mathematics

MDPI

Article

Estimation and Simultaneous Confidence Bands for Fixed-Effects Panel Data Partially Linear Models

Suigen Yang [1,*], Xiujuan Yang [2] and Xuefei Wang [3]

1 College of Accounting, Guangzhou College of Technology and Business, Guangzhou 528138, China
2 IT Department, China Southern Airlines, Guangzhou 510405, China; yangxiujuan@csair.com
3 School of Statistics, Beijing Normal University, Beijing 100875, China; xuefeiwang@mail.bnu.edu.cn
* Correspondence: yangsuigen@gzgs.edu.cn

Abstract: In this paper, we study the estimation and simultaneous confidence band (SCB) problems for fixed-effects panel data partially linear models. We remove the fixed effects and then obtain estimators for the parametric and nonparametric components, which do not depend on the fixed effects. We establish the asymptotic distribution of the maximum absolute deviation between the estimated nonparametric component and the true nonparametric component under some suitable conditions; hence, this result can be used to construct the simultaneous confidence band for the nonparametric component. Based on the asymptotic distribution, it becomes difficult to construct the simultaneous confidence band. The reason for this is that the asymptotic distribution involves estimators of the asymptotic bias and conditional variance, as well as the choice of bandwidth for estimating the second derivative of the nonparametric function. Clearly, this will result in a computational burden and accumulated errors. To overcome these problems, we propose a bootstrap method to construct the simultaneous confidence band. The Monte Carlo results indicate that the proposed bootstrap method exhibits better performance with limited samples. An empirical application is presented to evaluate the performance of the proposed method.

Keywords: partially linear model; simultaneous confidence band; panel data; fixed effects; asymptotic property

MSC: 62G05; 62G20

check for updates

Citation: Yang, S.; Yang, X.; Wang, X. Estimation and Simultaneous Confidence Bands for Fixed-Effects Panel Data Partially Linear Models. *Mathematics* **2024**, *12*, 3774. https://doi.org/10.3390/math12233774

Academic Editor: Jin-Ting Zhang

Received: 21 October 2024
Revised: 21 November 2024
Accepted: 29 November 2024
Published: 29 November 2024

1. Introduction

Panel data models are widely used in econometrics to analyze data that vary over time and across entities (e.g., individuals, firms, or countries). These models can be broadly categorized into parametric and nonparametric models.

Parametric panel data models assume a specific functional form for the relationship between the dependent and independent variables. Baltagi [1] and Hsiao [2] provided an excellent overview of parametric panel data model analysis. Papke and Wooldridge [3] proposed a test to determine whether controlling for fixed effects at a more aggregate level is sufficient for panel data.

To relax the strong assumptions of parametric models, nonparametric and semiparametric panel data models have gained significant attention in recent years. These models offer greater flexibility by allowing the relationship between variables to take any form without imposing a specific functional form. Some examples are as follows:

- Fixed-Effects Nonparametric Models: Henderson et al. [4] studied nonparametric estimation and testing. Li et al. [5] investigated the construction of simultaneous confidence bands for nonparametric regression. Feng et al. [6] studied testing for heteroskedasticity in panel data models.

- Random-Effects Nonparametric Models: Henderson and Ullah [7] and Lin and Ying [8] focused on nonparametric estimation in random-effects panel data models.
- Partially Linear Models: Li and Stengos [9] considered a partially linear panel data model with some regressors being endogenous via the IV approach. Su and Ullah [10] investigated a fixed-effects partially linear panel data model with exogenous regressors. Zhang et al. [11] explored empirical likelihood inference for fixed-effects partially linear panel data models.
- Varying-Coefficient Models: Sun et al. [12] studied the problem of estimating varying-coefficient panel data models with fixed effects using a local linear regression approach. Feng et al. [13] examined varying-coefficient panel data models with interactive fixed effects.
- Semiparametric Single-Index Models: Chen et al. [14,15] and Lai et al. [16] studied semiparametric estimation for single-index panel data models.

These studies, along with many others, highlight the growing interest and importance of nonparametric and semiparametric methods in panel data analysis.

Recently, fixed-effects models have frequently been used in econometrics and biometrics. For example, in labor economics, the wage equation often includes both parametric and nonparametric components to account for the complex relationship between education, experience, and wages (Baltagi [1]). Another example is in the health sciences, where fixed-effects models have been used to analyze panel data, studying the progression of chronic diseases such as diabetes in patients (Fitzmaurice et al. [17]). In this paper, we consider the following fixed-effects panel data partially linear model:

$$Y_{it} = \boldsymbol{X}_{it}^{\mathrm{T}}\boldsymbol{\beta} + g(Z_{it}) + \alpha_i + V_{it}, \quad i = 1, \cdots, n, \quad t = 1, \cdots, T, \tag{1}$$

where $\{\boldsymbol{X}_{it}\}$ represents a $p \times 1$ vector of observable regressors, $\{Z_{it}\}$ represents the explanatory variables in [0,1], $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, $g(\cdot)$ is an unknown smooth function in [0,1], $\{V_{it}\}$ represents random errors with zero mean, and $\{\alpha_i\}$ represents the fixed effects. In addition, T is the time-series length, and n is the cross-sectional size.

For model (1), we assume that $\{\alpha_i\}$ represents the unobserved time-invariant individual effects. Model (1) is called a partially linear fixed-effects model if $\{\alpha_i\}$ is correlated with $\{\boldsymbol{X}_{it}, Z_{it}\}$ with an unknown correlation structure. For identification purposes, we define $\sum_{i=1}^{n} \alpha_i = 0$. An application of fixed-effects models is the study of individual wage rates, where α_i represents the different unobserved abilities of individual i, such as the unmeasured skills or unobservable characteristics of individual i, which may correlate with some observed covariates, such as age, educational level, job grade, gender, and work experience (Baltagi [1]). As a special case, when $\{\alpha_i\}$ is uncorrelated with $\{\boldsymbol{X}_{it}, Z_{it}\}$, model (1) becomes a partially linear random-effects model.

Baltagi and Li [18] applied the first-order difference to eliminate the fixed effects, used the series method to estimate the parametric and nonparametric components, and then established the asymptotic properties. Su and Ullah [10] considered the estimation of partially linear panel data models with fixed effects. Zhang et al. [11] applied the empirical likelihood method to model (1).

For partially linear panel data models, the existing literature has considered the pointwise asymptotic normality of the estimator for the nonparametric component, and the result can be used to construct pointwise confidence bands. In practice, we need to construct the simultaneous confidence band for the nonparametric function in the model. The simultaneous confidence band is a powerful tool for checking the graphical representation of the nonparametric function during practical applications. Therefore, there is extensive literature on the construction of simultaneous confidence bands. For example, Fan and Zhang [19] and Zhang and Peng [20] considered the simultaneous confidence bands for the coefficient functions in varying-coefficient models. Li et al. [5] considered the simultaneous confidence bands for nonparametric fixed-effects panel data models.

Li et al. [21] studied the simultaneous confidence bands and hypothesis testing for the link function in single-index models. Additionally, several studies have focused on the construction and application of simultaneous confidence bands for various models. For instance, Yothers and Sampson [22] developed methods for the difference of segmented linear models. Liu et al. [23] proposed techniques for a finite set of contrasts of three, four, or five generally correlated normal means. Yang et al. [24] developed methods for single-index random-effects models with longitudinal data. These studies provide valuable insights into the construction and interpretation of simultaneous confidence bands, which are essential for our analysis.

In this paper, we combine the idea of the least-squares dummy-variable approach in parametric panel data models with the local linear regression technique in nonparametric models and use the profile least-squares dummy-variable method proposed by Su and Ullah [10] to remove the fixed effects. Then, we obtain the estimators of the parametric and nonparametric components, which do not depend on the fixed effects. Under some suitable conditions, we establish the asymptotic distribution of the maximum absolute deviation between the estimated nonparametric component and the true nonparametric component; hence, the result can be used to construct the simultaneous confidence band for the nonparametric component. In order to construct the simultaneous confidence band based on the asymptotic distribution, we first need to estimate the asymptotic bias and conditional variance, as well as choose the bandwidth for estimating the second derivative of the nonparametric function. These steps introduce significant computational challenges. Specifically, the repeated estimation of the bias and variance, along with the selection of appropriate bandwidths, requires substantial computational resources and time. Additionally, the accumulation of small errors during each estimation step can propagate and magnify, leading to less accurate final estimates. These issues become particularly pronounced as the sample size increases, making the construction of the simultaneous confidence band based on the asymptotic distribution difficult. To overcome these problems, we further propose a bootstrap method to construct the simultaneous confidence band for the nonparametric component in model (1). Bissantz et al. [25] extended the bootstrap to nonparametric regression. Vos and Stauskas [26] studied the bootstrap for common correlated effects regressions.

The rest of this paper is organized as follows. In Section 2, we use the profile least-squares dummy-variable approach to obtain the estimators of the parametric and nonparametric components and present the asymptotic properties. In Section 3, we propose the bootstrap method to construct the simultaneous confidence band. In Section 4, the Monte Carlo results and an empirical application are used to illustrate the proposed method with limited samples. The technical proofs of the main theorems are presented in Appendix A.

2. Estimation Procedure and Asymptotic Properties

2.1. Estimation Procedure

Let $\{(Y_{it}; \boldsymbol{X}_{it}^{\mathrm{T}}, Z_{it}), i = 1, \cdots, n, t = 1, \cdots, T\}$ be an independent identically distributed (i.i.d.) random sample that comes from model (1). In this paper, we consider the asymptotic theory by letting n approach infinity while keeping T fixed. In this section, we describe the estimation procedure to first remove the fixed effects and then obtain the efficient estimators of the parametric and nonparametric components.

