Statistica Sinica Preprint No: SS-2018-0248	
Title	Varying Coefficient Panel Data Model with Interactive Fixed Effects
Manuscript ID	SS-2018-0248
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0248
Complete List of Authors	Sanying Feng Gaorong Li Heng Peng and Tiejun Tong
Corresponding Author	Gaorong Li
E-mail	ligaorong@gmail.com
Notice: Accepted version subject to English editing.	

VARYING COEFFICIENT PANEL DATA MODEL WITH INTERACTIVE FIXED EFFECTS

Sanying Feng[1], Gaorong Li[2], Heng Peng[3] and Tiejun Tong[3]

[1] *Zhengzhou University,* [2] *Beijing Normal University,*

[3] *Hong Kong Baptist University*

Abstract: In this paper, we propose a varying coefficient panel data model with unobservable multiple interactive fixed effects that are correlated with the regressors. We approximate each coefficient function by B-splines, and propose a robust nonlinear iteration scheme based on the least squares method to estimate the coefficient functions of interest. We also establish the asymptotic theory of the resulting estimators under certain regularity assumptions, including the consistency, the convergence rate and the asymptotic distributions. To construct the pointwise confidence intervals for the coefficient functions, a residual-based block bootstrap method is proposed that reduces the computational burden and avoids the accumulative errors. We further extend our proposed procedure to the partially linear varying coefficient panel data model with unobservable multiple interactive fixed effects, and study the testing problem about constant coefficients versus function coefficients. Simulation studies and a real data analysis are also carried out to assess the performance of our proposed methods.

Key words and phrases: Bootstrap, B-spline, hypothesis testing, interactive fixed effect, panel data, partially linear varying coefficient model, varying coefficient model.

1. Introduction

Panel data models typically incorporate individual and time effects to control the heterogeneity in the cross-section and across the time-periods. Panel data analysis has attracted considerable attention in the literature. The methodology for parametric panel data analysis is quite mature, see, for example, Arellano (2003), Hsiao (2003), Baltagi (2005) and the references therein. The individual and time effects may enter the model additively, or they can interact multiplicatively that leads to the so-called interactive effects or a factor structure. Panel data models with interactive fixed effects are useful modelling paradigm. In macroeconomics, incorporating the interactive effects can account for the heterogenous impact of unobservable common shocks, while the regressors can be input such as labor and capital. Panel data models with interactive fixed effects are used to incorporate unmeasured skills or unobservable characteristics, or to study the individual wage rate (Su and Chen (2013)). In finance, a combination of unobserved factors and observed covariates can explain the excess returns of assets. Bai

(2009) considered the linear panel data model with interactive fixed effects:

$$Y_{it} = X_{it}^{\tau}\boldsymbol{\beta} + \lambda_i^{\tau} F_t + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1.1}$$

where X_{it} is a $p \times 1$ vector of observable regressors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, λ_i is an $r \times 1$ vector of factor loadings, F_t is an $r \times 1$ vector of common factors so that $\lambda_i^{\tau} F_t = \lambda_{i1} F_{1t} + \cdots + \lambda_{ir} F_{rt}$, and ε_{it} are idiosyncratic errors. In this model, λ_i, F_t and ε_{it} are all unobserved, and the dimension r of the factor loadings does not depend on the cross section size N or the time series length T.

A number of researchers have developed statistical methods to study panel data models with interactive fixed effects. For example, Holtz-Eakin et al. (1988) estimated model (1.1) by quasi-differencing and using lagged variables as instruments. Their approach, however, ruled out time constant regressors. Coakley et al. (2002) studied model (1.1) by augmenting the regression of Y on X with the principal components of the ordinary least squares residuals. Pesaran (2006) showed that the method of Coakley et al. (2002) is inconsistent unless the correlation between X_{it} and λ_i tends to be uncorrelated or fully correlated as N tends to infinity. As an alternative, Pesaran (2006) developed a correlated common effects (CCE) estimator, in which model (1.1) is augmented by the cross-sectional averages of X_{it}. Although Pesaran's estimator is consistent, it does not allow for time-invariant individual regressors. Ahn et al. (2001) developed a generalized method of moments (GMM) estimator for model (1.1). Their estimator is more efficient than the least squares estimator under a fixed T. However, the identification of their estimator requires that X_{it} is correlated with λ_i, and it is impossible to make testing for the interactive random effects assumption. Bai (2009) studied the identification, consistency, and limiting distribution of the principal component analysis (PCA) estimators and demonstrated that these estimators are $\sqrt{NT}$ consistent. Bai and Li (2014) investigated the maximum likelihood estimation of model (1.1). Wu and Li (2014) conducted several tests for the existence of individual effects and time effects of model (1.1). Li et al. (2016) studied the estimation and inference of common structural breaks in panel data models with interactive fixed effects using Lasso-type methods. More studies can be found in Moon and Weidner (2017), Lee et al. (2012), Su and Chen (2013), Moon and Weidner (2015), Lu and Su (2016), and many others.

Note that the aforementioned papers have focused on the linear specification of regression relationship in panel data models with interactive fixed effects. A natural extension of model (1.1) is to consider the varying coefficient panel data model with interactive fixed effects:

$$Y_{it} = X_{it}^{\tau}\boldsymbol{\beta}(U_{it}) + \lambda_i^{\tau} F_t + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1.2}$$

where $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_p(\cdot))^\tau$ is a $p \times 1$ vector of unknown coefficient functions to be estimated. We allow for $\{X_{it}\}$ and/or $\{U_{it}\}$ to be correlated with $\{\lambda_i\}$ alone or with $\{F_t\}$ alone, or simultaneously correlated with $\{\lambda_i\}$ and $\{F_t\}$, or correlated with an unknown correlation structure. In fact, X_{it} can be a nonlinear function of λ_i and F_t. Hence, model (1.2) is a fixed effects model, and assumes an interactive fixed effects linear model for each fixed time t but allows the coefficients to vary with the covariate U_{it}. This model is attractive because it has an intuitive interpretation, meanwhile it retains the unobservable multiple interactive fixed effects, the general nonparametric characteristics, and the explanatory power of the linear panel data model.

Model (1.2) is fairly general and it encompasses various panel data models as special cases. If $X_{it} \equiv 1$ and $p = 1$, model (1.2) reduces to the nonparametric panel data model with interactive fixed effects, which has received much attention in recent years. Huang (2013) studied the local linear estimation of nonparametric panel data models with interactive fixed effects. Su and Jin (2012) extended the CCE method of Pesaran (2006) from a linear model to a nonparametric model via the method of sieves. Jin and Su (2013) constructed a nonparametric test for poolability in nonparametric regression models with interactive fixed effects. Su et al. (2015) proposed a consistent nonparametric test for the linearity in large dimensional panel data model with interactive fixed effects.

If $r = 1$ and $F_t \equiv 1$, model (1.2) reduces to the fixed individual effects panel data varying coefficient model:

$$Y_{it} = X_{it}^{\tau}\boldsymbol{\beta}(U_{it}) + \lambda_i + \varepsilon_{it}.$$

This model has also been widely studied in the literature. For example, Sun et al. (2009) considered the estimation using the local linear regression and the kernel-based weights. Li et al. (2011) considered a nonparametric time varying coefficient model with fixed effects under the assumption of cross-sectional independence, and proposed two methods to estimate the trend function and the coefficient functions. Rodriguez-Poo and Soberon (2014) proposed a new technique to estimate the varying coefficient functions based on the first-order differences and the local linear regression. Rodriguez-Poo and Soberon (2015) investigated the model by using the mean transformation technique and the local linear regression. Li et al. (2015) considered the variable selection for the model using the basis function approximations and the group nonconcave penalized functions. Malikov et al. (2016) considered the problem of varying coefficient panel data model in the presence of endogenous selectivity and fixed effects. In addition, if $\lambda_i \equiv 0$ or $F_t \equiv 0$, model (1.2) reduces to the varying coefficient model with panel data. For the development of this model, one may refer to Chiang et al. (2001), Huang et al. (2002), Huang et al. (2004), Xue and Zhu (2007), Cai (2007), Cai and Li (2008), Wang et al. (2008), Wang and Xia (2009) and Noh and Park (2010). We note, however, that most of these papers were dealing with the "large N small T" setting.

