Jae Kwang Kim and Chris J. Skinner

Weighting in survey analysis under informative sampling

Article (Accepted version) (Refereed)

Original citation:
Kim, Jae Kwang and Skinner, Chris J. (2013) Weighting in survey analysis under informative sampling. Biometrika, 100 (2). pp. 385-398. ISSN 0006-3444

DOI: 10.1093/biomet/ass085

Weighting in survey analysis under informative sampling

Jae Kwang Kim
Department of Statistics
Iowa State University
Ames, Iowa 50011, U.S.A.

Chris Skinner
Department of Statistics
London School of Economics and Political Science
London, WC2A 2AE, U.K.

May 14, 2013

Abstract

Sampling related to the outcome variable of a regression analysis conditional on covariates is called informative sampling and may lead to bias in ordinary least squares estimation. Weighting by the reciprocal of the inclusion probability approximately removes such bias but may inflate variance. This paper investigates two ways of modifying such weights to improve efficiency while retaining consistency. One approach is to multiply the inverse probability weights by functions of the covariates. The second is to smooth the weights given values of the outcome variable and covariates. Optimal ways of constructing weights by these two approaches are explored. Both approaches require the fitting of auxiliary weight models. The asymptotic properties of the resulting estimators are investigated and linearisation variance estimators are obtained. The approach is extended to pseudo maximum likelihood estimation for generalized linear models. The properties of the different weighted estimators are compared in a limited simulation study. The robustness of the estimators to misspecification of the auxiliary weight model or of the regression model of interest is discussed.

Keywords: Complex sampling, Pseudo maximum likelihood estimation, Regression analysis, Sample likelihood.

1 Introduction

Survey data are often used to make inference about superpopulation models from which finite populations are assumed to be generated. When survey data are obtained from units selected with complex sample designs, the resulting analyses require different methods to those developed classically under random sampling assumptions. [?], [?] and [?] provide overviews of this topic. Regression is a key tool for statistical analyses that describe the structural relationship between survey variables. Survey weights are often used in regression analysis of survey data to ensure consistent estimation of parameters when sampling may be informative, that is when sample inclusion may be related to the outcome variable conditional on covariates [6.3]fuller09. Although weighting has this bias-correction advantage, it also brings the disadvantage of often leading to a loss of efficiency relative to an unweighted approach.

A number of authors have considered how the survey weights may be modified to improve efficiency while retaining their advantage of ensuring consistency under informative sampling. [?] showed that consistency is retained under any multiplication of the weights by a function of covariates and suggested how such a function might be chosen. [?] proposed a modification of the weights by a function of the covariates which minimized a prediction criterion. [?] discussed both approaches and extended their consideration of variance estimation. [?] extended their approach to generalized linear models. [6.3.2]fuller09 showed how efficiency could be maximized for a class of modified weights. [?] extended Fuller's approach in the context of a cross-national application. Addressing a rather different inferential problem, [?] proposed smoothing survey weights, with a modification which depends on the survey variables, in order to improve efficiency of descriptive estimation. In this paper, we show how the ideas of [?] and [?] may be integrated in the regression analysis of survey data by considering a weight modification which is a function of both the outcome variable and covariates and we develop and evaluate associated inferential methods, including variance estimation.

Alternative approaches, especially likelihood-based methods, have also been proposed for efficient inference about regression parameters in the presence of informative sampling. See [?], [?], [?] and [?] and references therein. However, in this paper we restrict attention to weighting methods, which are widely used in survey practice and for which various weight modifications are already familiar to survey data users.

2 Basic set-up

We consider the regression of a variable y on a vector of variables x. Let (x_i', y_i) denote the row vector of values of these variables for a unit with label i in the index set $U = \{1, \ldots, N\}$ of a finite population of size N and suppose that these values follow the regression model

$$y_i = x_i'\beta_0 + e_i, \tag{1}$$

where $E(e_i \mid x_i) = 0$. We assume a probability sampling design, where inclusion in the sample is represented by the indicator variables $I_i (i = 1, \ldots, N)$, where $I_i = 1$ if unit i is included in the sample and $I_i = 0$ otherwise and $\pi_i = (I_i = 1 \mid i)$ is the first-order inclusion probability. Then the ordinary least squares estimator of β_0 solves

$$\sum_{i=1}^{N} I_i \left(y_i - x_i'\beta\right) x_i = 0 \tag{2}$$

for β, and this estimator will generally be biased unless sampling is non-informative, that is I_i and y_i are pairwise independent conditional on x_i,

$$(I_i = 1 \mid y_i, x_i) = (I_i = 1 \mid x_i) \, . \tag{3}$$

In some circumstances it is possible to ensure that sampling is non-informative by including in x_i all of the design variables which explain variation in the π_i. Many surveys, however, exhibit variation in the π_i which, at least in part, is attributable to practical features of the survey implementation and cannot be wholly explained by variables which would be of scientific interest as covariates in the model. Thus, it is more realistic to write $\pi_i = \pi_i(X, Z)$ as a function of the population values $X = (x_1, \ldots, x_N)$ and $Z = (z_1, \ldots, z_N)$ of both x_i and a vector of design variables z_i, often unobserved, which may induce informative sampling via residual association between y_i and z_i given x_i. In such settings, the use of the design weights $d_i = \pi_i^{-1}$ in the weighted least squares estimator $\hat{\beta}_d$ which solves

$$\sum_{i=1}^{N} I_i d_i \left(y_i - x_i'\beta\right) x_i = 0 \tag{4}$$

for β is a standard approach to achieving consistent estimation of β_0.

