

LONDON  
SCHOOL *of*  
HYGIENE  
& TROPICAL  
MEDICINE



Computational Methods for D-optimal Design  
in Nonlinear Mixed Effects Models

Ketil Biering Tvermosegaard

Thesis submitted in accordance with the requirements for the degree of  
Doctor of Philosophy of the  
University of London

June 2020

Department of Non-communicable Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the European Community's Seventh Framework programme FP7/2011

# Abstract

This thesis aims to make a first step towards a foundation for a different, more practical approach to employing the principles of optimal experimental design in nonlinear mixed effects models. As an alternative to approaches which aim to mathematically account for parameter uncertainty and misspecification, it is proposed that the “space of possible parameter guesses” is investigated more directly, by visualising the resulting optimal designs and their relative performance. To provide some justification for the computational choices made in the packages, the thesis provides a comparison of two linearisation-based approaches to approximating the Fisher information matrix (First Order and Integrated First Order), a necessary step in computing  $D$ -optimality objective functions. This comparison is performed by utilising an approximation (Monte Carlo / Adaptive Gaussian Quadrature) which is not based on linearisation and which theoretically allows arbitrarily low error but at a high computational cost. It is concluded that the computationally cheaper First Order approximation is likely to be superior in all cases. A number of models taken from the applied and theoretical literature are introduced. Through these examples, it is shown how one can use the R-packages developed for this thesis (`doptim` and `randon`) to check robustness of proposed designs against parameter misspecification, in terms of information lost. A gentle introduction to using the packages is also provided, and it is demonstrated how to find  $D$ -,  $D_S$ - and  $D_A$ -optimal designs for nonlinear mixed effects models and, because the objective functions are made available to the user, how custom objective functions such as compound objective functions can also be generated and optimised.

# Acknowledgements

I am deeply grateful for the support and guidance of my supervisors Dave Woods, Sebastian Funk and John Whittaker. My office mates at LSHTM and my colleagues at GSK have provided invaluable boosts to motivation and morale. In particular, I owe thanks to Sean Cavany, Frank Sandmann and Matteo Quartagno as well as Steven Barrett, Nicholas Galwey, David Wille and Mark Lennon.

More than anyone, thanks to Pravina for her heart, her wisdom and her partnership.

# Contents

Introduction	25
1 D-optimal Design for Nonlinear Mixed Effects Models - Choosing Computational Approach	29
1.1 Introduction	31
1.2 FIM Approximation	39
1.3 Cases to be examined	45
1.4 Demonstration of the Gold Standard	49
1.5 Strategy for comparing FO and InFO	55
1.6 Finding <i>D</i> -optimal designs	58
1.7 Comparing FO and InFO <i>D</i> -Optimal Designs	64
1.8 Discussion	70
1.9 Further Work	74
2 A Practical Workflow Based On D-optimal Design	76
2.1 Introduction	77
2.2 Contents of this chapter	79
2.3 Methodological Background	80
2.4 Motivating Example: One-Compartment Model	84
2.5 Other Models	101
2.6 Discussion	126
3 R-packages: doptim and randon	130
3.1 Introduction to new R-packages doptim and randon	131
3.2 Example: Linear Regression with random intercept and slope	134
3.3 Example: exponential decay - first nonlinear example	140
3.4 Replicating Wang2012	144
3.5 Overview of the randon package	151
Conclusion	162



# List of Figures

Figure 1.1	Illustration of the MC/AGQ approximation of the FIM.	53
Figure 1.2	$\log\det(\text{FIM}_{\text{MC/AGQ}})$ as a (restricted) function of $\zeta$	54
Figure 1.3	Distribution of MC/AGQ $\log\det(\text{FIM})$	56
Figure 1.4	Distribution of $\log\det(\text{FIM}_{\text{FO}})$ and $\log\det(\text{FIM}_{\text{InFO}})$	57
Figure 1.5	InFO $D$ -optimality objective function	60
Figure 1.6	Design graph of FO $D$ -optimal designs	62
Figure 1.7	FO and InFO $D$ -optimal designs for all models	63
Figure 1.8	Comparing FO and InFO $D$ -optimal designs	71
Figure 2.1	Sensitivity bands for the One-Compartment model.	87
Figure 2.2	Design graphs for the One-Compartment model.	89
Figure 2.3	A single efficiency profile for the One-Compartment model	91
Figure 2.4	Efficiency profiles for the designs in Figure 2.2.	92
Figure 2.5	$D_s$ -optimal designs for the One-Compartment model.	95
Figure 2.6	Efficiency profiles for the $D_s$ -optimal designs in Figure 2.5.	96
Figure 2.7	$D_A$ -optimal designs for the One-Compartment model.	99
Figure 2.8	Efficiency profiles for the designs in Figure 2.7.	100

Figure 2.9 Design graphs generated from the Michaelis-Menten model taken from Bates and Watts (1988), Appendix A1.3. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (213, 0.0641)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0.02, 1.1]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . .106

Figure 2.10 Efficiency profiles generated from the Michaelis-Menten model taken from Bates and Watts (1988), Appendix A1.3. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (213, 0.0641)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0.02, 1.1]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.9. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.9) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .107



Figure 2.11 Design graphs generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.4. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(1 - \exp(\beta_2 t))$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (19.1, -0.531)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [1, 7]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . .108

Figure 2.12 Efficiency profiles generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.4. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(1 - \exp(\beta_2 t))$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (19.1, -0.531)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [1, 7]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.11. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.11) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .109

Figure 2.13 Design graphs generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.12. The regression function is  $\eta(\beta, t) = \beta_1 t (\beta_2 + t + \beta_3 t^2)^{-1}$  and the fixed effects parameter vector is  $\beta = (89800, 187, 0.0163)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \eta(\beta + \mathbf{b}_i, \zeta) + \epsilon_i$ . Here, it is assumed that  $\zeta \in [2.2, 170]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\epsilon_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$ ,  $\omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . . .110

Figure 2.14 Efficiency profiles generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.12. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 t (\beta_2 + t + \beta_3 t^2)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (89800, 187, 0.0163)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \zeta) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\zeta \in [2.2, 170]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$ ,  $\omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.13. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.13) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . .111

Figure 2.15 Design graphs generated from the compartmental gas adsorption model taken from Kitsos and Kolovos (2013), example (25). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 \beta_2 t \left( [1 - t][1 + (\beta_2 - 1)t] \right)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (2, 10)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0.05, 0.3]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . . .112

Figure 2.16 Efficiency profiles generated from the compartmental gas adsorption model taken from Kitsos and Kolovos (2013), example (25). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 \beta_2 t \left( [1 - t][1 + (\beta_2 - 1)t] \right)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (2, 10)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0.05, 0.3]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.15. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.15) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . .113

Figure 2.17 Design graphs generated from the reaction network model taken from Kitsos and Kolovos (2013), example (28). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(\beta_1 + \beta_2 - \beta_3)^{-1} \left( \exp[-\beta_3 t] - \exp[-(\beta_1 + \beta_2)t] \right)$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{2})^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0, 2]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$ ,  $\omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . . .114

Figure 2.18 Efficiency profiles generated from the reaction network model taken from Kitsos and Kolovos (2013), example (28). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(\beta_1 + \beta_2 - \beta_3)^{-1} \left( \exp[-\beta_3 t] - \exp[-(\beta_1 + \beta_2)t] \right)$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{2})^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [0, 2]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.17. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.17) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .115



Figure 2.19 Design graphs generated from Sigmoid Emax ( $\gamma = 1$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . . .116

Figure 2.20 Efficiency profiles generated from Sigmoid Emax ( $\gamma = 1$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.19. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.19) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .117

Figure 2.21 Design graphs generated from Sigmoid Emax ( $\gamma = 3$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . . .118

Figure 2.22 Efficiency profiles generated from Sigmoid Emax ( $\gamma = 3$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.21. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.21) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .119

Figure 2.23 Design graphs generated from Sigmoid Emax ( $\gamma = 5$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . . .120

Figure 2.24 Efficiency profiles generated from Sigmoid Emax ( $\gamma = 5$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.23. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.23) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .121

Figure 2.25 Design graphs generated from the One-Compartment model found in row one of Table 2.1. The regression function is  $\eta(\boldsymbol{\beta}, t) = \mu_2 \left[ \mu_3(\mu_2 - \mu_1) \right]^{-1} \left[ \exp(-\mu_1 t) - \exp(-\mu_2 t) \right]$  with  $\mu_j = \exp(\beta_j)$  for  $j = 1, 2, 3$ . and the fixed effects parameter vector is  $\boldsymbol{\beta} = (0, 1, 1)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0, 4]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . .122

Figure 2.26 Efficiency profiles generated from the One-Compartment model found in row one of Table 2.1. The regression function is  $\eta(\boldsymbol{\beta}, t) = \mu_2 \left[ \mu_3(\mu_2 - \mu_1) \right]^{-1} \left[ \exp(-\mu_1 t) - \exp(-\mu_2 t) \right]$  with  $\mu_j = \exp(\beta_j)$  for  $j = 1, 2, 3$ . and the fixed effects parameter vector is  $\boldsymbol{\beta} = (0, 1, 1)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0, 4]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$ ,  $\omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.25. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.25) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .123

Figure 2.27 Design graphs generated from the tetracycline metabolism model found in Bates and Watts (1988), Appendix A1.14. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_3 \left( \exp[-\beta_1(t - \beta_4)] - \exp[-\beta_2(t - \beta_4)] \right)$ . and the fixed effects parameter vector is  $\boldsymbol{\beta} = (0.149, 0.716, 2.65, 0.412)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0, 4]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_4(0, \text{diag}\{\omega_1, \omega_2, \omega_3, \omega_4\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3, \omega_4$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the  $x$ -axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the  $y$ -axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the  $y$ -axis. . . . .124

Figure 2.28 Efficiency profiles generated from the tetracycline metabolism model found in Bates and Watts (1988), Appendix A1.14. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_3 \left( \exp[-\beta_1(t - \beta_4)] - \exp[-\beta_2(t - \beta_4)] \right)$ . and the fixed effects parameter vector is  $\boldsymbol{\beta} = (0.149, 0.716, 2.65, 0.412)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [1, 16]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_4(0, \text{diag}\{\omega_1, \omega_2, \omega_3, \omega_4\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3, \omega_4$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.27. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the  $x$ -axis value indicates the  $D$ -optimal design which is being used (this can be located at the corresponding  $x$ -axis value in Figure 2.27) and the  $y$ -value is the relative FO  $D$ -efficiency of the chosen design relative to the FO  $D$ -optimal design corresponding to the true variance level. . . . . .125

Figure 3.1 A simple version of the design graphs found in earlier chapters. The plotting symbol is the index of the sampling time in the individual  $D_\beta$ -optimal design being plotted and the  $x$ -axis is the variance magnitude used to generate the design. Note that overplotting occurs, with time labels "1" through "4" being consistently placed on top of each other and time label "6" being on top of "5" for some of the variance values. . . . . .143

Figure 3.2	Profile objective functions for the One-Compartment model example. The hollow circles are the coordinates (time and objective function value) of the final sampling times of the optimal designs. The lines are the profiles of objective functions for the range of variance levels considered. The values of $i$ which indicate which of the $i = 1, \dots, 40$ variance levels is being used have been added to the end of the corresponding profiles. . . . .	.159
Figure 3.3	Efficiency profiles for the One-Compartment model example. Each profile is a plot of the $D$ -efficiency of a design which uses the first four sampling times of the $D$ -optimal design and then replaces the fifth sampling time with the value on the $x$ -axis. The values of $i$ which indicate which of the $i = 1, \dots, 40$ variance levels is being used have been added to the end of the corresponding profiles. . . . .	.160

## List of Tables

Table 0.1	Acronyms and symbols used in this thesis. . . . .	27
Table 1.1	Models considered in this chapter. . . . .	50
Table 1.2	Computation times. . . . .	69
Table 2.1	Other Models . . . . .	.105



# Introduction

THIS work focuses on optimal experimental design for nonlinear mixed effects models. The theoretical foundations of this field were laid almost a century ago, dealing with simple linear problems, but when it comes to nonlinear mixed effects models, progress has been dependent on parallel developments in the computational methods for maximum likelihood estimation as well as on the computational power available. While a variety of methods are now available in the literature and many of these are implemented in freely available software packages, the tools of optimal experimental design are not as widely used as one would expect, especially considering that the intended outcome of applying this theory is optimal allocation of sparse experimental resources, an outcome which has both economic and ethical implications.

A possible explanation for the relative lack of penetration of these tools is that they require a large amount of quite specific prior knowledge; nonlinear mixed effects models require the specification of a regression function as well as which parameters are presumed to vary between subjects. On top of this, for these models even the simplest criteria for optimal experimental design require the user to specify guesses as to the values of the parameters, which in the mixed effects case includes the variance parameters related to variation between subjects. The more advanced criteria require the user to specify distributions for the parameters in order, for instance, “robustify” against parameter misspecification or to implement Bayesian thinking.

The work presented here constitutes an attempt to lay a foundation for a practically attractive approach to experimental design in this setting. The strategy relies on finding optimal designs for a wide range of parameter values and then assessing their relative performances under parameter misspecification. Parameter values must still be chosen; to facilitate this, a scale for variance parameters is constructed which allows users to relate

the variation of parameters between subjects to its expected impact on the response profile, thus yielding a more intuitively accessible avenue of elicitation.

Because many designs must be found, a classic and computationally cheap optimal design criterion is employed. This criterion is compared with a newer, computationally more expensive one in order to assess the trade-off, in terms of information lost, related to this choice. In the cases examined, it turns out that the two criteria do not differ substantially.

To illustrate the application and general feasibility of the proposed workflow it is applied to a wide range of examples from the literature. Two R-packages have been developed which allow the user to apply the proposed variance scale, find optimal experimental designs and construct the proposed diagnostic graphs for user-specified nonlinear mixed effects models.

This thesis has two main chapters:

**Chapter 1** introduces the relevant theoretical background of optimal design and the two different design criteria. Methods for assessing the relative performance of the two criteria are introduced and applied to a selection of models. The results, together with a deliberation on the theoretical justification of the more expensive criterion, indicate that choosing the cheaper one may not be associated with any loss of quality in the resulting experimental design.

**Chapter 2** describes and applies the proposed workflow to a wide range of cases. This also works as a showcase for the R-packages that were developed. Finally, the advantages and weaknesses of the approach are discussed and a conclusion is drawn as to what has been achieved in this work.

**Table 0.1:** *Acronyms and symbols used in this thesis*

<b>Acronym</b>	<b>Explanation</b>
NLME	<i>Nonlinear Mixed Effects</i>
$\eta$	<i>Regression function in the NLME model</i>
$\boldsymbol{\eta}$	<i>Vectorised regression function in the NLME model</i>
$\boldsymbol{\beta}$	<i>Vector of fixed effects parameters in the NLME model</i>
$\mathbf{b}$	<i>Vector of random effects</i>
$\Omega$	<i>Variance matrix of the random effects vector in the NLME model</i>
$\omega$	<i>Random effect variance parameter</i>
	<i>Variance component in the NLME model</i>
$\varepsilon_{ij}$	<i>Error term (residual) in the NLME model</i>
	<i>Of measurement <math>j</math> in individual <math>i</math></i>
$\sigma^2$	<i>Variance of the error term in the NLME model (residual variance)</i>
$\xi$	<i>Experimental design vector</i>
MLE	<i>Maximum Likelihood Estimate (or Estimation)</i>
FO	<i>First Order</i>
	<i>Definition on page 41 in equation (1.5)</i>
FOCE	<i>First Order Conditional Expectation</i>
	<i>See (Retout and Mentre, 2001)</i>
	<i>and (Pinheiro and Bates, 1995)</i>
InFO	<i>Integrated First Order</i>
MC	<i>Monte Carlo</i>
AGQ	<i>Adaptive Gaussian Quadrature</i>
MC/AGQ	<i>Monte Carlo / Adaptive Gaussian Quadrature</i>
$L_{GQ}$	<i>Likelihood approximation using Gaussian Quadrature</i>
$L_{AGQ}$	<i>Likelihood approximation using Adaptive Gaussian Quadrature</i>
EC50	<i>"Effective Concentration, 50% of"</i>

$FIM$	<i>Fisher Information Matrix</i>
$FIM_{lin}$	<i>Fisher Information Matrix of linearised model</i>
$FIM_{FO}$	<i>Approximation of Fisher Information Matrix based on First Order likelihood approximation</i>
$FIM_{InFO}$	<i>Approximation of Fisher Information Matrix based on Integrated First Order likelihood approximation</i>
$FIM_{MC/AGQ}$	<i>MC/AGQ approximation of the FIM</i>
$h_{AGQ}$	<i>Term in the definition of <math>FIM_{MC/AGQ}</math></i>
$\phi_n$	<i>Standard normal density on <math>\mathbb{R}^n</math></i>
CV	<i>Coefficient of Variation</i>

# 1 D-optimal Design for Nonlinear Mixed Effects Models - Choosing Computational Approach

1.1	Introduction . . . . .	31
1.1.1	Nonlinear Mixed Effects (NLME) model . . . . .	34
1.1.2	Population Experimental Design . . . . .	37
1.1.3	Likelihood and Fisher Information . . . . .	38
1.1.4	<i>D</i> -optimality . . . . .	39
1.2	FIM Approximation . . . . .	39
1.2.1	Linearisation . . . . .	40
1.2.2	The First Order (FO) Method . . . . .	41
1.2.3	The Integrated First Order (InFO) Method . . . . .	41
1.2.4	Monte Carlo / Adaptive Gaussian Quadrature . . . . .	42
1.3	Cases to be examined . . . . .	45
1.3.1	Exponential Decay I and II . . . . .	45
1.3.2	One-Compartment . . . . .	46
1.3.3	Sigmoid Emax . . . . .	47
1.4	Demonstration of the Gold Standard . . . . .	49
1.4.1	Example I . . . . .	49
1.4.2	Example II . . . . .	52
1.5	Strategy for comparing FO and InFO . . . . .	55
1.6	Finding <i>D</i> -optimal designs . . . . .	58
1.6.1	Optimisation . . . . .	58
1.7	Comparing FO and InFO <i>D</i> -Optimal Designs . . . . .	64
1.7.1	Qualitative Comparison: Visual inspection . . . . .	64
1.7.2	Quantitative Comparison: Relative <i>D</i> -Efficiency . . . . .	65

1.7.3	Note on interpretation of relative $D$ -efficiency . . . . .	65
1.7.4	Bootstrap . . . . .	66
1.8	Discussion . . . . .	70
1.8.1	Remark concerning scope . . . . .	72
1.9	Further Work . . . . .	74

## 1.1 Introduction

*Nonlinear Mixed Effects* (NLME) modelling allows an investigator to *a priori* posit, e.g., a mechanistic model describing the phenomenon of interest; this is then used to specify the regression function. Between-individual variation in the shape of the response is taken into account by allowing the parameters which specify the regression function to vary between individuals; for each individual, a parameter is drawn from a distribution which is common to the whole population. Thus, under an NLME model, individual response profiles follow the same functional form but the parameters are allowed to be individual-specific. These parameters often have scientific interpretations due to the mechanistic derivation of the regression function, e.g., in pharmacokinetics, representing the absorption rates of a compound into the blood.

NLME models are widely used; examples of applications include modelling the growth of orange trees, intensity of current in semi-conductors, intensity of earthquakes and the cold tolerance of grass (see Pinheiro and Bates, 2000).

We will focus on examples drawn from the pharmaceutical industry, an area which has motivated seminal developments in NLME modelling and design. The models we consider have been used to, e.g. analyse chemical reactions, the dynamics of *in vivo* drug concentration or the effect of a drug on a subject (often described by receptor theory, i.e., a mechanistic description of how much of the drug is bound to the relevant receptors in the body at a given time).

We are concerned with optimal experimental design for NLME models. In practical terms, optimal experimental design constitutes an attempt to optimally allocate sparse experimental resources. Given this economic incentive, it is not surprising that optimal experimental design is widely used, especially in the pharmaceutical industry. In this context, optimal experimental design is of ethical as well as economic importance; e.g., avoiding unnecessary experimentation on humans and animals is self-evidently a desirable objective.

In broad terms, the idea of optimal experimental design can be described as follows: given a parameterised statistical model, we view the distribution

of the parameter estimator as a function of the experimental design and then optimise a specific property of that distribution.

A wide range of criteria have been developed which correspond to different objectives and perspectives. In this chapter, we focus on the classic *D-optimality* criterion.

For linear models, the *D-optimality* criterion picks the design which minimises the determinant of the covariance matrix of the maximum likelihood parameter estimator (MLE). This is equivalent to minimising the volume of the confidence region for the parameter estimates. In the linear case, this optimisation is easily performed in practice as explicit expressions for the the covariance matrix are available.

However, NLME models generally do not admit an explicit expression of the likelihood. This implies that the MLE covariance matrix also has no explicit expression. Instead, the inverse of the Fisher information matrix (FIM) is used as a proxy for the MLE covariance matrix. This is motivated by the Cramer-Rao bound for unbiased estimators, see Retout and Mentre (2003); Nyberg et al. (2012); Bazzoli et al. (2009); Nguyen and Mentre (2014).

Thus, *D-optimal* designs for NLME models are found by maximising the determinant of the FIM over the space of possible experimental designs, just as with the general linear model. It follows that accurate and efficient computation of the FIM is essential to the practical applicability of *D-optimal* designs.

In this chapter we compare two approaches to approximating the FIM for NLME models, both relying on linearisation of the model. The first, which we will refer to as the *first order* (FO) method, is the classical approach (see Mentre et al. (1997) and page 468 of Atkinson et al. (2007) and has the advantages of being both simple to implement and computationally inexpensive. Being well-established, it is also widely implemented in software, for instance PopED, Pfim, PkStaMp and PopDes (see Nyberg et al. (2015)). The second approach was initially proposed by Retout and Mentre (2003) but has not yet seen widespread use. We will refer to this approach as the *integrated first order* (InFO) method. We believe there are two factors which explain the failure of InFO to gain traction; firstly, it is computationally expensive in that it involves an integration of the



FIM of the linearised model over a set of parameter values. This results in substantially longer computation times when optimising, especially given the current implementations which rely on Monte Carlo integration (PopED). Secondly, while it has been suggested (Wang et al. (2012); Nyberg et al. (2012)) that InFO is superior to FO when random effect variance is large, to date no thorough investigation of when or how this superiority obtains has been performed.

We seek to address both these points. First, a computationally efficient (fast enough to make optimisation feasible for the models we investigate) implementation of InFO has been achieved through the use of Gaussian quadrature (assuming normal distribution of random effects). Second, we carry out a systematic investigation of four different models in which we find  $D$ -optimal designs using both methods and compare the resulting designs using a third FIM approximation, which we regard as a gold standard. This third FIM approximation is called *Monte Carlo with Adaptive Gaussian Quadrature* (MC/AGQ) and has the advantages that it avoids linearisation and theoretically can be made arbitrarily close to the true FIM. It was initially proposed by Nguyen and Mentre (2014) and subsequently an improved version was presented in Ueckert et al. (2015). While we regard it as a gold standard it is not feasible to use it for optimisation purposes due to its high computational cost.

To enable an initial investigation of the relative performance of FO and InFO, we have implemented both methods in generalised code and applied them to a number of industrially relevant cases. Our investigations focus on comparing the relative quality of designs that are  $D$ -optimal according to FO and InFO, respectively.

The remainder of the paper is arranged as follows. Section 1.1 introduces the formal framework and terminology; experimental designs, NLME models and  $D$ -optimality are defined. Section 1.2 explains the two linearisation-based methods of FIM approximation as well as MC/AGQ. Section 1.3 introduces the examples for which we compare the two methods. In Section 1.4 we demonstrate the gold standard method (MC/AGQ) and use it to provide a sanity check of the two existing methods. Section 1.5 provides an outline of the strategy we employ to compare the two methods. In Section 1.6 we identify  $D$ -optimal designs using both FO and

InFO and we investigate the effect of increasing the between-individual variance on the resulting designs. Sections 1.7 deals with comparing designs, qualitatively and quantitatively. The former is mainly done through visual inspection of the resulting designs and the latter is achieved through computing relative  $D$ -efficiency, a measure of relative quality. Section 1.8 is a discussion of the results and their implications for further research.

### 1.1.1 Nonlinear Mixed Effects (NLME) model

Consider a study in which  $N$  individuals participate. A response of interest, say concentration of a compound in the blood, is to be measured repeatedly for each individual. It is believed a priori that the response over time for a given individual can be modelled with a specific structural function. Such a function can be motivated by, say, a compartmental modelling approach (Bates and Watts, 1988, chap. 5).

Each individual is to be measured  $n$  times, using the same set of sampling times for all individuals. A set of such sampling times for a single individual is called an *elementary design*. The set of allowed elementary designs takes the form  $\Xi = \{\xi = (t_1, \dots, t_n) : T_1 \leq t_1 \leq t_2 \leq \dots \leq t_n < T_2\}$  for some  $T_2 > T_1 > 0$ . The response of the  $i$ th individual is a vector

$$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in})^\top,$$

of univariate observations where  $Y_{ij}$ ,  $1 \leq j \leq n$ , is the observation made at time  $t_j$  for individual  $i$ .  $Y_{ij}$  is assumed to be given by

$$Y_{ij} = \eta(\boldsymbol{\beta}_i, t_j) + \varepsilon_{ij} \tag{1.1}$$

where  $\eta : \mathbb{R}^p \times [0, \infty) \rightarrow \mathbb{R}$  is the structural function assumed to be shared across all individuals,  $\boldsymbol{\beta}_i$  is a vector of parameters specific to individual  $i$  and  $\varepsilon_{ij}$  is the error term associated with observation  $ij$ .

The function  $\eta$  is assumed to be nonlinear in its first argument (i.e., the parameter  $\boldsymbol{\beta}_i$  specific to individual  $i$ ). This is the “nonlinear” part of a Nonlinear Mixed Effects Model.

The distributional assumptions behind the model are that  $\varepsilon_{ij}$  are independent and normally distributed with zero mean and variance  $\sigma^2$  and

that  $\beta_i$  are independent and normally distributed with a population-level mean  $\beta \in \mathbb{R}^p$  and a  $p \times p$  variance matrix  $\Omega$ . Within an individual, the parameters are assumed independent, i.e.,  $\Omega$  is assumed to be diagonal with entries  $(\omega_1, \dots, \omega_p) \in (0, \infty)^p$ . Here, the entries of  $\Omega$ , also known as “variance components”, are the individual variances of the entries in  $\beta_i$ . They determine how much variation is associated with a given parameter, when sampling several individuals from the population.

In this setup,  $\beta_i$  determines the shape of the response profile for individual  $i$  and drawing  $\beta_i$  from a normal distribution allows the model to accommodate different shapes for different individuals. In the context of an experiment which measures concentration of compound in blood for instance,  $\beta_i$  could be related to the metabolism of individual  $i$  and thus determine how quickly the compound is cleared from her system.

In Davidian and Giltinan (2003), it is noted that  $\beta$ , the population-level mean of the distribution of  $\beta_i$ , represents the typical parameter value, not the typical response profile. For instance, assume that  $p = 1$ ,  $\beta = 0$ ,  $\Omega = 1$  and  $\eta(\beta_i, t) = (\beta_i)^2$ . And say we take “the typical response profile” to refer to the function of  $t \geq 0$  given by  $E[\eta(\beta_i, t)]$ . Then  $E[\eta(\beta_i, t)] = 1 \neq \eta(\beta, t) = 0$  for all  $t$ . Had we taken “typical” to refer to the mode instead of the mean, this would actually have yielded  $\text{mode}(\eta(\beta_i, t)) = \eta(\beta, t) = 0$ . But in that case, let  $p = 3$ ,  $\beta = (0, 0, 0)^\top$ ,  $\Omega = \mathbf{I}_3$  and  $\eta(\beta_i, t) = \beta_{i1}^2 + \beta_{i2}^2 + \beta_{i3}^2$ . Then the mode is equal to 1 while  $\eta(\beta, t) = 0$  still holds. This illustrates that it is important to understand what aspects of the structural function  $\beta_i$  actually controls.

In the cases treated in this thesis, the number of fixed effects parameters  $p$  will range from one to four. This number is determined by the complexity of the underlying structural model, i.e.,  $\eta$ . For instance, if the response is assumed to be well described by simple exponential decay then  $p = 1$ , and we set  $\eta(\beta, t) = \exp(-\beta t)$ . Conversely, if a one-compartment model including adsorption and elimination rates is used, then  $p = 3$  and we set

$$\eta(\beta, t) = \frac{e^{\beta_2 - \beta_3}}{e^{\beta_2} - e^{\beta_1}} \left( e^{-e^{\beta_1} t} - e^{-e^{\beta_2} t} \right)$$

The scientific rationale for these two regression functions will be given in a later section.

This thesis focuses on choosing  $\xi$  optimally for the purpose of estimating  $\beta$ , a part of  $\beta$  or a function of  $\beta$ . Typically, at least one of the entries of  $\beta$  has a clear scientific interpretation and is of greater interest to the investigators than the rest. A common example from pharmacokinetics is the “EC<sub>50</sub>” parameter. In a dose-response curve, EC<sub>50</sub> is “half the maximally effective dose”, i.e., the dose at which the response is predicted to be 50% of the upper asymptote. An example of this is the four-parameter logistic curve (Seber and Wild, 1989, p.338), given by

$$\eta(\beta, t) = \beta_1 + \frac{\beta_2 - \beta_1}{1 + e^{-\beta_3(t - \beta_4)}}$$

where it is assumed that  $\beta_2 > \beta_1 > 0$  and  $\beta_3, \beta_4 > 0$ . Upon inspection of this function, it is easy to see that  $\beta_1$  is the lower asymptote,  $\beta_2$  is the upper asymptote and  $\beta_4$  is in fact the EC<sub>50</sub>, i.e.,  $\eta(\beta, \beta_4) = \beta_1 + (\beta_2 - \beta_1)/2$ . It is very common to compare sets of dose-response curves using only this parameter. (Whether this is wise is debatable; it ignores the actual response window given by the interval  $(\beta_1, \beta_2)$ ).

For notational convenience in later developments, a vectorised formulation of the model may be employed. The response vector for the  $i$ th individual can be written as

$$Y_i = \eta(\beta_i, \xi) + \varepsilon_i \tag{1.2}$$

where the within-individual means and errors have been bundled into vectors

$$\begin{aligned} \eta(\beta_i, \xi) &= (\eta(\beta_i, t_1), \dots, \eta(\beta_i, t_n))^\top \\ \varepsilon_i &= (\varepsilon_{i1}, \dots, \varepsilon_{in})^\top \end{aligned}$$

And, also for convenience, the individual-level parameters  $\beta_i$  may be considered as given by normally distributed deviations (random effects) from the population-level parameter  $\beta$

$$\beta_i = \beta + \mathbf{b}_i$$

where  $\mathbf{b}_i \sim \mathcal{N}_p(0, \Omega)$  and  $\mathbf{b}_1, \dots, \mathbf{b}_N$  are independent.

A more general model could be used. The choice to allocate all individuals to the same design is a choice made for convenience and can in theory be relaxed. The same applies to the dimension of the covariate space, which has been reduced to only one dimension, namely time. Other covariates of interest could theoretically be included, be they part of the design or out of experimental control. In the context of reaction kinetics, examples of such additional covariates include temperature, pressure, starting concentrations and starting volumes. This point also serves to illustrate that “individual” here is simply a place-holder for some experimental unit; in the context of reaction kinetics, the experimental unit of interest is commonly a batch of reagent. The individual observations  $Y_{ij}$  could also be assumed multivariate, as in Nyberg et al. (2015). In their paper, they also allow for the relation between  $\beta$  and  $\mathbf{b}_i$  to be given by a nonlinear function. In Nyberg et al. (2015) they also note that  $\Omega$  need not be diagonal (though for purposes of convenience, they assume it to be so). Another popular extension is to include so-called “inter-occasion-variability” which essentially adds another random effect to the model which applies to groups of measurements (occasions) within individuals (Retout and Mentre, 2003; Nyberg et al., 2012).

### 1.1.2 Population Experimental Design

As stated in the previous section, this thesis defines an elementary design as an ordered vector of time points  $\zeta = (t_1, \dots, t_n)$ , with  $0 \leq t_1 \leq \dots \leq t_n < \infty$ . A *population experimental design* is generally speaking a collection of elementary designs

$$\mathbb{E} = \{\zeta_1, \dots, \zeta_N\}$$

which in the context of the previous section can be read as an experimental plan for a study of  $N$  individuals in which individual  $i$  is allocated to elementary design  $\zeta_i$  for  $i = 1, \dots, N$ . However, with the exception of a single example in Chapter 3, this thesis deals exclusively with the case where all individuals are allocated to the same elementary design, i.e.,  $\zeta_i = \zeta$  for all  $i = 1, \dots, N$ .

### 1.1.3 Likelihood and Fisher Information

The following assumes that the variance components  $\Omega$  and  $\sigma^2$  are known. Then the likelihood for the vectorised model given in equation 1.2 is simply the corresponding probability density, considered as a function of both the observation  $\mathbf{y} \in \mathbb{R}^n$  and the parameter  $\boldsymbol{\beta} \in \mathbb{R}^p$ . When  $\boldsymbol{\eta}$  is linear, the likelihood admits an explicit expression. That is, if  $\boldsymbol{\eta}(\boldsymbol{\beta}) = \mathbf{A}\boldsymbol{\beta}$  for some full rank  $n \times p$  matrix  $\mathbf{A}$ , then the likelihood of  $\mathbf{Y}_i$  is given by

$$L(\mathbf{y}; \boldsymbol{\beta}) = (2\pi)^{-\frac{n}{2}} \det(\mathbf{V})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{A}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{A}\boldsymbol{\beta})\right) \quad (1.3)$$

where  $\mathbf{V} = \text{Var}(\mathbf{Y}_i) = \text{Var}(\mathbf{A}(\boldsymbol{\beta} + \mathbf{b}_i) + \boldsymbol{\varepsilon}_i) = \mathbf{A}\Omega\mathbf{A}^\top + \sigma^2\mathbf{I}_n$ . In cases where  $\boldsymbol{\eta}$  is nonlinear, the likelihood generally does not admit an explicit expression and thus numerical approximation is usually required.

To find the maximum likelihood estimator (MLE)  $\hat{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$ , one solves the score equation

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log L(\mathbf{y}; \boldsymbol{\beta}) = 0$$

where the function on the left-hand side is known as the score function. The Fisher Information Matrix (FIM) is the variance of the score function. For [GET REF], the expectation of the score function is zero and thus the FIM is equal to the second moment of the score function. Thus, for a given elementary design  $\boldsymbol{\zeta} = (t_1, \dots, t_n)$  the FIM is the  $p \times p$  matrix

$$\text{FIM}(\boldsymbol{\zeta}) = \int_{\mathbb{R}^n} \left( \frac{\partial \log L(\mathbf{y}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) \left( \frac{\partial \log L(\mathbf{y}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)^\top d\mathbf{y}, \quad (1.4)$$

and for the population design  $\Xi$ , the FIM is

$$\text{FIM}(\Xi) = \sum_{i=1}^N \text{FIM}(\boldsymbol{\zeta}_i),$$

due to independence between individuals.

Intuitively, the FIM describes the degree to which the MLE  $\hat{\boldsymbol{\beta}}$  varies with the observations  $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ . It can also be thought of as the probability-weighted curvature of the log-likelihood when the log-likelihood is considered as a function of  $\boldsymbol{\beta}$  only. Loosely speaking, it follows that a “large”

Fisher Information corresponds to high curvature and thus an easily identifiable  $\hat{\beta}$ , i.e., a small confidence region. This will be formalised in the next section.

#### 1.1.4 *D*-optimality

According to Demidenko (1997), if the number of individuals goes to infinity and the number of samples per individual is bounded, then the maximum likelihood estimator is asymptotically normal,

$$\hat{\beta} \stackrel{as.}{\sim} \mathcal{N}(\beta, \text{FIM}(\Xi)^{-1}) .$$

A *D-optimal design*,  $\Xi^*$ , minimises the log determinant of the asymptotic variance of the MLE; this is equivalent to

$$\Xi^* \in \operatorname{argmax}_{\Xi} \{ \log[\det(\text{FIM}(\Xi))] \} .$$

From the asymptotic distribution of  $\hat{\beta}$ , it immediately follows that solving this problem yields a design which minimises the volume of the asymptotic confidence region for  $\hat{\beta}$ . Note that the FIM and the approximations of it we employ will generally depend on the true parameter. Since the true parameter is not known, a guess as to its value must be used. This leads to the problem of the sensitivity of the resulting design to error in the guess. While this is an interesting problem and there are several different approaches to “robustification” of the design criterion, e.g., minimax or pseudo-Bayesian designs, we will not address it further in this chapter. In all examples we present, the true parameter is used as the “guess”.

## 1.2 FIM Approximation

In this section, the following three FIM approximations are introduced. The First Order (FO) approximation, the Integrated First Order (InFO) approximation and the Monte Carlo / Adaptive Gaussian Quadrature (MC/AGQ) approximation.

Both FO and InFO rely on linearisation of the model. The linearised model admits an exact expression of the FIM which depends on the value

around which we linearise. FO and InFO differ in how they address this dependence.

By contrast, MC/AGQ avoids linearisation. It is an attempt to directly compute the FIM and relies on a combination of adaptive Gaussian quadrature, Monte Carlo integration and (in our implementation) numerical differentiation.

### 1.2.1 Linearisation

The vectorised form of the NLME model (equation 1.2) may be linearised around an arbitrary value  $\bar{\mathbf{b}} \in \mathbb{R}^p$ . Specifically, if we consider the design  $\zeta$  to be fixed, i.e.,  $\eta(\boldsymbol{\beta}) = \eta(\boldsymbol{\beta}, \zeta)$ , a first-order Taylor expansion of  $\eta$  around  $\boldsymbol{\beta} + \bar{\mathbf{b}}$  evaluated in  $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{b}_i$  yields

$$Y_i = \eta(\boldsymbol{\beta}_i) + \varepsilon_i \simeq \eta(\boldsymbol{\beta} + \bar{\mathbf{b}}) + \boldsymbol{\eta}'(\boldsymbol{\beta} + \bar{\mathbf{b}})(\mathbf{b}_i - \bar{\mathbf{b}}) + \varepsilon_i .$$

Following Retout and Mentre (2003), let  $E$  and  $V$  respectively denote the expectation and variance of the right-hand side. Then the FIM for the linearised model is given by

$$\begin{aligned} \text{FIM}_{lin}(\bar{\mathbf{b}})_{kl} = & \sum_{i=1}^N \left[ \frac{\partial E_i^\top}{\partial \beta_k} V_i^{-1} \frac{\partial E_i}{\partial \beta_l} \right. \\ & \left. + \frac{1}{2} \text{trace} \left( \frac{\partial V_i}{\partial \beta_l} V_i^{-1} \frac{\partial V_i}{\partial \beta_k} V_i^{-1} \right) \right] , \end{aligned}$$

for  $k, l = 1, \dots, p$ , with

$$E_i = \eta(\boldsymbol{\beta} + \bar{\mathbf{b}}) - \boldsymbol{\eta}'(\boldsymbol{\beta} + \bar{\mathbf{b}})\bar{\mathbf{b}}$$

and

$$V_i = \boldsymbol{\eta}'(\boldsymbol{\beta} + \bar{\mathbf{b}})^\top \Omega \boldsymbol{\eta}'(\boldsymbol{\beta} + \bar{\mathbf{b}}) + \sigma^2 I_n ,$$



for  $i = 1, \dots, N$ . Since all subjects are assigned to the same design, it holds that  $E_i = E_1$  and  $V_i = V_1$  for  $i = 1, \dots, N$ . Therefore, the above expression reduces to

$$\begin{aligned} \text{FIM}_{lin}(\bar{\mathbf{b}})_{kl} = N & \left[ \frac{\partial E_1^\top}{\partial \beta_k} V_1^{-1} \frac{\partial E_1}{\partial \beta_l} \right. \\ & \left. + \frac{1}{2} \text{trace} \left( \frac{\partial V_1}{\partial \beta_l} V_1^{-1} \frac{\partial V_1}{\partial \beta_k} V_1^{-1} \right) \right]. \end{aligned}$$

Up until this point,  $\bar{\mathbf{b}}$  has been left as an arbitrary choice of constant in  $\mathbb{R}^p$ . What distinguishes the two linearisation-based FIM approximations described in the subsequent sections is the way in which the problem of choosing  $\bar{\mathbf{b}}$  is approached. Both approaches are originally motivated by thinking of  $\bar{\mathbf{b}}$  as a “best guess” value of the realised random effect  $\mathbf{b}_i$ , the problem being that, at the time of guessing, the random effect has not yet been realised.

