Ann Inst Stat Math (2013) 65:473–490
DOI 10.1007/s10463-012-0379-4

Checking the adequacy of partial linear models with missing covariates at random

Wangli Xu · Xu Guo

Received: 10 November 2011 / Revised: 14 May 2012 / Published online: 30 September 2012

Abstract In this paper, we consider the goodness-of-fit for checking whether the nonparametric function in a partial linear regression model with missing covariate at random is a parametric one or not. We estimate the selection probability by using parametric and nonparametric approaches. Two score type tests are constructed with the estimated selection probability. The asymptotic distributions of the test statistics are investigated under the null and local alterative hypothesis. Simulation studies are carried out to examine the finite sample performance of the sizes and powers of the tests. We apply the proposed procedure to a data set on the AIDS clinical trial group (ACTG 315) study.

Keywords Partial linear model · Lack-of-fit test · Covariates missing at random · Inverse probability weights

1 Introduction

To fix notation, let Y denote a continuous outcome variable, X and T be the exposure variables with p and q-dimensional vectors, respectively, the partial linear model can

W. Xu · X. Guo
Center for Applied Statistics, School of Statistics, Renmin University of China,
Zhongguancun Street 59, Beijing 100872, China
e-mail: xwlbnu@163.com

X. Guo (✉)
Department of Mathematics, Hong Kong Baptist University, Fong Shu-Chuen Library 1110,
Kowloon Tong, Hong Kong
e-mail: liushengjunyi@163.com

be written as

$$Y = X^{\tau}\beta + g(T) + \varepsilon, \tag{1}$$

where β is an unknown parameter vector of dimension p, $g(\cdot)$ is an unknown measurable function, and ε is the error term with $E(\varepsilon|X,T) = 0$ and $E(\varepsilon^2|X,T) = \sigma^2(X,T) < \infty$. We use the superscript τ in (1) denote a transpose. Here, we assume that the variable T is a scalar for simplicity. The proposed methods in this paper can be extended to the situation when T is a multivariate variable.

When the data set is completed, there are many literature investigating the goodness-of-test for a partial linear model. For checking whether the partial linear model in (1) is correct for data fitting, i.e., $H_0 : E(Y|X,T) = X^{\tau}\beta + g(T)$ for some β and $g(\cdot)$, among others, Whang and Andrews (1993) and Yatchew (1992) used sample splitting to recommend ad hoc methods to handle this problem. To avoid the use of ad hoc modification and improve the power performance, Fan and Li (1996), based on a kernel estimator of the conditional expectation of residuals given (X,T), constructed a consistent test for the above H_0. To obtain some distribution-free test, Zhu and Ng (2003) proposed a residual marked process test. For the implementation of their procedure, they resorted to a variant of the wild bootstrap approximation called "Random Symmetrization".

In practice, we are also concerned with the question whether the nonparametric part in (1) is a parametric function, that is, $H_0 : g(\cdot) = g(\cdot,\theta)$ for some θ and a known function $g(\cdot,\theta)$. If the null hypothesis for some known function $g(\cdot,\theta)$ such as linear function holds, we can carry out more efficient statistical inference. On the other hand, if we misspecify the regression model, we are at the risk of getting biased estimator and unreliable inference. For this issue, among others, Li et al. (2011) proposed a test to check the linearity of the nonparametric portion by a linear interpolation and obtain the p value using the fiducial method. Liang (2006) developed a Crámer-von Mises statistic and likelihood ratio test for checking the linearity of nonparametric function. Li (2009) proposed two Wald-type spline-based test statistics to check the linearity of partially linear models.

It is quite common in practice that some covariates variable, denoted it as U with $X = (U,V)$, may be not available. Missing covariates data can arise due to various reasons such as limited budget to measure for the full study cohort, refusals to reply to certain question to supply the desired information, drop outs due to serious side effects, failure on the part of investigator to gather correct information, errors in the measuring apparatus, and so forth.

When the partial linear model is missing response at random, there are many investigations in the literature for the estimation of β and $g(\cdot)$, and goodness-of-fit problems. For estimation of the partial linear model, among others, Wang et al. (2004) defined a class of estimators including semiparametric regression imputation estimator, marginal average estimator and (marginal) propensity score weighted estimator for the marginal mean of response. Wang and Sun (2007) proposed imputation, semiparametric regression surrogate and inverse marginal probability weighted methods to estimate the parameters and nonparametric function, respectively. Liang et al. (2007) developed a class of semiparametric estimators for the partial linear model with missing response variables and error-prone covariates. For testing of the partial linear model, among others, Xu et al. (2012) constructed two completed data sets based on imputation and marginal inverse probability weighted methods, and developed two empirical process-based tests for checking whether the nonparametric part in (1) is a parametric function. Sun et al. (2009) proposed two empirical process-based tests for checking whether model (1) with missing response is plausible for data fitting.

When the covariates are missing, Liang et al. (2004) estimated the regression parameter by employing the augmented inverse weight probability method asbreak in Robins et al. (1994). Wang (2009) proposed a model calibration-based method and a weighted way to estimate the parameter and nonparametric function, respectively. To our knowledge, few works focus on the goodness-of-fit for the partial linear model with missing covariate data. Evidently, it is an interesting topic for testing whether the nonparametric function is a parametric form in model (1) with missing covariates, and the existing methods for complete data may not be used directly.

In this paper, for the model (1) with covariates missing at random, we consider testing

$$H_0 : g(\cdot) = g(\cdot, \theta) \tag{2}$$

for some θ and known function $g(\cdot, \theta)$. Our test statistics are based on a weighted version of the residual. We choose the commonly used inverse selection probability as the weighted function, which is estimated parametrically and nonparametrically, respectively. Based on the parametric and nonparametric estimation of the weight function, two types of the score test are constructed. We investigate the finite sample property of our proposed tests through simulation studies.

The rest of this paper is organized as follows. In Sect. 2, we construct the test statistics and derive their asymptotic properties under null hypothesis and local alternative hypothesis. In Sect. 3, some simulation analysis and a real data analysis is carried out to illustrated the proposed tests. The proofs of the asymptotic results are presented in the Appendix.

2 Test procedure

2.1 Construction of test statistics

Let $X = (U, V)$ and we assume that the covariate U is missing at random (MAR) throughout this paper, while Y, V and T are fully observed. Here U, V are p_1, p_2-dimensional random vectors, respectively. Let δ be the missing indicator for the individual. It is defined as $\delta = 1$ if U is observed and $\delta = 0$ if otherwise. MAR implies that δ and U are conditional independent given other variables Y, V and T. Thats,

$$P(\delta = 1|Y, U, V, T) = P(\delta = 1|Y, V, T) = \pi(Z),$$

here $Z = (Y, V, T)$. MAR is commonly assumed in the statistical analysis with missing data and is suitable in many practical situations, see Little and Rubin (1987).

