Statistica Sinica **31** (2021), 935-957
doi:https://doi.org/10.5705/ss.202018.0248

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 and* [3]*Hong Kong Baptist University*

Abstract: 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 using 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, convergence rate, and asymptotic distributions. To construct the pointwise confidence intervals for the coefficient functions, we propose a residual-based block bootstrap method that reduces the computational burden and avoids accumulative errors. We extend our proposed procedure to partially linear varying-coefficient panel-data models with unobservable multiple interactive fixed effects, and examine the problem of constant coefficients versus function coefficients. Simulation studies and a real-data analysis are used to assess the performance of the 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 periods. Panel-data analysis has attracted considerable attention in the literature. The methodology for a parametric panel-data analysis is relatively mature; see, for example, Arellano (2003), Hsiao (2003), Baltagi (2005), and the references therein. Individual and time effects may enter the model additively, or they can interact multiplicatively, leading to the so-called interactive effects or a factor structure. Panel-data models with interactive fixed effects are a useful modeling paradigm. In macroeconomics, incorporating interactive effects can account for the heterogenous effects of unobservable common shocks, while the regressors can be inputs, such as labor and

Corresponding author: Gaorong Li, School of Statistics, Beijing Normal University, Beijing 100875, P. R. China. E-mail: ligaorong@bnu.edu.cn.

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 following 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, such 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 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, Newey and Rosen (1988) estimated model (1.1) using quasi-differencing and lagged variables as instruments. Their approach, however, rules out time constant regressors. Coakley, Fuertes and Smith (2002) studied model (1.1) by augmenting the regression of Y on X with the principal components of the ordinary least squares residuals. However, Pesaran (2006) showed that this method is inconsistent unless X_{it} and λ_i tend 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 with the cross-sectional averages of X_{it}. Although Pesaran's estimator is consistent, it does not allow for time-invariant individual regressors. Ahn, Lee and Schmidt (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, being able to identify their estimator requires that X_{it} is correlated with λ_i, and it is impossible to test for the interactive random effects assumption. Bai (2009) studied the identification, consistency, and limiting distribution of the principal component analysis (PCA) estimators, showing that they 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 in model (1.1). Li, Qian and Su (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, Moon and Weidner (2012), Su and Chen (2013), Moon and Weidner (2015), Lu and Su (2016), and many others.

Note that the aforementioned works focus on linear specifications of the regression relationships in panel-data models with interactive fixed effects. A natural extension of model (1.1) is to consider the following 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, while retaining the unobservable multiple interactive fixed effects, general nonparametric characteristics, and explanatory power of the linear panel-data model.

Model (1.2) is fairly general, and 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 such models. Su and Jin (2012) extended the CCE method of Pesaran (2006) from a linear model to a nonparametric model using the method of sieves. Jin and Su (2013) constructed a nonparametric test for poolability in nonparametric regression models with interactive fixed effects. Su, Jin and Zhang (2015) proposed a consistent nonparametric test for linearity in a 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, Carroll and Li (2009) considered estimations using a local linear regression and kernel-based weights. Li, Chen and Gao (2011) considered a nonparametric time varying-coefficient model with fixed effects under the assumption of cross-sectional independence, and proposed methods for estimating the trend function and coefficient functions. Rodriguez-Poo and Soberon (2014) proposed a new technique to estimate the varying-coefficient functions based on the first-order differences and a local linear regression. Rodriguez-Poo and Soberon (2015) investigated the model using the mean transformation technique and a local linear regression. Li et al. (2015) considered variable selection for the model using the basis function approximations and the group nonconcave penalized functions. Malikov, Kumbhakar and Sun (2016) considered the problem of a 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, refer to Chiang, Rice and Wu (2001), Huang, Wu and Zhou (2002), Huang, Wu and Zhou (2004), Xue and Zhu (2007), Cai (2007), Cai and Li (2008), Wang, Li and Huang (2008), Wang and Xia (2009), and Noh and Park (2010). Note, however, that most of these studies focus on a "large N small T" setting.

Despite the rich literature on panel data models with interactive fixed effects, to the best of our knowledge, there are few works on varying-coefficient panel-data models with interactive fixed effects. As such, the main goals of this study are to estimate the coefficient functions $\boldsymbol{\beta}(\cdot)$, and to establish the asymptotic theory for varying-coefficient panel-data models with interactive fixed effects when both N and T tend to infinity and there exist 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), owing 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 make the estimation procedure and the asymptotic theory much more complicated than those in Huang, Wu and Zhou (2002). To apply the asymptotic normality to construct the pointwise confidence intervals for the coefficient functions, we need consistent estimators of the asymptotic biases and variances. To reduce the computational burden and to avoid accumulative errors, we propose a residual-based block bootstrap procedure to construct these confidence intervals.

Moreover, we extend the proposed estimation procedure to include partially linear varying coefficient models with interactive fixed effects, and show that the convergence rate for the estimation of the parametric components is of order $O_P((NT)^{-1/2})$. 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, we develop a residual-based block bootstrap procedure to construct the pointwise confidence intervals for the coefficient functions. In Section 5, we extend the estimation procedure to partially linear varying coefficient models 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. Technical details are given in the online Supplementary Material, along with simulation studies and a real application to demonstrate the efficacy of our proposed methods.

2. Methodology

To estimate the coefficient functions $\beta_k(\cdot)$, for $1 \leq k \leq p$, 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, selected using “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 interactive fixed effects.

Because 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, owing 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 these 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

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

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

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

Owing 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:

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

These two restrictions 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 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 in (2.4) with (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 $M_{\boldsymbol{F}} = I_T - \boldsymbol{F}(\boldsymbol{F}^\tau \boldsymbol{F})^{-1} \boldsymbol{F}^\tau = I_T - \boldsymbol{F}\boldsymbol{F}^\tau / T$ 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:

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

Because 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:

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

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

where V_{NT} is a diagonal matrix consisting of the r largest eigenvalues of the matrix $(NT)^{-1} \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i \hat{\boldsymbol{\gamma}})(\boldsymbol{Y}_i - \boldsymbol{R}_i \hat{\boldsymbol{\gamma}})^\tau$, arranged in decreasing order. As noted by Bai (2009), the iterated solution is somewhat sensitive to the initial values. Bai (2009) proposed starting with either the least squares estimator of $\boldsymbol{\gamma}$ or the principal components estimate of $\boldsymbol{F}$. 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 for the initial values. In general, poor initial values result in an exceptionally large number of iterations. By (2.5), (2.7), and (2.8), we have

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

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 $\beta_k(u)$ as

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

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

Step 3. Given $\hat{\boldsymbol{\gamma}}$, compute $\hat{\boldsymbol{F}}$ according to (2.8) (multiplied by $\sqrt{T}$, owing 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 asymptotic properties for the proposed estimators, we let $\mathcal{F} \equiv \{\boldsymbol{F} : \boldsymbol{F}^\tau \boldsymbol{F}/T = I_r\}$ and

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

where $a_{ij} = \lambda_i^\tau (\Lambda^\tau \Lambda / N)^{-1} \lambda_j$. To obtain a 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 Bai (2009)).

3.1. Regularity assumptions

In this section, we introduce a definition and present some regularity assumptions, which we use to establish 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 independent and identically distributed (i.i.d.) across 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 stationary 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 zero 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 zero 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 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), $E\left|N^{-1/2}\sum_{i=1}^{N}[\varepsilon_{it}\varepsilon_{is}-E(\varepsilon_{it}\varepsilon_{is})]\right|^4 \le M$.

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

(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 studies on varying-coefficient models with repeated measurements, such as those of Huang, Wu and Zhou (2002), Huang, Wu and Zhou (2004), and Wang, Li and Huang (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 study, whether F_t or λ_i has a zero mean is not crucial, because 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 is used for the system of general basis functions B_{kl}, which includes 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 Huang, Wu and Zhou (2002)). We define $\hat{\beta}_k(\cdot)$ as 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 asymptotic properties including consistency, a convergence rate, and an asymptotic distribution.

Theorem 1. *Suppose 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) $\hat{\beta}_k(\cdot)$*, for* $k = 1,\ldots,p$*, 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 unique estimators $\hat{\beta}_k(\cdot)$ for the unknown coefficient functions $\beta_k(\cdot)$ under some regularity assumptions, regardless of whether unobservable multiple interactive fixed effects exist 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 that guarantees 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 the following. 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 cross-sectional and serial correlation and heteroskedasticity, respectively. 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 $O_P\left((NT)^{-2d/(2d+1)}\right)$, which holds for the 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 one in the kth coordinate, and zero in all other coordinates, for $k = 1, \dots, p$. Then, the conditional variance of $\hat{\beta}_k(u)$ is

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

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 $(1/NT)\Sigma_{NT1}$; then, $(1/\sqrt{NT})\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{\varepsilon}_i \xrightarrow{L} N(\mathbf{0}, \Sigma_1)$, where "$\xrightarrow{L}$" denotes convergence in distribution.

Denote $\tilde{\Sigma} = D_0^{-1}\Sigma_1 D_0^{-1}$, where $D_0 = \mathrm{plim}(L_N/NT)\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{Z}_i$. 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 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 a statistical inference, we propose a bias-corrected procedure to remove the bias term $\boldsymbol{b}(u)$. Noting that 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 $\hat{D}_0 = (L_N/NT)\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}^\tau$, with $\boldsymbol{F}^0$, λ_i, and Λ replaced with $\hat{\boldsymbol{F}}$, $\hat{\lambda}_i$, 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 $\boldsymbol{S}_i^\tau\hat{\Omega}_k\boldsymbol{S}_i = C_{0i} + \sum_{\nu=1}^{q}\left[1 - \nu/(q+1)\right](C_{\nu i} + C_{\nu i}^\tau)$, $\boldsymbol{S}_i = (\boldsymbol{R}_i, \hat{\boldsymbol{F}})$, $C_{\nu i} = (1/T)\sum_{t=\nu+1}^{T}S_{it}\hat{\varepsilon}_{kt}\hat{\varepsilon}_{k,t-\nu}S_{i,t-\nu}$, and $q \to \infty$ and $q/T^{1/4} \to 0$ as $T \to \infty$. Thus, we define the bias-corrected estimator of $\boldsymbol{\beta}(u)$ as

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

The following theorem shows 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) In the absence of serial correlation and heteroskedasticity, $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij,tt} = \sigma_{ij}$, because it does not depend on t. It is easy to show that $W_2 = 0$. (2) In the absence of cross-sectional correlation and heteroskedasticity, $E(\varepsilon_{it}\varepsilon_{is}) = \sigma_{ii,ts} = \omega_{ts}$, because it does not depend on i, in which case, a simple calculation yields $W_1 = 0$. Let Π and Ξ be the probability limits, defined as, 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* $T/N\to 0$, $\Pi^{-1/2}(\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}(u))\xrightarrow{L}N(0,I_p)$.

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

For model (1.2) with unobservable multiple interactive fixed effects, Theorem 4 establishes the asymptotic normality for the bias-corrected estimator $\breve{\beta}_k(\cdot)$ 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 as

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

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 $\beta_k(\cdot)$ from Theorems 3 and 4. For Theorem 3, we first need to derive consistent estimators for the asymptotic biases and variances of the estimators $\hat{\beta}_k(\cdot)$, for $k=1,\ldots,p$. Nevertheless, because 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 their consistent and efficient estimators, 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 $\hat{W}_2$, because 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 cross-sectional and serial correlations exist within the group in model (1.2). In addition to increasing the computational burden and causing accumulative errors, they make it more difficult to construct the pointwise confidence intervals. To overcome these limitations, we propose a residual-based block bootstrap bias-correction procedure to construct the pointwise confidence intervals for $\beta_k(\cdot)$. The algorithm follows.

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

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

Step 2. Generate the bootstrap residuals ε_{it}^* by $\hat{\varepsilon}_{it}$ using the block bootstrap method with a two-step procedure: (i) Choose the block lengths. In our block bootstrap procedure, similarly to Inoue and Shintani (2006), we choose block lengths of $l_1 = cT^{1/3}$ and $l_2 = cN^{1/3}$, respectively, for some $c > 0$. (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}^* using 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 using the estimation procedure in Section 2.

Step 4. Repeat Steps 2 and 3 B times to obtain a size B bootstrap estimator $\hat{\boldsymbol{\beta}}^{(b)}(u)$, for $b = 1,\ldots,B$. 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

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

Intuitively, 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 $\beta_k(u)$ as

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

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 $\boldsymbol{\theta} = (\beta_{q+1}, \ldots, \beta_p)^\tau$.

Similarly 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 iterating between $\boldsymbol{\gamma}^{(1)}$, $\boldsymbol{\theta}$, and $\boldsymbol{F}$ using the following nonlinear equations:

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left[\sum_{i=1}^N \overline{\boldsymbol{X}}_i^\tau M_{\hat{\boldsymbol{F}}} \left\{ I_T - \underline{\boldsymbol{R}}_i \left(\sum_{i=1}^N \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \underline{\boldsymbol{R}}_i\right)^{-1} \sum_{i=1}^N \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \right\} \overline{\boldsymbol{X}}_i \right]^{-1} \\
&\quad \times \sum_{i=1}^N \overline{\boldsymbol{X}}_i^\tau M_{\hat{\boldsymbol{F}}} \left\{ I_T - \underline{\boldsymbol{R}}_i \left(\sum_{i=1}^N \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \underline{\boldsymbol{R}}_i\right)^{-1} \sum_{i=1}^N \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \right\} \boldsymbol{Y}_i, \\
\hat{\boldsymbol{\gamma}}^{(1)} &= \left(\sum_{i=1}^N \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} \underline{\boldsymbol{R}}_i\right)^{-1} \sum_{i=1}^N \underline{\boldsymbol{R}}_i^\tau M_{\hat{\boldsymbol{F}}} (\boldsymbol{Y}_i - \overline{\boldsymbol{X}}_i^\tau \hat{\boldsymbol{\theta}}), \\
\hat{\boldsymbol{F}} V_{NT} &= \left[\frac{1}{NT} \sum_{i=1}^N (\boldsymbol{Y}_i - \underline{\boldsymbol{R}}_i \hat{\boldsymbol{\gamma}}^{(1)} - \overline{\boldsymbol{X}}_i^\tau \hat{\boldsymbol{\theta}})(\boldsymbol{Y}_i - \underline{\boldsymbol{R}}_i \hat{\boldsymbol{\gamma}}^{(1)} - \overline{\boldsymbol{X}}_i^\tau \hat{\boldsymbol{\theta}})^\tau \right] \hat{\boldsymbol{F}}.
\end{aligned} \tag{5.3}$$

By the property of B-spline bases that $\sum_{l=1}^{L_k} B_{kl}(u) = 1$ 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 of one. With a slight abuse of notation, (5.2) can be rewritten as

$$Q(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\theta}, \boldsymbol{F}) = Q(\boldsymbol{\gamma}, \boldsymbol{F}) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i \boldsymbol{\gamma})^\tau M_{\boldsymbol{F}} (\boldsymbol{Y}_i - \boldsymbol{R}_i \boldsymbol{\gamma}), \tag{5.4}$$

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 $\sqrt{T}$, owing 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 notation. 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*

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

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

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

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 the 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, the convergence rate of $\hat{\boldsymbol{\theta}}$ is of order $O_P((NT)^{-1/2})$. Thus substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in model (5.1) will have little effect on the estimation of $\beta_j(u)$, for $j = 1, \ldots, q$. This implies that the estimator $\hat{\beta}_j(u)$ will have 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 that compares the residual sum of squares from least square fits under the null and alternative hypotheses.

We consider the null hypothesis that some of the coefficient functions are constants:

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

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 coefficient functions are allowed to vary with u, the residual sum of squares is defined by

$$\mathrm{RSS}_1 = \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i). \tag{6.1}$$

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

$$T_n = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \tag{6.2}$$

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

Theorem 6. *Suppose that the conditions of Theorem* 3 *are satisfied. Under the null hypothesis* H_0, $T_n \to 0$ *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* $T \to \infty$.

Because 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 the 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 many times to 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 the event $\{T_n^* \geq T_n\}$.

7. Conclusion

This study contributes to the literature by proposing an estimation procedure for a varying-coefficient panel-data model with interactive fixed effects. First, we use B-splines to approximate the coefficient functions for the model. With an appropriate choice of smoothing parameters, we propose a robust nonlinear iteration scheme based on the least squares method to estimate the coefficient functions. Then, we establish the asymptotic theory for the resulting estimators under some regularity assumptions, including their consistency, convergence rate, and asymptotic distribution. Second, to deal with the serial and cross-sectional correlation and heteroskedasticity within our model, which increases the computational burden and cause accumulative errors, we propose using a residual-based block bootstrap procedure to construct the pointwise confidence intervals for the coefficient functions. Third, we extend our proposed estimation procedure to include partially linear varying-coefficient models with interactive fixed effects, and study the asymptotic properties of the resulting estimator. In addition, we develop a test statistic for the constancy of the varying coefficient functions, and propose a bootstrap procedure to evaluate the null distribution of the test statistic. Finally, numerical studies demonstrate the satisfactory performance of our proposed methods in practice, and support our theoretical results.

Supplementary Material

The online Supplementary Material contains the numerical studies, proofs of Theorems 1–6 and Corollary 1, and Lemmas 1–7 and their proofs. In addition, we introduce the estimation procedure for a special model, namely, the varying-coefficient panel-data model with additive fixed effects.

Acknowledgments

The authors sincerely thank the Editor, Associate Editor, and two anonymous reviewers for their insightful comments and suggestions. Sanying Feng's research was supported by the National Statistical Science Research Project of China (No. 2019LY18), the Foundation of Henan Educational Committee (No.21A910004), and the Training Fund for Basic Research Program of Zhengzhou University (No. 32211591). Gaorong Li's research was supported by NSFC (Nos. 11871001 and 11971001), the Beijing Natural Science Foundation (No. 1182003), and the Fundamental Research Funds for the Central Universities (No. 2019NTSS18). Heng Peng's research was supported by GRF Grants of the Research Grants Council of Hong Kong (Nos. HKBU12302615 and HKBU 12303618) and NSFC (Nos. 11871409 and 11971018). Tiejun Tong's research was supported by NSFC (No. 11671338), RGC Grant (No. HKBU12303918), and FNRA Fund (No. RC-IG-FNRA/17-18/13).

References

Ahn, S. G., Lee, Y. H. and Schmidt, P. (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 University Press, Oxford.

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

Bai, J. S. and Li, K. P. (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*. Wiley, New York.

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

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

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

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

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

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

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

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

Huang, J. Z., Wu, C. O. and Zhou, L. (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 Shintani, M. (2006). Bootstrapping GMM estimators for time series. *J. Econometrics* **133**, 531–555.

Jin, S. N. and Su, L. J. (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., Moon, H. R. and Weidner, M. (2012). Analysis of interactive fixed effects dynamic linear panel regression with measurement error. *Econom. Lett.* **117**, 239–242.

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

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

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

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

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

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

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

Noh, H. S. and Park, B. U. (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 Soberon, A. (2014). Direct semi-parametric estimation of fixed effects panel data varying coefficient models. *Econom. J.* **17**, 107–138.

Rodriguez-Poo, J. M. and Soberon, A. (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 Chen, Q. H. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory* **29**, 1079–1135.

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

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

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

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

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

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

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

Sanying Feng

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

E-mail: fsy5801@zzu.edu.cn

Gaorong Li

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

E-mail: ligaorong@bnu.edu.cn

Heng Peng

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong.

E-mail: hpeng@hkbu.edu.hk

Tiejun Tong

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong.

E-mail: tongt@hkbu.edu.hk

(Received June 2018; accepted August 2019)

Statistica Sinica: Supplement

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*

Supplementary Material

This is a supplement to the paper "Varying-Coefficient Panel Data Model with Interactive Fixed Effects", in which it contains the numerical studies, proofs of Theorems 1–6 and Corollary 1, and Lemmas 1–7 and their proofs. In addition, we introduce the estimation procedure for a special model, namely, varying-coefficient panel-data model with additive fixed effects.

