On Time-varying Panel Data Models with Time-varying Interactive Fixed Effects*

Xia Wang[a], Yingxing Li[b], Junhui Qian[c], and Liangjun Su[d]

[a] School of Economics, Renmin University of China, China
[b] Wang Yanan Institute for Studies in Economics, Xiamen University, China
[c] Antai College of Economics and Management, Shanghai Jiao Tong University, China
[d] School of Economics and Management, Tsinghua University, China

March 29, 2023

Abstract

This paper introduces a time-varying (TV) panel data model with interactive fixed effects (IFEs) where both the coefficients and factor loadings are allowed to change smoothly over time. We propose a local version of the least squares and principal component method to estimate the TV coefficients, TV factor loadings, and common factors simultaneously. We provide a bias-corrected local least squares estimator for the TV coefficients and establish the limiting distributions and uniform convergence of the bias-corrected estimators, estimated factors, and factor loadings in the large N and large T framework. We also propose a BIC-type information criterion to determine the number of common factors in the IFEs, which is robust to the TV behavior in the coefficients and factor loadings. Based on the estimates, we propose three test statistics to gauge possible sources of TV features. We establish the limit null distributions and the asymptotic local power properties of our tests. Simulations are conducted to evaluate the finite sample performance of our information criterion, estimates, and tests. We apply our theoretical results to analyze the Phillips curve using the U.S. state-level unemployment rates and nominal wages, and document significant TV behavior in both the slope coefficient and factor loadings.

JEL Classification: C12, C14, C33, C38.

Key Words: Time varying; Panel data models; Interactive fixed effects; Specification test; Phillips curve

*Wang acknowledges financial support from the National Science Foundation of China (NSFC, No. 71873151), Su gratefully acknowledges the financial support from the NSFC (No. 72133002). Address Correspondence to Liangjun Su, Phone +86 10 62789506. Email:*sulj@sem.tsinghua.edu.cn*

1 Introduction

Since the seminal works of Pesaran (2006) and Bai (2009), panel data models with interactive fixed effects (IFEs) have been investigated extensively. See Moon and Weidner (2015, 2017), Chudik and Pesaran (20015), Lu and Su (2016), Li et al. (2020), among others. The IFEs could be regarded as a multiplicative form of individual effects and time effects, which allows for common shocks to affect the cross-sectional units with individual-specific sensitivities. In comparison with the traditional fixed effects models, the IFEs models not only allow flexible specifications of heterogeneity but also provide an effective way to model strong cross-sectional dependence, which is an essential feature of macroeconomic and financial data. Nevertheless, existing studies typically assume that the slope coefficients and factor loadings are time-invariant, which appears very restrictive given the long time span for such data. In fact, as technology progresses, preference changes, or economic transition occurs, the relationship between economic and financial variables may change over time. If a panel data model fails to account for such a time-varying (TV) behavior, statistical inferences, forecasting, and policy evaluations based on it can be misleading.

Numerous efforts have been devoted to modeling and testing structural changes in panel data models over the past decade. For example, De Wachter and Tzavalis (2012) propose a likelihood ratio test for a single structural break in dynamic panel data models with fixed effects in the large N and fixed T framework, where N and T denote the dimensions of the cross section and time, respectively. Qian and Su (2016) estimate multiple structural breaks in panel data models with fixed effects by adaptive group fused Lasso where N is large and T can be either large or fixed. Li et al. (2017) study the common correlated effects (CCE) estimation of heterogenous panels with unknown common breaks in the large N and large T framework; Baltagi et al. (2017) extend the model by allowing for nonstationary regressors. Apart from modeling, estimating, and testing for structural breaks in panel data models, the existing literature also investigates the case of smooth structural changes. Actually, the sources of structural changes, such as preference changes, technological progress, and institutional transformation, usually take effect gradually over time. Even if some policy switches occur immediately, it may take some time for economic agents to react. Thus, it is more realistic to assume smooth changes rather than abrupt breaks in many scenarios. Along this line of research, Li et al. (2011) introduce a TV functional-coefficient panel data model, Robinson (2012) proposes a nonparametric trending panel data model, both Chen (2019) and Su et al. (2019) study the estimation of a TV panel data model with latent group structures, and Chen and Huang (2018) develop two Hausman-type tests for smooth structural changes in panel data models. However, the aforementioned works focus on modeling and testing structural changes in panel data models with the usual additive fixed effects.

Given the appealing advantage of IFEs, some recent studies have considered the model instability problem of panel data models with IFEs. Li et al. (2016) consider Lasso-type estimation of panel data models with IFEs and multiple structural breaks in the large N and large T framework. Cheng et al. (2019) introduce a regime-switching panel data model with IFEs and develop an expectation and maximization (EM) algorithm to estimate the unknown parameters. Miao et al. (2020) study the estimation and inference in a panel threshold model with IFEs and propose a likelihood ratio test to check the existence of threshold effects. Dong et al. (2021) propose sieve estimation of a varying-coefficient panel data model with IFEs where the factor loadings are time-invariant. Similarly, Liu et al. (2019) consider the duple least squares estimation of a TV panel data model with individual-specific TV coefficients and time-invariant factor loadings. In contrast, Liu et al. (2021) study a TV panel with functional coefficients for both the regressors and factors and propose a hybrid of kernel and sieve methods to estimate the model.

In this paper, we model and test smooth structural changes in panel data models with IFEs under the local smoothing framework. We allow both the slope parameters and factor loadings to change over time and propose a local version of the least squares and principal component analysis (PCA) method to estimate the TV slope parameters, TV factor loadings, and unknown common factors simultaneously. Under the large N and large T framework, we establish the consistency and limiting distributions of the estimated slope parameters, the common factors, and factor loadings. Following Su and Wang (2017), we propose a BIC-type information criterion to determine the number of common factors, which is robust to the existence of structural changes in the slope coefficients and factor loadings. Based on the estimates, we propose three test statistics to gauge possible sources of TV features. We construct an L_2-distance-based test to check the stability of both slope coefficients and factor loadings. The basic idea is to estimate the TV panel data models with IFEs via the local least squares (LLS) method, and compare the estimated residuals with those obtained from the time-invariant model. If we reject the null hypothesis that both the slope coefficients and factor loadings are time-invariant, it is valuable to gauge the possible sources of rejection. We further propose two test statistics to check the stability of the slope coefficients and factor loadings separately. By construction, all of our tests can capture both smooth and abrupt structural changes, where neither the number of breaks nor break dates is unknown. The simulation studies show that the proposed LLS and local PCA method performs well, and our test statistics have reasonable size and excellent power. Empirically, we apply our modeling and testing framework to the Phillips curve using panel data of the U.S. state-level unemployment rates and nominal wages, and find significant evidence on the TV behavior of the Phillips curve and the factor loadings.

The rest of this paper is organized as follows. In Section 2, we introduce a TV panel data model with TV IFEs and the local PCA estimates of the model parameters. In Section 3, we study the asymptotic properties of the local PCA estimators of the slope coefficients, factors, and factor loadings. In Section 4, we construct various test statistics to check for the time-invariance of the slope coefficients or/and factor loadings, derive the asymptotic distributions of these test statistics under the null hypotheses, and investigate their local power properties. In Section 5, we study the finite sample performance of our estimators and test statistics via simulations. Section 6 provides an empirical application. Section 7 concludes. All proofs are relegated to an Online Supplement.

Notation. For an $m \times n$ real matrix A, we denote its transpose as A', its Frobenius norm as $\|A\|$ ($\equiv [\mathrm{tr}(AA')]^{1/2}$), its spectral norm as $\|A\|_{\mathrm{sp}}$ ($\equiv \sqrt{\mu_1 (A'A)}$) and its Moore-Penrose generalized inverse as A^+, where $\equiv$ signifies a definitional relationship and $\mu_s(\cdot)$ denotes the sth largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. Note that the two norms are equal when A is a vector. We will frequently use the submultiplicative property of these norms and the fact that $\|A\|_{\mathrm{sp}} \leq \|A\| \leq \|A\|_{\mathrm{sp}} \mathrm{rank}(A)^{1/2}$. We also use $\mu_{\max}(B)$ and $\mu_{\min}(B)$ to denote the largest and smallest eigenvalues of a symmetric matrix B, respectively. We use $B > 0$ to denote that B is positive definite. Let $P_A \equiv A(A'A)^+ A'$ and $M_A \equiv \mathbb{I}_m - P_A$, where $\mathbb{I}_m$ denotes an $m \times m$ identity matrix. Let $[a] = \{1, 2, ..., a\}$, when a is a positive integer. The operator $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, and plim probability limit. We use $(N, T) \to \infty$ to denote that N and T pass to infinity jointly.

2 The model and estimation

In this section, we propose a TV panel data model with TV IFEs and discuss the estimation of the parameters in the model.

2.1 The model

We consider the following TV panel data model with N cross-sectional units and T time periods:

$$Y_{it} = X_{it}' \beta_t + \lambda_{it}' F_t + \varepsilon_{it}, \; i \in [N], \; t \in [T], \tag{2.1}$$

where X_{it} is a $P \times 1$ vector of observable regressors, β_t is a $P \times 1$ vector of unknown TV coefficients, λ_{it} is an $R \times 1$ vector of TV factor loadings, F_t is an $R \times 1$ vector of unobserved common factors, and ε_{it} is an idiosyncratic error term with mean zero. The true values of β_t, F_t and λ_{it} are denoted as β_t^0, F_t^0 and λ_{it}^0, respectively, but we will frequently suppress the superscript 0 unless confusion may arise. At the moment, we assume the number of unobserved common factors, R, is known. We will introduce a BIC-type information criterion to determine it in Section 3.4.

The model in (2.1) generalizes Bai's (2009) panel data models with IFEs by allowing both the slope coefficients and factor loadings to vary over time, and it also extends the TV factor model of Su and Wang (2017) to allow for exogenous regressors. In order to capture various kinds of TV coefficients and factor loadings, we use a nonparametric local smoothing method to estimate both β_t and λ_{it} under some smoothness conditions. Specifically, we follow the nonparametric literature on TV models (see, e.g., Cai, 2007; Robinson, 2012; Chen et al., 2012; Su and Wang, 2017) and model β_t and λ_{it} as nonrandom functions of t/T:

$$\beta_t = \beta(t/T) \text{ and } \lambda_{it} = \lambda_i(t/T), \tag{2.2}$$

where $\beta(\cdot)$ and $\lambda_i(\cdot)$ are unknown smooth functions defined on $(0, 1]$.

The model in (2.1) includes some commonly used models as special cases. For example, when neither β_t nor λ_{it} varies over t, the model in (2.1) reduces to the model studied in Bai (2009), Lu and Su (2016), and Moon and Weidner (2015, 2017), among others. When $X_{it}'\beta_t$ is absent in (2.1), the model becomes the TV factor model studied in Su and Wang (2017). When the IFEs degenerate to the conventional additive fixed effects, the model in (2.1) becomes Li et al.'s (2011) TV coefficient panel data models with additive fixed effects.

2.2 Estimation

Following the literature on TV models, we propose to estimate the model in (2.1) by a local smoothing procedure.

For the moment, fix $r \in [T]$. Under the assumption that $\lambda_i(\cdot) : (0,1] \mapsto \mathbb{R}^R$ and $\beta(r) : (0,1] \mapsto \mathbb{R}^P$ are continuously differentiable up to the second order, for any t/T close to r/T, we have

$$\lambda_{it} = \lambda_i\left(\frac{t}{T}\right) = \lambda_i\left(\frac{r}{T}\right) + \left[\lambda_i\left(\frac{t}{T}\right) - \lambda_i\left(\frac{r}{T}\right)\right] \equiv \lambda_{ir} + d_i(t,r), \text{ and}$$
$$\beta_t = \beta\left(\frac{t}{T}\right) = \beta\left(\frac{r}{T}\right) + \left[\beta\left(\frac{t}{T}\right) - \beta\left(\frac{r}{T}\right)\right] \equiv \beta_r + d_0(t,r).$$

That is, we approximate λ_{it} by λ_{ir} with an approximation error $d_i(t,r)$ and β_t by β_r with an approximation error $d_0(t,r)$. It follows that

$$Y_{it} = X_{it}'\beta_r + \lambda_{ir}'F_t + \Delta_i(t,r) + \varepsilon_{it}. \tag{2.3}$$

where $\Delta_i(t,r) = X_{it}'d_0(t,r) + d_i(t,r)'F_t$. The term $\Delta_i(t,r)$ represents the combined approximation error in the regression model and it will generate some bias terms in the estimation of the TV slope coefficients and factor loadings.

To estimate $\{\lambda_{ir}\}_{i=1}^N$, $\{F_t\}_{t=1}^T$ and β_r, we consider the following LLS problem:

$$\min_{\{\lambda_{ir}\}_{i=1}^N,\{F_t\}_{t=1}^T,\beta_r} (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \left[Y_{it} - X_{it}'\beta_r - \lambda_{ir}'F_t\right]^2 K_h\left(\frac{t-r}{T}\right) \tag{2.4}$$

subject to some identification restrictions. Here, $K_h(x) = h^{-1}K(x/h)$ with kernel $K : \mathbb{R} \mapsto \mathbb{R}^+$ and bandwidth $h = h(T,N)$. Multiplying both sides of (2.3) by $k_{h,tr} \equiv h^{-1}K((t-r)/(Th))$ yields the transformed model:

$$k_{h,tr}^{1/2}Y_{it} \approx k_{h,tr}^{1/2}X_{it}'\beta_r + k_{h,tr}^{1/2}\lambda_{ir}'F_t + k_{h,tr}^{1/2}\varepsilon_{it} \quad \text{when } \frac{t}{T} \approx \frac{r}{T}. \tag{2.5}$$

Denote the $T \times N$ matrices $Y^{(r)} = (Y_1^{(r)}, \cdots, Y_N^{(r)})$, and $\varepsilon^{(r)} = (\varepsilon_1^{(r)}, \cdots, \varepsilon_N^{(r)})$, where $Y_i^{(r)} = (k_{h,1r}^{1/2}Y_{i1}, \cdots, k_{h,Tr}^{1/2}Y_{iT})'$ and $\varepsilon_i^{(r)} = (k_{h,1r}^{1/2}\varepsilon_{i1}, \cdots, k_{h,Tr}^{1/2}\varepsilon_{iT})'$. Denote $\mathbf{X}^{(r)}$ as a $T \times N \times P$ three-dimensional tensor with P sheets, the pth sheet of which is denoted as $\mathbf{X}_p^{(r)} = \{k_{h,tr}^{1/2}X_{it,p}\}$, a

$T \times N$ matrix, where $X_{it,p}$ is the pth element of X_{it} for $p \in [P]$. Note that the cross-product $\mathbf{X}^{(r)}\beta_r$ is a $T \times N$ matrix with the (t,i)th element as $k_{h,tr}^{1/2} X_{it}'\beta_r$. Let $X_i^{(r)} = (k_{h,1r}^{1/2} X_{i1}, ..., k_{h,Tr}^{1/2} X_{iT})'$, $F^{(r)} = (k_{h,1r}^{1/2} F_1, \cdots, k_{h,Tr}^{1/2} F_T)'$ and $\Lambda_r = (\lambda_{1r}, \cdots, \lambda_{Nr})'$. Then, we can rewrite the transformed model in (2.5) in vector form:

$$Y_i^{(r)} \approx X_i^{(r)}\beta_r + F^{(r)}\lambda_{ir} + \varepsilon_i^{(r)},$$

or in matrix notation: $Y^{(r)} \approx \mathbf{X}^{(r)}\beta_r + F^{(r)}\Lambda_r' + \varepsilon^{(r)}$. Then the minimization problem becomes

$$\min_{\Lambda_r, F^{(r)}, \beta_r} (NT)^{-1} \sum_{i=1}^{N} \left(Y_i^{(r)} - X_i^{(r)}\beta_r - F^{(r)}\lambda_{ir}\right)' \left(Y_i^{(r)} - X_i^{(r)}\beta_r - F^{(r)}\lambda_{ir}\right) \tag{2.6}$$

subject to the constraints that $F^{(r)\prime}F^{(r)}/T = \mathbb{I}_R$ and $\Lambda_r'\Lambda_r$ is diagonal with elements arranged in descending order along its main diagonal line. By concentrating Λ_r out, we obtain the following minimization problem

$$\min_{\beta_r, F^{(r)}} (NT)^{-1} \sum_{i=1}^{N} \left(Y_i^{(r)} - X_i^{(r)}\beta_r\right)' M_{F^{(r)}} \left(Y_i^{(r)} - X_i^{(r)}\beta_r\right). \tag{2.7}$$

where $M_{F^{(r)}} = \mathbb{I}_T - F^{(r)}(F^{(r)\prime}F^{(r)})^{-1}F^{(r)\prime} = \mathbb{I}_T - F^{(r)}F^{(r)\prime}/T \equiv \mathbb{I}_T - P_{F^{(r)}}$.

Given $F^{(r)}$, we can readily obtain the LLS estimator of β_r from (2.7). Given β_r, the variable $W_i^{(r)} = Y_i^{(r)} - X_i^{(r)}\beta_r$ has an approximate factor structure: $W_i^{(r)} \approx F^{(r)}\lambda_{ir} + \varepsilon_i^{(r)}$. Then we can estimate the normalized factor $F^{(r)}$ via the standard PCA. Such observations motivate us to obtain the LLS estimator $(\hat{\beta}_r, \hat{F}^{(r)})$ as the solution to the following set of nonlinear equations:

$$\hat{\beta}_r = \left(\sum_{i=1}^{N} X_i^{(r)\prime} M_{\hat{F}^{(r)}} X_i^{(r)}\right)^{-1} \sum_{i=1}^{N} X_i^{(r)\prime} M_{\hat{F}^{(r)}} Y_i^{(r)}, \tag{2.8}$$

$$\hat{F}^{(r)}\hat{V}_{NT}^{(r)} = \left[\frac{1}{NT}\sum_{i=1}^{N} \left(Y_i^{(r)} - X_i^{(r)}\hat{\beta}_r\right)\left(Y_i^{(r)} - X_i^{(r)}\hat{\beta}_r\right)'\right]\hat{F}^{(r)}, \tag{2.9}$$

where $\hat{V}_{NT}^{(r)}$ is a diagonal matrix that consists of the R largest eigenvalues of the matrix in the square brackets in (2.9), arranged in descending order along its main diagonal line.

Note that $\hat{F}^{(r)}/\sqrt{T}$ is a collection of the normalized eigenvectors of $\frac{1}{NT}\sum_{i=1}^{N}(Y_i^{(r)} - X_i^{(r)}\hat{\beta}_r)(Y_i^{(r)} - X_i^{(r)}\hat{\beta}_r)'$ associated with its R largest eigenvalues. Given $\hat{\beta}_r$ and $\hat{F}^{(r)}$, we obtain the estimator of $\Lambda^{(r)\prime}$ by $\hat{\Lambda}_r' = (\hat{F}^{(r)}\hat{F}^{(r)\prime})^{-1}\hat{F}^{(r)\prime}(Y_i^{(r)} - X_i^{(r)}\hat{\beta}_r) = \hat{F}^{(r)\prime}(Y_i^{(r)} - X_i^{(r)}\hat{\beta}_r)/T$ for $r \in [T]$. We use $\hat{\lambda}_{ir}$ to denote the ith column of $\hat{\Lambda}_r'$.

