Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Nonparametric propensity weighting for survey nonresponse through local polynomial regression

by Damião N. da Silva and Jean D. Opsomer

December 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2009
Vol. 35, No. 2, pp. 165-176
Statistics Canada, Catalogue No. 12-001-X

Nonparametric propensity weighting for survey nonresponse through local polynomial regression

Damião N. da Silva and Jean D. Opsomer [1]

Abstract

Propensity weighting is a procedure to adjust for unit nonresponse in surveys. A form of implementing this procedure consists of dividing the sampling weights by estimates of the probabilities that the sampled units respond to the survey. Typically, these estimates are obtained by fitting parametric models, such as logistic regression. The resulting adjusted estimators may become biased when the specified parametric models are incorrect. To avoid misspecifying such a model, we consider nonparametric estimation of the response probabilities by local polynomial regression. We study the asymptotic properties of the resulting estimator under quasi-randomization. The practical behavior of the proposed nonresponse adjustment approach is evaluated on NHANES data.

Key Words: Kernel regression; Missing data; Propensity scores; Unit nonresponse; Weighting adjustment.

1. Introduction

Propensity weighting is a procedure that is often applied in sampling surveys to compensate for unit nonresponse. Under this type of nonresponse, complete data collection is accomplished at only a part of the units selected to the sample, which are termed as the respondents. The propensity weighting procedure operates by increasing the sampling weights of the respondents in the sample using estimates of the probabilities that they responded to the survey. These probabilities are also referred to as response propensities in virtue of their analogy with the propensity score theory of Rosenbaum and Rubin (1983) for observational studies, incorporated into survey nonresponse problems by David, Little, Samuhel and Triest (1983).

General descriptions of propensity weighting to adjust classical survey estimators for nonresponse can be seen, for example, in Nargundkar and Joshi (1975), Cassel, Särndal and Wretman (1983) and Groves, Dillman, Eltinge and Little (2002). Traditionally, the way the procedure is implemented estimates the response probabilities with parametric regression curves, such as logistic, probit or exponential models. See Alho (1990), Folsom (1991), Ekholm and Laaksonen (1991) and Iannacchione, Milne and Folsom (1991) for earlier references. A recent theoretical account of the statistical properties of the procedure is given in Kim and Kim (2007). These parametric models are readily fitted as generalized linear models. However, an important and sometimes overlooked part of this procedure is the specification of the form of the link function to relate the response propensities and a linear predictor of the auxiliary information. If this function, which we shall refer to as the response propensity function, is misspecified, the resulting adjusted estimators of the population quantities are likely to be biased.

Another approach to estimate the response propensities is through nonparametric methods. The main motivation to use such methods is that the parametric form for the response propensity function need not be specified. In this sense, these methods offer an appealing alternative to the choice of a link function, as raised by Laaksonen (2006), or when a parametric model is difficult to specify a priori. In this context, Giommi (1984) proposed using kernel smoothing, in the form of the Nadaraya-Watson estimator, to estimate the response probabilities. Da Silva and Opsomer (2006) established the consistency of Giommi's estimator for the population mean and derived rates for the asymptotic bias and the variance. Theoretical properties of a Jackknife variance estimator were also studied.

In this article, we extend the results of Da Silva and Opsomer (2006) in two directions. First, we consider the estimation of the response propensities by local polynomial regression, a nonparametric technique described, for instance, in Wand and Jones (1995). Compared to kernel smoothing, local polynomial regression improves the local approximation to the unknown propensity function, which results in better practical and theoretical properties. It is also much more prevalent as a smoothing method in practice, with implementations available in most major statistical programs. Second, we apply the nonparametric propensity score estimation approach to data from the National Health and Nutrition Examination Survey (NHANES), which makes it possible to compare several nonresponse adjustment methods, both parametric and nonparametric, in a realistic setting.

In Section 2, we introduce the weighting procedure and the estimation of the response propensities. The theoretical properties of the adjusted estimators are discussed in Section 3. In section 4, we describe how to adapt a replication

1. Damião N. da Silva, Departamento de Estatística, Campus Universitário, Natal, RN 59078-970, Brazil. E-mail: damiao@ccet.ufrn.br; Jean D. Opsomer, Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, U.S.A. E-mail: jopsomer@stat.colostate.edu.

variance procedure to estimate the variance of the proposed adjusted estimators. Finally, in Section 5, we demonstrate the finite sample properties of the estimators by means of a simulation experiment using data from NHANES.

2. Weighting by local polynomial regression

Consider a population of N_v units, denoted by $U_v = \{1, 2, ..., N_v\}$. Suppose that a sample s_v is drawn from U_v, according to some probabilistic sampling design $p(s_v)$. Let n_v be the size of s_v and $\pi_i = \pi_{iv} = \Pr\{i \in s_v\} = \sum_{s_v:\, i \in s_v} p(s_v)$ be the inclusion probability of unit i, for all $i \in U_v$. It is of interest to estimate the population mean of a study variable y, namely $\bar{y}_{N_v} = N_v^{-1} \sum_{i \in U_v} y_i$, where y_i denotes the value of y for the i^{th} unit of U_v. We assume that the values x_i of an auxiliary variable x are fully observed throughout the sample. Let $\boldsymbol{y}_v = (y_1, ..., y_{N_v})$, and similarly for $\boldsymbol{x}_v$.

When the sample contains unit nonresponse, we only observe the values of the study variables for the units in a subset $r_v \subset s_v$. To account for the information lost in the estimation of the parameters of interest, it becomes necessary to model the response process. To define this response model, let R_i be an indicator variable assuming the value one if the unit i respond to the survey, and the value zero otherwise, for all $i \in s_v$. We assume that, given the sample, the response indicators are independent Bernouilli random variables with

$$\Pr\{R_i = 1 \mid i \in s_v, \boldsymbol{y}_v, \boldsymbol{x}_v\} = \phi(x_i) \equiv \phi_i, \text{ for all } i \in s_v, \quad (1)$$

where the exact form of the *response propensity function* $\phi(\cdot)$ is unspecified, but it is assumed to be a smooth function of x_i with $\phi(\cdot) \in (0, 1]$. The relationship in (1) defines a nonresponse process said to be ignorable, in the sense that the response propensities are independent of the values of any study variable, conditional on the covariate x (see Lohr 1999, page 265). The theory developed here, therefore, does not intend to handle non-ignorable response mechanisms.

If all response propensities were known, resulting weighting adjustments could be obtained by applying a two-phase estimation approach. For instance, two possible estimators of the population mean $\bar{y}_{N_v}$ would be given by

$$\bar{y}_{\pi\phi v} = \frac{1}{N_v} \sum_{i \in s_v} \pi_i^{-1} \phi_i^{-1} y_i R_i \quad (2)$$

and

$$\bar{y}_{\text{rat},\, \pi\phi v} = \sum_{i \in s_v} \pi_i^{-1} \phi_i^{-1} y_i R_i \Big/ \sum_{i \in s_v} \pi_i^{-1} \phi_i^{-1} R_i, \quad (3)$$

which are forms of adjustments for the Horvitz-Thompson and the Hájek estimators to compensate for the unit nonresponse. The same ideas can be used to obtain propensity weighting adjustments for the generalized regression estimator for estimation in the presence of nonresponse (Cassel *et al*. 1983).

Estimators (2) and (3) are unbiased and nearly unbiased for $\bar{y}_{N_v}$ respectively, under the quasi-randomization approach of Oh and Scheuren (1983), where the statistical properties are evaluated using the joint distribution of the sampling design and the response model. However, the response propensities are usually unknown in practice and we need to replace the ϕ_i in (2) and (3) by estimates $\hat{\phi}_i$, satisfying $0 < \hat{\phi}_i \leq 1$. The resulting propensity weighting estimators are therefore

$$\bar{y}_{\pi\hat{\phi}v} = \frac{1}{N_v} \sum_{i \in s_v} \pi_i^{-1} \hat{\phi}_i^{-1} y_i R_i \quad (4)$$

and

$$\bar{y}_{\text{rat},\, \pi\hat{\phi}v} = \sum_{i \in s_v} \pi_i^{-1} \hat{\phi}_i^{-1} y_i R_i \Big/ \sum_{i \in s_v} \pi_i^{-1} \hat{\phi}_i^{-1} R_i. \quad (5)$$

The latter formula has the advantage of being location-scale invariant, because the summation of its adjusted weights $\pi_i^{-1} \hat{\phi}_i^{-1} R_i / \sum_{i \in s_v} \pi_i^{-1} \hat{\phi}_i^{-1} R_i$ is equal to one, and does not require the population size N_v to be known.

