Journal of Multivariate Analysis 154 (2017) 96–111

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Improved estimation of fixed effects panel data partially linear models with heteroscedastic errors

Jianhua Hu [a,b,*], Jinhong You [a,b], Xian Zhou [c]

[a] *School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China*
[b] *Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China*
[c] *Department of Applied Finance and Actuarial Studies, Macquarie University, Sydney, NSW 2109, Australia*

ARTICLE INFO

Article history:
Received 6 April 2016
Available online 4 November 2016

AMS 2000 classifications:
primary 62H12
secondary 62A01

Keywords:
Consistent estimator
Fixed effects
Heteroscedastic errors
Incidental parameter
Partially linear

ABSTRACT

Fixed effects panel data regression models are useful tools in econometric and microarray analysis. In this paper, we consider statistical inferences under the setting of fixed effects panel data partially linear regression models with heteroscedastic errors. We find that the usual local polynomial estimator of the error variance function based on residuals is inconsistent, and develop a consistent estimator. Applying this consistent estimator of error variance and spline series approximation of the nonparametric component, we further construct a weighted semiparametric least squares dummy variables estimator for the parametric and nonparametric components. Asymptotic normality of the proposed estimator is derived and its asymptotic covariance matrix estimator is provided. The proposed estimator is shown to be asymptotically more efficient than those ignoring heteroscedasticity. Simulation studies are conducted to demonstrate the finite sample performances of the proposed procedure. As an application, a set of economic data is analyzed by the proposed method.

1. Introduction

Panel data refer to the pooling of observations on a cross-section of subjects, such as households, countries, firms, etc., over a time period, which can be achieved by surveying a sample of subjects and following them over time; see, e.g., Baltagi [4]. Such a two-dimensional information set enables researchers to estimate complex models and draw efficient statistical inferences that may not be possible using pure time-series data or cross-section data. Both theoretical developments and applied works in panel data analysis have become more popular in recent years.

Panel data parametric (mainly linear) regression models have been the dominant framework for analyzing panel data; see [1,4] for summaries of early work and [18] for a recent comprehensive survey. If correctly specified, the parametric model has the advantages of easy interpretation and efficient estimation. In practice, however, correct parameterization is often difficult or unavailable, and a misspecification of the model could lead to biased and misleading estimates of the underlying parameters. To address this issue, various more flexible models have been introduced in literature of statistics and econometrics. Among the most important is the panel data partially linear regression model, which allows unspecified relationship between the response variable and some covariate(s).

* Corresponding author at: School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China.
E-mail address: frank.jianhuahu@gmail.com (J. Hu).

http://dx.doi.org/10.1016/j.jmva.2016.10.010
0047-259X/

Let $(\mathbf{X}_{it}, \mathbf{U}_{it}, Y_{it})$ denote the observations collected from the ith subject, $i = 1, \ldots, n$, at time t, $t = 1, \ldots, T$, where $Y_{it} \in \mathbb{R}$ is the response of interest, $\mathbf{X}_{it} = (X_{it1}, \ldots, X_{itp})^\top \in \mathbb{R}^p$ is a p-vector of linear predictors and $\mathbf{U}_{it} = (U_{it1}, \ldots, U_{itq})^\top$ is a q-vector of nonlinear predictors. A typical panel data partially linear regression model has the form

$$Y_{it} = \mathbf{X}_{it}^\top \boldsymbol{\beta} + g(\mathbf{U}_{it}) + \varepsilon_{it}, \quad \varepsilon_{it} = \mu_i + \nu_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is an unknown parameter vector, $g(\cdot)$ is an unknown function, μ_i is the individual effect of subject i and ν_{it} is an *idiosyncratic* error. Depending on whether μ_i is correlated with the observed explanatory variables $(\mathbf{X}_{it}^\top, \mathbf{U}_{it}^\top)^\top$ or not, model (1.1) can be divided into two classes. One is the *random effects* model in which μ_i is uncorrelated with $(\mathbf{X}_{it}^\top, \mathbf{U}_{it}^\top)^\top$, so that $\mathrm{E}\{\mu_i(\mathbf{X}_{it}^\top, \mathbf{U}_{it}^\top)^\top\} = 0$, where E is the symbol of expectation; and another is the *fixed effects* model in which μ_i is correlated with $(\mathbf{X}_{it}^\top, \mathbf{U}_{it}^\top)^\top$, i.e., $\mathrm{E}\{\mu_i(\mathbf{X}_{it}^\top, \mathbf{U}_{it}^\top)^\top\} \neq 0$. Considering a large number of random draws from the cross section, it makes sense for us to treat the individual effects, μ_i, as random draws from the population. Fixed effects panel data model, however, is appropriate if the interest is in a specific set of subjects, such as specific firms, a set of OECD countries, all American states, and so on.

For model (1.1) with fixed effects, the individual effect is often viewed as a parameter to be estimated. It is typically assumed that T is finite and n is large. Consequently, the number of parameters grows with the sample size, and the ordinary least squares (OLS) or maximum likelihood estimator (MLE) would lead to inconsistent estimates of the common parameter of interest. This is well-known as the *incidental parameter problem*; see [19] for a general discussion on this problem. Due to the incidental parameter problem, it is a great challenge to construct consistent estimators for the parametric and nonparametric components in the fixed effects panel data partially linear regression model, and few results were available until Baltagi and Li [5], who proposed a difference-based series estimation (DSE) for the parametric component and nonparametric component. They established the asymptotic normality of the former and derived the convergence rate of the latter. This DSE, however; is not efficient when $T > 2$, see [2].

Fan et al. [9] found that the model (1.1) with fixed effects is also useful to conduct the microarray analysis of the neuroblastoma cell in response to macrophage migration inhibitory factor (MIF). Fan et al. [8] proposed a novel profile least squares estimation (PLSE) for the parametric component and a local linear estimation for the nonparametric component by the back fitting method. They established the asymptotic normality of the former and the MSE upper bound of the latter. In addition, the estimation problem of the model (1.1) with fixed effects was considered in [13,24], while the problem of estimating a varying-coefficient panel data model with fixed effects was studied in [21].

All above-mentioned results assumed that the idiosyncratic errors ν_{it} are independent and identically distributed (i.i.d.). In practice, however, the random errors are often *heteroscedastic* (with unequal variances). For example, heteroscedasticity has been found in gasoline demands across Organization for Economic Co-operation and Development (OECD) countries, in steam electricity generations across utilities of different sizes, in cost functions for US airline firms, and in medical expenditures [4]. It is well known that when heteroscedasticity is present, ignoring its impact will result in inefficient estimators of the regression coefficients and biased estimators of covariance. Under the setting of fixed effects panel data linear regression model, Kézdi [14] and Stock and Watson [20] investigated the consistent estimations of the regression parameters and asymptotic properties. They found that, due to the incidental parameter problem, the conventional heteroscedasticity-robust (HR) asymptotic covariance matrix estimator is inconsistent, and further provided a $\sqrt{n}$ consistently bias-adjusted HR estimator. However, [14,20] did not make any assumption about the heteroscedasticity. Therefore, the error variance in their case could not be estimated and the information of the heteroscedasticity could not be taken into account to improve the estimation of the mean parameter. There is another important situation of heteroscedasticity where the error variance is a function of some of the predictors, see [3,10,23]. In this situation, the error variance could usually be estimated and the information of heteroscedasticity could be taken into account to improve the estimation of the mean parameter, see Amemiya [3] for the cross-sectional data, Fan and Yao [10] for the time series data and You et al. [23] for the random effects panel data. To date, however, whether the error variance could be estimated consistently and whether the information of heteroscedasticity could be utilized to improve the estimation of the mean parameter remains unsolved even in the setting of fixed effects panel data linear regression model. In this paper, we address these issues under the more general partially linear model (1.1).

As in [8], we assume that the errors are heteroscedastic and the error variance is a smoothing function of $\mathbf{V}_{it}$ in the form:

$$\varepsilon_{it} = \mu_i + \sigma(\mathbf{V}_{it})\nu_{it}, \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T, \tag{1.2}$$

where $\sigma(\cdot)$ is an unknown function, $\mathrm{var}(\nu_{it}) = 1$, and $\mathbf{V}_{it} = (V_{it1}, \ldots, V_{itm})^\top$ is a known vector. $\mathbf{V}_{it}$ may be a function of $\mathbf{X}_{it}$ and $\mathbf{U}_{it}$, such as $\mathbf{V}_{it} = \mathbf{U}_{it}$ in [8]. We find that the usual residuals-based local polynomial estimator of the error variance function $\sigma(\cdot)$ is not consistent, and will propose an alternative consistent estimator. Applying the proposed estimator together with spline series approximation of the nonparametric component, we further construct a weighted semiparametric least squares dummy variables estimator for the parametric components and a weighted spline series estimator for nonparametric components. Asymptotic normal distributions for the proposed estimators are derived and asymptotic covariance matrix estimators are provided. The proposed estimator is shown to be asymptotically more efficient than those ignoring the heteroscedasticity. The results can be extended to more general situation, such as the case where the noise level may be a smoothing function of the mean, and so on.

Throughout this paper, we choose $\mathrm{E}\{g(\mathbf{U}_{it})\} = 0$ as our identification condition and assume that n is large and T is small and fixed with $T \geq 2$. The remainder of the paper is as follows. The pilot estimators of the parametric and nonparametric components are presented in Section 2. A consistent estimator of the error variance is constructed in Section 3. In Section 4 we construct a weighted semiparametric least squares dummy variables estimator of parametric regression coefficients and a weighted spline series estimator of nonparametric components. Section 5 reports some simulation results. An application is illustrated in Section 6, followed by concluding remarks in Section 7. Proofs of the main results are relegated to Appendix.

2. Pilot estimators

For convenience, we assume that $q = 1, m = 1$, and write $\mathbf{U}_{it} = U_{it}, \mathbf{V}_{it} = V_{it}$. It is not difficult to extend our investigation results to the case of $q > 1$ and $m > 1$. More discussions could be found in the concluding remarks. In this section we will neglect the heteroscedasticity to construct the estimators of the parametric and nonparametric components by combining the spline series approximation and least squares. Due to the identification condition $\mathrm{E}\{g(U_{it})\} = 0$, same as [15,16,22], we apply the centered B-spline basis to approximate the unknown function $g(U_{it})$. Specifically, let $\{\zeta_l(u), l = 1, 2, \ldots\}$ denote a sequence of the centered B-spline basis functions. In addition, let $K \equiv K(n)$ be a sequence of integers such that $K \to \infty$ as $n \to \infty$, $\boldsymbol{\zeta}^K(u) = (\zeta_1(u), \ldots, \zeta_K(u))^\top$, $\boldsymbol{\zeta}_{it} = \boldsymbol{\zeta}^K(U_{it})$, $\boldsymbol{\zeta}_i = (\boldsymbol{\zeta}_{i1}, \ldots, \boldsymbol{\zeta}_{iT})^\top$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^\top, \ldots, \boldsymbol{\zeta}_n^\top)^\top$. We have suppressed the dependence of $\boldsymbol{\zeta}_{it}$, $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}$ on K for ease of notation, where $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}$ are $T \times K$ and $nT \times K$ matrices, respectively.

Under fairly mild conditions, we can approximate $g(u)$ in model (1.1) very well by $\boldsymbol{\theta}^\top \boldsymbol{\zeta}^K(u)$ for some K-dimensional vector $\boldsymbol{\theta}$. As a result, model (1.1) can be approximated by

$$Y_{it} \approx \mathbf{X}_{it}^\top \boldsymbol{\beta} + \boldsymbol{\zeta}^{K\top}(U_{it})\boldsymbol{\theta} + \mu_i + \sigma(V_{it})\nu_{it}, \quad i = 1, \ldots, n,\ t = 1, \ldots, T. \tag{2.1}$$

Let $\mathbf{B} = \mathbf{I}_n \otimes \mathbf{1}_T$, where $\otimes$ denotes the Kronecker product, $\mathbf{Y} = (Y_{11}, \ldots, Y_{1T}, \ldots, Y_{nT})^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ and $\boldsymbol{\varepsilon} = (\sigma(V_{11})\nu_{11}, \ldots, \sigma(V_{1T})\nu_{1T}, \ldots, \sigma(V_{nT})\nu_{nT})^\top$. Then model (2.1) can be written in a matrix form as

$$\mathbf{Y} \approx \mathbf{B}\boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\zeta}\boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{2.2}$$

Define $\mathbf{M_B} = \mathbf{I}_{nT} - \mathbf{P_B}$ with $\mathbf{P_B} = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top$. Pre-multiplying (2.2) by $\mathbf{M_B}$ leads to

$$\mathbf{M_B Y} \approx \mathbf{M_B X}\boldsymbol{\beta} + \mathbf{M_B}\boldsymbol{\zeta}\boldsymbol{\theta} + \mathbf{M_B}\boldsymbol{\varepsilon}. \tag{2.3}$$

It is easy to see that

$$\mathbf{M_B}\boldsymbol{\varepsilon} = \left(\nu_{11} - \sum_{t=1}^T \nu_{1t}, \ldots, \nu_{1T} - \sum_{t=1}^T \nu_{1t}, \ldots, \nu_{nT} - \sum_{t=1}^T \nu_{nt}\right)^\top \quad \text{and} \quad \mathrm{E}(\mathbf{M_B}\boldsymbol{\varepsilon}) = \mathbf{0}_{nT}.$$

If we take $\mathbf{M_B}\boldsymbol{\varepsilon}$ as the residuals, then model (2.3) is a version of the usual linear regression. Based on (2.3), we can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ with a given K by minimizing the loss function

$$S(\boldsymbol{\beta}, \boldsymbol{\theta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\zeta}\boldsymbol{\theta})^\top \mathbf{M_B}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\zeta}\boldsymbol{\theta}). \tag{2.4}$$

The loss function (2.4) has a unique minimizer $(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_n)$ given by

$$\widehat{\boldsymbol{\beta}}_n = (\mathbf{X}^\top \mathbf{M_B M_{M_B\zeta} M_B X})^{-1}\mathbf{X}^\top \mathbf{M_B M_{M_B\zeta} M_B Y} \quad \text{and} \quad \widehat{\boldsymbol{\theta}}_n = (\boldsymbol{\zeta}^\top \mathbf{M_B}\boldsymbol{\zeta})^{-1}\boldsymbol{\zeta}^\top \mathbf{M_B}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_n).$$

It results a spline series estimator of $g(u)$, given by $\widehat{g}_n(u) = \boldsymbol{\zeta}^{K\top}(u)\widehat{\boldsymbol{\theta}}_n$. Here $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{g}_n(u)$ are called pilot estimators.

