01 Apr 2023

RANK-BASED INFERENCE FOR SURVEY SAMPLING DATA

Akim Adekpedjou
Missouri University of Science and Technology, akima@mst.edu

Huybrechts F. Bindele

Recommended Citation

A. Adekpedjou and H. F. Bindele, "RANK-BASED INFERENCE FOR SURVEY SAMPLING DATA," *Journal of Survey Statistics and Methodology*, vol. 11, no. 2, pp. 412 - 432, Oxford University Press; American Statistical Association, Apr 2023.
The definitive version is available at https://doi.org/10.1093/jssam/smab019

Journal of Survey Statistics and Methodology (2023) ***11***, 412–432

RANK-BASED INFERENCE FOR SURVEY SAMPLING DATA

AKIM ADEKPEDJOU*
HUYBRECHTS F. BINDELE

For regression models where data are obtained from sampling surveies, the statistical analysis is often based on approaches that are either non-robust or inefficient. The handling of survey data requires more appropriate techniques, as the classical methods usually result in biased and inefficient estimates of the underlying model parameters. This article is concerned with the development of a new approach of obtaining robust and efficient estimates of regression model parameters when dealing with survey sampling data. Asymptotic properties of such estimators are established under mild regularity conditions. To demonstrate the performance of the proposed method, Monte Carlo simulation experiments are carried out and show that the estimators obtained from the proposed methodology are robust and more efficient than many of those obtained from existing approaches, mainly if the survey data tend to result in residuals with heavy-tailed or skewed distributions and/or when there are few gross outliers. Finally, the proposed approach is illustrated with a real data example.

KEYWORDS: Rank estimator; Sampling; Weighting in survey.

1. INTRODUCTION

Data collected from surveys are often used to make inference about super population models, from which finite populations are assumed to be generated. Complex sample designs constructed from survey data require different statis-

Akim Adekpedjou is a Professor, Department of Mathematics and Statistics, Missouri University of Science and Technology, 400 W. 12th St., Rolla, MO 65409, USA. Huybrechts F. Bindele is an Associate Professor, Department of Mathematics and Statistics, University of South Alabama, 411 University Blvd. N, MSPB 316, Mobile, AL 36688, USA
*Address correspondence to Akim Adekpedjou, Department of Mathematics and Statistics, Missouri University of Science and Technology, 400 W. 12th St., Rolla, MO 65409, USA; E-mail: akima@mst.edu.

doi: 10.1093/jssam/smab019 Advance access publication 9 August 2021

tical methods to be analyzed than those developed classically under random sampling assumptions. An overview on this topic can be found in Chambers and Skinner (2003), Korn and Graubard (2011), and Steven, Heeringa, and West (2017). Regression analysis is a statistical tool that is key to describing the structural relationship that exists between survey variables. In sampling surveies, when sampling is related to the response variable of a regression analysis conditional on covariates, such sampling is called *informative sampling* and may lead to biased estimates in ordinary least squares estimation. To overcome this issue, weights are often introduced to obtain consistent estimators of the model parameters. Considering the sample inclusion indicator, such weights are often proportional to the reciprocal of the probability of including a unit in the sample. Parameter estimators are then obtained by minimizing the weighted least squares (LS) objective function or maximizing the quasi-likelihood objective function. Under similar framework, Kim and Skinner (2013) investigated two ways of defining such weights to improve the efficiency of their proposed estimators. They also explored two optimal ways of constructing weights by fitting of auxiliary weighted models. Furthermore, their approaches were extended to the pseudo maximum likelihood method for generalized linear models. It is worth pointing out that although weighting has a bias-correction advantage, it brings the disadvantage of often leading to a loss of efficiency relative to an unweighted approach. Moreover, weights also have the disadvantage of inflating the variance of the estimates. To address the variance inflation issue to further improve efficiency, Skinner and Mason (2012) modified surveys weights using Pfeffermann and Sverchkov (1999) ideas. Other weights were derived in Pfeffermann (1993) and Pfeffermann and Sverchkov (1999). Weights derived in Pfeffermann and Sverchkov (1999) were initially employed in linear regression and gained popularity in general regression. Beaumont (2008) and Magee (1998) proposed smoothing survey weights by considering a function that depends on the survey variables to address variance inflation issue. They also discussed how such a function may be chosen to achieve their goal. Several other authors have considered how the survey weights may be modified to improve efficiency of model parameters estimators under informative sampling. Magee (1998) showed how a weight function can be chosen to obtain a consistent estimator of the model parameter. Kim and Skinner (2013) proposed to estimate the regression parameter of the linear model using both the LS and the pseudo-likelihood approaches. It is worth pointing out that while many of the papers cited above used the LS approach, such an approach is known to result in non-robust and inefficient estimators when dealing with contaminated, heavy-tailed, and skewed model error distributions, and/or when data contain gross outliers. Alternative approaches include quantile regression, the likelihood-based methods, and other methods of moments. While quantile regression is robust, it comes with the drawback of selecting the correct quantile scale. As an example, when the quantile scale is set at $\tau = 0.5$, we have the least absolute deviation (LAD), which is known to be robust, but very often, results in inefficient estimators. On the other hand, likelihood-based methods are known to be vulnerable to model misspecification. For inference about regression parameters in the presence of informative sampling using likelihood-based methods, many contributions exist in the literature, which include Chambers (2003), Pfeffermann and Sverchkov (2009), Pfeffermann (2011), and Scott and Wild (2011) to cite a few. To overcome issues in LS and other methods of moments, in this article, we propose a rank-based approach that is robust and results in efficient estimators. Furthermore, the proposed approach has a simple geometric interpretation and does not require the specification of the model error distribution (Hettmansperger and McKean 2011).

Consider a finite population $\mathcal{V}$ of size N, where the units are labeled $i = 1, \ldots, N$ and let the row vector $(y_i, \mathbf{x}_i)$ denote the associated values of a pair of response y and explanatory variables $\mathbf{x}$ via some function $g(\cdot, \cdot)$ in a regression analysis settings. That is,

$$y_i = g(\mathbf{x}_i, \boldsymbol{\beta}_0) + \varepsilon_i, \quad i = 1, \ldots, N, \tag{1}$$

where $g(\cdot, \cdot)$ is a continuous regression function that can be linear or nonlinear, $\boldsymbol{\beta}_0 \in \boldsymbol{B}$ with $\boldsymbol{B}$ compact, $\mathbf{x}_i \in \mathbb{R}^p$, and ε_i are independent and identically distributed continuous random errors that have a zero mean and a positive variance bounded away from zero. The interest in this article is placed in a robust and efficient estimation of the true regression parameter $\boldsymbol{\beta}_0$ when data are obtained from a complex survey.

To this end, the remainder of the paper is organized as follows. Section 2 introduces our estimation approach. Asymptotic properties of the proposed estimator are discussed in section 3. The performance of our method is demonstrated from an extensive Monte Carlo simulation study and a real data example that are presented in section 4. The paper concludes with a short discussion. Technical details of the results stated in the paper are given in the appendix.

2. ESTIMATION

Consider the residuals $z_i(\boldsymbol{\beta}) = y_i - g(\mathbf{x}_i, \boldsymbol{\beta})$ and let $\delta_i = 1$, if unit i is in the sample, and $\delta_i = 0$, otherwise. We assume a probability sampling design in a sample s of size $n = \sum_{i=1}^{N} \delta_i$ with inclusion probability $\pi_i \equiv \pi(\mathbf{x}_i, y_i) = P(\delta_i | \mathbf{x}_i, y_i)$. Define the rank-based objective function as

$$D_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \delta_i d_i w(\mathbf{x}_i) \varphi\left(\frac{R(z_i(\boldsymbol{\beta}))}{N+1}\right) z_i(\boldsymbol{\beta}), \tag{2}$$

where d_i is a survey weight, $w(\mathbf{x})$ is a positive weight function introduced to take care of unusual observations (outliers, leverage points) in case they exist,

$\varphi : (0,1) \to \mathbb{R}$ is a continuous, nondecreasing and bounded score function and $R(t) = \sum_{j=1}^{N} (\delta_j/\pi_j) I\{z_j(\boldsymbol{\beta}) \leq t\}$. Note that for a sample s of size n selected from a finite population $\mathcal{V}$ of size N, as discussed in Arnab (2017), letting $F_{\boldsymbol{\beta}}(t) = \frac{1}{N}\sum_{i=1}^{N} I(z_i(\boldsymbol{\beta}) \leq t)$, an unbiased of $F_{\boldsymbol{\beta}}(t)$ is given by

$$\widehat{F}_{\boldsymbol{\beta}}(t) = \frac{1}{N}\sum_{i\in s}\frac{1}{\pi_i} I\{z_i(\boldsymbol{\beta}) \leq t\} = \frac{1}{N}\sum_{i=1}^{N}\frac{\delta_i}{\pi_i} I\{z_i(\boldsymbol{\beta}) \leq t\} = \frac{R(t)}{N}.$$

