https://doi.org/10.1016/j.jspi.2018.03.002

This is the Pre-Published Version.

Consistent model check of errors-in-variables varying-coefficient model with auxiliary variable [1]

Zhifan LIU[a], Chunling LIU[b], Zhihua SUN[a,c]

[a] *University of Chinese Academy of Sciences, Beijing, 100049, China*
[b] *The Hong Kong Polytechnic University Kowloon, Hong Kong*
[c] *Key Laboratory of Big Data Mining and Knowledge Management of CAS, Beijing, 100049, China*

Abstract

In this paper, we consider the adequacy check of the varying-coefficient model when covariates are measured with error and some auxiliary variable is available. With the help of auxiliary variable, we calibrate the measurement error and obtain an estimator of the unobservable true variable. The empirical-process-based test is built by applying the calibrated estimator of the model error. The asymptotic properties of the proposed test are rigorously investigated under the null hypothesis, local and global alternatives. It is shown that the proposed test is consistent and has good properties of power. We illustrate that the naive method cannot control Type I error and loses effect completely. But the proposed calibrated method performs well in terms of the empirical sizes close to the test level and high empirical powers. Simulation studies and two real data analyses are conducted to demonstrate the performance of the proposed approach.

KEY WORDS: Varying-coefficient model; Measurement error; Auxiliary variable; Model check; Empirical process

Short Title: Test of EV varying-coefficient model with auxiliary variable

[1]The corresponding author is Dr. Zhihua Sun. Email: sunzh@amss.ac.cn. The research was supported by the National Natural Science Foundation of China (Grant Nos. 11571340, 11401502, U1430103) and the Open Project of Key Laboratory of Big Data Mining and Knowledge Management, CAS.

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/

1. INTRODUCTION

The varying-coefficient model is a popular semiparametric model, which takes the form:

$$Y = \alpha^\top(U)\xi + \varepsilon, \tag{1.1}$$

where Y is a scalar response variable, ξ is a $p-$dimensional predictor and U is scalar. The functional coefficient $\alpha(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_p(\cdot))^\top$ is unknown and the model error ε satisfies $E(\varepsilon^2|\xi, U) < \infty$. Model (1.1) was introduced by Hastie & Tibshirani (1993) and further studied extensively. See Fan & Zhang (2000), Fan & Zhang (2007), Wang et al. (2008), Park et al. (2015) and the references within.

In this paper, we consider the adequacy check of the varying-coefficient model with errors in covariates. That is, we aim to test

$$\mathcal{H}_0 : \exists\ \alpha(U)\ \text{ s.t. }\quad E[Y|\xi, U] = \alpha^\top(U)\xi,\ \text{ a.s.} \tag{1.2}$$

against the alternative hypothesis that $\mathcal{H}_0$ is not true when the variable ξ is not observed for some reasons, such as measurement error. Instead of ξ, its surrogate $\tilde{\xi}$ is observed. We assume that a d-dimensional auxiliary variable V is available to remit ξ. And the unobservable true variable ξ, the observed surrogate variable $\tilde{\xi}$ and the auxiliary variable V are of the following relationship:

$$\xi = E(\tilde{\xi}|V) =: \xi(V). \tag{1.3}$$

Actually, the error structure (1.3) is a special case of the additive error model because (1.3) is equivalent to the model that $\tilde{\xi} = \xi + e$ with $E(e|V) = 0$. We further assume that $E(e|\xi, U) = 0$ but allow for an unknown covariance of the error variable e. So we aim to investigate the adequacy of the following model

$$\begin{cases} Y &= \alpha^\top(U)\xi + \varepsilon \\ \tilde{\xi} &= \xi(V) + e. \end{cases} \tag{1.4}$$

This type of measurement error was also discussed in Zhou & Liang (2009), Zhao & Xue (2010), Sun et al. (2015) and Zhang et al. (2017). We present two examples, which are similar to Examples 1-2 in Zhou & Liang (2009), to illustrate the rationality of Model (1.4).

Example 1: (Errors-in-variables model with validation data) Consider Model $Y = \alpha^\top(U)X + \eta$, where $E(\eta|X,U) = 0$ with predictors X and U. The variable $\tilde{X}$ is an observed vector associated with vector X. We have a primary data set: $\{Y_i, \tilde{X}_i, U_i, i = 1, \cdots, n\}$ and a validation data set: $\{X_j, \tilde{X}_j, U_j, j = n+1, \cdots, n+n_0\}$. Let $V = (\tilde{X}^\top, U)^\top$. Then the errors-in-variables varying-coefficient model with validation data can be written as

$$\begin{cases} Y &= \alpha^\top(U)E(X|V) + \varepsilon \\ \varepsilon &= \eta + \alpha^\top(U)\{X - E(X|V)\}. \end{cases} \tag{1.5}$$

Let $\xi(V) = E(X|V)$ and $\tilde{\xi} = X$. Then Model (1.5) is a sub-model of Model (1.4).

Example 2: (De-noise varying-coefficient model) The relation between the response variable Y and covariates (ξ, U) is described by $Y = \alpha^\top(U)\xi + \varepsilon$ with $E(\varepsilon|\xi, U) = 0$, where $\xi = \xi(t)$ is subject to measurement error at time t. The variable $\tilde{\xi}$ is observed and serves as a surrogate of the variable ξ. Then the following model is considered

$$\begin{cases} Y &= \alpha^\top(U)\xi + \varepsilon \\ \tilde{\xi} &= \xi(t) + e. \end{cases} \tag{1.6}$$

Similar De-noise linear model was employed by Cai et al. (2000) to analyze the relationship between awareness and television rating points of TV commercials for certain products.

Let $\{(Y_i, \tilde{\xi}_i, U_i, V_i), i = 1, 2, \cdots, n\}$ be an i.i.d. sample from the population $(Y, \tilde{\xi}, U, V)$. For Model (1.3), if the measurement error is ignored, it can be validated that the naive estimator of the coefficient function, denoted by $\hat{\alpha}_{naive}(u)$, will be biased. Let $\hat{\varepsilon}_{naive}(Y_i, \tilde{\xi}_i, U_i)$ be the naive estimator of the model error ε for the $i-$th subject. It can be decomposed into three parts:

$$\begin{aligned} \hat{\varepsilon}_{naive}(Y_i, \tilde{\xi}_i, U_i) &= Y_i - \hat{\alpha}_{naive}^\top(U_i)\tilde{\xi}_i \\ = \{Y_i - \alpha^\top(U_i)\xi_i\} &+ \{\alpha(U_i) - \hat{\alpha}_{naive}(U_i)\}^\top\xi_i + \hat{\alpha}_{naive}^\top(U_i)(\xi_i - \tilde{\xi}_i), \end{aligned} \tag{1.7}$$

for $i = 1, 2, \cdots, n$. We can validate that the first and third terms, $Y_i - \alpha^\top(U_i)\xi_i$ and $\hat{\alpha}_{naive}^\top(U_i)(\xi_i - \tilde{\xi}_i)$, have zero expectations. However, the expectation of $\{\alpha(U_i) - \hat{\alpha}_{naive}(U_i)\}^\top\xi_i$ doesn't converge to zero in that the expectation of $\hat{\varepsilon}_{naive}(Y_i, \tilde{\xi}_i, U_i)$ is not equal to zero. Actually, the term $\{\alpha(U_i) - \hat{\alpha}_{naive}(U_i)\}^\top\xi_i$ acts just as a deviation function. This causes the naive test to tend to reject the null hypothesis even if it is true. We conduct some simulation studies in Section 4, which show that the naive method yields empirical sizes larger than 0.5 in many scenarios. A reasonable test should be able to control Type I error. This motivates us to develop a model checking method for (1.2) based on the calibration of the measurement error.

The estimation of the regression models with errors in covariates has been studied extensively. See Carroll et al. (2006); Li & Greene (2008); Liang et al. (1999); Ma et al. (2006), among others. However, the lack-of-fit test of regression models with measurement error has not received enough attention that it deserves. Sporadic researches can be found in the literature: Hall et al. (2007); Koul & Song (2009); Ma et al. (2011); Sun et al. (2015). For the model checking problem (1.2), we first calibrate the model error and then build an empirical process (EP) test with simple indicator (SI) weighting function, which has many merits. First, it is consistent; second, it is free from the nonparametric smoothing of the estimated model error; third, it can detect the alternative hypothetical model converging to the null hypothetical model at the parametric rate. More details of the EP test with SI weighting function can refer to Ma et al. (2014); Sun et al. (2009); Xu & Zhu (2015); Zhu & Ng (2003), among others.

The rest of paper is organized as follows. In Section 2, we calibrate the model error and develop an empirical process test. The asymptotic properties of the test statistic are rigorously studied in Section 3. In Section 4, simulation studies and real data analyses are conducted to validate the performance of the proposed test. The proofs of the main results are presented in the Appendix.

2. THE TESTING METHOD

2.1. The estimation of the null hypothetical model

From the error structure (1.3), recalling that we have a random sample $\{(\tilde{\xi}_j, V_j), j = 1, \cdots, n\}$ from $(\tilde{\xi}, V)$. we can define an estimator of the true variable ξ_i by the local smoothing method (Härdle et al. (2012)):

$$\hat{\xi}_n(V_i) = \frac{\sum_{j=1}^{n} \tilde{\xi}_j K_v(V_i - V_j)}{\sum_{j=1}^{n} K_v(V_i - V_j)}, \quad i = 1, 2, \cdots, n, \tag{2.1}$$

where $K_v(\cdot) = 1/h_n^d \prod k(\cdot/h_n)$ with a univariate kernel function $k(\cdot)$ and a bandwidth h_n. Clearly, $\hat{\xi}_n(V_i)$ is a $p \times 1$ vector with components $\hat{\xi}_{nl}(V_i), l = 1, \cdots, p$ for $i = 1, \cdots, n$. Then an estimator of the coefficient function $\alpha(u) = (\alpha_1(u), \cdots, \alpha_p(u))^\top$ and their derivatives can be defined by solving

$$\text{argmin}_{\{(a_l,b_l),l=1,\cdots,p)\}} \sum_{j=1}^{n} [Y_j - \sum_{l=1}^{p} \{a_l + b_l(U_j - u)\} \hat{\xi}_{nl}(V_j)]^2 \lambda_u(U_j - u), \tag{2.2}$$

where $\lambda(u)$ is a kernel function and $\lambda_u(U_j - u) = 1/l_n \lambda\{(U_j - u)/l_n\}$ with the bandwidth l_n. Let $\hat{\theta}_n(u) = (\hat{a}_1(u), \cdots, \hat{a}_p(u), \hat{b}_1(u), \cdots, \hat{b}_p(u))^\top$ be the solution to (2.2). Then $\hat{\alpha}_n(u) = (\hat{a}_1(u), \cdots, \hat{a}_p(u))^\top$ is the estimator of $\alpha(u)$. Thus estimators of the model error are available:

$$\hat{\varepsilon}_n(Y_i, \tilde{\xi}_i, U_i, V_i) = Y_i - \hat{\alpha}_n^\top(U_i)\hat{\xi}_n(V_i), \ i = 1, 2, \cdots, n.$$

Notice that, after calibration of the measurement error based on the auxiliary information, we estimate the coefficient functions by the local linear regression technique, which is different from locally corrected score equations presented by Li & Greene (2008) in the same model. Our modeling allows that variance-covariance matrix of the error term e in equation (1.4) can be unknown whereas Li and Greene's method assumed that such matrix structure is known and needs no auxiliary data.

