Computational Statistics and Data Analysis xxx (xxxx) xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Model checking for general linear regression with nonignorable missing response[☆]

Xu Guo[a], Lianlian Song[b], Yun Fang[c,*], Lixing Zhu[a,d]

[a] School of Statistics, Beijing Normal University, Beijing, China
[b] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[c] Department of Mathematics, Shanghai Normal University, Shanghai, China
[d] Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

ARTICLE INFO

Article history:
Received 19 January 2018
Received in revised form 14 March 2019
Accepted 16 March 2019
Available online xxxx

Keywords:
Inverse probability weighting
Nonignorable missing response
Model checking

ABSTRACT

Model checking for the general linear regression model with nonignorable missing response is studied. Based on an exponential tilting model, two estimators are proposed for the unknown parameter in the regression model. Then, two empirical-process-based tests are constructed. The asymptotic properties of the proposed tests are investigated under the null and local alternative hypotheses in different scenarios. It is found that the two tests perform identically in the asymptotic sense. In addition, a nonparametric Monte Carlo test procedure is performed to obtain the critical values. Further, simulation studies are conducted to assess the performance of the proposed tests and compare them with other possible approaches. Finally, a real data set is analyzed for illustration.

1. Introduction

Due to its easy interpretation and well developed theories, the linear regression model is widely used to describe the relationship between scalar response Y and covariates X of dimension q. Consider the following general linear regression model with the form

$$Y = \phi^{\top}(X)\beta + \varepsilon, \tag{1.1}$$

where $\phi(\cdot)$ is a known smooth function with dimension p, and β is an unknown parameter to be estimated. Furthermore, ε is the error term satisfying $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = \sigma^2(X) < \infty$. The superscript $\top$ in (1.1) denotes the transpose. When $\phi(X) \equiv X$, model (1.1) becomes the classical linear model. Compared with the classical linear model, model (1.1) is more flexible and applicable because interaction and high-order terms of the covariates can be included.

To prevent incorrect conclusions and improve interpretations, any statistical analysis conducted using model (1.1) should be accompanied by a check of whether the hypothetical model holds. The literature contains a number of proposals for doing so when all the response measurements are available (González-Manteiga and Crujeiras, 2013; Guo and Zhu, 2017 for a review). However, missing responses are often encountered in practice. For instance, the response Y may be very expensive to measure, and, because of financial limitations, the response values are available for only some of the subjects; the sampled individuals may have refused to supply the desired information to some survey questions, or the

☆ The authors thank the editor, the associate editor and the anonymous referee for their constructive comments and suggestions that substantially improved an early manuscript.

* Corresponding author.
E-mail address: fangyun0919@shnu.edu.cn (Y. Fang).

https://doi.org/10.1016/j.csda.2019.03.009
0167-9473/

investigators may have failed to gather the correct information. Simply excluding the units with missing responses and conducting statistical analysis based on only the completely observed samples can often lead to biased and inefficient parameter estimates when the data are not missing completely at random (Little and Rubin, 1987; Kim and Shao, 2013). Consequently, we cannot directly extend the existing model checking procedures for complete data to deal with model checking with missing responses.

It is now realized that to develop a more accurate and useful methodology in the presence of missing response values, we often need to make some assumptions regarding the missing mechanism. Here we simply present the related concepts; for more discussions, see Ibrahim et al. (2001, 2005). Let δ be a missing indicator of whether Y is observed ($\delta = 1$) or not ($\delta = 0$). If given the fully observed covariates X the response Y is independent of the missing indicator δ, then the response is missing at random (MAR) or ignorable. If the missing depends on the value of the underlying unobserved response, then we call it nonignorable or not missing at random (NMAR). The NMAR case is common in social surveys, especially when the questions are sensitive. Consider a survey of personal income, to which people in low socioeconomic groups are more likely to refuse to provide the desired information.

The literature contains some proposals for model checking when the response value Y is MAR. González-Manteiga and Pérez-González (2006) extended the method due to Härdle and Mammen (1993), with the constructed test statistic being based on the L_2 distance between the nonparametric and parametric fits. For the general linear model (1.1) with MAR response, Sun and Wang (2009) imputed the incomplete observations by regression imputation and inverse probability weighting and then constructed two score-type tests and two empirical-process tests with the completed samples. Li (2012) proposed a test that is based on the minimum integrated squared distances between nonparametric and parametric fits, which can be viewed as an extension of the minimum-distance test proposed by Koul and Ni (2004). See also Guo et al. (2015), which extended the quadratic form conditional moment test due to Zheng (1996).

However, the aforementioned studies are limited to MAR response, and it would be difficult to extend these methods to nonignorable missing response scenarios because the missingness is also related to the unobserved responses. To deal with nonignorable missing response, some authors have assumed parametric missing mechanisms (e.g., Qin et al., 2002; Kott and Chang, 2010; Morikawa et al., 2017). For nonignorable missing response, the missing mechanism cannot be verified from the observed variables alone. It is then desirable to make the weakest possible model assumptions about the missing mechanism. To this aim, Kim and Yu (2011) proposed a semiparametric exponential tilting model for the missing mechanism. Based on the semiparametric missing model, they imputed the missing responses by nonparametric kernel regression imputation, whereupon the sample average of the observed and imputed responses is used to estimate the mean function. Based on regression imputation and augmented inverse probability weighting, Zhao et al. (2013) used the empirical likelihood method introduced by Owen (1988) to construct the confidence interval of the mean function. Niu et al. (2014) discussed the confidence interval for the parameters in a linear regression model with nonignorable missing response. Tang et al. (2014) investigated the inference on parameters in estimating equations by adopting the empirical likelihood. Zhong and Chen (2014) developed jackknife empirical likelihood inference for nonignorable missing response. Bindele and Zhao (2018) proposed rank-based estimators for a general parametric regression model. Other recently developed estimation approaches for nonignorable missing data include those of Wang et al. (2014), Shao and Wang (2016), and Zhao et al. (2017). For an up-to-date review of nonignorable missing data, see Tang and Ju (2018).

In this paper, we aim to perform a model checking procedure for model (1.1) with nonignorable missing response and completely observed covariates. In other words, we want to test the following hypothesis:

$$H_0 : E(Y|X) = \phi^\top(X)\beta \tag{1.2}$$

for some β and known $\phi(\cdot)$ with nonignorable missing response.

We begin by discussing how to estimate the unknown parameter β under the null hypothesis, and we suggest two estimators for β. The first is based on an imputed complete data set and the second is based on inverse probability weighting. We then construct two residual-marked empirical-process-based test statistics. Based on the exponential tilting model, we study in detail the asymptotic properties of our proposed tests under different situations.

