The Annals of Statistics
2000, Vol. 28, No. 4, 1026–1053

LOCAL POLYNOMIAL REGRESSION ESTIMATORS IN SURVEY SAMPLING[1]

BY F. JAY BREIDT AND JEAN D. OPSOMER

Colorado State University and Iowa State University

Estimation of finite population totals in the presence of auxiliary information is considered. A class of estimators based on local polynomial regression is proposed. Like generalized regression estimators, these estimators are weighted linear combinations of study variables, in which the weights are calibrated to known control totals, but the assumptions on the superpopulation model are considerably weaker. The estimators are shown to be asymptotically design-unbiased and consistent under mild assumptions. A variance approximation based on Taylor linearization is suggested and shown to be consistent for the design mean squared error of the estimators. The estimators are robust in the sense of asymptotically attaining the Godambe–Joshi lower bound to the anticipated variance. Simulation experiments indicate that the estimators are more efficient than regression estimators when the model regression function is incorrectly specified, while being approximately as efficient when the parametric specification is correct.

1. Introduction.

1.1. *Background.*

In many survey problems, auxiliary information is available for all elements of the population of interest. Population registers in some countries contain age and taxable income for all residents. Studies of labor force characteristics or household expenditure patterns might benefit from these auxiliary data. Geographic information systems may contain measurements derived from satellite imagery for all locations. These spatially explicit data can be useful in augmenting measurements obtained in agricultural surveys or natural resource inventories. Indeed, use of auxiliary information in estimating parameters of a finite population of study variables is a central problem in surveys.

One approach to this problem is the superpopulation approach, in which a working model ξ describing the relationship between the auxiliary variable x and the study variable y is assumed. Estimators are sought which have good efficiency if the model is true, but maintain desirable properties like asymptotic design unbiasedness (unbiasedness over repeated sampling from

---

Received June 1999; revised March 2000.

[1] Supported in part by cooperative agreement 68-3A75-43 between USDA Natural Resources Conservation Service and Iowa State University. Funds for computing equipment provided by NSF SCREMS Grant DMS-97-07740.

AMS 1991 *subject classifications*. Primary 62D05; secondary 62G08.

Key words and phrases. Calibration, generalized regression estimation, Godambe-Joshi lower bound, model-assisted estimation, nonparametric regression.

the finite population) and design consistency if the model is false. Typically, the assumed models are linear models, leading to the familiar ratio and regression estimators [e.g., Cochran (1977)], the best linear unbiased estimators [Brewer (1963), Royall (1970)], the generalized regression estimators [Cassel, Särndal, and Wretman (1977), Särndal (1980), Robinson and Särndal (1983)], and related estimators [Wright (1983), Isaki and Fuller (1982)].

The papers cited vary in their emphasis on design and model, but it is fair to say that all are concerned to some extent with behavior of the estimators under model misspecification. Given this concern with robustness, it is natural to consider a nonparametric class of models for ξ, because they allow the models to be correctly specified for much larger classes of functions. Kuo (1988), Dorfman (1992), Dorfman and Hall (1993), Chambers (1996) and Chambers, Dorfman and Wehrly (1993) have adopted this approach in constructing model-based estimators.

This paper describes theoretical properties of a new type of model-assisted nonparametric regression estimator for the finite population total, based on local polynomial smoothing. Local polynomial regression is a generalization of kernel regression. Cleveland (1979) and Cleveland and Devlin (1988) showed that these techniques are applicable to a wide range of problems. Theoretical work by Fan (1992, 1993) and Ruppert and Wand (1994) showed that they have many desirable theoretical properties, including adaptation to the design of the covariate(s), consistency and asymptotic unbiasedness. Wand and Jones (1995) provide a clear explanation of the asymptotic theory for kernel regression and local polynomial regression. The monograph by Fan and Gijbels (1996) explores a wide range of application areas of local polynomial regression techniques. However, the application of these techniques to model-assisted survey sampling is new.

In Section 1.2 we introduce the local polynomial regression estimator and in Section 1.3 we state assumptions used in the theoretical derivations of Section 2, in which our main results are described. Section 2.1 shows that the estimator is a weighted linear combination of study variables in which the weights are calibrated to known control totals. Section 2.2 contains a proof that the estimator is asymptotically design unbiased and design consistent, and Section 2.3 provides an approximation to its mean squared error and a consistent estimator of the mean squared error. Section 2.4 provides sufficient conditions for asymptotic normality of the local polynomial regression estimator and establishes a central limit theorem in the case of simple random sampling. We show that the estimator is robust in the sense of asymptotically attaining the Godambe–Joshi lower bound to the anticipated variance in Section 2.5, a result previously known only for the parametric case. Section 3 reports on a simulation study of the design properties of the estimator, which is competitive with the classical survey regression estimator when the population regression function is linear, but dominates the regression estimator when the regression function is not linear. Our estimator also performs well relative to other parametric and nonparametric estimators, both model-assisted and model-based.

1.2. *Proposed estimator.* Consider a finite population $U_N = \{1, \dots, i, \dots, N\}$. For each $i \in U_N$, an auxiliary variable x_i is observed. In this article, we explore the case in which the x_i are scalars. But it will be clear from the definition of the estimator later in this section that there is no inherent reason to do so, except to make the theory more tractable. Most of the properties of the estimator we will discuss are also expected to hold for multidimensional x_i, though the "curse of dimensionality" implies that practical applications would be complicated by sparseness of the regressors in the design space. Sparseness adjustments [e.g., Hall and Turlach (1997)] or dimension-reducing alternatives such as additive modeling [Hastie and Tibshirani (1990)] would then need to be employed. These topics will be explored elsewhere.

Let $t_x = \sum_{i \in U_N} x_i$. A probability sample s is drawn from U_N according to a fixed-size sampling design $p_N(\cdot)$, where $p_N(s)$ is the probability of drawing the sample s. Let n_N be the size of s. Assume $\pi_{iN} = \Pr\{i \in s\} = \sum_{s:\, i \in s} p_N(s) > 0$ and $\pi_{ijN} = \Pr\{i, j \in s\} = \sum_{s:\, i, j \in s} p_N(s) > 0$ for all $i, j \in U_N$. For compactness of notation we will suppress the subscript N and write π_i, π_{ij} in what follows. The study variable y_i is observed for each $i \in s$. The goal is to estimate $t_y = \sum_{i \in U_N} y_i$.

Let $I_i = 1$ if $i \in s$ and $I_i = 0$ otherwise. Note that $\mathrm{E}_p[I_i] = \pi_i$, where $\mathrm{E}_p[\cdot]$ denotes expectation with respect to the sampling design (i.e., averaging over all possible samples from the finite population). Using this notation, an estimator t_y^* of t_y is said to be design-unbiased if $\mathrm{E}_p\left[t_y^*\right] = t_y$. A well-known design-unbiased estimator of t_y is the Horvitz–Thompson estimator,

$$\hat{t}_y = \sum_{i \in s} \frac{y_i}{\pi_i} = \sum_{i \in U_N} \frac{y_i I_i}{\pi_i} \tag{1}$$

[Horvitz and Thompson (1952)]. The variance of the Horvitz–Thompson estimator under the sampling design is

$$\mathrm{Var}_p\left(\hat{t}_y\right) = \sum_{i, j \in U_N} \left(\pi_{ij} - \pi_i \pi_j\right) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j}. \tag{2}$$

Note that $\hat{t}_y$ does not depend on the $\{x_i\}$. It is of interest to improve upon the efficiency of the Horvitz–Thompson estimator by using the auxiliary information.

The estimator we propose is motivated by modeling the finite population of y_i's, conditioned on the auxiliary variable x_i, as a realization from an infinite superpopulation, ξ, in which

$$y_i = m(x_i) + \varepsilon_i,$$

where ε_i are independent random variables with mean zero and variance $v(x_i)$, $m(x)$ is a smooth function of x, and $v(x)$ is smooth and strictly positive. Given x_i, $m(x_i) = \mathrm{E}_\xi[y_i]$ and so is called the regression function, while $v(x_i) = \mathrm{Var}_\xi(y_i)$ and so is called the variance function.

Let K denote a continuous kernel function and let h_N denote the bandwidth. We begin by defining the local polynomial kernel estimator of degree q

based on the entire finite population. Let $\mathbf{y}_U = [y_i]_{i\in U_N}$ be the N-vector of y_i's in the finite population. Define the $N \times (q+1)$ matrix

$$\mathbf{X}_{Ui} = \begin{bmatrix} 1 & x_1 - x_i & \cdots & (x_1 - x_i)^q \\ \vdots & \vdots & & \vdots \\ 1 & x_N - x_i & \cdots & (x_N - x_i)^q \end{bmatrix} = \begin{bmatrix} 1 & x_j - x_i & \cdots & (x_j - x_i)^q \end{bmatrix}_{j\in U_N},$$

and define the $N \times N$ matrix,

$$\mathbf{W}_{Ui} = \operatorname{diag}\left\{ \frac{1}{h_N} K\left(\frac{x_j - x_i}{h_N} \right) \right\}_{j\in U_N}.$$

Let $\mathbf{e}_r$ represent a vector with a 1 in the rth position and 0 elsewhere. Its dimension will be clear from the context. The local polynomial kernel estimator of the regression function at x_i, based on the entire finite population, is then given by

$$m_i = \mathbf{e}_1'(\mathbf{X}_{Ui}'\mathbf{W}_{Ui}\mathbf{X}_{Ui})^{-1}\mathbf{X}_{Ui}'\mathbf{W}_{Ui}\mathbf{y}_U = \mathbf{w}_{Ui}'\mathbf{y}_U, \tag{3}$$

which is well defined as long as $\mathbf{X}_{Ui}'\mathbf{W}_{Ui}\mathbf{X}_{Ui}$ is invertible.