For $\hat{\alpha}_n(u)$, we have the following result.

Proposition 1. *Under the regular conditions in the Appendix, we have*

$$\sqrt{nl_n}\{\hat{\alpha}(u) - \alpha(u) - \frac{\mu_2}{2}\alpha^{(2)}(u)l_n^2\} = \frac{Q^{-1}(u)}{\sqrt{nl_n}} \sum_{i=1}^{n} \xi_i \lambda\left(\frac{U_i - u}{l_n}\right)\varepsilon_i$$
$$+ \frac{Q^{-1}(u)}{\sqrt{nl_n}} \sum_{i=1}^{n} \xi_i E[\lambda\left(\frac{U_i - u}{l_n}\right)|V = V_i]\alpha^\top(u)(\tilde{\xi}_i - \xi_i) + o_p(1),$$

where $\alpha^{(2)}(u)$ is the second order derivative of $\alpha(u)$, $\mu_2 = \int u^2\lambda(u)du$ and $Q(u) = E(\xi\xi^\top|U = u)$.

By letting $D(\xi, U) = 0$, the result of Proposition 1 can be validated from Lemma 1 in the Appendix. We omit the details of the proofs.

2.2. *Construct the test statistic*

Note that the null hypothesis (1.2) can be transformed into infinite unconditional expectations: $E[(Y-\alpha^{\top}(U)\xi)I(\xi\leq z, U\leq u)]=0$ for all $z\in R^{P}, u\in R$. We can construct an estimated empirical process marked by the estimated residuals:

$$\widehat{CR}_n(z,u)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{Y_i-\hat{\alpha}_n^{\top}(U_i)\hat{\xi}_n(V_i)\}I(\hat{\xi}_n(V_i)\leq z, U_i\leq u).$$

Here the representation of $\widehat{CR}_n(z,u)$ is actually, up to a constant, an accumulative calibrated residuals with an indicator function being the weight. Then a Crámer-von Mises type test statistic can be defined as

$$T_n=\int[\widehat{CR}_n(z,u)]^2dF_n(z,u), \tag{2.3}$$

where $F_n(\cdot,\cdot)$ is the empirical distribution function based on data $\{(\hat{\xi}_n(V_i),U_i), i=1,\cdots,n\}$.

For the sake of simplicity, denote $I(\xi(V)\leq z, U\leq u)$ by $I(W\leq w)$ with $W=(\xi^{\top}(V),U)^{\top}$ and $w=(z^{\top},u)^{\top}$. Let $f_u(u)$, $f_v(v)$ and $f_{u,v}(u,v)$ be the densities of U, V and (U,V), respectively. Further denote $Q(u)=E(\xi\xi^{\top}|U=u)$, $\Gamma_1(V)=E\{\alpha(U)I(\xi(V)\leq z, U\leq u)|V\}$, $\Gamma_2(U)=E\{\xi(V)\ I(\xi(V)\leq z)|U\}$ and $\Gamma_3=E\{\xi^{\top}(V)I(W\leq w)\}$. Denote $H_{(z,u)}(Y,\xi,U,V;z,u)=\varepsilon I(W\leq w)-\Gamma_1(V)e+\xi^{\top}Q^{-1}(u)\varepsilon I(U\leq u)\Gamma_2(U)+\Gamma_3\xi\frac{f_{u,v}(u,V)}{f_u(u)f_v(V)}e^{\top}a(u)$. By combining the theories of the local kernel method and the empirical process, we have the following results for the test statistic T_n.

Theorem 1. *Under the regular conditions in the Appendix, when the null hypothesis (1.2) holds, the estimated empirical process $\widehat{CR}_n(z,u)$ converges in distribution to $R(z,u)$ in the Skorohod space $D[-\infty,\infty]^{p+1}$, where $R(z,u)$ is a centered Gaussian process with the covariance function:*

$$Cov\{R(z_1,u_1),R(z_2,u_2)\} \;=\; E(H_{(z,u)}(Y,\xi,U,V;z_1,u_1)H_{(z,u)}(Y,\xi,U,V;z_2,u_2)).$$

Furthermore, we have $T_n\stackrel{L}{\longrightarrow}\int[R(z,u)]^2dF(z,u)$, where $F(z,u)$ is the distribution function of (ξ,U).

When T_n is large enough, the null hypothesis should be rejected. The estimated empirical process $\widehat{CR}_n(z,u)$ takes the form of an accumulative summation of the estimated model errors $\hat{\varepsilon}_n(Y_i,\tilde{\xi}_i,U_i,V_i), i = 1,\cdots,n$. Similarly to (1.7), we decompose $\hat{\varepsilon}_n(Y_i,\tilde{\xi}_i,U_i,V_i)$ into three parts:

$$\{Y_i - \alpha^\top(U_i)\xi_i\} + \{\alpha(U_i) - \hat{\alpha}_n(U_i)\}^\top \xi_i + \hat{\alpha}_n^\top(U_i)(\xi_i - \tilde{\xi}_i).$$

Proposition 1 shows that $\hat{\alpha}_n(u)$ is an asymptotically consistent estimator of $\alpha(u)$ under the null hypothesis. However, under the Pitman alternative models, this fact is not true. Therefore $\hat{\varepsilon}_n(Y_i,\tilde{\xi}_i,U_i,V_i)$ and $\widehat{CR}_n(z,u)$ have asymptotically zero expectations under $\mathcal{H}_0$ in (1.2) and asymptotically nonzero expectation under the Pitman local alternative models. However, the naive estimation of the model error $\hat{\varepsilon}_{naive}(Y_i,\tilde{\xi}_i,U_i)$ and the estimated empirical process $n^{-1/2}\sum_{i=1}^n\{Y_i - \hat{\alpha}_{naive}^\top(U_i)\tilde{\xi}_i\}I(\tilde{\xi}_i \le z, U_i \le u)$ have asymptotically nonzero expectations under both null and alternative hypotheses. Hence the naive test cannot distinguish the null hypothesis from the alternative hypothesis.

2.3. *Realization of the test*

In this section, following Wu (1986), Stute et al. (1998) and Sun et al. (2018), we resort to a wild bootstrap method to determine the critical value of the test. Let $\{\eta_i, i = 1,2,\cdots,n\}$ be i.i.d. random variables with mean 0 and variance 1. The scheme of the bootstrap method is listed in the following.

First calculate the test statistic $T_n = \int[\widehat{CR}_n(z,u)]^2 dF_n(z,u)$ from the sample $\{(Y_i,\tilde{\xi}_i,U_i,V_i), i = 1,2,\cdots,n\}$. Then generate random variables $\{\eta_i, i = 1,2,\cdots,n\}$ and compute the bootstrap response variables: $Y_i^* = \hat{\alpha}_n^\top(U_i)\hat{\xi}_n(V_i) + \{Y_i - \hat{\alpha}_n^\top(U_i)\hat{\xi}_n(V_i)\}\eta_i,\ \ i = 1,\cdots,n$. Replacing the sample $\{(Y_i,\tilde{\xi}_i,U_i,V_i), i = 1,2,\cdots,n\}$ by the bootstrap sample $\{(Y_i^*,\tilde{\xi}_i,U_i,V_i), i = 1,2,\cdots,n\}$, we obtain the bootstrap value of T_n, denoted by T_n^*. Repeat the above steps B times and obtain B values of T_n^*: $T_{n1}^*,\cdots,T_{nB}^*$. Calculate the $1-\alpha$ empirical quantile of $T_{n1}^*,\cdots,T_{nB}^*$, which is taken as the $\alpha-$level critical value. Theoretically, a large value of B is preferable. For computational expedience, B can be chosen to be 300 or 500.

The aforementioned bootstrap procedure is robust. It doesn't depend on the variance or distribution of the model error. Further, the proposed method is data-driven. When the sample data is available, we can drive a conclusion whether the varying-coefficient model is adequate or not for the data from the proposed test procedure.

3. Analyses of the power

In the following, we investigate the performance of the proposed test under the local and global alternatives. First, we consider the Pitman local alternative hypothetical models:

$$\mathcal{H}_{1n}: \ Y_i = \alpha^\top(U_i)\xi_i + n^{-1/2}D(\xi_i, U_i) + \varepsilon_i, \ \ i = 1, 2, \cdots, n \tag{3.1}$$

with some bounded measurable nonzero function $D(\xi, U)$, which cannot take the form of $\alpha^\top(U)\xi$ for any measurable function vector $\alpha(U)$. Let $\Delta = E\{D(\xi, U)I(W \leq w)\}$ $+E[E\{\xi D(\xi, U)|U\}^\top \xi(V)I(W \leq w)]$.

Theorem 2. *Under the regular conditions in the Appendix, when the alternative hypothesis (3.1) holds, we have*

$$T_n \ \stackrel{L}{\longrightarrow} \ \int [R(z, u) + \Delta]^2 dF(z, u)$$

with $R(z, u)$ defined in Theorem 1.