### 1.2.2 The First Order (FO) Method

The following approach was proposed by Mentre et al. (1997).

With no observations, a natural guess as to the value of the realised random effect is  $\bar{\mathbf{b}} = E(\mathbf{b}) = 0$ , leading to the FO FIM approximation

$$\text{FIM}_{FO} = \text{FIM}_{lin}(0). \quad (1.5)$$

### 1.2.3 The Integrated First Order (InFO) Method

InFO was proposed by Retout and Mentre (2003). It addresses the dependence on  $\bar{\mathbf{b}}$  by integrating it out with respect to the distribution of the random effect. This leads to the InFO FIM approximation,

$$\text{FIM}_{InFO} = \int_{\mathbb{R}^p} \text{FIM}_{lin}(\bar{\mathbf{b}}) \phi_p \left( \Omega^{-1/2} \bar{\mathbf{b}} \right) d\bar{\mathbf{b}}, \quad (1.6)$$

where  $\phi_p$  is the standard normal density on  $\mathbb{R}^p$ .  $\text{FIM}_{InFO}$  can heuristically be regarded as an attempt to provide a “robustified” alternative to  $\text{FIM}_{FO}$ , the intuition being that the more variable the random effect is, the less accurate a fixed value guess will be on average.

### 1.2.4 Monte Carlo / Adaptive Gaussian Quadrature

The following approach was initially proposed in Nguyen and Mentre (2014) and subsequently an improved version was presented in Ueckert et al. (2015). To avoid cumbersome notation, we suppress the index  $i$  and treat the following as if we had only one elementary design  $\Xi = \{\xi\}$ , with  $n$  observations.

As mentioned in a previous section, the NLME model does not generally admit a likelihood in closed form. However, the likelihood can be expressed as an integral,

$$L(\mathbf{y}; \boldsymbol{\beta}) = \int_{\mathbb{R}^p} p(\mathbf{y}|\mathbf{b}; \boldsymbol{\beta}) p(\mathbf{b}) d\mathbf{b} \quad (1.7)$$

where  $p(\mathbf{y}|\mathbf{b}; \boldsymbol{\beta})$  and  $p(\mathbf{b})$  are normal densities on  $\mathbb{R}^n$  and  $\mathbb{R}^p$ , respectively,

$$p(\mathbf{y}|\mathbf{b}; \boldsymbol{\beta}) = \phi_n \left( \frac{\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b})}{\sqrt{\sigma^2}} \right),$$

$$p(\mathbf{b}) = \phi_p \left( \boldsymbol{\Omega}^{-1/2} \mathbf{b} \right).$$

In words, equation 1.7 expresses the marginal density of  $\mathbf{Y}_i$  as the integral of the joint density of  $\mathbf{Y}_i$  and the random effect  $\mathbf{b}_i$  which in turn is expressed as the product of the conditional density of  $\mathbf{Y}_i$  given the random effect and the marginal density of the random effect. This decomposition is also used in Section 7.2.1 of Pinheiro and Bates (2000). In the same section, it is shown how, by approximating the integrand in (1.7) with a normal density, we can use adaptive Gaussian Quadrature (AGQ) to approximate the integral. Gaussian quadrature is a method of numerical integration with respect to a given kernel or probability density; for different families of densities different quadrature rules apply. Gaussian quadrature approximates the integral with a weighted sum of integrand values, with weights and abscissas (points at which the integrand is evaluated) that are predetermined and which depend on the nature of the integrand as well as the level of precision that is required (to increase precision, we increase the number of abscissas, also referred to as “nodes”). The specific quadrature rule we implement, *Gauss-hermite* quadrature, assumes that the integrand takes the form of the product of a standard normal density and

a polynomial. Thus, it can be used to compute moments of a standard normal random variable.

For a given number  $Q \in \mathbb{N}$  of quadrature points, the standard Gaussian quadrature approximation to the NLME likelihood is

$$L_{GQ}(\mathbf{y}; \boldsymbol{\beta}) = \sum_{q=1}^Q w_q p(\mathbf{y} | \mathbf{b}_q, \boldsymbol{\beta}),$$

where  $w_1, \dots, w_Q \in (0, \infty)$  are the quadrature weights and  $\mathbf{b}_1, \dots, \mathbf{b}_Q \in \mathbb{R}^p$  are the quadrature nodes. These can be obtained from tables in Abramowitz and Stegun (1964) or from the R-package `statmod`.

An extension of this rule is *adaptive Gaussian quadrature* (see Section 7.2.1 of Pinheiro and Bates, 2000). Simply put, adaptive Gaussian quadrature is an importance sampling version of Gaussian quadrature. It is clear that by a simple substitution, Gaussian quadrature allows the computation of integrals where the kernel is any normal density. Given an integral to be computed, adaptive Gaussian quadrature exploits this by identifying a Gaussian kernel which approximates the integrand. The choice of kernel is given by choosing the mean and the variance of the approximating normal distribution.

The following AGQ approximation to the NLME likelihood was constructed in (Section 7.2.1 of Pinheiro and Bates, 2000). For the mean of the approximating normal distribution, pick the mode (assumed to be unique) of the integrand

$$\hat{\mathbf{b}} = \operatorname{argmax}_{\mathbf{b} \in \mathbb{R}^p} \{p(\mathbf{y} | \mathbf{b}; \boldsymbol{\beta}) p(\mathbf{b})\}$$

and for the variance, choose

$$B = \left[ -\frac{\partial^2}{\partial \mathbf{b}^2} \ln[p(\mathbf{y} | \mathbf{b}; \boldsymbol{\beta}) p(\mathbf{b})] \Big|_{\mathbf{b}=\hat{\mathbf{b}}} \right]^{-1}.$$

For a given number of quadrature points  $Q \in \mathbb{N}$ , the AGQ approximation of the NLME likelihood is

$$L_{AGQ}(\mathbf{y}, \boldsymbol{\beta}) = \sum_{q=1}^Q w_q^+ p(\mathbf{y} | \mathbf{b}_q^+; \boldsymbol{\beta}), \quad (1.8)$$

where

$$\mathbf{b}_q^+ = \hat{\mathbf{b}} + B^{-1/2} \mathbf{b}_q$$

with corresponding weights

$$w_q^+ = B^{-1/2} \frac{\phi_p(\mathbf{b}_q^+)}{\phi_p(\mathbf{b}_q)} w_q.$$

The approximate likelihood  $L_{AGQ}$  can be used to construct an approximation of the FIM. First note that FIM for the NLME model can be written as

$$\text{FIM} = \int_{\mathbb{R}^n} h(\mathbf{y}; \boldsymbol{\beta}) L(\mathbf{y}; \boldsymbol{\beta}) d\mathbf{y}$$

where

$$\begin{aligned} h(\mathbf{y}; \boldsymbol{\beta}) &= \frac{\partial \log L(\mathbf{y}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \log L(\mathbf{y}; \boldsymbol{\beta})^\top}{\partial \boldsymbol{\beta}} \\ &= \frac{\partial L(\mathbf{y}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial L(\mathbf{y}; \boldsymbol{\beta})^\top}{\partial \boldsymbol{\beta}} \left( L(\mathbf{y}; \boldsymbol{\beta}) \right)^{-2}, \end{aligned}$$

with an application of the chain rule in the second equality. Let

$$h_{AGQ}(\mathbf{y}; \boldsymbol{\beta}) = \frac{\partial L_{AGQ}(\mathbf{y}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial L_{AGQ}(\mathbf{y}; \boldsymbol{\beta})^\top}{\partial \boldsymbol{\beta}} \left( L_{AGQ}(\mathbf{y}; \boldsymbol{\beta}) \right)^{-2}.$$

The FIM can thus be approximated by a combination of likelihood approximation and Monte Carlo integration

$$\text{FIM}_{\text{MC/AGQ}} = \frac{1}{M} \sum_{m=1}^M h_{AGQ}(\mathbf{y}_m; \boldsymbol{\beta}),$$

where  $\mathbf{y}_1, \dots, \mathbf{y}_M$  are simulations from the NLME model. The high computational cost of this method is in part from having to compute the derivatives; in this chapter, they are computed numerically with the `grad` function from the `numDeriv` package in R. By default, it uses Richard-

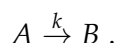
son's extrapolation method (Linfield and Penny, 1989). Another source of computational cost is that we have to find the mode  $\hat{\mathbf{b}}$  for each simulation.

### 1.3 Cases to be examined

We examine four industrially relevant examples from the literature. A summary of the models can be found in Table 1.1. The chosen parameter values are taken from the references indicated in the table. Each model is given by a regression function, a design region and a set of parameter values (value of  $\beta$ ,  $\omega$  and  $\sigma^2$ ), found in columns two, three and four of Table 1.1, respectively. Note that  $\omega_1$  has been omitted from the parameter values because we vary this parameter in our investigations.

#### 1.3.1 Exponential Decay I and II

Exponential decay provides a natural mechanistic description for a wide range of phenomena. For instance, it has been used to model radioactive decay, as well as irreversible first order reactions (Kitsos and Kolovos, 2013). On p.248-249 of Atkinson et al. (2007), exponential decay is used to describe the reaction scheme



That is, the compound A is transformed into another compound B at a rate  $k$  which is proportional to the concentration of A. In this model, the concentration  $[A]_t$  of A at time  $t$  is described by the differential equation

$$d[A]_t/dt = -k[A]_t ,$$

with analytical solution

$$[A]_t = [A]_0 \exp(-k \cdot t) ,$$

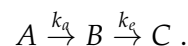
where  $[A]_0$  is the initial concentration of A and  $k$  is the *rate constant*. The square bracket notation for concentration is standard in the chemistry literature and its unit is number of moles per litre. The rate constant  $k$  is the proportion of material converted per second.

The model we call Exponential Decay I corresponds to assuming that the initial concentration of  $A$  is known (and assumed to take a value of one) and the rate constant  $k$  is to be estimated. Expressing this in the notation of equation 1.1, the regression function can be given as  $\eta(\beta, t) = [A]_0 \exp(-\beta t)$ . Exponential Decay II corresponds to the situation where both the rate constant and the initial concentration of  $A$  are assumed unknown. Expressing this in the notation of equation 1.1, the regression function can be given as  $\eta(\beta, t) = \beta_2 \exp(-\beta_1 t)$ .

### 1.3.2 One-Compartment

The One-Compartment model is one of the most popular examples in the literature of nonlinear regression models and optimal design for such models. It was used as a motivating example in seminal works such as Sheiner et al. (1972) and Mentre et al. (1997) as well as more recent works such as Wang et al. (2012).

As with exponential decay, the One-Compartment model is a solution to a differential equation which provides a mechanistic description of a wide range of phenomena. Examples of applications include the modelling of chemical reactions (Kitsos and Kolovos (2013)) and metabolisation of enzymes (Sjogren et al. (2011)). In a mechanistic pharmacokinetic interpretation of the model, it can be used to represent the concentration of a drug in an individual over time; the individual's body is considered a compartment which absorbs and eliminates the drug at rates that are proportional to the amount of drug yet to be absorbed and already absorbed but not yet eliminated, respectively. This can be represented by the following scheme



Specifically, an initial dose  $[A]_0$  of drug  $A$  is administered and it is absorbed by the body (or the blood) at a rate which is proportional to the current concentration  $[A]_t$  of unabsorbed  $A$ . As with Exponential Decay above,  $k_a$  and  $k_e$  are rate constants, with  $k_a$  being the rate of absorption and  $k_e$  being the rate of elimination. The absorption rate is thus given by the differential equation

$$d[A]_t/dt = -k_a[A]_t .$$

From the body compartment, the absorbed drug is eliminated at a rate which is proportional to the current concentration  $[B]_t$  of the drug in the compartment. The rate of change in the body compartment is thus given by

$$d[B]_t/dt = k_a[A]_t - k_e[B]_t ,$$

where  $k_e$  is the *elimination rate constant*. Solving this system of differential equations for  $[B]_t$  yields

$$[B]_t = \frac{k_a[A]_0}{k_e - k_a} \left( \exp(-k_a t) - \exp(-k_e t) \right) ,$$

which can then be used as the regression function in an NLME model. Expressing  $[B]_t$  in the notation of equation 1.1, the regression function can be given as

$$\eta(\boldsymbol{\beta}, t) = \frac{e^{\beta_2 - \beta_3}}{e^{\beta_2} - e^{\beta_1}} \left( e^{-e^{\beta_1} t} - e^{-e^{\beta_2} t} \right)$$

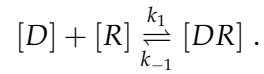
where  $\beta_1 = \log k_e$  and  $\beta_2 = \log k_a$  and  $\beta_3 = -\log[A]_0$ . Note that to make the translation consistent with Wang et al. (2012), the denominator of the first term has been multiplied by  $-1$  and the second term has been multiplied by  $-1$  as well, keeping the value of the expression unchanged.

### 1.3.3 Sigmoid Emax

The Sigmoid Emax model is a classic model in the area of pharmacodynamics. It provides a mechanistic description of a simple instance of “receptor theory”; the underlying idea is that the effect of a drug is indicated by the proportion of relevant receptors in the body the drug succeeds in binding itself to. This proportion is then modelled as a function of the concentration, or amount, of drug administered.

According to Mortensen et al. (2008), the model can be derived as follows. The effect of a drug is given by the number of drug molecules that have bound to relevant receptor sites. There is a number of available receptor sites  $[R]$  and a number of available drug molecules  $[D]$ . The number

of bound drug molecules (or *drug-receptor complexes*)  $[DR]$  is given by a reversible binding reaction



This is standard chemical reaction notation for an equilibrium reaction ( $\rightleftharpoons$  implies equilibrium), which indicates that the rate constant of the forward reaction (binding) is  $k_1$  and the rate constant of the backward reaction (unbinding) is  $k_{-1}$ . The rate of the forward reaction is thus  $\frac{d[DR]}{dt} = k_1[D][R]$  and the rate of the backward reaction is  $\frac{d([D]+[R])}{dt} = k_{-1}[DR]$ .

If the response of interest (the “effect”),  $E$ , is assumed proportional to  $[DR]$  by some proportionality constant  $\alpha > 0$ , that is, if  $E = \alpha[DR]$  then the maximal effect is  $E_{max} = \alpha([R] + [DR])$ . That is, the maximal effect is reached when all the receptors in the system are in drug-receptor complexes. In Mortensen et al. (2008), the forward and backward rate equations are solved under “steady state conditions”, the first of which is equality of the rates, i.e., equilibrium. This yields the solution

$$E = \frac{E_{max}[D]}{[D] + K_d}$$

where  $K_d = k_{-1}/k_1$ . This model is known as the Michaelis-Menten relationship.

The Sigmoid Emax model is an extension of the Michaelis-Menten model, given by

$$E = E_0 + \frac{E_{max}C^\gamma}{C^\gamma + EC_{50}^\gamma}$$

where  $C$  is concentration of the drug,  $EC_{50}$  is the concentration for which  $E = E_0 + \frac{1}{2}E_{max}$  and  $\gamma > 0$  is a parameter which allows the shape of the response curve to become more (or less) “s-shaped”. According to Salahudeen and Nishtala (2017), the quantity  $E_0$  represents a baseline for the measured effect in the absence of any drug and  $\gamma$ , known as the Hill coefficient, can be mechanistically justified by the system allowing for multiple ligands (drug molecules) to bind to the same receptor site. However, the use of  $\gamma$  is often empirically driven, i.e., scientists have experienced that including this particular parameter results in a substantially better fit.



Technically speaking, the instances of this model that are treated in this chapter are a special case of the Sigmoid Emax model, where  $\gamma = 1$ . This is also referred to as the Emax model. However, cases where  $\gamma \neq 1$  will be treated in the next chapter.

Expressing  $E$  in the notation of equation 1.1, and setting  $\gamma = 1$  we obtain the regression function

$$\eta(\boldsymbol{\beta}, t) = 5 + \frac{\beta_1 t}{\beta_2 + t} \quad (1.9)$$

where  $E_0 = 5$  has been assumed, to make the example consistent with Nguyen and Mentre (2014).

## 1.4 Demonstration of the Gold Standard

In this Section, we employ Monte Carlo with Adaptive Gaussian Quadrature (MC/AGQ) as the gold standard FIM approximation. Simple examples are given of its use and a brief sanity check is performed.

### 1.4.1 Example I

Recall that for the cases treated in this thesis, all random effects variance matrices are assumed to be diagonal  $V(\mathbf{b}_i) = \Omega = \text{diag}\{\omega_1, \dots, \omega_p\}$  for  $i = 1, \dots, N$ . Consider the Sigmoid Emax model described in Table 1.1, with  $\omega_1 = 0.3$ . In Figure 1.1 we plot the value of the MC/AGQ  $D$ -optimality objective function for the FO  $D$ -optimal design, as a function of the number of MC simulations. We have included 10,000 bootstrapped profiles of the MC/AGQ  $D$ -objective function value and constructed point-wise percentile intervals (2.5% and 97.5% quantiles) from these profiles. After about 2,000 simulations the percentile intervals narrow from an initial width of about 0.6 to a width of about 0.2. At 10,000 simulations the width is about 0.1. Whether these magnitudes of variation are acceptable can be judged only within some appropriate reference frame, which we will attempt to provide.

To provide a reference frame, we draw the reader's attention to Figure 1.2. Here, we plot profiles of the MC/AGQ  $D$ -optimality objective function

Model	Regression Function	Design Region**	Parameter values	Reference
Exponential Decay I	$\eta(\beta, t) = \exp(-\beta t)$	$T_1 = 0, T_2 = 5, n = 5$	$\beta = 1$ $\sigma^2 = 0.005^2$ $\omega_1 \in (0, 0.3]$	Atkinson et al. (2007)
Exponential Decay II	$\eta(\beta, t) = \beta_2 \exp(-\beta_1 t)$	$T_1 = 0, T_2 = 5, n = 5$	$\beta = (1, 1)^\top$ $\omega_2 = 0.1^2$ $\sigma^2 = 0.005^2$ $\omega_1 \in (0, 0.3]$	Kitsoos and Kolovos (2013)
One-Compartment	$\eta(\beta, t) = \frac{e^{\beta_2 - \beta_3} - e^{-\beta_1 t} - e^{-\beta_2 t}}{e^{\beta_2 - \beta_3} - e^{\beta_1}}$	$T_1 = 0, T_2 = 4, n = 5$	$\beta = (0, 1, 1)^\top$ $\omega_2 = 0.05$ $\omega_3 = 0.1$ $\sigma^2 = 0.05^2$ $\omega_1 \in (0, 1]$	Wang et al. (2012)
Sigmoid Emax*	$\eta(\beta, t) = 5 + \frac{\beta_1 t}{\beta_2 + t}$	$T_1 = 0, T_2 = 1000, n = 2$	$\beta = (30, 500)^\top$ $\omega_2 = 0.3$ $\sigma^2 = 4$ $\omega_1 \in (0, 4]$	Nguyen and Mentre (2014)

\* : In the Sigmoid Emax model, the random effects are not additive but exponential,

$$\beta_i = \begin{pmatrix} \beta_1 e^{b_{1i}} \\ \beta_2 e^{b_{2i}} \end{pmatrix}.$$

\*\* : The set of possible designs considered is given by the set  $\Xi = \{\xi = (t_1, \dots, t_n) : T_1 \leq t_1 \leq \dots \leq t_n \leq T_2\}$ .

**Table 1.1:** The specifications of the four models considered in this chapter. Note that  $\omega_1$  is varied across different ranges of values during our investigations. The first two models, Exponential Decay I and II, are not specified as NLME models in the references given, only as nonlinear regression models. Hence, the specification of the random effects for these two models is our own.

(5 quadrature nodes), corresponding to incremental changes from the FO  $D$ -optimal design. Each function evaluation is based on 10,000 simulations, and the red dots are 100 bootstrapped function evaluations based on those 10,000 simulations. The intervals are taken from Figure 1.1 and correspond to 1,000 simulations and 10,000 simulations, respectively.

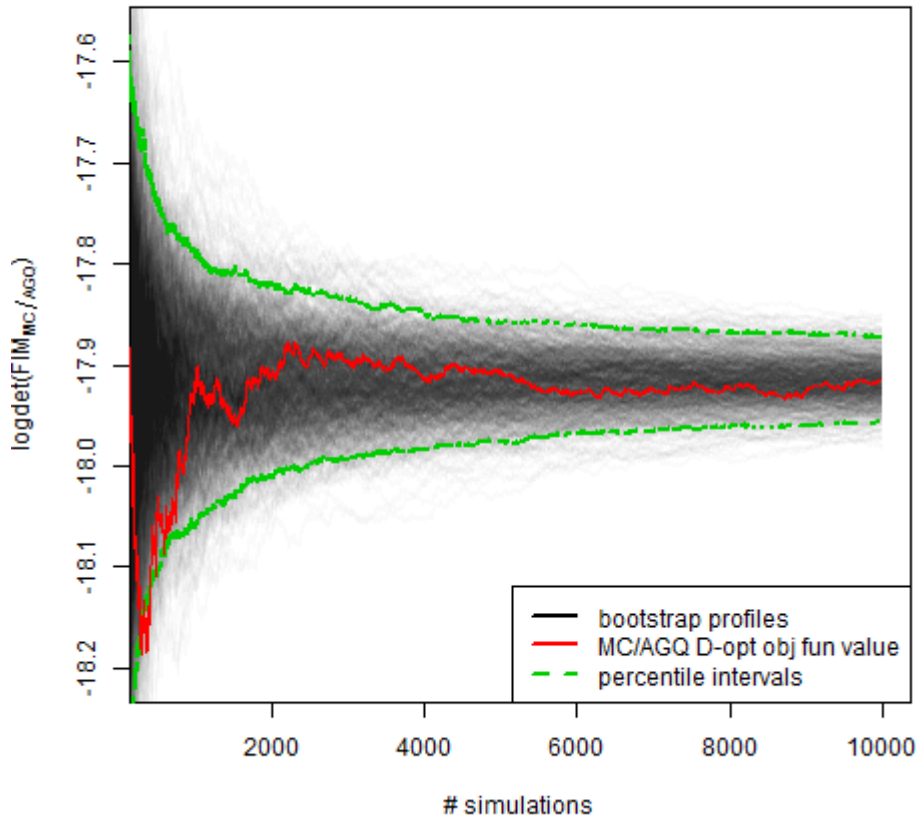
All three plots in Figure 1.2 give us the impression that the margin of error of the MC/AGQ  $D$ -optimality objective function is low enough that it will allow us to distinguish between radically different designs. This statement must be qualified in a number of ways. First, we observe that the objective function does seem to be flat around the optimum value and hence there is perhaps some reason to be concerned about how precisely we could determine the optimum if we were to use MC/AGQ for finding optimal designs. Second, this is a special case, both in terms of the model under consideration as well as the designs we evaluate the MC/AGQ  $D$ -optimality objective function for. It can therefore only serve as a proof of concept, not as a guarantee that MC/AGQ will always work. Third, these plots still only serve as rough visual checks, they do not provide a quantitative or absolute guide as to the quality of the approximation. With these caveats in mind, the plots allow us a certain degree of confidence that the approximation does not behave in an unexpected fashion and does not exhibit clearly unacceptable levels of imprecision with the chosen number of simulations and nodes.

### 1.4.2 Example II

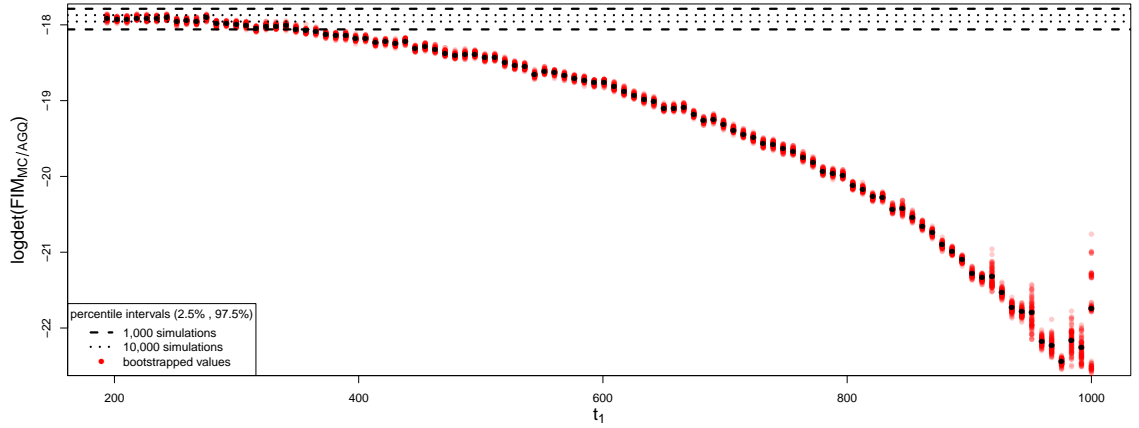
While FO and InFO have been used before in the literature, especially the former, the same is not true of the more novel MC/AGQ approach. For uses of FO and InFO see e.g. McGree et al. (2009), Strömberg et al. (2016) and Almquist et al. (2015). We therefore perform a quick sanity check; we randomly generate a number of designs uniformly from the design region and evaluate them with the MC/AGQ approach. If FO and InFO generate good designs, we would expect MC/AGQ to evaluate these designs as better than any randomly generated design.

The results have been plotted in Figure 1.3. We see the expected behaviour in the One-Compartment and Sigmoid Emax cases, i.e., the FO and InFO  $D$ -optimal designs have a higher value than any of the randomly generated designs when evaluated with the MC/AGQ  $D$ -optimality objective function. (Note that while the smoothed densities of the values of the random designs do cross the FO and InFO values that is an artefact of the smoothing; there are no random designs in that region). However, in the Exponential Decay cases, the FO and InFO  $D$ -optimal designs perform about as well as the randomly generated designs.

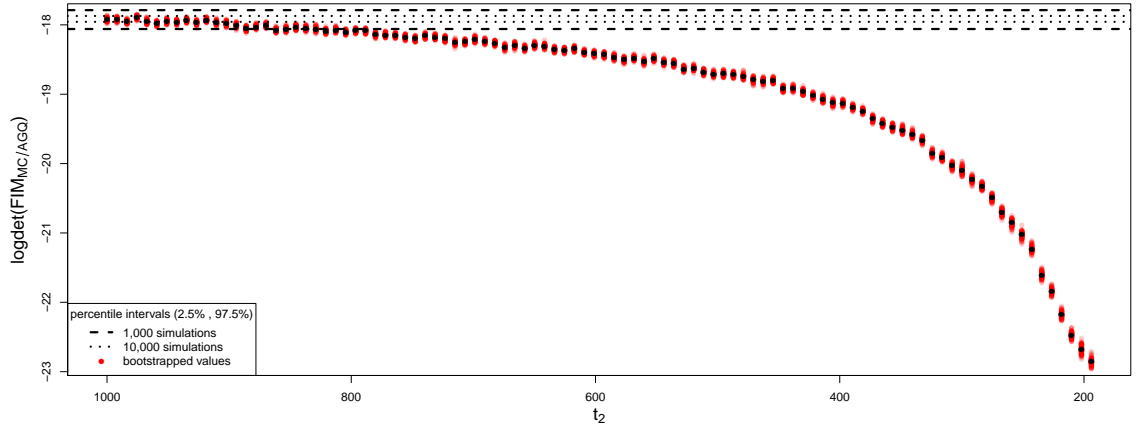
One possible explanation for the poor performance of the FO and InFO  $D$ -optimal designs in the case of Exponential Decay could be that the optimisation algorithm failed. To check this, consider Figure 1.4. Here, as in Figure 1.3, for each plot 10,000 design were sampled from the uniform distribution on the relevant design region. The densities plotted are values of the FO and InFO  $D$ -optimality objective functions. As a naive check, we observe that the optimisations seem to have worked well enough that none of the random designs are better than the optimal one (again, the densities that cross the relevant vertical lines are artefacts of smoothing). It seems there is good agreement between the InFO and FO objective functions in the One-Compartment and Sigmoid Emax models, but substantial disagreement in the Exponential Decay models. In the Exponential Decay models, the InFO density seems to be shifted to the right, in the other models the densities seem to overlap. In the Exponential Decay models, the InFO  $D$ -optimal design is worse than many of the random designs (especially in Exponential Decay I) and in the other models the overlap seems to



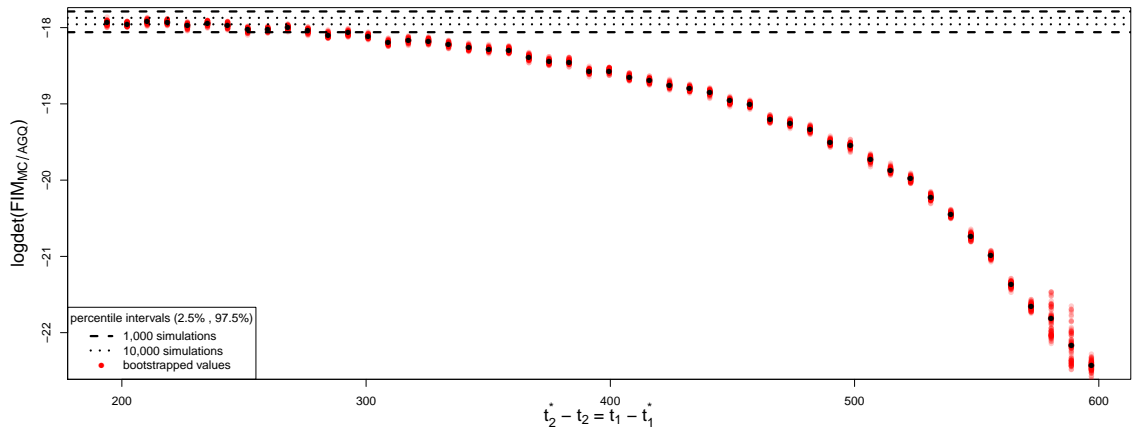
**Figure 1.1:** Illustration of the MC/AGQ approximation of the *FIM*. For the Sigmoid Emax model in Table 1.1 (with  $\omega_1 = 0.3$ ), we have computed the MC/AGQ approximation (5 quadrature nodes) of the *FIM* and plotted the value of its log-determinant in **red**, for an increasing number of MC simulations. The 10,000 simulations used in the approximation were subsequently bootstrapped and used to create bootstrapped profiles in faded black. Finally, point-wise percentile intervals (i.e., 2.5% and 97.5% quantiles of the bootstrapped distribution of  $\log\det(\text{FIM}_{\text{MC/AGQ}})$  for each number of simulations under consideration) were added in **green**.



(a) Moving  $t_1$  from  $t_1^*$  towards  $t_2^*$ .



(b) Moving  $t_2$  from  $t_2^*$  towards  $t_1^*$ .



(c) Moving  $t_1$  and  $t_2$  from  $\bar{\xi}^*$  towards each other.

**Figure 1.2:** Solid black points are the estimated values of  $\log\det(FIM_{MC/AGQ})$  as a function of (a)  $t_1$ , the first sampling time, keeping  $t_2 = t_2^*$ , (b)  $t_2$ , the second sampling time, keeping  $t_1 = t_1^*$ , (c) the distance between  $t_1$  and  $t_2$ . Note that  $(t_1^*, t_2^*) = \bar{\xi}^*$ , the FO D-optimal design. Dashed lines are the bounds of a percentile interval given by the 2.5% and 97.5% quantiles of the distribution of  $\log\det(FIM_{MC/AGQ}(\bar{\xi}^*))$ . For description of how these intervals are generated, see caption of Figure 1.1.

extend to the optimal designs. The scales of the function values are very different when compared with their MC/AGQ counterparts in Figure 1.3. The most remarkable cases of this are Exponential Decay II and Sigmoid Emax, which are orders of magnitude different from the MC/AGQ-based values in terms of the both location and spread. This is further evidence that there is a strong difference between the linearisation-based criteria and the MC/AGQ-based one. Recalling from Section 1.1.3 that the FIM is a measure of the curvature of the log-likelihood (as a function of  $\beta$ ) it can be concluded that the linearisation-based approaches seem to underestimate this curvature, in particular for Exponential Decay II.

This leads us to conclude that FO and InFO yield a poor FIM approximations in the Exponential Decay cases. It has previously been noted by Merlé and Tod (2001) that the linearisation-based methods are not always appropriate. Hence, it seems natural to include in any application of FO and InFO an assessment of whether the model is “too nonlinear”.

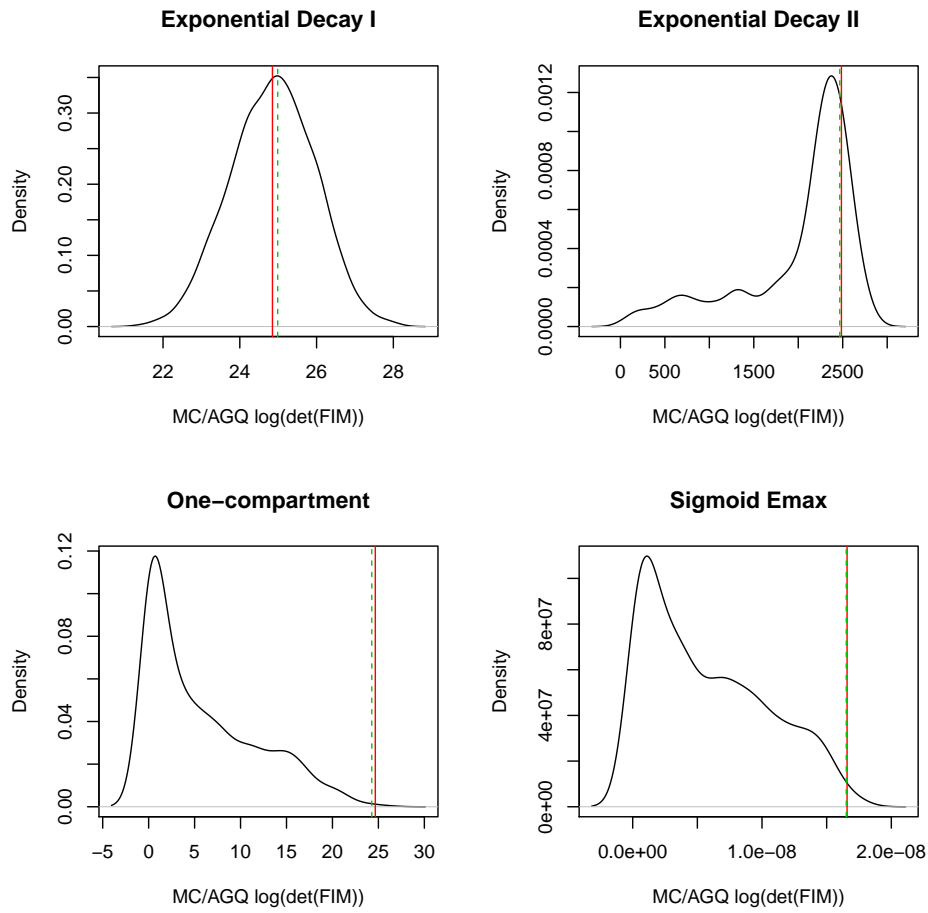
While addressing that question is outside the scope of this thesis, it is important; it may provide further motivation for exploring the feasibility of an MC/AGQ-based method for finding  $D$ -optimal designs.

## 1.5 Strategy for comparing FO and InFO

The existing evidence corroborating the claim that InFO is superior to FO for “large random effects” is sparse<sup>1</sup>; Wang et al. (2012) find FO and InFO  $D$ -optimal designs for a single model (One-Compartment), for a single value of  $\theta$ <sup>2</sup> and base the claim of InFO superiority on the fact that the value of the InFO  $D$ -optimality objective function is larger in the InFO  $D$ -optimal design than the value of the FO  $D$ -optimality objective function is in the FO  $D$ -optimal design. However, this is not sufficient evidence since two different  $D$ -optimality functions are being used; in fact, we have observed that there is often a systematic difference between the FO and InFO  $D$ -optimality objective functions.

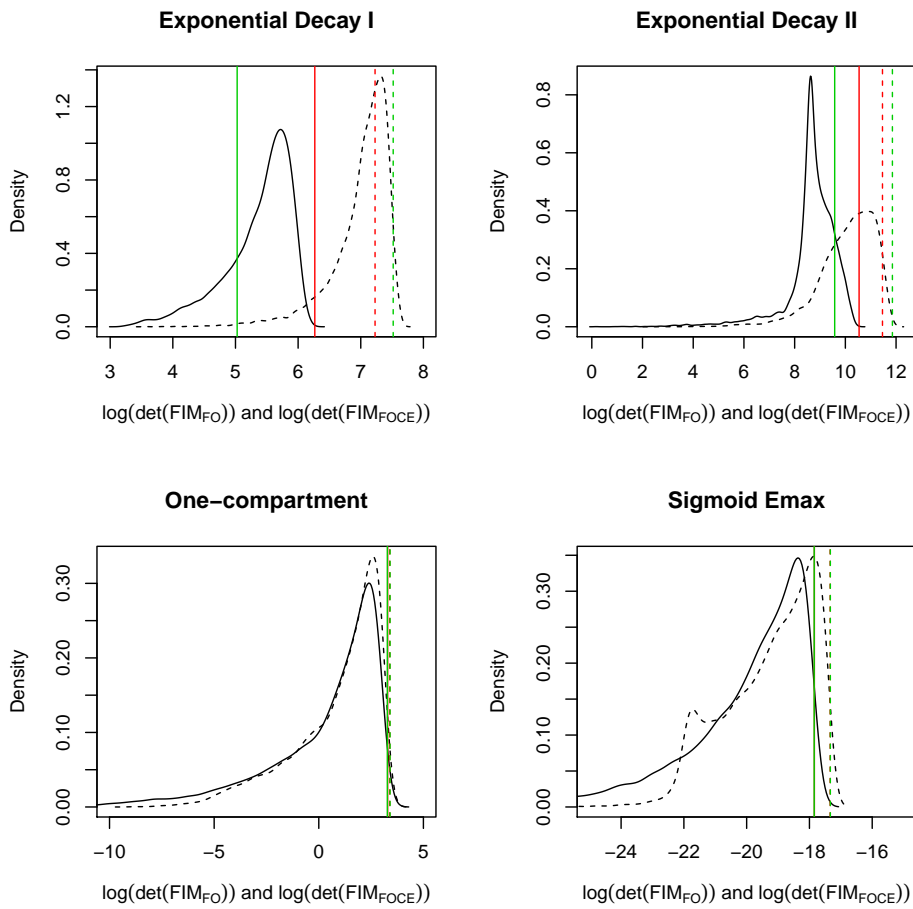
<sup>1</sup> Retout and Mentre (2003) actually observed no substantial difference between the approaches in the examples they considered.

<sup>2</sup> However, they vary the constraints on the design space and they also employ  $D_s$ -optimality, i.e.,  $D$ -optimality for subvectors of the parameter.



**Figure 1.3:** Smoothed histograms of the empirical distribution of the MC/AGQ  $\log(\det(\text{FIM}))$  value of designs distributed uniformly on the experimental region. The red vertical lines represent the MC/AGQ  $\log(\det(\text{FIM}))$  value of the FO D-optimal designs and the dashed green vertical lines represent the MC/AGQ  $\log(\det(\text{FIM}))$  value of the InFO D-optimal designs.





**Figure 1.4:** Smoothed histograms of the empirical distribution of the FO (solid lines) and InFO (dashed lines)  $\log\det(\text{FIM})$  value of designs distributed uniformly on the experimental region. The red vertical lines represent the  $\log\det(\text{FIM})$  value of the FO D-optimal designs and the green vertical lines represent the  $\log\det(\text{FIM})$  value of the InFO D-optimal designs. The red lines can't be seen in the One-Compartment and Sigmoid Emax plots because they are under the green lines (the solid red line is under the solid green line, the dashed red line is under the dashed green line).

Another method of comparing FO and InFO is employed by Retout and Mentre (2003); Nyberg et al. (2012) and Wang et al. (2012). For a given design, they compute the expected coefficients of variation (CVs) of the parameter estimators based on FO and based on InFO and then they compare the results with their empirical, simulation-based, counterparts. The idea is that the FIM approximation which generates CVs that are closer to the simulation-based CVs is the better method. However, they only consider a fairly low number of designs and small sample sizes for the simulation (five and 100, respectively, in the case of Wang et al. (2012)). Further it is not clear what constitutes “close” when comparing CVs, as these are marginal, parameter-wise, comparisons. The fact that CVs are marginal also means that it is not straightforward to use them to compare designs; some sort of evaluation of the joint CVs corresponding to a given design would have to be performed.