Despite the rich literature in panel data models with interactive fixed effects, to the best of our knowledge, there is little work on the varying coefficient panel data models with interactive fixed effects. Inspired by this, the main goals of this paper are to estimate the coefficient functions $\boldsymbol{\beta}(\cdot)$ and to establish the asymptotic theory for the varying coefficient panel data models with interactive fixed effects when both N and T tend to infinity and serial or cross-sectional correlations and heteroskedasticities of unknown form in ε_{it}. To achieve these goals, we first apply the B-spline expansion to estimate the smooth functions in model (1.2) due to its simplicity. We then introduce a novel iterative least squares procedure to estimate the coefficient functions and the factor loadings, and derive some asymptotic properties for the proposed estimators. Nevertheless, the existence of the unobservable interactive fixed effects and the weak correlations and heteroskedasticities of unknown form in both dimensions will make the estimation procedure and the asymptotic theory much more complicated than those in Huang et al. (2002). To apply the asymptotic normality for constructing the pointwise confidence intervals for the coefficient functions, we need some consistent estimators of the asymptotic biases and variances. To reduce the computational burden and to avoid the accumulative errors, we propose a residual-based block bootstrap procedure to construct the pointwise confidence intervals of the coefficient functions.

Moreover, we extend the proposed estimation procedure to the partially linear varying coefficient model with interactive fixed effects and show that the convergence rate for the estimation of the parametric components is order of $O_P((NT)^{-1/2})$. To determine whether a varying coefficient model or partially linear varying coefficient model is appropriate, we propose a test statistic to test between the two alternatives in practice. Numerical results confirm that our proposed estimation and testing procedures work well in a wide range of settings.

The remainder of the paper is organized as follows. In Section 2, we propose an estimation procedure for the coefficient functions and provide a robust iteration algorithm under the identification restrictions. In Section 3, we establish the asymptotic theory of the resulting estimators under some regularity assumptions as both N and T tend to infinity. In Section 4, a residual-based block bootstrap procedure is developed to construct the pointwise confidence intervals for the coefficient functions. In Section 5, we extend the estimation procedure to partially linear varying coefficient model and establish the asymptotic distribution of the estimator. In Section 6, a test statistic and the bootstrap procedure are developed. Finally, we conclude the paper in Section 7 with some remarks. To shorten the length of the article, the technical details are given in the supplementary material, and the simulation studies and a real application are carried out to demonstrate the efficacy of our proposed methods in the supplementary material.

2. Methodology

To estimate the coefficient functions $\beta_k(\cdot)$ for $1 \le k \le p$, we consider the widely used B-spline approximations. Let $B_k(u) = (B_{k1}(u), \ldots, B_{kL_k}(u))^\tau$ be the $(m+1)$th order B-spline basis functions, where $L_k = l_k + m + 1$ is the number of basis functions in approximating $\beta_k(u)$, l_k is the number of interior knots for $\beta_k(\cdot)$, and m is the degree of the spline. The interior knots of the splines can be either equally spaced or placed on the sample number of observations between any two adjacent knots. With the above basis functions, the coefficient functions $\beta_k(u)$ can be approximated by

$$\beta_k(u) \approx \sum_{l=1}^{L_k} \gamma_{kl} B_{kl}(u), \quad k = 1, \ldots, p, \tag{2.1}$$

where γ_{kl} are the coefficients, and L_k represent the smoothing parameters and they will be selected by the "leave-one-subject-out" cross validation.

Substituting (2.1) into model (1.2), we have the following approximation:

$$Y_{it} \approx \sum_{k=1}^{p} \sum_{l=1}^{L_k} \gamma_{kl} X_{it,k} B_{kl}(U_{it}) + \lambda_i^{\tau} F_t + \varepsilon_{it}, \; i = 1, \ldots, N, \; t = 1, \ldots, T. \tag{2.2}$$

Model (2.2) is a standard linear regression model with the interactive fixed effects. As each coefficient function $\beta_k(u)$ in model (1.2) is characterized by $\boldsymbol{\gamma}_k = (\gamma_{k1}, \ldots, \gamma_{kL_k})^{\tau}$, model (2.2) cannot be estimated directly due to the unobservable multiple interactive fixed effects. In what follows, we propose a robust nonlinear iteration scheme based on the least squares method to estimate the coefficient functions and deal with those fixed effects.

For the sake of convenience, we use vectors and matrices to present the model and perform the analysis. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT})^{\tau}$, $\boldsymbol{F} = (F_1, \ldots, F_T)^{\tau}$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^{\tau}$, and $\Lambda = (\lambda_1, \ldots, \lambda_N)^{\tau}$ be an $N \times r$ matrix. Let

$$\boldsymbol{B}(u) = \begin{pmatrix} B_{11}(u) & \cdots & B_{1L_1}(u) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ & \vdots & & & \vdots & & & \vdots & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & B_{p1}(u) & \cdots & B_{pL_p}(u) \end{pmatrix},$$

$R_{it} = (X_{it}^{\tau}\boldsymbol{B}(U_{it}))^{\tau}$, and $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{iT})^{\tau}$. Let also $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\tau}, \ldots, \boldsymbol{\gamma}_p^{\tau})^{\tau}$, where $\boldsymbol{\gamma}_k = (\gamma_{k1}, \ldots, \gamma_{kL_k})^{\tau}$. Then model (2.2) can be rewritten as

$$\boldsymbol{Y}_i \approx \boldsymbol{R}_i\boldsymbol{\gamma} + \boldsymbol{F}\lambda_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, N.$$

Due to potential correlations between the unobservable effects and the regressors, we treat F_t and λ_i as the fixed effects parameters to be estimated. To ensure the identifiability of the coefficient function $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_p(\cdot))^{\tau}$, we follow Bai (2009) and impose the following identification restrictions:

$$\boldsymbol{F}^{\tau}\boldsymbol{F}/T = I_r \quad \text{and} \quad \Lambda^{\tau}\Lambda = \text{diagonal}. \tag{2.3}$$

These two restrictions can uniquely determine Λ and $\boldsymbol{F}$. We then define the objective function as

$$Q(\boldsymbol{\gamma}, \boldsymbol{F}, \Lambda) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma} - \boldsymbol{F}\lambda_i)^{\tau}(\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma} - \boldsymbol{F}\lambda_i) \tag{2.4}$$

subject to the constraint (2.3). Taking partial derivatives of (2.4) with respect to λ_i and setting them equal to zero, we have

$$\widetilde{\lambda}_i = (\boldsymbol{F}^{\tau}\boldsymbol{F})^{-1}\boldsymbol{F}^{\tau}(\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma}) = T^{-1}\boldsymbol{F}^{\tau}(\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma}). \tag{2.5}$$

Replacing λ_i into (2.4) by (2.5), we have

$$\begin{aligned} Q(\boldsymbol{\gamma}, \boldsymbol{F}) &= \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma} - \boldsymbol{F}\widetilde{\lambda}_i)^{\tau}(\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma} - \boldsymbol{F}\widetilde{\lambda}_i) \\ &= \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma})^{\tau} M_{\boldsymbol{F}} (\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma}), \end{aligned}$$

where $\widetilde{\Lambda} = (\widetilde{\lambda}_1, \widetilde{\lambda}_2, \ldots, \widetilde{\lambda}_N)^\tau$, and $M_{\boldsymbol{F}} = I_T - \boldsymbol{F}(\boldsymbol{F}^\tau \boldsymbol{F})^{-1}\boldsymbol{F}^\tau = I_T - \boldsymbol{F}\boldsymbol{F}^\tau/T$ is a projection matrix. For each given $\boldsymbol{F}$, if $\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{R}_i$ is invertible, the least squares estimator of $\boldsymbol{\gamma}$ can be uniquely obtained by minimizing $Q(\boldsymbol{\gamma}, \boldsymbol{F})$ as follows:

$$\hat{\gamma}(\boldsymbol{F}) = \left(\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{R}_i\right)^{-1} \sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{Y}_i. \tag{2.6}$$

Since the least squares estimator (2.6) of $\boldsymbol{\gamma}$ depends on the unknown common factors $\boldsymbol{F}$, the final solution of $\boldsymbol{\gamma}$ can be obtained by iteration between $\boldsymbol{\gamma}$ and $\boldsymbol{F}$ using the following nonlinear equations:

$$\hat{\gamma} = \left(\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{R}_i\right)^{-1} \sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{Y}_i, \tag{2.7}$$

$$\hat{\boldsymbol{F}} V_{NT} = \left[\frac{1}{NT}\sum_{i=1}^N (\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\gamma})(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\gamma})^\tau\right] \hat{\boldsymbol{F}}, \tag{2.8}$$

where V_{NT} is a diagonal matrix consisting of the r largest eigenvalues of the matrix $(NT)^{-1}\sum_{i=1}^N (\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\gamma})(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\gamma})^\tau$ arranged in decreasing order. As noted by Bai (2009), the iterated solution is somewhat sensitive to the initial values. Bai (2009) proposed using either the least squares estimator of $\boldsymbol{\gamma}$ or the principal components estimate of $\boldsymbol{F}$ to start with. From the numerical studies in the supplementary material, we find that the procedure is more robust when the principal components estimator of $\boldsymbol{F}$ is used as the initial values. Generally, the poor initial values will result in an exceptionally large number of iterations. By (2.5), (2.7) and (2.8), we have

$$\hat{\Lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^\tau = T^{-1}\Big(\hat{\boldsymbol{F}}^\tau(\boldsymbol{Y}_1 - \boldsymbol{R}_1\hat{\boldsymbol{\gamma}}), \ldots, \hat{\boldsymbol{F}}^\tau(\boldsymbol{Y}_N - \boldsymbol{R}_N\hat{\boldsymbol{\gamma}})\Big)^\tau. \quad (2.9)$$

Once we obtain the estimator $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1^\tau, \ldots, \hat{\boldsymbol{\gamma}}_p^\tau)^\tau$ of $\boldsymbol{\gamma}$ with $\hat{\boldsymbol{\gamma}}_k = (\hat{\gamma}_{k1}, \ldots, \hat{\gamma}_{kL_k})^\tau$ for $k = 1, \ldots, p$, we can estimate $\beta_k(u)$ subsequently by

$$\hat{\beta}_k(u) = \sum_{l=1}^{L_k} \hat{\gamma}_{kl} B_{kl}(u), \quad k = 1, \ldots, p.$$

In what follows, we present a robust iteration algorithm for estimating the parameters $(\boldsymbol{\gamma}, \boldsymbol{F}, \Lambda)$.

