2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Doubly-robust weight smoothing models to smooth post-stratification weights in case of a Gaussian survey outcome

Promotor :
De heer Yannick VANDENDIJCK

Supervisor :
Prof.dr. Christel FAES

Adriana Rocio Reyes Sierra
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Doubly-robust weight smoothing models to smooth post-stratification weights in case of a Gaussian survey outcome

Promotor :
De heer Yannick VANDENDIJCK

Supervisor :
Prof.dr. Christel FAES

Adriana Rocio Reyes Sierra

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Abstract

In order to obtain unbiased estimates of a population quantity based on sample survey data, post-stratification techniques use external data to adjust the estimates during the analysis stage. Small sample sizes in any post-strata may yield highly variable estimator. The weight trimming method pools highly underrepresented units into a stratum with better representation but it is somehow arbitrary. In the same spirit, weight-smoothing approach treats post-stratum means as random-effects, inducing shrinkage across post-stratum means. To protect against the bias generated by possible misspecification of the mixed-model, a doubly-robust version is used as well as a nonparametric spline function for the underlying weight stratum means. I compare those approaches in a simulation study for the inference about the population mean of a normally distributed survey outcome with ordinal post-stratifying variable. None of the 9 estimators is uniformly best in all 24 scenarios considered but the nonparametric weight-smoothing doubly-robust is close to the best for a wide range of populations offering protection against unfavorable mean structures and model misspecification, therefore can be seen as a robust technique. The methods are illustrated by estimating the weekly working hours using data from the 2008 Quality of Life Survey in Colombia.

Key words: *Empirical Bayes estimation; random-effects model; post-stratification; simulation study; sample survey weights; nonparametric regression.*

1 Introduction

When working with sample surveys representativeness is difficult to achieve for all variables of interest. Several sampling designs are available in order to get valid inference from a population, including stratification, multi-stage and clustering. The former is intended to increase precision or allow sup-population level estimates, multi-stage facilitates a more efficient fieldwork at the cost of decrease in precision, clustering occurs when several dependent units are selected simultaneously. Stratification consists in partitioning the population in more homogeneous subgroups, according to the levels of an auxiliary variable which needs to be known prior to sampling, and sample each subpopulation (stratum) independently.

Individual responses may vary with several factors such as age, sex, education, none available prior to sampling for stratification purpose. However, if the population distribution is known at the aggregate level, e.g. from a census, the estimates of the outcome can be adjusted in the analysis stage. This method is called post-stratification, defined as a stratified analysis of a sample that was taken in an un-stratified way, or more general, is an analysis that uses more stratifying variables than at design stage (Molenberghs, 2013). Post-stratification reduces bias without improving precision as much as full stratification (i.e. stratification at design stage) because strata sample sizes are not fixed by design, adding a new source of variability. The stratification after selection is particularly useful in multi-purpose surveys where stratification factors selected prior to sampling may be weakly correlated with several secondary variables (Holt and Smith, 1979). Moreover, the post-stratification technique intends to restore representativeness in observational studies, where data are collected in a nonrandom fashion from a population and differential rates are more likely.

The usual post-stratified estimator presents high variability and the unweighted alternative is highly biased. The trade-off between variance and bias can lead to high mean squared error especially in the presence of large weights, i.e. highly underrepresented groups, or small sample sizes (Elliott and Little, 2000). An alternative method consists in using a model to predict the non-sampled outcomes, known as a model-based approach. The price to pay is the assumption regarding the distribution of the response. Misspecification is reduced by allowing the model to vary within strata. Moreover, smoothing of the weights is achieved when considering stratum means as random-effects, called weight-smoothing models (ibid.). The main advantage of the approach is the borrowing of strength between strata through the shrinkage process related to the sample sizes per stratum. A doubly-robust version includes the responders selection probability per stratum in an attempt to nullify the bias created by misspecification (Vandendijck et al., 2014).

To shield further against misspecification a smooth function is used for the underlying weight stratum means, the nonparametric model (Elliott and Little, 2000). The selected function is a cubic spline, a curve composed of sections of cubic polynomials joined together at points called the knots of the spline that is continuous in its second derivative (Wood, 2006).

Lazzeroni and Little (1998) and Elliott and Little (2000) have compared the performance of the design-based estimators and the weight-smoothing models under several conditions via simulations for normal outcomes and ordered post-strata. The former study extends the mixed-model with exchangeable assumption, i.e. no systematic relation between the outcome and the post-stratified variable, by assuming the post-stratum means follow a linear trend, and by considering an autoregressive covariance structure of the post-stratum means. They find the latest a useful approach. In turn, the second research extends the weight-smoothing technique by including a nonparametric spline function for the underlying weight stratum means. The authors stress the superiority of their proposal. More recently, Chen et al. (2010) proposed a Bayesian Penalized Spline Predictive estimator for a proportion in the presence of unequal sampling probabilities. The authors use a probit regression to incorporate the inclusion probabilities into the estimation of the proportion via penalized splines. They find their estimator advantageous in terms of efficiency, coverage, average length of credible intervals, and robustness to misspecification. Finally, Vandendijck et al. (2014) focus on prevalence and trend estimation of binary outcomes and propose a doubly-robust weight-smoothing method, as an extension of the weight-smoothing technique, in an attempt to protect against model misspecification. They find consistently a good performance of their estimator. Although design-based and model-based methods for normally distributed data have been well described in literature, no comparison has been made including the doubly-robust estimator.

This report presents a simulation study to evaluate the performance of nine methods in estimating the mean of a continuous, normally-distributed, survey outcome variable when two post-stratified weights are high and an ordinal post-stratifying variable is available. Additionally, an application using data from the 2008 Quality of Life Survey in Colombia is conducted; weekly working hours is the outcome of interest and age the post-stratified variable. Section 2 summarizes the methods, starting from the design-based methods which include the post-stratified, the unweighted mean and the weight-trimming; followed by the weight-smoothing techniques and finalizing with the promising doubly-robust estimators. The results of the simulation study under several conditions can be found in Section 3. Section 4 presents the application and Section 5 concludes.

2 Methodology

2.1 Notation

Let Y denote a continuous normally distributed survey outcome and X an ordinal post-stratifying variable with H strata. Let N_h and n_h denote the population size and the sample size, respectively, in each post-stratum h, $h = 1, 2, \ldots, H$. It is assumed that the population distribution, N_h, is known. Let $N = \sum_{h=1}^{H} N_h$ and $n = \sum_{h=1}^{H} n_h$ denote the total population and sample size, respectively.

Responders in each post-stratum are treated as a random sample, implying ignorable probability of inclusion in each stratum. The objective is to estimate the population mean, i.e. $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{N}\sum_{h=1}^{H} N_h \frac{\sum_{i=1}^{N_h} Y_i}{N_h} = \frac{1}{N}\sum_h N_h \bar{Y}_h$. The normalized weights are denoted by $w_h = \frac{N_h/N}{n_h/n}$. Note that $\sum_{i=1}^{n} w_h = \sum_{h=1}^{H}\sum_{i=1}^{n_h} w_h = \sum_{h=1}^{H} n_h w_h = n$. Finally, arranging in ascending order the strata according to the normalized weight values, let l denotes the pooling level such that $w_h < w_0$ if $h < l$ and $w_h \geq w_0$ if $h \geq l$ for a given cut-point w_0.

The entire simulation was conducted in R Development Core Team (2011) and the mixed-models were fitted using the `lme` function implemented in the `nlme` package (Pinheiro et al., 2011). For the bootstrap part of the simulation the function `rmvnorm` included in the `mvtnorm` package was used (Genz et al., 2012).

2.2 Design-Based Methods

Design-based estimators and their variances are shown in Table 1. The post-stratified mean is an unbiased estimator which weights each stratum mean by the relative size of that stratum in the population, correcting for unbalanced samples. However, its variance may explode when a post-stratum contains few observations, small n_h for some h, vanquishing the bias reduction. The unweighted mean is unbiased when $\bar{Y}_h$ is constant for all h and when the sample is truly representative of the population, i.e. $n_h/n = N_h/N$, otherwise it is biased (Holt and Smith, 1979). Using the means per stratum, $\bar{y}_h$, regardless of the number of observations per stratum, n_h, can lead to unstable estimates, alternative techniques solve the instability by taking into account the sample size distribution. The weight-trimming estimator fixes weights larger than a cut-point value, w_0, to w_0 and adjusts low weights upward by a constant, γ, to maintain the untrimmed weight sum. This way it pools all units with high weights into a stratum with reduced weight. The approach reduces variance at the expense of introducing some bias (Elliott and Little, 2000). The underlying weight pooling intention seems appealing but, at the same time, arbitrary in the selection of the cut-point. Potter (1990) proposes alternative methods to choose 'the optimal' w_0 based on data, the criteria attempt to minimize quantities related to the estimated mean squared error and bias. The methods include deriving the distribution of the sampling weights based on the assumption that the selection probabilities follow a standard beta distribution. The author says that a weight with extreme low probability of occurring can be trimmed to a specific probability of occurrence, e.g. using 0.01 a weight with value in excess of w_{op}, where $1 - F(w_{op}) = 0.01$, is trimmed to w_{op}.

