NISS

On the asymptotic normality and variance estimation for nondifferentiable survey estimators

Jianqiang C. Wang and J.D. Opsomer

Technical Report 176
November 2010

National Institute of Statistical Sciences
19 T.W. Alexander Drive
PO Box 14006
Research Triangle Park, NC
www.niss.org

Biometrika (2010), **xx**, x, *pp*. 1–16

Printed in Great Britain

On the asymptotic normality and variance estimation for nondifferentiable survey estimators

BY JIANQIANG C. WANG

National Institute of Statistical Sciences; Durham, North Carolina 27709, U.S.A.
jqwang@niss.org

AND J. D. OPSOMER

Department of Statistics, Colorado State University; Fort Collins, Colorado 80523, U.S.A.
jopsomer@stat.colostate.edu

SUMMARY

Survey estimators of population quantities such as distribution functions and quantiles contain nondifferentiable functions of estimated quantities. The theoretical properties of such estimators are substantially more complicated to derive than those of differentiable estimators. In this article, we provide a unified framework for obtaining the asymptotic design-based properties of two common types of nondifferentiable estimators. Estimators of the first type have an explicit expression, while those of the second are defined only as the solution to estimating equations. We propose both analytical and replication-based design consistent variance estimators for both cases, based on kernel regression. The practical behavior of the variance estimators is demonstrated in a simulation experiment. Our simulation suggests that the proposed variance estimators work reasonably well under the appropriate bandwidth.

Some key words: estimating equation, kernel regression, nondifferentiable estimator, replication variance estimation.

1. INTRODUCTION

A number of common survey estimators, including estimators of population distribution functions and quantiles, involve nondifferentiable functions of estimated quantities. Because of this nondifferentiability, these estimators do not follow the standard paradigm for obtaining the statistical properties of survey estimators, which relies on Taylor linearization. Statisticians wanting to work with this type of estimators are faced with the choice of either developing a customized approach for their particular estimator, or of glossing over the nondifferentiability. In this article, we shall consider two types of nondifferentiable estimators and provide a unified theoretical framework under which their properties can be obtained. The first comprises explicitly defined estimators, in which one or several estimated quantities are embedded inside a nondifferentiable function. The second comprises estimators defined as the solution to estimating equations, with the equations containing nondifferentiable components.

Examples of the first type of estimators include estimators of population distribution function using auxiliary information (Dunstan & Chambers, 1986; Rao et al., 1990; Chambers et al., 1992; Wang & Dorfman, 1996), estimators of population fraction above or below an estimated quantity (Shao & Rao, 1993; Binder & Kovacevic, 1995; Preston, 1996; Eurostat, 2000; Berger & Skinner, 2003), the endogenous post-stratification estimator (Breidt & Opsomer, 2008) and the estimator of the population distribution of distances to subpopulation centre (Wang & Opsomer, 2010). In all these cases, the estimation targets are finite population quantities, so we focus exclusively on design-based estimators. Many of the authors above obtained theoretical properties of their specific estimators, often taking advantage of the fact that the nondifferentiability is due to indicator functions. A more general treatment of nondifferentiable estimators in survey context is provided by Deville (1999), who described variance estimation for complex statistics using influence functions and introduced kernel smoothing in variance estimation. However, no formal proof was provided and there is no unified theoretical work establishing the asymptotic properties of this class of estimators under a complex survey design.

The second type of nondifferentiable estimators under consideration involves design-weighted estimating equations. Godambe & Thompson (2009) gave a general treatment of estimating equations in survey sampling, and showed how quantities of interest can be defined through estimating equations, including means, quantiles and generalized linear model parameters (see also Binder, 1983; Wu & Sitter, 2001). Section 1.3.4 of Fuller (2009) derived properties of estimators defined by estimating equations in complex surveys when the estimating function satisfies a differentiability condition. A specific example of an estimator defined as the solution to nondifferentiable estimating equations is the sample quantile. Kuk & Mak (1989) discussed median estimation using auxiliary information under simple random sampling. Rao et al. (1990) furnished a thorough treatment of estimating distribution functions and quantiles in the presence of auxiliary information under a general sampling design. Francisco & Fuller (1991) derived the design normality of both distribution function and quantile estimators, and proposed a number of confidence intervals for quantiles, including Woodruff confidence intervals, further examined by Sitter & Wu (2001). To our knowledge, a general theoretical treatment of survey estimators with nondifferentiable estimating equations is not available in literature.

Many of the above estimators are readily handled in a model-based context, where the sample observations can be treated as independent and identically realizations from a model. The seminal article by Randles (1982) gave a unified treatment of nondifferentiable functions with estimated parameters when the estimator can be written as a U-statistic. However, when these estimators are considered under an unequal-probability design-based paradigm in which the randomness comes from the sampling design and the population remains fixed, a corresponding unified treatment is not available. Given the prevalence of design-based inference for government and other surveys, we attempt to provide a unified approach for handling nondifferentiable survey estimators. For both types of estimators described above, we state a full set of design, population and estimator assumptions that are sufficient to obtain design consistency and asymptotic normality. We also propose design consistent variance estimators that use kernel regression to estimate the smooth limit of the nondifferentiable functions.

2. GENERAL DESIGN ASSUMPTIONS

In this section, we give assumptions on the sampling design and estimator that are sufficient to obtain the asymptotic properties of a Horvitz–Thompson estimator for a quantity with certain moment conditions. Additional assumptions for specific classes of estimators will be stated in later sections. We follow the framework of Isaki & Fuller (1982) in which the properties of estimators are established under a fixed sequence of populations and a corresponding sequence of random sampling designs. Suppose therefore that we have an increasing sequence of finite populations $\{U_N\}$ of size N, with $N \to \infty$. Associated with population element i is a vector of observations $y_i = (y_{i,1}, \ldots, y_{i,p})$, and let $\mathcal{F}_N$ denote $\{y_1, \ldots, y_N\}$ containing all the variables of interest in the Nth population.

We take a sample $\mathcal{S}$ of size n from population U_N, and the sampling design generating $\mathcal{S}$ may be a complex design with stratification or multi-stage sampling. Let $\pi_i = \mathrm{pr}(i \in \mathcal{S})$ represent the inclusion probability of the ith population element. We write $\bar{y}_N = N^{-1} \sum_{i=1}^{N} y_i$ for the population mean of variable y_i and $\bar{y}_\pi = N^{-1} \sum_{\mathcal{S}} {\pi_i}^{-1} y_i$ for its Horvitz–Thompson estimator.

We state three assumptions on the probability sampling design. Assumption 1 sets limits on the rate of the sample size, with a more restrictive version in Assumption 1(A) that is needed only for explicitly defined nondifferentiable estimators. Assumption 2 ensures design consistency and Assumption 3 guarantees asymptotic normality of our estimator under a general design.

ASSUMPTION 1. *The expected sample size* $n^* = E(n \mid \mathcal{F}_N) = O(N^\beta)$, *with* $1/2 < \beta \leq 1$.

ASSUMPTION 1 (A). *The expected sample size* $n^* = O(N^\beta)$, *with* $2p/(2p+1) < \beta \leq 1$, *where* p *denotes the dimension of study variable* y,

ASSUMPTION 2. *The following conditions hold for the inclusion probabilities* π_i *and the design variance of the Horvitz–Thompson estimator of the mean,*

1. $K_L \leq N\pi_i/n^* \leq K_U$ *for all* i, *where* K_L *and* K_U *are positive constants;*
2. *for any vector* z *with finite* $2+\delta$ *population moments with arbitrarily small* $\delta > 0$, *we assume* $\mathrm{var}(\bar{z}_\pi \mid \mathcal{F}_N) \leq c_1 n^{*-1}(N-1)^{-1} \sum_{i=1}^{N} (z_i - \bar{z}_N)(z_i - \bar{z}_N)^{\mathrm{T}}$, *for some constant* c_1.

