Chin. Ann. Math.
31B(2), 2010, 247–272
DOI: 10.1007/s11401-008-0153-3

Chinese Annals of Mathematics, Series B
© The Editorial Office of CAM and Springer-Verlag Berlin Heidelberg 2010

Weighted Profile Least Squares Estimation for a Panel Data Varying-Coefficient Partially Linear Model*****

Bin ZHOU*   Jinhong YOU**   Qinfeng XU***   Gemai CHEN****

Abstract This paper is concerned with inference of panel data varying-coefficient partially linear models with a one-way error structure. The model is a natural extension of the well-known panel data linear model (due to Baltagi 1995) to the setting of semiparametric regressions. The authors propose a weighted profile least squares estimator (WPLSE) and a weighted local polynomial estimator (WLPE) for the parametric and nonparametric components, respectively. It is shown that the WPLSE is asymptotically more efficient than the usual profile least squares estimator (PLSE), and that the WLPE is also asymptotically more efficient than the usual local polynomial estimator (LPE). The latter is an interesting result. According to Ruckstuhl, Welsh and Carroll (2000) and Lin and Carroll (2000), ignoring the correlation structure entirely and "pretending" that the data are really independent will result in more efficient estimators when estimating nonparametric regression with longitudinal or panel data. The result in this paper shows that this is not true when the design points of the nonparametric component have a closeness property within groups. The asymptotic properties of the proposed weighted estimators are derived. In addition, a block bootstrap test is proposed for the goodness of fit of models, which can accommodate the correlations within groups. Some simulation studies are conducted to illustrate the finite sample performances of the proposed procedures.

Keywords Semiparametric, Panel data, Local polynomial, Weighted estimation, Block bootstrap
2000 MR Subject Classification 62H12, 62A10

1 Introduction

Panel data arise frequently in biological and economic applications. Various parametric models and statistical tools have been developed for panel data analysis (see, for instance, [1, 15] and the references therein. It is true that parametric models are very useful for analyzing panel data and for providing a parsimonious description of the relationship between the response variable and its covariates. However, they are often subject to the risk of introducing modeling bias. To relax the assumptions on parametric forms, Ruckstuhl, Welsh and [12] proposed a nonparametric panel data regression model, which allows one to explore possible hidden structures in the data and to reduce modeling bias of the traditional parametric methods. Such nonparametric models, however, have several shortcomings including the curse of

Manuscript received April 22, 2008. Revised May 15, 2009. Published online February 2, 2010.
*School of Finance and Statistics, East China Normal University, Shanghai 200241, China.
**Department of Statistics, Shanghai University of Finance and Economics, Shanghai 200433, China.
***Corresponding Author. School of Management, Fudan University, Shanghai 200433, China. E-mail: qfxu@fudan.edu.cn
****Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, T2N 1N4, Canada.
*****Project supported by the Leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics (the 3rd phase) (No. B803) and the Shanghai Leading Academic Discipline Project (No. B210).

dimensionality, difficulty of interpretation, and lack of extrapolation capability. To overcome these shortcomings, semiparametric panel data regression models have been considered recently which embody a compromise between a general nonparametric specification and a fully parametric specification. In this paper, we focus on a panel data varying-coefficient partially linear model defined as

$$Y_{ij} = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}(U_{ij}) + \varepsilon_{ij}, \quad i = 1, \cdots, k,\ j = 1, \cdots, n_0, \tag{1.1}$$

with a one-way error structure

$$\varepsilon_{ij} = \mu_i + \nu_{ij}, \quad i = 1, \cdots, k,\ j = 1, \cdots, n_0. \tag{1.2}$$

In (1.1), $Y_{i1}, \cdots, Y_{in_0}$ represent repeated response measurements from individual i, $\mathbf{X}_{ij} = (X_{ij1}, \cdots, X_{ijp})^{\mathrm{T}}$, $\mathbf{Z}_{ij} = (Z_{ij1}, \cdots, Z_{ijq})^{\mathrm{T}}$and U_{ij} are referred to as the design points, $\boldsymbol{\beta}$ is an unknown p-dimensional vector of regression coefficients for the parametric component, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_q(\cdot))^{\mathrm{T}}$ are unknown functions to model the nonlinear and nonparametric components, and T denotes the transpose of a vector or matrix. In (1.2), μ_i and ν_{ij} are i.i.d. random variables with mean zero and variances σ_μ^2 and σ_ν^2, respectively. Moreover, μ_i and ν_{ij} are independent. With the one-way error structure (1.2), the observations $Y_{i1}, \cdots, Y_{in_0}$ from the same individual i share a common variable μ_i and thus are allowed to be dependent.

Obviously, model (1.1)–(1.2) includes many usual parametric, semiparametric and nonparametric regression models. When $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_q(\cdot))^{\mathrm{T}} \equiv (a_1, \cdots, a_q)^{\mathrm{T}}$, model (1.1)–(1.2) reduces to the traditional panel data linear model (see [1]), where $a_1, \cdots, a_q$ are unknown constants. When $q = 1$ and $\mathbf{Z}_{ij} = 1$, model (1.1)–(1.2) becomes the panel data partially linear model. Zeger and Diggle [19] used this model to study CDE cell number in HIV seroconverters. Other applications can be found in [8]. From the form of model (1.1), we can see that it permits the interaction between the covariates U and $\mathbf{Z}$ in such a way that a different level of covariate U is associated with a different linear model. This allows one to examine the extent to which the effect of covariate $\mathbf{Z}$ varies over different levels of the covariate U, making model (1.1)–(1.2) more flexible than the panel data partially linear model. When $n_0 = 1$, which corresponds to independent errors, model (1.1) reduces to the non-panel varying-coefficient partially linear model, which has been widely studied in the literature (see, for example, [9, 20]). When $p = 0$, model (1.1)–(1.2) becomes the panel data varying-coefficient regression, which has been studied by Fan and Zhang [5], Huang, Wu and Zhou [7] among others. Moreover, the nonparametric panel data regression model in [12] is a special case of model (1.1)–(1.2) as well.

In this paper, we study the inference for the panel data varying-coefficient partially linear model (1.1)–(1.2). We propose a weighted profile least squares estimator (WPLSE) of the parametric component and a weighted local polynomial estimator (WLPE) of the nonparametric component when random effects are present. We show that the WPLSE is asymptotically more efficient than the usual profile least squares estimator (PLSE), and that WLPE is also asymptotically more efficient than the usual local polynomial estimator (LPE). The latter is an interesting finding. Ruckstuhl, Welsh and Carroll [12] and Lin and Carroll [10] found that when estimating nonparametric regression with longitudinal or panel data, ignoring the correlation structure entirely and "pretending" that the data are really independent will result in more efficient estimators. Our finding shows that this is not always true. When the design points of the nonparametric component have a closeness property within groups, the WLPE has the same asymptotic bias as the LPE but smaller asymptotic covariance matrix. Here the closeness property means that $\max_{1\le i\le k} \big| \max_{1\le j\le n_0} U_{ij} - \min_{1\le j\le n_0} U_{ij} \big| = O(k^{-1}\log k)$. The following are two examples in which the closeness property is satisfied.

Example 1.1 U_{ij}'s are generated as $U_{ij} = F^{-1}\big(\frac{(i-1)k+j}{kn_0-1}\big)$, $i = 1, \cdots, k$, $j = 1, \cdots, n_0$, where $F(\cdot)$ is a distribution function on $\mathcal{U}$ which has a Lipschitz continuous density $p(\cdot) = \frac{\partial F(\cdot)}{\partial u}$ satisfying $0 < \inf\limits_{\mathcal{U}} p(\cdot) \le \sup\limits_{\mathcal{U}} p(\cdot) < \infty$, where $\mathcal{U}$ is the bounded support of U. According to [11], it holds that

$$O\Big\{\frac{\log(kn_0)}{kn_0}\Big\} \le \inf_{1\le i\le k}\inf_{2\le j\le n_0} |U_{ij} - U_{ij-1}| \le \sup_{1\le i\le k}\sup_{2\le j\le n_0} |U_{ij} - U_{ij-1}| = O\Big\{\frac{\log(kn_0)}{kn_0}\Big\}.$$

Example 1.2 Covariate U is measured repeatedly, that is, $U_{i1} = U_{i2} = \cdots = U_{in_0}$ for $i = 1, \cdots, k$.

An important issue in fitting model (1.1)–(1.2) is whether there exist parametric structures for $\alpha_j(\cdot)$, $j = 1, \cdots, q$. This amounts to testing if $\alpha_j(\cdot)$'s are with a certain parametric form. A test is proposed based on the comparison of the residual sum of squares under the null and alternative models, and a block bootstrap method is used to find the null distribution of the test statistic. This block bootstrap test can accommodate the one-way error structure. Our simulation shows that the resulting testing procedure is indeed powerful and the bootstrap method does give the right null distribution.

The rest of the paper is organized as follows. The unweighted profile least squares and local polynomial estimations of the parametric and nonparametric components are presented in Section 2. A deleting block cross-validation method for bandwidth selection is proposed in Section 3. Estimations of the error variances are discussed in Section 4. A weighted profile least squares estimator of the parametric component and a weighted local polynomial estimator of the nonparametric component are proposed in Section 5. A block bootstrap test for the goodness of fit of model (1.1)–(1.2) is developed in Section 6. Some simulation studies are conducted in Section 7. Some remarks are given in Section 8. Proofs of the main results are relegated to the Appendix.

2 Unweighted Profile Least Squares and Local Linear Estimation

Throughout this paper, we assume large k and small n_0. This is typical for labor or consumer panel data (see [11]). We also assume that the design points $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ are random and U_{ij} are fixed as in [6, 16] etc. Extending our results to the case of both $(\mathbf{X}_{ij}, \mathbf{Z}_{ij})$ and U_{ij} being fixed or random is conceptually straightforward.

If we neglect the error structure, we can construct a profile least squares estimator (PLSE) of the parametric component $\boldsymbol{\beta}$. Assume that $\{\mathbf{X}_{ij}^{\mathrm{T}}, \mathbf{Z}_{ij}^{\mathrm{T}}, U_{ij}, Y_{ij};\ i = 1, \cdots, k,\ j = 1, \cdots, n_0\}$ satisfy model (1.1). For any given $\boldsymbol{\beta}$, model (1.1) can be written as

$$Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}(U_{ij}) + \varepsilon_{ij}, \quad i = 1, \cdots, k,\ j = 1, \cdots, n_0, \tag{2.1}$$

which is a version of the usual varying-coefficient regression model. Among the various procedures available, we use a local linear regression technique to estimate the varying coefficient functions $\{\alpha_s(\cdot), s = 1, \cdots, q\}$ in (2.1). For u in a small neighborhood of u_0, we approximate $\alpha_s(u)$ locally by a linear function

$$\alpha_s(u) \approx \alpha_s(u_0) + \alpha_s'(u_0)(u - u_0) \equiv a_s + b_s(u - u_0), \quad s = 1, \cdots, q,$$

where $\alpha_s'(u) = \frac{\partial \alpha_s(u)}{\partial u}$. This leads to the following weighted local least squares problem: find

$\{(a_s, b_s), s = 1, \cdots, q\}$ to minimize

$$\sum_{i=1}^{k}\sum_{j=1}^{n_0}\Big[(Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}) - \sum_{s=1}^{q}\{a_s + b_s(U_{ij} - u_0)\}Z_{ijs}\Big]^2 K_h(U_{ij} - u_0), \tag{2.2}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(\cdot) = \frac{K(\frac{\cdot}{h})}{h}$. The solution to problem (2.2) is given by

$$(\widehat{\alpha}_1(u), \cdots, \widehat{\alpha}_q(u), \widehat{b}_1(u), \cdots, \widehat{b}_q(u))^{\mathrm{T}} = (\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),$$

where

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_0} \\ \vdots \\ Y_{kn_0} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_{11}^{\mathrm{T}} \\ \vdots \\ \mathbf{X}_{1n_0}^{\mathrm{T}} \\ \vdots \\ \mathbf{X}_{kn_0}^{\mathrm{T}} \end{pmatrix}, \quad \mathbf{D}_u = \begin{pmatrix} \mathbf{Z}_{11}^{\mathrm{T}} & (U_{11} - u)\mathbf{Z}_{11}^{\mathrm{T}} \\ \vdots & \vdots \\ \mathbf{Z}_{1n_0}^{\mathrm{T}} & (U_{1n_0} - u)\mathbf{Z}_{1n_0}^{\mathrm{T}} \\ \vdots & \vdots \\ \mathbf{Z}_{kn_0}^{\mathrm{T}} & (U_{kn_0} - u)\mathbf{Z}_{kn_0}^{\mathrm{T}} \end{pmatrix}$$

and

$$\boldsymbol{\omega}_u = \mathrm{diag}(K_h(U_{11} - u), \cdots, K_h(U_{1n_0} - u), \cdots, K_h(U_{kn_0} - u)).$$

Substituting $(\widehat{\alpha}_1(u), \cdots, \widehat{\alpha}_q(u))^{\mathrm{T}}$ into (2.1) as an estimate of $\boldsymbol{\alpha}(u)$, we obtain

$$\widehat{Y}_{ij} = \widehat{\mathbf{X}}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_{ij}^{\star}, \quad i = 1, \cdots, k, \ j = 1, \cdots, n_0, \tag{2.3}$$

where $\widehat{\mathbf{Y}} = (\widehat{Y}_{11}, \cdots, \widehat{Y}_{1n_0}, \cdots, \widehat{Y}_{kn_0})^{\mathrm{T}} = (\mathbf{I}_{kn_0} - \mathbf{S})\mathbf{Y}$, $\widehat{\mathbf{X}} = (\widehat{\mathbf{X}}_{11}, \cdots, \widehat{\mathbf{X}}_{1n_0}, \cdots, \widehat{\mathbf{X}}_{kn_0})^{\mathrm{T}} = (\mathbf{I}_{kn_0} - \mathbf{S})\mathbf{X}$, $\boldsymbol{\varepsilon}^{\star} = (\varepsilon_{11}^{\star}, \cdots, \varepsilon_{1n_0}^{\star}, \cdots, \varepsilon_{kn_0}^{\star})^{\mathrm{T}} = (\mathbf{I}_{kn_0} - \mathbf{S})\boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \cdots, \varepsilon_{1n_0}, \cdots, \varepsilon_{kn_0})^{\mathrm{T}}$,

$$\mathbf{S} = \begin{pmatrix} (\mathbf{Z}_{11}^{\mathrm{T}} \quad \mathbf{0}_q^{\mathrm{T}})(\mathbf{D}_{U_{11}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{11}}\mathbf{D}_{U_{11}})^{-1}\mathbf{D}_{U_{11}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{11}} \\ \vdots \\ (\mathbf{Z}_{1n_0}^{\mathrm{T}} \quad \mathbf{0}_q^{\mathrm{T}})(\mathbf{D}_{U_{1n_0}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{1n_0}}\mathbf{D}_{U_{1n_0}})^{-1}\mathbf{D}_{U_{1n_0}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{1n_0}} \\ \vdots \\ (\mathbf{Z}_{kn_0}^{\mathrm{T}} \quad \mathbf{0}_q^{\mathrm{T}})(\mathbf{D}_{U_{kn_0}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{kn_0}}\mathbf{D}_{U_{kn_0}})^{-1}\mathbf{D}_{U_{kn_0}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{kn_0}} \end{pmatrix}$$

and $\mathbf{0}_q$ is a $q \times 1$ vector of zeros.

