Theoretical properties of propensity weighting for survey nonresponse through local polynomial regression

Damião N. da Silva *   Jean D. Opsomer †

August 28, 2008

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 and probit analysis. The resulting adjusted estimators may become biased when the specified parametric models are incorrect. To avoid misspecifying such a model, we consider the nonparametric estimation of the response probabilities by local polynomial regression. We describe theoretical and practical properties of the proposed approach.

Key Words: missing data, propensity scores, unit nonresponse, kernel regression, weighting adjustment.

*Departamento de Estatística, Campus Universitário, Natal, RN 59078-970 Brazil, damiao@ccet.ufrn.br.

†Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA, jopsomer@stat.colostate.edu.

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 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 et al. (1983).

General descriptions of propensity weighting to adjust classical survey estimators for nonresponse can be seen, for example, in Nargundkar and Joshi (1975), Cassel et al. (1983) and Groves et al. (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 et al. (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 to be specified. In this sense, these methods offer an appealing alternative to having to choose a suitable 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.

Extensions for the results in Da Silva and Opsomer (2006) were studied by Da Silva and Opsomer (2008) with the estimation the response propensities by local polynomial regression. This nonparametric technique (see Wand and Jones (1995), for instance) improves the local approximation to the unknown propensity function, resulting in better practical and theoretical properties compared to kernel smoothing.

In this report, we give detailed derivations for the results stated in Da Silva and Opsomer (2008). In Section 2, we introduce the weighting procedure and the estimation of the response propensities.

The theoretical statistical properties of the adjusted estimators are discussed in Section 3. Finally, in the Appendix, we state the assumptions needed and give proofs for the main results in the paper and also for the required lemmas.

2 WEIGHTING BY LOCAL POLYNOMIAL REGRESSION

Consider a population of N_ν units, denoted by $U_\nu = \{1, 2, \ldots, N_\nu\}$. Suppose that a probabilistic sample s_ν is drawn from U_ν, according to some probabilistic sampling design $p(s_\nu)$. Let n_ν be the size of s_ν and $\pi_i = \pi_{i\nu} = \Pr\{i \in s_\nu\} = \sum_{s_\nu : i \in s_\nu} p(s_\nu)$ be the inclusion probability of unit i, for all $i \in U_\nu$. It is of interest to estimate the population mean of a study variable y, namely $\bar{y}_{N_\nu} = N_\nu^{-1} \sum_{i \in U_\nu} y_i$, where y_i denotes the value of y for the i–th unit of U_ν. We assume that the values x_i of an auxiliary variable x are fully observed throughout the sample.

When the sample contains unit nonresponse, then we only observe the values of the study variables for the units in a subset $r_\nu \subset s_\nu$. 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 $i \in s_\nu$. We assume that, given the sample is fixed, the response indicators are independent Bernouilli random variables with

$$\Pr\{R_i = 1 \,|\, i \in s_\nu, \boldsymbol{y}, \boldsymbol{x}\} = \phi(x_i) \equiv \phi_i, \quad \text{for all } i \in s_\nu, \tag{1}$$

where the exact form of $\phi(\cdot)$ is unspecified, but it is assumed to be a smooth function of x_i with $\phi(\cdot) \in (0, 1]$. As a consequence of the analogy between the response probabilities ϕ_i in (1) and the the propensity scores of the Rosenbaum and Rubin (1983), we shall refer to the function $\phi(\cdot)$ as the response propensity function.

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. The theory developed here, therefore, does not intend to handle non–ignorable response mechanisms. Another important aspect of condition (1) is that the ϕ_i are allowed to vary, depending on the auxiliary variable only through the value x_i. Thus, (1) defines a form of missing–at–random nonresponse model according to the taxonomy given in Little and Rubin (2002).

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

$$\bar{y}_{\pi\phi\nu} = \frac{1}{N_\nu} \sum_{i \in s_\nu} \pi_{i\nu}^{-1} \phi_i^{-1} y_i R_i \tag{2}$$

and

$$\bar{y}_{rat,\pi\phi\nu} = \sum_{i\in s_\nu} \pi_{i\nu}^{-1}\phi_i^{-1} y_i R_i \bigg/ \sum_{i\in s_\nu} \pi_{i\nu}^{-1}\phi_i^{-1} R_i, \tag{3}$$

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, Särndal, and Wretman 1983).

Estimators (2) and (3) are unbiased and nearly unbiased for $\bar{y}_{N_\nu}$ 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 $\widehat{\phi}_i$, satisfying $0 < \widehat{\phi}_i \leq 1$. The resulting propensity weighting estimators are therefore

$$\bar{y}_{\pi\widehat{\phi}\nu} = \frac{1}{N_\nu} \sum_{i\in s_\nu} \pi_i^{-1}\widehat{\phi}_i^{-1} y_i R_i \tag{4}$$

and

$$\bar{y}_{rat,\pi\widehat{\phi}\nu} = \sum_{i\in s_\nu} \pi_i^{-1}\widehat{\phi}_i^{-1} y_i R_i \bigg/ \sum_{i\in s_\nu} \pi_i^{-1}\widehat{\phi}_i^{-1} R_i. \tag{5}$$

The latter formula has the advantages of being location–scale invariant, because the summation of its adjusted weights $\pi_i^{-1}\widehat{\phi}_i^{-1}R_i / \sum_{i\in s_\nu} \pi_i^{-1}\widehat{\phi}_i^{-1}R_i$ is equal to one, and that it does not require the population size N_ν to be known.

In order to implement the propensity weighting estimators (4) and (5), it is necessary to estimate the response propensities $\widehat{\phi}_i$. 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_ν be its bandwidth. Define the $N_\nu \times (k+1)$ matrix

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

the $N_\nu \times N_\nu$ matrix

$$\mathbf{W}_{Ui} = \mathrm{diag}\left\{ \frac{1}{h_\nu} K\left(\frac{x_j - x_i}{h_\nu}\right) \,:\, 1 \leq j \leq N_\nu \right\}.$$

and population vector of response indicators $\mathbf{R}_U = (R_1, R_2, \ldots, R_{N_\nu})'$. The vector $\mathbf{R}_U$ would be known if, instead of the sample s_ν, a census was considered from the population U_ν. In this 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

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

where $\boldsymbol{e}_j$ 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

$$\widehat{\phi}_i^o = \boldsymbol{e}_1{}'\widehat{\mathbf{T}}_{si}^{-1}\widehat{\mathbf{t}}_{si} \tag{7}$$

where

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

and it is assumed that $\widehat{\mathbf{T}}_{si}$ is invertible. An special case of (7) is obtained by considering $k=0$, which implies in the fit of a constant locally. This procedure corresponds to the kernel regression response propensity estimator studied in detail by Da Silva and Opsomer (2006). Other potential 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 $\widehat{\mathbf{T}}_{si}$ happens to be singular, a simple procedure to insure that $\widehat{\phi}_i^o$ 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

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

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

$$\widehat{\phi}_i = \max\left\{\widehat{\phi}(x_i, k, h_\nu),\, \delta_2\left(N_\nu h_\nu\right)^{-1}\right\}, \tag{9}$$

for some constant $\delta_2 > 0$. This idea is related to the adjustment made by Da Silva and Opsomer (2006) regarding the local constant response probability 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. 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_ν ($n_\nu \geq n_{\nu-1}$) is selected according to a sampling design $p_\nu(\cdot)$. This framework is commonly adopted in asymptotic studies of survey estimators. See, for instance, Isaki and Fuller (1982), for an early reference. For what follows, we define $\mathbf{x}_\nu = (x_1, x_2, \ldots, x_{N_\nu})'$ and let $P_{\mathbf{X}}$ be the joint probability distribution of $(\mathbf{x}_1, \mathbf{x}_2, ...)'$. The results should be interpreted as statements with probability one regarding the possible sequences of values for the auxiliary variable x in the population. Although the statements depend on the distribution of x values, they reflect inferences for the finite population because the results are valid for almost any sequence of U_ν.