Theorem 2 shows that, under the alternative hypothesis (3.1), the proposed test statistic T_n is asymptotically equivalent in distribution to a random variable which contains an additional nonzero drift quantity Δ, compared to its asymptotic equivalent random variable under the null hypothesis (1.2). This implied that the proposed test statistic is powerful enough to detect the Pitman alternative hypothesis (3.1). To the best of our knowledge, other types of test statistics do not possess this power in detection of such Pitman local alternative hypothesis. For example, the test based on U-statistic can only detect the local alternative model converging to the null hypothetical model at the rate $n^{-p/2}l_n^{-p/4}$(Li & Wang, 1998; Niu et al., 2016). And the rate that the weighted integrated squared distance (WISD) test can detect is slower than $n^{-1/2}l_n^{-p/4}$ (Härdle & Mammen, 1993). Here l_n is the bandwidth and p is the dimension of all the predictors.

Next we consider the alternative hypothetical models:

$$\mathcal{H}_{2n}: \quad Y_i = \alpha^\top(U_i)\xi_i + L_n D(\xi_i, U_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n \tag{3.2}$$

with $0 < L_n < n^{-1/2}$. When $L_n = 1$, $\mathcal{H}_{2n}$ in (3.2) is the global alternative hypothesis.

Theorem 3. *Under the regular conditions in the Appendix and the alternatives (3.2) with* $0 < L_n < n^{-1/2}$, $\mathcal{T}_n$ *converges to* ∞ *as* $n \to \infty$.

From the result with $L_n = 1$, we conclude that the proposed test is consistent against the global alternative hypothesis. Also, from Theorems 2 and 3, it can be concluded that the proposed test has satisfying power properties theoretically.

4. SIMULATION STUDIES AND REAL DATA ANALYSES

4.1. Simulation studies

In this section, we conduct simulation studies to demonstrate the performance of the proposed test. We consider two different examples.

Example 1. *A two-dimensional varying-coefficient model is investigated:*

$$Y = \sin(2U)\xi_1 + \{1 - \exp(U/2)\}\xi_2 + d|\xi_1/3 - \xi_2| + \varepsilon, \tag{4.1}$$

where $\xi_1 = 4|V - 1|, \xi_2 = V^2 + 3V - 1$, *denoted as Case 1 and* $\xi_1 = 3\sin(V) + 2, \xi_2 = V^2 - 2V + \cos(V)$, *denoted as Case 2. Here* $\tilde{\xi}_l = \xi_l + e_l, l = 1, 2$, $V \sim \mathcal{N}(0, 0.25)$ *and* $U \sim \mathcal{U}(0, 1)$.

Example 2. *A three-dimensional varying-coefficient model is considered:*

$$Y = \ln(U+1)\xi_1 + (U/2 - 1)\xi_2 + \sqrt{U}\xi_3 + d|\xi_1\xi_2/2 - 2\xi_2\xi_3 + \xi_3^2| + \varepsilon. \tag{4.2}$$

Two generating processes of the true variable are considered: Case 1: $\xi_1 = 4\sin(2\pi V), \xi_2 = 2^{V+1}, \xi_3 = 4\cos(2\pi V) - V/3$; *Case 2:* $\xi_1 = 4\cos(V^2 + V), \xi_2 = \exp\{3\sin(2\pi V) + 1/2\}, \xi_3 = V^2 + 3V - 2$. *Let* $\tilde{\xi}_l = \xi_l + e_l, l = 1, 2, 3$, $V \sim \beta(2, 1)$ *and* $U \sim \exp(1)$.

For both examples, the model error ε and the measurement error e are assumed normal random variables with mean 0 and variance 0.36. The constant d is chosen to be $0, 1, 2, 3, 4$ and $0, 0.1, 0.2, 0.3, 0.4$ for Examples 1 and 2 respectively. When $d = 0$, the null hypothesis holds. And different values of $d \neq 0$ mean that different alternative hypothetical models are considered. We choose Epanechnikov kernel function $K(t) = \lambda_u(t) = \frac{3}{4}(1-t^2)I_{(|t|\leq 1)}$. Let the smoothing parameters $h_n = 2.34 \min\{\hat{\sigma}_v^2, 3/4\hat{q}_v\}n^{-1/3}$ and $l_n = 2.34 \min\{\hat{\sigma}_u^2, 3/4\hat{q}_u\}n^{-1/3}$, where $\hat{\sigma}^2$ and $\hat{q}$ denote the sample standard deviation and sample inter-quartile range, respectively. The nominal level is set to be 0.05 and 0.1.

We calculate the empirical sizes and powers of the proposed test based on 1000 replications. For each replication, the bootstrap process is repeated 300 times. The results are presented in Figures 1-4. In addition, we compute the empirical sizes and powers of the naive method. The results are also shown in Figures 1-4. The naive method ignores the measurement error and applies the data with measurement error directly.

Here insert Figures 1-4.

From the results in Figures 1-4, we can observe that the proposed test performs well in term of its empirical sizes close to the test levels and the empirical powers increasing with the values of d and sample sizes. For both examples, when the sample sizes and the values of d are large enough, the empirical powers of the proposed test tend to be one, which is consistent with the conclusion that the proposed test is consistent. It can be found that the empirical sizes of the naive method are very high, which are larger than 0.9 in some scenarios. Moreover, as shown in Figure 3, the empirical sizes of the naive test are larger than the empirical powers with $d = 1$. We can further find that in all situations, the empirical sizes of the naive method increase with the sample sizes. Therefore, the naive method can hardly control Type I error though its empirical powers are high. Comparatively, the proposed test method can control Type I error and is more reliable.

4.2. Real data analyses

In this section, we conduct two real data analyses: one is a diabetes data set and the other is a Duchenne Muscular Dystrophy (DMD) data set.

Example 3. *We consider a diabetes data set (http://www.stanford.edu/ hastie/Papers /LARS /diabetes.data), which contains 442 observations. Let the response variable Y be a quantitative measurement of disease progression one year after baseline. There are other observed variables: age, body mass index (*`BMI`*), blood pressure (*`BP`*), high-density lipoprotein (*`HDL`*) and glucose concentration (*`GLU`*). We are interested in checking the following relationship between the response variable Y and the predictors (ξ_1, ξ_2, ξ_3, U):*

$$E\{Y|\xi_1, \cdots, \xi_3, U\} = \alpha_1(U)\xi_1 + \alpha_2(U)\xi_2 + \alpha_3(U)\xi_3, \tag{4.3}$$

where ξ_1, ξ_2, ξ_3 are age, `BMI` *and* `BP` *respectively and the variable U is chosen to be* `HDL`*. We take* `GLU` *to be the auxiliary variable V. The kernel function and smoothing parameters are chosen by the same methods as those in the simulation studies. The bootstrap procedure is repeated 10000 times. The p-values of the proposed and naive methods are calculated to be 0.0108 and 0.0087, respectively. Hence the candidate varying-coefficient model is not adequate for this diabetes data set.*

Here inserts Figure 5.

We plot the scatter plots of the estimated residuals against $\hat{\alpha_n}^\top(U)\hat{\xi}_n(V)$ and $\hat{\alpha}_{naive}^\top(U)\tilde{\xi}$ in Figure 5. Two subplots of Figure 5 show that the estimated residuals of both the calibrated and naive methods are asymmetric and both residual plots look like an asymmetric parallelogram. We also plot the nonparametric fitted curves and pointwise confidence curves of the calibrated and naive estimated residuals in Figure 5. The fitted curves reveal apparent deviations from the abscissa axis. Therefore, it is reasonable to conclude that the varying coefficient model is not appropriate for this diabetes data set.

Example 4. *We consider a data set of Duchenne Muscular Dystrophy (DMD), which contains 209 observations. This data set can be found in Ziegel (1987). DMD is a devastating, progressive muscular disease, which can be transmitted from the mother to the children. Though the affected mother has no apparent symptom, fortunately, it is found that the carriers of DMD tend to exhibit high levels of some serum enzymes or proteins. For this data set, we want to validate whether the varying-coefficient model is adequate or not. That is, we aim for checking the regression model:*

$$E[Y|\xi_1, \xi_2, U] = \alpha_1(U)\xi_1 + \alpha_2(U)\xi_2. \tag{4.4}$$

Here we choose the response Y to be the level of lactate dehydrogenase (`LD`*) and the covariates ξ_1, ξ_2, U to be the level of creatine kinase (*`CK`*), hemopexin (*`H`*) and age of patient respectively. We employ the variable age to calibrate the measurement error in the variables $\tilde{\xi}_1, \tilde{\xi}_2$. The settings are similar to those of Zhou & Liang (2009). The p-values of the proposed and naive test methods are computed, which are 0.4103 and 0.0139 respectively. So the result of the proposed method shows that the varying-coefficient model is adequate for this data set. But the naive method rejects the null hypothetical varying-coefficient model. This result is consistent with the theoretical analysis that the naive method tends to reject the null hypothesis.*

Here inserts Figure 6.

We plot the scatter plots of the estimated residuals against $\hat{\alpha}_n^\top(U)\hat{\xi}_n(V)$ and $\hat{\alpha}_{naive}^\top(U)\tilde{\xi}$ in Figure 6. Figure 6 suggests that the estimated model errors of the calibrated method are distributed approximately uniformly with regard to the abscissa axis. However, the estimated model errors of the naive method have an obvious negative center. In Figure 6, we also plot the fitted residual and confidence curves for both the calibrated and naive methods. The fitted residual curve of the calibrated method and the abscissa axis coincide almost. However, the fitted residual curve of the naive method deviates from the abscissa axis apparently. From the example, it reveals that by calibrating the measurement errors, the varying-coefficient model is adequate for this data set.

5. Discussion

In this work, we propose a test method for the errors-in-variables varying-coefficient model in the presence of an auxiliary variable by calibrating the measurement error. The theoretical and numerical studies illustrate that the proposed test performs well. The naive method breaks down in this setting.

The calibration skill in the paper can be extended to the model adequacy check of other errors-in-variables semiparametric model. Another extension is to consider the empirical process test with other weighting function, instead of the simple indicator weighting function. Some weighting functions can result in test methods with dimension reduction effect (Escanciano, 2006; Stute & Zhu, 2002; Sun et al., 2017). It should be interesting to develop test methods suitable for moderate or high dimensional data with measurement error by applying some special weighting functions.

