Goodness-of-fit Test for Partial Functional Linear Model with Errors in Scalar Covariates

Tong Zhang[1,2], Zhihua Sun[1†] and Liuquan Sun[1,2]

[1] *University of Chinese Academy of Sciences, Beijing, 100049, P.R.China, sunzh@amss.ac.cn*

[2] *Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P.R.China,*

zt001062@mail.ustc.edu.cn slq@amt.ac.cn

Abstract. In this article, we study the adequacy test of the partial functional linear model when the scalar predictor is measured with additive errors. Based on a corrected profile least-squares estimation of the null model, estimated residuals are first constructed, and a U-statistic-based test is proposed. The asymptotic properties of the test are investigated under the null and the alternative hypotheses. It is shown that the proposed test can control the type I error well, and its power performance is satisfactory. The finite sample performance of the proposed test is demonstrated through simulation studies and a real data analysis.

Keywords: Additive measurement error; Functional covariate; Goodness-of-fit test; Partial functional linear model; U-statistic-based test.

† *Corresponding author.*

1. Introduction

In many fields, emerging data often has the characteristic of trajectory, and is called functional data. Horváth and Reeder (2013) presented several typical examples of functional data. There has been a vast literature in the field of functional data analysis (FDA). Most of the articles focus on the estimation of the functional regression models. Morris (2015) classified functional regression models into three categories: scalar responses and functional predictors; functional responses and scalar predictors; and functional responses and functional predictors. Reiss et al. (2017) presented a comprehensive overview of the first type regression models. For more details, we refer readers to Ferraty and Vieu (2007b); Ramsay and Silverman (2007); Hsing and Eubank (2015), and the references therein.

It is well known that statistical inferences often depend severely on correctly-specified models. If the candidate models are not adequate, the subsequent statistical analysis results tend to be misleading. Therefore, it is generally well recognized that the test of the candidate model's suitability is a fundamental and important topic. In recent years, the adequacy test of regression models with functional covariate has attracted increasing attention. The earliest literature that focused on adequacy test of regression models with functional covariate can be traced back to Cardot et al. (2003, 2004). Subsequently, the classic local test methods in Härdle et al. (1998) and Dette (1999) are applied to test the suitability of the functional linear model; see Delsol et al. (2011) and Bücher et al. (2011). The global projection-based empirical process test in Escanciano (2006) is also extended to test the adequacy of the functional linear models (García-Portugués et al., 2014). Recently, Cuesta-Albertos et al. (2019) and Patilea and Sánchez-Sellero (2020) extended the residual marked empirical process test in Stute (1997) and the U-statistic type test in Zheng (1996) to the functional data context, respectively. Besides, Hilgert et al. (2013) developed a minimax adaptive test for the functional linear model. However, all the above-mentioned works are limited to regression models with data observed accurately.

In many applications, variables often inevitably suffer from measurement errors owing to some causes (e.g. instrument imprecision, laboratory uncertainties, human inconsistencies). It is noted that measurement errors always cause misleading results, such as biased estimators and loss of efficiency. There is a large body of literature on how to eliminate the adverse effects of measurement errors. Sepanski and Lee (1995) and Tekwe et al. (2019) employed a validation data set and an instrumental variable to deal with measurement errors, respectively; Sun et al. (2006) and Wang et al. (2020) developed nonparametric-correction approaches to remove the disruptive influence originated from measurement errors when repeated measurements or observations exist. However, as far as we know, there are only sporadic existing studies on the treatment of functional regression models with measurement errors. Radchenko et al. (2015) and Ferraty et al. (2019) developed estimating procedures for a functional single index model and a nonparametric regression model when functional predictors suffer from measurement errors. Zhu et al. (2019b) studied the partial functional linear model with real-valued covariates measured with additive errors. Zhu et al. (2019a) addressed the estimation of the semi-functional linear regression models when both functional- and real-valued random variables are measured with additive errors.

In this article, for the candidate partial functional linear model with covariates measured with additive errors, the model suitability is investigated with the aid of a U-statistic-based test, a basic introduction about U-statistic theories and U-statistic-based test can be found in Zheng (1996) and reference therein. However, the complexity of constructing an appropriate test is doubled for the considered functional setting due to the existence of the functional predictor and the variable with measurement error. The test statistic is built by combining local kernel smoothing method for both the scalar and the functional predictors. To the best of our knowledge, the present study is the first to apply this local kernel method related to both scalar and functional predictors to test the model adequacy.

Model adequacy test with measurement error has been considered challenging because of the main difficulties in calculating the likelihood or forming residuals. See Ma et al. (2011) and Wang et al. (2020). In this study, we employ corrected profile least-squares method to yield asymptotic unbiased estimators of the regression parameters. To eliminate the adverse effects due to measurement error, a U-statistic-based test statistic is constructed through a kernel weighting technique only related to the exactly-observed scalar predictor and the functional variable. The developed test is free from complex deconvolution operations or minimum distance procedure, and therefore possesses the advantage of computational expedience.

The asymptotic properties of the test statistic under the null and the alternative hypothetical models are rigorously investigated. We validate that the proposed test can control the type I error well, along with satisfactory power performance. Numerical simulation and case analysis results show that the proposed method can effectively eliminate the adverse effects of measurement errors.

The remainder of this article is organized as follows. In Section 2, the model and test problem is introduced. In Section 3, an estimation procedure of the null hypothetical model is presented, and an intuitive explanation why the naive method fails is also listed. A corrected test procedure is proposed in Section 4. The asymptotic properties of the test and the determination of the critical value are presented in Sections 5 and 6, respectively. Some simulation studies and a real data analysis are conducted in Sections 7 and 8. Some assumptions and technical proofs of the main results are collected in the Appendixes.

2. Model and hypotheses

2.1. Model introduction

Consider the partial functional linear model of the following form:

$$Y = \mathbf{Z}^\top \beta + \int_{\mathfrak{T}} \alpha(t) X(t) dt + \varepsilon, \tag{1}$$

where Y is a scalar response, $\mathbf{Z}$ is a p-dimensional variable, and β is an unknown parameter vector. Both the functional covariate $X(\cdot)$ and the unknown slope function $\alpha(\cdot)$ are smooth and square integrable functions defined on a bounded closed interval $\mathfrak{T} \subset \mathbb{R}$. Without loss of generality, the interval $\mathfrak{T}$ is assumed to be $[0, 1]$. The variable ε is the random model error with zero mean and finite variance σ^2.

Due to the existence of the measurement errors, the covariate $\mathbf{Z}$ is unavailable. Instead, a surrogate variable $\mathbf{W}$ is observed. An additive measurement error structure $\mathbf{W} = \mathbf{Z}+\mathbf{U}$ is assumed with a measurement error variable $\mathbf{U}$ of zero mean and covariance matrix Σ_u. Temporarily, Σ_u is assumed to be known, and the case of unknown Σ_u will be discussed later. Furthermore, suppose that the measurement errors are caused by random rather than systematic factors, and the measurement error variable $\mathbf{U}$ is independent of the variables Y, $\mathbf{Z}$, and X.

2.2. The hypotheses

Estimation of model (1) has been addressed in Zhu et al. (2019b) via employing a corrected profile least-squares estimating procedure. When the method in Zhu et al. (2019b) is used for data analysis, a fundamental and important question is to make sure that the model is appropriate. However, there is no existing method to test the suitability of the candidate model in (1). This study aims for filling this gap by developing a U-statistic-based test procedure.

To be specific, the test problem is stated as

$$\mathcal{H}_0 : \Pr\left\{\mathbb{E}(Y|\mathbf{Z}, X) = \mathbf{Z}^\top \beta + \int_0^1 \alpha(t) X(t) dt\right\} = 1, \tag{2}$$

for some $\beta \in \mathbb{R}^p$ and some slope function $\alpha \in \mathbf{L}^2[0, 1]$. Here, $\mathbf{L}^2[0, 1]$ is a functional space composed of all functions $f : [0,1] \to \mathbb{R}$ satisfying $\int_0^1 f^2(t)dt < \infty$, with the norm $\|f\| = \langle f, f\rangle^{1/2}$ and the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$. Instead of directional alternative hypotheses, an omnibus alternative hypothesis is considered:

$$\mathcal{H}_1 : \Pr\left\{\mathbb{E}(Y|\mathbf{Z}, X) = \mathbf{Z}^\top\beta + \int_0^1 \alpha(t)X(t)dt\right\} < 1, \tag{3}$$

for all $\beta \in \mathbb{R}^p$ and $\alpha \in \mathbf{L}^2[0,1]$.

3. Two related preliminaries

3.1. Estimation of the null hypothetical model

Suppose that there is a random sample $\{(Y_i, \mathbf{W}_i, X_i), i = 1, \ldots, n\}$ from $(Y, \mathbf{W}, X)$. Since estimated model errors are often indispensable for a model diagnostic procedure, we first present an estimating method of the null hypothetical model.

Let $\gamma(s,t) = \mathtt{cov}(X(s), X(t))$ be the covariance operator of the stochastic process X. By Mercer's theorem, the spectral decomposition of $\gamma(s,t)$ yields $\gamma(s,t) = \sum_{j=1}^\infty \lambda_j\psi_j(s)\psi_j(t)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are eigenvalues of $\gamma(s,t)$, and $\psi_1, \psi_2, \ldots,$ are eigenfunctions, which constitute an orthogonal basis of the function space $\mathbf{L}^2[0,1]$. Denote the empirical counterparts of $\gamma(s,t)$, λ_j and ψ_j by $\hat{\gamma}$, $\hat{\lambda}_j$ and $\hat{\psi}_j$, for $j = 1, 2, \ldots$.

Let $Y = (Y_1, \ldots, Y_n)^\top$ and $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_n)^\top$. Denote the identity matrix by $\mathbf{I}$. From Zhu et al. (2019b), the estimators of the regression parameter β and the slope function $\alpha(t)$ are given by

$$\hat{\beta}_n = \{\mathbf{W}^\top(\mathbf{I}-\mathbf{S})\mathbf{W} - n\Sigma_u\}^{-1}\mathbf{W}^\top(\mathbf{I}-\mathbf{S})Y, \tag{4}$$

and

$$\hat{\alpha}_n(t) = \sum_{i=1}^m \tilde{\alpha}_i\hat{\psi}_i(t),$$

where $\mathbf{S} = \mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top$, $\tilde{\alpha} = (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top(Y - \mathbf{W}\hat{\beta}_n) = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_m)^\top$, $\mathbf{H} = (\hat{s}_{i,j})_{1\leq i\leq n, 1\leq j\leq m}$, and $\hat{s}_{i,j} = \langle X_i, \hat{\psi}_j\rangle$ for $1 \leq i \leq n$, $1 \leq j \leq m$. Herein, the cutoff parameter m is the number of the highest eigenvalues/eigenfunctions selected, and controls the smoothness degree of the estimator $\hat{\alpha}_n(t)$.

The above estimating procedure offers an advantage that the estimators have closed forms. Thus, even if the data come from the alternative model, the estimated procedure and the follow-up test procedure can be carried out.

3.2. Intuitive explanation of the failure of the naive method

The method taking errors-in-variables as true variables is called a naive method. Unfortunately, for the considered test problem in (2), the naive method fails.

Denote the estimated naive model error by $\hat{\varepsilon}_{ni}^{NA} = Y_i - \mathbf{W}_i^\top \hat{\beta}_n^{NA} - \int_0^1 \hat{\alpha}_n^{NA}(t) X_i(t)dt$ for $i = 1, \ldots, n$, where $\hat{\beta}_n^{NA}$ and $\hat{\alpha}_n^{NA}(t)$ are the naive estimators of β and $\alpha(t)$, respectively. Similarly to the proofs of Theorems 1 and 2 of Zhu et al. (2019b), by some lengthy derivation, it can be deduced that $\hat{\beta}_n^{NA} - \beta$ and $\hat{\alpha}_n^{NA}(t) - \alpha(t)$ have nonzero asymptotic means. The estimated naive error can be decomposed into four parts:

$$\begin{aligned}
\hat{\varepsilon}_{ni}^{NA} = & \left\{ Y_i - \mathbf{Z}_i^\top \beta - \int_0^1 \alpha(t) X_i(t) dt \right\} - \mathbf{U}_i^\top \hat{\beta}_n^{NA} \\
& + \int_0^1 \{\alpha(t) - \hat{\alpha}_n^{NA}(t)\} X_i(t) dt + \mathbf{Z}_i^\top (\beta - \hat{\beta}_n^{NA}), \quad i = 1, \ldots, n.
\end{aligned}$$

Under the null hypothesis, the expectations of the first two terms converge to zero. But the last two terms and therefore the naive error $\hat{\varepsilon}_{ni}^{NA}$ have nonzero asymptotic means because the naive estimators $\hat{\beta}_n^{NA}$ and $\hat{\alpha}_n^{NA}(t)$ are asymptotically biased.

The estimated naive errors tend to keep away from zero under both the null and the alternative hypotheses. Therefore, the naive test cannot distinguish the null hypothetical model from the alternative ones. For the considered test problem, it is imperative to develop a corrected test procedure.

4. A corrected test procedure

Let $\varepsilon = Y - \mathbf{Z}^\top \beta - \int_0^1 \alpha(t)X(t)dt$, and $\mathbf{V} = (\mathbf{Z}^\top, X)^\top$. The corrected model error is defined as $\hat{\varepsilon}_{ni} = Y_i - \mathbf{W}_i^\top \hat{\beta}_n - \int_0^1 \hat{\alpha}_n(t)X_i(t)dt,\ i = 1, \ldots, n$, with $\hat{\beta}_n$ and $\hat{\alpha}_n(t)$ given in Section 3. Further, denote $\check{K}_h(\mathbf{V}_i, \mathbf{V}_j) = k_{h_0}(d(X_i, X_j)) \prod_{l=1}^{p} k_{h_l}(\mathbf{Z}_{il} - \mathbf{Z}_{jl})$ for $i, j = 1, \ldots, n$. Herein, $k_{h_l}(\cdot)$ is defined as $k(\cdot/h_l)$ with a univariate kernel function k and bandwidths $h_l, l = 0, 1, \ldots, p$. The metric $d(X_1, X_2) = \|X_1 - X_2\|$ is defined on $\mathbf{L}^2[0, 1]$.

The test methods arising from the estimated model errors are broadly categorized into two groups: local tests requiring nonparametric local smoothing techniques and global tests based on the weighted averages of residuals. The U-statistic type test and the empirical process test are representative methods of the local and the global methods, respectively. It is generally agreed that the U-statistic type test and the empirical process test complement each other (Ma et al., 2014). A comparison of the advantages and the disadvantages of these two methods can be found in Guo and Zhu (2017). Sun et al. (2021) pointed that the U-statistic type test and the empirical process test are special cases of the ICM test with different weighting functions.

Following the line of Li and Wang (1998), we aim for developing a U-statistic type test for the considered hypothetical problem in (2). In general cases where ε is a model error and $\mathbf{V}$ is a predictor, the U-statistic type test originates from the equivalence of $\mathbb{E}[\varepsilon|\mathbf{V}] = 0$ and $\mathbb{E}[\varepsilon\mathbb{E}(\varepsilon|\mathbf{V})\omega(\mathbf{V})] = 0$ for some positive measurable weight function $w(v)$. Motivated by an estimator of $\mathbb{E}[\varepsilon\mathbb{E}(\varepsilon|\mathbf{V})\omega(\mathbf{V})]$, define $\mathcal{T}_n^v$ as follows:

$$\mathcal{T}_n^v = \frac{1}{n(n-1)} \frac{1}{\tilde{\varphi}(h)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \check{K}_h(\mathbf{V}_i, \mathbf{V}_j)\hat{\varepsilon}_{ni}\hat{\varepsilon}_{nj}, \tag{5}$$

where $\hat{\varepsilon}_{ni}$ is defined above. The function $\tilde{\varphi}(h)$ is equal to $\psi(h_0) \prod_{l=1}^{p} h_l$, where the positive function $\psi(\cdot)$ is given in Assumptions (C3) in Appendix A. However, $\mathcal{T}_n^v$ is not calculable, because the variable $\mathbf{V}$ is unavailable.

