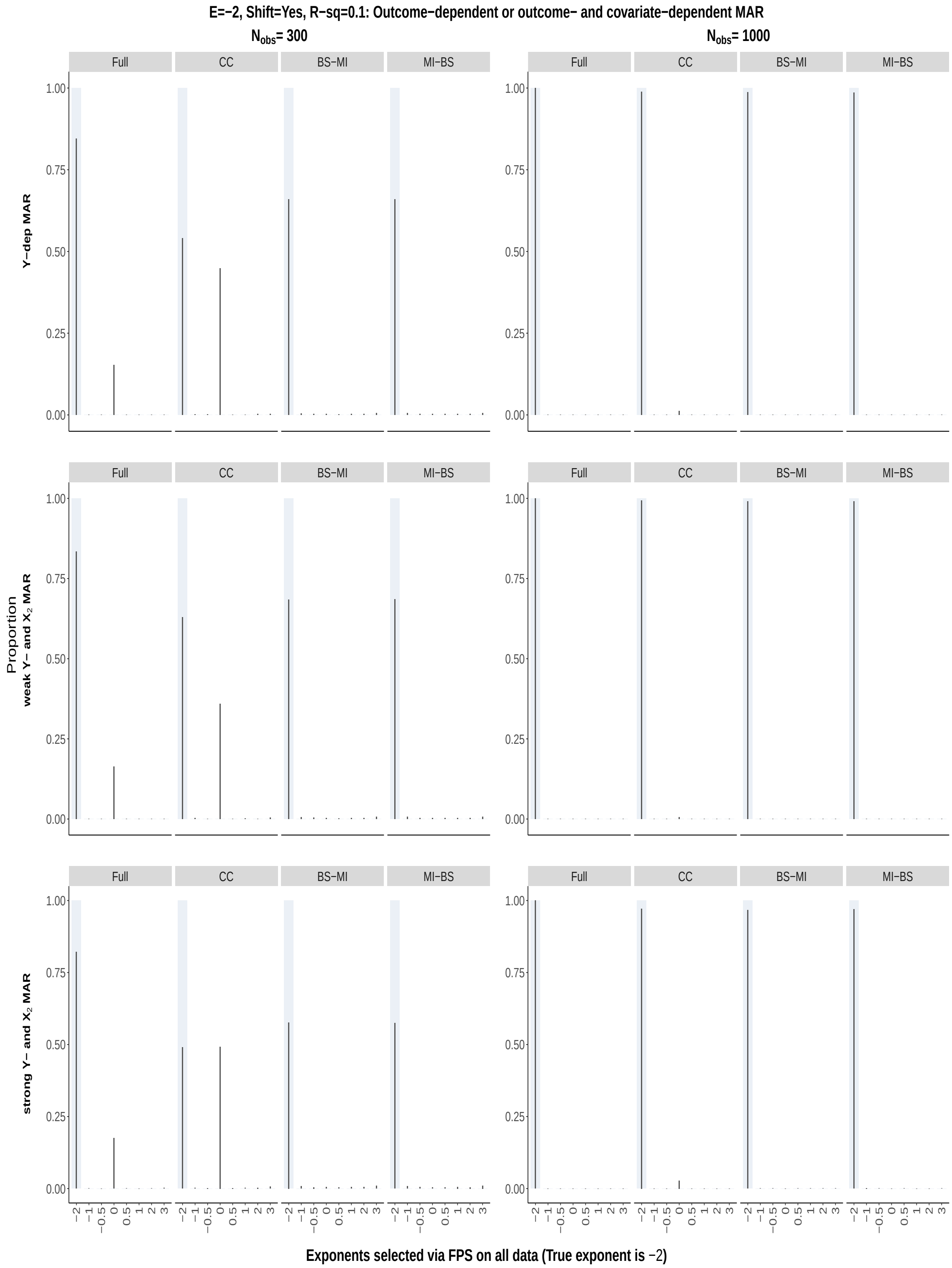

Figure S323: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

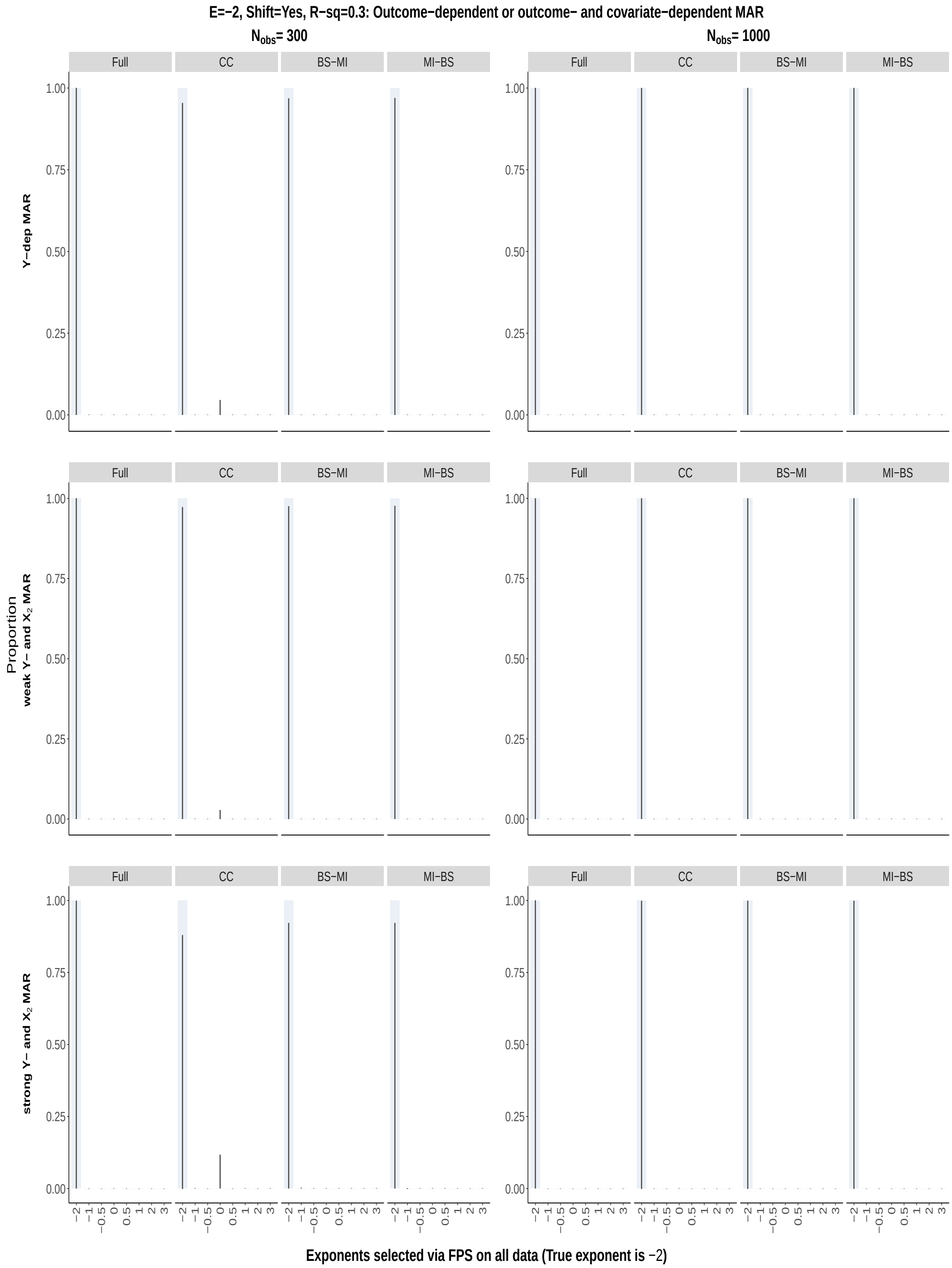

Figure S324: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S8.2.24 The 0.632 bootstrap, exponents selected using all the data: $\beta_2 = 0$, $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

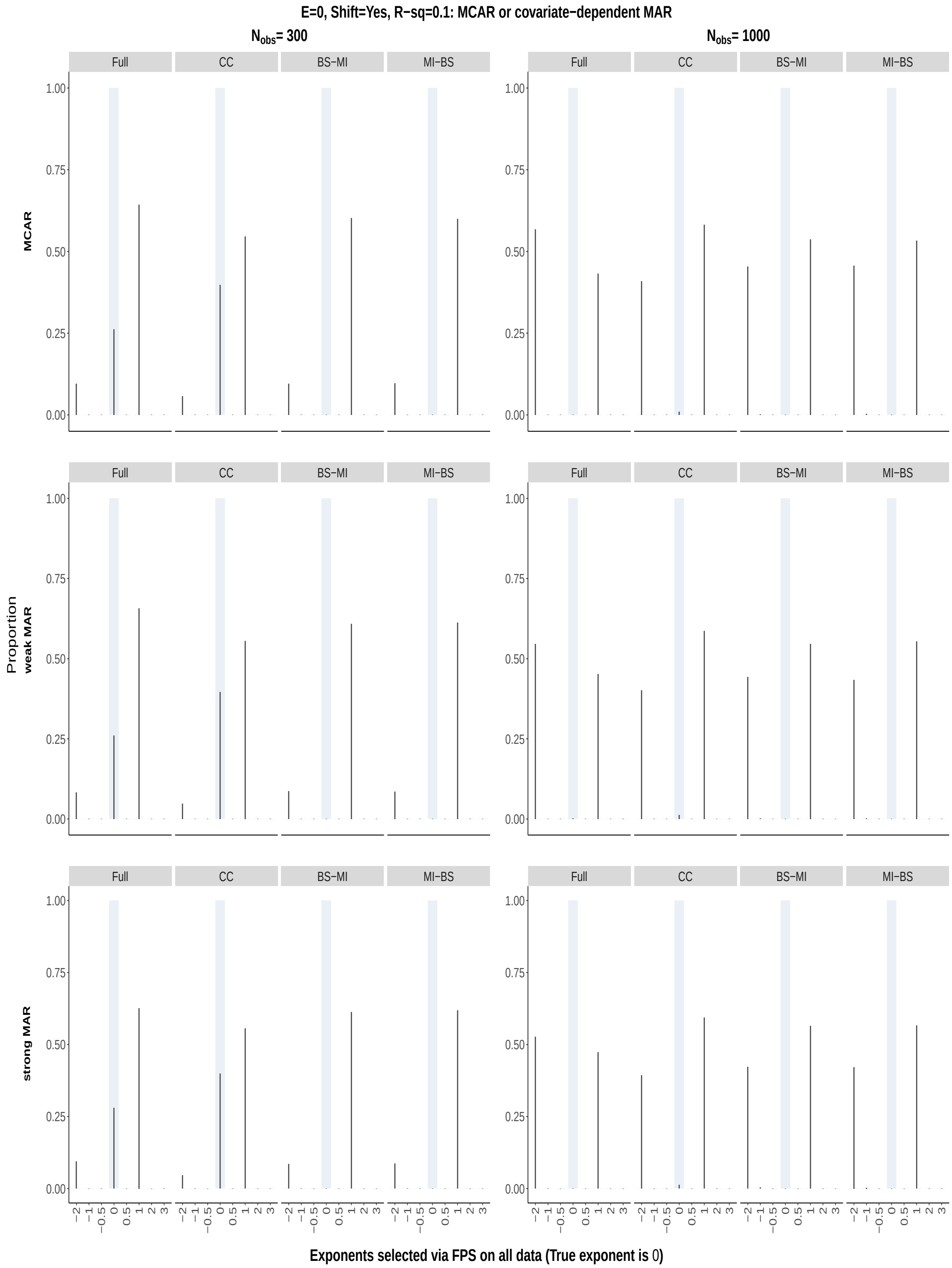

Figure S325: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

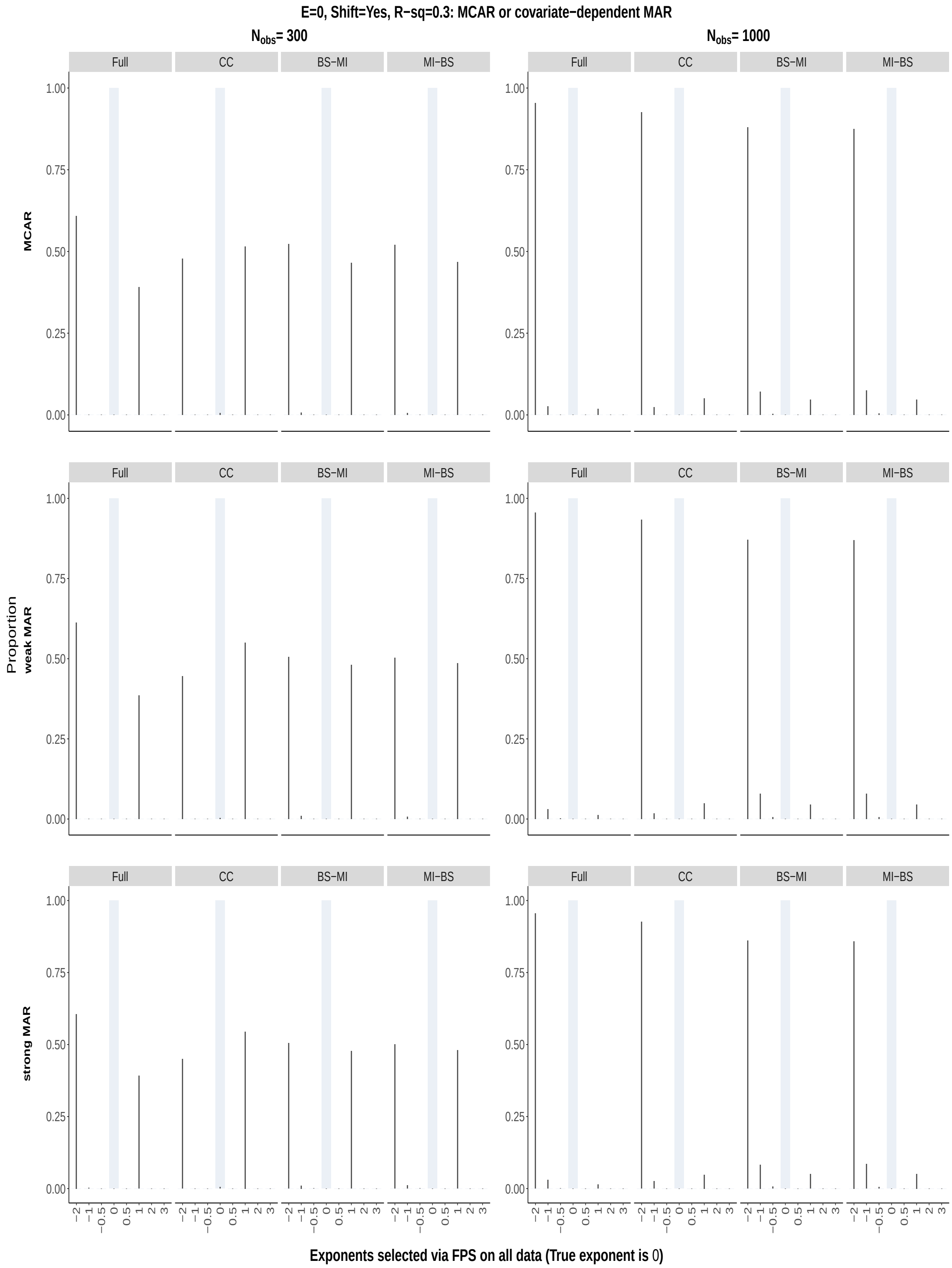

Figure S326: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

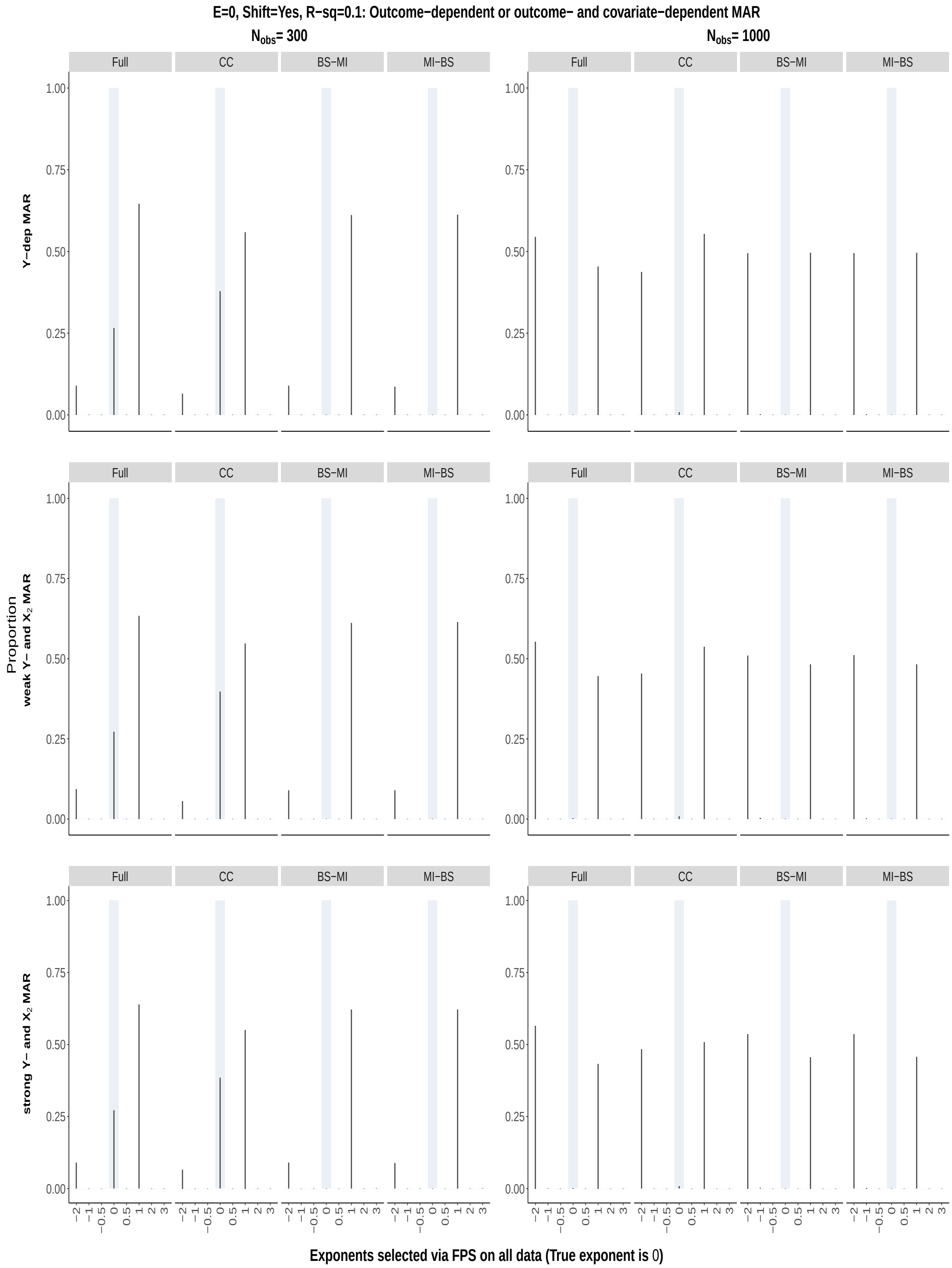

Figure S327: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

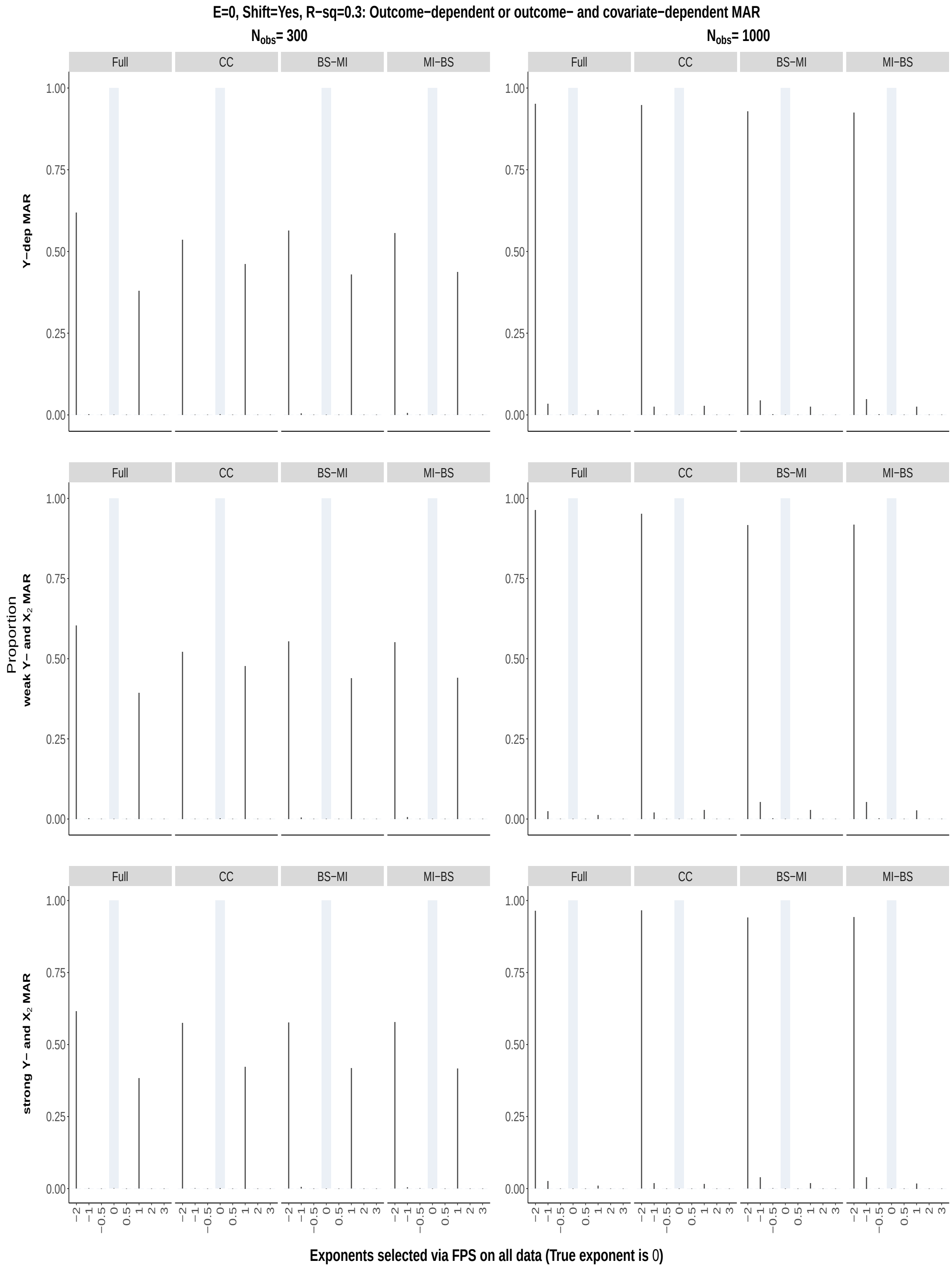

Figure S328: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

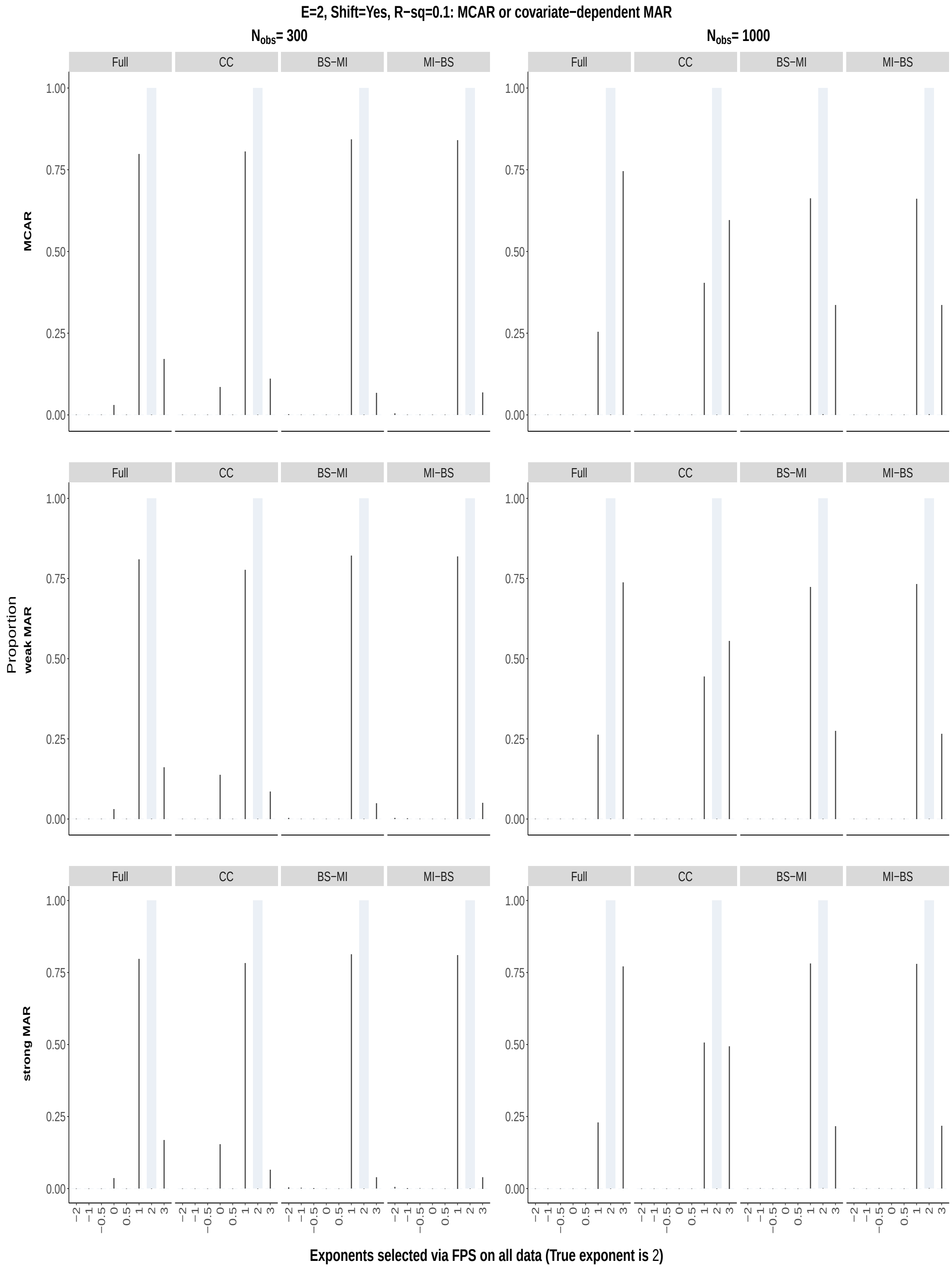

Figure S329: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

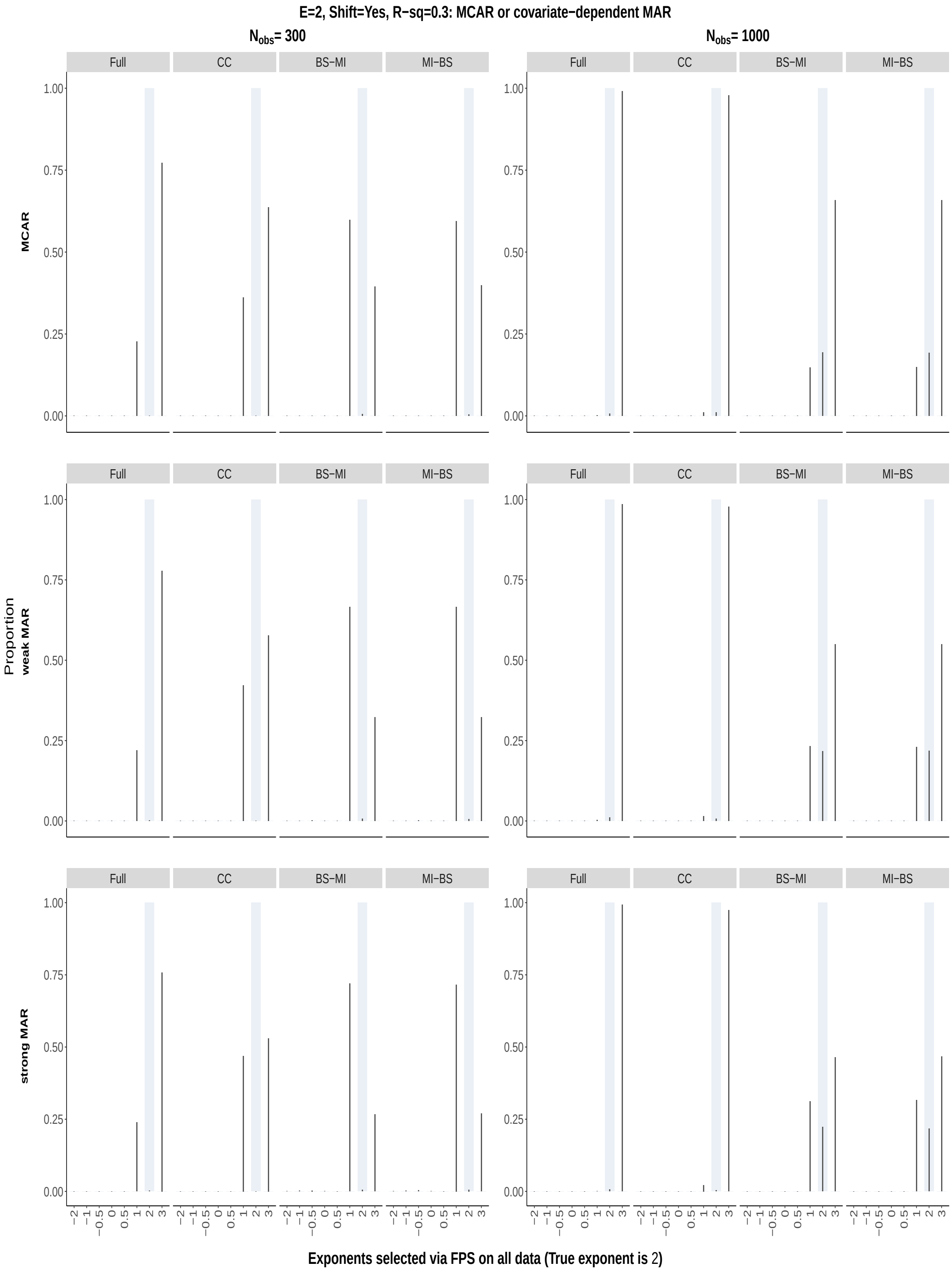

Figure S330: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

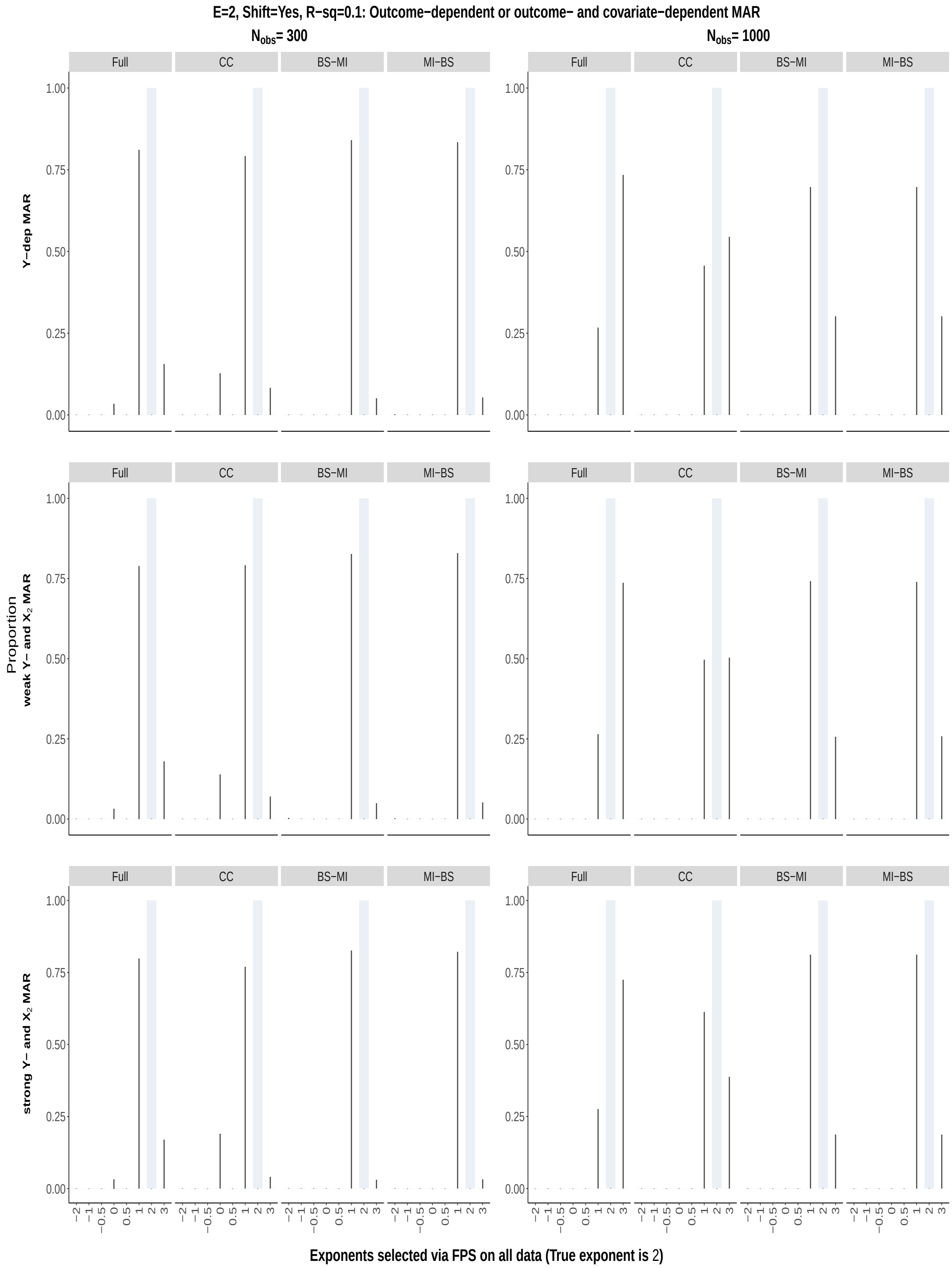

Figure S331: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

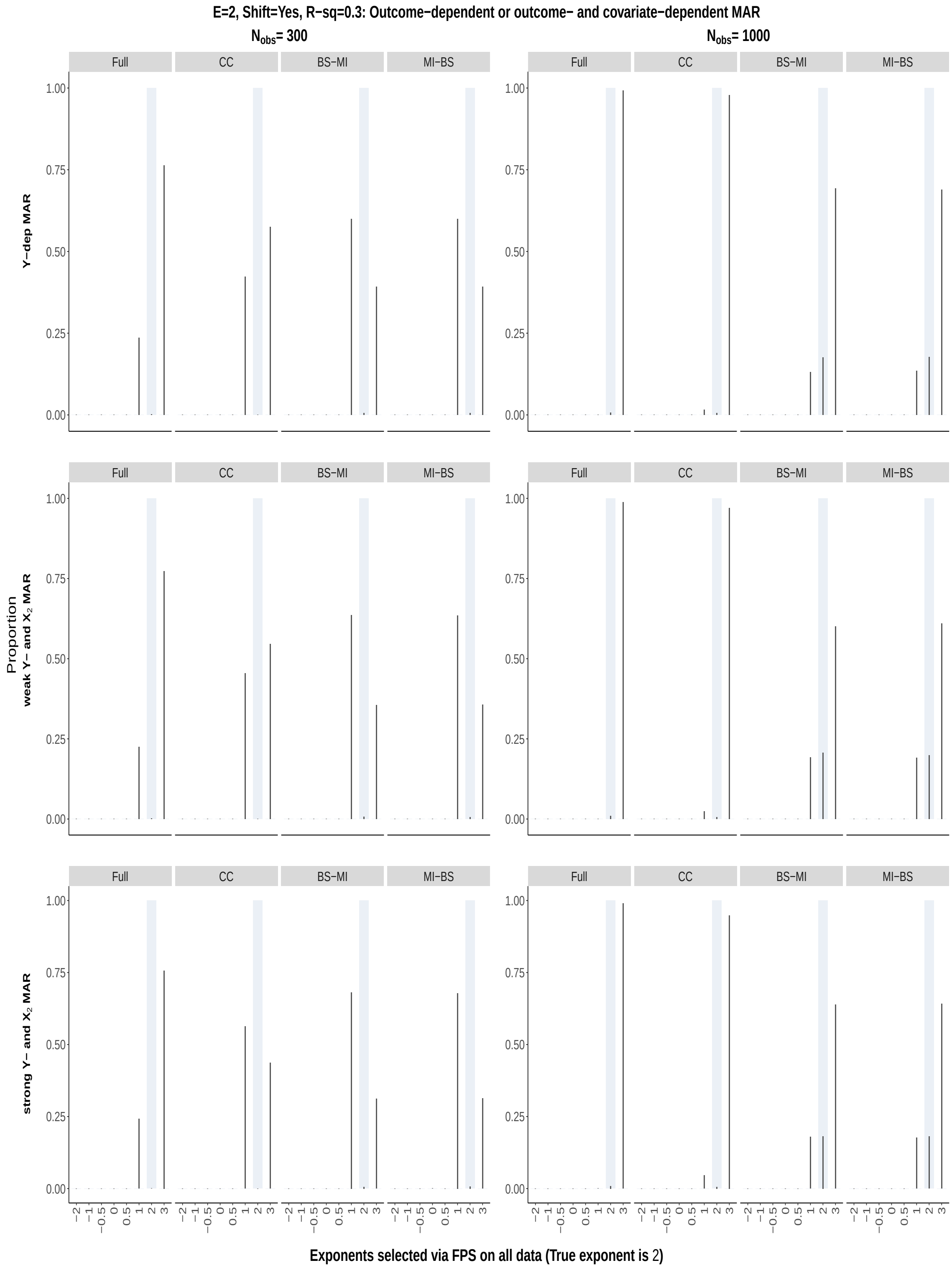

Figure S332: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

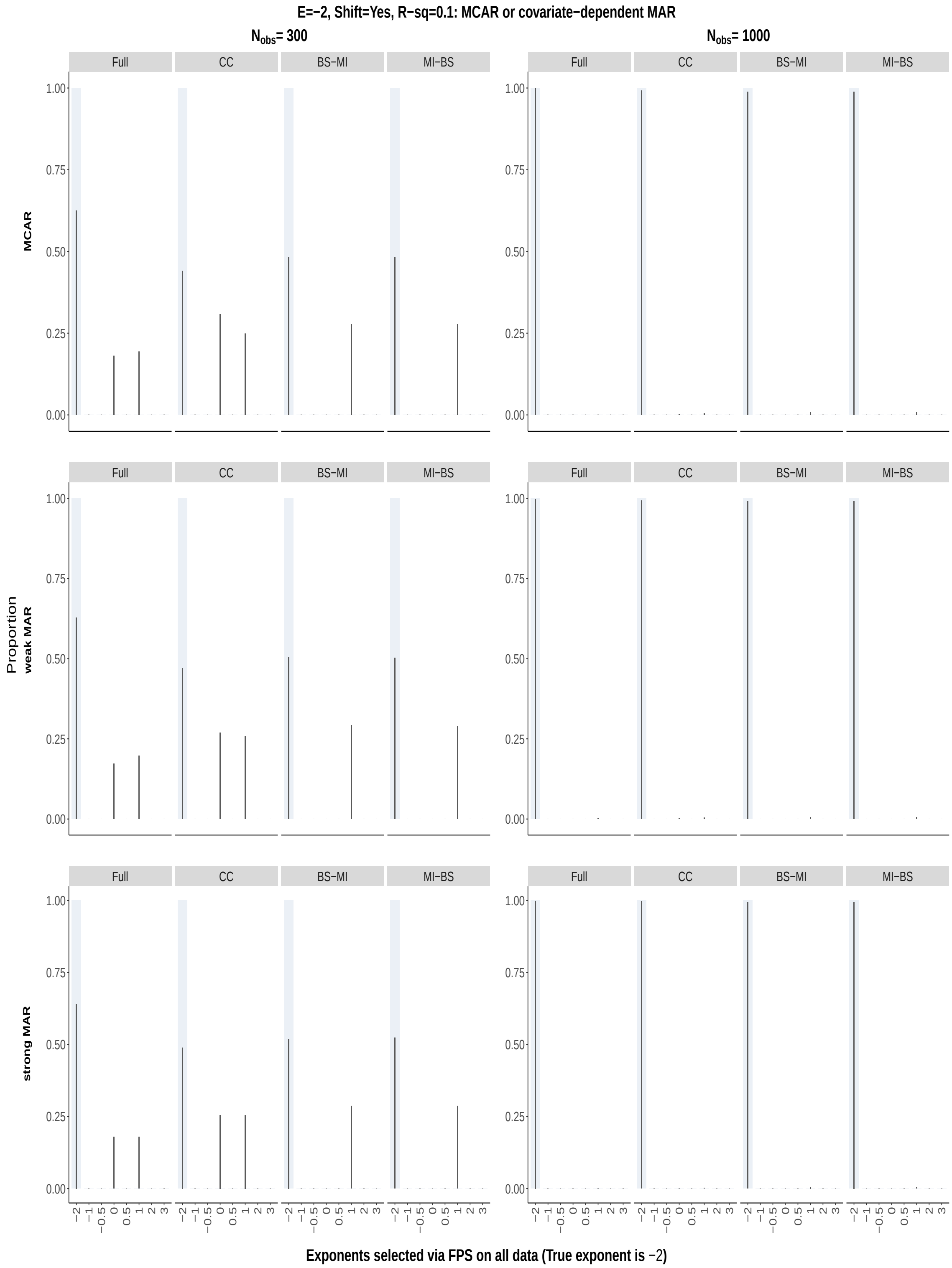

Figure S333: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

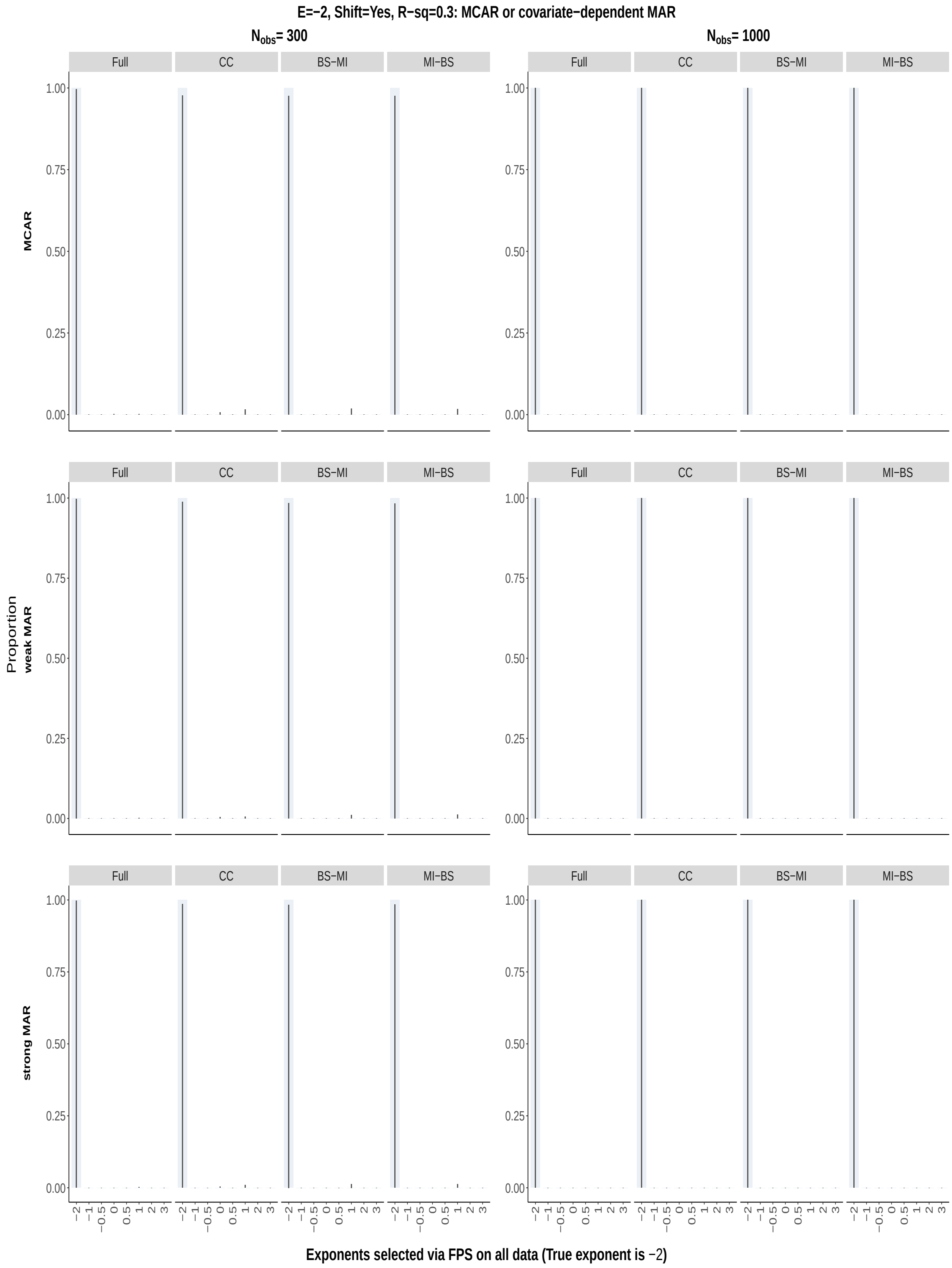

Figure S334: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

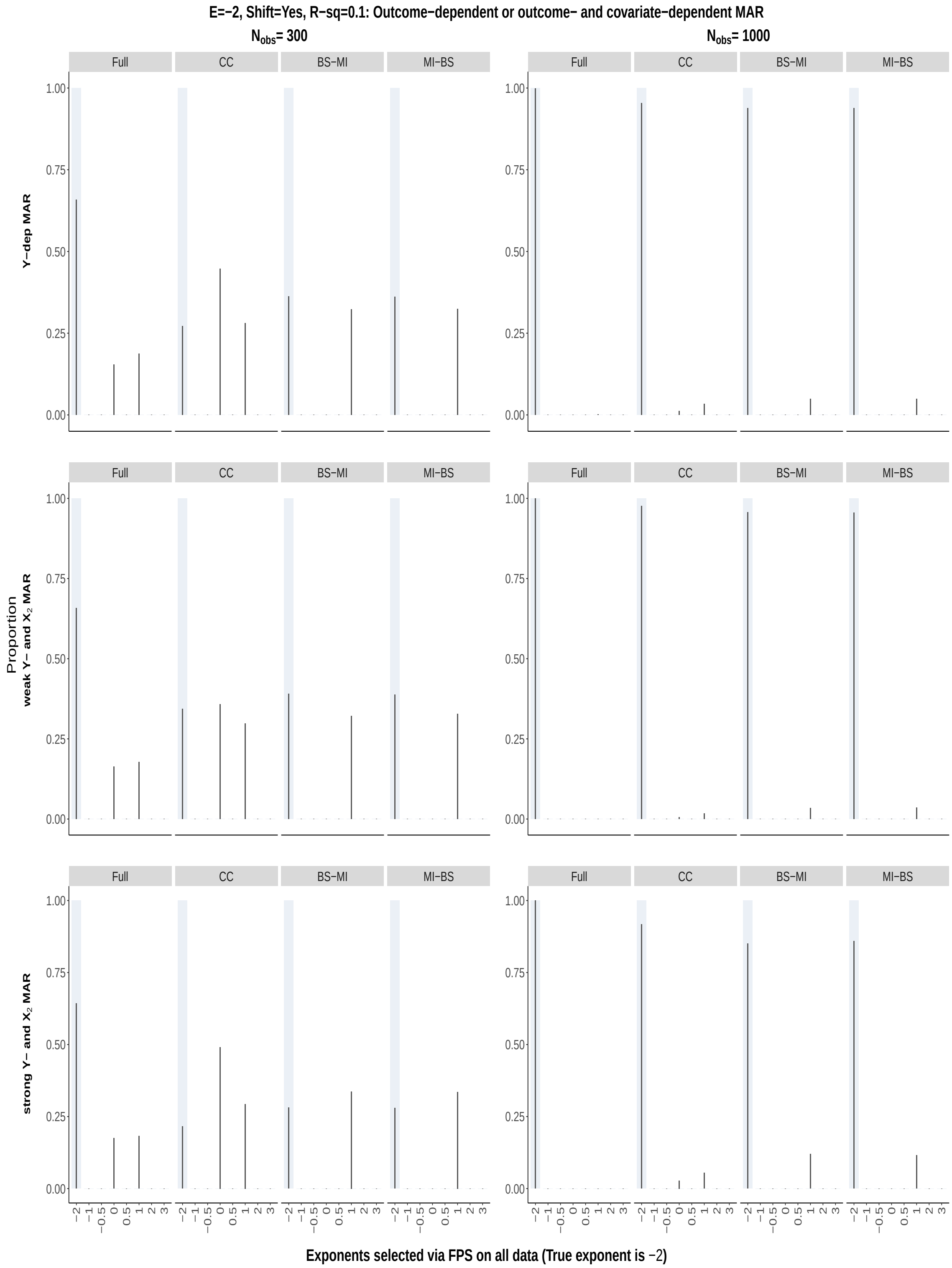

Figure S335: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

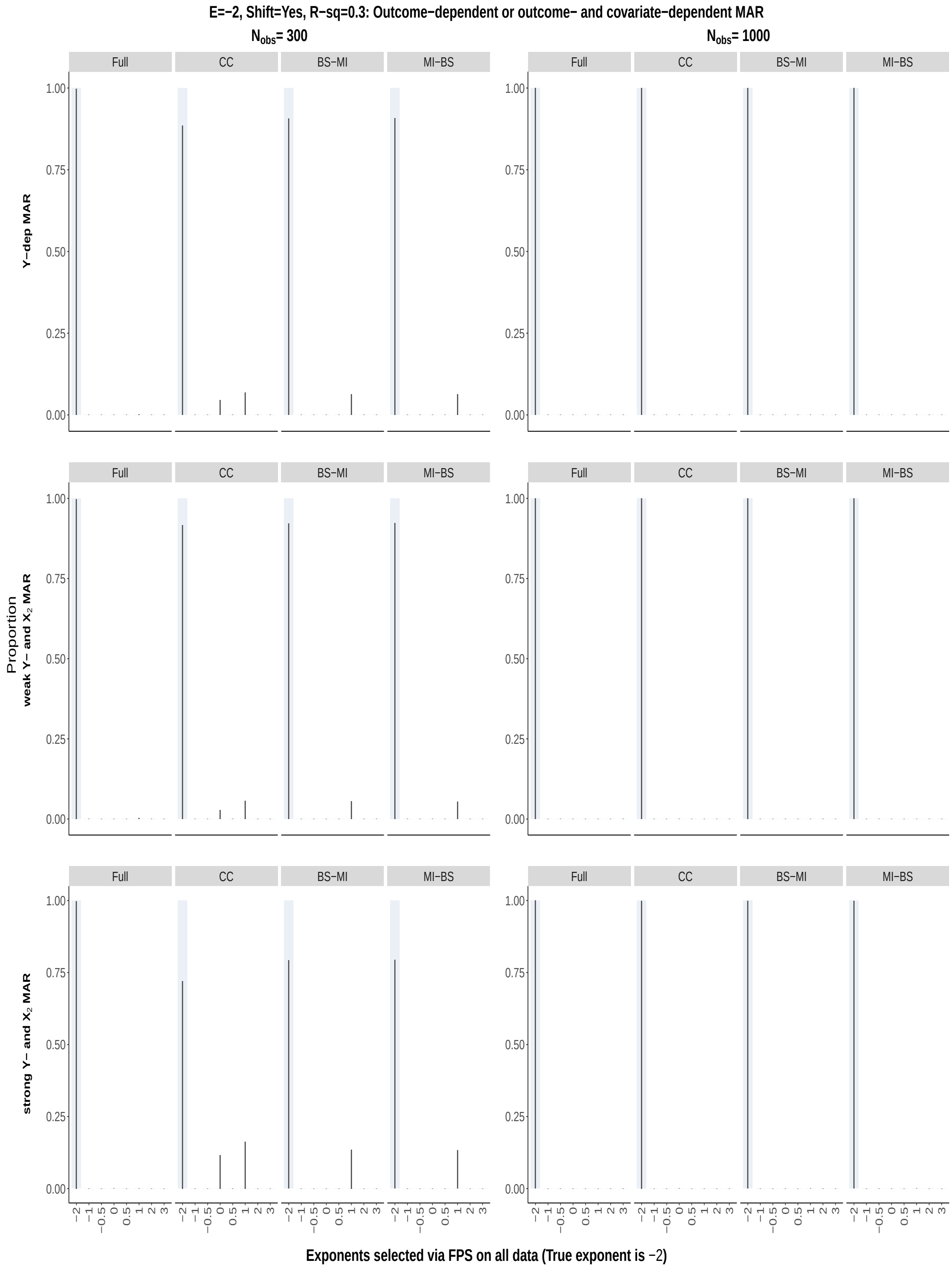

Figure S336: The proportion of times an exponent was selected via FPS post-imputing when $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The pale blue bar highlights the underlying 'true' exponent, while the black bars represent the proportion of times across the 2000 repetitions, M imputed datasets, that each exponent was selected via the FPS algorithm. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S9 Chapter 9: Simulation study results for MFP, covariate selection (Section 9.6)

S9.1 Cross-validation

S9.1.1 Covariate selection of X_2: $\beta_2 = 1$, $\alpha_E = 1$, no origin-shift

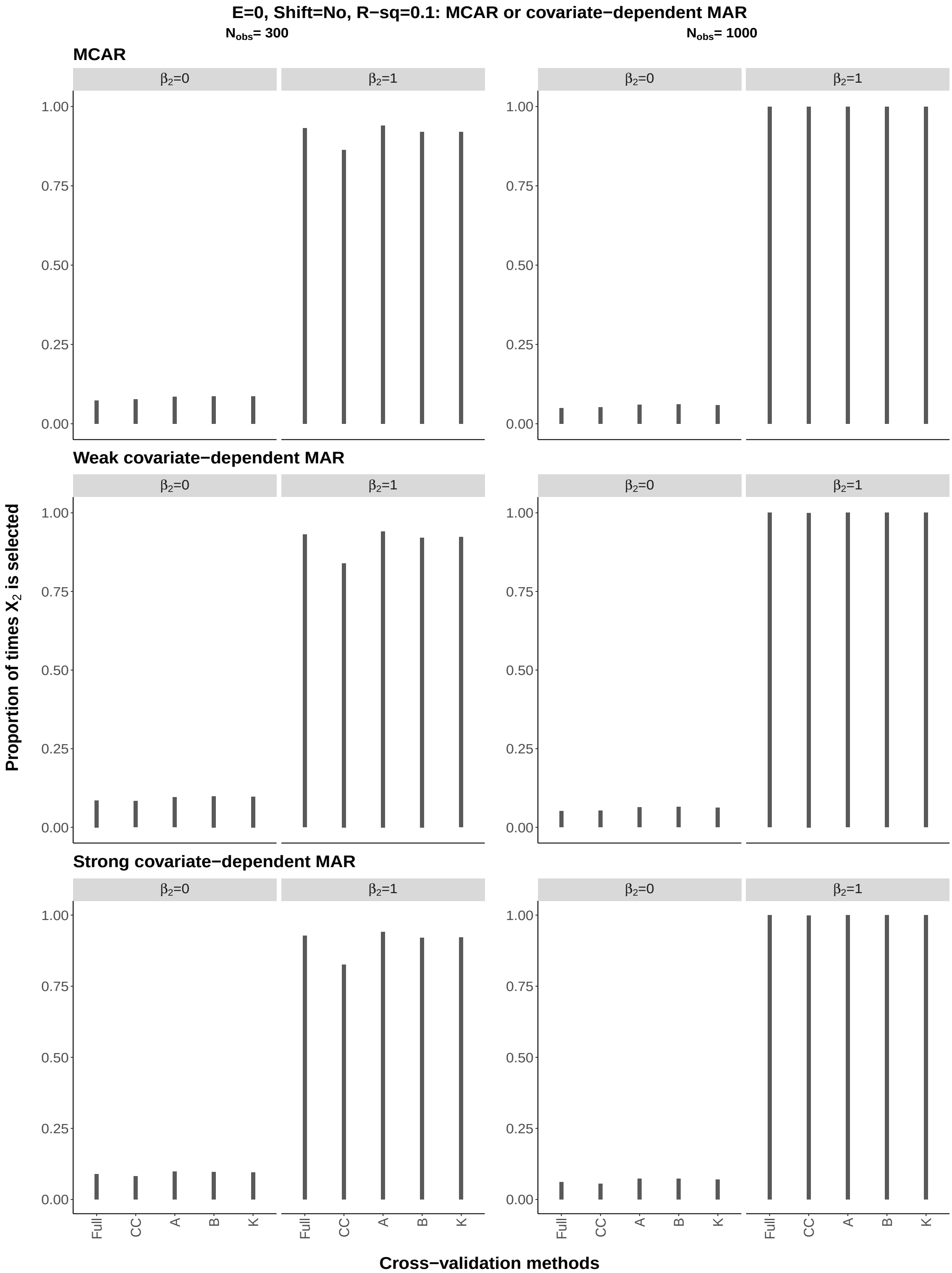

Figure S1: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was

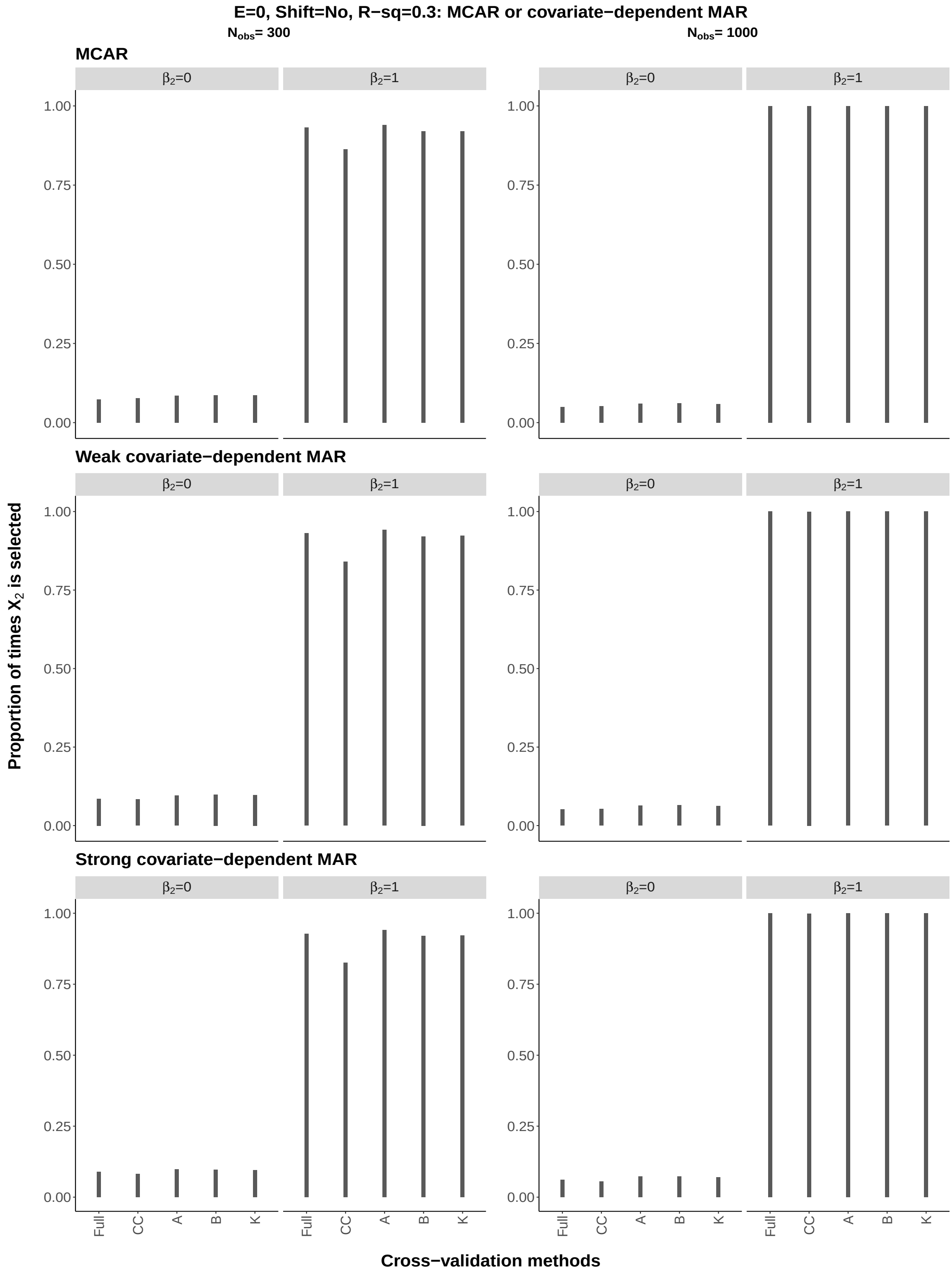

Figure S2: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

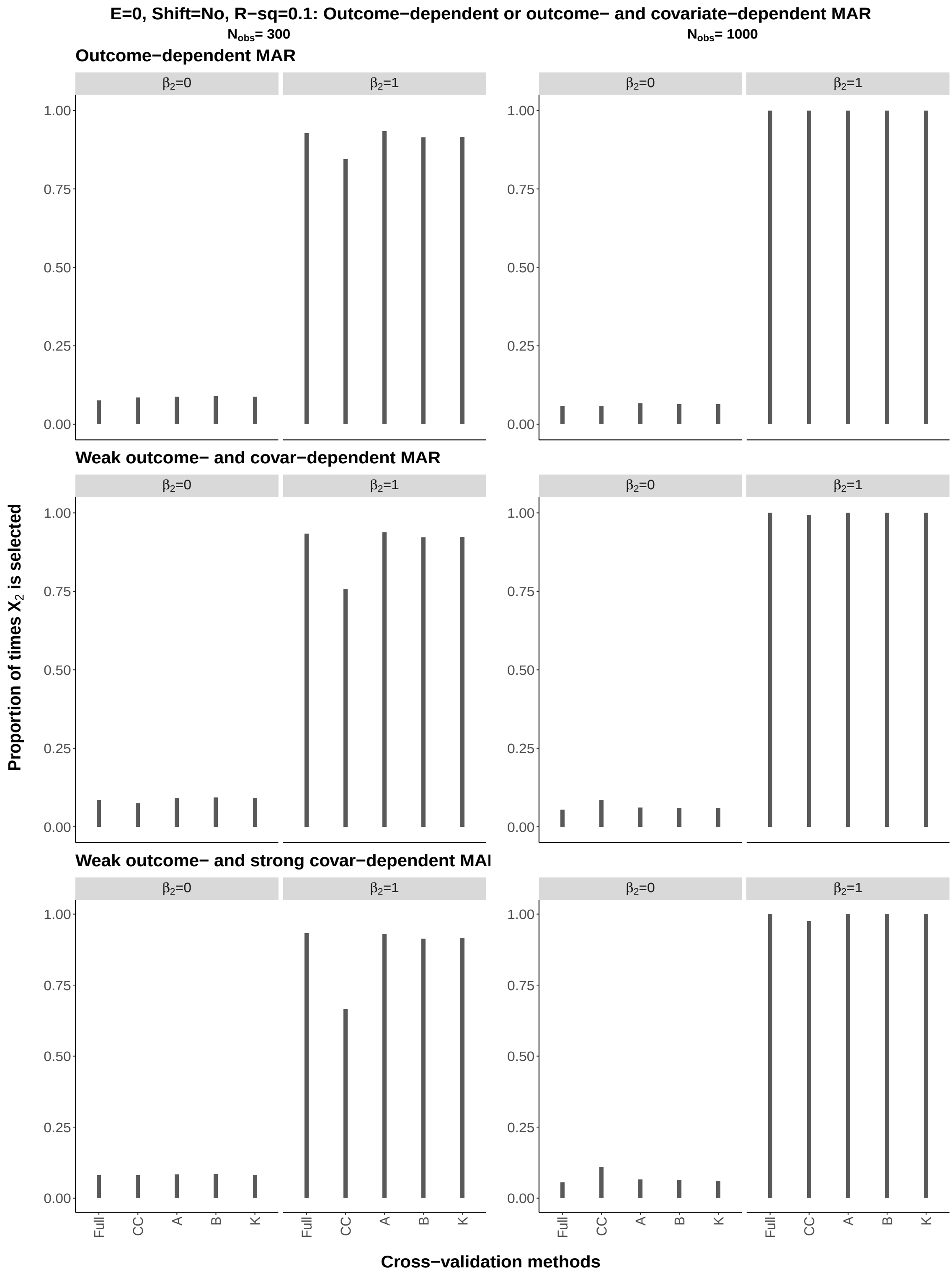

Figure S3: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

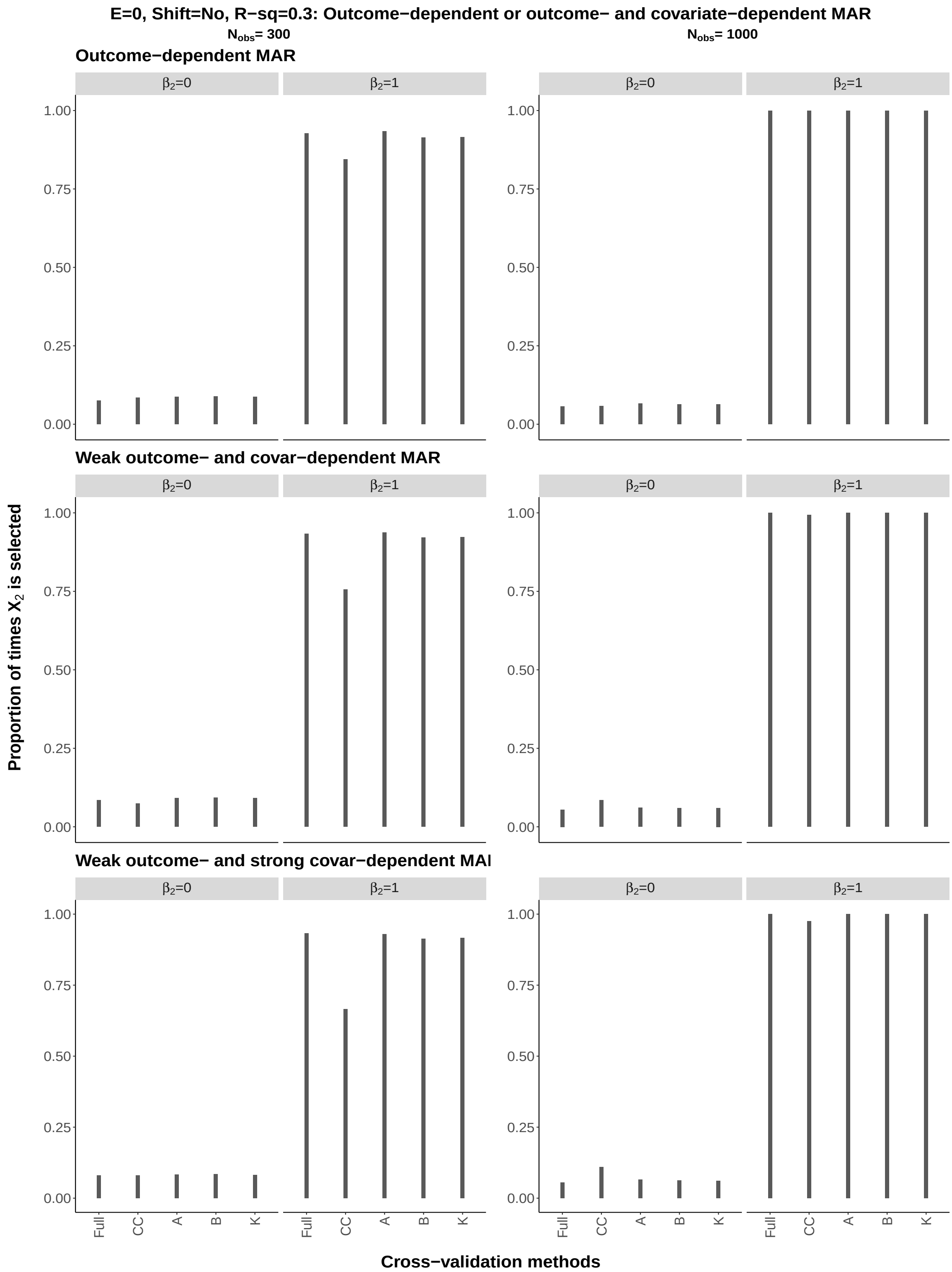

Figure S4: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

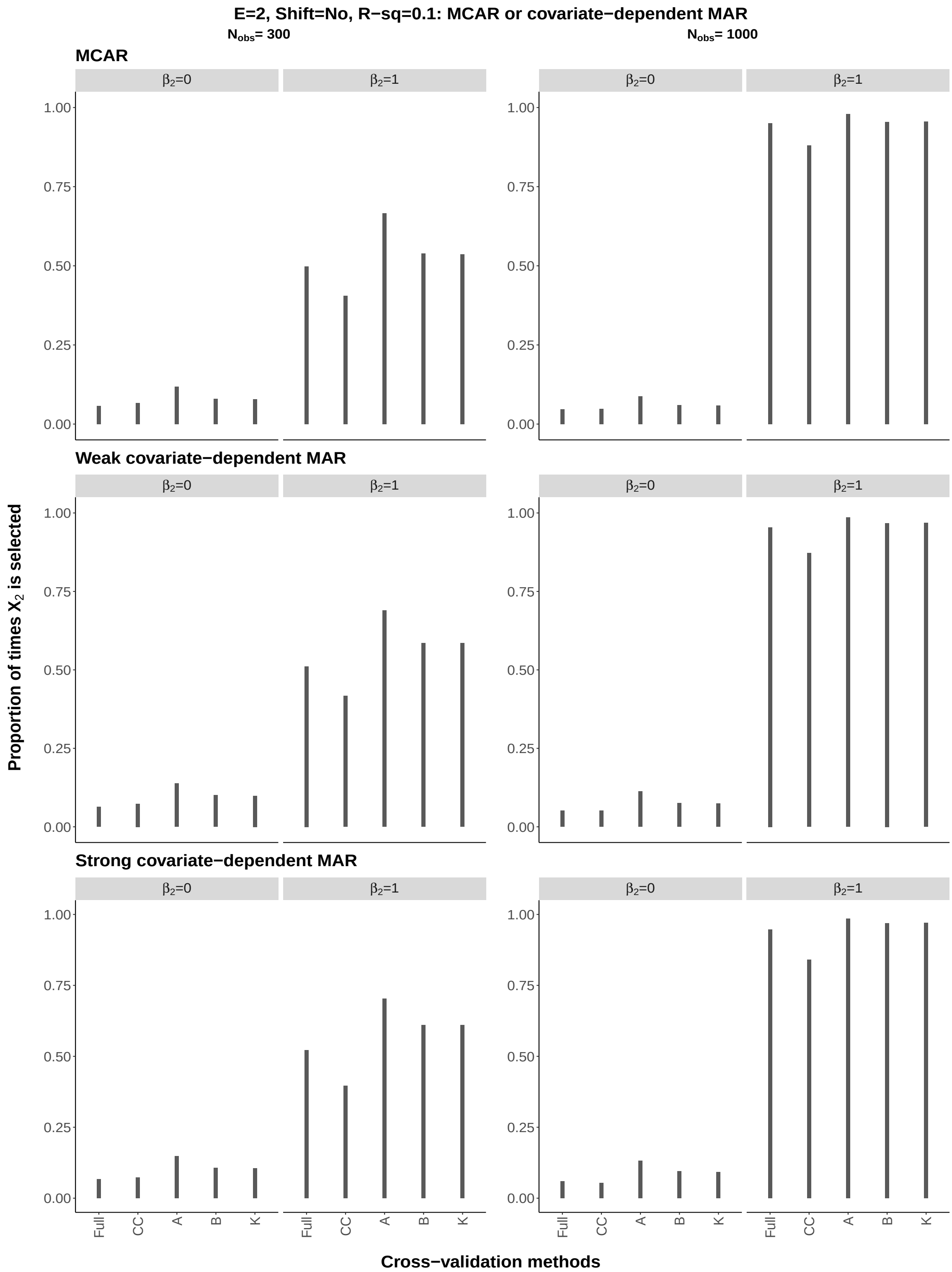

Figure S5: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

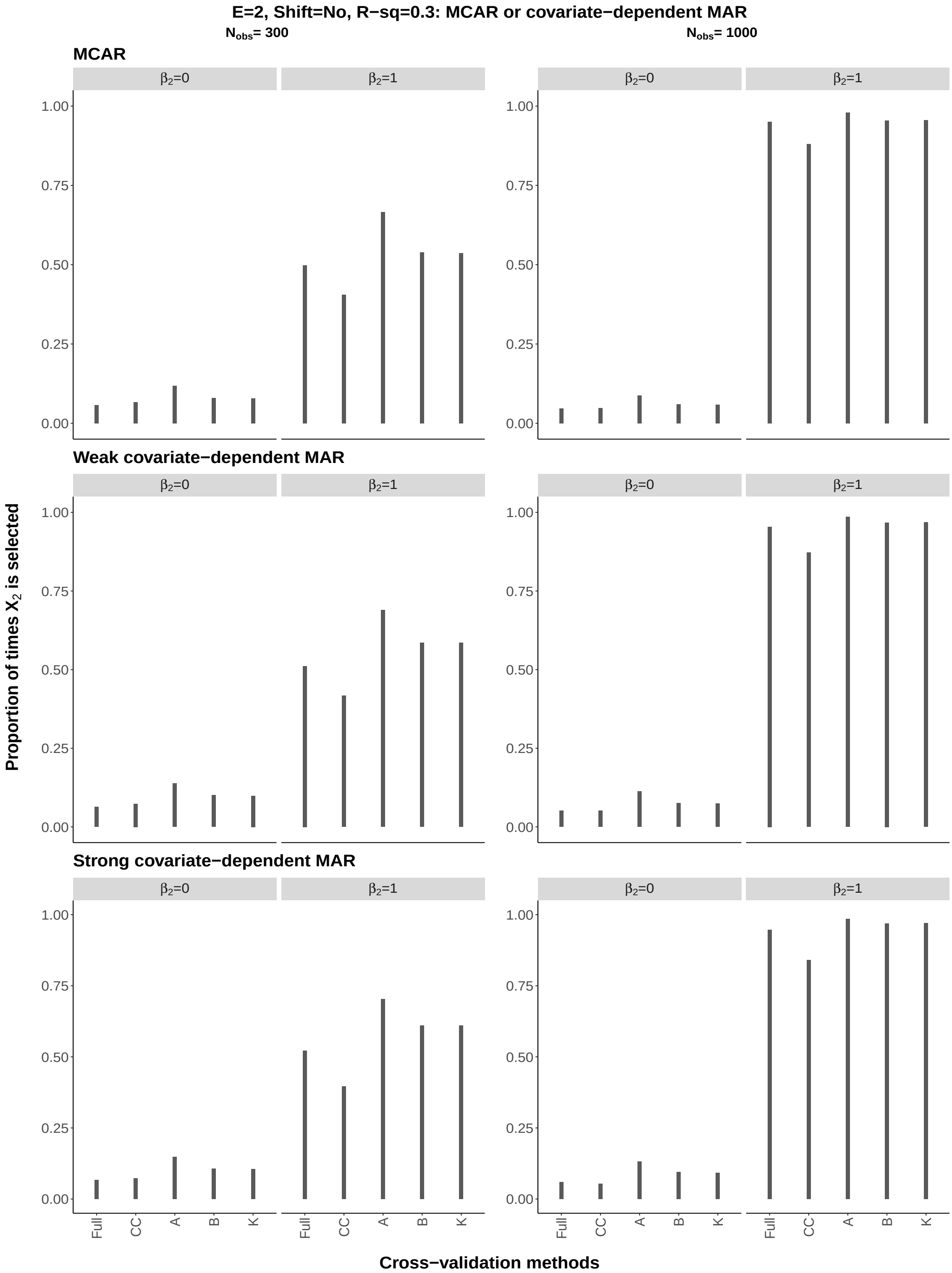

Figure S6: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

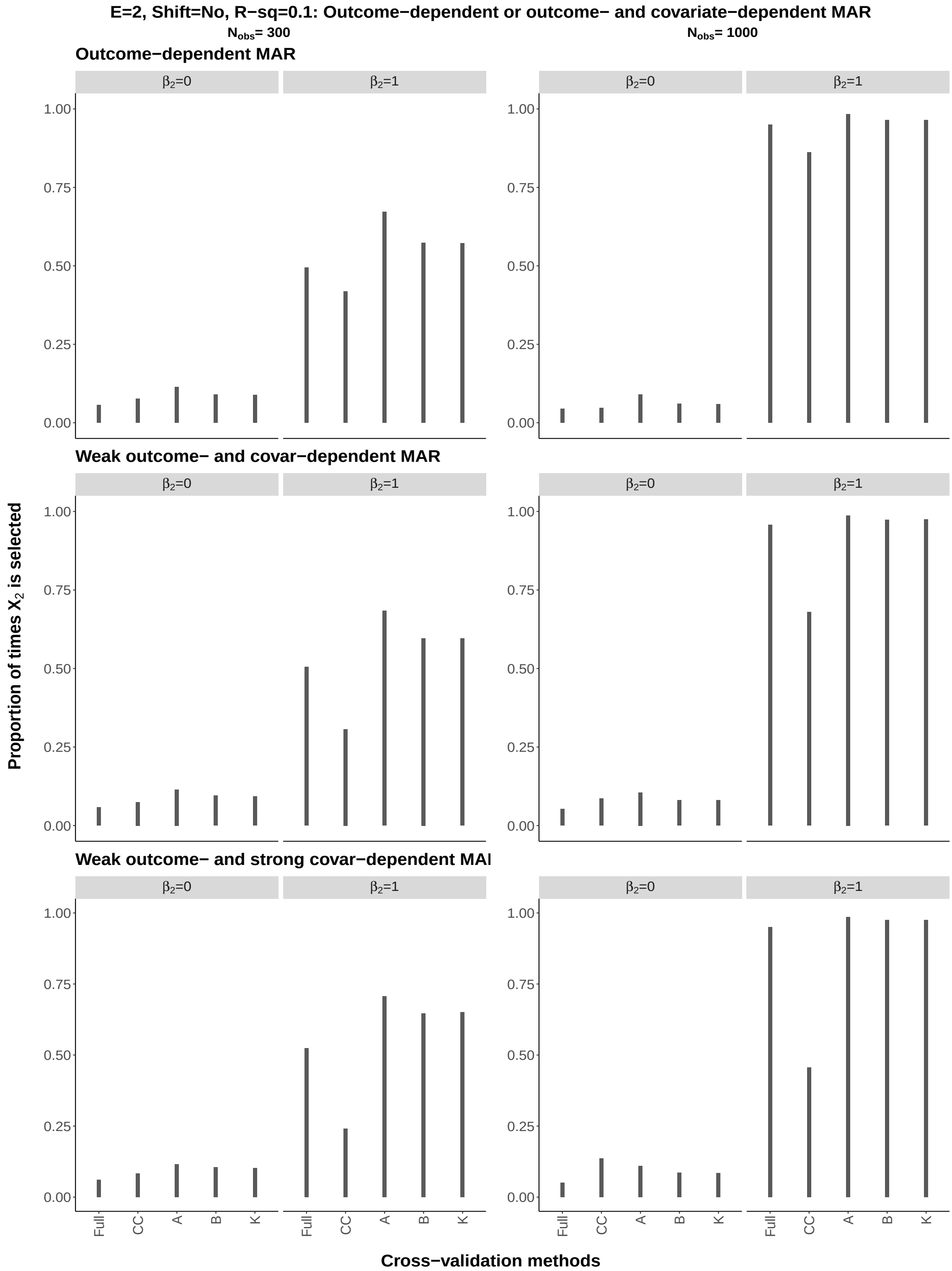

Figure S7: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

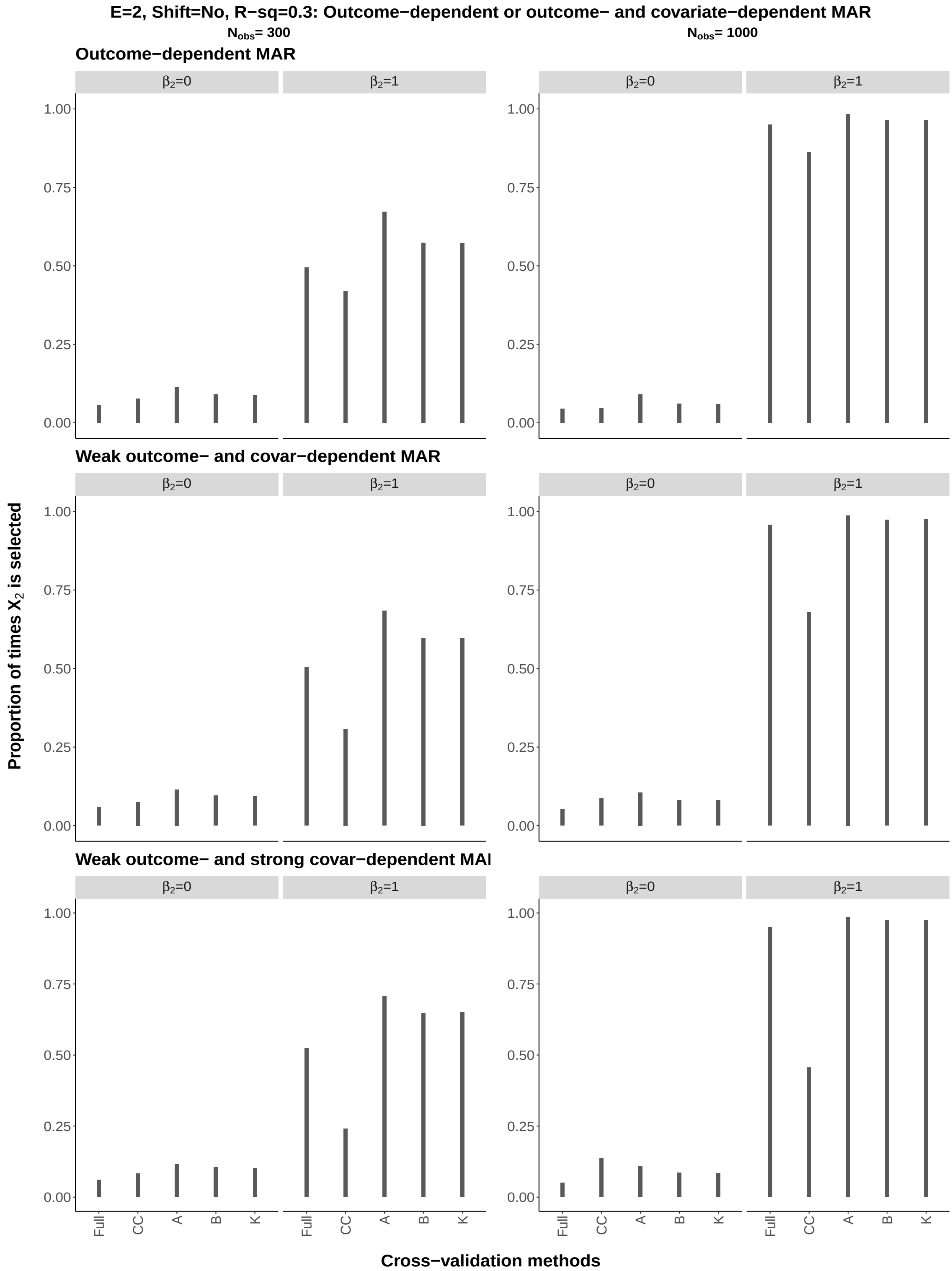

Figure S8: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

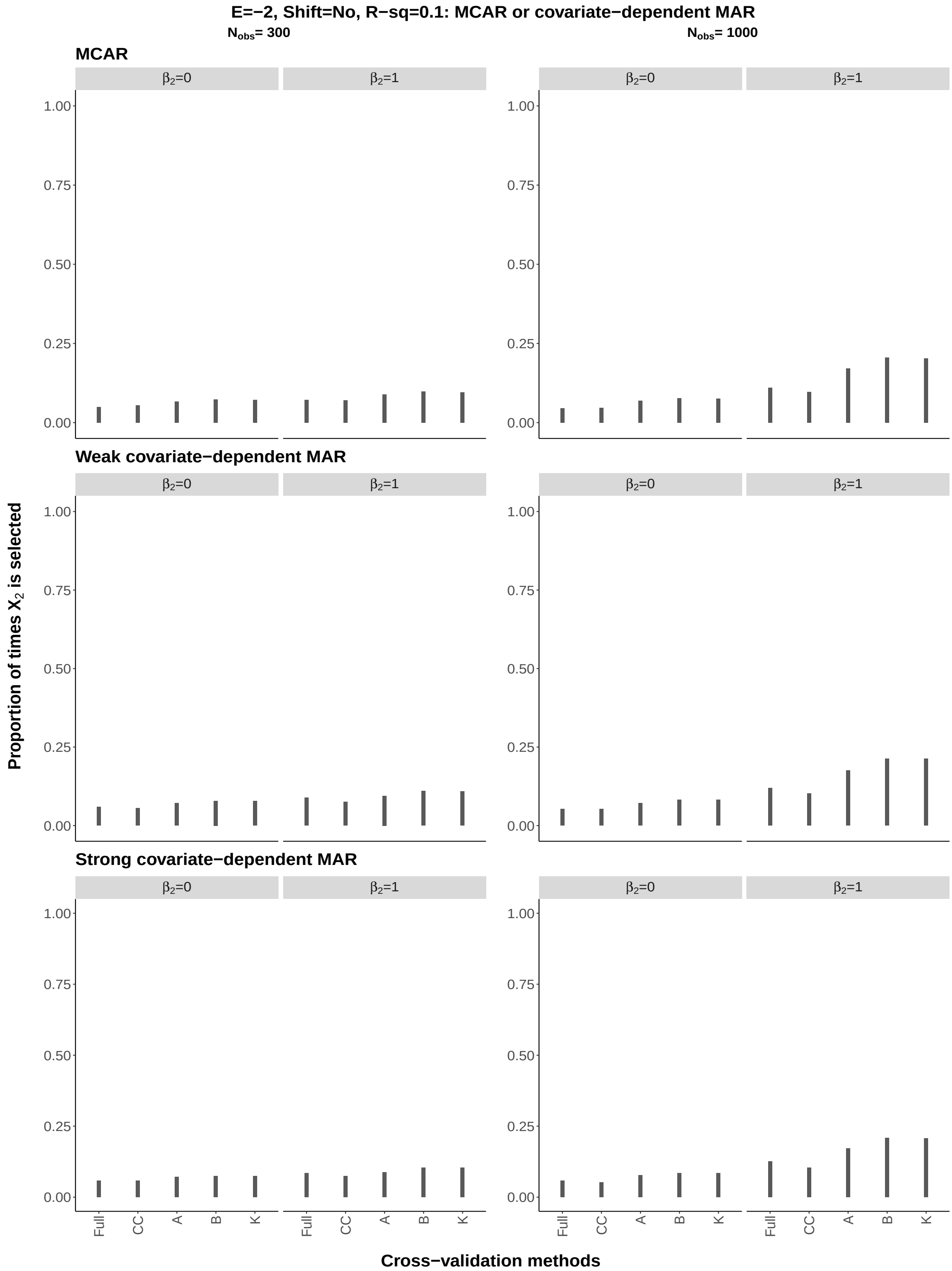

Figure S9: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

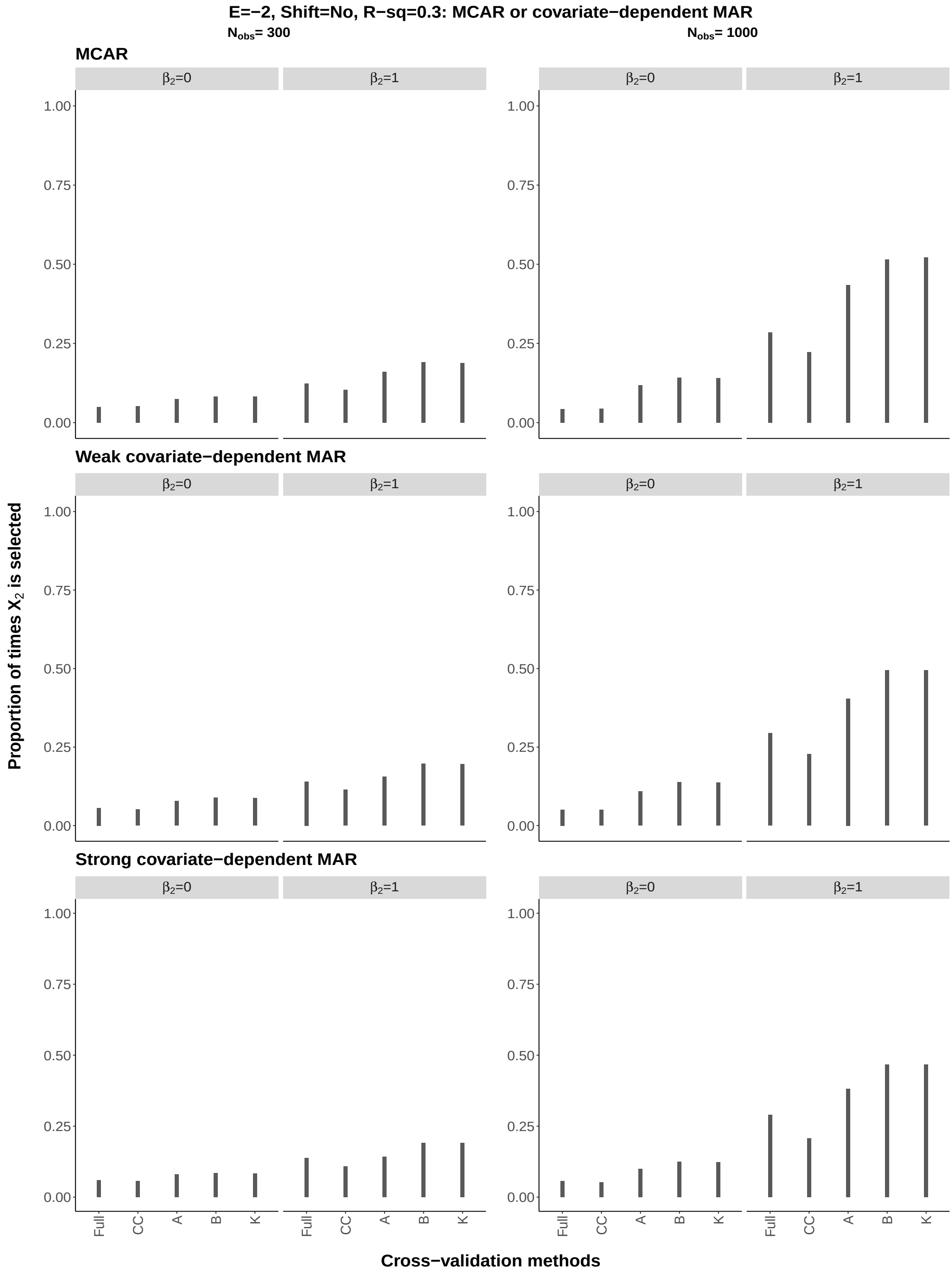

Figure S10: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

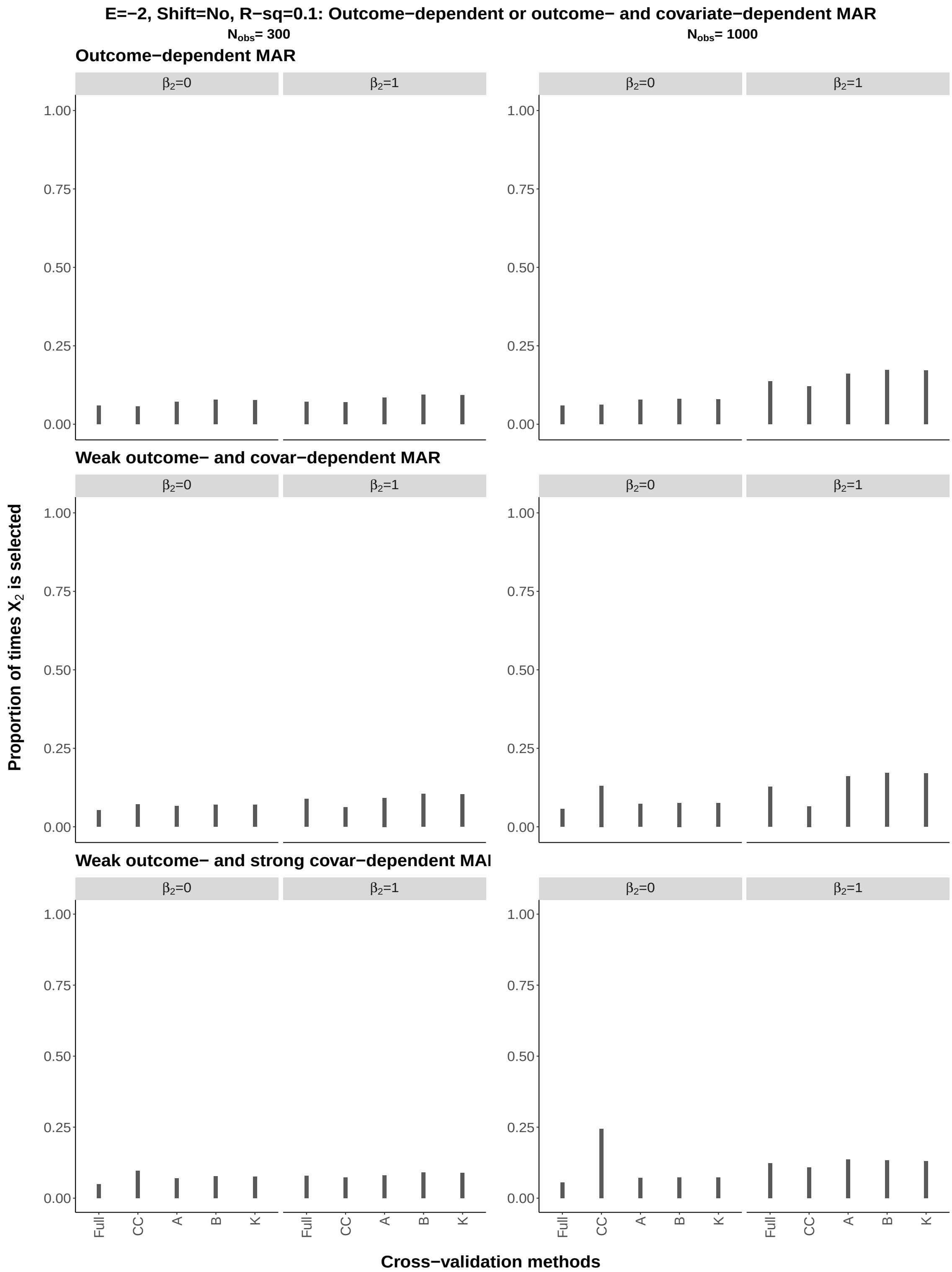

Figure S11: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

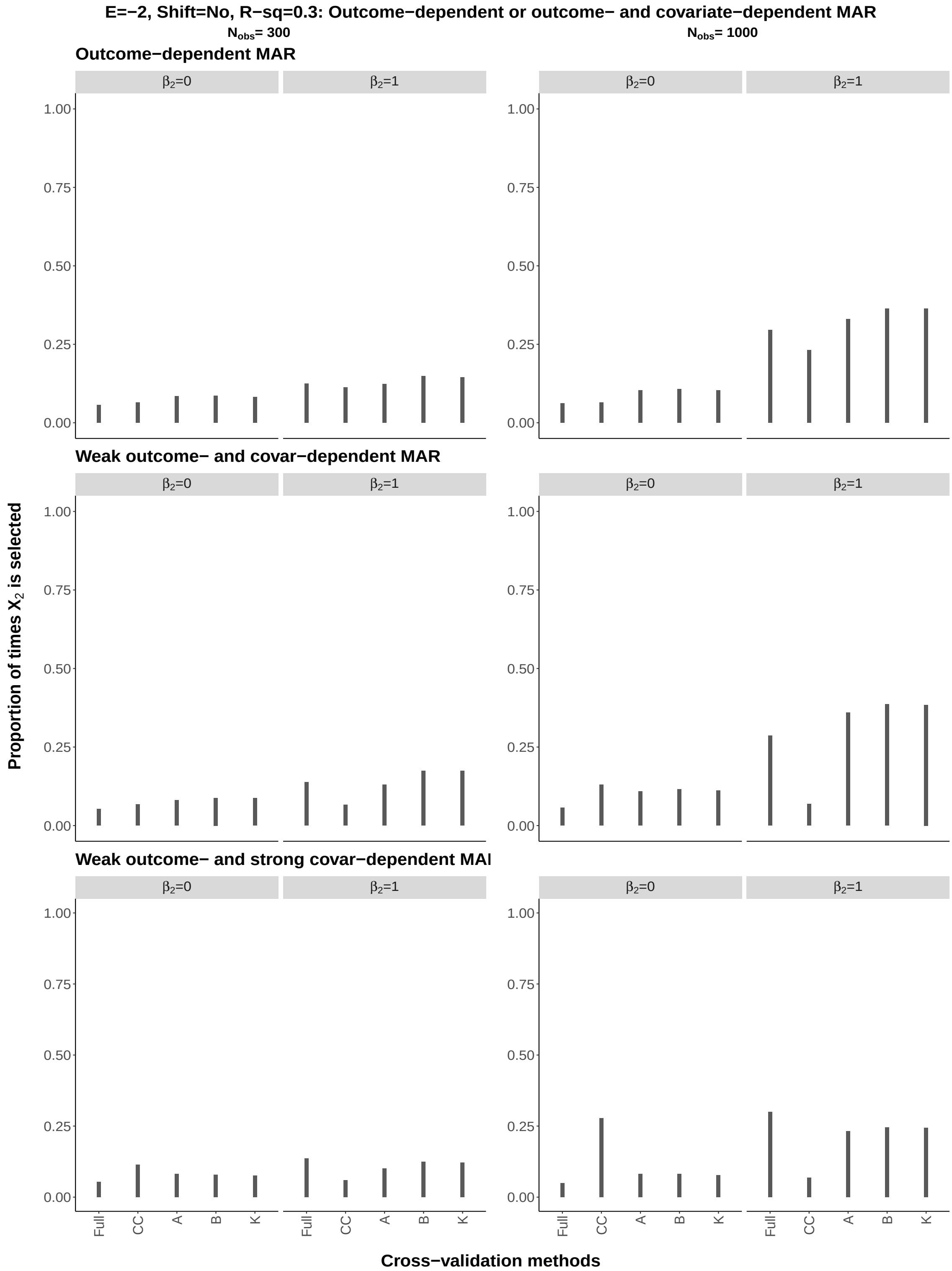

Figure S12: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.1.2 Covariate selection of X_2: $\beta_2 = 1$, $\alpha_E = 0.05$, no origin-shift

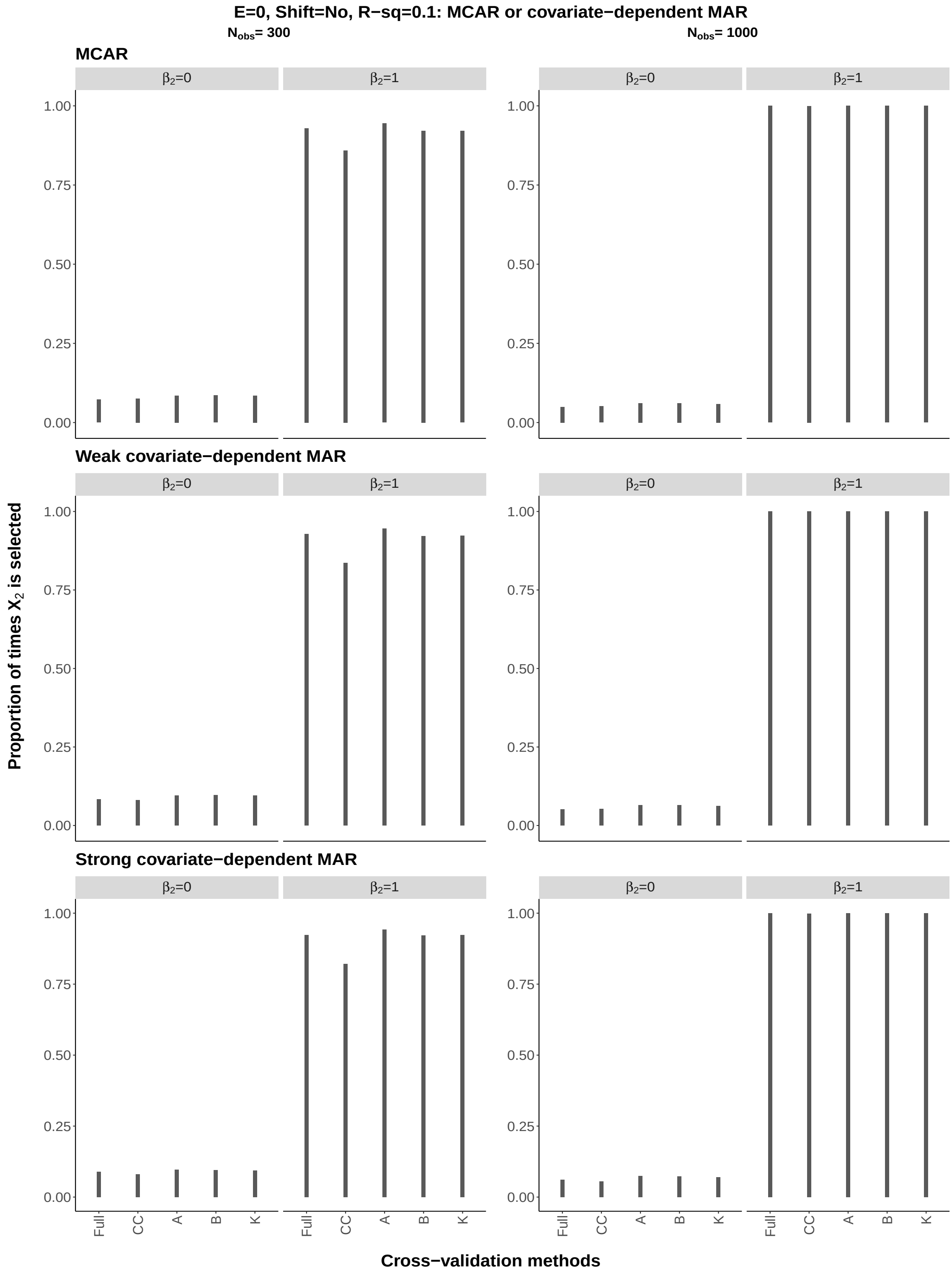

Figure S13: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

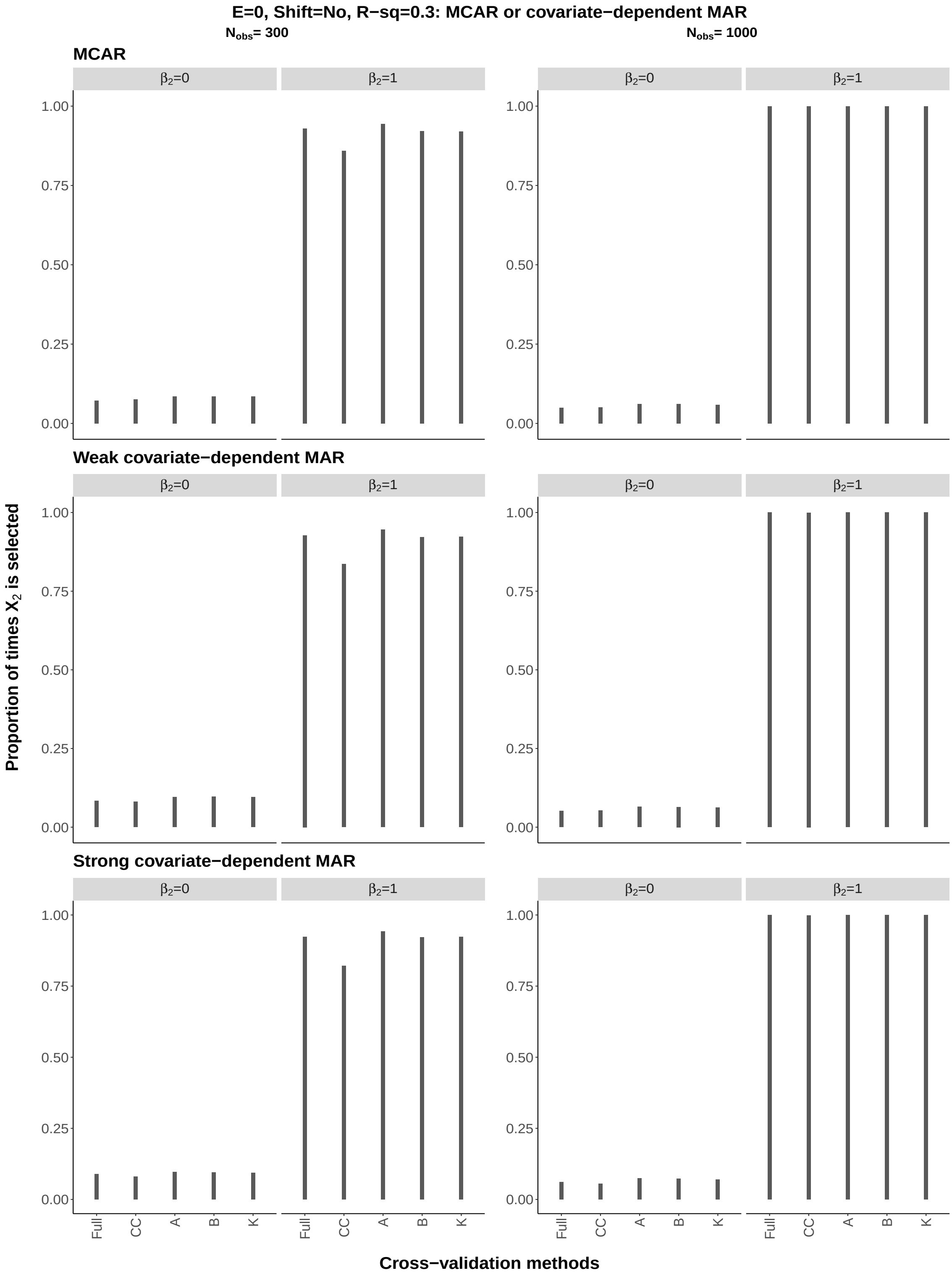

Figure S14: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

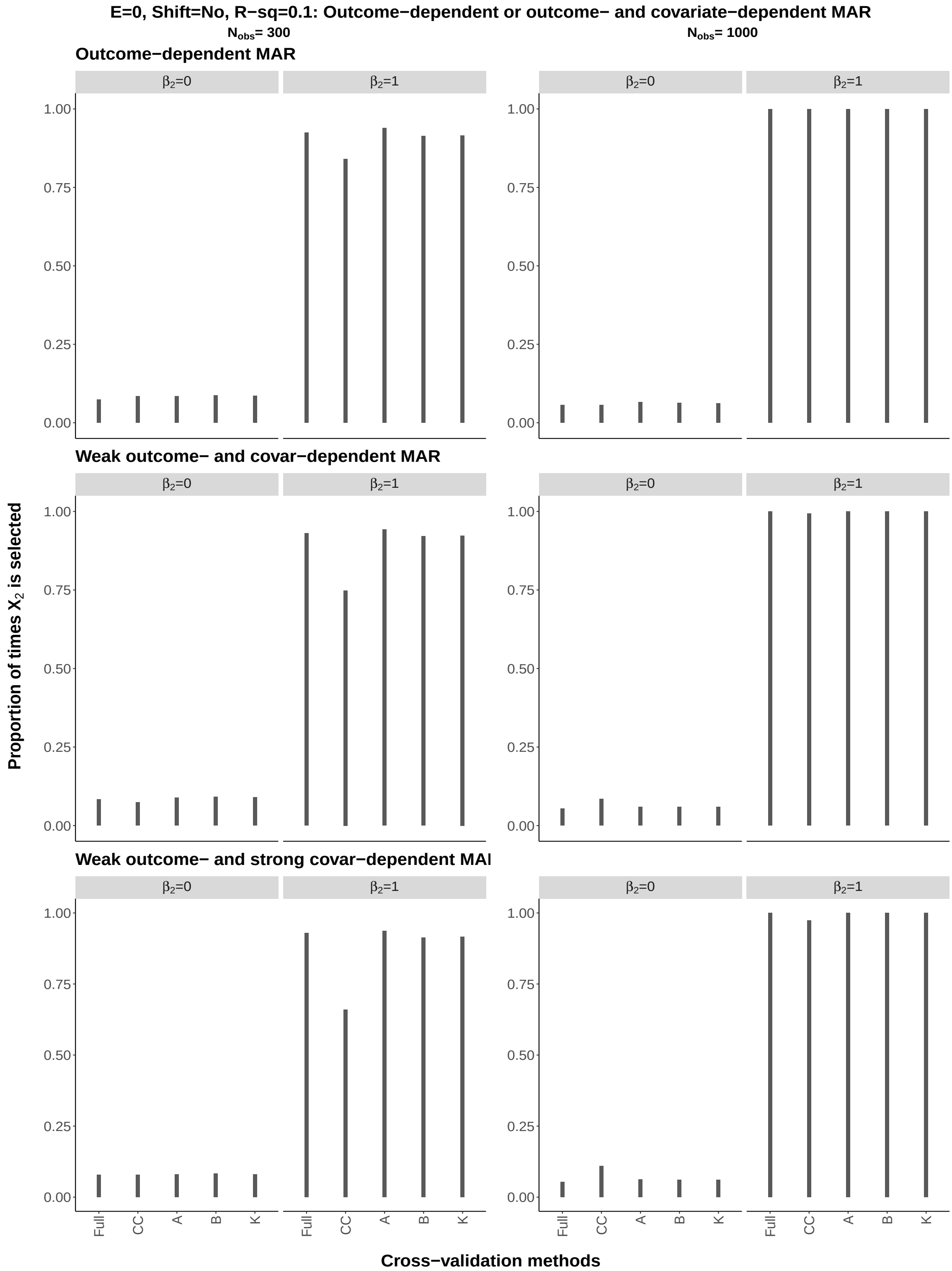

Figure S15: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

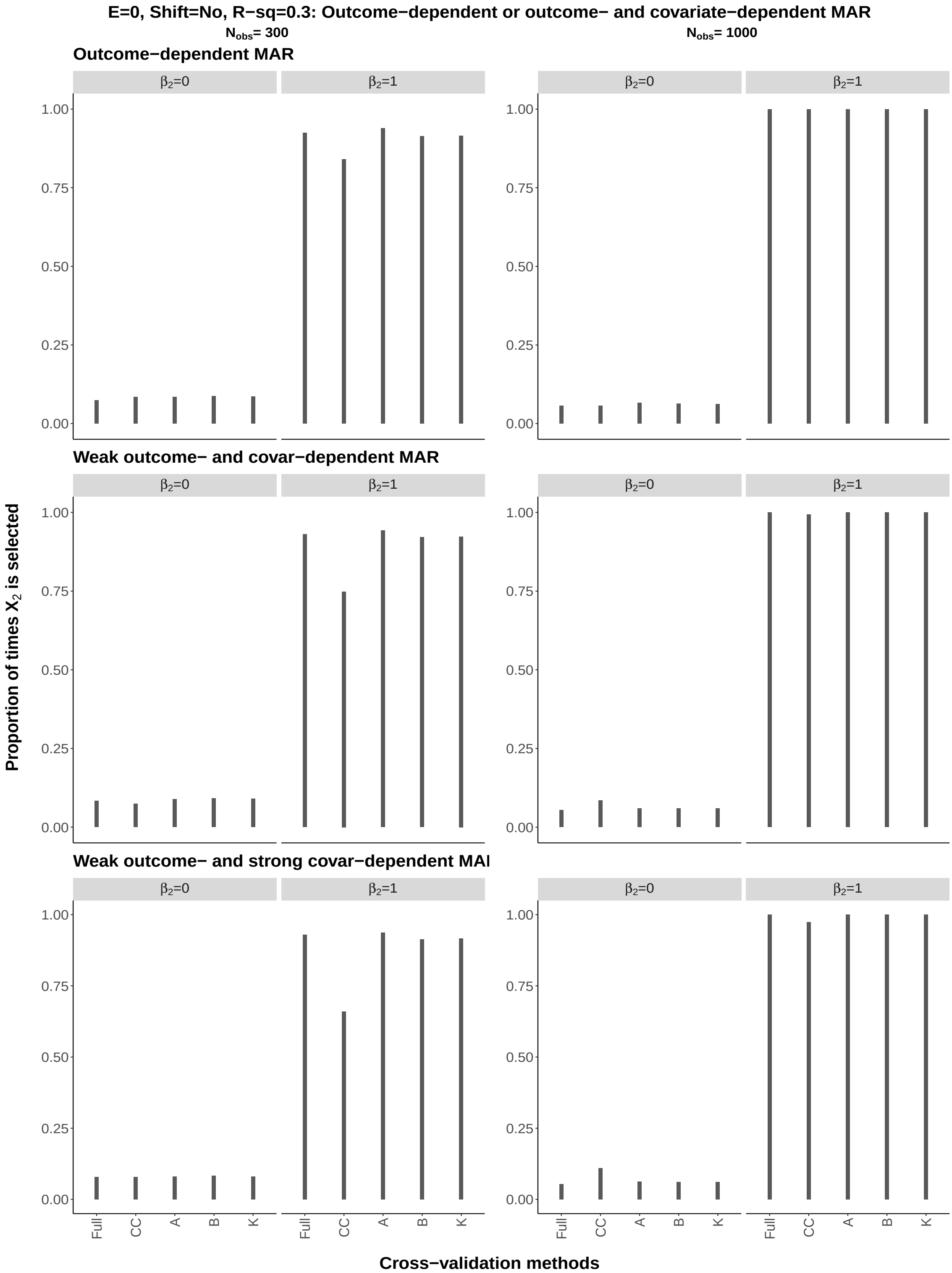

Figure S16: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

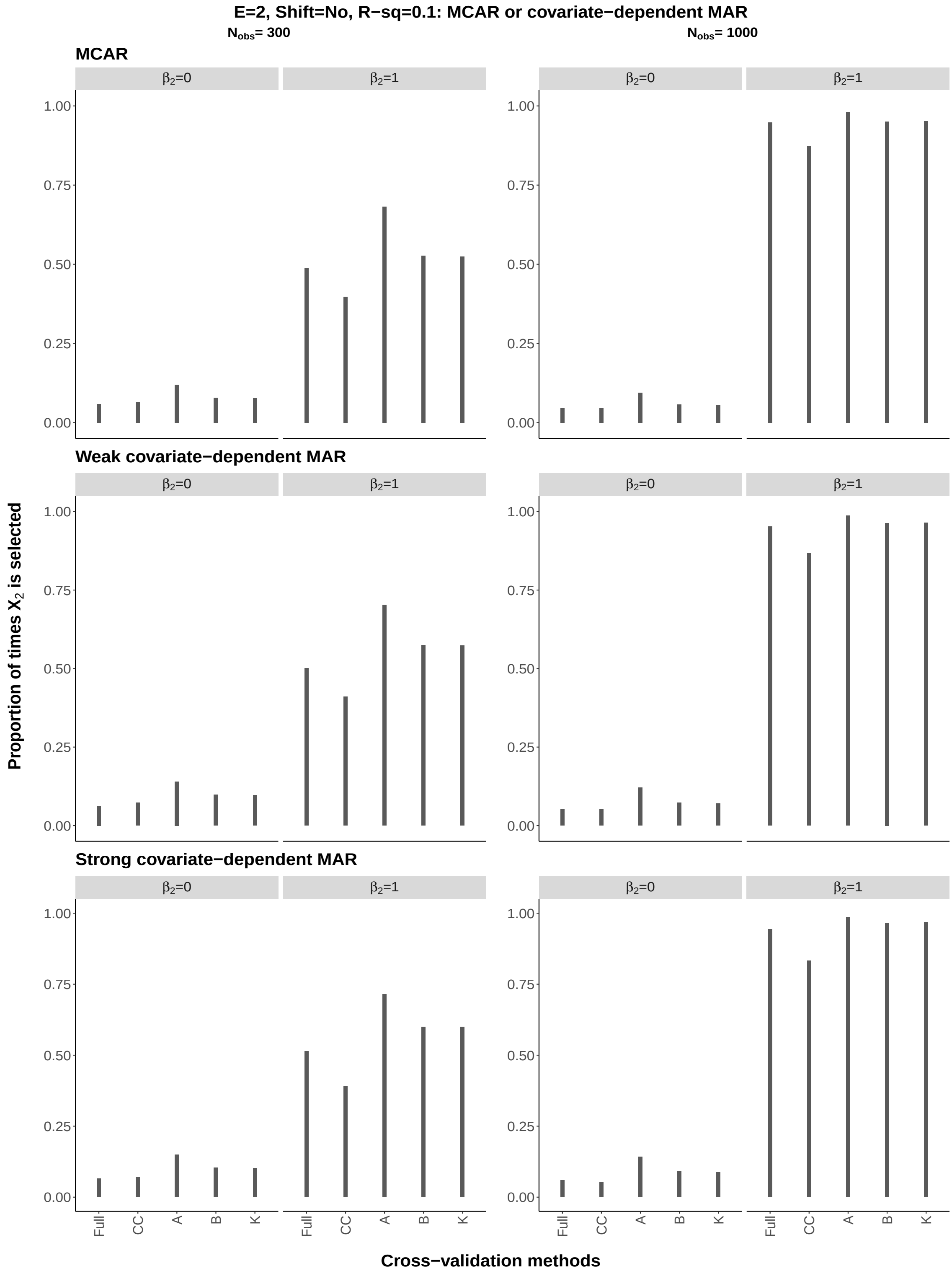

Figure S17: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

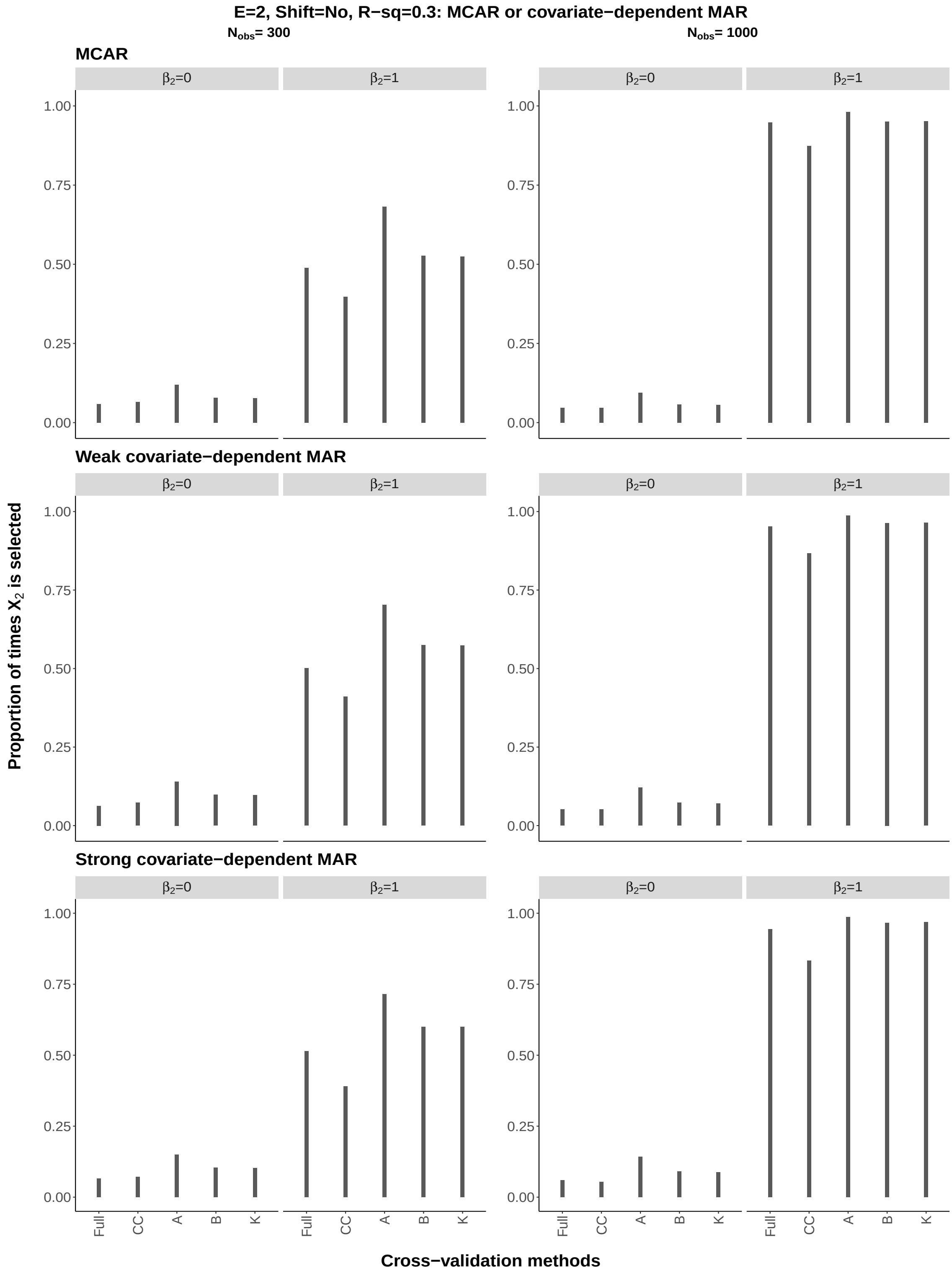

Figure S18: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

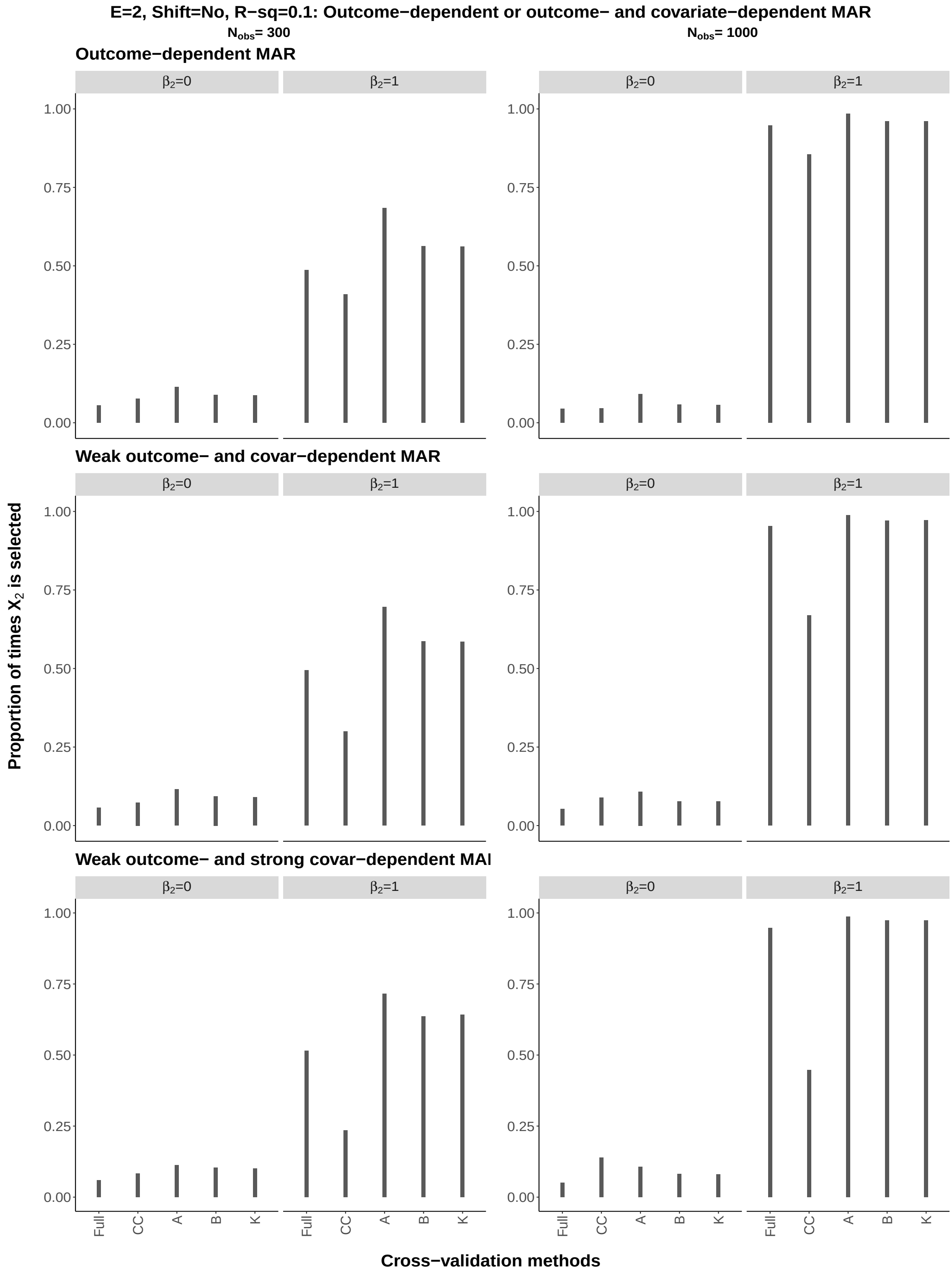

Figure S19: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

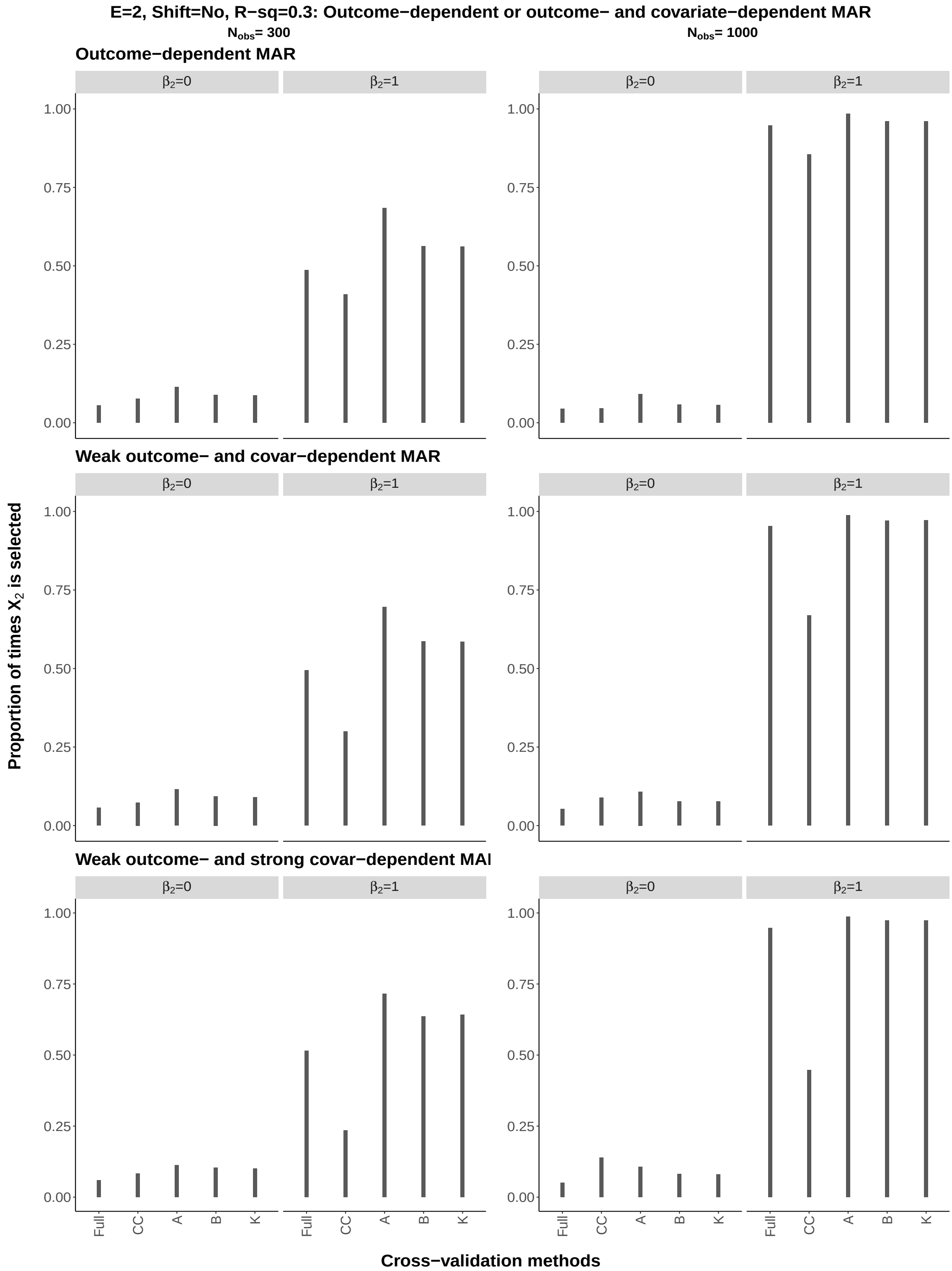

Figure S20: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

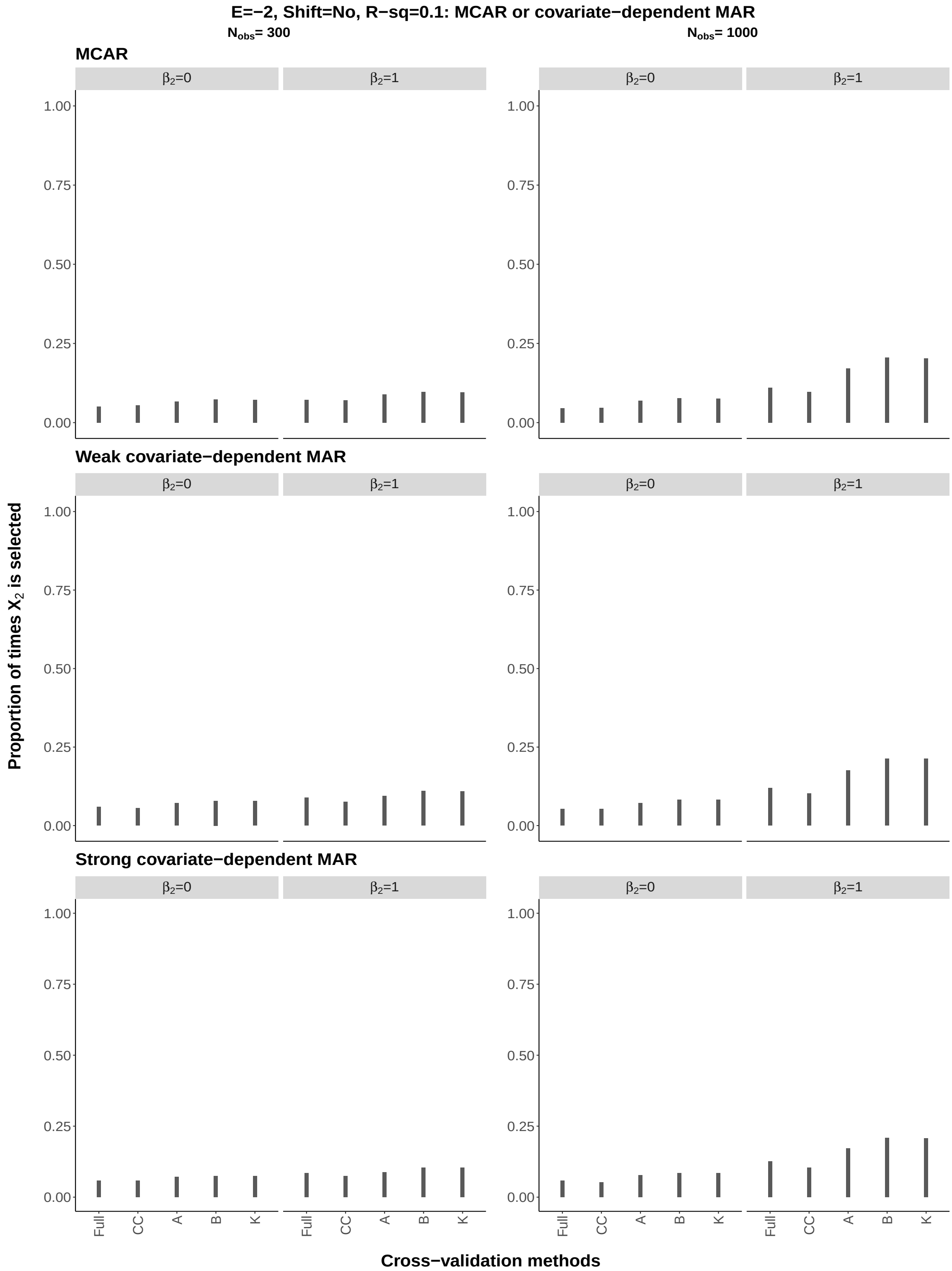

Figure S21: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

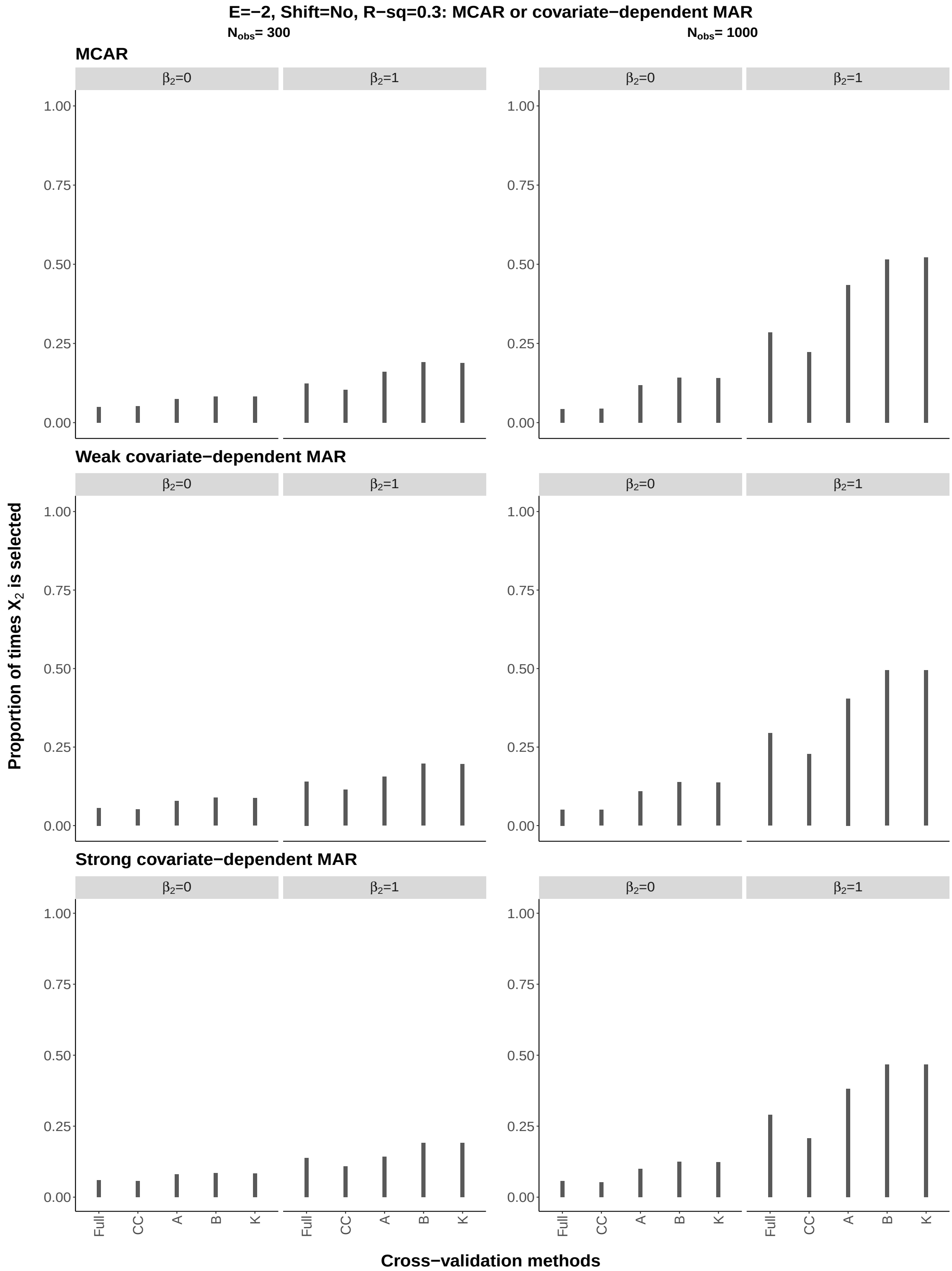

Figure S22: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

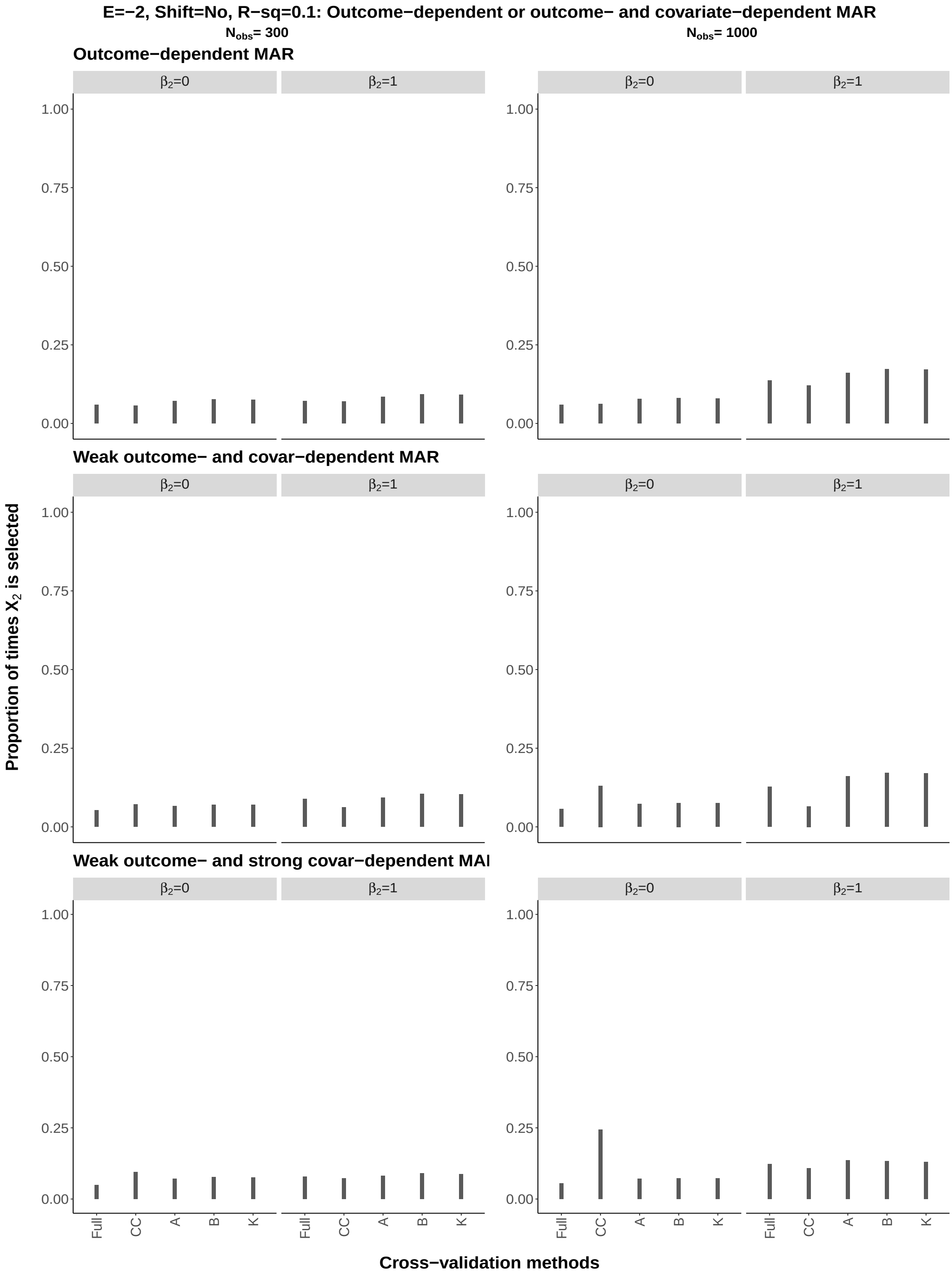

Figure S23: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

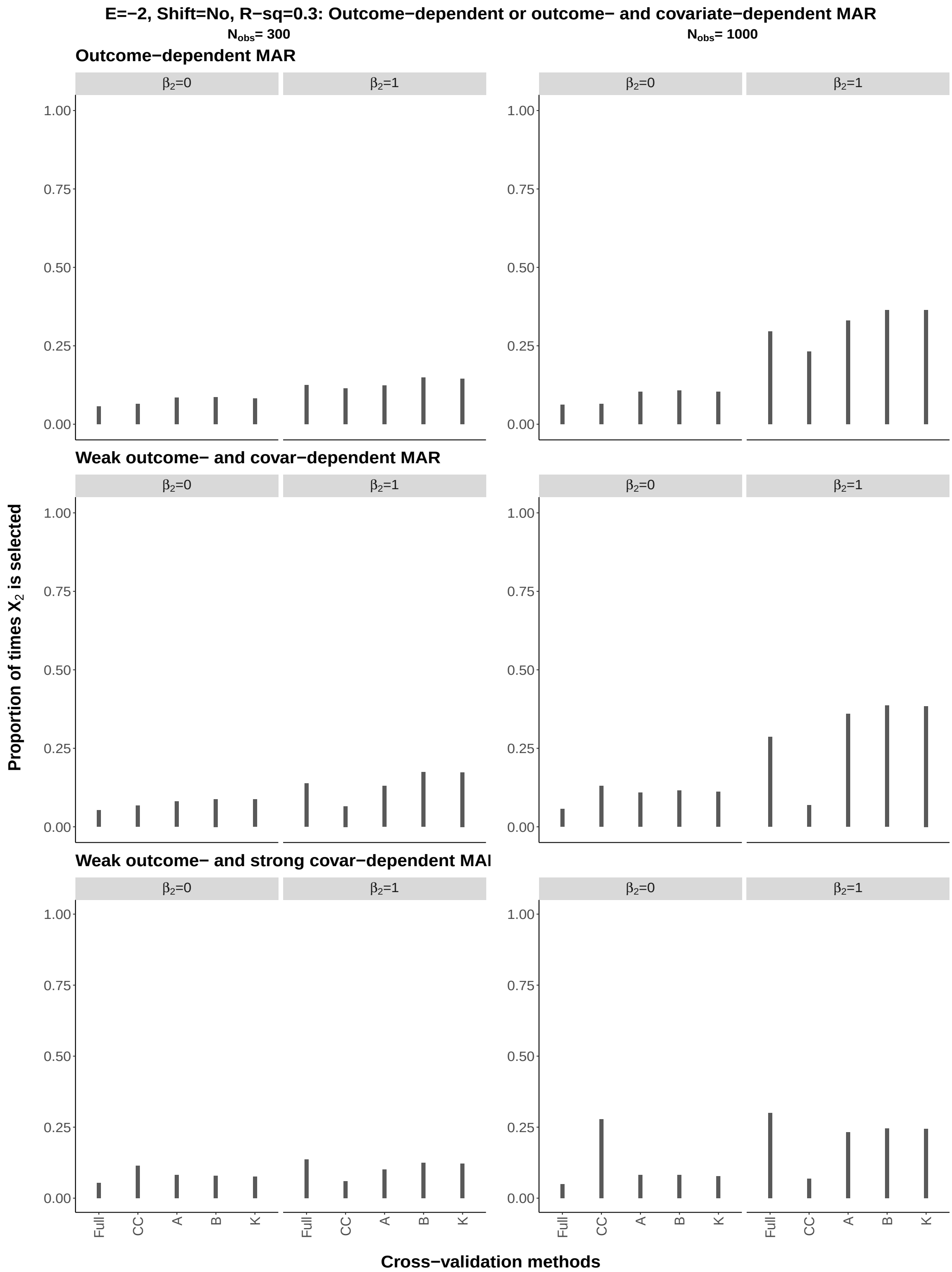

Figure S24: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.1.3 Covariate selection of X_2: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