The rest of this paper is organized as follows. In Section 2, we construct the test statistics. In Section 3, we derive their asymptotic properties under the null hypothesis and local alternative hypotheses when the tilting parameter is known or estimated. In Section 4, we report simulation results and analyze real data to illustrate the proposed tests. In Section 5, we make some concluding remarks. The conditions are described in the Appendix. The proofs of the theoretical results are given in the supplementary material.

2. Construction of test statistics

Denote by $P(\delta = 1|Y, X) = \pi(X, Y)$ the selection probability function, which is the probability that a unit is observed, given X and Y. Because the missing mechanism $\pi(X, Y)$ plays a key role in the analysis of nonignorable missing data, we begin with a short review of the methodology developed by Kim and Yu (2011).

Assume that $\pi(x_i, y_i)$ is a semiparametric logistic regression model in the form

$$\pi(x_i, y_i) = \frac{\exp\{g(x_i) - \gamma y_i\}}{1 + \exp\{g(x_i) - \gamma y_i\}}. \tag{2.1}$$

Here, $g(\cdot)$ is an unknown function and γ is an unknown parameter. Clearly, MAR can be viewed as a special case of model (2.1) by setting $\gamma = 0$.

Define

$$O(x_i, y_i) = \frac{P(\delta_i = 0|x_i, y_i)}{P(\delta_i = 1|x_i, y_i)} = \frac{1}{\pi(x_i, y_i)} - 1$$

as the conditional odds of nonresponse. Under model (2.1), we can rewrite the odds function as

$$O(x_i, y_i) = \exp\{-g(x_i) + \gamma y_i\}.$$

Define $\alpha(X; \gamma) := \exp\{-g(X)\} = O(X, Y)/\exp(\gamma Y)$, whereupon we have

$$\alpha(X; \gamma)E\{\delta \exp(\gamma Y)|X\} = E\{\delta O(X, Y)|X\} = E(1 - \delta|X).$$

Based on the above equation, we can estimate $\alpha(X; \gamma)$ as

$$\hat{\alpha}(x_i; \gamma) = \frac{\sum_{j=1}^{n}(1 - \delta_j)K_h(x_i, x_j)}{\sum_{j=1}^{n} \delta_j \exp(\gamma y_j)K_h(x_i, x_j)},$$

where $K_h(u, x) = K\{(u - x)/h\}/h^q$ with $K(\cdot)$ being a kernel function and h being a bandwidth. Under model (2.1), $\pi(x_i, y_i)$ can then be estimated by $\hat{\pi}(x_i, y_i; \gamma) = \hat{\pi}(x_i, y_i)$, where

$$\hat{\pi}(x_i, y_i) = \{1 + \hat{\alpha}(x_i; \gamma)\exp(\gamma y_i)\}^{-1}.$$

For the estimator of $m_0(x_i) = E(Y|x_i, \delta_i = 0)$, note that $m_0(X) = E[\delta \exp(\gamma Y)Y|X]/E[\delta \exp(\gamma Y)|X]$. Thus we can estimate $m_0(x_i)$ by

$$\hat{m}_0(x_i) := \hat{m}_0(x_i; \gamma) = \sum_{j=1}^{n} \omega_{j0}(x_i; \gamma)y_j, \tag{2.2}$$

where the weight

$$\omega_{j0}(x_i; \gamma) = \frac{\delta_j \exp(\gamma y_j)K_h(x_i, x_j)}{\sum_{j=1}^{n} \delta_j \exp(\gamma y_j)K_h(x_i, x_j)}. \tag{2.3}$$

We denote γ^* as the true value of γ in model (2.1). When γ^* is unknown, it can be estimated by either an independent survey or a validation sample, the latter being a subsample of the nonrespondents. Let $\hat{\gamma}$ be the corresponding estimator of γ^*. Then semiparametric estimators of $m_0(x_i)$ and $\pi(x_i, y_i)$ can be derived using $\hat{m}_0(x_i; \hat{\gamma})$ and $\hat{\pi}(x_i, y_i; \hat{\gamma})$, respectively.

After we get the estimators of $m_0(x_i)$, we can then impute the missing response based on regression imputation and construct one completed data set as follows:

$$\tilde{y}_{i1} = \delta_i y_i + (1 - \delta_i)\hat{m}_0(x_i).$$

Then a test statistic that is based on the residual-marked empirical process can be constructed as

$$R_{n1}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{\tilde{y}_{i1} - \phi^\top(x_i)\hat{\beta}_1\right\} I(x_i \le x), \tag{2.4}$$

where $\hat{\beta}_1$ is the estimator of β and is defined as

$$\hat{\beta}_1 = \left\{\sum_{i=1}^{n} \phi(x_i)\phi^\top(x_i)\right\}^{-1} \sum_{i=1}^{n} \phi(x_i)\tilde{y}_{i1}.$$

The Cramër–von-Mises-type test statistic is defined by

$$T_{n1} = \int R_{n1}^2(x)dF_n(x), \tag{2.5}$$

where $F_n(x)$ is the empirical distribution of X based on $x_1, x_2, \ldots, x_n$. For the use of Cramër–von-Mises-type test statistics in model checking, see Sun and Wang (2009), Xu and Guo (2013), and González-Manteiga and Crujeiras (2013) as examples.

Inspired by inverse probability weighting, we consider another approach. Note that under H_0, the following equation holds:

$$E\left[\frac{\delta}{\pi(X, Y)}\{Y - \phi^\top(X)\beta\}|X\right] = E\left\{Y - \phi^\top(X)\beta|X\right\} \equiv 0.$$

It is worth mentioning that our idea for constructing tests is to weight the observed residuals by the inverse probability function. The residual-marked empirical-process test statistic is constructed as

$$R_{n2}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(x_i, y_i)} \left\{ y_i - \phi^\top(x_i)\hat{\beta}_2 \right\} I(x_i \le x), \tag{2.6}$$

where $\hat{\beta}_2$ is the inverse-probability-weighted estimator of β in the form

$$\hat{\beta}_2 = \left\{ \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(x_i, y_i)} \phi(x_i)\phi^\top(x_i) \right\}^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(x_i, y_i)} \phi(x_i) y_i.$$

The test statistic is then defined as

$$T_{n2} = \int R_{n2}^2(x) dF_n(x). \tag{2.7}$$

3. Asymptotic behavior of test statistics

In this section, we investigate the asymptotic behaviors of the proposed test statistics under the null hypothesis as well as local alternatives. We then begin by discussing the case with known γ^*.