Moreover, the post-stratified, the unweighted mean and the weight-trimming can be written as $\sum_{i=1}^{n} w_{h(i)} y_i$ when $w_{h(i)}$ takes the values w_h, 1 and w_h^* respectively:

- Post-stratified: $\sum_{i=1}^{n} w_{h(i)} y_i = \frac{1}{n}\sum_{i=1}^{n} w_h y_i = \frac{1}{n}\sum_{h=1}^{H} \frac{n}{N} N_h \frac{\sum_{i=1}^{n_h} y_i}{n_h} = \frac{1}{N}\sum_{h=1}^{H} N_h \bar{y}_h$.
- Unweighted: $\sum_{i=1}^{n} w_{h(i)} y_i = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{h=1}^{H} n_h \frac{\sum_{i=1}^{n_h} y_i}{n_h} = \frac{1}{n}\sum_{h=1}^{H} n_h \bar{y}_h$.

- Weighed-trimming: $\sum_{i=1}^{n} w_{h(i)} y_i = \frac{1}{n}\sum_{i=1}^{n} w_h^* y_i = \frac{1}{n}\left[\sum_{h=1}^{l-1}\sum_{i=1}^{n_h}\gamma w_h y_i + \sum_{h=l}^{H}\sum_{i=1}^{n_h} w_0 y_i\right]$
$= \frac{1}{n}\left[\sum_{h=1}^{l-1}\gamma N_h \frac{n}{N}\frac{\sum_{i=1}^{n_h} y_i}{n_h} + \sum_{h=l}^{H} w_0 n_h \frac{\sum_{i=1}^{n_h} y_i}{n_h}\right] = \frac{\gamma}{N}\sum_{h=1}^{l-1} N_h \bar{y}_h + \frac{w_0}{n}\sum_{h=l}^{H} n_h \bar{y}_h.$

Table 1: *Mean and variance estimators for the design-based methods: post-stratified, unweighted and weight-trimming*

Estimator	Mean	Variance
Post-stratified	$\bar{y}_{ps} = \frac{1}{N}\sum_{h=1}^{H} N_h \bar{y}_h$	$V(\bar{y}_{ps}) = \frac{1}{N^2}\sum_{h=1}^{H}\left(\frac{N_h}{n_h}\right)^2\left(1-\frac{n_h}{N_h}\right) n_h s_h^2$
Unweighted	$\bar{y}_{uw} = \frac{1}{n}\sum_{h=1}^{H} n_h \bar{y}_h$	$V(\bar{y}_{uw}) = \frac{1}{n^2}\sum_{h=1}^{H}\left(1-\frac{n_h}{N_h}\right) n_h s_h^2$
Weight-trimming	$\bar{y}_{tr} = \frac{\gamma}{N}\sum_{h=1}^{l-1} N_h \bar{y}_h + \frac{w_0}{n}\sum_{h=l}^{H} n_h \bar{y}_h,$	$V(\bar{y}_{tr}) = \frac{1}{n^2}\sum_{h=1}^{H}\left(1-\frac{n_h}{N_h}\right) n_h w_h^* s_h^2,$
	where $\gamma = \frac{n - w_0 \sum_{h=l}^{H} n_h}{\sum_{h=1}^{l-1} n_h w_h}$	where $w_h^* = \begin{cases}\gamma w_h & \text{if } w_h \le w_0\\ w_0 & \text{if } w_h > w_0\end{cases}$

Notation:
$N = \sum_{h=1}^{H} N_h$; $n = \sum_{h=1}^{H} n_h$; $w_h = \frac{N_h/N}{n_h/n}$; $\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$; $s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)^2$
l: pooling level such that $w_h < w_0$ if $h < l$ and $w_h \geq w_0$ if $h \geq l$ for a given cut-point w_0

2.3 Weight-Smoothing Model

Design-based methods ignore the ordinal nature of the post-stratifying variable. As an alternative a model-based approach reflects the intrinsic order and allows to borrow strength from neighboring strata with more information. The method involves modeling the weight-stratum means as random-effects (Lazzeroni and Little, 1998):

$$y_{hi}|\mu_h \sim N(\mu_h, \sigma^2)$$
$$\boldsymbol{\mu} \sim N_H(\boldsymbol{\delta}, D)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_H)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_H)$ and D is an $H \times H$ covariance matrix of the post-stratum means. Written in the mixed-effect form (Laird and Ware, 1982) $\boldsymbol{y} = NX\boldsymbol{\beta} + NZ\boldsymbol{b} + \boldsymbol{\varepsilon}$; where N is an $n \times H$ incidence matrix (Green and Silverman, 1994, p. 65) that indicates to which stratum an observations belongs (it captures the connection between the observations and the ordered distinct values, the entries $n_{hi} = 1$ if $y_i \in$ stratum h and 0 otherwise), X is an $H \times p$ fixed-effects design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed-effects parameters, Z is an $H \times q$ random-effects design matrix, $\boldsymbol{b}$ is a $q \times 1$ vector of random-effects such that $\boldsymbol{b} \sim N_q(\mathbf{0}, G)$, and $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 I_n)$.

The estimator uses the strata means, $\bar{y}_h$, for the individuals in the sample, and the predicted mixed-model means, $\hat{\mu}_h$, for those not included in the sample (Lazzeroni and Little, 1998):

$$\bar{y}_{ws} = \frac{1}{N}\sum_{h=1}^{H}\left[n_h \bar{y}_h + (N_h - n_h)\hat{\mu}_h\right]$$

where $\hat{\mu}_h = E(\mu_h)$ shrinks the sample means $\bar{y}_h$ towards δ_h in an amount according to the stratum sample size n_h: the bigger the sample size the less shrinkage. Note that the estimator is design-consistent[1] since it reduces to the post-stratified mean when strata sample size increases. The unweighted mean is a special case when $D \to 0$ and the post-stratified is obtained when $D \to \infty$. The following three assumptions are considered for the underlaying weight stratum means:

[1] An estimator e is asymptotically design consistent for $\bar{Y}_t$ if $\lim_{t\to\infty} P[|e_t - \bar{Y}_t| > \epsilon] = 0$ for every $\epsilon > 0$ (Mukhopadhyay, 1996, p. 96)

- Exchangeable random-effects (XR): $\delta_h = \beta_0\ \forall h$, $D = \tau^2 I_H$ (Holt and Smith 1979, Ghosh and Meeden 1986, Little 1991, Lazzeroni and Little 1998).
  In the mixed-effect notation: $X = \mathbf{1}_H$, $\boldsymbol{\beta} = \beta_0$, $Z = I_H$, $G = \tau^2 I_H$. Therefore,
  $y_{hi} = \beta_0 + b_h + \varepsilon_{hi}$; $b_h \sim N(0,\tau^2)$; $\varepsilon_{hi} \sim N(0,\sigma^2)$.

- Linear (LI): $\delta_h = \beta_0 + \beta_1 h$, $D = \tau^2 I_H$ (Lazzeroni and Little, 1998).
  In the mixed-effect notation: $X = [\mathbf{1}_H|(1,2,\dots,H)']$, $\boldsymbol{\beta} = (\beta_0,\beta_1)'$, $Z = I_H$, $G = \tau^2 I_H$. Therefore,
  $y_{hi} = \beta_0 + \beta_1 h + b_h + \varepsilon_{hi}$; $b_h \sim N(0,\tau^2)$; $\varepsilon_{hi} \sim N(0,\sigma^2)$.

- Nonparametric (NP): $\delta_h = f(h)$, $D = \tau^2 I_H$ (Elliott and Little 2000, Zheng and Little 2004).
  In the mixed-effect notation: $X = [\mathbf{1}_H|(1,2,\dots,H)']$, $\boldsymbol{\beta} = (\beta_0,\beta_1)'$, $Z = [Z_1|Z_2]$, $\boldsymbol{b} = \left[\frac{\boldsymbol{b}_s}{\boldsymbol{b}_h}\right]$,
  $\boldsymbol{b}_s$ is a $H \times 1$ vector of the random-effects of the smooth nonparametric function, $\boldsymbol{b}_h$ is a $H \times 1$ vector of the post-stratum mean random-effects (similar to the vector of random-effects in XR and LI assumptions), $\boldsymbol{b} \sim N_{2H}(\mathbf{0}, G)$, $Z_2 = I_H$, $Z_1 = (S^{-1}Z^*)'$, $S = UE^{1/2}V'$, U and V are the components of the singular value decomposition of Z^* (i.e. $Z^* = UEV'$) and $E^{1/2}$ a diagonal matrix with elements the squared root of the elements in E,
  $$Z^* = \begin{pmatrix} |1-\kappa_1|^3 & |1-\kappa_2|^3 & \cdots & |1-\kappa_H|^3 \\ |2-\kappa_1|^3 & |2-\kappa_2|^3 & \cdots & |2-\kappa_H|^3 \\ \vdots & \vdots & \dots & \vdots \\ |H-\kappa_1|^3 & |H-\kappa_2|^3 & \cdots & |H-\kappa_H|^3 \end{pmatrix}, G = \begin{pmatrix} \tau_s^2 I_H & \mathbf{0}_{H\times H} \\ \mathbf{0}_{H\times H} & \tau^2 I_H \end{pmatrix}.$$ Therefore,
  $y_{hi} = f(h) + b_h + \varepsilon_{hi}$; $b_h \sim N(0,\tau^2)$; $\varepsilon_{hi} \sim N(0,\sigma^2)$; f is a natural cubic smoothing spline with radial basis functions $1, h, |h-\kappa_1|^3, \dots, |h-\kappa_H|^3$ (called radial since, in its general form, is radially symmetric around $\kappa_j, j = 1,\dots,H$) that can be written as $\hat{f}(h) = \hat{\beta}_0 + \hat{\beta}_1 h + \sum_{j=1}^{H} \hat{b}_{sj}|h-\kappa_j|^3$ with knots $(\kappa_1,\dots,\kappa_H)$ at $1,\dots,H$ and where $\hat{\beta}_0, \hat{\beta}_1, \hat{b}_{s1}, \dots, \hat{b}_{sH}$ minimize $\|\boldsymbol{y} - NX\boldsymbol{\beta} - NZ^*\boldsymbol{b}_s\|^2 + \lambda \boldsymbol{b}_s' Z^* \boldsymbol{b}_s$ subject to $X'\boldsymbol{b}_s = \mathbf{0}$ (Ruppert et al., 2003, p. 73); $b_{sh} \sim N(0,\tau_s^2)\ \forall h$ the random-effects of the smooth nonparametric function. The parameter $\lambda = \frac{\tau^2}{\tau_s^2}$ is the roughness penalty to control the trade-off between data fitting and smoothness of f, $\lambda \to 0$ implies $\bar{y}_{np} \to \bar{y}_{ps}$ and $\lambda \to \infty$ implies $\bar{y}_{np} \to \bar{y}_{li}$, so the nonparametric assumption can be viewed as a compromise between the linear assumption and the post-stratified mean (Elliott and Little, 2000).