In order to implement the propensity weighting estimators (4) and (5), it is necessary to estimate the response propensities $\hat{\phi}_i$. Da Silva and Opsomer (2006) used kernel regression for this purpose. The procedure we consider here is local polynomial regression, which can be described as follows. Let $K(\cdot)$ be a continuous and positive kernel function and h_v be its bandwidth. Define the $N_v \times (k + 1)$ matrix

$$\mathbf{X}_{Ui} = \begin{bmatrix} 1 & (x_1 - x_i) & \cdots & (x_1 - x_i)^k \\ \vdots & \vdots & & \vdots \\ 1 & (x_{N_v} - x_i) & \cdots & (x_{N_v} - x_i)^k \end{bmatrix},$$

the $N_v \times N_v$ matrix

$$\mathbf{W}_{Ui} = \text{diag}\left\{ \frac{1}{h_v} K\left(\frac{x_j - x_i}{h_v} \right) : 1 \leq j \leq N_v \right\}.$$

and population vector of response indicators $\mathbf{R}_U = (R_1, R_2, ..., R_{N_v})'$. The vector $\mathbf{R}_U$ would be known if, instead of the sample s_v, a census was considered from the population U_v. In that case, the local polynomial regression estimator of degree k of $\phi_i = \phi(x_i)$, based on the whole population, would be given by the fit

$$\hat{\phi}_{Ui} = \boldsymbol{e}_1'(\mathbf{X}_{Ui}'\mathbf{W}_{Ui}\mathbf{X}_{Ui})^{-1}\mathbf{X}_{Ui}'\mathbf{W}_{Ui}\mathbf{R}_U, \quad (6)$$

where e_j denotes the j^{th} column of the identity matrix of order $k+1$ and it is assumed that $\mathbf{X}_{Ui}'\mathbf{W}_{Ui}\mathbf{X}_{Ui}$ is non-singular.

Since the values of the response indicators are only observed for those units selected into the sample, the population fit (6) is unfeasible. However, defining $\mathbf{X}_{si}$ as the $n_\nu \times (k+1)$ matrix formed with the rows of $\mathbf{X}_{Ui}$ corresponding to the units $j \in s_\nu$,

$$\mathbf{W}_{si} = \text{diag}\left\{\frac{1}{\pi_j h_\nu} K\left(\frac{x_j - x_i}{h_\nu}\right):\ j \in s_\nu\right\}$$

and $\mathbf{R}_s = (R_j:\ j \in s_\nu)'$, then a sample-based local polynomial regression estimator of degree k of $\phi_i = \phi(x_i)$ is given by

$$\hat{\phi}_i^o = e_1'\hat{\mathbf{T}}_{si}^{-1}\hat{\mathbf{t}}_{si} \quad (7)$$

where

$$(\hat{\mathbf{T}}_{si},\ \hat{\mathbf{t}}_{si}) \equiv \left(\{\hat{T}_{si,\,pq}\}_{p,\,q=1}^{k+1},\ (\hat{t}_{si,\,p})_{p=1}^{k+1}\right) = (\mathbf{X}_{si}'\mathbf{W}_{si}\mathbf{X}_{si},\ \mathbf{X}_{si}'\mathbf{W}_{si}\mathbf{R}_s)$$

and it is assumed that $\hat{\mathbf{T}}_{si}$ is invertible. An special case of (7) is obtained by considering $k = 0$, which corresponds to the kernel regression estimator of Da Silva and Opsomer (2006). Other special cases from (7) are the local linear, the local quadratic and the local cubic response propensity estimators, which result from the local fit of polynomials of degree one, two and three, respectively.

In practice, when $\hat{\mathbf{T}}_{si}$ happens to be singular, a simple procedure to insure that $\hat{\phi}_i^o$ is well defined is choosing a bandwidth large enough to guarantee at least $k+1$ values of R_j in the window $[x_i - h_\nu,\ x_i + h_\nu]$, for all $i \in s_\nu$. If this window does not contain enough responses indicators and the bandwidth has to remain fixed, another approach has to be considered. To this purpose, we adopt here the adjustment made by Breidt and Opsomer (2000) and define the sample-based local polynomial regression estimator of degree k of $\phi_i = \phi(x_i)$ by

$$\hat{\phi}(x_i,\ k,\ h_\nu) = \boldsymbol{e}_1'\left(\hat{\mathbf{T}}_{si} + \text{diag}\left\{\frac{\delta_1}{N_\nu}\right\}\right)^{-1}\hat{\mathbf{t}}_{si},\ i \in s_\nu. \quad (8)$$

where δ_1 is some small positive constant. The smaller order terms δ_1/N_ν added to the main diagonal of $\hat{\mathbf{T}}_{si}$ are sufficient to make the resulting adjusted matrix invertible for any h_ν. As a consequence, $\hat{\phi}(x_i,\ k,\ h_\nu)$ will be well defined, for all $i \in s_\nu$. However, another technical difficulty to use $\hat{\phi}(x_i,\ k,\ h_\nu)$ as a propensity weighting adjustment arises because the response propensity estimator (8) can indeed become arbitrarily close to zero. To tackle this problem, we bound $\hat{\phi}(x_i,\ k,\ h_\nu)$ away from zero by considering the estimator

$$\hat{\phi}_i = \max\{\hat{\phi}(x_i,\ k,\ h_\nu),\ \delta_2(N_\nu h_\nu)^{-1}\}, \quad (9)$$

for some constant $\delta_2 > 0$. This idea is related to the adjustment made by Da Silva and Opsomer (2006) for the kernel regression estimator.

3. Asymptotic properties

In this section, we present the properties of the propensity weighting estimators (4) and (5) under estimation of the response propensities by the local polynomial estimator (9). The assumptions, lemmas and outlines of the proofs for the following results are given in the Appendix, and a complete theoretical investigation can be found in Da Silva and Opsomer (2008). The full derivations are not reported in this article, because they follow the general approach described in Da Silva and Opsomer (2006). We consider an asymptotic framework by which the population U_ν is embedded into the increasing sequence of populations $\{U_\nu:\ N_\nu < N_{\nu+1}\}_{\nu=1}^{\infty}$. From each U_ν, a sample s_ν of size $n_\nu (n_\nu \geq n_{\nu-1})$ is selected according to a sampling design $p_\nu(\cdot)$. This framework is commonly adopted in asymptotic studies of survey estimators. See Isaki and Fuller (1982) for an early reference.

As a population-based approximation for $\phi_i \equiv \phi(x_i)$, we shall consider in the derivation of most results in this section the population fit by local polynomial regression

$$\tilde{\phi}_i \equiv \tilde{\phi}(x_i,\ k,\ h_\nu) = \boldsymbol{e}_1'\mathbf{B}_i \equiv \boldsymbol{e}_1'\mathbf{T}_i^{-1}\mathbf{t}_i,\ i \in U_\nu, \quad (10)$$

where

$$(\mathbf{T}_i,\ \mathbf{t}_i) \equiv (\{T_{i,\,pq}\}_{p,\,q=1}^{k+1},\ (t_{i,\,p})_{p=1}^{k+1}) \equiv \text{E}(\hat{\mathbf{T}}_{si},\ \hat{\mathbf{t}}_{si}) = (\mathbf{X}_{Ui}'\mathbf{W}_{Ui}\mathbf{X}_{Ui},\ \mathbf{X}_{Ui}'\mathbf{W}_{Ui}\boldsymbol{\phi}_U),$$

the matrices $\mathbf{X}_{Ui}$ and $\mathbf{W}_{Ui}$ are as in (6) and $\boldsymbol{\phi}_U = (\phi(x_1),\ \phi(x_2),\ ...,\ \phi(x_{N_\nu}))'$. The following theorem states the asymptotic properties of $\bar{y}_{\pi\hat{\phi}\nu}$ under a set of assumptions in the Appendix. These assumptions are regularity conditions on the sampling design and the finite population, both of which are standard infinite population asymptotics, ignorability conditions on the nonresponse mechanism, and a set of standard regularity conditions related to the local polynomial regression of the response propensity function.