For ease of notation, let

$$
\begin{aligned}
\boldsymbol{Y} &= (Y_{11}, \cdots, Y_{1T}, Y_{21}, \cdots, Y_{2T}, \cdots, Y_{n1}, \cdots, Y_{nT})^{\mathrm{T}}, \\
\boldsymbol{g} &= \Big(g(Z_{11}), \cdots, g(Z_{1T}), g(Z_{21}), \cdots, g(Z_{2T}), \cdots, g(Z_{n1}), \cdots, g(Z_{nT})\Big)^{\mathrm{T}}, \\
\boldsymbol{V} &= (V_{11}, \cdots, V_{1T}, V_{21}, \cdots, V_{2T}, \cdots, V_{n1}, \cdots, V_{nT})^{\mathrm{T}}, \\
\boldsymbol{\alpha}_0 &= (\alpha_1, \cdots, \alpha_n)^{\mathrm{T}}
\end{aligned}
$$

and $\mathbf{X} = (\boldsymbol{X}_{11}, \cdots, \boldsymbol{X}_{1T}, \boldsymbol{X}_{21}, \cdots, \boldsymbol{X}_{2T}, \cdots, \boldsymbol{X}_{n1}, \cdots, \boldsymbol{X}_{nT})^{\mathrm{T}}$ is an $nT \times p$ matrix, where $\boldsymbol{X}_{it} = (X_{it1}, \cdots, X_{itp})^{\mathrm{T}}$. Then, model (1) can be written in the following matrix form:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{g} + (\mathbf{I}_n \otimes \boldsymbol{e}_T)\boldsymbol{\alpha}_0 + \boldsymbol{V}, \tag{2}$$

where $\mathbf{I}_n$ is an $n \times n$ identity matrix, $\boldsymbol{e}_T$ is a T-dimensional column vector with all elements equal to 1, and $\otimes$ denotes the Kronecker product. Furthermore, by the identification assumption $\sum_{i=1}^{n} \alpha_i = 0$, we have $\alpha_1 = -\sum_{i=2}^{n} \alpha_i$. Define the $(nT) \times (n-1)$ matrix $\mathbf{D} = [-\boldsymbol{e}_{n-1}, \mathbf{I}_{n-1}]^{\mathrm{T}} \otimes \boldsymbol{e}_T$, and $\boldsymbol{\alpha} = (\alpha_2, \cdots, \alpha_n)^{\mathrm{T}}$. Model (2) can then be rewritten as

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{g} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{V}. \tag{3}$$

Given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, model (3) is a version of the usual nonparametric fixed-effects panel data model

$$\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{D}\boldsymbol{\alpha} = \boldsymbol{g} + \boldsymbol{V}. \tag{4}$$

We first apply the local polynomial method (see the details in Fan and Gijbels [27]) to estimate the nonparametric function $g(\cdot)$. For Z_{it} in a small neighborhood of $z \in [0,1]$, approximate $g(Z_{it})$ by

$$g(Z_{it}) \approx g(z) + g'(z)(Z_{it} - z). \tag{5}$$

Let $K(\cdot)$ be a kernel function in $\mathbb{R}$, and define $K_h(z) = K(z/h)/h$, where h is the bandwidth. Let

$$\mathbf{Z}_z = \begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \\ Z_{11} - z & \cdots & Z_{1T} - z & \cdots & Z_{n1} - z & \cdots & Z_{nT} - z \end{pmatrix}^{\mathrm{T}},$$

and $\mathbf{W}_z = \mathrm{diag}(K_h(Z_{11} - z), \cdots, K_h(Z_{1T} - z), K_h(Z_{21} - z), \cdots, K_h(Z_{2T} - z), \cdots, K_h(Z_{n1} - z), \cdots, K_h(Z_{nT} - z))$ is an $(nT) \times (nT)$ diagonal matrix. Let $\boldsymbol{G}(z) = (g(z), (g'(z)))^{\mathrm{T}}$, and $\boldsymbol{\eta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$.

In what follows, we outline the estimation procedure for $\boldsymbol{\beta}$ and $g(\cdot)$. This step is similar to the approach described by Zhang et al. [11].

Given $\boldsymbol{\eta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$, we define the following weighted least-squares objective function:

$$(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_z\boldsymbol{G}(z) - \mathbf{D}\boldsymbol{\alpha})^{\mathrm{T}}\mathbf{W}_z(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_z\boldsymbol{G}(z) - \mathbf{D}\boldsymbol{\alpha}). \tag{6}$$

By minimizing the above objective function (6) with respect to $\boldsymbol{G}(z)$, we can obtain the solution for $\boldsymbol{G}(z)$ as follows:

$$\widetilde{\boldsymbol{G}}(z, \boldsymbol{\eta}) = (\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{D}\boldsymbol{\alpha}). \tag{7}$$

Define the smoothing operator as

$$\mathbf{M}(z) = (\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z.$$

Then, we can define the estimator of $g(z)$ as

$$\widetilde{g}(z, \boldsymbol{\eta}) = \boldsymbol{m}^{\mathrm{T}}(z)(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{D}\boldsymbol{\alpha}), \tag{8}$$

where $\boldsymbol{m}^{\mathrm{T}}(z) = \boldsymbol{e}^{\mathrm{T}}\mathbf{M}(z)$ and $\boldsymbol{e} = (1, 0)^{\mathrm{T}}$ is a 2×1 vector.

Since the fixed effects are an n-dimensional unobserved variable, it is difficult to obtain a consistent estimator for the fixed effects. Therefore, we first need to remove the fixed effects from the model and then obtain the estimators of the parametric and nonparametric components. By (8), we define the following objective function:

$$(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \widetilde{\boldsymbol{g}}_\eta(z) - \mathbf{D}\boldsymbol{\alpha})^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \widetilde{\boldsymbol{g}}_\eta(z) - \mathbf{D}\boldsymbol{\alpha})$$

$$\begin{aligned} &= [\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{M}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{D}\boldsymbol{\alpha}) - \mathbf{D}\boldsymbol{\alpha}]^{\mathrm{T}}[\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{M}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{D}\boldsymbol{\alpha}) - \mathbf{D}\boldsymbol{\alpha}] \quad (9)\\ &= (\widetilde{\boldsymbol{Y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta} - \widetilde{\mathbf{D}}\boldsymbol{\alpha})^{\mathrm{T}}(\widetilde{\boldsymbol{Y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta} - \widetilde{\mathbf{D}}\boldsymbol{\alpha}), \end{aligned}$$

where $\widetilde{g}_{\boldsymbol{\eta}}(z) = (\widetilde{g}(Z_{11}, \boldsymbol{\eta}), \cdots, \widetilde{g}(Z_{1T}, \boldsymbol{\eta}), \cdots, \widetilde{g}(Z_{n1}, \boldsymbol{\eta}), \cdots, \widetilde{g}(Z_{nT}, \boldsymbol{\eta}))$, $\widetilde{\boldsymbol{Y}} = (\mathbf{I}_{nT} - \mathbf{M})\boldsymbol{Y}$, $\widetilde{\mathbf{X}} = (\mathbf{I}_{nT} - \mathbf{M})\mathbf{X}$, $\widetilde{\mathbf{D}} = (\mathbf{I}_{nT} - \mathbf{M})\mathbf{D}$, $\widetilde{\mathbf{Q}} = \mathbf{I}_{nT} - \widetilde{\mathbf{D}}(\widetilde{\mathbf{D}}^{\mathrm{T}}\widetilde{\mathbf{D}})^{-1}\widetilde{\mathbf{D}}^{\mathrm{T}}$, and $\mathbf{M}$ is an $(nT) \times (nT)$ smoothing matrix, given by

$$\mathbf{M} = \begin{pmatrix} (1,0)(\mathbf{Z}_{Z_{11}}^{\mathrm{T}}\mathbf{W}_{Z_{11}}\mathbf{Z}_{Z_{11}})^{-1}\mathbf{Z}_{Z_{11}}^{\mathrm{T}}\mathbf{W}_{Z_{11}} \\ \vdots \\ (1,0)(\mathbf{Z}_{Z_{1T}}^{\mathrm{T}}\mathbf{W}_{Z_{1T}}\mathbf{Z}_{Z_{1T}})^{-1}\mathbf{Z}_{Z_{1T}}^{\mathrm{T}}\mathbf{W}_{Z_{1T}} \\ \vdots \\ (1,0)(\mathbf{Z}_{Z_{nT}}^{\mathrm{T}}\mathbf{W}_{Z_{nT}}\mathbf{Z}_{Z_{nT}})^{-1}\mathbf{Z}_{Z_{nT}}^{\mathrm{T}}\mathbf{W}_{Z_{nT}} \end{pmatrix}.$$

In addition, let $\mathbf{P} = (\mathbf{I}_{nT} - \mathbf{M})^{\mathrm{T}}(\mathbf{I}_{nT} - \mathbf{M})$ be an $(nT) \times (nT)$ matrix.

Taking the derivative of (10) with respect to $\boldsymbol{\alpha}$ and setting it equal to zero, we have

$$\widetilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = (\widetilde{\mathbf{D}}^{\mathrm{T}}\widetilde{\mathbf{D}})^{-1}\widetilde{\mathbf{D}}^{\mathrm{T}}(\widetilde{\boldsymbol{Y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta}). \quad (10)$$

Obviously, the estimator of the fixed effects depends on $\boldsymbol{\beta}$. Based on the idea of the least-squares dummy-variable approach in panel data parametric models and the nonparametric local linear regression technique, we then apply the profile least-squares dummy-variable method to estimate the parameter vector $\boldsymbol{\beta}$.

By plugging (10) into (10), we minimize the profile least-squares objective function with respect to $\boldsymbol{\beta}$. Thus, we obtain the profile least-squares estimator of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}} = (\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{Q}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{Q}}\widetilde{\boldsymbol{Y}}. \quad (11)$$

By (10) and (11), we have

$$\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_2, \cdots, \hat{\alpha}_n) = (\widetilde{\mathbf{D}}^{\mathrm{T}}\widetilde{\mathbf{D}})^{-1}\widetilde{\mathbf{D}}^{\mathrm{T}}(\widetilde{\boldsymbol{Y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\beta}}). \quad (12)$$

By $\sum_{i=1}^{n} \alpha_i = 0$ and (12), the estimator of α_1 is $\hat{\alpha}_1 = -\sum_{i=2}^{n} \hat{\alpha}_i$.

