On the asymptotic normality and variance estimation of nondifferentiable survey estimators

Jianqiang.C Wang

Visiting Assistant Professor, Department of Statistics

Colorado State University

Jean Opsomer

Professor, Department of Statistics

Colorado State University

July 1, 2009

Abstract

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 the second is defined only as the solution to estimating equations. We propose both analytical and replication-based design consistent variance estimators that use kernel regression to estimate limits of nondifferentiable functions.

Keywords: estimating equations, kernel regression, nondifferentiable estimators, replication variance estimation.

1 Introduction

A number of common survey estimators, including estimators of population distribution functions and population quantiles, involve nondifferentiable functions of estimated quantities. Because of this nondifferentiability, these estimators do not follow the standard paradigm for obtaining 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, often leading to a "reinvention of the wheel" scenario, or of glossing over the nondifferentiability aspects of their estimator. The former option is often technically challenging and requires assumptions on the population that might be difficult to validate with sample data.

In this article, we will consider two types of nondifferentiable estimators. The first is explicitly defined estimators, in which one or several estimated quantities are "embedded" inside a non-differentiable function. The second is estimators that are defined as the solution to estimating equations, with the equation containing non-differentiable components. Examples of the first type of estimators include estimators of a population distribution function using auxiliary information (Dunstan and Chambers 1986, Rao, Kovar, and Mantel 1990, Chambers, Dorfman, and Hall 1992, Wang and Dorfman 1996), estimators of a population fraction above or below an estimated quantity (Shao and Rao 1993, Binder and Kovacevic 1995, Preston 1996, Eurostat 2000, Berger and Skinner 2003), the endogenous post-stratification estimator (Breidt and Opsomer 2008) and an estimator for the population distribution of distances to a subpopulation center (Wang and Opsomer 2008).

Many of these authors obtain the 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 describes variance estimation for complex statistics using influence functions. He also introduced kernel smoothing in variance estimation. However, no formal proof is provided and there is little influential theoretical work establishing the asymptotic properties of this class of estimators under a complex survey design.

The second type of nondifferentiable estimators we will consider involves design-weighted estimating equations. Godambe and Thompson (2009) gave a general treatment of estimating equations in survey sampling, and show how many quantities of interest can be defined through estimating equations, including means, quantiles and generalized linear model parameters (see also Binder 1983, Wu and Sitter 2001). Chapter 1.3.4 of Fuller (2007) derived the properties of estimators defined by estimating equations in complex surveys when the estimating function satisfies some differentiability condition. To our knowledge, a full theoretical treatment of survey estimators with non-differentiable estimating equations is not available in the literature.

In the current article, we provide a unified approach for handling nondifferentiable survey estimators. 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-statistics and the data are independent and identically distributed. We extend those results to the survey setting, in which the randomness comes from the sampling design and the population remains fixed. For both types of estimators 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 limits of the nondifferentiable functions.

The ultimate goal of the article is to make available a set of tools that allows statisticians to avoid having to "reinvent the wheel" when faced with a new nondifferentiable survey estimator. Instead, they only need to check whether their design and estimation set-up falls within the assumptions provided here, after which the properties and variance estimation will follow from the results in this paper. To this end, we have aimed our treatment to be as general as possible without sacrificing simplicity and interpretability of the assumptions.

A brief overview of the article follows. Section 2 provides general design assumptions addressing design consistency and asymptotic normality of a general Horvitz-Thompson estimator. Section 3 presents further assumptions and the theoretical results for estimators that can be written as nondifferentiable functions of estimated quantities. Section 4 treats the case of estimators defined as solutions to nondifferetiable estimating equations. Variance estimation for both cases is addressed in Section 5, and we provide both analytic and a replication-based variance estimator versions. Section 6 describes a simulation study evaluating the practical properties of the variance estimators.

2 General design assumptions

In this section, we state assumptions on the sampling design and estimators, which address 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 and Fuller (1982) in which the properties of estimators are established under a fixed sequence of populations and a corresponding sequence of random samples. Suppose therefore that we have an increasing sequence of finite populations $\{U_N\}$ of size N, with $N \to \infty$. Associated with the i-th population element is a p-dimensional vector of observations

$$\boldsymbol{y}_i = (y_{i,1}, ..., y_{i,p}), \tag{1}$$

and let $\mathcal{F}_N$ be the power set of N-th finite population $\{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_N\}$.

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 $\pi_i = \Pr(i \in \mathcal{S})$ represent the inclusion probability of the ith population element. We write $\bar{\boldsymbol{y}}_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i$ for the population mean of variable $\boldsymbol{y}_i$ and $\bar{\boldsymbol{y}}_\pi = \frac{1}{N}\sum_{\mathcal{S}} \frac{\boldsymbol{y}_i}{\pi_i}$ for its Horvitz-Thompson estimator.

We state three assumptions. Assumption 2.1 sets limits on the rate of the sample size, Assumption 2.2 ensures the design consistency and Assumption 2.3 guarantees asymptotic normality of our estimator under a general design.

Assumption 2.1. *The expected sample size* $n^* = \mathrm{E}(n|\mathcal{F}_N) = O(N^\beta)$*, where we assume either*

1. $\frac{2p}{2p+1} < \beta \leq 1$*, with* p *the dimension of study variable* $\boldsymbol{y}$*, or*
2. $\frac{1}{2} < \beta \leq 1$*,*

Assumption 2.2. *The following conditions hold for inclusion probabilities* π_i *and design variance of Horvitz-Thompson estimator of the mean,*

1. $K_L \leq \frac{N}{n^*}\pi_i \leq K_U$ *for all* i*, where* K_L *and* K_U *are positive constants.*
2. *For any vector* $\boldsymbol{z}$ *with finite* $2+\delta$ *population monents, or equivalently,*

$$\frac{1}{N}\sum_{i=1}^{N} |\boldsymbol{z}|^{2+\delta} < \infty,$$

we assume

$$\mathrm{Var}(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_N) \le c_1 \mathrm{Var}_{SRS}(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_N), \tag{2}$$

for some constant c_1*, where* $\mathrm{Var}_{SRS}(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_N)$ *is the design variance-covariance matrix of* $\bar{\boldsymbol{z}}_\pi$ *under simple random sampling of size* $n^* = \mathrm{E}(n|\mathcal{F}_N)$*.*

It is readily shown that under Assumption 2.2(2), $\frac{n}{n^*} \xrightarrow{p} 1$ by bounding its design variance.

Assumption 2.3. *For any* $\boldsymbol{z}$ *with finite fourth population moment,*

$$[\mathrm{Var}(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_\mathrm{N})]^{-1/2} (\bar{\boldsymbol{z}}_\pi - \bar{\boldsymbol{z}}_N)|\mathcal{F}_N \xrightarrow{d} N(\mathbf{0}, \mathrm{I}_{p\times p}), \tag{3}$$

and

$$[\mathrm{Var}(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_\mathrm{N})]^{-1} \widehat{V}_{HT}\{\bar{\boldsymbol{z}}_\pi\} - \mathrm{I}_{p\times p} = O_p(n^{*-1/2}), \tag{4}$$

where $\mathrm{I}_{p\times p}$ *is the* $p\times p$ *identity matrix, the design variance-covariance matrix of* $\bar{\boldsymbol{z}}_\pi$*, denoted by* $\mathrm{Var}(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_\mathrm{N})$*, is positive definite, and* $\widehat{V}_{HT}\{\bar{\boldsymbol{z}}_\pi\}$ *is the Horvitz-Thompson estimator of* $\mathrm{Var}(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_\mathrm{N})$*.*

3 Explicitly defined nondifferentiable survey estimators

3.1 The Estimators

We assume that the population quantity takes the form of an order 1 U-statistic and that the sample estimator is a Horvitz-Thompson estimator of the population quantity but with estimated parameter(s). The sample estimator can be written in the form

$$\widehat{T}(\hat{\boldsymbol{\lambda}}) = \tfrac{1}{N}\sum_{i\in\mathcal{S}} \tfrac{1}{\pi_i} h(\boldsymbol{y}_i; \hat{\boldsymbol{\lambda}}), \tag{5}$$

where $\hat{\boldsymbol{\lambda}}$ is a sample based estimator of some q-dimensional population quantity $\boldsymbol{\lambda}_N$ and $h(\boldsymbol{y};\boldsymbol{\lambda}) : \Re^p \times \Re^q \to \Re$ is not necessarily a differentiable function of $\boldsymbol{\lambda}$. The two integers p and q represent the dimension of the target variable $\boldsymbol{y}_i$ and estimated parameter $\hat{\boldsymbol{\lambda}}$ respectively, and need not be the same. The estimator (5) targets the population quantity

$$T_N(\boldsymbol{\lambda}_N) = \frac{1}{N}\sum_{i=1}^{N} h(\boldsymbol{y}_i; \boldsymbol{\lambda}_N). \tag{6}$$

In what follows, we will use an arbitrary constant $\boldsymbol{\lambda}$ in $T_N(\cdot)$ and $\widehat{T}(\cdot)$ to emphasize that both quantities are functions of population quantity $\boldsymbol{\lambda}_N$ or sample estimator $\hat{\boldsymbol{\lambda}}$. The case when $h(\boldsymbol{y}; \boldsymbol{\lambda})$ is a smooth function of $\boldsymbol{\lambda}$ is easy to deal with, because we can apply Taylor linearization and obtain the ignorability of the remaining terms in the expansion using traditional arguments. But if $h(\boldsymbol{y}; \boldsymbol{\lambda})$ is a nondifferentiable function of $\boldsymbol{\lambda}$, we can not 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. But in the survey context, if $h(\boldsymbol{y}; \boldsymbol{\lambda})$ is a nonsmooth function of $\boldsymbol{\lambda}$, the expectation of $\widehat{T}(\boldsymbol{\lambda})$ under the design, namely $T_N(\boldsymbol{\lambda})$, remains as a nonsmooth function of $\boldsymbol{\lambda}$, so we need to modify the approach of Randles (1982) to extend the results to survey context.