Our tests are based on the consideration that under H_0, we have

$$E\Big(\frac{\delta}{\pi(Z)}(Y - X^\tau\beta - g(T,\theta))\Big) = 0, \tag{3}$$

while under alternative hypothesis with $\Pr(g(T) = g(T,\theta)) < 1$, we can obtain

$$E\Big(\frac{\delta}{\pi(Z)}(Y - X^\tau\beta - g(T,\theta))\Big) \neq 0.$$

Consequently, we can construct the following two residual-based test statistics based on the left hand side of the empirical version in (3)

$$T_{n1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i)}(y_i - x_i^\tau\hat{\beta} - g(t_i,\hat{\theta})), \tag{4}$$

$$T_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi(z_i,\hat{\alpha})}(y_i - x_i^\tau\hat{\beta} - g(t_i,\hat{\theta})), \tag{5}$$

where $\hat{\beta}, \hat{\theta}$ are the estimators of β, θ, and $\hat{\pi}(z_i)$ and $\pi(z_i,\hat{\alpha})$ are the parametric and nonparametric estimators of $\pi(z_i)$, respectively. The parameter estimators $\hat{\beta}$ and $\hat{\theta}$, and the function estimation $\hat{\pi}(z_i)$ and $\pi(z_i,\hat{\alpha})$ will be specified later. The difference between the test (4) and (5) is that the test (4) considers the estimator of $\pi(z_i)$ to be nonparametric function while (5) considers that to be parametric form.

Let $g_1(T) = E(\delta X|T)/E(\delta|T)$, $g_2(T) = E(\delta Y|T)/E(\delta|T)$, and their corresponding estimators are denoted by

$$\hat{g}_1(t) = \frac{\sum_{j=1}^{n}\delta_j x_j K_h(t-t_j)}{\sum_{j=1}^{n}\delta_j K_h(t-t_j)},\quad \hat{g}_2(t) = \frac{\sum_{j=1}^{n}\delta_j y_j K_h(t-t_j)}{\sum_{j=1}^{n}\delta_j K_h(t-t_j)}$$

here $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being a kernel function and h being a bandwidth. We estimate the regression parameters β by using the following expression:

$$\hat{\beta} = \left(\sum_{i=1}^{n}\delta_i(x_i - \hat{g}_1(t_i))(x_i - \hat{g}_1(t_i))^\tau\right)^{-1}\sum_{i=1}^{n}\delta_i(x_i - \hat{g}_1(t_i))(y_i - \hat{g}_2(t_i)).$$

Based on the above estimator for $\hat{\beta}$, the weighted least square estimator $\hat{\theta}$ of θ is defined as

$$\hat{\theta} = \arg\min_{\theta}\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i)}\Big(y_i - x_i^\tau\hat{\beta} - g(t_i,\theta)\Big)^2,$$

or

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\pi(z_i, \hat{\alpha})} \left(y_i - x_i^{\tau} \hat{\beta} - g(t_i, \theta) \right)^2,$$

which is dependent on whether the inverse probability function is estimated parametrically or nonparametrically.

When the dimension of Z is not high and we do not have the idea about the form of $\pi(Z)$, we can apply the following kernel estimation method to estimate $\pi(Z)$, thats,

$$\hat{\pi}(z_i) = \frac{\sum_{j=1}^{n} \delta_j K_h(z_i - z_j)}{\sum_{j=1}^{n} K_h(z_i - z_j)}. \tag{6}$$

However, as verified in the literature, when the dimension of Z is high, the fully nonparametric kernel estimator may suffer from the curse of dimensionality and impede its use in practice. At this time, if we have some a priori knowledge about the structure of $\pi(Z)$ which is also very common in practice, parametric estimation is another alterative method. For example, we assume that $\pi(Z, \alpha)$ to be a logistic function based on $\delta_i, y_i, t_i, i = 1, \ldots, n$. More specifically, we suppose $\pi(z_i, \alpha) = (1 + \exp(-\alpha_0 - \alpha_1 y_i - \alpha_2 t_i - \alpha_3^{\tau} v_i))^{-1}$ where $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)^{\tau}$ is an unknown vector parameter. We denote the maximum likelihood estimator of α as $\hat{\alpha} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)^{\tau}$, and the corresponding estimator of $\pi(z, \alpha)$ is

$$\pi(z_i, \hat{\alpha}) = (1 + \exp(-\hat{\alpha}_0 - \hat{\alpha}_1 y_i - \hat{\alpha}_2 t_i - \hat{\alpha}_3^{\tau} v_i))^{-1}. \tag{7}$$

When $\pi(Z)$ is not parametric, the parametric method may obtain inconsistent estimator. However, from the numerical analysis of the paper, the test (5) is robust to the missing mechanisms. The estimators $\hat{\pi}(z_i)$ and $\pi(z_i, \hat{\alpha})$ in (4) and (5) are from that in (6) and (7), respectively.

2.2 Asymptotic behavior of the test statistics

Let $\pi'(Z, \alpha) = \mathrm{grad}_{\alpha}(\pi(Z, \alpha))$, $g'(T, \theta) = \mathrm{grad}_{\theta}(g(T, \theta))$, $\Gamma = (1, Z)$, $\Sigma_{\alpha} = E(\pi(Z, \alpha)(1 - \pi(Z, \alpha)\Gamma^{\tau}\Gamma)$ and $\Sigma_{\theta} = E(g'(T, \theta)^{\tau} g'(T, \theta))$. Under mild conditions, see Jennrich (1969), we have

$$\sqrt{n}(\hat{\theta} - \theta) = \Sigma_{\theta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g'(t_i, \theta)^{\tau} \frac{\delta_i}{\hat{\pi}(z_i)} (y_i - x_i^{\tau} \hat{\beta} - g(t_i, \theta)) + o_p(1), \tag{8}$$

For the MLE estimator of α, $\hat{\alpha}$, we have

$$\sqrt{n}(\hat{\alpha} - \alpha) = \Sigma_{\alpha}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma^{\tau} (\delta_i - \pi(z_i, \alpha)) + o_p(1), \tag{9}$$

where the $\hat{\pi}(z_i)$ should be updated as $\pi(z_i, \hat{\alpha})$ when $\pi(z_i)$ is estimated parametrically.