Furthermore, in reality and in literature, another attention lies in the so-called Berkson measurement error in regression models(Wang (2003), Koul & Song (2009); among others). Take the varying coefficient model for instance, rather than the second equation in the model (1.4), one observes that the true variable ξ is related to an observed controlled variable $\tilde{\xi}$ via $\xi = \tilde{\xi} + e$. Consequently, the random measurement error e depends on the true variable ξ but is independent of the observed variable $\tilde{\xi}$. In linear regression models with Berkson measurement error, the least squares method can yield unbiased estimator of the regression parameters without calibration. However, this is not true for nonlinear regression. See Koul & Song (2008) for more details. A common tool to handle Berkson measurement error is the minimum distance moment estimating method. For varying-coefficient models in the presence of Berkson measurement error, the statistical inference is nontrivial. This no doubt deserves our future investigation separately.

References

Arcones, M. A. & Yu, B. (1994). Central limit theorems for empirical andu-processes of stationary mixing sequences. *Journal of Theoretical Probability* 7 47–71.

Cai, Z., Naik, P. A. & Tsai, C.-L. (2000). Denoised least squares estimators: an application to estimating advertising effectiveness. *Statistica Sinica* 10 1231–1242.

Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*. CRC press.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* 22 1030–1051.

Fan, J. & Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics* 27 715–731.

Fan, J. & Zhang, W. (2007). Statistical methods with varying coefficient models. *Statistics & Its Interface* 1 179–195.

Hall, P., Ma, Y. et al. (2007). Testing the suitability of polynomial models in errors-in-variables problems. *The Annals of Statistics* 35 2620–2638.

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

Härdle, W., Müller, M., Sperlich, S. & Werwatz, A. (2012). *Nonparametric and semiparametric models*. Springer Science & Business Media.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society* 55 757–796.

Koul, H. L. & Song, W. (2008). Regression model checking with berkson measurement errors. *Journal of Statistical Planning & Inference* 138 1615–1628.

KOUL, H. L. & SONG, W. (2009). Minimum distance regression model checking with berkson measurement errors. *Annals of Statistics* 37 132–156.

LI, L. & GREENE, T. (2008). Varying coefficients model with measurement error. *Biometrics* 64 519–526.

LI, Q. & WANG, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87 145–165.

LIANG, H., HÄRDLE, W. & CARROLL, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* 27 1519–1535.

MA, S., ZHANG, J., SUN, Z. & LIANG, H. (2014). Integrated conditional moment test for partially linear single index models incorporating dimension-reduction. *Electronic Journal of Statistics* 8 523–542.

MA, Y., CHIOU, J.-M. & WANG, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika* 93 75–84.

MA, Y., HART, J. D., JANICKI, R. & CARROLL, R. J. (2011). Local and omnibus goodness-of-fit tests in classical measurement error models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 81–98.

MASRY, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal* 17 571–599.

NIU, C., GUO, X., XU, W. & ZHU, L. (2016). Checking nonparametric component for partial linear regression model with missing response. *Journal of Statistical Planning and Inference* 168 1 – 19.

NOLAN, D. & POLLARD, D. (1988). Functional limit theorems for u-processes. *The Annals of Probability* 16 1291–1298.

PARK, B. U., MAMMEN, E., LEE, Y. K. & LEE, E. R. (2015). Varying coefficient regression models: a review and new developments. *International Statistical Review* 83 36–64.

STUTE, W., MANTEIGA, W. G. & QUINDIMIL, M. P. (1998). Bootstrap approximations in model checks for regression. *Publications of the American Statistical Association* 93 141–149.

STUTE, W. & ZHU, L.-X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics. Theory and Applications* 29 535–545.

SUN, Z., CHEN, F., ZHOU, X. & ZHANG, Q. (2017). Improved model checking methods for parametric models with responses missing at random. *Journal of Multivariate Analysis* 154 147–161.

SUN, Z., WANG, Q. & DAI, P. (2009). Model checking for partially linear models with missing responses at random. *Journal of Multivariate Analysis* 100 636–651.

SUN, Z., YE, X. & SUN, L. (2015). Consistent test of error-in-variables partially linear model with auxiliary variables. *Journal of Multivariate Analysis* 141 118–131.

SUN, Z., YE, X. & SUN, L. (2018). Consistent test for parametric models with right-censored data using projections. *Computational Statistics & Data Analysis* 119 112 – 125.

WANG, L. (2003). Estimation of nonlinear berkson-type measurement error models. *Statistica Sinica* 13 1201–1210.

WANG, L., LI, H. & HUANG, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association* 103 1556–1569.

WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14 1261–1295.

XU, W. & ZHU, L. (2015). Nonparametric check for partial linear errors-in-covariables models with validation data. *Annals of the Institute of Statistical Mathematics* 67 793–815.

ZHANG, J., FENG, Z., XU, P. & LIANG, H. (2017). Generalized varying coefficient partially linear measurement errors models. *Annals of the Institute of Statistical Mathematics* 69 97–120.

ZHAO, P. & XUE, L. (2010). Variable selection for semiparametric varying coefficient partially linear errors-in-variables models. *Journal of Multivariate Analysis* 101 1872–1883.

ZHOU, Y. & LIANG, H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *Annals of Statistics* 37 427–458.

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

ZIEGEL, E. (1987). Data: A collection of problems from many fields for the student and research worker. *Technometrics* 29 502–503.

APPENDIX

We begin this section by listing the conditions needed in the proofs of the theorems.

(C.1) The functions $\alpha(u)$ and $\xi(v)$ satisfy Lipschitz condition of order 2.

(C.2) The matrix $Q(u) = E(\xi\xi^\top|U = u)$ is positive definite; $E|\varepsilon|^{2+\delta} < \infty$ and $E|e|^{2+\delta} < \infty$ with $\delta > 0$.

(C.3) The densities of U and V, say $f_u(u)$ and $f_v(v)$, exist and satisfy $0 < \inf_u f_u(u) \leq \sup_u f_u(u) < \infty$ and $0 < \inf_v f_v(v) \leq \sup_v f_v(v) < \infty$.

(C.4) The univariate kernel functions $k(\cdot)$ and $\lambda(\cdot)$ are bounded kernel functions of order $j(\geq 2)$ with bounded support.

(C.5) (i) $nh_n^p \to \infty$ and $h_n \to 0$; (ii) $\ln(n)/(nl_n) \to 0$ and $nl_n^2 \to 0$. (iii) $l_n {h_n^{-p/2}} (\ln(n))^{-1} \to 0$ and $h_n^{2p} {l_n^{-1/2}} (\ln(n))^{-1} \to 0$.

Remark Conditions (C.1)-(C.2) are necessary for proving the asymptotic normality of the estimator of the coefficient function. Condition (C.3) aims at simplifying the proofs of the theorems. Otherwise, a truncation technique should be applied since some denominators are zeros. Conditions (C.4) and(C.5) are common to obtain the convergence rates of the nonparametric estimates.

Lemma 1. *Under Conditions (C.1)-(C.5) and the alternative model (3.1), we have*

$$\begin{aligned}\sqrt{nl_n}\{\hat{\alpha}(u)-\alpha(u)-\frac{\mu_2}{2}\alpha^{(2)}(u)l_n^2\} &= \frac{Q^{-1}(u)}{\sqrt{nl_n}f_u(u)}\sum_{i=1}^{n}\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\varepsilon_i \\ &+\frac{Q^{-1}(u)}{\sqrt{nl_n}f_u(u)}\sum_{i=1}^{n}\xi_i E[\lambda\left(\frac{U_i-u}{l_n}\right)|V=V_i]a^\top(u)(\tilde{\xi}_i-\xi_i) \\ &+l_n^{1/2}Q^{-1}(u)E\{\xi D(\xi,U)|U=u\}+o_p(1),\end{aligned}$$

where $\mu_2=\int u^2\lambda(u)du$.

Proof Denote $\theta(u)=(\alpha_1(u),\ldots,\alpha_p(u),l_nb_1(u),\ldots,l_nb_p(u))^\top$, where $b_i(u)$ is the derivative of $\alpha_i(u)$ for $i=1,\cdots,p$. From (2.2), by some simple calculations, we can obtain that

$$(nl_n)^{1/2}\{\hat{\theta}_n(u)-\theta(u)\}=:\hat{A}_n^{-1}B_n$$

with

$$\hat{A}_n = \frac{1}{nl_n}\sum_{i=1}^{n}\begin{pmatrix}\hat{\xi}_n(V_i)\hat{\xi}_n^\top(V_i) & \hat{\xi}_n(V_i)\hat{\xi}_n^\top(V_i)(U_i-u)/l_n \\ \hat{\xi}_n(V_i)\hat{\xi}_n^\top(V_i)(U_i-u)/l_n & \hat{\xi}_n(V_i)\hat{\xi}_n^\top(V_i)((U_i-u)/l_n)^2\end{pmatrix}\lambda(\frac{U_i-u}{l_n})$$

and

$$B_n=:\left\{\begin{array}{l}(nl_n)^{-1/2}\sum_{i=1}^n\hat{\xi}_n(V_i)\lambda\left(\frac{U_i-u}{l_n}\right)[Y_i-\sum_{j=1}^p\{\alpha_j(u)+b_j(u)(U_i-u)\}\hat{\xi}_{nj}(V_i)] \\ (nl_n)^{-1/2}\sum_{i=1}^n\hat{\xi}_n(V_i)(U_i-u)/l_n\lambda(\frac{U_i-u}{l_n})[Y_i-\sum_{j=1}^p\{\alpha_j(u)+b_j(u)(U_i-u)\}\hat{\xi}_{nj}(V_i)].\end{array}\right. .$$

Step 1. Consider A_n. By the fact that $\sup_v|\hat{\xi}_n(v)-\xi(v)|=O_P((\ln(n)/nh_v^d)^{1/2})+h_n^{2d}$ (Theorem 6 in Masry (1996)), the law of large numbers and Conditions (C.4) and(C.5), we can prove that

$$\hat{A}_n = \begin{pmatrix}E(\xi\xi^\top|U=u) & 0 \\ 0 & E(\xi\xi^\top|U=u)\mu_2\end{pmatrix}f_u(u)+o_p(1).$$

Thus it yields

$$(nl_n)^{1/2}\{\hat{\alpha}_n(u)-\alpha(u)\} = E^{-1}(\xi\xi^\top|U=u)f_u^{-1}(u)B_{n1}+o_p(1) \qquad \text{(A.1)}$$

with $B_{n1}=(nl_n)^{-1/2}\sum_{i=1}^n \hat{\xi}_n(V_i)\lambda\left(\frac{U_i-u}{l_n}\right)[Y_i-\sum_{j=1}^p\{a_j(u)+b_j(u)(U_i-u)\}\hat{\xi}_{nj}(V_i)]$.