For these reasons, we pursue a different strategy for comparing FO and InFO. For the cases described in Section 1.3, we find FO and InFO  $D$ -optimal designs for a range of values of  $\omega_1$ . This will allow us to establish whether the two methods generate qualitatively different designs. Subsequently, we use our chosen gold standard, the MC/AGQ  $D$ -optimality objective function, to compute the relative  $D$ -efficiency which is defined in Section 1.7.2. As explained in Section 1.7.2, the relative  $D$ -efficiency of two designs is a measure of how many times one of the designs must be replicated in order to yield as much information as the other design. In particular, this enables an investigation of whether increasing  $\omega_1$  eventually leads to the InFO  $D$ -optimal design clearly outperforming the FO  $D$ -optimal design, on the scale of relative  $D$ -efficiency.

## 1.6 Finding $D$ -optimal designs

### 1.6.1 Optimisation

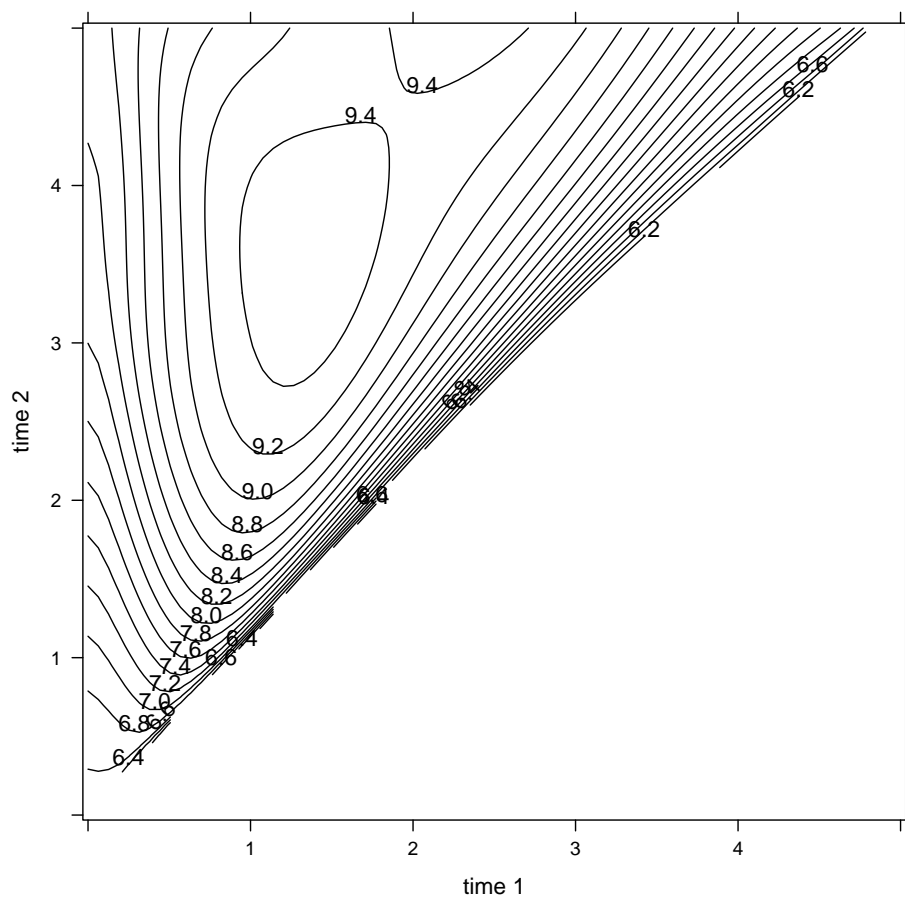
The FO  $D$ -optimality objective function is convex when approximate designs are allowed, see, e.g., page 61 of Fedorov and Leonov (2014) or page 75 of Schmelter (2007). However, in this work, only exact designs are considered and under this restriction the problem is not convex (see page

15 of Schmelter (2007)). The InFO  $D$ -optimality objective function is not convex in general either. This is illustrated in Figure 1.5; it is a plot of the objective function for the Exponential Decay II case described in Table 1.1, with  $\omega_1 = 0.15$  and with a reduced maximum sample size of  $n = 2$  instead of  $n = 5$  (so it becomes possible to plot the function). The plot clearly suggests that the function is not convex given that it appears to contain (at least) two separate local optima (manual of function values confirms this).

The R routines `nlminb` and `DEoptim` are employed for optimisation of the  $D$ -optimality objective functions. Note that both routines allow bounded optimisation, which is needed to ensure that the resulting design is within the design region. Another, classic, approach would be to transform the design region into an unbounded space, optimise there and transform back. This can be achieved using e.g. hyper-spherical coordinates. In Figure 1.6 we plot the FO  $D$ -optimal designs as a function of the between-individual variance for each of the four cases described in Table 1.1. These were found via local optimisation with the `nlminb` routine.

However, non-convexity necessitates a cautious approach to optimisation. This is illustrated by the first column of plots in Figure 1.7. In these plots, we have replicated the optimisation approach used to generate Figure 1.6. For both Exponential Decay cases, we see that the  $D$ -optimal design as a function of between-individual variance behaves somewhat erratically, with sudden discontinuities in an otherwise smooth graph. Furthermore, the bifurcation of the graph that we see in Exponential Decay II and in the One-Compartment model also suggests that there are several local optima in the objective function and that the optimisation method switches from locating one optimum to the other as  $\omega_1$  is increased.

For these reasons, we employ a global optimisation algorithm. We used the `DEoptim` routine from the R package by the same name. It implements a differential evolution algorithm for which we use the default settings except for the fact that we double the default maximum number of iterations and the population size parameter (`itermax=400` and `NP=100`). The resulting plots are presented in the second column of Figure 1.7. However, the design graphs seem to be highly discontinuous, especially in the Exponential Decay II case. This is not necessarily surprising, given that `DEoptim` make very few assumptions about the objective function, i.e., even if the true



**Figure 1.5:** InFO  $D$ -optimality objective function for Exponential Decay II, with  $(t_1, t_2) \in [0, 5]^2$ . Note that the NLME model differs from the Exponential Decay II model in Table 1.1 in that the number of sampling times is reduced to two (to allow the function to be plotted).

objective function is in fact smooth and convex, the optimisation algorithm does not assume it to be either and will, due to its random nature, produce solutions that vary around the true optimum, so that, even if all parameters are fixed, two independent runs of DEoptim will produce two (perhaps only slightly) different solutions. For the purposes of facilitating further enquiry, we make the heuristic assumption that the InFO  $D$ -optimality objective function is a smooth function without too many optima and that the DEoptim routine has produced designs that are located sufficiently near a global optimum that a follow-up local optimisation will produce a global optimum, i.e., a true InFO  $D$ -optimal design. Specifically, we use the results from the global optimisation as starting points for a follow-up local optimisation with nlminb and the resulting designs are assumed to be true InFO  $D$ -optimal designs. They are plotted in the third column of the Figure 1.7.

One could raise the concern that global optimisation ought to be used to find the FO  $D$ -optimal designs as well; the resulting graphs were identified and found to be essentially indistinguishable from the ones already presented.

For the sake of completeness, we formally define the *design graphs* presented in Figures 1.6 and 1.7.

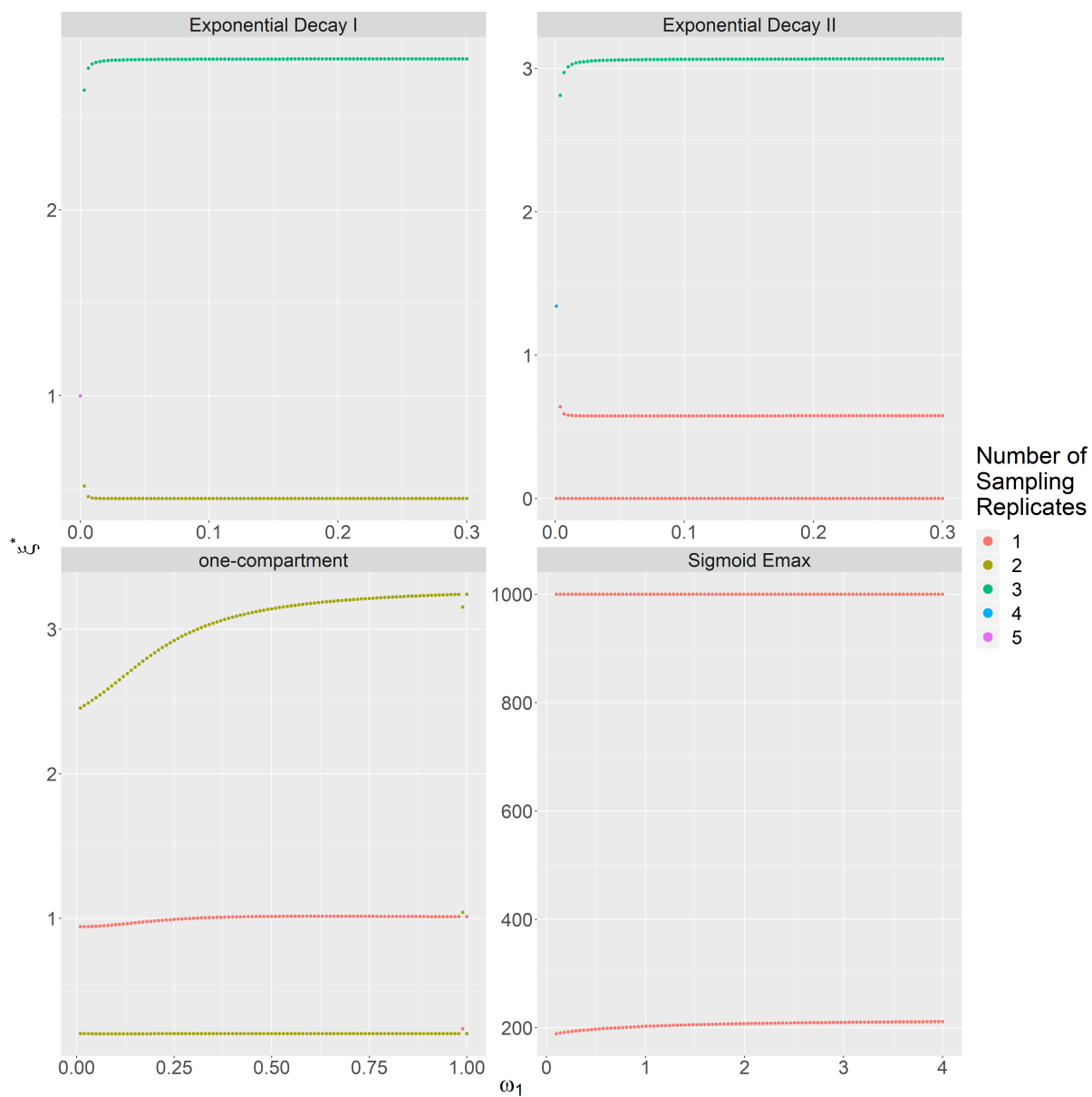
**Definition.** Given  $\theta = (\boldsymbol{\beta}^\top, \omega^\top, \sigma^2)^\top \in \Theta$  and a choice of optimality criterion, let  $\zeta^*(\theta)$  denote the optimal design. Let  $i \in \{1, \dots, p\}$  be given and define  $\omega^{(i)} : (0, \infty) \rightarrow \Theta$  as  $\omega^{(i)}(u) = \omega + e_i(u - \omega_i)$ , where  $e_i \in \{0, 1\}^p$  with

$$e_{ij} = \begin{cases} 0 & \text{if } j \neq i \\ 1 & \text{if } j = i \end{cases}$$

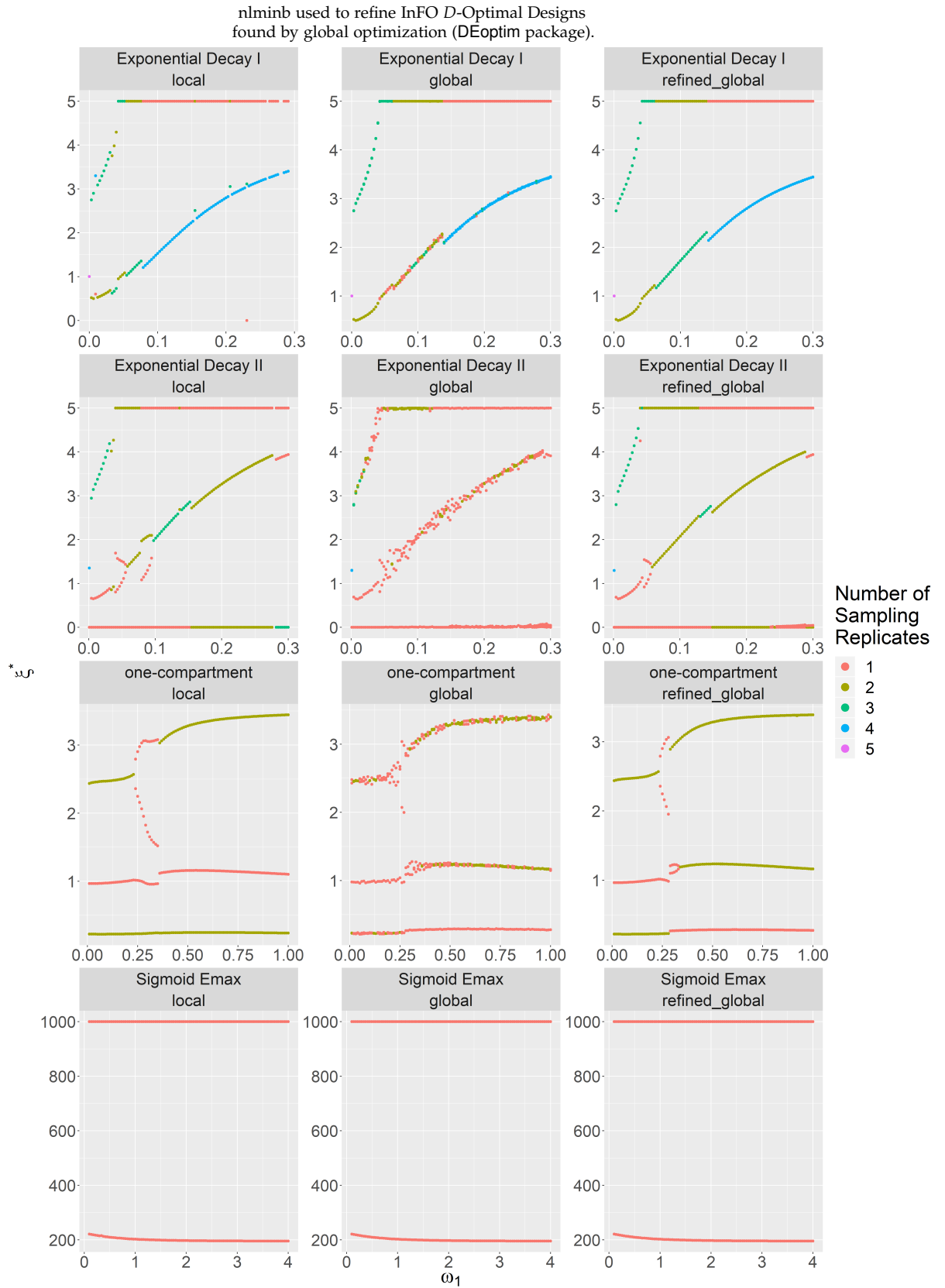
Then the design graph of  $\omega_i$  is given by the graph of the function

$$\zeta^*(u) = \zeta^*((\boldsymbol{\beta}^\top, \omega^{(i)}(u), \sigma^2)^\top).$$

That is, a vector  $\theta$  of true parameter values for both fixed effects and variance components is chosen. Then the optimal design  $\zeta^*$ , by some design criterion, is considered as a function of a single variance component, with all other parameter values fixed at the original value.



**Figure 1.6:** FO  $D$ -optimal designs for each of the models in Table 1.1. The optimal design is plotted as a function of the variance of the first random effect  $\omega_1 = \Omega_{11}$ .



**Figure 1.7:**

*First column:* InFO  $D$ -optimal designs with local optimization (function *nlminb*).

*Second column:* InFO  $D$ -optimal designs with global optimization (function *DEoptim*).

*Third column:* results of local optimisation (function *nlminb*) used with column 2 as starting point.

## 1.7 Comparing FO and InFO $D$ -Optimal Designs

### 1.7.1 Qualitative Comparison: Visual inspection

We have now succeeded in producing FO  $D$ -optimal designs, as presented in Figure 1.6, and we have also produced designs that can reasonably be expected to be InFO  $D$ -optimal, presented in the third column of Figure 1.7. This allows us to make a qualitative comparison of the methods via a visual inspection of the designs they yield. It is immediately apparent that the graphs of the  $D$ -optimal design as a function of  $\omega_1$  are qualitatively different for the two methods; there are differences in the shape as well as in the replication structure. However, in all cases, the FO  $D$ -optimal design for the smallest value of  $\omega_1$  closely resembles or is indistinguishable from the corresponding InFO  $D$ -optimal design. This is probably due to the fact that as  $\Omega \rightarrow 0$ , the two objective functions converge. For Exponential Decay I, the graph for FO (plot (a) of Figure 1.6) seems to quickly stabilise and then remain essentially constant whereas the corresponding graph for InFO (plot (c) of Figure 1.7) yields designs that are highly dependent on  $\omega_1$  both in terms of sampling times and replication structure. The same comments can be made regarding Exponential Decay II (plot (b) of Figure 1.6 versus plot (f) of Figure 1.7).

In the case of the One-Compartment model (comparing plot (c) of Figure 1.6 with plot (i) of Figure 1.7), FO does seem to yield designs that change smoothly with  $\omega_1$  without any tendency to stabilise within the range of values examined. The same is true of InFO. However, while some similarity can be traced in the set of sampling times yielded by FO and InFO, the graph for InFO has a bifurcation as well as a subsequent shift in replication structure; before the bifurcation, the graphs don't look very dissimilar, but afterwards, they are different in that FO yields designs which replicate the first and last sampling time twice whereas InFO yields designs that replicate the second and the last sampling times twice.

Lastly, in the case of the Sigmoid Emax model (plot (d) of Figure 1.6 versus plot (l) of Figure 1.7), all designs place the second sampling time at the extreme ( $t_2 = 1000$ ) while the first sampling time varies. In the case of



FO, the first sampling appears to be a concave function of  $\omega_1$  whereas for InFO, the opposite is the case.

There is thus ample reason to conclude that the two methods yield qualitatively different designs; the next question is whether one method yields better designs than the other. If we believe the claim that InFO outperforms FO when random effects are “large”, then we would expect that InFO would gradually start to outperform FO as  $\omega_1$  is increased. In the next section, we try to determine whether this is the case, by computing the relative  $D$ -efficiency of the InFO  $D$ -optimal design to the FO  $D$ -optimal design, as a function of  $\omega_1$ .

### 1.7.2 Quantitative Comparison: Relative $D$ -Efficiency

We propose to compare designs via relative  $D$ -efficiency, as defined below. Given two designs  $\xi, \xi'$ , the relative  $D$ -efficiency (of  $\xi$  to  $\xi'$ ) is

$$\text{relEff}(\xi, \xi') = \left\{ \frac{\det(\text{FIM}(\xi))}{\det(\text{FIM}(\xi'))} \right\}^{1/p} .$$

Relative  $D$ -efficiency is based on the notion of average variance; in the very simple case of a linear model and diagonal FIM's, the relative  $D$ -efficiency is in fact equal to the ratio of the geometric mean of the variance of the entries of  $\hat{\beta}$  under  $\xi'$  to the geometric mean under  $\xi$ . Given an approximation to the FIM, we can approximate the relative  $D$ -efficiency; in Section 1.7.4, we compute the MC/AGQ relative  $D$ -efficiency given by

$$\text{relEff}_{\text{MC/AGQ}}(\xi, \xi') = \left\{ \frac{\det(\text{FIM}_{\text{MC/AGQ}}(\xi))}{\det(\text{FIM}_{\text{MC/AGQ}}(\xi'))} \right\}^{1/p} .$$

### 1.7.3 Note on interpretation of relative $D$ -efficiency

Relative  $D$ -efficiency can also be defined for population designs, with the observation that if  $\Xi$  is a population design, then

$$\text{FIM}(\Xi) = \sum_{\xi \in \Xi} \text{FIM}(\xi) .$$

Let  $\zeta, \zeta'$  be given elementary designs and let  $\Xi$  and  $\Xi'$  be population designs where  $\Xi = \{\zeta\}$  and

$$\Xi' = \underbrace{\{\zeta', \dots, \zeta'\}}_{R \text{ times}}.$$

Then

$$\text{FIM}(\Xi') = R \cdot \text{FIM}(\zeta')$$

implying that

$$\det(\text{FIM}(\Xi')) = R^p \det(\text{FIM}(\zeta')).$$

In this case,

$$\begin{aligned} \text{relEff}(\Xi, \Xi') &= \frac{1}{R} \{\det(\text{FIM}(\zeta)) / \det(\text{FIM}(\zeta'))\}^{1/p} \\ &= \frac{1}{R} \text{relEff}(\zeta, \zeta'). \end{aligned}$$

So, if we (allowing non-integer number of individuals in population designs for the purposes of illustration) choose  $R$  such that

$$\text{relEff}(\Xi, \Xi') = 1$$

we get

$$\text{relEff}(\zeta, \zeta') = R.$$

In words, the relative  $D$ -efficiency of  $\zeta$  to  $\zeta'$  can be interpreted as “how many individuals must be allocated to design  $\zeta'$  in order to obtain the same information as we obtain from a *single* individual allocated to design  $\zeta$ ?”.

#### 1.7.4 Bootstrap

Since we compute the relative efficiency of FO and InFO  $D$ -optimal designs via MC/AGQ, it is subject to a random error associated with the Monte Carlo integration. In order to account for this error, we construct bootstrap confidence intervals for the relative efficiency, as per Efron and Tibshirani (1993). Because of the computational cost of constructing these intervals, we restrict our attention to ten pairs of designs for each model (one FO

and one InFO), with the designs corresponding to ten equi-distant values of  $\omega_1$ .

The bootstrap procedure for each value of  $\omega_1$  can be summarised as follows.

---

**Algorithm 1** Bootstrapped percentile intervals for relative MC/AGQ  $D$ -efficiency

---

**Input**  $\eta, \beta, \sigma^2$  and  $\omega_2, \dots, \omega_p \leftarrow$  from chosen row in Table 1.1  
**Input**  $\omega_1 \leftarrow$  desired value of  $\omega_1$   
Set  $\Omega = \text{diag}\{\omega_1, \dots, \omega_p\}$   
**for**  $a$  in (FO, InFO) **do**  
 $\zeta^{(a)} = (t_1^{(a)}, \dots, t_n^{(a)}) \leftarrow D$ -optimal design using approximation  $a$   
**for**  $i$  in 1:S **do**  $\triangleright S = 100,000$   
 $\varepsilon_i^{(a)} = (\varepsilon_{i1}^{(a)}, \dots, \varepsilon_{in}^{(a)}) \leftarrow$  drawn from  $\mathcal{N}_n(0, \sigma^2 \mathbf{I}_n)$   
 $\mathbf{b}_i^{(a)} = (b_{i1}^{(a)}, \dots, b_{ip}^{(a)}) \leftarrow$  drawn from  $\mathcal{N}_p(0, \Omega)$   
set  $\boldsymbol{\eta}_i^{(a)} = (\eta(\boldsymbol{\beta} + \mathbf{b}_i^{(a)}, t_1^{(a)}), \dots, \eta(\boldsymbol{\beta} + \mathbf{b}_i^{(a)}, t_n^{(a)}))$   
set  $\mathbf{y}_i^{(a)} = \boldsymbol{\eta}_i^{(a)} + \varepsilon_i^{(a)}$   
**end for**  
**end for**  
**for**  $r$  in 1:R **do**  $\triangleright R = 10,000$   
**for**  $i$  in 1:M **do**  $\triangleright M = 10,000$   
 $\tilde{i} \leftarrow$  sampled from uniform distribution on  $\{1, \dots, S\}$   
Set  $\tilde{\mathbf{y}}_i^{(FO)} = \mathbf{y}_{\tilde{i}}^{(FO)}$   
Set  $\tilde{\mathbf{y}}_i^{(InFO)} = \mathbf{y}_{\tilde{i}}^{(InFO)}$   
**end for**  
Set  $\text{FIM}_{\text{MC/AGQ}}^{(r)}(\zeta^{(FO)}) = \frac{1}{M} \sum_{m=1}^M h_{\text{AGQ}}(\tilde{\mathbf{y}}_m^{(FO)}, \boldsymbol{\beta})$   $\triangleright$  See Section 1.2.4  
Set  $\text{FIM}_{\text{MC/AGQ}}^{(r)}(\zeta^{(InFO)}) = \frac{1}{M} \sum_{m=1}^M h_{\text{AGQ}}(\tilde{\mathbf{y}}_m^{(InFO)}, \boldsymbol{\beta})$   
Set  $\text{relEff}_{\text{MC/AGQ}}^{(r)}(\zeta^{(FO)}, \zeta^{(InFO)}) = \left\{ \frac{\det(\text{FIM}_{\text{MC/AGQ}}^{(r)}(\zeta^{(FO)}))}{\det(\text{FIM}_{\text{MC/AGQ}}^{(r)}(\zeta^{(InFO)}))} \right\}^{1/p}$   
**end for**  
**return**  $(\text{relEff}_{\text{MC/AGQ}}^{(1)}(\zeta^{(FO)}, \zeta^{(InFO)}), \dots, \text{relEff}_{\text{MC/AGQ}}^{(R)}(\zeta^{(FO)}, \zeta^{(InFO)}))$

---

Algorithm 1 is applied to every model in Table 1.1, with ten different choices of  $\omega_1$  for each model. These values of  $\omega_1$  are equidistantly placed in the intervals indicated in Table 1.1, Specifically, they are generated by the R-command `seq(a,b,length.out=100)[1:10*10]` where  $a$  and  $b$  are the endpoints of the interval to which  $\omega_1$  belongs according to Table 1.1.

Figure 1.8 contains plots of the designs that are being compared (first and second row) as well as the corresponding bootstrap confidence intervals (third row). First note that most confidence intervals contain 1, so on a marginal basis no strong conclusion is evident. However, judging by the medians (or the inter-quartile range) of the sampled relative  $D$ -efficiencies, it seems that FO might be slightly worse than InFO in Exponential Decay I, at least for some values of  $\omega_1$ , though not for the highest. In Exponential Decay II, the two methods seem to yield designs of the same relative efficiency with the exception of the eighth and ninth intervals; both intervals are entirely above 1. This means that for each of these intervals, at least 95% of the 1,000 simulated MC/AGQ relative  $D$ -efficiencies of FO to InFO  $D$ -optimal designs are above 1, indicating superiority of the FO  $D$ -optimal design. This jump in relative  $D$ -efficiency coincides with a change in replication structure for the InFO  $D$ -optimal design from design seven to design eight (the second sampling time changes from being equal to zero to being strictly positive) and from design nine to design ten (the second sampling time becomes zero again and the third sampling time drops to a very small value close to zero). Hence, the most obvious explanation for the jump is that it is caused by the non-convexity of the InFO  $D$ -optimality objective function and the corresponding discontinuity in the design graph.

For the One-Compartment model, FO seems to perform as well as, or better than, InFO. For the Sigmoid Emax model, the two methods seem to perform equally well. Two further observations can be made: firstly, that designs that are quite different seem to have quite similar MC/AGQ relative  $D$ -efficiencies; this could be seen as an indication that, as with InFO, the MC/AGQ  $D$ -optimality objective function is not convex. Second, that the pairs of FO and InFO  $D$ -optimal designs with MC/AGQ relative  $D$ -efficiencies that are clearly different from 1 seem quite similar to pairs where the confidence interval comfortably contains 1. As an example of this latter phenomenon, consider the FO and InFO designs for the Exponential Decay I model, for  $\omega_1 = 0.21$  and  $\omega_1 = 0.24$ ; the FO designs for these two values of  $\omega_1$  are visually indistinguishable and the only appreciable difference between the InFO designs is that the first sampling time is slightly later for  $\omega_1 = 0.24$  than for  $\omega_1 = 0.21$ . Yet this is enough to cause

one of the most dramatic shifts relative  $D$ -efficiency in these plots. Another example is the Sigmoid Emax model for  $\omega_1 = 0.7$  and  $\omega_1 = 0.8$ .

These jumps in MC/AGQ relative  $D$ -efficiency, which are not accompanied by qualitatively different designs, could potentially be explained if the true objective function is subject to relatively radical changes in shape as a consequence of the change in  $\omega_1$ . That is, though we are evaluating pairs of designs that from a visual inspection seem to be close together, the effect of changing  $\omega_1$  could theoretically be such that the functions with which we evaluate these pairs are quite different. In the case of the Exponential Decay model, this could potentially be tied to the conjectured high degree of nonlinearity.

An alternative explanation is computational error; it is possible that the bootstrap confidence intervals does not reflect the true variance of the estimated

MC/AGQ relative  $D$ -efficiency. This would be a difficult question to resolve and is not within the scope of this work. A possible place to start would be with linear models given that they are the only case we know of for which analytical solutions are available.

Lastly, note that we chose to utilise bootstrap rather than simply increasing the number of simulations. This is motivated by computational cost. To illustrate the computational cost, consider the tabulated computation times in Table 1.2; here, we present the median time taken to compute the MC/AGQ  $D$ -optimality objective function (for a single design). The median is taken over the computation times for the designs considered in Figure 1.8.

Exp Dec I	Exp Dec II	Sigmoid Emax	One-comp
0.35 hrs	1.79 hrs	1.95 hrs	8.19 hrs

**Table 1.2:** *Computation times. Median time taken to compute the MC/AGQ  $D$ -optimality objective function for a single design using 10,000 simulations and five quadrature nodes.*

The computation times presented in Table 1.2 are for a single design. This indicates that optimising the MC/AGQ  $D$ -optimality function would

require a very large computational effort, even for these very simple models where the maximum number of parameters is three. In particular, note that it seems that the computation times increases exponentially with the number of parameters.

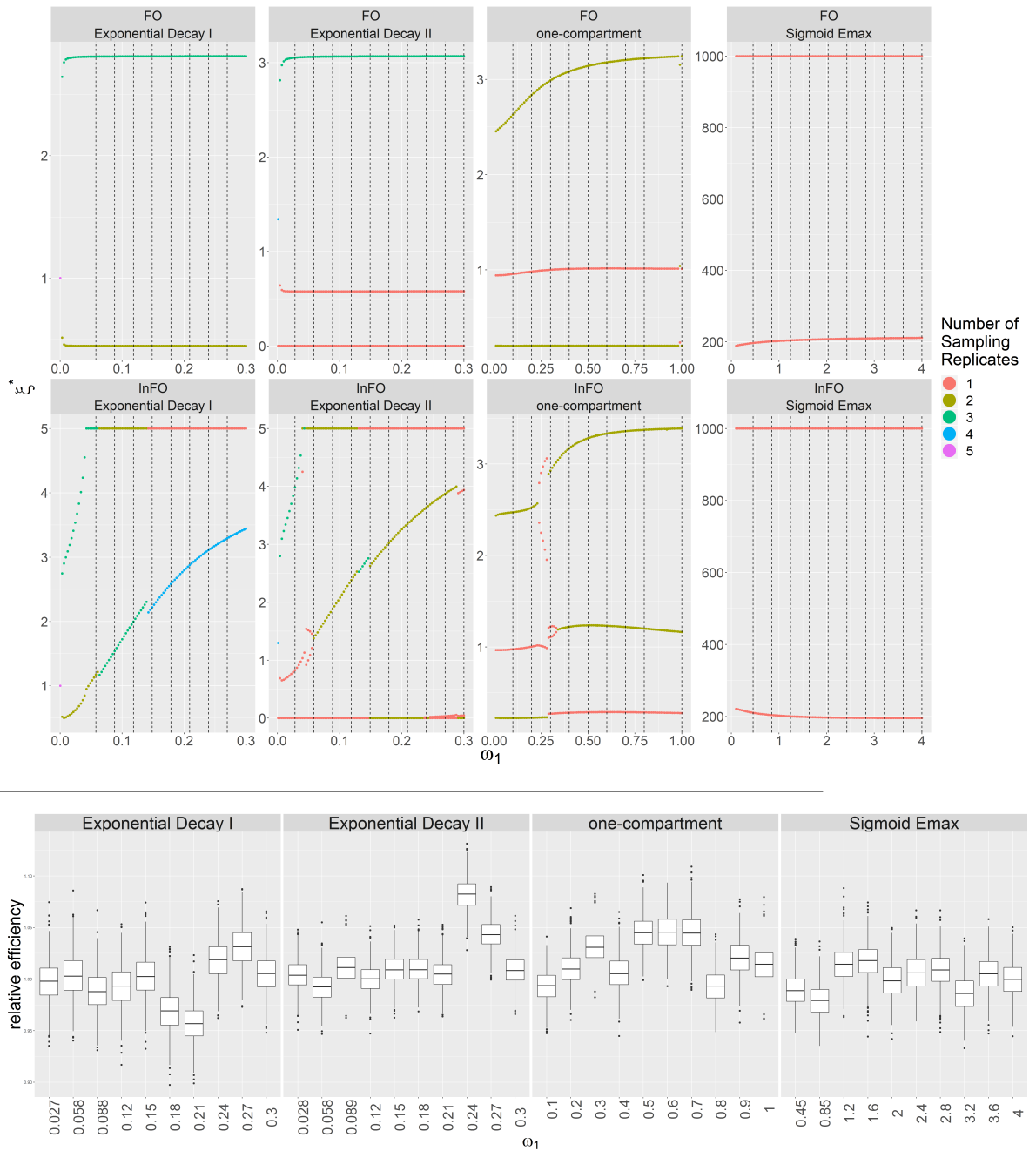
## 1.8 Discussion

In this chapter we present and implement a strategy to investigate the claim that InFO “performs better when between-individual variance is large”. We find FO and InFO  $D$ -optimal designs for a number of models commonly used in, e.g., the pharmaceutical industry and then establish that when between-individual variance is increased, FO and InFO yield qualitatively different designs. This motivates a further investigation and also necessitates a quantitative method of comparing designs.

FO and InFO  $D$ -optimal designs are quantitatively compared via MC/AGQ, a third FIM approximation. This is motivated by the fact that MC/AGQ avoids linearisation and is theoretically capable of arbitrarily high precision. For each model, we compute the MC/AGQ relative  $D$ -efficiency, along with bootstrap confidence intervals, for 10 pairs of FO and InFO  $D$ -optimal designs, each pair corresponding to a given value of between-individual variance. The resulting plots - of MC/AGQ relative  $D$ -efficiency against between-individual variance - indicate that InFO *does not* in general perform better than FO.

In terms of practical implications, this finding is in fact good news. InFO is computationally far more expensive than FO since it involves approximating a multi-dimensional integral. Moreover, the InFO  $D$ -optimality objective function is more prone to local optima and the implication is that optimisation compounds computational cost all the more.

A further point against InFO is an apparent lack of a theoretical justification for the choice of kernel, i.e., the distribution of  $\mathbf{b}$ , used to “robustify” the linearised FIM. The resulting quantity has no known interpretation as an approximation of anything. This point was elucidated further by Nyberg et al. (2012), who point out that FOCE relies on linearising around the mode of the subject-wise joint density, which InFO cannot do because



**Figure 1.8:** Final comparison of FO and InFO D-optimal designs and their relative efficiencies.

**First row:** FO D-optimal designs for each of the models in Table 1.1. The optimal design is plotted as a function of the variance of the first random effect  $V(b_1) = \omega_1 = \Omega_{11}$ .

**Second row:** InFO D-optimal designs.

**Third row:** Relative D-efficiency of FO D-optimal design to InFO D-optimal design. Computed for ten equi-distant values of  $\Omega_{11}$ . Confidence intervals constructed via percentile bootstrap intervals (1,000 samples with resample size 10,000).

it is not possible without data. However, a more appropriate way to compensate for the lack of data in this step would intuitively be to integrate with respect to the distribution of the mode, rather than the distribution of the random effect. An implementation of this idea is proposed in Nyberg et al. (2012).

This flaw in the construction of InFO could serve to explain why it does not seem to outperform FO in the cases we have examined. Moreover, it could explain some of the strange jumps in MC/AGQ relative  $D$ -efficiency that were noted in Figure 1.8. Two InFO  $D$ -optimality objective functions corresponding to two adjacent  $\omega_1$  values can have optima that are arbitrarily better or worse than the relatively constant FO  $D$ -optimal designs, when evaluated by MC/AGQ. This, as opposed to consistently improving in relative performance as  $\omega_1$  is increased, which was the expectation of the originators of the algorithm.

In our investigations, we find that both FO and InFO seem to perform poorly in the Exponential Decay models. It is possible that this is due to these models being “highly nonlinear”, i.e., that the linearised model is a poor approximation; if that is the case this suggests that in such cases neither FO nor InFO is appropriate. This might motivate a search for a third method for computing  $D$ -optimal designs. The MC/AGQ approach we used to compute relative  $D$ -efficiencies is a natural candidate as a third method but suffers from being extremely computationally expensive. Hence, optimisation of the MC/AGQ  $D$ -optimality objective function may not be feasible, especially given the fact that the function itself will be subject to random error and the optimisation method would have to take this into account.

### 1.8.1 Remark concerning scope

Our investigations are empirical in nature because we are dealing with problems where meaningful analytical results are typically unobtainable. In such cases, one must carefully consider the scope of one’s investigations. By the nature of the question, the potential number of cases and variants of the problem one might consider is infinite. However, when we are responding to a general claim that InFO is superior to FO for “large”



random effects, it makes sense to first establish that it holds in the simplest possible setting. If it is found to hold in the first set of problems, a follow-up investigation could be initiated to determine more specifically the gains that might be derived from implementing InFO. With this in mind, we consider this chapter an *initial* investigation, the purpose of which is to examine if we can expect InFO superiority to hold in *any* case at all.

We make two concrete simplifications to the general problem.

Firstly, we assume that only fixed effects are estimated and that variance components are known. The investigations we perform can all be carried out for the case where variance components are not known and have to be estimated. However, it would take somewhat more computational effort.

Second, for the purposes of this chapter we only treat problems in which all individuals are allocated to the *same* elementary experimental design. Hence, we do not treat cases where the notion of population design becomes relevant. According to Theorem 7.3 of Schmelter (2007), this restricted set of designs actually contains the optimal design if we allow *approximate* designs, i.e., designs with non-integer levels of replication. However, we restrict our search to exact designs, i.e., integer levels of replication, in which case the  $D$ -optimal population design may contain distinct elementary designs, cf. Example 7.4 of Schmelter (2007). It would require more complicated plots, as well as significantly more computational effort, to repeat these investigations under the more relaxed assumption of (potentially) distinct elementary designs.

In Chapter 3, where my R-packages (`doptim` and `randon`) are introduced, several examples are included where variance components are not assumed known. The relevant expansion of the theory is also covered in Chapter 3. Chapter 3 also includes an example of using `doptim` to find an “unbalanced population  $D$ -optimal design”, i.e. a design which is  $D$ -optimal over a set of population designs which can have different elementary designs for different individuals.

## 1.9 Further Work

Our work opens up a number of possible avenues of investigation. It also gives a framework for initial comparison of different methods for generating  $D$ -optimal designs.

In terms of further investigations, it would be highly desirable to develop a more systematic approach to evaluating when models are “too nonlinear”, especially given the severe impact of nonlinearity in the Exponential Decay cases. In fact, some measures of curvature have been suggested in the literature and O’Brien (2005) provides a nice overview of these. However, we did not find an NLME version of any of these, so this is may be an interesting area for futher development.

Investigating the feasibility of finding MC/AGQ  $D$ -optimal designs would also be of interest, in part exactly because it might provide a remedy for the cost of linearisation in cases when that cost is judged to be too high. A particular case, mentioned on page 30 of Nyberg (2011), is when the relationship between the fixed effects and the random effect is not linear, for instance if there is a power (Box-Cox) relationship instead (Pettersson et al., 2009). In this case, the FO approximation is not appropriate.