We first introduce some notations that are related to the limiting variance of the test statistic. Let $\Sigma_0 = E[\delta(X - g_1(T))(X - g_1(T))^\tau]$, and

$$\begin{aligned}
&A_1(t_i, \theta) = 1 - Eg'(T, \theta)\Sigma_\theta^{-1} g'(t_i, \theta)^\tau, \\
&A_2(\theta) = E(X^\tau) - Eg'(T, \theta)\Sigma_\theta^{-1} E(g'(T, \theta)^\tau X^\tau), \\
&M = E((1 - \pi(Z, \alpha))\Gamma A_1(T, \theta)\epsilon), \\
&L_1(x_i, t_i, \theta) = A_1(t_i, \theta) - A_2(\theta)\Sigma_0^{-1}(x_i - g_1(t_i))\pi(z_i), \\
&L_2(x_i, t_i, \theta) = A_1(t_i, \theta) - A_2(\theta)\Sigma_0^{-1}(x_i - g_1(t_i))\pi(z_i, \alpha)
\end{aligned}$$

Under null hypothesis, the asymptotic properties of $T_{ni} (i = 1, 2)$ in (4) and (5) are as follows.

Theorem 1 *Under H_0 and the conditions in appendix, we have*

$$T_{n1} \to N(0, V_1) \text{ and } T_{n2} \to N(0, V_2),$$

where

$$\begin{aligned}
V_1 &= E\Big(\frac{\delta\epsilon}{\pi(Z)} L_1(X, T, \theta) + \frac{\pi(Z) - \delta}{\pi(Z)} A_1(T, \theta)E(\epsilon|Z)\Big)^2, \\
V_2 &= E\Big(\frac{\delta\epsilon}{\pi(Z, \alpha)} L_2(X, T, \theta) - M\Sigma_\alpha^{-1}\Gamma^\tau(\delta - \pi(Z, \alpha))\Big)^2.
\end{aligned}$$

We now investigate the sensitive of the tests for a sequence of local alternatives with the form

$$H_{1n} : g(T) = g(T, \theta) + C_n G(T) + \eta, \tag{10}$$

where $E(\eta|T) = 0$ and the function $G(\cdot)$ satisfies $E(G^2(T)) < \infty$. Then we have the following theorem under H_{1n},

Theorem 2 *Assume the same conditions as Theorem 1, under local alternatives H_{1n}, we have,*

(i) *If $n^{1/2}C_n \to 1$, $T_{n1} \to N(\mu_1, V_1)$ and $T_{n2} \to N(\mu_2, V_2)$, where $\mu_1 = \mu_2 = E(G(T)A_1(T, \theta))$;*
(ii) *If $n^r C_n \to a$ with $0 < r < 1/2$ and $a \neq 0$, then $T_{n1} \to \infty$ and $T_{n2} \to \infty$.*

We realize, from Theorem 2, that, when local alternatives are distinct from the null hypothesis at the rate n^{-r} with $0 < r < 1/2$, the proposed test in the paper can have asymptotic power 1; when that are distinct from the null hypothesis at the rate $n^{-1/2}$, the test can also detect alternatives. The rate $n^{-1/2}$ is the possible fastest rate for lack-of-fit test.

The asymptotical properties of T_{n2} in Theorem 1 and 2 are from the assumption that $\pi(Z, \alpha)$ is specified correctly. When the assumption is violated, we denote the true probability function as $\pi_0(Z, \tilde{\alpha})$ instead of $\pi(Z, \alpha)$. In this case, the left hand side in (3) can be proved that

$$E\left(\frac{\delta}{\pi(Z,\alpha)}(Y - X^\tau\beta - g(T,\theta))\right) = E\left[\frac{\pi_0(Z,\tilde{\alpha})}{\pi(Z,\alpha)}E((Y - X^\tau\beta - g(T,\theta))|Z)\right]. \tag{11}$$

Note that $E((Y - X^\tau\beta - g(T,\theta))|Z) = E((Y - X^\tau\beta - g(T,\theta))|Y, V, T)$, which may not be equal to zero, the term in (11) cannot be proved to be zero consequently. Hence, when $\pi(z_i, \alpha)$ is misspecified, Theorem 1 and 2 are not corrected for the tests theoretically. In order to avoid this problem, in the paper, we also propose the test T_{n1} based on the nonparametric estimator of selection probability $\pi(Z)$.

Remark 1 We establish the asymptotic normality of the test T_{n2} in Theorem 1 and 2 when $\pi(Z, \alpha)$ is logistic regression. If $\pi(Z, \alpha)$ is any other parametric function instead of logistics regression function, we can similarly obtain the corresponding asymptotic property of the test T_{n2} by modifying the asymptotical behavior of $\hat{\alpha}$ in (9). Other popular parametric methods, such as generalized estimating equations (GEE), can also be applied to estimate the parameter α. We only need to update the asymptotic expansion of $\hat{\alpha}$ derived from other estimation procedures and the asymptotic result for the test T_{n2} should be changed correspondingly.

3 Numerical analysis

3.1 Simulation study

In this section, we report results from several simulation studies to evaluate the finite sample behavior of the proposed test statistics, and generate 2,000 simulated data sets for all simulations. We take $K(u) = 15/16(1 - u^2)^2$, if $|u| \leq 1$; 0 otherwise as the kernel function. Though bandwidth selection has been studied extensively in nonparametric estimation problem, it is still an open problem in model checking area as pointed out by Zhu and Ng (2003) and Zhu (2005). From our experience, a good empirical choice of the bandwidth is $h_0 = \hat{\sigma}(T)n^{-1/3}$ here $\hat{\sigma}(T)$ is the empirical estimator of the standard deviation of variable T. Evidently, this bandwidth satisfies condition (5) in Appendix. To investigate the sensitivity of the bandwidth selection, we also consider several bandwidth selection, that is, $h_0 = \hat{\sigma}(T)n^{-1/3}$, $h_1 = 0.5\hat{\sigma}(T)n^{-1/3}$ and $h_2 = 2\hat{\sigma}(T)n^{-1/3}$.

Study 1. The data was generated according to the following partial linear model

$$Y = \beta X + 1 + T + aT^2 + \varepsilon, \tag{12}$$

where $\beta = 1$, $X \sim N(0, 1)$, $T \sim U(0, 1)$, $\varepsilon \sim N(0, 0.4)$ and $g(T) = 1 + T + aT^2$. For model (12), the testing problem is whether $g(T)$ is a linear function of $1 + T$, i.e., $H_0 : g(T) = \theta(1 + T)$ when X is missing. It is clear that $a = 0$ is corresponding to the null hypothesis and $a \neq 0$ to alternatives.