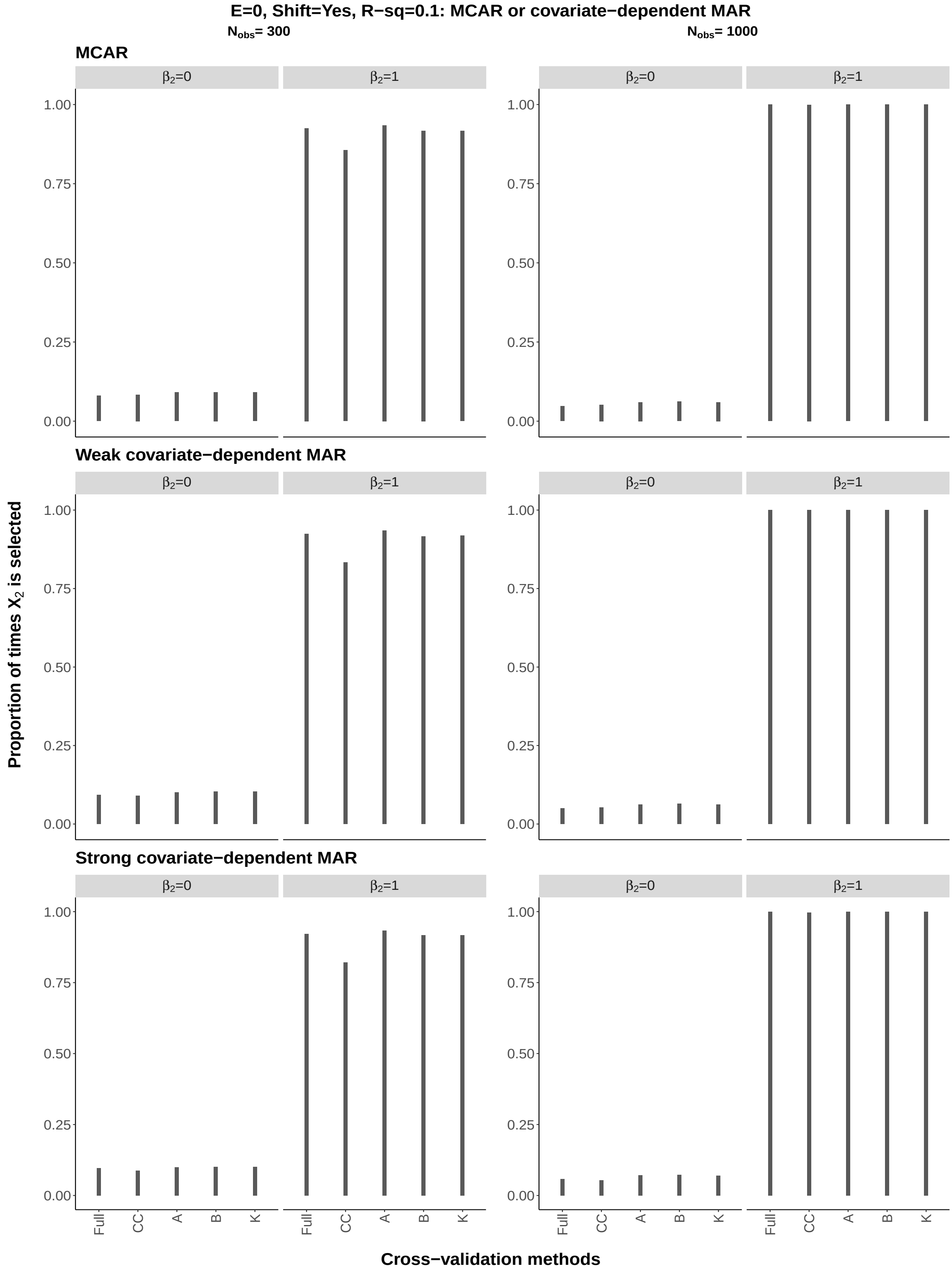

Figure S25: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

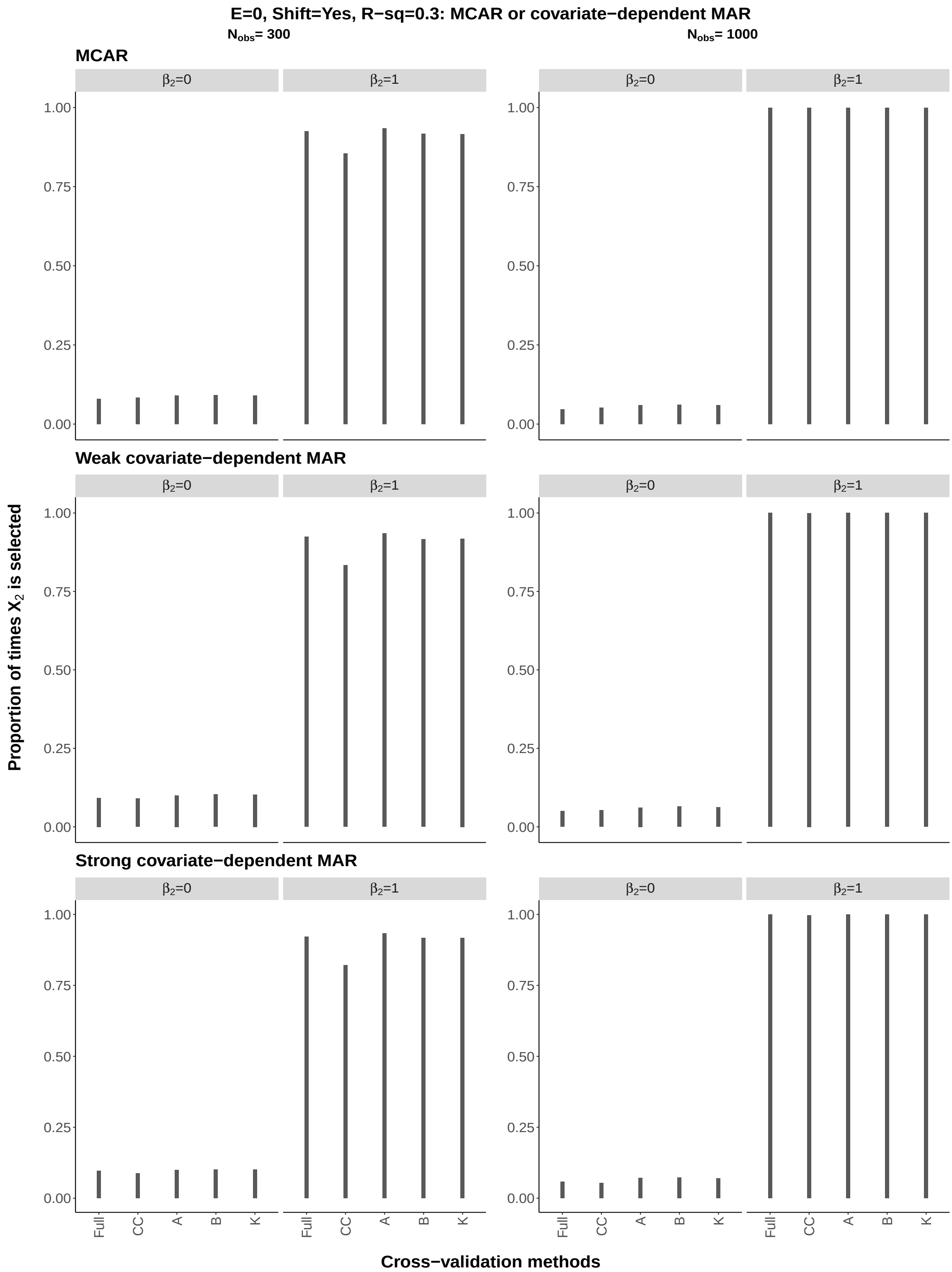

Figure S26: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

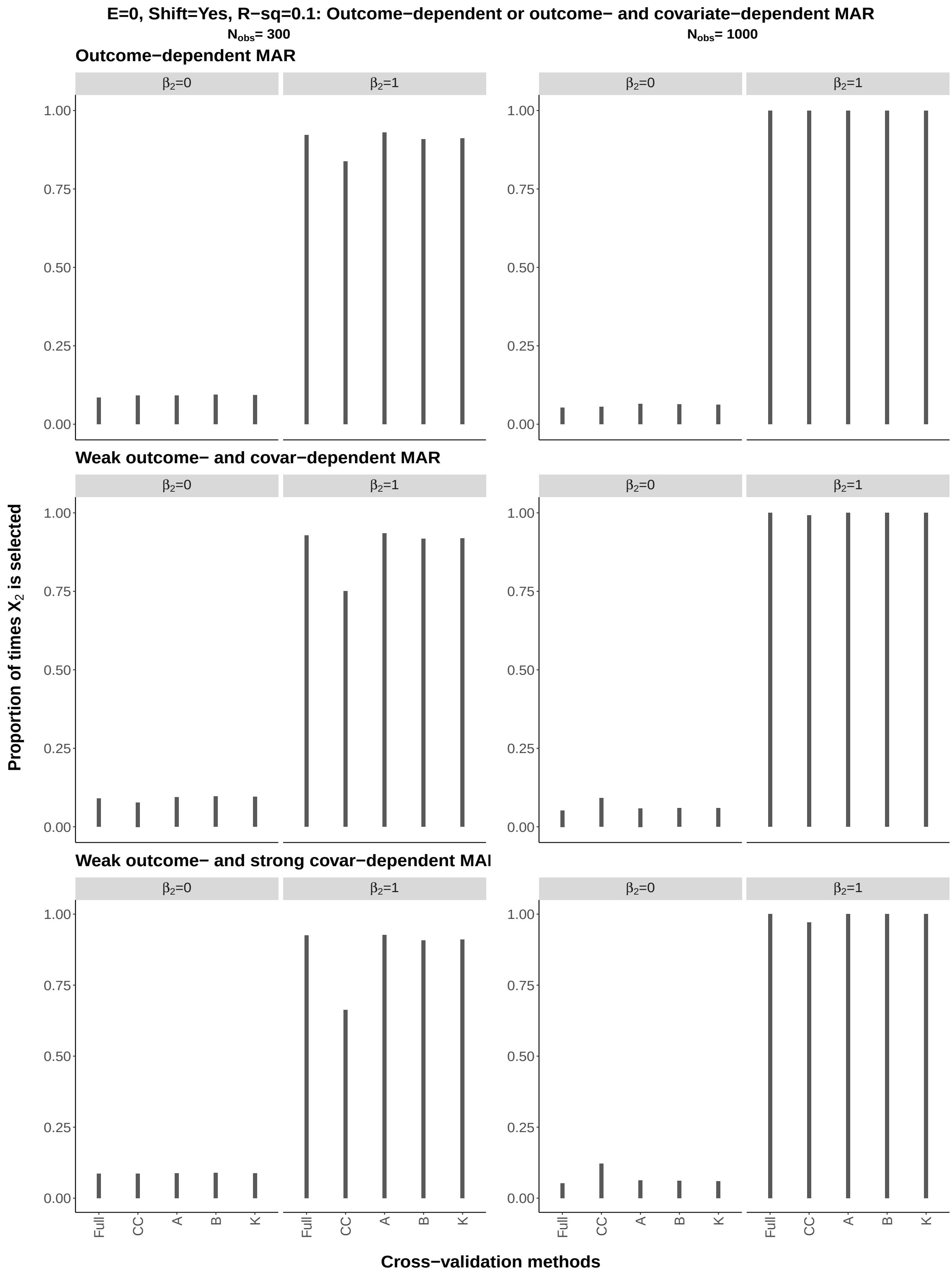

Figure S27: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

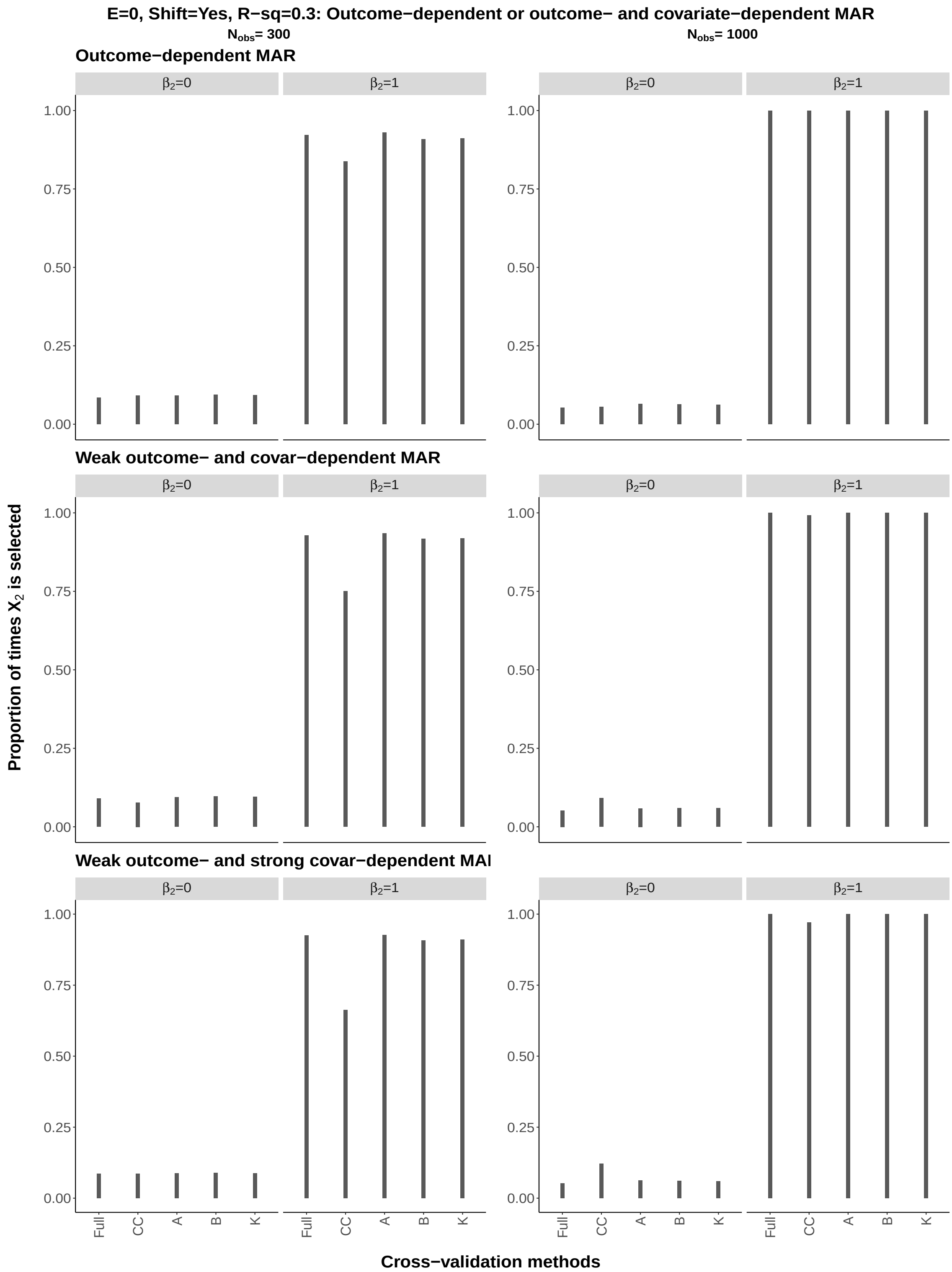

Figure S28: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

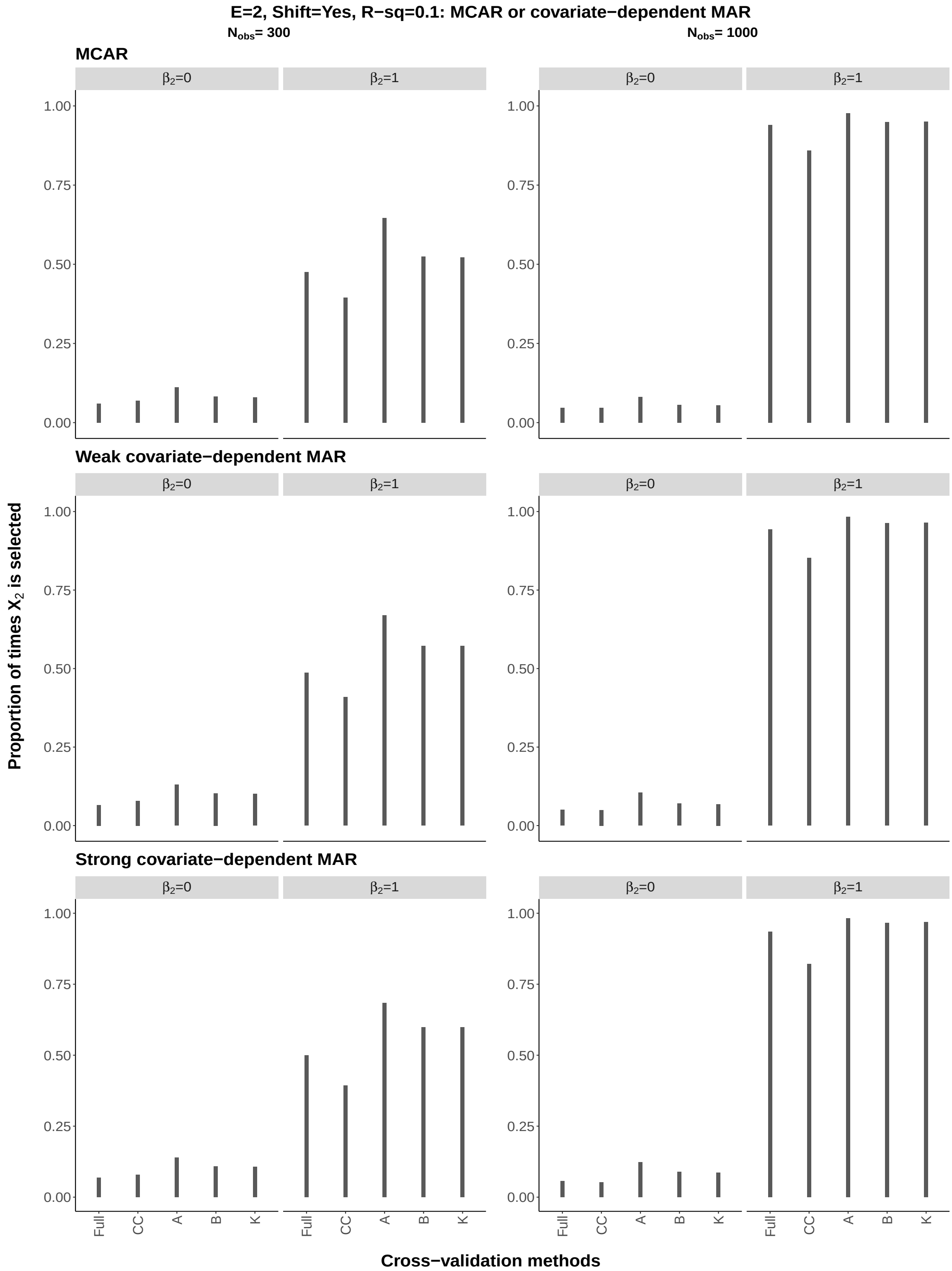

Figure S29: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

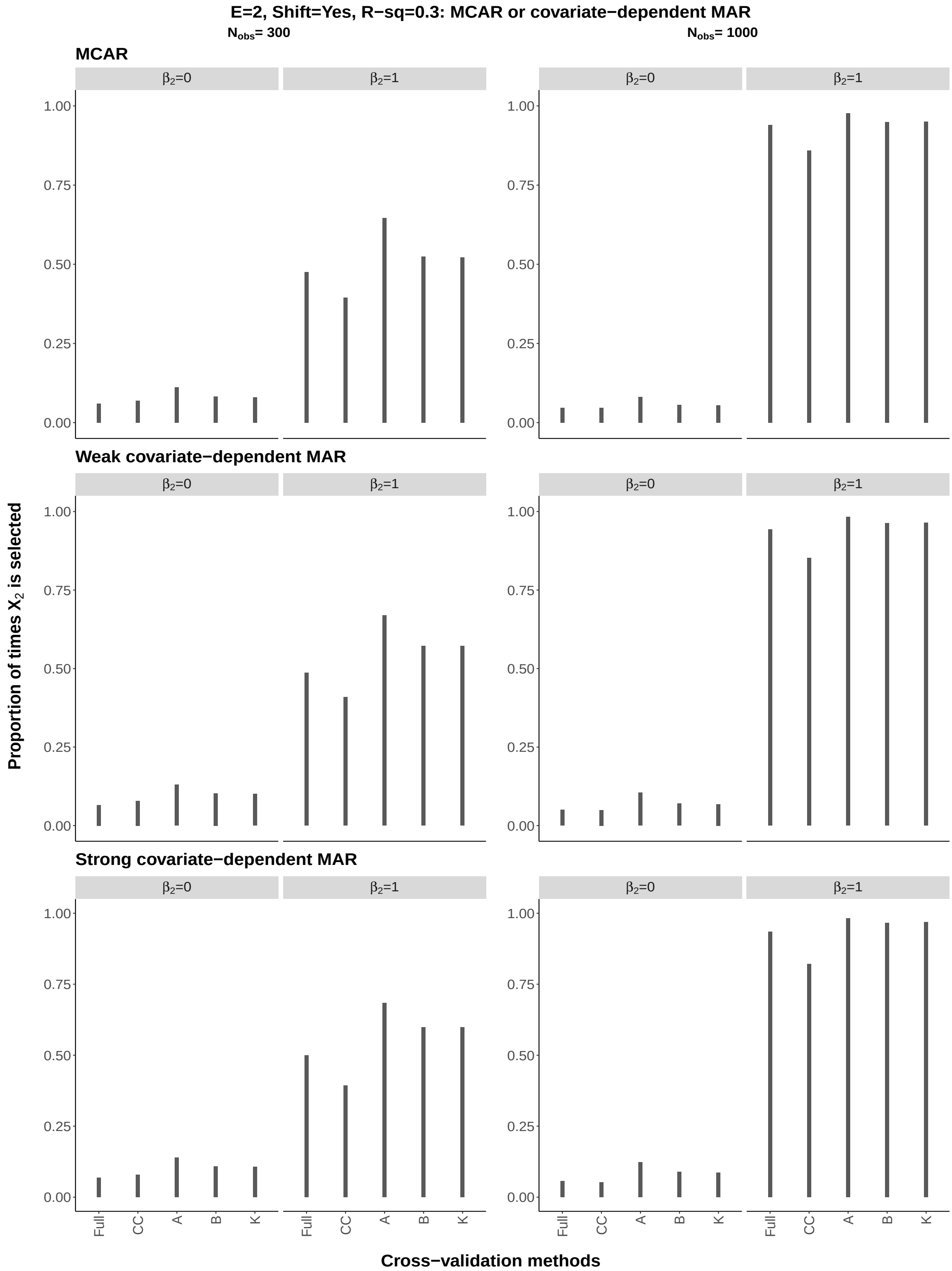

Figure S30: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

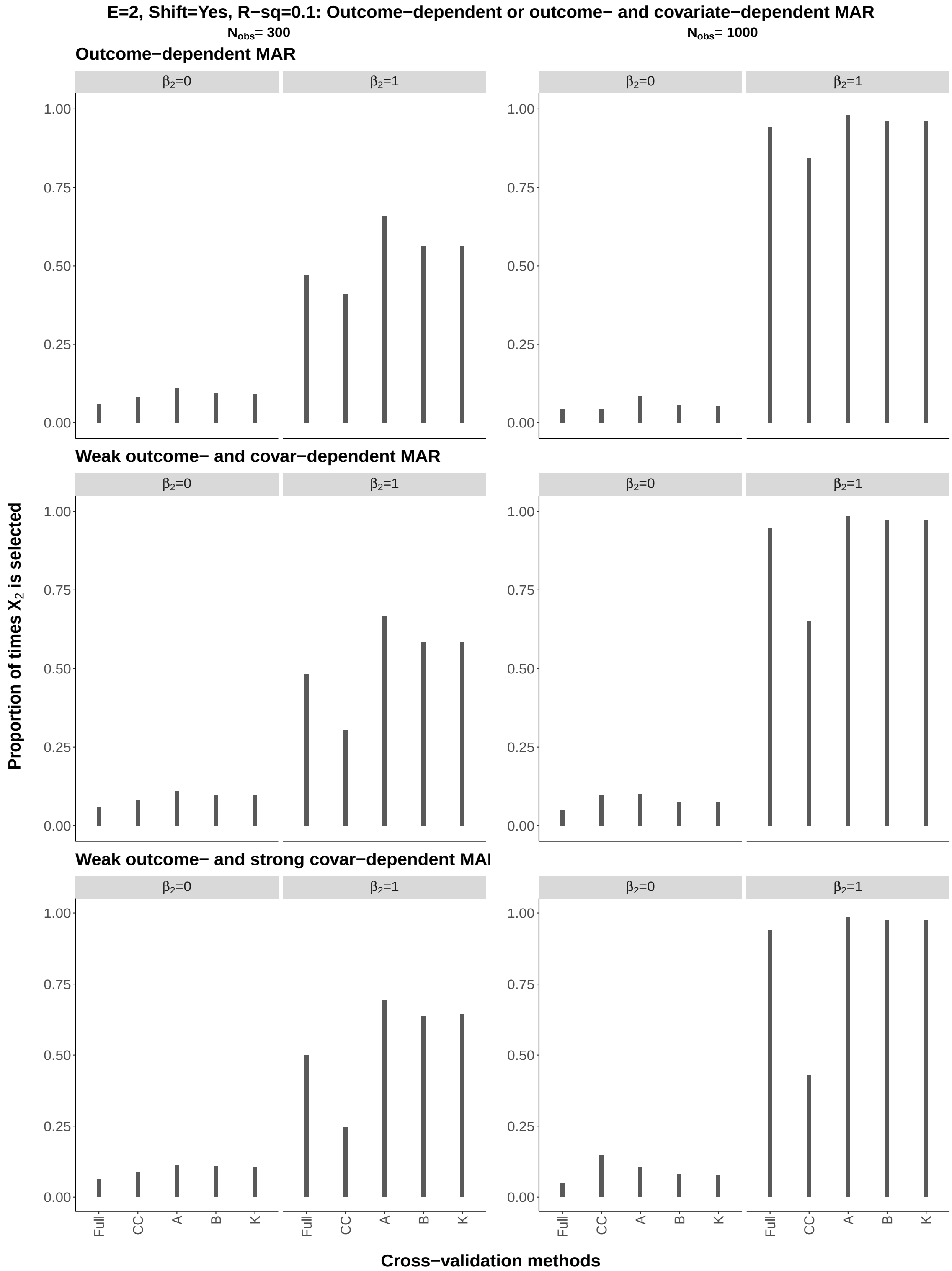

Figure S31: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

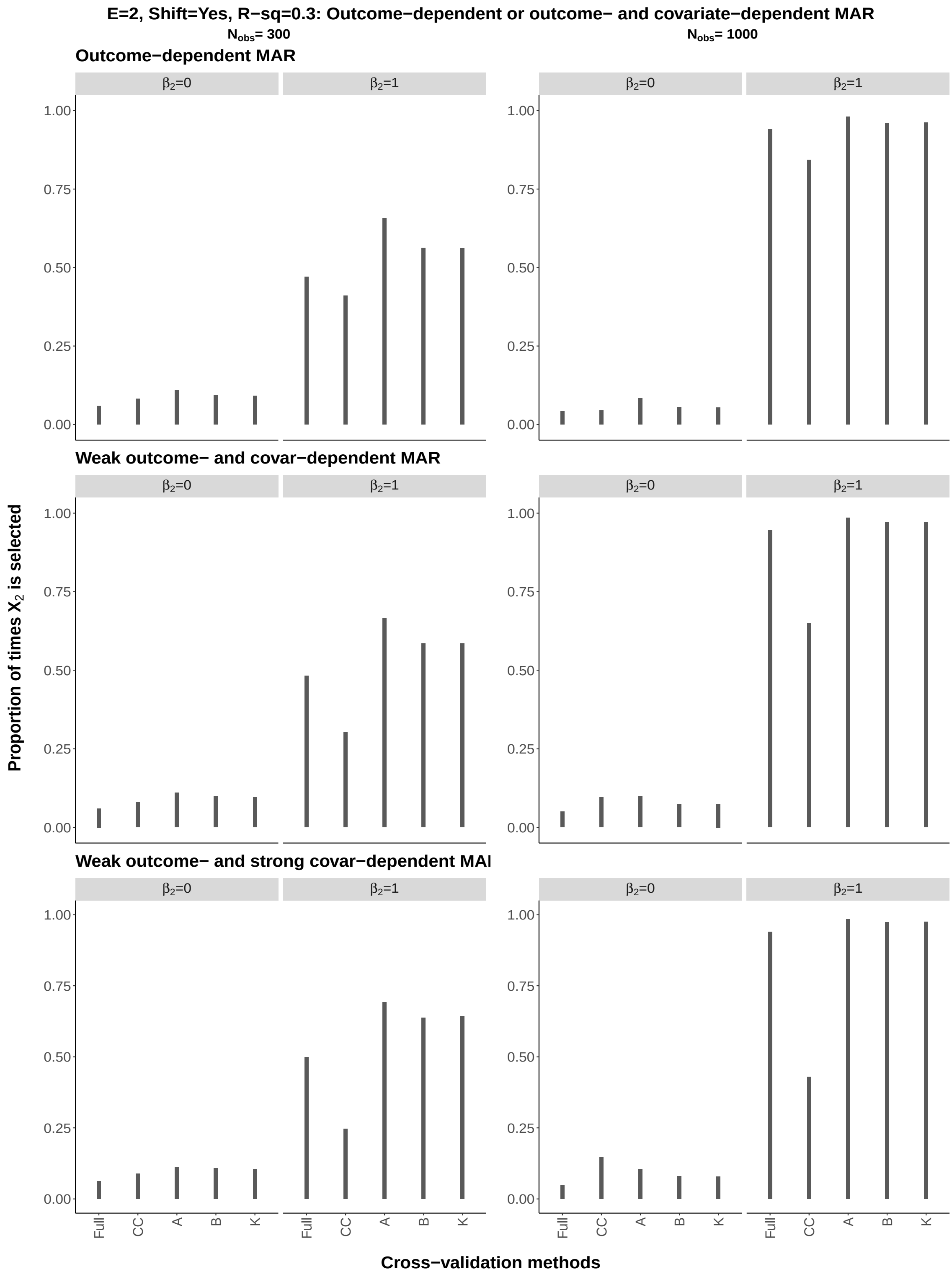

Figure S32: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

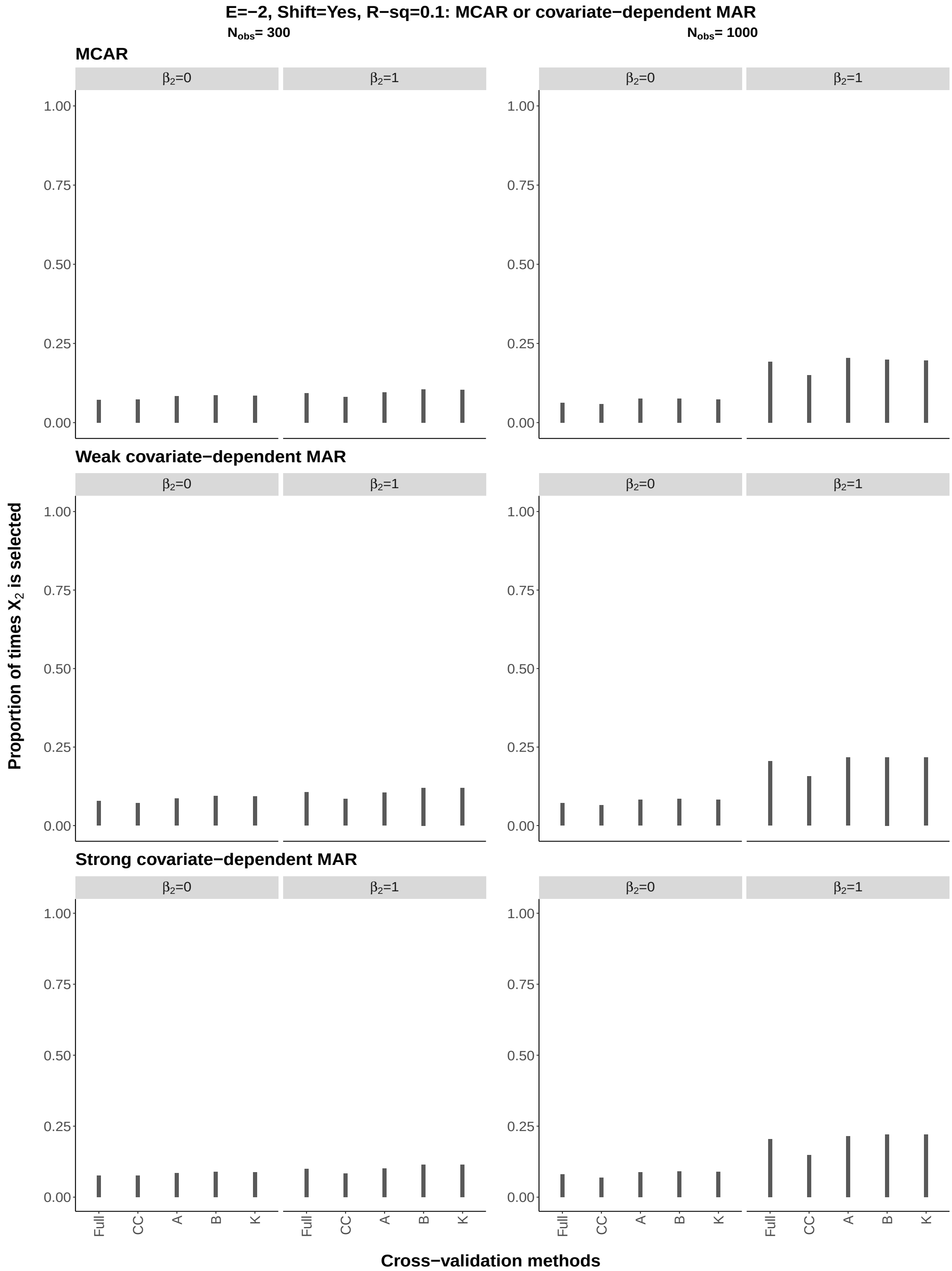

Figure S33: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

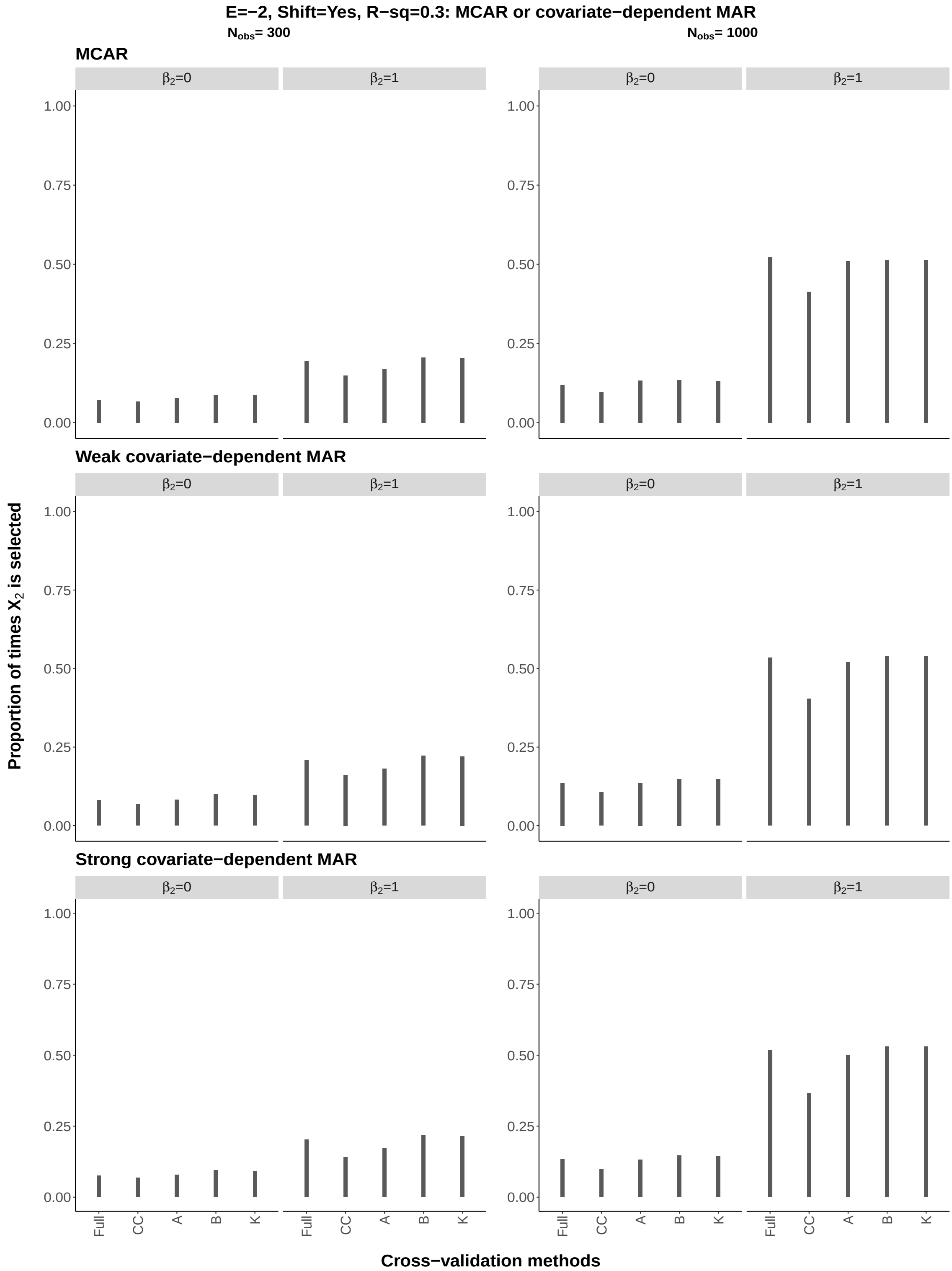

Figure S34: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

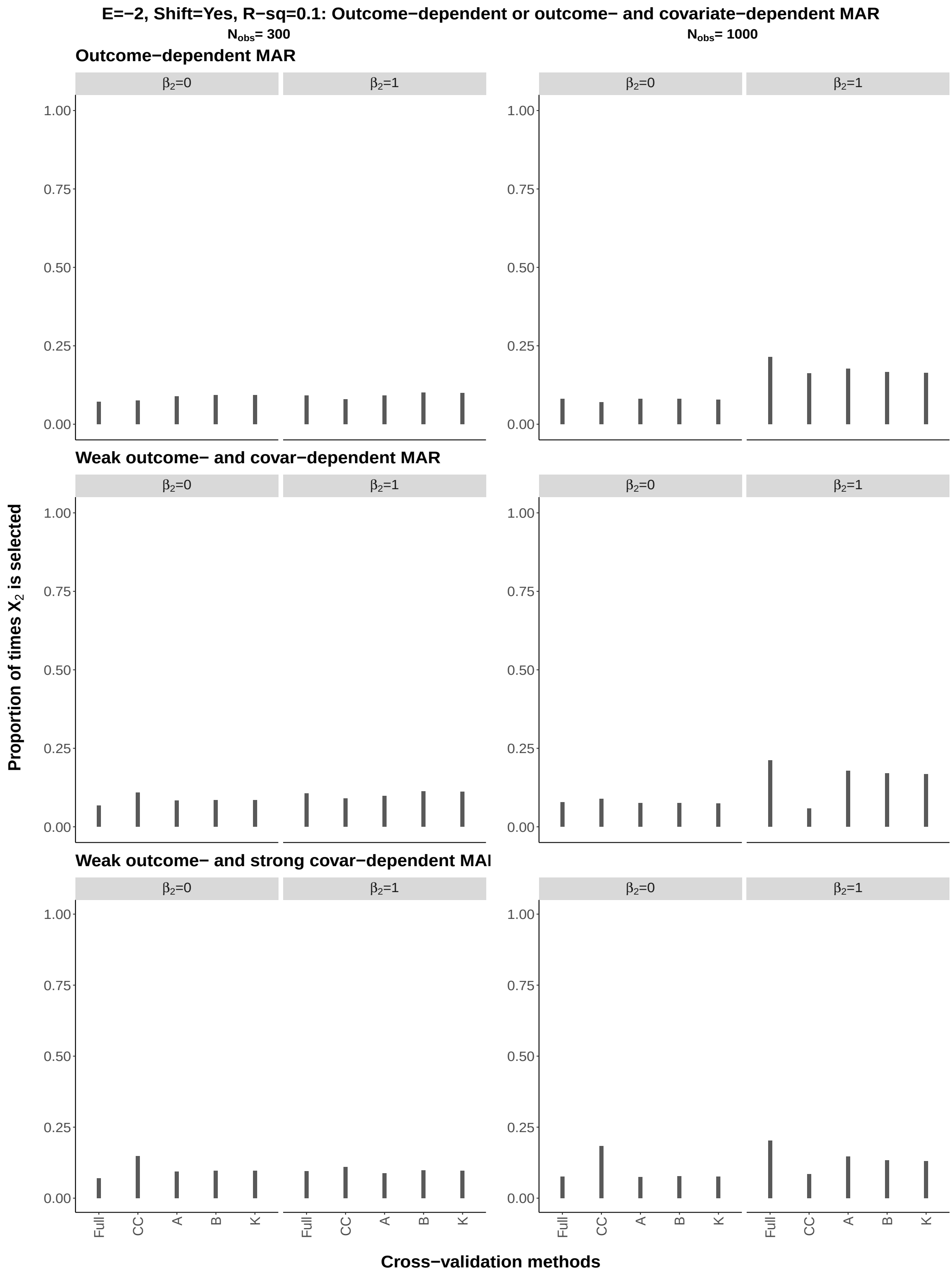

Figure S35: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

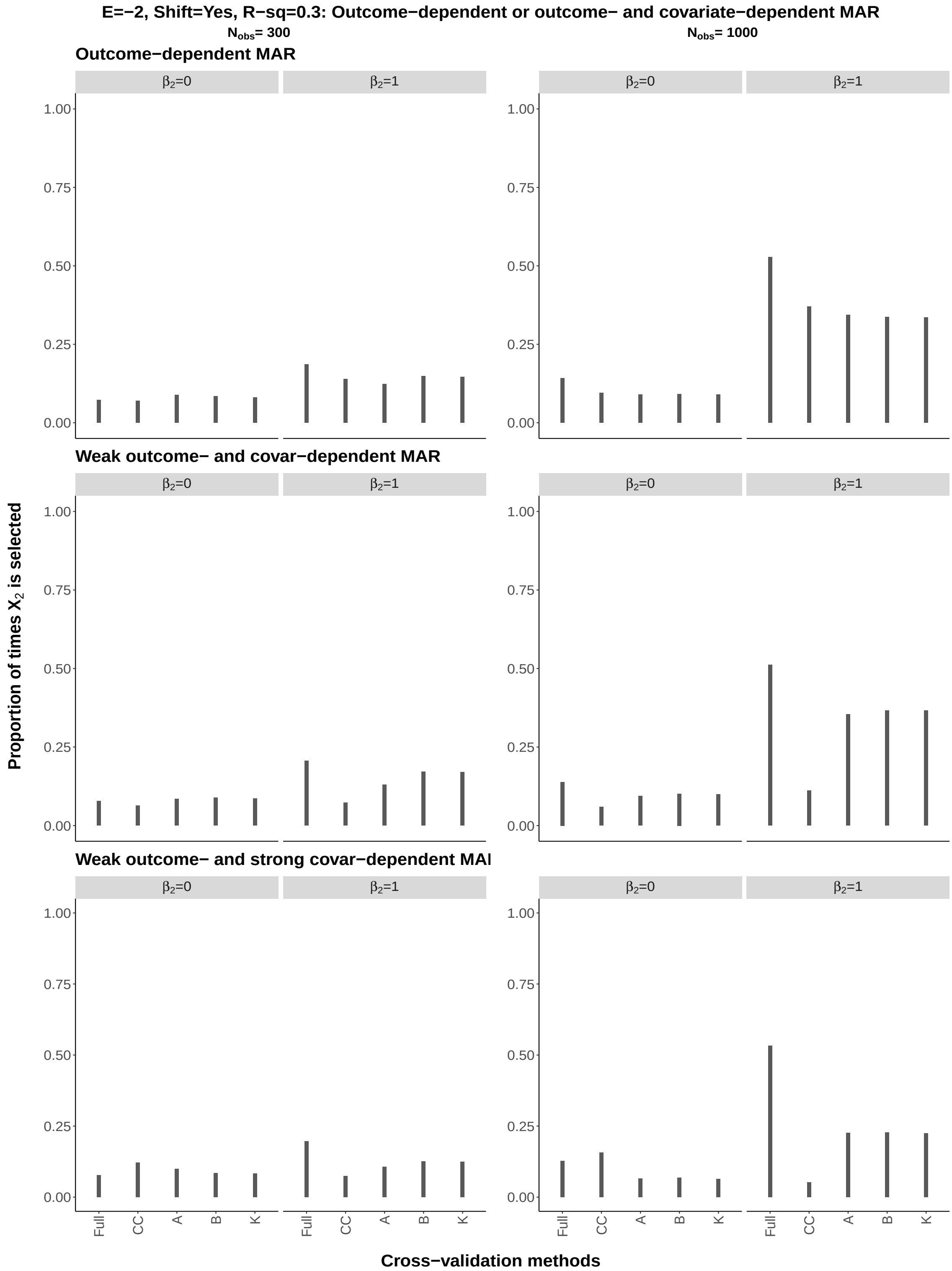

Figure S36: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.1.4 Covariate selection of X_2: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

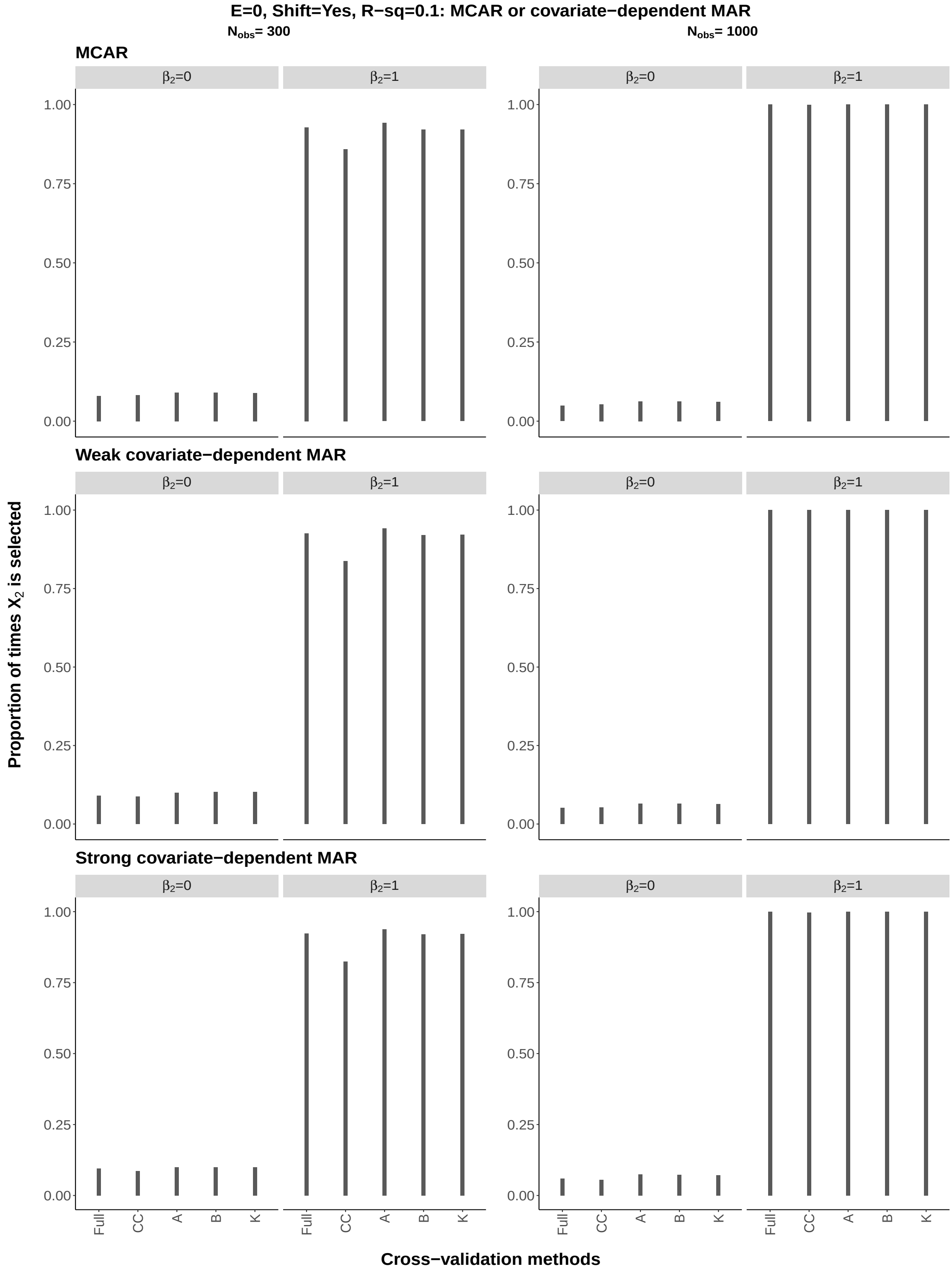

Figure S37: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

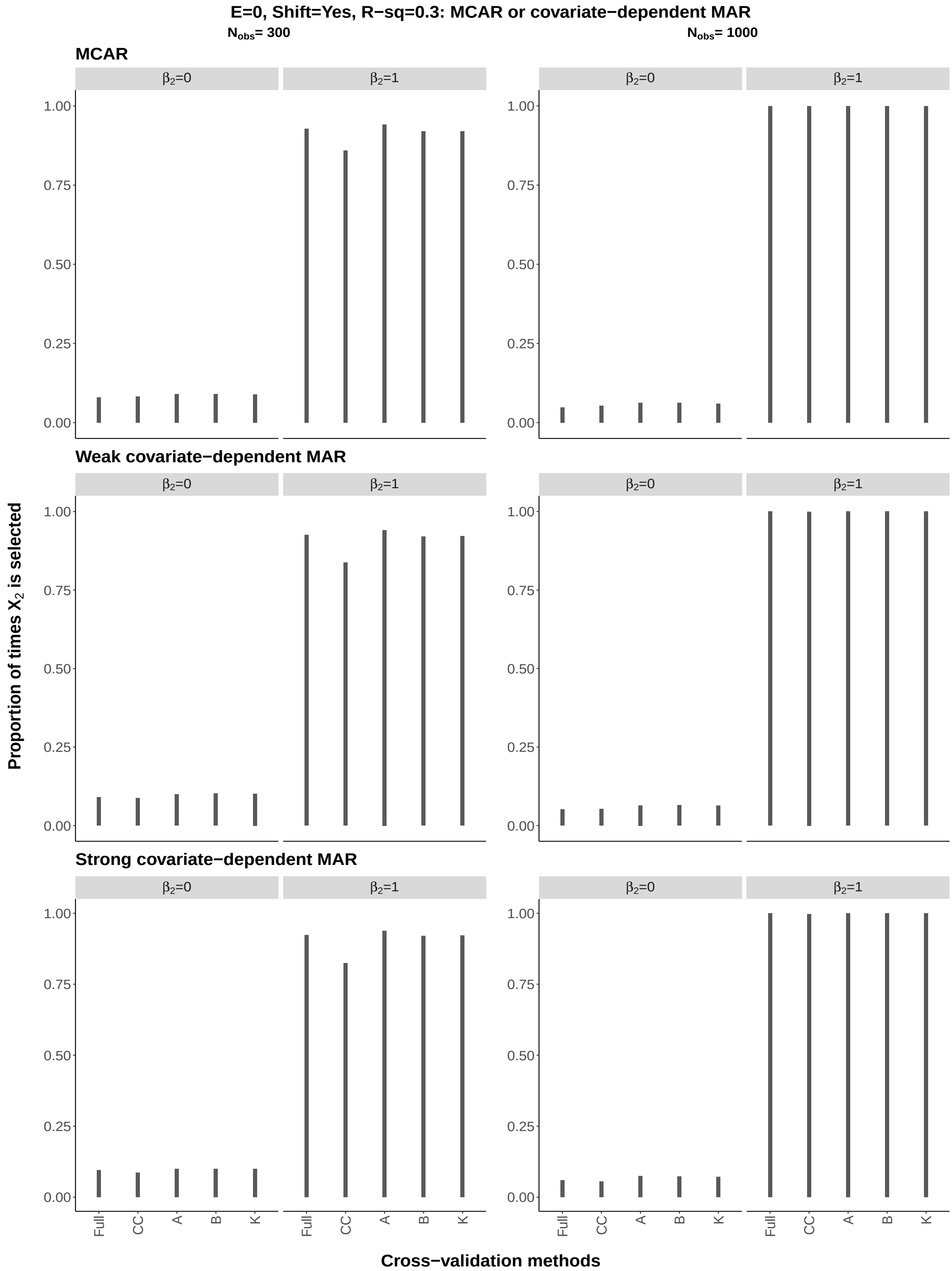

Figure S38: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

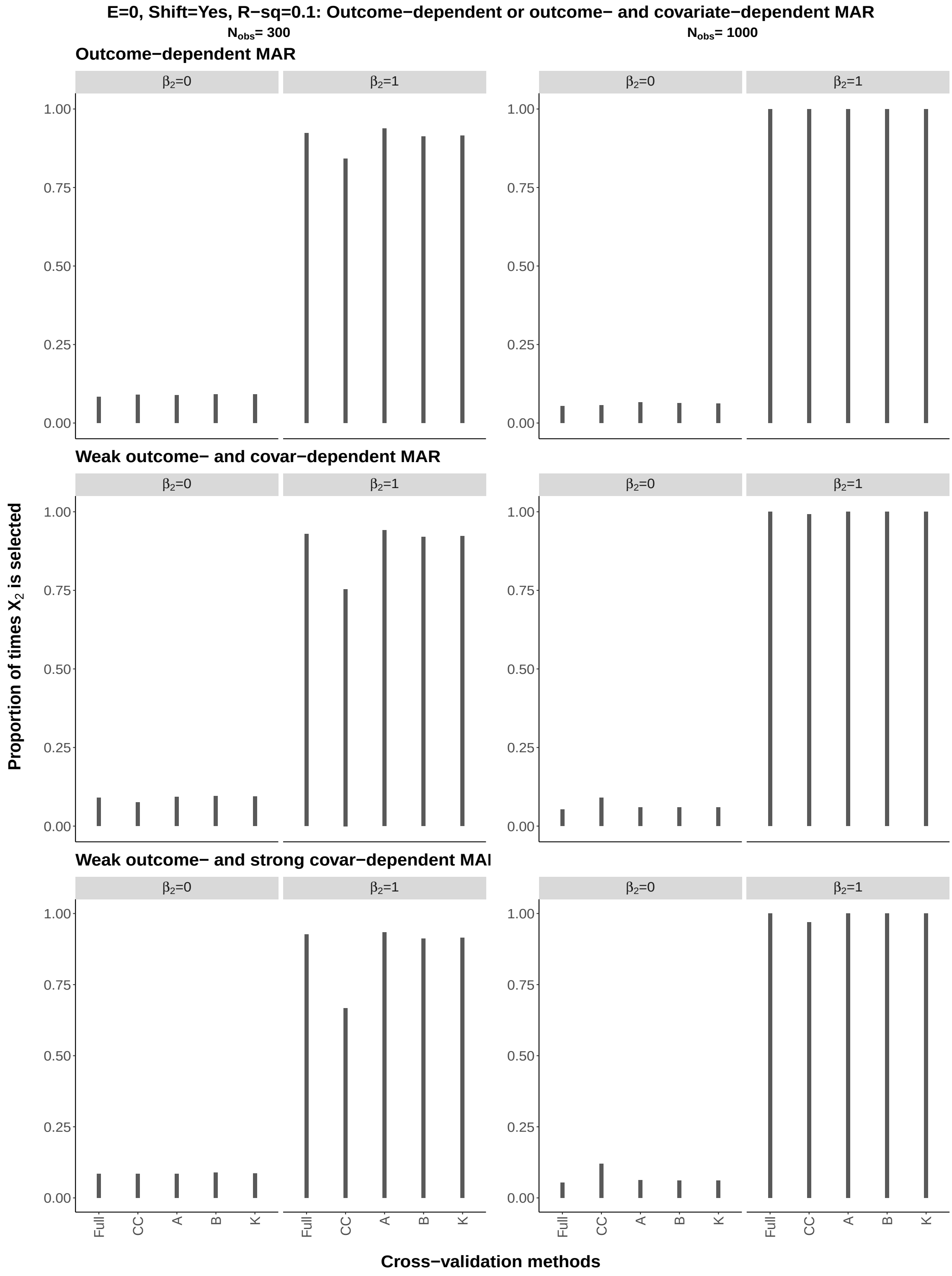

Figure S39: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

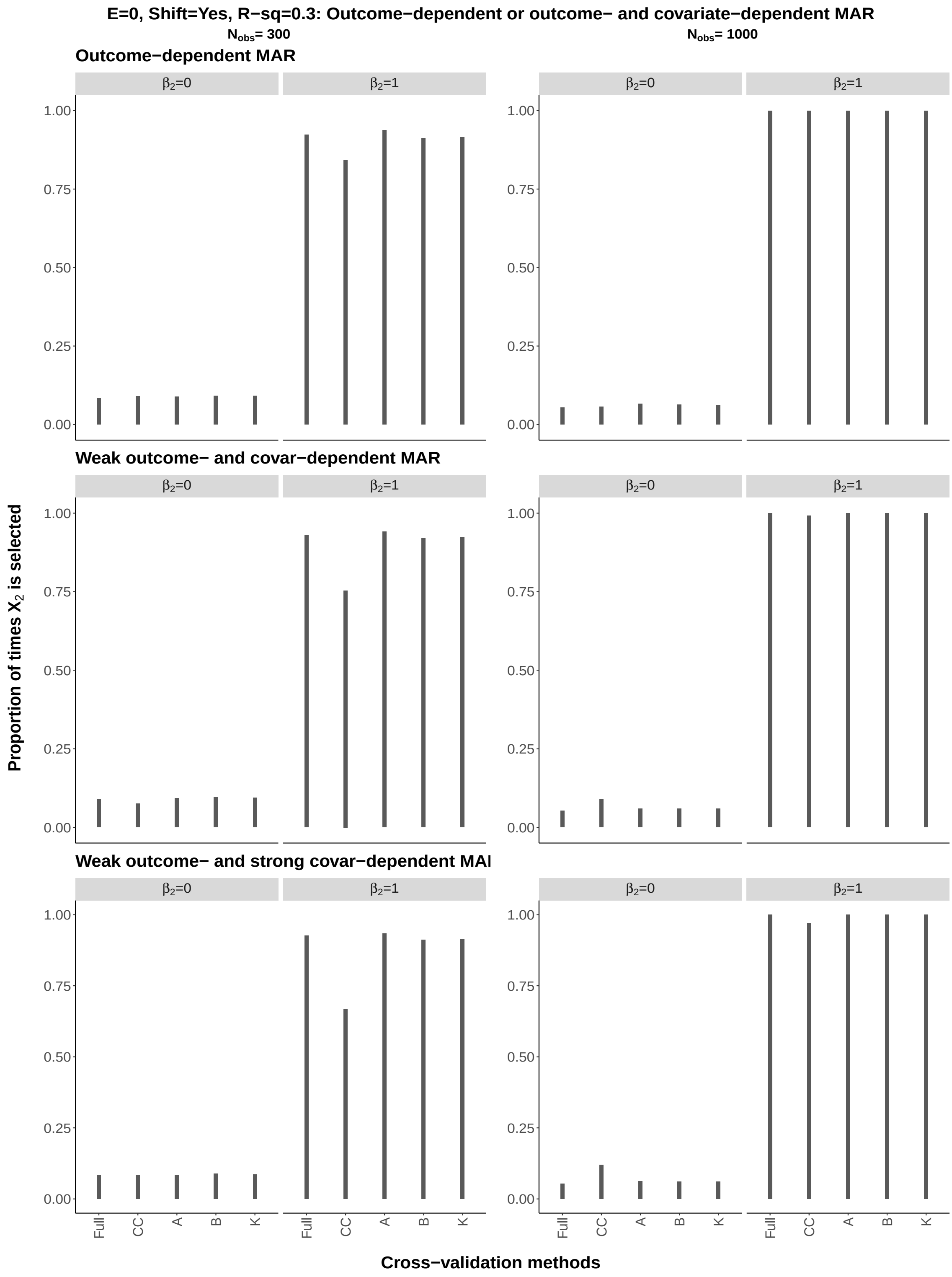

Figure S40: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

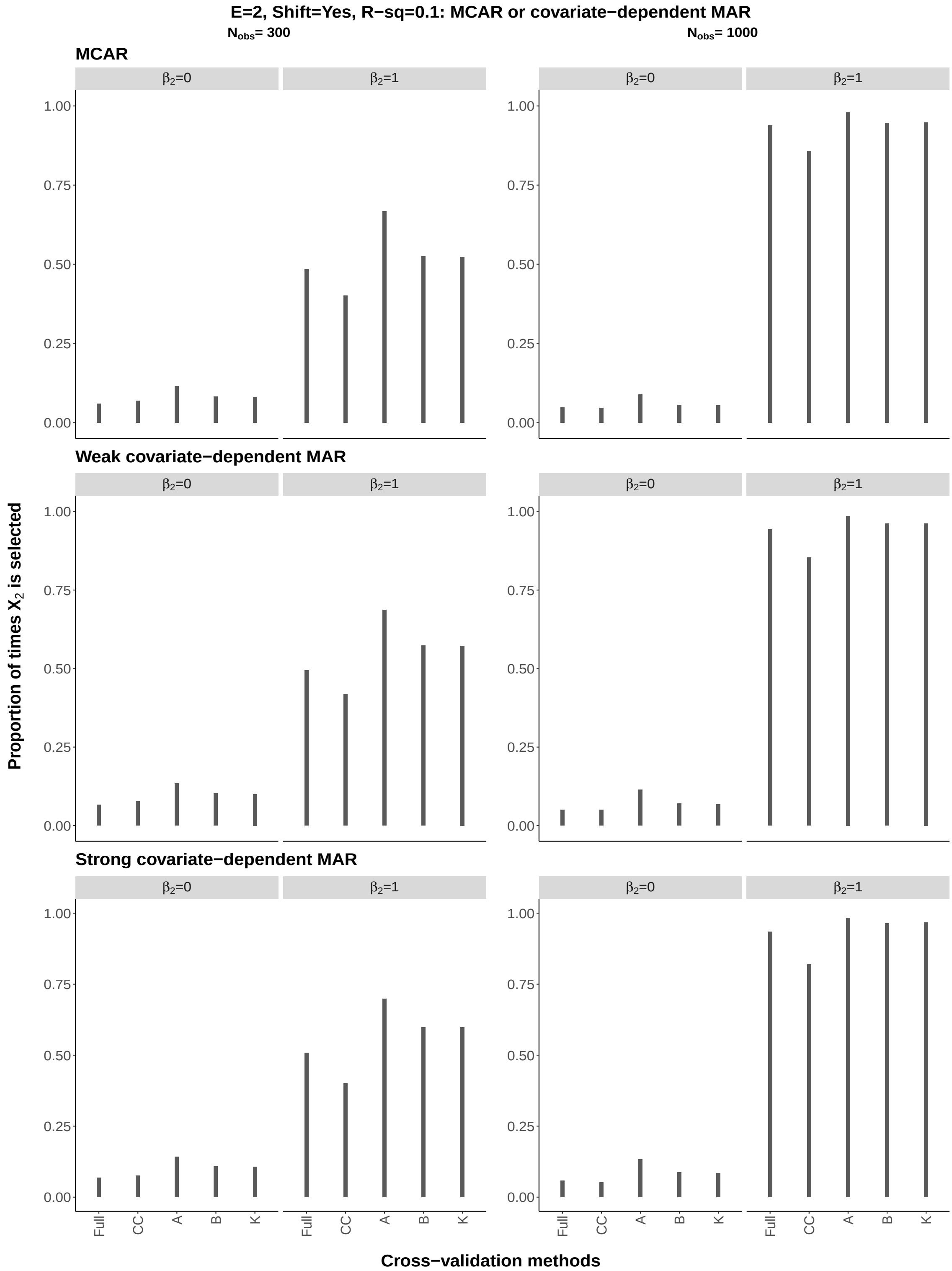

Figure S41: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

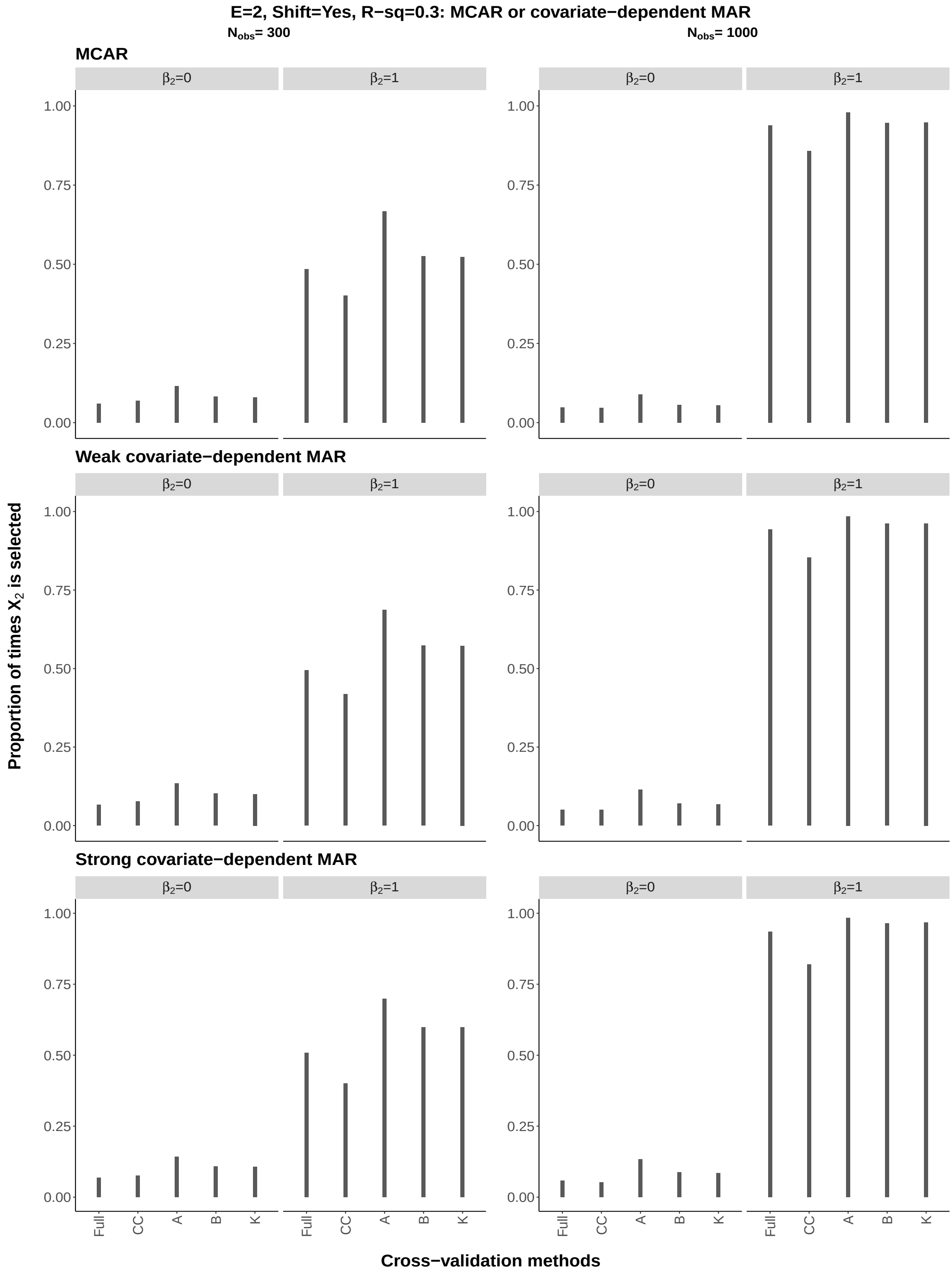

Figure S42: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

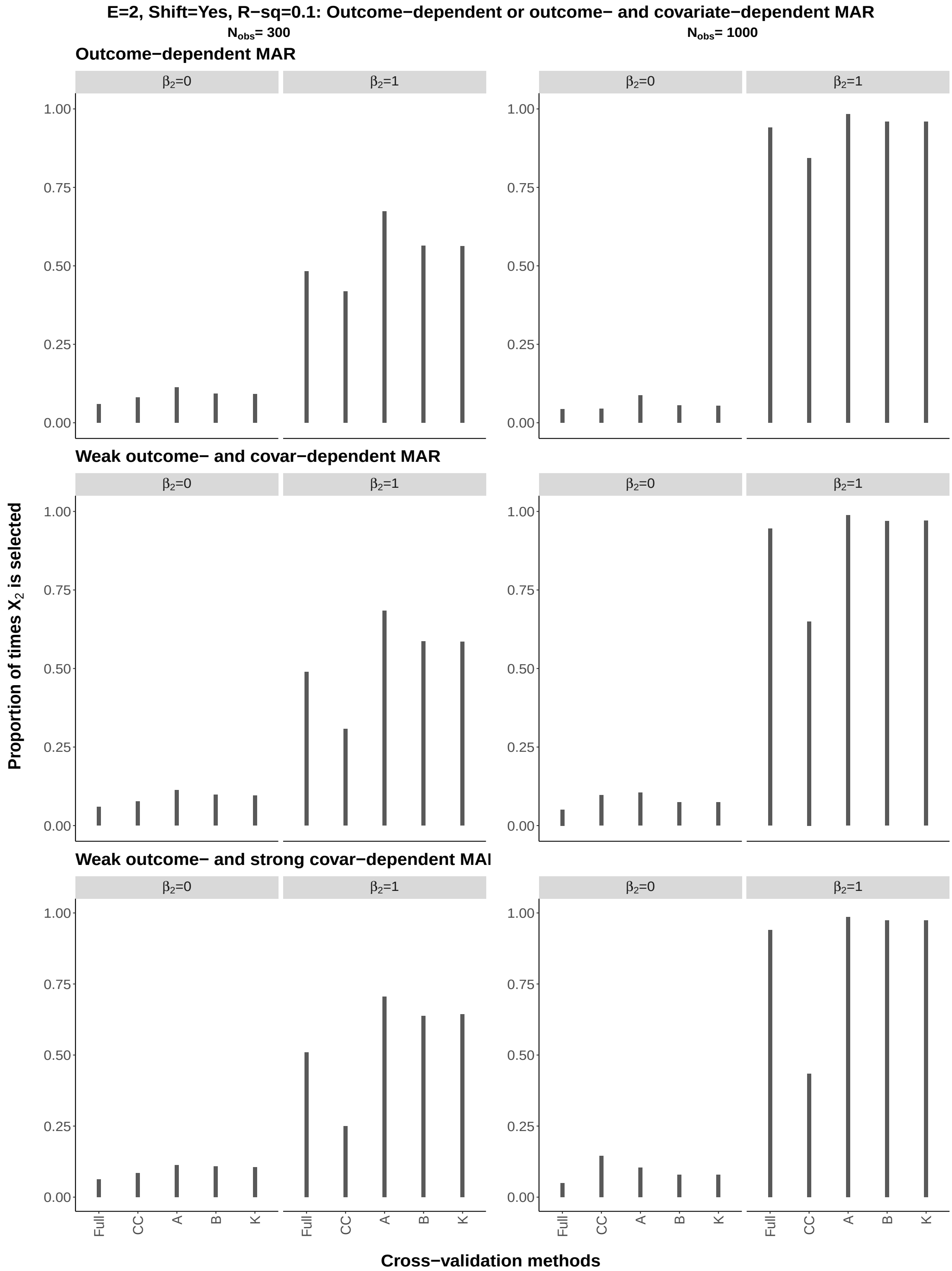

Figure S43: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

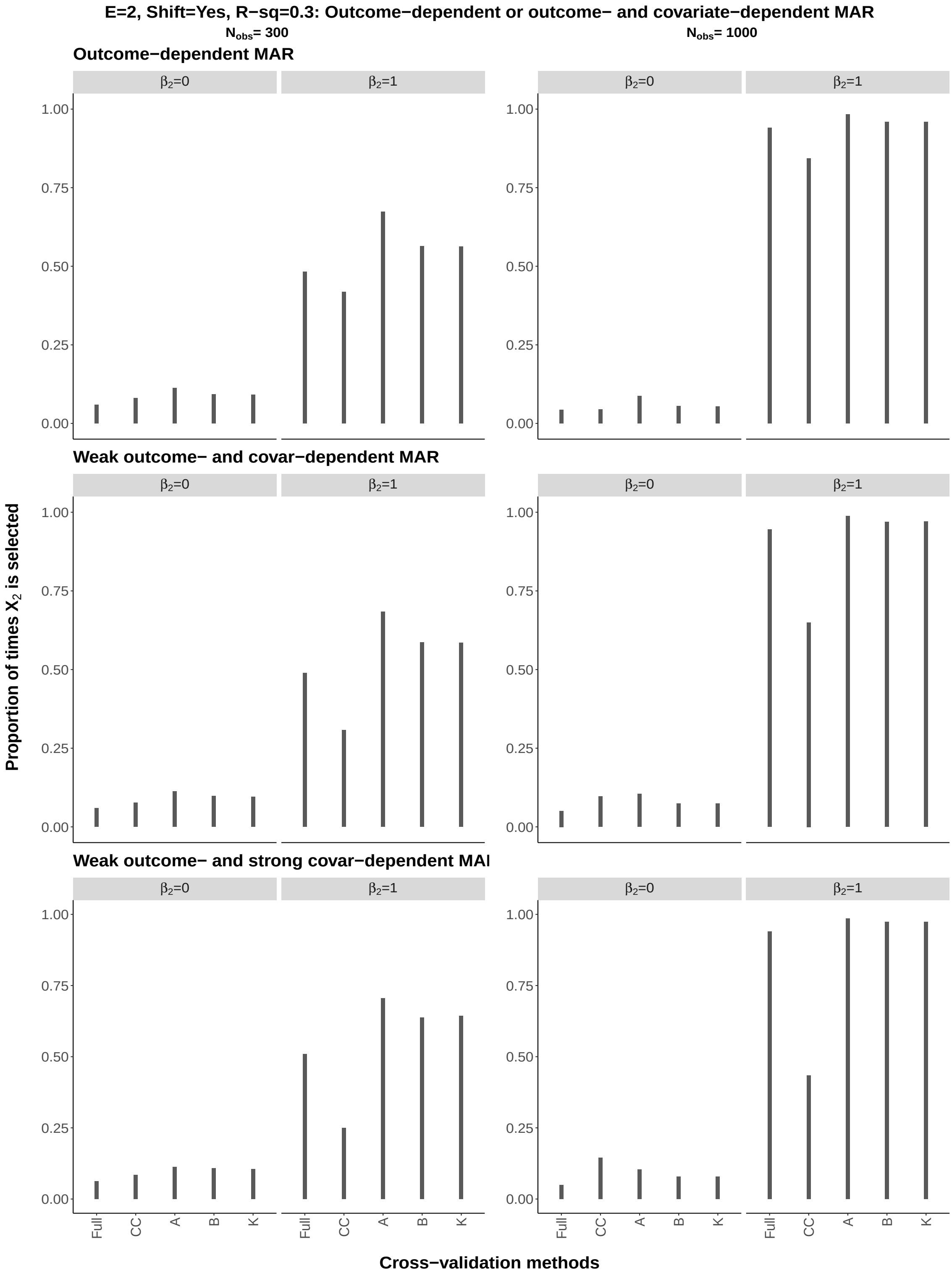

Figure S44: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

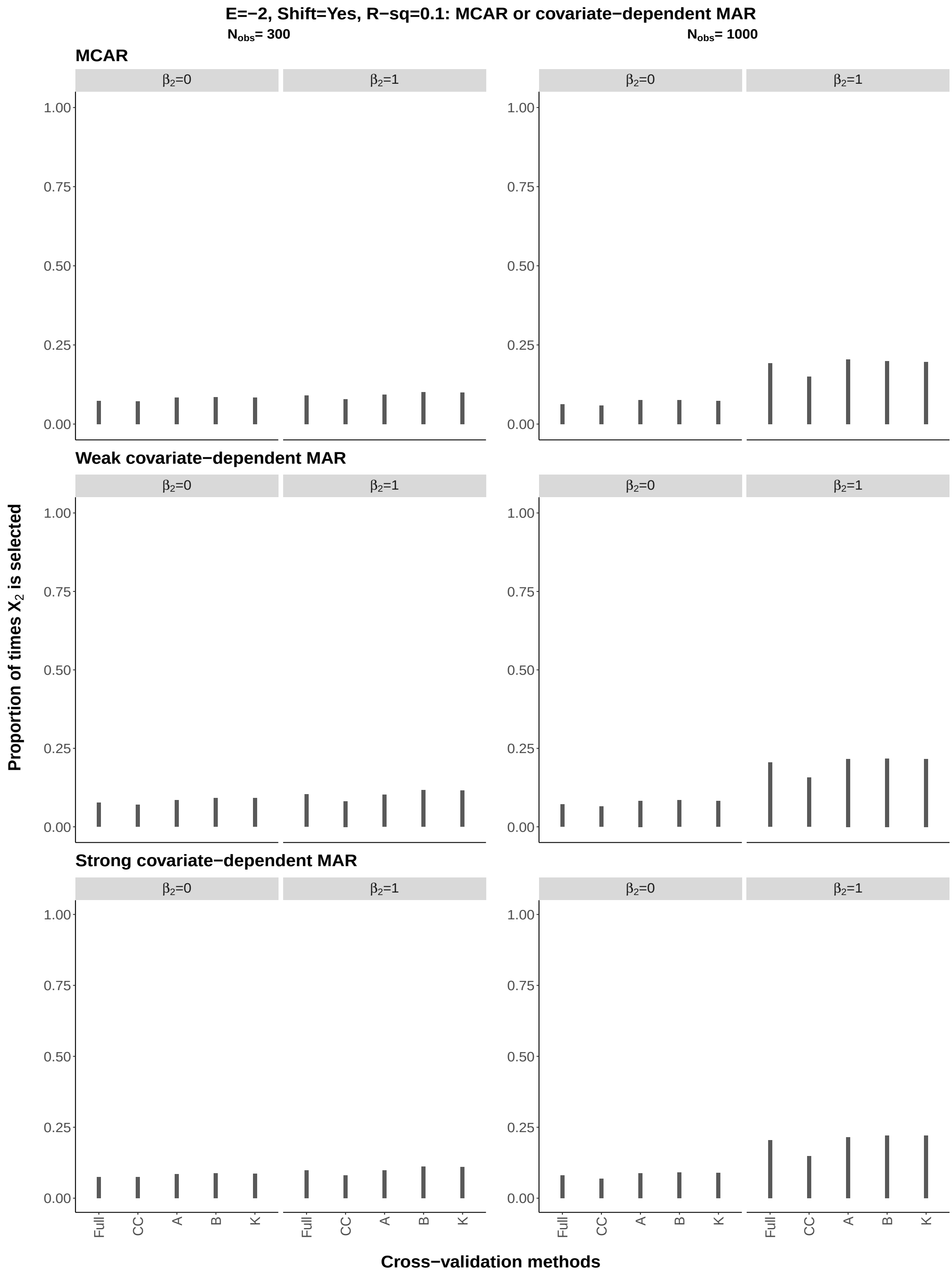

Figure S45: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

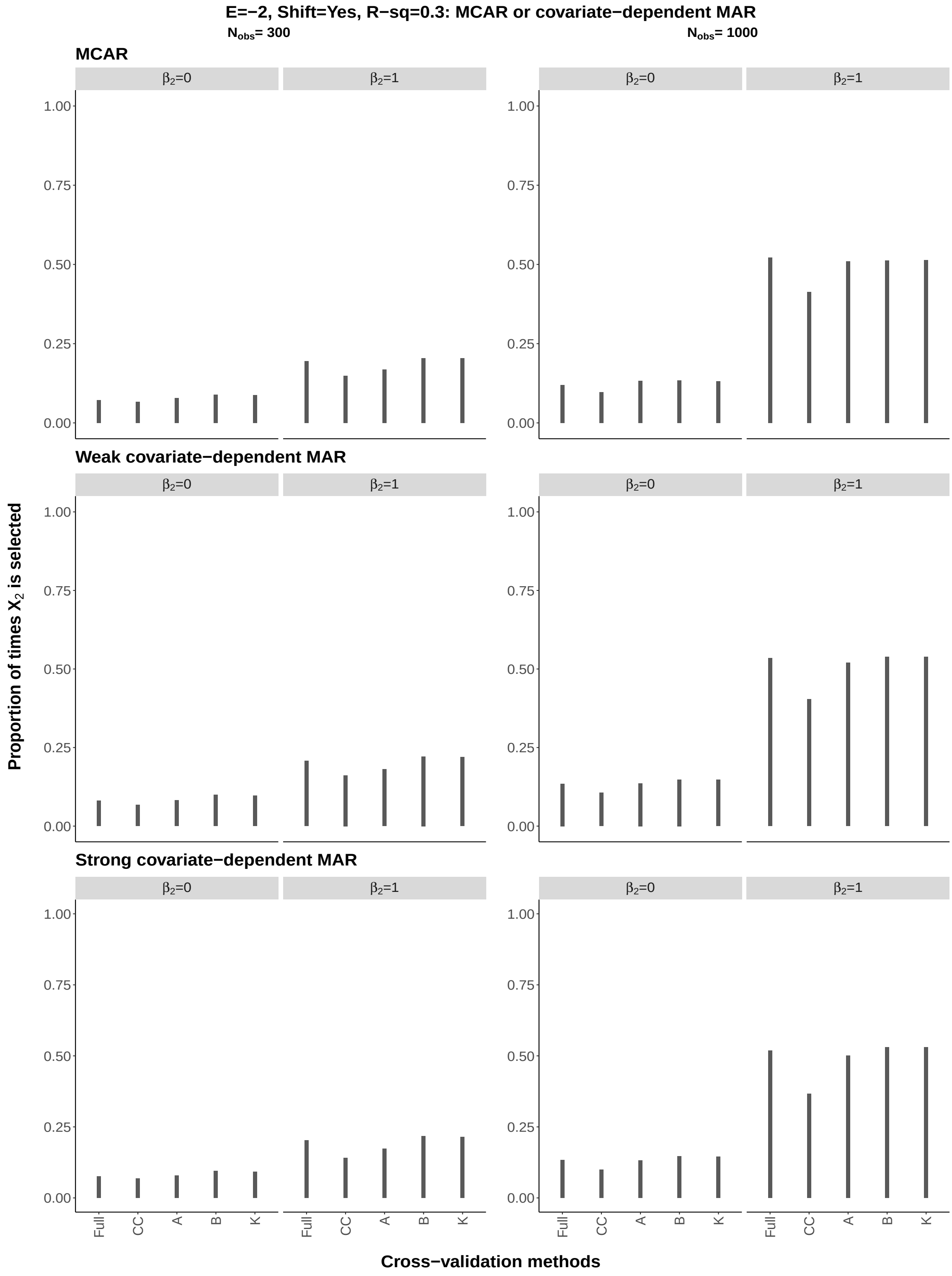

Figure S46: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

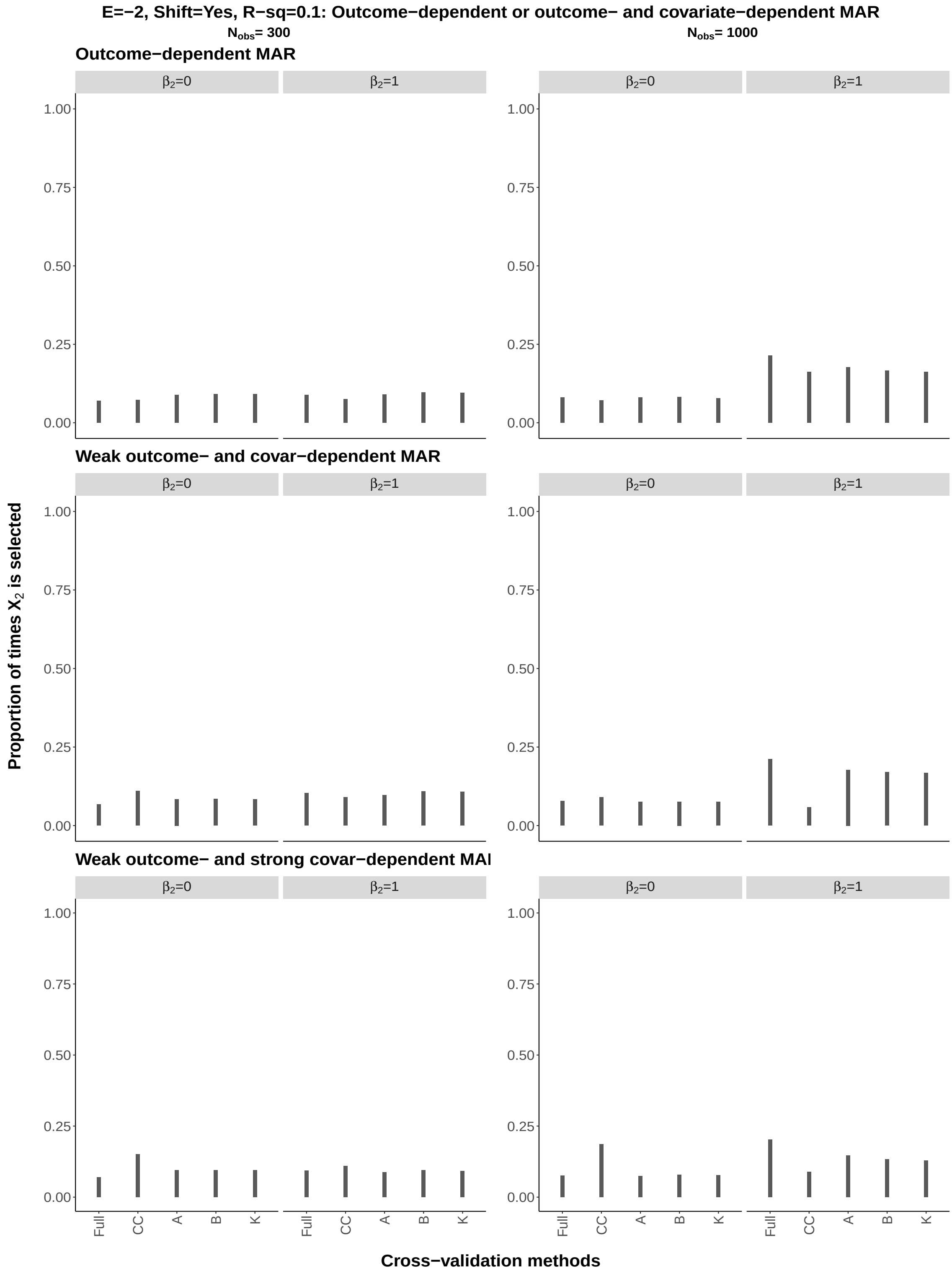

Figure S47: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

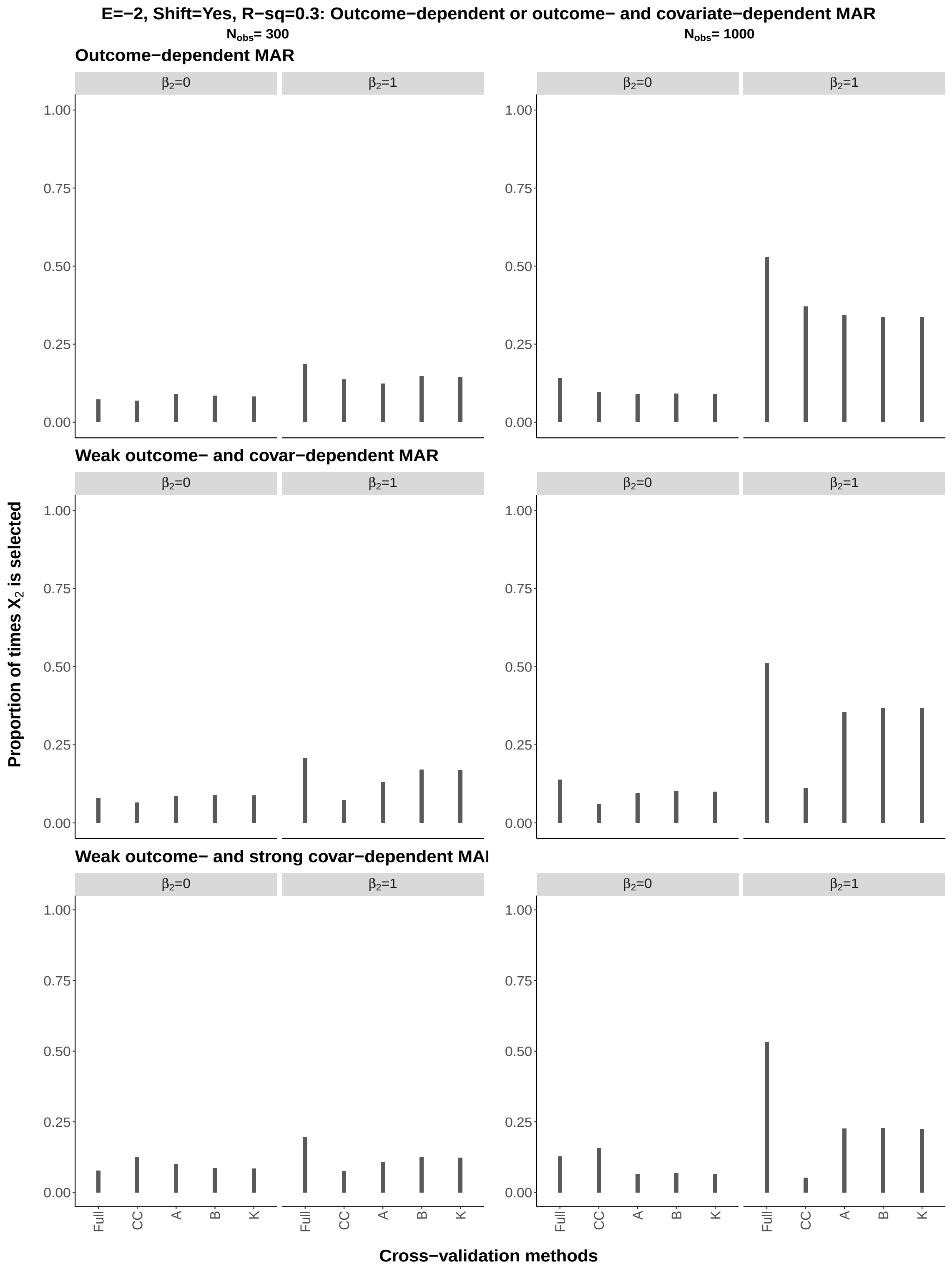

Figure S48: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.1.5 Covariate selection of X_1: $\beta_2 = 1$, $\alpha_E = 1$, no origin-shift

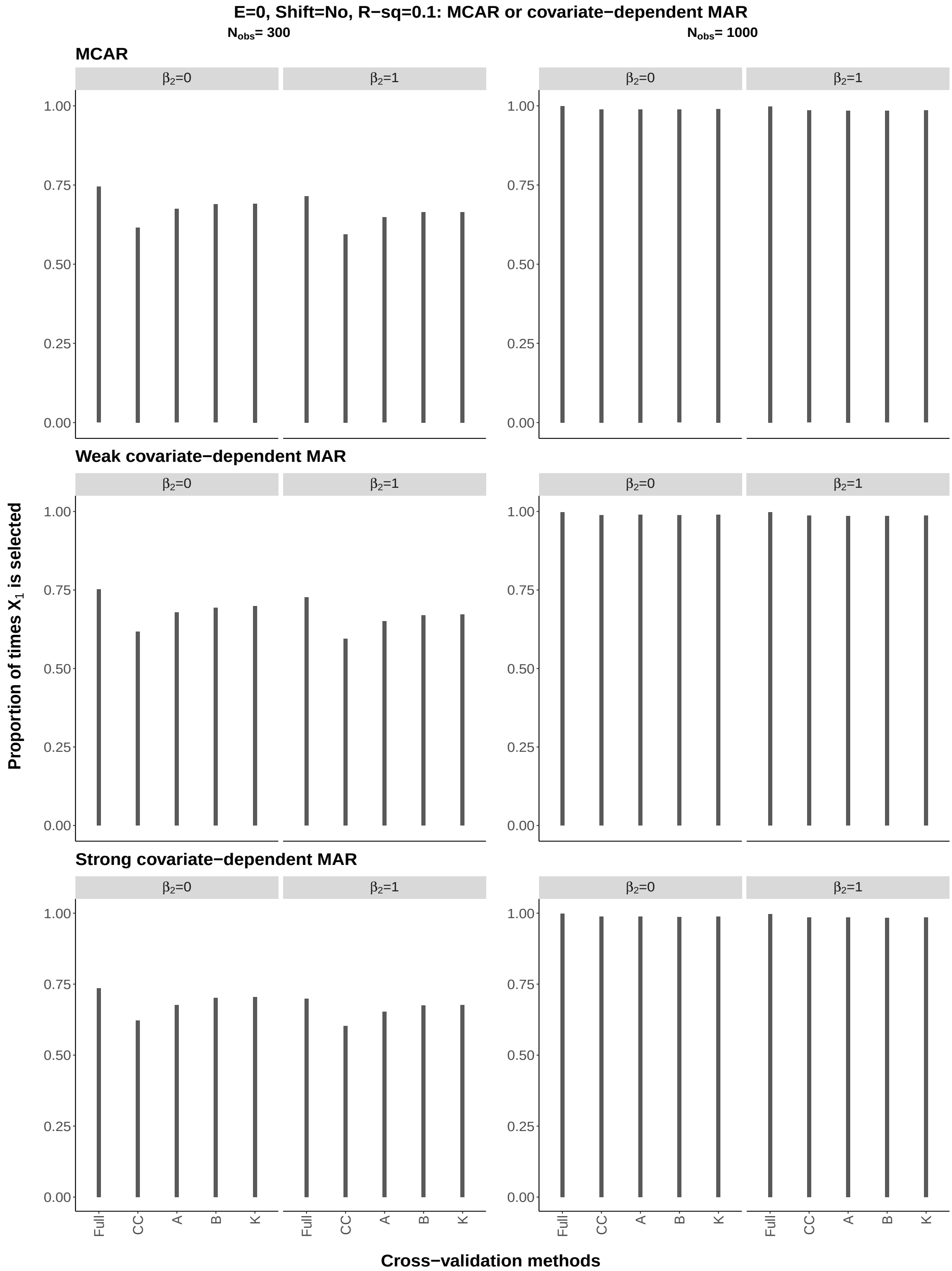

Figure S49: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

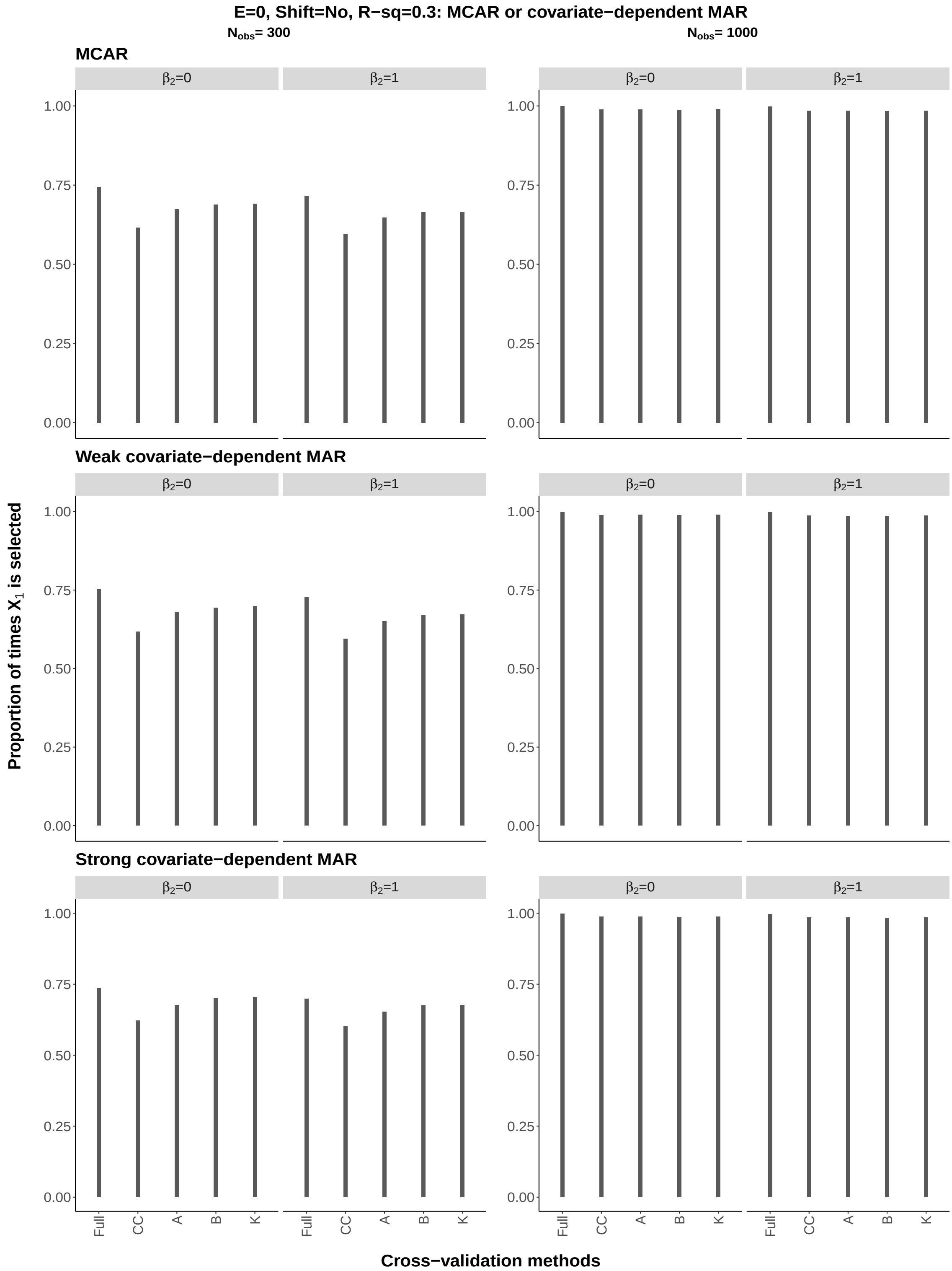

Figure S50: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

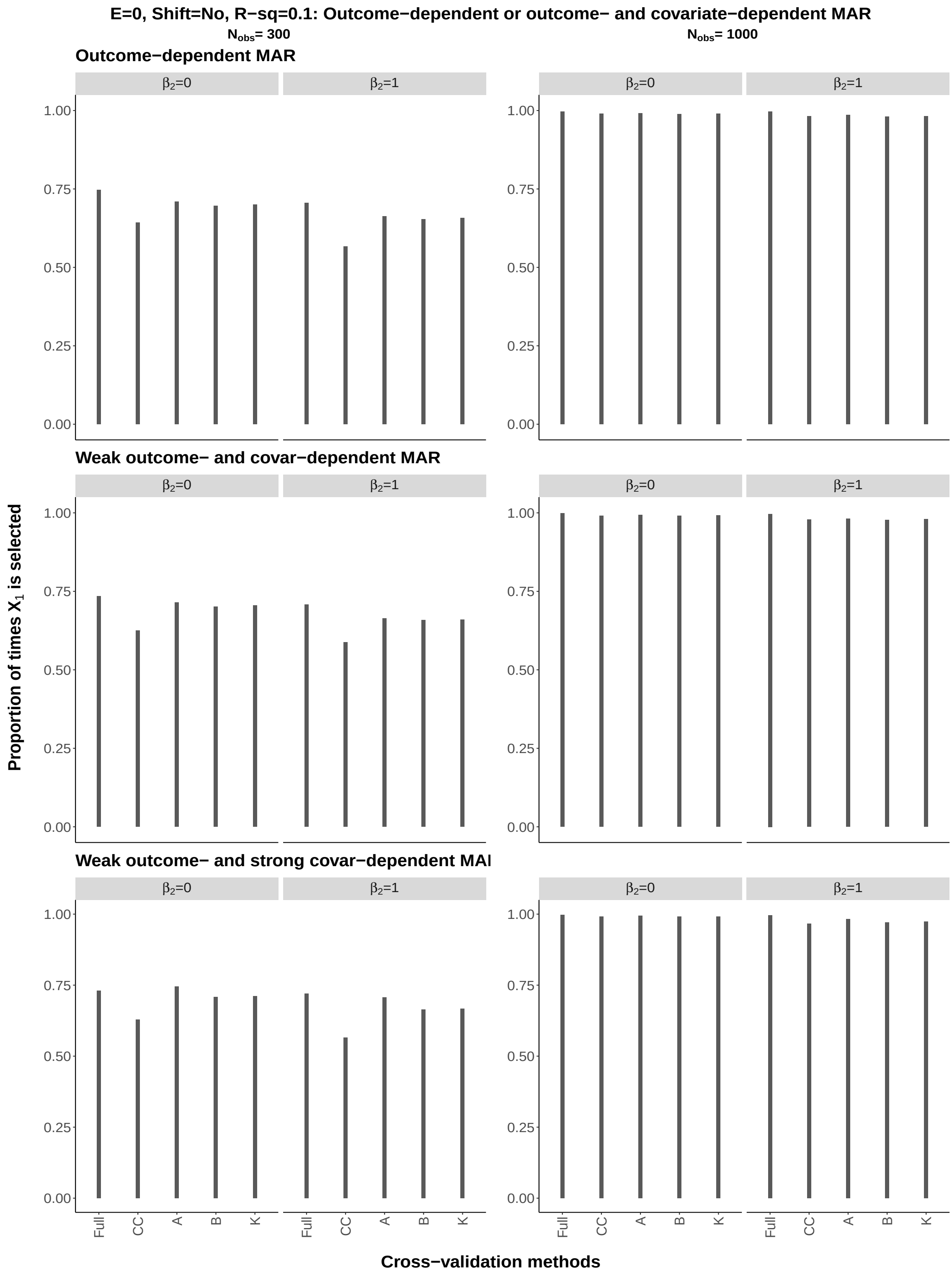

Figure S51: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

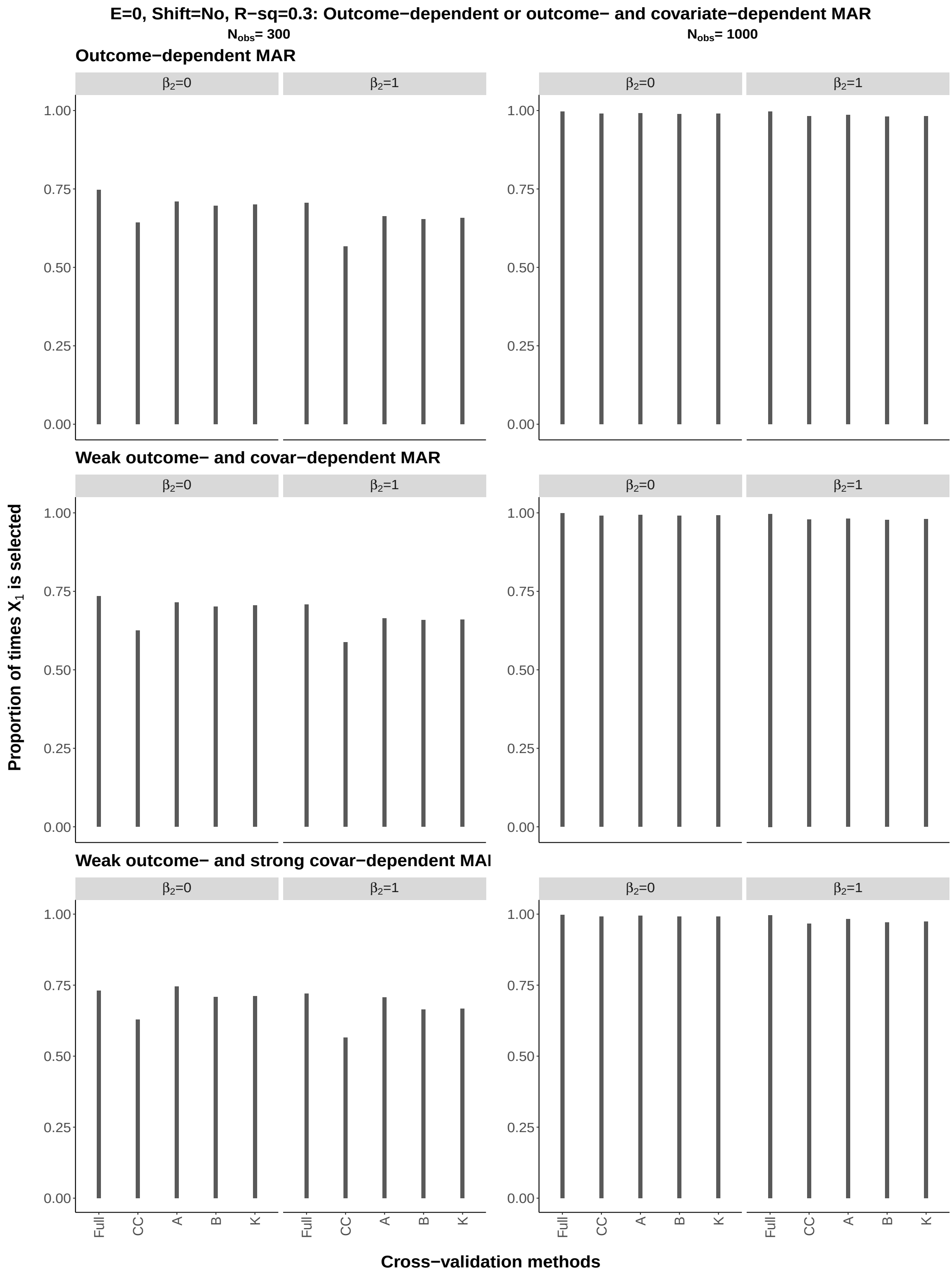

Figure S52: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

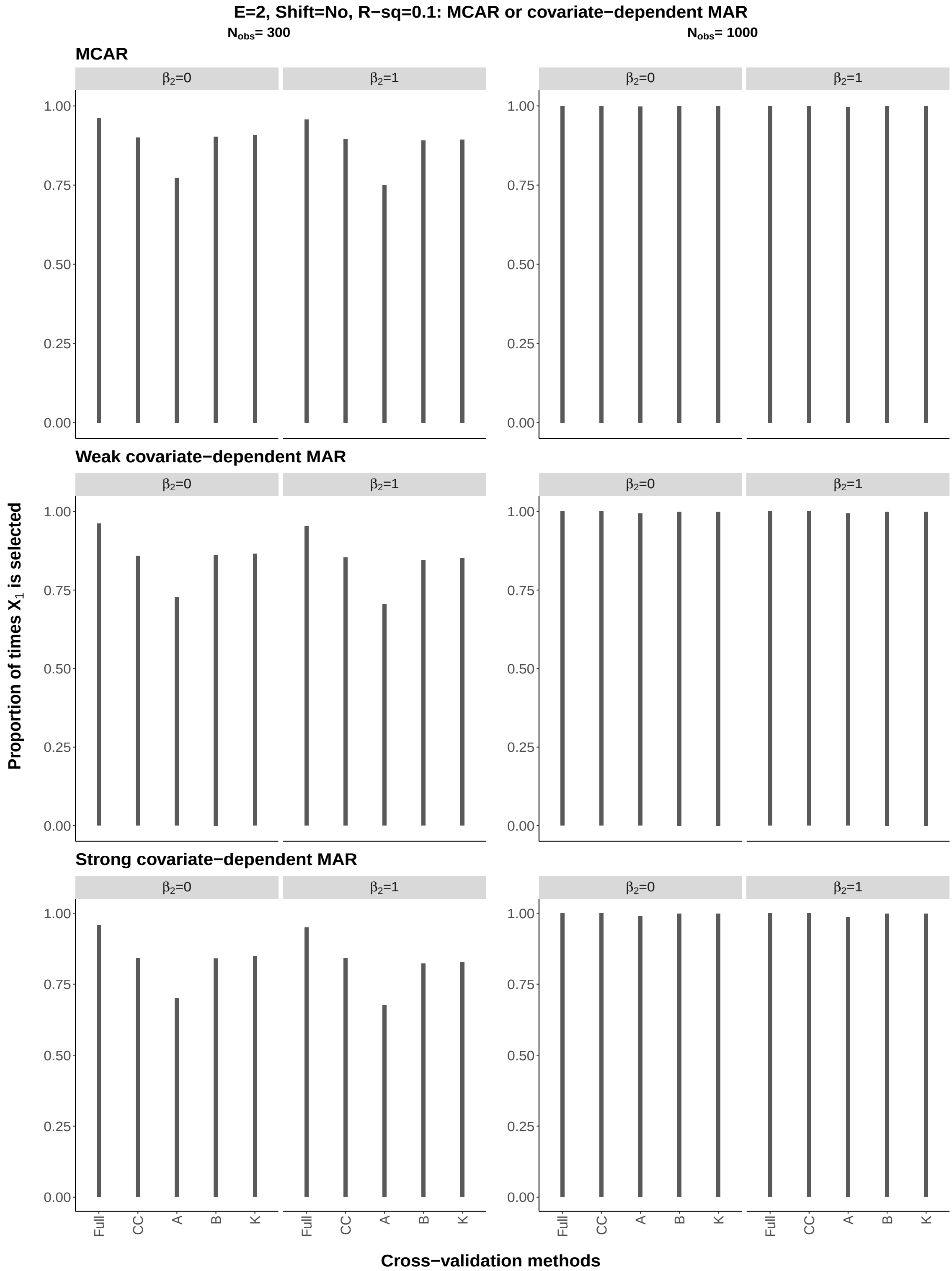

Figure S53: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

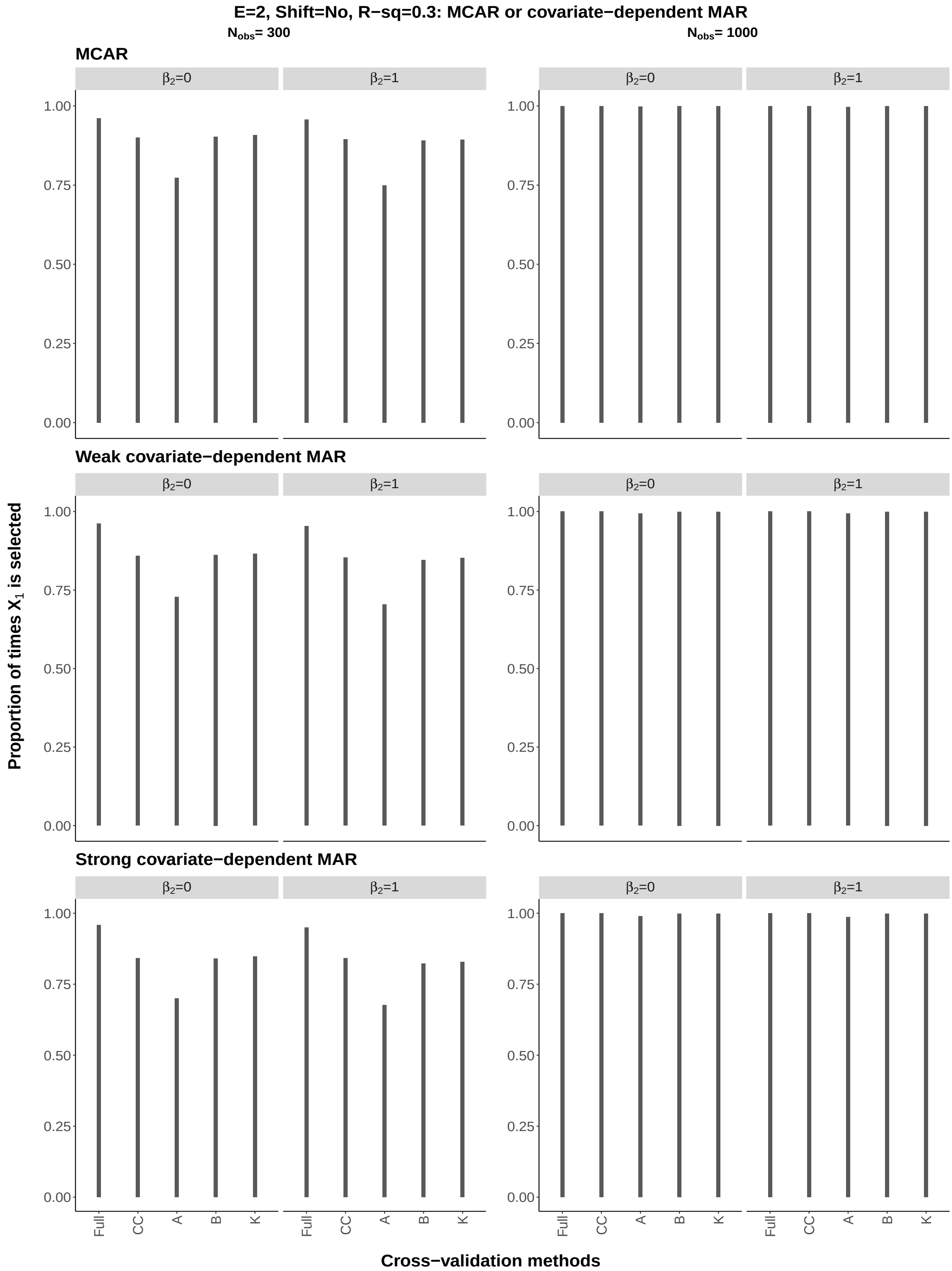

Figure S54: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

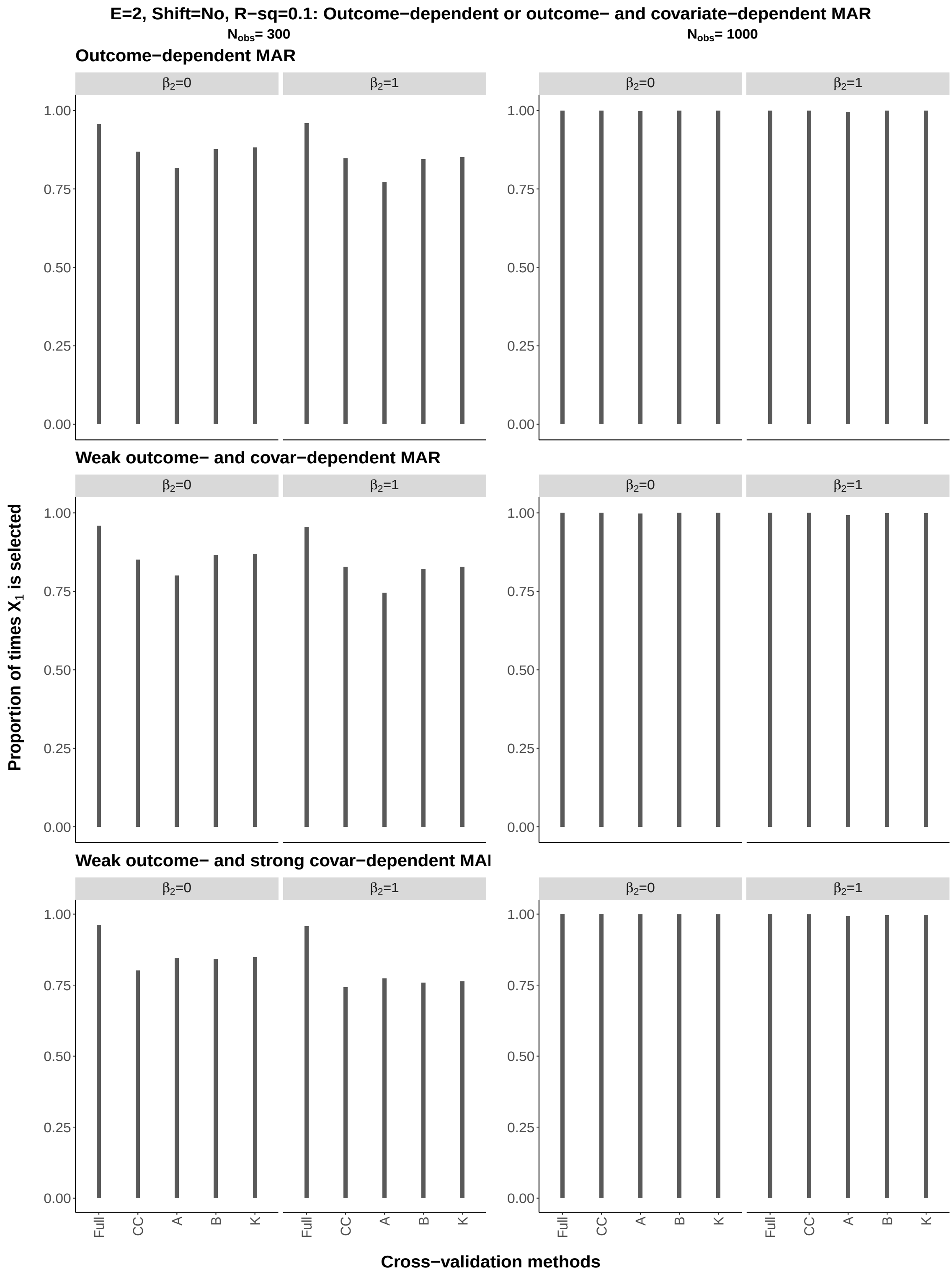

Figure S55: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

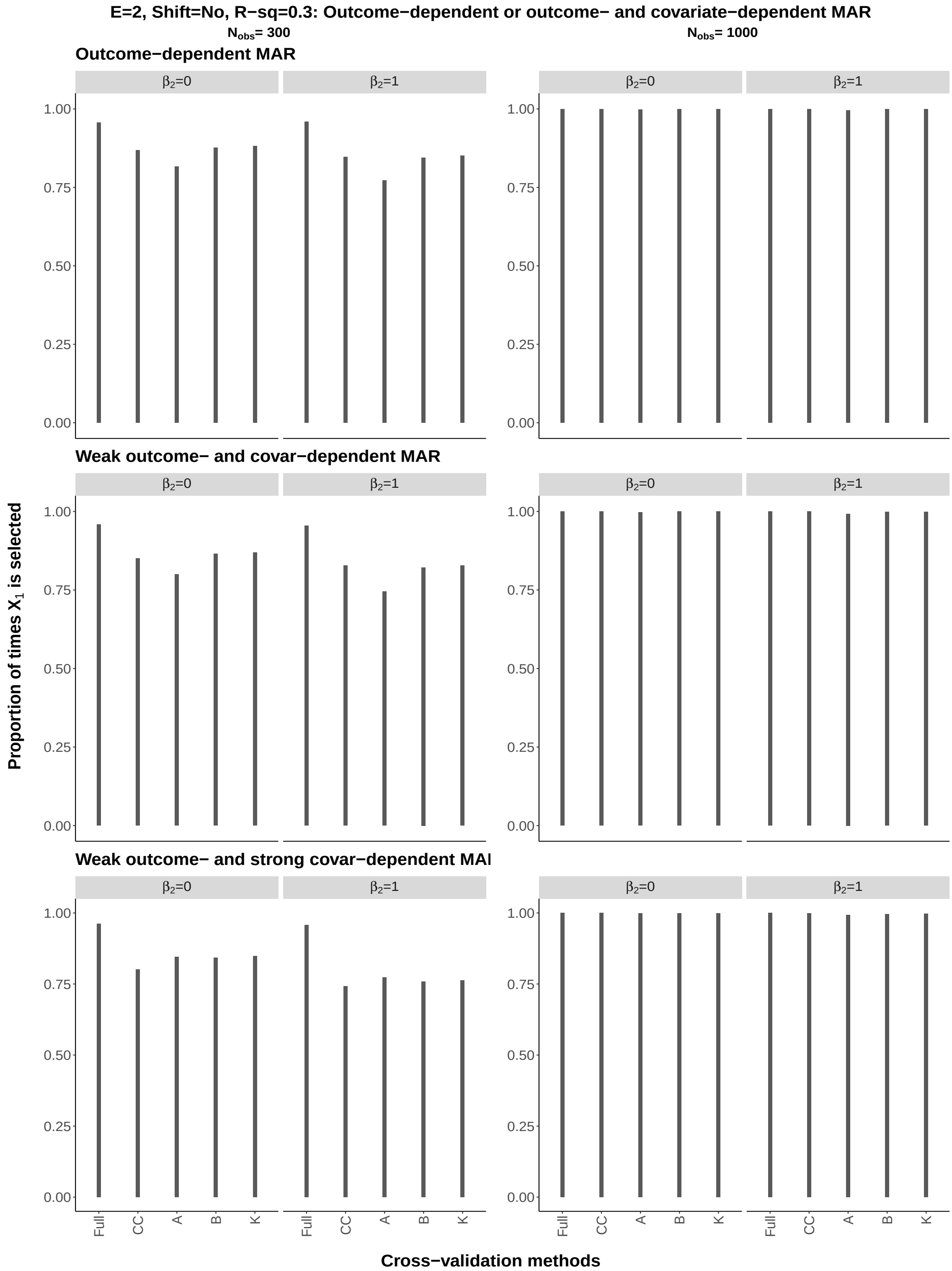

Figure S56: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

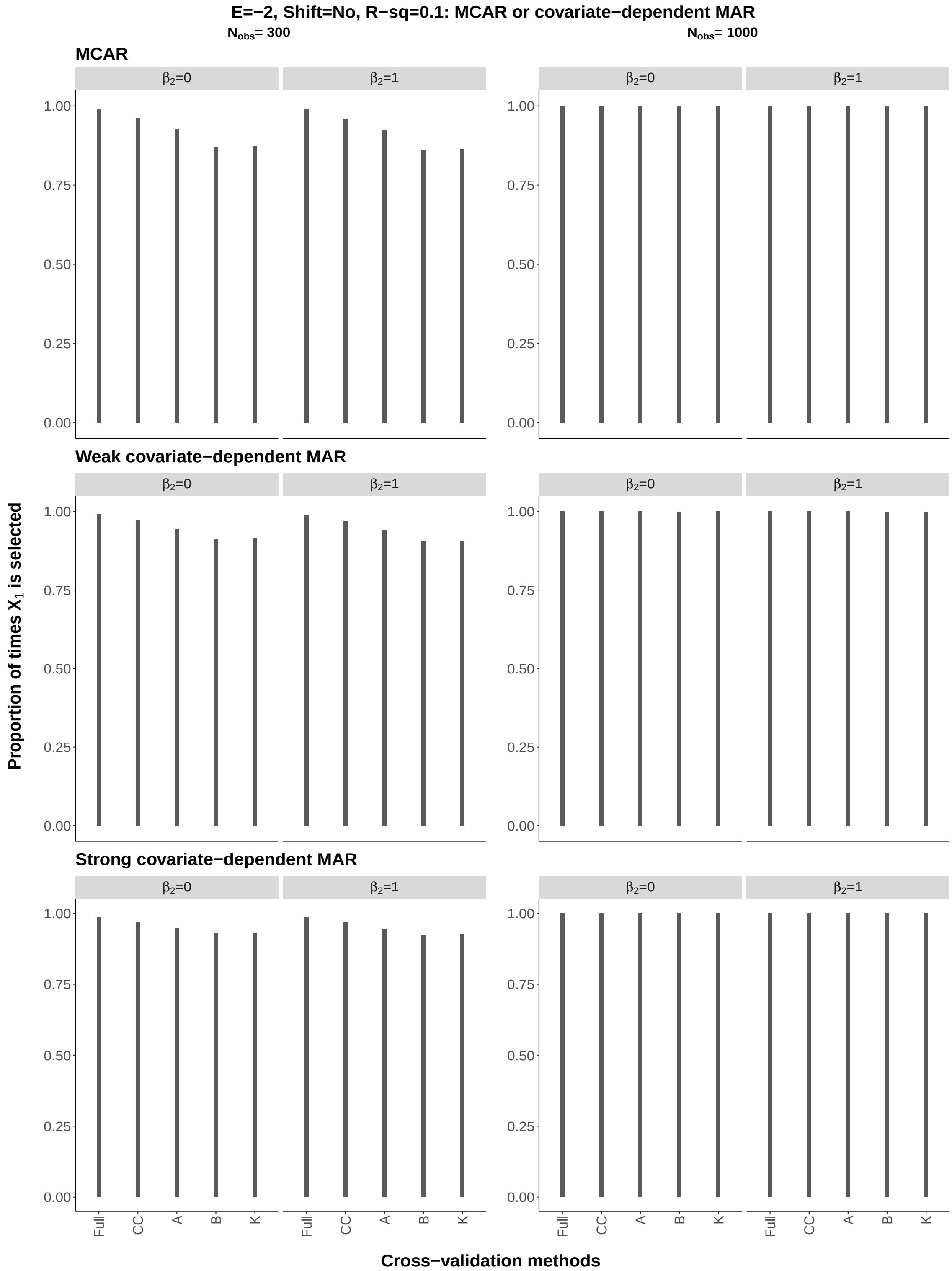

Figure S57: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

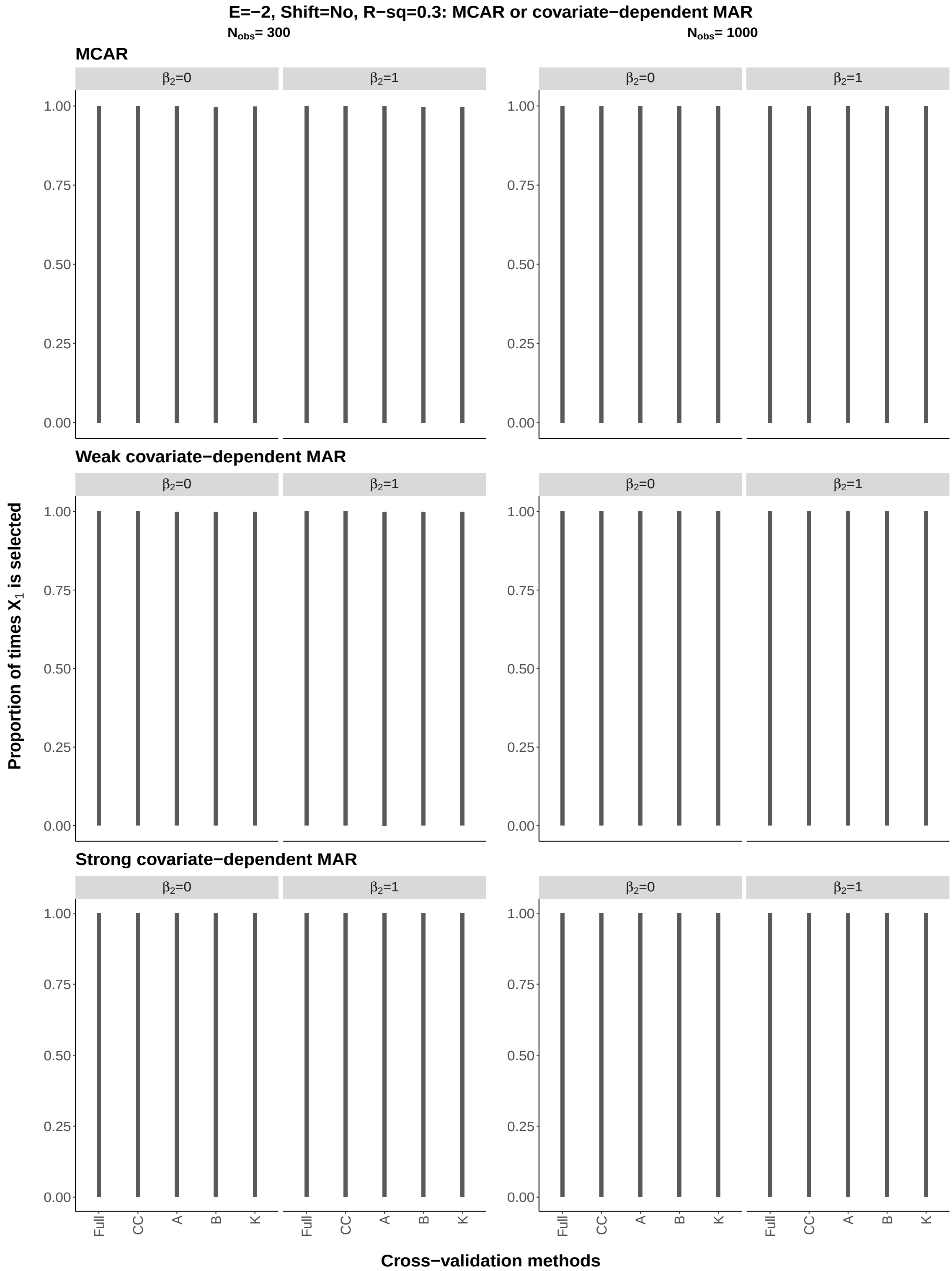

Figure S58: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

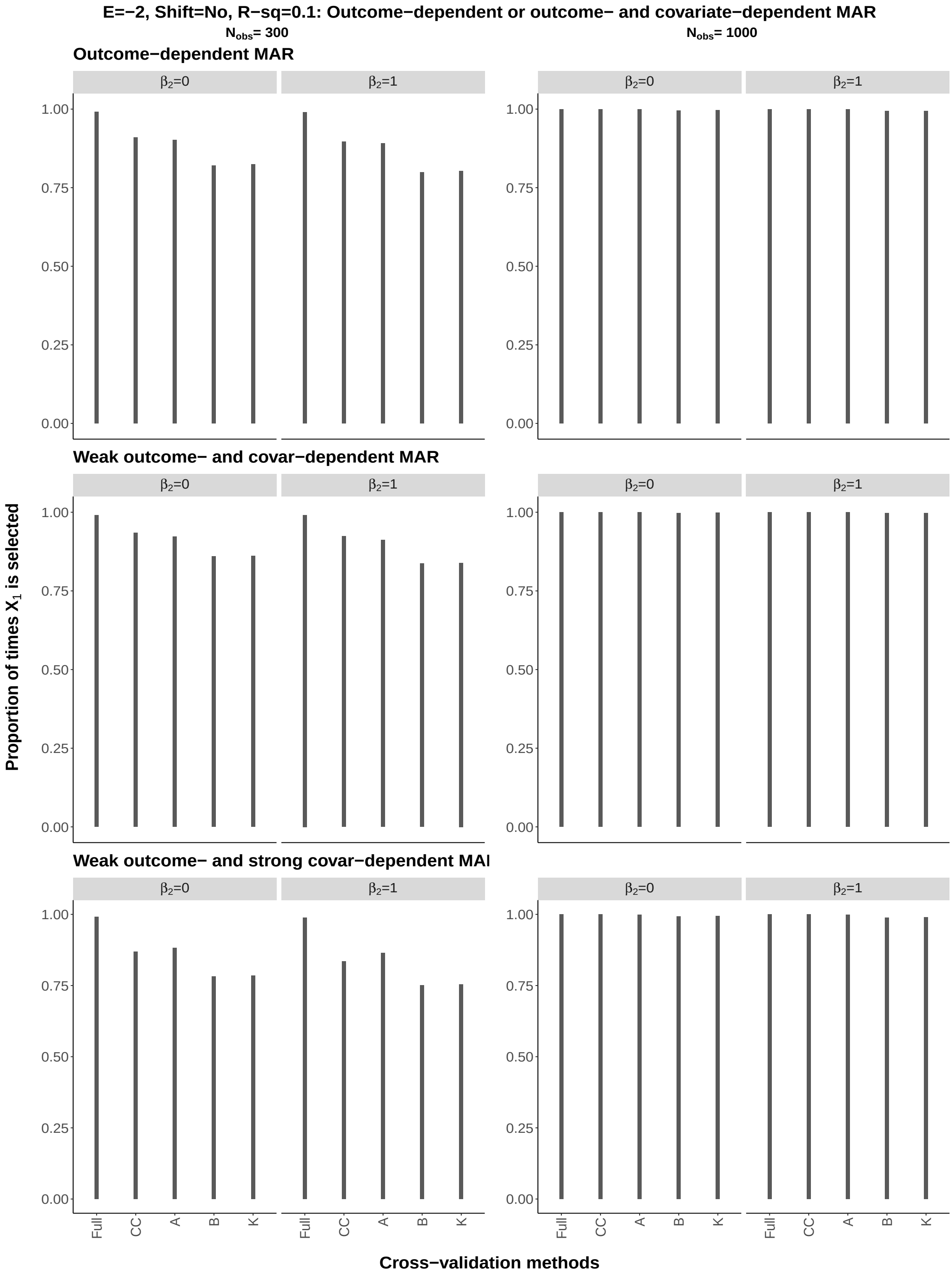

Figure S59: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

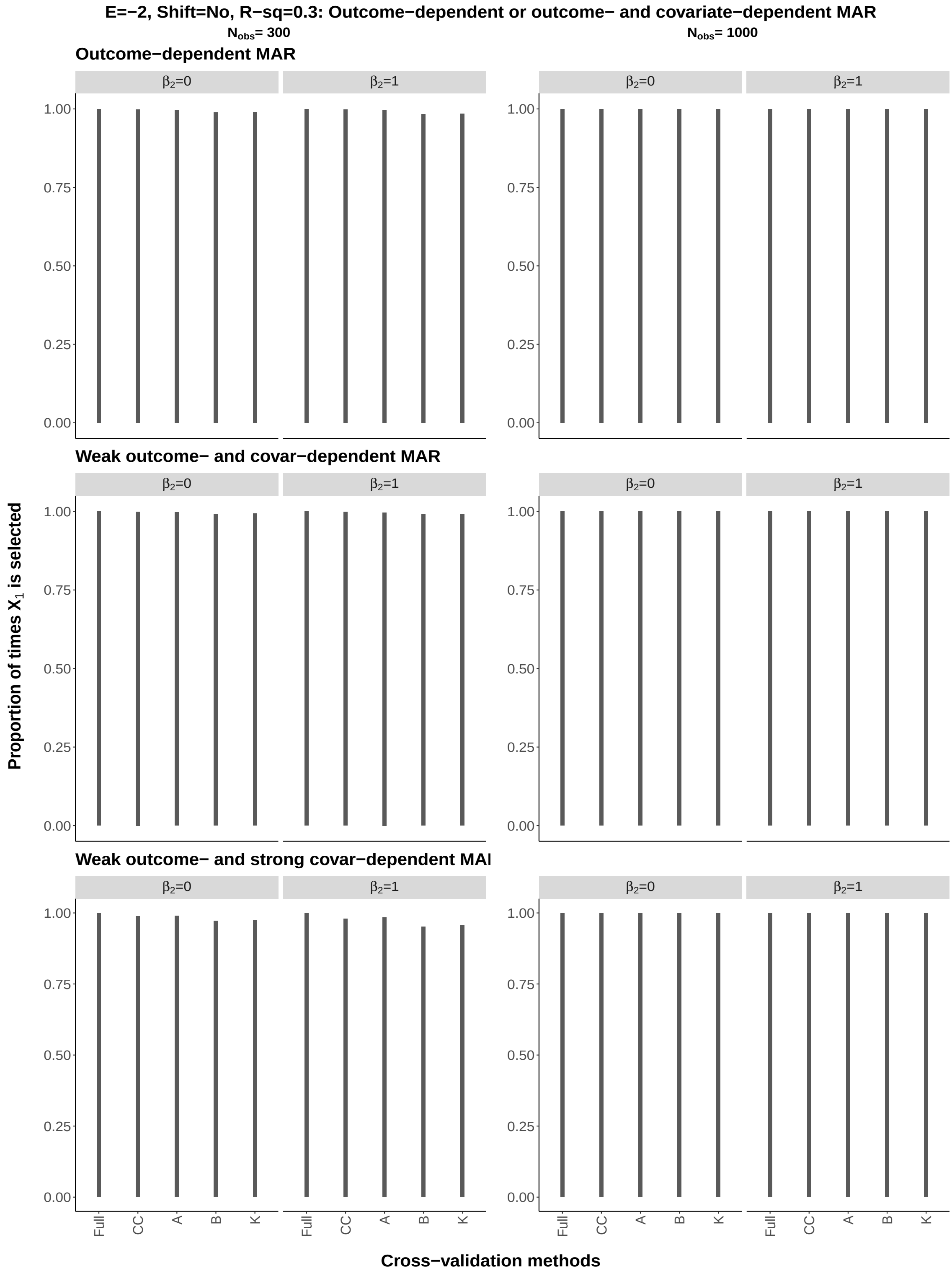

Figure S60: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.1.6 Covariate selection of X_1: $\beta_2 = 1$, $\alpha_E = 0.05$, no origin-shift