3.1. Asymptotic properties with known γ^*

To state the theorems, we introduce some notations. Denote $Z = (X, Y)$, $\Sigma = E\{\phi(X)\phi^\top(X)\}$, $L(X; x) = I(X \le x) - E\{\phi^\top(X) I(X \le x)\} \Sigma^{-1} \phi(X)$, and

$$J(\delta_i, x_i, y_i; x) = L(x_i; x) \frac{\delta_i \varepsilon_i + \{\pi(z_i) - \delta_i\} E(\varepsilon | x_i, \delta = 0)}{\pi(z_i)}.$$

The following theorem states the asymptotic behavior of the proposed tests under the null hypothesis (1.2).

Theorem 1. *Under H_0 and the conditions in the Appendix,*

$$R_{nk}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J(\delta_i, x_i, y_i; x) + o_p(1), \quad k = 1, 2,$$

converges in distribution to $R(x)$ in the Skorokhod space $D[-\infty, +\infty]$, where $R(x)$ is a centered continuous Gaussian process with covariance function

$$Cov(R(x_1), R(x_2)) = E\{J(\delta, X, Y; x_1) J(\delta, X, Y; x_2)\}$$

for any x_1 and x_2. Thus, T_{nk} converges in distribution to $T := \int R^2(x) dF(x)$ for $k = 1, 2$, where $F(\cdot)$ is the distribution function of X.

From Theorem 1, the two tests have the same limit under the null hypothesis. Interestingly, this continues to hold under the following local alternatives:

$$H_{1n} : Y = \phi^\top(X)\beta + C_n G(X) + \eta,$$

where $E(\eta|X) = 0$ and the function $G(\cdot)$ satisfies $E\{G^2(X)\} < \infty$. Furthermore, C_n is a constant sequence that converges to zero as n tends to infinity. Denoting $S(x) = E\{G(X)L(X; x)\}$, we have the following results.

Theorem 2. *Under H_{1n} and the conditions in the Appendix, we have that if $C_n\sqrt{n} \to 1$, then $R_{nk}(x)$ converges in distribution to $R(x) + S(x)$ for $k = 1, 2$, where $S(x)$ is a nonrandom shift function and T_{nk} converges in distribution to $\int \{R(x) + S(x)\}^2 dF(x)$. If $n^r C_n \to a$, $0 < r < 1/2$, then T_{nk} converges to infinity.*

From Theorem 2, we have that (i) if the local alternatives are distinct from the null hypothesis at the rate n^{-r} with $0 < r < 1/2$, then the asymptotic powers of the proposed tests are all one, and (ii) if the alternatives converge to the null hypothesis at the rate $n^{-1/2}$, then the proposed test can detect them also.

3.2. Asymptotic properties with estimated $\hat{\gamma}$ from validation sample

In Section 3.1, the asymptotic properties of the proposed test statistics are derived with known γ^*. However, in many cases γ^* is unknown and must be estimated. In general, we can obtain an estimator of γ^* from either an independent survey or a validation sample (Kim and Yu, 2011). To save space, we consider herein only the case in which a validation sample is selected randomly from the set of nonrespondents and the responses are obtained for all the units in the validation sample. The results for the case in which γ^* is estimated from an independent survey with sample size n have also been derived and are available on request.

An estimator $\hat{\gamma}$ of γ^* is defined as the solution of

$$\sum_{i=1}^{n}(1-\delta_i)r_i\{y_i - \hat{m}_0(x_i;\gamma)\} = 0,$$

where r_i is an indicator function that equals one if unit i belongs to the follow-up sample or zero otherwise.

Denote

$$M = E[(1-\delta)r\{E(Y^2|X,\delta=0) - m_0^2(X,\gamma^*)\}],$$
$$\Delta_i = \frac{1}{M}\{\eta_i - E(\eta|x_i,\delta=0)\}\left[(1-\delta_i)r_i - \delta_i\nu\left\{\frac{1}{\pi(z_i)} - 1\right\}\right],$$
$$H(x) = E\left[(1-\delta)\{Y-m_0(X)\}^2L(X;x)\right],$$
$$\tilde{J}(\delta_i, r_i, x_i, y_i; x) = J(\delta_i, x_i, y_i; x) + H(x)\Delta_i,$$

where $\nu = E(r|\delta=0)$ is the follow-up rate and is assumed to be a known constant. Compared with $J(\delta_i, r_i, x_i, y_i; x)$, there is an extra term in $\tilde{J}(\delta_i, r_i, x_i, y_i; x)$, which is the estimation effect of $\hat{\gamma}$. Under the null hypothesis, we have the following result.

Theorem 3. *Under H_0 and the conditions in the Appendix, we have that*

$$R_{nk}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{J}(\delta_i; r_i; x_i; y_i; x) + o_p(1), \quad k=1,2,$$

converges in distribution to $\tilde{R}(x)$ in the Skorokhod space $D[-\infty,+\infty]$, where $\tilde{R}(x)$ is a centered continuous Gaussian process with covariance function

$$Cov(\tilde{R}(x_1), \tilde{R}(x_2)) = E\{\tilde{J}(\delta, r, X, Y; x_1)\tilde{J}(\delta, r, X, Y; x_2)\}$$

for any x_1 and x_2. Thus, T_{nk} converges in distribution to $\tilde{T} := \int \tilde{R}^2(x)dF(x)$ for $k=1,2$.

For the power performances of the proposed tests under the local alternative hypotheses H_{1n}, the following theorem states the results.

Theorem 4. *Under H_{1n} and the conditions in the Appendix, the following results are obtained. If $C_n\sqrt{n} \to 1$, then $R_{nk}(x)$ converges in distribution to $\tilde{R}(x) + S(x)$ for $k=1,2$, and T_{nk} converges in distribution to $\int\{\tilde{R}(x)+S(x)\}^2dF(x), k=1,2$. If $n^rC_n \to a, 0<r<1/2$, then $T_{nk}, k=1,2$ converges to infinity.*

The above results imply that T_{nk} can still detect local alternative hypotheses, which converge to the null hypothesis at the rate of $n^{-1/2}$. Comparing Theorems 2 and 4, it is unclear whether the test statistics with known γ^* are more efficient than those with estimated $\hat{\gamma}$ because of the complicated asymptotic covariances of $R_{nk}(x)$. This issue is investigated numerically in Section 4.

3.3. Monte Carlo approximation

From Theorems 1 and 3, we can obtain the asymptotic covariances of $R_{nk}(x)$ in different situations. However, the variances of the test statistics T_{nk} are very complex and not easy to estimate, and thus the critical values are difficult to determine. To solve this problem, we adopt the nonparametric Monte Carlo approach; see Zhu and Neuhaus (2000) and Zhu (2005) for details. This procedure has some desirable features: for instance, the test procedure is self-scale invariant, whereupon we can determine the p-values with no additional standardization. We present the algorithm for T_{n1} when γ^* is estimated from a validation sample. For the other test T_{n2}, the algorithms are similar.