Theorem 1. *Assume the assumptions* (*A*1)-(*A*4), (*B*1)-(*B*3) *and* (*C*1)-(*C*5) *in the Appendix hold. Consider the estimation of the population mean* $\bar{y}_{N_\nu}$ *by the propensity weighting estimator* $\bar{y}_{\pi\hat{\phi}\nu}$ *defined in* (4), *and suppose the response propensities are estimated by* $\hat{\phi}_i$, *the local polynomial regression estimator of degree* k *in* (9). *Let*

$$\overline{y}_{\pi\hat{\psi}\nu} = \frac{1}{N_\nu} \sum_{i \in s_\nu} \pi_i^{-1} \hat{\psi}_i^{-1} y_i R_i, \tag{11}$$

where

$$\hat{\psi}_i^{-1} = \tilde{\phi}_i^{-1} - \tilde{\phi}_i^{-2} \boldsymbol{e}_1' \mathbf{T}_i^{-1} (\hat{\mathbf{t}}_{si} - \hat{\mathbf{T}}_{si} \mathbf{B}_i),$$

$\hat{\mathbf{t}}_{si}$ *and* $\hat{\mathbf{T}}_{si}$ *are given in* (7) *and* $\tilde{\phi}_i$, $\mathbf{B}_i$, $\mathbf{T}_i$ *are defined in* (10). *Then*,

$$\mathrm{E}[(\overline{y}_{\pi\hat{\phi}\nu} - \overline{y}_{\pi\hat{\psi}\nu})^2] = O\left(\frac{1}{n_\nu^2 h_\nu^2}\right) \tag{12}$$

and the bias and variance of $\overline{y}_{\pi\hat{\psi}\nu}$ *satisfy*

$$\begin{aligned} B_\nu &\equiv \mathrm{E}[\overline{y}_{\pi\hat{\psi}\nu} - \overline{y}_{N_\nu}] \\ &= \begin{cases} O\left(h_\nu^{k+(3/2)}\right) + O\left(\dfrac{1}{n_\nu h_\nu}\right) & k \text{ even}, \\ O\left(h_\nu^{k+1}\right) + O\left(\dfrac{1}{n_\nu h_\nu}\right) & k \text{ odd}, \end{cases} \end{aligned} \tag{13}$$

and

$$\mathrm{Var}[\overline{y}_{\pi\hat{\psi}\nu}] = O\left(\frac{1}{n_\nu h_\nu}\right). \tag{14}$$

Results (12) and (13) imply that the propensity weighting estimator $\overline{y}_{\pi\hat{\phi}\nu}$, using a response propensity estimator based on local polynomial regression, is asymptotically unbiased for the population mean $\overline{y}_{N_\nu}$ under the joint distribution of the sampling design and the response model (1). Combining this result with (14), then we obtain that

$$\hat{y}_{\pi\hat{\phi}\nu} = \overline{y}_{N_\nu} + O_p\left(\frac{1}{\sqrt{n_\nu h_\nu}}\right), \tag{15}$$

when the bandwidth satisfies

$$h_\nu = \begin{cases} O\left(n_\nu^{-\frac{1}{2k+4}}\right), & k \text{ even}, \\ O\left(n_\nu^{-\frac{1}{2k+3}}\right), & k \text{ odd}. \end{cases} \tag{16}$$

Hence, without assuming a parametric form for the response propensity function $\phi(\cdot)$, $\overline{y}_{\pi\hat{\phi}\nu}$ is consistent for the population mean with respect to the sampling design and the response model, as long as the response propensities are a smooth function of the covariate x. As a price paid for this robustness, the rate of convergence is of order $\sqrt{n_\nu h_\nu}$ instead of the usual parametric rate $\sqrt{n_\nu}$. However, as the degree of the local polynomial k increases, the rate of convergence improves. Since the kernel regression estimator in Da Silva and Opsomer (2006) is equivalent to the case $k = 0$, local polynomial regression with higher degree is asymptotically superior to kernel regression in the context of a nonresponse adjustment. This theoretical finding is consistent with that in other contexts (see *e.g.*, Wand and Jones 1995, page 130).

Expression (11) on Theorem 1 generalizes another finding from Da Silva and Opsomer (2006) to the case of local polynomial regression, which is that the asymptotic weights $\hat{\psi}_i^{-1}$ cannot be approximated by the inverse of response propensities ϕ_i^{-1} (or their population-level estimators $\tilde{\phi}_i^{-1}$). One immediate consequence is that the estimator $\overline{y}_{\pi\hat{\phi}\nu}$ is *not* asymptotically equivalent to $\overline{y}_{\pi\hat{\phi}\nu}$ in (2).

The following corollary provides an asymptotic distribution for $\overline{y}_{\pi\hat{\phi}\nu}$, assuming the asymptotic normality of $\overline{y}_{\pi\hat{\psi}\nu}$.

Corollary 1. *Assume the conditions of Theorem* 1 *hold. Suppose that the sampling design and the response model are such that*

$$\frac{\overline{y}_{\pi\hat{\psi}\nu} - \overline{y}_{N_\nu} - B_\nu}{[\mathrm{Var}(\overline{y}_{\pi\hat{\psi}\nu})]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \text{ as } \nu \to \infty,$$

where B_ν *is defined in* (13). *If additionally*

$$\lim_{\nu \to \infty} (n_\nu h_\nu) \mathrm{Var}(\overline{y}_{\pi\hat{\psi}\nu}) \in (0, \infty),$$

then

$$\frac{\overline{y}_{\pi\hat{\phi}\nu} - \overline{y}_{N_\nu} - B_\nu}{[\mathrm{Var}(\overline{y}_{\pi\hat{\psi}\nu})]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

We now discuss the properties of the ratio-based version of propensity weighting estimator given in (5). Based on the results for $\overline{y}_{\pi\hat{\phi}\nu}$, standard ratio estimation theory can be used to derive asymptotic results for $\overline{y}_{\mathrm{rat},\,\pi\hat{\phi}\nu}$. In particular, under the same assumptions the asymptotic rates for the approximate bias and variance of $\overline{y}_{\mathrm{rat},\,\pi\hat{\phi}\nu}$ are the same as those in Theorem 1, and the asymptotic distribution of $\overline{y}_{\mathrm{rat},\,\pi\hat{\phi}\nu}$ is given in the following result.

Theorem 2. *Assume the conditions of Theorem* 1 *hold. Suppose the population mean is to be estimated by the propensity weighted estimator* $\overline{y}_{\mathrm{rat},\,\pi\hat{\phi}\nu}$ *of* (5) *and the response propensities are estimated by* $\hat{\phi}_i$, *the local polynomial regression estimator of degree* k *defined in* (8). *Let*

$$\overline{e}_{\pi\hat{\psi}\nu} = \frac{1}{N_\nu} \sum_{i \in s_\nu} \pi_i^{-1} \hat{\psi}_i^{-1} (y_i - \overline{y}_{N_\nu}) R_i,$$

where the weights $\hat{\psi}_i^{-1}$ *are given in Theorem* 1. *Suppose that*

$$\frac{\overline{e}_{\pi\hat{\psi}\nu} - \mathrm{E}(\overline{e}_{\pi\hat{\psi}\nu})}{[\mathrm{Var}(\overline{e}_{\pi\hat{\psi}\nu})]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \text{ as } \nu \to \infty,$$

and

$$\lim_{\nu\to\infty}(n_\nu\, h_\nu)\mathrm{Var}(\overline{e}_{\pi\hat{\psi}\nu}) \in (0,\ \infty).$$

Then,

$$\frac{\overline{y}_{\mathrm{rat},\,\pi\hat{\phi}\nu} - \overline{y}_{N_\nu} - B_{\mathrm{rat},\,\nu}}{[\mathrm{Var}(\overline{e}_{\pi\hat{\psi}\nu})]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\ 1)$$

as $\nu \to \infty$, *where* $B_{\mathrm{rat},\,\nu} = O(h_\nu^{k+1})$, *if* k *is odd, and* $B_{\mathrm{rat},\,\nu} = O(h_\nu^{k+(3/2)})$, *if* k *is even.*

4. Variance estimation

As noted in Section 3, the estimator $\overline{y}_{\pi\hat{\phi}\nu}$ is not asymptotically equivalent to $\overline{y}_{\pi\phi\nu}$, so that approximating the asymptotic variance of the former by that of the latter is typically incorrect. In fact, a proof that the asymptotic variance of $\overline{y}_{\pi\phi\nu}$ overestimates the variance of $\overline{y}_{\pi\hat{\phi}\nu}$ is given by Kim and Kim (2007) when the response propensities are assumed to follow a parametric model. In the present context, the asymptotic variance of $\overline{y}_{\pi\hat{\phi}\nu}$ is

$$\mathrm{Var}[\,\overline{y}_{\pi\hat{\psi}\nu}] = \mathrm{Var}\left(\frac{1}{N_\nu}\sum_{i\in s_\nu}\pi_i^{-1}\,\hat{\psi}_i^{-1}R_i\,y_i\right),$$

with $\hat{\psi}_i^{-1}$ given in Theorem 1. As was previously noted in Da Silva and Opsomer (2006) for the simpler case of a zero degree polynomial, the high level of complexity in the expression makes direct estimation of this variance impractical, and a replication method was proposed instead. We briefly outline the procedure here, which is extended to local polynomials of degree k. We omit the theoretical derivations.