2.3 Boundary kernel

It is well known that a local constant estimator may suffer from the boundary bias problem when we estimate a conditional mean object. To avoid the boundary bias problem associated with the kernel estimation and to facilitate the study of some uniform convergence results, we follow Hong and Li (2005) and Li and Racine (2006, p.31) to apply the following boundary kernel:

$$k_{h,tr}^{*} = h^{-1}K_r^{*}\left(\frac{t-r}{Th}\right) = \begin{cases} h^{-1}K\left(\frac{t-r}{Th}\right) / \int_{-r/(Th)}^{1} K(u)du, & \text{if } r \in [1, \lfloor Th \rfloor) \\ h^{-1}K\left(\frac{t-r}{Th}\right), & \text{if } r \in [\lfloor Th \rfloor, T - \lfloor Th \rfloor] \\ h^{-1}K\left(\frac{t-r}{Th}\right) / \int_{-1}^{(1-r/T)/h} K(u)du, & \text{if } r \in (T - \lfloor Th \rfloor, T] \end{cases},$$

where $\lfloor Th \rfloor$ denotes the integer part of Th. Note that $k_{h,tr}^{*}$ coincides with $k_{h,tr}$ in the interior region but not in the boundary regions. By using this boundary kernel to replace $k_{h,tr} = K_h(\frac{t-r}{T})$ in (2.4)-(2.9), we obtain the estimators to be analyzed.

2.4 Updated estimation of the factors

The estimator $\hat{F}_t^{(r)}$ is only consistent for a rotational version of the weighted factor $F_t^{(r)} \equiv k_{h,tr}^{*1/2} F_t$. To obtain a consistent estimator of F_t after suitable rotation, we consider a two-stage estimation procedure. Based on the consistent estimators of λ_{it} and β_t obtained in the first stage, we can obtain the consistent estimators of F_t, $t \in [T]$, in the second stage, by considering the following least squares problem:

$$\min_{F_t \in \mathbb{R}^R} \sum_{i=1}^{N} \left[Y_{it} - X_{it}'\hat{\beta}_t - \hat{\lambda}_{it}' F_t \right]^2, \quad t \in [T].$$

The solution to the above problem is: $\hat{F}_t = \left(\sum_{i=1}^{N} \hat{\lambda}_{it}\hat{\lambda}_{it}'\right)^{-1} \sum_{i=1}^{N} \hat{\lambda}_{it}(Y_{it} - X_{it}'\hat{\beta}_t)$ for $t \in [T]$. Let $\hat{C}_{it} = \hat{\lambda}_{it}'\hat{F}_t$, which is an estimator of the common component $C_{it} \equiv \lambda_{it}' F_t$.

3 Asymptotic Properties of the Estimators

In this section, we study the asymptotic distributions of the estimators of the TV coefficients, factors, and TV factor loadings.

3.1 Basic assumptions

Define the $T \times P$ matrix

$$Z_i^{(r)} = M_{F^{(r)}} \left[X_i^{(r)} - \frac{1}{N} \sum_{j=1}^{N} X_j^{(r)} a_{ij}^{(r)} \right],$$

where $i \in [N]$, $a_{ij}^{(r)} = \lambda_{ir}' \left(\Lambda_r'\Lambda_r/N\right)^{-1} \lambda_{jr}$. Obviously, $a_{ij}^{(r)} = a_{ji}^{(r)}$, $N^{-1}\sum_{l=1}^{N} a_{il}^{(r)} a_{lj}^{(r)} = a_{ij}^{(r)}$, and $Z_i^{(r)}$ is equal to the deviation of $M_{F^{(r)}} X_i^{(r)}$ from its weighted average with the weighting vector $a_i^{(r)} = (a_{i1}^{(r)}, \cdots, a_{iN}^{(r)})'$. Define the $P \times P$ matrix

$$D^{(r)}(F^{(r)}) = \frac{1}{NT} \sum_{i=1}^{N} X_i^{(r)\prime} M_{F^{(r)}} X_i^{(r)} - \frac{1}{T} \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} X_i^{(r)\prime} M_{F^{(r)}} X_j^{(r)} a_{ij}^{(r)} \right] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} Z_{it}^{(r)} Z_{it}^{(r)\prime},$$

where $Z_{it}^{(r)\prime}$ denotes the tth row of $Z_i^{(r)}$. Apparently, $D^{(r)}(F^{(r)})$ is semi-positive definite (p.s.d.) for any r and $F^{(r)}$. Let $M_{\Lambda_r} = \mathbb{I}_N - \Lambda_r(\Lambda_r'\Lambda_r)^{-1}\Lambda_r' \equiv \mathbb{I}_N - P_{\Lambda_r}$. Recall that $\mathbf{X}_p^{(r)}$ denotes the pth sheet of the three dimensional tensor $\mathbf{X}^{(r)}$. Define

$$\bar{\mathbf{X}}_p^{(r)} = E(\mathbf{X}_p^{(r)}),\ \tilde{\mathbf{X}}_p^{(r)} = \mathbf{X}_p^{(r)} - \bar{\mathbf{X}}_p^{(r)},\ \text{and}\ \mathfrak{X}_p^{(r)} = M_{F^{(r)}}\bar{\mathbf{X}}_p^{(r)}M_{\Lambda_r} + \tilde{\mathbf{X}}_p^{(r)}.$$

Define a three-dimensional $T \times N \times P$ tensor $\mathbf{Z}^{(r)}$ whose pth sheet is denoted as $\mathbf{Z}_p^{(r)}$ (a $T \times N$ matrix) and given by

$$\mathbf{Z}_p^{(r)} = M_{F^{(r)}}\mathbf{X}_p^{(r)}M_{\Lambda_r} = \mathfrak{X}_p^{(r)} - P_{F^{(r)}}\tilde{\mathbf{X}}_p^{(r)} - \tilde{\mathbf{X}}_p^{(r)}P_{\Lambda_r} + P_{F^{(r)}}\tilde{\mathbf{X}}_p^{(r)}P_{\Lambda_r}.$$

Note that the elements of $\mathfrak{X}_p^{(r)}$, say $\mathfrak{X}_{p,it}^{(r)}$, are contemporaneously uncorrelated with the error terms $\varepsilon_{it}^{(r)}$ and ε_{it}.

Let $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$, $\gamma_N(s,t) = N^{-1}E(\varepsilon_s'\varepsilon_t)$, $\gamma_{N,F}(s,t) = N^{-1}E(F_s\varepsilon_s'\varepsilon_t)$, $\gamma_{N,FF}(s,t) = N^{-1}E(F_s\varepsilon_s'\varepsilon_tF_t')$, and $\zeta_{st} = N^{-1}[\varepsilon_s'\varepsilon_t - E(\varepsilon_s'\varepsilon_t)]$. Define

$$\begin{aligned}
\varpi_{NT,1}(r) &= \frac{h^{1/2}}{\sqrt{NT}}F^{(r)\prime}\varepsilon^{(r)}\Lambda_r = \frac{h^{1/2}}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}k_{h,tr}^{*}F_t\varepsilon_{it}\lambda_{ir}', \\
\varpi_{NT,2}(r,t) &= \frac{h^{1/2}}{\sqrt{NT}}[F^{(r)\prime}\varepsilon^{(r)}\varepsilon_t - E(F^{(r)\prime}\varepsilon^{(r)}\varepsilon_t)] = \frac{h^{1/2}}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}k_{h,sr}^{*}[F_s\varepsilon_{is}\varepsilon_{it} - E(F_s\varepsilon_{is}\varepsilon_{it})].
\end{aligned}$$

Let $\tau_{ij,s} = E(\varepsilon_{is}\varepsilon_{js}F_s'F_s)$. Let $C < \infty$ denote a generic positive constant that may vary from case to case. Let $\max_{i,t} = \max_{1\le i\le N}\max_{1\le t\le T}$. Define analogously $\max_{i,j}$, $\max_{t,s}$, $\max_i$, and $\max_t$, etc. Let $\sum_{i,j} = \sum_{i=1}^{N}\sum_{j=1}^{N}$ and $\sum_{t,s} = \sum_{t=1}^{T}\sum_{s=1}^{T}$. Let $\mathcal{F} = \{F^{(r)} : F^{(r)\prime}F^{(r)} = \mathbb{I}_R\}$ and $\rho_{\min}^{(r)} = \inf_{F^{(r)}\in\mathcal{F}} \mu_{\min}(D^{(r)}(F^{(r)}))$. Let $\mathcal{C} = \sigma(F^0)$, the minimal-sigma field generated from the true common factor $F^0 = (F_1^0, ..., F_T^0)'$. We use $E_{\mathcal{C}}(\cdot)$ and $\mathrm{Var}_{\mathcal{C}}(\cdot)$ to denote expectation and variance conditional on $\mathcal{C}$. Let $\beta^{(c)}(\cdot)$ and $\lambda_i^{(c)}(\cdot)$ denote the cth order derivative of $\beta(\cdot)$ and $\lambda_i(\cdot)$, respectively.

We note that the factors and factor loadings that appear in our assumptions below denote the true values F_t^0 and λ_{it}^0. But for notational simplicity, we suppress the superscript 0 hereafter unless confusion may arise. We make the following assumptions.

Assumption A.1 (i) $E(\varepsilon_{it}|X_{it}, F_t) = E_{\mathcal{C}}(\varepsilon_{it}) = 0$, and $\max_{i,t}\{E\|\varepsilon_{it}\|^{8+4\eta} + E\|F_t\|^{8+4\eta}\} \le C$ for some $\eta > 0$.

(ii) $T^{-1}\sum_{t=1}^{T}F_tF_t' = \Sigma_F + O_P(T^{-1/2})$ with $E(F_tF_t') = \Sigma_F > 0$, and $\max_t\sum_{s=1}^{T}|\mathrm{Cov}(F_{t,m}F_{t,n}, F_{s,m}F_{s,n})| \le C$ for $m, n \in [R]$, where $F_{t,m}$ denotes the mth element of F_t.

(iii) $\max_t\sum_{s=1}^{T}\|\gamma(s,t)\| + \max_s\sum_{t=1}^{T}\|\gamma(s,t)\| \le C$ for $\gamma = \gamma_N$, $\gamma_{N,F}$, and $\gamma_{N,FF}$.

(iv) $\max_{s,t}E|N^{1/2}\zeta_{st}|^4 + \max_{r,t}E\|N^{-1/2}\Lambda_r'\varepsilon_t\|^4 \le C$.

(v) $\varpi_{NT,1}(r) = O_P(1)$ and $\max_t E\|\varpi_{NT,2}(r,t)\|^2 \le C$ for each r.

(vi) $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij,t}$, $|\sigma_{ij,t}| \le \bar{\sigma}_{ij}$ for all t, such that $N^{-1}\sum_{i,j=1}^{N}\bar{\sigma}_{ij} \le C$.

(vii) $\max_{t\neq r} E\left\|N^{-1/2}F_t\varepsilon_t'\varepsilon_r F_r'\right\|^4 + \frac{1}{NT}\sum_{i,j}\sum_{s=1}^T |\tau_{ij,s}| \le C$.

Assumption A.2 (i) There exists a constant $C > 0$ such that $\rho_{\min}^{(r)} \ge C$ for $r \in [T]$.

(ii) Both $E(X_t)$ and $E(X_tF_t')$ are time-invariant. $\max_{i,t} E\|X_{it}\|^{8+4\eta} \le C$.

(iii) $\max_{i,j,t}\sum_{s=1}^T |\mathrm{Cov}(\varepsilon_{it}\varepsilon_{jt}, X_{is}'X_{js})| \le C$.

(iv) $\frac{h}{N^{1/2}T}\sum_{i=1}^N\sum_{t,s} k_{h,tr}^* k_{h,sr}^* F_{t,l}F_{s,q}\left[\tilde{X}_{is}\varepsilon_{it} - E_{\mathcal{C}}(\tilde{X}_{is}\varepsilon_{it})\right] = O_P(1)$ for $l,q \in [R]$ and $r \in [T]$.

Assumption A.3 (i) λ_{it} are nonrandom such that $\max_{i,t}\|\lambda_{it}\| \le \bar{c}_\lambda < \infty$ and $N^{-1}\Lambda_r'\Lambda_r = \Sigma_{\Lambda_r} + O(N^{-1/2})$ for some $R\times R$ positive definite matrix Σ_{Λ_r} and for all r.

(ii) For each r, the eigenvalues of the $R\times R$ matrix $\Sigma_F\Sigma_{\Lambda_r}$ are distinct.

(ii) $\lambda_i(\cdot)$ is third-order continuously differentiable with $\max_{i,t}||\lambda_i^{(c)}(t/T)|| \le \bar{c}_\lambda < \infty$ for $c \in [3]$.

(iii) $\beta(\cdot)$ is third-order continuously differentiable with $\max_t\left\|\beta^{(c)}(t/T)\right\| \le \bar{c}_\beta < \infty$ for $c \in [3]$.

Assumption A.4 (i) The kernel function $K:\mathbb{R}\to\mathbb{R}^+$ is a symmetric continuously differentiable probability density function (PDF) with compact support $[-1,1]$.

(ii) As $(N,T)\to\infty$, $Th^7 \to 0$, $Nh^6 \to 0$, $Nh/T \to 0$, $NTh^9 \to 0$, $Nh^2 \to \infty$, $Th/N^{1/2} \to \infty$, $Th^2/\ln T \to \infty$, and $N^3T^{-2}h^{-1}(\ln T)^{-2} \to \infty$.

Assumption A.5 As $(N,T)\to\infty$, $\frac{\sqrt{h}}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \mathfrak{X}_{it}^{(r)}\varepsilon_{it}^{(r)} \xrightarrow{d} N(0,\Omega_r)$, where $\Omega_r = \lim_{(T,N)\to\infty}\mathrm{Var}$ $(\frac{\sqrt{h}}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \mathfrak{X}_{it}^{(r)}\varepsilon_{it}^{(r)})$.

Assumption A.1 mainly imposes some restrictions on ε_{it} and F_t. It is comparable with the corresponding conditions in Assumption A.1 of Su and Wang (2017). The moment condition in Assumption A.1(i) is not needed for some of the theorems below but will be needed for the asymptotic results in Section 4. Assumption A.1(ii) assumes that $E(F_tF_t')$ is homogenous over t. The same assumption has been made in the literature on TV factor models; see, e.g., Stock and Watson (2002), Breitung and Eickmeier (2011), Han and Inoue (2015), and Su and Wang (2017, 2020b), among others. As remarked by Su and Wang (2017), this condition can be regarded as an identification condition for the TV factor loadings. The other conditions in Assumption A.1 require the cross-sectional dependence among $\{\varepsilon_{it}\}$ or serial dependence among $\{(F_t,\varepsilon_{it})\}$ to be weak.

Assumption A.2(i) is an identification condition for the regression coefficients. It is similar to Assumption A in Bai (2009) and rules out the low-rank (e.g., time-invariant) regressors in X_{it}. Assumption A.2(ii) imposes some moment conditions on $\{X_{it}\}$ and $\{X_{it}F_t'\}$. Note that we assume $E(X_{it})$ and $E(X_{it}F_t')$ to be time-invariant, which rules out TV dynamic panels. Assumptions A.2(iii)-(iv) require weak cross-sectional and serial dependence among $\{(X_{it},\varepsilon_{it})\}$.

Assumption A.3 imposes some restrictions on the TV coefficients and factor loadings. A.3(i) assumes that the factor loadings are nonrandom and uniformly bounded, as in Bai (2003) and Breitung and Eickmeier (2011). A.3(ii) is needed for the identification of the eigenvectors. A.3(iii) and (iv) require that $\beta(\cdot)$ and $\lambda_i(\cdot)$ are third-order continuously differentiable, which facilities the Taylor expansion up to the third order. Assumption A.4 imposes regularity conditions on the kernel function and bandwidth. Assumption A.5 is similar to Assumption E in Bai (2009). This assumption is used to establish the asymptotic normality of our LLS estimators and can be verified under some primitive conditions.

It is worth mentioning that we do not assume independence between $\{\varepsilon_{it}\}$ and $\{(X_{it}, F_t)\}$ (see Assumption D in Bai (2009)). This will allow for conditional heteroskedasticity of the type considered in Su and Chen (2013). In particular, the conditional variance of ε_{it} given (X_{it}, F_t) can be a function of both X_{it} and F_t. Such a generality will complicate our asymptotic analysis in various places by using conditional arguments instead of unconditional arguments.

3.2 Asymptotic properties of $\hat{\beta}_t$, $\hat{F}_t$, and $\hat{\lambda}_{it}$

Under the above regularity conditions, we now establish the asymptotic distributions of the estimators of the TV coefficients, common factors, and TV factor loadings that are obtained via our LLS and local PCA methods. As is well known, latent common factors and factor loadings are not separately identifiable. However, they can be identified up to an invertible $R \times R$ matrix $H^{(r)}$, where $H^{(r)} = (N^{-1}\Lambda_r'\Lambda_r)(T^{-1}F^{(r)\prime}\hat{F}^{(r)})\hat{V}_{NT}^{(r)-1}$ and $\hat{V}_{NT}^{(r)}$ is as defined below (2.9). That is, $\hat{F}^{(r)}$ is a consistent estimator of $H^{(r)}F^{(r)}$, and $\hat{\lambda}_{ir}$ is a consistent estimator of $H^{(r)-1}\lambda_{ir}$.

Before establishing the asymptotic distribution of these estimators, we first show that the estimators $\hat{\beta}_t$ and $\hat{F}^{(r)}$ are consistent.

Proposition 3.1 *(Consistency of $\hat{\beta}_t$ and $\hat{F}^{(r)}$) Suppose that Assumptions A.1–A.4 hold. Then as $(N, T) \to \infty$,*

(i) the estimator $\hat{\beta}_t$ is consistent such that $\hat{\beta}_t - \beta_t = O_P(T^{-1/4} + N^{-1/8} + h^{1/2})$;

(ii) the matrix $T^{-1}F^{(r)\prime}\hat{F}^{(r)}$ is invertible and $\left\| P_{\hat{F}^{(r)}} - P_{F^{(r)}} \right\| = O_P(T^{-1/4} + N^{-1/8} + h^{1/2})$.

In the above proposition, both β_t and $F^{(r)}$ denote the true values despite the fact that we suppress their superscript 0. Proposition 3.1(i) establishes the consistency of the LLS estimator $\hat{\beta}_t$. However, we can not deduce that $\hat{F}^{(r)}$ is consistent for $F^{(r)}H^{(r)}$ at this moment. This is because $F^{(r)}$ is a $T \times R$ matrix, where its dimension grows to infinity as $T \to \infty$. However, Proposition 3.1(ii) shows that the spaces spanned by $\hat{F}^{(r)}$ and $F^{(r)}H^{(r)}$ are asymptotically the same.