S1 Appendix A: Numerical studies

In Appendix A, some simulation examples and a real data are analyzed to augment the derived theoretical results in the main context.

S1.1 Choice of smoothing parameters

We develop a data-driven procedure to choose the smoothing parameters L_k, for $k = 1, \ldots, p$, where L_k control the smoothness of $\beta_k(u)$. In practice, various smoothing methods can be applied to select the smoothing parameters, such as the cross validation (CV), the generalized cross validation (GCV), or the Bayesian information criterion (BIC). Following Huang et al. (2002), we propose a modified "leave-one-subject-out" CV to automatically select the smoothing parameters L_k by minimizing the following CV score:

$$\mathrm{CV} = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i \hat{\boldsymbol{\gamma}}^{(-i)})^{\tau} M_{\hat{\boldsymbol{F}}^{(-i)}} (\boldsymbol{Y}_i - \boldsymbol{R}_i \hat{\boldsymbol{\gamma}}^{(-i)}), \tag{A.1}$$

where $\hat{\boldsymbol{\gamma}}^{(-i)}$ and $\hat{\boldsymbol{F}}^{(-i)}$ are the estimators defined by solving the nonlinear equations (2.7) and (2.8) from data with the ith subject deleted. In fact, the CV score in (A.1) can also be viewed as a weighted estimate of the true prediction error. The performance of the modified "leave-one-subject-out" CV procedure will be evaluated in the next section.

To determine the number r of the factors, we adopt BIC in Li et al. (2016):

$$\mathrm{BIC}(r) = \ln(V(r, \dot{\boldsymbol{\gamma}}_r)) + r \frac{(N+T)\sum_{k=1}^{p} L_k}{NT} \ln\left(\frac{NT}{N+T}\right), \tag{A.2}$$

where $\dot{\boldsymbol{\gamma}}_r$ is the estimator of $\boldsymbol{\gamma}$, and $V(r, \dot{\boldsymbol{\gamma}}_r)$ is defined as

$$V(r, \dot{\boldsymbol{\gamma}}_r) = \frac{1}{NT} \sum_{\varrho=r+1}^{T} \mu_{\varrho}\left(\sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{R}_i \dot{\boldsymbol{\gamma}}_r)(\boldsymbol{Y}_i - \boldsymbol{R}_i \dot{\boldsymbol{\gamma}}_r)^{\tau} \right). \tag{A.3}$$

In (A.3), $\mu_{\varrho}(A)$ denotes the ϱ-th largest eigenvalue of a symmetric matrix A by counting multiple eigenvalues multiple times. We set $r_{\max} = 8$, and choose the number r of the factors by minimizing the objective function $\mathrm{BIC}(r)$ in (A.2), that is, $\hat{r} = \arg\min_{0\le r\le r_{\max}} \mathrm{BIC}(r)$.

S1.2 Simulation studies

In this section, we conduct simulation studies to assess the finite sample performance of our proposed methods.

Example 1 (Varying-coefficient model). In this example, we generate data from the following model:

$$Y_{it} = X_{it,1}\beta_1(U_{it}) + X_{it,2}\beta_2(U_{it}) + \lambda_i^\tau F_t + \varepsilon_{it}, \qquad \text{(A.4)}$$

where $\lambda_i = (\lambda_{i1}, \lambda_{i2})^\tau$, $F_t = (F_{t1}, F_{t2})^\tau$, $\beta_1(u) = 2 - 5u + 5u^2, \beta_2(u) = \sin(u\pi)$, $U_{it} = \omega_{it} + \omega_{i,t-1}$, and ω_{it} are i.i.d. random errors from the uniform distribution on $[0, 1/2]$. As the regressors $X_{it,1}$ and $X_{it,2}$ are correlated with λ_i, F_t, and their product $\lambda_i^\tau F_t$, we generate them according to

$$X_{it,1} = 1 + \lambda_i^\tau F_t + \iota^\tau\lambda_i + \iota^\tau F_t + \eta_{it,1}, \quad X_{it,2} = 1 + \lambda_i^\tau F_t + \iota^\tau\lambda_i + \iota^\tau F_t + \eta_{it,2},$$

where $\iota = (1, 1)^\tau$, the effects λ_{ij}, F_{tj}, $j = 1,\ 2$, $\eta_{it,1}$ and $\eta_{it,2}$ are all independently from $N(0, 1)$. Lastly, the regression error ε_{it} are generated i.i.d. from $N(0, 4)$.

As a standard measure of the estimation accuracy, the performance of the estimator $\hat{\boldsymbol{\beta}}(\cdot)$ will be assessed by the integrated squared error (ISE):

$$\mathrm{ISE}(\hat{\beta}_k) = \int \{\hat{\beta}_k(u) - \beta_k(u)\}^2 f(u)\mathrm{d}u, \quad k = 1,\ 2.$$

We further approximate the ISE by the average mean squared error (AMSE):

$$\mathrm{AMSE}(\hat{\beta}_k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\hat{\beta}_k(U_{it}) - \beta_k(U_{it})]^2, \quad k = 1,\ 2. \qquad \text{(A.5)}$$

Throughout the simulations, we use the cubic B-splines as the basis functions. Thus $L_k = l_k + m + 1$, where l_k is the number of interior knots and $m = 3$ is the degree of the spline. For simplicity, we use the equally spaced knots for all numerical studies. To implement the estimation procedure, we select L_k by minimizing the modified "leave-one-subject-out" CV score in (A.1), and determine the number r of the factors using the BIC-type criterion (A.2).

For comparison, we compute the AMSEs in (A.5) by three estimation procedures, and report their numerical results in Table 1 based on 1000 repetitions. The column with label "IE" denotes the infeasible estimators, which are obtained by assuming observable F_t. The column with label "IFE" denotes the interactive fixed effects estimators obtained by our proposed procedure in Section 2. Finally, the column with label "LSDVE" denotes the least squares dummy variable estimators, which are obtained under the false assumption with additive fixed effects in model (A.4) by applying the least squares dummy variable method (see Section S4 for details).

Table 1: Finite sample performance of the estimators for model (A.4).

		IE		IFE		LSDVE	
N	T	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)
100	15	0.0091	0.0092	0.0102	0.0103	0.0947	0.0918
100	30	0.0045	0.0044	0.0047	0.0048	0.0878	0.0909
100	60	0.0021	0.0020	0.0022	0.0022	0.0844	0.0829
100	100	0.0012	0.0012	0.0013	0.0013	0.0830	0.0822
60	100	0.0020	0.0020	0.0021	0.0022	0.0848	0.0838
30	100	0.0043	0.0042	0.0047	0.0048	0.0864	0.0873
15	100	0.0082	0.0083	0.0102	0.0102	0.0946	0.0910

From Table 1, we note that both the infeasible estimators and the interactive fixed effects estimators are consistent, and the results of the latter are gradually closer to those of the former as both N and T increase. However, the least squares dummy variable estimators of the coefficient functions are biased and inconsistent. One possible reason is that the interactive fixed effects are correlated with the regressors and cannot be removed by the least squares dummy variable method. In addition, AMSEs decrease significantly as both N and T increase for the infeasible estimators and the interactive fixed effects estimators.

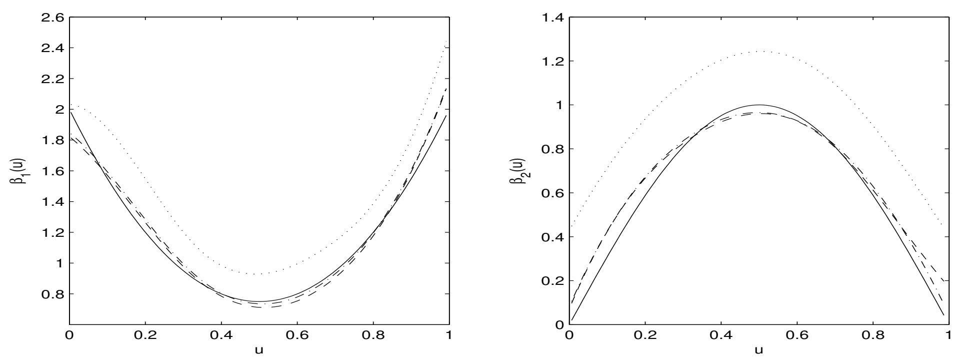

Figure 1: *Simulation results for model* (A.4) *when* $N = 100$, $T = 60$. *In each plot, the solid curves are for the true coefficient functions, the dash-dotted curves are for the interactive fixed effects estimators (IFE), the dashed curves are for the infeasible estimators (IE), the dotted curves are for the least squares dummy variable estimators (LSDVE).*

Figure 1 presents the estimated curves of $\beta_1(\cdot)$ and $\beta_2(\cdot)$ from a typical sample, in which the typical sample is selected such that its AMSE is equal to the median of the 1000 replications. It is also found that the infeasible estimators and the interactive fixed effects estimators are close to the true coefficient functions, whereas the least squares dummy variable estimators are biased.

To construct the 95% pointwise confidence intervals for $\beta_1(\cdot)$ and $\beta_2(\cdot)$ using the residual-based block bootstrap procedure in Section 4, we generate 1000 bootstrap samples based on the typical sample, and we choose the block length l by the criterion $l = T^{1/3}$. The 95% bootstrap pointwise confidence intervals of $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are given in Figure 2. Overall, the proposed residual-based block bootstrap procedure works quite well.

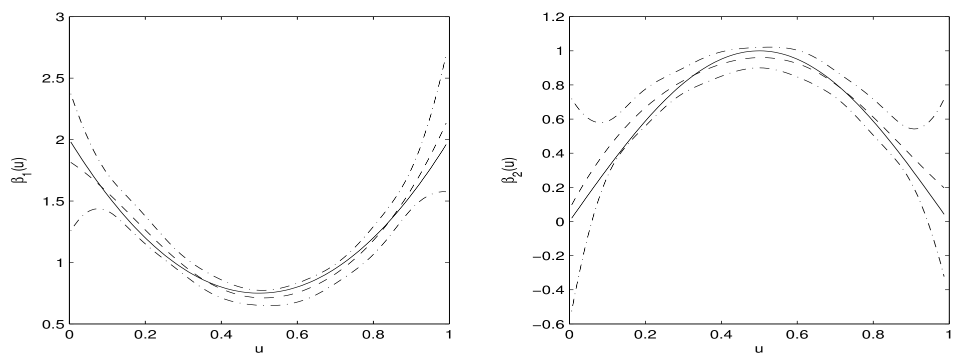

Figure 2: 95% *pointwise confidence intervals for* $\boldsymbol{\beta}(\cdot)$ *when* $N = 100$, $T = 60$. *In each plot, the solid curves are for the true coefficient functions, the dashed curves are for the interactive fixed effects estimators, the dash-dotted curves are for the* 95% *pointwise confidence intervals based on bootstrap procedure.*

Our next study is to investigate the performance of our proposed methods when the fixed effects are additive. Letting $\lambda_i = (\mu_i, 1)^\tau$ and $F_t = (1, \xi_t)^\tau$, we have $\lambda_i^\tau F_t = \mu_i + \xi_t$. We then consider the following varying-coefficient panel-data model with additive fixed effects:

$$Y_{it} = X_{it,1}\beta_1(U_{it}) + X_{it,2}\beta_2(U_{it}) + \mu_i + \xi_t + \varepsilon_{it}, \tag{A.6}$$

where $\beta_1(u)$, $\beta_2(u)$, U_{it}, and ε_{it} are the same as those in model (A.4). The regressors $X_{it,1}$ and $X_{it,2}$ are generated according to $X_{it,1} = 3+2\mu_i+2\xi_t+\eta_{it,1}$ and $X_{it,2} = 3 + 2\mu_i + 2\xi_t + \eta_{it,2}$, where $\eta_{it,j} \sim N(0, 1)$, $j = 1$, 2, and the fixed effects are generated by

$$\mu_i \sim N(0,1), \quad i = 2, \ldots, N \quad \text{and} \quad \mu_1 = -\sum_{i=2}^{N} \mu_i,$$

$$\xi_t \sim N(0,1), \quad t = 2, \ldots, T \quad \text{and} \quad \xi_1 = -\sum_{t=2}^{T} \xi_t.$$

With 1000 repetitions, we report the simulation results in Table 2, Figure 3 and Figure 4, respectively. To be specific, Table 2 presents the finite sample performance of the estimators for model (A.6) with additive fixed effects, Figure 3 displays the estimated curves of the three estimators for the coefficient functions, and Figure 4 displays the 95% bootstrap pointwise confidence intervals for $\beta_1(\cdot)$ and $\beta_2(\cdot)$ when $N = 100$ and $T = 60$.

Table 2: Finite sample performance of the estimators for model (A.6) with additive fixed effects.

		IE		IFE		LSDVE	
N	T	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)
100	15	0.0102	0.0102	0.0267	0.0260	0.0083	0.0083
100	30	0.0048	0.0048	0.0224	0.0216	0.0040	0.0040
100	60	0.0022	0.0023	0.0192	0.0198	0.0020	0.0019
100	100	0.0013	0.0013	0.0171	0.0176	0.0011	0.0011
60	100	0.0022	0.0022	0.0214	0.0226	0.0019	0.0019
30	100	0.0046	0.0045	0.0271	0.0281	0.0040	0.0040
15	100	0.0089	0.0090	0.0340	0.0343	0.0083	0.0083

Table 2 and Figure 3 show that the infeasible estimators, the interactive fixed effects estimators, and the least squares dummy variable estimators are all consistent. Our proposed interactive fixed effects estimators remain valid even for the varying-coefficient panel-data model with additive fixed effects. However, they are less efficient than the least squares dummy variable estimators. Finally, the 95% bootstrap pointwise confidence intervals

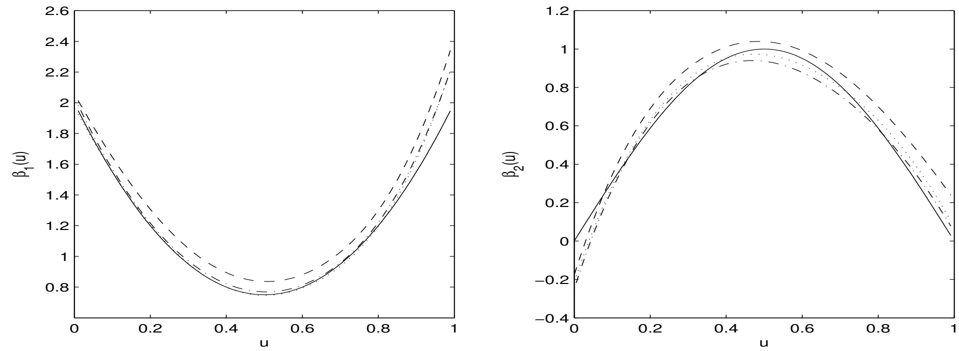

Figure 3: *Simulation results for model* (A.6) *with additive fixed effects when* $N = 100$, $T = 60$. *In each plot, the solid curves are for the true coefficient functions, the dash-dotted curves are for the interactive fixed effects estimators, the dashed curves are for the infeasible estimators, the dotted curves are for the least squares dummy variable estimators.*

for the typical estimates of $\beta_1(\cdot)$ and $\beta_2(\cdot)$ in Figure 4 demonstrate the validity and effectiveness of our proposed methods.

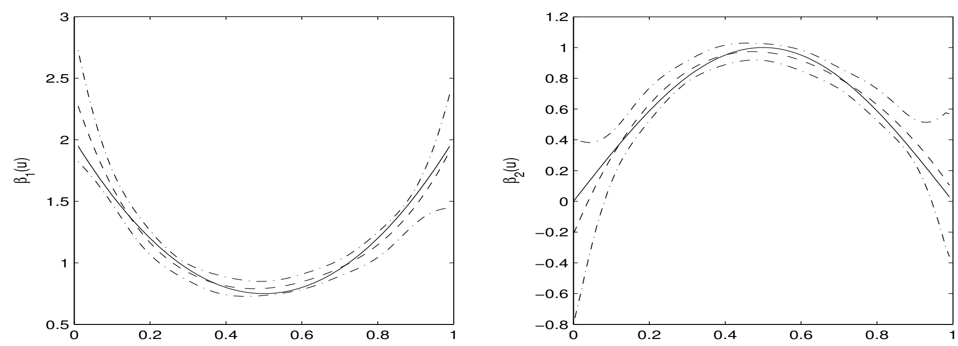

Figure 4: 95% *pointwise confidence intervals for* $\boldsymbol{\beta}(\cdot)$ *when* $N = 100$, $T = 60$. *In each plot, the solid curves are for the true coefficient functions, the dashed curves are for the interactive fixed effects estimators, the dash-dotted curves are for the* 95% *pointwise confidence intervals based on bootstrap procedure.*

Example 2 (Lagged dependent variables case). In this example, we consider the following varying-coefficient panel-data model with lagged dependent variables as follows:

$$Y_{it} = Y_{i,t-1}\alpha(U_{it}) + X_{it,1}\beta_1(U_{it}) + X_{it,2}\beta_2(U_{it}) + \lambda_i^\tau F_t + \varepsilon_{it}, \qquad \text{(A.7)}$$

where $i = 1, \ldots, N$, $t = 2, \ldots, T$, $\alpha(u) = \cos(u\pi)$, $X_{it,1}$, $X_{it,2}$, U_{it}, λ_i, and F_t are generated as in model (A.4). Table 3 presents the results for model (A.7), and the estimated results show that the proposed method works well even for model (A.7) with lagged dependent variables.

Table 3: Finite sample performance of the estimators for model (A.7).

		IE			IFE		
N	T	AMSE($\hat{\alpha}$)	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)	AMSE($\hat{\alpha}$)	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)
100	15	0.0114	0.0109	0.0105	0.0124	0.0117	0.0118
100	30	0.0073	0.0068	0.0069	0.0082	0.0078	0.0075
100	60	0.0039	0.0035	0.0033	0.0041	0.0041	0.0039
100	100	0.0022	0.0023	0.0024	0.0026	0.0027	0.0025
60	100	0.0038	0.0036	0.0032	0.0040	0.0043	0.0038
30	100	0.0071	0.0072	0.0067	0.0084	0.0078	0.0078
15	100	0.0112	0.0108	0.0106	0.0125	0.0116	0.0115

Example 3 (Partially linear varying-coefficient model). In this example, we generate data from the following model:

$$Y_{it} = X_{it,1}\beta_1(U_{it}) + X_{it,2}\beta_2 + X_{it,3}\beta_3 + \lambda_i^\tau F_t + \varepsilon_{it}, \qquad \text{(A.8)}$$

where $\beta_1(u) = \sin(u\pi)$, $\beta_2 = 3$, $\beta_3 = 2.5$ and $X_{it,3} = 2 + \lambda_i^\tau F_t + \iota^\tau \lambda_i + \iota^\tau F_t + \eta_{it,3}$ with $\eta_{it,3} \sim N(0, 1)$. The regression error ε_{it} is generated as AR(1) for each fixed i such that $\varepsilon_{it} = 0.7\varepsilon_{i,t-1} + \epsilon_{it}$, where ϵ_{it} is i.i.d. $N(0, 1)$. Further, we use the other settings in model (A.4). The summary of simulation results is reported in Table 4.

Table 4 indicates that, although there is serial correlation in the error terms, the interactive fixed effects estimators are gradually closer to the

Table 4: Finite sample performance of the estimators for model (A.8).