Step 1. Obtain an initial estimator $(\hat{\boldsymbol{F}}, \hat{\Lambda})$ of $(\boldsymbol{F}, \Lambda)$.

Step 2. Given $\hat{\boldsymbol{F}}$ and $\hat{\Lambda}$, compute $\hat{\gamma}(\hat{\boldsymbol{F}}, \hat{\Lambda}) = \Big(\sum_{i=1}^N \boldsymbol{R}_i^\tau \boldsymbol{R}_i\Big)^{-1} \sum_{i=1}^N \boldsymbol{R}_i^\tau(\boldsymbol{Y}_i - \hat{\boldsymbol{F}}\hat{\lambda}_i)$.

Step 3. Given $\hat{\gamma}$, compute $\hat{\boldsymbol{F}}$ according to (2.8) (multiplied by $\sqrt{T}$ due to the restriction that $\boldsymbol{F}^\tau\boldsymbol{F}/T = I_r$) and calculate $\hat{\Lambda}$ using formula (2.9).

Step 4. Repeat Steps 2 and 3 until $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{F}}, \hat{\Lambda})$ satisfy the given convergence criterion.

3. Regularity assumptions and asymptotic properties

To derive some asymptotic properties of the proposed estimators, we let $\mathcal{F} \equiv \{\boldsymbol{F} : \boldsymbol{F}^\tau \boldsymbol{F}/T = I_r\}$ and

$$D(\boldsymbol{F}) = \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{R}_i - \frac{1}{T}\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{R}_j a_{ij}\Big],$$

where $a_{ij} = \lambda_i^\tau(\Lambda^\tau\Lambda/N)^{-1}\lambda_j$. To obtain the unique estimator of $\boldsymbol{\gamma}$ with probability tending to one, we require that the first term of $D(\boldsymbol{F})$ on the right-hand side is positive definite when $\boldsymbol{F}$ is observable. The presence of the second term is because of the unobservable $\boldsymbol{F}$ and Λ. The reason for this particular form is the nonlinearity of the interactive effects (see details in Bai (2009)).

3.1 Regularity assumptions

In this section, we introduce a definition and present some regularity assumptions for establishing the asymptotic theory of the resulting estimators.

Definition 1. Let $\mathcal{H}_d$ define the collection of all functions on the support $\mathcal{U}$ whose mth order derivative satisfies the Hölder condition of order ν with $d \equiv m + \nu$, where $0 < \nu \leq 1$. That is, for each $h \in \mathcal{H}_d$, there exists a constant $M_0 \in (0, \infty)$ such that $|h^{(m)}(u) - h^{(m)}(v)| \leq M_0|u - v|^\nu$, for any $u, v \in \mathcal{U}$.

(A1) The random variable X_{it} is i.i.d. cross the N individuals, and there exists a positive M such that $|X_{it,k}| \leq M < \infty$ for all $k = 1, \ldots, p$. We further assume that $\{X_{it} : 1 \leq t \leq T\}$ is strictly stationarity for each i. The eigenvalues $\rho_1(u) \leq \cdots \leq \rho_p(u)$ of $\Omega(u) = E(X_{it}X_{it}^{\tau}|U_{it} = u)$ are bounded away from 0 and ∞ uniformly over $u \in \mathcal{U}$, that is, there exist positive constants ρ_0 and ρ^* such that $0 < \rho_0 \leq \rho_1(u) \leq \cdots \leq \rho_p(u) \leq \rho^* < \infty$ for $u \in \mathcal{U}$.

(A2) The observation variables U_{it} are chosen independently according to a distribution F_U on the support $\mathcal{U}$. Moreover, the density function of U, $f_U(u)$, is uniformly bounded away from 0 and ∞, and continuously differentiable uniformly over $u \in \mathcal{U}$.

(A3) $\beta_k(u) \in \mathcal{H}_d$ for all $k = 1, \ldots, p$.

(A4) Let $u_{k1}, \ldots, u_{kl_k}$ be the interior knots of the kth coefficient function over $u \in \mathcal{U} = [U_0, U_1]$ for $k = 1, \ldots, p$. Furthermore, let $u_{k0} = U_0$ and $u_{k(l_k+1)} = U_1$. There exists a positive constant C_0 such that

$$\frac{h_k}{\min_{1\leq i\leq l_k} h_{ki}} \leq C_0 \quad \text{and} \quad \frac{\max_{1\leq k\leq p} h_{ki}}{\min_{1\leq k\leq p} h_{ki}} \leq C_0,$$

where $h_{ki} = u_{ki} - u_{k(i-1)}$ and $h_k = \max_{1\leq i\leq l_k+1} h_{ki}$.

(A5) Suppose that $\inf_{\boldsymbol{F}\in\mathcal{F}} D(\boldsymbol{F}) > 0$.

(A6) $E\|F_t\|^4 \leq M$ and $\sum_{t=1}^{T} F_t F_t^{\tau}/T \xrightarrow{P} \Sigma_F > 0$ for some $r \times r$ matrix Σ_F, as $T \to \infty$, where "$\xrightarrow{P}$" denotes the convergence in probability.

(A7) $E\|\lambda_i\|^4 \leq M$ and $\Lambda^{\tau}\Lambda/N \xrightarrow{P} \Sigma_{\Lambda} > 0$ for some $r \times r$ matrix Σ_{Λ}, as $N \to \infty$.

(A8) (i) Suppose that ε_{it} are independent of X_{js}, U_{js}, λ_j and F_s for all i, t, j and s with zero mean and $E(\varepsilon_{it})^8 \leq M$.

(ii) Let $\sigma_{ij,ts} = E(\varepsilon_{it}\varepsilon_{js})$. $|\sigma_{ij,ts}| \leq \rho_{ij}$ for all (t, s) and $|\sigma_{ij,ts}| \leq \varrho_{ts}$ for all (i, j) such that

$$\frac{1}{N}\sum_{i,j=1}^{N} \rho_{ij} \leq M, \quad \frac{1}{T}\sum_{t,s=1}^{T} \varrho_{ts} \leq M, \quad \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} |\sigma_{ij,ts}| \leq M.$$

The smallest and largest eigenvalues of $\Omega_i = E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{\tau})$ are bounded uniformly for all i and t, where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^{\tau}$.

(iii) For every (t, s), $E\left|N^{-1/2}\sum_{i=1}^{N}[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})]\right|^4 \leq M$.

(iv) Moreover, we assume that $T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j} |\mathrm{cov}(\varepsilon_{it}\varepsilon_{is}, \varepsilon_{ju}\varepsilon_{jv})| \leq M$ and $T^{-1}N^{-2}\sum_{t,s}\sum_{i,j,m,l} |\mathrm{cov}(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ms}\varepsilon_{ls})| \leq M$.

(A9) $\limsup_{N,T}(\max_k L_k / \min_k L_k) < \infty$.

Assumptions (A1)–(A4) are mild conditions that can be validated in many practical situations. These conditions have been widely assumed in the context of varying coefficient models with repeated measurements, such as Huang et al. (2002), Huang et al. (2004) and Wang et al. (2008). Assumption (A5) is an identification condition for $\boldsymbol{\gamma}$, and $\boldsymbol{\gamma}$ can be uniquely determined by (2.7) if $D(\boldsymbol{F})$ is positive definite. Assumptions (A6) and (A7) imply the existence of r factors. In this paper, whether F_t or λ_i has zero mean is not crucial since they are treated as parameters to be estimated. Assumption (A8) allows for weak forms of both cross sectional dependence and serial dependence in the error processes. Assumption (A9) can also be found in Noh and Park (2010), and this condition is used for the system of general basis functions B_{kl} including orthonormal bases, non-orthonormal bases and B-splines.