To proceed, we add some notations. Let V_r denote the diagonal matrix consisting of the eigenvalues of $\Sigma_{\Lambda_r}^{1/2}\Sigma_F\Sigma_{\Lambda_r}^{1/2}$ in descending order with Υ_r being the corresponding (normalized) eigenvector matrix ($\Upsilon_r'\Upsilon_r = \mathbb{I}_R$). Let $Q_r = V_r^{1/2}\Upsilon_r^{-1}\Sigma_{\Lambda_r}^{-1/2}$. We use $H_0^{(r)}$ to denote the probability limit of $H^{(r)}$. Let $A_{1,tr} = X_t\beta_r^{(1)} + \Lambda_r^{(1)}F_t$ and $A_{2,tr} = \frac{1}{2}[X_t\beta_r^{(2)} + \Lambda_r^{(2)}F_t]$, where $X_t = (X_{1t}, ..., X_{Nt})'$, $\Lambda_r^{(l)} = (\lambda_{1r}^{(l)}, ..., \lambda_{Nr}^{(l)})'$, $\beta_r^{(l)}$ denotes the lth order derivative of $\beta(\cdot)$ evaluated at r/T, and $\lambda_{ir}^{(l)}$ denotes the lth order derivative of $\lambda_i(\cdot)$ evaluated at r/T for $l = 1, 2, 3$. Denote

$$
\begin{aligned}
C_{\ell t}^{(r)} &= \hat{V}_{NT}^{(r)-1}H^{(r)\prime}\Sigma_F\left(\Lambda_r'A_{\ell,tr}/N\right) \text{ for } \ell = 1, 2, \\
C_{3t}^{(r)} &= N^{-1}\hat{V}_{NT}^{(r)-1}H^{(r)\prime}E(F_tA_{2,tr}')\Lambda_rF_t, \text{ and} \\
C_{4t}^{(r)} &= \hat{V}_{NT}^{(r)-1}\frac{1}{TN}\sum_s(\frac{s-r}{Th})^2k_{h,sr}^*E[\bar{C}_{1s}^{(r)}A_{1,sr}']\Lambda_rF_t/\kappa_2, \qquad (3.1)
\end{aligned}
$$

where $\kappa_2 = \int u^2 K(u) du$. Let $\bar{C}_{1t}^{(r)} = V^{(r)-1} H_0^{(r)\prime} \Sigma_F \left(\Lambda_r' A_{1,tr} / N \right)$. Define $\bar{C}_{2t}^{(r)}$, $\bar{C}_{3t}^{(r)}$ and $\bar{C}_{4t}^{(r)}$ analogously to $C_{2t}^{(r)}$, $C_{3t}^{(r)}$, and $C_{4t}^{(r)}$ with $\hat{V}_{NT}^{(r)-1}$ and $H^{(r)}$ replaced by their probability limits $V^{(r)-1}$ and $H_0^{(r)}$. Let $A_{2,itr} = \frac{1}{2}[X_{it}' \beta_r^{(2)} + \lambda_{ir}^{(2)\prime} F_t]$, the ith element in $A_{2,tr}$. Let $V_i^{(r)} \equiv N^{-1} \sum_{k=1}^N a_{ik}^{(r)} X_k^{(r)} \equiv (V_{i1}^{(r)}, ..., V_{iT}^{(r)})'$, where $V_{it}^{(r)} = N^{-1} \sum_{k=1}^N a_{ik}^{(r)} X_{kt}^{(r)} \equiv k_{h,tr}^{*1/2} V_{it,r}$ and $V_{it,r} = N^{-1} \sum_{k=1}^N a_{ik}^{(r)} X_{kt}$.

Given the consistency result in Proposition 3.1, we can further establish the asymptotic normality of the coefficient estimator.

Theorem 3.2 *(Asymptotic normality of $\hat{\beta}_t$) Suppose Assumptions A.1–A.5 hold. As $(N,T) \to \infty$, we have for each $r \in [\lfloor Th \rfloor, T - \lfloor Th \rfloor]$,*

$$\sqrt{NTh} \left(\hat{\beta}_r - \beta_r - [D^{(r)}(F^{(r)})]^{-1} [B_{1\beta}^{(r)} + \frac{1}{Th} B_{2\beta}^{(r)} + \frac{1}{N} B_{3\beta}^{(r)} + \frac{1}{Th} B_{4\beta}^{(r)}] \right) \xrightarrow{d} N \left(0, \; D_0^{(r)-1} \Omega_r D_0^{(r)-1} \right),$$

where

$$\begin{aligned}
D_0^{(r)} &= \underset{(N,T)\to\infty}{plim} D^{(r)}(F^{(r)}), \\
B_{1\beta}^{(r)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T k_{h,tr}^* E\left[(X_{it} - V_{it,r}) A_{2,itr} \right] \left(\frac{t-r}{T}\right)^2 \\
&\quad + \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T k_{h,tr}^* k_{h,sr}^* (X_{it} - V_{it,r}) F_t' H^{(r)} [H^{(r)\prime} F_s A_{2,isr} + C_{1s}^{(r)} A_{1,isr}] \left(\frac{s-r}{T}\right)^2, \\
B_{2\beta}^{(r)} &= (B_{2\beta,1}^{(r)}, \cdots, B_{2\beta,P}^{(r)})' \; \textit{with} \; B_{2\beta,p}^{(r)} = -\frac{h}{N} tr\left(P_{F^{(r)}} E_{\mathcal{C}} [\mathbf{X}_p^{(r)} \varepsilon^{(r)\prime}] \right) \; \textit{for} \; p \in [P], \\
B_{3\beta}^{(r)} &= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{T} (X_i^{(r)} - V_i^{(r)})' F^{(r)} \left(\frac{1}{T} F^{(r)\prime} F^{(r)}\right)^{-1} \left(\frac{1}{N} \Lambda_r' \Lambda_r\right)^{-1} \lambda_{jr} \left(\frac{1}{T} \sum_{t=1}^T k_{h,tr}^* E_{\mathcal{C}}(\varepsilon_{it} \varepsilon_{jt}) \right), \; \textit{and} \\
B_{4\beta}^{(r)} &= -\frac{h}{N^2 T} \sum_{i=1}^N \sum_{j=1}^N X_i^{(r)\prime} M_{F^{(r)}} \varepsilon_j^{(r)} \varepsilon_j^{(r)\prime} F^{(r)} \left(\frac{1}{T} F^{(r)\prime} F^{(r)}\right)^{-1} \left(\frac{1}{N} \Lambda_r' \Lambda_r\right)^{-1} \lambda_{ir}.
\end{aligned}$$

Theorem 3.2 establishes the asymptotic normality of the LLS estimator $\hat{\beta}_r$ for each interior point r. In fact, it is a local version of Theorem 3 in Bai (2009) and Theorem 4.3 in Moon and Weidner (2017). We note that the bias term contains four components. The first component, $[D^{(r)}(F^{(r)})]^{-1} B_{1\beta}^{(r)}$, is introduced by the local approximation error and is $O_P(h^2)$ as in standard kernel regressions. For this reason, we can refer to this bias term as the *nonparametric kernel bias* hereafter. The second component $\frac{1}{Th}[D^{(r)}(F^{(r)})]^{-1} B_{2\beta}^{(r)}$ is due to the non-contemporaneous correlation of the error term and regressors along the time direction. This bias term is a local version of that in Moon and Weidner (2017), which generalizes Nickell's (1981) bias in panel data regressions with predetermined but not strictly exogenous regressors X_{it}. If we only consider strictly exogenous regressors such that $E_{\mathcal{C}}(X_{is} \varepsilon_{it}) = 0$ for all s and t, then this term will disappear. The third component $\frac{1}{N}[D^{(r)}(F^{(r)})]^{-1} B_{3\beta}^{(r)}$ is due to the cross-sectional correlation and heteroskedasticity over i. When the cross-sectional correlation and heteroskedasticity are absent conditional on $\mathcal{C}$, we have $E_{\mathcal{C}}(\varepsilon_{it}\varepsilon_{jt}) = \sigma_t^2 \mathbf{1}\{i=j\}$ with $\mathbf{1}\{\cdot\}$ being the usual indicator function. Then

$$B_{3\beta}^{(r)} = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\left(X_i^{(r)} - N^{-1}\sum_{k=1}^{N}a_{ik}^{(r)}X_k^{(r)}\right)' F^{(r)}(\frac{1}{T}F^{(r)\prime}F^{(r)})^{-1}(\frac{1}{N}\Lambda_r'\Lambda_r)^{-1}\lambda_{ir}\bar{\sigma}^{(r)2} = 0,$$

where $\bar{\sigma}^{(r)2} = \frac{1}{T}\sum_{t=1}^{T}k_{h,tr}^{*}\sigma_t^2$ and the last equality follows from the fact that $N^{-1}\sum_{i=1}^{N}a_{ik}^{(r)}\lambda_{ir} = \lambda_{kr}$. That is, $B_{3\beta}^{(r)}$ would disappear if the error terms are cross-sectionally uncorrelated and homoskedastic given $\mathcal{C}$. The fourth component is due to the time series heteroskedasticity and serial correlation. Let $\varepsilon_j = (\varepsilon_{j1}, ..., \varepsilon_{jT})'$. Unlike Bai (2009) and Moon and Weidner (2017), even if $E_{\mathcal{C}}(\varepsilon_j\varepsilon_j') = \sigma_j^2\mathbb{I}_T$, this bias term will not disappear because of the appearance of kernel weights $\{k_{h,tr}^{*1/2}\}$ in the definition of $\varepsilon_j^{(r)}$. In the special case where the "weighted error term" $\varepsilon_{jt}^{(r)} = k_{h,tr}^{*1/2}\varepsilon_{jt}$ is serially uncorrelated and homoskedastic over t conditional on $\mathcal{C}$, we can show that $B_{4\beta}^{(r)}$ is asymptotically negligible by noting that $M_{F^{(r)}}E_{\mathcal{C}}[\varepsilon_j^{(r)}\varepsilon_j^{(r)\prime}]F^{(r)}$ is now proportional to $M_{F^{(r)}}\mathbb{I}_T F^{(r)} = 0$.

It is easy to see that the contribution of $B_{1\beta}^{(r)}$, $\frac{1}{Th}B_{2\beta}^{(r)}$, $\frac{1}{N}B_{3\beta}^{(r)}$ and $\frac{1}{Th}B_{4\beta}^{(r)}$ to the asymptotic bias of $\hat{\beta}_r$ are respectively h^2, $\frac{1}{Th}$, $\frac{1}{N}$, and $\frac{1}{Th}$ in probability order. Since the last three bias terms can be corrected as in the standard parametric panel data models with time-invariant factor loadings, it is desirable to choose $h \propto (NT)^{-1/5}$ to achieve a balance between the asymptotic variance of $\hat{\beta}_r$ and the squared nonparametric kernel bias. With such a choice of bandwidth, $\hat{\beta}_r$, after correcting the other three bias terms, will achieve the desirable pointwise $(NTh)^{-1/2}$-rate of convergence. This rate is faster than the convergence rate of $\hat{F}_t$ and $\hat{\lambda}_{it}$ in Su and Wang's (2017) TV factor models. As a result, the estimation of β_r has asymptotically negligible impact on the limiting distributions of $\hat{F}_t$ and $\hat{\lambda}_{it}$.

To study the asymptotic distributions of the estimated TV factor loadings $\hat{\lambda}_{it}$, the estimated common factors $\hat{F}_t$, and the estimated common component $\hat{C}_{it} = \hat{\lambda}_{it}'\hat{F}_t$, we add a new assumption.

Assumption A.6 (i) $N^{-1/2}\Lambda_r'\varepsilon_t \xrightarrow{d} N(0, \Gamma_{rt})$ for each r, t, where $\Gamma_{rt} = \lim_{N\to\infty} N^{-1}\sum_{i,j}\lambda_{ir}\lambda_{jr}'E(\varepsilon_{it}\varepsilon_{jt})$.

(ii) $\frac{\sqrt{h}}{\sqrt{T}}\sum_{s=1}^{T}k_{h,sr}F_s\varepsilon_{is} \xrightarrow{d} N(0, \Omega_{i,r})$, where

$$\Omega_{i,r} = \lim_{T\to\infty}\left[\frac{h}{T}\sum_{s=1}^{T}k_{h,sr}^2 E\left(F_sF_s'\varepsilon_{is}^2\right) + \frac{2h}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}k_{h,sr}k_{h,tr}E\left(F_sF_t'\varepsilon_{is}\varepsilon_{it}\right)\right].$$

This assumption is the same as Assumption A.2 in Su and Wang (2017). Assumption A.6(i) extends Assumption F.3 in Bai (2003) to allow for the TV factor loadings and Assumption A.6(ii) is the kernel-weighted version of Assumption F in Bai (2003). Both parts are used to establish the asymptotic normality of our estimated common factors and factor loadings, and can be verified under some primitive conditions.

The following theorem shows the asymptotic distributions of $\hat{F}_t^{(r)}$, $\hat{\lambda}_{it}$, $\hat{F}_t$, and $\hat{C}_{it}$.

Theorem 3.3 *(Asymptotic normality of $\hat{F}_t^{(r)}$, $\hat{\lambda}_{it}$, $\hat{F}_t, \hat{C}_{it}$)*

(i) Suppose that Assumptions A.1–A.4 and A.6(i) hold. Then, for each $t \in [T]$ and $r \in [T]$ such that $|r-t| \le Th$, we have

$$K_r^*\left(\frac{r-t}{Th}\right)^{-1/2}\sqrt{Nh}\left[\hat{F}_t^{(r)} - H^{(r)\prime}F_t^{(r)} - B_t^{(r)}\right] \xrightarrow{d} N(0, V_r^{-1}Q_r\Gamma_{rt}Q_r'V_r^{-1}),$$

where $B_t^{(r)} = k_{h,tr}^{*1/2}[C_{1t}^{(r)}\frac{t-r}{T} + C_{2t}^{(r)}(\frac{t-r}{T})^2 + C_{3t}^{(r)}h^2\kappa_2 + C_{4t}^{(r)}h^2\kappa_2]$.

(ii) Suppose that Assumptions A.1–A.4 and A.6(ii) hold. Then for $r \in [\lfloor Th\rfloor, T-\lfloor Th\rfloor]$ we have

$$\sqrt{Th}\left[\hat{\lambda}_{ir} - H^{(r)-1}\lambda_{ir} - B_\Lambda(i,r)\right] \xrightarrow{d} N\left(0, \left(Q_r'\right)^{-1}\Omega_{i,r}Q_r^{-1}\right) \ \forall i \in [N] \text{ and } r \in [T],$$

where $B_\Lambda(i,r) = [E(\bar{C}_{1t}^{(r)}A_{1,itr}) + H^{(r)\prime}E\left(F_tA_{2,itr}\right)]\kappa_2h^2 - \left\{H^{(r)\prime}E[F_t(\bar{C}_{2t}^{(r)} + \bar{C}_{3t}^{(r)} + \bar{C}_{4t}^{(r)})'] + E(\bar{C}_{1t}^{(r)}\bar{C}_{1t}^{(r)\prime})\right\}$ $\times H^{(r)-1}\lambda_{ir}\kappa_2h^2$.

(iii) Suppose that Assumptions A.1–A.4 and A.6(i) hold. If in addition, we assume that $Nh^4 = o(1)$, then

$$\sqrt{N}\left[\hat{F}_t - H^{(t)\prime}F_t\right] \xrightarrow{d} N\left(0, \left(\Sigma_{\Lambda_t}^{-1}Q_t^{-1}\right)'\Gamma_{tt}\Sigma_{\Lambda_t}^{-1}Q_t^{-1}\right) \ \forall t \in [T].$$

(iv) Suppose that Assumptions A.1–A.4 and A.6 hold and $Nh^4 = o(1)$. Then

$$\left(\frac{1}{N}V_{1it} + \frac{1}{Th}V_{2it}\right)^{-1/2}\left[\hat{C}_{it} - C_{it}^0 - B_C(i,t)\right] \xrightarrow{d} N(0,1) \ \forall i \in [N] \text{ and } t \in [T],$$

where $V_{1it} = \lambda_{it}'\Sigma_{\Lambda_t}^{-1}\Gamma_{tt}\Sigma_{\Lambda_t}^{-1}\lambda_{it}$, $V_{2it} = F_t'\Sigma_F^{-1}\Omega_{i,t}\Sigma_F^{-1}F_t$, $B_C(i,t) = \left[\lambda_{it}'Q_t'B_F(t) + B_\Lambda(i,t)'(Q_t^{(-1)})'\right]F_t$, $B_F(t) = \left(Q_t\Sigma_{\Lambda_t}Q_t'\right)^{-1}\tilde{B}_F(t)\kappa_2h^2$ *and $\tilde{B}_F(t)$ is defined in the proof of (iii).*

The results in Theorem 3.3(i)–(iv) are comparable with those in Theorems 2.1-2.4 of Su and Wang (2020a), which correct the results in Su and Wang (2017). Theorem 3.3(i) establishes the asymptotic distribution of $\hat{F}_t^{(r)}$. When we treat $H^{(r)\prime}F_t^{(r)}$ as the pseudo-true factor, we note that the bias term $B_t^{(r)}$ consists of four parts. The first and second parts are related to $\frac{t-r}{T}$ and $(\frac{t-r}{T})^2$ that are of respective orders $O_P(h)$ and $O_P(h^2)$ and generated from the third-order Taylor expansion of $\Delta_{it}^{(r)} \equiv k_{h,tr}^{*1/2}\Delta_i(t,r)$, where recall that $\Delta_i(t,r) \equiv X_{it}'(\beta_t - \beta_r) + (\lambda_{it} - \lambda_{ir})'F_t$. In the eigenvalue analysis, there is no summation running over r or t so that terms associated with $\frac{t-r}{T}$ and $(\frac{t-r}{T})^2$ cannot be smoothed out. The third part in $B_t^{(r)}$, viz., $N^{-1}\hat{V}_{NT}^{(r)-1}H^{(r)\prime}E(F_tA_{2,tr}')\Lambda_rF_t$ is derived from the usual local constant estimation of the common factors while the last component in $B_t^{(r)}$ generated from the summation over a term associated with $C_{1t}^{(r)}\frac{t-r}{T}$ appearing in the derivation. Consequently, $B_t^{(r)}$ is $O_P(h)$, which is quite large but does not cause much trouble in the asymptotic analyses of $\hat{\lambda}_{ir}$ and $\hat{F}_t$ below.

Theorem 3.3(ii) establishes the asymptotic distribution of $\hat{\lambda}_{ir}$ for $r \in [\lfloor Th\rfloor, T-\lfloor Th\rfloor]$, and it is quite similar to that in Su and Wang (2017). When we treat $H^{(r)-1}\lambda_{ir}$ as the pseudo-true factor loading, Theorem 3.3(ii) indicates that the bias of $\hat{\lambda}_{ir}$ contains two terms associated with λ_{ir} and β_r and their first and second order derivatives. The first term is associated with the conventional nonparametric kernel estimation, and the second term is introduced by the bias terms in $\hat{F}_t^{(r)}$. If one uses the optimal bandwidth h in estimating β_r such that $NTh^5 \asymp 1$, it is easy to see that $Th^5 = o(1)$ and the asymptotic bias of $\hat{\lambda}_{ir}$ is asymptotically negligible.

Theorem 3.3(iii) reports the asymptotic distribution of the second stage estimator of the factor under the additional condition $Nh^4 = o(1)$. This condition ensures that the nonparametric kernel bias of $\hat{\beta}_r$, which is $O_P(h^2)$, and thus $o_P(N^{-1/2})$. Therefore it is asymptotically negligible in the asymptotic distribution of $\hat{F}_t$. Without this condition, the asymptotic kernel bias of $\hat{\beta}_r$ will be carried over to yield an asymptotically non-negligible bias for $\hat{F}_t$. Note that the result in Theorem 3.3(iii) is different from that in Theorem 2.3 of Su and Wang (2020a) because the latter paper does not impose the condition $Nh^4 = o(1)$. Without this condition, Su and Wang (2020a) demonstrate that $\hat{F}_t$ is asymptotically unbiased for $\tilde{H}^{(t)\prime} F_t$ in the pure TV factor model where $\tilde{H}^{(t)} = H^{(t)} + B_F(t)'$ with $B_F(t) = (Q_t \Sigma_{\Lambda_t} Q_t')^{-1} \tilde{B}_F(t) \kappa_2 h^2$ and $\tilde{B}_F(t)$ is defined in the proof of the theorem. Since the correction for $H^{(t)}$ is also $O_P(h^2)$ and thus $o_P(N^{-1/2})$ provided $Nh^4 = o(1)$, the centering around either $\tilde{H}^{(t)\prime} F_t$ or $H^{(t)\prime} F_t$ yields the same asymptotic distribution.