A natural idea is replacing the true variable $\mathbf{V} = (\mathbf{Z}^\top, X)^\top$ in (5) with the surrogate variable $\mathbf{V}^w = (\mathbf{W}^\top, X)^\top$. Hence, we obtain a calculable statistic:

$$\mathcal{T}_n^w = \frac{1}{n(n-1)} \frac{1}{\tilde{\varphi}(h)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \check{K}_h(\mathbf{V}_i^w, \mathbf{V}_j^w) \hat{\varepsilon}_{ni} \hat{\varepsilon}_{nj}.$$

The statistic $\mathcal{T}_n^w$ is a consistent estimator of $\mathbb{E}[\varepsilon \mathbb{E}(\varepsilon|\mathbf{V}^w)\omega(\mathbf{V}^w)]$ with some appropriate weight function. Observe that $\mathbb{E}[\varepsilon \mathbb{E}(\varepsilon|\mathbf{V}^w)\omega(\mathbf{V}^w)] = \mathbb{E}[\mathbb{E}^2(\varepsilon|\mathbf{V}^w)\omega(\mathbf{V}^w)]$. Under the null hypothesis, we cannot judge whether $\mathbb{E}[\varepsilon|\mathbf{V}^w]$, along with $\mathbb{E}[\varepsilon \mathbb{E}(\varepsilon|\mathbf{V}^w)\omega(\mathbf{V}^w)]$, is zero or not. Thus, the test $\mathcal{T}_n^w$ cannot control the type I error.

In practice, measurement errors are commonly known to occur in only one or several variables. Suppose that the first q $(0 \leq q \leq p)$ components in $\mathbf{Z}$ are observed accurately. Denote $\tilde{\mathbf{V}} = (\tilde{\mathbf{Z}}^\top, X)^\top$ with $\tilde{\mathbf{Z}} = (Z_1, \ldots, Z_q)^\top$. Under the null hypothesis, it is true that $\mathbb{E}[\varepsilon|\tilde{\mathbf{V}}] = 0$, which yields $\mathbb{E}[\varepsilon \mathbb{E}(\varepsilon|\tilde{\mathbf{V}})\omega(\tilde{\mathbf{V}})] = 0$. Intuitively, the test arising from $\mathbb{E}[\varepsilon|\tilde{\mathbf{V}}]$ should be able to control the type I error.

Based on the above considerations, defined $\mathcal{T}_n$ as follows:

$$\mathcal{T}_n = \frac{1}{n(n-1)} \frac{1}{\varphi(h)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j) \hat{\varepsilon}_{ni} \hat{\varepsilon}_{nj},$$

where $\varphi(h) = \psi(h_0) \prod_{l=1}^{q} h_l$, $K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j) = k_{h_0}(d(X_i, X_j)) \prod_{l=1}^{q} k_{h_l}(\mathbf{Z}_{il} - \mathbf{Z}_{jl})$ for $i, j = 1, \ldots, n$. The final test statistic is defined as

$$n\varphi(h)^{1/2}\mathcal{T}_n = \frac{1}{n-1} \frac{1}{\varphi(h)^{1/2}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j) \hat{\varepsilon}_{ni} \hat{\varepsilon}_{nj}.$$

The null hypothesis should be rejected when $n\varphi(h)^{1/2}\mathcal{T}_n$ is large enough.

If the weight function $\omega(\tilde{\mathbf{V}}_i)$ is taken as $\frac{1}{n-1} \sum_{j=1, j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)$, $i = 1, \ldots, n$, then the test statistic is actually the empirical estimator of $\mathbb{E}[\varepsilon \mathbb{E}(\varepsilon|\tilde{\mathbf{V}})\omega(\tilde{\mathbf{V}})]$ multiplied by a regularization factor $n\varphi(h)^{-1/2}$. Denote

$$\hat{\mathbb{E}}(\varepsilon_{ni}|\tilde{\mathbf{V}}_i) = \frac{\sum_{j=1, j\neq i}^{n} \hat{\varepsilon}_{nj} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)}{\sum_{j=1, j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)}.$$

To be specific, we can validate that

$$\begin{aligned}\hat{\mathbb{E}}[\varepsilon\mathbb{E}(\varepsilon|\tilde{\mathbf{V}})\omega(\tilde{\mathbf{V}})] &=: \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_{ni}\hat{\mathbb{E}}(\varepsilon_{ni}|\tilde{\mathbf{V}}_i)\omega(\tilde{\mathbf{V}}_i)\\ &= \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_{ni}\frac{\sum_{j=1,j\neq i}^{n}\hat{\varepsilon}_{nj}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)}{\sum_{j=1,j\neq i}^{n}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)}\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\hat{\varepsilon}_{ni}\hat{\varepsilon}_{nj}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j),\end{aligned}$$

and thus,

$$n\varphi(h)^{-1/2}\hat{\mathbb{E}}[\varepsilon\mathbb{E}(\varepsilon|\tilde{\mathbf{V}})\omega(\tilde{\mathbf{V}})] = \frac{1}{n-1}\frac{1}{\varphi(h)^{1/2}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\hat{\varepsilon}_{ni}\hat{\varepsilon}_{nj} = n\varphi(h)^{1/2}\mathcal{T}_n.$$

This test statistic is of a relatively simple form, and the case of zero-denominator is effectively avoided. This technique is commonly used in the local test methods (Li and Wang, 1998).

5. Asymptotic properties

For $t \in \mathbb{R}^+$, denote the small ball centered at X with radius t by $\mathcal{B}(X,t) = \{X' \in \mathbf{L}^2[0,1] : d(X,X') \leq t\}$. The small ball probability is defined as $p(t) = \Pr\{X' \in \mathcal{B}(X,t)\}$.

Theorem 1. *Under Assumptions (C1)–(C6) in Appendix A, when the null hypothesis (2) holds, we have*

$$n\varphi(h)^{1/2}\mathcal{T}_n \xrightarrow{d} \mathcal{N}(0,\sigma_0^2),$$

with σ_0^2 presented in Appendix B.

In the following, consider the local alternative hypotheses:

$$\mathcal{H}_{1n} : Y = \mathbf{Z}^\top\beta + \int_0^1 \alpha(t)X(t)dt + \delta_n\mathcal{D}(\mathbf{Z},X) + \varepsilon, \tag{6}$$

where the deviation function $\mathcal{D}(\cdot,\cdot)$ satisfies

$$\inf_{\beta\in\mathbb{R}^p,\alpha(t)\in\mathbf{L}^2[0,1]} \mathbb{E}\left\{\mathcal{D}(\mathbf{Z},X) - \mathbf{Z}^\top\beta - \int_0^1 \alpha(t)X(t)dt\right\}^2 > 0,$$

and δ_n is a constant.

Theorem 2. *Under the assumptions of Theorem 1, when the local alternative hypothesis (6) with* $\delta_n = n^{-1/2}\varphi(h)^{-1/4}$ *holds, we have*

$$n\varphi(h)^{1/2}\mathcal{T}_n \xrightarrow{d} \mathcal{N}(\Lambda, \sigma_0^2),$$

with Λ *and* σ_0^2 *presented in Appendix B.*

Theorem 2 shows that the proposed test has nonignorable power for the alternative hypothesis (6). Even if the predictors suffer from measurement errors, the proposed test has the same sensitivity as the method for accurately observed data. In the following, we investigate the convergence property of the test under the fixed alternative hypothesis in (3).

Corollary 1. *Under Assumptions (C1)–(C6) in Appendix A, if the alternative hypothesis (3) holds, then* $n\varphi^{1/2}(h)\mathcal{T}_n \to \infty$*, as* $n \to \infty$*.*

The existing model testing methods for measurement error data often suffer from inconsistency (Wang et al., 2020). Theorem 1 suggests that the proposed test has an asymptotic power one, and hence is consistent under some weak conditions.

6. Determination of the critical value

From the asymptotic normality in Theorem 1, the upper quantile is a possible choice to determine the critical value. Since it is difficult to estimate σ_0^2 directly, we provide a bootstrap procedure to determine the critical value. Recall that $\hat{\varepsilon}_{ni} = Y_i - \mathbf{W}_i^\top \hat{\beta}_n - \int_0^1 \hat{\alpha}_n(t)X_i(t)dt$, $i = 1, \ldots, n$, with $\hat{\beta}_n$ and $\hat{\alpha}_n(t)$ given in Section 3. The details are listed below.

Step 1: Generate an i.i.d. random variable sequence $\{e_1, \ldots, e_n\}$ with mean zero, variance one and a finite third moment. Let $Y_i^* = \mathbf{W}_i^\top \hat{\beta}_n + \int_0^1 \hat{\alpha}_n(t)X_i(t)dt + e_i\hat{\varepsilon}_{ni}$, $i =$

$1, \ldots, n$.

Step 2: Based on the bootstrap sample $\{(Y_i^*, \mathbf{W}_i, X_i), i = 1, \ldots, n\}$, calculate the test statistic $\mathcal{T}_n$, denoted by $\mathcal{T}_n^*$.

Step 3: Repeat **Step 2** ρ times and obtain $\{\mathcal{T}_{n1}^*, \ldots, \mathcal{T}_{n\rho}^*\}$. Calculate the 1-$\alpha$ empirical quantile based on $\{\mathcal{T}_{n1}^*, \ldots, \mathcal{T}_{n\rho}^*\}$, and take it as the α-level critical value.

The random variable sequence $\{e_i, i = 1, \ldots, n\}$ can be sampled from a Bernoulli distribution: $e_i = (1+\sqrt{5})/2$ corresponding to probability $(5-\sqrt{5})/10$, and $e_i = (1-\sqrt{5})/2$ corresponding to probability $(5+\sqrt{5})/10$. The choice of the distribution of e_i is suggested by Enno (1993) and Li and Wang (1998), and the condition of finite third moment is assumed to improve the rate of convergence of the bootstrap estimate (Liu, 1988; Härdle and Mammen, 1990). The above wild bootstrap method is widely applied in the model testing field (Escanciano, 2006; Wang et al., 2020).

7. Simulation studies

Simulation studies are implemented to evaluate the performance of the proposed test. Three data generating processes are considered. In Setting 1, we consider the adequacy test problem for a small-dimensional candidate model with all components of the scalar covariates error-contaminated. In Settings 2 and 3, two moderate-dimensional candidate models are taken into account, where at least one scalar covariate is error-free.

Setting 1. A two-dimensional predictor $\mathbf{Z} = (Z_1, Z_2)^\top$ is generated as follows: $Z_1 \sim \mathcal{N}(-1, 1)$, and $Z_2 \sim \mathcal{U}(0, \pi)$. The regression coefficient β is set to be $(-1.5, 1)^\top$. The measurement error variable $\mathbf{U}$ follows a two-dimensional normal distribution $\mathcal{N}((0, 0)^\top, \Sigma_u)$. The covariance matrix Σ_u is set to be Σ_1=diag$(0.8, 0.8)$ and Σ_2=diag$(0.4, 0.4)$. Let the functional coefficient $\alpha(t)$ be $\sqrt{2}\sin(0.5\pi t) + 3\sqrt{2}\sin(1.5\pi t)$. For the deviation function, we consider two cases: $\mathcal{D}(\mathbf{Z}, X) = 3\|X\|^2$ in Case 1, and $\mathcal{D}(\mathbf{Z}, X) = 1.5|Z_1 - Z_2| \cdot \|X\|$ in

Case 2.

Setting 2. A four-dimensional predictor $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)^\top$ is considered: Z_1 and Z_3 follow $\mathcal{N}(1,1)$; Z_2 and Z_4 follow $\mathcal{U}(0,\pi)$. The regression coefficient β is chosen to be $(1, 0.8, 0.5, 0.3)^\top$. The measurement error variable $\mathbf{U}$ follows $\mathcal{N}((0,0,0,0)^\top, \Sigma_u)$. We consider two choices for Σ_u: Σ_1=diag$(0.8, 0.8, 0.8, 0)$ and Σ_2=diag$(0.4, 0.4, 0.4, 0)$. The functional coefficient $\alpha(t)$ is set to be $2\sqrt{2}\sin(0.5\pi t) - \sqrt{2}\sin(1.5\pi t)$. For the deviation function, we consider two cases: $\mathcal{D}(\mathbf{Z}, X) = 1.2\big(\exp(\|X\|) - 1\big)$ in Case 1, and $\mathcal{D}(\mathbf{Z}, X) = 2\exp(|Z_4 - 1|)/3$ in Case 2.

Setting 3. We consider an eight-dimensional predictor $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_8)^\top$. The first four components follow $\mathcal{N}(1,1)$, and the remaining four components follow $\mathcal{U}(0,\pi)$. The regression coefficient β is chosen to be $(0.8, 0.6, 0.4, 0.2, 0.8, 0.6, 0.4, 0.2)^\top$. The measurement error variable $\mathbf{U}$ is of zero mean and covariance matrix Σ_u. We consider two choices for Σ_u: Σ_1=diag$(0.8, \ldots, 0.8, 0, 0)$ and Σ_2=diag$(0.4, \ldots, 0.4, 0, 0)$. The functional coefficient $\alpha(t)$ is set to be $\sqrt{2}\sin(0.5\pi t) + \sqrt{2}\sin(1.5\pi t)$. For the deviation function, we consider two cases: $\mathcal{D}(\mathbf{Z}, X) = 4\log(1 + \|X\|^2)$ in Case 1, and $\mathcal{D}(\mathbf{Z}, X) = Z_8^3/4$ in Case 2.

In all the settings above, the model error ε is generated from $\mathcal{N}(0, 0.5^2)$. The functional predictor X is chosen to be $\sum_{l=1}^{50} g_l \psi_l(t)$, where $\psi_l(t) = \sqrt{2}\sin((l-0.5)\pi t)$ and $g_l \sim \mathcal{N}(0, ((j-0.5)\pi)^{-2})$ for $l = 1, \ldots, 50$. The coefficient of the deviation function δ_n is set to be 0, 1, 1.5, 2 and 2.5. When $\delta_n = 0$, the null hypothetical model $Y = \mathbf{Z}^\top\beta + \int_0^1 \alpha(t)X(t)dt + \varepsilon$ holds. The nonzero different values of δ_n correspond to different alternative hypothetical models. We fix the sample size n to be 100 and 200, and take the test levels to be 0.05 and 0.10. The bootstrap resample size ρ is set to be 300. To calculate the integral, the interval $[0,1]$ is divided into 100 sub-intervals of equal length. As suggested by a reviewer, the kernel function is chosen to be the Epanechnikov function in Setting 1 and the standard normal density function in the last two settings, which shows that the results are not very sensitive to the choice of the kernel function.