Let $\|a\|_{L_2} = \{\int_{\mathcal{U}} a^2(u)\mathrm{d}u\}^{1/2}$ be the L_2 norm of any square integrable real-valued function $a(u)$ on $\mathcal{U}$, and let $\|A\|_{L_2} = \{\sum_{k=1}^{p} \|a\|_{L_2}^2\}^{1/2}$ be the L_2 norm of $A(u) = (a_1(u), \ldots, a_p(u))^\tau$, where $a_k(u)$ are real-valued functions on $\mathcal{U}$ (see details in Huang et al. (2002)). We define $\hat{\beta}_k(\cdot)$ to be a consistent estimator of $\beta_k(\cdot)$ if $\lim_{N,T\to\infty} \|\hat{\beta}_k(\cdot) - \beta_k(\cdot)\|_{L_2} = 0$ holds in probability. Define $\delta_{NT} = \min[\sqrt{N}, \sqrt{T}]$, and $L_N = \max_{1\le k\le p} L_k$, which tend to infinity as N or T tends to infinity. Let $\mathcal{D} = \{(X_{it}, U_{it}, \lambda_i, F_t), i = 1, \ldots, N, t = 1, \ldots, T\}$. We use $E_{\mathcal{D}}$ and $\mathrm{Var}_{\mathcal{D}}$ to denote the expectation and variance conditional on $\mathcal{D}$, respectively.

3.2 Asymptotic properties

Let $\boldsymbol{F}^0$ be the true value of $\boldsymbol{F}$. With an appropriate choice of L_k to balance the bias and variance, our proposed estimators have the asymptotic properties including the consistency, the convergence rate and the asymptotic distribution.

Theorem 1. *Suppose that assumptions (A1)–(A9) hold. If* $\delta_{NT}^{-2} L_N \log L_N \to 0$ *as* $N \to \infty$ *and* $T \to \infty$ *simultaneously, then*

(i) $\hat{\beta}_k(\cdot), k = 1, \ldots, p$, *are uniquely defined with probability tending to one.*

(ii) *The matrix* $\boldsymbol{F}^{0\tau} \hat{\boldsymbol{F}}/T$ *is invertible and* $\|P_{\hat{\boldsymbol{F}}} - P_{\boldsymbol{F}^0}\| \xrightarrow{P} 0$, *where* $P_A = A(A^\tau A)^{-1} A^\tau$ *for a given matrix* A.

Part (i) of Theorem 1 implies that, with probability tending to one, we can obtain the unique estimators $\hat{\beta}_k(\cdot)$ for the unknown coefficient functions $\beta_k(\cdot)$ under some regularity assumptions, no matter whether there exist unobservable multiple interactive fixed effects in model (1.2). Part (ii) of Theorem 1 indicates that the spaces spanned by $\hat{\boldsymbol{F}}$ and $\boldsymbol{F}^0$ are asymptotically consistent. This is a key result to guarantee that the estimators $\hat{\beta}_k(\cdot)$ have good asymptotic properties including the optimal convergence rate, consistency and asymptotic normality.

Theorem 2. *Under the assumptions of Theorem 1, we further have*

$$\|\hat{\beta}_k(u) - \beta_k(u)\|_{L_2}^2 = O_P\left(\frac{L_N}{NT} + \frac{L_N}{T^2} + \frac{L_N}{N^2} + L_N^{-2d}\right), \quad k = 1, \ldots, p.$$

Theorem 2 gives the convergence rate of $\hat{\beta}_k(u)$ for all $k = 1, \ldots, p$, and hence establishes the consistency of our proposed estimators under the condition $\delta_{NT}^{-2} L_N \log L_N \to 0$ as $N \to \infty$ and $T \to \infty$ simultaneously. From the proof of Theorem 2, we note that the first term in the convergence rate is caused by the stochastic error, the second and third terms are caused by the estimation error of the fixed effects $\boldsymbol{F}^0$ and the presence of the cross-sectional and serial correlation and heteroskedasticity, and the last term is the error due to the basis approximation. If we take the appropriate relative rate $T/N \to c > 0$ as $N \to \infty$ and $T \to \infty$ simultaneously, then we have a more accurate convergence rate as follows

$$\|\hat{\beta}_k(u) - \beta_k(u)\|_{L_2}^2 = O_P\left(\frac{L_N}{NT} + L_N^{-2d}\right), \quad k = 1, \ldots, p.$$

Furthermore, if we take $L_N = O((NT)^{1/(2d+1)})$, then

$$\|\hat{\beta}_k(u) - \beta_k(u)\|_{L_2}^2 = O_P\left((NT)^{-2d/(2d+1)}\right), \quad k = 1, \ldots, p.$$

This leads to the optimal convergence rate of order $O_P\left((NT)^{-2d/(2d+1)}\right)$ that holds for i.i.d. data in Stone (1982).

Next, we establish the asymptotic distribution of $\hat{\boldsymbol{\beta}}(u)$. Let $\boldsymbol{Z}_i = M_{\boldsymbol{F}^0}\boldsymbol{R}_i - N^{-1}\sum_{j=1}^N a_{ij} M_{\boldsymbol{F}^0}\boldsymbol{R}_j$. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}(u)$ conditioning on $\mathcal{D}$ is $\Sigma = \mathrm{Var}(\hat{\boldsymbol{\beta}}(u)|\mathcal{D}) = \boldsymbol{B}(u)\Phi\boldsymbol{B}(u)^\tau$, where Φ is the limit in probability of

$$\Phi^* = \left(\sum_{i=1}^{N} \boldsymbol{Z}_i^\tau \boldsymbol{Z}_i\right)^{-1} \Sigma_{NT1} \left(\sum_{i=1}^{N} \boldsymbol{Z}_i^\tau \boldsymbol{Z}_i\right)^{-1}$$

with $\Sigma_{NT1} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{ij,ts}Z_{it}Z_{js}^\tau$. Let ϖ_k denote the unit vector in $\mathbb{R}^p$ with 1 in the kth coordinate and 0 in all other coordinates for $k = 1,\ldots,p$. Then the conditional variance of $\hat{\beta}_k(u)$ is

$$\Sigma_{kk} = \mathrm{Var}(\hat{\beta}_k(u)|\mathcal{D}) = \varpi_k^\tau \Sigma \varpi_k, \quad k = 1,\ldots,p.$$

To study the asymptotic distribution of $\hat{\boldsymbol{\beta}}(u)$, we add the following assumption.

(A10) Let Σ_1 be the limit in probability of $\frac{1}{NT}\Sigma_{NT1}$, then $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{Z}_i^\tau\boldsymbol{\varepsilon}_i \xrightarrow{L} N(\boldsymbol{0},\Sigma_1)$, where "$\xrightarrow{L}$" denotes the convergence in distribution.

Denote $\tilde{\Sigma} = D_0^{-1}\Sigma_1 D_0^{-1}$, where $D_0 = \mathrm{plim}\frac{L_N}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i^\tau\boldsymbol{Z}_i$. The following theorem establishes the asymptotic distribution of $\hat{\boldsymbol{\beta}}(u)$.

Theorem 3. *Suppose that assumptions (A1)-(A10) hold. If* $\delta_{NT}^{-2}L_N\log L_N \to 0$, $L_N^{2d+1}/NT \to \infty$ *and* $T/N \to c$ *as* $N \to \infty$ *and* $T \to \infty$ *simultaneously, then*

$$\Sigma^{-1/2}(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)) \xrightarrow{L} N(\boldsymbol{b}(u), I_p),$$

where $\boldsymbol{b}(u) = \tilde{\Sigma}^{-1/2}c^{1/2}W_1^0 + \tilde{\Sigma}^{-1/2}c^{-1/2}W_2^0$*, and* W_1^0 *is the limit in probability of* W_1 *with*

$$W_1 = -\boldsymbol{B}(u)\left(L_N D(\boldsymbol{F}^0)\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\boldsymbol{R}_i - \boldsymbol{V}_i)^\tau \boldsymbol{F}^0}{T}\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1} \times \left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\sigma_{ij,tt}\right)$$

and W_2^0 *is the limit in probability of* W_2 *with*

$$W_2 = -\boldsymbol{B}(u)\left(L_N D(\boldsymbol{F}^0)\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau M_{\boldsymbol{F}^0}\Omega\boldsymbol{F}^0\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_i,$$

where $\boldsymbol{V}_i = N^{-1}\sum_{j=1}^{N}a_{ij}\boldsymbol{R}_j$ *and* $\Omega = N^{-1}\sum_{i=1}^{N}\Omega_i$*.*