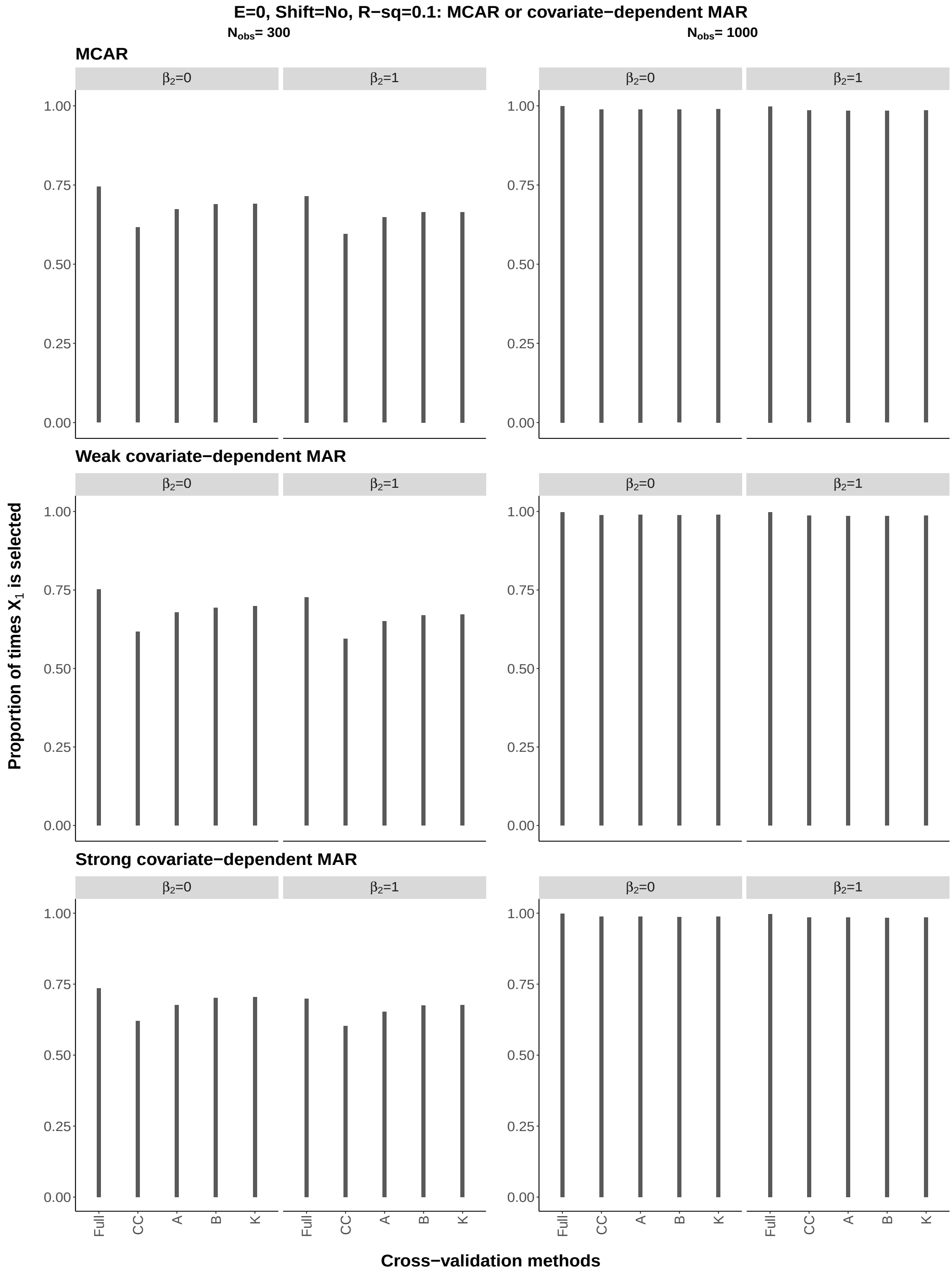

Figure S61: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

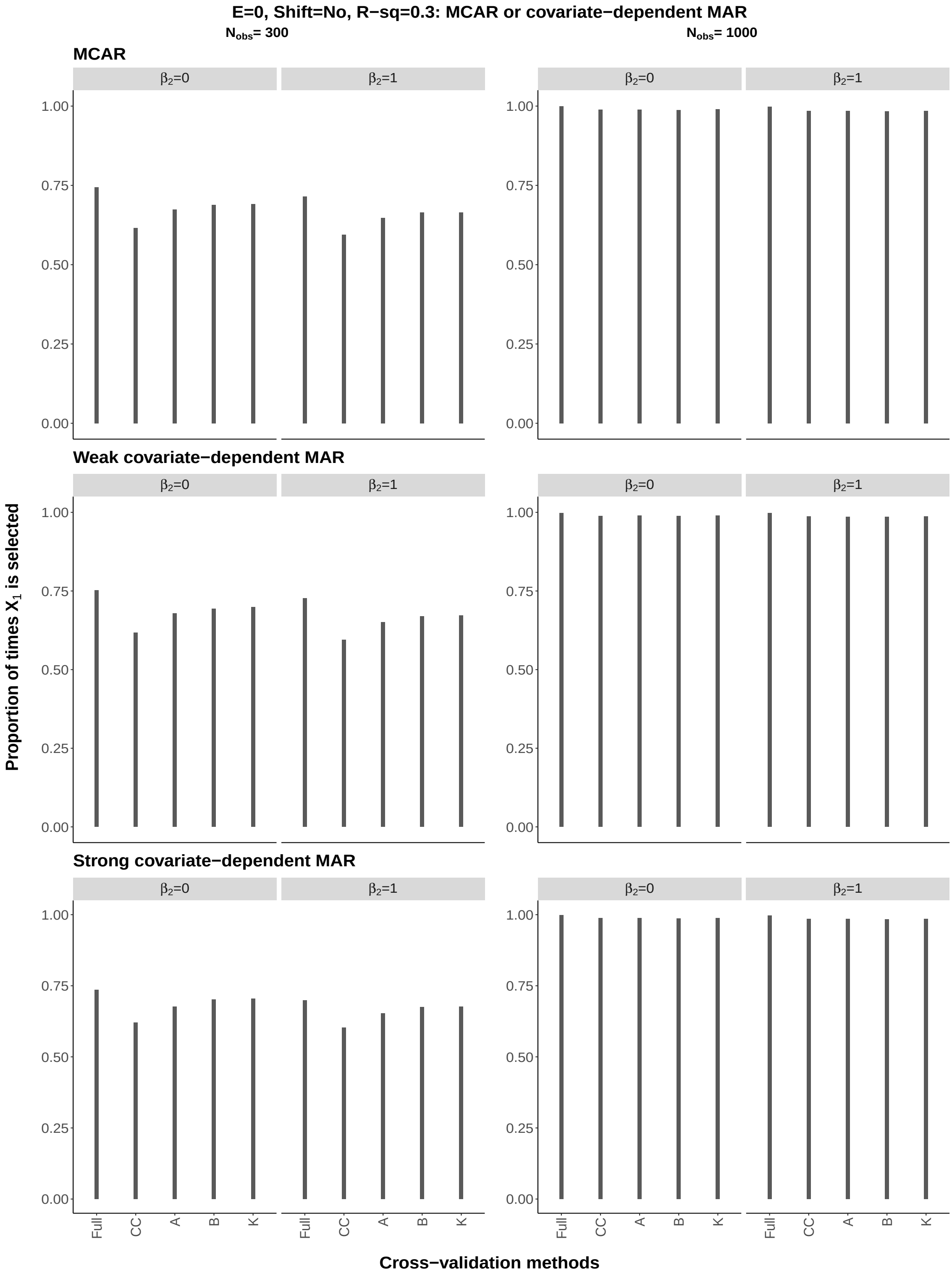

Figure S62: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

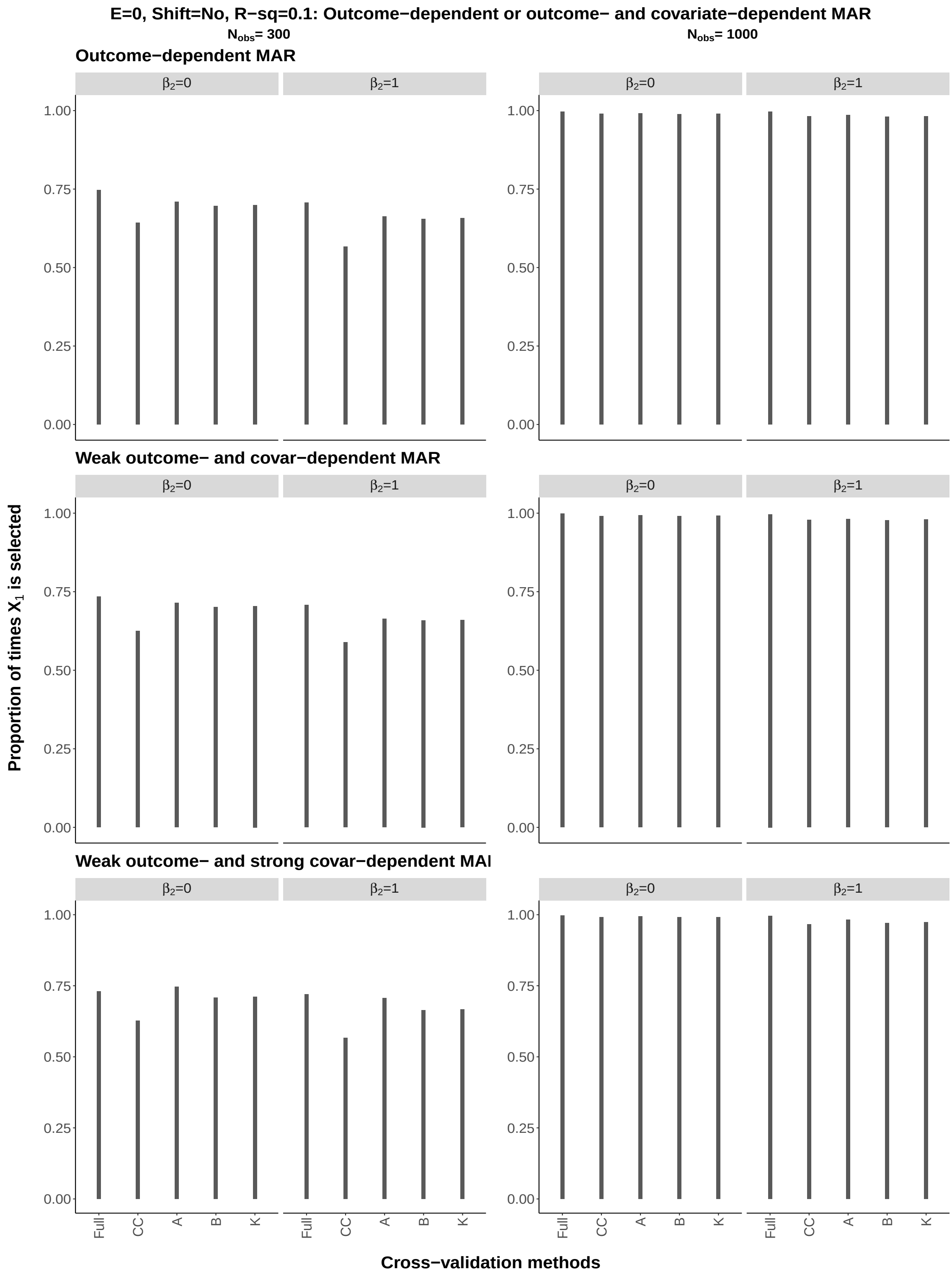

Figure S63: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

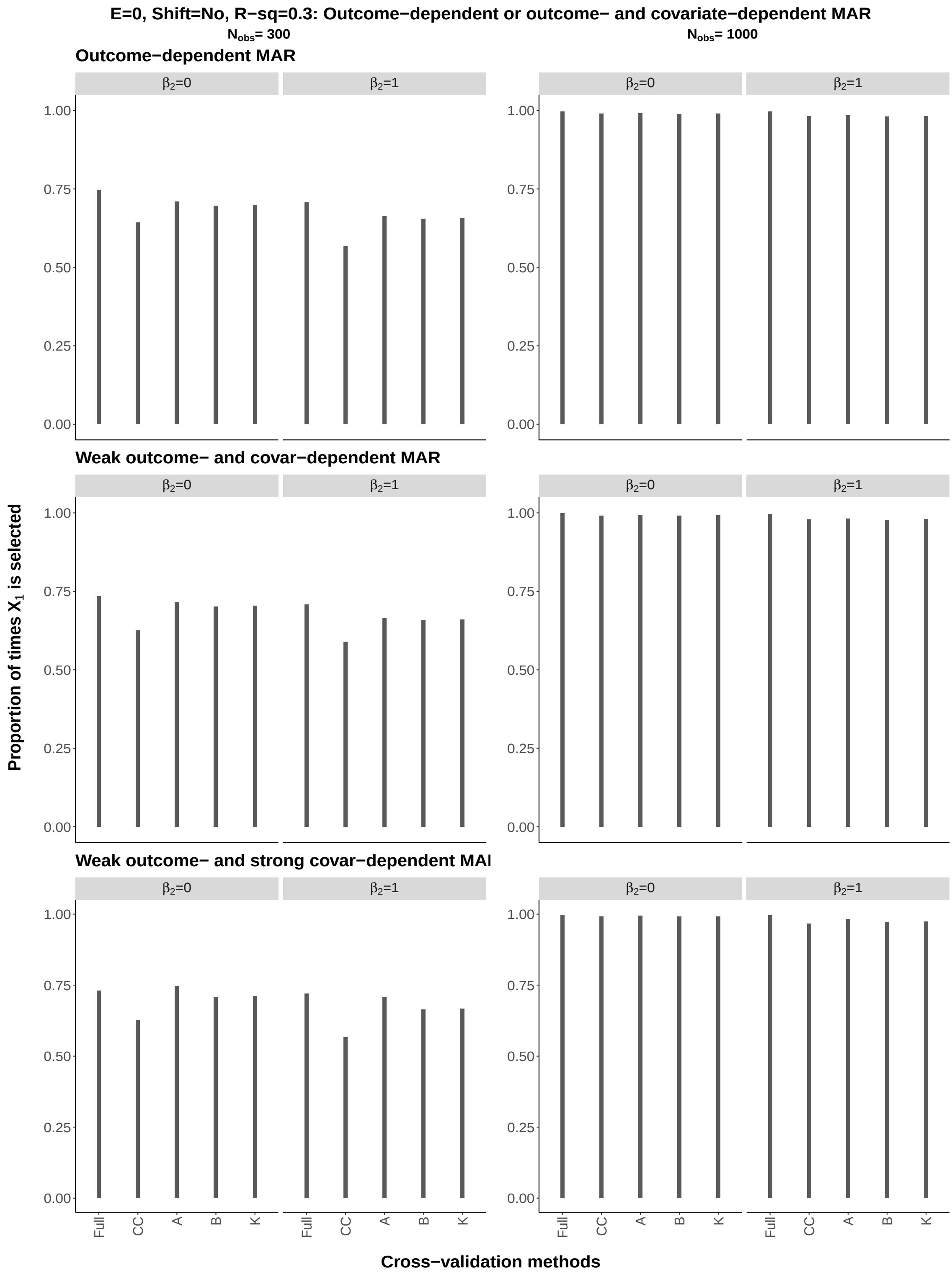

Figure S64: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

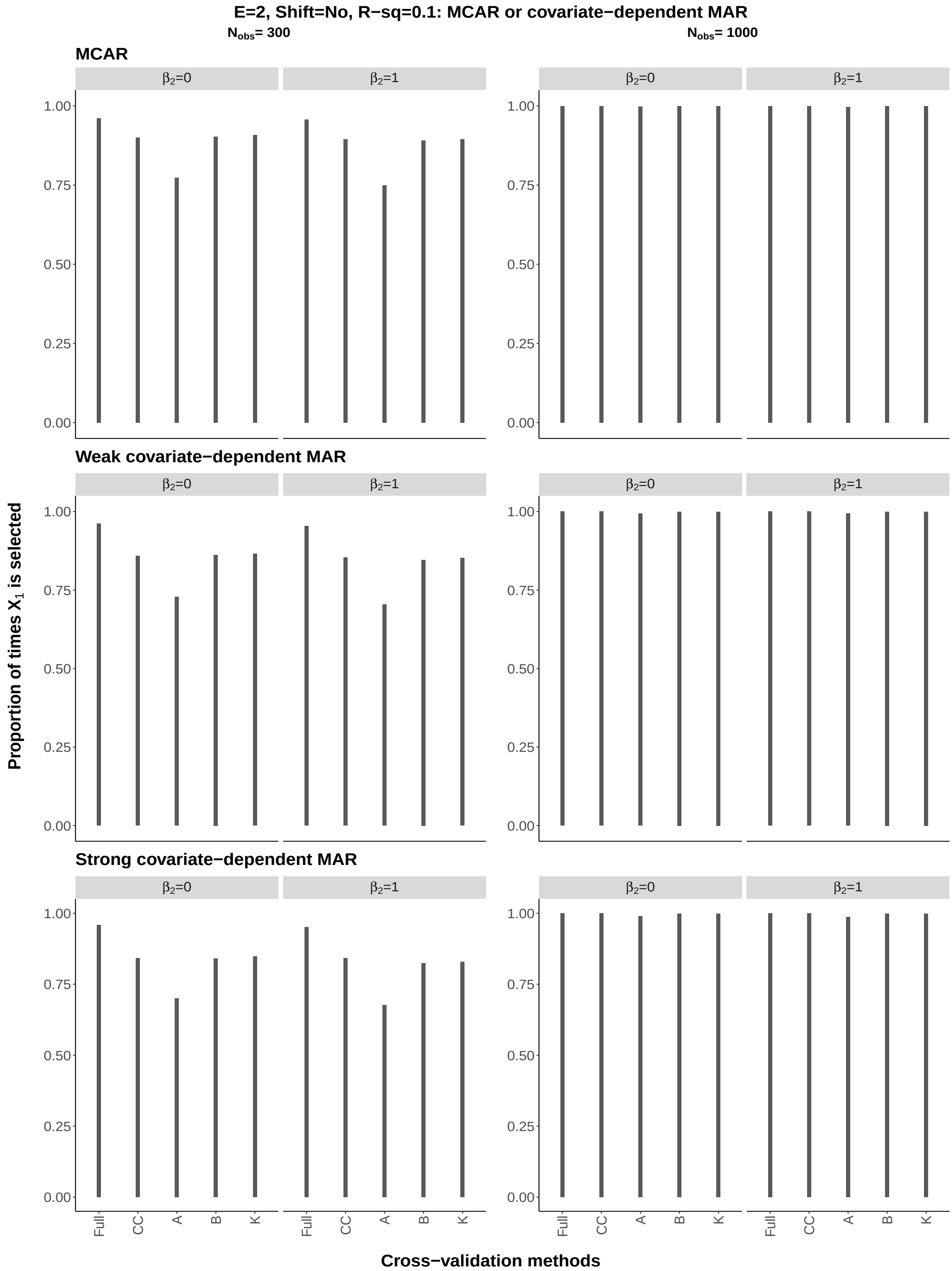

Figure S65: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

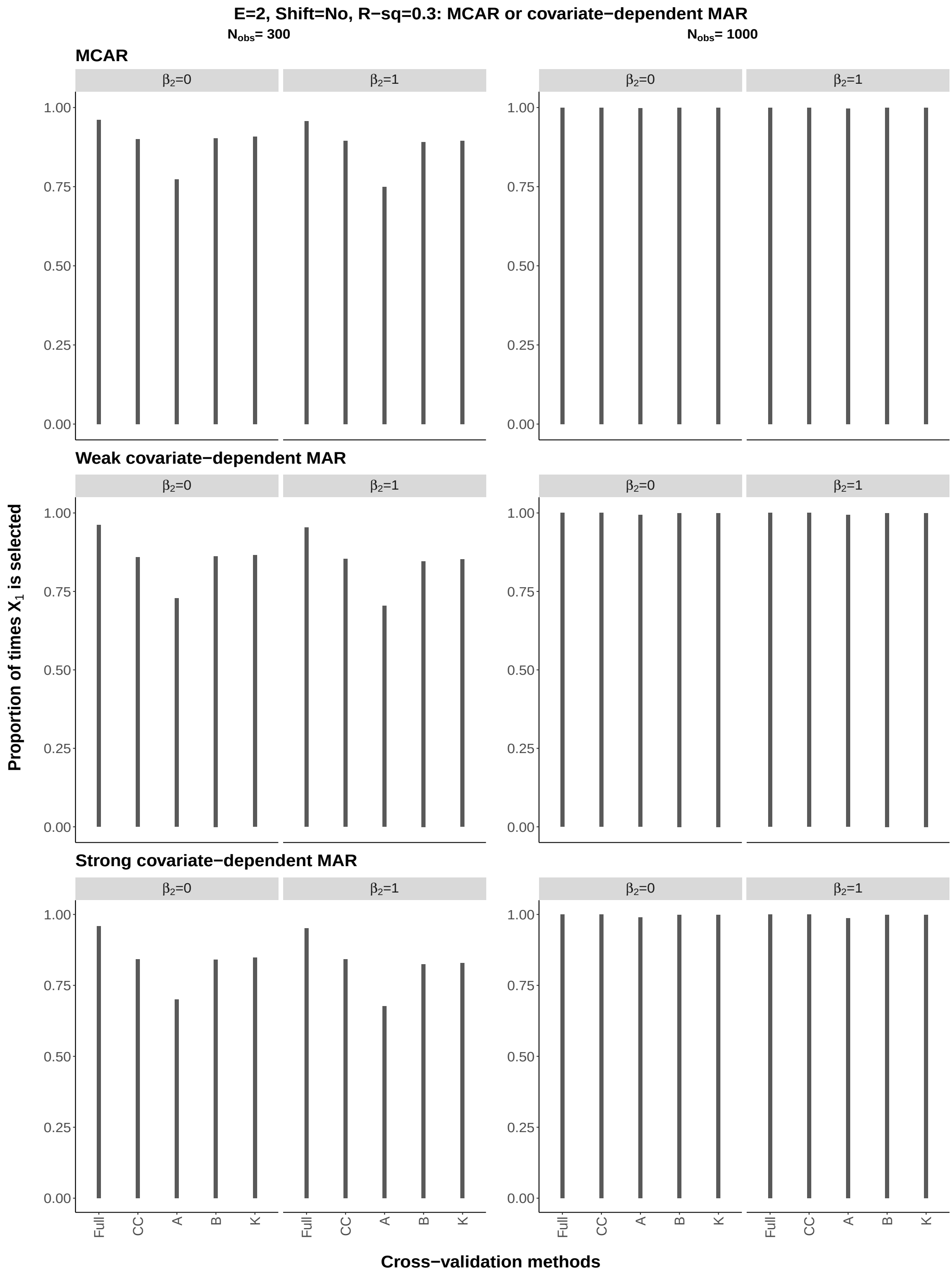

Figure S66: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

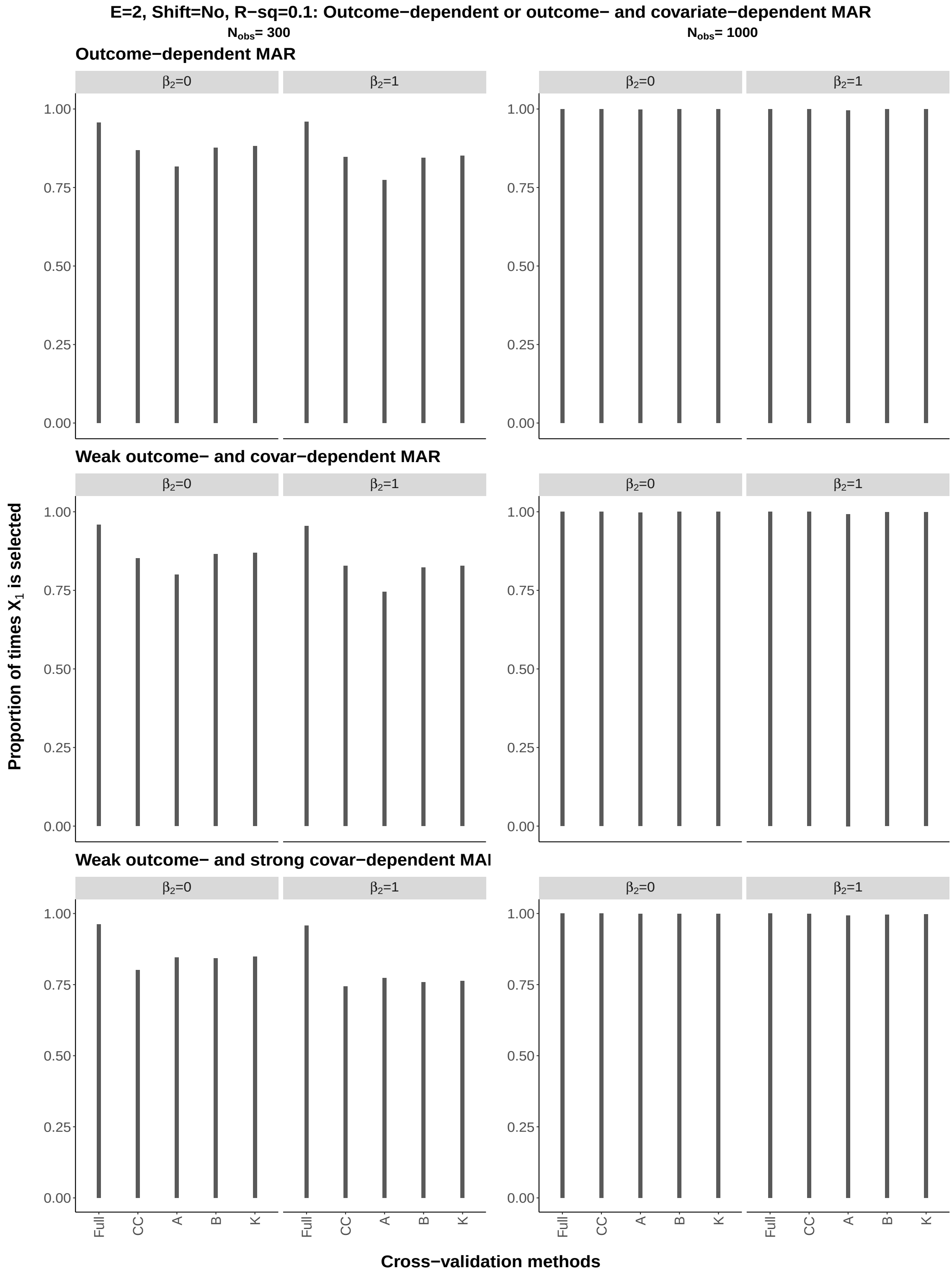

Figure S67: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

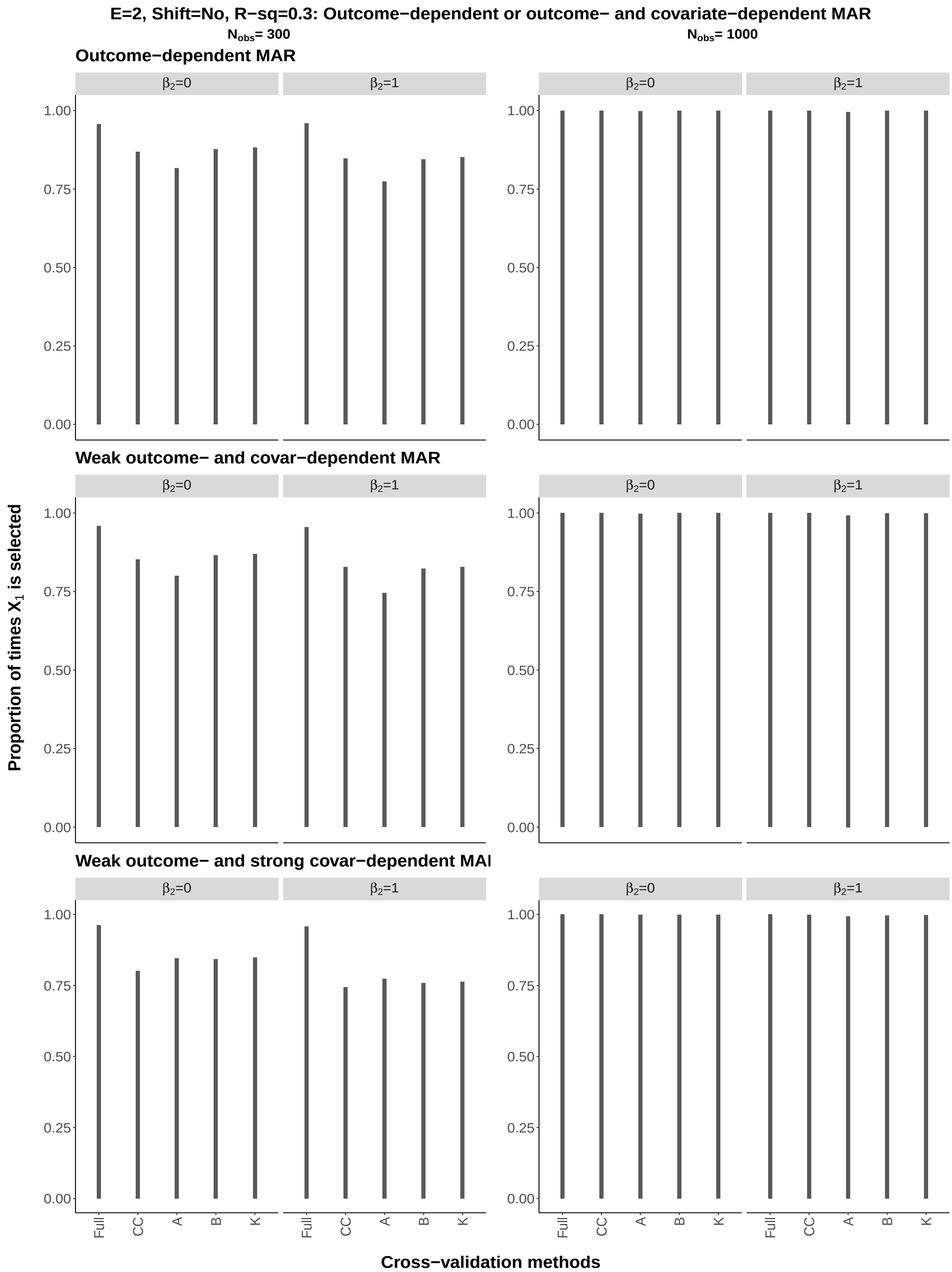

Figure S68: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

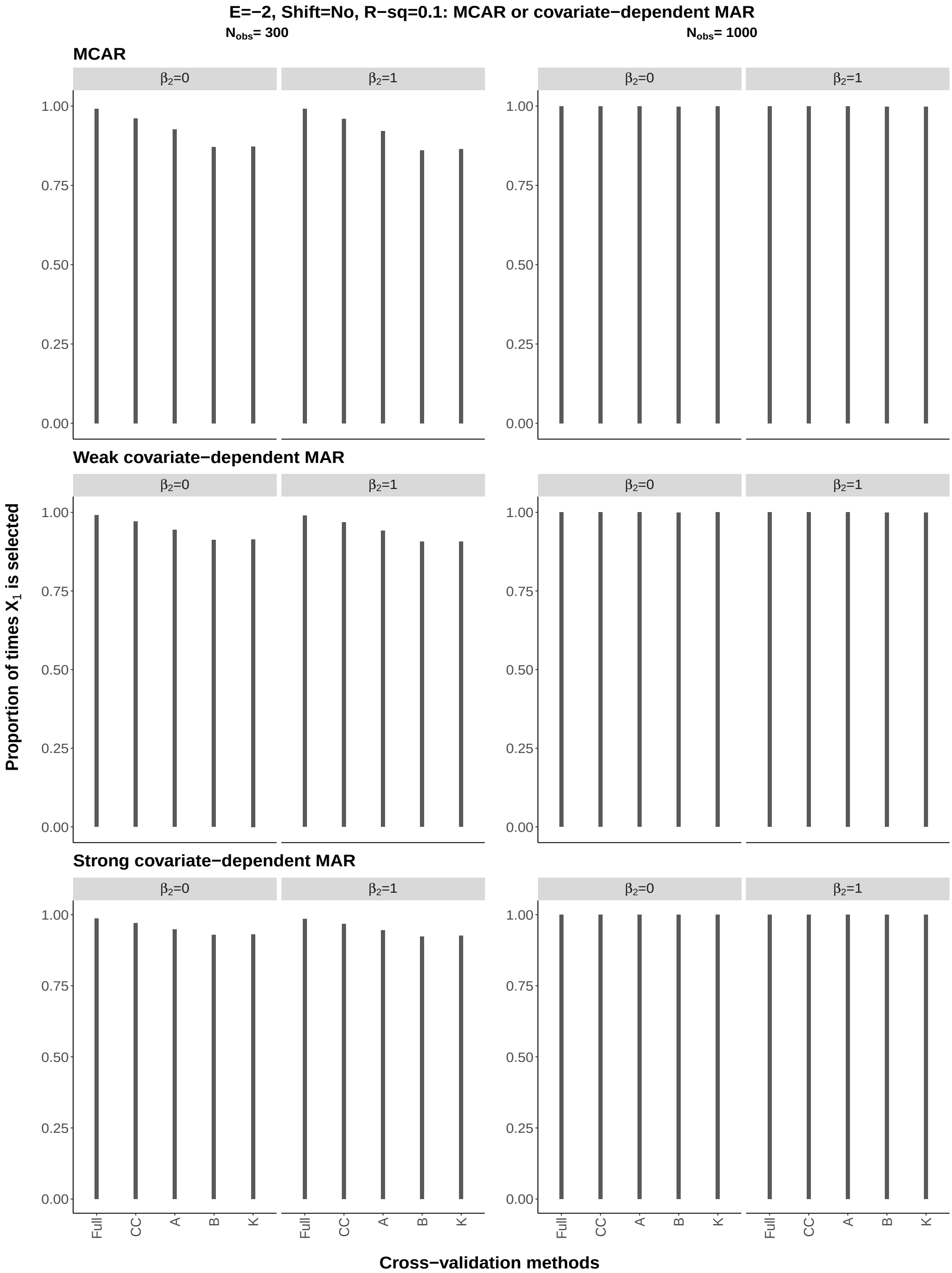

Figure S69: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

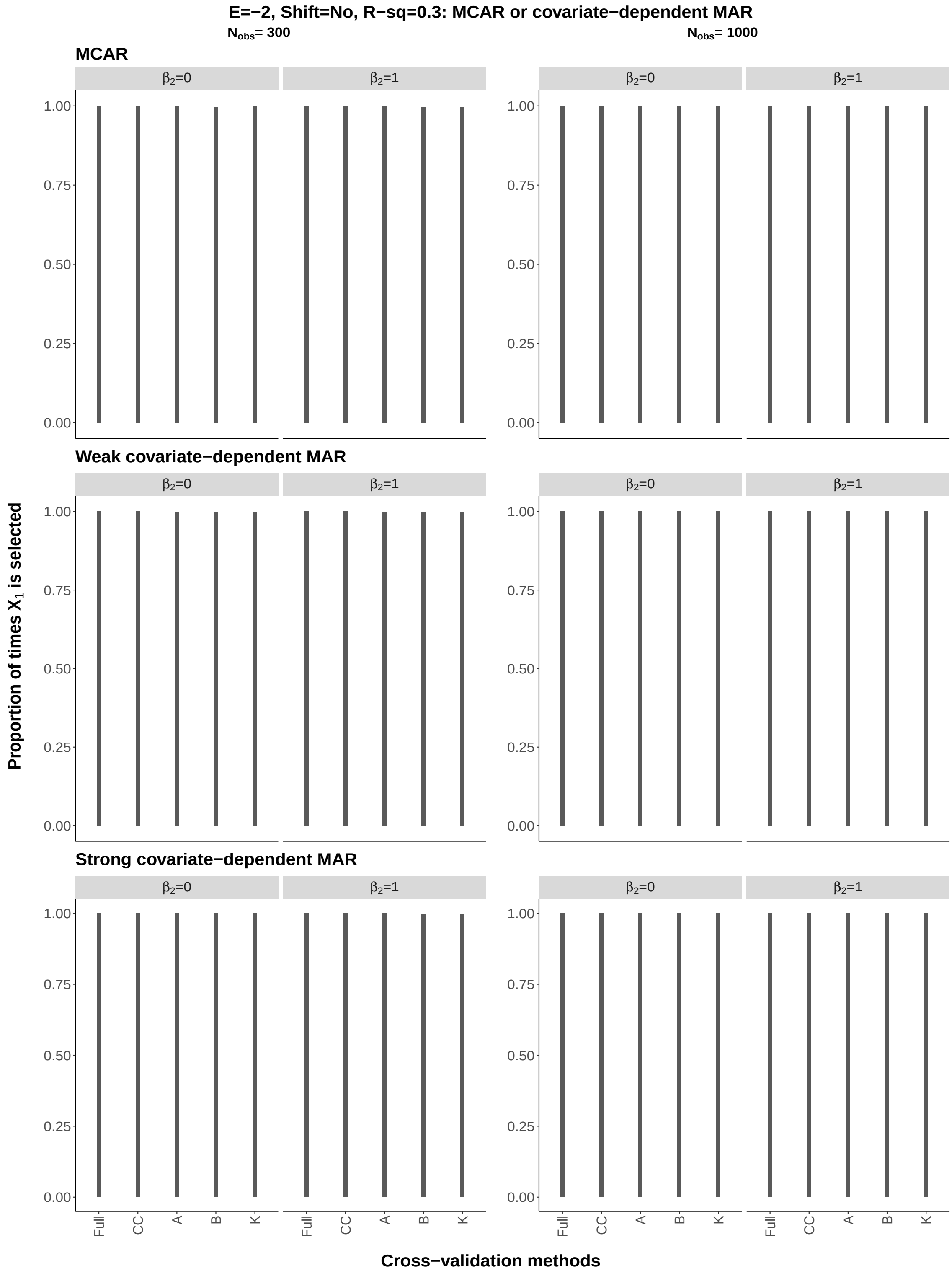

Figure S70: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

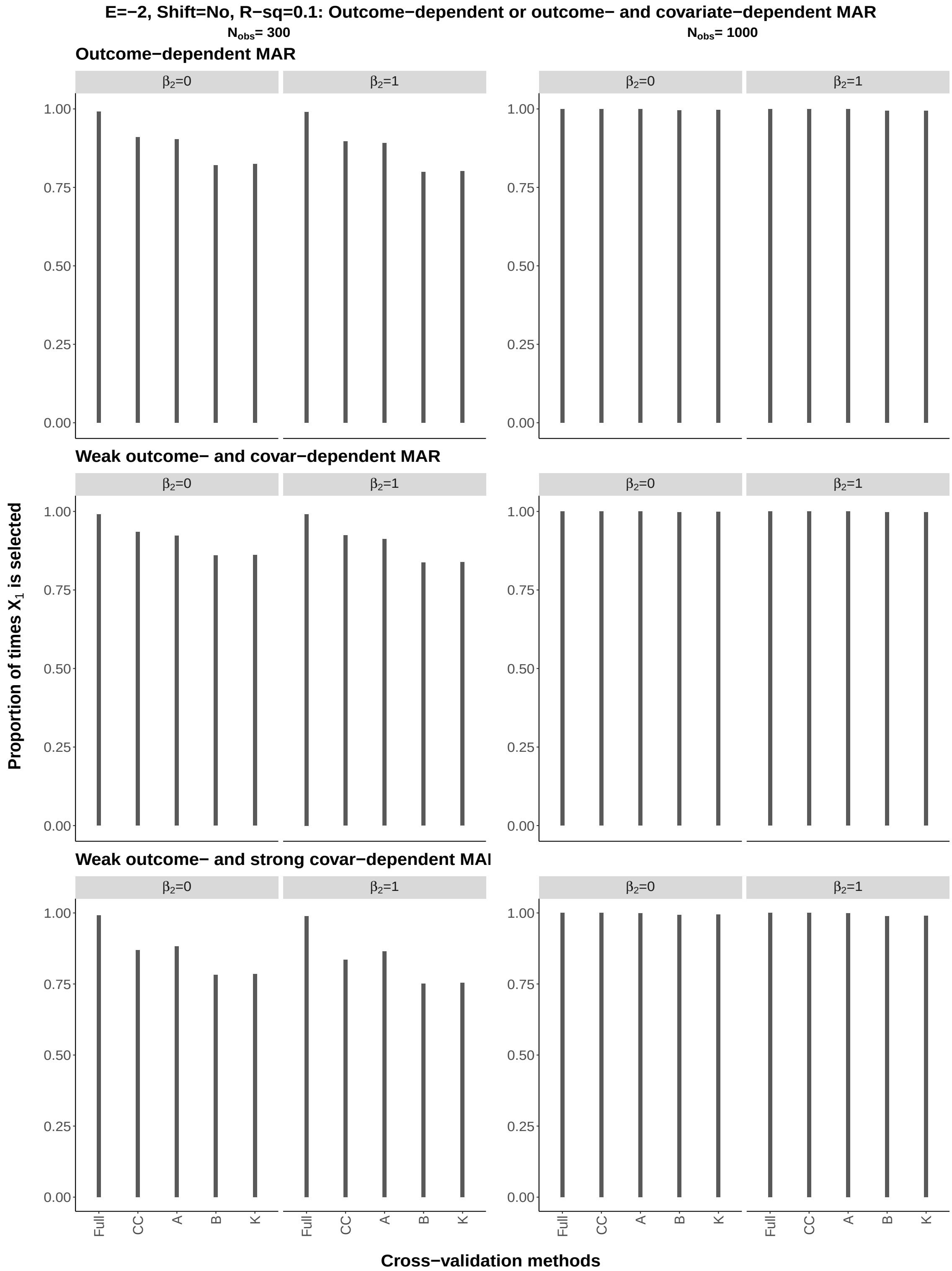

Figure S71: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

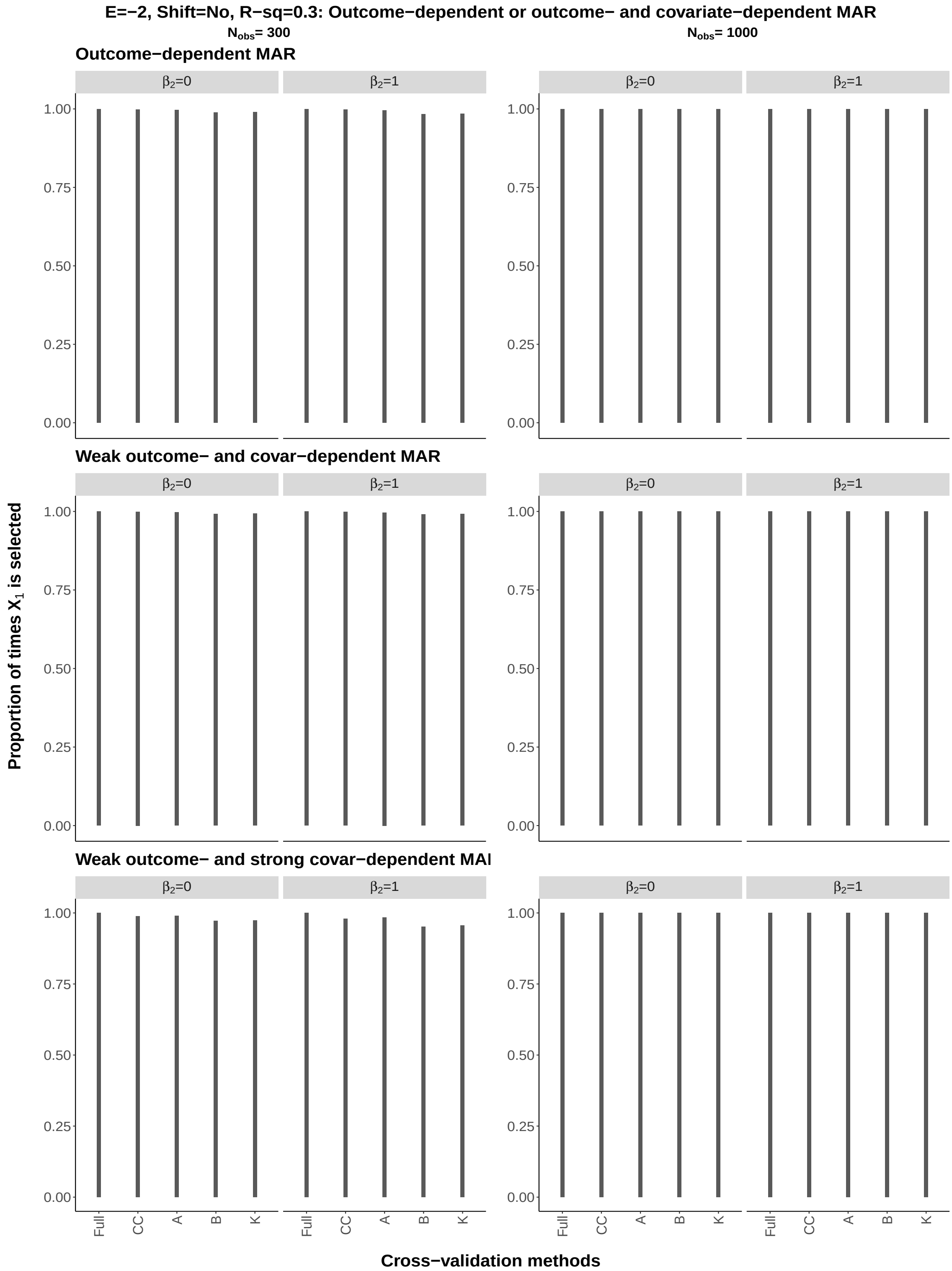

Figure S72: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.1.7 Covariate selection of X_1: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

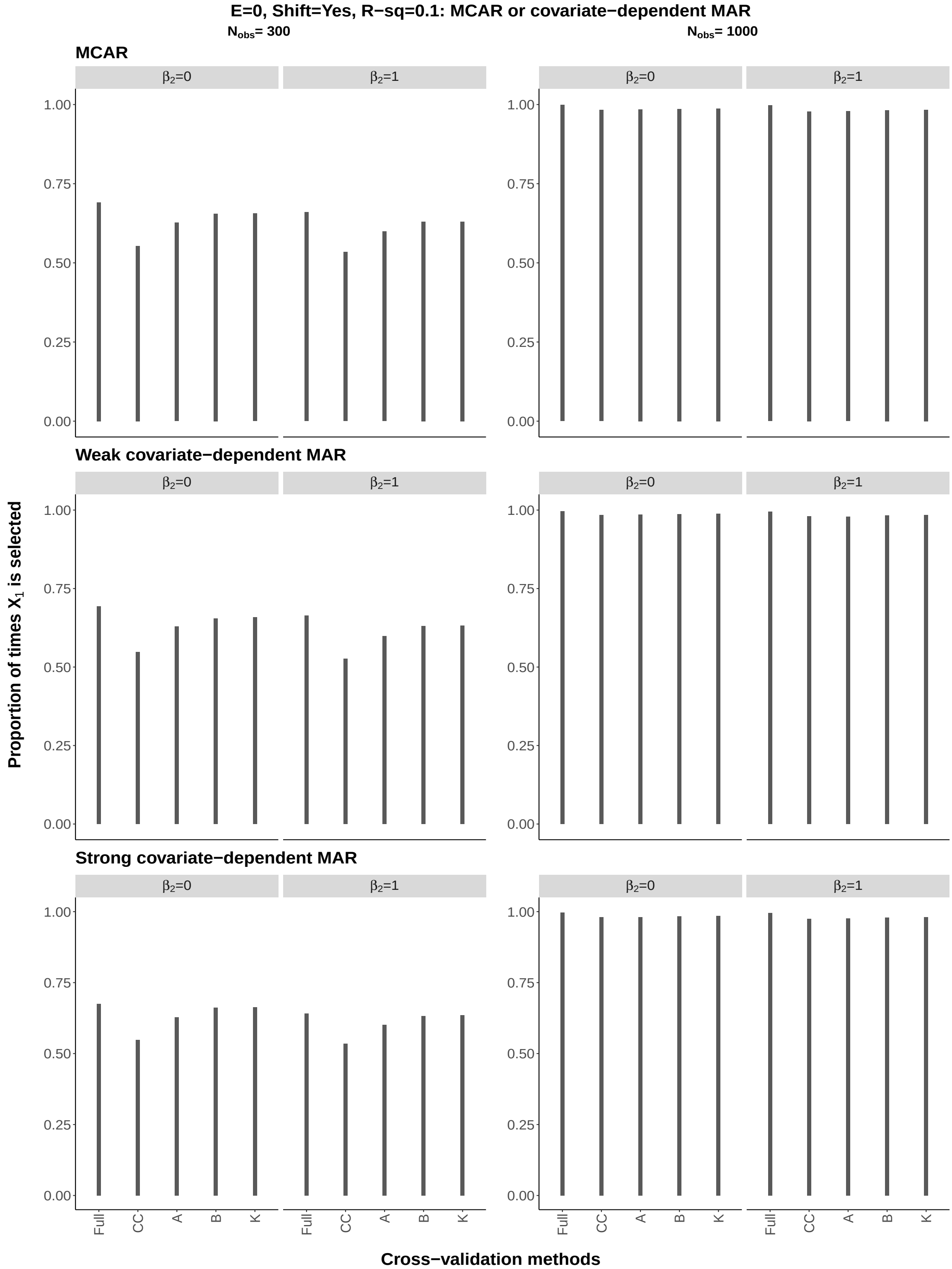

Figure S73: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

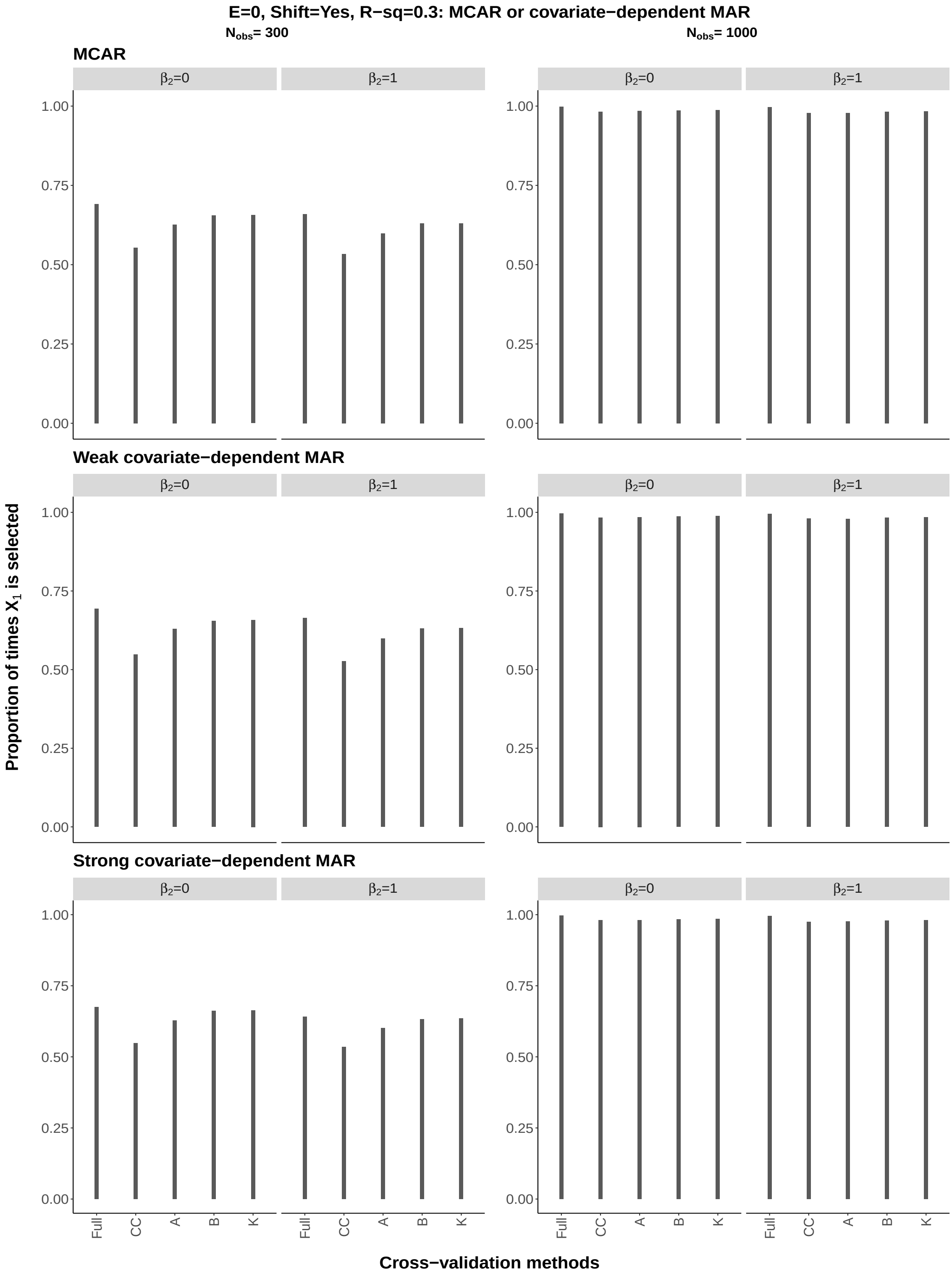

Figure S74: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

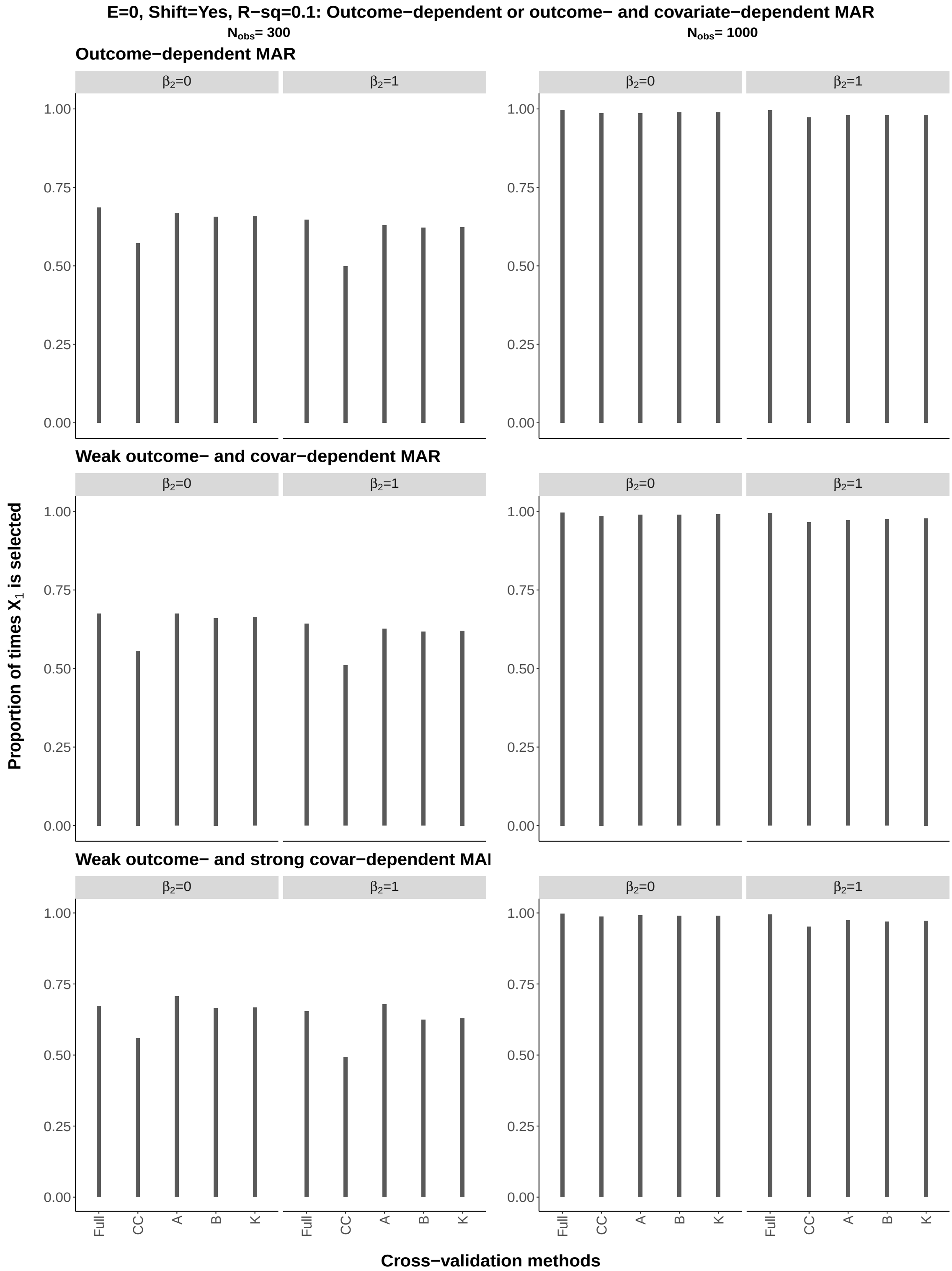

Figure S75: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

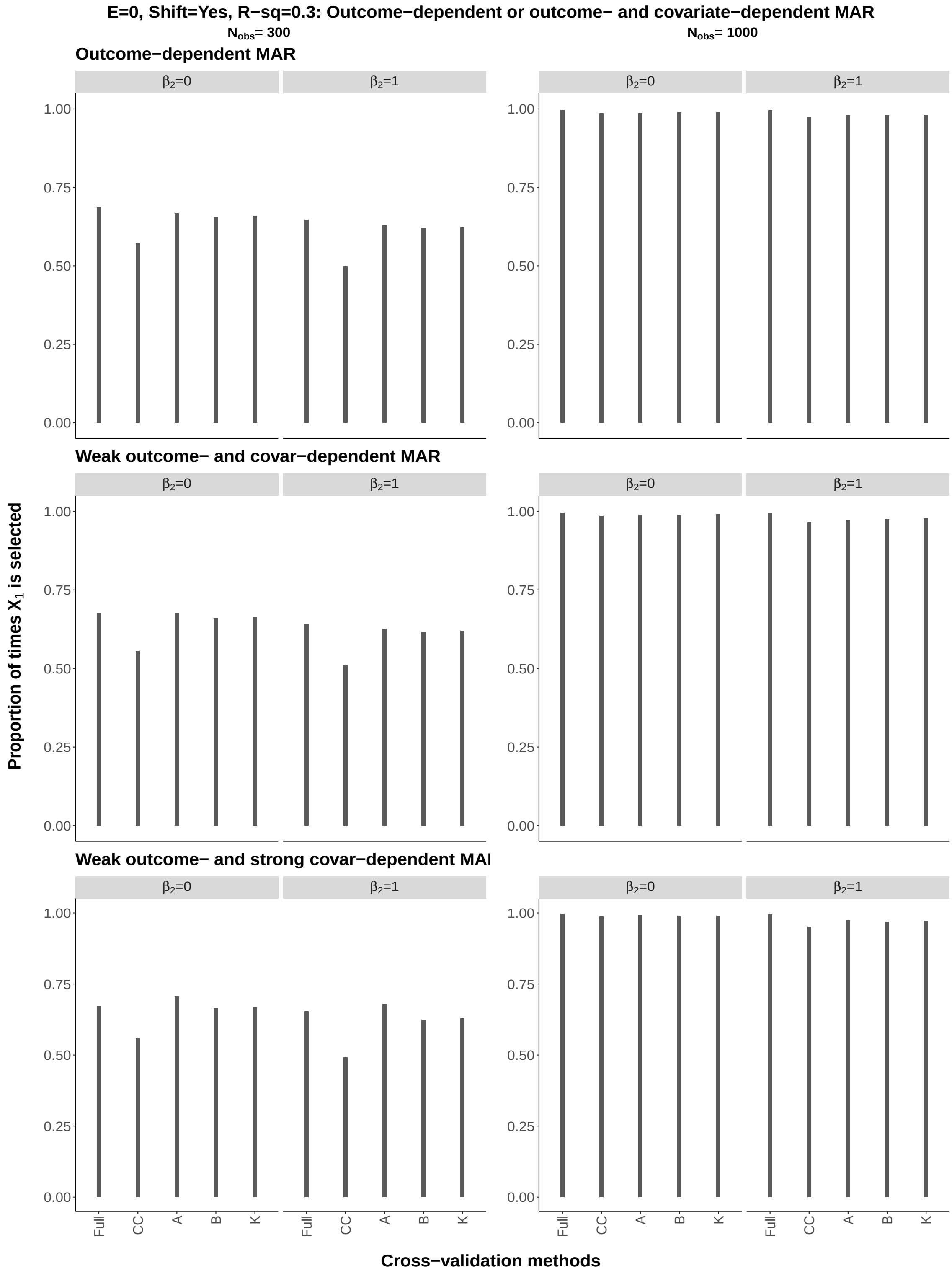

Figure S76: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

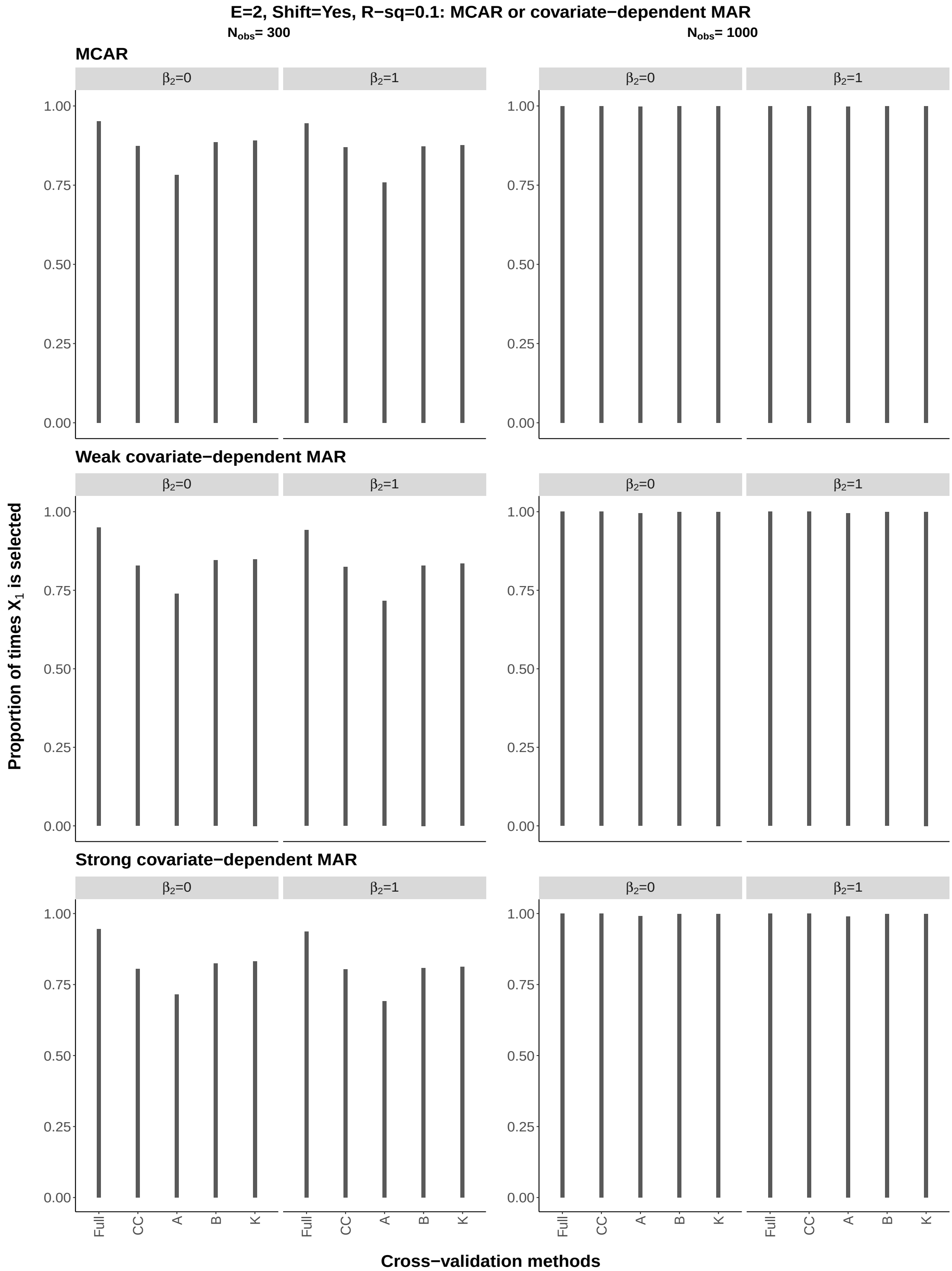

Figure S77: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

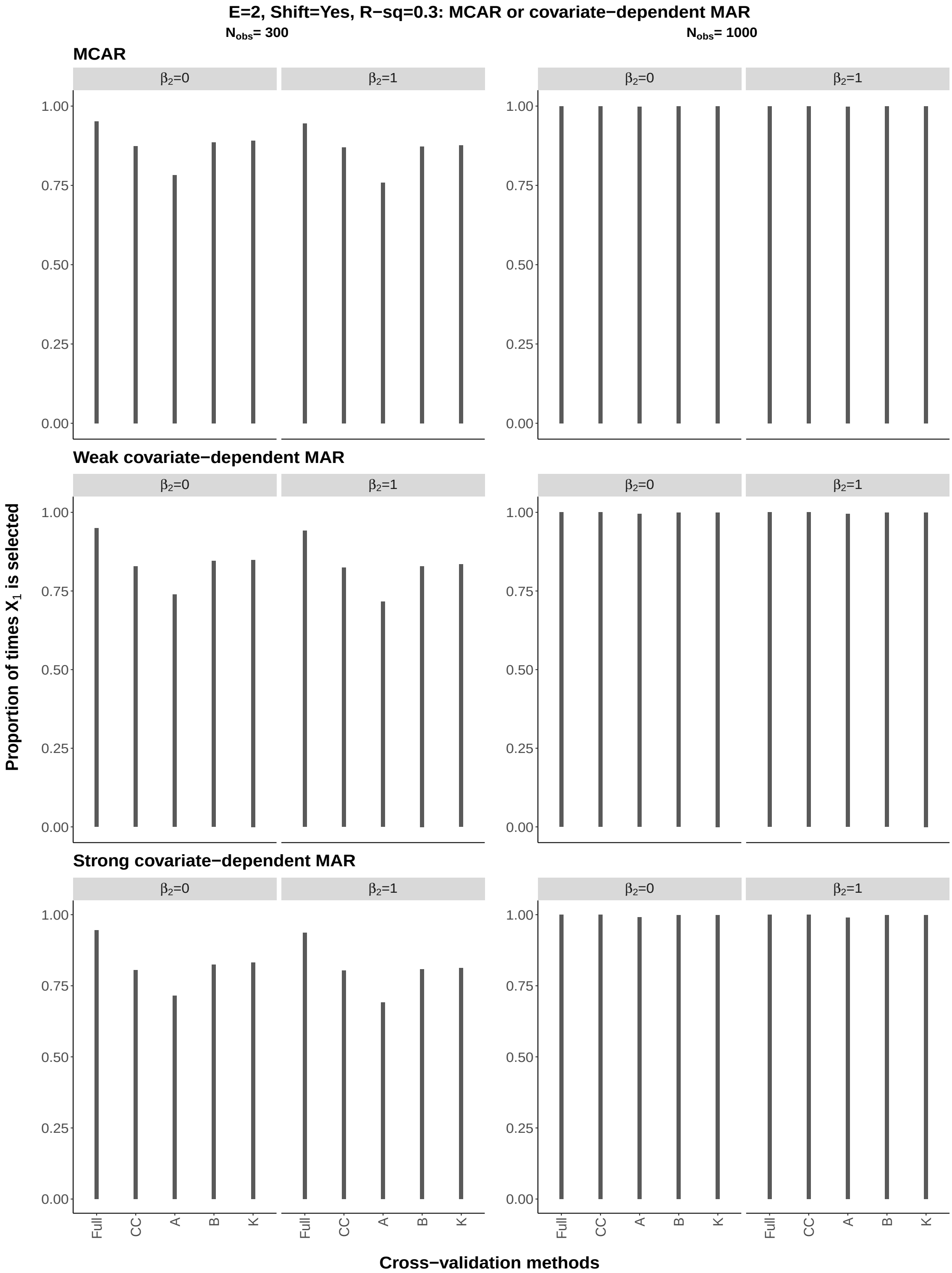

Figure S78: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

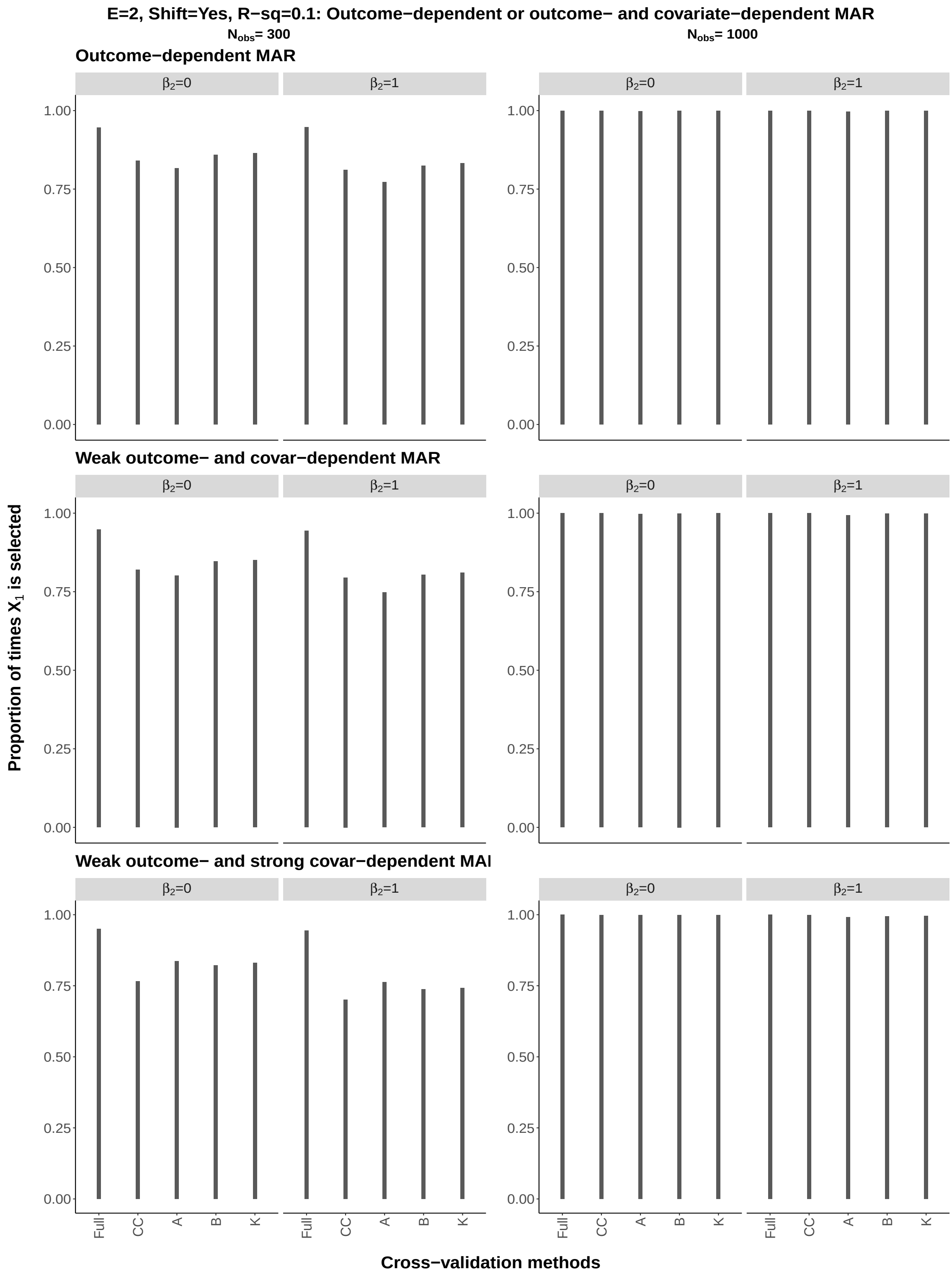

Figure S79: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

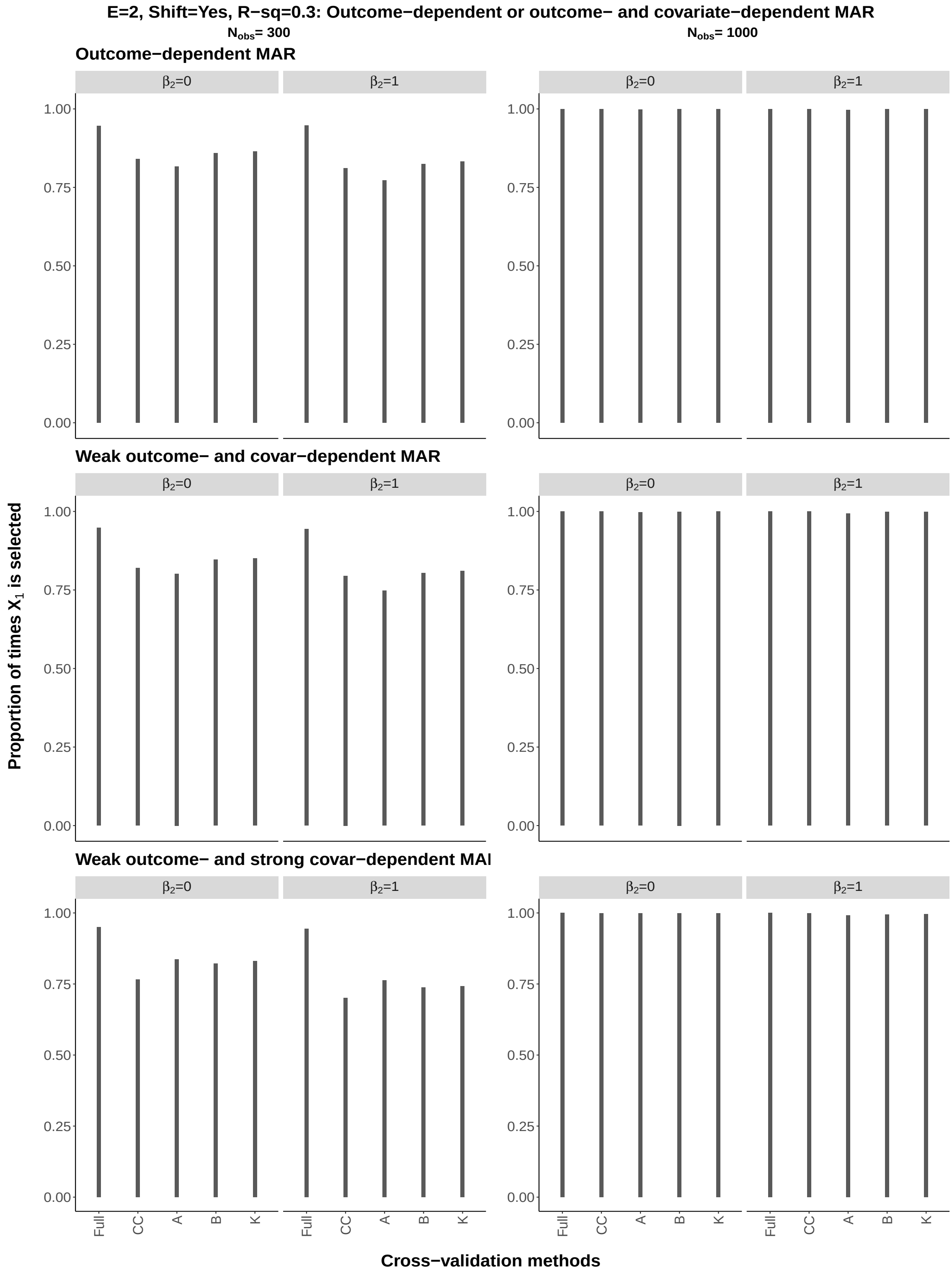

Figure S80: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

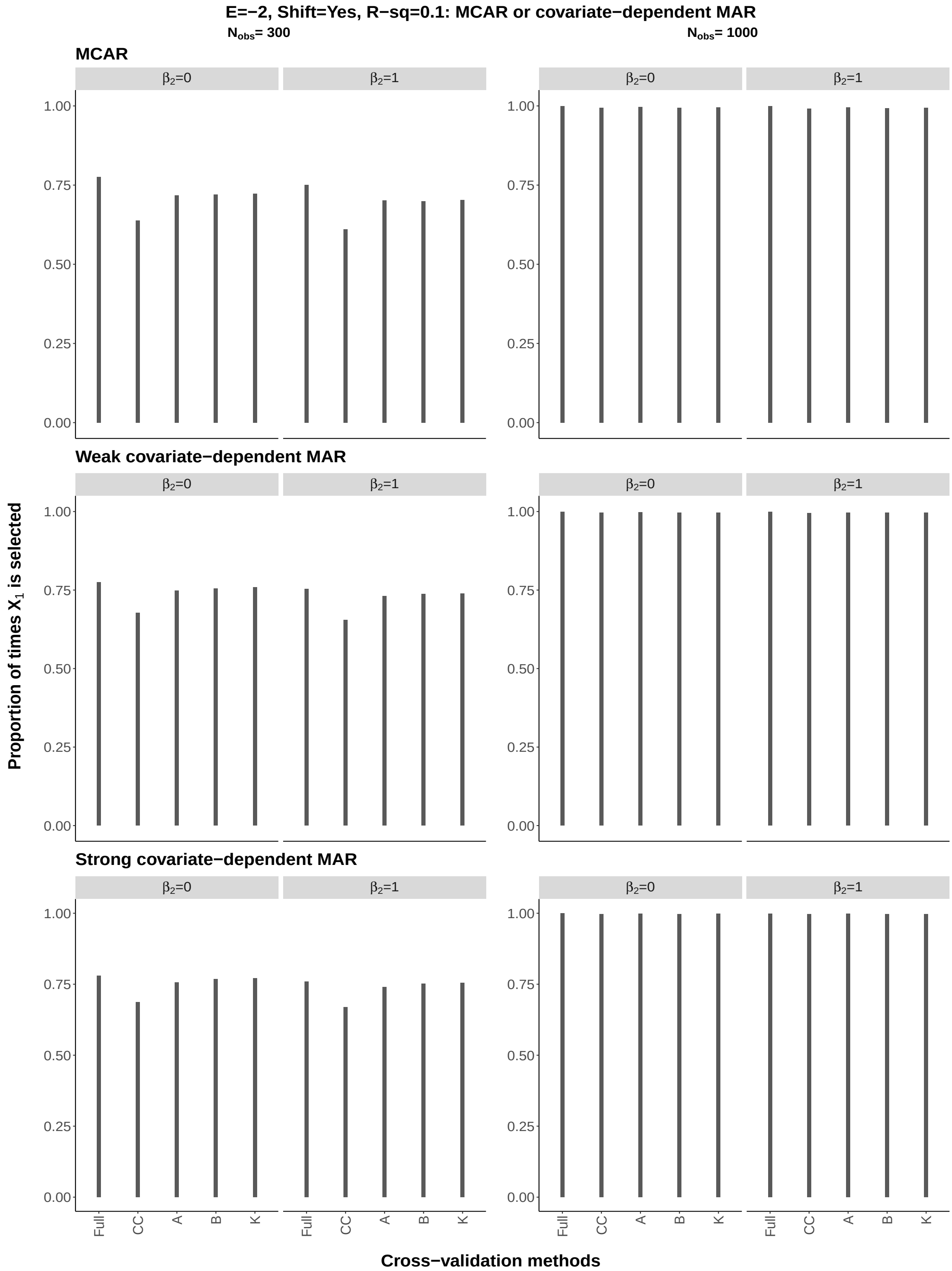

Figure S81: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

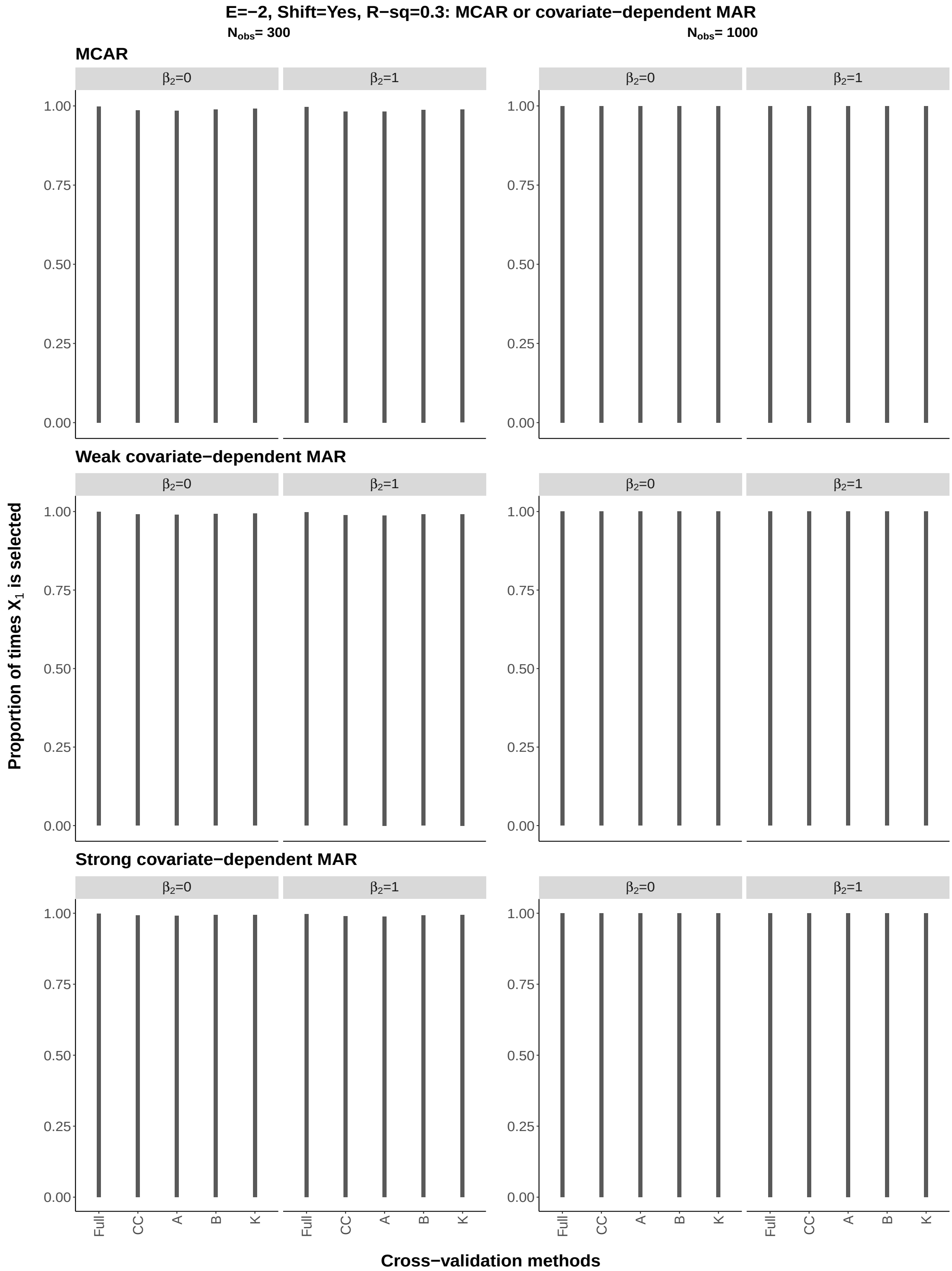

Figure S82: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

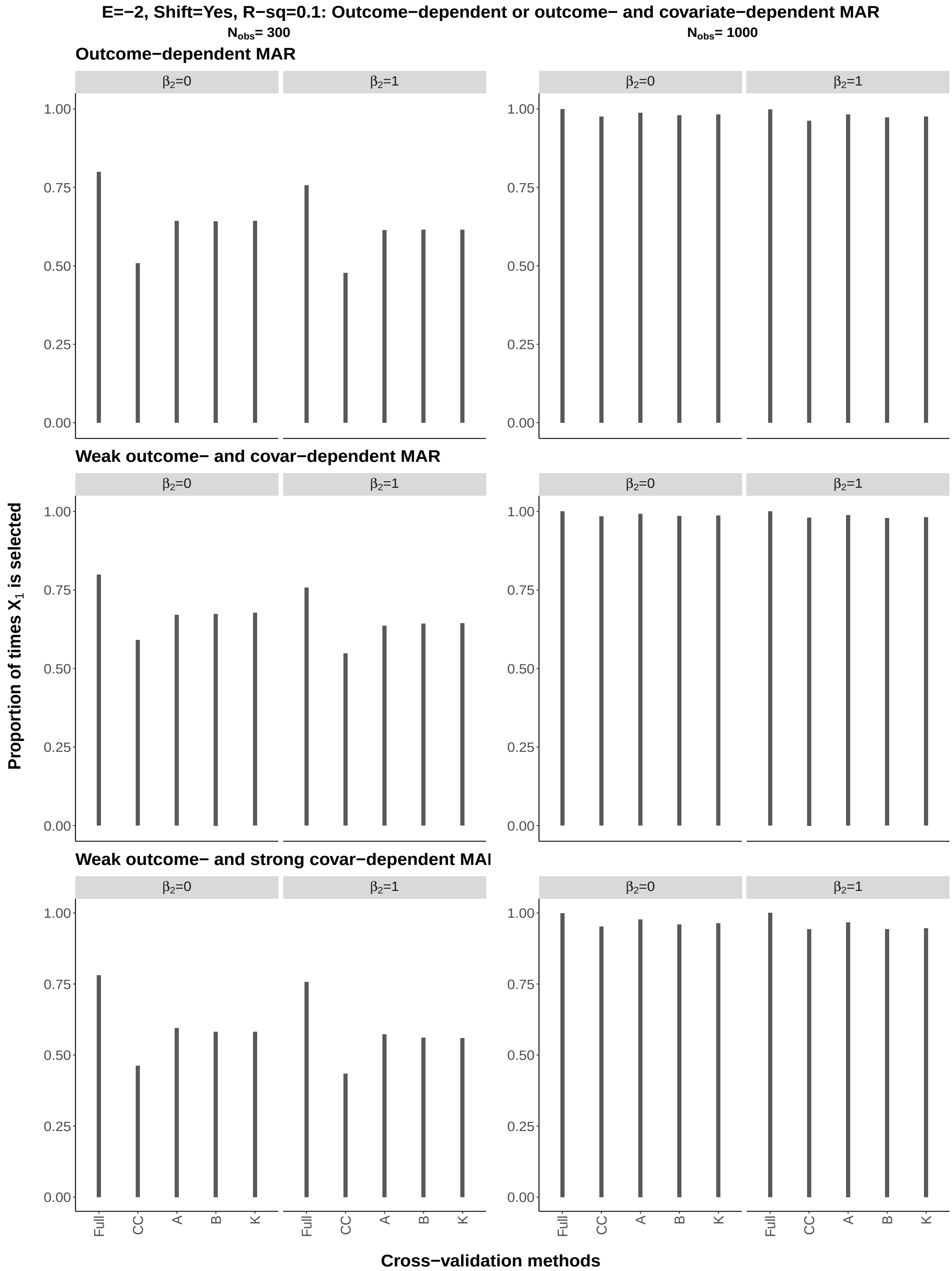

Figure S83: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

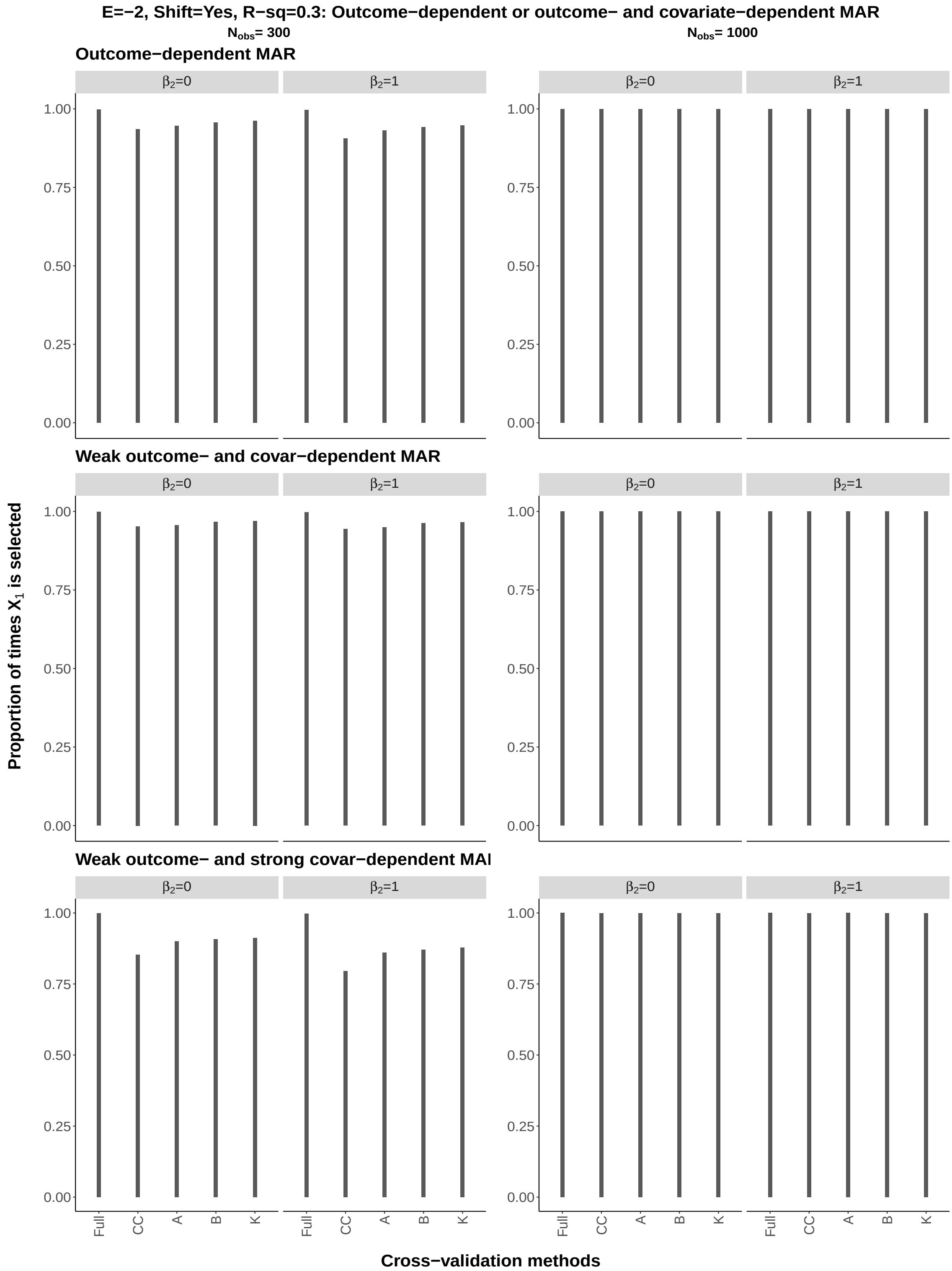

Figure S84: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.1.8 Covariate selection of X_1: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

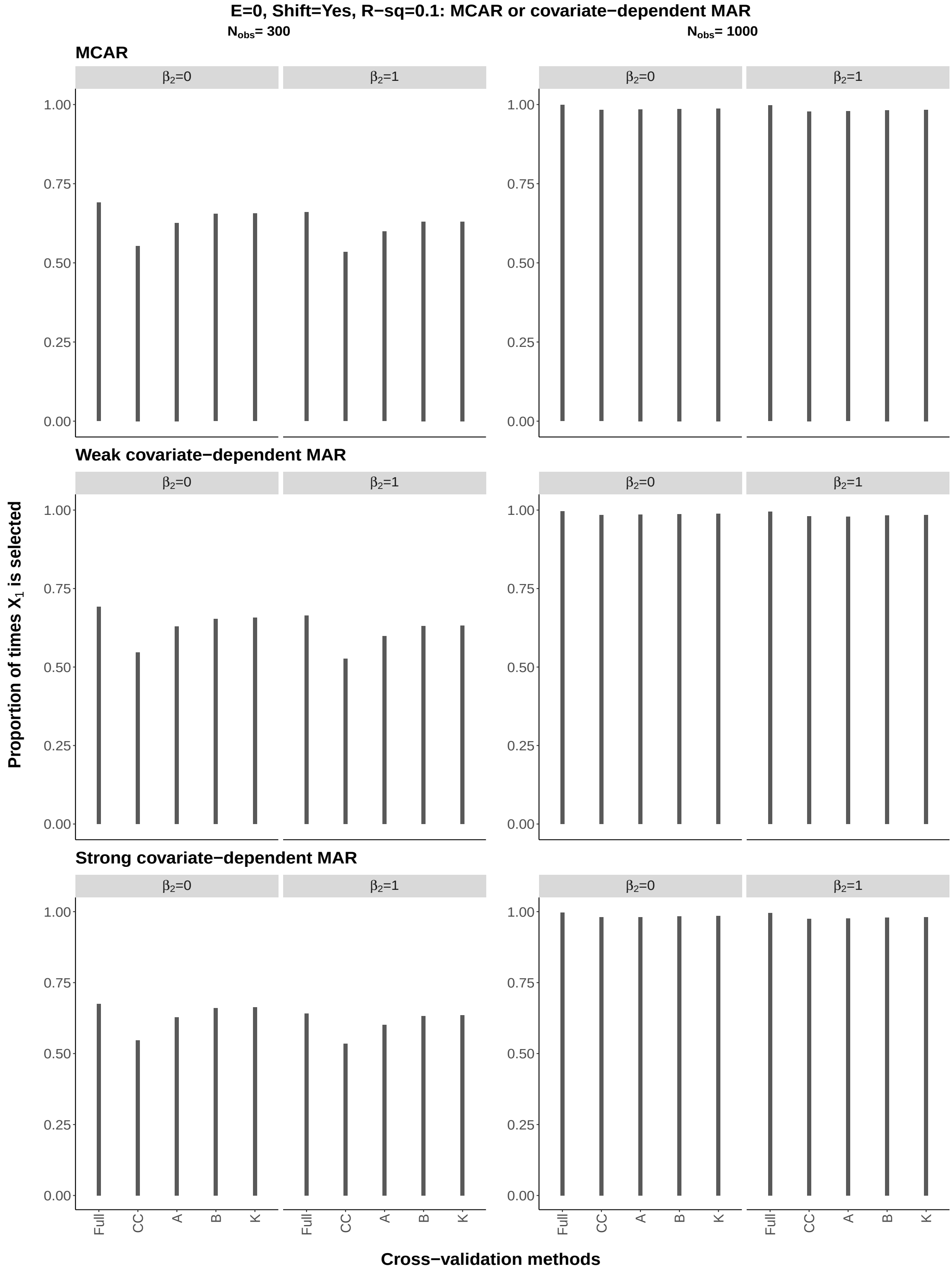

Figure S85: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

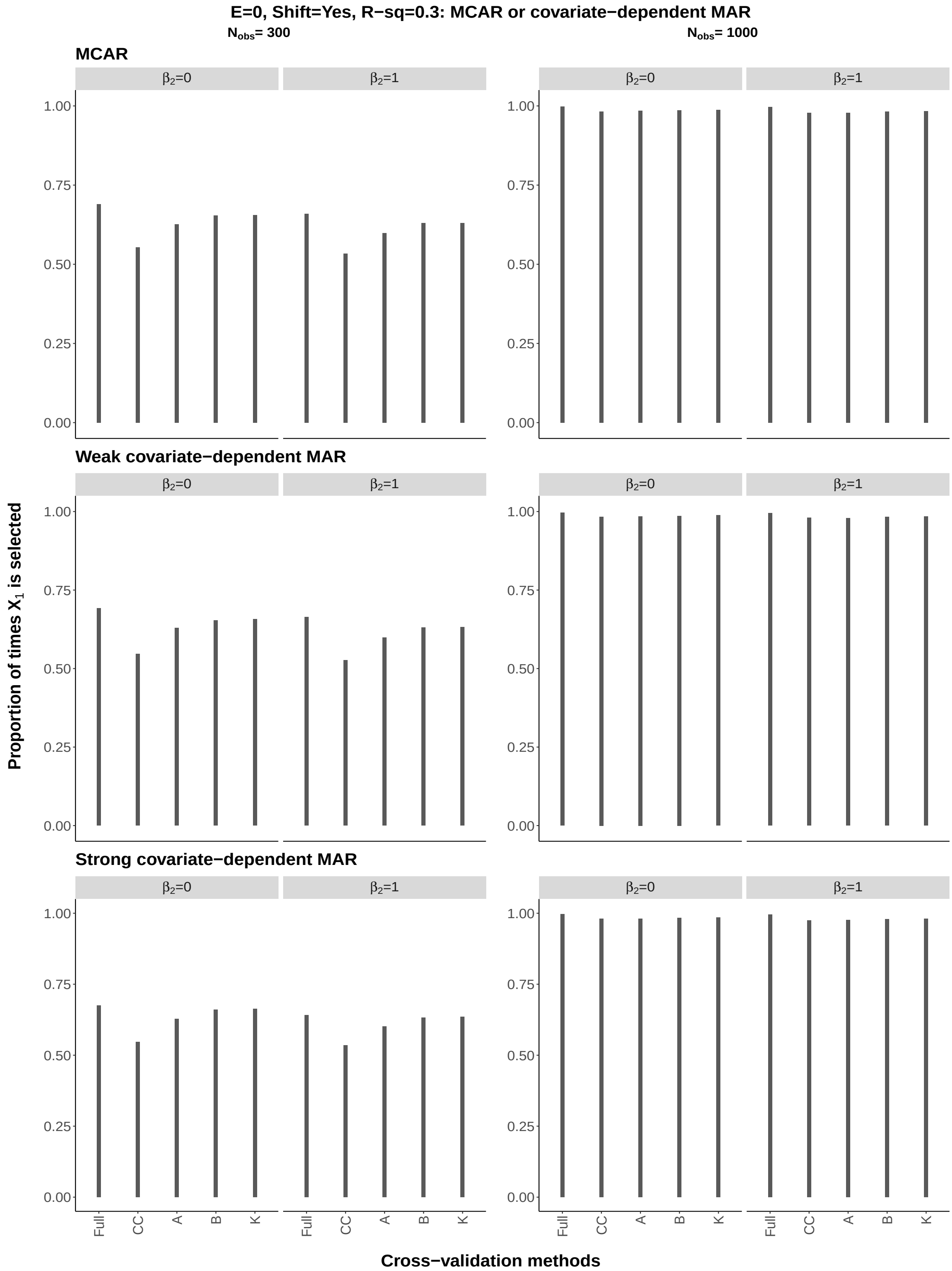

Figure S86: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

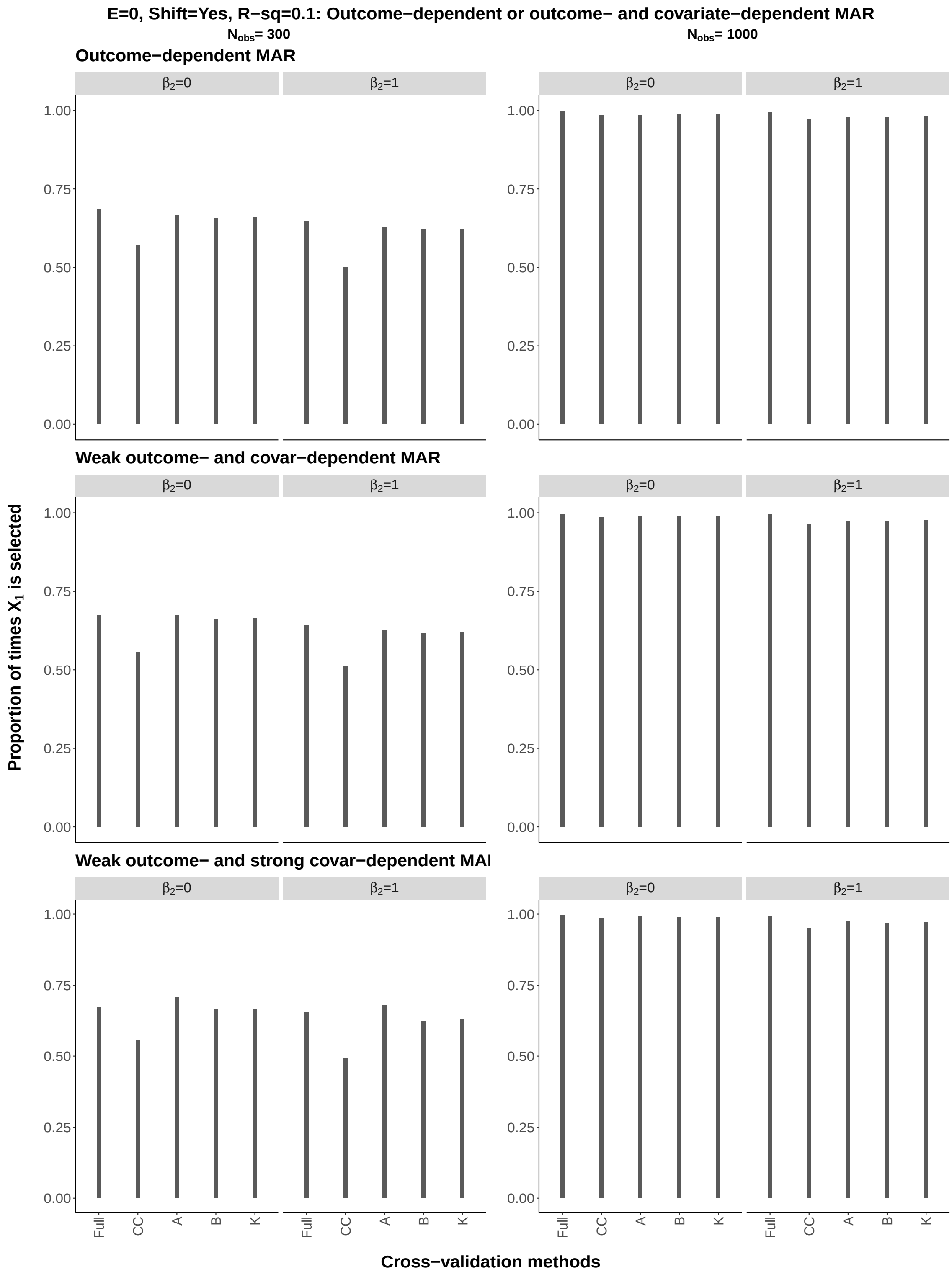

Figure S87: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

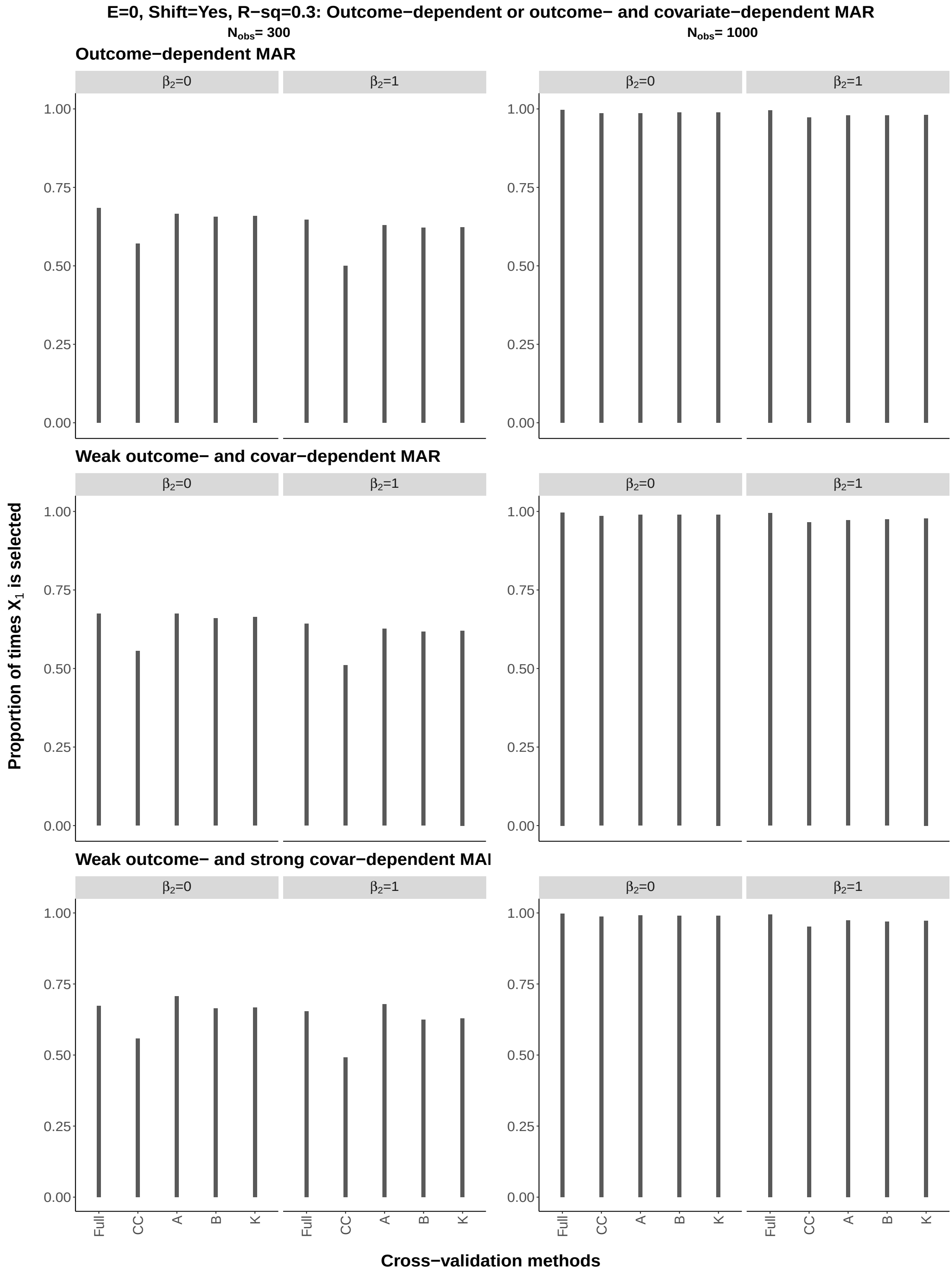

Figure S88: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

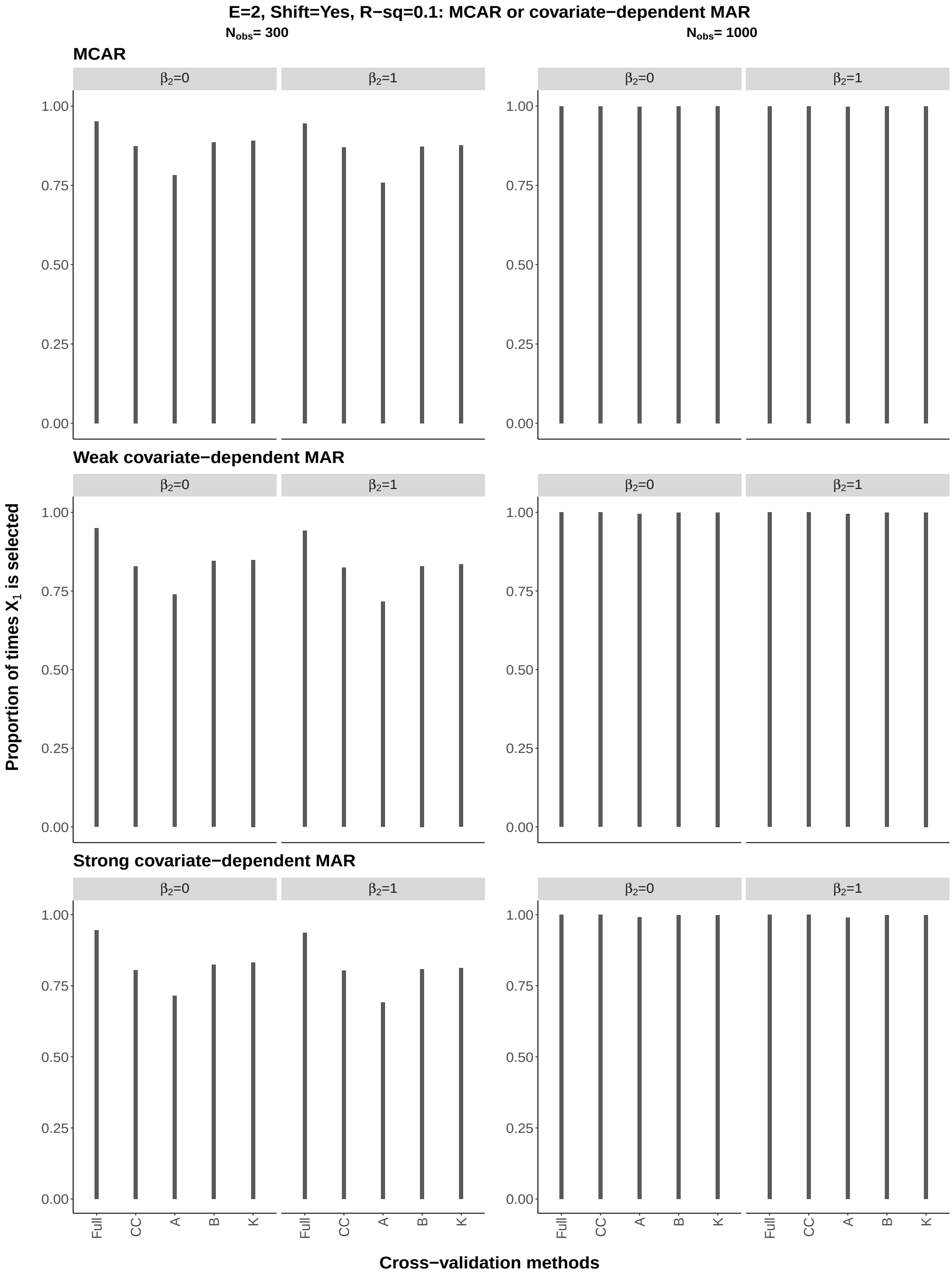

Figure S89: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

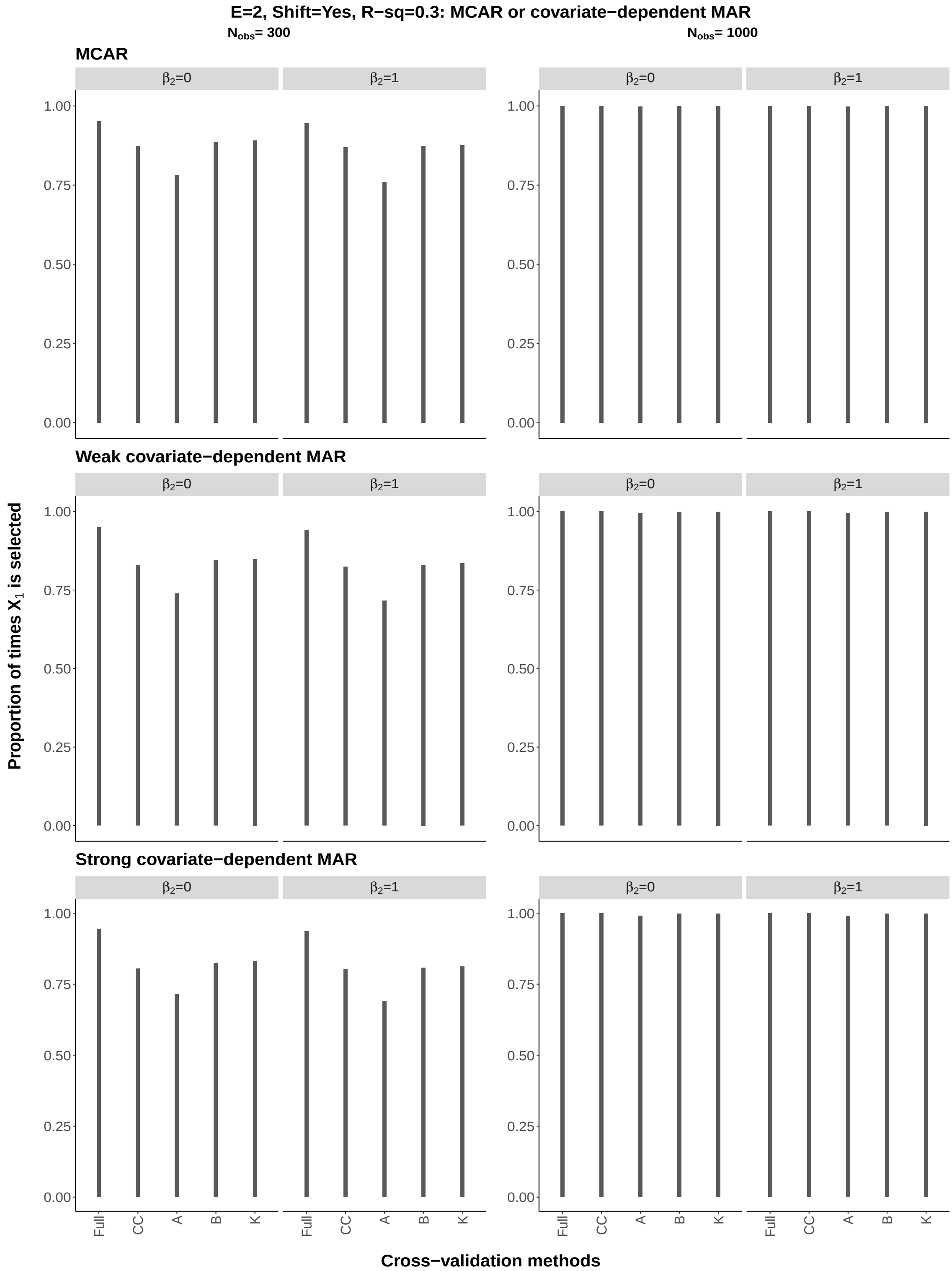

Figure S90: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

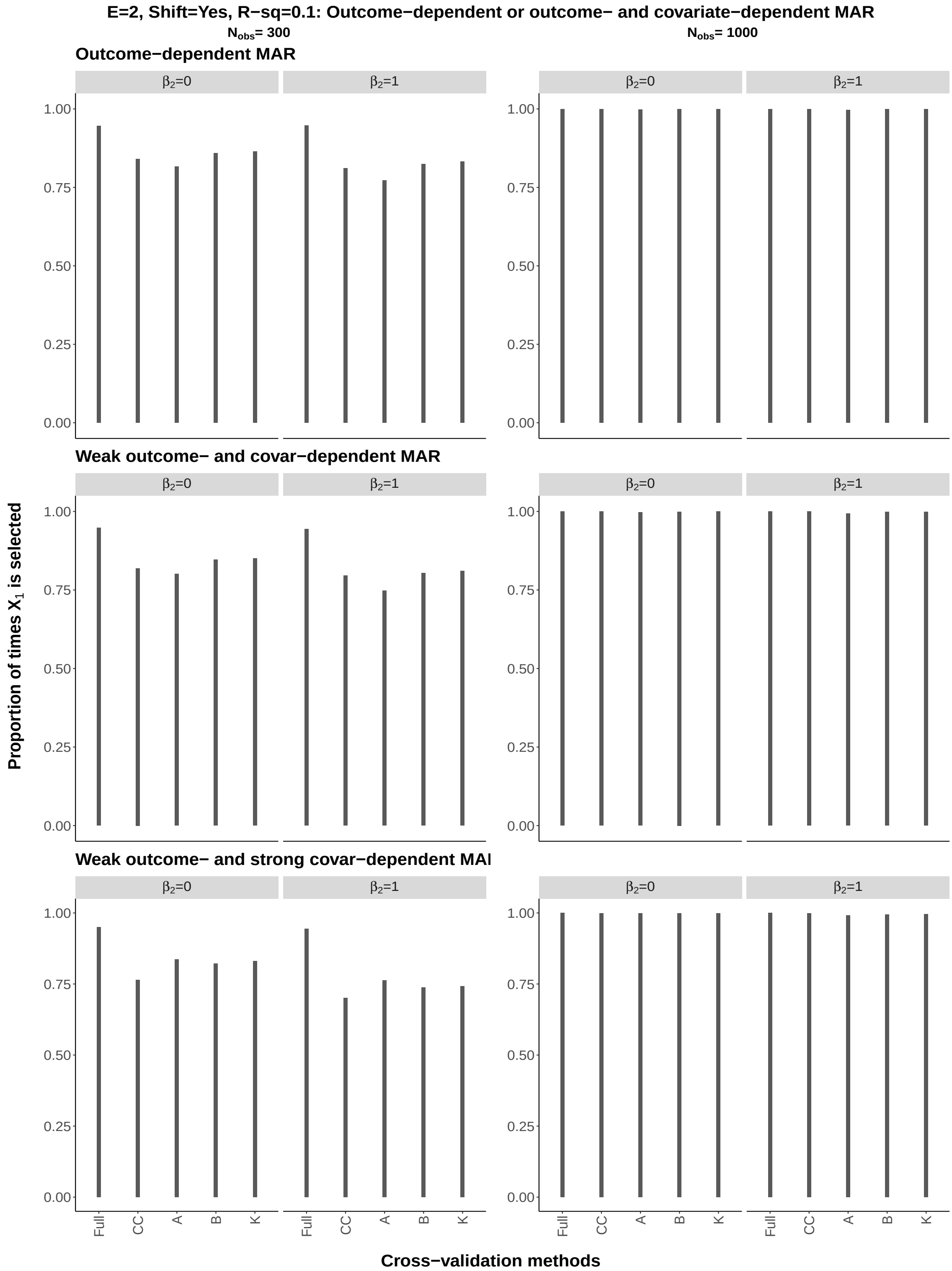

Figure S91: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

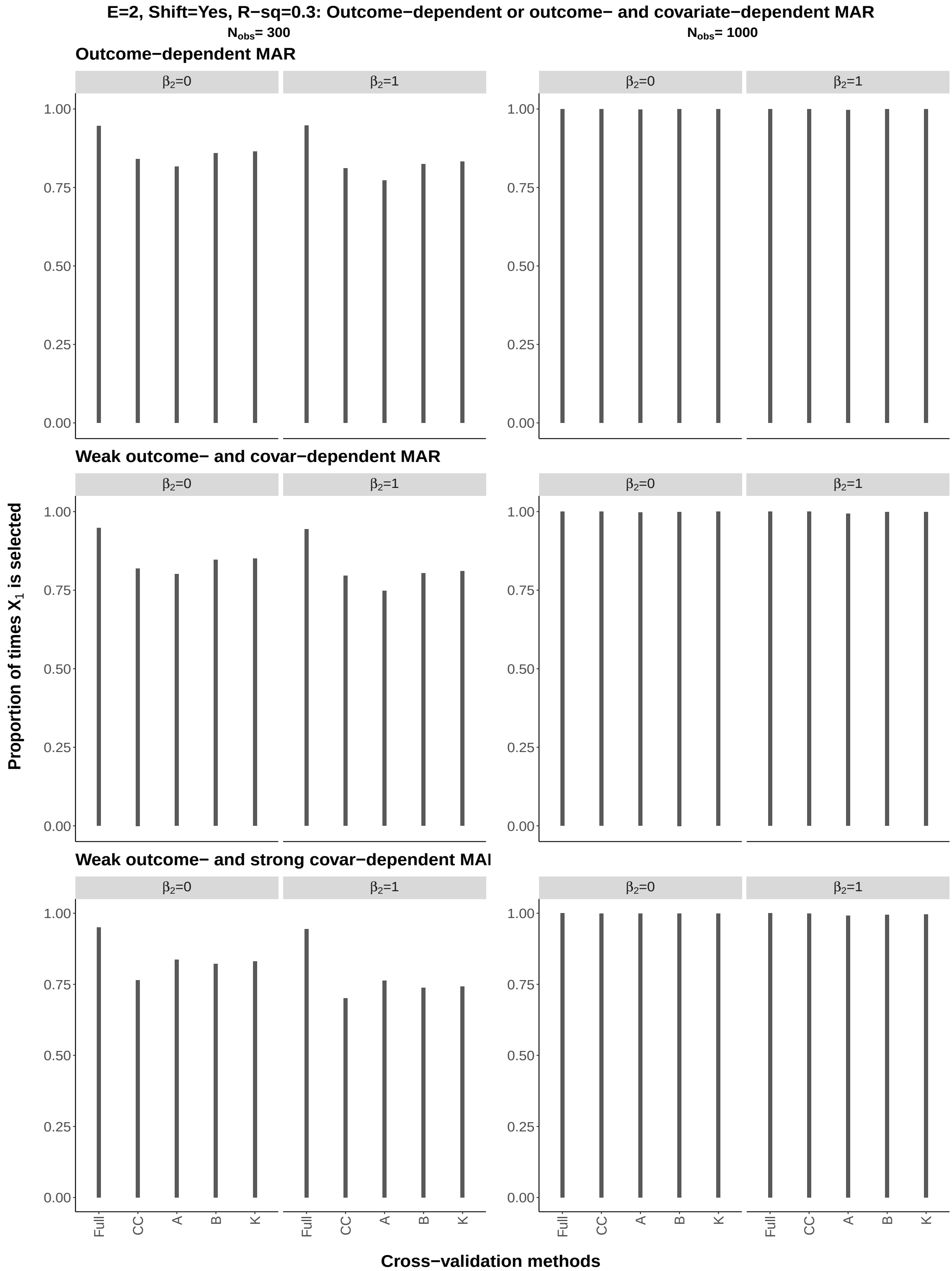

Figure S92: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

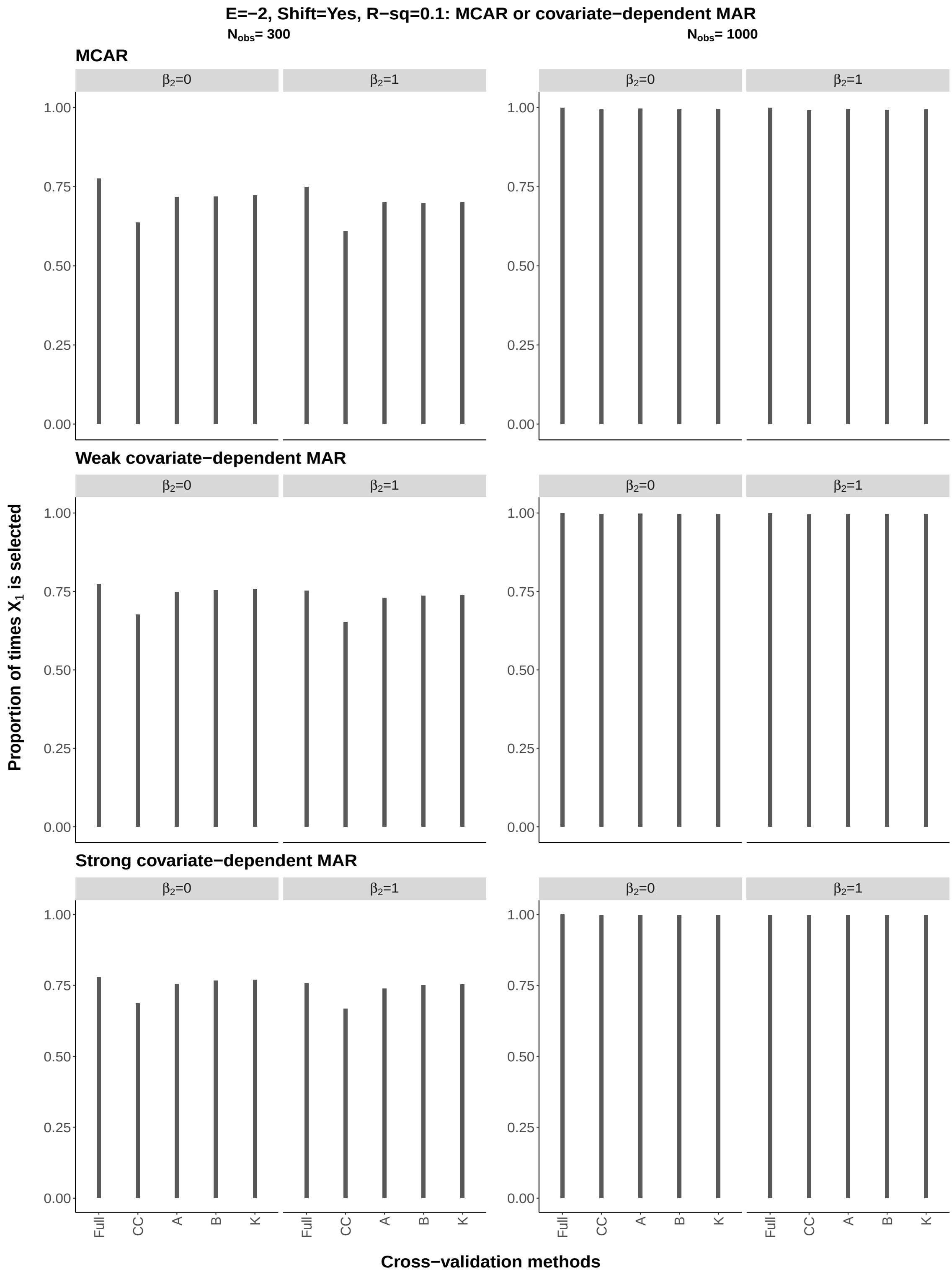

Figure S93: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

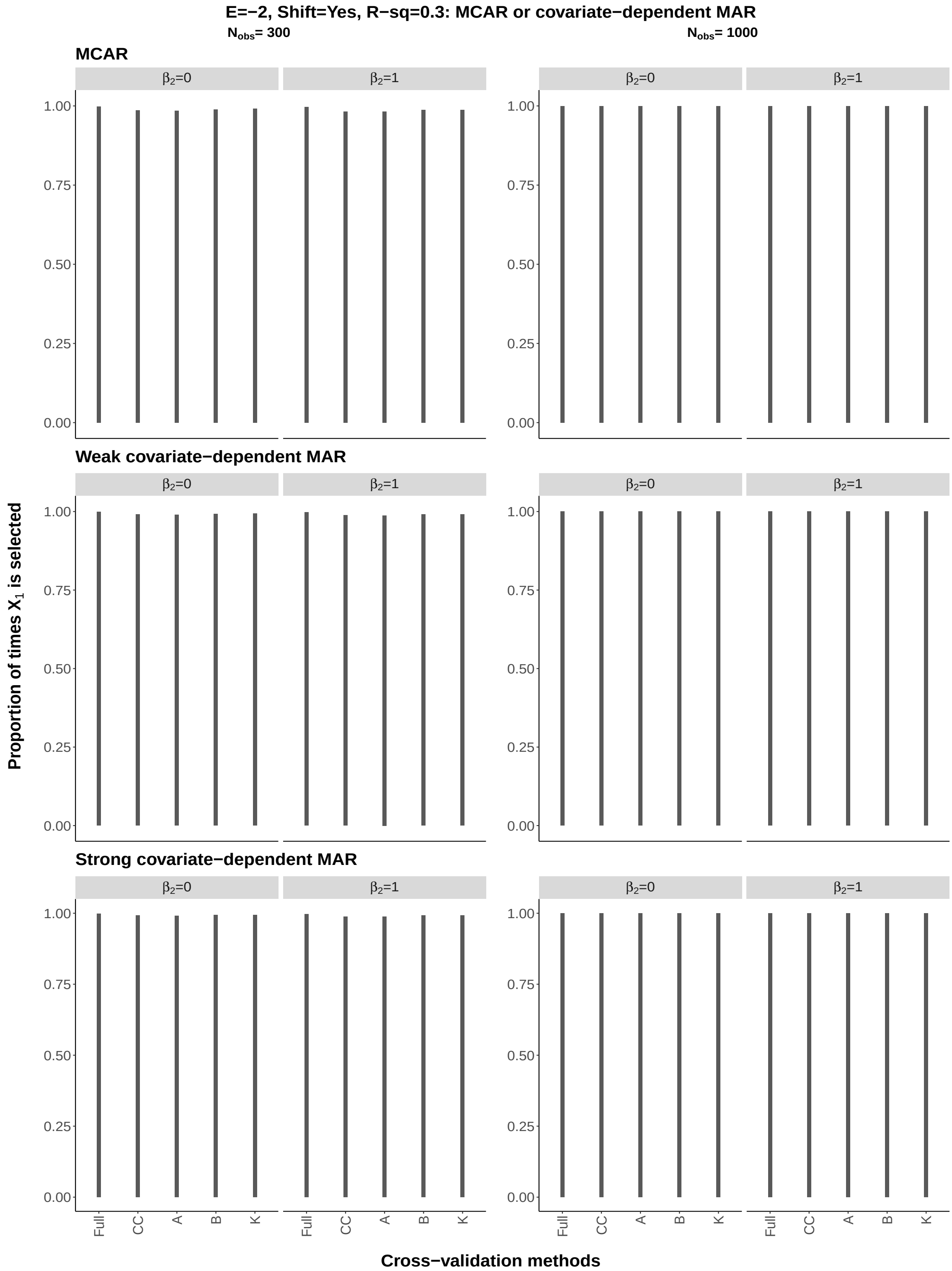

Figure S94: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

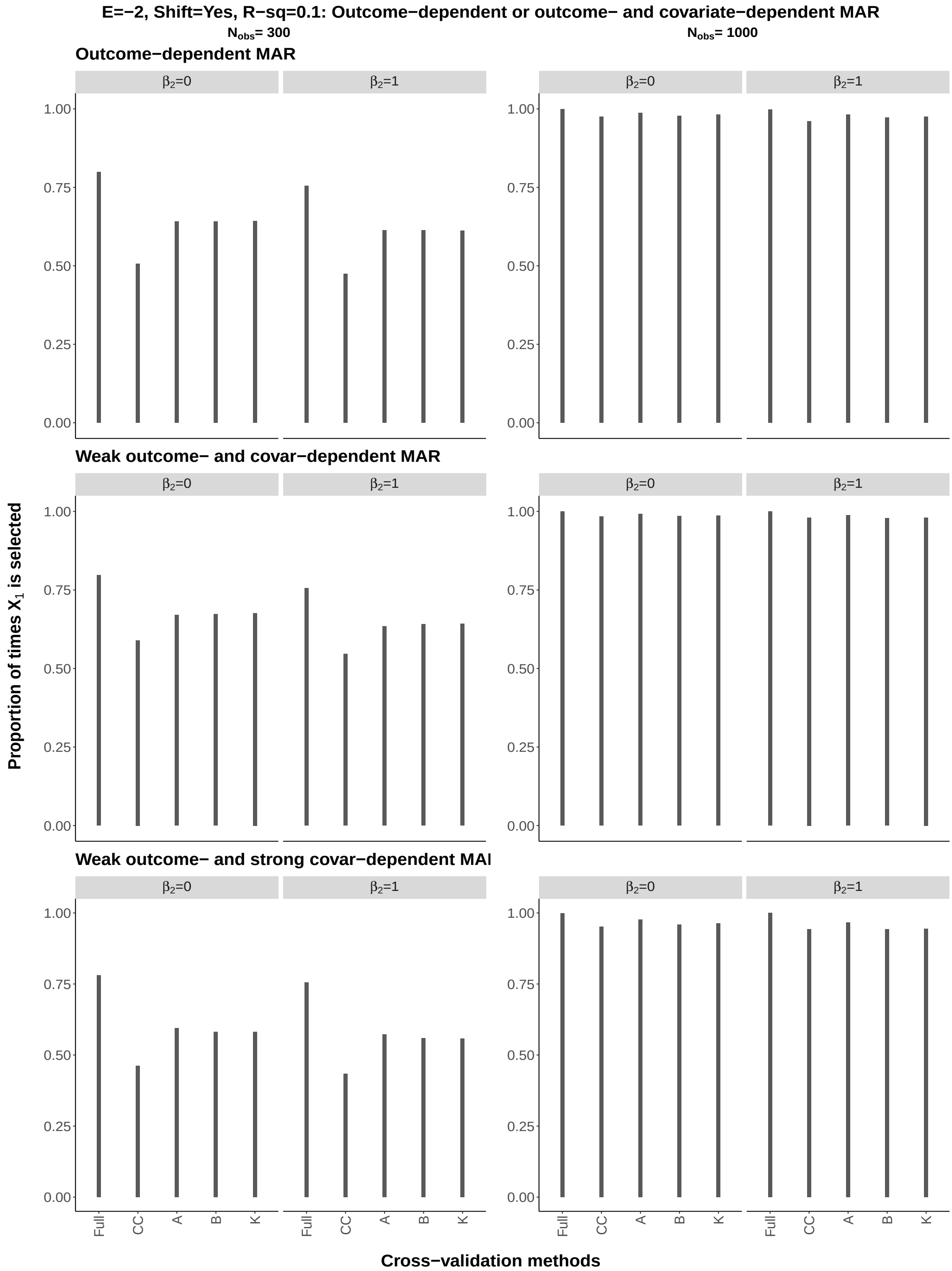

Figure S95: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

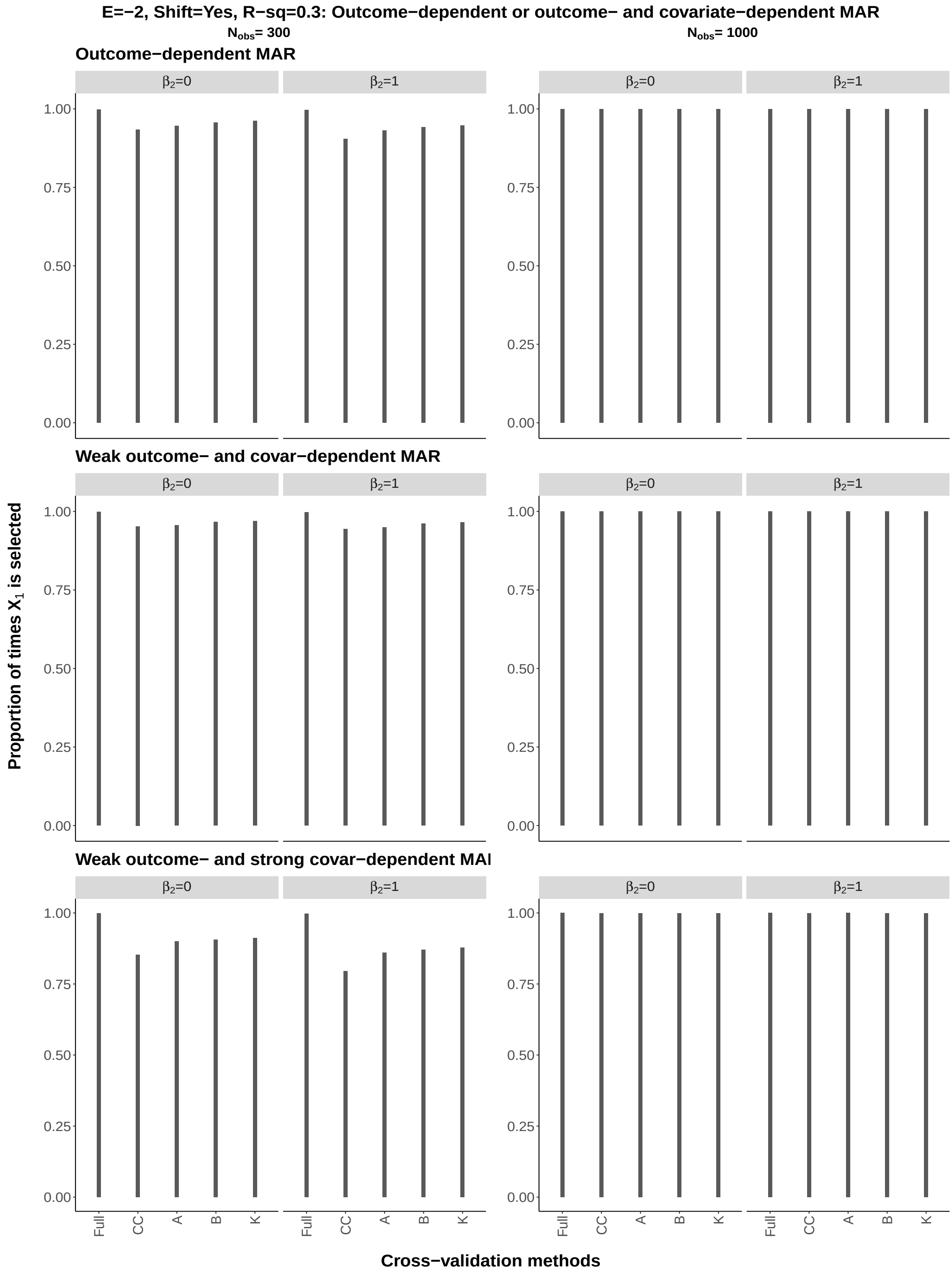

Figure S96: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, K folds and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2 The 0.632 bootstrap

S9.2.1 Covariate selection of X_2 using all data: $\beta_2 = 1$, $\alpha_E = 1$, no origin-shift