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

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

where

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

the matrices $\mathbf{X}_{Ui}$ and $\mathbf{W}_{Ui}$ are as in (6) and $\boldsymbol{\phi}_U = (\phi(x_1), \phi(x_2), \ldots, \phi(x_{N_\nu}))'$. Also, as a consequence of the Taylor expansion for $\widehat{\phi}_i^{-1}$, given in Lemma 5, we will use a limiting approximation for (4) based on the decomposition

$$\bar{y}_{\pi\widehat{\phi}\nu} \;=\; \bar{y}_{\pi\tilde{\phi}\nu} \;-\; \bar{d}_{\pi\tilde{\phi}\nu} \;+\; \bar{\epsilon}_{\pi\tilde{\phi}\nu} \;+\; O\left(\frac{1}{N_\nu^2 h_\nu^2}\right), \tag{11}$$

where

$$\left(\bar{y}_{\pi\tilde{\phi}\nu}, \bar{d}_{\pi\tilde{\phi}\nu}, \bar{\epsilon}_{\pi\tilde{\phi}\nu}\right)' = \frac{1}{N_\nu}\sum_{i\in s_\nu} \pi_i^{-1}\tilde{\phi}_i^{-1}\left(1, \tilde{\phi}_i^{-1}\boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si} - \widehat{\mathbf{T}}_{si}\mathbf{B}_i), \epsilon_{i\nu}\right)' y_i R_i,$$

$\epsilon_{i\nu}$ is the remainder term in a Taylor expansion and the smaller order term $O(N_\nu^{-2}h_\nu^{-2})$ hold uniformly over all samples and all response indicators.

Next, it follows the statistical properties of the estimator $\bar{y}_{\pi\widehat{\phi}\nu}$.

Theorem 1. *Assume the assumptions (A1)–(A4), (B1)–(B3) and (C1)–(C5) hold. Consider the estimation of the population mean* $\bar{y}_{N_\nu}$ *by the propensity weighting estimator* $\bar{y}_{\pi\widehat{\phi}\nu}$*, defined in (4),* *and suppose the response propensities are estimated by the local polynomial regression estimator of degree k, $\widehat{\phi}_i$, given in (9). Let*

$$\bar{y}_{\pi\widehat{\psi}\nu} = \frac{1}{N_\nu} \sum_{i \in s_\nu} \pi_i^{-1} \widehat{\psi}_i^{-1} y_i R_i, \tag{12}$$

where

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

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

$$\mathrm{E}\left[\left(\bar{y}_{\pi\widehat{\phi}\nu} - \bar{y}_{\pi\widehat{\psi}\nu} \right)^2 \,\middle|\, \mathbf{x}_\nu \right] \;=\; O\left(\frac{1}{n_\nu^2 h_\nu^2} \right), \tag{13}$$

with $P_{\mathbf{X}}$–probability one, and the bias and variance of $\bar{y}_{\pi\widehat{\psi}\nu}$ satisfy

$$B_\nu \;\equiv\; \mathrm{E}\left[\bar{y}_{\pi\widehat{\psi}\nu} - \bar{y}_{N_\nu} \,\middle|\, \mathbf{x}_\nu \right] \;=\; \begin{cases} O\left(h_\nu^{k+(3/2)} \right) \;+\; O\left(\frac{1}{n_\nu h_\nu} \right), & k \text{ even}, \\ O\left(h_\nu^{k+1} \right) \;+\; O\left(\frac{1}{n_\nu h_\nu} \right), & k \text{ odd}, \end{cases} \tag{14}$$

and

$$\mathrm{Var}\left[\bar{y}_{\pi\widehat{\psi}\nu} \,\middle|\, \mathbf{x}_\nu \right] \;=\; O\left(\frac{1}{n_\nu h_\nu} \right), \tag{15}$$

with $P_{\mathbf{X}}$–probability one.

Remark 1. Theorem 1 builds a limiting approximation for the local polynomial response propensity estimator $\bar{y}_{\pi\widehat{\phi}\nu}$, given by the random variable $\bar{y}_{\pi\widehat{\psi}\nu}$ in (12). The approximation consists of adjusting the sampling inclusion probabilities with the weights $\widehat{\psi}_i^{-1}$. The first component in the formula defining these weights is asymptotically equivalent to the inverse of true response propensities ϕ_i^{-1}. However, since the following term in $\widehat{\psi}_i^{-1}$ is not small enough to be discarded, approximating the properties of $\bar{y}_{\pi\widehat{\phi}\nu}$ by plugging in estimates $\widehat{\phi}_i$ in the properties of $\bar{y}_{\pi\phi\nu}$ of (2) might lead to invalid asymptotic inferences. One imediate consequence of this comment is that the simple approach of plugging in an estimate of ϕ_i in

$$\mathrm{Var}\left[\bar{y}_{\pi\phi\nu} \,\middle|\, \mathbf{x}_\nu \right] = \mathrm{Var}\left[\bar{y}_{\pi\nu} \right] + \frac{1}{N_\nu^2} \sum_{i \in s_\nu} \frac{(1 - \phi_i)}{\pi_i^2 \phi_i} y_i,$$

where the first component in the right–hand side is the design variance of the Horvitz–Thompson estimator, may not yield a satisfactory variance estimator.

Remark 2. Results (13) and (14) imply that the propensity weighting estimator $\bar{y}_{\pi\widehat{\phi}\nu}$, using a response propensity estimator based on local polynomial regression, is asymptotically unbiased for the population mean $\bar{y}_{N_\nu}$ under the joint distribution of the sampling design and the response model (1). This result follows directly from assumption (C4), because both terms in the right–hand side of (14) converges to zero, as $\nu \to \infty$, with $P_{\mathbf{X}}$–probability one. Combining this result with (15), then we obtain that

$$\widehat{y}_{\pi\widehat{\phi}\nu} = \bar{y}_{N_\nu} + O_p\left(\frac{1}{\sqrt{n_\nu h_\nu}}\right), \tag{16}$$

with $P_{\mathbf{X}}$–probability one, when the bandwidth satisfies

$$h_\nu = \begin{cases} O\left(n_\nu^{-\frac{1}{2k+4}}\right), & k \text{ even}, \\ O\left(n_\nu^{-\frac{1}{2k+3}}\right), & k \text{ odd}. \end{cases} \tag{17}$$

Hence, without assuming a parametric form for the response propensity function $\phi(\cdot)$, $\bar{y}_{\pi\widehat{\phi}\nu}$ is also consistent for the population mean, with respect to the sampling design and the response model. As a price paid, however, the rate of convergence here is of order $\sqrt{n_\nu h_\nu}$, therefore slower than the usual rate $\sqrt{n_\nu}$.

Remark 3. Two theoretical properties also given in Theorem 1 are the asymptotic rates by which the bias and variance of the limiting approximation for $\bar{y}_{\pi\widehat{\phi}\nu}$ converge to zero. Fixing the order of the polynomial fit, the bandwidth that minimizes the asympotic mean square error satisfies (17). When the order of the local polynomial estimator is even, the rate for the leading term in the asymptotic bias is at most $h_\nu^{k+3/2}$, as opposed to the usual rate h_ν^{k+2} of local polynomial regression estimation at a single x value. This result is a direct extension of the one given in Da Silva and Opsomer (2006) for the kernel regression case. It is a consequence of the fact the leading bias term is an average of the population values of study variable y_i and the local polynomial estimates of the response propensity function obtained for all interior and boundary x_i values. The disagreement betweeen the rates follows from the chosen framework for the y_i values in (A4), where it is assumed only that those values have a finite fourth population moment. However, by strengthening this assumption to uniform boundedness, the obtained rate will be the same as the usual one.