Some model-based estimators correspond to the standard design-based estimators, for example the unweighted mean coincides with XR under an equal-probability design where n_h are approximately constant across strata (Zheng and Little, 2004). Variance components $\sigma^2, \tau^2, \tau_s^2$ can be estimated by fitting the respective models via restricted maximum likelihood when maximizing $L_{\text{REML}}(\boldsymbol{\theta}) = |\sum_{i=1}^{N} X_i' V_i^{-1}(\boldsymbol{\alpha}) X_i|^{-1/2} L_{\text{ML}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ (vector containing variance components and fixed-effects). The fixed-effects are estimated by $\hat{\boldsymbol{\beta}} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\bar{\boldsymbol{y}}$ and the random-effects are predicted using the posterior mean $\hat{\boldsymbol{b}} = \hat{G}Z'\hat{V}^{-1}(\bar{\boldsymbol{y}} - X\hat{\boldsymbol{\beta}})$ where $\hat{V} = Z\hat{G}Z' + A$, $A = \hat{\sigma}^2 diag(\frac{1}{n_1},\dots,\frac{1}{n_H})$ and $\bar{\boldsymbol{y}} = (\bar{y}_1,\dots,\bar{y}_H)'$ (Zheng and Little, 2004).

The variance formula is (Zheng and Little 2004, Vandendijck et al. 2014):

$$V(\bar{y}_{ws}) = \frac{1}{N^2}(\boldsymbol{N} - \boldsymbol{n})'\boldsymbol{\Theta}(\boldsymbol{N} - \boldsymbol{n})$$

where $(\boldsymbol{N} - \boldsymbol{n}) = (N_1 - n_1, \dots, N_H - n_H)'$; $\boldsymbol{\Theta} = C(C'A^{-1}C + B)^{-1}C'$; $A = \sigma^2 diag\left(\frac{1}{n_1},\dots,\frac{1}{n_H}\right)$ and

- Exchangeable random-effects (XR): $C = [\mathbf{1}_H|I_H]$; $B = \begin{pmatrix} 0 & \mathbf{0}_{1\times H} \\ \mathbf{0}_{H\times 1} & \frac{1}{\tau^2} I_H \end{pmatrix}$.

- Linear (LI): $C = [\mathbf{1}_H|(1,\ldots,H)'|I_H]$; $B = \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times H} \\ \mathbf{0}_{H\times 2} & \frac{1}{\tau^2} I_H \end{pmatrix}$.
- Nonparametric (NP): $C = [\mathbf{1}_H|(1,\ldots,H)'|Z_1|I_H]$; $Z_1 = (S^{-1}Z^*)'$; $S = UE^{1/2}V'$; U and V are the components of the singular value decomposition of Z^* (i.e. $Z^* = UEV'$) and $E^{1/2}$ a diagonal matrix with elements the squared root of the elements in E;

$$Z^* = \begin{pmatrix} 0 & 1^3 & 2^3 & \cdots & |1-H|^3 \\ 1^3 & 0 & 1^3 & \cdots & |2-H|^3 \\ 2^3 & 1^3 & 0 & \cdots & |3-H|^3 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ |H-1|^3 & |H-2|^3 & |H-3|^3 & \cdots & 0 \end{pmatrix}; B = \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2H} \\ \mathbf{0}_{2H\times 2} & \begin{pmatrix} \frac{1}{\tau_s^2} I_H & \mathbf{0}_{H\times H} \\ \mathbf{0}_{H\times H} & \frac{1}{\tau^2} I_H \end{pmatrix} \end{pmatrix}.$$

2.4 Weight-Smoothing Doubly-Robust Estimator

The weight-smoothing estimator can be rewritten in the following form (Vandendijck et al., 2014):

$$\begin{aligned} \bar{y}_{ws} &= \frac{1}{N}\sum_{h=1}^{H}[n_h\bar{y}_h + (N_h - n_h)\hat{\mu}_h] = \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{n_h} y_{hi} + \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}\hat{\mu}_h - \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{n_h}\hat{\mu}_h \\ &= \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}\hat{\mu}_h + \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{n_h}(y_{hi} - \hat{\mu}_h) = \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}\hat{\mu}_h + \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h} r_{hi}\hat{\varepsilon}_{hi} \end{aligned}$$

where $\hat{\varepsilon}_{hi} = Y_{hi} - \hat{\mu}_h$ are the estimated residuals and $r_{hi} = 1$ if Y_{hi} is in the sample and zero otherwise. When the model is correct $\hat{\mu}_h$ is un unbiased estimator of $\bar{Y}_h$, i.e. $E(\hat{\mu}_h) = \bar{Y}_h$, implying $E(\hat{\varepsilon}_{hi}) = E(Y_{hi} - \hat{\mu}_h) = 0$. Therefore, $E(\bar{y}_{ws}) = E\left(\frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}\hat{\mu}_h\right) = \frac{1}{N}\sum_{h=1}^{H} N_h\bar{Y}_h = \bar{Y}$, i.e. $\bar{y}_{ws}$ is unbiased.

Model misspecification of the weigh-smoothing estimator can lead to bias. As an alternative, the weight-smoothing doubly-robust estimator proposed by Vandendijck et al. (2014) provides protection against misspecification. The authors present an estimator of the form $\bar{y}_{wsdr} = \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}\hat{\mu}_h + \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}\frac{r_{hi}}{\hat{\pi}_{hi}}\hat{\varepsilon}_{hi}$, where $\hat{\pi}_{hi}$ is the inclusion probability in the observed sample, assuming same inclusion probabilities within strata the estimator becomes:

$$\bar{y}_{wsdr} = \frac{1}{N}\sum_{h=1}^{H}\left[\frac{n_h}{\hat{\pi}_h}\bar{y}_h + \left(N_h - \frac{n_h}{\hat{\pi}_h}\right)\hat{\mu}_h\right]$$

where $\hat{\pi}_h = \frac{\hat{n}_h}{N_h}$; $\hat{n}_h = \begin{cases} \frac{N_h/N}{w_0/n} & \text{if } w_h > w_0 \\ \gamma n_h & \text{if } w_h \le w_0 \end{cases}$; $\gamma = \frac{n - \sum_{h=l}^{H}\hat{n}_h}{\sum_{h=1}^{l-1} n_h}$.

When the model is correct $E(\bar{y}_{wsdr}) = \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}E(\hat{\mu}_h) + \frac{1}{N}\sum_{h=1}^{H}\frac{n_h}{\hat{\pi}_h}E(\bar{y}_h) - \frac{1}{N}\sum_{h=1}^{H}\frac{n_h}{\hat{\pi}_h}E(\hat{\mu}_h) = \frac{1}{N}\sum_{h=1}^{H}N_h\bar{Y}_h + \frac{1}{N}\sum_{h=1}^{H}\frac{n_h}{\hat{\pi}_h}\bar{Y}_h - \frac{1}{N}\sum_{h=1}^{H}\frac{n_h}{\hat{\pi}_h}\bar{Y}_h = \bar{Y}$, so it is unbiased regardless of the inverse probability weights. When the model is misspecified but the weights are correct $\frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}E(\hat{\mu}_h) = \frac{1}{N}\sum_{h=1}^{H}\frac{n_h}{\hat{\pi}_h}E(\hat{\mu}_h)$, therefore $E(\bar{y}_{wsdr}) = \frac{1}{N}\sum_{h=1}^{H}\frac{n_h}{\hat{\pi}_h}\bar{Y}_h = \bar{Y}$. The estimator is called doubly-robust because, given its form, it attempts to nullify the bias coming from misspecification but no proof is provided.

The variance is estimated via bootstrap according to the following scheme:

1. Set the number of bootstrap populations and samples per population, b_{pop} and b_{sam} respectively. The total number of bootstrap samples is then $b_{pop} \times b_{sam}$.

2. For $k = 1, \ldots, b_{pop}$

   (a) Simulate a random vector $\boldsymbol{u}^{(k)} \sim N_H\left(\mathbf{0}, \hat{D}\right)$, where $\hat{D} = \hat{\tau}^2 I_H$ is the fitted covariance matrix of the post-stratum means (defined in Section 2.3).

   (b) Calculate $\boldsymbol{\delta}^{(k)} = \left(\delta_1^{(k)}, \ldots, \delta_H^{(k)}\right)' = X\hat{\boldsymbol{\beta}} + Z\boldsymbol{u}^{(k)}$ for XR and LI models, and $\boldsymbol{\delta}^{(k)} = X\hat{\boldsymbol{\beta}} + Z_1\hat{\boldsymbol{b}}_s + Z_2\boldsymbol{u}^{(k)}$ for the NP model (same notation as used in Section 2.3).

