Ann Inst Stat Math (2015) 67:793–815
DOI 10.1007/s10463-014-0476-7

Nonparametric check for partial linear errors-in-covariables models with validation data

Wangli Xu · **Lixing Zhu**

Received: 22 April 2013 / Revised: 8 May 2014 / Published online: 31 July 2014

Abstract In this paper, we investigate the goodness-of-fit test of partial linear regression models when the true variable in the linear part is not observable but the surrogate variable $\tilde{X}$, the variable in the non-linear part T and the response Y are exactly measured. In addition, an independent validation data set for X is available. By a transformation, it is found that we only need to check whether the linear model is plausible or not. We estimate the conditional expectation of X under a given the surrogate variable with the help of the validation sample. Finally, a residual-based empirical test for the partial linear models is constructed. A nonparametric Monte Carlo test procedure is used, and the null distribution can be well approximated even when data are from alternative models converging to the hypothetical model. Simulation results show that the proposed method works well.

Keywords Errors-in-variables model · Validation sample · Partial linear models · Goodness-of-fit testing

W. Xu (✉)
Center for Applied Statistics, School of Statistics, Renmin University of China,
No. 59 Zhongguancun Street, Haidian District, Beijing 100872, People's Republic of China
e-mail: wxu.stat@gmail.com

L. Zhu
Department of Mathematics, The Hong Kong Baptist University, FSC1208, Fong Shu Chuen Building,
Kowloon Tong, Hong Kong
e-mail: lzhu@hkbu.edu.hk

1 Introduction

The following partial linear model has received considerable attention:

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

where Y is a scaler response variable, X and T are respectively d-dimensional and m-dimensional covariates, β is a d-dimensional column vector denoting regression coefficients, $g(\cdot)$ is an unknown measurable function, and e is the random error with $E(e|X, T) = 0$. Here and afterward τ stands for the transpose. The expectation of X is assumed to be 0 without loss of generality, and m, the dimension of T, is assumed to be 1 for simplicity. All the theoretical results below can be extended to the case $m > 1$. There is a lot of literature on the estimation of this model. Among others, Wang (2003) developed an estimating theory when there is measurement error in the response and validation data are available.

To avoid wrong conclusions and make efficient inference, it is necessary to do a model checking before performing any further statistical analysis. There are several proposals available in the literature for model (1) under different scenarios. The null hypothesis is

$$H_0 : E(Y|X, T) = X^{\tau}\beta + g(T), \tag{2}$$

for some β and $g(\cdot)$. When all involved variables, (Y, X, T), are observable, Zhu and Ng (2003), among others, were the first to consider testing this null hypothesis. Relevant methods are found in Stute and Zhu (2005) for the single-index model, and Stute et al. (2008) for parametric models.

In practice, we may encounter the situation where the covariates are not fully observed. In the presence of validation sample, one observes independent replicates $(X_i, \tilde{X}_i, T_i)$, $1 \le i \le n$. Independent of these validation data, the primary data set in which the covariates measured with errors are sampled as $(Y_j, \tilde{X}_j, T_j)$, $n + 1 \le j \le n + N$ from $(Y, \tilde{X}, T)$ rather than (Y, X, T), where the relationship between $\tilde{X}$ and X is not specified in this paper. This type of data set for errors-in-covariates models with validation sampling may emerge when not all the covariates X for the full study cohort can be exactly measured due to limited budget. If a surrogate covariate $\tilde{X}$ for X exists, we can use it in the study.

Some strategies can be used to handle the scenarios with incompletely observed data. In the situation that the covariates (X, T) are exactly measured, but the responses are missing, Sun and Wang (2009) considered the testing problem for the null hypothesis of (2). Xu et al. (2012) constructed a test to check the parametric structure of $g(\cdot)$ when the responses are missing. Xu and Guo (2013) proposed a test for the case where covariates are missing at random. Dai et al. (2010) considered the goodness-of-fit test for a general linear model when the covariates are measured with errors and an independent validation data set is observable. Wang (2003) investigated the estimation of partial linear error-in-response models with validation data. For partial linear errors-in-covariates models with validation data, the testing problem has been paid less attention. This paper considers the goodness-of-fit test for the partial linear model, in which X is measured with error and both Y and T are measured exactly. In addition, there is a surrogate variable $\tilde{X}$ for X. For this purpose, a residual-marked process-based test is suggested and no condition about the relationship between X and $\tilde{X}$ is assumed. As the limiting null distribution is intractable, nonparametric Monte Carlo method (NMCT) proposed by Zhu (2005) is applied to calculate the p value. Similarly as the conclusions in Zhu and Ng (2003), when (Y, X, T) are exactly sampled, the proposed test also has the desirable features: (i) the test is consistent for the global alternatives; (ii) the local alternatives can be detected when it is distinct from the null hypothesis at rate close to $N^{-1/2}$ enough.

The rest of this paper is organized as follows. In Sect. 2, the test statistic will be constructed and the asymptotic properties under the null and local alternative will be investigated. Section 3 presents the NMCT procedure. In Sect. 4, numerical results are reported to examine the performance of the test. The proofs of the asymptotic results are postponed to the Appendix.

2 Test statistic construction

2.1 Motivation

Suppose that the validation data set $\{(X_i, \tilde{X}_i, T_i)_{i=1}^n\}$ is independent of the primary sample set $\{(Y_j, \tilde{X}_j, T_j)_{j=n+1}^{n+N}\}$. Further, similarly as the conditions in Wang (1999), suppose that $(\tilde{X}_i, T_i), i = 1, \dots, n$ and $(\tilde{X}_j, T_j), j = n+1, \dots, n+N$ are i.i.d. covariates, and that the connection between X and $\tilde{X}$ in the primary data set is the same as that in the validation data set, although in the primary data set X is not observable. We assume that $\tilde{X}$ and e in model (1) are independent. For instance, a popular measurement error is $\tilde{e}$ with $\tilde{X} = X + \tilde{e}$ which is independent of model error e. When there exist measurement errors in covariates, regression calibration transfers the errors-in-covariates model to a classical regression model (see among others Wang 1999 and Stute et al. 2007). Thus, model (1) can be rewritten as

$$Y = u^\tau(\tilde{V})\beta + g(T) + \eta, \tag{3}$$

here $\tilde{V} = (\tilde{X}, T)$, $u(\tilde{V}) = E(X|\tilde{V})$ and $\eta = e + X^\tau\beta - u^\tau(\tilde{V})\beta$. Note that under the null hypothesis, some elementary calculation yields that

$$\begin{aligned} E(\eta|\tilde{V}) &= E[E(e + X^\tau\beta - u^\tau(\tilde{V})\beta|X, \tilde{X}, T)|\tilde{X}, T] \\ &= E[E(e|X, T) + X^\tau\beta - u^\tau(\tilde{V})\beta|\tilde{X}, T] \\ &= u^\tau(\tilde{V})\beta - u^\tau(\tilde{V})\beta = 0. \end{aligned} \tag{4}$$

Also, by the condition that $E(e|X, T) = 0$ and the independence between e and $\tilde{X}$ when X is given, we can derive easily that $\mathrm{E}(e \mid \tilde{X}, T) = 0$ as

$$E(e|\tilde{X}, T) = E[E(e|X, \tilde{X}, T)|\tilde{X}, T] = E[E(e|X, T)|\tilde{X}, T] = 0.$$

Model (3) can be further re-written as

$$Y - g_1(T) = (u(\tilde{V}) - g_2(T))^\tau \beta + \eta, \tag{5}$$

where $g_1(T) = E(Y|T)$ and $g_2(T) = E(u(\tilde{V})|T)$. Hence, the testing problem is converted to testing whether the linear model in (5) is plausible or not. Equation (4) implies that $E[\eta|\tilde{X}, T] = 0$, and this yields that

$$E[\eta I(\tilde{X} \le \tilde{x}, T \le t)] = 0$$

for all $\tilde{x}$ and t.

Therefore, the corresponding empirical version of the left-hand side of (6) can be used as the basis for constructing a test statistic:

$$T_N(\tilde{x}, t) = \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} (Y_j - \hat{u}^\tau(\tilde{V}_j)\hat{\beta} - \hat{g}(T_j)) I(\tilde{X}_j \le \tilde{x}, T_j \le t), \tag{6}$$

where $\hat{g}(T) = \hat{g}_1(T) - \hat{g}_2(T)^\tau \hat{\beta}$. The terms $\hat{u}(\tilde{V})$, $\hat{\beta}$, $\hat{g}_1(T)$, $\hat{g}_2(T)$ and $\hat{g}(T)$ are the estimators of $u(\tilde{V})$, β, $g_1(T)$, $g_2(T)$ and $g(T)$, respectively, which will be specified later. The proposed test statistic is defined as

$$CV_N = \int (T_N(\tilde{X}, T))^2 dF_N(\tilde{X}, T), \tag{7}$$

where F_N is the empirical distribution based on $\{(\tilde{X}_{n+1}, T_{n+1}), \ldots, (\tilde{X}_{n+N}, T_{n+N})\}$. The null hypothesis will be rejected under large observed values of CV_N.