3.2 Assumptions

We provide a set of conditions that need to be satisfied by the parameter estimator $\hat{\boldsymbol{\lambda}}$, its population target $\boldsymbol{\lambda}_N$ and the population quantity (6). Together with design Assumptions 2.1–2.3, these will provide sufficient conditions to obtain the asymptotic properties of nondifferentiable survey estimator (5).

Assumptions 3.1 and 3.2 are conditions on the population parameter and its sample-based estimator. We also need a number of specific regularity conditions on the form and asymptotic behavior of the population quantity $T_N(\boldsymbol{\lambda}_N)$ as $N \to \infty$. In particular, Assumption 3.3 specifies a limiting smooth function of $T_N(\boldsymbol{\lambda})$, and Assumption 3.4 puts an important bound on the variation of a necessary population quantity as a function of its argument $\boldsymbol{\lambda}$.

Assumption 3.1. *The population parameter of interest $\boldsymbol{\lambda}_N \in C_{\boldsymbol{\lambda}}$ where $C_{\boldsymbol{\lambda}}$ is compact on $\Re^q$.*

Assumption 3.2. *We need the following conditions for $\hat{\boldsymbol{\lambda}}$ as an estimator of $\boldsymbol{\lambda}_N$,*

1. *$\hat{\boldsymbol{\lambda}}$ is $\sqrt{n^*}$-consistent for $\boldsymbol{\lambda}_N$.*

2. $\hat{\boldsymbol{\lambda}}$ *has the following linearization*

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}_N + \tfrac{1}{N} \sum_{i \in \mathcal{S}} \tfrac{1}{\pi_i} \boldsymbol{g}(\boldsymbol{y}_i) + o_p(n^{*-1/2}),$$

where $\boldsymbol{g}(\boldsymbol{y}_i)$ *has finite fourth population moments.*

Assumption 3.3. *1. The absolute value of* $h(\cdot;\cdot)$ *is bounded by a constant* c_h*.*

2. The population level function $T_N(\boldsymbol{\lambda})$ *converges to a limiting smooth function,*

$$\lim_{N \to \infty} T_N(\boldsymbol{\lambda}) = \mathcal{T}(\boldsymbol{\lambda}),$$

uniformly in a neighborhood of $\boldsymbol{\lambda}_\infty$ *where* $\boldsymbol{\lambda}_\infty = \lim_{N \to \infty} \boldsymbol{\lambda}_N$*.*

3. The limiting function $\mathcal{T}(\boldsymbol{\lambda})$ *is uniformly continuous for* $\boldsymbol{\lambda}$ *in a neighborhood of* $\boldsymbol{\lambda}_\infty$*, say* $C_{\boldsymbol{\lambda}}$*. Further,* $\mathcal{T}(\boldsymbol{\lambda})$ *has finite first and second derivatives with respect to* $\boldsymbol{\lambda}$*.*

Assumption 3.4. *The population quantity*

$$\sup_{\boldsymbol{s} \in C_{\boldsymbol{s}}} N^{\alpha} \left| T_N(\boldsymbol{\lambda}_N + N^{-\alpha}\boldsymbol{s}) - T_N(\boldsymbol{\lambda}_N) - \mathcal{T}(\boldsymbol{\lambda}_N + N^{-\alpha}\boldsymbol{s}) + \mathcal{T}(\boldsymbol{\lambda}_N) \right| \to 0 \tag{7}$$

where $C_{\boldsymbol{s}}$ *is a large enough compact set in* $\Re^q$ *and* $\alpha \in \left(\frac{1}{4}, \frac{1}{2}\right]$*.*

The reasonableness of the population requirements in Assumptions 3.3 and in particular 3.4 is somewhat difficult to evaluate as stated. Therefore, in Appendix A, a superpopulation model version of Assumption 3.3 is stated, under which the $\boldsymbol{y}_i$ 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 Assumption 3.4 to hold almost surely under the superpopulation model, as derived in Lemma A.1 in the Appendix. We here assume that the (fixed) population sequence from which we are sampling is such that these results hold, without the almost sure condition. In other words, under the model assumptions stated in the Appendix, the sequence of populations that we are sampling from satisfies Assumption 3.4 with probability one.

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 $\mathcal{T}(\boldsymbol{\lambda})$ instead of nonsmooth population quantity $T_N(\boldsymbol{\lambda})$ in asymptotic expansion. Then, we establish the asymptotic normality of estimator (5) in Theorem 1 and give the expression of its asymptotic variance.

Lemma 1. *Under Assumptions 2.1(1), 2.2, 3.2(1) and 3.3-3.4,*

$$N^{\beta/2}\left[\widehat{T}(\hat{\boldsymbol{\lambda}}) - \widehat{T}(\boldsymbol{\lambda}_N) - \mathcal{T}(\hat{\boldsymbol{\lambda}}) + \mathcal{T}(\boldsymbol{\lambda}_N)\right]\Big| \mathcal{F}_N \xrightarrow{p} 0.$$

Proof. See Appendix. □

Theorem 1. *Under Assumptions 2.1(1) and 2.2-3.4, the sample estimator $\widehat{T}(\hat{\boldsymbol{\lambda}})$ is design consistent for $T_N(\boldsymbol{\lambda}_N)$ and asymptotically normally distributed,*

$$\left[\mathrm{AV}(\widehat{T}(\hat{\boldsymbol{\lambda}}))\right]^{-1/2}\left(\widehat{T}(\hat{\boldsymbol{\lambda}}) - T_N(\boldsymbol{\lambda}_N)\right)\Big| \mathcal{F}_N \xrightarrow{d} N(0,1)$$

where

$$\mathrm{AV}(\widehat{T}(\hat{\boldsymbol{\lambda}})) = \left(1, [\boldsymbol{\zeta}(\boldsymbol{\lambda}_N)]^T\right)\mathrm{Var}\left(\bar{\boldsymbol{z}}_\pi|\mathcal{F}_\mathrm{N}\right)\begin{pmatrix} 1 \\ \boldsymbol{\zeta}(\boldsymbol{\lambda}_N)\end{pmatrix},$$

$\bar{\boldsymbol{z}}_\pi = \frac{1}{N}\sum_{i\in\mathcal{S}}\frac{1}{\pi_i}\begin{bmatrix} h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N) \\ \boldsymbol{g}(\boldsymbol{y}_i)\end{bmatrix}$, *and $\boldsymbol{\zeta}(\boldsymbol{\lambda})$ denotes the first derivative of $\mathcal{T}(\boldsymbol{\lambda})$.*

Proof. We have the following decomposition

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

where the last term is stochastically small by Lemma 1. After linearization of the second term, we obtain

$$\begin{aligned}&\left[\mathrm{AV}(\widehat{T}(\hat{\boldsymbol{\lambda}}))\right]^{-1/2}\left\{\widehat{T}(\hat{\boldsymbol{\lambda}}) - T_N(\boldsymbol{\lambda}_N)\right\} \\ &= \left[\mathrm{AV}(\widehat{T}(\hat{\boldsymbol{\lambda}}))\right]^{-1/2}\left\{\tfrac{1}{N}\sum_{i\in\mathcal{S}}\frac{1}{\pi_i}h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N) - \tfrac{1}{N}\sum_{i\in U}h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)\right\} \\ &\quad + \left[\mathrm{AV}(\widehat{T}(\hat{\boldsymbol{\lambda}}))\right]^{-1/2}[\boldsymbol{\zeta}(\boldsymbol{\lambda}_N)]^T\left\{\tfrac{1}{N}\sum_{i\in\mathcal{S}}\frac{1}{\pi_i}\boldsymbol{g}(\boldsymbol{y}_i) - \tfrac{1}{N}\sum_{i\in U}\boldsymbol{g}(\boldsymbol{y}_i)\right\} + o_p(1). \quad (8)\end{aligned}$$

Design consistency of the estimator follows immediately from Assumptions 3.2(2) and 3.3 and the design assumptons 2.1-2.2. Using the fact that $h(\boldsymbol{y}_i; \boldsymbol{\lambda}_N)$ and $\boldsymbol{g}(\boldsymbol{y}_i)$ have finite fourth population moments and Assumption 2.3, we also obtain the normality of the sample estimator. $\square$

Generally speaking, for nondifferentiable survey estimators with estimated parameters, we can first replace the estimated parameter $\hat{\boldsymbol{\lambda}}$ with an arbitrary constant $\boldsymbol{\lambda}$ in $C_{\boldsymbol{\lambda}}$, then take expectation with respect to sampling design to obtain population quantity $T_N(\boldsymbol{\lambda})$. The population quantity usually remains as a nondifferentiable function of $\boldsymbol{\lambda}$, but we can often reasonably assume a differentiable limit for $T_N(\boldsymbol{\lambda})$ as in Assumption 3.3. The differentiable limit can then be used in asymptotic expansions and variance expressions.

This section furnishes a theoretical treatment of nondifferentiable survey estimators with estimated parameters, assuming that the estimator admits expression (5). In practice, many complex estimators in ongoing surveys can not be written in the simple form of a survey weighted order-1 U-statistic, but are differentiable functions of estimators with expression (5). The corollary below considers an estimator that can be written as $m(\hat{T}(\hat{\boldsymbol{\lambda}}))$, where m is a smooth function. Properties for such estimators are a straightforward extension of Theorem 1, since the additional effect of m is easily handled by traditional methods. Additional extensions to multivariate functions are not explicitly discussed here but follow the same approach.

Corollary 1. *Under Assumptions 2.1(1) and 2.2-3.4, and for any function $m(\cdot)$ with continuous first derivative in a neighborhood of $T_N(\boldsymbol{\lambda}_N)$,*

1. *$m(\widehat{T}(\hat{\boldsymbol{\lambda}}))$ is $\sqrt{n^*}$−consistent for $m(T_N(\boldsymbol{\lambda}_N))$.*
2. *The quantity $m'(T_N(\boldsymbol{\lambda}_N))^{-1}\mathrm{AV}(\widehat{T}(\hat{\boldsymbol{\lambda}}))^{-1/2}\left(m(\widehat{T}(\hat{\boldsymbol{\lambda}})) - m(T_N(\boldsymbol{\lambda}_N))\right)$ is asymptotically normally distributed with mean 0 and variance 1.*

In the next section, we discuss three specific examples of nondifferentiable estimators with estimated parameters that have appeared in the survey literature, including estimators for population distribution functions, estimators for population fractions above or below a sample-estimated level, and the endogenous post-stratification estimator. In each case, we show how their asymptotic design properties follow from the general results we just derived.