   (c) Draw sample sizes per stratum from a multinomial distribution: $\left(n_1^{(k)}, \ldots, n_H^{(k)}\right) \sim \text{Multinom}\left[n, \left(\frac{n_1}{N_1}, \ldots, \frac{n_H}{N_H}\right)\right]$.

   (d) For $j = 1, \ldots, b_{sam}$ draw outcomes in each post-stratum from normal distribution: $y_{hi}^{(k,j)} \sim N\left(\delta^{(k)}, \hat{\sigma}^2\right)$, $i = 1, \ldots, n_h^{(k)}$. Where $\hat{\sigma}$ comes from the model fitted to the original data.

3. For each k and j fit the model to the sampled data and calculate the doubly-robust mean estimates for XR, LI and NP: $\bar{y}_{wsdr}^{(k,j)}$.

4. For each k calculate the variance of the estimates: $v^{(k)} = \frac{1}{b_{sam}-1} \sum_{j=1}^{b_{sam}} \left(\bar{y}_{wsdr}^{(k,j)} - \frac{1}{b_{sam}} \sum_{j=1}^{b_{sam}} \bar{y}_{wsdr}^{(k,j)}\right)^2$.

5. The estimated variance is obtained by averaging the variances of the previous step: $\hat{V}(\bar{y}_{wsdr}) = \frac{1}{b_{pop}} \sum_{k=1}^{b_{pop}} v^{(k)}$.

3 Simulation Study

A simulation, similar to that in Elliott and Little (2000), was conducted to evaluate the performance of the methods under different conditions. Table 2 presents the population and sample size distributions over the 10 strata used. The normalized post-stratified weights range from 0.1 to 13.9, same for both samples. The values of y_{hi} were generated as $y_{hi} = \delta_h + \varepsilon_{hi}$ considering four scenarios for the mean structure δ_h: δ^C, δ^D, δ^E, and δ^L (called closed, distant, equal and linear respectively, the first two according to its proximity to the last strata, the underrepresented ones) as Figure 1 shows; $\varepsilon_{hi} \sim N(0, \sigma^2)$ with three values for $\sigma = 1,\ 5,\ 10$. Parameters of the mean structure are such that $E\left(\bar{Y}|\delta^D\right) = E\left(\bar{Y}|\delta^E\right) = E\left(\bar{Y}|\delta^L\right) = 10.88$ and $E\left(\bar{Y}|\delta^C\right) = 0$. This way a total of $2 \times 4 \times 3$ (n's$\times\delta$'s $\times\sigma$'s) = 24 populations are considered. For each scenario 50 populations are randomly generated and in each 10 samples are drawn yielding a total of 500 replications per combination.

The overall population mean is estimated using the 9 methods described: unweighted $\bar{y}_{uw}$ (uw), post-stratified $\bar{y}_{ps}$ (ps), trimmed with a cut-off value $w_0 = 3$ $\bar{y}_{tr}$ (tr), weight-smoothing using exchangeable assumption $\bar{y}_{ws.xr}$ (ws.xr), weight-smoothing using linear assumption $\bar{y}_{ws.li}$ (ws.li), weight-smoothing using nonparametric assumption $\bar{y}_{ws.np}$ (ws.np), weight-smoothing doubly-robust using exchangeable assumption $\bar{y}_{wsdr.xr}$ (wsdr.xr), weight-smoothing doubly-robust using linear assumption $\bar{y}_{wsdr.li}$ (wsdr.li), and weight-smoothing doubly-robust using nonparametric assumption $\bar{y}_{wsdr.np}$ (wsdr.np). I used $w_0 = 3$ for all doubly-robust estimates.

For each method, I calculate the average bias, variability, mean squared error (MSE), coverage and 95% confidence interval (C.I.). The MSE$^{(p)}$ (MSE in population p) was obtained from the 10 estimates, one per sample, and the overall MSE was calculated by averaging over the 50 MSE$^{(p)}$ values as follows: let $\mu^{(p)}$ be the true mean in population p, $\hat{\mu}_j^{(p)}$ the estimate of $\mu^{(p)}$ in sample j, $j = 1, \ldots, 10$, and $\bar{\hat{\mu}}^{(p)} = \frac{1}{10}\sum_{j=1}^{10} \hat{\mu}_j^{(p)}$ the average of the 10 estimates in population p. The bias is defined as $\text{Bias}(\hat{\mu}^{(p)}) = \bar{\hat{\mu}}^{(p)} - \mu^{(p)}$, the variance is $V(\hat{\mu}^{(p)}) = \frac{1}{9}\sum_{j=1}^{10}\left(\hat{\mu}_j^{(p)} - \bar{\hat{\mu}}^{(p)}\right)^2$, the MSE per population is $\text{MSE}^{(p)} = V\left(\hat{\mu}^{(p)}\right) + \left(\text{Bias}\left(\hat{\mu}^{(p)}\right)\right)^2$ and the overall MSE is MSE$= \frac{1}{50}\sum_{p=1}^{50} \text{MSE}^{(p)}$. The 95% C.I. was calculated by $\hat{\mu}_j^{(p)} \pm z\sqrt{\hat{V}\left(\hat{\mu}_j^{(p)}\right)}$ in population p, sample j, $\hat{V}\left(\hat{\mu}_j^{(p)}\right)$ is the fitted variance of the estimate and z the 97.5th quantile of the standard normal distribution. The variances of the doubly-robust estimators were calculated using the bootstrap procedure with $b_{pop} = 25$ and $b_{sam} = 10$.

Table 2: *Population and sample sizes in the 10 strata used in the simulation study*

Stratum h	1	2	3	4	5	6	7	8	9	10	Total
N_h	800	1,000	1,200	1,500	2,000	3,000	4,000	5,000	7,500	10,000	36,000
$n_{1,h}$	90	80	70	60	50	50	40	30	20	10	500
$n_{2,h}$	18	16	14	12	10	10	8	6	4	2	100
w_h	0.1	0.2	0.2	0.3	0.6	0.8	1.4	2.3	5.2	13.9	

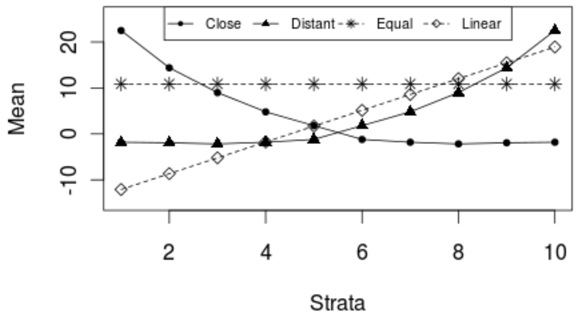

Figure 1: *Mean structure for the parameters (δ^C: close, δ^D: distant, δ^E: equal, δ^L: linear) used to generate samples for the simulation study*

Table 3 presents the MSE of the 9 estimators for all 24 scenarios. The unweighted mean has the biggest MSE for close, distant and linear mean structures due to the huge bias; it has the best performance for equal means and low variance in all scenarios (see Figure 3). The post-stratified estimator is unbiased and has the biggest variance specially for equal mean structure. The weighted-trimming has low variance, is notably biased for distant and linear scenarios and slightly biased for close mean structure. The exchangeable random-effects estimator performs best under equal mean structure, in the other scenarios the bias increases with large values of σ and small sample size. Linear estimator has the best performance under linear structure and behaves well under equal scenario; is biased under close and distant structure for small sample size and $\sigma \geq 5$, and for large sample size and $\sigma = 10$. The nonparametric estimator has the best performance under close structure, the second best under distant scheme and the third best under linear scenarios but it is slightly biased under distant structure specially when σ increases and sample size decreases. The doubly-robust version decreases the bias although it persist for large σ and small sample size under distant for XR, LI and NP, under close for XR and LI and under linear for XR. The doubly-robust estimators have also bigger variance than their counterparts, which is not compensated by the reduction in bias under equal structure and for some linear scenarios. The variance of the estimators increases with larger values of σ but mostly when sample size decreases; similarly for MSE but σ has the biggest impact (see Figures 2 and 3).

Regarding coverage of the nominal 95% confidence intervals, the post-stratified mean performs better for large sample size. The unweighted mean and the weighted-trimming method have bad coverage for all but equal structure, due to strong bias. The doubly-robust approaches improve the coverage, compared with its counterpart, in all but equal scheme. Among the models, the nonparametric consistently maintains the coverage even under large σ and small sample sizes. The factor that mostly affects coverage is the sample size (see Table 5 in the Appendix).

Finally, concerning average length of the confidence intervals, among the ones with at least 90% actual coverage, equal scheme yields short intervals. The length increases with larger values of σ as well as with smaller sample sizes. The post-stratified, the unweighted mean and the weighted-trimming present similar lengths regardless of the mean structure whereas for the weight-smoothing and the doubly-robust similarities are only predominant in close and distant schemes. Under equal structure, the unweighted mean has the smaller length followed by the XR weight-smoothing. Under linear scheme the weight-smoothing estimators have the smallest length, being LI model the best followed by NP. In close scenario weight-smoothing NP has the smallest length when $\sigma \leq 5$ and doubly-robust XR overcomes when $\sigma = 10$. Under distant scheme weight-smooth NP has the smaller length for $\sigma \leq 5$, except the combination ($\sigma = 5, n_2$) when the post-stratified mean wins, and for $\sigma = 10$ doubly-robust

LI is the best (see Table 6 in the Appendix).