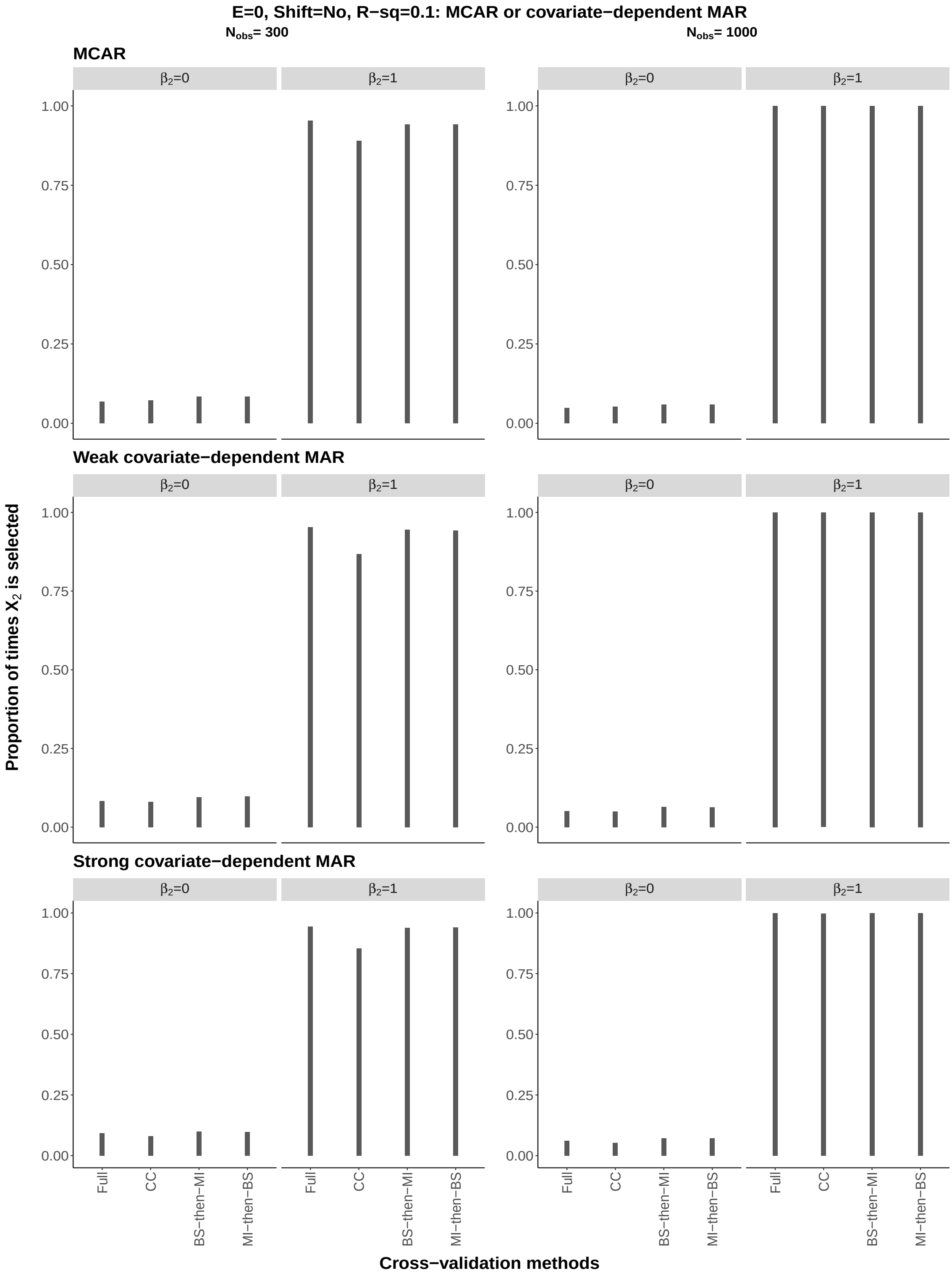

Figure S97: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

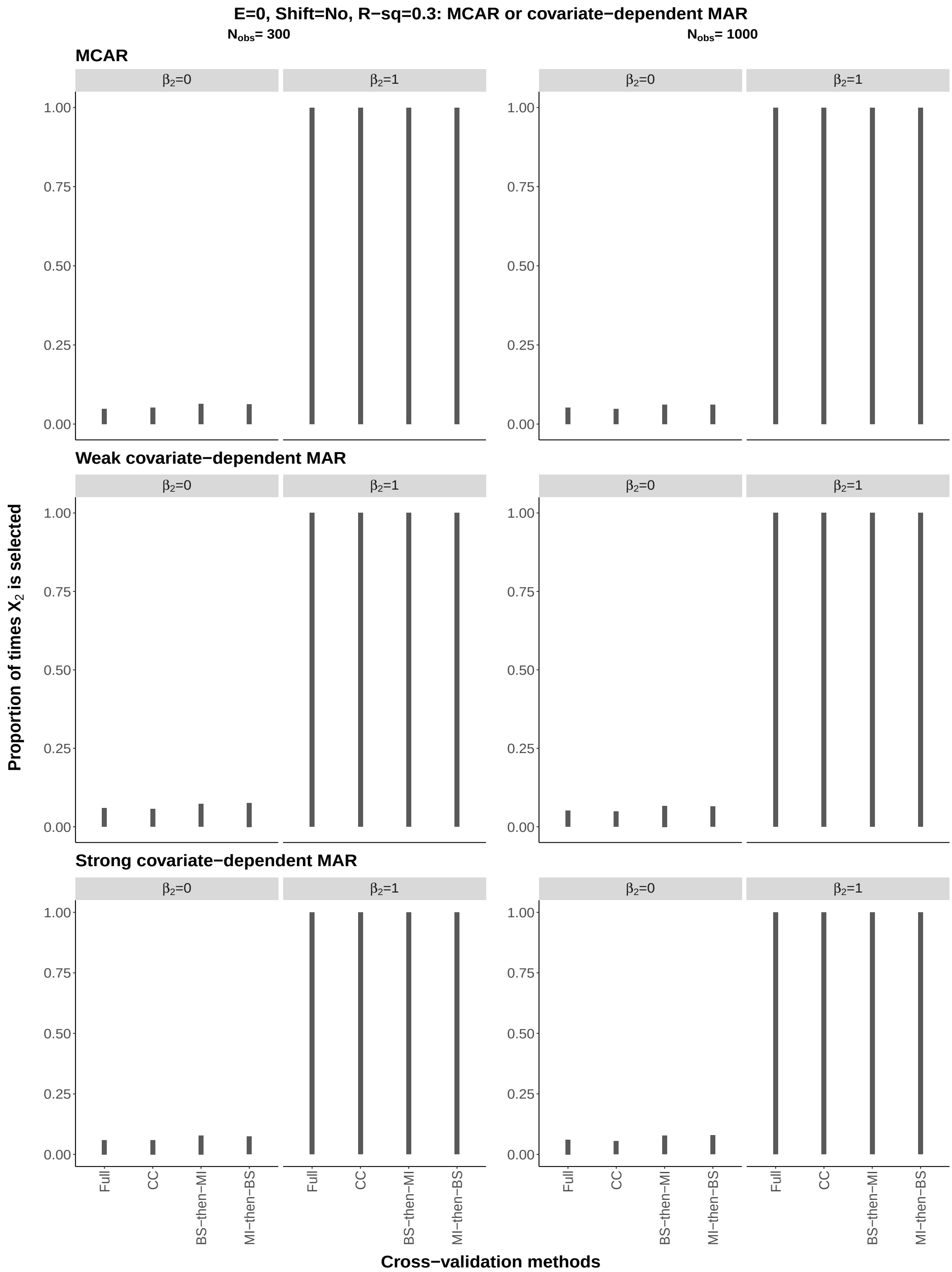

Figure S98: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

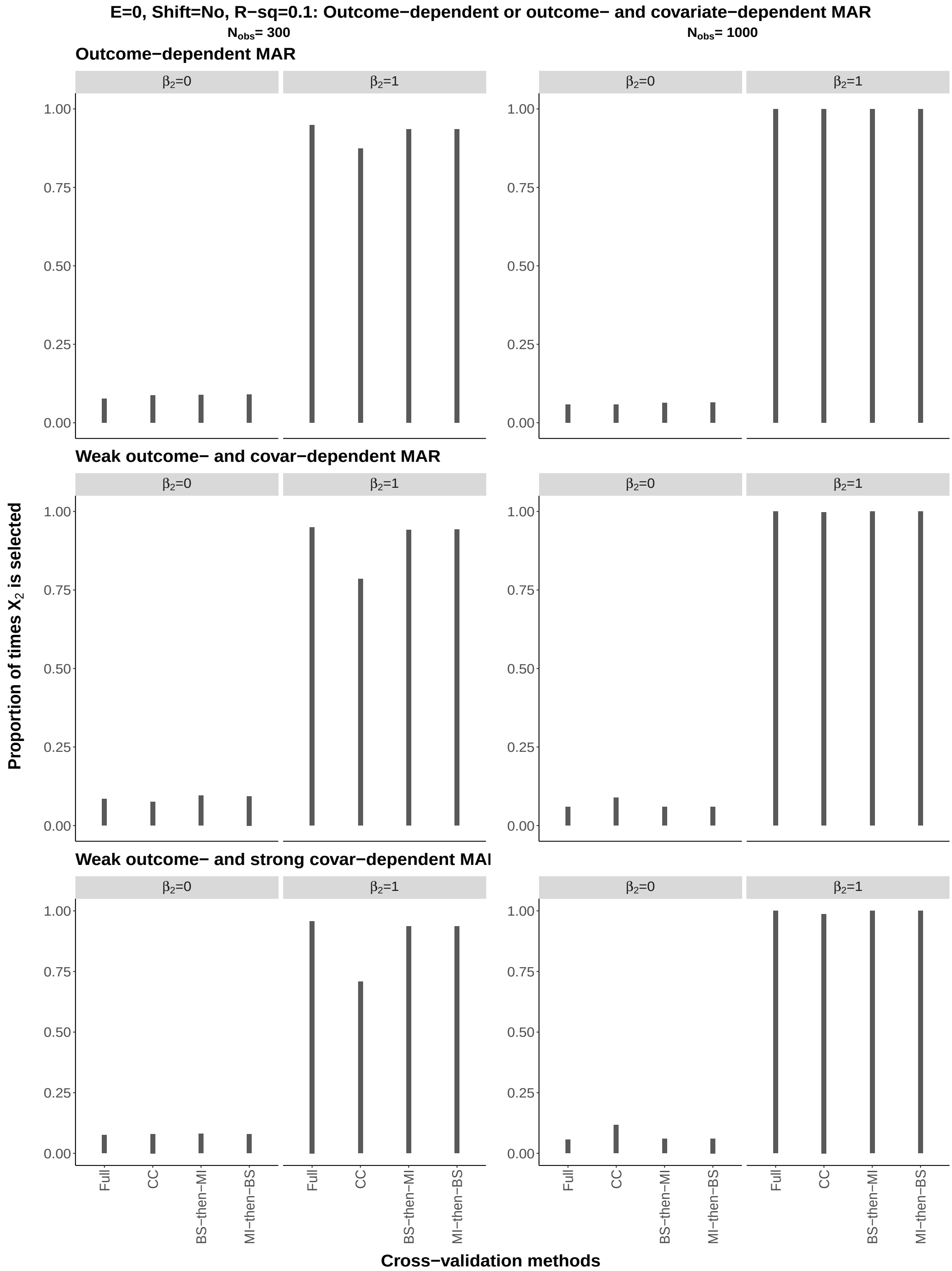

Figure S99: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

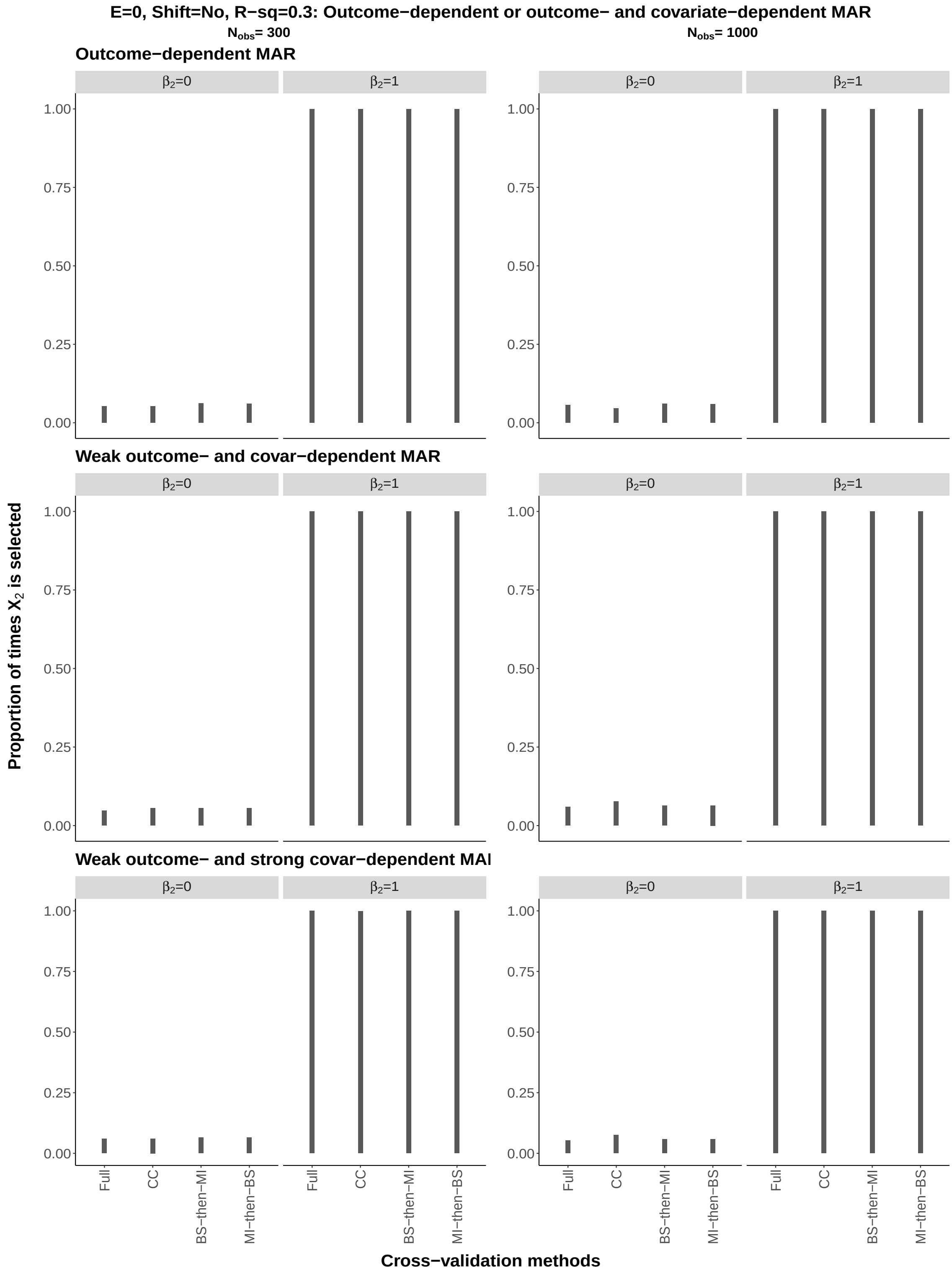

Figure S100: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

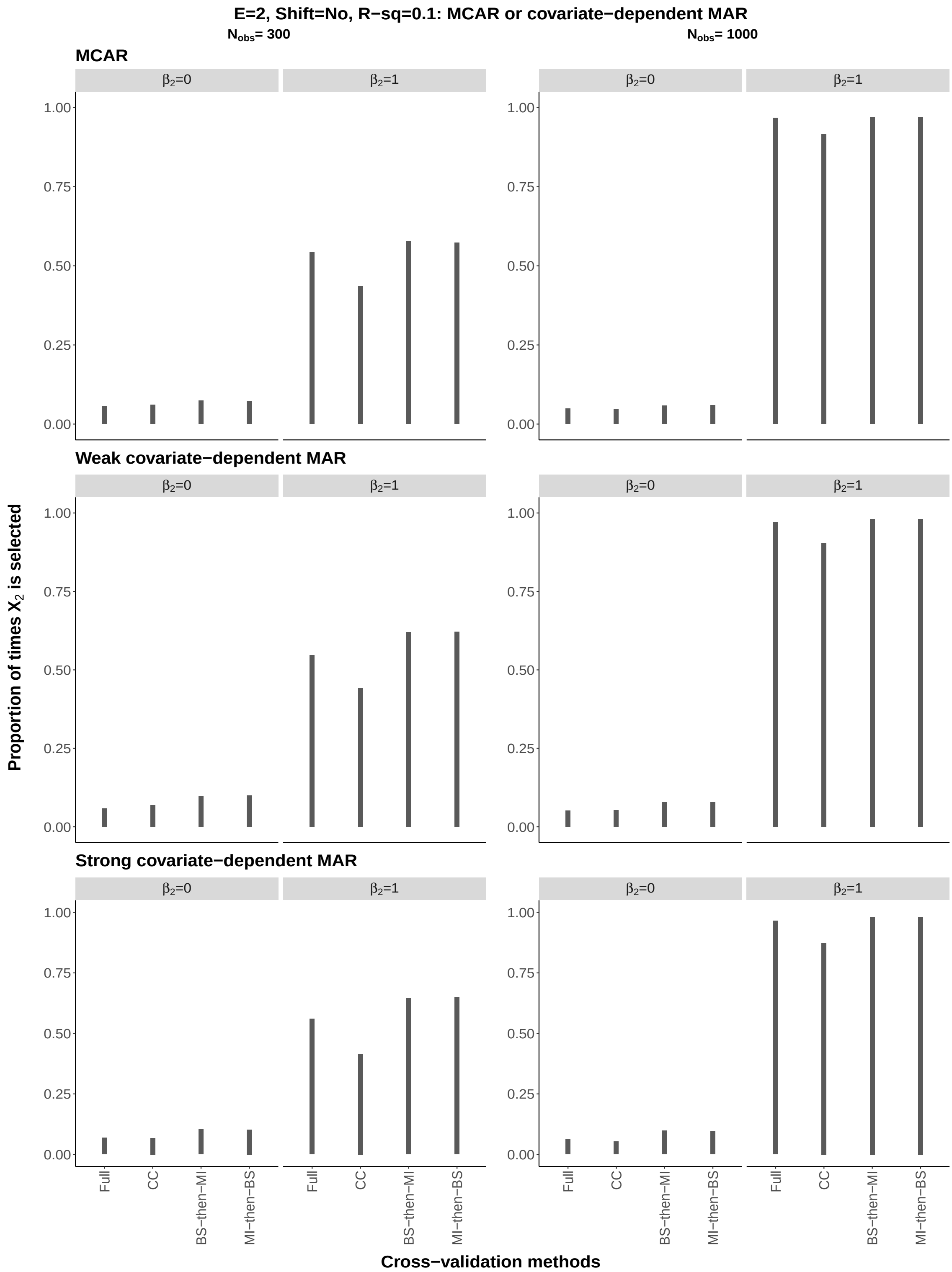

Figure S101: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

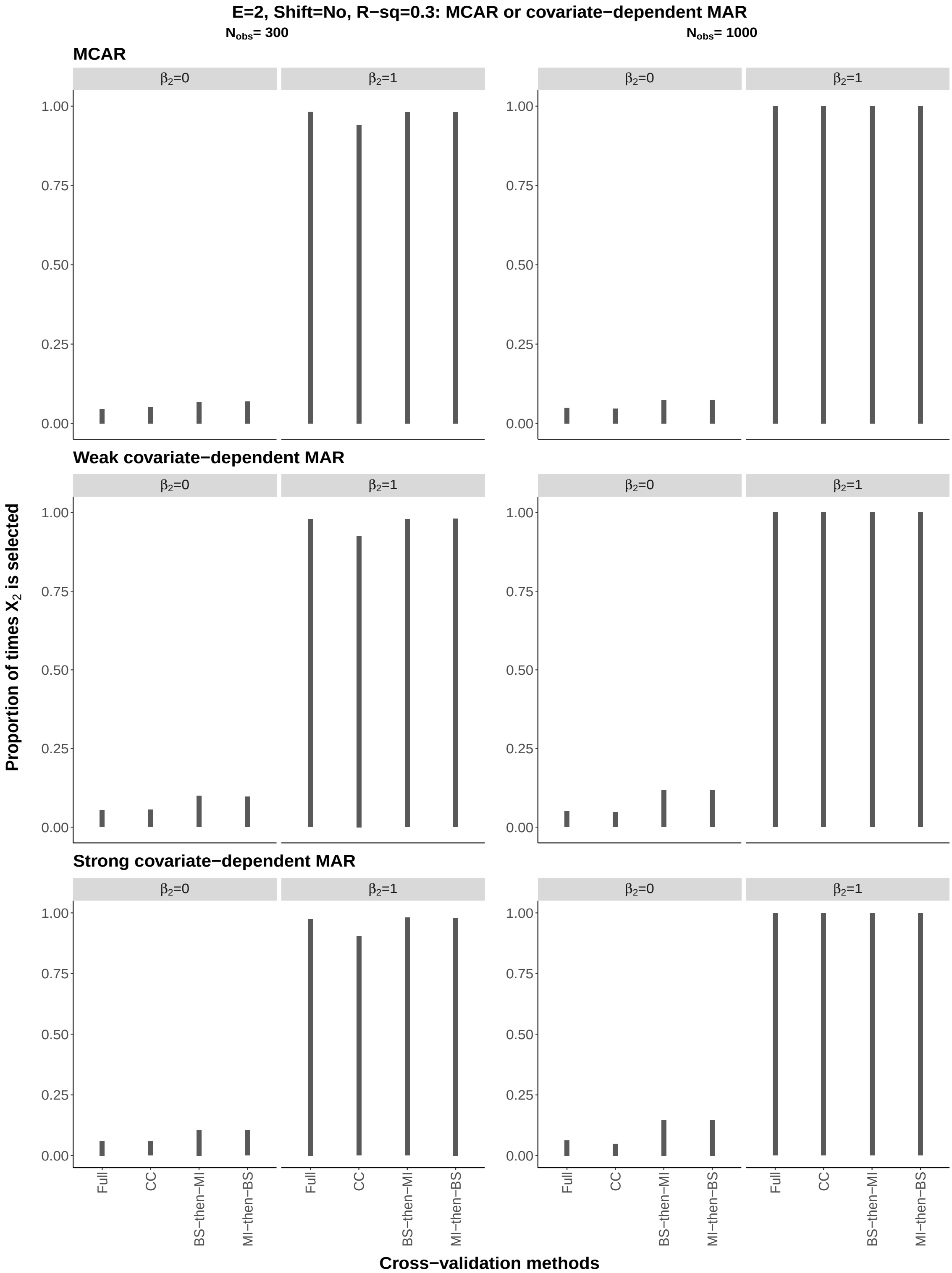

Figure S102: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

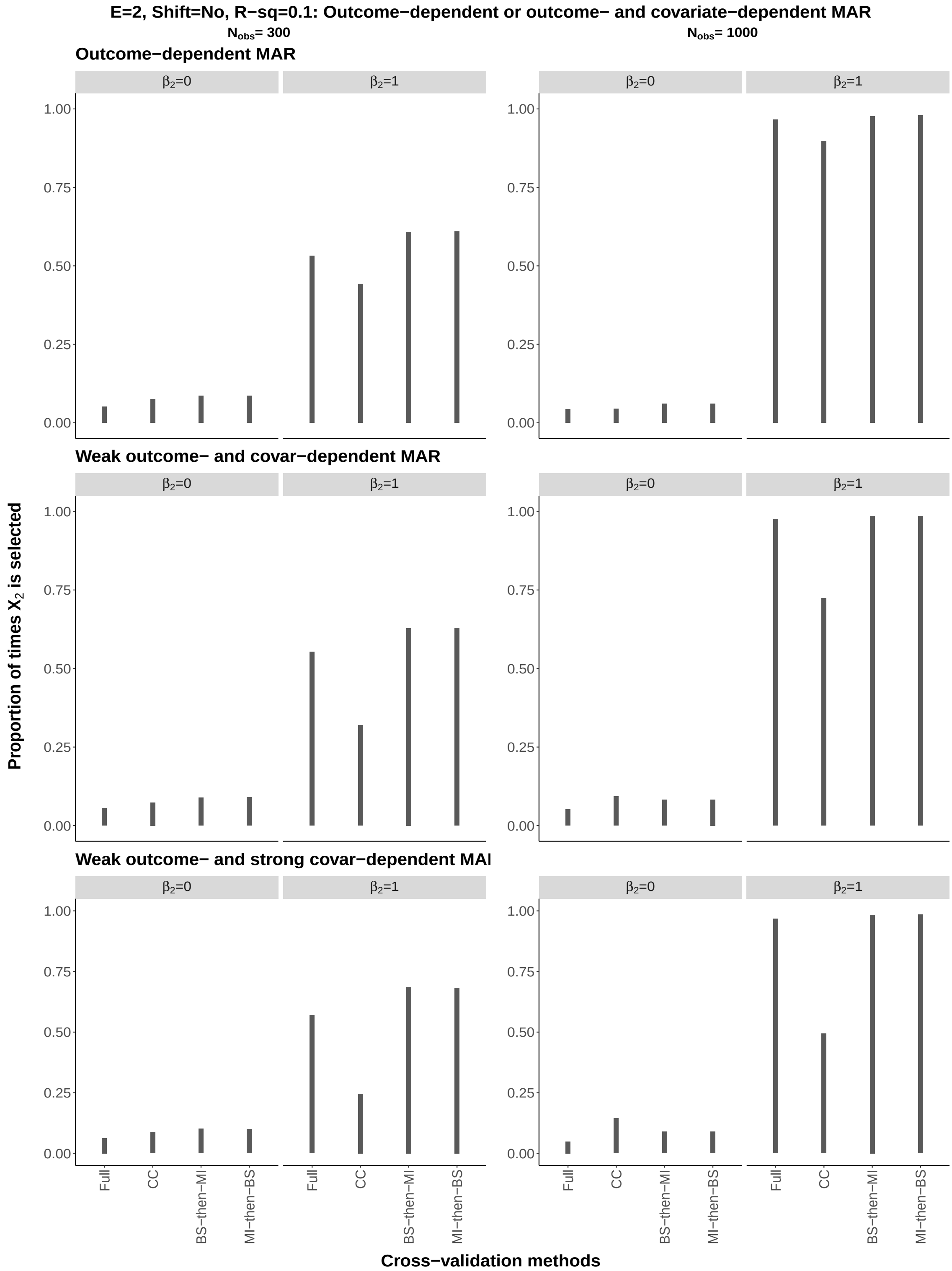

Figure S103: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

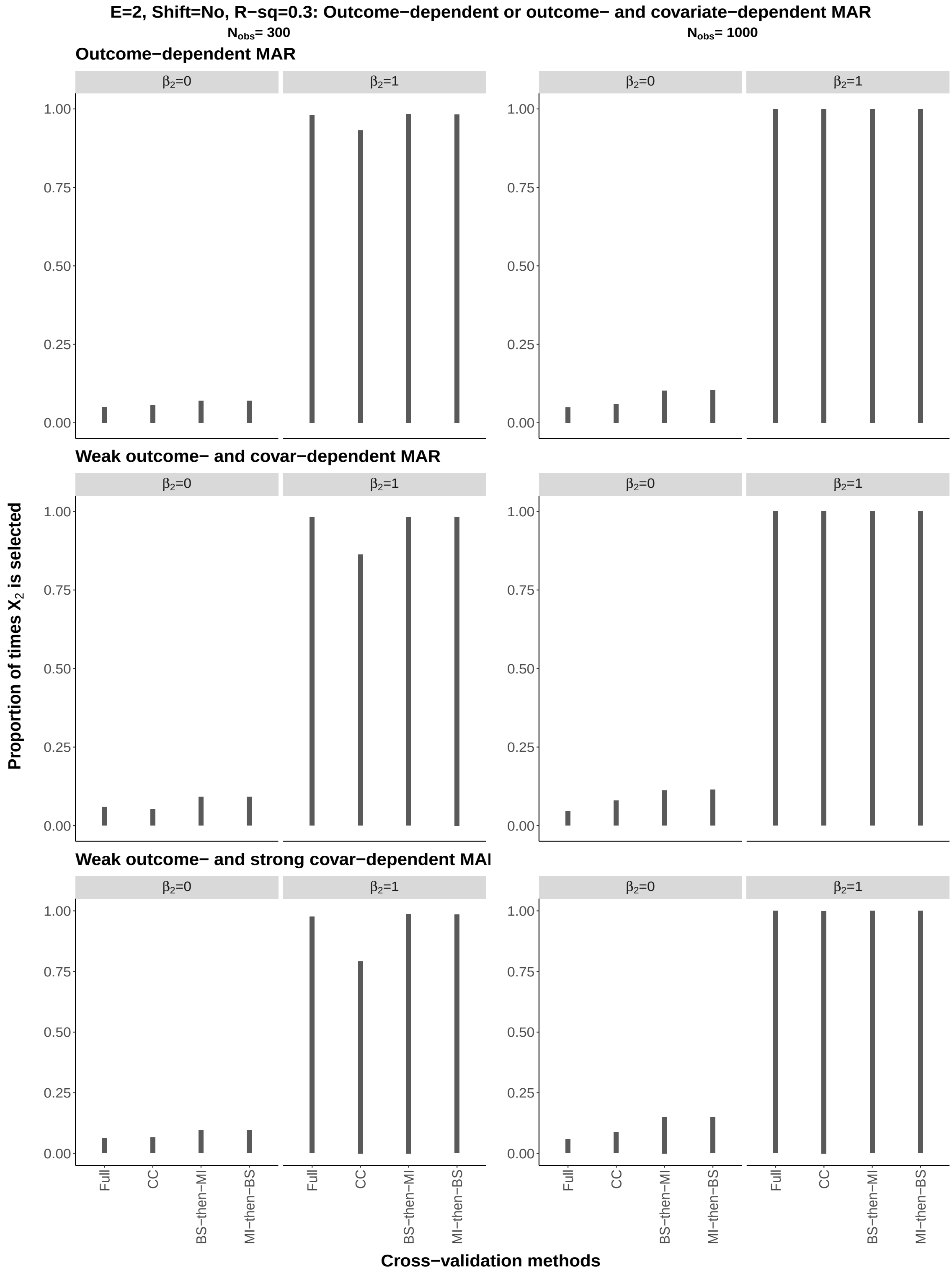

Figure S104: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

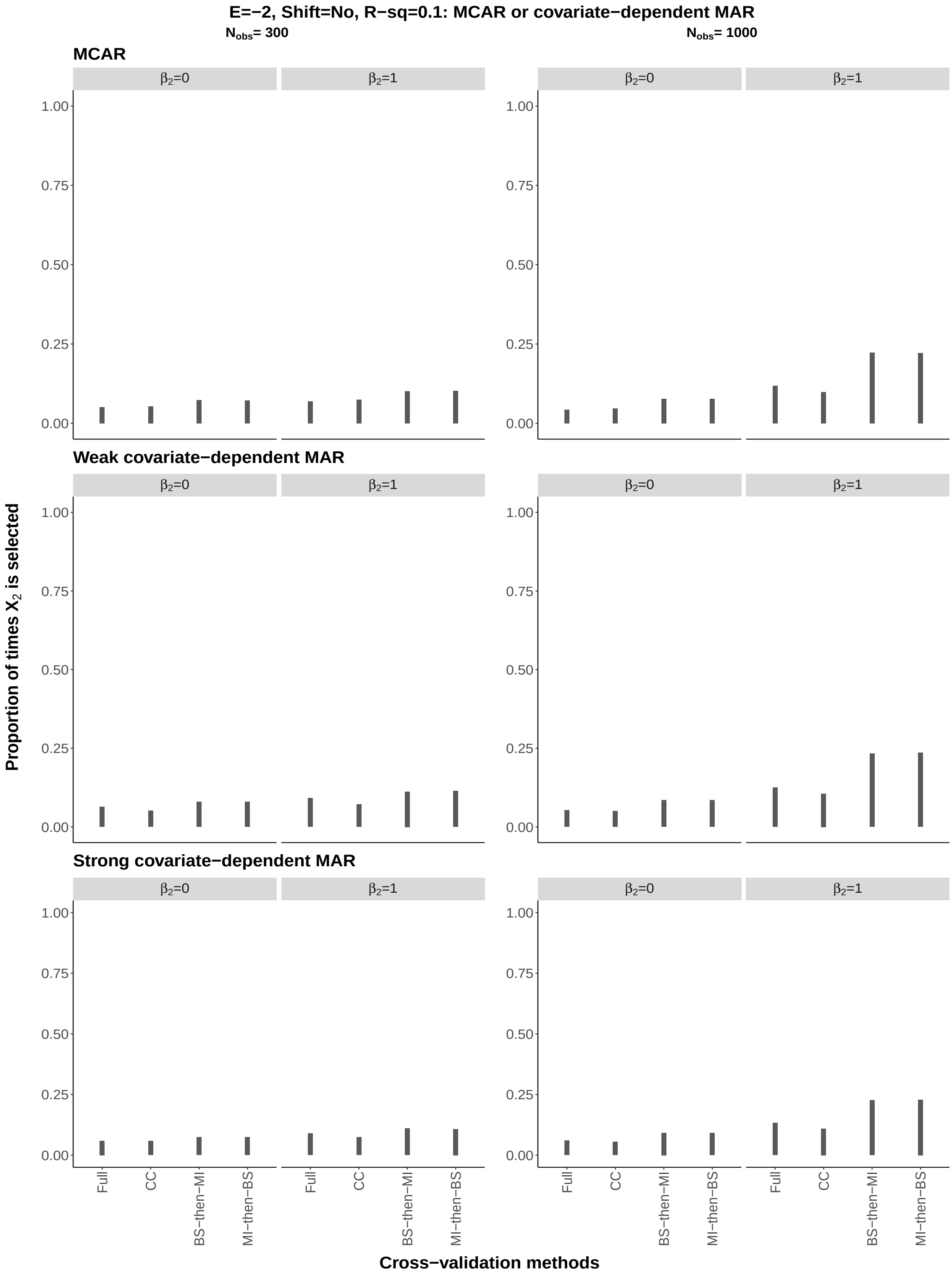

Figure S105: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

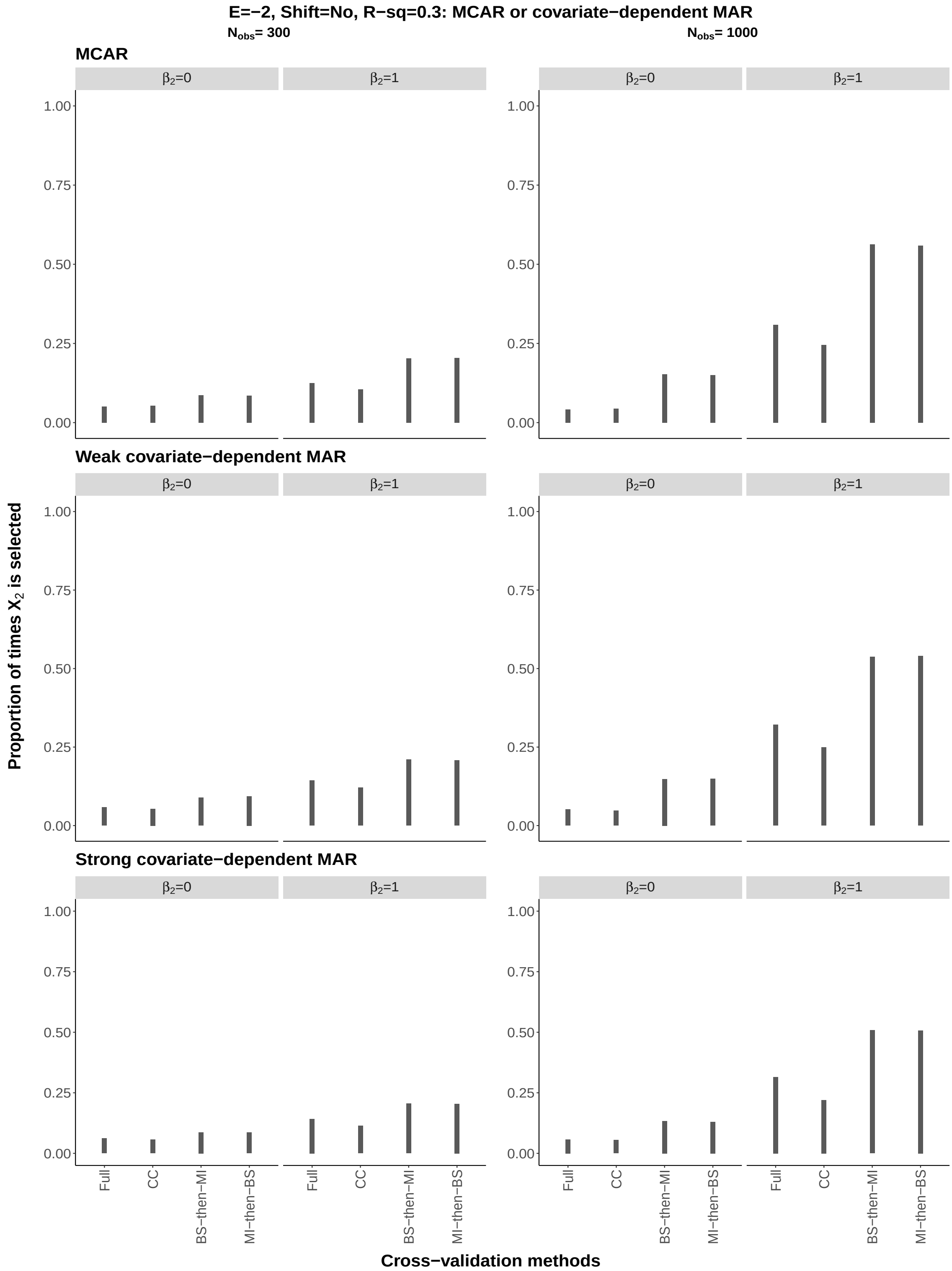

Figure S106: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

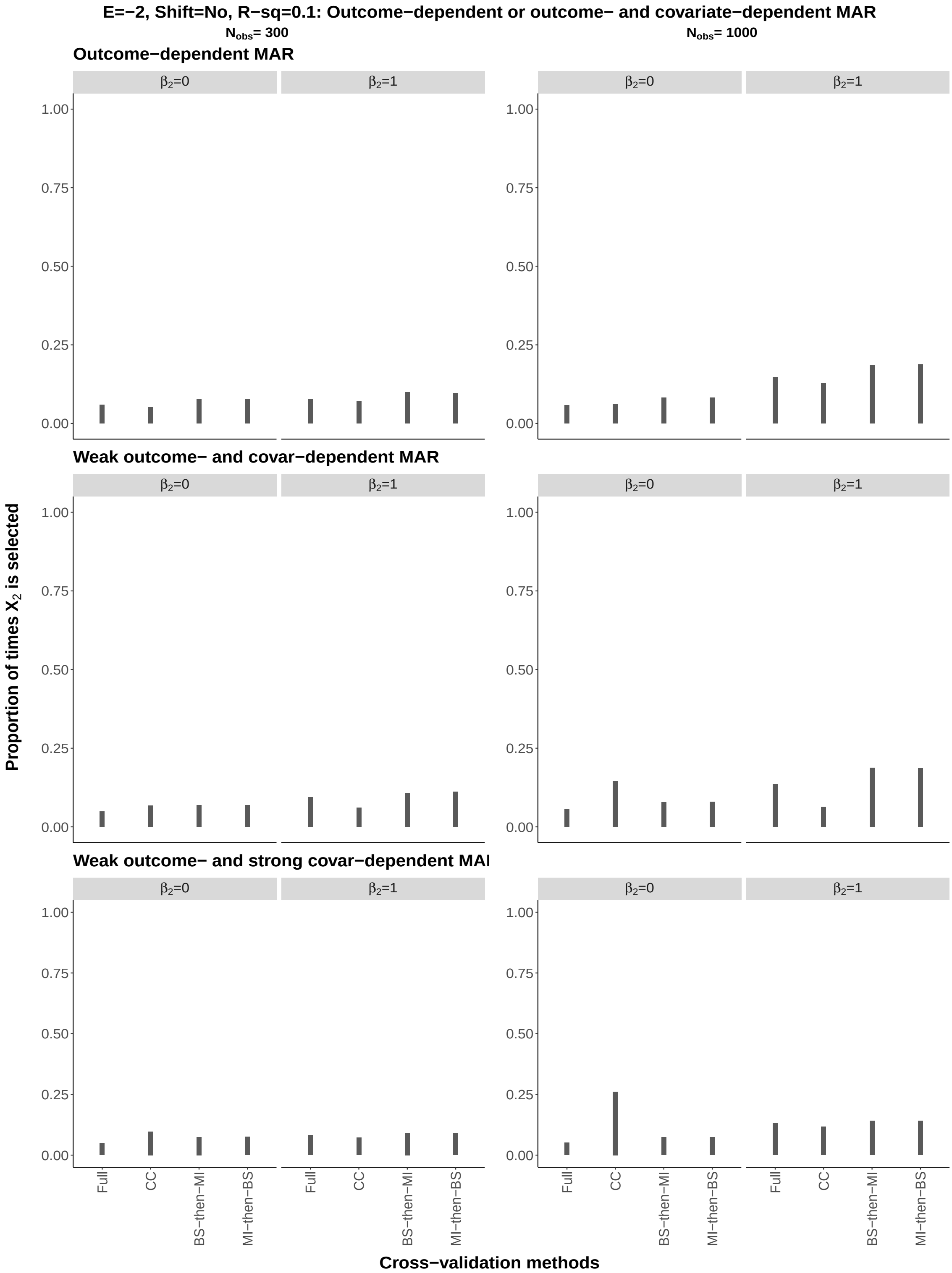

Figure S107: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

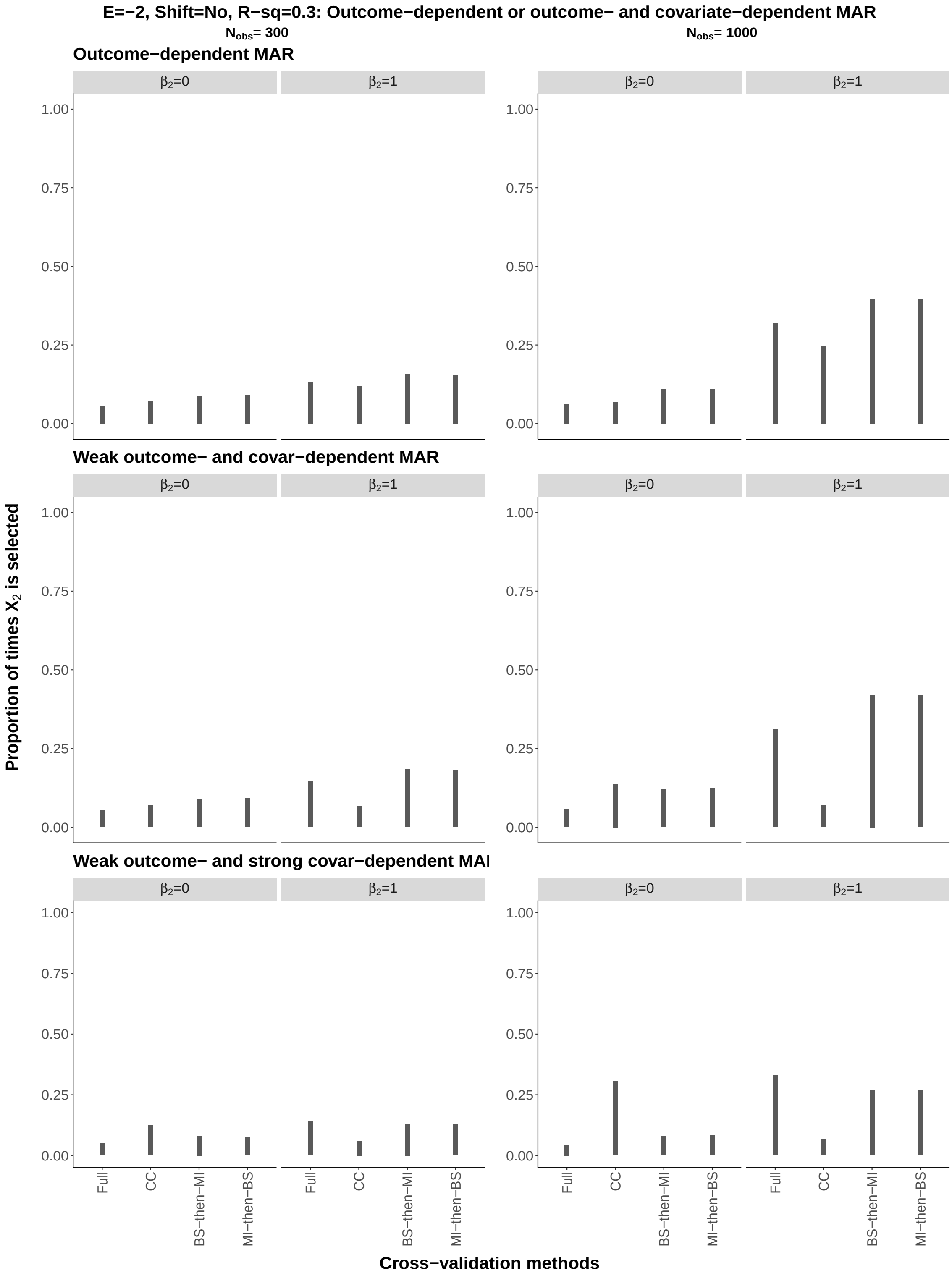

Figure S108: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.2 Covariate selection of X_2 using all data: $\beta_2 = 1$, $\alpha_E = 0.05$, no origin-shift

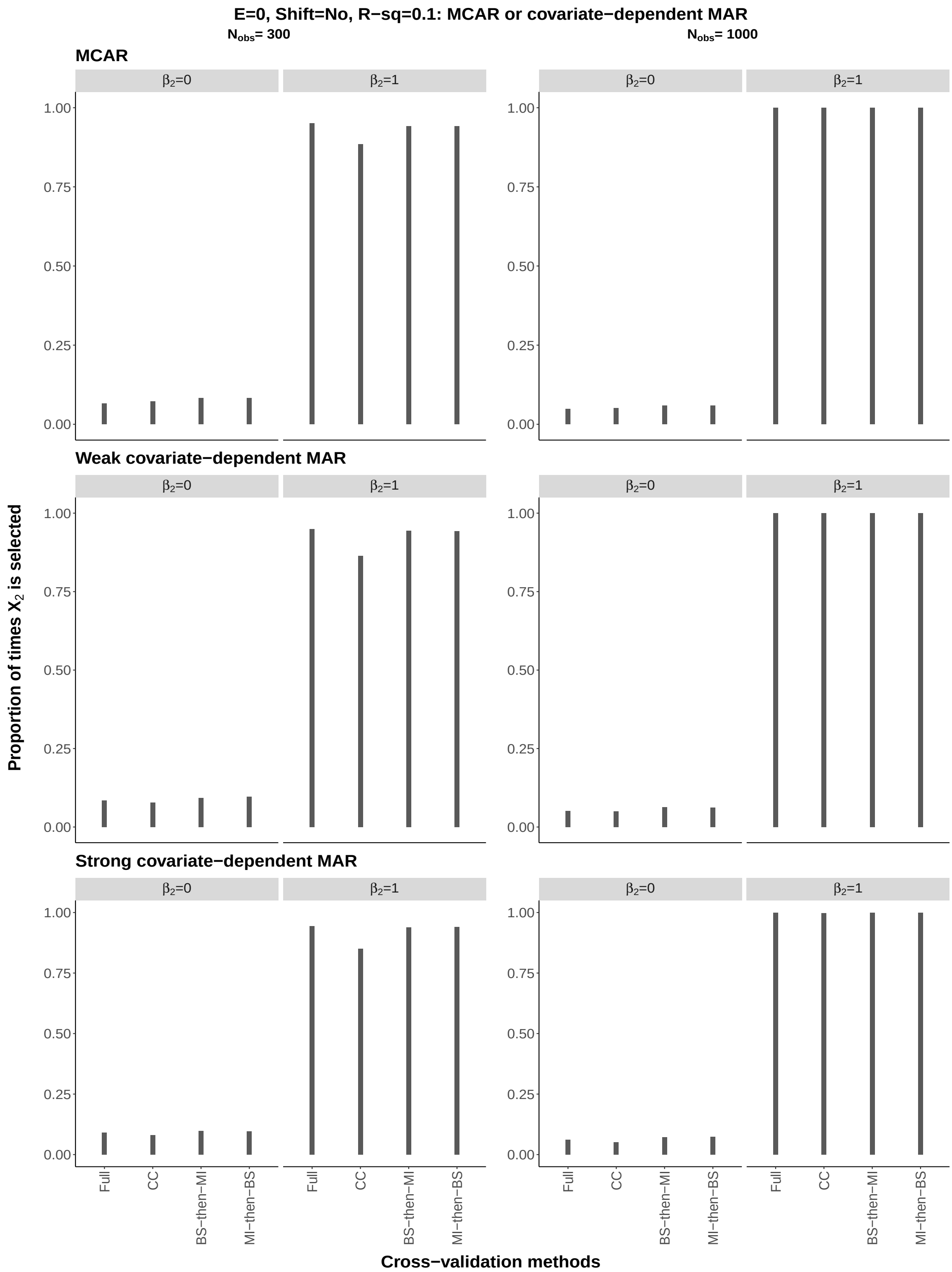

Figure S109: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

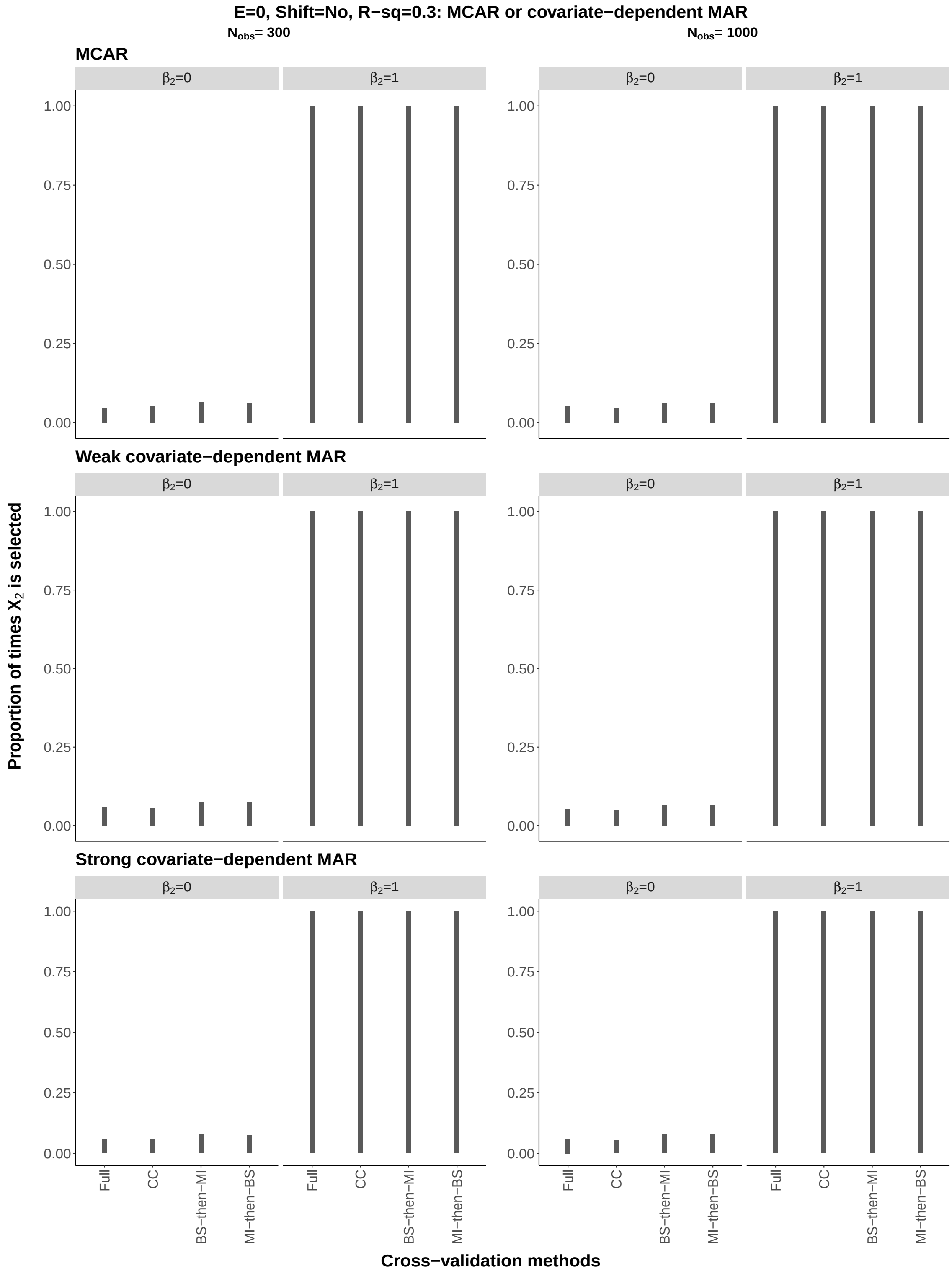

Figure S110: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

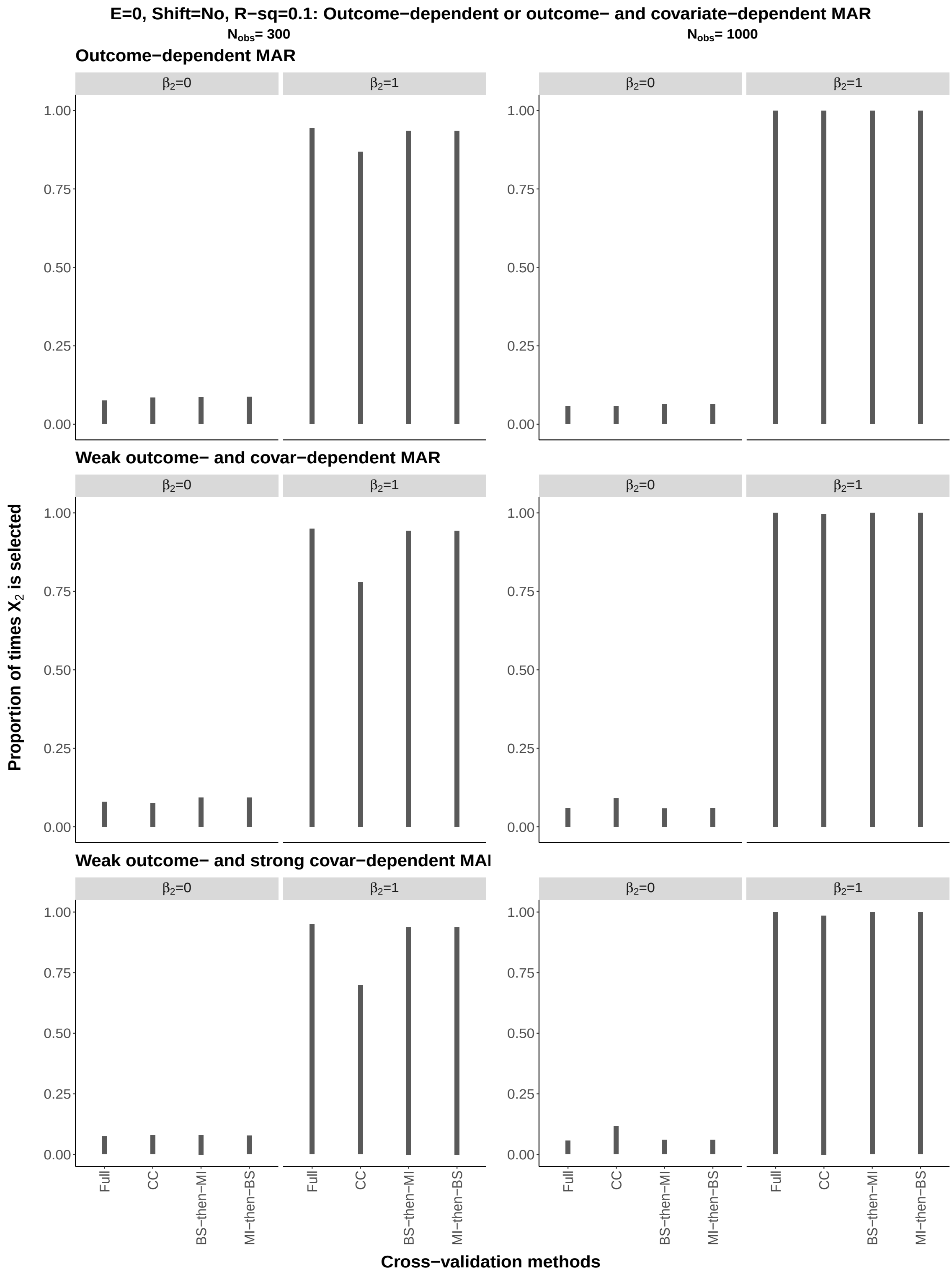

Figure S111: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

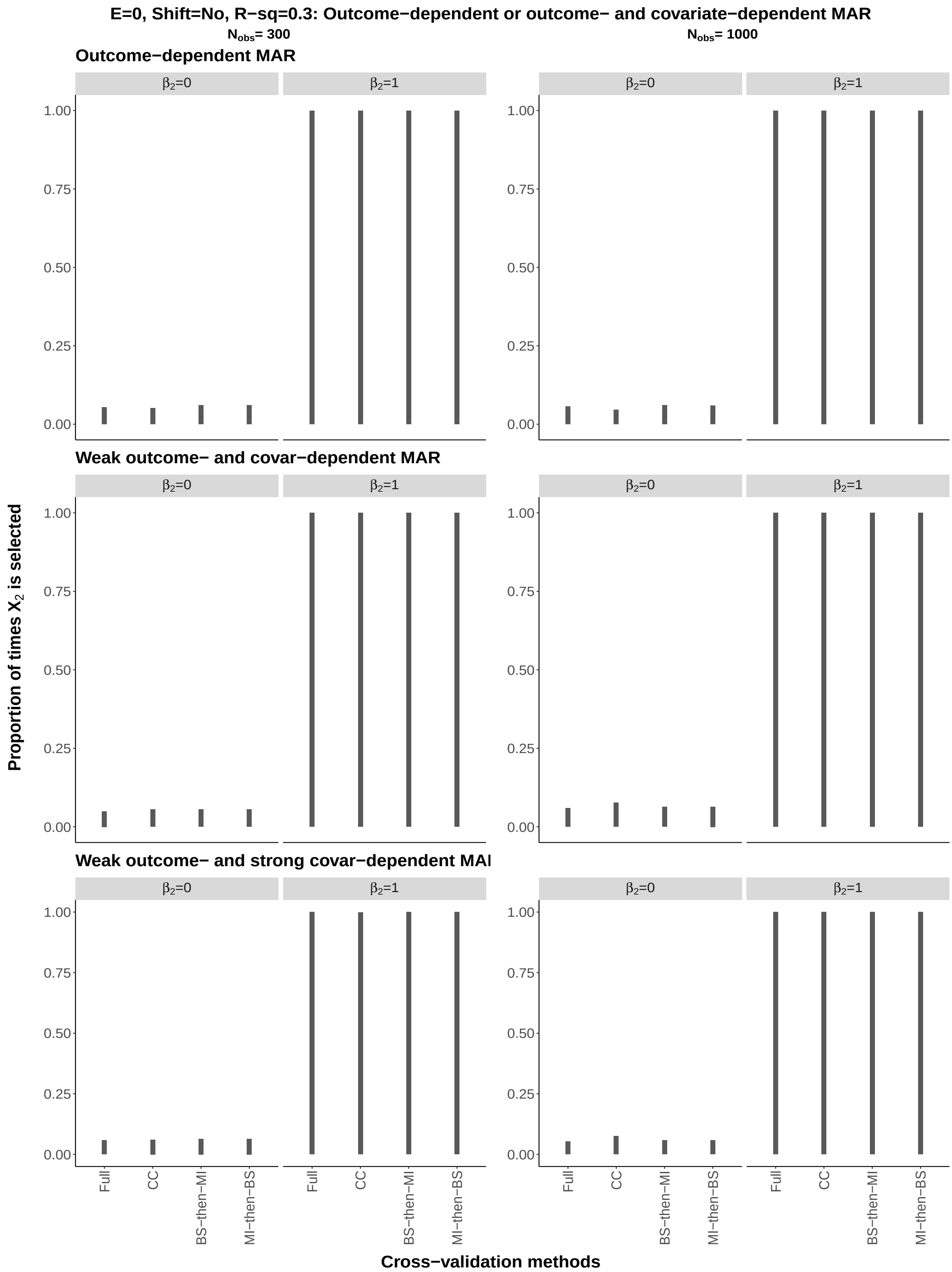

Figure S112: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

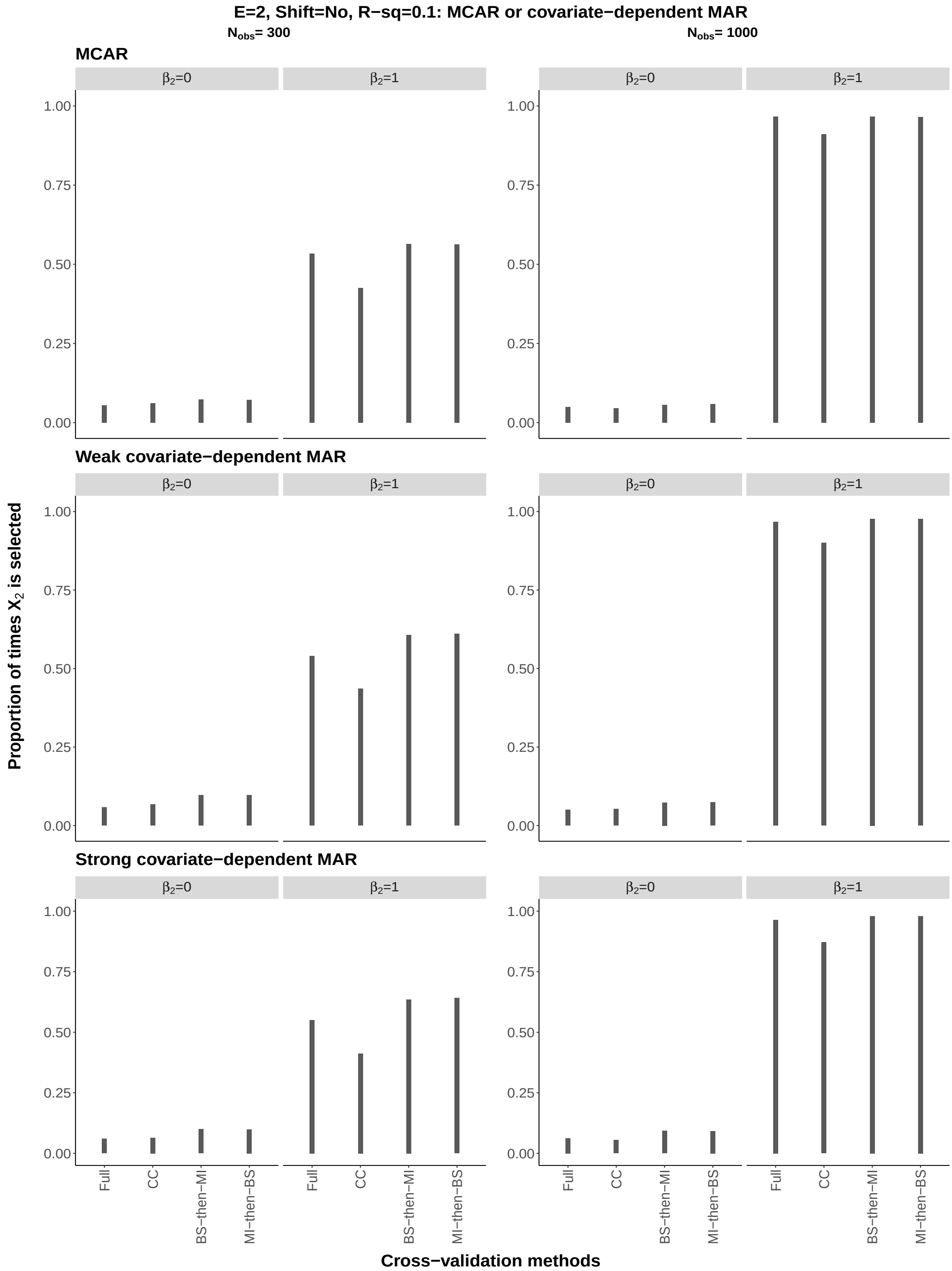

Figure S113: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

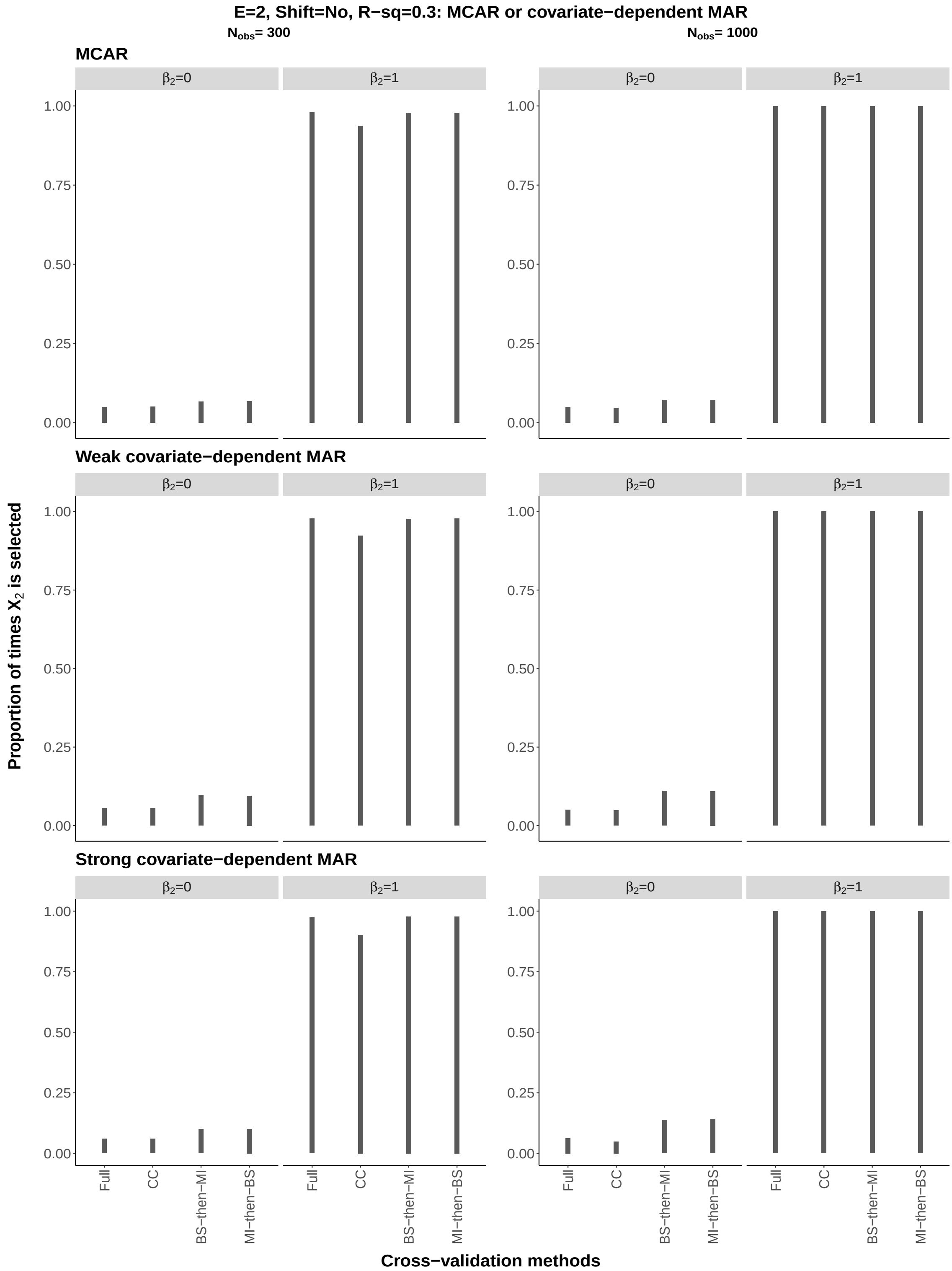

Figure S114: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

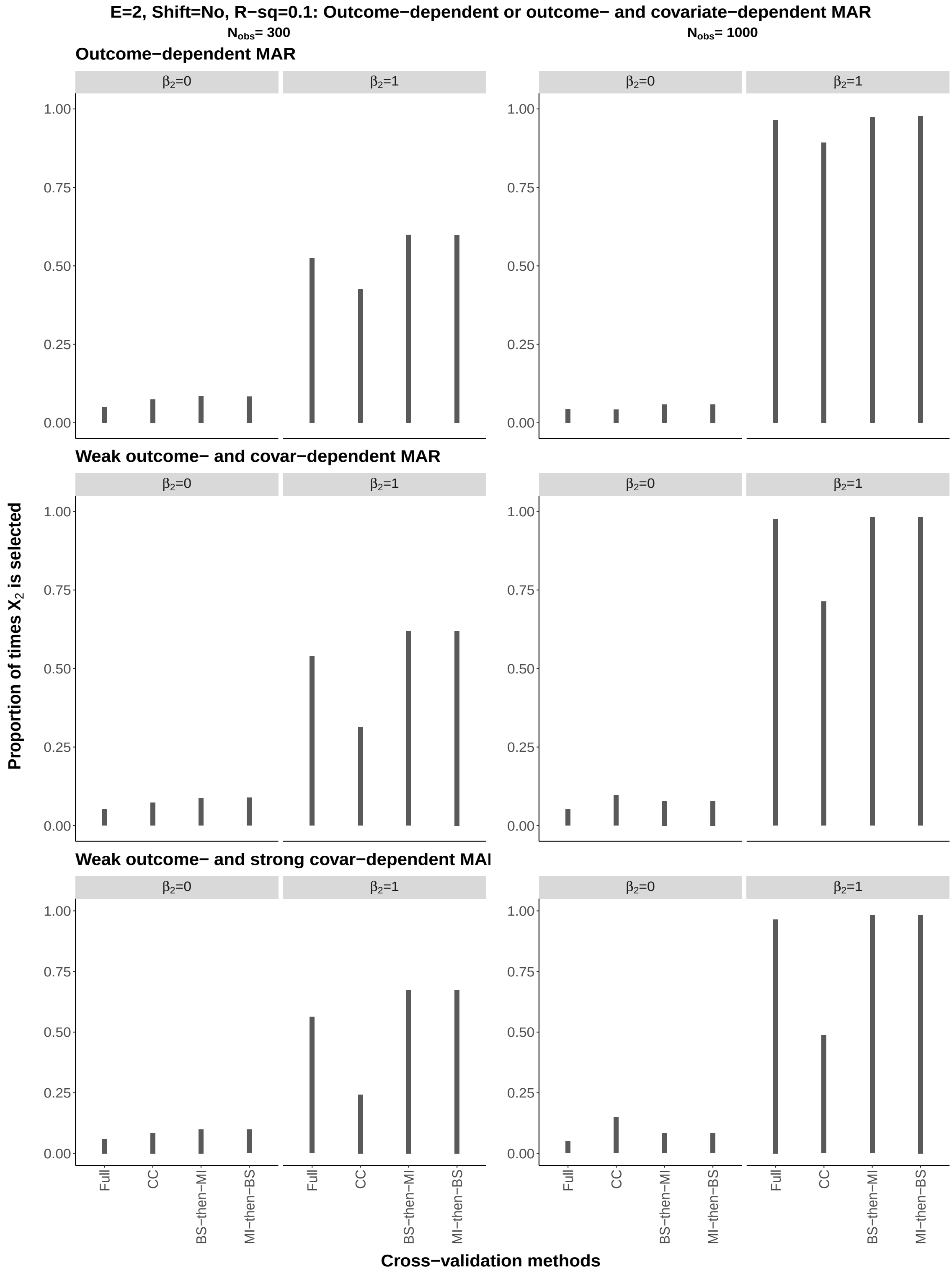

Figure S115: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

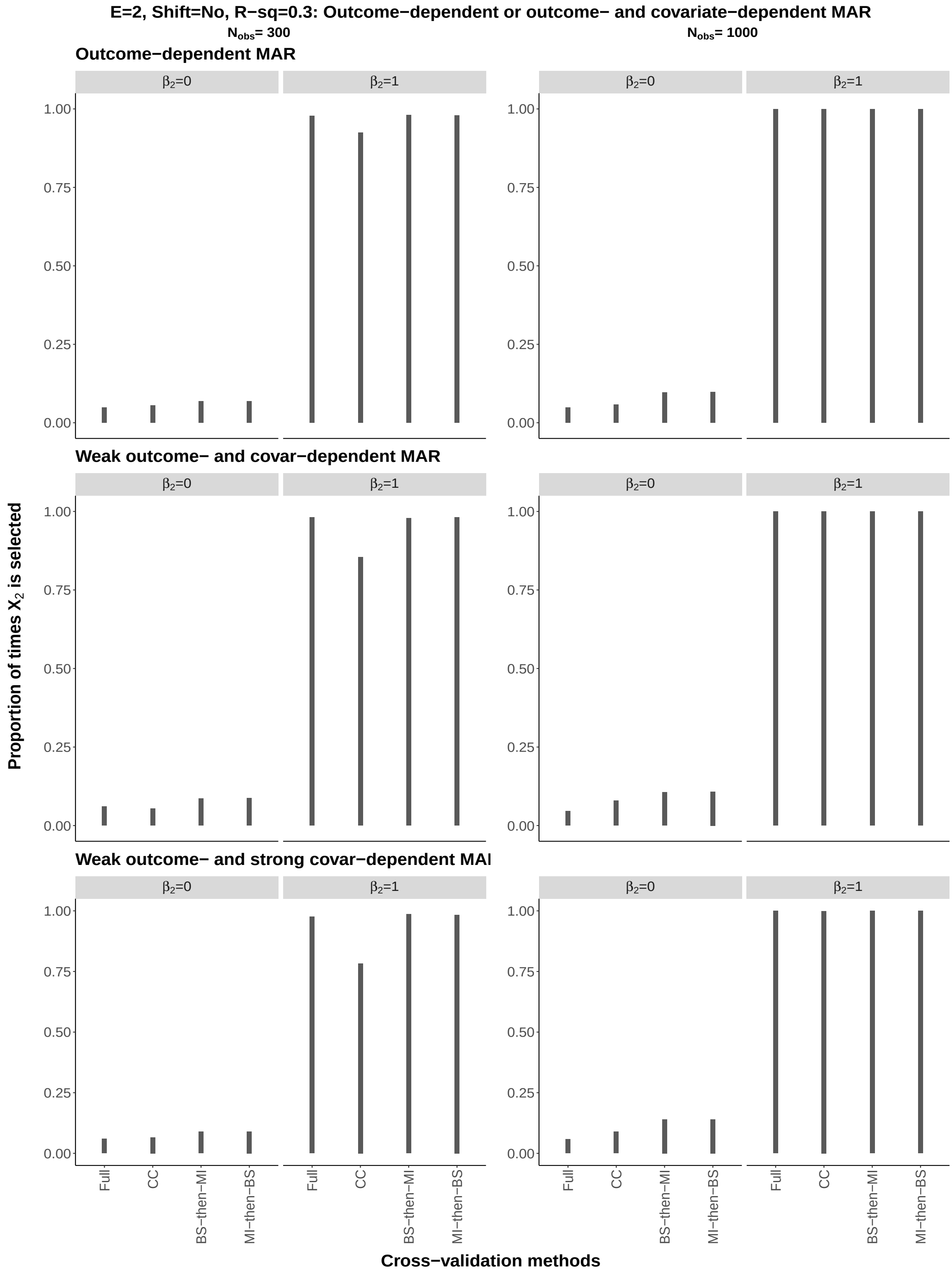

Figure S116: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

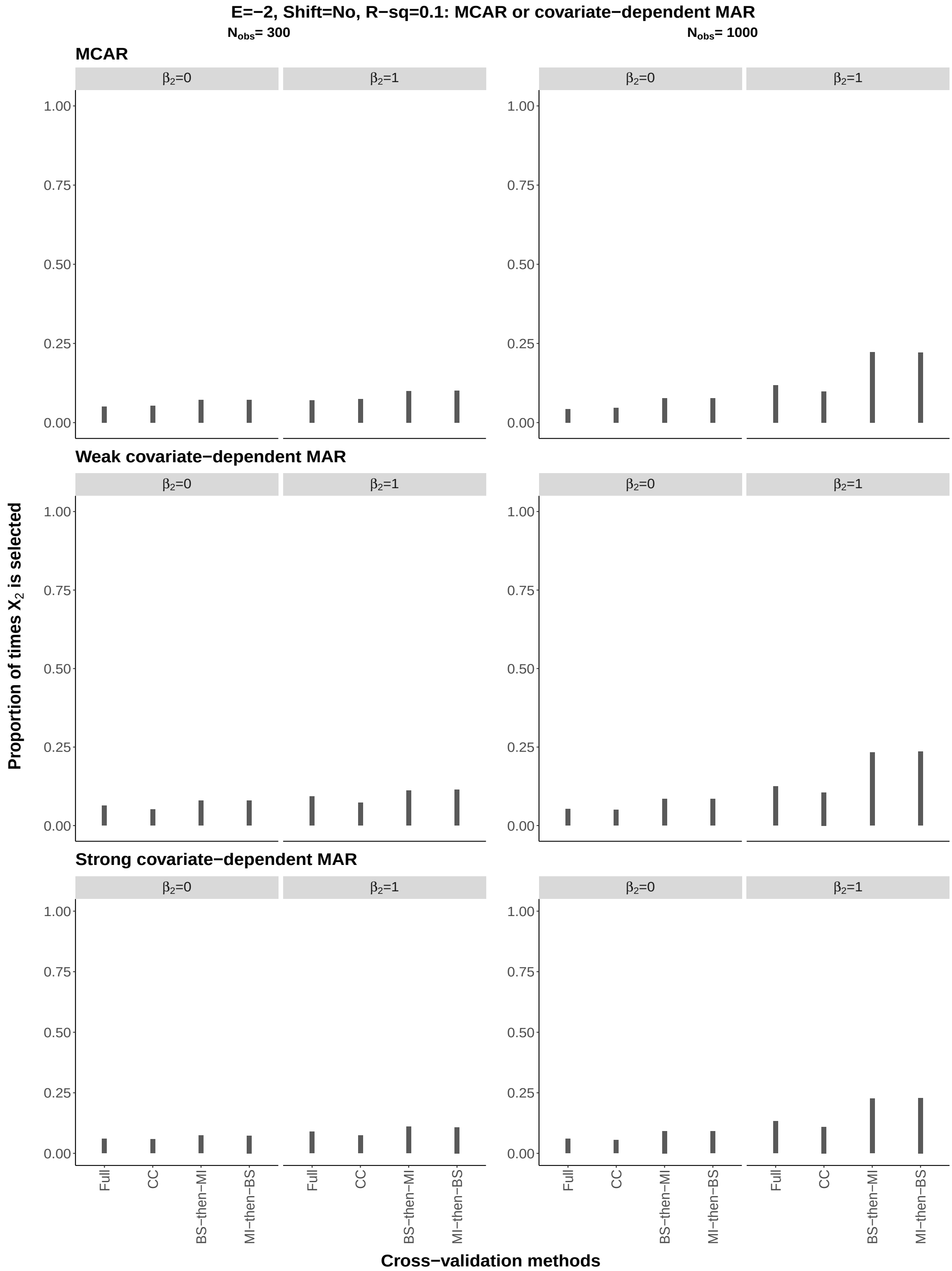

Figure S117: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

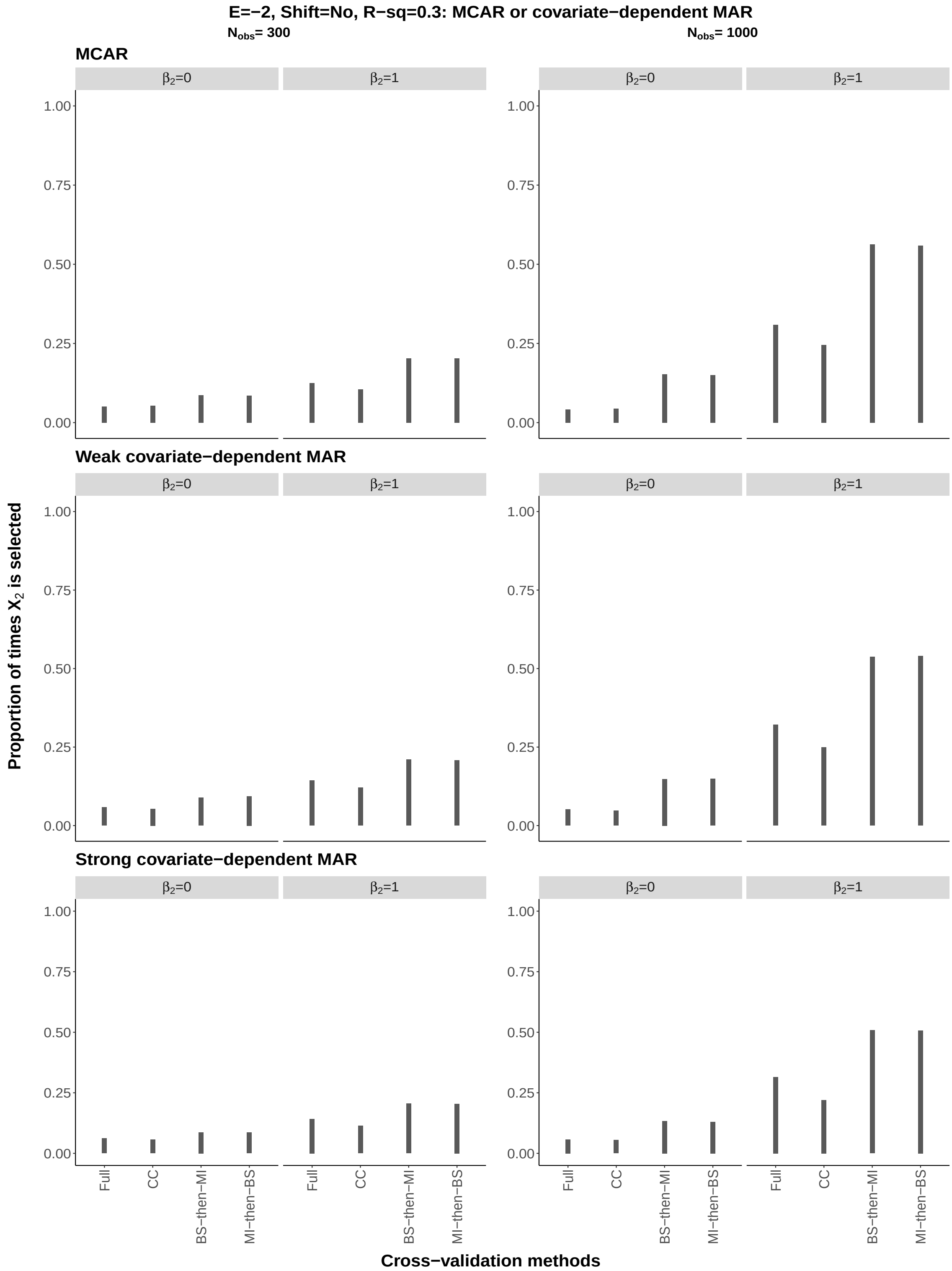

Figure S118: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

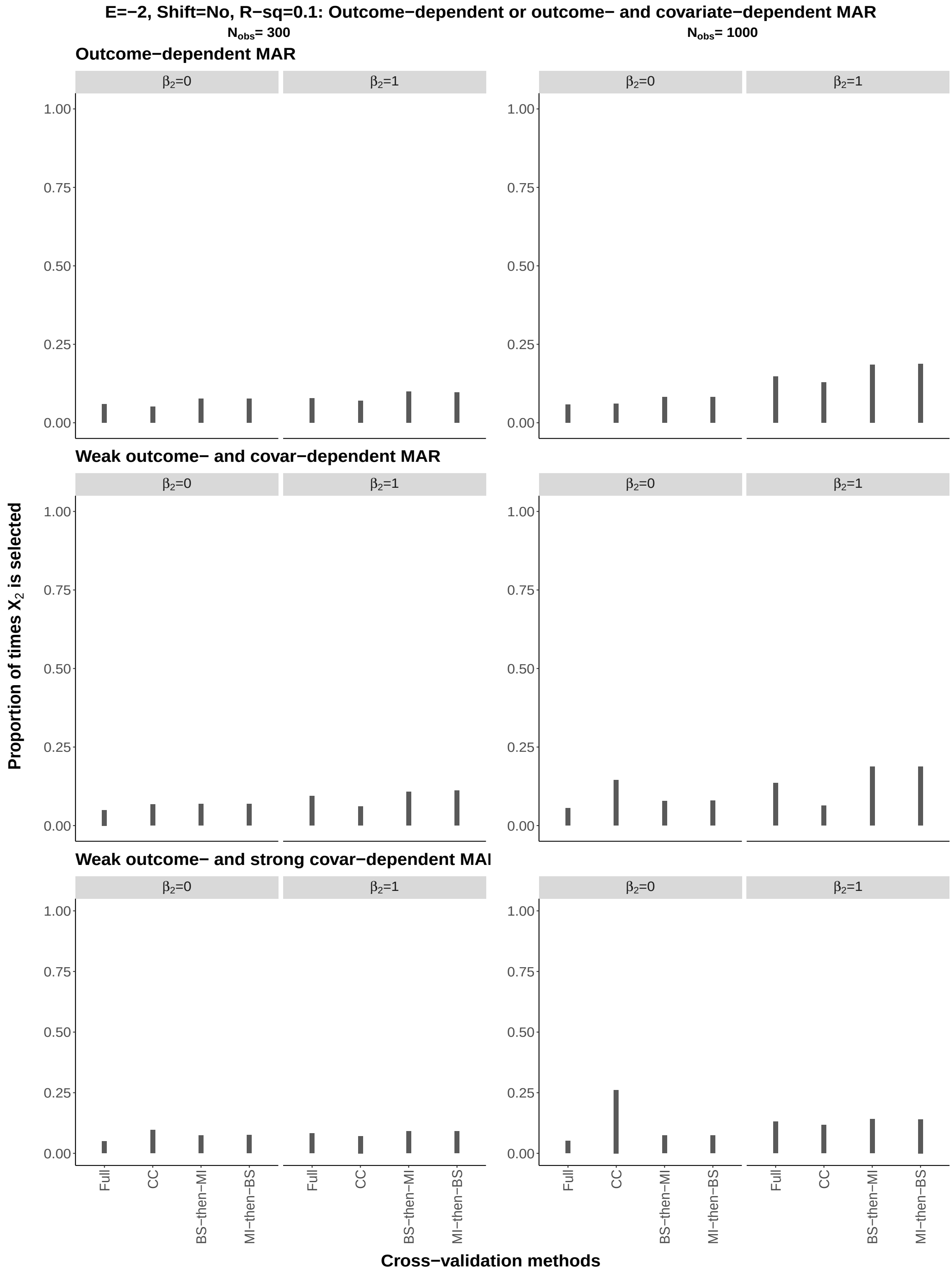

Figure S119: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

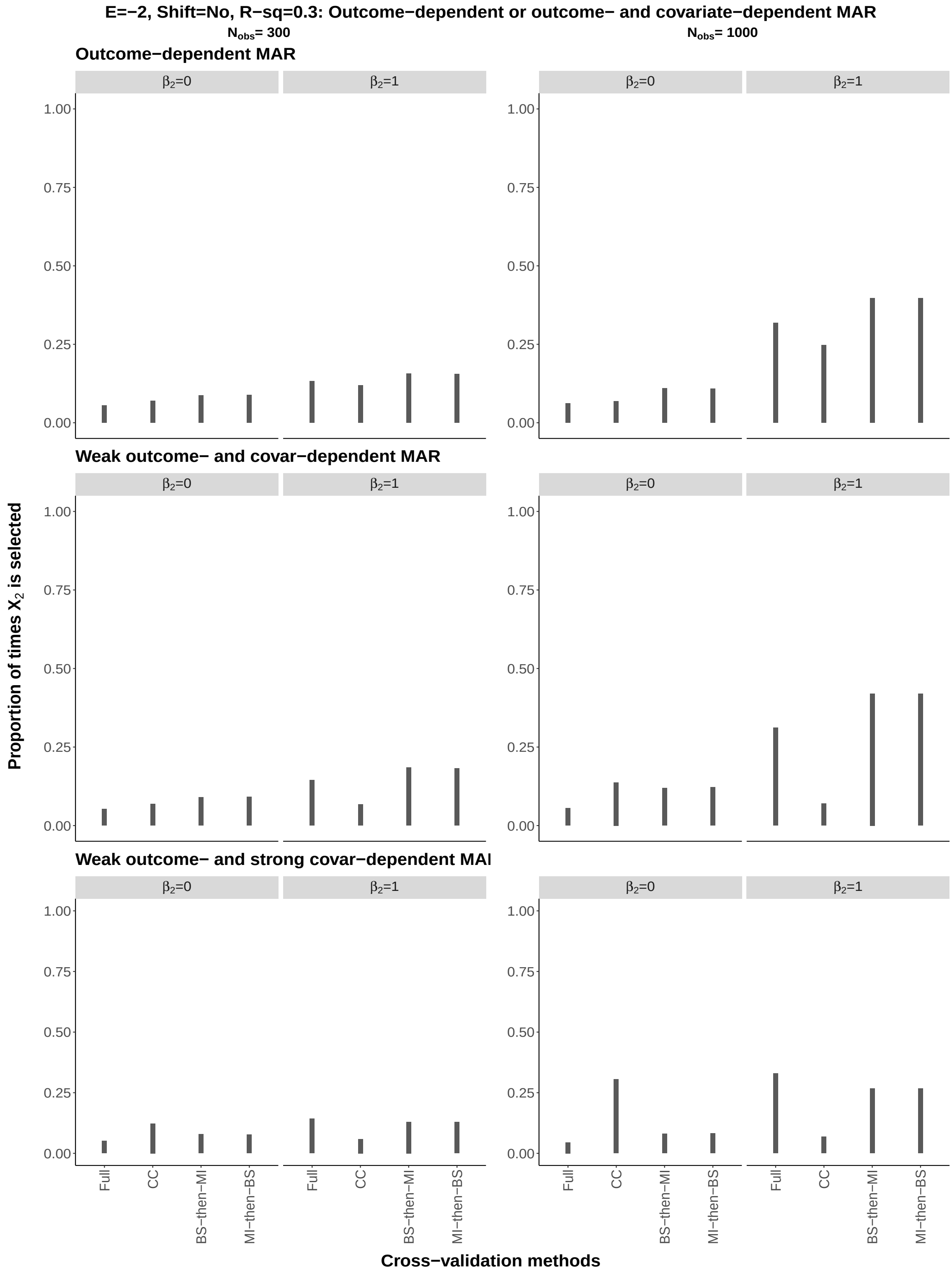

Figure S120: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.3 Covariate selection of X_2 using all data: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

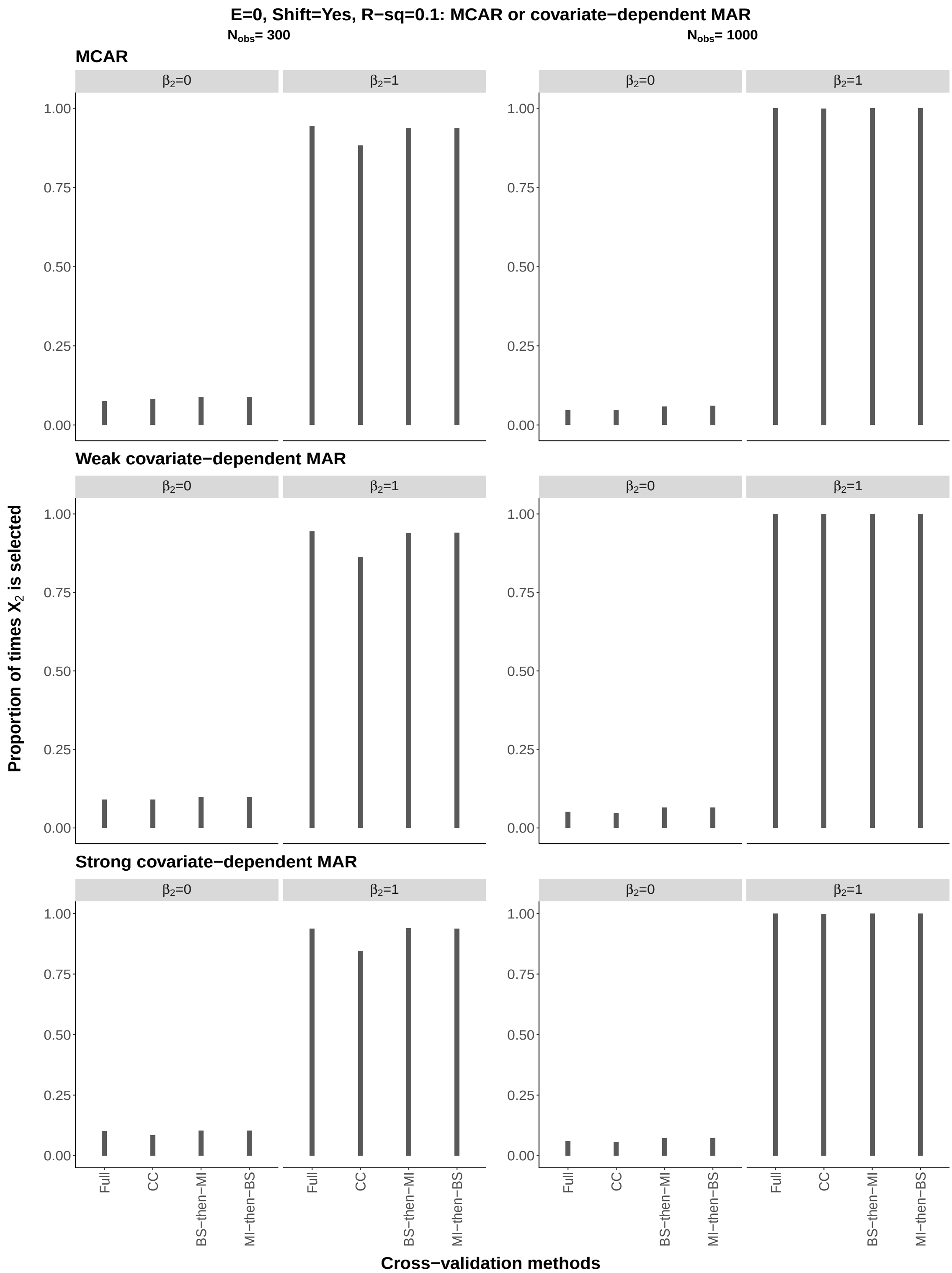

Figure S121: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.1763

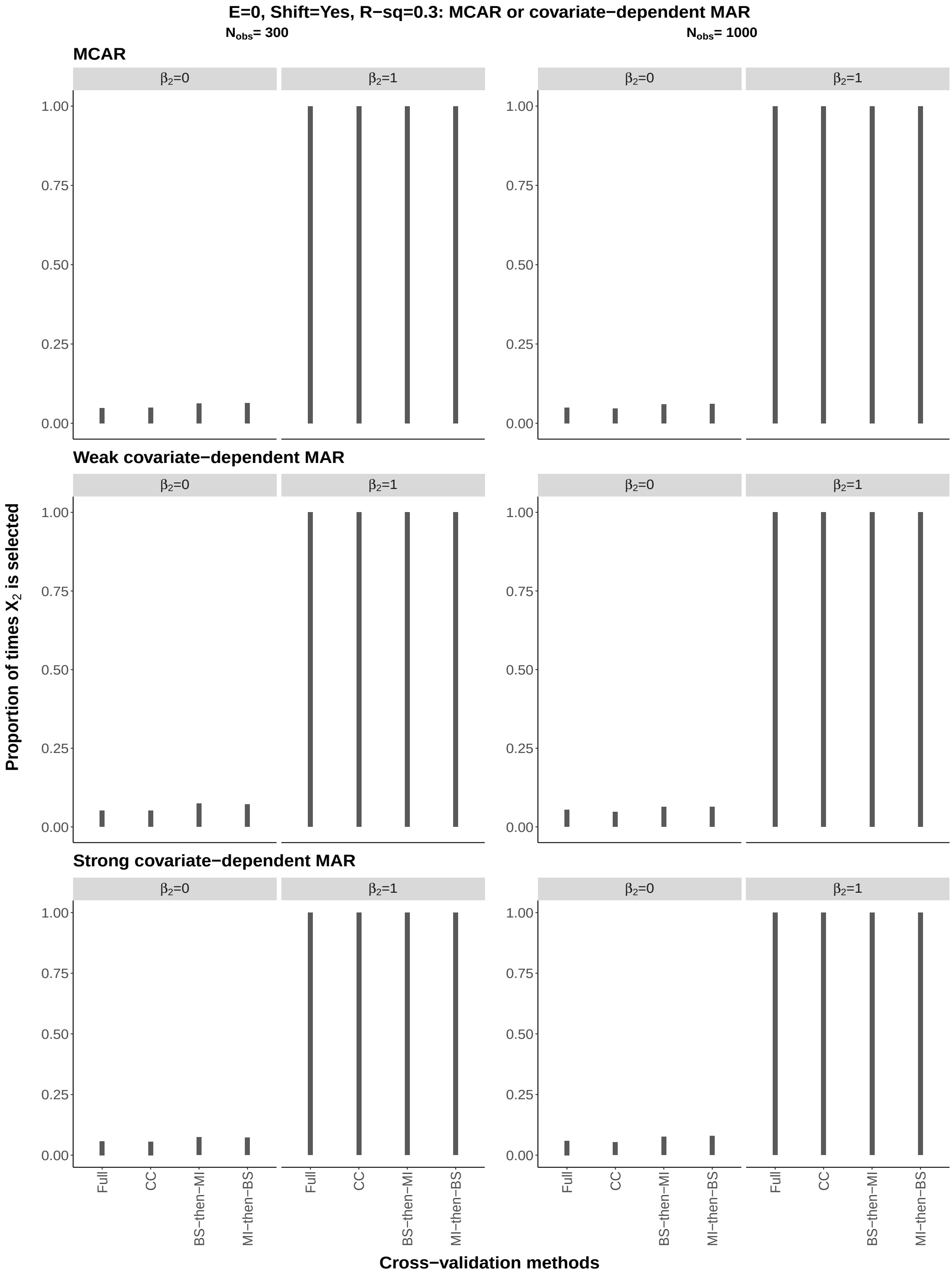

Figure S122: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

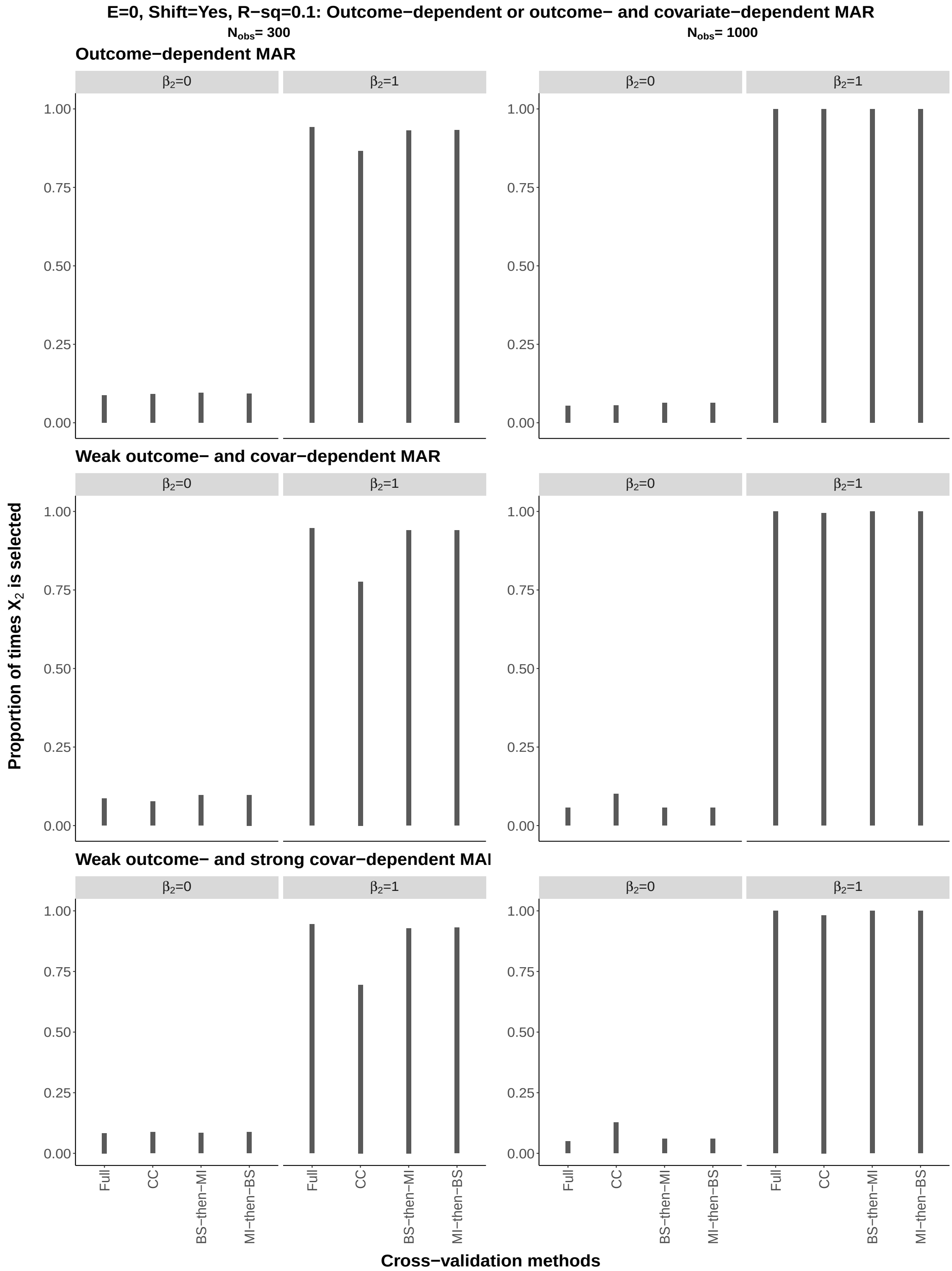

Figure S123: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

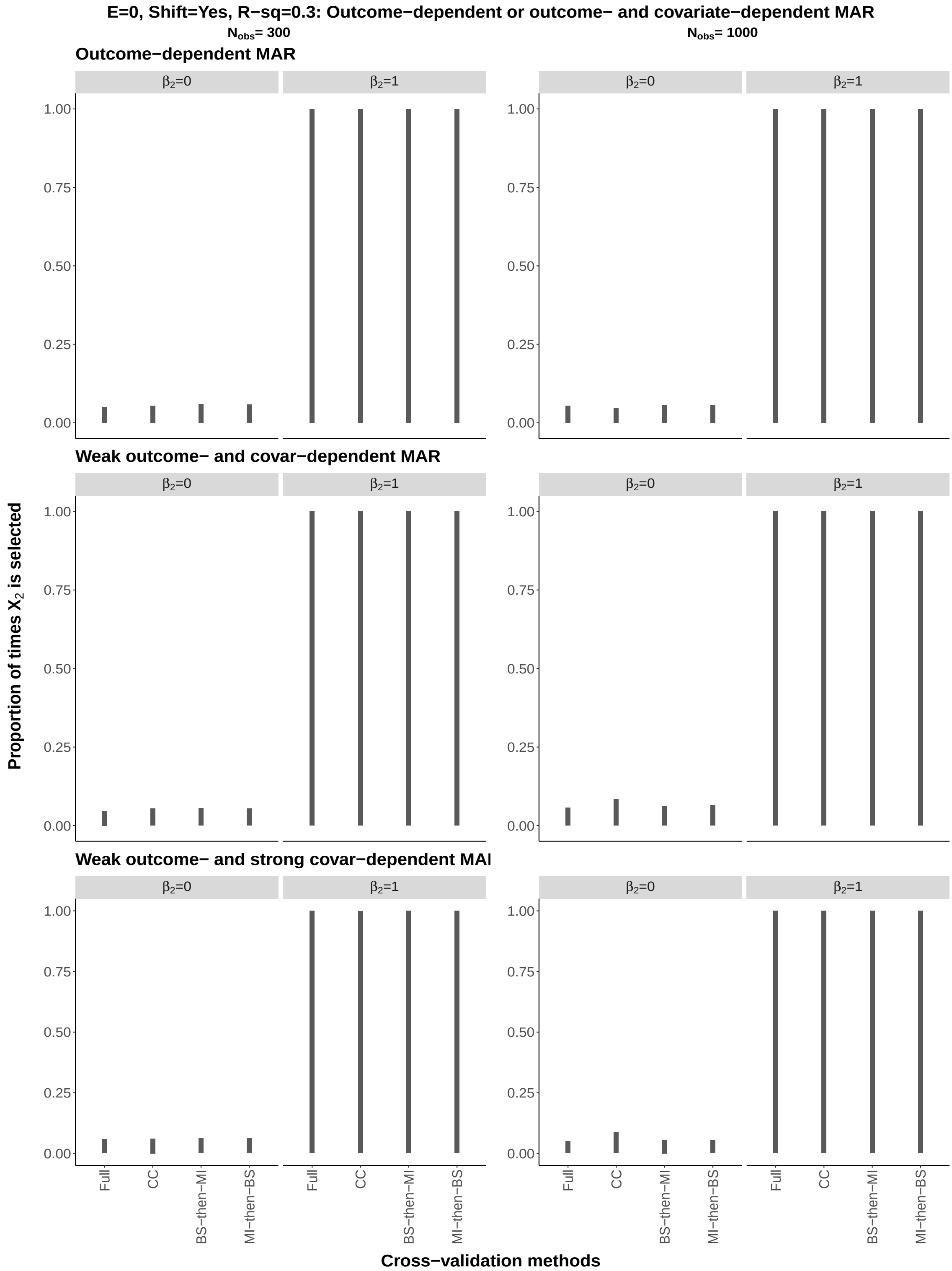

Figure S124: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

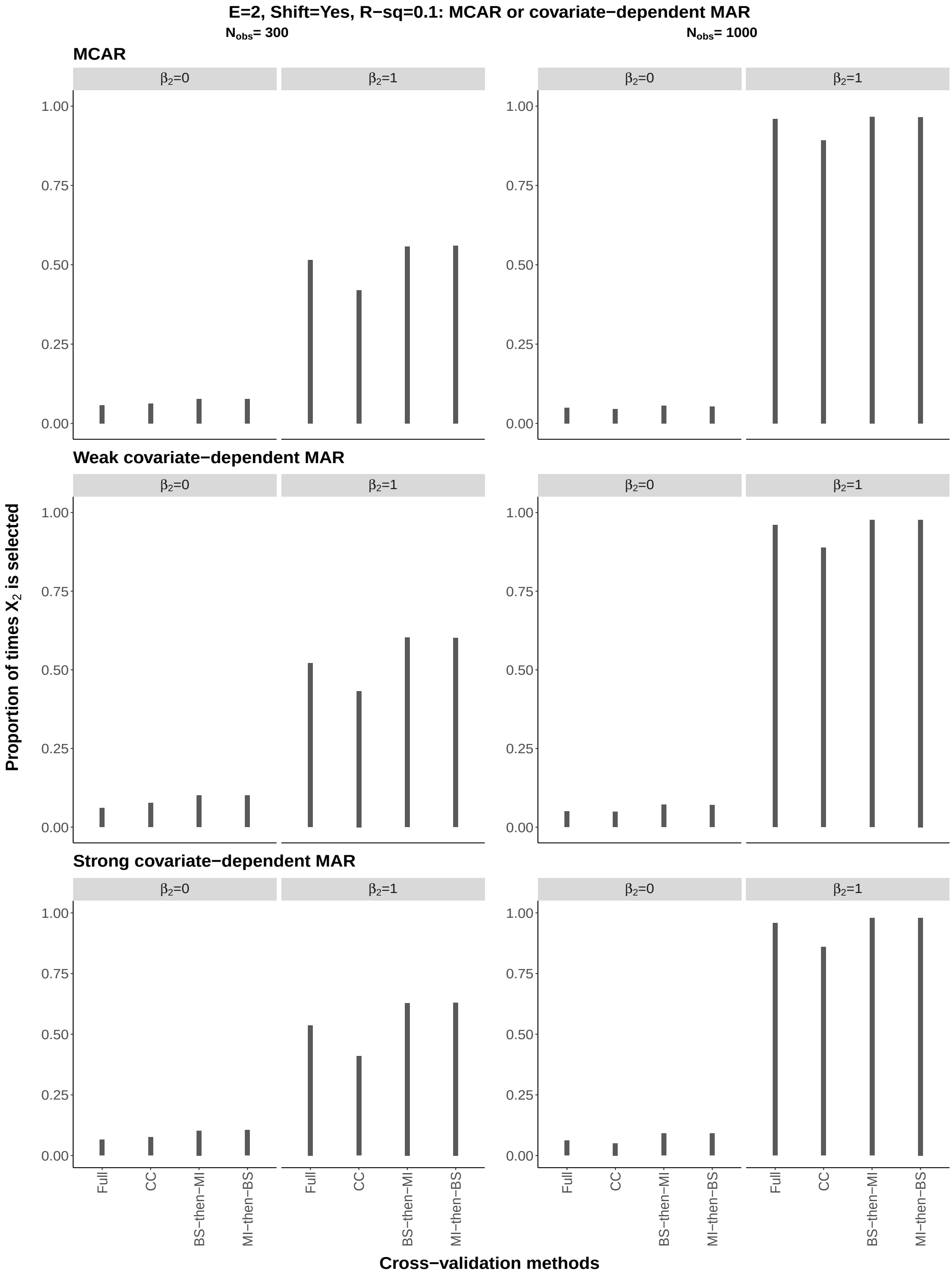

Figure S125: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

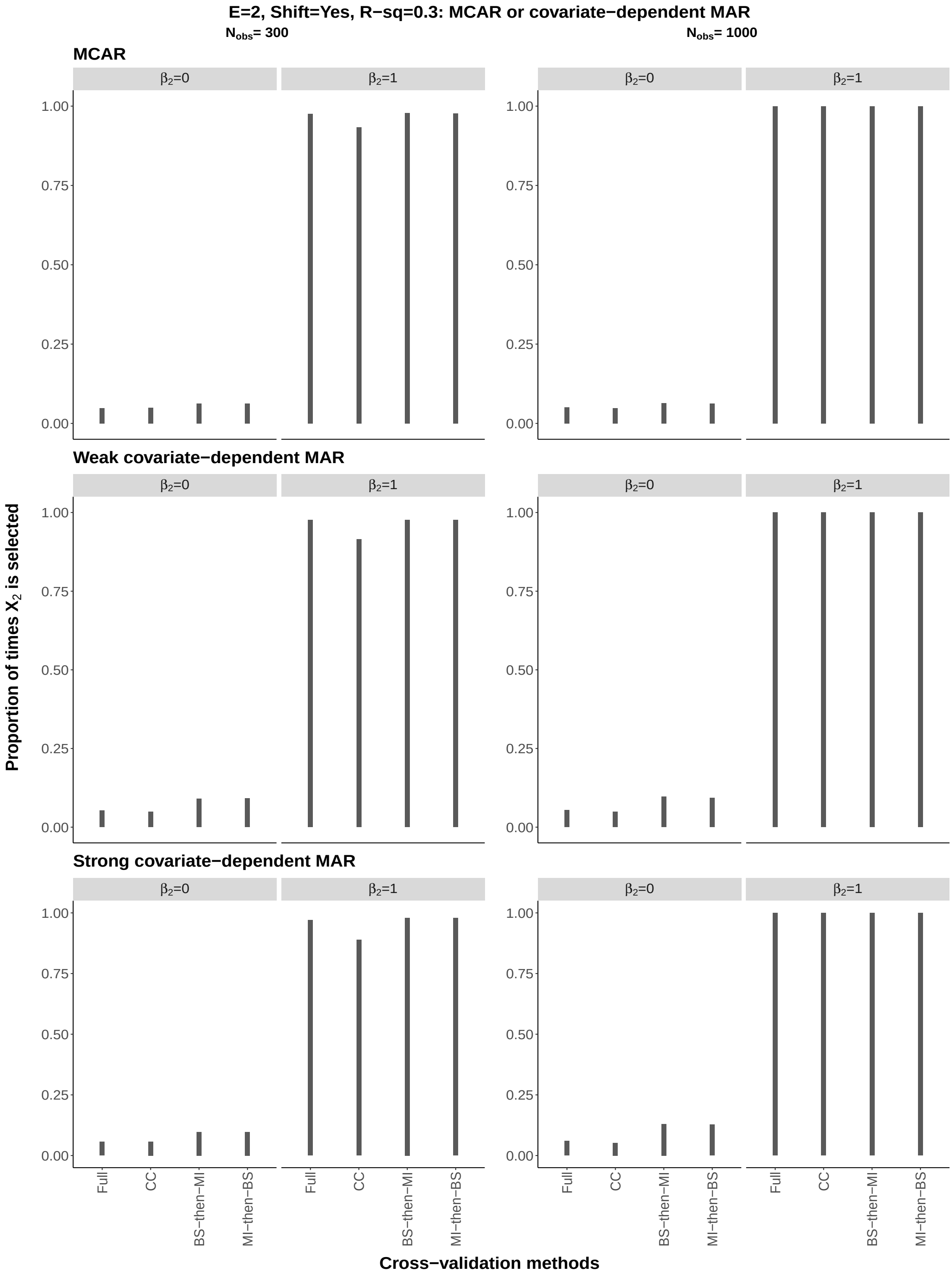

Figure S126: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

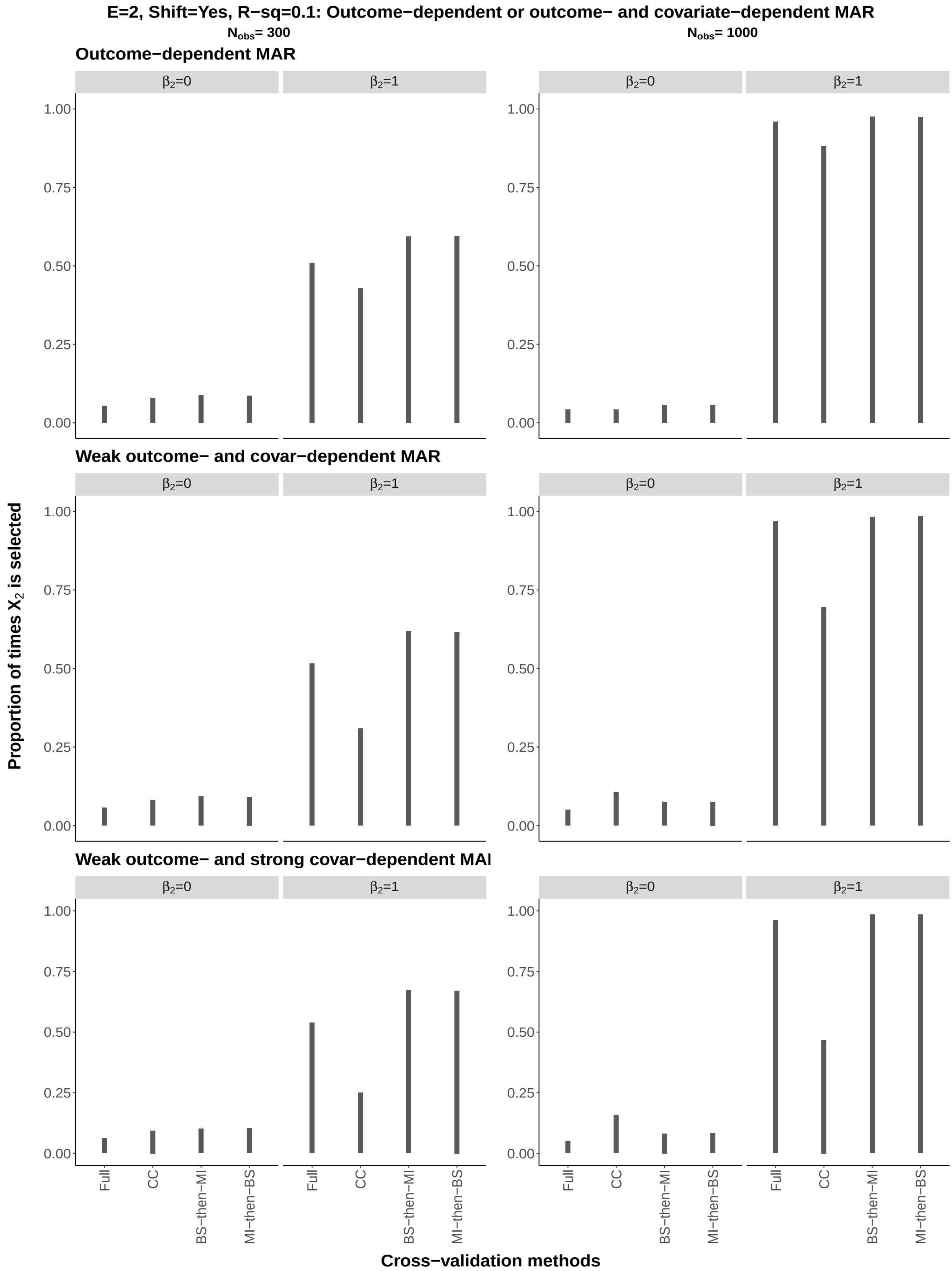

Figure S127: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

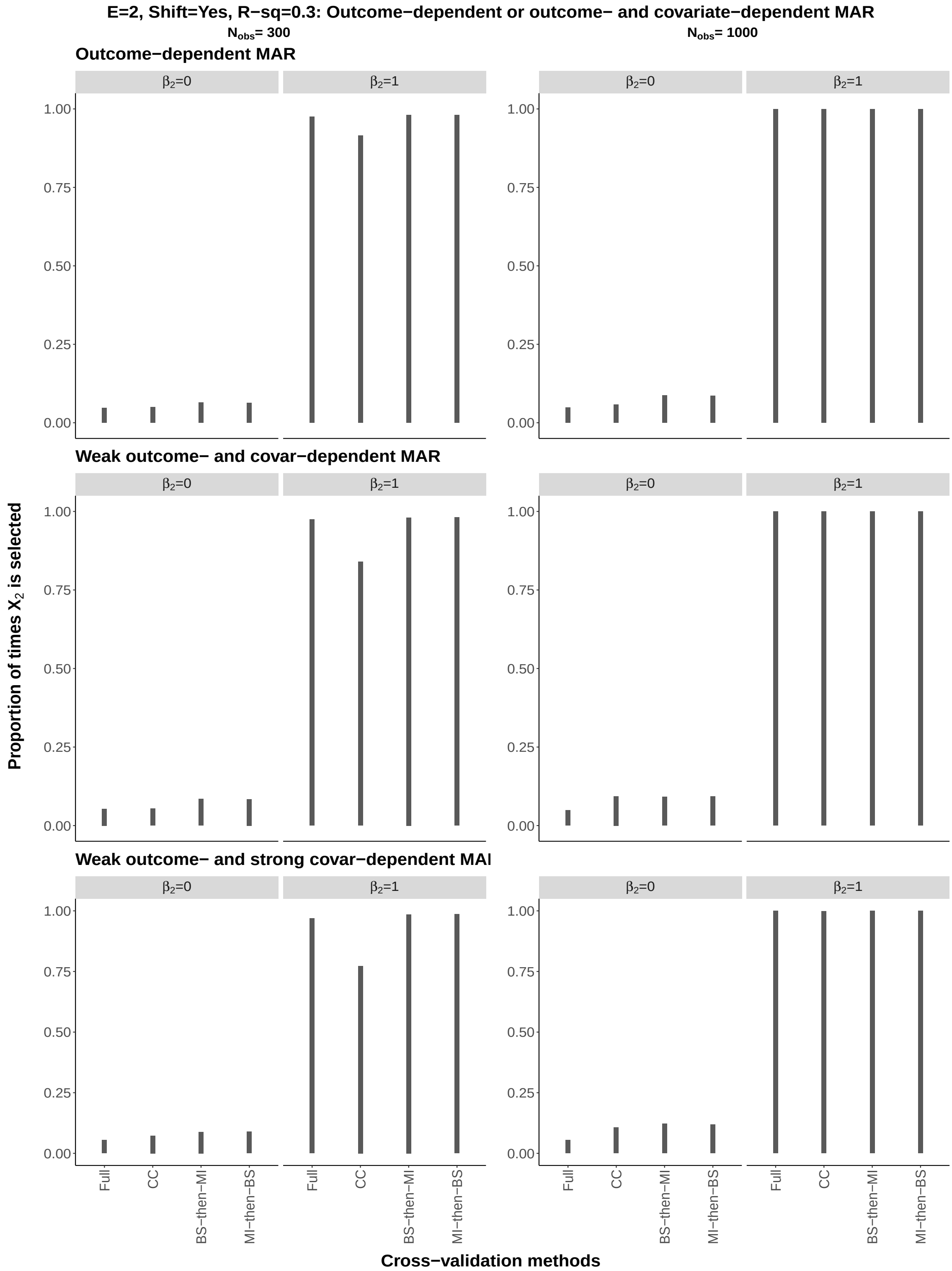

Figure S128: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

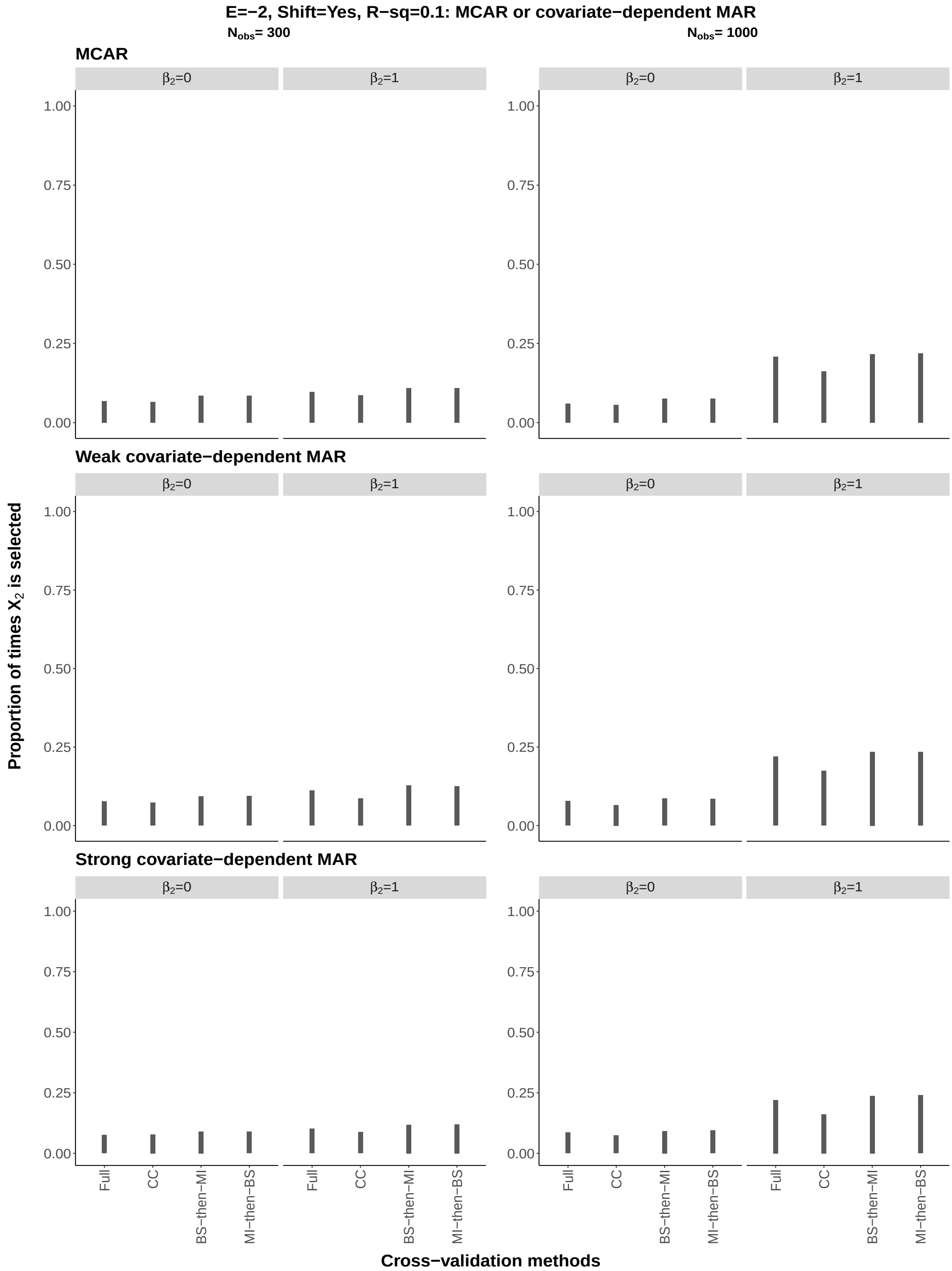

Figure S129: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

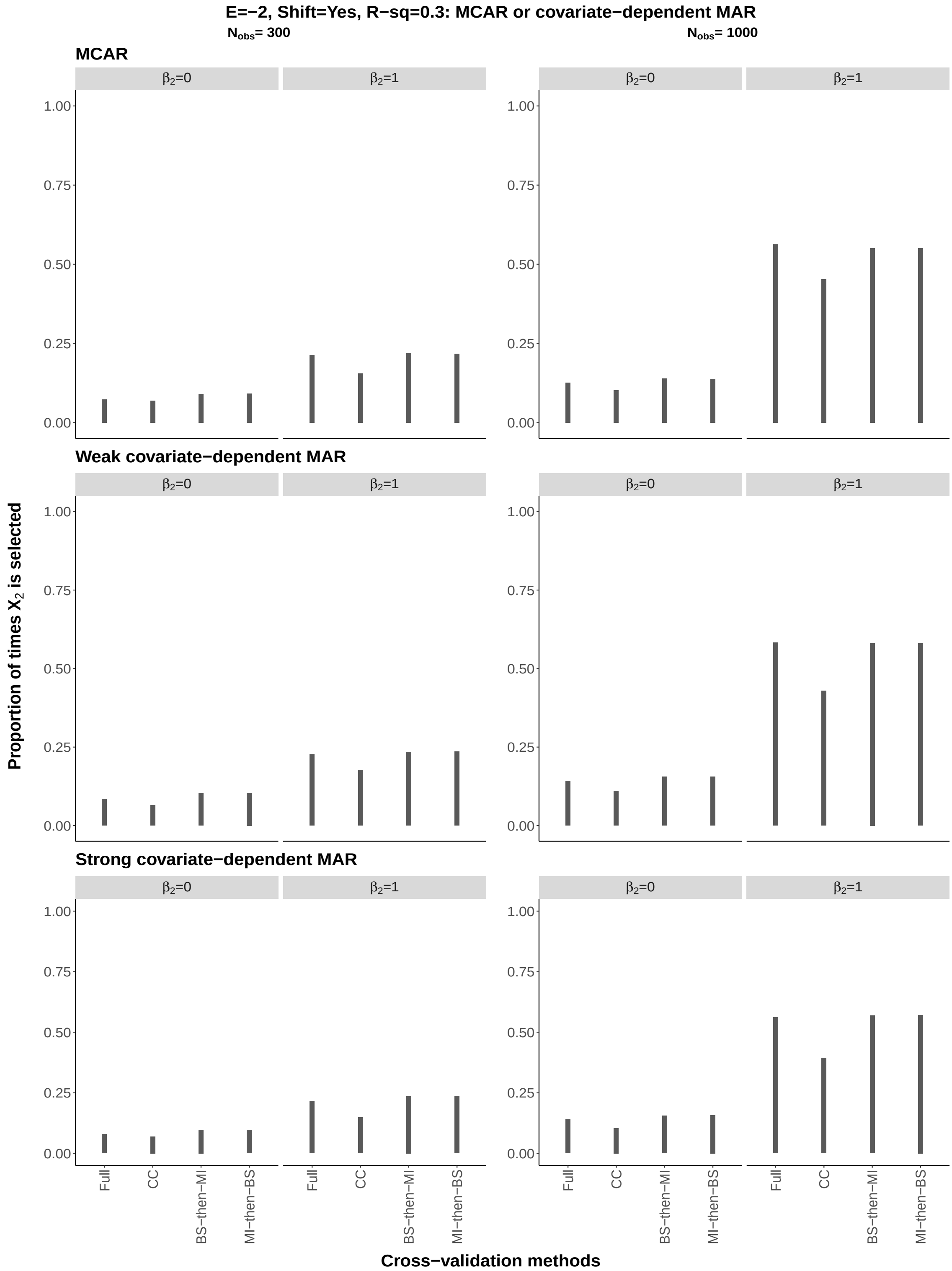

Figure S130: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

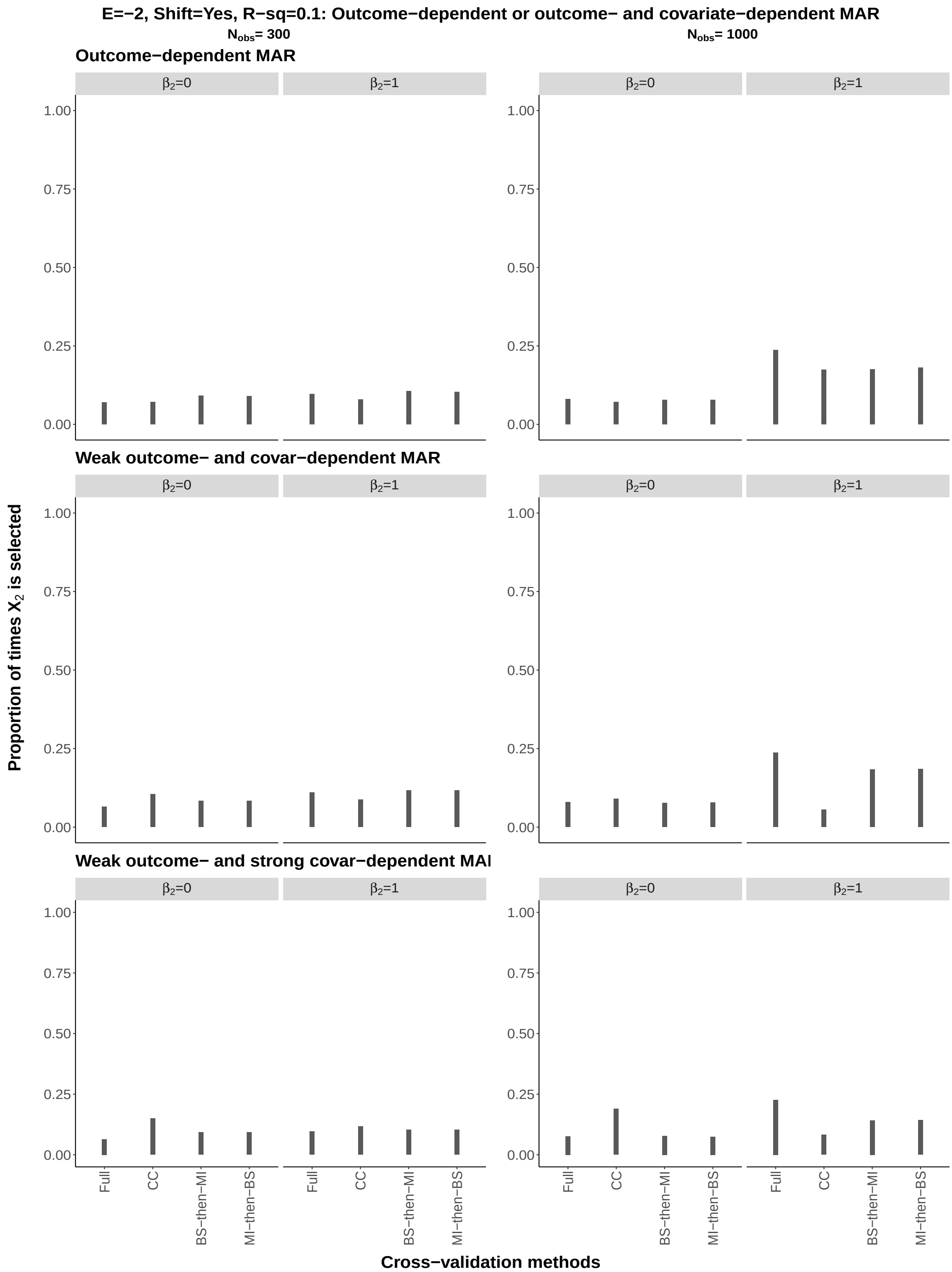

Figure S131: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

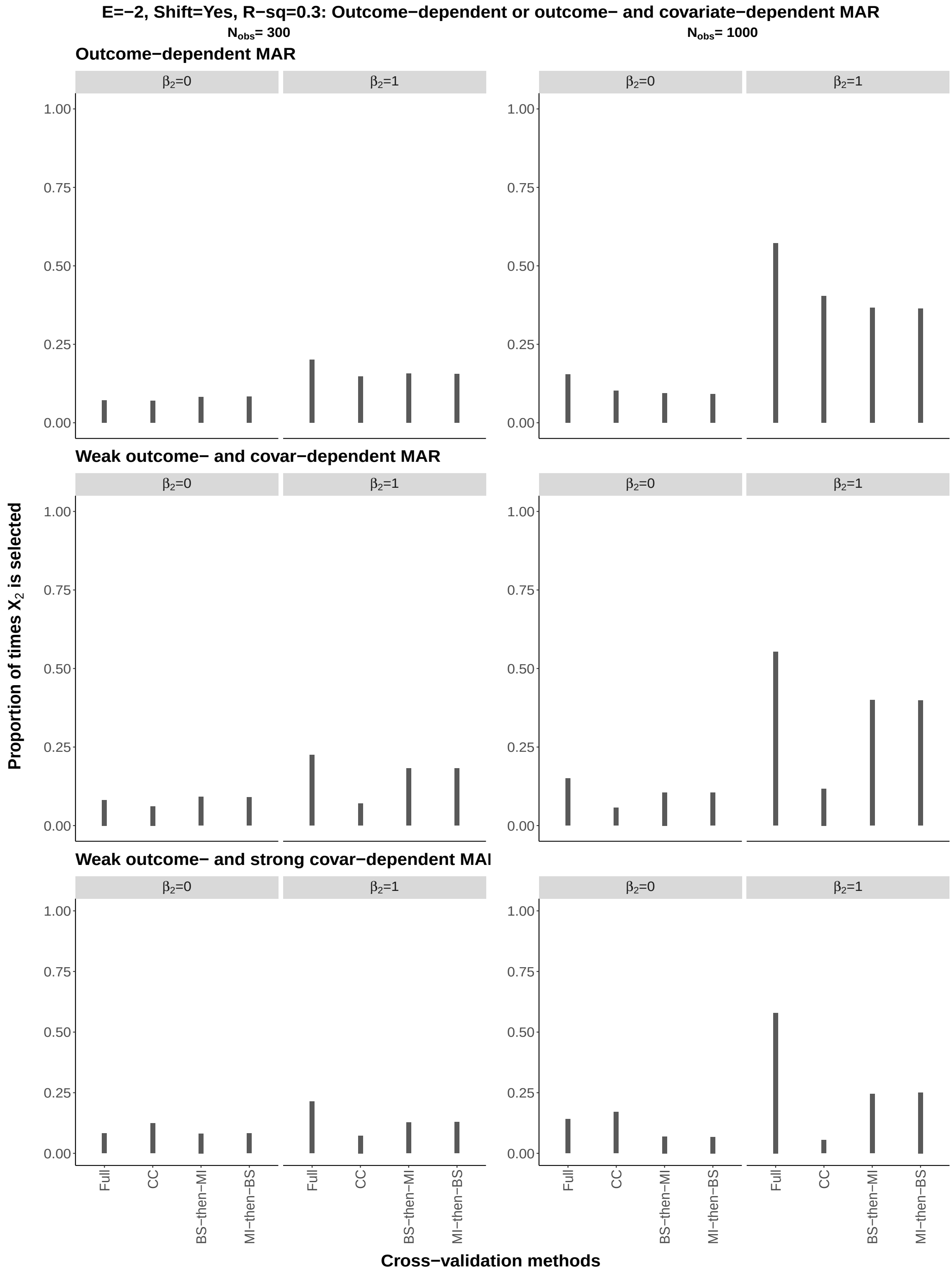

Figure S132: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.4 Covariate selection of X_2 using all data: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