If these m_i's were known, then a design-unbiased estimator of t_y would be the generalized difference estimator

$$t_y^* = \sum_{i\in s} \frac{y_i - m_i}{\pi_i} + \sum_{i\in U_N} m_i \tag{4}$$

[Särndal, Swensson and Wretman (1992), page 221]. The design variance of the estimator (4) is

$$\operatorname{Var}_p(t_y^*) = \sum_{i,\, j\in U_N} (\pi_{ij} - \pi_i\pi_j) \frac{y_i - m_i}{\pi_i} \frac{y_j - m_j}{\pi_j}, \tag{5}$$

which we would expect to be smaller than (2); the deviations $\{y_i - m_i\} = \{(m(x_i) - m_i) + \varepsilon_i\}$ will typically have smaller variation than the $\{y_i\}$ for any reasonable smoothing procedure under the model ξ.

The population estimator m_i is the traditional local polynomial regression estimator for the unknown function $m(\cdot)$, widely discussed in the nonparametric regression literature. In the present context, it cannot be calculated, because only the y_i in $s \subset U_N$ are known. Therefore, we will replace each m_i by a sample-based consistent estimator. Let $\mathbf{y}_s = [y_i]_{i\in s}$ be the n_N-vector of y_i's obtained in the sample. Define the $n_N \times (q+1)$ matrix

$$\mathbf{X}_{si} = \begin{bmatrix} 1 & x_j - x_i & \cdots & (x_j - x_i)^q \end{bmatrix}_{j\in s},$$

and define the $n_N \times n_N$ matrix

$$\mathbf{W}_{si} = \operatorname{diag}\left\{ \frac{1}{\pi_j h_N} K\left(\frac{x_j - x_i}{h_N} \right) \right\}_{j\in s}.$$

A sample design-based estimator of m_i is then given by

$$\hat{m}_i^o = \mathbf{e}_1'(\mathbf{X}_{si}'\mathbf{W}_{si}\mathbf{X}_{si})^{-1}\mathbf{X}_{si}'\mathbf{W}_{si}\mathbf{y}_s = \mathbf{w}_{si}^{o\prime}\mathbf{y}_s, \tag{6}$$

as long as $\mathbf{X}'_{si}\mathbf{W}_{si}\mathbf{X}_{si}$ is invertible. When we substitute the $\hat{m}_i^o$ into (4), we have the local polynomial regression estimator for the population total

$$\tilde{t}_y^o = \sum_{i \in s} \frac{y_i - \hat{m}_i^o}{\pi_i} + \sum_{i \in U_N} \hat{m}_i^o. \tag{7}$$

The sample estimator in (6) differs in one important way from the traditional local polynomial regression estimator. The presence of the inclusion probabilities in the "smoothing weights" $\mathbf{w}_{si}^o$ makes our sample-based estimator $\hat{m}_i$ a design-consistent estimator of the finite population smooth m_i, which is based on some (not necessarily optimal) bandwidth h_N, considered fixed here for any N. In real survey problems, h_N will rarely be optimal because a single bandwidth would be chosen and used to compute weights to be applied to all study variables. Despite this, the topics of theoretically optimal and practical bandwidth selection are clearly of some interest and will be addressed elsewhere. Regardless of the choice of h_N, m_i is a well-defined parameter of the finite population. Specifically, m_i is a function of finite population totals, each of which can be estimated consistently by their corresponding Horvitz–Thompson estimators. That is, we have included probability weights in the smoothing weights in order to construct asymptotically design-unbiased and design-consistent estimators of the finite population smooths m_i [not $m(x_i)$]. This is consistent with the development of the generalized regression estimator (GREG), which our procedure reverts to as the bandwidth becomes large.

In principle, the estimator (6) can be undefined for certain $i \in U_N$, even if the population estimator in (3) is defined everywhere: if for some sample s, there are less than $q+1$ observations in the support of the kernel at some x_i, then the matrix $\mathbf{X}'_{si}\mathbf{W}_{si}\mathbf{X}_{si}$ will be singular. This is not a problem in practice, because it can be avoided by selecting a bandwidth that is sufficiently large to make $\mathbf{X}'_{si}\mathbf{W}_{si}\mathbf{X}_{si}$ invertible at all locations x_i. However, that situation cannot be excluded theoretically as long as the bandwidth is considered fixed for a given population. Therefore, for the purpose of the theoretical derivations in Section 2, we will consider an adjusted sample estimator that is guaranteed to exist for any sample $s \subset U_N$.

The adjusted sample estimator for m_i is given by

$$\hat{m}_i = \mathbf{e}'_1 \left(\mathbf{X}'_{si}\mathbf{W}_{si}\mathbf{X}_{si} + \operatorname{diag}\left\{ \frac{\delta}{N^2} \right\}_{j=1}^{q+1} \right)^{-1} \mathbf{X}'_{si}\mathbf{W}_{si}\mathbf{y}_s = \mathbf{w}'_{si}\mathbf{y}_s \tag{8}$$

for some small $\delta > 0$. The terms δN^{-2} in the denominator are small order adjustments that ensure the estimator is well defined for all $s \subset U_N$. This adjustment was also used by Fan (1993) for the same reason when the x_i are considered random. Another possible adjustment would consist of replacing the usual choice of a kernel with compact support by one with infinite support such as the Gaussian kernel. In practice, however, such kernels have been found to increase the computational complexity of local polynomial fitting and result in less satisfactory fits compared to those obtained with compactly supported kernels. The adjustment proposed here maintains the sparseness of the smoothing vector $\mathbf{w}_{si}$, and its effect can be made arbitrarily small by choosing δ accordingly. We let

$$\tilde{t}_y = \sum_{i\in s} \frac{y_i - \hat{m}_i}{\pi_i} + \sum_{i\in U_N} \hat{m}_i \tag{9}$$

denote the local polynomial regression estimator that uses the adjusted sample smoother in (8). The remainder of this article will be concerned with studying the properties of $\tilde{t}_y$ (and $\tilde{t}_y^o$ when appropriate).

The development of the model-assisted local polynomial regression estimator above could clearly be followed for other kinds of smoothing procedures. We focus on the local polynomial methodology because it is of considerable practical interest. In the case $q = 0$, the estimator relies on kernel regression, and behaves like a classical poststratification estimator, but mixed over different possible stratum boundaries. As the bandwidth becomes large, the estimator reverts to the Hájek estimator, $N\hat{t}_y/\widehat{N}$, where $\widehat{N} = \sum_s 1/\pi_k$. In the local linear regression ($q = 1$) case, the estimator looks something like a poststratified regression estimator, and the estimator reverts to the classical regression estimator as the bandwidth becomes large.

1.3. *Notation and assumptions.* Our basic approach to studying the design and model properties of the estimators will be to use a Taylor linearization for the sample smoother $\hat{m}_i$. Note first that we can write $m_i = f(N^{-1}\mathbf{t}_i, 0)$ and $\hat{m}_i = f(N^{-1}\hat{\mathbf{t}}_i, \delta)$ for some function f, where the δ comes from the adjustment in (8) and vanishes in the population fit (3),

$$\mathbf{t}_i = \left[t_{ig}\right]_{g=1}^G = \left[\sum_{k\in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right) z_{igk}^{\dagger}\right]_{g=1}^G = \left[\sum_{k\in U_N} z_{igk}^{*}\right]_{g=1}^G$$

and

$$\hat{\mathbf{t}}_i = \left[\hat{t}_{ig}\right]_{g=1}^G = \left[\sum_{k\in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right) z_{igk}^{\dagger} \frac{I_k}{\pi_k}\right]_{g=1}^G = \left[\sum_{k\in U_N} z_{igk}^{*} \frac{I_k}{\pi_k}\right]_{g=1}^G$$

for suitable $z_{igk}^{\dagger}$. For local polynomial regression of degree q, $G = 3q+2$. If we let $G_1 = 2q+1$, we can write the $z_{igk}^{\dagger}$ as

$$z_{igk}^{\dagger} = \begin{cases} (x_k - x_i)^{g-1}, & g \le G_1, \\ (x_k - x_i)^{g-G_1-1} y_k, & g > G_1. \end{cases}$$

EXAMPLES. The kernel regression ($q = 0$) and local linear regression ($q = 1$) cases are of particular interest.

In the case $q = 0$,

$$t_{i1} = \sum_{k\in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right) \quad \text{and} \quad t_{i2} = \sum_{k\in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right) y_k,$$

so that (3) is the Nadaraya–Watson estimator, based on the entire finite population, of the model regression function: $m_i = t_{i1}^{-1} t_{i2}$. Ignoring the δ-adjustment, the corresponding sample-based estimator is then

$$\hat{m}_i^0 = \hat{t}_{i1}^{-1} \hat{t}_{i2}.$$

In the case of local linear regression ($q = 1$),

$$t_{i1} = \sum_{k \in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right), \qquad t_{i2} = \sum_{k \in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right)(x_k - x_i),$$

$$t_{i3} = \sum_{k \in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right)(x_k - x_i)^2, \qquad t_{i4} = \sum_{k \in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right) y_k$$

and

$$t_{i5} = \sum_{k \in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N}\right)(x_k - x_i) y_k.$$

Then

$$m_i = \frac{t_{i3} t_{i4} - t_{i2} t_{i5}}{t_{i1} t_{i3} - t_{i2}^2},$$

with the corresponding sample-based estimator

$$\hat{m}_i^0 = \frac{\hat{t}_{i3} \hat{t}_{i4} - \hat{t}_{i2} \hat{t}_{i5}}{\hat{t}_{i1} \hat{t}_{i3} - \hat{t}_{i2}^2}.$$

Using a Taylor approximation, define

$$R_{iN} = \hat{m}_i - m_i - \frac{1}{N} \sum_{k \in U_N} z_{ik} \left(\frac{I_k}{\pi_k} - 1\right) - \left.\frac{\partial \hat{m}_i}{\partial \delta}\right|_{\hat{\mathbf{t}}_i = \mathbf{t}_i,\, \delta = 0} \frac{\delta}{N^2}, \tag{10}$$

where

$$z_{ik} = \sum_{g=1}^{G} \left.\frac{\partial \hat{m}_i}{\partial (N^{-1} \hat{t}_{ig})}\right|_{\hat{\mathbf{t}}_i = \mathbf{t}_i,\, \delta = 0} z_{igk}^*.$$

To prove our theoretical results, we make the following assumptions.

(A1) (Distribution of the errors under ξ). The errors ε_i are independent and have mean zero, variance $v(x_i)$ and compact support, uniformly for all N.