The approach we have taken in this chapter could be replicated in comparing other methods of FIM approximation. For instance, InFO is introduced in Retout and Mentre (2003) as an attempt to mimic the First Order Conditional Estimate (FOCE) method of estimation in NLME models, as per Lindstrom and Bates (1990). But Nyberg et al. (2012) suggest that InFO does not mimic FOCE well and propose an alternative FOCE-mimicking FIM approximation. This alternative approximation could be compared with FO in exactly the same manner as we have done in this work, provided it is not too computationally expensive.

Another avenue for further work would be to compare designs by using the actual performance of the MLE. This would be a direct way of comparing different design criteria for a specific choice of model and parameter values. A concrete place to start would be to simulate whole datasets from the assumed model, compute the MLE for each simulated dataset and then consider the empirical distribution of the MLE. For instance, the empirical distribution of the empirical covariance matrix could be computed and

this could in turn be used to estimate the volume of the distribution of the MLE. Such a volume estimate could be used to directly compare different designs. However, as with many of the results in this thesis, due to the empirical nature of the simulation approach it is difficult to make more than case-specific claims on such a basis. Still, one may use the results as a starting point from which to form conjectures about how results might generalise. Simulation studies of the kind described above have been performed in many papers, though they typically use parameter-wise performance metrics, e.g. root squared error, coefficient of variation or related quantities. See e.g. Table 4 in Nyberg et al. (2015), Table 9 in Wang et al. (2012) or Table 4 in Bazzoli et al. (2010).

## 2 A Practical Workflow Based On D-optimal Design

2.1	Introduction . . . . .	77
2.2	Contents of this chapter . . . . .	79
2.3	Methodological Background . . . . .	80
2.3.1	$D_A$ -optimality . . . . .	80
2.3.2	average Proportional Marginal Sensitivity . . . . .	81
2.3.3	Design Graphs . . . . .	83
2.3.4	Efficiency Profiles . . . . .	83
2.4	Motivating Example: One-Compartment Model . . . . .	84
2.5	Other Models . . . . .	101
2.5.1	Results . . . . .	102
2.6	Discussion . . . . .	126

## 2.1 Introduction

What does a practical application of  $D$ -optimal design look like? For locally  $D$ -optimal designs, the components are an NLME model, which presumably is scientifically justified, and a parameter guess. The question of how to pick a value for the parameter guess has so far been left open. If we revisit the example of exponential decay as a naive model for metabolism, the metabolic rate constant (and its variation between individuals) could be guessed either by using historical data from other experiments or by relying on the expertise of investigators.

Once a parameter guess has been made, it is natural to be concerned about the consequences of *parameter misspecification*; how much is lost if the guess is wrong? The most popular approaches to addressing the concern of parameter misspecification involve “robustification”; *pseudo-Bayesian*  $D$ -optimal designs achieve this by putting a prior on the parameter and integrating the objective function with respect to this prior, thus achieving a sort of weighted objective function. This criterion has been implemented for NLME models in the software PopED by Nyberg et al. (2012), who refer to it as ElnD-optimality. Another way of achieving designs that are robust to parameter misspecification is to employ *maximin* designs (see page 59 of Silvey (1980)). As the name suggests, maximin designs maximise the minimal efficiency, over a chosen set of possible parameters.

Finally, in many applications, interest is not in estimation of the full parameter, but rather in some subset of it; for instance, in the case of dose-response curves, the parameter estimate of scientific interest is often that of the  $EC_{50}$  (the dose at which 50% of the maximal effect is achieved). An example where this parameter occurs, in the Sigmoid Emax model, was covered in Section 1.3.3. In such cases, one can use the  $D_s$ -optimality criterion to generate experimental designs which minimise the variance of the corresponding subset of the MLE. In other applications, the quantity of interest may be a function of the parameters rather than a simple subset. For instance, the *area under the curve* (AUC) of a One-Compartment model may be calculated as a measure of total exposure to the drug. For cases such as this, the  $D_A$ -optimality criterion can be used to generate designs

that minimise the variance of the MLE plug-in estimator of the parameter transformation of interest.

While the techniques mentioned above address their motivating practical problems in theory, they have their own inherent problems in practice.

Pseudo-Bayesian design necessitates the elicitation of a prior and this is highly non-trivial for an NLME model; one could use conjugate priors but these are generally not motivated by an actual belief in the prior but rather by mathematical convenience. An exception to this criticism would be when there are relevant historical data available (e.g. in adaptive designs where a sequence of experiments is conducted and it is possible to take data from one experiment and use it to inform the design of the next one). In the fixed effects case, work has been performed by Bornkamp (2012) to construct priors that are “uniform in the functional space”, however, it is not clear how to expand this to include random effects. Moreover, pseudo-Bayesian designs with uniform priors (of any kind) and minimax designs share the issue that undue weight may be given to cases which are highly unlikely, potentially leading to an unacceptable rate of information loss in practice. Some criteria try to address the problem of disproportionately informative parameter regions. For instance, the Bias-API criterion in Nyberg and Hooker (2012) tries to, in the words of the author, “spread the design by forcing each sample to be responsible for a portion of the parameter distribution”.

In addition to the problems outlined above, many robust approaches carry a heavier computational cost that may well make optimisation unfeasible for even moderate numbers of parameters or time points.

As for  $D_s$ - and  $D_A$ -optimal designs, it is often the case in practice that experimental design is an iterative process whereby the statistical formulation of the objective of the experiment only gradually becomes apparent; if the above-mentioned robustified approaches are used, this would almost inevitably entail a laborious process of customising code several times in order to reflect the updated version of the objective.

The issues mentioned above highlight a need for a “middle road” approach that allows the practitioner to address the questions of parameter misspecification and changing experimental objectives in a flexible manner,

coupled with an intuitively meaningful way of choosing ranges of possible parameter misspecification.

In this chapter such a workflow is proposed, together with a detailed description as well as a step by step demonstration for the case of the One-Compartment model.

A second purpose of this chapter is to function as a showcase for the R-packages `doptim` and `randon` that were developed as part of this work (see Tvermosegaard (2018a) and Tvermosegaard (2018b)). While the `doptim` package allows the use of both FO and InFO (as well as pseudo-Bayesian criteria), all results in this Chapter have been achieved using the FO approximation, motivated by the findings in Chapter 1.

## 2.2 Contents of this chapter

In Section 2.3, the methodological background to the envisioned workflow is explained in depth and local  $D_s$ - and  $D_A$ -optimality are defined.

The *average Proportional Marginal Sensitivity (aPMS)* is introduced as a means to developing a scale - the *variance level (VL)* - for the magnitude of the variance components which makes intuitive sense to the practitioner. *Sensitivity bands* help to visualise the variation in the regression function related to a given choice of variance level.

*Design graphs*, as those used in Chapter 1, enable the user to assess whether different values within the chosen range of variance levels result in different designs. Design graphs could equally well be drawn for the fixed effects parameters; the choice to focus on variance components is motivated by the relatively sparse investigations that have so far been done for this aspect of NLME models.

*Efficiency profiles* enable the user to assess the impact of misspecifying the variance level (or, equivalently, the variance component), in terms of information lost. Efficiency profiles can also be used to identify which designs are most robust to misspecification and, lastly, they provide a quality check on the optimisation used to generate the  $D$ -optimal designs in the design graph.

Section 2.4 uses the One-Compartment model as a motivating example. The tools in Section 2.3 are used in combination to form a practically attractive workflow; it allows a focused check of design robustness to parameter misspecification which avoids “over-compensating” for parameter uncertainty as well as the “black box” aspects of robustified criteria. Local  $D_s$ - and  $D_A$ -optimality are easily implemented in this framework and require minimal coding to apply in the packages `doptim` and `randon` developed for this work. The resulting graphs illustrate that  $D$ -,  $D_s$ - and  $D_A$ -optimal designs can be quite different from each other. Specifically, depending on which variance component is inflated and which parameter subset or transformation one designs for, a misspecified variance component can result in the loss of information in tens of percentage points and different guesses about the variance component can yield designs that are radically different (as was seen in Chapter 1).

In Section 2.5, design graphs and efficiency profiles are drawn for a number of models taken from the nonlinear regression literature. These plots allow an extensive investigation of the relative merits of ranges of designs and this serves to illustrate the wide applicability of the proposed workflow. It also demonstrates that the `randon` package makes these investigations feasible.

## 2.3 Methodological Background

### 2.3.1 $D_A$ -optimality

Note that  $D_s$ -optimality is a special case of  $D_A$ -optimality, so it suffices to define the latter. Following O’Brien (2005), suppose interest is in  $\phi(\boldsymbol{\beta})$ , where  $\phi : \mathbb{R}^p \rightarrow \mathbb{R}^s$  is some potentially nonlinear function and  $s \leq p$ . Let  $\mathbf{A} = \frac{\partial \phi}{\partial \boldsymbol{\beta}}$  and suppose that this  $s \times p$  matrix is of full rank.

Then the  $D_A$ -optimality objective function is

$$\psi_{\mathbf{A}}(\boldsymbol{\zeta}) = \log \det(\mathbf{A}^{\top} \text{FIM}(\boldsymbol{\zeta})^{-1} \mathbf{A}) \quad (2.1)$$

and so a  $D_A$ -optimal design  $\boldsymbol{\zeta}^*$  minimises this objective function,  $\boldsymbol{\zeta}^* \in \text{argmax}_{\boldsymbol{\zeta}} \psi_{\mathbf{A}}(\boldsymbol{\zeta})$ . A simple delta-method argument can be employed to



justify this objective function (see pages 49-50 of Schmelter (2007) as well as page 385 of Rao et al. (1973)): Suppose that

$$\hat{\beta} \stackrel{as.}{\sim} \mathcal{N}(\beta, \text{FIM}(\xi)^{-1}),$$

then the MLE plug-in estimator of  $\phi(\beta)$  is also asymptotically normal

$$\phi(\hat{\beta}) \stackrel{as.}{\sim} \mathcal{N}(\phi(\beta), \mathbf{A}^\top \text{FIM}(\xi)^{-1} \mathbf{A}).$$

The relative quality of two designs  $\xi, \xi'$  can be measured by their relative  $D_{\mathbf{A}}$ -efficiency. Given a transformation of interest  $\phi: \mathbb{R}^p \rightarrow \mathbb{R}^s$ , the relative  $D_{\mathbf{A}}$ -efficiency of  $\xi$  to  $\xi'$  is

$$\text{relEff}_{\mathbf{A}}(\xi, \xi') = \exp(\psi_{\mathbf{A}}(\xi') - \psi_{\mathbf{A}}(\xi))^{1/s}$$

If  $\xi'$  is the  $D_{\mathbf{A}}$ -optimal design, the ‘relative’ is dropped and one refers to the “ $D_{\mathbf{A}}$ -efficiency of  $\xi$ ”.

### 2.3.2 average Proportional Marginal Sensitivity

Here follows a constructive argument behind the definition of the *average Proportional Marginal Sensitivity (aPMS)*.

The partial derivative of the regression function with respect to a single fixed effects parameter is here referred to as the *marginal sensitivity* (the derivative of the regression function with respect to the full fixed effects parameter vector is sometimes referred to as the parameter sensitivity (Atkinson, p. 251)). The marginal sensitivity can be thought of as measuring the change in the regression function per unit change in the parameter.

For a given NLME model, the marginal sensitivity is given by

$$MS_i(t, v) = \left. \frac{\partial \eta}{\partial \beta_i} \right|_{\beta = \psi_i(v)},$$

where, for  $i, j = 1, \dots, p$ ,

$$\{\psi_i(v)\}_j = \begin{cases} \beta_j, & \text{for } j \neq i \\ v & \text{for } j = i \end{cases} .$$

That is, the marginal sensitivity of  $\eta$  with respect to the  $i$ th fixed effects parameter is defined as the partial derivative of  $\eta$  with respect to the  $i$ th fixed effects parameter, considered as a function of sampling time  $t$  and the  $i$ th fixed effects parameter  $v$ . The  $i$ th fixed effect is represented by  $v$  to highlight that it is not considered fixed whereas the other fixed effects parameters are considered fixed.

In order to construct a relative measure of change in the regression function, the marginal sensitivity is normalised with respect to the value of the regression function, i.e., it is transformed into the point-wise percentage change in the value of the regression function per unit change in the parameter. This yields the *proportional marginal sensitivity*  $PMS_i(t, v) = MS_i(t, v) / \eta(t, \psi_i(v))$ . Lastly, in order to arrive at a scalar-valued measure of change in the regression function, the proportional marginal sensitivity is averaged over the time interval of interest  $[a, b]$ , yielding the *average Proportional Marginal Sensitivity*

$${}_aPMS_i(v) = \frac{1}{b-a} \int_a^b PMS_i(t, v) dt . \quad (2.2)$$

Note that  $[a, b] \subseteq [0, \infty)$  and the specific values of  $a$  and  $b$  will relate to the context in which the NLME model is being applied. For instance, if a reaction described by a One-Compartment model is planned to be run for a set amount of time then  $a = 0$  and  $0 < b < \infty$  will be the time where the reaction is stopped. In other words,  $[a, b]$  will typically be the time interval from which sampling times may be taken. The aPMS is interpreted as a type of normalised rate of change, in the set of regression functions  $\{t \mapsto \eta(t, \beta) : \beta \in \mathbb{R}^p\}$ . Its intended application is as follows:

1. Let  $\rho$  be the variance level related to  $\beta_i$ . This value can e.g. be elicited by asking the practitioner how much between-individual variation in the response profile (in terms of percent deviation from the typical

profile) is expected to be related to between-individual variation in  $\beta_i$ .

2. One can then plug  $v$  into the left-hand side of Equation (2.2) and solve for  $v$ , yielding  $\delta$ . The solution  $\delta$  can be interpreted as the change in  $\beta_i$  which results in a deviation in the regression function of magnitude  $\rho$ , relative to the typical profile.
3. The variance of  $b_i$  is set to  $\frac{1}{2}\delta^2$ , so that twice the standard deviation of the random effect equals  $\delta$ .

Thus, the aPMS facilitates the elicitation of guesses pertaining to the value of the variance components by allowing the question to be framed in terms of between-individual deviation in profile related to each parameter.

Note that the R-implementation of aPMS uses the first order approximation

$$\int_a^b PMS_i(t, v) dt \simeq (v - \beta_i) \int_a^b PMS_i(t, \beta_i) dt ,$$

reducing the problem of solving for  $v$  to

$$\delta_i = (v - \beta_i) = \frac{\rho}{aPMS_i(\beta_i)} , \quad (2.3)$$

where  $v$  is the desired variance level.

### 2.3.3 Design Graphs

The same definition of design graph is used here as in Subsection 1.6.1. This chapter utilises the one-to-one correspondence between variance level  $\rho_i$  and variance component  $\omega_i$  to draw design graphs as a function of the variance level. In all examples in this chapter, all other parameters are kept fixed; the fixed effects parameters are fixed at model-specific values and the variance components are fixed at a variance level of ten percent.

### 2.3.4 Efficiency Profiles

Efficiency profiles are used to evaluate the robustness of designs to misspecification of a parameter. In the case of this chapter, the parameter in question is the variance of the random effect used to generate the design

graphs. For each variance level in a given design graph in Figure 2.2, the  $D$ -efficiency of all the other designs in the design graph is evaluated and this generates an efficiency profile. Thus one efficiency profile per design in the design graph can be drawn.

This can be outlined mathematically. Let  $\psi_{\mathbf{A}}(\rho, \xi) = \log \det[\mathbf{A} \mathbf{FIM}_{FO}(\rho, \xi) \mathbf{A}^{\top}]$  be the chosen approximation of the  $D_{\mathbf{A}}$ -optimality objective function, considered as a function of the variance level  $\rho$  of  $b_1$  and of the elementary design  $\xi$ . Let  $\xi^*(\rho)$  denote the  $D_{\mathbf{A}}$ -optimal design corresponding to  $\rho$ , i.e.,  $\xi^*(\rho) = \operatorname{argmax}_{\xi} \psi(\rho, \xi)$ . Then, for a given  $\rho' \in (0, \infty)$ , the efficiency profile for  $\xi^*(\rho')$  is given by the graph of the map

$$v \mapsto \exp \left[ \psi_{\mathbf{A}}(\rho', \xi(\rho)) - \psi_{\mathbf{A}}(\rho', \xi(\rho')) \right]^{1/s}.$$

In words, this graph addresses the question “if the correct guess as to parameter values has been made for all parameters, with the exception of  $\omega_1$  (via the specification of  $\rho$ ), what is the information loss from using a  $D_{\mathbf{A}}$ -optimal design for a different  $\rho$ ?”.

## 2.4 Motivating Example: One-Compartment Model

The One-Compartment model is used as a motivating example because it appears as such in many publications dealing with optimal experimental designs for nonlinear regression models.

In particular, note that Dokoumetzidis (2006) produced plots very similar to the design graphs found here, using the One-Compartment model.

Therefore it is well suited as an example to demonstrate the techniques employed in this chapter.

This example provides an idealised version of the imagined workflow:

1. A mechanistic model is chosen along with an initial parameter guess.
2. aPMS is used to draw sensitivity bands which illustrate the impact of a chosen variance level for the random effect on the variability in the response curve.

3. Optimal designs are identified, either for the full fixed effects parameter vector, a sub-vector of it, or a nonlinear transformation of it.
4. Efficiency profiles are drawn to enable assessment of the robustness of the identified designs to misspecification in the parameter of interest, as well as to provide a quality check on the optimisation itself.

### Regression Function

The regression function for this model, and the chosen value of  $\beta$  can be found in Table 2.1. Note that unless otherwise specified the values used for  $\omega$  and  $\sigma^2$  are chosen such that the corresponding variance levels equal ten percent.

### Sensitivity Bands

This is the first example of how sensitivity bands are constructed (a function to generate these plots is included in the `randon` package). Say that the consulting statistician is informed by the scientist client that batch-to-batch variation in the first rate constant ( $\exp \beta_1$ ) is likely to result in approximately ten percent variation in the response profile. Note that in eliciting this information from the scientist, the statistician must be careful to explain that the client is meant to think in terms of average relative variance across the entire response curve. The statistician can then solve equation (2.3) for  $\delta$ , with  $\rho = 0.1$ , in order to obtain the corresponding, say,  $\delta_{\beta_1}$ , the value which, when added to  $\beta_1$ , will result in a ten percent deviation from the population level regression function. Sensitivity bands are then given by evaluating the regression function  $\eta$  in  $\psi_1(\pm\delta_{\beta_1}) = (\beta_1 \pm \delta_{\beta_1}, \beta_2, \beta_3)^\top$  over the full range of  $t \in (0, 4)$ . In this case, the upper ten percent sensitivity band is given by

$$t \mapsto \eta(\psi_1(-0.064), t) = \eta((-0.064, 1, 1)^\top, t) \quad (t \in (0, 4))$$

and the lower band is given by

$$t \mapsto \eta(\psi_1(0.064), t) = \eta((0.064, 1, 1)^\top, t) \quad (t \in (0, 4)).$$

The first plot in Figure 2.1 contains sensitivity bands for the parameter  $\beta_1$ . [NB: language note: sensitivity is for parameters not for random effects; sensitivity is used to decide variance of additive random effects]. The bands are very narrow around the regression function before the peak, after which they widen a little and then seem to stabilise.

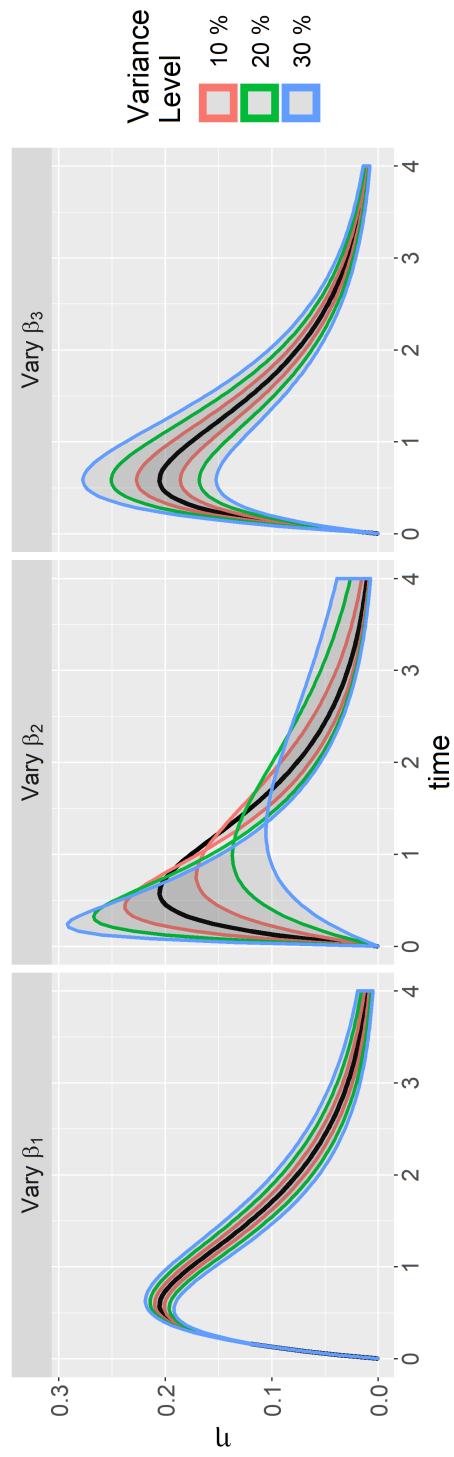
The second plot in Figure 2.1 contains sensitivity bands for the parameter  $\beta_2$  and illustrates that sensitivity bands do not necessarily contain the original profile. Moreover, the time of the maximum response clearly depends on the value of  $\beta_2$  and the 'flatter' band appears to be almost concave. In the first plot, all qualitative aspects of the profile remained the same for all variance levels.

This also serves as an occasion to make a cautionary statement regarding the use of variance levels; if we contrast with the case where we varied  $\beta_1$ , the variation in the second plot arguably looks larger, even though the same variance levels were used. This can be explained by a number of factors; for one, small (or large) function values can dominate the averaging step in aPMS, making bands seem counter-intuitively narrow (or wide). Further, there is a first order approximation (i.e., linearisation) step involved in the implemented version of aPMS which will induce some error. Thus, there is a limit to the direct interpretability of 'variance level'. A wise user will therefore always draw sensitivity bands such as these in order to visually assess the impact of the chosen level of variance.

The third plot in Figure 2.1 contains sensitivity bands for the parameter  $\beta_3$ . Given that varying  $\beta_3$  essentially corresponds to scaling the regression function (as  $\exp(-\beta_3)$  enters linearly into the regression function), one could have predicted these sensitivity bands.

### Design Graphs

In Figure 2.2, design graphs for the One-Compartment model are presented. As a guide to reading these plots, consider the left-hand panel: Here, the variance of the random effect on  $\beta_2$  and  $\beta_3$ , as well as the variance of the error term, are all set such that the corresponding variance level is



**Figure 2.1:** Sensitivity bands for the One-Compartment model. In all plots, the black line is the regression function for the true parameter,  $\boldsymbol{\beta} = (0, 1, 1)^\top$ .

ten percent. The variance of the random effect on  $\beta_1$  is set such that the variance level equals the value on the  $x$ -axis.

Thus, for each value on the  $x$ -axis, the full parameter vector for the model is specified. The  $D$ -optimal design for  $\beta$  can then be computed. The values (in time units) of the design points are given by the  $y$ -axis, with replication level indicated by the colour of the points.

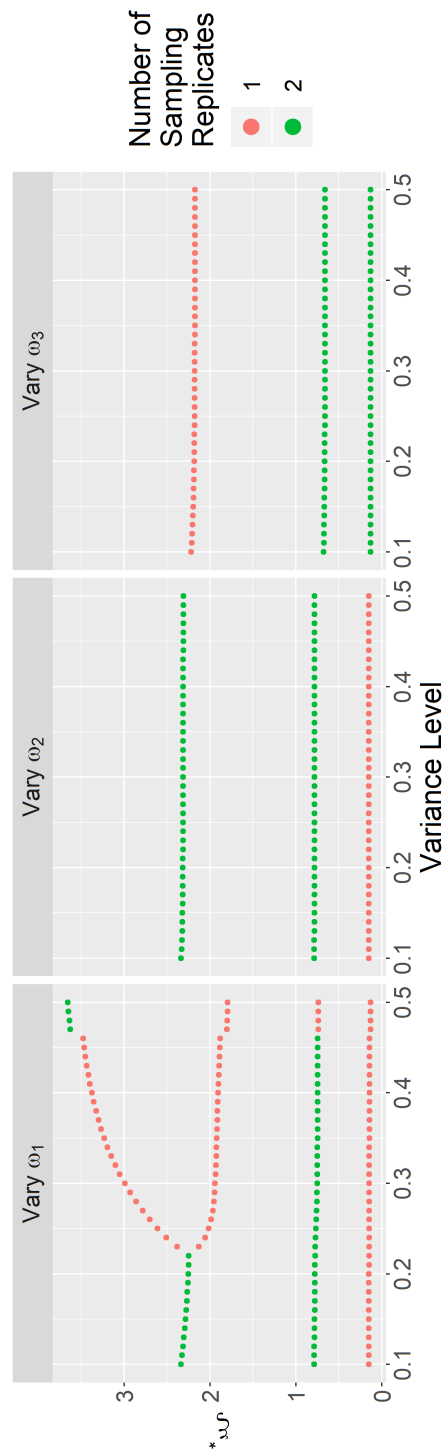
So the interpretation of the left-hand panel in Figure 2.2 is that the  $D$ -optimal design is fairly stable for variance levels of  $b_1$  between ten percent and a little more than twenty percent, with three distinct design points where the first design point is replicated once in the design and the two other design points are replicated twice. For variance levels of  $b_1$  between twenty and forty-five percent, the designs have four distinct design points, with the second earliest time being replicated twice and the remaining three times being replicated once. Lastly, for variance levels between forty-five and fifty percent, the replication structure (using the notation from Subsection 1.1.2) shifts from  $r = (1, 2, 1, 1)$  to  $r = (1, 1, 1, 2)$ , while the design points themselves remain almost unchanged (there is a slight discontinuous jump for the last two times, but it appears small in the plot).

This example thus nicely demonstrates three behaviours that are not uncommon with design graphs: bifurcation, replication shifts and discontinuities.

The other two panels in Figure 2.2, where the variance level is increased for  $\omega_2$  and  $\omega_3$  respectively, suggest that the  $D$ -optimal design is (almost) the same for all of the corresponding parameter choices. This is unsurprising in the case of  $\omega_3$  because  $\beta_3$  (and  $b_3$ ) enters linearly into the regression function; in the case of ordinary nonlinear regression it holds that for partially linear models, the value of the linear parameter does not influence the  $D$ -optimal design (see Khuri (1984)). Therefore, it is not surprising if similar behaviour manifests here.

For  $\omega_2$  it is interesting to note that even though  $\beta_2$  enters into the regression function in a way which is quite similar to  $\beta_1$ , the  $D$ -optimal design does not seem to be impacted by the variance level of  $\omega_2$ . Considering the sensitivity bands in Figure 2.1, one might have expected the reverse





**Figure 2.2:** Design graphs for the One-Compartment model. The x-axis gives the variance level in percent, for the random effect mentioned in the respective panels of the plot. The y-axis gives the design points in the corresponding D-optimal design for  $\beta$ , with the colour of the individual point indicating how many times the design point is replicated.

result, i.e., that the  $D$ -optimal design depends a lot on  $\omega_2$  and not very much on the variance on  $\omega_1$ . This serves as a demonstration of one of the motivations for using  $D$ -optimal designs: intuition is a poor guide in assessing the relationship between parameter values and their estimators.

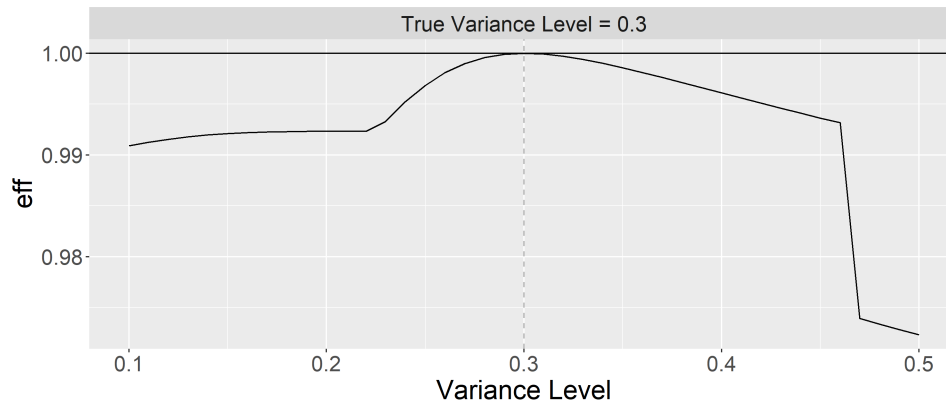
Once design graphs for the cases of interest have been computed, it is a good idea to check the robustness of the designs to misspecification of the parameters; one of the weaknesses of  $D$ -optimal designs for nonlinear models is that they depend on the parameter.

### Efficiency Profiles

Figure 2.3 is the efficiency profile for the design in the left-hand panel of Figure 2.2, where  $\rho = 30\%$ . Per construction, the efficiency of that design, indicated by the vertical dashed line in Figure 2.3, is 100%. The designs in the left-hand panel of Figure 2.2 which have the same form, i.e., which belong to the bifurcation with four distinct sampling times and replication structure  $r = (1, 2, 1, 1)$ , have efficiencies that are very close to 100%. The designs before the bifurcation, i.e., corresponding to the lowest variance levels, all have efficiency over 99%. Lastly, the designs that correspond to the highest variance levels, i.e., after the (discontinuous) replication shift, cause a sharp decrease in the efficiency although it still remains above 97%. Overall, the conclusion from this plot is that the efficiency profile indicates that if the true variance level is 30%, then all of the generated  $D$ -optimal designs are highly informative, i.e., misspecification of  $\rho$  is unproblematic.

Every design in a design graph can be used to generate an efficiency profile. In the first row of plots in Figure 2.4, all the efficiency profiles for the designs in the left-hand panel of Figure 2.2 have been plotted. They have been bundled together, for ease of visualisation, so that each panel contains several efficiency profiles.

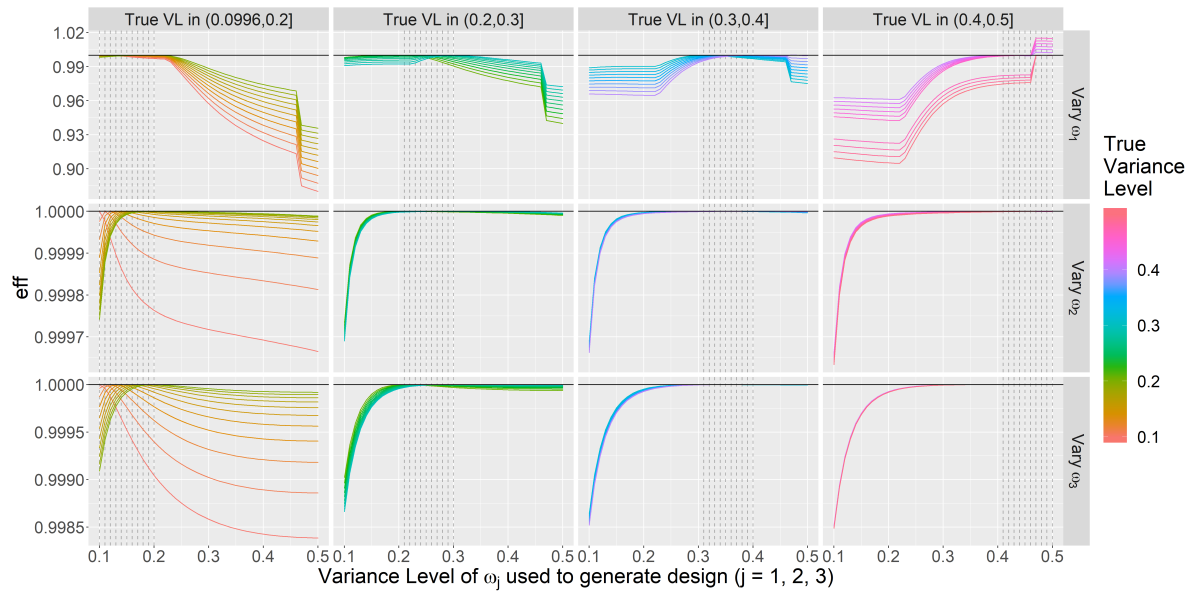
Taken together, the profiles reveal which designs result in the highest minimal efficiency. A second use of efficiency profiles is as a quality check of the optimisation; if any efficiency profile contains values over 100%, it may be concluded that the design which was used to generate the efficiency profile in question is not optimal, i.e., the optimisation failed.



**Figure 2.3:** *D*-efficiency of designs in the left-hand panel of Figure 2.2 when the true variance level is 30 %. The x-axis is the variance level used to generate the design under consideration; this design can be identified by looking up the design for the same variance level in the left-hand panel of Figure 2.2. The y-axis is the *D*-efficiency of the design under consideration. The dashed line goes through  $\rho = 30\%$ , i.e., the true variance level.

With this in mind, the considered profiles yield the following conclusions: unsurprisingly, the worst cases of misspecification lead to the highest loss of information; if the true variance level is low (left-most panel) and a design corresponding to the highest variance levels is used, the information loss is about 10%. If the variance level is truly high (right-most panel) and a low-variance design is used, the information loss is of a similar magnitude.

The right-most panel contains efficiency profiles with efficiencies over 100%. This implies that when the true variance level is above approximately 40%, the designs with the shifted replication structure  $r = (1, 1, 1, 2)$  have a higher value by the objective function than the design yielded by the optimisation and which is included as the *D*-optimal design in the design graph. This indicates that the optimisation procedure has trouble locating the appropriate region of the design space in terms of replication structure. The shifted designs are only evaluated as less than 2% more efficient however, so the cost in terms of obtaining maximal information is small. Also, with the knowledge provided by these efficiency profiles, the user



**Figure 2.4:** Efficiency profiles for the designs in Figure 2.2. The x-axis gives the variance level for the random effect mentioned in the caption of the relevant panel of Figure 2.2. The y-axis gives the D-efficiency of all the designs in the design graph mentioned in the sub-caption, relative to the D-optimal design for the true variance level. The dashed lines go through the variance levels which are used as the true variance levels for the plot.

can resort to the more efficient designs that the optimisation procedure failed to identify at the relevant variance level.

Rows two and three of Figure 2.4 are unsurprising in light of the near constant design graphs to which they correspond. The range of efficiency is between 99.9% and 100%, implying that any difference between the designs is negligible. While the almost imperceptible difference in the designs for the lowest variance levels does cause a corresponding non-constant set of efficiency profiles, there are no features of practical interest in these plots.

### Design graphs for $D_s$ -optimal designs

As mentioned in subsection 2.3.1, interest may be in estimating a subset of the parameter, or a transformation of the parameter. For such cases, the  $D_A$ -optimality objective function has been implemented in `randon` which allows the user to produce design graphs and efficiency profiles just as in the previous section where  $D$ -optimal designs for the full parameter vector  $\beta$  were identified.

In Figure 2.5, design graphs are plotted of  $D_s$ -optimal designs for the One-Compartment model. There is a column of plots per subset of size two of  $\beta$ , i.e., the first column contains  $D_s$ -optimal designs for  $(\beta_1, \beta_2)^\top$ , the second for  $(\beta_1, \beta_3)^\top$  and the third for  $(\beta_2, \beta_3)^\top$ . In each row of plots, a different variance component is varied to generate the design graph, in the first row,  $\omega_1$  is varied, in the second  $\omega_2$  and in the third it is  $\omega_3$ .

Similarly to Figure 2.2, the design graphs of  $D_s$ -optimal designs in Figure 2.5 are almost constant for variance components  $\omega_2$  and  $\omega_3$ . This is with exception of the design graph for varying  $\omega_2$  and  $s = (\beta_1, \beta_3)^\top$ , i.e., the design graph in the second row and second column of plots. In fact, this design graph resembles a horizontal translation of the design graph for varying  $\omega_1$  and same  $s$ . As previously mentioned,  $\beta_1$  and  $\beta_2$  appear almost symmetrically in  $\eta$  so it is tempting conjecture that varying  $\omega_1$  and varying  $\omega_2$  give rise to the same set of design graphs, up to a horizontal translation.

When  $\omega_1$  is varied, i.e., for the first row of design graphs, the choice of  $s$  clearly impacts the qualitative characteristics of the design graph. For the different choices of  $s$ , not one design is repeated between columns, nor do any two design graphs even contain the same replication structure. For

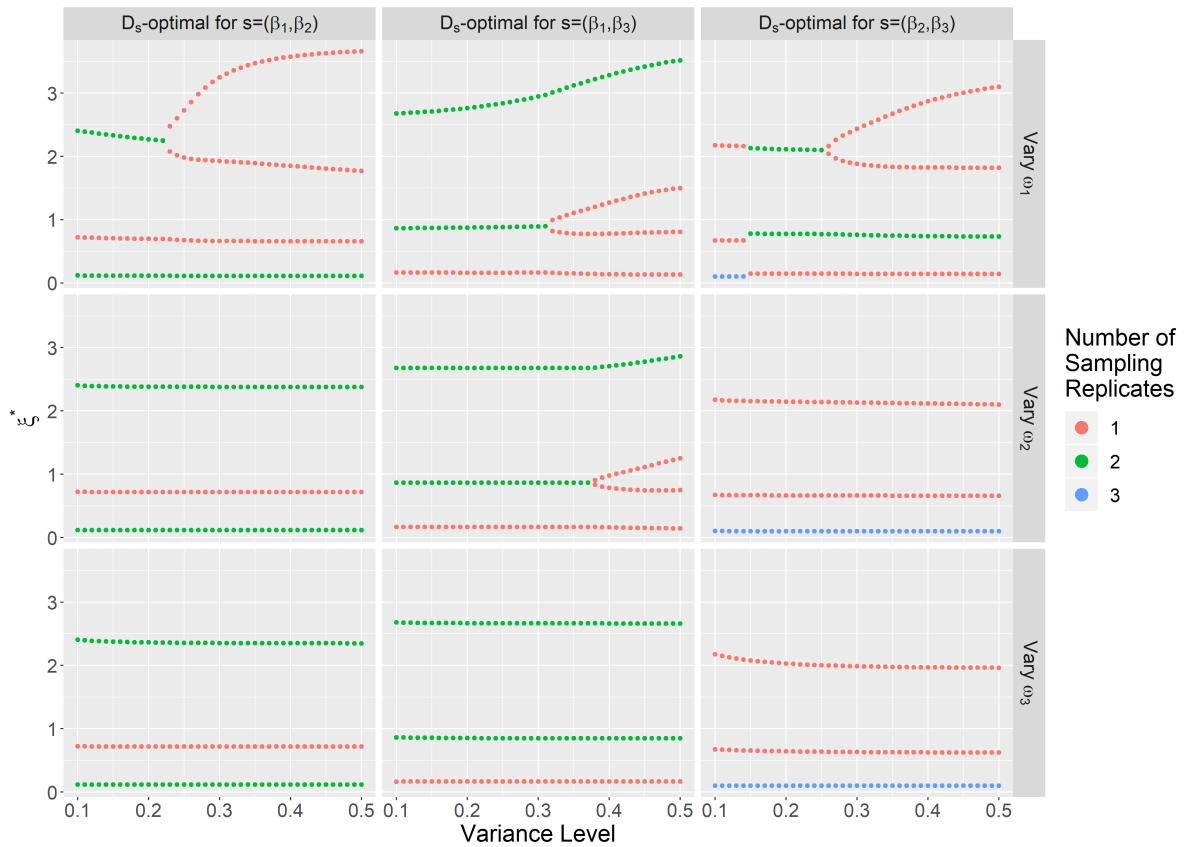
varying  $\omega_1$  and  $s = (\beta_1, \beta_2)^\top$ , the replication structures in the design graph are  $r = (2, 1, 2)$  and  $r = (2, 1, 1, 1)$ . For varying  $\omega_1$  and  $s = (\beta_1, \beta_3)^\top$ , the replication structures in the design graph are  $r = (1, 2, 2)$  and  $r = (1, 1, 1, 2)$ . For varying  $\omega_1$  and  $s = (\beta_2, \beta_3)^\top$ , the replication structures in the design graph are  $r = (3, 1, 1)$ ,  $r = (1, 2, 2)$  and  $r = (1, 2, 1, 1)$ .