The good choice for $w(\mathbf{x})$ is the one that makes the influence function resulting from $D_N(\boldsymbol{\beta})$, bounded. The rank-based estimator, say $\widehat{\boldsymbol{\beta}}_N$, is a minimizer of $D_N(\boldsymbol{\beta})$. In (2), d_i is usually introduced to correct the estimator's bias that is due to the inclusion probability π_i. Let E_m and E_s be expectations with respect to the model and the sampling scheme, respectively. As in Skinner and Mason (2012), $D_N(\boldsymbol{\beta})$ results in unbiased estimating equation if

$$E_m\left(E_s\left[\delta_i d_i w(\mathbf{x}_i)\varphi\left(\frac{R(z_i(\boldsymbol{\beta}))}{N+1}\right)\nabla_{\boldsymbol{\beta}} g(\mathbf{x}_i, \boldsymbol{\beta})\right]\right) = 0, \tag{3}$$

where $\nabla_{\boldsymbol{\beta}}$ is the gradient operator. This occurs if, with assumption (J_1) given below, ε_i and $\mathbf{x}_i$ are independent, and the sampling scheme is non-informative, that is, I_i and y_i are independent conditional on $\mathbf{x}_i$. This could also arise when the sample inclusion depends only on a set of design variables. The non-informative sampling condition translates into $P(\delta_i = 1|\mathbf{x}_i, y_i) = P(\delta_i = 1|\mathbf{x}_i)$. This non-informative sampling condition can be tested as discussed in DuMouchel and Duncan (1983) and Fuller (2011). It is worth pointing out that throughout this paper, we do not assume a non-informative sampling design since assumption (J_3) given below ensures that the resulting estimating equation is asymptotically unbiased. As we will see later, the Horvitz–Thompson estimator is just the weighted LS estimator from a simple location model with $d_i = 1/\pi_i$. Depending on the scenarios, the sampling probability $\pi(\mathbf{x}, y)$ may be unknown, which in this case needs to be estimated. Some authors have considered estimating $\pi(\mathbf{x}, y)$ using the logistic linear regression model, which requires the specification of the functional form of $\pi(\mathbf{x}, y)$. This approach is often vulnerable to model identifiability. To overcome this issue, we propose a fully nonparametric estimation of π_i. To do so, as in Bindele and Adekpedjou (2018), consider the nonparametric estimator of $\pi(\mathbf{z})$ given by $\widehat{\pi}(\mathbf{z}) = \sum_{j=1}^{N} w_{Nj}(\mathbf{z})\delta_j$, where $w_{Nj}(\mathbf{z})$ is defined by

$$w_{Nj}(\mathbf{z}) = \frac{K[||\mathbf{z} - \mathbf{z}_j||/b_N]}{\sum_{j=1}^{N} K[||\mathbf{z} - \mathbf{z}_j||/b_N]}, \tag{4}$$

where $\mathbf{z} = (\mathbf{x}, y)$, K is a univariate real valued kernel function with bandwidth b_N satisfying $b_N \to 0$ and $Nb_N \to \infty$, as $n \to \infty$. For non-informative sampling, $\mathbf{z} = \mathbf{x}$. Under mild conditions, it can be shown that as $n \to \infty$, $\widehat{\pi}(\mathbf{z}) \to \pi(\mathbf{z})$ *a.s.* (see Einmahl and Mason 2005).

From model (1), under the simple random sampling, if $g(\cdot,\cdot)$ is a constant function, we obtain a location model, from which the LS estimation approach will produce the Horvitz–Thompson estimator. That is, considering the model $y_i = \mu + \varepsilon_i$, $i = 1, \ldots, N$, the Horvitz–Thompson estimator of μ is given by

$$\widehat{\mu}_{HT} = \underset{\mu}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} \delta_i d_i e_i^2(\mu), \quad \text{where} \quad e_i(\mu) = y_i - \mu.$$

A simple algebra manipulation gives $\widehat{\mu}_{HT} = \frac{1}{N}\sum_{i=1}^{N} \delta_i d_i y_i$. Now, considering the objective function (2) with $w(\mathbf{x}) \equiv 1$ and $\varphi(t) = \sqrt{12}(t - 1/2)$, we obtain

$$D_N(\mu) = \frac{\sqrt{12}}{N}\sum_{i=1}^{N} \delta_i d_i \left(\frac{R(e_i(\mu))}{N+1} - \frac{1}{2}\right) e_i(\mu).$$

For the linear model, $D_N(\mu)$ is location equivariant (Jaeckel 1972; Hettmansperger and McKean 2011). Thus, the rank-based estimator is obtained as $\widehat{\mu}_R = \text{Med}(u_1, \ldots, u_N)$, the median of $u_1, \ldots, u_N$, where $u_i = I_i d_i y_i$.

Remark 1. For stratified sampling, the model defined in (1), would be

$$y_{ij} = g(\mathbf{x}_{ij}, \boldsymbol{\beta}) + \varepsilon_{ij}, \quad i = 1, \ldots, N_j, j = 1, \ldots, m,$$

where N_{j} is the population size of stratum j. From this, the objective function in (2) needs to be modified. To be more precise, suppose Ω is the sample space of size n that can be partitioned into m sample strata, that is, $\Omega = \cup_{j=1}^{m} \Omega_j$ and $\Omega_i \cap \Omega_j = \varnothing$, for $i \neq j$. Let $\delta_{ij} = 1$, if unit $i \in \Omega_j$, and $\delta_{ij} = 0$, if $i \notin \Omega_j$. Then, $\pi_{ij} = \pi(\mathbf{x}_{ij}, y_{ij}) = P(\delta_{ij}|\mathbf{x}_{ij}, y_{ij})$ and $d_{ij} = 1/\pi_{ij}$. From this, setting $N = \sum_{j=1}^{m} N_j$, the objective function given in (2) can be redefined as

$$D_N(\boldsymbol{\beta}) = \sum_{j=1}^{m} \frac{1}{N_j} \sum_{i=1}^{N_j} \delta_{ij} d_{ij} w(\mathbf{x}_{ij}) \varphi\left(\frac{R(z_{ij}(\boldsymbol{\beta}))}{N_j + 1}\right) z_{ij}(\boldsymbol{\beta}),$$

where $z_{ij} = y_{ij} - g(\mathbf{x}_{ij}, \boldsymbol{\beta})$.

Now, throughout this paper, we consider the following assumptions:

- (J_1) $\varphi(\cdot)$ is twice continuously differentiable with bounded derivatives. Moreover, $\varphi(\cdot)$ can be standardized as $\int_0^1 \varphi(u)du = 0$ and $\int_0^1 \varphi^2(u)du = 1$.
- (J_2) $g(\cdot,\cdot)$ is a three times continuously differentiable bounded function of $\boldsymbol{\beta}$, and there exists an arbitrary function $H(\mathbf{x})$ satisfying

$$||\nabla_{\boldsymbol{\beta}}^{r} g(\mathbf{x}, \boldsymbol{\beta})|| \leq H(\mathbf{x}),$$
$$r = 0, 1, 2, 3 \text{ and } E[H^{\ell}(\mathbf{X})] < \infty \text{ with } 1 \leq \ell \leq 4.$$

(J_3) For every fixed N, $\boldsymbol{\beta}_{0N} = \text{argmin}_{\beta \in \boldsymbol{B}} E[D_N(\boldsymbol{\beta})]$ is the unique minimizer such that $\lim_{N\to\infty} \boldsymbol{\beta}_{0N} = \boldsymbol{\beta}_0$. Moreover, there exist c_1 and c_2 such that $0 < c_1 \leq \pi_i N n^{-1} < c_2$, for all i. Also, $n = O(N^{\alpha})$, $\alpha \in (0, 1)$.

Remark 2. Assumption (J_1) is a regular condition in the framework of rank-based estimation (see Hettmansperger and McKean 2011). Assumption (J_2) is necessary to ensure the consistency of $\widehat{\beta}_N$ in the next theorem. Assumption (J_3) is the identifiability condition that is assumed in most regression problems. Under this same assumption, the resulting estimating equation is asymptotically unbiased. In addition, (J_3) is also necessary in ensuring the $\sqrt{n}$-asymptotic normality distribution of the proposed estimator.

3. ASYMPTOTIC PROPERTIES

In this section, we discuss the asymptotic properties (consistency and asymptotic normality) of the proposed estimator. Recall that a random sample of size n is drawn from a finite population of size N and the asymptotic results are being derived for $n \to \infty$. Note that since $N \geq n$, as $n \to \infty$, $N \to \infty$.

