Time-varying Factor-augmented Forecasting Models with Variable Selection*

Xiyuan Liu[a], Zongwu Cai[b], and Liangjun Su[a]

[a] School of Economics and Management, Tsinghua University, Beijing, Beijing 100084, China
[b] Department of Economics, University of Kansas, Lawrence, KS 66045, USA

September 26, 2025

Abstract

We study a novel time-varying (TV) factor-augmented (FA) forecasting model, where the forecast target is driven by a strict subset of all the latent factors driving the predictors. To consistently select the target-related factors and estimate the TV parameters simultaneously, we first obtain the unobserved common factors via the local principal component analysis. Next, we conduct a variable selection procedure via a time-varying weighted group least absolute shrinkage and selection operator to select relevant factors. The identification restrictions used in this paper permit asymptotically rotation-free estimation of both factors and loadings. The asymptotic properties, such as consistency, sparsity and the oracle property of these two-step estimators are established. Simulation studies demonstrate the excellent finite sample performance of the proposed estimators. In an empirical application to the U.S. macroeconomic dataset, we show that the penalized TV-FA forecasting model outperforms the conventional TV-FAVAR model in predicting certain key macroeconomic series.

Key words: Factor-augmented forecasting models; Local-linear smoothing; Structural change; Weighted group LASSO, Time-varying modeling

JEL Classification: C13, C23, C33, C38.

*The authors acknowledge the financial support from the National Natural Science Foundation of China (NSFC) under the grant No. 72133002 and 72033008. Address correspondence to: Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, 100084, China; E-mail: sulj@sem.tsinghua.edu.cn.

1 Introduction

Factor-augmented (FA) forecasting models and their variants have garnered increasing attention in finance and macroeconomics. By incorporating unobservable common factors into a linear regression framework, these models effectively summarize high-dimensional information and achieve significant dimension reduction. Empirical studies demonstrate that integrating latent common factors improves economic predictions and clarifies economic relationships (e.g., Stock and Watson (2002)). Seminal contributions include Bernanke et al. (2005)'s FA vector autoregressive model (FAVAR) for identifying monetary transmission mechanisms, Bai and Ng (2006)'s asymptotic theory for FA regressions, and Bai et al. (2016)'s work on identification and inference for FAVAR models. This body of literature provides a crucial theoretical foundation for FA forecasting.

However, most existing FA models rest on the assumption of time-invariant factor loadings and regression coefficients, an assumption often too restrictive for financial and macroeconomic datasets spanning long periods. Over such extended horizons, structural shifts in institutional frameworks, business cycles, and technology can alter the interdependencies among variables (e.g., Stock and Watson (2002), Stock and Watson (2009), Su and Wang (2017), and Pelger and Xiong (2022)). To better capture these evolving nonlinear relationships, it is essential to allow both factor loadings and regression coefficients to vary over time.

While structural change is a well-studied econometric topic in both linear regressions (e.g., Bai and Perron (1998) and Qu and Perron (2007)) and factor models (e.g., Breitung and Eickmeier (2011), Chen et al. (2014), Corradi and Swanson (2014), Han and Inoue (2015), Su and Wang (2017), Su and Wang (2020b), Baltagi et al. (2021), Fu et al. (2023) and Peng et al. (2025)), few studies allow both components of FA models to be time-varying (TV). Eickmeier et al. (2015) propose a TV-FAVAR model with parameters following random walks. More recently, Fu et al. (2024) developed a framework with smoothly evolving parameters and derived its asymptotic properties, while Wang et al. (2025) examined estimation and testing in TV panels with TV factor structures.

A common thread in these studies is the assumption that all latent factors driving the predictors are relevant to the forecast target. Yet, Kelly and Pruitt (2015) and Dendramis et al. (2023) argue that only a subset of factors may be influential. To address this issue, Dendramis et al. (2023) introduce a TV three-pass regression filter that dynamically selects only the relevant factors for forecasting, independent of the total number of common factors.

Motivated by revisiting an empirical study using a large dataset of U.S. macroeconomic variables, this article introduces a novel penalized TV-FA forecasting model that accommodates smoothly evolving factor loadings and regression coefficients. A key innovation of our approach is that it relaxes the common and often restrictive assumption that all latent factors are relevant to the forecast target, allowing instead for the selection of a sparse subset of influential factors. Our estimation procedure consists of two stages. In the first stage, we employ the local principal component analysis (LPCA) method of Su and Wang (2017) to consistently estimate the common factors in a TV framework. In the second stage, we augment a TV regression model with these estimated factors and apply a weighted group LASSO penalty within a local-linear smoothing framework. This enables the simultaneous estimation of time-varying parameters and consistent selection of the relevant factors. To ensure identifiability, we impose restrictions that permit asymptotically rotation-free estimation of both factors and loadings, following Cheung (2024). Also, we establish the asymptotic properties of our estimators, demonstrating the consistency of all estimators and the sparsity and oracle property of the second-stage estimator. Simulations indicate the good performance of our estimators, in comparison with the competitive estimators of Fu et al. (2024). Finally, in an empirical application using a large dataset of U.S. macroeconomic variables, we demonstrate that our penalized TV-FA model achieves superior forecasting performance for several key variables compared to the conventional TV-FAVAR approach of Fu et al. (2024).