It can be concluded that  $D_s$ -optimal designs for this model differ substantially depending on the subset of interest  $s$ , and that sensitivity of the  $D_s$ -optimal design to different variance components ranges from high (in the case of  $\omega_1$ ), to low (in the case of  $\omega_2$ ), to negligible (in the case of  $\omega_3$ ). Lastly, it is consistently the case that the number of sampling points increases (weakly) with the magnitude of the variance component. It is again noted that for the parameter with the “least interesting” looking sensitivity bands, namely  $\beta_1$ , the corresponding variance component,  $\omega_1$ , yields the “most interesting” looking design graphs.

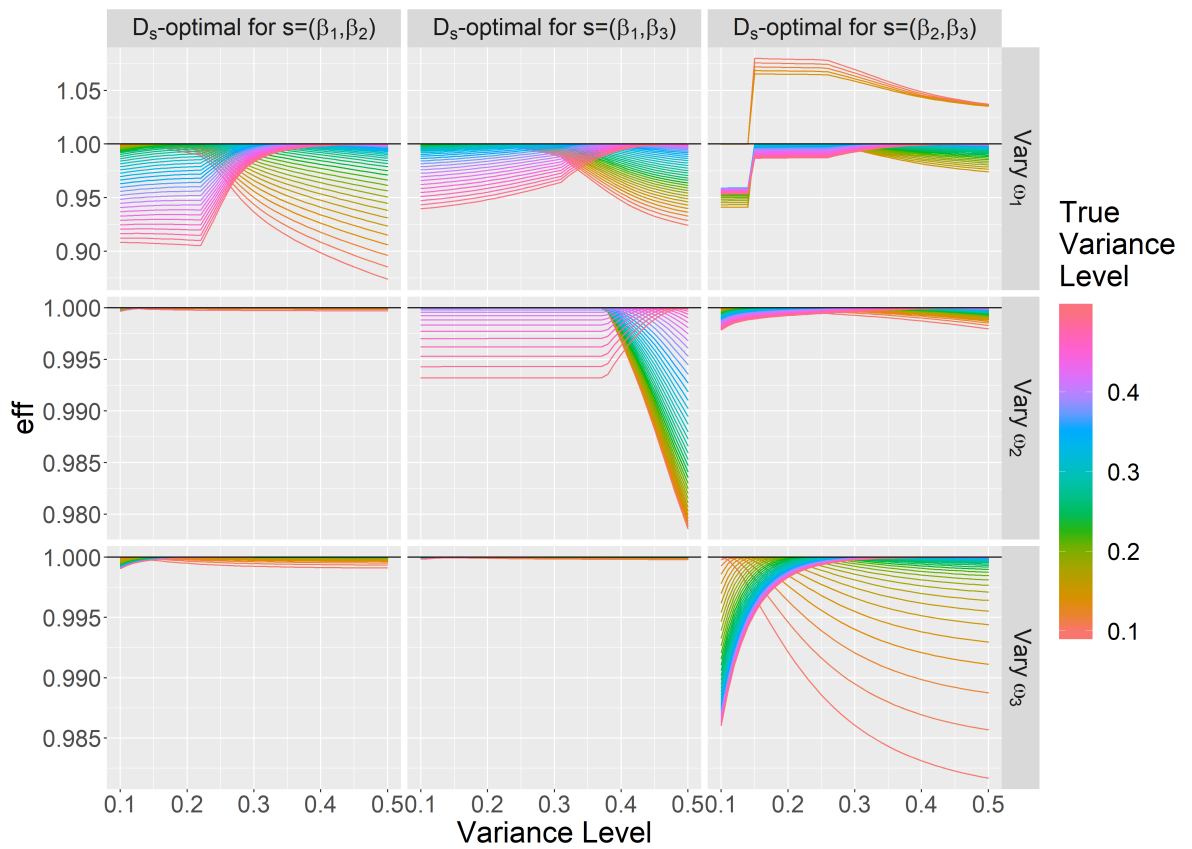
### Efficiency profiles for $D_s$ -optimal designs

As before, efficiency profiles can be drawn for the design graphs presented in Figure 2.5. These are presented in Figure 2.6, this time with all profiles for each design graph contained in one plot per design graph. The plot in the first row, third column, i.e., for varied  $\omega_1$  and  $s = (\beta_2, \beta_3)$  clearly indicates that the optimisation failed for the lower variance level because they contain efficiency values above one. It is clearly preferable to use designs corresponding to higher variance levels. A more thorough optimisation could be employed but it is likely to yield designs very similar to the ones with  $r = (1, 2, 2)$  directly to the right of the sub-optimal ones in the design graph.

Secondly, the plots reveal that when  $\omega_1$  is varied, the designs that are most robust to misspecification of  $\omega_1$  are near the middle of the design graph, i.e., corresponding to the variance levels where all efficiency profiles are relatively close to one (where there is a “peak” in the space under the efficiency profiles). Beyond these observations, there is little of interest to note in the efficiency profiles, especially given the near constant design graphs the other profiles correspond to.



**Figure 2.5:** Design graphs for  $D_s$ -optimal designs for the One-Compartment model. The x-axis gives the variance level in percent, for the random effect mentioned in the title of the respective panels. The y-axis gives the design points in the corresponding  $D_s$ -optimal design, with the colour of the individual point indicating how many times the design point is replicated. The subset of the parameter for which the  $D_s$ -optimal design is optimal is given by the top panels. The variance parameter  $\omega_j$  which is varied in order to generate the design graph is given on the side panels.



**Figure 2.6:** Efficiency profiles for the  $D_s$ -optimal designs in Figure 2.5. The  $x$ -axis is the variance level used to generate the design under consideration; this design can be identified by looking up the design for the same variance level in Figure 2.5. The  $y$ -axis is the  $D_s$ -efficiency of the design under consideration, when the true variance level is as indicated by the colour of the line. Each line is a plot of the  $D_s$ -efficiency of all the designs in the corresponding design graph, under the assumption that a particular variance level (indicated by the colour of the line) for a particular variance component (indicated by the side panel) is true. Variance components that are not varied are kept at a variance level of ten percent.



### Design graphs and Efficiency Profiles for $D_A$ -optimal designs

Transformations of interest for the One-Compartment model include the *Area Under the Curve* ( $AUC$ ) and the *time to maximum concentration* ( $t_{max}$ ). Both of these quantities admit explicit expressions and it can be shown that they are respectively given by

$$AUC = \frac{e^{\beta_2 - \beta_3}}{e^{\beta_2} - e^{\beta_1}} (e^{-\beta_1} - e^{-\beta_2}) \quad (2.4)$$

and

$$t_{max} = \frac{\beta_1 - \beta_2}{e^{\beta_1} - e^{\beta_2}}. \quad (2.5)$$

$D_A$ -optimal designs for  $AUC$ ,  $t_{max}$  and for the vector  $(AUC, t_{max})$  are presented in Figure 2.7 and efficiency profiles are presented in Figure 2.8.

The design graph for  $AUC$  has only two distinct sampling times in the first design, with replication structure  $r = (1, 4)$ . In nonlinear regression without fixed effects, this would usually imply that the fixed effects parameter  $\beta$  cannot be estimated (since the model under consideration has three fixed effects parameters) and so it must be assumed that the same thing holds for NLME models, at least in the case where only one individual is sampled. This behaviour is not exceptional; in (O'Brien, 2005), the author found several  $D_A$ - and  $D_s$ -optimal designs for nonlinear regression models that make estimation of  $\beta$  impossible.

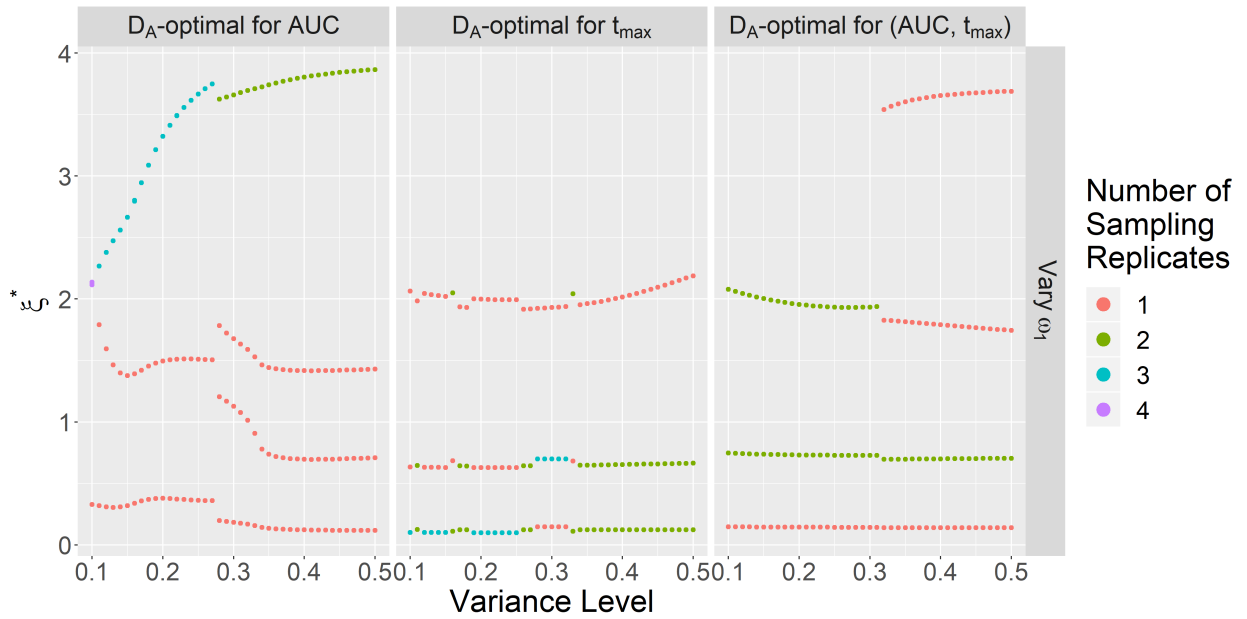
The design graph for  $t_{max}$  is somewhat noisy-looking. It suggests that there are several local optima with different replication structures, but quite similar sampling times, that are hard for the optimisation algorithm to distinguish between. Specifically, the graph seems to vacillate between the three replication structures  $r = (3, 1, 1)$ ,  $r = (2, 2, 1)$  and  $r = (1, 3, 1)$  before settling on the second one.

For variance levels below thirty percent, the design graph for the vector  $(AUC, t_{max})$  looks similar to the design graph for  $t_{max}$ , albeit with a different replication structure. However, it resembles the design graph for  $AUC$  in that at approximately thirty percent variance level, there is a discontinuous bifurcation, which adds a sampling time near the upper bound of the time interval.

The three efficiency profile plots in Figure 2.8 each exhibit their own interesting behaviour. In the case of  $AUC$ , the profiles suggest that misspecification of the variance level is more serious if one wrongly specifies a low variance level than if one wrongly specifies a high variance level; all designs generated for variance levels of twenty percent and above have efficiencies of more than 97 % regardless of the true variance level whereas the designs for the lowest variance levels risk up to twenty percent loss of efficiency in the most extreme cases. In Strömberg et al. (2016) it was hypothesised that  $D$ -optimal designs with more support points and less clustering would be more robust to parameter misspecification and this finding agrees with that hypothesis. The efficiency profiles for designs of about thirty percent and above all have values higher than one for the designs immediately before the discontinuous bifurcation and replication shift in the design graph. This means the optimisation failed to distinguish between the replication structures. However, the loss in efficiency appears small.

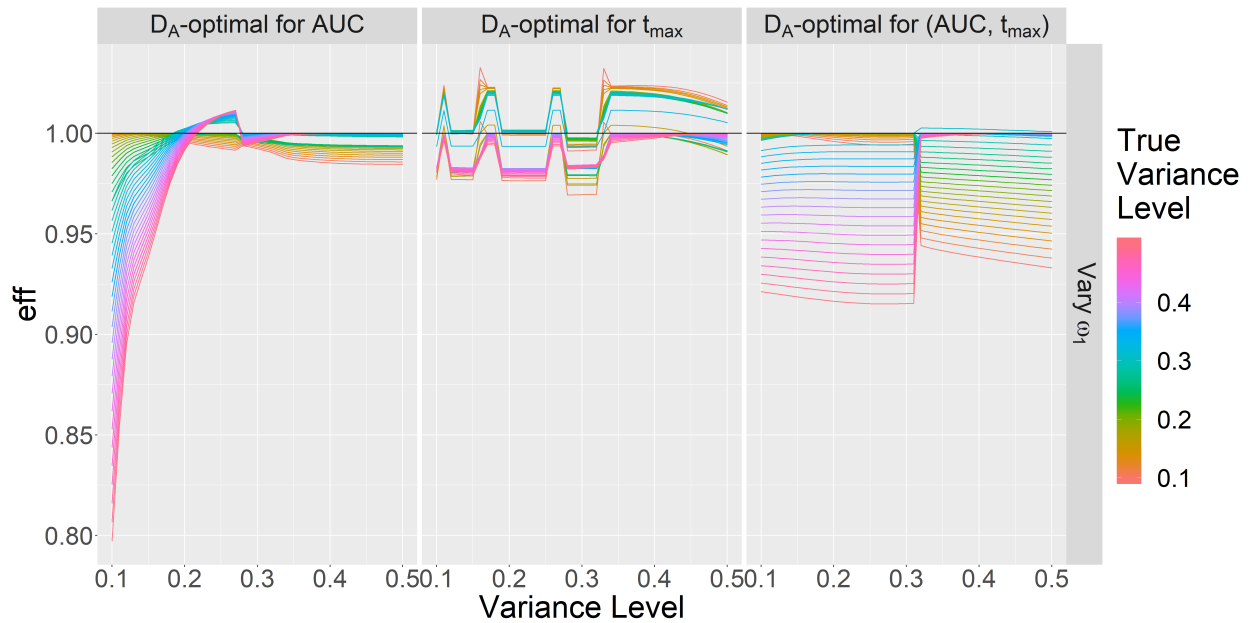
The efficiency profiles for  $t_{max}$  reflect the vacillation of the design graph between the three different replication structures. They also clearly reveal that the replication structure  $r = (2, 2, 1)$  is in fact best for all variance levels. Thus, it is tempting to suggest a follow-up optimisation constrained to this replication structure, but it is worth noting that the efficiency gain seems likely to be small given the relatively consistent sampling times in the design graph.

The efficiency profiles for  $(AUC, t_{max})$  display a new behaviour whereby the variance level can be clearly partitioned into two regions (above or below approximately thirty percent) and the crucial step is to specify which of these regions of variance one believes to be correct. If a design for a variance level below thirty percent is chosen and the true variance level is in fact below thirty percent, it makes a negligible difference which of the relevant designs in the design graph is chosen. However, if the true variance level is above thirty percent the cost in efficiency can be more than five percent. This behaviour is mirrored if a design for a variance level above thirty percent is chosen, and carries an efficiency cost of misspecifying the region of up to ten percent.



**Figure 2.7:** Design graphs for  $D_{\mathbf{A}}$ -optimal designs for the One-Compartment model. The left-hand panel contains  $D_{\mathbf{A}}$ -optimal designs for the AUC, the middle panel contains  $D_{\mathbf{A}}$ -optimal designs for  $t_{max}$  and the right-hand panel contains  $D_{\mathbf{A}}$ -optimal designs for the vector  $(AUC, t_{max})$ . The x-axis gives the variance level in percent for  $\omega_1$ , the variance parameter which is inflated. The y-axis gives the design points in the corresponding  $D_{\mathbf{A}}$ -optimal design, with the colour of the individual point indicating how many times the design point is replicated.

As before, the number of distinct sampling times increases with the magnitude of  $\omega_1$ .



**Figure 2.8:** Efficiency profiles for the designs in Figure 2.7. The x-axis is the variance level used to generate the design under consideration; this design can be identified by looking up the design for the same variance level in Figure 2.7. The y-axis is the  $D_A$ -efficiency of the design under consideration, when the true variance level is as indicated by the colour of the line. Each line is a plot of the  $D_A$ -efficiency of all the designs in the corresponding design graph, under the assumption that a particular variance level (indicated by the colour of the line) for a particular variance component (indicated by the side panel) is true. Variance components that are not varied are kept at a variance level of ten percent.

## 2.5 Other Models

The One-Compartment model is popular as a motivating example, but the workflow suggested in Section 2.1 must be seen to be applicable to other models.

In the preceding sections it is demonstrated that  $D$ -,  $D_s$ - and  $D_A$ -optimal designs can be found for the One-Compartment model and that the relevant design graphs can be very different in terms of sampling times and replication structure. Moreover, efficiency profiles are used both to check the success of the optimisation and to assess the robustness of designs to misspecification of the variance components. In this section, these methods are applied to the models used in Chapter 1 as well as models drawn from Kitsos and Kolovos (2013), a review article which summarises  $D$ -optimal designs for a wide range of nonlinear regression models. Furthermore, to illustrate the application of these methods when real data is involved, nonlinear regression models have been fitted to data sets from Appendix A of Bates and Watts (1988), with the models suggested in the appendix itself. Note that all of these examples are based on nonlinear regression without random effects; the variance components are all based on a choice of variance level (ten percent), with no prior estimates or guesses available.

The specifications of the full set of models investigated in this chapter can be found in Table 2.1. Here follows short descriptions of their respective origins (except for those already seen in Chapter 1).

*Kitsos (2013), example (25)*: This model describes the adsorption of gas molecules on a solid surface via so-called Brunauer-Emmett-Teller theory, which provides a mechanistic description of multilayer adsorption.

*Kitsos (2013), example (28)*: This model describes a reaction network of three irreversible first-order parallel reactions; it is effectively a compartmental model, similar to the One-Compartment model, but the compartmental scheme is different, leading to a different solution to the corresponding set of differential equations.

*Bates and Watts (1988), A1.3*: This model utilises a simple form of the classic Michaelis-Menten equation to describe the velocity of an enzymatic reaction, in this instance represented by counts per minute of a radioactive

product as a function of the substrate concentration. Responses were measured with and without the addition of a treatment, Purmoycin, to the enzyme. The fitted values are obtained from a fit which only considers the treated group.

*Bates and Watts (1988), A1.4:* This model is used to describe the decline over time in biochemical oxygen demand in stream water. The regression function is a variant of exponential decay.

*Bates and Watts (1988), A1.12:* This model was used to describe nitrite utilisation in bush beans as a function of light intensity. The regression function belongs to the class of rational polynomial models. According to page 85 of Piegorsch and Bailer (2005) this model can be thought of as an extension of the Michaelis-Menten model, where the addition of the quadratic term in the denominator " ...allows the mean reaction velocity to rise to a true maximum and then drop".

*Bates and Watts (1988), A1.14:* This model was used to describe the metabolism of a compound called tetracycline (see page 281 of Bates and Watts (1988)). The regression function solves a set of first order differential equations which provides a mechanistic description of metabolisation; more specifically, it is a 2-compartment model with an extra parameter ( $\beta_4$ ) to allow for so-called "dead time".

### 2.5.1 Results

Figures 2.9 to 2.26 contain design graphs and efficiency profiles for the models in Table 2.1, for all possible  $D_s$ -optimal designs. One exception is the model from Appendix A1.14 of Bates and Watts (1988) where only two of the 60 possible cases have been presented. Many of the design graphs are essentially constant, meaning the magnitude of the variance component does not change which design is  $D_s$ -optimal. In most of these cases, the corresponding efficiency profiles indicate that the range of possible efficiencies is extremely narrow around 100%. One exception to this is the Sigmoid Emax model for  $\gamma = 1$ , for  $s = (\beta_2)$  and varying  $\omega_2$ ; the efficiency profiles in Figure 2.20 indicate that the range of possible efficiencies varies from a width of about ten percent to a width of about two percent, depending on the design chosen.

The two most common ways for the design graphs to diverge from being constant is by a change in replication structure which approximately retains the sampling times or by bifurcation. In some cases, the change in replication structure does not look stable; for instance in several of the design graphs for the Sigmoid Emax ( $\gamma = 5$ ) model in Figure 2.23, there seems to be a sort of wavering between replication structures. This could be due to numerical properties of the problem, for instance a very flat objective function. Identifying the replication structure which is best is quite easy to do from the efficiency profiles in Figure 2.24; for instance, when  $s = (\beta_2)$  and  $\omega_1$  is inflated, the replication structure  $r = (2, 2, 1)$  is superior - all profiles are either at or above 100% efficiency for the designs with this replication structure. The replication structure  $r = (2, 1, 2)$  also has efficiency profiles which contain values above 100% but those profiles are even higher for designs with  $r = (2, 2, 1)$  (which can also be seen by the fact that the efficiency profiles for the corresponding variance levels are below 100% for the designs with  $r = (2, 1, 2)$ ). A similar conclusion can be drawn when  $s = (\beta_2)$  and  $\omega_2$  is inflated or when  $s = (\beta_1, \beta_2)$  and  $\omega_1$  is inflated; again, visual inspection of the efficiency profiles reveal that the replication structure  $r = (2, 2, 1)$  is superior.

By far the most serious cases of efficiency loss, in the magnitudes of tens of percent, happen in the cases where the design graph suddenly adds a sampling time; the designs with fewer design points rapidly lose efficiency as the variance level is increased. This is seen for the models from Appendix A1.3 and A1.4 of Bates and Watts (1988), the model from example (25) of Kitsos and Kolovos (2013) and for the Sigmoid Emax model with  $\gamma = 3$  as well as for  $\gamma = 5$ .

There are at least three cases where a sampling time is clearly dropped for a sub-interval of the investigated variance levels, as the variance level is increased. These are

- The model from Appendix A1.3 of Bates and Watts (1988), when  $s = (\beta_2)$  and  $\omega_2$  is varied.
- The model from Appendix A1.4 of Bates and Watts (1988) when  $s = (\beta_1, \beta_2)$  and  $\omega_2$  is varied. The case when  $s = (\beta_2)$  and  $\omega_2$  is varied could arguably also be included.

- The model from example (25) of Kitsos and Kolovos (2013), when  $s = (\beta_2)$  and  $\omega_2$  is varied.

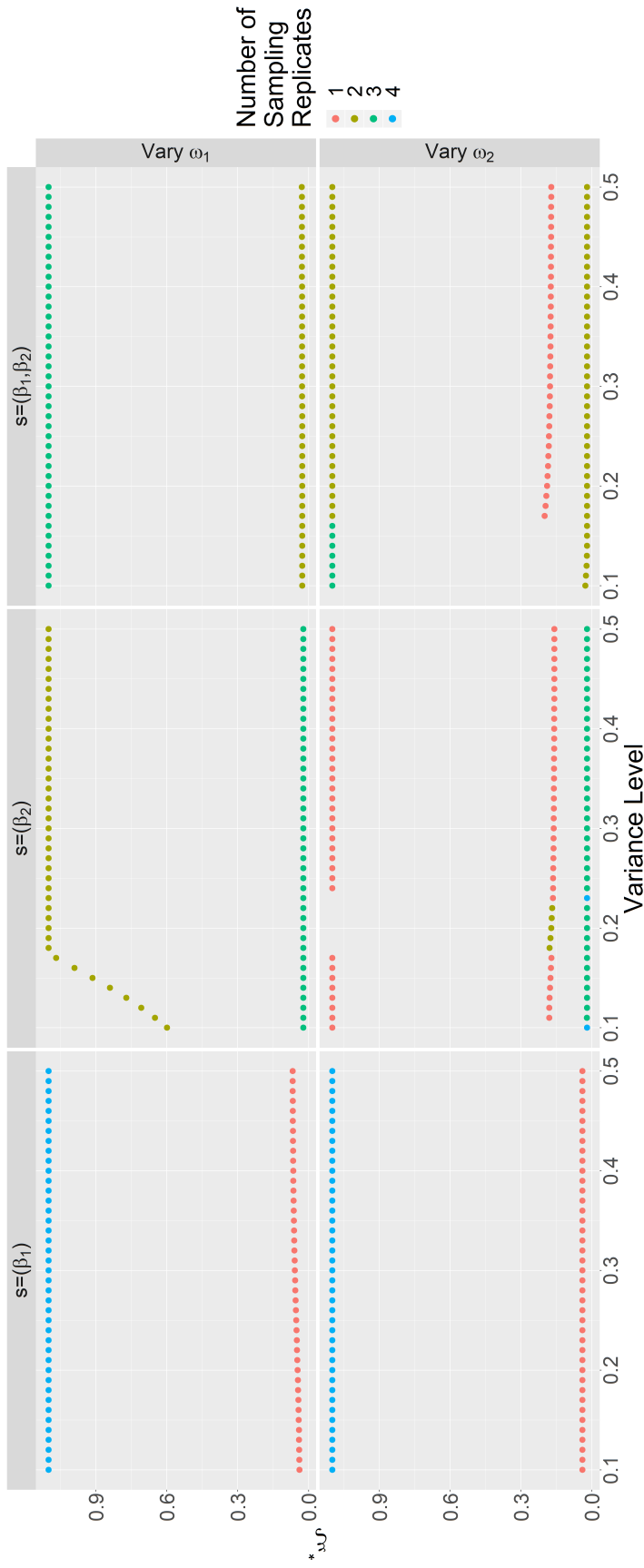
For all of these cases, the efficiency profiles indicate that the designs with more sampling times are (much) more robust than the optimal design. As noted, this corresponds well with the work of Strömberg et al. (2016). Also for all of these cases, there are efficiency profiles with values above 100 %, indicating a failed optimisation.

The last model examined is the one from Appendix A1.14 of Bates and Watts (1988); two of the more interesting design graphs are presented in Figure 2.27. As opposed to the other designs which were all for five-point designs, the design size is here increased to eight. The resulting design graphs combine several of the behaviours seen for other models: discontinuous shifts, adding and dropping sampling times and a subset of sampling times that is close to being approximately constant (visually) for all variance levels. However, the efficiency profiles in Figure 2.28 indicate that in the case where  $s = (\beta_4)$  and  $\omega_4$  is varied, the maximum possible efficiency loss from using a  $D_s$ -optimal design corresponding to the wrong variance level is about 1%. For the other case, where  $s = (\beta_2, \beta_3)$  and  $\omega_3$  is varied, the efficiency profiles form a pattern that is surprisingly similar to those one can see in the case where the design graph is nearly constant, and the design graph is clearly not constant. However, relative to the cases where the design graph is constant, the range of possible efficiencies is fairly wide; moreover, the profiles clearly suggest that the most robust design is for a variance level of about 18%.

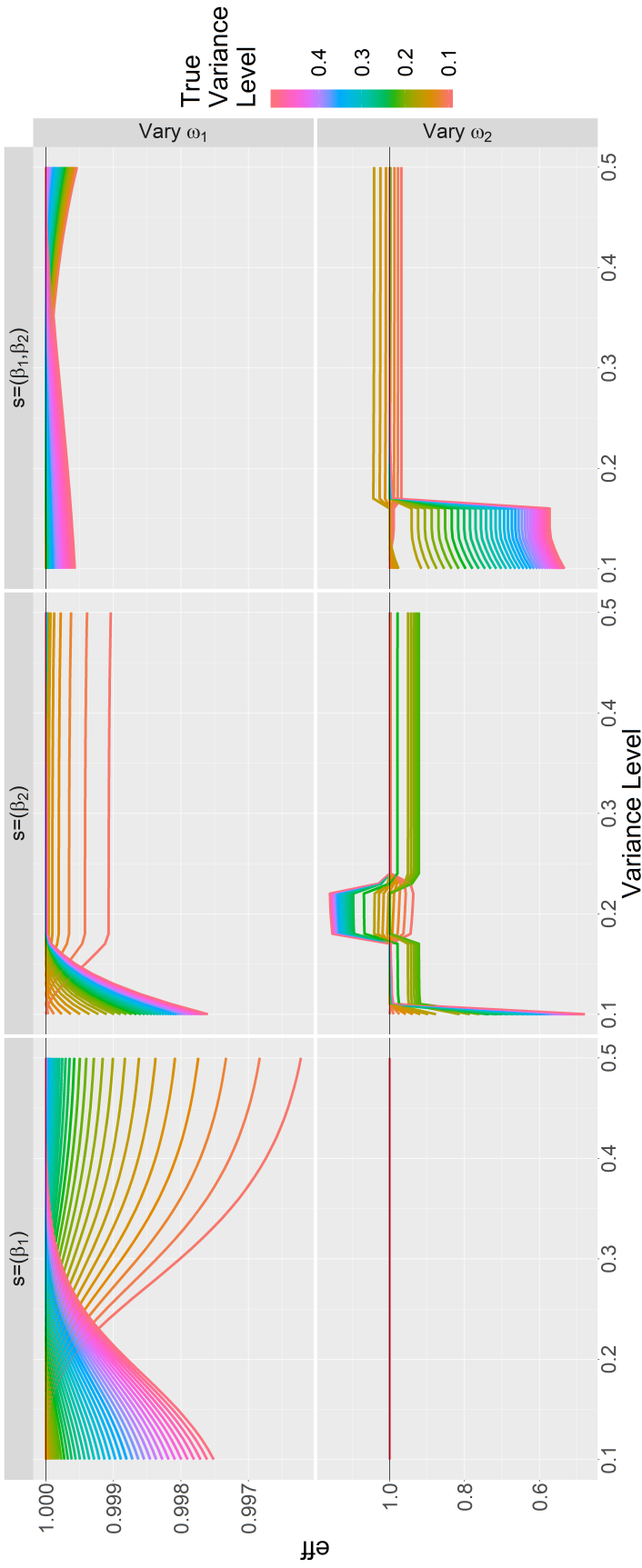


model	regression function ( $\eta$ )	parameter values
One-Compartment	$\mu_2 \left[ \mu_3(\mu_2 - \mu_1) \right]^{-1} \left[ \exp(-\mu_1 t) - \exp(-\mu_2 t) \right]$	$\mu_j = \exp(\beta_j)$ $\beta = (0, 1, 1)^\top$ $t \in [0, 4]$
Kitsos (2013), ex. (25)	$\beta_1 \beta_2 t \left( [1 - t][1 + (\beta_2 - 1)t] \right)^{-1}$	$\beta = (2, 10)^\top$ $t \in [0.05, 0.3]$
Kitsos (2013), ex. (28)	$\beta_1(\beta_1 + \beta_2 - \beta_3)^{-1} \left( \exp[-\beta_3 t] - \exp[-(\beta_1 + \beta_2)t] \right)$	$\beta = \left( \frac{1}{4}, \frac{1}{2}, \frac{1}{2} \right)^\top$ $t \in [0, 2]$
Sigmoid Emax( $\gamma$ ) for $\gamma = 1, 3, 5$	$5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$	$\beta = (30, 500)^\top$ $t \in [0, 1000]$
Bates and Watts A1.3	$\beta_1 t (\beta_2 + t)^{-1}$	$\beta = (213, 0.0641)^\top$ $t \in [0.02, 1.1]$
Bates and Watts A1.4	$\beta_1 (1 - \exp(\beta_2 t))$	$\beta = (19.1, -0.531)^\top$ $t \in [1, 7]$
Bates and Watts A1.12	$\beta_1 t (\beta_2 + t + \beta_3 t^2)^{-1}$	$\beta = (89800, 187, 0.0163)^\top$ $t \in [2.2, 170]$
Bates and Watts A1.14	$\beta_3 \left( \exp[-\beta_1(t - \beta_4)] - \exp[-\beta_2(t - \beta_4)] \right)$	$\beta = (0.149, 0.716, 2.65, 0.412)^\top$ $t \in [1, 16]$

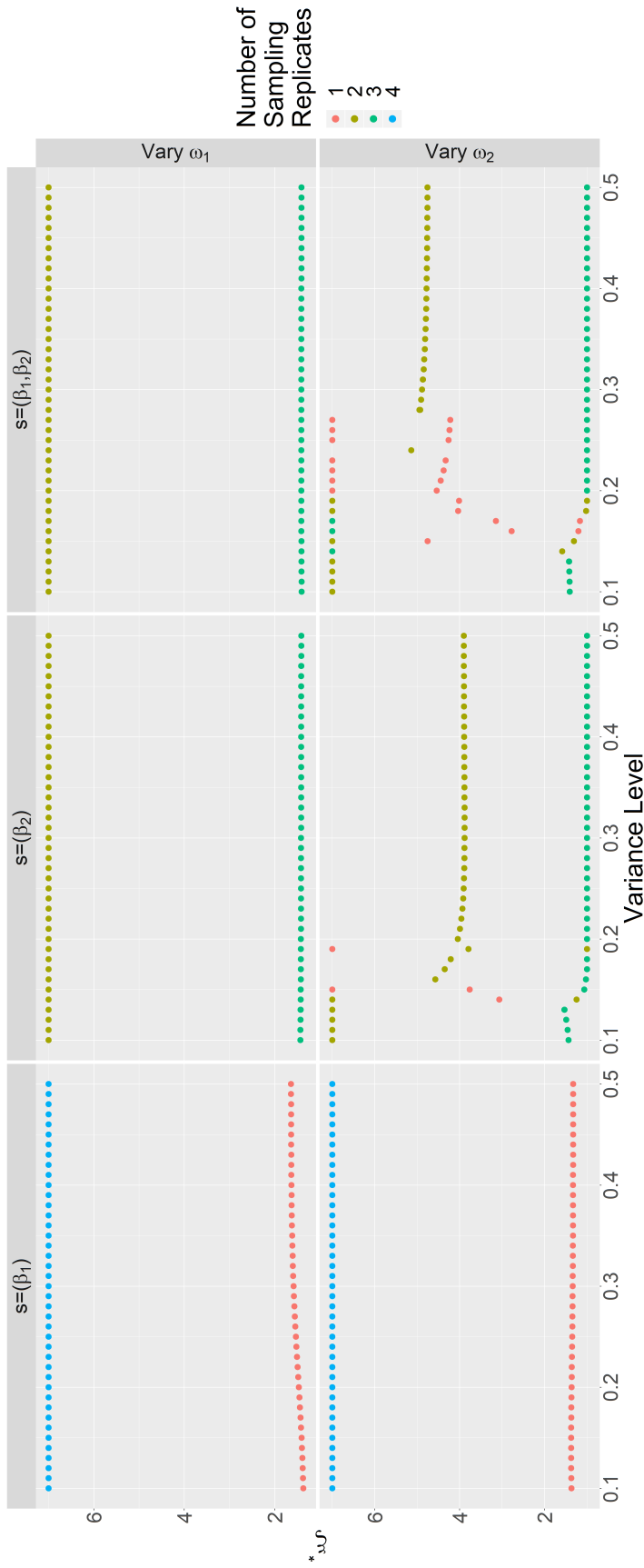
Table 2.1: Nonlinear regression models to be translated to NLME models.



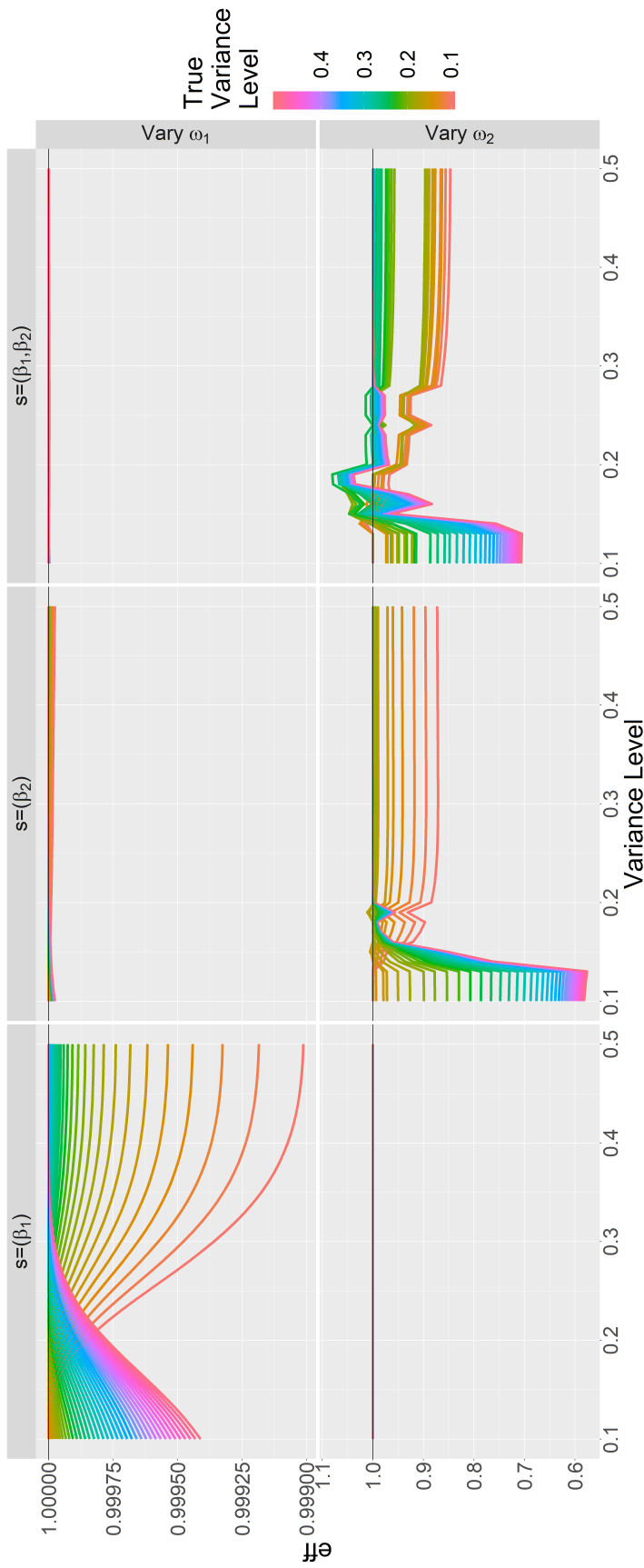
**Figure 2.9:** Design graphs generated from the Michaelis-Menten model taken from Bates and Watts (1988), Appendix A1.3. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (213, 0.0641)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [0.02, 1.1]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.



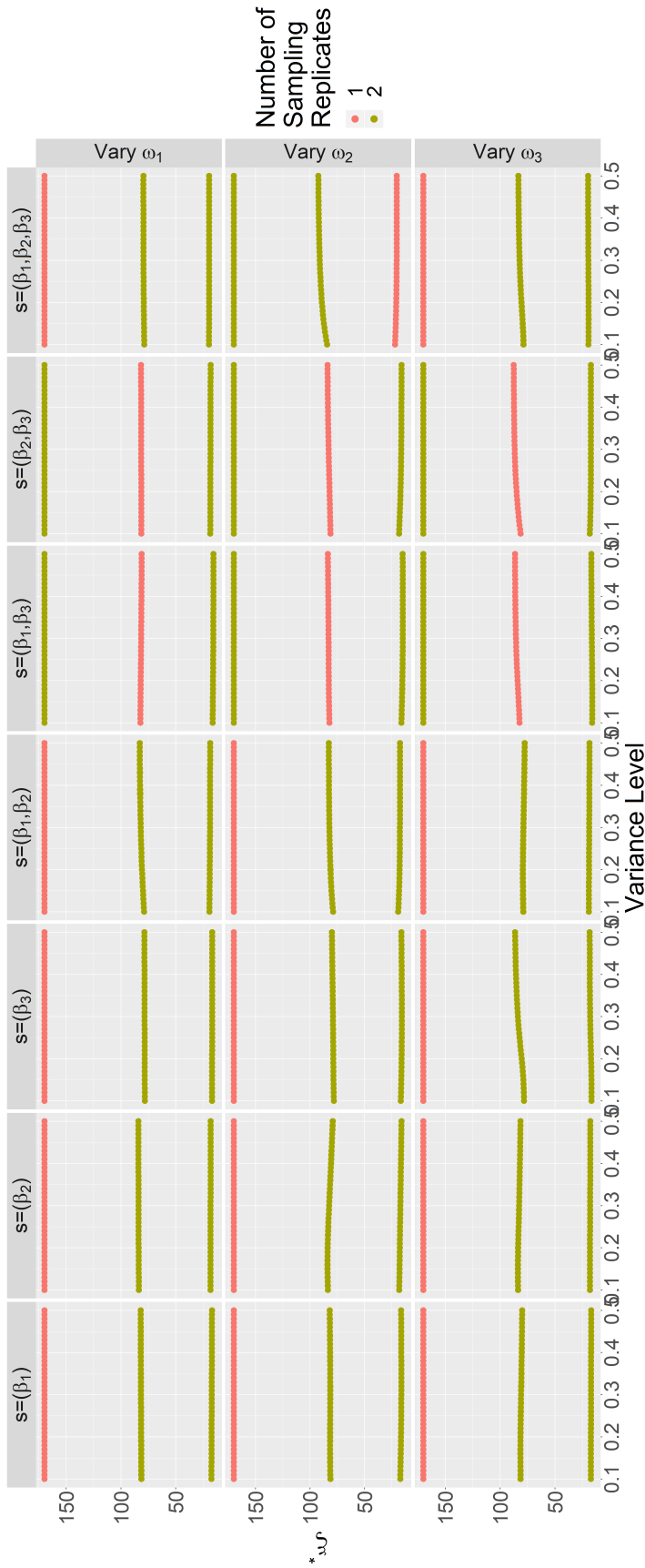
**Figure 2.10:** Efficiency profiles generated from the Michaelis-Menten model taken from Bates and Watts (1988), Appendix A1.3. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (213, 0.0641)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0.02, 1.1]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.9. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.9) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.