Corollary 1. *Assume the conditions of Theorem 1 hold. Suppose that the sampling design and the response model are such that, conditional on $\boldsymbol{x}_\nu$,*

$$\frac{\bar{y}_{\pi\widehat{\psi}\nu} - \bar{y}_{N_\nu} - B_\nu}{\left[\operatorname{Var}\left(\bar{y}_{\pi\widehat{\psi}\nu} \,\middle|\, \boldsymbol{x}_\nu\right)\right]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad a.s.\ P_{\boldsymbol{X}}, \quad as\ \nu \to \infty,$$

where B_ν is defined in (14). If additionally

$$\lim_{\nu\to\infty} (n_\nu h_\nu)\operatorname{Var}\left(\bar{y}_{\pi\widehat{\psi}\nu} \,\middle|\, \boldsymbol{x}_\nu\right) \in (0,\infty) \ \ a.s.\ P_{\boldsymbol{X}},$$

then, conditional on $\boldsymbol{x}_\nu$,

$$\frac{\bar{y}_{\pi\widehat{\phi}\nu} - \bar{y}_{N_\nu} - B_\nu}{\left[\mathrm{Var}\big(\bar{y}_{\pi\widehat{\psi}\nu} \,\big|\, \boldsymbol{x}_\nu\big)\right]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad a.s.\ P_{\boldsymbol{X}},$$

as $\nu \to \infty$.

Remark 4. Corollary 1 states a Central Limit Theorem result for $\bar{y}_{\pi\widehat{\phi}\nu}$. Hence, apart from being consistent as discussed in Remark 2, this estimator has a sampling distribution asymptotically normal. This result follows from the assumption that the linearized random variable $\bar{y}_{\pi\widehat{\psi}\nu}$ satisfies the conditions of the Central Limit Theorem, under the joint distribution of the sampling design and the response model.

We now discuss the properties of the ratio–based version of propensity weighting estimator given in (5). Note that since

$$\bar{y}_{rat,\pi\widehat{\phi}\nu} - \bar{y}_{N_\nu} = \frac{N_\nu^{-1} \sum\limits_{i \in s_\nu} \pi_i^{-1} \widehat{\psi}_i^{-1} R_i (y_i - \bar{y}_{N_\nu})}{N_\nu^{-1} \sum\limits_{i \in s_\nu} \pi_i^{-1} \widehat{\psi}_i^{-1} R_i},$$

then standard ratio estimation theory can be used to derive asymptotic results for $\bar{y}_{rat,\pi\widehat{\phi}\nu}$. There is no difference between the asymptotic rates for the approximate bias and variance of $\bar{y}_{rat,\pi\widehat{\phi}\nu}$ and the rates given in Theorem 1 for the $\bar{y}_{\pi\widehat{\phi}\nu}$. However, an explicit expression for the asymptotic bias $\bar{y}_{rat,\pi\widehat{\phi}\nu}$ is given by

$$\mathrm{Bias}\big(\bar{y}_{rat,\pi\widehat{\phi}\nu}\big) = \frac{\sum\limits_{i \in U_\nu} \tilde{\phi}_i^{-1} \phi_i (y_i - \bar{y}_{N_\nu})}{\sum\limits_{i \in U_\nu} \tilde{\phi}_i^{-1} \phi_i},$$

which shows that the weaker is the population correlation between the values of $\tilde{\phi}_i^{-1}\phi_i$ and the values of y_i, the bigger will be the reduction if bias of $\bar{y}_{rat,\pi\widehat{\phi}\nu}$ as compared to the bias of $\bar{y}_{\pi\widehat{\phi}\nu}$. The asymptotic variance of $\bar{y}_{rat,\pi\widehat{\phi}\nu}$ can be approximated by

$$\mathrm{Var}\big(\bar{y}_{rat,\pi\widehat{\psi}\nu} \,\big|\, \mathbf{x}_\nu\big) = \mathrm{Var}\left(\frac{1}{N_\nu} \sum_{i \in s_\nu} \pi_i^{-1} \widehat{\psi}_i^{-1} R_i (y_i - \bar{y}_{N_\nu}) \,\big|\, \mathbf{x}_\nu\right),$$

which, as remarked by Da Silva (2003) for the zero order local polynomial case, has a complex analytical expression. This expression, however, has the same form of the corresponding expression of the $\bar{y}_{\pi\widehat{\phi}\nu}$ estimator, with the exception is that the residuals $y_i - \bar{y}_{N_\nu}$ replace the y_i values, leading therefore to a generally smaller asymptotic variance. An asymptotic approximation for the distribution of $\bar{y}_{rat,\pi\widehat{\phi}\nu}$ is given next.

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

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

where the weights $\widehat{\psi}_i^{-1}$ *are given in Theorem 1. Suppose that, conditional on* $\boldsymbol{x}_\nu$,

$$\frac{\bar{e}_{\pi\widehat{\psi}\nu} - \mathrm{E}\big(\bar{e}_{\pi\widehat{\psi}\nu}\,\big|\,\boldsymbol{x}_\nu\big)}{\Big[\mathrm{Var}\big(\bar{e}_{\pi\widehat{\psi}\nu}\,\big|\,\boldsymbol{x}_\nu\big)\Big]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad a.s.\ P_{\boldsymbol{X}}, \quad as\ \nu\to\infty,$$

and

$$\lim_{\nu\to\infty}(n_\nu h_\nu)\mathrm{Var}\big(\bar{e}_{\pi\widehat{\psi}\nu}\,\big|\,\boldsymbol{x}_\nu\big) \in (0,\infty)\ \ a.s.\ P_{\boldsymbol{X}}.$$

Then, conditional on $\boldsymbol{x}_\nu$,

$$\frac{\bar{y}_{rat,\pi\widehat{\phi}\nu} - \bar{y}_{N_\nu} - B_{rat,\nu}}{\Big[\mathrm{Var}\big(\bar{e}_{\pi\widehat{\psi}\nu}\,\big|\,\boldsymbol{x}_\nu\big)\Big]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad a.s.\ P_{\boldsymbol{X}},$$

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

A 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 (2003). Consider the asymptotic framework of Section 3. Let $\boldsymbol{I}_\nu = (I_1, I_2, \cdots, I_{N_\nu})'$ 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

$$\Delta_{j_1,\cdots,j_k} \equiv \mathrm{E}_{\mathrm{d}}\Big(\prod\nolimits_{\ell=1}^{k}(I_{j_\ell} - \pi_{j_\ell})\Big) \tag{18}$$

denote higher moments for the sample inclusion indicators $I_{j_1}, I_{j_2}, \cdots, I_{j_k}$ with respect to the sampling design. We assume that there are positive constants $\lambda_1, \lambda_2, \ldots, \lambda_6$ 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 $j_1, j_2, \cdots, j_k \in U_\nu$, where $k = 2, 3, \cdots, 8$,

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

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

Let $\boldsymbol{R}_\nu = (R_1, R_2, \cdots, R_{N_\nu})'$ 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) $R_1, R_2, \cdots, R_{N_\nu}$ are independent random variables;

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

(B3) $\phi_i = \phi(x_i),\ \forall\ i \in U_\nu$, where $\phi(\cdot)$ is a $(k+2)$-th continuously differentiable function with $\lambda_6 < \phi(\cdot) \leq 1$. 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 $\nu \geq 1$, $x_1, x_2, \cdots, x_{N_\nu}$ are realizations of random variables $X_1, X_2, \cdots, X_{N_\nu}$ 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 $[a_X, b_X]$. Without loss of generality, we shall take $[a_X, b_X] \equiv [0,1]$;

(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 $\nu \geq 1$, $\{h_\nu\}$ is a sequence of bandwidths satisfying $0 < h_\nu \leq 1$, $h_\nu \to 0$, $n_\nu h_\nu^2 \to \infty$ 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) With $P_{\boldsymbol{X}}$–probability one, the matrix $N_\nu \mathbf{T}_i^{-1}$ is non–singular for all $i \in U_\nu$ and all $\nu \geq 1$.