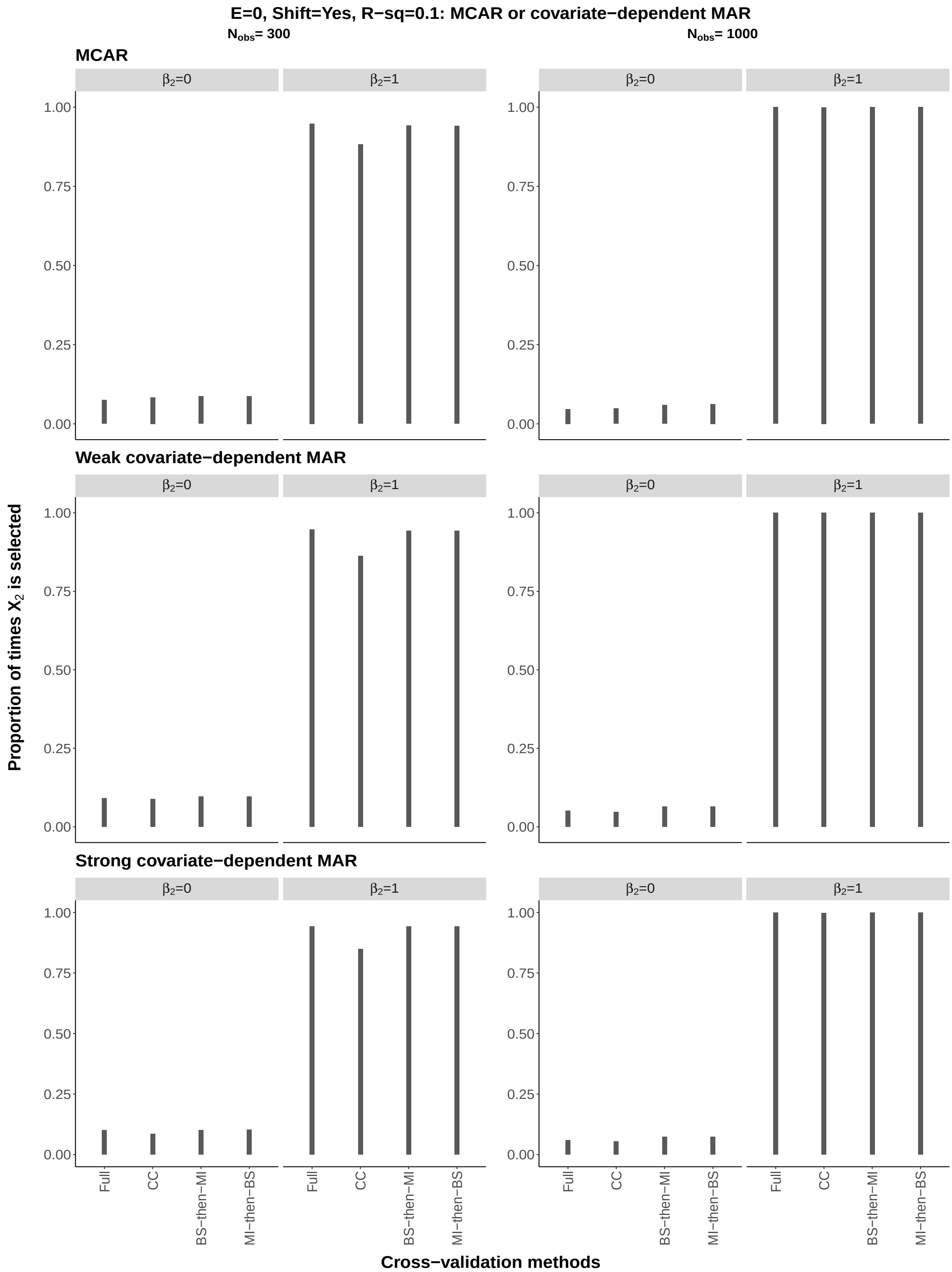

Figure S133: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

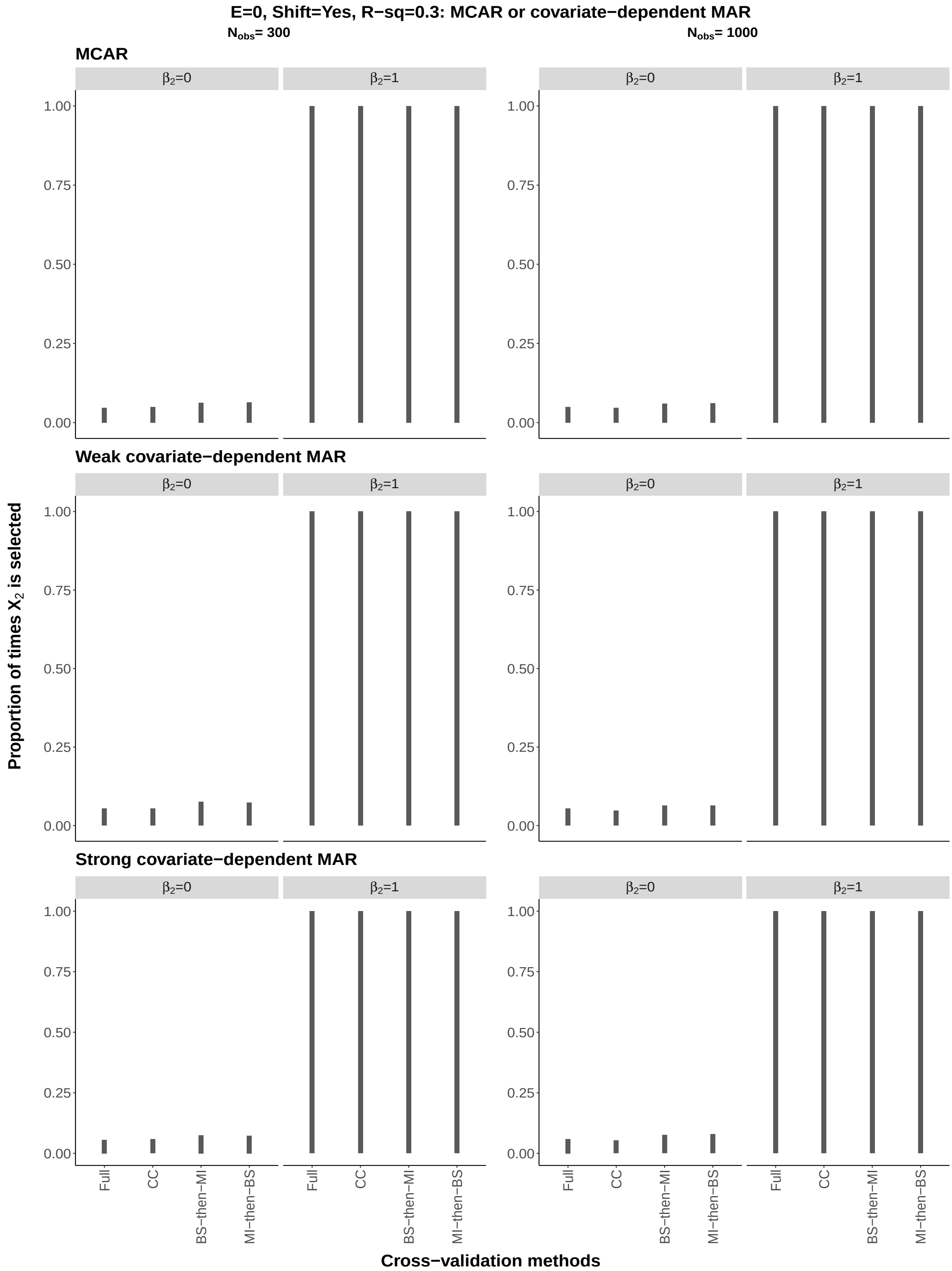

Figure S134: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

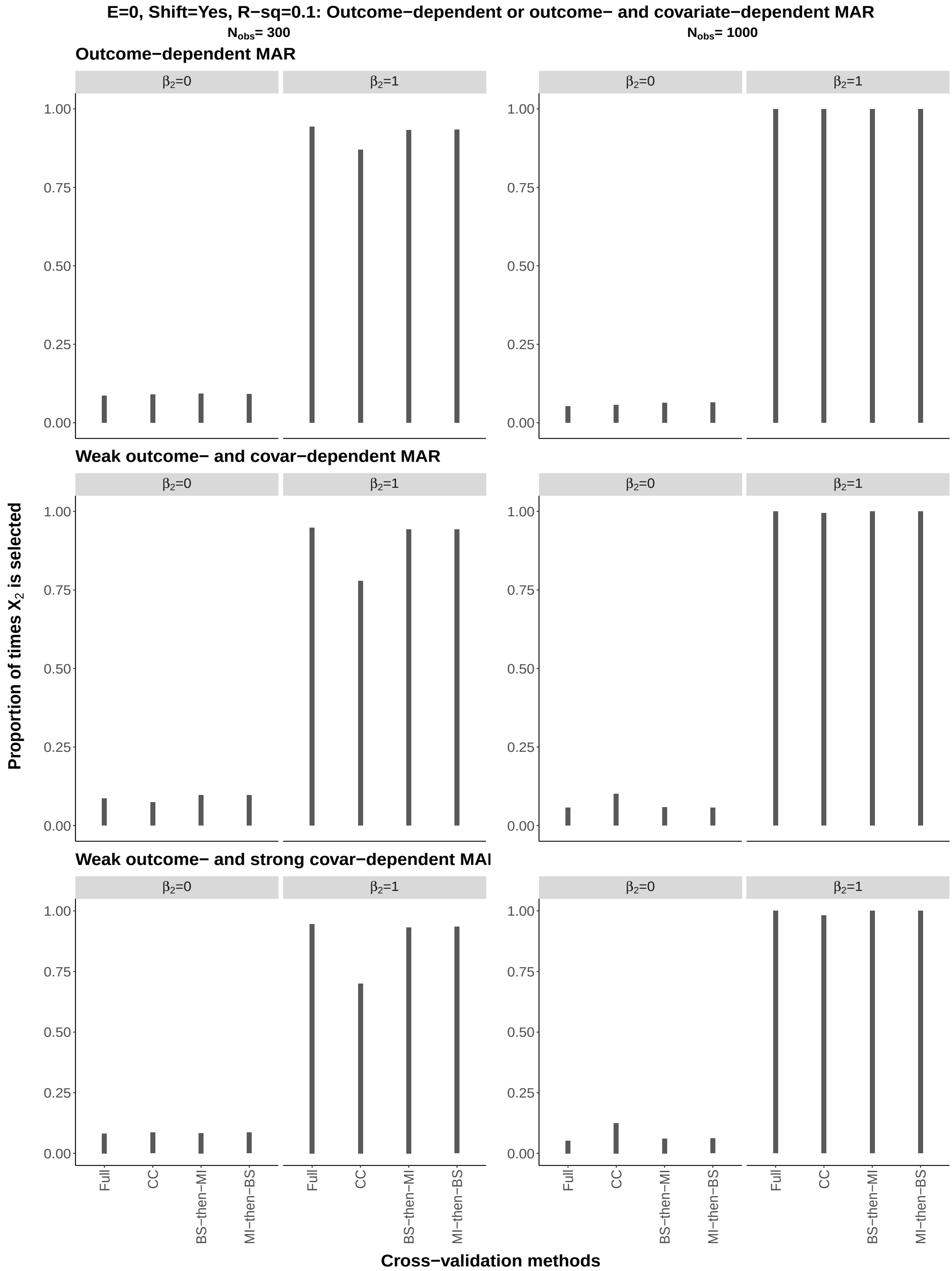

Figure S135: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

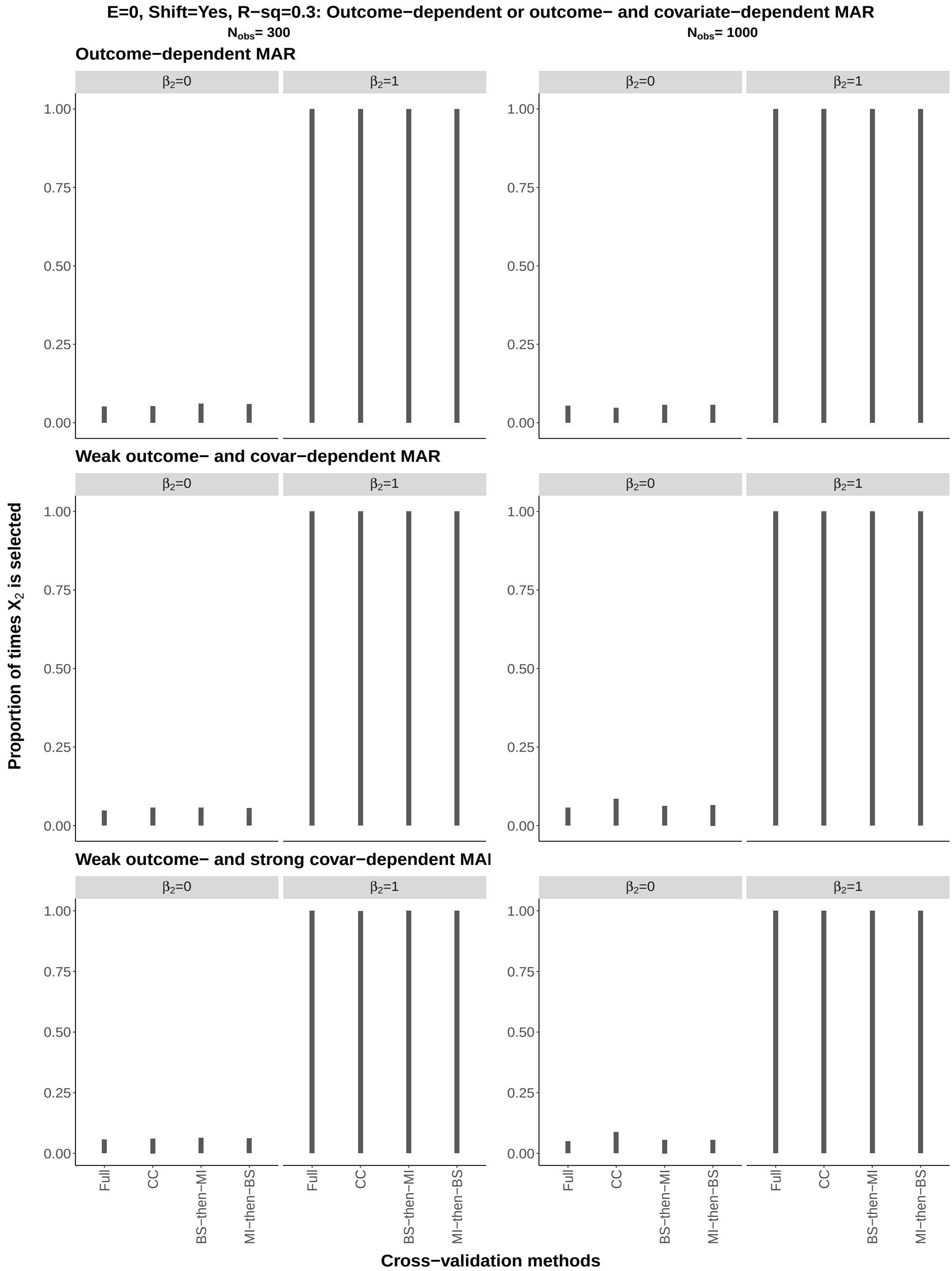

Figure S136: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

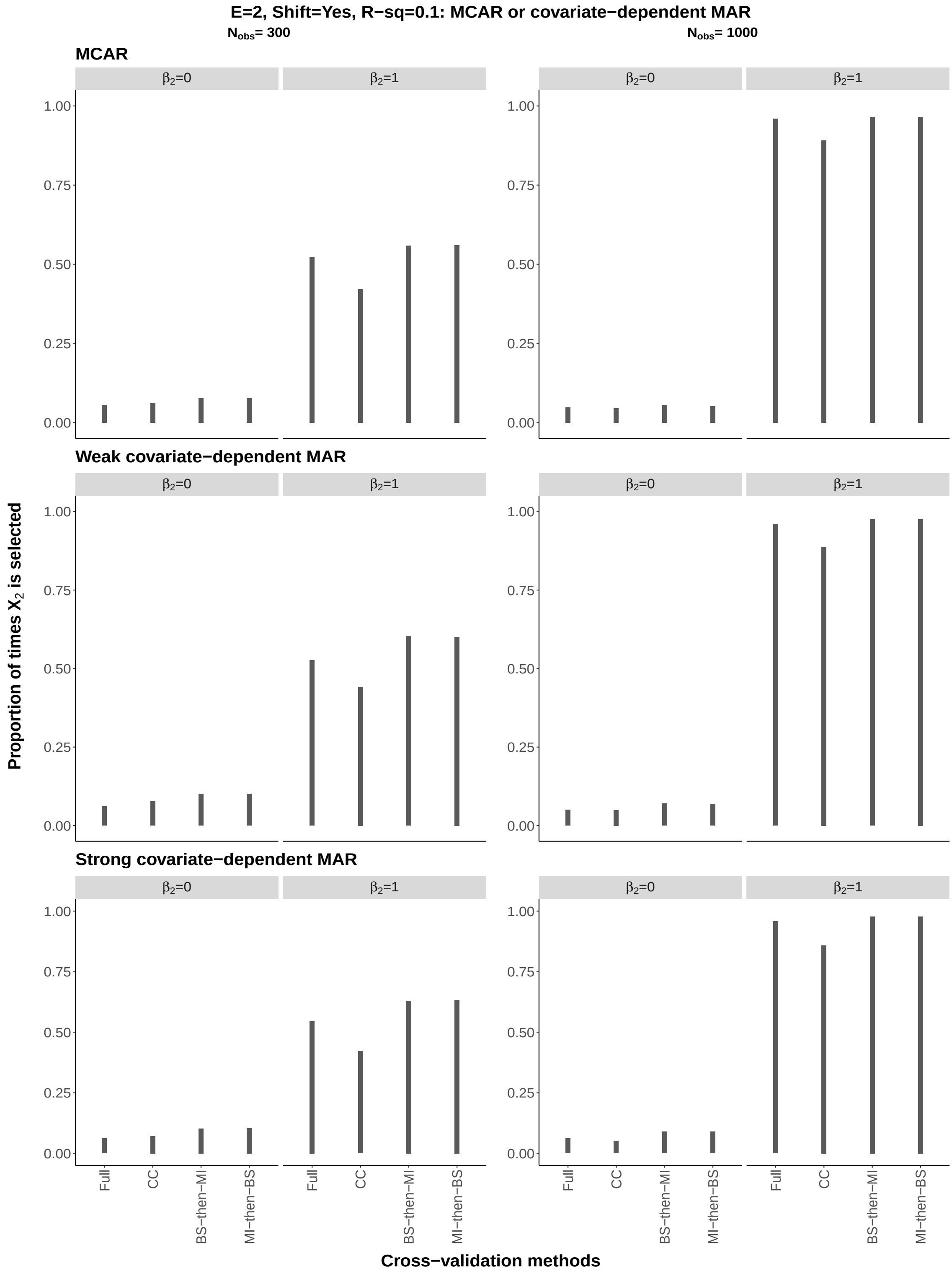

Figure S137: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

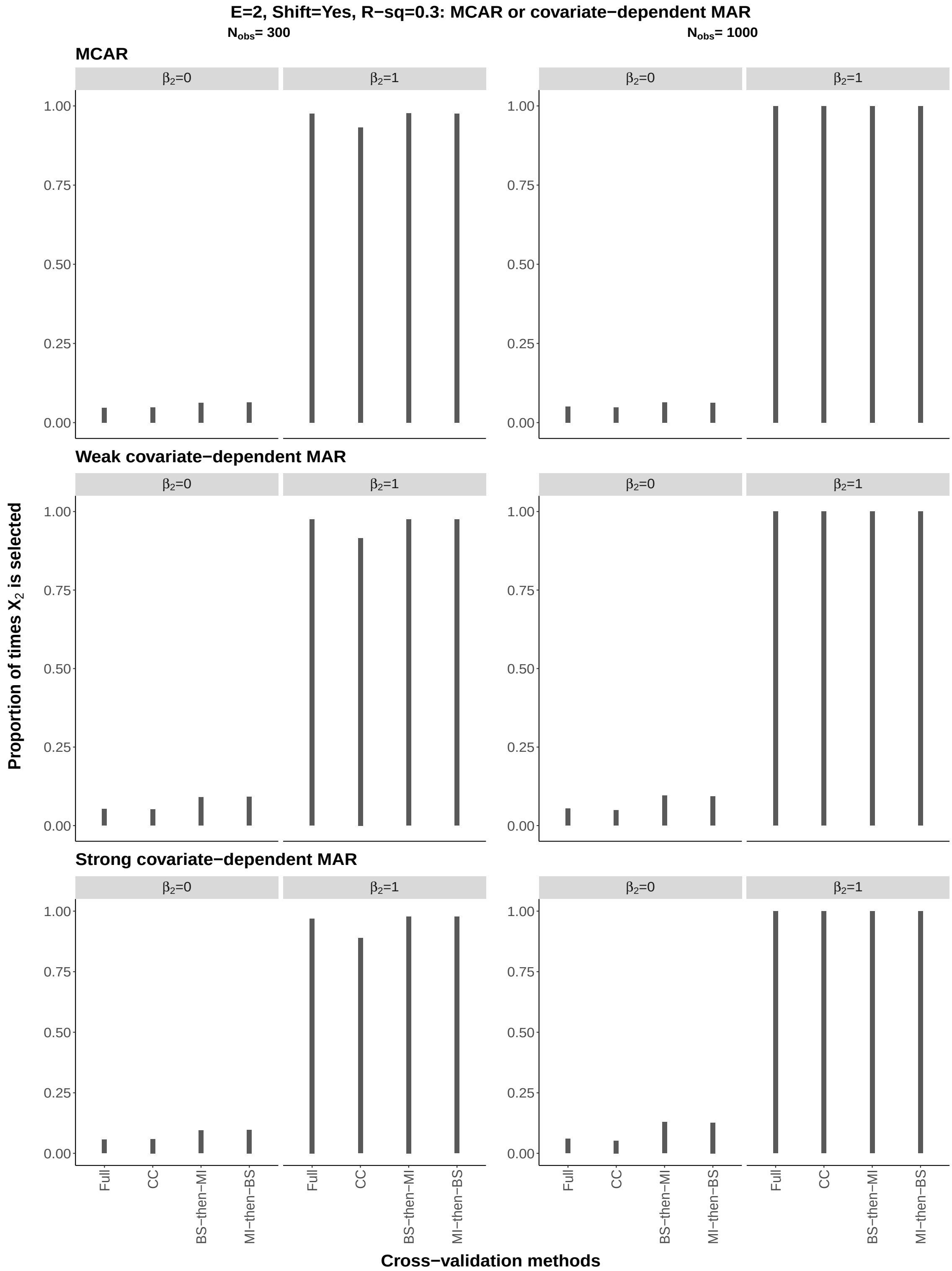

Figure S138: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

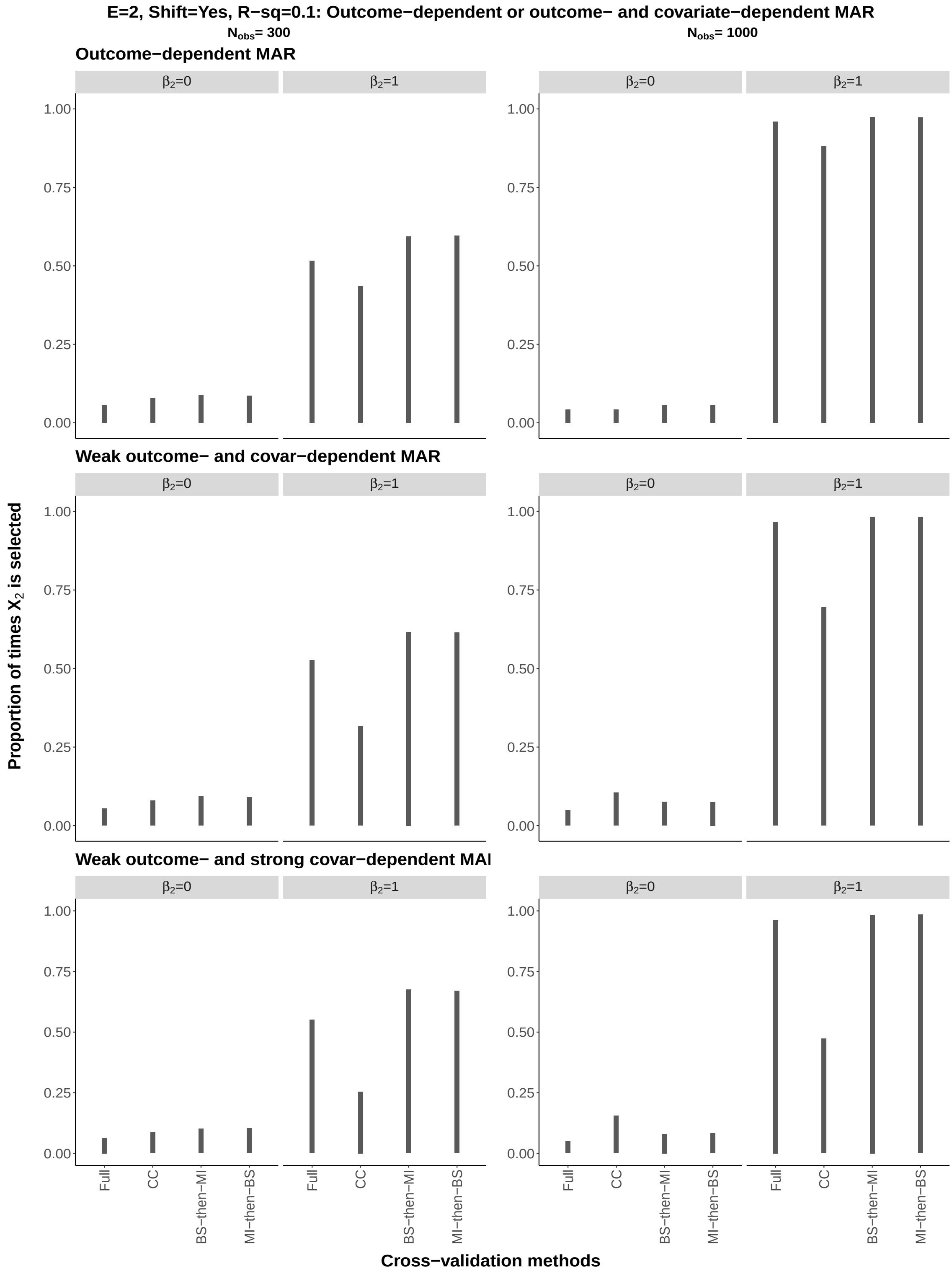

Figure S139: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

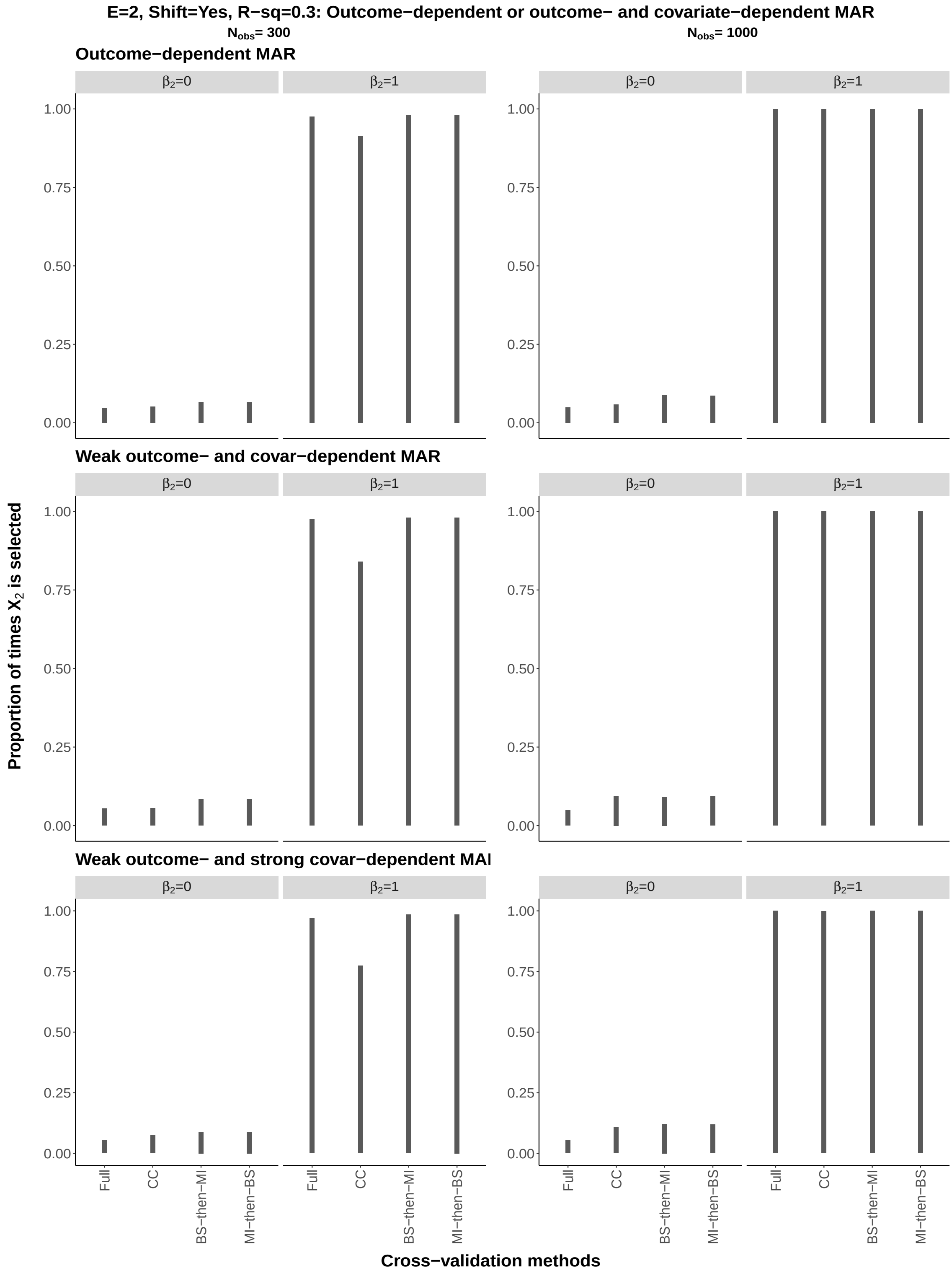

Figure S140: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

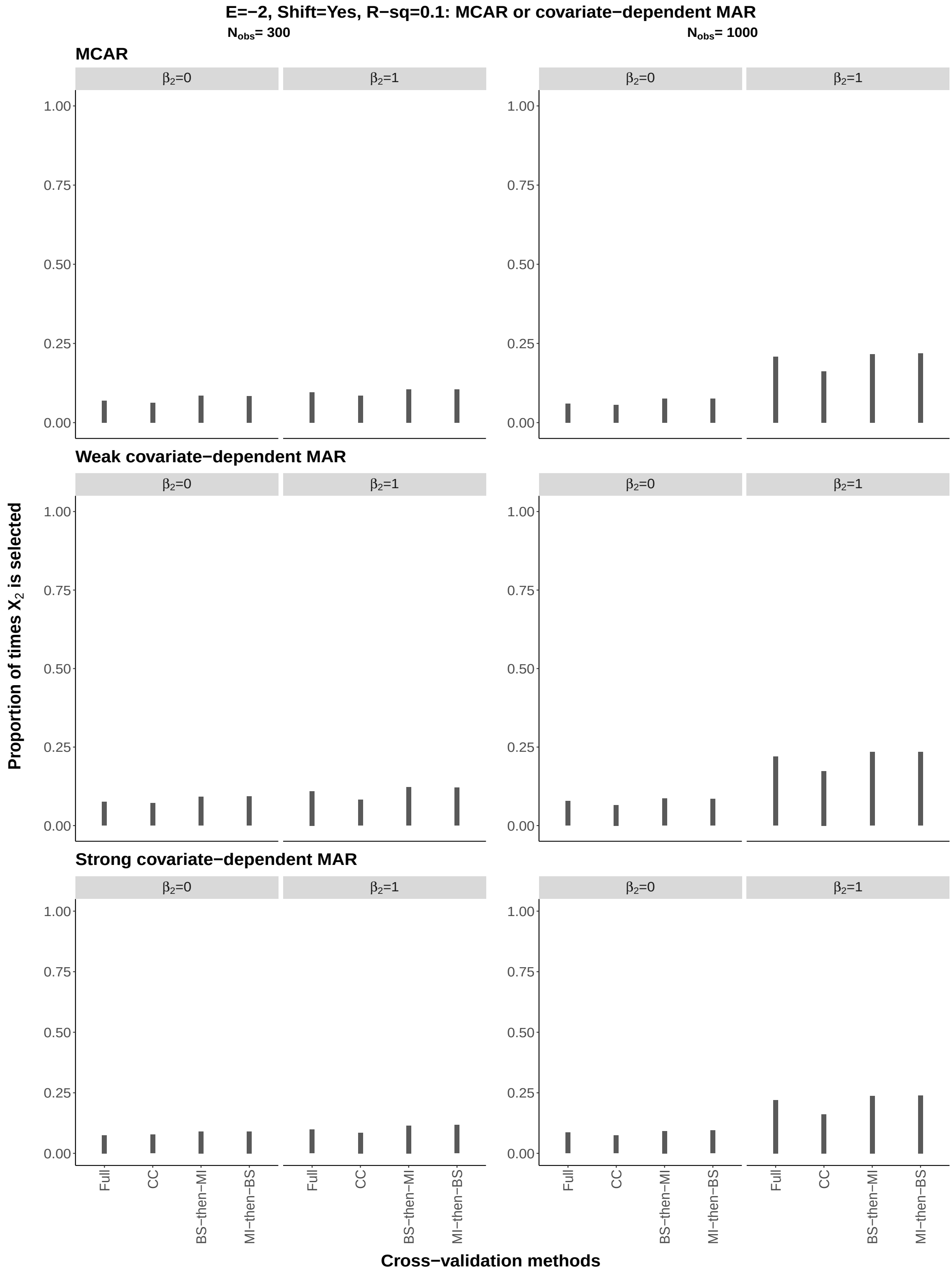

Figure S141: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

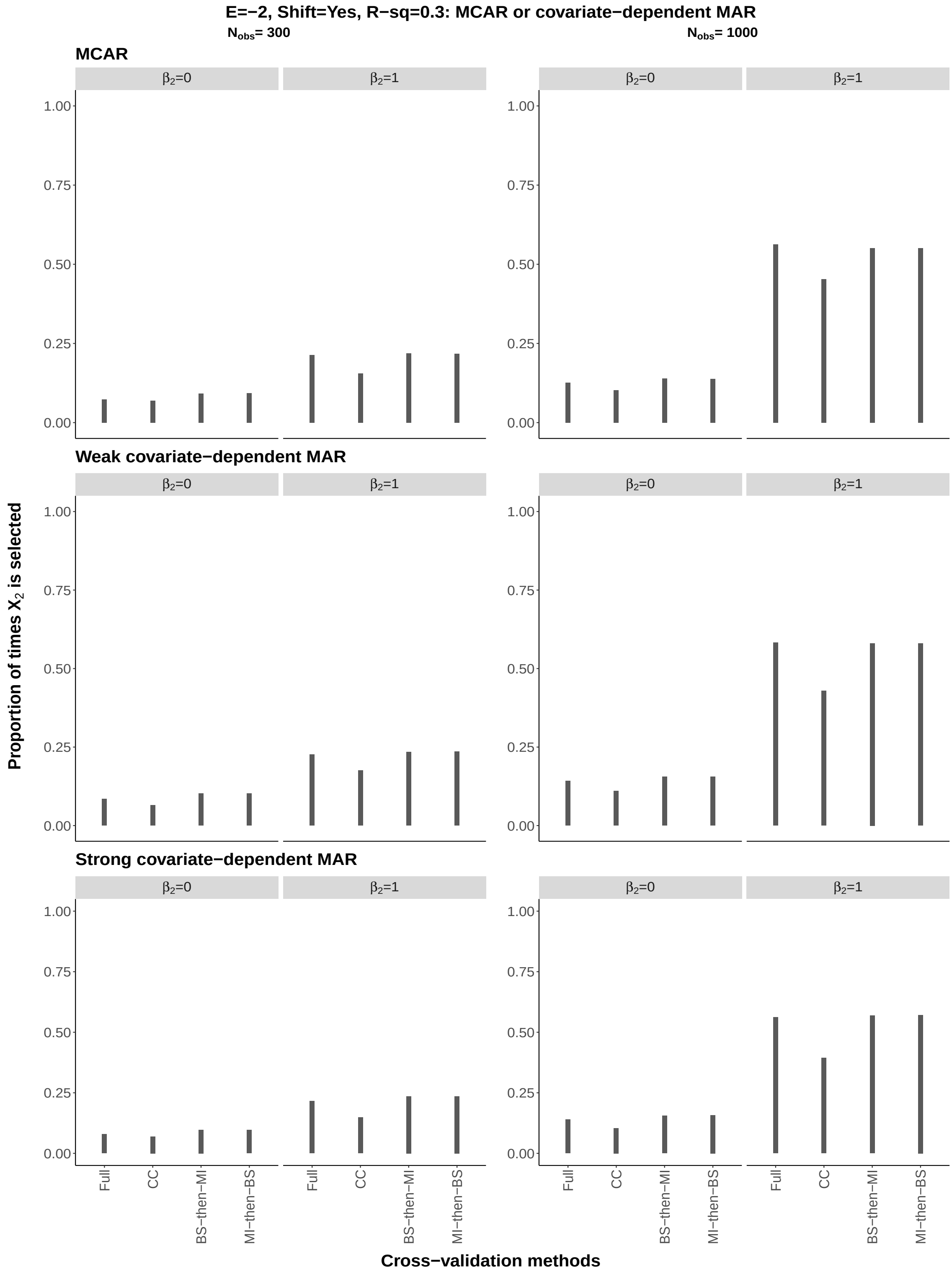

Figure S142: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

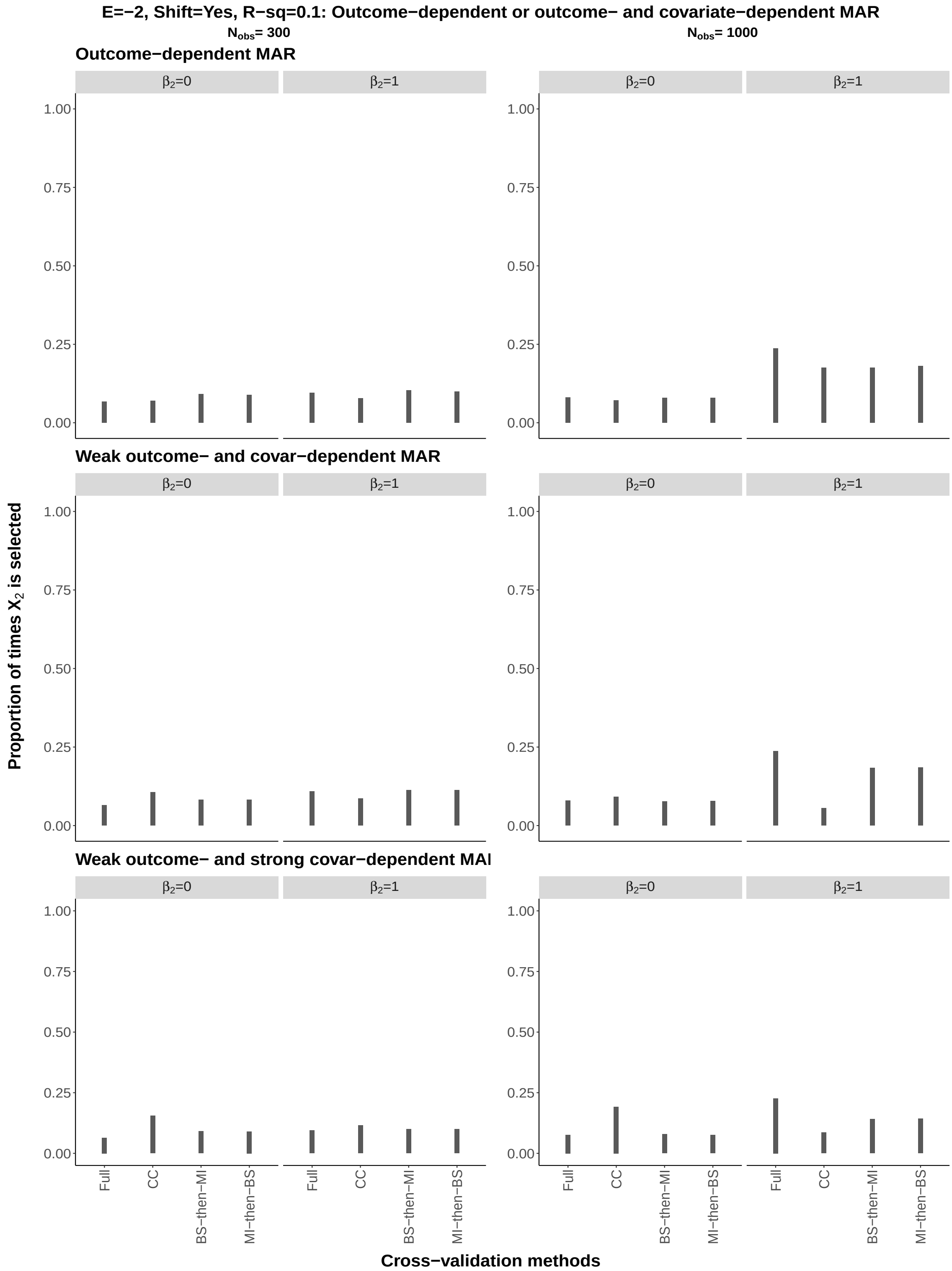

Figure S143: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

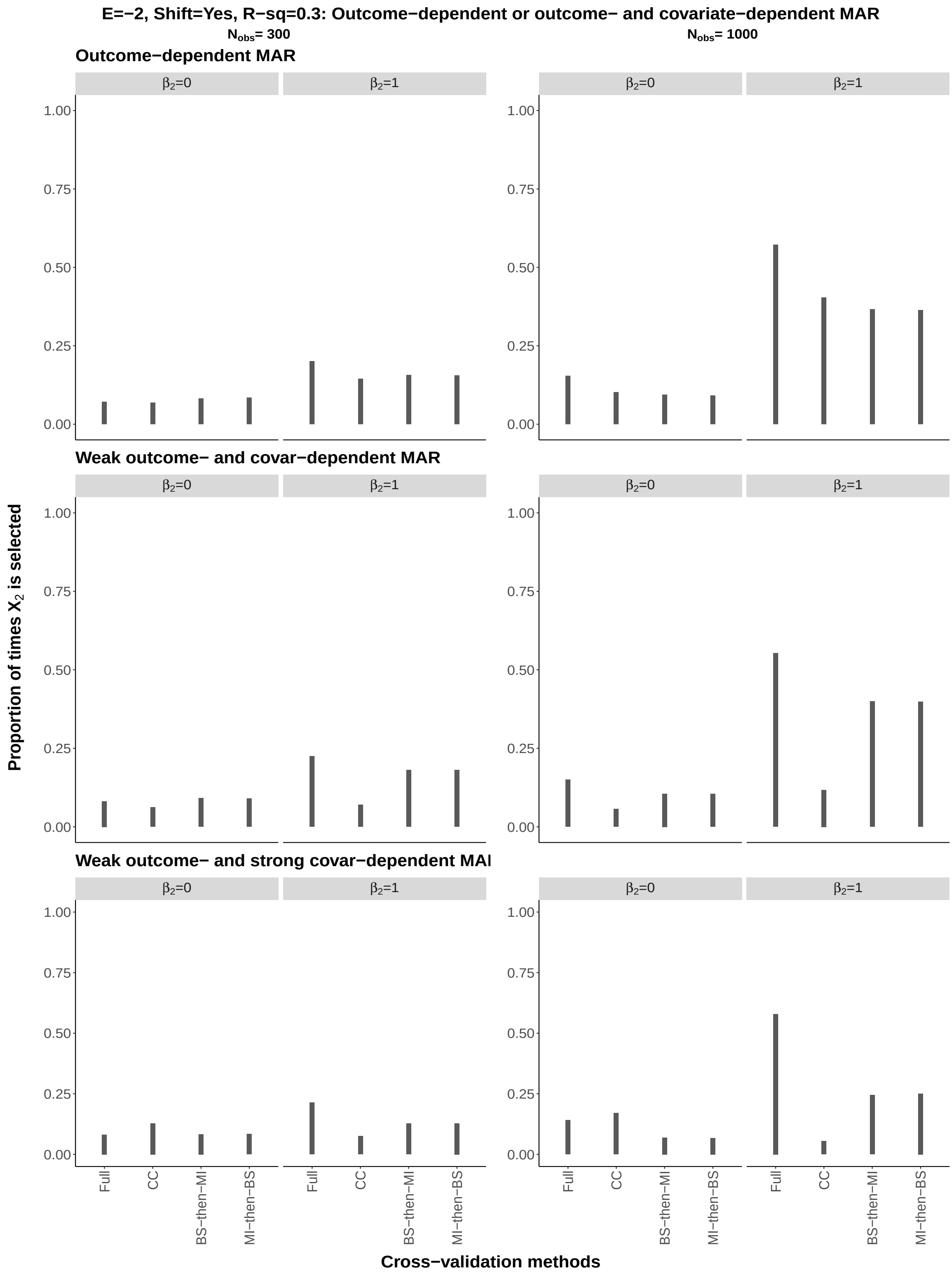

Figure S144: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.5 Covariate selection of X_1 using all data: $\beta_2 = 1$, $\alpha_E = 1$, no origin-shift

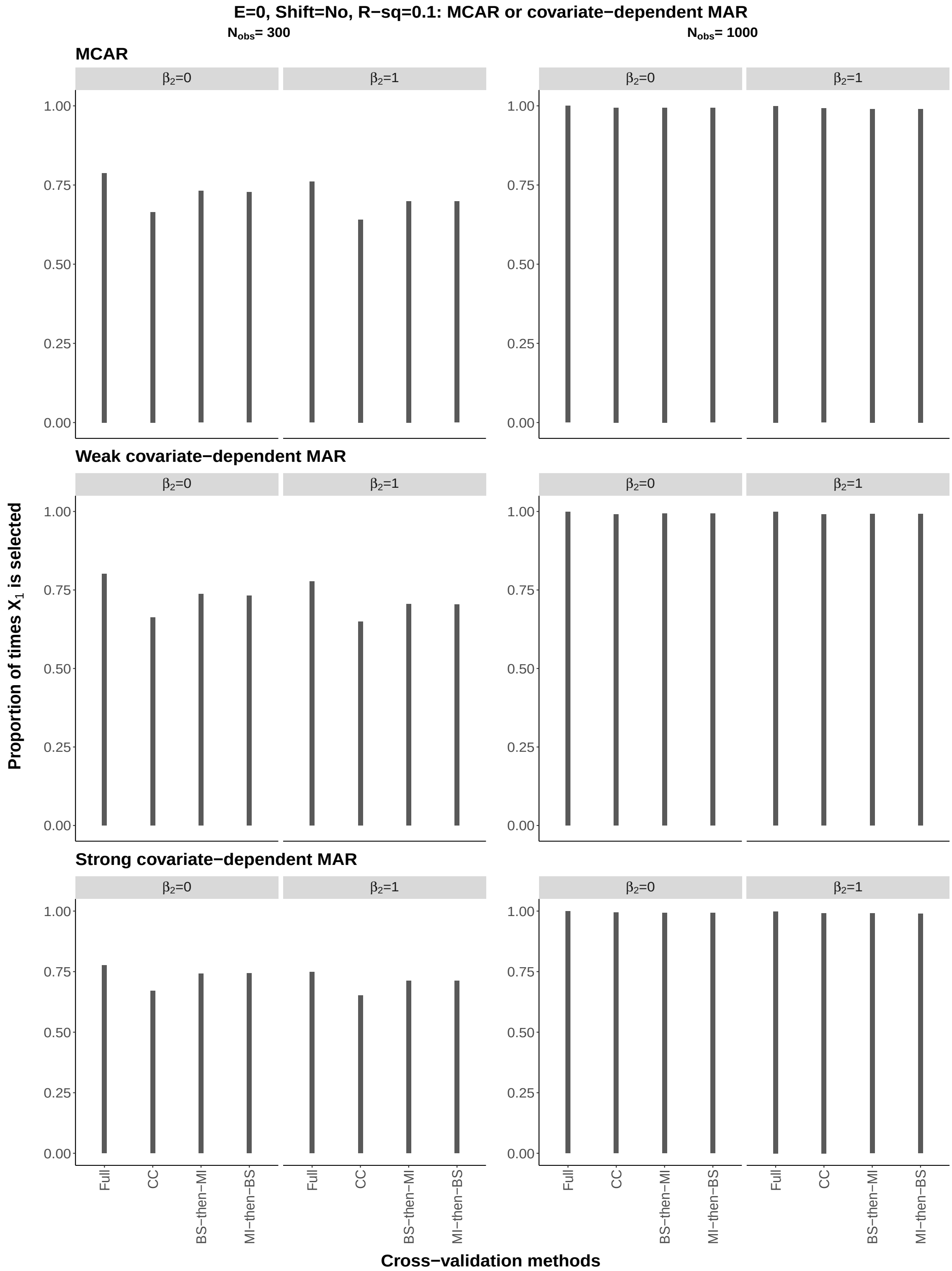

Figure S145: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

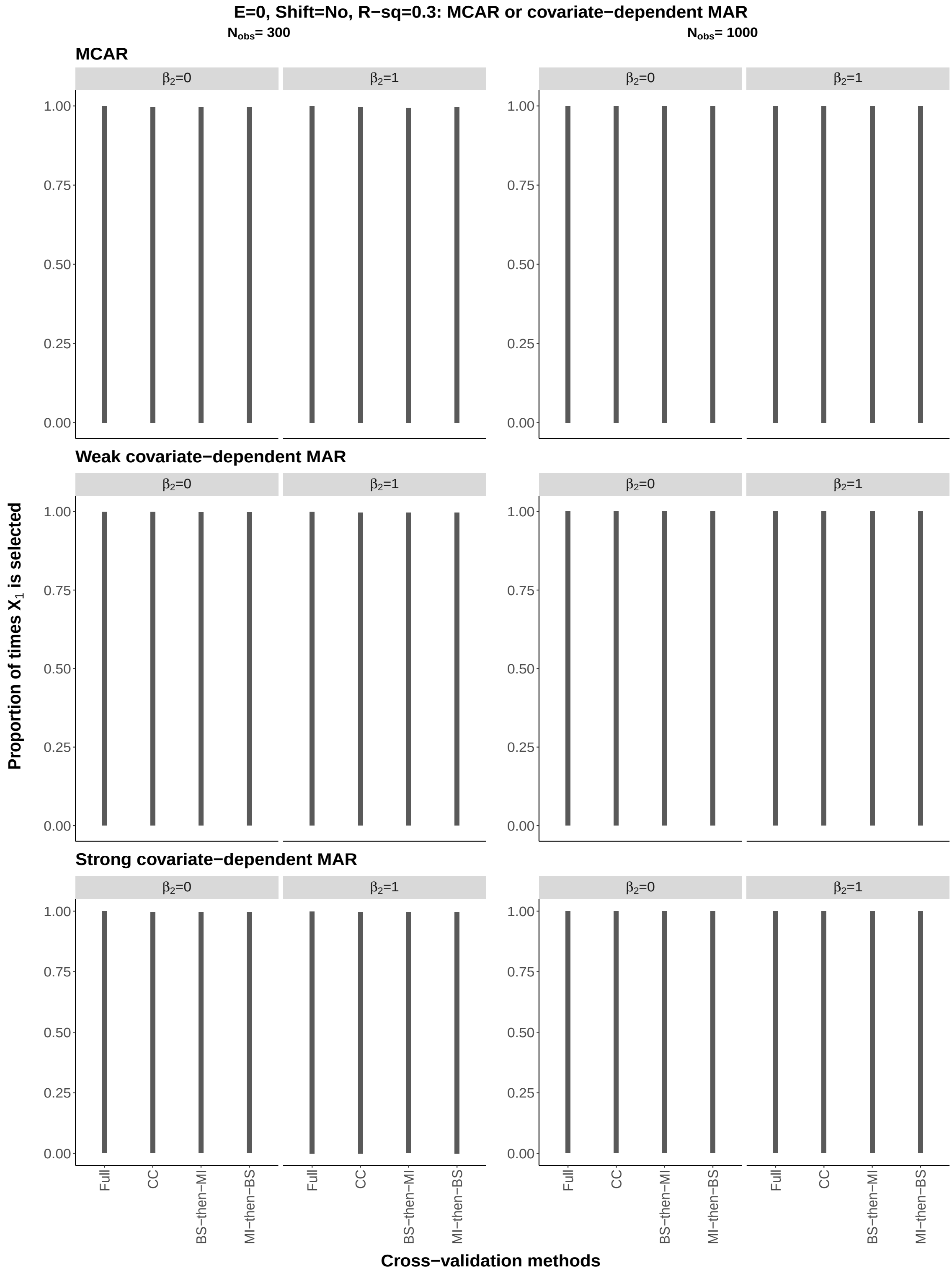

Figure S146: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

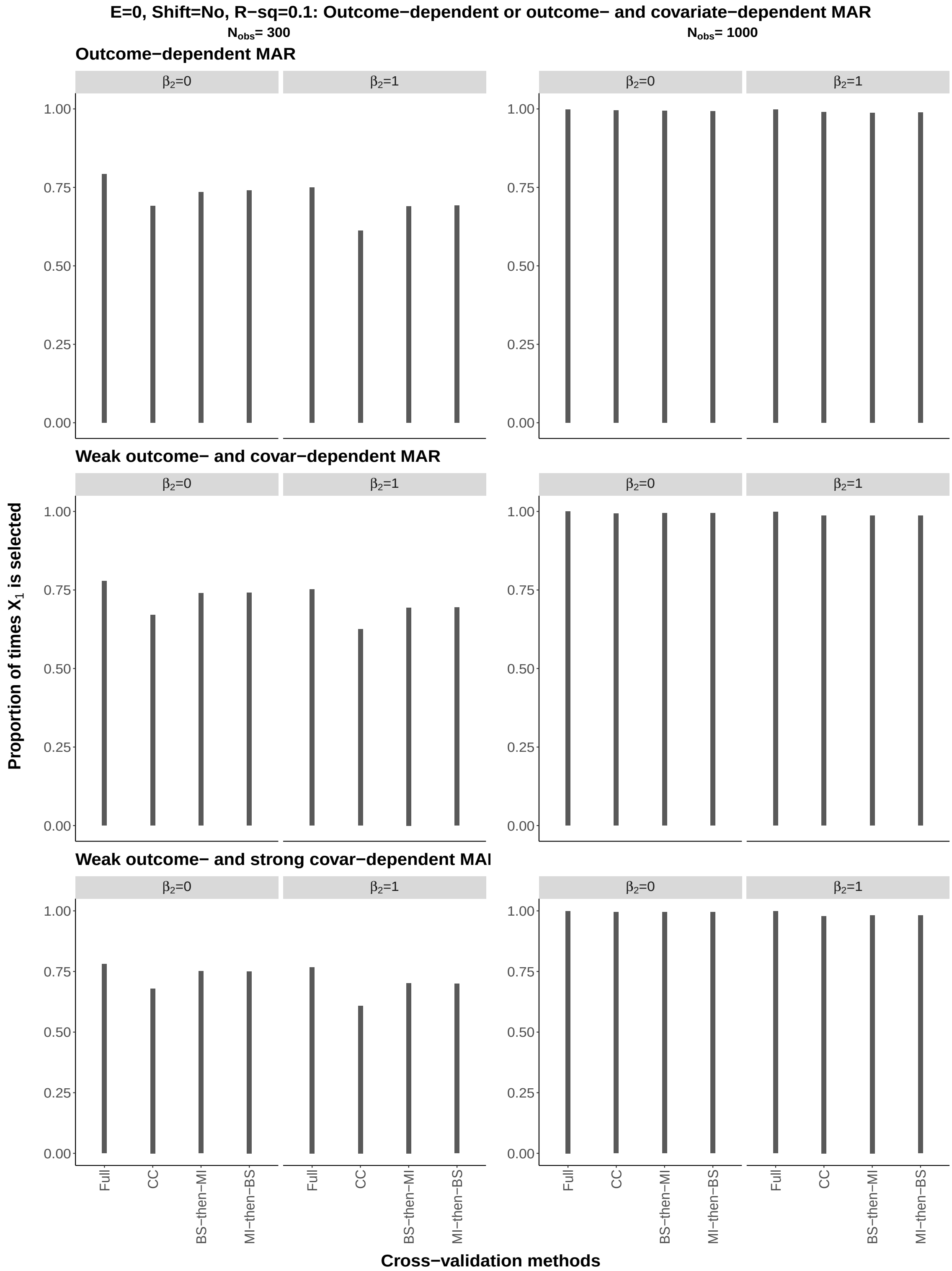

Figure S147: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

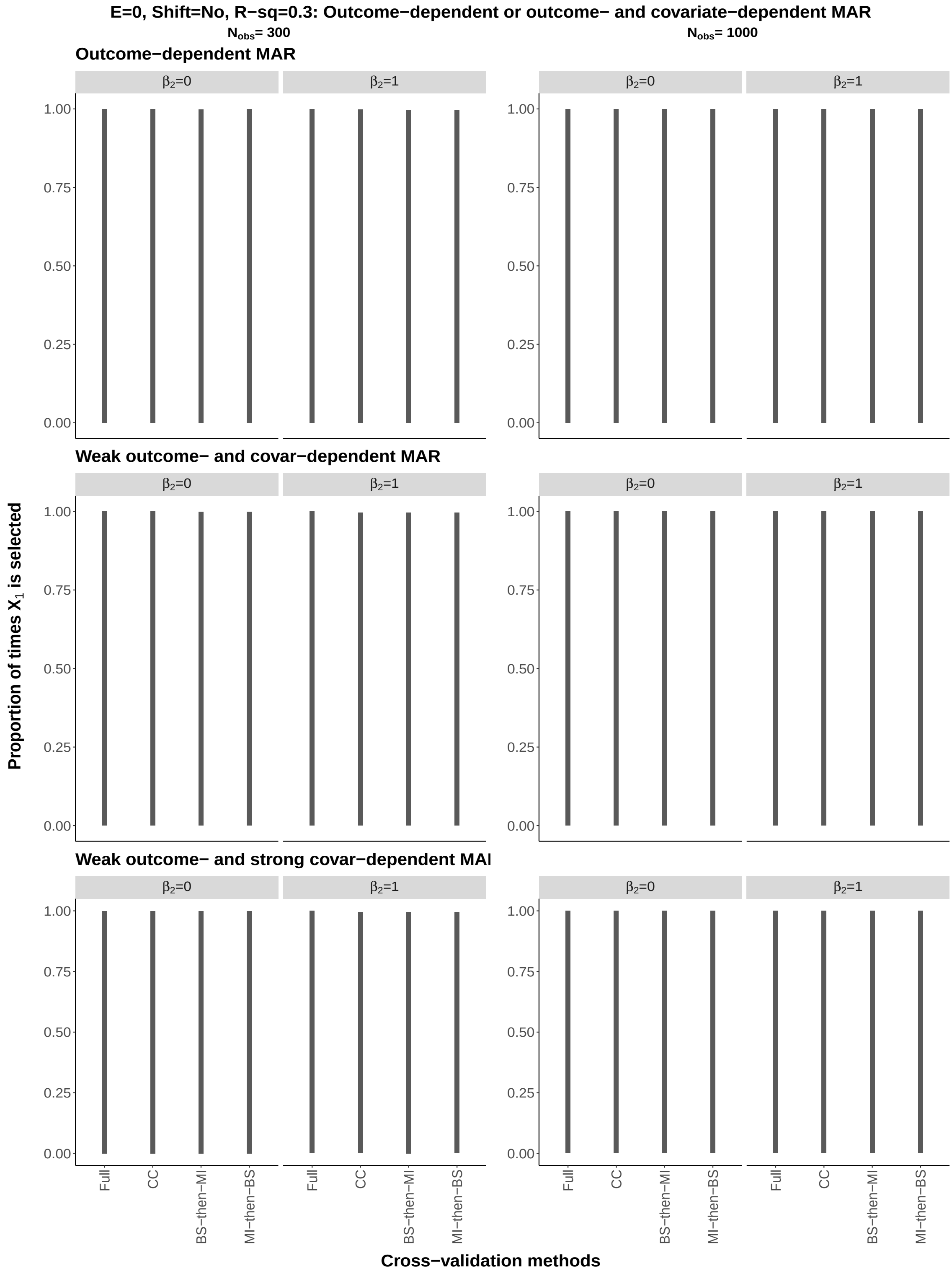

Figure S148: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

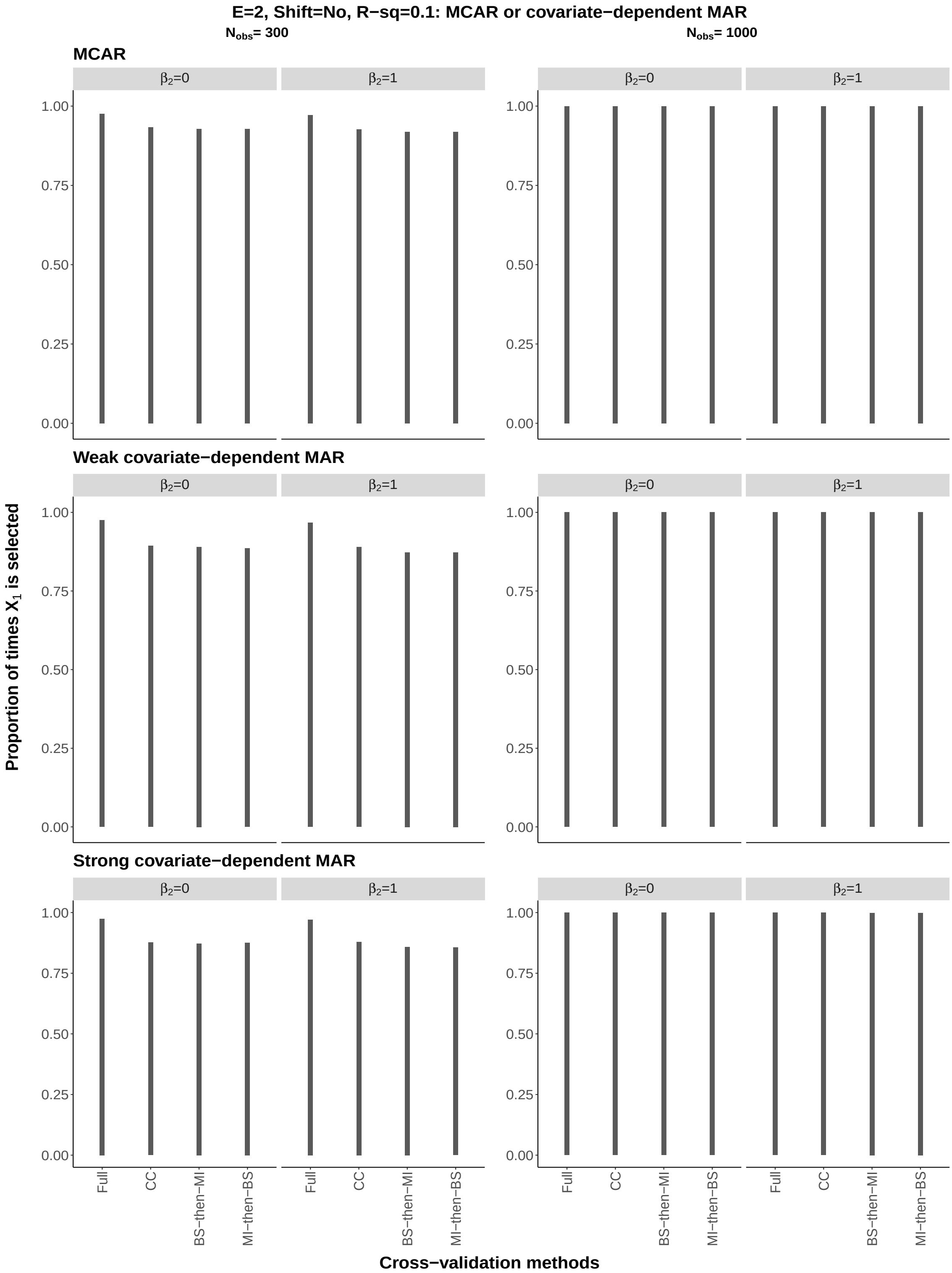

Figure S149: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

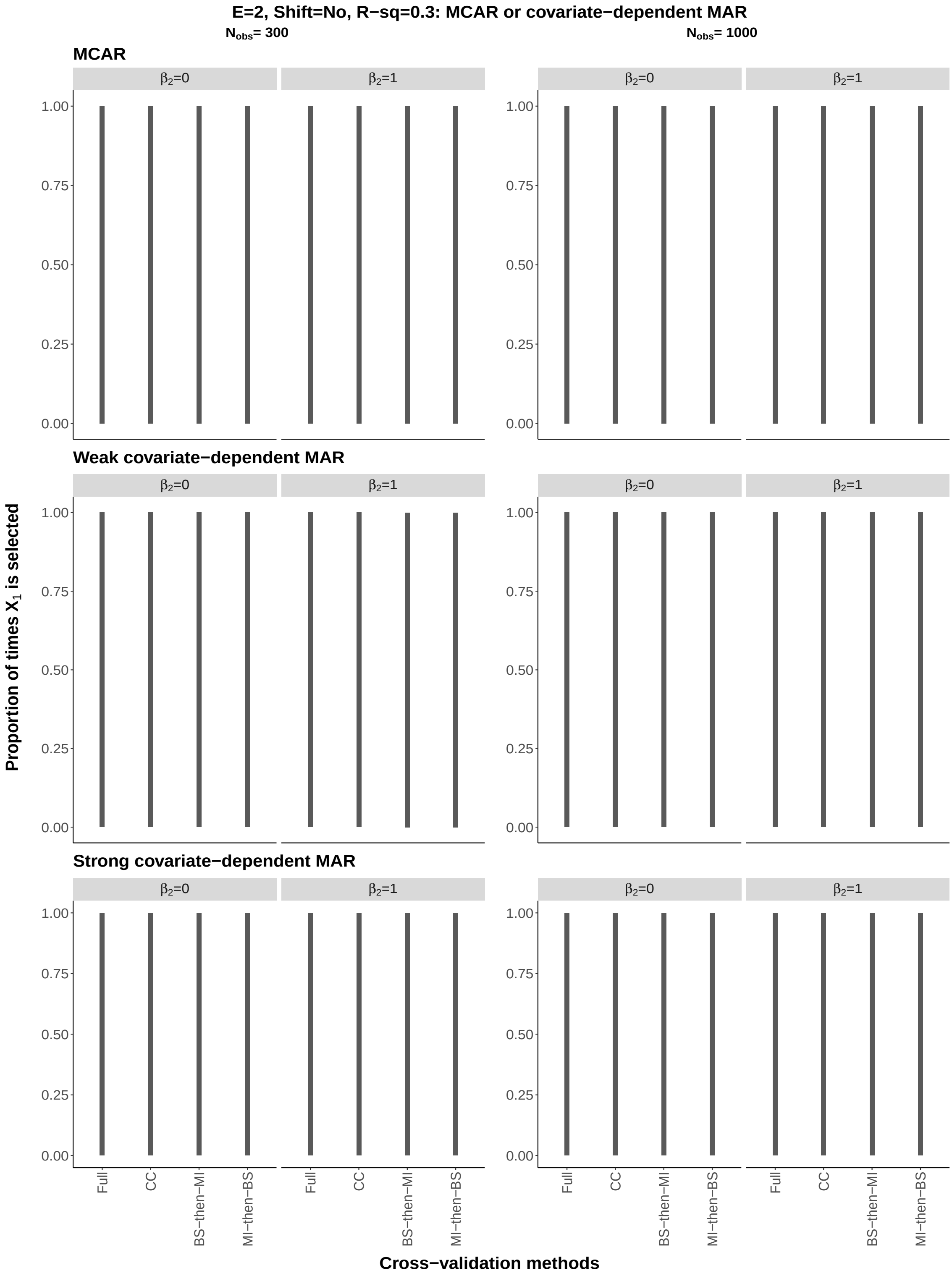

Figure S150: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

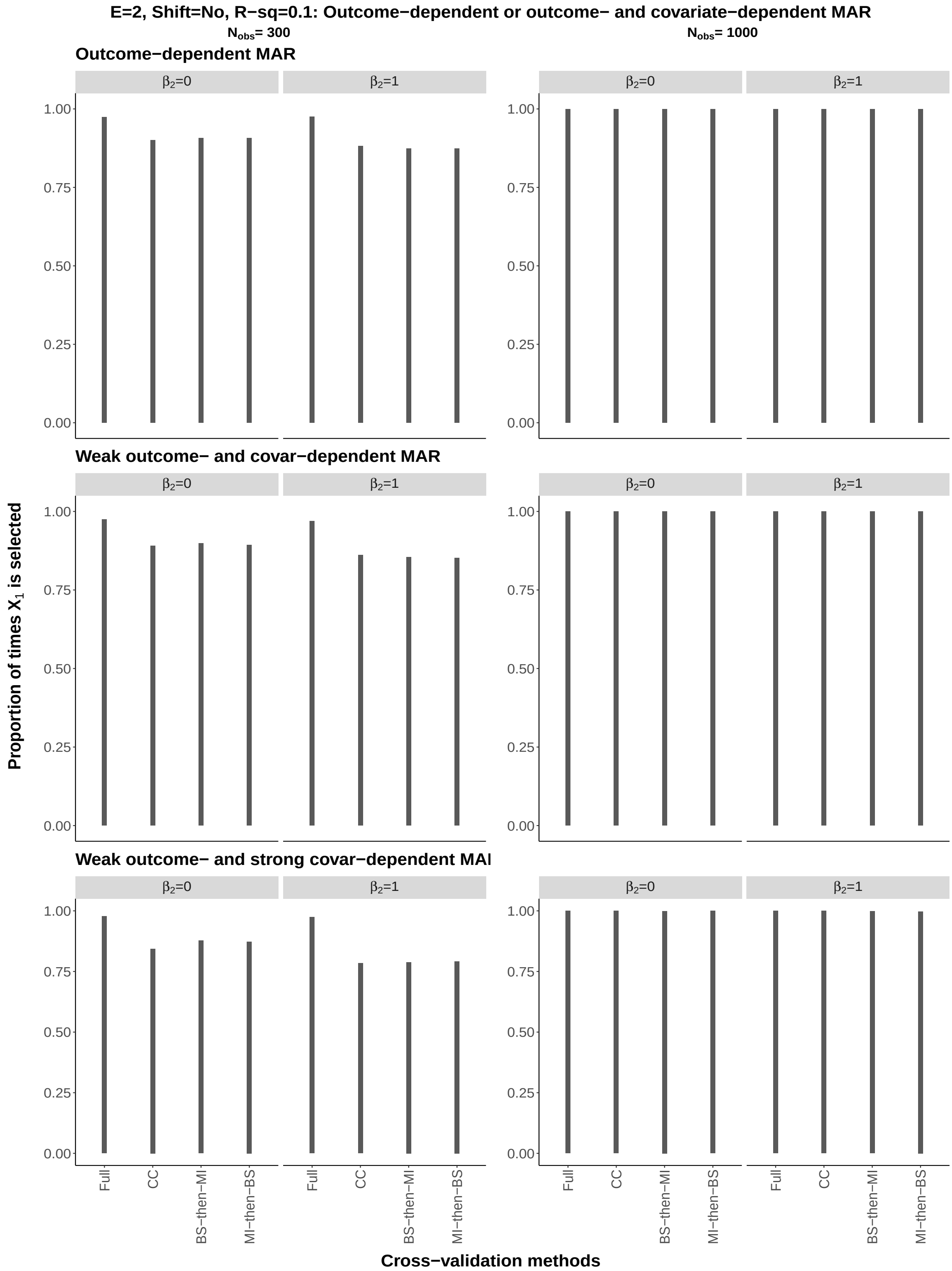

Figure S151: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

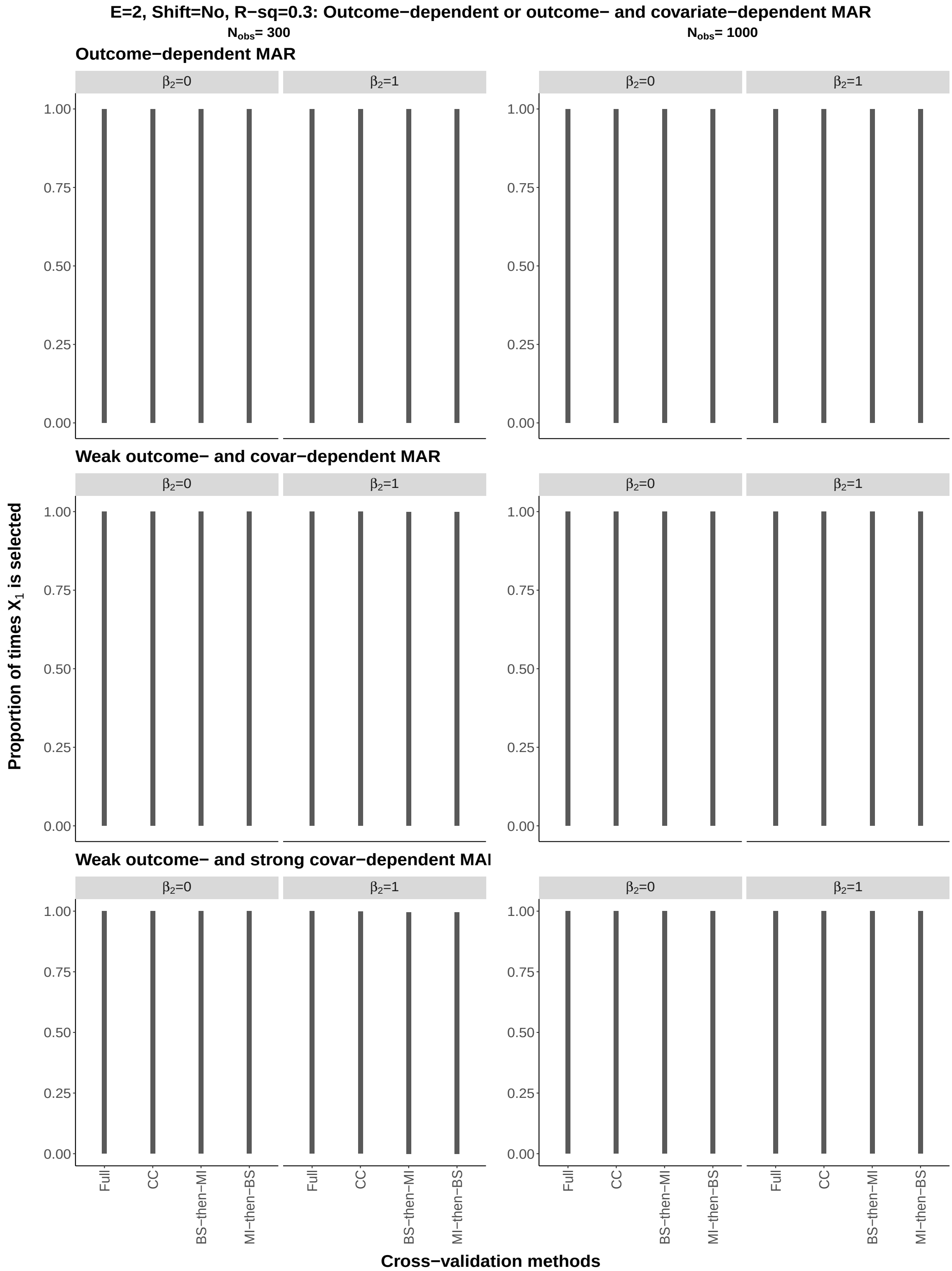

Figure S152: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

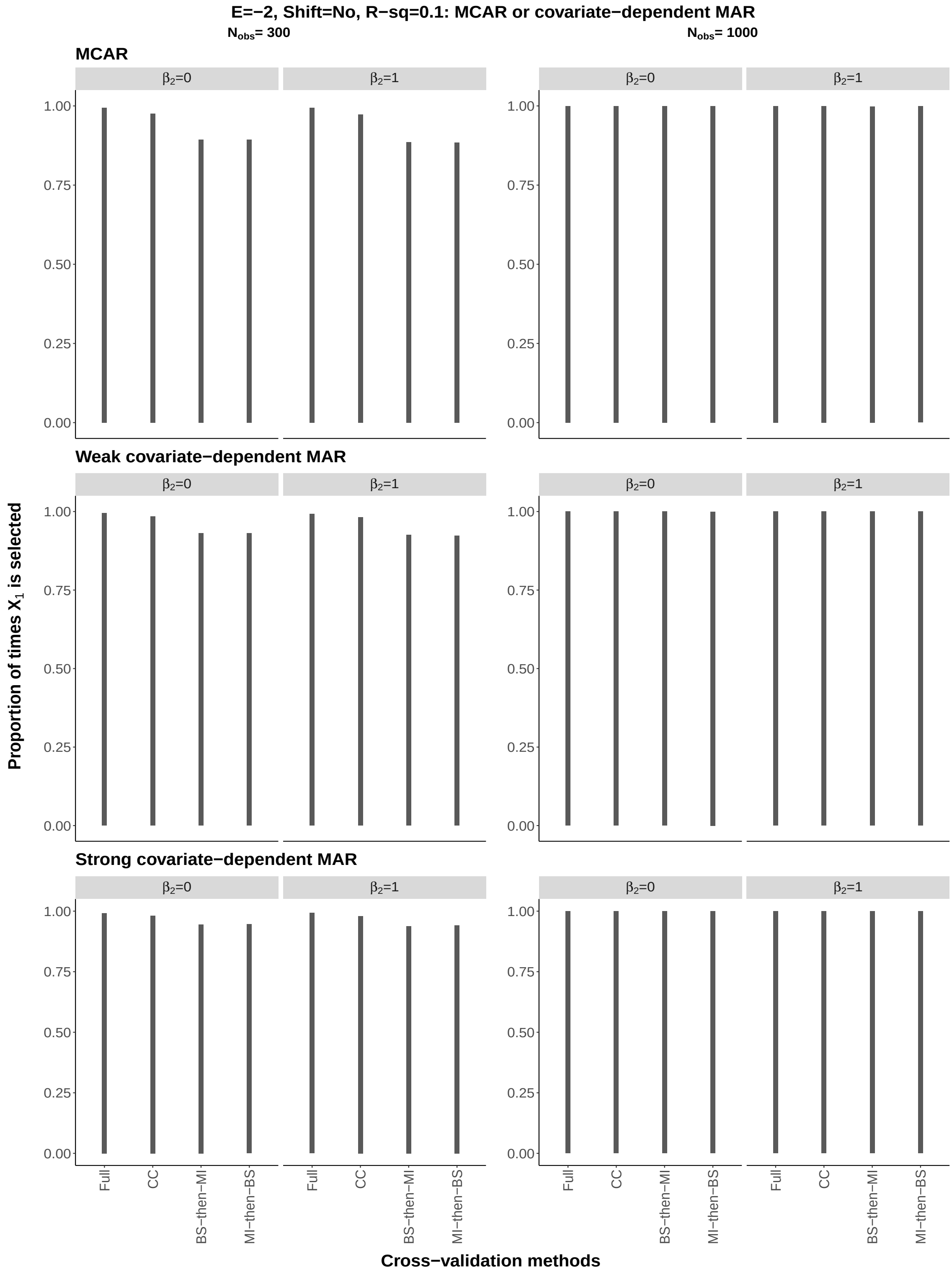

Figure S153: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

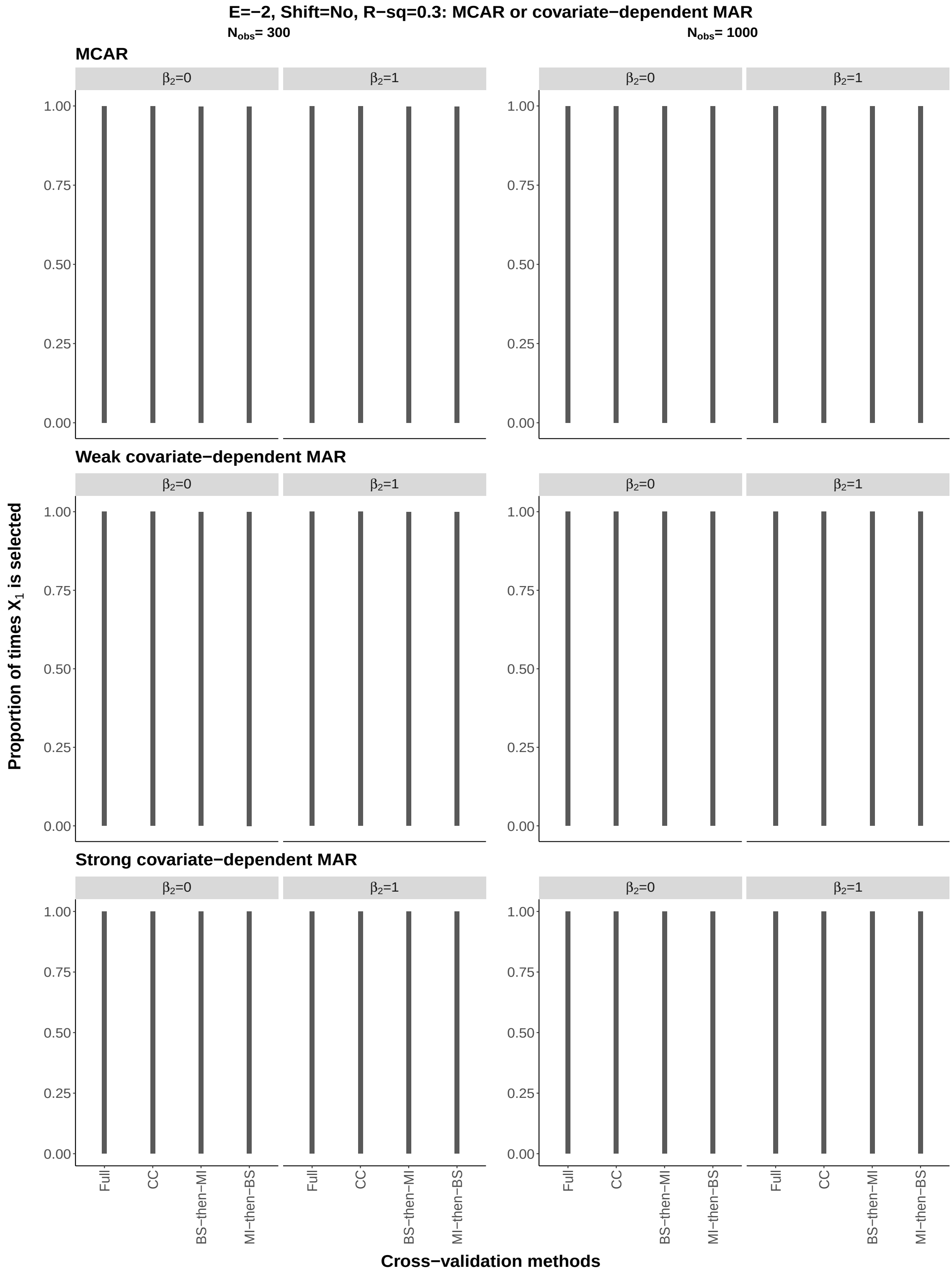

Figure S154: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

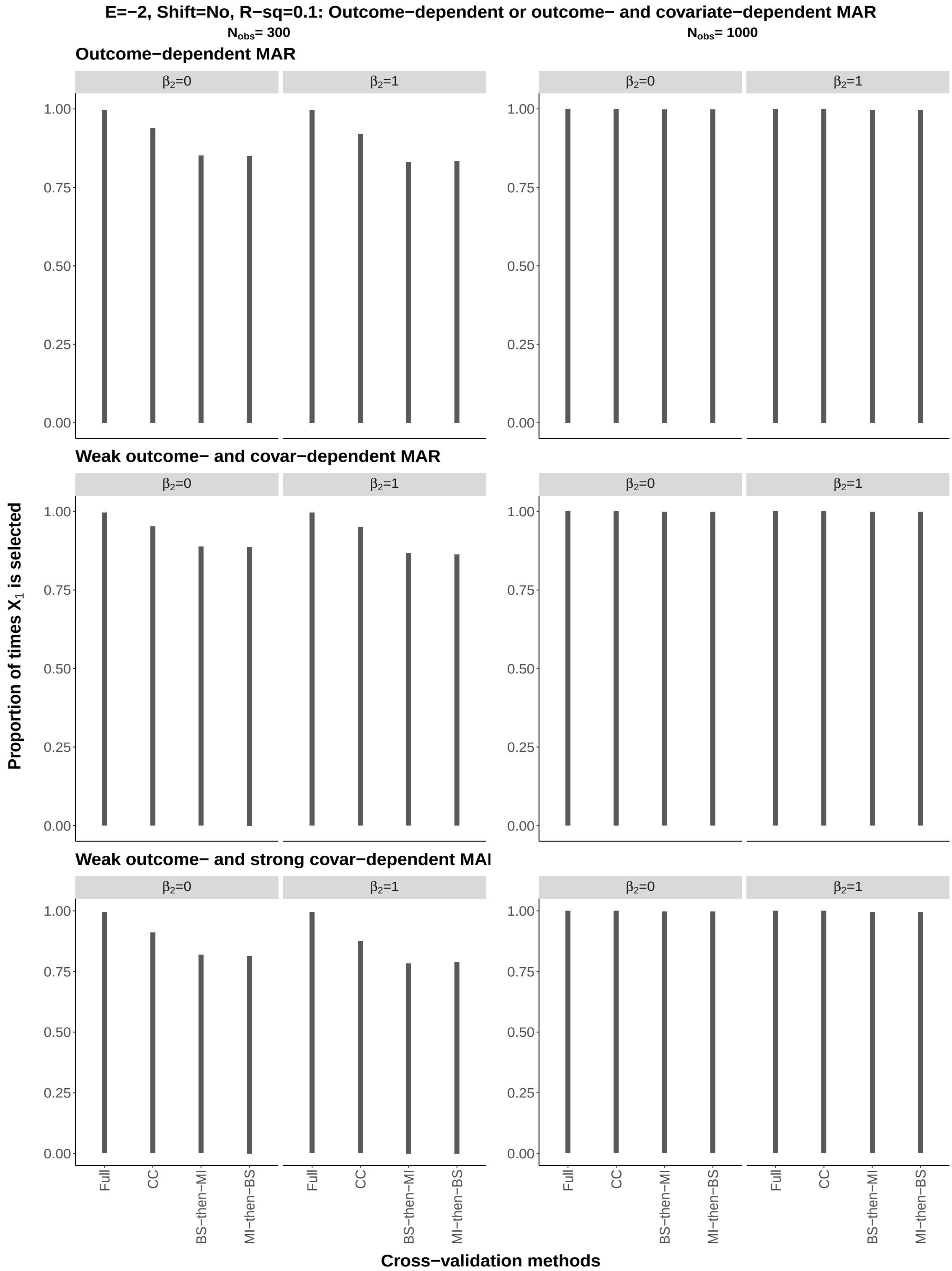

Figure S155: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

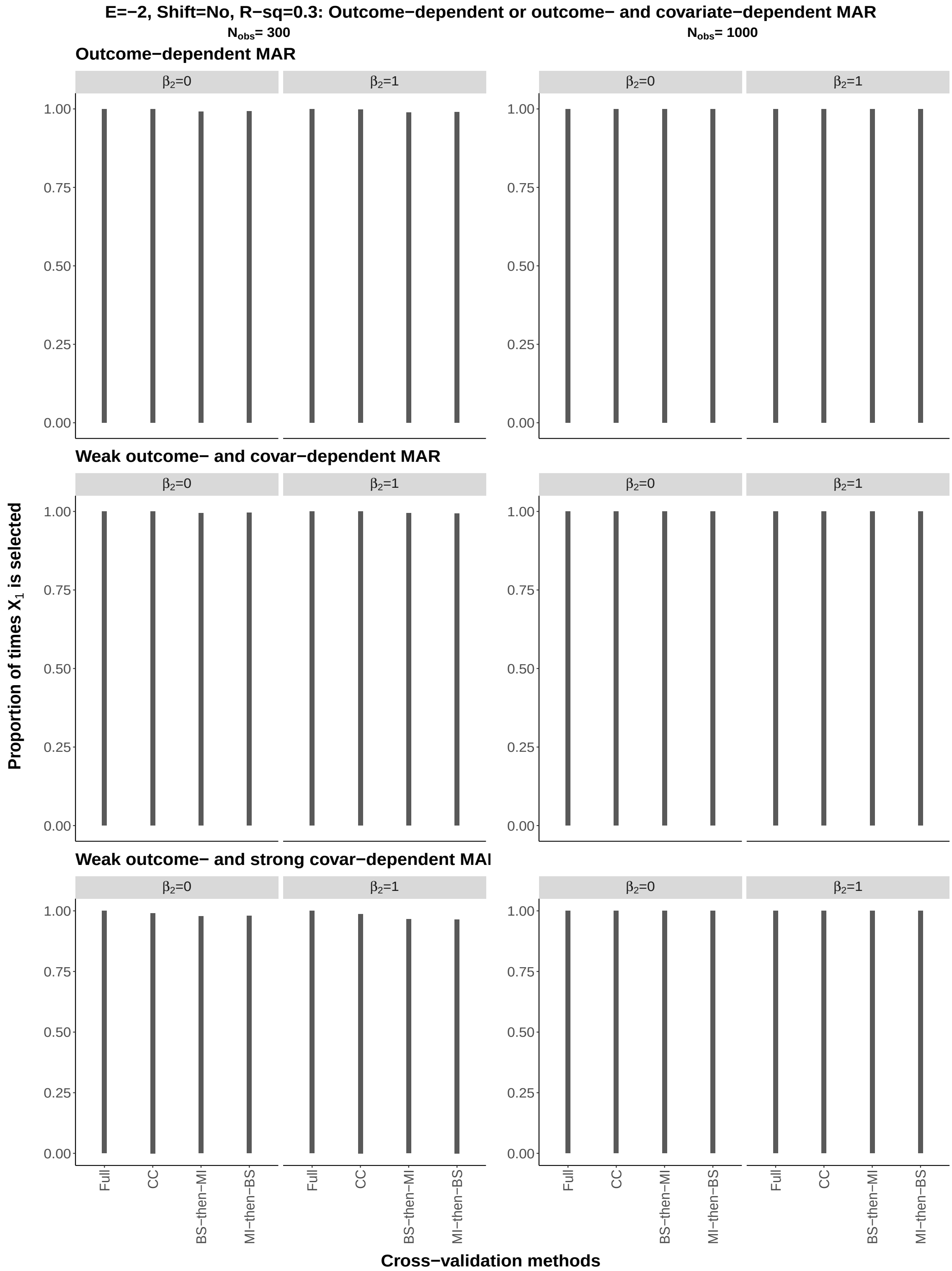

Figure S156: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.6 Covariate selection of X_1 using all data: $\beta_2 = 1$, $\alpha_E = 0.05$, no origin-shift

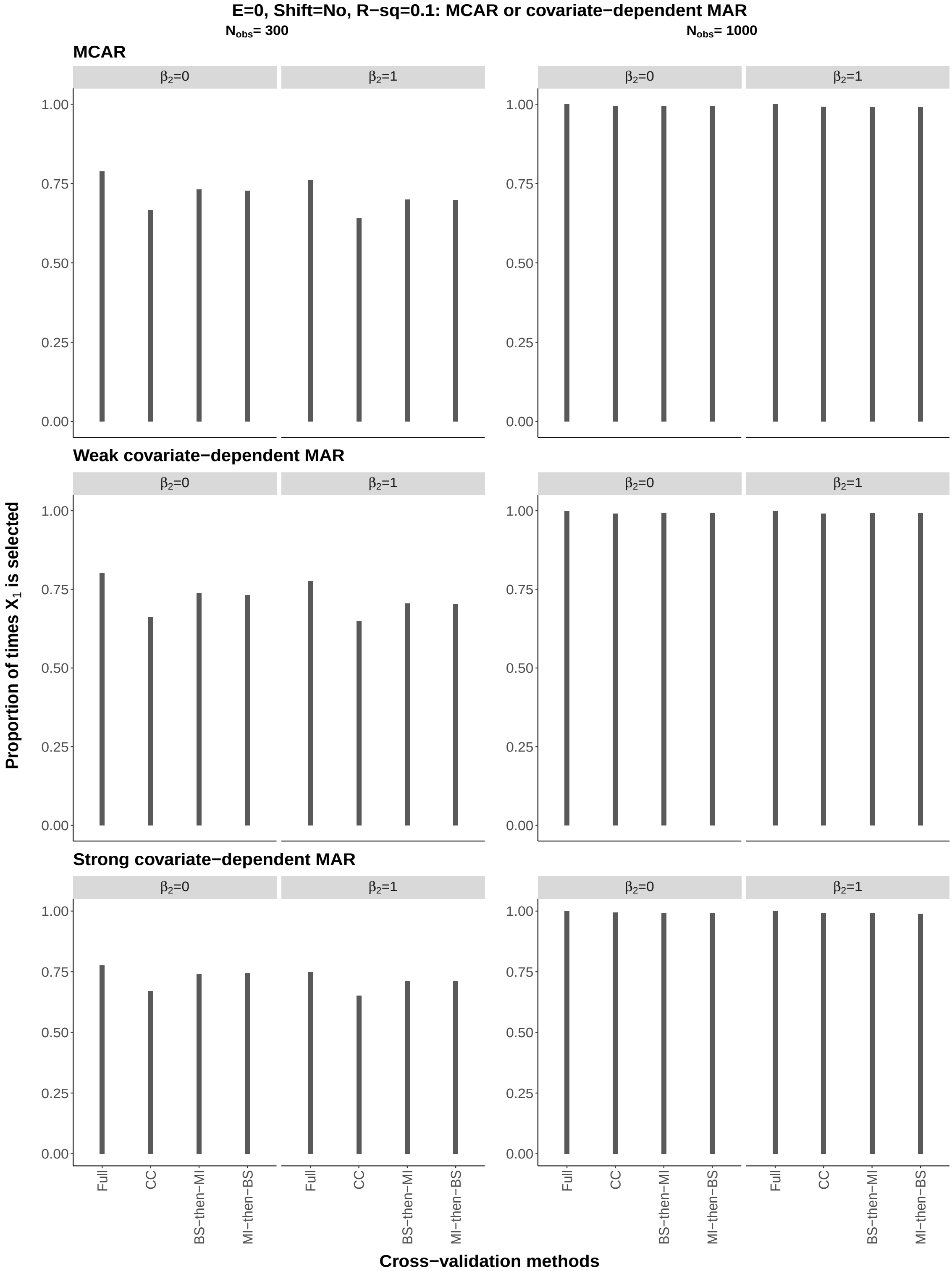

Figure S157: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

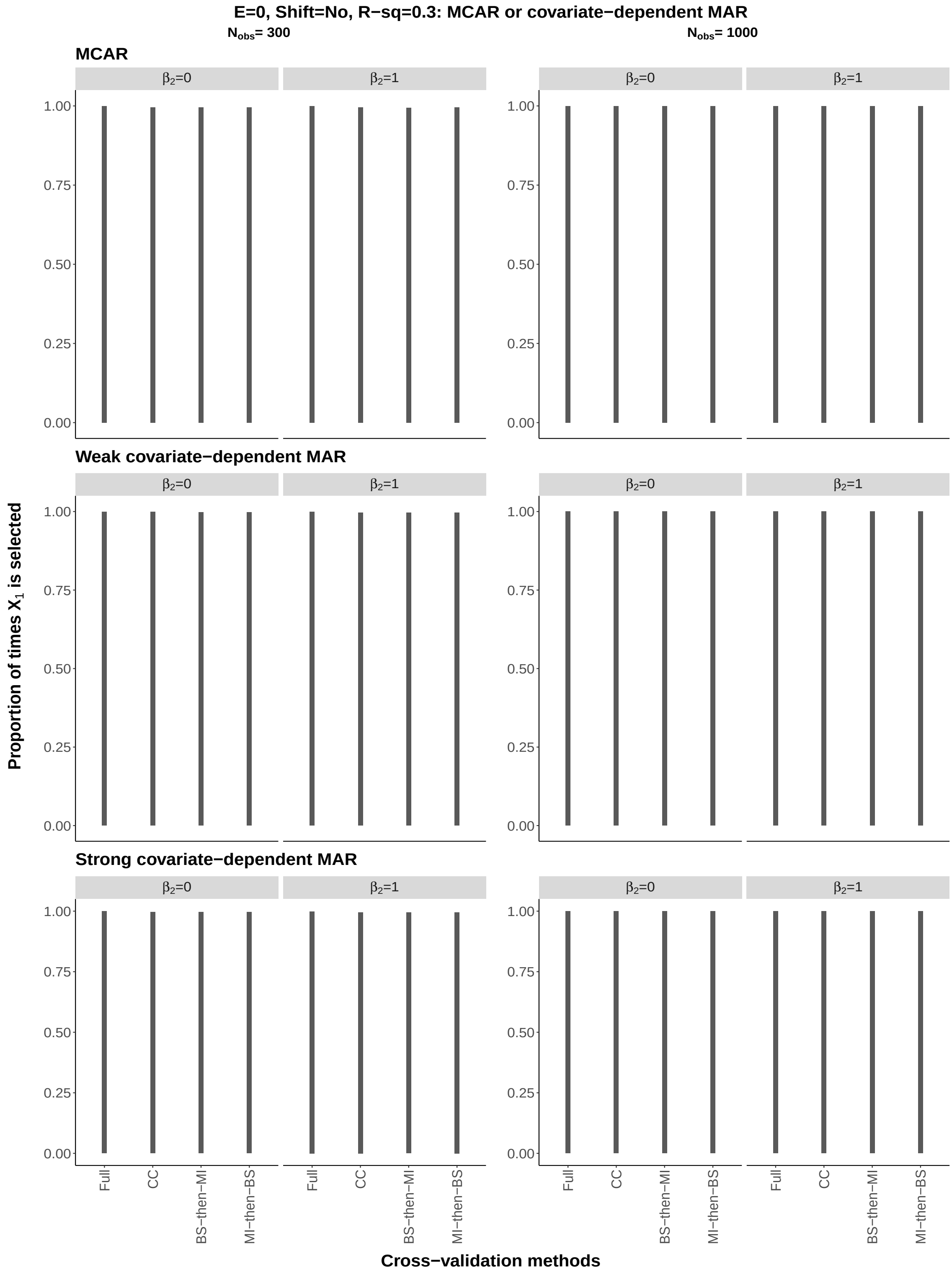

Figure S158: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

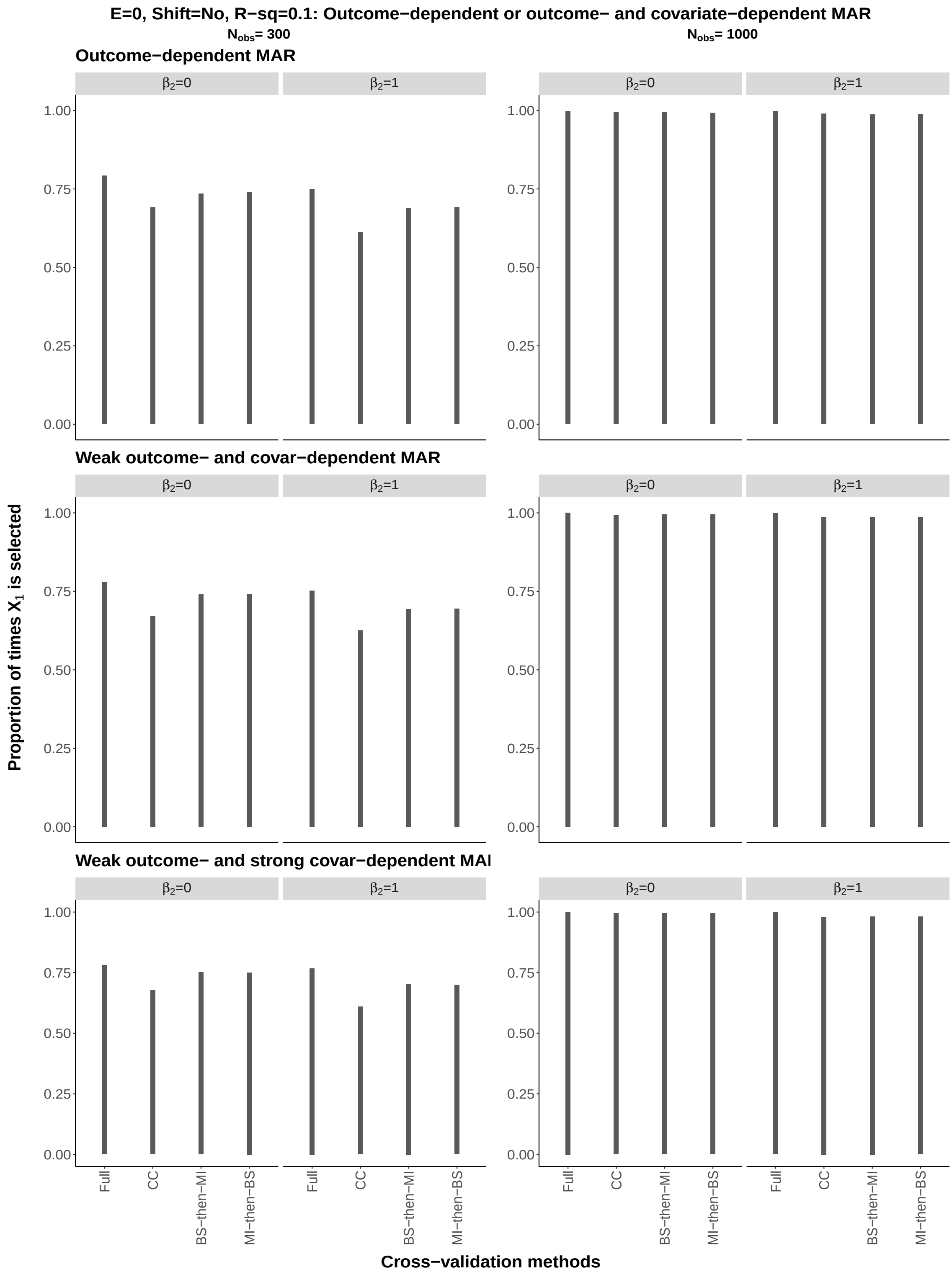

Figure S159: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

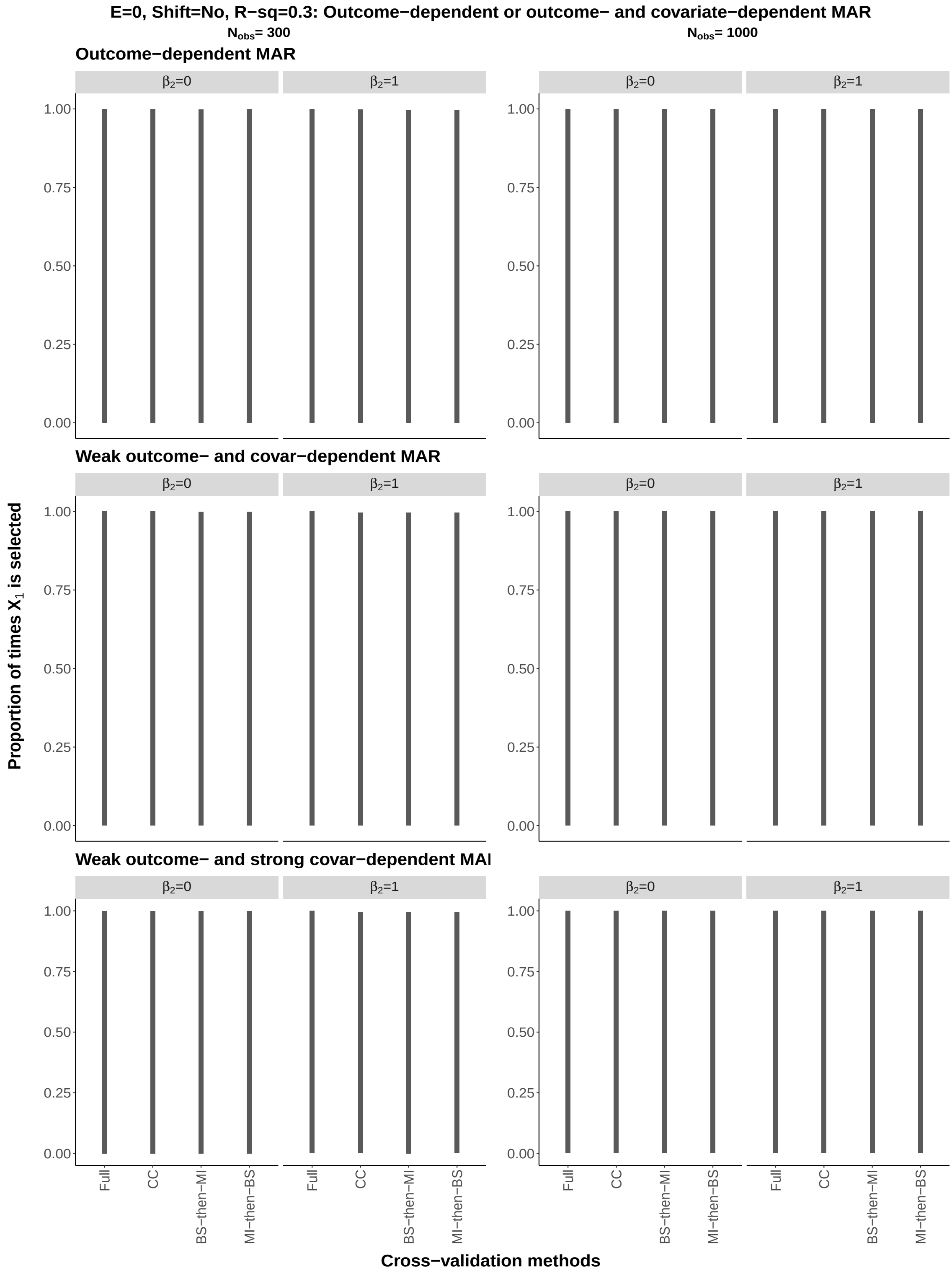

Figure S160: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

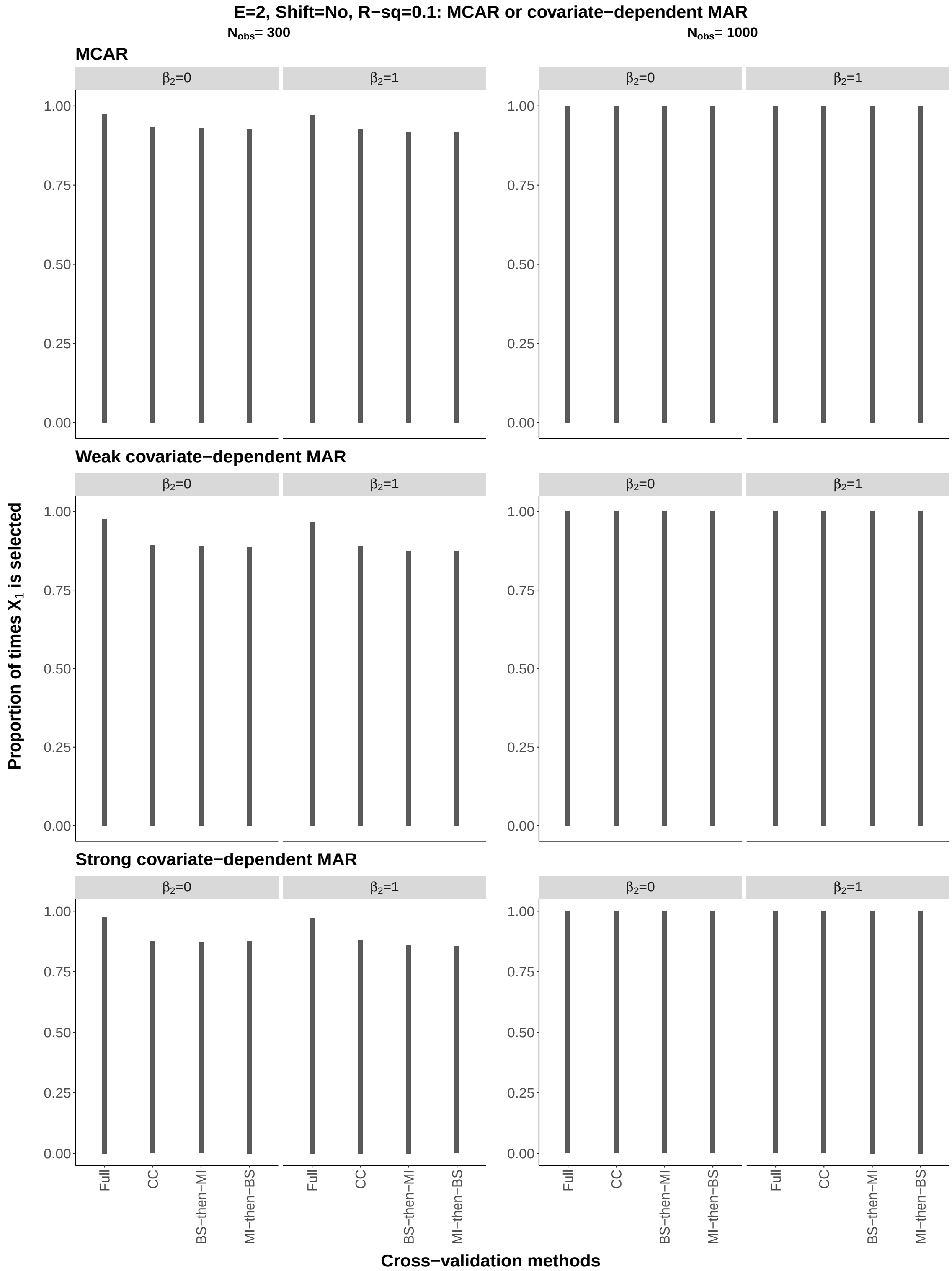

Figure S161: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

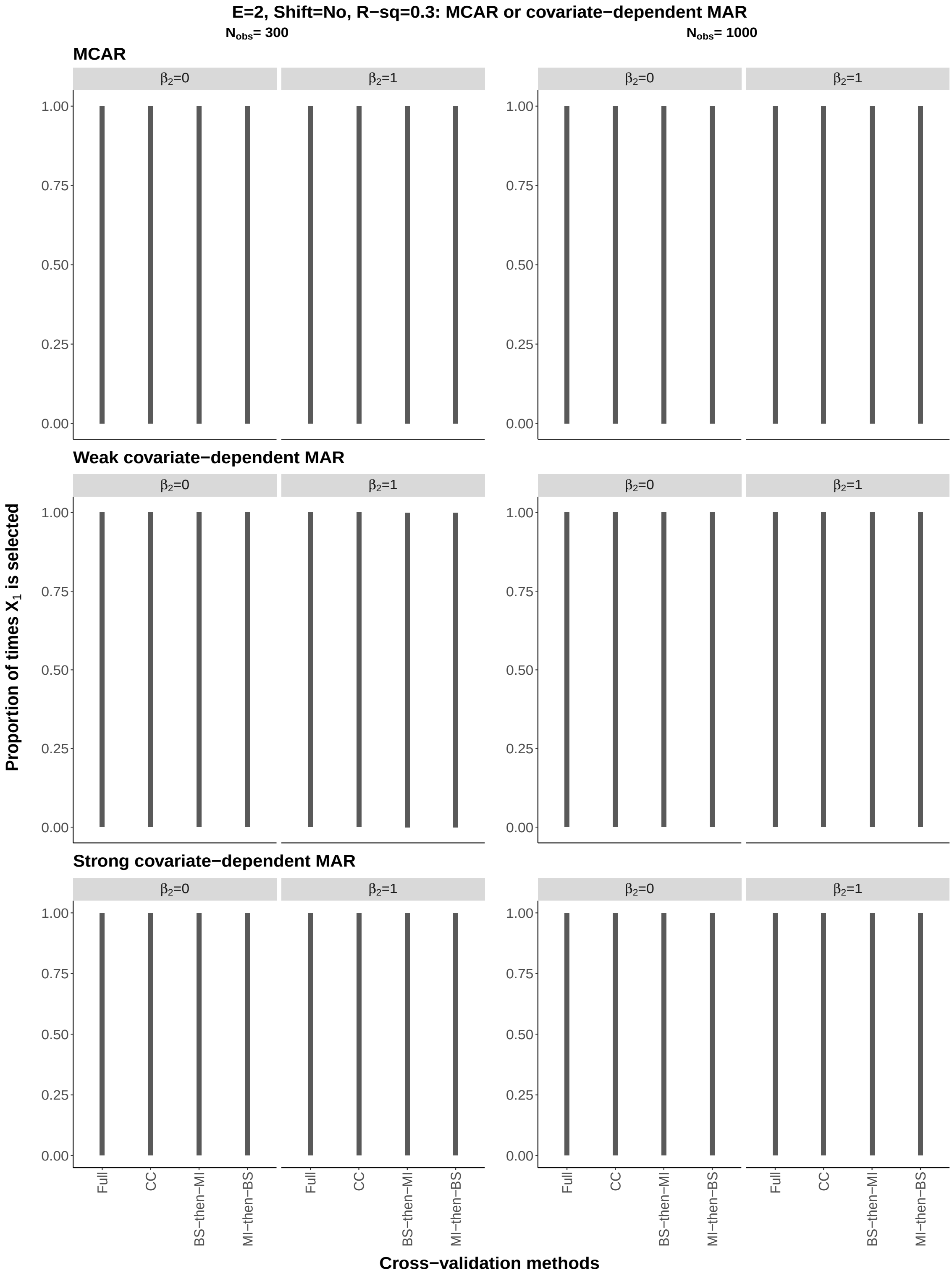

Figure S162: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

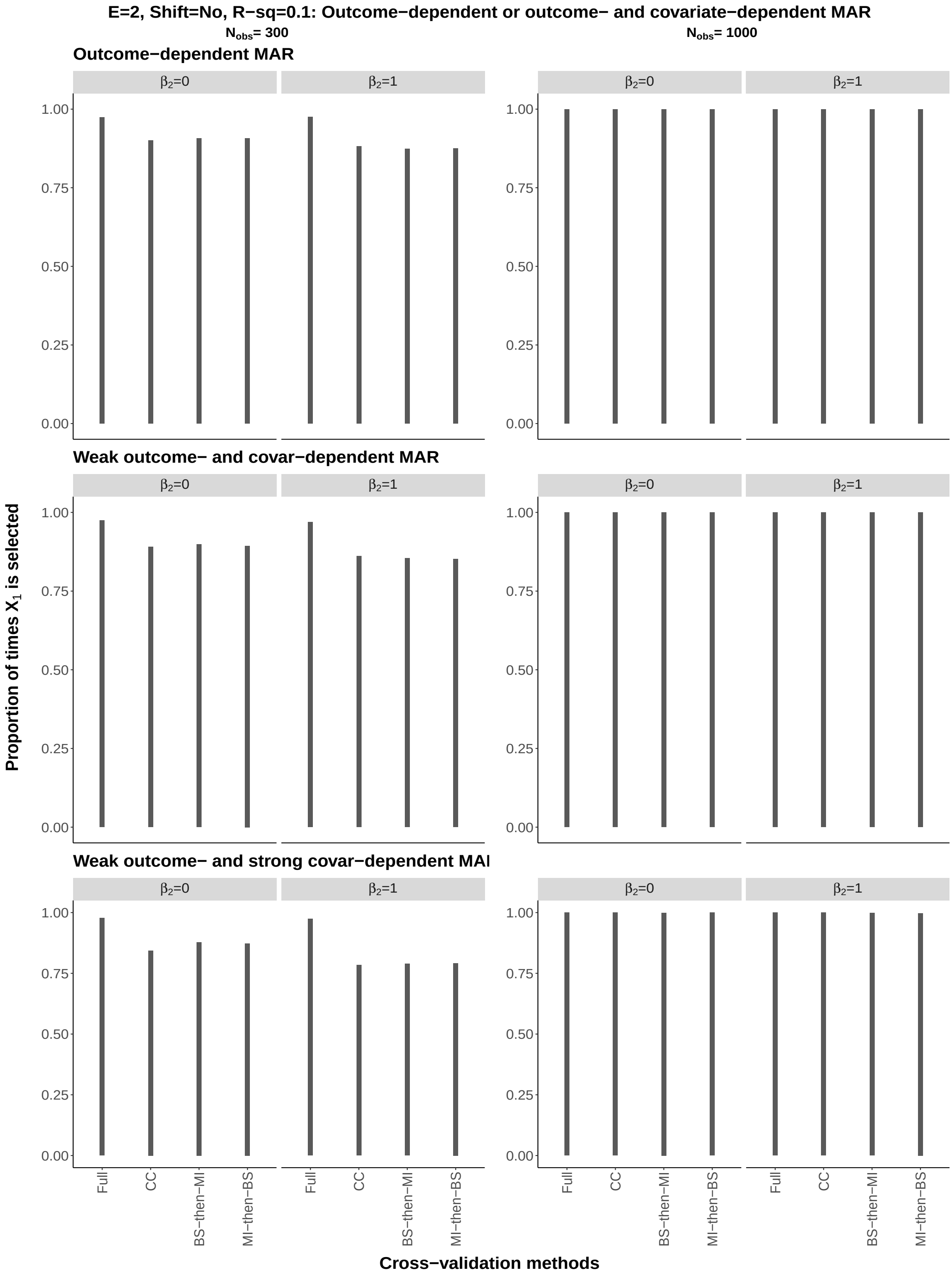

Figure S163: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

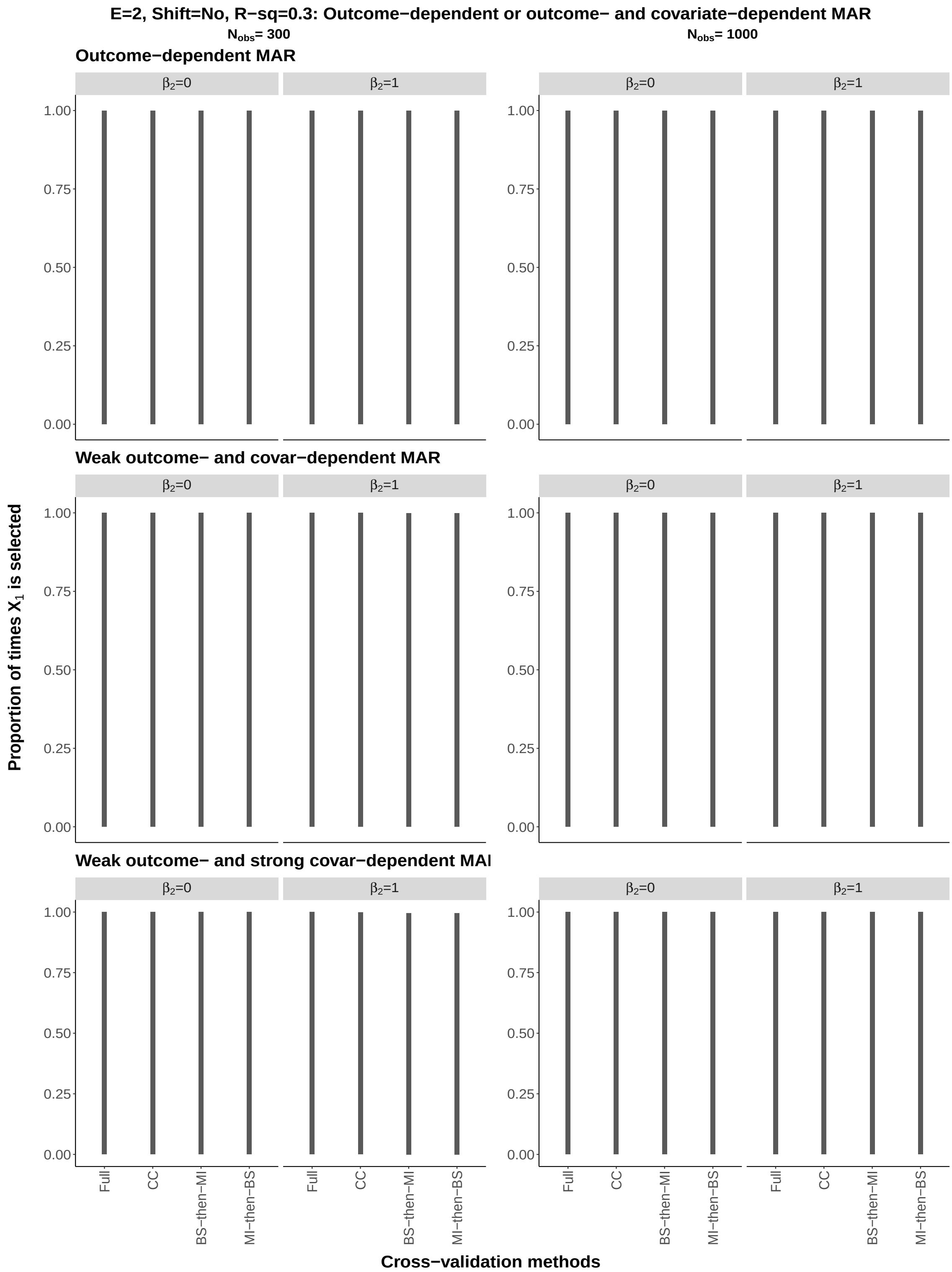

Figure S164: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

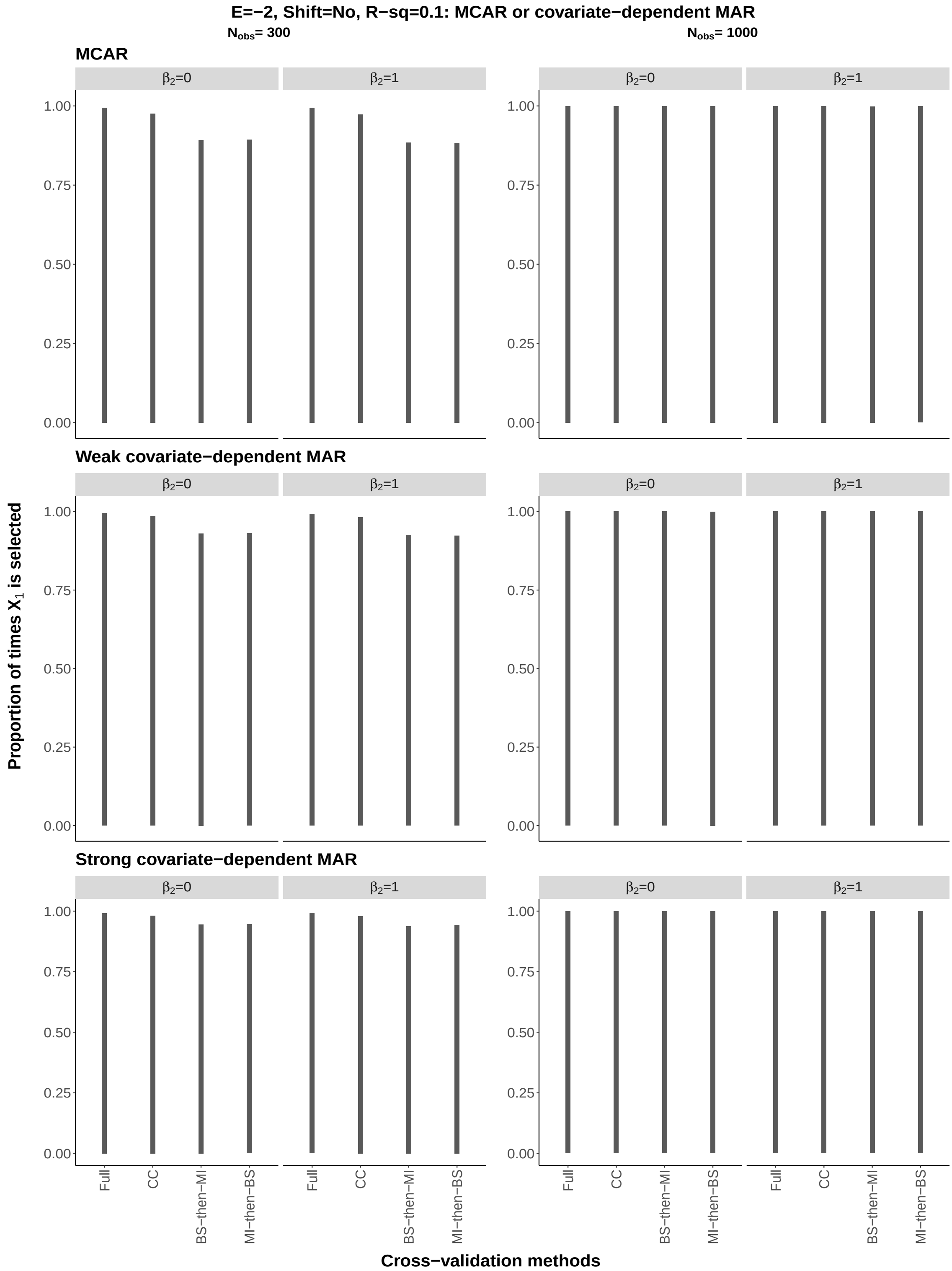

Figure S165: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

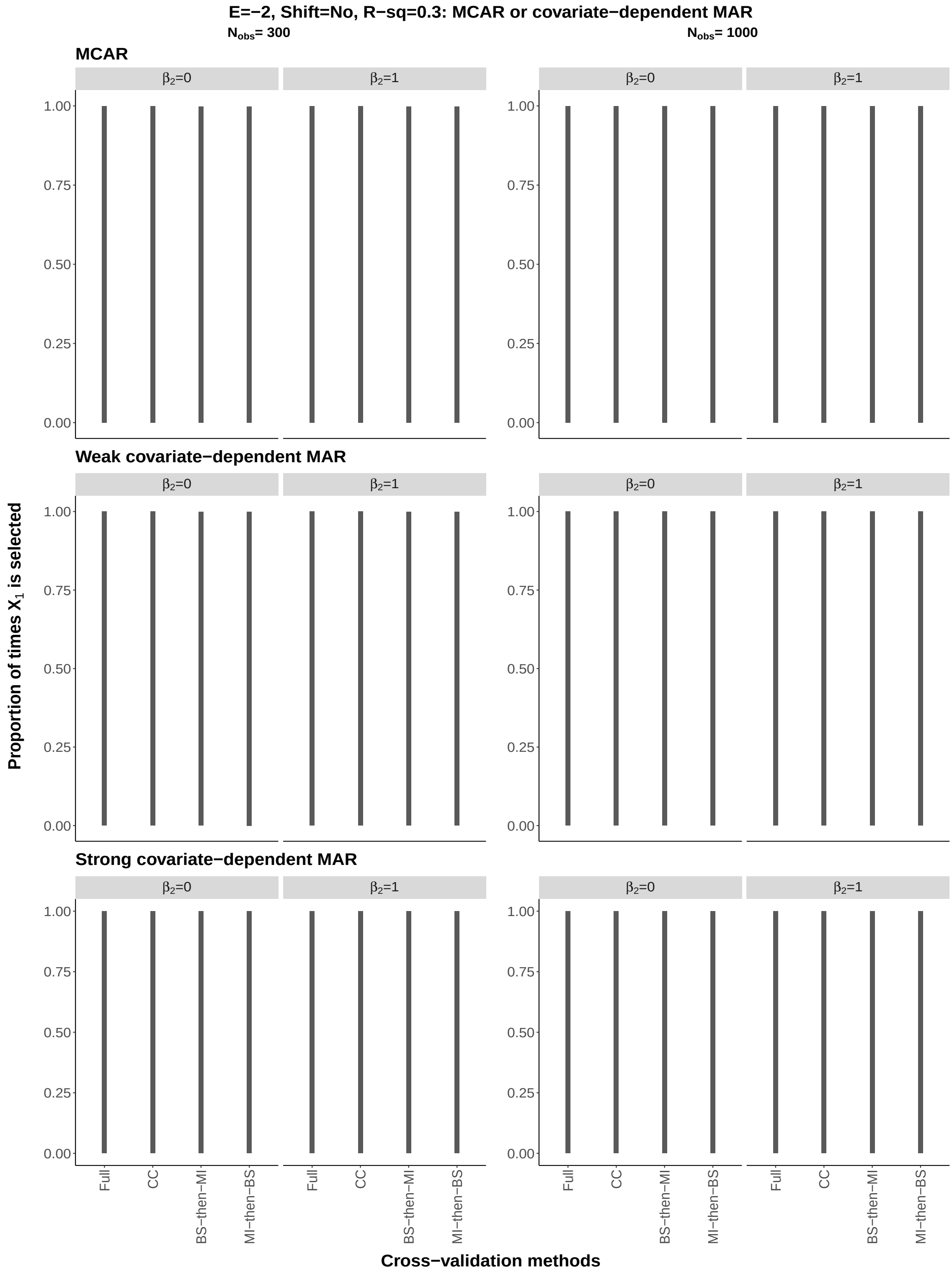

Figure S166: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

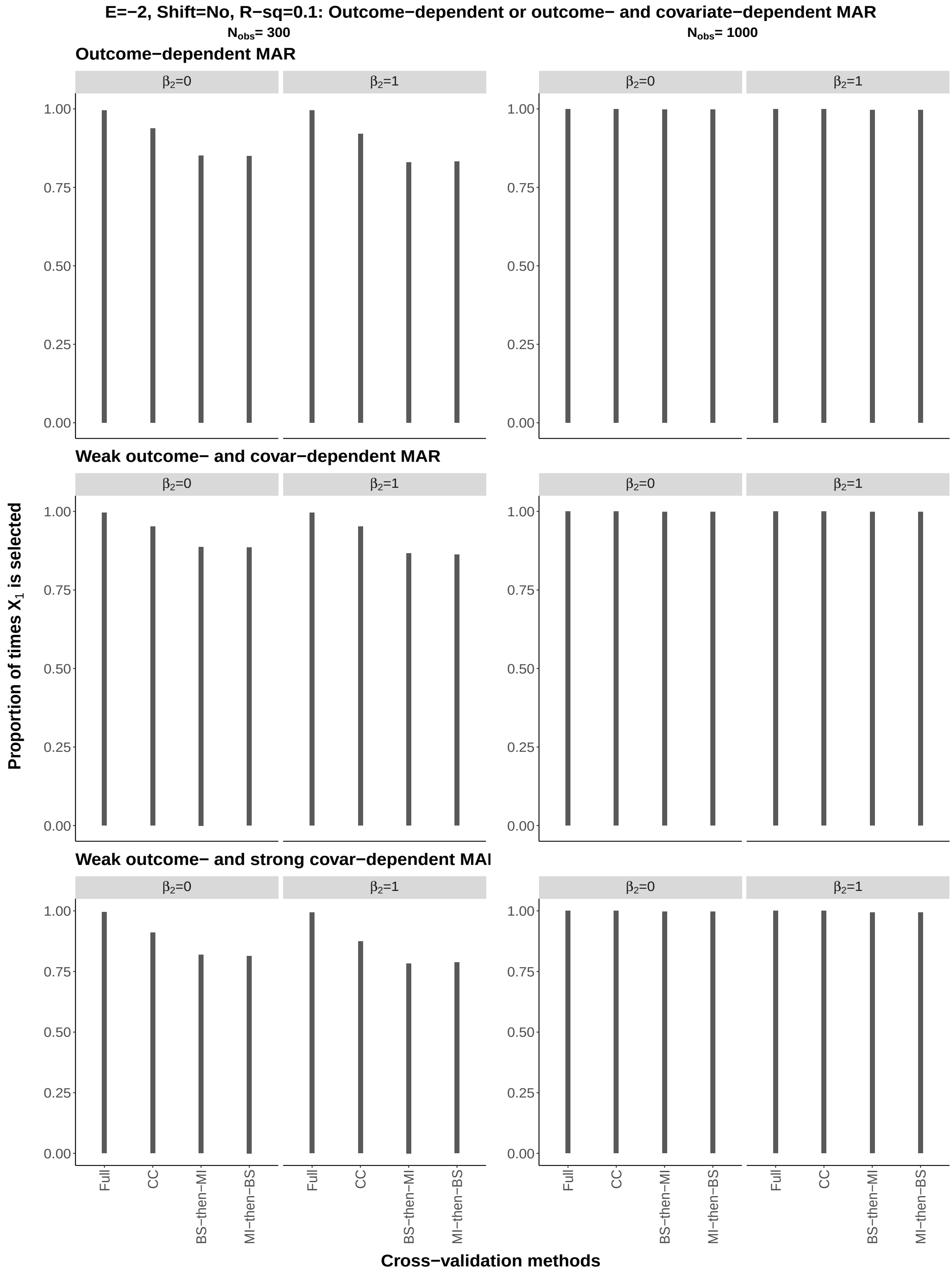

Figure S167: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

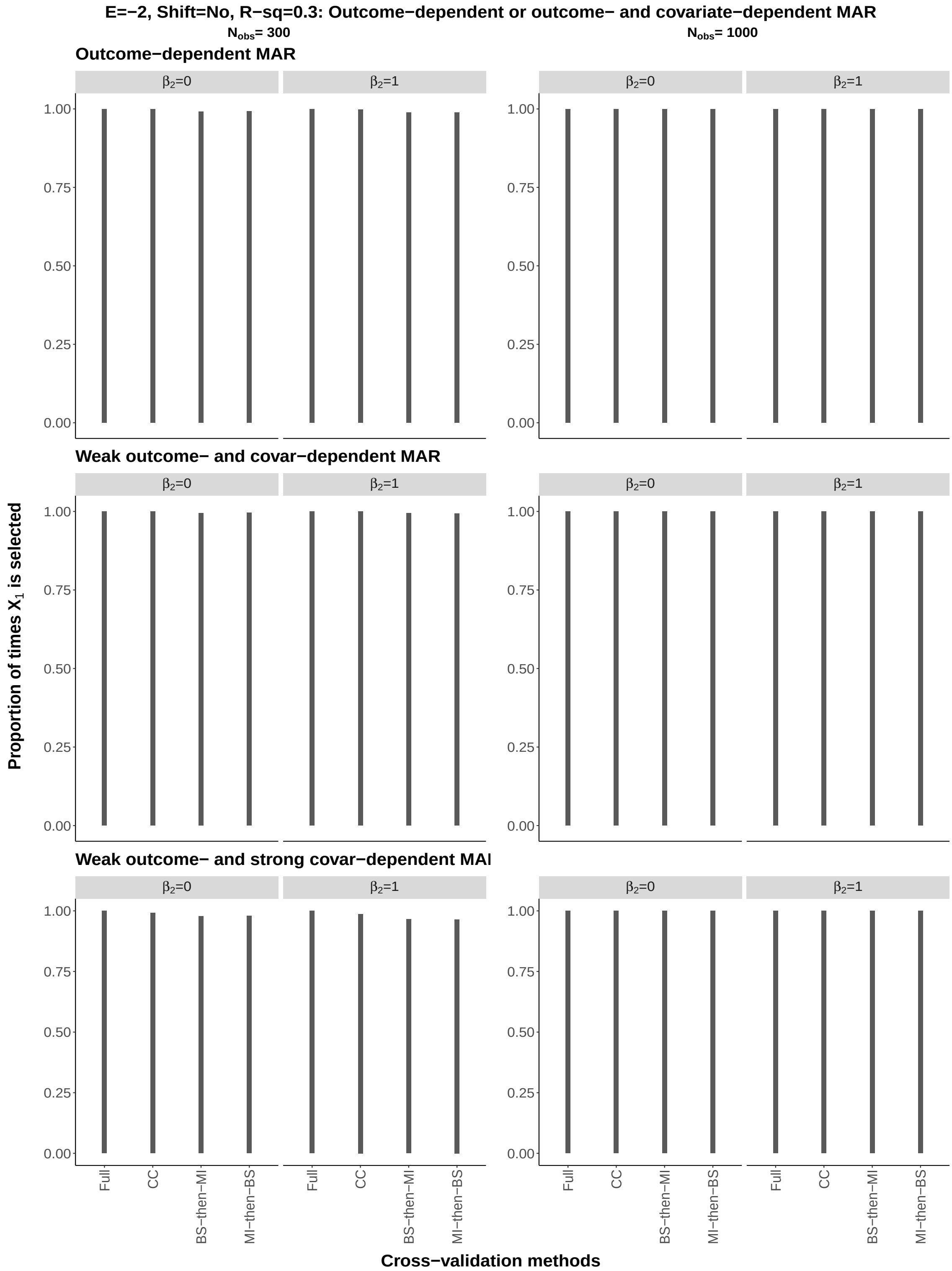

Figure S168: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.7 Covariate selection of X_1 using all data: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

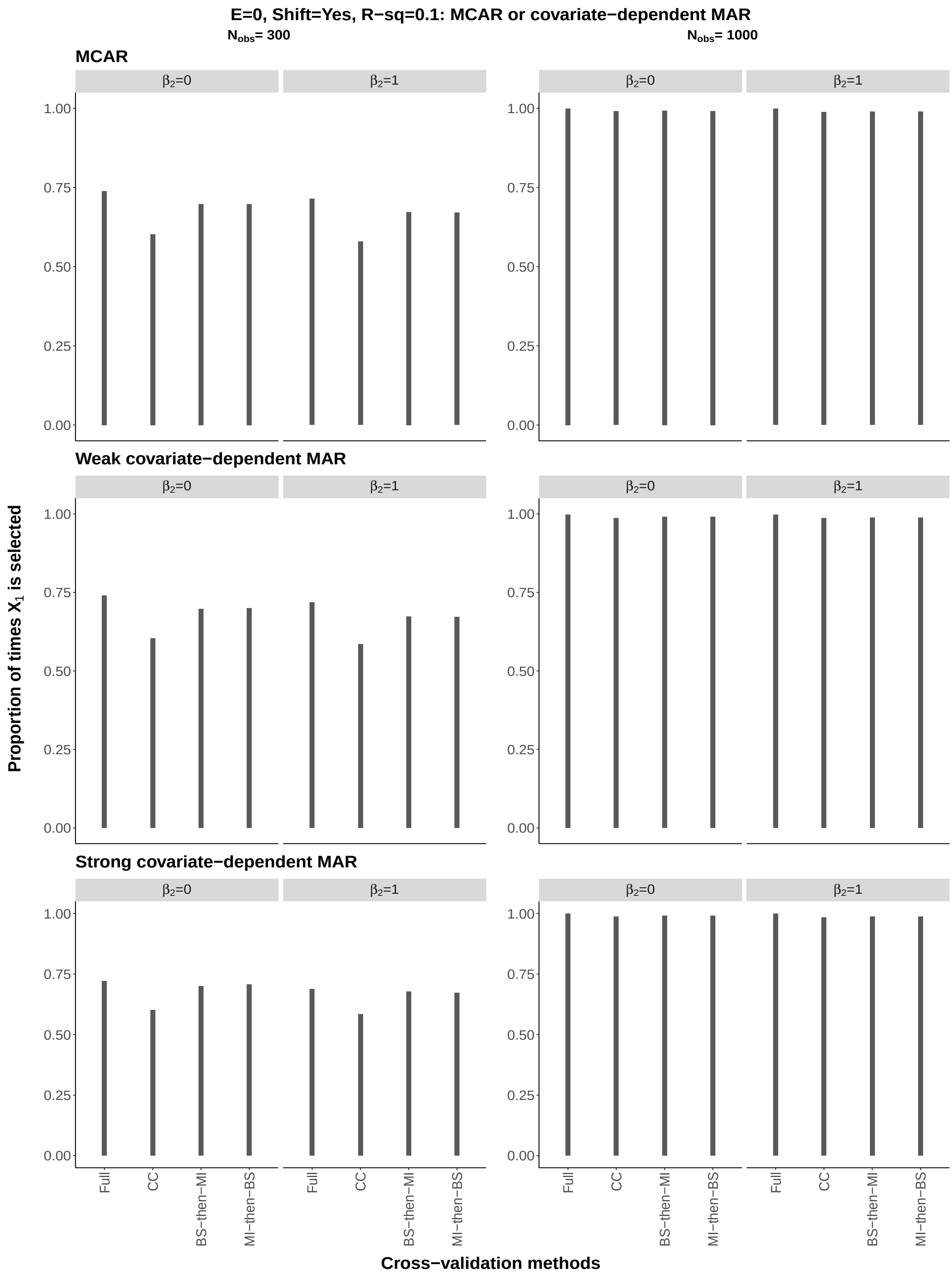

Figure S169: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.1827

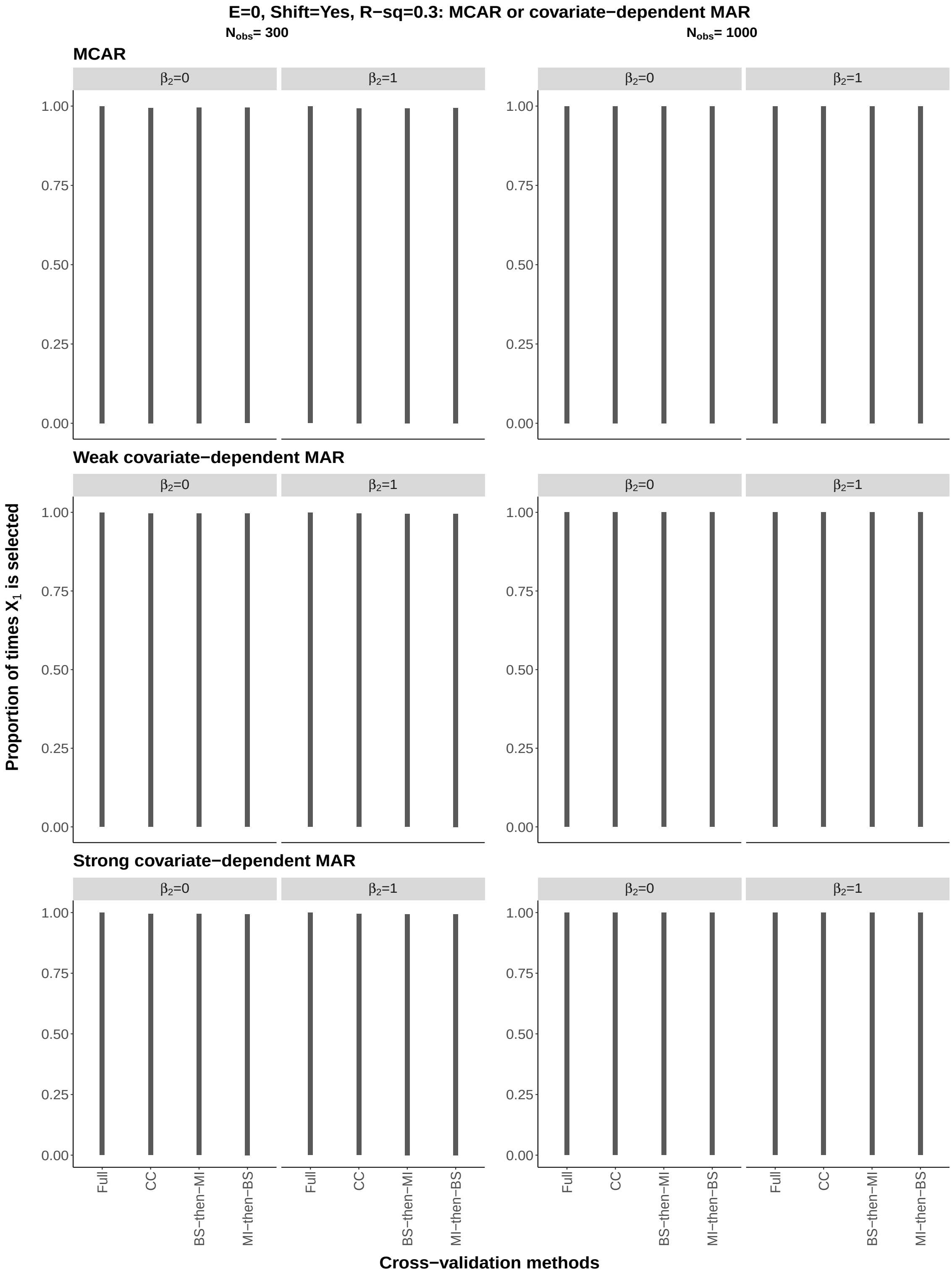

Figure S170: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

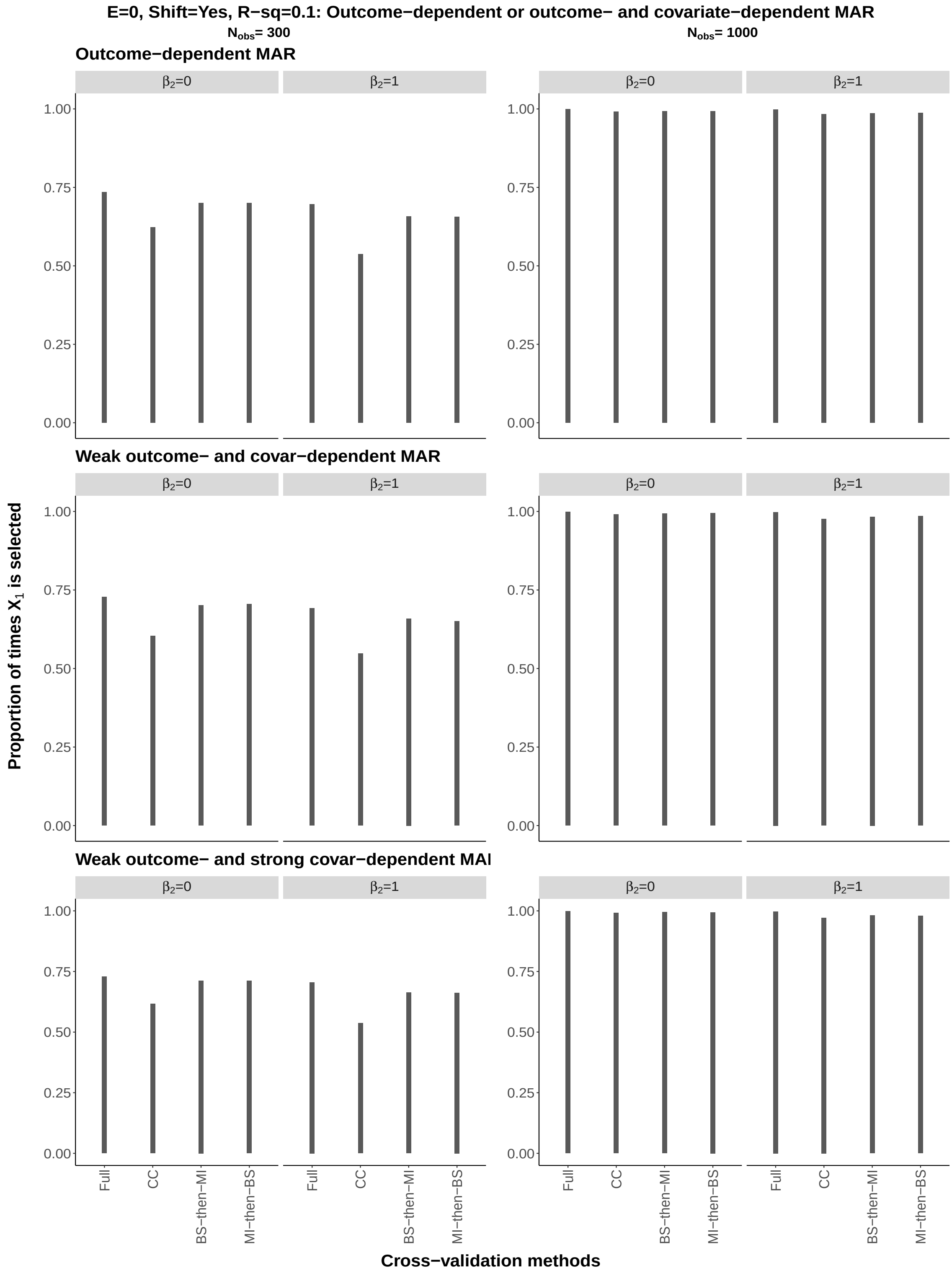

Figure S171: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

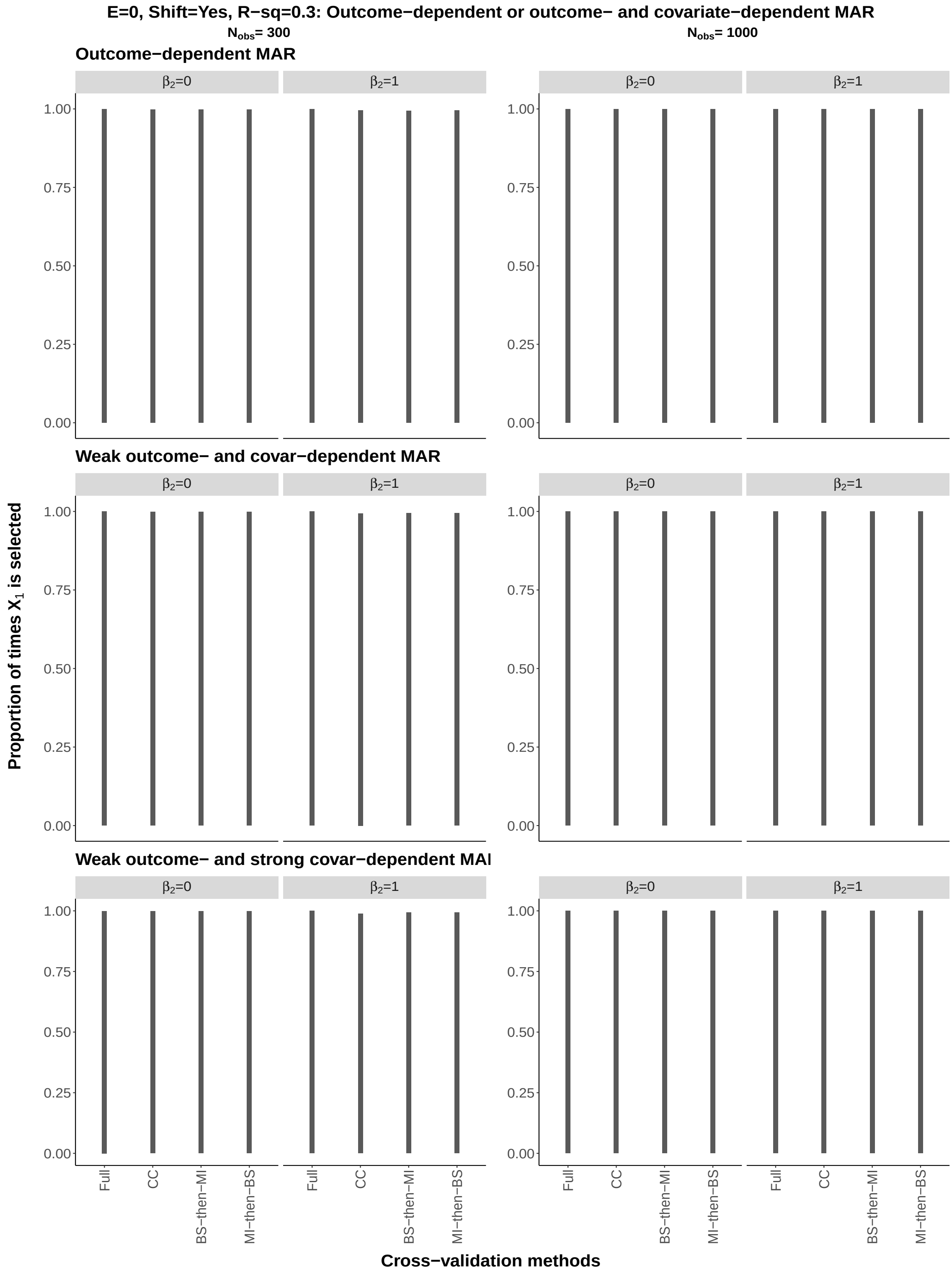

Figure S172: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

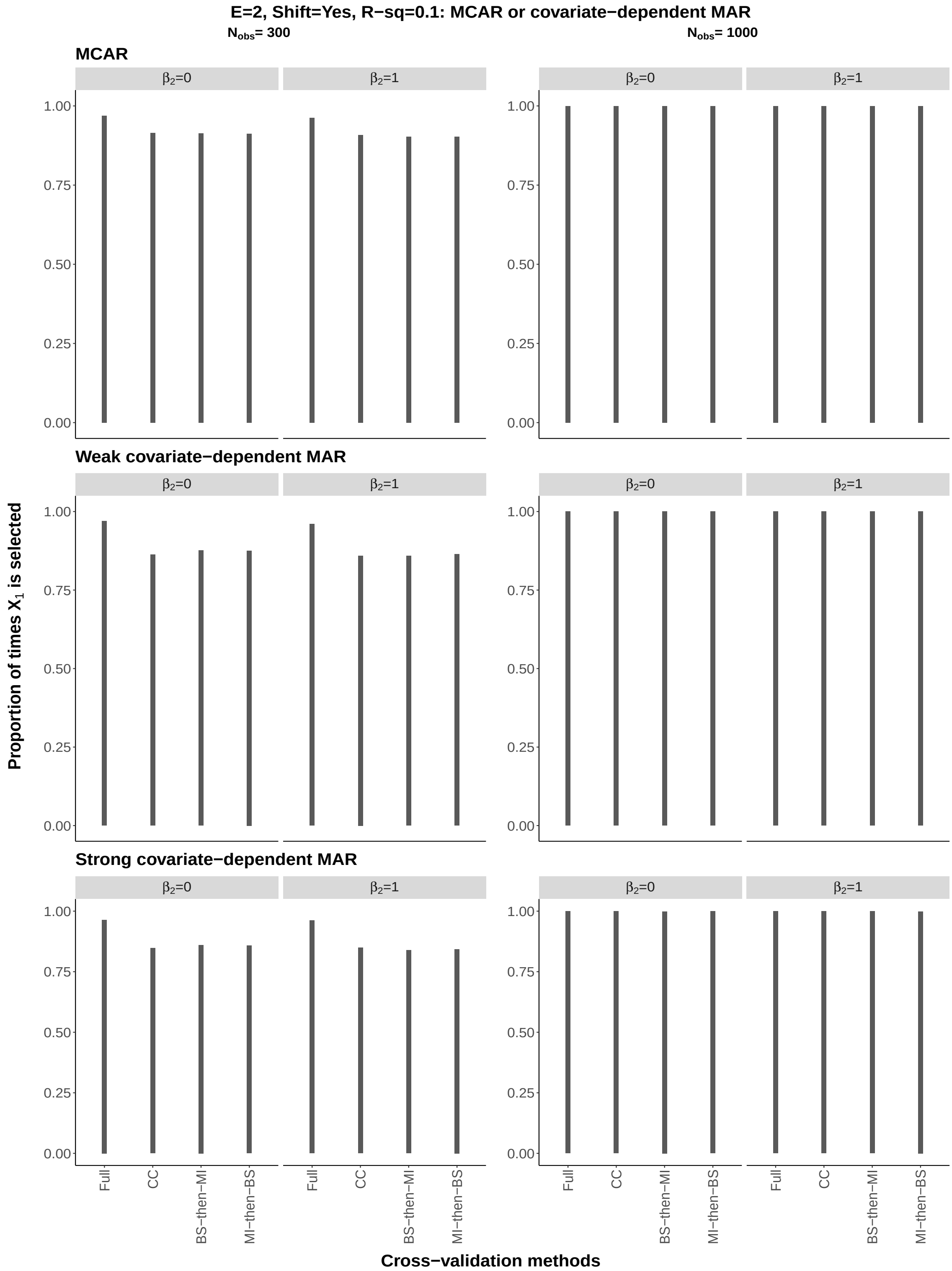

Figure S173: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

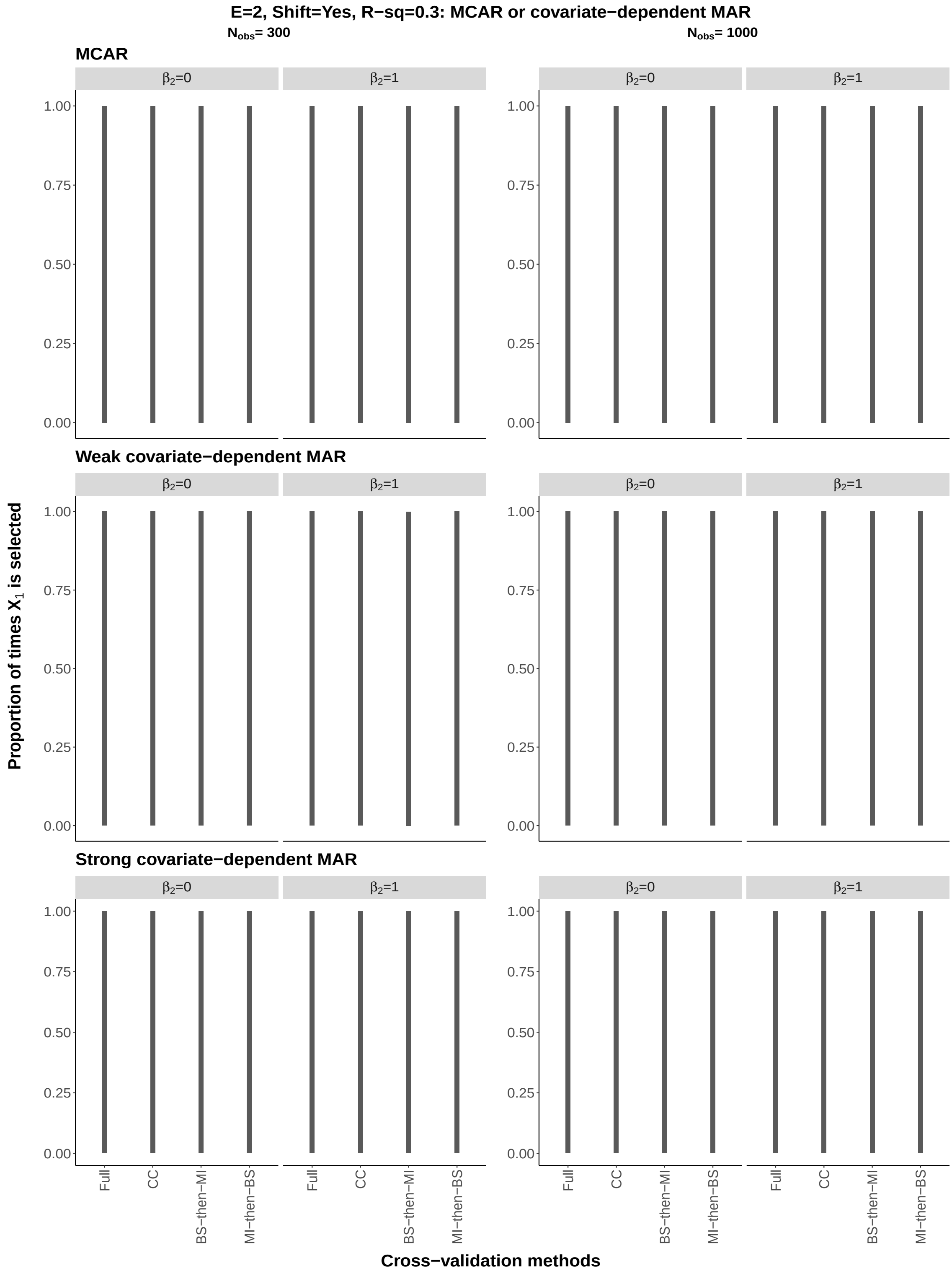

Figure S174: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

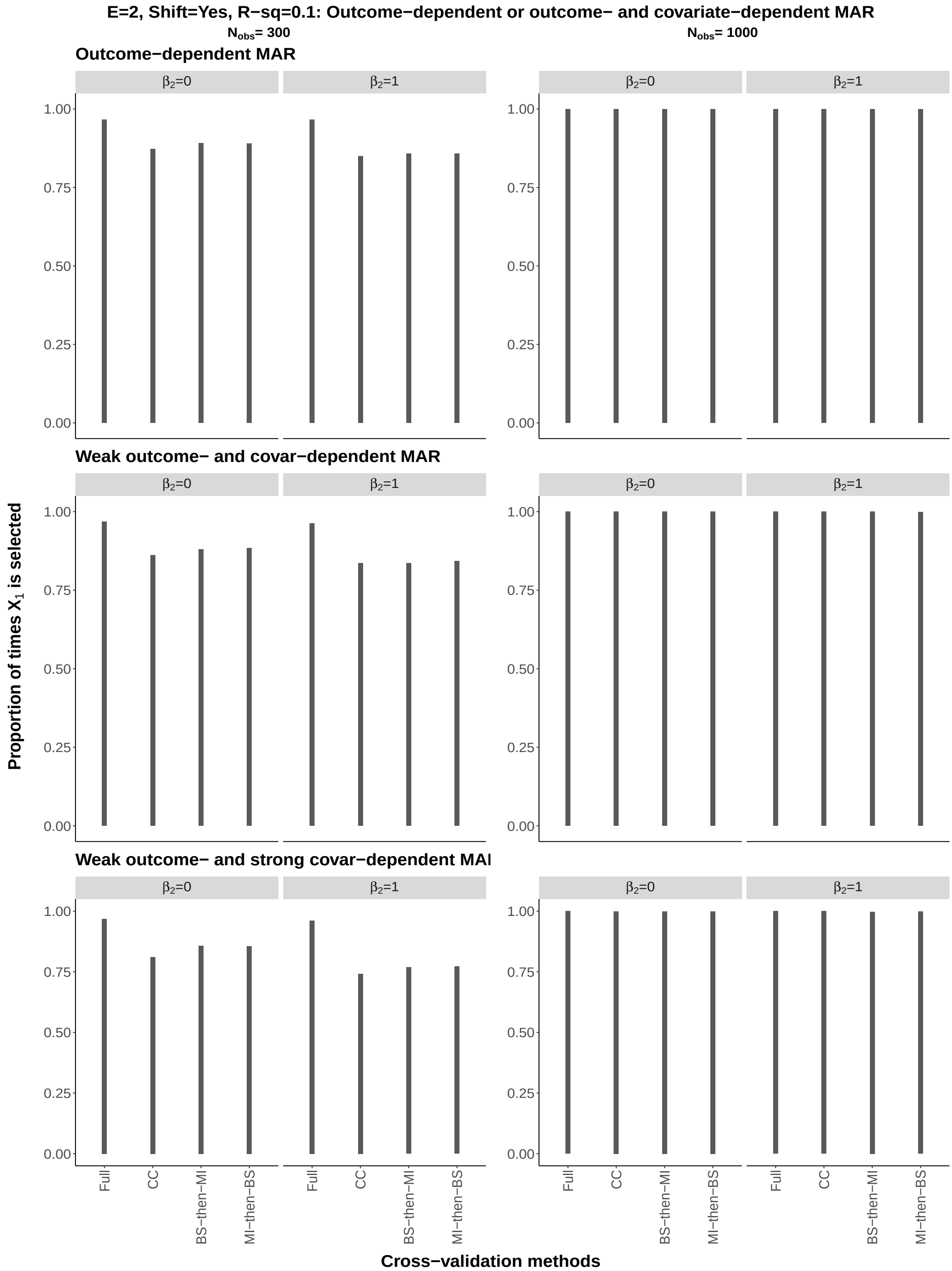

Figure S175: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

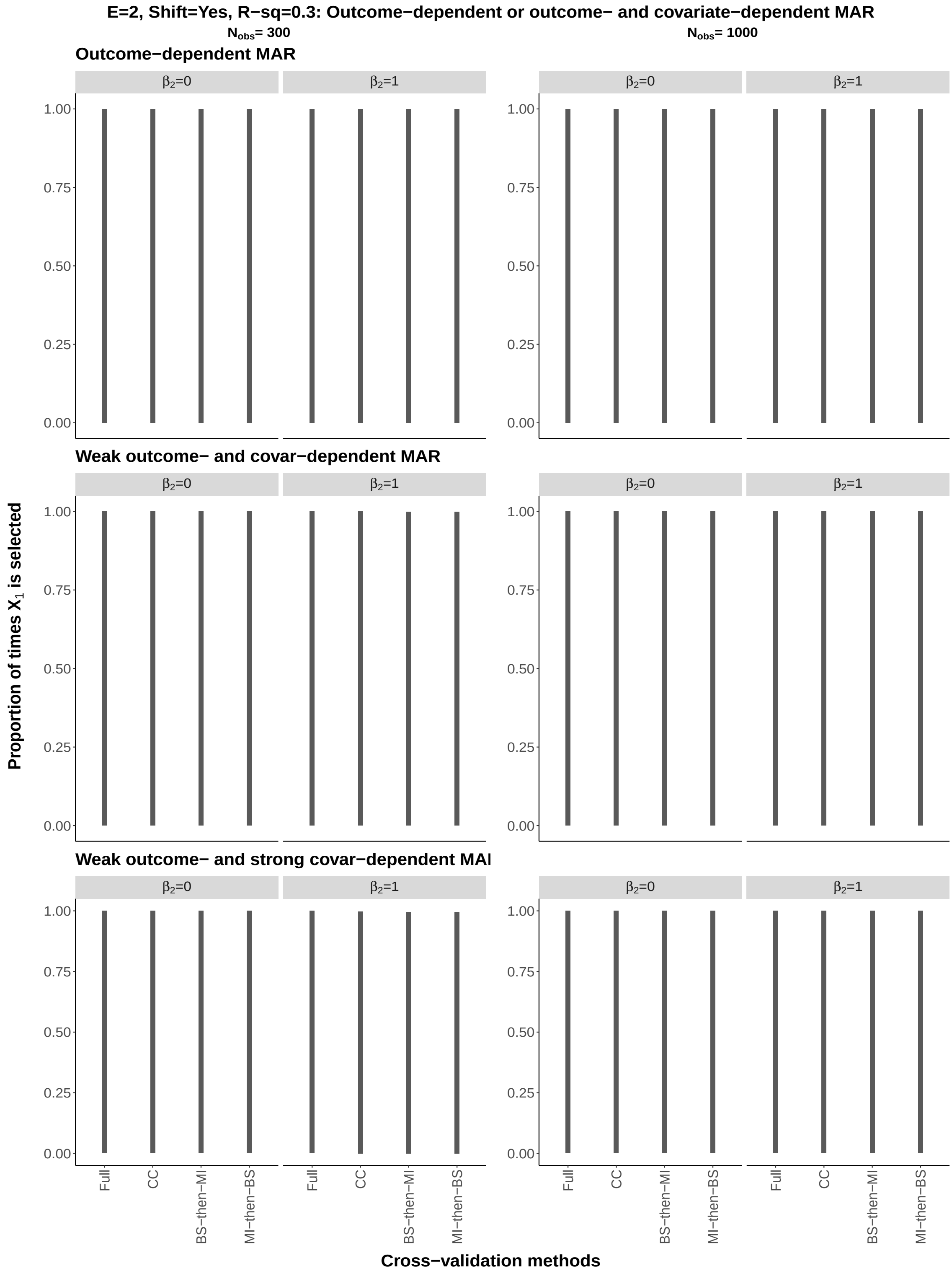

Figure S176: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

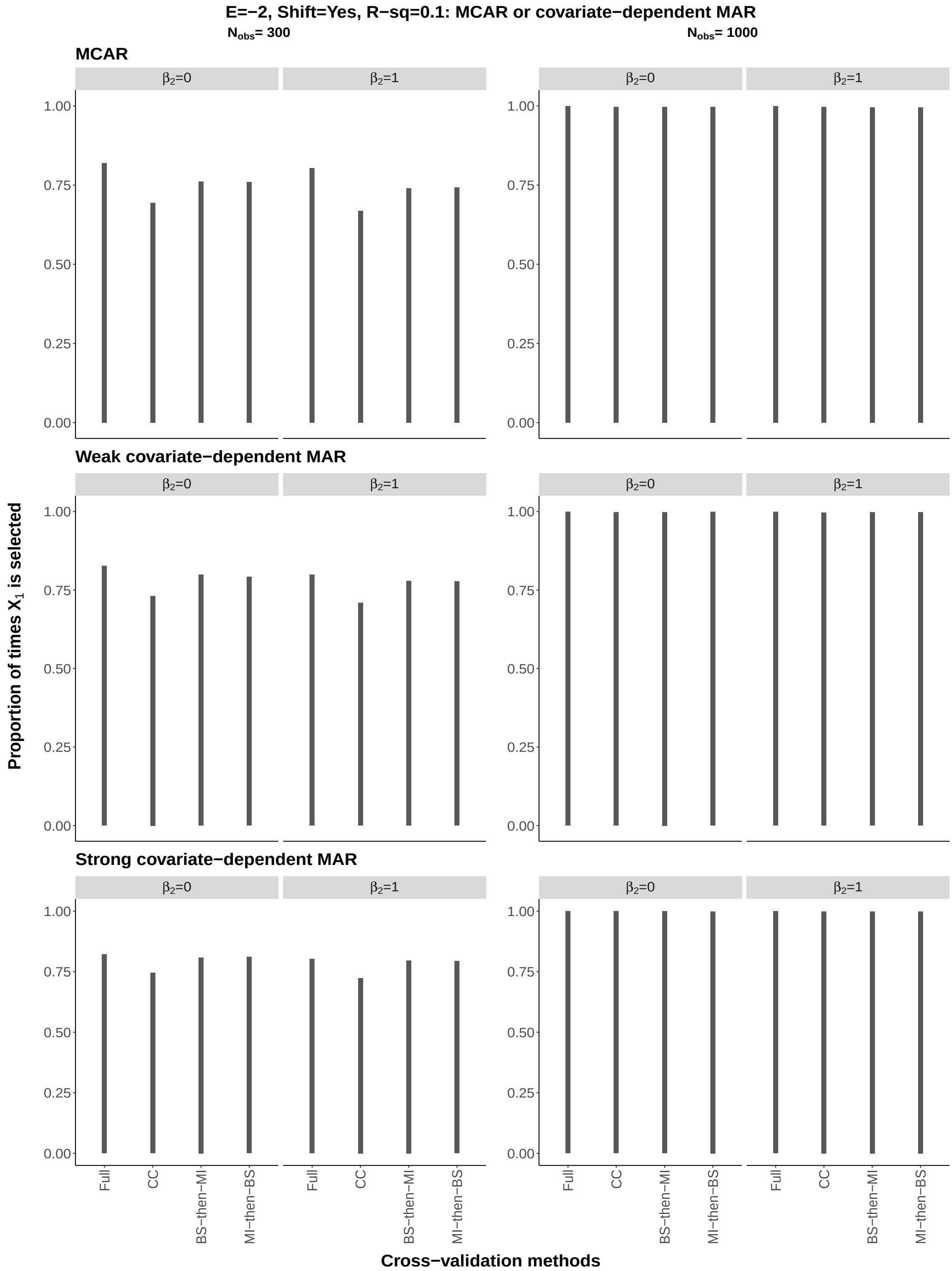

Figure S177: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

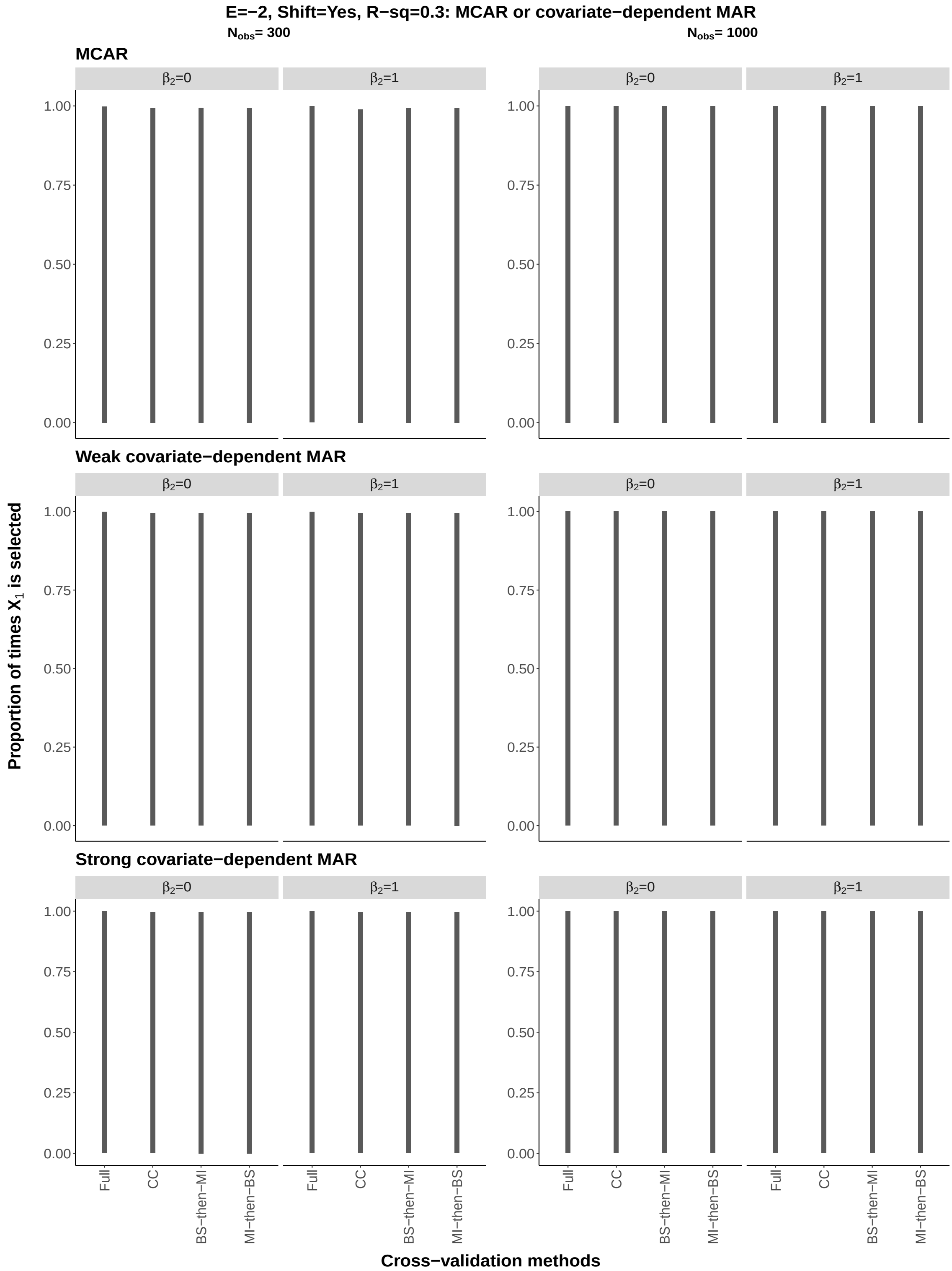

Figure S178: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

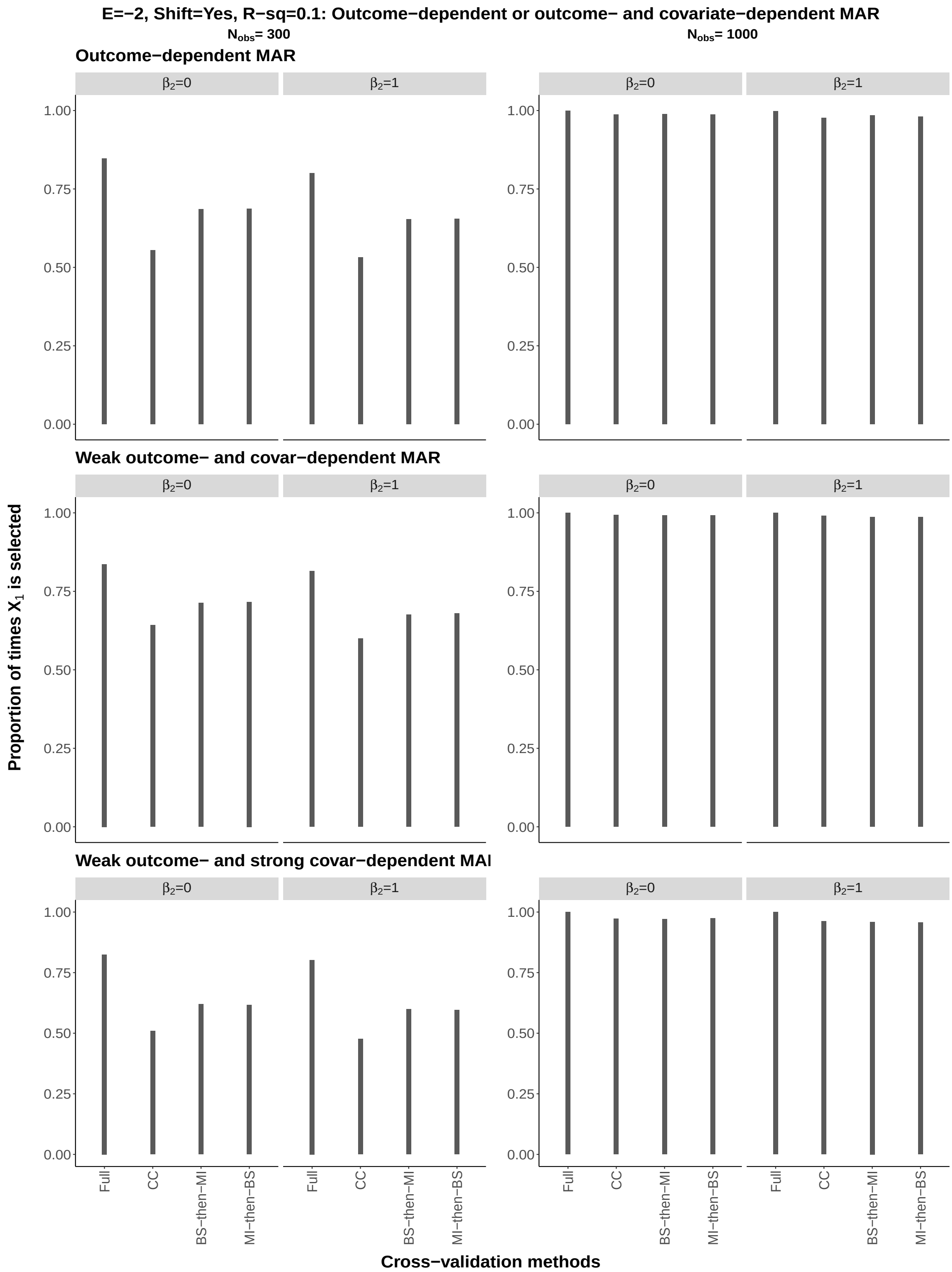

Figure S179: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

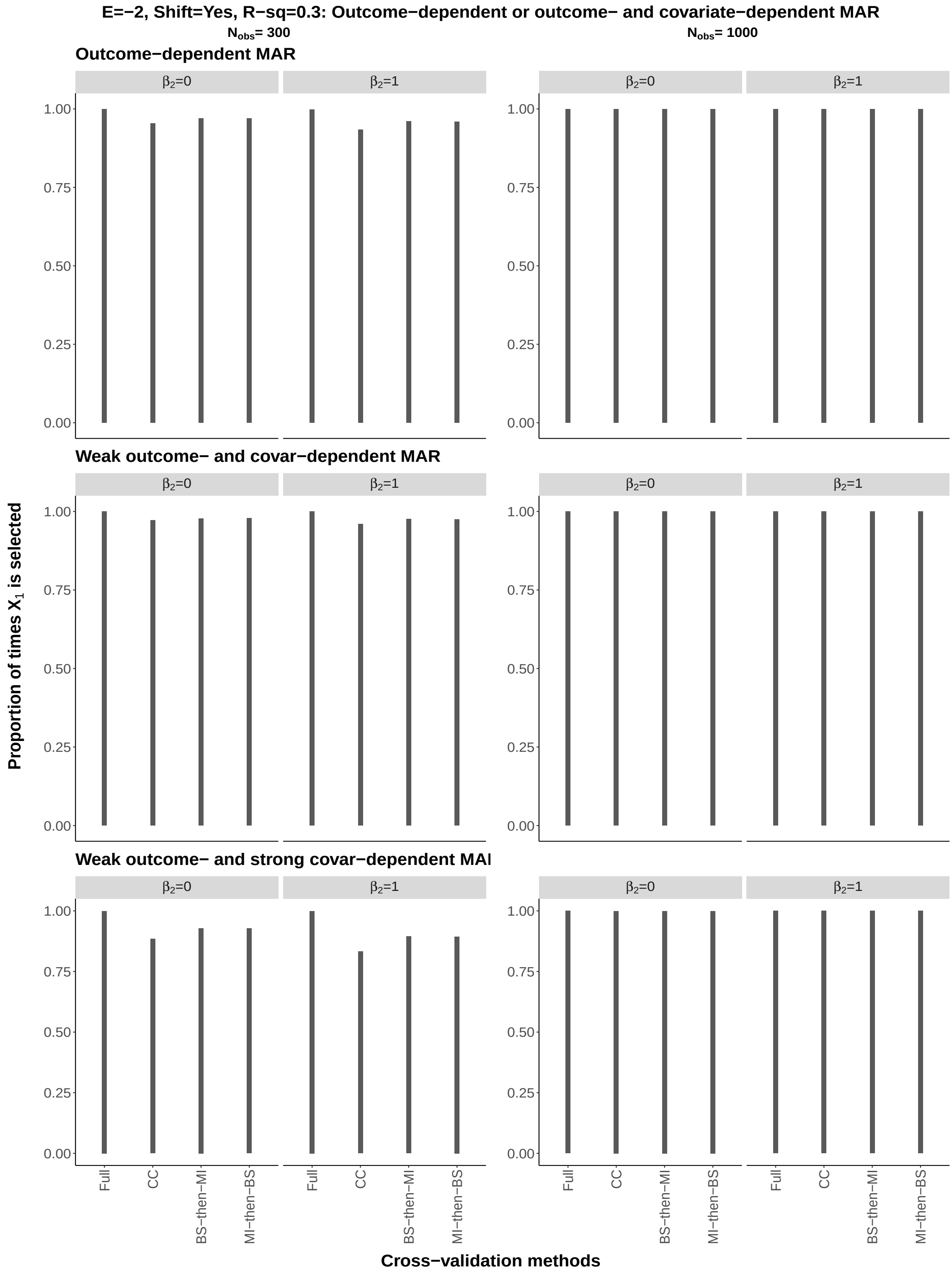

Figure S180: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.8 Covariate selection of X_1 using all data: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