3.4 Applications

Estimating distribution functions using auxiliary information

There is an extensive literature on estimating the population distribution function of a target variable when auxiliary information is present (e.g. Dunstan and Chambers 1986, Rao, Kovar, and Mantel 1990, Chambers, Dorfman, and Hall 1992, Wang and Dorfman 1996). 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 usually a nondifferentiable function of the estimated parameter(s), like the model-based estimator in Dunstan and Chambers (1986) or the ratio, difference and Rao-Kovar-Mantel (RKM) estimators in Rao, Kovar, and Mantel (1990). It was stated in Rao, Kovar, and Mantel (1990) that we can ignore the variation due to estimating parameters in the last three estimators, but no rigorous proof was presented. We will show that this is because the derivative $\boldsymbol{\zeta}(\boldsymbol{\lambda}_N)$ is either strictly zero or a smaller order term.

We consider the difference estimator of Rao, Kovar, and Mantel (1990) as an example, defined as

$$\hat{F}_d(t;\hat{R}) = \tfrac{1}{N}\left\{\sum_{\mathcal{S}} \tfrac{1}{\pi_i}\mathrm{I}_{(y_i\le t)} + \left[\sum_{U}\mathrm{I}_{(\hat{R}x_i\le t)} - \sum_{\mathcal{S}}\tfrac{1}{\pi_i}\mathrm{I}_{(\hat{R}x_i\le t)}\right]\right\},$$

where $\hat{R}$ is a parameter estimated from the sample data. If we replace $\hat{R}$ by an arbitrary constant λ to obtain $\hat{F}_d(t;\lambda)$ and take expectation with respect to design, this is an unbiased estimator of

$$F_{N,d}(t) = \tfrac{1}{N}\sum_{U}\mathrm{I}_{(y_i\le t)},$$

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 RKM estimator in Rao, Kovar, and Mantel (1990). There is a slight difference between the difference estimator and ratio estimator, in that the derivative $\boldsymbol{\zeta}(\boldsymbol{\lambda}_N)$ is exactly zero for the difference estimator but a smaller order term for ratio estimator due to ratio bias. To summarize, for all the three distribution function estimators, estimating the population parameter will not contribute to an increase in the leading term in the asymptotic variance. The original paper by Rao, Kovar, and Mantel (1990) also argued the asymptotic equivalence of $\hat{F}_d(t;\hat{R})$ and $\hat{F}_d(t;R_N)$ by referring to the theoretical results of Randles (1982), but did not formally derive the result in the design-based setting.

Estimating a fraction below/above an estimated quantity

Another estimator that follows our framework is an estimated fraction below or above an estimated level, which is commonly seen in social surveys. There is extensive literature on variance estimation for the proportion below an estimated level, as in Shao and Rao (1993), Binder and Kovacevic (1995), Preston (1996), Deville (1999), Eurostat (2000) and Berger and Skinner (2003). 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. This sample fraction with estimated median plugged in is a nondifferentiable function of the estimated parameter, and we can apply the previous results to this situation. with $h(y_i;\lambda) = \mathrm{I}_{(y_i \le \lambda)}$ and $\hat{\lambda}$ as sample-based estimator for the population median λ_N for the variable y_i.

If we assume that y_i's 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. Theorem 1 can then be applied as long as we have a linearization or an asymptotic variance for the sample-based median estimator $\hat{\lambda}$, since the variance component due to estimation of the median remains significant in this case. The estimation of quantiles like the median will be discussed in Section 4.

A closely related application is discussed in Wang and Opsomer (2008), where the distribution of distances relative to a subpopulation center is estimated. The "center" is defined as the multivariate median or mean in the space defined by the survey variables $\boldsymbol{y}_i$. In this case, $h(\boldsymbol{y}_i;\boldsymbol{\lambda}) = \mathrm{I}_{(\|\boldsymbol{y}_i - \boldsymbol{\lambda}\| \le t)}$ and

$$\widehat{T}(\hat{\boldsymbol{\lambda}}) = \frac{1}{N}\sum_{i\in\mathcal{S}} \frac{1}{\pi_i}\mathrm{I}_{(\|\boldsymbol{y}_i - \hat{\boldsymbol{\lambda}}\| \le t)},$$

The corresponding population quantity $T_N(\boldsymbol{\lambda})$ is a nondifferentiable function of the $p-$dimentional parameter $\boldsymbol{\lambda}$, but we assume it has a differentiable limit so that the results of Section 3.3 can be applied. For a detailed treatment of this case and an application in the detection of survey outliers, see Wang and Opsomer (2008).

Endogenous post-stratification estimator

The endogenous post-stratification estimator (EPSE) is a more complicated estimator with estimated parameters discussed in Breidt and Opsomer (2008). While these authors showed that the estimator was design consistent, no other design-based properties were provided and the asymptotic behavior of EPSE was studied under a model-based framework. We apply the methodology of Section 3.3 to the EPSE and examine it from a design-based perspective.

The EPSE is a type of post-stratified estimator where the population units are to be assigned to strata based on the value of an index $m(\boldsymbol{\lambda}'\boldsymbol{x}_i)$, with $\boldsymbol{x}_i$ a vector of covariates known for the population. The strata are defined based on a set of predetermined stratum boundaries $-\infty \leq \tau_0 < \tau_1 < \cdots < \tau_{H-1} < \tau_H \leq \infty$. If $\boldsymbol{\lambda}$ is fixed, the resulting estimator is the classical post-stratification estimator. When $\boldsymbol{\lambda}$ is unknown, it needs to be estimated from the sample, for instance by considering a particular survey variable z_i and fitting a logistic regression model with $\mathrm{E}(\mathrm{z}|\boldsymbol{x}) = \mathrm{m}(\boldsymbol{\lambda}'\boldsymbol{x})$ to the sample observations $(\boldsymbol{x}_i, z_i)$, with m denoting the inverse logit function. For any variable in the survey, the EPSE is then defined as

$$\hat{\mu}_y(\hat{\boldsymbol{\lambda}}) = \sum_{h=1}^{H} \frac{A_{Nh0}(\hat{\boldsymbol{\lambda}})}{\hat{A}_{Nh0}(\hat{\boldsymbol{\lambda}})} \hat{A}_{Nh1}(\hat{\boldsymbol{\lambda}}) \tag{9}$$

with $A_{Nhl}(\hat{\boldsymbol{\lambda}}) = \frac{1}{N}\sum_U y_i^l \mathrm{I}_{\left\{\tau_{h-1} < m(\hat{\boldsymbol{\lambda}}'\boldsymbol{x}_i) \leq \tau_h\right\}}$, $\hat{A}_{Nhl}(\hat{\boldsymbol{\lambda}}) = \frac{1}{N}\sum_S \frac{y_i^l}{\pi_i} \mathrm{I}_{\left\{\tau_{h-1} < m(\hat{\boldsymbol{\lambda}}'\boldsymbol{x}_i) \leq \tau_h\right\}}$, $l = 0, 1$.

By assuming uniformly smooth limits for $A_{Nhl}(\boldsymbol{\lambda}), l = 0, 1$, for fixed $\boldsymbol{\lambda}$ and under the remaining assumptions from Section 2 and 3.2, Theorem 1 and a multivariate version of Corollary 1 can be applied to obtain asymptotic design-based results for $\hat{\mu}_y(\hat{\boldsymbol{\lambda}})$. The model regularity conditions in Appendix A can also be used to ensure that these design-based assumptions are reasonable for the variables of interest.

4 Nondifferentiable estimating equations

4.1 The Estimators

Godambe and Thompson (2009) give an overview of survey estimators defined by estimating equations and study their asymptotic properties when the estimating function is differentiable. But literature on nondifferentiable estimating equations in survey context is rare except in quantile estimation. We discuss the more general case for a population parameter ξ_N defined by

$$\xi_N = \inf\{\gamma : S_N(\gamma) \geq 0\}, \tag{10}$$

where

$$S_N(\gamma) = \tfrac{1}{N} \sum_{i=1}^{N} \psi(y_i - \gamma) \tag{11}$$

and $\psi(\cdot)$ is a univariate real function. The population parameter ξ_N is estimated by $\hat{\xi}$, where

$$\hat{\xi} = \inf\{\gamma : \hat{S}(\gamma) \geq 0\} \tag{12}$$

with

$$\hat{S}(\gamma) = \tfrac{1}{N} \sum_{(i \in \mathcal{S})} \tfrac{1}{\pi_i} \psi(y_i - \gamma). \tag{13}$$

4.2 Assumptions

In addition to the design assumptions in Section 2, we will require a further number of regularity conditions on the sequence of finite populations. Assumption 4.1 assumes that the population quantity ξ_N lives in a closed interval on $\Re$, and Assumption 4.2 specifies conditions on the monotonicity and smoothness of $S_N(t)$ and its limit.

Assumption 4.1. *The population parameter of interest $\xi_N \in C_\xi$, where C_ξ is a closed interval on $\Re$.*

Assumption 4.2. *The population score function $S_N(\cdot)$ and $\psi(\cdot)$ satisfy the following conditions:*

1. *The score function $\psi(\cdot)$ is bounded.*

2. *$\lim_{n \to \infty} S_N(\gamma) = S(\gamma)$ uniformly on C_ξ, and the equation $S(\gamma) = 0$ has a unique root in the interior of C_ξ.*

3. *The limiting function $S(\gamma)$ is 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 quantity*

$$\sup_{\gamma \in C_\gamma} N^\alpha \left| S_N(\xi_N + N^{-\alpha}\gamma) - S_N(\xi_N) - S(\xi_N + N^{-\alpha}\gamma) + S(\xi_N) \right| \to 0, \tag{14}$$

where C_γ is a large enough compact set in $\Re$ and $\alpha \in \left(\frac{1}{4}, \frac{1}{2}\right]$.