Lazzeroni and Little (1998) considered all designed-based methods and from model-based weight-smoothing, the LI and XR assumptions. Similarly, they find the unweighted mean does well under equal scenario but very poorly when population means have an structure, due to serious bias. The gain of smoothing vanishes as the sample size increases since LI and XR converge towards the post-stratified estimator. The XR seriously undercovers for linear structure whilst the unweighted mean has poor coverage in all but equal scenario specially for large sample size where bias prevails over variance. The XR and LI models achieve a reduction in the average width of confidence intervals compared with the post-stratified; the smaller the sample size, the greater the reduction.

Elliott and Little (2000) studied all but the doubly-robust estimators with the difference that under the nonparametric model they assumed $D = 0$, implying the post-stratifying means do not vary around the spline function. In general, results go in the same direction. The authors conclude that unlike close mean structure, distant is unfavorable for trimming since the mean of the most underrepresented stratum differs substantially from the other and therefore pooling of the strata is not appropriate. The XR estimator does well under equal scenario but performs poorly, relative to $\bar{y}_{ps}$, in all other mean structures. The LIN works well under equal and linear schemes but is less efficient than XR in the former and has moderate coverage problems when the mean trend is not linear. Relative to the post-stratified, LIN works poorly in distant scenario. The NP performs nearly as well as LI under linear structures (δ^E, δ^L). Unlike current simulation, the authors find that under nonlinear scenarios NP mimics post-stratified for small σ^2 and LI when σ^2 increases; poor coverage of XR when means follow a linear trend and variance is moderate; and the post-stratified coverage is closed to nominal under all scenarios.

Table 3: *Mean squared error for the simulation study based on 9 estimators (Est. ps=post-stratified, uw=unweighted, tr=weight-trimmed, ws.xr=weight-smoothing exchangeable, ws.li=weight-smoothing linear, ws.np=weight-smoothing nonparametric, wsdr.xr=doubly-robust exchangeable, wsdr.li=doubly-robust linear, wsdr.np=doubly-robust nonparametric) and 500 subsamples for all 24 scenarios. Sample sizes:* $n_1 = 500$ *and* $n_2 = 100$*. Some subsamples discarded due to not invertible matrices in ws variance calculation*

	Close						Distant					
	$\sigma = 1$		$\sigma = 5$		$\sigma = 10$		$\sigma = 1$		$\sigma = 5$		$\sigma = 10$	
Est.	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
ps	0.013	0.053	0.303	1.462	1.230	5.488	0.011	0.061	0.366	1.526	1.156	5.999
uw	61.824	61.801	61.979	61.862	62.032	62.626	100.346	100.355	100.407	100.826	100.313	102.244
tr	1.204	1.214	1.334	1.754	1.597	3.137	27.064	27.076	26.993	28.083	27.298	29.233
ws.xr	0.013	0.053	0.307	1.486	1.211	5.745	0.012	0.064	0.405	3.189	2.038	23.682
ws.li	0.013	0.055	0.337	2.119	1.745	7.838	0.012	0.065	0.427	3.189	2.151	10.795
ws.np	0.013	0.052	0.283	1.355	1.119	4.806	0.012	0.063	0.361	1.622	1.208	6.306
wsdr.xr	0.013	0.053	0.299	1.368	1.145	4.439	0.012	0.062	0.384	2.412	1.614	13.338
wsdr.li	0.013	0.054	0.323	1.848	1.548	6.226	0.012	0.064	0.412	2.737	1.886	8.663
wsdr.np	0.013	0.052	0.283	1.342	1.100	4.663	0.012	0.062	0.357	1.564	1.176	5.981
	Equal						Linear					
	$\sigma = 1$		$\sigma = 5$		$\sigma = 10$		$\sigma = 1$		$\sigma = 5$		$\sigma = 10$	
Est.	$n_1^{(1)}$	$n_2^{(7)}$	$n_1^{(2)}$	$n_2^{(8)}$	$n_1^{(3)}$	$n_2^{(9)}$	$n_1^{(4)}$	$n_2^{(10)}$	$n_1^{(5)}$	$n_2^{(11)}$	$n_1^{(6)}$	$n_2^{(12)}$
ps	0.012	0.061	0.278	1.443	1.255	6.116	0.011	0.063	0.330	1.520	1.305	5.963
uw	0.002	0.012	0.058	0.280	0.228	1.152	150.882	150.900	150.983	151.489	151.300	153.182
tr	0.004	0.023	0.121	0.592	0.481	2.324	16.517	16.550	16.647	17.222	17.685	18.818
ws.xr	0.002	0.014	0.064	0.299	0.247	1.266	0.011	0.063	0.344	1.913	1.616	11.210
ws.li	0.006	0.034	0.177	0.877	0.696	3.573	0.006	0.030	0.178	0.741	0.727	3.583
ws.np	0.007	0.038	0.189	0.968	0.809	4.037	0.006	0.036	0.207	0.849	0.790	4.064
wsdr.xr	0.004	0.021	0.104	0.516	0.427	2.119	0.011	0.062	0.334	1.636	1.390	7.516
wsdr.li	0.007	0.035	0.182	0.904	0.746	3.722	0.006	0.032	0.189	0.782	0.751	3.791
wsdr.np	0.007	0.040	0.197	1.006	0.858	4.208	0.007	0.038	0.218	0.896	0.823	4.295

(1): 8 samples discarded. (2): 1 sample discarded. (3): 3 samples discarded. (4): 7 samples discarded.
(5): 4 samples discarded. (6): 6 samples discarded. (7): 22 samples discarded. (8): 14 samples discarded.
(9): 10 samples discarded. (10): 11 samples discarded. (11): 19 samples discarded. (12): 11 samples discarded.

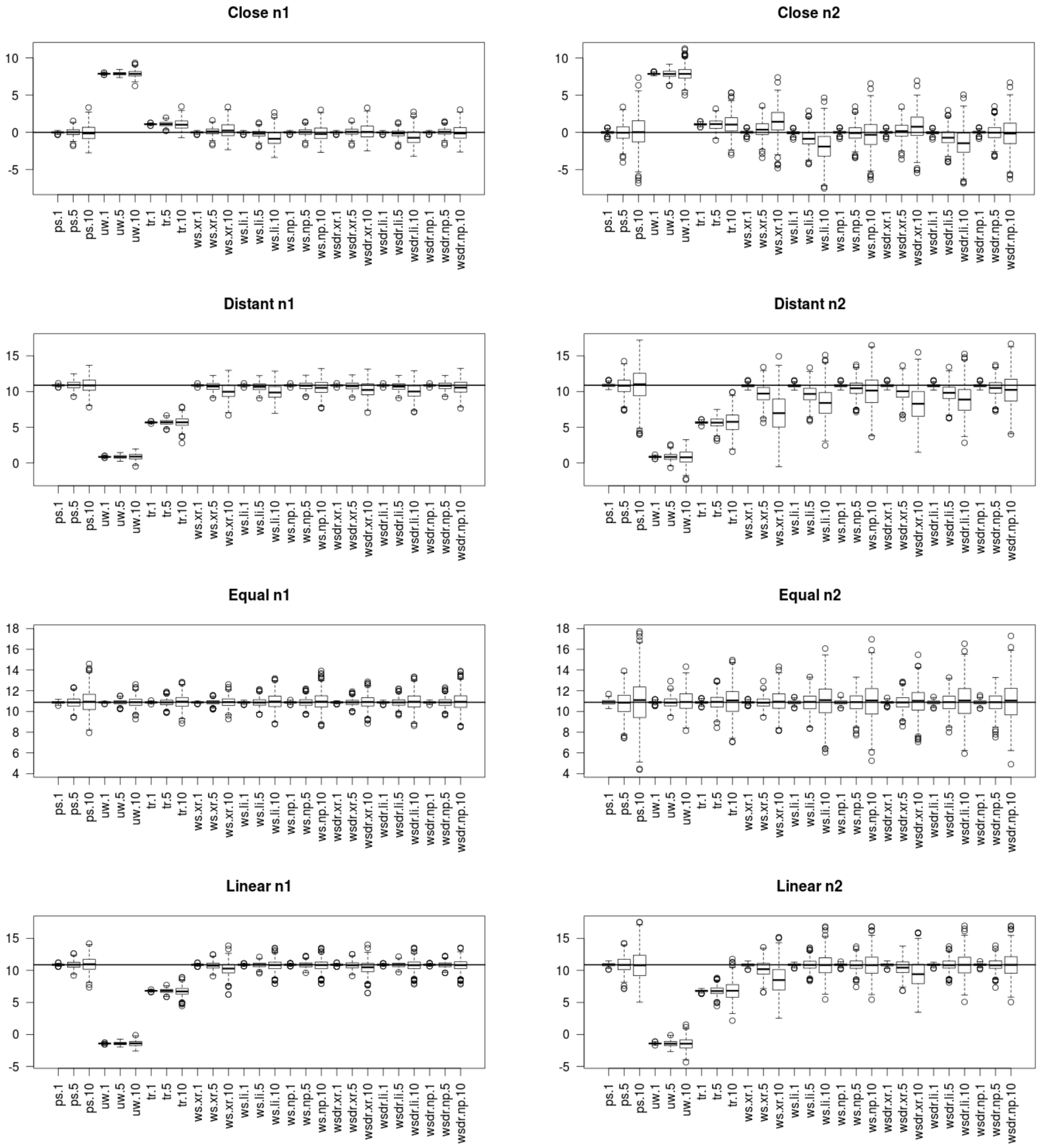

Figure 2: *Mean estimates for the simulation study based on 9 estimators (ps=post-stratified, uw=unweighted, tr=weight-trimmed, ws.xr=weight-smoothing exchangeable, ws.li=weight-smoothing linear, ws.np=weight-smoothing nonparametric, wsdr.xr=doubly-robust exchangeable, wsdr.li=doubly-robust linear, wsdr.np=doubly-robust nonparametric), four mean structures (close, distant, equal, linear) and three values of σ (1, 5 and 10). Sample sizes: $n_1 = 500$ and $n_2 = 100$*