3 Proposed weighting method

3.1 Introduction

We consider the class of weighted estimators $\hat{\beta}_w$ solving

$$\sum_{i=1}^{N} I_i w_i \left(y_i - x_i'\beta\right) x_i = 0, \tag{5}$$

for β, that is the solution of (4) with d_i replaced by a modified weight denoted w_i. We aim to choose w_i so that $\hat{\beta}_w$ has minimum asymptotic variance subject to being consistent for β_0. A sufficient condition for consistency is that w_i obeys

$$E\left(I_i w_i e_i \mid x_i\right) = 0, \tag{6}$$

and we shall restrict attention to the class of estimators $\hat{\beta}_w$ for which w_i meets this condition. Note that $E(.)$ and (.) in this paper will generally denote expectation with respect to both the model in (1) and the probability sampling scheme which is the source of randomness in the I_i. Moments with respect to just one of these distributions will be represented by appropriate conditioning. More details of the asymptotic properties of $\hat{\beta}_w$ are in the Appendix. The design-weighted estimator solving (4) is in the class of estimators $\hat{\beta}_w$ obeying (6) since

$$E\left(I_i d_i e_i \mid x_i\right) = E\left\{E\left(I_i \mid y_i, X, Z\right) d_i e_i \mid x_i\right\} = E\left(e_i \mid x_i\right) = 0,$$

by assumption and using the fact that $\pi_i = \pi_i(X, Z) = E\left(I_i \mid y_i, X, Z\right)$. The asymptotic variance of $\hat{\beta}_w$ may be expressed as

$$N^{-2} M_{xx\pi w,N}^{-1} \left(\hat{T} \mid X\right) M_{xx\pi w,N}^{-1} \tag{7}$$

where $M_{xx\pi w,N} = N^{-1} \sum_{i=1}^{N} \pi_i w_i x_i x_i'$ and $\hat{T} = \sum_{i=1}^{N} I_i w_i e_i x_i$.

Expression (7) may be decomposed into two parts by writing:

$$\left(\hat{T} \mid X\right) = \left\{E\left(\hat{T} \mid Y, X, I\right) \mid X\right\} + E\left\{\left(\hat{T} \mid Y, X, I\right) \mid X\right\}. \qquad (8)$$

where $Y = (y_1, \ldots, y_N)$ and $I = (I_1, \ldots, I_N)$. In the next subsection, we shall discuss how the second component of this expression may be removed by taking w_i as a smoothed version of d_i. In the following subsection we shall discuss the use of further weight modification to minimize the first component of this expression.

3.2 Weight smoothing

Before discussing the estimation of regression parameters, we first consider the simple case of estimating $\theta = E(Y)$, which is either the finite population mean of a single variable y_i or its model expectation. Following [**?**], the smoothed weight is defined as $\tilde{d}_i = E(d_i \mid y_i, I_i = 1)$. Equivalently, using identity (2.5a) from [**?**], we may write $\tilde{d}_i = \tilde{\pi}_i^{-1}$, where $\tilde{\pi}_i = E(\pi_i \mid y_i)$ is the conditional expectation of π_i given y_i. Let $\hat{\theta}_{\mathrm{HT}} = N^{-1} \sum_{i=1}^{N} I_i d_i y_i$ be the Horvitz–Thompson estimator of θ and let $\tilde{\theta}_{\mathrm{SHT}} = N^{-1} \sum_{i=1}^{N} I_i \tilde{d}_i y_i$ be the smoothed Horvitz–Thompson estimator that uses $\tilde{d}_i$. We shall distinguish in notation between $\hat{}$ for an estimator, such as $\hat{\theta}_{\mathrm{HT}}$, which depends only on observed data and $\tilde{}$, such as for $\tilde{\theta}_{\mathrm{SHT}}$, which depends also on a conditional expectation, such as $\tilde{d}_i$ which is not observed and the estimation of which we shall discuss in 4. The following lemma summarizes basic properties of the smoothed estimator presented in [**?**].

Lemma 3.1

The smoothed Horvitz–Thompson estimator $\tilde{\theta}_{\mathrm{SHT}} = N^{-1} \sum_{i=1}^{N} I_i \tilde{d}_i y_i$ *using*

$\tilde{d}_i = E(d_i \mid y_i, I_i = 1)$ *satisfies*

$$E\left(\tilde{\theta}_{\mathrm{SHT}}\right) = \theta \tag{9}$$

and

$$\mathrm{var}\left(\tilde{\theta}_{\mathrm{SHT}}\right) \leq \mathrm{var}\left(\hat{\theta}_{\mathrm{HT}}\right). \tag{10}$$

Proof. Result (9) is easy to establish. To show (10), note that

$$\left(\hat{\theta}_{\mathrm{HT}}\right) \quad \geq \quad \left\{E\left(\hat{\theta}_{\mathrm{HT}} \mid Y, I\right)\right\},$$

Then, (10) follows, provided $E(d_i \mid Y, I) = E(d_i \mid y_i, I_i)$, because

$$\begin{aligned} E\left(\hat{\theta}_{\mathrm{HT}} \mid Y, I\right) &= N^{-1} \sum_{i=1}^{N} I_i E(d_i \mid y_i, I_i = 1) y_i \\ &= N^{-1} \sum_{i=1}^{N} I_i \frac{1}{E(I_i \mid y_i)} y_i = \tilde{\theta}_{\mathrm{SHT}}. \end{aligned}$$

■

We now extend this smoothing idea to estimation in the regression model. We propose to replace the design weight by the smoothed weight $\tilde{d}_i = E\left(d_i \mid x_i, y_i, I_i = 1\right)$. We condition on both x_i and y_i in order to ensure that the consistency condition in (6) holds. This is the case since

$$E\left(I_i d_i e_i \mid x_i\right) = E\left\{I_i E\left(d_i \mid y_i, x_i, I_i = 1\right) e_i \mid x_i\right\} = E(I_i \tilde{d}_i e_i \mid x_i).$$

The resulting estimator $\tilde{\beta}_{Sd}$ of β_0, obtained by using $\tilde{d}_i$ in place of w_i in (5) is therefore consistent. Moreover, this choice of w_i removes the second term of (8) since $\tilde{d}_i$ is a function of x_i, y_i and I_i. Thus, by the same argument used to obtain (10), $\tilde{\beta}_{Sd}$ is more efficient than $\hat{\beta}_d$. The smoothed weight $\tilde{d}_i$ is not, however, observed and its estimation is discussed in 4. The weight $\tilde{d}_i$ was also derived, in the form $\tilde{d}_i = \{\mathrm{pr}(I_i = 1 \mid y_i, x_i)\}^{-1}$, using an empirical likelihood argument by [Equation 3.33]pfeffermann11.