Similar to Assumption 3.4 above, Assumption 4.2(4) is somewhat restrictive and difficult to interpret. In Appendix A, we show that it holds with probability one under suitable assumptions on the probability mechanism generating the y_i and on the function ψ.

4.3 Design-based results

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

Lemma 2. *Under Assumptions 2.1(2), 2.2(2) and 4.2, for any large enough closed interval $C \in \Re$,*

$$\sup_{\gamma \in C} \left| \hat{S}(\gamma) - S_N(\gamma) \right| \xrightarrow{p} 0.$$

Theorem 2. *Under Assumptions 2.1(2), 2.2, 4.1 and 4.2, and assuming the limiting function $S(\gamma)$ is strictly increasing, the sample estimator $\hat{\xi}$ is design consistent for the population quantity ξ_N.*

Theorem 3. *Under Assumptions 2.1(2), 2.2-2.3, 4.1-4.2, for any sequence of sample estimators $\hat{\xi}$ that is $\sqrt{n^*}$-consistent for ξ_N,*

1. *the sample estimator $\hat{\xi}$ admits the following linearization*

$$\hat{\xi} = \xi_N - \frac{\hat{S}(\xi_N) - S_N(\xi_N)}{S'(\xi_N)} + o_p(n^{*-1/2}).$$

2. *the sample estimator $\hat{\xi}$ is asymptotically normally distributed,* i.e.

$$\left[AV(\hat{\xi})\right]^{-1/2} (\hat{\xi} - \xi_N) \Big| \mathcal{F}_N \xrightarrow{d} N(0,1),$$

where

$$AV(\hat{\xi}) = \frac{\mathrm{Var}(\hat{S}(\xi_N) | \mathcal{F}_N)}{S'^2(\xi_N)}. \tag{15}$$

4.4 Applications

In this section, we discuss examples of nondifferentiable estimating equations and bring them into the theoretical framework established in the preceding section. The first example is the sample quantile. The estimation of quantiles using survey data has been well studied in the literature. Kuk and Mak (1989) discuss median estimation using auxiliary information under simple random sampling. Rao, Kovar, and Mantel (1990) furnish a thorough treatment of estimating distribution functions and quantiles in the presence of auxiliary information under general sampling design. Francisco and Fuller (1991) derive design normality for both distribution function and quantile estimators, and propose a number of confidence intervals for quantiles, which are carefully examined by Sitter and Wu (2001) under stratified cluster sampling.

For simplicity, consider the sample quantile estimator obtained by inverting the Hajek estimator of the cumulative distribution function. In this case, the estimating function for α quantile is

$$\psi(y_i - \gamma) = \mathrm{I}_{(y_i - \gamma \leq 0)} - \alpha,$$

with population estimating equation

$$S_{N,\alpha}(\gamma) = \tfrac{1}{N} \sum_{i=1}^{N} \mathrm{I}_{(y_i - \gamma \leq 0)} - \alpha.$$

The α sample quantile is defined as

$$\hat{\xi}_\alpha = \inf\{t : \hat{S}_\alpha(\gamma) \geq 0\} = \inf\left\{\gamma : \tfrac{1}{N} \sum_{i \in \mathcal{S}} \tfrac{1}{\pi_i} \mathrm{I}_{(y_i \leq \gamma)} \geq \alpha\right\}.$$

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 asymptotic variance of $\hat{\xi}$ using design variance $\mathrm{Var}(\hat{\mathrm{S}}_\alpha(\xi_\mathrm{N})|\mathcal{F}_\mathrm{N})$ 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

$$\begin{aligned}\psi(y_i-\gamma) &= \begin{cases} -k, & y_i-\gamma<k \\ y_i-\gamma, & |y_i-\gamma|\le k \\ k, & y_i-\gamma>k \end{cases} \\ &= (y_i-\gamma)\mathrm{I}_{(|y_i-\gamma|\le k)} - k\,\mathrm{I}_{(y_i-\gamma<-k)} + k\,\mathrm{I}_{(y_i-\gamma>k)}.\end{aligned}$$

The population score function is

$$S_N(\gamma)=\tfrac{1}{N}\sum_{i=1}^{N}(y_i-\gamma)\mathrm{I}_{(|y_i-\gamma|\le k)}+k\sum_{i=1}^{N}\left[\mathrm{I}_{(y_i-\gamma>k)}-\mathrm{I}_{(y_i-\gamma<-k)}\right],$$

and we assume $S_N(\gamma)$ converges a limit function $S(\gamma)$ which is differentiable in a neighborhood of ξ_N, where ξ_N is the population Winsorized mean as defined by (10). This population score function is nonincreasing, but we can use $-S_N(\gamma)$, and still define the parameter of interest as (10). Then we can define 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 and Hallock (2001) and Koenker (2005). There does not currently appear to exist any 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 will 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 $\widehat{T}(\hat{\boldsymbol{\lambda}})$ in Section 3 or $\hat{\xi}$ in Section 4, we need to estimate the derivatives $\boldsymbol{\zeta}(\boldsymbol{\lambda}_N)$ or $S'(\xi_N)$ of the limiting function $\mathcal{T}(\boldsymbol{\lambda})$ or $S(\gamma)$, respectively. We can start from some smoothed estimator of primitive functions $\mathcal{T}(\boldsymbol{\lambda})$ or $S(\gamma)$, and take derivatives with respect to $\boldsymbol{\lambda}$ or γ. Natural sample-based estimators are $\widehat{T}(\boldsymbol{\lambda})$ and $\hat{S}(\gamma)$, but being non-differentiable, they cannot be used directly to obtain derivatives. We therefore work with a smoothed version of those estimator. The two cases can be handled in similar fashion, so we only examine the first case to avoid duplication. This section describes a direct "plug-in" variance estimator, in which the derivative is estimated by a kernel-based estimator and plugged into the variance formula of Theorem 1. The next section shows how to integrate the kernel-based derivative estimator into a replication-based variance estimator.

We denote $K_q(\cdot)$ as a kernel function in $\Re^q$, and convolute the nonsmooth function $h(\boldsymbol{y}_i;\cdot)$ with $K_q(\cdot)$ using bandwidth b to obtain

$$h_i * K_q(\boldsymbol{\lambda}) = \int \cdots \int h(\boldsymbol{y}_i;\boldsymbol{x}) K_q\left(\tfrac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right) d\boldsymbol{x},$$

so that we can estimate $\mathcal{T}(\boldsymbol{\lambda})$ by

$$\tfrac{1}{N}\sum_{\mathcal{S}} \tfrac{1}{\pi_i} \int \cdots \int h(\boldsymbol{y}_i;\boldsymbol{x}) K_q\left(\tfrac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right) d\boldsymbol{x}. \tag{16}$$

Taking a derivative of (16) with respect to $\boldsymbol{\lambda}$, we obtain the estimator

$$\hat{\boldsymbol{\zeta}}(\boldsymbol{\lambda}) = \tfrac{1}{Nb^q}\sum_{\mathcal{S}} \tfrac{1}{\pi_i} \int \cdots \int h(\boldsymbol{y}_i;\boldsymbol{x}) K'_q\left(\tfrac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right) d\boldsymbol{x}, \tag{17}$$

which estimates the population quantity (for fixed $\boldsymbol{\lambda}$)

$$\boldsymbol{\zeta}_N(\boldsymbol{\lambda}) = \tfrac{1}{Nb^q}\sum_{U} \int \cdots \int h(\boldsymbol{y}_i;\boldsymbol{x}) K'_q\left(\tfrac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right) d\boldsymbol{x}. \tag{18}$$

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

Assumption 5.1. *The following conditions hold for the 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 $\int\cdots\int K_q(\boldsymbol{x})d\boldsymbol{x} = 1$.*
2. *The bandwidth $b \to 0$ and $Nb^q \to \infty$, as $N \to \infty$.*
3. *There exists a constant c, such that $\left\|\frac{1}{b^q}K'_q(\frac{\boldsymbol{x}_1}{b}) - \frac{1}{b^q}K'_q(\frac{\boldsymbol{x}_2}{b})\right\| \le c\,\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$ for any $\boldsymbol{x}_1, \boldsymbol{x}_2$, and b arbitrarily small.*

Assumption 5.2. *The deviance* $\|\boldsymbol{\zeta}_N(\boldsymbol{\lambda}) - \boldsymbol{\zeta}(\boldsymbol{\lambda})\| \to 0$ *uniformly for* $\boldsymbol{\lambda} \in C_{\boldsymbol{\lambda}}$.

As in the previous sections, we can justify Assumption 5.2 by showing that under some stated model regularity conditions on the $\boldsymbol{y}_i$, it holds with probability one for sufficiently large populations. This is done in Appendix A. Assumption 5.1 states conditions on the smoothness and tail behavior of the kernel functions. The assumptions we need are mild and popular kernel functions including Epanechnikov, Gaussian, and Triangle kernels all satisfy the required conditions.

Given Assumptions 5.1, 5.2 and previously stated regularity conditions on the sampling design, we can show the consistency of the kernel-based estimator (17) in Lemma 3 and of the resulting variance estimator in Theorem 4.

Lemma 3. *Under Assumptions 2.2, 3.1-3.2(1), 5.1(1-3) and 5.2, the estimator* $\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}})$ *is design consistent for* $\boldsymbol{\zeta}(\boldsymbol{\lambda}_N)$.