Step 1. Generate random variables $e_i (i=1,2,\ldots,n)$ independently with mean zero and variance one. Let $E_n := (e_1,\ldots,e_n)$ and define the conditional counterpart of R_{n1} as

$$\tilde{R}_{n1}(E_n, x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\hat{\tilde{J}}(\delta_i; r_i; x_i; y_i; x),$$

where $\hat{\tilde{J}}(\delta_i; r_i; x_i; y_i; x)$ is the estimator of $\tilde{J}(\delta_i; r_i; x_i; y_i; x)$, which is defined as

$$\hat{\tilde{J}}(\delta_i; r_i; x_i; y_i; x) = \hat{J}(\delta_i, x_i, y_i; x) + \hat{H}(x)\hat{\Delta}_i.$$

First, we obtain the estimator $\hat{J}(\delta_i, x_i, y_i; x)$ as

$$\hat{J}(\delta_i, x_i, y_i; x) = \hat{L}(x_i; x)\frac{\delta_i\hat{\varepsilon}_i + \{\hat{\pi}(z_i, \hat{\gamma}) - \delta_i\}\hat{E}(\varepsilon|x_i, \delta = 0)}{\hat{\pi}(z_i, \hat{\gamma})},$$

with

$$\hat{L}(x_i; x) = I(x_i \le x) - \sum_{i=1}^{n} \phi^\top(x_i)I(x_i \le x)\left\{\sum_{i=1}^{n}\phi(x_i)\phi^\top(x_i)\right\}^{-1}\phi(x_i);$$

$$\hat{\varepsilon}_i = y_i - \phi^\top(x_i)\hat{\beta}_1; \quad \hat{E}(\varepsilon|x_i, \delta = 0) = \sum_{j=1}^{n}\omega_{j0}(x_i, \hat{\gamma})\hat{\varepsilon}_j.$$

Here, $\hat{\gamma}$ is an estimator of γ^* by solving $\sum_{i=1}^{n}(1-\delta_i)r_i\{y_i - \hat{m}_0(x_i; \gamma)\} = 0$, and $\hat{\pi}(z_i, \hat{\gamma})$, $\hat{m}_0(x_i; \gamma)$, and $\omega_{j0}(x_i, \gamma)$ have been defined in Section 2.

Next, for the estimator of Δ_i, we have:

$$\hat{\Delta}_i = \frac{1}{\hat{M}}\{\hat{\varepsilon}_i - \hat{E}(\varepsilon|x_i, \delta = 0)\}\Big[(1-\delta_i)r_i - \delta_i\nu\{\frac{1}{\hat{\pi}(z_i, \hat{\gamma})} - 1\}\Big],$$

with

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)r_i\{\hat{E}(Y^2|x_i, \delta = 0) - \hat{m}_0^2(x_i, \hat{\gamma})\};$$

$$\hat{E}(Y^2|x_i, \delta = 0) = \sum_{j=1}^{n}\omega_{j0}(x_i, \hat{\gamma})y_j^2.$$

Next note that

$$\begin{aligned} H(x) &= E[(1-\delta)\{Y - m_0(X)\}^2L(X; x)] \\ &= E\left(E[(1-\delta)\{Y - m_0(X)\}^2L(X; x)|X]\right) \\ &= E\left(E[\{Y - m_0(X)\}^2L(X; x)|X, \delta = 0]P(\delta = 0|X)\right). \end{aligned}$$

Thus we can estimate $H(x)$ by

$$\hat{H}(x) = \frac{1}{n}\sum_{i=1}^{n}\hat{E}[\{Y - m_0(X)\}^2L(X; x)|x_i, \delta = 0]\hat{P}(\delta = 0|x_i),$$

with

$$\hat{E}[\{Y - m_0(X)\}^2L(X; x)|x_i, \delta = 0] = \sum_{j=1}^{n}\omega_{j0}(x_i, \hat{\gamma})\{y_j - \hat{m}_0(x_j, \hat{\gamma})\}^2\hat{L}(x_j; x);$$

$$\hat{P}(\delta = 0|x_i) = 1 - \frac{\sum_{j=1}^{n}\delta_jK_h(x_i, x_j)}{\sum_{j=1}^{n}K_h(x_i, x_j)}.$$

The resultant conditional test statistic is

$$\tilde{T}_{n1}(E_n) = \int \tilde{R}_{n1}^2(E_n, x)dF_n(x).$$

Step 2. Generate m sets of E_n, say $E_n^{(i)}$, $i = 1, \ldots, m$ and get m values of $\tilde{T}_{n1}(E_n)$, say $\tilde{T}_{n1}(E_n^{(i)})$, $i = 1, \ldots, m$.

Step 3. The p-value is estimated using $\hat{p}_k = n_k/(m+1)$, where n_k is the number of $\tilde{T}_{n1}(E_n^{(i)})$ that are larger than or equal to T_{n1}. Reject H_0 whenever $\hat{p}_k \le \alpha$ for a designed level α.

With the above algorithm, we expect the determination of the critical values to be unaffected by the choice of hypothesis (i.e., null or local alternative). Theorem 5 shows that the conditional distribution based on the Monte Carlo approximation converges to the limit null distribution.

Theorem 5. *Under either the null hypothesis or local alternatives with $n^rC_n \to a, 0 < r \le 1/2$ and the conditions in Theorem 3, we have that for almost all sequences*

$$\{(y_1, \delta_1, r_1, x_1); \cdots; (y_n, \delta_n, r_n, x_n), \ldots\},$$

the conditional distribution of $\tilde{T}_{n1}(E_n)$ converges to the limit null distribution of T_{n1}. Furthermore, under global alternatives with $C_n \to c > 0$, $\tilde{T}_{n1}(E_n)$ converges in probability to a finite limit that may be different from the limit null distribution of T_{n1}.

4. Numerical analysis

In this section, we carry out some simulation studies and present a real data example to illustrate the merits of our proposed test statistics.

4.1. Simulation studies

In this subsection, we examine the performance of the proposed test statistics through some simulation runs. We also compare with the MAR method and complete-case (CC) analysis. Here, the MAR method corresponds to set $\hat{\gamma} = 0$ regardless of the true value of γ^*. We also consider the oracle test statistic, which knows the true value of γ^*. Let T_n^{cc} be the test statistic based on the CC method. Denote the MAR methods corresponding to T_{n1}, T_{n2} as T_{n1}^{MAR} and T_{n2}^{MAR}, respectively, and denote the oracle test statistics as T_{n1}^{O} and T_{n2}^{O}. Clearly, the oracle tests are only applicable when the assumed exponential tilting model (2.1) holds. We begin by considering the following study design.

