

The Signed-rank estimator for nonlinear regression with responses missing at random

Huybrechts F. Bindele*

411 University Blvd. North, ILB 316, Mobile AL 36688-0002

Department of Mathematics & Statistics

University of South Alabama

e-mail: hbindele@southalabama.edu

Abstract: This paper is concerned with the study of the signed-rank estimator of the regression coefficients under the assumption that some responses are missing at random in the regression model. Strong consistency and asymptotic normality of the proposed estimator are established under mild conditions. To demonstrate the performance of the signed-rank estimator, a simulation study is conducted under different settings of model error's distributions, and shows that the proposed estimator is more efficient than the least squares estimator whenever the error distribution is heavy-tailed or contaminated. When the model error follows a normal distribution, the simulation experiment shows that the signed-rank estimator is more efficient than its least squares counterpart whenever a large proportion of the responses are missing.

MSC 2010 subject classifications: Primary 62J02, 62G05; secondary 62F12, 62G20.

Keywords and phrases: Signed-rank norm, strong consistency, asymptotic normality, imputation, missing at random.

Received June 2014.

1. Introduction

Nowadays the missing data problem is a troubleshooting problem in almost all statistical studies; it occurs for several reasons in a large array of applications. Such reasons may occur due to equipment malfunction, unfavorable weather conditions, incorrect entry of the data or non-participation of individuals due to illness. Examples of missing data problems include: users of a recommender system rate, extremely small number of available books, movies, or songs leading to massive amounts of missing data; in a clinical trial, participants may drop out during the course of study leading to missing observations at subsequent time points; while diagnosing a patient, all applicable tests may not be ordered by a doctor; a sensor in a remote sensor network may be damaged and cease to transmit data; certain regions of a gene microarray may fail to yield measurements of the underlying gene expressions due to scratches, fingerprints, or manufacturing defects; etc.

*Corresponding author.

The missingness scenario to be considered in this paper is that of the missing response in the context of regression analysis that often arises in various experimental settings such as opinion polls, socioeconomic investigations, medical studies and market research surveys.

This problem has been intensively investigated in the literature and appears to be a very challenging task. The challenge comes from the fact that in most cases, missing data contains either little or no information about the missing data mechanism (MDM). The most common assumption used for the MDM and also employed in this paper, is the missing at random (MAR) assumption, whose details can be found in [35].

When f is taken to be a linear function of θ , several regression models in the framework of missing data have been studied by many authors such as [6, 21, 46, 45, 30, 52, 50, 51, 42, 31], among others. All of them have done considerable work using different techniques, but most of them use the least squares approach in estimating the regression parameters.

In this paper, we consider the study of a general signed-rank estimator of the regression coefficients that is obtained as a minimizer of an objective function $D_{nj}(\theta)$ defined in the next section with the assumption that some responses are missing at random. The motivation for using the signed-rank approach comes from the fact that it provides robust estimators compared to the more common LS approach when dealing with situations where there are outlying observations in the response space, and/or where the error distribution is heavy-tailed. A particular case of the signed-rank approach, the LAD estimator has been studied by [28] and [14]. Although robust in many situations, the LAD has been shown to be less efficient than the estimator commonly referred to as the signed-rank estimator and, also, the general form of the signed-rank estimator defined as a minimizer of $D_{nj}(\theta)$, has been extensively studied in the literature by authors such as [16, 3], and [1] for models with i.i.d. errors. Applications of this approach to factorial designs under dependent observations are given in [4]. As the signed-rank procedure is based on rank scores, an important approach using such rank scores in estimating the true nonlinear regression parameter is developed in [17] and [18] for i.i.d. errors and [25] for time-series regression models. [25] provided a comprehensive review of different techniques of estimation in the literature including M-estimation, R-estimation and L-estimation. In the context of missing data with the MAR assumption and f taken to be a nonlinear function of θ , [8] considered the LAD estimator of the regression coefficients and established its asymptotic properties. Another important piece of work in the framework of missing data is given in [26], who provided the estimation of a function of the mean response with responses missing at random and also established the asymptotic properties of its proposed estimator.

In the framework of general nonlinear regression with responses missing at random, to the best of our knowledge, no asymptotic results have been provided for the general signed-rank estimator. The purpose of this paper is to provide conditions needed for the strong consistency and the \sqrt{n} -asymptotic normality of the resulting estimator toward statistical inference when some responses in the regression model are missing at random. To measure the accuracy of the

proposed estimator, we run a Monte Carlo simulation to obtain MSEs of the SR and compare them to those obtained via LS estimation. From the simulation study, it is shown that the signed-rank approach outperforms the least squares approach under different settings. This is also shown to be true even when the errors' distribution is normal, but with a high rate of missing information.

To this end, the rest of the paper is organized as follows: in section 2, we define the model, discuss the missing mechanism, motivate the imputation procedure and provide the estimation procedure of the regression coefficients. Section 3 and section 4 are concerned with the asymptotic properties of the signed-rank estimator. Settings of the Monte Carlo simulation are presented in section 5 with section 6 providing the discussion on the simulation results. A conclusion summarizing the findings of the paper is given in section 7. Finally, Appendix A contains complete proofs and some sketches of proofs of the theorems presented in the paper.

2. Model definition and estimation

Consider the nonlinear regression model

$$y_i = f(\mathbf{x}_i, \theta) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (2.1)$$

where $\theta \in \Theta \subset \mathbb{R}^p$ is a vector of parameters, \mathbf{x}_i 's are i.i.d. p -variable random covariate vectors, the function f is twice continuously differentiable as a function of θ and measurable as a function of \mathbf{x} . Θ is assumed to be compact and $E(\varepsilon_i|\mathbf{x}_i) = 0$, $\text{var}(\varepsilon_i|\mathbf{x}_i) < \infty$. In this paper, we are interested in inference on the true parameter θ_0 when there are missing responses in the nonlinear model (2.1). Specifically, we consider the case where some values of y in the sample of size n may be missing, but \mathbf{x} is fully observed. That is, we obtain the following incomplete observations $\{y_i, \delta_i, \mathbf{x}_i\}_{i=1,2,\dots,n}$ from (2.1), where \mathbf{x}_i 's are observed, and,

$$\delta_i = \begin{cases} 0, & y_i \text{ is missing;} \\ 1, & \text{otherwise.} \end{cases}$$

One major problem of dealing with missing data is that it is usually unknown how nonresponse for each variable is generated, that is, the distribution $P(\delta = 1|y, \mathbf{x})$, referred to as the nonresponse mechanism is unknown. It is often necessary to make assumptions about this distribution, which are almost impossible to verify [19, 27, 22, 24]. To that extent, we assume that y is missing at random (MAR). The MAR assumption implies that δ and y are conditionally independent given \mathbf{x} , that is $P(\delta = 1|y, \mathbf{x}) = P(\delta = 1|\mathbf{x})$. Although usually very difficult to test in practice, this is a common assumption for statistical analysis with missing data and is reasonable in many practical situations, see chapter 1 of [23]. Under the MAR assumption, [45] developed inference tools in the missing response case for the mean of y based on the least squares estimation approach, and [47] studied the least squares estimator of the regression

coefficient θ in the semi-parametric linear model. One method for constructing the confidence interval for the true mean of y is that of the empirical likelihood method introduced by [29]. This approach is used to investigate a variety of statistical problems by [13, 6, 21, 30, 52, 50, 51], just to mention a few. These works demonstrate that the method of empirical likelihood has a number of advantages over methods such as those based on normal approximations or the bootstrap. Studies which used this approach to study (2.1) under the MAR assumption include [46, 45, 42], and [49]. Other missing mechanisms include missing completely at random (MCAR) and missing not at random (MNAR), and are discussed in [36].

Imputation in general makes use of a number of auxiliary variables that are statistically related to the variable in which the item-nonresponse occurs by means of an *imputation model* [22, 37]. The main reason for carrying out imputation is to reduce bias, which occurs because the distribution of the missing values, assuming it was known, generally differs from the distribution of observed items. When imputation is used properly, it is possible to recreate a balanced design such that procedures used in tackling complete data can be applied in many situations. Rather than deleting cases that are subject to item-nonresponse, the sample size is maintained resulting in a potentially higher efficiency than case deletion analysis. Imputation usually makes use of observed auxiliary information for cases with nonresponse maintaining high precision [38]. However, such an imputation may result in a negative impact if the imputed values are not properly used. Since the variance estimation is subject to the imputation procedure, an adequate estimation requires adjustment methods such as those proposed by [26] to correct for the increase in variability due to nonresponse and imputation. In most of the regression problems, the commonly used approaches include linear regression imputation [15], nonparametric kernel regression imputation [7, 46], and semi-parametric regression imputation [45, 47]. Another approach for handling missing data is the inverse probability weighting. This approach has gained considerable attention as a way of dealing with missing data problems. For a discussion of this approach, see [34, 54, 44, 45] and references therein. As pointed out by [47], for missing problems, the inverse probability weighting approach usually depends on high dimensional smoothing for estimating the completely unknown propensity score function. This suffers from the *curse of dimensionality* that may restrict the use of the resulting estimator. One way to avoid such a problem is to use the inverse marginal probability weighted method proposed by [45].

To this end, let us introduce the following notations: $\sigma^2(\mathbf{x}) = E(\varepsilon^2|\mathbf{x})$, $\Delta(x) = P(\delta = 1|\mathbf{x} = x)$. Taking advantage of the imputation procedure in [45] and [47], that is, by simple imputation and inverse marginal weighting probability, define

$$y_{ij} = \begin{cases} \delta_i y_i + (1 - \delta_i) f(\mathbf{x}_i, \theta) & j = 1 \\ \frac{\delta_i}{\Delta(\mathbf{x}_i)} y_i + \left(1 - \frac{\delta_i}{\Delta(\mathbf{x}_i)}\right) f(\mathbf{x}_i, \theta) & j = 2. \end{cases}$$

Under the MAR assumption and the fact that $E(\varepsilon|\mathbf{x}_i) = 0$, it can easily be

shown that $E(y_{ij}|\mathbf{x}_i) = f(\mathbf{x}_i, \theta) = E(y_i|\mathbf{x}_i)$.