If we take $\varepsilon_{ij}^{\star}$ as the residual errors, then (2.3) is a version of the ordinary linear regression model. Applying the least squares method to (2.3) results in the following (unweighted) profile least squares estimator (PLSE) of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}_k = \Big(\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widehat{\mathbf{X}}_{ij}\widehat{\mathbf{X}}_{ij}^{\mathrm{T}}\Big)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widehat{\mathbf{X}}_{ij}\widehat{Y}_{ij}. \tag{2.4}$$

Correspondingly, $\boldsymbol{\alpha}(\cdot)$ is estimated by a local linear estimator (LLE)

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_k(u) &= (\mathbf{I}_q, \mathbf{0}_{q\times q})[\widehat{\alpha}_1(u), \cdots, \widehat{\alpha}_q(u), \widehat{b}_1(u), \cdots, \widehat{b}_q(u)]^{\mathrm{T}} \\ &= (\mathbf{I}_q, \mathbf{0}_{q\times q})(\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_k),\end{aligned}$$

where $\mathbf{0}_{q\times q}$ is a $q \times q$ matrix of zeros.

In order to derive the asymptotic properties of $\widehat{\boldsymbol{\beta}}_k$, $\widehat{\boldsymbol{\alpha}}_k(\cdot) = (\widehat{\alpha}_1(\cdot), \cdots, \widehat{\alpha}_q(\cdot))^{\mathrm{T}}$ and other estimators proposed in the following sections, we make the following assumptions. These assumptions are quite mild and can be easily satisfied. They are also used by Fan and Huang [4] except Assumption 2.1(ii).

Assumption 2.1 (i) *U_{ij}'s are generated from a distribution with bounded support $\mathcal{U}$ and density function $p(\cdot)$ that is Lipschitz continuous and satisfies $0 < \inf_{u\in\mathcal{U}} p(u) \le \sup_{u\in\mathcal{U}} p(u) < \infty$.*

(ii) *U_{ij}'s satisfy a closeness property within groups, namely,*

$$\max_{1\le i\le k}\Big|\max_{1\le j\le n_0} U_{ij} - \min_{1\le j\le n_0} U_{ij}\Big| = O(k^{-1}\log k).$$

Assumption 2.2 *$\mathbf{Z}_i = (\mathbf{Z}_{i1},\cdots,\mathbf{Z}_{in_0})^{\mathrm{T}}$ and $\mathbf{X}_i = (\mathbf{X}_{i1},\cdots,\mathbf{X}_{in_0})^{\mathrm{T}}$ $(i=1,\cdots,k)$ are i.i.d. random matrices and $E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_1)$ is non-singular.*

Assumption 2.3 *There is an $s>2$ such that $E\|\mathbf{X}_1\|^{2s} < \infty$ and $E\|\mathbf{Z}_1\|^{2s} < \infty$ and $k^{2\delta-1}h \to \infty$ as $k\to\infty$ for some $\delta < 2 - s^{-1}$.*

Assumption 2.4 *$\{\alpha_s(\cdot),\ s=1,\cdots,q\}$ have continuous second derivatives on $\mathcal{U}$. Let $\boldsymbol{\alpha}'(\cdot) = (\alpha_1'(\cdot),\cdots,\alpha_q'(\cdot))^{\mathrm{T}}$, and $\boldsymbol{\alpha}''(\cdot) = (\alpha_1''(\cdot),\cdots,\alpha_q''(\cdot))^{\mathrm{T}}$.*

Assumption 2.5 *The kernel function $K(\cdot)$ is a density function with compact support and the bandwidth h satisfies $k^{\frac{1}{2}}h^8 \to 0$ and $\frac{kh^2}{(\log k)^2} \to \infty$ as $k\to\infty$.*

Theorem 2.1 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.5 *hold. Then we have*

$$\sqrt{k}(\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}) \to_D N(0, \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}), \quad \text{as } k\to\infty,$$

where $\to_D$ denotes convergence in distribution,

$$\begin{aligned}
\boldsymbol{\Sigma}_1 &= E(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1) - E(\mathbf{X}_1^{\mathrm{T}}\mathbf{Z}_1)(E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_1))^{-1}E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{X}_1),\\
\boldsymbol{\Sigma}_2 &= E[\{\mathbf{X}_1^{\mathrm{T}} - E(\mathbf{X}_1^{\mathrm{T}}\mathbf{Z}_1)(E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_1))^{-1}\mathbf{Z}_1^{\mathrm{T}}\}\cdot(\sigma_\mu^2\boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\mathrm{T}} + \sigma_\nu^2\mathbf{I}_{n_0})\\
&\quad\cdot\{\mathbf{X}_1^{T} - E(\mathbf{X}_1^{\mathrm{T}}\mathbf{Z}_1)(E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_1))^{-1}\mathbf{Z}_1^{\mathrm{T}}\}^{\mathrm{T}}],
\end{aligned}$$

where $\boldsymbol{\iota}_{n_0}$ is an $n_0\times 1$ vector of 1's, and σ_μ^2 and σ_ν^2 are the variances of μ_i and ν_{ij}, respectively.

Theorem 2.2 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.5 *hold. Then we have*

$$\sqrt{kn_0h}\Big\{\widehat{\boldsymbol{\alpha}}_k(u_0) - \boldsymbol{\alpha}(u_0) - \frac{h^2}{2}\frac{(\mu_2^0)^2 - \mu_1^0\mu_3^0}{\mu_2^0 - (\mu_1^0)^2}\boldsymbol{\alpha}''(u_0)\Big\} \to_D N(0,\boldsymbol{\Sigma}_3), \quad \text{as } k\to\infty,$$

provided $p(u_0)\ne 0$, where

$$\boldsymbol{\Sigma}_3 = \frac{(c_0^2\nu_0^0 + 2c_0c_1\nu_1^0 + c_1^2\nu_2^0)n_0}{p(u_0)}(E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_1))^{-1}E\{\mathbf{Z}_1^{\mathrm{T}}(\sigma_\mu^2\boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\mathrm{T}} + \sigma_\nu^2\mathbf{I}_{n_0})\mathbf{Z}_1\}(E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_1))^{-1},$$

$$c_0 = \frac{\mu_2^0}{\mu_2^0 - (\mu_1^0)^2}, \quad c_1 = \frac{-\mu_1^0}{\mu_2^0 - (\mu_1^0)^2}, \quad \mu_j^0 = \int_{-\infty}^{\infty} u^jK(u)\mathrm{d}u, \quad \nu_j^0 = \int_{-\infty}^{\infty} u^jK^2(u)\mathrm{d}u.$$

Theorem 2.3 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.5 *hold. Then we have*

$$\sup_{u\in\mathcal{U}}\|\widehat{\boldsymbol{\alpha}}_k(u) - \boldsymbol{\alpha}(u)\| = O_p\Big\{h^2 + \Big(\frac{\log k}{kh}\Big)^{\frac{1}{2}}\Big\}.$$

Since $\widehat{\boldsymbol{\beta}}_k$ and $\widehat{\boldsymbol{\alpha}}_k(\cdot)$ do not take the correlation within groups into account, they may not be asymptotically efficient. However, according to Theorems 2.1 and 2.2 they are consistent estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}(\cdot)$. Therefore, based on $\widehat{\boldsymbol{\beta}}_k$ and $\widehat{\boldsymbol{\alpha}}_k(\cdot)$ we can estimate the residuals in (1.1) and use them to test the one-way error structure (1.2), to fit the error structure of (1.2), and to construct asymptotically more efficient estimators. In the following sections, we discuss these issues.

3 A Leave-One-Subject Cross-Validation Criterion

The PLSE and LLE depend on the choice of the bandwidth. We note that there exist correlations within groups. Therefore, we propose to use the leave-one-subject cross-validation technique to determine an appropriate value of the bandwidth. Let $\widehat{\boldsymbol{\alpha}}_{h,-i}(\cdot)$ and $\widehat{\boldsymbol{\beta}}_{h,-i}$ be the local linear and profile least squares estimates, omitting the ith subject. Define the delete block squares cross-validation function by

$$CV(h) = (kn_0)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{h,-i} - \mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_{h,-i}(U_{ij}))^2. \tag{3.1}$$

Depending on the bandwidth h, $CV(h)$ is used as an overall measure of the effectiveness of the estimation scheme. The leave-one-subject cross-validation bandwidth selector is the one that minimizes (3.1), namely

$$\widehat{h}_{CV} = \arg\min_h CV(h).$$

4 Estimating the Error Structure

The variances σ_μ^2 and σ_ν^2 describe the noise level. Apart from the intrinsic interest as parameters of the model, estimating them is essential in the construction of efficient estimators of the regression coefficients, confidence regions, model-based tests, model selection procedures, signal-to-noise ratio and so on. Since $E(\varepsilon_{ij_1}\varepsilon_{ij_2}) = \sigma_\mu^2$ when $j_1 \neq j_2$ and $E(\varepsilon_{ij}^2) = \sigma_\mu^2 + \sigma_\nu^2$, we estimate σ_μ^2 and σ_ν^2 by

$$\widehat{\sigma}_\mu^2 = \frac{1}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\neq j_1}\widehat{\varepsilon}_{ij_1}\widehat{\varepsilon}_{ij_2} \quad \text{and} \quad \widehat{\sigma}_\nu^{\,2} = \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widehat{\varepsilon}_{ij}^{\,2} - \widehat{\sigma}_\mu^2, \tag{4.1}$$

respectively, where $\widehat{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k - \mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(U_{ij})$, $i = 1, \cdots, k$ and $j = 1, \cdots, n_0$.

The theorem below provides the asymptotic normality of the estimators $\widehat{\sigma}_\mu^{\,2}$ and $\widehat{\sigma}_\nu^{\,2}$.

Theorem 4.1 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.5 *hold. Then it holds that*

$$k^{\frac{1}{2}}(\widehat{\sigma}_\mu^{\,2} - \sigma_\mu^2) \to_D N\Big(0, \mathrm{Var}(\mu_1^2) + \frac{4\sigma_\mu^2\sigma_\nu^2}{n_0} + \frac{2\sigma_\nu^4}{n_0(n_0-1)}\Big), \quad \textit{as } k \to \infty \tag{4.2}$$

and

$$(kn_0)^{\frac{1}{2}}(\widehat{\sigma}_\nu^{\,2} - \sigma_\nu^2) \to_D N\Big(0, \mathrm{Var}(\nu_{11}^2) + \frac{2\sigma_\nu^4}{n_0-1}\Big), \quad \textit{as } k \to \infty. \tag{4.3}$$

In order to use Theorem 4.1 to make statistical inferences, we need consistent estimators of the asymptotic variances of $\widehat{\sigma}_\mu^{\,2}$ and $\widehat{\sigma}_\nu^{\,2}$.

Define

$$\widehat{\mu}^{\,4} = \frac{1}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\neq j_1}\widehat{\varepsilon}_{ij_1}^{\,2}\widehat{\varepsilon}_{ij_2}^{\,2} \quad \text{and} \quad \widehat{\nu}^{\,4} = \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widehat{\varepsilon}_{ij}^{\,4} - \widehat{\mu}^{\,4}.$$

The following theorem gives consistent estimators of the asymptotic variances of $k^{\frac{1}{2}}(\widehat{\sigma}_\mu^{\,2} - \sigma_\mu^2)$ and $(kn_0)^{\frac{1}{2}}(\widehat{\sigma}_\nu^{\,2} - \sigma_\nu^2)$.

Theorem 4.2 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.5 *hold. Then we have*

$$\widehat{\mu}^{\,4} - \widehat{\sigma}^{\,4}_{\mu} + \Big[\frac{2}{n_0(n_0-1)} - 1\Big]\widehat{\sigma}^{\,4}_{\nu} + \Big(\frac{4}{n_0} - 2\Big)\widehat{\sigma}^{\,2}_{\mu}\widehat{\sigma}^{\,2}_{\nu} \to_p \mathrm{Var}(\mu_1^2) + \frac{4\sigma_\mu^2\sigma_\nu^2}{n_0} + \frac{2\sigma_\nu^4}{n_0(n_0-1)} \tag{4.4}$$

and

$$\widehat{\nu}^{\,4} - 4\widehat{\sigma}^{\,2}_{\nu}\widehat{\sigma}^{\,2}_{\mu} + \frac{2\widehat{\sigma}^{\,4}_{\nu}}{n_0-1} \to_p \mathrm{Var}(\nu_{11}^2) + \frac{2\sigma_\nu^4}{n_0-1}, \tag{4.5}$$

as $k \to \infty$.

Combining Theorem 4.2 with Theorem 4.1, we get the following corollary, which can be used to construct tests and confidence regions for σ_μ^2 and σ_ν^2.

Corollary 4.1 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.5 *hold. Then we have*

$$\frac{k^{\frac{1}{2}}(\widehat{\sigma}^{\,2}_{\mu} - \sigma_\mu^2)}{\sqrt{\widehat{\mu}^{\,4} - \widehat{\sigma}^{\,4}_{\mu} + \widehat{\sigma}^{\,4}_{\nu}[\frac{2}{n_0(n_0-1)} - 1] + \widehat{\sigma}^{\,2}_{\mu}\widehat{\sigma}^2_{\nu}(\frac{4}{n_0} - 2)}} \to_D N(0,1), \quad \textit{as } k \to \infty$$

and

$$\frac{(kn_0)^{\frac{1}{2}}(\widehat{\sigma}^{\,2}_{\nu} - \sigma_\nu^2)}{\sqrt{\widehat{\nu}^{\,4} - 4\widehat{\sigma}^{\,2}_{\nu}\widehat{\sigma}^{\,2}_{\mu} + \frac{2\widehat{\sigma}^{\,4}_{\nu}}{n_0-1}}} \to_D N(0,1), \quad \textit{as } k \to \infty.$$

5 Weighted Profile Least Squares and Local Linear Estimations

The unweighted profile least squares estimator $\widehat{\boldsymbol{\beta}}_k$ and local linear estimator $\widehat{\boldsymbol{\alpha}}_k(\,\cdot\,)$ are not asymptotically efficient since they do not take into account the error structure. In order to construct more efficient estimators, we first derive the inverse of the covariance matrix of $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \cdots, \varepsilon_{1n_0}, \cdots, \varepsilon_{kn_0})$. According to (1.2), we have

$$\boldsymbol{\Omega} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \mathbf{I}_k \otimes (\sigma_\mu^2 \boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\mathrm{T}} + \sigma_\nu^2 \mathbf{I}_{kn_0}), \tag{5.1}$$

where $\otimes$ denotes Kronecker product. Let $\boldsymbol{\varpi}$ be the inverse of $\sigma_\mu^2 \boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\mathrm{T}} + \sigma_\nu^2 \mathbf{I}_{kn_0}$, i.e.,

$$\boldsymbol{\varpi} = \eta^{\,-2}\frac{\boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\mathrm{T}}}{n_0} + \sigma_\nu^{-2}\mathbf{E}_{n_0}, \tag{5.2}$$

where $\eta^{\,2} = n_0\sigma_\mu^2 + \sigma_\nu^2$ and $\mathbf{E}_{n_0} = \mathbf{I}_{n_0} - \frac{\boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\mathrm{T}}}{n_0}$. It is easy to see that the inverse of $\boldsymbol{\Omega}$ is

$$\boldsymbol{\Omega}^{-1} = \mathbf{I}_k \otimes \boldsymbol{\varpi}. \tag{5.3}$$

