Cover Page

Quality and Quantity

Submission date: 7th April 2015

Title: Entropy Balancing: A maximum-entropy reweighting scheme to adjust for coverage error

Authors:
Samantha Watson,
Department of Global Health & Development, London School of Hygiene & Tropical Medicine
Email: Samantha.watson@lshtm.ac.uk
Telephone: 00 44 (0) 7903020860

Mark Elliot, Centre for Census & Survey Research, University of Manchester, UK
Email: mark.elliot@.manchester.ac.uk
Telephone: 00 44 (0)161 275-4257

Author Biographies

Samantha Watson is a Lecturer in the Department of Global Health & Development at the London School of Hygiene & Tropical Medicine. Her research interests centre on method in social inquiry, philosophy of science, and theories of social change. She conducts "mixed methods" research on development policy, labour, and social movements in the Global South.

Mark Elliot is a Senior Lecturer at the University of Manchester's Cathy Marsh Institute for Social Research. He is currently the Postgraduate Director of the School of Social Sciences. He has published extensively - and has managed numerous research projects in the fields of statistical confidentiality, privacy, and disclosure, and attitudes research.

Entropy Balancing: A maximum-entropy reweighting scheme to adjust for coverage error

Abstract: This paper presents a newly available technique to adjust for bias in non-probabilistically selected samples. To date, applications of this innovative technique – termed *entropy balancing* – have been restricted to evaluation settings, where the goal is to reduce model dependence prior to the estimation of treatment effects. In a novel application, we demonstrate the technique's utility in cases where the goal is to correct for sample bias originating in coverage error.

The appeal of entropy balancing in this latter setting lies in its capacity to optimise the twin goals of improved balance in covariate distribution and maximum retention of information. Entropy balancing combines the opportunity to incorporate a large set of moment conditions in the calculation of weights, with the ability to directly implement exact balance. The technique thus builds upon the theoretical appeal of the more widely known and applied *propensity score adjustment* method, while addressing that method's practical limitations.

We demonstrate the utility of the entropy balancing technique empirically, through an example using the *Young Lives Project* survey data for rural Andhra Pradesh, South India. We conclude by summarising the potential of this procedure to contribute to robust survey-based research more widely.

Key words: *causal inference, coverage error, design based inference, entropy balance, propensity score adjustment*

1. Introduction

In many fields of survey analysis, generalisation from sample to population rests on design based inference, the plausibility of which depends on the adoption of randomisation procedures in sample selection and the application of survey weights to produce estimates that are unbiased, or at least "approximately unbiased" (Kalton 2002: 129). Under this mode of inference, survey procedures are ideally designed and implemented to permit generalisation of findings beyond the surveyed sample of respondents "n", to a defined population of interest "N". Various sources of error can undermine this ideal and so compromise the external validity of findings. Here our interest is in a newly available technique to compensate for bias originating in coverage error, which occurs when a portion of the target population is excluded from the sampling frame or a sampling frame is unavailable or unutilised.

The advent of internet-based surveys has led to an increase in methodological work to adjust for coverage error (see for example Schonlau *et al*. 2009; Steinmetz and Tijdens 2009, Steinmetz *et al.* 2014). Techniques developed in this setting have applicability to any sample design that employs non-random methods to select respondents, where selection bias may be an issue (Stuart *et al.* 2011). They have particular relevance for research utilising data for the Global South, where out-dated, inadequate, or absent sampling frames tend to be more commonly encountered (UN 2006; Wilson *et al* 2006). One increasingly prominent technique to adjust for bias in such settings involves the extension of *propensity score adjustment* (PSA) methods beyond their traditional evaluative application to the separate but related field of survey sample weighting. The remainder of this section outlines the theoretical appeal and practical limitations of PSA methods in this context, before introducing *entropy balancing* as an alternative approach.

To date, *propensity score adjustment* (PSA) has been applied to a number of substantive studies employing survey samples in which bias originating in coverage error is present or suspected. Examples include work by Isakson and Forsman (2003) and Duffy *et al* (2005) to predict election results from non-probability sample surveys canvassing political opinions, by Yoshimura (2004) to generalise findings on consumption patterns beyond an internet survey sample, by Frölich (2007) for analysis of the UK gender wage gap, and by Stuart *et al.* (2011) to assess the generalizabilty of results from a randomised trial to evaluate the impact of an education intervention. In each of these examples, the aim is the analysis of one or more outcome(s) in the non-probability survey sample "n", given the distribution of covariates in the target population "N".

The application of PSA to adjust for coverage error relies upon the general approach developed by Rosenbaum and Rubin (1983) for the evaluation of causal effects, and shares its key assumptions. In the traditional evaluative application, propensity score pre-processing methods are employed to reweight or remove survey units to equate (or "balance") the distribution of covariates in "treatment" and "control" groups prior to the estimation of treatment effects. The propensity score, P(X), is calculated as the conditional probability e(x) of each observation i, being exposed to a "treatment" z = 1, as a function of a vector of observed covariates x. Under *strong ignorability* (also termed *conditional independence*), z is independent of x, and the propensity score is constant (Rosenbaum and Rubin, 1983). The estimation process orthogonalizes the treatment indicator to the covariate moments included in the reweighting - in theory reducing model dependence prior to treatment effect estimation (see Sekhon 2009, Abadie and Imbens 2011 for example applications of this principle). In their seminal 1983 paper Rosenbaum and Rubin formally demonstrate that, under *strong ignorability*, balance across a large number of covariates can be achieved by weighting or matching on the propensity score alone, so substituting a potentially large vector of discrete covariates with a one-dimensional probability. This ability to incorporate complete auxiliary information without incurring high dimensionality explains the appeal of propensity score methods in their traditional evaluation application (Rosenbaum and Rubin 1983, Heckman *et al* 1998, Rivers 2007; Vavreck and Rivers 2008; Hainmueller 2012).

The extension of propensity score methods to adjust for coverage error is similarly motivated by the theoretical appeal of applying a one-dimensional measure to balance a large number of covariates. In common with traditional survey weighting techniques, the expectation is that estimates for unknown characteristics of the target population (represented by the reference sample) can be improved through the introduction of auxiliary information about the target population's known characteristics. In traditional sample weighting schemes the set of auxiliary information is usually – in practice – limited to a small number of known totals, since weighting on a large number of discrete variables often leads to small or zero cell counts (Heckman *et al* 1998). This can be a shortcoming in cases where the plausibility of *strong ignorability* demands the inclusion of many confounders. Where propensity score adjustment differs from traditional weighting schemes is in its ability to incorporate complete auxiliary information without incurring high dimensionality.