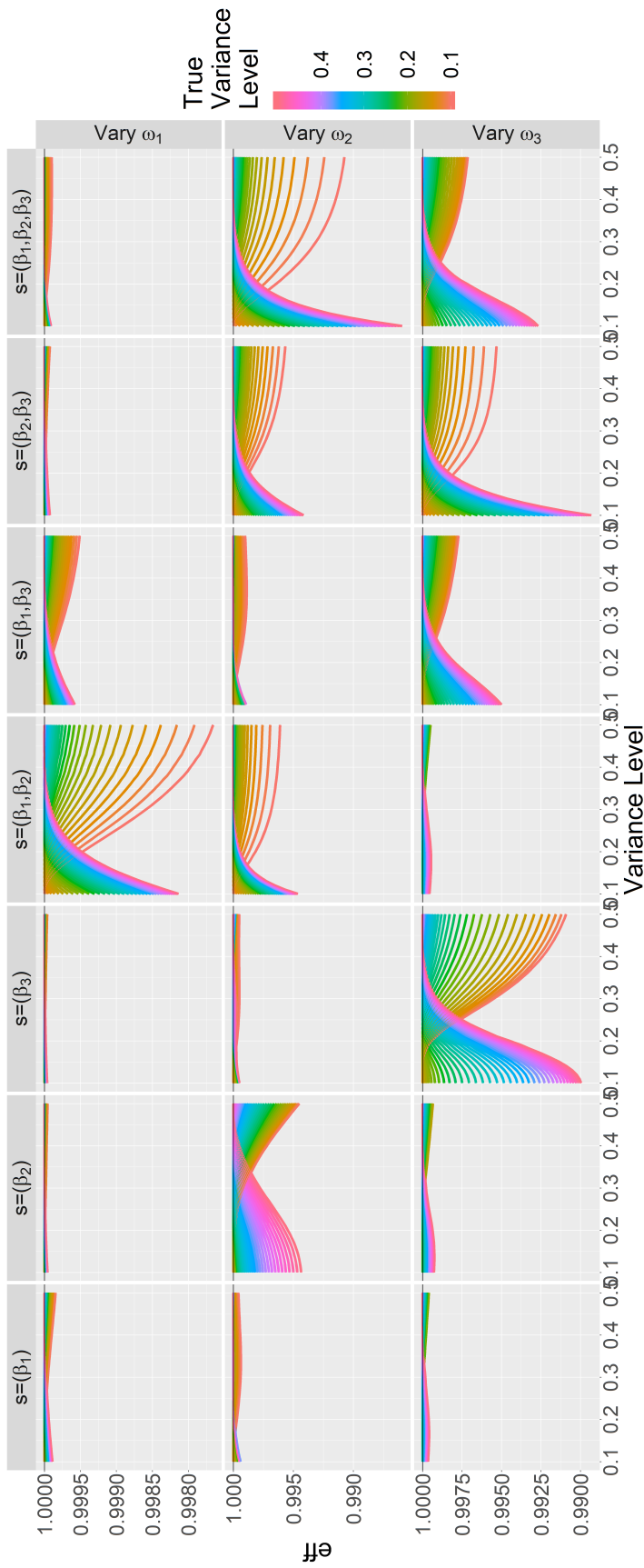
**Figure 2.11:** Design graphs generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.4. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(1 - \exp(\beta_2 t))$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (19.1, -0.531)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [1, 7]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.



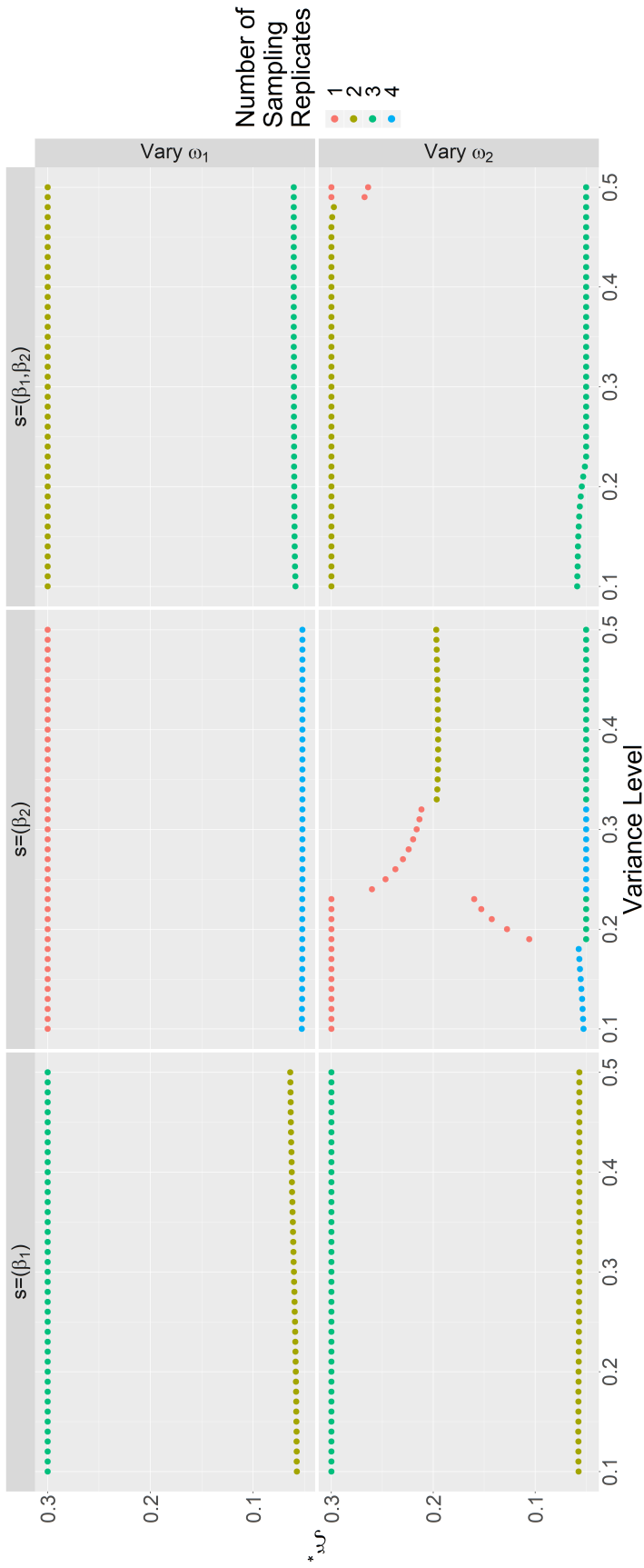
**Figure 2.12:** Efficiency profiles generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.4. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(1 - \exp(\beta_2 t))$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (19.1, -0.531)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [1, 7]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.11. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.11) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.



**Figure 2.13:** Design graphs generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.12. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 t (\beta_2 + t + \beta_3 t^2)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (89800, 187, 0.0163)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [2.2, 170]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$ ,  $\omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.

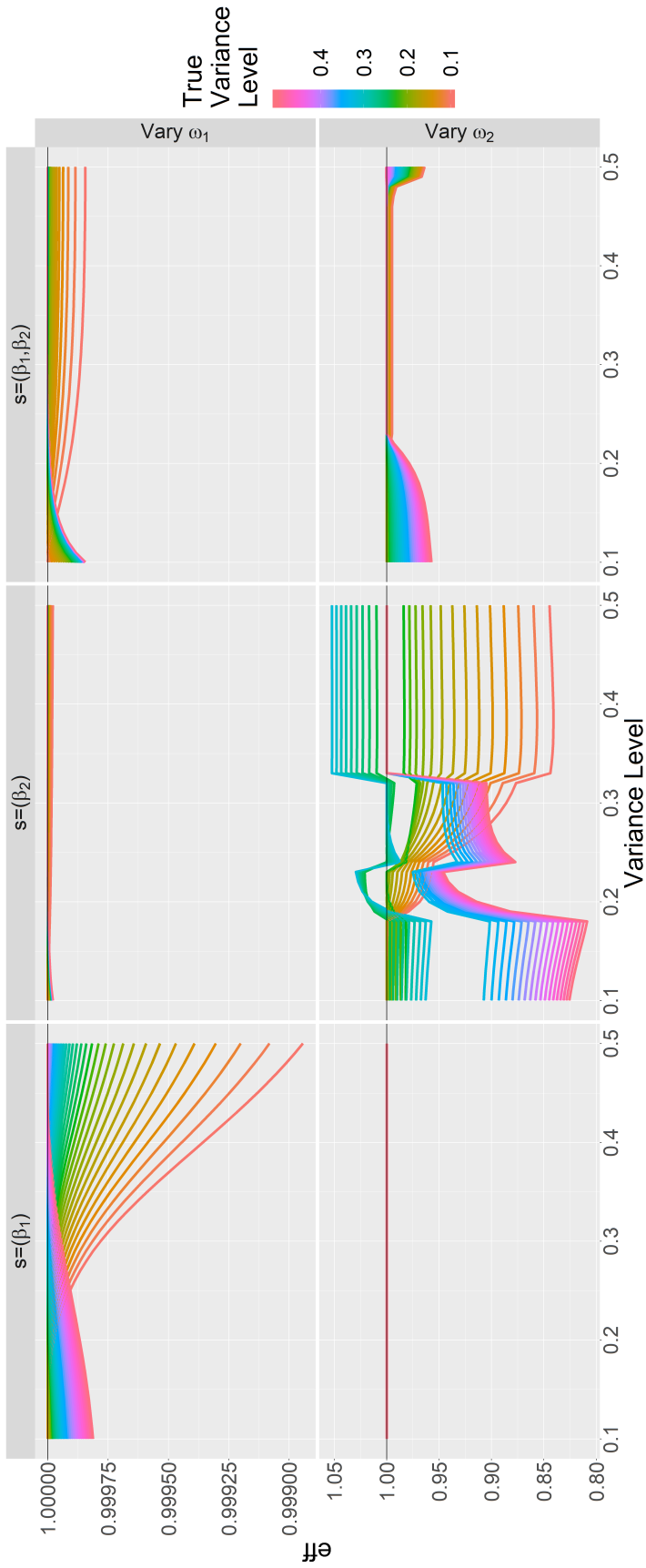


**Figure 2.14:** Efficiency profiles generated from the (variant of a) exponential decay model taken from Bates and Watts (1988), Appendix A1.12. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 t (\beta_2 + t + \beta_3 t^2)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (89800, 187, 0.0163)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [2.2, 170]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$ ,  $\omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.13. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.13) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.

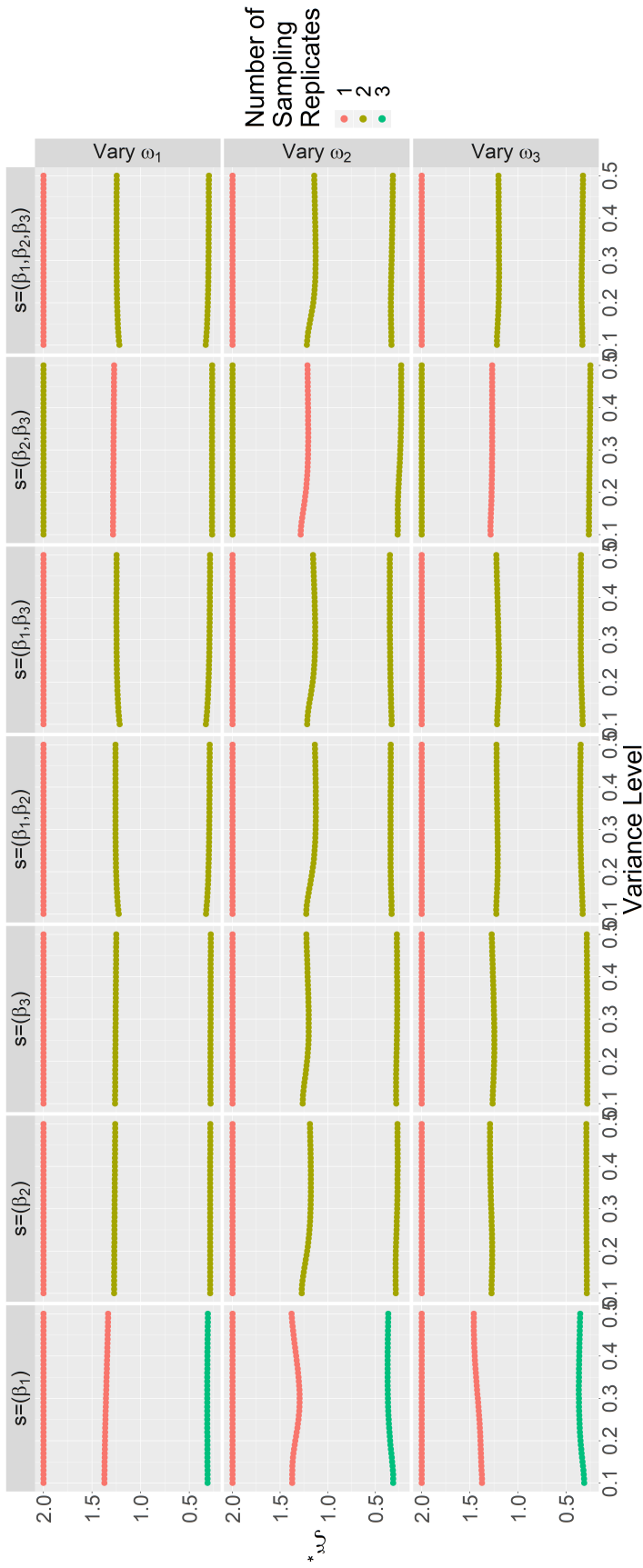


**Figure 2.15:** Design graphs generated from the compartmental gas adsorption model taken from Kitsoos and Kolovos (2013), example (25). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 \beta_2 t \left( [1 - t][1 + (\beta_2 - 1)t] \right)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (2, 10)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [0.05, 0.3]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.

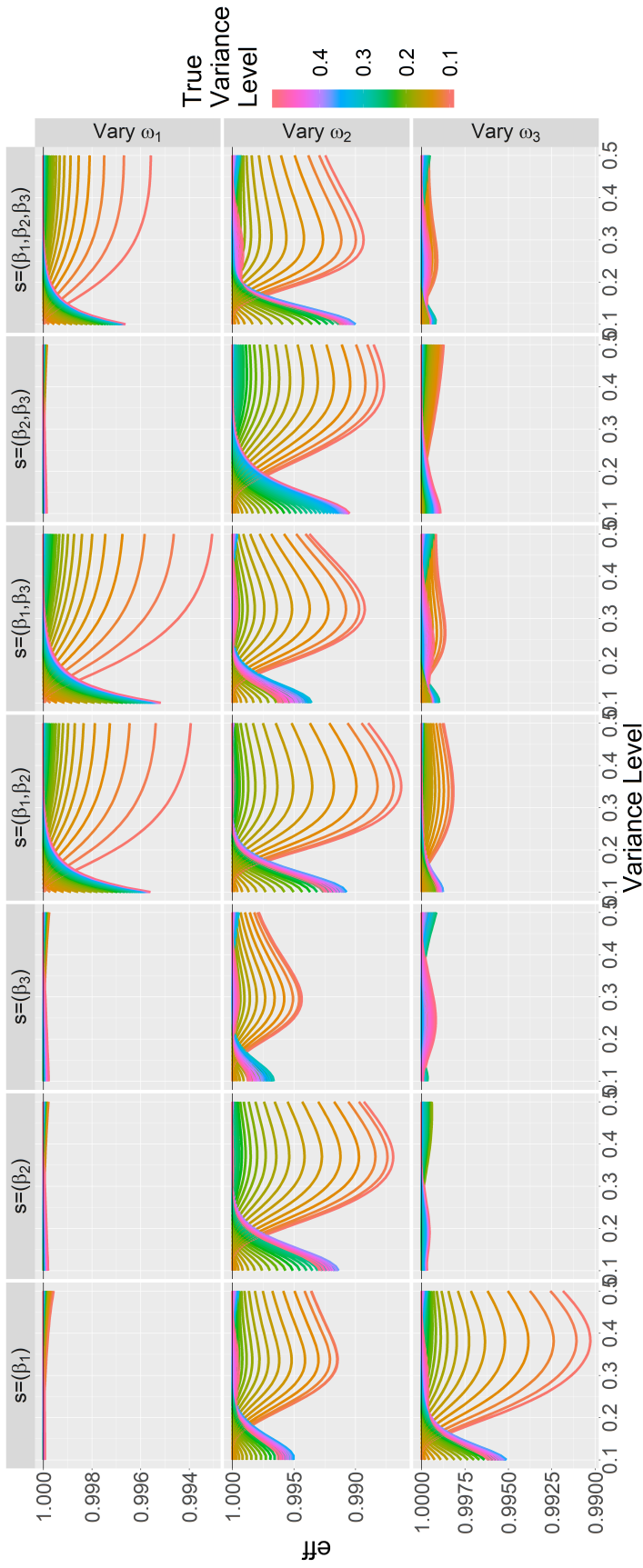




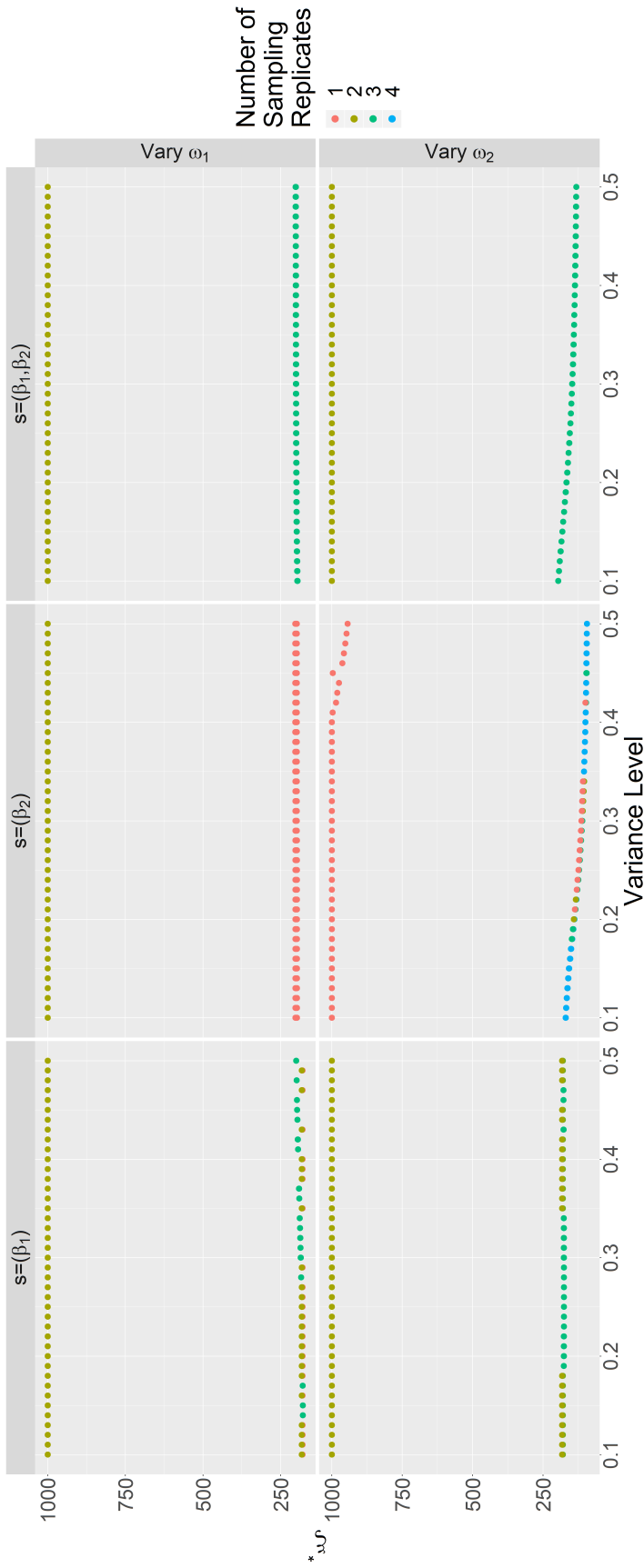
**Figure 2.16:** Efficiency profiles generated from the compartmental gas adsorption model taken from Kitsos and Kolovos (2013), example (25). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1 \beta_2 t \left( [1 - t][1 + (\beta_2 - 1)t] \right)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (2, 10)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [0.05, 0.3]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.15. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.15) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.



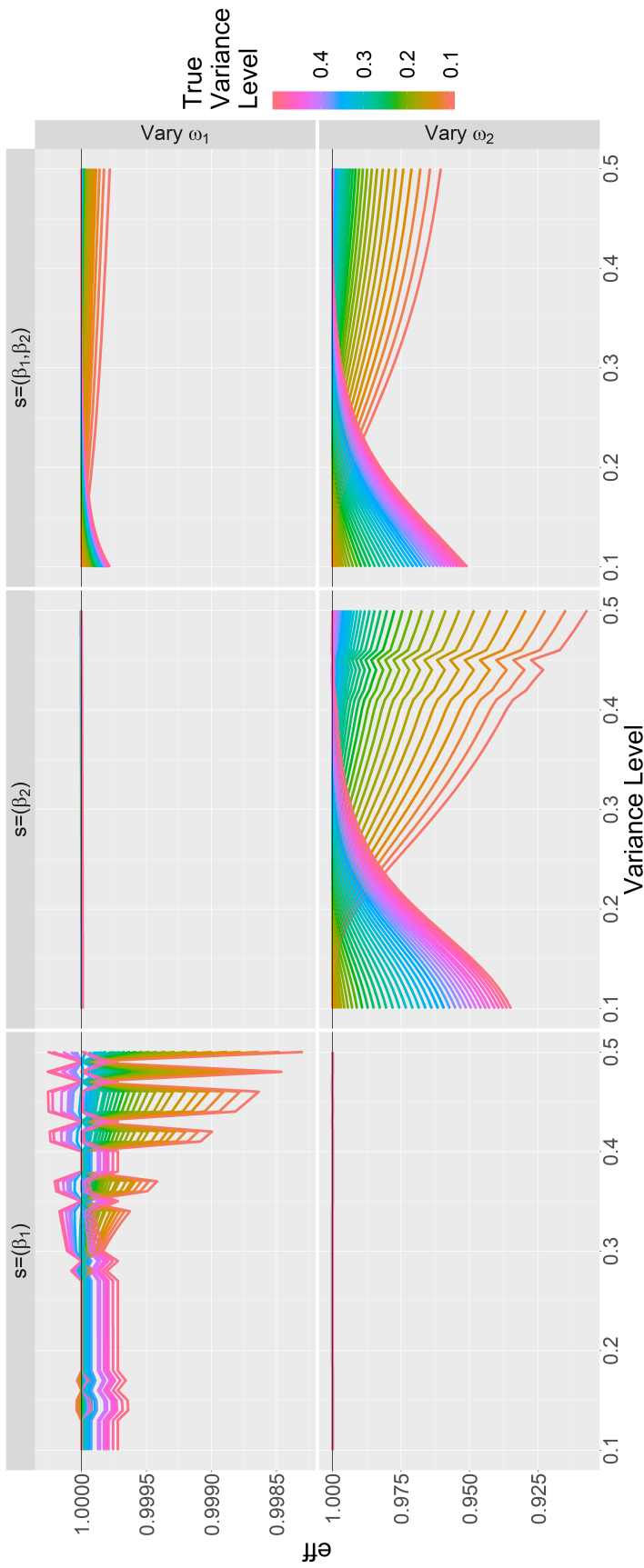
**Figure 2.17:** Design graphs generated from the reaction network model taken from Kitsos and Kolovos (2013), example (28). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(\beta_1 + \beta_2 - \beta_3)^{-1} \left( \exp[-\beta_3 t] - \exp[-(\beta_1 + \beta_2)t] \right)$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = \left( \frac{1}{4}, \frac{1}{2}, \frac{1}{2} \right)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [0, 2]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.



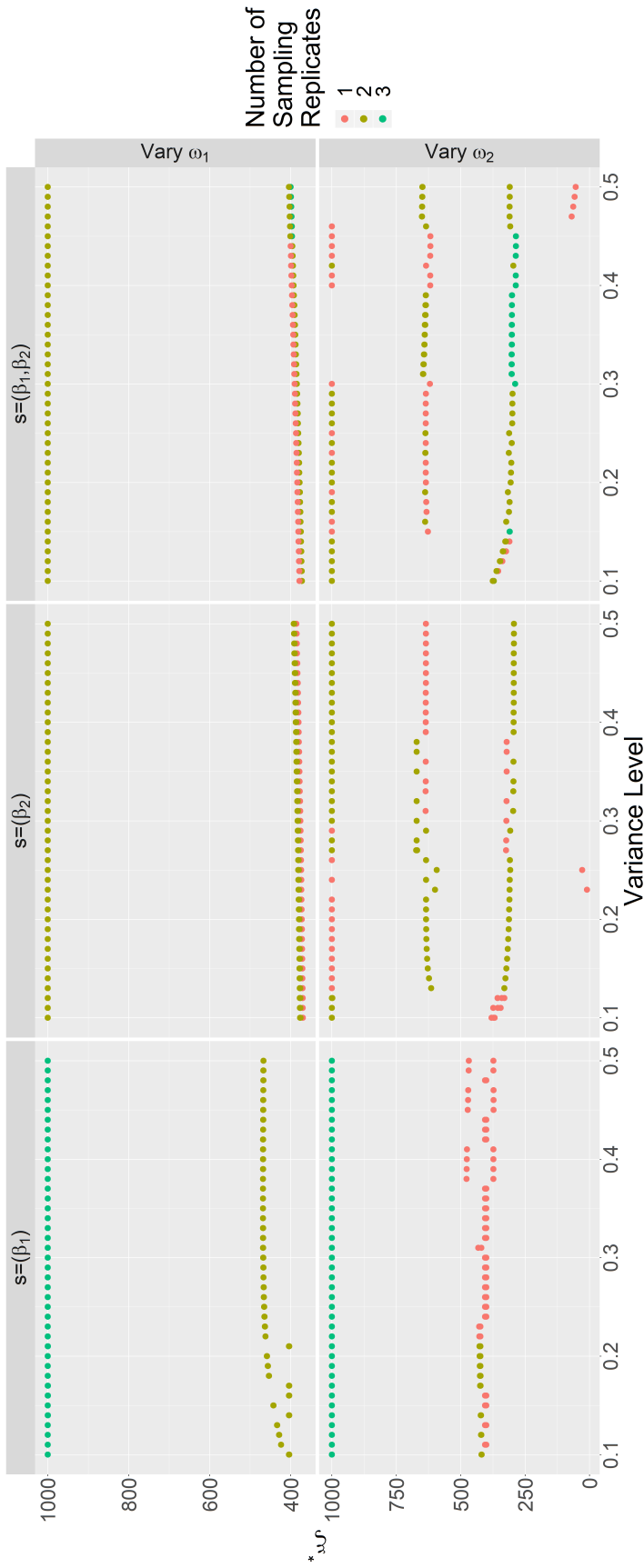
**Figure 2.18:** Efficiency profiles generated from the reaction network model taken from Kitsoos and Kolovos (2013), example (28). The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_1(\beta_1 + \beta_2 - \beta_3)^{-1} \left( \exp[-\beta_3 t] - \exp[-(\beta_1 + \beta_2)t] \right)$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{2})^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [0, 2]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.17. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.17) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.



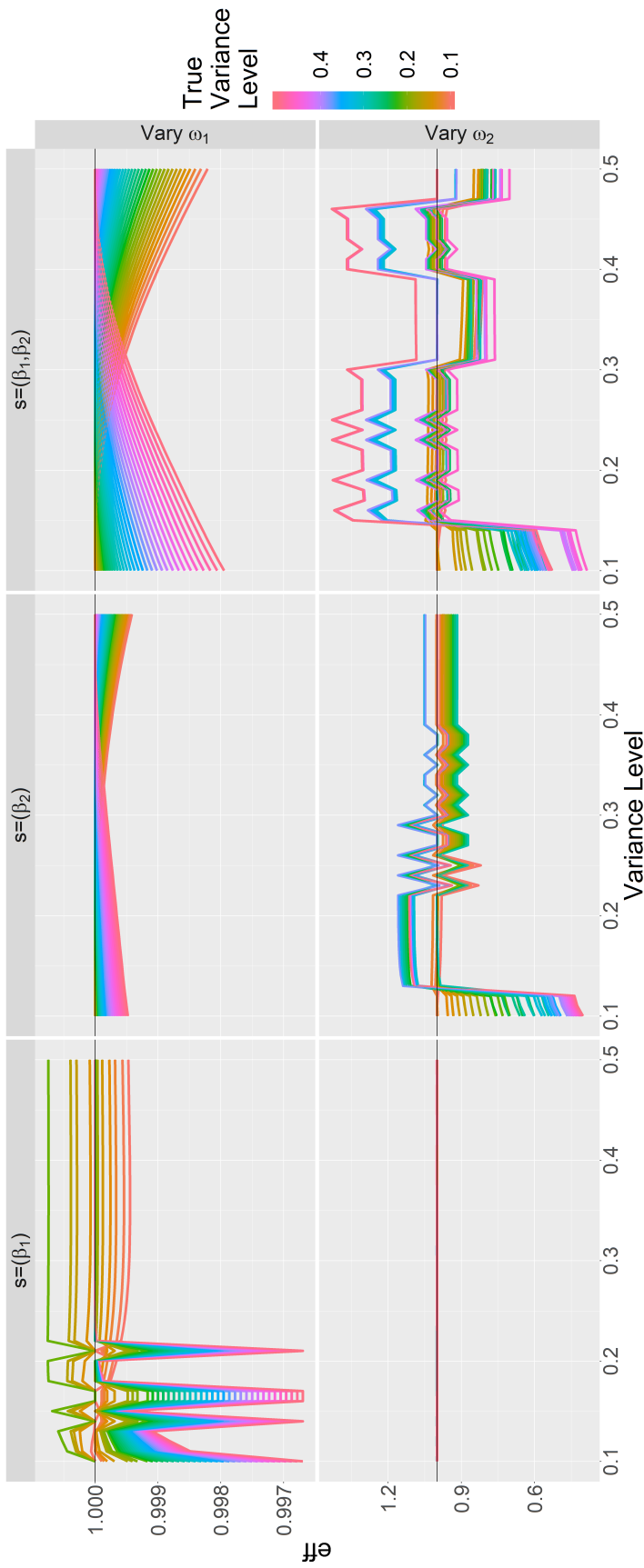
**Figure 2.19:** Design graphs generated from Sigmoid Emax ( $\gamma = 1$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.



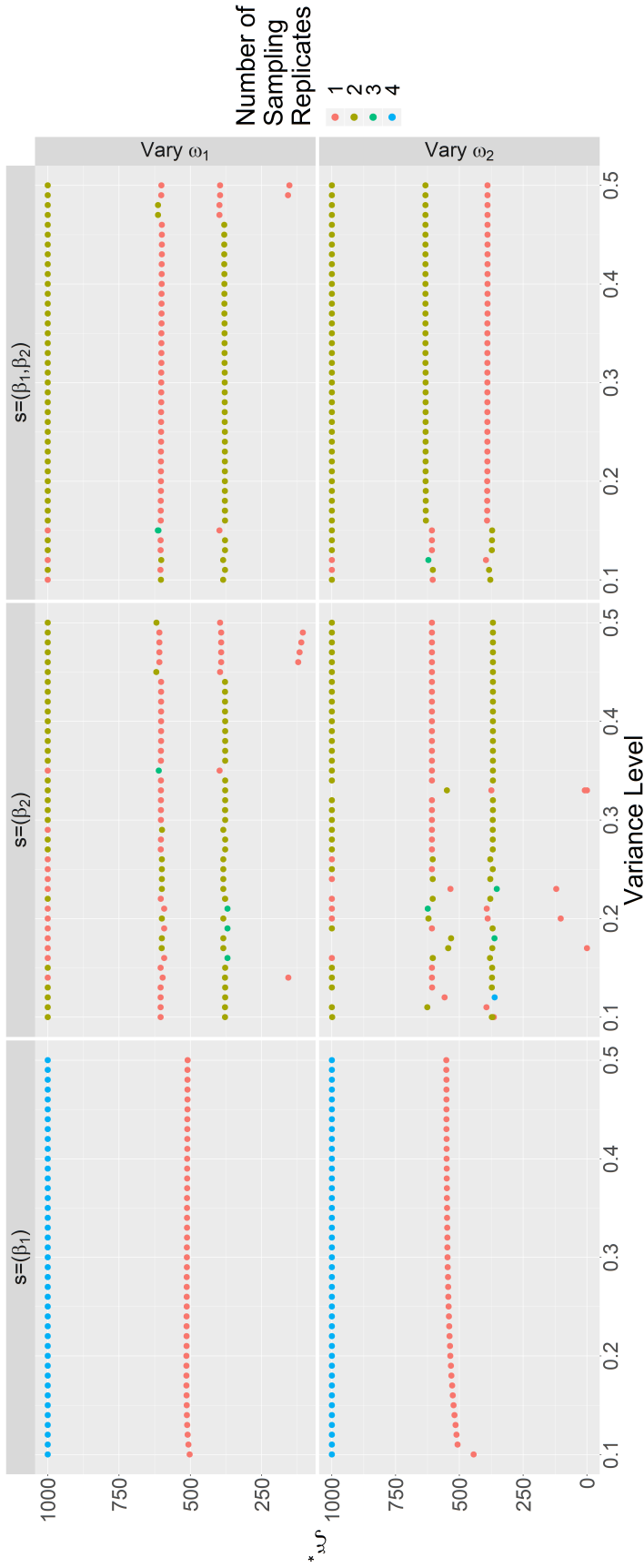
**Figure 2.20:** Efficiency profiles generated from Sigmoid Emax ( $\gamma = 1$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1 t(\beta_2 + t)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\xi} \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.19. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.19) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.



**Figure 2.2.1:** Design graphs generated from Sigmoid Emax ( $\gamma = 3$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.

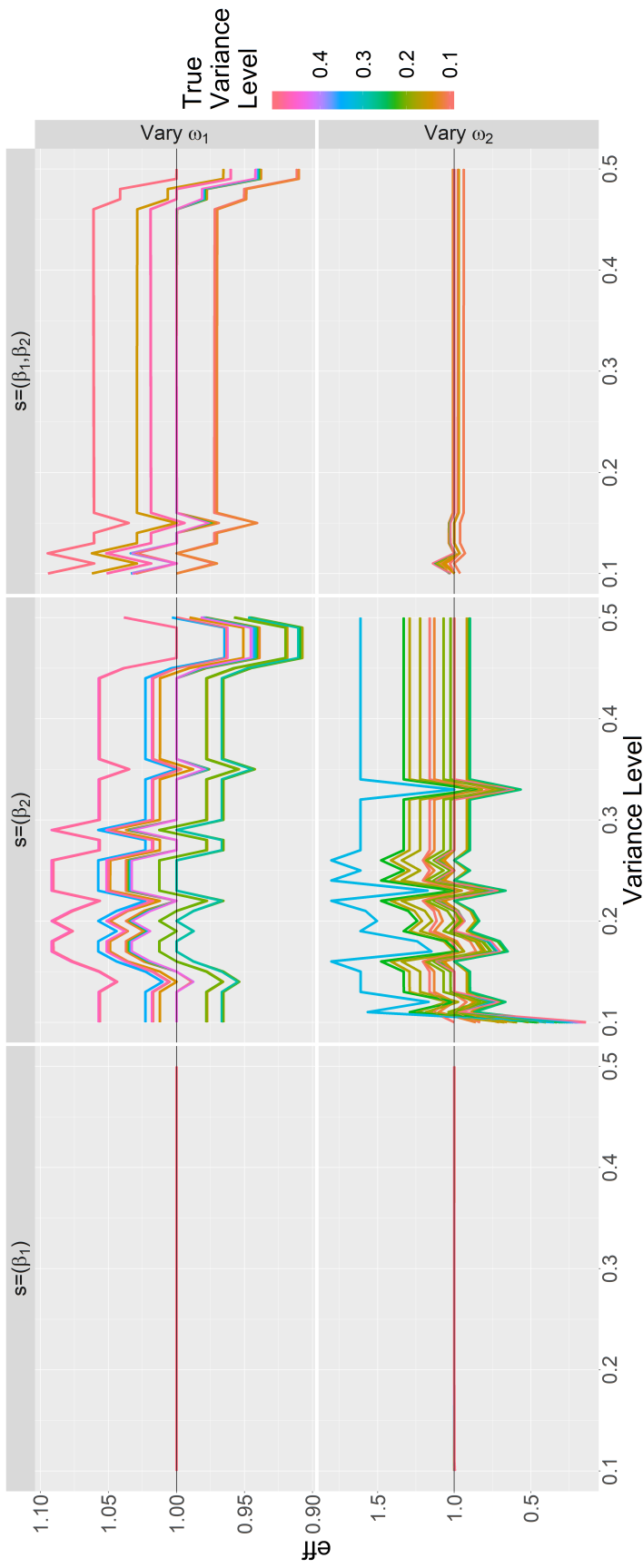


**Figure 2.22:** Efficiency profiles generated from Sigmoid Emax ( $\gamma = 3$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.21. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.21) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.

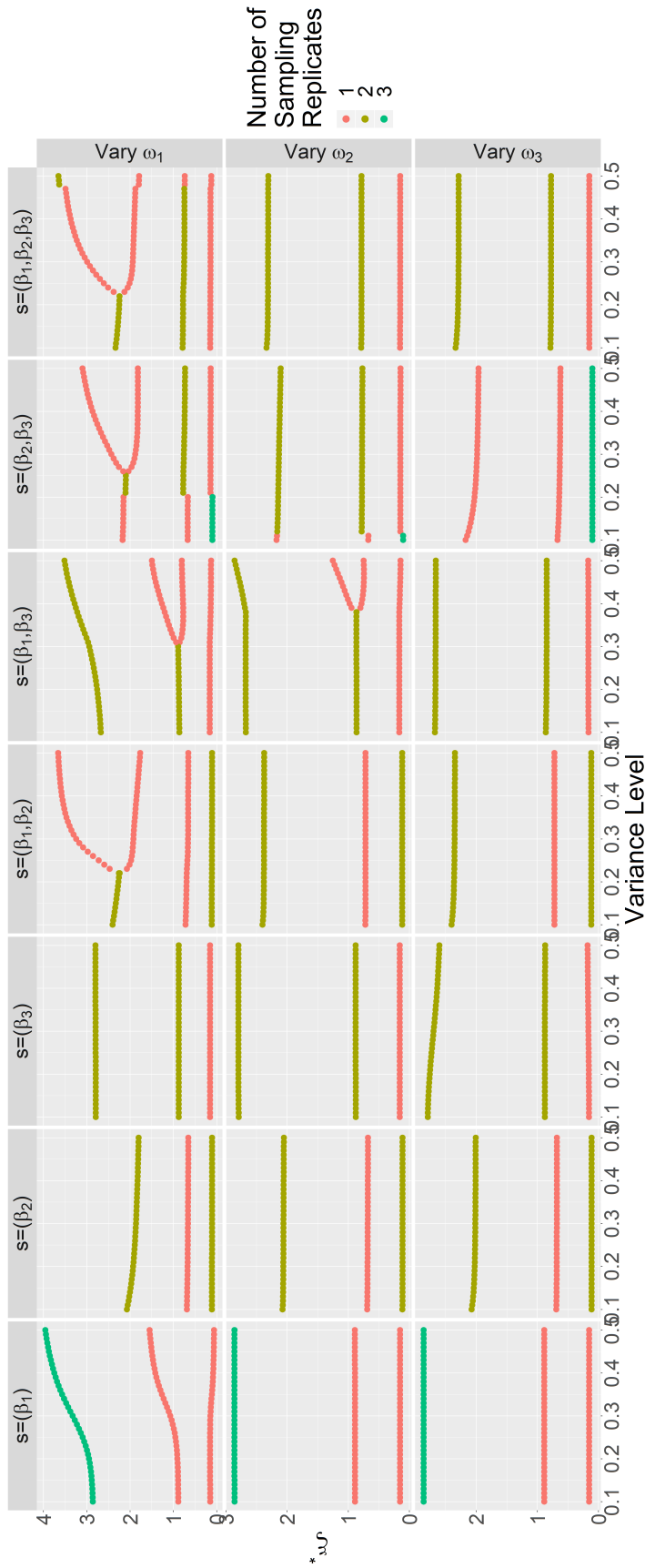


**Figure 2.23:** Design graphs generated from Sigmoid Emax ( $\gamma = 5$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.