It is worth mentioning that this test statistic is not scale-invariant, which seems to violate a very important requirement for test statistic construction. However, we will see that when we use a Monte Carlo procedure to determine p values, it is self-standardized. This is a desirable feature as we need not estimate variance with a complex structure. We will see this in the next section.

2.2 Estimation of β and $g(T)$

As described above, several unknowns need to be estimated under the assistance of the validation data. First the nonparametric kernel estimator $\hat{u}(\tilde{V})$ of $u(\tilde{V})$ in (3) is defined as

$$\hat{u}(\tilde{V}) = \frac{\sum_{i=1}^n X_i K_1((\tilde{V}_i - \tilde{V})/b_n)}{\sum_{i=1}^n K_1((\tilde{V}_i - \tilde{V})/b_n)} \tag{8}$$

for any $\tilde{V}$. Here, $K_1(\cdot)$ is a $d+1$-dimensional kernel function, and b_n is a bandwidth to be selected.

The estimators for $g_1(T)$ and $g_2(T)$ we use are defined as

$$\hat{g}_1(T) = \sum_{j=n+1}^{n+N} W_j(T)Y_j, \quad \text{and} \quad \hat{g}_2(T) = \sum_{j=n+1}^{n+N} W_j(T)\hat{u}(\tilde{V}_j),$$

where

$$W_j(T) = \frac{K_2((T-T_j)/h_N)}{\sum_{j=n+1}^{n+N} K_2((T-T_j)/h_N)},$$

$K_2(\cdot)$ is a one-dimensional kernel function, and h_N is also a bandwidth to be selected later.

For β, we use an estimator $\hat{\beta}$ which is the minimizer of the following over all β:

$$\sum_{j=n+1}^{n+N} \left[Y_j - \hat{u}^\tau(\tilde{V}_j)\beta - \hat{g}(T_j)\right]^2.$$

It has a closed form as

$$\hat{\beta} = \left(\frac{1}{N}\sum_{j=n+1}^{n+N} (\hat{u}(\tilde{V}_j) - \hat{g}_2(T_j))(\hat{u}(\tilde{V}_j) - \hat{g}_2(T_j))^\tau\right)^{-1} \times \frac{1}{N}\sum_{j=n+1}^{n+N} (\hat{u}(\tilde{V}_j) - \hat{g}_2(T_j))(Y_j - \hat{g}_1(T_j)).$$

Consequently, an estimator of $g(T)$ is

$$\hat{g}(T) = \hat{g}_1(T) - \hat{g}_2^\tau(T)\hat{\beta}.$$

Remark 1 Du et al. (2011) developed an estimating approach for nonparametric function with measurement errors in covariates, under the case that the explanatory variable is univariate. This paper does not investigate how to get the nonparametric estimator when the dimension of explanatory variable is more than one. In our paper, in the case that T is not observable, the corresponding data set is then constructed by a validation dataset $\{(X_i, \tilde{X}_i, \tilde{T}_i, T_i)_{i=1}^n\}$ and a primary sample set $\{(Y_j, \tilde{X}_j, \tilde{T}_j)_{j=n+1}^{n+N}\}$. Corresponding to (2.1) in the paper for constructing statistics, under this data set with t unobservable, we have

$$Y = E(X|\tilde{X}, \tilde{T})\beta + E(g(T)|\tilde{X}, \tilde{T}) + \tilde{\eta}, \tag{9}$$

where $\tilde{\eta} = e + X^\tau\beta + g(T) - E(X|\tilde{X}, \tilde{T})\beta - E(g(T)|\tilde{X}, \tilde{T})$. Because both the terms $E(X|\tilde{X}, \tilde{T})$ and $E(g(T)|\tilde{X}, \tilde{T})$ in (9) are nonparametric functions of $\tilde{X}, \tilde{T}$, it is difficult to estimate them. As a result, it is not easy to check the corresponding model when T is not observable and it deserves further study. Another issue is about curse of dimensionality when we use nonparametric estimation with high-dimensional $\tilde{V} = (\tilde{X}, T)$ such as that in (8). In our approach, the estimation still suffers from this problem, and thus it is of importance to investigate this in a further study.

2.3 Asymptotic properties of the test statistics

We now state the asymptotic properties of $T_N(\tilde{x}, t)$ and CV_N. To well present the results, let $U_I(\tilde{x}, t) = E[u^\tau(\tilde{V}) I(\tilde{X} \le \tilde{x}, T \le t)]$, $\Sigma = E\{(u(\tilde{V}) - g_2(T))(u(\tilde{V}) - g_2(T))^\tau\}$ and

$$\begin{aligned}
&J_1(Y, \tilde{X}, T; \tilde{x}, t) \\
&\quad = (Y - u^\tau(\tilde{V})\beta - g(T)) I(\tilde{X} \le \tilde{x}, T \le t) - (Y - g_1(T)) F[\tilde{X}|T] I(T \le t) \\
&\qquad - U_I(\tilde{x}, t)\Sigma^{-1}(u(\tilde{V}) - g_2(T))[u(\tilde{V}) - g_2(T)]^\tau \beta, \\
&J_2(\tilde{X}, X, T; \tilde{x}, t) \\
&\quad = (X - u(\tilde{V}))^\tau \beta I(\tilde{X} \le \tilde{x}, T \le t) - U_I(\tilde{x}, t)\Sigma^{-1}(u(\tilde{V}) - g_2(T)) \\
&\qquad \{(X - u(\tilde{V}))^\tau \beta\}, \\
&J(Y_k, \tilde{X}_k, T_k, \tilde{x}, t) \\
&\quad = \frac{\sqrt{n+N}}{\sqrt{N}} J_1(Y_k, \tilde{X}_k, T_k; \tilde{x}, t) I(k > n) \\
&\qquad - \frac{\sqrt{N(n+N)}}{n} J_2(\tilde{X}_k, X_k, T_k; \tilde{x}, t) I(k \le n),
\end{aligned}$$

where $F(\tilde{X}|T)$ is the conditional distribution of $\tilde{X}$ given T.

Theorem 1 *Under H_0 and the conditions in "Appendix", we have that*

$$\begin{aligned}
T_N(\tilde{x}, t) &= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} J_1(Y_j, \tilde{X}_j, T_j; \tilde{x}, t) - \frac{\sqrt{N}}{n} \sum_{i=1}^{n} J_2(\tilde{X}_i, X_i, T_i; \tilde{x}, t) \\
&= \frac{1}{\sqrt{n+N}} \sum_{k=1}^{n+N} \Bigg[\frac{\sqrt{n+N}}{\sqrt{N}} J_1(Y_k, \tilde{X}_k, T_k; \tilde{x}, t) I(k > n) \\
&\qquad - \frac{\sqrt{N(n+N)}}{n} J_2(\tilde{X}_k, X_k, T_k; \tilde{x}, t) I(k \le n) \Bigg] + o_p(1) \\
&= \frac{1}{\sqrt{n+N}} \sum_{k=1}^{n+N} J(Y_k, \tilde{X}_k, T_k, \tilde{x}, t) + o_p(1)
\end{aligned}$$

converges in distribution to $T(\tilde{x}, t)$ as N goes to infinity in the Skorokhod space $D[-\infty, +\infty]^{p+1}$, where $T(\tilde{x}, t)$ is a centered continuous Gaussian process with the covariance function

$$E(T(\tilde{x}_1, t_1)T(\tilde{x}_2, t_2)) = E(J(Y, \tilde{X}, T, \tilde{x}_1, t_1)J(Y, \tilde{X}, T, \tilde{x}_2, t_2)).$$

Therefore, CV_N *converges in distribution to* $CV := \int T(\tilde{X}, T)^2 dF(\tilde{X}, T)$ *with* $F(\cdot, \cdot)$ *being the distribution function of* $(\tilde{X}, T)$*.*

To study the power performance of the test, consider the following sequence of alternative models:

$$H_{1n}: \; Y = X^\tau \beta + g(T) + C_N G(X, T) + \varepsilon, \tag{10}$$

where $E(\varepsilon|X, T) = 0$ and the function $G(\cdot)$ satisfies $E(G^2(X, T)) < \infty$. Let $\tilde{G}(\tilde{V}) = E(G(X, T)|\tilde{V})$, model (10) can be rewritten as

$$\begin{cases} Y = u^\tau(\tilde{V})\beta + g(T) + C_N\tilde{G}(\tilde{V}) + \zeta, \\ \zeta = \varepsilon + X^\tau\beta - u^\tau(\tilde{V})\beta + C_N G(X, T) - C_N\tilde{G}(\tilde{V}), \\ u(\tilde{V}) = E(X|\tilde{V}). \end{cases} \tag{11}$$

The following theorem shows how sensitive the test is against the local alternatives.