Consider the signed-rank objective function,

$$D_{jn}(\theta) = \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{R_{ij}}{n+1}\right) |z_{ij}(\theta)|, \quad (2.2)$$

where $z_{ij}(\theta) = y_{ij} - f(\mathbf{x}_i, \theta)$, R_{ij} is the i^{th} rank of $|z_{ij}(\theta)|$ and φ is a differentiable positive function with bounded derivative. It can also be shown that $E(z_{ij}(\theta)|\mathbf{x}_i) = 0$. The signed-rank estimator $\hat{\theta}_{jn}$ of θ_0 is the minimizer of $D_{jn}(\theta)$, that is

$$\hat{\theta}_{jn} = \underset{\theta \in \Theta}{\text{Argmin}} D_{jn}(\theta). \quad (2.3)$$

From the definition of y_{ij} , $\hat{\theta}_{jn}$ is obtained based on complete observations, that is, ignoring observations with missing responses. Note that $D_{jn}(\theta)$ is a norm (see [16]) and hence convex as a function of the residuals. The existence of $\hat{\theta}_{jn}$ is ensured by the continuity of $D_{jn}(\theta)$ on Θ compact. Also, setting $\varphi(t) = 1$ for all $t \in (0, 1)$ in (2.2), $\hat{\theta}_{nj}$ becomes the LAD estimator.

Remark 2.1. For the complete case which considers observations without missing responses, [3] give a brief discussion about the relative efficiency of the SR with respect to the LS and the least absolute deviation (LAD). They show that the SR is robust and more efficient than the LS and the LAD when dealing with heavy-tailed model error's distributions and/or outliers in the response space. The same paper provides a weighted version of the signed rank that is robust and more efficient than the LS and the LAD when dealing with high leverage points (outliers in the design space). It is also well-known that when dealing with heavy-tailed error distributions and/or outliers in the response space, the SR is robust and more efficient compared to the LS, LAD and MLE, see [16].

In the expression of y_{ij} , we have the function $\Delta(\mathbf{x}) = P(\delta = 1|\mathbf{x})$ that might be unknown and then needs to be estimated. Being a probability and under the assumption that $\inf_x \Delta(x) > 0$, one can estimate $\Delta(x)$ non-parametrically. To this end, define

$$\tilde{y}_{ijn} = \begin{cases} \delta_i y_i + (1 - \delta_i) f(\mathbf{x}_i, \hat{\theta}_{jn}) & j = 1 \\ \frac{\delta_i}{\hat{\Gamma}(\mathbf{x}_i)} y_i + \left(1 - \frac{\delta_i}{\hat{\Gamma}(\mathbf{x}_i)}\right) f(\mathbf{x}_i, \hat{\theta}_{jn}) & j = 2, \end{cases}$$

where $\hat{\Gamma}(x) = \sum_{j=1}^n \omega_{nj}(x) \delta_j$, with

$$\omega_{nj}(x) = \frac{K\left(\frac{x - \mathbf{x}_j}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x - \mathbf{x}_j}{h_n}\right)},$$

where K is a kernel function and h_n a bandwidth sequence satisfying $h_n \rightarrow 0$ as $n \rightarrow \infty$. $\hat{\Gamma}(x)$ defined above is known as a kernel smoother and is used in the literature for different purposes including the machine learning literature

under the banner of semi-supervised learning, see [9]. Also, under the MAR condition and the fact that $E(\varepsilon|\mathbf{x}_i) = 0$, we have $E(\tilde{y}_{ijn}|\mathbf{x}_i) = f(\mathbf{x}_i, \theta)$. Note that under assumptions (I_1) and (I_3) given in Theorem 2.1 below, $\widehat{\Gamma}(x) \rightarrow \Delta(x)$ a.s. as $n \rightarrow \infty$. This together with the assumption that $\inf_x \Delta(x) > 0$, we have $\frac{1}{\widehat{\Gamma}(x)} \rightarrow \frac{1}{\Delta(x)}$ a.s. as $n \rightarrow \infty$. Now define $\widetilde{D}_{jn}(\theta)$ as

$$\widetilde{D}_{jn}(\theta) = \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{\widetilde{R}_{ij}}{n+1}\right) |\tilde{z}_{ij}(\theta)|, \quad (2.4)$$

where $\tilde{z}_{ij}(\theta) = \tilde{y}_{ijn} - f(\mathbf{x}_i, \theta)$, \widetilde{R}_{ij} is the i^{th} rank of $|\tilde{z}_{ij}(\theta)|$. Note that the objective function given in Equation (2.4) is defined on dependent residuals, as the imputation procedures introduce a dependence structure. Therefore, the asymptotic properties (consistency and asymptotic normality) of the minimizer of $\widetilde{D}_{jn}(\theta)$ cannot be established following [3]. [2] established such results, but strong mixing conditions were imposed on the model error. One way to avoid such strong mixing conditions is to establish the following theorem whose proof is provided in the Appendix.

Theorem 2.1. *Under the following assumptions:*

- (I_1) $K(\cdot)$ is a regular kernel function of order $r > 2$ and the bandwidth h_n satisfies $nh_n \rightarrow \infty$, $nh_n^{4r} \rightarrow 0$ and $C(\ln n/n)^\gamma < h_n < b_n < 1$, for any $C > 0$, $\gamma = 1 - 2/p$, $p > 2$ and $b_n \rightarrow 0$ as $n \rightarrow \infty$.
- (I_2) φ is a bounded, twice continuously differentiable score function with bounded derivatives, defined on $(0, 1)$. Also, $\inf_x \Delta(x) > 0$.
- (I_3) $\sup_x E(\varepsilon^p|\mathbf{x} = x) < \infty$, for $p \geq 2$.

we have

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\widetilde{D}_{jn}(\theta) - D_{jn}(\theta)| = 0$$

Remark 2.2. Assumptions (I_1) and (I_3) ensure the almost sure uniform convergence of $\widehat{\Gamma}_n(x)$ to $\Delta(x)$. The optimal bandwidth for practical purposes can be chosen to lie in the interval $[an^{-1/5}, bn^{-1/5}]$, for $0 < a < b < \infty$. More discussion about this fact can be found in [12]. If we set

$$\widetilde{\theta}_{jn} = \underset{\theta \in \Theta}{\text{Argmin}} \widetilde{D}_{jn}(\theta),$$

Theorem 2.1 stands for the asymptotic equivalence between $\widetilde{\theta}_{jn}$ and $\widehat{\theta}_{jn}$; that is, in the neighborhood of θ_0 , they share the same asymptotic properties (strong consistency, \sqrt{n} -consistency and asymptotic normality).

3. Consistency

Let (Ω, \mathcal{F}, P) be a probability space. For $i = 1, \dots, n$, assume that ε_i and \mathbf{x}_i are independent. Also, let $\mathbf{x}_i \sim H$, $\varepsilon_i \sim G$ and $|z_{ij}(\theta)| \sim G_\theta^j$. Consider the following assumptions:

- (I₄) $P(d_i(\theta) = 0) < 1$ for all $\theta \neq \theta_0$, where $d_i(\theta) = f(\mathbf{x}_i, \theta) - f(\mathbf{x}_i, \theta_0)$.
- (I₅) $E(|d_i(\theta)|^p) < \infty$, for some $p \geq 2$.
- (I₆) The conditional density of ε_i given \mathbf{x}_i is symmetric about 0 and strictly decreasing on \mathbb{R}^+ .

Remark 3.1. Assumption (I₄) stands for the identifiability condition. Under this assumption, if B is an open neighborhood of θ_0 , it can be shown that

$$\lim_{n \rightarrow \infty} \inf_{\theta \in B^c} [E(D_{jn}(\theta)) - E(D_{jn}(\theta_0))] \geq \alpha$$

for some $\alpha > 0$. Both (I₄) and (I₆) are necessary to prove that $\inf_{\theta \in \Theta^*} \mu_j(\theta) > \mu_j(\theta_0)$ in Lemma 3.1 below. (I₃) and (I₄) ensure the uniform convergence in Lemma 3.1. Also, if we restrict ourselves just to the consistency result, continuity is enough for f .

Theorem 3.1. Under assumptions (I₁) – (I₆), $\tilde{\theta}_{jn} \rightarrow \theta_0$ a.s. as $n \rightarrow \infty$.

The following Lemma is essential for the proof of Theorem 3.1. Its proof can be constructed along the lines given in [3]. For sake of brevity, it will not be included here.

Lemma 3.1. Under assumptions (I₂)–(I₆), $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} D_{jn}(\theta) = \mu_j(\theta)$ a.s., where $\mu_j(\theta) = \int \varphi(t) dG_\theta^j(t)$ is a continuous function of θ satisfying

$$\inf_{\theta \in \Theta^*} \mu_j(\theta) > \mu_j(\theta_0)$$

and Θ^* is a closed subset of Θ not containing θ_0

Remark 3.2. $\hat{\theta}_{jn}$ being also a strong consistent estimator of θ_0 [1], and by continuity of f , $f(\mathbf{x}_i, \hat{\theta}_{jn}) \rightarrow f(\mathbf{x}_i, \theta)$, and this implies that $\tilde{y}_{ijn} \rightarrow y_{ij}$ as $n \rightarrow \infty$. From this, we have as $n \rightarrow \infty$

$$\tilde{\theta}_{jn} - \hat{\theta}_{jn} \rightarrow 0 \quad a.s.$$

and from Remark 2.2, $\sqrt{n}(\tilde{\theta}_{jn} - \hat{\theta}_{jn}) \rightarrow 0$ a.s.