		IE					IFE				
N	T	AMSE($\hat{\beta}_1$)	Mean($\hat{\beta}_2$)	SD($\hat{\beta}_2$)	Mean($\hat{\beta}_3$)	SD($\hat{\beta}_3$)	AMSE($\hat{\beta}_1$)	Mean($\hat{\beta}_2$)	SD($\hat{\beta}_2$)	Mean($\hat{\beta}_3$)	SD($\hat{\beta}_3$)
100	15	0.0109	2.9891	0.0960	2.4872	0.0891	0.0152	3.2104	0.1269	2.6517	0.1174
100	30	0.0069	3.0096	0.0715	2.5081	0.0712	0.0106	3.1017	0.0922	2.5953	0.0918
100	60	0.0044	2.9912	0.0482	2.5066	0.0473	0.0058	3.0192	0.0541	2.5153	0.0536
100	100	0.0028	3.0051	0.0256	2.5039	0.0237	0.0030	3.0079	0.0363	2.4966	0.0344
60	100	0.0032	3.0068	0.0325	2.5052	0.0331	0.0037	3.0087	0.0391	2.5074	0.0395
30	100	0.0051	3.0079	0.0433	2.5068	0.0442	0.0060	3.0098	0.0494	2.5091	0.0497
15	100	0.0092	3.0091	0.0558	2.4917	0.0563	0.0097	3.0112	0.0607	2.5135	0.0618

infeasible estimators as both N and T increase. However, for small T, the estimators are inconsistent. The simulation results are consistent with the theoretical results.

To demonstrate the power of the test, for model (A.8), we consider the null hypothesis H_0: $\beta_2(u) = 3$, $\beta_3(u) = 2.5$, against the alternative hypothesis H_1: $\beta_2(u) = 3 + c_0(2 - 5u + 5u^2)$, $\beta_3(u) = 2.5 + c_0 \cos(\pi u)$, where c_0 determines the extent that $\beta_j(u)$ varies with u. We set $c_0 = 0,\ 0.06,\ 0.12,\ \ldots,\ 0.66$. If $c_0 = 0$, the alternative hypothesis becomes the null hypothesis. For sample size N=100 and $T = 60$, we generate 1000 samples under H_1, and use 1000 bootstrap replications for the bootstrap procedure in Section 6. Figure 5 reports the estimated power function curves with the significance level $\alpha_0 = 0.05$.

From Figure 5, we have the following results. (1) The size of our test is close to the nominal 5% when the null hypothesis holds ($c_0 = 0$). This demonstrates that the bootstrap estimate of the null distribution is approximately correct. (2) When the alternative hypothesis is true ($c_0 > 0$), the power functions increase rapidly as c_0 increases. These results show that

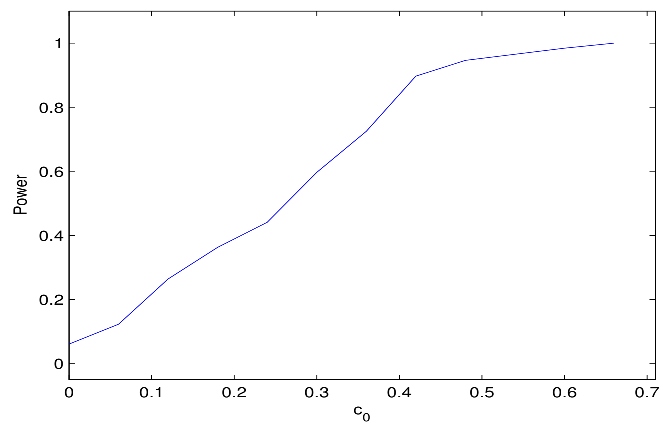

Figure 5: *The simulated power function for sample size* $N = 100$ *and* $T = 60$.

the proposed test statistic performs well.

S1.3 Application to a real dataset

We apply our proposed methods to a real dataset from the UK Met Office that contains the monthly mean maximum temperatures (in Celsius degrees), the mean minimum temperatures (in Celsius degrees), the days of air frost (in days), the total rainfall (in millimeters), and the total sunshine duration (in hours) from 37 stations. For this dataset, one main goal is to investigate the impact of other factors on the mean maximum temperatures across different stations. Li et al. (2011) analyzed the effect of the total rainfall and the sunshine duration on the mean maximum temperatures. By contrast, we take into account the days of air frost. Data from 21 stations during the period of January 2005 to December 2014 are selected while, as the record values for the other stations missed too much, we drop them from further analysis.

Because there exists the seasonal variation in this dataset, our first step is to remove the seasonality from the observations. We impose the additive decomposition on time series objects and then subtract the seasonal term from the corresponding time series objects. Let Y_{it} be the seasonally adjusted monthly mean maximum temperatures in the tth month in station i, $X_{it,1}$ be the seasonally adjusted monthly days of air frost, $X_{it,2}$ be the seasonally adjusted monthly total rainfall, and $X_{it,3}$ be the seasonally adjusted monthly total sunshine duration. To analyze the dataset, we consider the following varying-coefficient panel-data model with interactive fixed effects:

$$Y_{it} = X_{it,1}\beta_1(t/T) + X_{it,2}\beta_2(t/T) + X_{it,3}\beta_3(t/T) + \lambda_i^{\tau} F_t + \varepsilon_{it}, \quad \text{(A.9)}$$

where $1 \leq i \leq 21$, $1 \leq t \leq 120$, and the multi-factor error structure $\lambda_i^{\tau} F_t + \varepsilon_{it}$ is used to control the heterogeneity and to capture the unobservable common effects.

Note that the objectives of the study are to estimate the trend effects of the days of air frost, the monthly total rainfall and the sunshine duration over time. To achieve the goals, we fit model (A.9) using the cubic splines with equally spaced knots, and select the numbers of interior knots for the unknown coefficient functions by minimizing the modified "leave-one-subject-out" CV score in (A.1). Moreover, the number r of the factors is determined according to the BIC-type criterion (A.2). The estimated curves and 95% bootstrap pointwise confidence intervals of $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $\beta_3(\cdot)$ are plotted in Figure 6 based on the proposed methods.

The estimated trend curve in Figure 6 shows that the estimate of $\beta_1(\cdot)$ is almost flat, thus we assume that the effect of $X_{it,1}$ is time-invariant and test the constancy of the coefficient function $\beta_1(\cdot)$. Based on the proposed

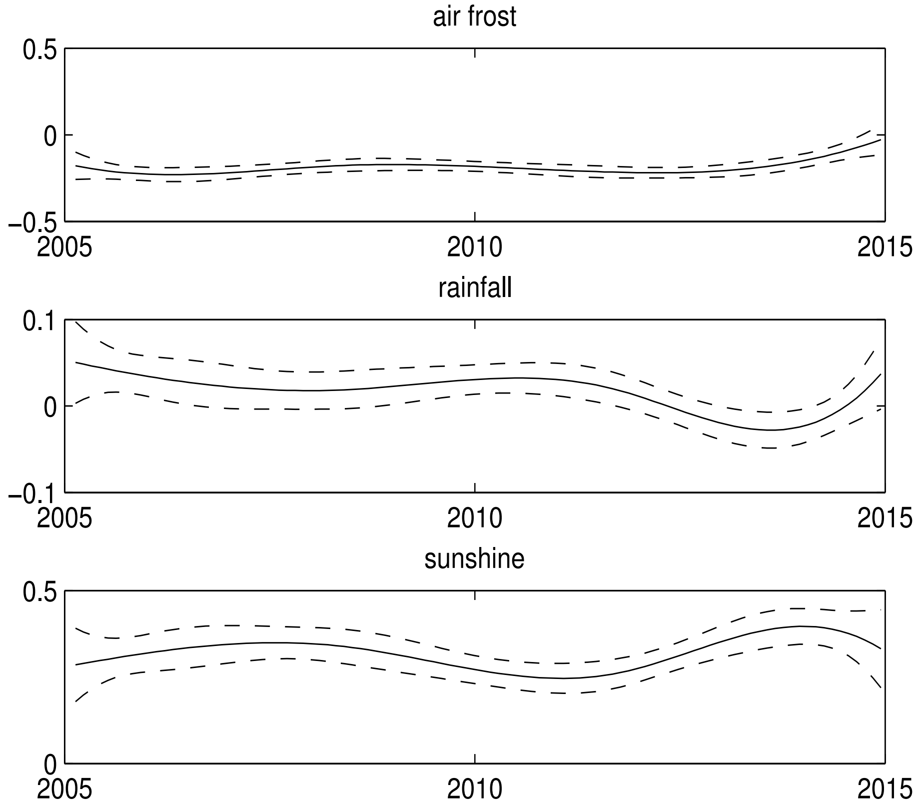

Figure 6: *The estimated curves and 95% pointwise confidence intervals of* $\beta_1(\cdot)$*,* $\beta_2(\cdot)$ *and* $\beta_3(\cdot)$*. In each plot, the solid curves are for the interactive fixed effects estimators, the dashed curves denote the* 95% *pointwise confidence intervals.*

bootstrap test procedure, we generate 1000 bootstrap samples and obtain the p-value of the test is 0.133 at the significance level 5%. This motivates us consider the following partially linear varying-coefficient panel-data model with interactive fixed effects:

$$Y_{it} = X_{it,1}\beta_1 + X_{it,2}\beta_2(t/T) + X_{it,3}\beta_3(t/T) + \lambda_i^{\tau} F_t + \varepsilon_{it}, \qquad \text{(A.10)}$$

We apply the proposed estimation procedure in Section 5 to model

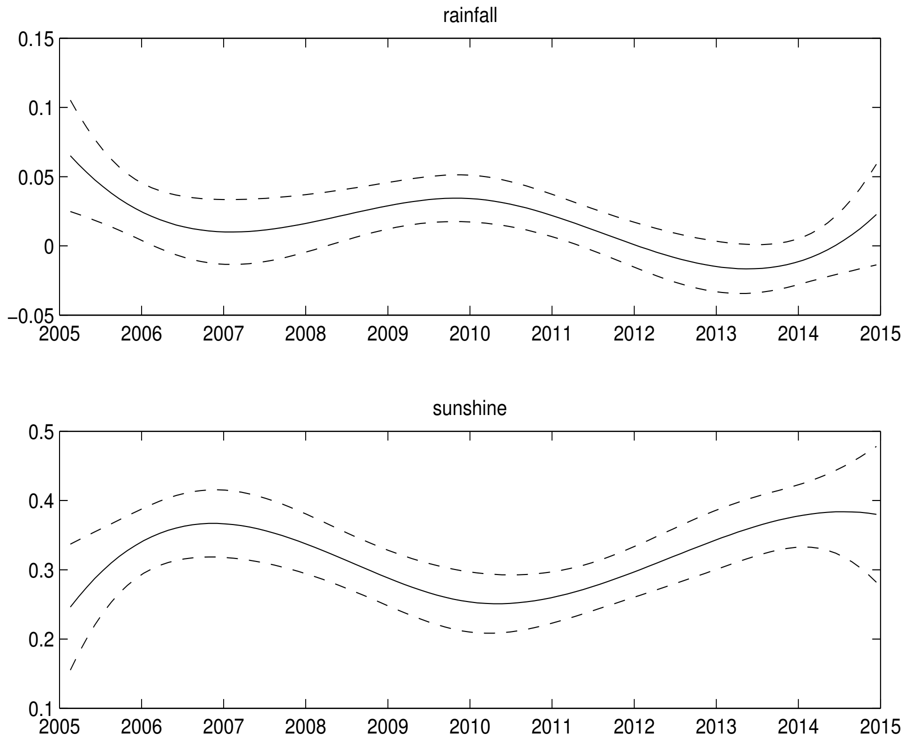

Figure 7: *The estimated curves and 95% pointwise confidence intervals of* $\beta_2(\cdot)$ *and* $\beta_3(\cdot)$ *in model (A.10) . In each plot, the solid curves are for the interactive fixed effects estimators, the dashed curves denote the* 95% *pointwise confidence intervals.*

(A.10) and obtain that the estimate of β_1 is -0.1915, which means there is a negative effect of monthly days of air frost on monthly mean maximum temperatures. The estimated curves and 95% bootstrap pointwise confidence intervals of $\beta_2(\cdot)$ and $\beta_3(\cdot)$ are given in Figure 7. From Figure 7, we can see that the estimated curves of $\beta_2(\cdot)$ and $\beta_3(\cdot)$ are all oscillating over time, and the effect of the monthly total sunshine duration is obviously above zero, which shows that the monthly total sunshine duration has an overall positive effect on the monthly mean maximum temperatures.

S2 Appendix B: Proofs of theorems

We provide the proofs of Theorems 1–6 and Corollary 1 in Appendix B.

For the ease of the presentation, let C denote some positive constants not depending on N and T, but which may assume different values at each appearance. In the proof, we use the following properties of B-spline (see de Boor (2001)): (1) $B_{kl}(u) \geq 0$ and $\sum_{l=1}^{L_k} B_{kl}(u) = 1$, for $u \in \mathcal{U}$ and $k = 1, \ldots, p$. (2) There exist constants $0 < M_1, M_2 < \infty$, not depending on L_k, such that

$$M_1 L_k^{-1} \sum_{l=1}^{L_k} \gamma_{kl}^2 \leq \int_{\mathcal{U}} \Big[\sum_{l=1}^{L_k} \gamma_{kl} B_{kl}(u) \Big]^2 \mathrm{d}u \leq M_2 L_k^{-1} \sum_{l=1}^{L_k} \gamma_{kl}^2,$$

for any sequence $\{\gamma_{kl} \in \mathbb{R} : l = 1, \ldots, L_k\}$.

From Assumptions (A1)–(A4) and Corollary 6.21 in Schumaker (1981), there exists a constant $M > 0$ such that

$$\begin{aligned} \beta_k(u) &= \sum_{l=1}^{L_k} \widetilde{\gamma}_{kl} B_{kl}(u) + Re_k(u), \\ \sup_{u \in \mathcal{U}} |Re_k(u)| &\leq M L_k^{-d}, \quad k = 1, \ldots, p. \end{aligned} \tag{B.1}$$

Let $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{iT})^\tau$ with $e_{it} = \sum_{k=1}^p Re_k(U_{it}) X_{it,k}$, and $\widetilde{\boldsymbol{\gamma}} = (\widetilde{\boldsymbol{\gamma}}_1^\tau, \ldots, \widetilde{\boldsymbol{\gamma}}_p^\tau)^\tau$ with $\widetilde{\boldsymbol{\gamma}}_k = (\widetilde{\gamma}_{k1}, \ldots, \widetilde{\gamma}_{kL_k})^\tau$. Then $\boldsymbol{Y}_i = \boldsymbol{R}_i \widetilde{\boldsymbol{\gamma}} + \boldsymbol{F}^0 \lambda_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i$, for $i = 1, \ldots, N$. We use the following facts throughout the paper: $\|\boldsymbol{F}^0\| = O_P(T^{1/2})$, $\|\boldsymbol{R}_i\| = O_P(T^{1/2})$ for all i, and $(NT)^{-1} \sum_{i=1}^N \|\boldsymbol{R}_i\|^2 = O_P(1)$. Note that $\|\hat{\boldsymbol{F}}\| = T^{1/2}\sqrt{r}$. For ease of notation, we define $\delta_{NT} = \min[\sqrt{N}, \sqrt{T}]$ and $\zeta_{Ld} = \sum_{k=1}^p L_k^{-2d}$. Following the notation of Huang et al. (2004), we write $a_n \asymp b_n$ if both a_n and b_n are positive and a_n/b_n and b_n/a_n are bounded for all n.

Proof We only give the proof of $\|\boldsymbol{R}_i\| = O_P(T^{1/2})$, and omit the proofs of $\|\boldsymbol{F}^0\| = O_P(T^{1/2})$ and $(NT)^{-1}\sum_{i=1}^N \|\boldsymbol{R}_i\|^2 = O_P(1)$.

$$
\begin{aligned}
E(\|\boldsymbol{R}_i\|^2) &= E\Big(\mathrm{tr}(\boldsymbol{R}_i\boldsymbol{R}_i^\tau)\Big) = E\left(\sum_{t=1}^T \|X_{it}^\tau \boldsymbol{B}(U_{it})\|^2\right) \\
&= E\left(\sum_{t=1}^T\sum_{k=1}^p\sum_{l=1}^{L_k} X_{it,k}^2 B_{kl}^2(U_{it})\right) = \sum_{t=1}^T\sum_{k=1}^p\sum_{l=1}^{L_k} E\Big(X_{it,k}^2 B_{kl}^2(U_{it})\Big).
\end{aligned}
$$

By Assumption (A1), we have $E\Big(X_{it,k}^2 B_{kl}^2(U_{it})\Big) \le CE\Big(B_{kl}^2(U_{it})\Big)$. Moreover, by the properties of B-spline, we can get that

$$
\sum_{l=1}^{L_k} B_{kl}^2(u) \le \left(\sum_{l=1}^{L_k} B_{kl}(u)\right)^2 = 1.
$$

Then we have $E(\|\boldsymbol{R}_i\|^2) = O(T)$, which implies that $\|\boldsymbol{R}_i\| = O_P(T^{1/2})$, for all i. □

S2.1 Proof of Theorem 1

Without loss of generality, we assume that $\boldsymbol{\beta}(\cdot) = 0$. Then $\boldsymbol{Y}_i = \boldsymbol{F}^0\lambda_i + \boldsymbol{\varepsilon}_i$, for $i = 1, \ldots, N$. By Lemma 2, we have

$$
\begin{aligned}
Q_{NT}(\boldsymbol{\gamma}, \boldsymbol{F}) &= \frac{1}{NT}\sum_{i=1}^N (\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma})^\tau M_{\boldsymbol{F}}(\boldsymbol{Y}_i - \boldsymbol{R}_i\boldsymbol{\gamma}) \\
&= \boldsymbol{\gamma}^\tau\left(\frac{1}{NT}\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}}\boldsymbol{R}_i\right)\boldsymbol{\gamma} + \mathrm{tr}\left[\left(\frac{\boldsymbol{F}^{0\tau}M_{\boldsymbol{F}}\boldsymbol{F}^0}{T}\right)\left(\frac{\Lambda^\tau\Lambda}{N}\right)\right] \\
&\quad -\frac{2}{NT}\boldsymbol{\gamma}^\tau\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}}\boldsymbol{F}^0\lambda_i - \frac{2}{NT}\boldsymbol{\gamma}^\tau\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i \\
&\quad +\frac{2}{NT}\sum_{i=1}^N \lambda_i^\tau\boldsymbol{F}^{0\tau}M_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^N \boldsymbol{\varepsilon}_i^\tau M_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i \\
&=: \widetilde{Q}_{NT}(\boldsymbol{\gamma}, \boldsymbol{F}) + o_P(1),
\end{aligned}
$$

uniformly over bounded $\boldsymbol{\gamma}$ and over $\boldsymbol{F}$ such that $\boldsymbol{F}^\tau \boldsymbol{F}/T = I$, where

$$
\begin{aligned}
\widetilde{Q}_{NT}(\boldsymbol{\gamma}, \boldsymbol{F}) &= \boldsymbol{\gamma}^\tau \left(\frac{1}{NT} \sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{R}_i \right) \boldsymbol{\gamma} + \mathrm{tr} \left[\left(\frac{\boldsymbol{F}^{0\tau} M_{\boldsymbol{F}} \boldsymbol{F}^0}{T} \right) \left(\frac{\Lambda^\tau \Lambda}{N} \right) \right] \\
& \quad - \frac{2}{NT} \boldsymbol{\gamma}^\tau \sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{F}^0 \lambda_i.
\end{aligned}
$$

Let $\eta = \mathrm{vec}(M_{\boldsymbol{F}} \boldsymbol{F}^0)$, and

$$
A_1 = \frac{1}{NT} \sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}} \boldsymbol{R}_i, \quad A_2 = \left(\frac{\Lambda^\tau \Lambda}{N} \otimes I_T \right), \quad A_3 = \frac{1}{NT} \sum_{i=1}^N (\lambda_i^\tau \otimes M_{\boldsymbol{F}} \boldsymbol{R}_i).
$$

Then,

$$
\begin{aligned}
\widetilde{Q}_{NT}(\boldsymbol{\gamma}, \boldsymbol{F}) &= \boldsymbol{\gamma}^\tau A_1 \boldsymbol{\gamma} + \eta^\tau A_2 \eta - 2 \boldsymbol{\gamma}^\tau A_3^\tau \eta \\
&= \boldsymbol{\gamma}^\tau (A_1 - A_3^\tau A_2^{-1} A_3) \boldsymbol{\gamma} + (\eta^\tau - \boldsymbol{\gamma}^\tau A_3^\tau A_2^{-1}) A_2 (\eta - A_2^{-1} A_3 \boldsymbol{\gamma}) \\
&=: \boldsymbol{\gamma}^\tau D(\boldsymbol{F}) \boldsymbol{\gamma} + \theta^\tau A_2 \theta,
\end{aligned}
$$

where $\theta = \eta - A_2^{-1} A_3 \boldsymbol{\gamma}$. By Assumption (A5), $D(\boldsymbol{F})$ is a positive-definite matrix and A_2 is also a positive-definite matrix, which show that $\widetilde{Q}_{NT}(\boldsymbol{\gamma}, \boldsymbol{F}) \geq 0$. By the similar argument as in Bai (2009), it is easy to show that $\widetilde{Q}_{NT}(\boldsymbol{\gamma}, \boldsymbol{F})$ achieves its unique minimum at $(0, \boldsymbol{F}^0 H)$ for any $r \times r$ invertible matrix H. Thus, $\hat{\beta}_k(\cdot), k = 1, \ldots, p$, are uniquely defined. This completes the proof of part (i).