From the asymptotic normality in Theorem 3, we can find that $\hat{\boldsymbol{\beta}}(u)$ has a bias term $\boldsymbol{b}(u)$, and $\boldsymbol{b}(u)$ has a complex structure. In order to improve the efficiency of statistical inference, we propose a bias-corrected procedure to remove the bias term $\boldsymbol{b}(u)$. Noting that both cross-sectional and serial dependence and heteroskedasticity are allowed in the error terms, we first estimate W_1 and W_2 as follows:

$$\hat{W}_1 = -\boldsymbol{B}(u)\hat{D}_0^{-1}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(\boldsymbol{R}_i - \hat{\boldsymbol{V}}_i)^\tau\hat{\boldsymbol{F}}}{T}\left(\frac{\hat{\Lambda}^\tau\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_j\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{jt}\right),$$

$$\hat{W}_2 = -\boldsymbol{B}(u)\hat{D}_0^{-1}\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{N}\sum_{k=1}^{N}\left(\boldsymbol{R}_i^\tau\hat{\Omega}_k\hat{\boldsymbol{F}} - T^{-1}\hat{\boldsymbol{F}}\hat{\boldsymbol{F}}^\tau\hat{\Omega}_k\hat{\boldsymbol{F}}\right)\left(\frac{\hat{\Lambda}^\tau\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i,$$

where n satisfies $n/N \to 0$, $n/T \to 0$, and $\hat{D}_0 = \frac{L_N}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}^\tau$ with $\boldsymbol{F}^0$, λ_i and Λ being replaced by $\hat{\boldsymbol{F}}$, $\hat{\lambda}_i$ and $\hat{\Lambda}$ in $\hat{Z}_{it}$, respectively. Note that

$\boldsymbol{R}_i^\tau \hat{\Omega}_k \hat{\boldsymbol{F}} = (I_{p_0}, \boldsymbol{0})(\boldsymbol{S}_i^\tau \hat{\Omega}_k \boldsymbol{S}_i)(\boldsymbol{0}^\tau, I_r)^\tau$ and $\hat{\boldsymbol{F}}^\tau \hat{\Omega}_k \hat{\boldsymbol{F}} = (\boldsymbol{0}, I_r)(\boldsymbol{S}_i^\tau \hat{\Omega}_k \boldsymbol{S}_i)(\boldsymbol{0}^\tau, I_r)^\tau$, where $p_0 = \sum_{k=1}^p L_k$ and $\boldsymbol{S}_i^\tau \hat{\Omega}_k \boldsymbol{S}_i = C_{0i} + \sum_{\nu=1}^q [1 - \nu/(q+1)] (C_{\nu i} + C_{\nu i}^\tau)$, $\boldsymbol{S}_i = (\boldsymbol{R}_i, \hat{\boldsymbol{F}})$, $C_{\nu i} = \frac{1}{T}\sum_{t=\nu+1}^T S_{it}\hat{\varepsilon}_{kt}\hat{\varepsilon}_{k,t-\nu}S_{i,t-\nu}$, and $q \to \infty$ and $q/T^{1/4} \to 0$ as $T \to \infty$. Thus, we define the bias-corrected estimator of $\boldsymbol{\beta}(u)$ by

$$\breve{\boldsymbol{\beta}}(u) = \hat{\boldsymbol{\beta}}(u) - \frac{L_N}{N}\hat{W}_1 - \frac{L_N}{T}\hat{W}_2.$$

The following theorem show that there is no bias term in the asymptotic distribution of the bias-corrected estimator $\breve{\boldsymbol{\beta}}(u)$.

Theorem 4. *Suppose that assumptions (A1)-(A10) hold. If $\delta_{NT}^{-2} L_N \log L_N \to 0$, $L_N^{2d+1}/NT \to \infty$ and $T/N \to c$ as $N \to \infty$ and $T \to \infty$ simultaneously, then*

$$\Sigma^{-1/2}(\breve{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)) \xrightarrow{L} N(0, I_p).$$

In particular, we have $\Sigma_{kk}^{-1/2}(\breve{\beta}_k(u) - \beta_k(u)) \xrightarrow{L} N(0,1)$ for $k = 1, \ldots, p$.

Next we consider some special cases where the asymptotic bias can be simplified. (1) If there is the absence of serial correlation and heteroskedasticity, then $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij,tt} = \sigma_{ij}$ since it does not depend on t. It is easy to show that $W_2 = 0$. (2) If there is the absence of cross-sectional correlation and heteroskedasticity, $E(\varepsilon_{it}\varepsilon_{is}) = \sigma_{ii,ts} = \omega_{ts}$ since it does not depend on i, a simple calculation yields $W_1 = 0$. Let Π and Ξ be the probability limits, and they are defined by, respectively,

$$\Pi = \text{plim}\boldsymbol{B}(u)\left(\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{Z}_i\right)^{-1}\Sigma_{NT2}\left(\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{Z}_i\right)^{-1}\boldsymbol{B}(u)^{\tau},$$

$$\Xi = \text{plim}\boldsymbol{B}(u)\left(\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{Z}_i\right)^{-1}\Sigma_{NT3}\left(\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{Z}_i\right)^{-1}\boldsymbol{B}(u)^{\tau},$$

where $\Sigma_{NT2} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}\sum_{t=1}^{T}Z_{it}Z_{jt}^{\tau}$ and $\Sigma_{NT3} = \sum_{t=1}^{T}\sum_{s=1}^{T}\omega_{ts}\sum_{i=1}^{N}Z_{it}Z_{is}^{\tau}$.

Corollary 1. *Suppose that assumptions (A1)-(A10) hold. If* $\delta_{NT}^{-2}L_N\log L_N \to 0$ *and* $L_N^{2d+1}/NT \to \infty$ *as* $N \to \infty$ *and* $T \to \infty$ *simultaneously, we have the following results:*

(i) *In the absence of serial correlation and heteroskedasticity and* $T/N \to 0$*, then* $\Pi^{-1/2}(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)) \xrightarrow{L} N(0, I_p)$.

(ii) *In the absence of cross-sectonal correlation and heteroskedasticity and* $N/T \to 0$*, then* $\Xi^{-1/2}(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)) \xrightarrow{L} N(0, I_p)$.

For model (1.2) with unobservable multiple interactive fixed effects, Theorem 4 establishes the asymptotic normality for the bias-corrected estimators $\breve{\beta}_k(\cdot)$ of $\beta_k(\cdot)$. Hence, if we can obtain a consistent estimator $\hat{\Sigma}_{kk}$ of Σ_{kk}, the asymptotic pointwise confidence intervals for $\beta_k(u)$ can be constructed by

$$\breve{\beta}_k(u) \pm z_{\alpha/2}\hat{\Sigma}_{kk}^{-1/2}, \quad k = 1, \ldots, p,$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

4. A residual-based block bootstrap procedure

In theory, we can construct the pointwise confidence intervals for the coefficient functions $\beta_k(\cdot)$ by Theorems 3 and 4. For Theorem 3, we first need to derive the consistent estimators of the asymptotic biases and variances of the estimators $\hat{\beta}_k(\cdot)$ for $k = 1, \ldots, p$. Nevertheless, as the asymptotic biases and variances involve the unknown fixed effects $\boldsymbol{F}$ and the covariance matrices Ω_i of $\boldsymbol{\varepsilon}_i$, it is difficult to obtain the consistent and efficient estimators of the asymptotic biases and variances even if the plug-in method is used. For Theorem 4, it is difficult to show the consistency of the estimators $\hat{W}_1$ and $\hat{W}_2$ since both cross-sectional and serial dependence and heteroskedasticity are allowed in the error terms.

Therefore, the standard nonparametric bootstrap procedure cannot be applied to construct the pointwise confidence intervals directly because there exist the cross-sectional and serial correlations within the group in model (1.2). They will not only increase the computational burden and cause the accumulative errors, but also make it more difficult to construct the pointwise confidence intervals. To overcome the limitations, we hereby propose a residual-based block bootstrap bias correction procedure to construct the pointwise confidence intervals for $\beta_k(\cdot)$ with the detailed algorithm as follows.