3.3 Weight optimization

We initially leave aside the use of weight smoothing and consider the class of estimators $\hat{\beta}_{dq}$ with $w_i = d_i q_i$, where $q_i = q(x_i)$ is an arbitrary function of x_i. Condition (6) holds regardless of the choice of function $q(\cdot)$ and so the corresponding weighted least squares estimator $\hat{\beta}_{dq}$ is consistent for β_0 for any such function, with respect to the joint distribution induced by the model and the probability sampling scheme. [Appendix A]magee98 proves consistency of this estimator with respect to the conditional distribution given the realized sample. We should like to identify an estimator within the class of estimators of the form $\hat{\beta}_{dq}$ which has minimum variance. In order to obtain a practical solution, we shall approximate the variance to be minimized, since this will not affect consistency nor the validity of inferences using the chosen estimator. We begin with the expression for the asymptotic variance in (7), with w_i set equal to $d_i q_i$, and make the approximation that the variance of the sum $\hat{T}$ given X in the central expression is equal to the sum of the variances of the terms terms $I_i d_i q_i e_i x_i$ given x_i , that is we shall act as if the $I_i d_i e_i$ are independent given the x_i. Perhaps the most obvious reason for questioning this assumption will occur when the survey is clustered, but we do not explore here the departure from optimality arising from possible intra-cluster correlation of $I_i d_i e_i$. In the absence of clustering, an assumption of independence of the e_i between units will often be made in many survey settings. Independence of the $d_i I_i$ would occur under Poisson sampling with $\pi_i = \pi_i(z_i)$ and independent z_i, as considered by [p. 359]fuller09 in deriving a similar approximately efficient estimator. In the simulation study described in 7 we compared the relative efficiency of alternative estimators under Poisson sampling and a without replacement scheme with the same inclusion probabilities and found very little difference. Thus, we suggest that the departure from optimality arising from this approximation will often be small for single stage sampling schemes which arise in practice. The variance of $I_i d_i q_i e_i x_i$ given x_i may be expressed as

$$\begin{aligned}
(I_i d_i q_i e_i x_i \mid x_i) &= E\left\{(I_i d_i q_i e_i x_i \mid y_i, X, Z) \mid x_i\right\} + \left\{E\left(I_i d_i e_i q_i x_i \mid y_i, X, Z\right) \mid x_i\right\} \\
&= E\left\{(d_i - 1) e_i^2 q_i^2 x_i x_i' \mid x_i\right\} + (e_i q_i x_i \mid x_i) \\
&= E\left(d_i e_i^2 \mid x_i\right) q_i^2 x_i x_i'.
\end{aligned}$$

Writing $E(d_i {e_i}^2 \mid x_i) = v_i$, the asymptotic variance of $\hat{\beta}_{dq}$ under our approximation can be expressed as

$$\left(\sum_{i=1}^{N} q_i x_i x_i'\right)^{-1} \sum_{i=1}^{N} v_i q_i^2 x_i x_i' \left(\sum_{i=1}^{N} q_i x_i x_i'\right)^{-1}.$$

Thus, the choice of $q_i^* = v_i^{-1} = E(d_i e_i^2 \mid x_i)^{-1}$ will minimize the variance of any linear combination of the elements of $\hat{\beta}_{dq}$, which is consistent with the suggestion of [Ch. 6]fuller09 for scalar x.

Let us now consider modifying the smoothed weight in a similar way using $w_i = \tilde{d}_i q(x_i)$, where again $q(x_i)$ is a function of x_i. The corresponding estimator $\tilde{\beta}_{\mathrm{S}dq}$ can be expressed as the solution to

$$\tilde{U}_{\mathrm{S}dq}(\beta) \equiv \sum_{i=1}^{N} I_i \tilde{d}_i q(x_i) \left(y_i - x_i'\beta\right) x_i = 0 \tag{11}$$

for β. The weight $\tilde{d}_i q_i$ can be shown to obey the consistency condition (6) by combining the arguments used to show that each of $\tilde{d}_i$ and $d_i q_i$ obey this condition. The estimator $\tilde{\beta}_{\mathrm{S}dq}$ is consistent for β_0 regardless of the choice of function $q(\cdot)$. By a similar argument to that used for $w_i = d_i q_i$, the optimal estimator $\tilde{\beta}_{\mathrm{S}dq^*}$ can be obtained by using $q_i^* = E(\tilde{d}_i e_i^2 \mid x_i)^{-1}$. This is our proposed estimator, subject to the need to estimate q_i^*, which is considered in 4.

For comparison, we also consider the semi-parametric method proposed by [**?**], which is a particular case of an estimator of form $\hat{\beta}_{dq}$ with $q_i = E(d_i \mid x_i, I_i = 1)^{-1}$. The implied estimator $\tilde{\beta}_{\mathrm{PS}}$ of β_0 is the solution of

$$\tilde{U}_{\mathrm{PS}}(\beta) = \sum_{i=1}^{N} I_i \frac{d_i}{E_s(d_i \mid x_i)} (y_i - x_i'\beta)\, x_i = 0, \tag{12}$$

where $E_s\,(d_i \mid x_i) = E\,(d_i \mid x_i, I_i = 1)$. Applying smoothing to (12) gives a particular version of (11). The resulting estimator $\tilde{\beta}_{\mathrm{SPS}}$ solves

$$\tilde{U}_{\mathrm{SPS}}(\beta) \equiv \sum_{i=1}^{N} I_i \frac{E_s\,(d_i \mid x_i, y_i)}{E_s\,(d_i \mid x_i)} (y_i - x_i'\beta)\, x_i = 0, \tag{13}$$

where $E_s\,(d_i \mid x_i, y_i) = E\,(d_i \mid x_i, y_i, I_i = 1)$. Models for $E_s\,(d_i \mid x_i, y_i)$ and $E_s\,(d_i \mid x_i)$ and methods for their estimation will be discussed in 4. Note that

$$E\left\{\tilde{U}_{\mathrm{PS}}(\beta) \mid X, Y, I\right\} = \tilde{U}_{\mathrm{SPS}}(\beta).$$

Thus, by the same argument as for (10), the solution to (13) is more efficient than the solution to (12). In particular, if the sampling design is non-informative in the sense that (3) holds then $E_s(d_i \mid x_i, y_i) = E_s(d_i \mid x_i)$ and $\tilde{\beta}_{\mathrm{SPS}}$ from (13) reduces to the unweighted least squares estimator in (2).