(A2) For each N, the x_i are considered fixed with respect to the superpopulation model ξ. The x_i are independent and identically distributed $F(x) = \int_{-\infty}^{x} f(t)\,dt$, where $f(\cdot)$ is a density with compact support $[a_x, b_x]$ and $f(x) > 0$ for all $x \in [a_x, b_x]$.

(A3) (Mean and variance functions m, v on $[a_x, b_x]$). The mean function $m(\cdot)$ is continuous and has $q + 1$ continuous derivatives, and the variance function $v(x)$ is continuous and strictly positive.

(A4) (Kernel K). The kernel $K(\cdot)$ has compact support $[-1,1]$, is symmetric and continuous, and satisfies

$$\int_{-1}^{1} K(u)\,du = 1.$$

(A5) (Sampling rate $n_N N^{-1}$ and bandwidth h_N). As $N \to \infty, n_N N^{-1} \to \pi \in (0,1)$, $h_N \to 0$ and $N h_N^2/(\log\log N) \to \infty$.

(A6) (Inclusion probabilities π_i and π_{ij}). For all $N, \min_{i \in U_N} \pi_i \geq \lambda > 0$, $\min_{i,\, j \in U_N} \pi_{ij} \geq \lambda^* > 0$ and

$$\limsup_{N\to\infty} n_N \max_{i,\, j \in U_N:\ i \neq j} |\pi_{ij} - \pi_i \pi_j| < \infty.$$

(A7) Additional assumptions involving higher-order inclusion probabilities:

$$\lim_{N\to\infty} n_N^2 \max_{(i_1, i_2, i_3, i_4) \in D_{4,N}} \left| \mathrm{E}_p[(I_{i_1} - \pi_{i_1})(I_{i_2} - \pi_{i_2})(I_{i_3} - \pi_{i_3})(I_{i_4} - \pi_{i_4})] \right| < \infty,$$

where $D_{t,N}$ denotes the set of all distinct t-tuples $(i_1, i_2, \dots, i_t)$ from U_N,

$$\lim_{N\to\infty} \max_{(i_1, i_2, i_3, i_4) \in D_{4,N}} \left| \mathrm{E}_p[(I_{i1} I_{i_2} - \pi_{i_1 i_2})(I_{i_3} I_{i_4} - \pi_{i_3 i_4})] \right| = 0$$

and

$$\limsup_{N\to\infty} n_N \max_{(i_1, i_2, i_3) \in D_{3,N}} \left| \mathrm{E}_p\big[(I_{i_1} - \pi_{i_1})^2 (I_{i_2} - \pi_{i_2})(I_{i_3} - \pi_{i_3})\big] \right| < \infty.$$

REMARKS. (i) The $\{x_i\}$ are kept fixed with respect to the model to make the results in later sections look like traditional (nonasymptotic) finite population results. Assumption (A2), however, ensures that the $\{x_i\}$ are a random sample from a continuous distribution. In order to maintain the article's emphasis on the model ξ and sampling design p_N, the conditioning on the x_i's will be suppressed in what follows.

(ii) The assumption of compactly supported errors in A1 is made to simplify bounding arguments used extensively in the proofs. It is possible to obtain the results using finite population moment assumptions of the form

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{i \in U_N} \varepsilon_i^k < \infty \text{ with } \xi\text{-probability } 1,$$

though this significantly complicates the arguments. The assumption of compactly supported errors is also used to establish uniform integrability needed to allow taking expectations through Taylor approximations. Alternatively, it is possible to modify (A3) and (A4) to include additional assumptions about smoothness of the first derivatives of the various functions, which are normally not required for local polynomial regression. Assumption (A5) would then also be modified by a factor of $\log\log N/\log N$ in the bandwidth rate.

These adjustments, together with moment assumptions on the ε_k, would guarantee uniform convergence of the nonparametric regression components of the estimator. Such assumptions were used in Opsomer and Ruppert (1997) for the same purpose in the context of additive model fitting. The assumptions are based on the uniform convergence results of Pollard (1984).

(iii) Assumption (A6) is similar to assumptions used by Robinson and Särndal (1983), who examined the parametric regression case. (A7) extends those assumptions. Assumptions (A6) and (A7) involve first through fourth-order inclusion probabilities of the design. These assumptions hold for simple random sampling without replacement. Let ρ_k denote the kth order inclusion probability of k distinct elements under simple random sampling without replacement. Then (A6) is well known, and it is easy to check that the first expression in (A7) becomes

$$N^2(\rho_4 - 4\rho_1\rho_3 + 6\rho_1^2\rho_2 - 3\rho_1^4) = O(1),$$

the second one becomes

$$\rho_4 - \rho_2^2 = O(N^{-1})$$

and the third expression becomes

$$n_N(\rho_3 - 2\rho_1\rho_2 + \rho_1^3 - 2\rho_1\rho_3 + 5\rho_1^2\rho_2 - 3\rho_1^4) = O(1).$$

(iv) By similar arguments, (A6) and (A7) will hold for stratified simple random sampling with fixed stratum boundaries determined by the x_i's and for related designs. If clustering is a significant feature of the design, then there are at least two possibilities for auxiliary information: availability at the element level and availability at the cluster level. (A6) and (A7) will generally not hold (at the element level) for designs with nontrivial clustering. The first case, however, is rare in practice because a clustered design is unlikely to be used when such detailed element-level frame information is available. In the second case, elements may be fully enumerated within selected clusters or they may be subsampled. If they are fully enumerated, the assumptions above apply directly to the sample of clusters with cluster-level auxiliary information. If they are subsampled, then the framework above would require extensions to describe the subsampling procedure. Though such extensions are beyond the scope of the present investigation, we believe that results similar to those described below would continue to hold under reasonable assumptions.

These remarks indicate that it should be possible to obtain the results we describe under a variety of asymptotic formulations. The specific assumptions we describe are only one possibility, though we believe they are reasonable and they give some insight into when the asymptotic approximations we describe would be expected to work.

2. Main results.

2.1. *Weighting and calibration.* Note from (7) that

$$\begin{aligned}\tilde{t}_y^o &= \sum_{i\in s}\frac{y_i}{\pi_i} + \sum_{j\in U_N}\left(1-\frac{I_j}{\pi_j}\right)\mathbf{w}_{sj}^{o\prime}\mathbf{y}_s \\ &= \sum_{i\in s}\left\{\frac{1}{\pi_i} + \sum_{j\in U_N}\left(1-\frac{I_j}{\pi_j}\right)\mathbf{w}_{sj}^{o\prime}\mathbf{e}_i\right\}y_i \qquad (11)\\ &= \sum_{i\in s} w_{is}y_i.\end{aligned}$$

Thus $\tilde{t}_y^o$ is a linear combination of the sample y_i's, where the weights are the inverse inclusion probabilities, suitably modified to reflect the information in the auxiliary variable x_i. The same reasoning applies directly to $\tilde{t}_y$.

Because the weights are independent of y_i, they can be applied to any study variable of interest. In particular, they can be applied to the auxiliary variables $1, x_i, \dots, x_i^q$. Then it is straightforward to verify that for the local polynomial regression estimator $\tilde{t}_y^o$,

$$\sum_{i\in s}\omega_{is}x_i^l = \sum_{i\in U_N} x_i^l$$

for $l = 0, 1, \dots, q$. That is, the weights are exactly *calibrated* to the $q+1$ known control totals $N, t_x, \dots, t_{x^q}$. Calibration is a highly desirable property for survey weights and in fact motivates the class of estimators considered by Deville and Särndal (1992). Part of the desirability of the calibration property comes from the fact that if $m(x_i)$ is exactly a qth degree polynomial function of x_i, then $\tilde{t}_y^o$ is exactly model-unbiased. In addition, the control totals are often published in official tables or otherwise widely disseminated as benchmark values, so reproducing them from the sample is reassuring to the user.

While the local polynomial regression estimator $\tilde{t}_y$ is no longer exactly calibrated, it remains approximately so, in the sense that its weights reproduce the control totals to terms of $o(\delta N^{-1})$. We omit the proof.

2.2. *Asymptotic design unbiasedness and consistency.* The price for using $\hat{m}_i$'s in place of m_i's in the generalized difference estimator (4) is design bias. The estimator $\tilde{t}_y$ is, however, asymptotically design unbiased and design consistent under mild conditions, as the following theorem demonstrates.

THEOREM 1. *Assume* (A1)–(A7). *Then the local polynomial regression estimator*

$$\tilde{t}_y = \sum_{i\in U_N}\left\{(y_i - \hat{m}_i)\frac{I_i}{\pi_i} + \hat{m}_i\right\}$$

is asymptotically design unbiased (ADU) *in the sense that*

$$\lim_{N\to\infty} \mathrm{E}_p\left[\frac{\tilde{t}_y - t_y}{N}\right] = 0 \quad \textit{with } \xi\textit{-probability } 1,$$

and is design consistent in the sense that

$$\lim_{N\to\infty} \mathrm{E}_p\left[I_{\{|\tilde{t}_y - t_y| > N\eta\}}\right] = 0 \quad \textit{with } \xi\textit{-probability } 1$$

for all $\eta > 0$.

The proof of this and following theorems rely on several technical lemmas which are gathered in the Appendix.