Then a consistent estimator of $\boldsymbol{\Omega}^{-1}$ is given by

$$\widehat{\boldsymbol{\Omega}}^{-1} = \mathbf{I}_k \otimes \Big(\widehat{\eta}^{\,-2}\frac{\boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\mathrm{T}}}{n_0} + \widehat{\sigma}^{\,-2}_{\nu}\mathbf{E}_{n_0}\Big),$$

with $\widehat{\eta}^{\,2} = n_0\widehat{\sigma}^{\,2}_{\mu} + \widehat{\sigma}^{\,2}_{\nu}$. From (1.1) and the procedure of constructing the PLSE $\widehat{\boldsymbol{\beta}}_k$ in Section 2, we propose a weighted profile least squares estimator (WPLSE) of $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}}_k^{w} = (\widehat{\mathbf{X}}^{w\mathrm{T}}\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{X}}^{w})^{-1}\widehat{\mathbf{X}}^{w\mathrm{T}}\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{Y}}^{w},$$

where $\widehat{\mathbf{X}}^w = (\mathbf{I}_{kn_0} - \mathbf{S}^w)\mathbf{X}$, $\widehat{\mathbf{Y}}^w = (\mathbf{I}_{kn_0} - \mathbf{S}^w)\mathbf{Y}$ and

$$\mathbf{S}^w = \begin{pmatrix} (\mathbf{Z}_{11}^{\mathrm{T}} \quad \mathbf{0}_q^{\mathrm{T}})(\mathbf{D}_{U_{11}}^{\mathrm{T}} \boldsymbol{\omega}_{U_{11}} \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{D}_{U_{11}})^{-1} \mathbf{D}_{U_{11}}^{\mathrm{T}} \boldsymbol{\omega}_{U_{11}} \widehat{\boldsymbol{\Omega}}^{-1} \\ \vdots \\ (\mathbf{Z}_{1n_0}^{\mathrm{T}} \quad \mathbf{0}_q^{\mathrm{T}})(\mathbf{D}_{U_{1n_0}}^{\mathrm{T}} \boldsymbol{\omega}_{U_{1n_0}} \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{D}_{U_{1n_0}})^{-1} \mathbf{D}_{U_{1n_0}}^{\mathrm{T}} \boldsymbol{\omega}_{U_{1n_0}} \widehat{\boldsymbol{\Omega}}^{-1} \\ \vdots \\ (\mathbf{Z}_{kn_0}^{\mathrm{T}} \quad \mathbf{0}_q^{\mathrm{T}})(\mathbf{D}_{U_{kn_0}}^{\mathrm{T}} \boldsymbol{\omega}_{U_{kn_0}} \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{D}_{U_{kn_0}})^{-1} \mathbf{D}_{U_{kn_0}}^{\mathrm{T}} \boldsymbol{\omega}_{U_{kn_0}} \widehat{\boldsymbol{\Omega}}^{-1} \end{pmatrix}.$$

Correspondingly, the nonparametric component $\boldsymbol{\alpha}(\cdot)$ is estimated by

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_k^w(u) &= (\mathbf{I}_q, \mathbf{0}_{q\times q})[\widehat{\alpha}_1^w(u), \cdots, \widehat{\alpha}_q^w(u), \widehat{b}_1^w(u), \cdots, \widehat{b}_q^w(u)]^{\mathrm{T}} \\ &= (\mathbf{I}_q, \mathbf{0}_{q\times q})(\mathbf{D}_u^{\mathrm{T}} \boldsymbol{\omega}_u \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{D}_u)^{-1} \mathbf{D}_u^{\mathrm{T}} \boldsymbol{\omega}_u \widehat{\boldsymbol{\Omega}}^{-1} (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_k^w).\end{aligned}$$

Remark 5.1 From its definition, we see that $\widehat{\boldsymbol{\beta}}_k^w$ is not a simple weighted estimator from the transformed model (2.3). Besides, by adding $\widehat{\boldsymbol{\Omega}}^{-1}$ into the construction, the form of $\widehat{\mathbf{X}}$ is also changed into $\widehat{\mathbf{X}}^w$.

Theorem 5.1 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.5 *hold. Then we have*

$$\sqrt{k}(\widehat{\boldsymbol{\beta}}_k^w - \boldsymbol{\beta}) \to_D N(0, \boldsymbol{\Sigma}_4^{-1}), \quad \textit{as } k \to \infty,$$

where

$$\boldsymbol{\Sigma}_4 = E(\mathbf{X}_1^{\mathrm{T}} \boldsymbol{\varpi} \mathbf{X}_1) - E(\mathbf{X}_1^{\mathrm{T}} \boldsymbol{\varpi} \mathbf{Z}_1)(E(\mathbf{Z}_1^{\mathrm{T}} \boldsymbol{\varpi} \mathbf{Z}_1))^{-1} E(\mathbf{Z}_1^{\mathrm{T}} \boldsymbol{\varpi} \mathbf{X}_1).$$

Remark 5.2 Let $\mathbf{Z} = (\mathbf{Z}_{11}, \cdots, \mathbf{Z}_{1n_0}, \cdots, \mathbf{Z}_{kn_0})^{\mathrm{T}}$. For any $n_0 \times n_0$ positive definite matrix $\boldsymbol{\Sigma}$, there is

$$\begin{aligned}&\begin{pmatrix} \mathbf{X}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1})\mathbf{X} & \mathbf{X}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1})\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1})\mathbf{X} & \mathbf{Z}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1})\mathbf{Z} \end{pmatrix}^{-1} \\ &\leq \begin{pmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\mathbf{Z} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma})\mathbf{X} & \mathbf{X}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma})\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma})\mathbf{X} & \mathbf{Z}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma})\mathbf{Z} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\mathbf{Z} \end{pmatrix}^{-1} \quad \text{a.s.}\end{aligned}$$

From

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}_{11.2}^{-1} & -\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11.2}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{pmatrix},$$

where $\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$, we have

$$\begin{aligned}&\{\mathbf{X}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1})\mathbf{X} - \mathbf{X}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1})\mathbf{Z}[\mathbf{Z}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1})\mathbf{Z}]^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{I}_k \otimes \boldsymbol{\Sigma})\mathbf{X}\}^{-1} \\ &\leq [\mathbf{X}^{\mathrm{T}}\mathbf{X} - \mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}]^{-1}[\mathbf{X}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}] \\ &\quad \cdot (\mathbf{I}_k \otimes \boldsymbol{\Sigma})[\mathbf{X}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}]^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{X} - \mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}]^{-1}.\end{aligned}$$

This implies $\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1} \geq \boldsymbol{\Sigma}_4^{-1}$. Therefore, $\widehat{\boldsymbol{\beta}}_k^w$ is asymptotically more efficient than $\widehat{\boldsymbol{\beta}}_k$ in the sense of having a smaller asymptotic covariance matrix. In fact, Carroll, Ruppert and Welsh [3] showed that $\boldsymbol{\Sigma}_4^{-1}$ is the semiparametric information bound. Hence, the weighted profile least squares estimator $\widehat{\boldsymbol{\beta}}_k^w$ is semiparametrically efficient.

Theorem 5.2 *Suppose that Assumptions* 2.1 *to* 2.5 *hold. Then we have*

$$\sqrt{kn_0h}\Big\{\widehat{\boldsymbol{\alpha}}_k^w(u_0)-\boldsymbol{\alpha}(u_0)-\frac{h^2}{2}\frac{(\mu_2^0)^2-\mu_1^0\mu_3^0}{\mu_2^0-(\mu_1^0)^2}\boldsymbol{\alpha}''(u_0)\Big\}\to_D N(0,\boldsymbol{\Sigma}_5),\quad as\ k\to\infty,$$

where $\boldsymbol{\Sigma}_5=\big\{(c_0^2\nu_0^0+2c_0c_1\nu_1^0+c_1^2\nu_2^0)\frac{n_0}{p(u_0)}\big\}\{E(\mathbf{Z}_1^{\rm T}\boldsymbol{\varpi}\mathbf{Z}_1)\}^{-1}$, $\mu_1^0,\mu_2^0,\mu_3^0,\nu_0^0,\nu_1^0,\nu_2^0,c_0,c_1$ *are defined in Theorem* 2.2.

Theorem 5.3 *Suppose that Assumptions* 2.1 *to* 2.5 *hold. Then we have*

$$\sup_{u\in\mathcal{U}}\|\widehat{\boldsymbol{\alpha}}_k^w(u)-\boldsymbol{\alpha}(u)\|=O_p\Big\{h^2+\Big(\frac{\log k}{kh}\Big)^{\frac12}\Big\}.$$

Remark 5.3 From the definition of $\boldsymbol{\varpi}$ and

$$\{E(\mathbf{Z}_1^{\rm T}\boldsymbol{\varpi}\mathbf{Z}_1)\}^{-1}\le\{E(\mathbf{Z}_1^{\rm T}\mathbf{Z}_1)\}^{-1}E(\mathbf{Z}_1^{\rm T}\boldsymbol{\varpi}^{-1}\mathbf{Z}_1)\{E(\mathbf{Z}_1^{\rm T}\mathbf{Z}_1)\}^{-1},$$

we have $\boldsymbol{\Sigma}_5\le\boldsymbol{\Sigma}_3$. Therefore, the weighted local linear estimator $\widehat{\boldsymbol{\alpha}}_k^w$ is asymptotically more efficient than the local linear estimator $\widehat{\boldsymbol{\alpha}}_k$. This is an interesting result. According to [10, 12], ignoring the correlation structure entirely and "pretending" that the data are really independent will result in more efficient estimators when estimating nonparametric regression with longitudinal or panel data. This observation was termed "working independence". However, we find this not true when the design points of the nonparametric component have the closeness property within groups.

To estimate $\boldsymbol{\Sigma}_4$ and $\boldsymbol{\Sigma}_5$ consistently, let $\widehat{\boldsymbol{\varpi}}=\widehat{\eta}^{-2}\frac{\boldsymbol{\iota}_{n_0}\boldsymbol{\iota}_{n_0}^{\rm T}}{n_0}+\widehat{\sigma}_\nu^{-2}\mathbf{E}_{n_0}$ and define

$$\widehat{\boldsymbol{\Sigma}}_4=\frac1k\sum_{i=1}^k\widehat{\mathbf{X}}_i^{\rm T}\widehat{\boldsymbol{\varpi}}\widehat{\mathbf{X}}_i-\frac1k\sum_{i=1}^k\widehat{\mathbf{X}}_i^{\rm T}\widehat{\boldsymbol{\varpi}}\widehat{\mathbf{Z}}_i\Big(\frac1k\sum_{i=1}^k\widehat{\mathbf{Z}}_i^{\rm T}\widehat{\boldsymbol{\varpi}}\widehat{\mathbf{Z}}_i\Big)^{-1}\Big(\frac1k\sum_{i=1}^k\widehat{\mathbf{X}}_i^{\rm T}\widehat{\boldsymbol{\varpi}}\widehat{\mathbf{Z}}_i\Big)^{\rm T}$$

and

$$\widehat{\boldsymbol{\Sigma}}_5=\frac{(c_0^2\nu_0^0+2c_0c_1\nu_1^0+c_1^2\nu_2^0)}{p(u_0)}\Big(\frac{1}{kn_0}\sum_{i=1}^k\mathbf{Z}_i^{\rm T}\widehat{\boldsymbol{\varpi}}\mathbf{Z}_i\Big)^{-1}.$$

Theorem 5.4 *Suppose that Assumptions* 2.1 *and* 2.4 *hold. Then we have* $\widehat{\boldsymbol{\Sigma}}_4\to_p\boldsymbol{\Sigma}_4$ *and* $\widehat{\boldsymbol{\Sigma}}_5\to_p\boldsymbol{\Sigma}_5$ *as* $k\to\infty$.

6 A Block Bootstrap Goodness of Fit Test

To test whether model (1.2) holds with a specified parametric form such as panel data linear regression models, we conduct a goodness of fit test based on the comparison of the residual sum of squares (RSS) from both parametric and semiparametric fittings.

Consider the null hypothesis

$$H_0:\alpha_j(u)=a_j(u,\boldsymbol{\theta}),\quad j=1,\cdots,q,\tag{6.1}$$

where $a_j(\,\cdot\,,\boldsymbol{\theta})$ is a given family of functions indexed by an unknown parameter vector $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}_k$ be an estimator of $\boldsymbol{\theta}$. The residual sum of squares under the null hypothesis is

$$\mathrm{RSS}_0=(kn_0)^{-1}\sum_{i=1}^k\sum_{j=1}^{n_0}(Y_{ij}-\mathbf{X}_{ij}^{\rm T}\widehat{\boldsymbol{\beta}}_k-\mathbf{Z}_{ij}^{\rm T}\mathbf{a}(U_{ij},\widehat{\boldsymbol{\theta}}_k))^2,$$

where $\mathbf{a}(U_{ij},\widehat{\boldsymbol{\theta}}_k)=(a_1(U_{ij},\widehat{\boldsymbol{\theta}}_k),\cdots,a_q(U_{ij},\widehat{\boldsymbol{\theta}}_k))^{\mathrm{T}}$, and the residual sum of squares corresponding to model (1.1) is

$$\mathrm{RSS}_1=(kn_0)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(Y_{ij}-\mathbf{X}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k-\mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(U_{ij}))^2.$$

We define our statistic to be

$$D_k=\frac{\mathrm{RSS}_0-\mathrm{RSS}_1}{\mathrm{RSS}_1}=\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1}-1 \tag{6.2}$$

and reject the null hypothesis (6.1) for large values of D_k. The following theorem supports our test.

Theorem 6.1 *Suppose that Assumptions* 2.1(i) *and* 2.2 *to* 2.4 *hold. Then under* H_0, $D_k\to 0$ *in probability as* $k\to\infty$. *Otherwise, if* $\inf_{\boldsymbol{\theta}}[\int_{\mathcal{U}}\|\boldsymbol{\alpha}(u)-\mathbf{a}(u,\boldsymbol{\theta})\|^2\mathrm{d}u]^{\frac12}>0$, *then there exists a* $\delta>0$ *such that, with probability tending to one,* $D_k>\delta$.

Since there exist correlations within groups, the usual nonparametric bootstrap approach (see [2]) can not be used to evaluate p-values of the test. We here propose to use the block bootstrap approach to evaluate p-values of the test which can accommodate the correlations within groups.

(1) By fitting the panel data varying-coefficient partially linear model we estimate the residuals $\{\boldsymbol{\varepsilon}_i\}_{i=1}^k$ by $\{\widehat{\boldsymbol{\varepsilon}}_i\}_{i=1}^k$ where $\boldsymbol{\varepsilon}_i=(\varepsilon_{i1},\cdots,\varepsilon_{in_0})^{\mathrm{T}}$, $\widehat{\boldsymbol{\varepsilon}}_i=(\widehat{\varepsilon}_{i1},\cdots,\widehat{\varepsilon}_{in_0})^{\mathrm{T}}$ and

$$\widehat{\varepsilon}_{ij}=Y_{ij}-\mathbf{X}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k-\mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(U_{ij}),\quad i=1,\cdots,k,\ j=1,\cdots,n_0.$$

(2) Generate bootstrap residuals $\{\boldsymbol{\varepsilon}_i^\star\}_{i=1}^k$ from the empirical distribution of the centered residuals $\{\widehat{\boldsymbol{\varepsilon}}_i-\overline{\widehat{\boldsymbol{\varepsilon}}}\}_{i=1}^k$ where $\overline{\widehat{\boldsymbol{\varepsilon}}}=k^{-1}\sum\limits_{i=1}^{k}\widehat{\boldsymbol{\varepsilon}}_i$. Define

$$Y_{ij}^\star=\mathbf{X}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k+\mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(U_{ij})+\varepsilon_{ij}^\star\quad\text{for } i=1,\cdots,k \text{ and } \ j=1,\cdots,n_0,$$

where $\varepsilon_{ij}^\star$ is the jth entry of $\boldsymbol{\varepsilon}_i^\star$.

(3) Calculate the bootstrap test statistic $D_k^\star$ based on the sample $\{U_{ij},\mathbf{X}_{ij},\mathbf{Z}_{ij},Y_{ij}^\star\}$ and (6.2).

(4) Reject the null hypothesis H_0 at level α when D_k is greater than the upper-α percentage point of the conditional distribution of $D_k^\star$ given $\{U_{ij},\mathbf{X}_{ij},\mathbf{Z}_{ij},Y_{ij}\}$.

The p-value of the test is simply the relative frequency of the event $\{D_k^\star\ge D_k\}$ in the replications of the bootstrap sampling. For the sake of simplicity, we use the same bandwidth in calculating $D_k^\star$ as that in D_k. Note that we bootstrap the centralized residuals from the semiparametric fit instead of the parametric fit, because the semiparametric estimate of the residuals is always consistent, no matter which of the null or the alternative hypothesis is true.

7 Some Simulation Studies

In this section, we carry out some simulation studies to demonstrate the finite sample performances of the estimators and tests proposed in the previous sections.