Remark 3.1 *The estimator $\tilde{\beta}_{\mathrm{SPS}}$ that solves (13) can be justified by a prediction argument rather than an efficiency argument. Let the parameter of* *interest, β_0, be defined as the unique minimizer of the population prediction mean squared error*

$$Q(\beta) = \int (y - x'\beta)^2 f(y \mid x)\, dy.$$

By

$$f(y \mid x, I = 1) = f(y \mid x) \frac{\mathrm{pr}(I = 1 \mid x, y)}{\mathrm{pr}(I = 1 \mid x)},$$

we can write

$$Q(\beta) = \int (y - x'\beta)^2 f(y \mid x, I = 1) \frac{\mathrm{pr}(I = 1 \mid x)}{\mathrm{pr}(I = 1 \mid x, y)} dy.$$

Thus, a consistent estimator of β_0 can be obtained by minimizing

$$Q_{\mathrm{SPS}}(\beta) = \sum_{I_i=1} (y_i - x_i'\beta)^2 \frac{E(\pi_i \mid x_i)}{E(\pi_i \mid x_i, y_i)}.$$

Using equality (2.5a) of [?], we have

$$E(\pi_i \mid x_i, y_i) = E(d_i \mid x_i, y_i, I_i = 1)^{-1}. \tag{14}$$

Thus, we can write

$$Q_{\mathrm{SPS}}(\beta) = \sum_{I_i=1} (y_i - x_i'\beta)^2 \frac{E_s(d_i \mid x_i, y_i)}{E_s(d_i \mid x_i)}, \tag{15}$$

and the solution to (13) is obtained by minimizing (15).

4 Auxiliary weight models

In order to apply the proposed estimator $\tilde{\beta}_{Sdq^*}$ we need to estimate $\tilde{d}_i$ and we propose to do this using an auxiliary model for $E(d \mid x, y, I = 1)$. If both x and y are categorical then a fully nonparametric model for $E(d \mid x, y, I = 1)$

can be used, that is we can partition the sample $A = \{i \in U \mid I_i = 1\}$ into G groups $A = A_1 \cup \cdots \cup A_G$ such that

$$E(d_i \mid x_i, y_i, I_i = 1) = \tilde{d}_g, \quad \text{if } i \in A_g$$

and (x_i, y_i) is constant for $i \in A_g$. In this case, we can estimate $\tilde{d}_g$ by the simple group mean of the d_i in A_g. Furthermore, $E_s(d_i \mid x_i)$ can be computed similarly in order to construct $\hat{\beta}_{SPS}$ from (13).

If x or y is continuous, we can specify a parametric model $\tilde{d}_i \equiv E(d_i \mid x_i, y_i, I_i = 1) \equiv \tilde{d}(x_i, y_i; \phi)$, indexed by an unknown parameter ϕ. For example, one may consider the following parametric model

$$\tilde{d}(x_i, y_i; \phi) = c + \exp(-\phi_1 x_i - \phi_2 y_i) \tag{16}$$

for some $\phi = (\phi_1, \phi_2)$, where c is assumed to be known. Note that c is the minimum value of the weight and we may often set $c = 1$. By (14), model (16) is equivalent to assuming the logistic model

$$Pr(I_i = 1 \mid x_i, y_i) = \frac{\exp(\phi_1 x_i + \phi_2 y_i)}{1 + c \exp(\phi_1 x_i + \phi_2 y_i)}. \tag{17}$$

In principle, models such as (16) may be checked with sample data, especially since the conditioning on $I = 1$ implies that the model applies to the sample, and the sensitivity of results to alternative well-fitting models could be investigated.

Given the specification of the model in (16), the parameter vector ϕ can be estimated by minimizing

$$\sum_{i=1}^{N} I_i \left\{ d_i - \tilde{d}(x_i, y_i; \phi) \right\}^2 \frac{1}{v_{1i}(\phi)}$$

for some $v_{1i}(\phi)$. The optimal choice of $v_{1i}(\phi)$ is $v_{1i}(\phi) = var(d_i \mid x_i, y)$ which requires additional assumption about the form of $var(d_i \mid x_i, y_i)$. Under the assumption that the conditional distribution of $(d_i - c)$ given x_i and y_i follows a log-normal distribution, we have $var(d_i \mid x_i, y_i) \propto \{\tilde{d}(x_i, y_i; \phi) - c\}^2$. In this case, the optimal estimating equation for ϕ is

$$\sum_{i=1}^{N} I_i \left\{ \frac{d_i - c}{\tilde{d}(x_i, y_i; \phi) - c} - 1 \right\} (x_i, y_i) = (0, 0). \tag{18}$$

Using the resulting estimator $\hat{\phi}$ we obtain $\hat{d}_i = \tilde{d}(x_i, y_i; \hat{\phi})$ as an estimimator of $\tilde{d}_i$. We can also estimate $q_i^* = 1/E\{\tilde{d}_i e_i^2 \mid x_i\}$ as follows.

1. Obtain consistent estimators of β and ϕ by solving (4) and (18), respectively.

2. Let $q_i^*(\phi, \beta) = 1/E\left\{\tilde{d}(x_i, y_i; \phi)(y - x_i\beta)^2 \mid x_i\right\}$. Using the current values $\hat{\phi}$ and $\hat{\beta}$ of the parameter estimates, the consistent estimator $\hat{q}_i^* = q_i^*(\hat{\phi}, \hat{\beta})$ of q_i^* can be expressed as

$$\hat{q}_i^* = \left[\int \tilde{d}(x_i, y; \hat{\phi})(y - x_i\hat{\beta})^2 f(y \mid x_i; \hat{\beta}) dy \right]^{-1}. \tag{19}$$

   and (19) can be computed using Monte Carlo sampling.

3. Use equation (11) to obtain $\hat{\beta}$. Go to Step 2. Continue until convergence.

We can use a similar Monte Carlo approach to compute an estimator of $E_s(d_i \mid x_i)$ from $\hat{d}_i$ and obtain the smoothed estimator $\hat{\beta}_{SPS}$ defined by (13). Alternatively, $E_s(d_i \mid x_i)$ can be obtained more directly by assuming a further parametric model, such as $E_s(d_i \mid x_i) \equiv \tilde{d}(x_i;\phi^*) = c + \exp(-\phi^* x)$ and estimating ϕ^* by solving

$$\sum_{i=1}^{N} I_i \left\{ \frac{d_i - c}{\tilde{d}(x_i;\phi^*) - c} - 1 \right\} x_i = 0. \tag{20}$$