We start from a set of replicate weights in the absence of nonresponse, defined for estimating the variance of a linear estimator

$$\hat{\theta} = \frac{1}{N_\nu}\sum_{i\in s_\nu} w_i\, y_i.$$

The replicate variance estimator for $\hat{\theta}$ is defined as

$$\hat{\mathrm{V}}(\hat{\theta}) = \sum_{\ell=1}^{L_\nu} c_\ell\,(\hat{\theta}^{(\ell)} - \hat{\theta})^2, \qquad (17)$$

where

$$\hat{\theta}^{(\ell)} = \frac{1}{N_\nu}\sum_{i\in s_\nu} w_i^{(\ell)}\, y_i,\ \ell = 1,\ 2,\ ...,\ L_\nu,$$

denotes a set of L_ν replicates for $\hat{\theta}$, $w_i^{(\ell)}$ are sampling weights associated with the ℓ^{th} replicate and c_ℓ is factor that depends on the replication procedure. Examples of replication procedures satisfying (17) use variants of the Jackknife method or the Balanced Repeated Replication technique. The process to adapt the replication procedure to estimating the variance of $\overline{y}_{\pi\hat{\phi}\nu}$ and $\overline{y}_{\mathrm{rat},\,\pi\hat{\phi}\nu}$ is straightforward. The needed replicates of these adjusted estimators, namely $\overline{y}^{(\ell)}_{\pi\hat{\phi}\nu}$ and $\overline{y}^{(\ell)}_{\mathrm{rat},\,\pi\hat{\phi}\nu}$, are obtained by replacing the $w_i = \pi_i^{-1}$ by $w_i^{(\ell)}$ in (4) and (5), respectively, and also in the computations needed to produce the $\hat{\phi}_i$ in (9). In section 5.4 below, we evaluate the practical performance of the replication variance procedure on NHANES data.

5. Application to NHANES data

5.1 The NHANES design

We evaluate the performance of the local polynomial adjusted estimators on real data. We consider the 2005-2006 release of the National Health and Nutrition Examination Survey (NHANES), which is conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention (NCHS/CDC), of the U.S. Department of Health and Human Services. This survey consists of a stratified, multistage sample of the U.S. civilian non-institutionalized population. A general overview of the sample formation is as follows:

(i) within each stratum, primary sampling units (PSUs) consisting of counties or grouped smaller counties are selected by sampling with probabilities proportional to a measure of size;
(ii) from the sampled PSUs, groups of city blocks (segments) containing clusters of households are selected also by sampling with probability proportional to size;
(iii) in the selected segments, clusters of households are randomly selected with varying selection probabilities to oversample groups of age, ethnic, or income in certain geographic areas; and
(iv) in the selected households, one or more participants are selected randomly.

The public release of NHANES data has two important aspects. First, to reduce disclosure risks, the stratified, four-stage survey is condensed in a stratified one-stage design, with neither the new stratum variable nor the new PSU variable corresponding to the same variables in the original design. Secondly, the base sampling weights, obtained by reciprocal of the inclusion probabilities of the survey participants, are not released. The weights provided reflect adjustments made to the base weights to account for unit nonresponse, in the interview and exam portions of the survey, and to produce estimates satisfying known population controls.

5.2 The simulation experiment

In order to empirically evaluate the local polynomial estimators as adjustments for nonresponse in complex surveys, we will apply an artificially generated source of unit nonresponse to the public-release NHANES dataset. The nonresponse mechanism will be taken as a smooth function of the age in years of the survey participant (AGE). For this comparison, we chose as study variables four characteristics related to heart diseases, namely the systolic blood pressure (SBP), the diastolic blood pressure (DBP), the indicator of hypertension (HTN) and the indicator of high serum total cholesterol (HTC). All of these were measured on survey participants who were 18 years or older. The systolic and diastolic variables were obtained as the average of the corresponding measurements in a set of up to four readings. Hypertension was defined for individuals having systolic blood pressure of 140 mm Hg or higher or a mean diastolic blood pressure of 90 mm Hg or higher or currently taking medication to lower high blood pressure. High serum total cholesterol was considered when the individual had a total serum cholesterol greater than or equal to 240 mg/dL. The unweighted sample correlations among these and the AGE variable are 0.481 (SBP), 0.118 (DBP), 0.552 (HTN) and 0.060 (HTC), respectively. Hence, it is reasonable to postulate that unit nonresponse related to age is likely to have different effects on survey estimators for these four variables.

The total number of eligible individuals in the NHANES dataset is 4,727. We generated unit nonresponse for the four variables of interest according to two logistic response propensity functions of the auxiliary variable x taken by the age (in years) of the survey participant minus 18. These functions consider a linear and a nonlinear predictor of x as follows

Linear predictor:

$$\phi_I(x) = \{1 + \exp[-(\beta_0 + \beta_1 x)]\}^{-1}$$

Nonlinear predictor:

$$\phi_{II}(x) = \{1+\exp[-(\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 \cos(\beta_4 x^2/\pi)\sin(\beta_5 x/\pi))]\}^{-1},$$

where the regression coefficients $\beta_0, ..., \beta_5$ were chosen so that the response propensity functions give an overall nonresponse rate of about 30% when applied to the sample values of x. In both cases, we kept the NHANES sample fixed and generated $B = 1{,}000$ independent response indicator vectors by Poisson sampling.

The following six nonresponse adjustments were evaluated on these data. Note that in all cases we reported the ratio versions (5) of the estimators, because they were found to be much more precise than the Horvitz-Thompson versions.

1. True response probabilities: $\hat{\phi}_i = \phi(x_i)$, $i \in s_v$.
2. Logistic regression adjustment: $\hat{\phi}_i$ obtained as the estimated probabilities from a logistic regression of each response vector on x, using a polynomial in x of degree one as the linear predictor.
3. Weighted local polynomial regression of degree k and bandwidth h_v: $\hat{\phi}_i = \hat{\phi}(x_i, k, h_v)$ given by (8), with $i \in s_v$, $k = 0, 1, 2, 3$, $h_v = 0.15, 0.25, 0.50$ and the Epanechnikov kernel function
$$K(t) = (3/4)(1 - x^2) I\{|x| \le 1\}.$$
4. Unweighted local polynomial regression of degree k and bandwidth h_v: the same as above but not including the sampling weights in (8) to obtain the $\hat{\phi}_i = \hat{\phi}(x_i, k, h_v)$. This might be somewhat easier to compute in practice and should lead to similar results, even if it does not, strictly speaking, follow the pseudo-randomization theory of Section 3.
5. Weighting within cell: within each stratum, respondents and nonrespondents were classified into four classes of age based on the sample quartiles of this variable. This procedure subdivided the sample in a total of 60 cells. Let s_g and s_{rg} denote respectively the set of sampled elements and the set of responding elements in the g^{th} cell. Then, the WC adjustment is defined by taking
$$\hat{\phi}_i = \frac{\sum_{i \in s_{rg}} w_i}{\sum_{i \in s_g} w_i},$$
for all respondents $i \in s_{rg}$.
6. Naive: $\hat{\phi}_i = 1$, $i \in s_v$.

5.3 Bias and robustness against a misspecified response propensity function

When the full sample without artificial nonresponse is used, the Hájek estimated means for the four study variables are respectively SBP = 122.19 mm Hg, DBP = 70.29 mm Hg, HTN = 29.04% and HTC = 15.76%. Table 1 gives the percentage bias relative to those means across response sets obtained for every adjustment procedure in this simulation experiment. For both weighted and unweighted Local Polynomial Regression adjusted estimators, we only display the results for the bandwidth $h_v = 0.25$, but those for other bandwidth values are similar. We instead show the results for different degrees of the local polynomial, so that the effect of moving from local constant to higher order polynomials can be evaluated.