4. Asymptotic normality

Setting $\tilde{S}_{jn}(\theta) = -\nabla \tilde{D}_{jn}(\theta)$, it is clear that $\tilde{\theta}_{jn}$ is a zero of $\tilde{S}_{jn}(\theta)$. With the following notations for the gradient and the Hessian matrix, respectively:

$$\nabla_\theta f(\mathbf{x}, \theta) = \nabla_\theta f(\mathbf{x}), \quad \text{and} \quad \nabla_\theta^2 f(\mathbf{x}, \theta) = \nabla_\theta^2 f(\mathbf{x}),$$

we have,

$$\tilde{S}_{jn}(\theta) = \begin{cases} \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{\tilde{R}_{ij}}{n+1}\right) \delta_i \nabla_\theta f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i), & \text{if } j = 1 \\ \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{\tilde{R}_{ij}}{n+1}\right) \frac{\delta_i}{\Gamma(\mathbf{x}_i)} \nabla_\theta f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i), & \text{if } j = 2. \end{cases}$$

Also, setting

$$\lambda_{ij} = \begin{cases} \delta_i, & \text{if } j = 1 \\ \frac{\delta_i}{\widehat{\Gamma}(\mathbf{x}_i)}, & \text{if } j = 2, \end{cases}$$

we have, $\widetilde{S}_{jn}(\theta) = \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{\widetilde{R}_{ij}}{n+1}\right) \lambda_{ij} \nabla_{\theta} f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i)$. Putting $e_i = y_i - f(\mathbf{x}_i, \theta_0)$, we have the following theorem,

Theorem 4.1. For any θ in the neighborhood of θ_0 and with probability 1,

$$\sqrt{n}[\widetilde{S}_{jn}(\theta) - \widetilde{S}_{jn}(\theta_0) - \mathbf{G}_n^T(\theta - \theta_0)] = o(1), \quad (4.1)$$

where $\mathbf{G}_n = \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{R_{ij}^*}{n+1}\right) \lambda_{ij} \nabla_{\xi}^2 f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i)$, with $\xi = \lambda\theta_0 + (1 - \lambda)\theta$ for some $\lambda \in (0, 1)$.

The proof of this theorem is provided in the Appendix, and is just a slight modification of that given in [2] taking into account the strong consistency of $\widehat{\Gamma}(\mathbf{x})$.

Now assuming that the true parameter θ_0 satisfies

$$\theta_0 = \underset{\theta \in \Theta}{\text{Argmin}} E[\widetilde{D}_{jn}(\theta)],$$

we have

$$E[\widetilde{S}_{jn}(\theta_0)] = 0.$$

Set $G_{ij}^+(t) = P(|\widetilde{z}_{ij}(\theta_0)| \leq t)$ for $j = 1, 2$ and $i = 1, \dots, n$. Also, set

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \nu_{ij} G_{ij}^+(t) \quad \text{and} \quad \widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \nu_{ij} I_{[0, \infty)}(t - |\widetilde{z}_{ij}(\theta_0)|),$$

$$J_n(t) = \frac{1}{n} \sum_{i=1}^n G_{ij}^+(t) \quad \text{and} \quad \widehat{J}_n(t) = \frac{1}{n} \sum_{i=1}^n I_{[0, \infty)}(t - |\widetilde{z}_{ij}(\theta_0)|),$$

where $\nu_{ij} = \lambda_{ij} \nabla_{\theta_0} f(\mathbf{x}_i)$. Then,

$$\widetilde{S}_{jn}(\theta) = \frac{1}{n} \sum_{i=1}^n \nu_{ij} \varphi\left(\frac{\widetilde{R}_{ij}}{n+1}\right) \times \text{sgn}(\widetilde{z}_{ij}(\theta_0)) = \int \text{sgn}(z) \varphi\left(\frac{n}{n+1} \widehat{J}_n(|z|)\right) \widehat{F}_n(dz).$$

On the other hand, note that $E[\widetilde{S}_{jn}(\theta_0)] = \int \text{sgn}(z) \varphi\left(J_n(|z|)\right) F_n(dz) = 0$. Set

$$\Sigma_{jn} = n^2 E[(\widetilde{S}_{jn}(\theta_0))(\widetilde{S}_{jn}(\theta_0))^T]. \quad (4.2)$$

Theorem 4.2. Let τ_{jn} represent the minimum eigenvalue of Σ_{jn} . Then assuming that

$$(I_7) \lim_{n \rightarrow \infty} \frac{n^2}{\tau_{jn}} = 0,$$

$\tilde{S}_{jn}(\theta_0)$ is asymptotically normal $N(\mathbf{0}, \Sigma_{jn})$, where $\mathbf{0}$ is a vector of zeros.

The proof of this theorem can be constructed along the lines that of Theorem 5.2 of [4]. Hence for sake of brevity, it will not be included here.

From (4.1) and the fact that $\tilde{S}_{jn}(\tilde{\theta}_{jn}) = 0$, we have

$$\sqrt{n}(\tilde{\theta}_{jn} - \theta_0) = (\sqrt{n}\mathbf{G}_n)^{-1}\tilde{S}_{jn}(\theta_0). \quad (4.3)$$

To this end, we will assume that

$$(I_8) \sqrt{n}\mathbf{G}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi\left(\frac{R_{ij}^*}{n+1}\right) \lambda_{ij} \nabla_{\xi}^2 f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i) \rightarrow \mathcal{W} \quad a.s. \quad as \quad n \rightarrow \infty,$$

where \mathcal{W} is a positive definite matrix. Hence, we have the following asymptotic normality theorem.

Theorem 4.3. Under assumptions (I₁) – (I₈), we have,

$$\sqrt{n}(\tilde{\theta}_{jn} - \theta_0) \sim N(\mathbf{0}, \mathcal{W}^{-1}\Sigma_{jn}\mathcal{W}^{-1}),$$

where Σ_{jn} is defined by equation (4.2) and \mathcal{W} is obtained from assumption (I₈).

Remark 4.1. This result allows us to make statistical inferences (confidence intervals and hypothesis tests) about the true regression parameters. To achieve this in practice, the covariance matrix in Theorem 4.3 above needs to be estimated. This turns out to be a complicated task in the nonlinear regression setting, as it depends on the true parameters and the imputation procedure. One way to avoid this problem is to use a plug-in or sandwich estimator as discussed in [3] for the complete case and [26] in the missing responses case under the MAR assumption. To this end, it needs to be stressed that standard variance estimation for a point estimator valid for complete data may, in many cases, lead to severe underestimation of the true variance if applied to observed and imputed data [33]. Standard variance estimation techniques are therefore not adequate in the presence of imputation. In general, all imputation methods require adjustments to the variance estimation formula to reflect the additional variability due to nonresponse and imputation correctly. Various methods exist for estimating the variance of an estimator under imputation, including multiple imputation [36, 42, 20], two-phase approaches [32, 40], model-assisted approaches [11, 5] and replication methods, such as jackknife variance estimators [33, 39, 53, 41] and references therein.

5. Simulation

To illustrate the performance of the signed-rank estimator, a simulation study is conducted using R. In model (2.1), two scenarios are being considered:

Scenario 1: We set $f(\theta, x) = \beta_0 x / (\beta_1 + x)$, where x is generated from the normal distribution with mean 1 and standard deviation 1, $\theta = (\beta_0, \beta_1)^\tau$ with

the true parameter set at $\theta_0 = (5, 0)^\tau$. The random error is generated from a set of two different distributions: a contaminated normal distribution $\mathcal{CN}(\epsilon) = (1 - \epsilon)N(0, 1) + \epsilon N(0, \sigma^2)$ with $\sigma = 3$ and ϵ taken to be a proportion of contamination, and the t -distribution with various degrees of freedom (df). In this scenario the sample size is set at $n = 30$.

Scenario 2: In signal processing, one important model is the sinusoidal model defined via model (2.1) by setting $f(x, \theta) = C + A \sin((\pi\sqrt{12}/1.645)\omega x + \phi)$, where $\theta = (C, A, \omega, \phi)^\tau$. C is a constant defining the mean level, A is an amplitude for the sine wave, ω is the frequency, x is a time variable to be generated from the uniform distribution in the interval $[0, 2\pi]$, and ϕ is the scale parameter known as the phase. For simplicity, we set $C = 0$, $A = 1$, $\phi = (1.645/\sqrt{12} - \sqrt{3}/2)/(3.29/\sqrt{12}) \approx -0.412$, and we are interested in the estimation of the true frequency set at $\omega_0 = 1/\sqrt{3}$. To investigate the effect of the sample size (n) and the tail thickness of the distribution, the random error is generated from the Laplace distribution with different sample sizes ($n = 20, 35, 50, 70, 100$).

As pointed out in [20], under the MAR assumption and when the distribution of the model error is fully parametric, when it comes to estimating the regression parameters, it is enough to consider the complete case as it results in consistent estimators. The same remark is valid for the considered approach, as is shown in Theorem 3.1, and discussed in Remark 3.1 and Remark 3.2. In practice, unfortunately, with high rates of missingness, it is almost impossible to fully specify the distribution of the model error. Also, when making inference (hypothesis testing, confidence intervals) about regression parameters, it is often of interest to estimate the asymptotic variance of the corresponding estimators. Such an estimation of the variance turns out to be dependent upon the imputation procedure; see [45, 47, 42] and [26] regarding a discussion of this fact. To this end, we show through the simulation below, based on both imputation procedures, that the SR provides more efficient estimators compared to the LS.