Theorem 4. *Let* $\widehat{V}_{HT}(\bar{\boldsymbol{z}}_\pi)$ *be the Horvitz-Thompson variance estimator for* $\bar{\boldsymbol{z}}_\pi$ *as defined in Theorem 1. Under Assumptions 2.2-2.3, 3.1-3.2, 3.4(1), 5.1-5.2, the estimator*

$$\widehat{V}(\widehat{T}(\hat{\boldsymbol{\lambda}})) = \left(1, \left(\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}})\right)^T\right) \widehat{V}_{HT}(\bar{\boldsymbol{z}}_\pi) \begin{pmatrix} 1 \\ \hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}}) \end{pmatrix} \tag{19}$$

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

Proof. The proof easily follows from Assumption 2.3, Lemma 3, and the unbiasedness of $\widehat{V}_{HT}(\bar{\boldsymbol{z}}_\pi)$. □

Similarly, one can obtain the design consistency of the estimator

$$\widehat{V}(\hat{\xi}) = \frac{\widehat{V}_{HT}(\widehat{S}(\hat{\xi}))}{(\hat{S}'(\hat{\xi}))^2},$$

for AV($\hat{\xi}$) in (15), where

$$\hat{S}'(\gamma) = \frac{1}{Nb} \sum_{\mathcal{S}} \frac{1}{\pi_i} \int \psi(y_i - x) K' \left(\frac{\gamma - x}{b}\right) dx \tag{20}$$

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

5.2 Jackknife variance estimator

We start by assuming 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 is also used by Fuller and Kim (2005) and Da Silva and Opsomer (2006). We will use a number of regularity assumptions on the replication method that are stated in the latter article, and not repeat them here fully for the sake of brevity.

Theorem 5. *We define sampling weight $w_i = \frac{1}{N\pi_i}$ for the i-th population unit 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 $l = 1, 2, \cdots, L$ and $w_i^{(l)}$ is the replication weight for unit i in replicate l. The replication variance estimator is defined as

$$\widehat{V}_{JK}(\hat{\theta}) = \sum_{l=1}^{L} c_l \left(\hat{\theta}^{(l)} - \hat{\theta}\right)^2, \tag{21}$$

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

1. *For explicit nondifferentiable survey estimators, we define the l-th jackknife replicate as*

$$\widehat{T}^{(l)}(\hat{\boldsymbol{\lambda}}) = \sum_{\mathcal{S}} w_i^{(l)} h(\boldsymbol{y}_i; \hat{\boldsymbol{\lambda}}) + [\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}})]^T (\hat{\boldsymbol{\lambda}}^{(l)} - \hat{\boldsymbol{\lambda}}), \tag{22}$$

where the design variance of $\hat{\boldsymbol{\lambda}}$ can be consistently estimated by $\sum_{l=1}^{L} c_l \left(\hat{\boldsymbol{\lambda}}^{(l)} - \hat{\boldsymbol{\lambda}}\right)^2$, and $\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}})$ is a kernel estimator as defined in (17). Then the jackknife variance estimator

$$\widehat{V}_{JK}\left(\widehat{T}(\hat{\boldsymbol{\lambda}})\right) = \sum_{l=1}^{L} c_l \left(\widehat{T}^{(l)}(\hat{\boldsymbol{\lambda}}) - \widehat{T}(\hat{\boldsymbol{\lambda}})\right)^2 \tag{23}$$

is design consistent for $AV\left(\widehat{T}(\hat{\boldsymbol{\lambda}})\right)$ in Theorem 1.

2. *For estimators defined by nondifferentiable estimating equations, we use the following jackknife replicate,*

$$\hat{\xi}^{(l)} = \frac{1}{\hat{S}'(\hat{\xi})} \sum_{i \in \mathcal{S}} w_i^{(l)} \psi(y_i - \hat{\xi}), \tag{24}$$

where $\hat{S}'(\hat{\xi})$ is defined in (20). Then the jackknife variance estimator

$$\widehat{V}_{JK}\left(\hat{\xi}\right) = \sum_{l=1}^{L} c_l \left(\hat{\xi}^{(l)} - \frac{1}{L}\sum_{l=1}^{L} \hat{\xi}^{(l)}\right)^2 \tag{25}$$

is design consistent for $AV(\hat{\xi})$ 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 (23) is readily interpreted by considering the composition of the replicate in (22). Ignoring the second term in (22), the resulting jackknife variance estimator consistently estimates the asymptotic variance of $\widehat{T}(\boldsymbol{\lambda}_N)$, with population parameter $\boldsymbol{\lambda}_N$ substituted. The second term in (22) 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 $\sum_{i \in \mathcal{S}} w_i^{(l)} \psi(y_i - \hat{\xi})$ allow us to consistently estimate $\mathrm{Var}(\hat{\mathrm{S}}(\xi_{\mathrm{N}})|\mathcal{F}_{\mathcal{N}})$, and thus the whole jackknife estimator is consistent for the target asymptotic variance.

In many practical situations, a significant advantage of the replication-based approach is that we do not need to compute explicit design-based variance-covariance matrices, which can be complicated in a large-scale complex survey. In jackknife variance estimation, we start from an existing jackknife procedure for simple linear estimators, as could for instance be provided by a statistical agency as part of the survey dataset. We then estimate the gradient vector using kernel smoothing on the whole sample only once, while the linearizable terms are computed for each replicate. Hence, starting from an existing replication method, it is straightforward to obtain variance estimates for complicated estimators as proposed in Sections 3 and 4.

6 Simulation study

We perform a simulation study to examine the performance of the proposed analytic and jackknife variance estimators. We consider two survey estimators: the estimator of the proportion below an estimated threshold (Section 3.4.2) and the sample quantile defined by inverting the sample CDF (Section 4.4).

We generate a fixed finite population of size $N = 2000$, and generate the study variable $\boldsymbol{y}$ 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 $z_i = y_i + \frac{1}{\sqrt{y_i}} + 5 + \epsilon_i$ 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 $n_1 = \frac{n^*}{4}$ 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 $n^* = 200, 400$ 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:

1. population α-quantiles ξ_α with $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$.

2. 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 CDF, defined as

$$\hat{F}(\gamma) = \sum_{h=1}^{3} \frac{N_h}{N} \frac{\sum_{\mathcal{S}_h} \frac{1}{\pi_i} \mathrm{I}_{(y_i \leq \gamma)}}{\sum_{\mathcal{S}_h} \frac{1}{\pi_i}},$$

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

$$\widehat{T}(c\hat{\xi}_{0.5}) = \sum_{h=1}^{3} \frac{N_h}{N} \frac{\sum_{\mathcal{S}_h} \frac{1}{\pi_i} \mathrm{I}_{(y_i \leq c\hat{\xi}_{0.5})}}{\sum_{\mathcal{S}_h} \frac{1}{\pi_i}}. \tag{26}$$

We will compare three variance estimators: the analytic variance estimator from Theorem 4 (denoted $\widehat{V}_{AN}$), the jackknife estimator from Theorem 5 ($\widehat{V}_{JK}$) and a "naive" jackknife variance estimator that ignores the special structure of the estimator and calculates a sample quantile for each replicate ($\widehat{V}_{NVJK}$). Note that the estimator (26) contains two nested non-differentiable functions, so that the estimation of $\mathrm{AV}(\widehat{T}(c\hat{\xi}_{0.5}))$ requires the estimation of the density of $\boldsymbol{y}$ at $\hat{\xi}_{0.5}$ and at $c\hat{\xi}_{0.5}$. We use kernel regression with Gaussian kernel and the same bandwidth $h = 0.1, 0.2, 0.4$ in estimating the density at both locations.

Table 1: Relative bias of three variance estimators for sample quantiles under three different bandwidths h and two sample size n^*.

Variance estimator	h	$n^* = 200$					$n^* = 400$				
		$\hat{\xi}_{0.1}$	$\hat{\xi}_{0.25}$	$\hat{\xi}_{0.5}$	$\hat{\xi}_{0.75}$	$\hat{\xi}_{0.9}$	$\hat{\xi}_{0.1}$	$\hat{\xi}_{0.25}$	$\hat{\xi}_{0.5}$	$\hat{\xi}_{0.75}$	$\hat{\xi}_{0.9}$
$\widehat{V}_{AN}$	0.1	-8%	4%	1%	-2%	-6%	-3%	16%	1%	1%	-2%
$\widehat{V}_{JK}$		-7%	5%	3%	0%	-1%	-3%	16%	2%	1%	0%
$\widehat{V}_{AN}$	0.2	0%	9%	3%	-7%	-1%	5%	21%	3%	-6%	0%
$\widehat{V}_{JK}$		2%	10%	4%	-5%	4%	5%	21%	3%	-4%	3%
$\widehat{V}_{AN}$	0.4	26%	28%	3%	-11%	4%	32%	41%	3%	-9%	8%
$\widehat{V}_{JK}$		28%	29%	4%	-9%	8%	33%	42%	4%	-8%	9%
$\widehat{V}_{NVJK}$		186%	204%	156%	221%	253%	206%	229%	182%	208%	232%

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	$n^* = 200$			$n^* = 400$		
		$\widehat{T}(0.25\,\hat{\xi}_{0.5})$	$\widehat{T}(0.4\,\hat{\xi}_{0.5})$	$\widehat{T}(0.6\,\hat{\xi}_{0.5})$	$\widehat{T}(0.25\,\hat{\xi}_{0.5})$	$\widehat{T}(0.4\,\hat{\xi}_{0.5})$	$\widehat{T}(0.6\,\hat{\xi}_{0.5})$
$\widehat{V}_{AN}$	0.1	-1%	1%	4%	-1%	-2%	3%
$\widehat{V}_{JK}$		0%	2%	5%	-1%	-2%	3%
$\widehat{V}_{AN}$	0.2	-3%	0%	0%	-3%	-3%	0%
$\widehat{V}_{JK}$		-2%	1%	1%	-2%	-2%	1%
$\widehat{V}_{AN}$	0.4	-4%	-2%	-5%	-4%	-5%	-4%
$\widehat{V}_{JK}$		-3%	-1%	-4%	-3%	-4%	-3%
$\widehat{V}_{NVJK}$		137%	157%	304%	181%	173%	357%

The results in Table 1 show that 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. The proposed analytic and jackknife variance estimators provide satisfactory results except in a number of cases under bandwidth $h = 0.4$. Overall, the simulation results suggest that the two proposed variance estimators work reasonably well under appropriate bandwidth, although the preformance of the variance estimators depend on bandwidth h to some extent. The case with $h = 0.4$ appears to result in substantial oversmoothing of the data, at least for the smaller quantiles. Table 2 shows the same results for the sample estimator defined in (26). The proposed variance estimators have relative bias less than 5% for all bandwidths, while the naive variance estimator is again severely biased.

A Appendix: 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 evaluate 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 $F(\boldsymbol{y})$, and we show that these assumptions on the sequences of finite populations hold with probability one. The proofs are in Appendix B.