The proof of (ii) is similar to that of Proposition 1 (ii) in Bai (2009). To save space, we do not present the detailed proof. □

S2.2 Proof of Theorem 2

Since $\hat{\beta}_k(u) = \sum_{l=1}^{L_k} \hat{\gamma}_{kl} B_{kl}(u)$ and $\widetilde{\beta}_k(u) = \sum_{l=1}^{L_k} \widetilde{\gamma}_{kl} B_{kl}(u)$, by the properties of B-spline and (C.2), we have

$$\|\hat{\beta}_k(\cdot) - \beta_k(\cdot)\|_{L_2}^2 \leq 2\|\hat{\beta}_k(\cdot) - \widetilde{\beta}_k(\cdot)\|_{L_2}^2 + ML_k^{-2d},$$

and

$$\|\hat{\beta}_k(\cdot) - \widetilde{\beta}_k(\cdot)\|_{L_2}^2 = \|\hat{\boldsymbol{\gamma}}_k - \widetilde{\boldsymbol{\gamma}}_k\|_H^2 \asymp L_k^{-1}\|\hat{\boldsymbol{\gamma}}_k - \widetilde{\boldsymbol{\gamma}}_k\|^2, \ k = 1, \ldots, p, \tag{B.2}$$

where $\|\boldsymbol{\gamma}_k\|_H^2 = \boldsymbol{\gamma}_k^\tau \boldsymbol{H}_k \boldsymbol{\gamma}_k$, and $\boldsymbol{H}_k = (h_{ij})_{L_k \times L_k}$ is a matrix with entries $h_{ij} = \int_{\mathcal{U}} B_{ki}(u) B_{kj}(u) \mathrm{d}u$. Summing over k for (B.2), we obtain that

$$\|\hat{\boldsymbol{\beta}}(\cdot) - \widetilde{\boldsymbol{\beta}}(\cdot)\|_{L_2}^2 = \sum_{k=1}^{p} \|\hat{\boldsymbol{\gamma}}_k - \widetilde{\boldsymbol{\gamma}}_k\|_H^2 \asymp L_N^{-1}\|\hat{\gamma} - \widetilde{\boldsymbol{\gamma}}\|^2.$$

By (2.7) and $\boldsymbol{Y}_i = \boldsymbol{R}_i \widetilde{\boldsymbol{\gamma}} + \boldsymbol{F}^0 \lambda_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i$, for $i = 1, \ldots, N$, we have

$$\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}} = \Big(\sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{R}_i\Big)^{-1} \sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} (\boldsymbol{F}^0 \lambda_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i),$$

or equivalently,

$$\begin{aligned} &\Big(\sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{R}_i\Big)(\hat{\gamma} - \widetilde{\gamma}) \\ = \ &\sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{F}^0 \lambda_i + \sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{\varepsilon}_i + \sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{e}_i. \end{aligned} \tag{B.3}$$

We first deal with the third term of the right hand in (B.3). By Assumption (A1) and (C.2), and using the similar proofs to Lemma A.7 in Huang et al. (2004), and Lemmas 2 and 3, it is easy to show that

$$\left\| \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{e}_i \right\|^2 = O_P\Big(L_N^{-1} \zeta_{Ld}\Big). \tag{B.4}$$

For the first term of the right hand in (B.3), by noting that $M_{\hat{\boldsymbol{F}}}\hat{\boldsymbol{F}} = 0$, we have $M_{\hat{\boldsymbol{F}}}\boldsymbol{F}^0 = M_{\hat{\boldsymbol{F}}}(\boldsymbol{F}^0 - \hat{\boldsymbol{F}}H^{-1})$. By (B.3), we have

$$\boldsymbol{F}^0 - \hat{\boldsymbol{F}}H^{-1} = -(B_1 + B_2 + \cdots + B_{15})G, \tag{B.5}$$

where $H = (\Lambda^\tau\Lambda/N)(\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}/T)V_{NT}^{-1}$, $G = (\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}/T)^{-1}(\Lambda^\tau\Lambda/N)^{-1}$ is a matrix of fixed dimension and does not vary with i, and $B_1, \ldots, B_{15}$ are defined in Lemma 3. By (B.5), we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{F}^0\lambda_i &= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}(\boldsymbol{F}^0 - \hat{\boldsymbol{F}}H^{-1})\lambda_i \\
&= -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}(B_1 + B_2 + \cdots + B_{15})G\lambda_i \\
&=: J_1 + J_2 + \cdots + J_{15}.
\end{aligned}$$

It is easy to see that J_1–J_{15} depend on B_1–B_{15} respectively. For J_2, we have

$$\begin{aligned}
J_2 &= -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\left[\frac{1}{NT}\sum_{j=1}^{N}\boldsymbol{R}_j(\widetilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}})\lambda_j^\tau\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}\right]\left(\frac{\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}}{T}\right)^{-1}\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_i \\
&= \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{R}_j)\left[\lambda_j^\tau\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_i\right](\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}) \\
&= \frac{1}{T}\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{R}_j a_{ij}\right](\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}),
\end{aligned}$$

where $a_{ij} = \lambda_i^\tau(\Lambda^\tau\Lambda/N)^{-1}\lambda_j$. For J_1, we have

$$J_1 = -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}B_1G\lambda_i = o_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|).$$

For J_3, we have

$$J_3 = \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{R}_j\left(\frac{\boldsymbol{\varepsilon}_j^\tau\hat{\boldsymbol{F}}}{T}\right)G\lambda_i(\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}).$$

By Lemma 3 and some elementary calculations, we have

$$\begin{aligned} T^{-1}\boldsymbol{\varepsilon}_j^\tau \hat{\boldsymbol{F}} &= T^{-1}\boldsymbol{\varepsilon}_j^\tau \boldsymbol{F}^0 H + T^{-1}\boldsymbol{\varepsilon}_j^\tau (\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H) \\ &= O_P(T^{-1/2}) + T^{-1/2} O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2} T^{-1/2}\right). \end{aligned}$$

Using the above result and the similar argument as the proof of Lemma 2, it is easy to verify that $J_3 = o_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|)$. Similarly, we can obtain that $J_5 = o_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|)$. For J_4, we have

$$J_4 = -\frac{1}{N^2 T} \sum_{i=1}^N \sum_{j=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{F}^0 \lambda_j (\widetilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}})^\tau \left(\frac{\boldsymbol{R}_j^\tau \hat{\boldsymbol{F}}}{T} \right) G \lambda_i.$$

Noting that $M_{\hat{\boldsymbol{F}}} \boldsymbol{F}^0 = M_{\hat{\boldsymbol{F}}}(\boldsymbol{F}^0 - \hat{\boldsymbol{F}} H^{-1})$, and using Lemma 3 (i), that is, $T^{-1/2}\|\boldsymbol{F}^0 - \hat{\boldsymbol{F}} H^{-1}\| = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-1}) + O_P(\zeta_{Ld}^{1/2})$, we can obtain that $J_4 = o_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|)$. For J_6, noting that G is a matrix of fixed dimension and does not vary with i, and by $M_{\hat{\boldsymbol{F}}} \boldsymbol{F}^0 = M_{\hat{\boldsymbol{F}}}(\boldsymbol{F}^0 - \hat{\boldsymbol{F}} H^{-1})$, we have

$$\begin{aligned} J_6 &= -\frac{1}{N^2 T} \sum_{i=1}^N \sum_{j=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \boldsymbol{F}^0 \lambda_j \left(\frac{\boldsymbol{\varepsilon}_j^\tau \hat{\boldsymbol{F}}}{T} \right) G \lambda_i \\ &= -\frac{1}{NT} \sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} (\boldsymbol{F}^0 - \hat{\boldsymbol{F}} H^{-1}) \left[\frac{1}{N} \sum_{j=1}^N \lambda_j \left(\frac{\boldsymbol{\varepsilon}_j^\tau \hat{\boldsymbol{F}}}{T} \right) \right] G \lambda_i. \end{aligned}$$

By (B.6) and Lemma 3, we have

$$\begin{aligned} \frac{1}{NT} \sum_{j=1}^N \lambda_j \boldsymbol{\varepsilon}_j^\tau \hat{\boldsymbol{F}} &= \frac{1}{NT} \sum_{j=1}^N \lambda_j \boldsymbol{\varepsilon}_j^\tau \boldsymbol{F}^0 H + \frac{1}{NT} \sum_{j=1}^N \lambda_j \boldsymbol{\varepsilon}_j^\tau (\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H) \\ &= O_P((NT)^{-1/2}) + (TN)^{-1/2} O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(N^{-1}) \\ &\quad + N^{-1/2} O_P(\delta_{NT}^{-2}) + N^{-1/2} O_P\left(\zeta_{Ld}^{1/2}\right) \\ &= O_P((NT)^{-1/2}) + O_P(N^{-1}) + N^{-1/2} O_P(\delta_{NT}^{-2}) \\ &\quad + N^{-1/2} O_P\left(\zeta_{Ld}^{1/2}\right). \end{aligned}$$

By Lemma 3 (v), then

$$\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)=O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-2})+O_P(\zeta_{Ld}^{1/2}).$$

Moreover, the matrix G does not depend on i and $\|G\|=O_P(1)$, then

$$\begin{aligned}
J_6 &= \left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-2})+O_P\left(\zeta_{Ld}^{1/2}\right)\right]\\
&\quad\times\left[O_P((NT)^{-1/2})+O_P(N^{-1})+N^{-1/2}O_P(\delta_{NT}^{-2})+N^{-1/2}O_P\left(\zeta_{Ld}^{1/2}\right)\right]\\
&= o_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+o_P((NT)^{-1/2})+N^{-1}O_P(\delta_{NT}^{-2})+N^{-1/2}O_P(\delta_{NT}^{-4})\\
&\quad+N^{-1}O_P\left(\zeta_{Ld}^{1/2}\right)+N^{-1/2}O_P\left(\zeta_{Ld}\right).
\end{aligned}$$

For J_7, we have

$$J_7=-\frac{1}{N^2T}\sum_{i=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\left[\sum_{j=1}^{N}\boldsymbol{\varepsilon}_j\lambda_j^{\tau}\left(\frac{\Lambda^{\tau}\Lambda}{N}\right)^{-1}\right]\lambda_i=-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_j,$$

where $a_{ij}=\lambda_i^{\tau}(\Lambda^{\tau}\Lambda/N)^{-1}\lambda_j$. For J_8, by Assumption (A8), and the same argument as in the Proposition A.2 of Bai (2009), and Lemma 5, we have

$$\begin{aligned}
J_8 &= -\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}\hat{\boldsymbol{F}}G\lambda_i\\
&= -\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\Omega_j\hat{\boldsymbol{F}}G\lambda_i-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}-\Omega_j)\hat{\boldsymbol{F}}G\lambda_i\\
&=: A_{NT}+O_P(1/(T\sqrt{N}))+(NT)^{-1/2}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\left(\zeta_{Ld}^{1/2}\right)\right]\\
&\quad+\frac{1}{\sqrt{N}}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\left(\zeta_{Ld}^{1/2}\right)\right]^2,
\end{aligned}$$

where $A_{NT}=-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\Omega_j\hat{\boldsymbol{F}}G\lambda_i$. For J_9 and J_{10}, which depend on $\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}$. Using the same argument, it is easy to prove that J_9 and J_{10} are bounded in the Euclidean norm by $o_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)$. For J_{11}, using $M_{\hat{\boldsymbol{F}}}\boldsymbol{F}^0 = M_{\hat{\boldsymbol{F}}}(\boldsymbol{F}^0 - \hat{\boldsymbol{F}}H^{-1})$ again, and letting $\widetilde{\boldsymbol{W}}_j = \boldsymbol{e}_j^\tau \hat{\boldsymbol{F}}/T$ and $\|\widetilde{\boldsymbol{W}}_j\| = \|\boldsymbol{e}_j\|\sqrt{r}/\sqrt{T} = O_P(\zeta_{Ld}^{1/2})$, and using Lemma 3 (v), we have

$$\begin{aligned} J_{11} &= -\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{F}^0\lambda_j\left(\frac{\boldsymbol{e}_j^\tau\hat{\boldsymbol{F}}}{T}\right)G\lambda_i \\ &= -\frac{1}{NT}\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}(\boldsymbol{F}^0 - \hat{\boldsymbol{F}}H^{-1})\left[\frac{1}{N}\sum_{j=1}^N\lambda_j\left(\frac{\boldsymbol{e}_j^\tau\hat{\boldsymbol{F}}}{T}\right)\right]G\lambda_i \\ &= O_P\left(\zeta_{Ld}^{1/2}\right)\left[O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}\right)\right]. \end{aligned}$$

For J_{12}, similar to (B.4), we have

$$\begin{aligned} J_{12} &= -\frac{1}{N^2T}\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\left[\sum_{j=1}^N \boldsymbol{e}_j\lambda_j^\tau\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\right]\lambda_i \\ &= -\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N a_{ij}\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{e}_j = O_P\left(L_N^{-1/2}\zeta_{Ld}^{1/2}\right), \end{aligned}$$

where $a_{ij} = \lambda_i^\tau(\Lambda^\tau\Lambda/N)^{-1}\lambda_j$. Using the similar argument, it is easy to see that $J_{13} = (NT)^{-1/2}O_P(\zeta_{Ld}^{1/2})$.

For J_{14}, by (B.6) we have

$$\begin{aligned} J_{14} &= -\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{e}_j\left(\frac{\boldsymbol{\varepsilon}_j^\tau\hat{\boldsymbol{F}}}{T}\right)G\lambda_i \\ &= -\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{e}_j\left(\frac{\boldsymbol{\varepsilon}_j^\tau\boldsymbol{F}^0H}{T}\right)G\lambda_i \\ &\quad -\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}\boldsymbol{e}_j\left(\frac{\boldsymbol{\varepsilon}_j^\tau(\hat{\boldsymbol{F}} - \boldsymbol{F}^0H)}{T}\right)G\lambda_i. \end{aligned}$$

Similarly, we can prove that the first term of the above equation is bounded by $T^{-1/2}O_P(\zeta_{Ld}^{1/2})$. For the second term, by a similar argument and Lemma 4, we can prove that the second term is bounded above by

$$O_P\left(\zeta_{Ld}^{1/2}\right)\left[T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}T^{-1/2}\right)\right].$$

For J_{15}, by $M_{\hat{\boldsymbol{F}}}\hat{\boldsymbol{F}} = 0$ and some simple calculations, we have

$$J_{15} = -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\Big(\frac{\boldsymbol{e}_j\boldsymbol{e}_j^{\tau}}{T}\Big)\hat{\boldsymbol{F}}G\lambda_i = o_P(\zeta_{Ld}).$$

Summarizing the above results, we can obtain that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\boldsymbol{F}^0\lambda_i \\
= \;& J_2 + J_7 + A_{NT} + o_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + o_P((NT)^{-1/2}) + O_P\Big(\frac{1}{T\sqrt{N}}\Big) \\
&+N^{-1/2}O_P(\delta_{NT}^{-2}) + O_P\Big(T^{-1/2}\zeta_{Ld}^{1/2}\Big) + O_P\Big(L_N^{-1/2}\zeta_{Ld}^{1/2}\Big).
\end{aligned}$$

This leads to

$$\begin{aligned}
&\left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\boldsymbol{R}_i + o_P(1)\right)(\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}) - J_2 \\
= \;& \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i + J_7 + A_{NT} + o_P((NT)^{-1/2}) + O_P\Big(\frac{1}{T\sqrt{N}}\Big) \\
&\quad + N^{-1/2}O_P(\delta_{NT}^{-2}) + O_P\Big(T^{-1/2}\zeta_{Ld}^{1/2}\Big) + O_P\Big(L_N^{-1/2}\zeta_{Ld}^{1/2}\Big).
\end{aligned}$$

Multiplying $L_N(L_ND(\hat{\boldsymbol{F}}))^{-1}$ on each side of the above equation, and by Lemma 6, we have

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}} \;=\;& \Big(L_ND(\hat{\boldsymbol{F}})\Big)^{-1}\frac{L_N}{NT}\sum_{i=1}^{N}\left[\boldsymbol{R}_i^{\tau}M_{\boldsymbol{F}^0} - \frac{1}{N}\sum_{j=1}^{N}a_{ij}\boldsymbol{R}_j^{\tau}M_{\boldsymbol{F}^0}\right]\boldsymbol{\varepsilon}_i + \frac{L_N}{T}\Lambda_{NT} \\
&+\frac{L_N}{N}\Big(L_ND(\hat{\boldsymbol{F}})\Big)^{-1}\xi_{NT}^{*} + \Big(L_ND(\hat{\boldsymbol{F}})\Big)^{-1}O_P\left(L_N(NT)^{-1/2}\right) \\
&+\Big(L_ND(\hat{\boldsymbol{F}})\Big)^{-1}O_P\Big(\frac{L_N}{T\sqrt{N}}\Big) + L_NN^{-1/2}\Big(L_ND(\hat{\boldsymbol{F}})\Big)^{-1}O_P(\delta_{NT}^{-2}) \\
&+\Big(L_ND(\hat{\boldsymbol{F}})\Big)^{-1}O_P\Big(L_NT^{-1/2}\zeta_{Ld}^{1/2}\Big) + \Big(L_ND(\hat{\boldsymbol{F}})\Big)^{-1}O_P\Big(L_N^{1/2}\zeta_{Ld}^{1/2}\Big),
\end{aligned}$$

where

$$\xi_{NT}^{*} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\boldsymbol{R}_i - \boldsymbol{V}_i)^{\tau}\boldsymbol{F}^0}{T}\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\left(\frac{\Lambda^{\tau}\Lambda}{N}\right)^{-1}\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right) = O_P(1),$$

and

$$\Lambda_{NT} = -\left(L_N D(\hat{\boldsymbol{F}})\right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}} \Omega \hat{\boldsymbol{F}} G \lambda_i,$$

with $\Omega = \frac{1}{N}\sum_{j=1}^{N} \Omega_j$ and $\Omega_j = E(\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^\tau)$. By Lemmas 1 and 7, it can be shown that $D(\hat{\boldsymbol{F}}) = D(\boldsymbol{F}^0) + o_P(1)$ and the minimum and maximum eigenvalues of $L_N D(\hat{\boldsymbol{F}})$ are bounded with probability tending to 1. In addition, by Lemma 1 and Lemma A.6 in Bai (2009), it is easy to verify that $\Lambda_{NT} = O_P(1)$. Using the same argument for Lemma 2, we have

$$\begin{aligned} & \left\| D(\boldsymbol{F}^0)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \left[\boldsymbol{R}_i^\tau M_{\boldsymbol{F}^0} - \frac{1}{N} \sum_{j=1}^{N} a_{ij} \boldsymbol{R}_j^\tau M_{\boldsymbol{F}^0} \right] \boldsymbol{\varepsilon}_i \right\|^2 \\ \asymp & \left\| \frac{L_N}{NT} \sum_{i=1}^{N} \left[\boldsymbol{R}_i^\tau M_{\boldsymbol{F}^0} - \frac{1}{N} \sum_{j=1}^{N} a_{ij} \boldsymbol{R}_j^\tau M_{\boldsymbol{F}^0} \right] \boldsymbol{\varepsilon}_i \right\|^2 = O_P(L_N^2 (NT)^{-1}), \end{aligned}$$

uniformly for $\boldsymbol{F}^0$. By the above results, together with Lemma 1 and $\delta_{NT}^{-2} L_N \log L_N \to 0$ as $N, T \to \infty$, we have

$$\begin{aligned} \|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\| &= O_P(L_N (NT)^{-1/2}) + O_P(L_N T^{-1}) + O_P(L_N N^{-1}) \\ &\quad + O_P\left(L_N T^{-1/2} \zeta_{Ld}^{1/2}\right) + O_P\left(L_N^{1/2} \zeta_{Ld}^{1/2}\right). \end{aligned}$$

Summarizing the above results, we finish the proof of Theorem 2. □

S2.3 Proof of Theorem 3

Note that $\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u) = \boldsymbol{B}(u)^\tau(\hat{\gamma} - \tilde{\gamma}) + \boldsymbol{B}(u)^\tau \tilde{\gamma} - \boldsymbol{\beta}(u)$. By (C.2), we have

$$\|\boldsymbol{B}(u)^\tau \tilde{\gamma} - \boldsymbol{\beta}(u)\|_\infty = O_P(\zeta_{Ld}^{1/2}).$$

By Assumptions (A1) and (A8), Lemma 1, and the properties of B-spline, similarly to the proof of Corollary 1 in Huang et al. (2004), we can obtain that

$$\begin{aligned}&\varpi_k^\tau \boldsymbol{B}(u)\left(\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{Z}_i\right)^{-1}\Sigma_{NT1}\left(\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{Z}_i\right)^{-1}\boldsymbol{B}(u)^\tau \varpi_k\\ \gtrsim\ & C\frac{L_N}{NT}\sum_{l=1}^{L_k} B_{kl}^2(u) \gtrsim \frac{L_N}{NT}.\end{aligned}$$

Then, as $L_N^{2d+1}/NT \to \infty$, we have $\sup_{u\in\mathcal{U}}\left|\Sigma^{-1/2}(\boldsymbol{B}(u)^\tau\tilde{\gamma} - \boldsymbol{\beta}(u))\right| = o_P(1)$.