In order to present the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n, \widehat{g}_n(u)$ and other estimators proposed in subsequent sections, we need the following notations and technical assumptions.

Write $\boldsymbol{\Pi}_{it} = \mathbf{X}_{it} - \mathrm{E}(\mathbf{X}_{it}|U_{it})$, $\hbar_j(U_{it}) = \mathrm{E}(X_{itj}|U_{it}), j = 1, \ldots, p$, $\mathbf{M}_{\mathbf{1}_T} = \mathbf{I}_T - \mathbf{P}_{\mathbf{1}_T}$, $\boldsymbol{\Delta}_i = \mathrm{diag}(\sigma(V_{i1}), \ldots, \sigma(V_{iT}))$, $\mathbf{Q}_1 = T^{-1}\mathrm{E}(\boldsymbol{\zeta}_i^\top \mathbf{M}_{\mathbf{1}_T}\boldsymbol{\zeta}_i)$, $\mathbf{Q}_2 = T^{-1}\mathrm{E}(\boldsymbol{\zeta}_i^\top \mathbf{M}_{\mathbf{1}_T}\boldsymbol{\Delta}_i\mathbf{M}_{\mathbf{1}_T}\boldsymbol{\zeta}_i)$, $\boldsymbol{\Sigma}_1 = T^{-1}\mathrm{E}(\boldsymbol{\Pi}_i^\top \mathbf{M}_{\mathbf{1}_T}\boldsymbol{\Pi}_i)$ and $\boldsymbol{\Sigma}_2 = T^{-1}\mathrm{E}(\boldsymbol{\Pi}_i^\top \mathbf{M}_{\mathbf{1}_T}\boldsymbol{\Delta}_i\mathbf{M}_{\mathbf{1}_T}\boldsymbol{\Pi}_i)$.

Assumption 1. $(\mathbf{X}_{i1}^\top, \ldots, \mathbf{X}_{iT}^\top, U_{i1}, \ldots, U_{iT}, V_{i1}, \ldots, V_{iT})^\top$ are independent and identically distributed (i.i.d.) over $i = 1, \ldots, n$, and ν_{it} are i.i.d. over $i = 1, \ldots, n$, $t = 1, \ldots, T$, with mean 0 and variance 1. In addition, $\sum_{t=1}^T \mathrm{E}(\|\boldsymbol{\Pi}_{1t}\|^{2+\delta}) \le c < \infty$ and $\mathrm{E}(|\nu_{11}|^{2+\delta}) \le c < \infty$, $t = 1, \ldots, T$, for some $\delta > 0$ and constant $c > 0$.

Assumption 2. U_{it}'s are generated from a distribution which has a bounded support $\mathcal{U}_t$ and Lipschitz continuous density function $p_t^u(\cdot)$ such that $0 < \inf_{\mathcal{U}_t} p_t^u(\cdot) \le \sup_{\mathcal{U}_t} p_t^u(\cdot) < \infty$.

Assumption 3. V_{it}'s are generated from a distribution which has a bounded support $\mathcal{V}_t$ and Lipschitz continuous density function $p_t^v(\cdot)$ such that $0 < \inf_{\mathcal{V}_t} p_t^v(\cdot) \le \sup_{\mathcal{V}_t} p_t^v(\cdot) < \infty$.

Assumption 4. $g(\cdot), \sigma(\cdot)$ and $\hbar_j(\cdot)$ have the continuous second derivatives, $j = 1, \ldots, p$.

Assumption 5. $K = o(n^{1/2})$ and $n^{1/2}K^{-4} = o(1)$.

Assumption 6. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}, \mathbf{Q}_1$ are positive definite and $\boldsymbol{\zeta}^{K\top}(u)\mathbf{Q}_1^{-1}\mathbf{Q}_2\mathbf{Q}_1^{-1}\boldsymbol{\zeta}^K(u) > 0$ for all $u \in \bigcup_{t=1}^T \mathcal{U}_t$.

Denote $\Im = \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\right)^{-1/2}$ and $\eta_n(u) = \sqrt{nT}\left\{\boldsymbol{\zeta}^{K\top}(u)\mathbf{Q}_1^{-1}\mathbf{Q}_2\mathbf{Q}_2^{-1}\boldsymbol{\zeta}^K(u)\right\}^{-1/2}$. The asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$ is established in the following theorem.

Theorem 1 (*Asymptotic Normality*). *Under Assumptions 1–6,*

$$\sqrt{nT}\,\Im\,(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \rightsquigarrow \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p) \quad \text{as } n \to \infty,$$

where $\rightsquigarrow$ denotes convergence in distribution.

The next theorem provides the convergence rates and asymptotic normality of $\widehat{g}_n(u)$.

Theorem 2 (*Convergence Rate*). *Under Assumptions 1–6,*

(i) $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| = O_p(\sqrt{K/n} + K^{-2})$;
(ii) $\sum_{t=1}^{T}\int_{u\in\mathcal{U}_t}\{\widehat{g}_n(u) - g(u)\}^2\, p_t(u) = O_p(K/n + K^{-4})$;
(iii) *If* $\sqrt{n}/K^2 \to 0$ *as* $n \to \infty$, *then* $\eta_n(u)\{\widehat{g}_n(u) - g(u)\} \rightsquigarrow \mathcal{N}(0, 1)$ *as* $n \to \infty$.

The proofs of Theorems 1 and 2 are the same as those of Theorems 1 and 2 in [24].

Since the pilot estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{g}_n(u)$ do not take the heteroscedasticity into account, they may not be asymptotically efficient. They are, however, consistent estimators of $\boldsymbol{\beta}$ and $g(u)$, respectively, according to Theorems 1 and 2. Based on $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{g}_n(u)$, we can estimate the residuals in models (1.1)–(1.2), then use them to estimate the error variance and construct asymptotically more efficient estimators of the parametric and nonparametric components. We will investigate these in subsequent sections.

3. Consistent estimator of error variance function

After $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{g}_n(\cdot)$ are obtained, a natural estimator of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ is given by

$$\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_n)^\top = (\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_n - \widehat{\mathbf{G}}_n)$$

with $\widehat{\mathbf{G}}_n = (\widehat{g}_n(U_{11}), \ldots, \widehat{g}_n(U_{1T}), \ldots, \widehat{g}_n(U_{nT}))^\top$. Based on $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{g}_n(\cdot)$, the errors can be estimated by $\widehat{r}_{it} = Y_{it} - \widehat{\mu}_i - \mathbf{X}_{it}^\top\widehat{\boldsymbol{\beta}}_n - \widehat{g}_n(U_{it})$, $i = 1, \ldots, n$, $t = 1, \ldots, T$. Using the data set $(V_{it}, \widehat{r}_{it})$, a typical estimator of the error variance function $\sigma^2(v)$ is

$$\tilde{\sigma}_n^2(v) = \sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{nit}(v)\widehat{r}_{it}^2,$$

where

$$\omega_{nit}(\cdot) = \frac{(nh)^{-1}K\{h^{-1}(V_{it} - \cdot)\}\{A_{n,2}(\cdot) - (V_{it} - \cdot)A_{n,1}(\cdot)\}}{A_{n,0}(\cdot)A_{n,2}(\cdot) - A_{n,1}^2(\cdot)},$$

$$A_{n,j}(\cdot) = \frac{1}{nh}\sum_{i=1}^{n}\sum_{t=1}^{T}K\left(\frac{V_{it} - \cdot}{h}\right)(V_{it} - \cdot)^j, \quad j = 0, 1, 2,$$

and $K(\cdot)$ is a kernel function with bandwidth h; see [10] for more details.

In order to establish the asymptotic property of $\tilde{\sigma}_n^2(u)$, the following assumptions on $K(\cdot)$ and h are needed.

Assumption 7. The function $K(\cdot)$ is a symmetric density function with compact support.

Assumption 8. The bandwidth h satisfies $nh^8 \to 0$ and $nh^2/(\ln n)^2 \to \infty$ as $n \to \infty$.

Theorem 3. *Under Assumptions 1–8,*

$$\tilde{\sigma}_n^2(v) = \left(1 - \frac{2}{T} + \frac{1}{T^2}\right)\sigma^2(v) + \frac{T-1}{T^3}\sum_{t=1}^{T}\mathrm{E}\{\sigma^2(V_{it})\} + O_p\left(h^2 + \frac{1}{nh}\right) + o_p(h^2).$$

Theorem 3 shows that $\tilde{\sigma}_n^2(v)$ is not a consistent estimator of the error variance function $\sigma^2(v)$ for a bounded T. Hence the typical error variance estimator is not consistent in the context of fixed effects panel data partially linear models, which is caused by incidental parameters. Nevertheless, Theorem 3 suggests that, if we have a consistent estimator of $\sum_{t=1}^{T}\mathrm{E}\{\sigma^2(V_{it})\}$, then we can construct a consistent estimator of the error variance function $\sigma^2(v)$ by correcting $\tilde{\sigma}_n^2(v)$. So we now focus on constructing a consistent estimator of $\sum_{t=1}^{T}\mathrm{E}\{\sigma^2(V_{it})\}$.

Define $\widehat{\zeta}_n = \sum_{i=1}^n \sum_{t=1}^T \widehat{r}_{it}^2/(nT)$. After some algebraic calculations, we can express $\widehat{\zeta}_n$ as

$$\widehat{\zeta}_n = \frac{1}{T}\left(1 - \frac{1}{T}\right)\sum_{t=1}^T \mathrm{E}\{\sigma^2(V_{it})\} + o_p(n^{-1/2}).$$

As a result, we propose an estimator of the error variance function $\sigma^2(v)$ as follows.

$$\widehat{\sigma}_n^2(v) = \left(1 - \frac{2}{T} + \frac{1}{T^2}\right)^{-1}\left\{\tilde{\sigma}_n^2(v) - \frac{1}{T}\widehat{\zeta}_n\right\},$$

which is shown to be a consistent estimator of $\sigma^2(v)$ in the following theorem.

Theorem 4. *Under Assumptions 1–8,*

$$\widehat{\sigma}_n^2(v) = \sigma^2(v) + O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right) + o_p(h^2).$$

Remark 1. From Theorem 4, we can see that $\widehat{\sigma}_n^2(v)$ not only is consistent but also achieves the optimal convergence rate; see [10,12].

The estimator $\widehat{\sigma}_n^2(v)$ depends on the choice of bandwidth h. Since variance functions tend to have less structure than models for the mean in the regression models, we can use rather simple rule of thumb (ROT) to select h. The details can be found in [25].

4. Weighted semiparametric least squares dummy variables estimation

In this section, we construct asymptotically efficient estimators of the parametric and nonparametric components by dividing the two sides of (2.1) by $\sqrt{\widehat{\sigma}_n^2(\cdot)}$. This yields

$$\frac{Y_{it}}{\sqrt{\widehat{\sigma}_n^2(V_{it})}} \approx \frac{\mathbf{X}_{it}^\top \boldsymbol{\beta}}{\sqrt{\widehat{\sigma}_n^2(V_{it})}} + \frac{\boldsymbol{\zeta}^{K\top}(U_{it})\boldsymbol{\theta}}{\sqrt{\widehat{\sigma}_n^2(V_{it})}} + \frac{\mu_i}{\sqrt{\widehat{\sigma}_n^2(V_{it})}} + \frac{\sigma(U_{it})\nu_{it}}{\sqrt{\widehat{\sigma}_n^2(V_{it})}} \tag{4.1}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. Let

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(1/\sigma(V_{11}), \ldots, 1/\sigma(V_{1T}), \ldots, 1/\sigma(V_{nT})\right),$$

$$\widehat{\boldsymbol{\Lambda}} = \mathrm{diag}\left(1/\sqrt{\widehat{\sigma}_n^2(V_{11})}, \ldots, 1/\sqrt{\widehat{\sigma}_n^2(V_{1T})}, \ldots, 1/\sqrt{\widehat{\sigma}_n^2(V_{nT})}\right), \tag{4.2}$$

$$\boldsymbol{\nu}^{\mathrm{w}} = \widehat{\boldsymbol{\Lambda}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu}, \qquad \mathbf{X}^{\mathrm{w}} = \widehat{\boldsymbol{\Lambda}}\mathbf{X}, \qquad \boldsymbol{\zeta}^{\mathrm{w}} = \widehat{\boldsymbol{\Lambda}}\boldsymbol{\zeta}, \qquad \mathbf{Y}^{\mathrm{w}} = \widehat{\boldsymbol{\Lambda}}\mathbf{Y} \tag{4.3}$$

and

$$\mathbf{B}^{\mathrm{w}} = \begin{pmatrix} 1/\sqrt{\widehat{\sigma}_n^2(V_{11})} & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 1/\sqrt{\widehat{\sigma}_n^2(V_{1T})} & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 1/\sqrt{\widehat{\sigma}_n^2(V_{n1})} \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 1/\sqrt{\widehat{\sigma}_n^2(V_{nT})} \end{pmatrix}. \tag{4.4}$$