Theorem 2 *Assume that the conditions of Theorem 1 hold. Then under* H_{1n} *of* (10)*, if* $C_N\sqrt{N} \to 1$*,* $T_N(\tilde{x}, t)$ *converges in distribution to* $T(\tilde{x}, t) + G_*(\tilde{x}, t)$*, where*

$$\begin{aligned} G_*(\tilde{x}, t) = -E[u^\tau(\tilde{V})I(\tilde{X} \le \tilde{x}, T \le t)]\Sigma^{-1}\Big(&E\{\tilde{G}(\tilde{V})(X - u(\tilde{V})]^\tau\beta\} \\ &+E\{\tilde{G}(\tilde{V})[u(\tilde{V}) - g_2(T)]\}\Big) + E\{\tilde{G}(\tilde{V})I(\tilde{X}_j \le \tilde{x}, T_j \le t)\} \end{aligned}$$

is a non-random shift function and CV_N *converges in distribution to* $\int (CV(\tilde{x}, t) + G_*(\tilde{x}, t))^2 dF(\tilde{x}, t)$*. If* $C_N N^r \to a$ *with* $-1/2 < r < 0$ *and* $a \neq 0$*,* $T_N(\tilde{x}, t)$ *converge to infinity in probability.*

Theorem 2 means that the test can distinct the local alternatives from the null hypothesis at the rate N^{-r} with $0 < r \le 1/2$. This rate is the possible fastest rate in goodness-of-fit testing.

3 Monte Carlo approximation for null limiting distribution

Theorem 1 shows that the limiting null distribution is intractable. The assistance from Monte Carlo approximation is often helpful, and the nonparametric Monte Carlo test (NMCT) procedure in Zhu (2005) is promising. This method has been successfully applied to model checking such as Zhu and Ng (2003). An important feature is that NMCT is a self-scale invariant procedure, and thus standardization of test statistic is not necessary and then the variance need not to be estimated.

The following three steps present the procedure for determining p values.

(I) Denote $e_k (k = 1, 2, \ldots, n + N)$ as independent variables with mean zero and variance one, and let $E_{n+N} := (e_1, \ldots, e_{n+N})$. The conditional counterpart of $T_N(\tilde{x}, t)$ is defined as

$$\tilde{T}_N(E_N; \tilde{x}, t) = \frac{1}{\sqrt{n+N}} \sum_{k=1}^{n+N} e_k \hat{J}(Y_k, \tilde{X}_k, T_k, \tilde{x}, t)$$

where $\hat{J}(Y_k, \tilde{X}_k, T_k, \tilde{x}, t)$ is the estimator of $J(Y_k, \tilde{X}_k, T_k, \tilde{x}, t)$, which is defined as

$$\begin{aligned}\hat{J}(Y_k, \tilde{X}_k, T_k, \tilde{x}, t) = {} & \frac{\sqrt{n+N}}{\sqrt{N}} \hat{J}_1(Y_k, \tilde{X}_k, T_k; \tilde{x}, t) I(k > n) \\ & - \frac{\sqrt{N(n+N)}}{n} \hat{J}_2(\tilde{X}_k, X_k, T_k; \tilde{x}, t) I(k \le n),\end{aligned}$$

where

$$\begin{aligned}& \hat{J}_1(Y, \tilde{X}, T; \tilde{x}, t) \\ & = (Y - \hat{u}^\tau(\tilde{V})\hat{\beta} - \hat{g}(T)) I(\tilde{X} \le \tilde{x}, T \le t) - (Y - \hat{g}_1(T)) \hat{F}[\tilde{X}|T] I(T \le t) \\ & \quad - \hat{U}_I \hat{\Sigma}^{-1} (\hat{u}(\tilde{V}) - \hat{g}_2(T)) [\hat{u}(\tilde{V}) - \hat{g}_2(T)]^\tau \hat{\beta}, \hat{J}_2(\tilde{X}, X, T; \tilde{x}, t) \\ & = (X - \hat{u}(\tilde{V}))^\tau \hat{\beta} I(\tilde{X} \le \tilde{x}, T \le t) - \hat{U}_I \hat{\Sigma}^{-1} \\ & \quad \times (\hat{u}(\tilde{V}_i) - \hat{g}_2(T_i)) \{(X_i - \hat{u}(\tilde{V}_i))^\tau \hat{\beta}\},\end{aligned}$$

and $\hat{U}_I$, $\hat{\Sigma}$ and $\hat{F}[\tilde{X}|T]$ are the estimators of U_I, Σ and $F[\tilde{X}|T]$, respectively. That is,

$$\begin{aligned}\hat{U}_I(\tilde{x}, t) &= \frac{1}{N} \sum_{k=n+1}^{n+N} \hat{u}^\tau(\tilde{V}_k) I(\tilde{X}_k \le \tilde{x}, T_k \le t), \\ \hat{\Sigma} &= \frac{1}{N} \sum_{k=n+1}^{n+N} (\hat{u}(\tilde{V}_k) - \hat{g}_2(T_k))(\hat{u}(\tilde{V}_k) - \hat{g}_2(T_k))^\tau, \\ \hat{F}[\tilde{X}|T] &= \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} W_k(T) I(\tilde{X}_k \le \tilde{X}).\end{aligned}$$

The resultant conditional test statistic is

$$\widetilde{CV}_N(E_n) = \int \tilde{T}_N(E_N; \tilde{x}, t)^2 dF_N(\tilde{x}, t).$$

(II) Generate m sets of E_{n+N}, say $E_{n+N}^{(i)}, i = 1, \ldots, m$, and get the corresponding m values of $\widetilde{CV}_N(E_{n+N})$. Without loss of generality, denote them as $\widetilde{CV}_N(E_{n+N}^{(i)}), i = 1, \ldots, m$.

(III) The p value is estimated by $\hat{p} = n_k/(m+1)$. Here $n_k = \sum_{i=1}^{m} I_{\{\widetilde{CV}_N(E_{n+N}^{(i)}) \geq CV_N\}}$ with $I_{\{\widetilde{CV}_N(E_{n+N}^{(i)}) \geq CV_N\}}$ being an indicator function. Reject H_0 when $\hat{p}_k \leq \alpha$ for a pre-specified level α.

For the consistency of this NMCT, we can refer to Zhu (2005) for details.

The following Theorem 3 indicates that the conditional distribution based on the Monte Carlo algorithm can avoid this trouble at least under local alternatives.

Theorem 3 *Suppose that the conditions in Theorem 1 hold. Under H_0 and H_{1n} with $C_N \to 0$, for almost all sequences*

$$\{(Y_{n+1}, \tilde{X}_{n+1}, T_{n+1}), \ldots, (Y_{n+N}, \tilde{X}_{n+N}, T_{n+N}), \ldots\},$$

the conditional distribution of $\tilde{CV}_N(E_{n+N})$ converges to the limiting null distribution CV in Theorem 1.

This theorem indicates that the NMCT makes the approximate distribution close to the null distribution, even under the local alternatives. This is an desirable feature in goodness-of-fit testing.

4 Numerical analysis

To examine the performance of the test, we conduct several simulation studies. Let $K(t) = (15/16)(1-t^2)^2 I(t^2 \leq 1)$, $K_1(\tilde{v}) = K(t)\prod_{i=1}^{d} K(\tilde{x}_i)$ and $K_2(t) = K(t)$ as the kernel functions. See Härdle and Mammen (1993), and Zhu and Ng (2003) for the use of these kernels. Note that selecting an optimal bandwidth in hypothesis testing is still an open problem and is beyond the scope of this paper. In this section, we choose the bandwidth as $h_N = \hat{\sigma}(T)N^{-1/3}$ with $\hat{\sigma}(T)$ being the empirical estimator of the standard deviation of variable T, which satisfies condition C.5 in Appendix. Similarly, the bandwidth b_n is chosen to be $b_n = \hat{\sigma}(\tilde{V})n^{-1/3}$.

To examine the sensitivity of the bandwidth selection, we consider several values of the bandwidth: $h_{N0} = h_N = \hat{\sigma}(T)N^{-1/3}$, $h_{N1} = 0.5\hat{\sigma}(T)N^{-1/3}$ and $h_{N2} = 2\hat{\sigma}(T)N^{-1/3}$ for fixed $b_n = \hat{\sigma}(\tilde{V})n^{-1/3}$.