Invoking Lemmas 1 and 7, from the proof of Theorem 2, it is easy to show that

$$\begin{aligned}\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}} \ =\ & \left(L_N D(\boldsymbol{F}^0)\right)^{-1}\frac{L_N}{NT}\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{\varepsilon}_i + \frac{L_N}{N}\left(L_N D(\boldsymbol{F}^0)\right)^{-1}\tilde{\xi}_{NT}\\ &+\frac{L_N}{T}\left(L_N D(\boldsymbol{F}^0)\right)^{-1}\tilde{\Lambda}_{NT} + \left(L_N D(\boldsymbol{F}^0)\right)^{-1} O_P\left(L_N (NT)^{-1/2}\right)\\ &+\left(L_N D(\boldsymbol{F}^0)\right)^{-1} O_P\Big(L_N^{1/2}\zeta_{Ld}^{1/2}\Big), \qquad \text{(B.6)}\end{aligned}$$

where

$$\tilde{\xi}_{NT} = -\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \frac{(\boldsymbol{R}_i - \boldsymbol{V}_i)^\tau \boldsymbol{F}^0}{T}\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_j\left(\frac{1}{T}\sum_{t=1}^T \sigma_{ij,tt}\right),$$

and

$$\tilde{\Lambda}_{NT} = -\frac{1}{NT}\sum_{i=1}^N \boldsymbol{R}_i^\tau M_{\boldsymbol{F}^0}\Omega \boldsymbol{F}^0\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_i.$$

Under the assumptions that $\delta_{NT}^{-2}L_N\log L_N \to 0$, $L_N^{2d+1}/NT \to \infty$, and $T/N \to c$, we have

$$\Sigma^{-1/2}\boldsymbol{B}(u)\frac{L_N}{N}\left(L_N D(\boldsymbol{F}^0)\right)^{-1}\tilde{\xi}_{NT} \xrightarrow{P} \tilde{\Sigma}^{-1/2}c^{1/2}W_1^0,$$

$$\Sigma^{-1/2}\boldsymbol{B}(u)\frac{L_N}{N}\left(L_N D(\boldsymbol{F}^0)\right)^{-1}\tilde{\Lambda}_{NT} \xrightarrow{P} \tilde{\Sigma}^{-1/2}c^{-1/2}W_2^0,$$

where W_1^0 and W_2^0 are given in Theorem 3. Combining with Assumption (A10), we finish the proof of Theorem 3. □

S2.4 Proof of Theorem 4

Similarly to the argument of Bai and Ng (2006), it is easy to show that $\hat{W}_1$ is consistent for W_1. Similarly to the argument of Newey and West (1987) and Bai (2003), we can obtain that $\hat{W}_2$ is consistent for W_2. Thus, Theorem 4 follows. □

S2.5 Proof of Corollary 1

Invoking (B.6), similarly to the proof of Theorem 2 in Bai (2009), we can prove Corollary 1, and hence omit the details of proof. □

S2.6 Proof of Theorem 5

Since $Q(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\theta}, \boldsymbol{F}) = Q(\boldsymbol{\gamma}, \boldsymbol{F})$ attains the minimal value at $(\hat{\gamma}^{(1)\tau}, \hat{\beta}_{q+1}\mathbf{1}^{\tau}_{L_{q+1}}, \ldots, \hat{\beta}_p\mathbf{1}^{\tau}_{L_p})^{\tau}$, where $\hat{\gamma}^{(1)} = (\hat{\gamma}^{\tau}_1, \ldots, \hat{\gamma}^{\tau}_q)^{\tau}$. Similarly to the proof of Theorem

2, invoking Lemmas 3–7 and $\sum_{l=1}^{L_k} B_{kl}(u)=1$, we can get that

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\underline{\boldsymbol{R}}_i(\hat{\boldsymbol{\gamma}}^{(1)}-\widetilde{\boldsymbol{\gamma}}^{(1)}) \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\overline{\boldsymbol{X}}_i(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\overline{\boldsymbol{X}}_j a_{ij}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) \\
& + \frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\underline{\boldsymbol{R}}_j^{\tau} M_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\boldsymbol{e}_i \\
& + \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\underline{\boldsymbol{R}}_j a_{ij}(\hat{\boldsymbol{\gamma}}^{(1)}-\widetilde{\boldsymbol{\gamma}}^{(1)}) - \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\Omega_j\hat{\boldsymbol{F}}G\lambda_i \\
& + o_P(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + o_P(\hat{\boldsymbol{\gamma}}^{(1)}-\widetilde{\boldsymbol{\gamma}}^{(1)}) + N^{-1/2}O_P(\delta_{NT}^{-2}) + O_P(\zeta_{Ld}) \\
& + o_P((NT)^{-1/2}) + O_P(T^{-1/2}\zeta_{Ld}^{1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\overline{\boldsymbol{X}}_i(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\underline{\boldsymbol{R}}_i(\widetilde{\boldsymbol{\gamma}}^{(1)}-\hat{\boldsymbol{\gamma}}^{(1)}) - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\underline{\boldsymbol{R}}_j a_{ij}(\widetilde{\boldsymbol{\gamma}}^{(1)}-\hat{\boldsymbol{\gamma}}^{(1)}) \\
& + \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\overline{\boldsymbol{X}}_j^{\tau} M_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\boldsymbol{e}_i \\
& + \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\overline{\boldsymbol{X}}_j a_{ij}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) - \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\hat{\boldsymbol{F}}}\Omega_j\hat{\boldsymbol{F}}G\lambda_i \\
& + o_P(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) + o_P(\widetilde{\boldsymbol{\gamma}}^{(1)}-\hat{\boldsymbol{\gamma}}^{(1)}) + N^{-1/2}O_P(\delta_{NT}^{-2}) + O_P(\zeta_{Ld}) \\
& + o_P((NT)^{-1/2}) + O_P(T^{-1/2}\zeta_{Ld}^{1/2}).
\end{aligned}
$$

Let $\overline{\boldsymbol{Z}}_i = M_{\boldsymbol{F}^0}\overline{\boldsymbol{X}}_i - \frac{1}{N}\sum_{j=1}^{N} M_{\boldsymbol{F}^0}\overline{\boldsymbol{X}}_j a_{ij}$ and $\underline{\boldsymbol{Z}}_i = M_{\boldsymbol{F}^0}\underline{\boldsymbol{R}}_i - \frac{1}{N}\sum_{j=1}^{N} M_{\boldsymbol{F}^0}\underline{\boldsymbol{R}}_j a_{ij}$, a simple calculation yields that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{Z}}_i^{\tau}\underline{\boldsymbol{Z}}_i(\hat{\boldsymbol{\gamma}}^{(1)}-\widetilde{\boldsymbol{\gamma}}^{(1)})\\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{R}}_i^{\tau}\overline{\boldsymbol{Z}}_i(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})+\frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{Z}}_i^{\tau}\boldsymbol{\varepsilon}_i+\frac{1}{NT}\sum_{i=1}^{N}\underline{\boldsymbol{R}}_i^{\tau}M_{\boldsymbol{F}^0}\boldsymbol{e}_i\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\underline{\boldsymbol{R}}_i-\underline{\boldsymbol{V}}_i)^{\tau}\boldsymbol{F}^0}{T}G^0\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right)\\
&-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\underline{\boldsymbol{R}}_i^{\tau}M_{\boldsymbol{F}^0}\Omega_j\boldsymbol{F}^0G^0\lambda_i+o_P(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\\
&+o_P(\hat{\boldsymbol{\gamma}}^{(1)}-\widetilde{\boldsymbol{\gamma}}^{(1)})+N^{-1/2}O_P(\delta_{NT}^{-2})+O_P(\zeta_{Ld})\\
&+o_P((NT)^{-1/2})+O_P(T^{-1/2}\zeta_{Ld}^{1/2})+O_P(N^{-1/2}\zeta_{Ld}^{1/2}), \qquad \text{(B.7)}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{Z}}_i^{\tau}\overline{\boldsymbol{Z}}_i(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{X}}_i^{\tau}\underline{\boldsymbol{R}}_i(\widetilde{\boldsymbol{\gamma}}^{(1)}-\hat{\boldsymbol{\gamma}}^{(1)})+\frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{Z}}_i^{\tau}\boldsymbol{\varepsilon}_i+\frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{X}}_i^{\tau}M_{\boldsymbol{F}^0}\boldsymbol{e}_i\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\overline{\boldsymbol{X}}_i-\overline{\boldsymbol{V}}_i)^{\tau}\boldsymbol{F}^0}{T}G^0\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right)\\
&-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\overline{\boldsymbol{X}}_i^{\tau}M_{\boldsymbol{F}^0}\Omega_j\boldsymbol{F}^0G^0\lambda_i+o_P(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\
&+o_P(\widetilde{\boldsymbol{\gamma}}^{(1)}-\hat{\boldsymbol{\gamma}}^{(1)})+N^{-1/2}O_P(\delta_{NT}^{-2})+O_P(\zeta_{Ld})\\
&+o_P((NT)^{-1/2})+O_P(T^{-1/2}\zeta_{Ld}^{1/2})+O_P(N^{-1/2}\zeta_{Ld}^{1/2}), \qquad \text{(B.8)}
\end{aligned}$$

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

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

$$\underline{\Xi}_1 = \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\underline{\boldsymbol{R}}_i^{\tau} M_{\boldsymbol{F}^0}\Omega_j \boldsymbol{F}^0 G^0\lambda_i,$$

$$\overline{\Xi}_1 = \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\overline{\boldsymbol{X}}_i^{\tau} M_{\boldsymbol{F}^0}\Omega_j \boldsymbol{F}^0 G^0\lambda_i,$$

$$\underline{\Xi}_2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\underline{\boldsymbol{R}}_i - \underline{\boldsymbol{V}}_i)^{\tau}\boldsymbol{F}^0}{T} G^0\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right),$$

$$\overline{\Xi}_2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{V}}_i)^{\tau}\boldsymbol{F}^0}{T} G^0\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right),$$

$$\Psi = \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{X}}_i^{\tau}\underline{\boldsymbol{Z}}_i = \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{Z}}_i^{\tau}\underline{\boldsymbol{R}}_i = \frac{1}{NT}\sum_{i=1}^{N}\overline{\boldsymbol{Z}}_i^{\tau}\underline{\boldsymbol{Z}}_i.$$

Then we get

$$\begin{aligned}(\hat{\boldsymbol{\gamma}}^{(1)} - \widetilde{\boldsymbol{\gamma}}^{(1)}) \;=\;& (\underline{\Phi} + o_P(1))^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + o_P(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \\ & -(\underline{\Phi} + o_P(1))^{-1}\underline{\Xi}_1 - (\underline{\Phi} + o_P(1))^{-1}\underline{\Xi}_2 \\ & +(\underline{\Phi} + o_P(1))^{-1}\frac{1}{NT}\sum_{i=1}^{N}(\underline{\boldsymbol{Z}}_i^{\tau}\boldsymbol{\varepsilon}_i + \underline{\boldsymbol{R}}_i^{\tau}M_{\boldsymbol{F}^0}\boldsymbol{e}_i) \\ & +o_P(\hat{\boldsymbol{\gamma}}^{(1)} - \widetilde{\boldsymbol{\gamma}}^{(1)}) + N^{-1/2}O_P(\delta_{NT}^{-2}) + O_P(\zeta_{Ld}) \\ & +o_P((NT)^{-1/2}) + O_P(T^{-1/2}\zeta_{Ld}^{1/2}) + O_P(N^{-1/2}\zeta_{Ld}^{1/2}). \quad \text{(B.9)}\end{aligned}$$

Substituting (B.9) into (B.8), and a simple calculation yields that

$$\begin{aligned}& (\overline{\Phi} - \Psi\underline{\Phi}^{-1}\Psi^{\tau} + o_P(1))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ =\;& \frac{1}{NT}\sum_{i=1}^{N}(\overline{\boldsymbol{Z}}_i^{\tau}\boldsymbol{\varepsilon}_i + \overline{\boldsymbol{X}}_i^{\tau}M_{\boldsymbol{F}^0}\boldsymbol{e}_i) - \overline{\Xi}_1 - \overline{\Xi}_2 + \Psi(\underline{\Phi}^{-1} + o_P(1))\underline{\Xi}_1 \\ & +\Psi(\underline{\Phi}^{-1} + o_P(1))\underline{\Xi}_2 - \Psi(\underline{\Phi}^{-1} + o_P(1))\frac{1}{NT}\sum_{i=1}^{N}(\underline{\boldsymbol{Z}}_i^{\tau}\boldsymbol{\varepsilon}_i + \underline{\boldsymbol{R}}_i^{\tau}M_{\boldsymbol{F}^0}\boldsymbol{e}_i) \\ & +N^{-1/2}O_P(\delta_{NT}^{-2}) + O_P(\zeta_{Ld}) + o_P((NT)^{-1/2}) \\ & +O_P(T^{-1/2}\zeta_{Ld}^{1/2}) + O_P(N^{-1/2}\zeta_{Ld}^{1/2}).\end{aligned}$$

Thus we have

$$\begin{aligned}
&(\overline{\Phi} - \Psi\underline{\Phi}^{-1}\Psi^\tau + o_P(1))\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\
= \;& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(\overline{\boldsymbol{Z}}_i - \underline{\boldsymbol{Z}_i\Phi}^{-1}\Psi^\tau)^\tau \boldsymbol{\varepsilon}_i + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(\overline{\boldsymbol{X}}_i - \underline{\boldsymbol{R}_i\Phi}^{-1}\Psi^\tau)^\tau M_{\boldsymbol{F}^0}\boldsymbol{e}_i \\
& -\sqrt{NT}(\overline{\Xi}_1 - \Psi(\underline{\Phi}^{-1} + o_P(1))\underline{\Xi}_1) - \sqrt{NT}(\overline{\Xi}_2 - \Psi(\underline{\Phi}^{-1} + o_P(1))\underline{\Xi}_2) + o_P(1).
\end{aligned}$$

By Assumption (A1) and (C.2), and using the similar proofs of Lemma A.7 in Huang et al. (2004), and Lemmas 2 and 3, it is easy to show that

$$\left\| \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(\overline{\boldsymbol{X}}_i - \underline{\boldsymbol{R}_i\Phi}^{-1}\Psi^\tau)^\tau M_{\boldsymbol{F}^0}\boldsymbol{e}_i \right\| = o_P(1).$$

Using the central limits theorem, we can obtain that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(\overline{\boldsymbol{Z}}_i - \underline{\boldsymbol{Z}_i\Phi}^{-1}\Psi^\tau)^\tau \boldsymbol{\varepsilon}_i \xrightarrow{L} N(0, \Pi_2).$$

In addition, by the law of large numbers, we have

$$\overline{\Phi} - \Psi\underline{\Phi}^{-1}\Psi^\tau \xrightarrow{P} \Pi_1.$$

Invoking the Slutsky Theorem, we complete the proof of Theorem 5. □

S2.7 Proof of Theorem 6

By a simple calculation, we have

$$\begin{aligned}
\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^*)\\
&= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i + \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{F}}^*\hat{\lambda}_i^* + \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau\\
&\quad\times(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i + \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{F}}^*\hat{\lambda}_i^* + \hat{\boldsymbol{F}}\hat{\lambda}_i)\\
&= \mathrm{RSS}_1 + \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}})^\tau(\boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}})\\
&\quad+\frac{1}{NT}\sum_{i=1}^{N}(\hat{\boldsymbol{F}}^*\hat{\lambda}_i^* - \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau(\hat{\boldsymbol{F}}^*\hat{\lambda}_i^* - \hat{\boldsymbol{F}}\hat{\lambda}_i)\\
&\quad+\frac{2}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau(\boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}})\\
&\quad-\frac{2}{NT}\sum_{i=1}^{N}(\boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}})^\tau(\hat{\boldsymbol{F}}^*\hat{\lambda}_i^* - \hat{\boldsymbol{F}}\hat{\lambda}_i)\\
&\quad-\frac{2}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau(\hat{\boldsymbol{F}}^*\hat{\lambda}_i^* - \hat{\boldsymbol{F}}\hat{\lambda}_i).
\end{aligned}$$

For the second term of the above equation, by the properties of B-spline, we have

$$\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}})^\tau(\boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \underline{\boldsymbol{R}_i}\hat{\boldsymbol{\gamma}}^{(1)*} - \overline{\boldsymbol{X}}_i\hat{\boldsymbol{\theta}}) \asymp \|\hat{\boldsymbol{\beta}}(u) - \check{\boldsymbol{\beta}}(u)\|_{L_2}^2,$$

where $\check{\boldsymbol{\beta}}(u) = \boldsymbol{R}_i\check{\boldsymbol{\gamma}}$ with $\check{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}^{(1)*\tau}, \hat{\beta}_{q+1}\mathbf{1}_{L_{q+1}}^\tau, \ldots, \hat{\beta}_p\mathbf{1}_{L_p}^\tau)^\tau$. Then, under H_0, we have

$$\|\hat{\boldsymbol{\beta}}(u) - \check{\boldsymbol{\beta}}(u)\|_{L_2} \leq \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\|_{L_2} + \|\check{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\|_{L_2} \xrightarrow{P} 0,$$

where $\boldsymbol{\beta}(u) = (\beta_1(u), \ldots, \beta_q(u), \beta_{q+1}, \ldots, \beta_p)^\tau$. For the third term, a simple calculation yields that

$$\hat{\boldsymbol{F}}^* \hat{\lambda}_i^* - \hat{\boldsymbol{F}} \hat{\lambda}_i = \hat{\boldsymbol{F}}^* \hat{\lambda}_i^* - \boldsymbol{F}^0 \lambda_i + \boldsymbol{F}^0 \lambda_i - \hat{\boldsymbol{F}} \hat{\lambda}_i,$$

$$\boldsymbol{F}^0 \lambda_i - \hat{\boldsymbol{F}} \hat{\lambda}_i = (\boldsymbol{F}^0 H - \hat{\boldsymbol{F}}) H^{-1} \lambda_i - \hat{\boldsymbol{F}}(\hat{\lambda}_i - H^{-1}\lambda_i),$$

$$\hat{\boldsymbol{F}}^* \hat{\lambda}_i^* - \boldsymbol{F}^0 \lambda_i = (\hat{\boldsymbol{F}}^* - \boldsymbol{F}^0 H) H^{-1} \lambda_i + \hat{\boldsymbol{F}}^*(\hat{\lambda}_i^* - H^{-1}\lambda_i).$$