The application of PSA to the sample-reweighting setting relies on the availability of probability-sampled survey data representative of the target population and gathered reasonably contemporaneously with the non-random sample data (Lee 2006). This, the *reference sample*, provides the benchmark covariate distributions for the non-random sample. Typically the reference sample includes extensive auxiliary data for the population of interest and variables related to participation in / selection into the non-random sample, but lacks the key variable(s) of interest included in the non-random sample. The two samples are merged to form a single dataset. In applied settings, propensity scores are typically calculated via a logistic or probit regression model of selection into the non-random sample based on a set of observable characteristics common to both datasets. The inverse of an observation's probability of selection is then applied as a weight to adjust for differences in the sample distributions to reduce bias on observed (and associated unobserved) characteristics. Ideally, the application of the PSA weights balances the covariate distributions in the non-probability sample to match those of the reference sample. However, while theoretically appealing, attempts to utilise PSA to adjust for coverage error have tended to produce disappointing results (see for example: Isakson and Forsman 2003; Yoshimura 2004; Duffy *et al*. 2005; Stuart *et al.* 2011).

In practice, PSA's theoretical promise has proven difficult to translate to applied settings. While certain limitations are acknowledged in the literature, others have tended to be overlooked. The estimation of propensity scores may itself be a source of difficulty. Where the propensity score is estimated non-parametrically the problem of high dimensionality is in fact incurred. Where estimated parametrically, the sensitivity of the estimated treatment effects to the specifications of the propensity score becomes an issue. A further issue is that practical applications of the PSA method have tended to depart from classical sample weighting techniques in their use of representative survey data in place of census-based known population totals. This departure is commonly motivated by an absence of key auxiliary data, or of variables thought indicative of selection into the non-random sample, in available census data. While unavoidable in many applications, the use of weighted representative survey data necessarily increases the variance associated with the estimates. This risks undermining any gains from improved balance and associated bias-reduction, a point we return to below. As Zhao (2005) points out the impact on variance, and the sensitivity of results to estimation procedures, are rarely acknowledged in the applied literature on PSA weighting.

As its applications have extended, another practical limitation of PSA have also gained attention. Whereas traditional weighting methods directly adjust sampling weights to exactly reproduce known population totals, PSA involves the researcher in a time consuming back-and-forwards process of propensity score estimation, matching, and balance checking in an attempt to identify the algorithm that results in the most balanced covariate distribution. This rarely succeeds in simultaneously balancing all of the covariates, with improved balance on one covariate often only achieved at the expense of another (Ho *et al*. 2007; Stuart *et al.* 2011; Hainmueller 2012).

Entropy balancing provides a means to overcome the practical limitations of PSA, so realising the latter's (theoretical) ability to incorporate complete auxiliary information without incurring high dimensionality. In the remainder of the article, we outline the advantages of the entropy balancing approach to sample weighting relative to PSA, and demonstrate its utility in a worked example.

2. Entropy balance reweighting as an alternative to propensity score adjustment

Hainmueller and Xu (2012, 2013) describe entropy balancing as a generalization of the propensity score adjustment approach, though in practice the procedures are the inverse of one another (Heinmueller and Xu 2013). Whereas propensity scores are typically calculated via a series of logistic or probit regressions, entropy balancing, in common with many traditional survey weighting schemes, directly calculates weights to adjust for known sample distributions, so integrating covariate balance directly into the weights and avoiding the tedious back-and-forwards algorithm-checking required by PSA. Although Heinmueller's interest is in the "evaluation problem", he acknowledges, without elaborating further that it is possible that a modified entropy balancing procedure could be used to reweight a single sample to some known features of the target population (Heinmueller 2012) and this is what we investigate in this paper.

The entropy balance method employs a maximum-entropy reweighting scheme whereby covariate balance is directly incorporated into the weight function used to adjust the non-random samples units in line with the reference sample. A condensed version of the theoretical framework is presented in appendix one[1]. Weights are estimated directly from pre-specified balance constraints as a log-linear function of the known target sample moment conditions. This entails that the adjusted sample moments of the non-random sample exactly match the corresponding moments in the reference sample. The entropy balancing solution seeks a set of scalar unit weights that simultaneously satisfy the pre-assigned balance constraints while minimising the distance from uniform (or sample-design based) base weights, so retaining maximum information. The distance between the distribution of the estimated balance weights and the distribution of the pre-set uniform base weights is measured by the loss function (employing a directed entropy divergence distance metric). The loss function is non-negative and decreases as the estimated balance weights approach the base weights. The procedure accommodates high dimensionality to assign one weight to each control unit by reducing the balancing scheme to a sequence of non-linear equations in R Lagrange multipliers. The solution therefore allows that, by adjusting the unit weights in line with known sample moments, exact matching is obtained for finite samples.

The method's principal advantage over the logistic / probit algorithms typically used to calculate propensity scores is its ability to directly implement exact balance. By calculating weights to be as similar as possible to base weights, the entropy balance weighting procedure optimising the twin goals of improved balance in covariate distribution and maximum retention of information (the latter is enhanced by the entropy

[1] See Hainmueller (2012) for a comprehensive presentation of the theoretical framework.

approach's ability to vary weights smoothly across units). A further advantage of the entropy balancing method is that it allows for survey weights pre-assigned to the reference sample to be easily incorporated into the calculation of all moment conditions for reweighting.

Where traditional survey weighting techniques usually necessarily limit the size of the vector of auxiliary information to evade the "curse of dimensionality" (Heckman *et al* 1998), the entropy balance reweighting procedure (potentially) permits all available data from the reference sample to be incorporated (including higher moments and co-moments as interaction effects), generating an inclusive vector of moment conditions. This permits the density of X in the reweighted non-probability sample to mirror very closely that of the reference sample. In contrast to the PSA method, however, entropy balancing precludes and balance decreases on the specified moments by directly adjusting weights to known sample moments. Application of the entropy balance weights to the non-probability sample results in more weight being given to under-represented groups and less weight to over-represented groups, adjusting for unequal probability of sample selection and creating a 'pseudo-population' with characteristics in line with the reference sample[2].