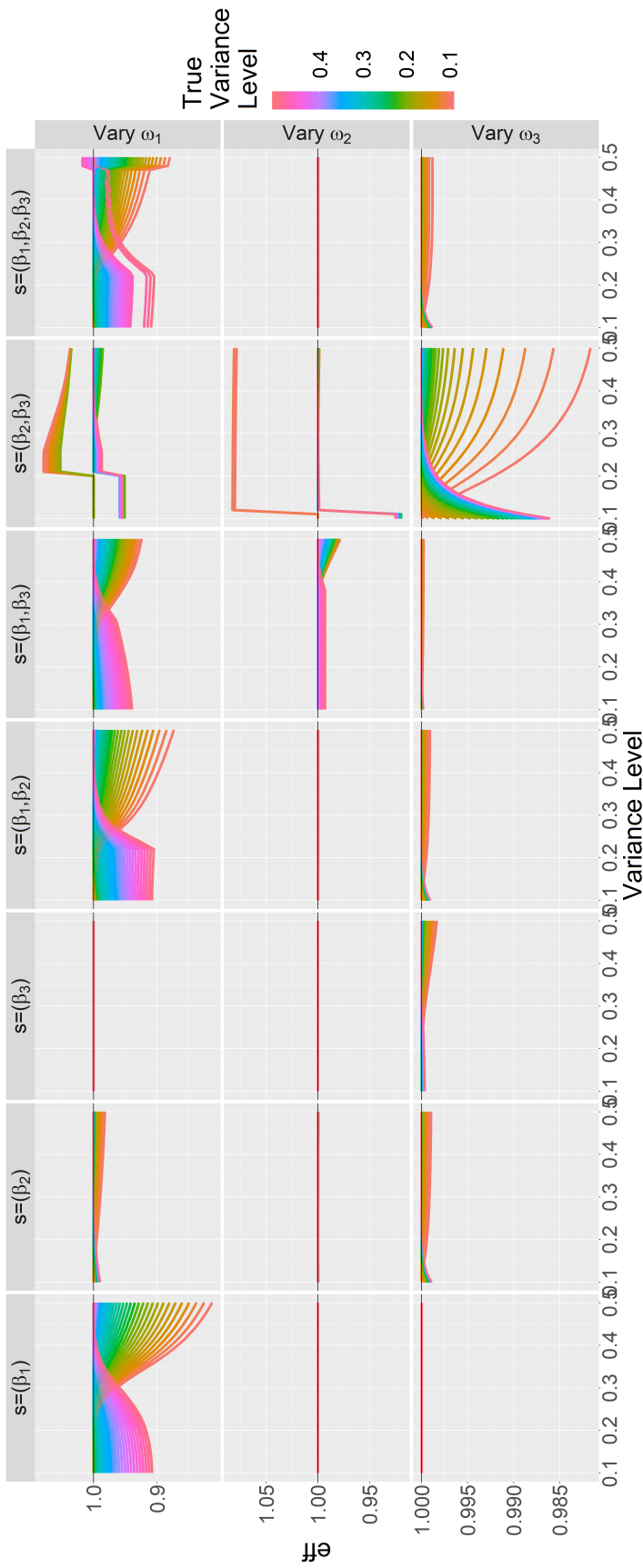




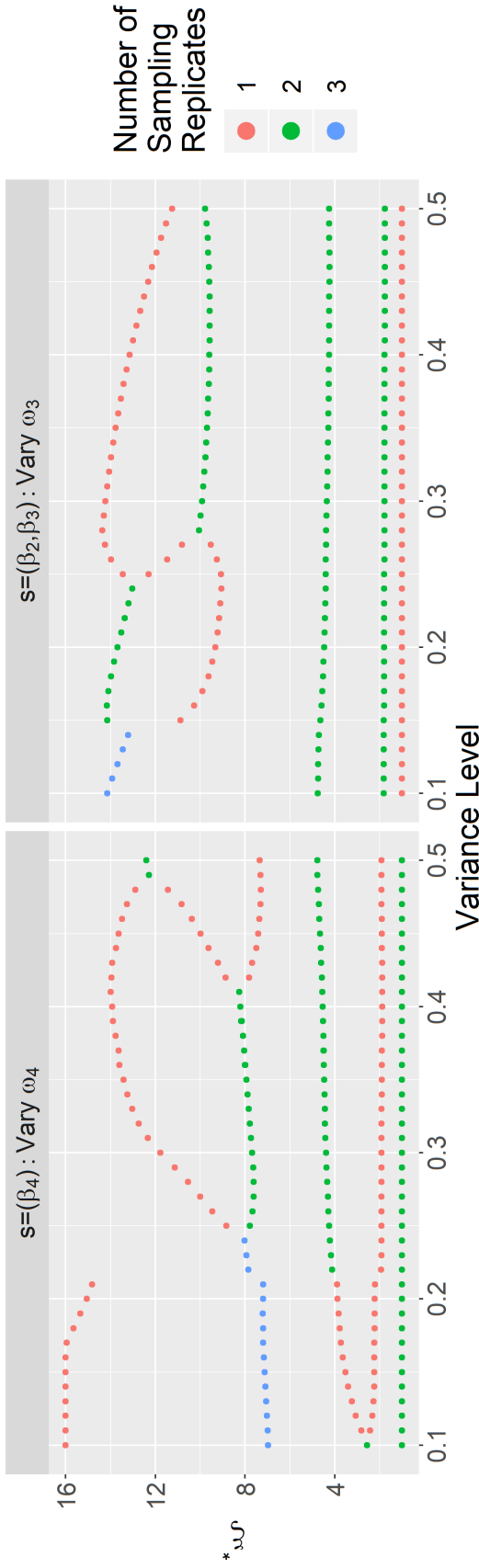
**Figure 2.24:** Efficiency profiles generated from Sigmoid Emax ( $\gamma = 5$ ). The regression function is  $\eta(\boldsymbol{\beta}, t) = 5 + \beta_1^\gamma t^\gamma (\beta_2^\gamma + t^\gamma)^{-1}$  and the fixed effects parameter vector is  $\boldsymbol{\beta} = (30, 500)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 1000]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_2(0, \text{diag}\{\omega_1, \omega_2\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.23. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.23) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.



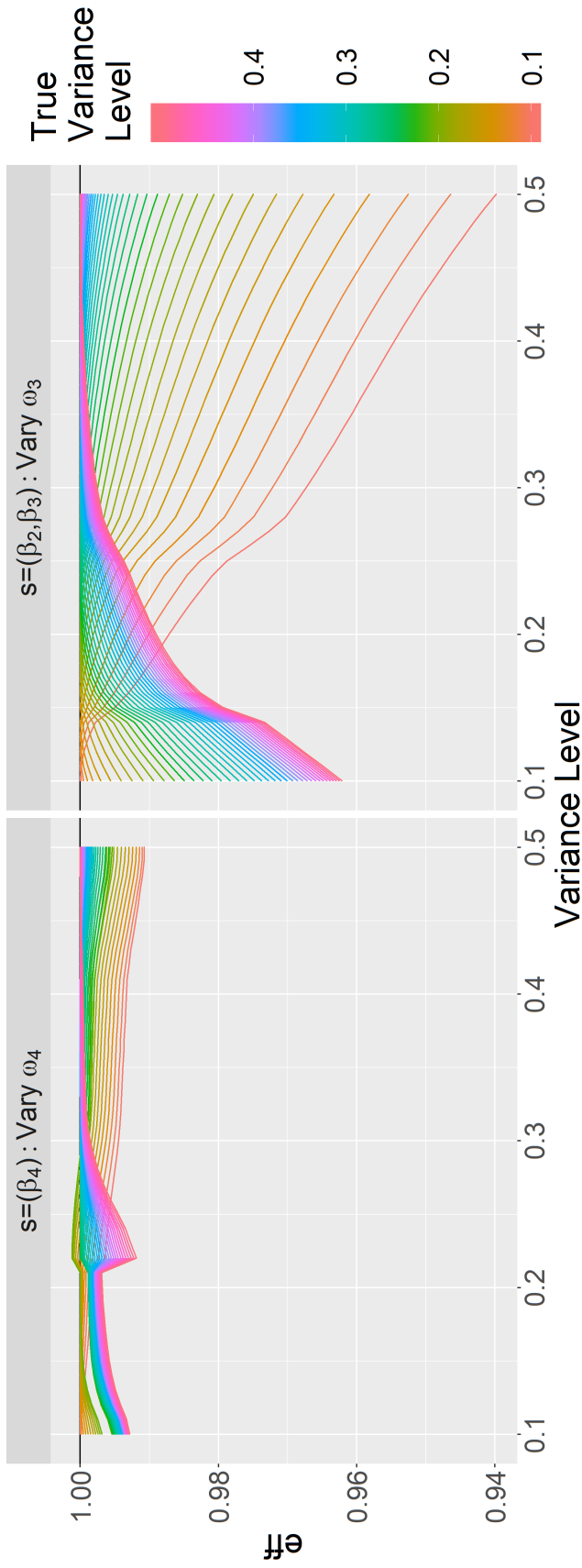
**Figure 2.25:** Design graphs generated from the One-Compartment model found in row one of Table 2.1. The regression function is  $\eta(\boldsymbol{\beta}, t) = \mu_2 \left[ \mu_3(\mu_2 - \mu_1) \right]^{-1} \left[ \exp(-\mu_1 t) - \exp(-\mu_2 t) \right]$  with  $\mu_j = \exp(\beta_j)$  for  $j = 1, 2, 3$ . and the fixed effects parameter vector is  $\boldsymbol{\beta} = (0, 1, 1)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \xi) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\xi \in [0, 4]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2$ ,  $\omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.



**Figure 2.26:** Efficiency profiles generated from the One-Compartment model found in row one of Table 2.1. The regression function is  $\eta(\beta, t) = \mu_2 \left[ \mu_3(\mu_2 - \mu_1) \right]^{-1} \left[ \exp(-\mu_1 t) - \exp(-\mu_2 t) \right]$  with  $\mu_j = \exp(\beta_j)$  for  $j = 1, 2, 3$ . and the fixed effects parameter vector is  $\beta = (0, 1, 1)^T$ . The NLME model which is used to generate this design graph has the form  $y_i = \eta(\beta + \mathbf{b}_i, \xi) + \varepsilon_i$ . Here, it is assumed that  $\xi \in [0, 4]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_3(0, \text{diag}\{\omega_1, \omega_2, \omega_3\})$  and  $\varepsilon_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.25. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.25) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.



**Figure 2.27:** Design graphs generated from the tetracycline metabolism model found in Bates and Watts (1988), Appendix A1.14. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_3 \left( \exp[-\beta_1(t - \beta_4)] - \exp[-\beta_2(t - \beta_4)] \right)$ . and the fixed effects parameter vector is  $\boldsymbol{\beta} = (0.149, 0.716, 2.65, 0.412)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i, \zeta) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\zeta \in [0, 4]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_4(0, \text{diag}\{\omega_1, \omega_2, \omega_3, \omega_4\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3, \omega_4$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. A given value on the x-axis corresponds to the true value of  $\omega_1$ , given by the corresponding variance level. The value on the y-axis of points with the same  $\omega_1$ -value give the sampling times of the corresponding  $D_s$ -optimal design. The number of sampling replicates of a given sampling time is indicated by the colour code used for the point. Thus, a single  $D$ -optimal design is plotted per variance level, captured in the set of points along the y-axis.



**Figure 2.28:** Efficiency profiles generated from the tetracycline metabolism model found in Bates and Watts (1988), Appendix A1.14. The regression function is  $\eta(\boldsymbol{\beta}, t) = \beta_3 \left( \exp[-\beta_1(t - \beta_4)] - \exp[-\beta_2(t - \beta_4)] \right)$ , and the fixed effects parameter vector is  $\boldsymbol{\beta} = (0.149, 0.716, 2.65, 0.412)^\top$ . The NLME model which is used to generate this design graph has the form  $\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta} + \mathbf{b}_i; \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}_i$ . Here, it is assumed that  $\boldsymbol{\zeta} \in [1, 16]^5$  and  $\mathbf{b}_i \sim \mathcal{N}_4(0, \text{diag}\{\omega_1, \omega_2, \omega_3, \omega_4\})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_5(0, \sigma^2 \mathbf{I}_5)$ . The values of  $\omega_2, \omega_3, \omega_4$  and  $\sigma^2$  are fixed at values corresponding to a variance level of ten percent. To interpret this plot, one must concurrently refer to Figure 2.27. Each curve is an efficiency profile; the colour indicates the true value of  $\omega_1$ , the x-axis value indicates the D-optimal design which is being used (this can be located at the corresponding x-axis value in Figure 2.27) and the y-value is the relative FO D-efficiency of the chosen design relative to the FO D-optimal design corresponding to the true variance level.

## 2.6 Discussion

The results described above demonstrate that several different characteristics of practical interest can be uncovered about the optimal design through the suggested methodology: if the design graph is essentially constant (and the corresponding efficiency profiles cluster around 100%), one can rest easy about misspecifying the variance level. For a non-constant design graph, one can use efficiency profiles to check robustness of the designs and possibly pick alternative designs that are more robust or that are acceptably robust and more practically feasible. In cases where the optimisation has failed, in particular when there are two competing replication structures that are hard for the optimisation algorithm to distinguish, the efficiency profiles provide a check of optimality. If the efficiency profiles indicate that slightly different designs have surprisingly different efficiencies, one may conclude that either great control must be exercised when the experiment is conducted or some sort of robustified approach may have to be sought.

Moreover, considering different parameter subsets per default, creates a foundation for deciding whether some parameter subsets are more important to estimate than others. For instance, if the  $D$ -optimal design has design graphs which contain very different designs and the efficiency profiles do not reveal any of those designs to be particularly robust, then it becomes all the more interesting if it turns out that the parameter subset of primary interest does in fact have easily identifiable and acceptably robust designs associated with them. Note that this aspect of the workflow relies on designs being relatively cheap to compute.

Currently, there's a range of vulnerabilities in the implementation of the suggested workflow, from the quality of the FO approximation to the reliability of the optimisation algorithm. As was seen in Chapter 1 in the case of exponential decay, the FO approximation can fail, with the suggested possible cause being a high degree of nonlinearity of the model. It is hard to envision a FIM approximation which does not entail linearisation and is simultaneously as computationally cheap as FO, so in case FO fails the suggested workflow might have to be abandoned (or allowed a much larger computational budget).

A warning is also warranted when it comes to using aPMS and variance levels to specify the variance component. It allows specification to be done easily, but by that same token it can also be done too easily; it is all too simple to copy a basis variance level from a reference in the literature (this Chapter for instance) and then draw design graphs and efficiency profiles without questioning whether one is looking at the right ranges or starting from the right basis. This can of course be alleviated if one has access to prior information in the form of, say, historical data or knowledge specific to the parameters (for instance, chemical engineers might have a reasonable a priori idea of activation energies in reaction kinetics as well as their typical variability between reactions).

Problems with the optimisation algorithm were also seen in Chapter 1 but the prevalence of this problem has been confirmed in this chapter; several design graphs and efficiency profiles indicated problems with distinguishing between local optima. It is not trivial to prescribe a general solution to this problem; a naive approach is to increase the computational budget (population size, number of iterations) given to the global optimisation problem. Alternatively one could investigate whether the objective functions (or special cases) have attractive properties that can be used to pick a more specialised algorithm; for instance, the FO objective function is almost certainly smooth - this makes it a candidate for a global optimiser for smooth functions, such as the convex relaxation regression suggested by Azar et al. (2016). Lastly, the investigations in this chapter suggest certain heuristics might be worthwhile; for instance, since the largest efficiency losses due to optimisation failure are associated with a failure to identify the correct replication structure, one could consider simply optimising different versions of the objective function, each restricted to a particular replication structure. However, this would in all likelihood require a manageable (short) list of candidate replication structures in order to be feasible. In cases where two "competing" replication structures are discovered, the sampling times are often roughly identical, so one could simply fix those and then optimise over the replication structure. Also, since certain designs have fewer design points than there are unknown parameters in the model, this suggests a restriction (to designs of size greater than or equal to  $p$ ) of the set over which one optimises is desirable. Suggestions that reflect

practical constraints include sampling windows (i.e., relax the requirement that sampling times be exact) or imposing a minimum amount of time between samples, reflecting the fact that often simultaneous sample-taking is not realistic. A final point against this method is that design graphs are constructed by varying one parameter at a time; thus, the impact of varying several parameters at a time is not investigated. In this regard, most robustified optimality criteria are superior, in that the criteria usually entail optimisation over the joint distribution (or set) of the parameters.

We have discussed failed optimisation, but another kind of failure is that the FO approximation is simply a poor one and in particular, its minimum is far from the minimum of the true objective function. This was seen in the Exponential Decay models in Chapter 1. When this is the case, both design graphs and efficiency profiles may be highly misleading and it may not be possible to detect any problems with the figures themselves. Thus, if it is suspected that the FO approximation may be inappropriate, a user could replicate the sanity check in Figure 1.3. That is, use MC/AGQ to compute an FIM approximation, both for the FO  $D$ -optimal design as well as for a simulated set of random designs and then check the MC/AGQ  $D$ -optimality objective function value of the FO  $D$ -optimal design against those of the random designs. If the FO approximation is any good, then the FO generated design should be superior to all the random designs.

Another limitation of our investigations in this chapter is that while the problem of singular designs (which contain fewer sampling times than parameters to be estimated) is encountered in, e.g., the  $D_A$ -optimal design for the AUC in the One-Compartment model, we do not propose solutions. A related problem is *sparse designs*, which are population designs where different individuals may be allocated to different elementary designs but some or all elementary designs are restricted to contain fewer sampling times than the number of parameters to be estimated. However, it will be seen in Chapter 3 that the `doptim` package which is a part of this work at least theoretically allows the user to address such problems. Because the  $D$ -optimality objective function is made available to the user, she can optimise it over a restricted set which forces a minimal number of distinct sampling times. This would be one (potentially naive) way of avoiding singular designs. As for sparse designs, `doptim` should theoretically allow



the user to generate the relevant  $D$ -optimality objective function, though the closest example treated in Chapter 3 is an unbalanced population  $D$ -optimal design. This has the first feature of sparse designs in that it allows different individuals to be allocated to different elementary designs but the number of distinct sampling times for a given individual is not restricted to be less than the number of parameters to be estimated.

In summary, the conclusion of this chapter is that the proposed workflow is feasible and yields useful conclusions with important practical implications.

## 3 R-packages: doptim and randon

3.1	Introduction to new R-packages doptim and randon . . .	131
3.1.1	“Reduced FIM” . . . . .	132
3.2	Example: Linear Regression with random intercept and slope . . . . .	134
3.3	Example: exponential decay - first nonlinear example . . .	140
3.4	Replicating Wang2012 . . . . .	144
3.4.1	$D_\beta$ -optimal design for a single individual . . . . .	144
3.4.2	Composite $D$ -optimal design . . . . .	146
3.4.3	Population $D$ -optimal design . . . . .	148
3.4.4	Population $D$ -optimal design, unbalanced . . . . .	148
3.4.5	$D_A$ -optimal design . . . . .	150
3.5	Overview of the randon package . . . . .	151
3.5.1	plot_sensBands . . . . .	152
3.5.2	genOptimDesnRange . . . . .	153
3.5.3	Discussion . . . . .	161

### 3.1 Introduction to new R-packages doptim and randon

In this chapter the new R-packages doptim and randon are introduced and described. The doptim package is for generating objective functions for user-specified problems and the randon package is a wrapper for doptim which enables the creation of design graphs and efficiency profiles. In the sections below, code examples are provided in which doptim is used to solve a series of problems of increasing complexity, starting from a simple linear model and ending with a One-Compartment model. Code using randon to reproduce some of the figures in chapter 2 will also be provided and explained.

**The randon package is intended to encapsulate the “middle road” workflow described in the introduction to Chapter 2.** By writing a minimal amount of code, intuitively accessible graphs (design graphs and efficiency profiles) are generated. These should allow a practitioner to explore how knowledge (and uncertainty) concerning parameters translates into different experimental designs and different exposures to risk of inefficiency of those designs. The randon package also implements the computation of aPMS which is used per default to provide solutions where variance is investigated as “variance level”, as discussed in the previous chapter. That is, the first axis of design graphs and efficiency profiles generated using randon is “variance level” (variance magnitude viewed as average perturbation in profile caused by the relevant random effect, proportional to the magnitude of the typical profile). All of the design graphs and efficiency profiles in Chapter 2 were generated using randon.

**The new R-packages can solve problems not treated in previous chapters.** The doptim package is the “computational core” of this work. The genObjFun in doptim can be used to generate objective functions for problems like those in previous chapters. For users who want a finer level of control, genObjFun can also be used to address the following additional problems:

- Creating objective functions for the case where interest is in estimating variance components

- Generating so-called composite objective functions (reflecting a weighted interest in estimating variance components and fixed effects, see Wang et al. (2012))
- Allowing different elementary designs for different individuals in a population experimental design

All the use-cases described above will be illustrated in the sections below.

**It is easy to generate objective functions with `doptim`.** The `genObjFun` function is used as follows:

- The user specifies an NLME model and a (vector of) desired size(s) of the elementary design(s) to which individuals are to be allocated. This is easiest to do using a list of arguments, which can be amended as the problem is redefined or expanded
- The arguments are supplied to `genObjFun`. This generates an objective function
- The objective function is optimised using an appropriate optimiser available in R, e.g., `nlsminb`

**The user must make choices regarding FIM approximation and the type of criterion being employed.** The user can choose between

- First Order (FO) and Integrated First Order (InFO)  $D$ -optimality. Definitions can be found on page 41 and page 41, respectively.
- Using reduced and full FIM (see below)
- Returning  $\log[\det(\text{FIM})]$  or the whole FIM. This is of interest when singular FIMs occur or when custom objective functions are desired.
- $D$ - or  $D_A$ -optimality, i.e., the user can specify a function of  $\theta$  as the parameter of interest (see pages 39 and 80 for formal definitions)
- Local or pseudo-Bayesian  $D$ -optimality (see below)

Among the features mentioned above, we have yet to explain what is meant by “reduced FIM” and “pseudo-Bayesian  $D$ -optimality”.

### 3.1.1 “Reduced FIM”

In Section 1.2.1, the FIM approximation from Retout and Mentre (2003) was given, for the fixed effects part of the FIM (i.e., the sub-matrix involving

derivatives with respect to entries of  $\beta$ ). In the same paper, they in fact give an approximation for the FIM in the case where all of  $\theta = (\beta, \omega, \sigma^2)$  is unknown, i.e., fixed effects, variance components and the residual variance all need to be estimated. In this case, the linearised model has FIM

$$\text{FIM}_{lin}(\bar{\mathbf{b}}) = \sum_{i=1}^n \begin{bmatrix} \mathbf{A}_i(E, V) & \mathbf{C}_i(E, V) \\ \mathbf{C}_i^\top(E, V) & \mathbf{B}_i(E, V) \end{bmatrix}$$

Here, with  $E$  and  $V$  as in Section 1.2.1, the blocks are given by

$$\begin{aligned} (\mathbf{A}_i(E, V))_{kl} &= \frac{\partial E_i^\top}{\partial \beta_k} V_i^{-1} \frac{\partial E_i}{\partial \beta_l} + \frac{1}{2} \text{trace} \left( \frac{\partial V_i}{\partial \beta_l} V_i^{-1} \frac{\partial V_i}{\partial \beta_k} V_i^{-1} \right) \text{ for } k, l = 1, \dots, p \\ (\mathbf{B}_i(E, V))_{kl} &= \text{trace} \left( \frac{\partial V_i}{\partial \omega_k} V_i^{-1} \frac{\partial V_i}{\partial \omega_l} V_i^{-1} \right) \text{ for } k, l = 1, \dots, p + 1 \\ (\mathbf{C}_i(E, V))_{kl} &= \text{trace} \left( \frac{\partial V_i}{\partial \omega_k} V_i^{-1} \frac{\partial V_i}{\partial \beta_l} V_i^{-1} \right) \text{ for } k = 1, \dots, p + 1 \text{ and } l = 1, \dots, p \end{aligned}$$

where we use the convention that  $\omega_{p+1} = \sigma^2$ , for notational convenience. Following Nyberg et al. (2012), one can assume “independence between  $\beta$  and  $V_i$ ”, for computational simplicity. Mathematically, this means assuming  $\frac{\partial V_i}{\partial \beta_k} = 0$  for  $k = 1, \dots, p$ . Then the approximation above simplifies to

$$\text{FIM}_{lin}(\bar{\mathbf{b}}) = \sum_{i=1}^n \begin{bmatrix} \mathbf{A}_i^*(E, V) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_i(E, V) \end{bmatrix},$$

where the trace term has been dropped from the fixed effects part of the matrix, i.e.,

$$(\mathbf{A}_i^*(E, V))_{kl} = \frac{\partial E_i^\top}{\partial \beta_k} V_i^{-1} \frac{\partial E_i}{\partial \beta_l} \text{ for } k, l = 1, \dots, p$$

If the boolean argument `reduced` is set to `TRUE` in `genObjFun`, the approximation described above will be used, in place of the “full FIM”. Note that all results in Chapters 1 and 2 were produced using `reduced=FALSE`.

The next sections can be considered to constitute a beginner’s guide to using `doptim`.

### 3.2 Example: Linear Regression with random intercept and slope

This example covers the basic usage of the `genObjFun` function in the `doptim` package, namely to generate a  $D$ -optimality objective function for a user-specified NLME model. In this case, the model is a simple linear regression with random effects.

For this model,  $\eta$  is of the form

$$\eta(\boldsymbol{\beta}, t) = \beta_1 + \beta_2 t = (\beta_1 + \beta_2) \begin{pmatrix} 1 \\ t \end{pmatrix},$$

with  $t \in [a, b]$  for a fixed choice of  $a, b > 0$  with  $a < b$ . Since  $\eta$  is linear in  $\boldsymbol{\beta}$  this will technically lead to an LME model, which is not representative of the models this thesis is otherwise concerned with. However, an LME model is merely a special case of an NLME model (recall,  $\eta$  is allowed but not required to be nonlinear in  $\boldsymbol{\beta}$ ).

One reason to consider such a simple model to begin with is that it is one of the few cases where explicit results regarding  $D$ -optimal designs are available. In the fixed effects case, and with an even number of samples, it is known (see e.g. page 70 of Atkinson et al. (2007)) that the  $D$ -optimal design assigns half of the samples to  $t = a$  and the other half of the samples to  $t = b$ . Therefore, it might be expected that when  $\omega_1$  and  $\omega_2$  are both small, the  $D$ -optimal design for estimating  $\boldsymbol{\beta}$  will be of a similar form.

The linear mixed effects model has two more features which enable a direct sanity check of the code. First, the FIM admits an explicit expression (Wand, 2007). Second, the FO approximation is exact. Therefore, we can compute the FIM both ways and check that the results agree.

The parameters of the model are assumed to be

$$\boldsymbol{\beta} = (3, 5) \quad \Omega = \text{diag}\{.1, .1\} \quad \sigma^2 = 0.25^2,$$

and the first task is to find a  $D_{\boldsymbol{\beta}}$ -optimal design for a single individual ( $N = 1$ ) for whom six samples are taken ( $k = 6$ ) in the time interval  $[0, 10]$ .

To generate the  $D_{\boldsymbol{\beta}}$ -optimality objective function run the following R-code:

```

1 | objFun <- genObjFun(
2 |   eta = expression(
3 |     (beta1 + b1) + (beta2+b2) * t
4 |   ),
5 |   Ebeta = c(0, 1),
6 |   Vb = c(Vb1 = 1, Vb2 = 1)/10,
7 |   Veps = .25^2,
8 |   phi = "fixed",
9 |   desvarNames = "t",
10 |  noSamples = c(6)
11 | )

```

The arguments to `genObjFun` clearly specify the model described above as well as some aspects of the design problem. Setting `phi = "fixed"` indicates that the objective function is based on the  $D_\beta$ -optimality criterion. Setting `noSamples = c(6)` indicates that the objective function expects designs with six sampling times, for a single individual.

The objective function can now be evaluated in a design of the user's choice, provided it is of the correct length:

```

1 | Xi0 <- c(1, 2, 4, 6, 8, 10)
2 | objFun(Xi0)

```

```
# returns #
```

```
[1] -4.276033
```

One aspect of the design problem which is not specified above is the experimental region, i.e., that  $[a, b] = [0, 10]$ . That is addressed in the optimisation step where the user is free to employ any optimisation algorithm they deem appropriate. Given that this is a bounded optimisation problem, the `nlmminb` function (available in the pre-loaded package `stats`) can be used:

```

1 | nlmminb(
2 |   start = Xi0,
3 |   objective = objFun,
4 |   lower = 0,

```

```

5 |     upper = 10
6 | )

# returns #

$`par`
[1] 0 0 0 0 0 10

$objective
[1] -4.480053

$convergence
[1] 0

$iterations
[1] 20

$evaluations
function gradient
20      121

$message
[1] "both X-convergence and relative convergence (5)"

```

The list of outputs from `nlsminb` includes `'par'` which is the optimum identified by the algorithm. In this case, the optimum identified by the algorithm is different from the theoretical optimum in the fixed effects case. A simple way to investigate this further is to reduce the magnitude of the variance component and re-run the code. Since it takes at least seven arguments to specify a *D*-optimality problem, it is advisable to save these arguments in a dedicated list so as to avoid duplicating code:

```

1 | args <- list(
2 |   eta = expression(
3 |     (beta1 + b1) + (beta2+b2) * t
4 |   ),

```



```

5 | Ebeta = c(0, 1),
6 | Vb = c(Vb1 = 1, Vb2 = 1)/10,
7 | Veps = .25^2,
8 | phi = "fixed",
9 | desvarNames = "t",
10 | noSamples = c(6)
11 | )

```

Now it is simple to reduce the variance components and re-run the problem with the rest of the problem retained:

```

1 | args$Vb = c(Vb1 = 1, Vb2 = 1)/10000
2 |
3 | objFun <- do.call(genObjFun, args)
4 |
5 | nlminb(
6 |   start = Xi0,
7 |   objective = objFun,
8 |   lower = 0,
9 |   upper = 10
10 | )

```

```
# returns #
```

```
$`par`
```

```
[1] 0 0 0 10 10 10
```

```
$objective
```

```
[1] -11.94752
```

```
$convergence
```

```
[1] 0
```

```
$iterations
```

```
[1] 12
```

```
$evaluations
```

```

function gradient
12         72

$message
[1] "both X-convergence and relative convergence (5) "

```

With the variance component reduced, `nlminb` yields the same optimum as the fixed effects version of the problem, i.e., it splits the samples evenly between the two boundaries of the time interval.

We can also perform the sanity check mentioned above, i.e., check that the explicit form of the FIM agrees with the FO approximation. Since this is just for the purpose of illustration, we will perform the check for a single, but arbitrary design, produced by sampling six times from the uniform distribution on  $[0, 10]$ .

In the code below, we first generate the random design and then the objective function for the same problem as above, but with one difference: We set `returnFIM=TRUE` in `genObjFun` to make the resulting function return the FO FIM approximation.

```

1  # fix seed for reproducibility
2  set.seed(123)
3  # generate random design
4  Xi_random <- runif(n = 6, min = 0, max = 10)
5  # reset the variance parameters
6  args$Vb <- c(1, 1)/10
7  # argument to return FIM rather than objective
   function value
8  args$returnFIM = TRUE
9  # generate ``objective function''
10 objFun <- do.call(genObjFun, args)
11 # evaluate objective function
12 objFun(Xi_random)

# returns #
[[1]]
[,1]      [,2]      [,3]      [,4]      [,5]

```

```
[1,] 7.163227 0.374001 0.00000000 0.00000000 0.00000000
[2,] 0.374001 9.926123 0.00000000 0.00000000 0.00000000
[3,] 0.000000 0.000000 25.65590797 0.06993839 16.1444590
[4,] 0.000000 0.000000 0.06993839 49.26396022 0.4747473
[5,] 0.000000 0.000000 16.14445898 0.47474729 522.6655929
```

The code below implements the explicit expression for the FIM as found in Wand (2007).

```
1 # fixed effects part
2 X <- cbind(1, Xi_random)
3 Omega <- diag(args$Vb)
4 V <- X %*% Omega %*% t(X) + diag(args$Veps, 6)
5 A <- t(X) %*% solve(V) %*% X
6
7 # variance component part
8 B <- matrix(NA, 3, 3)
9 for(i in 1:3){
10   for(j in 1:3){
11     if( i < 3) Zi = X[,i]
12     if( j < 3) Zj = X[,j]
13     if (i == 3) Zi = diag(1, 6)
14     if (j == 3) Zj = diag(1, 6)
15     tmp <- t(Zi) %*% solve(V) %*% Zj
16     tmp <- tmp %*% t(tmp)
17     if(dim(tmp)[2] > 1) tmp <- diag(tmp)
18     B[i, j] <- 1/2 * sum(tmp)
19   }
20 }
21
22 # to prettify
23 dimnames(A) = NULL
24 # to print together
25 list(A, B)
```

# returns #

```

[[1]]
[,1]      [,2]
[1,] 7.163227 0.374001
[2,] 0.374001 9.926123

[[2]]
[,1]      [,2]      [,3]
[1,] 25.65590797 0.06993839 16.1444590
[2,] 0.06993839 49.26396022 0.4747473
[3,] 16.14445898 0.47474729 522.6655929

```

Which equals the previous result. As a sanity check, this provides confidence that the code works.

### 3.3 Example: exponential decay - first nonlinear example

A simple, but important, example where the regression function is nonlinear is exponential decay. This is another case where a general expression for the  $D$ -optimal design has been found for the fixed effects case (Kitsos and Kolovos, 2013).

The regression function is

$$\eta(\boldsymbol{\beta}, t) = \beta_1 \exp(\beta_2 t),$$

the parameters of the model are assumed to be

$$\boldsymbol{\beta} = (1, -1/10) \quad \Omega = \text{diag}\{1/100, 1/100\} \quad \sigma^2 = 1/1000$$

and the experimental region is  $t \in [0, 30]$ . To specify and solve the problem, again searching for a single elementary  $D_\beta$ -optimal design of length six:

```

1 | args <- list(
2 |   eta = expression(

```

```
3     (beta1 + b1) * exp((beta2 + b2)*t)
4   ),
5   Ebeta = c(1, - 1/10),
6   Vb = c(1, 1)/100,
7   Veps = 1/1000,
8   phi = "fixed",
9   desvarNames = "t",
10  noSamples = c(6)
11 )
12
13
14 objFun <- do.call(genObjFun, args)
15
16 Xi0 = 1:6/10
17
18 nlmminb(
19   start = Xi0,
20   objective = objFun,
21   lower = 0,
22   upper = 30
23 )

# returns #

$`par`
[1] 0.000000 0.000000 0.000000 3.894444 3.894445 25.562844

$objective
[1] -14.13597

$convergence
[1] 0

$iterations
[1] 23
```

```
$evaluations
function gradient
24      157
```

```
$message
[1] "relative convergence (4)"
```

According to Kitsos and Kolovos (2013), the  $D$ -optimal design for the fixed effects case has support points  $(0, -1/\beta_1) = (0, 10)$ . Again, one can check if decreasing the magnitude of the variance component produces designs that are closer to the fixed effects case. In the code below, a decreasing sequence of variance component magnitudes is used to generate a sequence of  $D_\beta$ -optimal designs:

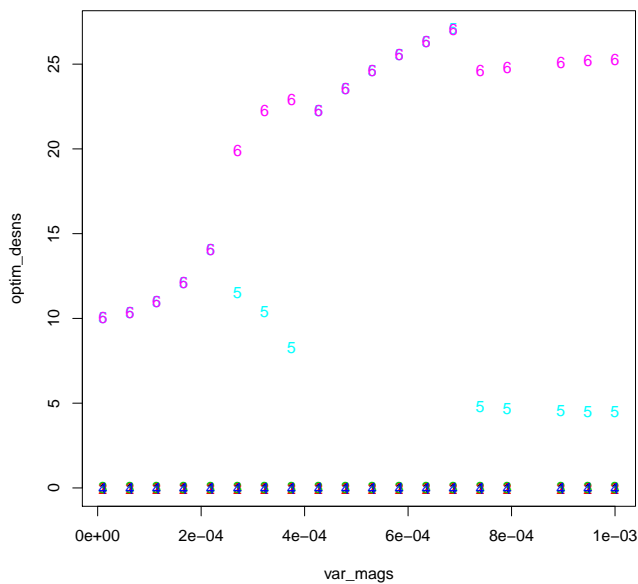
```
1 | var_mags <- seq(1/10^3, 1/10^5, length.out = 20)
2 |
3 | optim_desns <- t(sapply(var_mags,
4 |   function(x){
5 |     args$Vb <- c(1, 1) * x
6 |     objFun <- do.call(genObjFun, args)
7 |     res <- try(nlminb(
8 |       start = Xi0,
9 |       objective = objFun,
10 |      lower = 0,
11 |      upper = 30
12 |    )$par)
13 |     if(class(res) == "try-error") return(rep(
14 |       NA, length(Xi0)))
15 |     return(res)
   |   )))
   |
# returns #
```

```
Error in chol.default(x) :
the leading minor of order 2 is not positive definite
```

Note that in a single case, numerical issues caused the matrix inversion of the variance of the linearised model to fail, hence the `chol.default` error message. This does happen from time to time and the solution will be context specific - there might be a problem with the starting design provided or the optimisation problem might be genuinely hard. While it can be cause to investigate further we will not do so in this demonstration. Base R can be used to generate a simple design graph with the following code:

```
1 | matplot(var_mags, optim_desns)
```

Which produces the plot in Figure 3.1 for the 19 successfully computed designs.



**Figure 3.1:** A simple version of the design graphs found in earlier chapters. The plotting symbol is the index of the sampling time in the individual  $D_\beta$ -optimal design being plotted and the x-axis is the variance magnitude used to generate the design. Note that over-plotting occurs, with time labels “1” through “4” being consistently placed on top of each other and time label “6” being on top of “5” for some of the variance values.

As can be seen from Figure 3.1, when the magnitude of the variance component is reduced sufficiently, the  $D_\beta$ -optimal designs have two sampling times, the first at zero and the second at a time which converges to 10. This is consistent with the above-mentioned result from Kitsos and Kolovos (2013).

### 3.4 Example: replicating results from Wang et al. (2012)

In Wang et al. (2012), several kinds of  $D$ -optimal design are found for a specific NLME model. In this example, some of those results are reproduced using the `doptim` package. The regression function is motivated by a One-Compartment model:

$$\eta(\boldsymbol{\beta}, t) = \frac{\exp(\beta_1 - \beta_3)}{\exp(\beta_1) - \exp(\beta_2)} \{ \exp(-\exp(\beta_2)t) - \exp(-\exp(\beta_1)t) \} .$$

Note that the model is parametrised so as to ensure positivity of rate constants ( $\exp(\beta_1)$  and  $\exp(\beta_2)$ ) and volume of distribution ( $\exp(\beta_3)$ ).