Study 1. Generate data from the model

$$Y = \phi^\top(X)\beta + aG(X) + \varepsilon, \tag{4.1}$$

where $\phi(X) = 1 + X^2$ with $X \sim U(0, 1)$, $\beta = 1$, $G(X) = X^3$ and $\varepsilon \sim N(0, 0.5)$. For model (4.1), the null hypothesis $H_0 : E(Y|X) = \phi^\top(X)\beta$ for some β corresponds to $a = 0$. For this model, we assume that Y is NMAR. The follow-up rate is taken to be 30%.

We consider the following four missing mechanisms for model (4.1):

Case 1. $\pi_1(x, y) = 1/[1 + \exp\{-(1 + 0.3x + 0.3y)\}]$;
Case 2. $\pi_2(x, y) = 1/[1 + \exp\{-(-0.3 + 0.3x + 0.3y)\}]$;
Case 3. $\pi_3(x, y) = 1/[1 + \exp\{-(1 + 0.3x + x^2 + 0.3y)\}]$;
Case 4. $\pi_4(x, y) = 1/[1 + \exp\{-(-0.3 + 0.3x + 0.3y + 0.3y^2)\}]$.

For the above different cases, the mean response rates are $E\pi_1(X, Y) \approx 0.82, E\pi_2(X, Y) \approx 0.56, E\pi_3(X, Y) \approx 0.86$, and $E\pi_4(X, Y) \approx 0.69$, respectively. Cases 1 and 2 are both linear nonignorable missing mechanisms, case 3 is a nonlinear nonignorable missing mechanism with quadratic form in x, and case 4 does not follow the assumed exponential tilting model (2.1) and is used here to examine the robustness of the proposed tests against the failure of the assumed missing mechanism model (2.1). In cases 1–3, the true value of γ^* is -0.3.

The kernel function is taken to be the Gaussian kernel function $K(u) = \exp(-u^2/2)/(2\pi)^{1/2}$. From one of our simulations not reported here, the bandwidths have little impact for the proposed test statistics in terms of empirical sizes and powers. Following the suggestions of Kim and Yu (2011), Zhao et al. (2013), and Niu et al. (2014), we take the bandwidth h to be $n^{-1/5}$ in the following studies. The simulation results are based on 1000 replications. For each sample, the Monte Carlo approximation is based on 500 sets of reference data generated by the Monte Carlo method. The nominal level is set to be $\alpha = 0.05$, and the performances of the proposed tests under model (4.1) are investigated for different values of a, different sample sizes $n = 100$ and $n = 200$, and different missing mechanisms $\pi_i(x, y)(i = 1, \ldots, 4)$. The results are presented in Tables 1–4.

According to intuition, all the test statistics have larger powers when the sample size increases. From Table 1, we see that for the missing mechanism of case 1, all seven test statistics perform similarly. The powers of T_{n1} and T_{n2} are slightly larger than those of the MAR methods T_{n1}^{MAR} and T_{n2}^{MAR} and also the CC method. For instance, when $a = 0.6$ and $n = 200$, the empirical powers of T_{n1} and T_{n2} are 0.5880 and 0.5870, respectively, and the corresponding powers of T_{n1}^{MAR}, T_{n2}^{MAR}, and T_n^{cc} are 0.5400, 0.5410, and 0.5540, respectively. Meanwhile, T_{n1} and T_{n2} have almost the same powers as those of T_{n1}^{O} and T_{n2}^{O}, respectively. For the missing mechanism of case 2, the powers of T_{n1} and T_{n2} are much larger than those of T_{n1}^{MAR}, T_{n2}^{MAR}, and T_n^{cc}. For instance, when $a = 0.6$ and $n = 200$, the empirical powers of T_{n1} and T_{n2} are both 0.46, while the corresponding powers of T_{n1}^{MAR}, T_{n2}^{MAR}, and T_n^{cc} are 0.345, 0.344, and 0.367, respectively. Comparing the results in Tables 1 and 2, it is also clear that with a larger missing proportion, the powers of the test statistics decrease. From Table 3, we can have similar findings. Next, we turn our attention to Table 4, from which we see that neither the MAR method nor the CC method can control the empirical size. This problem becomes even more severe as the sample size increases. However, they are not very powerful to the alternatives. For instance, when $a = 0.3$ and $n = 200$, the empirical powers of T_{n1} and T_{n2} are 0.184 and 0.183, respectively, while the corresponding powers of T_{n1}^{MAR}, T_{n2}^{MAR}, and T_n^{cc} are all approximately 0.06. This implies that using the MAR method and the CC method mistakenly in the nonignorable missing situation would result in misleading conclusions. Moreover, from Table 4 we also conclude that our proposed test statistics T_{n1} and T_{n2} are robust against failure of the assumed missing mechanism model (2.1).

To study the performance of the tests against high-frequency alternatives, we conduct the following simulation study.

Study 2. Generate data from the model

$$Y = \phi^\top(X)\beta + aG(X) + \varepsilon. \tag{4.2}$$

Table 1
Empirical sizes and powers of test statistics for study 1, with $n = 100, 200$ and missing mechanism $\pi_1(x, y)$.

a	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	T_{n1}		T_{n2}	
0.0	0.0535	0.0485	0.0535	0.0480
0.3	0.1130	0.1830	0.1130	0.1820
0.6	0.3520	0.5880	0.3510	0.5870
0.9	0.6480	0.9120	0.6490	0.9110
1.2	0.8690	0.9940	0.8710	0.9940
	T_{n1}^{MAR}		T_{n2}^{MAR}	
0.0	0.0500	0.0555	0.0500	0.0555
0.3	0.1010	0.1510	0.1010	0.1500
0.6	0.3260	0.5400	0.3270	0.5410
0.9	0.6230	0.8880	0.6240	0.8880
1.2	0.8560	0.9890	0.8570	0.9890
	T_{n1}^{O}		T_{n2}^{O}	
0.0	0.0435	0.0485	0.0435	0.0485
0.3	0.1110	0.1820	0.1110	0.1820
0.6	0.3600	0.6020	0.3580	0.6020
0.9	0.6490	0.9170	0.6510	0.9170
1.2	0.8790	0.9940	0.8790	0.9940
	T_{n}^{cc}			
0.0	0.0460	0.0560		
0.3	0.1020	0.1520		
0.6	0.3290	0.5540		
0.9	0.6320	0.8960		
1.2	0.8670	0.9930		

Table 2
Empirical sizes and powers of test statistics for study 1, with $n = 100, 200$ and missing mechanism $\pi_2(x, y)$.