Theorem 1 (Consistency). Under a probability sampling design in a random sample of size n and under assumption $(\boldsymbol{J}_1) - (\boldsymbol{J}_2)$, $\{\boldsymbol{D}_{\boldsymbol{N}}(\boldsymbol{\beta})\}_{\boldsymbol{N}\geq 1}$ is stochastically equicontinuous and $\lim_{\boldsymbol{N}\to\infty}[\boldsymbol{D}_{\boldsymbol{N}}(\boldsymbol{\beta}) - \boldsymbol{E}(\boldsymbol{D}_{\boldsymbol{N}}(\boldsymbol{\beta}))] = 0\, \boldsymbol{a.s.}$ Moreover, adding $(\boldsymbol{J}_3)$ to the two previous assumptions, as $\boldsymbol{N} \to \infty$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{N}} \to \boldsymbol{\beta}_0\, \boldsymbol{a.s.}$
Proof: See the appendix.
Let $\boldsymbol{S}_{\boldsymbol{N}}(\boldsymbol{\beta}) = -\nabla_{\boldsymbol{\beta}} \boldsymbol{D}_{\boldsymbol{N}}(\boldsymbol{\beta})$. That is,

$$S_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \delta_i d_i w(\mathbf{x}_i) \varphi\left(\frac{R(z_i(\boldsymbol{\beta}))}{N+1}\right) \nabla_{\boldsymbol{\beta}} g(\mathbf{x}_i, \boldsymbol{\beta}).$$

Since $\widehat{\boldsymbol{\beta}}_N = \text{argmin}_{\boldsymbol{\beta} \in \boldsymbol{B}} D_N(\boldsymbol{\beta})$, then $S_N(\widehat{\boldsymbol{\beta}}_N) = 0$. Define $T_N(\boldsymbol{\beta})$ by

$$T_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \delta_i d_i w(\mathbf{x}_i) \varphi(F_{i\boldsymbol{\beta}}(z_i(\boldsymbol{\beta}))) \nabla_{\boldsymbol{\beta}} g(\mathbf{x}_i, \boldsymbol{\beta}),$$

where $F_{i\boldsymbol{\beta}}(t) = P(z_i(\boldsymbol{\beta}) \leq t)$. The following theorem pertains to the asymptotic equivalence of $S_N(\boldsymbol{\beta})$ and $T_N(\boldsymbol{\beta})$.

Theorem 2. Under a probability sampling design in a random sample of size n and under $(\boldsymbol{J}_1)$ and $(\boldsymbol{J}_2)$, $\lim_{\boldsymbol{n}\to\infty} \sup_{\boldsymbol{\beta}\in\boldsymbol{B}} ||\boldsymbol{T}_N(\boldsymbol{\beta}) - \boldsymbol{S}_N(\boldsymbol{\beta})|| = 0$, $\boldsymbol{a.s}$.

Thus, $S_N(\boldsymbol{\beta}) = T_N(\boldsymbol{\beta}) + o_p(1/n)$. On the other hand, it can be shown in a straightforward manner (Bindele, Abebe, and Meyer 2018) that for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\Big(\sqrt{n}||S_N(\boldsymbol{\beta}_0) - T_N(\boldsymbol{\beta}_0)|| > \epsilon\Big) = 0,$$

which implies that $\sqrt{n}S_N(\boldsymbol{\beta}_0)$ and $\sqrt{n}T_N(\boldsymbol{\beta}_0)$ have the same asymptotic distribution. Now, applying a Taylor expansion to $T_N(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$, we have

$$T_N(\boldsymbol{\beta}) = T_N(\boldsymbol{\beta}_0) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\nabla_{\boldsymbol{\beta}} T_N(\boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\nabla^2_{\boldsymbol{\beta}} T_N(\boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

where $\boldsymbol{\beta}^*$ lies on the line segment joining $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$, and $\nabla^2_{\boldsymbol{\beta}}$ being the Laplacian operator. Then,

$$S_N(\boldsymbol{\beta}) = T_N(\boldsymbol{\beta}_0) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\nabla_{\boldsymbol{\beta}} T_N(\boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\nabla^2_{\boldsymbol{\beta}} T_N(\boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1/n), \quad (5)$$

From this, we have the following theorem.

Theorem 3. Under a probability sampling design in a random sample of size n and under assumptions $(\boldsymbol{J}_1) - (\boldsymbol{J}_3)$, we have that as $\boldsymbol{n} \to \infty$,

(i) $E(T_N(\boldsymbol{\beta}_0)) \to 0$,
(ii) $\nabla_{\boldsymbol{\beta}} T_N(\boldsymbol{\beta}_0) \to \Gamma(\boldsymbol{\beta}_0)$ *a.s., where* $\Gamma(\boldsymbol{\beta}_0)$ *is a* $p \times p$ *matrix defined as*

$$\Gamma(\beta_0) := -E\{\gamma(\mathbf{X})[\nabla_\beta g(\mathbf{X}, \beta_0)][\nabla_\beta g(\mathbf{X}, \beta_0)]' f(\varepsilon)\varphi'(F(\varepsilon))\} + E\{\gamma(\mathbf{X})[\nabla^2_\beta g(\mathbf{X}, \beta_0)]\varphi(F(\varepsilon))\},$$

with $\gamma(\mathbf{x}_i) = \delta_i d_i w(\mathbf{x}_i)$. *Furthermore,* $\nabla^2_{\boldsymbol{\beta}} T_N(\boldsymbol{\beta}^*)$ *is bounded in probability, and*
(iii) $\sqrt{n}T_N(\boldsymbol{\beta}_0) \to \mathcal{N}_p(0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0})$, *where* $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0} = \lim_{n\to\infty} \mathrm{Var}\Big(\sqrt{n}T_N(\boldsymbol{\beta}_0)\Big)$.

The next theorem provides the asymptotic normality property of the proposed estimator. In addition to the previous assumptions, consider the following assumption, which is necessary to ensure the invertibility of $\Gamma(\boldsymbol{\beta}_0)$.

(J_4) $\Gamma(\boldsymbol{\beta}_0)$ is positive definite.

This assumption could be relaxed by considering its generalized inverse.

Theorem 4. Under a probability sampling design in a rample sample of size n and under assumptions $(\boldsymbol{J}_1) - (\boldsymbol{J}_4)$, as $\boldsymbol{n} \to \infty$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \to \mathcal{N}_p(0, \Gamma^{-1}(\boldsymbol{\beta}_0)\Sigma_{\boldsymbol{\beta}_0}\Gamma^{-1}(\boldsymbol{\beta}_0)).$$

Remark 3. For stratified sampling, these asymptotic properties with slight modification of asymptotic expressions will be valid under the additional assumption that the strata are pairwise independent. For further statistical inference about $\boldsymbol{\beta}_0$, the estimation of the covariance matrix $\Gamma^{-1}(\boldsymbol{\beta}_0)\Sigma_{\boldsymbol{\beta}_0}$ $\Gamma^{-1}(\boldsymbol{\beta}_0)$ is necessary. From the complex form of this covariance matrix, its estimation can be done by considering sandwitch type estimators (Brunner and Denker 1994; Bindele 2015). This can be done as follows: for a given random sample of size n, setting $\widehat{\boldsymbol{e}}_{\boldsymbol{in}} = \boldsymbol{z_i}(\widehat{\boldsymbol{\beta}}_N)$, $\lambda_i = \gamma(\mathbf{x}_i)\nabla_{\boldsymbol{\beta}}\boldsymbol{g}(\mathbf{x}_i, \boldsymbol{\beta}_0)$ and $\widehat{\lambda}_i = \gamma(\mathbf{x}_i)\nabla_{\boldsymbol{\beta}}\boldsymbol{g}(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_N)$, define

$$\widehat{A}_n = \frac{1}{n}\sum_{i=1}^{n}\widehat{\lambda}_i\varphi'\left(\frac{R(\widehat{e}_{in})}{n+1}\right)R(\widehat{e}_{in}) \quad \text{and}$$

$$A_N(\beta_0) = N\int\varphi(J_N(t))\widehat{F}_n(dt) + \int\varphi'(J_N(t))\widehat{J}_{jn}(t)F_N(dt),$$

from which putting $F_i(t) = P(z_i(\boldsymbol{\beta}_0) \le t)$,

$$J_N(s) = \frac{1}{N}\sum_{i=1}^{N}F_i(s) \quad \text{and} \quad \widehat{J}_n(s) = \frac{1}{n}\sum_{i=1}^{n}I(z_i(\beta_0) \le s),$$

$$F_N(s) = \frac{1}{N}\sum_{i=1}^{N}\lambda_i F_i(s) \quad \text{and} \quad \widehat{F}_n = \frac{1}{n}\sum_{i=1}^{n}\lambda_i I(z_i(\beta_0) \le s).$$

From this, following theorem 4.1 of Brunner and Denker (1994) and setting

$$\widehat{\boldsymbol{\Sigma}} = [\widehat{A}_n - E(A_N(\boldsymbol{\beta}))][\widehat{A}_n - E(A_N(\boldsymbol{\beta}))]',$$

under mild conditions, $\widehat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}$ are asymptotically equivalent in L^2-norm. Moreover, Bindele (2015) showed how $A_N(\boldsymbol{\beta}_0)$ can be approximated by a Riemann sum. On the other hand, $\Gamma(\boldsymbol{\beta}_0)$ can be approximated by

$$\widehat{\Gamma}_n = -\frac{1}{n}\sum_{i=1}^{n}\gamma(\mathbf{x}_i)[\nabla_\beta g(\mathbf{x}_i, \widehat{\beta}_N)][\nabla_\beta g(\mathbf{x}_i, \widehat{\beta}_N)]'\varphi'\left(\frac{R(\widehat{e}_{in})}{n+1}\right)R(\widehat{e}_{in})$$
$$+\frac{1}{n}\sum_{i=1}^{n}\gamma(\mathbf{x}_i)[\nabla^2_\beta g(\mathbf{x}_i, \widehat{\beta}_N)]\varphi\left(\frac{R(\widehat{e}_{in})}{n+1}\right).$$

To this end, $\Gamma^{-1}(\boldsymbol{\beta}_0)\Sigma_{\boldsymbol{\beta}_0}\Gamma^{-1}(\boldsymbol{\beta}_0)$ can be estimated by $\widehat{\Gamma}_n^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\Gamma}_n^{-1}$.