The remainder of the article illustrates the application of the entropy balancing procedure to reweight a non-probabilistically sampled survey to a reference sample representative of its target population, before evaluating the methods effectiveness in adjusting for coverage bias relative to the better-known PSA technique[3].

3. Example problem

We draw on two independent sample datasets. The Young Lives Project sample (n_0) is a non-probabilistically sampled survey separately undertaken in Ethiopia, Peru, Vietnam, and Andhra Pradesh (AP), South India in four planned rounds of data collection[4]. The dataset that we are using is the second round for rural AP - collected in 2005/6. The dataset includes information for 2,196 households and 14,110 individuals[5]. The data are primarily intended to provide a means to study the changing dynamics of childhood and household wellbeing. The population of interest is families with young children. The YLP provides a rich source of data on household demographics and individual characteristics, assets, market and non-market labour activities, and attitudes. The substantive purpose of our own study was to use this dataset to analyse the relationships among rural women's participation in poverty amelioration schemes, gender norms, and labour profiles at the individual, household, and community levels.

The YLP survey's sampling procedure is described in Wilson *et al* (2006). The use of non-probabilistic sampling was prompted by the absence of "effective, accessible and accurate sampling frames of households with qualifying children in [the] study countries" (Wilson *et al.* 2006: 356). Consequently, the study adopted a *sentinel site surveillance system* (Galeb *et al* 2003; Kumra 2008)[6]. In AP, 20 study sites (5 urban and 15 rural), each an administrative zone, or *mandal*, were selected across the State's three agro-climactic regions. Here we limit analysis to the 15 rural sites. Sites were selected on the basis of relative wealth, in line with the study's aim to oversample income poor households, while enabling comparisons to be made between poor and non-poor (Wilson *et al.* 2006).

---

[2] In cases where only marginal population probabilities are available (from summarised census data for example) the ebalance procedure allows for values to be manually specified to reweight the non-probability sample covariates in line with available known population targets.

[3] All analysis is conducted in STATA 13 software; Hainmueller's "ebalance" suite of commands to perform the entropy balance procedure can be imported to STATA in the usual manner, i.e. "ssc install ebalance, all replace".

[4] The survey was sponsored by the UK *Department for International Development* (DFID), and is led by the *Oxford Department of International Development* at the *University of Oxford*, in collaboration with academic institutions in each of the four project countries.

[5] In the second round of data collection all individuals resident in a selected household were included in the survey.

[6] Andersson (1996) discusses the general method of sentinel site sampling in some detail.

The reference sample is drawn from the *All India National Sample Survey* (n_1). The NSS is a weighted, probabilistically sampled survey representative of the national population. The survey is designed and collected by the *National Sample Survey Organisation* (NSSO), a department of the *Ministry of Statistics and Programme Implementation* (MSPI). The NSS has been conducted annually since 1950. Here we utilise the employment and unemployment schedule (schedule ten) as, importantly for our purposes, it contains information relevant to the selection mechanisms informing inclusion in the YLP sample. Schedule 10 is incorporated quinquennially. We use round 61 of the NSS. Data for this survey year was collected in 2004 - 2005, overlapping with data collection for round two of the YLP. The NSS dataset includes information for 5,550 households and 22,591 individuals in rural Andhra Pradesh[7]. The survey employs a probabilistic stratified, multi-stage sample design. Briefly, the NSS stratifies by geographic region, urban-rural area, population density, and household affluence; with each stratum designed to be non-overlapping and proportional (based on projected population figures from the 2001 national census taking into account decadal growth rated between 1991 and 2001) (MSPI 2006: 82). Full details of the sampling methodology can be found in the NSSO's documentation for the 61st round (NSSO 2004). The NSSO, in line with the practice of most nationally representative sample survey organisations uses adjustment weights at the household level based on extrapolations of the 2001 census to account for unequal sampling rates in the strata. Samples are selected from each stratum independently. Unequal sampling rates in the strata are corrected for (in order to produce an unbiased mean estimator). In this example, the appropriate sampling weights are drawn from probabilities of selection (MSPI 2008). The weights are uniform within households, with all individuals resident in a household included in the survey.

4. The application of entropy balancing

As a first step, we define a subpopulation comparable with the YLP's target population within the NSS sample to include only households in AP with children in the target population age range. Next, covariates common to both datasets are identified and operationalised. Table one presents the covariates common to the two datasets and their values across the two datasets. The entropy balancing scheme permits the inclusion of both continuous and categorical data, taking advantage of all available common information.

>>>>>Table 1: Sample characteristics prior to entropy balance procedure**<<<<<**

Table one demonstrates that the densities of characteristics recorded in the YLP sample deviate substantially from those of the target population. The YLP sample has selected a roughly even number of households from each of the State's three agro-climactic regions, with households in Rayalaseema oversampled relative to those in Coastal Andhra and Telangana. "Forward" caste households are significantly under-represented, likely as a result of the oversampling of poor households, Adivasi households are significantly over-represented. Over-sampling is practiced inconsistently, however, with religious minorities substantially under-sampled. Casual daily wage labour households are very under-represented, while marginal and mid-size farming households are over-represented[8]. Households are generally larger in the YLP sample than the target population. Heads of household are younger, disproportionately male, and more literate in the YLP sample than the target population.

In order to apply the entropy reweighting scheme, a single indicator variable is generated in both the reference and non-probability datasets, coded 1 for all observations in the NSS and 0 for all those in the YLP. The two datasets are then merged to form a single dataset, necessary for the calculation of entropy balance weights across higher moments. As detailed above, a set of balance constraints can now be specified for each of the covariates, equating the moments of the covariate distribution between the reference and target samples. Recall that possible moment constraints include the mean (the first moment), variance (the second moment), and skewness (the third moment). In the case of binary variables (for example, gender of household head) adjustment of the first moment is, in practice, sufficient to match

[7] At the all India level a total of 124,680 households and 602,833 individuals took part in the survey for schedule 10 of the 61st round of the NSS.

[8] Household class is calculated on the basis of household landholding and dominant labour relations.

higher moments. Moment constraints may be separately defined for each covariate. The specification of interaction terms allows covariates to be balanced across key subsample groups (in this case, we balance across the subsample caste).

The “ebalance” algorithm computes the values of the specified moments in the reference sample (n_1), in this case the NSS, and seeks a set of entropy weights that can adjust the YLP sample to match. Convergence occurs once all the specified moments are matched across the data sources, within the specified number of iterations and tolerance level[9]. Though rare in practice, the inclusion of too many collinear moment constraints may in theory prevent convergence (Hainmueller 2012). Specifying fewer moment constraints, either via the removal of implicated covariates or a reduction in their specified moment constraints (mean, variance, skewness), can remedy this. Alternatively (or additionally), the tolerance level can be relaxed. Tables 2a and 2b present the results of the entropy balance procedure.