Table 1
Relative biases (%) of nonresponse-adjusted estimators for mean systolic blood pressure (SBP), diastolic blood pressure (DBP), indicator of hypertension (HTN) and indicator of high serum total cholesterol (HTC), based on 1,000 response sets for two propensity functions of the age of the survey participant in NHANES 2005-2006

	Logistic propensity function (linear predictor)				Logistic propensity function (nonlinear predictor)			
Type of adjustment	**SBP**	**DBP**	**HTN**	**HTC**	**SBP**	**DBP**	**HTN**	**HTC**
True Response Propensities	0.01	0.01	-0.01	0.04	-0.00	-0.00	0.01	-0.22
Logistic Regression	0.01	0.00	-0.03	0.03	0.47	-1.67	6.49	-6.76
Weighted Local Polynomial Regression:								
Degree 0	0.27	0.34	3.39	2.41	-0.20	-0.39	-1.20	-2.27
Degree 1	0.00	0.04	-0.03	0.20	-0.01	-0.49	0.34	-2.36
Degree 2	0.01	0.01	0.03	0.07	0.03	-0.05	0.51	-0.27
Degree 3	0.01	0.01	-0.02	0.04	-0.03	-0.05	-0.24	-0.44
Unweighted Local Polynomial Regression:								
Degree 0	0.11	0.24	1.34	1.53	-0.17	-0.47	-0.98	-2.70
Degree 1	0.01	0.05	-0.00	0.25	-0.01	-0.57	0.34	-2.69
Degree 2	0.01	0.01	-0.00	0.07	0.01	-0.07	0.26	-0.40
Degree 3	0.00	0.01	-0.06	0.03	-0.03	-0.06	-0.29	-0.48
Weighting Within Cell	0.08	0.08	0.84	0.69	-0.11	-0.07	-0.84	-0.48
Naive	1.62	0.80	20.49	8.04	-1.30	-1.60	-15.61	-10.77

Among the estimators affected by the generated nonresponse, the worst bias performances are clearly for the unadjusted "Naive" estimator. As displayed in the last row of Table 1, the biases are higher in the estimation of the prevalence of hypertension and the mean systolic blood pressure, as these are the characteristics of the study variables with higher correlations with the AGE variable, and also for the prevalence of high serum total cholesterol. The biases of the Naive estimator can be successfully reduced with the true response propensity estimator, any of the local polynomial regression adjusted estimators, the weighting-within cell estimator or with the logistic adjusted estimator, if the model for the propensity function is correctly specified. The best performances in terms of small bias are obtained using the estimator adjusted by the true response propensities, because it is conditionally unbiased for the full sample estimates. The logistic adjustment when it is applied under the correct model, given by the propensity function with a linear predictor, also gives nearly unbiased estimates. For the second propensity function, where the form of the predictor is not well captured by the logistic regression fit of a regression line, this adjustment yields a conditionally biased estimator.

The averages of the local polynomial regression estimates become generally closer to the full sample estimates by increasing the degree of the polynomial fitted, with the largest jump when moving from a local constant to a local linear estimator. Hence, it seems that local polynomial regression is indeed superior to kernel regression in this context. There is very little difference between the weighted and unweighted forms of this adjustment and both procedures have overall smaller conditional biases than the biases of the weighting-within cell estimator, when they are implemented by fitting locally a polynomial of order greater than zero to estimate the response propensities. The zero degree propensity weighted and unweighted adjusted estimators have smaller biases at smaller bandwidths, as we observed with the bandwidth 0.15, for instance, but smaller bandwidths tend to increase the variance of the estimators. Overall, both weighted and unweighted local polynomial regression adjustments outperform the parametric logistic adjustment when the response model is misspecified. By implementing the local polynomial adjustments with degrees above one, their performances are similar to the one of the logistic adjustment under the correct specification of the response model.

5.4 Variance and variance estimation

Table 2 shows the variance of the adjustment methods considered here across the nonresponse replicates, and we normalized them by the variance for the true response propensity adjustment for clarity. Interestingly, there appears to be an inverse relationship between the magnitude of the relative biases in Table 1 and the variances in this table. In those cases where the relative bias was small (the weighted and unweighted local polynomial regression, the weighting within cell as well as the logistic regression adjustment for the linear propensity function), all the methods appear to result in roughly similar variances. There is a tendency for higher degree local polynomials to be more variable than lower degree ones, and this is particularly noticeable for the nonlinear propensity function, where a clear jump is seen when one moves from degree 1 (local linear) to 2 (local quadratic). Overall, it seems that local linear regression, either weighted or unweighted, offers a good compromise between the bias and the variance of the nonresponse adjustment procedure.

Table 2
Normalized Monte Carlo variances of nonresponse-adjusted estimators for mean systolic blood pressure (SBP), diastolic blood pressure (DBP), indicator of hypertension (HTN) and indicator of high serum total cholesterol (HTC), based on 1,000 response sets for two propensity functions of the age of the survey participant in NHANES 2005-2006

	Logistic propensity function (linear predictor)				Logistic propensity function (nonlinear predictor)			
Type of adjustment	**SBP**	**DBP**	**HTN**	**HTC**	**SBP**	**DBP**	**HTN**	**HTC**
True Response Propensities	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Logistic Regression	85.9	92.4	79.5	96.5	63.9	61.5	54.1	52.0
Weighted Local Polynomial Regression:								
Degree 0	74.9	81.2	65.7	92.1	70.3	67.0	67.4	75.0
Degree 1	81.8	89.5	66.2	92.7	73.6	69.8	68.9	76.0
Degree 2	81.3	89.8	65.5	94.0	90.3	81.7	88.0	96.1
Degree 3	82.3	90.2	65.8	93.1	90.1	82.2	87.7	96.2
Unweighted Local Polynomial Regression:								
Degree 0	82.2	85.8	77.6	95.8	71.9	69.2	70.7	74.7
Degree 1	85.6	90.1	79.4	95.7	74.4	71.1	71.2	74.6
Degree 2	86.6	91.3	79.3	96.1	91.8	84.5	91.8	96.8
Degree 3	87.3	91.5	78.5	95.0	91.2	84.7	91.2	96.9
Weighting Within Cell	79.7	89.1	62.1	91.6	82.5	77.0	81.1	92.3
Naive	71.3	58.0	81.7	74.6	48.6	48.7	45.5	45.1

The above simulation results showed the behavior of several nonresponse adjustments in the NHANES setting. We now consider the replication variance estimation approach of Section 4 and evaluate its usefulness as a sample-based measure of uncertainty for the nonresponse-adjusted estimators in the same setting. We implemented (17) with the Jackknife method. Since NHANES does not provide information on the joint sample inclusion probabilities, we could not apply a full Jackknife variance estimator as in, for instance, Berger and Skinner (2005), as a means to account for the selection of units with varying probabilities in the survey. Because of this, we assumed the within-stratum designs in NHANES could be approximated by cluster sampling with replacement and rewrite (17) in the form proposed by Rust (1985),

$$\hat{\mathrm{V}}_{JK}(\hat{\theta}) = \sum_{t=1}^{T} c_t \sum_{j \in s_t} (\hat{\theta}^{(tj)} - \hat{\theta})^2, \qquad (18)$$

where s_t denote the set of units in sample from the t^{th} NHANES stratum, $t = 1, 2, ..., T$, n_t be the number of units selected to s_t, $c_t = (n_t - 1)/n_t$ and $\hat{\theta}^{(tj)}$ is obtained from (5) by replacing the w_i with the replication weights

$$w_{i(tj)} = \begin{cases} 0, & \text{for a survey participant } i \in \text{PSU } j,\ j \in s_t \\ n_t/(n_t - 1)\, w_i, & \text{for a survey participant } i \in \text{PSU } j',\ j' \in s_t\ (j' \neq j) \\ w_i, & \text{for a survey participant } i \notin s_t. \end{cases}$$

These weights were also applied in the estimation of the response propensities for the weighted local polynomial regression procedure adjustment procedure.