Study 1. Consider the model

$$\begin{cases} Y = X\beta + (T^2 - 1/3) + aX^2 + e \\ \tilde{X} = X + \zeta, \end{cases}$$

where X, T, e and ζ are independent respectively from the standard normal distribution, the uniform distribution on [0, 1], the standard normal distribution and a normal distribution with mean zero and standard deviation 0.50. The value of parameter $\beta = 1$ in the simulation. The null hypothesis holds if and only if $a = 0.0$.

Table 1 Simulated size and power with the sample size $N = 100$ and $n = 100$, and different a in Study 1 for the model $Y = X\beta + (T^2 - 1/3) + aX^2 + e$, $\tilde{X} = X + \zeta$

a	h_{N1}	h_{N0}	h_{N2}
0.000	0.058	0.054	0.052
0.100	0.078	0.064	0.076
0.200	0.086	0.088	0.084
0.300	0.174	0.160	0.168
0.400	0.250	0.230	0.238
0.500	0.368	0.344	0.384
0.600	0.508	0.462	0.478
0.700	0.636	0.658	0.628
0.800	0.760	0.716	0.738
0.900	0.810	0.794	0.798
1.000	0.884	0.878	0.856

Each simulation experiment is repeated 1,000 times, and to determine critical values, the NMCT procedure is repeated 1,000 times. In general, the results show that the test is not very sensitive to different bandwidth. For space saving, we only present the simulated results with $n = 100$ and $N = 100$ in Table 1. From it, we can see that the power with smaller bandwidth h_{N1} is slightly higher than that with other bandwidths, but not significantly.

The size and power of the test are simulated with $a = 0.0, 0.1, 0.2, \ldots, 1.0$, and the sample sizes $N = 100, 200, 300$ and $n = 100, 150$. Also, each simulation experiment is repeated 1,000 times, and to determine critical values, the NMCT procedure is repeated 1,000 times. $h_{N0} = h_N = \hat{\sigma}(T)N^{-1/3}$ is used. The simulation results are summarized in Table 2. We can see that the test can maintain the significance level well in general. For power performance, the larger the value of a is, the more the power. This is reasonable. For fixed n and a, the power is higher under larger sample size N. For example, in the case $n = 50$ and $a = 0.5$, the values of the power are respectively 0.206, 0.400 and 0.520 under $N = 100$, $N = 200$ and $N = 300$. Similarly, for fixed N and a, the power performance is also better under larger validation data sample size n.

To examine the power performance when X and T are correlated, we further consider the above setting with the Pearson correlation coefficient 0.50 between X and T. The results are reported in Table 3. The results suggest that the trend is very similar as that in Table 2. It is reasonable that the larger the parameter a is, the more powerful the test is. Also large sample size leads to higher power. Compared with Table 2, we can see that when N is small, the correlation does not seem to significantly affect the power performance, whereas it does when N is large, say, $N = 300$.

Study 2. Consider a high-frequency model in the nonparametric component:

$$\begin{cases} Y = X\beta + \sin(2\pi T) + aX^2 + e \\ \tilde{X} = X + \zeta, \end{cases}$$

Table 2 Simulated size and power with the sample size $n = 50$ and $n = 100$, and different a in Study 1 for the model $Y = X\beta + (T^2 - 1/3) + aX^2 + e$, $\tilde{X} = X + \zeta$

a	$n = 50$			$n = 100$		
	$N = 100$	$N = 200$	$N = 300$	$N = 100$	$N = 200$	$N = 300$
0.000	0.056	0.058	0.059	0.054	0.056	0.056
0.100	0.068	0.076	0.116	0.064	0.070	0.088
0.200	0.084	0.100	0.124	0.088	0.152	0.198
0.300	0.118	0.142	0.254	0.160	0.296	0.362
0.400	0.138	0.248	0.366	0.230	0.456	0.636
0.500	0.206	0.400	0.520	0.344	0.702	0.838
0.600	0.296	0.494	0.688	0.462	0.836	0.968
0.700	0.354	0.642	0.808	0.658	0.940	0.986
0.800	0.480	0.800	0.892	0.716	0.974	0.992
0.900	0.556	0.830	0.956	0.794	0.988	1.000
1.000	0.618	0.926	0.982	0.878	0.996	1.000

Table 3 Simulated size and power when X and T are correlated in Study 1

a	$n = 50$			$n = 100$		
	$N = 100$	$N = 200$	$N = 300$	$N = 100$	$N = 200$	$N = 300$
0.000	0.058	0.052	0.047	0.058	0.052	0.046
0.100	0.081	0.062	0.064	0.088	0.073	0.060
0.200	0.086	0.073	0.070	0.120	0.128	0.133
0.300	0.097	0.130	0.132	0.163	0.209	0.265
0.400	0.116	0.202	0.219	0.256	0.376	0.546
0.500	0.231	0.306	0.352	0.357	0.578	0.777
0.600	0.254	0.448	0.574	0.440	0.801	0.913
0.700	0.370	0.608	0.771	0.601	0.886	0.990
0.800	0.439	0.752	0.908	0.688	0.960	0.996
0.900	0.586	0.876	0.953	0.796	0.997	1.000
1.000	0.642	0.916	0.990	0.883	0.994	1.000

where X, T, e and ζ are respectively from the t distribution with freedom 4, the standard normal, standard normal and normal distribution with mean zero and standard deviation 0.50. The value of parameter is also $\beta = 1$. The null hypothesis holds if and only if $a = 0.0$.

The simulation results are summarized in Table 4. The conclusions may be very similar to those obtained from Table 2 of Study 1. We then do not repeat the relevant comments.

For readers who are interested in our method, the Matlab codes will be provided upon request.

Table 4 Simulated size and power with the sample size $n = 50$ and $n = 100$, and different a in Study 2 for the model $Y = X\beta + \sin(2\pi T) + aX^2 + e$, $\tilde{X} = X + \zeta$

a	$n = 50$			$n = 100$		
	$N = 100$	$N = 200$	$N = 300$	$N = 100$	$N = 200$	$N = 300$
0.000	0.038	0.058	0.058	0.036	0.052	0.054
0.100	0.060	0.096	0.160	0.064	0.070	0.140
0.200	0.100	0.142	0.210	0.102	0.200	0.268
0.300	0.150	0.264	0.368	0.174	0.304	0.518
0.400	0.174	0.380	0.550	0.240	0.474	0.784
0.500	0.214	0.464	0.672	0.300	0.714	0.882
0.600	0.292	0.556	0.788	0.438	0.800	0.922
0.700	0.316	0.662	0.816	0.488	0.858	0.952
0.800	0.380	0.752	0.888	0.566	0.862	0.940
0.900	0.416	0.766	0.882	0.616	0.902	0.946
1.000	0.458	0.768	0.890	0.620	0.906	0.950

5 Appendix: Proofs of the theorems

The following conditions are assumed.

C.1 $\Sigma = E\{(u(\tilde{V}) - g_2(T))(u(\tilde{V}) - g_2(T))^\tau\}$ is a positive definite matrix;

C.2 $g_1(t), g_{2r}(t)$ (the rth component of $g_2(t)$ for $r = 1, \ldots, d$), $g(t)$ and $u(\tilde{V})$ satisfy the Lipschitz condition of order one;

C.3 The density of T, say $r(t)$, exists and satisfies

$$0 < \inf_{0 \le t \le 1} r(t) \le \sup_{0 \le t \le 1} r(t) < \infty.$$

C.4 There exists a constant $C > 0$ such that $N/n \le C$.

C.5 As $N \to \infty$, $\sqrt{N}h_N \to \infty$, $\sqrt{N}h_N^2 \to 0$. As $n \to \infty$, $nb_n^{2(d+1)} \to \infty$ and $nb_n^{2k} \to 0$ for $k > d + 1$.

C.6 (i) The density of $\tilde{V}$, say $f_{\tilde{V}}(\tilde{v})$, exists and satisfies

$$\sum_{N=1}^{\infty} NP[f_{\tilde{V}}(\tilde{V}) \le \eta_N] < \infty$$

for a positive constants sequence $\eta_N > 0$. (ii) $f_{\tilde{V}}(\tilde{v})$ has bounded partial derivative of order one.

C.7 $\sup_{\tilde{v}} E[Y^2|\tilde{V} = \tilde{v}] < \infty$ and $E(\|X\|^2) < \infty$.

C.8 The kernel functions $K_1(\cdot)$ and $K_2(\cdot)$ are bounded with bounded support, and $K_1(\cdot)$ and $K_2(\cdot)$ are kernels of order $k (k \ge 2)$.

Remark 2 Conditions C.1 and C.7 are the necessary conditions for consistency of relevant parametric estimators. Conditions C.2, C.8 and C.5 are typically needed for the asymptotic property of the nonparametric estimators. Condition C.3 avoids the cumbersome proofs for the theorems. Without it, a truncation technique is needed because certain denominators in the estimator are close to zero.