Then (4.1) can be re-written as

$$\mathbf{Y}^{\mathrm{w}} \approx \mathbf{X}^{\mathrm{w}}\boldsymbol{\beta} + \boldsymbol{\zeta}^{\mathrm{w}}\boldsymbol{\theta} + \mathbf{B}^{\mathrm{w}}\boldsymbol{\mu} + \boldsymbol{\nu}^{\mathrm{w}}. \tag{4.5}$$

Define $\mathbf{M}_{\mathbf{B}^{\mathrm{w}}} = \mathbf{I}_{nT} - \mathbf{P}_{\mathbf{B}^{\mathrm{w}}}$ with $\mathbf{P}_{\mathbf{B}^{\mathrm{w}}} = \mathbf{I}_{nT} - \mathbf{B}^{\mathrm{w}}(\mathbf{B}^{\mathrm{w}\top}\mathbf{B}^{\mathrm{w}})^{-1}\mathbf{B}^{\mathrm{w}\top}$. Then pre-multiplying (4.5) by $\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}$ leads to

$$\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{Y}^{\mathrm{w}} \approx \mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{X}^{\mathrm{w}}\boldsymbol{\beta} + \mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\boldsymbol{\zeta}^{\mathrm{w}}\boldsymbol{\theta} + \mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\boldsymbol{\nu}^{\mathrm{w}}. \tag{4.6}$$

If we take $\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\boldsymbol{\nu}^{\mathrm{w}}$ as the residuals, model (4.6) is a version of the usual linear regression. Thus $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ with a given K can be estimated by minimizing the following loss function

$$S^{\mathrm{w}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = (\mathbf{Y}^{\mathrm{w}} - \mathbf{X}^{\mathrm{w}}\boldsymbol{\beta} - \boldsymbol{\zeta}^{\mathrm{w}}\boldsymbol{\theta})^\top \mathbf{M}_{\mathbf{B}^{\mathrm{w}}}(\mathbf{Y}^{\mathrm{w}} - \mathbf{X}^{\mathrm{w}}\boldsymbol{\beta} - \boldsymbol{\zeta}^{\mathrm{w}}\boldsymbol{\theta}). \tag{4.7}$$

The loss function (4.7) has a unique minimizer $(\widehat{\boldsymbol{\beta}}_n^{\mathrm{w}}, \widehat{\boldsymbol{\theta}}_n^{\mathrm{w}})$ with

$$\widehat{\boldsymbol{\beta}}_n^{\mathrm{w}} = (\mathbf{X}^{\mathrm{w}\top}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\zeta^{\mathrm{w}}}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{X}^{\mathrm{w}})^{-1}\mathbf{X}^{\mathrm{w}\top}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\zeta^{\mathrm{w}}}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{Y}^{\mathrm{w}}$$

and $\widehat{\boldsymbol{\theta}}_n^{\mathrm{w}} = (\boldsymbol{\zeta}^{\mathrm{w}\top}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\boldsymbol{\zeta}^{\mathrm{w}})^{-1}\boldsymbol{\zeta}^{\mathrm{w}\top}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}(\mathbf{Y}^{\mathrm{w}} - \mathbf{X}^{\mathrm{w}}\widehat{\boldsymbol{\beta}}_n^{\mathrm{w}})$. In this paper, $\widehat{\boldsymbol{\beta}}_n^{\mathrm{w}}$ is said to be a weighted semiparametric least squares dummy variables (WSLSDV) estimator of the unknown parameter vector $\boldsymbol{\beta}$. Furthermore, a weighted spline series estimator of $g(u)$ is given by $\widehat{g}_n^{\mathrm{w}}(u) = \boldsymbol{\zeta}^{K\top}(u)\widehat{\boldsymbol{\theta}}_n^{\mathrm{w}}$.

Let $\boldsymbol{\Pi}_i^{\mathrm{w}} = (1/\sigma(U_{i1})\boldsymbol{\Pi}_{i1}, \ldots, 1/\sigma(U_{iT})\boldsymbol{\Pi}_{iT})^\top$, $\boldsymbol{\iota}_i^{\mathrm{w}} = (1/\sigma(U_{i1}), \ldots, 1/\sigma(U_{iT}))^\top$, $\mathbf{M}_{\iota_i^{\mathrm{w}}} = \mathbf{I}_T - \boldsymbol{\iota}_i^{\mathrm{w}}(\boldsymbol{\iota}_i^{\mathrm{w}\top}\boldsymbol{\iota}_i^{\mathrm{w}})^{-1}\boldsymbol{\iota}_i^{\mathrm{w}\top}$, $\boldsymbol{\Sigma}_3 = \mathrm{E}(\boldsymbol{\Pi}_i^{\mathrm{w}\top}\mathbf{M}_{\iota_i^{\mathrm{w}}}\boldsymbol{\Pi}_i^{\mathrm{w}})/T$, $\mathbf{Q}_3 = \mathrm{E}(\zeta_i^{\mathrm{w}\top}\mathbf{M}_{\iota_i^{\mathrm{w}}}\zeta_i^{\mathrm{w}})/T$, $\Im^{\mathrm{w}} = \boldsymbol{\Sigma}_3^{1/2}$ and $\eta^{\mathrm{w}}(u) = \{\boldsymbol{\zeta}^{K\top}(u)\mathbf{Q}_3\boldsymbol{\zeta}^K(u)\}^{-1/2}$.

Assumption 9. $\boldsymbol{\Sigma}_3$ is a positive definite matrix and $\boldsymbol{\zeta}^{K\top}(u)\mathbf{Q}_3\boldsymbol{\zeta}^K(u) > 0$ for all $u \in \cup_{t=1}^T \mathcal{U}_t$.

The following two theorem establishes the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n^{\mathrm{w}}$ and $\widehat{g}_n^{\mathrm{w}}(u)$.

Theorem 5. *Under Assumptions 1–9,* $\sqrt{nT}\Im^{\mathrm{w}}(\widehat{\boldsymbol{\beta}}_n^{\mathrm{w}} - \boldsymbol{\beta}) \rightsquigarrow \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$ *as* $n \to \infty$.

Theorem 6. *Under Assumptions 1–9,*

(i) $\|\widehat{\boldsymbol{\theta}}_n^{\mathrm{w}} - \boldsymbol{\theta}\| = O_p(\sqrt{K/n} + K^{-2})$;
(ii) $\sum_{t=1}^T \int_{u\in\mathcal{U}_t} \left\{\widehat{g}_n^{\mathrm{w}}(u) - g(u)\right\}^2 p_t(u) = O_p(K/n + K^{-4})$;
(iii) *If* $\sqrt{n}/K^2 \to 0$ *as* $n \to \infty$, *then* $\sqrt{nT}\eta^{\mathrm{w}}(u)\left\{\widehat{g}_n^{\mathrm{w}}(u) - g(u)\right\} \rightsquigarrow \mathcal{N}(0, 1)$ *as* $n \to \infty$.

Remark 2. We can show that the WSLSDV estimator $\widehat{\boldsymbol{\beta}}_n^{w}$ is asymptotically more efficient than $\widehat{\boldsymbol{\beta}}_n$ in the sense that $\widehat{\boldsymbol{\beta}}_n^{w}$ has a smaller asymptotic covariance matrix than $\widehat{\boldsymbol{\beta}}_n$, i.e., $\boldsymbol{\Sigma}_3^{-1} \le \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$. Let

$$\mho = G^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top\boldsymbol{\Delta}_i^{1/2} - H^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top\boldsymbol{\Delta}_i^{-1/2},$$

where $G = \mathrm{E}\left\{(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top(\boldsymbol{\Pi}_i, \mathbf{1}_T)\right\}$ and $H = \mathrm{E}\left\{(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top\boldsymbol{\Delta}_i^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)\right\}$. Then

$$\begin{aligned}\mho\mho^\top &= G^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top\boldsymbol{\Delta}_i(\boldsymbol{\Pi}_i, \mathbf{1}_T)G^{-1} - G^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top(\boldsymbol{\Pi}_i, \mathbf{1}_T)H^{-1}\\ &\quad - H^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top(\boldsymbol{\Pi}_i, \mathbf{1}_T)G^{-1} + H^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top\boldsymbol{\Delta}_i^{-1}(\boldsymbol{\Pi}_i, \mathbf{1}_T)H^{-1}.\end{aligned}$$

Since $\mho\mho^\top$ is nonnegative definite,

$$\mathrm{E}(\mho\mho^\top) = G^{-1}\mathrm{E}\left\{(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top\boldsymbol{\Delta}_i(\boldsymbol{\Pi}_i, \mathbf{1}_T)\right\}G^{-1} - H^{-1} \ge \mathbf{0}.$$

Hence

$$\boldsymbol{\Sigma}_3^{-1} = (\mathbf{I}_p, \mathbf{0}_p)H^{-1}(\mathbf{I}_p, \mathbf{0}_p)^\top \le \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1} = (\mathbf{I}_p, \mathbf{0}_p)G^{-1}\mathrm{E}\left\{(\boldsymbol{\Pi}_i, \mathbf{1}_T)^\top\boldsymbol{\Delta}_i(\boldsymbol{\Pi}_i, \mathbf{1}_T)\right\}G^{-1}(\mathbf{I}_p, \mathbf{0}_p)^\top.$$

By the same argument, we can show that $\widehat{g}_n^{\mathrm{w}}(\cdot)$ is asymptotically more efficient than $\widehat{g}_n(\cdot)$.

In order to use Theorems 5 and 6 to make statistical inference on $\boldsymbol{\beta}$ and $g(\cdot)$, we need consistent estimators of $\Im^{\mathrm{w}}$ and $\eta^{\mathrm{w}}(u)$, which are given respectively by

$$\widehat{\Im}^{\mathrm{w}} = \frac{1}{nT}\left(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\zeta^{\mathrm{w}}}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\mathbf{X}^{\mathrm{w}}\right)^{1/2} \quad \text{and} \quad \widehat{\eta}^{\mathrm{w}}(u) = \left\{\boldsymbol{\zeta}^{K\top}(u)(\boldsymbol{\zeta}^{\mathrm{w}\top}\mathbf{M}_{\mathbf{B}^{\mathrm{w}}}\boldsymbol{\zeta}^{\mathrm{w}})^{-1}\boldsymbol{\zeta}^K(u)\right\}^{1/2},$$

as shown in the following theorem.

Theorem 7. *Under Assumptions 1–9,* $\widehat{\Im}^{\mathrm{w}}(\Im^{\mathrm{w}})^{-1} \to \mathbf{I}_p$ *and* $\widehat{\eta}^{\mathrm{w}}(u)\{\eta^{\mathrm{w}}(u)\}^{-1} \to 1$ *as* $n \to \infty$, *where* $u \in \bigcup_{t=1}^T \mathcal{U}_t$.

5. Simulation studies

In this section, we conduct some simulation studies to investigate the finite sample performance of the proposed procedures in previous sections. The data are generated from the following fixed effects panel data partially linear regression model

$$Y_{it} = X_{it1}\beta_1 + X_{it2}\beta_2 + g(U_{it}) + \mu_i + \sigma(V_{it})\nu_{it}, \quad i = 1, \ldots, n,\ t = 1, \ldots, T,$$

where $X_{it1} \sim$ *i.i.d.* $\mathcal{N}(1, 1)$, $X_{it2} \sim$ *i.i.d.* $\mathcal{N}(0, 2.25)$, $\nu_{it} \sim$ *i.i.d.* $\mathcal{N}(0, 1)$, $U_{it} \sim$ *i.i.d.* $\mathcal{U}(0, 1)$, $\beta_1 = 1.5$, $\beta_2 = 2$, $\sigma(V_{it}) = \sqrt{0.2 + 1.5\{\cos(\pi V_{it})\}^2}$, $V_{it} = U_{it}$,

$$g(U_{it}) = \sqrt{U_{it}(1 - U_{it})}\sin\{2.1\pi/(U_{it} + 0.65)\} - \mathrm{E}[\sqrt{U_{it}(1 - U_{it})}\sin\{2.1\pi/(U_{it} + 0.65)\}]$$

Table 1
The finite-sample performance of the estimators of the parametric component, nonparametric component and error variance function.