>>>>>>Table 2a: Variable moment conditions prior to entropy balance procedure**<<<<<<**

>>>>>>Table 2b: Variable moment conditions after entropy balance procedure**<<<<<<**

Figure one presents measures of the standardised differences in means for the two data samples before and after entropy balance reweighting.

>>>Figure 1: Covariate balance for all moment conditions before and after entropy balance reweighting**<<<**

The results demonstrate that the adjustment has a dramatic effect. The entropy balance derived weights have adjusted the YLP sample’s distribution such that it now reflects rural AP’s population densities as reported in the weighted reference sample. Following the reweighting procedure, differences between the non-probability and reference samples, across all moment conditions (mean, variance, skewness) for all matching variables are now effectively zero and are non-significant.

Figure three compares the results obtained through the entropy balance reweighting procedure with those obtained via the PSA method, demonstrating the superior results achieved with the former. The reported PSA results are the best obtained through an extensive back-and-forwards process of estimation, matching, and balance checking. The weights derived through the PSA procedure improve balance on some covariates (specifically religion, head of household literacy rates, and some categories of household class), but – in line with the tendency widely reported in the literature on applied PSA methods - this comes at the expense of balance on other covariates. Notably the PSA derived weight exacerbates the extent and / or significance of the original differences in some cases. In contrast, the entropy balance derived weights result in simultaneous balance across all of the specified covariates.

>Figure 2: Comparison of equivalent results obtained by entropy balance and PSA weighting**<**

Table three presents the effects of the ebalance weighting procedure on key outcome variables in the pre-adjusted and adjusted non-random sample data. The results demonstrate that the application of the ebalance derived weights modifies the distribution of key outcome mean estimates. There is, however, a trade-off between bias reduction and variance increase. The weighted estimates have increased standard errors and substantially increased confidence errors in comparison with the unweighted data.

>Table 3: Comparison of equivalent results obtained by entropy balance and PSA weighting**<**

5. Discussion

[9] The default iteration number is 20, the default tolerance level 0.015, and both can be increased if convergence fails.

In this article we have introduced, and applied in modified form, an innovative means to adjust for selection bias in non-probabilistically selected samples. We have demonstrated its benefits in relation to the more widely known and applied propensity score adjustment method.

The entropy balance reweighting scheme permits many of the difficulties encountered with PSA-based reweighting to be overcome, negating the need for the time consuming and often unsatisfactory iterative process of propensity score adjustment. Whereas the PSA reweighting procedure rarely succeeds in simultaneously balancing all of the covariates, entropy balancing directly calculates weights to adjust for known sample distributions, integrating covariate balance directly into the weights. The entropy balance reweighting procedure (potentially) permits all available parallel data from a reference sample to be incorporated in calculating the non-probability sample's population weights. This enables the density of X in the reweighted non-probability sample to be made to mirror that of the reference sample very closely. The ability to include a large set of moment conditions results in a covariate density for the reweighted sample consistent with the population of interest (as defined by the reference sample). However, the extent of the trade-off exacted between bias reduction and variance increase remains an important consideration. By incorporating design weights for the reference sample we introduce a source of variance that, though shared by the PSA approach, is absent in traditional calibration procedures utilising census counts. We should not, however, discount the possibility that some of the increase in variance in fact corrects for bias present in the unweighted non-random sample. Since we are balancing on the second moment condition it may be that the increased standard errors represent less a loss of precision than a correction for inaccurate estimates of precision in the original data. Further research is needed to assess the trade-off exacted between bias reduction and variance increase in different settings.

As with all reweighting schemes, the effectiveness of the process will depend ultimately on the quality and applicability of the reference sample. Similarly, the entropy balance scheme can only correct for bias resulting from unobserved confounders to the extent that they are associated with the recorded balance constraints. The extent (and degree) of covariate equivalence across the reference and non-probability samples needs to be assessed on a case by case basis, and a sufficient number of units must be available in each to permit adequate overlap in the covariate distributions. Whilst it is possible to increase the iteration number and tolerance level in the pursuit of convergence, it is important the balance constraints are realistic and consistent. Bearing in mind these caveats, the example application demonstrates that remarkable results can be obtained through the entropy balance reweighting scheme. Following the reweighting procedure, differences between the non-probability and reference samples, across all moment conditions for all matching variables are reduced to effectively zero. It is anticipated that similar results can be obtained for any sample dataset where coverage error is known or suspected and an appropriate reference sample is available.

References

Abadie, Alberto and Guido W. Imbens. 2011. "Bias-corrected matching estimators for average treatment effects." *Journal of Business & Economic Statistics* (29)1: 1-11.

Andersson, N. 1996. *Evidence-based planning: The philosophy and methods of sentinel community surveillance.* Washington DC: Economic Development Institute of the World Bank.

Duffy, Bobby., George Terhanian, John Bremer, and Kate Smith. 2005. "Comparing data from online and face-to-face surveys." *International Journal of Market Research* 47(6): 615-39.

Frölich, Markus. 2007. "Propensity score matching without conditional independence assumption—with an application to the gender wage gap in the United Kingdom." *The Econometrics Journal* 10(2): 359-407.

Galab, S., M Gopinath Reddy, Piush Antony, Anne McCoy, C Ravi, D Sreerama Raju, K Mayuri, and P Prudhivikar Reddy. 2003. *Young Lives Preliminary Country Report: India*. Oxford, Young Lives. Retrieved October 12 2013 (http://www.younglives.org.uk/files/country-reports/country-report-1-india-2003).

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1): 25–46.

Hainmueller, Jens and Yiqing Xu. 2013. "Ebalance: A Stata Package for Entropy Balancing" *Journal of Statistical Software* 54(7). Retrieved March 3 2014 (http://www.jstatsoft.org/v54/i07/paper).

Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1998. "Matching as an econometric evaluation estimator." *The Review of Economic Studies* 65(2): 261-294.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political analysis* 15(3): 199-236.

Kalton, Graham. 2002. "Models in the Practice of Survey Sampling (Revisited)." *Journal of Official Statistics* 18(2): 129-154.