ASSUMPTION 3. *For any* z *with finite fourth population moment and conditional on* $\mathcal{F}_N$,

$$\mathrm{var}(\bar{z}_\pi \mid \mathcal{F}_N)^{-1/2}(\bar{z}_\pi - \bar{z}_N) \to N(0, \mathrm{I}_p), \tag{1}$$

in distribution, and

$$\mathrm{var}(\bar{z}_\pi \mid \mathcal{F}_N)^{-1} \widehat{V}_{\mathrm{HT}}(\bar{z}_\pi) - \mathrm{I}_p = O_p(n^{*-1/2}), \tag{2}$$

where I_p *is the* $p \times p$ *identity matrix, the design variance-covariance matrix of* $\bar{z}_\pi$, $\mathrm{var}(\bar{z}_\pi \mid \mathcal{F}_N)$, *is positive definite, and* $\widehat{V}_{\mathrm{HT}}(\bar{z}_\pi)$ *is the Horvitz–Thompson estimator of* $\mathrm{var}(\bar{z}_\pi \mid \mathcal{F}_N)$.

In this article and unless specifically indicated otherwise, convergence results are to be interpreted as being with respect to the sequence of sampling designs, conditional on $\mathcal{F}_N$.

3. EXPLICITLY DEFINED NONDIFFERENTIABLE SURVEY ESTIMATORS

3·1. *The Estimators*

Assume that a population quantity of interest takes the form of a U-statistic of order one,

$$T_N(\lambda_N) = \frac{1}{N} \sum_{i=1}^{N} h(y_i; \lambda_N), \tag{3}$$

where λ_N represents a q-dimensional population quantity and $h(y;\lambda) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ is not necessarily a differentiable function of λ. The two integers p and q represent the dimension of the target variable y and parameter λ respectively, and need not be the same. The sample estimator of $T_N(\lambda_N)$ is the following Horvitz–Thompson estimator with estimated parameter(s),

$$\widehat{T}_N(\hat{\lambda}_N) = \frac{1}{N} \sum_{i \in \mathcal{S}} \frac{1}{\pi_i} h(y_i; \hat{\lambda}_N), \tag{4}$$

where $\hat{\lambda}_N$ is a sample-based estimator of λ_N.

The case when $h(y;\lambda)$ is a smooth function of λ is easy to deal with, because we can apply Taylor linearization and obtain the ignorability of the remainder terms in the expansion using traditional arguments. But if $h(y;\lambda)$ is a nondifferentiable function of λ, we cannot express the extra variation by a direct linearization, so that further steps need to be taken to study the asymptotic properties of the estimator. Randles (1982) gave a general treatment of nondifferentiable estimators in a nonsurvey setting. In the current context, if $h(y;\lambda)$ is a nonsmooth function of λ, the expectation of $\widehat{T}_N(\lambda)$ under the design, $T_N(\lambda)$, remains as a nonsmooth function of λ, so we need to modify the approach of Randles (1982) to extend the results to the survey context.

3·2. *Assumptions*

We provide a set of conditions that need to be satisfied by the parameter estimator $\hat{\lambda}_N$, its population target λ_N and the population quantity (3). Specifically, Assumptions 4 and 5 are conditions on the population parameter and its sample-based estimator. We also need a number of regularity conditions on the form and asymptotic behaviour of the population quantity $T_N(\lambda_N)$ as $N \to \infty$. In particular, Assumption 6 specifies a limiting smooth function for $T_N(\lambda)$ and Assumption 7 puts an important bound on the variation of necessary population quantities.

ASSUMPTION 4. *The population parameter of interest λ_N lies in a compact set C_λ.*

ASSUMPTION 5. *1. The estimator $\hat{\lambda}_N$ is $n^{*1/2}$-consistent for λ_N.*
2. The estimator $\hat{\lambda}_N$ can be linearized as $\hat{\lambda}_N = \lambda_N + N^{-1}\sum_{i\in\mathcal{S}} {\pi_i}^{-1} g(y_i) + o_p(n^{*-1/2})$, where $g(y_i)$ has finite fourth population moments.

ASSUMPTION 6. *1. The absolute value of $h(\cdot;\cdot)$ is bounded by a constant c_h.*
2. The population level function $T_N(\lambda)$ converges to a smooth function, $\lim_{N\to\infty} T_N(\lambda) = \mathcal{T}(\lambda)$, uniformly in C_λ defined in Assumption 4.
3. The limiting function $\mathcal{T}(\lambda)$ is uniformly continuous in λ in a neighbourhood of $\lambda_\infty = \lim_{N\to\infty}\lambda_N$, say C_λ. Further, $\mathcal{T}(\lambda)$ has finite first and second derivatives.

ASSUMPTION 7. *The population quantities satisfy*

$$\sup_{s\in C_s} N^\alpha \left|T_N(\lambda_N + N^{-\alpha}s) - T_N(\lambda_N) - \mathcal{T}(\lambda_N + N^{-\alpha}s) + \mathcal{T}(\lambda_N)\right| \to 0, \tag{5}$$

and

$$\sup_{s\in C_s} N^{-1}\sum_{i=1}^{N}\left|h(y_i;\lambda_N + N^{-\alpha}s) - h(y_i;\lambda_N)\right| = O(N^{-\alpha}), \tag{6}$$

where C_s is a large enough compact set in $\mathbb{R}^q$ and $\alpha \in (1/4, 1/2]$.

The reasonableness of the population requirements in Assumption 7 is somewhat difficult to evaluate as stated. Therefore, in Appendix A.1, a superpopulation model version of Assumption 7 is stated under which the y_i are generated through a probabilistic mechanism. Based on that assumption, a number of model results can be shown to hold with probability one. In particular, we can show that (5) and (6) hold almost surely under the superpopulation model as in Lemma 4. We here assume that the fixed population sequence from which we are sampling is such that these results hold, without the almost sure condition.

3·3. *Design-based results*

The key intermediate result we need in this section is stated in Lemma 1, which allows us to use the limiting smooth function $\mathcal{T}(\lambda)$ instead of nonsmooth population quantity $T_N(\lambda)$ in asymptotic expansions. We then establish the asymptotic normality of estimator (4) in Theorem 1. Proofs are deferred to Appendix A.2.

LEMMA 1. *Under Assumptions 1(A), 2, 5(1), and 6-7, conditional on $\mathcal{F}_N$,*

$$N^{\beta/2}\left\{\widehat{T}_N(\hat{\lambda}_N) - \widehat{T}_N(\lambda_N) - \mathcal{T}(\hat{\lambda}_N) + \mathcal{T}(\lambda_N)\right\} = o_p(1).$$

THEOREM 1. *Under Assumptions 1(A) and 2-7, the sample estimator $\widehat{T}_N(\hat{\lambda}_N)$ is design consistent for $T_N(\lambda_N)$ and asymptotically normally distributed, i.e. conditional on $\mathcal{F}_N$,*

$$\left[AV\{\widehat{T}_N(\hat{\lambda}_N)\}\right]^{-1/2}\left\{\widehat{T}_N(\hat{\lambda}_N) - T_N(\lambda_N)\right\} \to N(0,1),$$

in distribution, where $AV\{\widehat{T}_N(\hat{\lambda}_N)\} = \left\{1, \zeta(\lambda_N)^{\mathrm{T}}\right\} \mathrm{var}\left(\bar{z}_\pi \mid \mathcal{F}_N\right) \left\{1, \zeta(\lambda_N)^{\mathrm{T}}\right\}^{\mathrm{T}}$, $\zeta(\lambda)$ *denotes the first derivative of* $\mathcal{T}(\lambda)$ *and* $\bar{z}_\pi = N^{-1}\sum_{i\in\mathcal{S}} \pi_i{}^{-1}\left\{h(y_i;\lambda_N), g(y_i)^{\mathrm{T}}\right\}^{\mathrm{T}}$.

Generally speaking, for nondifferentiable survey estimators with estimated parameters, we can first replace the estimated parameter $\hat{\lambda}_N$ with an arbitrary constant λ in C_λ, then take expectation with respect to sampling design to obtain the population quantity $T_N(\lambda)$. The population quantity usually remains as a nondifferentiable function of λ, but we can often reasonably assume a differentiable limit for $T_N(\lambda)$ as in Assumption 6. The differentiable limit is then used in asymptotic expansion and variance expression.

In practice, many complex survey estimators cannot be written in the simple form of a survey weighted order-1 U-statistic, but are differentiable functions of estimators with expression (4). Properties of such estimators are straightforward extensions of Theorem 1, since the additional effect of the differentiable function is easily handled by traditional methods.