Estimation of Σ_u. The covariance matrix of the measurement error variable Σ_u is often unknown. An estimator based on the repeated measurement data is provided below:

$$\hat{\Sigma}_u = \frac{1}{n(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} (\mathbf{W}_{ij} - \bar{\mathbf{W}}_i)(\mathbf{W}_{ij} - \bar{\mathbf{W}}_i)^\top,$$

where $\mathbf{W}_{i1}, \ldots, \mathbf{W}_{ik}$ denote k repeated observations of $\mathbf{W}_i$, and $\bar{\mathbf{W}}_i = k^{-1} \sum_{j=1}^{k} \mathbf{W}_{ij}$ for $i = 1, \ldots, n$. The number of repeated observations k is set to be 3. The above technique can also be found in Wang et al. (2020).

Choice of m. Note that a cutoff parameter m is brought into the estimating procedure in Section 3. We determine this nuisance parameter m via minimizing the following leave-one-out cross-validation (CV) function:

$$\mathbf{CV}(m) = \frac{1}{n} \sum_{i=1}^{n} \left[\{Y_i - \mathbf{W}_i^\top \hat{\beta}_n^{(-i)} - \mathbf{H}_i \hat{\alpha}_n^{(-i)}(t)\}^2 - (\hat{\beta}_n^{(-i)})^\top \Sigma_u \hat{\beta}_n^{(-i)} \right],$$

where $\hat{\beta}_n^{(-i)}$ and $\hat{\alpha}_n^{(-i)}(t)$ are respectively the leave-one-out estimators of β and $\alpha(t)$. Here, $\mathbf{H}_i$ denotes the ith row of $\mathbf{H}$ for $i = 1, \ldots, n$. For simplicity, we assume the covariance matrix of the measurement error variable Σ_u to be known in the process of determining m.

For each setting, only the case with the sample size $n = 100$ and $\delta_n = 0$ is considered. To ease the computational burden, as suggested by Zhu et al. (2019b), the chosen cutoff parameter m is used in other cases. Let m take values from 1 to 10 with a fixed interval equal to 1. Based on 1000 repetitions, the mean values of the CV function are calculated. Dot graphs with error bars are presented in Figure 1, where the bars are at ± 1 sd from the mean. Since the objective CV functions reach minima at the point $m = 2$ as shown in Figure 1, we should choose $m = 2$ for all settings following the leave-one-out cross-validation criterion. The result is consistent with the number of sine curves which build the functional predictor, which shows that the leave-one-out cross-validation method is efficient.

Both known and unknown covariance matrices of the measurement error variable are considered. For clarity, the corresponding tests are marked by $\mathcal{T}_n^{KN}$ and $\mathcal{T}_n^{UN}$. The bandwidth is chosen by the rule of thumb. The simulation is performed 500 times for all settings. Empirical sizes ($\delta_n = 0$) and empirical powers ($\delta_n > 0$) of the proposed test are listed in Tables 1–3. For comparisons, the method for accurately observed data and the naive test, denoted by $\mathcal{T}_n^{BE}$ and $\mathcal{T}_n^{NA}$, respectively, are considered. The test $\mathcal{T}_n^{BE}$ serves as the gold standard. The detailed results are presented in Tables 1–3. In each table, the column with $\delta_n = 0$ is highlighted in bold as this is the key point of the tables.

From Tables 1–3, it is observed that the proposed tests $\mathcal{T}_n^{KN}$ and $\mathcal{T}_n^{UN}$ yield the empirical sizes close to the test levels. With the increase of δ_n, the true model deviates from the null hypothetical model farther, and the empirical powers of $\mathcal{T}_n^{KN}$ and $\mathcal{T}_n^{UN}$ tend to be larger. Both $\mathcal{T}_n^{KN}$ and $\mathcal{T}_n^{UN}$ perform a little worse than the benchmark method. This behaviour is expected. These results show that our proposed test performs well even if all components of the scalar covariate are error-contaminated and the candidate model is moderate-dimensional.

The most prominent deficit of the naive method is that it cannot control the type I error. In some cases, the empirical sizes are shown to be 0.466 and 0.470, which are produced in Setting 2, $\Sigma_u = \Sigma_1$, $\alpha = 0.10$, $n = 200$, in Case 1 and Case 2, respectively. A test should be able to control the type I error. Otherwise, this test cannot be employed in practice. It should be noted that the naive method yields higher empirical powers. These results are natural since the estimated naive errors tend to keep away from zero. Because the naive method cannot control the type I error, its higher empirical powers are meaningless.

Overall, our proposed method can control type I error and can detect alternative models effectively. For the considered test problem, it should be a valuable tool.

8. A real data analysis

For illustration purposes, we apply the proposed method to analyze Tecator data set of 215 food individuals. For each individual, the values of moisture, fat and protein contents in percentage are recorded. As for the functional part, 100 channel spectra of absorbance are recorded in the wavelength range 850–1050 nm by the near-infrared transmission (NIT) principle. More details about the Tecator data set can be found at http://lib.stat.cmu.edu/datasets/tecator.

We turn to consider the adequacy test of the following partial functional linear model:

$$Y = \beta_1 Z_1 + \beta_2 Z_2 + \int_{850}^{1050} \alpha(t) X(t) dt + \varepsilon, \tag{7}$$

where Y denotes the fat content in percentage, $\mathbf{Z} = (Z_1, Z_2)^\top$ denotes the content of protein and moisture. The predictor $X(t)$ denotes the spectrometric curve and is regarded as a functional variable. Extra measurement error $\mathbf{U} = (U_1, U_2)^\top$ with zero mean and covariance matrix $\mathrm{diag}(3, 0)$ is added to $\mathbf{Z}$. The number of repeated measurements k is chosen to be 2. This data reprocessing technique is also employed in Zhu et al. (2019b). Similarly to the pretreatment in Boente and Vahnovan (2017), outliers, atypical observations, and observations of identical trajectories are removed from the original data set, and the remaining 181 trajectories are taken into account.

All variables are centered. The cutoff parameter m is chosen to be 3 according to the leave-one-out CV procedure. The bootstrap procedure is repeated 1000 times. The p-values of the tests $\mathcal{T}_n^{KN}$ and $\mathcal{T}_n^{UN}$ are calculated to be 0.869 and 0.675, respectively. Therefore, the candidate model (7) is suitable for this dataset. For comparisons, we also compute the p-value of the naive test $\mathcal{T}_n^{NA}$, which is 0.056.

To provide further graphical evidence, we plot residual plots of the proposed method $\mathcal{T}_n^{KN}$ and $\mathcal{T}_n^{UN}$ in Figure 2. From Figure 2, we observe that for the tests $\mathcal{T}_n^{KN}$ and $\mathcal{T}_n^{UN}$, the estimated model errors are distributed randomly with regard to the abscissa axis and there is no clear indication of the relation between the residuals and $\mathbf{W}^\top \hat{\beta}_n + \int_{850}^{1050} \hat{\alpha}_n(t) X(t) dt$. Therefore, the results drawn from the test methods are consistent with the residual plots.

9. Discussions

In this article, we studied the adequacy test of the partial functional linear model when the scalar predictor was measured with additive measurement error. The estimated residuals were constructed, and a U-statistic-based test was proposed. The asymptotic properties of the test were investigated under the null and the alternative hypotheses. The simulation results showed that the proposed test work well for the situations considered. The proposed method can be borrowed to study the adequacy test of other semiparametric functional regression models with errors-in-variables.

Besides the additive measurement errors, other types of measurement errors, such as distortion measurement errors, also arise frequently. However, relevant studies are highly limited. It should be of great value to investigate the statistical inference problems, for example (but not limited to) the model estimation and test, for functional regression models with other types of measurement errors.

Acknowledgements

The authors thank the Executive Editors, the Associate Editor and the reviewers for their insightful comments and suggestions that greatly improved the article. Zhihua Sun's research was supported by Beijing Natural Science Foundation (Z200001, 1202001) and the National Natural Science Foundation of China (12071457, 11971045). Liuquan Sun's research was partially supported by the National Natural Science Foundation of China (12171463).

Appendix A: Assumptions

(C1) (i) $\|X\|$ is bounded;

(ii) $\mathbb{E}\|\mathbf{Z}\|^4 < \infty$ and $\mathbb{E}\|\mathbf{U}\|^4 < \infty$, where $\|\cdot\|$ is the $\mathbf{L}^2$ norm in $\mathbb{R}^p$.

(C2) (i) $k(\cdot)$ is a univariate kernel function of order two with support $[0,1]$, and its derivative $k'(\cdot)$ is continuous and bounded.

(ii) $h_l \to 0$ $(l = 1, \ldots, q)$ and $n\varphi(h) \to \infty$, as $n \to \infty$.

(C3) (i) There exist a strictly positive and increasing function $\psi(\cdot)$ satisfying $\lim_{t\to 0} \psi(t) = 0$, and a nonnegative bounded function $f_x(\cdot)$ such that $\Pr\{d(x, X) \leq t\} = \psi(t)f_x(x) + o(\psi(t))$, as $t \to 0$.

(ii) There exists a nondecreasing bounded function $\mu_0(t)$ such that uniformly in $t \in [0, 1]$, $\psi(ht)/\psi(h) = \mu_0(t) + o(1)$, as $h \to 0$, and $\int_0^1 (k^j(t))' \mu_0(t)dt < \infty$ for $j \geq 1$.

(C4) (i) There exist some constants $a > 1$ and $b > a/2+1$ such that $C^{-1}k^{-a} \leq \lambda_k \leq Ck^{-a}$, $\lambda_k - \lambda_{k+1} \geq C^{-1}k^{-a-1}$, $k = 1, 2, \ldots$.

(ii) For each k, $|\langle \mathrm{cov}(Z_l, X), \psi_k\rangle| < Ck^{-(a+b)}$ for any $l \in \{1, 2, \ldots, p\}$;

(iii) $\lim_{n\to\infty} m/n^{1/(a+2b)}$ is a nonzero constant.

(iv) $\lim_{n\to\infty} \varphi(h)^{1/2} n^{1-(2b-1)/(a+2b)} = 0$.

(C5) (i) For any $l \in \{1, \ldots, p\}$, $\mathbb{E}[Z_l|X] = \langle g_l, X\rangle$, where $g_l = \sum_{j=1}^{+\infty} \langle \mathrm{cov}(Z_l, X), \psi_j\rangle \psi_j/\lambda_j \in \mathbf{L}^2[0, 1]$.

(ii) Let $\mathcal{Z}_X = (Z_1 - \mathbb{E}[Z_1|X], \ldots, Z_p - \mathbb{E}[Z_p|X])^\top$, and $\mathcal{B} = \mathbb{E}[\mathcal{Z}_X \mathcal{Z}_X^\top]$. Then, $\mathcal{B} = \mathrm{cov}(\mathbf{Z}) - \sum_{j=1}^{+\infty} \langle \mathrm{cov}(\mathbf{Z}, X), \psi_j\rangle \langle \mathrm{cov}(X, \mathbf{Z}), \psi_j\rangle/\lambda_j$ is positive definite.

(C6) $\sup_{u\in B(x,h_0)} |\rho(\tilde{\mathbf{z}}, x) - \rho(\tilde{\mathbf{z}}, u)| \leq Cd(u, x)^\theta$ for some constant $\theta > 0$, and $\rho(\tilde{\mathbf{z}}, x) = \rho(\tilde{\mathbf{v}})$ is one of $\sigma_1(\tilde{\mathbf{v}})$, $\sigma_2(\tilde{\mathbf{v}})$ and $m_w(\tilde{\mathbf{v}})$, where $\sigma_1(\tilde{\mathbf{V}}) = \mathbb{E}[\varepsilon_u^2|\tilde{\mathbf{V}}]$, $\sigma_2(\tilde{\mathbf{V}}) = \mathbb{E}[\varepsilon_u^4|\tilde{\mathbf{V}}]$, $m_w(\tilde{\mathbf{V}}) = \mathbb{E}[\mathbf{W}|\tilde{\mathbf{V}}]$, and $\varepsilon_{ui} = \varepsilon_i - \beta^\top \mathbf{U}_i$, for $i = 1, \ldots, n$.

Remark 1. More details about the above assumptions can be found in Ferraty et al. (2007a), Louani (2010) and Zhu et al. (2019b). Assumption (C4) guarantees that $1/2 < (2b-1)/(a+2b) < 1$. It should be noted that the constant C may be different in different scenarios.

Appendix B: Proofs of Theorems

To prove Theorems 1 and 2, we need the following Lemmas. Lemmas 1 and 2 give the asymptotic properties of $\hat{\beta}_n$ and $\hat{\alpha}_n$ under the null hypothetical model (2) and the alternative hypothetical model (6), respectively.

Lemma 1. *Under the null hypothetical model (2) and assumptions of Theorem 1, we have*

(i) $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \mathcal{B}^{-1}\Gamma\mathcal{B}^{-1})$, *where* $\Gamma = \mathbb{E}[(\varepsilon - \mathbf{U}^\top\beta)^2\mathcal{B}] + \mathbb{E}[\{(\mathbf{U}\mathbf{U}^\top - \Sigma_u)\beta\}^{\otimes 2}] + \Sigma_u\sigma^2$, *and* $A^{\otimes 2} = AA^\top$ *for any matrix* A*;*

(ii) $\|\hat{\alpha}_n - \alpha\|^2 = O_p(n^{-(2b-1)/(a+2b)})$.

The proof can be found in Zhu et al. (2019b).

Lemma 2. *Under the alternative hypothetical model (6) with* $\delta_n = n^{-1/2}$ *and the assumptions of Theorem 1, we have*

(i) $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}\left(\mathbb{E}\left[\{\mathbf{Z} - \mathbb{E}[\mathbf{Z}|X]\}\mathcal{D}(\mathbf{Z}, X)\right], \mathcal{B}^{-1}\Gamma\mathcal{B}^{-1}\right)$;

(ii) $\|\hat{\alpha}_n - \alpha\|^2 = O_p(n^{-(2b-1)/(a+2b)})$.

Proof of Lemma 2. Define

$$\widehat{K}_w = \frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_i\mathbf{W}_i^\top, \quad \widehat{K}_{wx} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_iX_i, \quad \widehat{K}_{xw} = \widehat{K}_{wx}^\top,$$

$$\widehat{K}_{zx} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_iX_i, \quad \widehat{K}_{ux} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{U}_iX_i.$$

Consider the alternative hypothetical model $\mathcal{H}_{1n}: Y = \mathbf{Z}^\top\beta + \int_0^1 \alpha(t)X(t)dt + n^{-1/2}\mathcal{D}(\mathbf{Z}, X) + \varepsilon$. From (4), we have the following decomposition:

$$\sqrt{n}(\hat{\beta} - \beta) = \widehat{B}_n^{-1}(A_{1n} + A_{2n}), \tag{A.1}$$

where $\widehat{B}_n = \widehat{K}_w - \sum_{j=1}^m \hat{\lambda}_j^{-1} \langle \widehat{K}_{wx}, \hat{\psi}_j \rangle \langle \widehat{K}_{xw}, \hat{\psi}_j \rangle - \Sigma_u$,

$$\begin{aligned}
A_{1n} =& \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\mathbf{Z}_i - \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} \langle \widehat{K}_{zx}, \hat{\psi}_j \rangle \langle X_i, \hat{\psi}_j \rangle \Big) \big(\langle \alpha, X_i \rangle + \varepsilon_i \big) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{U}_i \left(\langle \alpha, X_i \rangle + \varepsilon_i \right) + \sqrt{n} \Big(- \frac{1}{n} \sum_{i=1}^n \mathbf{W}_i \mathbf{U}_i^\top \beta + \Sigma_u \beta \Big) \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} \langle \widehat{K}_{ux}, \hat{\psi}_j \rangle \langle X_i, \hat{\psi}_j \rangle \big(\langle \alpha, X_i \rangle + \varepsilon_i \big) \\
&+ \sqrt{n} \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} \langle \widehat{K}_{wx}, \hat{\psi}_j \rangle \langle \widehat{K}_{ux}, \hat{\psi}_j \rangle^\top \beta,
\end{aligned}$$

and

$$\begin{aligned}
A_{2n} =& \frac{1}{n} \sum_{i=1}^n \Big(\mathbf{Z}_i - \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} \langle \widehat{K}_{zx}, \hat{\psi}_j \rangle \langle X_i, \hat{\psi}_j \rangle \Big) \mathcal{D}(X_i, \mathbf{Z}_i) \\
&+ \frac{1}{n} \sum_{i=1}^n \mathbf{U}_i \mathcal{D}(X_i, \mathbf{Z}_i) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} \langle \widehat{K}_{ux}, \hat{\psi}_j \rangle \langle X_i, \hat{\psi}_j \rangle \mathcal{D}(X_i, \mathbf{Z}_i) \\
=: & \ A_{2n,1} + A_{2n,2} - A_{2n,3}.
\end{aligned}$$

It has been shown in Zhu et al. (2019b) that

$$\widehat{B}_n = \mathcal{B} + o_p(1), \quad \widehat{B}_n^{-1} A_{1n} \xrightarrow{d} \mathcal{N}(0, \mathcal{B}^{-1} \Gamma \mathcal{B}^{-1}), \tag{A.2}$$

where $\mathcal{B}$ and Γ are given in Assumption (C5) and Lemma 1, respectively.