Step 1. Fit model (1.2) using the proposed methods in Section 2, and estimate the residuals ε_{it} by

$$\hat{\varepsilon}_{it} = Y_{it} - \sum_{k=1}^{p}\sum_{l=1}^{L_k} \hat{\gamma}_{kl} X_{it,k} B_{kl}(U_{it}) + \hat{\lambda}_i^\tau \hat{F}_t,\ i = 1, \ldots, N,\ t = 1, \ldots, T.$$

Step 2. Generate the bootstrap residuals ε_{it}^* by $\hat{\varepsilon}_{it}$ using the block bootstrap method by a two-step procedure: (i) Choose the block lengths. In our block bootstrap procedure, similar to Inoue and Shintani (2006), we choose the block lengths by $l_1 = cT^{1/3}$ and $l_2 = cN^{1/3}$ for some $c > 0$, respectively. (ii) Resample the blocks and generate the bootstrap samples. The blocks can be overlapping or non-overlapping. According to Lahiri (1999), there is little difference in the performance for these two methods. We hence adopt the non-overlapping method for simplicity. Then we first divide the $N \times T$ residual matrix $\hat{\boldsymbol{\varepsilon}}$ into $m_1 = T/l_1$ blocks by column, and generate the bootstrap samples $N \times T$ matrix $\tilde{\boldsymbol{\varepsilon}}$ by resampling with replacement the m_1 blocks of columns of $\hat{\boldsymbol{\varepsilon}}$. Next, we divide $\tilde{\boldsymbol{\varepsilon}}$ into $m_2 = N/l_2$ blocks by row and generate the bootstrap samples matrix $\boldsymbol{\varepsilon}^*$ by resampling with replacement the m_2 blocks of rows of $\tilde{\boldsymbol{\varepsilon}}$.

Step 3. We generate the bootstrap sample Y_{it}^* by the following model:

$$Y_{it}^* = \sum_{k=1}^{p}\sum_{l=1}^{L_k} \hat{\gamma}_{kl} X_{it,k} B_{kl}(U_{it}) + \hat{\lambda}_i^\tau \hat{F}_t + \varepsilon_{it}^*,\ i = 1, \ldots, N,\ t = 1, \ldots, T,$$

where $\hat{\gamma}_{kl}$, $\hat{F}_t$ and $\hat{\lambda}_i$ are the respective estimators of γ_{kl}, F_t and λ_i using the estimation procedure in Section 2. Based on the bootstrap sample $\{(Y_{it}^*, X_{it}, U_{it}), i = 1, \ldots, N, t = 1, \ldots, T\}$, we calculate the bootstrap estimator $\hat{\boldsymbol{\beta}}^{(b)}(\cdot)$ also by the estimation procedure in Section 2.

Step 4. Repeat Steps 2 and 3 for B times to get a size B bootstrap estimators $\hat{\boldsymbol{\beta}}^{(b)}(u)$, $b = 1, \ldots, B$. The bootstrap estimator $\mathrm{Var}^*(\hat{\boldsymbol{\beta}}(u)|\mathcal{D})$ of $\Sigma = \mathrm{Var}(\hat{\boldsymbol{\beta}}(u)|\mathcal{D})$ is taken as the sample variance of $\hat{\boldsymbol{\beta}}^{(b)}(u)$. Next, the bootstrap bias corrected estimator of $\hat{\beta}_k(u)$ can be defined as

$$\breve{\beta}_k(u) = \hat{\beta}_k(u) - \left(\frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_k^{(b)}(u) - \hat{\beta}_k(u) \right) = 2\hat{\beta}_k(u) - \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_k^{(b)}(u).$$

The intuition is that the bias of a bootstrap estimator is a good approximation to that of a true coefficient function estimator. Finally, we construct the asymptotic pointwise confidence intervals for $\beta_k(u)$ by

$$\breve{\beta}_k(u) \pm z_{\alpha/2}\{\mathrm{Var}^*(\hat{\beta}_k(u)|\mathcal{D})\}^{1/2}, \quad k = 1, \ldots, p,$$

where $z_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

5. Partially linear varying coefficient model

In this section, we consider a special case of model (1.2), where some components $\underline{X}_{it} = (X_{it,1}, \ldots, X_{it,q})^\tau$ of X_{it} are constant effects and the rest $\overline{X}_{it} = (X_{it,q+1}, \ldots, X_{it,p})^\tau$ are varying effects for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, then model (1.2) becomes the following partially linear varying coefficient model with interactive fixed effects:

$$Y_{it} = \underline{X}_{it}^\tau \boldsymbol{\beta}^{(1)}(U_{it}) + \overline{X}_{it}^\tau \boldsymbol{\theta} + \lambda_i^\tau F_t + \varepsilon_{it}, \tag{5.1}$$

where $\boldsymbol{\beta}^{(1)}(u) = (\beta_1(u)), \ldots, \beta_q(u))^\tau$ and $\boldsymbol{\theta} = (\beta_{q+1}, \ldots, \beta_p)^\tau$.

Similar to the proposed estimation procedure in Section 2, we can define the following objective function:

$$Q(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\theta}, \boldsymbol{F}) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \underline{\boldsymbol{R}}_i \boldsymbol{\gamma}^{(1)} - \overline{\boldsymbol{X}}_i \boldsymbol{\theta})^\tau M_{\boldsymbol{F}} (\boldsymbol{Y}_i - \underline{\boldsymbol{R}}_i \boldsymbol{\gamma}^{(1)} - \overline{\boldsymbol{X}}_i \boldsymbol{\theta}). \tag{5.2}$$

Thus, the estimators of $\boldsymbol{\gamma}^{(1)}$ and $\boldsymbol{\theta}$ can be obtained by iteration between $\boldsymbol{\gamma}^{(1)}$, $\boldsymbol{\theta}$ and $\boldsymbol{F}$ using the following nonlinear equations:

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left[\sum_{i=1}^{N} \overline{\boldsymbol{X}}_i^\tau M_{\hat{\boldsymbol{F}}} \left\{ I_T - \underline{\boldsymbol{R}}_i \left(\sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \underline{\boldsymbol{R}}_i \right)^{-1} \sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \right\} \overline{\boldsymbol{X}}_i \right]^{-1} \\
&\quad \times \sum_{i=1}^{N} \overline{\boldsymbol{X}}_i^\tau M_{\hat{\boldsymbol{F}}} \left\{ I_T - \underline{\boldsymbol{R}}_i \left(\sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \underline{\boldsymbol{R}}_i \right)^{-1} \sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \right\} \boldsymbol{Y}_i, \\
\hat{\gamma}^{(1)} &= \left(\sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \underline{\boldsymbol{R}}_i \right)^{-1} \sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} (\boldsymbol{Y}_i - \overline{\boldsymbol{X}}_i^\tau \hat{\boldsymbol{\theta}}), \\
\hat{\boldsymbol{F}} V_{NT} &= \left[\frac{1}{NT} \sum_{i=1}^{N} (\boldsymbol{Y}_i - \underline{\boldsymbol{R}}_i \hat{\gamma}^{(1)} - \overline{\boldsymbol{X}}_i^\tau \hat{\boldsymbol{\theta}}) (\boldsymbol{Y}_i - \underline{\boldsymbol{R}}_i \hat{\gamma}^{(1)} - \overline{\boldsymbol{X}}_i^\tau \hat{\boldsymbol{\theta}})^\tau \right] \hat{\boldsymbol{F}}.
\end{aligned}$$

By the property of B-spline bases: $\sum_{l=1}^{L_k} B_{kl}(u) = 1$, if $\beta_k(u)$ is a constant β_k, we set $\gamma_k = \beta_k \mathbf{1}_{L_k}$, where $\mathbf{1}_{L_k}$ is an $L_k \times 1$ vector with entries 1.

With little abuse of notation, then (5.2) can be rewritten as

$$Q(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\theta}, \boldsymbol{F}) = Q(\boldsymbol{\gamma}, \boldsymbol{F}) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i \boldsymbol{\gamma})^\tau M_{\boldsymbol{F}} (\boldsymbol{Y}_i - \boldsymbol{R}_i \boldsymbol{\gamma}), \tag{5.3}$$

where $\boldsymbol{\gamma} = (\gamma_1^\tau, \ldots, \gamma_q^\tau, \beta_{q+1}\mathbf{1}_{L_{q+1}}^\tau, \ldots, \beta_p \mathbf{1}_{L_p}^\tau)^\tau = (\boldsymbol{\gamma}^{(1)\tau}, \beta_{q+1}\mathbf{1}_{L_{q+1}}^\tau, \ldots, \beta_p \mathbf{1}_{L_p}^\tau)^\tau$. For each $k = q+1, \ldots, p$, we treat β_k as a function and apply the estimation procedure in Section 2 to obtain the initial estimators of $\hat{\boldsymbol{\gamma}}^{(1)}$, $\hat{\boldsymbol{F}}$ and $\hat{\Lambda}$. Then we propose the following robust iteration algorithm for estimating the parameters $(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\theta}, \boldsymbol{F}, \Lambda)$.

Step 1. Start with an initial estimator $(\hat{\boldsymbol{\gamma}}^{(1)}, \hat{\boldsymbol{F}}, \hat{\Lambda})$.