The Jackknife variance estimator (18) was applied to each response vector from the two propensity functions, yielding estimates $\hat{v}_{JK}(\hat{\theta}(b))$, $b = 1, 2, ..., B$, for all adjusted estimators in the experiment. For the sake of comparison, it would be informative to produce estimates of the corresponding variances by the Monte Carlo method. However, as the NHANES sample is fixed, the Monte Carlo variance of the point estimates $\hat{\theta}(b)$ across response vectors estimates only the conditional variance $\mathrm{Var}(\hat{\theta}|s_{\mathrm{v}})$ with respect to the response model. Since

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\mathrm{E}(\hat{\theta}|s_{\mathrm{v}})) + \mathrm{E}(\mathrm{Var}(\hat{\theta}|s_{\mathrm{v}})),$$

where the "inner" moments are taken with respect to the response model given the sample and the "outer" moments are with respect to the sampling design, the design variance of $\mathrm{E}(\hat{\theta}|s_{\mathrm{v}})$ needs to be accounted for in order to have a valid estimation target for $\hat{\mathrm{V}}_{JK}(\hat{\theta})$. Using the fact that weighted and unweighted local polynomial regression and weighting within cell all produce approximately conditionally unbiased estimators of the full sample estimator, $\bar{y}_{\pi,\,\mathrm{rat}} = \sum_{i \in s_{\mathrm{v}}} w_i y_i / \sum_{i \in s_{\mathrm{v}}} w_i$, for the two response propensities functions, we decided to use the Jackknife variance estimator of $\bar{y}_{\pi,\,\mathrm{rat}}$ as a "proxy" for $\mathrm{Var}(\mathrm{E}(\hat{\theta}|s_{\mathrm{v}}))$. Hence, our "comparison variance" will be defined as

$$\hat{v}_C(\hat{\theta}) = \hat{v}_{JK}(\bar{y}_{\pi,\,\mathrm{rat}}) + \frac{1}{B-1}\sum_{b=1}^{B} (\hat{\theta}(b) - \hat{\theta})^2.$$

Using $\hat{v}_C(\hat{\theta})$ instead of the true variance will tend to underestimate any bias issues associated with the use of the jackknife variance estimator for the full sample estimator. However, it will show how well the replication procedure manages to capture the nonresponse variability.

Table 3 gives relative biases of the Jackknife variance estimators obtained in this experiment. The results show that the jackknife variance estimator performs reasonably well for both nonresponse mechanisms and all estimators considered. The weighted local polynomial regression adjusted procedure appears to yield estimated variances in greater consonance with the comparison variance than when the procedure is implemented by its unweighted version. The results for the nonlinear predictor function exhibit more bias than those for the linear predictor, with more pronounced positive and negative biases present for the former for all the variables. As discussed in Da Silva and Opsomer (2006), replication methods for nonresponse-adjusted estimators often ignore a component of the total variance, which includes the effect of both sampling and the response mechanism. We therefore conjecture that the different bias behaviors exhibited for the different variables could be due to this missing variance component.

6. Concluding remarks

In this article, we studied properties of nonparametric propensity weighting as an adjustment procedure for survey nonresponse. The local polynomial regression technique is seen to offer a flexible way of constructing new survey adjustments for nonresponse. The results in the article extend those in Da Silva and Opsomer (2006) by allowing the use of local polynomials of arbitrary degree, which offers both theoretical and practical advantages over zero-degree kernel regression.

In addition to its good theoretical properties, the estimator was shown in the simulation experiment to be competitive with an estimator based on a correctly specified parametric model in terms of bias and variance, while protecting against a potentially misspecified model. The weighting-cell estimator is similarly robust against model misspecification, but a particular advantage of nonparametric regression methods over weighting cell approaches is the connection to broad classes of modeling techniques available in the non-survey literature. Extensions of the methodology we described here to semiparametric and (generalized) additive models (Hastie and Tibshirani 1986) are readily formulated and should work well in a wide range of potential response model scenarios, including situations with multiple covariates that are both categorical and continuous. A detailed discussion of these extensions is beyond the scope of the current paper, however.

In Section 5, we applied the nonparametric nonresponse adjustment to NHANES data by modeling the response probability as a smooth function of the age of the respondents, and weighting the data by the inverses of the estimated response probabilities. The same approach can be used in other survey datasets whenever continuous covariates related to the response probability are available for all elements in the original sample. This provides a viable alternative to the commonly used weighting-within-cell approach for situations in which cells are constructed by "binning" one or several continuous variables.

Table 3
Relative biases (%) of the Jackknife variance estimators of estimators of the mean systolic blood pressure (SBP), diastolic blood pressure (DBP), indicator of hypertension (HTN) and indicator of high serum total cholesterol (HTC), based on 1,000 response sets for two propensity functions of the age of the survey participant in NHANES 2005-2006

	Logistic propensity function (linear predictor)				Logistic propensity function (nonlinear predictor)			
Type of adjustment	**SBP**	**DBP**	**HTN**	**HTC**	**SBP**	**DBP**	**HTN**	**HTC**
True Response Propensities	0.55	-0.47	-0.06	0.16	0.92	-0.26	-1.03	-2.76
Weighted Local Polynomial Regression:								
Degree 0	-0.66	2.33	2.74	4.44	1.63	-2.27	-5.12	-9.44
Degree 1	-0.31	-1.03	0.31	1.87	5.27	4.03	2.60	-9.95
Degree 2	-0.14	-0.76	0.41	0.49	0.25	0.65	-2.60	-3.60
Degree 3	-0.27	-1.03	0.39	0.48	0.19	0.45	-2.19	-3.02
Unweighted Local Polynomial Regression:								
Degree 0	2.00	2.77	3.57	5.56	5.73	0.31	1.83	-10.22
Degree 1	2.02	1.06	2.63	2.61	7.46	5.57	4.33	-10.43
Degree 2	2.26	1.07	2.88	1.36	4.16	3.81	1.62	-2.94
Degree 3	2.21	1.01	2.94	1.46	3.45	3.65	0.96	-2.63
Weighting Within Cell	-1.15	1.70	-0.47	5.16	2.69	-6.91	3.06	-5.88

There are still a number of open issues that need to be further investigated with respect to implementation of the method in actual surveys, whether in the univariate case described in detail here or in the various model extensions just mentioned. An important practical issue is the selection of estimator settings such as the degree of the local polynomial and the bandwidth. As noted in the non-parametric literature (*e.g.*, Fan and Gijbels 1996, page 77) and also confirmed in the simulations, higher degree polynomials reduce the bias but increase the variance, so that polynomials of degree $k = 1$ or 2 are generally recommended as a good compromise. More critical is the choice of bandwidth parameter. In our simulations, the results were only modestly sensitive to the choice of bandwidth within a "reasonable" range of values, *i.e.*, ones ensuring that the number of observations used for estimating $\phi(x)$ at any x does not become too small (see discussion at the end of Section 2), or that is so large that the fit cannot capture changes in $\phi(\cdot)$ over the range of x. As a rule of thumb, we would recommend considering values for h that are within 20% and 50% of the range of x as a good place to start, and making a final determination by looking at both model diagnostics for the model fit $\hat{\phi}(x)$ and weight diagnostics for the adjusted survey weights $(\pi_i \hat{\phi}_i)^{-1}$, similarly as would be done when constructing cell-based weights.

Acknowledgements

We thank the Associate Editor and two referees for their useful comments. The first author was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil, under the grant Projeto Universal 480518/2004-1.