Next, we consider A_{2n}. It follows from the law of large numbers that $A_{2n,2} = o_p(1)$. Define $\hat{\Phi}_k(g) = \sum_{j=1}^m \left(\langle \widehat{K}_{z_k x}, \hat{\psi}_j \rangle / \hat{\lambda}_j \right) \langle \hat{\psi}_j, g \rangle$, and $\Phi_k(g) = \sum_{j=1}^\infty \Big(\langle K_{z_k x}, \psi_j \rangle / \lambda_j \Big) \langle \psi_j, g \rangle$, for $g \in \mathbf{L}^2([0,1])$ and $k = 1, \ldots, p$, where $\widehat{K}_{z_k x} = n^{-1} \sum_{i=1}^n Z_{ik} X_i$ and $K_{z_k x} = \mathbb{E}[Z_k X]$. From Lemma 2 in Shin (2009), we have $\left\| \Phi_k - \hat{\Phi}_k \right\|_\infty = O_p(n^{-(2b-1)/2(a+2b)})$, $k = 1, \ldots, p$.

Observe that

$$
\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}\Big\{\sum_{j=1}^{m}\frac{1}{\hat{\lambda}_j}\langle\widehat{K}_{zx},\hat{\psi}_j\rangle\langle X_i,\hat{\psi}_j\rangle\mathcal{D}(X_i,\mathbf{Z}_i)-\sum_{j=1}^{\infty}\frac{1}{\lambda_j}\langle K_{zx},\psi_j\rangle\langle X_i,\psi_j\rangle\mathcal{D}(X_i,\mathbf{Z}_i)\Big\}\|\\
&=\|\{(\hat{\Phi}_k-\Phi_k)(\frac{1}{n}\sum_{i=1}^{n}X_i\mathcal{D}(X_i,\mathbf{Z}_i))\}_{k\in\{1,\ldots,p\}}\|\\
&\le\left[\sum_{k=1}^{p}\left\|\Phi_k-\hat{\Phi}_k\right\|_{\infty}\left\|\frac{1}{n}\sum_{i=1}^{n}X_i\mathcal{D}(X_i,\mathbf{Z}_i)\right\|\right]^{\frac{1}{2}}\\
&=O_p(n^{-(2b-1)/2(a+2b)})O_p(1)=o_p(1),
\end{aligned}
$$

where $K_{zx}=\mathbb{E}[\mathbf{Z}X]$. Thus, we get

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{\hat{\lambda}_j}\langle\widehat{K}_{zx},\hat{\psi}_j\rangle\langle X_i,\hat{\psi}_j\rangle\mathcal{D}(X_i,\mathbf{Z}_i)\\
=&\frac{1}{n}\sum_{i=1}^{n}\mathcal{D}(X_i,\mathbf{Z}_i)\sum_{j=1}^{\infty}\frac{1}{\lambda_j}\langle K_{zx},\psi_j\rangle\langle X_i,\psi_j\rangle+o_p(1)\\
=&\frac{1}{n}\sum_{i=1}^{n}\mathcal{D}(X_i,\mathbf{Z}_i)\mathbb{E}[\mathbf{Z}|X_i]+o_p(1).
\end{aligned}
$$

Therefore, we conclude that

$$
\begin{aligned}
A_{2n,1}&=\frac{1}{n}\sum_{i=1}^{n}\Big(\mathbf{Z}_i-\mathbb{E}[\mathbf{Z}|X_i]\Big)\mathcal{D}(X_i,\mathbf{Z}_i)+o_p(1)\\
&=\mathbb{E}[\{\mathbf{Z}-\mathbb{E}(\mathbf{Z}|X)\}\mathcal{D}(\mathbf{Z},X)]+o_p(1).
\end{aligned}
$$

For $A_{2n,3}$, by the fact that $K_{wx}=K_{zx}$ with $K_{wx}=\mathbb{E}[\mathbf{W}X]$, we have

$$
\begin{aligned}
A_{2n,3}&=\frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=1}^{m}\frac{1}{\hat{\lambda}_j}\langle\widehat{K}_{wx},\hat{\psi}_j\rangle\langle X_i,\hat{\psi}_j\rangle-\sum_{j=1}^{m}\frac{1}{\hat{\lambda}_j}\langle\widehat{K}_{zx},\hat{\psi}_j\rangle\langle X_i,\hat{\psi}_j\rangle\Big)\mathcal{D}(X_i,\mathbf{Z}_i)\\
&=\frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=1}^{\infty}\frac{1}{\lambda_j}\langle K_{wx},\psi_j\rangle\langle X_i,\psi_j\rangle-\sum_{j=1}^{\infty}\frac{1}{\lambda_j}\langle K_{zx},\psi_j\rangle\langle X_i,\psi_j\rangle\Big)\mathcal{D}(X_i,\mathbf{Z}_i)+o_p(1)\\
&=o_p(1).
\end{aligned}
$$

Hence, we obtain

$$
A_{2n}=\mathbb{E}[\{\mathbf{Z}-\mathbb{E}(\mathbf{Z}|X)\}\mathcal{D}(\mathbf{Z},X)]+o_p(1). \tag{A.3}
$$

It then follows from (A.1)-(A.3) that

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}\left(\mathbb{E}[\{\mathbf{Z} - \mathbb{E}(\mathbf{Z}|X)\}\mathcal{D}(\mathbf{Z}, X)], \mathcal{B}^{-1}\Gamma\mathcal{B}^{-1}\right).$$

The result that $\|\hat{\alpha}_n - \alpha\|^2 = O_p(n^{-(2b-1)/(a+2b)})$ can be proved by the similar arguments as in Zhu et al. (2019b).

■

Let $f_{\tilde{z}}(\tilde{\mathbf{z}})$ be the density function of $\tilde{\mathbf{Z}}$, $\kappa_{0j} = \int k^j(u)du$ for $j = 2, 4$, and $\mathcal{M}_j = k^j(1) - \int_0^1 (k^j)'(t)\mu_0(t)dt$ for $j = 1, 2, 4$.

The following technical lemma plays essential role in proving Theorem 1.

Lemma 3. *Under Assumptions (C2), (C3) and (C6), we have that for* $\tilde{\mathbf{v}}_1 = (\tilde{\mathbf{z}}_1^\top, x_1)^\top$,

$$\frac{1}{\varphi(h)}\mathbb{E}[K_h(\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2)|\tilde{\mathbf{V}}_1 = \tilde{\mathbf{v}}_1] = \mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{z}}_1)f_x(x_1) + o(1), \tag{A.4}$$

$$\frac{1}{\varphi(h)}\mathbb{E}[K_h^2(\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2)|\tilde{\mathbf{V}}_1 = \tilde{\mathbf{v}}_1] = \mathcal{M}_2 f_{\tilde{z}}(\tilde{\mathbf{z}}_1)\kappa_{02}^q f_x(x_1) + o(1), \tag{A.5}$$

$$\frac{1}{\varphi(h)}\mathbb{E}[K_h^2(\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2)\rho(\tilde{\mathbf{V}}_2)|\tilde{\mathbf{V}}_1 = \tilde{\mathbf{v}}_1] = \mathcal{M}_2 f_{\tilde{z}}(\tilde{\mathbf{z}}_1)\kappa_{02}^q f_x(x_1)\rho(\tilde{\mathbf{v}}_1) + o(1), \tag{A.6}$$

$$\frac{1}{\varphi(h)}\mathbb{E}[K_h^4(\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2)\rho(\tilde{\mathbf{V}}_2)|\tilde{\mathbf{V}}_1 = \tilde{\mathbf{v}}_1] = \mathcal{M}_4 f_{\tilde{z}}(\tilde{\mathbf{z}}_1)\kappa_{04}^q f_x(x_1)\rho(\tilde{\mathbf{v}}_1) + o(1), \tag{A.7}$$

and

$$\frac{1}{n\varphi(h)}\sum_{i=1}^n K_h(\tilde{\mathbf{v}}, \tilde{\mathbf{V}}_i)\mathbf{W}_i = \mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{z}})f_x(x)m_w(\tilde{\mathbf{v}}) + o_p(1). \tag{A.8}$$

Proof of Lemma 3. We only prove the results (A.4), (A.6) and (A.8). The results (A.5) and (A.7) can be proved similarly. Denote $\prod_{l=1}^q k\left(\frac{\mathbf{z}_{1l}-\mathbf{z}_l}{h_l}\right)$ and $dz_1\cdots dz_q$ by $\tilde{K}\left(\frac{\tilde{\mathbf{z}}_1-\tilde{\mathbf{z}}}{h_n}\right)$ and $d\tilde{\mathbf{z}}$, respectively. Recalling the definition of $K_h(\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2)$, we have

$$\begin{aligned} & \frac{1}{\varphi(h)}\mathbb{E}[K_h\Big(\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2\Big)|\tilde{\mathbf{V}}_1 = \tilde{\mathbf{v}}_1] \\ = \; & \frac{1}{\varphi(h)}\int_0^1\int_{-\infty}^{\infty} k(t)\tilde{K}\Big(\frac{\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}}{h_n}\Big)f_{\tilde{z}}(\tilde{\mathbf{z}})d\tilde{\mathbf{z}}dP\Big(\frac{d(x_1, x)}{h_0} \le t\Big) \end{aligned}$$

$$= \frac{1}{\varphi(h)} \int_0^1 k(t) dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) \int_{-\infty}^{\infty} \tilde{K}\Big(\frac{\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}}{h_n}\Big) f_{\tilde{z}}(\tilde{\mathbf{z}}) d\tilde{\mathbf{z}} + o_p(1).$$

Note that $\varphi(h) = \psi(h_0)\prod_{l=1}^{q} h_l$. It then follows from Lemma 1 in Louani (2010) that $\{\psi(h_0)\}^{-1}\int_0^1 k(t)dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) = \mathcal{M}_1 f_x(x_1) + o(1)$. Further, it is easy to prove that $(\prod_{l=1}^{q} h_l)^{-1}\int_{-\infty}^{\infty} \tilde{K}\Big(\frac{\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}}{h_n}\Big) f_{\tilde{z}}(\tilde{\mathbf{z}})d\tilde{\mathbf{z}} = f_{\tilde{z}}(\tilde{\mathbf{z}}_1) + o(1)$. This completes the proof of (A.4).

Some straightforward calculation yields

$$\{\varphi(h)\}^{-1}\int_{-\infty}^{\infty} \tilde{K}^2\Big(\frac{\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}}{h_n}\Big)\rho(\tilde{\mathbf{v}}) f_{\tilde{z}}(\tilde{\mathbf{z}}) d\tilde{\mathbf{z}} = \kappa_{02}^q f_{\tilde{z}}(\tilde{\mathbf{z}}_1)\rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x_1) + o(1).$$

Then it follows that

$$\begin{aligned}
& \frac{1}{\varphi(h)}\mathbb{E}[K_h^2(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\rho(\tilde{\mathbf{V}})|\tilde{\mathbf{V}}_1 = \tilde{\mathbf{v}}_1] \\
= & \frac{1}{\varphi(h)}\int_0^1\int_{-\infty}^{\infty} k^2(t)\tilde{K}^2\Big(\frac{\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}}{h_n}\Big)\rho(\tilde{\mathbf{v}}) f_{\tilde{z}}(\tilde{\mathbf{z}}) d\tilde{\mathbf{z}} dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) \\
= & \frac{1}{\varphi(h)}\int_0^1 k^2(t)\Big(\int_{-\infty}^{\infty} \tilde{K}^2\Big(\frac{\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}}{h_n}\Big)\rho(\tilde{\mathbf{v}}) f_{\tilde{z}}(\tilde{\mathbf{z}}) d\tilde{\mathbf{z}}\Big) dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) \\
= & \frac{\kappa_{02}^q f_{\tilde{z}}(\tilde{\mathbf{z}}_1)}{\psi(h)}\int_0^1 k^2(t)\rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x) dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) + o(1) \\
= & \frac{\kappa_{02}^q f_{\tilde{z}}(\tilde{\mathbf{z}}_1)}{\psi(h)}\int_0^1 k^2(t)\rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x_1) dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) \\
& + \frac{\kappa_{02}^q f_{\tilde{z}}(\tilde{\mathbf{z}}_1)}{\psi(h)}\int_0^1 k^2(t)\{\rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x) - \rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x_1)\} dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) + o(1).
\end{aligned}$$

The first term is $\kappa_{02}^q \mathcal{M}_2 f_{\tilde{z}}(\tilde{\mathbf{z}}_1) f_x(x_1)\rho(\tilde{\mathbf{v}}_{\mathbf{1}}) + o(1)$. For the second term, it follows from Assumption (C6) that

$$\begin{aligned}
& \Big|\frac{\kappa_{02}^q f_{\tilde{z}}(\tilde{\mathbf{z}}_1)}{\varphi(h)}\int_0^1 k^2(t)\{\rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x) - \rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x_1)\} dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big)\Big| \\
\le & \frac{f_{\tilde{z}}(\tilde{\mathbf{z}}_1)\kappa_{02}^q}{\varphi(h)}\int_0^1 \sup_{x\in B(x_1,h_0)} |\rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x) - \rho(\tilde{\mathbf{z}}_{\mathbf{1}}, x_1)| k^2(t) dP\Big(\frac{d(x_1,x)}{h_0} \le t\Big) \\
= & O(h_0^\theta) = o(1).
\end{aligned}$$

Thereby the result (A.6) is proved.

Observe that

$$\frac{1}{n\varphi(h)}\sum_{i=1}^{n} K_h(\tilde{\mathbf{v}},\tilde{\mathbf{V}}_i)\mathbf{W}_i = \mathbb{E}[\frac{1}{\varphi(h)}K_h(\tilde{\mathbf{v}},\tilde{\mathbf{V}})\mathbf{W}] + o_p(1)$$

$$= \mathbb{E}[\frac{1}{\varphi(h)} K_h(\tilde{\mathbf{v}}, \tilde{\mathbf{V}}) m_w(\tilde{\mathbf{V}})] + o_p(1).$$

Then (A.8) can be proved following similar arguments as in the proof of (A.6).