Step 2. Given $\hat{\boldsymbol{\gamma}}^{(1)}$, $\hat{\boldsymbol{F}}$ and $\hat{\Lambda}$, compute

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\gamma}}^{(1)}, \hat{\boldsymbol{F}}, \hat{\Lambda}) = \left(\sum_{i=1}^{N} \overline{\boldsymbol{X}}_i^\tau \overline{\boldsymbol{X}}_i \right)^{-1} \sum_{i=1}^{N} \overline{\boldsymbol{X}}_i^\tau (\boldsymbol{Y}_i - \underline{\boldsymbol{R}}_i \hat{\boldsymbol{\gamma}}^{(1)} - \hat{\boldsymbol{F}} \hat{\lambda}_i).$$

Step 3. Given $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{F}}$ and $\hat{\Lambda}$, compute

$$\hat{\boldsymbol{\gamma}}^{(1)}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{F}}, \hat{\Lambda}) = \left(\sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau \underline{\boldsymbol{R}}_i \right)^{-1} \sum_{i=1}^{N} \underline{\boldsymbol{R}}_i^\tau (\boldsymbol{Y}_i - \overline{\boldsymbol{X}}_i \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{F}} \hat{\lambda}_i).$$

Step 4. Given $\hat{\boldsymbol{\gamma}}^{(1)}$ and $\hat{\boldsymbol{\theta}}$, compute $\hat{\boldsymbol{F}}$ according to (5.3) (multiplied by $\sqrt{T}$ due to the restriction that $\boldsymbol{F}^\tau \boldsymbol{F}/T = I_r$) and calculate $\hat{\Lambda}$ using formula (2.9) with $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}^{(1)\tau}, \hat{\beta}_{q+1}\mathbf{1}_{L_{q+1}}^\tau, \ldots, \hat{\beta}_p \mathbf{1}_{L_p}^\tau)^\tau$.

Step 5. Repeat Steps 2—4 until $(\hat{\boldsymbol{\gamma}}^{(1)}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{F}}, \hat{\Lambda})$ satisfy the given convergence criterion.

In order to give the following asymptotic distribution, we first introduce some notations. Let

$$\overline{\boldsymbol{Z}}_i = M_{\boldsymbol{F}^0}\overline{\boldsymbol{X}}_i - \frac{1}{N}\sum_{j=1}^{N} M_{\boldsymbol{F}^0}\overline{\boldsymbol{X}}_j a_{ij}, \quad \underline{\boldsymbol{Z}}_i = M_{\boldsymbol{F}^0}\underline{\boldsymbol{R}}_i - \frac{1}{N}\sum_{j=1}^{N} M_{\boldsymbol{F}^0}\underline{\boldsymbol{R}}_j a_{ij},$$

$$\overline{\Phi} = \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{Z}}_i^{\tau}\overline{\boldsymbol{Z}}_i, \quad \underline{\Phi} = \frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{Z}}_i^{\tau}\underline{\boldsymbol{Z}}_i, \quad \Psi = \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{Z}}_i^{\tau}\underline{\boldsymbol{Z}}_i$$

and $\check{\boldsymbol{Z}}_i = \overline{\boldsymbol{Z}}_i - \underline{\boldsymbol{Z}}_i\underline{\Phi}^{-1}\Psi^{\tau}$. In addition, we define the following probability limits

$$\Pi_1 = \text{plim}\frac{1}{NT}\sum_{i=1}^{N}\check{\boldsymbol{Z}}_i^{\tau}\check{\boldsymbol{Z}}_i = \text{plim}(\overline{\Phi} - \Psi\underline{\Phi}^{-1}\Psi^{\tau}),$$

$$\Pi_2 = \text{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{ij,ts}\check{Z}_{it}\check{Z}_{js}^{\tau}.$$

The following theorem gives the asymptotic normality of the parametric components.

Theorem 5. *Suppose that assumptions (A1)-(A10) hold. If* $\delta_{NT}^{-2}L_N\log L_N \to 0$*,* $L_N^{2d+1}/NT \to \infty$ *and* $T/N \to c$ *as* $N \to \infty$ *and* $T \to \infty$ *simultaneously, then*

$$(NT)^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{L} N(\boldsymbol{b}, \Pi_1^{-1}\Pi_2\Pi_1^{-1}),$$

where $\boldsymbol{b} = c^{1/2}\check{S}_1^0 + c^{-1/2}\check{S}_2^0$*, and* $\check{S}_1^0$ *is the probability limit of* $\check{S}_1$ *with*

$$\begin{aligned}\check{S}_1 \;&=\; -(\overline{\Phi} - \Psi\underline{\Phi}^{-1}\Psi^{\tau})^{-1}\Bigg[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{V}}_i)^{\tau}\boldsymbol{F}^0}{T}G^0\lambda_j\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\Big)\\ &\quad -\Psi\underline{\Phi}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\underline{\boldsymbol{R}}_i - \underline{\boldsymbol{V}}_i)^{\tau}\boldsymbol{F}^0}{T}G^0\lambda_j\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\Big)\Bigg]\end{aligned}$$

and $\check{S}_2^0$ *is the probability limit of* $\check{S}_2$ *with*

$$\begin{aligned}\check{S}_2 &= -(\overline{\Phi} - \Psi\underline{\Phi}^{-1}\Psi^\tau)^{-1}\Bigg(\frac{1}{NT}\sum_{i=1}^N \overline{\boldsymbol{X}}_i^\tau M_{\boldsymbol{F}^0}\Omega\boldsymbol{F}^0G^0\lambda_i \\ &\quad -\Psi\underline{\Phi}^{-1}\frac{1}{NT}\sum_{i=1}^N \underline{\boldsymbol{R}}_i^\tau M_{\boldsymbol{F}^0}\Omega\boldsymbol{F}^0G^0\lambda_i\Bigg),\end{aligned}$$

where $G^0 = (\boldsymbol{F}^{0\tau}\boldsymbol{F}^0/T)^{-1}(\Lambda^\tau\Lambda/N)^{-1}$ *and* $\overline{\boldsymbol{V}}_i = N^{-1}\sum_{j=1}^N a_{ij}\overline{\boldsymbol{X}}_j$.

It is easy to show that $\check{S}_1^0 = 0$ in bias term $\boldsymbol{b}$ if the cross-sectional correlation and heteroskedasticity are absent. Similarly, $\check{S}_2^0 = 0$ if the serial correlation and heteroskedasticity are absent. We also show that both $\check{S}_1^0 = \check{S}_2^0 = 0$ if ε_{it} are i.i.d. over i and t. From Theorem 5, we can see that the convergence rate of $\hat{\boldsymbol{\theta}}$ is order of $O_P((NT)^{-1/2})$, the substitution of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in model (5.1) will have little impact on the estimation of $\beta_j(u)$, $j = 1,\ldots,q$, which implies the estimator $\hat{\beta}_j(u)$ will have the similar asymptotic distributions in Theorems 3 and 4.

6. Hypothesis testing

In practice, it is often of interest to test whether one or several coefficient functions are nonzero constants or are identically zero. We here propose a goodness-of-fit test based on the comparison of the residual sum of squares from least square fits under both the null hypothesis and the alternative.

We consider the following null hypothesis that some of the coefficient functions are constants, i.e.,

$$H_0: \quad \beta_{q+1}(u) = \beta_{q+1}, \ldots, \beta_p(u) = \beta_p$$

for all $u \in \mathcal{U}$, where β_k $(k = q+1, \ldots, p)$ are unknown constants. Under H_0, model (1.2) reduces to the partially linear varying coefficient panel data model (5.1). Let $\hat{\boldsymbol{\gamma}}^{(1)*}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{F}}^*$ and $\hat{\lambda}_i^*$ be the consistent estimators of $\boldsymbol{\gamma}^{(1)}$, $\boldsymbol{\theta}$, $\boldsymbol{F}$ and λ_i, respectively. Thus, the residual sum of squares under the null hypothesis H_0 is

$$\mathrm{RSS}_0 = \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{F}}^*\hat{\lambda}_i^*)^\tau(\boldsymbol{Y}_i - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{F}}^*\hat{\lambda}_i^*).$$

Under the general alternative that all the coefficient functions are allowed to be vary with u, the residual sum of squares is defined by

$$\mathrm{RSS}_1 = \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i). \tag{6.4}$$

We extend the generalized likelihood ratio in Fan et al. (2001) to the current setting, and construct the test statistic under null hypothesis H_0 as follows

$$T_n = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \tag{6.5}$$

where $\mathrm{RSS}_0 - \mathrm{RSS}_1$ indicates the difference of fit under the null hypothesis and alternative hypothesis. If T_n is larger than an appropriate critical value, we will reject the null hypothesis H_0. Let t_0 be the observed value of T_n, the p-value of the test is defined by $p_0 = P_{H_0}(T_n > t_0)$, which denotes the probability of the event $\{T_n > t_0\}$. For a given significance level α_0, the null hypothesis H_0 will be rejected if $p_0 \leq \alpha_0$.