		$n = 100$				$n = 200$				$n = 300$			
		$T = 3$	$T = 4$	$T = 5$	$T = 10$	$T = 3$	$T = 4$	$T = 5$	$T = 10$	$T = 3$	$T = 4$	$T = 5$	$T = 10$
$\widehat{\beta}_{n1}$	sm	1.5001	1.4965	1.5012	1.5000	1.5004	1.4970	1.4987	1.4991	1.5002	1.4998	1.5019	1.5007
	std	0.0702	0.0571	0.0490	0.0329	0.0482	0.0388	0.0348	0.0228	0.0397	0.0314	0.0267	0.0187
	mstd	0.0698	0.0558	0.0480	0.0323	0.0486	0.0395	0.0342	0.0228	0.0397	0.0322	0.0279	0.0187
	cp	0.9530	0.9420	0.9500	0.9470	0.9500	0.9520	0.9470	0.9440	0.9430	0.9580	0.9590	0.9520
$\widehat{\beta}_{n2}$	sm	2.0002	1.9991	1.9999	2.0002	2.0000	1.9991	1.9984	1.9996	2.0013	2.0007	2.0008	1.9991
	std	0.0484	0.0361	0.0324	0.0220	0.0329	0.0268	0.0238	0.0147	0.0267	0.0215	0.0184	0.0121
	se	0.0466	0.0372	0.0319	0.0215	0.0324	0.0263	0.0228	0.0152	0.0265	0.0215	0.0186	0.0125
	cp	0.9390	0.9570	0.9500	0.9350	0.9440	0.9470	0.9470	0.9500	0.9480	0.9520	0.9580	0.9580
$\widehat{\beta}^{\mathrm{w}}_{n1}$	sm	1.5011	1.4975	1.5021	1.5000	1.4998	1.4980	1.4992	1.4993	1.4995	1.5003	1.5011	1.5001
	std	0.0592	0.0483	0.0401	0.0267	0.0414	0.0326	0.0297	0.0183	0.0349	0.0275	0.0230	0.0152
	mstd	0.0597	0.0469	0.0397	0.0263	0.0418	0.0335	0.0286	0.0184	0.0344	0.0272	0.0232	0.0151
	cp	0.9570	0.9500	0.9430	0.9420	0.9530	0.9530	0.9440	0.9500	0.9450	0.9470	0.9570	0.9520
$\widehat{\beta}^{\mathrm{w}}_{n2}$	sm	1.9997	1.9993	1.9996	2.0002	2.0002	1.9994	1.9993	1.9997	2.0010	2.0007	2.0007	1.9989
	std	0.0420	0.0303	0.0263	0.0178	0.0278	0.0229	0.0194	0.0123	0.0228	0.0184	0.0153	0.0098
	mstd	0.0399	0.0313	0.0265	0.0175	0.0278	0.0223	0.0191	0.0123	0.0229	0.0182	0.0155	0.0100
	cp	0.9270	0.9550	0.9540	0.9510	0.9490	0.9510	0.9460	0.9490	0.9480	0.9500	0.9530	0.9600
$\widehat{g}_n(\cdot)$	sm	0.1803	0.1490	0.1322	0.0921	0.1319	0.1103	0.0970	0.0695	0.1104	0.0917	0.0826	0.0609
	std	0.0508	0.0403	0.0348	0.0238	0.0355	0.0282	0.0250	0.0164	0.0293	0.0227	0.0204	0.0134
$\widehat{g}^{\mathrm{w}}_n(\cdot)$	sm	0.1739	0.1433	0.1273	0.0893	0.1269	0.1062	0.0934	0.0679	0.1063	0.0884	0.0802	0.0599
	std	0.0487	0.0392	0.0344	0.0228	0.0342	0.0274	0.0239	0.0160	0.0285	0.0217	0.0195	0.0131
$\tilde{\sigma}^2_n(\cdot)$	sm	0.4573	0.3641	0.3093	0.1886	0.4476	0.3501	0.2955	0.1703	0.4444	0.3482	0.2866	0.1624
	std	0.0797	0.0707	0.0630	0.0458	0.0570	0.0524	0.0514	0.0360	0.0458	0.0466	0.0409	0.0306
$\widehat{\sigma}^2_n(\cdot)$	sm	0.3174	0.2542	0.2236	0.1510	0.2348	0.1912	0.1646	0.1114	0.1978	0.1576	0.1368	0.0958
	std	0.1070	0.0807	0.0763	0.0448	0.0781	0.0584	0.0496	0.0315	0.0570	0.0445	0.0377	0.0262

and

$$\mu_i = \sum_{t=1}^{T}(X_{it1} + X_{it2})/T - \sum_{t=1}^{T}\mathrm{E}(X_{it1} + X_{it2})/T,$$

which are correlated with (X_{it1}, X_{it2}). The sample sizes are taken as $n = 100$, 200 and 300, and $T = 3$, 4, 5 and 10, and the replication number for the simulation is 1000. For the spline basis functions, we take the uniform knots and five knots.

For the WSLSDV estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{w}}_n = (\widehat{\beta}^{\mathrm{w}}_{n1}, \widehat{\beta}^{\mathrm{w}}_{n2})^\top$ of the parametric components $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, given a sample size, the sample mean (sm), standard deviation (std), mean of the estimate of the standard deviation (mstd) and coverage percentage (cp) of the 95% nominal confidence intervals are summarized in Table 1. In this table, we also present the sm, std, mstd and cp of the unweighted semiparametric least squares dummy variables estimator $\widehat{\boldsymbol{\beta}}_n = (\widehat{\beta}_{n1}, \widehat{\beta}_{n2})^\top$, which ignores the heteroscedasticity.

Table 1 provides the following observations:

- The proposed WSLSDV estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{w}}_n$ and the unweighted SLSDV estimator $\widehat{\boldsymbol{\beta}}_n$ are asymptotically unbiased, but $\widehat{\boldsymbol{\beta}}^{\mathrm{w}}_n$ has smaller standard deviation than $\widehat{\boldsymbol{\beta}}_n$.
- The standard deviations of the proposed WSLSDV estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{w}}_n$ and the unweighted SLSDV estimator $\widehat{\boldsymbol{\beta}}_n$ decrease as the sample size n increases.
- When $n \times T$ is fixed (for example, the cases $(n, T) = (100, 10)$ and $(n, T) = (200, 5)$), the standard deviations of the proposed WSLSDV estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{w}}_n$ and the unweighted SLSDV estimator $\widehat{\boldsymbol{\beta}}_n$ decrease as T increases.
- The means of the standard error estimates closely agree with the simulation standard errors for the proposed WSLSDV estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{w}}_n$.
- The estimated confidence interval attains a coverage close to the nominal 95% level.

For the nonparametric component, its estimators are assessed via the Square-Root of Averaged Squared Errors (RASE):

$$\mathrm{RASE}(\widehat{g}_n) = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\widehat{g}_n(U_{it}) - g(U_{it})\}^2\right]^{1/2}.$$

For a given sample size, the sm and std of the RASEs of $\widehat{g}_n(\cdot)$ and $\widehat{g}^{\mathrm{w}}_n(\cdot)$ are calculated. The results are shown in Table 1 as well, which show that the RASEs of both $\widehat{g}_n(\cdot)$ and $\widehat{g}^{\mathrm{w}}_n(\cdot)$ decrease as the sample size n increases, and when nT is fixed, the RASEs of $\widehat{g}_n(\cdot)$ and $\widehat{g}^{\mathrm{w}}_n(\cdot)$ decrease as T increases. More importantly, when n and T are fixed, $\widehat{g}^{\mathrm{w}}_n(\cdot)$ has smaller RASE than $\widehat{g}_n(\cdot)$.

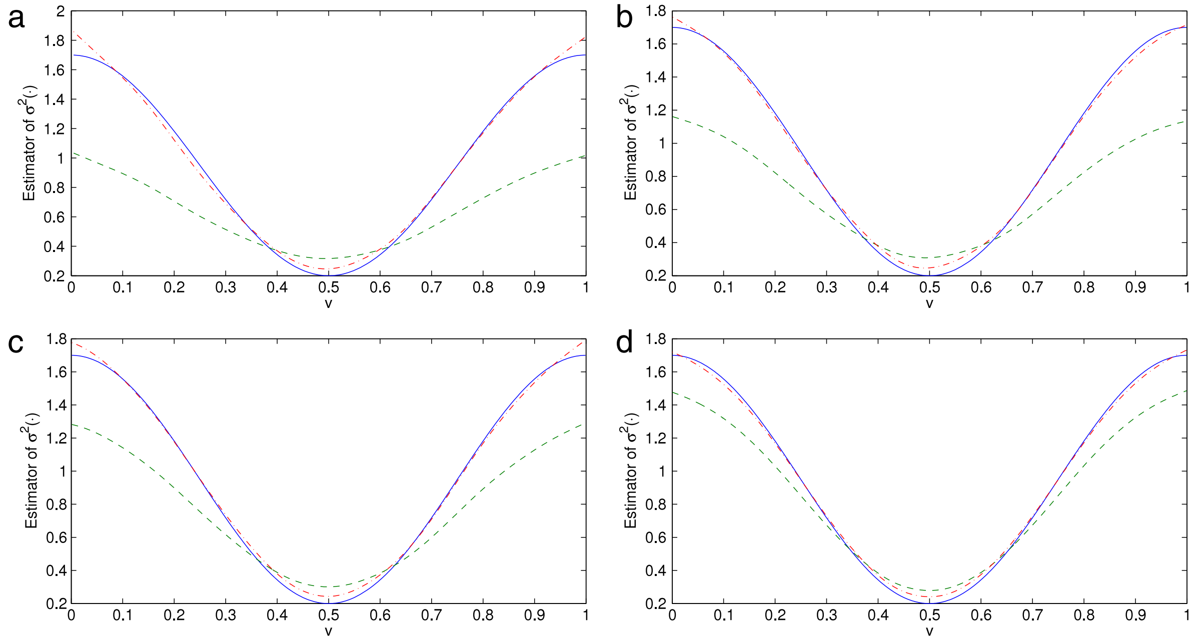

Fig. 1. (a) $n = 100$ and $T = 3$; (b) $n = 100$ and $T = 4$; (c) $n = 100$ and $T = 5$; (d) $n = 100$ and $T = 10$. $\sigma^2(\cdot)$ (solid curve), $\tilde{\sigma}_n^2(\cdot)$ (dashed curve) and $\widehat{\sigma}_n^2(\cdot)$ (dash-dotted curve).

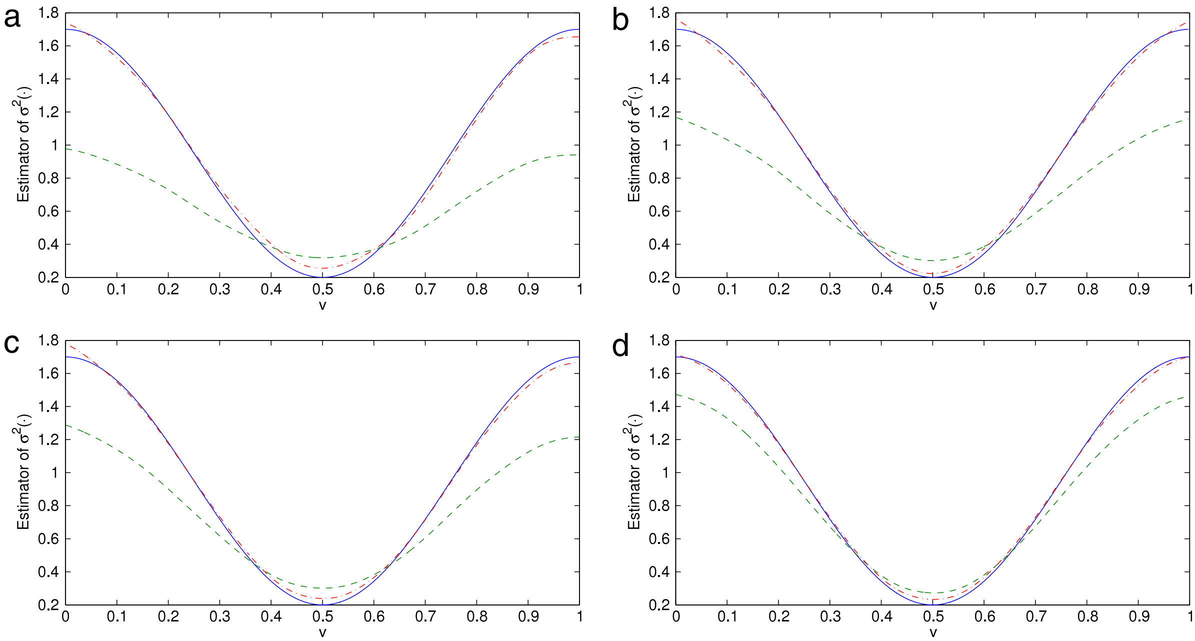

Fig. 2. (a) $n = 200$ and $T = 3$; (b) $n = 200$ and $T = 4$; (c) $n = 200$ and $T = 5$; (d) $n = 200$ and $T = 10$. $\sigma^2(\cdot)$ (solid curve), $\tilde{\sigma}_n^2(\cdot)$ (dashed curve) and $\widehat{\sigma}_n^2(\cdot)$ (dash-dotted curve).

For the error variance function, its estimators are also assessed via the Square-Root of Averaged Squared Errors (RASE). For a given sample size, the sm and std of the RASEs of $\tilde{\sigma}_n^2(\cdot)$ and $\widehat{\sigma}_n^2(\cdot)$ are calculated. The results are also shown in Table 1. The RASE of $\tilde{\sigma}_n^2(\cdot)$ decreases significantly with increasing T, but not obvious with increasing n. The RASE of $\widehat{\sigma}_n^2(\cdot)$, on the other hand, decreases with either increasing n or increasing T. Figs. 1–3 also illustrate these phenomena, which are consistent with Theorems 3 and 4. It is apparent from Figs. 1–3 that the proposed variance estimator $\widehat{\sigma}_n^2(\cdot)$ is much closer to the true variance function $\sigma^2(\cdot)$ than the typical residuals-based estimator $\tilde{\sigma}_n^2(\cdot)$ that ignores the heteroscedasticity.