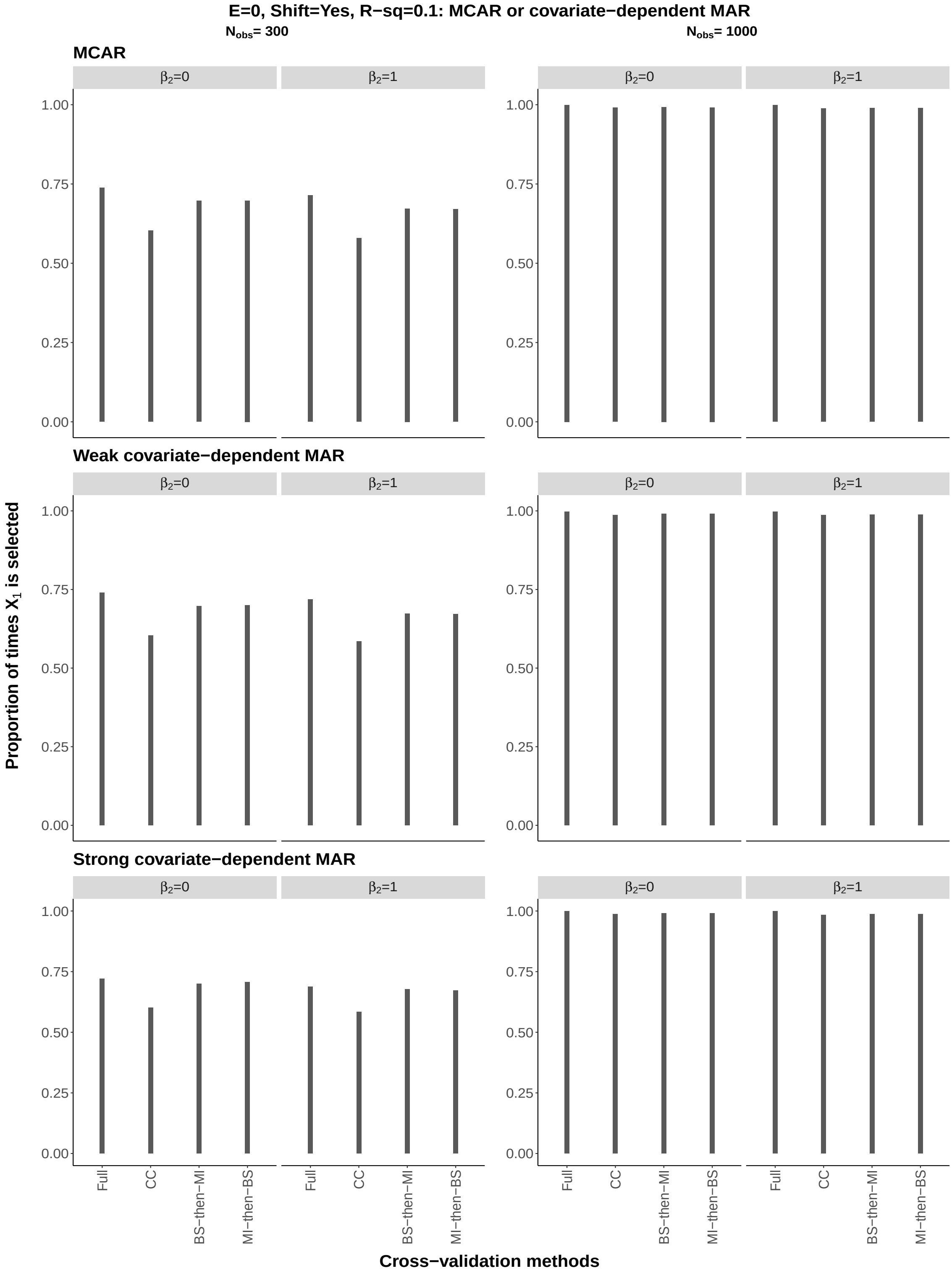

Figure S181: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

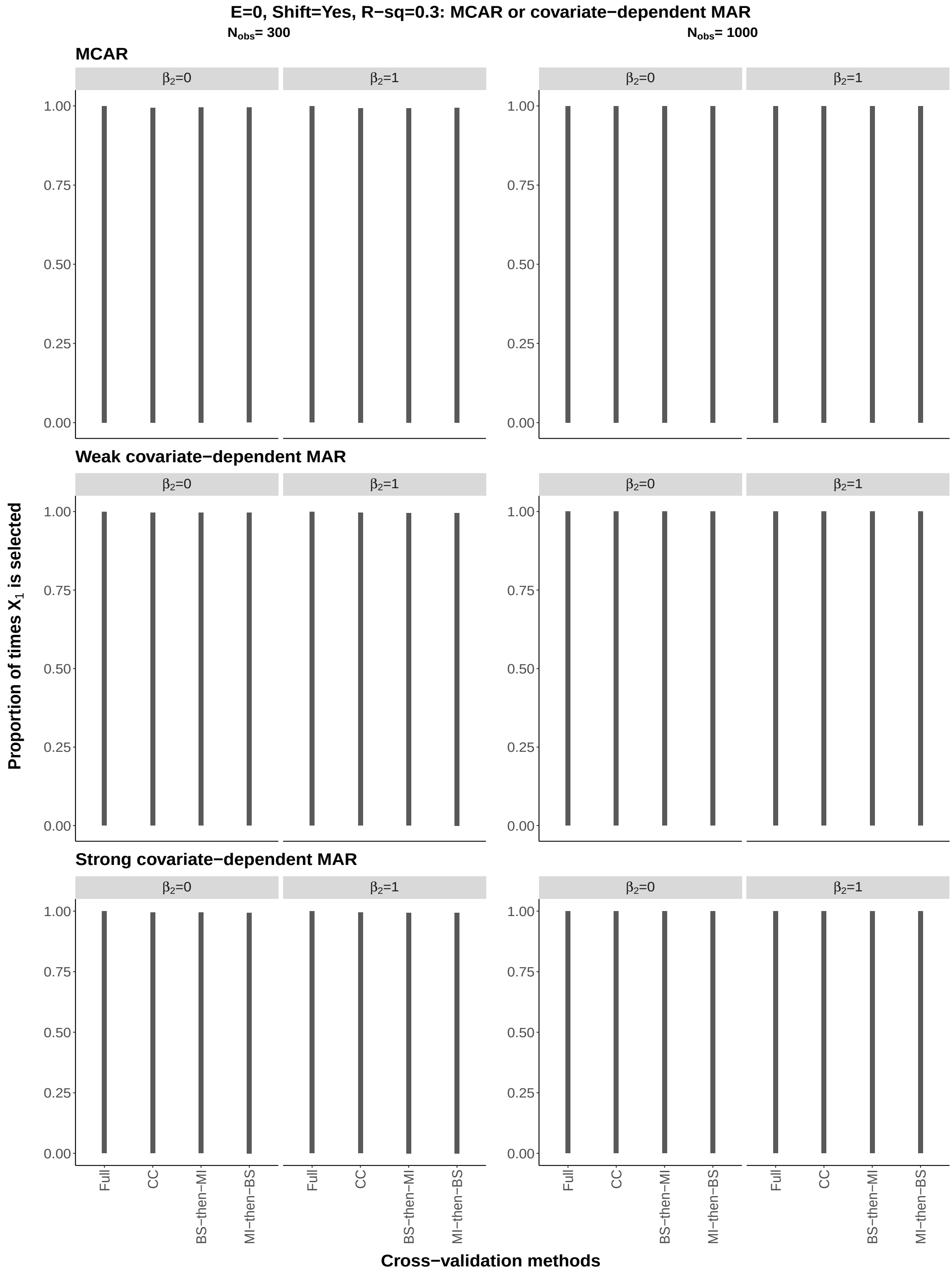

Figure S182: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

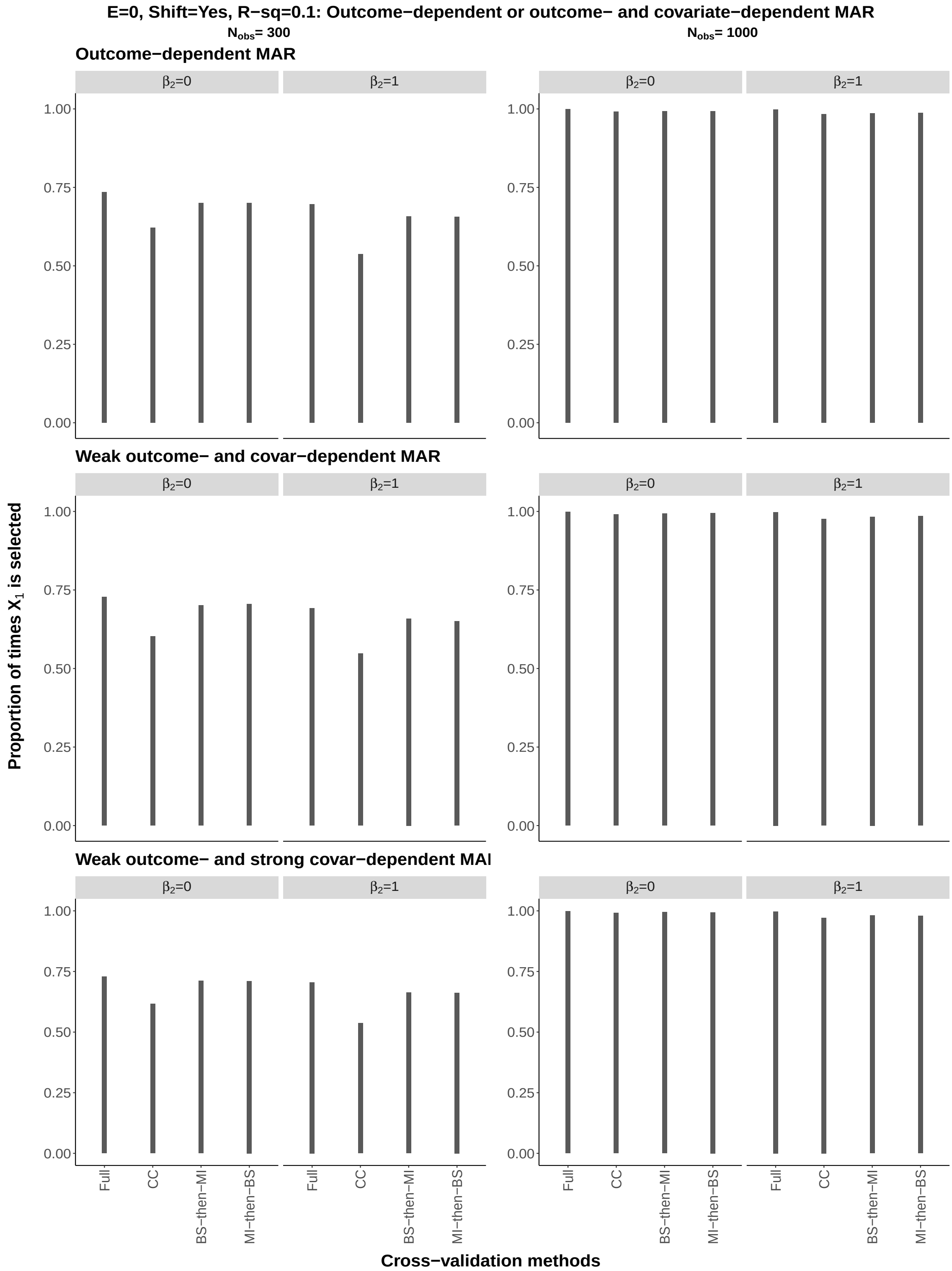

Figure S183: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

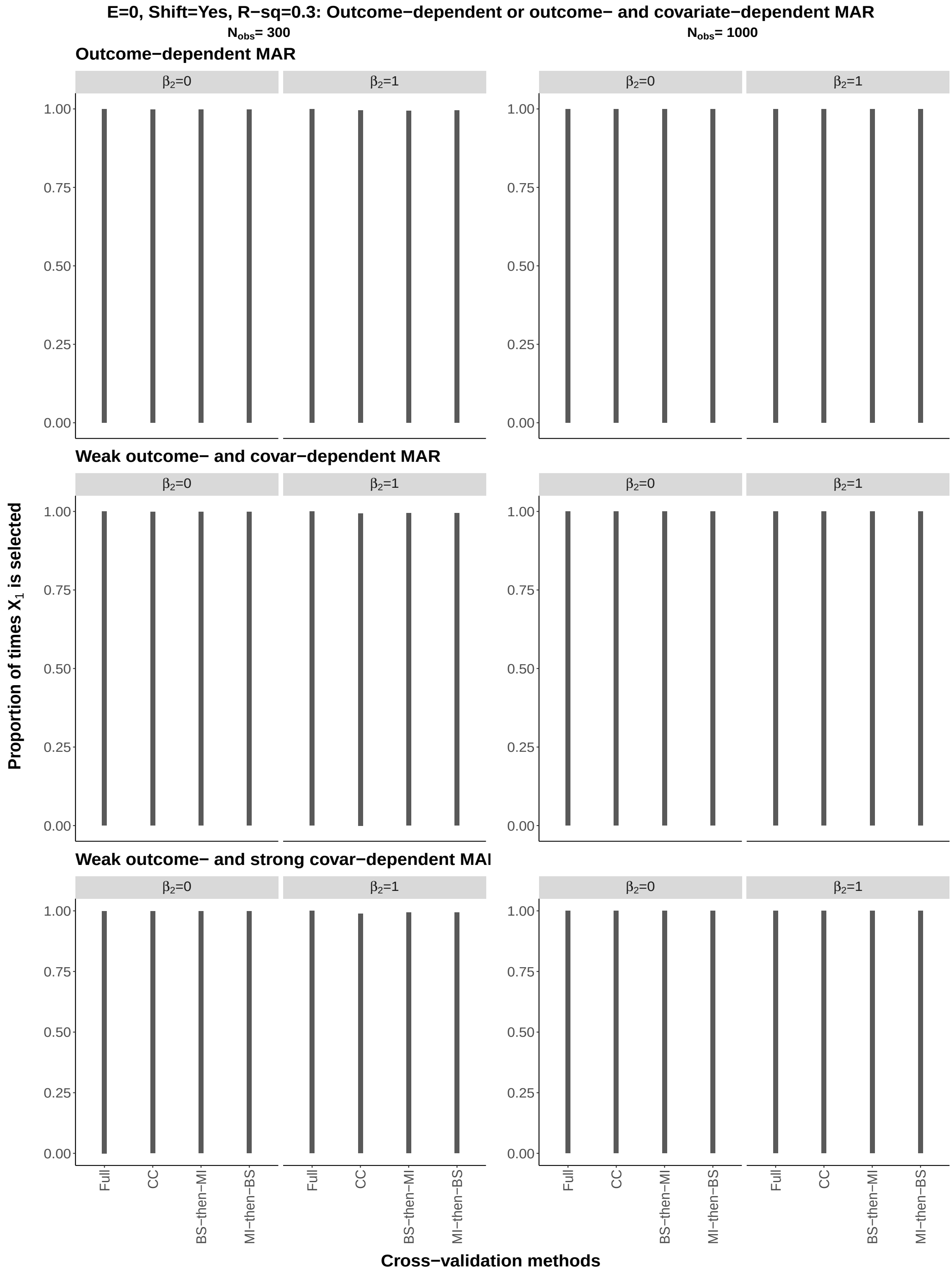

Figure S184: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

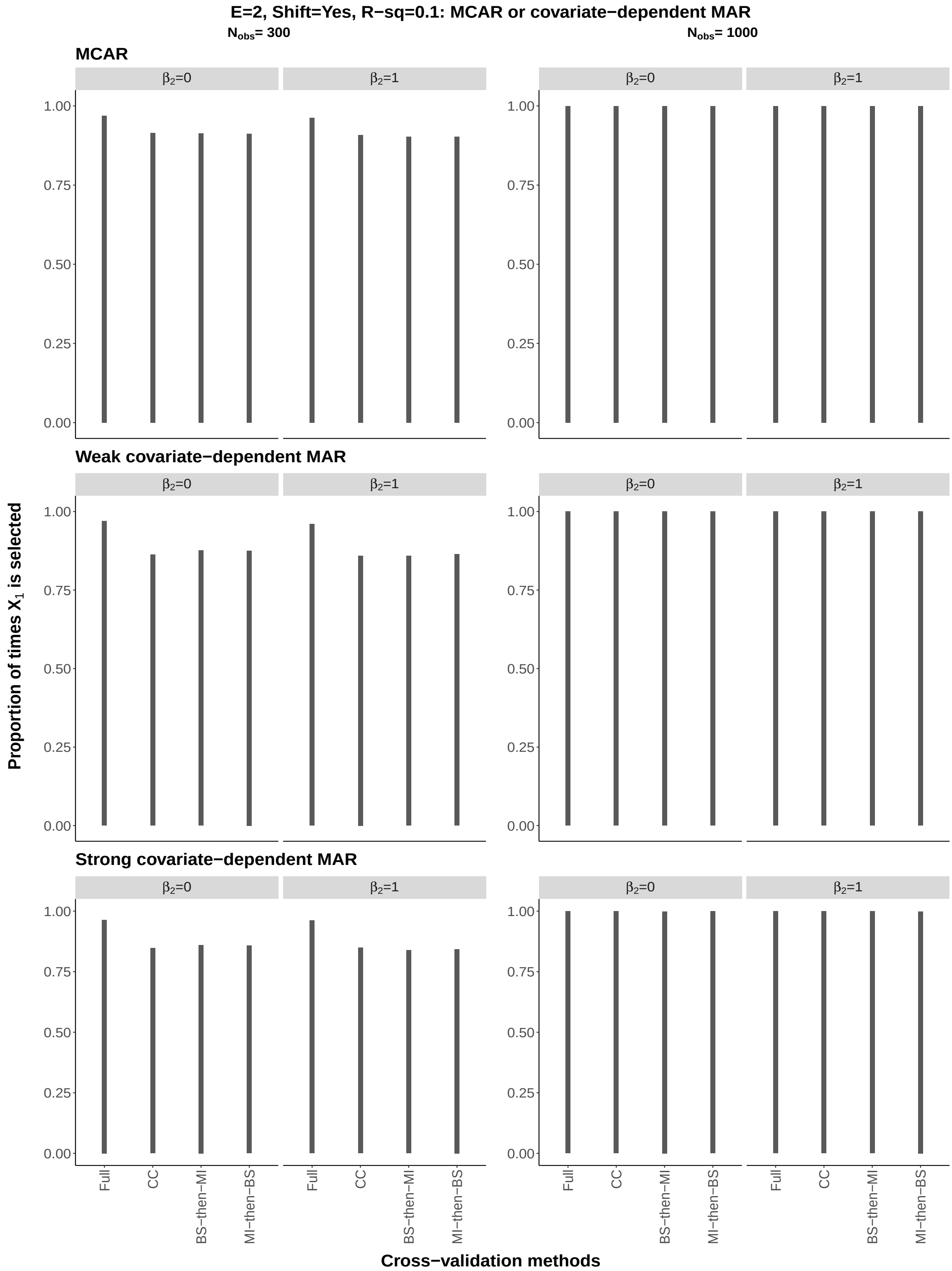

Figure S185: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

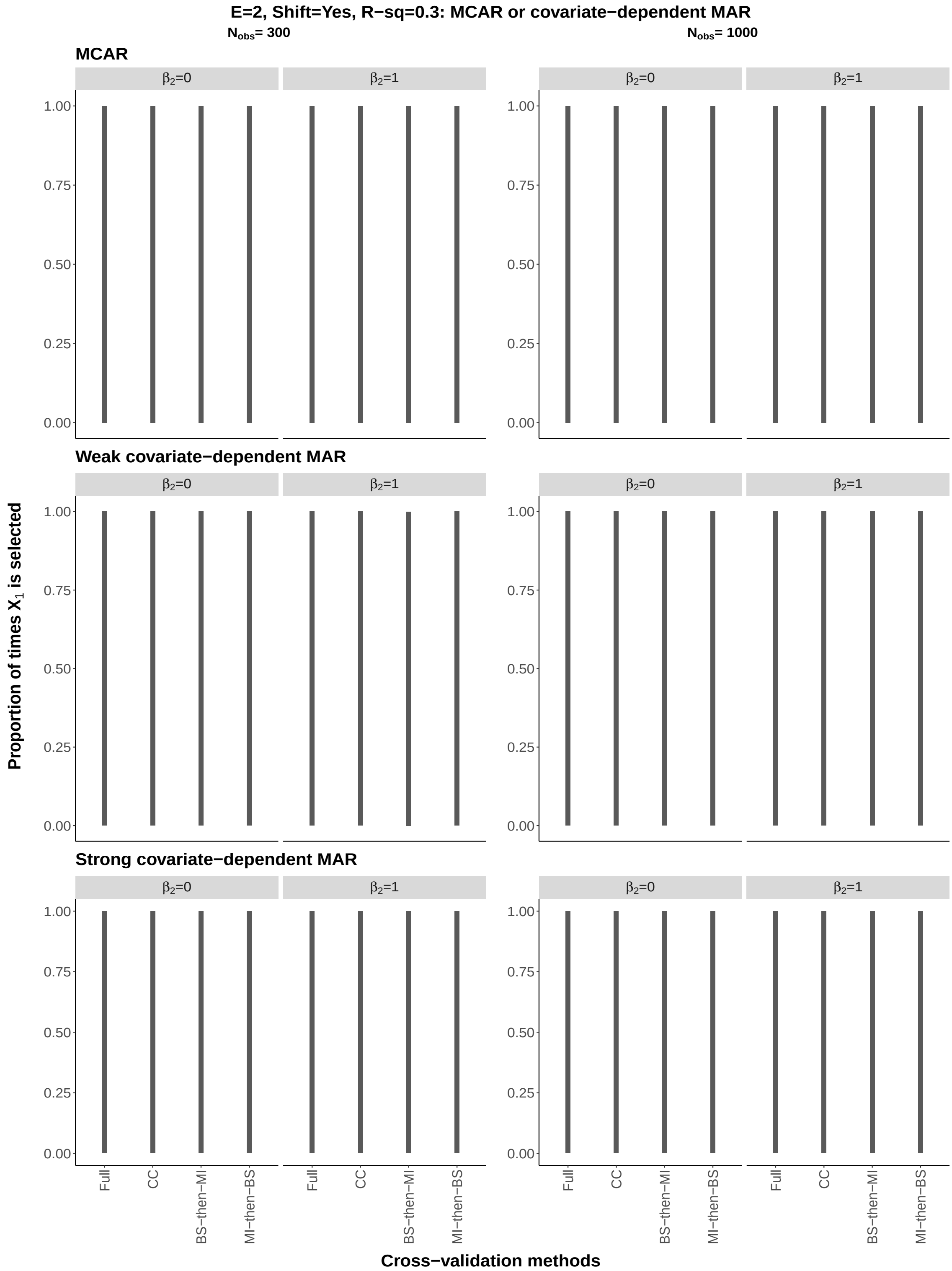

Figure S186: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

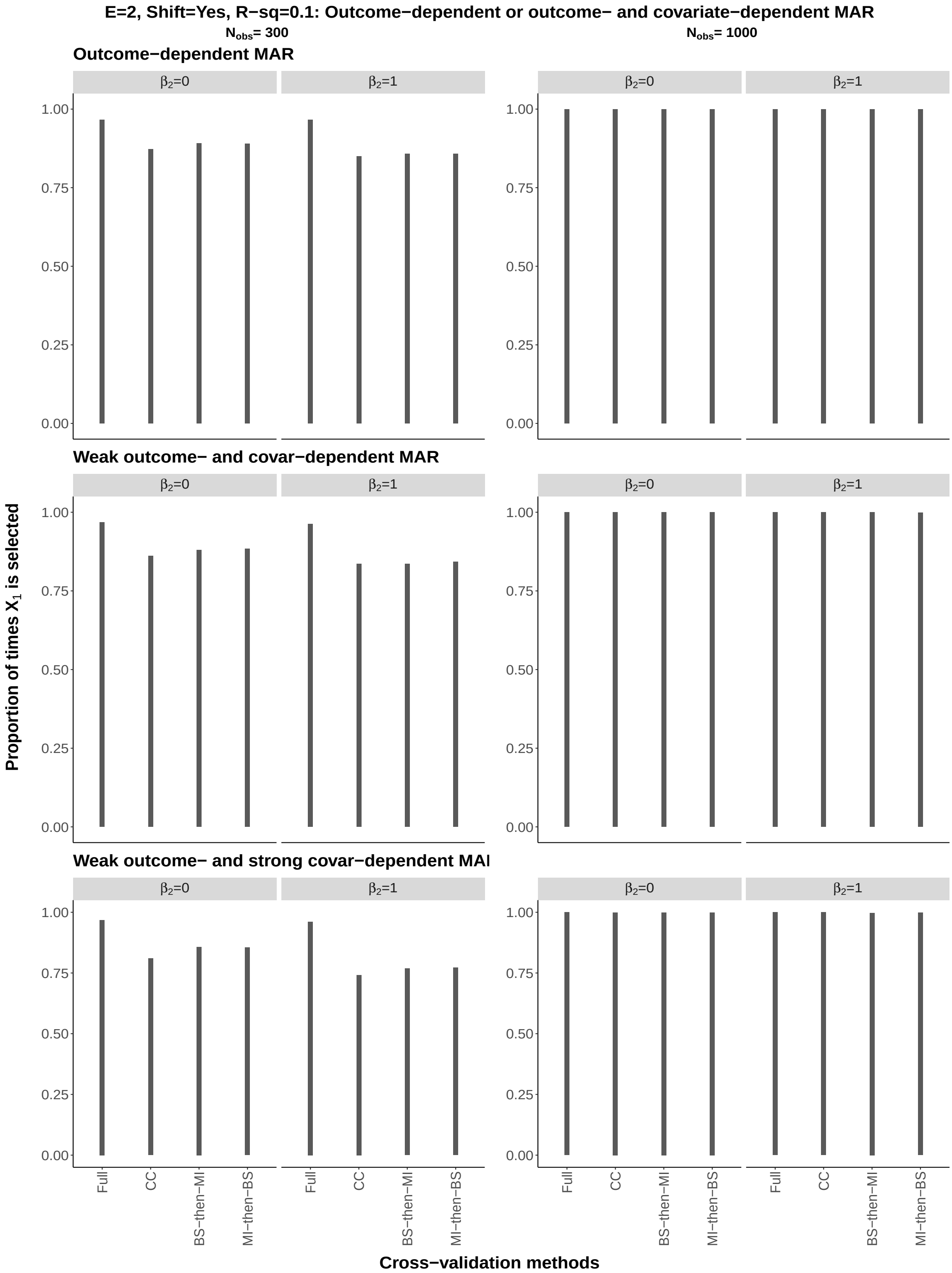

Figure S187: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

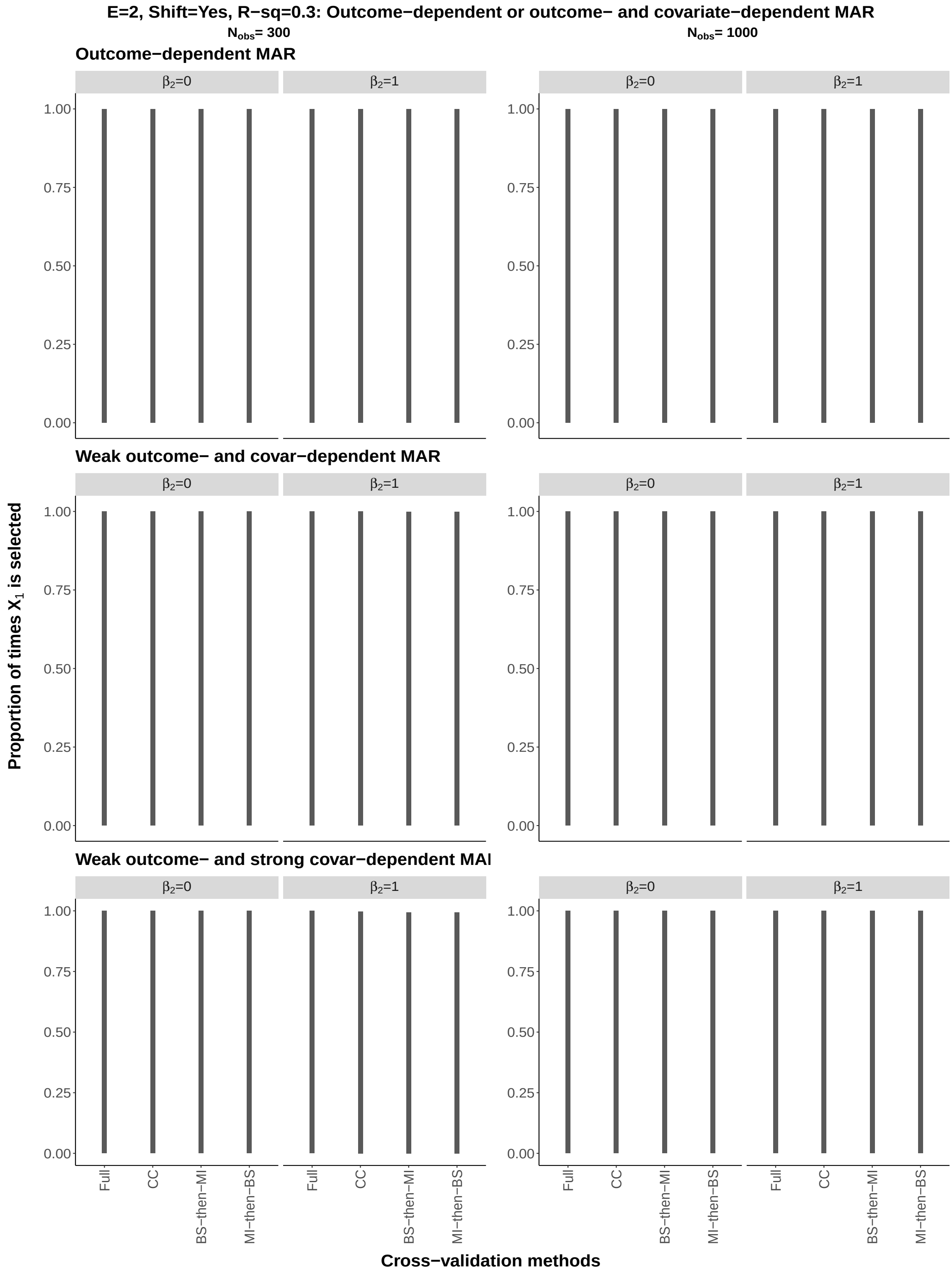

Figure S188: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

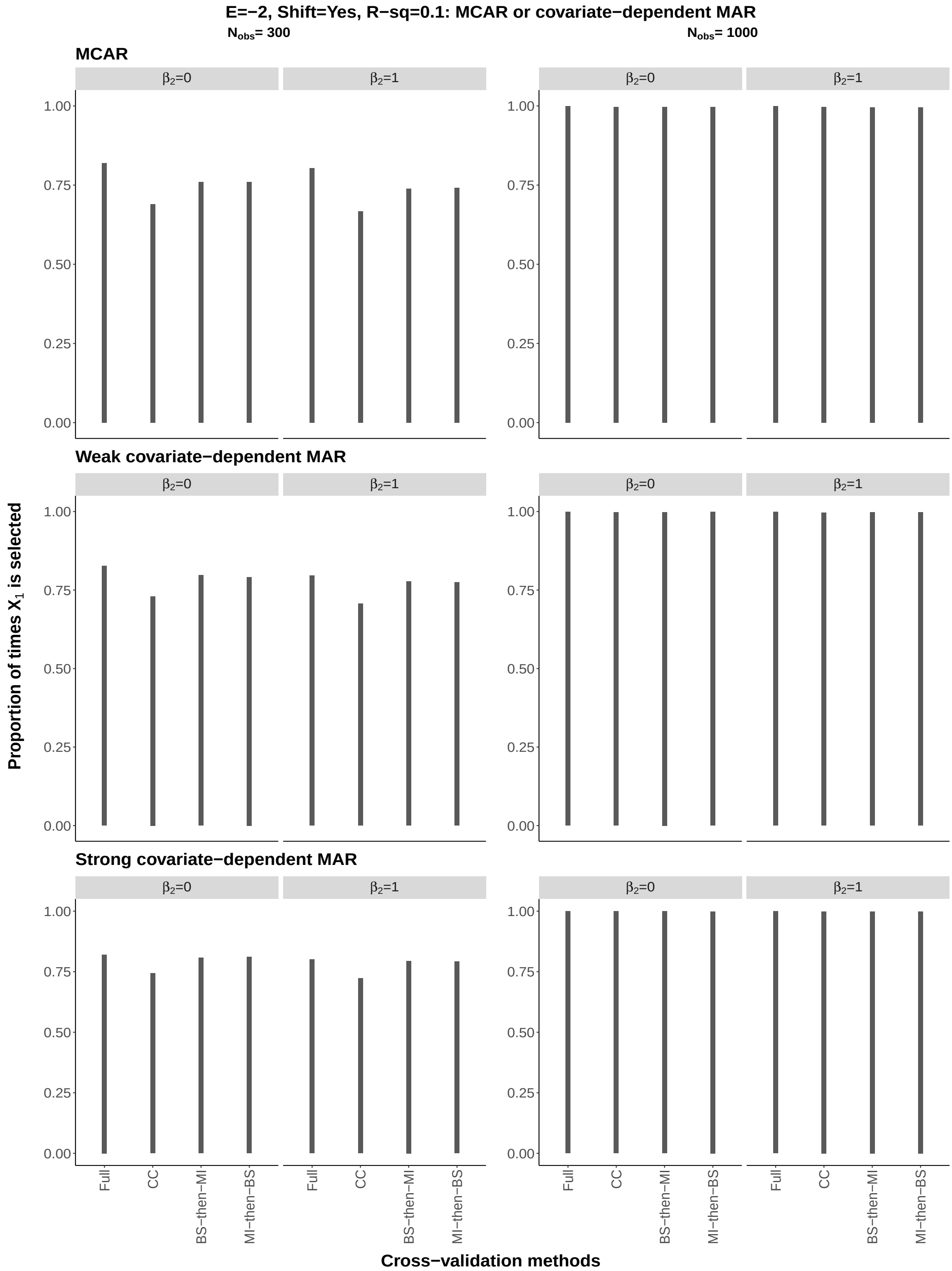

Figure S189: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

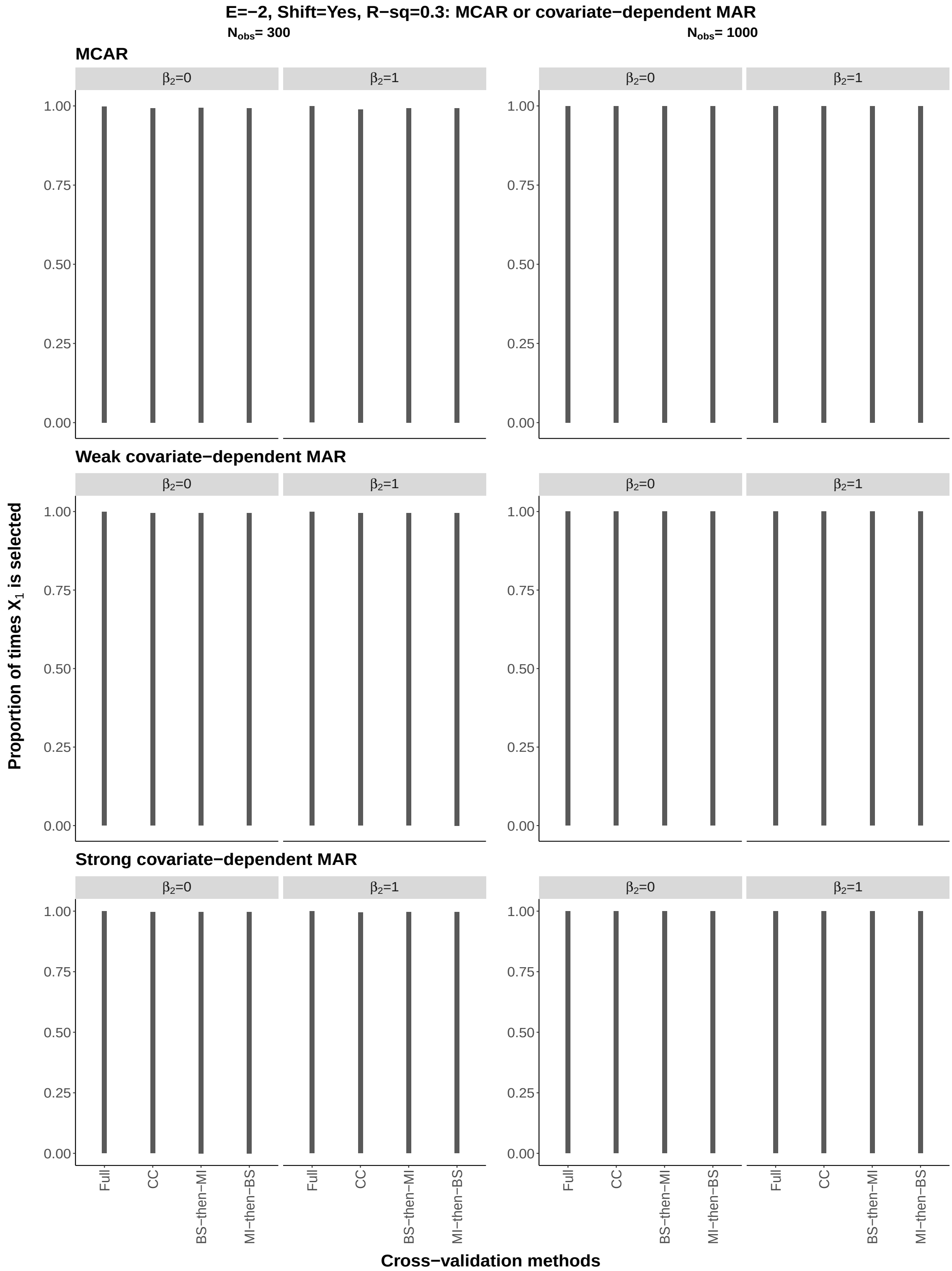

Figure S190: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

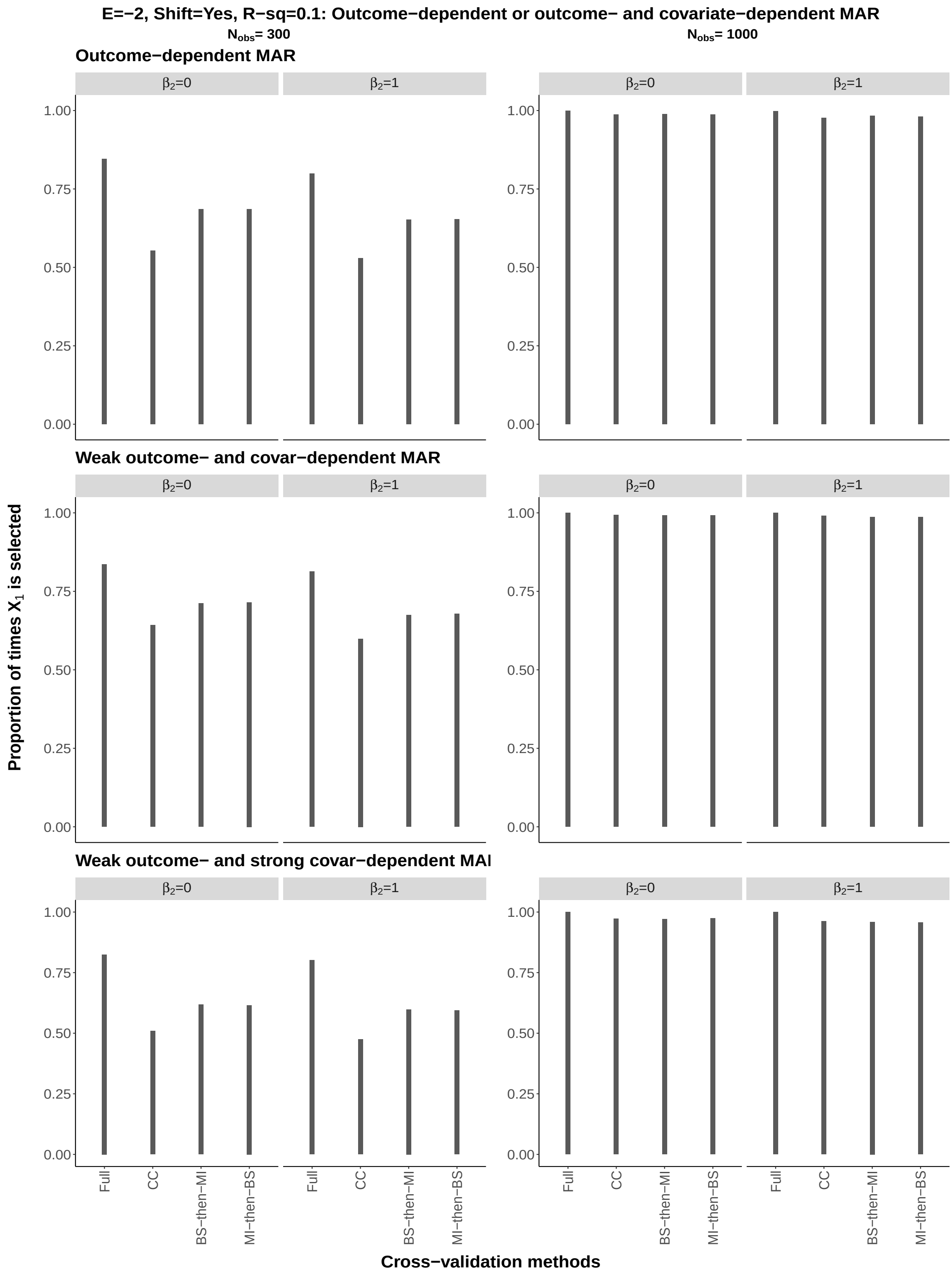

Figure S191: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

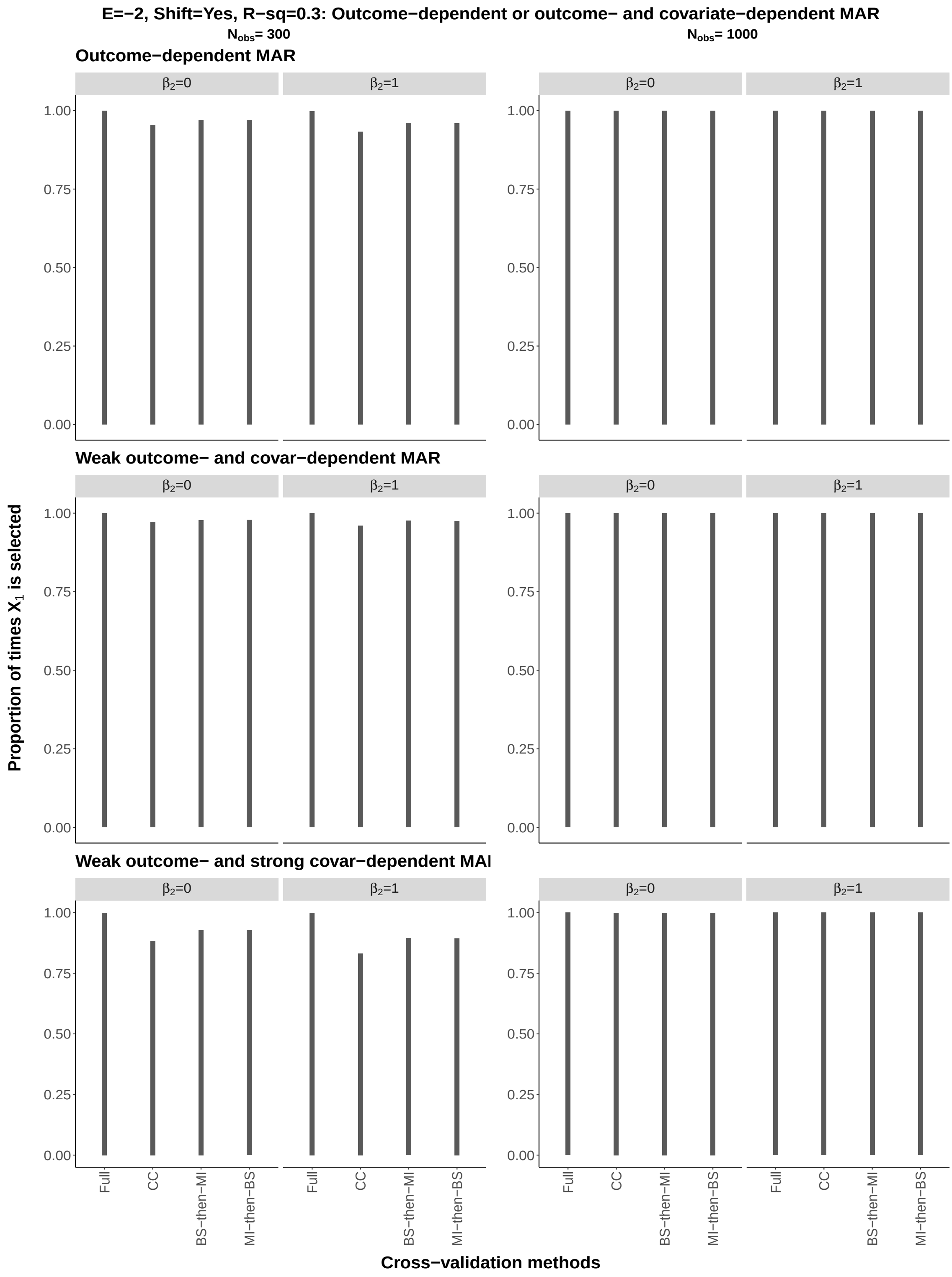

Figure S192: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.9 Covariate selection of X_2 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 1$, no origin-shift

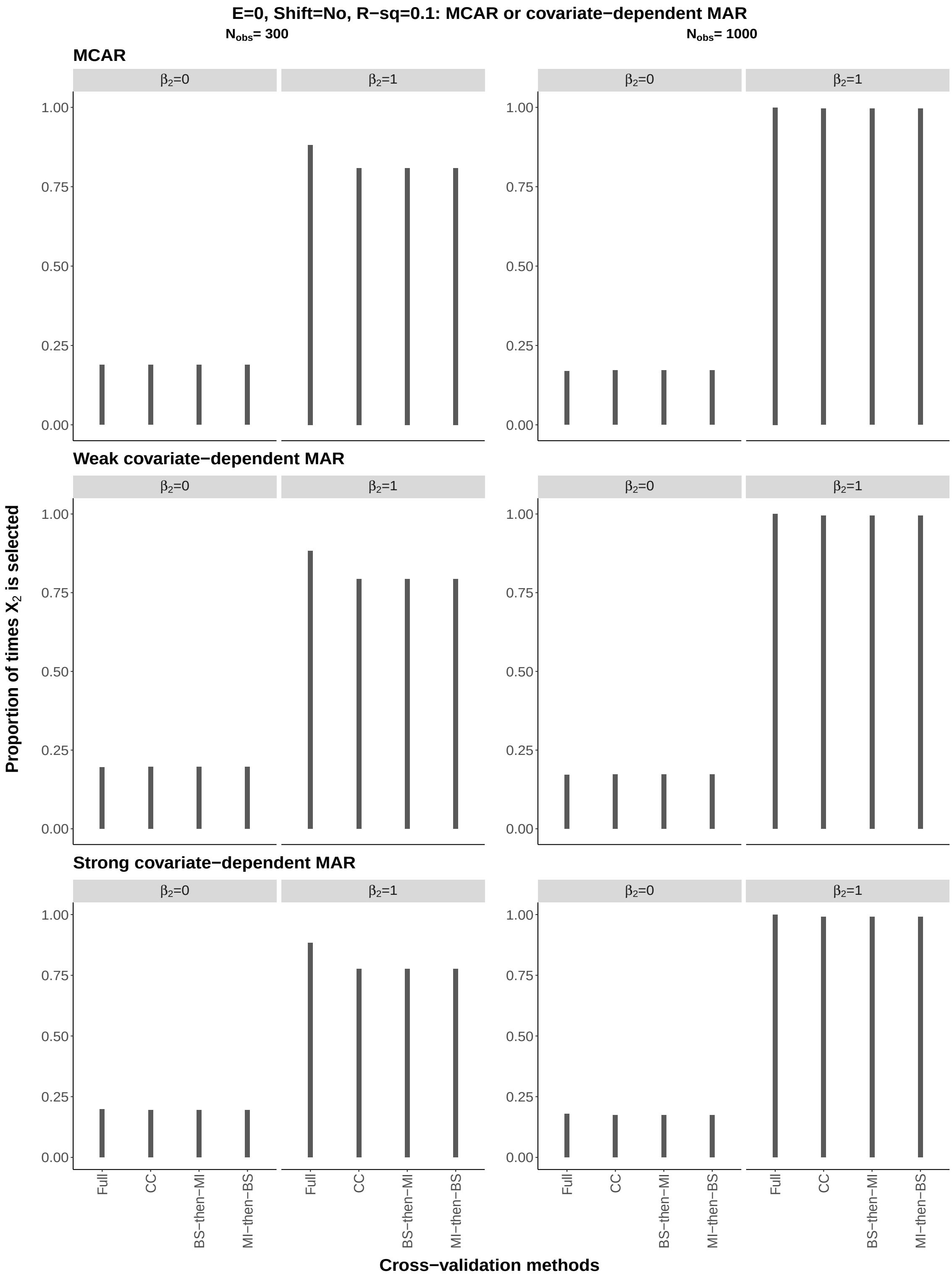

Figure S193: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

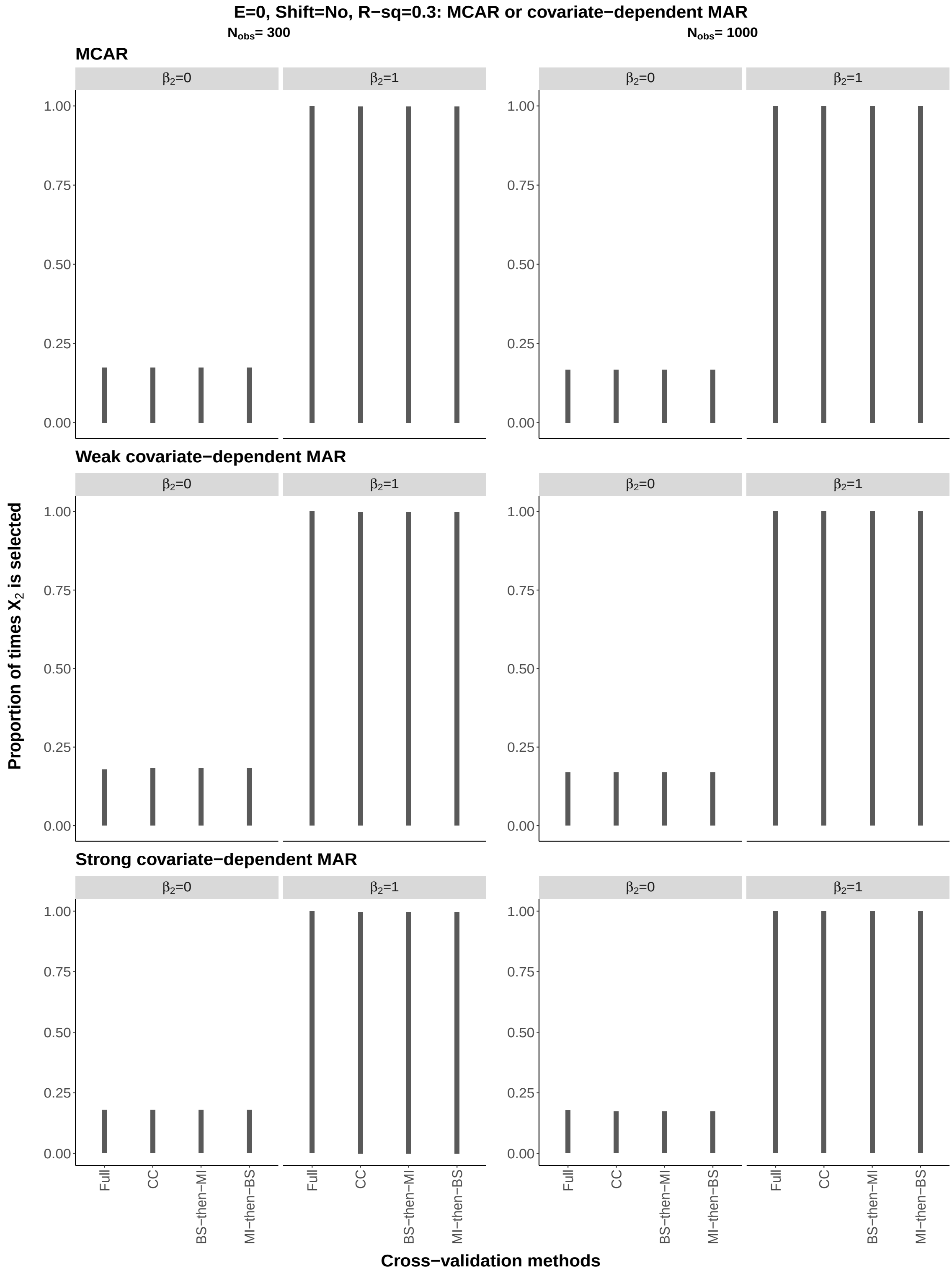

Figure S194: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

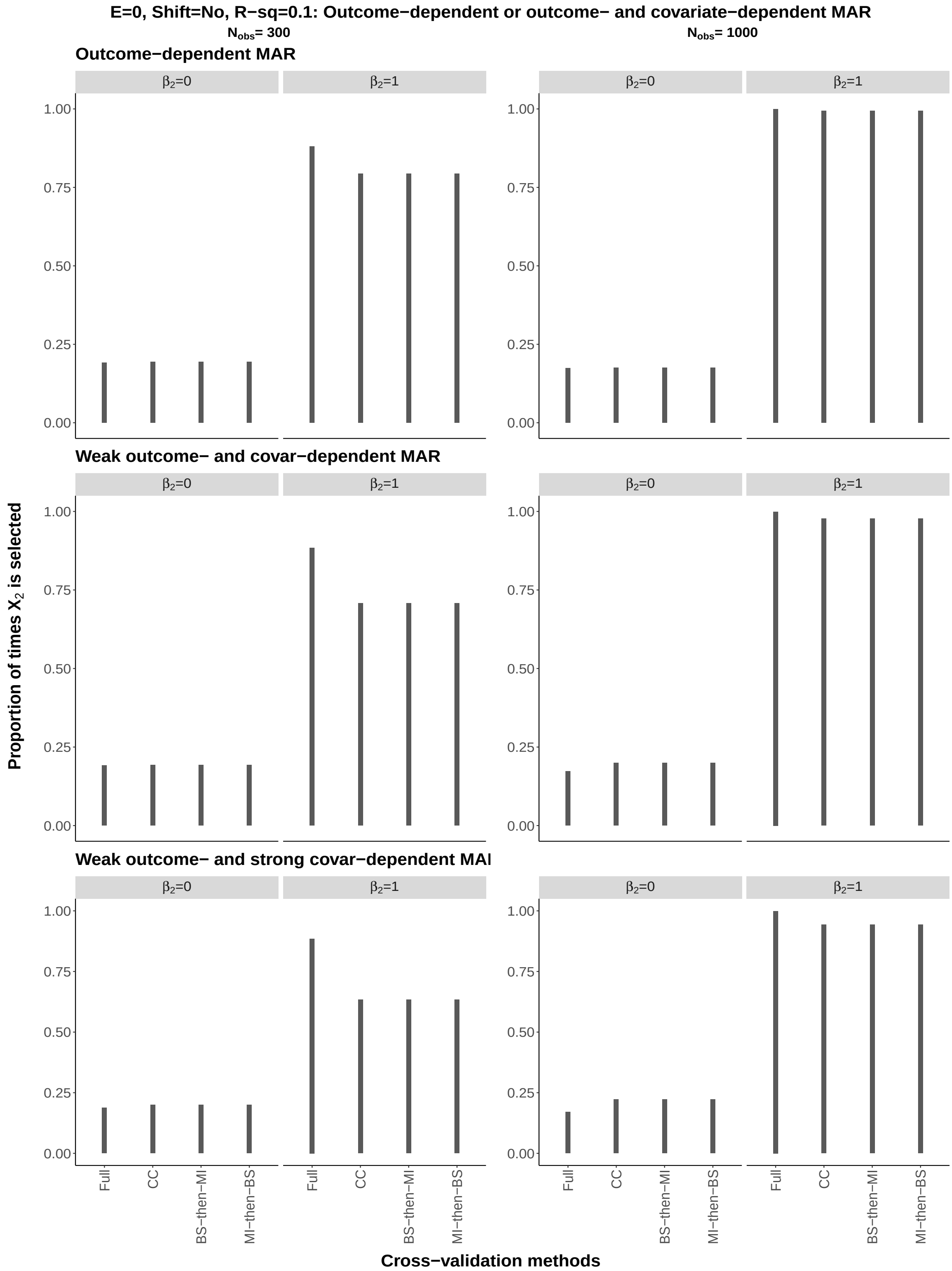

Figure S195: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

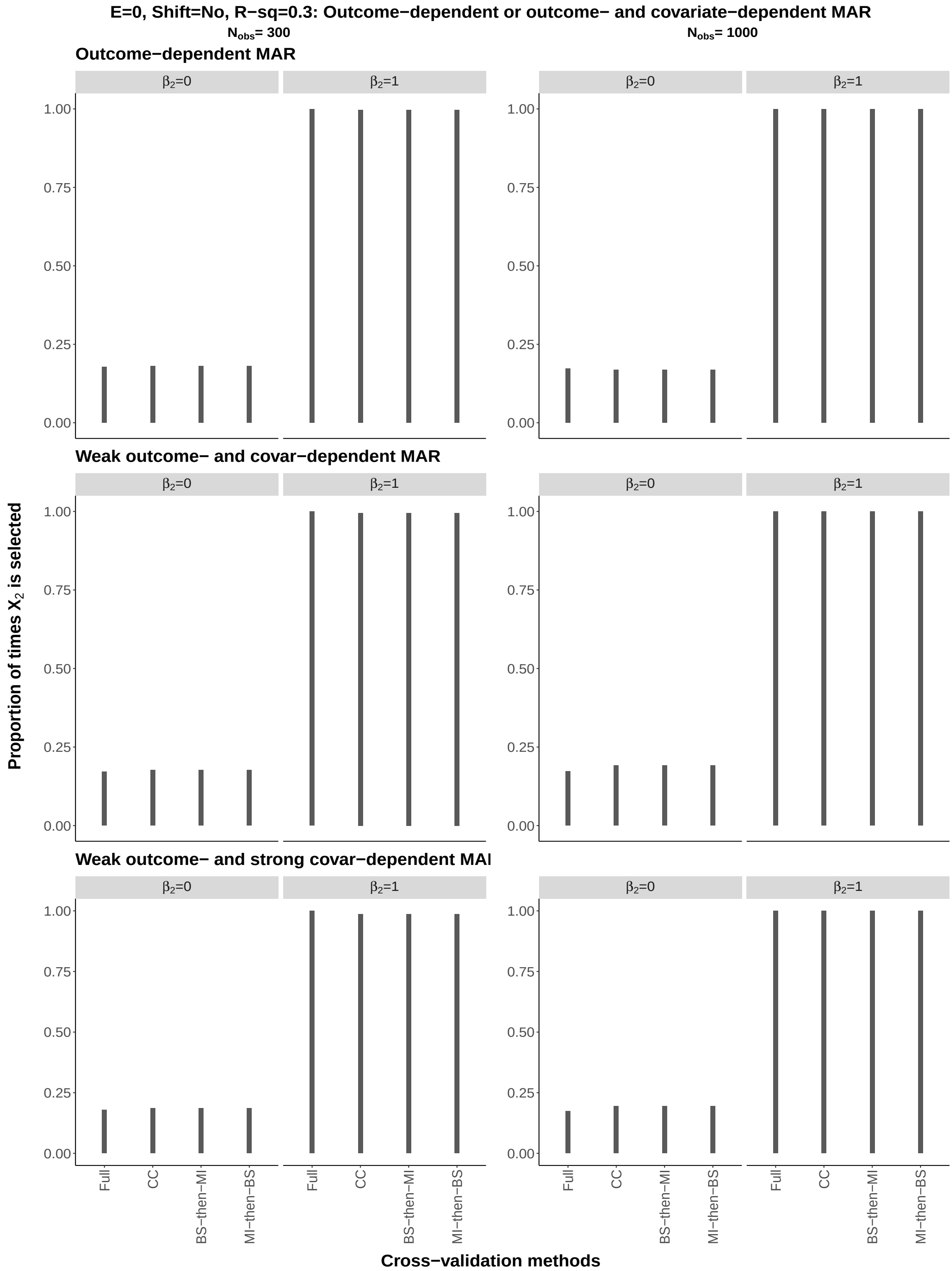

Figure S196: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

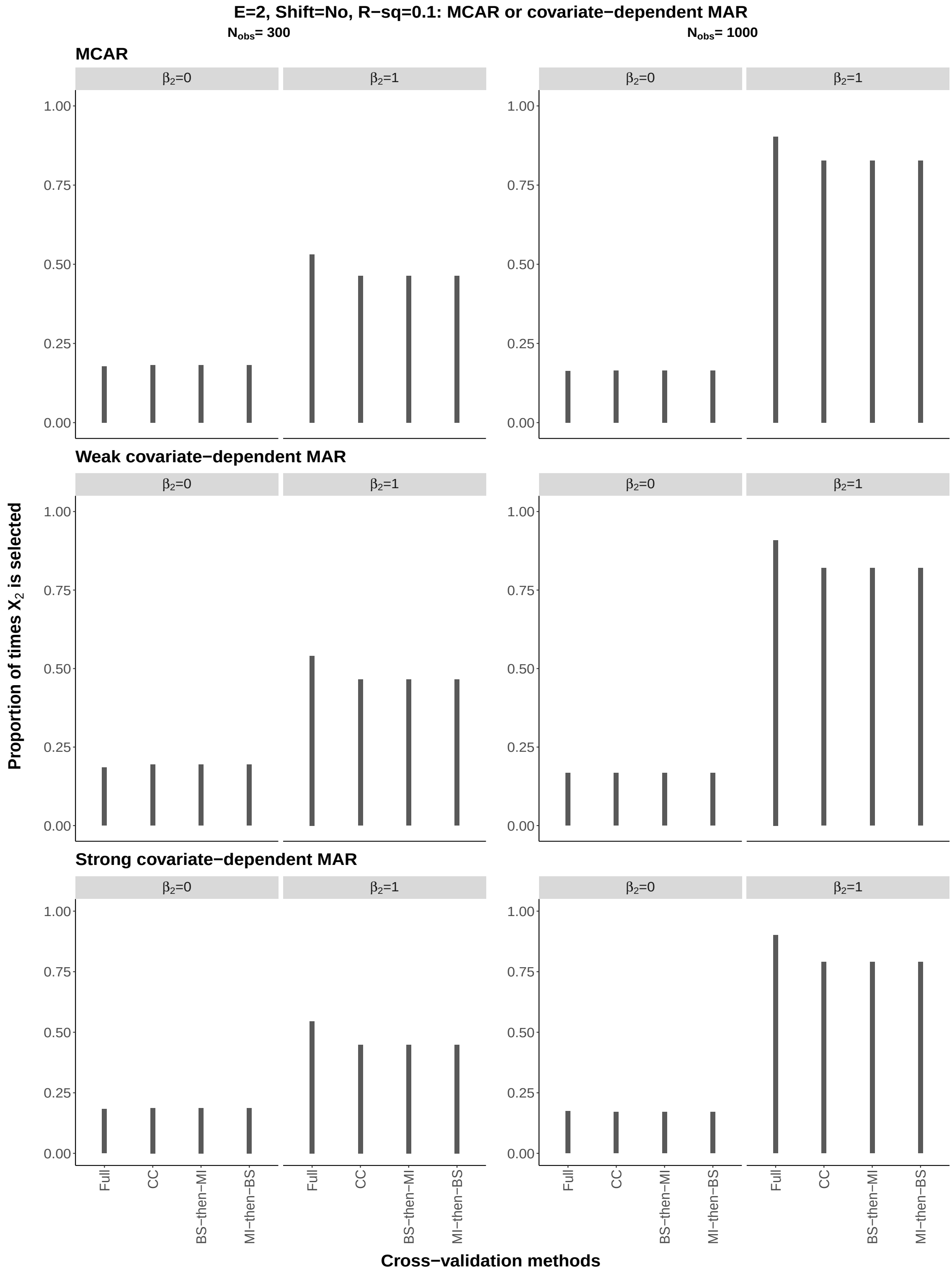

Figure S197: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

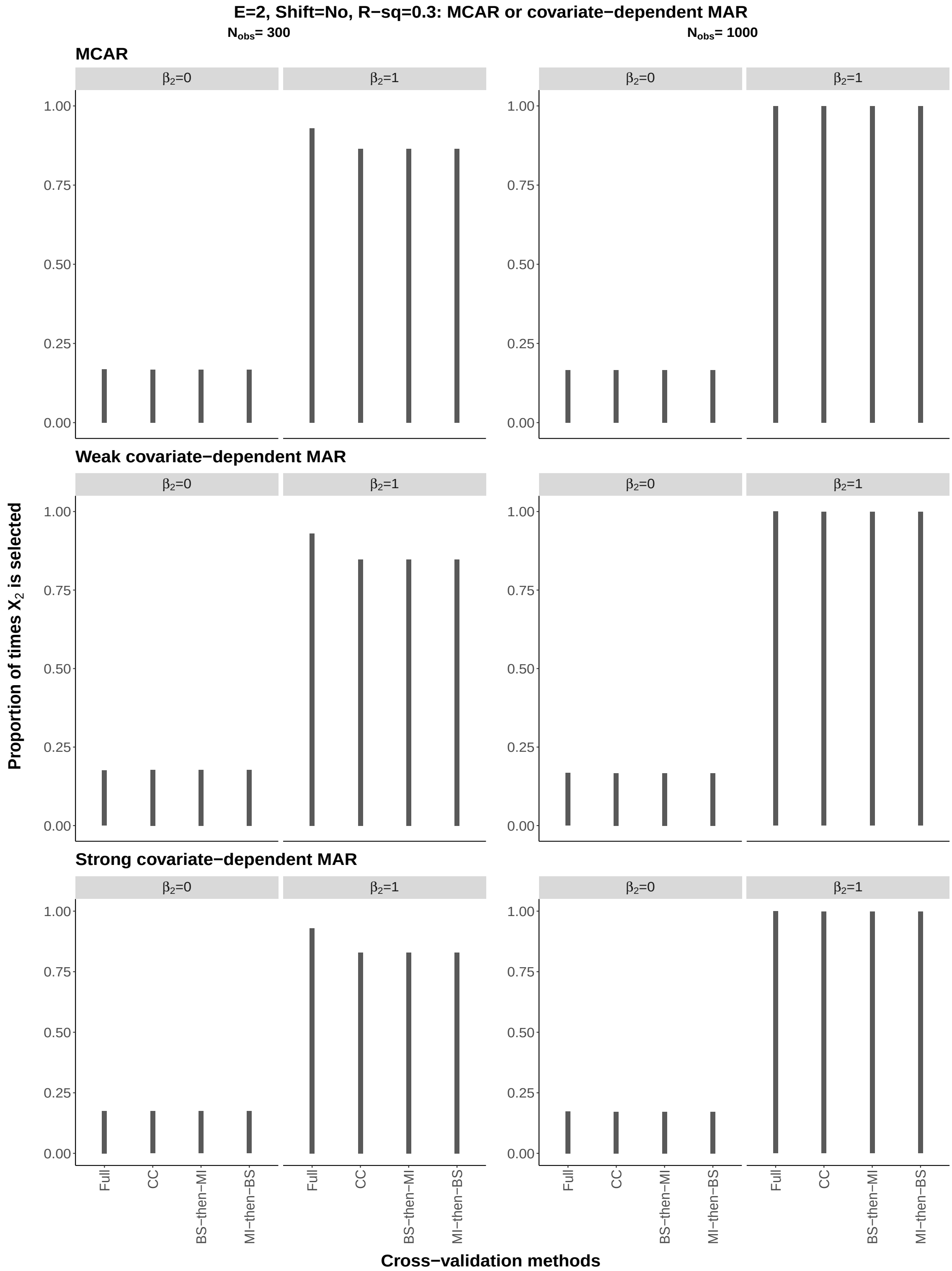

Figure S198: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

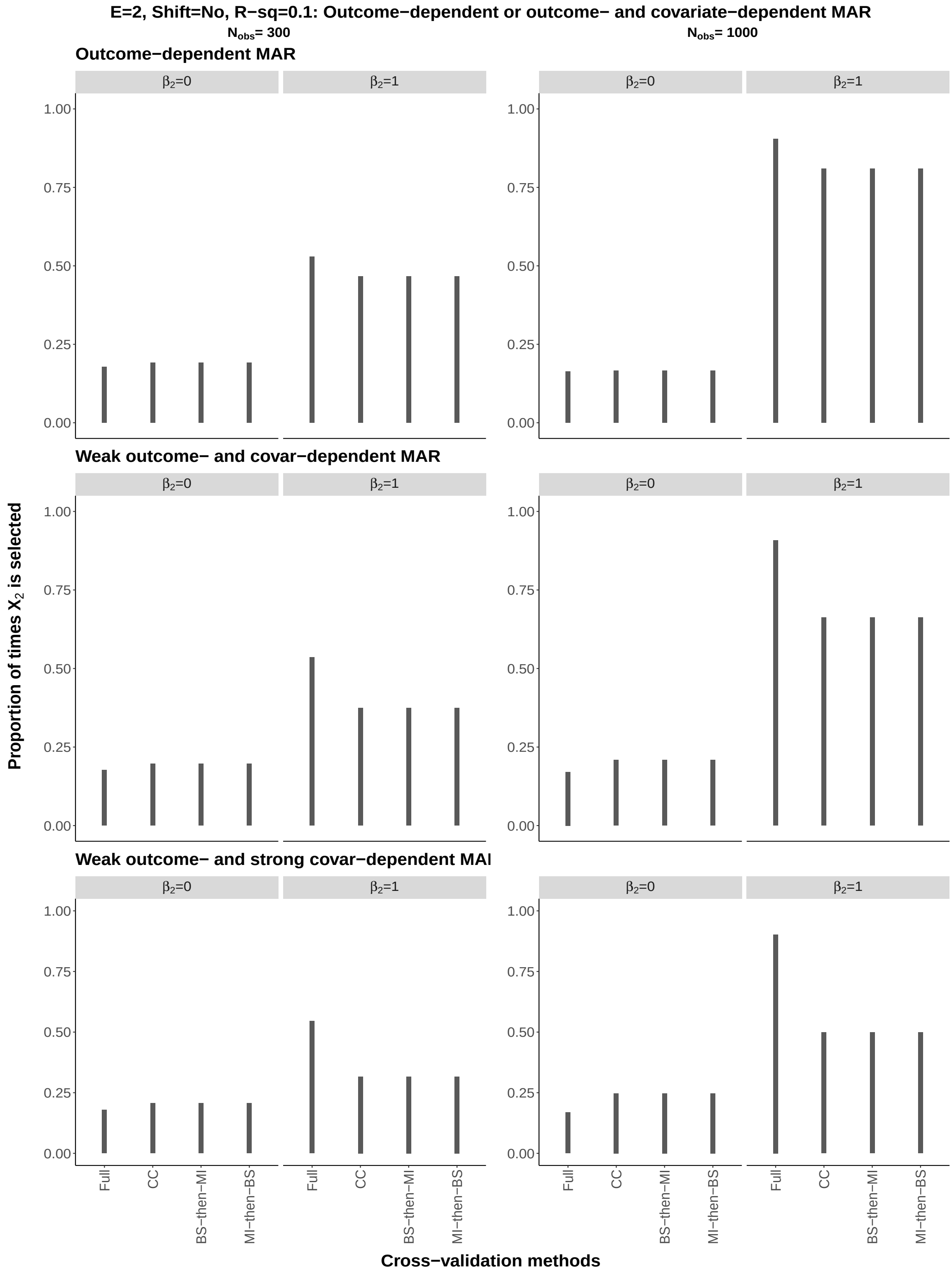

Figure S199: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

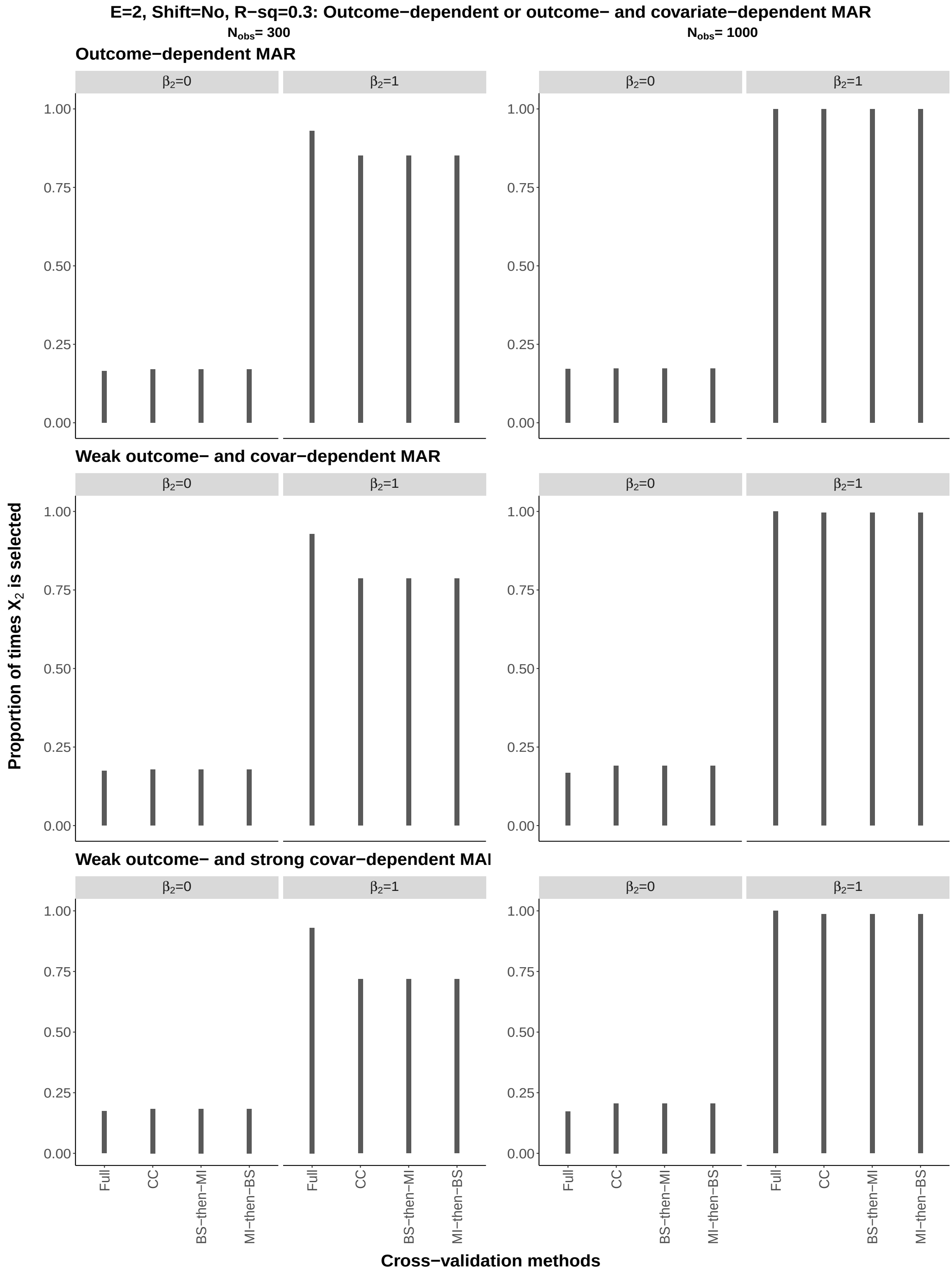

Figure S200: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

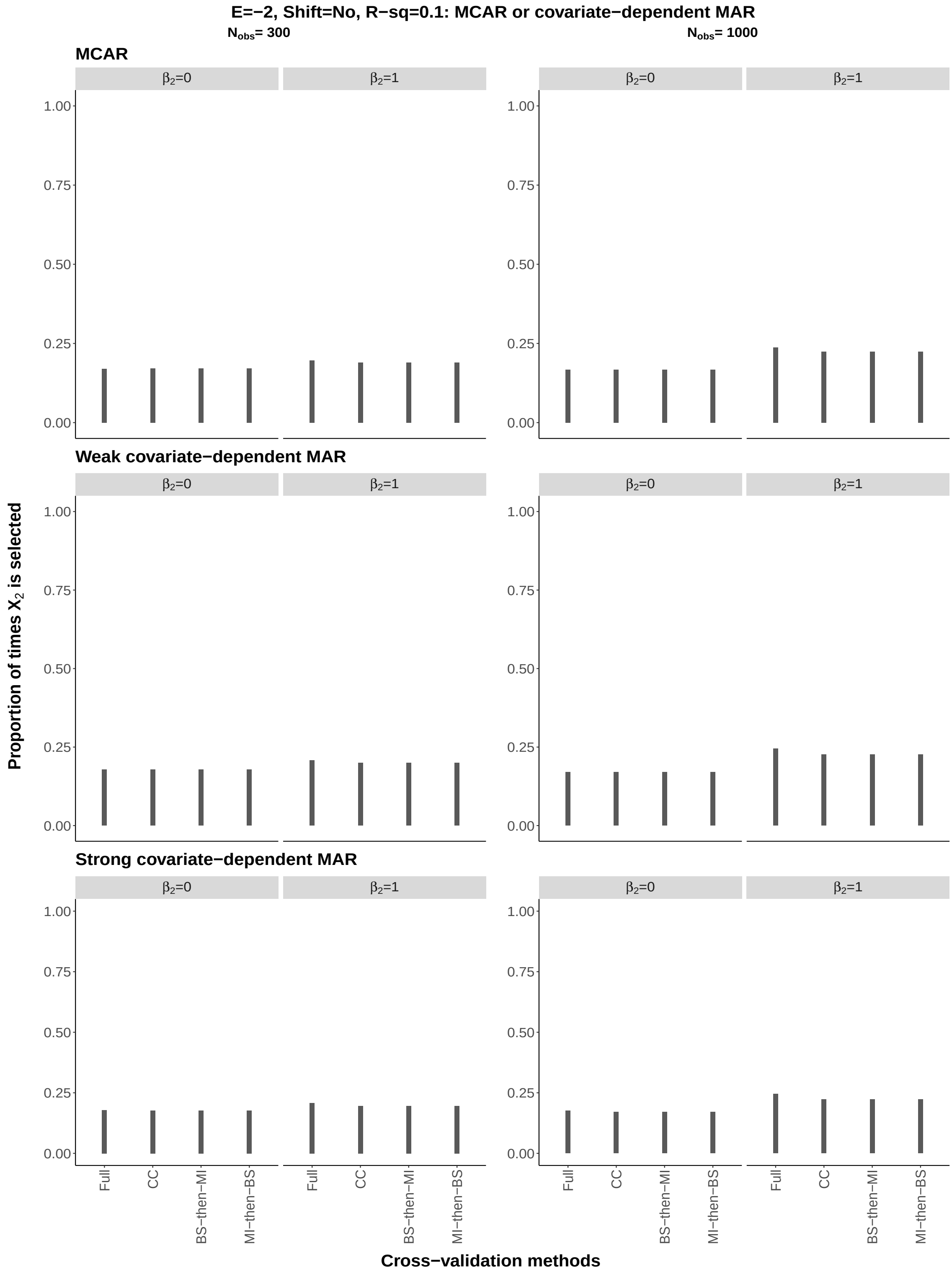

Figure S201: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

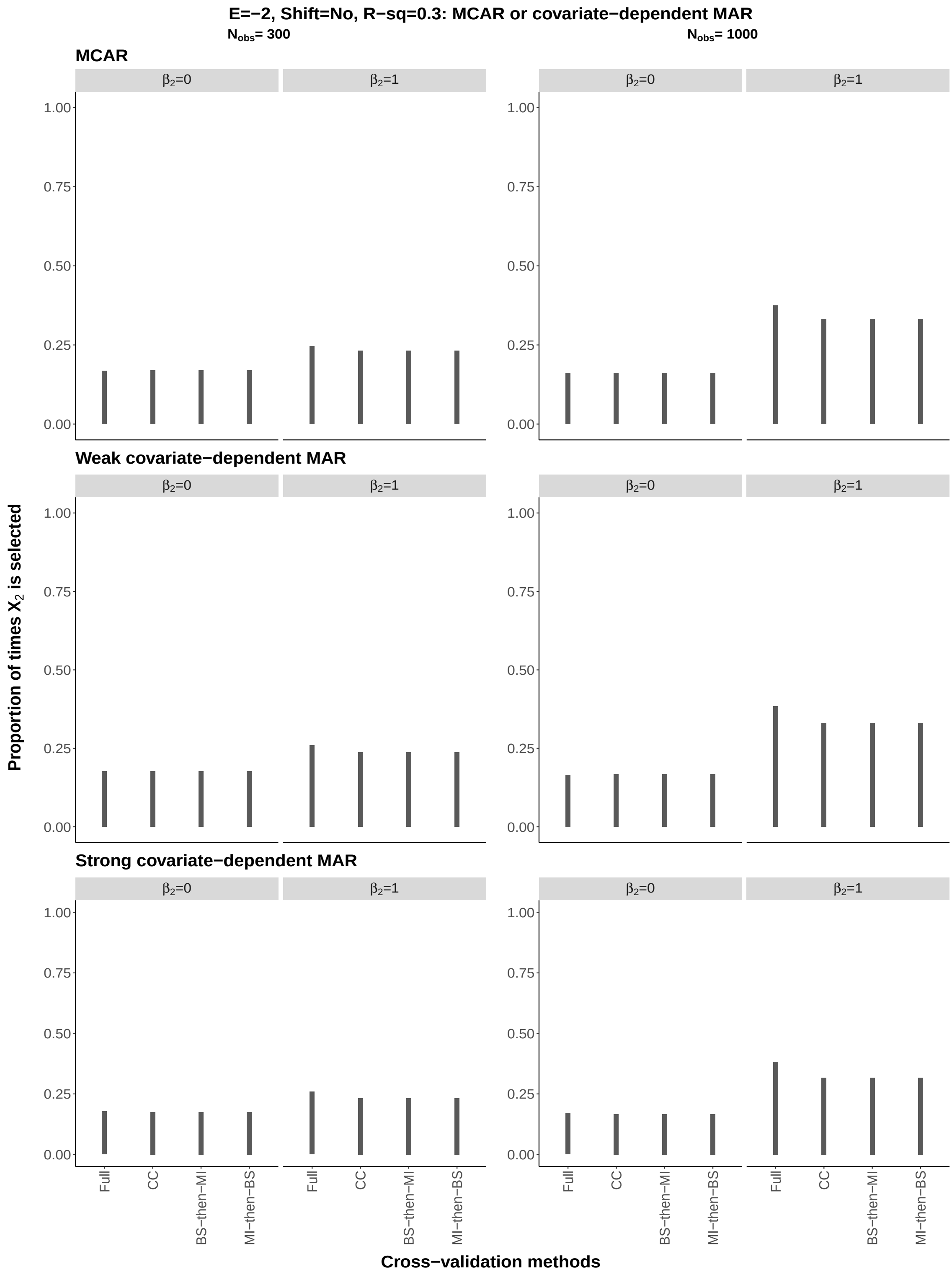

Figure S202: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

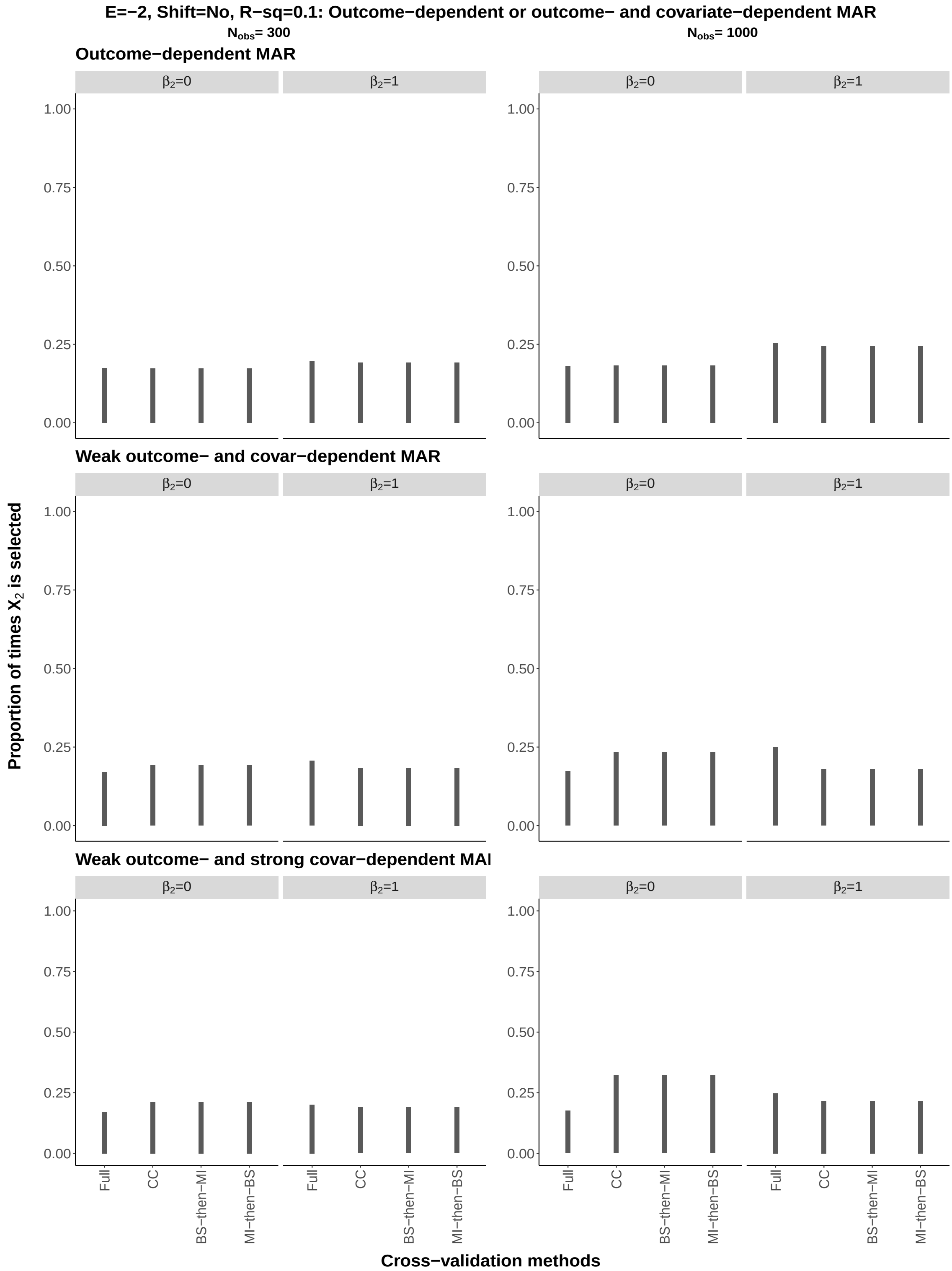

Figure S203: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

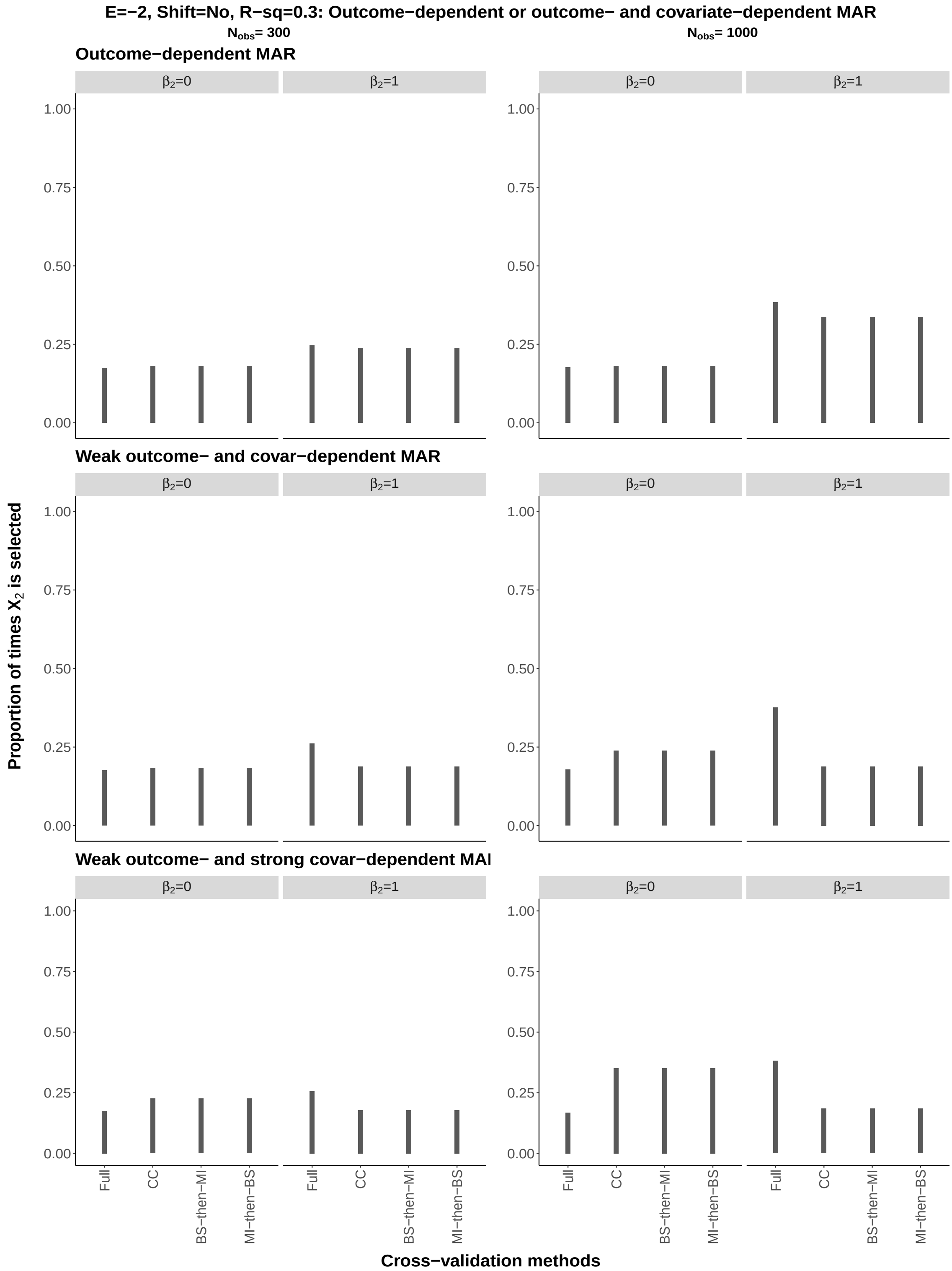

Figure S204: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.10 Covariate selection of X_2 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 0.05$, no origin-shift

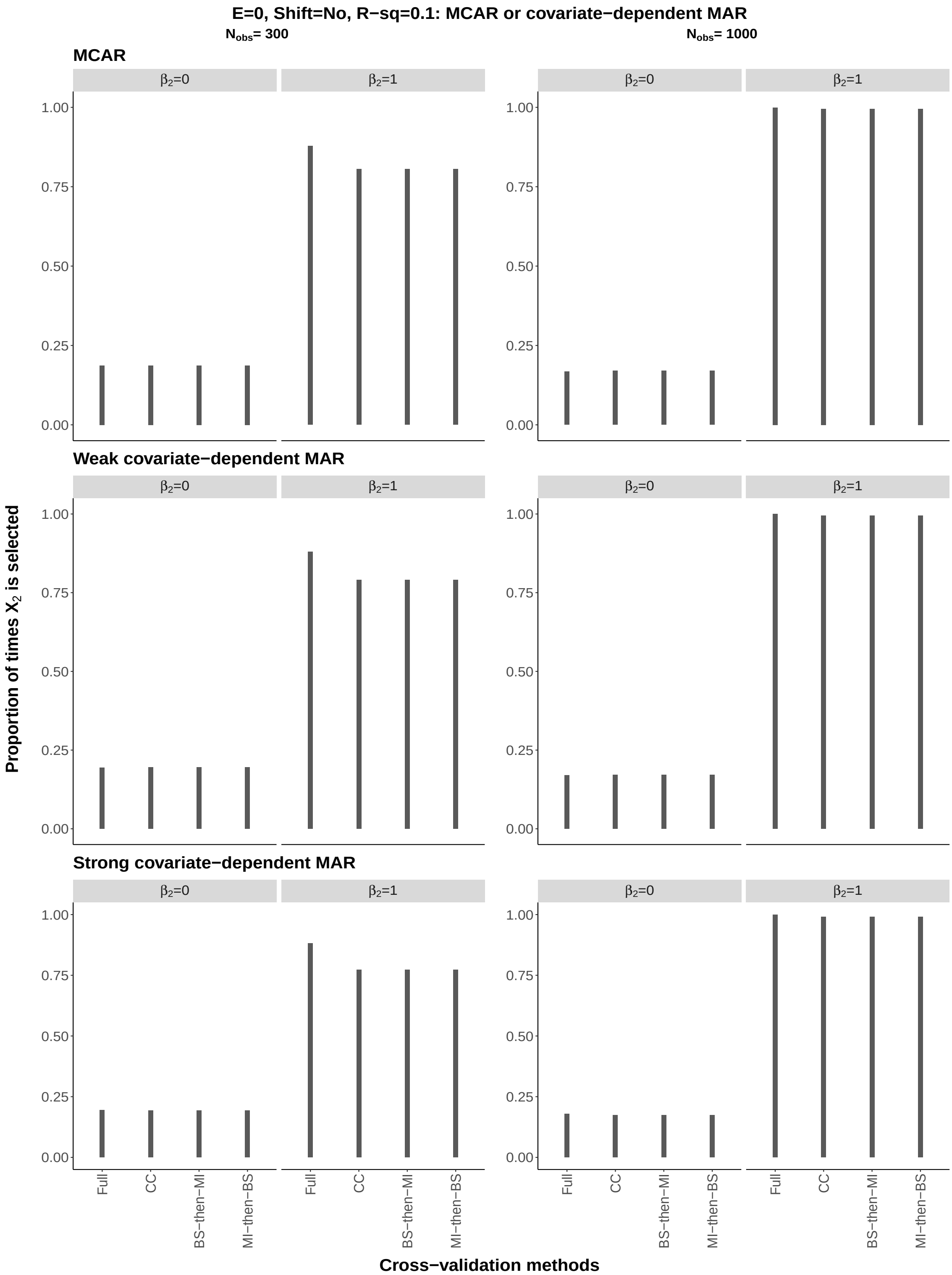

Figure S205: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

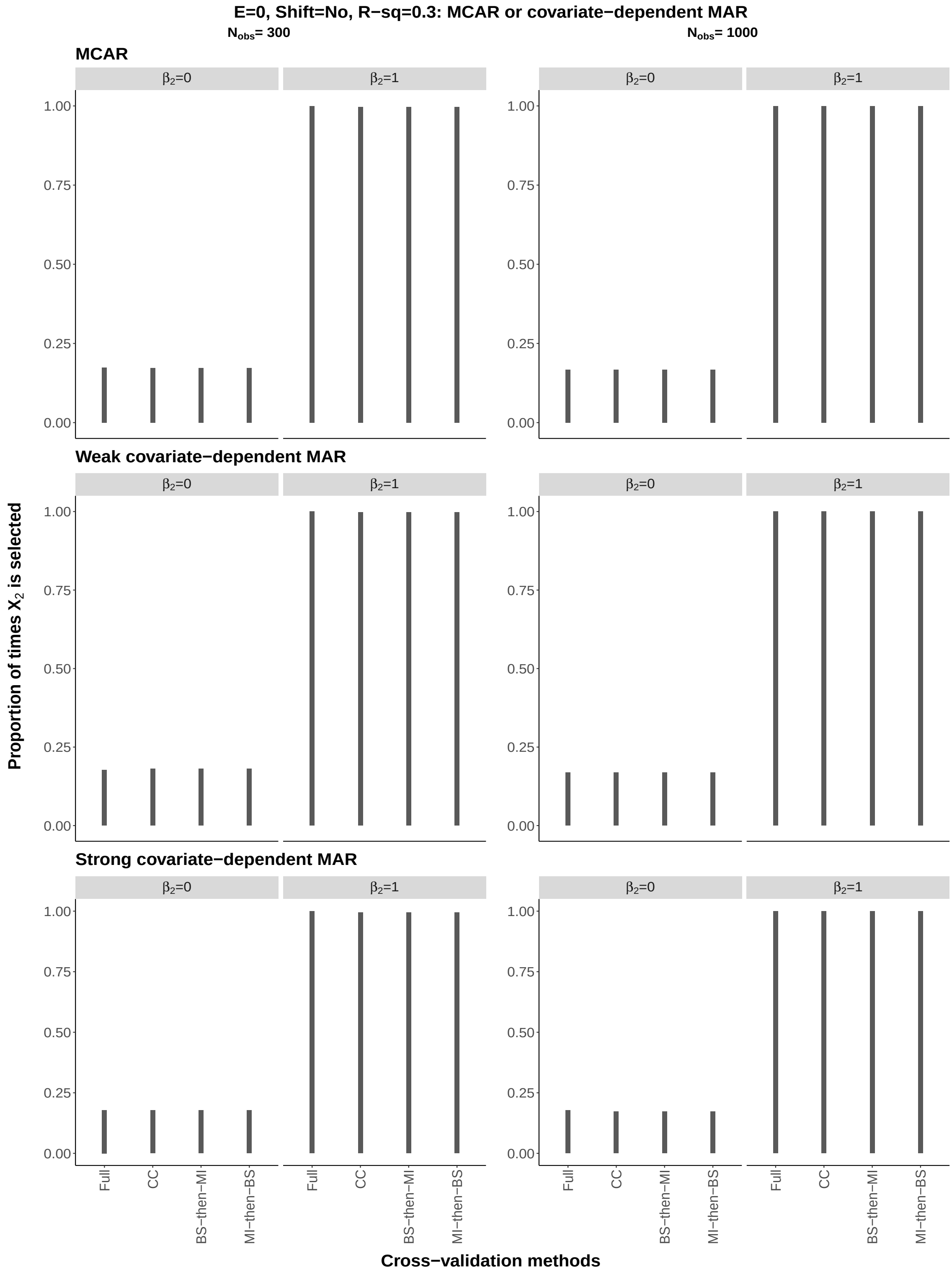

Figure S206: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

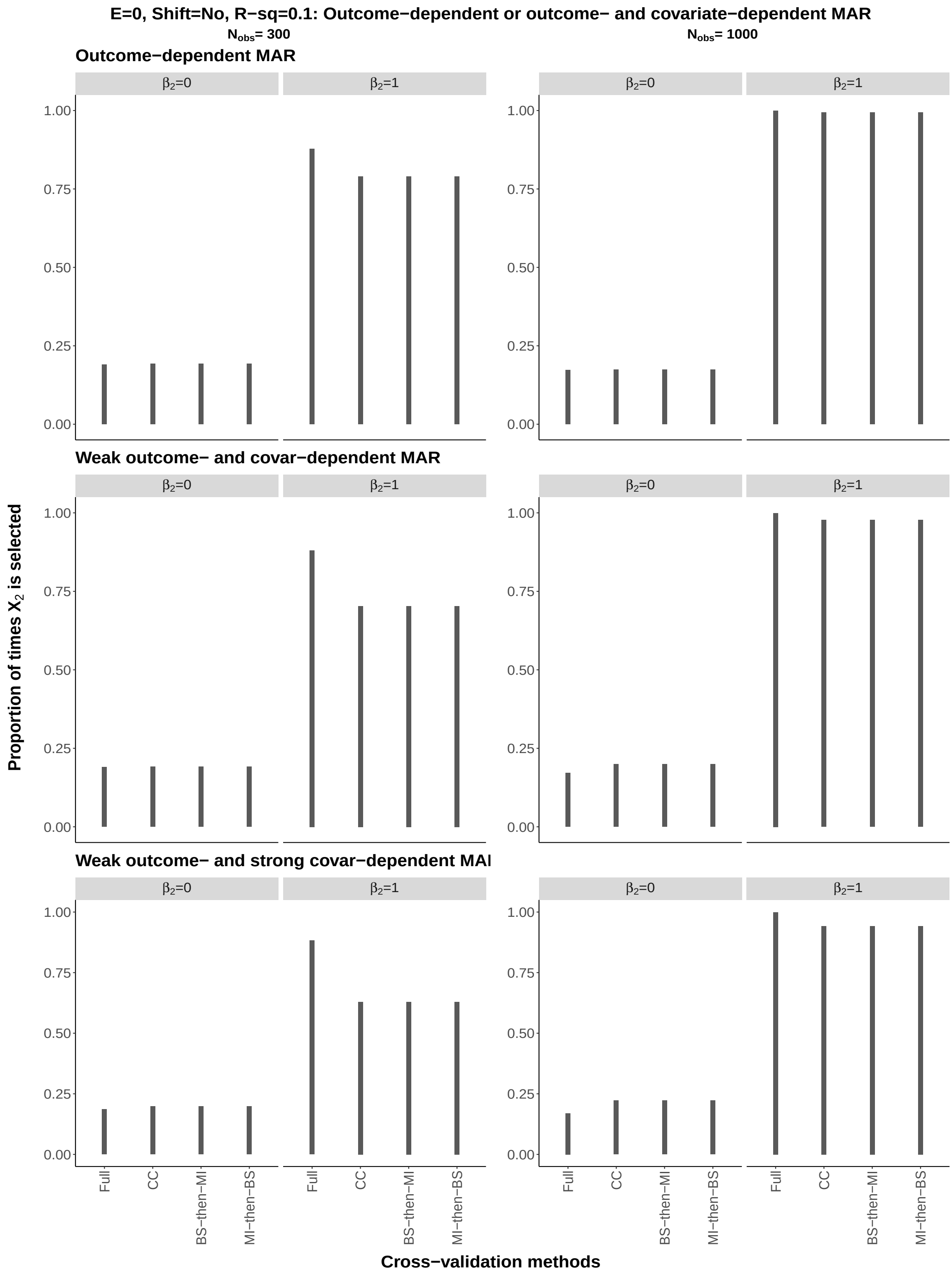

Figure S207: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

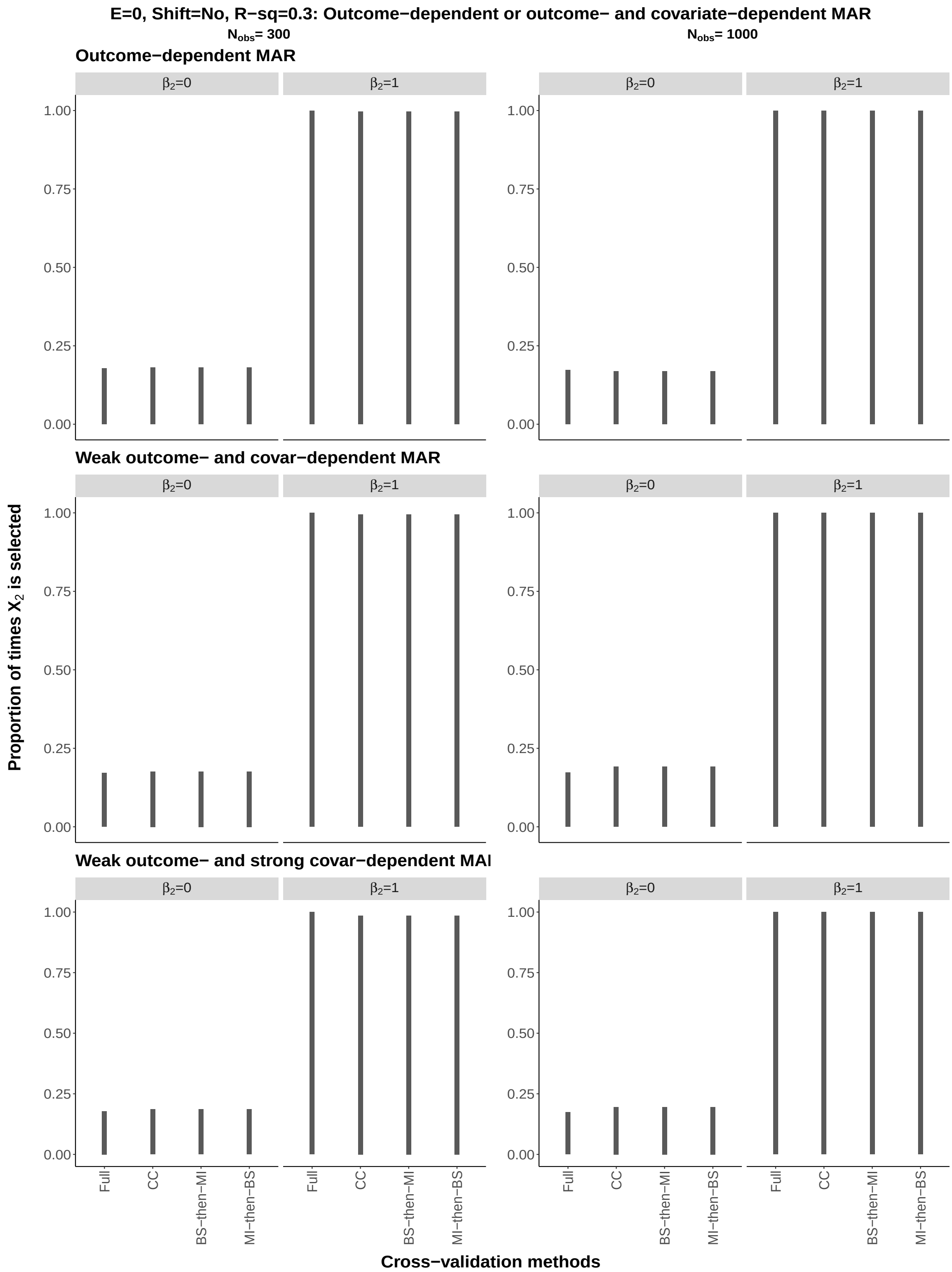

Figure S208: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

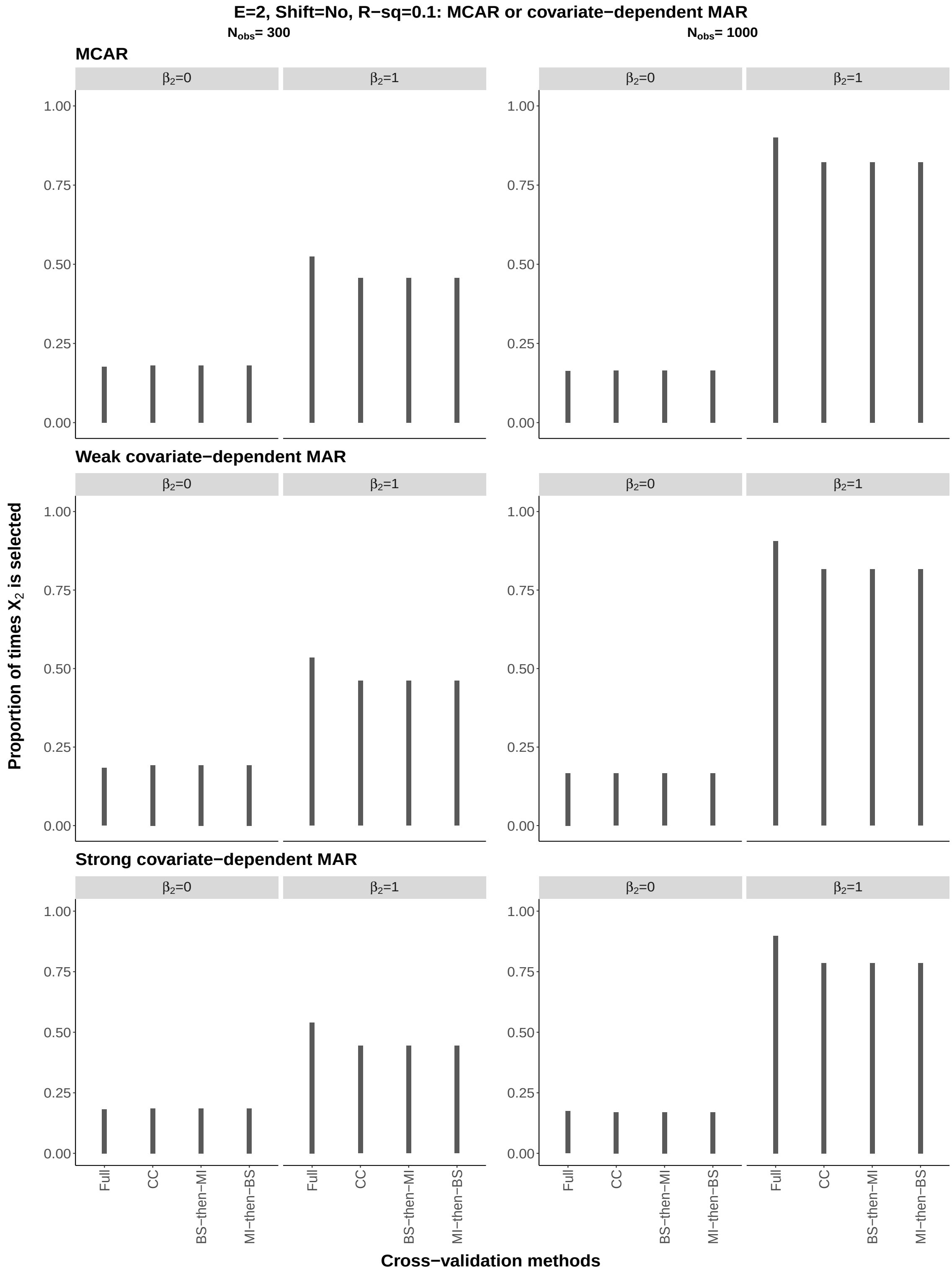

Figure S209: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

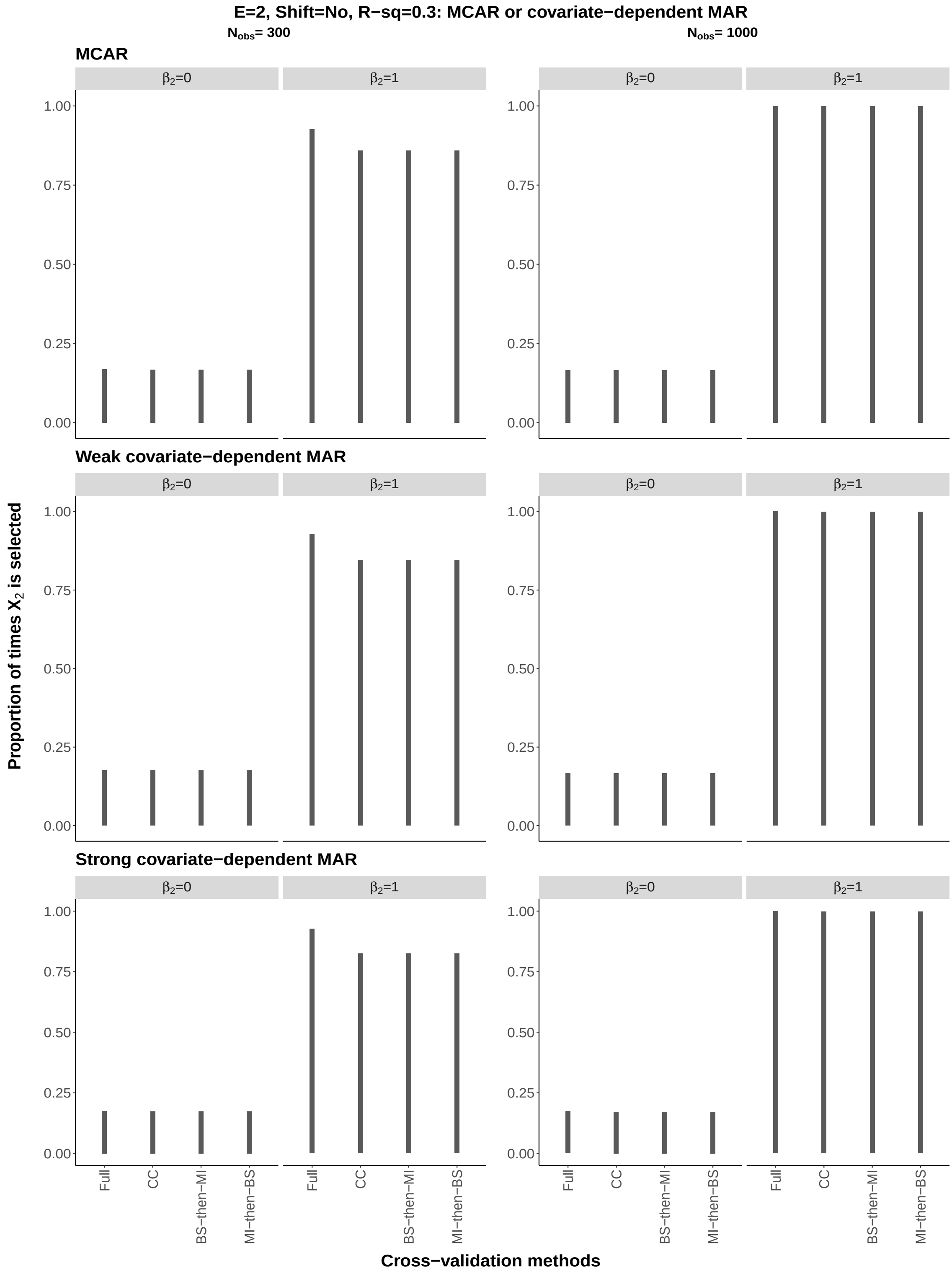

Figure S210: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

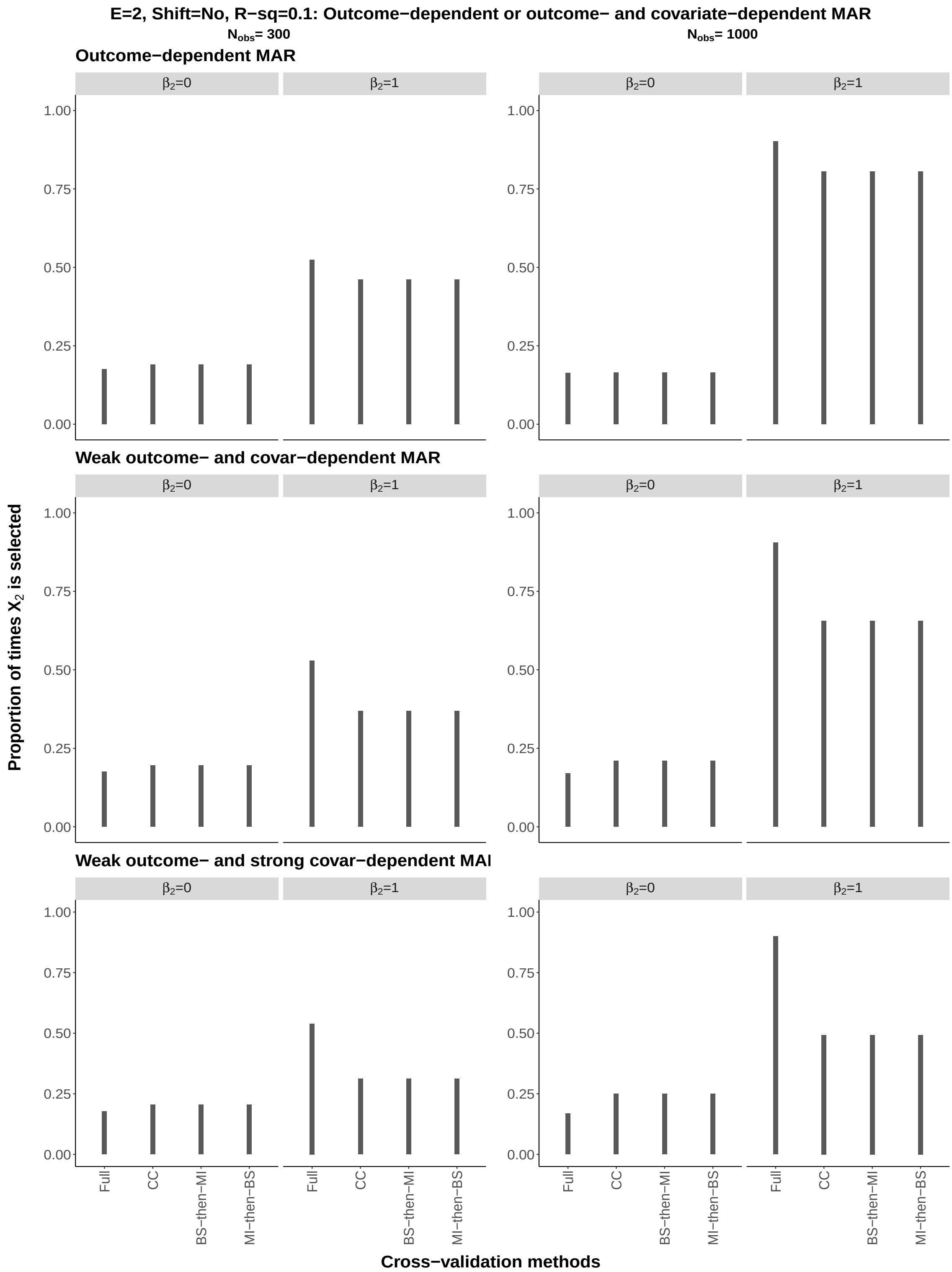

Figure S211: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

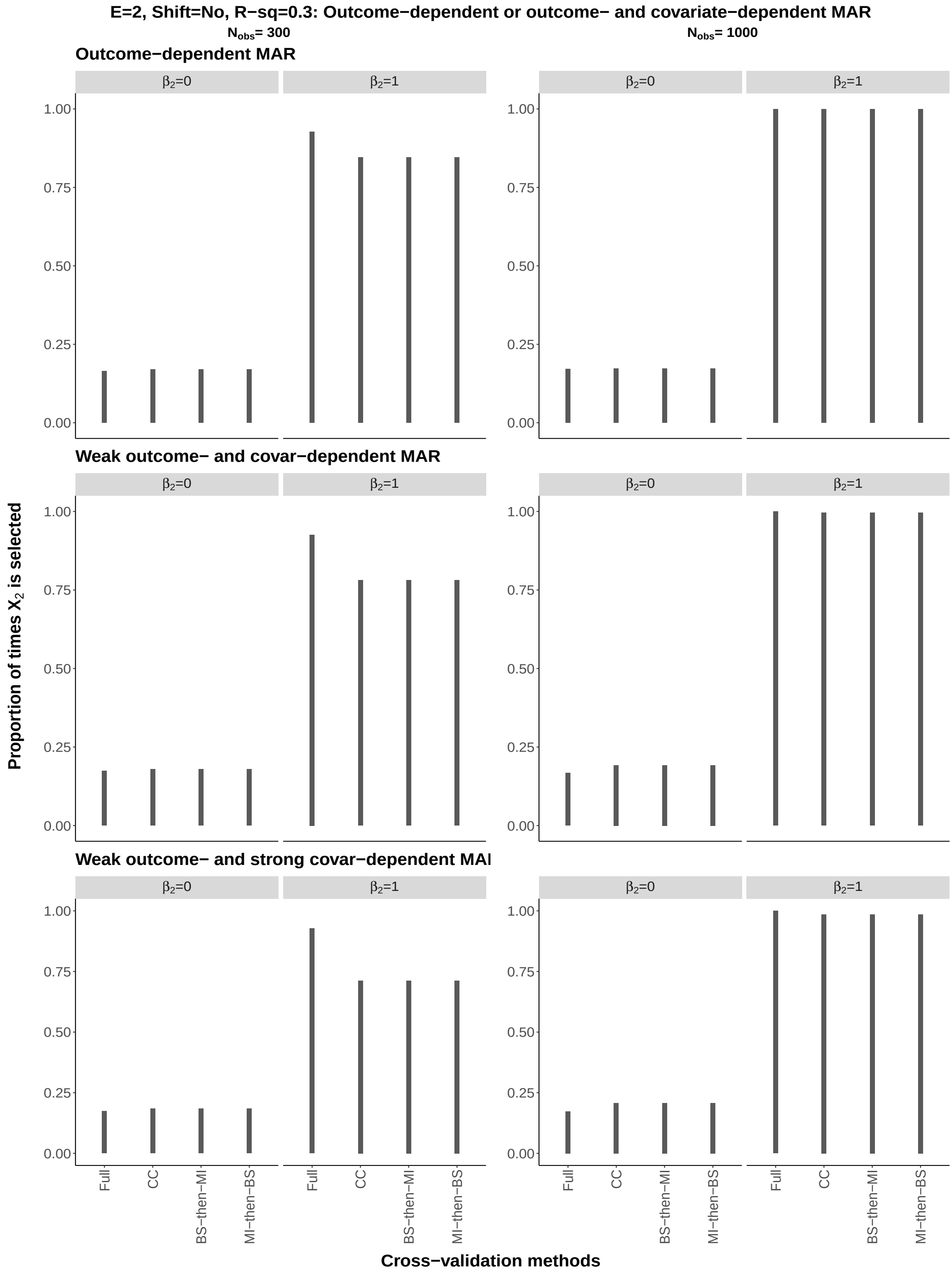

Figure S212: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

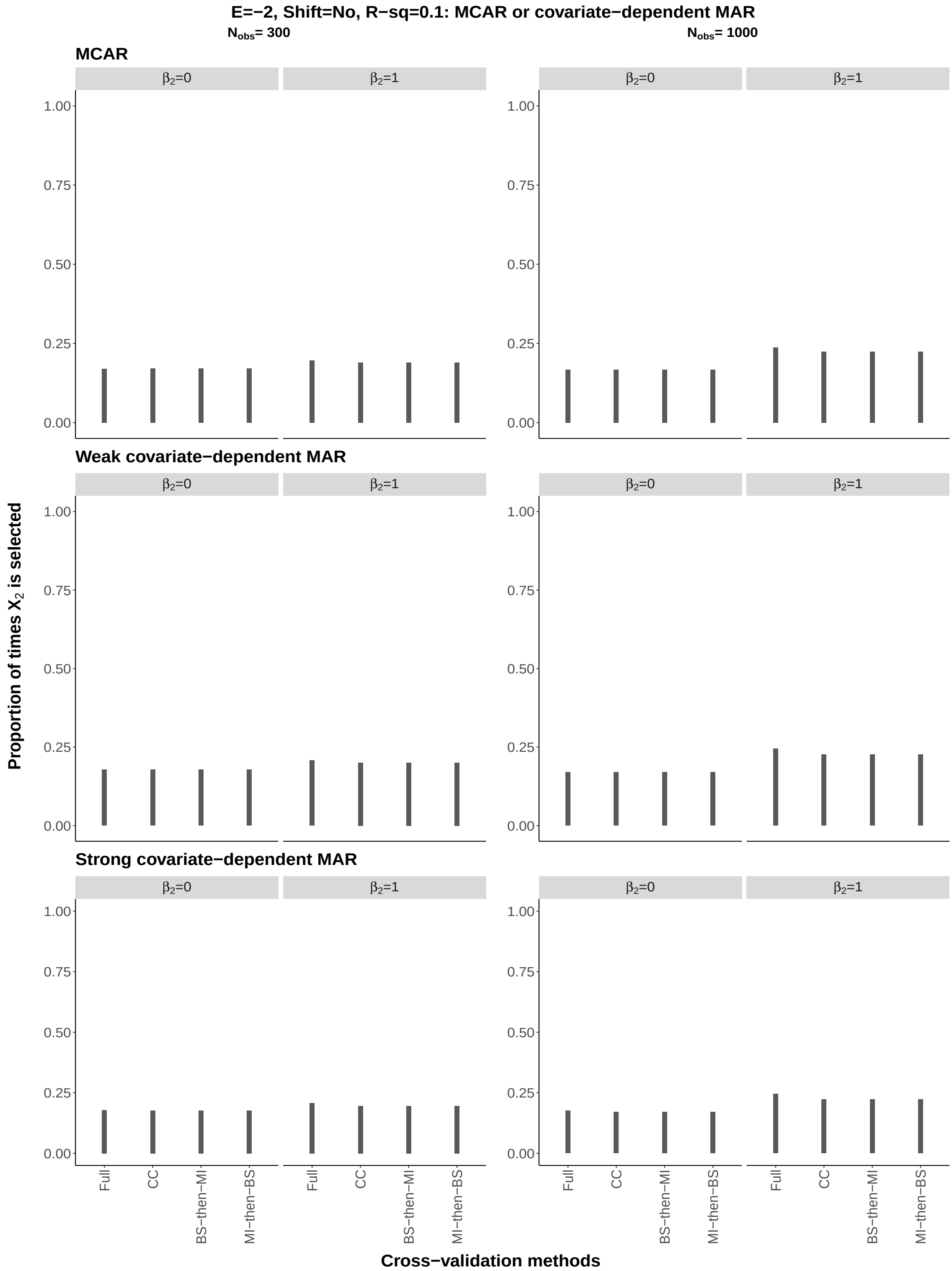

Figure S213: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

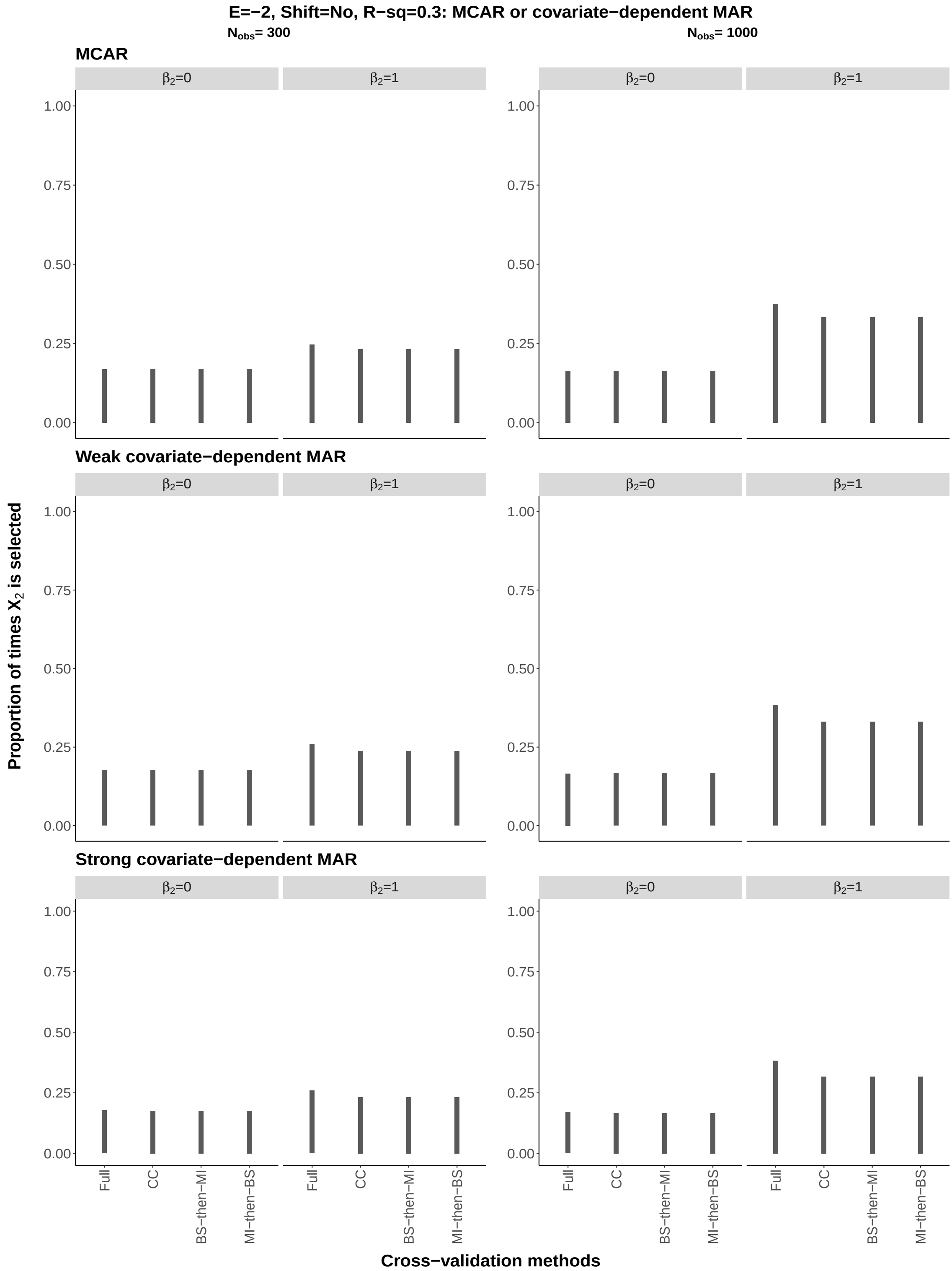

Figure S214: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

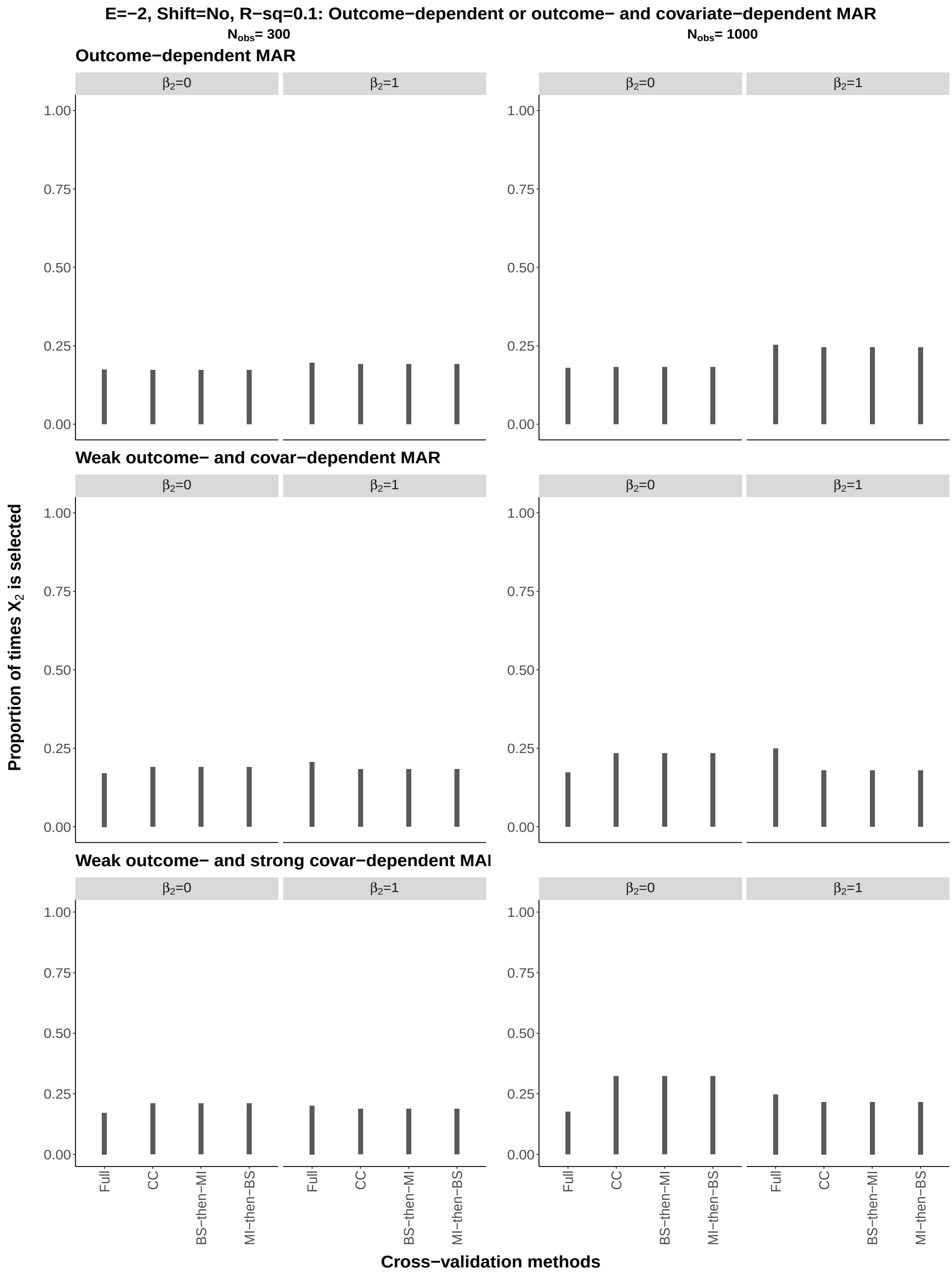

Figure S215: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

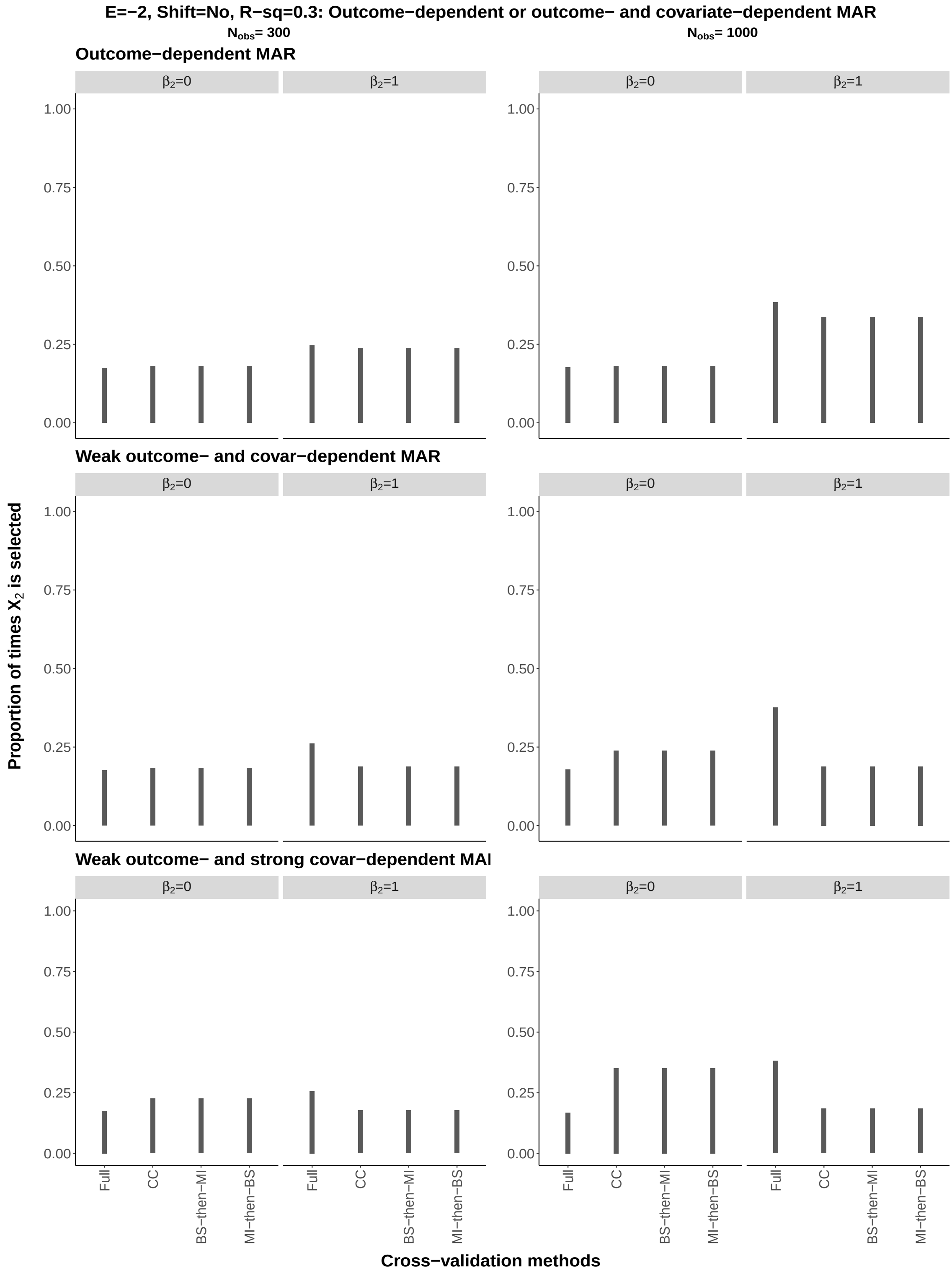

Figure S216: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.11 Covariate selection of X_2 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