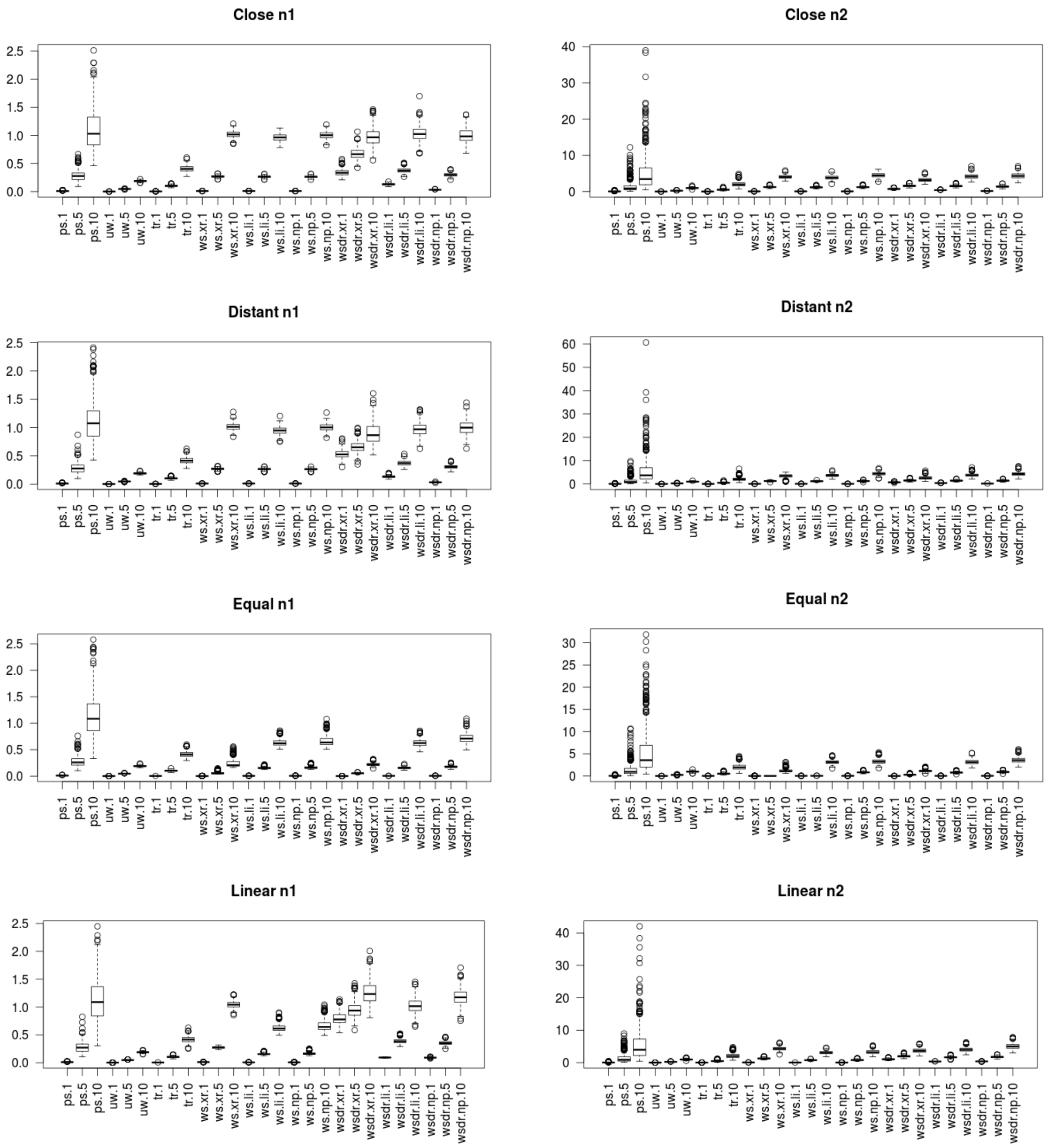

Figure 3: *Variance estimates for the simulation study based on 9 estimators (ps=post-stratified, uw=unweighted, tr=weight-trimmed, ws.xr=weight-smoothing exchangeable, ws.li=weight-smoothing linear, ws.np=weight-smoothing nonparametric, wsdr.xr=doubly-robust exchangeable, wsdr.li=doubly-robust linear, wsdr.np=doubly-robust nonparametric), four mean structures (close, distant, equal, linear) and three values of* σ *(1, 5 and 10). Sample sizes:* $n_1 = 500$ *and* $n_2 = 100$

4 Application

The 2008 Colombian Quality of Life Survey (QLS) was used to illustrate the methods. Performed by the National Administrative Department of Statistics during August the 11th and October the 18th nationwide with the aim of gathering information regarding socio-economic conditions of households. A total of 50,542 people were interviewed. The aim of the application is to estimate the average weekly working hours for 20 years or older Colombians using as post-stratifying variable age groups of 5 years. Younger than 20 years where excluded because it is expected most of them are studying rather than working. The population age distribution (N_h, $h = 1, \ldots, 13$) was taken from the website of United Nations (*www.un.org*). The survey is representative of the total population and therefore it nicely reflects the age distribution: normalized weights range from 0.69 to 1.16 (see Figure 4).

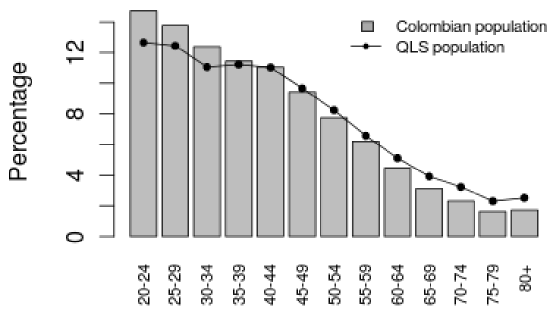

Figure 4: *Age distribution of the Quality of Life Survey (QLS) population and the overall Colombian population (older than 20 years) stratified by 13 age intervals of 5 years*

In order to see the advantage of the methods a simulation was performed. Unlike the preceding simulation study, this is more realistic since the values are drawn from a real rather than a simulated distribution. The survey was then considered as the true population and 100 subsamples where taken according to the following two schemes: subsample type 1 ($n_1 = 317$) with highest normalized weights 13.1 and 5.2 for underrepresented age groups 50-54 and 55-59 respectively; and subsample type 2 ($n_2 = 300$) which highest normalized weights are 9.5 and 6.2 for underrepresented age groups 20-24 and 25-29 respectively, Figure 5 shows the age distributions.

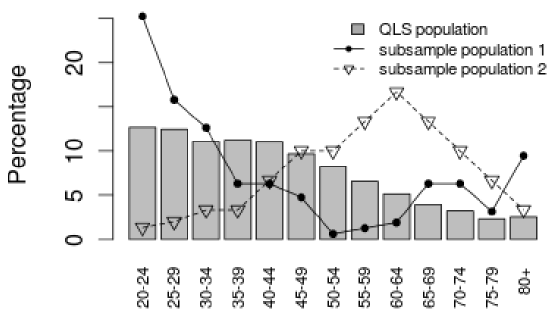

Figure 5: *Age distribution of the 2008 Colombian Quality of Life Survey (QLS) population and the sub-sampled populations (older than 20 years) stratified by 13 age intervals of 5 years*

In total $N = 30,118$ people aged from 20 to 104 were sampled. According to Figure 6 in the Appendix the age profile of the mean structure presents a quadratic curvature. Table 4 shows the results from the application. The best method in terms of MSE is the weight-trimming in both subsample types but in the first case the doubly-robust estimators are nearly as good, particularly the nonparametric one. The unweighted mean is by far the worst in the two scenarios despite its low variance, reflected in the smallest C.I. length, due to the high bias. In the second scenario all estimates are more variable and all but trimming and XR are more biased. In terms of bias, post-stratified is the best method in the first scheme but is also the most variable and the coverage is just 89%; whereas in the second scheme is outperformed by trimming and doubly-robust XR in terms of bias but remains the most variable. For detailed insight see Figure 7 in the Appendix.

Table 4: *Coverage, mean squared error (MSE), average bias (A. Bias), and average length of the 95% confidence interval (A.L.C.I.) of the mean weekly working hours (older than 20 years) for the 2008 Colombian Quality of Life Survey based on 9 estimators and 100 subsamples. Subsample type 1: 50-59 age groups highly underrepresented. Subsample type 2: 20-29 age groups highly underrepresented*

	Subsample type 1				Subsample type 2			
Estimator	Coverage (%)	MSE	A. Bias	A.L.C.I.	Coverage (%)	MSE	A. Bias	A.L.C.I.
ps	89	5.33	0.15	8.66	91	7.78	-0.72	10.72
uw	24	16.70	3.83	5.48	6	27.27	5.04	5.58
tr	96	3.16	0.21	6.95	97	3.95	0.02	7.74
$ws.xr$	88	4.56	1.12	7.15	90	5.77	0.91	8.46
$ws.li$	79	6.19	1.74	6.53	66	17.30	-3.31	8.72
$ws.np$	80	5.75	1.64	6.38	73	15.66	-2.98	8.88
$wsdr.xr$	92	3.82	0.44	7.20	87	5.54	-0.24	7.79
$wsdr.li$	89	3.81	0.65	7.05	80	12.74	-2.50	9.12
$wsdr.np$	91	3.66	0.60	7.21	86	11.80	-2.23	9.59

The unexpected bad performance of the nonparametric estimators in the second scenario, where they were worse than the post-stratified one, is due to $\tau_s^2 \to 0$ implying $\lambda \to \infty$ and $\bar{y}_{np} \to \bar{y}_{li}$. Figure 8 in the Appendix shows the predicted means in each post-stratum for three subsamples. They reflect that some cubic spline functions are extremely smoothed, because the roughness penalty is too high, resembling a linear function, not capturing the curvature and therefore biasing the estimator upwards.