By (7), (11), and (12), as well as some simple calculations, we can obtain the estimator of $\boldsymbol{G}(z)$ as follows:

$$\begin{aligned} \hat{\boldsymbol{G}}(z) &= \hat{\boldsymbol{G}}(z, \hat{\boldsymbol{\eta}}) = \mathbf{M}(z)(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{D}\hat{\boldsymbol{\alpha}}) \\ &= \mathbf{M}(z)[\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{D}(\widetilde{\mathbf{D}}^{\mathrm{T}}\widetilde{\mathbf{D}})^{-1}\widetilde{\mathbf{D}}^{\mathrm{T}}(\widetilde{\boldsymbol{Y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\beta}})] \quad (13)\\ &= \mathbf{M}(z)[\mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}](\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \end{aligned}$$

By (8) and (14), we obtain the estimator of $g(z)$ as

$$\hat{g}(z) = \boldsymbol{m}^{\mathrm{T}}(z)[\mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}](\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (14)$$

Remark 1. *From (11) and (14), it is easy to see that the estimators of $\boldsymbol{\beta}$ and $g(\cdot)$ do not depend on the fixed effects.*

2.2. Asymptotic Properties

Let $\mu_l = \int z^l K(z)dz$ and $\nu_l = \int z^l K^2(z)dz$ for $l = 0, 1, 2$. Define the observed covariate set as $\mathcal{D} = \{\boldsymbol{X}_{it}, Z_{it}, 1 \le i \le n, 1 \le t \le T\}$. In order to obtain the main results, we first present the following technical conditions:

(C1) $(\alpha_i, \boldsymbol{V}_i, \mathbf{X}_i, \boldsymbol{Z}_i), i = 1, \cdots, n$, are i.i.d., where $\boldsymbol{V}_i = (V_{i1}, V_{i2}, \cdots, V_{iT})^{\mathrm{T}}$, and $\mathbf{X}_i$ and $\boldsymbol{Z}_i$ can be defined similarly. $E\|\boldsymbol{X}_{it}\|^{2+\delta} < \infty$ and $E\|V_{it}\|^{2+\delta} < \infty$ for some $\delta > 0$.

Let $\sigma^2(\boldsymbol{x}, z) = \mathrm{Var}(Y_{it}|\boldsymbol{X}_{it} = \boldsymbol{x}, Z_{it} = z)$, $\sigma^2(z) = \mathrm{Var}(Y_{it}|Z_{it} = z)$, and $0 < \sigma^2(\boldsymbol{x}, z), \sigma^2(z) < \infty$.

(C2) $E(Y_{it}|\mathbf{X}_i, \mathbf{Z}_i, \alpha_i) = E(Y_{it}|\boldsymbol{X}_{it}, Z_{it}, \alpha_i) = \boldsymbol{X}_{it}^{\mathrm{T}}\boldsymbol{\beta} + g(Z_{it}) + \alpha_i, i = 1, \cdots, n, t = 1, \cdots, T$.

(C3) Let $f(z) = \sum_{t=1}^{T} f_t(z)$, where $f_t(z)$ is the continuous density function of Z_{it}, and $f_t(z)$ is bounded away from zero and infinity on $[0, 1]$ for each $t = 1, \cdots, T$. Let $\widetilde{V}_{it} = V_{it} - \frac{1}{T}\sum_{s=1}^{T} V_{is}$, $\sigma_t^2(z) = E[\widetilde{V}_{it}^2|Z_{it} = z]$, and $\bar{\sigma}^2(z) = \sum_{t=1}^{T}\sigma_t^2(z)f(z)$.

(C4) Let $\boldsymbol{p}(z) = E(\boldsymbol{X}_{it}|Z_{it} = z)$. The functions $g(\cdot)$ and $\boldsymbol{p}(\cdot)$ have bounded and continuous second derivatives on the interval $[0, 1]$.

(C5) The kernel function $K(\cdot)$ is a symmetric density function and is absolutely continuous on its support set $[-A, A]$:

- (C5a) $K(A) \neq 0$;
- (C5b) $K(A) = 0$, $K(t)$ is absolutely continuous, and $K^2(t)$, $[K'(t)]^2$ are integrable on $(-\infty, +\infty)$.

(C6) The bandwidth h satisfies the conditions that $nh^3/\log n \to \infty$ and $nh^5 \log n \to 0$ as $n \to \infty$.

Conditions (C1)–(C3) are actually quite mild and can be easily satisfied. These conditions can also be found in the work of Zhang et al. [11]. Condition (C4) is the standard smoothness assumption, and a similar assumption can be found in the work of Li et al. [21]. Condition (C5) is a set of mild assumptions on the kernel function that have been used by many authors, such as Li et al. [5], Zhang et al. [11], Li et al. [21], and Fan and Gijbels [27]. Condition (C6) is the bandwidth assumption that can be found in the work of Fan and Gijbels [27].

Theorem 1. *Assume that conditions (C1)–(C6) hold. Let* $b(z) = h^2\mu_2 g''(z)/2$, $\Sigma_g = \nu_0\bar{\sigma}^2(z)f^{-2}(z)$, *and* $\Sigma_{g'} = \nu_2\bar{\sigma}^2(z)/(f^2(z)\mu_2^2)$, *uniformly for* $z \in [0, 1]$. *We have*

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| = O_p(n^{-1/2})$$

and

$$\sqrt{nh}\{\hat{g}(z) - g(z) - b(z)\} \xrightarrow{L} N(0, \Sigma_g),$$

$$\sqrt{nh^3}\{\hat{g}'(z) - g'(z)\} \xrightarrow{L} N(0, \Sigma_{g'}),$$

where "$\xrightarrow{L}$*" denotes the convergence in the distribution.*

Theorem 1 shows the $\sqrt{n}$-consistency of the parameter estimator $\widehat{\boldsymbol{\beta}}$ and the asymptotic normalities of the nonparametric component $\widehat{g}(\cdot)$ and its derivative $\widehat{g}'(\cdot)$. Theorem 1 can be used to make statistical inferences for the parametric component $\boldsymbol{\beta}$ and the nonparametric component $g(\cdot)$.

Theorem 2. *Assume that conditions (C1)–(C6) hold and* $h = O(n^{-\rho})$ *for* $1/5 \leq \rho < 1/3$. *For all* $z \in [0, 1]$, *we have*

$$P\left\{(-2\log h)^{1/2}\left(\sup_{z\in[0,1]}\left|(nh\Sigma_g^{-1})^{1/2}(\hat{g}(z) - g(z) - b(z))\right| - d_n\right) < u\right\}$$
$$\longrightarrow \exp(-2\exp(-u)), \quad \text{as } n \to \infty,$$

where if $K(A) \neq 0$,

$$d_n = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}}\left\{\log\frac{K^2(A)}{\nu_0\pi^{1/2}} + \frac{1}{2}\log\log h^{-1}\right\},$$

and if $K(A) = 0$,

$$d_n = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}} \log\left\{ \frac{1}{4\nu_0\pi} \int (K'(z))^2 dz \right\}.$$

Theorem 2 gives the asymptotic distribution of the maximum absolute deviation between the estimated nonparametric component $\hat{g}(\cdot)$ and the true nonparametric component $g(\cdot)$ when the estimator of $\boldsymbol{\beta}$ is $\sqrt{n}-$consistent. It provides the theoretical foundation for constructing the simultaneous confidence band for the nonparametric function in model (1).

Remark 2. *If the supremum in Theorem 2 is taken over an interval of $[c,d]$ instead of $[0,1]$, Theorem 2 still holds under certain conditions through a transformation. The asymptotic distribution is given by*

$$P\left\{ (-2\log h/(d-c))^{1/2} \Big(\sup_{z\in[c,d]} \left| (nh\Sigma_g^{-1})^{1/2} (\hat{g}(z) - g(z) - b(z)) \right| - \widetilde{d}_n \Big) < u \right\}$$
$$\longrightarrow \exp(-2\exp(-u)),$$

where $\widetilde{d}_n$ is the same as d_n in Theorem 2 except that h is replaced by $h/(d-c)$.

Theorem 3. *Assume that conditions (C1)–(C6) hold and $\Sigma_{g'} = \nu_2\bar{\sigma}^2(z)/(f^2(z)\mu_2^2)$. For all $z \in [0,1]$, we have*

$$P\left\{ (-2\log h)^{1/2} \Big(\sup_{z\in[0,1]} \left| (nh^3\Sigma_{g'}^{-1})^{1/2} (\hat{g}'(z) - g'(z)) \right| - d_{n_1} \Big) < u \right\}$$
$$\longrightarrow \exp(-2\exp(-u)), \quad \text{as } n \to \infty,$$

where $d_{n_1} = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}} \log\left\{ \frac{1}{2\pi\sqrt{\nu_2}} (\int z^2 (K'(z))^2 dz)^{1/2} \right\}$. If $K(c_0) = 0$, $K(z)$ is absolutely continuous, and $K^2(z)$, $(K'(z))^2$ are integrable on $(-\infty, +\infty)$.

Theorem 3 gives the asymptotic distribution of the maximum absolute deviation for $\hat{g}'(\cdot)$. The result in Theorem 3 provides the theoretical foundation for constructing the simultaneous confidence band for the derivative of the nonparametric component.

2.3. Simultaneous Confidence Band for the Nonparametric Function

Since the asymptotic bias and variance of $\hat{g}(\cdot)$ in Theorem 2 involve some unknown quantities, we cannot apply Theorem 2 to construct the simultaneous confidence band for $g(\cdot)$ directly. In order to construct the simultaneous confidence band for $g(\cdot)$, we first need to obtain the consistent estimators of the asymptotic bias and variance of $\hat{g}(\cdot)$. By Theorem 1, the asymptotic bias of $\hat{g}(z)$ is

$$(h^2\mu_2/2)g''(z)(1 + o_p(1)).$$

Thus, the consistent estimator of the asymptotic bias is $\widehat{\text{bias}}(\hat{g}(z)) = h^2\mu_2\hat{g}''(z)/2$, where the estimator $\hat{g}''(z)$ of $g''(z)$ is obtained using a local cubic fit with an appropriate pilot bandwidth $h_* = O(n^{-1/7})$, which is optimal for estimating $g''(z)$ and can be chosen by the residual squares criterion proposed by Fan and Gijbels [27].