A.2 Proof of Theorems in Section 3

Proof of Theorem 1: From (12) and the decomposition (11), we have that

$$\bar{y}_{\pi\widehat{\psi}\nu} \;=\; \bar{y}_{\pi\tilde{\phi}\nu} - \bar{d}_{\pi\tilde{\phi}\nu},$$

Hence, (13)–(15) follow from Lemma 6, by noting that

$$\left|\mathrm{E}\big[\,(\bar{y}_{\pi\widehat{\phi}\nu} - \bar{y}_{\pi\widehat{\psi}\nu})^2 \,\big|\, \mathbf{x}_\nu\big]\right| \;\leq\; 2\left\{\mathrm{E}\big[\,\bar{\epsilon}^2_{\pi\tilde{\phi}\nu} \,\big|\, \mathbf{x}_\nu\big] \;+\; O\left(\frac{1}{N_\nu^4 h_\nu^4}\right)\right\},$$

$$\left|\mathrm{E}\left[\,\bar{y}_{\pi\hat{\psi}\nu} - \bar{y}_{N_\nu}\,\middle|\,\mathbf{x}_\nu\right]\right| \;\;\leq\;\; \left|\mathrm{E}\left[\,\bar{y}_{\pi\tilde{\phi}\nu} - \bar{y}_{N_\nu}\,\middle|\,\mathbf{x}_\nu\right]\right| \;+\; \left|\mathrm{E}\left[\,\bar{d}_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right]\right|$$

and

$$\begin{aligned}\mathrm{Var}\left[\,\bar{y}_{\pi\hat{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right] \;\;\leq\;\; & \mathrm{Var}\left[\,\bar{y}_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right] \;+\; \mathrm{E}\left[\,\bar{d}^2_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right] \;+ \\ & \mathrm{E}\left[\,\bar{\epsilon}^2_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right] \;+\; 2\left\{\mathrm{Var}\left[\,\bar{y}_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right]\mathrm{E}\left[\,\bar{d}^2_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right]\right\}^{1/2} \;+ \\ & 2\left\{\mathrm{Var}\left[\,\bar{y}_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right]\mathrm{E}\left[\,\bar{\epsilon}^2_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right]\right\}^{1/2} \;+\; 2\left\{\mathrm{E}\left[\,\bar{d}^2_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right]\mathrm{E}\left[\,\bar{\epsilon}^2_{\pi\tilde{\phi}\nu}\,\middle|\,\mathbf{x}_\nu\right]\right\}^{1/2}.\end{aligned}$$

□

A.3 Lemmas

Lemma 1. *Assume that assumptions (C1)–(C5) hold. Consider* $\mu_\ell(K,x) = \int_{D_{x,h}} z^\ell K(z)dz$, *where* $D_{x,h_\nu} = \{t : (x+ht) \in \mathrm{supp}(f_X)\} \cap \mathrm{supp}(K)$. *Then, for all* $\ell = 0, 1, \ldots, k+2$,

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

with $P_{\mathbf{X}}$*–probability one, where*

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

Proof of Lemma 1: This lemma corresponds to a slight generalization of some results in Opsomer (1995) who used the theory developed by Pollard (1984) to establish uniform convergence of kernel–based averages to their corresponding population moments. The idea of the proof is as follows: first fix $\ell = 0, 1, \cdots, k+2$. Then, applying Lemma A.4 of Opsomer (1995, p. 114) and assumptions (C1), (C2), C(4) and (C5), the conditions of Theorem 37 (Pollard 1984, p. 34) are verified. This implies that, as $\nu \to \infty$,

$$\sup_{x\in\mathrm{supp}(f_X)} \left|\frac{1}{N_\nu h_\nu}\sum_{j\in U_\nu} K\left(\frac{x_j - x}{h_\nu}\right)(x_j - x)^\ell - \frac{1}{h_\nu}\int_{\mathrm{supp}(f_X)} K\left(\frac{t-x}{h_\nu}\right)(t-x)^\ell f_X(t)dt\right| \;\;\to\;\; 0,$$

over all possible sequences $\{\mathbf{x}_\nu\}_{\nu\geq 1}$ with $P_{\mathbf{X}}$–probability one. Because the kernel function $K(\cdot)$ has finite moments, by (C3), and $f_X''(\cdot)$ is bounded on $\mathrm{supp}(f_X)$, by (C5), standard kernel regression arguments give directly that

$$\frac{1}{h_\nu}\int_{\mathrm{supp}(f_X)} K\left(\frac{t-x}{h_\nu}\right)(t-x)^\ell f_X(t)dt - E_\nu(x,\ell) = o(h_\nu^{\ell+1}),$$

uniformly for all $x \in \mathrm{supp}(f_X)$. Hence, the result follows.

□ (**Lemma 1**)

Lemma 2. *Assume that assumptions (C1)–(C5) 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* $\mathrm{supp}(f_X)$

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

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

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

where all the smaller order terms hold uniformly in $i \in U_\nu$*, a.s* $P_{\mathbf{X}}$*.*

Proof of Lemma 2: This lemma summarizes the rates by which the bias of $\tilde{\phi}_i$ as a population–based estimator of $\phi(x_i)$ converges to zero, as ν goes to infinity. These results can be derived following almost exaxtly the same steps as those in the the proof of Theorem 4.1 and the Remark 7 in Ruppert and Wand (1994), by adopting the extended definition of the kernel moments $\mu_\ell(K, x)$, given in Lemma 1. One difference here is the uniform convergence in $i \in U_\nu$ of the smaller order terms, which follows from Lemma 1. For simplicity, we omitted in the notation for $c_1(x_i)$, $c_2(x_i)$ and $c_3(x_i)$ their analytic expressions as well as their dependence on the kernel function $K(\cdot)$, the derivatives of the response propensity function $\phi(\cdot)$, the density $f_X(\cdot)$ and its first derivative $f'_X(\cdot)$. $\square$ (**Lemma 2**)

Lemma 3. *Assume that assumptions (C1) and (C4) hold. Then, with* $P_{\boldsymbol{X}}$*–probability one,*

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

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

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* $\nu \ge \nu^*$*,*

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

Proof of Lemma 3: Parts (i)–(iii) are given by (ii)–(iv) of Lemma 3.2 in Da Silva (2003). For part (iv), see Lemma 2(ii) in Breidt and Opsomer (2000). $\square$

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

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* $p, q = 1, \cdots, k+1$*;*

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

$$\max_{1 \le p,q \le k+1} \mathrm{E}\left[\left(\frac{\widehat{T}_{si,pq} - T_{i,pq}}{N_\nu}\right)^8 \,\middle|\, \mathbf{x}_\nu\right], \quad \max_{1 \le p \le k+1} \mathrm{E}\left[\left(\frac{\widehat{t}_{si,p} - t_{i,p}}{N_\nu}\right)^8 \,\middle|\, \mathbf{x}_\nu\right] = O\left(\frac{1}{n_\nu^4 h_\nu^4}\right),$$

uniformly in $i \in U_\nu$*;*

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

$$\max_{i \in U_\nu} \mathrm{E}\left\{\left[\boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si} - \widehat{\mathbf{T}}_{si}\mathbf{B}_i) I_i R_i\right] \,\middle|\, \mathbf{x}_\nu\right\} = O\left(\frac{1}{n_\nu h_\nu}\right) \tag{19}$$

and

$$\max_{i \in U_\nu} \mathrm{E}\left\{\left[\boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si} - \widehat{\mathbf{T}}_{si}\mathbf{B}_i)\right]^4 \,\middle|\, \mathbf{x}_\nu\right\} = O\left(\frac{1}{n_\nu^2 h_\nu^2}\right). \tag{20}$$