PROOF. By Markov's inequality, it suffices to show that

$$\lim_{N\to\infty} \mathrm{E}_p\left|\frac{\tilde{t}_y - t_y}{N}\right| = 0.$$

Write

$$\frac{\tilde{t}_y - t_y}{N} = \sum_{i\in U_N} \frac{y_i - m_i}{N}\left(\frac{I_i}{\pi_i} - 1\right) + \sum_{i\in U_N} \frac{\hat{m}_i - m_i}{N}\left(1 - \frac{I_i}{\pi_i}\right).$$

Then

$$\begin{aligned} \mathrm{E}_p\left|\frac{\tilde{t}_y - t_y}{N}\right| \le{}& \mathrm{E}_p\left|\sum_{i\in U_N} \frac{y_i - m_i}{N}\left(\frac{I_i}{\pi_i} - 1\right)\right| \\ &+ \left\{\mathrm{E}_p\left[\sum_{i\in U_N} \frac{(\hat{m}_i - m_i)^2}{N}\right] \mathrm{E}_p\left[\sum_{i\in U_N} \frac{(1 - \pi_i^{-1} I_i)}{N}\right]\right\}^{1/2}. \end{aligned} \tag{12}$$

Under (A1)–(A6) and using the fact that

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{i\in U_N} (y_i - m_i)^2 < \infty$$

by Lemma 2(iv) in the Appendix, the first term on the right of (12) converges to zero as $N \to \infty$, following the argument of Theorem 1 in Robinson and Särndal (1983). Under (A6),

$$\mathrm{E}_p\left[\sum_{i\in U_N} \frac{(1 - \pi_i^{-1} I_i)^2}{N}\right] = \sum_{i\in U_N} \frac{\pi_i(1 - \pi_i)}{N\pi_i^2} \le \frac{1}{\lambda}.$$

Combining this with Lemma 4, the second term on the right of (12) converges to zero as $N \to \infty$, and the theorem follows. □

2.3. *Asymptotic mean squared error.* In this section we derive an asymptotic approximation to the mean squared error of the local polynomial regression estimator and propose a consistent variance estimator. We first show that the asymptotic mean squared error of the local polynomial regression estimator is equivalent to the variance of the generalized difference estimator, given in (5).

THEOREM 2. *Assume* (A1)–(A7). *Then*

$$n_N \mathrm{E}_p\left(\frac{\tilde{t}_y - t_y}{N}\right)^2 = \frac{n_N}{N^2} \sum_{i,\, j \in U_N} (y_i - m_i)(y_j - m_j) \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} + o(1). \tag{13}$$

PROOF. Let

$$a_N = n_N^{1/2} \sum_{i \in U_N} \frac{y_i - m_i}{N} \left(\frac{I_i}{\pi_i} - 1\right) \quad \text{and} \quad b_N = n_N^{1/2} \sum_{i \in U_N} \frac{m_i - \hat{m}_i}{N} \left(\frac{I_i}{\pi_i} - 1\right).$$

Then

$$\begin{aligned} \mathrm{E}_p[a_N^2] &= \frac{n_N}{N^2} \sum_{i,\, j \in U_N} (y_i - m_i)(y_j - m_j) \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \\ &\le \left(\frac{1}{\lambda} + \frac{n_N \max_{i,\, j \in U_N:\ i \neq j} |\pi_{ij} - \pi_i \pi_j|}{\lambda^2}\right) \sum_{i \in U_N} \frac{(y_i - m_i)^2}{N}, \end{aligned}$$

so that $\limsup_{N \to \infty} \mathrm{E}_p[a_N^2] < \infty$ by (A6). By Lemma 5, $\mathrm{E}_p[b_N^2] = o(1)$, so that

$$\mathrm{E}_p[a_N b_N] \le \left(\mathrm{E}_p[a_N^2] \mathrm{E}_p[b_N^2]\right)^{1/2} = o(1).$$

Hence,

$$n_N \mathrm{E}_p\left(\frac{\tilde{t}_y - t_y}{N}\right)^2 = \mathrm{E}_p\left[a_N^2\right] + 2\mathrm{E}_p[a_N b_N] + \mathrm{E}_p\left[b_N^2\right] = \mathrm{E}_p\left[a_N^2\right] + o(1),$$

and the result is proved. □

The next result shows that the asymptotic mean squared error in (13) can be estimated consistently under mild assumptions.

THEOREM 3. *Assume* (A1)–(A7). *Then*

$$\lim_{N \to \infty} n_N \mathrm{E}_p \left| \widehat{V}(N^{-1} \tilde{t}_y) - \mathrm{AMSE}(N^{-1} \tilde{t}_y) \right| = 0,$$

where

$$\widehat{V}(N^{-1} \tilde{t}_y) = \frac{1}{N^2} \sum_{i,\, j \in U_N} (y_i - \hat{m}_i)(y_j - \hat{m}_j) \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \frac{I_i I_j}{\pi_{ij}} \tag{14}$$

and

$$\mathrm{AMSE}(N^{-1} \tilde{t}_y) = \frac{1}{N^2} \sum_{i,\, j \in U_N} (y_i - m_i)(y_j - m_j) \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}.$$

Therefore, $\widehat{V}(N^{-1}\tilde{t}_y)$ *is asymptotically design unbiased and design consistent for* $\mathrm{AMSE}(N^{-1}\tilde{t}_y)$.

PROOF. Write

$$A_N = n_N \mathrm{E}_p \left| \frac{1}{N^2} \sum_{i,\, j \in U_N} (y_i - m_i)(y_i - m_j) \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \frac{I_i I_j - \pi_{ij}}{\pi_{ij}} \right|.$$

Now

$$\begin{aligned}
&n_N^2 \mathrm{E}_p \left(\frac{1}{N^2} \sum_{i,\, j \in U_N} (y_i - m_i)(y_j - m_j) \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \frac{I_i I_j - \pi_{ij}}{\pi_{ij}} \right)^2 \\
&\quad = n_N^2 \sum_{i,\, k \in U_N} \frac{1 - \pi_i}{\pi_i} \frac{1 - \pi_k}{\pi_k} \frac{(y_i - m_i)^2 (y_k - m_k)^2}{N^4} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} \\
&\qquad + 2 n_N^2 \sum_{i \in U_N} \sum_{k,\, l \in U_N : k \neq l} \frac{1 - \pi_i}{\pi_i} \frac{\pi_{kl} - \pi_k \pi_l}{\pi_k \pi_l} \frac{(y_i - m_i)^2 (y_k - m_k)(y_l - m_l)}{N^4} \\
&\qquad \times \mathrm{E}_p \left[\frac{I_i - \pi_i}{\pi_i} \frac{I_k I_l - \pi_{kl}}{\pi_{kl}} \right] + n_N^2 \sum_{i,\, j \in U_N,\, i \neq j} \sum_{k,\, l \in U_N : k \neq l} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \frac{\pi_{kl} - \pi_k \pi_l}{\pi_k \pi_l} \\
&\qquad \times \frac{(y_i - m_i)(y_j - m_j)(y_k - m_k)(y_l - m_l)}{N^4} \mathrm{E}_p \left[\frac{I_i I_j - \pi_{ij}}{\pi_{ij}} \frac{I_k I_l - \pi_{kl}}{\pi_{kl}} \right] \\
&\quad = a_{1N} + a_{2N} + a_{3N}.
\end{aligned}$$

But

$$\begin{aligned}
a_{1N} &\leq n_N^2 \sum_{i \in U_N} \frac{(y_i - m_i)^4}{\lambda^3 N^4} + n_N^2 \sum_{i,\, k \in U_N : i \neq k} \frac{(y_i - m_i)^2 (y_k - m_k)^2 |\pi_{ik} - \pi_i \pi_k|}{\lambda^4 N^4} \\
&\leq \left(\frac{1}{N \lambda^3} + \frac{n_N \max_{i,\, k \in U_N : i \neq k} |\pi_{ik} - \pi_i \pi_k|}{N \lambda^4} \right) \sum_{i \in U_N} \frac{(y_i - m_i)^4}{N},
\end{aligned}$$

which goes to zero as $N \to \infty$, and

$$\begin{aligned}
a_{3N} &\leq \frac{(n_N \max_{i,\, k \in U_N : i \neq k} |\pi_{ik} - \pi_i \pi_k|)^2}{\lambda^4 \lambda^{*2}} \sum_{i,\, j \in U_N : i \neq j} \sum_{k,\, l \in U_N : k \neq l} \\
&\qquad \times \frac{|(y_i - m_i)(y_j - m_j)(y_k - m_k)(y_l - m_l)|}{N^4} \left| \mathrm{E}_p \left[\frac{I_i I_j - \pi_{ij}}{\pi_{ij}} \frac{I_k I_l - \pi_{kl}}{\pi_{kl}} \right] \right| \\
&\leq O(N^{-1}) + \frac{(n_N \max_{i,\, k \in U_N : i \neq k} |\pi_{ik} - \pi_i \pi_k|)^2}{\lambda^4 \lambda^{*2}} \\
&\qquad \times \max_{(i,\, j,\, k,\, l) \in D_{4,\, N}} \left| \mathrm{E}_p \left[\frac{I_i I_j - \pi_{ij}}{\pi_{ij}} \frac{I_k I_l - \pi_{kl}}{\pi_{kl}} \right] \right| \sum_{i \in U_N} \frac{(y_i - m_i)^4}{N},
\end{aligned}$$

which converges to zero as $N \to \infty$ by (A7). The Cauchy–Schwarz inequality may then be applied to show that $a_{2N} \to 0$ as $N \to \infty$, and it follows that $A_N \to 0$ as $N \to \infty$.

Next, write

$$\begin{aligned} B_N &= n_N \mathrm{E}_p \left| \frac{1}{N^2} \sum_{i,\, j \in U_N} \{2(y_i - m_i)(m_j - \hat{m}_j) \right. \\ &\qquad \left. + (m_i - \hat{m}_i)(m_j - \hat{m}_j)\} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \frac{I_i I_j}{\pi_{ij}} \right| \\ &\le \left(\frac{2 n_N \max_{i,\, j \in U_N : i \neq j} |\pi_{ij} - \pi_i \pi_j|}{\lambda^2 \lambda^*} + \frac{2 n_N}{\lambda^2 N} \right) \\ &\quad \times \left\{ \frac{\sum_{i \in U_N} (y_i - m_i)^2}{N} \frac{\sum_{i \in U_N} \mathrm{E}_p[(m_i - \hat{m}_i)^2]}{N} \right\}^{1/2} \\ &\quad + \left(\frac{n_N \max_{i,\, j \in U_N : i \neq j} |\pi_{ij} - \pi_i \pi_j|}{\lambda^2 \lambda^*} + \frac{n_N}{\lambda^2 N} \right) \\ &\quad \times \frac{\sum_{i \in U_N} \mathrm{E}_p[(m_i - \hat{m}_i)^2]}{N} \\ &\to 0 \end{aligned}$$

as $N \to \infty$ using (A6) and Lemma 4. The result then follows because

$$n_N \mathrm{E}_p | \widehat{V}(N^{-1} \tilde{t}_y) - \text{ AMSE } (N^{-1} \tilde{t}_y)| \le A_N + B_N. \qquad \square$$

An alternative variance estimator could be constructed by replacing the term $\pi_i^{-1} \pi_j^{-1}$ in (14) with the product of weights $\omega_{is} \omega_{js}$ from (11). This is the analogue of the weighted residual technique [Särndal, Swensson and Wretman (1989)] for estimating the variance of the general regression estimator, which they propose to improve the conditional and small sample properties of the variance estimator.