The data are generated from the following panel data varying coefficient partially linear model:

$$y_{ij}=x_{ij1}\beta_1+x_{ij2}\beta_2+z_{ij}\alpha(u_{ij})+\varepsilon_{ij},\quad \varepsilon_{ij}=\mu_i+\nu_{ij},\quad i=1,\cdots,k,\ j=1,\cdots,n_0,$$

where $x_{ij1} \sim$ i.i.d. $N(2,1)$, $x_{ij2} \sim$ i.i.d. $N(2,1)$, $u_{ij} = \frac{(i-1)n_0+j-0.5}{kn_0}$, $\beta_1 = 1.5$, $\beta_2 = 2$, $\mu_i \sim$ i.i.d. $N(0,\sigma_\mu^2)$ and $\nu_i \sim$ i.i.d. $N(0,\sigma_\nu^2)$. Moreover, we take $\alpha(u) = 2\sin(2\pi u), 2\sin(4\pi u)$, $n_0 = 2, 4$ and $(\sigma_\mu^2, \sigma_\nu^2) = (2.25, 0.25)$, $(2.25, 1)$, respectively.

In each case, the number of simulated realizations is $10,000$. We use Gaussian kernel, and select bandwidth by the leave-one-subject cross-validation criterion proposed in Section 3.

7.1 The finite sample performance of the weighted estimation

Samples of size $kn_0 = 200$, 400 and 600 are drawn repeatedly. We calculate the sample means and standard deviations (SD) of the PLSE and WPLSE for the parametric components β_1 and β_2, and of the estimators of the error variances σ_μ^2 and σ_ν^2. We also calculate the sample means and SDs of the benchmark estimator of β_1 and β_2, which is defined as

$$\widetilde{\boldsymbol{\beta}}_k^w = (\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}\widetilde{\mathbf{X}}^w)^{-1}\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}\widetilde{\mathbf{Y}}^w, \tag{7.1}$$

where $\widetilde{\mathbf{X}}^w$ and $\widetilde{\mathbf{Y}}^w$ have the same definitions as $\widehat{\mathbf{X}}^w$ and $\widehat{\mathbf{Y}}^w$ except that $\widehat{\boldsymbol{\Omega}}$ is replaced by $\boldsymbol{\Omega}$. Moreover, we report the relative efficiencies (RE) of the PLSE and WPLSE defined as the ratio of the mean square error (MSE) of the estimator in question to that of the benchmark estimator. Some representative results are listed in Table 1 and Table 2. We see that

(1) The PLSE, WPLSE and the estimators for the error variances are asymptotically unbiased.

(2) The WPLSE has smaller standard deviation than that of the PLSE. This is consistent with our theoretical results.

(3) The relative efficiency of the WPLSE is very close to 1. As for the PLSE, since it ignores the error structure, its relative efficiency is not close to 1. In some cases, the MSE of the PLSE is more than twice as large as that of the WPLSE.

(4) In general, increasing k and/or n_0 improves the performance of the WPLSE, but this is not true for the PLSE.

In addition, we also demonstrate the finite sample performances of the estimators of the nonparametric components. An estimator $\widetilde{\boldsymbol{\alpha}}(\cdot)$ of $\boldsymbol{\alpha}(\cdot)$ is assessed via the square root of average squared errors (SRASE) defined by

$$\mathrm{SRASE} = \Big[(kn_0)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\{\widetilde{\alpha}(u_{ij}) - \alpha(u_{ij})\}^2\Big]^{\frac{1}{2}}.$$

Some representative results are listed in Table 3, in which R1 is for the SRASE of the unweighted local polynomial estimator, R2 for the SRASE of the weighted local polynomial estimator with estimated weights, and R3 for the SRASE of the weighted local polynomial estimator with the known weights. From Table 3, we see that the WLPE outperforms the LPE.

7.2 The power of the block bootstrap goodness of fit test

To study the power of the proposed bootstrap test, we consider the following null hypothesis:

$$H_0 : \alpha(u_{ij}) = \theta \quad \text{for all } i = 1, \cdots, k,\ j = 1, \cdots, n_0,$$

namely a linear regression model, against the alternative

$$H_1 : \alpha(u_{ij}) \neq \theta \quad \text{for at least one pair of } i \text{ and } j.$$

Table 1 Finite sample performance of the estimators of the parametric components and error variances with $n_0 = 2$

		$\widehat{\beta}_{1k}$	$\widehat{\beta}_{2k}$	$\widehat{\beta}^w_{1k}$	$\widehat{\beta}^w_{2k}$	$\widetilde{\beta}_{1k}$	$\widetilde{\beta}_{2k}$	$\widehat{\sigma}^2_\mu$	$\widehat{\sigma}^2_\nu$
$\alpha(u) = 2\sin(2\pi u)$	$\sigma^2_\mu = 2.25,\ \sigma^2_\nu = 0.25$								
$k = 100$	Mean	1.4993	1.9989	1.5021	1.9973	1.5025	1.9976	1.9921	0.3184
	SD	0.0838	0.0866	0.0503	0.0496	0.0510	0.0500	0.3070	0.0549
	RE	1.6429	1.7318	0.9862	0.9920	–	–	–	–
$k = 200$	Mean	1.4995	2.0000	1.4981	2.0007	1.4980	2.0007	2.1275	0.2892
	SD	0.0626	0.0629	0.0328	0.0319	0.0329	0.0320	0.2293	0.0334
	RE	1.9025	1.9656	0.9969	0.9969	–	–	–	–
$k = 300$	Mean	1.4959	1.9994	1.5037	2.0025	1.5041	2.0031	2.2303	0.2820
	SD	0.0434	0.0480	0.0242	0.0253	0.0244	0.0250	0.1908	0.0274
	RE	1.7782	1.9193	0.9917	1.0119	–	–	–	–
	$\sigma^2_\mu = 2.25, \sigma^2_\nu = 1.0$								
$k = 100$	Mean	1.4883	2.0085	1.4901	2.0041	1.4896	2.0041	1.9970	1.0153
	SD	0.0999	0.1006	0.0779	0.0795	0.0777	0.0796	0.3710	0.1446
	RE	1.2857	1.2645	1.0024	0.9987	–	–	–	–
$k = 200$	Mean	1.5022	1.9905	1.4938	1.9951	1.4937	1.9956	2.1429	1.0305
	SD	0.0645	0.0693	0.0520	0.0502	0.0522	0.0504	0.2585	0.1016
	RE	1.2348	1.3763	0.9961	0.9961	–	–	–	–
$k = 300$	Mean	1.5049	1.9903	1.4984	1.9949	1.4982	1.9950	2.1735	1.0111
	SD	0.0610	0.0632	0.0458	0.0468	0.0460	0.0467	0.2323	0.0904
	RE	1.3265	1.3546	0.9956	1.0022	–	–	–	–
$\alpha(u) = 2\sin(4\pi u)$	$\sigma^2_\mu = 2.25, \sigma^2_\nu = 0.25$								
$k = 100$	Mean	1.5106	2.0030	1.5031	2.0002	1.5020	1.9993	2.0129	0.3519
	SD	0.0960	0.0886	0.0548	0.0520	0.0553	0.0531	0.3235	0.0727
	RE	1.7379	1.6687	0.9911	0.9793	–	–	–	–
$k = 200$	Mean	1.5007	1.9999	1.4964	2.0001	1.4962	1.9996	2.1430	0.3274
	SD	0.0628	0.0621	0.0331	0.0343	0.0336	0.0350	0.2554	0.0476
	RE	1.8683	1.7743	0.9851	0.9800	–	–	–	–
$k = 300$	Mean	1.4997	2.0011	1.5007	1.9998	1.5006	1.9996	2.1090	0.2917
	SD	0.0502	0.0491	0.0290	0.0277	0.0295	0.0275	0.1952	0.0252
	RE	1.7017	1.7855	0.9831	1.0073	–	–	–	–
	$\sigma^2_\mu = 2.25, \sigma^2_\nu = 1.0$								
$k = 100$	Mean	1.4864	2.0004	1.4869	2.0007	1.4870	2.0011	1.8170	1.0147
	SD	0.1020	0.1050	0.0847	0.0843	0.0857	0.0844	0.3625	0.1552
	RE	1.1900	1.2441	0.9884	0.9988	–	–	–	–
$k = 200$	Mean	1.5000	1.9990	1.5002	2.0001	1.5006	1.9996	2.0106	1.0256
	SD	0.0718	0.0733	0.0540	0.0572	0.0537	0.0573	0.2591	0.1039
	RE	1.3370	1.2792	1.0056	0.9983	–	–	–	–
$k = 300$	Mean	1.5141	1.9861	1.5102	1.9920	1.5104	1.9920	2.0933	1.0204
	SD	0.0471	0.0491	0.0395	0.0357	0.0395	0.0358	0.1672	0.0946
	RE	1.1942	1.3744	0.9999	0.9972	–	–	–	–

Note Mean, SD and RE denote the sample mean, standard deviation and relative efficiency, respectively. $\widehat{\beta}_{sk}$, $\widehat{\beta}^w_{sk}$ and $\widetilde{\beta}^w_{sk}$ are the PLSE, WPLSE and benchmark estimators of β_s, respectively, where $s = 1$ and 2.

Table 2 Finite sample performance of the estimators of the parametric components and error variances with $n_0 = 4$

		$\widehat{\beta}_{1k}$	$\widehat{\beta}_{2k}$	$\widehat{\beta}^w_{1k}$	$\widehat{\beta}^w_{2k}$	$\widetilde{\beta}_{1k}$	$\widetilde{\beta}_{2k}$	$\widehat{\sigma}^2_\mu$	$\widehat{\sigma}^2_\nu$
$\alpha(u) = 2\sin(2\pi u)$	$\sigma^2_\mu = 2.25, \sigma^2_\nu = 0.25$								
$k = 50$	Mean	1.5051	1.9964	1.4996	2.0030	1.4995	2.0030	1.9882	0.3183
	SD	0.0981	0.0964	0.0443	0.0485	0.0445	0.0496	0.4364	0.0467
	RE	2.2039	1.9445	0.9947	0.9783	–	–	–	–
$k = 100$	Mean	1.4893	2.0023	1.4964	1.9983	1.4963	1.9982	2.0994	0.2896
	SD	0.0634	0.0603	0.0324	0.0302	0.0322	0.0304	0.3039	0.0294
	RE	1.9744	1.9802	1.0071	0.9925	–	–	–	–
$k = 150$	Mean	1.5016	1.9938	1.4995	1.9998	1.4996	2.0000	2.1225	0.2853
	SD	0.0545	0.0570	0.0256	0.0245	0.0257	0.0245	0.2372	0.0238
	RE	2.1234	2.3252	0.9980	0.9996	–	–	–	–
	$\sigma^2_\mu = 2.25, \sigma^2_\nu = 1.0$								
$k = 50$	Mean	1.5012	1.9951	1.4957	1.9985	1.4948	1.9983	1.9586	1.0268
	SD	0.1052	0.1120	0.0719	0.0790	0.0724	0.0789	0.4904	0.1248
	RE	1.4518	1.4197	0.9925	1.0018	–	–	–	–
$k = 100$	Mean	1.5029	1.9988	1.5015	2.0013	1.5017	2.0014	2.1111	1.0174
	SD	0.0782	0.0681	0.0524	0.0438	0.0525	0.0440	0.3394	0.0857
	RE	1.4892	1.5483	0.9987	0.9947	–	–	–	–
$k = 150$	Mean	1.4968	2.0006	1.5002	2.0088	1.4999	2.0089	2.1140	1.0297
	SD	0.0523	0.0531	0.0370	0.0371	0.0371	0.0372	0.2526	0.0668
	RE	1.4105	1.4234	0.9985	0.9969	–	–	–	–
$\alpha(u) = 2\sin(4\pi u)$	$\sigma^2_\mu = 2.25, \sigma^2_\nu = 0.25$								
$k = 50$	Mean	1.5080	1.9930	1.4998	1.9980	1.4994	1.9980	1.9360	0.3533
	SD	0.1103	0.0944	0.0524	0.0486	0.0523	0.0493	0.4435	0.0649
	RE	2.1113	1.9158	1.0022	0.9859	–	–	–	–
$k = 100$	Mean	1.5022	1.9959	1.5018	1.9959	1.5018	1.9959	2.0399	0.3092
	SD	0.0664	0.0625	0.0337	0.0319	0.0339	0.0319	0.3133	0.0328
	RE	1.9613	1.9572	0.9964	0.9987	–	–	–	–
$k = 150$	Mean	1.4846	2.0082	1.4989	1.9985	1.4992	1.9986	2.0839	0.2904
	SD	0.0511	0.0525	0.0283	0.0295	0.0283	0.0293	0.2477	0.0257
	RE	1.8156	1.7966	1.0002	1.0098	–	–	–	–
	$\sigma^2_\mu = 2.25, \sigma^2_\nu = 1.0$								
$k = 50$	Mean	1.5025	2.0010	1.5018	1.9981	1.5022	1.9979	1.9264	1.0503
	SD	0.1102	0.1018	0.0710	0.0706	0.0716	0.0708	0.4132	0.1176
	RE	1.5402	1.4377	0.9924	0.9964	–	–	–	–
$k = 100$	Mean	1.4979	2.0003	1.4992	2.0001	1.4992	2.0002	2.0699	1.0357
	SD	0.0747	0.0758	0.0520	0.0483	0.0522	0.0481	0.3270	0.0856
	RE	1.4304	1.5777	0.9957	1.0056	–	–	–	–
$k = 150$	Mean	1.5284	1.9920	1.5068	2.0026	1.5070	2.0030	2.1862	1.0229
	SD	0.0539	0.0663	0.0344	0.0440	0.0346	0.0440	0.2847	0.0529
	RE	1.5819	1.5059	0.9951	0.9992	–	–	–	–

Note Mean, SD and RE denote the sample mean, standard deviation and relative efficiency, respectively. $\widehat{\beta}_{sk}$, $\widehat{\beta}^w_{sk}$ and $\widetilde{\beta}^w_{sk}$ are the PLSE, WPLSE and benchmark estimators of β_s, respectively, where $s = 1$ and 2.

Table 3 Finite sample performances of the estimators of the nonparametric components

σ_μ^2	σ_ν^2	k	n_0	$\alpha(u)$	Mean(R1)	SD(R1)	Mean(R2)	SD(R2)	Mean(R3)	SD(R3)
2.25	0.25	200	2	$2\sin(2\pi u)$	0.0536	0.0239	0.0130	0.0061	0.0130	0.0061
2.25	1.00	200	2	$2\sin(2\pi u)$	0.0746	0.0378	0.0411	0.0251	0.0414	0.0251
2.25	0.25	200	2	$2\sin(4\pi u)$	0.0930	0.0338	0.0248	0.0088	0.0248	0.0086
2.25	1.00	200	2	$2\sin(4\pi u)$	0.1097	0.0367	0.0685	0.0250	0.0683	0.0249
2.25	0.25	100	4	$2\sin(2\pi u)$	0.0720	0.0381	0.0105	0.0045	0.0105	0.0045
2.25	1.00	100	4	$2\sin(2\pi u)$	0.0886	0.0528	0.0332	0.0150	0.0334	0.0158
2.25	0.25	100	4	$2\sin(4\pi u)$	0.0804	0.0283	0.0231	0.0080	0.0231	0.0079
2.25	1.00	100	4	$2\sin(4\pi u)$	0.1063	0.0555	0.0513	0.0207	0.0515	0.0207

Note R1, R2 and R3 denote the SRASE of the LLE, WLLE with estimated weights and WLLE with known weights, respectively.