5 Asymptotic properties

We now discuss asymptotic properties of the proposed estimator. We assume that (1) holds with $E(e_i \mid x_i) = 0$. We first consider an estimator obtained from (11) which takes the form

$$\tilde{\beta}_{Sdq} = \left(\sum_{i=1}^{N} I_i \tilde{d}_i x_i x_i' q_i \right)^{-1} \sum_{i=1}^{N} I_i \tilde{d}_i x_i y_i q_i \tag{21}$$

where $\tilde{d}_i$ is assumed to be known and $q_i = q(x_i)$. Now, writing

$$\left(\tilde{M}_{xxq}, \tilde{M}_{xyq}, \tilde{M}_{xeq} \right) = N^{-1} \sum_{i=1}^{N} I_i \tilde{d}_i q_i \left(x_i x_i', x_i y_i, x_i e_i \right)$$

and

$$\left(M_{xxq,N}, M_{xyq,N}, M_{xeq,N} \right) = N^{-1} \sum_{i=1}^{N} q_i \left(x_i x_i', x_i y_i, x_i e_i \right),$$

we have

$$\begin{aligned} E\left(\tilde{M}_{xxq}\right) &= E\left\{ N^{-1} \sum_{i=1}^{N} I_i E(d_i \mid x_i, y_i, I_i = 1) x_i x_i' q_i \right\} \\ &= E\left\{ E\left(N^{-1} \sum_{i=1}^{N} I_i d_i x_i x_i' q_i \mid X, Y, I \right) \right\} \\ &= E\left\{ E\left(N^{-1} \sum_{i=1}^{N} I_i d_i x_i x_i' q_i \mid X, Y, Z \right) \right\} \\ &= E\left(M_{xxq,N} \right), \end{aligned}$$

where $Z = (z_1, \cdots, z_N)$. Similarly, we have $E(\tilde{M}_{xeq}) = E(M_{xeq,N}) = 0$. Under regularity conditions, we have $var(\tilde{M}_{xxq}) = O(n^{-1})$ and so $\tilde{M}_{xxq} - M_{xxq,N} = O_p(n^{-1/2})$. Similarly, we have $\tilde{M}_{xeq} = O_p(n^{-1/2})$ and so

$$\begin{aligned} \tilde{\beta}_{Sdq} - \beta_0 &= \tilde{M}_{xxq}^{-1}\tilde{M}_{xeq} \\ &= M_{xxq,N}^{-1}\tilde{M}_{xeq} + O_p(n^{-1}). \end{aligned}$$

The asymptotic distribution of $\tilde{\beta}_{Sdq} - \beta_0$ is equal to the asymptotic distribution of $M_{xxq,N}^{-1}\tilde{M}_{xeq}$.

To consider variance estimation, note that we can write $\tilde{M}_{xeq} = N^{-1}\sum_{i=1}^{N} I_i d_i (\tilde{d}_i/d_i) u_i$, where $u_i = x_i e_i q_i$. Since $\tilde{d}_i$ is a fixed quantity conditional on X and Y, we have

$$\begin{aligned} var\left(\tilde{M}_{xeq}\right) &= var\left(\tilde{M}_{xeq} - M_{xeq,N}\right) + var\left(M_{xeq,N}\right) \\ &= E\left\{var\left(\tilde{M}_{xeq} - M_{xeq,N} \mid X, Y, Z\right)\right\} \\ &\quad + var\left\{E\left(\tilde{M}_{xeq} - M_{xeq,N} \mid X, Y, Z\right)\right\} + var\left(M_{xeq,N}\right). \end{aligned}$$

The first term is the sampling variance of $\tilde{M}_{xeq} = N^{-1}\sum_{i=1}^{N} I_i d_i \tilde{u}_i$ where $\tilde{u}_i = (\tilde{d}_i/d_i)u_i$ and can be easily estimated by applying a standard variance estimation formula for $\hat{\theta}_{HT} = N^{-1}\sum_{i=1}^{N} I_i d_i y_i$ with y_i replaced by $\tilde{u}_i = (\tilde{d}_i/d_i)u_i$. Since

$$E\left(\tilde{M}_{xeq} - M_{xeq,N} \mid X, Y, Z\right) = N^{-1}\sum_{i=1}^{N}\left(\tilde{u}_i - u_i\right)$$

the second term will be of order $O(N^{-1})$ and will be negligible if n/N is negligible. An unbiased estimator of the second term can be easily computed by

$$\hat{V}_2 = N^{-2}\sum_{i=1}^{N} I_i d_i \left(\frac{\tilde{d}_i}{d_i} - 1\right)^2 x_i x_i' \hat{e}_i^2 q_i^2, \tag{22}$$

where $\hat{e}_i = y_i - x_i'\hat{\beta}$. In addition, we need to estimate the third term, which is order $O(N^{-1})$ and will be negligible if n/N is negligible, and it is consistently estimated by

$$\hat{V}_3 = N^{-2} \sum_{i=1}^{N} I_i d_i x_i x_i' \hat{e}_i^2 q_i^2. \tag{23}$$

In practice, we use

$$\hat{\beta}_{Sdq} = \left(\sum_{i=1}^{N} I_i \hat{d}_i x_i x_i' \hat{q}_i \right)^{-1} \sum_{i=1}^{N} I_i \hat{d}_i x_i y_i \hat{q}_i, \tag{24}$$

where $\hat{d}_i = \tilde{d}(x_i, y_i; \hat{\phi})$ is a consistent estimator of $\tilde{d}_i$ and $\hat{q}_i = q(x_i; \hat{\alpha})$ is a consistent estimator of $q_i = q(x_i; \alpha)$. The estimator $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2)$ may be computed by solving an estimation equation $\tilde{U}(\phi) = 0$, such as (18). Without loss of generality, we may write

$$\tilde{U}(\phi) = N^{-1} \sum_{i=1}^{N} I_i \left(\frac{d_i - c}{\tilde{d}(x_i, y_i; \phi) - c} - 1 \right) W_i \zeta_i$$

for some $W_i = W(x_i, y_i; \phi)$, where $\zeta_i = \partial \tilde{d}(x_i, y_i; \phi)/\partial \phi$. Writing

$$\left(\hat{M}_{xxq}, \hat{M}_{xyq}, \hat{M}_{xeq} \right) = N^{-1} \sum_{i=1}^{N} I_i \hat{d}_i q_i \left(x_i x_i', x_i y_i, x_i e_i \right),$$

we have, under some regularity conditions,

$$\hat{M}_{xxq} = \tilde{M}_{xxq} + O_p(n^{-1/2}).$$

Also, by first order Taylor linearization,

$$\begin{aligned} \hat{M}_{xeq} &= \tilde{M}_{xeq} + k'\tilde{U}(\phi) + O_p(n^{-1}) \qquad (25) \\ &= N^{-1} \sum_{i=1}^{N} I_i \tilde{d}_i \left\{ q_i x_i e_i + k' \tilde{d}_i^{-1} \left(\frac{d_i - c}{\tilde{d}_i - c} - 1 \right) W_i \zeta_i \right\} + O_p(n^{-1}) \end{aligned}$$

where

$$k' = -E\left\{\frac{\partial}{\partial \phi'}\tilde{M}_{xeq}\right\}\left[E\left\{\frac{\partial}{\partial \phi'}\tilde{U}(\phi)\right\}\right]^{-1}.$$