Simplified versions of (13) and (14) are given in Corollary 1 below for the case of simple random sampling.

2.4. *Asymptotic normality.* The local polynomial regression estimator inherits the limiting distributional properties of the generalized difference estimator, as we now demonstrate.

THEOREM 4. *Assume that* (A1)–(A7) *hold and let* t_y^* *and* $\mathrm{Var}_p(t_y^*)$ *be as defined in* (4) *and* (5), *respectively. Then,*

$$\frac{N^{-1}(t_y^* - t_y)}{\mathrm{Var}_p^{1/2}(N^{-1} t_y^*)} \xrightarrow{\mathscr{L}} N(0, 1)$$

as $N \to \infty$ *implies*

$$\frac{N^{-1}(\tilde{t}_y - t_y)}{\widehat{V}^{1/2}(N^{-1}\tilde{t}_y)} \xrightarrow{\mathscr{L}} N(0,1)$$

as $N \to \infty$, *where* $\widehat{V}(N^{-1}\tilde{t}_y)$ *is given in* (14).

PROOF. From the proof of Theorem 2,

$$N^{-1}(\tilde{t}_y - t_y) = \sum_{i \in U_N} \frac{y_i - m_i}{N}\left(\frac{I_i}{\pi_i} - 1\right) + o_p(n_N^{-1/2}) = N^{-1}(t_y^* - t_y) + o_p(n_N^{-1/2}).$$

Further, $\widehat{V}(N^{-1}\tilde{t}_y)/\mathrm{AMSE}(N^{-1}\tilde{t}_y) \to_p 1$ by Theorem 3, so the result is established. □

Thus, establishing a central limit theorem (CLT) for the local polynomial regression estimator is equivalent to establishing a CLT for the generalized difference estimator, which in turn is essentially the same problem as establishing a CLT for the Horvitz–Thompson estimator. Additional conditions on the design beyond those of Theorem 3 are generally needed; for example, conditions which ensure that the design is well approximated by unequal probability Bernoulli sampling conditioned to the fixed sample size n_N, or by successive sampling with stable draw-to-draw selection probabilities [e.g., Sen (1988), Thompson (1997), page 62]. These conditions can be verified on a design-by-design basis. In the following corollary, we establish a central limit theorem for the pivotal statistic under simple random sampling.

COROLLARY 1. *Assume that the design is simple random sampling without replacement*, *and assume that* (A1)–(A7) *hold*. *Then*

$$\frac{N^{-1}(\tilde{t}_y - t_y)}{\widehat{V}^{1/2}(N^{-1}\tilde{t}_y)} \xrightarrow{\mathscr{L}} N(0,1)$$

as $N \to \infty$, *where* $\widehat{V}(N^{-1}\tilde{t}_y)$ *can be written as*

$$\widehat{V}(N^{-1}\tilde{t}_y) = \left(1 - \frac{n_N}{N}\right)\frac{\sum_{i\in s}(y_i - \hat{m}_i)^2 - n_N^{-1}[\sum_{i\in s}(y_i - \hat{m}_i)]^2}{n_N(n_N - 1)}.$$

PROOF. From the assumptions and Lemma 2(iv),

$$\limsup_{N\to\infty} N^{-1}\sum_{i\in U_N}(y_i - m_i)^4 < \infty,$$

from which the Lyapunov condition (3.25) of Thompson (1997) can be deduced. Note that

$$\mathrm{Var}_p\left(N^{-1}t_y^*\right) = \left(1 - \frac{n_N}{N}\right)\frac{\sum_{i\in U_N}(y_i - m_i)^2 - N^{-1}\left[\sum_{i\in U_N}(y_i - m_i)\right]^2}{n_N(N-1)}.$$

From Theorem 3.2 of Thompson (1997),

$$\frac{N^{-1}(t_y^* - t_y)}{\left(\operatorname{Var}_p(N^{-1}t_y^*)\right)^{1/2}} \xrightarrow{\mathscr{L}} N(0,1),$$

so that the result follows from Theorem 4. □

2.5. *Robustness*. In this section we consider the behavior of the *anticipated variance*,

$$\operatorname{Var}\left(N^{-1}(\tilde{t}_y - t_y)\right) = \mathrm{E}\left(N^{-1}(\tilde{t}_y - t_y)\right)^2 - \mathrm{E}^2\left[N^{-1}(\tilde{t}_y - t_y)\right],$$

where the expectation is taken over both design, p_N, and model, ξ. It can be shown from previous results that

$$\mathrm{E}^2\left[N^{-1}(\tilde{t}_y - t_y)\right] = o\left(n_N^{-1}\right),$$

so that the model-averaged design mean squared error and the anticipated variance are asymptotically equivalent in this case.

Godambe and Joshi (1965) showed that for any estimator $\hat{t}_y$ satisfying

$$\mathrm{E}\left[N^{-1}(\hat{t}_y - t_y)\right] = 0,$$

the following inequality holds:

$$\mathrm{E}\left(\frac{\hat{t}_y - t_y}{N}\right)^2 \geq \frac{1}{N^2} \sum_{i \in U_N} v(x_i) \frac{1 - \pi_i}{\pi_i}.$$

The right-hand side of the above expression is the Godambe–Joshi lower bound, which attains its minimum value when $\pi_i \propto v^{1/2}(x_i)$. Conditions under which generalized regression estimators asymptotically attain this lower bound have been studied by Wright (1983), Tam (1988) and others. In what follows, we prove that the local polynomial regression estimator is robust in the sense that it asymptotically attains the Godambe–Joshi lower bound.

THEOREM 5. *Under* (A1)–(A7), $\tilde{t}_y$ *asymptotically attains the Godambe–Joshi lower bound*, *in the sense that*

$$n_N \mathrm{E}\left(\frac{\tilde{t}_y - t_y}{N}\right)^2 = \frac{n_N}{N^2} \sum_{i \in U_N} v(x_i) \frac{1 - \pi_i}{\pi_i} + o(1).$$

PROOF. Write

$$b_N = \frac{n_N^{1/2}}{N} \sum_{i \in U_N} (m_i - \hat{m}_i)\left(\frac{I_i}{\pi_i} - 1\right),$$

$$c_N = \frac{n_N^{1/2}}{N} \sum_{i \in U_N} (y_i - m(x_i))\left(\frac{I_i}{\pi_i} - 1\right),$$

$$d_N = \frac{n_N^{1/2}}{N} \sum_{i \in U_N} (m(x_i) - m_i)\left(\frac{I_i}{\pi_i} - 1\right).$$

Then

$$n_N \mathrm{E}\left(\frac{\tilde{t}_y - t_y}{N}\right)^2 = \mathrm{E}[b_N^2] + \mathrm{E}[c_N^2] + \mathrm{E}[d_N^2] + 2\mathrm{E}[b_N c_N] + 2\mathrm{E}[b_N d_N] + 2\mathrm{E}[c_N d_N].$$

By Lemma 8, $\mathrm{E}[b_N^2] \to 0$ as $N \to \infty$. Next,

$$\begin{aligned}\mathrm{E}[d_N^2] &= \frac{n_N}{N^2} \sum_{i,\, j \in U_N} \mathrm{E}[(m_i - m(x_i))(m_j - m(x_j))] \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \\ &\le \left(\frac{n_N \max_{i,\, j \in U_N : i \ne j} |\pi_{ij} - \pi_i \pi_j|}{\lambda^2} + \frac{1}{\lambda}\right) \sum_{i \in U_N} \frac{\mathrm{E}(m_i - m(x_i))^2}{N} \\ &\to 0\end{aligned}$$

as $N \to \infty$ by Lemma 6. Note that

$$\mathrm{E}[c_N^2] = \frac{n_N}{N^2} \sum_{i \in U_N} v(x_i) \frac{1 - \pi_i}{\pi_i}$$

so that

$$\limsup_{N \to \infty} \mathrm{E}[c_N^2] \le \limsup_{N \to \infty} \frac{1}{N\lambda} \sum_{i \in U_N} v(x_i) < \infty$$

by (A3). The cross product terms go to zero as $N \to \infty$ by application of the Cauchy–Schwarz inequality, and the result is proved. □

3. Simulation results. In this section, we report on some simulation experiments comparing the performance of several estimators:

HT	Horvitz–Thompson	Equation (1)
REG	Linear regression	Cochran [(1977), page 193]
REG3	Cubic regression	
PS	Poststratification	Cochran [(1977), page 134]
LPR0	Local polynomial with $q = 0$	Equation (7)
LPR1	Local polynomial with $q = 1$	Equation (7)
KERN	Model-based nonparametric	Dorfman (1992)
CDW	Bias-calibrated nonparametric	Chambers, Dorfman and Wehrly (1993)

The first four estimators are parametric estimators (corresponding to constant, linear, cubic and piecewise constant mean functions) and the last four are nonparametric. The poststratification estimator is based on a division of the x-range into ten equally-sized strata. In practice, a survey designer with full auxiliary information $\{x_i\}_{i \in U_N}$ could implement an efficient stratification, as a referee has pointed out. Direct comparison of a prestratification approach to our estimation approach is difficult in a design-based setting, so we instead use poststratification. The number of poststrata was chosen to ensure a very small probability of empty poststrata. Note that poststratification can also be regarded as a special case of the empirical likelihood procedure described in Chen and Qin (1993), using auxiliary information on the deciles of the x-distribution.

Of the four nonparametric procedures, two are model-assisted (LPR0 and LPR1) and two are model-based (KERN and CDW). In KERN, the estimated mean function from a nonparametric procedure is used to predict each nonsampled y_i. The CDW estimator involves an additional bias calibration step, which requires specification of a working parametric model. We take the working model to be $m(x) = \beta x$, $v(x) = \sigma^2$ (this is the correct model for the first of our study populations). In KERN and CDW, we use the Nadaraya–Watson estimator, which is also used in LPR0 under equal-probability sampling.