In general, Theorem 3.3(iv) shows that the estimated common component, $\hat{C}_{it}$, also exhibits bias that is $O_P(h^2)$, carried over from the estimates $\hat{\lambda}_{ir}$ and $\hat{F}_t$. If one imposes the additional condition $NTh^5 = O(1)$, then both Nh^4 and Th^5 are $o(1)$ under Assumption A.4(ii) so that the bias term $B_C(i,t)$ can be removed from the result in Theorem 3.3(iv).

Since we focus on the inference of the regression coefficient β_r, we will impose the condition $NTh^5 = O(1)$ in the subsequent study. This condition, along with the condition that $Th^2/\ln T \to \infty$ in Assumption A.4(ii), automatically ensures $Nh^4 = o(1)$. In practice, if one is also interested in the inferences on the factors and factor loadings, a multiple-step approach would be recommended. In the first step, one chooses $h \propto (NT)^{-1/5}$ to obtain estimates $\{\hat{\beta}_r\}$ as in Section 2. In the second step, one considers fitting a TV factor model to the residuals $\{\hat{W}_{it}\}$, where $\hat{W}_{it} = Y_{it} - X_{it}'\hat{\beta}_r$ by specifying another bandwidth, say, $\tilde{h}$, which is proportional to $T^{-1/5}$, the optimal rate for the estimation of the factor loadings.

3.3 Bias-corrected estimator and uniform convergence

By Theorem 3.2, the estimator $\hat{\beta}_t$ has four bias terms. The first term has been referred to as a kernel bias. There are various ways to correct such a bias term, including the simplest but inefficient method of undersmoothing, the plug-in method which requires explicit estimation of certain first and second-order derivative objects in our context, and some bootstrap methods. For brevity, we do not address the correction of this bias term here. Instead, we focus on the correction of the other three bias terms.

We will define a bias-corrected estimator $\hat{\beta}_t^{bc}$ based on the estimators $\hat{\beta}_t$, $\hat{F}_t$ and $\hat{\lambda}_{it}$. For the common factor F_t, we use the second stage estimator $\hat{F}_t$, and we estimate $F_s^{(t)} (= k_{h,st}^{*1/2} F_s^{(t)})$ by $\check{F}_s^{(t)} = k_{h,st}^{*1/2} \hat{F}_s$. Let $\check{F}^{(t)} = (\check{F}_1^{(t)}, \cdots, \check{F}_T^{(t)})'$.

First, we define the estimated $T \times N \times P$ tensor $\hat{\mathbf{Z}}^{(t)}$, whose pth sheet is given by $\hat{\mathbf{Z}}_p^{(t)} =$

$M_{\check{F}^{(t)}}\mathbf{X}_p^{(t)}M_{\hat{\Lambda}_t}$ for $p \in [P]$, where $M_{\check{F}^{(t)}} = \mathbb{I}_T - \check{F}^{(t)}(\frac{1}{T}\check{F}^{(t)\prime}\check{F}^{(t)})^{-1}\check{F}^{(t)\prime}/T$ and $M_{\hat{\Lambda}_t} = \mathbb{I}_T - \hat{\Lambda}_t(\hat{\Lambda}_t'\hat{\Lambda}_t)^{-1}\hat{\Lambda}_t'$. Note that $\frac{1}{T}\check{F}^{(t)\prime}\check{F}^{(t)}$ may not be exactly an identity matrix. As before, let $\hat{Z}_i^{(t)}$ denote the $T \times P$ submatrix of $\hat{\mathbf{Z}}^{(t)}$ by fixing the cross-section unit i :

$$\hat{Z}_i^{(t)} = M_{\check{F}^{(t)}}[X_i^{(t)} - \frac{1}{N}\sum_{j=1}^{N} X_j^{(t)}\hat{a}_{ij}^{(t)}] = M_{\check{F}^{(t)}}[X_i^{(t)} - \hat{V}_i^{(t)}],$$

where $\hat{a}_{ij}^{(t)} = \hat{\lambda}_{it}'(N^{-1}\hat{\Lambda}_t'\hat{\Lambda}_t)^{-1}\hat{\lambda}_{jt}$ and $\hat{V}_i^{(t)} = N^{-1}\sum_{j=1}^{N}\hat{a}_{ij}^{(t)}X_j^{(t)}$. Define

$$\hat{D}^{(t)}(\hat{F}^{(t)}) = \frac{1}{NT}\sum_{i=1}^{N}\hat{Z}_i^{(t)\prime}\hat{Z}_i^{(t)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\hat{Z}_{is}^{(t)}\hat{Z}_{is}^{(t)\prime},$$

where $\hat{Z}_{is}^{(t)\prime}$ denotes the sth row of $\hat{Z}_i^{(t)}$.

Note that $E_{\mathcal{C}}(X_{is}\varepsilon_{it}) = 0$ if $s \leq t$ and it can be nonzero otherwise under our conditions. Let $\Gamma(\tau) = \mathbf{1}\{0 \leq \tau \leq 1\}$. Define

$$\hat{B}_{2\beta}^{(t)} = -\frac{h}{N}\sum_{i=1}^{N}\sum_{r=1}^{T-1}\sum_{s=r+1}^{T}\Gamma(\frac{s-r}{M})\left[P_{\hat{F}^{(t)}}\right]_{r,s}\hat{\varepsilon}_{ir}^{(t)}X_{i,s}^{(t)},$$

$$\hat{B}_{3\beta}^{(t)} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{T}[X_i^{(t)} - \hat{V}_i^{(t)}]'\hat{F}^{(t)}(\frac{1}{N}\hat{\Lambda}_t'\hat{\Lambda}_t)^{-1}\hat{\lambda}_{jt}\left[\frac{1}{T}\sum_{s=1}^{T}k_{h,st}^{*}\hat{\varepsilon}_{is}\hat{\varepsilon}_{js}\right],$$

$$\hat{B}_{4\beta}^{(t)} = -\frac{h}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}X_i^{(t)\prime}M_{\hat{F}^{(t)}}\hat{\varepsilon}_j^{(t)}\hat{\varepsilon}_j^{(t)\prime}\hat{F}^{(t)}(\frac{1}{N}\hat{\Lambda}_t'\hat{\Lambda}_t)^{-1}\hat{\lambda}_{it},$$

which are estimators of $B_{2\beta}^{(t)}$, $B_{3\beta}^{(t)}$, and $B_{4\beta}^{(t)}$, respectively. Here M is a truncation parameter such that $M \equiv M(T) \to \infty$ as $T \to \infty$.

Then, we define the bias-corrected estimator as follows:

$$\hat{\beta}_t^{bc} = \hat{\beta}_t - \hat{D}^{(t)}(\hat{F}^{(t)})^{-1}\left[\frac{1}{Th}\hat{B}_{2\beta}^{(t)} + \frac{1}{N}\hat{B}_{3\beta}^{(t)} + \frac{1}{Th}\hat{B}_{4\beta}^{(t)}\right].$$

To study the uniform convergence of $\hat{\beta}_t^{bc}$, $\hat{\lambda}_{it}$, $\hat{F}_t$ and $\hat{C}_{it}$, we add the following assumptions.

Assumption A.7 (i) $\|\varepsilon\|_{\text{sp}} = O_P\left(N^{1/2} + T^{1/2}\right)$ and $\max_t\left|\frac{1}{N}\sum_{i=1}^{N}\left[\varepsilon_{it}^2 - E(\varepsilon_{it}^2)\right]\right| = O_P(N^{-1/2}(\ln T)^{1/2})$.

(ii) $\max_r\left|\frac{1}{N}\Lambda_r'\Lambda_r' - \Sigma_{\Lambda_r}\right| = o(1)$, and the eigenvalues of Σ_{Λ_r} are bounded below from 0 and above from infinity uniformly in r.

(iii) $\max_{i,r}\left\|\frac{1}{T}\sum_{t=1}^{T}k_{h,tr}^{*}F_t\varepsilon_{it}\right\| + \max_r\left\|\frac{1}{T}\sum_{t=1}^{T}k_{h,tr}^{*}(F_tF_t' - \Sigma_F)\right\| + \max_r\left|\frac{1}{T}\sum_{t=1}^{T}k_{h,tr}^{*}[\|F_t\| - E\|F_t\|]\right|$ $= O_P((Th)^{-1/2}(\ln T)^{1/2})$.

(iv) $\max_{s,t}\left\|N^{-1}\Lambda_s'\varepsilon_t\right\| + \max_s\left\|N^{-1}\Lambda_s'\varepsilon_sF_s\right\| = O_P(N^{-1/2}(\ln T)^{1/2})$.

(v) $\max_r\|\varpi_{NT,1}(r)\| + \max_{r,t}\|\varpi_{NT,2}(r,t)\| = O_P((\ln T)^{1/2})$.

(vi) $\max_r\left\|\frac{1}{NT}\sum_{t=1}^{T}\Lambda_r'\varepsilon_t^{(r)}F_t^{(r)\prime}\right\| = \max_r\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}k_{h,tr}^{*}\lambda_{ir}\varepsilon_{it}F_t'\right\| = O_P\left((NTh)^{-1/2}(\ln T)^{1/2}\right)$.

Assumption A.8 As $(N,T)\to\infty$, $MN^{1/2}(Th)^{-1/2}(N^{-1/2}+(Th)^{-1/2}+h^2)\to 0$ and $M^2(Th)^{-1}\to 0$.

Assumption A.7 imposes some high level conditions. It strengthens Assumption A.4 in Su and Wang (2017). Assumption A.8 imposes conditions on the truncation parameter M.

Theorem 3.4 *(Uniform convergence of $\hat{\beta}_t^{bc}$). Suppose that Assumptions A.1–A.8 hold. Suppose that $NTh^5 = O(1)$. Then as $(N,T)\to\infty$, we have*
(i) $\sqrt{NTh}\left(\hat{\beta}_t^{bc}-\beta_t-[D^{(t)}(F^{(t)})]^{-1}B_{1\beta}^{(t)}\right)\xrightarrow{d} N\left(0,\ D_0^{(t)-1}\Omega_t D_0^{(t)-1}\right)$,
(ii) $\max_t\left\|\hat{\beta}_t^{bc}-\beta_t\right\| = O_P\left((NTh/\ln T)^{-1/2}\right)$.

Theorem 3.4(i) establishes the asymptotic normality for the bias-corrected estimator $\hat{\beta}_t^{bc}$. Since $B_{1\beta}^{(t)} = O\left(h^2\right)$, the MSE-optimal rate of bandwidth h should be proportional to $(NT)^{-1/5}$ as in standard local constant estimation when a second-order kernel is applied. When $h\propto (NT)^{-1/5}$, the remaining asymptotic bias $\hat{\beta}_t^{bc}$ is of order $h^2 = O((NT)^{-2/5}) = o\left((NTh/\ln T)^{-1/2}\right)$. This explains why the role of the asymptotic bias appears to vanish in Theorem 3.4(ii) when the optimal-rate of bandwidth is employed. In the following theorem, we study the uniform convergence rates of $\hat{\lambda}_{it}$, $\hat{F}_t$, and $\hat{C}_{it}$.

Theorem 3.5 *(Uniform convergence of $\hat{\lambda}_{it}$, $\hat{F}_t, \hat{C}_{it}$) Suppose that Assumptions A.1–A.8 hold. Then*
(i) $\max_{i,t}\left\|\hat{\lambda}_{it}-H^{(t)-1}\lambda_{it}\right\| = O_P\left((Th/\ln T)^{-1/2}+h^2\right)$,
(ii) $\max_t\left\|\hat{F}_t-H^{(t)\prime}F_t\right\| = O_P\left((N/\ln T)^{-1/2}+h^2\right)$,
(iii) $\max_{i,t}\left|\hat{C}_{it}-C_{it}\right| = o_P\left((Th/\ln T)^{-1/2}T^{1/8}+h^2T^{1/8}\right)$.
If, in addition, $h = O((NT)^{-1/5})$, then the terms associated with h^2 in (i)-(iii) are asymptotically of smaller order than the other term.

Theorem 3.5 suggests that the uniform convergence rates are the same as those obtained for the pure TV factor model in Su and Wang (2017). This is due to the faster $\sqrt{NTh}$-convergence rate of $\hat{\beta}_t^{bc}$ to $\beta_t+[D^{(t)}(F^{(t)})]^{-1}B_{1\beta}^{(t)}$. When the optimal-rate bandwidth is applied in the estimation of β_t, the effect of $B_{1\beta}^{(t)}$ is asymptotically negligible. Otherwise, its effect is reflected in the terms associated with h^2 in Theorem 3.5(i)-(iii).

3.4 Determination of the number of factors

In the above analysis, we assume that the number of factors, R, is known. In practice, one has to determine R from the data. Following Bai (2009), Moon and Weidner (2015), and Lu and Su (2016), when R is overspecified, the convergence rate of the regression coefficient estimator does not change, but the estimates would be less efficient than in the case of correct specification in finite samples. The same argument holds for our TV cases. In order to choose the appropriate number of common factors, we follow Su and Wang (2017) to introduce an information criterion (IC) to determine the number of common factors. Here we assume that the true value of R, denoted as R_0, is bounded from above by a finite integer $R_{\max}$.

Let $\hat{\beta}_t^{bc}(R)$, $\hat{F}_t(R)$, and $\hat{\lambda}_{it}(R)$ denote the LLS estimators of the coefficients, factors and factor loadings given in Section 2. We define the sum of squared residuals (SSR) as follows:

$$V(R) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[Y_{it} - X_{it}'\hat{\beta}_t^{bc}(R) - \hat{\lambda}_{it}'(R)\hat{F}_t(R)\right]^2. \tag{3.2}$$

As in Bai and Ng (2002), we propose the following BIC-type IC to determine R_0 :

$$IC(R) = \ln V(R) + \rho_{NT}R, \tag{3.3}$$

where ρ_{NT} plays the role of $\ln(NT)/(NT)$ in the case of BIC and $2/(NT)$ in the case of AIC. Let $\hat{R} = \arg\min_{1\le R\le R_{\max}} IC(R)$.

When we prove the consistency of the above IC, we follow Su and Wang (2017) to use the following normalization rule instead of that used in Section 2.2:

$$N^{-1}\Lambda_r'\Lambda_r = \mathbb{I}_R \text{ and } T^{-1}F^{(r)\prime}F^{(r)} \text{ is a diagonal matrix with decreasing diagonal elements.} \tag{3.4}$$

As Bai and Ng (2002) and Su and Wang (2017) remark, $V(R)$ in (3.2) does not depend on which version of normalization is used. However, the normalization in (3.4) will facilitate the proof of Theorem 3.6 below.

We add the following two assumptions.

Assumption A.9. (i) $\max_{s,t}\{E\left\|N^{-1/2}\Lambda_s'\varepsilon_t F_t'\right\|^4 + E\left\|N^{-1/2}\left[F_s\varepsilon_s'\varepsilon_t F_t' - E\left(F_s\varepsilon_s'\varepsilon_t F_t'\right)\right]\right\|^2\} \le C.$

(ii) $\max_r E\left\|\frac{h^{1/2}}{(NT)^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}k_{h,tr}\left[F_t\varepsilon_{it}\varepsilon_{ir}F_r' - E\left(F_t\varepsilon_{it}\varepsilon_{ir}F_r'\right)\right]\right\|^2 \le C.$

Assumption A.10. As $(N,T)\to\infty$, $\rho_{NT}\to 0$ and $\rho_{NT}C_{NT}^2\to\infty$, where $C_{NT} = \min(\sqrt{Th}, \sqrt{N}, h^{-2})$.

Assumption A.9 is new and needed for the proof of Theorem 3.6. The conditions on ρ_{NT} in Assumption A.10 are typical conditions in order to estimate the number of factors consistently. The penalty coefficient ρ_{NT} has to shrink to zero at an appropriate rate to avoid both overfitting and underfitting.

Theorem 3.6 *(Consistency of the IC) Suppose that Assumptions A.1–A.10 hold. Then*

$$P\left(\hat{R} = R_0\right) \to 1 \text{ as } (N,T)\to\infty.$$

Theorem 3.6 shows that the class of information criteria defined by $IC(R)$ in (3.3) can consistently estimate R_0. To implement the information criterion, one needs to choose the penalty coefficient ρ_{NT}. Following the lead of Bai and Ng (2002) and Su and Wang (2017), we suggest setting $\rho_{NT} = \frac{N+Th}{NTh}\ln\left(\frac{NTh}{N+Th}\right)$ or $\rho_{NT} = \frac{N+Th}{NTh}\ln C_{NT}^2$. Intuitively, as the estimators of the regression slopes converge to the true values faster than those of the factors and factor loadings, the estimation of the slope coefficients has asymptotically negligible effect on the determination of the number of factors. Such an intuition is used in the proof of Theorem 3.6.

4 Specification Testing

In this section, we test the specifications of the commonly used panel data models with time-invariant slope coefficients or factor loadings.

4.1 Hypotheses

Given the asymptotic results in the last section, it is natural to consider testing the constancy of the slope coefficients β_t and factor loadings λ_{it}. We first consider the following null hypothesis:

$$\mathbb{H}_0^{(1)} : \beta_t = \beta_0 \text{ for some } \beta_0 \in \mathbb{R}^P \text{ for all } t \text{ and } \lambda_{it} = \lambda_{i0} \text{ for some } \lambda_{i0} \in \mathbb{R}^R \text{ for all } (i,t) .$$

The alternative hypothesis $\mathbb{H}_A^{(1)}$ is the negation of $\mathbb{H}_0^{(1)}$:

$$\mathbb{H}_A^{(1)} : \beta_t \neq \beta_0 \text{ for some } t \text{ or } \lambda_{it} \neq \lambda_{i0} \text{ for some } (i,t) ,$$

where β_0 and λ_{i0} are the unknown slope coefficients and the time-invariant factor loadings, respectively. We allow $\beta_t = \beta(t/T)$ and $\lambda_{it} = \lambda_i(t/T)$ to be piece-wise smooth functions on (0,1] with finite numbers of discontinuities under $\mathbb{H}_A^{(1)}$. Under $\mathbb{H}_0^{(1)}$, both the slope coefficients and the factor loadings are time-invariant. The model degenerates to the traditional time-invariant panel data model with the usual IFEs, which has been studied by Bai (2009), Lu and Su (2016), and Moon and Weidner (2017), among others. Under $\mathbb{H}_A^{(1)}$, either the slope coefficients, or the factor loadings, or both can vary over time so that the traditional PCA method typically fails to yield consistent estimators of the model parameters.