■

Lemma 4 is used for estimating $\mathcal{T}_{4n}$, which is considered in Lemma 8.

Lemma 4. *Under the same assumptions in Theorem 1, it follows that*

$$\frac{1}{(n-1)\varphi(h)^{1/2}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)(\mathbf{W}_i \varepsilon_j + \mathbf{W}_j \varepsilon_i) = O_p(1).$$

The proof is similar to that of Theorem 1, and we omit the details.

Define

$$\begin{aligned} H_n(\Delta_i, \Delta_j) &= \frac{1}{\varphi(h)^{1/2}} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j) \varepsilon_{ui} \varepsilon_{uj}, \\ G_n(\Delta_1, \Delta_2) &= \mathbb{E}[H_n(\Delta_3, \Delta_1) H_n(\Delta_3, \Delta_2) | \Delta_1, \Delta_2], \end{aligned}$$

and $\Delta_i = (\varepsilon_i, \mathbf{U}_i^\top, \tilde{\mathbf{V}}_i^\top)$, for $i = 1, \ldots, n$.

The next lemma verifies the condition of the central limit theorem in Theorem 1.

Lemma 5. *Under the same assumptions in Theorem 1, we have*

$$\frac{\mathbb{E}[G_n^2(\Delta_1, \Delta_2)] + n^{-1}\mathbb{E}[H_n^4(\Delta_1, \Delta_2)]}{\{\mathbb{E}[H_n^2(\Delta_1, \Delta_2)]\}^2} = o(1).$$

Proof of Lemma 5. It is easy to check that

$$\begin{aligned} & G_n(\Delta_1, \Delta_2) = \mathbb{E}[H_n(\Delta_3, \Delta_1) H_n(\Delta_3, \Delta_2) | \Delta_1, \Delta_2] \\ =\ & \mathbb{E}\Big[\frac{1}{\varphi(h)} K_h(\tilde{\mathbf{V}}_3, \tilde{\mathbf{V}}_1) K_h(\tilde{\mathbf{V}}_3, \tilde{\mathbf{V}}_2) \varepsilon_{u3}^2 \varepsilon_{u1} \varepsilon_{u2} \Big| \Delta_1, \Delta_2] \\ =\ & \varepsilon_{u1}\varepsilon_{u2} \mathbb{E}\Big[\frac{1}{\varphi(h)} K_h(\tilde{\mathbf{V}}_3, \tilde{\mathbf{V}}_1) K_h(\tilde{\mathbf{V}}_3, \tilde{\mathbf{V}}_2) \varepsilon_{u3}^2 | \Delta_1, \Delta_2\Big] \\ =\ & \varepsilon_{u1}\varepsilon_{u2} \mathbb{E}\Big[\frac{1}{\varphi(h)} K_h(\tilde{\mathbf{V}}_3, \tilde{\mathbf{V}}_1) K_h(\tilde{\mathbf{V}}_3, \tilde{\mathbf{V}}_2)(\varepsilon_3^2 - 2\beta^\top \mathbf{U}_3 \varepsilon_3 \end{aligned}$$

$$
\begin{aligned}
&\quad +\beta^\top \mathbf{U}_3\mathbf{U}_3^\top\beta)|\Delta_1,\Delta_2\Big] \\
&= \varepsilon_{u1}\varepsilon_{u2}\mathbb{E}\Big[\frac{1}{\varphi(h)}K_h(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_1)K_h(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_2)\sigma_1(\tilde{\mathbf{V}}_3)|\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2\Big],
\end{aligned}
$$

where $\sigma_1(\tilde{\mathbf{V}})=\mathbb{E}[\varepsilon_u^2|\tilde{\mathbf{V}}]=\mathbb{E}[\varepsilon^2|\tilde{\mathbf{V}}]+\beta^\top\mathbb{E}[\mathbf{U}\mathbf{U}^\top|\tilde{\mathbf{V}}]\beta$. Moreover, we validate that

$$
\begin{aligned}
&\mathbb{E}^2\Big[\frac{1}{\varphi(h)}K_h(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_1)K_h(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_2)\sigma_1(\tilde{\mathbf{V}}_3)|\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2\Big] \\
\leq\ &\varphi(h)\mathbb{E}\big[\sigma_1(\tilde{\mathbf{V}}_3)\frac{K_h^2(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_1)}{\varphi(h)}|\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2\big]\mathbb{E}\big[\sigma_1(\tilde{\mathbf{V}}_3)\frac{K_h^2(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_2)}{\varphi(h)}|\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2\big]\Big] \\
=\ &\kappa_{02}^{2q}\mathcal{M}_2^2\varphi(h)\sigma_1(\tilde{\mathbf{V}}_1)\sigma_1(\tilde{\mathbf{V}}_2)f_{\tilde{z}}(\tilde{\mathbf{Z}}_1)f_{\tilde{z}}(\tilde{\mathbf{Z}}_2)f_x(X_1)f_x(X_2)+o(1).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\mathbb{E}[G_n^2(\Delta_1,\Delta_2)] \\
=\ &\mathbb{E}\Bigg[\varepsilon_{u1}^2\varepsilon_{u2}^2\mathbb{E}^2\Big[\frac{1}{\varphi(h)}K_h(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_1)K_h(\tilde{\mathbf{V}}_3,\tilde{\mathbf{V}}_2)\sigma_1(\tilde{\mathbf{V}}_3)|\tilde{\mathbf{V}}_\mathbf{1},\tilde{\mathbf{V}}_\mathbf{2}\Big]\Bigg] \\
\leq\ &\kappa_{02}^{2q}\mathcal{M}_2^2\varphi(h)\mathbb{E}\Big[\{\sigma_1(\tilde{\mathbf{V}}_1)\}^2\{\sigma_1(\tilde{\mathbf{V}}_2)\}^2f_{\tilde{z}}(\tilde{\mathbf{Z}}_1)f_{\tilde{z}}(\tilde{\mathbf{Z}}_2)f_x(X_1)f_x(X_2)\Big]+o(1).
\end{aligned}
$$

Next, consider $\mathbb{E}[H_n^2(\Delta_1,\Delta_2)]$, and a straightforward calculation gives

$$
\begin{aligned}
\mathbb{E}[H_n^2(\Delta_1,\Delta_2)] &= \mathbb{E}\Big[\frac{1}{\varphi(h)^{1/2}}K_h(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\varepsilon_{u1}\varepsilon_{u2}\Big]^2 \\
&= \mathbb{E}\Big[\frac{1}{\varphi(h)}K_h^2(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\sigma_1(\tilde{\mathbf{V}}_1)\sigma_1(\tilde{\mathbf{V}}_2)\Big] \\
&= \mathbb{E}\Big[\sigma_1(\tilde{\mathbf{V}}_1)\mathbb{E}[\frac{1}{\varphi(h)}K_h^2(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\sigma_1(\tilde{\mathbf{V}}_2)|\tilde{\mathbf{V}}_1]\Big] \\
&= \mathcal{M}_2\kappa_{02}^q\mathbb{E}\Big[\{\sigma_1(\tilde{\mathbf{V}}_1)\}^2f_{\tilde{z}}(\tilde{\mathbf{Z}}_1)f_x(X_1)\Big]+o(1).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\mathbb{E}[H_n^4(\Delta_1,\Delta_2)] &= \mathbb{E}\Big[\frac{1}{\varphi^2(h)}K_h^4(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\varepsilon_{u1}^4\varepsilon_{u2}^4\Big] \\
&= \mathbb{E}\Big[\frac{1}{\varphi^2(h)}K_h^4(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\sigma_2(\tilde{\mathbf{V}}_1)\sigma_2(\tilde{\mathbf{V}}_2)\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[\frac{1}{\varphi^2(h)}K_h^4(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\sigma_2(\tilde{\mathbf{V}}_1)\sigma_2(\tilde{\mathbf{V}}_2)|\tilde{\mathbf{V}}_1]\Big] \\
&= \frac{1}{\varphi(h)}\mathbb{E}\Big[\sigma_2(\tilde{\mathbf{V}}_1)\mathbb{E}\Big[\frac{1}{\varphi(h)}K_h^4(\tilde{\mathbf{V}}_1,\tilde{\mathbf{V}}_2)\sigma_2(\tilde{\mathbf{V}}_2)|\tilde{\mathbf{V}}_1\Big]\Big] \\
&= \frac{\mathcal{M}_4\kappa_{04}^q}{\varphi(h)}\mathbb{E}\big[\{\sigma_2(\tilde{\mathbf{V}}_1)\}^2f_{\tilde{z}}(\tilde{\mathbf{Z}}_1)f_x(X_1)\Big]+o(\frac{1}{\varphi(h)}).
\end{aligned}
$$

Thus, we obtain

$$\frac{\mathbb{E}[G_n^2(\Delta_1,\Delta_2)]+n^{-1}\mathbb{E}[H_n^4(\Delta_1,\Delta_2)]}{\mathbb{E}^2[H_n^2(\Delta_1,\Delta_2)]} = O\big(\varphi(h)+\frac{1}{n\varphi(h)}\big) = o(1).$$

■

Let $\Upsilon_i^{x\alpha} = \int_0^1 \{\alpha(t)-\hat{\alpha}_n(t)\}X_i(t)dt,\ i=1,\dots,n.$ Define

$$\begin{aligned}
\mathcal{T}_{1n} &= \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j<i} K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\varepsilon_i\varepsilon_j,\\
\mathcal{T}_{2n} &= \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j<i} K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)(\mathbf{Z}_i^\top\beta-\mathbf{W}_i^\top\hat{\beta}_n)(\mathbf{Z}_j^\top\beta-\mathbf{W}_j^\top\hat{\beta}_n),\\
\mathcal{T}_{3n} &= \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j<i} K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\Upsilon_i^{x\alpha}\Upsilon_j^{x\alpha},\\
\mathcal{T}_{4n} &= \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j<i} K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\left\{(\mathbf{Z}_i^\top\beta-\mathbf{W}_i^\top\hat{\beta}_n)\varepsilon_j+(\mathbf{Z}_j^\top\beta-\mathbf{W}_j^\top\hat{\beta}_n)\varepsilon_i\right\},\\
\mathcal{T}_{5n} &= \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\varepsilon_i\Upsilon_j^{x\alpha},\\
\mathcal{T}_{6n} &= \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)(\mathbf{Z}_i^\top\beta-\mathbf{W}_i^\top\hat{\beta}_n)\Upsilon_j^{x\alpha}.
\end{aligned}$$

The next lemmas 6–8 establish the asymptotic behaviors of the terms $\mathcal{T}_{ln}$, for $l=2,\dots,6$.

Lemma 6. *Under the same assumptions in Theorem 1, we have*

$$\mathcal{T}_{2n} = \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j<i} K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\beta^\top\mathbf{U}_i\mathbf{U}_j^\top\beta + o_p(1). \tag{A.9}$$

Proof of Lemma 6. It is easy to check that

$$\mathbf{Z}_i^\top\beta-\mathbf{W}_i^\top\hat{\beta}_n = -\beta^\top\mathbf{U}_i-\mathbf{W}_i^\top(\hat{\beta}_n-\beta),\ i=1,\dots,n. \tag{A.10}$$

Then we split $\mathcal{T}_{2n}$ into four parts:

$$\mathcal{T}_{2n} = \frac{1}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\beta^\top\mathbf{U}_i\mathbf{U}_j^\top\beta$$

$$
\begin{aligned}
&+\frac{\beta^\top}{\sqrt{n}(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\mathbf{U}_i\mathbf{W}_j^\top\sqrt{n}(\hat{\beta}_n-\beta)\\
&+\frac{\beta^\top}{\sqrt{n}(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\mathbf{U}_j\mathbf{W}_i^\top\sqrt{n}(\hat{\beta}_n-\beta)\\
&+\frac{(\hat{\beta}_n-\beta)^\top}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\mathbf{W}_i\mathbf{W}_j^\top(\hat{\beta}_n-\beta)\\
=:&\ \mathcal{T}_{2n,1}+\mathcal{T}_{2n,2}+\mathcal{T}_{2n,3}+\mathcal{T}_{2n,4}.
\end{aligned}
$$

By Lemma 1 and (A.8), we have

$$
\begin{aligned}
\mathcal{T}_{2n,2} &= \frac{\varphi(h)^{1/2}\beta^\top}{\sqrt{n}}\sum_{i=1}^n\mathbf{U}_i\Big\{\frac{1}{(n-1)\varphi(h)}\sum_{j=1,j\neq i}^n K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\mathbf{W}_j^\top\Big\}\sqrt{n}(\hat{\beta}_n-\beta)\\
&= \frac{\mathcal{M}_1\varphi(h)^{1/2}\beta^\top}{\sqrt{n}}\sum_{i=1}^n\mathbf{U}_i f_{\tilde{z}}(\tilde{\mathbf{Z}}_i)f_x(X_i)m_w(\tilde{\mathbf{V}}_i)^\top\sqrt{n}(\hat{\beta}_n-\beta).
\end{aligned}
$$

Note that $n^{-1/2}\mathcal{M}_1\beta^\top\sum_{i=1}^n\mathbf{U}_i f_{\tilde{z}}(\tilde{\mathbf{Z}}_i)f_x(X_i)m_w(\tilde{\mathbf{V}}_i)=O_p(1)$, and $\sqrt{n}(\hat{\beta}_n-\beta)=O_p(1)$. Then $\mathcal{T}_{2n,2}=O_p(\varphi(h)^{1/2})=o_p(1)$. The result that $\mathcal{T}_{2n,3}=o_p(1)$ can be proved similarly.

For $\mathcal{T}_{2n,4}$, it follows that

$$
\begin{aligned}
\mathcal{T}_{2n,4} &= \sqrt{n}(\hat{\beta}_n-\beta)^\top\frac{\varphi(h)^{1/2}}{n}\sum_{i=1}^n\mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{Z}}_i)f_x(X_i)m_w(\tilde{\mathbf{V}}_i)\mathbf{W}_i^\top\sqrt{n}(\hat{\beta}_n-\beta)+o_p(1)\\
&= O_p(\varphi(h)^{1/2})=o_p(1).
\end{aligned}
$$

Thus, (A.9) is proved.