#### 3.4.1 $D_\beta$ -optimal design for a single individual

In the code below the model is defined, with parameter values as in Wang et al. (2012). To begin with, the problem is given as optimizing the design for a single individual (`noSamples` has length one) and the focus is on finding a  $D_\beta$ -optimal design (`fixed` is set equal to `TRUE`). The search is restricted to designs consisting of up to five distinct sampling times (`noSamples` is set equal to five).

```

1 | args0 <- list(
2 |   eta = expression(
3 |     exp((beta2 + b2) - (beta3 + b3)) /
4 |     (exp(beta2+b2) - exp(beta1+b1)) *
5 |     (exp(-exp(beta1+b1)*t) - exp(-exp(beta2 + b2)
6 |       *t))
   ),
```



```

7 | Ebeta = c(0, 1, 1),
8 | Vb = c(0.05, 0.05, 0.1),
9 | Veps = 0.05^2,
10 | phi = "fixed",
11 | desvarNames = "t",
12 | noSamples = c(5),
13 | reduced = TRUE
14 | )

```

After this the objective function can be generated and optimised.

```

1 | objFun <- do.call(genObjFun, args0)
2 |
3 | Xi0 <- seq(0, 4, length.out = 5)
4 | nlminb(start = Xi0, objective = objFun, lower = 0,
   |       upper = 4)

```

# returns #

```

$`par`
[1] 0.1961514 0.1961512 0.9437235 2.5528683 2.5528691

```

```

$objective
[1] -3.170249

```

```

$convergence
[1] 0

```

```

$iterations
[1] 16

```

```

$evaluations
function gradient
19      107

```

```

$message
[1] "relative convergence (4)"

```

The resulting design, along with the value of the objective function, correspond exactly with the first row of Table 1 in Wang et al. (2012). Note the argument `reduced = TRUE` in the code; this indicates that in the approximation of the FIM, the "trace term" should be set to zero. More specifically, recall from Section 1.2 that the linearised model had FIM

$$\text{FIM}_{lin}(\bar{\mathbf{b}})_{kl} = \sum_{i=1}^n \left[ \frac{\partial E_i^\top}{\partial \beta_k} V_i^{-1} \frac{\partial E_i}{\partial \beta_l} + \frac{1}{2} \text{trace} \left( \frac{\partial V_i}{\partial \beta_l} V_i^{-1} \frac{\partial V_i}{\partial \beta_k} V_i^{-1} \right) \right].$$

The "reduced" version omits the second term of this approximation. In Wang et al. (2012), this is recommended for cases where the variance component is deemed "small".

### 3.4.2 Composite $D$ -optimal design

In Wang et al. (2012), the *composite  $D$ -optimality criterion* is introduced, which allows the user to express a weighted interest in the fixed effects and variance components, respectively. Specifically, for a chosen  $\lambda \in [0, 1]$ , the composite  $D$ -optimality criterion is to find the design  $\xi^* \in \Xi$  which maximizes

$$\lambda \cdot \log \det \text{FIM}(\xi^*)_{\beta} + (1 - \lambda) \cdot \log \det \text{FIM}(\xi^*)_{\omega, \sigma^2}$$

The following code replicates a result from Wang et al. (2012), second row of Table 2. For the same model as before, the composite  $D$ -optimality objective function is created, with  $\lambda = 0.5$ , and the optimal design is identified:

```

1 | args <- args0
2 | args$returnFIM <- TRUE
3 | objFun <- do.call(genObjFun, args)
4 |
5 | f1 <- function(x) - log(det(objFun(x)[[1]][1:3,
   |   1:3]))

```

```

6 | f2 <- function(x) - log(det(objFun(x)[[1]][-(1:3),
   |   -(1:3)]))
7 | f <- function(x) .5 * f1(x) + .5*f2(x)
8 |
9 | Xi0 <- seq(0, 4, length.out = 5)
10 | nlminb(start = Xi0, objective = f, lower = 0, upper
    |   = 4)[c("par", "objective")]

```

```
# returns #
```

```
$`par`
```

```
[1] 0.2331545 0.2331545 1.0261762 2.1961446 2.1961435
```

```
$objective
```

```
[1] -11.75895
```

The second line of code adds a new argument, `returnFIM`, which is set to `TRUE`. This changes the return value of the generated function from the log-determinant of the FIM to simply the FIM approximation itself. The remaining lines of code directly specify the composite  $D$ -optimality objective function, followed by the optimisation step.

To replicate the first row of Table 2 in Wang et al. (2012), where  $\lambda = 0$ , simply run

```

1 | nlminb(start = Xi0, objective = f2, lower = 0,
   |   upper = 4)[c("par", "objective")]

```

```
# returns #
```

```
$`par`
```

```
[1] 0.2748855 0.2748866 1.6424521 1.6424451 1.6424473
```

```
$objective
```

```
[1] -20.88068
```

### 3.4.3 Population $D$ -optimal design

The `genObjFun` can also handle population  $D$ -optimal designs. In this case, the design problem involves several individuals who may be allocated to different elementary designs. In the code below, the `args` list is modified to specify that two individuals are to be allocated to two different elementary designs, each with three sampling times.

```

1 | args <- args0
2 | args$noSamples <- c(3, 3)
3 | objFun <- do.call(genObjFun, args)
4 | Xi0 <- c(1:3, 1:3)
5 | nlmnb(start = Xi0, objective = objFun, lower = 0,
      |     upper = 4)[c("par", "objective")]

# returns #

$`par`
[1] 0.2192545 0.9634978 2.4872472 0.2192544 0.9634977 2.4872472

$objective
[1] -4.297543

```

Note that the `noSamples` argument is used to indicate the number of individuals (by the length of the vector provided, i.e., two) and the sizes of the elementary designs (by the integer valued entries of the vector provided, i.e., both three). The first three values of the optimisation result constitute the first elementary design of the population  $D$ -optimal design and the last three values constitute the second elementary design. This example replicates the first row of Table 3 in Wang et al. (2012). They note that while the two elementary designs are identical in the optimal design for this problem, that does not generally have to be the case.

### 3.4.4 Population $D$ -optimal design, unbalanced

This case is an extension of the previous case with multiple individuals. Here, three groups of different sizes (with group sizes chosen by the user)

are allocated to elementary designs of different lengths (with the design lengths chosen by the user). Specifically, the first group consists of 30 individuals allocated to a design of length three, the second group of 4 individuals allocated to a design of length two and the last group of 2 individuals allocated to a single point design. The design problem is now to optimally choose the six design points that make up the three elementary designs. Specifically, a  $D_\beta$ -optimal design is sought and the trace term is omitted (so `reduced = TRUE`).

```

1  args <- args0
2  args$returnFIM <- TRUE
3
4  args1 <- args
5  args1$noSamples <- 3
6  f1 <- function(Xi) do.call(genObjFun, args1)(Xi)
   [[1]][1:3, 1:3]
7
8  args2 <- args
9  args2$noSamples <- 2
10 f2 <- function(Xi) do.call(genObjFun, args2)(Xi)
   [[1]][1:3, 1:3]
11
12 args3 <- args
13 args3$noSamples <- 1
14 f3 <- function(Xi) do.call(genObjFun, args3)(Xi)
   [[1]][1:3, 1:3]
15
16
17 f <- function(Xi) -log(det(
18   30*f1(Xi[1:3]) + 4*f2(Xi[4:5]) + 2*f3(Xi[6])
19   ))
20
21 case <- DEoptim::DEoptim(f, lower = rep(0, 6),
   upper = rep(4, 6))
22 nlminb(start = case$optim$bestmem, objective = f,
   lower = 0, upper = 4)[c("par", "objective")]

```

```
# returns #

$`par`
par1      par2      par3      par4      par5      par6
0.2170318 0.9784380 2.4956097 2.4309390 0.7556319 0.2550840

$objective
[1] -12.74138
```

NB: the reason we need to define `args1`, `args2` and `args3` is lazy evaluation in R; if we instead kept updating `args`, the resulting functions would *all* use the last definition of `args`.

The above results replicate row one of Table 4 in Wang et al. (2012). Note that the FIM for the group of individuals is equal to the sum of the elementary FIMs, due to independence between individuals.

### 3.4.5 $D_A$ -optimal design

The following example is not included in Wang et al. (2012), but it serves to illustrate one last capability of `doptim`, namely generating  $D_A$ -optimality objective functions. In the code below, the element `phi` in `args`, the list of arguments, is modified to contain a list of expressions. The first element of this list is the expression for Area Under the Curve (AUC), which was given in Equation 2.4 and the second is the expression for the time to maximum response ( $t_{max}$ ), given in Equation 2.5. Thus, the code below generates and optimises a  $D_A$ -optimality objective function for the parameter transformation  $\phi = (AUC, t_{max})$ .

```
1 | args <- args0
2 | args$phi <- list(
3 |   expression(exp(beta2-beta3)/(exp(beta2) - exp(
4 |     beta1)) * (exp(-beta1) - exp(-beta2))),
5 |   expression((beta1 - beta2)/(exp(beta1)-exp(beta2)
6 |     )))
5 | objFun <- do.call(genObjFun, args)
6 | nlmnb(start = Xi0, objective = objFun, lower = 0,
   |       upper = 4)[c("par", "objective")]
```

```
# returns #

$par
[1] 0.1702486 0.1702488 1.0757593 2.4337100 2.4337135

$objective
[1] -6.685038
```

### 3.5 Overview of the randon package

The sensitivity band plots, design graphs and efficiency profiles in Chapter 2 were created using the `randon` package. In this section, the reader will be taken through simple examples of how to use `randon` to generate similar results for their own problem.

The package makes three main functions available to the user:

<code>calc_aPMS</code>	Compute a first order approximation of aPMS for a given nonlinear regression function, parameter and interval. The aPMS is an attempt to measure the average deviation in the regression function over the interval (in NLME terms, the population level profile), caused by a perturbation in a single fixed effects parameter. It can be used as a starting point to find realistic priors for marginal variances of (additive) random effects.
<code>plot_sensBands</code>	Plot regression function with aPMS sensitivity bands
<code>genOptimDesnRange</code>	A wrapper to generate a range of optimal designs for given model, parameter transformation and range of variance levels.

Of these functions, only `plot_sensBands` and `genOptimDesnRange` are envisioned as part of the workflow suggested in this thesis; the `calc_aPMS`

function is used by the other two functions in `randon` to translate between variance level and magnitude of variance components.

### 3.5.1 `plot_sensBands`

This function generates plots of the regression function, with added sensitivity bands, as seen in Figure 2.1. The following code partially recreates the plot in the first panel of the Figure, i.e., 10% sensitivity bands for the One-Compartment model.

```

1 | m <- list(
2 |   eta = expression(
3 |     (exp(beta2) / exp(beta3)) /
4 |     (exp(beta2) - exp(beta1)) *
5 |     (exp(-exp(beta1) * t) - exp(-exp(beta2) *
6 |       t))
7 |   ),
8 |   beta = list(beta1 = 0, beta2 = 1, beta3 = 1),
9 |   tInt = c(0, 4)
10 | )
11 | plot_sensBands(model = m, VLrange_variedParam = .1)

```

The `model` argument must be a list with three named elements

- `eta` is an expression which defines the regression function, but *without* any random effects specified. The parameters must be named `beta1`, `beta2`, etc. The covariate must be named `t`.
- `beta` is a list with entry names matching the corresponding fixed effects parameters. The entries are scalars setting the values of the parameters.
- `tInt` is a vector of length two, defining the (time) interval which the covariate `t` is restricted to for the purposes of the sensitivity plot.

By supplying the `VLrange_variedParam` argument, the user specifies which variance levels she wishes to visualise for the parameter being varied. Per default, the parameter being varied is "beta1", because it always exists.



If the user wishes to visualise several variance levels in the same plot, she can supply a vector-valued `VRange_variedParam` argument, as follows.

```
1 | plot_sensBands(model = m,
2 |               VRange_variedParam = c(.1, .2, .3))
```

This code recreates the first panel of Figure 2.1. Lastly, to fully recreate Figure 2.1, one can supply a vector of parameter names to the function.

```
1 | plot_sensBands(model = m,
2 |               VRange_variedParam = c(.1, .2, .3),
3 |               paramName_variedParam = c("beta1", "
      beta2", "beta3"))
```

This generates three plots, one per entry in the vector `c("beta1", "beta2", "beta3")` supplied to `paramName_variedParam`, each with three sensitivity bands, one per entry in the vector `c(.1, .2, .3)` supplied to `VRange_variedParam`.

### 3.5.2 `genOptimDesnRange`

This function was used to generate the design graphs and efficiency profiles in Chapter 2. The seed has been set for reproducibility; the first optimisation is a global optimisation using `DEoptim` which is a stochastic algorithm.

#### Reproducing plots from the One-Compartment model example

To generate a range of optimal designs for the One-Compartment model specified above, use the code

```
1 | optDes_beta1 <- genOptimDesnRange(m, noSamples = 5)
```

Here, the only arguments supplied to the function are the model definition, `m`, and the `noSamples` argument, the latter of which indicates the size of the design. If `noSamples` is not supplied, the value defaults to the number of fixed effects parameters in the model. When the line of code is run, some descriptive output is printed to the screen and a list of objects is returned by the function and saved in the `optDes_beta1` variable. In the

standard case, which this should be, the descriptive output has two distinct sections. First, the diagnostics from the global optimisation which is used to generate the first  $D$ -optimal design are printed; one line per iteration. These are standard outputs from the DEoptim function from the package of the same name. It has the following form:

```

startDesn is NULL. Performing global optimisation to identify appropriate startDesn.
Iteration: 1 bestvalit: -24.308162 bestmemit: 0.669019 2.730682 1.151220 0.258034 1.978325
Iteration: 2 bestvalit: -24.308162 bestmemit: 0.669019 2.730682 1.151220 0.258034 1.978325
Iteration: 3 bestvalit: -24.308162 bestmemit: 0.669019 2.730682 1.151220 0.258034 1.978325
Iteration: 4 bestvalit: -24.308162 bestmemit: 0.669019 2.730682 1.151220 0.258034 1.978325
Iteration: 5 bestvalit: -24.308162 bestmemit: 0.669019 2.730682 1.151220 0.258034 1.978325
Iteration: 6 bestvalit: -24.318806 bestmemit: 0.267714 1.994231 0.320108 0.861529 2.147230
Iteration: 7 bestvalit: -24.318806 bestmemit: 0.267714 1.994231 0.320108 0.861529 2.147230
.
.
Iteration: 196 bestvalit: -24.440543 bestmemit: 0.945508 2.357908 0.259394 0.941260 2.389271
Iteration: 197 bestvalit: -24.440543 bestmemit: 0.945508 2.357908 0.259394 0.941260 2.389271
Iteration: 198 bestvalit: -24.440543 bestmemit: 0.945508 2.357908 0.259394 0.941260 2.389271
Iteration: 199 bestvalit: -24.440543 bestmemit: 0.945508 2.357908 0.259394 0.941260 2.389271
Iteration: 200 bestvalit: -24.440543 bestmemit: 0.945508 2.357908 0.259394 0.941260 2.389271

```

Note that if the user supplies a `startDesn` argument, which should be a vector of the appropriate length (`noSamples`) and with values in the appropriate range (`m$tlnt`) this is instead used as the starting design. The second output which is printed to the screen is a counter which indicates the progress along the chosen range of variance levels, i.e., how many of the desired  $D$ -optimal designs have been computed. It takes the following form:

```

[1] "Vb_val 1 of 41"
[1] "Vb_val 2 of 41"
[1] "Vb_val 3 of 41"
.
.
.
[1] "Vb_val 39 of 41"
[1] "Vb_val 40 of 41"
[1] "Vb_val 41 of 41"

```

The `genOptimDesnRange` function returns a list, which in the example above is stored in the variable `optDes_beta1`. The entries of the list are named as follows.

```
> names(optDes_beta1)
[1] "eta_supplied"          "model"          "paramName_variedParam"
[4] "Vlrange_variedParam"  "Vb_range"       "optimDesign"
[7] "objFuns"              "desnGraph"      "repcount"
[10] "effProfiles"
```

In `optDes_beta1$desnGraph` and `optDes_beta1$effProfiles` the design graph and efficiency profiles can be found, respectively. The plots are presented in Figures 2.25 and 2.26 bearing column label “ $s=(\beta_1, \beta_2, \beta_3)$ ” and row label “Vary  $\omega_1$ ”.

### Exploiting access to objective functions: Spot check of optimality and plotting efficiency

Much of the output listed above is for bookkeeping purposes (for instance, `Vb_range` contains a matrix with all the variance values that have been used, one row per variance level). However, one item worth noting is the entry named `objFuns`. This is a list (“wrapped” in a matrix, for better printing) of the objective functions used to generate the optimal designs in the design graph (they are also stored as a matrix in the output, as `optimDesign`). These objective functions can be accessed directly by the user, just as the output of `genObjFun` can.

As an example of how this might be convenient, consider a scenario in which the user wishes to inspect the optimality function for the purpose of checking that the optimisation actually yielded globally optimal designs. A first step in such an inspection could be to consider profiles of the relevant objective functions, with all but one design point fixed at the optimal design and the final design point being varied. In the code below, the objective function  $\Psi^{(i)} : [0, 4]^5 \rightarrow \mathbb{R}$  for each variance level  $i$  is extracted from `optDes_beta1` from the previous example. Profile functions  $\Psi_{\cdot}^{(i)} : [0, 4] \rightarrow \mathbb{R}$  are then defined by fixing the first four arguments of  $\Psi^{(i)}$  at the

first four values of the  $i$ th optimal design and letting the final sampling time  $t_5$  vary. These profile functions are then plotted.

```

1  # optimal designs from previous optimisation
2  od <- optDes_beta1$optimDesign
3
4  # objective functions used to generate od
5  ofs <- unlist(optDes_beta1$objFuns)
6
7  # range of alternative sampling times for t5
8  t5s <- seq(0, 4, length.out = 100)
9
10 # value of objective functions in range of t5s
11 vof_t5s = sapply(1:nrow(od), function(i){
12   # i'th objective function
13   of_i = ofs[[i]]
14   # i'th optimal design
15   od_i = od[i,]
16   # value of i'th objective function
17   # when modifying t5 in optimal design
18   vof_i_t5s = sapply(t5s, function(t5){
19     of_i(c(od_i[-5], t5))
20   })
21   return(vof_i_t5s)
22 })
23
24 matplot(t5s, vof_t5s,
25         type = "l",
26         col = 1,
27         lty = 1,
28         xlab = expression(t[5]),
29         ylab = "",
30         xlim = c(0, 4.4))
31 text(cbind(4.2, vof_t5s[100,]), paste0("i = ", 1:
      nrow(od)), cex = .6)

```

```

32 | title(ylab=expression({psi[.]}^{(i)}~"("~t[5]~") "
    | ), line=2, cex.lab=1.2)

```

And the final sampling times from the optimal designs can be added to the plot with this code.

```

1 | # value of objective functions in optimal t5s
2 | vof_od = sapply(1:nrow(od), function(i){
3 |   # i'th objective function
4 |   of_i = ofs[[i]]
5 |   # i'th optimal design
6 |   od_i = od[i,]
7 |   # value of i'th objective function
8 |   # when t5 is optimal
9 |   vof_i_od_i = of_i(od_i)
10 |   return(vof_i_od_i)
11 | })
12 |
13 | points(od[,5], vof_od)

```

The resulting plot is presented in Figure 3.2. From a visual inspection, it appears that the final sampling times in the optimal designs stored in `optDes_beta1$optimDesign`, are in fact globally optimal for the profile objective functions. However, the plot also reveals two additional local optima in every profile. So had the local optimiser been supplied with a different starting value, a sub-optimal design could have been identified. Extending this approach (plotting profile objective functions for a single sampling time) to involve pairs of sampling times is straightforward, but three or more sampling times at a time may present a challenge.

While it is tempting to make conjectures about the implications of the relative flatness of some of the profiles in the plot, it is more meaningful to discuss this on the scale of  $D$ -efficiency. The approach taken in the code above can easily be extended to compute the relevant efficiencies. This will also allow the user to address problems where experimenters are considering changing a sampling time, due to feasibility or resourcing constraints.

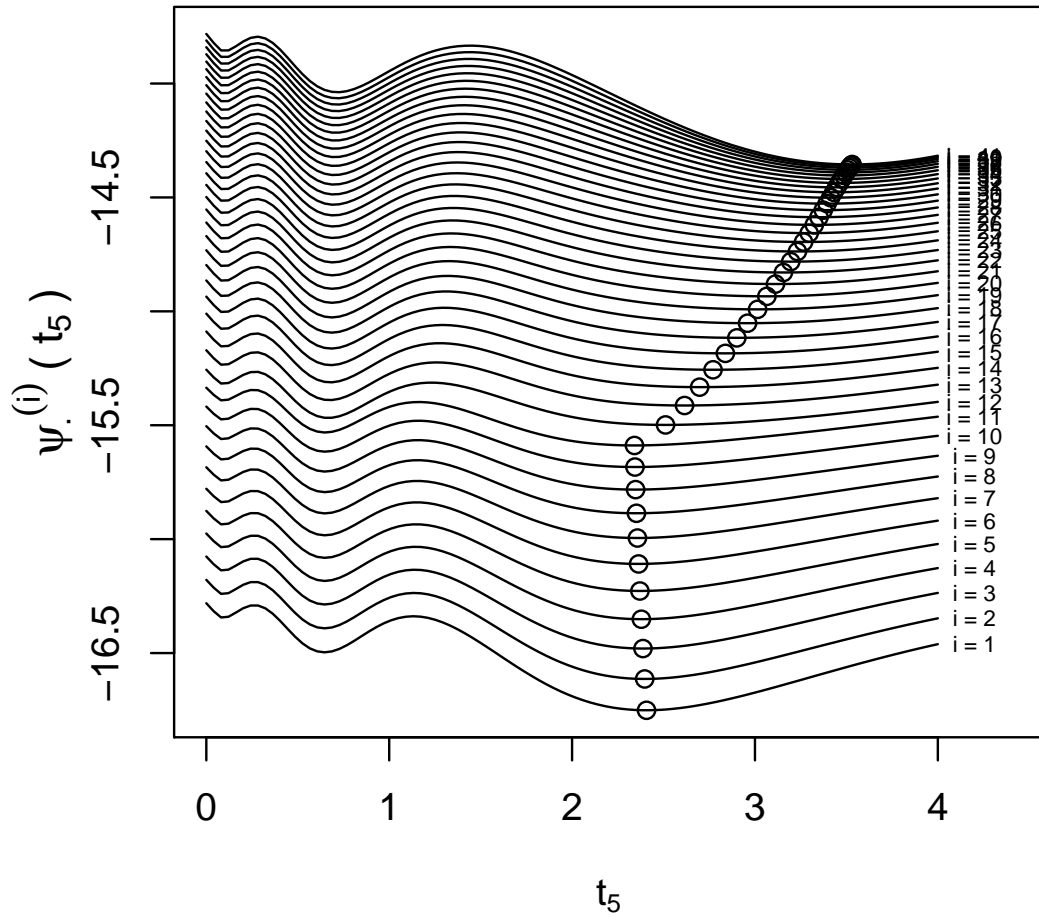
In the code below, the  $D$ -efficiency profiles of the designs under consideration are computed and plotted (with one profile per variance level). The resulting plot is given in Figure 3.3.

```

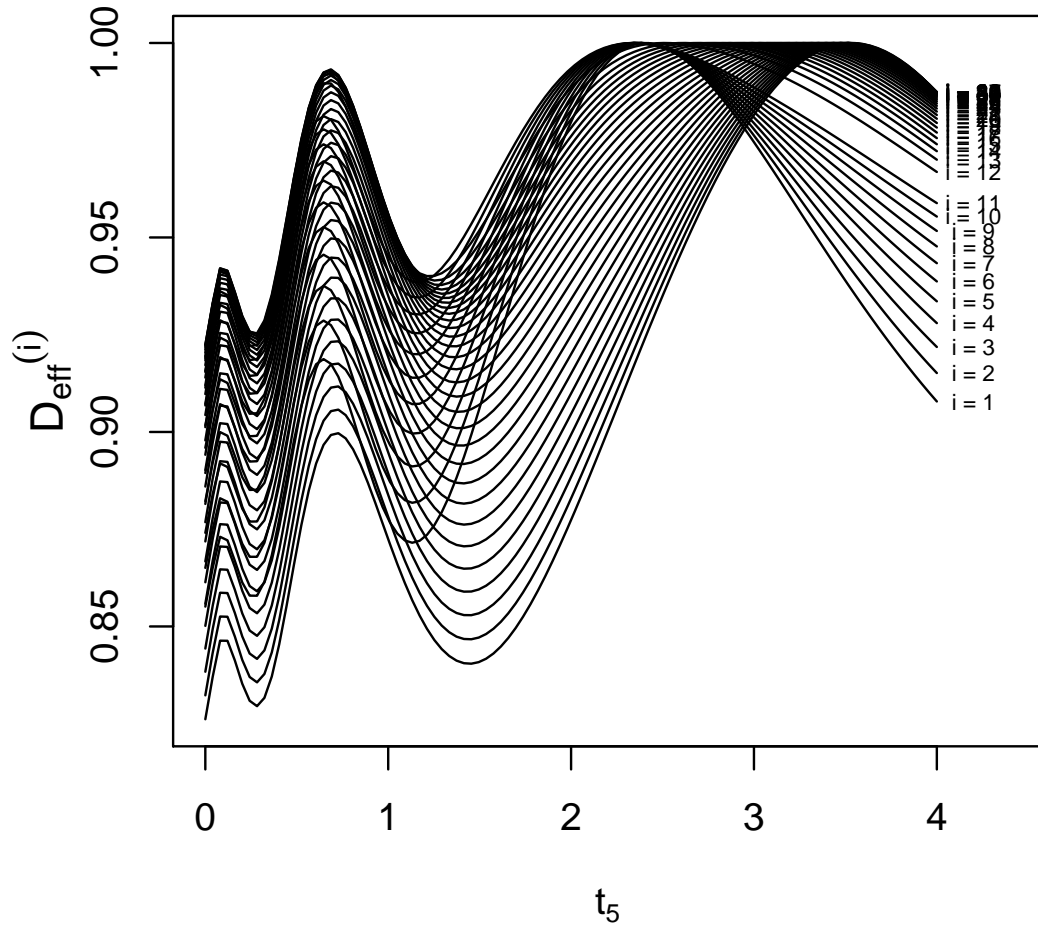
1   # D-efficiency profiles in range of t5s
2   dEff_t5s = sapply(1:nrow(od), function(i){
3     # i'th objective function
4     of_i = ofs[[i]]
5     # i'th optimal design
6     od_i = od[i,]
7     # value of i'th objective function
8     # when modifying t5 in optimal design
9     dEff_i_t5s = sapply(t5s, function(t5){
10      exp(of_i(od_i) - of_i(c(od_i[-5], t5)))^(1/
11        length(optDes_beta1$model$beta))
12    })
13    return(dEff_i_t5s)
14  })
15  matplot(t5s, dEff_t5s,
16          type = "l",
17          lty = 1,
18          col = 1,
19          xlab = expression(t[5]),
20          ylab = "",
21          xlim = c(0, 4.4))
22  text(cbind(4.2, dEff_t5s[100,]), paste0("i = ",
23      1:nrow(od)), cex = .6)
24  title(ylab = expression({D[eff]}^{(i)}), line =
25      2, cex.lab = 1.2)

```

Inspection of Figure 3.3 reveals that the apparent flatness of the objective function profiles in Figure 3.2 is indeed misleading insofar as judging efficiency loss is concerned. Depending on the chosen value of the fifth sampling time, efficiency losses of more than 10% are clearly quite common.



**Figure 3.2:** Profile objective functions for the One-Compartment model example. The hollow circles are the coordinates (time and objective function value) of the final sampling times of the optimal designs. The lines are the profiles of objective functions for the range of variance levels considered. The values of  $i$  which indicate which of the  $i = 1, \dots, 40$  variance levels is being used have been added to the end of the corresponding profiles.



**Figure 3.3:** Efficiency profiles for the One-Compartment model example. Each profile is a plot of the D-efficiency of a design which uses the first four sampling times of the D-optimal design and then replaces the fifth sampling time with the value on the x-axis. The values of  $i$  which indicate which of the  $i = 1, \dots, 40$  variance levels is being used have been added to the end of the corresponding profiles.



### 3.5.3 Discussion

The purpose of this chapter is to equip the reader with sufficient knowledge to make extensive use of the packages `doptim` and `randon`.

In the first part of this chapter, the basic capabilities of the `doptim` package are covered. In particular, it is worth noting that because the user is able to easily construct objective functions, the package provides a great deal of flexibility: for instance, it allows the user to expand or restrict the design space, customise the optimisation method and create composite objective functions.

Building on this, note that once an appropriate FIM approximation is accessible, then any criterion which relies on such an approximation may be fairly easily implemented. For instance, *A*-optimality which minimises the trace of the FIM or *E*-optimality which maximises the minimum eigenvalue of the FIM. Thus, while these criteria are not utilised in this thesis, extending the usage of `doptim` to cover them is straightforward.

In the second part of this chapter, the `randon` package is demonstrated and the reader is shown how to generate sensitivity bands, design graphs and efficiency profiles for a desired range of variance levels for a chosen variance component. This is achieved by guiding the reader through reproducing some of the plots used in Chapter 2. The advantage of having access to the objective functions used to generate the design graphs is also illustrated in a simple example.

With the material presented in these two sections, the reader is empowered to employ the workflow which we suggest and demonstrate in Chapter 2.

# Conclusion

The purpose of this work is to establish the foundation for a workflow involving  $D$ -optimal design when NLME models are used. In the preceding chapters, this workflow is outlined, implemented and applied and the relevant tools, which have been made available in the form of R-packages, are described in detail.

Chapter 1 deals primarily with the problem of choosing an approximation to the FIM that makes the computation of large numbers of  $D$ -optimal designs feasible while trying to assess the cost incurred from choosing an inexpensive approach, in terms of information lost. Happily, it is found that relative to the lnFO approximation, the FO approximation does not seem to perform noticeably worse in any of the cases examined. Along the way, it is illustrated that optimisation of the objective function is generally a non-trivial task; the lnFO approximation is shown to lead to an objective function with multiple local optima. In Chapter 2 this problem is shown to apply to the FO approximation as well, though theory suggests that working with approximate designs might alleviate it. However, Chapter 1 leaves some questions open which are not answered later on: are there other FIM approximations which do fare better while remaining computationally feasible? What is the cost of linearisation and can the cases where this cost is too high be identified?

Chapter 2 addresses the main problem; how can the advantages of  $D$ -optimal designs be delivered to the end user? The two biggest obstacles are intimately related; the practical issue that, for NLME models,  $D$ -optimal design necessitates the elicitation of prior beliefs about parameter values and the theoretical issue of addressing parameter misspecification in the elicited values. The former issue has, to this author's knowledge, only been addressed on an application - to - application basis and often it is implicitly supposed that prior data exists so that the problem at hand comes to

resemble that of sequential design. In Chapter 2, a simple translation of the variance scale is developed which makes it more feasible to elicit priors directly from practitioners. The issue of parameter misspecification has been addressed extensively in the literature, but primarily via theoretically advanced solutions that tend to either complicate the elicitation of priors, or be overly sceptical (trying to minimise the maximal loss) or both. Chapter 2 takes a different tack and proposes the use of efficiency profiles to systematically evaluate and compare the relative costs of parameter misspecification and, because of the cheap computational cost, this approach makes it possible to, as a default, extend the scenarios considered to cases where only subsets of parameters or transformations of parameters are of interest. A set of examples is treated in a fairly comprehensive manner; design graphs and efficiency profiles are drawn for a wide range of scenarios. With this, it is illustrated that the proposed workflow is indeed feasible and reveals important characteristics of the generated designs, such as their relative robustness.

Chapter 3 provides a gentle introduction to the R-packages which I developed, `doptim` and `randon`. A potential user is shown, through a natural progression of examples, how to generate  $D$ -optimality objective functions for a fairly flexible class of NLME models. The regression function must be explicitly supplied; regression functions which are specified as solutions to ODEs are within the theoretical framework utilised in this thesis but are not implemented in the packages. It is shown how `doptim` can be used to replicate results from the literature;  $D_\beta$ -optimal designs, fully  $D$ -optimal designs as well as composite  $D$ -optimal designs and unbalanced population  $D$ -optimal designs. For completeness, an example of how to generate  $D_A$ -optimal designs is also provided.

The `randon` package is meant to encapsulate an approach to optimal design which balances ease of use, interpretability and insight into robustness properties of investigated designs. Thus, Chapter 3 is meant to illustrate how easily the functions in `randon` can be applied to a given nonlinear regression function of interest to generate plots such as those found in Chapter 2. The natural progression of generating sensitivity bands to check a priori specifications of variance components, through to generating design graphs and efficiency profiles for parameter transformations and

variance components of interest is captured in a few simple examples. Further, the power given to the user by this package through making the resulting set of objective functions available, is demonstrated in an example where information loss due to moving a single design point in an optimal design is investigated simultaneously for a range of variance levels.

To summarise, we recommend that the FO approximation is used as a default when computing FIM approximations. The caveat is that the user should consider whether there are specific reasons to worry about linearisation, in particular whether the model may in fact be so nonlinear that the FO approximation is poor. A simple sanity check is provided in Chapter 1, but a more rigorous approach is lacking and is a good area for further work. Then, the `randon` package can be employed both to help form parameter guesses through generating sensitivity bands and also to check robustness of designs to misspecified variance components. Thus, a consulting statistician should be able to easily use these tools in collaboration with their client, to quickly generate insights into their problem.

In terms of further work and future applications, it was mentioned in Chapter 3 that other criteria which rely on the FIM should be simple to investigate using the approaches outlined in this thesis. The scope of Chapter 2 could also easily have been wider. There was nothing theoretically stopping us from expanding to other sizes of designs, different parameter transformations or parameter transformations which also involve variance components. As mentioned in Chapter 1 when we defined the NLME model, a more general model could be used, with multivariate responses, multiple covariates and more complex variance structures. A lot of the work here would be simple to extend to these cases.

Looking forward, many problems remain; e.g., what effect does simultaneous variation of parameters have on the design graph and what is the impact of a multivariate parameter misspecification?

While the gravity of these issues is yet to be explored and, if necessary, addressed, this work makes possible a new kind of consultation which one may argue is more closely aligned with the scientific approach to experiments in many applied settings. Therefore, issues aside, it is hoped that for settings where nonlinear mixed effects are appropriate and relevant, the

perceived gap between the theory and application of optimal experimental design has been narrowed, if only a little.

# Bibliography

- Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. NBS.
- Almquist, J., Leander, J., and Jirstrand, M. (2015). Using sensitivity equations for computing gradients of the foce and focei approximations to the population likelihood. *Journal of pharmacokinetics and pharmacodynamics*, 42(3):191–209.
- Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press.
- Azar, M. G., Dyer, E., and Kording, K. (2016). Convex relaxation regression: Black-box optimization of smooth functions by learning their convex envelopes. *arXiv preprint arXiv:1602.02191*.
- Bates, D. M. and Watts, D. G. (1988). *Nonlinear regression analysis and its applications*, volume 2. Wiley Online Library.
- Bazzoli, C., Retout, S., and Mentre, F. (2009). Fisher information matrix for nonlinear mixed effects multiple response models: Evaluation of the appropriateness of the first order linearization using a pharmacokinetic/pharmacodynamic model. *Statistics in Medicine*, 28:1940–1956.
- Bazzoli, C., Retout, S., and Mentre, F. (2010). Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Computer Methods and Programs in Biomedicine*, 98:55–65.
- Bornkamp, B. (2012). Functional uniform priors for nonlinear modelling.
- Davidian, M. and Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. 8:387–419.

- Demidenko, E. (1997). Asymptotic properties of nonlinear mixed-effects models.
- Dokoumetzidis, A. (2006). Numerical investigation of the number of design points in bayesian d-optimal designs in respect to prior uncertainty.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*.
- Fedorov, V. and Leonov, S. (2014). *Optimal Design for Nonlinear Response Models*. CRC Press.
- Khuri, A. (1984). A note on d-optimal designs for partially nonlinear regression models. *Technometrics*, 26(1):59–61.
- Kitsos, C. and Kolovos, K. (2013). A compilation of the D-optimal designs in chemical kinetics. *Chemical Engineering Communications*, 200:185–204.
- Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46:673–687.
- Linfield, G. R. and Penny, J. E. T. (1989). *Microcomputers in Numerical Analysis*.
- McGree, J. M., Eccleston, J. A., and Duffull, S. B. (2009). Simultaneous versus sequential optimal design for pharmacokinetic-pharmacodynamic models with FO and FOCE considerations. *Journal of Pharmacokinetics and Pharmacodynamics*, 36:101–123.
- Mentre, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika*, 84:429–442.
- Merlé, Y. and Tod, M. (2001). Impact of pharmacokinetic-pharmacodynamic model linearization on the accuracy of population information matrix and optimal design. *Journal of pharmacokinetics and pharmacodynamics*, 28(4):363–388.
- Mortensen, S. B., Jónsdóttir, A. H., Klim, S., and & Madsen, H. (2008). Introduction to pk/pd modelling - with focus on pk and stochastic differential equations. Technical report, Technical University of Denmark, DTU Informatics.

- Mullen, K., Ardia, D., Gil, D., Windover, D., and Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26.
- Nguyen, T. T. and Mentre, F. (2014). Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature. *Computational Statistics and Data Analysis*, 80:57–69.
- Nyberg, J. (2011). Practical optimal experimental design in drug development and drug treatment using nonlinear mixed effects models.
- Nyberg, J., Bazzoli, C., Ogungbenro, K., Aliev, A., Leonov, S., Duffull, S., Hooker, A. C., and Mentré, F. (2015). Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. 79:6–17.
- Nyberg, J. and Hooker, A. C. (2012). The robustness of global optimal designs. *PAGE 2012, poster*.
- Nyberg, J., Ueckert, S., Strömberg, E. A., Hennig, S., Karlsson, M. O., and Hooker, A. C. (2012). PopED: An extended, parallelized, nonlinear mixed effects models optimal design tool. *Computer Methods and Programs in Biomedicine*, 108:789–805.
- O'Brien, T. E. (2005). Designing for parameter subsets in Gaussian nonlinear regression models. *Journal of Data Science*, 3:179–197.
- Petersson, K. J. F., Hanze, E., Savic, R. M., and Karlsson, M. O. (2009). Semiparametric distributions with estimated shape parameters. 26:2174–2185.
- Piegorsch, W. W. and Bailer, A. J. (2005). *Analyzing environmental data*. John Wiley & Sons.
- Pinheiro, J. and Bates, D. (1995). Approximations to the Loglikelihood Function in the Nonlinear Mixed Effects Model. *Journal of Computational Graphics and Statistics*, 4:12–35.
- Pinheiro, J. and Bates, D. (2000). *Mixed-Effects Models in S and S-plus*. Springer.



- Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R., and Rao, C. R. (1973). *Linear statistical inference and its applications*, volume 2. Wiley New York.
- Retout, S., D. S. and Mentre, F. (2001). Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Computer Methods and Programs in Biomedicine*, 65:141–151.
- Retout, S. and Mentre, F. (2003). Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, 13:209–227.
- Salahudeen, M. S. and Nishtala, P. S. (2017). An overview of pharmacodynamic modelling, ligand-binding approach and its application in clinical practice. *Saudi pharmaceutical journal*, 25(2):165–175.
- Schmelter, T. (2007). *Experimental Design for Mixed Models with Application to Population Pharmacokinetic Studies*. PhD thesis, Otto-von-Guericke-Universität Magdeburg.
- Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear regression*. Wiley series in probability and mathematical statistics. Wiley, New York, NY [u.a.]. Literaturverz. S. 711- 743.
- Sheiner, L. B., Rosenberg, B., and Melmon, K. L. (1972). Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research*, 5:441–459.
- Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall.
- Sjogren, E., Nyberg, J., Magnusson, M. O., Lennernas, H., Hooker, A., and Bredberg, U. (2011). Optimal experimental design for assessment of enzyme kinetics in a drug discovery screening environment. *Drug Metabolism and Disposition*, 39:858–863.
- Strömberg, E. A., Nyberg, J., and Hooker, A. C. (2016). The effect of fisher information matrix approximation methods in population optimal design calculations. *Journal of pharmacokinetics and pharmacodynamics*, 43(6):609–619.

- Tvermosegaard, K. B. (2018a). doptim: An r-package for d-optimal design for nlme models, doi: 10.5281/zenodo.3545819.
- Tvermosegaard, K. B. (2018b). randon: An r-package to generate design graphs and efficiency profiles for d-optimal design in nlme models, based on the doptim package, doi: 10.5281/zenodo.1154963.
- Ueckert, S., Riviere, M., and Mentre, F. (2015). Evaluation of the fisher information matrix in non-linear mixed effects models witwith linearization. In *PODE*.
- Wand, M. P. (2007). Fisher information for generalised linear mixed models. 98:1412–1416.
- Wang, Y., Eskridge, K. M., and Nadarajah, S. (2012). Optimal design of mixed-effects PK/PD models based on differential equations. *Journal of Biopharmaceutical Statistics*, 22:180–205.