To carry out the imputation for missing responses, δ was generated from the Bernoulli ($\Delta(x)$) distribution, where three cases were considered:

- Case 1: $\Delta(x) = \eta$ is taken to be a sequence of proportion of no missingness starting from 50% to 100% with steps of 10%.
- Case 2: $\Delta(x) = 0.8 + 0.2|x - 1|$ if $|x - 1| \leq 1$ and 0.95 elsewhere. The expected value of $\Delta(x)$ is 0.91 while the empirical mean of $\widehat{\Gamma}(x)$ is 0.905. These values were also obtained by [8] under the same setting on x .
- Case 3: $\Delta(x) = 0.9 - 0.2|x - 1|$ if $|x - 1| \leq 4.5$ and 0.1 elsewhere. The expected value of $\Delta(x)$ is 0.79 while the empirical mean of $\widehat{\Gamma}(x)$ is 0.78.

The first case is used to investigate the effect of missing completely at random, as δ is independent of x , while the last two are used for the effect of missing at random, as δ and x are dependent via $\Delta(x)$. Also, while all three cases are being

TABLE 1
MSEs×10⁻³ of the SR & the LS against the rate of contamination (Case 2), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the contaminated normal distribution with different rates of contamination (ϵ)

ϵ	Ker		Log		SI		
	LS	SR	LS	SR	LS	SR	
β_0	0.00	41.820	44.568	46.363	45.391	56.731	58.446
	0.01	56.079	46.922	97.741	50.703	76.134	62.866
	0.05	73.554	56.118	145.066	62.716	99.597	74.396
	0.10	94.143	62.560	197.740	73.862	126.480	83.796
	0.15	107.621	71.155	232.612	88.947	141.219	92.324
	0.25	125.318	79.198	267.321	103.529	164.860	104.619
β_1	0.00	0.011	0.012	0.018	0.031	0.497	0.040
	0.01	0.021	0.016	0.676	0.131	0.544	0.047
	0.05	0.033	0.019	1.655	0.187	0.679	0.088
	0.10	0.051	0.030	1.857	0.814	0.780	0.111
	0.15	0.051	0.037	3.063	1.277	0.886	0.144
	0.25	0.078	0.048	3.311	1.589	1.317	0.220

used in scenario 1, just the last two cases are used in scenario 2. Furthermore, in scenario 2, setting $\zeta(x) = 1 - \Delta(x)$, the expected value of $\zeta(x)$ can be obtained from that of $\Delta(x)$ and the empirical average of $\hat{\zeta}(x) = 1 - \hat{\Gamma}(x)$ is obtained from that of $\hat{\Gamma}(x)$. This leads to a high missingness rate of about 91% for Case 2 and 79% for Case 3 to be investigated under the Laplace model error with different sample sizes ($n = 50, 70, 100$). From these settings, θ_0 is estimated using simple imputation (**SI**) and the inverse marginal probability (**IP**) procedures. For the inverse probability weighting imputation the function $K(x)$ is chosen in two ways: first, it is taken to be a Gaussian kernel (**Ker**), defined by

$$K(x) = \frac{1}{2\pi} e^{-x^2/2}$$

and, secondly, it is taken to be the logistic propensity score (**Log**), defined by

$$K(x) = \frac{1}{1 + e^{-x}},$$

as in [26]. The kernel estimation involves the choice of the bandwidth. This can be done by either a joint minimization of $\tilde{D}_{nj}(h, \theta)$ as a function of the bandwidth h and θ , or a simultaneous minimization of $\tilde{D}_{nj}(h, \theta)$ by first choosing $h \in \{h : an^{-1/5} \leq h \leq bn^{-1/5}\}$ for $a, b > 0$, and obtaining the estimator of θ , and next obtaining the estimator of h when θ is replaced by its estimator. As shown in [10], the two minimization procedures are equivalent. Also, as pointed out in [12] and [10], the estimated bandwidth is proportional to $n^{-1/5}$. So for simplicity, in our computations, the bandwidth is set at $n^{-1/5}$. From 5000 simulations, MSEs of the signed-rank (SR) estimator of θ_0 are reported and compared to those of the least squares (LS) under different settings. Tables 1–8, summarize the results of the simulation study which are discussed in the next section.

TABLE 2

$MSEs \times 10^{-3}$ of the SR & LS estimators against the degree of freedom (Case 2), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the t -distribution with different degrees of freedom (df)

df	Ker		Log		SI		
	LS	SR	LS	SR	LS	SR	
β_0	5	70.429	56.205	77.509	57.189	94.254	75.042
	10	50.482	49.036	57.558	50.191	67.250	64.590
	15	48.520	48.064	52.429	49.927	63.691	62.890
	20	47.679	46.010	51.685	48.988	61.526	61.047
	25	46.428	45.931	49.793	47.144	61.274	60.104
	30	44.316	45.750	49.651	46.866	59.883	59.753
β_1	5	0.021	0.019	0.154	0.052	0.559	0.079
	10	0.019	0.016	0.039	0.037	0.498	0.059
	15	0.015	0.014	0.036	0.028	0.411	0.057
	20	0.014	0.013	0.033	0.027	0.255	0.053
	25	0.013	0.012	0.028	0.025	0.251	0.046
	30	0.011	0.012	0.023	0.024	0.213	0.044

6. Discussion

Scenario 1: In Case 1, based on simple imputation and inverse marginal probability procedures with either the Gaussian kernel or the logistic propensity score under the contaminated normal distribution model error, the SR performs better than the LS as the rate of contamination increases (Table 3). Although, their MSEs are comparable for the normal distribution model error ($\mathcal{CN}(0)$), the SR is still more efficient than the LS for the considered proportions of missingness. A similar observation is made under the t -distribution model error with the SR being superior to LS for the considered degrees of freedom (from heavier tails to the situation when n gets larger), see Table 4.

In Case 2, once again under contaminated distribution model error, the SR outperforms the LS as the proportion of contamination increases (Table 1), whereas under the t -distribution model error, based on **SI** and **IP(Ker, Log)**, the SR is more efficient than the LS for the used degrees of freedom. As expected, these MSEs become comparable with the increase of the degree of freedom. See Table 2.

Finally, in Case 3, either under the contaminated normal distribution or the t -distribution model error, we have a similar observation as in Case 2. See Tables 5 and 6.

In this entire scenario, as expected, the MSEs increase as the rate of contamination increases for the contaminated normal distribution model error and decrease as the degrees of freedom increase for the t -distribution model error. Also, for the inverse marginal probability imputation, we note that the MSEs obtained based on the Gaussian kernel are generally smaller than those obtained based on the logistic propensity score. We suspect that this is due to the fact that the given logistic propensity score is not a density. It is worth pointing out that the higher the proportion of no missingness, the smaller the MSEs are for both SR and LS.

TABLE 3
MSEs × 10⁻³ of the SR & LS estimators against the proportion of missingness (η) & the rate of contamination (Case 1), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the contaminated normal distribution with different rates of contamination (ϵ)

		β_0					β_1				
$\epsilon \setminus \eta$.50	.60	.70	.80	.90	.50	.60	.70	.80	.90
LS-Ker	.00	96.80	73.42	59.74	49.37	41.14	1.57	0.85	0.43	0.48	0.23
	.01	125.15	98.90	80.59	66.36	58.62	1.69	1.12	0.83	0.49	0.29
	.05	154.66	127.89	104.91	85.23	74.28	2.14	1.42	1.05	0.70	0.42
	.10	188.25	154.36	128.07	107.96	92.55	2.19	1.76	1.09	0.84	0.45
	.15	207.78	168.21	145.56	124.71	111.54	2.77	1.88	1.31	0.85	0.64
	.25	234.93	194.67	168.18	138.10	122.11	2.99	1.93	1.55	0.94	0.79
SR-Ker	.00	89.04	67.59	59.59	49.39	42.00	0.82	0.49	0.33	0.26	0.21
	.01	98.41	76.05	63.24	54.77	50.55	0.84	0.56	0.42	0.34	0.24
	.05	112.02	90.06	73.41	63.29	54.67	1.04	0.66	0.51	0.37	0.32
	.10	127.93	104.75	87.37	75.01	62.78	1.17	0.84	0.56	0.48	0.33
	.15	144.49	111.15	94.99	81.86	71.53	1.41	0.98	0.68	0.52	0.39
	.25	157.39	125.01	107.29	88.30	78.78	1.68	1.04	0.73	0.56	0.49
LS-Log	.00	106.07	87.66	75.86	70.73	64.33	1.56	0.90	0.47	0.58	0.42
	.01	136.02	114.52	97.87	91.83	86.68	1.84	1.23	0.95	0.74	0.68
	.05	163.97	141.79	124.49	113.46	109.54	2.29	1.58	1.25	1.08	0.77
	.10	197.23	170.61	149.42	137.78	129.51	2.42	2.20	1.30	1.28	1.22
	.15	218.97	187.86	172.31	158.54	158.61	3.29	1.95	1.75	1.38	1.42
	.25	244.34	211.77	203.73	172.85	168.17	3.29	2.44	2.01	1.67	1.88
SR-Log	.00	89.43	67.76	60.14	51.51	44.97	0.79	0.51	0.34	0.31	0.28
	.01	98.23	76.24	63.59	55.73	53.34	0.84	0.57	0.45	0.38	0.30
	.05	111.94	90.37	74.20	64.36	58.33	1.08	0.70	0.52	0.43	0.40
	.10	127.97	105.35	87.57	76.41	65.16	1.26	0.91	0.59	0.54	0.46
	.15	144.75	111.24	95.69	82.56	74.60	1.48	1.06	0.71	0.59	0.47
	.25	156.52	123.45	107.76	88.85	81.41	1.75	1.10	0.74	0.64	0.67
LS-SI	.00	94.31	72.05	59.41	48.71	40.95	1.58	0.80	0.42	0.48	0.23
	.01	123.62	97.00	79.12	65.98	58.48	1.74	1.13	0.80	0.49	0.29
	.05	151.52	125.74	103.44	84.35	73.98	2.11	1.42	1.04	0.70	0.42
	.10	184.10	152.02	127.80	107.18	92.11	2.17	1.78	1.10	0.84	0.45
	.15	205.32	165.63	144.04	123.78	111.22	2.74	1.83	1.31	0.84	0.64
	.25	230.66	191.90	166.17	136.58	121.03	2.86	1.89	1.43	0.93	0.78
SR-SI	.00	88.71	67.89	59.63	49.30	41.93	0.78	0.50	0.33	0.26	0.21
	.01	98.26	75.83	63.24	54.77	50.44	0.84	0.55	0.41	0.34	0.24
	.05	112.13	90.68	73.38	63.13	54.68	1.06	0.70	0.51	0.39	0.31
	.10	127.43	104.59	87.43	74.84	62.55	1.22	0.85	0.56	0.48	0.33
	.15	144.07	110.65	94.51	82.16	71.70	1.48	0.92	0.69	0.51	0.39
	.25	157.31	124.70	107.28	88.09	78.40	1.71	1.02	0.72	0.56	0.49