Invoking Proposition A.1 (ii) and Lemma A.10 in Bai (2009), Lemma 3 (i), and Assumptions (A6)–(A7), we have $\frac{1}{NT}\sum_{i=1}^N (\hat{\boldsymbol{F}}^* \hat{\lambda}_i^* - \hat{\boldsymbol{F}} \hat{\lambda}_i)^\tau (\hat{\boldsymbol{F}}^* \hat{\lambda}_i^* - \hat{\boldsymbol{F}} \hat{\lambda}_i) = o_P(1)$. Similarly, it is easy to show that

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

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

$$\frac{1}{NT}\sum_{i=1}^N (\boldsymbol{Y}_i - \boldsymbol{R}_i \hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}} \hat{\lambda}_i)^\tau (\hat{\boldsymbol{F}}^* \hat{\lambda}_i^* - \hat{\boldsymbol{F}} \hat{\lambda}_i) = o_P(1).$$

On the other hand, under H_1, because $\|\hat{\boldsymbol{\beta}}(u) - \check{\boldsymbol{\beta}}(u)\|_{L_2} \geq \|\check{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\|_{L_2} - \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\|_{L_2}$. As $N \to \infty$ and $T \to \infty$, we have

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}(u) - \check{\boldsymbol{\beta}}(u)\|_{L_2} &\geq \sum_{k=1}^p \|\check{\beta}_k(u) - \beta_k(u)\|_{L_2} - o_P(1) \\ &\geq \sum_{k=q+1}^p \inf_{a \in \mathbb{R}} \|\beta_k(u) - a\|_{L_2} - o_P(1). \end{aligned}$$

Then, by the Cauchy-Schwarz inequality, a simple calculation yields that

$$\mathrm{RSS}_0 - \mathrm{RSS}_1 \geq \sum_{k=q+1}^p \inf_{a \in \mathbb{R}} \|\beta_k(u) - a\|_{L_2} + o_P(1).$$

It remains to show that, with probability tending to one, RSS_1 is bounded away from zero and infinity. By some elementary calculations, we have

$$\begin{aligned}
\mathrm{RSS}_1 &= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau(\boldsymbol{Y}_i - \boldsymbol{R}_i\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{F}}\hat{\lambda}_i) \\
&= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{\varepsilon}_i + \boldsymbol{e}_i + \boldsymbol{R}_i(\tilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}) + \boldsymbol{F}^0\lambda_i - \hat{\boldsymbol{F}}\hat{\lambda}_i)^\tau \\
&\quad \times(\boldsymbol{\varepsilon}_i + \boldsymbol{e}_i + \boldsymbol{R}_i(\tilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}) + \boldsymbol{F}^0\lambda_i - \hat{\boldsymbol{F}}\hat{\lambda}_i) \\
&= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{\varepsilon}_i + \boldsymbol{e}_i + \boldsymbol{R}_i(\tilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}) + (\boldsymbol{F}^0 H - \hat{\boldsymbol{F}})H^{-1}\lambda_i - \hat{\boldsymbol{F}}(\hat{\lambda}_i - H^{-1}\lambda_i))^\tau \\
&\quad \times(\boldsymbol{\varepsilon}_i + \boldsymbol{e}_i + \boldsymbol{R}_i(\tilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}) + (\boldsymbol{F}^0 H - \hat{\boldsymbol{F}})H^{-1}\lambda_i - \hat{\boldsymbol{F}}(\hat{\lambda}_i - H^{-1}\lambda_i)) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{\varepsilon}_i + o_P(1).
\end{aligned}$$

Thus, it suffices to show that, with probability tending to one, $\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{\varepsilon}_i$ is bounded away from zero and infinity. By Assumption (A8), we have

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{\varepsilon}_i\right) &= \frac{1}{N^2T^2}\mathrm{Cov}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2, \sum_{j=1}^{N}\sum_{s=1}^{T}\varepsilon_{js}^2\right) \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left(\varepsilon_{it}^2, \varepsilon_{js}^2\right) \to 0.
\end{aligned}$$

The Chebyshev inequality then implies that, as $N \to \infty$ and $T \to \infty$,

$$\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{\varepsilon}_i - E\left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{\varepsilon}_i\right) \to 0$$

in probability. Since $E(\varepsilon_{it}^2)$ is bounded away from 0 and infinity, the result follows. □

S3 Appendix C: Some lemmas and their proofs

In order to prove Theorems 1–6, we provide Lemmas 1–7 in Appendix C.

Lemma 1 *Let $\rho_{\min}$ and $\rho_{\max}$ be the minimum and maximum eigenvalues of $L_N D(\boldsymbol{F})$ respectively. Then there exist two positive constants M_3 and M_4 such that $M_3 \leq \rho_{\min} \leq \rho_{\max} \leq M_4$.*

Proof The proof of Lemma 1 follows the same lines as Lemma A.3 in Huang et al. (2004), Lemma 3.2 in He and Shi (1994), and Lemma 3 in Tang and Cheng (2009). We hence omit the proof of Lemma 1. □

Lemma 2 *Assume that assumptions (A1), (A2), (A4)–(A8) hold. We have*

$$\sup_{\boldsymbol{F}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{R}_i^{\tau} M_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i \right\| = o_P(1),$$

$$\sup_{\boldsymbol{F}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{\tau} \boldsymbol{F}^{\tau} M_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i \right\| = o_P(1),$$

$$\sup_{\boldsymbol{F}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{\varepsilon}_i^{\tau} P_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i \right\| = o_P(1).$$

Proof Using $P_{\boldsymbol{F}} = \boldsymbol{F}\boldsymbol{F}^{\tau}/T$, we have

$$\frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{R}_i^{\tau} M_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i = \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{R}_i^{\tau} \boldsymbol{\varepsilon}_i - \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{R}_i^{\tau} P_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i.$$

By Assumptions (A1) and (A8), together with the properties of B-spline, it is easy to show that $\frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{R}_i^{\tau} \boldsymbol{\varepsilon}_i = O_P((NT)^{-1/2}) = o_P(1)$. Now we show that $\sup_{\boldsymbol{F}} \frac{1}{NT}\sum_{i=1}^N \boldsymbol{R}_i^\tau P_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i = o_P(1)$. Note that

$$\begin{aligned}\frac{1}{NT}\left\|\sum_{i=1}^N \boldsymbol{R}_i^\tau P_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i\right\| &= \left\|\frac{1}{N}\sum_{i=1}^N\left(\frac{\boldsymbol{R}_i^\tau \boldsymbol{F}}{T}\right)\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\| \\ &\le \frac{1}{N}\sum_{i=1}^N\left\|\frac{\boldsymbol{R}_i^\tau \boldsymbol{F}}{T}\right\|\cdot\left\|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\|. \qquad \text{(C.1)}\end{aligned}$$

By $T^{-1/2}\|\boldsymbol{F}\| = \sqrt{r}$, we have $T^{-1}\|\boldsymbol{R}_i^\tau\boldsymbol{F}\| \le T^{-1}\|\boldsymbol{R}_i\|\|\boldsymbol{F}\| = \sqrt{r}T^{-1/2}\|\boldsymbol{R}_i\|$. By Cauchy-Schwarz inequality, (C.1) is bounded above by

$$\sqrt{r}\left(\frac{1}{N}\sum_{i=1}^N\frac{1}{T}\sum_{t=1}^T\|R_{it}\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\|^2\right)^{1/2}.$$

By $T^{-1/2}\|\boldsymbol{R}_i\| = O_P(1)$, the first term of the above expression is of order $O_P(1)$. Similarly to the proof of Lemma A.1 in Bai (2009), it is easy to show that the order of the second term is $o_P(1)$ uniformly in $\boldsymbol{F}$.

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\|^2 &= \operatorname{tr}\left(\frac{1}{N}\sum_{i=1}^N\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s^\tau\varepsilon_{it}\varepsilon_{is}\right) \\ &= \operatorname{tr}\left(\frac{1}{N}\sum_{i=1}^N\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s^\tau[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})]\right) \\ &\quad +\operatorname{tr}\left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s^\tau\frac{1}{N}\sum_{i=1}^N E(\varepsilon_{it}\varepsilon_{is})\right).\end{aligned}$$

Note that $T^{-1}\sum_{t=1}^T\|F_t\|^2 = \|\boldsymbol{F}^\tau\boldsymbol{F}/T\| = r$. By Cauchy-Schwarz inequality and Assumption (A8), we obtain that

$$\begin{aligned}&\operatorname{tr}\left(\frac{1}{N}\sum_{i=1}^N\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s^\tau[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})]\right) \\ \le\ &\left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\|F_t\|^2\|F_s\|^2\right)^{1/2} N^{-1/2}\left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})]\right]^2\right)^{1/2} \\ =\ &rN^{-1/2}O_P(1).\end{aligned}$$

Next, by Assumption (A8)(ii), we have $|N^{-1}\sum_{i=1}^{N}\sigma_{ii,ts}| \le \varrho_{ts}$, where $\sigma_{ii,ts} = E(\varepsilon_{it}\varepsilon_{is})$. Again using the Cauchy-Schwarz inequality,

$$\begin{aligned}
& \operatorname{tr}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_tF_s^\tau\frac{1}{N}\sum_{i=1}^{N}E(\varepsilon_{it}\varepsilon_{is})\right) \\
\le & \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\|F_t\|^2\|F_s\|^2\right)^{1/2}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varrho_{ts}^2\right)^{1/2} \\
\le & \ rT^{-1/2}C\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varrho_{ts}\right)^{1/2} \\
= & \ rO(T^{-1/2}).
\end{aligned}$$

This shows that

$$\sup_{\boldsymbol{F}}\left\|\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau M_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i\right\| = O_P((NT)^{-1/2}) = o_P(1).$$

The proofs of the second and third results are similar to the proof of the first one, and hence are omitted. □

Lemma 3 *Assume that assumptions (A1)–(A9) hold. For ease of notation, let* $H = (\Lambda^\tau\Lambda/N)(\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}/T)V_{NT}^{-1}$. *We have*

(i) $T^{-1/2}\|\hat{\boldsymbol{F}} - \boldsymbol{F}^0H\| = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-1}) + O_P\left(\zeta_{Ld}^{1/2}\right)$,

(ii) $T^{-1}\boldsymbol{F}^{0\tau}(\hat{\boldsymbol{F}} - \boldsymbol{F}^0H) = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}\right)$,

(iii) $T^{-1}\hat{\boldsymbol{F}}^\tau(\hat{\boldsymbol{F}} - \boldsymbol{F}^0H) = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}\right)$,

(iv) $T^{-1}\boldsymbol{R}_j^\tau(\hat{\boldsymbol{F}} - \boldsymbol{F}^0H) = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}\right)$, *for all* j,

(v) $\dfrac{1}{NT}\displaystyle\sum_{j=1}^{N}\boldsymbol{R}_j^\tau M_{\hat{\boldsymbol{F}}}(\hat{\boldsymbol{F}} - \boldsymbol{F}^0H) = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}\right)$,

(vi) $HH^\tau - (T^{-1}\boldsymbol{F}^{0\tau}\boldsymbol{F}^0)^{-1} = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}\right)$.

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

$$\sup_{u\in\mathcal{U}}|Re_k(u)| \le ML_k^{-d}, \quad k = 1,\ldots,p. \tag{C.2}$$

In addition, noting that $\boldsymbol{Y}_i = \boldsymbol{R}_i\widetilde{\boldsymbol{\gamma}} + \boldsymbol{F}^0\lambda_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i$, for $i = 1,\ldots,N$, we have the following expansion:

$$\begin{aligned}
\hat{\boldsymbol{F}} V_{NT} \;&=\; \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i(\widetilde{\gamma}-\hat{\gamma})(\widetilde{\gamma}-\hat{\gamma})^\tau\boldsymbol{R}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i(\widetilde{\gamma}-\hat{\gamma})\lambda_i^\tau\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}} \\
&+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i(\widetilde{\gamma}-\hat{\gamma})\boldsymbol{\varepsilon}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}^0\lambda_i(\widetilde{\gamma}-\hat{\gamma})^\tau\boldsymbol{R}_i^\tau\hat{\boldsymbol{F}} \\
&+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i(\widetilde{\gamma}-\hat{\gamma})^\tau\boldsymbol{R}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}^0\lambda_i\boldsymbol{\varepsilon}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\lambda_i^\tau\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}} \\
&+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i(\widetilde{\gamma}-\hat{\gamma})\boldsymbol{e}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i(\widetilde{\gamma}-\hat{\gamma})^\tau\boldsymbol{R}_i^\tau\hat{\boldsymbol{F}} \\
&+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}^0\lambda_i\boldsymbol{e}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\lambda_i^\tau\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{e}_i^\tau\hat{\boldsymbol{F}} \\
&+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{\varepsilon}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{e}_i^\tau\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}^0\lambda_i\lambda_i^\tau\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}} \\
=:&\;\; B_1 + B_2 + B_3 + \cdots + B_{16},
\end{aligned}$$

where $B_{16} = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}^0\lambda_i\lambda_i^\tau\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}} = \boldsymbol{F}^0(\Lambda^\tau\Lambda/N)(\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}/T)$. This leads to

$$\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H = (B_1 + B_2 + \cdots + B_{15})V_{NT}^{-1}. \tag{C.3}$$

Noting that $T^{-1/2}\|\hat{\boldsymbol{F}}\| = \sqrt{r}$ and $\|\boldsymbol{R}_i\| = O_P(T^{1/2})$, we have

$$\begin{aligned}
T^{-1/2}\|B_1\| \;&\le\; \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\|\boldsymbol{R}_i\|^2}{T}\right)\|\hat{\gamma}-\widetilde{\gamma}\|^2\sqrt{r} = O_P(\|\hat{\gamma}-\widetilde{\gamma}\|^2) = o_P(\|\hat{\gamma}-\widetilde{\gamma}\|), \\
T^{-1/2}\|B_2\| \;&\le\; \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\|\boldsymbol{R}_i\|}{\sqrt{T}}\right)\|\hat{\gamma}-\widetilde{\gamma}\|\|\lambda_i\|\|\boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}/T\| = O_P(\|\hat{\gamma}-\widetilde{\gamma}\|).
\end{aligned}$$

Using the same argument, it is easy to show that $T^{-1/2}\|B_l\| = O_P(\|\hat{\gamma}-\widetilde{\gamma}\|)$, for $l = 3$, 4 and 5, and $T^{-1/2}\|B_l\| = O_P(\delta_{NT}^{-1})$, for $l = 6$, 7 and 8. For B_9, using the same argument, and by (C.2) and Assumption (A1), we have

$$\begin{aligned} T^{-1/2}\|B_9\| &\leq T^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{\|\boldsymbol{R}_i\|}{\sqrt{T}}\Big)\|\hat{\gamma}-\widetilde{\gamma}\|\Big(\frac{\|\hat{\boldsymbol{F}}\|}{\sqrt{T}}\Big)\sqrt{\sum_{t=1}^{T}e_{it}^2} \\ &\leq O_P(\|\hat{\gamma}-\widetilde{\gamma}\|)\cdot M\zeta_{Ld}^{1/2}. \end{aligned}$$

Similarly, we can prove that $T^{-1/2}\|B_{10}\| = O_P(\|\hat{\gamma}-\widetilde{\gamma}\|)\cdot M\zeta_{Ld}^{1/2}$. For B_{11}, we have

$$T^{-1/2}\|B_{11}\| \leq T^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{\|\boldsymbol{F}^0\|}{\sqrt{T}}\Big)\|\lambda_i\|\sqrt{r\sum_{t=1}^{T}e_{it}^2} = O_P\Big(\zeta_{Ld}^{1/2}\Big).$$

Similarly, it yields that $T^{-1/2}\|B_{12}\| = O_P(\zeta_{Ld}^{1/2})$. For B_{13}, we have

$$T^{-1/2}\|B_{13}\| \leq \frac{1}{NT}\sum_{i=1}^{N}\|\boldsymbol{\varepsilon}_i\|\sqrt{r\sum_{t=1}^{T}e_{it}^2} = O_P\Big(\zeta_{Ld}^{1/2}\delta_{NT}^{-1}\Big).$$

Similarly, it yields that $T^{-1/2}\|B_{14}\| = O_P(\zeta_{Ld}^{1/2}\delta_{NT}^{-1})$. For B_{15}, we have

$$T^{-1/2}\|B_{15}\| \leq \frac{1}{NT}\sum_{i=1}^{N}\Big(\sum_{t=1}^{T}e_{it}^2\Big)\sqrt{r} = O_P(\zeta_{Ld}).$$

Following the same arguments as in the proof of Proposition A.1 in Bai (2009), together with the above results, we have

$$T^{-1/2}\|\hat{\boldsymbol{F}}-\boldsymbol{F}^0H\| = O_P(\|\hat{\gamma}-\widetilde{\gamma}\|) + O_P(\delta_{NT}^{-1}) + O_P\Big(\zeta_{Ld}^{1/2}\Big).$$

(ii) By (C.3), we have the following decomposition:

$$T^{-1}\boldsymbol{F}^{0\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H) = T^{-1}\boldsymbol{F}^{0\tau}(B_1+B_2+\cdots+B_{15})V_{NT}^{-1}.$$

Invoking the similar arguments as in the proof of Lemma A.3 (i) in Bai (2009s) to the first eight terms, we can obtain that

$$T^{-1}\boldsymbol{F}^{0\tau}(B_1+B_2+\cdots+B_8)V_{NT}^{-1}=O_P(\|\hat{\gamma}-\widetilde{\gamma}\|)+O_P(\delta_{NT}^{-2}).$$

For the other terms, we can show that $T^{-1}\boldsymbol{F}^{0\tau}B_9V_{NT}^{-1}$ and $T^{-1}\boldsymbol{F}^{0\tau}B_{10}V_{NT}^{-1}$ are of order $O_P(\|\hat{\gamma}-\widetilde{\gamma}\|\zeta_{Ld}^{1/2})$, $T^{-1}\boldsymbol{F}^{0\tau}B_{11}V_{NT}^{-1}$ and $T^{-1}\boldsymbol{F}^{0\tau}B_{12}V_{NT}^{-1}$ are of order $O_P(\zeta_{Ld}^{1/2})$, $T^{-1}\boldsymbol{F}^{0\tau}B_{13}V_{NT}^{-1}$ and $T^{-1}\boldsymbol{F}^{0\tau}B_{14}V_{NT}^{-1}$ are of order $O_P(\zeta_{Ld}^{1/2}\delta_{NT}^{-1})$, and $T^{-1}\boldsymbol{F}^{0\tau}B_{15}V_{NT}^{-1}=O_P(\zeta_{Ld})$. This finishes the proof of (ii).