Next, we estimate the asymptotic variance of $\hat{g}(z)$. For simplicity, suppose that the random errors V_{it} are i.i.d. for all i and t. By the proofs of the theorem, we have

$$\text{Var}\{\hat{g}(z)|\mathcal{D}\} = (1,0)(\mathbf{Z}_z^{\text{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}(\mathbf{Z}_z^{\text{T}}\mathbf{W}_z\mathbf{Q}_1\Phi_1\mathbf{Q}_1\mathbf{W}_z\mathbf{Z}_z)(\mathbf{Z}_z^{\text{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}(1,0)^{\text{T}},$$

where $\mathbf{Q}_1 = (\mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}^{\text{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\text{T}}\mathbf{P})$ and $\Phi_1 = \text{diag}(\sigma^2(Z_{11}), \cdots, \sigma^2(Z_{1T}), \sigma^2(Z_{21}), \cdots, \sigma^2(Z_{2T}), \cdots, \sigma^2(Z_{n1}), \cdots, \sigma^2(Z_{nT}))$. Using the similar approximate local homoscedasticity proposed by Li and Tong [5], the asymptotic variance of $\hat{g}(z)$ is defined as

$$\mathrm{Var}\{\hat{g}(z)|\mathcal{D}\} = (1,0)(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Q}_1\mathbf{W}_z\mathbf{Z}_z)(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}(1,0)^{\mathrm{T}}\sigma^2(z).$$

Let $\hat{\boldsymbol{V}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$ be the residual, where $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{g}} + \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{D}\hat{\boldsymbol{\alpha}}$. By (11), (12), and (14), we have

$$\begin{aligned}
\hat{\boldsymbol{V}} = \;& \boldsymbol{Y} - \hat{\boldsymbol{g}} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{D}\hat{\boldsymbol{\alpha}} \\
=\;& \boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{D}\hat{\boldsymbol{\alpha}} - \mathbf{M}(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{D}\hat{\boldsymbol{\alpha}}) \\
=\;& (\mathbf{I}_{nT} - \mathbf{M})(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{D}\hat{\boldsymbol{\alpha}}) \\
=\;& (\mathbf{I}_{nT} - \mathbf{M})(\mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P})(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
=\;& (\mathbf{I}_{nT} - \mathbf{M})\mathbf{Q}_1(\mathbf{I}_{nT} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{Q}_1\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{Q}_1)\boldsymbol{Y} \\
=\;& (\mathbf{I}_{nT} - \mathbf{M})\mathbf{Q}_1\mathbf{Q}_2\boldsymbol{Y},
\end{aligned} \tag{15}$$

where $\mathbf{Q}_2 = \mathbf{I}_{nT} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{Q}_1\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{Q}_1$. Obviously, the residual $\hat{\boldsymbol{V}}$ does not depend on the fixed effects and is a linear function of $\boldsymbol{Y}$. By the normalized weighted residual sum of squares, $\sigma^2(z)$ can be estimated by

$$\hat{\sigma}^2(z) = \frac{\hat{\boldsymbol{V}}^{\mathrm{T}}\hat{\boldsymbol{V}}}{\mathrm{tr}(\mathbf{Q}_2^{\mathrm{T}}\mathbf{Q}_1^{\mathrm{T}}\mathbf{P}\mathbf{Q}_1\mathbf{Q}_2)} = \frac{\boldsymbol{Y}^{\mathrm{T}}(\mathbf{Q}_2^{\mathrm{T}}\mathbf{Q}_1^{\mathrm{T}}\mathbf{P}\mathbf{Q}_1\mathbf{Q}_2)\boldsymbol{Y}}{\mathrm{tr}(\mathbf{Q}_2^{\mathrm{T}}\mathbf{Q}_1^{\mathrm{T}}\mathbf{P}\mathbf{Q}_1\mathbf{Q}_2)}.$$

Theorem 4. *Under the conditions in Theorem 2, and assuming that $\hat{g}^{(3)}(\cdot)$ is continuous on $[0,1]$ and the pilot bandwidth h_* satisfies $h_* = O(n^{-1/7})$, for all $z \in [0,1]$, we have*

$$P\left\{(-2\log h)^{1/2}\Bigg(\sup_{z\in[0,1]}\left|\frac{\hat{g}(z) - g(z) - \widehat{\mathrm{bias}}(\hat{g}(z)|\mathcal{D})}{[\widehat{\mathrm{Var}}\{\hat{g}(z)|\mathcal{D}\}]^{1/2}}\right| - d_n\Bigg) < u\right\} \longrightarrow \exp(-2\exp(-u)),$$

where d_n is defined in Theorem 2.

By Theorem 4, we construct the $(1-\alpha)\times 100\%$ simultaneous confidence band for the nonparametric function $g(z)$ as

$$\Big(\hat{g}(z) - \widehat{\mathrm{bias}}(\hat{g}(z)|\mathcal{D}) \pm \Delta_{1,\alpha}(z)\Big), \tag{16}$$

where $\Delta_{1,\alpha}(z) = \Big(d_n + [\log 2 - \log\{-\log(1-\alpha)\}](-2\log h)^{-1/2}\Big)\Big[\widehat{\mathrm{Var}}\{\hat{g}(z)|\mathcal{D}\}\Big]^{1/2}$.

3. The Bootstrap Method

Despite the fact that Theorem 4 provides the asymptotic distribution to construct the simultaneous confidence band (16) for the nonparametric component, we still need to estimate the asymptotic bias and the asymptotic conditional variance. First, the estimator of the asymptotic bias involves the estimator of the second derivative $g''(\cdot)$ and the choice of the pilot bandwidth h_* for estimating the second derivative $g''(\cdot)$. The estimator of the second derivative $g''(\cdot)$ has a slow convergence rate and is very sensitive to the pilot bandwidth h_*, which influences the estimator of the asymptotic bias. Second, the asymptotic variance estimation is very complicated, especially for the panel data semiparametric fixed-effects model. Finally, the asymptotic critical value c_α depends on the double exponential distribution and the estimators of the asymptotic bias and the asymptotic conditional variance. These not only result in a computational burden and accumulated errors but also make constructing the simultaneous confidence band difficult. To overcome these problems, we extend the bootstrap method used by Li et al. [5] to the partially linear panel data fixed-effects model (1).

Now, we discuss how to use the bootstrap procedure to construct the simultaneous confidence band for $g(\cdot)$. Let

$$T = \sup_{z\in[0,1]} \frac{|\hat{g}(z) - g(z)|}{\{\mathrm{Var}(\hat{g}(z|\mathcal{D}))\}^{1/2}}.$$

Suppose that the upper α quantile of T is c_α. If c_α and $\mathrm{Var}(\hat{g}(z|\mathcal{D}))$ are known, the simultaneous confidence band for $g(\cdot)$ with $(1-\alpha)\times 100\%$ on the interval $[0,1]$ should be

$$\hat{g}(z) \pm \{\mathrm{Var}(\hat{g}(z|\mathcal{D}))\}^{1/2} c_\alpha.$$

However, c_α and $\mathrm{Var}(\hat{g}(z|\mathcal{D}))$ are unknown. We obtain their estimators using the bootstrap method. Suppose that we have the estimators $\hat{c}_\alpha$ and $\mathrm{Var}^*(\hat{g}(z|\mathcal{D}))$ of c_α and $\mathrm{Var}(\hat{g}(z|\mathcal{D}))$, respectively. Then, we can obtain the $(1-\alpha)\times 100\%$ simultaneous confidence band for $g(\cdot)$ as follows:

$$\hat{g}(z) \pm \{\mathrm{Var}^*(\hat{g}(z|\mathcal{D}))\}^{1/2} \hat{c}_\alpha. \tag{17}$$

The bootstrap procedure is as follows:

(1) By (15), obtain the residuals $\hat{\boldsymbol{V}} = (\mathbf{I}_{nT} - \mathbf{M})\mathbf{Q}_1\mathbf{Q}_2\boldsymbol{Y}$, where $\hat{\boldsymbol{V}} = (\hat{V}_{11}, \cdots, \hat{V}_{1T}, \hat{V}_{21}, \cdots, \hat{V}_{2T}, \cdots, \hat{V}_{n1}, \cdots, \hat{V}_{nT})^{\mathrm{T}}$.

(2) For each $i = 1, \cdots, n$, $t = 1, \cdots, T$, obtain the bootstrap error $V^*_{it} = \hat{V}_{it}\varepsilon_{it}$, where ε_{it} are i.i.d. $\sim N(0,1)$ across i and t. Generate the bootstrap sample member Y^*_{it} by $Y^*_{it} = \hat{Y}_{it} + V^*_{it}$, $i = 1, \cdots, n$, $t = 1, \cdots, T$.

(3) Given the bootstrap resample $\{(Y^*_{it}, \boldsymbol{X}_{it}, Z_{it}), i = 1, \cdots, n,\ t = 1, \cdots, T\}$, obtain the estimators of $\boldsymbol{\beta}$ and $g(\cdot)$, and denote the resulting estimates by $\hat{\boldsymbol{\beta}}^*$ and $\hat{g}^*(\cdot)$ as the bootstrap estimators of $\boldsymbol{\beta}$ and $g(\cdot)$, respectively.

(4) Repeat (2)–(3) N times to obtain a size N bootstrap sample of $g(\cdot)$, $\hat{g}^*_k(\cdot), k = 1, \cdots, N$. The estimator $\mathrm{Var}^*(\hat{g}(z))$ of $\mathrm{Var}(\hat{g}(\cdot))$ is taken as the sample variance of $\hat{g}^*_k(\cdot)$.

(5) Compute the bootstrap sample of T by

$$T^*_k = \sup_{z\in[0,1]} \frac{|\hat{g}^*_k(z) - \hat{g}(z)|}{\{\mathrm{Var}^*(\hat{g}(z|\mathcal{D}))\}^{1/2}}, \qquad k = 1, \cdots, N.$$

Use the upper α percentile $\hat{c}_\alpha$ of $T^*_k, k = 1, \cdots, N$, to estimate the upper α quantile c_α of T.