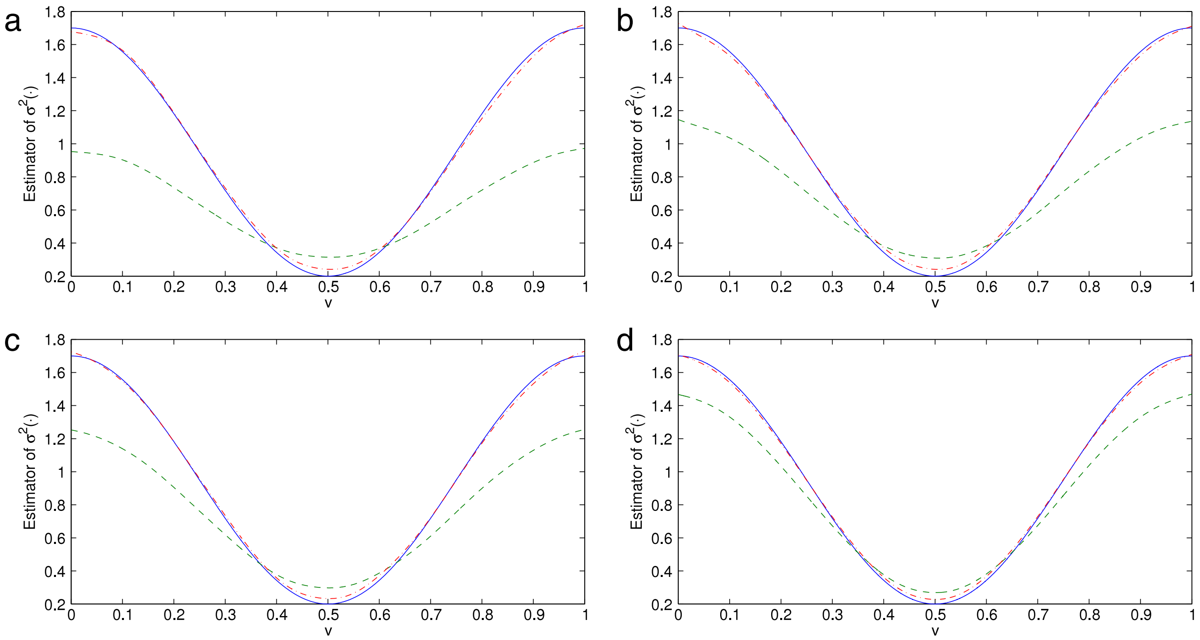

Fig. 3. (a) $n = 300$ and $T = 3$; (b) $n = 300$ and $T = 4$; (c) $n = 300$ and $T = 5$; (d) $n = 300$ and $T = 10$. $\sigma^2(\cdot)$ (solid curve), $\tilde{\sigma}_n^2(\cdot)$ (dashed curve) and $\widehat{\sigma}_n^2(\cdot)$ (dash-dotted curve).

6. An application to economical data

We now demonstrate the application of the proposed estimation procedures to analyze a set of economical data. The data were extracted from the STARS database of the World Bank. From this database we obtained measures of Gross Domestic Product (GDP) and the aggregate physical capital stock, both were denominated in constant local currency units at the end of period 1987 (converted into US dollars at the end of period 1987) for 81 countries over the period from 1960 to 1987. We excluded one country whose workforce has around 15 years of schooling, which is much higher than others. The database also provided the number of individuals in the workforce between ages 15 and 64, and the mean years of schooling for the workforce, e.g., Duffy and Papageorgiou [7]. Fig. 2(a)–(c) present the scatter plots of the logarithm of the real GDP against the logarithms of real capital, labor supply and mean years of schooling for the workforce. This data set has also been analyzed by [24], but without heteroscedasticity.

According to [24], on the log scale, a linear relationship is reasonable between GDP and capital or labor supply. However, the relationship between GDP and mean years of schooling for the workforce is clearly nonlinear. Therefore, we use the following fixed effects panel data partially linear regression model to fit this data set by taking the heteroscedasticity into account.

$$Y_{it} = X_{it1}\beta_1 + X_{it2}\beta_2 + g(U_{it}) + \mu_i + \sigma(V_{it})\nu_{it}, \quad i = 1, \dots, 81, \; t = 1, \dots, 28,$$

where Y_{it} is the log real GDP of country i in year t (with $t = 1$ for year 1960, and so on), μ_i is individual affect of each country, X_{it1} is the log of real capital, X_{it2} is the log of labor supply, U_{it} is the log of mean years of schooling for the workforce and $V_{it} = U_{it}$. The unweighted SLSDV estimate of $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ is $\widehat{\boldsymbol{\beta}}_n = (\widehat{\beta}_{n1}, \widehat{\beta}_{n2})^\top = (0.5495, 0.2670)$ with standard deviation $(0.0105, 0.0347)$ and the weighted SLSDV estimate is $\widehat{\boldsymbol{\beta}}_n^{\mathrm{w}} = (\widehat{\beta}_{n1}^{\mathrm{w}}, \widehat{\beta}_{n2}^{\mathrm{w}}) = (0.5685, 0.3195)$ with standard deviation $(0.0077, 0.0282)$. Both estimates imply significant and positive effects of capital and labor supply on the GDP (at 5% level). As a result, the real GDP is strongly and positively correlated with both real capital and labor supply. In addition, the weighted SLSDV has smaller standard deviation than the unweighted SLSDV.

Furthermore, the fitted nonparametric component curve $g(\cdot)$ and error variance $\sigma^2(\cdot)$ are plotted in Fig. 4(b) and (c). Fig. 4(b) shows that the GDP changes little for log mean years of schooling between −2.5 and 1 (corresponding approximately to mean years of schooling between 0 and 3 years). When the log mean years of schooling is over 1 (about 3 years), however, the GDP increases with the mean years of schooling rapidly at a nonlinear and accelerated pace. All these results are generally consistent with those obtained by previous studies. Fig. 4(c) shows that error variance $\sigma(\cdot)$ is a nonlinear function of U_{it}. In addition, the individual effect (μ_i) of each country is provided in Fig. 4(d) as well.

7. Concluding remarks

In this paper, we have studied the statistical inference of the fixed effects panel data partially linear regression model with heteroscedastic errors whose variance is a smooth function of some covariate(s). We found that the typical

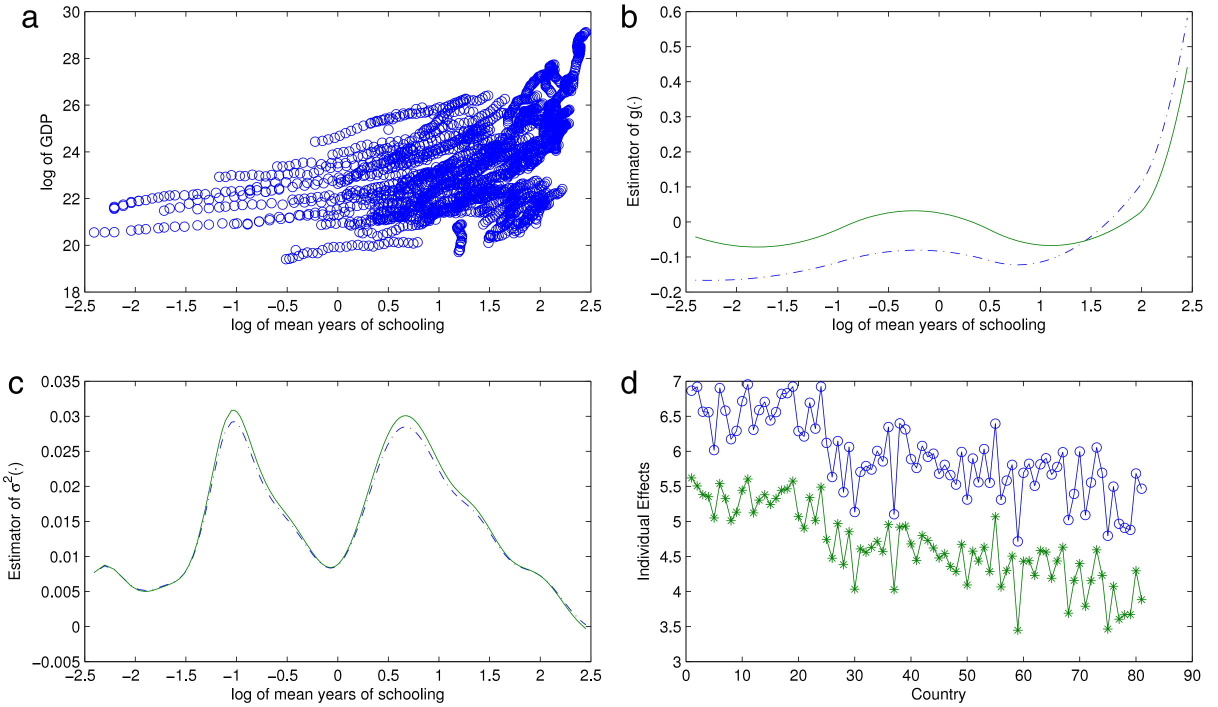

Fig. 4. (a) The scatter plot of log real GDP against log mean years of schooling for the workforce; (b) The estimators of $g(\cdot)$. Solid curve: the weighted series/sieve estimator and dash-dotted curve: the unweighted series/sieve estimator; (c) The estimators of $\sigma^2(\cdot)$. Solid curve: $\widehat{\sigma}_n^2(\cdot)$ and dash-dotted curve: $\tilde{\sigma}_n^2(\cdot)$; (d) Individual effects. Solid-cycle curve: based on weighted fitting and solid-star curve: based on unweighted fitting.

residuals-based local polynomial estimator of the error variance function is not consistent, apparently due to the incidental parameter problem. We then proposed a consistent alternative estimator of the error variance function. Applying the proposed estimator and series/sieve approximation of the nonparametric component, we further constructed a weighted semiparametric least squares dummy variables estimator for the parametric and nonparametric components. Asymptotic normal distributions for the proposed estimators were derived and asymptotic covariance matrix estimators were provided as well. Both estimators are shown to be asymptotically more efficient than the those ignoring the heteroscedasticity.

It is not difficult to extend our results to the scenario of $q > 1$ by tensor product spline series technique and extend our results to the scenario of $m > 1$ by local linear multivariate regression technique.

Acknowledgments

The authors are grateful to the Editor-in-Chief, an Associate Editor and anonymous referees for their constructive comments which led to the present greatly improved version. Hu's research was supported by National Natural Science Foundation of China (NSFC) Grant 11571219. You's research was supported by NSFC Grant 11071154. The work is also partially supported by Program IRT13077 for Changjiang Scholars and Innovative Research Team in University and by Program 2016110391 for Innovative Research Team of Shanghai University of Finance and Economics.

Appendix. Proofs of main results

In order to prove other results we first present the following three lemmas.

Lemma A.1. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d random vectors, where the Y_i's are scalar random variables. Further assume that* $\mathrm{E}|Y_i|^4 < \infty$ *and* $\sup_x \int |y|^4 f(x, y)dy < \infty$, *where $f(\cdot, \cdot)$ denotes the joint density of (X, Y). Let $K(\cdot)$ be a bounded positive function with a bounded support, and satisfies Lipschitz's condition. Then if $nh^8 \to 0$ and $nh^2/(\ln n)^2 \to \infty$, it holds that*

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^{n} [K_h(X_i - x)Y_i - \mathrm{E}\{K_h(X_i - x)Y_i\}] \right| = O_p\left(\left\{ \frac{\ln(1/h)}{nh} \right\}^{1/2} \right).$$

The proof of Lemma A.1 follows immediately from the result of Mack and Silverman [17].

Define $\mathbf{B}^*$, $\boldsymbol{\Lambda}$, $\mathbf{X}^*$, $\boldsymbol{\zeta}^*$ and $\mathbf{Y}^*$ the same way as $\mathbf{B}^{\mathrm{w}}$, $\widehat{\boldsymbol{\Lambda}}$, $\mathbf{X}^{\mathrm{w}}$, $\boldsymbol{\zeta}^{\mathrm{w}}$ and $\mathbf{Y}^{\mathrm{w}}$ in (4.2)–(4.4), except with $\sigma^2(\cdot)$ in place of $\widehat{\sigma}^2(\cdot)$. Further define

$$\begin{aligned}
\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*} &= \mathbf{I}_{nT} - \mathbf{P}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*} = \mathbf{I}_{nT} - \mathbf{M}_{\mathbf{B}^*}\boldsymbol{\zeta}^*(\boldsymbol{\zeta}^{*\top}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\zeta}^*)^{-}\boldsymbol{\zeta}^{*\top}\mathbf{M}_{\mathbf{B}^*},\\
\breve{\boldsymbol{\beta}}_n^w &= (\mathbf{X}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\mathbf{Y}^*, \quad \text{and}\\
\breve{\boldsymbol{\theta}}_n^w &= (\boldsymbol{\zeta}^{*\top}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\zeta}^*)^{-1}\boldsymbol{\zeta}^{*\top}\mathbf{M}_{\mathbf{B}^*}(\mathbf{Y}^* - \mathbf{X}^*\breve{\boldsymbol{\beta}}_n^w) \quad \text{and} \quad \breve{g}_n^w(u) = \boldsymbol{\zeta}^{K\top}(u)\breve{\boldsymbol{\theta}}^w.
\end{aligned}$$

Lemma A.2. *Under Assumptions 1–9,*

(i) $\left\{(nT)^{-1}\mathbf{X}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\mathbf{X}^*\right\}\boldsymbol{\Sigma}_3^{-1} \to_p \mathbf{I}_p$ *as* $n \to \infty$;
(ii) *For any fixed K-vector* λ, $\left\{(nT)^{-1}\lambda^\top\boldsymbol{\zeta}^{*\top}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\zeta}^*\lambda\right\}(\lambda^\top\mathbf{Q}_3\lambda)^{-1} \to_p 1$ *as* $n \to \infty$, *where* $\boldsymbol{\Sigma}_3$ *and* $\mathbf{Q}_3$ *are defined in Section 4.*

Proof. Let $\boldsymbol{\Pi}^* = \boldsymbol{\Lambda}\boldsymbol{\Pi}$ and $\mathbf{H} = (\mathbf{H}_1, \ldots, \mathbf{H}_p)$ with $\mathbf{H}_j = (\hbar_j(U_{11}), \ldots, \hbar_j(U_{1T}), \ldots, \hbar_j(U_{nT}))^\top$. Then $\mathbf{X}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\mathbf{X}^*$ can be written as

$$\begin{aligned}
\mathbf{X}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\mathbf{X}^* = {}& \boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Pi}^* + \mathbf{H}^\top\boldsymbol{\Lambda}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Lambda}\mathbf{H}\\
&+ \boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Lambda}\mathbf{H} + \mathbf{H}^\top\boldsymbol{\Lambda}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Pi}^*.
\end{aligned}$$