Appendix

A.1 Assumptions

We now state the assumptions needed to derive our main results. A detailed discussion of these assumptions is provided in Da Silva and Opsomer (2008). Consider the asymptotic framework of Section 3. Let $\boldsymbol{I}_\nu = (I_1, I_2, ..., I_{N_\nu})'$ be the sample inclusion indicator vector for the ν^{th} population. Suppressing the ν for ease of notation, let $\pi_i = \Pr(I_i = 1)$, and let

$$\Delta_{j_1, ..., j_k} \equiv \mathrm{E_d}\left(\prod_{\ell=1}^{k} (I_{j_\ell} - \pi_{j_\ell})\right) \qquad (19)$$

denote higher moments for the sample inclusion indicators $I_{j_1}, I_{j_2}, ..., I_{j_k}$ with respect to the sampling design. We assume that there are positive constants $\lambda_1, \lambda_2, ..., \lambda_6$ such that:

(A1) $\lambda_1 < N_\nu n_\nu^{-1} \pi_i < \lambda_2 < \infty, \forall\, i \in U_\nu$;

(A2) $N_\nu^{-1} n_\nu \to \pi$, for some $0 < \pi < 1$, as $\nu \to \infty$;

(A3) For distinct $j_1, j_2, ..., j_k \in U_\nu$, where $k = 2, 3, ..., 8$,

$$|\Delta_{j_1, ..., j_k}| \le \begin{cases} \left[\prod_{\ell=1}^{k} (N - \ell + 1)\right]^{-1} n_\nu^{\frac{k}{2}} \lambda_3 & \text{if } k \text{ is even,} \\ \left[\prod_{\ell=1}^{k} (N - \ell + 1)\right]^{-1} n_\nu^{\frac{k-1}{2}} \lambda_4 & \text{if } k \text{ is odd} \end{cases}$$

(A4) $\lim_{\nu \to \infty} N_\nu^{-1} \sum_{i \in U_\nu} y_i = \mu \in (-\infty, \infty)$ and $N_\nu^{-1} \sum_{i \in U_\nu} |y_i|^4 \le \lambda_5$, for all $\nu \ge 1$.

Let $\boldsymbol{R}_\nu = (R_1, R_2, ..., R_{N_\nu})'$ denote the response indicator vector for the ν-th population. In addition to the assumptions on the sampling design and the population distribution of the variable Y, we will also need the following assumptions on the response mechanism:

(B1) $R_1, R_2, ..., R_{N_\nu}$ are independent random variables;

(B2) $\Pr\{R_i = 1 \mid \boldsymbol{I}_\nu, \boldsymbol{y}_\nu, \boldsymbol{x}_\nu\} = \Pr\{R_i = 1 \mid \boldsymbol{x}_\nu\} \equiv \phi_i, \forall\, i \in U_\nu$;

(B3) $\phi_i = \phi(x_i), \forall\, i \in U_\nu$, where $\phi(\cdot)$ is a $(k+2)^{\text{th}}$ continuously differentiable function with $\lambda_6 < \phi(\cdot) \le 1$. The first derivative $\phi'(\cdot)$ has a finite number of sign changes.

Regarding the distribution of the x_i and the kernel estimator, we assume that:

(C1) For all $\nu \ge 1$, $x_1, x_2, ..., x_{N_\nu}$ are realizations of random variables $X_1, X_2, ..., X_{N_\nu}$ independent and identically distributed with distribution $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$, where $f_X(\cdot)$ is a continuous and positive probability density function on a compact set $[a_X, b_X]$;

(C2) The kernel function $K(\cdot)$ is a bounded and continuous probability density, which is symmetric around zero and supported on [-1, 1];

(C3) $\int_{-1}^{1} |z|^{k+4} K(z)\, dz < \infty$;

(C4) For all $\nu \ge 1$, $\{h_\nu\}$ is a sequence of bandwidths satisfying $0 < h_\nu \le 1$, $h_\nu \to 0$, $n_\nu h_\nu^2 \to \infty$ and $N_\nu h_\nu / \log N_\nu \to \infty$, as $\nu \to \infty$;

(C5) The first derivative $f_X'(\cdot)$ is continuously differentiable and contains a finite number of sign changes on $\mathrm{supp}(f_X)$. The first derivative $K'(\cdot)$ has a finite number of sign changes on $\mathrm{supp}(K)$;

(C6) The matrix $N_\nu \mathbf{T}_i^{-1}$ is non-singular for all $i \in U_\nu$ and all $\nu \geq 1$.

A.2 Technical derivations

Complete proofs are in Da Silva and Opsomer (2008). The proof of Theorem 1 relies on bounding the moments of the difference $\overline{y}_{\pi\hat{\psi}\nu} - \overline{y}_{\pi\tilde{\phi}\nu}$ under the combined design and response model probability mechanism, followed by deriving the rates of convergence for the bias and variance of the linearized estimator $\overline{y}_{\pi\hat{\psi}\nu}$. This is done in a series of six lemmas, which are stated here without proof. The proof of Theorem 2 is based on the result of Theorem 1, followed by an additional linearization of the ratio form.

For notational simplicity in what follows, we suppressed the fact that the results are conditional on the sequences $\boldsymbol{x}_\nu = (x_1, ..., x_{N_\nu})$ in the populations U_ν. However, the results in these lemmas are shown to hold with probability one over these sequences in Da Silva and Opsomer (2008), as was also done in Da Silva and Opsomer (2006). Hence, the results can be interpreted to hold for all population sequences, except on a set of probability 0 with respect to the distribution of the $\boldsymbol{x}_\nu$.

Lemma 1. *Assume that assumptions* (*C*1)-(*C*5) *hold. Consider* $\mu_\ell(K, x) = \int_{D_{x,h}} z^\ell K(z)\,dz$, *where* $D_{x,h_\nu} = \{t: (x + ht) \in \text{supp}(f_X)\} \cap \text{supp}(K)$. *Then, for all* $\ell = 0, 1, ..., k + 2$,

$$\sup_{x \in \text{supp}(f_X)} \left| \frac{1}{N_\nu h_\nu} \sum_{j \in U_\nu} K\left(\frac{X_j - x}{h_\nu}\right)(X_j - x)^\ell - E_\nu(x, \ell) \right| \underset{\nu \to \infty}{\to} 0,$$

where

$$E_\nu(x, \ell) = f_X(x)\mu_\ell(K, x)h_\nu^\ell + f_X'(x)\mu_{\ell+1}(K, x)h_\nu^{\ell+1} + o(h_\nu^{\ell+1}).$$

Lemma 2. *Assume that assumptions* (*C*1)-(*C*5) *hold. Consider the population fit* $\tilde{\phi}_i = \tilde{\phi}(x_i, k, h_\nu)$, $i \in U_\nu$, *defined in* (10). *Hence, for all* $i \in U_\nu$, *there exists positive bounded terms* $c_1(x_i)$, $c_2(x_i)$ *and* $c_3(x_i)$, *such that if* x_i *in an interior point of* $\text{supp}(f_X)$

$$\tilde{\phi}_i - \phi(x_i) = \begin{cases} c_1(x_i)h_\nu^{k+2} + o(h_\nu^{k+2}) & k \text{ is even} \\ c_2(x_i)h_\nu^{k+1} + o(h_\nu^{k+1}) & k \text{ is odd} \end{cases}$$

and if x_i *in a boundary point of* $\text{supp}(f_X)$

$$\tilde{\phi}_i - \phi(x_i) = c_3(x_i)h_\nu^{k+1} + o(h_\nu^{k+1}),$$

where all the smaller order terms hold uniformly in $i \in U_\nu$.

Lemma 3. *Assume that assumptions* (*C*1) *and* (*C*4) *hold. Then*,

i) *For* $p \in [0, \infty)$ *fixed*,

$$\limsup_{\nu \to \infty} \left(\frac{1}{N_\nu h_\nu} \sum_{j \in U_\nu} I_{\{x - h_\nu \leq x_j \leq x + h_\nu\}} \right)^p < \infty,$$

uniformly in x;

ii) $\limsup_{\nu \to \infty} \frac{1}{2N_\nu h_\nu} \sum_{j \in U_\nu} I_{\{x_j \in [0, h_\nu] \cup (1 - h_\nu, 1]\}} < \infty$;

iii) $\limsup_{\nu \to \infty} \frac{1}{N_\nu} \sum_{j \in U_\nu} I_{\{x_j \in (h_\nu, 1 - h_\nu]\}} < \infty$.