a	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	T_{n1}		T_{n2}	
0.0	0.0540	0.0550	0.0540	0.0550
0.3	0.0830	0.1650	0.0830	0.1640
0.6	0.2260	0.4600	0.2230	0.4600
0.9	0.4690	0.7620	0.4680	0.7580
1.2	0.6940	0.9490	0.6940	0.9500
	T_{n1}^{MAR}		T_{n2}^{MAR}	
0.0	0.0660	0.0680	0.0660	0.0680
0.3	0.0820	0.1060	0.0810	0.1070
0.6	0.1880	0.3450	0.1880	0.3440
0.9	0.4190	0.6790	0.4180	0.6800
1.2	0.6500	0.9190	0.6490	0.9200
	T_{n1}^{O}		T_{n2}^{O}	
0.0	0.0620	0.0450	0.0610	0.0460
0.3	0.0940	0.1570	0.0930	0.1570
0.6	0.2290	0.4300	0.2270	0.4290
0.9	0.4710	0.7600	0.4680	0.7600
1.2	0.7110	0.9460	0.7110	0.9470
	T_{n}^{cc}			
0.0	0.0660	0.0560		
0.3	0.0780	0.1150		
0.6	0.2010	0.3670		
0.9	0.4530	0.7260		
1.2	0.6920	0.9400		

We use the same settings as study 1 except for the alternatives. In this study, $G(X) = \cos(2\pi X)$ is used as a high-frequency alternative. From study 1, we also observe that the sizes and powers of T_{n1} and T_{n2}, T_{n1}^{MAR} and T_{n2}^{MAR}, and T_{n1}^{O} and T_{n2}^{O} are almost the same, which is consistent with the theories developed in Section 3. To save space, we consider only T_{n2}, T_{n2}^{MAR}, T_{n}^{cc}, and/or T_{n2}^{O} in this example. Besides the missing mechanisms of cases 2 and 4, we consider the following two missing mechanisms:

Table 3
Empirical sizes and powers of test statistics for study 1, with $n = 100, 200$ and missing mechanism $\pi_3(x, y)$.

a	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	T_{n1}		T_{n2}	
0.0	0.0485	0.0445	0.0485	0.0445
0.3	0.0925	0.2040	0.0930	0.2040
0.6	0.3340	0.6120	0.3340	0.6120
0.9	0.6240	0.9120	0.6240	0.9110
1.2	0.8650	0.9940	0.8650	0.9940
	T_{n1}^{MAR}		T_{n2}^{MAR}	
0.0	0.0480	0.0510	0.0480	0.0510
0.3	0.0845	0.1560	0.0845	0.1560
0.6	0.3010	0.5460	0.3000	0.5470
0.9	0.6040	0.8910	0.6040	0.8910
1.2	0.8480	0.9940	0.8480	0.9940
	T_{n1}^{O}		T_{n2}^{O}	
0.0	0.0425	0.0405	0.0425	0.0405
0.3	0.0970	0.2070	0.0965	0.2070
0.6	0.3380	0.6270	0.3380	0.6280
0.9	0.6440	0.9290	0.6440	0.9290
1.2	0.8740	0.9960	0.8740	0.9960
	T_{n}^{cc}			
0.0	0.0505	0.0490		
0.3	0.0880	0.1610		
0.6	0.3090	0.5690		
0.9	0.6240	0.9120		
1.2	0.8610	0.9960		

Table 4
Empirical sizes and powers of test statistics for study 1, with $n = 100, 200$ and missing mechanism $\pi_4(x, y)$.

a	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	T_{n1}		T_{n2}	
0.0	0.0420	0.0460	0.0410	0.0460
0.3	0.0840	0.1840	0.0820	0.1830
0.6	0.2330	0.5650	0.2290	0.5640
0.9	0.5300	0.8850	0.5290	0.8850
1.2	0.7930	0.9890	0.7940	0.9890
	T_{n1}^{MAR}		T_{n2}^{MAR}	
0.0	0.1060	0.1250	0.1060	0.1260
0.3	0.0510	0.0600	0.0510	0.0600
0.6	0.1110	0.2330	0.1120	0.2320
0.9	0.3210	0.5870	0.3190	0.5880
1.2	0.6050	0.8970	0.6040	0.8970
	T_{n}^{cc}			
0.0	0.0905	0.1240		
0.3	0.0500	0.0620		
0.6	0.1240	0.2660		
0.9	0.3590	0.6540		
1.2	0.6650	0.9360		

Case 5. $\pi_5(x, y) = 1/[1 + \exp\{-(-0.3 + 0.3x)\}]$;
Case 6. $\pi_6(x, y) = \Phi(-0.3 + 0.3x + 0.3y)$.

Here, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. The missing mechanism of case 5 corresponds to the missing at random situation, and thus γ^* is now equal to zero. The missing mechanism of case 6 does not follow the assumed exponential tilting model (2.1). The mean response rates are $E\pi_5(X, Y) \approx 0.46$ and $E\pi_6(X, Y) \approx 0.60$, respectively. The results are presented in Figs. 1 and 2, from which we see that T_{n2} is generally more powerful than T_{n2}^{MAR} and T_{n}^{cc}. Furthermore, it performs almost as well as the oracle test statistic T_{n2}^{O}. Again, this study also shows that the proposed tests are robust against failure of the assumed missing mechanism model (2.1). Moreover, even with a very large missing proportion ($\pi_5(X, Y)$), our proposed test statistics can still detect the alternative hypotheses efficiently, whereas the MAR method and the CC method may not control the empirical size well in some situations.

Please cite this article as: X. Guo, L. Song, Y. Fang et al., Model checking for general linear regression with nonignorable missing response. Computational Statistics and Data Analysis (2019), https://doi.org/10.1016/j.csda.2019.03.009.