■

Lemma 7. *Under the same assumptions in Theorem 1, we have*

$$
\mathcal{T}_{3n}=o_p(1),\quad \mathcal{T}_{5n}=o_p(1).
$$

Proof of Lemma 7. Observed that by the Cauchy-Schwarz inequality,

$$
\Upsilon_i^{x\alpha}=\int_0^1\{\alpha(t)-\hat{\alpha}_n(t)\}X_i(t)dt\leq\|\hat{\alpha}_n-\alpha\|\|X_i\|\leq C\|\hat{\alpha}_n-\alpha\|.
$$

It then follows from Lemmas 1–2 and Assumption (C4) that

$$
\mathcal{T}_{3n} = \frac{2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^n\sum_{j<i}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\Upsilon_i^{x\alpha}\Upsilon_j^{x\alpha}
$$

$$
\begin{aligned}
&\leq \frac{C\|\hat{\alpha}_n-\alpha\|^2}{(n-1)\varphi(h)^{1/2}}\sum_{i=1}^{n}\sum_{j<i}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\\
&= C\|\hat{\alpha}_n-\alpha\|^2\varphi(h)^{1/2}\Big(\sum_{i=1}^{n}\{\frac{1}{n\varphi(h)}\sum_{j=1}^{n}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\}+O_p(1)\Big)\\
&= C\|\hat{\alpha}_n-\alpha\|^2\varphi(h)^{1/2}\Big(\mathcal{M}_1\sum_{i=1}^{n}f_{\tilde{z}}(\tilde{\mathbf{Z}}_i)f_x(X_i)+O_p(1)\Big)\\
&= O_p((\varphi(h)^{1/2}n^{1-(2b-1)/(a+2b)})=o_p(1).
\end{aligned}
$$

For $\mathcal{T}_{5n}$, noting that $\varepsilon_1,\ldots,\varepsilon_n$ are independent and have zero mean, it can be checked that

$$
\begin{aligned}
\mathbb{E}[\mathcal{T}_{5n}]^2 &= \frac{4}{(n-1)^2\varphi(h)}\mathbb{E}[\sum_{i=1}^{n}\varepsilon_i\{\sum_{j=1,j\neq i}^{n}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\Upsilon_j^{x\alpha}\}]^2\\
&= \frac{4}{(n-1)^2\varphi(h)}\sum_{i=1}^{n}\mathbb{E}[\varepsilon_i^2\{\sum_{j=1,j\neq i}^{n}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\Upsilon_j^{x\alpha}\}^2]\\
&= \frac{4}{(n-1)^2\varphi(h)}\sum_{i=1}^{n}\sum_{j_1=1,j_1\neq i}^{n}\sum_{j_2=1,j_2\neq i,j_1}^{n}\mathbb{E}[\varepsilon_i^2K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_1})K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_2})\Upsilon_{j_1}^{x\alpha}\Upsilon_{j_2}^{x\alpha}]\\
&\quad+\frac{4}{(n-1)^2\varphi(h)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\mathbb{E}[\varepsilon_i^2K_h^2(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)(\Upsilon_j^{x\alpha})^2].
\end{aligned}
$$

It can be proved that

$$
\frac{4}{(n-1)^2\varphi(h)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\mathbb{E}[\varepsilon_i^2K_h^2(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)(\Upsilon_j^{x\alpha})^2]=O(n^{-(2b-1)/(a+2b)})=o(1).
$$

Thus we have

$$
\begin{aligned}
\mathbb{E}[\mathcal{T}_{5n}]^2 &= \frac{4}{(n-1)^2\varphi(h)}\sum_{i=1}^{n}\sum_{j_1=1,j_1\neq i}^{n}\sum_{j_2=1,j_2\neq i,j_1}^{n}\mathbb{E}\left[\mathbb{E}\{\varepsilon_i^2|\tilde{\mathbf{V}}_i\}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_1})K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_2})\Upsilon_{j_1}^{x\alpha}\Upsilon_{j_2}^{x\alpha}\right]\\
&\quad+o(1)\\
&\leq \frac{4}{(n-1)^2\varphi(h)}\sum_{i=1}^{n}\sum_{j_1=1,j_1\neq i}^{n}\sum_{j_2=1,j_2\neq i,j_1}^{n}\mathbb{E}\Big[\mathbb{E}\{\varepsilon_i^2|\tilde{\mathbf{V}}_i\}\mathbb{E}\{K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_1})|\Upsilon_{j_1}^{x\alpha}||\tilde{\mathbf{V}}_i\}\\
&\quad\times\mathbb{E}\{K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_2})|\Upsilon_{j_2}^{x\alpha}||\tilde{\mathbf{V}}_i\}\Big]+o(1)\\
&\leq \frac{4}{(n-1)^2\varphi(h)}\sum_{i=1}^{n}\mathbb{E}\Bigg[\|\hat{\alpha}_n-\alpha\|^2\mathbb{E}\{\varepsilon_i^2|\tilde{\mathbf{V}}_i\}\sum_{j_1=1,j_1\neq i}^{n}\mathbb{E}\{K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_1})\|X_{j_1}\||\tilde{\mathbf{V}}_i\}\\
&\quad\times\sum_{j_2=1,j_2\neq i,j_1}^{n}\mathbb{E}\{K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_{j_2})\|X_{j_2}\||\tilde{\mathbf{V}}_i\}\Bigg]+o(1)
\end{aligned}
$$

$$= \quad O(\varphi(h)n^{1-(2b-1)/(a+2b)}) = o(1).$$

Hence, we have $\mathcal{T}_{5n} = o_p(1)$.

■

Lemma 8. *Under the same assumptions in Theorem 1, we have*

$$\mathcal{T}_{4n} = -\frac{2\beta^\top}{(n-1)\varphi(h)^{1/2}} \sum_{i=1}^{n} \sum_{j=1,j<i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\{\mathbf{U}_i\varepsilon_j + \mathbf{U}_j\varepsilon_i\} + o_p(1), \tag{A.11}$$

and

$$\mathcal{T}_{6n} = o_p(1).$$

Proof of Lemma 8. It follows from (A.10) that

$$\begin{aligned}\mathcal{T}_{4n} \quad &= \quad -\frac{\beta^\top}{(n-1)\varphi(h)^{1/2}} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)(\mathbf{U}_i\varepsilon_j + \mathbf{U}_j\varepsilon_i) \\ &\quad + \frac{1}{(n-1)\varphi(h)^{1/2}} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)(\mathbf{W}_i\varepsilon_j + \mathbf{W}_j\varepsilon_i)^\top(\beta - \hat{\beta}_n).\end{aligned}$$

The second term is proved to be $O_p(n^{-1/2}) = o_p(1)$ from Lemmas 4 and 1. Hence (A.11) holds.

Observe that

$$\begin{aligned}\mathcal{T}_{6n} \quad &= \quad -\frac{2}{(n-1)\varphi(h)^{1/2}} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\beta^\top\mathbf{U}_i\Upsilon_j^{x\alpha} \\ &\quad -\frac{2}{(n-1)\varphi(h)^{1/2}} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\mathbf{W}_i^\top(\hat{\beta}_n - \beta)\Upsilon_j^{x\alpha} \\ &=: \quad \mathcal{T}_{6n,1} + \mathcal{T}_{6n,2}.\end{aligned}$$

Similarly to prove $\mathcal{T}_{5n} = o_p(1)$, we can verify that $\mathcal{T}_{6n,1} = o_p(1)$ via proving $\mathbb{E}[\mathcal{T}_{6n,1}^2] = o(1)$. For $\mathcal{T}_{6n,2}$, it follows from Lemma 3 and Assumption (C4) that

$$\mathcal{T}_{6n,2} \quad = \quad -2\varphi(h)^{1/2} \sum_{j=1}^{n} \Upsilon_j^{x\alpha}\{\frac{1}{(n-1)\varphi(h)} \sum_{i=1,i\neq j} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\mathbf{W}_i\}^\top(\hat{\beta}_n - \beta)$$

$$
\begin{aligned}
&= -2\varphi(h)^{1/2}\sum_{j=1}^{n}\Upsilon_j^{x\alpha}\mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{Z}}_j)f_x(X_j)m_w(\tilde{\mathbf{V}}_j)^\top(\hat{\beta}_n-\beta)+o_p(1)\\
&= O_p(\varphi(h)^{1/2}n^{1/2-(2b-1)/2(a+2b)}) = o_p(1).
\end{aligned}
$$

Thus, it is proved that $\mathcal{T}_{6n} = o_p(1)$.

■

Proof of Theorem 1. Note that

$$
\hat{\varepsilon}_{ni} = \varepsilon_i + (\mathbf{Z}_i^\top\beta - \mathbf{W}_i^\top\hat{\beta}_n) + \int_0^1\{\alpha(t)-\hat{\alpha}_n(t)\}X_i(t)dt.
$$

Some simple calculations yield $n\varphi(h)^{1/2}\mathcal{T}_n = \sum_{l=1}^{6}\mathcal{T}_{ln}$. From Lemmas 6–8, we have

$$
n\varphi(h)^{1/2}\mathcal{T}_n = \frac{2}{(n-1)}\sum_{i=1}^{n}\sum_{j<i}H_n(\Delta_i,\Delta_j)+o_p(1).
$$

It is easy to show that $\mathbb{E}[H_n(\Delta_1,\Delta_2)|\Delta_1] = 0$. Further, by Lemma 3.2 in Zheng (1996) and Lemma 5, $n\varphi(h)^{1/2}\mathcal{T}_n$ is asymptotically normally distributed with mean zero and variance σ_0^2, where $\sigma_0^2 = \mathbb{E}[2H_n^2(\Delta_1,\Delta_2)] = 2\mathcal{M}_2\kappa_{02}^q\mathbb{E}[\{\sigma_1(\tilde{\mathbf{V}})\}^2 f_{\tilde{z}}(\tilde{\mathbf{Z}})f_x(X)]$.

■

Proof of Theorem 2. Define

$$
\begin{aligned}
\mathcal{T}_{7n} &= \frac{2}{\sqrt{n}(n-1)\varphi(h)^{3/4}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\varepsilon_i\mathcal{D}(X_j,\mathbf{Z}_j),\\
\mathcal{T}_{8n} &= \frac{1}{n(n-1)\varphi(h)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\mathcal{D}(X_i,\mathbf{Z}_i)\mathcal{D}(X_j,\mathbf{Z}_j),\\
\mathcal{T}_{9n} &= \frac{2}{\sqrt{n}(n-1)\varphi(h)^{3/4}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)(\mathbf{Z}_i^\top\beta-\mathbf{W}_i^\top\hat{\beta}_n)\mathcal{D}(X_j,\mathbf{Z}_j),\\
\mathcal{T}_{10n} &= \frac{2}{\sqrt{n}(n-1)\varphi(h)^{3/4}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}K_h(\tilde{\mathbf{V}}_i,\tilde{\mathbf{V}}_j)\mathcal{D}(X_j,\mathbf{Z}_j)\Upsilon_i^{x\alpha}.
\end{aligned}
$$

Recall that $\hat{\varepsilon}_{ni} = Y_i - \mathbf{W}_i^\top\hat{\beta}_n - \int_0^1\hat{\alpha}_n(t)X_i(t)dt,\ i=1,\ldots,n$. Under the local alternative hypothesis (6) with $\delta_n = n^{-1/2}\varphi(h)^{-1/4}$, we have

$$
\hat{\varepsilon}_{ni} = \varepsilon_i + (\mathbf{Z}_i^\top\beta - \mathbf{W}_i^\top\hat{\beta}_n) + \int_0^1\{\alpha(t)-\hat{\alpha}_n(t)\}X_i(t)dt + n^{-1/2}\varphi(h)^{-1/4}\mathcal{D}(X_i,\mathbf{Z}_i).
$$

Some simple calculations yield

$$n\varphi(h)^{1/2}\mathcal{T}_n = \sum_{l=1}^{10} \mathcal{T}_{ln}. \tag{A.12}$$

In view of the proof of Theorem 1, we obtain

$$\sum_{l=1}^{6} \mathcal{T}_{ln} \xrightarrow{d} \mathcal{N}(0, \sigma_0^2). \tag{A.13}$$

By the law of large numbers, it follows that

$$\begin{aligned}
\frac{1}{n\varphi(h)} \sum_{j=1}^{n} K_h(\tilde{\mathbf{v}}, \tilde{\mathbf{V}}_j)\mathcal{D}(X_j, \mathbf{Z}_j) &= \mathbb{E}\left[\mathbb{E}[\frac{1}{\varphi(h)} K_h(\tilde{\mathbf{v}}, \tilde{\mathbf{V}})\mathcal{D}(\mathbf{Z}, X)|\tilde{\mathbf{V}}]\right] + o_p(1) \\
&= \mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{z}}) f_x(x)\mathcal{D}(\tilde{\mathbf{v}}) + o_p(1),
\end{aligned}$$

where $\mathcal{D}(\tilde{\mathbf{V}}) = \mathbb{E}[\mathcal{D}(\mathbf{Z}, X)|\tilde{\mathbf{V}}]$. For $\mathcal{T}_{7n}$, we have

$$\begin{aligned}
\mathcal{T}_{7n} &= \frac{2\varphi(h)^{1/4}}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{n\varphi(h)} \sum_{j=1}^{n} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\mathcal{D}(X_j, \mathbf{Z}_j) \right\} \varepsilon_i + o_p(1) \\
&= \frac{2\varphi(h)^{1/4}}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{Z}}_i) f_x(X_i)\mathcal{D}(\tilde{\mathbf{V}}_i)\varepsilon_i + o_p(1) \\
&= O_p(\varphi(h)^{1/4}) = o_p(1).
\end{aligned} \tag{A.14}$$

Next, for $\mathcal{T}_{8n}$, we obtain

$$\begin{aligned}
\mathcal{T}_{8n} &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{(n-1)\varphi(h)} \sum_{j=1, j\neq i} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\mathcal{D}(X_j, \mathbf{Z}_j) \right\} \mathcal{D}(X_i, \mathbf{Z}_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{Z}}_i) f_x(X_i)\mathcal{D}(\tilde{\mathbf{V}}_i)\mathcal{D}(X_i, \mathbf{Z}_i) + o_p(1) \\
&= \mathcal{M}_1 \mathbb{E}[f_{\tilde{z}}(\tilde{\mathbf{Z}}) f_x(X)\mathcal{D}(\tilde{\mathbf{V}})\mathcal{D}(\mathbf{Z}, X)] + o_p(1).
\end{aligned} \tag{A.15}$$

For $\mathcal{T}_{9n}$, recalling (A.10), we have

$$\begin{aligned}
\mathcal{T}_{9n} &= -\frac{2}{\sqrt{n}(n-1)\varphi(h)^{3/4}} \sum_{i=1}^{n} \sum_{j=1, j\neq i} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\mathcal{D}(X_j, \mathbf{Z}_j)\mathbf{W}_i^\top(\hat{\beta}_n - \beta) \\
&- \frac{2}{\sqrt{n}(n-1)\varphi(h)^{3/4}} \sum_{i=1}^{n} \sum_{j=1, j\neq i} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j)\mathcal{D}(X_j, \mathbf{Z}_j)\mathbf{U}_i^\top \beta
\end{aligned}$$

$$=: \quad -\mathcal{T}_{9n,1} - \mathcal{T}_{9n,2}.$$

For the first term, it is verified that

$$\begin{aligned}
\mathcal{T}_{9n,1} &= \frac{2\varphi(h)^{1/4}}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{(n-1)\varphi(h)} \sum_{j=1, j\neq i} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j) \mathcal{D}(X_j, \mathbf{Z}_j) \right\} \mathbf{W}_i^\top (\hat{\beta}_n - \beta) \\
&= \frac{2\varphi(h)^{1/4}}{n} \sum_{i=1}^{n} \mathcal{M}_1 f_{\tilde{z}}(\tilde{\mathbf{Z}}_i) f_x(X_i) \mathcal{D}(\tilde{\mathbf{V}}_i) \mathbf{W}_i^\top \sqrt{n}(\hat{\beta}_n - \beta) + o_p(1).
\end{aligned}$$

By the law of large numbers and $\sqrt{n}(\hat{\beta}_n - \beta) = O_p(1)$, we have that $\mathcal{T}_{9n,1} = O_p(\varphi(h)^{1/4}) = o_p(1)$. For the second term $\mathcal{T}_{9n,2}$, it can be shown that

$$\begin{aligned}
\mathcal{T}_{9n,2} &= \frac{\mathcal{M}_1 \varphi(h)^{1/4}}{n} \sum_{i=1}^{n} f_{\tilde{z}}(\tilde{\mathbf{Z}}_i) f_x(X_i) \mathcal{D}(\tilde{\mathbf{V}}_i) \mathbf{U}_i^\top \beta + o_p(1) \\
&= O_p(n^{-\frac{1}{2}} \varphi(h)^{1/4}) = o_p(1).
\end{aligned}$$

Therefore, we conclude that

$$\mathcal{T}_{9n} = o_p(1). \tag{A.16}$$

For $\mathcal{T}_{10n}$, we get

$$\begin{aligned}
\mathcal{T}_{10n} =& \frac{2\varphi(h)^{1/4}}{\sqrt{n}} \sum_{i=1}^{n} \{ \frac{1}{(n-1)\varphi(h)} \sum_{j=1, j\neq i} K_h(\tilde{\mathbf{V}}_i, \tilde{\mathbf{V}}_j) \mathcal{D}(X_j, \mathbf{Z}_j) \} \\
& \times \int_0^1 \{\alpha(t) - \hat{\alpha}_n(t)\} X_i(t) dt \\
=& \frac{2\mathcal{M}_1 \varphi(h)^{1/4}}{n} \sum_{i=1}^{n} f_{\tilde{z}}(\tilde{\mathbf{Z}}_i) f_x(X_i) \mathcal{D}(\tilde{\mathbf{V}}_i) \int_0^1 \{\alpha(t) - \hat{\alpha}_n(t)\} X_i(t) dt + o_p(1) \\
=& O_p(n^{-\frac{2b-1}{2(a+2b)}} \varphi(h)^{1/4}) = o_p(1).
\end{aligned} \tag{A.17}$$

It follows from (A.14)–(A.17) that

$$\sum_{l=7}^{10} \mathcal{T}_{ln} = \mathcal{M}_1 \mathbb{E}[f_{\tilde{z}}(\tilde{\mathbf{Z}}) f_x(X) \mathcal{D}(\tilde{\mathbf{V}}) \mathcal{D}(\mathbf{Z}, X)] + o_p(1). \tag{A.18}$$

Combining (A.12), (A.13), and (A.18), we conclude that

$$n\varphi(h)^{1/2} \mathcal{T}_n \xrightarrow{d} \mathcal{N}(\Lambda, \sigma_0^2),$$

where $\Lambda = \mathcal{M}_1\mathbb{E}[f_{\tilde{z}}(\tilde{\mathbf{Z}})f_x(X)\mathcal{D}(\tilde{\mathbf{V}})\mathcal{D}(\mathbf{Z}, X)]$.