Proof of Lemma 4: (i) This corresponds to Lemma 2(iii) in Breidt and Opsomer (2000). (ii) Since, conditional on $\mathbf{x}_\nu$,

$$\frac{1}{N_\nu^8}\left(\widehat{T}_{si,pq} - T_{i,pq}\right)^8 = \frac{1}{N_\nu^8 h_\nu^8}\left[\sum_{j \in U_\nu} K\left(\frac{x_j - x_i}{h_\nu}\right)(x_j - x_i)^{p+q-2}\left(\frac{I_j - \pi_j}{\pi_j}\right)\right]^8$$

and that

$$\frac{1}{N_\nu^8}\left(\widehat{t}_{si,p} - t_{i,p}\right)^8 = \frac{1}{N_\nu^8 h_\nu^8}\left[\sum_{j \in U_\nu} K\left(\frac{x_j - x_i}{h_\nu}\right)(x_j - x_i)^{p-1}\left(\frac{I_j R_j - \pi_j \phi_j}{\pi_j}\right)\right]^8,$$

the prescribed order for their expected values is obtained by the same method applied in the proof of Lemma 3.4 of Da Silva (2003). (iii) Let $\mathbf{X}'_{si,j}$ denote the j–th row of $\mathbf{X}_{si}$. From (7) and (10),

$$\mathrm{E}\left[\frac{1}{N_\nu}(\widehat{\mathbf{t}}_{si} - \mathbf{t}_i) I_i R_i \,\middle|\, \mathbf{x}_\nu\right] = \frac{1}{N_\nu h_\nu}\sum_{j \in U_\nu} K\left(\frac{x_j - x_i}{h_\nu}\right)\mathbf{X}_{si,j}\frac{\mathrm{E}\left[I_i R_i (I_j R_j - \pi_j \phi_j) \,\middle|\, \mathbf{x}_\nu\right]}{\pi_j}$$

and

$$\mathrm{E}\left[\frac{1}{N_\nu}(\widehat{\mathbf{T}}_{si} - \mathbf{T}_i) I_i R_i \mathbf{B}_i \,\middle|\, \mathbf{x}_\nu\right] = \frac{1}{N_\nu h_\nu}\sum_{j \in U_\nu} K\left(\frac{x_j - x_i}{h_\nu}\right)\mathbf{X}_{si,j}\mathbf{X}'_{si,j}\mathbf{B}_i\frac{\mathrm{E}\left[I_i R_i (I_j - \pi_j) \,\middle|\, \mathbf{x}_\nu\right]}{\pi_j},$$

and evaluating the moments in these two expressions with respect to the quasi-randomization model, we can write

$$\begin{aligned}
\mathrm{E}\left\{\left[\boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si}-\widehat{\mathbf{T}}_{si}\mathbf{B}_i)I_iR_i\right]\,\middle|\,\mathbf{x}_\nu\right\} &= \mathrm{E}\left\{\boldsymbol{e}_1{}'[N_\nu\mathbf{T}_i^{-1}]\frac{1}{N_\nu}\left[(\widehat{\mathbf{t}}_{si}-\mathbf{t}_i)-(\widehat{\mathbf{T}}_{si}-\mathbf{T}_i)\mathbf{B}_i\right]I_iR_i\,\middle|\,\mathbf{x}_\nu\right\} \\
&= \frac{K(0)}{N_\nu h_\nu}\left(\boldsymbol{e}_1{}'[N_\nu\mathbf{T}_i^{-1}]\boldsymbol{e}_1\right)\frac{\pi_i\phi_i\left[(1-\pi_i\phi_i)-(1-\pi_i)\tilde{\phi}_i\right]}{\pi_i}+ \\
&\quad \frac{1}{N_\nu h_\nu}\sum_{j\in U_\nu, j\neq i}K\left(\frac{x_j-x_i}{h_\nu}\right)\frac{\Delta_{ij}}{\pi_j}\phi_i\left(\boldsymbol{e}_1{}'[N_\nu\mathbf{T}_i^{-1}]\mathbf{X}_{si,j}\right)\varepsilon_{i,j},
\end{aligned}$$

where $\varepsilon_{i,j}=\phi_j-\mathbf{X}'_{si,j}\mathbf{B}_i=\phi(x_j)-\sum_{q=1}^{k+1}(x_j-x_i)^{q-1}B_{i,q}$. Since the π_i, ϕ_i and $\tilde{\phi}_i$ are terms uniformly bounded in $i\in U_\nu$, by (A1), (B3) and Lemma 2, the $[N_\nu\mathbf{T}_i^{-1}]_{p,q}$ and $B_{i,q}$ terms are continuous functions of the uniformly bounded $T_{si,pq}$ and $t_{si,p}$, by (i), with denominators bounded away from zero, by Lemma 3(iv), and because of the compactness of $f_x(\cdot)$, by (C1),

$$\max_{i\in U_\nu}\left|\mathrm{E}\left\{\left[\boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si}-\widehat{\mathbf{T}}_{si}\mathbf{B}_i)I_iR_i\right]\,\middle|\,\mathbf{x}_\nu\right\}\right|\leq O\left(\frac{K(0)}{N_\nu h_\nu}\frac{N_\nu}{n_\nu}+\frac{\max\limits_{i\neq j\in U_\nu}|\Delta_{ij}|}{\min\limits_{i\in U_\nu}\pi_i}\left(\sum_{j\in U_\nu}\frac{I_{\{|x_j-x_i|\leq h_\nu\}}}{N_\nu h_\nu}\right)\right).$$

Hence, (19) follows by noting that the second term in the right–hand side is, by (A1)–(A3) and Lemma 3(i), of order at most n_ν^{-1} uniformly in $i\in U_\nu$, with $P_{\boldsymbol{X}}$–probability one.

To establish (20), apply the C_r inequality (Sen and Singer 1993, pp. 21) to show that

$$\begin{aligned}
\mathrm{E}\left\{\left[\boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si}-\widehat{\mathbf{T}}_{si}\mathbf{B}_i)\right]^4\,\middle|\,\mathbf{x}_\nu\right\} &\leq c_r\left\{\sum_{p=1}^{k+1}|[N_\nu\mathbf{T}_i^{-1}]_{1,p}|^4\mathrm{E}\left[\frac{1}{N_\nu^4}(\widehat{t}_{si,p}-t_{i,p})^4\,\middle|\,\mathbf{x}_\nu\right]+\right. \\
&\quad \left.\sum_{p=1}^{k+1}\sum_{q=1}^{k+1}|[N_\nu\mathbf{T}_i^{-1}]_{1,p}|^4\mathrm{E}\left[\frac{1}{N_\nu^4}(\widehat{T}_{si,pq}-T_{i,pq})^4\,\middle|\,\mathbf{x}_\nu\right]B_{i,q}^4\right\},
\end{aligned} \tag{21}$$

for all $i\in U_\nu$, where $c_r=[(k+1)(k+2)]^3$. Now, by the same reasoning as in (iii), the $[N_\nu\mathbf{T}_i^{-1}]_{1,p}$ and $B_{i,q}$ are uniformly bounded terms in $i\in U_\nu$. Hence, the result follows by applying (ii) and the Liapounov's inequality to conclude that the fourth moments in the right–hand side of (21) are of order $\left[O\left(n_\nu^{-4}h_\nu^{-4}\right)\right]^{1/2}$ uniformly in $i\in U_\nu$, with $P_{\boldsymbol{X}}$–probability one. $\square$

Lemma 5. *Suppose the assumptions of Theorem 1 hold. Then, with $P_{\boldsymbol{X}}$–probability one, 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 $\widehat{\phi}_i^{-1}$ of orders one up to four, when evaluated at $\widehat{\mathbf{T}}_{si}=\mathbf{T}_i$, $\widehat{\mathbf{t}}_{si}=\mathbf{t}_i$, $\delta_1=0$ and $\delta_2=0$, are uniformly bounded in $i\in U_\nu$;

iii) $\mathrm{E}\left[\widehat{\phi}_i^{-4}\,|\,\mathbf{x}_\nu\right]$ is uniformly bounded in $i\in U_\nu$;

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

$$\widehat{\phi}_i^{-1} = \tilde{\phi}_i^{-1} - \tilde{\phi}_i^{-2}\boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si} - \widehat{\mathbf{T}}_{si}\mathbf{B}_i) + \epsilon_{i\nu} + O\left(\frac{1}{N_\nu^2 h_\nu^2}\right), \tag{22}$$

uniformly in $i \in U_\nu$, where the $\epsilon_{i\nu}$ are random variables such that

$$\max_{i\in U_\nu} \mathrm{E}\left[\epsilon_{i\nu}^2 \,|\, \mathbf{x}_\nu\right] = O\left(\frac{1}{n_\nu^2 h_\nu^2}\right).$$

Proof of Lemma 5:

(i) This is a consequence of Lemma 2, because when $\nu \to \infty$, $\tilde{\phi}_i^{-1}$ converges with $P_{\boldsymbol{X}}$–probability one to $\phi(x_i)^{-1} \in [1, \lambda_6^{-1}]$, by (B3), for all $i \in U_\nu$.