Lemma 1 *Under H_0 and H_{1n} with all the above conditions, we have*

$$\hat{\Sigma} \to \Sigma \ a.s. \tag{12}$$

Proof of Lemma 1 Under H_0, Wang (1999) proved the relevant asymptotic property of $\hat{\Sigma}$ in (12). We now prove that in (12) under H_{1n}. Note that

$$\hat{\Sigma} = \frac{1}{N} \sum_{j=n+1}^{n+N} (\hat{u}(\tilde{V}_j) - \hat{g}_2(T_j))(\hat{u}(\tilde{V}_j) - \hat{g}_2(T_j))^{\tau},$$

with

$$\hat{u}(\tilde{V}) = \frac{\sum_{i=1}^{n} X_i K_1((\tilde{V}_i - \tilde{V})/b_n)}{\sum_{i=1}^{n} K_1((\tilde{V}_i - \tilde{V})/b_n)}, \quad \hat{g}_2(T) = \sum_{j=n+1}^{n+N} W_j(T)\hat{u}(\tilde{V}_j),$$

we know that $\hat{\Sigma}$ depends only on the covariates, but not the response. Hence, the limiting behavior of $\hat{\Sigma}$ under H_{1n} is the same as that under H_0. The proof is finished. □

From the following lemma, we can also get the asymptotic representation of $\sqrt{N}(\hat{\beta} - \beta)$ under H_0 when $C_N = 0$ although it is about the asymptotic result under H_{1n}.

Lemma 2 *Under H_{1n} and the above conditions, we have*

$$\begin{aligned}
&\sqrt{N}(\hat{\beta} - \beta) \\
&= \frac{1}{\sqrt{N}} \Sigma^{-1} \sum_{j=n+1}^{n+N} [u(\tilde{V}_j) - g_2(T_j)][C_N \tilde{G}(\tilde{V}_j) + \zeta_j - (Y_j - g_1(\tilde{T}_j))] \\
&\quad - \frac{\sqrt{N}}{n} \Sigma^{-1} \sum_{i=1}^{n} [u(\tilde{V}_i) - g_2(T_i) - C_N \tilde{G}(\tilde{V}_i)](X_i - u(\tilde{V}_i))^{\tau} \beta \\
&\quad + \frac{1}{\sqrt{N}} \left(\hat{\Sigma}^{-1} - \Sigma^{-1}\right) \sum_{j=n+1}^{n+N} C_N (u(\tilde{V}_j) - g_2(T_j)) \tilde{G}(\tilde{V}_j) \\
&\quad + \frac{1}{\sqrt{N}} \left(\hat{\Sigma}^{-1} - \Sigma^{-1}\right) \sum_{i=1}^{n} C_N \tilde{G}(\tilde{V}_i)(X_i - u(\tilde{V}_i))^{\tau} \beta + o_p(1).
\end{aligned} \tag{13}$$

Proof of Lemma 2. Denote $\sqrt{N}(\hat{\beta}-\beta)=\hat{\Sigma}^{-1}A$ and $\tilde{G}(\tilde{V}_j)=E(G(X,T)|\tilde{V}_j)$. We have

$$\begin{aligned}
A &= \frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}(\hat{u}(\tilde{V}_j)-\hat{g}_2(T_j))(Y_j-\hat{g}_1(T_j)-[\hat{u}(\tilde{V}_j)-\hat{g}_2(T_j)]^{\tau}\beta)\\
&= \frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}(u(\tilde{V}_j)-g_2(T_j)+\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)+g_2(T_j)-\hat{g}_2(T_j))\\
&\quad\times\Big(C_N\tilde{G}(\tilde{V}_j)+\zeta_j+g_1(T_j)-\hat{g}_1(T_j)-[\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)\\
&\quad+g_2(T_j)-\hat{g}_2(T_j)]^{\tau}\beta\Big),
\end{aligned}$$

which can be further decomposed as

$$\begin{aligned}
A &= \frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}\big([u(\tilde{V}_j)-g_2(T_j)][C_N\tilde{G}(\tilde{V}_j)+\zeta_j]+[\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)]\\
&\quad\times[C_N\tilde{G}(\tilde{V}_j)+\zeta_j]\\
&\quad+[g_2(T_j)-\hat{g}_2(T_j)][C_N\tilde{G}(\tilde{V}_j)+\zeta_j]+[u(\tilde{V}_j)-g_2(T_j)][g_1(T_j)-\hat{g}_1(T_j)]\\
&\quad+[\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)][g_1(T_j)-\hat{g}_1(T_j)]+[g_2(T_j)-\hat{g}_2(T_j)][g_1(T_j)-\hat{g}_1(T_j)]\\
&\quad-[u(\tilde{V}_j)-g_2(T_j)][\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)]^{\tau}\beta-[\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)][\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)]^{\tau}\beta\\
&\quad-[g_2(T_j)-\hat{g}_2(T_j)][\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)]^{\tau}\beta+[u(\tilde{V}_j)-g_2(T_j)][\hat{g}_2(T_j)\\
&\quad-g_2(T_j)]^{\tau}\beta\\
&\quad+[\hat{u}(\tilde{V}_j)-u(\tilde{V}_j)][\hat{g}_2(T_j)-g_2(T_j)]^{\tau}\beta+[g_2(T_j)-\hat{g}_2(T_j)][\hat{g}_2(T_j)\\
&\quad-g_2(T_j)]^{\tau}\beta\big)\\
&=: A_1+A_2+A_3+A_4+A_5+A_6-A_7-A_8-A_9+A_{10}+A_{11}+A_{12}. \qquad (14)
\end{aligned}$$

For the term A_2 in (14), we note that

$$\begin{aligned}
\hat{u}(\tilde{V}_k)-u(\tilde{V}_k) &= \frac{\hat{l}(\tilde{V}_k)}{\hat{f}_{\tilde{V}}(\tilde{V}_k)}-\frac{l(\tilde{V}_k)}{f_{\tilde{V}}(\tilde{V}_k)}\\
&= \frac{\hat{l}(\tilde{V}_k)-u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k)}{f_{\tilde{V}}(\tilde{V}_k)}-\frac{(\hat{l}(\tilde{V}_k)-u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k))(\hat{f}_{\tilde{V}}(\tilde{V}_k)-f_{\tilde{V}}(\tilde{V}_k))}{f_{\tilde{V}}(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k)},
\end{aligned}$$

where

$$\hat{l}(\tilde{v})=\frac{1}{nb_n^{d+1}}\sum_{i=1}^{n}X_iK_1\Big(\frac{\tilde{V}_i-\tilde{v}}{b_n}\Big),\quad \hat{f}_{\tilde{V}}(\tilde{v})=\frac{1}{nb_n^{d+1}}\sum_{i=1}^{n}K_1\Big(\frac{\tilde{V}_i-\tilde{v}}{b_n}\Big).$$

Then

$$
\begin{aligned}
A_2 &= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \frac{\hat{l}(\tilde{V}_j) - u(\tilde{V}_j)\hat{f}_{\tilde{V}}(\tilde{V}_j)}{f_{\tilde{V}}(\tilde{V}_j)} [C_N \tilde{G}(\tilde{V}_j) + \zeta_j] + o_p(1) \\
&= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \frac{1}{n b_n^{d+1}} \sum_{i=1}^{n} \frac{(X_i - u(\tilde{V}_j)) K_1\left(\frac{\tilde{V}_i - \tilde{V}_j}{b_n}\right)}{f_{\tilde{V}}(\tilde{V}_j)} [C_N \tilde{G}(\tilde{V}_j) + \zeta_j] + o_p(1) \\
&= \frac{\sqrt{N}}{n} \sum_{i=1}^{n} \sum_{j=n+1}^{n+N} \frac{1}{N b_n^{d+1}} \frac{(X_i - u(\tilde{V}_j)) K_1\left(\frac{\tilde{V}_i - \tilde{V}_j}{b_n}\right)}{f_{\tilde{V}}(\tilde{V}_j)} [C_N \tilde{G}(\tilde{V}_j) + \zeta_j] + o_p(1) \\
&= \frac{\sqrt{N}}{n} \sum_{i=1}^{n} C_N \tilde{G}(\tilde{V}_i)(X_i - u(\tilde{V}_i))^{\tau} \beta + o_p(1). \qquad (15)
\end{aligned}
$$