iv) *there exists* ν^*, *independent of* x, *such that whenever* $\nu \geq \nu^*$,

$$\sum_{j \in U_\nu} I_{\{|x_j - x| \leq h_\nu\}} \geq k + 1;$$

Lemma 4. *Suppose the assumptions of Theorem 1 hold. Consider the matrices* $\hat{\mathbf{T}}_{si} = \{\hat{T}_{si,\,pq}\}$ *and* $\mathbf{T}_i = \{T_{si,\,pq}\}$ *and the vectors* $\hat{\mathbf{t}}_{si} = \{\hat{t}_{si,\,p}\}$, $\mathbf{t}_i = \{t_{i,\,p}\}$ *and* $\mathbf{B}_i = \{B_{i,\,p}\}$ *given in* (7) *and* (10). *Then*,

i) *the* $N_\nu^{-1} T_{i,\,pq}$ *and* $N_\nu^{-1} t_{i,\,p}$ *are uniformly bounded in* $i \in U_\nu$, *for all* $p, q = 1, ..., k + 1$;

ii) *the* $\hat{T}_{si,\,pq}$ *and* $\hat{t}_{si,\,p}$ *satisfy*

$$\max_{1 \leq p,\, q \leq k+1} \text{E}\left(\frac{\hat{T}_{si,\,pq} - T_{i,\,pq}}{N_\nu} \right)^8 = O\left(\frac{1}{n_\nu^4 h_\nu^4} \right) \text{ and}$$

$$\max_{1 \leq p \leq k+1} \text{E}\left(\frac{\hat{t}_{si,\,p} - t_{i,\,p}}{N_\nu} \right)^8 = O\left(\frac{1}{n_\nu^4 h_\nu^4} \right),$$

uniformly in $i \in U_\nu$;

iii) *the random variable* $\boldsymbol{e}_1' \mathbf{T}_i^{-1}(\hat{\mathbf{t}}_{si} - \hat{\mathbf{T}}_{si}\mathbf{B}_i)$ *satisfies*

$$\max_{i \in U_\nu} \text{E}\,(\boldsymbol{e}_1' \mathbf{T}_i^{-1}(\hat{\mathbf{t}}_{si} - \hat{\mathbf{T}}_{si}\mathbf{B}_i) I_i R_i) = O\left(\frac{1}{n_\nu h_\nu} \right) \quad (20)$$

and

$$\max_{i \in U_\nu} \text{E}\,(\boldsymbol{e}_1' \mathbf{T}_i^{-1}(\hat{\mathbf{t}}_{si} - \hat{\mathbf{T}}_{si}\mathbf{B}_i))^4 = O\left(\frac{1}{n_\nu^2 h_\nu^2} \right). \quad (21)$$

Lemma 5. *Suppose the assumptions of Theorem 1 hold. Then, for all* $\nu \geq 1$

i) *the reciprocal of* $\tilde{\phi}_i$ *is uniformly bounded in* $i \in U_\nu$;

ii) *the partial derivatives of* $\hat{\phi}_i^{-1}$ *of orders one up to four, when evaluated at* $\hat{\mathbf{T}}_{si} = \mathbf{T}_i$, $\hat{\mathbf{t}}_{si} = \mathbf{t}_i$, $\delta_1 = 0$ *and* $\delta_2 = 0$, *are uniformly bounded in* $i \in U_\nu$;

iii) $\text{E}(\hat{\phi}_i^{-4})$ *is uniformly bounded in* $i \in U_\nu$;

iv) *the reciprocal of* $\hat{\phi}_i$ *satisfies*

$$\hat{\phi}_i^{-1} = \tilde{\phi}_i^{-1} - \tilde{\phi}_i^{-2}\, \boldsymbol{e}_1'\, \mathbf{T}_i^{-1}(\hat{\mathbf{t}}_{si} - \hat{\mathbf{T}}_{si}\, \mathbf{B}_i) + \varepsilon_{iv} + O\left(\frac{1}{N_v^2\, h_v^2}\right), \tag{22}$$

uniformly in $i \in U_v$, *where the* ε_{iv} *are random variables such that*

$$\max_{i \in U_v} \mathrm{E}\left(\varepsilon_{iv}^2\right) = O\left(\frac{1}{n_v^2 h_v^2}\right).$$

Lemma 6. *Suppose the assumptions of Theorem 1 hold. Define the random variables* $\overline{y}_{\pi\tilde{\phi}v}$, $\overline{d}_{\pi\tilde{\phi}v}$ *and* $\overline{\varepsilon}_{\pi\tilde{\phi}v}$ *as*

$$(\overline{y}_{\pi\tilde{\phi}v},\ \overline{d}_{\pi\tilde{\phi}v},\ \overline{\varepsilon}_{\pi\tilde{\phi}v})' = \frac{1}{N_v} \sum_{i \in s_v} \pi_i^{-1} \tilde{\phi}_i^{-1} (1,\ \tilde{\phi}_i^{-1} e_1' \mathbf{T}_i^{-1} (\hat{\mathbf{t}}_{si} - \hat{\mathbf{T}}_{si} \mathbf{B}_i),\ \varepsilon_{iv})' y_i R_i.$$

Then,

$$\mathrm{E}(\overline{y}_{\pi\tilde{\phi}v} - \overline{y}_{N_v}) = \begin{cases} O(h_v^{k+(3/2)}) & k \text{ even}, \\ O(h_v^{k+1}) & k \text{ odd}, \end{cases} \tag{23}$$

$$\mathrm{Var}(\overline{y}_{\pi\tilde{\phi}v}) = O\left(\frac{1}{n_v}\right), \tag{24}$$

$$(\mathrm{E}[\overline{d}_{\pi\tilde{\phi}v}],\ \mathrm{E}[\overline{d}_{\pi\tilde{\phi}v}^2 \hat{A}])' = O\left(\frac{1}{n_v h_v}\right) \tag{25}$$

and

$$\mathrm{E}(\overline{\varepsilon}_{\pi\tilde{\phi}v}^2) = O\left(\frac{1}{n_v^2\, h_v^2}\right). \tag{26}$$

References

Alho, J.M. (1990). Adjusting for nonresponse bias using logistic regression. *Biometrika*, 617-624.

Berger, Y.G., and Skinner, C.J. (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society*, Series B: *Statistical Methodology*, 67(1), 79-89.

Breidt, F.J., and Opsomer, J.D. (2000). Local polynomial regression estimators in survey sampling. *Annals of Statistics*, 1026-1053.

Cassel, C.-M., Särndal, C.-E. and Wretman, J.H. (1983). Some uses of statistical models in connection with the nonresponse problem. In *Incomplete data in sample surveys*: *Theory and bibliographies*, (Eds., W.G. Madow, I. Olkin and D.B. Rubin). Academic Press, New York: London, 3, 143-160.

Da Silva, D.N., and Opsomer, J.D. (2006). A kernel smoothing method of adjusting for unit non-response in sample surveys. *The Canadian Journal of Statistics*, 4, 563-579.

Da Silva, D.N., and Opsomer, J.D. (2008). Theoretical properties of propensity weighting for survey nonresponse through local polynomial regression. Technical Report #2008/6, Department of Statistics, Colorado State University.

David, M.H., Little, R., Samuhel, M. and Triest, R. (1983). Imputation models based on the propensity to respond. In *ASA Proceedings of the Business and Economic Statistics Section*, 168-173.

Ekholm, A., and Laaksonen, S. (1991). Weighting via response modeling in the Finnish Household Budget Survey. *Journal of Official Statistics*, 325-337.

Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.

Folsom, R.E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In *ASA Proceedings of the Social Statistics Section*, 197-202.

Giommi, A. (1984). A simple method for estimating individual response probabilities in sampling from finite populations. *Metron*, 4, 185-200.

Groves, R., Dillman, D., Eltinge, J. and Little, R.J.A. (2002). *Survey Nonresponse*. New York: John Wiley & Sons, Inc.

Hastie, T.J., and Tibshirani, R.J. (1986). Generalized additive models. *Statistical Science*, 297-318.

Iannacchione, V.G., Milne, J.G. and Folsom, R.E. (1991). Response probability weight adjustments using logistic regression. In *ASA Proceedings of the Section on Survey Research Methods*, 637-642.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 89-96.

Kim, J.K., and Kim, J.J. (2007). Nonresponse weighting adjustment using estimated response probability. *The Canadian Journal of Statistics*, 4, 501-514.

Laaksonen, S. (2006). Does the choice of link function matter in response propensity modelling? *Model Assisted Statistics and Applications*, 2, 95-100.

Lohr, S.L. (1999). *Sampling*: *Design and Analysis*. Pacific Grove: Duxbury Press.

Nargundkar, M., and Joshi, G.B. (1975). Non-response in sample surveys. In *40th Session of the ISI*, *Warsaw 1975*, *Contributed papers*, 626-628.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustments for unit non-response. In *Incomplete data in sample surveys*, (Eds., W.G. Madow, I. Olkin and D.B. Rubin), *Theory and bibliographies*, Academic Press, New York: London. 2, 143-184.

Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 41-55.

Rust, K. (1985). Variance estimation for complex estimators in sample survey. *Journal of Official Statistics*, 381-397.

Wand, M.P., and Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman and Hall.