(ii) This result follows from Lemma 3(iv) and Lemma 4(i) because the designated derivatives of $\widehat{\phi}_i^{-1}$ are continuous functions of the $\tilde{\phi}_i$ and of the $N_\nu^{-1}T_{i,pq}$ and $N_\nu^{-1}t_{i,pq}$.

(iii) Let $g_1(a) = a^{-4}$ and note that for all a in the support set of $\widehat{\phi}_i$, $\delta_2/(N_\nu h_\nu) \le a \le 1$ for all $i \in U_\nu$, as a result of the adjustment in (9). Hence, with $P_{\boldsymbol{X}}$–probability one, $\widehat{\phi}_i^{-4}$ is a continuous function of $\widehat{\phi}_i$ that is bounded by $N_\nu^4 h_\nu^4/\delta_2^4 = O\left(n_\nu^4 h_\nu^4\right)$, by (A2), for all $i \in U_\nu$. As $\widehat{\phi}_i$ is a function of the sample weighted means $N_\nu^{-1}\widehat{T}_{si,pq}$ and $N_\nu^{-1}\widehat{t}_{si,p}$, $p, q = 1, \cdots, k+1$, then applying Lemma 4(ii) it follows that $\{\widehat{\phi}_i^{-4} \;:\; i \in U_\nu \text{ is fixed}\}_{\nu\ge 1}$ satisfies the conditions of Theorem 5.4.4 (with $r = 8$, $\eta = 1$ and $p = 4$) of Fuller (1996, pp. 247) uniformly in $i \in U_\nu$, with $P_{\boldsymbol{X}}$–probability one. Therefore, the result follows.

(iv) The reciprocal of the local polynomial regression estimator (9) can be written as

$$\widehat{\phi}_i^{-1} = \left\{\max\left[\boldsymbol{e}_1'\left(\widehat{\mathbf{T}}_{si} + \mathrm{diag}\left\{\delta_1/N_\nu\right\}\right)^{-1}\widehat{\mathbf{t}}_{si},\ \delta_2/(N_\nu h_\nu)\right]\right\}^{-1} \equiv g_2\left(\widehat{\mathbf{a}}_i\right).$$

where $\widehat{\mathbf{a}}_i = \left((\mathrm{vec}\,\widehat{\mathbf{T}}_{si})', \widehat{\mathbf{t}}_{si}', \delta_1, \delta_2\right)'$. Following closely Särndal et al. (1992, pp. 205-206), a Taylor expansion of g_2 around $\mathbf{a}_i = ((\mathrm{vec}\,\mathbf{T}_i)', \mathbf{t}_i', 0, 0)'$ gives

$$\begin{aligned}\widehat{\phi}_i^{-1} \;=\; \tilde{\phi}_i^{-1} \;&+\; \sum_{p=1}^{k+1}\sum_{q=1}^{k+1} \frac{\partial\widehat{\phi}_i^{-1}}{\partial\widehat{T}_{si,pq}}\bigg|_{\mathbf{a}_i=\mathbf{a}_{i0}}\left(\widehat{T}_{si,pq} - T_{i,pq}\right) \;+\; \sum_{p=1}^{k+1}\frac{\partial\widehat{\phi}_i^{-1}}{\partial\widehat{t}_{si,p}}\bigg|_{\mathbf{a}_i=\mathbf{a}_{i0}}\left(\widehat{t}_{si,p} - t_{i,p}\right) \;+ \\ &\frac{\partial\widehat{\phi}_i^{-1}}{\partial\delta_1}\bigg|_{\mathbf{a}_i=\mathbf{a}_{i0}}\left(\frac{\delta_1}{N_\nu} - 0\right) \;+\; \frac{\partial\widehat{\phi}_i^{-1}}{\partial\delta_2}\bigg|_{\mathbf{a}_i=\mathbf{a}_{i0}}\left(\frac{\delta_2}{N_\nu h_\nu} - 0\right) \;+\; \epsilon_{i\nu},\end{aligned} \tag{23}$$

where $\epsilon_{i\nu}$ is the remainder term. The derivatives in (23) are obtained as follows: applying Theorem 10.8.10 of Graybill (1983, pp. 357–358),

$$\frac{\partial\widehat{\phi}_i^{-1}}{\partial\widehat{T}_{si,pq}} = \widehat{\phi}_i^{-2}\boldsymbol{e}_1'\left(\widehat{\mathbf{T}}_{si} + \mathrm{diag}\left\{\delta_1/N_\nu\right\}\right)^{-1}\Lambda_{pq}\left(\widehat{\mathbf{T}}_{si} + \mathrm{diag}\left\{\delta_1/N_\nu\right\}\right)^{-1}\widehat{\mathbf{t}}_{si},$$

where the $(k+1)\times(k+1)$ Λ_{pq} matrix has a value one in positions (p,q) and (q,p) and a value zero otherwise. By direct differentiation,

$$\frac{\partial\widehat{\phi}_i^{-1}}{\partial\widehat{t}_{si,p}} = -\widehat{\phi}_i^{-2}\boldsymbol{e}_1'\left(\widehat{\mathbf{T}}_{si} + \mathrm{diag}\left\{\delta_1/N_\nu\right\}\right)^{-1}\boldsymbol{e}_p$$

and, by similar steps to those in the proof of Theorem 10.8.10 of Graybill (1983),

$$\frac{\partial\widehat{\phi}_i^{-1}}{\partial\delta_1} = \frac{1}{N_\nu}\widehat{\phi}_i^{-2}\boldsymbol{e}_1'\left[\left(\widehat{\mathbf{T}}_{si} + \mathrm{diag}\left\{\delta_1/N_\nu\right\}\right)^{-1}\right]^2\widehat{\mathbf{t}}_{si}.$$

Finally,

$$\frac{\partial\widehat{\phi}_i^{-1}}{\partial\delta_2} = -\widehat{\phi}_i^{-2}\frac{1}{N_\nu h_\nu}.$$

Evaluating those derivatives at $\mathbf{a}_i$, we obtain that

$$\begin{aligned}
\widehat{\phi}_i^{-1} &= \tilde{\phi}_i^{-1} + \tilde{\phi}_i^{-2}\sum_{p=1}^{k+1}\sum_{q=1}^{k+1}\boldsymbol{e}_1'\mathbf{T}_i^{-1}\Lambda_{pq}\mathbf{B}_i\left(\widehat{T}_{si,pq} - T_{i,pq}\right) - \\
&\quad \tilde{\phi}_i^{-2}\sum_{p=1}^{k+1}\boldsymbol{e}_1'\mathbf{T}_i^{-1}\boldsymbol{e}_p\left(\widehat{t}_{si,p} - t_{i,p}\right) + o\left(\frac{1}{N_\nu^2}\right) + O\left(\frac{1}{N_\nu^2 h_\nu^2}\right) + \epsilon_{i\nu}, \\
&= \tilde{\phi}_i^{-1} + \tilde{\phi}_i^{-2}\boldsymbol{e}_1'\mathbf{T}_i^{-1}\left(\widehat{\mathbf{T}}_{si} - \mathbf{T}_i\right)\mathbf{B}_i - \tilde{\phi}_i^{-2}\boldsymbol{e}_1'\mathbf{T}_i^{-1}\left(\widehat{\mathbf{t}}_{si} - \mathbf{t}_i\right) + \\
&\quad o\left(\frac{1}{N_\nu^2}\right) + O\left(\frac{1}{N_\nu^2 h_\nu^2}\right) + \epsilon_{i\nu} \\
&= \tilde{\phi}_i^{-1} - \tilde{\phi}_i^{-2}\boldsymbol{e}_1'\mathbf{T}_i^{-1}\left(\widehat{\mathbf{t}}_{si} - \widehat{\mathbf{T}}_{si}\mathbf{B}_i\right) + O\left(\frac{1}{N_\nu^2 h_\nu^2}\right) + \epsilon_{i\nu}.
\end{aligned}$$