When a nonparametric estimation of $\pi(\mathbf{z})$ as discussed earlier is involved, the bandwidth needs to be carefully selected. To do so, a leave-one-out cross validation can be considered. That is, $\widehat{b}_N$ can be chosen as

$$\widehat{b}_N = \underset{b_N}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} \delta_i \widehat{d}_i w(\mathbf{x}_i)\varphi\left(\frac{R(z_{-i}(\widehat{\boldsymbol{\beta}}_N))}{N+1}\right) z_{-i}(\widehat{\boldsymbol{\beta}}_N),$$

where $z_{-i}(\widehat{\boldsymbol{\beta}}_N)$ is the leave-one-out version of $z_i(\widehat{\boldsymbol{\beta}}_N)$ and $\widehat{d}_i = 1/\widehat{\pi}(\mathbf{x}_i, y_i)$ with $\widehat{\pi}(\mathbf{x}_i, y_i)$ defined before (4).

4. SIMULATION AND REAL DATA APPLICATION

4.1 Simulation

To assess the performance of the proposed approach, we performed a simulation study for the basic linear regression model using the same settings as in Kim and Skinner (2013), for a direct comparison. That is, we repeatedly generated 2,000 finite populations of size $N = 5{,}000$ with values (x_i, y_i, z_i, π_i), $(i = 1, \ldots, N)$, where $x_i = 0.5 + \tilde{x}_i$, with $\tilde{x}_i$ generated from an exponential distribution with mean 1, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $(\beta_0, \beta_1) = (-2, 1)$ and, ε_i is generated from four different distributions: $\varepsilon_i \sim N(0, 0.8)$, $\varepsilon_i \sim N(0, 0.5^2 x_i^2)$ as in Kim and Skinner (2013), $\varepsilon_i \sim t_3$, the t-distribution with three degree of freedom and $\varepsilon_i \sim CN(0.10, 10) = 0.9N(0, 1) + 0.1N(0, 10^2)$. These last two cases are included to study the effect of tail thickness and contamination on the resulting estimators. The inclusion probabilities were generated as $\pi_i = nk_i / \sum_{j=1}^{N} k_j$, where $n = 100$, $k_i = \{1 + \exp(2.5 - 0.5z_i)\}^{-1}$ and $z_i \sim N(1 + y_i, 0.52)$. From each of the finite populations generated, independent samples were drawn by Poisson sampling, where the sample indicator I_i followed a $Ber(\pi_i)$ distribution. From each sample, estimators of (β_0, β_1) are obtained under different sampling weights as discussed in Kim and Skinner (2013). From 2,000 replications, biases and standard errors of the rank and LS estimators are calculated. Tables 1 and 2 below display the results of our simulation study. Moreover, to evaluate the performance of the proposed sandwich variance estimator, under the considered model error distributions, 95 percent coverage probabilities are computed and results are displayed in table 3.

From tables 1 and 2, it is observed that the proposed rank-based method outperforms the LS method when the model error is heteroscedastic, contaminated or heavy-tailed. Under $N(0, 0.8)$ however, as would be expected, the LS performs better than the proposed method. This is not surprising since the LS is known to be superior than any other method when $\varepsilon \sim N(0, 1)$ (see tables 1 and 2). It is worth pointing out that our simulation results corroborate with those of Kim and Skinner (2013) (see table 1). It is also noted that while all weights have slightly similar performance, the fully nonparametric sampling weight has the advantage of not requiring the knowledge of the functional of the sampling inclusion probability.

Table 1. Biases ×10 (SEs ×10) of the Regression Parameter Estimates for Different Weights Schemes When $\varepsilon \sim N(0, 0.8)$ and $\varepsilon \sim N(0, 0.5^2 x_i^2)$

Method	Estimates	Weights	$N(0, 0.8)$	$N(0, 0.5^2 x_i^2)$
LS		Design	0.15 (1.84)	0.00 (1.29)
		Pfefferman–Sverchkov	0.08 (1.67)	−0.13 (1.28)
		Unsmoothed optimal	0.14 (1.65)	−0.02 (1.26)
	$\widehat{\beta}_0$	Smoothed design	0.20 (1.82)	−0.05 (1.25)
		Smoothed Pfefferman–Sverchkov	0.17 (1.64)	−0.15 (1.24)
		Smoothed optimal	0.15 (1.62)	−0.06 (1.23)
		Nonparametric	0.18 (1.67)	−0.10 (1.26)
		Design	−0.05 (0.78)	0.04 (1.25)
		Pfefferman–Sverchkov	−0.02 (0.64)	0.15 (1.21)
		Unsmoothed optimal	−0.02 (0.65)	0.06 (1.20)
	$\widehat{\beta}_1$	Smoothed design	−0.08 (0.76)	0.17 (1.20)
		Smoothed Pfefferman–Sverchkov	−0.05 (0.62)	0.25 (1.19)
		Smoothed optimal	−0.05 (0.62)	0.17 (1.21)
		Nonparametric	−0.06 (0.65)	0.20 (1.24)
Rank		Design	0.17 (1.89)	0.01 (1.11)
		Pfefferman–Sverchkov	0.11 (1.72)	−0.15 (1.08)
		Unsmoothed optimal	0.16 (1.68)	−0.03 (1.07)
	$\widehat{\beta}_0$	Smoothed design	0.21 (1.85)	−0.07 (1.07)
		Smoothed Pfefferman–Sverchkov	0.19 (1.66)	−0.17 (1.08)
		Smoothed optimal	0.18 (1.65)	−0.09 (1.06)
		Nonparametric	0.21 (1.70)	−0.13 (1.08)
		Design	−0.07 (0.81)	0.06 (1.08)
		Pfefferman–Sverchkov	−0.04 (0.67)	0.18 (1.05)
		Unsmoothed optimal	−0.05 (0.69)	0.07 (1.01)
	$\widehat{\beta}_1$	Smoothed design	−0.10 (0.79)	0.19 (1.03)
		Smoothed Pfefferman–Sverchkov	−0.06 (0.64)	0.27 (1.01)
		Smoothed optimal	−0.07 (0.65)	0.19 (1.02)
		Nonparametric	−0.08 (0.67)	0.21 (1.06)

When it comes to the variance estimator, it is observed that the proposed sandwich variance estimator performs fairly well as the coverage probabilities under different model error distributions are close to the nominal level (see table 3).

4.2 Real Data

To illustrate our methodology, we consider the dataset on the population biology of Abalone (Dua and Graff, 2017). The interest in this dataset is in predicting the age of Abalone from physical measurements. The age of abalone is

Table 2. Biases ×10 (SEs ×10) of the Regression Parameter Estimates for Different Weight Schemes under t_3 and $CN(0.10, 10)$

Method	Estimates	Weights	t_3	$CN(0.10, 10)$
LS		Design	0.97 (2.95)	0.94 (2.36)
		Pfefferman–Sverchkov	0.86 (2.69)	−0.93 (2.24)
		Unsmoothed optimal	0.92 (2.35)	−0.89 (2.17)
	$\widehat{\beta}_0$	Smoothed design	0.99 (2.67)	−0.97 (2.31)
		Smoothed Pfefferman–Sverchkov	0.94 (2.21)	−0.93 (2.27)
		Smoothed optimal	0.88 (2.41)	−0.91 (2.26)
		Nonparametric	0.95 (2.48)	−0.93 (2.36)
		Design	−0.78 (1.77)	0.76 (2.14)
		Pfefferman–Sverchkov	−0.71 (1.72)	0.87 (2.24)
		Unsmoothed optimal	−0.68 (1.75)	0.76 (2.18)
	$\widehat{\beta}_1$	Smoothed design	−0.79 (1.83)	0.81 (2.19)
		Smoothed Pfefferman–Sverchkov	−0.73 (1.66)	0.89 (2.16)
		Smoothed optimal	−0.72 (1.64)	0.87 (2.20)
		Nonparametric	−0.75 (1.73)	0.90 (2.27)
Rank		Design	0.13 (1.77)	0.00 (1.28)
		Pfefferman–Sverchkov	0.08 (1.67)	−0.11 (1.26)
		Unsmoothed optimal	0.12 (1.49)	−0.01 (1.23)
	$\widehat{\beta}_0$	Smoothed design	0.18 (1.78)	−0.03 (1.21)
		Smoothed Pfefferman–Sverchkov	0.15 (1.53)	−0.13 (1.19)
		Smoothed optimal	0.12 (1.56)	−0.04 (1.17)
		Nonparametric	0.15 (1.59)	−0.07 (1.21)
		Design	−0.03 (0.69)	0.02 (1.19)
		Pfefferman–Sverchkov	−0.01 (0.58)	0.11 (1.16)
		Unsmoothed optimal	−0.01 (0.55)	0.04 (1.14)
	$\widehat{\beta}_1$	Smoothed design	−0.05 (0.67)	0.14 (1.16)
		Smoothed Pfefferman–Sverchkov	−0.02 (0.53)	0.19 (1.12)
		Smoothed optimal	−0.02 (0.54)	0.14 (1.18)
		Nonparametric	−0.03 (0.58)	0.16 (1.19)

determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope—a boring and time-consuming task. Other measurements, which are easier to obtain, may also be used to predict the age. Further information, such as weather patterns and location (hence food availability) may also be required to solve the problem. For this data $N = 4{,}177$ and from this population, we draw a simple random sample of size $n = 500$. The data contain eight predictors that are: x_1 = "Sex," where $x_1 = 1$, if male and $x_1 = 0$, otherwise, x_2 = "Length," x_3 = "Diameter," x_4 = "Height," x_5 = "Whole weight," x_6 = "Shucked weight," x_7 = "Viscera weight," and x_8 = "Shell weight." The outcome of interest is y = "the number of rings that gives