When we reject $\mathbb{H}_0^{(1)}$, it is of further interest to consider testing either the null hypothesis of time-invariant slope coefficients:

$$\mathbb{H}_0^{(2)} : \beta_t = \beta_0 \text{ for some } \beta_0 \in \mathbb{R}^P \text{ for all } t,$$

or the null hypothesis of time-invariant factor loadings:

$$\mathbb{H}_0^{(3)} : \lambda_{it} = \lambda_{i0} \text{ for some } \lambda_{i0} \in \mathbb{R}^R \text{ for all } (i,t) ,$$

where the alternative hypotheses $\mathbb{H}_A^{(2)}$ and $\mathbb{H}_A^{(3)}$ are the negations of $\mathbb{H}_0^{(2)}$ and $\mathbb{H}_0^{(3)}$, respectively. Note that we allow the factor loadings to be TV under $\mathbb{H}_0^{(2)}$ and the slope coefficients to be TV under $\mathbb{H}_0^{(3)}$, so that the null models under $\mathbb{H}_0^{(1)}$, $\mathbb{H}_0^{(2)}$, and $\mathbb{H}_0^{(3)}$ are distinct from each other. Under $\mathbb{H}_0^{(2)}$, we have the time-invariant slope coefficient and TV factor loadings; under $\mathbb{H}_0^{(3)}$, we have the TV panel data models with the usual IFEs. Obviously, the model under $\mathbb{H}_0^{(1)}$ is nested in the model under either $\mathbb{H}_0^{(2)}$ or $\mathbb{H}_0^{(3)}$.

In the absence of regressors, the model in (2.1) becomes a pure factor model and various tests have been proposed to test the null hypothesis $\mathbb{H}_0^{(3)}$. See Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), Yamamoto and Tanaka (2015), Cheng et al. (2016), Su and Wang (2017, 2020b), Fu et al. (2023), among others.

4.2 The test statistics

There are many ways to construct the test statistics for testing $\mathbb{H}_0^{(1)}$, $\mathbb{H}_0^{(2)}$, and $\mathbb{H}_0^{(3)}$. Here we propose some convenient test statistics based on the asymptotic results obtained in the last section.

Under $\mathbb{H}_0^{(1)}$, we can follow Bai (2009) and Moon and Weidner (2017) to estimate the conventional time-invariant panel data models with the standard IFEs to obtain the constrained estimators $\tilde{\beta}_0$, $\tilde{\lambda}_{i0}$ and $\tilde{F}_t$ of β_0, λ_{i0} and F_t^0, respectively. We use $\tilde{\beta}_0^{bc}$ to denote the bias-corrected version of $\tilde{\beta}_0$. One may be tempted to construct a test statistic based on the squared L_2-distance between the unrestricted estimates $(\hat{\beta}_t^{bc}, \hat{\lambda}_{it})$ under $\mathbb{H}_A^{(1)}$ and the restricted estimates $(\tilde{\beta}_t^{bc}, \tilde{\lambda}_i)$ under $\mathbb{H}_0^{(1)}$. But due to the difference of the rotational matrices appearing in the probability limits of $\hat{\lambda}_{it}$ and $\tilde{\lambda}_i$, it is extremely difficult to study such a test statistic. For this reason, we propose to consider the squared L_2-distance between $X_{it}'\hat{\beta}_{it}^{bc} + \hat{\lambda}_{it}'\hat{F}_t$ and $X_{it}'\tilde{\beta}_{i0}^{bc} + \tilde{\lambda}_{i0}'\tilde{F}_t$, where the rotational matrices will not play a role asymptotically. Define the residuals under $\mathbb{H}_0^{(1)}$ and $\mathbb{H}_A^{(1)}$ respectively as

$$\tilde{\varepsilon}_{it} = Y_{it} - X_{it}'\tilde{\beta}_0^{bc} - \tilde{\lambda}_{i0}'\tilde{F}_t \text{ and } \hat{\varepsilon}_{it} = Y_{it} - X_{it}'\hat{\beta}_t^{bc} - \hat{\lambda}_{it}'\hat{F}_t.$$

Noting that the distance between $X_{it}'\hat{\beta}_t^{bc} + \hat{\lambda}_{it}'\hat{F}_t$ and $X_{it}'\tilde{\beta}_0^{bc} + \tilde{\lambda}_{i0}'\tilde{F}_t$ is the same as that between $\hat{\varepsilon}_{it}$ and $\tilde{\varepsilon}_{it}$, we propose to consider the following test statistic

$$\hat{M}^{(1)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it} - \tilde{\varepsilon}_{it}\right)^2$$

to test $\mathbb{H}_0^{(1)}$ against $\mathbb{H}_A^{(1)}$.

If we reject $\mathbb{H}_0^{(1)}$, it is valuable to gauge the possible sources of rejection by testing $\mathbb{H}_0^{(2)}$ and $\mathbb{H}_0^{(3)}$ against their respective alternatives. Under $\mathbb{H}_0^{(2)}$, we note that β_t is time-invariant. Since $\hat{\beta}_t^{bc}$ is a consistent estimator for $\beta_t = \beta_0$, it centers around β_0 and so is its average $\bar{\hat{\beta}}^{bc} \equiv \frac{1}{T}\sum_{t=1}^{T}\hat{\beta}_t^{bc}$. Under $\mathbb{H}_A^{(2)}$, $\hat{\beta}_t^{bc}$ centers around different values for different t. This motivates us to consider the following test statistic

$$\hat{M}^{(2)} = \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\beta}_t^{bc} - \bar{\hat{\beta}}^{bc}\right\|^2$$

to test $\mathbb{H}_0^{(2)}$ against $\mathbb{H}_A^{(2)}$.

To test $\mathbb{H}_0^{(3)}$ against $\mathbb{H}_A^{(3)}$, we define $\hat{W}_{it} = Y_{it} - X_{it}'\hat{\beta}_t^{bc}$. Then, we estimate the pure factor model $\hat{W}_{it} = \lambda_{it}' F_t + \eta_{it}$ via the conventional PCA as in Bai (2003) and the local PCA of Su and Wang (2017) to obtain the restricted estimates $(\tilde{\lambda}_{i0}^W, \tilde{F}_t^W)$ and unrestricted estimates $(\hat{\lambda}_{it}^W, \hat{F}_t^W)$, respectively. Then we consider the following test statistic

$$\hat{M}^{(3)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\lambda}_{it}^{W\prime}\hat{F}_t^W - \tilde{\lambda}_{i0}^{W\prime}\tilde{F}_t^W\right)^2.$$

Obviously, $\hat{M}^{(l)}$, $l = 1, 2, 3$, converge to zero in probability under the respective null hypotheses and to a positive value under the respective global alternatives. We will study their asymptotic distributions under their respective null hypotheses and sequences of local alternatives in the following subsections.

It is worth mentioning that here we focus on test statistics based on the squared L_2-distance between constrained and unconstrained estimates, which are suitable for non-sparse alternatives. Of course, one can also consider the supremum-type statistics that appear to be more appropriate to detect sparse alternatives; see, e.g., Xu (2022) for testing the time-invariance of factor loadings in a pure factor model.

4.3 Asymptotic null distributions

In this subsection, we study the asymptotic distributions of $\hat{M}^{(1)}$, $\hat{M}^{(2)}$, and $\hat{M}^{(3)}$ under their respective null hypotheses.

Under $\mathbb{H}_0^{(1)}$, we can apply Bai (2009) and Moon and Weidner (2017) to estimate the null model to obtain the bias-corrected estimator $\hat{\beta}^{bc}$ of β_0 and the restricted estimators $\tilde{F} = (\tilde{F}_1, ..., \tilde{F}_T)'$ and $\tilde{\Lambda} = (\tilde{\lambda}_1, ..., \tilde{\lambda}_N)'$ of the factor matrix F and factor loading matrix $\Lambda_0 = (\lambda_{10}, ..., \lambda_{N0})'$. Let $\tilde{V}_{NT}$ denote the $R \times R$ diagonal matrix of the first R largest eigenvalues of $(NT)^{-1}\tilde{W}\tilde{W}'$ in decreasing order, where $\tilde{W} = (\tilde{W}_1, \cdots, \tilde{W}_N)$ with $\tilde{W}_i = Y_i - X_i\hat{\beta}^{bc}$, and $H = (N^{-1}\Lambda_0'\Lambda_0)(T^{-1}F'\tilde{F})\tilde{V}_{NT}^{-1}$. Let H_0 denote the probability limit of H under $\mathbb{H}_0^{(1)}$. Note that it is also the probability limit of $H^{(t)}$ under $\mathbb{H}_0^{(1)}$ (or the local alternative $\mathbb{H}_0^{(1)}(a_{NT})$ specified in the next subsection). Let $L_{st} = k_{h,st}^* H^{(t)}H^{(t)\prime} - HH'$. Let $\xi_{h,sr} \equiv \frac{1}{T}\sum_{t=1}^{T} F_r' H_0 H_0' k_{h,st}^* k_{h,rt}^* F_t F_t' H_0 H_0' F_s$. Let $f_{sr} = F_r'\Sigma_F^{-1}F_s$. Let $\mathcal{X}_{i,s}^{(t)} = X_{is} - \frac{1}{N}\sum_{j=1}^{N} E_{\mathcal{C}}(X_{js})a_{ji}^{(t)} - \frac{1}{T}\sum_{r=1}^{T} k_{h,rt}^* f_{sr} E_{\mathcal{C}}(X_{jr}) + \frac{1}{NT}\sum_{r=1}^{T}\sum_{j=1}^{N} k_{h,rt}^* f_{sr}^{(t)} a_{ji}^{(t)} E_{\mathcal{C}}(X_{jr})$, $\varsigma_{ij,sr} = \frac{1}{T}\sum_{t=1}^{T} k_{h,st}^* k_{h,rt}^* \mathcal{X}_{is}^{(t)\prime}[D_0^{(t)}]^{-1}\,[D_0^{(t)}]^{-1}\mathcal{X}_{jr}^{(t)}$, and $\mathbb{D}^{(t)} = [D^{(t)}(F^{(t)})]^{-1}[D^{(t)}(F^{(t)})]^{-1}$. Define

$$\begin{aligned}
\mathbb{B}_{NT}^{(1)} &= \frac{h^{1/2}}{N^{1/2}T^2}\sum_{i=1}^{N}\sum_{t,s,r=1}^{T} F_t' L_{st} F_s F_r' L_{rt}' F_t E_{\mathcal{C}}(\varepsilon_{is}\varepsilon_{ir}), \\
\mathbb{B}_{1,NT}^{(2)} &= \frac{h^{1/2}}{NT}\sum_{i=1}^{N}\sum_{s,r=1}^{T}\varepsilon_{is}\varepsilon_{ir}\varsigma_{ii,sr}, \ \mathbb{B}_{2,NT}^{(2)} = \frac{2h^{1/2}}{NT^2}\sum_{1\le j<i\le N}\sum_{t=1}^{T}\sum_{s=1}^{T} k_{h,st}^{*2}\varepsilon_{is}\varepsilon_{js}\mathcal{X}_{is}^{(t)\prime}\mathbb{D}^{(t)}\mathcal{X}_{js}^{(t)}, \\
\mathbb{B}_{NT}^{(2)} &= \mathbb{B}_{1,NT}^{(2)} + \mathbb{B}_{2,NT}^{(2)},
\end{aligned}$$

$$\mathbb{V}_{NT}^{(1)} = \frac{h}{N}\sum_{i=1}^{N} E_{\mathcal{C}}\left(\frac{1}{T}\sum_{s,r=1}^{T}\xi_{h,sr}\varepsilon_{is}\varepsilon_{ir}\right)^{2}, \text{ and } \mathbb{V}_{NT}^{(2)} = \frac{4h}{N^2T^2}\sum_{1\leq i<j\leq N} E_{\mathcal{C}}\left(\sum_{1\leq s\neq r\leq T}\varepsilon_{is}\varepsilon_{jr}\varsigma_{ij,sr}\right)^{2},$$

To proceed, we impose an additional condition.

Assumption A.11 (i) For each $i \in [N]$, $\{(X_{it}, \varepsilon_{it}) : t = 1, 2, ...\}$ is conditionally strong mixing given $\mathcal{C}$ with mixing coefficients $\{\alpha_{NT,i}^{\mathcal{C}}(\cdot)\}$. $\alpha_{\mathcal{C}}(\cdot) \equiv \alpha_{NT}^{\mathcal{C}}(\cdot) \equiv \max_{1\leq i\leq N}\alpha_{NT,i}^{\mathcal{C}}(\cdot)$ satisfies $\sum_{j=1}^{\infty} j^6 [\alpha_{\mathcal{C}}(j)]^{\frac{\eta}{2+\eta}} \leq C < \infty$ a.s.

(ii) (ε_i, X_i), $i \in [N]$, are mutually independent of each other conditional on $\mathcal{C}$.

Assumption A.11(i) imposes that the process $\{(X_{it}, \varepsilon_{it})\}$ is conditionally strong mixing. For the definition of conditional strong mixing, see Prakasa Rao (2009). For the application of this concept in econometrics, see Hahn and Kuersteiner (2011), Su and Chen (2013), and Lu and Su (2016), among others. Assumption A.11(ii) imposes conditional independence of (ε_i, X_i) across i given $\mathcal{C}$. This allows for unconditional cross-sectional dependence of (ε_i, X_i) via $\mathcal{C}$. For example, if $\varepsilon_{it} = \sigma_0(F_t)\epsilon_{it}$ for some measurable function $\sigma_0(\cdot)$ and zero-mean process $\{\epsilon_{it}\}$ that are independent over i, ε_{it} can be cross-sectionally independent conditional on $\mathcal{C}$ as along as ϵ_{it}'s are independent across i. But unconditionally, ε_{it}'s are dependent across i.

The following theorem states the asymptotic null distributions of our test statistics $\hat{M}^{(l)}$, $l = 1, 2, 3$, after being properly normalized.

Theorem 4.1 *(Asymptotic null distributions) Suppose that Assumptions A.1–A.11 hold. Let $\mathbb{B}_{NT}^{(3)} = \mathbb{B}_{NT}^{(1)}$ and $\mathbb{V}_{NT}^{(3)} = \mathbb{V}_{NT}^{(1)}$. Suppose $NTh^5 = O(1)$ in (i) and (iii) below and $NTh^{9/2} = o(1)$ in (ii) below. Then*

(i) $J_{NT}^{(1)} \equiv [\mathbb{V}_{NT}^{(1)}]^{-1/2}\left(TN^{1/2}h^{1/2}\hat{M}^{(1)} - \mathbb{B}_{NT}^{(1)}\right) \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0^{(1)}$,

(ii) $J_{NT}^{(2)} \equiv [\mathbb{V}_{NT}^{(2)}]^{-1/2}\left(TNh^{1/2}\hat{M}^{(2)} - \mathbb{B}_{NT}^{(2)}\right) \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0^{(2)}$,

(iii) $J_{NT}^{(3)} \equiv [\mathbb{V}_{NT}^{(3)}]^{-1/2}\left(TN^{1/2}h^{1/2}\hat{M}^{(3)} - \mathbb{B}_{NT}^{(3)}\right) \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0^{(3)}$.

Theorem 4.1 indicates that all of the three normalized test statistics follow the standard normal distribution asymptotically under the corresponding null hypotheses. The asymptotic bias and variance of $TN^{1/2}h^{1/2}\hat{M}^{(3)}$ are the same as those of $TN^{1/2}h^{1/2}\hat{M}^{(1)}$. This is due to the faster convergence rate of the estimator of β_t than that of the factors and factor loadings under the additional bandwidth condition $NTh^5 = O(1)$. As Theorem 3.4(ii) implies, when $NTh^5 = O(1)$, $\max_t \left\|\hat{\beta}_t^{bc} - \beta_t\right\| = O_P\left((NTh/\ln T)^{-1/2}\right)$. With this result, we can show the estimation error of $\hat{\beta}_t^{bc}$ does not contribute to the asymptotic bias of $\hat{M}^{(1)}$ and $\hat{M}^{(3)}$.

Theorem 4.1(ii) requires a stronger condition on the bandwidth, namely, $NTh^{9/2} = o(1)$. This condition helps to eliminate the effect of the asymptotic bias due to the nonparametric kernel estimation in both $\hat{\beta}_t^{bc}$ and $\hat{\lambda}_{it}$. Intuitively, the asymptotic bias of such estimators, which is of order $O(h^2)$, has to be controlled as $o\left((NT)^{-1/2}h^{-1/4}\right)$ in order to ignore their effect on the asymptotic distribution of $\hat{M}^{(2)}$. We emphasize that even under $\mathbb{H}_0^{(2)}$, the asymptotic bias of $\hat{\beta}_t^{bc}$ is still $O\left(h^2\right)$ because we do not restrict λ_{it} to be time-invariant here and the estimation of λ_{it}, whose bias term is $O(h^2)$, enters the bias of $\hat{\beta}_t^{bc}$. The use of undersmoothing bandwidth to eliminate the effect of a kernel-based estimator's bias is standard in the nonparametric literature. In fact, Su and Hoshino (2016) also use such an idea to eliminate the effect of the bias of sieve estimators in testing the constancy of functional coefficients in a cross-section setup.