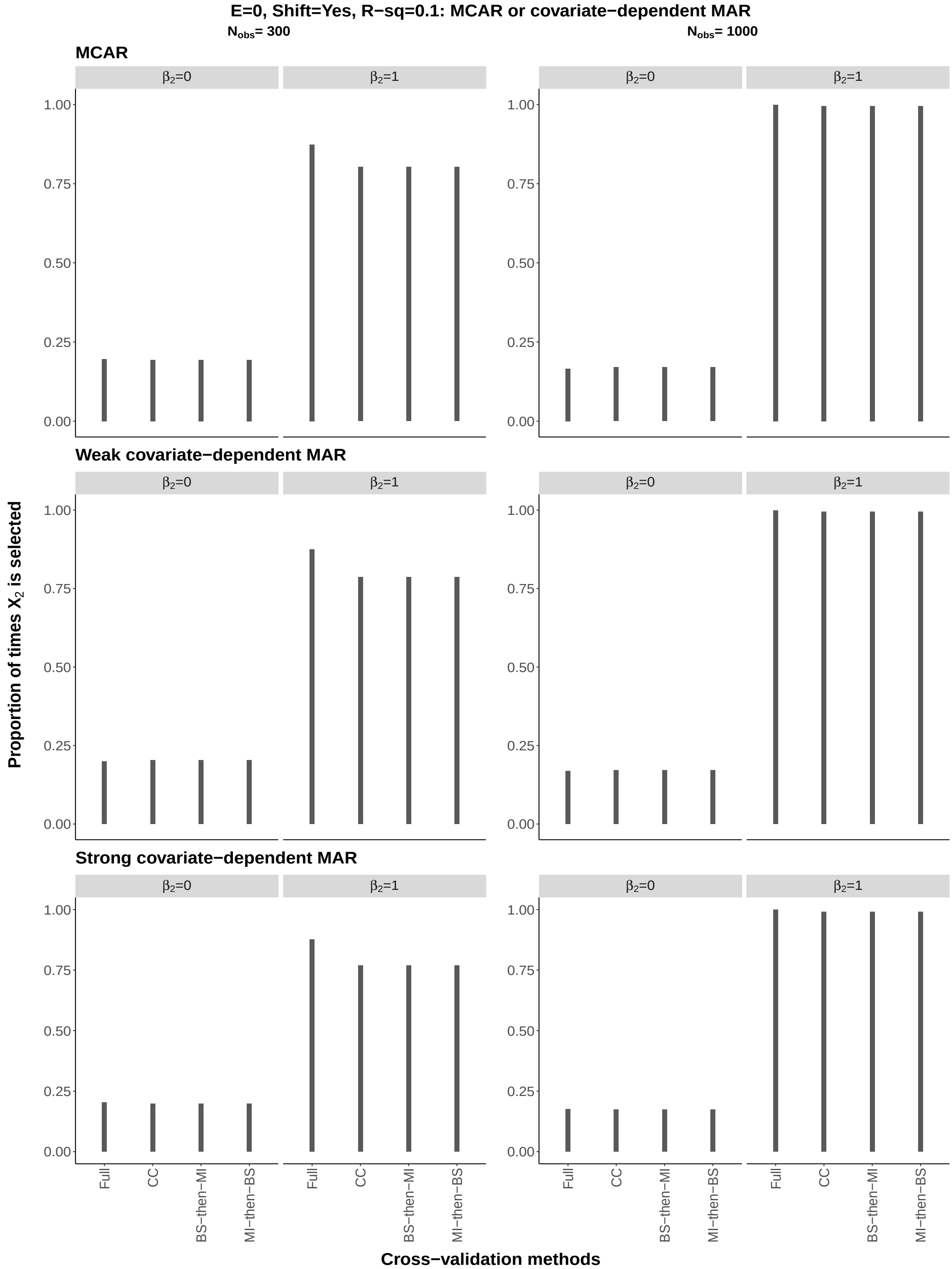

Figure S217: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

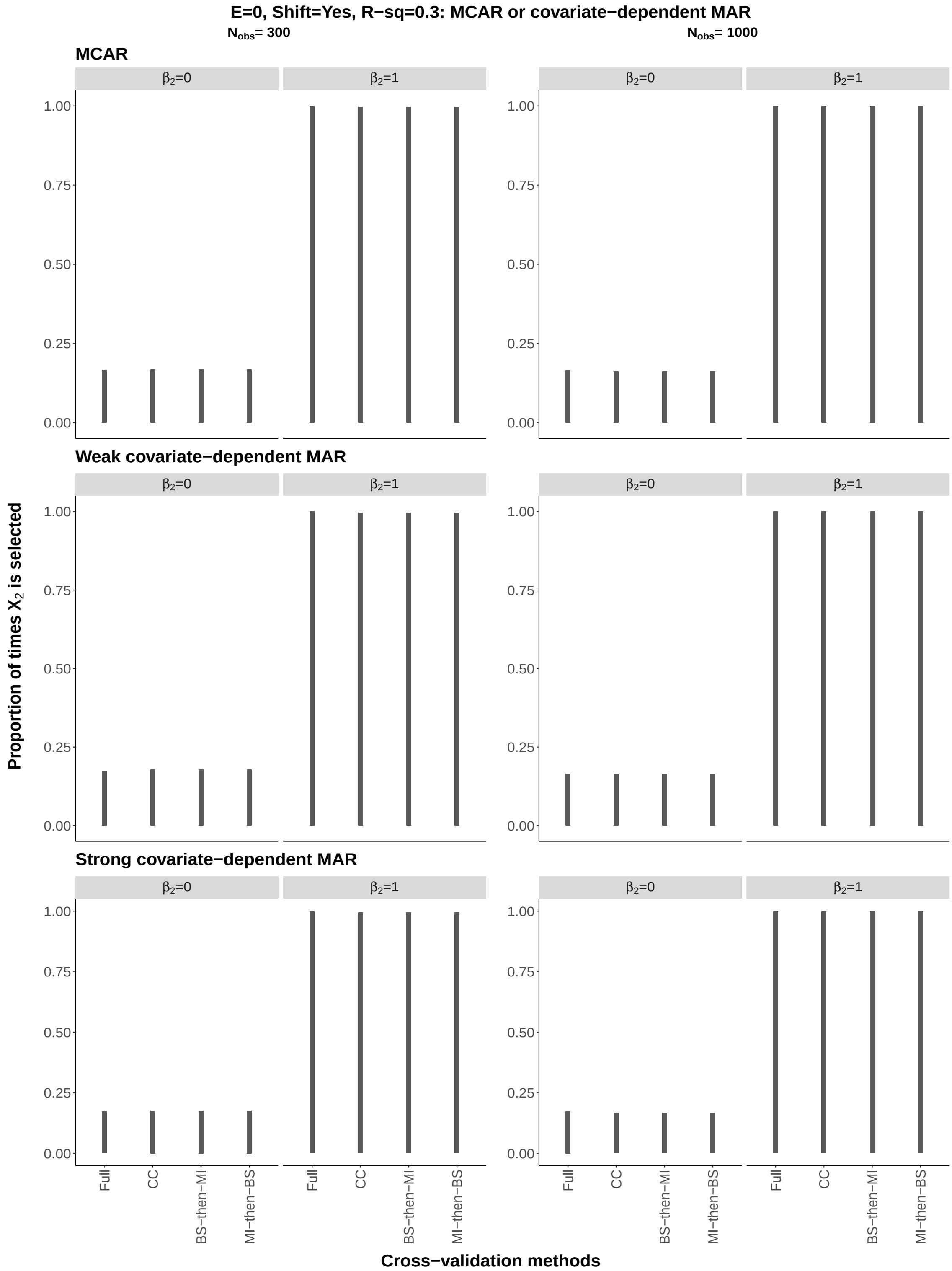

Figure S218: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

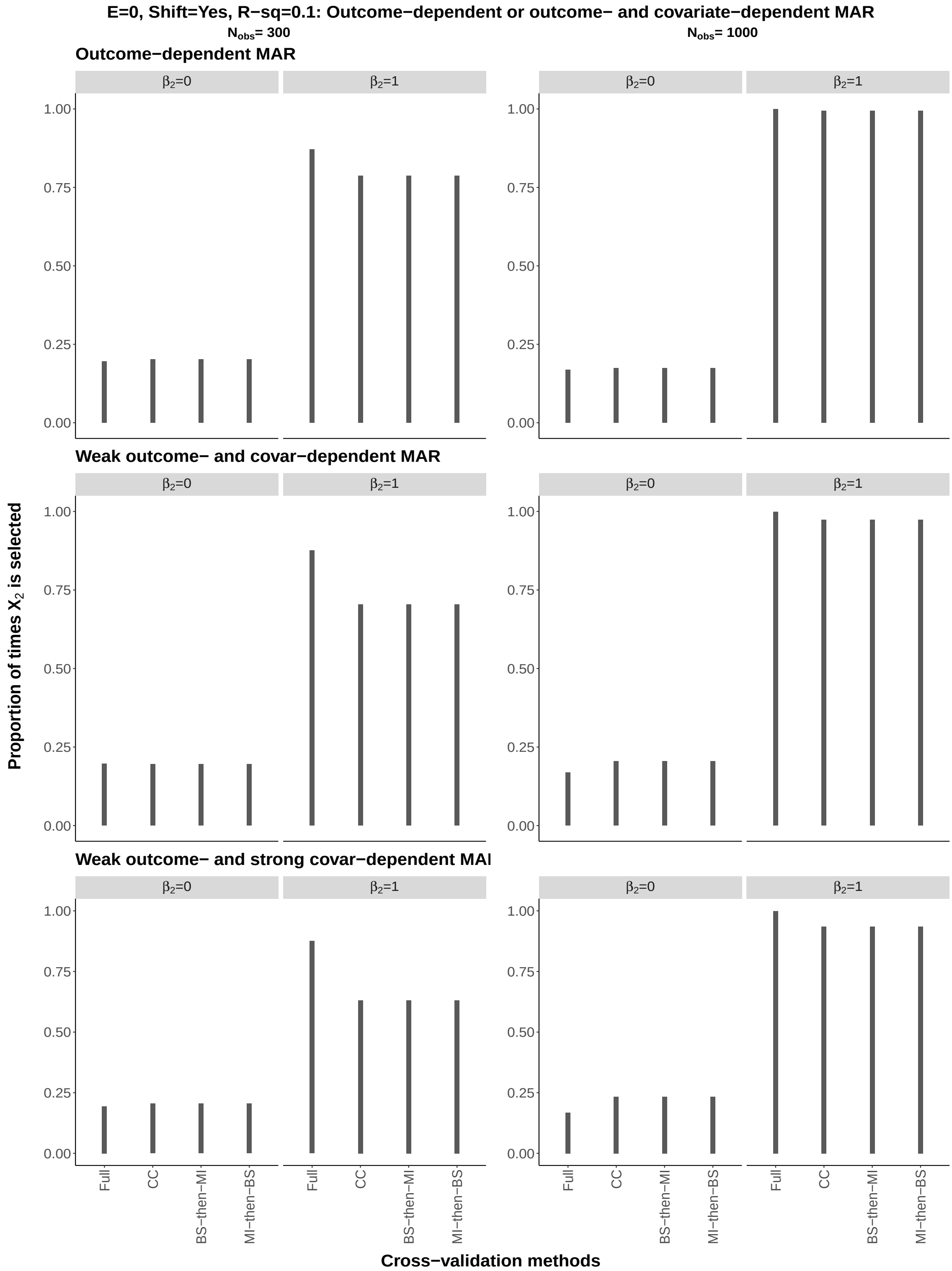

Figure S219: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

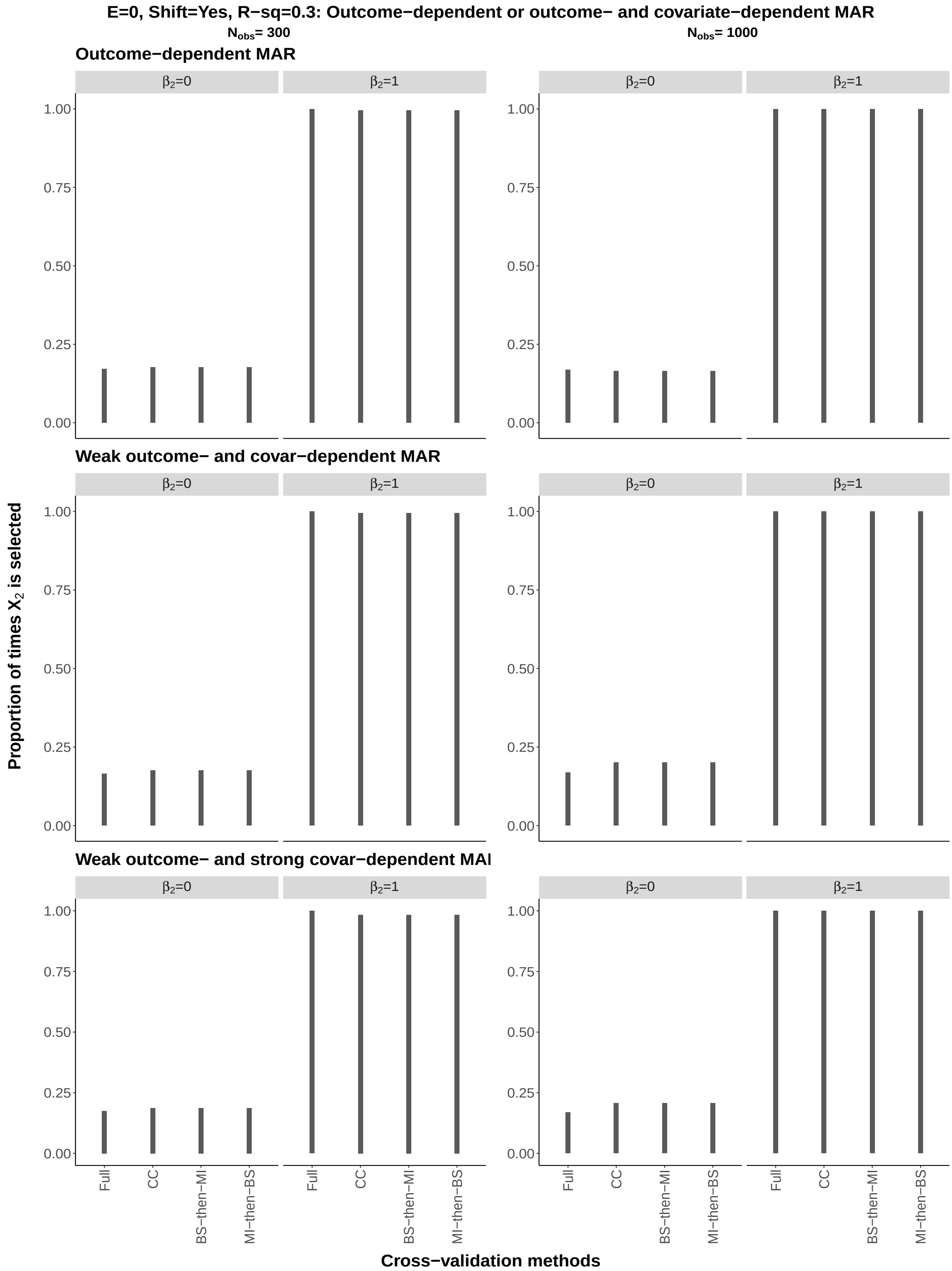

Figure S220: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

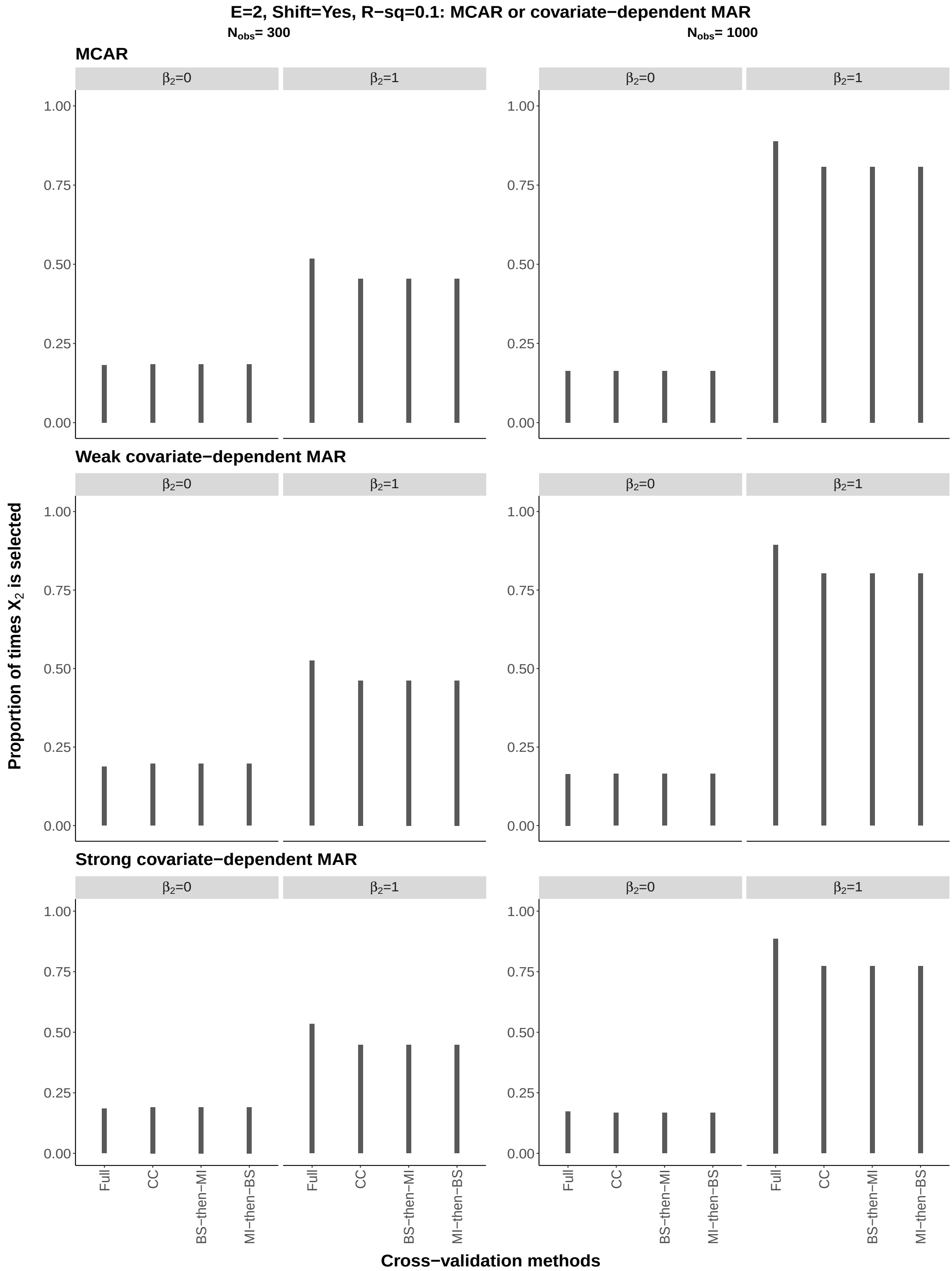

Figure S221: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

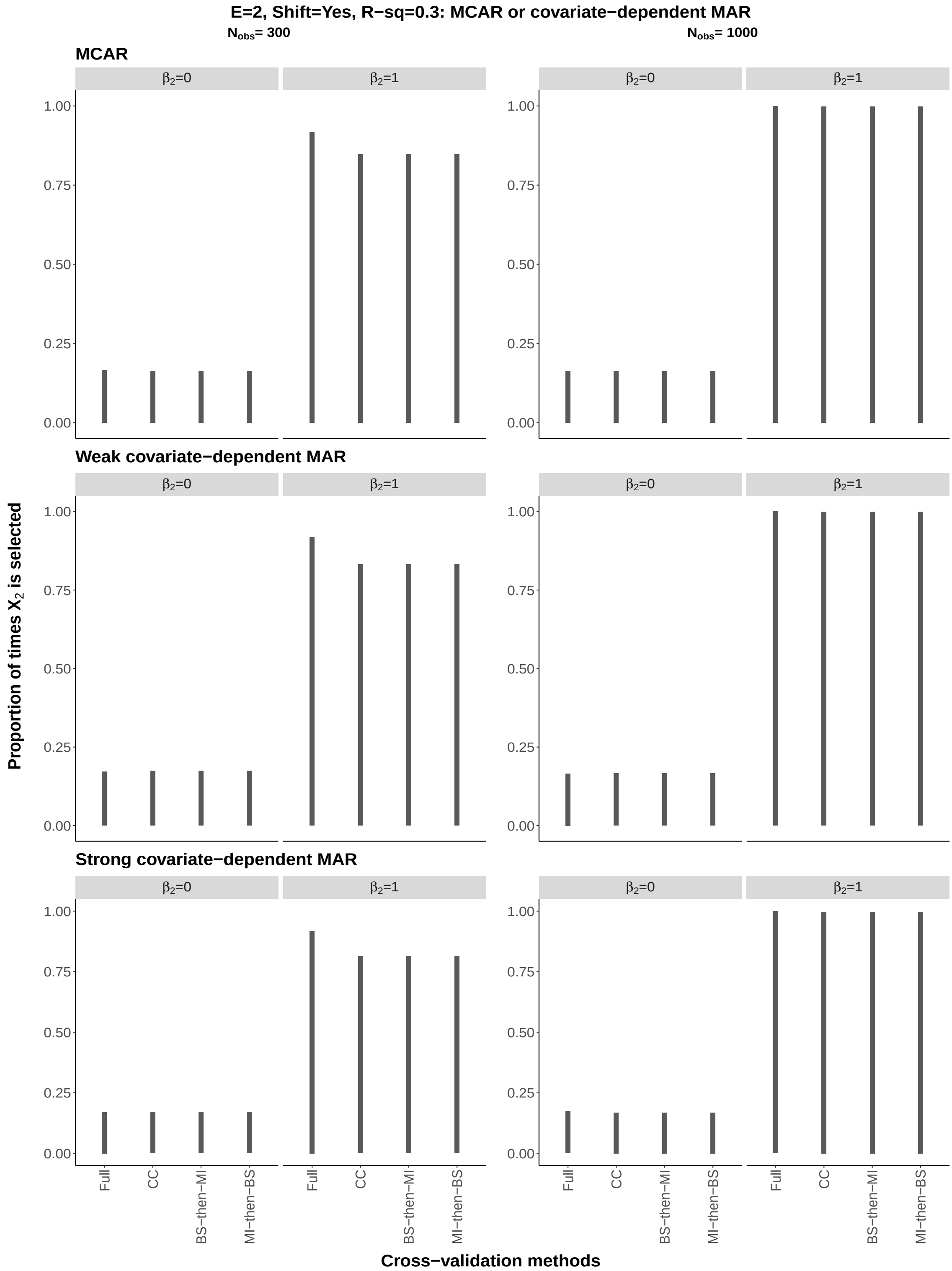

Figure S222: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

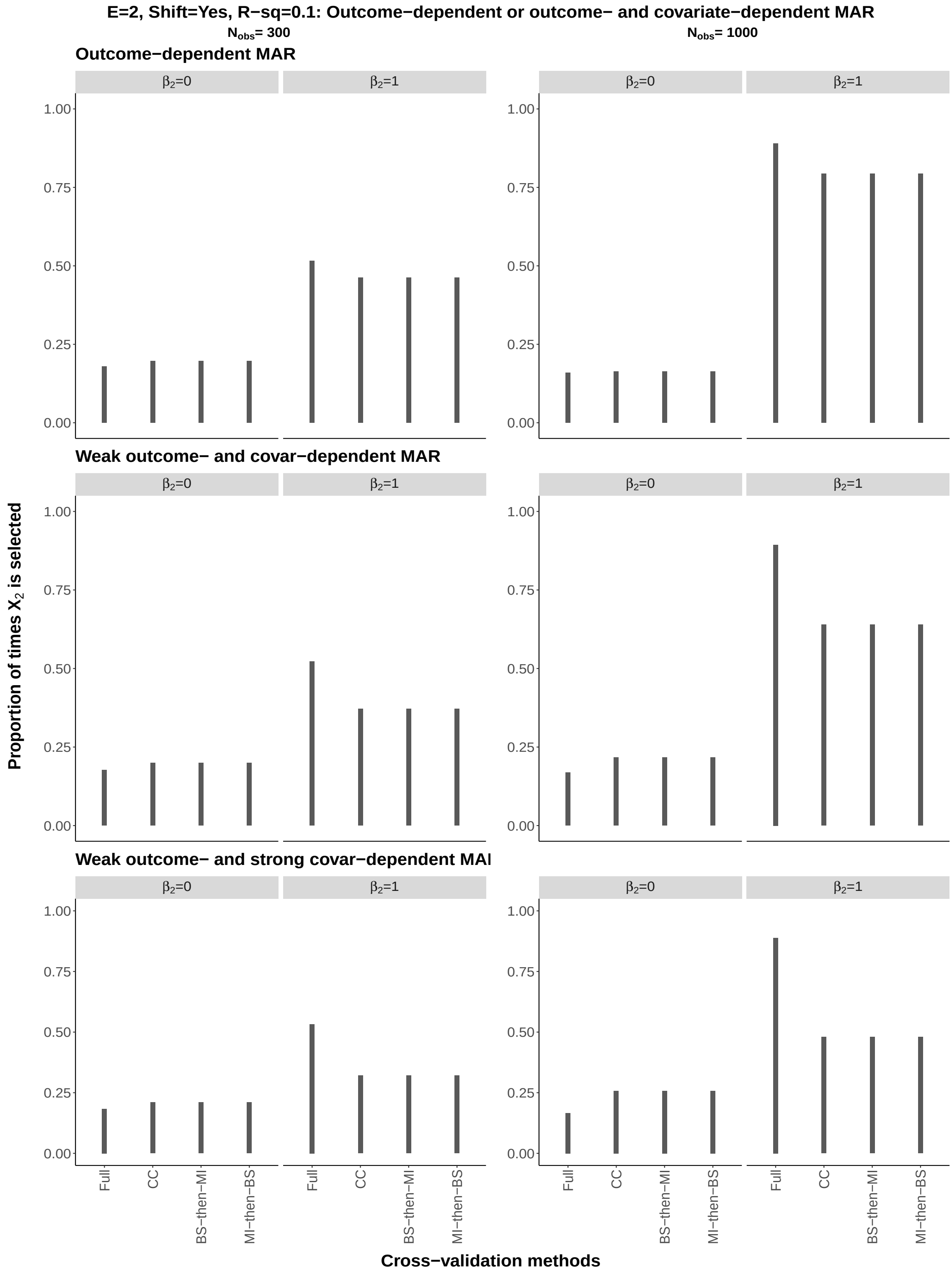

Figure S223: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

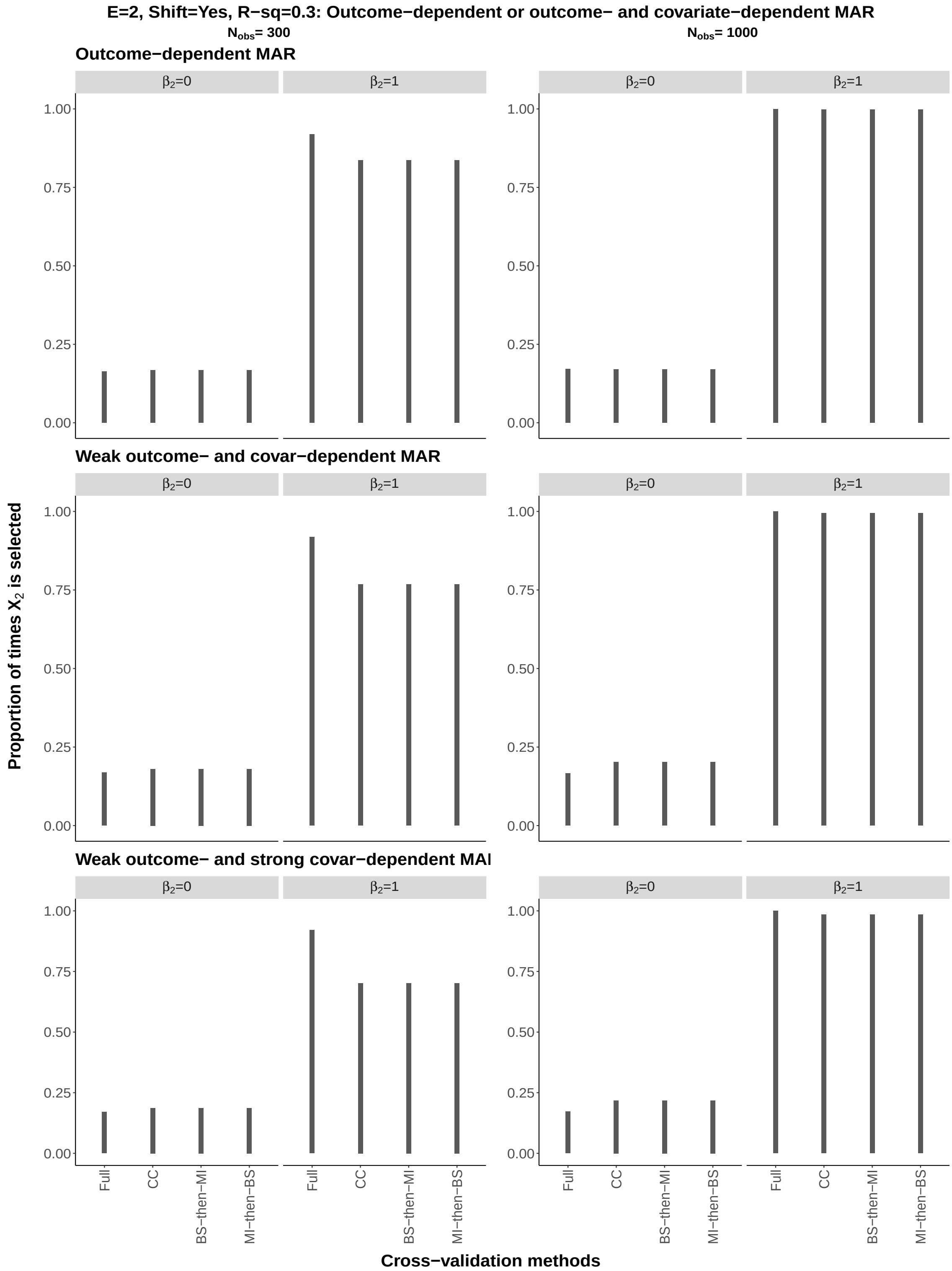

Figure S224: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

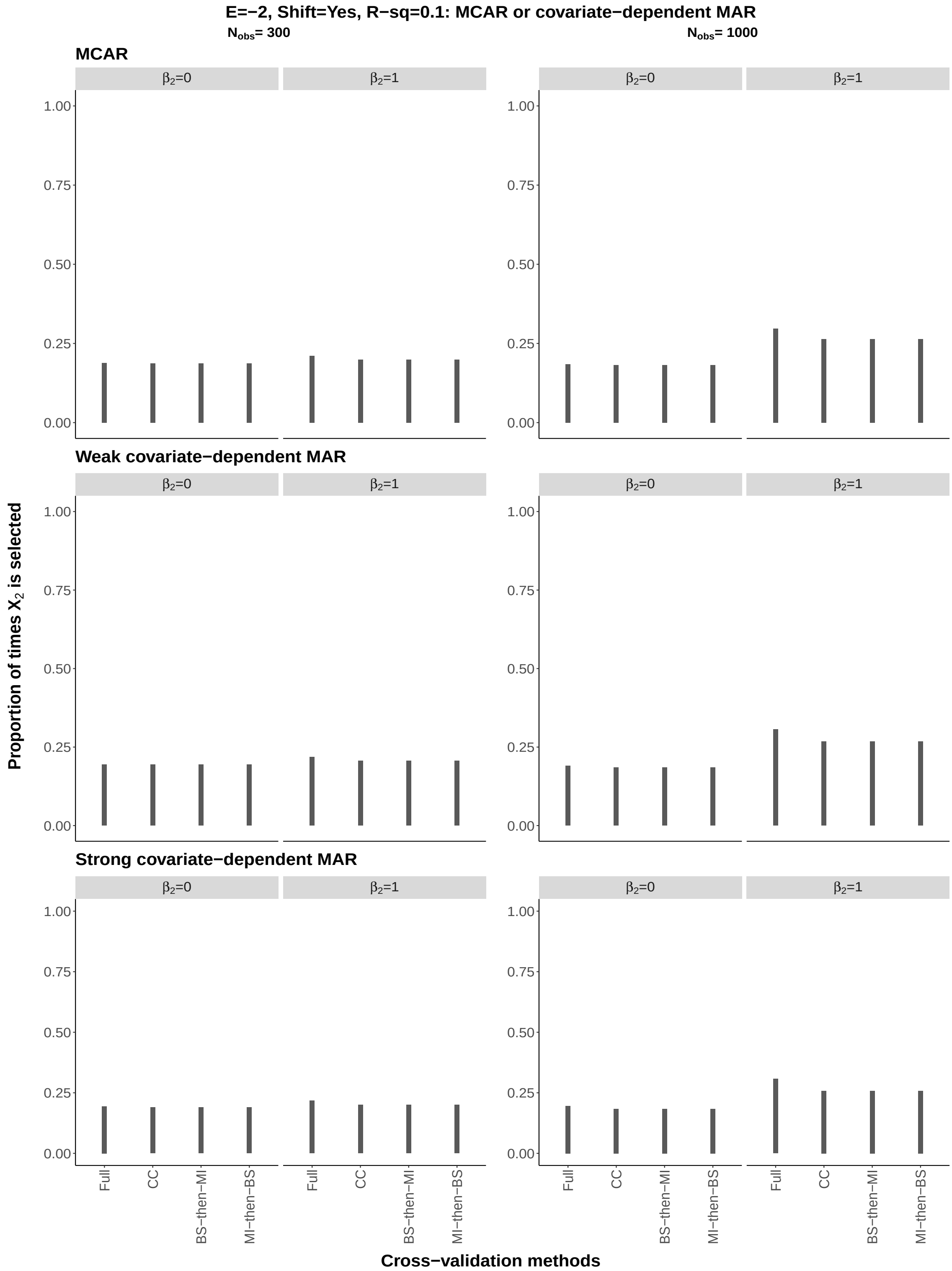

Figure S225: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

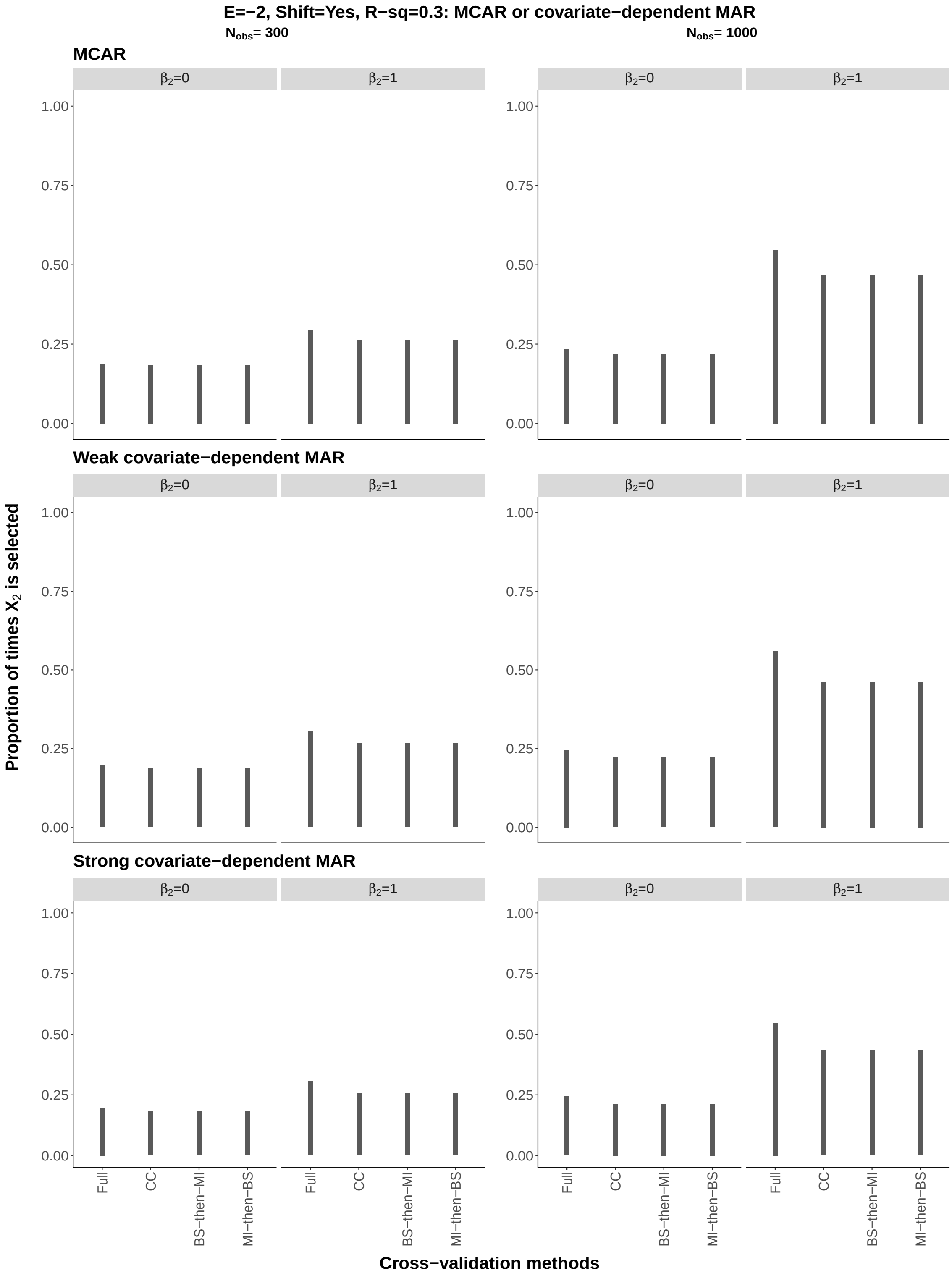

Figure S226: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

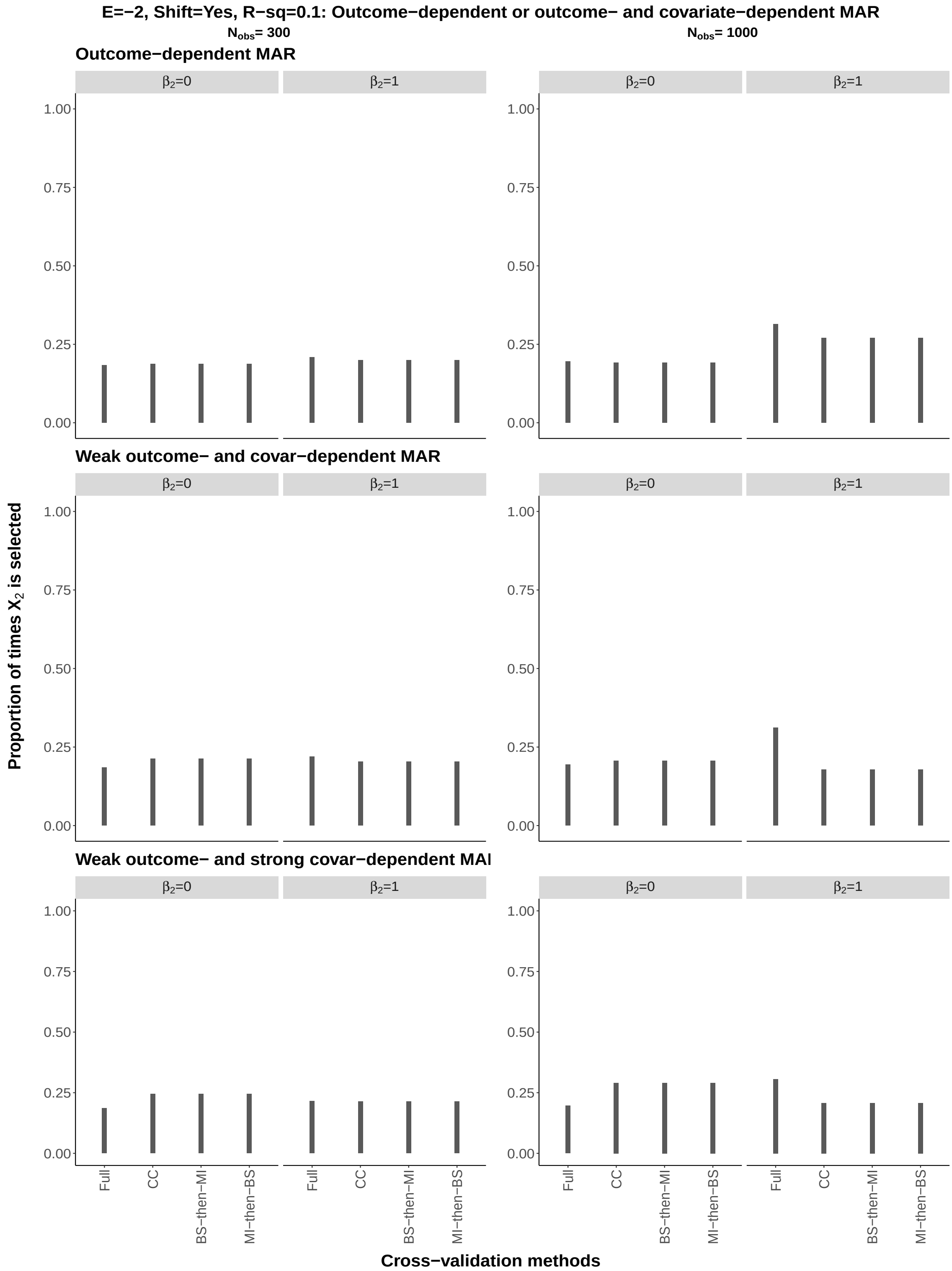

Figure S227: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

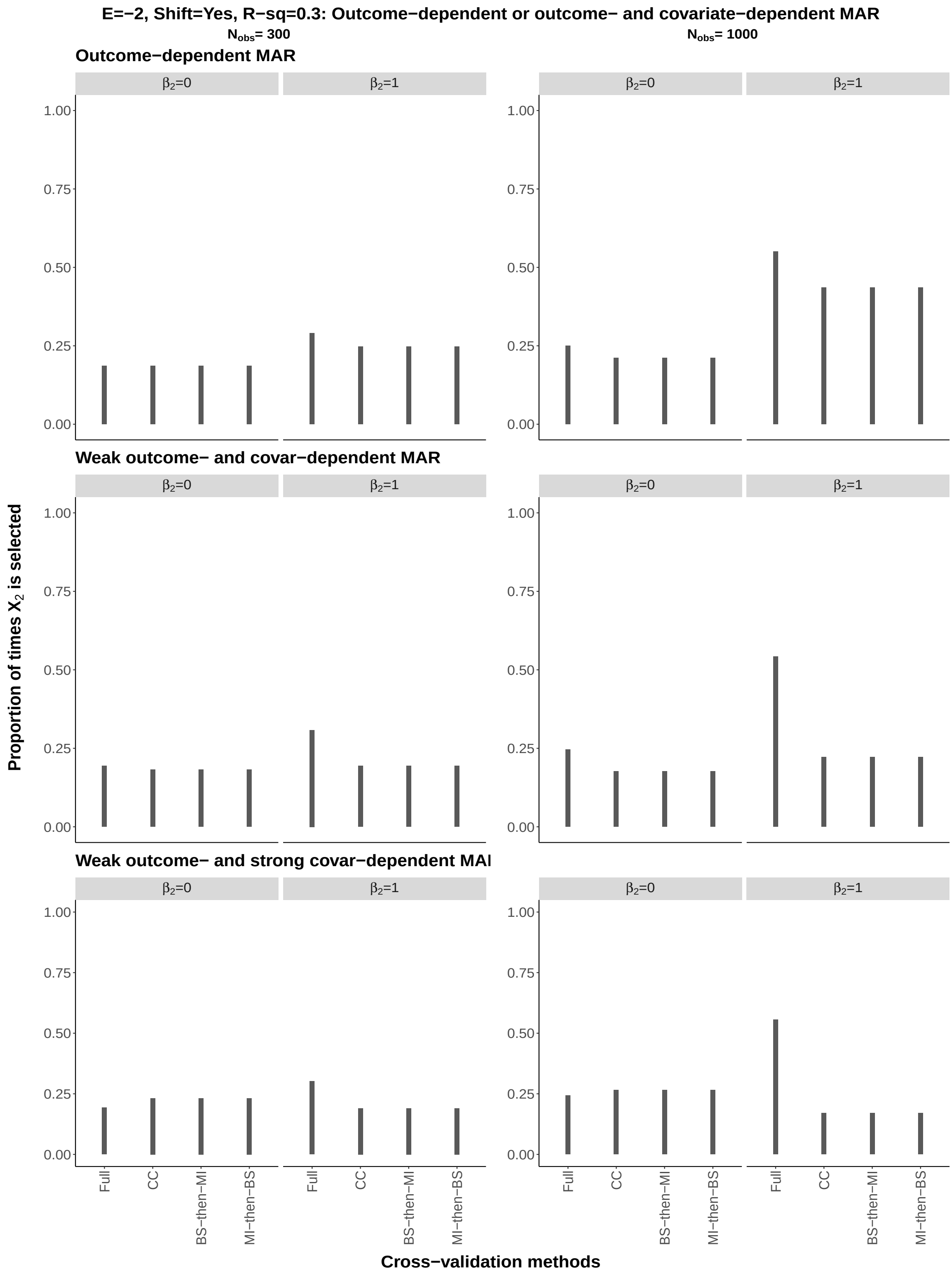

Figure S228: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.12 Covariate selection of X_2 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

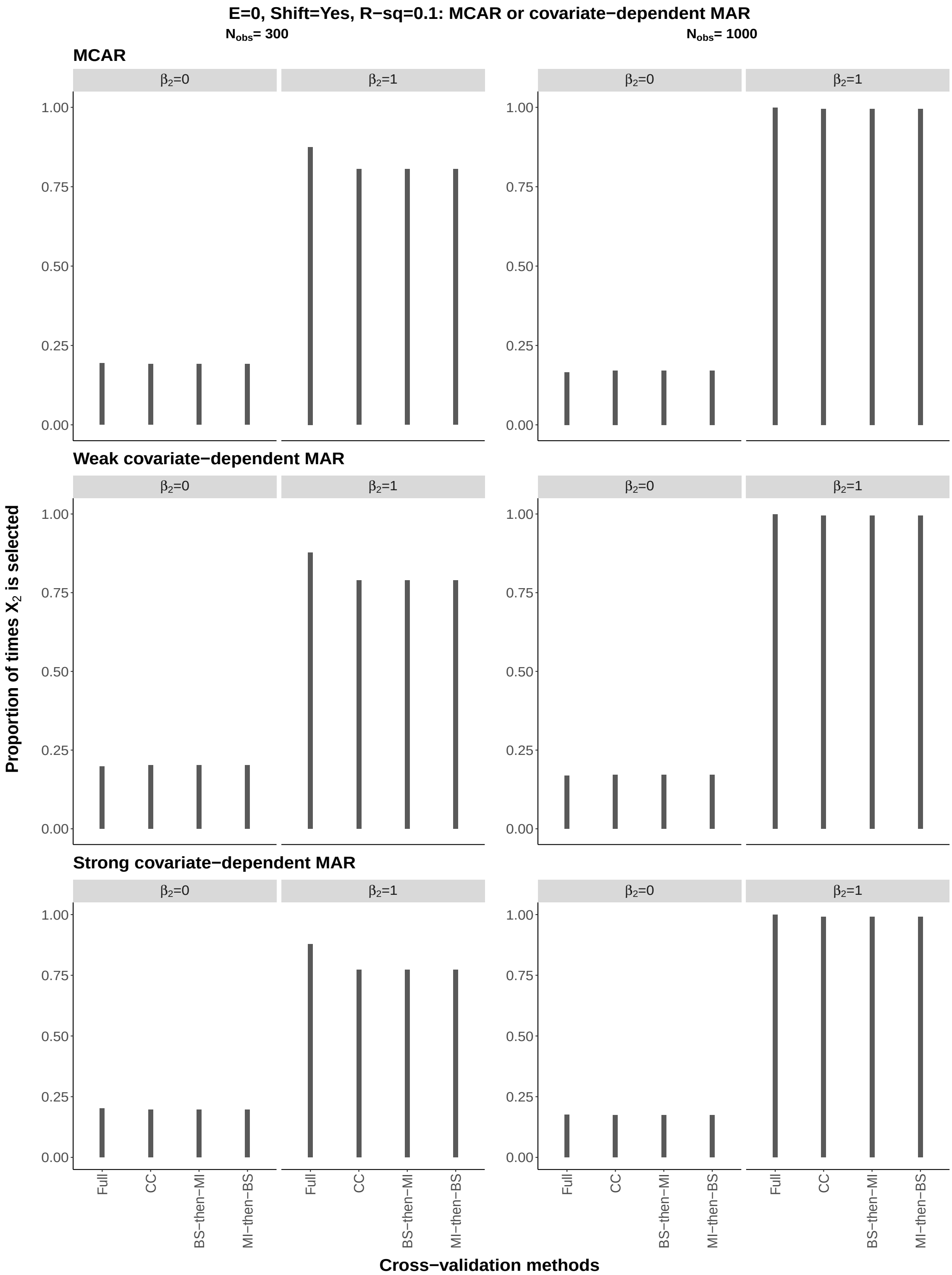

Figure S229: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

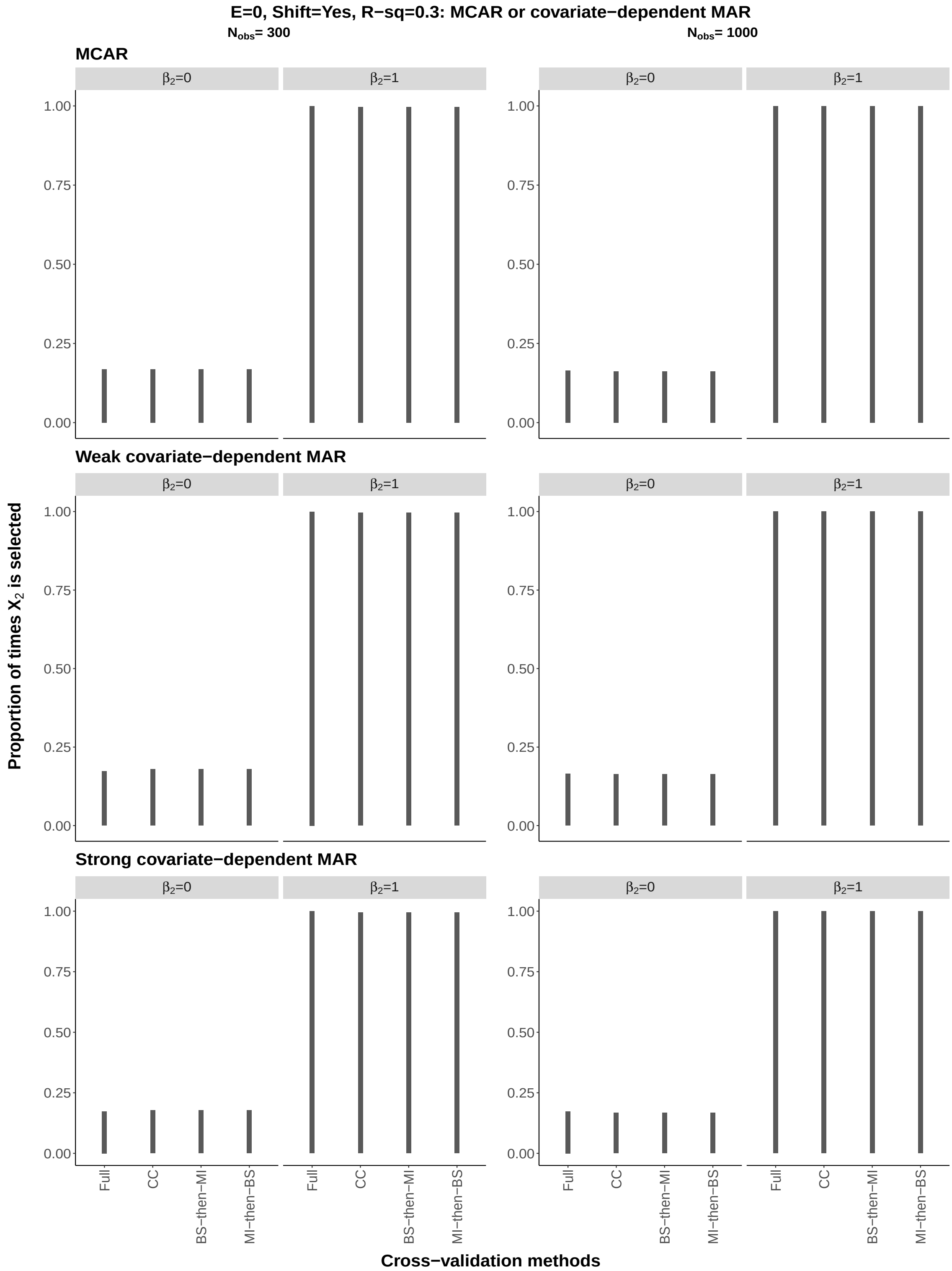

Figure S230: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

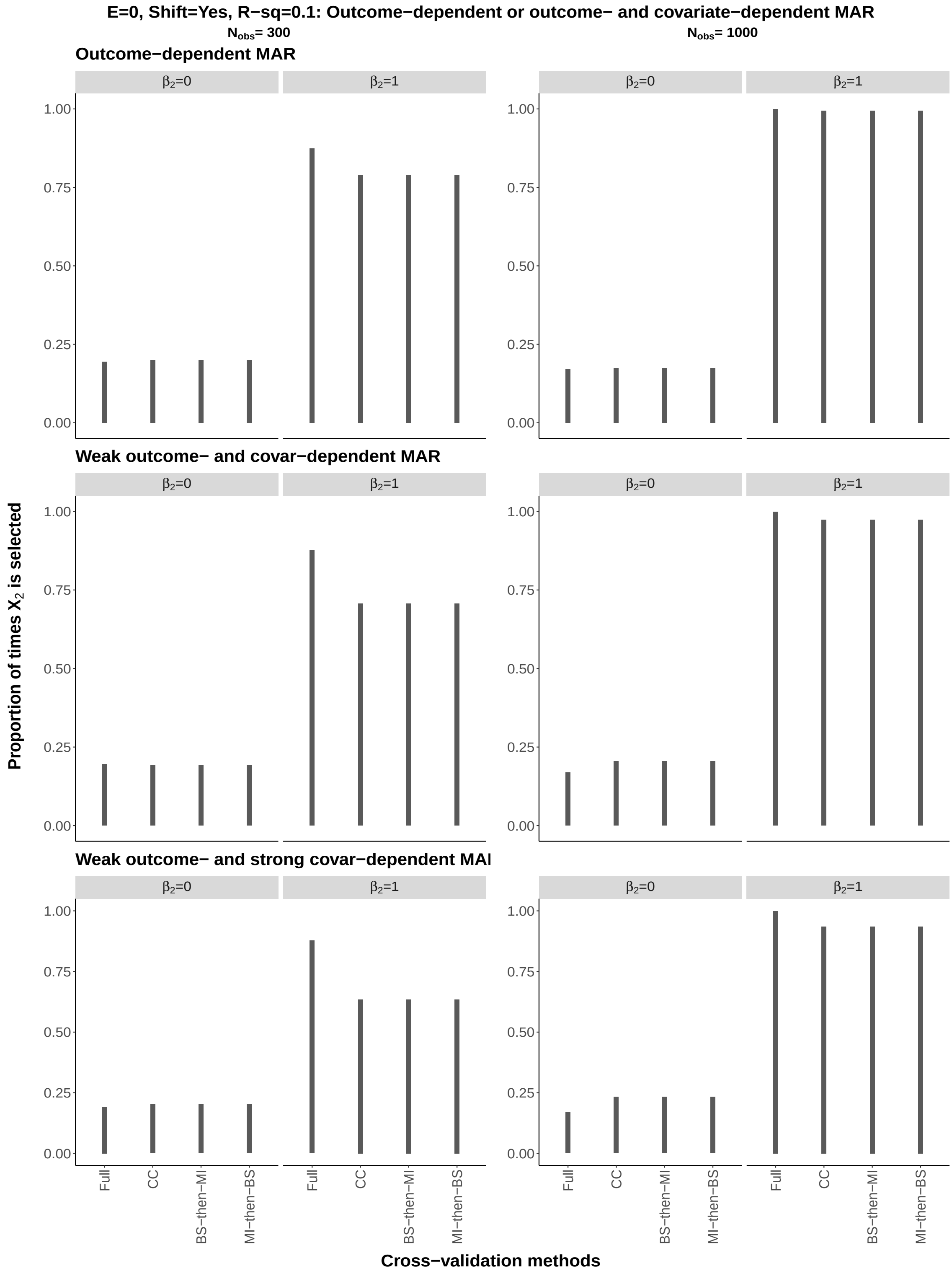

Figure S231: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

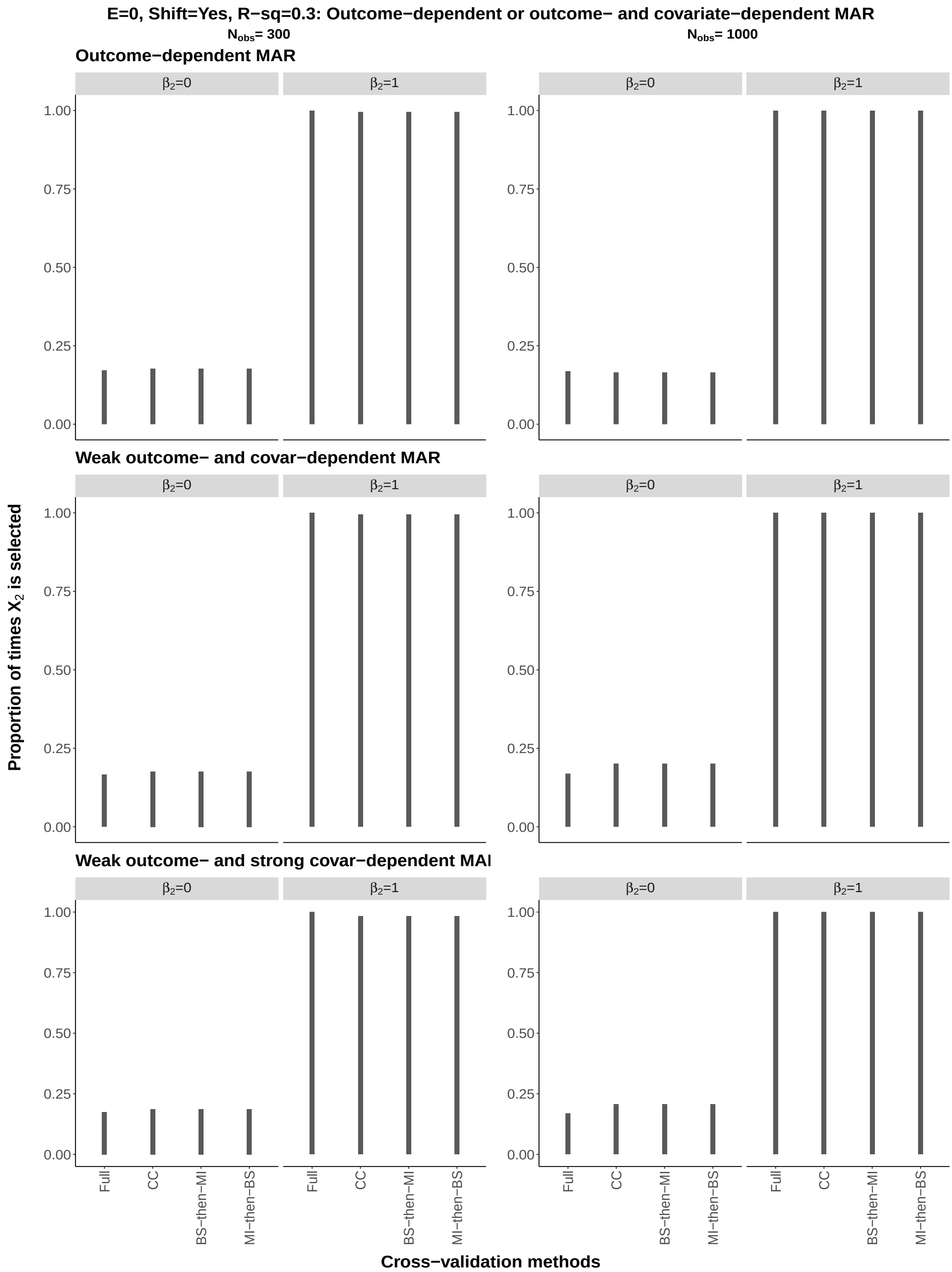

Figure S232: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

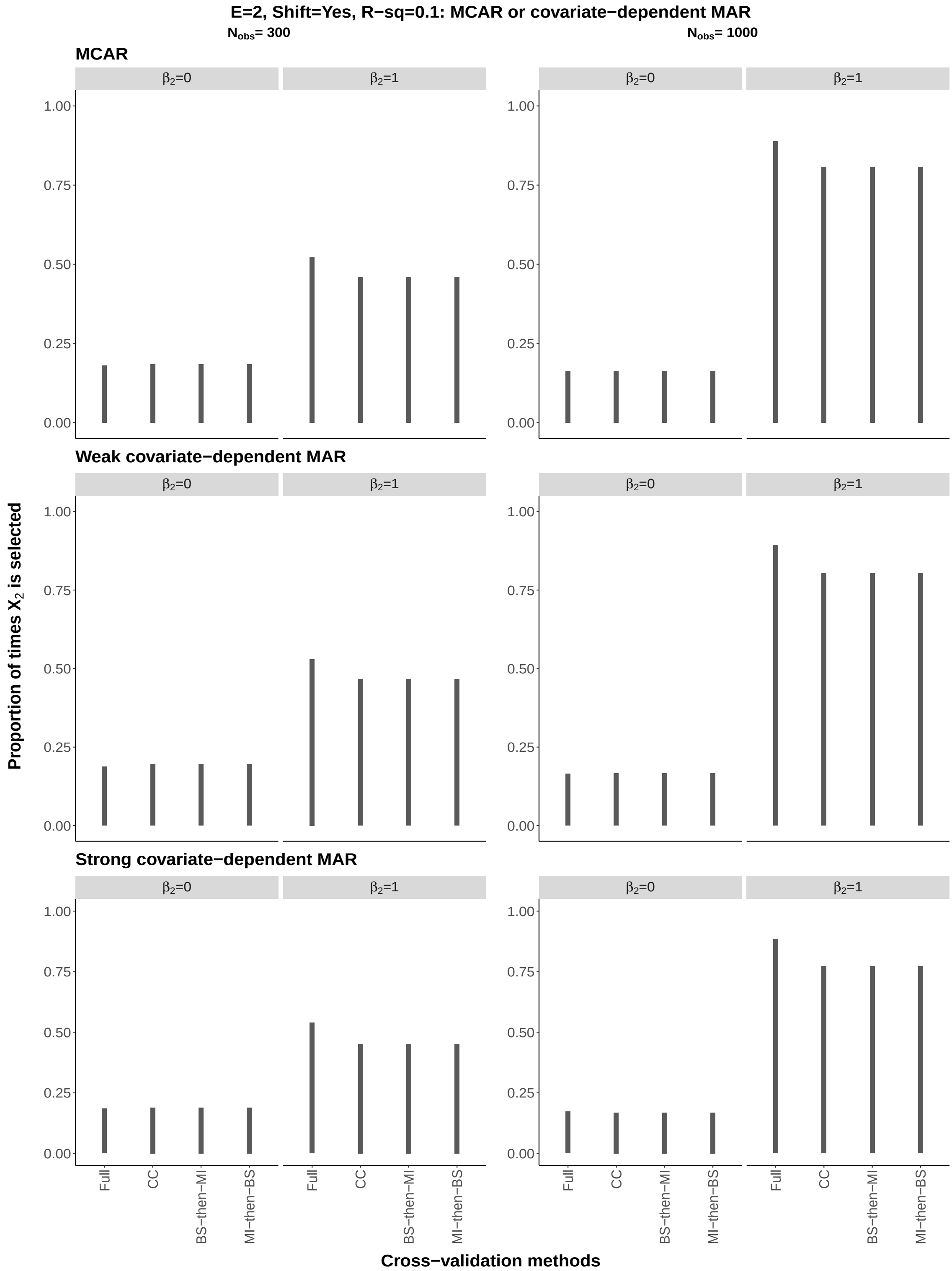

Figure S233: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

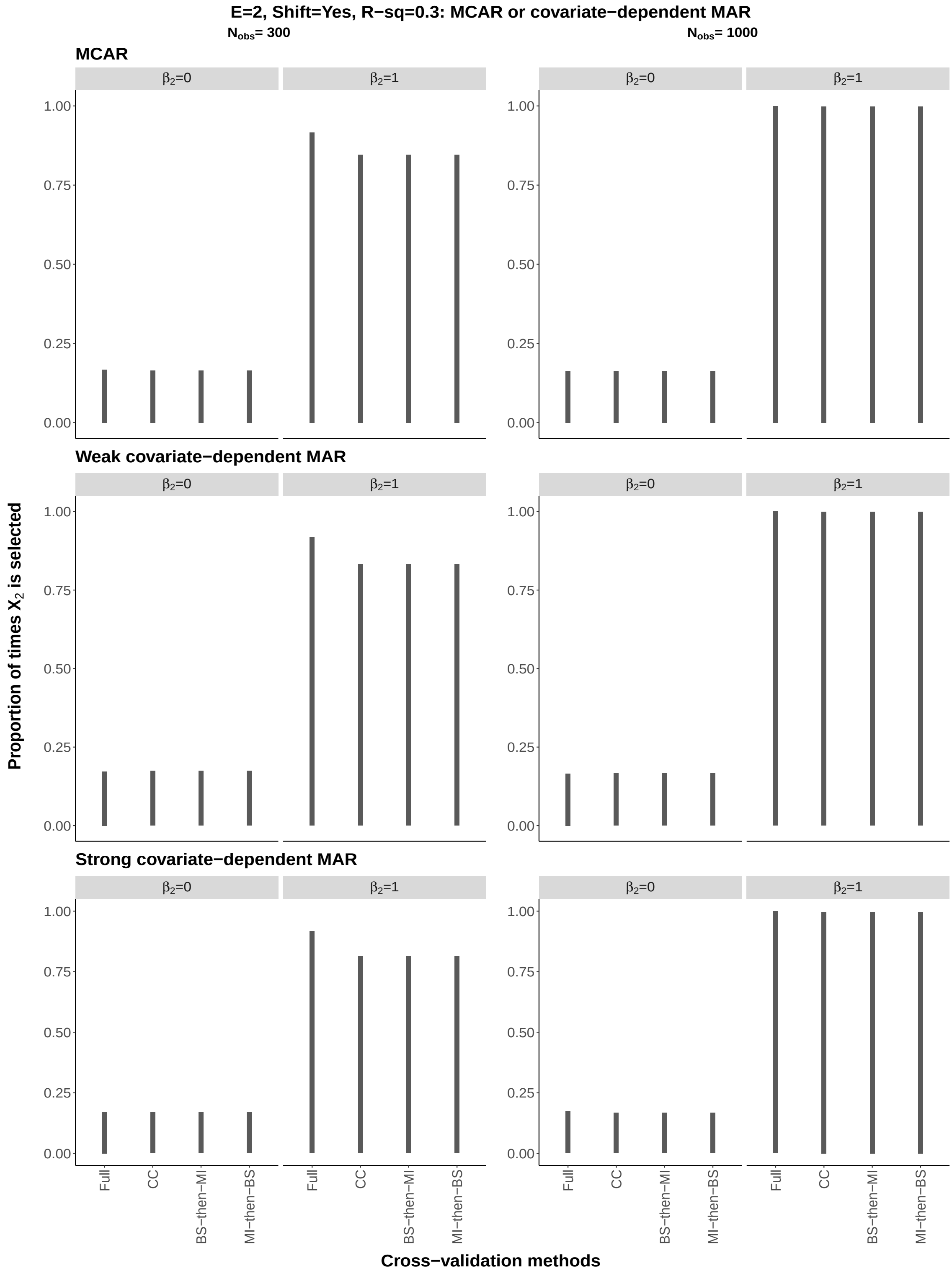

Figure S234: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

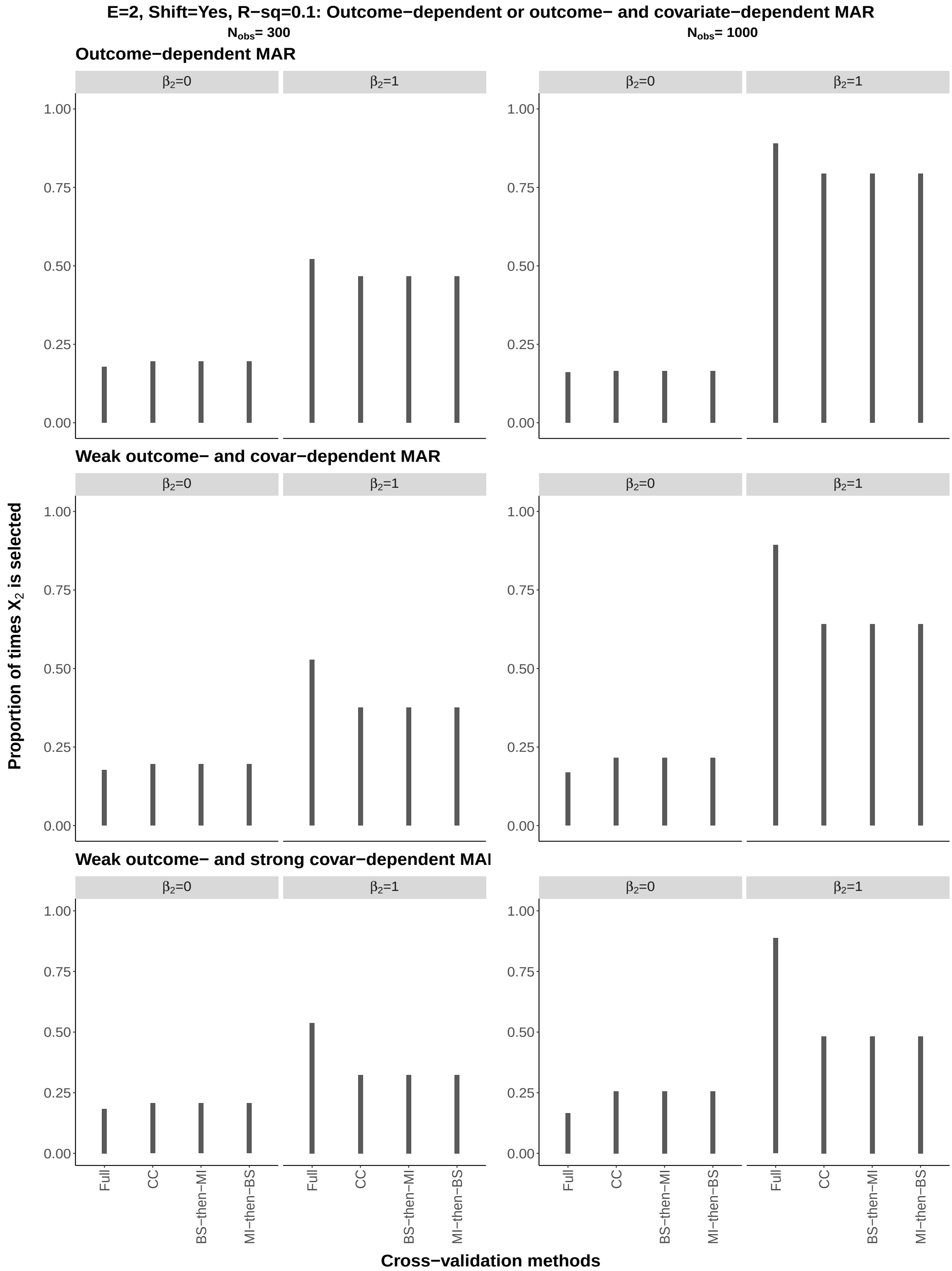

Figure S235: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

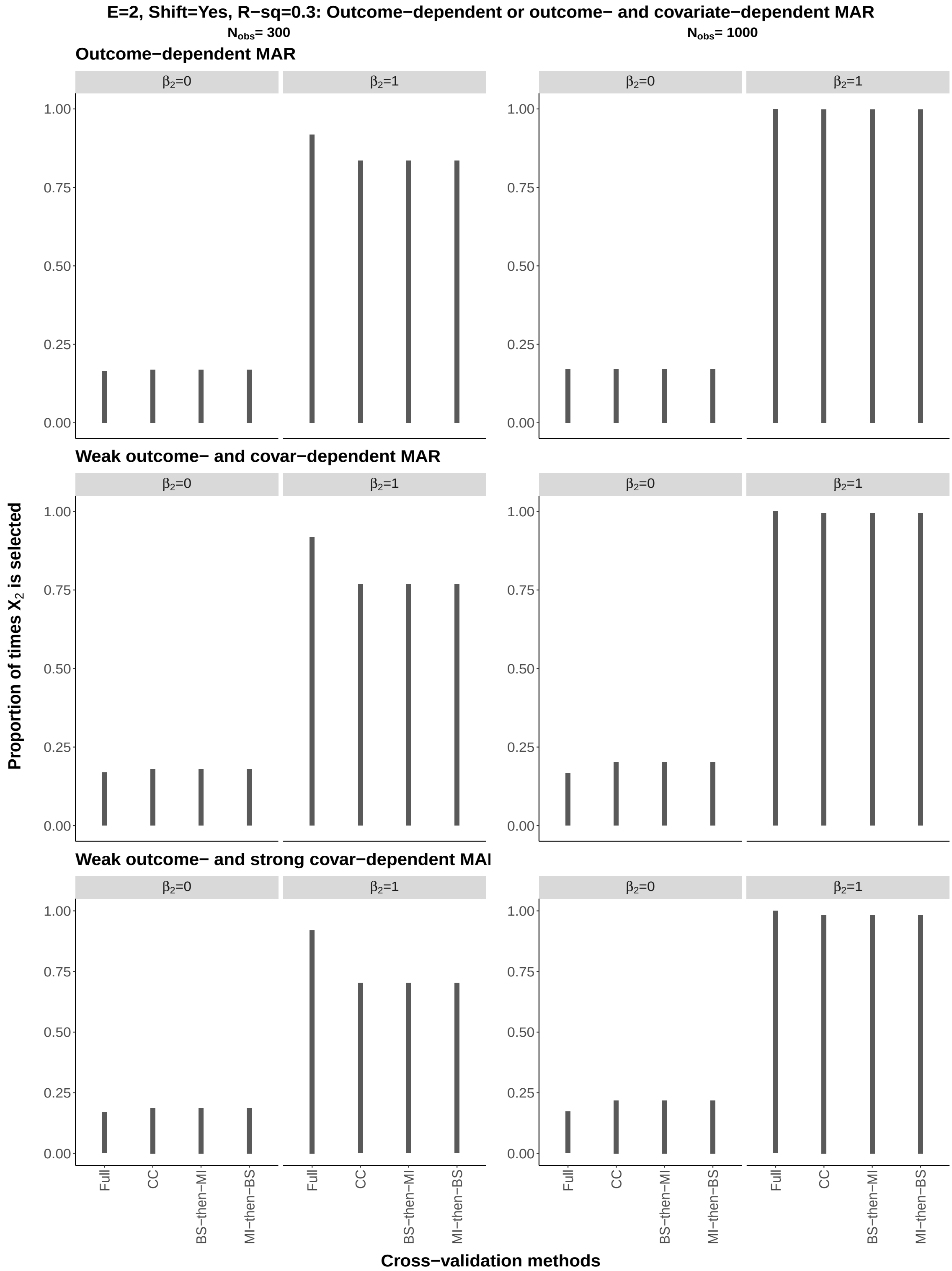

Figure S236: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

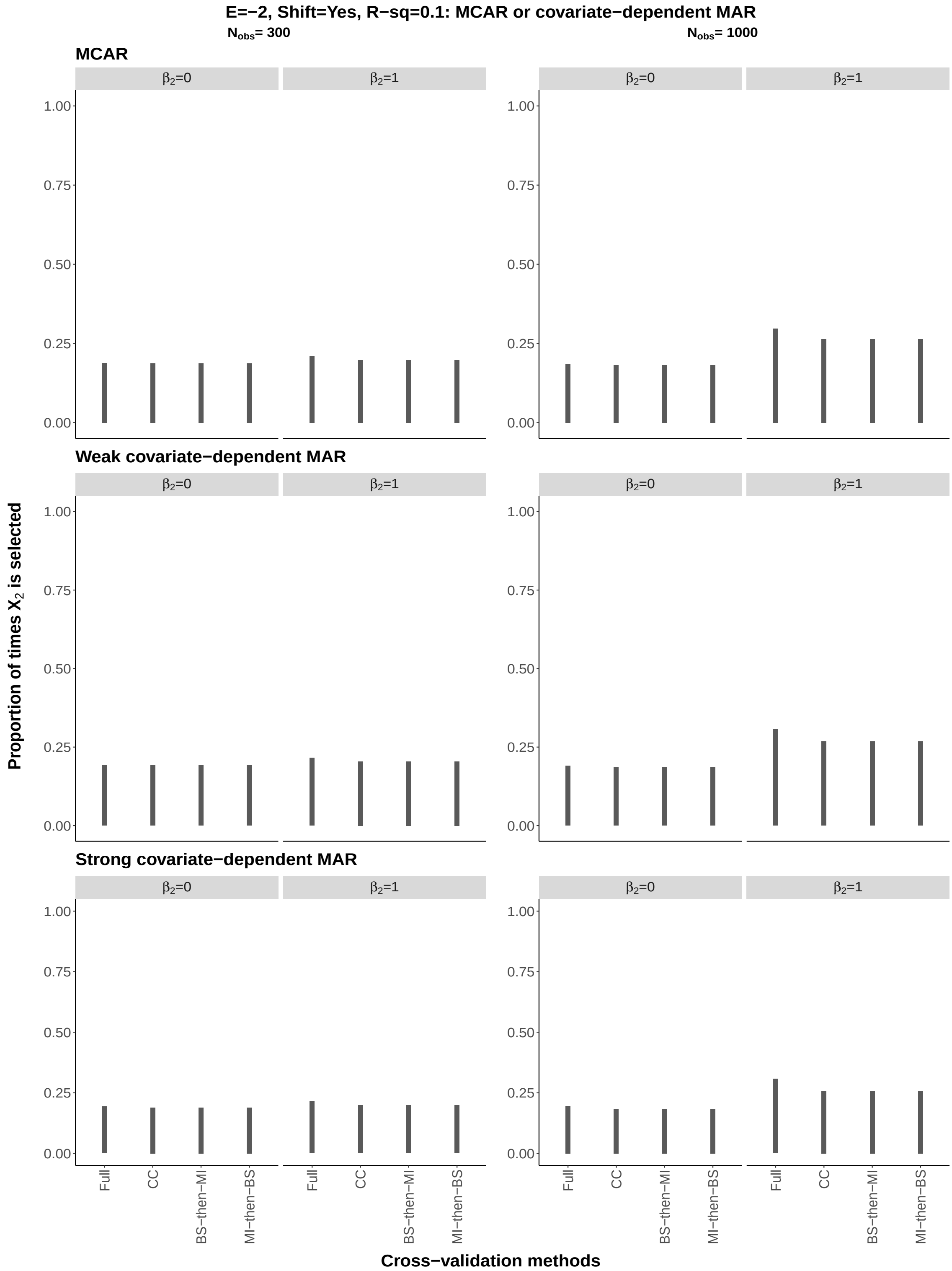

Figure S237: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

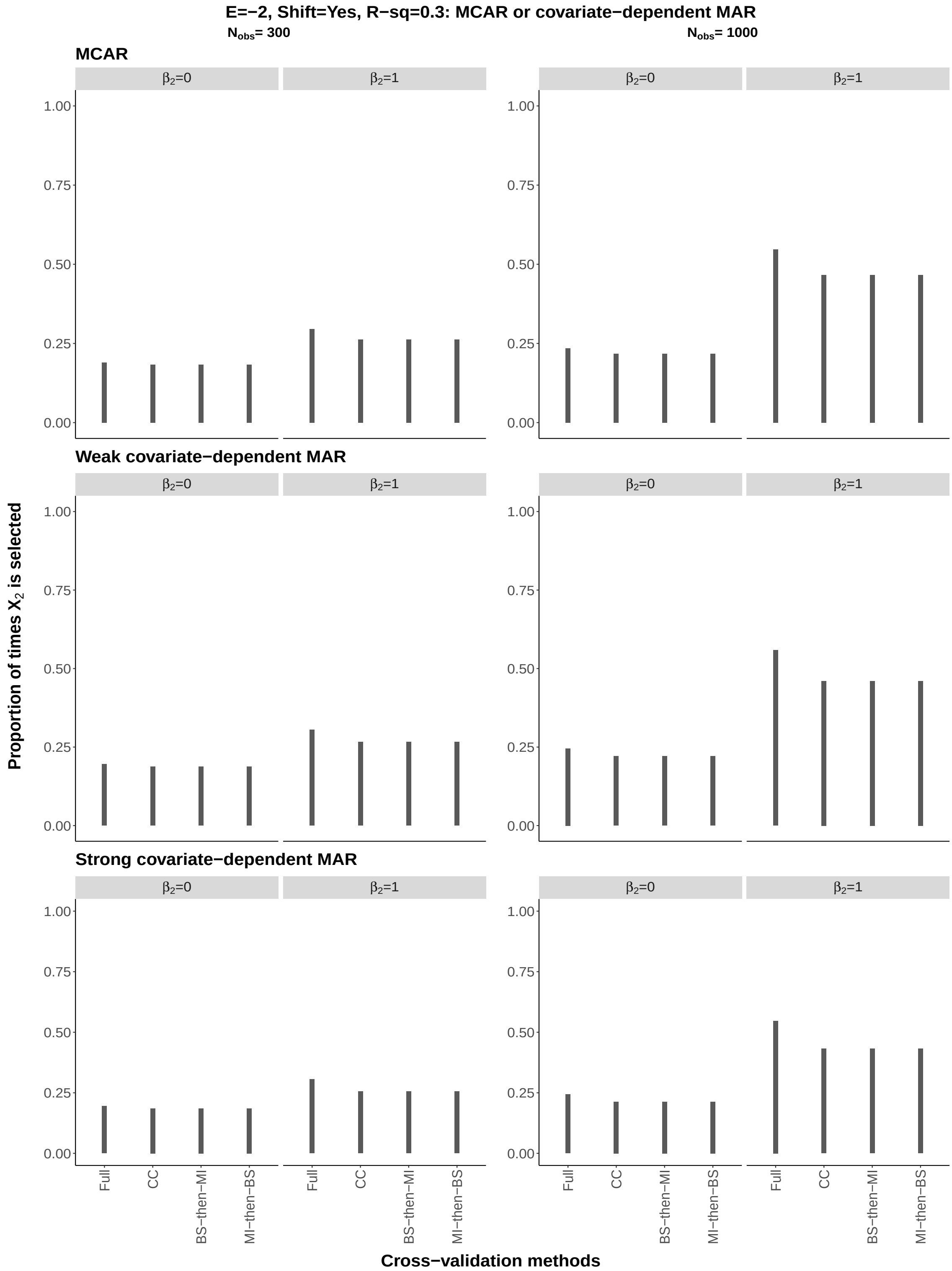

Figure S238: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

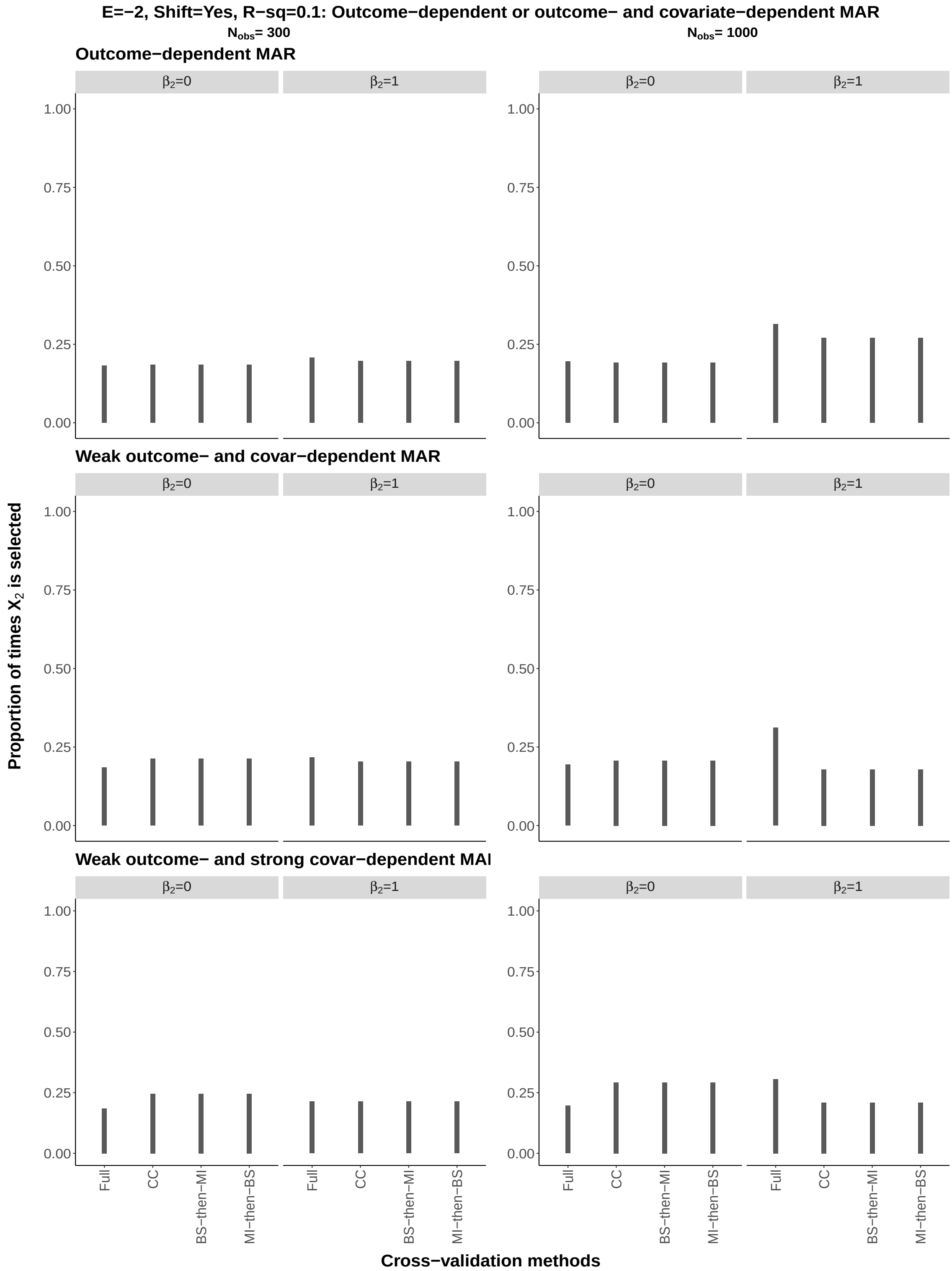

Figure S239: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

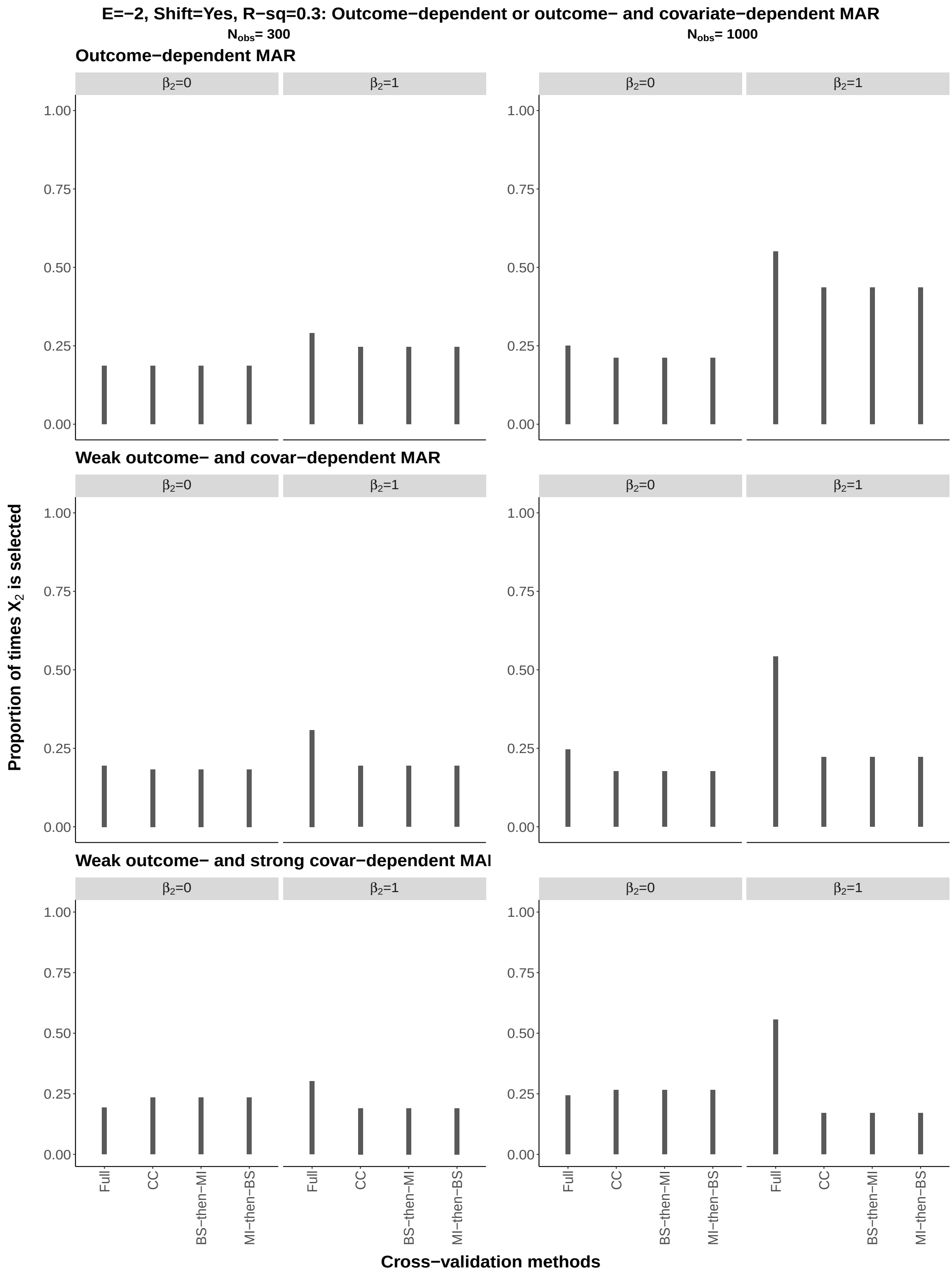

Figure S240: The proportion of times covariate X_2 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_2 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.13 Covariate selection of X_1 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 1$, no origin-shift

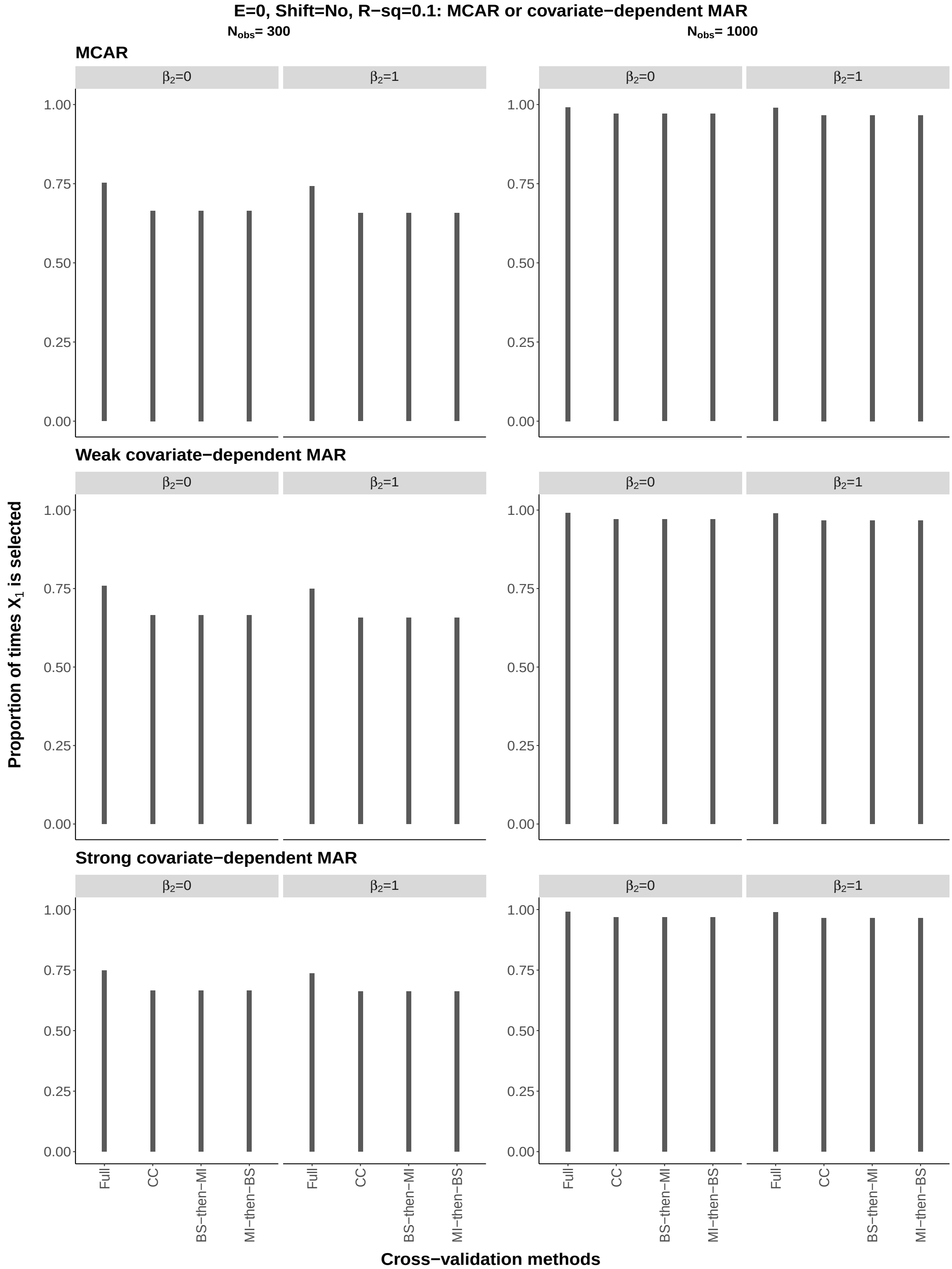

Figure S241: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

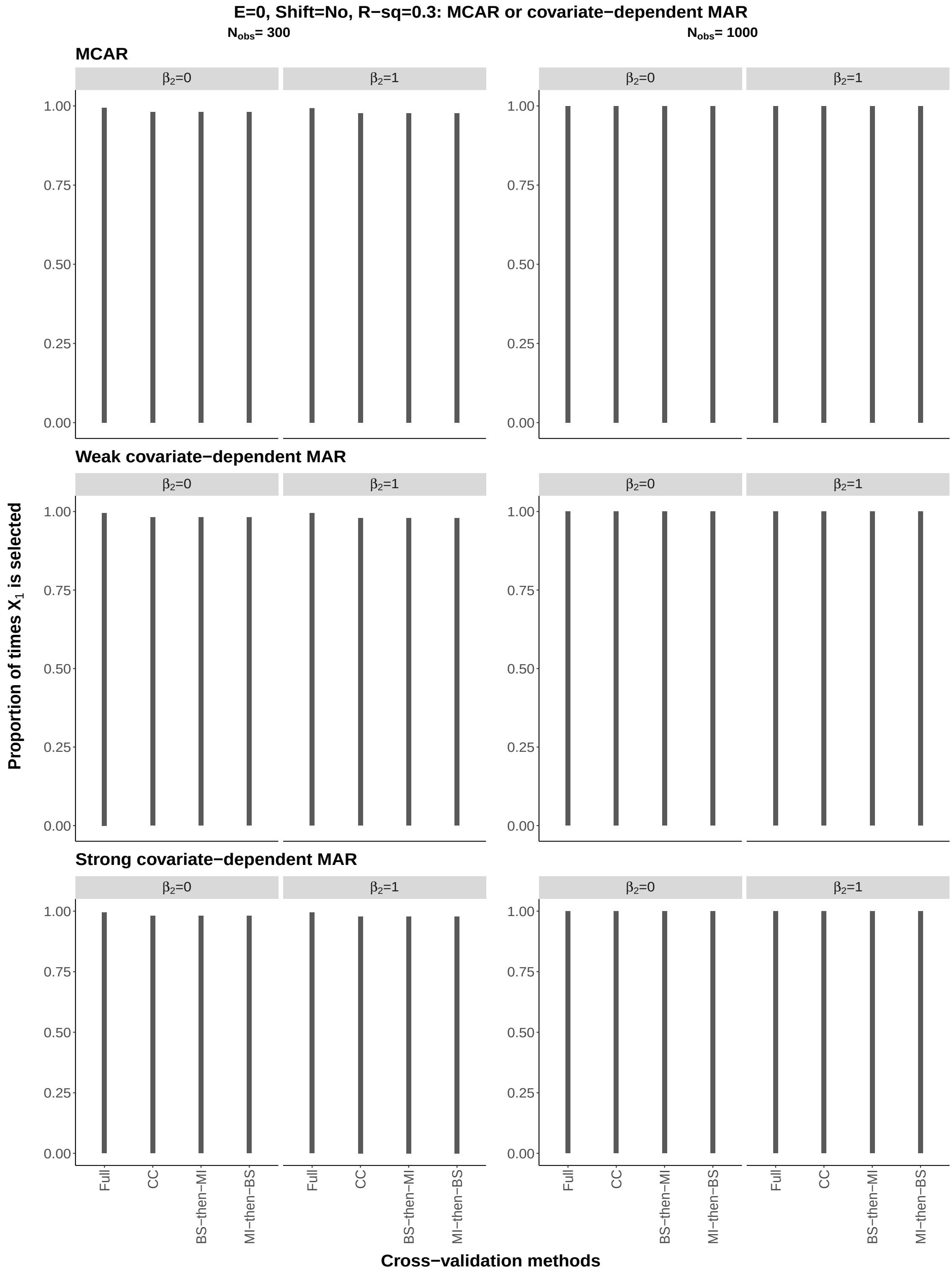

Figure S242: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

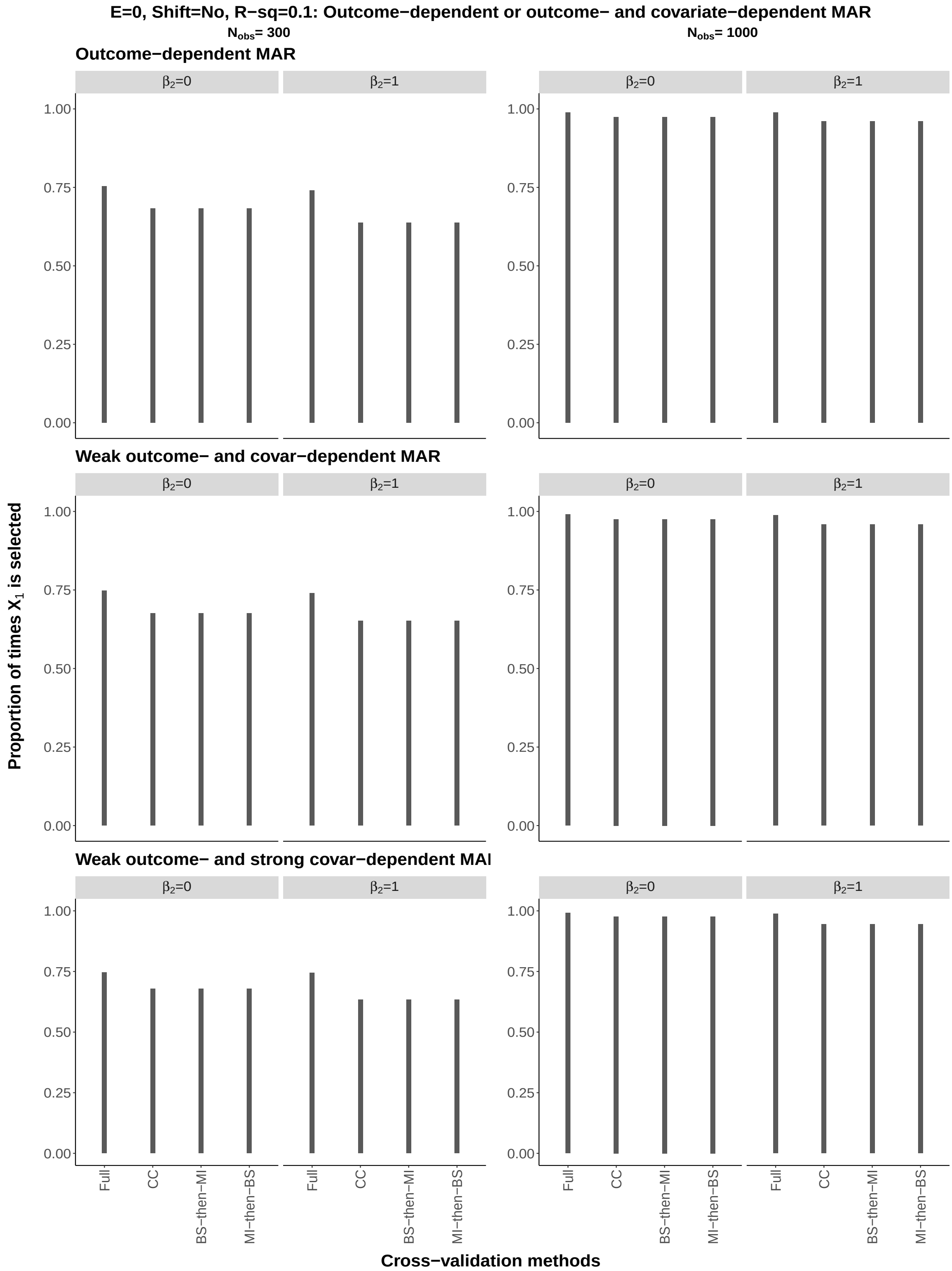

Figure S243: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

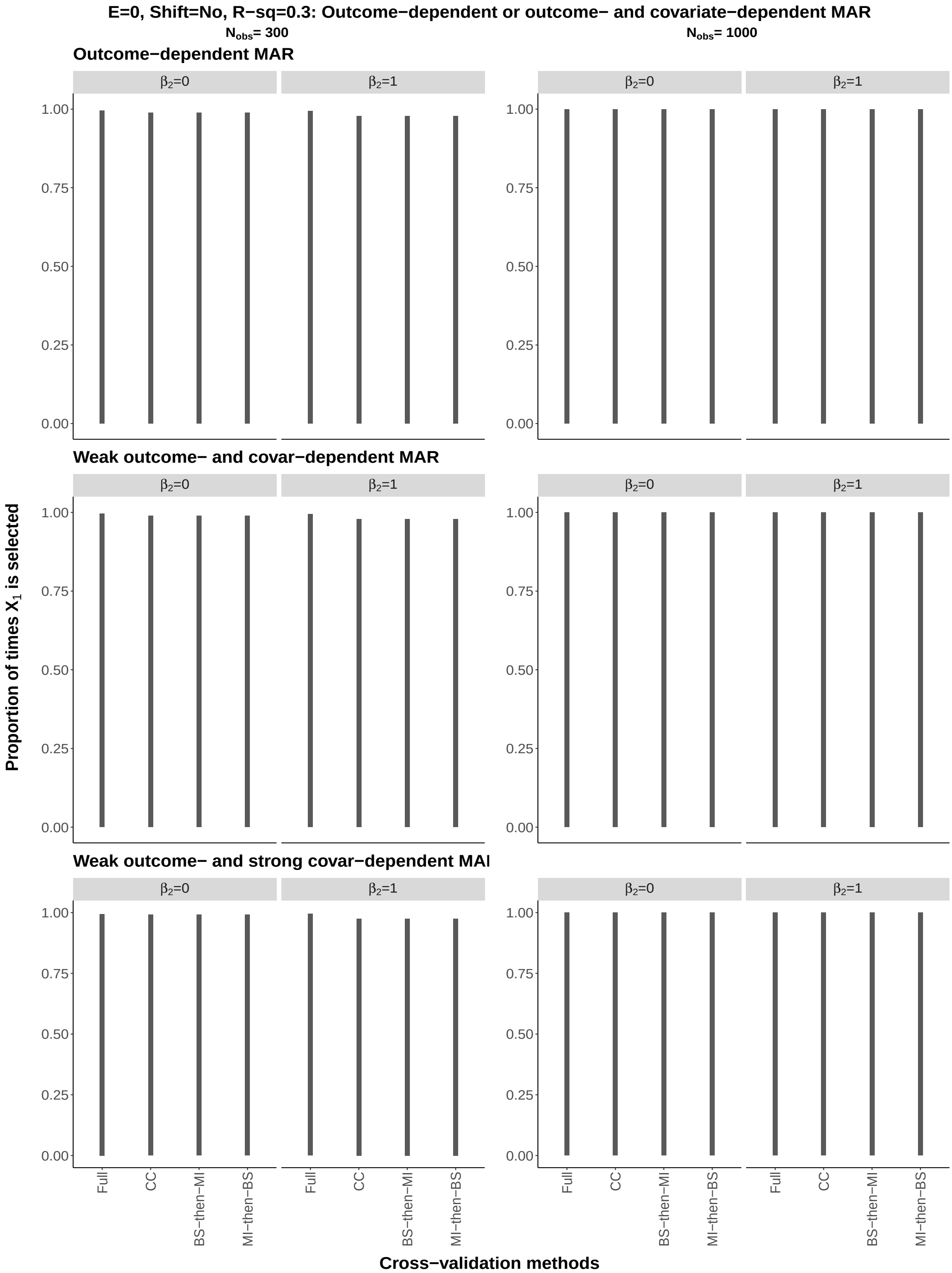

Figure S244: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

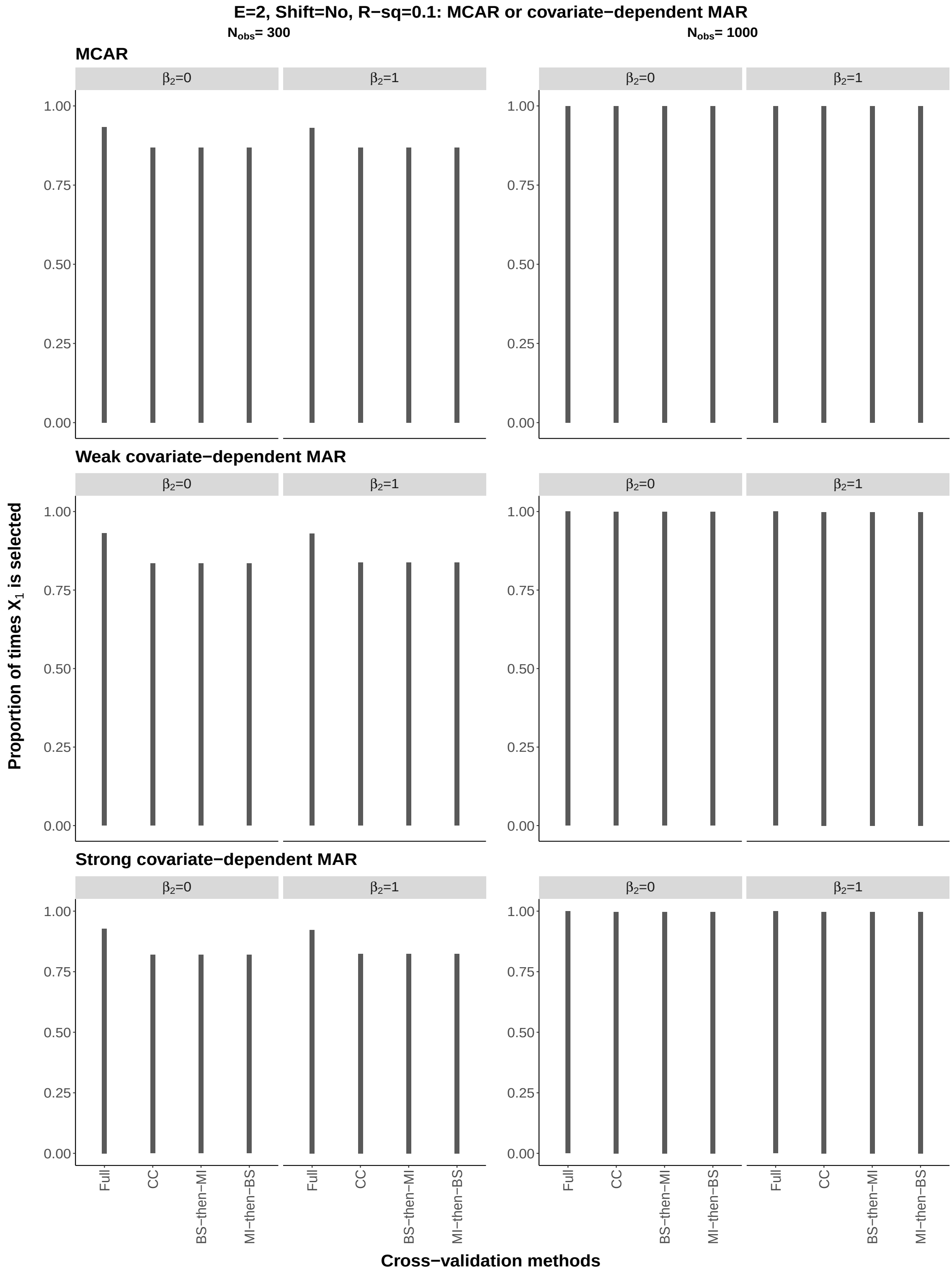

Figure S245: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

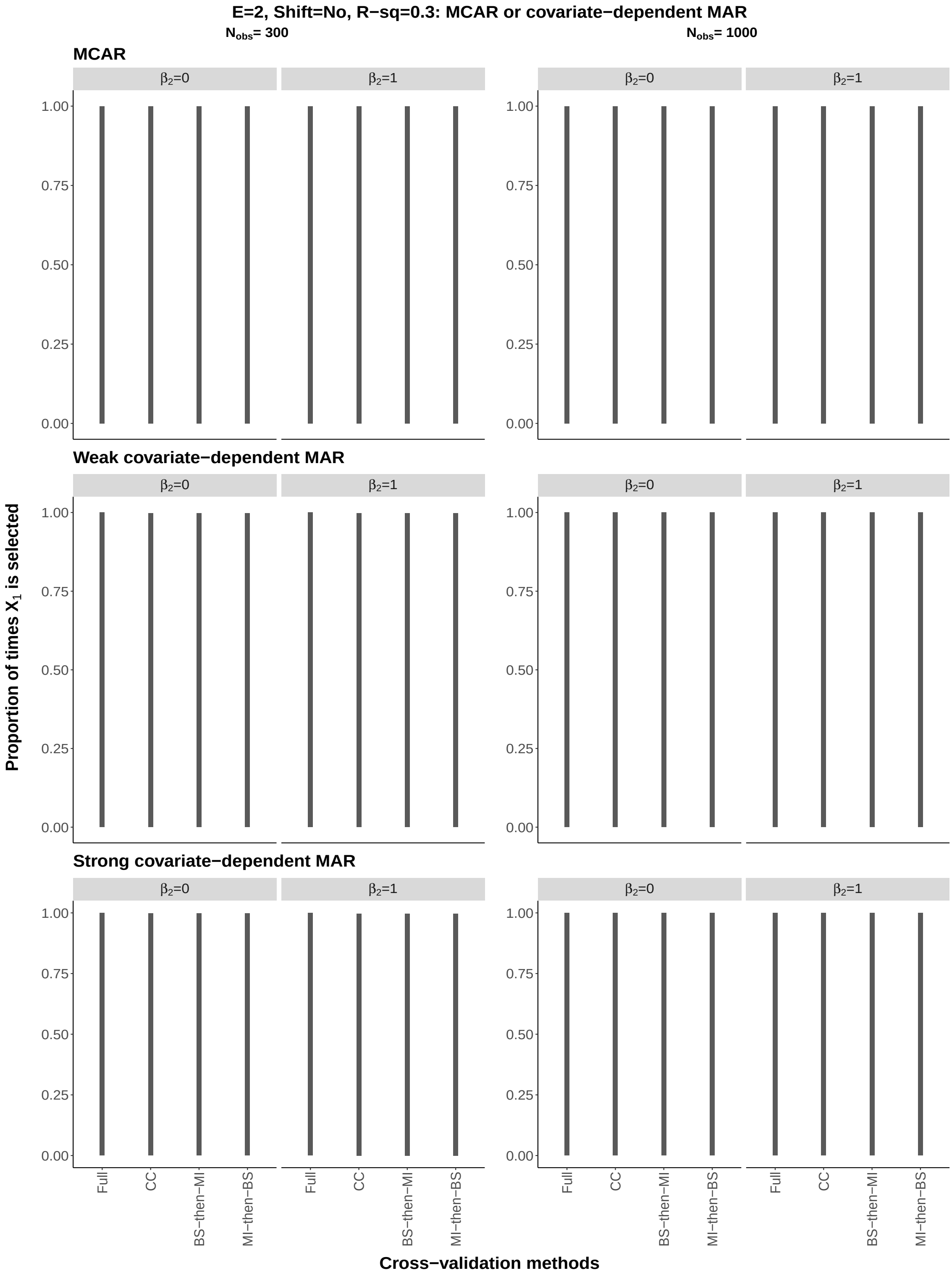

Figure S246: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

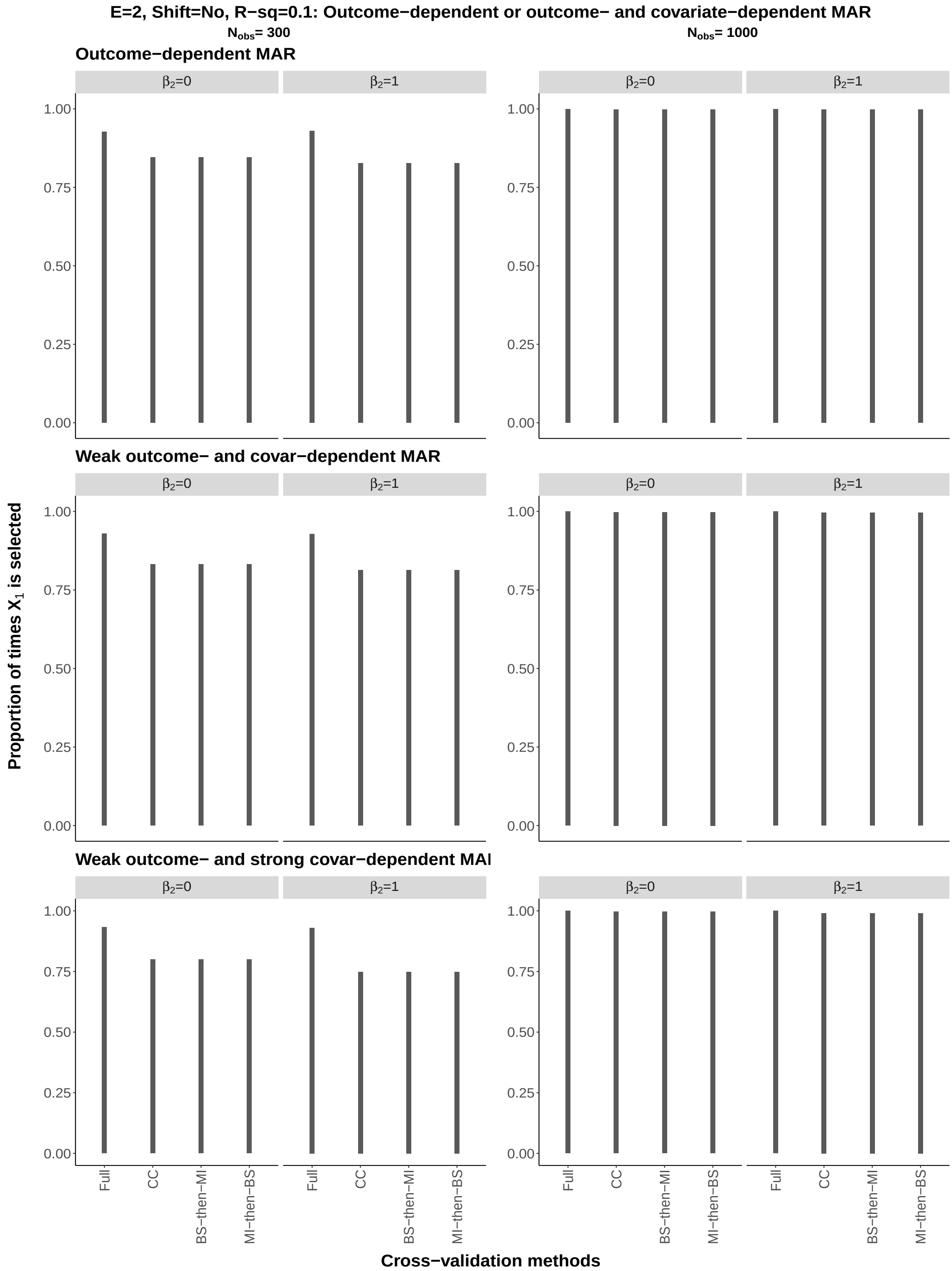

Figure S247: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

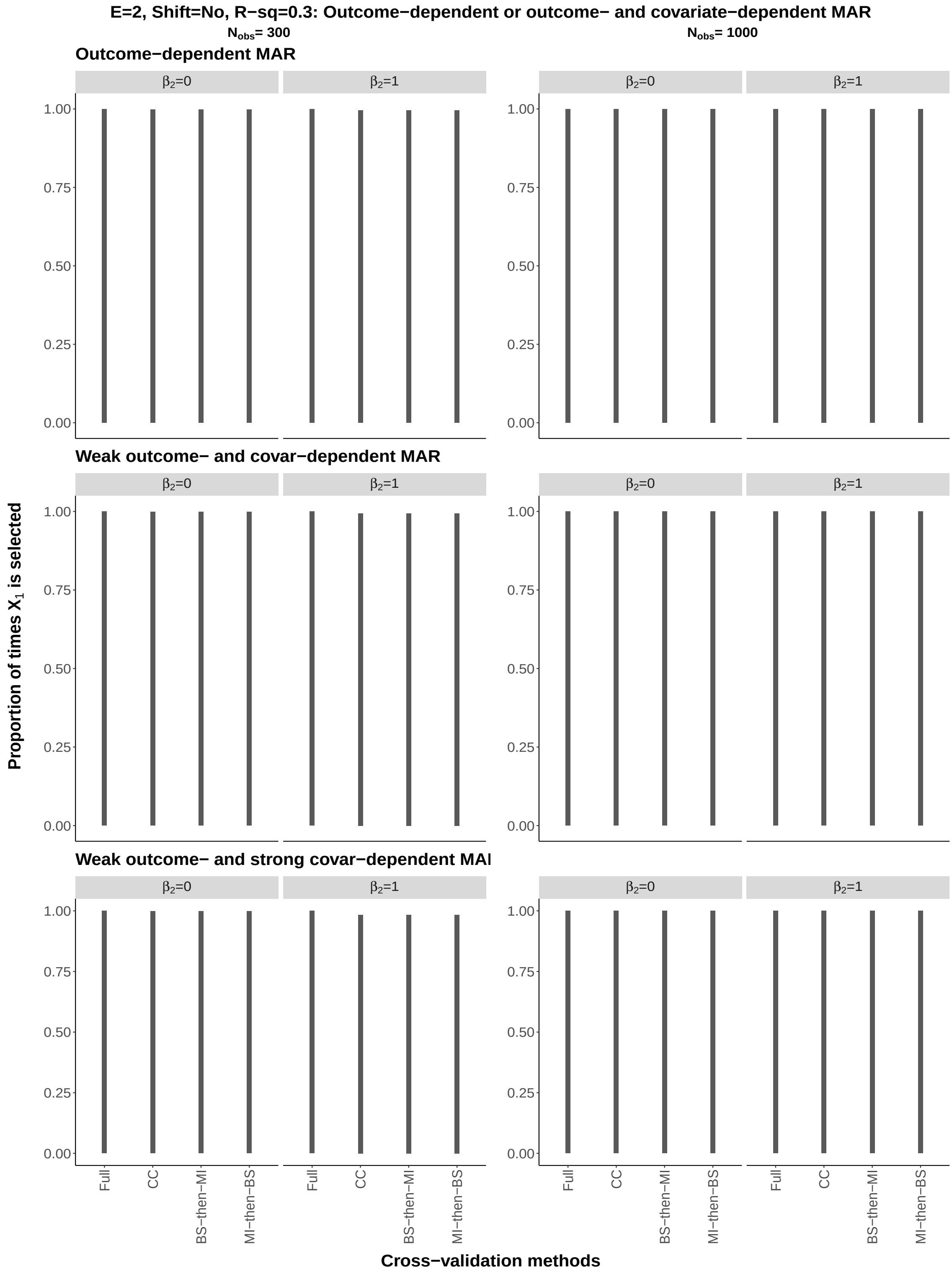

Figure S248: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

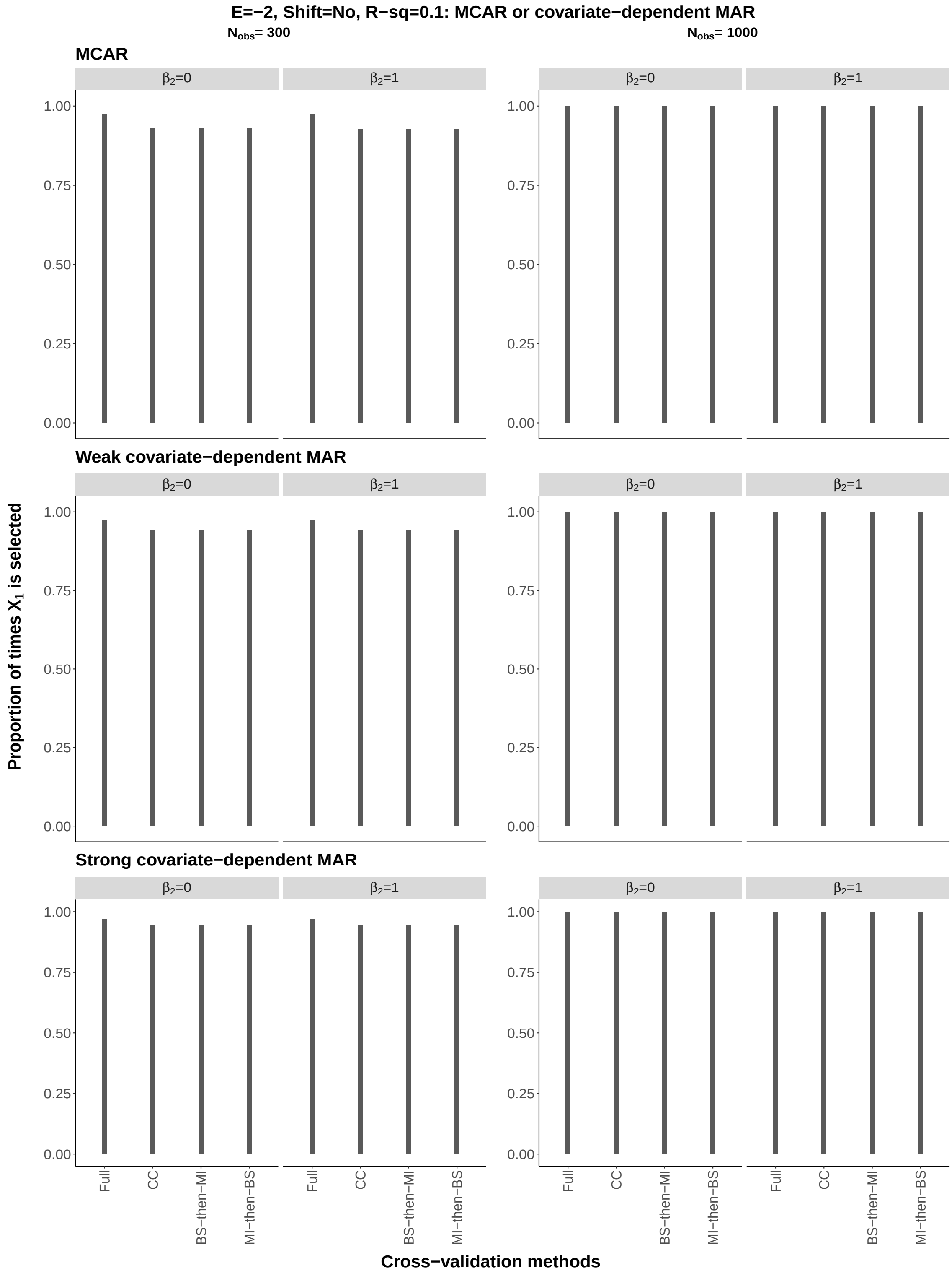

Figure S249: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

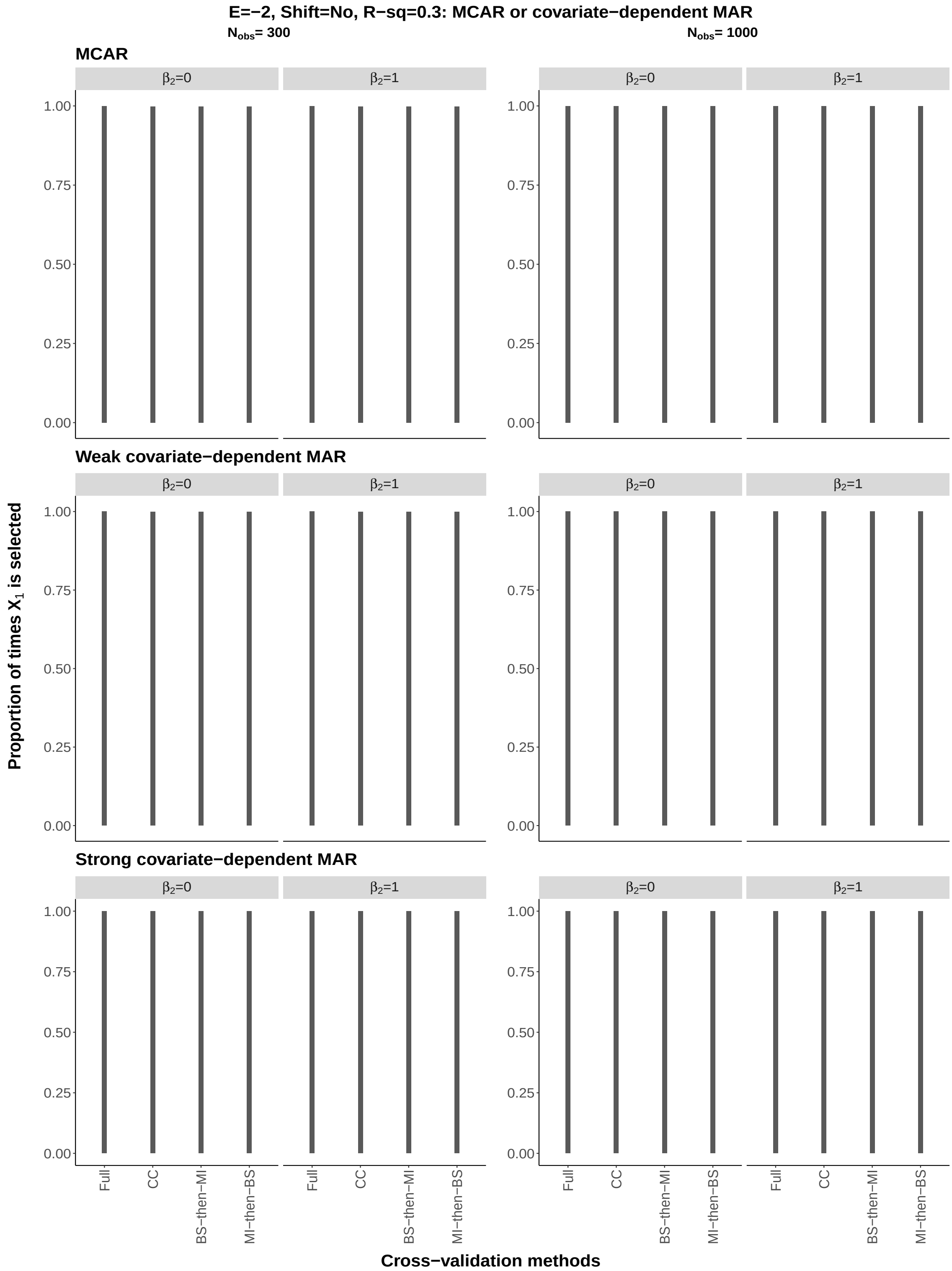

Figure S250: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

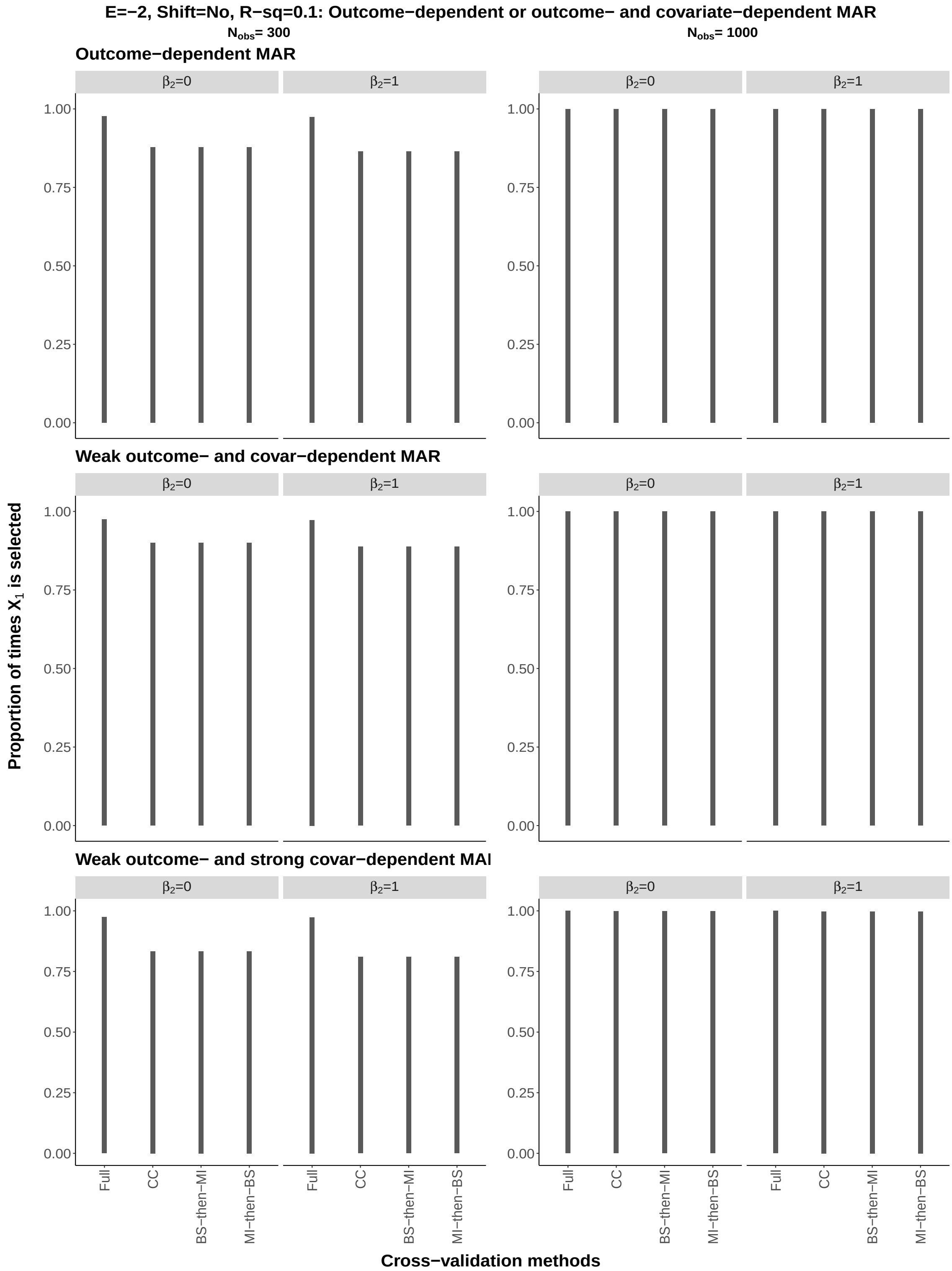

Figure S251: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

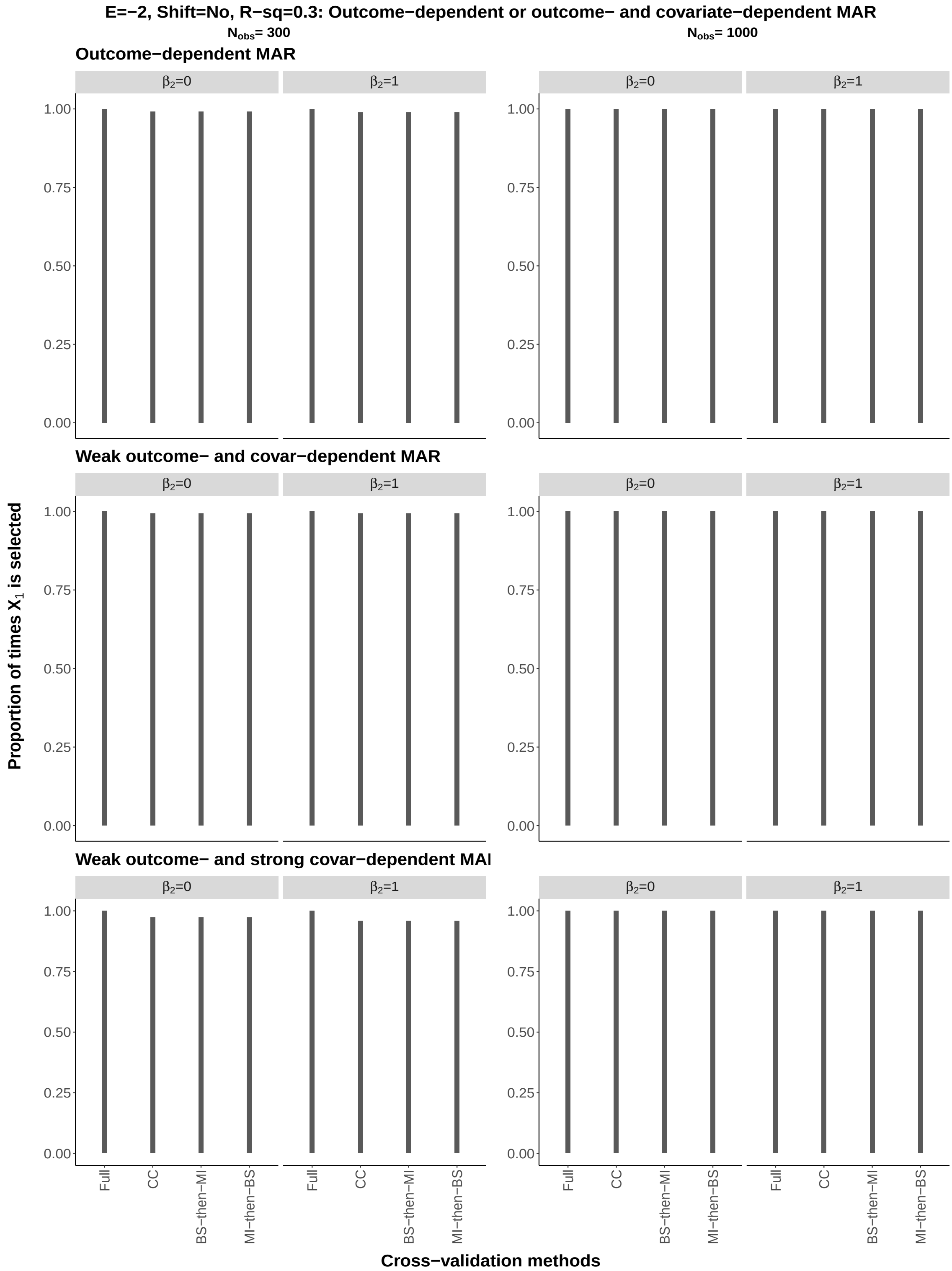

Figure S252: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.14 Covariate selection of X_1 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 0.05$, no origin-shift