Step 2. In the following, we focus on B_{n1}. We can split B_{n1} into two parts:

$$\begin{aligned}
B_{n1} &= \frac{1}{\sqrt{nl_n}}\sum_{i=1}^n \xi_i\lambda\left(\frac{U_i-u}{l_n}\right)[Y_i-\sum_{j=1}^p\{\alpha_j(u)+b_j(u)(U_i-u)\}\hat{\xi}_{nj}(V_i)] \\
&\quad +\frac{1}{\sqrt{nl_n}}\sum_{i=1}^n(\hat{\xi}_n(V_i)-\xi_i)\lambda\left(\frac{U_i-u}{l_n}\right)[Y_i-\sum_{j=1}^p\{\alpha_j(u)+b_j(u)(U_i-u)\}\hat{\xi}_{nj}(V_i)] \\
&=: B_{n1}^{[1]}+B_{n1}^{[2]}. \qquad \text{(A.2)}
\end{aligned}$$

Step 2.1 Consider $B_{n1}^{[1]}$. We have

$$\begin{aligned}
B_{n1}^{[1]} &= \frac{1}{\sqrt{nl_n}}\sum_{i=1}^n \xi_i\lambda\left(\frac{U_i-u}{l_n}\right)[Y_i-\sum_{j=1}^p\{\alpha_j(u)+b_j(u)(U_i-u)\}\xi_{ij}] \\
&\quad +\frac{1}{\sqrt{nl_n}}\sum_{i=1}^n \xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^p\{\alpha_j(u)+b_j(u)(U_i-u)\}(\hat{\xi}_{nj}(V_i)-\xi_{ij}) \\
&= B_{n1,1}^{[1]}+B_{n1,2}^{[1]}. \qquad \text{(A.3)}
\end{aligned}$$

Step 2.1.1 We deal with the first term $B_{n1,1}^{[1]}$:

$$\begin{aligned}
B_{n1,1}^{[1]} &= \frac{1}{\sqrt{nl_n}}\sum_{i=1}^n \xi_i\lambda\left(\frac{U_i-u}{l_n}\right)(Y_i-\sum_{j=1}^p\alpha_j(U_i)\xi_{ij}) \\
&\quad +\frac{1}{\sqrt{nl_n}}\sum_{i=1}^n \xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^p\{\alpha_j(U_i)-\alpha_j(u)-b_j(u)(U_i-u)\}\xi_{ij} \\
&= \frac{1}{\sqrt{nl_n}}\sum_{i=1}^n \xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\varepsilon_i+(nl_n)^{-1}l_n^{1/2}\sum_{i=1}^n\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)D(\xi_i,U_i) \\
&\quad +\frac{1}{2\sqrt{nl_n}}\sum_{i=1}^n \xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^p\alpha_j^{(2)}(u)(U_i-u)^2\xi_{ij}+o_p((nl_n)^{1/2}l_n^2).
\end{aligned}$$

Note that $(nl_n)^{-1/2}\sum_{i=1}^n\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^p\alpha_j^{(2)}(u)(U_i-u)^2\xi_{ij} = (nl_n)^{-1/2}\sum_{i=1}^n\xi_i\xi_i^\top\alpha^{(2)}(u)$ $(U_i-u)^2\lambda(\frac{U_i-u}{l_n}) = (nl_n)^{1/2}E(\xi\xi^\top|U=u)\alpha^{(2)}(u)\mu_2 l_n^2 f_u(u)+o_p((nl_n)^{1/2})l_n^2$. Furthermore, we can prove that $(nl_n)^{-1}\sum_{i=1}^n\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)D(\xi_i,U_i)=E\{\xi D(\xi,U)|U=u\}f_u(u)+o_p(1)$. Then it yields

$$B^{[1]}_{n1,1} = \frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\varepsilon_i + l_n^{1/2}E\{\xi D(\xi,U)|U=u\}f_u(u)$$
$$+\sqrt{nl_n}l_n^2E(\xi\xi^\top|U=u)\alpha^{(2)}(u)\mu_2 f_u(u)+o_p(1). \quad (A.4)$$

Step 2.1.2. We deal with $B^{[1]}_{n1,2}$. Recalling the definition of $\hat{\xi}_{nj}(V_i)$, we have

$$B^{[1]}_{n1,2} = \frac{1}{\sqrt{n^3l_nh_n^{2d}}}\sum_{i=1}^{n}\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^{p}a_j(u)/f_v(V_i)\sum_{l=1}^{n}(\tilde{\xi}_{lj}-\xi_{ij})K\left(\frac{V_i-V_l}{h_n}\right)$$
$$+\frac{1}{\sqrt{n^3l_nh_n^{2d}}}\sum_{i=1}^{n}\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^{p}b_j(u)(U_i-u)/f_v(V_i)\sum_{l=1}^{n}(\tilde{\xi}_{lj}-\xi_{ij})K\left(\frac{V_i-V_l}{h_n}\right)$$
$$+o_p(1)$$
$$=: B^{[1]}_{n1,21}+B^{[1]}_{n1,22}+o_p(1).$$

Note the fact that $(nh_n^d)^{-1}\sum_{i=1}^n\{\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)/f_v(V_i)\}K\left(\frac{V_i-V_l}{h_n}\right) = E[\xi\lambda\left(\frac{U_l-u}{l_n}\right)|V=V_l] + o_p(1)$, we have

$$B^{[1]}_{n1,21} = \frac{1}{\sqrt{nl_n}}\sum_{l=1}^{n}\sum_{j=1}^{p}a_j(u)\{1/(nh_n^d)\sum_{i=1}^{n}\frac{\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)}{f_v(V_i)}K\left(\frac{V_i-V_l}{h_n}\right)\}(\tilde{\xi}_{jl}-\xi_{jl})$$
$$+o_p(1).$$
$$= \frac{1}{\sqrt{nl_n}}\sum_{l=1}^{n}\sum_{j=1}^{p}a_j(u)E[\xi\lambda(\frac{U_l-u}{l_n})|V=V_l](\tilde{\xi}_{jl}-\xi_{jl})+o_p(1)$$
$$= \frac{1}{\sqrt{nl_n}}\sum_{l=1}^{n}\xi_lE[\lambda\left(\frac{U-u}{l_n}\right)|V=V_l]\alpha^\top(u)(\tilde{\xi}_l-\xi_l)+o_p(1).$$

We can further prove that $E[B^{[1]}_{n1,22}]^2 = O(l_n^2)$. By Condition (4)，it yields $B^{[1]}_{n1,22} = o_p(1)$. Therefore

$$B^{[1]}_{n1,2} = \frac{1}{\sqrt{nl_n}}\sum_{l=1}^{n}\xi_lE[\lambda\left(\frac{U-u}{l_n}\right)|V=V_l]a^\top(u)(\tilde{\xi}_l-\xi_l)+o_p(1).$$

This, together with (A.3) and (A.4)，we can get

$$B^{[1]}_{n1} = \frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\varepsilon_i+(nl_n)^{1/2}E(\xi\xi^\top|U=u)\alpha^{(2)}(u)\mu_2l_n^2$$
$$+l_n^{1/2}E\{\xi D(\xi,U)|U=u\}+\frac{1}{\sqrt{nl_n}}\sum_{l=1}^{n}\xi_lE[\lambda(\frac{U_l-u}{l_n})|V=V_l]a^\top(u)(\tilde{\xi}_l-\xi_l)$$

$$+o_p(1). \tag{A.5}$$

Step 2.2. We consider $B_{n1}^{[2]}$:

$$\begin{aligned}
B_{n1}^{[2]} &= \frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}(\hat{\xi}_n(V_i)-\xi_i)\lambda\left(\frac{U_i-u}{l_n}\right)(Y_i-\alpha^\top(U_i)\xi_i) \\
&+\frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}(\hat{\xi}_n(V_i)-\xi_i)\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^{p}\{a_j(U_i)-\{a_j(u)+b_j(u)(U_i-u)\}\}\xi_{ji} \\
&+\frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}(\hat{\xi}_n(V_i)-\xi_i)\lambda\left(\frac{U_i-u}{l_n}\right)\sum_{j=1}^{p}\{a_j(u)+b_j(u)(U_i-u)\}(\xi_{ji}-\hat{\xi}_{nj}(V_i)) \\
&= B_{n1,1}^{[2]}+B_{n1,2}^{[2]}+B_{n1,3}^{[2]}.
\end{aligned}$$

For $B_{n1,1}^{[2]}$, we can obtain that

$$\begin{aligned}
B_{n1,1}^{[2]} &= \frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}\lambda(\frac{U_i-u}{l_n})(Y_i-\alpha^\top(U_i)\xi_i)/(nh_n^d f_v(V_i))\sum_{l=1}^{n}(\tilde{\xi}_l-\xi_l)K\left(\frac{V_i-V_l}{h_n}\right) \\
&+\frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}\lambda(\frac{U_i-u}{l_n})(Y_i-\alpha^\top(U_i)\xi_i)/(nh_n^d f_v(V_i))\sum_{l=1}^{n}(\xi_l-\xi_i)K\left(\frac{V_i-V_l}{h_n}\right) \\
&=: B_{n1,11}^{[2]}+B_{n1,12}^{[2]}.
\end{aligned}$$

We further can prove that $E[B_{n1,11}^{[2]}]=O(1/(nh_n^d))\to 0$. Thus we can get $B_{n1,11}^{[2]}=o_p(1)$. Note that $\xi=\xi(V)$ which satisfies Lipschitz condition. The we can prove $E[B_{n1,12}^{[2]}]=O(h_n^{2d})\to 0$. So we can obtain $B_{n1,12}^{[2]}=o_p(1)$ and then $B_{n1,1}^{[2]}=o_p(1)$. Similarly, we can prove that $B_{n1,2}^{[2]}=o_p(1)$ and $B_{n1,3}^{[2]}=o_p(1)$. Thus it follows that

$$B_{n1}^{[2]}=o_p(1). \tag{A.6}$$

By (A.2), (A.5) and (A.6), we can obtain that

$$\begin{aligned}
B_{n1} &= \frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}\xi_i\lambda\left(\frac{U_i-u}{l_n}\right)\varepsilon_i+(nl_n)^{1/2}l_n^2E(\xi\xi^\top|U=u)\alpha^{(2)}(u)\mu_2 f_u(u) \\
&+l_n^{1/2}E\{\xi D(\xi,U)|U=u\}f_u(u)+\frac{1}{\sqrt{nl_n}}\sum_{i=1}^{n}\xi_iE[\lambda(\frac{U_i-u}{l_n})|V=V_i]\alpha^\top(u)(\tilde{\xi}_i-\xi_i) \\
&+o_p(1).
\end{aligned}$$

This, together with (A.1), can prove the lemma.