3·4. *Applications*

We discuss two examples of nondifferentiable estimators with estimated parameters that have appeared in the survey literature. As noted in Section 1, there exists extensive literature on estimating the population distribution function of a target variable when auxiliary information is present. To incorporate auxiliary information in estimating a distribution function, we generally estimate some model or population parameter(s) first and then substitute the estimated parameter(s) into an indicator function to construct a distribution function estimator. The sample distribution estimator is a nondifferentiable function of the estimated parameter(s), like the model-based estimator in Dunstan & Chambers (1986) or the ratio, difference and Rao–Kovar–Mantel estimators in Rao et al. (1990).

Rao et al. (1990) stated that one can ignore the variation due to estimating parameters in the last three estimators, but no rigorous proof was presented. We shall show that this is because the derivative $\zeta(\lambda_N)$ in Theorem 1 is either strictly zero or a smaller order term. Consider the difference estimator of Rao et al. (1990) defined as

$$\widehat{F}_{N,d}(t;\widehat{R}_N) = \frac{1}{N}\left[\sum_{\mathcal{S}}\frac{1}{\pi_i}I_{(y_i\le t)} + \left\{\sum_U I_{(\widehat{R}_N x_i\le t)} - \sum_{\mathcal{S}}\frac{1}{\pi_i}I_{(\widehat{R}_N x_i\le t)}\right\}\right],$$

where $\widehat{R}_N$ is a parameter estimated from sample data, and $I_{(\cdot)}$ is an indicator function, $I_{(c)} = 1$ if c is true and 0 otherwise. If we replace $\widehat{R}_N$ by an arbitrary constant λ to obtain $\widehat{F}_{N,d}(t;\lambda)$ and take expectation with respect to design, this is an unbiased estimator of $F_{N,d}(t) = N^{-1}\sum_U I_{(y_i\le t)}$, which does not depend on parameter λ. Therefore, the derivative of the limiting function with respect to λ is zero and, by the results in Theorem 1, the extra variance due to estimating population parameter R_N can be ignored in the asymptotic distribution. This resembles the asymptotic normality result (1.5) of Randles (1982). Similarly, the extra variance is negligible in the ratio estimator and Rao–Kovar–Mantel estimator in Rao et al. (1990), where it can be shown that $\zeta(\lambda_N)$ is of smaller order.

Another estimator that follows our framework is an estimated fraction below or above an estimated level, which is regularly seen in social surveys. A specific example is to estimate the fraction of households in poverty when the poverty line is draw at, say, 50% of the median income (Shao & Rao, 1993). This sample fraction with estimated median plugged in is a non-differentiable function of the estimated parameter, and we can apply the previous results to this situation, with $h(y_i; \lambda) = I_{(y_i \le \lambda)}$ and $\hat{\lambda}_N$ as sample-based estimator for the population median λ_N for the variable y_i. If we assume that the population observations y_i are independent and identically distributed random variables with distribution function $F_Y(\cdot)$, the limit of $T_N(\lambda)$ equals $F_Y(\lambda)$ almost surely, using the results in Appendix A.1. Theorem 1 can then be applied as long as we have a linearization or an asymptotic variance for the sample-based median estimator $\hat{\lambda}_N$, since the variance component due to estimation of the median remains significant in this case. The estimation of quantiles with the median as a special case will be discussed in Section 4.

4. NONDIFFERENTIABLE ESTIMATING EQUATIONS

4·1. *The Estimators*

We consider a population parameter ξ_N defined as

$$\xi_N = \inf\{\gamma : S_N(\gamma) \ge 0\}, \tag{7}$$

where $S_N(\gamma) = N^{-1} \sum_{i=1}^{N} \psi(y_i - \gamma)$ and $\psi(\cdot)$ is a univariate real function. The population parameter ξ_N is estimated by $\hat{\xi}_N$,

$$\hat{\xi}_N = \inf\{\gamma : \widehat{S}_N(\gamma) \ge 0\} \tag{8}$$

with $\widehat{S}_N(\gamma) = N^{-1} \sum_{i \in \mathcal{S}} \pi_i^{-1} \psi(y_i - \gamma)$.

4·2. *Assumptions*

In addition to the design assumptions in Section 2, we require regularity conditions on the sequence of finite populations. Assumption 8 assumes that the population quantity ξ_N lies in a closed interval on $\mathbb{R}$, and Assumption 9 specifies conditions on the monotonicity and smoothness of $S_N(\gamma)$ and its limit.

ASSUMPTION 8. *The population parameter ξ_N lies in a closed interval C_ξ on $\mathbb{R}$.*

ASSUMPTION 9. *The population estimating function $S_N(\cdot)$ and the function $\psi(\cdot)$ satisfy:*

1. *the function $\psi(\cdot)$ is bounded;*
2. *the population estimating function $S_N(\gamma)$ converges to $S(\gamma)$ uniformly on C_ξ as $N \to \infty$, and the equation $S(\gamma) = 0$ has a unique root in the interior of C_ξ;*
3. *the limiting function $S(\gamma)$ is strictly increasing and absolutely continuous with finite first derivative in C_ξ, and the derivative $S'(\gamma)$ is bounded away from 0 for γ in C_ξ;*
4. *the population quantities*

$$\sup_{\gamma \in C_\gamma} N^{\alpha} \left| S_N(\xi_N + N^{-\alpha}\gamma) - S_N(\xi_N) - S(\xi_N + N^{-\alpha}\gamma) + S(\xi_N) \right| \to 0, \tag{9}$$

and $\sup_{\gamma\in C_\gamma} N^{-1}\sum_{i=1}^N |\psi(y_i-\xi_N-N^{-\alpha}\gamma)-\psi(y_i-\xi_N)| = O(N^{-\alpha})$, *where* C_γ *is a large enough compact set in* $\mathbb{R}$ *and* $\alpha \in (1/4, 1/2]$.

Assumption 9(4) is somewhat restrictive and difficult to interpret. In Appendix A.1, we show that (9) holds with probability one under suitable assumptions on the probability mechanism generating the y_i and on the function ψ. Assumption 9(4) resembles Assumption 7 and their reasonableness will be discussed in the Appendix.

4·3. *Design-based results*

The main results for estimating equations are presented in this section, where Lemma 2 shows that $\widehat{S}_N(\gamma)$ converges in design probability to its population counterpart, Theorem 2 states the design consistency of sample estimator $\hat{\xi}_N$, and Theorem 3 states design-based asymptotic normality of the sample estimator. All proofs are given in Appendix A.2.

LEMMA 2. *Under Assumptions 1, 2(2) and 9, for any large enough closed interval* $C \in \mathbb{R}$,

$$\sup_{\gamma\in C}\left|\widehat{S}_N(\gamma)-S_N(\gamma)\right| = o_p(1).$$

THEOREM 2. *Under Assumptions 1, 2, 8 and 9, the estimator* $\hat{\xi}_N$ *is design consistent for the population quantity* ξ_N.

THEOREM 3. *Under Assumptions 1, 2-3, 8-9, for any sequence of estimators* $\hat{\xi}_N$ *that is* $n^{*1/2}$*-consistent for* ξ_N, *the estimator* $\hat{\xi}_N$ *can be linearized as* $\hat{\xi}_N = \xi_N - \{\widehat{S}_N(\xi_N) - S_N(\xi_N)\}/S'(\xi_N) + o_p(n^{*-1/2})$, *and is asymptotically normally distributed, i.e. conditional on* $\mathcal{F}_N$,

$$\left\{AV(\hat{\xi}_N)\right\}^{-1/2}(\hat{\xi}_N-\xi_N)\rightarrow N(0,1),$$

in distribution, where $AV(\hat{\xi}_N) = \mathrm{var}\{\widehat{S}_N(\xi_N)\mid\mathcal{F}_N\}/S'^2(\xi_N)$.