Table 1 Simulated size and power under sample size $n = 100$, missing mechanisms $\pi_1(y, t)$, and different a for Study 1

a	T_{n1}			T_{n2}		
	h_1	h_0	h_2	h_1	h_0	h_2
0.0	0.054	0.052	0.061	0.051	0.055	0.055
0.2	0.060	0.065	0.072	0.102	0.100	0.105
0.4	0.186	0.185	0.200	0.259	0.238	0.248
0.6	0.398	0.390	0.410	0.477	0.482	0.480
0.8	0.652	0.659	0.660	0.719	0.737	0.739
1.0	0.840	0.868	0.856	0.879	0.885	0.900
1.2	0.950	0.950	0.952	0.963	0.963	0.971
1.4	0.987	0.987	0.983	0.995	0.994	0.989
1.6	0.997	0.999	0.998	0.995	0.998	0.997
1.8	0.999	1.000	1.000	0.999	1.000	0.999
2.0	1.000	1.000	1.000	0.999	1.000	1.000

Two missing probability mechanisms are chosen as follows:

$$\pi_1(y, t) = P(\delta = 1|Y = y, T = t) = 1/(1 + \exp(-(1 + y + t))),$$
$$\pi_2(y, t) = P(\delta = 1|Y = y, T = t) = 1/(1 + 0.5y^2/(y^2 + t^2)).$$

For the above two cases, the mean response rates are $E\pi_1(y, t) \approx 0.91$ and $E\pi_2(y, t) \approx 0.71$, respectively. Furthermore, $\pi_1(y, t)$ is a parametric function with logistic form, while $\pi_1(y, t)$ does not have this form.

In this simulation, by varying the values of a in (12), the sample size $n = 100, 200$ and missing mechanism $\pi_i(y, t)$ (i = 1, 2), we study the size and power performance of the proposed tests. Also, the effect of the bandwidth on the performance of the test is studied by choosing $h = h_0, h_1$ and h_2. According to our simulation results, the size and power of the test statistics are not too sensitive for bandwidth $h = h_0, h_1, h_2$, we report the simulation results with sample size $n = 100$ and missing mechanism $\pi_1(y, t)$ in Table 1 for space consideration. All the simulation runs are shown in Fig. 1 for fixed bandwidth $h = h_0$.

From Table 1, we can know clearly that the bandwidth has little influence on the behavior of both tests. For example, the powers for T_{n1} with $n = 100$ and $a = 0.80$ are 0.652, 0.659 and 0.660, which are corresponding to $h = h_1, h_0$ and h_2, respectively. While for T_{n2}, the corresponding powers at this time are 0.719, 0.737 and 0.739, which are corresponding to $h = h_1, h_0$ and h_2, respectively.

Figure 1a and b present the plots with missing probability $\pi_1(y, t)$, while the plots with $\pi_2(y, t)$ are depicted in Fig. 1c and d. From these two figures, we can observe that our proposed tests T_{n1} and T_{n2} maintain the significance level very well. For the alternative hypothesis, we can notice that the power increases quickly as a in (12) increases, i.e., the tests are very sensitive to the alternatives. Also, we can find that when the sample size is $n = 200$, the power performances of both tests improve much compared with that under sample size $n = 100$. Further, the proposed tests with

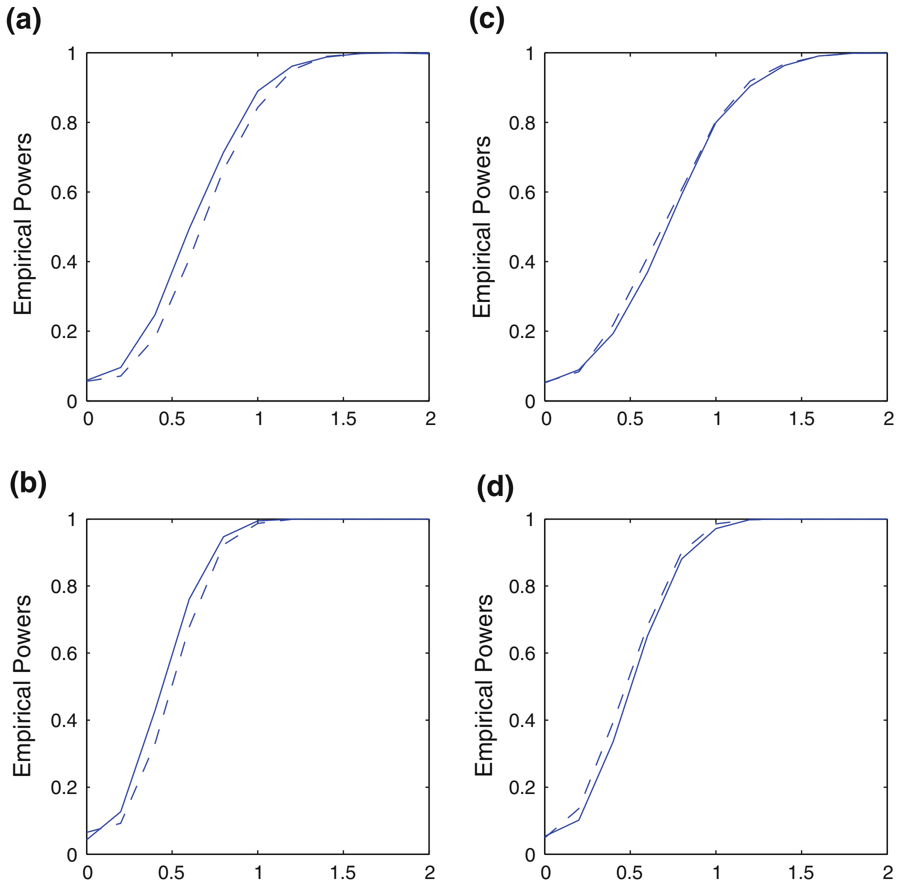

Fig. 1 Empirical powers of tests for Study 1 with $n = 100$ and $n = 200$: **a** for $\pi_1(y, t)$ and $n = 100$; **b** for $\pi_1(y, t)$ and $n = 200$; **c** for $\pi_2(y, t)$ and $n = 100$; **d** for $\pi_2(y, t)$ and $n = 200$; The *dashed line* is for T_{n1}, and the *solid line* is for T_{n2}. The model for generating data is $Y = \beta X + 1 + T + aT^2 + \varepsilon$, where $\beta = 1$, $X \sim N(0, 1)$, $T \sim U(0, 1)$, $\varepsilon \sim N(0, 0.4)$ and $g(T) = 1 + T + aT^2$. The null hypothesis is $H_0 : g(T) = \theta(1 + T)$

missing probability mechanisms $\pi_1(y, t)$ is more efficient than that with $\pi_2(y, t)$. Note that the mean response rates of $\pi_1(y, t)$ is larger than that of $\pi_2(y, t)$; we can conclude that larger sample size or more information generally improves the performance of the tests.

We turn to compare the tests T_{n1} and T_{n2}. We find that when the missing probability is $\pi_1(y, t)$, which is a logistics function, T_{n2} is more powerful than T_{n1}, while they are going in the opposite way with nonparametric missing mechanism $\pi_2(y, t)$. When we know the parametric form of the missing mechanism, using this information can generally produce more efficient checking procedure. On the other hand, if we misspecify the missing mechanism, parametric procedure may result in inferior power performance. However, we also notice that in this case, the performance of T_{n2} is still comparable to that of T_{n1}.