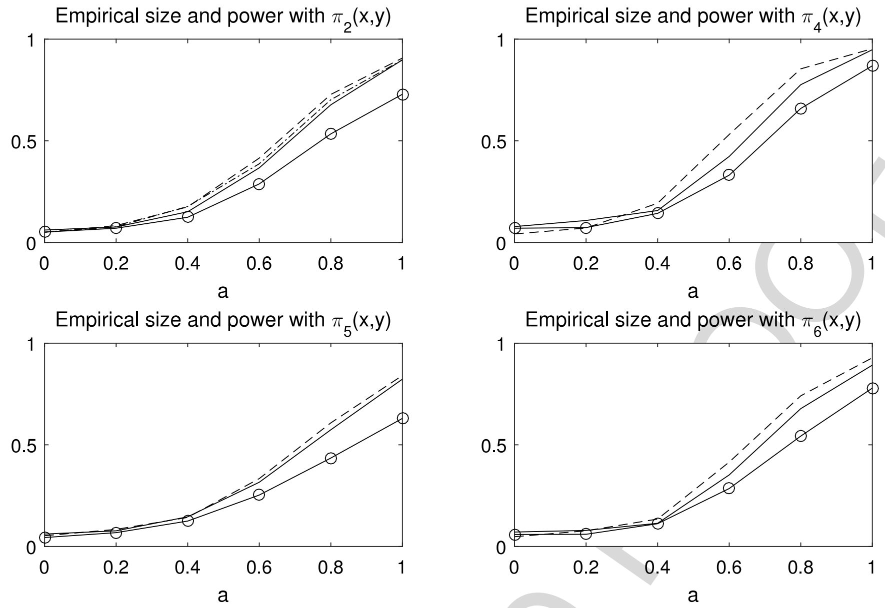

Fig. 1. Empirical sizes and powers of T_{n2}, T_{n2}^{MAR}, T_{n2}^{O}, and T_n^{cc} for study 2 with $n = 100$: (1) for $\pi_2(x, y)$; (2) for $\pi_4(x, y)$; (3) for $\pi_5(x, y)$; (4) for $\pi_6(x, y)$. The dashed line, the dashed–dotted line, the solid line, and the solid line marked by 'o' represent T_{n2}, T_{n2}^{O}, T_{n2}^{MAR}, and T_n^{cc}, respectively.

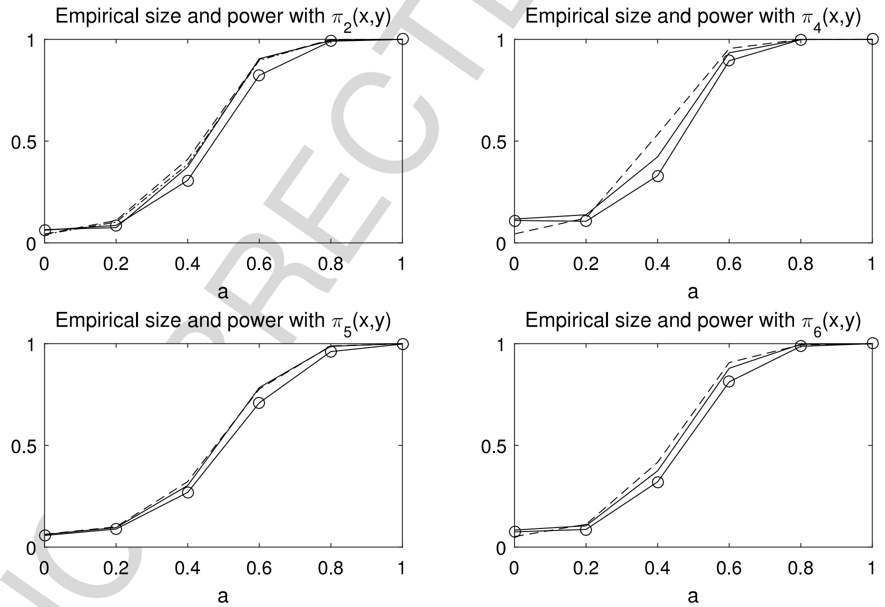

Fig. 2. Empirical sizes and powers of T_{n2}, T_{n2}^{MAR}, T_{n2}^{O}, and T_n^{cc} for study 2 with $n = 200$: (1) for $\pi_2(x, y)$; (2) for $\pi_4(x, y)$; (3) for $\pi_5(x, y)$; (4) for $\pi_6(x, y)$. The dashed line, the dashed–dotted line, the solid line, and the solid line marked by 'o' represent T_{n2}, T_{n2}^{O}, T_{n2}^{MAR}, and T_n^{cc}, respectively.

4.2. Real data analysis

We use a data set regarding the persistence of maternal smoking to illustrate our proposed test statistics. This data set is a subsample from the Six Cities Study of the health effects of air pollution and has been analyzed by Ware et al. (1984), Lipsitz et al. (2004), and Niu et al. (2014). There are 574 observations, of which 208 subjects have missing responses. Niu et al. (2014) used this data set to construct confidence intervals for the parameters in a linear regression model, and Lipsitz et al. (2004) also used the linear regression model. The response variable Y and the covariate X are the square root of the number of maternal cigarettes smoked per day when the child is 10 years old and when the child is 9 years old, respectively. X is completely observed and Y is regarded as NMAR. Niu et al. (2014) and Lipsitz et al. (2004) fitted the linear regression $Y = \beta_0 + \beta_1 X + \varepsilon$, but for their analysis to be correct it is critical to check whether the assumption of linearity holds. That is, we must check whether the following hypothesis is true:

$$H_0 : E(Y|X) = \beta_0 + \beta_1 X.$$

To that end, we use $T_{n1}, T_{n2}, T_{n1}^{MAR}, T_{n2}^{MAR}$, and T_n^{cc} to test this hypothesis. First, an estimate $\hat{\gamma}$ of the exponential tilting parameter γ can be obtained by solving $\sum_{i=1}^n (1-\delta_i) r_i \{y_i - \hat{m}_0(x_i;\gamma)\} = 0$. The follow-up rate used here is 20%. The Gaussian kernel function $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ is applied. The p-values of the five test statistics are all obviously larger than 0.1, and thus the linear regression model is plausible for this data set. The conclusions from Lipsitz et al. (2004) and Niu et al. (2014) should be sound.

5. Conclusions

In this paper, we investigate model checking for the general linear regression model with nonignorable missing response. Based on an exponential tilting model, we propose two residual-marked empirical-process-based tests. The asymptotic properties of the proposed tests are investigated under the null and local alternative hypotheses in different scenarios. It is found that the two tests perform identically in the asymptotic sense. To compute the critical values, a nonparametric Monte Carlo test procedure is adopted. Simulation studies are conducted and a real data example is analyzed to examine the performance of the proposed tests.

Herein, we are interested in the general linear regression model. However, the methodologies developed in this paper can be extended to other regression models, such as the partial linear regression model or single-index models. It would be interesting to check whether the nonparametric part in the partial linear regression model can be modeled parametrically and whether the link functions in single-index models are known. These topics have been studied previously for completely observed data sets, but there is as yet no literature for regression models with nonignorable missing response. We leave these topics for future work.

Acknowledgments

The research described herein is supported by the National Natural Science Foundation of China NSFC (11601227, 11701034, 11671042, and 71602089), the National Social Science Fund of China (NSSFC-16BTJ013 and NSSFC-16ZDA010), Shanghai Normal University Research Fund (KF201905), Sichuan project of Science and technology (2017JY0273), the Natural Science Foundation of Jiangsu Province (BK20160785), and a grant from the University Grants Council of Hong Kong.

Appendix. Conditions of the theorems

The following conditions are required for the theorems in Section 3. Note that the detailed proofs of the theoretical results are given in the supplementary material.