4·4. *Applications*

The first example is the sample quantile. For simplicity, consider the sample quantile estimator obtained by inverting the Hájek estimator of the cumulative distribution function. In this case, the estimating function for the αth quantile is $\psi(y_i-\gamma) = I_{(y_i-\gamma\le 0)} - \alpha$, with population estimating equation $S_{N,\alpha}(\gamma) = N^{-1}\sum_{i=1}^N I_{(y_i-\gamma\le 0)} - \alpha$. The sample α-quantile is defined as

$$\hat{\xi}_{N,\alpha} = \inf\{\gamma : \widehat{S}_{N,\alpha}(\gamma)\ge 0\} = \inf\left\{\gamma : \frac{1}{\widehat{N}}\sum_{i\in\mathcal{S}}\frac{1}{\pi_i}I_{(y_i\le\gamma)}\ge\alpha\right\},$$

where $\widehat{N} = \sum_{i\in\mathcal{S}}\pi_i^{-1}$. The function $\widehat{S}_{N,\alpha}(\gamma)$ is a Hájek estimator and it is asymptotically equivalent to a function with the same form of $\widehat{S}_N(\gamma)$. The limiting function of $S_{N,\alpha}(\gamma)$ is denoted as $S_\alpha(\gamma) = F(\gamma)-\alpha$, where $F(\gamma)$ can be taken to be the distribution function of y_i if we assume the y_i's are identically distributed and independent (or weakly dependent). Following the approach described earlier in this section, we directly obtain the asymptotic variance of $\hat{\xi}_N$ using design variance $\mathrm{var}\{\widehat{S}_{N,\alpha}(\xi_N)\mid\mathcal{F}_N\}$ and derivative $F'(\xi_N)$.

A second example is the Winsorized mean introduced by Huber (1964), where the estimating function $\psi(\cdot)$ is defined as $\psi(y_i-\gamma) = (y_i-\gamma)I_{(|y_i-\gamma|\le k)} - k\,I_{(y_i-\gamma<-k)} + k\,I_{(y_i-\gamma>k)}$

for some constant k. The population estimating function is

$$S_N(\gamma) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \gamma)I_{(|y_i-\gamma|\leq k)} + \frac{k}{N}\sum_{i=1}^{N}\left\{I_{(y_i-\gamma>k)} - I_{(y_i-\gamma<-k)}\right\},$$

and we assume $S_N(\gamma)$ converges to a limit function $S(\gamma)$ which is differentiable in a neighbourhood of ξ_N, where ξ_N is the population Winsorized mean as defined by (7). This population estimating function is nonincreasing, but we can use $-S_N(\gamma)$ and still define the parameter of interest as (7). Then we can define the sample estimating equation and estimator, and show its asymptotic properties as before.

Another possible application area for the theory presented in this section is quantile regression for survey data. There is growing interest in this topic in econometrics, see e.g. Koenker & Hallock (2001) and Koenker (2005). Currently there seem to be no references on how to use design information in quantile regression modelling. One could, in principle, incorporate survey weights in the equations that define the quantile model, and solve the estimating equations using linear programming. But the estimating equation itself is nondifferentiable, and traditional theory that requires differentiable estimating functions fails. Although we shall not do so here, our theoretical framework for nondifferentiable estimating equations could certainly be extended to this estimation setting.

5. VARIANCE ESTIMATION

5·1. *Analytic variance estimation*

To estimate the design variance of $\widehat{T}_N(\hat{\lambda}_N)$ in Section 3 or $\hat{\xi}_N$ in Section 4, we need to estimate the derivatives $\zeta(\lambda)$ or $S'(\gamma)$ of the limiting functions $\mathcal{T}(\lambda)$ or $S(\gamma)$, respectively. Natural sample-based estimators of the latter limiting functions are $\widehat{T}_N(\lambda)$ and $\widehat{S}_N(\gamma)$, but being nondifferentiable, cannot be used directly to obtain derivatives. We therefore work with a smoothed version of these estimators. This section describes a direct plug-in variance estimator with a kernel-based estimator, and the next section shows how to integrate the kernel-based derivative estimator into a replication-based variance estimator. The intrinsic idea of replacing the nonsmooth function with a convoluted smooth function in derivative estimation for design variance dates back at least to Deville (1999), but the choice of convoluting kernel was not made clear in Deville's article and no theoretical justifications have been provided in literature.

Define $K_q(\cdot)$ as a kernel function in $\mathbb{R}^q$, and convolute the nonsmooth function $h(y_i;\cdot)$ with $K_q(\cdot)$ using bandwidth b to obtain $h_i * K_q(\lambda) = \int\cdots\int h(y_i;x)K_q\{(\lambda-x)/b\}\,dx$, so that we can estimate $\mathcal{T}(\lambda)$ by

$$\frac{1}{N}\sum_{\mathcal{S}}\frac{1}{\pi_i}\int\cdots\int h(y_i;x)K_q\left(\frac{\lambda-x}{b}\right)dx. \quad (10)$$

Taking a derivative of (10) with respect to λ, we obtain the estimator

$$\hat{\zeta}(\lambda) = \frac{1}{Nb^q}\sum_{\mathcal{S}}\frac{1}{\pi_i}\int\cdots\int h(y_i;x)K_q'\left(\frac{\lambda-x}{b}\right)dx, \quad (11)$$

which estimates the population quantity

$$\zeta_N(\lambda) = \frac{1}{Nb^q}\sum_{U}\int\cdots\int h(y_i;x)K_q'\left(\frac{\lambda-x}{b}\right)dx, \quad (12)$$

for fixed λ.

We use $\|\cdot\|$ to denote the L_2 norm in $\mathbb{R}^q$ in assessing divergence, and we state a set of assumptions for obtaining the design consistency of $\hat{\zeta}(\hat{\lambda}_N)$ for $\zeta(\lambda_N)$.

ASSUMPTION 10. *The following conditions hold for kernel function $K_q(\cdot)$ and bandwidth b,*

1. *the kernel function $K_q(\cdot)$ is absolutely continuous with nonzero finite derivative $K_q'(\cdot)$ and $\int \cdots \int K_q(x)dx = 1$;*
2. *the bandwidth $b \to 0$ and $Nb^q \to \infty$, as $N \to \infty$;*
3. *there exists a constant c, such that $\left\|b^{-q}K_q'(x_1/b) - b^{-q}K_q'(x_2/b)\right\| \leq c\,\|x_1 - x_2\|$ for any x_1, x_2, and b arbitrarily small.*

ASSUMPTION 11. *The deviation $\|\zeta_N(\lambda) - \zeta(\lambda)\| \to 0$ uniformly for $\lambda \in C_\lambda$.*

Assumption 10 states conditions on the smoothness and tail behaviour of the kernel functions. Popular kernel functions including Epanechnikov, Gaussian, and triangle kernels all satisfy the required conditions. As in the previous sections, we can justify Assumption 11 by showing that under some stated model regularity conditions on the y_i, it holds with probability one for sufficiently large populations.

Given Assumptions 10, 11 and previously stated regularity conditions on the sampling design, we show the consistency of the kernel-based estimator (11) and the resulting variance estimator in Lemma 3 and Theorem 4, respectively. The proofs are provided in Appendix A.2.

LEMMA 3. *Under Assumptions 2, 4-5(1), 10(1-3) and 11, the estimator $\hat{\zeta}(\hat{\lambda}_N)$ is design consistent for $\zeta(\lambda_N)$.*

THEOREM 4. *Let $\widehat{V}_{\mathrm{HT}}(\bar{z}_\pi)$ be the Horvitz–Thompson variance estimator for $\bar{z}_\pi$ defined in Theorem 1. Under Assumptions 2-3, 4-5, 7(1), 10-11, the estimator*

$$\widehat{V}\{\widehat{T}_N(\hat{\lambda}_N)\} = \left\{1, \hat{\zeta}(\hat{\lambda}_N)^{\mathrm{T}}\right\} \widehat{V}_{\mathrm{HT}}(\bar{z}_\pi) \left\{1, \hat{\zeta}(\hat{\lambda}_N)^{\mathrm{T}}\right\}^{\mathrm{T}} \tag{13}$$

is design consistent for AV$\{\widehat{T}_N(\hat{\lambda}_N)\}$ as defined in Theorem 1.