Note that

$$\begin{aligned}\frac{\partial}{\partial \phi'}\tilde{M}_{xeq} &= N^{-1}\sum_{i=1}^{N} I_i q_i x_i e_i \zeta_i' \\ \frac{\partial}{\partial \phi'}\tilde{U}(\phi) &= -N^{-1}\sum_{i=1}^{N} I_i (d_i - c)(\tilde{d}_i - c)^{-2} W_i \zeta_i \zeta_i'.\end{aligned}$$

Thus, we can estimate k from the sample by

$$\hat{k}' = \left\{\sum_{i=1}^{N} I_i \hat{q}_i x_i \hat{e}_i \hat{\zeta}_i'\right\}\left\{\sum_{i=1}^{N} I_i (d_i - c)(\hat{d}_i - c)^{-2}\hat{W}_i \hat{\zeta}_i \hat{\zeta}_i'\right\}^{-1}, \tag{26}$$

where $\hat{e}_i, \hat{\zeta}_i, \hat{q}_i, \hat{d}_i$ and $\hat{W}_i$ are computed by evaluating $e_i = y_i - x_i'\beta$, $\zeta_i = \zeta_i(\phi)$, $q_i = q_i(\alpha)$, $\tilde{d}_i = \tilde{d}_i(\phi)$ and $W_i = W_i(\phi)$ at $(\beta, \phi, \alpha) = (\hat{\beta}, \hat{\phi}, \hat{\alpha})$.

To summarize, a consistent variance estimator of $\hat{\beta}_{Sdq}$ in (24) can be obtained by

$$\hat{V}(\hat{\beta}_{Sdq}) = \hat{M}_{xxq}^{-1}\left\{\hat{v}\left(\bar{b}_{HT}\right) + \hat{V}_2 + \hat{V}_3\right\}\hat{M}_{xxq}^{-1} \tag{27}$$

where $\hat{v}\left(\bar{b}_{HT}\right)$ is an estimator of the design variance of $\bar{b}_{HT} = N^{-1}\sum_{i=1}^{N} I_i d_i b_i$ calculated with $\hat{b}_i = (\hat{d}_i/d_i)\left[x_i q_i \hat{e}_i + \hat{k}'\hat{d}_i^{-1}\{(d_i - c)/(\hat{d}_i - c) - 1\}W_i \hat{\zeta}_i\right]$ and $\hat{e}_i = y_i - x_i'\hat{\beta}_{Sdq}$, and $\hat{V}_2$ and $\hat{V}_3$ are defined in (22) and (23), respectively.

6 Pseudo maximum likelihood estimation

We now extend the proposed method to a generalized linear model setting, where the finite population values $(x_i, y_i); i = 1, \cdots, N$ are generated independently with probability density $f(y_i \mid x_i; \theta)h(x_i)$ and the conditional distribution of y_i given x_i is in the exponential family

$$f\left(y_i \mid x_i\right) = \exp\left\{\frac{y_i \gamma_i - b(\gamma_i)}{\tau^2} - c(y_i, \tau)\right\},$$

where γ_i is the canonical parameter. By the theory of the generalized linear models, we have

$$E\left(y_i \mid x_i\right) \equiv \mu_i = \partial b(\gamma_i)/\partial \gamma_i$$

and we assume that $g(\mu_i) = x_i'\beta_0$. We are interested in estimating β_0.

Under this setup, the pseudo maximum likelihood estimator (Skinner, 1989) for β_0 is the solution of

$$\sum_{I_i=1} d_i S(\beta; x_i, y_i) = 0 \tag{28}$$

where

$$S(\beta; x_i, y_i) = \frac{1}{\tau^2}\left(y_i - \mu_i\right)\left\{v(\mu_i) g_\mu(\mu_i)\right\}^{-1} x_i$$

and $g_\mu(\mu_i) = \partial g(\mu_i)/\partial \mu_i$. To derive an optimal estimator under informative sampling, consider a class of estimators of β that solves

$$\sum_{I_i=1} d_i S(\beta; x_i, y_i) q(x_i) = 0 \tag{29}$$

where $q(x_i)$ is to be determined. The solution to (29) is consistent regardless of the choice of $q(x_i)$ because $E\{S(\beta; x_i, y_i) \mid x_i\} = 0$. By Taylor linearization, the solution to (29) satisfies

$$\hat{\beta} = \beta_0 + \left\{\sum_{i=1}^{N}\left(\frac{\partial \mu_i}{\partial \beta}\right)\left\{v(\mu_i) g_\mu(\mu_i)\right\}^{-1} x_i' q(x_i)\right\}^{-1} \sum_{I_i=1} d_i e_i \left\{v(\mu_i) g_\mu(\mu_i)\right\}^{-1} x_i q(x_i).$$

where $e_i = y_i - \mu_i$. Since $\partial \mu_i/\partial \beta = \{g_\mu(\mu_i)\}^{-1} x_i$, the asymptotic variance of $\hat{\beta}$ obtained from (29) is

$$J_q^{-1} var\left\{\sum_{I_i=1} d_i e_i \left\{v(\mu_i) g_\mu(\mu_i)\right\}^{-1} x_i q(x_i)\right\}(J_q')^{-1} \tag{30}$$

where $J_q = \sum_{i=1}^{N} \left\{ v(\mu_i) g_\mu^2(\mu_i) \right\}^{-1} x_i x_i' q(x_i)$. Using the same argument as in Section 3.2, we have

$$var \left\{ I_i d_i e_i \left\{ v(\mu_i) g_\mu(\mu_i) \right\}^{-1} q_i x_i \right\} = E\left(d_i e_i^2 \mid x_i \right) q_i^2 \left\{ v(\mu_i) g_\mu(\mu_i) \right\}^{-2} x_i x_i'$$

and the optimal choice that minimizes (30), assuming independence between the terms in the summation in this expression, is

$$q_i^* = v(\mu_i) \left\{ E(d_i e_i^2 \mid x_i) \right\}^{-1},$$

Because $v(\mu_i) = E(e_i^2 \mid x_i)$, we can write

$$q_i^* = \frac{E(e_i^2 \mid x_i)}{E(d_i e_i^2 \mid x_i)}.$$

If the smoothed weights $\tilde{d}_i$ are used in (28) instead of the original weights then the optimal choice becomes

$$\tilde{q}_i^* = \frac{E(e_i^2 \mid x_i)}{E(\tilde{d}_i e_i^2 \mid x_i)}. \tag{31}$$

To compute (31), we need to evaluate $E(\tilde{d}_i e_i^2 \mid x_i)$ which depends on the conditional distribution of y_i given x_i. An EM-type algorithm can be obtained as

$$\hat{\beta}^{(t+1)} \longleftarrow \sum_{I_i=1} \tilde{d}_i S(\beta; x_i, y_i) q_i^*(\hat{\beta}^{(t)}) = 0, \tag{32}$$

where $q_i^*(\hat{\beta}^{(t)})$ is the value of (31) evaluated at $\beta = \hat{\beta}^{(t)}$. In practice, we use $\hat{d}_i$ instead of $\tilde{d}_i$ in (32).