■

Proof of Corollary 1. Denote $\mathbb{E}[Y|\mathbf{Z}, X]$ by $G(\mathbf{Z}, X)$. When the alternative hypothesis (3) holds, the true model can be written as

$$Y = G(\mathbf{Z}, X) + \varepsilon,$$

where

$$\inf_{\beta\in\mathbb{R}^p,\alpha(t)\in\mathbf{L}^2[0,1]} \mathbb{E}[G(\mathbf{Z}, X) - \mathbf{Z}^\top\beta - \int_0^1 \alpha(t)X(t)dt]^2 > 0.$$

We rewrite the above model as

$$Y = \mathbf{Z}^\top\beta + \int_0^1 \alpha(t)X(t)dt + \mathcal{D}^*(X, \mathbf{Z}) + \varepsilon,$$

where

$$\mathcal{D}^*(X, \mathbf{Z}) = G(\mathbf{Z}, X) - \mathbf{Z}^\top\beta - \int_0^1 \alpha(t)X(t)dt.$$

It holds that

$$\inf_{\beta\in\mathbb{R}^p,\alpha(t)\in\mathbf{L}^2[0,1]} \mathbb{E}[\mathcal{D}^*(X, \mathbf{Z})]^2 > 0.$$

We further rewrite the model as follows:

$$Y = \mathbf{Z}^\top\beta - \int_0^1 \alpha(t)X(t)dt + n^{-1/2}\varphi(h)^{-1/4}\mathcal{D}^{**}(X, \mathbf{Z}) + \varepsilon,$$

where $\mathcal{D}^{**}(X, \mathbf{Z}) = n^{1/2}\varphi(h)^{1/4}\mathcal{D}^*(X, \mathbf{Z})$. Similarly to the proofs of Theorems 1 and 2, we obtain that $n\varphi(h)^{1/2}\mathcal{T}_n$ has an asymptotic mean of $\mathcal{M}_1\mathbb{E}[f_{\tilde{z}}(\tilde{\mathbf{Z}})f_x(X)\mathcal{D}^{**}(\tilde{\mathbf{V}})\mathcal{D}^{**}(\mathbf{Z}, X)]$, where $\mathcal{D}^{**}(\tilde{\mathbf{V}}) = \mathbb{E}[\mathcal{D}^{**}(\mathbf{Z}, X)|\tilde{\mathbf{V}}]$. Under the alternative hypothesis (3), it is easy to verify that

$$\mathcal{M}_1\mathbb{E}[f_{\tilde{z}}(\tilde{\mathbf{Z}})f_x(X)\mathcal{D}^{**}(\tilde{\mathbf{V}})\mathcal{D}^{**}(\mathbf{Z}, X)] \to \infty,$$

as $n \to \infty$. This completes the proof of Corollary 1.

■

References

Boente, G. and Vahnovan, A. (2017). Robust estimators in semi-functional partial linear regression models. *Journal of Multivariate Analysis*, **154**, 59-84.

Bücher, A., Dette, H. and Wieczorek, G. (2011). Testing model assumptions in functional regression models. *Journal of multivariate analysis*, **102**, 1472-1488.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, **30**, 241-255.

Cardot, H., Goia, A. and Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Communications in Statistics-Simulation and Computation*, **33**, 179-199.

Cuesta-Albertos, J. A., García-Portugués, E., Febrero-Bande, M. and González-Manteiga, W. (2019). Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. *The Annals of Statistics*, **47**, 439-467.

Delsol, L., Ferraty, F. and Vieu, P. (2011). Structural test in regression on functional variables. *Journal of Multivariate Analysis*, **102**, 422-447.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *The Annals of Statistics*, **27**, 1012-1040.

Enno, M. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *The Annals of Statistics*, **21**, 255-285.

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

Ferraty, F., Mas, A. and Vieu, P. (2007a). Nonparametric regression on functional data: Inference and practical aspects. *Australian and New Zealand Journal of Statistics*, **49**, 267-286.

Ferraty, F. and Vieu, P. (2007b). *Nonparametric Functional Data Analysis*, New York: Springer.

Ferraty, F., Zullo, A. and Fauvel, M. (2019). Nonparametric regression on contaminated functional predictor with application to hyperspectral data. *Econometrics and statistics*, **9**, 95-107.

García-Portugués, E., González-Manteiga, W. and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, **23**, 761-778.

Guo, X. and Zhu, L. (2017). *A review on dimension-reduction based tests for regressions.* In: From statistics to mathematical finance. Springer, Cham.

Härdle, W. and Mammen, E. (1990). Comparing nonparametric versus parametric regression fit. *The Annals of Statistics*, **21**, 1926-1947.

Härdle, W., Mammen, E. and Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, **93**, 1461-1474.

Hilgert, N., Mas, A. and Verzelen, N. (2013). Minimax adaptive tests for the functional linear model. *The Annals of Statistics*, **31**, 838-869.

Horváth, L. and Reeder, R. (2013). A test of significance in functional quadratic regression. *Bernoulli*, **19**, 2120-2151.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, John Wiley & Sons.

Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, **87**, 145-165.

Liu, R. (1988). Bootstrap procedures under some non i.i.d. models. *The Annals of Statistics*, **16**, 1696-1708.

Louani, D. (2010). Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *Journal of Multivariate Analysis*, **10**, 2266-2281.

Ma, S. , Zhang, J., Sun, Z. and 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., Hart, J. D., Janicki, R. and 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.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, **2**, 321-359.

Patilea, V. and Sánchez-Sellero, C. (2020). Testing for lack-of-fit in functional regression models against general alternatives. *Journal of Statistical Planning and Inference*, **209**, 229-251.

Radchenko, P., Qiao, X. and James, G. M. (2015). Index models for sparsely sampled functional data. *Journal of the American Statistical Association*, **110**, 824-836.

Ramsay, J. O. and Silverman, B. W. (2007). *Applied functional data analysis: methods and case studies*, New York: Springer.

Reiss, P. T., Goldsmith, J., Shang, H. and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review*, **85**, 228-249.

Sepanski, J. H. and Lee, L.-F. (1995). Semiparametric estimation of nonlinear errors-in-variables models with validation study. *Journaltitle of Nonparametric Statistics*, **4**, 365-394.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference*, **139**, 3405-3418.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, **25**, 613-641.

Sun, L., Zhang, Z. and Sun, J. (2006). Additive hazards regression of failure time data with covariate measurement errors. *Statistica Neerlandica*, **60**, 497-509.

Sun, Z., Chen, F., Liang, H., and Ruppert, D. (2021). A projection-based consistent test incorporating dimension-reduction in partially linear models. *Statistica Sinica*, **31**, 1489-1508.

Tekwe, C. D., Zoh, R. S., Yang, M., Carroll, R. J., Honvoh, G., Allison, D. B., Benden, M. and Xue, L. (2019). Instrumental variable approach to estimating the scalar-on-function regression model with measurement error with application to energy expenditure assessment in childhood obesity. *Statistics in Medicine*, **38**, 3764-3781.

Wang, M., Liu, C., Xie, T. and Sun, Z. (2020). Data-driven model checking for errors-in-variables varying-coefficient models with replicate measurements. *Computational Statistics & Data Analysis*, **141**, 12-27.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, **75**, 263-289.

Zhu, H., Zhang, R. and Li, H. (2019a). Estimation on semi-functional linear errors-in-variables models. *Communications in Statistics-Theory and Methods*, **48**, 4380-4393.

Zhu, H., Zhang, R., Yu, Z., Lian, H. and Liu, Y. (2019b). Estimation and testing for partially functional linear errors-in-variables models. *Journal of Multivariate Analysis*, **170**, 296-314.

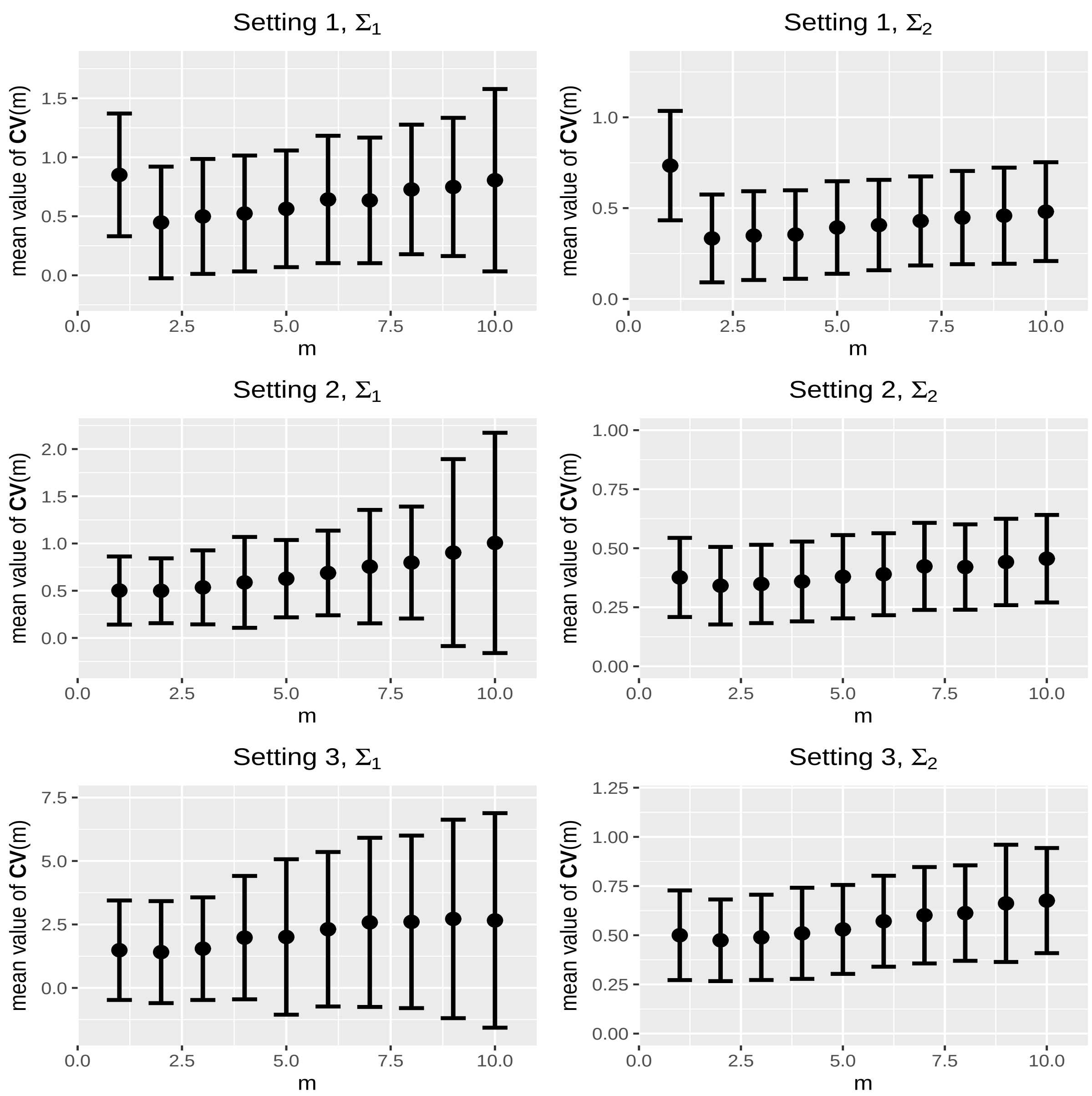

Figure 1: Dot graphs with error bars of the values of the CV function versus the nuisance parameter m in Settings 1–3 for $n = 100$, $\delta = 0$, $\Sigma_u = \Sigma_1$ or Σ_2.

Table 1: Frequencies of rejecting the null hypothesis in Setting 1 under different sample sizes and test levels with $\mathcal{D}(\mathbf{Z}, X) = 3\|X\|^2$ in Case 1, and $\mathcal{D}(\mathbf{Z}, X) = 1.5|Z_1 - Z_2| \cdot \|X\|$ in Case 2. $\mathcal{T}_n^{BE}$: the test for accurately observed data; $\mathcal{T}_n^{KN}$: the proposed test with known covariance matrix of the measurement error variable; $\mathcal{T}_n^{UN}$: the proposed test with unknown covariance matrix of the measurement error variable; $\mathcal{T}_n^{NA}$: the naive test.