Due to $\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*} = \mathbf{M}_{\mathbf{B}^*} - (\mathbf{I}_n - \mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*})$, we have

$$\frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Pi}^* = \frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Pi}^* - \frac{1}{nT}\boldsymbol{\Pi}^{*\top}(\mathbf{I}_n - \mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*})\boldsymbol{\Pi}^*.$$

The (ℓ, k)th entry of $(nT)^{-1}\boldsymbol{\Pi}^{*\top}(\mathbf{I}_n - \mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*})\boldsymbol{\Pi}^*$ is

$$\begin{aligned}
&\frac{1}{nT}\mathrm{E}\left\{\boldsymbol{\Pi}_{(k)}^{*\top}(\mathbf{I}_n - \mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*})\boldsymbol{\Pi}_{(k)}^*\right\}\\
&\quad = \frac{1}{n}\mathrm{E}\left[\mathrm{tr}\left\{\mathrm{E}(\boldsymbol{\Pi}_{(\ell)}\boldsymbol{\Pi}_{(k)}^T|U_{11}, \ldots, U_{1T}, \ldots, U_{nT})\boldsymbol{\Lambda}(\mathbf{I}_n - \mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*})\boldsymbol{\Lambda}\right\}\right] = O(n^{-1}K) = o(1),
\end{aligned}$$

where $\boldsymbol{\Pi}_{(\ell)}$ is the ℓth column of $\boldsymbol{\Pi}$. Moreover, by Assumption 1 we can show that

$$\frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Pi}^* \to_p \boldsymbol{\Sigma}_3 \quad \text{as } n \to \infty.$$

Given that $\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*} \le \mathbf{M}_{\mathbf{B}^*}$ and $\mathbf{M}_{\mathbf{B}^*}$ is an idempotent matrix, the maximum eigenvalue of $\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}$ is 1. Thus

$$\begin{aligned}
\frac{1}{nT}\mathbf{H}_{(\ell)}^\top\boldsymbol{\Lambda}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Lambda}\mathbf{H}_{(k)} &= \frac{1}{nT}(\mathbf{H}_{(\ell)} - \boldsymbol{\zeta}\boldsymbol{\xi}_{\mathbf{h}_\ell})^\top\boldsymbol{\Lambda}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Lambda}(\mathbf{H}_{(k)} - \boldsymbol{\zeta}\boldsymbol{\xi}_{\mathbf{h}_k})\\
&\le \frac{1}{nT}(\mathbf{H}_{(\ell)} - \boldsymbol{\zeta}\boldsymbol{\xi}_{\mathbf{h}_\ell})^\top(\mathbf{H}_{(k)} - \boldsymbol{\zeta}\boldsymbol{\xi}_{\mathbf{h}_k}) = O_p(K^{-4}) = o_p(1),
\end{aligned}$$

where $\mathbf{H}_{(\ell)}$ is the ℓth column of $\mathbf{H}$. It follows from Markov's inequality that

$$\frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Lambda}\mathbf{H} = O_p(K^{-2}) \quad \text{and} \quad \frac{1}{nT}\mathbf{H}^\top\boldsymbol{\Lambda}\mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\Pi}^* = O_p(K^{-2}).$$

This proves (i). Following the same line we can prove (ii) as well. □

Lemma A.3. *Under Assumptions 1–9,*

(i) $\sqrt{nT}\,\Im^{\mathrm{w}}(\breve{\boldsymbol{\beta}}_n^{\mathrm{w}} - \boldsymbol{\beta}) \rightsquigarrow \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$ *as* $n \to \infty$;
(ii) $\|\breve{\boldsymbol{\theta}}_n^{\mathrm{w}} - \boldsymbol{\theta}\| = O_p(\sqrt{K/n} + K^{-2})$;
(iii) $\sum_{t=1}^T \int_{u\in\mathcal{U}_t}\{\breve{g}_n^{\mathrm{w}}(u) - g(u)\}p_t(u) = O_p(K/n + K^{-4})$;
(iv) *if we further have* $\sqrt{n}/K^2 \to 0$ *as* $n \to \infty$, *then* $\sqrt{nT}\,\eta^{\mathrm{w}}(u)\{\breve{g}_n^{\mathrm{w}}(u) - g(u)\} \rightsquigarrow \mathcal{N}(0, 1)$ *as* $n \to \infty$.

Proof. (i) According to the definition of $\breve{\boldsymbol{\beta}}_n^{\mathrm{w}}$,

$$\begin{aligned}
\breve{\boldsymbol{\beta}}_n^{\mathrm{w}} &= (\mathbf{X}^{*\top}\mathbf{M}^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\mathbf{M}^*(\mathbf{X}^*\boldsymbol{\beta} + \mathbf{B}^* + \boldsymbol{\varepsilon}^* + \boldsymbol{\Lambda}\mathbf{G})\\
&= \boldsymbol{\beta} + (\mathbf{X}^{*\top}\mathbf{M}^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\mathbf{M}^*\boldsymbol{\varepsilon}^* + (\mathbf{X}^{*\top}\mathbf{M}^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\mathbf{M}^*\boldsymbol{\Lambda}(\mathbf{G} - \boldsymbol{\zeta}\boldsymbol{\theta})\\
&= \boldsymbol{\beta} + J_1 + J_2, \quad \text{say},
\end{aligned}$$

where $\mathbf{M}^* = \mathbf{M}_{\mathbf{B}^*}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}$. It is easy to see that $\mathbf{X}^{*\top}\mathbf{M}^*\boldsymbol{\varepsilon}^* = \boldsymbol{\Pi}^{*\top}\mathbf{M}^*\boldsymbol{\varepsilon}^* + \mathbf{H}\boldsymbol{\Lambda}\mathbf{M}^*\boldsymbol{\varepsilon}^*$. Note that $\mathrm{E}(\boldsymbol{\Pi}) = \mathbf{0}$. Similar to the proof in Lemma A.1, we can show that

$$\frac{1}{nT}\mathbf{H}\boldsymbol{\Lambda}\mathbf{M}^*\boldsymbol{\varepsilon}^* = O_p(n^{-1/2}K^{-2}).$$

Moreover, we have

$$\begin{aligned}\frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}^*\boldsymbol{\varepsilon}^* &= \frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\varepsilon}^* - \frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\mathbf{P}_{\mathbf{M}_{\mathbf{B}^*}\zeta^*}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\varepsilon}^* \\ &= \frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}_{\mathbf{B}^*}\boldsymbol{\varepsilon}^* + o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Pi}_i^{\mathrm{w}\top}\mathbf{M}_{\iota_i^{\mathrm{w}}}\boldsymbol{\Pi}_i^{\mathrm{w}} + o_p(n^{-1/2}).\end{aligned}$$

Let λ be a constant p-vector, and write

$$\ell_n(\lambda) = \lambda^\top\sum_{i=1}^{n}\boldsymbol{\Pi}_i^{\mathrm{w}\top}\mathbf{M}_{\iota_i^{\mathrm{w}}}\boldsymbol{\Pi}_i^{\mathrm{w}}.$$

Obviously, $\ell_n(\lambda)$ is a sum of independent random variables with variance $n\lambda^\top\mathrm{E}(\boldsymbol{\Pi}_i^{\mathrm{w}\top}\mathbf{M}_{\iota_i^{\mathrm{w}}}\boldsymbol{\Pi}_i^{\mathrm{w}})\lambda = n\lambda^T\boldsymbol{\Sigma}_3\lambda$. Therefore, by Lemma A.1, it is easy to show that $\sqrt{nT}J_2 \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Sigma}_3^{-1})$ as $n\to\infty$. In addition,

$$\begin{aligned}\frac{1}{nT}\mathbf{X}^{*\top}\mathbf{M}^*\boldsymbol{\Lambda}(\mathbf{G}-\boldsymbol{\zeta\theta}) &= \frac{1}{nT}\boldsymbol{\Pi}^{*\top}\mathbf{M}^*\boldsymbol{\Lambda}(\mathbf{G}-\boldsymbol{\zeta\theta}) + \frac{1}{nT}(\mathbf{H}-\boldsymbol{\zeta\theta}_{\mathbf{H}})^\top\boldsymbol{\Lambda}\mathbf{M}^*\boldsymbol{\Lambda}(\mathbf{G}-\boldsymbol{\zeta\theta}) \\ &= O_p(n^{-1/2}K^{-2}) + O_p(K^{-4}) = o_p(n^{-1/2}).\end{aligned}$$

This completes the proof of (i). Parts (ii)–(iv) are standard spline series approximating results. □

Proof of Theorem 3. By the definition of $\widehat{r}_{it}$, for $i = 1,\dots,n,\ t = 1,\dots,T$,

$$\widehat{r}_{it} = A(V_{it},\varepsilon_{it}) + \widetilde{X}_{it}^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n) + G(U_{it}),$$

where $A(V_{it},\varepsilon_{it}) = \sigma(V_{it})\varepsilon_{it} - T^{-1}\sum_{t_1}\sigma(V_{it})\varepsilon_{it}$, $\widetilde{\mathbf{X}}_{it} = \mathbf{X}_{it} - T^{-1}\sum_{t_1}\mathbf{X}_{it_1}$ and $G(U_{it}) = g(U_{it}) - \widehat{g}_n(U_{it}) - T^{-1}\sum_{t_1}\{g(U_{it_1}) - \widehat{g}_n(U_{it_1})\}$. Therefore, $\widehat{\sigma}_n^2(v)$ can be decomposed as

$$\begin{aligned}\widehat{\sigma}_n^2(v) &= \sum_{i,t}\omega_{nit}(v)(v)A^2(V_{it},\nu_{it}) + \sum_{i,t}\omega_{nit}(v)\{\widetilde{X}_{it}^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)\}^2 + \sum_{i,t}\omega_{nit}(v)G^2(U_{it}) \\ &\quad + 2\sum_{i,t}\omega_{nit}(v)\left\{\sigma(V_{it})\varepsilon_{it} - \frac{1}{T}\sum_{t_1}\sigma(V_{it_1})\nu_{it_1}\right\}\{\widetilde{X}_{it}^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)\} \\ &\quad + 2\sum_{i,t}\omega_{nit}(v)\{\widetilde{X}_{it}^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)\}G(U_{it}) + 2\sum_{i,t}\omega_{nit}(v)A(\sigma(V_{it}),\nu_{it})G(U_{it}) \\ &= J_1+J_2+J_3+J_4+J_5+J_6, \quad \text{say},\end{aligned}$$

where $\sum_{i,t}$, $\sum_i$ and $\sum_t$ denote $\sum_{i=1}^n\sum_{t=1}^T$, $\sum_{i=1}^n$ and $\sum_{t=1}^T$, respectively. J_1 can be further decomposed as

$$\begin{aligned}J_1 &= \sum_{i,t}\omega_{nit}(v)\sigma^2(V_{it})\nu_{it}^2 + \frac{1}{T^2}\sum_{i,t}\omega_{nit}(v)\left\{\sum_{t_1}\sigma(V_{it_1})\nu_{it_1}\right\}^2 - \frac{2}{T}\sum_{i,t}\psi_{nit}\sum_{t_1}\sigma(V_{it_1})\nu_{it_1} \\ &= J_{11}+J_{12}+J_{13}, \quad \text{say},\end{aligned}$$

where $\psi_{nit} = \omega_{nit}(v)\sigma(V_{it})\nu_{it}$. Following Chiou and Müller [6], we have

$$\omega_{nit}(v) = \frac{1}{nh}\sum_t K\left(\frac{V_{it}-v}{h}\right)\frac{h^2\varsigma_2^2 p_t(v) + O(h^3)}{h^2\varsigma_2 p_t^2(v) + O(h^4) + O(h/n)}, \tag{A.1}$$

where $\varsigma_2 = \int v^2K(v)dv < \infty$. Hence similar to [10], we can show that

$$\begin{aligned}J_{11} &= \sum_{i,t}\omega_{nit}(v)\sigma^2(V_{it})(\nu_{it}^2-1) + \sum_{i,t}\omega_{nit}(v)\left\{\sigma^2(V_{it}) - \sigma^2(v)\right\} + \sum_{i,t}\omega_{nit}(v)\sigma^2(v) \\ &= \sigma^2(v) + O_p(h^2) + O_p(n^{-1/2}h^{-1/2}).\end{aligned}$$