5 Discussion

This report compares several post-stratification techniques and extensions for a Gaussian survey outcome when an ordinal factor is available and some groups are highly underrepresented. Standard methods such as post-stratified, unweighted mean and weight trimming perform poorly under unfavorable configurations due to either considerable bias or large variance. Alternative methods, under the mixed-model framework, show substantial improvement by taking into account the ordinal nature of the post-stratifying variable and the strata sample size when borrowing strength from well represented neighboring groups, hence overcoming the instability of the post-stratified estimator. Given that when strata sample size increase the weight-smoothing estimates tend to the post-stratified mean, they are design consistent. To avoid misspecification issues overshadow the appeal of the model-based approach, a doubly-robust version was tested. As model assumptions exchangeable random-effects, linear and nonparametric were considered.

When means among strata are equal, the simple unweighted design-based method is superior although all techniques work relatively fine given that all models allow for equal means. However, since in practical situations means differ, the unweighted estimator is not useful for routine use. As expected, under linear scheme, linear model stood out. Under non-linear scenarios, i.e. close and distant mean structures, nonparametric estimators showed superiority, although post-stratified performed well for some distant conditions. Another advantage of the NP assumption is its higher plausibility when the post-stratifier is nominal rather than ordinal. In general, the doubly-robust estimators are better than their counterparts because the increase in variance is compensated by the reduction in bias. None of the estimators was superior in all 24 scenarios considered but the weight-smooth doubly-robust nonparametric was consistently among the best.

The weight-trimming estimator showed favorable results in the application due to the highest weight strata have a mean not substantially different from the other strata in both scenarios but is not recommended for routine use due to the lack of consistent good performance perceived in the simulation.

Among the model-based techniques, the XR assumption is more parsimonious (because only one fixed parameter is estimated) but at the expense of the strong exchangeability assumption which is questionable given the ordinal nature of the post-stratifying variable when a systematic relationship with the outcome of interest might be expected (Lazzeroni and Little, 1998). The LI and NP assumptions come into the rescue by adding parameters to the mean structure but paying the price in efficiency, once more a trade-off between robustness and efficiency is involved (Elliott and Little, 2000).

The empirical Bayes method, used under the model-based approaches, underestimates the posterior variance when fixing the parameters $\sigma^2, \tau^2, \tau_s^2$ to their restricted maximum likelihood estimates, ignoring part of the uncertainty. This explains partly the undercoverage of the weight-smoothing confidence intervals under some nonlinear schemes. Nevertheless, the discrepancy is not severe for many samples sizes found in practice (Zheng and Little, 2004). Full Bayesian analysis allows to incorporate the uncertainty in estimating the variance components. Alternatively, Lazzeroni and Little (1998) used a t-correction fraction to widen the confidence intervals by 20-30%, although Elliott and Little (2000) ruled out it because they find it overly conservative when H is small.

The methodology is better suited for observational surveys, i.e. nonprobability samples, sample designs which include intended subsampling, or when the interest is in subsamples, inter alia, where some normalized weights are large. When representative surveys are available, i.e. when the true underlying model is more simple, complex estimators are unnecessary because they come with the cost of intense computational requirements and possible loss of efficiency.

The performance of the methods relative to their robustness against deviations from the normal distribution assumption was not tested in the simulation study, besides only one post-stratifying variable was considered. As an extension, contaminated normal errors, more than one post-stratifier and estimation of other than mean, e.g. population regression coefficients, can be considered in further researches.

References

Chen, Q., Elliott, M. R., and Little, R. J. A. (2010). Bayesian penalized spline model-based inference for finite population proportion in unequal probability sampling. *Survey Methodology*, 36(1):23–34.

Elliott, M. R. and Little, R. J. A. (2000). Model-based alternatives to trimming survey weights. *Journal of Official Statistics*, 16(3):191–209.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2012). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9993.

Ghosh, M. and Meeden, G. (1986). Empirical bayes estimation in finite population sampling. *Journal of the American Statistical Association*, 81(396):1058–1062.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman & Hall.

Holt, D. and Smith, T. M. F. (1979). Post stratification. *Journal of the Royal Statistical Society*, Series A 142(1):33–46.

Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

Lazzeroni, L. and Little, R. (1998). Random-effects models for smoothing poststratification weights. *Journal of Official Statistics*, 14(1):61–78.

Little, R. (1991). Inference with survey weights. *Journal of Official Statistics*, 7(4):405–424.

Molenberghs, G. (2013). Survey methods. Slides of a lecture given at Universiteit Hasselt, Belgium.

Mukhopadhyay, P. (1996). *Inferential problems in survey sampling*. New Age International Publishers.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Development Core Team (2011). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-102.

Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In: Proceedings of the American Statistical Association, Survey Research Methods Section, 225-230.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Vandendijck, Y., Faes, C., and Hens, N. (2014). Doubly-robust weight smoothing modelling to smooth post-stratification weights for prevalence and trend estimation. *Working paper*.

Wood, S. (2006). *Generalized Additive Models: an introduction with R*. Chapman and Hall/CRC.

Zheng, H. and Little, R. J. A. (2004). Penalized spline nonparametric mixed models for inference about a finite population mean from two-stage samples. *Survey Methodology*, 30(2):209–218.

Appendix

Table 5: *Coverage of the nominal 95% confidence intervals for the simulation study based on 9 estimators (Est. ps=post-stratified, uw=unweighted, tr=weight-trimmed, ws.xr=weight-smoothing exchangeable, ws.li=weight-smoothing linear, ws.np=weight-smoothing nonparametric, wsdr.xr=doubly-robust exchangeable, wsdr.li=doubly-robust linear, wsdr.np=doubly-robust nonparametric) and 500 subsamples for all 24 scenarios. Sample sizes:* $n_1 = 500$ *and* $n_2 = 100$. *Some subsamples discarded due to not invertible matrices in ws variance calculation*

	Close						Distant					
	$\sigma = 1$		$\sigma = 5$		$\sigma = 10$		$\sigma = 1$		$\sigma = 5$		$\sigma = 10$	
Est.	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
ps	91.8	87.6	94.2	85.4	93.2	85.2	94.2	86.2	91.0	86.2	94.0	86.0
uw	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
tr	0.0	0.0	5.8	67.0	63.0	88.4	0.0	0.0	0.0	0.0	0.0	5.4
ws.xr	92.8	97.2	94.4	93.2	93.6	91.2	95.2	93.8	89.2	78.8	84.8	47.2
ws.li	92.4	96.6	92.4	87.4	87.2	82.0	95.2	93.6	87.6	75.6	81.4	70.8
ws.np	92.8	97.0	95.8	94.4	94.8	93.2	95.2	93.6	92.2	91.2	93.2	89.8
wsdr.xr	100.0	100.0	99.8	96.4	93.2	91.6	100.0	100.0	99.2	85.8	85.2	56.4
wsdr.li	100.0	100.0	97.8	93.6	91.8	90.2	100.0	100.0	94.4	82.4	84.4	79.4
wsdr.np	100.0	99.8	97.2	95.8	94.4	94.4	100.0	100.0	93.4	93.0	93.2	90.8
	Equal						Linear					
	$\sigma = 1$		$\sigma = 5$		$\sigma = 10$		$\sigma = 1$		$\sigma = 5$		$\sigma = 10$	
Est.	$n_1^{(1)}$	$n_2^{(7)}$	$n_1^{(2)}$	$n_2^{(8)}$	$n_1^{(3)}$	$n_2^{(9)}$	$n_1^{(4)}$	$n_2^{(10)}$	$n_1^{(5)}$	$n_2^{(11)}$	$n_1^{(6)}$	$n_2^{(12)}$
ps	93.9	83.5	93.2	86.0	93.6	86.1	93.9	83.8	92.7	85.0	94.1	86.1
uw	95.7	93.5	93.0	95.5	94.0	93.9	0.0	0.0	0.0	0.0	0.0	0.0
tr	95.7	92.7	93.6	92.8	93.8	93.1	0.0	0.0	0.0	0.0	0.0	19.0
ws.xr	96.1	94.3	93.8	43.2	94.6	94.7	96.2	94.3	93.5	90.6	89.3	75.5
ws.li	96.5	93.7	95.2	32.9	94.8	93.7	96.2	96.7	95.4	96.3	94.7	94.7
ws.np	95.9	93.7	94.8	94.4	93.4	93.5	96.2	96.1	94.2	95.2	94.7	94.3
wsdr.xr	88.0	86.6	86.8	87.2	87.1	85.9	100.0	100.0	100.0	96.5	94.3	81.0
wsdr.li	96.3	94.1	94.2	93.8	93.2	93.3	100.0	100.0	99.8	99.6	97.6	96.5
wsdr.np	95.3	94.1	94.2	93.8	93.4	94.1	100.0	100.0	98.6	99.2	97.6	97.1

(1): 8 samples discarded. (2): 1 sample discarded. (3): 3 samples discarded.
(4): 7 samples discarded. (5): 4 samples discarded. (6): 6 samples discarded.
(7): 22 samples discarded. (8): 14 samples discarded. (9): 10 samples discarded.
(10): 11 samples discarded. (11): 19 samples discarded. (12): 11 samples discarded.