Kumra, Neha. 2008. "An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India." *Young Lives technical note* 2. Oxford. Young Lives. Retrieved March 11 2014 (http://www.younglives.org.uk/files/technical-notes/an-assessment-of-the-young-lives-sampling-approach-in-andhra-pradesh-india).

Lee, Sunghee. 2006. "Propensity score adjustment as a weighting scheme for volunteer panel web surveys." *Journal of Official Statistics*, 22(2): 329–349.

MSPI 2006. *Annexure II - Population Projection. New Delhi*, National Sample Survey Organisation. New Delhi, National Sample Survey Organisation.

MSPI 2008. *How to Use Unit Level Data*. New Delhi, Ministry of Statistics and Programme Implementation.

NSSO 2004. *Note on Estimation Procedure of NSS 61st Round*. New Delhi, Government of India.

Rivers, D. 2007. "Sampling for Web Surveys" *Joint Statistical Meetings*, Salt Lake City, UT, Retrieved December 1 2012 (http://www.laits.utexas.edu/txp_media/html/poll/files/Rivers_matching.pdf).

Vavreck, L. and D. Rivers. 2008, "The 2006 Cooperative Congressional Election Study" *Journal of Elections* 18(4): 355-366.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70(1): 41-55.

Schonlau, Matthias, Arthur Van Soest, Arie Kapteyn, and Mick Couper. 2009. "Selection bias in web surveys and the use of propensity scores." *Sociological Methods & Research* 37(3): 291-318.

Sekhon, Jasjeet S. 2009. "Opiates for the matches: Matching methods for causal inference." *Annual Review of Political Science* 12: 487-508.

Steinmetz, S.M., and K. Tijdens 2009. "Can weighting improve the representativeness of volunteer online panels? Insights from the German Wage indicator data." *Concepts and Methods* 5(1): 7-11. Retrieved December 9 2012 (http://www.concepts-methods.org/newsletters/20091215_55_C&M_Newsletter_2009_2.pdf).

Steinmetz, S.M., A. Bianchi, K.G. Tijdens and S. Biffignandi. 2014. "Improving web survey quality - Potentials and constraints of propensity score weighting". in *Online Panel research: A Data Quality Perspective*, edited by M. Callegaro, R. Baker, J. Bethlehem, A. Göritz, J. Krosnick and P. Lavrakas. New York. Wiley.

Stuart, E., S. R. Cole, Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf. 2011. "The use of propensity scores to assess the generalizability of results from randomized trials." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174(2): 369-386.

UN. 2006. *Household Sample Surveys in Developing and Transition Countries*, United Nations, New York. Retrieved March 2 2014 (http://unstats.un.org/unsd/hhsurveys/).

Wilson, Ian, Sharon RA Huttly, and Bridget Fenn. 2006. "A case study of sample design for longitudinal research: Young Lives." *International Journal of Social Research Methodology* 9(5): 351-365.

Yoshimura, Osamu. 2004. "Adjusting Responses in a Non-probability Web Panel Survey by the Propensity Score Weighting". *Proceedings of the American Statistical Association.* Retrieved March 11 2014 (http://www.amstat.org/sections/srms/Proceedings/y2004f.html).

Zhao, Z. (2005) "Sensitivity of Propensity Score Methods to the Specifications" IZA Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labour) Discussion Paper No. 1873, December 2005, Bonn (Available from: *ftp.iza.org/dp1873.pdf*)

Appendix 1: A condensed version of the theoretical framework for entropy balancing

Under ebalance, weights are selected to minimize the entropy distance metric:

$$\min_{w_i} H(w) = \sum_{\{i|D=0\}} w_i \log(w_i / q_i) \qquad \text{equation 1}$$

Where w_i is the weight selected for each non-random sample unit.

$D_i \in \{1,0\}$ is a binary indicator coded 1 if unit i is drawn from the reference sample or 0 if it is drawn from the non-random sample. $q_i = 1/n_0$ and is a base weight

The selection of weights is subject to the balance constraint defined in equation 2.1, the normalising constraint defined in equation 2.2, and the non-negativity constraint defined in equation 2.3:

$$\sum_{\{i|D=0\}} w_i c_{ri}(X_i) = m_r \quad \text{with } r \in 1,\ldots,R \qquad \text{equation 2.1}$$

$$\sum_{\{i|D=0\}} w_i = 1 \qquad \text{equation 2.2}$$

$$w_i \geq 0 \text{ for all } i \text{ such that } D = 0 \qquad \text{equation 2.3}$$

X is a matrix that contains the data of J exogenous pre-treatment covariates with X_{ij} denoting the value of the j-th covariate characteristic for unit i.
$c_{ri}(X_i) = m_r$ describes a set of R balance constraints imposed on the covariate moments of the reweighted non-random sample group.

The ebalance approach accommodates high dimensionality to assign one weight to each control unit. The weights that solve the entropy balancing scheme are computed from a dual problem that is unconstrained and reduced to a system of non-linear equations in R Lagrange multipliers. The dual problem is given by:

$$\min_{Z} L^{d} = \log(Q' \exp(-C' Z)) + M' Z \qquad \text{equation 3}$$

Where $Z = \{\lambda_1, \ldots, \lambda_R\}'$ is a vector (Z^*) of Lagrange multipliers for the balance constraints, rewritten in matrix form as CW = M with the $(R \times n_0)$ constraint matrix, $C = [c_1(X_i), \ldots, c_R(X_i]'$, and the moment vector, $M = [m_1, \ldots, m_R]'$. The vector Z^* that solves the dual problem also solves the primal problem. The solution weights are recovered using:

$$W^* = \frac{Q \cdot \exp(-C' Z^*)}{Q' \exp(-C' Z) \cdot} \qquad \text{equation 4}$$

An iterative Levenberg-Marquardt scheme exploits second order information to solve the dual problem:

$$Z^{new} = Z^{old} - l \nabla^2_Z (L^d)^{-1} \nabla_Z (L^d) \qquad \text{equation 5}$$

Here, l is a scalar denoting the step length. The optimal step length (either the full Newton step or l) is selected for each iteration.