To implement these tests, we need to estimate both the asymptotic biases $\mathbb{B}_{NT}^{(l)}$ and the asymptotic variances $\mathbb{V}_{NT}^{(l)}$ for $l = 1, 2, 3$. Let $\hat{\varepsilon}_{is} = Y_{is} - X_{is}'\hat{\beta}_s^{bc} - \hat{\lambda}_{is}'\hat{F}_s$, $\hat{\xi}_{h,sr} = \frac{1}{T}\sum_{t=1}^T k_{h,st}^* k_{h,rt}^* \tilde{F}_r'\hat{F}_t\hat{F}_t'\tilde{F}_s$, $\hat{\eta}_{it,s} = (k_{h,st}^*\hat{F}_s'\hat{F}_t - \tilde{F}_s'\tilde{F}_t)\hat{\varepsilon}_{is}$, $\hat{\Gamma}_{it,j} = T^{-1}\sum_{s=j+1}^T \hat{\eta}_{it,s}\hat{\eta}_{it,s-j}$, and $\hat{\Xi}_{it} = \hat{\Gamma}_{it,0} + 2\sum_{j=1}^{l_T} w_{Tj}\hat{\Gamma}_{it,j}$. Here, l_T is a truncation parameter and w_{Tj} is a weighting function. We will impose some restrictions later. We propose to estimate $\mathbb{B}_{NT}^{(1)}$ and $\mathbb{V}_{NT}^{(1)}$ by

$$\begin{aligned}
\hat{\mathbb{B}}_{NT}^{(1)} &= \hat{\mathbb{B}}_{NT}^{(3)} = \frac{h^{1/2}}{N^{1/2}T}\sum_{i=1}^N\sum_{t=1}^T \hat{\Xi}_{it} = \frac{h^{1/2}}{N^{1/2}T^2}\sum_{i=1}^N\sum_{t,s=1}^T \hat{\eta}_{it,s}^2 + \frac{2h^{1/2}}{N^{1/2}T^2}\sum_{i=1}^N\sum_{t=1}^T\sum_{j=1}^{l_T} w_{Tj}\sum_{s=j+1}^T \hat{\eta}_{it,s}\hat{\eta}_{it,s-j}, \\
\hat{\mathbb{V}}_{NT}^{(1)} &= \hat{\mathbb{V}}_{NT}^{(3)} = \frac{h}{N}\sum_{i=1}^N\left(\frac{1}{T}\sum_{s,r=1}^T \hat{\xi}_{h,sr}\hat{\varepsilon}_{ir}\hat{\varepsilon}_{is}\right)^2.
\end{aligned}$$

Let $\bar{X}_j = \frac{1}{T}\sum_{t=1}^T X_{jt}$, $\hat{a}_{ji}^{(t)} = \hat{\lambda}_{jt}'(\hat{\Lambda}_t'\hat{\Lambda}_t/N)^{-1}\hat{\lambda}_{it}$, $\hat{f}_{sr} = \hat{F}_r'\hat{F}_s$, and $\hat{\mathcal{X}}_{i,s}^{(t)} = X_{is} - N^{-1}\sum_{j=1}^N \bar{X}_j\hat{a}_{ji}^{(t)} - T^{-1}\sum_{r=1}^T k_{h,rt}\hat{f}_{sr}\bar{X}_i + N^{-1}T^{-1}\sum_{r=1}^T\sum_{j=1}^N k_{h,rt}\hat{f}_{sr}\hat{a}_{ji}^{(t)}\bar{X}_j$. Let $\hat{\varsigma}_{ij,sr} = \frac{1}{T}\sum_{t=1}^T k_{h,st}^* k_{h,rt}^*\hat{\mathcal{X}}_{is}^{(t)\prime}\hat{\mathbb{D}}^{(t)}\hat{\mathcal{X}}_{jr}^{(t)}$ where $\hat{\mathbb{D}}^{(t)} = [\hat{D}^{(t)}(\hat{F}^{(t)})]^{-1}[\hat{D}^{(t)}(\hat{F}^{(t)})]^{-1}$. We propose to estimate $\mathbb{B}_{NT}^{(2)}$ and $\mathbb{V}_{NT}^{(2)}$ respectively by

$$\hat{\mathbb{B}}_{NT}^{(2)} = \hat{\mathbb{B}}_{1,NT}^{(2)} + \hat{\mathbb{B}}_{2,NT}^{(2)} \text{ and } \hat{\mathbb{V}}_{NT}^{(2)} = \frac{2h}{N^2T^2}\sum_{1\le i\ne j\le N}\left(\sum_{s,r=1}^T \hat{\varepsilon}_{is}\hat{\varepsilon}_{jr}\hat{\varsigma}_{ij,sr}\right)^2,$$

where $\hat{\mathbb{B}}_{1,NT}^{(2)} = \frac{h^{1/2}}{NT}\sum_{i=1}^N\sum_{s,r=1}^T \hat{\varepsilon}_{is}\hat{\varepsilon}_{ir}\hat{\varsigma}_{ii,sr}$ and $\hat{\mathbb{B}}_{2,NT}^{(2)} = \frac{2h^{1/2}}{NT^2}\sum_{1\le j<i\le N}\sum_{t,s=1}^T k_{h,st}^{*2}\hat{\varepsilon}_{is}\hat{\varepsilon}_{js}\hat{\mathcal{X}}_{is}^{(t)\prime}\hat{\mathbb{D}}^{(t)}\hat{\mathcal{X}}_{js}^{(t)}$. We define the feasible test statistics as follows:

$$\hat{J}_{NT}^{(l)} = \left[\hat{\mathbb{V}}_{NT}^{(l)}\right]^{-1/2}\left(TN^{1/2}h^{1/2}\hat{M}^{(l)} - \hat{\mathbb{B}}_{NT}^{(l)}\right) \text{ for } l = 1, 3 \text{ and } \hat{J}_{NT}^{(2)} = \left[\hat{\mathbb{V}}_{NT}^{(2)}\right]^{-1/2}\left(TNh^{1/2}\hat{M}^{(2)} - \hat{\mathbb{B}}_{NT}^{(2)}\right).$$

To guarantee the consistency of the estimated bias and variance terms, we introduce the following assumption.

Assumption A.12 (i) $\sup_j |w_{Tj}| \le c_w < \infty$ and $\lim_{T\to\infty} w_{Tj} = 1$ for each j.

(ii) As $(N,T)\to\infty$, $l_T(Nh)^{1/2}C_{NT}^{-2} = o(1)$, and $l_T^3/T = o(1)$.

(iii) There exists $a_0 > 0$ such that $(Nh)^{1/2}l_T^{-a_0} = o(1)$, and $\sum_{j=l_T+1}^\infty j^{a_0}\left[\alpha_{\mathcal{C}}(j)\right]^{\frac{3+2\eta}{4+2\eta}} \le C < \infty$ a.s.

Assumption A.12 imposes some conditions on w_{Tj} and l_T. The following theorem establishes the consistency of $\hat{\mathbb{B}}_{NT}^{(l)}$ and $\hat{\mathbb{V}}_{NT}^{(l)}$ and the asymptotic normality of $\hat{J}_{NT}^{(l)}$ with $l = 1, 2, 3$.

Theorem 4.2 *(Feasible test statistics) Suppose that Assumptions A.1–A.12 hold. Suppose* $NTh^5 = O(1)$ *in (i) and (iii) below and* $NTh^{9/2} = o(1)$ *in (ii) below. Then*

(i) $\hat{\mathbb{B}}_{NT}^{(1)} = \mathbb{B}_{NT}^{(1)} + o_P(1)$, $\hat{\mathbb{V}}_{NT}^{(1)} = \mathbb{V}_{NT}^{(1)} + o_P(1)$, *and* $\hat{J}_{NT}^{(1)} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0^{(1)}$,

(ii) $\hat{\mathbb{B}}_{NT}^{(2)} = \mathbb{B}_{NT}^{(2)} + o_P(1)$, $\hat{\mathbb{V}}_{NT}^{(2)} = \mathbb{V}_{NT}^{(2)} + o_P(1)$, *and* $\hat{J}_{NT}^{(2)} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0^{(2)}$,

(iii) $\hat{\mathbb{B}}_{NT}^{(3)} = \mathbb{B}_{NT}^{(3)} + o_P(1)$, $\hat{\mathbb{V}}_{NT}^{(3)} = \mathbb{V}_{NT}^{(3)} + o_P(1)$, *and* $\hat{J}_{NT}^{(3)} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0^{(3)}$.

Theorem 4.2 shows that all of our test statistics follow the asymptotic standard normal distribution and are asymptotically pivotal. We can compare the results of $\hat{J}_{NT}^{(l)}$ with $l = 1, 2, 3$ with the critical value z_α, the upper α-percentile of the $N(0,1)$ distribution, as the tests are one-sided, and reject the null hypothesis at significance level α when $\hat{J}_{NT}^{(l)} > z_\alpha$. Alternatively, to improve the finite sample performance of the tests, we propose suitable bootstrap procedures to obtain the bootstrap critical values or p-values. See Section 4.5 below.

4.4 Asymptotic local power properties

To study the asymptotic local power properties of our tests, we consider the following sequences of local alternatives:

$$\begin{aligned}
&\mathbb{H}_A^{(1)}(a_{NT}) : \lambda_{it} = \lambda_{i0} + a_{1NT} g_i\left(\frac{t}{T}\right) \text{ and } \beta_t = \beta_0 + a_{2NT} g_0\left(\frac{t}{T}\right) \text{ for each } i \text{ and } t, \\
&\mathbb{H}_A^{(2)}(a_{2NT}) : \beta_t = \beta_0 + a_{2NT} g_0\left(\frac{t}{T}\right) \text{ for each } t, \\
&\mathbb{H}_A^{(3)}(a_{1NT}) : \lambda_{it} = \lambda_{i0} + a_{1NT} g_i\left(\frac{t}{T}\right) \text{ for each } i \text{ and } t,
\end{aligned}$$

where $a_{NT} = (a_{1NT}, a_{2NT})$, $a_{1NT} \to 0$ and $a_{2NT} \to 0$ as $(N,T) \to \infty$, a_{1NT} and a_{2NT} control the speeds at which the local alternatives converge to the null hypotheses, and $g_0(\cdot)$ and $g_i(\cdot)$ are vector-valued piecewise smooth functions with finite numbers of discontinuity points. Noting that $\lambda_{i0} + a_{1NT} g_i(\frac{t}{T}) = (\lambda_{i0} + c_{i,NT}) + a_{1NT}[g_i(\frac{t}{T}) - c_{i,NT}/a_{1NT}]$ for any $c_{i,NT} = O(a_{1NT})$, below we will assume that $\int_0^1 g_i(u)\,du = 0$ for the location normalization purpose. Similarly, we also assume $\int_0^1 g_0(u)\,du = 0$. With this normalization, both (λ_{i0}, β_0) and $(g_i(\cdot), g_0(\cdot))$ can depend on the sample sizes N and T. But for notational simplicity, we continue to write them as λ_{i0}, β_0, $g_i(\cdot)$, and $g_0(\cdot)$ instead of $\lambda_{i0,NT}$, $\beta_{0,NT}$, $g_{i,NT}(\cdot)$, and $g_{0,NT}(\cdot)$.

Let $\Lambda_0 = (\lambda_{10}, ..., \lambda_{N0})'$ and $a_{ij}^{(0)} = \lambda_{i0}'(N^{-1}\Lambda_0'\Lambda_0)^{-1}\lambda_{j0}$. Define

$$D(F) = \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i - \frac{1}{T}\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} X_i' M_F X_j a_{ij}^{(0)}\right].$$

Let $g_{it} = g_i(\frac{t}{T})$ and $g_{it}^\dagger = F_t' g_i(\frac{t}{T})$ for $i = 0, 1, ..., N$. Let $g_t^\dagger = (g_{1t}^\dagger, ..., g_{Nt}^\dagger)'$. Let $\pi_{NT} = (\pi_{NT,1}, ..., \pi_{NT,P})'$ where $\pi_{NT,p} = \frac{1}{NT}\mathrm{tr}(M_F \mathbf{X}_p M_{\Lambda_0} \Delta_0')$ for $p \in [P]$, and Δ_0 is a $T \times N$ matrix with the (t,i)th element given by $X_{it}' g_{0t}$. Define

$$\Pi_1^{(1)} = \underset{(N,T)\to\infty}{\text{plim}} T^{-1} \sum_{t=1}^{T} \text{tr} \left[\left(N^{-1} \Lambda_0' g_t^{\dagger} \right) \left(N^{-1} g_t^{\dagger\prime} \Lambda_0 \right) \left(H_0^{-1} \right)' V_0^{-1} H_0^{-1} \Sigma_{\Lambda_0} \left(H_0^{-1} \right)' V_0^{-1} H_0^{-1} \right],$$

$$\Pi_2^{(1)} = \lim_{(N,T)\to\infty} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \text{tr} \left(\Sigma_F g_{it} g_{it}' \right),$$

$$\Pi_3^{(1)} = \underset{(N,T)\to\infty}{\text{plim}} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ X_{it}' [g_{0t} - D(F)^{-1} \pi_{NT}] \right\}^2, \text{ and}$$

$$\Pi^{(2)} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \| g_{0t} \|^2 .$$

To study the asymptotic power properties of $\hat{J}_{NT}^{(l)}$, $l = 1, 2, 3$, we impose the following assumption:

Assumption A.13 (i) For each $i \in [N]$, $g_i(\cdot)$ and $g_0(\cdot)$ are piecewise continuous functions with finite numbers of discontinuous points on (0,1];

(ii) $\max_{1\le r\le T} \left\| \frac{1}{NT} \sum_{s=1}^{T} \sum_{i=1}^{N} k_{h,sr} F_s \varepsilon_{is} g_{ir}' \right\| = O_P((NTh/\ln(NT))^{-1/2})$.

(iii) The (probability) limits $\Pi_1^{(1)}$, $\Pi_2^{(1)}$, $\Pi_3^{(1)}$ and $\Pi^{(2)}$ exist and are finite.

Assumption A.13 allows the slope coefficients and factor loadings to change smoothly over time or abruptly at a finite number of unknown discontinuity points. In either case, we assume that the factor loadings and the slope coefficients are uniformly bounded to facilitate the asymptotic analysis.

The following theorem studies the asymptotic local power property of $\hat{J}_{NT}^{(l)}$ for $l = 1, 2, 3$.

Theorem 4.3 *(Asymptotic local powers) Suppose that Assumptions A.1–A.13 hold. Let* $a_{1NT} = N^{-1/4} T^{-1/2} h^{-1/4}$ *and* $a_{2NT} = (NT)^{-1/2} h^{-1/4}$. *Suppose* $NTh^5 = O(1)$ *in (i) and (iii) below and* $NTh^{9/2} = o(1)$ *in (ii) below. Let* $\Pi^{(1)} = \Pi_1^{(1)} + \Pi_2^{(1)} + \Pi_3^{(1)}$, $\Pi^{(3)} = \Pi_1^{(1)} + \Pi_2^{(1)}$, $\mathbb{V}_0^{(l)} = \lim_{(N,T)\to\infty} \mathbb{V}_{NT}^{(l)}$ *and* $\pi^{(l)} = \Pi^{(l)} / (\mathbb{V}_0^{(l)})^{1/2}$ *for* $l \in [3]$. *Then*

(i) $\hat{\mathbb{B}}_{NT}^{(1)} = \mathbb{B}_{NT}^{(1)} + o_P(1)$, $\hat{\mathbb{V}}_{NT}^{(1)} = \mathbb{V}_{NT}^{(1)} + o_P(1)$, *and* $\hat{J}_{NT}^{(1)} \xrightarrow{d} N\left(\pi^{(1)}, 1\right)$ *under* $\mathbb{H}_A^{(1)}(a_{1NT})$,

(ii) $\hat{\mathbb{B}}_{NT}^{(2)} = \mathbb{B}_{NT}^{(2)} + o_P(1)$, $\hat{\mathbb{V}}_{NT}^{(2)} = \mathbb{V}_{NT}^{(2)} + o_P(1)$, *and* $\hat{J}_{NT}^{(2)} \xrightarrow{d} N\left(\pi^{(2)}, 1\right)$ *under* $\mathbb{H}_A^{(2)}(a_{2NT})$,

(iii) $\hat{\mathbb{B}}_{NT}^{(3)} = \mathbb{B}_{NT}^{(3)} + o_P(1)$, $\hat{\mathbb{V}}_{NT}^{(3)} = \mathbb{V}_{NT}^{(3)} + o_P(1)$, *and* $\hat{J}_{NT}^{(3)} \xrightarrow{d} N\left(\pi^{(3)}, 1\right)$ *under* $\mathbb{H}_A^{(3)}(a_{1NT})$.

Theorem 4.3 indicates that $\hat{J}_{NT}^{(1)}$ and $\hat{J}_{NT}^{(3)}$ can detect local alternatives converging to the respective null hypotheses at the rate of $N^{-1/4} T^{-1/2} h^{-1/4}$, which is also attainable in Su and Wang (2017) for testing time-invariant factor loadings. This rate is also comparable with the rate $N^{-1/4} T^{-1/2}$ of Su and Chen's (2013) parametric test for the homogeneity of slope coefficients in panel data models with IFEs. Interestingly, $\hat{J}_{NT}^{(2)}$ has power to detect local alternatives converging to $\mathbb{H}_0^{(2)}$ at a faster rate $(NT)^{-1/2} h^{-1/4}$ under a more restrictive condition on the bandwidth than the other two tests. This rate is comparable with the rate of local alternatives detected by Su and Hoshino's (2016) sieve-based test for the constancy of functional coefficient in the cross-section setup.

4.5 Bootstrap versions of the tests

As is well known, the kernel-based nonparametric test may be oversized in finite samples, as the asymptotic null distribution may not approximate its finite sample distribution well. Therefore, we propose bootstrap versions of our tests in this subsection.

Since we assume conditional cross-sectional independence in Assumption A.11(ii) and allow for conditional heteroskedasticity when we derive the asymptotic distributions of our test statistics under the null and local alternatives. we can follow the lead of Hansen (2000) and Gonçalves and Kilian (2004) and consider the fixed-regressor wild bootstrap to generate the bootstrap samples.

For our $\hat{J}_{NT}^{(1)}$ test, we consider the following bootstrap procedure:

1. Estimate the restricted model $Y_{it} = X_{it}'\beta_0 + \lambda_{i0}'F_t + \varepsilon_{it}^{\dagger}$ by Bai's (2009) method and the unrestricted model $Y_{it} = X_{it}'\beta_t + \lambda_{it}'F_t + \varepsilon_{it}$ by the LLS method to obtain the two sets of estimates $\{\tilde{\beta}_0^{bc}, \tilde{\lambda}_{i0}, \tilde{F}_t\}$ and $\{\hat{\beta}_t^{bc}, \hat{\lambda}_{it}, \hat{F}_t\}$. Let $\tilde{\varepsilon}_{it}$ denote the restricted residuals. Construct the test statistic $\hat{J}_{1NT}$ as in Section 4.2.

2. For $i \in [N]$ and $t \in [T]$, obtain the bootstrap error $\varepsilon_{it}^* = \tilde{\varepsilon}_{it}\varsigma_{it}$, where ς_{it} are i.i.d. $N(0,1)$ across i and t. Generate $Y_{it}^* = X_{it}'\tilde{\beta}_0^{bc} + \tilde{\lambda}_{i0}'\tilde{F}_t + \varepsilon_{it}^*$.

3. Use $\{Y_{it}^*, X_{it}\}$ to run the restricted and unrestricted models to obtain the bootstrap versions $\{\tilde{\beta}_0^{bc,*}, \tilde{\lambda}_{i0}^*, \tilde{F}_t^*\}$ and $\{\hat{\beta}_t^{bc,*}, \hat{\lambda}_{it}^*, \hat{F}_t^*\}$ of $\{\tilde{\beta}_0^{bc}, \tilde{\lambda}_{i0}, \tilde{F}_t\}$ and $\{\hat{\beta}_t^{bc}, \hat{\lambda}_{it}, \hat{F}_t\}$, respectively. Calculate the bootstrap test statistic $\hat{J}_{1NT}^*$, a bootstrap version of $\hat{J}_{1NT}$.

4. Repeat steps 2 and 3 for B times and index the bootstrap test statistics as $\{\hat{J}_{1NT,l}^*\}_{l=1}^B$. The bootstrap p-value is calculated by $p_1^* \equiv B^{-1}\sum_{l=1}^B \mathbf{1}\{\hat{J}_{1NT,l}^* > \hat{J}_{1NT}\}$.

For our $\hat{J}_{NT}^{(2)}$ test, we consider the following bootstrap procedure:

1. Estimate the unrestricted model $Y_{it} = X_{it}'\beta_t + \lambda_{it}'F_t + \varepsilon_{it}$ by the LLS method to obtain the estimates $\{\hat{\beta}_t^{bc}, \hat{\lambda}_{it}, \hat{F}_t\}$, and denote $\bar{\hat{\beta}}^{bc} = T^{-1}\sum_{t=1}^T \hat{\beta}_t^{bc}$. Let $\tilde{\varepsilon}_{it}$ denote the restricted residuals. Based on these estimates, construct the test statistic $\hat{J}_{2NT}$ as in Section 4.2.

2. For $i \in [N]$ and $t \in [T]$, obtain the bootstrap error $\varepsilon_{it}^* = \tilde{\varepsilon}_{it}\varsigma_{it}$, where ς_{it} are i.i.d. $N(0,1)$ across i and t. Generate $Y_{it}^* = X_{it}'\bar{\hat{\beta}}^{bc} + \hat{\lambda}_{it}'\hat{F}_t + \varepsilon_{it}^*$.

3. Use $\{Y_{it}^*, X_{it}\}$ to run the unrestricted model to obtain the bootstrap versions $\{\hat{\beta}_t^{bc,*}, \hat{\lambda}_{it}^*, \hat{F}_t^*\}$ of $\{\hat{\beta}_t^{bc}, \hat{\lambda}_{it}, \hat{F}_t\}$. Calculate the bootstrap test statistic $\hat{J}_{2NT}^*$, a bootstrap version of $\hat{J}_{2NT}$.