Under a model version of Assumption 3.3 and an additional mild regularity condition on function $h(\cdot;\cdot)$, we show that the statement in Assumption 3.4 holds with probability one under the superpopulation model.

Assumption A.1. *Let $\boldsymbol{Y}$ be a random vector with absolutely continuous cumulative distribution function $F(\boldsymbol{y})$, and denote $\mathcal{T}(\boldsymbol{\lambda}) \equiv \mathrm{E}(\mathrm{h}(\boldsymbol{Y};\boldsymbol{\lambda}))$. Further, $\mathcal{T}(\boldsymbol{\lambda})$ is a continuous function of $\boldsymbol{\lambda}$, with finite first and second derivatives.*

Assumption A.2. *There exists a finite integer m_1, such that $h(\boldsymbol{y};\boldsymbol{\lambda}) = \sum_{j=1}^{m_1} h_j(\boldsymbol{y};\boldsymbol{\lambda})$, where $h_j(\boldsymbol{y};\boldsymbol{\lambda})$ is monotone in $\boldsymbol{y}$ for every $\boldsymbol{\lambda}$.*

Lemma A.1. *Under Assumptions 3.3(1), A.1 and A.2, the population quantity*

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

uniformly for $\boldsymbol{s}\in C_{\boldsymbol{s}}$, a large enough compact set in $\Re^q$, $\boldsymbol{\lambda}\in C_{\boldsymbol{\lambda}}$, where $C_{\boldsymbol{\lambda}}$ is defined in Assumption 3.1 and $\alpha\in\left(\frac{1}{4},\frac{1}{2}\right]$.

Next, we justify Assumption 4.2(4) under a probabilistic model. 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 A.3. *Let Y be a random variable with absolutely continuous cumulative distribution function, and denote $S(\gamma)\equiv\mathrm{E}\left[\psi(Y-\gamma)\right]$. The score function $S(\gamma)$ is strictly increasing with finite first derivative.*

Assumption A.4. *The function $\psi(\cdot)$ is bounded and has a finite number of monotonicity changes.*

Lemma A.2. *Under Assumptions A.3 and A.4, the population quantity*

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

uniformly for $s\in C_s$, a large enough closed interval in $\Re$.

Finally, we address Assumption 5.2. In addition to the previous assumption A.1, we require additional regularity conditions on the function h and the kernel K_q.

Assumption A.5. *There exists a finite integer m_2, such that $K'_q(\boldsymbol{\lambda})=\sum_{j=1}^{m_2}K'_{q,j}(\boldsymbol{\lambda})$, where $K'_{q,j}(\boldsymbol{\lambda})$ has no change of sign for any j, i.e., $K'_{q,j}(\boldsymbol{\lambda})$ is either nonnegative or nonpositive for any $\boldsymbol{\lambda}$.*

Lemma A.3. *Under Assumptions 5.1(1,2), A.1, A.2 and A.5, the deviance*

$$\sup_{\boldsymbol{\lambda}\in C_{\boldsymbol{\lambda}}}\|\boldsymbol{\zeta}_N(\boldsymbol{\lambda})-\boldsymbol{\zeta}(\boldsymbol{\lambda})\|\overset{a.s.}{\to}0.$$

B Appendix: Proofs

Proof of Lemma 1

Proof. We define

$$Q_n(\boldsymbol{s}) = n^{*1/2}\left(\widehat{T}(\boldsymbol{\lambda}_N + n^{*-1/2}\boldsymbol{s}) - \widehat{T}(\boldsymbol{\lambda}_N) - \mathcal{T}(\boldsymbol{\lambda}_N + n^{*-1/2}\boldsymbol{s}) + \mathcal{T}(\boldsymbol{\lambda}_N)\right),$$

and we need to show that

$$\sup_{\boldsymbol{s}\in C} |Q_n(\boldsymbol{s})| \xrightarrow{p} 0,$$

where C is a sufficiently large compact region in $\Re^q$. We have

$$\begin{aligned}
|Q_n(\boldsymbol{s})| \leq\ & n^{*1/2}\left|\widehat{T}(\boldsymbol{\lambda}_N + n^{*-1/2}\boldsymbol{s}) - \widehat{T}(\boldsymbol{\lambda}_N) - T_N(\boldsymbol{\lambda}_N + n^{*-1/2}\boldsymbol{s}) + T_N(\boldsymbol{\lambda}_N)\right| \\
& + n^{*1/2}\left|T_N(\boldsymbol{\lambda}_N + n^{*-1/2}\boldsymbol{s}) - T_N(\boldsymbol{\lambda}_N) - \mathcal{T}(\boldsymbol{\lambda}_N + n^{*-1/2}\boldsymbol{s}) + \mathcal{T}(\boldsymbol{\lambda}_N)\right|,
\end{aligned}$$

where the supremum of the second term converges to zero by Assumption 3.4. Now we need to show the supremum of the first term converges to zero. For any $1-\beta < \xi < \frac{\beta}{2p}$, where β is defined in Assumption 2.1, partition the compact region C into,

$$C = C_1 \cup C_2 \cup \cdots \cup C_{N^{\xi p}}, C_j \cap C_{j'} = \emptyset, \forall j \neq j',$$

where $\mathrm{Diam}(C_j) = O(N^{-\xi}), \forall j = 1, 2, \cdots, N^{\xi p}$. For any set of $\boldsymbol{s}_j \in C_j, j = 1, 2, \cdots, N^{\xi p}$, we have

$$
\begin{aligned}
&\mathrm{P}\left(\max_j \left|\frac{n^{*1/2}}{N}\sum_{i=1}^N\left(\frac{\mathrm{I}_{(i\in\boldsymbol{s})}}{\pi_i}-1\right)\left[h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)\right]\right|>\epsilon\right)\\
\leq\ & \sum_{j=1}^{N^{\xi p}}\mathrm{P}\left(\left|\frac{n^{*1/2}}{N}\sum_{i=1}^N\left(\frac{\mathrm{I}_{(i\in\boldsymbol{s})}}{\pi_i}-1\right)\left[h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)\right]\right|>\epsilon\right)\\
\leq\ & \frac{c_1}{\epsilon^2}\sum_{j=1}^{N^{\xi p}}\frac{1}{N-1}\sum_{i=1}^N\left[h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)\right]^2\\
\leq\ & \frac{4c_1c_h}{\epsilon^2}\sum_{j=1}^{N^{\xi p}}\frac{1}{N}\sum_{i=1}^N\left|h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)\right|\\
\leq\ & \frac{4c_1c_h}{\epsilon^2}\sum_{j=1}^{N^{\xi p}}\frac{1}{N}\sum_{i=1}^N\left|h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)-\mathcal{T}(\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)+\mathcal{T}(\boldsymbol{\lambda}_N)\right|\\
& +\frac{4c_1c_h}{\epsilon^2}\sum_{j=1}^{N^{\xi p}}\frac{1}{N}\sum_{i=1}^N\left|\mathcal{T}(\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-\mathcal{T}(\boldsymbol{\lambda}_N)\right|\\
\leq\ & \frac{4c_1c_h}{\epsilon^2}N^{\xi p}\sup_j\frac{1}{N}\sum_{i=1}^N\left|h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)-\mathcal{T}(\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)+\mathcal{T}(\boldsymbol{\lambda}_N)\right|\\
& +O(N^{\xi p-\frac{\beta}{2}}),\\
=\ & O(N^{\xi p-\frac{\beta}{2}})
\end{aligned}
$$

where c_1 is defined in (2), c_h is defined in Assumption 3.3(1) and the last term goes to zero as $\xi<\frac{\beta}{2p}$. Additionally,

$$
\begin{aligned}
&\max_j\sup_{\boldsymbol{s}\in C_j}\left|\frac{n^{*1/2}}{N}\sum_{i=1}^N\left(\frac{\mathrm{I}_{(i\in\mathcal{S})}}{\pi_i}-1\right)\left[h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s})-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)\right]\right|\\
\leq\ & \frac{N}{n^{*1/2}}\max_j\sup_{\boldsymbol{s}\in C_j}\frac{1}{N}\sum_{i=1}^N\left|\left[h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s})-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)\right]\right|\\
\leq\ & \frac{1}{\sqrt{n^*}}\sup_{j,\boldsymbol{s}\in C_j}\sum_{i=1}^N\left|h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+\frac{\boldsymbol{s}}{\sqrt{n^*}})-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+\frac{\boldsymbol{s}_j}{\sqrt{n^*}})-\mathcal{T}(\boldsymbol{\lambda}_N+\frac{\boldsymbol{s}}{\sqrt{n^*}})+\mathcal{T}(\boldsymbol{\lambda}_N+\frac{\boldsymbol{s}_j}{\sqrt{n^*}})\right|\\
& +\frac{N}{n^{*1/2}}\max_j\sup_{\boldsymbol{s}\in C_j}\frac{1}{N}\sum_{i=1}^N\left|\mathcal{T}(\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-\mathcal{T}(\boldsymbol{\lambda}_N)\right|\\
\to\ & 0,
\end{aligned}
$$

by Assumption 3.4. So the first term in (27) can be bounded by

$$\mathrm{P}\left(\max_j \left|\frac{n^{*1/2}}{N}\sum_{i=1}^N\left(\frac{\mathrm{I}_{(i\in \boldsymbol{s})}}{\pi_i}-1\right)\left[h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N)\right]\right|>\epsilon\right)$$

and

$$\max_j \sup_{\boldsymbol{s}\in C_j}\left|\frac{n^{*1/2}}{N}\sum_{i=1}^N\left(\frac{\mathrm{I}_{(i\in \mathcal{S})}}{\pi_i}-1\right)\left[h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s})-h(\boldsymbol{y}_i;\boldsymbol{\lambda}_N+n^{*-1/2}\boldsymbol{s}_j)\right]\right|,$$

and thus the proof is completed. □

Proof of Lemma 2

Proof. The lemma is proved via covering approach. We partition C into N^ν equal sub-intervals,

$$C=\cup_{k=1}^{N^\nu}C_k,$$

with $\frac{1}{2}<\nu<\beta$ and select an arbitrary point $t_k\in C_k, k=1,2,\cdots,N^\nu$. Then,