We can construct the $(1-\alpha)\times 100\%$ simultaneous confidence band for $g(\cdot)$ by (17) when we obtain the estimators of c_α and $\mathrm{Var}(\hat{g}(z|\mathcal{D}))$.

4. Monte Carlo Results and Empirical Application

4.1. Monte Carlo Results

We conducted simulation studies to assess the performance of our proposed method. Our simulated data were generated from the following model:

$$Y_{it} = \boldsymbol{X}^{\mathrm{T}}_{it}\boldsymbol{\beta} + 0.8\cos(\pi Z_{it}) + \alpha_i + V_{it},\ \ i = 1, \cdots, n,\ \ t = 1, \cdots, T, \tag{18}$$

where $\boldsymbol{\beta} = (-1, 3, 5)^{\mathrm{T}}$, $\boldsymbol{X}_{it}$ are three-dimensional i.i.d. random variables from uniform [−1,1], Z_{it} are i.i.d. from uniform [−1,1], and the random errors V_{it} are i.i.d. from $N(0,1)$. In this simulation, we only considered α_i to be correlated with the covariate $Z_{i.}$ and generated $\alpha_i = \varepsilon_i + cZ_{i\cdot}, i = 2, \cdots, n$, where $\varepsilon_i \sim N(0,1), Z_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} Z_{it}$, and $\alpha_1 = -\sum_{i=2}^{n}\alpha_i$, $i = 1, \cdots, n$. We considered three cases for $c = 0, 0.5, 1$. When $c \neq 0$, Z_{it} and α_i are correlated, and model (18) is the partially linear fixed-effects model. When $c = 0$, model (18) leads to the usual partially linear random-effects model.

In our simulation studies, we applied the Epanechnikov kernel $K(z) = 0.75(1 - z^2)_+$ to estimate the nonparametric function. Finding an appropriate bandwidth can be of both theoretical and practical interest. To implement the estimation procedure described in Section 2, we need to choose the bandwidth h. One can select h by minimizing the generalized cross-validation criterion. Here, we use the following cross-validation method to automatically select the optimal bandwidth h_{CV}:

$$\mathrm{CV}(h) = \sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}^{-it})^2 = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{Y_{it} - \hat{Y}_{it}}{1 - l_{kk}}\right)^2 = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{\hat{V}_{it}}{1 - l_{kk}}\right)^2, \tag{19}$$

where Y_{it}^{-it} denotes the fitted values that are computed from the data, with the measurements of the $\{Y_{it}, \mathbf{X}_{it}\}$ observation deleted. $k = (i-1)T + t$, $\hat{V}_{it} = Y_{it} - \hat{Y}_{it}$, and l_{kk} is the (k,k) element of the matrix $[\mathbf{I}_{nT} - (\mathbf{I}_{nT} - \mathbf{M})\mathbf{Q}_1\mathbf{Q}_2]$. The cross-validation bandwidth h_{CV} is then defined as the minimizer of $\mathrm{CV}(h)$.

We fixed $T = 5$ and examined the finite sample performance of the proposed method when the sample size was set to $n = 100, 150$, and 200. For each case, 1000 replicates of the simulated realizations were generated, and the nominal level was set to $1 - \alpha = 0.95$. The results are given in Tables 1 and 2 and Figure 1. To compare the performance of the parameter estimation in the cases where Z_{it} and α_i are correlated and uncorrelated, Table 1 gives the bias, the standard deviation, and the mean squared error of the estimator $\hat{\boldsymbol{\beta}}$ for $c = 0$ and $c = 1$. In order to save space, we omit the results for $c = 0.5$; however, the results for $c = 0.5$ were similar to those for $c = 1$. In Table 1, we can see that the bias, the standard deviation, and the mean squared error decreased as the sample size n increased for both cases. For the same sample size n, the results for $c = 1$ were better than those for $c = 0$. Model (18) reduced to a partially linear random-effects model when $c = 0$. From (11) and (14), it is easy to see that, in order to remove the fixed effects from the model, some sample information was lost when obtaining the estimators of the parametric and nonparametric components. So, the profile least-squares dummy-variable method is not suitable for the partially linear random-effects model, and the resulting estimators of the parametric and nonparametric components are not efficient. Thus, we need to develop an effective estimation procedure to estimate the random-effects models, such as the generalized profile least-squares method or the generalized estimating equation (GEE).

Table 1. The bias, standard deviation (SD), and mean squared error (MSE) of $\hat{\boldsymbol{\beta}}$ for $n = 100, 150, 200$ and $c = 0, 1$.

		$c = 0$			$c = 1$		
$\hat{\beta}$		**100**	**150**	**200**	**100**	**150**	**200**
$\hat{\beta}_1$	Bias	0.0063	0.0059	0.0048	0.0045	0.0046	0.0023
	SD	0.0859	0.0720	0.0682	0.0841	0.0647	0.0635
	MSE	0.0074	0.0052	0.0046	0.0071	0.0042	0.0040
$\hat{\beta}_2$	Bias	0.0057	0.0046	0.0031	0.0053	0.0027	0.0022
	SD	0.0901	0.0696	0.0620	0.0906	0.0687	0.0601
	MSE	0.0081	0.0049	0.0038	0.0082	0.0048	0.0036
$\hat{\beta}_3$	Bias	0.0062	0.0049	0.0042	0.0041	0.0029	0.0026
	SD	0.0912	0.0770	0.0650	0.0857	0.0679	0.0545
	MSE	0.0083	0.0059	0.0042	0.0074	0.0046	0.0031

Based on the asymptotic distribution and the bootstrap method, Table 2 gives the average probabilities of the simultaneous confidence band for the nonparametric function $g(\cdot)$ when the nominal level is $1 - \alpha = 0.95$, where "method one" denotes the method based on the asymptotic distribution and "bootstrap" denotes the method based on the bootstrap procedure. For the bootstrap procedure, we used $M = 200$ bootstrap replications to estimate c_α and $\mathrm{Var}(\hat{g}(z|\mathcal{D}))$.

In Table 2, it is clear that the average coverage probabilities of the simultaneous confidence band for the nonparametric function obtained by the two methods tend to 0.95 as the sample size n increases across the three cases. When $c = 0$, the average coverage probabilities are lower than those for $c = 0.5$ and 1. In addition, we can see that the average coverage probabilities based on the asymptotic distribution are lower than those of the bootstrap method. Furthermore, Figure 1 demonstrates that the simultaneous confidence band of the bootstrap method is narrower than that of method one, yet it covers the true curve more frequently, indicating the superiority of the bootstrap method. The reason for this is that the bootstrap method avoids estimating the asymptotic bias and variance.

Table 2. Coverage probabilities of the nonparametric component with the nominal level 95% for $n = 100, 150, 200$ and $c = 0, 0.5, 1$, respectively.

	n	$c = 0$	$c = 0.5$	$c = 1$
Method one	100	0.926	0.933	0.941
	150	0.933	0.940	0.949
	200	0.946	0.951	0.953
Bootstrap	100	0.928	0.934	0.942
	150	0.937	0.946	0.950
	200	0.948	0.952	0.954

Based on the asymptotic distribution and the bootstrap method, Figure 1 shows the 95% pointwise confidence bands for $g(\cdot)$ for $n = 100, 150, 200$ and $c = 0, 0.5, 1$, respectively. Figure 1 reveals that the performance of the asymptotic confidence bands is no worse than that based on the bootstrap procedure. In addition, the confidence bands obtained by the two methods become narrower as the sample size n increases across the three cases. In Table 2 and Figure 1, it is clear that, although the bootstrap method works better than the method based on the asymptotic distribution, the proposed asymptotic distribution method is comparable to the bootstrap method. However, bootstrap methods are computationally expensive. In Table 3, it is clear that the runtime of method one is relatively insensitive to the sample size n, remaining consistently around 0.0006 s. In contrast, the runtime of the bootstrap method is thousands of times longer than that of method one, and it increases significantly with larger sample sizes. Therefore, as the sample size increases, the bootstrap method becomes considerably more computationally expensive.

Table 3. Program running times (seconds) for $n = 100, 150, 200$ and $c = 0, 0.5, 1$, respectively.

	n	$c = 0$	$c = 0.5$	$c = 1$
Method one	100	4.9×10^{-4}	6.7×10^{-4}	5.7×10^{-4}
	150	6.0×10^{-4}	6.0×10^{-4}	5.7×10^{-4}
	200	8.8×10^{-4}	8.0×10^{-4}	6.8×10^{-4}
Bootstrap	100	8.6	8.4	8.8
	150	22.7	22.6	23.4
	200	57.3	63.2	47.6