We next demonstrate that the second moment of the $\epsilon_{i\nu}$ is of order $O(n_\nu^{-2}h_\nu^{-2})$ uniformly in $i \in U_\nu$, with $P_{\boldsymbol{X}}$–probability one. This property will be established by showing the sequence $\{\epsilon_{i\nu}^2 \equiv f_\nu(\widehat{\mathbf{a}}_i)\}_{\nu\geq 1}$, with $i \in U_\nu$ fixed, satisfies the five conditions of Theorem 5.4.3 of Fuller (1996, pp. 244-245), for almost all realizations of $\{\boldsymbol{x}_\nu\}$. The first condition follows (with $\alpha = 2$, $s = 4$ and $a_\nu = O((n_\nu h_\nu)^{-1/2})$) from Lemma 4(ii). Now, since by the C_r inequality (Sen and Singer 1993, pp. 21)

$$\mathrm{E}\big[f_\nu(\widehat{\mathbf{a}}_i)^2\,|\,\mathbf{x}_\nu\big] \leq 4^3\left[\mathrm{E}\big[\widehat{\phi}_i^{-4}\,|\,\mathbf{x}_\nu\big] + \tilde{\phi}_i^{-4} + \tilde{\phi}_i^{-8}\mathrm{E}\big[D_i^4\,|\,\mathbf{x}_\nu\big] + O\left(\frac{1}{N_\nu^8 h_\nu^8}\right)\right] \tag{24}$$

for all $i \in U_\nu$, where $D_i = \boldsymbol{e}_1'\mathbf{T}_i^{-1}\big(\widehat{\mathbf{t}}_{si} - \widehat{\mathbf{T}}_{si}\mathbf{B}_i\big)$, applying (iii), Lemma 4(iii) and (C4) it follows then that (24) is uniformly bounded in $i \in U_\nu$ a. s. $P_{\boldsymbol{X}}$ and the second condition is verified (with $\delta = 1$). The remaining conditions are also satisfied by (i) and (ii). Theferore, as $f_\nu(\widehat{\mathbf{a}}_i)$ and its first three derivatives with respect to the components of $\widehat{\mathbf{a}}_i$ are equal to zero when they are evaluated at $\widehat{\mathbf{a}}_i = \mathbf{a}_i$, the result follows.

□

Lemma 6. *Suppose the assumptions of Theorem 1 hold. Consider the population fit $\tilde{\phi}_i$, given in (10), and the random variables $\bar{y}_{\pi\tilde{\phi}\nu}$, $\bar{d}_{\pi\tilde{\phi}\nu}$ and $\bar{\epsilon}_{\pi\tilde{\phi}\nu}$, given in (11). Then, with $P_{\mathbf{X}}$–probability one,*

$$\mathrm{E}\left[\bar{y}_{\pi\tilde{\phi}\nu} - \bar{y}_{N_\nu}\,\middle|\,\mathbf{x}_\nu\right] = \begin{cases} O\left(h_\nu^{k+(3/2)}\right), & k \text{ even}, \\ O\left(h_\nu^{k+1}\right), & k \text{ odd}, \end{cases} \tag{25}$$

$$\mathrm{Var}\left[\bar{y}_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu\right] \;=\; O\left(\frac{1}{n_\nu}\right) \tag{26}$$

$$\left(\mathrm{E}\big[\bar{d}_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu\big],\,\mathrm{E}\big[\bar{d}^2_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu\big]\right)' \;=\; O\left(\frac{1}{n_\nu h_\nu}\right) \tag{27}$$

and

$$\mathrm{E}\big[\bar{\epsilon}^2_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu\big] \;=\; O\left(\frac{1}{n_\nu^2 h_\nu^2}\right). \tag{28}$$

Proof of Lemma 6: The bias of $\bar{y}_{\pi\tilde{\phi}\nu}$ satisfies

$$\begin{aligned}\left|\mathrm{E}\left[\,\bar{y}_{\pi\tilde{\phi}\nu}-\bar{y}_{N_\nu}\,|\,\mathbf{x}_\nu\right]\right| \;&=\; \left|\frac{1}{N_\nu}\sum_{i\in U_\nu}\left(\frac{\phi_i}{\tilde{\phi}_i}-1\right)y_i\right| \\ &\le\; \left[\min_{i\in U_\nu}|\tilde{\phi}_i|^{-1}\right]\left[\frac{1}{N_\nu}\sum_{i\in U_\nu}y_i^2\right]^{1/2}\left[\frac{1}{N_\nu}\sum_{i\in U_\nu}\left(\tilde{\phi}_i-\phi_i\right)^2\right]^{1/2}.\end{aligned}$$

Writing

$$\begin{aligned}\frac{1}{N_\nu}\sum_{i\in U_\nu}\left(\tilde{\phi}_i-\phi_i\right)^2 \;&=\; \frac{2h_\nu}{2N_\nu h_\nu}\sum_{i\in U_\nu}\left(\tilde{\phi}_i-\phi_i\right)^2 I_{[0,h_\nu]\cup(1-h_\nu,1]}(x_i)\;+ \\ &\quad\;\frac{1}{N_\nu}\sum_{i\in U_\nu}\left(\tilde{\phi}_i-\phi_i\right)^2 I_{(h_\nu,1-h_\nu]}(x_i),\end{aligned}$$

Lemma 2 and Lemma 3(i–iii) gives that

$$\frac{1}{N_\nu}\sum_{i\in U_\nu}\left(\tilde{\phi}_i-\phi_i\right)^2 \;=\; \begin{cases} O\left(h_\nu^{2k+3}\right), & k \text{ even}, \\ O\left(h_\nu^{2(k+1)}\right), & k \text{ odd}.\end{cases}$$

Hence, (25) follows from Lemma 5(i) and (A5).

To prove (26), note first that

$$\begin{aligned}\mathrm{Var}\big[\,\bar{y}_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu\big] \;\le\;& \left[N_\nu\min_{i\in U_\nu}\pi_i\right]^{-1}\left[\max_{i\in U_\nu}\left(\frac{\phi_i}{\tilde{\phi}_i}\right)^2\right]\left[\frac{1}{N_\nu}\sum_{j\in U_\nu}y_i^2\right] \;+ \\ & \left[\max_{i'\neq i\in U_\nu}\frac{|\Delta_{ii'}|}{\pi_i\pi_{i'}}\right]\left[\max_{i'\neq i\in U_\nu}\frac{\phi_i\phi_{i'}}{\tilde{\phi}_i\tilde{\phi}_{i'}}\right]\left[\frac{1}{N_\nu}\sum_{j\in U_\nu}|y_i|\right]^2 \;+ \\ & \left[N_\nu\min_{i\in U_\nu}\pi_i\right]^{-1}\left[\max_{i\in U_\nu}\left(\frac{\phi_i(1-\phi_i)}{\tilde{\phi}_i^2}\right)\right]\left[\frac{1}{N_\nu}\sum_{j\in U_\nu}y_i^2\right]\end{aligned}$$

and, by (A1)–(A3), $[N_\nu\min_{i\in U_\nu}\pi_i]^{-1}$ and $\max_{i'\neq i\in U_\nu}\pi_i^{-1}\pi_{i'}^{-1}|\Delta_{ii'}|$ are $O(n_\nu^{-1})$. Since all the other terms are $O(1)$ a.s. $P_{\mathbf{X}}$, by Lemma 2, (A5) and (B3), then the variance of $\bar{y}_{\pi\tilde{\phi}\nu}$ has the prescribed order.