Example 6.1 *Assume that y_i follows from a Bernoulli distribution with mean $p(x_i;\beta) = \left\{1 + \exp(-x_i'\beta)\right\}^{-1}$. The pseudo maximum likelihood estimator of β can be obtained by solving*

$$\sum_{I_i=1} d_i \left\{ y_i - p(x_i;\beta) \right\} x_i = 0.$$

In this case, $g(\mu_i) = logit(\mu_i) = x_i'\beta$ *and so* $g_\mu(\mu_i) = \{\mu_i(1-\mu_i)\}^{-1}$*. Note that we can write*

$$\begin{aligned} E\left\{\tilde{d}(x_i, y_i)e_i^2 \mid x_i\right\} &= \tilde{d}(x_i, 1)\,(1-p_i)^2\, p_i + \tilde{d}(x_i, 0)\,(0-p_i)^2\,(1-p_i) \\ &= p_i(1-p_i)\left\{\tilde{d}(x_i, 1)(1-p_i) + \tilde{d}(x_i, 0)p_i\right\}, \end{aligned}$$

where $p_i = p(x_i, \beta)$*. Thus, the EM algorithm in (32) can be implemented by solving*

$$\sum_{I_i=1} w_i(\hat{\beta}^{(t)})\,\{y_i - p(x_i;\beta)\}\, x_i = 0$$

for β*, where*

$$w_i(\hat{\beta}) = \tilde{d}(x_i, y_i)\left\{\tilde{d}(x_i, 1)(1-\hat{p}_i) + \tilde{d}(x_i, 0)\hat{p}_i\right\}^{-1}$$

and $\hat{p}_i = p(x_i; \hat{\beta})$*.*

For variance estimation, note that the solution $\hat{\beta}_{Sdq}$ can be obtained by solving

$$\sum_{I_i=1} \tilde{d}_i S(\beta; x_i, y_i) q(x_i) = 0.$$

The Pfeffermann-Sverchkov-type estimator uses $q(x_i) = 1/E(\tilde{d}_i \mid x_i, I_i = 1)$. Using the argument in Section 5, we can show that a consistent estimator of the variance of $\hat{\beta}_{Sdq}$ is

$$\hat{v}(\hat{\beta}_{Sdq}) = \hat{M}_{hhq}^{-1}\left\{\hat{v}\left(\bar{b}_{HT}\right) + \hat{V}_2 + \hat{V}_3\right\}\hat{M}_{hhq}^{-1} \tag{33}$$

where

$$\hat{M}_{hhq} = N^{-1}\sum_{I_i=1} \hat{d}_i H(\hat{\beta}_{Sdq}; x_i, y_i) q(x_i),$$

$H(\beta; x_i, y_i) = -\partial S(\beta; x_i, y_i)/\partial\beta$, $\hat{v}\left(\bar{b}_{HT}\right)$ is an estimator of the design variance of $\bar{b}_{HT} = N^{-1}\sum_{i=1}^N I_i d_i b_i$ calculated with $\hat{b}_i = (\hat{d}_i/d_i)[q_i\hat{s}_i + \hat{k}_s'\hat{d}_i^{-1}\{(d_i -$

$c)/(\hat{d}_i - c) - 1\}W_i\hat{\zeta}_i]$ and $\hat{k}_s$ is computed by (26) with $x_i\hat{e}_i$ replaced by $\hat{s}_i = S(\hat{\beta}_{Sdq}; x_i, y_i)$, and $\hat{V}_2$ and $\hat{V}_3$ are computed by (22) and (23), respectively, with $x_i\hat{e}_i$ replaced by $\hat{s}_i$.

7 Simulation studies

7.1 Simulation 1

To compare the performance of the estimators, we performed two limited simulation studies. In the first simulation, we repeatedly generated $B = 2,000$ Monte Carlo samples of finite populations of size $N = 10,000$ with values (x_i, y_i, z_i, π_i), where $x_i \sim U(0, 2)$,

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

$(\beta_0, \beta_1) = (-2, 1)$, $e_i \sim N(0, 0.5^2)$, $z_i \sim N(1 + y_i, 0.8^2)$ and $\pi_i = \{1 + \exp(3.5 - 0.5z_i)\}$. From each finite population, a sample was drawn by Poisson sampling where the sample indicator I_i follows a Bernoulli(π_i) distribution. The average sample size in this situation is about 335.

In addition to the above model, called Model A, we generated another set of Monte Carlo samples from a different model, called Model B, where the simulation setup is the same as for Model A except that the e_i were generated from $e_i \sim N(0, 0.5^2 x_i^2)$, thus allowing for heteroscedasticity.

From each sample, we computed five estimators of (β_0, β_1) using (5) with the following alternative choices of weights w_i:

1. design weights d_i;
2. Pfeffermann-Sverchkov weights, as in (12), using the semi-parametric method of Pfeffermann and Sverchkov (1999);

3. smoothed design weights, as in (18) and (21) with $q_i = 1$;

4. smoothed Pfeffermann-Sverchkov weights, as in (21) with $q_i = 1/E(d_i \mid x_i, I_i = 1)$, where $E(d_i \mid x_i, I_i = 1) = 1 + \exp(-\phi_0 - \phi_1 x_i)$ was estimated by solving (20);

5. smoothed optimal weights, as in (32) with optimal q_i^* in (31) under normality of $f(y \mid x)$.

To estimate the smoothed weights, we used

$$E(d \mid x, y, I = 1) = 1 + \exp\left(-\phi_0 - \phi_1 x - \phi_2 y\right). \quad (34)$$

with $(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2)$ computed by (18).