Deal with the term A_3. It can be calculated that

$$
\begin{aligned}
A_3 &= -\frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} [\hat{g}_2(T_j) - g_2(T_j)][C_N \tilde{G}(\tilde{V}_j) + \zeta_j] \\
&= -\frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \left[\sum_{k=n+1}^{n+N} W_k(T_j)\{\hat{u}(\tilde{V}_k) - u(\tilde{V}_k)\} \right] [C_N \tilde{G}(\tilde{V}_j) + \zeta_j] \\
&\quad -\frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \left[\sum_{k=n+1}^{n+N} W_k(T_j) u(\tilde{V}_k) - g_2(T_j) \right] [C_N \tilde{G}(\tilde{V}_j) + \zeta_j] \\
&= -\frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \sum_{k=n+1}^{n+N} W_k(T_j) \frac{\hat{l}(\tilde{V}_k) - u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k)}{\hat{f}_{\tilde{V}}(\tilde{V}_k)} [C_N \tilde{G}(\tilde{V}_j) + \zeta_j] + o_p(1). \\
&= A_{3,1} + A_{3,2} + A_{3,3} + A_{3,4} + o_p(1),
\end{aligned}
$$

where the definitions of the terms $A_{3,i} (i = 1, 2, 3, 4)$ are as follows

$$
\begin{aligned}
A_{3,1} &= -\frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \sum_{k=n+1}^{n+N} W_k(T_j) \frac{(1/n b_n^{d+1}) \sum_{i=1}^{n} (X_i - u(\tilde{V}_i)) K_1((\tilde{V}_i - \tilde{V}_k)/b_n)}{\hat{f}_{\tilde{V}}(\tilde{V}_k)} \\
&\qquad \times [C_N \tilde{G}(\tilde{V}_j) + \zeta_j] I[\hat{f}_{\tilde{V}}(\tilde{V}_k) > \frac{1}{2} f_{\tilde{V}}(\tilde{V}_k) \geq \frac{1}{2}\eta'_N], \\
A_{3,2} &= -\frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \sum_{k=n+1}^{n+N} W_k(T_j) \\
&\qquad \times \frac{(1/n b_n^{d+1}) \sum_{i=1}^{n} (u(\tilde{V}_i) - u(\tilde{V}_k)) K_1((\tilde{V}_i - \tilde{V}_k)/b_n)}{\hat{f}_{\tilde{V}}(\tilde{V}_k)}
\end{aligned}
$$

$$\times[C_N\tilde{G}(\tilde{V}_j)+\zeta_j]I[\hat{f}_{\tilde{V}}(\tilde{V}_k)>\frac{1}{2}f_{\tilde{V}}(\tilde{V}_k)\geq\frac{1}{2}\eta'_N],$$

$$A_{3,3}=-\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}\sum_{k=n+1}^{n+N}W_k(T_j)\frac{\hat{l}(\tilde{V}_k)-u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k)}{\hat{f}_{\tilde{V}}(\tilde{V}_k)}[C_N\tilde{G}(\tilde{V}_j)+\zeta_j]$$
$$\times I[\hat{f}_{\tilde{V}}(\tilde{V}_k)<\frac{1}{2}f_{\tilde{V}}(\tilde{V}_k),\,f_{\tilde{V}}(\tilde{V}_k)\geq\eta'_N],$$

$$A_{3,4}=-\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}\sum_{k=n+1}^{n+N}W_k(T_j)\frac{\hat{l}(\tilde{V}_k)-u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k)}{\hat{f}_{\tilde{V}}(\tilde{V}_k)}[C_N\tilde{G}(\tilde{V}_j)+\zeta_j]$$
$$\times I[f_{\tilde{V}}(\tilde{V}_k)\leq\eta'_N].$$

Following the derivations for $D^{[r]}_{N11}$, $D^{[r]}_{N12}$, $D^{[r]}_{N13}$ and $D^{[r]}_{N14}$ on Page 58 of Wang (1999), we can obtain that $A_{3,1}=o_p(1)$, $A_{3,2}=o_p(1)$, $A_{3,3}=o_p(1)$ and $A_{3,4}=o_p(1)$, respectively. Consequently,

$$A_3=o_p(1). \tag{16}$$

Denote

$$\hat{q}(t)=\frac{1}{Nh_N}\sum_{k=n+1}^{n+N}Y_kK_2\Big(\frac{T_k-t}{h_N}\Big),\quad \hat{f}_t(t)=\frac{1}{Nh_N}\sum_{k=n+1}^{n+N}K_2\Big(\frac{T_k-t}{h_N}\Big).$$

For A_4 in (14), it can be decomposed as

$$\begin{aligned}A_4&=-\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}[u(\tilde{V}_j)-g_2(T_j)]\frac{(\hat{q}(T_j)-g_1(T_j)\hat{f}_t(T_j))}{f_t(T_j)}+o_p(1)\\&=-\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}[u(\tilde{V}_j)-g_2(T_j)]\frac{1}{Nh_N}\sum_{k=n+1}^{n+N}\frac{(Y_k-g_1(T_j))K_2\Big(\frac{T_k-t}{h_N}\Big)}{f_t(T_j)}+o_p(1)\\&=-\frac{1}{\sqrt{N}}\sum_{k=n+1}^{n+N}[u(\tilde{V}_k)-g_2(T_k)](Y_k-g_1(T_k))+o_p(1).\end{aligned} \tag{17}$$

Consider A_7 in (14). We have

$$\begin{aligned}
A_7 &= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} [u(\tilde{V}_j) - g_2(T_j)][\hat{u}(\tilde{V}_j) - u(\tilde{V}_j)]^\tau \beta \\
&= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} [u(\tilde{V}_j) - g_2(T_j)] \left(\frac{\hat{l}(\tilde{V}_j) - u(\tilde{V}_j) \hat{f}_{\tilde{V}}(\tilde{V}_j)}{f_{\tilde{V}}(\tilde{V}_j)} \right)^\tau \beta + o_p(1) \\
&= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \frac{1}{n b_n^{d+1}} \sum_{i=1}^{n} [u(\tilde{V}_j) - g_2(T_j)] \left(\frac{(X_i - u(\tilde{V}_j)) K_1\left(\frac{\tilde{V}_i - \tilde{V}_j}{b_n}\right)}{f_{\tilde{V}}(\tilde{V}_j)} \right)^{\tau\beta + o_p(1)} \\
&= \frac{\sqrt{N}}{n} \sum_{i=1}^{n} [u(\tilde{V}_i) - g_2(T_i)](X_i - u(\tilde{V}_i))^\tau \beta. \qquad (18)
\end{aligned}$$

We can similarly prove that all the other terms in (14) are equal to $o_p(1)$:

$$A_i = o_p(1), \quad i = 5, 6, 8, 9, \ldots, 12. \qquad (19)$$

From (15), (16), (17), (18) and (19), we obtain

$$\begin{aligned}
A = \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} & [u(\tilde{V}_j) - g_2(T_j)][C_N \tilde{G}(\tilde{V}_j) + \zeta_j - (Y_j - g_1(\tilde{T}_j))] \\
& - \frac{\sqrt{N}}{n} \sum_{i=1}^{n} [u(\tilde{V}_i) - g_2(T_i) - C_N \tilde{G}(\tilde{V}_i)](X_i - u(\tilde{V}_i))^\tau \beta + o_p(1).
\end{aligned}$$

Combining with Lemma 1 for the asymptotic property of $\hat{\Sigma}$, we have

$$\begin{aligned}
\sqrt{N}(\hat{\beta} - \beta) = \frac{1}{\sqrt{N}} \Sigma^{-1} & \sum_{j=n+1}^{n+N} [u(\tilde{V}_j) - g_2(T_j)][C_N \tilde{G}(\tilde{V}_j) + \zeta_j - (Y_j - g_1(\tilde{T}_j))] \\
& - \frac{\sqrt{N}}{n} \Sigma^{-1} \sum_{i=1}^{n} [u(\tilde{V}_i) - g_2(T_i) - C_N \tilde{G}(\tilde{V}_i)](X_i - u(\tilde{V}_i))^\tau \beta \\
& + \frac{1}{\sqrt{N}} (\hat{\Sigma}^{-1} - \Sigma^{-1}) \sum_{j=n+1}^{n+N} C_N (u(\tilde{V}_j) - g_2(T_j)) \tilde{G}(\tilde{V}_j) \\
& + \frac{1}{\sqrt{N}} (\hat{\Sigma}^{-1} - \Sigma^{-1}) \sum_{i=1}^{n} C_N \tilde{G}(\tilde{V}_i)(X_i - u(\tilde{V}_i))^\tau \beta + o_p(1).
\end{aligned}$$