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

and it suffices to show the two terms on the RHS are both stochastically small. Let us first calculate the design variance of $\hat{S}(\gamma)-S_N(\gamma)$:

$$\begin{aligned}
\mathrm{Var}(\hat{\mathrm{S}}(\gamma)-\mathrm{S_N}(\gamma)|\mathcal{F}_N) &= \frac{1}{N^2}\sum\sum(\pi_{ij}-\pi_i\pi_j)\frac{\psi(y_i-\gamma)-S_N(\gamma)}{\pi_i}\frac{\psi(y_j-\gamma)-S_N(\gamma)}{\pi_j}\\
&\le c_1\frac{1}{n^*}\frac{1}{N-1}\sum_{i=1}^N[\psi(y_i-\gamma)-S_N(\gamma)]^2\\
&\le \frac{c_2}{n^*}
\end{aligned}$$

by Assumption 4.2(1), so that

$$\mathrm{P}(\max_k\left|\hat{S}(\gamma_k)-S_N(\gamma_k)\right|\ge\epsilon)\le\sum_{k=1}^{N^\nu}\frac{\mathrm{Var}(\hat{\mathrm{S}}(\gamma_\mathrm{k})-\mathrm{S_N}(\gamma_\mathrm{k})|\mathcal{F}_N)}{\epsilon^2}\to 0.$$

Now let us show that

$$\max_k\sup_{\gamma\in C_k}\left|\left[\hat{S}(\gamma)-S_N(\gamma)\right]-\left[\hat{S}(\gamma_k)-S_N(\gamma_k)\right]\right|\xrightarrow{p}0.$$

This follows from the fact that

$$
\begin{aligned}
&\max_k \sup_{\gamma \in C_k} \left| \left[\hat{S}(\gamma) - S_N(\gamma) \right] - \left[\hat{S}(\gamma_k) - S_N(\gamma_k) \right] \right| \\
= \;& \max_k \sup_{\gamma \in C_k} \left| \frac{1}{N} \sum_{i=1}^{n} \frac{1}{\pi_i} (\psi(y_i - \gamma) - \psi(y_i - \gamma_k)) - [S_N(\gamma) - S_N(\gamma_k)] \right| \\
= \;& \max_k \sup_{\gamma \in C_k} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathrm{I}_i}{\pi_i} - 1 \right) (\psi(y_i - \gamma) - \psi(y_i - \gamma_k)) \right| \\
\leq \;& c_4 \frac{N}{n^*} \max_k \sup_{\gamma \in C_k} \frac{1}{N} \sum_{i=1}^{N} |\psi(y_i - \gamma) - \psi(y_i - \gamma_k)| \\
= \;& O(N^{\frac{1}{2} - \beta}),
\end{aligned}
$$

where c_4 is a positive constant, and the last equation follows from Assumption 4.2(4). □

Proof of Lemma 3:

Proof. The triangle inequality implies that

$$
\begin{aligned}
&\|\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}}) - \boldsymbol{\zeta}(\boldsymbol{\lambda}_N)\| \\
&\leq \|\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}}) - \hat{\boldsymbol{\zeta}}(\boldsymbol{\lambda}_N)\| + \|\hat{\boldsymbol{\zeta}}(\boldsymbol{\lambda}_N) - \boldsymbol{\zeta}_N(\boldsymbol{\lambda}_N)\| + \|\boldsymbol{\zeta}_N(\boldsymbol{\lambda}_N) - \boldsymbol{\zeta}(\boldsymbol{\lambda}_N)\|,
\end{aligned}
$$

where the third term goes to zero by Assumption 5.2. For the first term,

$$
\begin{aligned}
\|\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\lambda}}) - \hat{\boldsymbol{\zeta}}(\boldsymbol{\lambda}_N)\| &= \left\| \tfrac{1}{Nb^q} \sum_{\mathcal{S}} \tfrac{1}{\pi_i} \int \cdots \int h(\boldsymbol{y}_i; \boldsymbol{x}) \left[K_q' \left(\tfrac{\hat{\boldsymbol{\lambda}} - \boldsymbol{x}}{b} \right) - K_q' \left(\tfrac{\boldsymbol{\lambda}_N \boldsymbol{x}}{b} \right) \right] d\boldsymbol{x} \right\| \\
&= \left\| \tfrac{1}{N} \sum_{\mathcal{S}} \tfrac{1}{\pi_i} \int \cdots \int h(\boldsymbol{y}_i; \boldsymbol{x}) \frac{1}{b^q} \left[K_q' \left(\tfrac{\hat{\boldsymbol{\lambda}} - \boldsymbol{x}}{b} \right) - K_q' \left(\tfrac{\boldsymbol{\lambda}_N \boldsymbol{x}}{b} \right) \right] d\boldsymbol{x} \right\| \\
&\leq c \, \|\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_N\| \cdot \sup_{\boldsymbol{x}} |h(\boldsymbol{y}_i; \boldsymbol{x})| \cdot \tfrac{1}{N} \sum_{\mathcal{S}} \tfrac{1}{\pi_i} \cdot \int \cdots \int d\boldsymbol{x} \\
&\xrightarrow{p} 0
\end{aligned}
$$

by the conditions in Assumption 5.1. Finally, the second term also converges to zero in probability because its asymptotic variance goes to zero by Assumption 2.2(2). □

Proof of Theorem 2:

Proof. We prove this by the method of negation. Assume that $\exists\, \delta_1, \epsilon_1 > 0$ such that

$$\mathrm{P}(\xi_N - \hat{\xi} > \delta_1) > \epsilon_1, \tag{27}$$

for any N, n. The monotonicity of $S(\gamma)$ implies that, for N large enough, $\forall\, \delta_1 > 0$, and $\gamma_2 > \gamma_1 + \delta_1$ where γ_1 and γ_2 are in a neighbourhood of ξ_N, $\exists\, N(\delta_1; \gamma_1, \gamma_2)$ such that $\forall\, N > N(\delta_1; \gamma_1, \gamma_2)$,

$$S_N(\gamma_2) - S_N(\gamma_1) > \delta_2$$

for some $\delta_2 > 0$. Together with (27), this implies that

$$\lim_{n\to\infty} \mathrm{P}(S_N(\xi_N) - S_N(\hat{\xi}) > \delta_2) \geq \epsilon_1. \tag{28}$$

It is easy to verify that $S_N(\xi_N) \to 0$ and (28) imply that, for some $\delta_2' > 0$,

$$\lim_{n\to\infty} \mathrm{P}(S_N(\hat{\xi}) < -\delta_2') \geq \epsilon_1. \tag{29}$$

Lemma 2 implies that

$$\lim_{n\to\infty} \mathrm{P}\left(\left|S_N(\hat{\xi}) - \hat{S}(\hat{\xi})\right| > \frac{\delta_2'}{2}\right) = 0. \tag{30}$$

Equations (29) and (30) imply that $\hat{S}(\hat{\xi}) < -\frac{\delta_2'}{2}$ with positive probability, contradicting the definition of $\hat{\xi}$. In a similar fashion, one can show that there does not exist $\delta_1, \epsilon_1 > 0$ such that

$$\mathrm{P}(\xi_N - \hat{\xi} < -\delta_1) > \epsilon_1,$$

for all N, n. Thus the proof of design consistency is completed. $\square$

Proof of Theorem 3:

Proof. Under Assumptions 2.1(2), 2.2, 4.1, 4.2 and assuming sample estimator $\hat{\xi}$ that is $\sqrt{n^*}$-consistent for ξ_N, we can obtain an asymptotic exapnsion similar to (8),

$$\hat{S}(\hat{\xi}) - S_N(\xi_N) = \left(\hat{S}(\xi_N) - S_N(\xi_N)\right) + S'(\xi_N)\left(\hat{\xi} - \xi_N\right) + o_p(n^{*-1/2}). \tag{31}$$

The smoothness condition of $S(\gamma)$ implies that $S_N(\xi_N) = O\left(\frac{1}{N}\right)$ and

$$\hat{S}(\hat{\xi}) = O_p\left(\frac{1}{N}\sup\left(\frac{1}{\pi_i}\right)\right) = O_p\left(\frac{1}{N}\frac{N}{n^*}\right) = O_p\left(\frac{1}{n^*}\right),$$

so that $(\hat{S}(\hat{\xi}) - S_N(\xi_N)) = o_p(n^{*-1/2})$. Then we divide by $S'(\xi_N)$ on both sides of (31) to obtain

$$\hat{\xi} = \xi_N - \frac{\hat{S}(\xi_N) - S_N(\xi_N)}{S'(\xi_N)} + o_p(n^{*-1/2}).$$

The asymptotic normality of $\hat{\xi}$ easily follows from the linear expansion above. □

Proof of Lemma A.1

Proof. Letting $X_i = h(\boldsymbol{y}_i; \boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - h(\boldsymbol{y}_i; \boldsymbol{\lambda}) - \mathcal{T}(\boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) + \mathcal{T}(\boldsymbol{\lambda})$, we need to show that

$$N^{\alpha} \left| \frac{1}{N} \sum_{i=1}^{N} X_i \right| \stackrel{a.s.}{\to} 0$$

uniformly for $\boldsymbol{\lambda} \in C_{\boldsymbol{\lambda}}$ and $\boldsymbol{s} \in C_{\boldsymbol{s}}$, a large enough compact set in $\Re^q$. Here, $\mathrm{E(X_i) = 0}$ and $\mathrm{E(h(\boldsymbol{y}_i; \boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - h(\boldsymbol{y}_i; \boldsymbol{\lambda})) = O(N^{-\alpha})}$. Without loss of generality, assume

$$\left| h(\boldsymbol{y}_i; \boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - h(\boldsymbol{y}_i; \boldsymbol{\lambda}) \right| \leq 1,$$

and

$$\mathrm{E} \left[h(\boldsymbol{y}_i; \boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - h(\boldsymbol{y}_i; \boldsymbol{\lambda}) \right]^2 \leq N^{-\alpha}.$$