Proof of Theorem 1 The results of Theorem 1 can be obtained from Theorem 2 by setting $D(\xi, U) = 0$. We omit the details. □

Proof of Theorem 2 We first consider $\widehat{CR}_n(z, u)$:

$$\begin{aligned}
\widehat{CR}_n(z,u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - \alpha^\top(U_i)\xi(V_i)\} I(\hat{\xi}_n(V_i) \le z, U_i \le u) \\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^n \alpha^\top(U_i)\{\hat{\xi}_n(V_i) - \xi(V_i)\} I(\hat{\xi}_n(V_i) \le z, U_i \le u) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\hat{\alpha}(U_i) - \alpha(U_i)\}^\top \xi(V_i) I(\hat{\xi}_n(V_i) \le z, U_i \le u) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\hat{\alpha}(U_i) - \alpha(U_i)\}^\top \{\hat{\xi}_n(V_i) - \xi(V_i)\} I(\hat{\xi}_n(V_i) \le z, U_i \le u) \\
&=: \sum_{j=1}^4 \widehat{CR}_{nj}(z,u). \qquad \text{(A.7)}
\end{aligned}$$

Step 1. Consider $\widehat{CR}_{n1}(z, u)$. We have the following decomposition:

$$\widehat{CR}_{n1}(z,u) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i I(U_i \le u) I(\hat{\xi}_n(V_i) \le z) + \frac{1}{n}\sum_{i=1}^n D(\xi_i, U_i) I(U_i \le u) I(\hat{\xi}_n(V_i) \le z).$$

Note that $I(\hat{\xi}_n(V_i) \le z) = I(\xi(V_i) \le z) - I(\xi(V_i) \le z \le \hat{\xi}_n(V_i)) + I(\hat{\xi}_n(V_i) \le z \le \xi(V_i))$. Then for the first term of $\widehat{CR}_{n1}(z, u)$, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i I(U_i \le u) I(\hat{\xi}_n(V_i) \le z) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i I(U_i \le u) I(\xi(V_i) \le z) \\
- \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i I(\xi(V_i) \le z \le \hat{\xi}_n(V_i)) &+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i I(\hat{\xi}_n(V_i) \le z \le \xi(V_i))
\end{aligned}$$

We consider the term $\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i I(\xi(V_i) \le z \le \hat{\xi}_n(V_i))$. It is easy to prove that $E[\varepsilon_i\varepsilon_j I(\xi(V_i) \le z \le \hat{\xi}_n(V_i)) I(\xi(V_j) \le z \le \hat{\xi}_n(V_j)) = 0, i \ne j$ by the condition that $E(\varepsilon|X, \tilde{X}, V) = 0$ and the independence of the sample. Then we have $E[n^{-1/2}\sum_{i=1}^n \varepsilon_i I(\xi(V_i) \le z \le \hat{\xi}_n(V_i))]^2 \le n^{-1}\sum_{i=1}^n E[\varepsilon_i^2] E[I(\xi(V_i) \le z \le \hat{\xi}_n(V_i))]$. By the fact that $\sup_v |\hat{\xi}_n(v) - \xi(v)| \to 0$, we have $\forall v, \forall \eta > 0, \exists M\ s.t.\ |\hat{\xi}_n(v) - \xi(v)| < \eta$ when $n > M$. Thus for large n, $n^{-1}\sum_{i=1}^n E[\varepsilon_i^2] E[I(\xi(V_i) \le z \le \hat{\xi}_n(V_i))] = n^{-1}\sum_{l=1}^M E[\varepsilon_i^2] E[I(\xi(V_i) \le z \le \hat{\xi}_n(V_i))] + n^{-1}\sum_{l=M}^n E[\varepsilon_i^2] E[I(\xi(V_i) \le z \le \hat{\xi}_n(V_i))]$. We can observe that $n^{-1}\sum_{l=M}^n E[\varepsilon_i^2] E[I(z + \xi(V_i) - \hat{\xi}_n(V_i) \le \xi(V_i) \le z] \le$

$n^{-1}\sum_{l=M}^{n} E[\varepsilon_i^2]\ E[I(z-\eta \le \xi(V_i) \le z)] \le C\eta$ where C is some constant related to the variance of ε and the upper bound of the density function of X. Therefore we have $E[n^{-1/2}\sum_{i=1}^{n}\varepsilon_i I(\xi(V_i) \le z \le \hat{\xi}_n(V_i))]^2$ converges to zero as $n \to \infty$. Thus we can get $n^{-1/2}\sum_{i=1}^{n}\varepsilon_i I(\xi(V_i) \le z \le \hat{\xi}_n(V_i)) = o_p(1)$. Similarly, we can prove that $n^{-1/2}\sum_{i=1}^{n}\varepsilon_i$ $I(\hat{\xi}_n(V_i) \le z \le \xi(V_i)) = o_p(1)$. To avoid tedious proofs, we replace $I(\hat{\xi}_n(v) \le z)$ by $I(\xi(V_i) \le z)$ in the following proofs. For $\widehat{CR}_{n1}(z,u)$, we have

$$\widehat{CR}_{n1}(z,u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i I(W_i \le w) + E\{D(\xi,U)I(W \le w)\} + o_p(1). \quad \text{(A.8)}$$

Step 2. We consider $\widehat{CR}_{n2}(z,u)$. It can be split into two parts:

$$\begin{aligned}
\widehat{CR}_{n2}(z,u) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\alpha^\top(U_i)\{\hat{\xi}_n(V_i)-\xi(V_i)\}I(W_i \le w) + o_p(1)\\
&= -\frac{1}{n^{3/2}}\sum_{i=1}^{n}\alpha^\top(U_i)\sum_{j=1}^{n}\{\tilde{\xi}_j-\xi(V_j)\}K_v(V_i-V_j)/f_v(V_i)I(W_i \le w)\\
&\quad -\frac{1}{n^{3/2}}\sum_{i=1}^{n}\alpha^\top(U_i)\sum_{j=1}^{n}\{\xi(V_j)-\xi(V_i)\}K_v(V_i-V_j)/f_v(V_i)I(W_i \le w) + o_p(1)\\
&=: \widehat{CR}_{n2}^{[1]}(z,u) + \widehat{CR}_{n2}^{[2]}(z,u).
\end{aligned}$$

For the first term $\widehat{CR}_{n2}^{[1]}(z,u)$, we have

$$\begin{aligned}
\widehat{CR}_{n2}^{[1]}(z,u) &= -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{\sum_{i=1}^{n}\alpha(U_i)K_v(V_i-V_j)/f_v(V_i)I(W_i \le w)\}^\top\{\tilde{\xi}_j-\xi(V_j)\}\\
&= -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\{\alpha(U)I(W \le w)|V=V_j\}^\top\{\tilde{\xi}_j-\xi(V_j)\} + o_p(1).
\end{aligned}$$

Furthermore, we can prove that $\widehat{CR}_{n2}^{[2]}(z,u) = O_p(n^{1/2}h_n^{2d}) = o_p(1)$ by Condition (C4). Thus

$$\widehat{CR}_{n2}(z,u) = -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\{\alpha(U)I(W \le w)|V=V_j\}^\top\{\tilde{\xi}_j-\xi(V_j)\} + o_p(1). \quad \text{(A.9)}$$

Step 3. We consider $\widehat{CR}_{n3}(z,u)$ and $\widehat{CR}_{n4}(z,u)$. For $\widehat{CR}_{n3}(z,u)$, we have

$$\begin{aligned}
\widehat{CR}_{n3}(z,u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{nl_nf_u(u)}\sum_{j=1}^{n}\xi_j^\top Q^{-1}(u)\lambda(\frac{U_j-U_i}{l_n})\varepsilon_j\xi(V_i)I(W_i \le w)\\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mu_2}{2}\alpha^{(2)}(u)^\top l_n^2\xi(V_i)I(W_i \le w) + \frac{1}{n}\sum_{i=1}^{n}E\{\xi D(\xi,U)|U=U_i\}^\top\xi(V_i)I(W_i \le w)
\end{aligned}$$

$$
+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{nl_nf_u(u)}\sum_{l=1}^{n}\xi_l^\top Q^{-1}(u)E[\lambda(\frac{U_l-u}{l_n})|V=V_l]\alpha(u)(\tilde{\xi}_l-\xi_l)^\top\xi(V_i)I(W_i\leq w)+o_p(1)
$$
$$
=\sum_{l=1}^{4}\widehat{CR}_{n3,l}(z,u)+o_p(1).
$$

For $\widehat{CR}_{n3,1}(z,u)$, we have

$$
\begin{aligned}
\widehat{CR}_{n3,1}(z,u) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\xi_j^\top Q^{-1}(u)\varepsilon_j\frac{1}{nl_nf_u(u)}\sum_{i=1}^{n}\lambda(\frac{U_j-U_i}{l_n})\xi(V_i)I(W_i\leq w)\} \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\xi_j^\top\varepsilon_jQ^{-1}(u)I(U_j\leq u)E\{\xi(V)I(\xi(V)\leq z)|U=U_j\}+o_p(1).
\end{aligned}
$$