Table 3. 95 Percent Coverage Probabilities (CP) of the Sandwich Variance Estimator, as Discussed in Remark 3 for Different Model Error (ε) Distributions

E	$N(0, 0.8)$	$N(0, 0.5^2 x_i^2)$	t_3	$CN(0.10, 10)$
CP	0.939	0.935	0.941	0.938

the age of the Abalone." The sampling inclusion probabilities are obtained in two ways:

- **Setup 1:** First, as in our simulation study by setting $\pi_i = nk_i / \sum_{i=1}^{N} k_i$, k is estimated using a linear logistic regression model, with the outcome being either the unit is included in the sample or not. That is,

$$\frac{k = P(I = 1|\mathbf{x}, y) = \exp\{\alpha_0 + \mathbf{x}^\tau \boldsymbol{\alpha} + \gamma y\}}{1 + \exp\{\alpha_0 + \mathbf{x}^\tau \boldsymbol{\alpha} + \gamma y\},}$$

  where $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_8)'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_8)'$ and, $I = 1$, if the unit is included in the sample and, $I = 0$, if not. From the linear logistic regression analysis, the sampling weight is estimated as $\widehat{d}_i = 1/\widehat{\pi}_i$, where $\widehat{\pi}_i = n\widehat{k}_i / \sum_{i=1}^{N} \widehat{k}_i$, with

$$\widehat{k}_i = \frac{\exp\{\widehat{\alpha}_0 + \mathbf{x}_i^\tau \widehat{\boldsymbol{\alpha}} + \widehat{\gamma} y_i\}}{1 + \exp\{\widehat{\alpha}_0 + \mathbf{x}_i^\tau \widehat{\boldsymbol{\alpha}} + \widehat{\gamma} y_i\}},$$

$\widehat{\alpha}_0 = -0.82$, $\widehat{\boldsymbol{\alpha}} = (0.14, 1.98, -3.90, -2.30, -0.72, 0.31, 2.49, 1.30)'$ and $\widehat{\gamma} = 0.02$.

- **Setup 2:** Second, we consider a fully nonparametric estimation of the inclusion probability, as discussed earlier by setting $K(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z^2\right\}$.

The regression analysis results are displayed in table 4.

From figure 1, it is clearly observed that the response is right skewed with multiple outliers, which suggest that the LS may not be suitable for this data. On the other hand the LAD, which is equivalent to the quantile regression with quantile scale $\tau = 0.5$, may be an alternative choice but is known to result in inefficient estimators. From table 4, we observe that the three different approaches give different estimates. But the proposed method is more efficient as it gives estimates with smaller SEs. Our method appears most trustworthy, as it does not require the knowledge of the model error distribution. As observed from our simulation study, we see that when the sampling weights are estimated using the linear logistic regression model, all approaches give better estimates than those obtained when considering a fully nonparametric estimation of the sampling weights. However, the fully nonparametric sampling weights under the proposed method still outperform the LS and LAD.

Table 4. Estimates (Est.) and Standard Errors (SEs) of the LS, LAD, and Rank Estimators

		LS		LAD		Rank	
Weight	Variables	Est.	SE	Est.	SE	Est.	SE
Setup 1	x_1	0.72	0.21	0.47	0.19	0.68	0.08
	x_2	−3.42	4.02	−2.43	3.56	1.45	1.42
	x_3	12.46	5.00	6.03	4.62	8.06	1.75
	x_4	25.70	4.61	22.41	4.43	14.21	1.21
	x_5	13.66	1.61	11.53	1.90	7.86	0.57
	x_6	−24.36	1.85	−19.98	1.92	−17.28	0.64
	x_7	−15.89	2.70	−15.40	2.71	−9.60	1.02
	x_8	1.34	2.33	2.03	2.37	7.45	0.88
Setup 2	x_1	0.81	0.31	0.59	0.25	0.73	0.11
	x_2	−3.63	4.81	−2.65	3.89	1.51	1.63
	x_3	12.71	5.76	6.37	4.95	8.12	1.88
	x_4	25.92	5.09	22.67	4.92	14.27	1.32
	x_5	13.88	2.57	11.74	2.15	7.49	0.71
	x_6	−24.59	2.38	−20.09	2.22	−17.36	0.78
	x_7	−16.10	3.07	−15.63	2.98	−9.73	1.14
	x_8	1.47	2.85	2.17	2.68	7.54	0.96

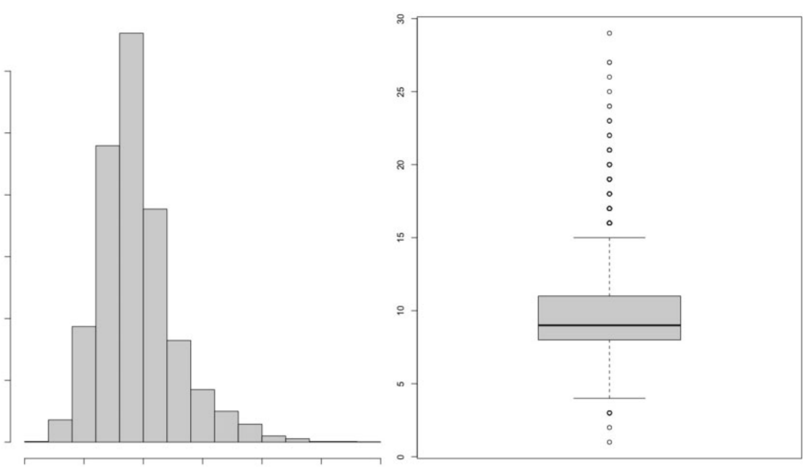

Figure 1. Histogram and Boxplot of the Number of Rings.

5. DISCUSSION

This article is concerned with a robust and efficient estimation of the regression parameters when data are collected from complex survey. Asymptotic properties of the proposed estimator are established under mild regularity conditions. A simulation study has demonstrated that under heteroscedastic, contaminated, or heavy-tailed model error distribution, the proposed approach outperforms the LS in terms of robustness and efficiency. It is also seen that while several weights exist to correct the bias introduced by the sampling scheme, most of them usually require the specification of the functional form of the sampling probability. Moreover, the fully nonparametric estimator of the sampling weight proposed in this article performs well and has the advantage of not requiring any functional form specification.

ACKNOWLEDGMENT

We thank the anonymous reviewers and the Associate Editor for their suggestions and comments that helped to greatly improve the paper.

APPENDIX

This appendix contains technical details of the theoretical results stated through this paper.

PROOFS

Proof of theorem 1. Recall that $F_{i\boldsymbol{\beta}}(z) = P(z_i(\boldsymbol{\beta}) \leq z)$. Note that if $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, $z_i(\boldsymbol{\beta})$ are independent but not necessarily identically distributed. Recall that for $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $z_i(\boldsymbol{\beta}_0) \equiv \varepsilon_i$ and $F_{i\boldsymbol{\beta}_0}(\cdot) \equiv F(\cdot)$. Set $a_{iN}(\boldsymbol{\beta}) = R(z_i(\boldsymbol{\beta}))/(N+1)$. For any $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \boldsymbol{B}$,

$$\begin{aligned}
|D_N(\beta_1) - D_N(\beta_2)| &\leq |\frac{1}{N}\sum_{i=1}^{N}\delta_i d_i w(\mathbf{x}_i)[\varphi(a_{iN}(\beta_1)) - \varphi(F_{i\beta_1}(z_i(\beta_1)))]| \\
&+|\frac{1}{N}\sum_{i=1}^{N}\delta_i d_i w(\mathbf{x}_i)[\varphi(F_{i\beta_1}(z_i(\beta_1))) - \varphi(F_{i\beta_2}(z_i(\beta_2)))]| \\
&+|\frac{1}{N}\sum_{i=1}^{N}\delta_i d_i w(\mathbf{x}_i)[\varphi(a_{iN}(\beta_2)) - \varphi(F_{i\beta_2}(z_i(\beta_2)))]| \\
&= I_{1N} + I_{2N} + I_{3N}.
\end{aligned}$$