Define the graph of $g_{\boldsymbol{\lambda},\boldsymbol{s}}(\boldsymbol{y}) \equiv h(\boldsymbol{y}_i; \boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - h(\boldsymbol{y}_i; \boldsymbol{\lambda})$ as

$$\begin{aligned} \mathrm{gr}(g_{\boldsymbol{\lambda},\boldsymbol{s}}) &= \{(\boldsymbol{y}, t) |\, 0 \leq t \leq g_{\boldsymbol{\lambda},\boldsymbol{s}}(\boldsymbol{y}) \,\mathrm{or}\, g_{\boldsymbol{\lambda},\boldsymbol{s}}(\boldsymbol{y}) \leq t \leq 0\} \\ &= \{(\boldsymbol{y}, t) |\, 0 \leq t \leq g_{\boldsymbol{\lambda},\boldsymbol{s}}(\boldsymbol{y})\} \cup \{(\boldsymbol{y}, t) |\, g_{\boldsymbol{\lambda},\boldsymbol{s}}(\boldsymbol{y}) \leq t \leq 0\}. \end{aligned}$$

Assumption A.2 implies that the set of functions

$$\left\{ (\boldsymbol{\lambda}, \boldsymbol{s}) \in \Re^q \times C_{\boldsymbol{s}} : h(\boldsymbol{y}; \boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - h(\boldsymbol{y}; \boldsymbol{\lambda}) \right\}$$

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 $\{(\boldsymbol{\lambda}, \boldsymbol{s}) \in C_{\boldsymbol{\lambda}} \times C_{\boldsymbol{s}} : h(\boldsymbol{y}; \boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - h(\boldsymbol{y}; \boldsymbol{\lambda})\}$ 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 that

$$\sup_{(\boldsymbol{\lambda},\boldsymbol{s}) \in C_{\boldsymbol{\lambda}} \times C_{\boldsymbol{s}}} N^{\alpha} \left| T_N(\boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) - T_N(\boldsymbol{\lambda}) - \mathcal{T}(\boldsymbol{\lambda} + N^{-\alpha}\boldsymbol{s}) + \mathcal{T}(\boldsymbol{\lambda}) \right| \stackrel{a.s.}{\to} 0,$$

and thus the result follows. □

Proof of Lemma A.2

Proof. We need to show that the set of functions

$$\{\psi(y-\gamma-n^{*-1/2}s)-\psi(y-\gamma)\,|\,(\gamma,s)\in C_\xi\times C_s\}$$

has polynomial discrimination, and observing that

$$n^{*1/2}S_N(\xi_N)\stackrel{a.s.}{\to}0,$$

then the rest of the proof continues as that for Lemma A.1. We define the following set of functions $g_{\gamma,s}(\cdot)$,

$$g_{\gamma,s}(y)=\psi(y-\gamma-n^{*-1/2}s)-\psi(y-\gamma),$$

and the graph of $g_{\gamma,s}$ is the subset

$$\mathrm{gr}(g_{\gamma,s})=\{(y,t)|\,0\le t\le g_{\gamma,s}(y),\mathrm{or}\,g_{\gamma,s}(y)\le t\le 0\}$$

of $\Re\times\Re$. The graphs in $\mathrm{gr}_+(g_{\gamma,s})=\{(y,t)|\,0\le t\le g_{\gamma,s}(y)\}$ can be partitioned into a finite number of mutually exclusive translational classes of sets, and each translational class has polynomial discrimination by Assumption A.4 and by Lemma B.1 of Opsomer and Ruppert (1997). Then the graphs in $\mathrm{gr}_+(g_{\gamma,s})$ have polynomial discrimination by Lemma II.15 of Pollard (1984). Similarly, we can argue that $\mathrm{gr}_-(g_{\gamma,s})=\{(y,t)|\,g_{\gamma,s}(y)\le t\le 0\}$ has polynomial discrimination. Thus we have shown that $\mathrm{gr}(g_{\gamma,s})$ has polynomial discrimination and the rest of the proof follows easily. □

Proof of Lemma A.3:

Proof. By triangle inequality,

$$\|\boldsymbol{\zeta}_N(\boldsymbol{\lambda})-\boldsymbol{\zeta}(\boldsymbol{\lambda})\|\le\|\boldsymbol{\zeta}_N(\boldsymbol{\lambda})-\mathrm{E}\boldsymbol{\zeta}_N(\boldsymbol{\lambda})\|+\|\mathrm{E}\boldsymbol{\zeta}_N(\boldsymbol{\lambda})-\boldsymbol{\zeta}(\boldsymbol{\lambda})\|\,,$$

where

$$\begin{aligned}
\mathrm{E}\boldsymbol{\zeta}_N(\boldsymbol{\lambda}) &= \frac{1}{Nb^q}\mathrm{E}\int\cdots\int h(\boldsymbol{y}_i;\boldsymbol{x})K_q'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)d\boldsymbol{x} \\
&= \frac{1}{Nb^q}\int\cdots\int \mathcal{T}(\boldsymbol{x})K_q'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)d\boldsymbol{x} \\
&= \boldsymbol{\zeta}(\boldsymbol{\lambda}) + o(1)
\end{aligned}$$

and the remainder converges to 0 uniformly for $\boldsymbol{\lambda} \in C_{\boldsymbol{\lambda}}$, so that $\sup_{\boldsymbol{\lambda}\in C_{\boldsymbol{\lambda}}}\|\mathrm{E}\boldsymbol{\zeta}_N(\boldsymbol{\lambda}) - \boldsymbol{\zeta}(\boldsymbol{\lambda})\| \to 0$.

It is left to show that $\sup_{\boldsymbol{\lambda}\in C_{\boldsymbol{\lambda}}}\|\boldsymbol{\zeta}_N(\boldsymbol{\lambda}) - \mathrm{E}\boldsymbol{\zeta}_N(\boldsymbol{\lambda})\| \overset{a.s.}{\to} 0$. Defining the quantity $g_{\boldsymbol{\lambda}}(\boldsymbol{y}) = \int\cdots\int h(\boldsymbol{y};\boldsymbol{x})K_q'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)d\boldsymbol{x}$, it is equivalent to show that

$$\sup_{\boldsymbol{\lambda}\in C_{\boldsymbol{\lambda}}}\left\|\tfrac{1}{N}\sum_{i=1}^{N} g_{\boldsymbol{\lambda}}(\boldsymbol{y}_i) - \mathrm{E}g_{\boldsymbol{\lambda}}(\boldsymbol{y}_i)\right\| \overset{a.s.}{\to} 0.$$

Assumption A.5 implies that

$$\begin{aligned}
g_{\boldsymbol{\lambda}}(\boldsymbol{y}) &= \int\cdots\int h(\boldsymbol{y};\boldsymbol{x})K_q'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)d\boldsymbol{x} \\
&= \int\cdots\int\left[\sum_{j=1}^{m_1} h_j(\boldsymbol{y};\boldsymbol{x})\right]\left[\sum_{j=1}^{m_2} K_{q,j}'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)\right]d\boldsymbol{x} \\
&= \sum_{j_1}\sum_{j_2}\int\cdots\int h_{j_1}(\boldsymbol{y};\boldsymbol{x})K_{q,j_2}'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)d\boldsymbol{x}.
\end{aligned}$$

Each $K_{q,j_2}'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)$ has no sign change by Assumption A.5 and $h_{j_1}(\boldsymbol{y};\boldsymbol{x})$ is monotone in $\boldsymbol{y}$ for every $\boldsymbol{x}$ by Assumption A.2, thus the integral of each $h_{j_1}(\boldsymbol{y};\boldsymbol{x})K_{q,j_2}'\left(\frac{\boldsymbol{\lambda}-\boldsymbol{x}}{b}\right)$ is monotone for any $\boldsymbol{\lambda}$. It is then possible to show that the graphs of $g_{\boldsymbol{\lambda}}(\boldsymbol{y})$ have polynomial discrimination by Lemma II.15 of Pollard (1984). Finally, we can again apply the results from empirical process theory to show the uniform convergence above, similarly to the proof of Lemma A.1. $\square$

References

Berger, Y. G. and C. J. Skinner (2003). Variance estimation for a low income proportion. *Journal of the Royal Statistical Society, Series C: Applied Statistics 52*(4), 457–468.

Binder, D. and M. Kovacevic (1995). Estimating some measures of income inequality from survey data: an application of the estimating equations approach. *Survey Methodology 21*, 137–145.

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

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

Chambers, R. L., A. H. Dorfman, and P. Hall (1992). Properties of estimators of the finite population distribution function. *Biometrika 79*, 577–582.

Da Silva, D. and J. Opsomer (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. and R. L. Chambers (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. and W. A. Fuller (1991). Quantile estimation with a complex survey design. *The Annals of Statistics 19*, 454–469.

Fuller, W. (2007). *Sampling Statistics*.

Fuller, W. A. and J. K. Kim (2005). Hot deck imputation for the response model. *Survey Methodology 31*(2), 139–149.

Godambe, V. and M. Thompson (2009). Estimating functions and survey sampling. In C. Rao and D. P. (editors) (Eds.), *Handbook of Statistics Volume 29: Sample Surveys: Inference and Analysis,*, pp. 669–687. Elsevier/North-Holland.

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

Isaki, C. and W. Fuller (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. and K. F. Hallock (2001). Quantile regression. *The Journal of Economic Perspectives 15*(4), 143–156.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. To appear in Springer Series in Statistics.

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

Opsomer, J. D. and D. Ruppert (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics 25*(1), 186–211.

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

Shao, J. and J. Rao (1993). Standard errors for low income proportions estimated from stratified multistage samples. *Sankhya B 55*, 393–414.

Sitter, R. R. and C. Wu (2001). A note on Woodruff confidence intervals for quantiles. *Statistics & Probability Letters 52*(4), 353–358.

Wang, J. and J. D. Opsomer (2008). Estimating the distance distribution of subpopulations for a large scale complex survey.

Wang, S. and A. H. Dorfman (1996). A new estimator for the finite population distribution function. *Biometrika 83*, 639–652.

Wu, C. and R. R. Sitter (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association 96*(453), 185–193(9).