The proof is concluded. □

Proof of Theorem 1 Under H_0, the test statistic in (6) can be decomposed as

$$\begin{aligned}
T_N(\tilde{x},t) &= \frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}\varepsilon_j I(\tilde{X}_j\le\tilde{x},T_j\le t)\\
&\quad -\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}(\hat{u}(\tilde{V}_j)-u(\tilde{V}_j))^\tau\beta I(\tilde{X}_j\le\tilde{x},T_j\le t)\\
&\quad -\left(\frac{1}{N}\sum_{j=n+1}^{n+N}u^\tau(\tilde{V}_j)I(\tilde{X}_j\le\tilde{x},T_j\le t)\right)\sqrt{N}(\hat{\beta}-\beta)\\
&\quad -\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}(\hat{g}(T_j)-g(T_j))I(\tilde{X}_j\le\tilde{x},T_j\le t)\\
&\quad -\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}(\hat{u}(\tilde{V}_j)-u(\tilde{V}_j))^\tau(\hat{\beta}-\beta)I(\tilde{X}_j\le\tilde{x},T_j\le t)\\
&=: T_{N,1}-T_{N,2}-T_{N,3}-T_{N,4}-T_{N,5}.
\end{aligned}\tag{20}$$

We have, for $T_{N,2}$ in (20),

$$\begin{aligned}
T_{N,2} &= \frac{1}{\sqrt{N}}\sum_{k=n+1}^{n+N}\frac{(\hat{l}(\tilde{V}_k)-u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k))^\tau\beta}{f_{\tilde{V}}(\tilde{V}_k)}I(\tilde{X}_k\le\tilde{x},T_k\le t)\\
&\quad -\frac{1}{\sqrt{N}}\sum_{k=n+1}^{n+N}\frac{(\hat{f}_{\tilde{V}}(\tilde{V}_k)-f_{\tilde{V}}(\tilde{V}_k))(\hat{l}(\tilde{V}_k)-u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k))^\tau\beta}{f_{\tilde{V}}(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k)}I(\tilde{X}_k\le\tilde{x},T_k\le t)\\
&= \frac{1}{\sqrt{N}}\sum_{k=n+1}^{n+N}\frac{(\hat{l}(\tilde{V}_k)-u(\tilde{V}_k)\hat{f}_{\tilde{V}}(\tilde{V}_k))^\tau\beta}{f_{\tilde{V}}(\tilde{V}_k)}I(\tilde{X}_k\le\tilde{x},T_k\le t)+o_p(1)\\
&= \frac{1}{\sqrt{N}}\sum_{k=n+1}^{n+N}\frac{1}{nb_n^{d+1}}\sum_{i=1}^{n}\frac{K_1((\tilde{V}_i-\tilde{V}_k)/b_n)(X_i-u(\tilde{V}_k))^\tau\beta}{f_{\tilde{V}}(\tilde{V}_k)}\\
&\quad \times I(\tilde{X}_k\le\tilde{x},T_k\le t)+o_p(1)\\
&= \frac{\sqrt{N}}{n}\sum_{i=1}^{n}(X_i-u(\tilde{V}_i))^\tau\beta I(\tilde{X}_i\le\tilde{x},T_i\le t)+o_p(1).
\end{aligned}\tag{21}$$

Recall the definition of $\hat{q}(t)$ and $\hat{f}_t(t)$ as follows,

$$\hat{q}(t)=\frac{1}{Nh_N}\sum_{j=n+1}^{n+N}Y_jK_2\Big(\frac{T_j-t}{h_N}\Big),\quad \hat{f}_t(t)=\frac{1}{Nh_N}\sum_{j=n+1}^{n+N}K_2\Big(\frac{T_j-t}{h_N}\Big).$$

For $T_{N,4}$ in (20), we have the following decomposition:

$$\begin{aligned}
T_{N,4} &= \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} (\hat{g}_1(T_k) - g_1(T_k)) I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&\quad - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} (\hat{g}_2(T_k) - g_2(T_k))^\tau \beta I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&= \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \frac{(\hat{q}(T_k) - g_1(T_k)\hat{f}_t(T_k))}{f_t(T_k)} I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&\quad - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \frac{(\hat{f}_t(T_k) - f_t(T_k))(\hat{q}(T_k) - g_1(T_k)\hat{f}_t(T_k))}{f_t(T_k)\hat{f}_t(T_k)} I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&\quad - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} (\hat{g}_2(T_k) - g_2(T_k))^\tau \beta I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&= \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \frac{(\hat{q}(T_k) - g_1(T_k)\hat{f}_t(T_k))}{f_t(T_k)} I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&\quad - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} (\hat{g}_2(T_k) - g_2(T_k))^\tau \beta I(\tilde{X}_j \le \tilde{x}, T_j \le t) + o_p(1) \\
&=: T_{N,41} - T_{N,42} + o_p(1).
\end{aligned}$$

For $T_{N,42}$ in (22), it can be proved that

$$\begin{aligned}
T_{N,42} &= -\frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_j(T_k)(E[u(\tilde{V}_k)|T_k] - E[u(\tilde{V}_j)|T_j])^\tau \beta I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&\quad - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_j(T_k)(E[u(\tilde{V}_j)|T_j] - u(\tilde{V}_j))^\tau \beta I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&\quad + \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_j(T_k)(\hat{u}(\tilde{V}_j) - u(\tilde{V}_j))^\tau \beta I(\tilde{X}_j \le \tilde{x}, T_j \le t) \\
&= o_p(1).
\end{aligned}$$

As a result, the term T_{N4} in (20) can be further derived as

$$\begin{aligned}
&T_{N,4} \\
&= \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \frac{(\hat{q}(T_k) - g_1(T_k)\hat{f}_t(T_k))}{f_t(T_k)} I(\tilde{X}_j \le \tilde{x}, T_j \le t) + o_p(1)
\end{aligned}$$

$$= \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \frac{1}{nb_n^{d+1}} \sum_{i=1}^{n} \frac{K_1((\tilde{V}_i - \tilde{V}_k)/b_n)(X_i - u(\tilde{V}_k))^\tau \beta}{f_{\tilde{V}}(\tilde{V}_k)}$$
$$\times I(\tilde{X}_k \le \tilde{x}, T_k \le t) + o_p(1)$$
$$= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} (Y_j - g_1(T_j)) E[I(\tilde{X} \le \tilde{x}) | T_j] I(T_j \le t) + o_p(1). \tag{22}$$

For $T_{N,5}$ in (20), we have

$$T_{N,5} = o_p(1). \tag{23}$$

Together with (20), (21), (22), (23) and Lemma 2 under H_0, it can be obtained that

$$T_N(\tilde{x}, t)$$
$$= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \varepsilon_j I(\tilde{X}_j \le \tilde{x}, T_j \le t) - \frac{\sqrt{N}}{n} \sum_{i=1}^{n} (X_i - u(\tilde{V}_i))^\tau \beta I(\tilde{X}_i \le \tilde{x}, T_i \le t)$$
$$- \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} (Y_j - g_1(T_j)) E[I(\tilde{X} \le \tilde{x}) | T_j] I(T_j \le t)$$
$$- E[u^\tau(\tilde{V}) I(\tilde{X} \le \tilde{x}, T \le t)] \sqrt{N}(\hat{\beta} - \beta) + o_p(1)$$
$$= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \varepsilon_j I(\tilde{X}_j \le \tilde{x}, T_j \le t) - \frac{\sqrt{N}}{n} \sum_{i=1}^{n} (X_i - u(\tilde{V}_i))^\tau \beta I(\tilde{X}_i \le \tilde{x}, T_i \le t)$$
$$- \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} (Y_j - g_1(T_j)) E[I(\tilde{X} \le \tilde{x}) | T_j] I(T_j \le t)$$
$$- E[u^\tau(\tilde{V}) I(\tilde{X} \le \tilde{x}, T \le t)] \Sigma^{-1} \Big(\frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} (u(\tilde{V}_j) - g_2(T_j)) \varepsilon_j$$
$$- \frac{\sqrt{N}}{n} \sum_{i=1}^{n} (u(\tilde{V}_i) - g_2(T_i))(X_i - u(\tilde{V}_i))^\tau \beta$$
$$- \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} (u(\tilde{V}_j) - g_2(T_j))(Y_j - g_1(\tilde{T}_j)) \Big) + o_p(1)$$
$$= \frac{1}{\sqrt{N}} \sum_{j=n+1}^{n+N} \Big(\varepsilon_j I(\tilde{X}_j \le \tilde{x}, T_j \le t) - (Y_j - g_1(T_j)) F[\tilde{X} | T_j] I(T_j \le t)$$
$$- E[u^\tau(\tilde{V}) I(\tilde{X} \le \tilde{x}, T \le t)] \Sigma^{-1} (u(\tilde{V}_j) - g_2(T_j)) [\varepsilon_j - (Y_j - g_1(\tilde{T}_j))] \Big)$$
$$- \frac{\sqrt{N}}{n} \sum_{i=1}^{n} \Big((X_i - u(\tilde{V}_i))^\tau \beta I(\tilde{X}_i \le \tilde{x}, T_i \le t) - E[u^\tau(\tilde{V}) I(\tilde{X} \le \tilde{x}, T \le t)] \Sigma^{-1}$$

$$
\begin{aligned}
&\times(u(\tilde V_i)-g_2(T_i))\{(X_i-u(\tilde V_i))^\tau\beta\}\Big)+o_p(1)\\
&=\frac{1}{\sqrt{n+N}}\sum_{k=1}^{n+N}J(Y_k,\tilde X_k,T_k,\tilde x,t)+o_p(1),
\end{aligned}
\tag{24}
$$

where $J(Y_k,\tilde X_k,T_k,\tilde x,t)$ is defined in Theorem 1. Hence $T_N(\tilde x,t)$ converges to $T(\tilde x,t)$ in distribution in the Skorokhod space $D[-\infty,+\infty]^{p+1}$. Here $T(\tilde x,t)$ is a centered continuous Gaussian process with the covariance function

$$
E(T(\tilde x_1,t_1)T(\tilde x_2,t_2))=E(J(Y,\tilde X,T,\tilde x_1,t_1)J(Y,\tilde X,T,\tilde x_2,t_2)).
$$