(Heinmueller and Xu 2013)

Table 1: Sample characteristics prior to entropy balance procedure

	NSS (reference sample, n_1)				YLP (non-random sample, n_0)				Difference	
	mean	S.E	[95% CI]		mean	S.E	[95% CI]		value	p
Household head gender (women)	9.175	0.009	7.313	11.037	7.016	0.005	5.947	8.085	-2.159	0.000
Household head age	38.079	0.424	37.247	38.911	40.123	0.245	39.643	40.603	2.044	0.000
Household head literate	41.831	0.014	39.005	44.656	50.410	0.011	48.318	52.502	8.579	0.000
Household size (adjusted)	4.932	0.007	4.919	4.945	6.425	0.058	6.313	6.538	1.493	0.000
“Forward” castes	21.471	0.012	19.163	23.779	14.299	0.007	12.834	15.763	-7.173	0.000
Dalit	20.159	0.012	17.846	22.473	21.220	0.009	19.510	22.931	1.061	0.232
Adivasi	12.440	0.011	10.281	14.599	15.073	0.008	13.576	16.570	2.633	0.001
"Other backward" castes (base)	45.927	0.002	45.585	46.268	49.408	0.011	47.316	51.500	3.481	0.001
Muslim	6.207	0.006	4.945	7.470	2.368	0.003	1.732	3.004	-3.840	0.000
Christian	2.047	0.004	1.300	2.794	0.865	0.002	0.478	1.253	-1.182	0.000
Hindu (base)	91.741	0.001	91.552	91.929	96.767	0.004	96.027	97.507	5.026	0.000
Household class: non-farm PCP, service, trade	13.425	0.009	11.693	15.156	14.390	0.007	12.921	15.858	0.965	0.204
Household class: marginal farming	3.935	0.006	2.733	5.137	18.670	0.008	17.040	20.301	14.736	0.000
Household class: small-scale farming	6.582	0.007	5.168	7.996	7.969	0.006	6.836	9.102	1.387	0.017
Household class: mid-size farming	13.836	0.011	11.738	15.934	23.087	0.009	21.325	24.850	9.252	0.000
Household class: capitalist farming	3.738	0.005	2.769	4.707	4.508	0.004	3.640	5.376	0.770	0.085
Household class: regular salaried employment	5.408	0.006	4.230	6.586	6.421	0.005	5.395	7.446	1.013	0.058
Household class: casual daily wage labour (base)	53.078	0.002	52.736	53.419	24.954	0.009	23.144	26.765	-28.123	0.000
Household landholding (acres)	2.076	0.091	1.897	2.254	2.196	0.080	2.039	2.352	0.120	0.140
Household landholding (log acres)	-0.484	0.049	-0.579	-0.388	-0.190	0.033	-0.255	-0.125	0.293	0.000
Region: Coastal	42.968	0.014	40.138	45.798	34.335	0.010	32.349	36.322	-8.633	0.000
Region: Rayalaseema	17.877	0.011	15.788	19.966	32.423	0.010	30.464	34.381	14.546	0.000
Region: Telengana (base)	39.149	0.002	38.815	39.483	33.242	0.010	31.272	35.213	-5.907	0.000

Source: Data Sources: All India National Sample Survey 2004 / 2005: round 55 / schedule 10: Employment & Unemployment & Young Lives Project; round two (2005 / 2006) n = 4172 (NSS n = 1,976) (YLP n = 2,196). Satterthwaite t-tests are applied to calculate the p value of the equality of means in the two samples (recommended when the population variances cannot be assumed to be equal)

Table 2a: Variable moment conditions prior to entropy balance procedure

	NSS (reference sample, n_1)			YLP (non-random sample, n_0)			Difference			
	mean	**variance**	**skewness**	**mean**	**variance**	**skewness**	**mean**	**variance**	**skewness**	**P**
Household head gender (women)	9.175	0.083	2.828	7.016	0.065	3.366	-2.159	-0.018	0.538	0.000
Household head age	38.080	178.200	0.598	40.120	131.600	1.206	2.040	-46.600	0.608	0.000
Household head literate	41.830	0.243	0.331	50.390	0.250	-0.015	8.560	0.007	-0.347	0.000
Household size (adjusted)	4.932	3.699	1.720	6.425	7.266	2.021	1.493	3.567	0.301	0.000
"Forward" castes	21.470	0.169	1.390	14.310	0.123	2.039	-7.160	-0.046	0.649	0.000
Dalit	20.160	0.161	1.488	21.230	0.167	1.407	1.070	0.006	-0.081	0.232
Adivasi	12.440	0.109	2.276	15.080	0.128	1.952	2.640	0.019	-0.324	0.001
Muslim	6.207	0.058	3.630	2.369	0.023	6.264	-3.838	-0.035	2.634	0.000
Christian	2.047	0.020	6.773	0.866	0.009	10.610	-1.181	-0.011	3.837	0.000
Household class: non-farm pcp, services, trade	13.420	0.116	2.146	14.350	0.123	2.034	0.930	0.007	-0.112	0.204
Household class: marginal farming	3.935	0.038	4.739	18.680	0.152	1.607	14.745	0.114	-3.132	0.000
Household class: small-scale farming	6.582	0.062	3.502	7.973	0.073	3.103	1.391	0.012	-0.399	0.017
Household class: mid-size farming	13.840	0.119	2.095	23.100	0.178	1.277	9.260	0.058	-0.818	0.000
Household class: capitalist farming	3.738	0.036	4.878	4.510	0.043	4.384	0.772	0.007	-0.494	0.085
Household class: regular salaried employment	5.408	0.051	3.943	6.424	0.060	3.555	1.016	0.009	-0.388	0.058
Household landholding (acres)	2.076	14.540	7.269	2.197	14.010	4.885	0.121	-0.530	-2.384	0.140
Household landholding (log acres)	-0.484	2.834	0.321	-0.189	2.429	-0.024	0.294	-0.405	-0.345	0.000
Region: Coastal	42.970	0.245	0.284	34.310	0.226	0.661	-8.660	-0.020	0.377	0.000
Region: Rayalaseema	17.880	0.147	1.677	32.440	0.219	0.750	14.560	0.072	-0.927	0.000
"Forward" castes *Coastal	10.940	0.097	2.504	3.235	0.031	5.287	-7.705	-0.066	2.783	0.000
Dalit*Coastal	8.502	0.078	2.976	3.508	0.034	5.054	-4.994	-0.044	2.078	0.000
Adivasi*Coastal	4.074	0.039	4.646	10.300	0.092	2.613	6.226	0.053	-2.033	0.000
"Forward" castes *Rayalaseema	5.908	0.056	3.740	7.927	0.073	3.115	2.019	0.017	-0.625	0.001
Dalit caste*Rayalaseema	3.305	0.032	5.224	8.702	0.079	2.930	5.397	0.048	-2.294	0.000
Adivasi*Rayalaseema	0.638	0.006	12.400	0.866	0.009	10.610	0.228	0.002	-1.790	0.258
"Forward" castes *household landholding	0.696	8.765	14.360	0.512	6.793	10.650	-0.184	-1.972	-3.710	0.001
Dalit*household landholding	0.186	1.066	10.780	0.241	0.938	8.417	0.054	-0.128	-2.363	0.010
Adivasi*household landholding	0.291	1.846	6.781	0.291	1.454	7.829	-0.001	-0.392	1.048	0.959