Scenario 2: Considering either Case 2 or Case 3 under the Laplace distribution based either on **SI** or **IP(Ker, Log)**, it is observed that the SR outperforms its LS counterpart by providing smaller MSEs. Also, as expected, these MSEs decrease with increase of the sample size. See Table 7. A similar observation is made when considering these same two cases with high rates of missingness (Table 8).

TABLE 4
MSEs × 10⁻³ of the SR & LS estimators against the proportion of no missingness (η) & the rate of contamination (Case 1), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the *t*-distribution with different degrees of freedom (*df*)

<i>df</i> \ η	β_0					β_1					
	.50	.60	.70	.80	.90	.50	.60	.70	.80	.90	
LS-Ker	5	152.54	119.30	100.37	81.91	69.34	1.96	1.35	1.01	0.49	0.44
	10	118.18	91.21	74.79	63.18	54.42	1.91	1.23	0.93	0.48	0.38
	15	105.70	87.11	68.98	59.04	47.49	1.73	1.12	0.79	0.47	0.32
	20	108.59	81.00	66.50	56.02	46.48	1.56	1.10	0.78	0.43	0.29
	25	105.94	79.63	66.44	54.19	45.52	1.47	1.03	0.62	0.41	0.26
	30	105.17	77.77	63.68	53.40	46.95	1.37	0.88	0.57	0.36	0.24
SR-Ker	5	115.58	92.27	78.29	64.86	56.62	1.11	0.76	0.59	0.38	0.29
	10	100.06	79.93	67.79	58.27	50.31	0.88	0.55	0.45	0.32	0.27
	15	95.53	77.07	63.59	56.13	46.97	0.83	0.54	0.38	0.29	0.23
	20	94.24	73.64	62.96	53.79	46.10	0.82	0.50	0.37	0.29	0.23
	25	92.77	71.72	60.22	52.55	46.04	0.81	0.49	0.36	0.28	0.22
	30	91.54	70.98	60.02	52.32	44.91	0.77	0.49	0.36	0.27	0.21
LS-Log	5	163.36	135.86	121.59	112.03	101.47	2.08	1.47	1.33	0.86	0.91
	10	126.96	106.80	95.15	86.69	80.22	2.04	1.30	1.00	0.65	0.61
	15	118.73	101.48	85.34	81.76	73.53	1.80	1.23	0.99	0.63	0.47
	20	115.63	92.98	84.26	75.99	73.27	1.60	1.15	0.88	0.55	0.46
	25	114.13	92.88	82.36	75.90	69.50	1.51	1.15	0.71	0.55	0.44
	30	114.00	92.49	80.10	72.25	69.41	1.39	0.93	0.66	0.54	0.39
SR-Log	5	115.18	92.55	78.44	66.08	58.35	1.12	0.78	0.70	0.42	0.35
	10	99.49	80.01	68.27	59.21	52.90	0.89	0.58	0.47	0.35	0.34
	15	95.70	77.67	64.25	57.01	49.57	0.85	0.54	0.42	0.33	0.28
	20	93.60	73.20	63.69	54.90	48.84	0.83	0.52	0.41	0.33	0.27
	25	92.86	71.93	61.08	54.50	48.40	0.80	0.51	0.41	0.33	0.26
	30	91.37	71.70	61.00	53.09	47.30	0.80	0.47	0.39	0.29	0.26
LS-SI	5	150.66	119.78	97.00	80.48	73.43	2.21	1.59	0.88	0.71	0.43
	10	116.69	93.02	73.65	61.86	51.49	1.62	1.20	0.81	0.65	0.33
	15	108.24	86.29	67.91	58.33	47.77	1.55	1.10	0.69	0.56	0.32
	20	105.02	85.07	63.99	56.26	47.24	1.50	1.03	0.67	0.50	0.31
	25	104.70	79.86	63.12	56.08	46.47	1.44	0.96	0.62	0.47	0.28
	30	100.43	74.44	62.19	52.58	43.52	1.39	0.74	0.59	0.46	0.24
SR-SI	5	116.22	95.39	76.81	64.21	58.63	1.12	0.78	0.48	0.38	0.32
	10	101.24	79.47	65.01	56.30	49.04	0.89	0.63	0.48	0.32	0.26
	15	95.32	77.66	64.84	53.95	47.31	0.87	0.58	0.42	0.31	0.22
	20	93.51	77.44	60.22	53.26	46.15	0.86	0.56	0.38	0.29	0.21
	25	93.26	76.68	59.40	52.64	45.97	0.81	0.54	0.37	0.28	0.21
	30	88.84	71.57	59.26	51.39	43.54	0.72	0.47	0.37	0.27	0.20

Generally, while the MSEs of the SR obtained via **SI** and **IP(Ker)** for the two scenarios considered are comparable, those obtained via **IP(Log)** are larger. This suggests that for practical purposes the SR obtained via **SI** ($j = 1$) because of its simplicity in the way the imputation is carried out, can be preferred to that obtained via **IP** ($j = 2$). However, when it comes to estimating the variance, SR via **IP** is more flexible as the nonparametric estimator of $\Delta(x)$ can be

TABLE 5
MSEs $\times 10^{-3}$ of the SR & the LS against the rate of contamination (Case 3), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the contaminated normal distribution with different rates of contamination (ϵ)

ϵ	Ker		Log		SI		
	LS	SR	LS	SR	LS	SR	
β_0	0.00	58.846	58.605	60.269	58.598	41.609	44.469
	0.01	79.278	62.778	80.713	62.844	55.760	46.877
	0.05	104.257	74.415	105.193	74.482	73.139	56.066
	0.10	133.325	83.798	131.702	83.500	93.627	62.483
	0.15	147.074	92.000	146.636	91.807	107.250	71.289
	0.25	169.987	104.313	172.300	104.465	124.743	79.492
β_1	0.00	0.497	0.041	0.521	0.045	0.011	0.012
	0.01	0.544	0.048	0.547	0.052	0.021	0.016
	0.05	0.781	0.091	0.912	0.114	0.033	0.022
	0.10	0.781	0.127	0.931	0.139	0.050	0.030
	0.15	0.936	0.138	1.307	0.252	0.051	0.039
	0.25	1.218	0.284	1.645	0.296	0.083	0.050

TABLE 6
MSEs $\times 10^{-3}$ of the SR & LS estimators against the proportion of missingness (η) & the rate of contamination (Case 3), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the t -distribution with different degrees of freedom (df)

df	Ker		Log		SI		
	LS	SR	LS	SR	LS	SR	
β_0	5	97.161	75.048	552.834	546.500	63.241	52.920
	10	70.523	64.616	547.132	544.049	46.394	42.321
	15	67.284	62.889	541.121	541.101	41.983	40.568
	20	64.416	62.179	540.989	539.725	41.868	39.768
	25	63.717	60.880	540.488	535.668	40.094	38.993
	30	63.212	59.856	537.408	532.051	39.632	37.933
β_1	5	0.559	0.080	153.280	147.418	0.133	0.123
	10	0.461	0.060	148.269	146.704	0.097	0.089
	15	0.451	0.059	146.354	146.361	0.088	0.081
	20	0.256	0.054	145.802	146.099	0.088	0.079
	25	0.251	0.046	145.012	143.072	0.087	0.077
	30	0.214	0.045	143.964	142.513	0.082	0.075

further adjusted to result in a more efficient estimator of the variance.

Overall, from this simulation study, we see that the SR performs better under heavy-tailed error's distributions and for cases containing contaminations. Although more efficient than LS, even for the normal distribution model error when dealing with high rates of missingness, it is generally comparable to LS under normality ($\mathcal{CN}(0)$) as the proportion of missingness decreases, or under the t distribution model error as the degrees of freedom increase. This makes the SR estimation extremely appealing for situations where we encounter high rates of missing data.