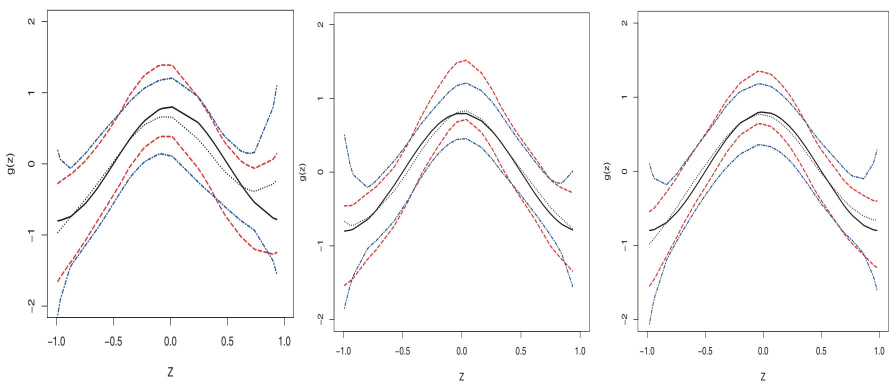

Figure 1. *Cont.*

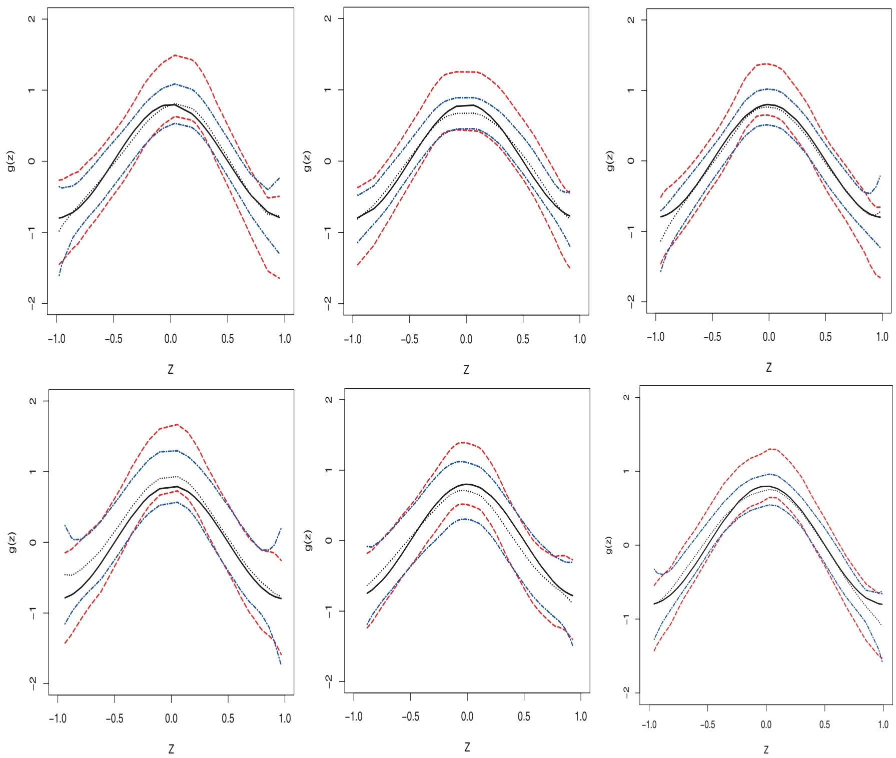

Figure 1. The solid lines denote the functional relationship, the dotted lines denote the estimated functional relationship, the long-dashed lines denote the 95% simultaneous confidence bands based on the asymptotic distribution, and the dash-dotted lines denote the 95% simultaneous confidence bands based on the bootstrap procedure for $g(\cdot)$. The figures are displayed for $c = 0, 0.5, 1$ from top to bottom and for the sample sizes $n = 100, 150, 200$ from left to right, respectively.

4.2. Empirical Application

The dataset comes from a cigarette consumption study [1], which is a panel of 46 observations from 1963 to 1992. The data contain nine variables: state, year, price per pack of cigarettes (price), population (pop), population above the age of 16 (pop16), consumer price index (cpi), per capita disposable income (ndi), cigarette sales in packs per capita (sales), and minimum price in adjoining states per pack of cigarettes (pimin). The variables were scaled using minimum-maximum normalization, with each feature scaled to a fixed interval [0,1]. Because the pairwise correlation coefficients of price, cpi, ndi, and pimin all exceeded 0.9, price was selected. Because the correlation coefficient of pop and pop16 exceeded 0.9, pop16 was selected. In this study, we analyzed the relationship between cigarette sales (sales), the price of cigarettes (price), and the logarithm of the population aged 16 and over (lpop16). The dataset consisted of panel data from multiple states over several years. The primary goal was to understand how changes in the price of cigarettes and the population aged 16 and over affect cigarette sales. We considered the following fixed-effects panel data partially linear model:

$$\boldsymbol{sales}_{it} = \boldsymbol{\beta} \cdot \boldsymbol{price}_{it} + g(\boldsymbol{lpop16}_{it}) + \boldsymbol{\alpha}_i + V_{it}, \quad i = 1, \cdots, 46, \quad t = 1, \cdots, 30. \tag{20}$$

When the nonparametric part $g(\cdot)$ is assumed to be a linear function, the model reduces to the classical fixed-effects linear model:

$$\boldsymbol{sales}_{it} = \boldsymbol{\beta_1} \cdot \boldsymbol{price}_{it} + \boldsymbol{\beta_2} \cdot \boldsymbol{lpop16}_{it} + \boldsymbol{\alpha}_i + V_{it}, \quad i = 1, \cdots, 46, \quad t = 1, \cdots, 30. \tag{21}$$

We estimated the fixed-effects linear model using the plm function from the plm package in R and obtained the following results: the price coefficient $\boldsymbol{\beta_1} = -0.232$ and the lpop16 coefficient $\boldsymbol{\beta_2} = 0.725$. The estimated coefficient for the price of cigarettes was negative, indicating an inverse relationship between price and sales. The population aged 16 and over has a significant positive effect on cigarette sales.

The estimation procedure proposed in Section 2 was used to estimate model (20). During the estimation process, we applied the Epanechnikov kernel $K(z) = 0.75(1 - z^2)_+$ to estimate the nonparametric function, with the bandwidth $h = 0.2 * (46 * 30)^{-1/5} \approx 0.05$, and the bootstrap was repeated $N = 100$ times.

The estimated coefficient was $\hat{\beta} = -0.254$. The estimated coefficient for the price of cigarettes was negative, indicating an inverse relationship between price and sales. Both models showed a significant negative effect of price on sales, with similar coefficient estimates. This consistency supports the robustness of our findings. It suggests that higher prices lead to reduced cigarette consumption, which is consistent with economic theory and previous empirical studies.

Using the proposed estimation procedure, the scatter plots of the variable $\boldsymbol{lpop16}_{it}$ versus $\boldsymbol{sales}_{it} - \hat{\beta} \cdot \boldsymbol{price}_{it} - \hat{\boldsymbol{\alpha}}_i$, the estimated function $\hat{g}(\cdot)$, and the 95% SCB for the function $g(\cdot)$ are shown in Figure 2. In Figure 2, it can be seen that as $\boldsymbol{lpop16}$ increases, $\boldsymbol{sales}$ increases nonlinearly. The 95% SCB for the function $g(\cdot)$ covers most of the scatter points, especially in the middle section, where the SCB uses a very narrow interval to cover most of the scatter points. When $\boldsymbol{lpop16}$ approaches 1, the variance increases due to fewer sample points, leading to a wider confidence interval. The classical model estimates a positive linear effect of the population aged 16 and over on sales, while our model captures a nonlinear effect. The 95% SCB in our model provides a more comprehensive understanding of the relationship.

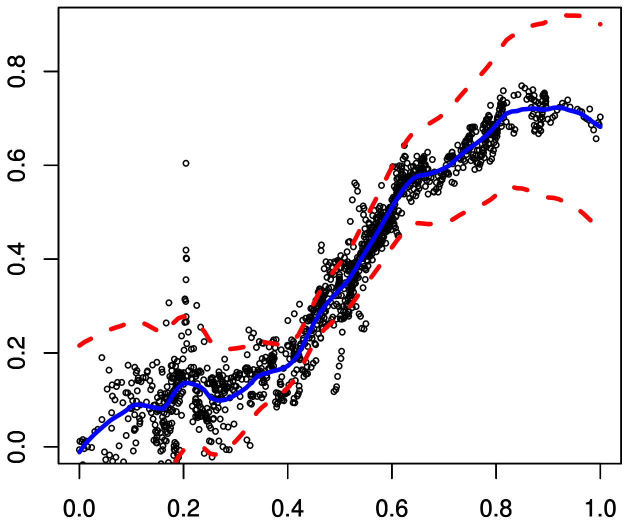

Figure 2. The scatter plot of the variable $\boldsymbol{lpop16}_{it}$ versus $\boldsymbol{sales}_{it} - \hat{\beta} \cdot \boldsymbol{price}_{it} - \hat{\boldsymbol{\alpha}}_i$. The solid line denotes the estimated function $\hat{g}(\cdot)$. The dotted lines denote the 95% simultaneous confidence bands for $g(\cdot)$.

5. Conclusions

In this paper, we consider the estimation and simultaneous confidence band (SCB) problems for fixed-effects panel data partially linear models. We remove the fixed effects and then obtain estimators for the parametric and nonparametric components, which do not depend on the fixed effects. We establish the asymptotic distribution of the maximum absolute deviation between the estimated nonparametric component and the true nonparametric component under some suitable conditions; hence, this result can be used to construct the simultaneous confidence band for the nonparametric component. We propose a bootstrap method to construct the simultaneous confidence band. Simulation studies indicate that the proposed bootstrap method exhibits better performance with limited samples. A real data analysis illustrates the practical use of the constructed confidence band. The methodology in this paper is general and widely applicable; therefore, we expect further research along these lines to yield deep theoretical results with interesting applications for other nonparametric or semiparametric models with fixed-effects panel data models.

However, the bootstrap method, while providing robust and accurate results, has some limitations. The computational burden of the bootstrap method is substantial, especially as the sample size increases. This limitation restricts the applicability of the model to large-scale datasets. So, we should focus on improving the computational efficiency of the bootstrap method to enhance its feasibility for large sample data. Going forward, we plan to further test and optimize our approach under appropriate conditions to promote its application in real-world scenarios.

Author Contributions: Conceptualization, S.Y. and X.Y.; methodology, S.Y. and X.Y.; software, S.Y.; formal analysis, S.Y. and X.W.; writing—original draft preparation, S.Y. and X.Y.; writing—review and editing, S.Y. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used to support the findings of this study are openly available through the R package plm, version 2.6-4.

Acknowledgments: The authors are deeply grateful to Gaorong Li for providing critical feedback and guidance during the writing of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Proofs of the Main Results

Let $\mathbf{P} = (\mathbf{I}_{nT} - \mathbf{M})^{\mathrm{T}}(\mathbf{I}_{nT} - \mathbf{M})$ and $\Phi = \sum_{t=1}^{T}\sum_{s=1}^{T} E\{\widetilde{\boldsymbol{X}}_{it}[\widetilde{\boldsymbol{X}}_{is} - \sum_l \widetilde{\boldsymbol{X}}_{il}/T]^{\mathrm{T}} V_{it} V_{is}\}$. Lemmas A1–A5 play a very important role in proving the main results of Theorems 1–4. The details of the proofs can be found in the works of Su and Ullah [10] and Zhang et al. [11]. So, we omit the details here.