The same argument also leads to

$$J_{13} = -\frac{2}{T}\sum_{i,t}\omega_{nit}(v)\sigma^2(V_{it})\nu_{it}^2 - \frac{2}{T}\sum_{i,t}\omega_{nit}(v)\sigma(V_{it})\nu_{it}\sum_{t_1\neq t}\sigma(V_{it_1})\nu_{it_1}$$
$$= -\frac{2}{T}\sigma^2(v) + O_p(h^2) + O_p(n^{-1/2}h^{-1/2})$$

and

$$J_{12} = \frac{1}{T^2}\sum_{i,t}\omega_{nit}(v)\sum_{t_1}\sigma^2(V_{it_1})\nu_{it_1}^2 + \frac{1}{T^2}\sum_{i,t}\omega_{nit}(v)\sum_{t_1}\sum_{t_2\neq t_1}\sigma(V_{it_1})\nu_{it_1}\sigma(V_{it_2})\nu_{it_2}$$
$$= \frac{1}{T^2}\sum_{i,t}\omega_{nit}(v)\left\{\sigma^2(V_{it})\nu_{it}^2 + \sum_{t_1\neq t}\sigma^2(V_{it_1})\nu_{it_1}^2\right\} + O_p(h^2) + O_p(n^{-1/2}h^{-1/2})$$
$$= \frac{1}{T^2}\sigma^2(v) + \frac{1}{T^2}\sum_{i,t}\omega_{nit}(v)\sum_{t_1\neq t}\sigma^2(V_{it_1})\nu_{it_1}^2 + O_p(h^2) + O_p(n^{-1/2}h^{-1/2})$$
$$= \frac{1}{T^2}\sigma^2(u) + J_{121} + O_p(h^2) + O_p(n^{-1/2}h^{-1/2}),$$

where $J_{121} = T^{-2}\sum_{i,t}\omega_{nit}(v)\sum_{t_1\neq t}\sigma^2(V_{it_1})\nu_{it_1}^2$. By Eq. (A.1), J_{121} can be further written as

$$J_{121} = \frac{1}{T^2}\sum_t \frac{h^2\varsigma_2^2 p_t(v) + O(h^3)}{h^2\varsigma_2 p^2(v)_t + O(h^4) + O(h/n)}\sum_i \frac{1}{nh}K\left(\frac{V_{it}-v}{h}\right)\sum_{t_1\neq t}\sigma^2(V_{it_1})\nu_{it_1}^2$$
$$= \frac{(T-1)}{T^3}\sum_t \mathrm{E}\{\sigma^2(V_{it})\}\frac{h^2\varsigma_2^2 p_t(v) + O(h^3)}{h^2\varsigma_2 p_t^2(v) + O(h^4) + O(h/n)}\sum_{i,t}\frac{1}{nh}K\left(\frac{V_{it}-v}{h}\right)$$
$$+ \frac{1}{T^2}\sum_t \frac{h^2\varsigma_2^2 p_t(v) + O(h^3)}{h^2\varsigma_2 p_t^2(v) + O(h^4) + O(h/n)}\sum_i \frac{1}{nh}K\left(\frac{V_{it}-v}{h}\right)\sum_{t_1\neq t}B(V_{it_1}, \nu_{it_1})$$
$$= \frac{(T-1)\sum_t \mathrm{E}\{\sigma^2(V_{it})\}}{T^3} + O_p(h^2) + O_p(n^{-1/2}h^{-1/2}),$$

where $B(V_{it_1}, \nu_{it_1}) = \left[\sigma^2(V_{it_1})\nu_{it_1}^2 - \mathrm{E}\{\sigma^2(V_{it_1})\}\right]$. It follows that

$$J_1 = \left(1 - \frac{2}{T} + \frac{1}{T^2}\right)\sigma^2(v) + \frac{T-1}{T^2}\sum_t \mathrm{E}\{\sigma^2(V_{it})\} + O_p(h^2) + O_p(n^{-1/2}h^{-1/2}).$$

Therefore, in order to complete the proof, it suffices to show that

$$J_s = O_p(h^2) + O_p(n^{-1/2}h^{-1/2}), \quad s = 2, \ldots, 6. \tag{A.2}$$

Combining Theorems 1 and 2 with Lemma A.1 and Eq. (A.1) yields $J_2 = O_p(1/n) = O_p(h^2) + O_p\{(nh)^{-1/2}\}$, $J_3 = O_p(K/n) + O_p(K^{-4}) = O_p(h^2) + O_p\{(nh)^{-1/2}\}$, $J_4 = O_p(n^{-1/2}) = O_p(h^2) + O_p\{(nh)^{-1/2}\}$ and $J_5 = O_p(n^{-1/2})\{O_p(\sqrt{K/n}) + O_p(K^{-2})\} = O_p(h^2) + O_p\{(nh)^{-1/2}\}$. Finally for J_6, we have

$$\sum_{i,t}\psi_{nit}\{g(U_{it}) - \widehat{g}_n(U_{it})\} = \sum_{i,t}\psi_{nit}\{g(U_{it}) - \boldsymbol{\zeta}^{K\top}(u)\widehat{\boldsymbol{\theta}}_n\}$$
$$= \sum_{i,t}\psi_{nit}\{g(U_{it}) - \boldsymbol{\zeta}^{K\top}(u)\boldsymbol{\theta}\} + \sum_{i,t}\psi_{nit}D(\boldsymbol{\zeta\theta} - \mathbf{G})$$
$$+ \sum_{i,t}\psi_{nit}D\boldsymbol{\nu} + \sum_{i,t}\omega_{nit}(v)\sigma(V_{it})\varepsilon_{it}D\mathbf{X}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n)$$
$$= Q_1 + Q_2 + Q_3 + Q_4, \quad \text{say},$$

where $D = \boldsymbol{\zeta}^{K\top}(u)(\boldsymbol{\zeta}^\top\mathbf{M_B}\boldsymbol{\zeta})^{-1}\boldsymbol{\zeta}^\top\mathbf{M_B}$ and $\mathbf{G} = (g(U_{11}), \ldots, g(U_{nT}))^\top$. Applying Assumption 3, Theorem 1, Lemma A.1 and Eq. (A.1), it is not difficult to show that $Q_1 + Q_2 + Q_4 = O_p(h^2) + O_p\{(nh)^{-1/2}\}$. In addition, Assumption 4 and Bernstein's inequality lead to $Q_3 = O_p(h^2) + O_p\{(nh)^{-1/2}\}$. Similarly, we can show that

$$\sum_{i,t}\omega_{nit}(v)\sigma(V_{it})\nu_{it}\frac{1}{T}\sum_{t_1}\{g(U_{it_1}) - \widehat{g}(U_{it_1})\} = O_p(h^2) + O_p(n^{-1/2}h^{-1/2})$$

and

$$\sum_{i,t}\omega_{nit}(v)\sum_{t_1}\sigma(V_{it_1})v_{it_1}G(U_{it})=O_p(h^2)+O_p(n^{-1/2}h^{-1/2}).$$

Thus $J_6=O_p(h^2)+O_p\{(nh)^{-1/2}\}$. Now all equations in Eq. (A.2) hold, hence the proof is complete. □

Proof of Theorem 4. By the same argument as the proof of Theorem 3, we have

$$\begin{aligned}\widehat{\zeta}_n&=\frac{1}{nT}\sum_{i,t}\left\{\sigma(V_{it})v_{it}-\frac{1}{T}\sum_{t_1}\sigma(V_{it_1})v_{it_1}\right\}^2+O_p(n^{-1/2})\\&=\frac{1}{T}\left(1-\frac{1}{T}\right)\sum_t \mathrm{E}\{\sigma^2(V_{it})\}+O_p(n^{-1/2}).\end{aligned}$$

Combining this with Theorem 3 completes the proof of Theorem 4. □

Proof of Theorem 5. According to the definition of $\widehat{\boldsymbol\beta}_n^{\mathrm{w}}$, we have

$$\begin{aligned}\sqrt{nT}\,(\widehat{\boldsymbol\beta}_n^{\mathrm{w}}-\boldsymbol\beta)=\sqrt{nT}(\check{\boldsymbol\beta}_n^{\mathrm{w}}-\boldsymbol\beta)&+(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\mathbf{X}^{\mathrm{w}})^{-1}(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\widehat{\boldsymbol\Lambda}\mathbf{G}-\mathbf{X}^{*\top}\mathbf{M}^{*}\boldsymbol\Lambda\mathbf{G})\\&+\{(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\mathbf{X}^{\mathrm{w}})^{-1}-(\mathbf{X}^{*\top}\mathbf{M}^{*}\mathbf{X}^{*})^{-1}\}\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\widehat{\boldsymbol\Lambda}\mathbf{G}\\&+(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\mathbf{X}^{\mathrm{w}})^{-1}(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\boldsymbol\varepsilon^{\mathrm{w}}-\mathbf{X}^{*\top}\mathbf{M}^{*}\boldsymbol\varepsilon^{*})\\&+\{(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\mathbf{X}^{\mathrm{w}})^{-1}-(\mathbf{X}^{*\top}\mathbf{M}^{*}\mathbf{X}^{*})^{-1}\}\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\boldsymbol\varepsilon^{\mathrm{w}},\end{aligned}$$

where $\mathbf{M}^{w}=\mathbf{M}_{\mathbf{B}^{w}}\mathbf{M}_{\mathbf{M}_{\mathbf{B}^{w}}\zeta^{w}}\mathbf{M}_{\mathbf{B}^{w}}$. Therefore by Lemma A.3(i), it suffices to show that

$$\frac{1}{nT}(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\mathbf{X}^{\mathrm{w}}-\mathbf{X}^{*\top}\mathbf{M}^{*}\mathbf{X}^{*})=o_p(1),\tag{A.3}$$

$$\frac{1}{nT}(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\widehat{\boldsymbol\Lambda}\mathbf{G}-\mathbf{X}^{*\top}\mathbf{M}^{*}\boldsymbol\Lambda\mathbf{G})=o_p(n^{-1/2})\tag{A.4}$$

and

$$\frac{1}{nT}(\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\boldsymbol\varepsilon\mathrm{w}-\mathbf{X}^{*\tau}\mathbf{M}^{*}\boldsymbol\varepsilon^{*})=o_p(n^{-1/2}).\tag{A.5}$$

By the proof of Lemma A.1, the left side of Eq. (A.3) is equal to

$$\begin{aligned}&\frac{1}{nT}\sum_i(\boldsymbol\Pi_i^{\mathrm{w}\top}\mathbf{M}_{\iota_i^{\mathrm{w}}}\boldsymbol\Pi_i^{\mathrm{w}}-\boldsymbol\Pi_i^{*\top}\mathbf{M}_{\iota_i^{*}}\boldsymbol\Pi_i^{*})+o_p(1)\\&\quad=\frac{1}{nT}\sum_i(\boldsymbol\Pi_i^{\mathrm{w}\top}\boldsymbol\Pi_i^{\mathrm{w}}-\boldsymbol\Pi_i^{*\top}\boldsymbol\Pi_i^{*})+\frac{1}{nT}\sum_i(\boldsymbol\Pi_i^{\mathrm{w}\top}P_{\iota_i^{\mathrm{w}}}\boldsymbol\Pi_i^{\mathrm{w}}-\boldsymbol\Pi_i^{*\top}P_{\iota_i^{*}}\boldsymbol\Pi_i^{*})+o_p(1)\\&\quad=J_{11}+J_{12}+o_p(1),\quad\text{say}.\end{aligned}$$

It follows from Theorem 3 that

$$J_{11}=(nT)^{-1}\sum_{i,t}\boldsymbol\Pi_{it}^{\top}\boldsymbol\Pi_{it}\{\widehat{\sigma}_n^2(V_{it})-\sigma^2(V_{it})\}=O_p\left(h^2+n^{-1}h^{-1}\right)=o_p(1).$$

Similarly,

$$\begin{aligned}J_{12}&=\frac{1}{nT}\sum_i(\boldsymbol\Pi_i^{\mathrm{w}\top}-\boldsymbol\Pi_i^{*\top})P_{\iota_i^{\mathrm{w}}}\boldsymbol\Pi_i^{\mathrm{w}}+\frac{1}{nT}\sum_i\boldsymbol\Pi_i^{*\top}P_{\iota_i^{*}}(\boldsymbol\Pi_i^{\mathrm{w}}-\boldsymbol\Pi_i^{*})+\frac{1}{nT}\sum_i\boldsymbol\Pi_i^{*\top}(P_{\iota_i^{\mathrm{w}}}-P_{\iota_i^{*}})\boldsymbol\Pi_i^{\mathrm{w}}\\&=O_p\left(h^2+\frac{1}{nh}\right)+\frac{1}{nT}\sum_i\boldsymbol\Pi_i^{*\top}(\iota_i^{\mathrm{w}}-\iota_i^{*})(\iota_i^{\mathrm{w}\top}\iota_i^{\mathrm{w}})^{-1}\iota_i^{\mathrm{w}\top}\boldsymbol\Pi_i^{\mathrm{w}}\\&\quad+\frac{1}{nT}\sum_i\boldsymbol\Pi_i^{*\top}\iota_i^{*}\{(\iota_i^{\mathrm{w}\top}\iota_i^{\mathrm{w}})^{-1}-(\iota_i^{*\top}\iota_i^{*})^{-1}\}\iota_i^{*\top}\boldsymbol\Pi_i^{\mathrm{w}}+\frac{1}{nT}\sum_i\boldsymbol\Pi_i^{*\top}\iota_i^{*}(\iota_i^{*\top}\iota_i^{*})^{-1}(\iota_i^{\mathrm{w}\top}-\iota_i^{*\top})\boldsymbol\Pi_i^{\mathrm{w}}\\&=o_p(1).\end{aligned}$$

This proves Eq. (A.3). Furthermore, by Theorem 3 and the proof of Lemma A.3, we get

$$\frac{1}{nT}\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\widehat{\boldsymbol\Lambda}\mathbf{G}=\frac{1}{nT}\mathbf{X}^{\mathrm{w}\top}\mathbf{M}^{w}\widehat{\boldsymbol\Lambda}(\mathbf{G}-\boldsymbol\zeta\boldsymbol\xi)=o_p(n^{-1/2})$$

and

$$\frac{1}{nT}\mathbf{X}^{*\top}\mathbf{M}^*\boldsymbol{\Lambda}\mathbf{G} = \frac{1}{nT}\mathbf{X}^{*\top}\mathbf{M}^*\boldsymbol{\Lambda}(\mathbf{G}-\boldsymbol{\zeta}\boldsymbol{\xi}) = o_p(n^{-1/2}).$$