Table 6: *Average length of the 95% confidence intervals for the simulation study based on 9 estimators (Est. ps=post-stratified, uw=unweighted, tr=weight-trimmed, ws.xr=weight-smoothing exchangeable, ws.li=weight-smoothing linear, ws.np=weight-smoothing nonparametric, wsdr.xr=doubly-robust exchangeable, wsdr.li=doubly-robust linear, wsdr.np=doubly-robust nonparametric) and 500 subsamples for all 24 scenarios. Sample sizes:* $n_1 = 500$ *and* $n_2 = 100$*. Some subsamples discarded due to not invertible matrices in ws variance calculation*

	Close						Distant					
	$\sigma = 1$		$\sigma = 5$		$\sigma = 10$		$\sigma = 1$		$\sigma = 5$		$\sigma = 10$	
Est.	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
ps	0.41	0.85	2.06	4.14	4.04	8.09	0.41	0.82	2.06	4.10	4.09	8.41
uw	0.17	0.39	0.85	1.93	1.70	3.88	0.17	0.39	0.85	1.95	1.71	3.88
tr	0.25	0.56	1.26	2.78	2.50	5.51	0.25	0.56	1.25	2.77	2.53	5.55
ws.xr	0.41	0.92	2.04	4.36	3.95	7.88	0.41	0.92	2.04	4.37	3.94	7.22
ws.li	0.41	0.92	2.02	4.23	3.84	7.66	0.41	0.92	2.02	4.21	3.81	7.48
ws.np	0.41	0.92	2.01	4.33	3.93	8.29	0.41	0.91	2.01	4.35	3.92	8.18
wsdr.xr	2.28	3.31	3.20	4.93	3.86	7.02	2.85	3.31	3.17	4.73	3.68	6.23
wsdr.li	1.42	2.46	2.40	4.84	3.96	8.01	1.42	2.44	2.39	4.65	3.85	7.60
wsdr.np	0.71	1.57	2.14	4.61	3.90	8.16	0.71	1.57	2.16	4.63	3.91	8.10
	Equal						Linear					
	$\sigma = 1$		$\sigma = 5$		$\sigma = 10$		$\sigma = 1$		$\sigma = 5$		$\sigma = 10$	
Est.	$n_1^{(1)}$	$n_2^{(7)}$	$n_1^{(2)}$	$n_2^{(8)}$	$n_1^{(3)}$	$n_2^{(9)}$	$n_1^{(4)}$	$n_2^{(10)}$	$n_1^{(5)}$	$n_2^{(11)}$	$n_1^{(6)}$	$n_2^{(12)}$
ps	0.41	0.80	2.04	4.25	4.12	8.44	0.41	0.85	2.05	4.17	4.10	8.54
uw	0.17	0.39	0.85	1.96	1.71	3.88	0.17	0.39	0.85	1.95	1.70	3.88
tr	0.25	0.56	1.26	2.80	2.52	5.53	0.25	0.56	1.26	2.77	2.53	5.57
ws.xr	0.19	0.42	0.95	0.56	1.91	4.28	0.41	0.92	2.05	4.46	4.00	8.11
ws.li	0.31	0.69	1.55	0.76	3.11	6.91	0.31	0.69	1.55	3.46	3.10	6.89
ws.np	0.32	0.71	1.59	3.57	3.19	7.08	0.32	0.71	1.59	3.54	3.19	7.07
wsdr.xr	0.18	0.41	0.92	2.10	1.84	4.18	3.49	4.02	3.80	5.49	4.38	7.53
wsdr.li	0.31	0.69	1.55	3.47	3.10	6.93	1.20	2.33	2.43	5.08	3.96	7.83
wsdr.np	0.33	0.73	1.65	3.72	3.32	7.38	1.18	2.44	2.33	5.19	4.25	8.79

(1): 8 samples discarded. (2): 1 sample discarded. (3): 3 samples discarded.
(4): 7 samples discarded. (5): 4 samples discarded. (6): 6 samples discarded.
(7): 22 samples discarded. (8): 14 samples discarded. (9): 10 samples discarded.
(10): 11 samples discarded. (11): 19 samples discarded. (12): 11 samples discarded.

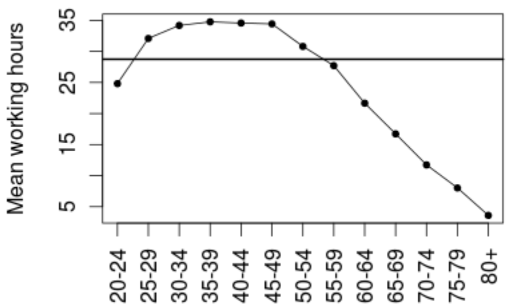

Figure 6: *2008 Colombian Quality of Life Survey. Mean age profile of weekly working hours (older than 20 years) and overall mean as horizontal line*

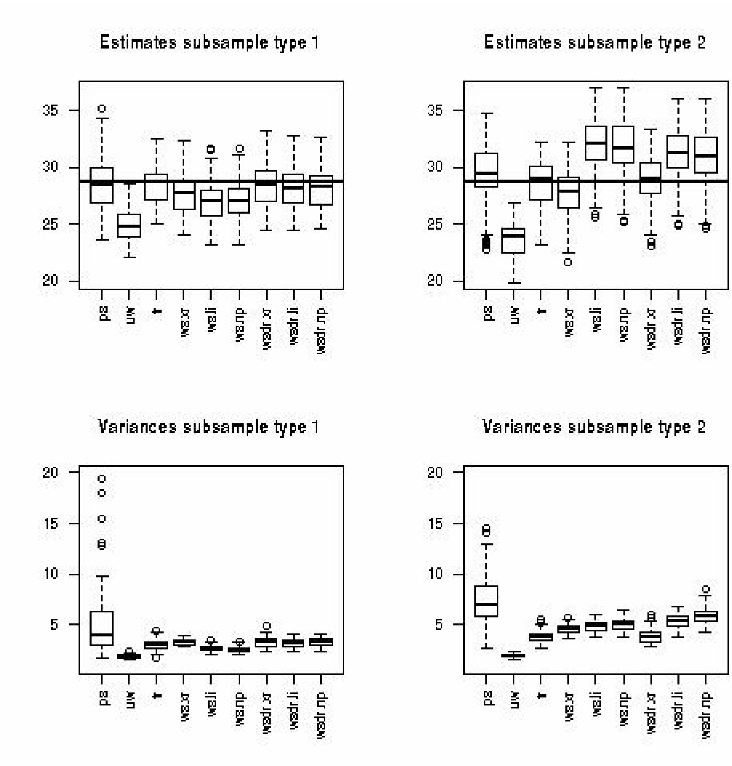

Figure 7: *2008 Colombian Quality of Life Survey. Estimates and variances of weekly working hours (older than 20 years) for the 100 subsamples of each type. Overall mean as horizontal line. Subsample type 1: 50-59 age groups highly underrepresented. Subsample type 2: 20-29 age groups highly underrepresented*

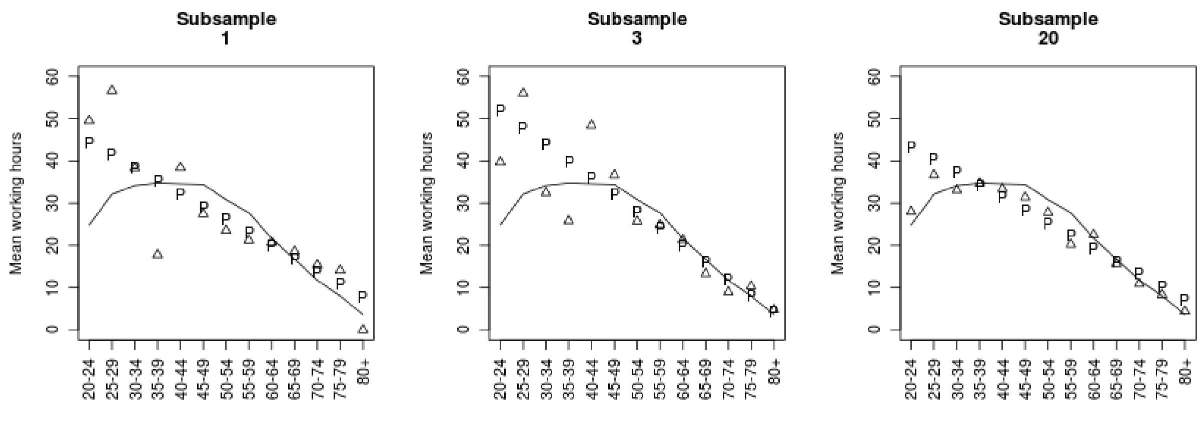

Figure 8: *Colombian Quality of Life Survey. Mean weekly working hours (older than 20 years) stratified by 13 age intervals of 5 years. True means (from survey) solid lines, subsample means triangles and predicted means by the nonparametric model 'P'*

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Doubly-robust weight smoothing models to smooth post-stratification weights in case of a Gaussian survey outcome

Richting: **Master of Statistics-Biostatistics**
Jaar: **2014**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Reyes Sierra, Adriana Rocio

Datum: **23/09/2014**