		Σ_u	$n = 100$					$n = 200$				
	Test	δ_n	0	1	1.5	2	2.5	0	1	1.5	2	2.5
$\alpha = 0.05$	$\mathcal{T}_n^{BE}$		**0.058**	0.944	0.960	0.962	0.966	**0.050**	0.990	0.992	0.996	0.998
	$\mathcal{T}_n^{KN}$	Σ_1	**0.046**	0.348	0.528	0.650	0.708	**0.048**	0.506	0.740	0.852	0.914
		Σ_2	**0.054**	0.526	0.718	0.808	0.844	**0.052**	0.760	0.936	0.952	0.972
Case 1	$\mathcal{T}_n^{UN}$	Σ_1	**0.060**	0.356	0.512	0.582	0.712	**0.052**	0.504	0.730	0.834	0.910
		Σ_2	**0.048**	0.540	0.750	0.798	0.860	**0.050**	0.746	0.876	0.960	0.978
	$\mathcal{T}_n^{NA}$	Σ_1	**0.196**	0.346	0.528	0.614	0.676	**0.292**	0.562	0.800	0.848	0.894
		Σ_2	**0.096**	0.522	0.666	0.760	0.856	**0.136**	0.704	0.906	0.950	0.962
	$\mathcal{T}_n^{BE}$		**0.064**	0.994	0.996	0.996	0.998	**0.046**	0.998	1.000	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.048**	0.262	0.420	0.522	0.606	**0.044**	0.388	0.618	0.696	0.794
		Σ_2	**0.050**	0.500	0.712	0.806	0.834	**0.054**	0.726	0.866	0.932	0.966
Case 2	$\mathcal{T}_n^{UN}$	Σ_1	**0.054**	0.262	0.432	0.528	0.586	**0.052**	0.370	0.522	0.734	0.776
		Σ_2	**0.048**	0.524	0.674	0.778	0.834	**0.050**	0.698	0.858	0.956	0.962
	$\mathcal{T}_n^{NA}$	Σ_1	**0.194**	0.234	0.412	0.442	0.532	**0.322**	0.390	0.594	0.718	0.770
		Σ_2	**0.086**	0.400	0.610	0.710	0.762	**0.126**	0.582	0.790	0.900	0.958
$\alpha = 0.10$	$\mathcal{T}_n^{BE}$		**0.102**	0.972	0.984	0.988	0.990	**0.096**	0.996	0.998	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.106**	0.466	0.600	0.736	0.790	**0.096**	0.572	0.808	0.920	0.944
		Σ_2	**0.102**	0.630	0.820	0.884	0.922	**0.108**	0.834	0.930	0.970	0.986
Case 1	$\mathcal{T}_n^{UN}$	Σ_1	**0.116**	0.448	0.642	0.706	0.798	**0.098**	0.564	0.788	0.906	0.946
		Σ_2	**0.106**	0.644	0.824	0.906	0.936	**0.104**	0.834	0.944	0.972	0.984
	$\mathcal{T}_n^{NA}$	Σ_1	**0.268**	0.514	0.676	0.720	0.804	**0.376**	0.670	0.814	0.918	0.958
		Σ_2	**0.170**	0.672	0.796	0.904	0.944	**0.190**	0.812	0.946	0.980	0.982
	$\mathcal{T}_n^{BE}$		**0.114**	0.996	1.000	1.000	1.000	**0.100**	1.000	1.000	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.094**	0.390	0.544	0.582	0.692	**0.096**	0.486	0.702	0.798	0.864
		Σ_2	**0.104**	0.626	0.770	0.812	0.874	**0.104**	0.764	0.944	0.952	0.978
Case 2	$\mathcal{T}_n^{UN}$	Σ_1	**0.098**	0.374	0.490	0.576	0.628	**0.104**	0.472	0.668	0.778	0.876
		Σ_2	**0.114**	0.596	0.758	0.840	0.872	**0.100**	0.804	0.912	0.954	0.972
	$\mathcal{T}_n^{NA}$	Σ_1	**0.286**	0.362	0.490	0.602	0.606	**0.388**	0.540	0.648	0.788	0.844
		Σ_2	**0.154**	0.510	0.722	0.784	0.868	**0.224**	0.704	0.896	0.956	0.960

Table 2: Frequencies of rejecting the null hypothesis in Setting 2 under different sample sizes and test levels with $\mathcal{D}(\mathbf{Z}, X) = 1.2\big(\exp(\|X\|) - 1\big)$ in Case 1, and $\mathcal{D}(\mathbf{Z}, X) = 2\exp(|Z_4 - 1|)/3$ in Case 2. The meanings of $\mathcal{T}_n^{BE}$, $\mathcal{T}_n^{KN}$, $\mathcal{T}_n^{UN}$ and $\mathcal{T}_n^{NA}$ are the same as those in Table 1.

		Σ_u	$n = 100$					$n = 200$				
	Test	δ_n	0	1	1.5	2	2.5	0	1	1.5	2	2.5
$\alpha = 0.05$	$\mathcal{T}_n^{BE}$		**0.046**	0.980	0.988	0.994	0.998	**0.048**	0.998	1.000	1.000	1.000
Case 1	$\mathcal{T}_n^{KN}$	Σ_1	**0.060**	0.214	0.520	0.672	0.734	**0.054**	0.542	0.854	0.940	0.974
		Σ_2	**0.056**	0.492	0.790	0.898	0.922	**0.052**	0.842	0.986	0.996	1.000
	$\mathcal{T}_n^{UN}$	Σ_1	**0.040**	0.264	0.496	0.652	0.772	**0.042**	0.524	0.828	0.926	0.984
		Σ_2	**0.046**	0.528	0.798	0.884	0.938	**0.052**	0.802	0.990	0.992	0.998
	$\mathcal{T}_n^{NA}$	Σ_1	**0.198**	0.760	0.918	0.962	0.968	**0.402**	0.992	0.998	1.000	1.000
		Σ_2	**0.096**	0.772	0.948	0.970	0.982	**0.136**	0.986	0.998	0.998	1.000
Case 2	$\mathcal{T}_n^{BE}$		**0.044**	0.990	1.000	1.000	1.000	**0.050**	1.000	1.000	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.038**	0.326	0.646	0.858	0.934	**0.058**	0.666	0.944	0.998	1.000
		Σ_2	**0.040**	0.620	0.896	0.982	0.994	**0.054**	0.948	0.998	1.000	1.000
	$\mathcal{T}_n^{UN}$	Σ_1	**0.060**	0.280	0.614	0.834	0.896	**0.042**	0.648	0.952	1.000	1.000
		Σ_2	**0.056**	0.586	0.894	0.986	0.998	**0.052**	0.940	1.000	1.000	1.000
	$\mathcal{T}_n^{NA}$	Σ_1	**0.204**	0.662	0.922	0.990	0.996	**0.442**	0.974	1.000	1.000	1.000
		Σ_2	**0.110**	0.750	0.970	0.994	0.998	**0.174**	0.980	1.000	1.000	1.000
$\alpha = 0.10$	$\mathcal{T}_n^{BE}$		**0.108**	0.990	0.996	0.998	1.000	**0.100**	1.000	1.000	1.000	1.000
Case 1	$\mathcal{T}_n^{KN}$	Σ_1	**0.114**	0.348	0.598	0.750	0.864	**0.108**	0.670	0.834	0.968	0.984
		Σ_2	**0.092**	0.630	0.824	0.938	0.970	**0.104**	0.894	0.994	0.996	1.000
	$\mathcal{T}_n^{UN}$	Σ_1	**0.090**	0.390	0.586	0.716	0.806	**0.092**	0.626	0.890	0.972	0.982
		Σ_2	**0.106**	0.584	0.848	0.924	0.950	**0.098**	0.866	0.994	0.998	1.000
	$\mathcal{T}_n^{NA}$	Σ_1	**0.250**	0.860	0.942	0.978	0.990	$\boxed{\mathbf{0.466}}$	0.988	0.998	1.000	1.000
		Σ_2	**0.178**	0.876	0.954	0.976	0.998	**0.252**	0.988	0.998	1.000	1.000
Case 2	$\mathcal{T}_n^{BE}$		**0.106**	0.996	1.000	1.000	1.000	**0.104**	1.000	1.000	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.108**	0.366	0.698	0.894	0.950	**0.094**	0.762	0.974	1.000	1.000
		Σ_2	**0.106**	0.694	0.932	0.992	0.996	**0.102**	0.962	1.000	1.000	1.000
	$\mathcal{T}_n^{UN}$	Σ_1	**0.112**	0.388	0.698	0.896	0.942	**0.108**	0.720	0.972	0.998	1.000
		Σ_2	**0.108**	0.700	0.958	0.992	0.994	**0.106**	0.968	1.000	1.000	1.000
	$\mathcal{T}_n^{NA}$	Σ_1	**0.280**	0.786	0.950	0.992	1.000	$\boxed{\mathbf{0.470}}$	0.982	1.000	1.000	1.000
		Σ_2	**0.154**	0.836	0.974	1.000	1.000	**0.226**	0.998	1.000	1.000	1.000

Table 3: Frequencies of rejecting the null hypothesis in Setting 3 under different sample sizes and test levels with $\mathcal{D}(\mathbf{Z}, X) = 4\log(1 + \|X\|^2)$ in Case 1, and $\mathcal{D}(\mathbf{Z}, X) = Z_8^3/4$ in Case 2. The meanings of $\mathcal{T}_n^{BE}$, $\mathcal{T}_n^{KN}$, $\mathcal{T}_n^{UN}$ and $\mathcal{T}_n^{NA}$ are the same as those in Table 1.

		Σ_u	$n = 100$					$n = 200$				
	Test	δ_n	0	1	1.5	2	2.5	0	1	1.5	2	2.5
$\alpha = 0.05$	$\mathcal{T}_n^{BE}$		**0.044**	0.974	0.998	0.998	1.000	**0.052**	0.998	1.000	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.052**	0.178	0.366	0.504	0.584	**0.048**	0.492	0.796	0.946	0.968
		Σ_2	**0.044**	0.470	0.766	0.850	0.926	**0.054**	0.802	0.974	0.992	0.996
Case 1	$\mathcal{T}_n^{UN}$	Σ_1	**0.046**	0.188	0.324	0.432	0.514	**0.048**	0.480	0.770	0.892	0.932
		Σ_2	**0.056**	0.460	0.688	0.864	0.920	**0.044**	0.806	0.974	0.982	0.998
	$\mathcal{T}_n^{NA}$	Σ_1	**0.116**	0.656	0.864	0.930	0.970	**0.204**	0.916	0.992	0.998	1.000
		Σ_2	**0.072**	0.730	0.892	0.968	0.980	**0.088**	0.958	0.998	1.000	1.000
	$\mathcal{T}_n^{BE}$		**0.060**	0.910	0.974	0.982	0.988	**0.058**	0.998	0.998	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.060**	0.240	0.512	0.732	0.756	**0.044**	0.612	0.942	0.986	0.994
		Σ_2	**0.040**	0.524	0.820	0.898	0.932	**0.046**	0.876	0.996	0.996	1.000
Case 2	$\mathcal{T}_n^{UN}$	Σ_1	**0.048**	0.234	0.450	0.594	0.668	**0.058**	0.598	0.934	0.988	0.996
		Σ_2	**0.060**	0.482	0.792	0.912	0.922	**0.052**	0.894	0.996	0.998	1.000
	$\mathcal{T}_n^{NA}$	Σ_1	**0.118**	0.398	0.776	0.918	0.960	**0.184**	0.760	0.998	0.998	1.000
		Σ_2	**0.066**	0.548	0.866	0.944	0.956	**0.096**	0.886	0.998	1.000	1.000
$\alpha = 0.10$	$\mathcal{T}_n^{BE}$		**0.104**	0.990	0.998	1.000	1.000	**0.092**	1.000	1.000	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.112**	0.306	0.416	0.582	0.666	**0.104**	0.602	0.862	0.946	0.990
		Σ_2	**0.096**	0.628	0.826	0.912	0.940	**0.102**	0.900	0.984	0.996	1.000
Case 1	$\mathcal{T}_n^{UN}$	Σ_1	**0.110**	0.282	0.412	0.510	0.610	**0.104**	0.534	0.842	0.936	0.964
		Σ_2	**0.100**	0.580	0.804	0.900	0.936	**0.102**	0.886	0.994	1.000	1.000
	$\mathcal{T}_n^{NA}$	Σ_1	**0.180**	0.758	0.910	0.970	0.986	**0.328**	0.958	0.998	1.000	1.000
		Σ_2	**0.142**	0.788	0.954	0.984	0.994	**0.172**	0.970	1.000	1.000	1.000
$\alpha = 0.10$	$\mathcal{T}_n^{BE}$		**0.110**	0.968	0.992	0.992	0.994	**0.100**	0.998	0.998	1.000	1.000
	$\mathcal{T}_n^{KN}$	Σ_1	**0.112**	0.318	0.600	0.776	0.838	**0.098**	0.698	0.966	0.992	1.000
		Σ_2	**0.094**	0.610	0.916	0.948	0.964	**0.104**	0.946	1.000	1.000	1.000
Case 2	$\mathcal{T}_n^{UN}$	Σ_1	**0.110**	0.318	0.564	0.722	0.770	**0.094**	0.692	0.966	0.994	1.000
		Σ_2	**0.112**	0.598	0.854	0.942	0.974	**0.096**	0.954	0.998	1.000	1.000
	$\mathcal{T}_n^{NA}$	Σ_1	**0.188**	0.506	0.830	0.956	0.974	**0.284**	0.840	0.992	1.000	1.000
		Σ_2	**0.142**	0.694	0.904	0.978	0.988	**0.164**	0.918	1.000	1.000	1.000

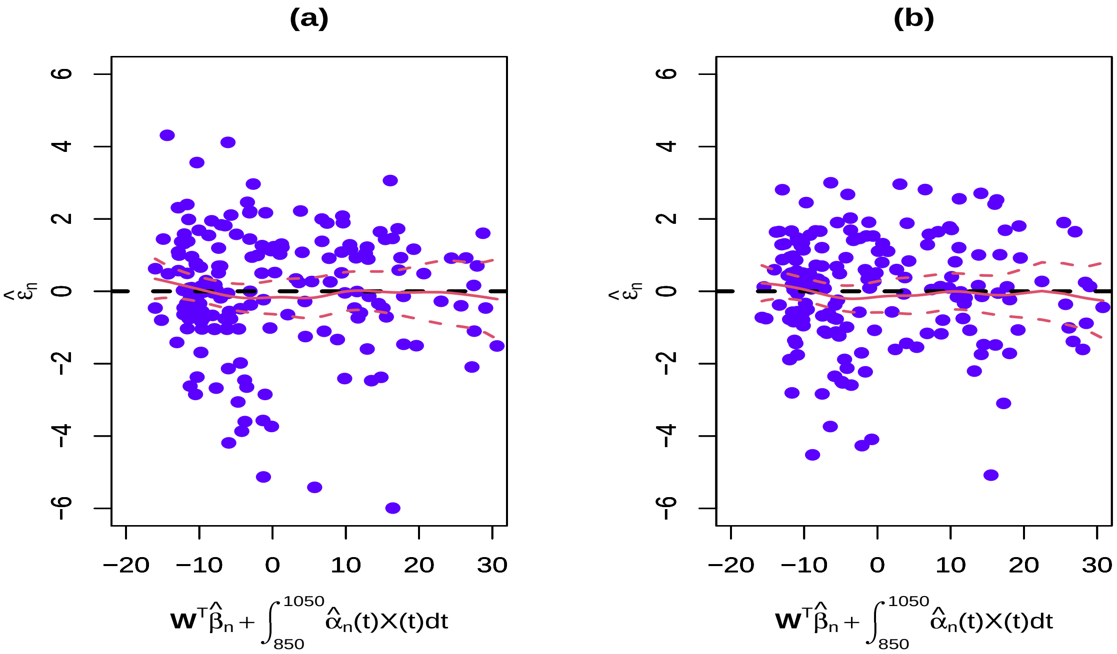

Figure 2: (a) Scatter plot of the calibrated model error estimator $\hat{\varepsilon}_n$ versus $\mathbf{W}^\top\hat{\beta}_n + \int_{850}^{1050} \hat{\alpha}_n(t)X(t)dt$ for Tecator data with Σ_u =diag(3, 0) by treating Σ_u as known; (b) Scatter plot of the calibrated model error estimator $\hat{\varepsilon}_n$ versus $\mathbf{W}^\top\hat{\beta}_n + \int_{850}^{1050} \hat{\alpha}_n(t)X(t)dt$ for Tecator data with Σ_u =diag(3, 0) by treating Σ_u as unknown.