Study 2. We generate the data according to the following model

$$Y = \beta_1 X_1 + \beta_1 X_2 + \beta_2 X_3 + 1 + T + a\sin(2\pi T) + \varepsilon, \tag{13}$$

where $\beta_1 = \beta_1 = \beta_2 = 1$, $X_1, X_2, X_3 \sim N(0, 1)$, $T \sim U(0, 1)$, $\varepsilon \sim N(0, 0.4)$, and $g(T) = 1 + T + a\sin(2\pi T)$. For model (13), we check whether $g(T)$ is a linear function, i.e., $H_0 : g(T) = \theta(1 + T)$. The null hypothesis is true when $a = 0$, and $a \neq 0$ is corresponding to the alternative hypothesis. We also assume X_1 missing is missing according to the following missing mechanisms:

$$\pi_1(y, x_2, x_3, t) = 1/(1 + 0.25|y/(y + x_2 + x_3 + t)|);$$
$$\pi_2(y, x_2, x_3, t) = 1/(1 + (y^2/(y^2 + x_2^2 + x_3^2 + t^2))).$$

At this time, the mean response rates are $E\pi_1(y, x_2, x_3, t) \approx 0.83$ and $E\pi_2(y, x_2, x_3, t) \approx 0.67$, respectively. For this study, we investigate the effect of dimension of the variables $Z = (Y, X_2, X_3, T)$ on T_{n1} and the robustness of T_{n2} to the nonparametric missing mechanisms.

We show the results in Fig. 2a and b with $\pi_1(y, x_2, x_3, t)$ and in Fig. 2c and d with $\pi_2(y, x_2, x_3, t)$. From this Figure, we can get the similar conclusions as Study 1, except the following findings. When the dimension of Z is four, the performance of T_{n1} is comparable. In other words, T_{n1} does not suffer the curse of dimension too much. Meanwhile, the test T_{n2} is robust to the missing mechanism, and it can perform very well even when the missing mechanism is not parametric.

3.2 Real data analysis

In this section, we apply our method to the analysis of an AIDS clinical trial group (ACTG 315) study. There are 53 HIV-1-infected patients treated, of which five patients quitted the study due to drug intolerance and other problems. Hence, 48 evaluable patients enrolled in ACTG 315. The response variable is viral load and the covariates are CD4+ cell counts and treatment time. All the patients are repeatedly measured, and a total of 317 observations are available from 48 patients with 64 CD4+ cell counts missing. That is, the missing proportion is 20.19 %. The data set has been studied by Wu and Wu (2001, 2002), and Yang et al. (2009) etc.

In general, immunologic response (measured by CD4+ cell count) and the virologic response (measured by viral load) are negatively correlated during antiviral treatments. It is suggested by Liang et al. (2004) that the impact of CD4+ cell count and treatment time on the viral load is linear and nonparametric, respectively. However, it is important to check whether the assumption about the nonparametric relationship between the viral load and the treatment time is validated or not. Note that this data set is actually longitudinal. We ignore the correlation structure for hypothesis testing since the proposed tests can be extended to a working independence analysis of longitudinal data, that is, the correlation structure is ignored as in Liang et al. (2004) and Yang et al. (2009).

Let Y be the viral load, T be the treatment time and X be the CD4+ cell counts. Denote $\delta = 0$ if CD4+ is missing, otherwise $\delta = 1$. The following model is considered

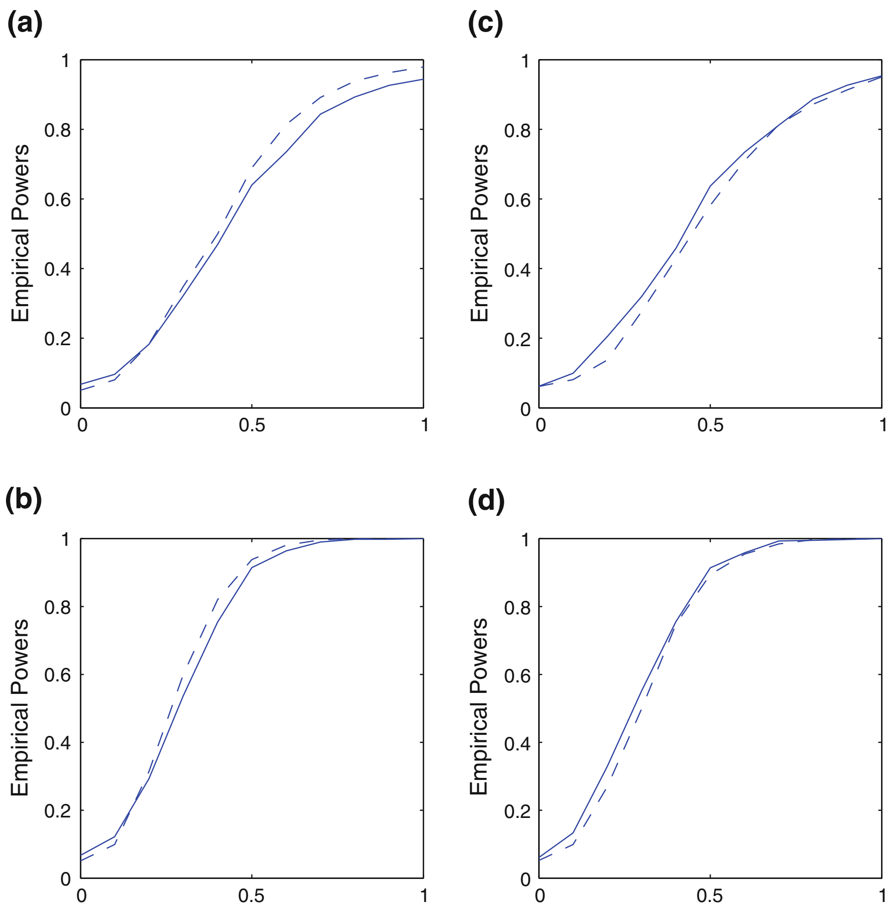

Fig. 2 Empirical powers of tests for Study 2 with $n = 50$ and $n = 100$: **a** for $\pi_1(y, x_2, x_3, t)$ and $n = 50$; **b** for $\pi_1(y, x_2, x_3, t)$ and $n = 100$; **c** for $\pi_2(y, x_2, x_3, t)$ and $n = 50$; **d** for $\pi_2(y, x_2, x_3, t)$ and $n = 100$; The *dashed line* is for T_{n1}, and the *solid line* is for T_{n2}. The model for generating data is $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_2 X_3 + 1 + T + a \sin(2\pi T) + \varepsilon$, where $\beta_1 = \beta_2 = \beta_3 = 1,\ X_i \sim N(0, 1), (i = 1, 2, 3),\ T \sim U(0, 1),\ \varepsilon \sim N(0, 0.4)$ and $g(T) = 1+T+a\sin(2\pi T)$. The null hypothesis is $H_0 : g(T) = \theta(1 + T)$

for data fitting:

$$Y = X\beta + g(T) + \varepsilon. \tag{14}$$

For models (14), we want to check whether the term $g(T)$ in (14) is linear or not. Both the p values for T_{n1} and T_{n2} are 0.000 for the null hypothesis $H_0 : g(T) = \theta T$. Thus we can reject the null hypothesis, and the linear form of $g(T)$ is not feasible. The result suggests that the nonparametric function $g(T)$ is appropriate in model (14) as in Liang et al. (2004) and Yang et al. (2009).