TABLE 7

$MSEs \times 10^{-3}$ of the SR & LS estimators in scenario 2 (Case 2, Case 3), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the Laplace distribution with different sample sizes (n)

		ω_0				
		20	35	50	70	100
LS-Case2	SI	240.272	155.681	106.004	68.481	36.342
	Ker	247.936	154.942	103.985	67.809	36.621
	Log	338.420	175.772	113.542	71.108	39.629
SR-Case2	SI	52.203	22.352	13.127	8.641	5.639
	Ker	52.573	21.929	13.104	8.619	5.531
	Log	55.993	22.589	12.436	8.525	5.388
LS-Case3	SI	259.589	158.781	112.118	65.244	36.184
	Ker	266.596	159.838	110.956	64.025	35.582
	Log	278.192	176.343	113.576	69.985	38.804
SR-Case3	SI	55.177	22.989	13.027	8.475	5.366
	Ker	52.630	22.623	13.007	8.352	5.362
	Log	53.523	21.728	13.116	8.140	5.355

TABLE 8

$MSEs \times 10^{-3}$ of the SR & LS estimators of ω_0 in scenario 2 (Case 2, Case 3), under simple imputation (SI) and inverse marginal probability (IP) based on the kernel (Ker) and the logistic function (Log) for the Laplace distribution with different sample sizes (n)

		ω_0					
		Case 2			Case 3		
		50	70	100	50	70	100
LS	SI	354.876	318.222	299.459	327.135	310.413	262.084
	Ker	354.153	318.931	300.843	329.611	306.509	264.282
	Log	396.379	349.802	322.074	335.332	345.022	257.676
SR	SI	267.205	243.276	199.536	220.961	192.445	146.080
	Ker	231.695	217.298	173.874	184.401	149.142	105.981
	Log	254.874	236.600	191.041	210.110	176.385	136.561

7. Conclusion

This paper provides asymptotic properties (strong consistency and asymptotic normality) of the signed-rank estimator under the assumption that some responses in the regression model are missing at random. Simulation studies demonstrate the performance of SR when dealing with heavy-tailed or contaminated model error's distributions and even for the normal model error's distribution when dealing with a high rate of missing responses.

Acknowledgements

The author is grateful to the anonymous referees and to the editor for their valuable comments and suggestions that helped to improve the earlier version of the manuscript.

Appendix A: Proofs

This appendix provides sketches of proofs of some results stated in this paper.

Proof of Theorem 2.1. In this proof, L is taken to be an arbitrary positive constant. Set $b_{nij} = \tilde{R}_{ij}/(n + 1)$ and $a_{nij} = R_{ij}/(n + 1)$, and without loss of generality, assume that $a_{nij} < b_{nij}$. By the mean value theorem, there exists α_{nij} such that $\varphi(b_{nij}) - \varphi(a_{nij}) = \varphi'(\alpha_{nij})(b_{nij} - a_{nij})$.

$$\begin{aligned} |\tilde{D}_{jn}(\theta) - D_{jn}(\theta)| &= \left| \frac{1}{n} \sum_{i=1}^n \left[\varphi(b_{nij})|\tilde{z}_{ij}(\theta)| - \varphi(a_{nij})|z_{ij}(\theta)| \right] \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n \left[\varphi(b_{nij})|\tilde{z}_{ij}(\theta)| \right. \right. \\ &\quad \left. \left. + \{ \varphi'(\alpha_{nij})(b_{nij} - a_{nij}) - \varphi(b_{nij}) \} |z_{ij}(\theta)| \right] \right| \end{aligned}$$

Also, by uniform continuity of φ (since φ' is bounded with bound $L > 0$) and for fixed n , one can choose $\epsilon(n) = \xi(n)/(L + 1) > 0$, such that for $\xi(n) \rightarrow 0$ as $n \rightarrow \infty$ and

$$|a_{nij} - b_{nij}| \leq \xi(n) \implies |\varphi(a_{nij}) - \varphi(b_{nij})| < \epsilon(n).$$

From this, we have,

$$\begin{aligned} |\tilde{D}_{jn}(\theta) - D_{jn}(\theta)| &\leq \frac{1}{n} \sum_{i=1}^n |\varphi(b_{nij})| \left| |\tilde{z}_{ij}(\theta)| - |z_{ij}(\theta)| \right| \\ &\quad + \frac{1}{n} \sum_{i=1}^n |\varphi'(\alpha_{nij})| |a_{nij} - b_{nij}| |z_{ij}(\theta)| \end{aligned}$$

From the boundedness of φ , there exists a constant $L > 0$ such that $|\varphi(b_{nij})| \leq L$. Then,

$$|\tilde{D}_{jn}(\theta) - D_{jn}(\theta)| \leq \frac{L}{n} \sum_{i=1}^n \left| |\tilde{z}_{ij}(\theta)| - |z_{ij}(\theta)| \right| + \xi(n) \frac{L}{n} \sum_{i=1}^n |z_{ij}(\theta)|$$

Note that $\left| |\tilde{z}_{ij}(\theta)| - |z_{ij}(\theta)| \right| \leq |\tilde{z}_{ij}(\theta) - z_{ij}(\theta)|$ and

$$\tilde{z}_{ij}(\theta) - z_{ij}(\theta) = \begin{cases} 0 & , \quad \text{if } j = 1 \\ \left(\frac{1}{\hat{\Gamma}(\mathbf{x}_i)} - \frac{1}{\Delta(\mathbf{x}_i)} \right) \delta_i \varepsilon_i(\theta) & , \quad \text{if } j = 2. \end{cases}$$

Then,

$$|\tilde{D}_{jn}(\theta) - D_{jn}(\theta)| \leq \begin{cases} \xi(n) \frac{L}{n} \sum_{i=1}^n |z_{ij}(\theta)|, & \text{if } j = 1 \\ \frac{L}{n} \sum_{i=1}^n \left| \frac{1}{\widehat{\Gamma}(\mathbf{x}_i)} - \frac{1}{\Delta(\mathbf{x}_i)} \right| \delta_i |\varepsilon_i(\theta)| + \xi(n) \frac{L}{n} \sum_{i=1}^n |z_{ij}(\theta)|, & \text{if } j = 2. \end{cases}$$

This implies that

$$\sup_{\theta \in \Theta} |\tilde{D}_{jn}(\theta) - D_{jn}(\theta)| \leq \begin{cases} \xi(n) \frac{L}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} |z_{ij}(\theta)|, & \text{if } j = 1 \\ \frac{L}{n} \sum_{i=1}^n \left| \frac{1}{\widehat{\Gamma}(\mathbf{x}_i)} - \frac{1}{\Delta(\mathbf{x}_i)} \right| \delta_i \sup_{\theta \in \Theta} |\varepsilon_i(\theta)| + \xi(n) \frac{L}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} |z_{ij}(\theta)|, & \text{if } j = 2. \end{cases}$$

$$z_{ij}(\theta) = \begin{cases} \delta_i \varepsilon_i(\theta), & \text{if } j = 1 \\ \frac{\delta_i}{\Delta(\mathbf{x}_i)} \varepsilon_i(\theta), & \text{if } j = 2. \end{cases}$$

f being continuous on a compact set Θ , we have $\sup_{\theta \in \Theta} f(\mathbf{x}_i, \theta) < \infty$, and then, $E[\sup_{\theta \in \Theta} |z_{ij}(\theta)| \mid \mathbf{x}_i] < \infty$ since $\text{var}(\varepsilon_i \mid \mathbf{x}_i) < \infty$. Now, under the MAR assumption and conditioning on \mathbf{x}_i , the Strong Law of Large Numbers implies that

$$\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} |z_{ij}(\theta)| \rightarrow E\left[\sup_{\theta \in \Theta} |z_{ij}(\theta)| \mid \mathbf{x}_i\right] < \infty \quad \text{a.s. as } n \rightarrow \infty$$

and then,

$$\xi(n) \frac{L}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} |z_{ij}(\theta)| \rightarrow 0 \quad \text{since } \xi(n) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

On the other hand, note that a direct application of the Dominated Convergence Theorem gives

$$E\left[\left|\frac{1}{\widehat{\Gamma}(\mathbf{x}_i)} - \frac{1}{\Delta(\mathbf{x}_i)}\right| \delta_i \sup_{\theta \in \Theta} |\varepsilon_i(\theta)| \mid \mathbf{x}_i\right] \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

since

$$\left|\frac{1}{\widehat{\Gamma}(\mathbf{x}_i)} - \frac{1}{\Delta(\mathbf{x}_i)}\right| \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

Hence, again by the Strong Law of Large Numbers,

$$\frac{L}{n} \sum_{i=1}^n \left|\frac{1}{\widehat{\Gamma}(\mathbf{x}_i)} - \frac{1}{\Delta(\mathbf{x}_i)}\right| \delta_i \sup_{\theta \in \Theta} |\varepsilon_i(\theta)| \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

Consequently,

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\tilde{D}_{jn}(\theta) - D_{jn}(\theta)| \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

□

Proof of Theorem 3.1. As pointed out in [48], for the result to follow, it is sufficient to show that

$$\lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (\tilde{D}_{jn}(\theta) - \tilde{D}_{jn}(\theta_0)) > 0 \quad a.s.$$

Note that

$$\tilde{D}_{jn}(\theta) - \tilde{D}_{jn}(\theta_0) = \tilde{D}_{jn}(\theta) - D_{jn}(\theta) + D_{jn}(\theta) - D_{jn}(\theta_0) + D_{jn}(\theta_0) - \tilde{D}_{jn}(\theta_0)$$

and so,

$$\begin{aligned} \lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (\tilde{D}_{jn}(\theta) - \tilde{D}_{jn}(\theta_0)) &\geq \lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (\tilde{D}_{jn}(\theta) - D_{jn}(\theta)) \\ &+ \lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (D_{jn}(\theta) - D_{jn}(\theta_0)) \\ &+ \lim_{n \rightarrow \infty} (D_{jn}(\theta_0) - \tilde{D}_{jn}(\theta_0)) \end{aligned}$$