Lemma A1. *Assume that conditions (C1)–(C6) hold. Let C be a positive constant and $m(Z_{it}, z) = \boldsymbol{e}^{\mathrm{T}}(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}Z_{z_{it}}K_h(Z_{it} - z)$, where $Z_{z_{it}}$ is a typical column of $\mathbf{Z}_z$. Then, we have the following:*

(i) $m(Z_{it}, z) = n^{-1}K_h(Z_{it} - z)f^{-1}(z)\{1 + o_p(1)\}$*, where* $f(z) = \sum_{t=1}^{T} f_t(z)$*;*

(ii) $\lim_{n\to\infty} P_n\left\{\sup_{z\in[0,1]}\max_{1\le i\le n, 1\le t\le T}|m(Z_{it}, z)| \le C(nh)^{-1}\right\} = 1.$

Lemma A2. *Assume that conditions (C1)–(C6) hold. Then, we have*

$$(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1} = (\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1} + O_p(\zeta_n) = T^{-1}\mathbf{I}_{n-1} + O_p(\zeta_n),$$

where $\zeta_n = (\boldsymbol{e}_{n-1}\boldsymbol{e}_{n-1}^{\mathrm{T}})(nh)^{-1}\sqrt{\ln n}$.

Lemma A3. *Assume that conditions (C1)–(C6) hold. Then, we have the following:*

(i) $\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{X} \xrightarrow{P} \sum_{t=1}^{T} E[(\boldsymbol{X}_{it} - \boldsymbol{p}(Z_{it}))(\boldsymbol{X}_{it} - \boldsymbol{p}(Z_{it}))^{\mathrm{T}}]$*;*

(ii) $\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-\mathbf{1}}\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{X} \xrightarrow{P} \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E[(\boldsymbol{X}_{it} - \boldsymbol{p}(Z_{it}))(\boldsymbol{X}_{is} - \boldsymbol{p}(Z_{is}))^{\mathrm{T}}]$*;*

(iii) $\frac{1}{n}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{Q}}\widetilde{\mathbf{X}} \xrightarrow{P} \Phi$.

Lemma A4. *Assume that conditions (C1)–(C6) hold. Then, we have*

$$\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{Q}}(\mathbf{I}_n - \mathbf{M})\boldsymbol{g}(\mathbf{Z}) = o_p(1).$$

Lemma A5. *Assume that conditions (C1)–(C6) hold. Then, we have*

$$\frac{1}{\sqrt{n}}\mathbf{X}^{\mathrm{T}}\mathbf{P}\boldsymbol{V}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{X}_{it}-\boldsymbol{p}(Z_{it}))V_{it}+o_p(1),$$

$$\frac{1}{\sqrt{n}}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}\boldsymbol{V}=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}(\boldsymbol{X}_{it}-\boldsymbol{p}(Z_{it}))V_{is}+o_p(1).$$

Proof of Theorem 1. The proofs of Theorem 1 can immediately be obtained from the works of Su and Ullah [10] and Zhang et al. [11] by Lemmas A1–A5. So, we omit the details here. □

Proof of Theorem 2. Note that $(\mathbf{I}_{nT}-\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P})\mathbf{D}\boldsymbol{\alpha}=0$. By (12), (14), and Lemma A2, we have

$$\begin{aligned}\hat{g}(z)=&\ \boldsymbol{m}^{\mathrm{T}}(z)(\boldsymbol{Y}-\mathbf{D}\hat{\boldsymbol{\alpha}}-\mathbf{X}\hat{\boldsymbol{\beta}})\\=&\ \boldsymbol{m}^{\mathrm{T}}(z)[\mathbf{I}_{nT}-\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}](\boldsymbol{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})\\=&\ \boldsymbol{m}^{\mathrm{T}}(z)[\mathbf{I}_{nT}-\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}](\boldsymbol{g}+\boldsymbol{V}-\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}))\\=&\ \boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1[\boldsymbol{g}+\boldsymbol{V}-\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})].\end{aligned}\tag{A1}$$

Invoking the Taylor expansion, we have

$$g(Z_{it})\approx g(z)+g'(z)(Z_{it}-z)+\frac{1}{2}g''(z)(Z_{it}-z)^2,\tag{A2}$$

where Z_{it} is close to $z\in[0,1]$. By (A1) and (A2), we have

$$\begin{aligned}\hat{g}(z)\approx&\ \boldsymbol{m}^{\mathrm{T}}(z)[\mathbf{I}_{nT}-\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}]g(z)\boldsymbol{e}_{nT}+\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1g'(z)\mathbf{Z}_z\\&+\frac{1}{2}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1g''(z)\mathbf{Z}_z^2+\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1\boldsymbol{V}-\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\\=&\ \boldsymbol{m}^{\mathrm{T}}(z)\mathbf{I}_{nT}g(z)\boldsymbol{e}_{nT}-\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}g(z)\boldsymbol{e}_{nT}+\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1g'(z)\mathbf{Z}_z\\&+\frac{1}{2}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1g''(z)\mathbf{Z}_z^2+\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1\boldsymbol{V}-\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}),\end{aligned}\tag{A3}$$

where $\mathbf{Z}_z=(Z_{11}-z,\cdots,Z_{1T}-z,Z_{21}-z,\cdots,Z_{2T}-z,\cdots,Z_{n1}-z,\cdots,Z_{nT}-z)^{\mathrm{T}}$. For ease of notation, let $S_{nT,l}(z)=\sum_{i=1}^{n}\sum_{t=1}^{T}K_h(Z_{it}-z)(Z_{it}-z)^l$, $l=0,1,2$. For the first term of (A3), some simple calculations yield that

$$\begin{aligned}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{I}_{nT}g(z)\boldsymbol{e}_{nT}=&\ (1,0)(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{I}_{nT}\boldsymbol{e}_{nT}g(z)\\=&\ (1,0)\begin{pmatrix}S_{nT,0}(z)&S_{nT,1}(z)\\S_{nT,1}(z)&S_{nT,2}(z)\end{pmatrix}^{-1}\begin{pmatrix}S_{nT,0}(z)\\S_{nT,1}(z)\end{pmatrix}g(z)\\=&\ (1,0)\begin{pmatrix}S_{nT,2}(z)&-S_{nT,1}(z)\\-S_{nT,1}(z)&S_{nT,0}(z)\end{pmatrix}\begin{pmatrix}S_{nT,0}(z)\\S_{nT,1}(z)\end{pmatrix}g(z)\\&\times\left(S_{nT,0}(z)S_{nT,2}(z)-S_{nT,1}^2(z)\right)^{-1}\\=&\ (1,0)\begin{pmatrix}S_{nT,0}(z)S_{nT,2}(z)-S_{nT,1}^2(z)\\0\end{pmatrix}g(z)\\&\times\left(S_{nT,0}(z)S_{nT,2}(z)-S_{nT,1}^2(z)\right)^{-1}\\=&\ g(z).\end{aligned}\tag{A4}$$

By (A3), (A4), and some calculations, we have

$$\begin{aligned}
\sqrt{nh}(\hat{g}(z)-g(z)) \quad &\approx \quad \sqrt{nh}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1 g'(z)\mathbf{Z}_z + \frac{\sqrt{nh}}{2}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1 g''(z)\mathbf{Z}_z^2 + \sqrt{nh}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1\boldsymbol{V} \\
&\quad -\sqrt{nh}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}g(z)\boldsymbol{e}_{nT} - \sqrt{nh}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\
&=: \quad J_{11}+J_{12}+J_{13}-J_{14}-J_{15}.
\end{aligned} \tag{A5}$$

From the results of Lemmas A1–A4, it is easy to show that $J_{11}=o_p(1)$ and $J_{14}=o_p(1)$. Again, by invoking the results of Lemmas A1–A3 and $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|=O_p(n^{-1/2})$ in Theorem 1, we can prove that $J_{15}=o_p(1)$.

Now, we consider J_{12} and J_{13}. Let $\boldsymbol{M}(Z_{it},z)$ be a typical column of $\mathbf{M}(z)$, where $\mathbf{M}(z)=(\boldsymbol{M}(Z_{11},z),\cdots,\boldsymbol{M}(Z_{1T},z),\boldsymbol{M}(Z_{21},z),\cdots,\boldsymbol{M}(Z_{2T},z),\cdots,\boldsymbol{M}(Z_{n1},z),\cdots,\boldsymbol{M}(Z_{nT},z))$. For J_{12}, by Lemma A1 and some calculations, we can show that

$$\begin{aligned}
J_{12} &\approx \frac{\sqrt{nh}}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}(1,0)\boldsymbol{M}(Z_{it},z)g''(z)(Z_{it}-z)^2 \\
&= \frac{\sqrt{nh}}{2}\frac{1}{nf(z)}\sum_{i=1}^{n}\sum_{t=1}^{T}K_h(Z_{it}-z)g''(z)(Z_{it}-z)^2+o_p(1) \\
&= \frac{\sqrt{nh}}{2}\frac{1}{nf(z)}g''(z)\int z^2K(z)dz+o_p(h^2) \\
&= \frac{\sqrt{nh}}{2}b(z)+o_p(h^2).
\end{aligned} \tag{A6}$$

By Lemma A2 and Lemma A5, and using the same argument for J_{13}, as well as some simple calculations, we can show that

$$\begin{aligned}
J_{13} \quad &= \quad \sqrt{nh}\boldsymbol{m}^{\mathrm{T}}(z)\mathbf{Q}_1\boldsymbol{V} \\
&= \quad \sqrt{nh}\frac{1}{nf(z)}\sum_{i=1}^{n}\sum_{t=1}^{T}K_h(Z_{it}-z)\widetilde{V}_{it}+o_p(1) \\
&\xrightarrow{L} \quad N(0,\Sigma_g),
\end{aligned} \tag{A7}$$

where $\widetilde{V}_{it}=V_{it}-\frac{1}{T}\sum\limits_{s=1}^{T}V_{is}$ and $\Sigma_g=\nu_0\bar{\sigma}^2(z)f^{-2}(z)$.