4. Repeat steps 2 and 3 for B times and index the bootstrap test statistics as $\{\hat{J}_{2NT,l}^*\}_{l=1}^B$. The bootstrap p-value is calculated by $p_2^* \equiv B^{-1}\sum_{l=1}^B \mathbf{1}\{\hat{J}_{2NT,l}^* > \hat{J}_{2NT}\}$.

For our $\hat{J}_{NT}^{(3)}$ test, we consider the following bootstrap procedure:

1. Estimate the unrestricted model $Y_{it} = X_{it}'\beta_t + \lambda_{it}'F_t + \varepsilon_{it}$ by the LLS method to obtain the estimates $\{\hat{\beta}_t^{bc}, \hat{\lambda}_{it}, \hat{F}_t\}$. Denote $W_{it} = Y_{it} - X_{it}'\hat{\beta}_t^{bc}$. Then estimate the restricted pure factor model $W_{it} = \lambda_{i0}'F_t + \varepsilon_{it}^{\dagger}$ by the PCA method to obtain the estimates $\{\tilde{\lambda}_{i0}, \tilde{F}_t\}$. Let $\tilde{\varepsilon}_{it}$ denote the restricted residuals. Based on these estimates, construct the test statistic $\hat{J}_{3,NT}$ as in Section 4.2.

2. For $i \in [N]$ and $t \in [T]$, obtain the bootstrap error $\varepsilon_{it}^* = \tilde{\varepsilon}_{it}\varsigma_{it}$, where ς_{it} are i.i.d. $N(0,1)$ across i and t. Generate $Y_{it}^* = X_{it}'\hat{\beta}_t^{bc} + \tilde{\lambda}_{it}'\tilde{F}_t + \varepsilon_{it}^*$.

3. Use $\{Y_{it}^*, X_{it}\}$ to run the restricted and unrestricted models to obtain the bootstrap versions $\{\tilde{\beta}_0^{bc,*}, \tilde{\lambda}_{i0}^*, \tilde{F}_t^*\}$ and $\{\hat{\beta}_t^{bc,*}, \hat{\lambda}_{it}^*, \hat{F}_t^*\}$ of $\{\tilde{\beta}_0^{bc}, \tilde{\lambda}_{i0}, \tilde{F}_t\}$ and $\{\hat{\beta}_t^{bc}, \hat{\lambda}_{it}, \hat{F}_t\}$, respectively. Calculate the bootstrap test statistic $\hat{J}_{3NT}^*$, a bootstrap version of $\hat{J}_{3NT}$.

4. Repeat steps 2 and 3 for B times and index the bootstrap test statistics as $\{\hat{J}_{3NT,l}^*\}_{l=1}^B$. The bootstrap p-value is calculated by $p_3^* \equiv B^{-1}\sum_{l=1}^B \mathbf{1}\{\hat{J}_{3NT,l}^* > \hat{J}_{3NT}\}$.

Although we allow serial correlation in the error terms in the original data, we do not mimic such a serial dependence structure in the error term in the bootstrap world. We can do so because our normalized statistics are asymptotically pivotal. Alternatively, one may consider the dependent wild bootstrap (DWB) procedure that was proposed and studied by Shao (2010) and Leucht and Neumann (2013) to mimic the serial dependence in the error terms. See Fu et al. (2023) for an application of DWB in specification tests for time series models.

The following theorem establishes the asymptotic validity of the above bootstrap procedures.

Theorem 4.4 *(Asymptotic validity of the bootstrap procedures) Suppose that Assumptions A.1–A.8 and A.11–A.12 hold. Suppose that (i)* $\frac{1}{T}\sum_{t=1}^T \left\|\tilde{F}_t\right\|^8 = O_P(1)$ *and (ii)* $\frac{1}{N}\sum_{i=1}^N \left\|\tilde{\lambda}_{i0}\right\|^8 = O_P(1)$*. Then* $\hat{J}_{lNT}^* \overset{D^*}{\to} N(0,1)$ *in probability for* $l \in [3]$*, where* $\overset{D^*}{\to}$ *denotes weak convergence under the bootstrap probability measure conditional on the observed sample* Y *and* X*.*

Theorem 4.4 shows that the above bootstrap procedures provide asymptotic valid approximations to the limit null distributions of $\hat{J}_{lNT}$, $l \in [3]$. This holds because we generate the bootstrap data by imposing the respective null hypotheses. If the null hypothesis does not hold in the original data, then we expect $\hat{J}_{1NT}$ and $\hat{J}_{3NT}$ to explode at the rate $T^{1/2}N^{1/4}h^{1/4}$ and $\hat{J}_{2NT}$ to explode at the rate $T^{1/2}N^{1/2}h^{1/4}$, which delivers the consistency of the bootstrap-based test $\hat{J}_{lNT}^*$, $l \in [3]$. The extra conditions (i)-(ii) in the above theorem can be easily verified if the original data satisfies either the null hypotheses or the local alternative hypotheses studied above.

5 Monte Carlo Simulations

In this section, we study the finite sample performance of our estimates and tests through Monte Carlo simulations.

5.1 Data generating processes (DGPs)

We generate the data using the following model:

$$Y_{it} = X_{it}'\beta_t + \lambda_{it}'F_t + \varepsilon_{it},$$

where $X_{it} = (X_{it,1}, X_{it,2})'$ is the observed covariate vector, $F_t = (F_{t1}, F_{t2})'$ is the unobserved factor, and ε_{it} is the error term. Both F_{t1} and F_{t2} are autoregressive processes of order one (AR(1)) with unit variances: $F_{t1} = 0.6F_{t-1,1} + u_{t1}$ with $u_{t1} \sim i.i.d. N(0, 1 - 0.6^2)$, and $F_{t2} = 0.3F_{t-1,2} + u_{t2}$ with $u_{t2} \sim i.i.d. N(0, 1 - 0.3^2)$. X_{it} may or may not be independent of F_t. We consider three types of error terms: (1) the $i.i.d.$ case, where $\varepsilon_{it} \sim i.i.d.\ N(0, 1)$; (2) the heteroskedastic case, where $\varepsilon_{it} = \sigma_i v_{it}$, with $\sigma_i \sim i.i.d.\ U(0.5, 1.5)$ and $v_{it} \sim i.i.d.\ N(0, 1)$; (3) the serially dependent case, where $\varepsilon_{it} = 0.5\varepsilon_{i,t-1} + \epsilon_t$ with $\epsilon_t \sim i.i.d.\ N(0, 1 - 0.5^2)$.

For the coefficients $\beta_t = (\beta_{1t}, \beta_{2t})'$ and the factor loadings $\lambda_{it} = (\lambda_{it,1}, \lambda_{it,2})'$, we consider the following six DGPs. Note that when we say $X_{it} \sim \text{AR}(1)$, we mean that $X_{it,1} = 0.7X_{i,t-1,1} + u_{it,1}$ with $u_{it,1} \sim i.i.d.\ N(0, 1 - 0.7^2)$ and $X_{it,2} = 0.4X_{i,t-1,2} + u_{it,2}$ with $u_{it,2} \sim i.i.d.\ N(0, 1 - 0.4^2)$.

DGP 1 (Time-invariant slope and factor loadings, X_{it} is independent of F_t)

$$\begin{aligned}
&\beta_{1t} = 0.75, \beta_{2t} = 0.25, \lambda_{it} = \lambda_{i,0} \sim i.i.d.\ N(0, I_2),\\
&X_{it} \sim \text{AR}(1),\ \text{independent of } F_t.
\end{aligned}$$

DGP 2 (Time-invariant slope, TV factor loadings, X_{it} is independent of F_t)

$$\begin{aligned}
&\beta_{1t} = 0.75, \beta_{2t} = 0.25,\\
&\lambda_{it,1} = \lambda_{i0,1} \sim i.i.d.\ N(0, 1), \lambda_{it,2} = \cos(\pi(t/T + i/N)),\\
&X_{it} \sim \text{AR}(1),\ \text{independent of } F_t.
\end{aligned}$$

DGP 3 (Time-invariant slope, TV factor loadings, X_{it} depends on F_t)

$$\begin{aligned}
&\beta_{1t} = 0.75, \beta_{2t} = 0.25,\\
&\lambda_{it,1} = \lambda_{i0,1} \sim i.i.d.\ N(0, 1), \lambda_{it,2} = \cos(\pi(t/T + i/N)),\\
&X_{it} = 0.5\gamma_i' F_t + 0.5u_{it},\ \text{with } \gamma_i \sim i.i.d.\ N(0, 1) \text{ and } u_{it} \sim i.i.d.\ N(0, 1).
\end{aligned}$$

DGP 4 (TV slope, time-invariant factor loadings, X_{it} is independent of F_t)

$$\begin{aligned}
&\beta_{1t} = \sin(0.5\pi t/T), \beta_{2t} = 2(t/T - 0.8)^2,\\
&\lambda_{it} = \lambda_{i0} \sim i.i.d.\ N(0, I_2),\\
&X_{it} \sim \text{AR}(1),\ \text{independent of } F_t.
\end{aligned}$$

DGP 5 (TV slope, TV factor loadings, X_{it} is independent of F_t)

$$\begin{aligned}&\beta_{1t} = \sin(0.5\pi t/T), \beta_{2t} = 2(t/T - 0.8)^2,\\&\lambda_{it,1} = \lambda_{i0,1} \sim i.i.d.\ N(0,1), \lambda_{it,2} = \cos(\pi(t/T + i/N)),\\&X_{it} \sim \text{AR}(1),\ \text{independent of } F_t.\end{aligned}$$

DGP 6 (TV slope, TV factor loadings, X_{it} depends on F_t)

$$\begin{aligned}&\beta_{1t} = \sin(0.5\pi t/T), \beta_{2t} = 2(t/T - 0.8)^2,\\&\lambda_{it,1} = \lambda_{i0,1} \sim i.i.d.\ N(0,1), \lambda_{it,2} = \cos(\pi(t/T + i/N)),\\&X_{it} = 0.5\gamma_i' F_t + 0.5u_{it},\ \text{with } \gamma_i \sim i.i.d.\ N(0,1) \text{ and } u_{it} \sim i.i.d.\ N(0,1).\end{aligned}$$

The above six DGPs describe various specifications of the regression coefficients, factor loadings, and dependence structure between X_{it} and F_t. DGP 1 specifies a panel data model with time-invariant coefficients and factor loadings, which is also considered in Bai (2009). DGPs 2-3 have time-invariant coefficients and TV factor loadings. DGPs 4-6 are panel data models with TV coefficients, in which the factor loadings are specified as time-invariant in DGP 4 and TV in DGPs 5-6. We note that in DGPs 3 and 6, X_{it} is correlated with F_t, which is a reasonable setting for macroeconomic datasets. In this case, we expect that Bai's (2009) estimator should yield an inconsistent estimate since it treats factor loadings as time-invariant, and thus the error term may contain variations in the factors.

5.2 Determination of the number of common factors

In this subsection, we examine the finite sample performance of the IC in (3.3) to determine the number of common factors in the TV panel model with TV IFEs. Specifically, we consider the following two information criteria:

$$\begin{aligned}IC_1(R) &= \ln V(R) + R\left(\frac{N+Th}{NTh}\right)\ln\left(\frac{NTh}{N+Th}\right), \text{ and}\\IC_2(R) &= \ln V(R) + R\left(\frac{N+Th}{NTh}\right)\ln C_{NT}^2,\end{aligned}$$

where $V(R)$ is the objective function given in (3.2). For each DGP, we generate 500 data sets with sample sizes $N = T = 40, 60, 80, 100$. We use the Epanechnikov kernel and the Silverman's rule of thumb (RoT) bandwidth $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$. We have also tried the Uniform kernel and the Quartic kernel, and the RoT bandwidth with different tuning parameters. Our simulation studies show that the choice of kernel function and the bandwidth have little impact on the performance of our information criteria.

We use two measures to evaluate the IC, i.e., the average number of common factors and the empirical probability of correct selection over 500 replications. Tables 1 and 2 report the empirical probability of correct selection and the average number of common factors, respectively. We note that our information criteria work well when the sample sizes N and T are large. When the sample sizes are small, they tend to overestimate the number of common factors. The poor performance of our information criteria in the small T case can be attributed to the use of nonparametric estimation for the TV slope coefficient and TV factor loadings. In addition, we find that IC_2 tends to outperform IC_1 when the sample sizes are small. Hence, we suggest using IC_2 if the sample sizes are not large enough.

Table 1: Performance of information criteria in determining the number of factors: empirical probabilities of correct selection

		$N = T = 40$		$N = T = 60$		$N = T = 80$		$N = T = 100$	
Error	DGP	IC_1	IC_2	IC_1	IC_2	IC_1	IC_2	IC_1	IC_2
	1	0.038	0.156	0.300	0.730	0.860	0.986	0.998	1.000
	2	0.204	0.532	0.790	0.964	0.992	1.000	1.000	1.000
(1)	3	0.214	0.466	0.788	0.954	0.992	0.998	1.000	1.000
	4	0.054	0.262	0.492	0.862	0.948	1.000	1.000	1.000
	5	0.366	0.664	0.872	0.986	1.000	1.000	1.000	1.000
	6	0.266	0.558	0.850	0.962	0.994	1.000	1.000	1.000
	1	0.010	0.076	0.198	0.572	0.730	0.966	0.988	1.000
	2	0.100	0.316	0.540	0.846	0.932	0.988	0.998	1.000
(2)	3	0.118	0.310	0.522	0.814	0.922	0.996	1.000	1.000
	4	0.016	0.138	0.268	0.642	0.836	0.986	0.988	1.000
	5	0.162	0.400	0.696	0.902	0.974	0.996	0.998	0.998
	6	0.126	0.366	0.604	0.842	0.950	0.996	1.000	1.000
	1	0.022	0.102	0.098	0.368	0.304	0.712	0.642	0.904
	2	0.190	0.420	0.522	0.778	0.802	0.944	0.940	0.992
(3)	3	0.240	0.444	0.576	0.820	0.864	0.970	0.962	0.996
	4	0.036	0.164	0.220	0.572	0.526	0.870	0.782	0.960
	5	0.258	0.540	0.656	0.876	0.902	0.984	0.990	0.998
	6	0.252	0.476	0.636	0.850	0.894	0.972	0.974	0.994

Notes: (i) IC_1 and IC_2 denote the information criteria with $\rho_{NT,1} = \frac{N+Th}{NTh} \ln(\frac{NTh}{N+Th})$ and $\rho_{NT,2} = \frac{N+Th}{NTh} \ln C_{NT}^2$ respectively; (ii) The main entries report the empirical probability of correct selection based on 500 replications.

5.3 The estimation of slope coefficients

We now compare LLS estimator with Bai's (2009) least squares (LS) estimator, which assumes time-invariant slope coefficients and factor loadings. Specifically, we evaluate the estimation performance

Table 2: Performance of information criteria in determining the number of factors: average number of factors

		$N = T = 40$		$N = T = 60$		$N = T = 80$		$N = T = 100$	
Error	DGP	IC_1	IC_2	IC_1	IC_2	IC_1	IC_2	IC_1	IC_2
(1)	1	2.962	2.844	2.700	2.27	2.140	2.014	2.002	2
	2	2.796	2.468	2.210	2.036	2.008	2	2	2
	3	2.786	2.534	2.212	2.046	2.008	2.002	2	2
	4	2.946	2.738	2.508	2.138	2.052	2	2	2
	5	2.634	2.328	2.128	2.014	2	2	2	2
	6	2.730	2.438	2.150	2.038	2.006	2	2	2
(2)	1	2.990	2.924	2.802	2.428	2.270	2.034	2.012	2
	2	2.900	2.676	2.460	2.154	2.068	2.012	2.002	2
	3	2.882	2.690	2.478	2.182	2.078	2.004	2	2
	4	2.984	2.862	2.732	2.358	2.164	2.014	2.012	2
	5	2.838	2.600	2.304	2.098	2.026	2.004	2.002	2.002
	6	2.874	2.630	2.396	2.158	2.050	2.004	2	2
(3)	1	2.978	2.898	2.902	2.632	2.696	2.288	2.358	2.096
	2	2.810	2.580	2.478	2.222	2.198	2.056	2.060	2.008
	3	2.760	2.556	2.424	2.180	2.136	2.030	2.038	2.004
	4	2.964	2.836	2.780	2.428	2.474	2.130	2.218	2.040
	5	2.742	2.456	2.344	2.124	2.098	2.016	2.010	2.002
	6	2.748	2.524	2.364	2.150	2.106	2.028	2.026	2.006

Notes: (i) IC_1 and IC_2 denote the information criteria with $\rho_{NT,1} = \frac{N+Th}{NTh}\ln(\frac{NTh}{N+Th})$ and $\rho_{NT,2} = \frac{N+Th}{NTh}\ln C_{NT}^2$ respectively; (ii) The main entries report the average number of factors based on 500 replications.

using mean normed error (MNE) of the estimators for the slope matrix $\beta = (\beta_1, \ldots, \beta_T)'$. The normed error is calculated using the Frobenius norm. Specifically, we calculate MNE of a slope coefficient matrix estimator $\hat{\beta}$ by $\|\hat{\beta} - \beta^0\|/\sqrt{T}$. For each DGP, we simulate 500 data sets with sample sizes $N, T = 40, 60, 80, 100$. We use the Epanechnikov kernel and Silverman's rule of thumb (RoT) bandwidth $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$. For simplicity, we use the true number of factors ($R = 2$) here.

Table 3: Finite-Sample Performance of the Slope Estimators

		$N=T=40$		$N=T=60$		$N=T=80$		$N=T=100$	
Type of Error	DGP	LLS	LS	LLS	LS	LLS	LS	LLS	LS
	1	0.0442	0.0239	0.0293	0.0150	0.0224	0.0115	0.0183	0.0096
	2	0.0454	0.0275	0.0302	0.0179	0.0227	0.0131	0.0181	0.0101
(1)	3	0.0903	0.1774	0.0590	0.1776	0.0452	0.1896	0.0361	0.1958
	4	0.0640	0.3336	0.0478	0.3363	0.0406	0.3378	0.0357	0.3387
	5	0.0646	0.3337	0.0483	0.3363	0.0411	0.3378	0.0359	0.3387
	6	0.1003	0.3393	0.0703	0.3407	0.0552	0.3412	0.0458	0.3411
	1	0.0463	0.0244	0.0308	0.0158	0.0231	0.0120	0.0189	0.0099
	2	0.0471	0.0274	0.0309	0.0179	0.0238	0.0137	0.0191	0.0106
(2)	3	0.0960	0.1751	0.0625	0.1813	0.0468	0.1912	0.0371	0.2064
	4	0.0658	0.3336	0.0494	0.3363	0.0409	0.3378	0.0360	0.3387
	5	0.0667	0.3338	0.0495	0.3364	0.0411	0.3378	0.0362	0.3387
	6	0.1068	0.3401	0.0709	0.3405	0.0567	0.3406	0.0464	0.3414
	1	0.0479	0.0264	0.0326	0.0174	0.0246	0.0128	0.0200	0.0101
	2	0.0491	0.0296	0.0324	0.0187	0.0246	0.0142	0.0202	0.0117
(3)	3	0.0878	0.1679	0.0592	0.1809	0.0441	0.1854	0.0356	0.1961
	4	0.0666	0.3336	0.0496	0.3363	0.0415	0.3378	0.0364	0.3387
	5	0.0679	0.3339	0.0500	0.3364	0.0417	0.3378	0.0367	0.3387
	6	0.1010	0.3391	0.0683	0.3398	0.0552	0.3410	0.0461	0.3411

Notes: The main entries report the mean normed errors (MNEs) based on 500 replications.