It can further be proved that

$$
\widehat{CR}_{n3,2}(z,u) = O_p(\sqrt{n}l_n^2)=o_p(1)
$$

and

$$
\widehat{CR}_{n3,3}(z,u) = E[E\{\xi D(\xi,U)|U\}^\top\xi(V)I(W\leq w)]+o_p(1).
$$

For the fourth term, $1/l_nE\{\lambda(\frac{U_l-u}{l_n})|V=V_l\}=1/l_nE\{\lambda(\frac{U_l-u}{l_n})\}=\frac{f_{u,v}(u,V_l)}{f_v(V_l)}+o_p(1)$. Thus we can obtain that

$$
\begin{aligned}
\widehat{CR}_{n3,4}(z,u) &= n^{-3/2}\sum_{i=1}^{n}\sum_{l=1}^{n}\xi^\top(V_i)\xi_la^\top(u)\frac{f_{u,v}(u,V_l)}{f_u(u)f_v(V_l)}(\tilde{\xi}_l-\xi_l)I(W_i\leq w)+o_p(1) \\
&= \frac{E\{\xi^\top(V)I(W\leq w)\}}{\sqrt{n}}\sum_{l=1}^{n}\xi_l(\tilde{\xi}_l-\xi_l)^\top\frac{f_{u,v}(u,V_l)\alpha(u)}{f_u(u)f_v(V_l)}+o_p(1).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\widehat{CR}_{n3}(z,u) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\xi_j^\top Q^{-1}(u)\varepsilon_jI(U_j\leq u)E\{\xi(V)I(\xi(V)\leq z)|U=U_j\} \\
&\quad+\frac{E\{\xi^\top(V)I(W\leq w)\}}{\sqrt{n}}\sum_{l=1}^{n}\xi_l\frac{f_{u,v}(u,V_l)}{f_u(u)f_v(V_l)}(\tilde{\xi}_l-\xi_l)^\top a(u) \\
&\quad+E[E\{\xi D(\xi,U)|U\}^\top\xi(V)I(W\leq w)]+o_p(1). \qquad\qquad (A.10)
\end{aligned}
$$

We can further prove that

$$
\widehat{CR}_{n4}(z,u) = o_p(1). \qquad\qquad (A.11)
$$

From (A.7)-(A.11), we have

$$\widehat{CR}_n(z,u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_{(z,u)}(Y_i,\xi_i,U_i;z,u) + \Delta + o_p(1) \tag{A.12}$$

with $H_{(z,u)}(Y,\xi,U;z,u)$ and Δ defined in Sections 2 and 3. Because the indicator function is monotone, it is easy to prove that $G_u = \{H_{(z,u)}(Y,\xi,U;z,u) : z \in R^p, u \in R\}$ is a V-C class of functions. See Nolan & Pollard (1988). By Theorem 3.1 of Arcones & Yu (1994), we can show that $\widehat{CR}_n(z,u)$ converges to a Gaussian process. Further by the continuous mapping theorem, we can prove the result for T_n. □

Proof of Theorem 3 If we denote $D_{new}(\xi,U) = n^{1/2}L_nD(\xi,U)$, then (3.2) can be transformed into

$$H_{1n}: \ \ Y_i = \alpha^\top(U_i)\xi_i + n^{-1/2}D_{new}(\xi_i,U_i) + \varepsilon_i, \ \ i = 1,2,\cdots,n. \tag{A.13}$$

By the similar method to prove Theorem 2, we can prove that

$$\widehat{CR}_n(z,u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_{(z,u)}(Y_i,\xi_i,U_i;z,u) + \Delta_{new} + o_p(1) \tag{A.14}$$

with $H_{(z,u)}(Y,\xi,U;z,u)$ defined in Section 2 and $\Delta_{new} = E\{D_{new}(\xi,U)I(W \le w)$ $+E[E\{\xi$ $D_{new}(\xi,U)|U\}^\top\ \xi(V)I(W \le w)] \to \infty$ as $n \to \infty$. So T_n converges to ∞ as $n \to \infty$. □

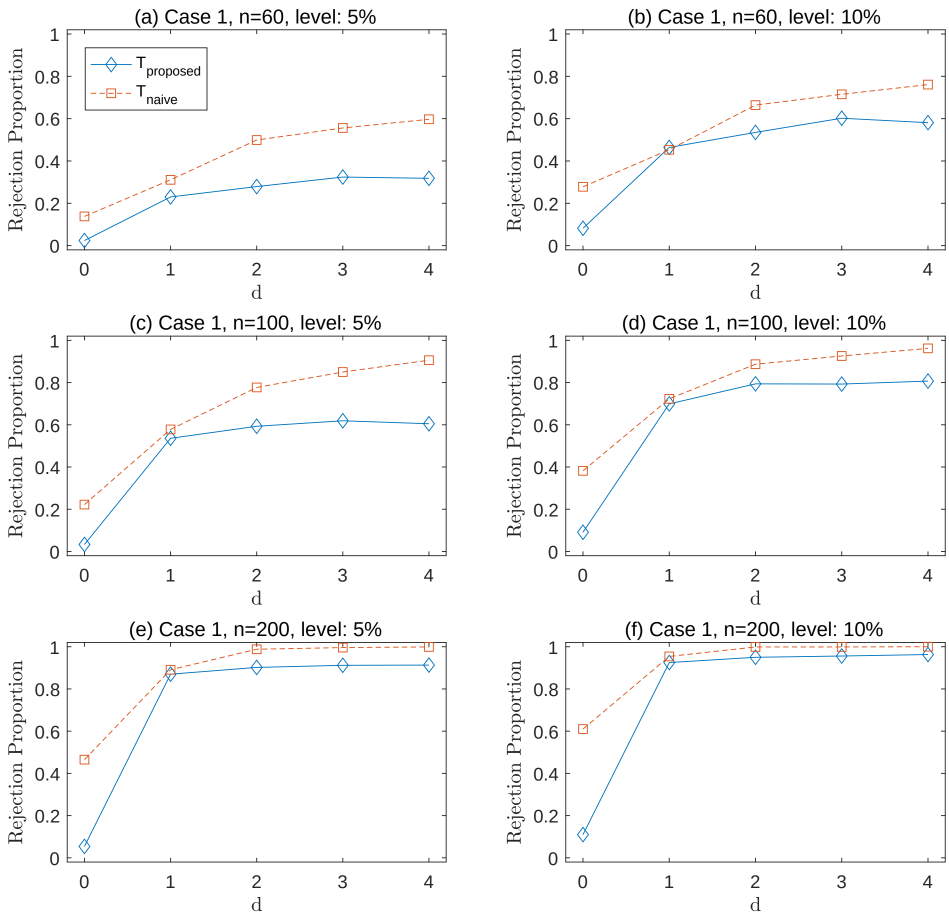

Figure 1: Plots of rejection frequency for Case 1 in Example 1 under different sample sizes and test levels 0.05 and 0.1. $T_{proposed}$: the proposed test; T_{naive}: the naive method.

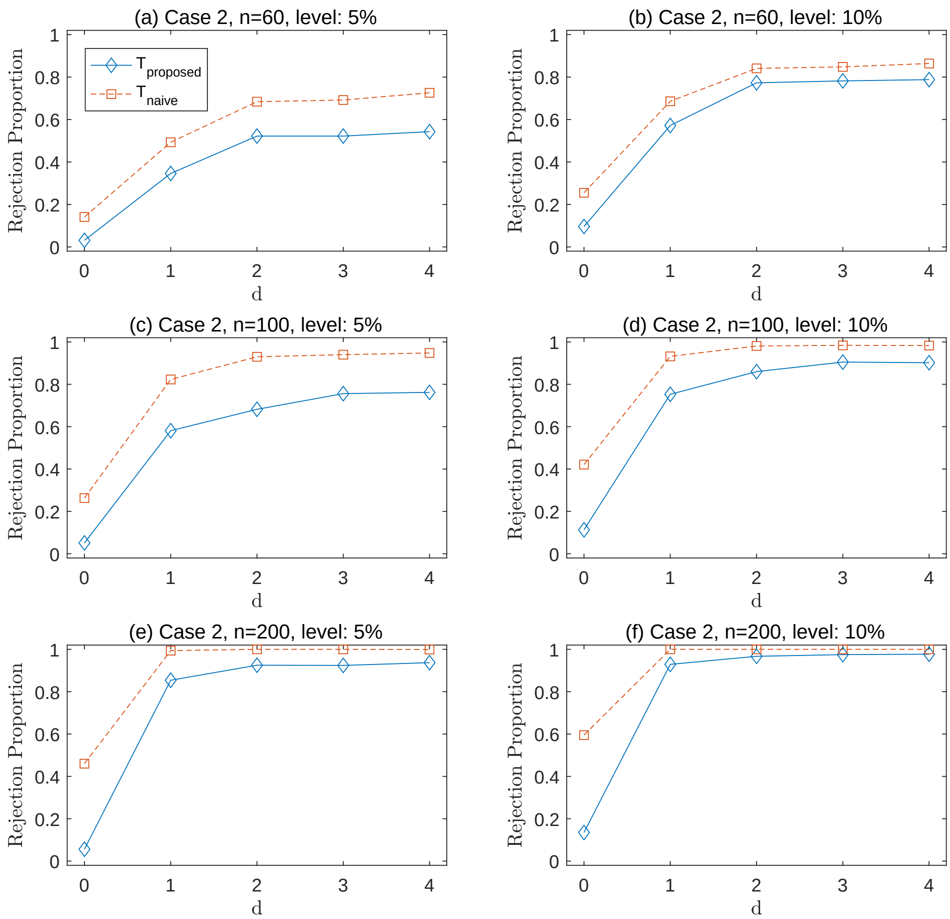

Figure 2: Plots of rejection frequency for Case 2 in Example 1 under different sample sizes and test levels 0.05 and 0.1. $T_{proposed}$: the proposed test; T_{naive}: the naive method.