Thus Eq. (A.4) follows. It remains to prove Eq. (A.5). By Theorem 3 and the proof of Lemma A.3, the left side of Eq. (A.4) is equal to

$$\begin{aligned}
&\frac{1}{nT}\sum_i(\boldsymbol{\Pi}_i^{\mathrm{w}\top}\mathbf{M}_{\iota_i^{\mathrm{w}}}\boldsymbol{\varepsilon}_i^{\mathrm{w}} - \boldsymbol{\Pi}_i^{*\top}\mathbf{M}_{\iota_i^*}\boldsymbol{\varepsilon}_i^*) + o_p(n^{-1/2})\\
&\quad= \frac{1}{nT}\sum_i(\boldsymbol{\Pi}_i^{\mathrm{w}\top}\boldsymbol{\varepsilon}_i^{\mathrm{w}} - \boldsymbol{\Pi}_i^{*\top}\boldsymbol{\varepsilon}_i^*) + \frac{1}{nT}\sum_i(\boldsymbol{\Pi}_i^{\mathrm{w}\top} - \boldsymbol{\Pi}_i^{*\top})P_{\iota_i^{\mathrm{w}}}\boldsymbol{\varepsilon}_i^{\mathrm{w}}\\
&\qquad+ \frac{1}{nT}\sum_i\boldsymbol{\Pi}_i^{*\top}(\boldsymbol{\iota}_i^{\mathrm{w}}-\boldsymbol{\iota}_i^*)(\boldsymbol{\iota}_i^{\mathrm{w}\top}\boldsymbol{\iota}_i^{\mathrm{w}})^{-1}\boldsymbol{\iota}_i^{\mathrm{w}\top}\boldsymbol{\varepsilon}_i^{\mathrm{w}} + \frac{1}{nT}\sum_i\boldsymbol{\Pi}_i^{*\top}\boldsymbol{\iota}_i^*\left\{(\boldsymbol{\iota}_i^{\mathrm{w}\top}\boldsymbol{\iota}_i^{\mathrm{w}})^{-1} - (\boldsymbol{\iota}_i^{*\top}\boldsymbol{\iota}_i^*)^{-1}\right\}\boldsymbol{\iota}_i^{*\top}\boldsymbol{\varepsilon}_i^{\mathrm{w}}\\
&\qquad+ \frac{1}{nT}\sum_i\boldsymbol{\Pi}_i^{*\top}\boldsymbol{\iota}_i^*(\boldsymbol{\iota}_i^{*\top}\boldsymbol{\iota}_i^*)^{-1}(\boldsymbol{\iota}_i^{\mathrm{w}\top}-\boldsymbol{\iota}_i^{*\top})\boldsymbol{\varepsilon}_i^{\mathrm{w}}\\
&\quad= J_{21}+J_{22}+J_{23}+J_{24}+J_{25}, \quad \text{say.}
\end{aligned}$$

We can decompose J_{21} as

$$\begin{aligned}
J_{21} &= \frac{1}{nT}\sum_{i,t}\boldsymbol{\Pi}_{it}\varepsilon_{it}\{\widehat{\sigma}_n^{-2}(\cdot) - \sigma^{-2}(\cdot)\}\\
&= \frac{1}{nT}\sum_{i,t}\boldsymbol{\Pi}_{it}\varepsilon_{it}\{\widehat{\sigma}_n^2(\cdot)-\sigma^2(\cdot)\}\sigma^{-4}(\cdot) + \frac{1}{nT}\sum_{i,t}\boldsymbol{\Pi}_{it}\varepsilon_{it}\{\widehat{\sigma}_n^2(\cdot)-\sigma^2(\cdot)\}\{\widehat{\sigma}_n^{-2}(\cdot)-\sigma^{-2}(\cdot)\}\sigma^{-2}(\cdot)\\
&= J_{211}+J_{212}, \quad \text{say,}
\end{aligned}$$

where $\cdot$ denotes V_{it}. Furthermore,

$$\begin{aligned}
J_{211} &= \frac{1}{nT}\sum_{i,t}\frac{\boldsymbol{\Pi}_{it}\varepsilon_{it}}{\sigma^4(V_{it})}\left[a_T^2\left\{\tilde{\sigma}_n^2(v) - \frac{1}{T}\widehat{\zeta}_n\right\} - \sigma^2(V_{it})\right]\\
&= \frac{1}{nT}\sum_{i,t}\frac{\boldsymbol{\Pi}_{it}\varepsilon_{it}}{\sigma^4(V_{it})}\left[a_T^2\left\{\sum_{i_1,t_1}\omega_{ni_1t_1}(V_{it})\widehat{r}_{i_1t_1}^2 - \frac{1}{nT^2}\sum_{i_1,t_1}\widehat{r}_{i_1t_1}^2\right\} - \sigma_n^2(V_{it})\right]\\
&= \frac{1}{nT}\sum_{i,t}\frac{\boldsymbol{\Pi}_{it}\varepsilon_{it}}{\sigma^4(V_{it})}\left[a_T^2\sum_{i_1,t_1}\left\{\omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2}\right\}\widehat{r}_{i_1t_1}^2 - \sigma_n^2(V_{it})\right]\\
&= \frac{1}{nT}\sum_{i,t}\frac{\boldsymbol{\Pi}_{it}\varepsilon_{it}}{\sigma^4(V_{it})}\left[a_T^2\sum_{i_1,t_1}\left\{\omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2}\right\}\varepsilon_{i_1t_1}^{*2} - \sigma_n^2(V_{it})\right]\\
&\quad+ \frac{1}{nT}\sum_{i,t}\frac{a_T^2\boldsymbol{\Pi}_{it}\varepsilon_{it}}{\sigma^4(V_{it})}\sum_{i_1,t_1}\left\{\omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2}\right\}\left(\nabla_{i_1t_1} - \frac{1}{T}\sum_{t_2}\nabla_{i_1t_2}\right)^2\\
&\quad+ \frac{1}{nT}\sum_{i,t}\frac{a_T^2\boldsymbol{\Pi}_{it}\varepsilon_{it}}{\sigma^4(V_{it})}\sum_{i_1,t_1}\left\{\omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2}\right\}\left(\nabla_{i_1t_1} - \frac{1}{T}\sum_{t_2}\nabla_{i_1t_2}\right)\varepsilon_{i_1t_1}^*\\
&= J_{1211}+J_{1212}+J_{1213}, \quad \text{say,}
\end{aligned}$$

where $\varepsilon_{it}^* = \varepsilon_{it} - T^{-1}\sum_{t_1}\varepsilon_{it_1}$, $a_T = T/(T-1)$ and $\nabla_{it} = \mathbf{X}_{it}^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n) + g(U_{it}) - \widehat{g}_n(U_{it})$. Combining Lemma 2 in Gao [11] with Lemma A.1 and the proof of Theorem 3, we get

$$\begin{aligned}
J_{1211} &= \frac{1}{nT}\sum_{i,t}\boldsymbol{\Pi}_{it}\varepsilon_{it}\sum_{i_1,t_1}\left\{\omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2}\right\}\{\varepsilon_{i_1t_1}^{*2} - \mathrm{E}(\varepsilon_{i_1t_1}^{*2})\}O_p(1)\\
&\quad+ \frac{1}{nT}\sum_{i,t}\boldsymbol{\Pi}_{it}\varepsilon_{it}\left[a_T^2\sum_{i_1,t_1}\left\{\omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2}\right\}\mathrm{E}(\varepsilon_{i_1t_1}^{*2}) - \sigma_n^2(V_{it})\right]\\
&= o_p(n^{-1/2}).
\end{aligned}$$

By Theorems 1 and 2 we can show that $J_{1212} = o_p(n^{-1/2})$. In follows from Theorems 1 and 2 together with Lemma A.1 and the proof of Theorem 3 that

$$\begin{aligned}
&\max_{\substack{1\le i\le n\\1\le t\le T}} \left| \sum_{i_1,t_1} \left\{ \omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2} \right\} \left(\nabla_{i_1t_1} - \frac{1}{T} \sum_{t_2} \nabla_{i_1t_2} \right) \varepsilon^*_{i_1t_1} \right| \\
&\quad \le \max_{\substack{1\le i\le n\\1\le t\le T}} \left| \sum_{i_1,t_1} \left\{ \omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2} \right\} \left(\mathbf{X}_{i_1t_1} - \frac{1}{T} \sum_{t_2} \mathbf{X}_{i_1t_2} \right)^{\top} (\widehat{\boldsymbol\beta}_n - \boldsymbol\beta)\varepsilon^*_{i_1t_1} \right| \\
&\qquad + \max_{\substack{1\le i\le n\\1\le t\le T}} \left| \sum_{i_1,t_1} \left\{ \omega_{ni_1t_1}(V_{it}) - \frac{1}{nT^2} \right\} \left[\widehat{g}_n(U_{i_1t_1}) - g(U_{i_1t_1}) - \frac{1}{T} \sum_{t_2} \{\widehat{g}_n(U_{i_1t_2}) - g(U_{i_1t_2})\} \right] \varepsilon^*_{i_1t_1} \right| \\
&\quad = o_p(n^{-1/2}),
\end{aligned}$$

which implies $J_{1213} = O_p(n^{-1/2})$ and so $J_{211} = O_p(n^{-1/2})$. By Theorem 3,

$$|J_{212}| \le O(k) \left[\max_{\substack{1\le i\le n\\1\le t\le T}} \{\widehat{\sigma}_n^2(V_{it}) - \sigma^2(V_{it})\} \right]^2 \frac{1}{nT} \sum_{i,t} \|\boldsymbol\Pi_{it}\varepsilon_{it}\| = o_p(n^{-1/2}).$$

It follows that $J_{21} = o_p(n^{-1/2})$. By the same arguments we can show that $J_{2s} = o_p(n^{-1/2})$ for $s = 2, 3, 4, 5$ as well. Hence Eq. (A.5) holds, and the proof of Theorem 5 is complete. □

Proof of Theorem 6. This follows from the same arguments as in the proof of Theorem 5. □

Proof of Theorem 7. It follows from Theorem 4, Lemma A.2 and the proof of Theorem 5. □

References

[1] S.C. Ahn, P. Schmidt, Estimation of linear panel data models using GMM, in: L. Mátyás (Ed.), Generalized Method of Moments Estimation, Cambridge University Press, 2000, (Chapter 8).
[2] C. Ai, J. You, Y. Zhou, Statistical inference using a weighted difference-based series approach for partially linear models, J. Multivariate Anal. 102 (2011) 601–618.
[3] T. Amemiya, A note on a heteroscedastic model, J. Econometrics 6 (1977) 365–377.
[4] B.H. Baltagi, Econometric Analysis of Panel Data, Wiley, New York, 1995.
[5] B.H. Baltagi, D. Li, Series estimation of partially linear panel data models with fixed effects, Ann. Econ. Finance 3 (2002) 103–116.
[6] J.M. Chiou, H.G. Müller, Nonparametric quasi-likelihood, Ann. Statist. 27 (1999) 36–64.
[7] J. Duffy, C. Papageorgiou, A Cross-country empirical investigation of the aggregate production function apecification, J. Econ. Growth 5 (2000) 87–120.
[8] J. Fan, H. Peng, T. Huang, Semilinear high-dimensional model for normalization of microarray data: a theoretical analysis and partial consistency, J. Amer. Statist. Assoc. 100 (2005) 781–813.
[9] J. Fan, P. Tam, G. Vande Woude, Y. Ren, Normalization and analysis of cDNA micro-arrays using within-array replications applied to neuroblastoma cell response to a cytokine, Proc. Natl. Acad. Sci. (2004) 1135–1140.
[10] J. Fan, Q. Yao, Efficient estimation of conditional variance functions in stochastic regression, Biometrika 85 (1998) 645–660.
[11] J. Gao, The laws of the iterated logarithm of some estimates in partially linear models, Statist. Probab. Lett. 25 (1995) 153–162.
[12] P. Hall, R.J. Carroll, Variance function estimation in regression: the effect of estimating the mean, J. R. Stat. Soc. Ser. B Stat. Methodol. 51 (1989) 3–14.
[13] D.J. Henderson, R.J. Carroll, Q. Li, Nonparametric estimation and testing of fixed effects panel data models, J. Econometrics 144 (2008) 257–276.
[14] G. Kézdi, Robust standard error estimaton in fixed-effects panel models. Manuscript, Budapest University of Economics, IE/HAS and CEU, 2003.
[15] R. Liu, L. Yang, Spline-backfitted kernel smoothing of additive coefficient model, Econometric Theory 26 (2010) 29–59.
[16] R. Liu, L. Yang, W. Härdle, Oracally efficient two-step estimation of generalized additive model, J. Amer. Statist. Assoc. 108 (2013) 619–631.
[17] Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61 (1982) 405–415.
[18] L. Mátyás, P. Sevestre, The Econometrics of Panel Data, Springer, Berlin, 2008.
[19] J. Neyman, E. Scott, Consistent estimates based on partially consistent observations, Econometrica 16 (1948) 1–32.
[20] J.H. Stock, M.W. Watson, Heteroskedasticity-robust standard errors for fixed effects panel data regression, Econometrica 76 (2008) 155–174.
[21] Y. Sun, J. Carroll, D. Li, Semiparametric estimation of fixed-effects panel data varying coefficient models, in: Q. Li, J.S. Racine (Eds.), Nonparametric Econometric Methods, in: Advances in Econometrics, vol. 25, Emerald Group Publishing Limited, 2009, pp. 101–129.
[22] L. Wang, L. Yang, Spline-backfitted kernel smoothing of nonlinear additive autoregression model, Ann. Statist. 35 (2007) 2474–2503.
[23] J.H. You, X. Zhou, Y. Zhou, Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors, J. Multivariate Anal. 101 (2010) 1079–1101.
[24] J.H. You, Y. Zhou, X. Zhou, Series estimation in partially linear in-slide regression models, Scand. J. Statist. 38 (2011) 89–107.
[25] K. Yu, M.C. Jones, Likelihood-based local linear estimation of the conditional variance function, J. Amer. Statist. Assoc. 99 (2004) 139–144.