The linearization term $g(y_i)$ that is part of $\bar{z}_\pi$ may also depend on unknown population parameters. Theorem 4 is written for the case in which these parameters are set at their population values, and in practice they would have to replaced by sample estimators. The above result will continue to hold by further linearization if $g(.)$ is differentiable. If that is not the case, then the theory from Section 3 would again need to be applied in order to approximate $\widehat{V}_{\mathrm{HT}}(\bar{z}_\pi)$. We do not explore this further here. The applications described in Section 3.4 all correspond to the case where $g(.)$ is differentiable.

Similarly to Theorem 4, one can obtain the design consistency of estimator $\widehat{V}(\hat{\xi}_N) = \widehat{V}_{\mathrm{HT}}\{\widehat{S}(\hat{\xi}_N)\}/\widehat{S}_N'^2(\hat{\xi}_N)$ for $AV(\hat{\xi}_N)$, where

$$\widehat{S}_N'(\gamma) = \frac{1}{Nb}\sum_{\mathcal{S}} \frac{1}{\pi_i} \int \psi(y_i - x) K'\left(\frac{\gamma - x}{b}\right) dx \tag{14}$$

is a kernel-based estimator of $S'(\gamma)$.

5·2. *Jackknife variance estimator*

We assume there already exists a design consistent jackknife variance estimator for simple linear estimators, then define jackknife replicates in our case and establish design consistency of the proposed variance estimator. This approach was also used by Fuller & Kim (2005) and Da Silva & Opsomer (2006). We apply a number of regularity assumptions on the replication method that were stated in the latter article, and do not repeat them here for the sake of brevity.

THEOREM 5. *Let $w_i = N^{-1}\pi_i^{-1}$ and let $\hat{\theta}$ be a linear estimator with $\hat{\theta} = \sum_{\mathcal{S}} w_i z_i$, where z_i has bounded $4+\delta$ population moments. Assume there is a jackknife replication procedure that generates L replicated estimates $\hat{\theta}^{(l)} = \sum_{\mathcal{S}} w_i^{(l)} z_i$, with $l = 1, 2, \cdots, L$. The replication variance estimator is defined as*

$$\widehat{V}_{\mathrm{JK}}(\hat{\theta}) = \sum_{l=1}^{L} c_l (\hat{\theta}^{(l)} - \hat{\theta})^2, \tag{15}$$

where $c_1, \ldots, c_L$ is a set of constants. Assumptions similar to (D1)-(D4) and (D6) in Da Silva & Opsomer (2006) are made.

1. *For explicit nondifferentiable survey estimators, we define the l-th jackknife replicate as*
$$\widehat{T}_N^{(l)}(\hat{\lambda}_N) = \sum\nolimits_{\mathcal{S}} w_i^{(l)} h(y_i; \hat{\lambda}_N) + \hat{\zeta}(\hat{\lambda}_N)^{\mathrm{T}}(\hat{\lambda}_N^{(l)} - \hat{\lambda}_N), \tag{16}$$
where the design variance of $\hat{\lambda}_N$ can be consistently estimated by $\sum_{l=1}^{L} c_l(\hat{\lambda}_N^{(l)} - \hat{\lambda}_N)^2$, and $\hat{\zeta}(\hat{\lambda}_N)$ is a kernel estimator as defined in (11). Then the jackknife variance estimator $\widehat{V}_{\mathrm{JK}}\{\widehat{T}_N(\hat{\lambda}_N)\}$ defined by (15) is design consistent for AV$\{\widehat{T}_N(\hat{\lambda}_N)\}$ in Theorem 1.
2. *For estimators defined by nondifferentiable estimating equations, we use the following jackknife replicate,*
$$\hat{\xi}_N^{(l)} = \frac{1}{\widehat{S}'_N(\hat{\xi}_N)} \sum_{i \in \mathcal{S}} w_i^{(l)} \psi(y_i - \hat{\xi}_N), \tag{17}$$
where $\widehat{S}'_N(\hat{\xi}_N)$ is defined in (14). Then the jackknife variance estimator $\widehat{V}_{\mathrm{JK}}(\hat{\xi}_N)$ is design consistent for AV$(\hat{\xi}_N)$ in Theorem 3.

The formal proof is omitted but follows by straightforward asymptotic bounding arguments from the assumptions. To see this for explicitly defined estimators, the replication variance estimator is readily interpreted by considering the composition of the replicate in (16). Ignoring the second term in (16), the resulting jackknife variance estimator consistently estimates the asymptotic variance of $\widehat{T}_N(\lambda_N)$, with population parameter λ_N substituted. The second term in (16) uses the combination of the kernel estimator and the replication method to estimate the effect of estimating the parameter. For implicitly defined estimators, the replicates $\sum_{i\in\mathcal{S}} w_i^{(l)}\psi(y_i - \hat{\xi}_N)$ allow us to consistently estimate $\mathrm{var}\{\widehat{S}_N(\xi_N) \mid \mathcal{F}_N\}$, and thus the whole jackknife estimator is consistent for the target asymptotic variance. Hence, in both situations, it is straightforward to modify existing replication variance estimation procedures to obtain variance estimates for the types of estimators proposed in Sections 3 and 4.

6. SIMULATION STUDY

For a fixed finite population of size $N = 2000$, we generate the variable y as independent realizations of a $\Gamma(2, 1)$ distribution. Then we repeatedly draw probability samples under a complex design with 3 strata. We create a stratification variable $z_i = y_i + {y_i}^{-1/2} + 5 + \epsilon_i$ with $\epsilon_i \sim N(0, 4)$, and use 7 and 9.5 as cutoff points on z for determining stratum membership for each element in finite population. Stratum 1 contains the elements where $z_i \leq 7$, and we draw a simple random sample without replacement with sample size $n_1 = n^*/4$ where n^* is desired total sample size. Stratum 2 contains the elements with $7 < z_i < 9.5$, where we partition the range of z_i into 150 intervals of equal length to form clusters, and select clusters using simple random sampling with the number of clusters equal to $n^*/(2N_c)$ with N_c denoting the average cluster size. Finally, we draw a Poisson sample with expected sample size $= n^*/4$ from stratum 3, with selection probability is proportional to z_i. We consider $n^* = 200$ and 400 and for each value of n^*, 2000 samples were drawn from the population using this design.

We examine estimators of the following target population quantities: population α-quantiles ξ_α with $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$, and proportions of points below $c \times \xi_{0.5}$, with $c = 0.25, 0.4$ and 0.6. Let $\mathcal{S}_h$ denote the sample drawn in stratum h. The sample quantiles are estimated by inverting the separate ratio estimator of the population cumulative distribution function, defined as

$$\widehat{F}_N(\gamma) = \sum_{h=1}^{3} \frac{N_h}{N} \frac{\sum_{\mathcal{S}_h} {\pi_i}^{-1} I_{(y_i \leq \gamma)}}{\sum_{\mathcal{S}_h} {\pi_i}^{-1}},$$

and the sample estimator of the proportion below $c \times \xi_{0.5}$ is defined as

$$\widehat{T}_{N,c} = \sum_{h=1}^{3} \frac{N_h}{N} \frac{\sum_{\mathcal{S}_h} {\pi_i}^{-1} I_{(y_i \leq c\hat{\xi}_{N,0.5})}}{\sum_{\mathcal{S}_h} {\pi_i}^{-1}}. \quad (18)$$

We compute the analytic variance estimator from Theorem 4 , the jackknife estimator from Theorem 5 and a naive jackknife variance estimator that calculates a sample quantile for each replicate. We incorporate finite population corrections into all three variance estimators. The estimators are compared in terms of relative bias, $(E\widehat{V} - V)/V$, where $\widehat{V}$ denotes the variance estimator and V denotes the true design variance simulated by Monte Carlo. The estimator (18) contains two nested non-differentiable functions, so the estimation of $AV(\widehat{T}_{N,c})$ requires estimation of the density of y at both $\hat{\xi}_{N,0.5}$ and $c\hat{\xi}_{N,0.5}$. These are obtained by kernel regression using a Gaussian kernel with bandwidth values $h = 0.1, 0.2, 0.4$.