Considering I_{1N} and applying the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
I_{1N} &\leq \frac{1}{N}\sum_{i=1}^{N}\delta_i d_i w(\mathbf{x}_i)|\varphi(a_{iN}(\beta_1))) - \varphi(F_{i\beta_1}(z_i(\beta_1)))| \\
&\leq \left[\frac{1}{N}\sum_{i=1}^{N}\delta_i d_i^2 w^2(\mathbf{x}_i)\right]^{1/2} [\max_{1\leq i\leq N}|\varphi(a_{iN}(\beta_1)) - \varphi(F_{i\beta_1}(z_i(\beta_1)))|^2]^{1/2}.
\end{aligned}
$$

Note that $U_i = F_{i\boldsymbol{\beta}}(z_i(\boldsymbol{\beta}))$ are independent and uniformly distributed in (0, 1). Following Hájek, Šidák, and Sen (1999), it is obtained that as $N \to \infty$, $a_{iN}(\boldsymbol{\beta}) - U_i \to 0\, a.s.$, for all i and for all $\boldsymbol{\beta} \in \boldsymbol{B}$. From the continuity of $\varphi(\cdot)$, applying the generalized continuous mapping theorem (Whitt 2002), we have $\varphi(a_{iN}(\boldsymbol{\beta})) - \varphi(U_i) = o(1/N)\, a.s.$, for all i and for all $\boldsymbol{\beta} \in \boldsymbol{B}$. Moreover, with $\{y_i, \mathbf{x}_i, I_i\}_{i=1}^N$ being a random sample, $\delta_i d_i^2 w^2(\mathbf{x}_i)$, $i = 1, \ldots,$ N are independent. So, from the Strong Law of Large Numbers (SLLN), we have that as $N \to \infty$,

$$
\frac{1}{N}\sum_{i=1}^{N}[\delta_i d_i^2 w^2(\mathbf{x}_i) - E(\delta_i d_i^2 w^2(\mathbf{X}_i))] \to 0\, a.s. \quad \text{and} \quad E(\delta_i d_i^2 w^2(\mathbf{X}_i)) < \infty \quad \text{by}(J_2) - (J_3).
$$

Thus, as $N \to \infty$, $I_{1N} \to 0\, a.s.$ and $I_{3N} \to 0\, a.s.$ Note that $F_{i\boldsymbol{\beta}}(z)$ is continuous and almost surely differentiable. Applying the mean value theorem to $\varphi(F_{i\boldsymbol{\beta}}(z_i(\boldsymbol{\beta}))$, there exists $\boldsymbol{\beta}^*$ lying on the line segment joining $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ such that

$$
\begin{aligned}
&\varphi\Big(F_{\beta_1}(z_i(\beta_1))\Big) - \varphi\Big(F_{\beta_2}(z_i(\beta_2))\Big) \\
&= \nabla_\beta g(\mathbf{x}_i, \beta^*)\varphi'(F_{i\beta^*}(z_i(\beta^*)))f_{i\beta^*}(z_i(\beta^*))(\beta_1 - \beta_2),
\end{aligned}
$$

where $f_{i\boldsymbol{\beta}}(\cdot)$ is the Radon–Nykodim derivative of $F_{i\boldsymbol{\beta}}(\cdot)$. Since $f_{i\boldsymbol{\beta}}(\cdot)$ is a probability density function, it is almost surely bounded. From the fact that $\varphi'(x)$ is bounded, so is $\varphi'(F(x))$. Therefore, there exists a positive constant c such that $\varphi'(F_{i\boldsymbol{\beta}^*}(z_i(\boldsymbol{\beta}^*)))f_{i\boldsymbol{\beta}^*}(z_i(\boldsymbol{\beta}^*)) \leq c\, a.s.$ Thus, with probability 1, by (J_2) and the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
I_{2N} &= |\frac{1}{N}\sum_{i=1}^{N}\delta_i d_i w(\mathbf{x}_i)[\nabla_\beta g(\mathbf{x}_i, \beta^*)\varphi'(F_{i\beta^*}(z_i(\beta^*)))f_{i\beta^*}(z_i(\beta^*))](\beta_1 - \beta_2)| \\
&\leq \left[\frac{c}{N}\sum_{i=1}^{N}d_i w(\mathbf{x}_i)||\nabla_\beta g(\mathbf{x}_i, \beta)||\right]||\beta_1 - \beta_2|| \\
&\leq c\left(\frac{1}{N}\sum_{i=1}^{N}d_i^2 w^2(\mathbf{x}_i)\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}H^2(\mathbf{x}_i)\right)^{1/2}||\beta_1 - \beta_2||.
\end{aligned}
$$

A direct application of the SLLN gives

$$A_N := c\left(\frac{1}{N}\sum_{i=1}^{N} d_i^2 w^2(\mathbf{x}_i)\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N} H^2(\mathbf{x}_i)\right)^{1/2}$$
$$\to c\Big(E(d^2w^2(\mathbf{X}))\Big)^{1/2}\Big(E(H^2(\mathbf{X}))\Big)^{1/2}\ a.s.,$$

by $(J_1)-(J_2)$, with $c\Big(E(d^2w^2(\mathbf{X}))\Big)^{1/2}\Big(E(H^2(\mathbf{X}))\Big)^{1/2} < \infty$. This ensures that A_N is almost surely bounded. Thus, with probability 1,

$$|D_N(\boldsymbol{\beta}_1) - D_N(\boldsymbol{\beta}_2)| \le A_N||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||.$$

Hence, $\{D_N(\boldsymbol{\beta})\}_{N\ge 1}$ is stochastically equicontinuous.

Under assumptions (J_1) and (J_2), it is obtained that $E(D_N(\boldsymbol{\beta})) < \infty$. Also, $\varphi(\cdot)$ and $g(\cdot)$ being continuous functions of $\boldsymbol{\beta}$, so is $E(D_N(\boldsymbol{\beta}))$. Thus, since $\boldsymbol{B}$ is compact, under (J_3), $\boldsymbol{\beta}_0$ exists. Set $\bar{F}_{\boldsymbol{\beta}}(\cdot) = \frac{1}{N}\sum_{i=1}^{N} F_{i\boldsymbol{\beta}}(\cdot)$, $V_i(t) = I\{z_i(\boldsymbol{\beta}) \le t\}$, $\gamma(\mathbf{x}_i) = \delta_i d_i w(\mathbf{x}_i)$, and let $\mathbf{x}_i \sim G(\cdot)$. We have,

$$E(D_N(\boldsymbol{\beta})) = E\left[\frac{1}{N}\sum_{i=1}^{N}\delta_i d_i w(\mathbf{x}_i)\varphi\Big(a_{iN}(\boldsymbol{\beta})\Big)z_i(\boldsymbol{\beta})\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N} E[\gamma(\mathbf{x}_i)\varphi\Big(a_{iN}(\boldsymbol{\beta})\Big)z_i(\boldsymbol{\beta})]$$
$$= \frac{1}{N}\sum_{i=1}^{N} E\left[\gamma(\mathbf{x}_i)\varphi\left(\frac{NF_{N,\beta}(z_i(\beta))}{N+1}\right)z_i(\beta)\right] \approx \int\int_0^1 \gamma(\mathbf{x})\varphi_N(u)dF_{N,\boldsymbol{\beta}}(u)dudG(\mathbf{x}),$$

where $\varphi_N(u) = \sum_{i=1}^{N}\varphi(i/(N+1))I_{((i-1)/N,i/N]}(u)$ and $F_{N,\boldsymbol{\beta}}(u) = \frac{1}{N}\sum_{i=1}^{N} V_i(u)$. From the boundedness of $\varphi(\cdot)$, we have, as $N\to\infty$, $\varphi_N(u) - \varphi(u) \to 0\, a.s.$ Also, since $z_i(\boldsymbol{\beta})$, $i = 1,\ldots,N$ are independent, $\{V_N\}_{N\ge 1}$ is a sequence of independent random variables satisfying $\sum_{N=1}^{\infty}\mathrm{var}(V_N)/N^2 < \infty$. Thus, by the SLLN, we have, as $N\to\infty$, $F_{N,\boldsymbol{\beta}}(u) - \bar{F}_{\boldsymbol{\beta}}(u) \to 0\, a.s.$, for any $\boldsymbol{\beta}\in\boldsymbol{B}$. A direct application of the Dominated Convergence Theorem gives, as $N\to\infty$,

$$\frac{1}{N}\sum_{i=1}^{N} E[\gamma(\mathbf{X}_i)\varphi\Big(a_{iN}(\boldsymbol{\beta})\Big)z_i(\boldsymbol{\beta})] - \int\int_0^1 \gamma(\mathbf{x})\varphi(u)d\bar{F}_{\boldsymbol{\beta}}(u)dudG(\mathbf{x}) \to 0.$$