The Epanechnikov kernel,

$$K(t) = \frac{3}{4}(1 - t^2) I_{\{|t| \le 1\}},$$

is used for all four nonparametric estimators. Two different bandwidths are considered: $h = 0.1$ and 0.25. The first bandwidth is equal to the poststratum width and the second is based on an ad hoc rule of 1/4th the data range. We have also considered but do not report on results for $h = 1.0$, which is a large bandwidth relative to the data range. Results for this third case confirm that as the bandwidth becomes large, LPR0 and KERN become numerically equivalent to HT under equal-probability sampling, while LPR1 becomes numerically equivalent to REG.

We consider eight mean functions:

Linear: $m_1(x) = 1 + 2(x - 0.5)$,
Quadratic: $m_2(x) = 1 + 2(x - 0.5)^2$,
Bump: $m_3(x) = 1 + 2(x - 0.5) + \exp(-200(x - 0.5)^2)$,
Jump: $m_4(x) = \{1 + 2(x - 0.5) I_{\{x \le 0.65\}}\} + 0.65 I_{\{x > 0.65\}}$,
cdf: $m_5(x) = \Phi(\frac{1.5 - 2x}{\sigma})$, where Φ is the standard normal cdf,
Exponential: $m_6(x) = \exp(-8x)$,
Cycle1: $m_7(x) = 2 + \sin(2\pi x)$,
Cycle4: $m_8(x) = 2 + \sin(8\pi x)$,

with $x \in [0, 1]$. These represent a range of correct and incorrect model specifications for the various estimators considered. For m_1, REG is expected to be the preferred estimator, since the assumed model is correctly specified. It is therefore interesting to see how much efficiency, if any, is lost by only assuming that the underlying model is smooth instead of linear. The remaining mean functions represent various departures from the linear model. For m_2, the trend is quadratic, so that an assumed linear model would be misspecified over the whole range of the x_k, but would be reasonable locally. The function m_3 is linear over most of its range, except for a "bump" present for a small portion of the range of x_k. The mean function m_4 is not smooth. The sigmoidal function m_5 is the mean of a binary random variable described below, and m_6 is an exponential curve. The function m_7 is a sinusoid completing one full cycle on $[0, 1]$, while m_8 completes four full cycles.

The population x_k are generated as independent and identically distributed (iid) uniform (0,1) random variables. The population values $y_{ik}\,(i = 1, \dots, 8)$ are generated from the mean functions by adding iid $N(0, \sigma^2)$ errors in all cases except cdf. The cdf population consists of binary measurements generated from the linear population via

$$y_{5k} = I_{\{y_{1k} \le 1.5\}}.$$

Note that the finite population mean of y_5 is $N^{-1} \sum_{k \in U_N} I_{\{y_{1k} \le 1.5\}}$, the finite population cdf of y_1, $F_1(t)$, at the point $t = 1.5$.

We evaluate two possible values for the standard deviation of the errors: $\sigma = 0.1$ and 0.4. The population is of size $N = 1000$. Samples are generated by simple random sampling using sample size $n = 100$. Other sample sizes of $n = 200$ and 500 have been considered but are not reported here. The effect of increasing sample size is similar to the effect of decreasing error standard deviation.

For each combination of mean function, standard deviation and bandwidth, 1000 replicate samples are selected and the estimators are calculated. Note that for each sample, a single set of weights is computed and applied to all eight study variables, as would be common practice in applications.

As the population is kept fixed during these 1000 replicates, we are able to evaluate the design-averaged performance of the estimators. Specifically, we estimate the design bias, design variance and design mean squared error. For nearly all cases in this simulation, the percent relative design biases

$$\frac{\mathrm{E}_p\big[\hat{t}_y\big] - t_y}{t_y} \times 100\%$$

were less than one percent for all estimators, and are not presented. (The exceptions were percent relative design biases, in rare cases, of up to 12% for the model-based nonparametric procedures.)

Table 1 shows the ratios of MSEs for the various estimators to the MSE for the local polynomial regression estimator with $q = 1$ (LPR1). Generally, both parametric and nonparametric regression estimators perform better than HT, regardless of whether the underlying model is correctly specified or not, but that effect decreases as the model variance increases.

With a few exceptions, the model-based nonparametric estimators KERN and CDW have similar behavior in this study. With respect to design MSE, the model-assisted estimators LPR0 and LPR1 are sometimes much better and never much worse (MSE ratio ≥ 0.95) than the model-based estimators KERN and CDW. Similarly, LPR1 is sometimes much better and never much worse than LPR0 in this study, and so LPR1 emerges as the best among the nonparametric estimators considered here.

TABLE 1

Ratio of MSE *of Horvitz–Thompson* (HT), *linear regression* (REG), *cubic regression* (REG3), *poststratification* (PS), *local constant regression* (LPR0), *model-based kernel* (KERN) *and bias-calibrated nonparametric* (CDW) *estimators to local linear regression* (LPR1) *estimator*

Population	**σ**	**h**	**HT**	**REG**	**REG3**	**PS**	**LPR0**	**KERN**	**CDW**
linear	0.1	0.10	33.35	0.09	0.94	1.43	0.98	0.98	0.95
	0.1	0.25	35.41	0.95	1.00	1.52	1.45	1.68	0.98
	0.4	0.10	2.97	0.90	0.94	1.05	0.96	0.95	0.95
	0.4	0.25	3.16	0.95	1.00	1.12	1.00	1.02	0.98
quadratic	0.1	0.10	2.96	3.02	0.94	1.15	0.99	1.02	1.08
	0.1	0.25	3.08	3.14	0.98	1.20	1.29	2.16	2.70
	0.4	0.10	1.01	1.02	0.94	1.03	0.96	0.95	0.96
	0.4	0.25	1.07	1.08	1.00	1.10	1.00	1.07	1.11
bump	0.1	0.10	25.06	4.90	4.00	2.08	1.13	1.50	1.47
	0.1	0.25	8.57	1.67	1.37	0.71	1.13	1.32	1.14
	0.4	0.10	3.43	1.35	1.30	1.15	0.98	1.01	1.01
	0.4	0.25	2.94	1.16	1.11	0.98	1.01	1.07	1.03
jump	0.1	0.10	7.12	5.24	2.08	1.51	1.00	1.08	1.07
	0.1	0.25	4.88	3.59	1.42	1.03	1.13	1.53	1.46
	0.4	0.10	1.47	1.30	1.05	1.07	0.96	0.95	0.95
	0.4	0.25	1.51	1.33	1.07	1.09	1.00	1.05	1.04
cdf	0.1	0.10	6.37	3.00	1.64	1.20	1.00	1.02	1.02
	0.1	0.25	5.00	2.35	1.29	0.94	1.09	1.44	1.58
	0.4	0.10	1.58	1.05	0.97	1.05	0.98	0.97	0.97
	0.4	0.25	1.64	1.09	0.01	1.08	1.00	1.04	1.08
exponential	0.1	0.10	5.21	2.90	1.07	1.36	1.10	1.19	1.19
	0.1	0.25	4.88	2.72	1.00	1.27	1.64	2.44	2.45
	0.4	0.10	1.14	1.02	0.96	1.05	0.97	0.96	0.96
	0.4	0.25	1.20	1.07	1.00	1.10	1.03	1.10	1.10
cycle1	0.1	0.10	46.58	18.68	1.36	2.73	1.24	1.46	1.25
	0.1	0.25	16.15	6.48	0.47	0.94	1.66	2.55	2.34
	0.4	0.10	3.91	2.08	0.99	1.15	0.97	0.97	0.96
	0.4	0.25	3.67	1.95	0.93	1.08	1.09	1.26	1.23
cycle4	0.1	0.10	3.85	3.79	3.56	1.83	1.21	1.97	2.02
	0.1	0.25	0.98	0.96	0.91	0.46	1.09	1.00	1.09
	0.4	0.10	2.27	2.26	2.17	1.35	1.07	1.42	1.45
	0.4	0.25	0.97	0.97	0.93	0.58	1.07	1.01	1.08

Based on 1000 replications of simple random sampling from eight fixed populations of size $N = 1000$.

Sample size is $n = 100$.

Nonparametric estimators are computed with bandwidth h and Epanechnikov kernel.

Among the parametric estimators in this study, the higher-order parametric estimators (REG3 and PS) generally perform better than REG, except in the linear population. In most cases, LPR1 is competitive or better than the parametric estimators (MSE ratios ≥ 0.95). In several cases, the parametric estimators are somewhat better than LPR1 (MSE ratios 0.90–0.94). This is due to undersmoothing when the population is linear or quadratic and is due to oversmoothing in other cases. Finally, the PS estimator for the bump and cycle4 populations and the REG3 estimator for cycle1 are substantially better than the oversmoothed LPR1 estimator. In each of these cases, however, the LPR1 estimator at the smaller bandwidth is much better than the corresponding parametric estimator.

Overall, then, the performance of the LPR1 estimator is consistently good, particularly at the smaller bandwidth. LPR1 loses a small amount of efficiency relative to REG for a linear population, but dominates REG for other populations. It dominates the other nonparametric estimators considered here and it dominates the higher-order parametric estimators provided it is not oversmoothed.

One common concern when using nonparametric regression techniques is how sensitive the results are to the choice of the smoothing parameter. This is especially important in the context of survey sampling, because the same set of regression weights (with a single choice for the bandwidth) are often used for a large number of different variables, as was done in the simulation experiment described above. A variable-specific bandwidth selection procedure based on cross-validation is currently being investigated by the authors and might be appropriate when the primary emphasis of a study is on achieving the best possible precision for a small number of variables. The investigation of such automated bandwidth selection procedures will be reported on elsewhere.