Table 1. *Relative bias (%) of three variance estimators for sample quantiles under three different bandwidths h and two sample sizes n^*. Estimators are: $\widehat{V}_{\text{AN}}$, analytic variance estimator; $\widehat{V}_{\text{JK}}$, proposed jackknife variance estimator; $\widehat{V}_{\text{NVJK}}$, naive jackknife estimator.*

Variance estimator	h	$n^* = 200$ $\hat{\xi}_{N,.1}$	$\hat{\xi}_{N,.25}$	$\hat{\xi}_{N,.5}$	$\hat{\xi}_{N,.75}$	$\hat{\xi}_{N,.9}$	$n^* = 400$ $\hat{\xi}_{N,.1}$	$\hat{\xi}_{N,.25}$	$\hat{\xi}_{N,.5}$	$\hat{\xi}_{N,.75}$	$\hat{\xi}_{N,.9}$
$\widehat{V}_{\text{AN}}$	0.1	–7.7	3.9	1.3	–1.9	–5.7	–3.3	15.7	1.5	0.7	–1.6
$\widehat{V}_{\text{JK}}$		–6.6	5.1	2.8	0.5	–1.4	–2.8	16.1	2.0	1.5	0.1
$\widehat{V}_{\text{AN}}$	0.2	0.4	8.6	2.6	–7.2	–0.5	4.5	20.6	2.7	–5.5	0.3
$\widehat{V}_{\text{JK}}$		1.6	9.9	4.2	–4.8	3.8	5.0	21.2	3.3	–4.3	2.9
$\widehat{V}_{\text{AN}}$	0.4	26.3	27.8	2.8	–11.2	3.6	32.0	41.2	3.1	–9.3	7.6
$\widehat{V}_{\text{JK}}$		27.9	29.4	4.4	–8.8	7.7	32.6	41.8	3.6	–8.4	9.3
$\widehat{V}_{\text{NVJK}}$		186.4	204.3	156.1	221.3	252.9	206.2	228.9	181.5	208.0	231.9

Table 2. *Relative bias (%) of three variance estimators for fraction below an estimated quantity under three different bandwidths and two sample sizes n^*.*

Variance estimator	h	$n^* = 200$			$n^* = 400$		
		$\widehat{T}_{N,.25}$	$\widehat{T}_{N,.4}$	$\widehat{T}_{N,.6}$	$\widehat{T}_{N,.25}$	$\widehat{T}_{N,.4}$	$\widehat{T}_{N,.6}$
$\widehat{V}_{\mathrm{AN}}$	0.1	–1.3	0.9	3.7	–1.2	–2.2	2.7
$\widehat{V}_{\mathrm{JK}}$	0.1	–0.1	2.0	5.0	–0.7	–1.8	3.3
$\widehat{V}_{\mathrm{AN}}$	0.2	–3.1	0.1	0.4	–2.7	–2.6	0.2
$\widehat{V}_{\mathrm{JK}}$	0.2	–1.9	1.0	0.8	–2.2	–2.2	0.7
$\widehat{V}_{\mathrm{AN}}$	0.4	–4.3	–2.4	–5.3	–3.7	–4.6	–3.5
$\widehat{V}_{\mathrm{JK}}$	0.4	–3.0	–1.4	–4.2	–3.2	–4.2	–3.1
$\widehat{V}_{\mathrm{NVJK}}$		136.9	157.3	304.4	181.3	172.6	357.4

The results in Table 1 show that, as expected, the naive jackknife variance estimator is severely biased, indicating that the special structure of the estimator needs to be taken into account in variance estimation (Shao & Wu, 1989). The proposed analytic and jackknife variance estimators provide satisfactory results except in a number of cases under bandwidth $h = 0.4$. This bandwidth value appears to result in substantial oversmoothing of the data, at least for the smaller quantiles. At the suggestion of a referee, we also evaluated the proposed variance estimators with the Woodruff variance estimator, which is defined as the length of the Woodruff confidence interval divided by $2z_{1-\alpha/2}$. The bias of proposed analytic and jackknife variance estimators with $h = 0.1$ or 0.2 are comparable to those of the Woodruff procedure. However, we found that the proposed variance estimators appear to be more efficient for the majority of the cases, although this depended on the choice of bandwidth. Further numerical and theoretical comparisons would be necessary to make more conclusive statements. Table 2 shows the same results for the sample estimator defined in (18). The proposed variance estimators have relative bias less than 5% for all bandwidths, while the naive variance estimator is again severely biased.

Overall, the simulation results suggest that the two proposed variance estimators work reasonably well under appropriate bandwidth, although the performance of the variance estimators depends on bandwidth h to some extent. Bandwidth selection for these estimators is a current topic of research.

Acknowledgements

This work was supported by a cooperative agreement between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University. The authors thank the editor, associate editor and the referees for helpful comments and suggestions that greatly improved the paper.

Appendix

A.1. Model-based justifications for population assumptions

In Sections 3·2, 4·2 and 5·1, we assumed some regularity conditions on the sequence of fixed finite populations to obtain design-based results. Here, we provide sufficient conditions under a superpopulation model to assess the reasonableness of these population-level regularity conditions. In the model-based context, we assume that the finite population is an independent and identicallly distributed sample from a superpopulation model with cumulative distribution function $F(y)$, and we show that the stated assumptions on the sequence of finite populations hold with probability one. Proofs are in Appendix A.2.

Under a model version of Assumption 6 and additional regularity conditions on $h(\cdot;\cdot)$, we show that the statements in Assumption 7 hold with probability one under the superpopulation model.

ASSUMPTION 12. *Let Y be a random vector with absolutely continuous cumulative distribution function $F(y)$, and denote $\mathcal{T}(\lambda) \equiv E\{h(Y;\lambda)\}$. Further, $\mathcal{T}(\lambda)$ is a continuous function of λ, with finite first and second derivatives.*

ASSUMPTION 13. *There exists a finite integer m_1, such that $h(y;\lambda) = \sum_{j=1}^{m_1} h_j(y;\lambda)$, where $h_j(y;\lambda)$ is monotone in y for every λ.*

ASSUMPTION 14. *Let Y be a random vector with cumulative distribution function $F(y)$. The following moment conditions are satisfied,*

$$E\left|h(Y;\lambda+N^{-\alpha}s)-h(Y;\lambda)\right| = O(N^{-\alpha}) \ \ \textit{and} \ \ \mathrm{var}\left\{h(Y;\lambda+N^{-\alpha}s)-h(Y;\lambda)\right\} = O(N^{-\alpha}),$$

for $\lambda \in C_\lambda$, $s \in C_s$ and $0 < \alpha \leq 1/2$.

LEMMA 4. *Under Assumptions 6(1), 12-14, the population quantities*

$$N^{\alpha}\left|T_N(\lambda+N^{-\alpha}s)-T_N(\lambda)-\mathcal{T}(\lambda+N^{-\alpha}s)+\mathcal{T}(\lambda)\right| \to 0$$

almost surely, and

$$N^{-1}\sum_{i=1}^{N}\left|h(y_i;\lambda_N+N^{-\alpha}s)-h(y_i;\lambda_N)\right| = O(N^{-\alpha}) \tag{A1}$$

almost surely, uniformly for $s \in C_s$, a compact set in $\mathbb{R}^q$, and for $\lambda \in C_\lambda$, where C_λ is defined in Assumption 4 and $\alpha \in (1/4, 1/2]$.

Next, we justify Assumption 9(4) under a probabilistic model. For the sake of brevity, we only explicitly state the result for the first statement. We assume the population characteristics y_i are independent and identically distributed realizations of a random variable Y, and place additional restrictions on ψ to obtain almost sure convergence.

ASSUMPTION 15. *Let Y be a random variable with absolutely continuous cumulative distribution function, and denote $S(\gamma) \equiv E\{\psi(Y-\gamma)\}$. The estimating function $S(\gamma)$ is strictly increasing with finite first derivative.*

ASSUMPTION 16. *The function $\psi(\cdot)$ is bounded and has a finite number of monotonicity changes.*

LEMMA 5. *Under Assumptions 15 and 16, the population quantity*

$$n^{*1/2}\left\{S_N(\xi_N+n^{*-1/2}s)-S(\xi_N+n^{*-1/2}s)+S(\xi_N)\right\} \to 0$$

almost surely, uniformly for $s \in C_s$, a closed interval in $\mathbb{R}$.

Finally, we address Assumption 11. In addition to the previous assumption 12, we require additional regularity conditions on the function h and the kernel K_q.