(iii) By (i) and (ii) and some elementary calculations, we have

$$\begin{aligned}\|T^{-1}\hat{\boldsymbol{F}}^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)\| &\leq T^{-1}\|\hat{\boldsymbol{F}}-\boldsymbol{F}^0H\|^2+\|H\|T^{-1}\|\boldsymbol{F}^{0\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)\| \\ &= O_P(\|\hat{\gamma}-\widetilde{\gamma}\|)+O_P(\delta_{NT}^{-2})+O_P\left(\zeta_{Ld}^{1/2}\right).\end{aligned}$$

(iv) The proof of (iv) is similar to that for (ii), and hence is omitted.

(v) Noting that $M_{\hat{\boldsymbol{F}}}=I_T-\hat{\boldsymbol{F}}\hat{\boldsymbol{F}}^{\tau}/T$, we have

$$\begin{aligned}&\frac{1}{NT}\sum_{j=1}^{N}\boldsymbol{R}_j^{\tau}M_{\hat{\boldsymbol{F}}}(\hat{\boldsymbol{F}}-\boldsymbol{F}H) \\ =&\ \frac{1}{N}\sum_{j=1}^{N}\frac{1}{T}\boldsymbol{R}_j^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}H)-\frac{1}{N}\sum_{j=1}^{N}\frac{\boldsymbol{R}_j^{\tau}\hat{\boldsymbol{F}}}{T}T^{-1}\hat{\boldsymbol{F}}^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}H) \\ =:&\ I_1+I_2.\end{aligned}$$

Since I_1 is an average of $\frac{1}{T}\boldsymbol{R}_j^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)$ over j, it is easy to verify that $I_1=O_P(\|\hat{\gamma}-\widetilde{\gamma}\|)+O_P(\delta_{NT}^{-2})+O_P(\zeta_{Ld}^{1/2})$. For I_2, by (iii) we have

$$\begin{aligned}\|I_2\| &\leq \frac{1}{N}\sum_{j=1}^{N}\frac{\|\boldsymbol{R}_j\|}{\sqrt{T}}\sqrt{r}\|T^{-1}\hat{\boldsymbol{F}}^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)\| \\ &= O_P(\|\hat{\gamma}-\widetilde{\gamma}\|)+O_P(\delta_{NT}^{-2})+O_P\left(\zeta_{Ld}^{1/2}\right).\end{aligned}$$

This completes the proof of (v).

(vi) By (ii), we have

$$\begin{aligned} & \boldsymbol{F}^{0\tau}\hat{\boldsymbol{F}}/T - (\boldsymbol{F}^{0\tau}\boldsymbol{F}^{0}/T)H \\ = \;& O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\Big(\zeta_{Ld}^{1/2}\Big). \end{aligned} \tag{C.4}$$

By (iii) and the fact that $\hat{\boldsymbol{F}}^{\tau}\hat{\boldsymbol{F}}/T = I_r$, we have

$$I_r - (\hat{\boldsymbol{F}}^{\tau}\boldsymbol{F}^{0}/T)H = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\Big(\zeta_{Ld}^{1/2}\Big). \tag{C.5}$$

Left-multiplying by H^{τ} in (C.4), and using the transpose for (C.5), we have

$$I_r - H^{\tau}(\boldsymbol{F}^{0\tau}\boldsymbol{F}^{0}/T)H = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\Big(\zeta_{Ld}^{1/2}\Big),$$

which shows that (vi) holds. □

Lemma 4 *Assume that assumptions (A1)–(A9) hold. We have*

(i) $$T^{-1}\boldsymbol{\varepsilon}_j^{\tau}(\hat{\boldsymbol{F}} - \boldsymbol{F}^{0}H) = T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P\Big(\zeta_{Ld}^{1/2}T^{-1/2}\Big), \text{ for all } j = 1, \ldots, N,$$

(ii) $$\frac{1}{T\sqrt{N}}\sum_{j=1}^{N}\boldsymbol{\varepsilon}_j^{\tau}(\hat{\boldsymbol{F}} - \boldsymbol{F}^{0}H) = T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + N^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(N^{-1/2}) + O_P(\delta_{NT}^{-2}) + O_P\Big(\zeta_{Ld}^{1/2}\Big),$$

(iii) $$\frac{1}{NT}\sum_{j=1}^{N}\lambda_j\boldsymbol{\varepsilon}_j^{\tau}(\hat{\boldsymbol{F}} - \boldsymbol{F}^{0}H) = (TN)^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(N^{-1}) + N^{-1/2}O_P(\delta_{NT}^{-2}) + N^{-1/2}O_P\Big(\zeta_{Ld}^{1/2}\Big).$$

Proof (i) By (C.3), we have

$$T^{-1}\boldsymbol{\varepsilon}_j^{\tau}(\hat{\boldsymbol{F}} - \boldsymbol{F}^{0}H) = T^{-1}\boldsymbol{\varepsilon}_j^{\tau}(B_1 + B_2 + \cdots + B_{15})V_{NT}^{-1}. \tag{C.6}$$

Invoking the similar arguments as in the proof of Lemma A.4 (i) in Bai (2009s) to the first eight terms, we can obtain that

$$T^{-1}\boldsymbol{\varepsilon}_j^\tau(B_1+B_2+\cdots+B_8)V_{NT}^{-1}=T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-2}).$$

For the other terms in (C.6), similarly to the proof of (i) in Lemma 3, we only need to show that the dominant terms $T^{-1}\boldsymbol{\varepsilon}_j^\tau B_{11}V_{NT}^{-1}$ and $T^{-1}\boldsymbol{\varepsilon}_j^\tau B_{12}V_{NT}^{-1}$ are the same order as $O_P(\zeta_{Ld}^{1/2}T^{-1/2})$. For $T^{-1}\boldsymbol{\varepsilon}_j^\tau B_{11}V_{NT}^{-1}$, we have

$$\left\|T^{-1}\boldsymbol{\varepsilon}_j^\tau B_{11}V_{NT}^{-1}\right\|\leq\frac{1}{\sqrt{T}}\frac{\|\boldsymbol{\varepsilon}_j^\tau\boldsymbol{F}^0\|}{\sqrt{T}}\frac{1}{N\sqrt{T}}\sum_{i=1}^N\|\lambda_i\|\|V_{NT}^{-1}\|\sqrt{r\sum_{t=1}^T e_{it}^2}=O_P\Big(\zeta_{Ld}^{1/2}T^{-1/2}\Big).$$

This leads to $T^{-1/2}\|\boldsymbol{\varepsilon}_j^\tau\boldsymbol{F}^0\|=O_P(1)$. Similarly, $\|T^{-1}\boldsymbol{\varepsilon}_j^\tau B_{12}V_{NT}^{-1}\Big\|=O_P(\zeta_{Ld}^{1/2}T^{-1/2})$. Thus, we finish the proof of (i).

(ii) By $\boldsymbol{F}^0-\hat{\boldsymbol{F}}H^{-1}=-(B_1+B_2+\cdots+B_{15})G$, we have

$$\begin{aligned}\frac{1}{T\sqrt{N}}\sum_{j=1}^N\boldsymbol{\varepsilon}_j^\tau(\hat{\boldsymbol{F}}H^{-1}-\boldsymbol{F}^0) &= \frac{1}{T\sqrt{N}}\sum_{j=1}^N\boldsymbol{\varepsilon}_j^\tau(B_1+B_2+\cdots+B_{15})G\\ &=: a_1+\cdots+a_{15}.\end{aligned}$$

Next we derive the orders of the fifteen terms, respectively. For the first four terms, we have

$$\begin{aligned}\|a_1\| &\leq T^{-1/2}\|G\|\left(\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{\sqrt{NT}}\sum_{j=1}^N\sum_{t=1}^T\varepsilon_{jt}R_{it}\right\|\Big(\frac{\|\boldsymbol{R}_i\|^2}{T}\Big)\right)\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|^2\\ &= T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|^2),\\ a_2 &= \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{j=1}^N\sum_{i=1}^N\boldsymbol{\varepsilon}_j^\tau\boldsymbol{R}_i(\widetilde{\boldsymbol{\gamma}}-\hat{\boldsymbol{\gamma}})\lambda_i^\tau\Big(\frac{\Lambda^\tau\Lambda}{N}\Big)^{-1}\\ &= \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^N\frac{1}{\sqrt{NT}}\sum_{j=1}^N\sum_{t=1}^T\varepsilon_{jt}R_{it}(\widetilde{\boldsymbol{\gamma}}-\hat{\boldsymbol{\gamma}})\lambda_i^\tau\Big(\frac{\Lambda^\tau\Lambda}{N}\Big)^{-1}\\ &= T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|),\end{aligned}$$

$$\begin{aligned}
\|a_3\| &\leq T^{-1/2}\|G\|\left(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{jt}R_{it}\Big\|\Big(\frac{\|\boldsymbol{\varepsilon}_i\|^2}{T}\Big)\right)\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\| \\
&= T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|), \\
\|a_4\| &\leq T^{-1/2}\|G\|\left\|\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{jt}F_t^{\tau}\right\|\left\|\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{\boldsymbol{R}_i^{\tau}\hat{\boldsymbol{F}}}{T}\Big)\right\|\|\lambda_i\|\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\| \\
&= T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|).
\end{aligned}$$

For a_5, let $\boldsymbol{W}_i = \boldsymbol{R}_i^{\tau}\hat{\boldsymbol{F}}/T$. It is easy to verify that $\|\boldsymbol{W}_i\|^2 \leq \|\boldsymbol{R}_i\|^2/T = O_P(1)$. Further,

$$\begin{aligned}
a_5 &= \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_j^{\tau}\boldsymbol{\varepsilon}_i(\widetilde{\boldsymbol{\gamma}}-\hat{\boldsymbol{\gamma}})^{\tau}\boldsymbol{W}_iG \\
&= \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\varepsilon_{jt}\Big)\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}(\widetilde{\boldsymbol{\gamma}}-\hat{\boldsymbol{\gamma}})^{\tau}\boldsymbol{W}_i\Big)G \\
&= N^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|).
\end{aligned}$$

For a_6, we have

$$\begin{aligned}
a_6 &= \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\boldsymbol{\varepsilon}_j^{\tau}\boldsymbol{F}^0\sum_{i=1}^{N}\lambda_i\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}}G \\
&= \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\boldsymbol{\varepsilon}_j^{\tau}\boldsymbol{F}^0\sum_{i=1}^{N}\lambda_i\boldsymbol{\varepsilon}_i^{\tau}\boldsymbol{F}^0HG + \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\boldsymbol{\varepsilon}_j^{\tau}\boldsymbol{F}^0\sum_{i=1}^{N}\lambda_i\boldsymbol{\varepsilon}_i^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)G \\
&=: a_{6.1}+a_{6.2}.
\end{aligned}$$

By the proof of Lemma A.4 in Bai (2009s), $a_{6.1} = O_P(T^{-1}N^{-1/2})$. Also,

$$a_{6.2} = T^{-1/2}\left(\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{jt}F_t^{0\tau}\right)\frac{1}{NT}\sum_{i=1}^{N}\lambda_i\boldsymbol{\varepsilon}_i^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)G.$$

By (i) of Lemma 3 and some elementary calculations, we have

$$\begin{aligned}
\|a_{6.2}\| &\leq T^{-1/2}O_P(1)\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|\|T^{-1/2}\boldsymbol{\varepsilon}_i\|\frac{\|\hat{\boldsymbol{F}}-\boldsymbol{F}^0H\|}{\sqrt{T}}\|G\| \\
&= T^{-1/2}\Big[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\Big].
\end{aligned}$$

Since a_7 and a_8 have the same structures as a_7 and a_8 in Bai (2009s), we can prove that $a_7 = O_P(N^{-1/2})$ and $a_8 = O_P(T^{-1}) + O_P((NT)^{-1/2}) + N^{-1/2}[O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-1}) + O_P(\zeta_{Ld}^{1/2})]$. For a_9, by (C.2) we have

$$\begin{aligned} \|a_9\| &\leq \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{jt}R_{it}\right\|T^{-1/2}\sqrt{r\sum_{t=1}^{T}e_{it}^2}\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|\|G\| \\ &= T^{-1/2}O_P\Big(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|\zeta_{Ld}^{1/2}\Big). \end{aligned}$$

Similarly, $a_{10} = T^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|\zeta_{Ld}^{1/2})$. For a_{11}, we have

$$\begin{aligned} \|a_{11}\| &\leq T^{-1/2}\left\|\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{jt}F_t^{\tau}\right\|\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|T^{-1/2}\sqrt{r\sum_{t=1}^{T}e_{it}^2}\|G\| \\ &= T^{-1/2}O_P\Big(\zeta_{Ld}^{1/2}\Big). \end{aligned}$$

For a_{12}, we have

$$\begin{aligned} a_{12} &= \frac{1}{\sqrt{N}}\frac{1}{NT}\sum_{j=1}^{N}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_j^{\tau}\boldsymbol{e}_i\lambda_i^{\tau}\Big(\frac{\Lambda^{\tau}\Lambda}{N}\Big)^{-1} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\varepsilon_{jt}\right)\left(\frac{1}{N}\sum_{i=1}^{N}e_{it}\lambda_i^{\tau}\right)\right]\Big(\frac{\Lambda^{\tau}\Lambda}{N}\Big)^{-1} \\ &= O_P\Big(\zeta_{Ld}^{1/2}\Big). \end{aligned}$$

For a_{13}, let $\widetilde{\boldsymbol{W}}_i = \boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}}/T$. Then we have $\|\widetilde{\boldsymbol{W}}_i\| = \|\boldsymbol{e}_i\|\sqrt{r}/\sqrt{T} = O_P(\zeta_{Ld}^{1/2})$ and

$$\begin{aligned} a_{13} &= \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\varepsilon_{jt}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}\widetilde{\boldsymbol{W}}_i\right)\right]G \\ &= N^{-1/2}O_P\Big(\zeta_{Ld}^{1/2}\Big). \end{aligned}$$

Finally, we can obtain that

$$a_{14} = N^{-1/2}O_P\Big(\zeta_{Ld}^{1/2}\Big) \;\text{ and }\; a_{15} = O_P(\zeta_{Ld}).$$

Summarizing the above results, we finish the proof of (ii).

(iii) Part (iii) follows immediately from (ii) by noting that the presence of λ_j does not alter the results. □

Lemma 5 *Assume that assumptions (A1)–(A9) hold. We have*

$$\begin{aligned}
&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}-\Omega_j)\hat{\boldsymbol{F}}G\lambda_i\\
= \;& O_P(1/(T\sqrt{N}))+(NT)^{-1/2}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\right]\\
&+\frac{1}{\sqrt{N}}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\right]^2.
\end{aligned}$$

Proof Some elementary calculations yield that

$$\begin{aligned}
&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}M_{\hat{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}-\Omega_j)\hat{\boldsymbol{F}}G\lambda_i\\
= \;& \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}-\Omega_j)\hat{\boldsymbol{F}}G\lambda_i\\
&-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}\Big(\frac{\hat{\boldsymbol{F}}\hat{\boldsymbol{F}}^{\tau}}{T}\Big)(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}-\Omega_j)\hat{\boldsymbol{F}}G\lambda_i\\
=: \;& I+II.
\end{aligned}$$

For the first term, by some basic calculations we have

$$\begin{aligned}
I \;=\;& \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}-\Omega_j)\boldsymbol{F}^0HG\lambda_i\\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{R}_i^{\tau}(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\tau}-\Omega_j)(\hat{\boldsymbol{F}}-\boldsymbol{F}^0H)G\lambda_i\\
=: \;& I_1+I_2.
\end{aligned}$$

For I_1, invoking Lemma A.2 (i) in Bai (2009) and Assumption (A8)(iv), it is easy to show that

$$\begin{aligned} I_1 &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\{\sum_{t=1}^{T}\sum_{s=1}^{T}R_{it}[\varepsilon_{jt}\varepsilon_{js}-E(\varepsilon_{jt}\varepsilon_{js})]F_s^{0\tau}HG\lambda_i\Big\} \\ &= \frac{1}{T\sqrt{N}}\frac{1}{N}\sum_{i=1}^{N}\Big\{\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\frac{1}{T}R_{it}[\varepsilon_{jt}\varepsilon_{js}-E(\varepsilon_{jt}\varepsilon_{js})]F_s^{0\tau}\Big\}HG\lambda_i \\ &= O_P\left(\frac{1}{T\sqrt{N}}\right). \end{aligned}$$

Let

$$a_s = \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{t=1}^{T}R_{it}[\varepsilon_{jt}\varepsilon_{js}-E(\varepsilon_{jt}\varepsilon_{js})] = O_P(1).$$

Then we have

$$I_2 = \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{s=1}^{T}a_s(\hat{F}_s-F_s^0H)^\tau G\lambda_i.$$

By Cauchy-Schwarz inequality and Lemma 3 (i), we have

$$\begin{aligned} \Big\|\frac{1}{T}\sum_{s=1}^{T}a_s(\hat{F}_s-F_s^0H)\Big\| &\le \Big(\frac{1}{T}\sum_{s=1}^{T}\|a_s\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{s=1}^{T}\|\hat{F}_s-F_s^0H\|^2\Big)^{1/2} \\ &= O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big). \end{aligned}$$

This leads to

$$I_2 = (NT)^{-1/2}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\right].$$

For the second term, by the similar proof of Lemma A.4 (ii) in Bai

(2009), we have

$$
\begin{aligned}
\|II\| &\le \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{\boldsymbol{R}_i^\tau \hat{\boldsymbol{F}}}{T}\Big\|\|G\lambda_i\|\Big\|\frac{1}{NT^2}\sum_{j=1}^{N}\hat{\boldsymbol{F}}^\tau(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^\tau-\Omega_j)\hat{\boldsymbol{F}}\Big\| \\
&= O_P(1)\Big\|\frac{1}{NT^2}\sum_{j=1}^{N}\hat{\boldsymbol{F}}^\tau(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^\tau-\Omega_j)\hat{\boldsymbol{F}}\Big\| \\
&= O_P(1/(T\sqrt{N}))+(NT)^{-1/2}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\right] \\
&\quad +\frac{1}{\sqrt{N}}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\right]^2.
\end{aligned}
$$

Summarizing the above results, we finish the proof of Lemma 5. □

Lemma 6 *Assume that assumptions (A1)–(A9) hold. We have*

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\left[\boldsymbol{R}_i^\tau M_{\hat{\boldsymbol{F}}}-\frac{1}{N}\sum_{j=1}^{N}a_{ij}\boldsymbol{R}_j^\tau M_{\hat{\boldsymbol{F}}}\right]\boldsymbol{\varepsilon}_i \\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\left[\boldsymbol{R}_i^\tau M_{\boldsymbol{F}^0}-\frac{1}{N}\sum_{j=1}^{N}a_{ij}\boldsymbol{R}_j^\tau M_{\boldsymbol{F}^0}\right]\boldsymbol{\varepsilon}_i+N^{-1}\xi_{NT}^*+N^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|^2) \\
&+(NT)^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+N^{-1/2}O_P(\delta_{NT}^{-2})+N^{-1/2}O_P\Big(\zeta_{Ld}^{1/2}\Big),
\end{aligned}
$$

where

$$
\xi_{NT}^*=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\boldsymbol{R}_i-\boldsymbol{V}_i)^\tau\boldsymbol{F}^0}{T}\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right)=O_P(1),
$$

with $\boldsymbol{V}_i=N^{-1}\sum_{j=1}^{N}a_{ij}\boldsymbol{R}_j$.