Table 4 The size and power of bootstrap goodness of test of models

			c	0	0	0	0.35	0.50	0.65
level of significance				5%	10%	20%	5%	5%	5%
σ_μ^2	σ_ν^2	k	n_0		size			power	
2.25	1.00	200	2	5.8%	8.7%	23.3%	59.7%	87.1%	96.7%
2.25	1.00	100	4	5.2%	10.2%	21.8%	66.6%	92.5%	100%

Specifically, we evaluate power under a sequence of alternative models indexed by c,

$$H_1 : \alpha(u_{ij}) = \overline{\alpha}^0 + c(\alpha^0(u_{ij}) - \overline{\alpha}^0) \quad \text{for all } i = 1, \cdots, k,\ j = 1, \cdots, n_0,$$

where $\alpha^0(u_{ij}) = 2\sin(2\pi u_{ij})$ and $\overline{\alpha}^0 = (kn_0)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_0} \alpha^0(u_{ij})$. For each realization, we repeat bootstrap sampling 500 samples. The results are listed in Table 4. From Table 4, we can see that the proposed bootstrap test behaves well under the null hypothesis and is powerful under the alternative hypotheses.

8 Concluding Remarks

Semiparametric regression modeling has become a mainstay recently due to its flexibility. This paper studies the test and estimation problems of a semiparametric panel data varying coefficient partially linear model with one-way error structure. For the parametric component we have proposed a weighted profile least squares estimator (WPLSE) and shown that it is asymptotically more efficient than the usual profile least squares estimator (PLSE). For the nonparametric component, we have proposed a weighted local polynomial estimator (WLPE) and shown that the WLPE is also asymptotically more efficient than the usual local polynomial estimator (LPE). This is an interesting result which is a complement of the finding by Ruckstuhl, Welsh and Carroll [12] and Lin and Carroll [10]. The asymptotic properties of these weighted estimators have been established. In addition, we have proposed a block bootstrap test for the goodness of fit of models. These results can be used to make asymptotically valid statistical inferences.

We focused on the case of equal error variance. In practice, however, the error variances may differ substantially. For example, heteroscedasticity in the disturbance term has been observed in gasoline demand across Organization for Economic Co-Operation and Development (OECD) countries. Interesting topics for future study include extending our results to look after heteroscedasticity.

Appendix Proofs of the Main Results

In order to prove the main results, we first introduce several lemmas.

Lemma A.1 *Suppose that Assumptions* 2.1(i), 2.2 *to* 2.4 *hold. Then, as* $k \to \infty$ *we have*

$$\sup_{u\in\mathcal{U}}\Big\|\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\Big[K_h(U_{ij_1}-u)\Big(\frac{U_{ij_1}-u}{h}\Big)^s\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^T-p(u)E(\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^T)\mu_s^0\Big]\Big\|=O(c_k)\quad a.s.$$

and

$$\sup_{u\in\mathcal{U}}\Big\|\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}K_h(U_{ij_1}-u)\Big(\frac{U_{ij_1}-u}{h}\Big)^s\mathbf{Z}_{ij_1}\varepsilon_{ij_2}\Big\|=O\Big[\Big(\frac{\log k}{kh}\Big)^{\frac{1}{2}}\Big]\quad \text{a.s.},$$

where $s=0,1,2,4$, $c_k=h^2+\big(\frac{\log k}{kh}\big)^{\frac{1}{2}}$.

Lemma A.2 *Suppose that Assumptions* 2.1(i), 2.2 *to* 2.4 *hold. Then, as* $k\to\infty$, *we have*

$$\frac{1}{n_0k}\sum_{i_1=1}^{k}\sum_{j_1=1}^{n_0}\sum_{i_2\neq i_1}^{k}\sum_{j_2=1}^{n_0}K_h(U_{i_2j_2}-U_{i_1j_1})\mathbf{d}_{2q,s_1}^{\mathrm{T}}(\mathbf{D}_{U_{i_1j_1}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{i_1j_1}}\mathbf{D}_{U_{i_1j_1}})^{-1}$$

$$\cdot\,\mathbf{d}_{2q,s_2}\zeta'_{i_1j_1s_1}\zeta'_{i_2j_2s_2}=o(k^{-\frac{1}{2}})\quad \text{a.s.},\tag{A.1}$$

where $1\le s_1,s_2\le q$, $\mathbf{d}_{2q,s}$ *is a* $2q$*-vector with the* s*th entry being* 1 *and other entries being* 0, $\zeta_{ijs}=Z_{ijs}\varepsilon_{ij}$ *and* $\zeta'_{ijs}=\zeta_{ijs}I_{\{|\zeta_{ijs}|\le i^{\frac{1}{4}}\}}-E(\zeta_{ijs}I_{\{|\zeta_{ijs}|\le i^{\frac{1}{4}}\}}|Z_{ijs})$.

Lemma A.3 *Let* $e_1,\cdots,e_k$ *be independent random variables with mean zero and finite variance, and* $\sup_{1\le i\le k}E|e_i|^r\le c<\infty$ $(r>2)$. *Assume* $\{a_{ij}^{(k)},i,j=1,\cdots,k\}$ *to be a sequence of positive numbers such that* $\sup_{1\le i,j\le k}|a_{ij}^{(k)}|\le k^{-p_1}$ *for some* $0\le p_1\le 1$ *and* $\sum_{i=1}^{k}|a_{ij}^{(k)}|=O(k^{p_2})$ *for* $p_2\ge\max(0,\frac{2}{r}-p_1)$. *Then*

$$\max_{1\le j\le k}\Big|\sum_{i=1}^{k}a_{ij}^{(k)}e_i\Big|=O(k^{-s}\log k)\quad for\ s=\frac{1}{2}(p_1-p_2)\quad \text{a.s.}\tag{A.2}$$

Lemma A.4 *Let* $\{\xi_i\}_{i=1}^k$ *be independent random variables with zero means and finite absolute moments of order* $s\ge 2$. *Then*

$$E\Big|\sum_{i=1}^{k}\xi_i\Big|^s\le c_sk^{\frac{s-2}{2}}\sum_{i=1}^{k}E|\xi_i|^s,$$

where c_s *is a constant.*

The proofs of Lemmas A.1–A.3 are similar to those of Lemma A2 in [17], Lemma 4.3 in [21] and Lemma 1 in [13], respectively. We here omit the details. Lemma A.4 can be found in [14].

Lemma A.5 *Suppose that Assumptions* 2.1 *to* 2.4 *hold. Then we have*

$$\frac{1}{k}\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}} \to_p \mathbf{\Sigma}_1 \quad \textit{and} \quad \frac{1}{k}\widetilde{\mathbf{X}}^{w\mathrm{T}}\mathbf{\Omega}^{-1}\widetilde{\mathbf{X}}^{w} \to_p \mathbf{\Sigma}_4, \quad \textit{as } k\to\infty, \tag{A.3}$$

where $\mathbf{\Sigma}_1$ *and* $\mathbf{\Sigma}_4$ *are defined in Theorems* 2.1 *and* 5.1, *respectively,* $\widetilde{\mathbf{X}}^w = (\mathbf{I}_{kn_0} - \widetilde{\mathbf{S}}^w)\mathbf{X}$, $\widetilde{\mathbf{S}}^w$ *has the same definition as* $\mathbf{S}^w$ *in Section* 5 *except that* $\widehat{\mathbf{\Omega}}$ *is replaced by* $\mathbf{\Omega}$.

Proof By applying Lemma A.1 and the form of Ω^{-1}, it is easy to show that Lemma A.5 holds. We here omit the details.

Denote $\widetilde{\boldsymbol{\alpha}}_k^w(u) = (\mathbf{I}_q, \mathbf{0}_{q\times q})(\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u\mathbf{\Omega}^{-1}\mathbf{D}_u)^{-1}\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u\mathbf{\Omega}^{-1}(\mathbf{Y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}_k^w)$ with $\widetilde{\boldsymbol{\beta}}_k^w$ as (7.1). For $\widetilde{\boldsymbol{\beta}}_k^w$ and $\widetilde{\boldsymbol{\alpha}}_k^w(u) = (\widetilde{\alpha}_1^w(u), \cdots, \widetilde{\alpha}_q^w(u))^{\mathrm{T}}$, we have the following results.

Lemma A.6 *Suppose that Assumptions* 2.1 *to* 2.5 *hold. Then we have*

$$\sqrt{k}(\widetilde{\boldsymbol{\beta}}_k^w - \boldsymbol{\beta}) \to_D N(0, \mathbf{\Sigma}_4^{-1}), \quad \textit{as } k\to\infty \tag{A.4}$$

and

$$\sqrt{kn_0h}\Big\{\widetilde{\boldsymbol{\alpha}}_k^w(u_0) - \boldsymbol{\alpha}(u_0) - \frac{h^2}{2}\frac{(\mu_2^0)^2-\mu_1^0\mu_3^0}{\mu_2^0-(\mu_1^0)^2}\boldsymbol{\alpha}''(u_0)\Big\} \to_D N(0,\mathbf{\Sigma}_5), \quad \textit{as } k\to\infty, \tag{A.5}$$

where $\mathbf{\Sigma}_4$ *and* $\mathbf{\Sigma}_5$ *are defined in Section* 5.

Proof Here, we write $\mathbf{\Phi} = (E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{Z}_1))^{-1}$ and $\boldsymbol{\varpi} = (\sigma^{j_1j_2})_{j_1,j_2=1}^{n_0}$ for convenience. According to the definition of $\widetilde{\boldsymbol{\beta}}_k^w$, it can be verified that

$$\widetilde{\boldsymbol{\beta}}_k^w - \boldsymbol{\beta} = (\widetilde{\mathbf{X}}^{w\mathrm{T}}\mathbf{\Omega}^{-1}\widetilde{\mathbf{X}}^w)^{-1}\{\widetilde{\mathbf{X}}^{w\mathrm{T}}\mathbf{\Omega}^{-1}(\mathbf{I}_{kn_0}-\widetilde{\mathbf{S}}^w)\mathbf{M} + \widetilde{\mathbf{X}}^{w\mathrm{T}}\mathbf{\Omega}^{-1}(\mathbf{I}_{kn_0}-\widetilde{\mathbf{S}}^w)\varepsilon\},$$

where $\mathbf{M} = (\mathbf{Z}_{11}^{\mathrm{T}}\boldsymbol{\alpha}(U_{11}), \cdots, \mathbf{Z}_{1n_0}^{\mathrm{T}}\boldsymbol{\alpha}(U_{1n_0}), \cdots, \mathbf{Z}_{kn_0}^{\mathrm{T}}\boldsymbol{\alpha}(U_{kn_0}))^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}$ is defined in Section 5. Since $\widetilde{\mathbf{X}}^{w\mathrm{T}}\mathbf{\Omega}^{-1}(\mathbf{I}_{kn_0}-\widetilde{\mathbf{S}}^w)\mathbf{M} = [\mathbf{X}+(\widetilde{\mathbf{X}}^w-\mathbf{X})]^{\mathrm{T}}\mathbf{\Omega}^{-1}(\mathbf{I}_{kn_0}-\widetilde{\mathbf{S}}^w)\mathbf{M}$, the sth entry of this matrix can be written as

$$\begin{aligned}
&\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{X_{ij_1s}\mathbf{Z}_{ij_2}^{\mathrm{T}} - E(\mathbf{d}_{p,s}^{\mathrm{T}}\mathbf{X}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{Z}_1)\mathbf{\Phi}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\}\\
&\cdot\{\boldsymbol{\alpha}(U_{ij_2}) - (\mathbf{I}_q,\mathbf{0}_{q\times q})(\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\mathbf{\Omega}^{-1}\mathbf{D}_{U_{ij_2}})^{-1}\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\mathbf{\Omega}^{-1}\mathbf{M}\}\\
&+\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{\widetilde{X}_{ij_1s}^w - X_{ij_1s} + \mathbf{Z}_{ij_1}^{\mathrm{T}}\mathbf{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_{p,s})\}\\
&\cdot\{\mathbf{Z}_{ij_2}^{\mathrm{T}}\boldsymbol{\alpha}(U_{ij_2}) - (\mathbf{Z}_{ij_2}^{\mathrm{T}},\mathbf{0}_q^{\mathrm{T}})(\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\mathbf{\Omega}^{-1}\mathbf{D}_{U_{ij_2}})^{-1}\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\mathbf{\Omega}^{-1}\mathbf{M}\} = \mathrm{I}_1 + \mathrm{I}_2,
\end{aligned}$$

where $\widetilde{X}_{ij_1s}^w$ is the sth entry of $\widetilde{\mathbf{X}}_{ij_1}^w$.

Since $\sum\limits_{j_1=1}^{n_0}\sum\limits_{j_2=1}^{n_0}\sigma^{j_1j_2}\{X_{ij_1s}\mathbf{Z}_{ij_2}^{\mathrm{T}} - E(\mathbf{d}_{p,s}^{\mathrm{T}}\mathbf{X}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{Z}_1)\mathbf{\Phi}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\}$'s are i.i.d. random vectors with mean zero and finite covariance matrix and

$$\max_{1\le i\le k}\max_{1\le j_2\le n_0}\|\boldsymbol{\alpha}(U_{ij_2}) - (\mathbf{I}_q,\mathbf{0}_{q\times q})(\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\mathbf{\Omega}^{-1}\mathbf{D}_{U_{ij_2}})^{-1}\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\mathbf{\Omega}^{-1}\mathbf{M}\| = O_p(c_k),$$

it holds that $\mathrm{I}_1 = o_p(k^{\frac{1}{2}})$. Further,

$$\widehat{X}_{ij_1s} - X_{ij_1s} + \mathbf{Z}_{ij_1}^{\mathrm{T}}(E(\mathbf{Z}_{1j_1}\mathbf{Z}_{2j_2}^{\mathrm{T}}))^{-1}E(\mathbf{Z}_{ij_2}X_{ij_1s})$$

$$= O_p\Big\{\Big(\frac{\log k}{kh}\Big)^{\frac{1}{2}}\Big\}\{X_{ij_1s} - \mathbf{Z}_{ij_1}^{\mathrm{T}}\boldsymbol{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_{p,s})\}$$

and

$$\mathbf{Z}_{ij_2}^{\mathrm{T}}\boldsymbol{\alpha}(U_{ij_2}) - (\mathbf{Z}_{ij_2}^{\mathrm{T}}, \mathbf{0})(\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\mathbf{D}_{U_{ij_2}})^{-1}\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\mathbf{M} = O_p(c_k)\cdot \mathbf{Z}_{ij_2}^{\mathrm{T}}\boldsymbol{\alpha}(U_{ij_2}).$$

This implies $\mathrm{I}_2 = o_p(k^{\frac{1}{2}})$ as well. So we have

$$\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{I}_{kn_0} - \widetilde{\mathbf{S}}^w)\mathbf{M} = o_p(k^{\frac{1}{2}}). \tag{A.6}$$

Moreover, the sth entry of $\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{I}_{kn_0} - \widetilde{\mathbf{S}}^w)\boldsymbol{\varepsilon}$ can be written as

$$\begin{aligned}
&\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{X_{ij_1s} - \mathbf{Z}_{ij_1}^{\mathrm{T}}\boldsymbol{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_{p,s})\}\varepsilon_{ij_2}\\
&-\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{X_{ij_1s}\mathbf{Z}_{ij_2}^{\mathrm{T}} - \mathbf{Z}_{ij_1}^{\mathrm{T}}\boldsymbol{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_{p,s})\mathbf{Z}_{ij_2}^{\mathrm{T}}\}\\
&\cdot(\mathbf{I}_q, \mathbf{0}_{q\times q})(\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\mathbf{D}_{U_{ij_2}})^{-1}\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\\
&+\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{\widetilde{X}_{ij_1s}^{w} - X_{ij_1s} + \mathbf{Z}_{ij_1}^{\mathrm{T}}\boldsymbol{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_{p,s})\}\varepsilon_{ij_2}\\
&-\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{\widetilde{X}_{ij_1s}^{w} - X_{ij_1s} + \mathbf{Z}_{ij_1}^{\mathrm{T}}\boldsymbol{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_{p,s})\}\\
&\cdot(\mathbf{Z}_{ij_2}^{\mathrm{T}}, \mathbf{0})(\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\mathbf{D}_{U_{ij_2}})^{-1}\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} = \mathrm{K}_1 + \mathrm{K}_2 + \mathrm{K}_3 + \mathrm{K}_4.
\end{aligned}$$