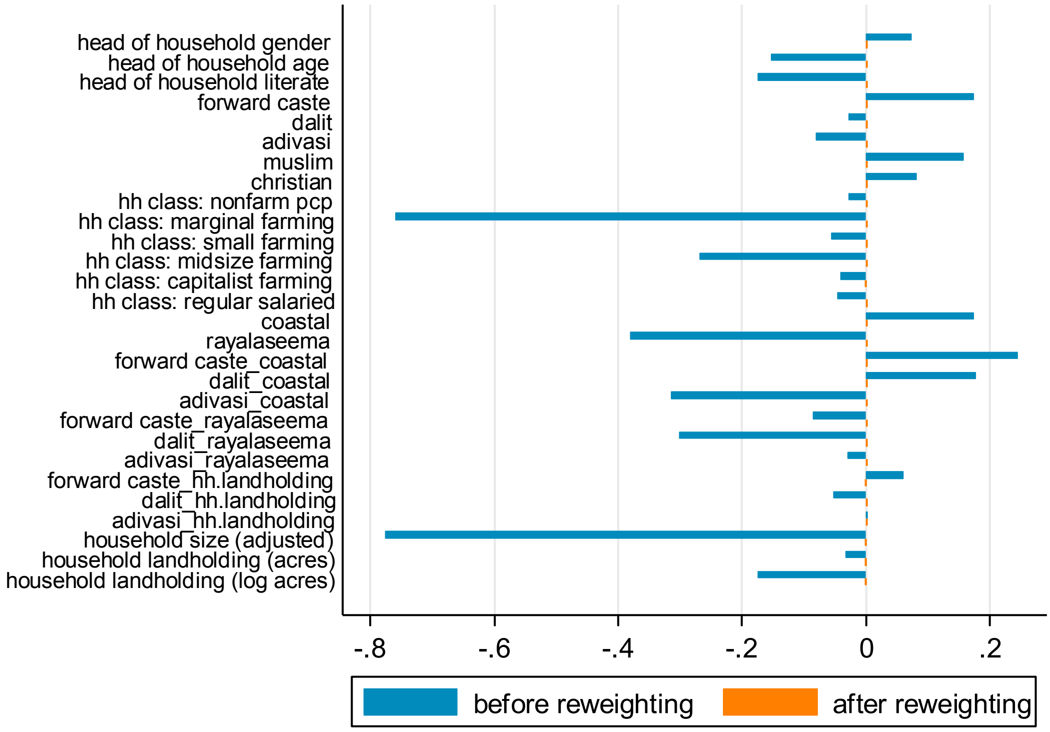
Source: Data Sources: All India National Sample Survey 2004 / 2005: round 55 / schedule 10: Employment & Unemployment & Young Lives Project; round two 2005 / 2006 n = 4172 (NSS n = 1,976) (YLP n = 2,196).

Table 2b: Variable moment conditions after entropy balance procedure

	NSS (reference sample, n_1)			YLP (non-random sample, n_0)			Difference			
	mean	**variance**	**skewness**	**mean**	**variance**	**skewness**	**mean**	**variance**	**skewness**	**P**
Household head gender (women)	9.175	0.083	2.828	9.174	0.083	2.829	-0.001	0.000	0.001	1.000
Household head age	38.080	178.200	0.598	38.080	178.200	0.598	0.000	0.000	0.000	0.999
Household head literate	41.830	0.243	0.331	41.830	0.243	0.331	0.000	0.000	0.000	0.999
Household size (adjusted)	4.932	3.699	1.720	4.933	3.704	1.722	0.001	0.005	0.002	0.995
“Forward” castes	21.470	0.169	1.390	21.480	0.169	1.389	0.010	0.000	-0.001	0.998
Dalit	20.160	0.161	1.488	20.160	0.161	1.488	0.000	0.000	0.000	0.999
Adivasi	12.440	0.109	2.276	12.440	0.109	2.276	0.000	0.000	0.000	0.999
Muslim	6.207	0.058	3.630	6.209	0.058	3.629	0.002	0.000	-0.001	0.999
Christian	2.047	0.020	6.773	2.046	0.020	6.774	-0.001	0.000	0.001	1.000
Household class: non-farm pcp, services, trade	13.420	0.116	2.146	13.430	0.116	2.146	0.010	0.000	0.000	1.000
Household class: marginal farming	3.935	0.038	4.739	3.934	0.038	4.739	-0.001	0.000	0.000	0.999
Household class: small-scale farming	6.582	0.062	3.502	6.582	0.062	3.502	0.000	0.000	0.000	1.000
Household class: mid-size farming	13.840	0.119	2.095	13.840	0.119	2.095	0.000	0.000	0.000	0.999
Household class: capitalist farming	3.738	0.036	4.878	3.749	0.036	4.870	0.011	0.000	-0.008	0.991
Household class: regular salaried employment	5.408	0.051	3.943	5.406	0.051	3.944	-0.002	0.000	0.001	0.999
Household landholding (acres)	2.076	14.540	7.269	2.077	14.550	7.263	0.001	0.010	-0.006	0.995
Household landholding (log acres)	-0.484	2.834	0.321	-0.483	2.835	0.362	0.001	0.001	0.040	0.996
Region: Coastal	42.970	0.245	0.284	42.970	0.245	0.284	0.000	0.000	0.000	0.999
Region: Rayalaseema	17.880	0.147	1.677	17.880	0.147	1.677	0.000	0.000	0.000	1.000
“Forward” castes*Coastal	10.940	0.097	2.504	10.940	0.097	2.503	0.000	0.000	-0.001	0.999
Dalit*Coastal	8.502	0.078	2.976	8.501	0.078	2.976	-0.001	0.000	0.000	0.999
Adivasi*Coastal	4.074	0.039	4.646	4.074	0.039	4.647	0.000	0.000	0.001	0.999
“Forward” castes*Rayalaseema	5.908	0.056	3.740	5.910	0.056	3.739	0.002	0.000	-0.001	0.999
Dalit caste*Rayalaseema	3.305	0.032	5.224	3.304	0.032	5.225	-0.001	0.000	0.001	0.999
Adivasi*Rayalaseema	0.638	0.006	12.400	0.638	0.006	12.400	0.000	0.000	0.000	1.000
“Forward” castes*household landholding	0.696	8.765	14.360	0.697	8.772	14.340	0.001	0.007	-0.020	0.995
Dalit*household landholding	0.186	1.066	10.780	0.186	1.066	10.780	0.000	0.000	0.000	1.000
Adivasi*household landholding	0.291	1.846	6.781	0.291	1.846	6.781	0.000	0.000	0.000	1.000