Appendix: Proof of the theorems

The following conditions are required for the theorems in Sect. 2.

1. $g(\cdot,\theta)$ is continuously differentiable with respect to θ in the interior set of Θ, and $g_1(\cdot)$ and $g_2(\cdot)$ satisfy Lipschitz condition of order 1;
2. $\pi(z)$ has bounded partial derivatives up to order $k(>2)$ almost surely;
3. Σ_0, Σ_α, Σ_θ are all positive definite matrix;
4. $\sup E(\varepsilon^2|X=x, T=t) < c_1$ for some c_1 and all x and t, $E|X|^4 < \infty$, and $E|Y|^4 < \infty$;
5. As $n\to\infty$, $\sqrt{n}h^2 \to 0$, and $\sqrt{n}h \to \infty$;
6. The density of Z, say $f(z)$ on support $\mathcal{C}$, exists and has bounded derivatives up to order 2 and satisfies

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

7. The continuous kernel function $K(\cdot)$ satisfies: (i) the support of $K(\cdot)$ is the interval $[-1,1]$; (ii) $K(\cdot)$ is symmetric about 0; iii) $\int_{-1}^{1} K(u)\mathrm{d}u = 1$ and $\int_{-1}^{1} |u|K(u)\mathrm{d}u \neq 0$.

Remark 2 Conditions (1), (5) and (7) are typical for obtaining convergence rates when nonparametric estimation is applied. Condition (2) is a common assumption in missing data study, which is also used in Wang et al. (2004), and so on. The conditions (3) and (4) are necessary for the asymptotic normality of the least squares estimator. Condition (6) is a typical condition for avoiding the boundary effect for nonparametric estimate.

Lemma 1 *Under conditions 1–7 in the Appendix, the asymptotic properties of $\sqrt{n}(\hat\beta - \beta)$ under the null hypothesis in (2) or the local alternatives in (10) are the same as follows*

$$\sqrt{n}(\hat\beta-\beta) = \frac{\Sigma_0^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \delta_i (x_i - g_1(t_i))\epsilon_i + o_p(1).$$

Proof for Lemma 1 Under the null hypothesis, we have

$$\begin{aligned}
\sqrt{n}(\hat\beta-\beta) &= \frac{\Sigma_0^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \delta_i (x_i - g_1(t_i))\{y_i - g_2(t_i) - (x_i - g_1(t_i))^\tau \beta\} + o_p(1) \\
&= \frac{\Sigma_0^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \delta_i (x_i - g_1(t_i))(y_i - x_i^\tau\beta - g(t_i,\theta)) + o_p(1) \\
&= \frac{\Sigma_0^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \delta_i (x_i - g_1(t_i))\epsilon_i + o_p(1).
\end{aligned}$$

Under the local alternatives, we have

$$\begin{aligned}\sqrt{n}(\hat{\beta}-\beta) &= \frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(x_i-g_1(t_i))\{y_i-g_2(t_i)-(x_i-g_1(t_i))^\tau\beta\}+o_p(1)\\ &= \frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(x_i-g_1(t_i))(y_i-x_i^\tau\beta-g(t_i,\theta)-C_nG(t_i))+o_p(1)\\ &= \frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(x_i-g_1(t_i))\epsilon_i+o_p(1).\end{aligned}$$

Thus, Lemma 1 is proved. □

Proof of Theorem 1 First, we prove the asymptotical distribution of T_{n1} under null hypothesis. It can be verified that

$$\begin{aligned}T_{n1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i)}(y_i-x_i^\tau\beta-g(t_i,\theta))-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i)}x_i^\tau(\hat{\beta}-\beta)\\ &\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i)}g'(t_i,\theta)(\hat{\theta}-\theta)\\ &= T_{n11}-T_{n12}-T_{n13}.\end{aligned} \tag{15}$$

For T_{n12} in (15), it can be verified that

$$\begin{aligned}T_{n12} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi(z_i)}x_i^\tau(\hat{\beta}-\beta)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\pi(z_i)-\hat{\pi}(z_i))\delta_i}{\pi^2(z_i)}x_i^\tau(\hat{\beta}-\beta)+o_p(1)\\ &= E(X^\tau)\sqrt{n}(\hat{\beta}-\beta)\\ &\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\sum_{j=1}^{n}(\pi(z_i)-\delta_j)K_h(z_j-z_i)\delta_i x_i^\tau}{\pi^2(z_i)nf(z_i)}(\hat{\beta}-\beta)+o_p(1)\\ &= E(X^\tau)\sqrt{n}(\hat{\beta}-\beta)+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\pi(z_j)-\delta_j}{\pi(z_j)}E(X^\tau|z_j)(\hat{\beta}-\beta)+o_p(1)\\ &= E(X^\tau)\sqrt{n}(\hat{\beta}-\beta)+o_p(1).\end{aligned} \tag{16}$$

The last equation follows from the fact that $n^{-1}\sum_{i=1}^{n}(\pi(z_i)-\delta_i)\pi(z_i)^{-1}E(X^\tau|z_i)=o_p(1)$ and $\sqrt{n}(\hat{\beta}-\beta)=O_p(1)$.