Now, taking $D_i = \boldsymbol{e}_1{}'\mathbf{T}_i^{-1}(\widehat{\mathbf{t}}_{si} - \widehat{\mathbf{T}}_{si}\mathbf{B}_i)$, we have that

$$\begin{aligned}
\left|\mathrm{E}\big(\bar{d}_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu\big)\right| &= \left|\frac{1}{N_\nu}\sum_{i\in U_\nu} y_i \frac{\mathrm{E}\big[D_i I_i R_i\big]}{\pi_i \tilde{\phi}_i^2}\right| \\
&\leq \left[\min_{i\in U_\nu}\pi_i\right]^{-1}\left[\max_{i\in U_\nu}\frac{\left|\mathrm{E}\big[D_i I_i R_i\,|\,\mathbf{x}_\nu\big]\right|}{\tilde{\phi}_i^2}\right]\left[\frac{1}{N_\nu}\sum_{j\in U_\nu}|y_i|\right] \\
&\leq O\left(1\right)O\left(\frac{1}{n_\nu h_\nu}\right)O\left(1\right) \;=\; O\left(\frac{1}{n_\nu h_\nu}\right),
\end{aligned}$$

a.s. $P_{\mathbf{X}}$, by (A1), (A2), (A5) and Lemmas 2 and 4(iii). By similar arguments,

$$\begin{aligned}
\mathrm{E}\big(\bar{d}^2_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu\big) \;\leq\; & \left[N_\nu\min_{i\in U_\nu}\pi_i^2\right]^{-1}\left[\max_{i\in U_\nu}\frac{\mathrm{E}\big[D_i^2\,|\,\mathbf{x}_\nu\big]}{\tilde{\phi}_i^4}\right]\left[\frac{1}{N_\nu}\sum_{j\in U_\nu}y_i^2\right] \;+ \\
& \left[\min_{i'\neq i\in U_\nu}\pi_i\pi_{i'}\right]^{-1}\left[\max_{i'\neq i\in U_\nu}\frac{\big\{\mathrm{E}\big[D_i^2\,|\,\mathbf{x}_\nu\big]\mathrm{E}\big[D_{i'}^2\,|\,\mathbf{x}_\nu\big]\big\}^{1/2}}{\tilde{\phi}_i^2\tilde{\phi}_{i'}^2}\right]\left[\frac{1}{N_\nu}\sum_{j\in U_\nu}|y_i|\right]^2 \\
\leq\; & O\left(\frac{1}{n_\nu}\right)O\left(\frac{1}{n_\nu h_\nu}\right)O\left(1\right) \;+\; O\left(1\right)O\left(\frac{1}{n_\nu h_\nu}\right)O\left(1\right) \;=\; O\left(\frac{1}{n_\nu h_\nu}\right),
\end{aligned}$$

a.s. $P_{\mathbf{X}}$ and (iii) is established.

Finally, note that

$$\begin{aligned}
\mathrm{E}(\bar{\epsilon}^2_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu) \;=\; & \mathrm{E}\left(\frac{1}{N_\nu^2}\sum_{i\in U_\nu}\pi_i^{-2}\epsilon_i^2 y_i^2 I_i R_i \,\middle|\, \mathbf{x}_\nu\right) \;+ \\
& \mathrm{E}\left(\frac{1}{N_\nu^2}\sum_{i\in U_\nu}\sum_{i'\neq i\in U_\nu}\pi_i^{-1}\pi_{i'}^{-1}\epsilon_i\epsilon_{i'}y_i y_{i'} I_i I_{i'} R_i R_{i'} \,\middle|\, \mathbf{x}_\nu\right).
\end{aligned} \tag{29}$$

The first term on the right–hand side of (29) is bounded by

$$\left(N_\nu\min_{i\in U_\nu}\pi_i^2\right)^{-1}\max_{i\in U_\nu}\mathrm{E}\left(\epsilon_i^2\,|\,\mathbf{x}_\nu\right)\frac{1}{N_\nu}\sum_{i\in U_\nu}y_i^2 = O\left(\frac{1}{n_\nu}\right)O\left(\frac{1}{n_\nu^2 h_\nu^2}\right)O\left(1\right)$$

a.s. $P_{\mathbf{X}}$, by assumptions (A1), (A2), (A5) and Lemma 5. Similarly, the following term is dominated by

$$\begin{aligned}
& \left(\min_{i\in U_\nu}\pi_i\min_{i'\neq i\in U_\nu}\pi_{i'}\right)^{-1}\max_{i'\neq i\in U_\nu}\left(\mathrm{E}\left(\epsilon_i^2\,|\,\mathbf{x}_\nu\right)\mathrm{E}\left(\epsilon_{i'}^2\,|\,\mathbf{x}_\nu\right)\right)^{1/2}\frac{1}{N_\nu}\sum_{i\in U_\nu}y_i^2 \\
&= O\left(1\right)O\left(\frac{1}{n_\nu^2 h_\nu^2}\right)O\left(1\right)
\end{aligned}$$

a.s. $P_{\mathbf{X}}$. Thus,

$$\mathrm{E}(\bar{\epsilon}^2_{\pi\tilde{\phi}\nu}\,|\,\mathbf{x}_\nu) \;=\; O\left(\frac{1}{n_\nu h_\nu}\right) \quad \text{a.s. } P_{\mathbf{X}}$$

and (28) follows. That concludes the proof of the Lemma. □

Acknowledgments

The first author was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil, under the grant Projeto Universal 480518/2004–1.

References

Alho, J. M. (1990). Adjusting for nonresponse bias using logistic regression. *Biometrika 77*, 617–624.

Breidt, F. J. and J. D. Opsomer (2000). Local polynomial regression estimators in survey sampling. *Annals of Statistics 28*, 1026–1053.

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

Da Silva, D. N. (2003). *Adjustments for survey unit nonresponse under nonparametric response mechanisms*. Ph. D. thesis, Iowa State University.

Da Silva, D. N. and J. D. Opsomer (2006). A kernel smoothing method of adjusting for unit non-response in sample surveys. *The Canadian Journal of Statistics/La revue canadienne de statistique 34*(4), 563–579.

Da Silva, D. N. and J. D. Opsomer (2008). Nonparametric propensity weighting for survey nonresponse through local polynomial regression. *Working paper*.

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

Ekholm, A. and S. Laaksonen (1991). Weighting via response modeling in the Finnish Household Budget Survey. *Journal of Official Statistics 7*, 325–337.

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

Fuller, W. A. (1996). *Introduction to Statistical Time Series (Second Edition)*. Wiley.

Giommi, A. (1984). A simple method for estimating individual response probabilities in sampling from finite populations. *Metron 42*(4), 185–200.

Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Duxbury Press.

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

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

Isaki, C. T. and W. A. Fuller (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association 77*, 89–96.

Kim, J. K. and J. J. . Kim (2007). Nonresponse weighting adjustment using estimated response probability. *The Canadian Journal of Statistics 35*(4), 501–514.

Laaksonen, S. (2005–2006). Does the choice of link function matter in response propensity modelling? *Model Assisted and Applications 1*(2), 95–100.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis With Missing Data*. Wiley.

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

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

Opsomer, J.-D. (1995). *Optimal Bandwidth Selection for Fitting an Additive Model by Local Polynomial Regression*. Ph. D. thesis, Cornell University.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York, NY: Springer-Verlag.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

Ruppert, D. and M. P. Wand (1994). Multivariate locally weighted least squares regression. *Annals of Statistics 22*, 1346–1370.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Sen, P. K. and J. d. M. Singer (1993). *Large sample methods in statistics: an introduction with applications*. Chapman and Hall Ltd.

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