We first show that $\mathrm{K}_2 = o_p(k^{\frac{1}{2}})$. Denote

$$\zeta_{ij_1j_2s} = X_{ij_1s}\mathbf{Z}_{ij_2} - \mathbf{Z}_{ij_1}^{\mathrm{T}}\boldsymbol{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_s)\mathbf{Z}_{ij_2}$$

and

$$\xi_{ij_2} = (\mathbf{I}_q, \mathbf{0}_{q\times q})(\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\mathbf{D}_{U_{ij_2}})^{-1}\mathbf{D}_{U_{ij_2}}^{\mathrm{T}}\boldsymbol{\omega}_{U_{ij_2}}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}.$$

Then we have

$$-\mathrm{K}_2 = \sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sum_{s_1=1}^{q}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\xi_{ij_2s_1} = \sum_{i=1}^{k}\sum_{j_2=1}^{n_0}\sum_{s_1=1}^{q}\xi_{ij_2s_1}\sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1},$$

where $\zeta_{ij_1j_2ss_1}$ and $\xi_{ij_2s_1}$ are the s_1th entry of $\zeta_{ij_1j_2s}$ and ξ_{ij_2}, respectively. It is easy to see

$$\max_{1\le s_1\le q}\max_{1\le i\le k}\max_{1\le j_2\le n_0}|\xi_{ij_2s_1}| = O_p\Big\{\frac{\log k}{(kh)^{\frac{1}{2}}}\Big\}.$$

For $i = 1, \cdots, k$, $j_2 = 1, \cdots, n_0$ and $s_1 = 1, \cdots, q$, put $\vartheta_{ij_2s_1} = \xi_{ij_2s_1}\sum\limits_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}$,

$$\Big(\sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)' = \sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}I_{\big\{\big|\sum\limits_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\big|\le\delta^2 i^{\frac{1}{2}}\big\}}$$

and

$$\Big(\sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)'' = \sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}I_{\big\{\big|\sum\limits_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\big|>\delta^2 i^{\frac{1}{2}}\big\}}$$

for any $\delta>0$, where I_B denotes the indicator function B. Then

$$\vartheta_{ij_2s_1} = \xi_{ij_2s_1}\Big(\sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)' + \xi_{ij_2s_1}\Big(\sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)''.$$

Since $\max\limits_{1\le i\le k}\max\limits_{1\le j_2\le n_0}\max\limits_{1\le s_1\le q}E\Big|\sum\limits_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big|^2<\infty$, by the three-series theorem we obtain $\sum\limits_{i=1}^{k}\sum\limits_{j_2=1}^{n_0}\Big|\Big(\sum\limits_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)''\Big|<\infty$, and consequently,

$$k^{-\frac{1}{2}}\Big|\sum_{i=1}^{k}\sum_{j_2=1}^{n_0}\Big(\sum_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)''\Big| = o(1) \quad \text{a.s.}$$

For $i=1,\cdots,k$, $j_2=1,\cdots,n_0$ and $s_1=1,\cdots,q$, let $\vartheta'_{ij_2s_1}=\xi_{ij_2s_1}\Big(\sum\limits_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)'$. Note that $\{\vartheta'_{11s_1},\cdots,\vartheta'_{1n_0s_1}\},\cdots,\{\vartheta'_{k1s_1},\cdots,\vartheta'_{kn_0s_1}\}$ are independent of the given $\boldsymbol{\varepsilon}$. It is easy to show that $E(\vartheta_{ij_2s_1}|\boldsymbol{\varepsilon})=0$,

$$\max_{1\le i\le k}\max_{1\le j_2\le n_0}\max_{1\le s_1\le q}|\vartheta_{ij_2s_1}| \le \max_{1\le i\le k}\max_{1\le j_2\le n_0}\max_{1\le s_1\le q}|\xi_{ij_2s_1}|\delta^2 k^{\frac{1}{2}}$$

and $E(\vartheta^2_{ij_2s_1}|\boldsymbol{\varepsilon})=\xi^2_{ij_2s_1}E\Big\{\Big(\sum\limits_{j_1=1}^{n_0}\sigma^{j_1j_2}\zeta_{ij_1j_2ss_1}\Big)^2\Big\}$. By the Bernstein inequality, we have

$$\begin{aligned}
P\Big\{\bigcup_{k\ge m}\frac{1}{\sqrt{k}}\Big|\sum_{i=1}^{k}\sum_{j_2=1}^{n_0}\sum_{s_1=1}^{q}\vartheta'_{ij_2s_1}\Big|\ge\delta\Big\} &\le \sum_{k\ge m}E\Big[\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\sum_{j_2=1}^{n_0}\sum_{s_1=1}^{q}P\{|\vartheta'_{ij_2s_1}|\ge\delta|\boldsymbol{\varepsilon}\}\Big]\\
&\le 2\sum_{k\ge m}\sum_{i=1}^{k}\sum_{j_2=1}^{n_0}\sum_{s_1=1}^{q}E\Big[\exp\Big\{-\frac{1}{2\delta^2k^{\frac{1}{2}}\zeta_k}\Big\}\Big]\\
&\le 2\sum_{k\ge m}k^{-2}\to 0, \quad \text{as } m\to\infty.
\end{aligned}$$

By combining Borel-Cantelli lemma, it holds that $\Big|k^{-\frac{1}{2}}\sum\limits_{i=1}^{k}\sum\limits_{j_2=1}^{n_0}\sum\limits_{s_1=1}^{q}\vartheta'_{ij_2s_1}\Big|\to 0$ a.s. So $\mathrm{K}_2=o_p(k^{\frac{1}{2}})$. By the same argument we can show $\mathrm{K}_3=o_p(k^{\frac{1}{2}})$. Moreover, by applying Lemma A.1, it is easy to see that $\mathrm{K}_4=o_p(k^{\frac{1}{2}})$. Therefore, we have

$$\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{I}_{kn_0}-\widetilde{\mathbf{S}}^w)\boldsymbol{\varepsilon} = \sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{\mathbf{X}_{ij_1}-E(\mathbf{X}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{Z}_1)\boldsymbol{\Phi}\mathbf{Z}_{ij_1}\}\varepsilon_{ij_2}+o_p(k^{\frac{1}{2}}). \quad \text{(A.7)}$$

For any nonzero vector $\boldsymbol{\Lambda}=(\lambda_1,\cdots,\lambda_p)^{\mathrm{T}}$, we have

$$\frac{1}{\sqrt{k}}\boldsymbol{\Lambda}^{\mathrm{T}}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{\mathbf{X}_{ij_1}-E(\mathbf{X}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{Z}_1)\boldsymbol{\Phi}\mathbf{Z}_{ij_1}\}\varepsilon_{ij_2} = \frac{1}{\sqrt{k}}\sum_{s=1}^{p}\lambda_s\sum_{i=1}^{k}\psi_{is},$$

where $\psi_{is} = \sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\{X_{ij_1s} - \mathbf{Z}_{ij_1}^{\mathrm{T}}\mathbf{\Phi}E(\mathbf{Z}_1^{\mathrm{T}}\boldsymbol{\varpi}\mathbf{X}_1\mathbf{d}_{p,s})\}\varepsilon_{ij_2}$. Note that for $i = 1, \cdots, k$, ψ_{is}'s are independent random variables with mean zero. It is easy to check that $\{\psi_{is}\}_{i=1}^{k}$ satisfy the Linderberg condition. Moreover,

$$\mathrm{Var}\Big\{\frac{1}{\sqrt{k}}\mathbf{\Lambda}^{\mathrm{T}}\widetilde{\mathbf{X}}^{w\mathrm{T}}\mathbf{\Omega}^{-1}(\mathbf{I}_{kn_0} - \widetilde{\mathbf{S}}^{w})\boldsymbol{\varepsilon}\Big\} \to \mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Sigma}_4\mathbf{\Lambda}, \quad \text{as } k \to \infty. \tag{A.8}$$

Therefore, combining Lemma A.5 and (A.6)–(A.8), we see that (A.4) holds.

Now, we begin to prove the result for $\widetilde{\boldsymbol{\alpha}}_k^w(u)$. Let $\overline{\boldsymbol{\alpha}}_k^w(u) = (\overline{\alpha}_1^w(u), \cdots, \overline{\alpha}_q^w(u))^{\mathrm{T}}$, $\overline{\mathbf{b}}_k^w(u) = (\overline{b}_1^w(u), \cdots, \overline{b}_q^w(u))^{\mathrm{T}}$ and

$$((\overline{\boldsymbol{\alpha}}_k^w(u))^{\mathrm{T}}, (\overline{\mathbf{b}}_k^w(u))^{\mathrm{T}})^{\mathrm{T}} = (\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u\mathbf{\Omega}^{-1}\mathbf{D}_u)^{-1}\mathbf{D}_u^{\mathrm{T}}\boldsymbol{\omega}_u\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

By the root-k consistency of $\widetilde{\boldsymbol{\beta}}_k^w$, it is easy to show $\overline{\boldsymbol{\alpha}}_k^w(u) - \widetilde{\boldsymbol{\alpha}}_k^w(u) = o(h^2)$. Therefore, in order to complete the proof, we just need to show

$$\sqrt{kn_0h}\Big\{\overline{\boldsymbol{\alpha}}_k^w(u_0) - \boldsymbol{\alpha}(u_0) - \frac{h^2}{2}\frac{(\mu_2^0)^2 - \mu_1^0\mu_3^0}{\mu_2^0 - (\mu_1^0)^2}\boldsymbol{\alpha}''(u_0)\Big\} \to_D N(0, \mathbf{\Sigma}_5), \quad \text{as } k \to \infty.$$

We write $\frac{1}{kn_0}\mathbf{D}_{u_0}^{\mathrm{T}}\boldsymbol{\omega}_{u_0}\mathbf{\Omega}^{-1}\mathbf{D}_{u_0} = \left(\begin{smallmatrix}\mathbf{\Psi}_1 & \mathbf{\Psi}_2\\ \mathbf{\Psi}_3 & \mathbf{\Psi}_4\end{smallmatrix}\right)$. According to Assumption 2.5 on kernel function $K(\cdot)$, Assumption 2.1 on the mechanism of generating U_{ij}, and Lemma A.1, we have the following results. First,

$$\mathbf{\Psi}_1 = \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1} - u_0) = \frac{p(u_0)}{n_0}\mathbf{\Phi}^{-1} + O_p\Big\{\Big(\frac{\log k}{kh}\Big)^{\frac{1}{2}}\Big\},$$

where $\mathbf{\Phi}^{-1} = \sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}E(\mathbf{Z}_{1j_1}\mathbf{Z}_{1j_2}^{\mathrm{T}})\sigma^{j_1j_2}$ according to the definition of $\mathbf{\Phi}$. Then

$$\begin{aligned}\mathbf{\Psi}_2 &= \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1} - u_0)(U_{ij_2} - u_0)\\ &= \frac{h}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1} - u_0)\Big(\frac{U_{ij_1} - u_0}{h}\Big)\\ &\quad + \frac{h}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1} - u_0)\Big(\frac{U_{ij_2} - U_{ij_1}}{h}\Big) = \mathrm{J}_1 + \mathrm{J}_2.\end{aligned}$$

Obviously, by Lemma A.1,

$$\mathrm{J}_1 = \frac{hp(u_0)}{n_0}\mu_1^0\mathbf{\Phi}^{-1} + O_p\Big\{\Big(\frac{\log k}{kh}\Big)^{\frac{1}{2}}\Big\} \quad \text{and} \quad \mathrm{J}_2 = O\Big(\frac{1}{kh}\Big) + O_p\Big(\frac{\log k}{kh^2}\Big) = o_p(1).$$

This implies

$$\mathbf{\Psi}_2 = \frac{hp(u_0)}{n_0}\mu_1^0\mathbf{\Phi}^{-1} + o_p(1).$$

By the same argument, we can show

$$\mathbf{\Psi}_3 = \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1} - u_0)(U_{ij_1} - u_0) = \frac{hp(u_0)}{n_0}\mu_1^0\mathbf{\Phi}^{-1} + o_p(1).$$

Further,

$$\begin{aligned}\boldsymbol{\Psi}_4 &= \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1}-u_0)(U_{ij_1}-u_0)(U_{ij_2}-u_0)\\ &= \frac{h^2}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1}-u_0)\Big(\frac{U_{ij_1}-u_0}{h}\Big)^2\\ &\quad+\frac{h^2}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\sigma^{j_1j_2}K_h(U_{ij_1}-u_0)\Big(\frac{U_{ij_1}-u_0}{h}\Big)\Big(\frac{U_{ij_2}-U_{ij_1}}{h}\Big)=\mathrm{J}_3+\mathrm{J}_4.\end{aligned}$$

By the same reason as before, we can show

$$\mathrm{J}_3\to_p\frac{h^2p(u_0)}{n_0}\mu_2^0\boldsymbol{\Phi}^{-1},\quad\text{as } k\to\infty\quad\text{and}\quad\mathrm{J}_4=O\Big(\frac{1}{kh}\Big)+O_p\Big(\frac{\log k}{kh^2}\Big).$$

So it holds that

$$\frac{1}{kn_0}\mathbf{D}_{u_0}^{\mathrm{T}}\omega_{u_0}\boldsymbol{\Omega}^{-1}\mathbf{D}_{u_0}\to_p\frac{p(u_0)}{n_0}\begin{pmatrix}1 & h\mu_1^0\\ h\mu_1^0 & h^2\mu_2^0\end{pmatrix}\otimes\boldsymbol{\Phi}^{-1},\quad\text{as } k\to\infty. \tag{A.9}$$

Since the coefficient functions $\alpha_j(u)$'s are conducted in the neighborhood of $|U_{ij}-u_0|<h$, by Taylor's expansion,

$$\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}(U_{ij})=\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}(u_0)+(U_{ij}-u_0)\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}'(u_0)+\frac{h^2}{2}\Big(\frac{U_{ij}-u_0}{h}\Big)^2\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}''(u_0)+o_p(h^2).$$

This implies

$$\frac{1}{kn_0}\mathbf{D}_{u_0}^{\mathrm{T}}\boldsymbol{\omega}_{u_0}\boldsymbol{\Omega}^{-1}\mathbf{M}=\frac{1}{kn_0}\mathbf{D}_{u_0}^{\mathrm{T}}\boldsymbol{\omega}_{u_0}\boldsymbol{\Omega}^{-1}\mathbf{D}_{u_0}\begin{pmatrix}\boldsymbol{\alpha}(u_0)\\ \boldsymbol{\alpha}'(u_0)\end{pmatrix}+\frac{1}{2}\begin{pmatrix}\mathbf{A}_1\\ \mathbf{A}_2\end{pmatrix}\boldsymbol{\alpha}''(u_0)+o_p(h^2), \tag{A.10}$$

where

$$\begin{aligned}\mathbf{A}_1&=\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}K_h(U_{ij_1}-u_0)\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}(U_{ij_2}-u_0)^2\\ &=\frac{h^2}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}K_h(U_{ij_1}-u_0)\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}\Big[\Big(\frac{U_{ij_1}-u_0}{h}\Big)^2\\ &\quad+\Big\{\Big(\frac{U_{ij_2}-u_0}{h}\Big)^2-\Big(\frac{U_{ij_1}-u_0}{h}\Big)^2\Big\}\Big]=\mathrm{J}_3+\mathrm{J}_5\end{aligned}$$

and

$$\mathbf{A}_2=\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}K_h(U_{ij_1}-u_0)(U_{ij_1}-u_0)\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\mathbf{Z}_{ij_1}\mathbf{Z}_{ij_2}^{\mathrm{T}}(U_{ij_2}-u_0)^2.$$