Table 3 reports the MNEs for Bai's (2009) LS estimator and our LLS estimator with various types of error terms based on 500 replications. We summarize some important findings from Table 3. First, as shown in the table, the MNEs of the LLS estimators decline rapidly for all six DGPs as the sample sizes (N, T) increase, confirming the consistent property of the LLS estimator. Second, when both the regression coefficients and factor loadings are time-invariant as in DGP 1, the LS estimator significantly outperforms the LLS estimator as expected. Third, in panel data models with constant regression coefficients and TV factor loadings, the LS estimators may or may not outperform the LLS estimator depending on whether the regressors and factors are uncorrelated. When the regressors and factors are uncorrelated as in DGP 2, the LS estimator typically outperforms the LLS estimator because, in such a scenario, one can readily show the bias-corrected LS estimator is still $\sqrt{NT}$-consistent and is asymptotically more efficient than the LLS estimator. In contrast, when the observed regressors are correlated with the latent factors as in DGP 3, the LS estimator yields an inconsistent estimator of the slope coefficient and underperforms vastly the consistent LLS estimator. Fourth, when the slope coefficients are TV as in DGPs 4-6, the LLS estimator substantially outperforms the LS estimator. Fifth, the above findings are also true irrespective of whether the error terms are serially correlated or heteroskedastic.

5.4 The performance of the tests

To implement our tests, we apply the Epanechnikov kernel and the Silverman's rule-of-thumb bandwidth $h = (2.35/\sqrt{12})T^{-1/4}N^{-1/4}$. We have also tried the Uniform kernel and the Quartic kernel, and the rule-of-thumb bandwidth with different tuning parameters. Our simulation studies show that the choice of kernel function has little impact on the performance of our test. However, the empirical sizes are a bit sensitive to bandwidth selection. To alleviate this problem, we apply the bootstrap procedure proposed in Section 4.5. We consider 500 replications with $B = 200$ bootstrap resamples for the bootstrap-based test. Since the bootstrap procedure, combined with our LLS and local PCA, is rather time-consuming, we consider two sample sizes here: $N = T = 40$, and 60. These sample sizes are comparable to that of the empirical application in Section 6.

Table 4 reports the empirical rejection rates for the bootstrap-based tests. Note that $\mathbb{H}_0^{(1)}$ is true for DGP 1, $\mathbb{H}_0^{(2)}$ is true for DGPs 1, 2, and 3, and $\mathbb{H}_0^{(3)}$ is true for DGPs 1 and 4. The rejection rates in these cases serve as empirical sizes, while the rest are empirical powers. We summarize some important findings from Table 4. First, the bootstrap tests can control the size reasonably well, although our results exhibit some mild undersize distortions for some of the DGPs and test statistics. Second, all three test statistics exhibit satisfactory empirical power even when the sample size is small ($N = T = 40$). When the sample size increases to $N = T = 60$, the empirical power increases fast and it approaches 1 in most cases. Third, when the error term is heteroskedastic or serially correlated (error types (2) and (3), respectively), the results are qualitatively the same, meaning that the tests are robust to heteroskedasticity and serial correlation in the error term.

6 An Empirical Application to the Phillips Curve

We apply our model and methodology to the analysis of the Phillips curve using a panel data set of the US state-level unemployment rates and nominal wages. Specifically, we consider the following regression:

$$wage\_growth_{it} = \beta_t \; unemploy\_rate_{it} + \lambda_{it}' F_t + \varepsilon_{it}, \tag{6.1}$$

Table 4: Empirical Rejection Rates of the Bootstrapped Test Statistics

		$\mathbb{H}_0^{(1)}$				$\mathbb{H}_0^{(2)}$				$\mathbb{H}_0^{(3)}$			
		$N=T=40$		$N=T=60$		$N=T=40$		$N=T=60$		$N=T=40$		$N=T=60$	
Error	DGP	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
(1)	1	0.026	0.060	0.032	0.062	0.054	0.098	0.054	0.086	0.020	0.048	0.030	0.056
	2	0.738	0.816	0.954	0.972	0.060	0.124	0.050	0.106	0.732	0.820	0.958	0.974
	3	0.620	0.716	0.892	0.946	0.040	0.086	0.042	0.082	0.694	0.790	0.918	0.966
	4	0.994	0.998	1	1	1	1	1	1	0.022	0.036	0.034	0.058
	5	0.990	0.996	1	1	1	1	1	1	0.696	0.790	0.950	0.978
	6	0.794	0.902	0.974	0.984	1	1	1	1	0.620	0.720	0.916	0.938
(2)	1	0.022	0.068	0.042	0.092	0.046	0.088	0.052	0.086	0.024	0.060	0.032	0.072
	2	0.634	0.758	0.972	0.984	0.058	0.106	0.052	0.094	0.640	0.760	0.992	1
	3	0.462	0.556	0.910	0.952	0.046	0.082	0.038	0.072	0.548	0.652	0.942	0.984
	4	0.994	0.998	1	1	1	1	1	1	0.020	0.042	0.032	0.050
	5	0.954	0.976	1	1	1	1	1	1	0.562	0.684	0.984	0.996
	6	0.682	0.786	0.984	0.990	1	1	1	1	0.502	0.612	0.932	0.944
(3)	1	0.042	0.086	0.064	0.142	0.056	0.102	0.048	0.092	0.028	0.092	0.060	0.132
	2	0.732	0.846	0.988	1	0.052	0.096	0.062	0.132	0.744	0.862	0.980	0.998
	3	0.624	0.772	0.942	0.968	0.028	0.072	0.038	0.082	0.732	0.824	0.962	0.996
	4	0.992	1	1	1	1	1	1	1	0.040	0.082	0.042	0.078
	5	0.968	0.992	1	1	1	1	1	1	0.702	0.782	0.978	0.988
	6	0.782	0.874	0.990	0.994	1	1	1	1	0.642	0.740	0.952	0.978

Note: The main entries report the empirical rejection rates under each DGP based on 500 replications and 200 bootstrap resamples.

where $wage\_growth$ represents the year-over-year growth rate of nominal hourly wages and $unemploy\_rate$ represents the unemployment rate. The subscript i denotes state and t denotes quarter. F_t is the $R \times 1$ vector of unobserved common factors that affect wage growth. We allow the factor loading λ_{it} to be varying over time. Note that (6.1) is the panel-data version of the original formulation of the Phillips curve (Phillips, 1958). The coefficient β_t gives us the slope of the Phillips curve. Recently, there is a heated debate in the policy circle on the shape and the evolution of the Phillips curve (e.g., Hooper et al. , 2020; Del Negro et al., 2020). To shed some light on the TV nature of the Phillips curve, we allow the slope coefficient to change over time.

We obtain the data from the Federal Reserve Economic Data (FRED). Monthly data are available for the state-level unemployment rates (since 1976) and hourly wages (since 2007). From the latter, we obtain the year-over-year wage growth rate. We transform monthly observations into quarterly data, which is less noisy than the monthly series, by averaging the monthly observations. The balanced panel data has 51 cross-sections over the time span of 57 quarters from 2008Q1 to 2022Q1.

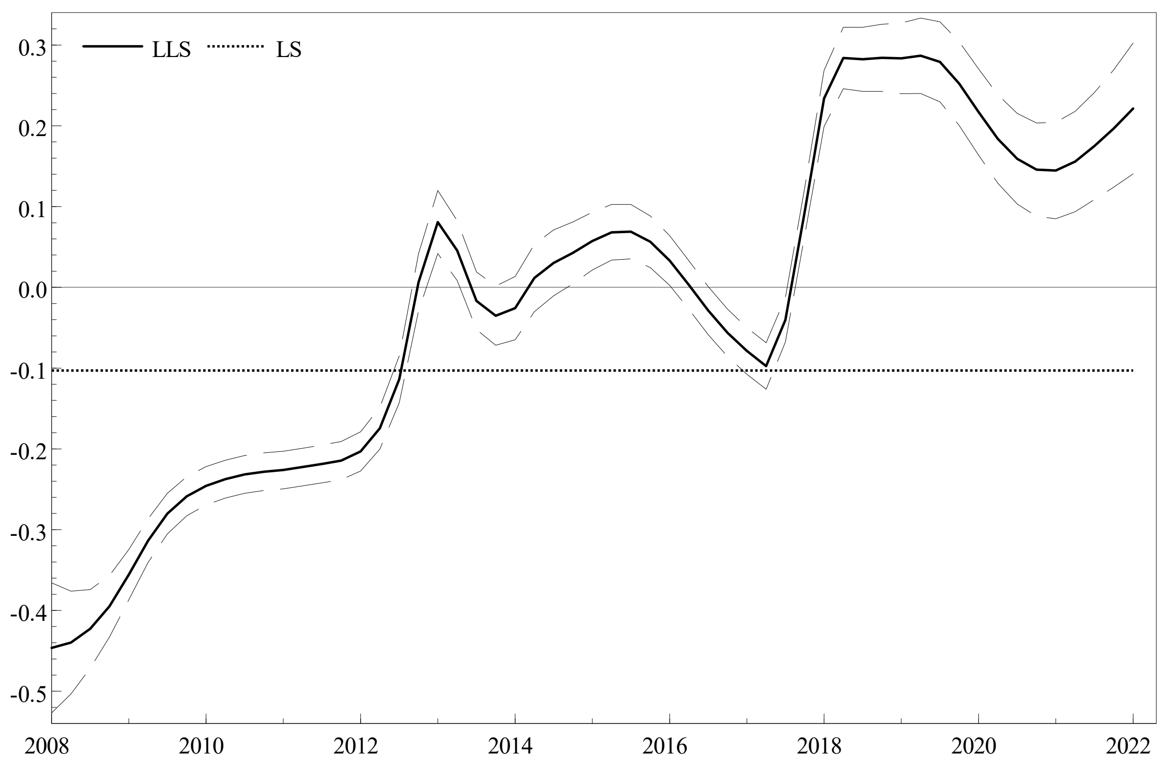

Figure 1: Slope Estimates of the Phillips Curve

We use the information criterion in (3.3) to determine the number of factors. Setting ρ_{NT} to be either $\frac{N+Th}{NTh}\ln(\frac{NTh}{N+Th})$ or $\frac{N+Th}{NTh}\ln C_{NT}^2$ with $C_{NT} = \min\{\sqrt{Th}, \sqrt{N}, h^{-2}\}$ in (3.3) yields the same estimate of R as $\hat{R} = 2$. Using $h = (2.35/\sqrt{12})(NT)^{-1/4}$ for $\hat{M}^{(1)}$ and $\hat{M}^{(3)}$ and $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$ for $\hat{M}^{(2)}$, we obtain the bootstrap p-values for testing $\mathbb{H}_0^{(1)}$, $\mathbb{H}_0^{(2)}$, and $\mathbb{H}_0^{(3)}$ respectively as 0.0000, 0.0401, and 0.0261. That is, the bootstrap versions of the test statistics reject all three null hypotheses at the 5% significance level. Overall, the US state-level data provide some strong support for a TV slope of the Phillips curve and TV factor loadings. Figure 1 shows the estimates of the slope of the Phillips curve obtained by LLS, LS, and the 95% pointwise confidential interval for the LLS estimate by using a undersmoothing bandwidth to eliminate the nonparametric kernel bias. The LS estimation treats both the slope and the factor loading as constant over time. The estimated slope is almost flat, consistent with the conventional wisdom that the Phillips curve has flattened after the 2008 global financial crisis (GFC). From the LLS estimates, we observe a mild negative-sloping Phillips curve in the wake of GFC. The curve gradually flattens and the slope fluctuates closely around zero between 2012 and 2017. Then the slope pivots to the positive territory, indicating that the shocks to the economy in the past five years are mainly from the supply side (e.g., the Sino-US trade war and the supply-chain disruptions during the Covid-19 pandemic).

7 Conclusion

The panel data models with IFEs have been extensively investigated in the literature. However, the conventional panel data models with IFEs assume the slope coefficients and the factor loadings to be time-invariant, which may not be true in practice. In this paper, we introduce a TV dynamic panel data model with TV unobservable IFEs, where the coefficients and factor loadings are allowed to change smoothly over time. We propose a local version of the least squares and PCA method to estimate the TV coefficients, TV factor loadings, and common factors simultaneously. We provide a bias-corrected LLS estimator for the TV coefficients and establish the uniform convergence and limiting distributions for all of the estimators in the large N and large T framework. We also propose a BIC-type information criterion to determine the number of common factors in the IFEs, which is robust to the structural changes in the coefficients and factor loadings. Based on the estimates, we propose three test statistics to check the stability of slope coefficients and/or factor loadings. We first construct an L_2-distance-based test to check the stability of both slope coefficients and factor loadings. If we reject the null hypothesis, it is meaningful to gauge possible sources of rejection. We further propose two test statistics to check the stability of slope coefficients and factor loadings, respectively. By construction, our tests can capture both smooth and abrupt structural changes in the factor loadings and slope coefficients without knowing the number of breaks in the data.

Monte Carlo studies demonstrate the excellent performance of our estimators and the BIC-type information criterion in determining the number of common factors. We also show the reasonable size and excellent power of our tests in checking the time-invariance of slope coefficients and/or factor loadings. In an application to the analysis of the Phillips curve using the U.S. state-level unemployment rates and nominal wages, we find some evidence of the TV behavior of both the factor loadings and slope coefficient.

Several extensions are possible. First, one can extend our model by allowing for the cross-sectional heterogeneity in the slope coefficient. Since the panel data usually cover individual units sampled from different backgrounds and with different individual characteristics, it is possible that the slope coefficients may exhibit cross-sectional heterogeneity. One can extend our slope coefficient from β_t to β_{it} and propose relevant tests to check the existence of cross-sectional heterogeneity. Second, it may be valuable to identify the types of structural changes, i.e., the smooth structural changes versus abrupt structural breaks, in the loadings or the slope coefficients. Third, it is interesting to apply our model and method to program evaluations and compare it with some existing methods such as difference in difference and synthetic control. We will pursue some of these extensions in subsequent studies.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71, 135-172.

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

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 90, 191-221.

Bai, J., & Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74(4), 1133-1150.

Baltagi, B. H., Feng, Q., & Kao, C. (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, 191(1), 176-195.

Baltagi, B. H., Kao, C., & Liu, L. (2017). Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term. *Econometric Reviews* 36(1-3), 85-102.

Breitung, J. & Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163, 71-84.

Chen, B., & Huang, L. (2018). Nonparametric testing for smooth structural changes in panel data models. *Journal of Econometrics* 202(2), 245-267.

Chen, J. (2019). Estimating latent group structure in time-varying coefficient panel data models. *The Econometrics Journal* 22(3), 223-240.

Chen, L., J. Dolado, & J. Gonzalo (2014) Detecting big structural breaks in large factor models. *Journal of Econometrics* 180, 30-48.

Cheng, T., Gao, J., & Yan, Y. (2019). Regime switching panel data models with interactive fixed effects. *Economics Letters*, 177, 47-51.

Cheng, X., Z. Liao, & F. Schorfheide (2016) Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies* 83, 1511-1543.

Chudik, A., & Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2), 393-420.

Del Negro, M., Lenza, M., Primiceri, G. E., & Tambalotti, A. (2020). What's up with the Phillips Curve? . *National Bureau of Economic Research (No. w27003).*

De Wachter, S., & Tzavalis, E. (2012). Detection of structural breaks in linear dynamic panel data models. *Computational Statistics & Data Analysis* 56(11), 3020-3034.

Dong, C., Gao, J., & Peng, B. (2021). Varying-coefficient panel data models with nonstationarity and partially observed factor structure. *Journal of Business & Economic Statistics*, 39(3), 700-711,

Feng, Q., Kao, C., & Lazarová, S. (2008). Estimation and identification of change points in panel models. *Working paper*, Nanyang Technological University.

Fu, Z., Hong, Y., & Wang, X. (2023). Testing for structural changes in large dimensional factor models via discrete Fourier transform. Journal of Econometrics, 233(1), 302-331.

Fu, Z., Su, L., & Wang, X. (2023). Time-Varying FAVAR Model Estimation and Inference, *Journal of Business & Economic Statistics*, forthcoming.

Gonçalves, S. & L. Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89-120.

Hahn, J., Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152-1191.

Han, X. & A. Inoue (2015) Tests for parameter instability in dynamic factor models. *Econometric Theory* 31, 1117-1152.

Hansen, B. E. (2000) Testing for structural change in conditional models. *Journal of Econometrics* 97, 93-115.

Hong, Y., & Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rate. *Review of Financial Studies* 18, 37-84.

Hooper, P., Mishkin, F. S., & Sufi, A., (2020). Prospects for inflation in a high pressure economy: Is the Phillips curve dead or is it just hibernating? *Research in Economics*, 74(1), 26-62.

Leucht, A., & Neumann, M. H. (2013). Dependent wild bootstrap for degenerate U-and V-statistics. *Journal of Multivariate Analysis*, 117, 257-280.

Li, D., Chen, J., & Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *Econometrics Journal* 14(3), 387-408.

Li, D., Qian, J., & Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516), 1804-1819.

Li, K., Cui, G., & Lu, L. (2020). Efficient estimation of heterogeneous coefficients in panel data models with common shocks. *Journal of Econometrics*, 216(2), 327-353.

Li, Q., & Racine, J.S. (2007). *Nonparametric Econometrics: Theory and Practice.* Princeton University Press.

Liu, F., Gao, J., & Yang, Y. (2019). Nonparametric estimation in panel data models with heterogeneity and time-varyingness. *Working paper*, Monash University.

Liu, F., Gao, J., & Yang, Y. (2021). Time-varying panel data models with an additive factor structure. *Working paper*, Monash University.

Lu, X., & Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics* 190(1), 148-175.

Miao, K., Li, K., & Su, L. (2020). Panel threshold models with interactive fixed effects. *Journal of Econometrics,* 219(1), 137-170..

Moon, H. R., & Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83(4), 1543-1579.

Moon, H. R., & Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* 33(1), 158-195.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4), 967-1012.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. *Economica*, 25(100), 283-299.

Prakasa Rao, B. L. S. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* 61, 441-460.

Qian, J., & Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics* 191(1), 86-109.

Robinson, P. M. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics* 169(1), 4-14.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489), 218-235.

Stock, J.H., & Watson, M.W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167-1179.

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

Su, L. and T. Hoshino, (2016). Sieve instrumental variable quantile regression estimation of functional coefficient models. *Journal of Econometrics* 191, 231-254.

Su, L., & Wang, X. (2017). On time-varying factor models: estimation and testing. *Journal of Econometrics* 198(1), 84-101.

Su, L., & Wang, X. (2020a). Corrigendum to "On time-varying factor models: Estimation and testing" [J. Econometrics 198 (2017) 84-101]. *Working paper.*

Su, L., & Wang, X. (2020b). Testing for structural changes in factor models via a nonparametric regression. *Econometric Theory*, 36(6), 1127-1158.

Su, L., Wang, X., & Jin, S. (2019). Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics*, 37(2), 334-349.

Xu, W. (2022). Testing for time-varying factor loadings in high-dimensional factor models. *Econometric Reviews*, 41(8), 918-965.

Yamamoto, Y. & S. Tanaka (2015) Testing for factor loading structural change under common breaks. *Journal of Econometrics* 189, 187-206.