(1) $\pi(X, Y)$ has bounded partial derivatives up to order 2, $\pi(X, Y) \geq c_0 > 0$, and $p(X) = E(\delta|X) \neq 1$.
(2) $E(Y^2) < \infty$ and $E\{\exp(2\gamma Y)\} < \infty$. $\Sigma = E\{\phi(X)\phi^\top(X)\}$ is nonsingular.
(3) $nh^q \to \infty$ and $h \to 0$.
(4) The density of X, say $f(x)$ on support $\mathcal{C}$, exists and has bounded derivatives up to order 2 and satisfies

$$0 < \inf_{x\in\mathcal{C}} f(x) \leq \sup_{x\in\mathcal{C}} f(x) < \infty;$$

(5) The kernel function $K(\cdot)$ is a bounded symmetric probability density function with bounded support, and of order 2.

Remark 1. Conditions 3 and 5 are typical for obtaining convergence rates when nonparametric estimation is applied. Condition 1 is a common assumption in missing-data studies (e.g., Sun and Wang, 2009; Kim and Yu, 2011). Condition 2 is necessary for the asymptotic normality of the least-squares estimator. Condition 4 aims at avoiding tedious proofs of the theorems (e.g., Xue, 2009); without it, we would have to resort to some truncation technique to control small values in the denominators.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2019.03.009.

Please cite this article as: X. Guo, L. Song, Y. Fang et al., Model checking for general linear regression with nonignorable missing response. Computational Statistics and Data Analysis (2019), https://doi.org/10.1016/j.csda.2019.03.009.

References

Bindele, H.F., Zhao, Y.C., 2018. Rank-based estimating equation with non-ignorable missing responses via empirical likelihood. Statist. Sinica 28, 1787–1820.

González-Manteiga, W., Crujeiras, R.M., 2013. An updated review of Goodness-of-Fit tests for regression models. Test 22, 361–411.

González-Manteiga, W., Pérez-González, A., 2006. Goodness-of-fit tests for linear regression models with missing response data. Canad. J. Statist. 34, 149–170.

Guo, X., Xu, W.L., Zhu, L.X., 2015. Model checking for parametric regressions with response missing at random. Ann. Inst. Statist. Math. 67, 229–259.

Guo, X., Zhu, L.X., 2017. A review on dimension-reduction based tests for regressions. In: From Statistics to Mathematical Finance. Springer, Cham, pp. 105–125.

Härdle, W., Mammen, E., 1993. Comparing non-parametric versus parametric regression fits. Ann. Statist. 21, 1926–1947.

Ibrahim, J.G., Chen, M.H., Lipsitz, S.R., 2001. Missing responses in generalised linearmixed models when the missing data mechanism is nonignorable. Biometrika 88, 551–564.

Ibrahim, J.G., Chen, M.H., Lipsitz, S.R., Herring, A.H., 2005. Missing data methods for generalized linear models: a comparative review. J. Amer. Statist. Assoc. 100, 332–346.

Kim, J.K., Shao, J., 2013. Statistical Methods for Handling Incomplete Data. Chapman & Hall / CRC.

Kim, J.K., Yu, C.L., 2011. A semiparametric estimation of mean functionals with nonignorable missing data. J. Amer. Statist. Assoc. 106, 157–165.

Kott, P.S., Chang, T., 2010. Using calibration weighting to adjust for nonignorable unit nonresponse. J. Amer. Statist. Assoc. 105, 1265–1275.

Koul, H.L., Ni, P.P., 2004. Minimum distance regression model checking. J. Statist. Plann. Inference 119, 109–141.

Li, X.Y., 2012. Lack-of-fit testing of regression model with response missing at random. J. Statist. Plann. Inference 142, 155–170.

Lipsitz, S.R., Molenberghs, G., Fitzmaurice, G.M., Ibrahim, J.G., 2004. Protective estimate for linear regression with nonignorably missing Gaussian outcomes. Statist. Modell. 4, 3–17.

Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. Wiley, New York.

Morikawa, K., Kim, J.K., Kano, Y., 2017. Semiparametric maximum likelihood estimation with data missing not at random. Canad. J. Statist. 45 (4), 393–409.

Niu, C.Z., Guo, X., Xu, W.L., Zhu, L.X., 2014. Empirical likelihood inference in linear regression with nonignorable missing response. Comput. Statist. Data Anal. 79, 91–112.

Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.

Qin, J., Leung, D., Shao, J., 2002. Estimation with survey data under nonignorable nonresponse or informative sampling. J. Amer. Statist. Assoc. 97, 193–200.

Shao, J., Wang, L., 2016. Semiparametric inverse propensity weighting for nonignorable missing data. Biometrika 103 (1), 175–187.

Sun, Z.H., Wang, Q.H., 2009. Checking the adequacy of a general linear model with responses missing at random. J. Statist. Plann. Inference 139, 3588–3604.

Tang, N.S., Ju, Y.Y., 2018. Statistical inference for nonignorable missing data problems: a selective review. Statist. Theory Field 2, 105–133.

Tang, N.S., Zhao, P.Y., Zhu, H.T., 2014. Empirical likelihood for estimating equations with nonignorably missing data. Statist. Sinica 24, 723–747.

Wang, S., Shao, J., Kim, J.K., 2014. An instrumental variable approach for identification and estimation with nonignorable nonresponse. Statist. Sinica 24 (3), 1097–1116.

Ware, J.H., Dockery, D.W., Spiro, A., Speizer, F.E., Ferris, B.G.J., 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. Amer. Rev. Respiratory Dis. 129, 366–374.

Xu, W.L., Guo, X., 2013. Nonparametric checks for varying coefficient models with missing response at random. Metrika 76, 459–482.

Xue, L.G., 2009. Empirical likelihood for linear models with missing responses. J. Multivariate Anal. 100, 1353–1366.

Zhao, P.Y., Tang, N.S., Qu, A., Jiang, D.P., 2017. Semiparametric estimating equations inference with nonignorable nonresponse. Statist. Sinica 27, 89–113.

Zhao, H., Zhao, P.Y., Tang, N.S., 2013. Empirical likelihood inference for mean functionals with nonignorably missing response data. Comput. Statist. Data Anal. 66, 101–116.

Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. J. Econometrics 75, 263–289.

Zhong, P.S., Chen, S.X., 2014. Jackknife empirical likelihood inference with regression imputation and survey data. J. Multivariate Anal. 129, 193–205.

Zhu, L.X., 2005. Nonparametric Monte Carlo Tests and their Applications. Springer, New York.

Zhu, L.X., Neuhaus, G., 2000. Nonparametric Monte Carlo tests for multivariate distributions. Biometrika 87, 919–928.

Please cite this article as: X. Guo, L. Song, Y. Fang et al., Model checking for general linear regression with nonignorable missing response. Computational Statistics and Data Analysis (2019), https://doi.org/10.1016/j.csda.2019.03.009.