Now, applying the SLLN of functions of order statistics (Helmers 1977; Wellner 1977; Van Zwet 1980), we have $D_N(\boldsymbol{\beta}) - \int\int_0^1 \gamma(\mathbf{x})\varphi(u)d\bar{F}_{\boldsymbol{\beta}}(u)dG(\mathbf{x}) \to 0\, a.s.$

and thus, $D_N(\boldsymbol{\beta}) - E(D_N(\boldsymbol{\beta})) \to 0\, a.s.$ Moreover, by theorem 1 of Bindele (2017), we have $\widehat{\boldsymbol{\beta}}_N \to \boldsymbol{\beta}_0\, a.s.$. □

Proof of theorem 2. From their definitions and the Cauchy–Schwartz inequality, we have

$$\begin{aligned}||S_N(\boldsymbol{\beta}) - T_N(\boldsymbol{\beta})|| &\le \frac{1}{N}\sum_{i=1}^{N}\gamma(\mathbf{x}_i)||\nabla_{\boldsymbol{\beta}} g(\mathbf{x}_i, \boldsymbol{\beta})|||\varphi\Big(a_{iN}(\boldsymbol{\beta})\Big) - \varphi(F_{i\boldsymbol{\beta}}(z_i(\boldsymbol{\beta})))| \\ &\le \left[\frac{1}{N}\sum_{i=1}^{N}\gamma^2(\mathbf{x}_i)H^2(\mathbf{x}_i)\right]^{1/2}[\sup_{\boldsymbol{\beta}\in\boldsymbol{B}}\max_{1\le i\le N}|\varphi\Big(a_{iN}(\boldsymbol{\beta})\Big) - \varphi(F_{i\boldsymbol{\beta}}(z_i(\boldsymbol{\beta})))|^2]^{1/2}.\end{aligned}$$

As in the proof of the previous theorem, the SLLN gives that as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}\gamma^2(\mathbf{x}_i)H^2(\mathbf{x}_i) \to E[\gamma^2(\mathbf{X})H^2(\mathbf{X})]\, a.s.,$$

with $E[\gamma^2(\mathbf{X})H^2(\mathbf{X})] < \infty$ and $\sup_{\boldsymbol{\beta}\in\boldsymbol{B}}\max_{1\le i\le N}|\varphi\Big(a_{iN}(\boldsymbol{\beta})\Big) - \varphi(F_{i\boldsymbol{\beta}}(z_i(\boldsymbol{\beta})))|^2 \to 0\, a.s.$ Thus,

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{B}}||S_N(\boldsymbol{\beta}) - T_N(\boldsymbol{\beta})|| \to 0\, a.s.$$

□

Proof of theorem 3. (i) Assuming we can interchange differentiation and integral, observe that $E(S_N(\boldsymbol{\beta})) = E[\nabla_{\boldsymbol{\beta}} D_N(\boldsymbol{\beta})] = \nabla_{\boldsymbol{\beta}} E[D_N(\boldsymbol{\beta})]$, by (J_1) and (J_2). Now, by (J_3), we have $E(S_N(\boldsymbol{\beta}_0)) = \nabla_{\boldsymbol{\beta}} E[D_N(\boldsymbol{\beta}_0)] = 0$. From the fact that $S_N(\boldsymbol{\beta}_0) = T_N(\boldsymbol{\beta}_0) + o_p(n^{-1/2})$, we have $E[T_N(\boldsymbol{\beta}_0)] \to 0$, as $n \to \infty$.

(ii) A direct differentiation of $T_N(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ gives

$$\nabla_\beta T_N(\beta) = -\frac{1}{N}\sum_{i=1}^{N}\gamma(\mathbf{x}_i)[\nabla_\beta g(\mathbf{x}_i, \beta)][\nabla_\beta g(\mathbf{x}_i, \beta)]' f_{i\beta}(z_i(\beta))\varphi'(F_{i\beta}(z_i(\beta)))$$
$$+\frac{1}{N}\sum_{i=1}^{N}\gamma(\mathbf{x}_i)[\nabla^2_\beta g(\mathbf{x}_i, \beta)]\varphi(F_{i\beta}(z_i(\beta))).$$

The previous display evaluated at $\boldsymbol{\beta}_0$ gives

$$\begin{aligned}\nabla_\beta T_N(\beta_0) = &-\frac{1}{N}\sum_{i=1}^{N}\gamma(\mathbf{x}_i)[\nabla_\beta g(\mathbf{x}_i, \beta_0)][\nabla_\beta g(\mathbf{x}_i, \beta_0)]' f(\varepsilon_i)\varphi'(F(\varepsilon_i)) \\ &+\frac{1}{N}\sum_{i=1}^{N}\gamma(\mathbf{x}_i)[\nabla^2_\beta g(\mathbf{x}_i, \beta_0)]\varphi(F(\varepsilon_i)).\end{aligned}$$

By (J_1) and (J_2), $E\Big(\gamma(\mathbf{X})[\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)][\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)]' f(\varepsilon)\varphi'(F(\varepsilon))\Big) < \infty$ and also, $E[\gamma(\mathbf{X})[\nabla^2_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)]\varphi(F(\varepsilon))] < \infty$. From the fact that $\omega_i = (\mathbf{x}_i, \varepsilon_i)$, $i = 1,\ldots,N$ are independent and identically distributed, we have by the SLLN, $\nabla_{\boldsymbol{\beta}} T_N(\boldsymbol{\beta}_0) = \Gamma(\boldsymbol{\beta}_0) + o(1/n)\, a.s.$, where

$$\Gamma(\beta_0) := -E\{\gamma(\mathbf{X})[\nabla_{\beta} g(\mathbf{X},\beta_0)][\nabla_{\beta} g(\mathbf{X},\beta_0)]' f(\varepsilon)\varphi'(F(\varepsilon))\}$$
$$+E\{\gamma(\mathbf{X})[\nabla^2_{\beta} g(\mathbf{X},\beta_0)]\varphi(F(\varepsilon))\}.$$

In addition, if ε is independent of **x**, we have

$$E\Big(\gamma(\mathbf{X})[\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)][\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)]' f(\varepsilon)\varphi'(F(\varepsilon))\Big)$$
$$= -\gamma_{\zeta}^{-1} E\Big(\gamma(\mathbf{X})[\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)][\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)]'\Big),$$

where $\zeta_{\varphi}^{-1} = E[f(\varepsilon)\varphi'(F(\varepsilon))] = \int_0^1 \varphi(u)\varphi_f(u)du$ with $\varphi_f(u) = f'(F^{-1}(u))/f(F^{-1}(u))$. Also,

$$E\Big(\gamma(\mathbf{X})[\nabla^2_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)]\varphi(F(\varepsilon))\Big) = E\Big(\gamma(\mathbf{X})[\nabla^2_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)]\Big)E[\varphi(F(\varepsilon))] = 0,$$

since $E[\varphi(F(\varepsilon))] = 0$, by (J_1). Thus, $\Gamma(\boldsymbol{\beta}_0)$ reduces to

$$\Gamma(\boldsymbol{\beta}_0) := \gamma_{\zeta}^{-1} E\Big(\gamma(\mathbf{X})[\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)][\nabla_{\boldsymbol{\beta}} g(\mathbf{X},\boldsymbol{\beta}_0)]'\Big).$$

(iii) Let $Q_N(\boldsymbol{\beta}_0) = S_N(\boldsymbol{\beta}_0) - E[S_N(\boldsymbol{\beta}_0)]$. Consider the following lemma, whose proof can be constructed along the lines as that of corollary 3.8 of Brunner and Denker (1994).

Lemma 1. Using definitions in remark 3, let ς_N be the minimum eigenvalue of $\mathbf{W}_N = Var(U_n)$ with U_n given by

$$U_n = \int \varphi(J_N(s))(\widehat{F}_n - F_N)(ds) + \int \varphi'(J_N(s))(\widehat{J}_n(s) - J_N(s))F_N(ds)\,.$$

Suppose that $\varsigma_N \geq CN^a$ for some constants $C, a \in \mathbb{R}$ and $m(n)$ is such that $M_0 N^{\gamma} \leq m(n) \leq M_1 N^{\gamma}$ for some constants $0 < M_0 \leq M_1 < \infty$ and $0 < \gamma < (a+1)/2$. Then $m(n)\mathbf{W}_N^{-1/2} Q_N(\boldsymbol{\beta}_0)$ is asymptotically standard multivariate normal, provided φ is twice continuously differentiable with bounded second derivative.