Theorem 6. *Suppose that the conditions of Theorem 3 are satisfied. Under the null hypothesis H_0, then $T_n \to 0$ in probability as $N \to \infty$ and $T \to \infty$. Otherwise, if $\inf_{a\in\mathbb{R}} \|\beta_k(u) - a\|_{L_2} > 0$ for some $k = q+1, \ldots, p$, then there exists a constant t_0 such that $T_n > t_0$ with probability approaching one as $N \to \infty$ and $T \to \infty$.*

Since it is difficult to develop the asymptotic null distribution of the statistic T_n, we use the following bootstrap procedure to evaluate the null distribution of T_n and compute p-values of the test.

Step 1. We generate the bootstrap sample $\{(Y_{it}^*, X_{it}, U_{it}), i = 1, \ldots, N, t = 1, \ldots, T\}$ as described in Section 4, and calculate the bootstrap test statistic T_n^*.

Step 2. We repeat Step 1 a large number of times and compute the bootstrap distribution of T_n^*.

Step 3. When the observed test statistic T_n is greater than or equal to the $\{100(1-\alpha_0)\}$th percentile of the empirical distribution T_n^*, we reject the null hypothesis H_0 at the significance level α_0. The p-value of the test is the empirical probability of event $\{T_n^* \geq T_n\}$.

7. Conclusion

In this paper, our contribution is to develop the estimation procedure for the varying coefficient panel data model with interactive fixed effects. Firstly, we use the B-splines to approximate the coefficient functions for the varying coefficient panel data model with interactive fixed effects. With an appropriate choice of the smoothing parameters, we propose a robust nonlinear iteration scheme based on the least squares method to estimate the coefficient functions, and then establish the asymptotic theory for the resulting estimators under some regularity assumptions, including the consistency, the convergence rate and the asymptotic distribution. Secondly, to deal with the serial and cross-sectional correlations and heteroskedasticity within our model that will increase the computational burden and cause the accumulative errors, we propose the residual-based block bootstrap procedure to construct the pointwise confidence intervals for the coefficient functions. Thirdly, we also extend our proposed estimation procedure to the partially linear varying coefficient model with interactive fixed effects, and study the asymptotic properties of the resulting estimator. In addition, we also develop a test statistic for testing the constancy of the varying coefficient functions, and propose a bootstrap procedure to evaluate the null distribution of the test statistic. Finally, numerical studies are also carried out to demonstrate the satisfactory performance of our proposed methods in practice and to also support the derived theoretical results.

Supplementary Materials

Supplement to "Varying Coefficient Panel Data Model with Interactive Fixed Effects". To save space, the numerical studies, the proofs of Theorems 1–6 and Corollary 1, and Lemmas 1–7 and their proofs are provided in the supplementary material. In addition, we also introduce the estimation procedure for a special model: the varying coefficient panel data model with additive fixed effects.

Acknowledgements

The authors sincerely thank the Editor, Associate Editor and two anonymous reviewers for their insightful comments and suggestions that have dramatically improved the paper. Sanying Feng's research was supported by NSFC (No. 11501522) and the Excellent Youth Foundation of Zhengzhou University (No. 32210452). Gaorong Li's research was supported by NSFC (Nos. 11871001 and 11471029) and the Beijing Natural Science Foundation (No. 1182003). Heng Peng's research was supported by CEGR Grant of the Research Grants Council of Hong Kong (Nos. HKBU12302615 and HKBU 12303618), FRG Grants from Hong Kong Baptist University (No. FRG2/16-17/042) and NSFC (No. 11871409). Tiejun Tong's research was supported by NSFC Grant (No. 11671338), RGC Grant (No. HKBU12303918) and HMRF Fund (No. 04150476).

References

Ahn, S. G., Y. H. Lee, and P. Schmidt (2001). GMM estimation of linear panel data models with time-varying individual effects. *J. Econometrics 102*, 219–255.

Arellano, M. (2003). *Panel Data Econometrics.* Oxford: Oxford University Press.

Bai, J. S. (2009). Panel data models with interactive fixed effects. *Econometrica 77*, 1229–1279.

Bai, J. S. and K. P. Li (2014). Theory and methods of panel data models with interactive effects. *Ann. Statist. 42*, 142–170.

Baltagi, B. H. (2005). *Econometrics Analysis of Panel Data.* New York: Wiley.

Cai, Z. W. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics 136*, 163–188.

Cai, Z. W. and Q. Li (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory 24*, 1321–1342.

Chiang, C. T., J. A. Rice, and C. O. Wu (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc. 96*, 605–619.

Coakley, J., A.-M. Fuertes, and R. P. Smith (2002). A principal components approach to cross-section dependence in panels. *No B5-3, 10th International Conference on Panel Data, Berlin, July 5-6*, 1–28.

Fan, J. Q., C. M. Zhang, and J. Zhang (2001). Generalized likelihood ratio statistics and wilks phenomenon. *Ann. Statist. 29*(1), 153–193.

Holtz-Eakin, D., W. Newey, and H. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*, 1371–1395.

Hsiao, C. (2003). *Analysis of Panel Data.* Cambridge: Cambridge University Press.

Huang, J. Z., C. O. Wu, and L. Zhou (2002). Varying-coefficient models and basis function approximations for the analysis of the analysis of repeated measurements. *Biometrika 89*, 111–128.

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

Huang, X. (2013). Nonparametric estimation in large panels with cross-sectional dependence. *Econometric Rev. 32*, 754–777.

Inoue, A. and M. Shintani (2006). Bootstrapping GMM estimators for time series. *J. Econometrics 133*, 531–555.

Jin, S. N. and L. J. Su (2013). A nonparametric poolability test for panel data models with cross section dependence. *Econometric Rev. 32*, 469–512.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist. 27*, 386–404.

Lee, N., H. R. Moon, and M. Weidner (2012). Analysis of interactive fixed effects dynamic linear panel regression with measurement error. *Economics Letters 117*, 239–242.

Li, D. G., J. Chen, and J. T. Gao (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *Economet. J. 14*, 387–408.

Li, D. G., J. H. Qian, and L. J. Su (2016). Panel data models with interactive fixed effects and multiple structural breaks. *J. Amer. Statist. Assoc. 111*, 1804–1819.

Li, G. R., H. Lian, P. Lai, and H. Peng (2015). Variable selection for fixed effects varying coefficient models. *Acta Math. Sin. (Engl. Ser.) 31*, 91–110.

Lu, X. and L. J. Su (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *J. Econometrics 190*, 148–175.

Malikov, E., S. C. Kumbhakar, and Y. Sun (2016). Varying coefficient panel data model in the presence of endogenous selectivity and fixed effects. *J. Econometrics 190*, 233–251.

Moon, H. R. and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica 83*, 1543–1579.

Moon, H. R. and M. Weidner (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory 33*, 158–195.

Noh, H. S. and B. U. Park (2010). Sparse varying coefficient models for longitudinal data. *Statist. Sinica 20*, 1183–1202.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*, 967–1012.

Rodriguez-Poo, J. M. and A. Soberon (2014). Direct semi-parametric estimation of fixed effects panel data varying coefficient models. *Economet. J. 17*, 107–138.

Rodriguez-Poo, J. M. and A. Soberon (2015). Nonparametric estimation of fixed effects panel data varying coefficient models. *J. Multivariate Anal. 133*, 95–122.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist. 10*, 1348–1360.

Su, L. J. and Q. H. Chen (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory 29*, 1079–1135.

Su, L. J. and S. N. Jin (2012). Sieve estimation of panel data models with cross section dependence. *J. Econometrics 169*, 34–47.

Su, L. J., S. N. Jin, and Y. H. Zhang (2015). Specification test for panel data models with interactive fixed effects. *J. Econometrics 186*, 222–244.

Sun, Y. G., R. J. Carroll, and D. D. Li (2009). Semiparametric estimation of fixed effects panel data varying coefficient models. *Adv. Econom. 25*, 101–129.

Wang, H. S. and Y. C. Xia (2009). Shrinkage estimation of the varying coefficient model. *J. Amer. Statist. Assoc. 104*, 747–757.

Wang, L. F., H. Z. Li, and J. Z. Huang (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J. Amer. Statist. Assoc. 103*, 1556–1569.

Wu, J. H. and J. C. Li (2014). Testing for individual and time effects in panel data models with interactive effects. *Economics Letters 125*, 306–310.

Xue, L. G. and L. X. Zhu (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *J. Amer. Statist. Assoc. 102*, 642–652.

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, P. R. China

E-mail: fsy5801@zzu.edu.cn

School of Statistics, Beijing Normal University, Beijing 100875, P. R. China

E-mail: ligaorong@bnu.edu.cn

Department of Mathematics, Hong Kong Baptist University, Hong Kong

E-mail: hpeng@math.hkbu.edu.hk

Department of Mathematics, Hong Kong Baptist University, Hong Kong

E-mail: tongt@hkbu.edu.hk

(Received June 26, 2018; accepted August 1, 2019)