ASSUMPTION 17. *There exists a finite integer m_2, such that $K'_q(\lambda) = \sum_{j=1}^{m_2} K'_{q,j}(\lambda)$, where $K'_{q,j}(\lambda)$ has no change of sign for any j, i.e., $K'_{q,j}(\lambda)$ is either nonnegative or nonpositive for any λ.*

LEMMA 6. *Under Assumptions 10(1,2), 12, 13 and 17, the deviation $\sup_{\lambda\in C_\lambda}\|\zeta_N(\lambda)-\zeta(\lambda)\| \to 0$, almost surely.*

A.2. Technical details

Proof of Lemma 1. Define $Q_n(s) = n^{*1/2}\{\widehat{T}_N(\lambda_N + n^{*-1/2}s) - \widehat{T}_N(\lambda_N) - \mathcal{T}(\lambda_N + n^{*-1/2}s) + \mathcal{T}(\lambda_N)\}$. We need to show that $\sup_{s\in C}|Q_n(s)| \to 0$ weakly, where C is a compact region in $\mathbb{R}^q$. We have

$$\begin{aligned}|Q_n(s)| \le n^{*1/2}&\left|\widehat{T}_N(\lambda_N + n^{*-1/2}s) - \widehat{T}_N(\lambda_N) - T_N(\lambda_N + n^{*-1/2}s) + T_N(\lambda_N)\right| \\ &+ n^{*1/2}\left|T_N(\lambda_N + n^{*-1/2}s) - T_N(\lambda_N) - \mathcal{T}(\lambda_N + n^{*-1/2}s) + \mathcal{T}(\lambda_N)\right|, \qquad \text{(A2)}\end{aligned}$$

where the supremum of the second term converges to zero by Assumption 7. Now we need to show the supremum of the first term converges to zero in probability. For any $1-\beta < \xi < \beta/(2p)$, where β is defined in Assumption 1, partition the compact region C into $C = C_1 \cup C_2 \cup \cdots \cup C_{N^{\xi p}}, C_j \cap C_{j'} = \emptyset$, for any $j \ne j'$, where $\mathrm{Diam}(C_j) = O(N^{-\xi})$, for any $j = 1, 2, \ldots, N^{\xi p}$. For any set of $s_j \in C_j, j = 1, 2, \ldots, N^{\xi p}$, define

$$X_{1j} = \left|\frac{n^{*1/2}}{N}\sum_{i=1}^{N}\left\{\frac{I_{(i\in s)}}{\pi_i} - 1\right\}\left\{h(y_i;\lambda_N + n^{*-1/2}s_j) - h(y_i;\lambda_N)\right\}\right|$$

and

$$X_{2j} = \sup_{s\in C_j}\left|\frac{n^{*1/2}}{N}\sum_{i=1}^{N}\left\{\frac{I_{(i\in\mathcal{S})}}{\pi_i} - 1\right\}\left\{h(y_i;\lambda_N + n^{*-1/2}s) - h(y_i;\lambda_N + n^{*-1/2}s_j)\right\}\right|.$$

Since $\sup_{s\in C}|Q_n(s)| \le \max_j|X_{1j}| + \max_j|X_{2j}|$, it suffices to show that both $\max_j|X_{1j}|$ and $\max_j|X_{2j}|$ converge to zero in probability. We have $\max_j|X_{1j}| \to 0$ weakly, since

$$\begin{aligned}\mathrm{pr}(\max_j|X_{1j}| > \epsilon) &\le \sum_j \mathrm{pr}(|X_{1j}| > \epsilon) \\ &\le \frac{4c_1c_h}{\epsilon^2}N^{\xi p}\max_j\frac{1}{N}\sum_{i=1}^{N}\left|h(y_i;\lambda_N + n^{*-1/2}s_j) - h(y_i;\lambda_N)\right| \\ &= O(N^{\xi p - \beta/2}),\end{aligned}$$

where c_1 is a positive constant, c_h is defined in Assumption 6(1) and the last term goes to zero as $\xi < \beta/(2p)$. The proof of $\max_j|X_{2j}| \to 0$ weakly follows from Assumptions 2(1) and 7. □

Proof of Theorem 1. We have the following identity

$$\begin{aligned}\left\{\widehat{T}_N(\hat\lambda_N) - T_N(\lambda_N)\right\} = &\left\{\widehat{T}_N(\lambda_N) - T_N(\lambda_N)\right\} + \left\{\mathcal{T}(\hat\lambda_N) - \mathcal{T}(\lambda_N)\right\} \\ &+ \left\{\widehat{T}_N(\hat\lambda_N) - \widehat{T}_N(\lambda_N) - \mathcal{T}(\hat\lambda_N) + \mathcal{T}(\lambda_N)\right\}\end{aligned}$$

where the last term is stochastically small by Lemma 1. The second term can now be linearized since the limiting function is differentiable. Design consistency of the estimator follows immediately from Assumptions 5(1) and 6 and design assumptions 2.1–2.2. Using Assumption 3 and the fact that $h(y_i;\lambda_N)$ and $g(y_i)$ have finite fourth population moments, we also obtain the normality of the sample estimator. □

Proof of Lemma 2. Partition C into N^ν equal sub-intervals $C = \cup_{k=1}^{N^\nu}C_k$, with $1/2 < \nu < \beta$ and select an arbitrary point $\gamma_k \in C_k, k = 1, 2, \ldots, N^\nu$. Then,

$$\sup_{\gamma\in C}\left|\widehat{S}_N(\gamma) - S_N(\gamma)\right| \le \max_k\left|\widehat{S}_N(\gamma_k) - S_N(\gamma_k)\right| + \max_k\sup_{\gamma\in C_k}\left|\left\{\widehat{S}_N(\gamma) - S_N(\gamma)\right\} - \left\{\widehat{S}_N(\gamma_k) - S_N(\gamma_k)\right\}\right|,$$

and it suffices to show the two terms on the RHS are both stochastically small. By Assumption 2 for some constant c_2, $\mathrm{var}\{\widehat{S}_N(\gamma) \mid \mathcal{F}_N\}$ is bounded by c_2/n^*, and hence $\mathrm{pr}\{\max_k|\widehat{S}_N(\gamma_k) - S_N(\gamma_k)| \ge \epsilon\} \to 0$.

The result then follows from

$$\max_k \sup_{\gamma \in C_k} \left| \left\{ \widehat{S}(\gamma) - S_N(\gamma) \right\} - \left\{ \widehat{S}(\gamma_k) - S_N(\gamma_k) \right\} \right|$$
$$\leq c_4 \frac{N}{n^*} \max_k \sup_{\gamma \in C_k} \frac{1}{N} \sum_{i=1}^{N} |\psi(y_i - \gamma) - \psi(y_i - \gamma_k)| = O(N^{1/2-\beta}),$$

where c_4 is a positive constant and the last equation follows from Assumption 9(4). □

Proof of Lemma 3. The triangle inequality implies that

$$\|\hat{\zeta}(\hat{\lambda}_N) - \zeta(\lambda_N)\| \leq \|\hat{\zeta}(\hat{\lambda}_N) - \hat{\zeta}(\lambda_N)\| + \|\hat{\zeta}(\lambda_N) - \zeta_N(\lambda_N)\| + \|\zeta_N(\lambda_N) - \zeta(\lambda_N)\|,$$

where the first term converges to zero weakly by the uniform continuity of $\hat{\zeta}(\cdot)$ is a neighbourhood around λ_N, the second term converges to zero in probability because its asymptotic variance goes to zero by Assumption 2(2) and the third term goes to zero by Assumption 11. □

Proof of Theorem 2. The proof follows from standard arguments on existence and consistency of M-estimators (Serfling, 1980). □

Proof of Theorem 3. Under Assumptions 1, 2, 8, 9 and assuming that the sample estimator $\hat{\xi}_N$ is $n^{*1/2}$-consistent for ξ_N, we obtain the asymptotic expansion

$$\widehat{S}_N(\hat{\xi}_N) - S_N(\xi_N) = \left\{ \widehat{S}_N(\xi_N) - S_N(\xi_N) \right\} + S'(\xi_N)(\hat{\xi}_N - \xi_N) + o_p(n^{*-1/2}). \tag{A3}$$