From its definition,

$$S_N(\beta_0) = \frac{1}{N}\sum_{i=1}^{N} \gamma(\mathbf{x}_i)\varphi\left(\frac{R(z_i(\beta_0))}{N+1}\right) = \int \varphi\left(\frac{n}{N+1}\widehat{J}_n)dF_N\right.$$

Now defining B_n by

$$B_n = -\int(\widehat{F}_n - F_N)d\varphi(J_N) + \int(\widehat{J}_n - J_N)\frac{dF_N}{dJ_N}d\varphi(J_N),$$

following Brunner and Denker (1994), it can be shown that $\mathbf{W}_N = N^2 Var(B_n)$. From the fact that $\boldsymbol{\beta}_0 = \text{argmin}_{\boldsymbol{\beta}\in\mathcal{B}} E(D_N(\boldsymbol{\beta}))$, it can be seen that $E[S_N(\boldsymbol{\beta}_0)] = 0$. This implies that $E[Q_N(\boldsymbol{\beta}_0)] = 0$ and $Q_N(\boldsymbol{\beta}_0) = S_N(\boldsymbol{\beta}_0)$. Now, under assumptions (J_2) and (J_3), for theorem 3(iii) to hold, it suffices to show that the conditions of lemma 1 are satisfied. To that end, since $Var(\varepsilon_i|\mathbf{x}_i) > 0$, there exists $\zeta > 0$ such that the minimum eigenvalue of $Var(B_n)$, say μ_N satisfies $\mu_N > \zeta N^b$, for $0 < b < 1/2$. This is obtained under the assumption that $\varsigma_N/N \to \infty$ putting $\mu_N \to \infty$ as $N \to \infty$, see Brunner and Denker (1994) for more discussion. Once again by Brunner and Denker (1994), putting $U_n = NB_n$, we have $Var(U_n) = N^2 Var(B_n)$. Thus, $\varsigma_N = N^2\mu_N \geq \zeta N^{2+b}$. From the fact that π_i is a probability, $\pi_i \leq 1$, for all i. By (J_3), $c_1 N n^{-1} \leq \pi_i N n^{-1} < c_2$, from which $n > \frac{c_1}{c_2}N$. Then, $\sqrt{n} > \sqrt{\frac{c_1}{c_2}}N^{\alpha/2}$, with $\alpha < 1$. Moreover, by (J_3) again, $\sqrt{n} \leq \sqrt{c_2}N^{\alpha/2}$. Thus, $\sqrt{\frac{c_1}{c_2}}N^{\alpha/2} < \sqrt{n} \leq \sqrt{c_2}N^{\alpha/2}$. Hence, putting $a = 2 + b$, $\gamma = \alpha/2$, $M_0 = \sqrt{\frac{c_1}{c_2}}$, $M_1 = \sqrt{c_2}$, $C = \zeta$ and $m(n) = \sqrt{n}$, conditions of lemma 1 are satisfied. This shows that $\mathbf{W}_N^{-1/2}\sqrt{n}Q_N(\boldsymbol{\beta}_0)$ is asymptotically multivariate standard normal. Therefore, $\sqrt{n}S_N(\boldsymbol{\beta}_0)$ is asymptotically multivariate normal with mean 0 and covariance matrix $\mathbf{W}_N$, with $\mathbf{W}_N = Var(\sqrt{n}S_N(\boldsymbol{\beta}_0)) \to \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}$, as $n \to \infty$. □

Proof of theorem 4. From the fact that $S_N(\boldsymbol{\beta}) = T_N(\boldsymbol{\beta}) + o_p(n^{-1})$, using (5), we have

$$0 = T_N(\beta_0) + (\widehat{\beta}_N - \beta_0)'\nabla_\beta T_N(\beta_0) + \frac{1}{2}(\widehat{\beta}_N - \beta_0)'\nabla^2_\beta T_N(\beta^*_N)(\widehat{\beta}_N - \beta_0) + o_p(1/n),$$

where $\boldsymbol{\beta}^*_N$ lies on the segment joining $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}_N$. Under assumptions $(J_1) - (J_3)$, it can be shown in a straight forward manner that $\nabla^2_{\boldsymbol{\beta}}T_N(\boldsymbol{\beta}^*_N)$ is bounded in probability and so, from the consistency of $\widehat{\boldsymbol{\beta}}_N$, we have that $(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0)'\nabla^2_{\boldsymbol{\beta}}T_N(\boldsymbol{\beta}^*_N) \to 0\, a.s.$ Thus, using $\nabla_{\boldsymbol{\beta}}T_N(\boldsymbol{\beta}_0) \to \Gamma(\boldsymbol{\beta}_0)\, a.s.$, we have

$$0 = T_N(\boldsymbol{\beta}_0) + (\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0)'\Gamma(\boldsymbol{\beta}_0) + o_p(1/n).$$

This implies that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) = -\Gamma^{-1}(\boldsymbol{\beta}_0)\sqrt{n}T_N(\boldsymbol{\beta}_0) + o_p(1/\sqrt{n})$, from which applying theorem 3(iii) yields $\sqrt{n}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \to \mathcal{N}_p(0, \Gamma^{-1}(\boldsymbol{\beta}_0)\Sigma_{\boldsymbol{\beta}_0}\Gamma^{-1}(\boldsymbol{\beta}_0))$. □

REFERENCES

Arnab, R. (2017), "Estimation of Distribution Functions and Quantiles," in *Survey Sampling Theory and Applications*, ed. R. Arnab, pp. 747–772, New York: Academic Press (Chapter 23).

Beaumont, J.-F. (2008), "A New Approach to Weighting and Inference in Sample Surveys," *Biometrika*, 95, 539–553.

Bindele, H. F. (2015), "The Signed-Rank Estimator for Nonlinear Regression with Responses Missing at Random," *Electronic Journal of Statistics*, 9, 1424–1448.

———. (2017), "Strong Consistency of the General Rank Estimator," *Communications in Statistics—Theory and Methods*, 46, 532–539.

Bindele, H. F., A. Abebe, and K. N. Meyer (2018), "General Rank-Based Estimation for Regression Single Index Models," *Annals of the Institute of Statistical Mathematics*, 70, 1115–1146.

Bindele, H. F., and A. Adekpedjou (2018), "Rank-Based Inference with Responses Missing Not at Random," *Canadian Journal of Statistics*, 46, 501–528.

Brunner, E., and M. Denker (1994), "Rank Statistics under Dependent Observations and Applications to Factorial Designs," *Journal of Statistical Planning and Inference*, 42, 353– 378.

Chambers, R. L. (2003), "Introduction to Part A," in *Analysis of Survey Data*, ed. R. L. Chambers and C. J. Skinner, pp. 277–287, Chichester: Wiley.

Chambers, R. L., and C. J. Skinner (2003), *Analysis of Survey Data*, Chichester: John Wiley & Sons.

Dua, D., and C. Graff (2017), *UCI: Machine Learning Repository*, Available at https://archive.ics.uci.edu/ml/index.php.

DuMouchel, W. H., and G. J. Duncan (1983), "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples," *Journal of the American Statistical Association*, 78, 535–543.

Einmahl, U., and D. M. Mason (2005), "Uniform in Bandwidth Consistency of Kernel-Type Function Estimators," *The Annals of Statistics*, 33, 1380–1403.

Fuller, W. A. (2011), *Sampling Statistics, Volume 560*, Canada and Hoboken, NJ: John Wiley & Sons.

Hájek, J., Z. Šidák, and P. Sen (1999), *Theory of Rank Tests (Probability and Mathematical Statistics)*, New York: Academic Press.

Helmers, R. (1977), *A Strong Law of Large Numbers for Linear Combinations of Order Statistics*, Report SW 50/77, Amsterdam: Mathematisch Centrum.

Hettmansperger, T. P., and J. W. McKean (2011), *Robust Nonparametric Statistical Methods, Volume 119 of Monographs on Statistics and Applied Probability* (2nd edition), Boca Raton, FL: CRC Press.

Jaeckel, L. A. (1972), "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals," *The Annals of Mathematical Statistics*, 43, 1449–1458.

Kim, J. K., and C. J. Skinner (2013), "Weighting in Survey Analysis under Informative Sampling," *Biometrika*, 100, 385–398.

Korn, E., and B. Graubard (2011), *Analysis of Health Surveys*, New York: Wiley Series in Survey Methodology, Wiley.

Magee, L. (1998). "Improving Survey-Weighted Least Squares Regression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60, 115–126.

Pfeffermann, D. (1993), "The Role of Sampling Weights When Modelling Survey Data," *International Statistical Review*, 61, 317–337.

———. (2011), “Modelling of Complex Survey Data: Why Model? Why Is It a Problem? How Can We Approach It,” *Survey Methodology*, 37, 115–136.

Pfeffermann, D., and M. Sverchkov (1999), “Parametric and Semi-Parametric Estimation of Regression Models Fitted to Survey Data,” *Sankhyā B*, 61, 166–186.

———. (2009), “Inference under Informative Sampling,” in *Handbook of Statistics, Volume*, vol. 29, pp. 455–487, Elsevier.

Scott, A. J., and C. J. Wild (2011), “Fitting Regression Models with Response-Biased Samples,” *Canadian Journal of Statistics*, 39, 536.

Skinner, C., and B. Mason (2012). “Weighting in the Regression Analysis of Survey Data with a Cross-National Application,” *Canadian Journal of Statistics*, 40, 697–711.

Steven, G., B. T. Heeringa, and P. A. B. West (2017), *Applied Survey Data Analysis* (2nd ed.), Boca Raton, FL: Chapman and Hall/CRC Statistics in the Social and Behavioral Sciences.

Van Zwet, W. R. (1980), “A Strong Law for Linear Functions of Order Statistics,” *The Annals of Probability*, 8(5), 986–990.

Wellner, J. A. (1977), “A Glivenko-Cantelli Theorem and Strong Laws of Large Numbers for Functions of Order Statistics,” *The Annals of Statistics*, 5(3), 473–480.

Whitt, W. (2002), *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*, New York: Springer Science & Business Media.