For T_{n13} in (15), we have

$$\begin{aligned}T_{n13} &= E(g'(T,\theta))\sqrt{n}(\hat{\theta}-\theta)\\ &= E(g'(T,\theta))\Sigma_\theta^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g'(t_i,\theta)^\tau \frac{\delta_i}{\hat{\pi}(z_i)}(y_i - x_i^\tau\hat{\beta} - g(t_i,\theta)) + o_p(1)\\ &= E(g'(T,\theta))\Sigma_\theta^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i}{\hat{\pi}(z_i)}g'(t_i,\theta)^\tau(\epsilon_i - x_i^\tau(\hat{\beta}-\beta)) + o_p(1)\\ &= E(g'(T,\theta))\Sigma_\theta^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g'(t_i,\theta)^\tau \frac{\delta_i\epsilon_i}{\hat{\pi}(z_i)}\\ &\quad - E(g'(T,\theta))\Sigma_\theta^{-1}E(g'(T,\theta)^\tau X^\tau)\sqrt{n}(\hat{\beta}-\beta) + o_p(1).\end{aligned} \tag{17}$$

Based on the expressions (15), (16) and (17), and note that

$$\begin{aligned}A_1(t_i,\theta) &= 1 - Eg'(T,\theta)\Sigma_\theta^{-1}g'(t_i,\theta)^\tau;\\ A_2(\theta) &= E(X^\tau) - Eg'(T,\theta)\Sigma_\theta^{-1}E(g'(T,\theta)^\tau X^\tau),\end{aligned}$$

we can get

$$T_{n1} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i\epsilon_i}{\hat{\pi}(z_i)}A_1(t_i,\theta) - A_2(\theta)\sqrt{n}(\hat{\beta}-\beta) + o_p(1). \tag{18}$$

According to Lemma 1, for $\hat{\beta}$, we have

$$\sqrt{n}(\hat{\beta}-\beta) = \Sigma_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i(x_i - g_1(t_i))\epsilon_i + o_p(1).$$

Note that,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i\epsilon_i}{\hat{\pi}(z_i)}A_1(t_i,\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i\epsilon_i + (\pi(z_i)-\delta_i)E(\epsilon|z_i)}{\pi(z_i)}A_1(t_i,\theta) + o_p(1),$$

the expression T_{n1} in (18) can be further derived as:

$$\begin{aligned}T_{n1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i\epsilon_i}{\pi(z_i)}\{A_1(t_i,\theta) - A_2(\theta)\Sigma_0^{-1}(x_i - g_1(t_i))\pi(z_i)\}\\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\pi(z_i)-\delta_i}{\pi(z_i)}A_1(t_i,\theta)E(\epsilon|z_i) + o_p(1).\end{aligned}$$

Then the asymptotical distribution of T_{n1} in Theorem 1 is proved.

Below we prove the asymptotical distribution of T_{n2}, it can be easily derived that

$$\begin{aligned}
T_{n2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i,\alpha)}(y_i - x_i^{\tau}\beta - g(t_i,\theta)) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i,\alpha)}x_i^{\tau}(\hat{\beta}-\beta)\\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(z_i,\alpha)}g'(t_i,\theta)(\hat{\theta}-\theta)\\
&= T_{n21} - T_{n22} - T_{n23}.
\end{aligned} \tag{19}$$

For T_{n22} in (19), we have

$$\begin{aligned}
T_{n22} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\frac{\delta_i x_i^{\tau}}{\pi(z_i,\alpha)}(\hat{\beta}-\beta) - \frac{(\pi(z_i,\hat{\alpha})-\pi(z_i,\alpha))\delta_i x_i^{\tau}}{\pi^2(z_i,\alpha)}(\hat{\beta}-\beta)\Big] + o_p(1)\\
&= T_{n22,1} - T_{n22,2} + o_p(1).
\end{aligned} \tag{20}$$

Note that $\Gamma_i = (1, z_i)$, $\sqrt{n}(\hat{\alpha}-\alpha) = O_p(1)$ and $\pi(z_i,\hat{\alpha}) - \pi(z_i,\alpha) = \pi'(z_i,\alpha)(\hat{\alpha} - \alpha) + o_p(n^{-1/2}) = \pi(z_i,\alpha)(1-\pi(z_i,\alpha))\Gamma_i(\hat{\alpha}-\alpha) + o_p(n^{-1/2})$, for $T_{n22,2}$, we have

$$T_{n22,2} = E((1-\pi(Z,\alpha))\Gamma X^{\tau})\sqrt{n}(\hat{\alpha}-\alpha)(\hat{\beta}-\beta) + o_p(1) = o_p(1).$$

Consequently, $T_{n22} = E(X^{\tau})\sqrt{n}(\hat{\beta}-\beta)$.

For T_{n23} in (19), it can be proved that

$$\begin{aligned}
T_{n23} &= Eg'(T,\theta)\sqrt{n}(\hat{\theta}-\theta)\\
&= Eg'(T,\theta)\Sigma_{\theta}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g'(t_i,\theta)^{\tau}\frac{\delta_i\epsilon_i}{\pi(z_i,\hat{\alpha})}\\
&\quad -Eg'(T,\theta)\Sigma_{\theta}^{-1}E(g'(T,\theta)^{\tau}X^{\tau})\sqrt{n}(\hat{\beta}-\beta) + o_p(1).
\end{aligned} \tag{21}$$

According to the Eqs. (19), (20) and (21), we have

$$\begin{aligned}
T_{n2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\pi(z_i,\hat{\alpha})}A_1(t_i,\theta) - A_2(\theta)\sqrt{n}(\hat{\beta}-\beta) + o_p(1)\\
&= \bar{T}_{n21} - \bar{T}n22 + o_p(1).
\end{aligned} \tag{22}$$

For $\bar{T}_{n21}$, it can be verified that

$$\begin{aligned}
\bar{T}_{n21} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_1(t_i,\theta)\Big[\frac{\delta_i\epsilon_i}{\pi(z_i,\alpha)} - \frac{(\pi(z_i,\hat{\alpha})-\pi(z_i,\alpha))\delta_i\epsilon_i}{\pi^2(z_i,\alpha)}\Big] + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_1(t_i,\theta)\frac{\delta_i\epsilon_i}{\pi(z_i,\alpha)} - M\sqrt{n}(\hat{\alpha}-\alpha),
\end{aligned} \tag{23}$$

where $M = E((1-\pi(Z,\alpha))\Gamma A_1(T,\theta)\epsilon)$.