It is easy to see

$$\mathrm{J}_5=O_p(k^{-1}h^{-2})=o_p(1).$$

This implies that as $k\to\infty$,

$$\mathbf{A}_1\to_p\frac{h^2p(u_0)}{n_0}\mu_2^0\boldsymbol{\Phi}^{-1}\quad\text{and}\quad\mathbf{A}_2\to_p\frac{h^3p(u_0)}{n_0}\mu_3^0\boldsymbol{\Phi}^{-1}. \tag{A.11}$$

Therefore, by (A.9)–(A.11) we have

$$\begin{aligned}
\begin{pmatrix}\overline{\boldsymbol{\alpha}}^w(u_0)\\ \overline{\mathbf{b}}^w(u_0)\end{pmatrix} =&\Big(\frac{1}{kn_0}\mathbf{D}_{u_0}^{\mathrm{T}}\omega_{u_0}\boldsymbol{\Omega}^{-1}\mathbf{D}_{u_0}\Big)^{-1}\Big(\frac{1}{kn_0}\mathbf{D}_{u_0}^{\mathrm{T}}\boldsymbol{\omega}_{u_0}\boldsymbol{\Omega}^{-1}\Big)(\mathbf{M}+\boldsymbol{\varepsilon})\\
=&\begin{pmatrix}\boldsymbol{\alpha}(u_0)\\ \boldsymbol{\alpha}''(u_0)\end{pmatrix}+\frac{1}{2}\left[\begin{pmatrix}1 & h\mu_1^0\\ h\mu_1^0 & h^2\mu_2^0\end{pmatrix}^{-1}\begin{pmatrix}h^2\mu_2^0\\ h^3\mu_3^0\end{pmatrix}\otimes\mathbf{I}_q\right]\boldsymbol{\alpha}''(u_0)\\
&+\left[\frac{n_0}{p(u_0)}\begin{pmatrix}1 & h\mu_1^0\\ h\mu_1^0 & h^2\mu_2^0\end{pmatrix}^{-1}\otimes\boldsymbol{\Phi}\right]\mathbf{T}_k^{\star}+o_p(h^2),
\end{aligned}\tag{A.12}$$

where

$$\begin{aligned}
\mathbf{T}_k^* =& \begin{pmatrix}\mathbf{T}_{k,0}^*\\ \mathbf{T}_{k,1}^*\end{pmatrix}=\frac{1}{kn_0}\mathbf{D}_{u_0}^{\mathrm{T}}\boldsymbol{\omega}_{u_0}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\\
=&\frac{1}{kn_0}\begin{pmatrix}\sum\limits_{i=1}^{k}\sum\limits_{j_1=1}^{n_0}\mathbf{Z}_{ij_1}K_h(U_{ij_1}-u_0)\sum\limits_{j_2=1}^{n_0}\sigma^{j_1j_2}\varepsilon_{ij_2}\\ \sum\limits_{i=1}^{k}\sum\limits_{j_1=1}^{n_0}\mathbf{Z}_{ij_1}K_h(U_{ij_1}-u_0)(U_{ij_1}-u_0)\sum\limits_{j_2=1}^{n_0}\sigma^{j_1j_2}\varepsilon_{ij_2}\end{pmatrix}.
\end{aligned}$$

From (A.12), we know

$$\overline{\boldsymbol{\alpha}}_k^w(u_0)-\boldsymbol{\alpha}(u_0)=\frac{n_0}{p(u_0)}\frac{\mu_2^0\boldsymbol{\Phi}\mathbf{T}_{k,0}^{\star}-\mu_1^0h^{-1}\boldsymbol{\Phi}\mathbf{T}_{k,1}^{\star}}{\mu_2^0-(\mu_1^0)^2}+\frac{h^2}{2}\frac{(\mu_2^0)^2-\mu_1^0\mu_3^0}{\mu_2^0-(\mu_1^0)^2}\boldsymbol{\alpha}''(u_0)+o_p(h^2),$$

the second term of which is the asymptotic bias of $\overline{\boldsymbol{\alpha}}_k^w(u_0)$, obviously. Let

$$\mathbf{Q}_k=\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\Big\{c_0+c_1\Big(\frac{U_{ij_1}-u_0}{h}\Big)\Big\}\mathbf{Z}_{ij_1}K_h(U_{ij_1}-u_0)\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\varepsilon_{ij_2},$$

where $c_0=\frac{\mu_2^0}{\mu_2^0-(\mu_1^0)^2}$ and $c_1=-\frac{\mu_1^0}{\mu_2^0-(\mu_1^0)^2}$. It follows that

$$\sqrt{kn_0h}\Big\{\overline{\boldsymbol{\alpha}}_k^w(u_0)-\boldsymbol{\alpha}(u_0)-\frac{h^2}{2}\frac{(\mu_2^0)^2-\mu_1^0\mu_3^0}{\mu_2^0-(\mu_1^0)^2}\boldsymbol{\alpha}''(u_0)\Big\}=\frac{n_0}{p(u_0)}\boldsymbol{\Phi}\sqrt{kn_0h}\mathbf{Q}_k+o_p(1).$$

Denote $\mathbf{P}=\{\mathbf{Z}_{11}(\frac{U_{11}-u_0}{h}),\cdots,\mathbf{Z}_{1n_0}(\frac{U_{1n_0}-u_0}{h}),\cdots,\mathbf{Z}_{kn_0}(\frac{U_{kn_0}-u_0}{h})\}$. The covariance matrix of $\sqrt{kn_0h}\,\mathbf{Q}_k$ is

$$\begin{aligned}
\mathrm{Cov}(\sqrt{kn_0h}\,\mathbf{Q}_k)=&\frac{hc_0^2}{kn_0}E\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\omega}_{u_0}(\mathbf{I}_k\otimes\boldsymbol{\varpi})\boldsymbol{\omega}_{u_0}\mathbf{Z}\}+\frac{hc_1^2}{kn_0}E\{\mathbf{P}\boldsymbol{\omega}_{u_0}(\mathbf{I}_k\otimes\boldsymbol{\varpi})\boldsymbol{\omega}_{u_0}\mathbf{P}^{\mathrm{T}}\}\\
&+\frac{2hc_1c_0}{kn_0}E\{\mathbf{P}\boldsymbol{\omega}_{u_0}(\mathbf{I}_k\otimes\boldsymbol{\varpi})\boldsymbol{\omega}_{u_0}\mathbf{Z}\}\\
=&\frac{p(u_0)}{n_0}(c_0^2\nu_0^0+2c_0c_1\nu_1^0+c_1^2\nu_2^0)\boldsymbol{\Phi}^{-1}+o(1).
\end{aligned}$$

For any nonzero q-vector $\boldsymbol{\Lambda}$, let

$$a_i=\sqrt{h}\sum_{j_1=1}^{n_0}\Big\{c_0+c_1\Big(\frac{U_{ij_1}-u_0}{h}\Big)\Big\}\boldsymbol{\Lambda}^{\mathrm{T}}\mathbf{Z}_{ij_1}K_h(U_{ij_1}-u_0)\sum_{j_2=1}^{n_0}\sigma^{j_1j_2}\varepsilon_{ij_2}$$

and $B_k^2 = \sum_{i=1}^{k} Ea_i^2$. Then

$$B_k^2 = kp(u_0)(c_0^2\nu_0^0 + 2c_0c_1\nu_1^0 + c_1^2\nu_2^0)\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Phi}^{-1}\mathbf{\Lambda} + o(k).$$

Simple calculation shows

$$\sum_{i=1}^{k} E|a_i|^3 \le O(1)\sum_{i=1}^{k} h^{\frac{3}{2}} \sum_{j_1=1}^{n_0} \Big[|c_0| + |c_1| \cdot \Big|\frac{U_{ij_1}-u_0}{h}\Big|\Big] K_h^3(U_{ij_1}-u_0) = O(h^{-\frac{1}{2}}).$$

It follows that $\lim_{k\to\infty} B_k^{-3} \sum_{i=1}^{k} E|a_i^3| = 0$. By the central limit theorem the proof of (A.5) is complete.

We are now ready to prove the main results. The proofs of Theorems 2.1 and 2.2 are similar to Lemma A.5. We here omit the details. By applying the $\sqrt{k}$-consistency of $\widehat{\boldsymbol\beta}_k$, the proof of Theorem 2.3 is the same as that of Theorem 3.1 in [17].

Proof of Theorem 4.1 According to the definition of $\widehat\sigma_\mu^2$, it can be written as

$$\begin{aligned}
\widehat\sigma_\mu^2 = {} & \frac{1}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\varepsilon_{ij_1}\varepsilon_{ij_2} + \frac{1}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\mathbf{X}_{ij_1}^{\mathrm{T}}(\boldsymbol\beta-\widehat{\boldsymbol\beta}_k)\mathbf{X}_{ij_2}^{\mathrm{T}}(\boldsymbol\beta-\widehat{\boldsymbol\beta}_k)\\
& + \frac{1}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\mathbf{Z}_{ij_1}^{\mathrm{T}}(\boldsymbol\alpha(U_{ij_1})-\widehat{\boldsymbol\alpha}_k(U_{ij_1}))\mathbf{Z}_{ij_2}^{\mathrm{T}}(\boldsymbol\alpha(U_{ij_2})-\widehat{\boldsymbol\alpha}_k(U_{ij_2}))\\
& + \frac{2}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\mathbf{X}_{ij_1}^{\mathrm{T}}(\boldsymbol\beta-\widehat{\boldsymbol\beta}_k)\varepsilon_{ij_2}\\
& + \frac{2}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\mathbf{Z}_{ij_1}^{\mathrm{T}}(\boldsymbol\alpha(U_{ij_1})-\widehat{\boldsymbol\alpha}_k(U_{ij_1}))\varepsilon_{ij_2}\\
& + \frac{2}{kn_0(n_0-1)}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\mathbf{X}_{ij_1}^{\mathrm{T}}(\boldsymbol\beta-\widehat{\boldsymbol\beta}_k)\mathbf{Z}_{ij_2}^{\mathrm{T}}(\boldsymbol\alpha(U_{ij_2})-\widehat{\boldsymbol\alpha}_k(U_{ij_2})).
\end{aligned}$$

By combining the root-k consistency of $\widehat{\boldsymbol\beta}_k$ and $\sup_{u\in\mathcal{U}}\|\boldsymbol\alpha(u)-\widehat{\boldsymbol\alpha}_k(u)\| = O_p\{h^2 + (\frac{\log k}{kh})^{\frac{1}{2}}\}$, it is easy to show that

$$\begin{aligned}
\sqrt{k}(\widehat\sigma_\mu^2 - \sigma_\mu^2) &= \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\Big\{(\mu_i^2-\sigma_\mu^2) + \frac{2}{n_0}\sum_{j=1}^{n_0}\mu_i\nu_{ij} + \frac{1}{n_0(n_0-1)}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\nu_{ij_1}\nu_{ij_2}\Big\} + o_p(1)\\
&= \frac{1}{\sqrt{k}}\sum_{i=1}^{k} a_i + o_p(1).
\end{aligned}$$

It is easy to show that

$$Ea_i^2 = \mathrm{Var}(\mu_1^2) + \frac{4}{n_0}\sigma_\mu^2\sigma_\nu^2 + \frac{2}{n_0(n_0-1)}\sigma_\nu^4.$$

Hence $B_k^2 = \sum_{i=1}^{k} Ea_i^2 = ck$ for some positive constant c. From Lemma A.4, $E|\mu_1|^{4+\delta_2} < \infty$ and

$E|\nu_{11}|^{4+\delta_2} < \infty$, simple calculations show

$$E|\mu_i^2 - \sigma_\mu^2|^{2+\frac{\delta_2}{2}} \le c, \quad E\Big|\frac{2}{n_0}\sum_{j=1}^{n_0}\mu_i\nu_{ij}\Big|^{2+\frac{\delta_2}{2}} < c \ \text{ and } \ E\Big|\frac{1}{n_0(n_0-1)}\sum_{j_1=1}^{n_0}\sum_{j_2\ne j_1}\nu_{ij_1}\nu_{ij_2}\Big|^{2+\frac{\delta_2}{2}} < c$$

for some positive constant δ_2. Therefore, $\sum\limits_{i=1}^{k} Ea_i^{2+\frac{\delta_2}{2}} = O(k)$. It follows that

$$\frac{1}{B_k^{2+\frac{\delta_2}{2}}}\sum_{i=1}^{k} Ea_i^{2+\frac{\delta_2}{2}} = O(k^{-\frac{\delta_2}{2}}) \to 0, \quad \text{as } k \to \infty.$$

Thus the first result of Theorem 4.1 follows the central limit theorem.

By the same argument, we can show that (4.3) holds.

Proof of Theorem 4.2 The proof of Theorem 4.2 is trivial. We here omit the details.

Proof of Theorem 5.1 According to the definition of $\widehat{\boldsymbol\beta}_k^w$, $\widetilde{\boldsymbol\beta}_k^w$ and the fact $a_1a_2 - b_1b_2 = (a_1-b_1)(a_2-b_2) + (a_1-b_1)b_2 + b_1(a_2-b_2)$, we have

$$\begin{aligned}
\widehat{\boldsymbol\beta}_k^w - \boldsymbol\beta = \widetilde{\boldsymbol\beta}_k^w - \boldsymbol\beta &+ \{(\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}\widehat{\mathbf X}^w)^{-1} - (\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w)^{-1}\}\\
&\cdot\{(\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}(\mathbf I_{kn_0} - \mathbf S^w) - \widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\}\mathbf M\\
&+ \{(\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}\widehat{\mathbf X}^w)^{-1} - (\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w)^{-1}\}\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{S}^w)\mathbf M\\
&+ (\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w)^{-1}\{\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}(\mathbf I_{kn_0} - \mathbf S^w) - \widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\}\mathbf M\\
&+ \{(\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}\widehat{\mathbf X}^w)^{-1} - (\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w)^{-1}\}\\
&\cdot\{(\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}(\mathbf I_{kn_0} - \mathbf S^w) - \widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\}\boldsymbol\varepsilon\\
&+ \{(\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}\widehat{\mathbf X}^w)^{-1} - (\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w)^{-1}\}\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\boldsymbol\varepsilon\\
&+ (\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w)^{-1}\{\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}(\mathbf I_{kn_0} - \mathbf S^w) - \widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\}\boldsymbol\varepsilon,
\end{aligned}$$

where $\widetilde{\boldsymbol\beta}_k^w$ is defined in (7.1), $\mathbf M$ and $\boldsymbol\varepsilon$ is defined in the proof of Lemma A.6, and $\mathbf S^w$ is defined in Section 2. Therefore, in order to prove Theorem 5.1, combining Lemma A.5 and the following fact

$$(\mathbf A + a\mathbf B)^{-1} = \mathbf A^{-1} - a\mathbf A^{-1}\mathbf B\mathbf A^{-1} + O(a^2), \quad \text{as } a \to 0,$$

we only need to show

$$\frac{1}{kn_0}(\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}\widehat{\mathbf X}^w - \widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w) = O_p(k^{-\frac12}), \tag{A.13}$$

$$\frac{1}{kn_0}\{\widehat{\mathbf X}^{wP\mathrm T}\widehat{\boldsymbol\Omega}^{-1}(\mathbf I_{kn_0} - \mathbf S^w) - \widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\}\mathbf M = o_p(k^{-\frac12}), \tag{A.14}$$

$$\frac{1}{kn_0}\{\widehat{\mathbf X}^{w\mathrm T}\widehat{\boldsymbol\Omega}^{-1}(\mathbf I_{kn_0} - \mathbf S^w) - \widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\}\boldsymbol\varepsilon = o_p(k^{-\frac12}), \tag{A.15}$$

$$kn_0(\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}\widetilde{\mathbf X}^w)^{-1} = O_p(1) \tag{A.16}$$

and

$$\frac{1}{kn_0}\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\mathbf M = o_p(k^{-\frac12}), \quad \frac{1}{kn_0}\widetilde{\mathbf X}^{w\mathrm T}\boldsymbol\Omega^{-1}(\mathbf I_{kn_0} - \widetilde{\mathbf S}^w)\boldsymbol\varepsilon = O_p(k^{-\frac12}). \tag{A.17}$$