From the fact that $B^c \subset \Theta$, by Theorem 2.1, we have

$$\lim_{n \rightarrow \infty} (D_{jn}(\theta_0) - \tilde{D}_{jn}(\theta_0)) = \lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (\tilde{D}_{jn}(\theta) - D_{jn}(\theta)) = 0 \quad a.s.$$

To complete the proof, it is sufficient to show that

$$\lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (D_{jn}(\theta) - D_{jn}(\theta_0)) > 0 \quad a.s.$$

To this end, decompose $D_{jn}(\theta) - D_{jn}(\theta_0)$ as

$$D_{jn}(\theta) - D_{jn}(\theta_0) = D_{jn}(\theta) - \mu_j(\theta) + \mu_j(\theta) - \mu_j(\theta_0) + \mu_j(\theta_0) - D_{jn}(\theta_0)$$

and so,

$$\begin{aligned} \lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (D_{jn}(\theta) - D_{jn}(\theta_0)) &\geq \lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (D_{jn}(\theta) - \mu_j(\theta)) \\ &+ \inf_{\theta \in B^c} (\mu_j(\theta) - \mu_j(\theta_0)) \\ &+ \lim_{n \rightarrow \infty} (\mu_j(\theta_0) - D_{jn}(\theta_0)) \end{aligned}$$

Now by Lemma 3.1,

$$\lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (D_{jn}(\theta) - \mu_j(\theta)) = \lim_{n \rightarrow \infty} (\mu_j(\theta_0) - D_{jn}(\theta_0)) = 0$$

and,

$$\inf_{\theta \in B^c} (\mu_j(\theta) - \mu_j(\theta_0)) > 0.$$

Hence

$$\lim_{n \rightarrow \infty} \inf_{\theta \in B^c} (D_{jn}(\theta) - D_{jn}(\theta_0)) > 0 \quad a.s.$$

which completes the proof. \square

Proof of Theorem 4.1. As in the proof of Theorem 2.1, set $b_{nij} = \tilde{R}_{ij}/(n+1)$ and $b_{nij}^* = \tilde{R}_{ij}^*/(n+1)$, where \tilde{R}_{ij}^* is the i^{th} rank of $|\lambda_{ij}e_i|$. Then,

$$\tilde{S}_{jn}(\theta) - \tilde{S}_{jn}(\theta_0) = \frac{1}{n} \sum_{i=1}^n [\varphi(b_{nij})\lambda_{ij}\nabla_{\theta}f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i) - \varphi(b_{nij}^*)\lambda_{ij}\nabla_{\theta_0}f(\mathbf{x}_i) \times \text{sgn}(e_i)]. \quad (\text{A.1})$$

Now without loss of generality, assume that $b_{nij} > b_{nij}^*$. Then, by the mean value theorem, φ being differentiable with bounded derivative, we have

$$\varphi(b_{nij}) - \varphi(b_{nij}^*) = \varphi'(\alpha_{nij})(b_{nij} - b_{nij}^*) \quad (\text{A.2})$$

where $\alpha_{nij} \in (b_{nij}^*, b_{nij})$. On the other hand, f being twice continuously differentiable, again by the mean value theorem, we have

$$\nabla_{\theta}f(\mathbf{x}_i) - \nabla_{\theta_0}f(\mathbf{x}_i) = (\nabla_{\xi}^2f(\mathbf{x}_i))^T(\theta - \theta_0). \quad (\text{A.3})$$

To this end, (A.2) and (A.3) in (A.1) give

$$\begin{aligned} \tilde{S}_{jn}(\theta) - \tilde{S}_{jn}(\theta_0) &= \frac{1}{n} \sum_{i=1}^n \varphi(b_{nij}^*)\lambda_{ij}\nabla_{\theta_0}f(\mathbf{x}_i) \times [\text{sgn}(\varepsilon_i) - \text{sgn}(e_i)] \\ &+ \frac{1}{n} \sum_{i=1}^n \varphi(b_{nij}^*)\lambda_{ij}(\nabla_{\xi}^2f(\mathbf{x}_i))^T(\theta - \theta_0) \times \text{sgn}(\varepsilon_i) \\ &+ \frac{1}{n(n+1)} \sum_{i=1}^n \lambda_{ij}\varphi'(\alpha_{nij})(\tilde{R}_{ij} - \tilde{R}_{ij}^*)\nabla_{\theta}f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i) \end{aligned}$$

and so,

$$\begin{aligned} &\sqrt{n}(\tilde{S}_{jn}(\theta) - \tilde{S}_{jn}(\theta_0) - \mathbf{G}_n^T(\theta - \theta_0)) \\ &= \frac{\sqrt{n}}{n} \sum_{i=1}^n \varphi(b_{nij}^*)\lambda_{ij}\nabla_{\theta_0}f(\mathbf{x}_i) \times [\text{sgn}(\varepsilon_i) - \text{sgn}(e_i)] \\ &+ \frac{1}{\sqrt{n}(n+1)} \sum_{i=1}^n \lambda_{ij}\varphi'(\alpha_{ij})(\tilde{R}_{ij} - \tilde{R}_{ij}^*)\nabla_{\theta}f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i) \end{aligned}$$

By assumptions $E(\varepsilon|\mathbf{x}) = E(e|\mathbf{x}) = 0$. Also, $\text{var}(\varepsilon|\mathbf{x}) = \text{var}(e|\mathbf{x})$. Then ε and e have the same distribution. Now taking into consideration that $\infty \times 0 = 0$, we have,

$$E[\varphi(b_{nij}^*)\lambda_{ij}\nabla_{\theta_0}f(\mathbf{x}_i) \times (\text{sgn}(\varepsilon_i) - \text{sgn}(e_i))] = 0$$

and by the Strong Law of Large numbers for functions of order statistics, we get

$$\frac{1}{n} \sum_{i=1}^n \varphi(b_{nij}^*)\lambda_{ij}\nabla_{\theta_0}f(\mathbf{x}_i) \times [\text{sgn}(\varepsilon_i) - \text{sgn}(e_i)] \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

Thus,

$$\frac{\sqrt{n}}{n} \sum_{i=1}^n \varphi(b_{nij}^*) \lambda_{ij} \nabla_{\theta_0} f(\mathbf{x}_i) \times [\text{sgn}(\varepsilon_i) - \text{sgn}(e_i)] \rightarrow \infty \times 0 = 0 \quad a.s. \quad \text{as } n \rightarrow \infty. \tag{A.4}$$

On the other hand,

$$\begin{aligned} & \sup_{\theta \in B} \left| \frac{1}{\sqrt{n}(n+1)} \sum_{i=1}^n \lambda_{ij} \varphi'(\alpha_{ij}) (\tilde{R}_{ij} - \tilde{R}_{ij}^*) \nabla_{\theta} f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i) \right| \\ & \leq \frac{2M}{\sqrt{n}(n+1)} \sum_{i=1}^n \lambda_{ij} \tilde{R}_{ij} \sup_{\theta \in B} \|\nabla_{\theta} f(\mathbf{x}_i)\|_1 \\ & \leq \frac{2M}{\sqrt{n}(n+1)} \sum_{i=1}^n \lambda_{ij} \tilde{R}_{ij} J(\mathbf{x}_i) \end{aligned}$$

where $|\varphi'(\alpha_{ij})| \leq M$ for some positive constant M , J is an integrable function of \mathbf{x} such that $\|\nabla_{\theta} f(\mathbf{x}_i)\|_1 \leq J(\mathbf{x})$ for any $\theta \in B$ and $\|\cdot\|_1$ represents the L_1 -norm; that is, $\int J(\mathbf{x}) dH(\mathbf{x}) < \infty$. Again, applying the strong law of large numbers for functions of order statistics [43],

$$\frac{1}{n} \sum_{i=1}^n \lambda_{ij} \tilde{R}_{ij} J(\mathbf{x}_i) \rightarrow \int \int J(\mathbf{x}) dF(\varepsilon) dH(\mathbf{x}) < \infty \quad a.s. \quad \text{as } n \rightarrow \infty.$$

where

$$F(\varepsilon) = \begin{cases} \delta G(\varepsilon), & \text{if } j = 1 \\ \frac{\delta}{\Delta(\mathbf{x})} G(\varepsilon), & \text{if } j = 2. \end{cases}$$

and so,

$$\begin{aligned} & \frac{2M}{\sqrt{n}(n+1)} \sum_{i=1}^n \lambda_{ij} \tilde{R}_{ij} J(\mathbf{x}_i) \rightarrow 0 \quad a.s. \quad \text{as } n \rightarrow \infty \\ & \lim_{n \rightarrow \infty} \sup_{\theta \in B} \left| \frac{1}{\sqrt{n}(n+1)} \sum_{i=1}^n \lambda_{ij} \varphi'(\alpha_{ij}) (\tilde{R}_{ij} - \tilde{R}_{ij}^*) \nabla_{\theta} f(\mathbf{x}_i) \times \text{sgn}(\varepsilon_i) \right| = 0 \quad a.s. \end{aligned} \tag{A.5}$$

Now combining (A.4) and (A.5), we obtain

$$\lim_{n \rightarrow \infty} \sup_{\theta \in B} \left| \sqrt{n} (\tilde{S}_{jn}(\theta) - \tilde{S}_{jn}(\theta_0) - \mathbf{G}_n^T(\theta - \theta_0)) \right| = 0 \quad a.s.$$

which completes the proof. □

Proof of Theorem 4.3. From (4.3) and for n large enough, we have

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = (\sqrt{n}\mathbf{G}_n)^{-1} \tilde{S}_{jn}(\theta_0) \approx \mathcal{W}^{-1} \tilde{S}_{jn}(\theta_0).$$