We consider two parameters: β_0 and β_1. Table 1 presents the Monte Carlo biases and standard errors of the five point estimators considered. All the point estimators considered are found to be nearly unbiased. In term of efficiency, the estimators using smoothed weights are generally more efficient than the estimators using the original weights. The smoothed Pfeffermann-Sverkov estimator and smoothed optimal estimator perform similarly under Model A, since the homoscedasticity in the error variance makes the former estimator nearly optimal. On the other hand, under model B, the smoothed Pfeffermann-Sverchkov estimator is markedly inferior because it does not take into account unequal variances. In addition to point estimation, we have also computed variance estimators using the linearization method discussed in Section 5. All the variance estimators show negligible relative biases, less than 5% of the absolute values.

Table 1. Properties of alternative weighted point estimators in Simulation 1, based on 2,000 Monte Carlo samples

		Model A		Model B	
Parameter	Weight	Bias	S.E.	Bias	S.E.
Intercept	Design	0.00	0.0720	0.00	0.0445
	Pfeffermann-Sverchkov	0.00	0.0704	0.00	0.0617
	Smoothed design	0.00	0.0683	0.00	0.0413
	Smoothed Pfeffermann-Sverkov	0.00	0.0668	0.00	0.0573
	Smoothed optimal	0.00	0.0668	0.00	0.0375
Slope	Design	0.00	0.0574	0.00	0.0639
	Pfeffermann-Sverchkov	0.00	0.0554	0.00	0.0766
	Smoothed design	0.00	0.0538	0.00	0.0616
	Smoothed Pfeffermann-Sverkov	0.00	0.0520	0.00	0.0731
	Smoothed optimal	0.00	0.0520	0.00	0.0571

7.2 Simulation 2

In the second simulation study a finite population of size $N = 10,000$ was generated with values (x_i, y_i, π_i), where $x_i \sim N(4,1)$, $y_i \sim$ Bernoulli(p_i), logit$(p_i) = \beta_0 + \beta_1 x_i$, $(\beta_0, \beta_1) = (-2, 1)$ and

$$\pi_i = \frac{\exp(-4 + 0.3x_i + 0.3y_i + 0.3u_i)}{1 + \exp(-4 + 0.3x_i + 0.3y_i + 0.3u_i)},$$

where $u_i \sim N(0,1)$. From the finite population, we repeatedly generated $B = 2,000$ samples by Poisson sampling where the sample indicator I_i follows a Bernoulli(π_i) distribution. The average sample size in this study is about 787.

For each sample, we computed five estimators of (β_0, β_1) by solving

$$\sum_{i=1}^{N} I_i w_i \{y_i - p(x_i; \beta)\}(1, x_i) = (0, 0),$$

where logit$p(x_i;\beta) = \beta_0 + \beta_1 x_i$ and the alternative choices of weights w_i are as follows:

1. design weights d_i;
2. Pfeffermann-Sverchkov semiparametric weights $d_i q_i$, where $q_i = 1/E(d_i \mid x_i, I_i = 1)$ was obtained using (20);
3. smoothed design weights $\tilde{d}_i$ computed using (18);
4. smoothed Pfeffermann-Sverkov weights $\tilde{d}_i q_i$, where $q_i = 1/E(d_i \mid x_i, I_i = 1)$ was obtained using (20), as in Simulation 1;
5. smoothed optimal weights using the EM-type algorithm (32), as discussed in Example 6.1.

Table 2 presents the Monte Carlo biases, standard errors, and root mean squared errors of the five weighted estimators of β_0 and β_1 considered. As expected, the smoothed optimal estimator shows the smallest standard error. Variance estimators computed from (33) were all nearly unbiased in the simulation.

8 Concluding Remarks on Robustness

We have shown how the efficiency of weighted estimation of regression coefficients under informative sampling may be improved by two approaches to modifying the survey weights: smoothing and multiplication by a function of the covariate values. Both approaches, in their optimal forms, depend on fitting an auxiliary regression model to the weights. An important difference

Table 2. Properties of alternative weighted estimators in Simulation 2, based on 2,000 Monte Carlo samples

Parameter	Weight	Bias	S.E.	RMSE
Intercept	Design	-0.00	0.586	0.586
	Pfeffermann-Sverchkov	-0.01	0.565	0.565
	Smoothed design	-0.00	0.567	0.567
	Smoothed Pfeffermann-Sverkov	-0.01	0.546	0.546
	Smoothed optimal	-0.01	0.544	0.544
Slope	Design	0.00	0.158	0.158
	Pfeffermann-Sverchkov	0.01	0.153	0.153
	Smoothed design	0.00	0.153	0.153
	Smoothed Pfeffermann-Sverkov	0.01	0.147	0.147
	Smoothed optimal	0.01	0.147	0.147

between the approaches is that the consistency of estimation of the regression coefficients of interest depends on specifying the weight model correctly for the smoothing approach, but holds under arbitrary misspecification of the weight model for the second approach.

We conclude that weight smoothing is only likely to be appealing in practice if the gain in efficiency it offers is appreciably superior to that offered by the second approach alone. The simulation studies illustrate how this may be the case. An illustration of how this may not be the case is provided by a common kind of survey of businesses or organizations where the principal source of variation in inclusion probabilities is the size of the organization. If this size variation is captured well by the x vector then smoothing is unlikely to offer much additional gain. For example, Fuller (2009, Example 6.3.3) analyses data from the Canadian Workplace and Employee Survey where disproportionate stratification is applied according to a measure of size based on 1998 tax records. A single x variable is based on the total employment at the workplace in 1999, which captures a major source of variation in the weights. Fuller (2009) finds that the second approach to weight modification does provide appreciable gains compared to a design-weighted approach. We find, when analysing these data, that there is little to be gained further by weight smoothing. Indeed, Fuller (2009) shows that the standard errors achieved by the second approach are close to those for unweighted least squares and, since this effectively represents a lower bound, smoothing will be unable to do any better.

A further consideration is the question of robustness of these approaches under misspecification of the underlying regression model of interest. Under such misspecification, the alternative weighting methods will no longer provide consistent estimation of a common parameter. The different weighted estimators will, in general, converge to different limits. Comparison of the different weighting approaches will therefore need to take account of both the appropriateness of these different limits as well as efficiency. We suggest that the appropriateness of limits will depend on the nature of the scientific application and that none of these approaches will always be superior in this respect, in line with the conclusion drawn by Scott and Wilde (2002) in the case of logistic regression modelling of case-control data.