By (A6) and (A6), it is easy to obtain

$$\begin{aligned}
\hat{g}(z)-g(z)-b(z) = \quad &\boldsymbol{m}^{\mathrm{T}}(z)[\mathbf{I}_{nT}-\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{P}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{P}]\boldsymbol{V}+o_p(1) \\
\approx \quad &\boldsymbol{m}^{\mathbf{T}}(z)\widetilde{\boldsymbol{V}}+o_p(1) \\
= \quad &(1,0)(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\widetilde{\boldsymbol{V}}+o_p(1) \\
=: \quad &I_1(z)+o_p(1),
\end{aligned} \tag{A8}$$

where $\widetilde{\boldsymbol{V}}=(\widetilde{V}_{11},\cdots,\widetilde{V}_{1T},\widetilde{V}_{21},\cdots,\widetilde{V}_{2T},\cdots,\widetilde{V}_{n1},\cdots,\widetilde{V}_{nT})^{\mathrm{T}}$ and $\widetilde{V}_{it}=V_{it}-\frac{1}{T}\sum_{s=1}^{T}V_{is}$.

Next, we approximate the process $I_1(z)$ as follows. Note that

$$\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z=\begin{pmatrix}\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}K_h(Z_{it}-z) & \sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}K_h(Z_{it}-z)(Z_{it}-z) \\ \sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}K_h(Z_{it}-z)(Z_{it}-z) & \sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}K_h(Z_{it}-z)(Z_{it}-z)^2\end{pmatrix}.$$

By Lemma A1, we have

$$n\mathbf{H}(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{H}=f^{-1}(z)\Omega^{-1}+O_p(h+(\log n/nh)^{1/2}), \tag{A9}$$

where $\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix}$ and $\Omega = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$.

By Lemma A1, we further obtain

$$\left\| \frac{1}{n}\mathbf{H}^{-1}\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\widetilde{\boldsymbol{V}} \right\|_\infty = O_p(h + (\log n/nh)^{1/2}). \tag{A10}$$

By (A9) and (A10), we have

$$\left\| I_1(z) - \frac{1}{nf(z)}(1,0)\Omega^{-1}\mathbf{H}^{-1}\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\widetilde{\boldsymbol{V}} \right\|_\infty = O_p\left(h(\log n/nh)^{1/2} + (\log n/nh)\right). \tag{A11}$$

Let

$$\begin{aligned} I_2(z) &=: \frac{1}{nf(z)}(1,0)\Omega^{-1}\mathbf{H}^{-1}\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\widetilde{\boldsymbol{V}} \\ &= \frac{1}{nf(z)}\sum_{i=1}^{n}\sum_{t=1}^{T}K_h(Z_{it}-z)\widetilde{V}_{it}. \end{aligned}$$

By invoking Theorem 1 and Lemma 1 in the work by Fan and Zhang (2000), for $h = n^{-\rho}$, where $1/5 \le \rho \le 1/3$, we have

$$P\left\{(-2\log h)^{1/2}\left(\left\|(nh\Sigma_g^{-1})^{1/2}I_2(z)\right\|_\infty - d_n\right) < u\right\} \longrightarrow \exp(-2\exp(-u)), \tag{A12}$$

where $\Sigma_g = \nu_0\bar{\sigma}^2(z)f^{-2}(z)$ is defined in Theorem 1 and d_n is defined in Theorem 2. By (A10)–(A12), we complete the proof of Theorem 2. □

Proof of Theorem 3. Similar to the proof of Theorem 2, it is easy to prove Theorem 3. Thus, we omit the details of the proof. □

Proof of Theorem 4. To prove Theorem 4, we need to derive the rate of convergence for the bias and variance estimators. First, we consider the difference between bias$(\hat{g}(z))$ and $b(z) = \frac{1}{2}h^2\mu_2 g''(z)$. By (A9) and similar arguments, we have

$$\left\| \widehat{\text{bias}}(\hat{g}(z)|\mathcal{D}) - b(z) \right\|_\infty = O_p(h^2\{\sqrt{\log n/nh_*^5}\}) = O_p\left(h^2(n^{-1/7}\log^{1/2} n)\right), \tag{A13}$$

where $h_* = O(n^{-1/7})$.

Furthermore, by Lemmas A1–A2 and a similar argument to (A10), we have

$$\left\| \frac{h}{n}\mathbf{H}^{-1}(\mathbf{Z}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{Q}_1\mathbf{W}_z\mathbf{Z}_z)\mathbf{H}^{-1} - f(z)\Lambda \right\|_\infty = o_p(1),$$

where $\Lambda = \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix}$. By a similar argument, it is easy to check that $\left\|\hat{\sigma}^2(z) - \sigma^2(z)\right\|_\infty = o_p(1)$. These results, together with Theorem 2, show that, uniformly for $z \in [0,1]$,

$$\left\| nh\widehat{\text{Var}}\{\hat{g}(z)|\mathcal{D}\} - \Sigma_g \right\|_\infty = o_p(1). \tag{A14}$$

By (A13) and (A14), and invoking the result of Theorem 2, we complete the proof of Theorem 4. □

References

1. Baltagi, B.H. *Econometrics Analysis of Panel Data*, 4th ed.; Wiley: New York, NY, USA, 2018.
2. Hsiao, C. *Analysis of Panel Data*; Cambridge University Press: Cambridge, UK, 2022.
3. Papke, L.E.; Wooldridge, J. A simple, robust test for choosing the level of fixed effects in linear panel data models. *Empir. Econ.* **2022**, *64*, 2683–2701. [CrossRef]

4. Henderson, D.J.; Carroll, R.J.; Li, Q. Nonparametric estimation and testing of fixed effects panel data models. *J. Econom.* **2008**, *144*, 257–275. [CrossRef]
5. Li, G.R.; Peng, H.; Tong, T. Simultaneous confidence band for nonparametric fixed effects panel data models. *Econ. Lett.* **2013**, *119*, 229–232. [CrossRef]
6. Feng, S.Y.; Li, G.R.; Tong, T.J.; Luo, S.H. Testing for heteroskedasticity in two-way fixed effects panel data models. *J. Appl. Stat.* **2020**, *47*, 91–116. [CrossRef]
7. Henderson, D.J.; Ullah, A. A nonparametric random effects estimator. *Econ. Lett.* **2005**, *88*, 403–407. [CrossRef]
8. Lin, D.Y.; Ying, Z. Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *J. Am. Stat. Assoc.* **2001**, *96*, 103–126. [CrossRef]
9. Li, Q.; Stengos, T. Semiparametric estimation of partially linear panel data models. *J. Econom.* **1996**, *71*, 289–397. [CrossRef]
10. Su, L.J.; Ullah, A. Profile likelihood estimation of partially linear panel data models with fixed effects. *Econ. Lett.* **2006**, *92*, 75–81. [CrossRef]
11. Zhang, J.H.; Feng, S.Y.; Li, G.R.; Lian, H. Empirical likelihood inference for partially linear panel data models with fixed effects. *Econ. Lett.* **2011**, *113*, 165–167. [CrossRef]
12. Sun, Y.; Carroll, R.J.; Li, D. Semiparametric estimation of fixed effects panel data varying coefficient models. *Adv. Econom.* **2009**, *25*, 101–129.
13. Feng, S.Y.; Li, G.R.; Peng, H.; Tong, T.J. Varying-coefficient panel data model with interactive fixed effects. *Stat. Sin.* **2021**, *31*, 935–957. [CrossRef]
14. Chen, J.; Gao, J.; Li, D. Estimation in a single-index panel data models with heterogeneous link functions. *Econom. Rev.* **2013**, *33*, 928–955. [CrossRef]
15. Chen, J.; Gao, J.; Li, D. Estimation in partially linear single-index panel data models with fixed effects. *J. Bus. Econ. Stat.* **2013**, *31*, 1–42. [CrossRef]
16. Lai, P.; Li, G.R.; Lian, H. Semiparametric estimation of fixed effects panel data single-index model. *Stat. Probab. Lett.* **2013**, *83*, 1595–1602. [CrossRef]
17. Fitzmaurice, G.M.; Laird, N.M.; Ware, J.H. *Applied Longitudinal Analysis*; Wiley: New York, NY, USA, 2004.
18. Baltagi, B.H.; Li, D. Series estimation of partially linear panel data models with fixed effects. *Ann. Econ. Financ.* **2002**, *3*, 103–116.
19. Fan, J.Q.; Zhang, W.Y. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **2000**, *27*, 715–731. [CrossRef]
20. Zhang, W.Y.; Peng, H. Simultaneous confidence band and hypothesis test in generalized varying-coefficient models. *J. Multivar. Anal.* **2010**, *101*, 1656–1680. [CrossRef]
21. Li, G.R.; Peng, H.; Dong, K.; Tong, T. Simultaneous confidence bands and hypothesis testing in single-index models. *Stat. Sin.* **2014**, *24*, 937–955. [CrossRef]
22. Yothers, G.; Sampson, A.R. Simultaneous confidence band for the difference of segmented linear models. *J. Stat. Plan. Inference* **2011**, *141*, 1059–1068. [CrossRef]
23. Liu, W.; Ah-Kine, P.; Bretz, F.; Hayter, A.J. Exact simultaneous confidence intervals for a finite set of contrasts of three, four or five generally correlated normal means. *Comput. Stat. Data Anal.* **2013**, *57*, 141–148. [CrossRef]
24. Yang, S.; Xue, L.; Li, G. Simultaneous confidence band for single-index random effects models with longitudinal data. *Stat. Probab. Lett.* **2014**, *85*, 6–14. [CrossRef]
25. Bissantz, N.; Dümbgen, L.; Holzmann, H.; Munk, A. Non-parametric confidence bands in deconvolution density estimation. *Tech. Rep.* **2007**, *69*, 483–506. [CrossRef]
26. Vos, I.D.; Stauskas, O. Cross-section bootstrap for cce regressions. *J. Econom.* **2024**, *240*, 6–14.
27. Fan, J.; Gijbels, I. *Local Polynomial Modeling and Its Applications*; Chapman and Hall: London, UK, 1996.