As the bandwidth becomes large, the local linear regression estimator becomes equivalent to the classical regression estimator and the MSEs converge. Clearly, the bandwidth has an effect on the MSE of LPR1, but Table 1 suggests that large gains in efficiency over other estimators can be gained for a variety of bandwidth choices. In particular, for either of the bandwidths considered here, LPR1 essentially dominates HT for all populations and essentially dominates REG for all populations except linear, where it is competitive. This again shows that the local linear regression estimator is likely to be an improvement over Horvitz–Thompson and classical regression estimation when the relationship between the auxiliary variable and the variable of interest is nonlinear.

We have investigated the performance of the estimated variances from Corollary 1 relative to simulation variances. The ratios are generally close to 1, though there is a fairly large amount of variability for sample size $n = 100$. The performance of the variance estimators for LPR1 is similar to the performance of variance estimators for REG in the cases we have examined. Further investigation of variance estimation for local polynomial regression estimators, including the use of the weighted residual technique [Särndal, Swensson and Wretman (1989)], will be reported elsewhere.

APPENDIX

Lemmas.

LEMMA 1. *Assume* (A2) *and* (A5). *Then*

$$\left|\frac{F_N(x+h_N)-F_N(x-h_N)}{2h_N}-f(x)\right|\to 0 \tag{15}$$

as $N\to\infty$, *uniformly in* x.

PROOF. Define $D_N = \sup_x |F_N(x) - F(x)|$. By (A2) and the law of the iterated logarithm [Serfling (1980), page 62], D_N satisfies

$$\limsup_{N\to\infty}\frac{2N^{1/2}D_N}{(2\log\log N)^{1/2}} = 1. \tag{16}$$

Let $\epsilon > 0$ be given. Then there exists $\eta > 0$ such that $|x - x'| < \eta$ implies that $|f(x) - f(x')| < \epsilon/2$, by uniform continuity of f on $[a_x, b_x]$. Using (16) and (A5), choose N^* so large that $N \geq N^*$ implies that

$$h_N < \frac{\eta}{2},\ \frac{2N^{1/2}D_N}{(2\log\log N)^{1/2}} < 1+\epsilon \quad\text{and}\quad \frac{(2\log\log N)^{1/2}}{2N^{1/2}h_N} < \frac{\epsilon}{2(1+\epsilon)}. \tag{17}$$

[Note that $x' \in (x - h_N, x + h_N)$ implies that $|x - x'| < 2h_N < \eta$.] It follows that for any x,

$$\begin{aligned}
&\left|\frac{F_N(x+h_N)-F(x+h_N)}{2h_N}+\frac{F(x-h_N)-F_N(x-h_N)}{2h_N}\right.\\
&\qquad\qquad\left.+\frac{F(x+h_N)-F(x-h_N)}{2h_N}-f(x)\right|\\
&\quad\leq \frac{D_N}{2h_N}+\frac{D_N}{2h_N}+\left|\frac{F(x+h_N)-F(x-h_N)}{2h_N}-f(x)\right| \qquad (18)\\
&\quad= \frac{2N^{1/2}D_N}{(2\log\log N)^{1/2}}\frac{(2\log\log N)^{1/2}}{2N^{1/2}h_N}+|f(x_N)-f(x)|\\
&\quad\leq (1+\epsilon)\frac{\epsilon}{2(1+\epsilon)}+\frac{\epsilon}{2}=\epsilon,
\end{aligned}$$

where the equality in (18) holds for some $x_N \in (x - h_N, x + h_N)$ by the mean value theorem. □

The uniform convergence in Lemma 1 has a number of useful consequences.

LEMMA 2. *Under* (A1)–(A6):

(i) *For* $k \geq 0$,

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{i\in U_N} \left(\frac{1}{2Nh_N} \sum_{j\in U_N} I_{\{x_i - h_N \leq x_j \leq x_i + h_N\}} \right)^k < \infty.$$

(ii) *There exists* N^*, *independent of* x, *such that* $N \geq N^*$ *implies*

$$\sum_{k\in U_N} I_{\{|x-x_k|\leq h_N\}} \geq q+1.$$

(iii) *The* $N^{-1}t_{ig}$ *are uniformly bounded in* i *and the* $N^{-1}\hat{t}_{ig}$ *are uniformly bounded in* i *and* s.

(iv) *The* m_i *are uniformly bounded in* i *and the* $\hat{m}_i$ *are uniformly bounded in* i *and* s.

(v) *The first, second, third and fourth order mixed partials of* $\hat{m}_i$ *with respect to* $N^{-1}t_{ig}$ *and* δ, *evaluated at* $\hat{\mathbf{t}}_i = \mathbf{t}_i$, $\delta = 0$, *are uniformly bounded in* i.

(vi) *The* R^2_{iN} *are uniformly bounded in* i *and* s.

PROOF. (i) By Lemma 1,

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \sum_{i\in U_N} \left(\frac{1}{2Nh_N} \sum_{j\in U_N} I_{\{x_i - h_N \leq x_j \leq x_i + h_N\}} \right)^k \\
&\quad \leq \limsup_{N\to\infty} \frac{1}{N} \sum_{i\in U_N} \left\{ \left| \frac{F_N(x_i+h_N) - F_N(x_i-h_N)}{2h_N} - f(x_i) \right| + f(x_i) \right\}^k \\
&\quad \leq \limsup_{N\to\infty} \frac{1}{N} \sum_{i\in U_N} \{\epsilon + f(x_i)\}^k \\
&\quad < \infty.
\end{aligned}$$

(ii) If not, then we could set $\epsilon = \min_x f(x)/2 > 0$ by compactness of the support and continuity of f, and choose N^* so large that $N \geq N^*$ satisfies (17) and implies $(q+1)/(2Nh_N) < \epsilon$. For some x and some $N \geq N^*$, $\sum_{k\in U_N} I_{\{|x-x_k|\leq h_N\}} < q+1$, so that

$$\begin{aligned}
f(x) - \frac{\sum_{k\in U_N} I_{\{|x-x_k|\leq h_N\}}}{2Nh_N} &> f(x) - \frac{q+1}{2Nh_N} \\
&> \min_x f(x) - \frac{\min_x f(x)}{2} = \epsilon,
\end{aligned}$$

contradicting the uniform convergence in (15).

(iii) Under the given assumptions,

$$\begin{aligned}
\limsup_{N\to\infty} |N^{-1}\hat{t}_{ig}| &= \limsup_{N\to\infty} \left| \sum_{k\in U_N} \frac{1}{Nh_N} K\left(\frac{x_k - x_i}{h_N} \right) (x_k - x_i)^{p_1} y_k^{p_2} \frac{I_k}{\pi_k} \right| \\
&\leq \limsup_{N\to\infty} \sum_{k\in U_N} \frac{c}{Nh_N\lambda} I_{\{x_i - h_N \leq x_k \leq x_i + h_N\}},
\end{aligned}$$

which does not depend on s, and is bounded independently of i by Lemma 1. Since $\lambda < 1$, the same uniform bound works for $N^{-1}t_{ig}$.

(iv), (v), (vi) The m_i are continuous functions of the uniformly bounded t_{ig}, with denominators uniformly bounded away from zero by (ii) above. Similarly, the $\hat{m}_i$ and their derivatives are continuous functions of the uniformly bounded $\hat{t}_{ig}$, with denominators uniformly bounded away from zero by the adjustment in (8). Combining these results and using the definition in (10), the R_{iN}^2 are uniformly bounded in i and s. $\square$

LEMMA 3. *Assume* (A1)–(A7). *For the Taylor linearization remainders of the sample local polynomial residuals in* (10),

$$\frac{n_N}{N}\sum_{i\in U_N}\mathrm{E}_p\Big[R_{iN}^2\Big] = O\bigg(\frac{1}{n_N h_N^2}\bigg).$$

PROOF. Note that

$$\begin{aligned}
&\frac{n_N^2 h_N^2}{N}\sum_{i\in U_N}\mathrm{E}_p|N^{-1}(\hat{t}_{ig}-t_{ig})|^4\\
&\quad\le \frac{cn_N^2h_N^2}{N^5h_N^4}\sum_{i,\,j,\,k,\,l,\,m\in U_N} I_{\{x_i-h_N\le x_j,\,x_k,\,x_l,\,x_m\le x_i+h_N\}}\\
&\qquad\qquad\qquad\times\Big|\mathrm{E}_p[(I_j-\pi_j)(I_k-\pi_k)(I_l-\pi_l)(I_m-\pi_m)]\Big|\\
&\quad\le c_1h_N^2n_N^2\max_{(j,\,k,\,l,\,m)\in D_{4,\,N}}\Big|\mathrm{E}_p[(I_j-\pi_j)(I_k-\pi_k)(I_l-\pi_l)(I_m-\pi_m)]\Big|\\
&\qquad\times\frac{1}{N}\sum_{i\in U_N}\bigg(\sum_{j\in U_N}\frac{I_{\{x_i-h_N\le x_j\le x_i+h_N\}}}{Nh_N}\bigg)^4\\
&\qquad+c_2\frac{n_Nh_N^2}{Nh_N}n_N\max_{(j,\,k,\,l)\in D_{3,\,N}}\Big|\mathrm{E}_p\Big[(I_j-\pi_j)^2(I_k-\pi_k)(I_l-\pi_l)\Big]\Big|\\
&\qquad\times\frac{1}{N}\sum_{i\in U_N}\bigg(\sum_{j\in U_N}\frac{I_{\{x_i-h_N\le x_j\le x_i+h_N\}}}{Nh_N}\bigg)^3\\
&\qquad+c_3\frac{n_N^2h_N^2}{N^2h_N^2}\frac{1}{N}\sum_{i\in U_N}\bigg(\sum_{j\in U_N}\frac{I_{\{x_i-h_N\le x_j\le x_i+h_N\}}}{Nh_N}\bigg)^2\\
&\qquad+c_4\frac{n_N^2h_N^2}{N^3h_N^3}\frac{1}{N}\sum_{i\in U_N}\bigg(\sum_{j\in U_N}\frac{I_{\{x_i-h_N\le x_j\le x_i+h_N\}}}{Nh_N}\bigg),
\end{aligned}$$

which remains bounded by (A5), (A7) and Lemma 2(i).