The smoothness condition of $S(\gamma)$ implies that $S_N(\xi_N) = O\left(N^{-1}\right)$ and $\widehat{S}_N(\hat{\xi}_N) = O_p\left(n^{*-1}\right)$, so that $\{\widehat{S}_N(\hat{\xi}_N) - S_N(\xi_N)\} = o_p(n^{*-1/2})$. Dividing by $S'(\xi_N)$ on both sides of (A3), we obtain linearization $\hat{\xi}_N = \xi_N - \{\widehat{S}_N(\xi_N) - S_N(\xi_N)\}/S'(\xi_N) + o_p(n^{*-1/2})$. Asymptotic normality of $\hat{\xi}_N$ follows directly. □

Proof of Lemma 4. Letting $X_i = h(y_i; \lambda + N^{-\alpha}s) - h(y_i; \lambda) - \mathcal{T}(\lambda + N^{-\alpha}s) + \mathcal{T}(\lambda)$, we need to show that $N^{\alpha}\left|N^{-1}\sum_{i=1}^{N} X_i\right| \to 0$ almost surely, uniformly for $\lambda \in C_\lambda$ and $s \in C_s$. Here, $E(X_i) = 0$ and $E\{h(y_i; \lambda + N^{-\alpha}s) - h(y_i; \lambda)\} = O(N^{-\alpha})$. Without loss of generality, assume $|h(y_i; \lambda + N^{-\alpha}s) - h(y_i; \lambda)| \leq 1$, and $E\{h(y_i; \lambda + N^{-\alpha}s) - h(y_i; \lambda)\}^2 \leq N^{-\alpha}$. Define the graph of $g_{\lambda,s}(y) \equiv h(y_i; \lambda + N^{-\alpha}s) - h(y_i; \lambda)$ as

$$\mathrm{gr}(g_{\lambda,s}) = \{(y,t) \mid 0 \leq t \leq g_{\lambda,s}(y)\} \cup \{(y,t) \mid g_{\lambda,s}(y) \leq t \leq 0\}.$$

Assumption 13 implies that the set of functions $\{(\lambda, s) \in \mathbb{R}^q \times C_s : h(y; \lambda + N^{-\alpha}s) - h(y; \lambda)\}$ can be written as the summation of a finite number of monotone function classes. Lemmas 9.9 and 9.11 of Kosorok (2008) imply that $\{(\lambda, s) \in C_\lambda \times C_s : h(y; \lambda + N^{-\alpha}s) - h(y; \lambda)\}$ is a VC class and thus has polynomial discrimination. Everything is set up for Theorem II.37 of Pollard (1984). Letting $\alpha_N = 1$ and $\delta_N^2 = N^{-\alpha}$, we obtain $\sup_{(\lambda,s)\in C_\lambda \times C_s} N^{\alpha} |T_N(\lambda + N^{-\alpha}s) - T_N(\lambda) - \mathcal{T}(\lambda + N^{-\alpha}s) + \mathcal{T}(\lambda)| \to 0$ almost surely. The almost sure convergence of (A1) is established similarly but omitted here for brevity. □

Proof of Lemma 5. Similar to the proof of Lemma 4. □

Proof of Lemma 6. Since $\|\zeta_N(\lambda) - \zeta(\lambda)\| \leq \|\zeta_N(\lambda) - E\zeta_N(\lambda)\| + \|E\zeta_N(\lambda) - \zeta(\lambda)\|$ and $\sup_{\lambda \in C_\lambda} \|E\zeta_N(\lambda) - \zeta(\lambda)\| \to 0$, we only need to show that $\sup_{\lambda \in C_\lambda} \|\zeta_N(\lambda) - E\zeta_N(\lambda)\| \to 0$ almost surely. Letting $g_\lambda(y) = \int \cdots \int h(y; x) K_q' \{(\lambda - x)/b\} \, dx$, it is equivalent to show that $\sup_{\lambda \in C_\lambda} \|N^{-1}\sum_{i=1}^{N} g_\lambda(y_i) - Eg_\lambda(y_i)\| \to 0$ almost surely. Assumption 17 implies that

$$g_\lambda(y) = \sum_{j_1}\sum_{j_2} \int \cdots \int h_{j_1}(y; x) K'_{q,j_2}\left(\frac{\lambda - x}{b}\right) dx.$$

By Assumptions 13 and 17, the integral of each $h_{j_1}(y;x)K'_{q,j_2}\{(\lambda - x)/b\}$ is monotone for any λ. It is then possible to show that the graphs of $g_\lambda(y)$ have polynomial discrimination by Lemma II.15 of Pollard (1984). The remainder of the proof is as for Lemma 4. □

Proof of Theorem 4. The result follows from Assumption 3, Lemma 3 and consistency of $\widehat{V}_{\mathrm{HT}}(\bar{z}_\pi)$. □

REFERENCES

BERGER, Y. G. & SKINNER, C. J. (2003). Variance estimation for a low income proportion. *Journal of the Royal Statistical Society, Series C: Applied Statistics* **52**, 457–468.

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* **51**, 279–292.

BINDER, D. A. & KOVACEVIC, M. S. (1995). Estimating some measures of income inequality from survey data: an application of the estimating equations approach. *Survey Methodology* **21**, 137–145.

BREIDT, F. J. & OPSOMER, J. D. (2008). Endogenous post-stratification in surveys: classifying with a sample-fitted model. *Annals of Statistics* **36**, 403–427.

CHAMBERS, R. L., DORFMAN, A. H. & HALL, P. (1992). Properties of estimators of the finite population distribution function. *Biometrika* **79**, 577–582.

DA SILVA, D. & OPSOMER, J. D. (2006). A kernel smoothing method to adjust for unit nonresponse in sample surveys. *the Canadian Journal of Statistics* **34**, 563–579.

DEVILLE, J.-C. (1999). Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Survey Methodology* **25**, 193–203.

DUNSTAN, R. & CHAMBERS, R. L. (1986). Model-based confidence intervals in multipurpose surveys. *Applied Statistics* **35**, 276–280.

EUROSTAT (2000). Low-wage employees in EU countries. *Statistics in Focus: Population and Social Conditions.* .

FRANCISCO, C. A. & FULLER, W. A. (1991). Quantile estimation with a complex survey design. *The Annals of Statistics* **19**, 454–469.

FULLER, W. A. (2009). *Sampling Statistics*. John Wiley & Sons.

FULLER, W. A. & KIM, J. K. (2005). Hot deck imputation for the response model. *Survey Methodology* **31**, 139–149.

GODAMBE, V. P. & THOMPSON, M. E. (2009). Estimating functions and survey sampling. In *Handbook of Statistics Volume 29B: Sample Surveys: Inference and Analysis,*, C. Rao & D. P. (editors), eds. Elsevier/North-Holland, pp. 83–101.

HUBER, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* **35**, 73–101.

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

KOENKER, R. (2005). *Quantile Regression*. Cambridge University Press.

KOENKER, R. & HALLOCK, K. F. (2001). Quantile regression. *The Journal of Economic Perspectives* **15**, 143–156.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag Inc.

KUK, A. Y. C. & MAK, T. K. (1989). Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society, Series B: Methodological* **51**, 261–269.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag Inc.

PRESTON, I. (1996). Sampling distributions of relative poverty statistics. *Applied Statistics* **45**, 399.

RANDLES, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics* **10**, 462–474.

RAO, J., KOVAR, J. G. & MANTEL, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika* **77**, 365–375.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

SHAO, J. & RAO, J. N. K. (1993). Standard errors for low income proportions estimated from stratified multistage samples. *Sankhya B* **55**, 393–414.

SHAO, J. & WU, C. F. J. (1989). A general theory for jackknife variance estimation. *The Annals of Statistics* **17**, 1176–1197.

SITTER, R. R. & WU, C. (2001). A note on Woodruff confidence intervals for quantiles. *Statistics & Probability Letters* **52**, 353–358.

WANG, J. C. & OPSOMER, J. D. (2010). Estimating the distance distribution of subpopulations for a large scale complex survey. Under revision.

WANG, S. & DORFMAN, A. H. (1996). A new estimator for the finite population distribution function. *Biometrika* **83**, 639–652.

WU, C. & SITTER, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association* **96**, 185–193.