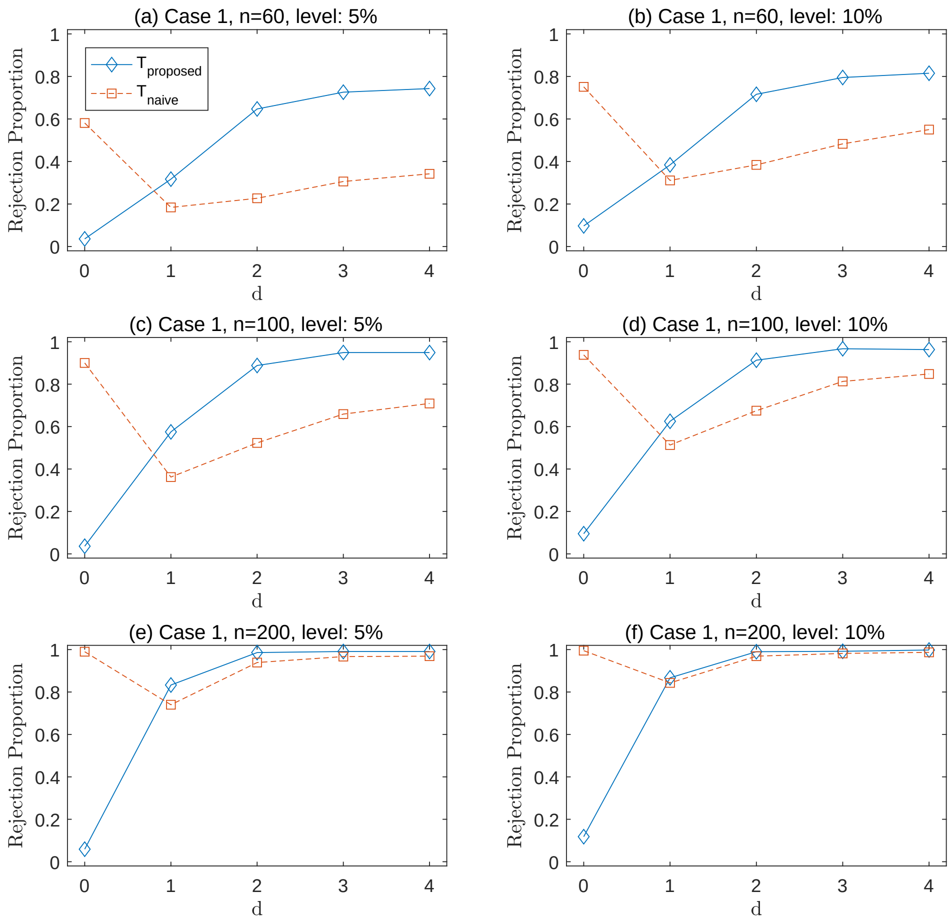

Figure 3: Plots of rejection frequency for Case 1 in Example 2 under different sample sizes and test levels 0.05 and 0.1. $T_{proposed}$: the proposed test; T_{naive}: the naive method.

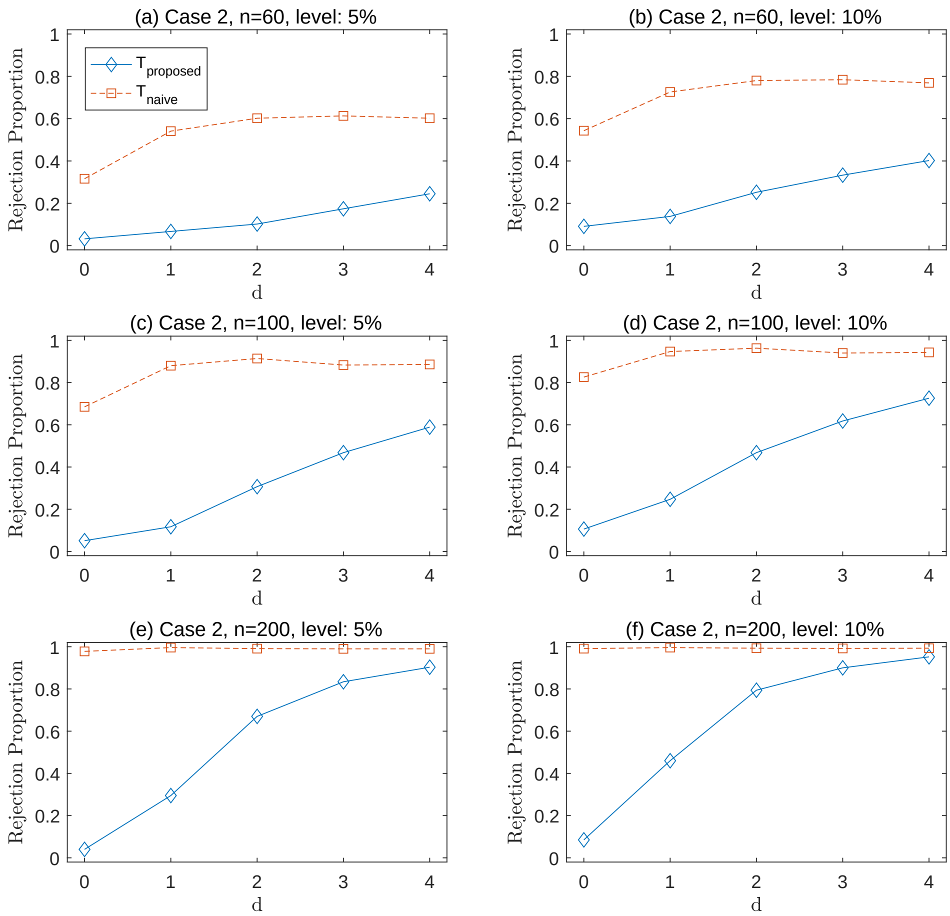

Figure 4: Plots of rejection frequency for Case 2 in Example 2 under different sample sizes and test levels 0.05 and 0.1. $T_{proposed}$: the proposed test; T_{naive}: the naive method.

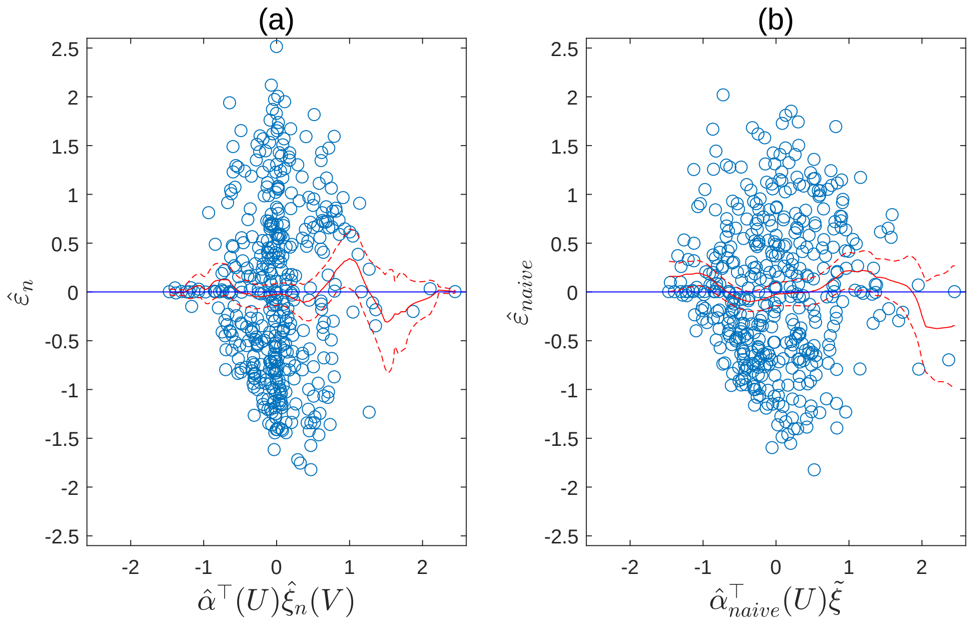

Figure 5: (a) Scatter plot of the calibrated model error estimator $\hat{\varepsilon}_n$ versus $\hat{\alpha}^\top(U)\hat{\xi}_n(V)$ in Example 3; (b) Scatter plot of the naive model error estimator $\hat{\varepsilon}_{naive}$ versus $\hat{\alpha}^\top_{naive}(U)\tilde{\xi}$ in Example 3.

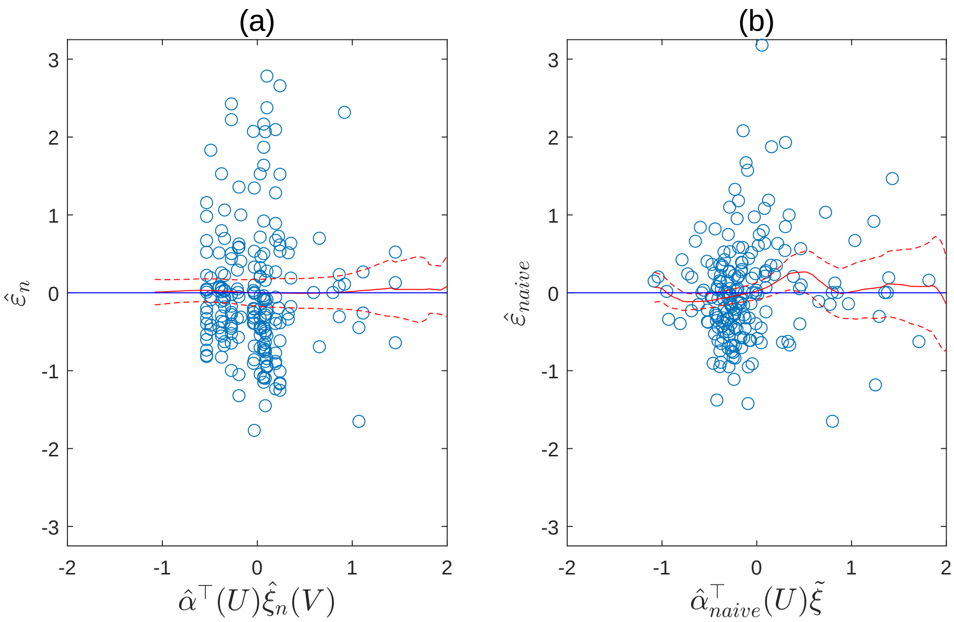

Figure 6: (a) Scatter plot of the calibrated model error estimator $\hat{\varepsilon}_n$ versus $\hat{\alpha}^\top(U)\hat{\xi}_n(V)$ in Example 4; (b) Scatter plot of the naive model error estimator $\hat{\varepsilon}_{naive}$ versus $\hat{\alpha}^\top_{naive}(U)\tilde{\xi}$ in Example 4.