The assumptions of Theorem 5.4.3 of Fuller (1996) with $\alpha = 1$, $s = 4$, $a_N = O((n_Nh_N)^{-2})$ and expectation $N^{-1}\sum_{i\in U_N}\mathrm{E}_p[\cdot]$ are then met for the sequence

$\{R_{iN}^2\}$. Since this function and its first three derivatives with respect to the elements of $(N^{-1}\mathbf{t}_i, \delta)$ evaluate to zero, we conclude that

$$\frac{n_N}{N}\sum_{i\in U_N}\mathrm{E}_p\Big[R_{iN}^2\Big] = O\bigg(\frac{1}{n_N h_N^2}\bigg),$$

which goes to zero by (A5). □

LEMMA 4. *Assume* (A1)–(A7). *Then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i\in U_N}\mathrm{E}_p(\hat{m}_i - m_i)^2 = 0.$$

PROOF. By (10),

$$\begin{aligned}\frac{1}{N}\sum_{i\in U_N}\mathrm{E}_p(\hat{m}_i - m_i)^2 &= \frac{1}{N^3}\sum_{i\in U_N}\sum_{k,\,l\in U_N} z_{ik}z_{il}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l}\\ &\quad+\frac{2}{N^2}\sum_{i,\,k\in U_N} z_{ik}\,\mathrm{E}_p\bigg[\bigg(1-\frac{I_k}{\pi_k}\bigg)R_{iN}\bigg]\\ &\quad+\frac{1}{N}\sum_{i\in U_N}\mathrm{E}_p\Big[R_{iN}^2\Big] + o\bigg(\frac{\delta}{N^2}\bigg),\end{aligned}\tag{19}$$

where the remainder term $o(\delta N^{-2})$ comes from the expansion (10) and does not depend on the sample. By Lemma 1 and Lemma 2(v),

$$\frac{1}{N^3}\sum_{i,\,k\in U_N} z_{ik}^2 \le \frac{c}{N^3 h_N^2}\sum_{i,\,k\in U_N} I_{\{x_i - h_N\le x_k\le x_i + h_N\}} \to 0$$

as $N\to\infty$. Thus, following the argument of Theorem 1 in Robinson and Särndal (1983), the first term of (19) is

$$\begin{aligned}&\frac{1}{N^3}\sum_{i\in U_N}\sum_{k,\,l\in U_N} z_{ik}z_{il}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l}\\ &\quad=\frac{1}{N^3}\sum_{i\in U_N}\sum_{k\in U_N} z_{ik}^2\frac{1-\pi_k}{\pi_k} + \frac{1}{N^3}\sum_{i\in U_N}\sum_{k\neq l} z_{ik}z_{il}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l}\\ &\quad\le\frac{1}{\lambda N^3}\sum_{i,\,k\in U_N} z_{ik}^2 + \frac{N\max_{i,\,j\in U_N:\,i\neq j}|\pi_{ij}-\pi_i\pi_j|}{\lambda^2 N^3}\sum_{i,\,k\in U_N} z_{ik}^2,\end{aligned}$$

which converges to zero using (A6).

The last term of (19) converges to zero by Lemma 3, and the second term converges to zero by an application of the Cauchy–Schwarz inequality. □

LEMMA 5. *Assume* (A1)–(A7). *Then*

$$\lim_{N\to\infty}\frac{n_N}{N^2}\mathrm{E}_p\bigg[\sum_{i,\,j\in U_N}(\hat{m}_i - m_i)(\hat{m}_j - m_j)\bigg(1-\frac{I_i}{\pi_i}\bigg)\bigg(1-\frac{I_j}{\pi_j}\bigg)\bigg] = 0.$$

PROOF. By (10),

$$\begin{aligned}
&\frac{n_N}{N^2}\mathrm{E}_p\left[\sum_{i,\,j\in U_N}(\hat{m}_i - m_i)(\hat{m}_j - m_j)\left(1-\frac{I_i}{\pi_i}\right)\left(1-\frac{I_j}{\pi_j}\right)\right]\\
&\quad= \frac{n_N}{N^4}\sum_{i,\,j,\,k,\,l\in U_N} z_{ik}z_{jl}\mathrm{E}_p\left[\left(1-\frac{I_i}{\pi_i}\right)\left(1-\frac{I_j}{\pi_j}\right)\left(1-\frac{I_k}{\pi_k}\right)\left(1-\frac{I_l}{\pi_l}\right)\right]\\
&\qquad+\frac{2n_N}{N^3}\sum_{i,\,j,\,k\in U_N} z_{ik}\mathrm{E}_p\left[R_{jN}\left(1-\frac{I_i}{\pi_i}\right)\left(1-\frac{I_j}{\pi_j}\right)\left(1-\frac{I_k}{\pi_k}\right)\right]\\
&\qquad+\frac{n_N}{N^2}\sum_{i,\,j\in U_N}\mathrm{E}_p\left[R_{iN}R_{jN}\left(1-\frac{I_i}{\pi_i}\right)\left(1-\frac{I_j}{\pi_j}\right)\right]+o(1)\\
&\quad= b_{1N}+b_{2N}+b_{3N}+o(1).
\end{aligned}$$

The remainder term $o(1)$ comes from the $O(\delta N^{-2})$ term in (10), using the Cauchy–Schwarz inequality for its two cross-products. In b_{1N}, we consider separately the cases of one, two, three and four distinct elements in (i, j, k, l). Straightforward bounding arguments like those in Lemma 3 show that each such case converges to zero. We omit the details. The term b_{3N} converges to zero by Lemma 3 and (A6). The cross-product term b_{2N} goes to zero by an application of the Cauchy–Schwarz inequality, and the result is proved. □

LEMMA 6. *Under* (A1)–(A5),

$$\lim_{N\to\infty} N^{-1}\sum_{i\in U_N}\mathrm{E}(m_i - m(x_i))^2 = 0.$$

PROOF. This follows directly from standard local polynomial regression theory [e.g., Wand and Jones (1995), page 125]. □

LEMMA 7. *Assume* (A1)–(A7). *Then*,

$$\lim_{N\to\infty}\frac{n_N}{N}\sum_{i\in U_N}\mathrm{E}\left[R_{iN}^2\right] = 0.$$

PROOF. The proof is identical to that of Lemma 3, after replacing the expectation operator with $N^{-1}\sum_{i\in U_N}\mathrm{E}[\cdot]$, because the uniformity results of Lemma 1 and Lemma 2 hold not only for all $i \in U_N$ and s, but also across realizations from ξ. □

LEMMA 8. *Assume* (A1)–(A7) *hold. Then*,

$$\lim_{N\to\infty}\frac{n_N}{N^2}\mathrm{E}\left[\sum_{i,\,j\in U_N}(\hat{m}_i - m_i)(\hat{m}_j - m_j)\left(1-\frac{I_i}{\pi_i}\right)\left(1-\frac{I_j}{\pi_j}\right)\right] = 0.$$

PROOF. The result follows from the assumptions and Lemma 7, using bounding arguments exactly as in Lemma 5. □

Acknowledgments. We are grateful to four anonymous referees for numerous constructive suggestions. The work of the first author was conducted while he was Associate Professor of Statistics of Iowa State University.

REFERENCES

BREWER, K. R. W. (1963). Ratio estimation in finite populations: some results deductible from the assumption of an underlying stochastic process. *Austral. J. Statist.* **5** 93–105.

CASSEL, C.-M., SÄRNDAL, C.-E. and WRETMAN, J. H. (1977). *Foundations of Inference in Survey Sampling.* Wiley, New York.

CHAMBERS, R. L. (1996). Robust case-weighting for multipurpose establishment surveys. *J. Official Statist.* **12** 3–32.

CHAMBERS, R. L., DORFMAN, A. H. and WEHRLY, T. E. (1993). Bias robust estimation in finite populations using nonparametric calibration. *J. Amer. Statist. Assoc.* **88** 268–277.

CHEN, J. and QIN, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80** 107–116.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

CLEVELAND, W. S. and DEVLIN, S. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83** 596–610.

COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York.

DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382.

DORFMAN, A. H. (1992). Nonparametric regression for estimating totals in finite populations. *Proceedings of the Section on Survey Research Methods* 622–625. Amer. Statist. Assoc., Alexandria, VA.

DORFMAN, A. H. and HALL, P. (1993). Estimators of the finite population distribution function using nonparametric regression. *Ann. Statist.* **21** 1452–1475.

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman and Hall, London.

FULLER, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed. Wiley, New York.

GODAMBE, V. P. and JOSHI, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations I. *Ann. Math. Statist.* **36** 1707–1722.

HALL, P. and TURLACH, B. A. (1997). Interpolation methods for adapting to sparse design in nonparametric regression. *J. Amer. Statist. Assoc.* **92** 466–472.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685.

ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96.

KUO, L. (1998). Classical and prediction approaches to estimating distribution functions from survey data. *Proceedings of the Section on Survey Research Methods* 280–285. Amer. Statist. Assoc., Alexandria, VA.

OPSOMER, J.-D. and RUPPERT, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25** 186–211.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

ROBINSON, P. M. and SÄRNDAL, C.-E. (1983). Asymptotic properties of the generalized regression estimation in probability sampling. *Sankhyā. Ser. B* **45** 240–248.

ROYALL, R. M. (1970). On finite population sampling under certain linear regression models. *Biometrika* **57** 377–387.

RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370.

SÄRNDAL, C.-E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika* **67** 639–650.

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika* **76** 527–537.

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling.* Springer, New York.

SEN, P. K. (1988). Asymptotics in finite population sampling. In *Handbook of Statistics* (P. R. Krishnaiah and C. R. Rao, eds.) **6** 291–331. North-Holland, Amsterdam.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

TAM, S. M. (1988). Some results on robust estimation in finite population sampling. *J. Amer. Statist. Assoc.* **83** 242–248.

THOMPSON, M. E. (1997). *Theory of Sample Surveys.* Chapman and Hall, London.

WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

WRIGHT, R. L. (1983). Finite population sampling with multivariate auxiliary information. *J. Amer. Statist. Assoc.* **78** 879–884.

DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523-1877
E-MAIL: jbreidt@stat.colostate.edu

STATISTICAL LABORATORY
AND DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY
221 SNEDECOR HALL
AMES, IOWA 50011-1210
E-MAIL: jopsomer@iastate.edu