Combining this to Theorem 4.2, we complete the proof. □

References

- [1] ABEBE, A., MCKEAN, J. W. and BINDELE, H. F. (2012). On the consistency of a class of nonlinear regression estimators. *PJSOR* **8** 543–555. [MR2975218](#)
- [2] BINDELE, H. F. (2014). Asymptotics of the signed-rank estimator under dependent observations. *Journal of Statistical Planning and Inference* **146** 49–55. [MR3132478](#)
- [3] BINDELE, H. F. and ABEBE, A. (2012). Bounded influence nonlinear signed-rank regression. *Canadian Journal of Statistics* **40** 172–189. [MR2896936](#)
- [4] BRUNNER, E. and DENKER, M. (1994). Rank statistics under dependent observations and applications to factorial designs. *Journal of Statistical Planning and Inference* **42** 353–378. [MR1309630](#)
- [5] CHEN, J. and SHAO, J. (2001). Jackknife variance estimation for nearest-neighbor imputation. *Journal of the American Statistical Association* **96** 260–269. [MR1952736](#)
- [6] CHEN, S. X. and HALL, P. (1993). Smoothed Empirical Likelihood Confidence Intervals for Quantiles. *The Annals of Statistics* **21** pp. 1166–1181. [MR1241263](#)
- [7] CHENG, P. E. (1994). Nonparametric Estimation of Mean Functionals with Data Missing at Random. *Journal of the American Statistical Association* **89** pp. 81–87.
- [8] CIUPERCA, G. (2011). Empirical likelihood for nonlinear models with missing responses. *Journal of Statistical Computation and Simulation* **0** 1–20.
- [9] CULP, M. and MICHAELIDIS, G. (2008). An iterative algorithm for extending learners to a semi-supervised setting. *Journal of Computational and Graphical Statistics* **17** 545–571. [MR2451341](#)
- [10] DELECROIX, M., HRISTACHE, M. and PATILEA, V. (2006). On semiparametric-estimation in single-index regression. *Journal of Statistical Planning and Inference* **136** 730–769. [MR2181975](#)
- [11] DEVILLE, J.-C. and SÄRNDAL, C.-E. (1994). Variance estimation for the regression imputed Horvitz-Thompson estimator. *Journal of Official Statistics-Stockholm-* **10** 381–381.
- [12] EINMAHL, U. and MASON, D. M. (2005). Uniform in Bandwidth Consistency of Kernel-Type Function Estimators. *The Annals of Statistics* **33** pp. 1380–1403. [MR2195639](#)
- [13] HALL, P. and LA SCALA, B. (1990). Methodology and Algorithms of Empirical Likelihood. *International Statistical Review* **58** 109–127.
- [14] HAUPT, H. and OBERHOFER, W. (2009). On asymptotic normality in nonlinear regression. *Statist. Probab. Lett.* **79** 848–849. [MR2724695](#)
- [15] HEALY, M. and WESTMACOTT, M. (1956). Missing Values in Experiments Analysed on Automatic Computers. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **5** pp. 203–206.
- [16] HETTMANSPERGER, T. P. and MCKEAN, J. W. (1998). *Robust non-parametric statistical methods*. *Kendall's Library of Statistics* **5**. Edward Arnold, London. [MR1604954 \(99c:62075\)](#)

- [17] JURECKOVA, J. (1971). Nonparametric Estimate of Regression Coefficients. *The Annals of Mathematical Statistics* **42** pp. 1328–1338. [MR0295487](#)
- [18] JUREČKOVÁ, J. (2008). Regression rank scores in nonlinear models. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Inst. Math. Stat. Collect.* **1** 173–183. Inst. Math. Statist., Beachwood, OH. [MR2462205](#) (2010d:62099)
- [19] KALTON, G. (1983). Compensating for missing survey data.
- [20] KIM, J. K. and SHAO, J. (2014). *Statistical methods for handling incomplete data*. Chapman & Hall/CRC, Boca Raton, FL. [MR3307946](#)
- [21] KITAMURA, Y. (1997). Empirical Likelihood Methods with Weakly Dependent Processes. *The Annals of Statistics* **25** pp. 2084–2102. [MR1474084](#)
- [22] LESSLER, J. T. and KALSBECK, W. D. (1992). *Nonsampling error in surveys. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. John Wiley & Sons, Inc., New York A Wiley-Interscience Publication. [MR1193229](#) (94a:62014)
- [23] LITTLE, R. J. A. and RUBIN, D. B. (1987). *Statistical analysis with missing data. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. John Wiley & Sons Inc., New York. [MR0890519](#) (88i:62042)
- [24] LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical analysis with missing data*, second ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. [MR1925014](#) (2003g:62007)
- [25] MUKHERJEE, K. (1999). Asymptotics of Quantiles and Rank Scores in Nonlinear Time Series. *Journal of Time Series Analysis* **20** 173–192. [MR1701053](#)
- [26] MÜLLER, U. U. (2009). Estimating linear functionals in nonlinear regression with responses missing at random. *Ann. Statist.* **37** 2245–2277. [MR2543691](#) (2010k:62265)
- [27] NORDHOLT, E. S. (1998). Imputation: methods, simulation experiments and practical examples. *International Statistical Review* **66** 157–180.
- [28] OBERHOFER, W. (1982). The Consistency of Nonlinear Regression Minimizing the L1-Norm. *The Annals of Statistics* **10** pp. 316–319. [MR0642745](#)
- [29] OWEN, A. (1990). Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics* **18** pp. 90–120. [MR1041387](#)
- [30] PENG, L. (2004). Empirical-Likelihood-Based Confidence Interval for the Mean with a Heavy-Tailed Distribution. *The Annals of Statistics* **32** pp. 1192–1214. [MR2065202](#)
- [31] QIN, Y., LI, L. and LEI, Q. (2009). Empirical likelihood for linear regression models with missing responses. *Statistics & Probability Letters* **79** 1391–1396. [MR2750886](#)
- [32] RAO, J. N. and SITTER, R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika* **82** 453–460. [MR1354243](#)
- [33] RAO, J. N. K. and SHAO, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79** 811–822. [MR1209480](#) (94c:62023)

- [34] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association* **89** pp. 846–866. [MR1294730](#)
- [35] RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. [MR0455196](#)
- [36] RUBIN, D. B. (2004). *Multiple imputation for nonresponse in surveys*. *Wiley Classics Library*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ Reprint of the 1987 edition [John Wiley & Sons, Inc., New York; MR899519]. [MR2117498](#)
- [37] SCHAFER, J. L. (1997). *Analysis of incomplete multivariate data*. *Mono-graphs on Statistics and Applied Probability* **72**. Chapman & Hall, London. [MR1692799](#) ([2000d:62003](#))
- [38] SCHAFER, J. L. and GRAHAM, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods* **7** 147.
- [39] SHAO, J. and SITTER, R. R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association* **91** 1278–1288. [MR1424624](#)
- [40] SHAO, J. and STEEL, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association* **94** 254–265. [MR1689230](#)
- [41] SKINNER, C. J. and RAO, J. (2002). Jackknife variance estimation for multivariate statistics under hot-deck imputation from common donors. *Journal of Statistical Planning and Inference* **102** 149–167. [MR1885199](#)
- [42] SUN, Z., WANG, Q. and DAI, P. (2009). Model checking for partially linear models with missing responses at random. *J. Multivar. Anal.* **100** 636–651. [MR2478187](#)
- [43] VAN ZWET, W. R. (1980). A Strong Law for Linear Functions of Order Statistics. *The Annals of Probability* **8** pp. 986–990. [MR0586781](#)
- [44] WANG, C. Y., WANG, S., ZHAO, L.-P. and OU, S.-T. (1997). Weighted Semiparametric Estimation in Regression Analysis With Missing Covariate Data. *Journal of the American Statistical Association* **92** pp. 512–525. [MR1467845](#)
- [45] WANG, Q., LINTON, O. and HÄRDLE, W. (2004). Semiparametric Regression Analysis with Missing Response at Random. *Journal of the American Statistical Association* **99** pp. 334–345. [MR2062820](#)
- [46] WANG, Q. and RAO, J. N. K. (2002). Empirical Likelihood-Based Inference under Imputation for Missing Response Data. *The Annals of Statistics* **30** pp. 896–924. [MR1922545](#)
- [47] WANG, Q. and SUN, Z. (2007). Estimation in partially linear models with missing responses at random. *Journal of Multivariate Analysis* **98** 1470–1493. [MR2364130](#)
- [48] WU, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513. [MR0615427](#) ([83c:62101](#))
- [49] XUE, L. (2009). Empirical Likelihood Confidence Intervals for Response Mean with Data Missing at Random. *Scandinavian Journal of Statistics* **36** 671–685. [MR2573302](#)

- [50] XUE, L. and ZHU, L. (2007). Empirical Likelihood Semiparametric Regression Analysis for Longitudinal Data. *Biometrika* **94** 921–937. [MR2416799](#)
- [51] XUE, L. and ZHU, L. (2007). Empirical Likelihood for a Varying Coefficient Model With Longitudinal Data. *Journal of the American Statistical Association* **102** 642–654. [MR2370858](#)
- [52] XUE, L.-G. and ZHU, L. (2006). Empirical likelihood for single-index models. *Journal of Multivariate Analysis* **97** 1295–1312. [MR2279674](#)
- [53] YUNG, W. and RAO, J. (2000). Jackknife variance estimation under imputation for estimators using poststratification information. *Journal of the American Statistical Association* **95** 903–915. [MR1804447](#)
- [54] ZHAO, L. P., LIPSITZ, S. and LEW, D. (1996). Regression Analysis with Missing Covariate Data Using Estimating Equations. *Biometrics* **52** pp. 1165–1182.