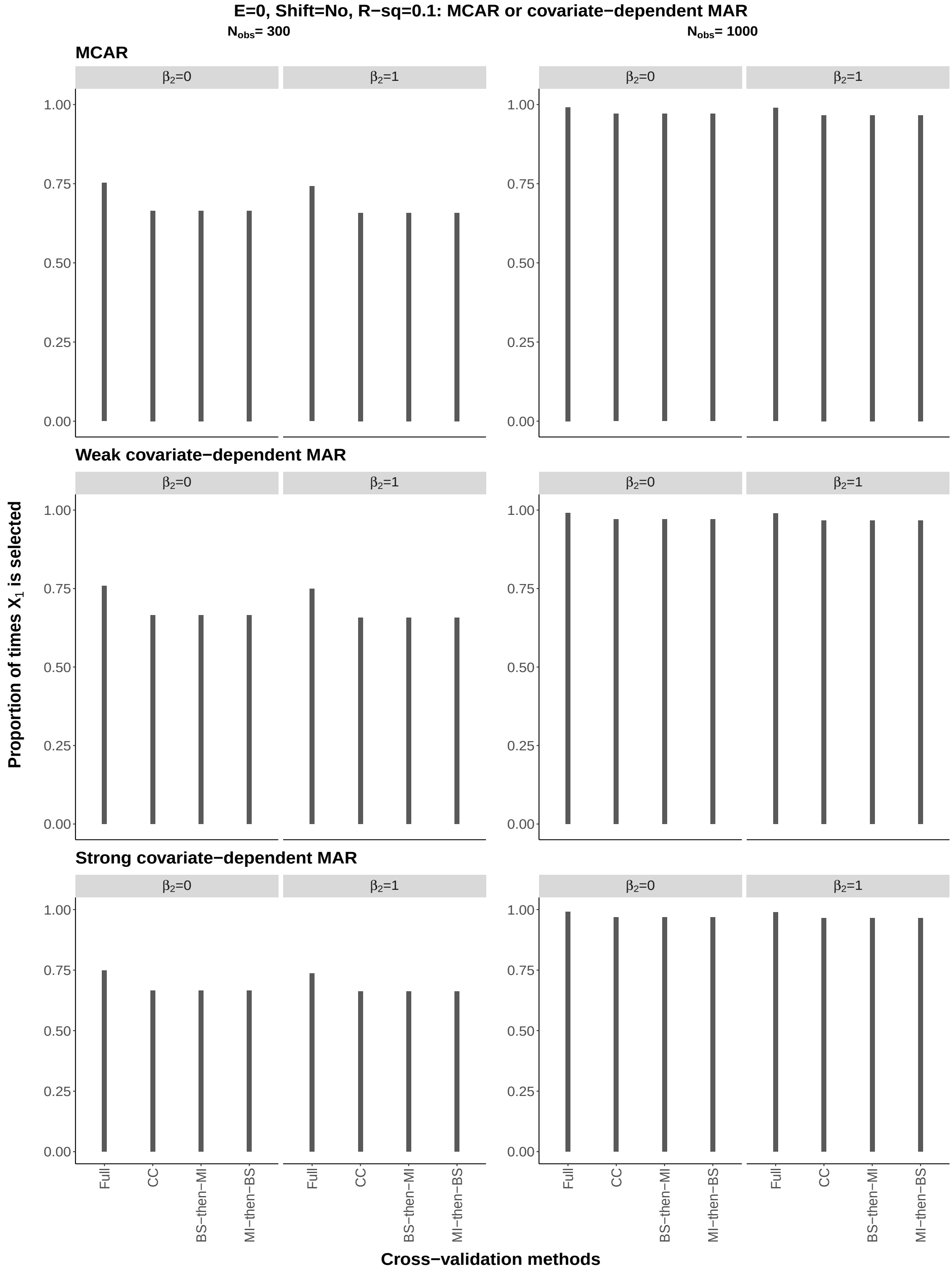

Figure S253: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

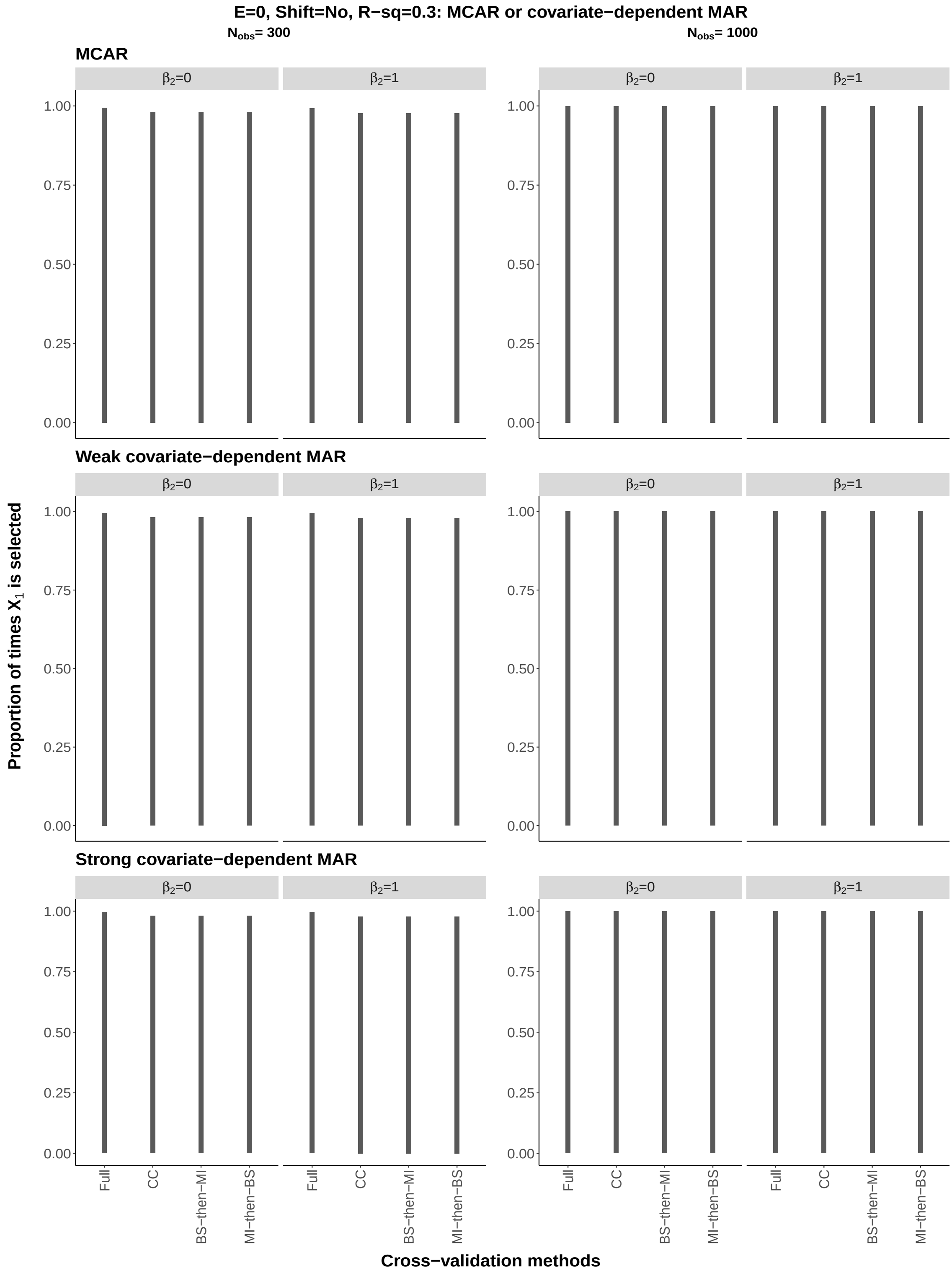

Figure S254: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

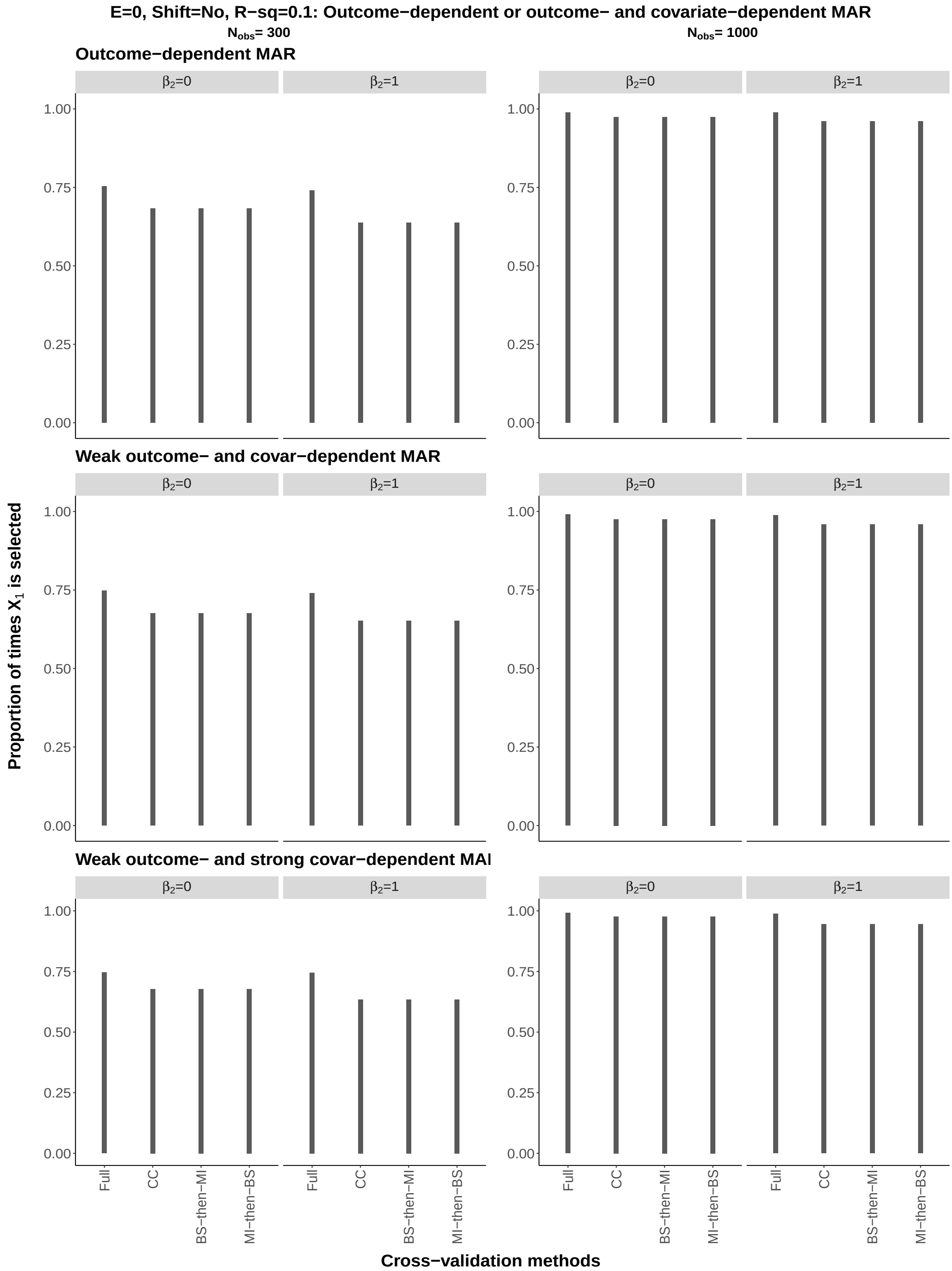

Figure S255: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

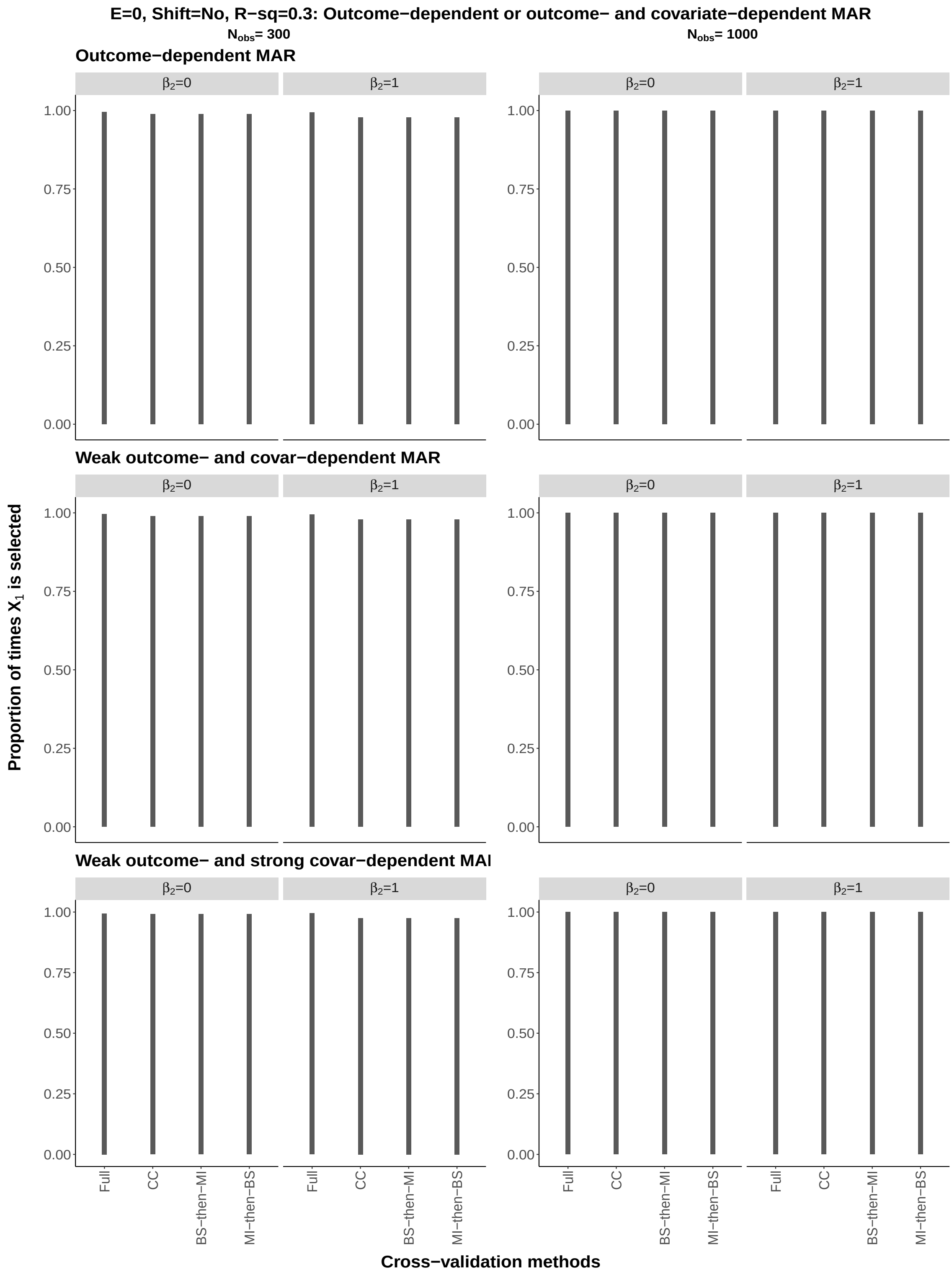

Figure S256: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 0

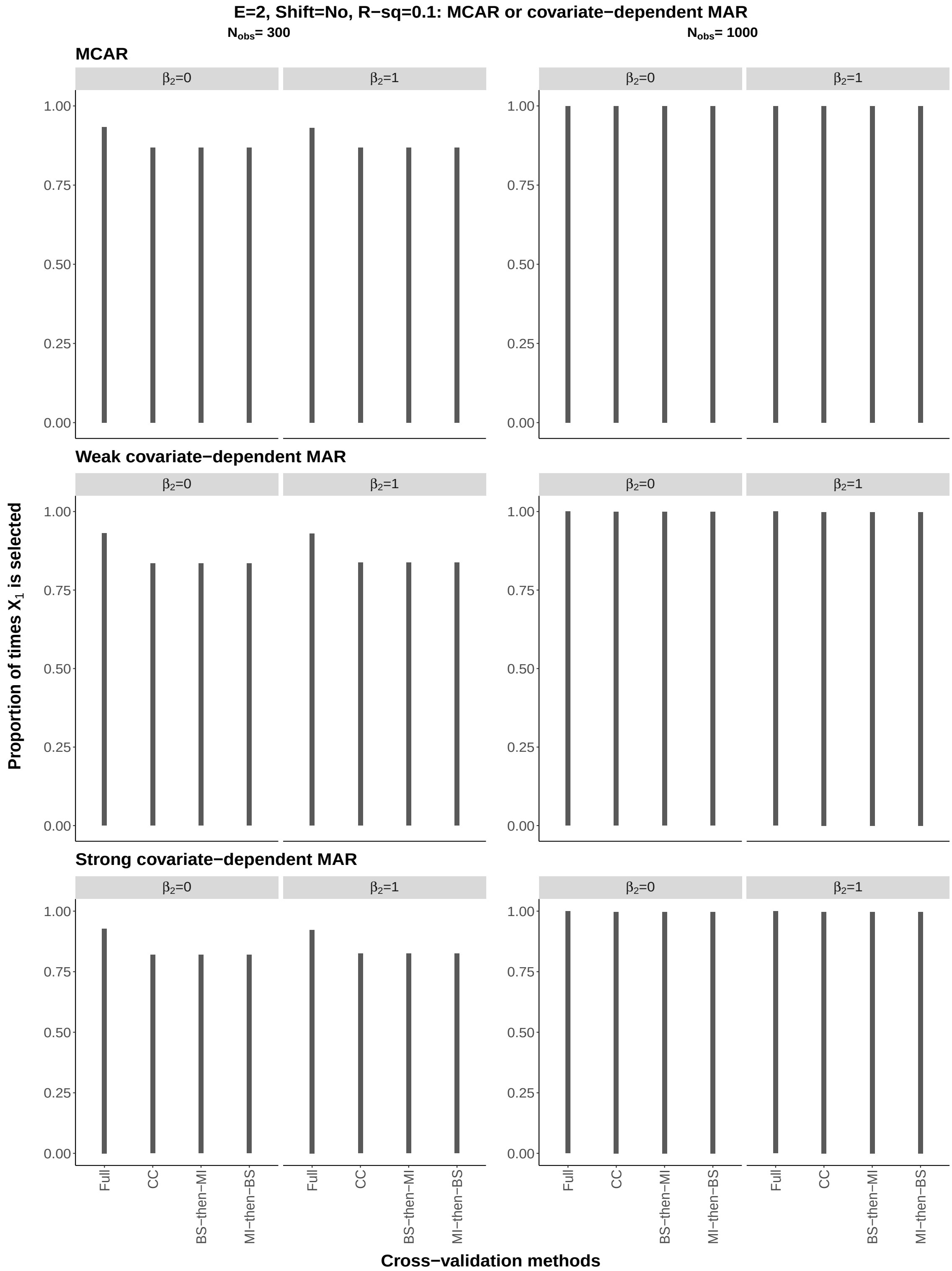

Figure S257: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

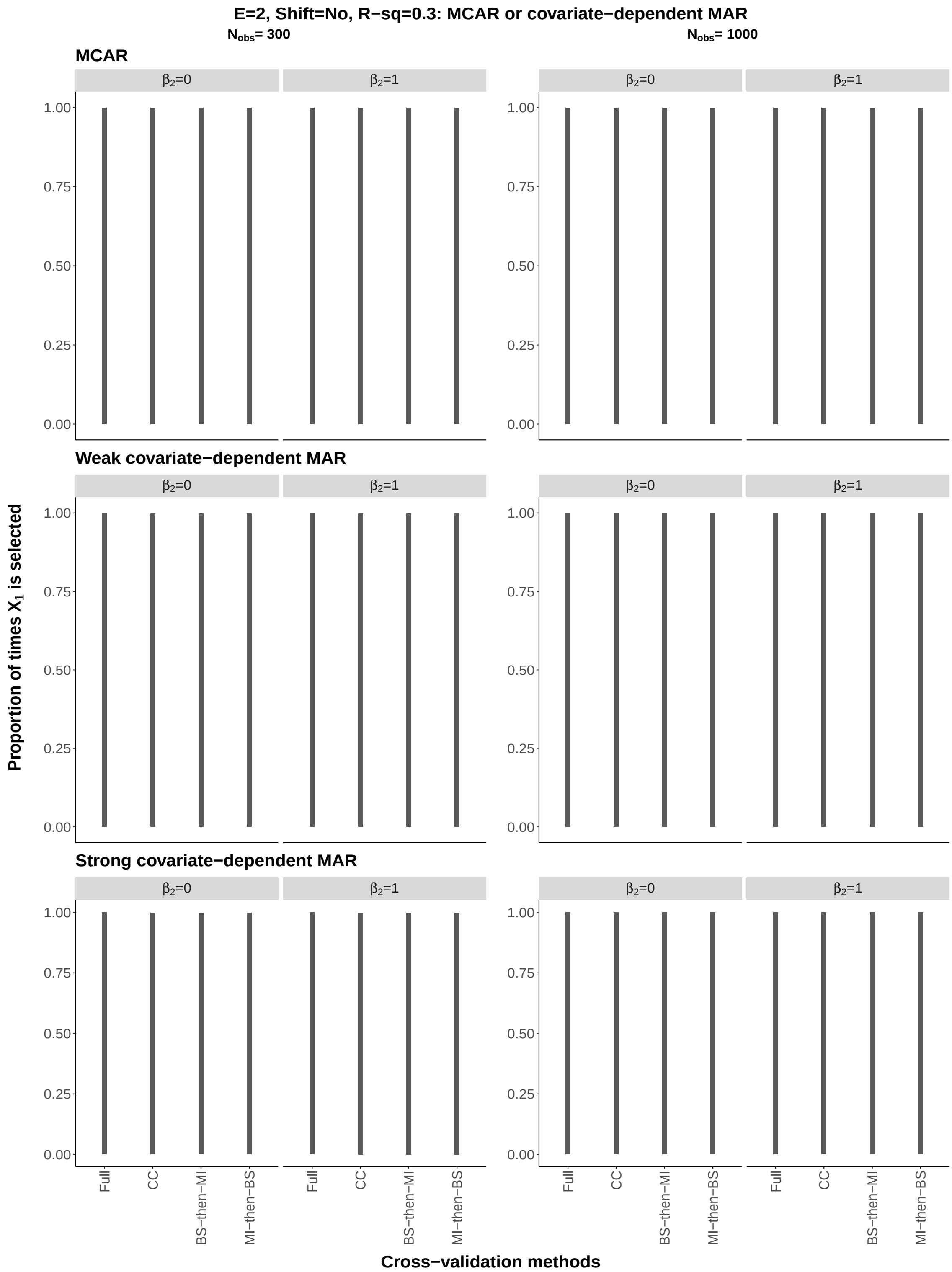

Figure S258: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

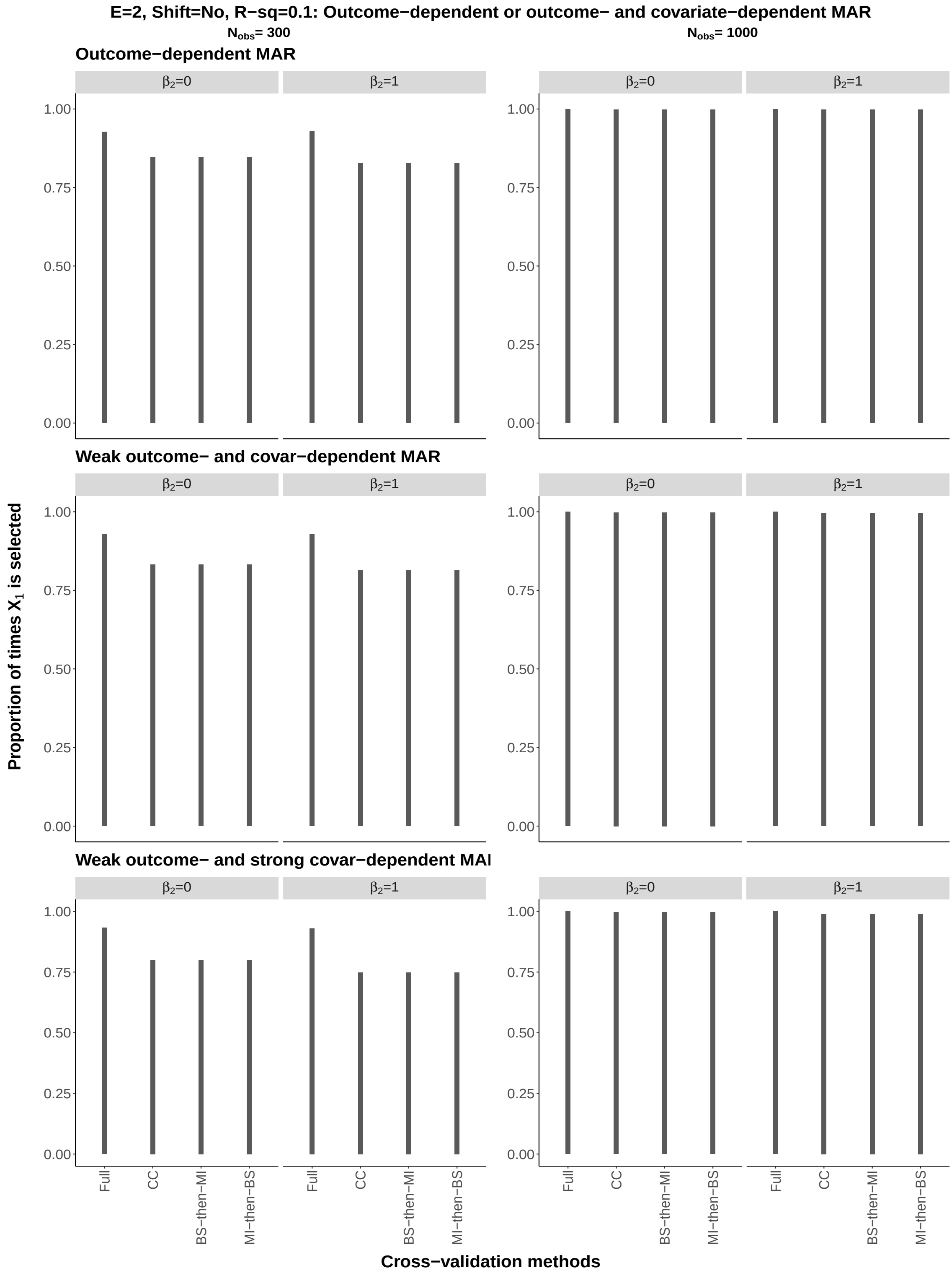

Figure S259: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

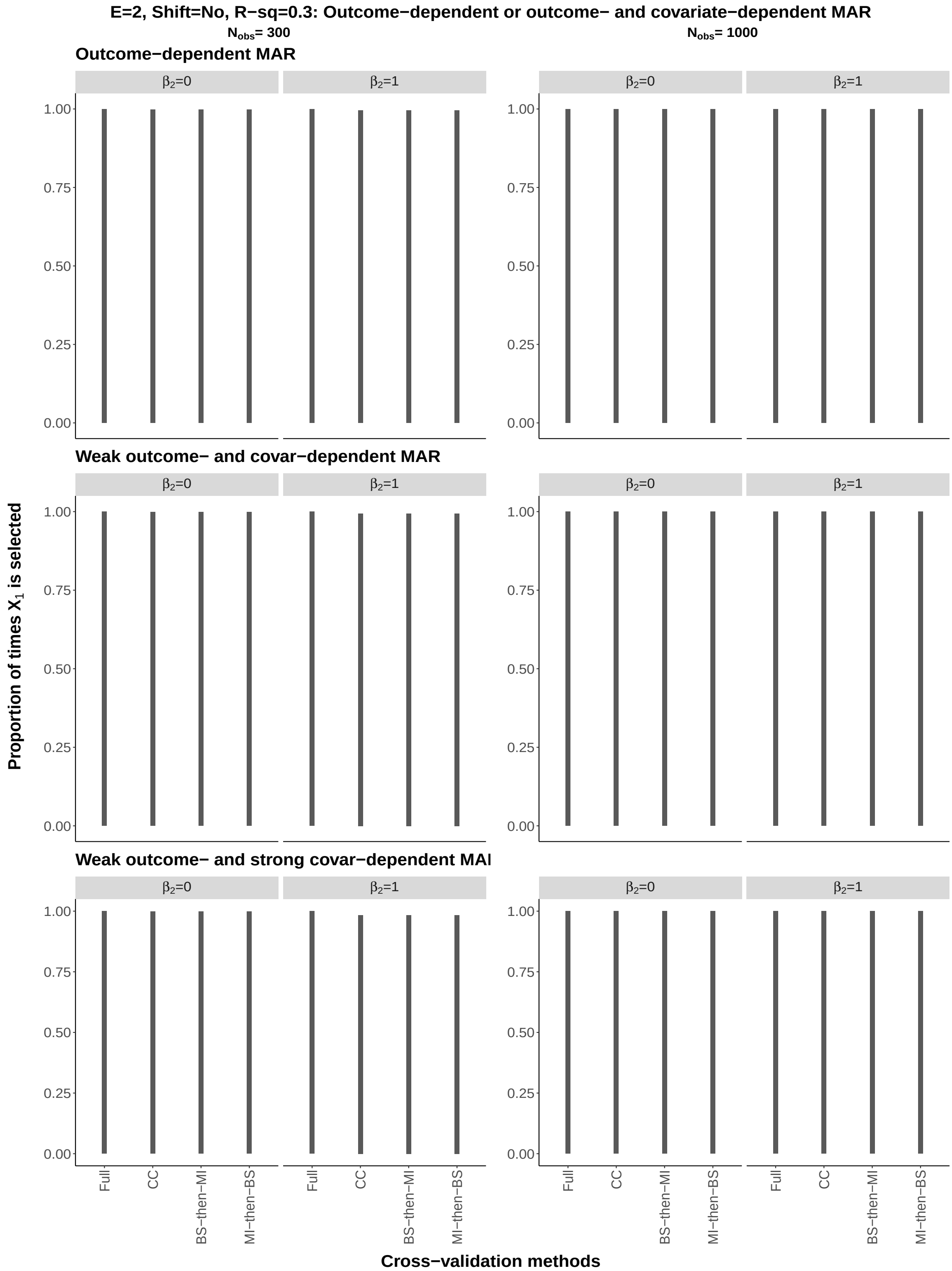

Figure S260: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

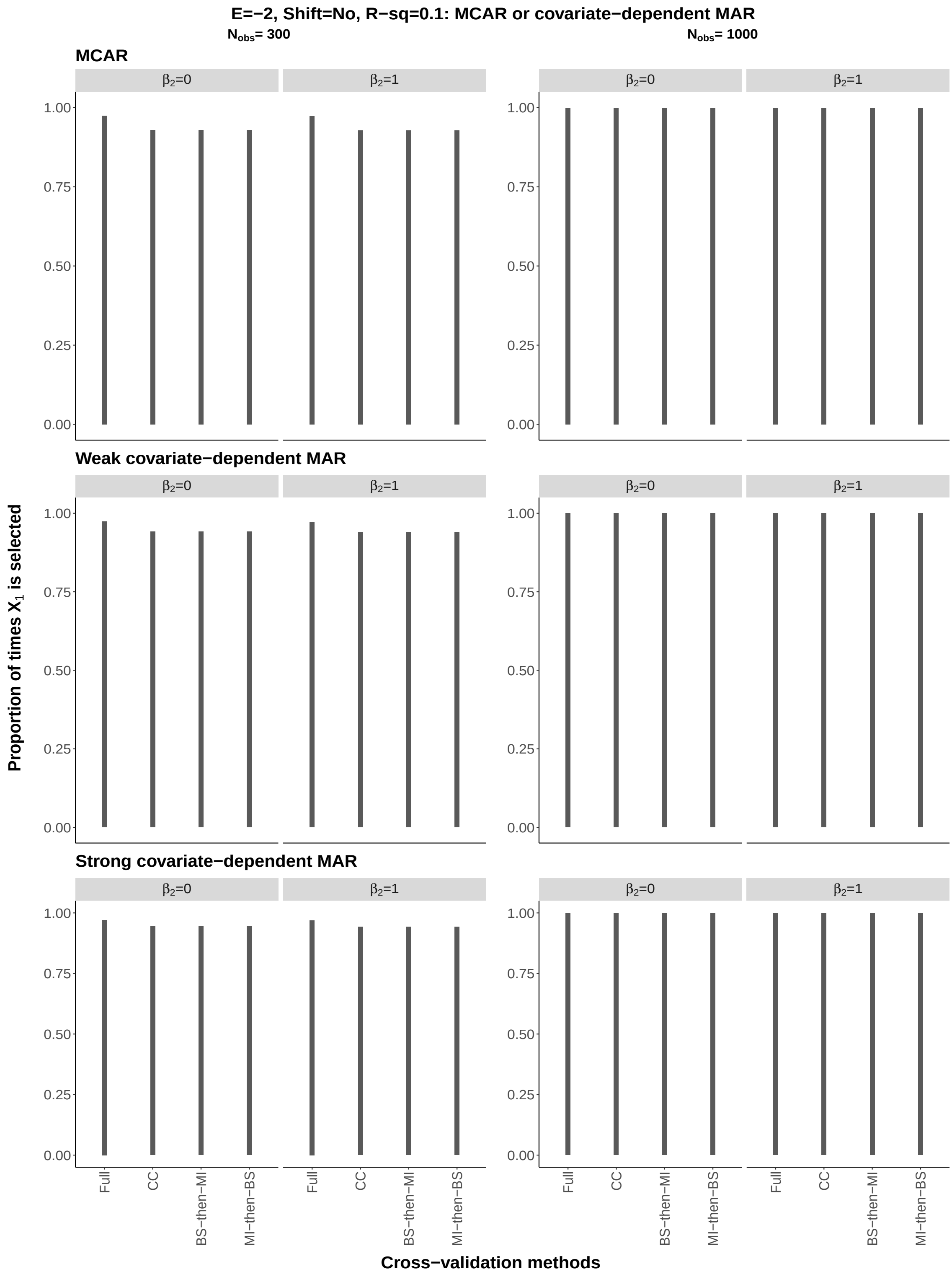

Figure S261: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

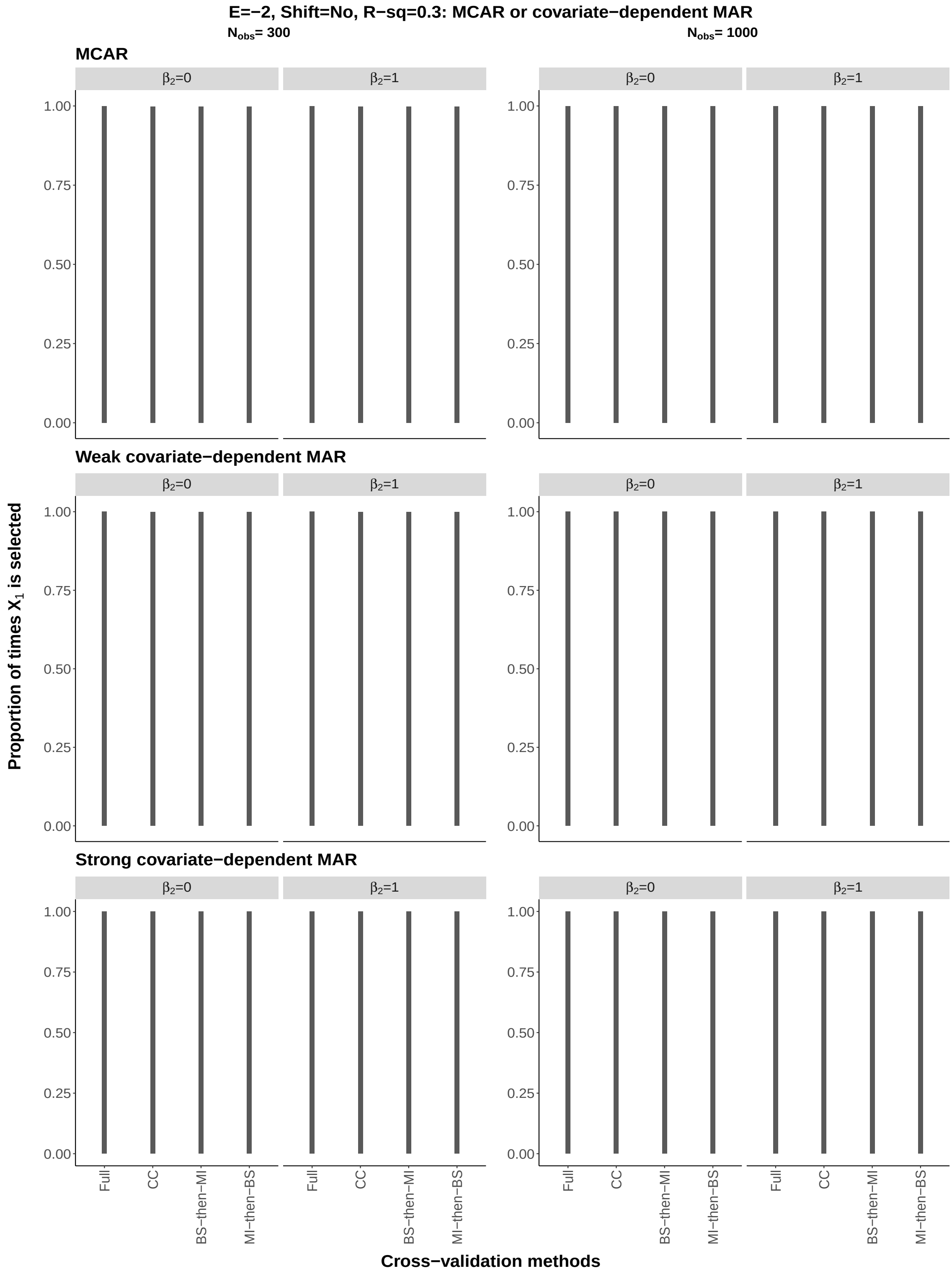

Figure S262: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

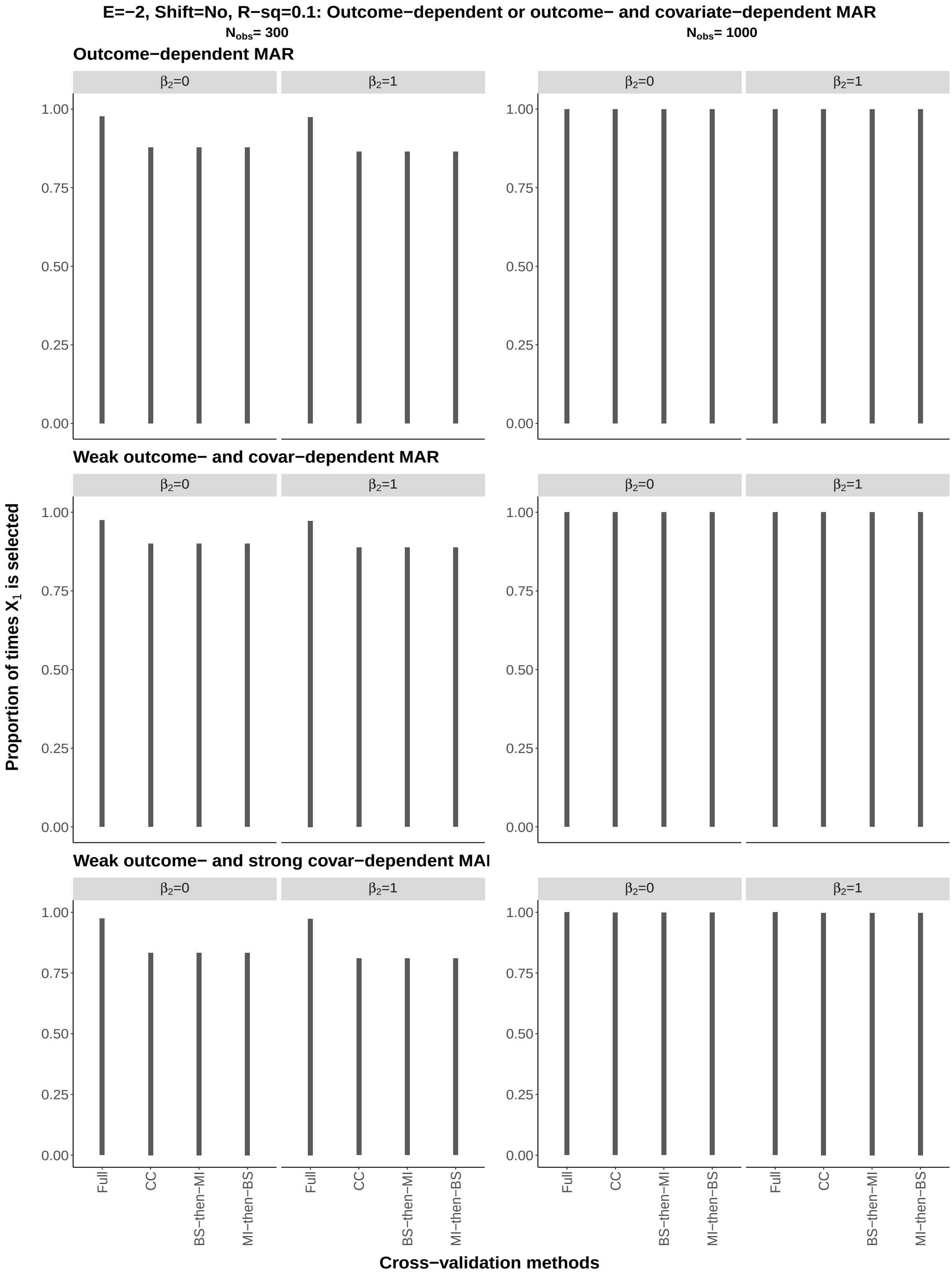

Figure S263: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

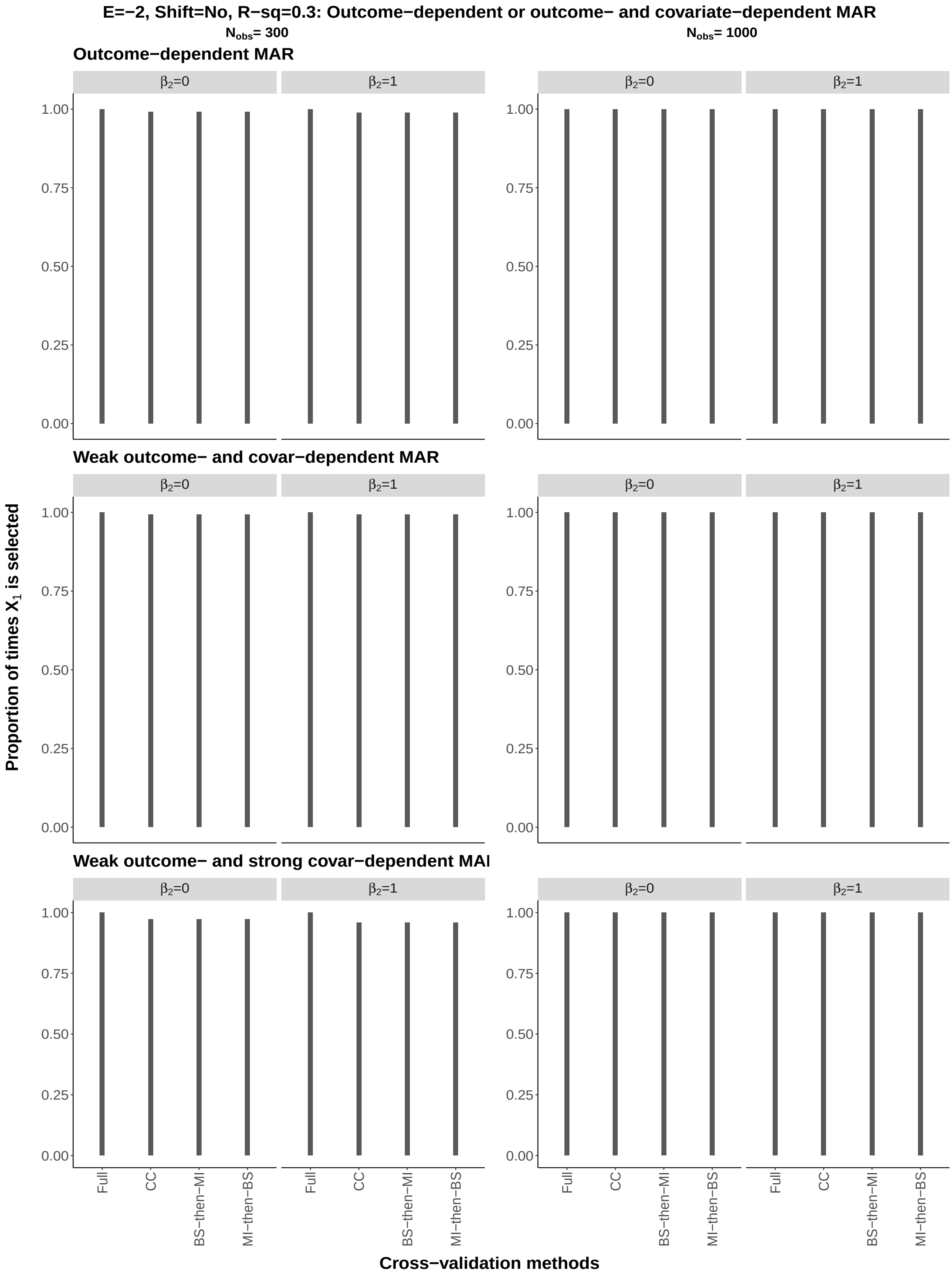

Figure S264: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has not been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

S9.2.15 Covariate selection of X_1 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 1$ and an origin-shift has been applied

True exponent is 0

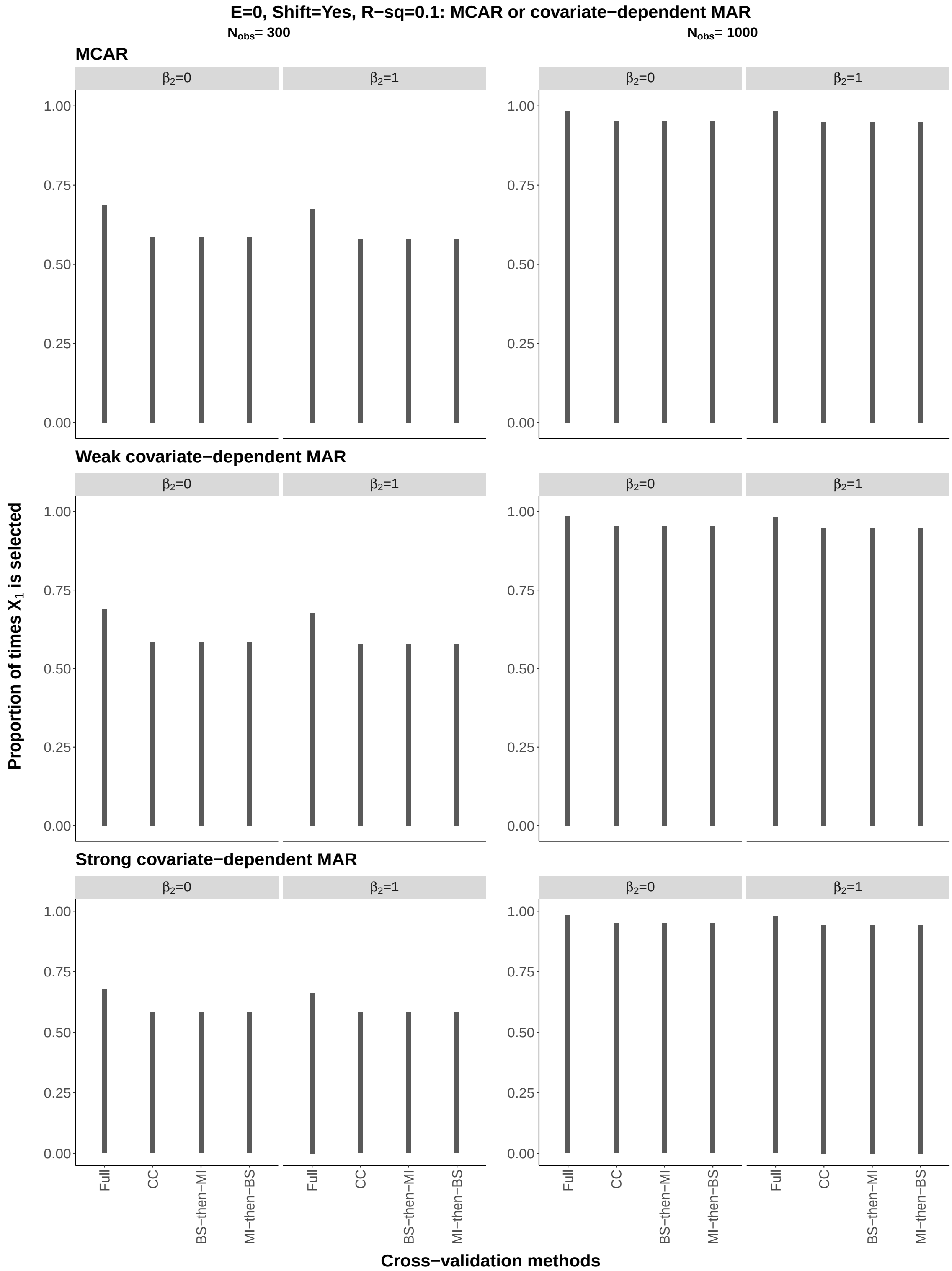

Figure S265: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

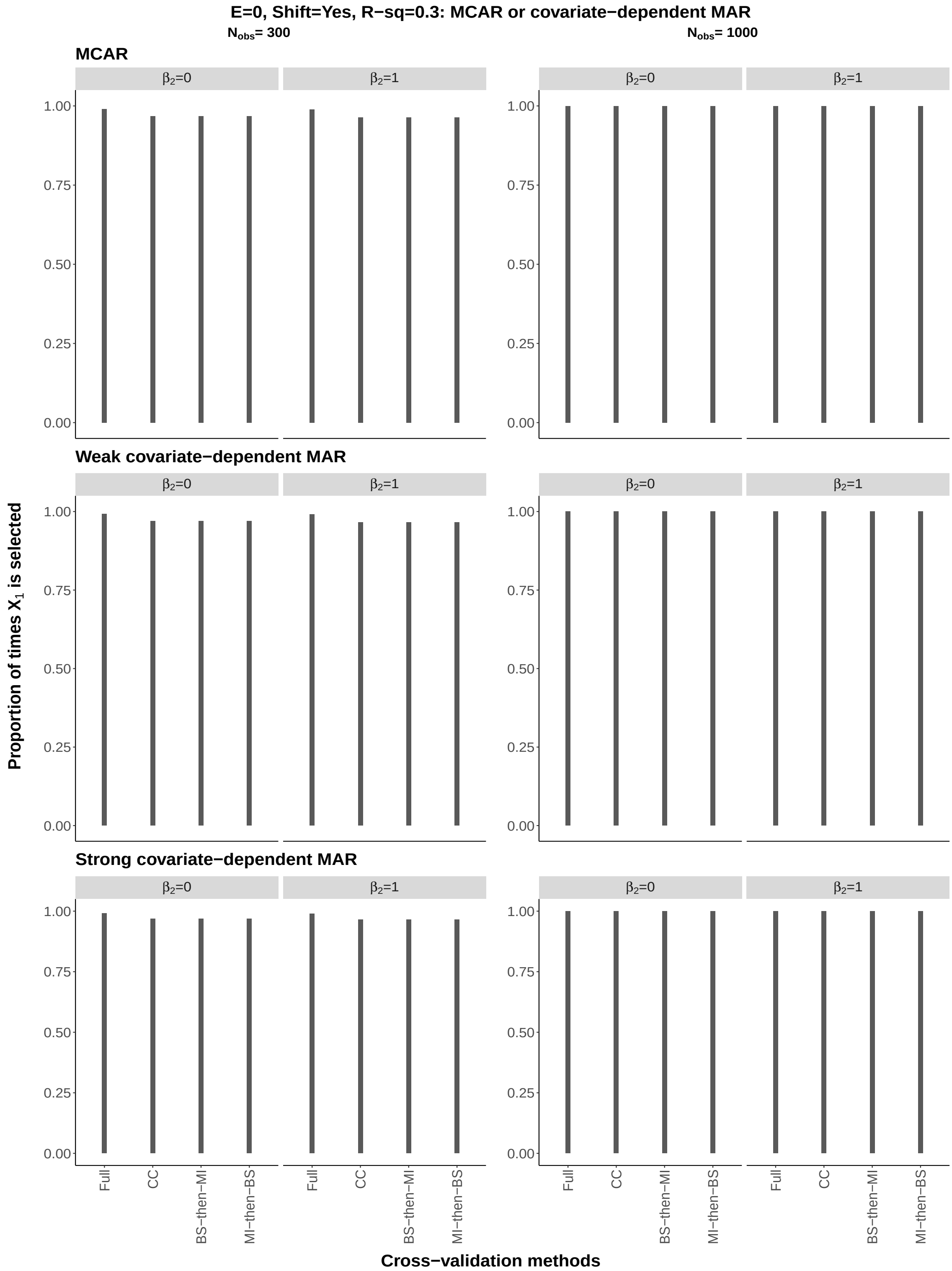

Figure S266: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

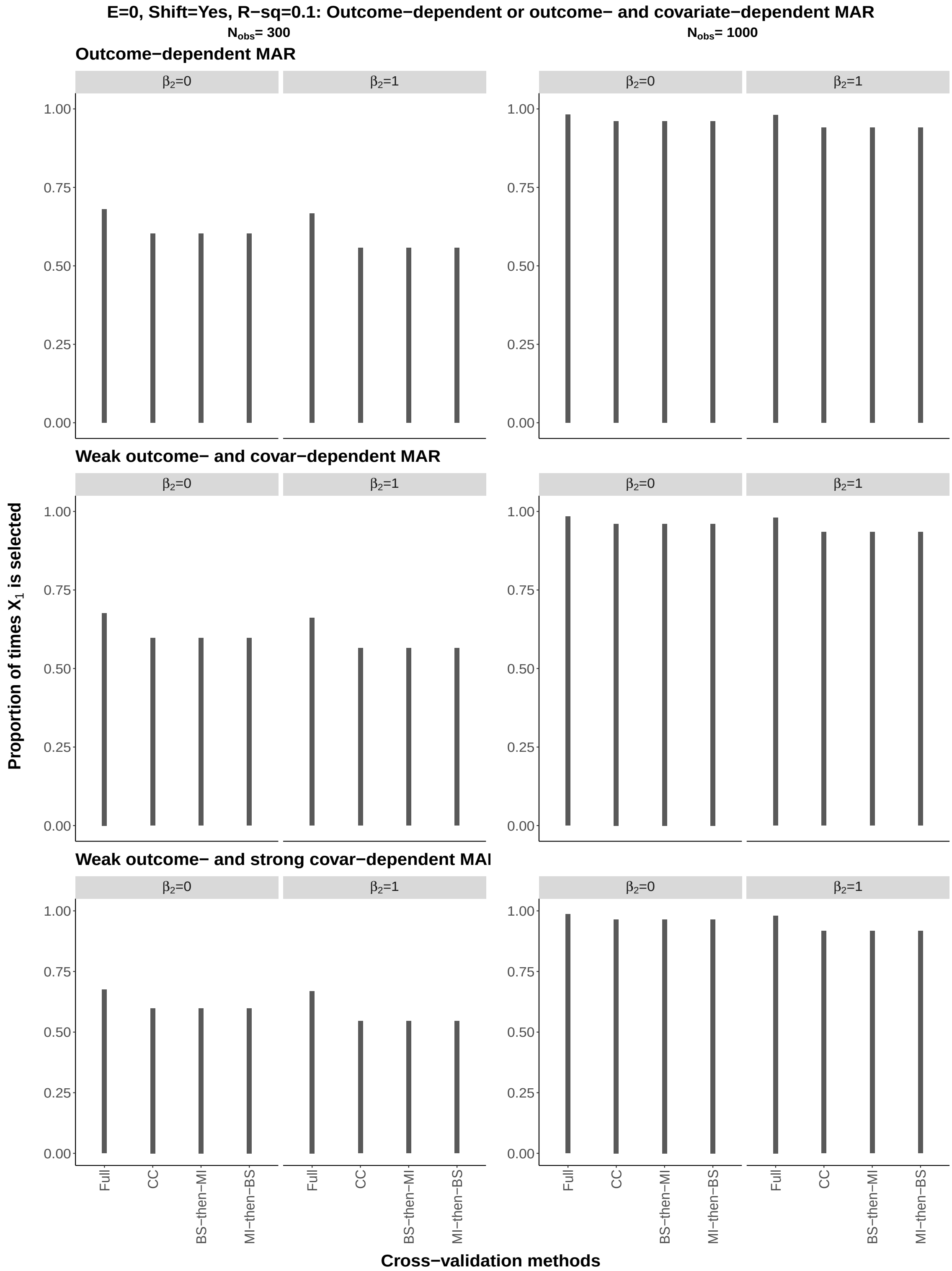

Figure S267: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

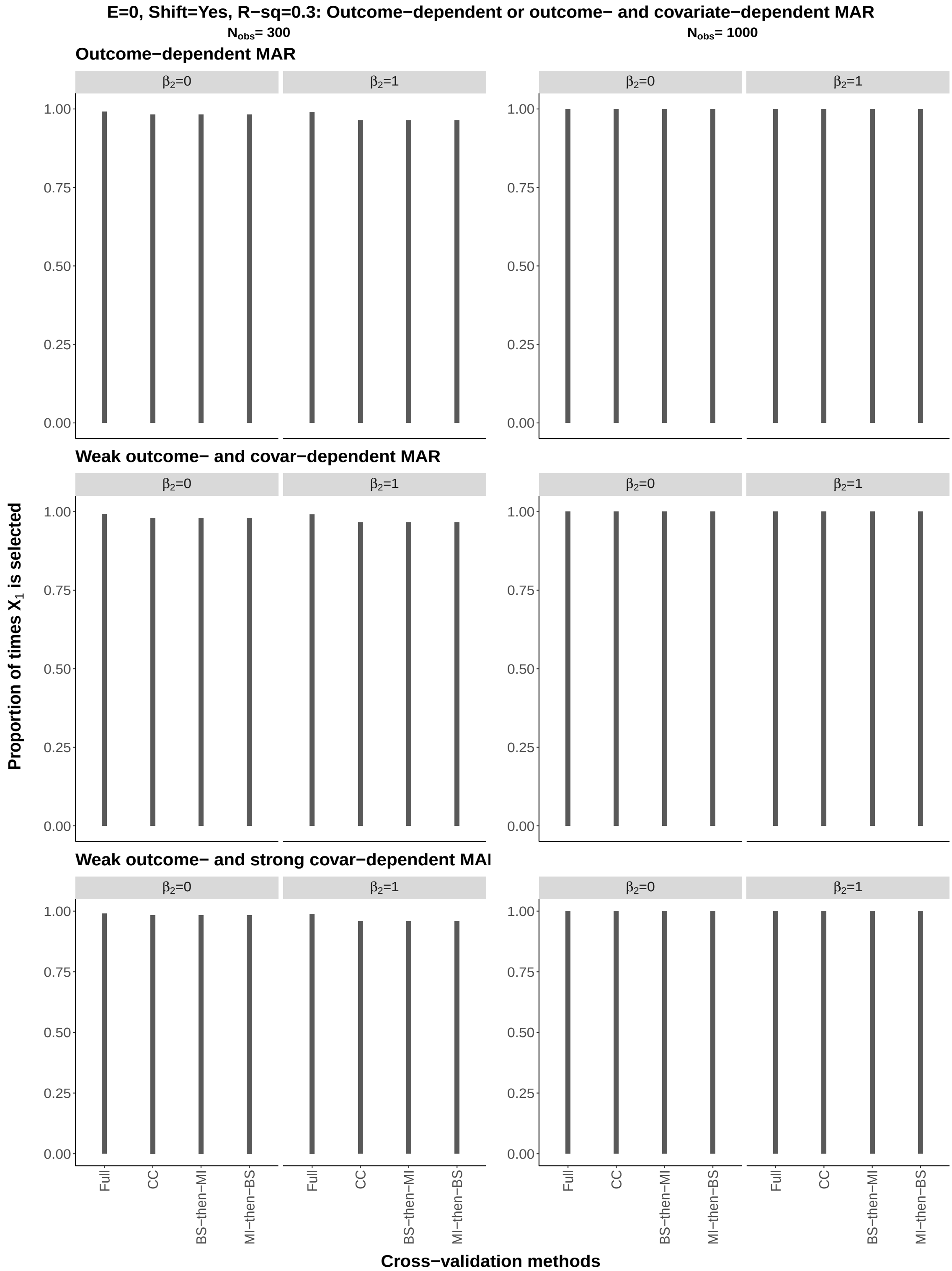

Figure S268: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

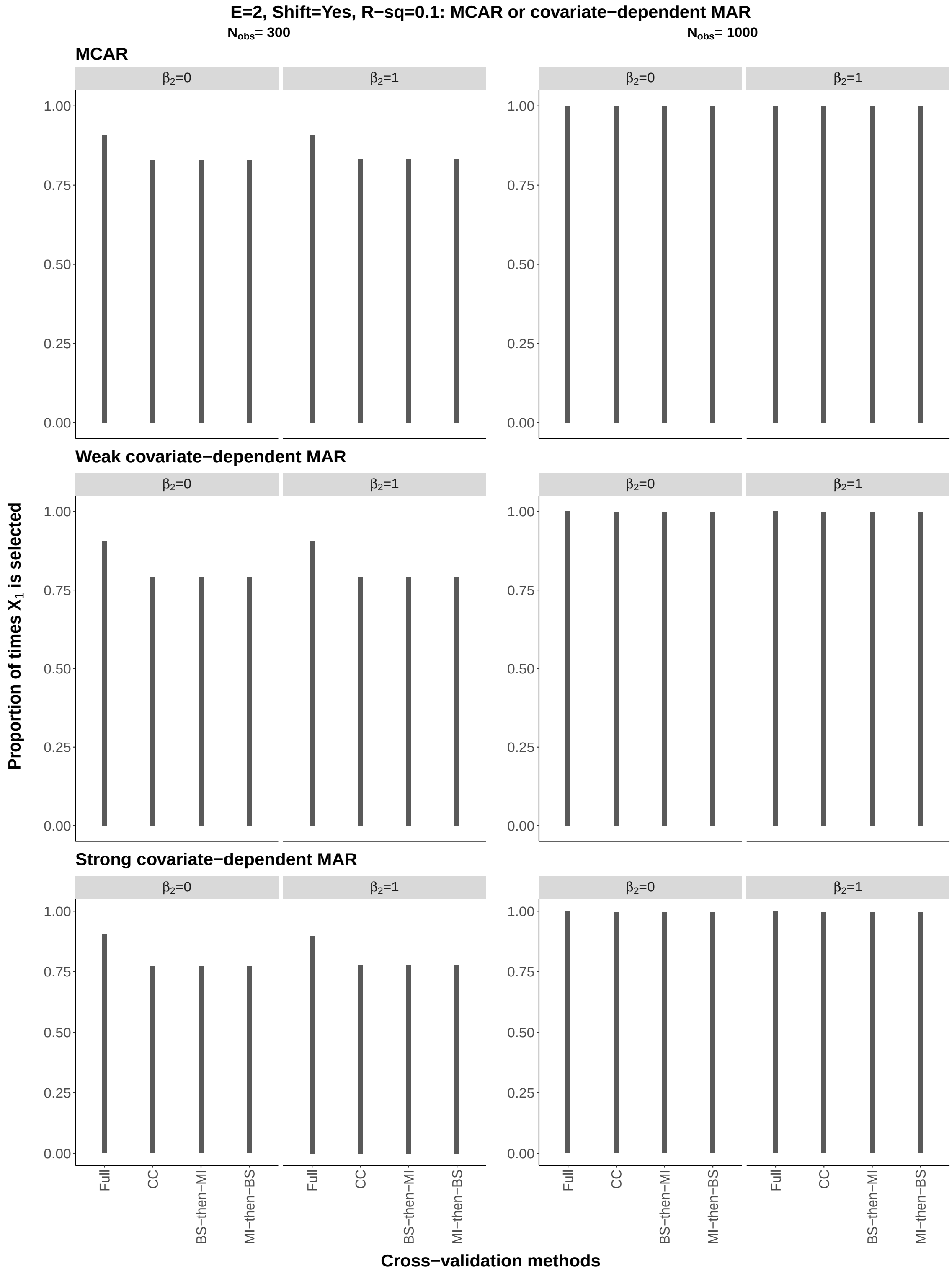

Figure S269: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

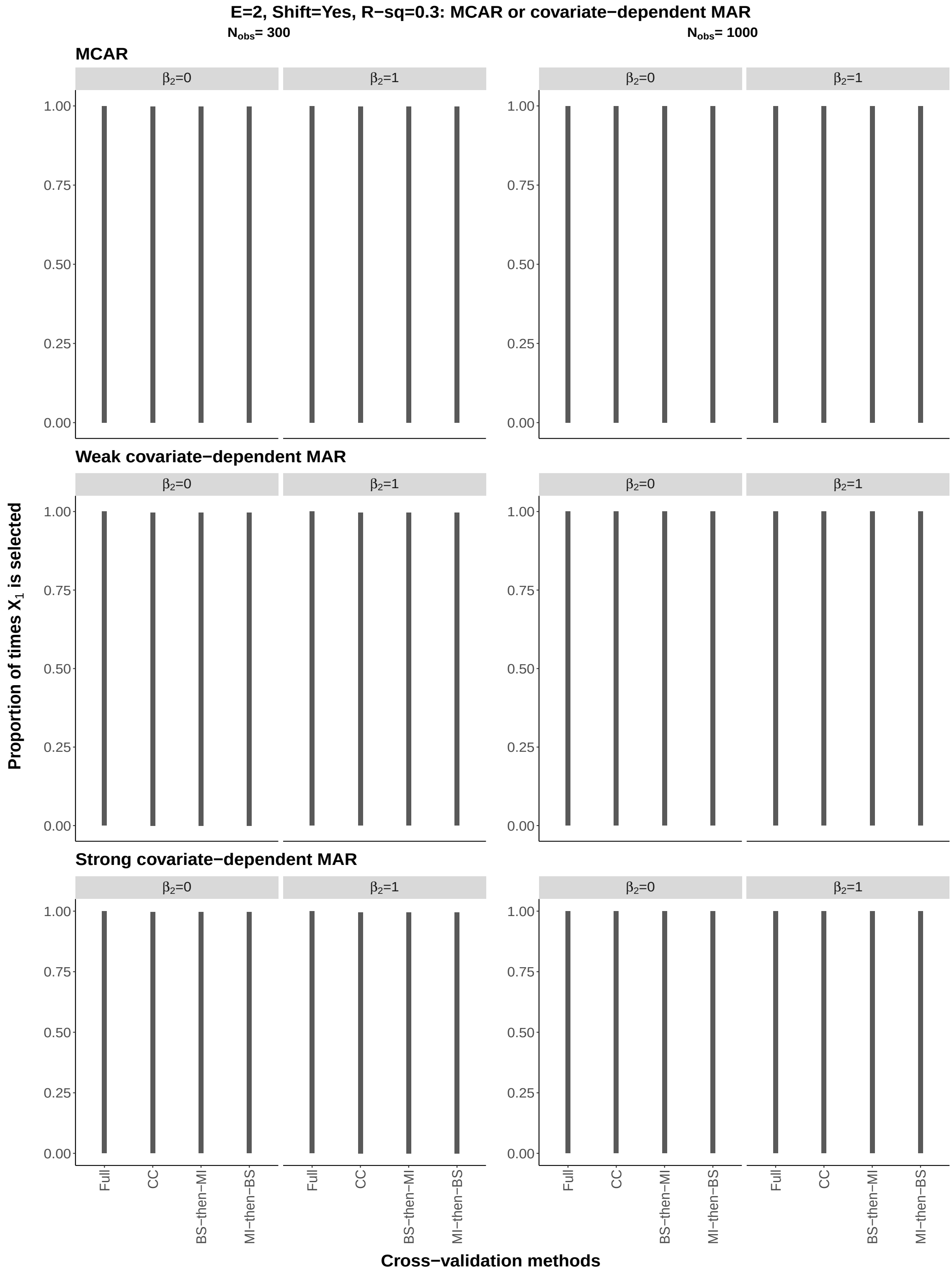

Figure S270: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

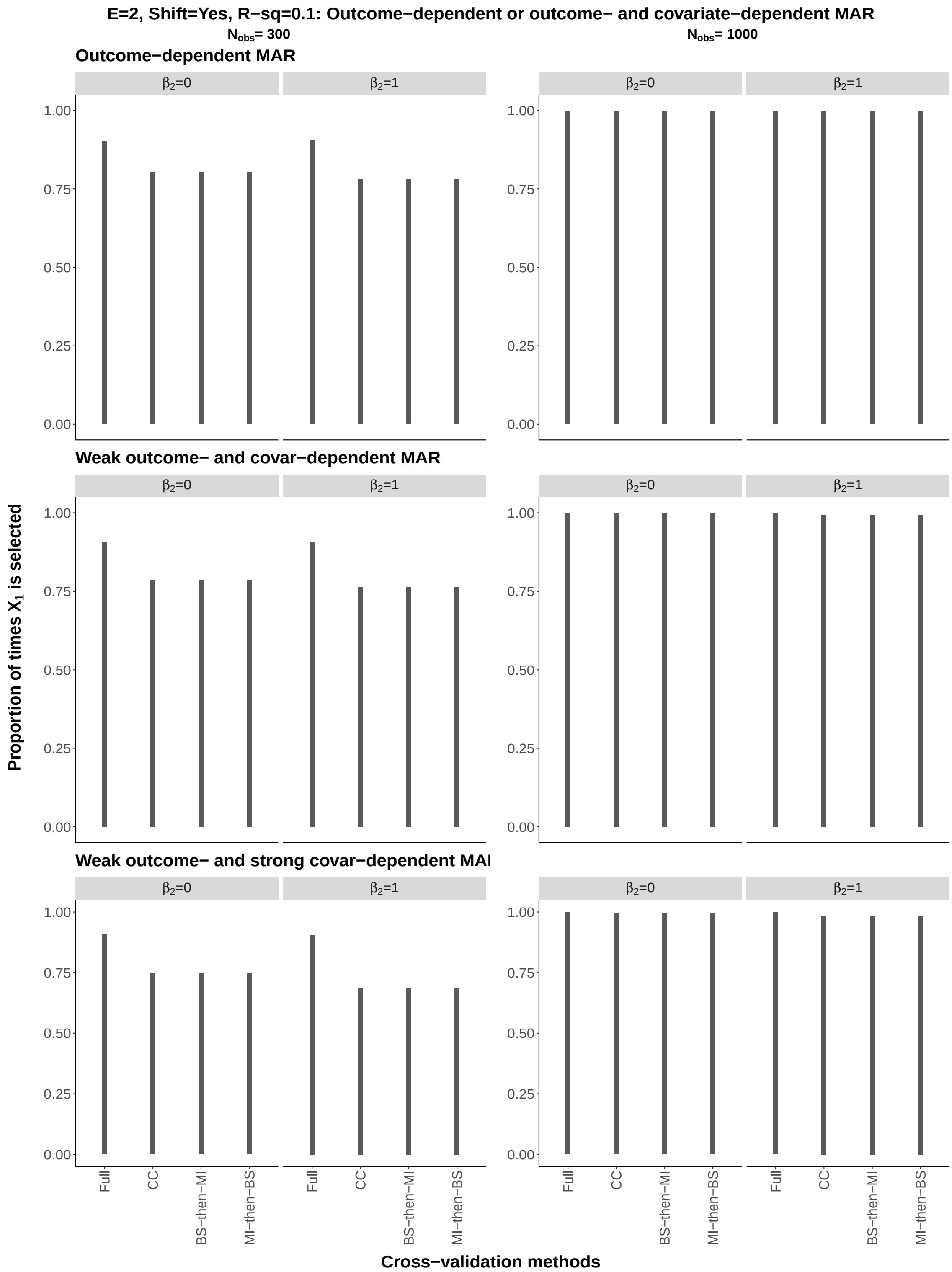

Figure S271: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

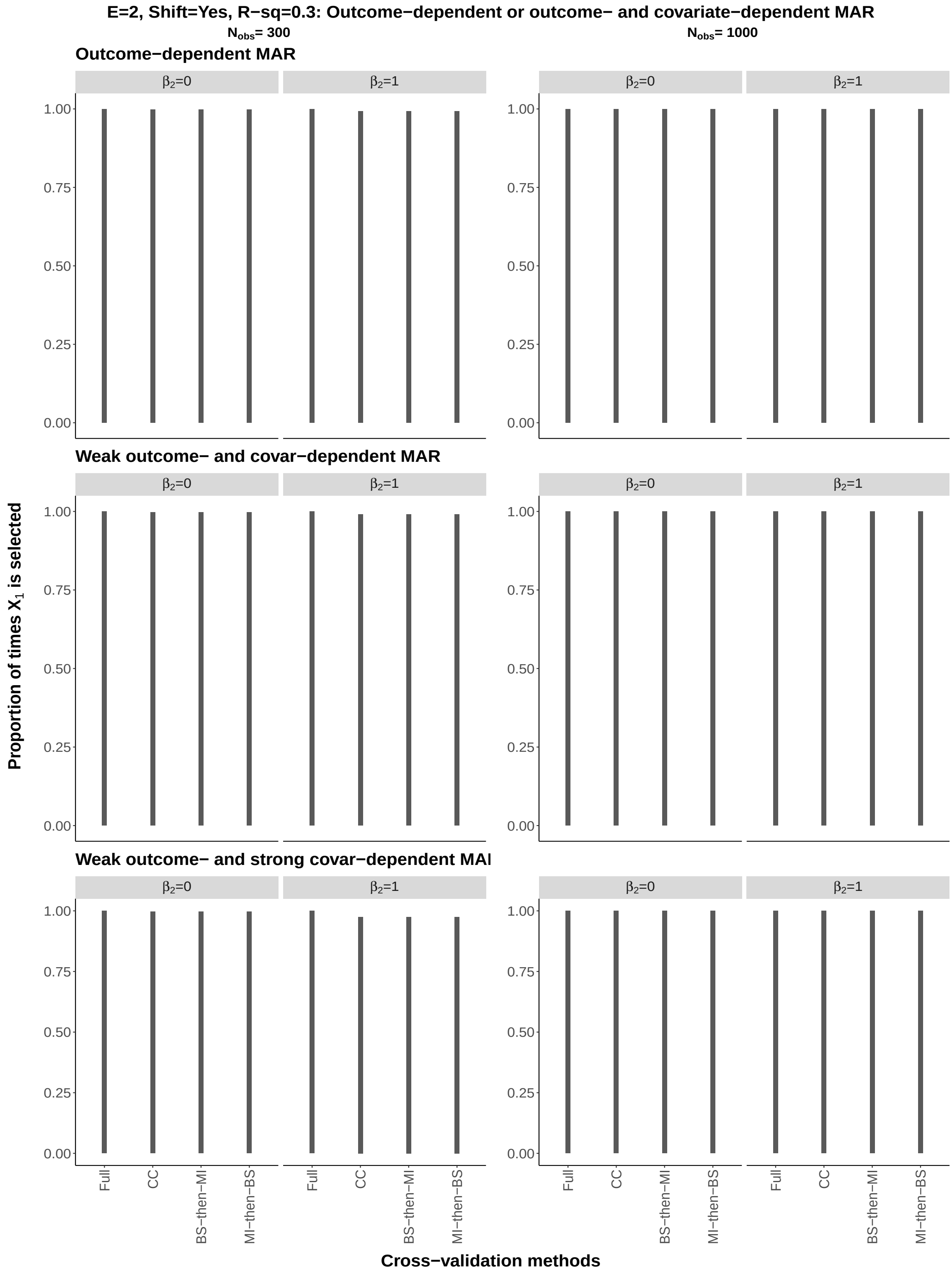

Figure S272: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

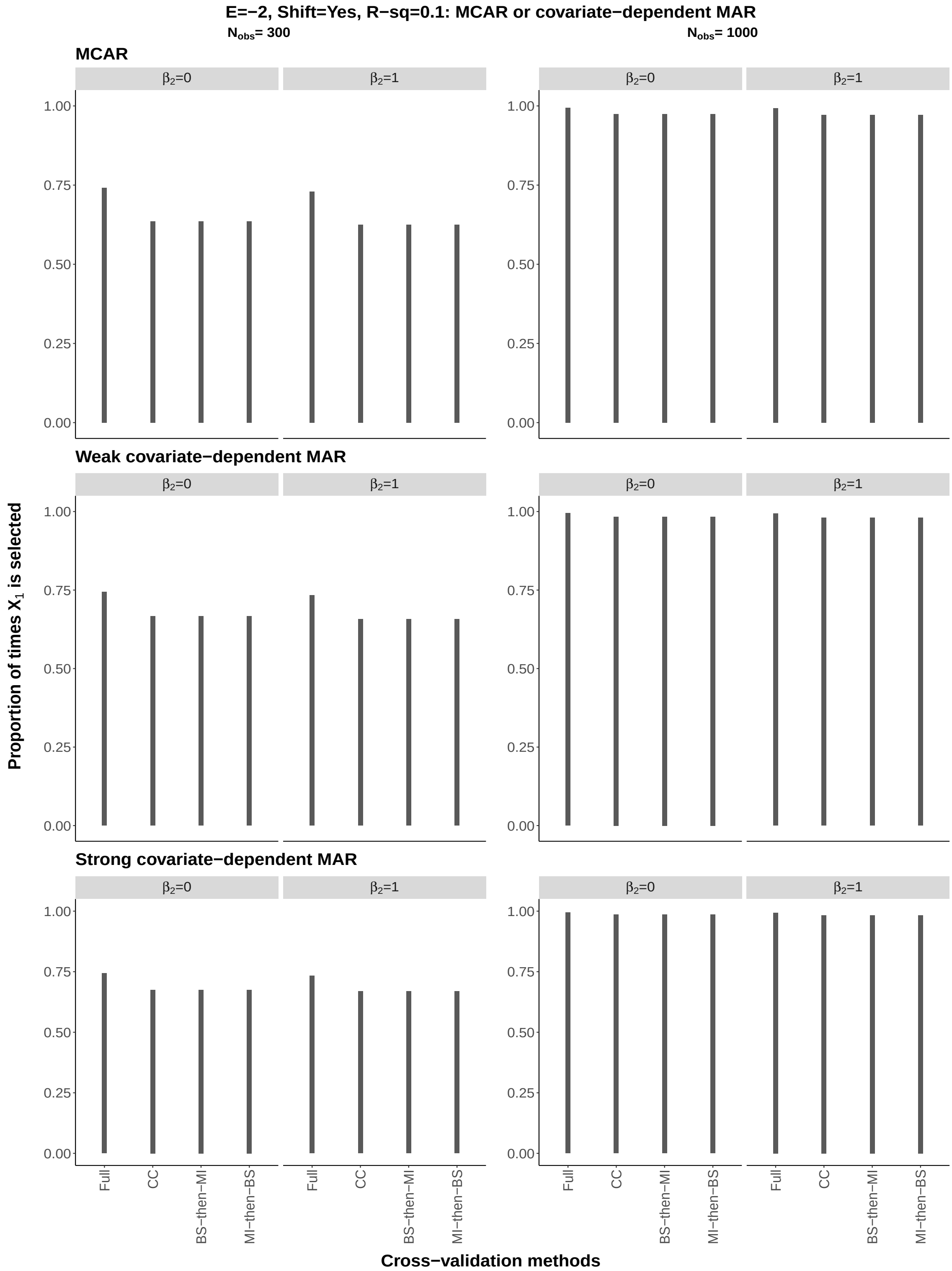

Figure S273: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

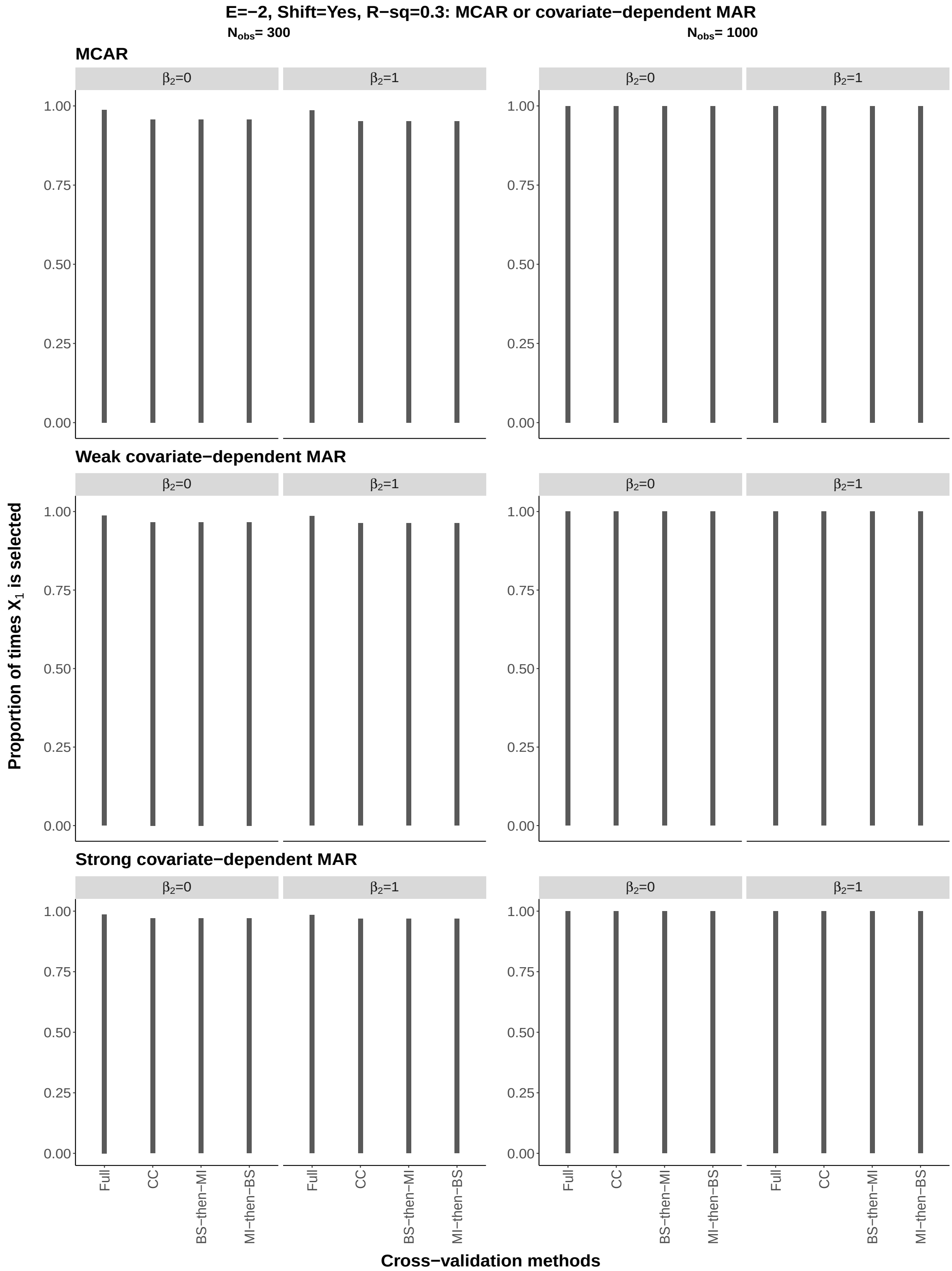

Figure S274: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

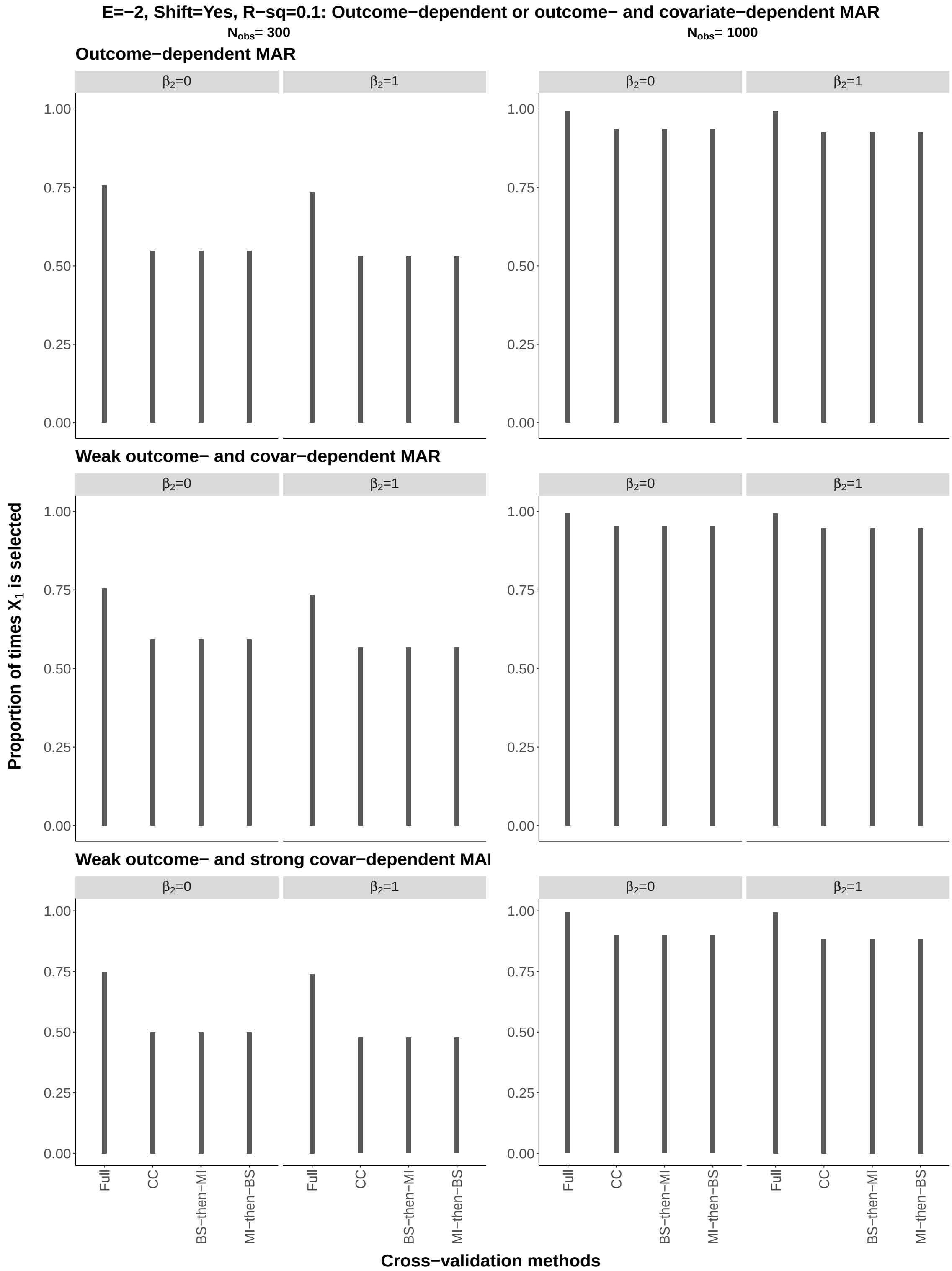

Figure S275: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

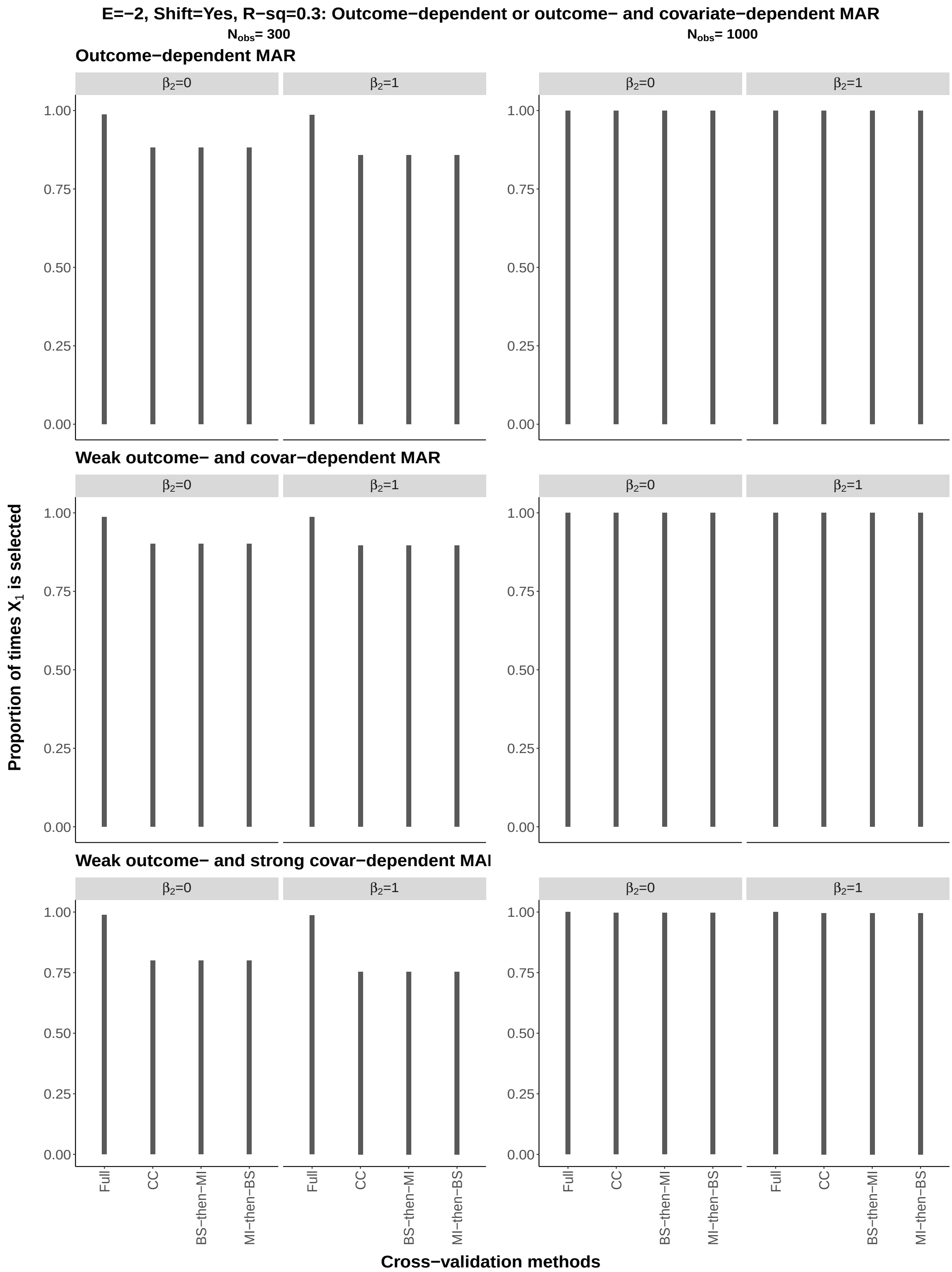

Figure S276: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 1$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

S9.2.16 Covariate selection of X_1 in the bootstrap samples: $\beta_2 = 1$, $\alpha_E = 0.05$ and an origin-shift has been applied

True exponent is 0

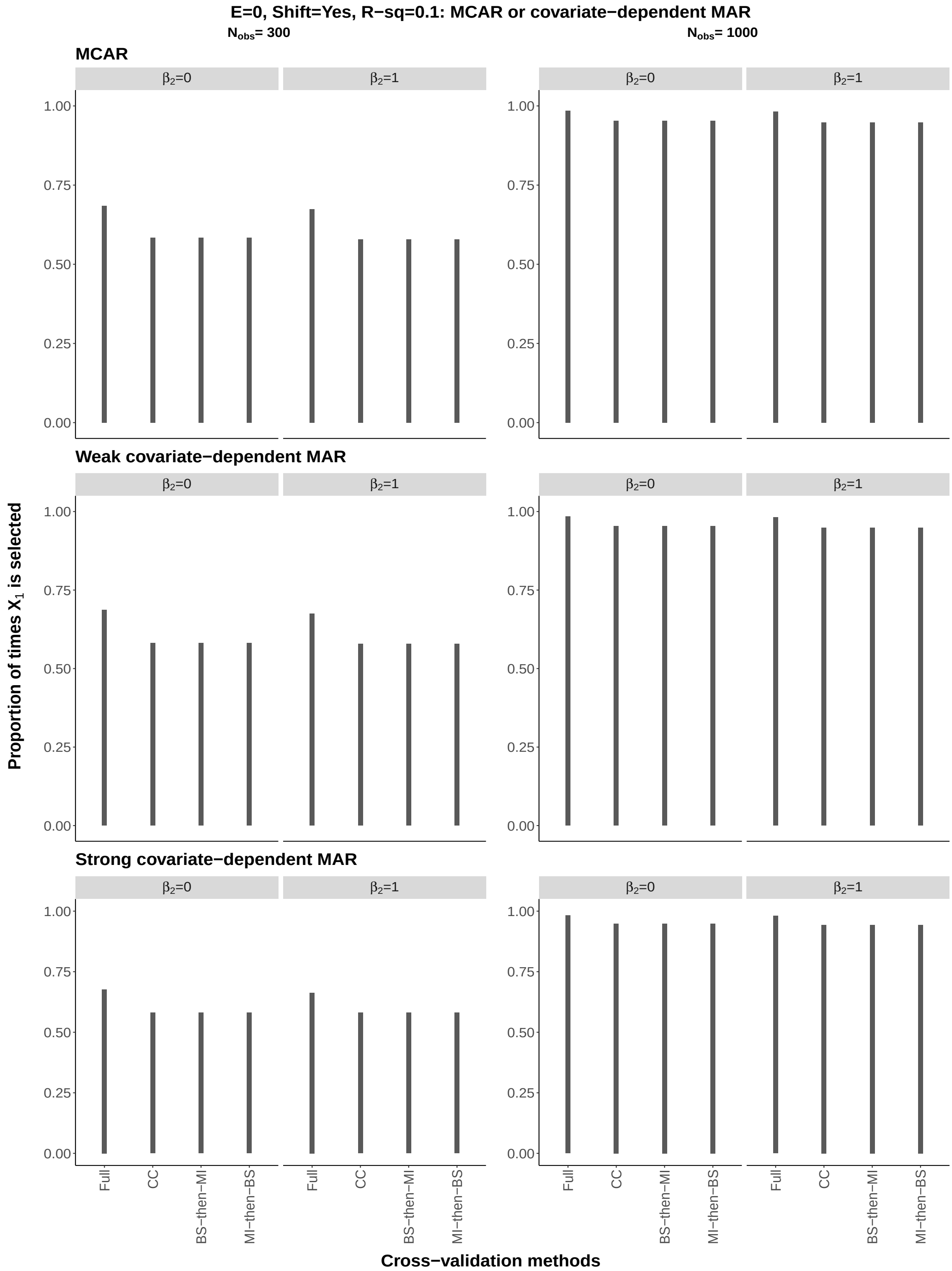

Figure S277: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

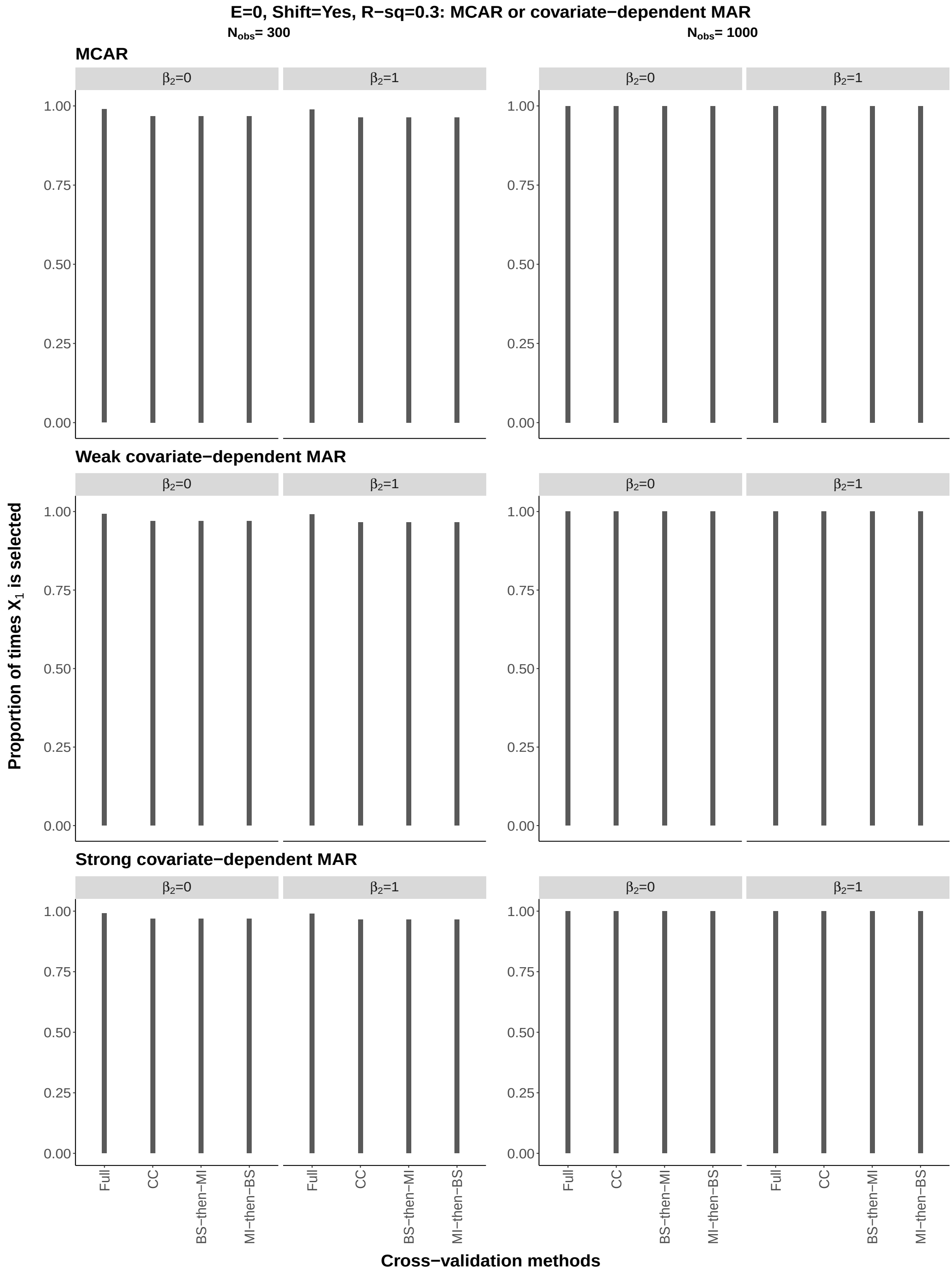

Figure S278: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

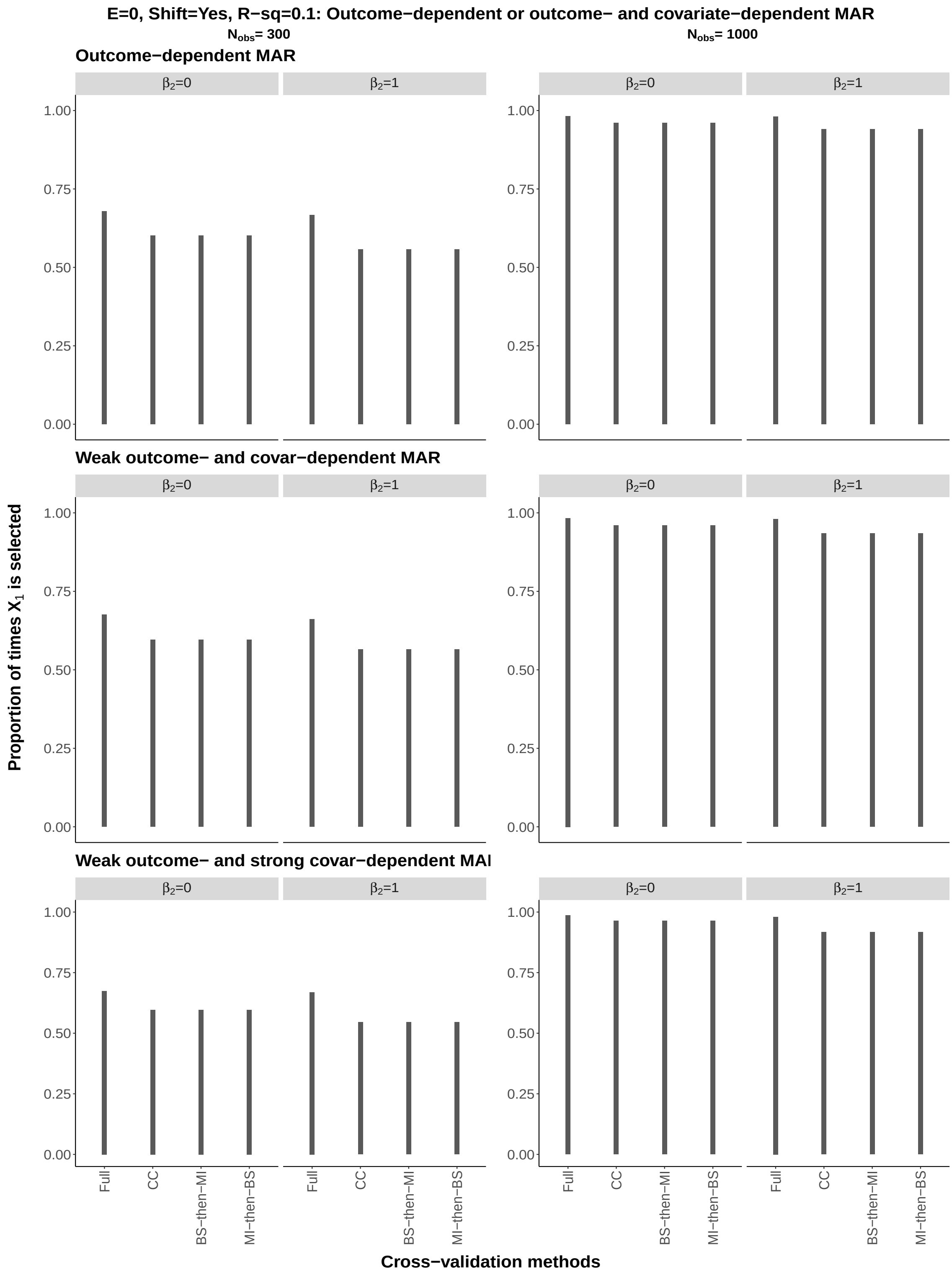

Figure S279: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

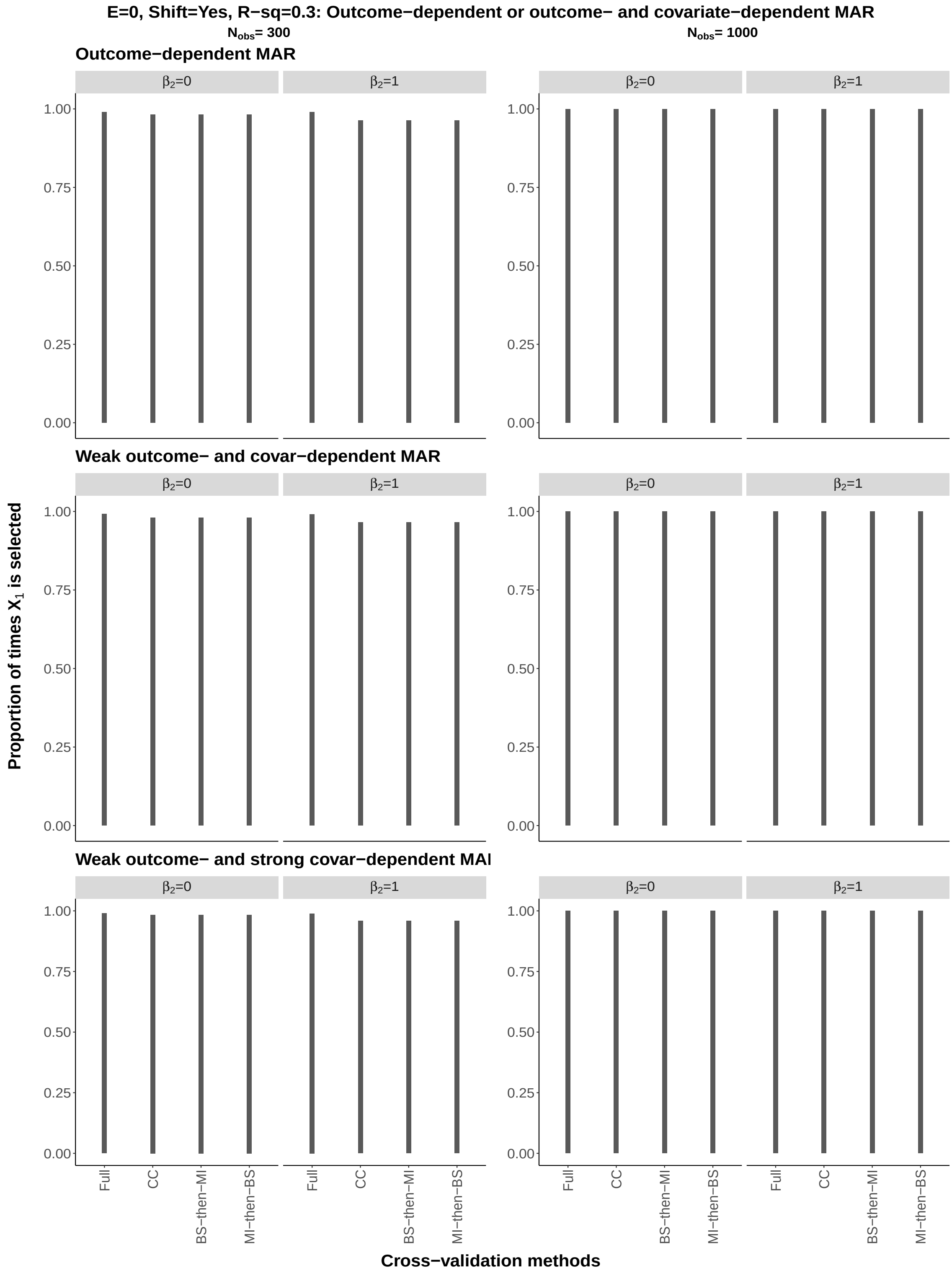

Figure S280: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 0 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is 2

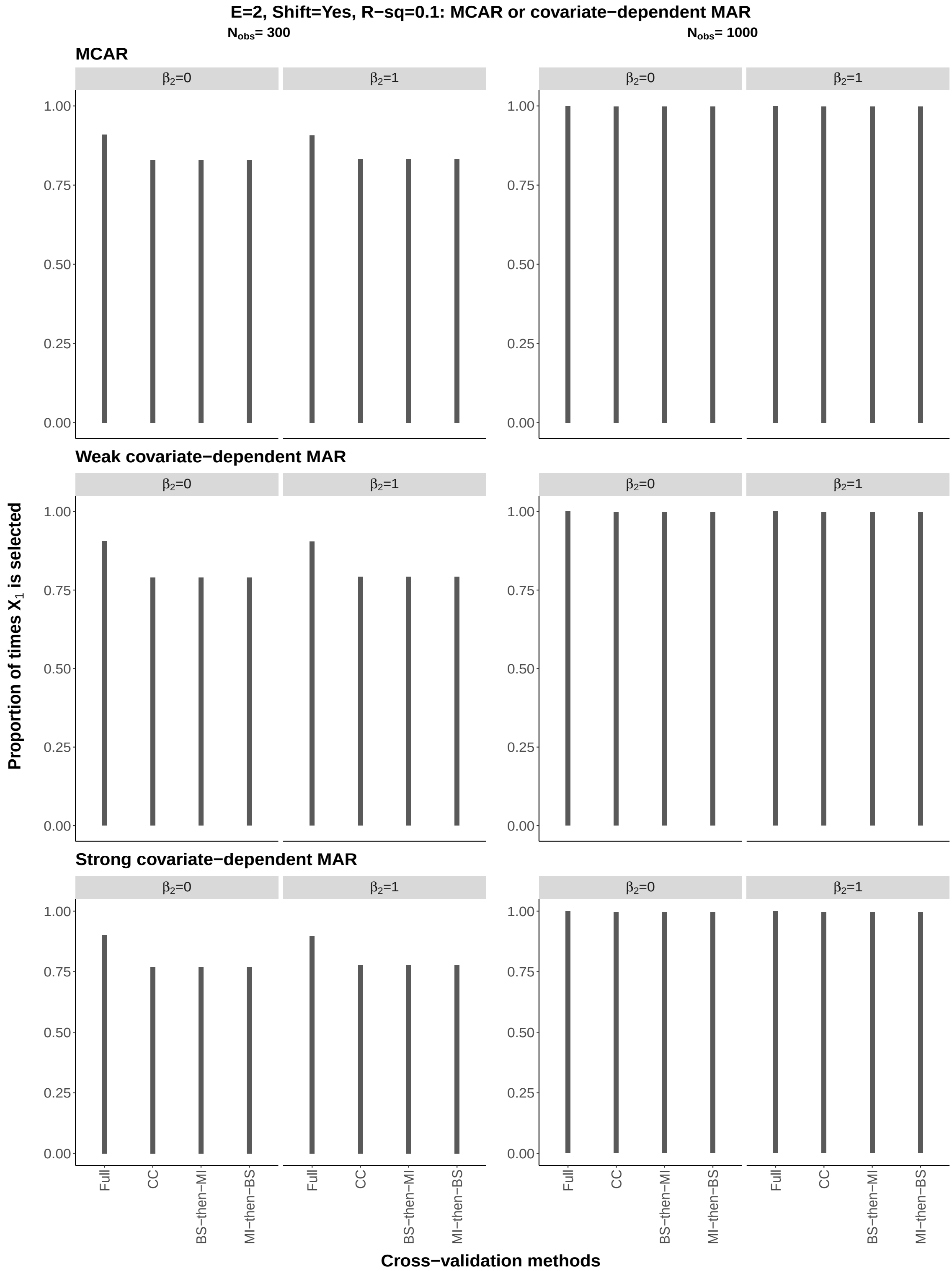

Figure S281: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

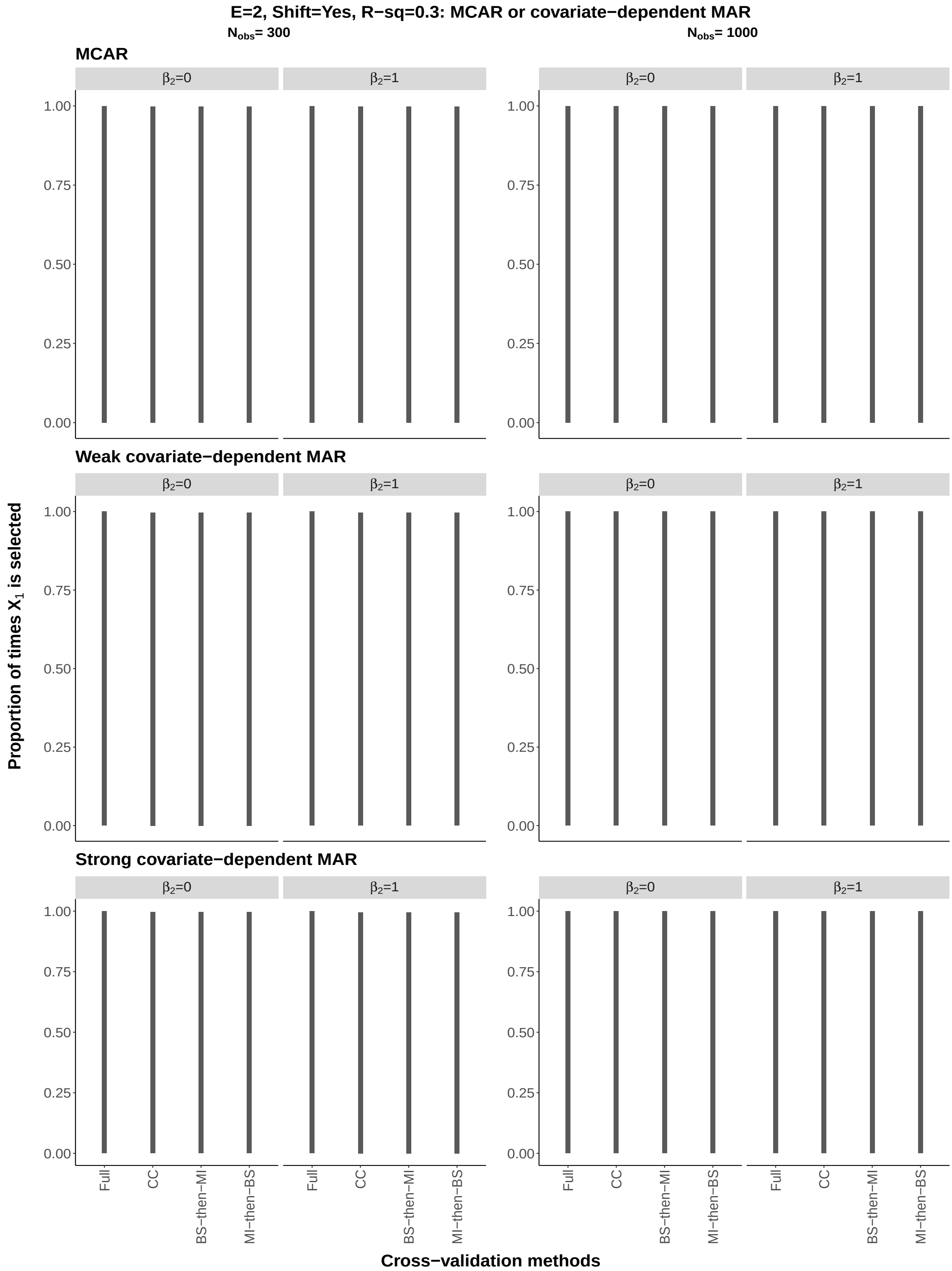

Figure S282: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

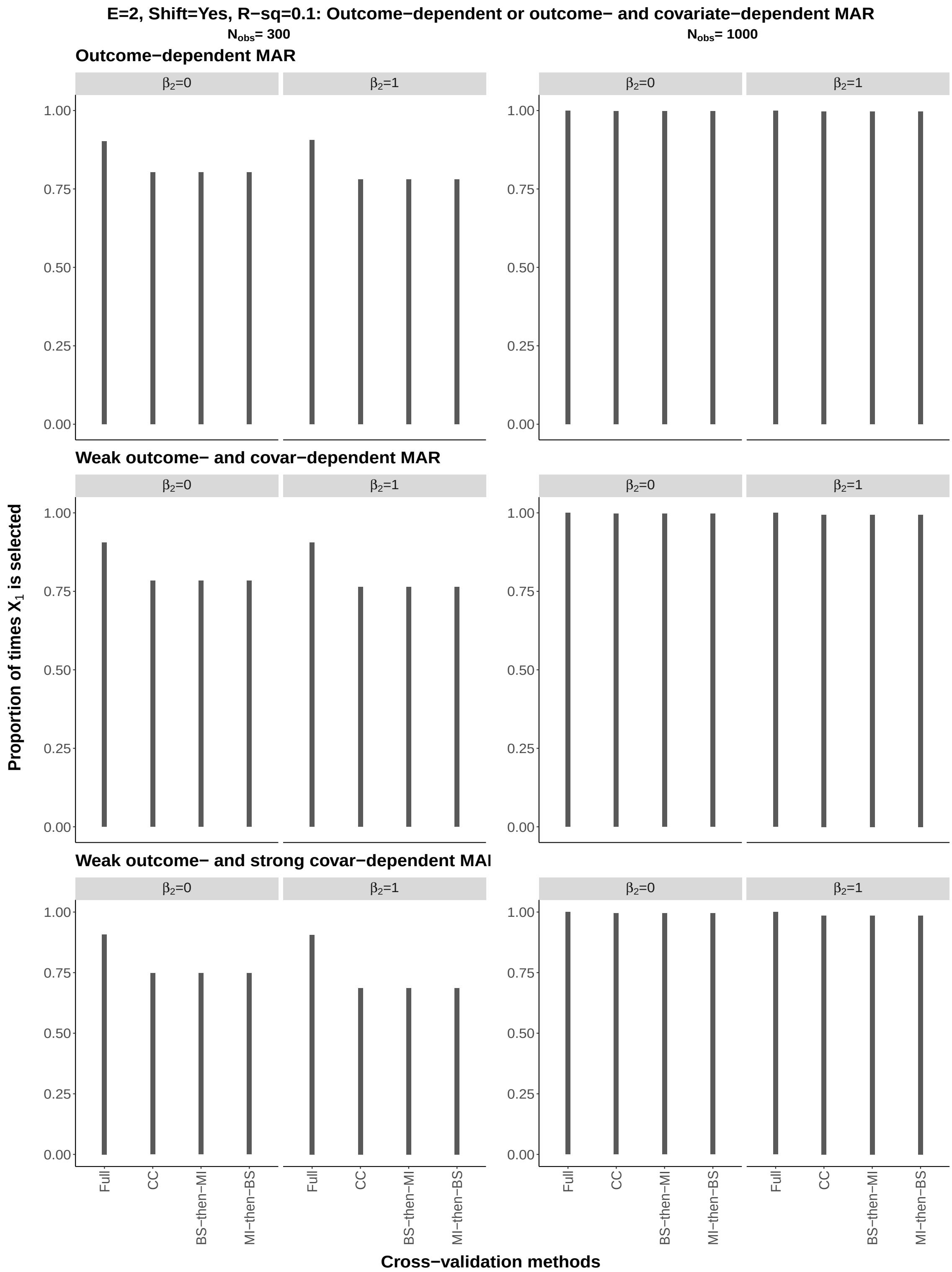

Figure S283: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

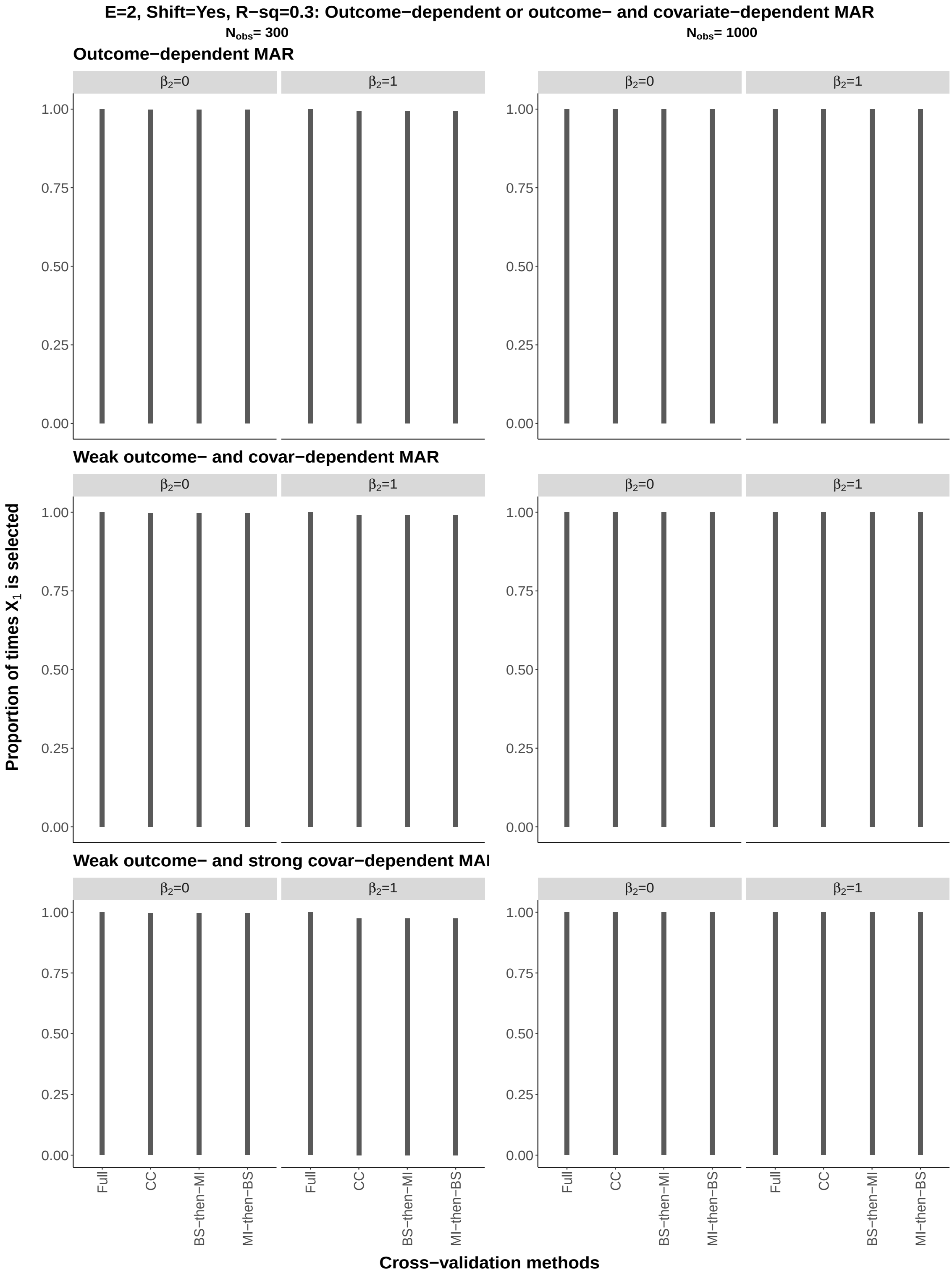

Figure S284: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is 2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

True exponent is -2

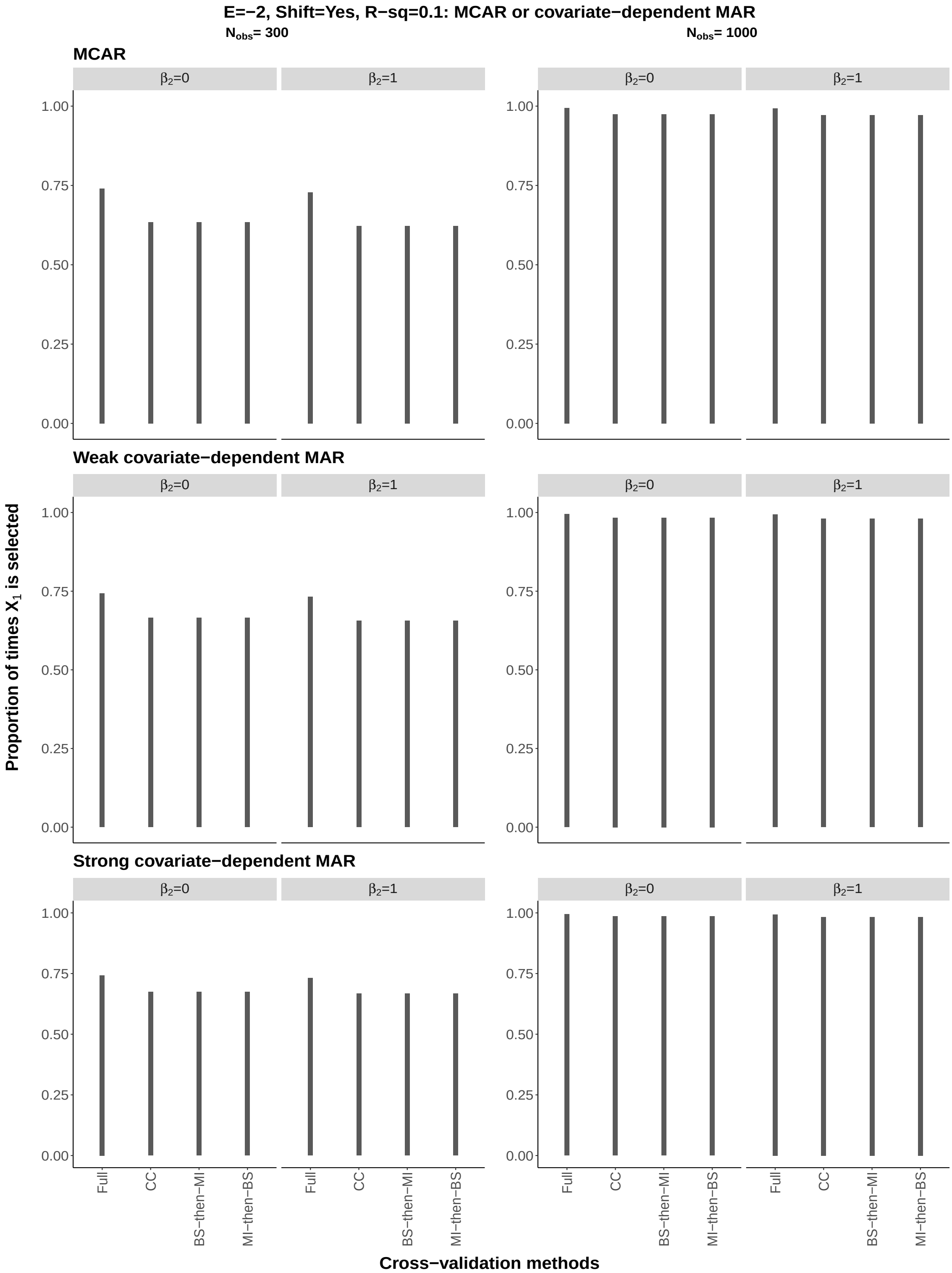

Figure S285: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

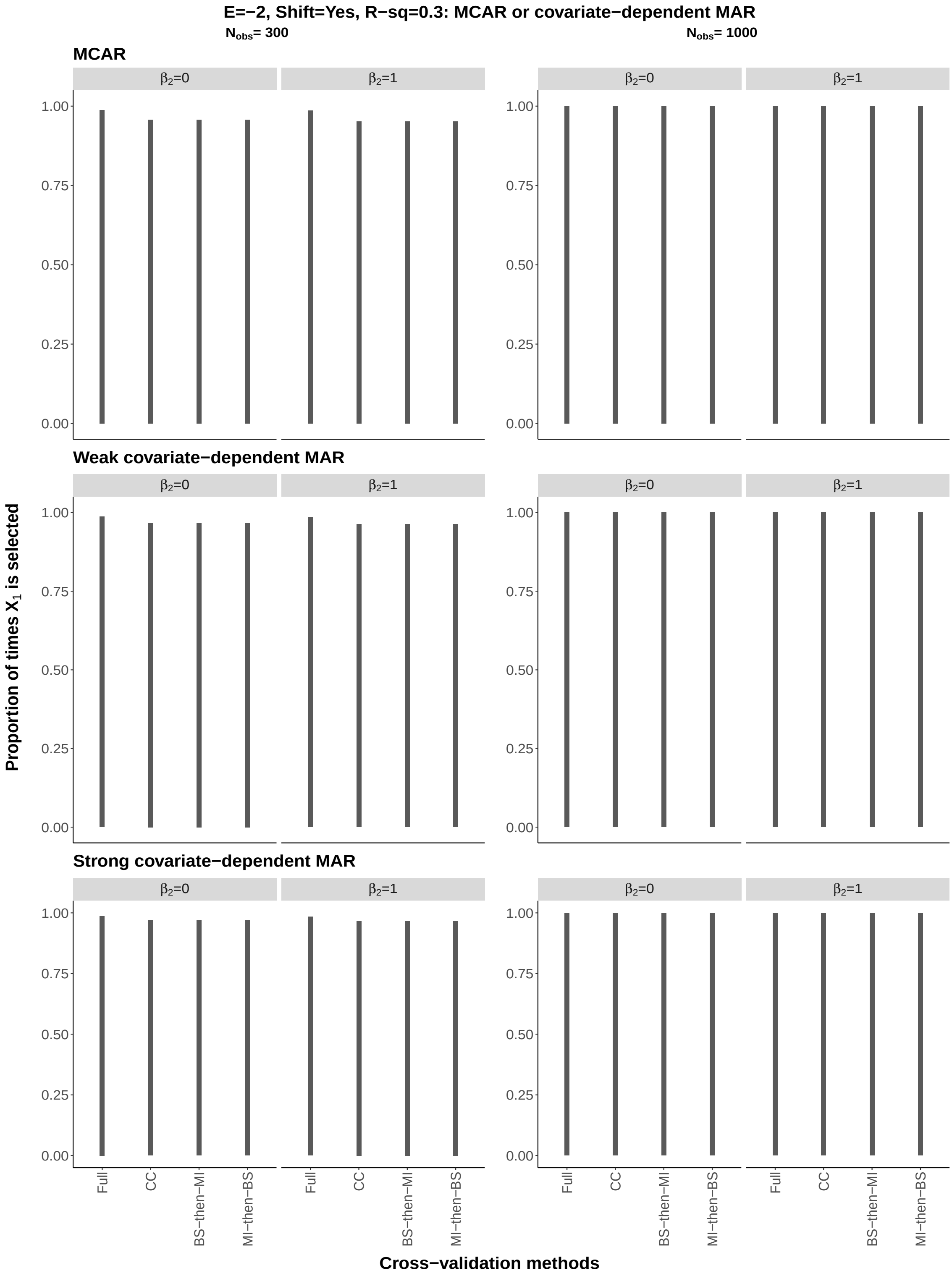

Figure S286: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are MCAR or covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

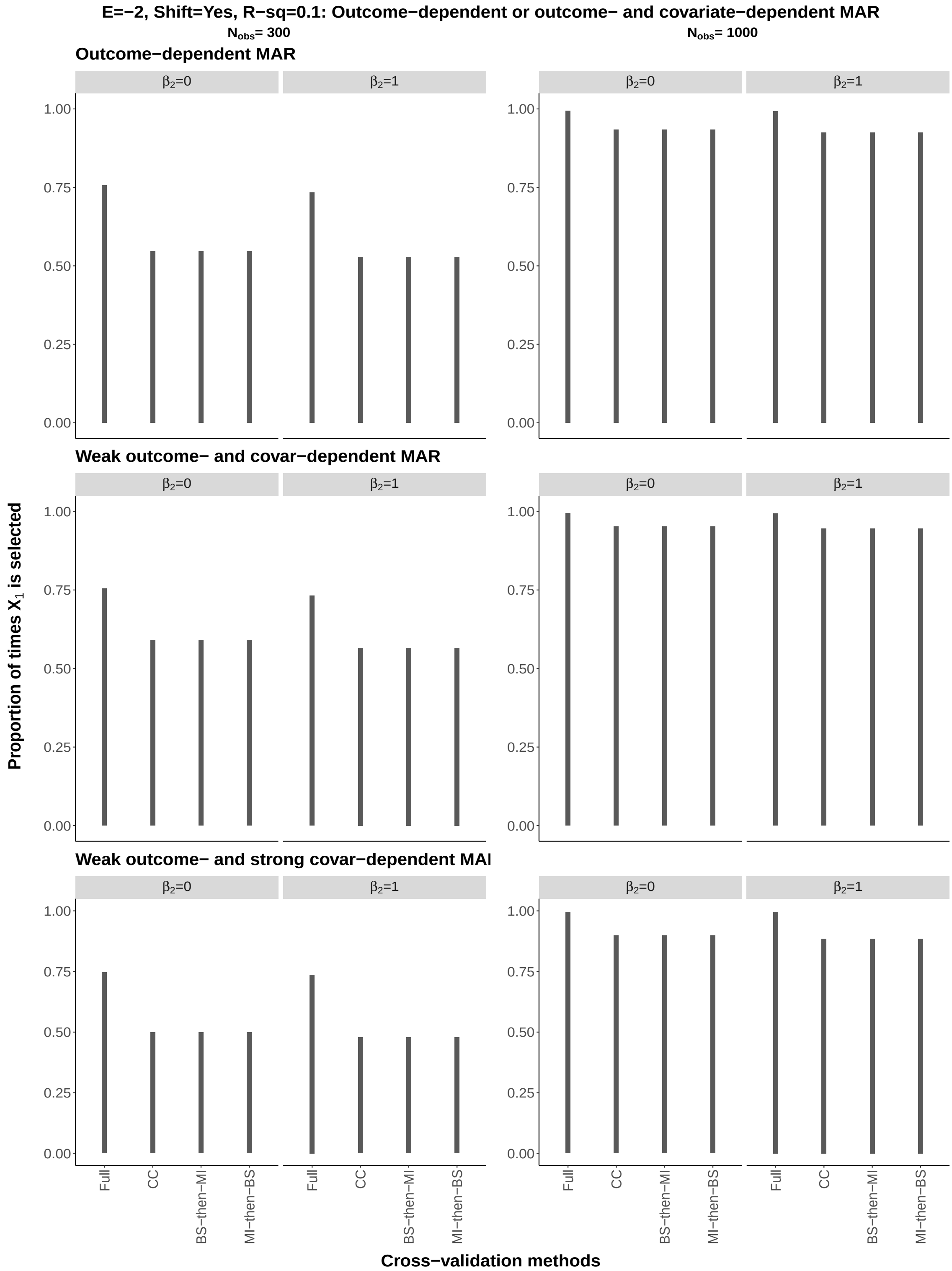

Figure S287: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.1$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.

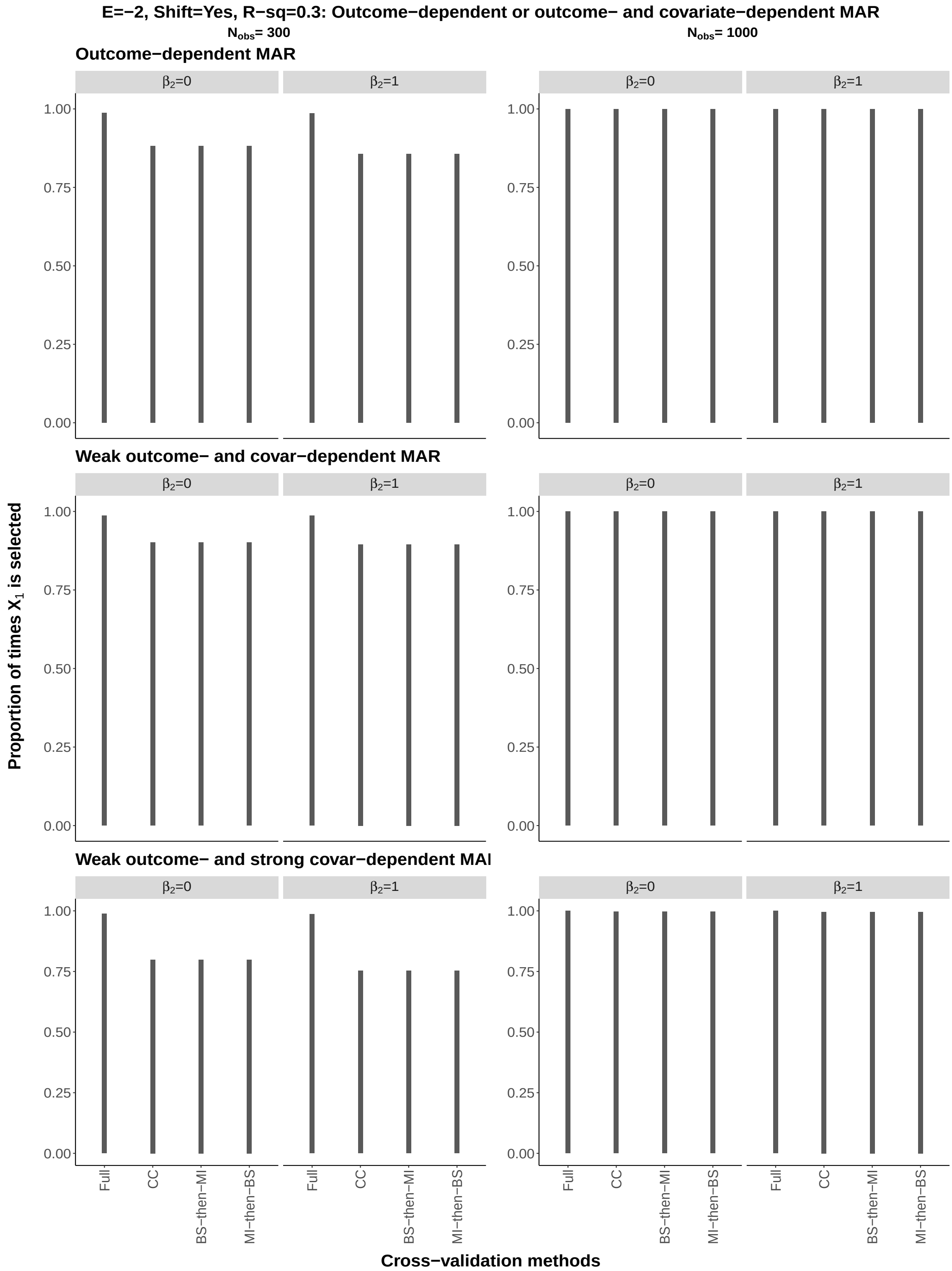

Figure S288: The proportion of times covariate X_1 is selected for inclusion to the prediction model using the MFP algorithm. The results presented are for the scenario where $\alpha_E = 0.05$ and an origin-shift transformation has been used. The results are for the scenario where data are outcome-dependent or outcome- and covariate-dependent MAR, the true exponent E is -2 and $R^2 = 0.3$. The black bars represent the proportion of times across the 2000 repetitions, B bootstrap samples and M imputed datasets, that X_1 was selected into a prediction model when its parameter, β_2 is 0 or 1. CC (complete-case); CV and bootstrap methods are described in Sections 7.4 and 7.5.