According to the continuous mapping theorem, CV_N converges in distribution to $CV:=\int T(\tilde X,T)^2dF(\tilde X,T)$. Hence, Theorem 1 is proved.

Proof of Theorem 2. Under the local alternatives H_{1n} in (11), similarly as the derivation under H_0, we have

$$
\begin{aligned}
T_N(\tilde x,t)=&\frac{1}{\sqrt N}\sum_{j=n+1}^{n+N}(\zeta_j+C_N\tilde G(\tilde V_j))I(\tilde X_j\le\tilde x,T_j\le t)\\
&-\frac{\sqrt N}{n}\sum_{i=1}^{n}(X_i-u(\tilde V_i))^\tau\beta I(\tilde X_i\le\tilde x,T_i\le t)\\
&-\frac{1}{\sqrt N}\sum_{j=n+1}^{n+N}(Y_j-g_1(T_j))E[I(\tilde X\le\tilde x)|T_j]I(T_j\le t)\\
&-E[u^\tau(\tilde V)I(\tilde X\le\tilde x,T\le t)]\sqrt N(\hat\beta-\beta)+o_p(1).
\end{aligned}
\tag{25}
$$

Note the asymptotic property of $\sqrt N(\hat\beta-\beta)$ in (13) under H_{1n}. We have

$$
\begin{aligned}
&T_N(\tilde x,t)\\
&=\frac{1}{\sqrt N}\sum_{j=n+1}^{n+N}(\zeta_j+C_N\tilde G(\tilde V_j))I(\tilde X_j\le\tilde x,T_j\le t)\\
&\quad-\frac{\sqrt N}{n}\sum_{i=1}^{n}(X_i-u(\tilde V_i))^\tau\beta I(\tilde X_i\le\tilde x,T_i\le t)\\
&\quad-\frac{1}{\sqrt N}\sum_{j=n+1}^{n+N}(Y_j-g_1(T_j))E[I(\tilde X\le\tilde x)|T_j]I(T_j\le t)\\
&\quad-E[u^\tau(\tilde V)I(\tilde X\le\tilde x,T\le t)]\\
&\quad\times\Big(\frac{1}{\sqrt N}\Sigma^{-1}\sum_{j=n+1}^{n+N}[u(\tilde V_j)-g_2(T_j)][C_N\tilde G(\tilde V_j)+\zeta_j-(Y_j-g_1(\tilde T_j))]
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\sqrt{N}}{n}\Sigma^{-1}\sum_{i=1}^{n}[u(\tilde{V}_i)-g_2(T_i)-C_N\tilde{G}(\tilde{V}_i)](X_i-u(\tilde{V}_i))^\tau\beta \\
&+\frac{1}{\sqrt{N}}(\hat{\Sigma}^{-1}-\Sigma^{-1})\sum_{j=n+1}^{n+N}C_N(u(\tilde{V}_j)-g_2(T_j))\tilde{G}(\tilde{V}_j) \\
&+\frac{1}{\sqrt{N}}(\hat{\Sigma}^{-1}-\Sigma^{-1})\sum_{i=1}^{n}C_N\tilde{G}(\tilde{V}_i)(X_i-u(\tilde{V}_i))^\tau\beta\Big)+o_p(1). \qquad (26)
\end{aligned}
$$

In the case that $C_N = N^{-1/2}$, the term $T_N(\tilde{x}, t)$ in (26) can be further calculated as follows:

$$
\begin{aligned}
&T_N(\tilde{x}, t) \\
&=\frac{1}{\sqrt{N}}\sum_{j=n+1}^{n+N}\Big(\zeta_j I(\tilde{X}_j\le\tilde{x}, T_j\le t)-(Y_j-g_1(T_j))F[\tilde{X}|T_j]I(T_j\le t) \\
&\quad -E[u^\tau(\tilde{V})I(\tilde{X}\le\tilde{x}, T\le t)]\Sigma^{-1}(u(\tilde{V}_j)-g_2(T_j))[\zeta_j-(Y_j-g_1(\tilde{T}_j))]\Big) \\
&\quad -\frac{\sqrt{N}}{n}\sum_{i=1}^{n}\Big((X_i-u(\tilde{V}_i))^\tau\beta I(\tilde{X}_i\le\tilde{x}, T_i\le t)-E[u^\tau(\tilde{V}) \\
&\quad \times I(\tilde{X}\le\tilde{x}, T\le t)]\Sigma^{-1}(u(\tilde{V}_i)-g_2(T_i))\{(X_i-u(\tilde{V}_i))^\tau\beta\}\Big) \\
&\quad -E[u^\tau(\tilde{V})I(\tilde{X}\le\tilde{x}, T\le t)]\Sigma^{-1}\Big(E\{\tilde{G}(\tilde{V})(X-u(\tilde{V}))^\tau\beta\} \\
&\quad +E\{\tilde{G}(\tilde{V})[u(\tilde{V})-g_2(T)]\}\Big)+E\{\tilde{G}(\tilde{V})I(\tilde{X}_j\le\tilde{x}, T_j\le t)\}+o_p(1) \\
&=T(\tilde{x}, t)+G_*(\tilde{x}, t)+o_p(1). \qquad (27)
\end{aligned}
$$

Here, the definition of $G_*(\tilde{x}, t)$ is in Theorem 2. As a result, when $C_N = N^{-1/2}$, $T_N(\tilde{x}, t)$ converges to $T(\tilde{x}, t) + G_*(\tilde{x}, t)$ in distribution. In the case that $C_N = N^r$ with $r > -1/2$, we have

$$
T_N(\tilde{x}, t) \to \infty. \qquad (28)
$$

According to (27) with $C_N = N^{-1/2}$ and (28) with $C_N = N^r (r > -1/2)$, Theorem 2 is proved. □

Acknowledgments The research described herein was supported by the Fundamental Research Funds for the Central Universities, a research Fund from Renmin University of China (No. 12XNI004), and a grant from Research Grants Council of Hong Kong, Hong Kong. The authors thank the associate editor and referees for their constructive comments and suggestions which led to the improvement of an early manuscript in presentation.

References

Dai, P. J., Sun, Z. H., Wang, P. (2010). Model Checking for general linear error-in-covariates model with validation data. *Journal of Systems Science and Complexity*, *23*, 1153–1166.

Du, L. L., Zou, C. L., Wang, Z. J. (2011). Nonparametric regression function estimation for errors-in-variables models with validation data. *Statistica Sinica*, *21*, 1093–1113.

Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, *21*, 1926–1947.

Stute, W., Zhu, L. X. (2005). Nonparametric checks for single-index models. *Annals of Statistics*, *33*, 1048–1083.

Stute, W., Xue, L. G., Zhu, L. X. (2007). Empirical likelihood inference in nonlinear errors-in-covariates models with validation data. *Journal of the American Statistical Association*, *102*, 332–346.

Stute, W., Xu, W. L., Zhu, L. X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika*, *95*, 451–467.

Sun, Z. H.,Wang, Q. H. (2009). Checking the adequacy of a general linear model with responses missing at random. *Journal of Statistical Planning and Inference*, *139*, 3588–3604.

Wang, Q. H. (1999). Estimation of partial linear error-in-variables models with validation data. *Jounal of Multivariate Analysis*, *69*, 30–64.

Wang, Q. H. (2003). Estimation of partial linear error-in-response models with validation data. *Annals of Institute of Statistical Mathematics*, *55*, 21–39.

Xu, W. L., Guo, X. (2013). Checking the adequacy of partial linear models with missing covariates at random. *Annals of Institute of Statistical Mathematics*, *65*, 473–490.

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.

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.