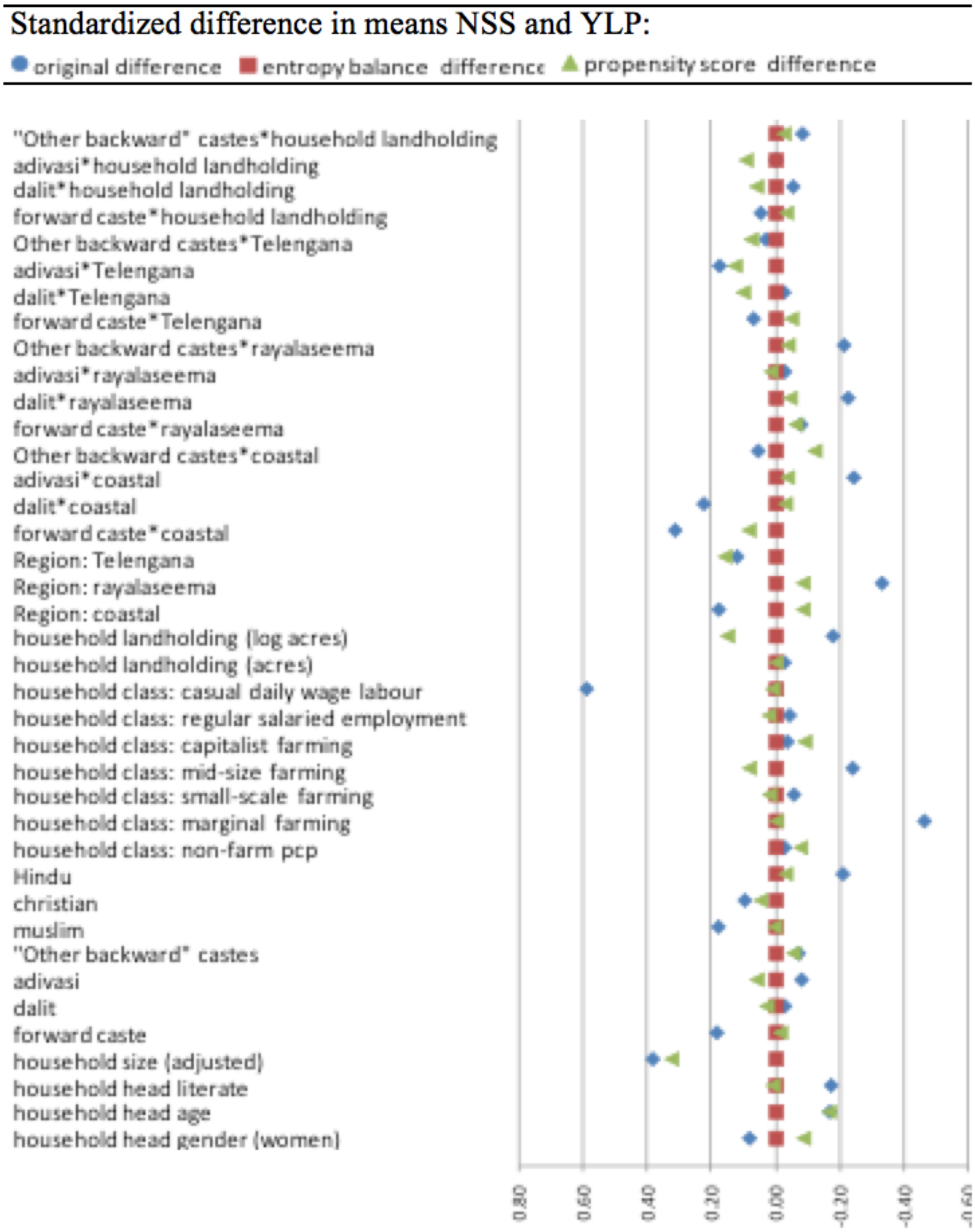
Source: Data Sources: All India National Sample Survey 2004 / 2005: round 55 / schedule 10: Employment & Unemployment & Young Lives Project; round two 2005 / 2006 n = 4172 (NSS n = 1,976) (YLP n = 2,196).

Figure 1: Covariate balance for all moment conditions before and after entropy balance reweighting



Source: Data Sources: All India National Sample Survey 2004 / 2005: round 55 / schedule 10: Employment & Unemployment & Young Lives Project; round two 2005 / 2006 n = 4172 (NSS n = 1,976) (YLP n = 2,196).

Figure 2: Comparison of equivalent results obtained by entropy balance and propensity score weighting



Source: Young Lives Project; round two 2005 / 2006 (YLP n = 2,196)

Table 3: Target estimates in the non-weighted and weighted non-random sample

Target estimate	unweighted YLP data					weighted YLP data			
	n	mean	S.E	[95% CI]		mean	S.E	[95% CI]	
% of women (over 15) participating in rural self help groups	4,242	28.41	2.09	24.03	32.78	38.91	4.22	30.07	47.76
% of hhs with children (under 15's) undertaking market labour	2,196	23.63	1.54	20.42	26.85	17.04	2.21	12.41	21.67
% of children (12year cohort) with moderate or severe stunting†	745	34.50	1.88	30.55	38.44	28.99	4.69	19.15	38.84
% of children (4year old cohort) with moderate or severe stunting†	1,451	30.46	1.21	28.09	32.83	25.34	4.58	15.74	34.93

Source: Young Lives Project; round two 2005 / 2006 (YLP n = 2,196 households, 14,110 individuals) † "moderate or severe stunting" is defined by a height to weight z-score of below -2.