Following lemma 1 and the Eqs. (22) and (23), we obtain

$$T_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\pi(z_i,\alpha)}\Big[A_1(t_i,\theta) - A_2(\theta)\Sigma_0^{-1}(x_i - g_1(t_i))\pi(z_i,\alpha)\Big]$$
$$-M\Sigma_\alpha^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma^\tau(\delta_i - \pi(z_i,\alpha)) + o_p(1).$$

Then the asymptotical distribution of T_{n2} in Theorem 1 is proved, and Theorem 1 is proved. □

Proof of Theorem 2 For T_{n1}, it can be divided as

$$T_{n1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(\eta_i + C_n G(t_i))}{\hat{\pi}(z_i)}A_1(t_i,\theta) - A_2(\theta)\sqrt{n}(\hat{\beta} - \beta) + o_p(1)$$
$$= \tilde{T}_{n11} + \tilde{T}_{n12} - \tilde{T}_{n13} + o_p(1).$$

For $\tilde{T}_{n12}$, we can easily obtain

$$\tilde{T}_{n12} = \frac{C_n}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i G(t_i)}{\pi(z_i)}A_1(t_i,\theta) + \frac{C_n}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\pi(z_i) - \delta_i)G(t_i)}{\pi(z_i)}A_1(t_i,\theta) + o_p(1)$$
$$= C_n\sqrt{n}E(G(T)A_1(T,\theta)) + o_p(1),$$

where the last equation follows according to the fact that

$$Var\left(\frac{C_n}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\pi(z_i) - \delta_i)G(t_i)A_1(t_i,\theta)}{\pi(z_i)}\right) = C_n^2 Var\left(\frac{(\pi(Z) - \delta)G(T)A_1(T,\theta)}{\pi(Z)}\right) \to 0.$$

From Lemma 1, we know

$$\sqrt{n}(\hat{\beta} - \beta) = \Sigma_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(x_i - G_1(t_i))\epsilon_i + o_p(1).$$

As a result, if $n^{1/2}C_n \to 1$, we can obtain

$$T_{n1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\pi(z_i)}\{A_1(t_i,\theta) - A_2(\theta)\Sigma_0^{-1}(x_i - g_1(t_i))\pi(z_i)\}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\pi(z_i) - \delta_i}{\pi(z_i)}A_1(t_i,\theta)E(\epsilon|z_i) + E(G(T)A_1(T,\theta)) + o_p(1).$$

If $n^r C_n \to a, 0 < r < 1/2$, then it yields $\sqrt{n}C_n \to \infty$, as $n \to \infty$. As a result, we have $T_{n1} \to \infty$.

We investigate the asymptotic property of T_{n2} below. Note that

$$\begin{aligned} T_{n2} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i(\eta_i + C_n G(t_i))}{\pi(z_i, \hat{\alpha})} A_1(t_i, \theta) - A_2(\theta)\sqrt{n}(\hat{\beta} - \beta) + o_p(1) \\ &= \tilde{T}_{n21} + \tilde{T}_{n22} - \tilde{T}_{n23} + o_p(1). \end{aligned}$$

For $\tilde{T}_{n22}$, it can be derived that

$$\begin{aligned} \tilde{T}_{n22} &= \frac{C_n}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i G(t_i)}{\pi(z_i, \alpha)} A_1(t_i, \theta) \\ &\quad - \frac{C_n}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\pi(z_i, \hat{\alpha}) - \pi(z_i, \alpha))\delta_i G(t_i)}{\pi^2(z_i, \alpha)} A_1(t_i, \theta) + o_p(1) \\ &= C_n \sqrt{n} E(G(T) A_1(T, \theta)) \\ &\quad - C_n E[(1 - \pi(Z, \alpha))G(T)A_1(T, \theta)\Gamma]\sqrt{n}(\hat{\alpha} - \alpha) + o_p(1) \\ &= C_n \sqrt{n} E(G(T) A_1(T, \theta)) + o_p(1). \end{aligned}$$

The last equation follows because $\sqrt{n}(\hat{\alpha} - \alpha) = O_p(1)$ and $C_n \to 0$.

Consequently, according to Lemma 1, we have

$$\begin{aligned} T_{n2} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i \epsilon_i}{\pi(z_i, \alpha)} \Big[A_1(t_i, \theta) - A_2(\theta)\Sigma_0^{-1}(x_i - g_1(t_i))\pi(z_i, \alpha) \Big] \\ &\quad - M\Sigma_\alpha^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma^\tau(\delta_i - \pi(z_i, \alpha)) + C_n \sqrt{n} E(G(T) A_1(T, \theta)). \end{aligned}$$

The asymptotic properties of T_{n2} can be obtained based on the above equation. Thus, Theorem 2 is proved. □

Acknowledgments The authors would like to thank two referees for their constructive comments which led to a much improved presentation of the paper. The research was supported by the National Natural Science Foundation of China (No. 11071253) and Beijing Nova Programme (2010B066).

References

Fan, Y., Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica, 64*, 865–890.

Jennrich, R. I. (1969). Asymptotic properties of non-least squares estimators. *Annals of Mathematical Statistics, 40*, 633–643.

Li, C. S. (2009). Using P-splines to test the linearity of partially linear models. *Statistical Methodology, 6*, 542–552.

Li, N., Xu, X. Z., Jin, P. (2011). Testing the linearity in partially linear models. *Journal of Nonparametric Statistics, 23*, 99–114.

Liang, H. (2006). Checking linearity of nonparametric component in partially linear models with an application in systemic inflammatory response syndrome study. *Statistical Methods in Medical Research, 15*, 273–284.

Liang, H., Wang, S. J., Robins, J., Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. *Journal of the American Statistical Association, 99*, 357–367.

Liang, H., Wang, S. J., Carroll, R. J. (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika, 94*, 185–198.

Little, R. J. A., Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Robins, J., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association, 89*, 846–866.

Sun, Z. H., Wang, Q. H., Dai, P. J. (2009). Model checking for partially linear models with missing responses at random. *Journal of Multivariate Analysis, 100*, 636–651.

Wang, Q. H. (2009). Statistical estimation in partial linear models with covariables data missing at random. *Annals of the Institute of Statistical Mathematics, 61*, 47–84.

Wang, Q. H., Sun, Z. H. (2007). Estimation in partially linear models with missing responses at random. *Journal of Multivariate Analysis, 98*, 1470–1493.

Wang, Q. H., Lindon, O., Hädle, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association, 99*, 334–345.

Whang, Y., Andrews, D. W. K. (1993). Tests of specification for parametric and semiparametric models. *Journal of Econometrics, 57*, 277–318.

Wu, H., Wu, L. (2001). A multiple imputation method for missing covariates in nonlinear mixed-effect models, with application to HIV dynamics. *Statistics in Medicine, 20*, 1755–1769.

Wu, L., Wu, H. (2002). Nonlinear mixed-effect models with missing time-dependent covariates, with application to HIV viral dynamics. *Journal of the Royal Statistical Society C., 51*, 297–318.

Xu, W. L., Guo, X., Zhu, L. X. (2012). Goodness-of-fitting for partial linear model with missing response at random. *Journal of Nonparametric Statistics, 24*, 103–118.

Yang, Y. P., Xue, L. G., Cheng, W. H. (2009). Empirical likelihood for a partially linear model with covariate data missing at random. *Journal of Statistical Planning and Inference, 139*, 4143–4153.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory, 8*, 435–451.

Zhu, L. X. (2005). *Nonparametric Monte Carlo tests and their applications*. New York: Springer.

Zhu, L. X., Ng, K. W. (2003). Checking the adequacy of a partial linear model. *Statistica Sinica, 13*, 763–781.