The (s_1, s_2)th entry of the matrix in the left-hand side of (A.6) is

$$\frac{1}{kn_0}\Big\{(\widehat{\eta}^{\,-2}+\widehat{\sigma}_{\nu}^{\,-2})\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\widehat{X}^w_{ij_1s_1}\widehat{X}^w_{ij_2s_2}-(\eta^{-2}+\sigma_\nu^{-2})\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\widetilde{X}^w_{ij_1s_1}\widetilde{X}^w_{ij_2s_2}\Big\}$$
$$+\frac{1}{kn_0}\Big(\widehat{\sigma}_{\nu}^{\,-2}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widehat{X}^w_{ijs_1}\widehat{X}^w_{ijs_2}-\sigma_\nu^{-2}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widetilde{X}^w_{ijs_1}\widetilde{X}^w_{ijs_2}\Big)=\mathrm{J}_1+\mathrm{J}_2,$$

where $\widehat{X}^w_{ijs}$ and $\widetilde{X}^w_{ijs}$ are the sth entry of $\widehat{\mathbf{X}}^w_{ij}$ and $\widetilde{\mathbf{X}}^w_{ij}$, respectively. It is easy to see that

$$\begin{aligned}\mathrm{J}_2&=\frac{1}{kn_0}\widehat{\sigma}_{\nu}^{\,-2}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(\widehat{X}^w_{ijs_1}-\widetilde{X}^w_{ijs_1})(\widehat{X}^w_{ijs_2}-\widetilde{X}^w_{ijs_2})+\frac{1}{kn_0}\widehat{\sigma}_{\nu}^{\,-2}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(\widehat{X}^w_{ijs_1}-\widetilde{X}^w_{ijs_1})\widetilde{X}^w_{ijs_2}\\&\quad+\frac{1}{kn_0}\widehat{\sigma}_{\nu}^{\,-2}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widetilde{X}^w_{ijs_1}(\widehat{X}^w_{ijs_2}-\widetilde{X}^w_{ijs_2})+\frac{1}{kn_0}(\widehat{\sigma}_{\nu}^{\,-2}-\sigma_\nu^{-2})\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widetilde{X}^w_{ijs_1}\widetilde{X}^w_{ijs_2}\\&=\mathrm{J}_{21}+\mathrm{J}_{22}+\mathrm{J}_{23}+\mathrm{J}_{24}.\end{aligned}$$

Denote

$$\begin{aligned}\widehat{\mathbf{A}}_1&=(\mathbf{D}^{\mathrm{T}}_{U_{ij}}\boldsymbol{\omega}_{U_{ij}}\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{D}_{U_{ij}})^{-1},\quad \widehat{\mathbf{A}}_2=\mathbf{D}^{\mathrm{T}}_{U_{ij}}\boldsymbol{\omega}_{U_{ij}}\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\mathbf{d}_{p,s},\\ \mathbf{A}_1&=(\mathbf{D}^{\mathrm{T}}_{U_{ij}}\boldsymbol{\omega}_{U_{ij}}\boldsymbol{\Omega}^{-1}\mathbf{D}_{U_{ij}})^{-1},\quad \mathbf{A}_2=\mathbf{D}^{\mathrm{T}}_{U_{ij}}\boldsymbol{\omega}_{U_{ij}}\boldsymbol{\Omega}^{-1}\mathbf{X}\mathbf{d}_{p,s}.\end{aligned}$$

By the root-k consistency of $\widehat{\sigma}_{\mu}^{\,2}$ and $\widehat{\sigma}_{\nu}^{\,2}$, we can show

$$\begin{aligned}&\max_{1\le i\le k}\max_{1\le j\le n_0}k\|\widehat{\mathbf{A}}_1-\mathbf{A}_1\|=O_p(k^{-\frac12}),\quad \max_{1\le i\le k}\max_{1\le j\le n_0}k^{-1}\|\widehat{\mathbf{A}}_2-\mathbf{A}_2\|=O_p(k^{-\frac12}),\\&\max_{1\le i\le k}\max_{1\le j\le n_0}k\mathbf{A}_1=O_p(1),\qquad\qquad\qquad \max_{1\le i\le k}\max_{1\le j\le n_0}k^{-1}\mathbf{A}_2=O_p(1).\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{J}_{22}&=\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}[(\mathbf{Z}^{\mathrm{T}}_{ij}\mathbf{0}^{\mathrm{T}}_q)(\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_2-\mathbf{A}_1\mathbf{A}_2)]\widetilde{X}^w_{ijs_2}\\&\le\max_{1\le i\le k}\max_{1\le j\le n_0}(\|\widehat{\mathbf{A}}_1-\mathbf{A}_1\|\cdot\|\widehat{\mathbf{A}}_2-\mathbf{A}_2\|+\|\widehat{\mathbf{A}}_1-\mathbf{A}_1\|\cdot\|\mathbf{A}_2\|+\|\mathbf{A}_1\|\cdot\|\widehat{\mathbf{A}}_2-\mathbf{A}_2\|)\\&\quad\cdot\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(\mathbf{Z}^{\mathrm{T}}_{ij}\mathbf{Z}_{ij}+\widetilde{\mathbf{X}}^{w\mathrm{T}}_{ij}\widetilde{\mathbf{X}}^w_{ij})=O_p(k^{-\frac12}).\end{aligned}$$

By the same argument we can show that $\mathrm{J}_{21}=O_p(k^{-\frac12})$ and $\mathrm{J}_{23}=O_p(k^{-\frac12})$. Moreover, combining the root-k consistency of $\widehat{\sigma}_{\nu}^{\,2}$ and Lemma A.5 we can show that $\mathrm{J}_{24}=O_p(k^{-\frac12})$. This implies that $\mathrm{J}_2=O_p(k^{-\frac12})$. Following the same line, we can obtain $\mathrm{J}_1=O_p(k^{-\frac12})$. Thus, (A.13) holds.

According to the proof of (A.4) and the root-k consistency of $\widehat{\sigma}_{\mu}^{\,2}$ and $\widehat{\sigma}_{\nu}^{\,2}$, we have

$$\frac{1}{kn_0}\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{I}_{kn_0}-\widetilde{\mathbf{S}}^w)\mathbf{M}=o_p(k^{-\frac12}),\quad \frac{1}{kn_0}\widehat{\mathbf{X}}^{w\mathrm{T}}\widehat{\boldsymbol{\Omega}}^{-1}(\mathbf{I}_{kn_0}-\mathbf{S}^w)\mathbf{M}=o_p(k^{-\frac12})$$

and

$$\frac{1}{kn_0}\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}\widetilde{\mathbf{S}}^w\boldsymbol{\varepsilon}=o_p(k^{-\frac12}),\quad \frac{1}{kn_0}\widehat{\mathbf{X}}^{w\mathrm{T}}\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{S}^w\boldsymbol{\varepsilon}=o_p(k^{-\frac12}),\quad \frac{1}{kn_0}\widetilde{\mathbf{X}}^{w\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}=O_p(k^{-\frac12}).$$

This implies that (A.14) and (A.17) hold. In order to prove (A.15), it is sufficient to show

$$\frac{1}{kn_0}\widehat{\mathbf{X}}^{w\mathrm{T}}\widehat{\mathbf{\Omega}}^{-1}\boldsymbol{\varepsilon} - \frac{1}{kn_0}\widetilde{\mathbf{X}}^{w\mathrm{T}}\mathbf{\Omega}^{-1}\boldsymbol{\varepsilon} = o_p(k^{-\frac{1}{2}}). \tag{A.18}$$

Note that the sth entry of (A.18) is

$$\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j_1=1}^{n_0}\sum_{j_2=1}^{n_0}\{(\widehat{\eta}^{-2}+\widehat{\sigma}_{\nu}^{-2})\widehat{X}_{ij_1s}^{w} - (\eta^{-2}+\sigma_{\nu}^{-2})\widetilde{X}_{ij_1s}^{w}\}(\mu_i+\nu_{ij_2})$$
$$+\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(\widehat{\sigma}_{\nu}^{-2}\widehat{X}_{ijs}^{w} - \sigma_{\nu}^{-2}\widetilde{X}_{ijs}^{w})(\mu_i+\nu_{ij}) = \mathrm{J}_3+\mathrm{J}_4$$

and

$$\mathrm{J}_4 = (\widehat{\sigma}_{\nu}^{-2}-\sigma_{\nu}^{-2})\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}\widehat{X}_{ijs}^{w}(\mu_i+\nu_{ij}) + \widehat{\sigma}_{\nu}^{-2}\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(\widetilde{X}_{ijs}^{w}-\widehat{X}_{ijs}^{w})(\mu_i+\nu_{ij}) = \mathrm{J}_{41}+\mathrm{J}_{42}.$$

By the root-k consistency of $\widehat{\sigma}_{\nu}^{-2}$ and the proof of Lemma A.6 we have $\mathrm{J}_{41} = O_p(k^{-1}) = o_p(k^{-\frac{1}{2}})$. Moreover, it is easy to see that

$$\mathrm{J}_{42} = \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(\mathbf{Z}_{ij}^{\mathrm{T}},\mathbf{0}_q^{\mathrm{T}})\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_2(\mu_i+\nu_{ij}) - \frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(\mathbf{Z}_{ij}^{\mathrm{T}},\mathbf{0}_q^{\mathrm{T}})\mathbf{A}_1\mathbf{A}_2(\mu_i+\nu_{ij}) = o_p(k^{-\frac{1}{2}}).$$

So (A.15) holds.

At last, Lemma A.5 implies that (A.16) holds. Thus, the proof is complete.

Proof of Theorem 5.2 According to the root-k consistency of $\widehat{\boldsymbol{\beta}}_k$, $\widehat{\sigma}_{\mu}^{2}$ and $\widehat{\sigma}_{\nu}^{2}$ and (A.5) in Lemma A.6, it is easy to complete the proof.

Proof of Theorem 5.3 Applying the $\sqrt{k}$-consistency of $\widehat{\boldsymbol{\beta}}_k^{w}$, the proof of Theorem 5.3 is the same as that of Theorem 3.1 in [17].

Proof of Theorem 5.4 According to the root-k consistency of $\widehat{\sigma}_{\mu}^{2}$ and $\widehat{\sigma}_{\nu}^{2}$ and Lemma A.5 we can complete the proof.

Proof of Theorem 6.1 It is easy to see that

$$\begin{aligned}
\mathrm{RSS}_0-\mathrm{RSS}_1 = {} & (kn_0)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}[\mathbf{X}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_k)+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{a}(U_{ij},\boldsymbol{\theta})-\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{a}(U_{ij},\widehat{\boldsymbol{\theta}}_k)]^2\\
& -(kn_0)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}[\mathbf{X}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_k)+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{a}(U_{ij},\boldsymbol{\theta})-\mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(U_{ij})]^2\\
& +2(kn_0)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}[\mathbf{X}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_k)+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{a}(U_{ij},\boldsymbol{\theta})-\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{a}(U_{ij},\widehat{\boldsymbol{\theta}}_k)]^{\mathrm{T}}\varepsilon_{ij}\\
& -2(kn_0)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_0}[\mathbf{X}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_k)+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{a}(U_{ij},\boldsymbol{\theta})-\mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(U_{ij})]^{\mathrm{T}}\varepsilon_{ij}\\
= {} & \mathrm{J}_1-\mathrm{J}_2+\mathrm{J}_3-\mathrm{J}_4.
\end{aligned}$$

Under the null hypothesis, applying the consistency of $\widehat{\boldsymbol{\beta}}_k$, $\widehat{\boldsymbol{\theta}}_k$ and $\widehat{\boldsymbol{\alpha}}_k(\cdot)$, it is easy to show that $J_s \to_p 0$ as $k \to \infty$ for $s = 1, 2, 3$ and 4. This implies that under the null hypothesis $\mathrm{RSS}_0 - \mathrm{RSS}_1 \to_p 0$. Therefore, in order to complete the proof of the first result, we just need to show that RSS_1 is bounded away from zero and infinity. According to the proof of Theorems 4.1 and 4.2, we can show

$$\mathrm{RSS}_1 = (kn_0)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_0} \varepsilon_{ij}^2 + o_p(1) \to_p \sigma_\mu^2 + \sigma_\nu^2.$$

Thus, the proof of the first result is complete. The proof of the second result is the same and we omit the details here.

References

[1] Baltagi, B. H., Econometric Analysis of Panel Data, John Wiley & Sons, New York, 1995.

[2] Cai, Z., Fan, J. and Yao, Q., Functional-coefficient regression models for nonlinear time series, *J. Amer. Statist. Assoc.*, **95**, 2000, 941–956.

[3] Carroll, R. J., Ruppert, D. and Welsh, A. H., Nonparametric estimation via local estimating equations, *J. Amer. Statist. Assoc.*, **93**, 1998, 214–227.

[4] Fan, J. Q. and Huang, T., Profile likelihood inferences on semiparametric varying-coefficient partially linear models, *Bernoulli*, **11**(6), 2005, 1031–1057.

[5] Fan, J. Q. and Zhang, W., Statistical estimation in varying coefficient models, *Ann. Statist.*, **27**, 1999, 1491–1518.

[6] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P., Nonparametric smoothing estimates of time-varying coeffiocient models with longitudinal data, *Biometrika*, **85**, 1998, 809–822.

[7] Huang, J. Z., Wu, C. O. and Zhou, L., Polynomial spline estimation and inference for varying coefficient models with longitudinal data, *Statist. Sinica*, **14**, 2004, 763–788.

[8] Kniesner, T. and Li, Q., Semiparametric panel data models with dynamic adjustment: Theoretical considerations and an application to labor supply, University of Guelph, manuscript, 1994.

[9] Li, Q., Huang, C. J., Li, D. and Fu, T. T., Semiparametric smooth coefficient models, *J. Business and Econ. Statist.*, **20**, 2002, 412–422.

[10] Lin, X. H. and Carroll, R. J., Nonparametric function estimation for clustered data when the predictor is measured without/with error, *J. Amer. Statist. Assoc.*, **95**, 2000, 520–534.

[11] Müller, H. G. and Zhao, P., On a semiparametric variance function model and a test for heteroscedasticity, *Ann. Statist.*, **23**, 1995, 946–967.

[12] Ruckstuhl, A. F., Welsh, A. H. and Carroll, R. J., Nonparametric function estimation of the relationship between two repeatedly measured variables, *Statist. Sinica*, **10**, 2000, 51–71.

[13] Shi, J. and Lau, T. S., Empirical likelihood for partially linear models, *J. Multivariate Anal.*, **72**, 2000, 132–149.

[14] Stout, W. F., Almost Sure Convergence, Academic Press, New York, 1974.

[15] Verbeke, G. and Molenberghs, G., Linear Mixed Models for Longitudinal Data, Springer Series in Statistics, Springer-Verlag, New York, 2000.

[16] Wu, C. O., Chiang, C. T. and Hoover, D. R., Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data, *J. Amer. Statist. Assoc*, **93**, 1998, 1388–1402.

[17] Xia, Y. and Li, W. K., On the estimation and testing of functional-coefficient linear models, *Statistica Sinica*, **9**, 1999, 737–757.

[18] You, J. H., Xu, Q. F. and Zhou, B., Statistical inference for partially linear regression models with measurement errors, *Chin. Ann. Math.*, **29B**(2), 2008, 207–222.

[19] Zeger, S. L. and Diggle, P. J., Semiparametric model for longitudinal data with application to CD4 cell numbers in HIV seroconvertiers, *Biometrics*, **50**, 1994, 689–699.

[20] Zhang, W., Lee, S. Y. and Song, X., Local polynomial fitting in semivarying coefficient models, *J. Multivariate Anal.*, **82**, 2002, 166–188.

[21] Zhou, Y. and You, J. H., Strong convergence rates of several estimators in varying-coefficient partially linear models, University of North Carolina at Chapel Hill, manuscript, 2003.