Proof For the term $\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau(M_{\boldsymbol{F}}-M_{\hat{\boldsymbol{F}}})\boldsymbol{\varepsilon}_i$, we consider the following decomposition:

$$\begin{aligned} M_{\boldsymbol{F}^0} - M_{\hat{\boldsymbol{F}}} &= P_{\hat{\boldsymbol{F}}} - P_{\boldsymbol{F}^0} \\ &= T^{-1}(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)H^\tau \boldsymbol{F}^{0\tau} + T^{-1}(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)^\tau \\ &\quad + T^{-1}\boldsymbol{F}^0 H(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)^\tau \\ &\quad + T^{-1}\boldsymbol{F}^0[HH^\tau - (T^{-1}\boldsymbol{F}^{0\tau}\boldsymbol{F}^0)^{-1}]\boldsymbol{F}^{0\tau}, \end{aligned}$$

for any invertible matrix H. Therefore, we have

$$\begin{aligned} &\frac{1}{NT}\sum_{i=1}^N \boldsymbol{R}_i^\tau (M_{\boldsymbol{F}^0} - M_{\hat{\boldsymbol{F}}})\boldsymbol{\varepsilon}_i \\ = &\frac{1}{NT}\sum_{i=1}^N \frac{\boldsymbol{R}_i^\tau(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)}{T} H^\tau \boldsymbol{F}^{0\tau}\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^N \frac{\boldsymbol{R}_i^\tau(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)}{T}(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)^\tau \boldsymbol{\varepsilon}_i \\ &+ \frac{1}{NT}\sum_{i=1}^N \frac{\boldsymbol{R}_i^\tau \boldsymbol{F}^0 H}{T}(\hat{\boldsymbol{F}} - \boldsymbol{F}^0 H)^\tau \boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^N \frac{\boldsymbol{R}_i^\tau \boldsymbol{F}^0}{T}[HH^\tau - (T^{-1}\boldsymbol{F}^{0\tau}\boldsymbol{F}^0)^{-1}]\boldsymbol{F}^{0\tau}\boldsymbol{\varepsilon}_i \\ =: &\ s_1 + s_2 + s_3 + s_4. \end{aligned}$$

For s_1, noting that $(\hat{F}_s - H^\tau F_s^0)^\tau H^\tau F_t^0$ is scalar, we have

$$s_1 = \frac{1}{\sqrt{NT}}\frac{1}{T}\sum_{s=1}^T (\hat{F}_s - H^\tau F_s^0)^\tau H^\tau \left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T F_t^0 R_{is}\varepsilon_{it}\right).$$

Further, we can derive that

$$\begin{aligned} \|s_1\| &\leq \frac{1}{\sqrt{NT}}\left[\frac{1}{T}\sum_{s=1}^T \|\hat{F}_s - H^\tau F_s^0\|^2\right]^{1/2} \|H\| \left[\frac{1}{T}\sum_{s=1}^T \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T F_t^0 R_{is}\varepsilon_{it}\right\|^2\right]^{1/2} \\ &= \frac{1}{\sqrt{NT}}\left[O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-1}) + O_P\left(\zeta_{Ld}^{1/2}\right)\right] O_P(1) \\ &= o_P((NT)^{-1/2}). \end{aligned}$$

Similarly, we can obtain that

$$s_2 = \frac{1}{\sqrt{N}}\frac{1}{T^2}\sum_{s=1}^T\sum_{t=1}^T (\hat{F}_s - H^\tau F_s^0)^\tau(\hat{F}_t - H^\tau F_t^0)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N R_{is}\varepsilon_{it}\right),$$

and

$$\begin{aligned}\|s_2\| &\leq \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H^\tau F_t^0\|^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}R_{is}\varepsilon_{it}\right\|^2\right)^{1/2}\\ &= \frac{1}{\sqrt{N}}\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-1})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\right]^2 O_P(1).\end{aligned}$$

For s_3, by some simple calculations we have

$$\begin{aligned}s_3 &= \frac{1}{NT}\sum_{i=1}^{N}\frac{\boldsymbol{R}_i^\tau\boldsymbol{F}^0}{T}HH^\tau(\hat{\boldsymbol{F}}H^{-1}-\boldsymbol{F}^0)^\tau\boldsymbol{\varepsilon}_i\\ &= \frac{1}{NT}\sum_{i=1}^{N}\frac{\boldsymbol{R}_i^\tau\boldsymbol{F}^0}{T}\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}(\hat{\boldsymbol{F}}H^{-1}-\boldsymbol{F}^0)^\tau\boldsymbol{\varepsilon}_i\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\frac{\boldsymbol{R}_i^\tau\boldsymbol{F}^0}{T}\left[HH^\tau-\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\right](\hat{\boldsymbol{F}}H^{-1}-\boldsymbol{F}^0)^\tau\boldsymbol{\varepsilon}_i\\ &=: s_{3.1}+s_{3.2}.\end{aligned}$$

Let $Q = HH^\tau - (\boldsymbol{F}^{0\tau}\boldsymbol{F}^0/T)^{-1}$. By Lemma 4 (iii) and Lemma 3 (vi), we have

$$\begin{aligned}s_{3.2} &= \left(\frac{1}{NT}\sum_{i=1}^{N}\left[\boldsymbol{\varepsilon}_i^\tau(\hat{\boldsymbol{F}}H^{-1}-\boldsymbol{F}^0)\otimes\left(\frac{\boldsymbol{R}_i^\tau\boldsymbol{F}^0}{T}\right)\right]\right)\mathrm{vec}(Q)\\ &= \left[(TN)^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(N^{-1})+N^{-1/2}O_P(\delta_{NT}^{-2})+N^{-1/2}O_P\Big(\zeta_{Ld}^{1/2}\Big)\right]\\ &\quad\times\left[O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+O_P(\delta_{NT}^{-2})+O_P\Big(\zeta_{Ld}^{1/2}\Big)\right]\\ &= N^{-1}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+N^{-1}O_P(\delta_{NT}^{-2})+N^{-1/2}O_P(\delta_{NT}^{-4})+N^{-1}O_P\Big(\zeta_{Ld}^{1/2}\Big).\end{aligned}$$

Similarly to the proof of c_1 in Lemma A.8 in Bai (2009s), we have

$$s_{3.1} = N^{-1}\psi_{NT}+(NT)^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}}-\widetilde{\boldsymbol{\gamma}}\|)+N^{-1/2}O_P(\delta_{NT}^{-2})+N^{-1/2}O_P\Big(\zeta_{Ld}^{1/2}\Big),$$

where

$$\psi_{NT} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\boldsymbol{R}_i^\tau\boldsymbol{F}^0}{T}\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right) = O_P(1).$$

For s_4, note that $Q = HH^\tau - (\boldsymbol{F}^{0\tau}\boldsymbol{F}^0/T)^{-1}$. Then,

$$\begin{aligned} s_4 &= \frac{1}{NT}\sum_{i=1}^{N}\left[\boldsymbol{\varepsilon}_i^\tau \boldsymbol{F}^0 \otimes \left(\frac{\boldsymbol{R}_i^\tau \boldsymbol{F}^0}{T}\right)\right]\mathrm{vec}(Q) \\ &= \frac{1}{\sqrt{NT}}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t^0 \varepsilon_{it} \otimes \left(\frac{\boldsymbol{R}_i^\tau \boldsymbol{F}^0}{T}\right)\right]\mathrm{vec}(Q) \\ &= o_P(1), \end{aligned}$$

by the facts that $\mathrm{vec}(Q) = O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) + O_P(\delta_{NT}^{-2}) + O_P(\zeta_{Ld}^{1/2})$ and

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t^0 \varepsilon_{it} \otimes \left(\frac{\boldsymbol{R}_i^\tau \boldsymbol{F}^0}{T}\right) = O_P(1).$$

In summary, we have

$$\begin{aligned} & \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{R}_i^\tau (M_{\boldsymbol{F}^0} - M_{\hat{\boldsymbol{F}}})\boldsymbol{\varepsilon}_i \\ = \; & N^{-1}\psi_{NT} + N^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|^2) + (NT)^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) \\ & + N^{-1/2}O_P(\delta_{NT}^{-2}) + N^{-1/2}O_P\left(\zeta_{Ld}^{1/2}\right). \end{aligned} \tag{C.7}$$

Let $\boldsymbol{V}_i = N^{-1}\sum_{j=1}^{N} a_{ij}\boldsymbol{R}_j$. Replacing $\boldsymbol{R}_i$ with $\boldsymbol{V}_i$, by the same argument, we have

$$\begin{aligned} & \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{V}_i^\tau (M_{\boldsymbol{F}^0} - M_{\hat{\boldsymbol{F}}})\boldsymbol{\varepsilon}_i \\ = \; & N^{-1}\psi_{NT}^* + N^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|^2) + (NT)^{-1/2}O_P(\|\hat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\|) \\ & + N^{-1/2}O_P(\delta_{NT}^{-2}) + N^{-1/2}O_P\left(\zeta_{Ld}^{1/2}\right), \end{aligned} \tag{C.8}$$

where $\psi_{NT}^* = O_P(1)$ is defined as

$$\psi_{NT}^* = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\boldsymbol{V}_i^\tau \boldsymbol{F}^0}{T}\left(\frac{\boldsymbol{F}^{0\tau}\boldsymbol{F}^0}{T}\right)^{-1}\left(\frac{\Lambda^\tau\Lambda}{N}\right)^{-1}\lambda_j\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right).$$

Letting $\xi_{NT}^* = \psi_{NT} - \psi_{NT}^*$, and together with (C.7) and (C.8), we finish the proof of Lemma 6. □

Lemma 7 *Assume that assumptions (A1)–(A9) hold. We have*

$$D(\hat{\boldsymbol{F}})^{-1} - D(\boldsymbol{F}^0)^{-1} = o_P(1).$$

Proof Similarly to the proof of Lemma A.7 (ii) in Bai (2009), we can show that

$$\|P_{\hat{\boldsymbol{F}}} - P^0_{\boldsymbol{F}}\| = O_P(\|\hat{\gamma} - \tilde{\gamma}\|) + O_P(\delta_{NT}^{-2}) + O_P\left(\zeta_{Ld}^{1/2}\right). \tag{C.9}$$

This leads to

$$\begin{aligned} & D(\hat{\boldsymbol{F}}) - D(\boldsymbol{F}^0) \\ = & \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{R}_i^{\tau}(M_{\hat{\boldsymbol{F}}} - M_{\boldsymbol{F}^0})\boldsymbol{R}_i - \frac{1}{T}\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{R}_i^{\tau}(M_{\hat{\boldsymbol{F}}} - M_{\boldsymbol{F}^0})\boldsymbol{R}_j a_{ij}\Big] \\ = & \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{R}_i^{\tau}(P_{\hat{\boldsymbol{F}}} - P_{\boldsymbol{F}^0})\boldsymbol{R}_i - \frac{1}{T}\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{R}_i^{\tau}(P_{\hat{\boldsymbol{F}}} - P_{\boldsymbol{F}^0})\boldsymbol{R}_j a_{ij}\Big]. \end{aligned}$$

The norm of the first term in the above expression is bounded above by

$$\left\|\frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{R}_i^{\tau}(P_{\hat{\boldsymbol{F}}} - P_{\boldsymbol{F}^0})\boldsymbol{R}_i\right\| \le \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\|\boldsymbol{R}_i\|^2}{T}\right)\|P_{\hat{\boldsymbol{F}}} - P_{\boldsymbol{F}^0}\| = o_P(1).$$

Similarly, the order of the second term is also $o_P(1)$. Noting that $[D(\hat{\boldsymbol{F}}) + o_P(1)]^{-1} = D(\hat{\boldsymbol{F}})^{-1} + o_P(1)$, we complete the proof of Lemma 7. □

S4 Appendix D: Additive fixed effects model

In Appendix D, we also consider an important special case of model (1.2). By letting $\lambda_i = (\mu_i, 1)^{\tau}$ and $F_t = (1, \xi_t)^{\tau}$, model (1.2) reduces to the varying-coefficient panel-data model with additive fixed effects:

$$Y_{it} = X_{it}^{\tau}\boldsymbol{\beta}(U_{it}) + \mu_i + \xi_t + \varepsilon_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T. \tag{D.1}$$

Similar to (2.3), for the purpose of identification, we assume that

$$\sum_{i=1}^{N} \mu_i = 0 \quad \text{and} \quad \sum_{t=1}^{T} \xi_t = 0. \tag{D.2}$$

Invoking (2.1), we have

$$Y_{it} \approx \sum_{k=1}^{p} \sum_{l=1}^{L_k} \gamma_{kl} X_{it,k} B_{kl}(U_{it}) + \mu_i + \xi_t + \varepsilon_{it}. \tag{D.3}$$

Note that, if we further assume that $\sum_{t=1}^{T} \xi_t^2 = T$, then $\boldsymbol{\gamma}$ can be estimated by the iteration procedure described in Section 2. However, we need to estimate the fixed effects F_t and λ_i, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. In order to avoid estimating the fixed effects F_t and λ_i, we propose to remove the unknown fixed effects by a least squares dummy variable method based on the identification condition (D.2). The estimation procedure is described in what follows.

Let $\mathbf{1}_N$ denote an $N \times 1$ vector with all elements being ones, $\boldsymbol{Y} = (\boldsymbol{Y}_1^\tau, \ldots, \boldsymbol{Y}_N^\tau)^\tau$, $\mathbf{R} = (\boldsymbol{R}_1^\tau, \ldots, \boldsymbol{R}_N^\tau)^\tau$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\tau, \ldots, \boldsymbol{\varepsilon}_N^\tau)^\tau$, $\boldsymbol{\mu} = (\mu_2, \ldots, \mu_N)^\tau$ and $\boldsymbol{\xi} = (\xi_2, \ldots, \xi_T)^\tau$. By the identification condition (D.2), we have

$$\mathbf{D} = [-\mathbf{1}_{N-1} \; I_{N-1}]^\tau \otimes \mathbf{1}_T \quad \text{and} \quad \mathbf{S} = \mathbf{1}_N \otimes [-\mathbf{1}_{T-1} \; I_{T-1}]^\tau,$$

where $\otimes$ denotes the Kronecker product. Then model $(D.3)$ can be rewritten as the matrix form:

$$\boldsymbol{Y} \approx \mathbf{R}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\mu} + \mathbf{S}\boldsymbol{\xi} + \boldsymbol{\varepsilon}.$$

Next, we solve the following optimization problem:

$$\min_{\boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\xi}} (\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma} - \mathbf{D}\boldsymbol{\mu} - \mathbf{S}\boldsymbol{\xi})^\tau (\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma} - \mathbf{D}\boldsymbol{\mu} - \mathbf{S}\boldsymbol{\xi}). \tag{D.4}$$

Taking partial derivatives of (D.4) with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$, and setting them equal to zero, we have

$$\mathbf{D}^{\tau}(\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma} - \mathbf{D}\boldsymbol{\mu} - \mathbf{S}\boldsymbol{\xi}) = 0,$$

$$\mathbf{S}^{\tau}(\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma} - \mathbf{D}\boldsymbol{\mu} - \mathbf{S}\boldsymbol{\xi}) = 0.$$

By a simple calculation, we can obtain that

$$\begin{aligned} \tilde{\boldsymbol{\xi}} &= (\mathbf{S}^{\tau}\mathbf{S})^{-1}\mathbf{S}^{\tau}(\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma}), \\ \tilde{\boldsymbol{\mu}} &= (\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\left[\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma} - \mathbf{S}(\mathbf{S}^{\tau}\mathbf{S})^{-1}\mathbf{S}^{\tau}(\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma})\right]. \end{aligned}$$

Replacing $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ in (D.4) by $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\xi}}$ respectively, the parameter $\boldsymbol{\gamma}$ can be estimated by minimizing $(\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma})^{\tau}\boldsymbol{\Gamma}(\boldsymbol{Y} - \mathbf{R}\boldsymbol{\gamma})$, where $\boldsymbol{\Gamma} = \mathbf{H}(I_{NT} - \mathbf{S}(\mathbf{S}^{\tau}\mathbf{S})^{-1}\mathbf{S}^{\tau})$ and $\mathbf{H} = I_{NT} - \mathbf{D}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}$. Specifically, the least squares estimator of $\boldsymbol{\gamma}$ is

$$\check{\boldsymbol{\gamma}} = \left(\mathbf{R}^{\tau}\boldsymbol{\Gamma}\mathbf{R}\right)^{-1}\mathbf{R}^{\tau}\boldsymbol{\Gamma}\boldsymbol{Y}.$$

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

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

S5 Appendix E: Simulation studies

In Appendix E, we consider the following varying-coefficient panel-data model with individual fixed effects:

$$Y_{it} = X_{it,1}\beta_1(U_{it}) + X_{it,2}\beta_2(U_{it}) + \mu_i + \varepsilon_{it}, \tag{E.1}$$

where $\beta_1(u)$, $\beta_2(u)$, U_{it}, and ε_{it} are the same as those in model (7.2). The regressors $X_{it,1}$ and $X_{it,2}$ are generated according to

$$X_{it,1} = 3 + 2\mu_i + \eta_{it,1}, \quad X_{it,2} = 3 + 2\mu_i + \eta_{it,2},$$

where $\eta_{it,j} \sim N(0,1)$, $j = 1, 2$, and the fixed effects are generated by

$$\mu_i \sim N(0,1), \quad i = 2, \ldots, N \quad \text{and} \quad \mu_1 = -\sum_{i=2}^{N} \mu_i.$$

With 1000 repetitions, we report the simulation results in Table 5 and Figure 8, respectively.

Table 5: Finite sample performance of the estimators for model (E.1) with additive fixed effects.

		IFE		LSDVE	
N	T	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)	AMSE($\hat{\beta}_1$)	AMSE($\hat{\beta}_2$)
100	15	0.0115	0.0118	0.0093	0.0095
100	30	0.0048	0.0058	0.0044	0.0050
100	60	0.0024	0.0023	0.0021	0.0020
100	100	0.0012	0.0013	0.0011	0.0011
60	100	0.0024	0.0025	0.0020	0.0021
30	100	0.0052	0.0053	0.0047	0.0046
15	100	0.0127	0.0110	0.0108	0.0101

From Table 5 and Figure 8 we can see that the interactive fixed effects estimators and the least squares dummy variable estimators are all consistent. The interactive fixed effects estimators remain valid even for the

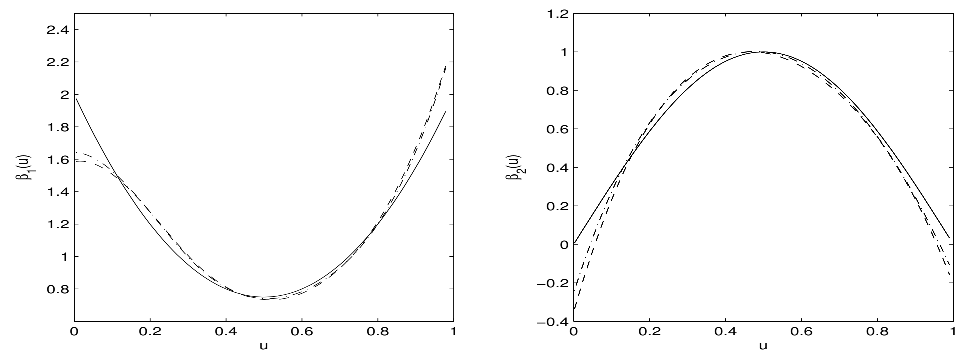

Figure 8: *Simulation results for model (E.1) when* $N = 100$, $T = 60$. *In each plot, the solid curves are for the true coefficient functions, the dashed curves are for the interactive fixed effects estimators, the dash-dotted curves are for the least squares dummy variable estimators.*

general fixed effects model. However, they are less efficient than the least squares dummy variable estimators.

Bibliography

Bai, J. S. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–171.

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

Bai, J. S. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*, 1133–1150.

de Boor, C. (2001). *A Practical Guide to Splines.* New York: Springer-Verlag.

He, X. M. and P. D. Shi (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *J. Nonparametr. Stat. 3*, 299–308.

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.

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.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Schumaker, L. L. (1981). *Spline Functions: Basic Theory.* New York: Wiley.

Tang, Q. G. and L. S. Cheng (2009). B-spline estimation for varying coefficient regression with spatial data. *Sci. China Ser. A: Mathematics 52*, 2321–2340.

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 (Corresponding author)

Department of Mathematics, Hong Kong Baptist University, Hong Kong

E-mail: hpeng@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)