The remainder of this article is organized as follows. Section 2 introduces the TV-FA forecasting model and its identification restrictions. Section 3 details the two-stage estimation procedure for simultaneous factor selection and parameter estimation. Section 4 establishes the asymptotic properties of the estimators, including consistency, sparsity, and the oracle property. Section 5 presents the computational algorithm for shrinkage estimation and the criteria for selecting tuning parameters. The finite-sample performance of our estimators is evaluated via simulations in Section 6, and Section 7 is devoted to an empirical application. Section 8 concludes the paper. All proofs are contained in the online supplement.

Notation. For an $m \times n$ real matrix A, we denote its transpose as $A^\top$, its Frobenius norm as $\|A\|$ ($\equiv [tr(AA^\top)]^{1/2}$), its spectral norm as $\|A\|_{sp}$ ($\equiv \sqrt{\mu_1(A^\top A)}$), and its Moore-Penrose generalized inverse as A^+, where $\equiv$ means "is defined as" and $\mu_s(\cdot)$ denotes the sth largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. We will frequently use the submultiplicative property of these norms and the fact that $\|A\|_{sp} \leq \|A\| \leq \|A\|_{sp} rank(A)^{1/2}$. 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 $P_A \equiv A(A^\top A)^+ A^\top$ and $M_A \equiv I_m - P_A$, where I_m denotes an $m \times m$ identity matrix. Let $0_{m\times n}$ be a $m \times n$ zero matrix and $[n] \equiv \{1, 2, ..., n\}$. The operator $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, and $p\lim$ probability limit. We use $(N, T) \to \infty$ to denote that N and T pass to infinity jointly.

2 Model Setup

In this section we first introduce the model and consider some identification conditions for the estimation of the factors and loadings.

2.1 Model

Let $y = (y_1, \ldots, y_T)^\top$ be the $T \times 1$ vector of target variable time series. Let $X = (x_1, \ldots, x_N)$ be the $T \times N$ matrix of predictors, with $x_i = (x_{i1}, \ldots, x_{iT})^\top$. Assume that the data are generated as follows

$$y_{t+1} = \beta_{0,t} + \beta_{f,t}^\top f_t + e_{y,t+1} \quad \text{and} \quad x_{it} = \phi_{f,it}^\top f_t + \phi_{g,it}^\top g_t + e_{x,it} \equiv \phi_{it}^\top F_t + e_{x,it}, \tag{2.1}$$

where $i \in [N]$, $t \in [T]$, $\phi_{f,it} = (\phi_{1it}, \ldots, \phi_{R_f it})^\top$, $\phi_{g,it} = (\phi_{1it}, \ldots, \phi_{R_g it})^\top$, f_t and g_t are $R_f \times 1$ and $R_g \times 1$ vectors of relevant and irrelevant factors in the prediction of y_{t+1}, respectively, $e_{y,t}$ and $e_{x,t} = (e_{x,1t}, \ldots, e_{x,Nt})^\top$ are zero mean error terms such that $E(e_{y,t+1}|F_t) = 0$ and $E(e^2_{y,t+1}|F_t) = \sigma^2_t(F_t)$, where $F_t = (f_t^\top, g_t^\top)^\top$ is an $R \times 1$ vector of common factors with $R = R_f + R_g$. In addition, let $\phi_{f,t} = (\phi_{f,1t}, \ldots, \phi_{f,Nt})^\top$ and $\phi_{g,t} = (\phi_{g,1t}, \ldots, \phi_{g,Nt})^\top$. Finally, denote $\beta_{0,t}$ and $\beta_{f,t} = (\beta_{f,1t}, \ldots, \beta_{f,R_f t})^\top$ as the target factor loadings. Then, we may write (2.1) as

$$y_{t+1} = \beta_t^\top F_t^a + e_{y,t+1} \quad \text{and} \quad x_{it} = \phi_{it}^\top F_t + e_{x,it}, \tag{2.2}$$

where $\beta_t = (1, \beta_{f,t}^\top, 0_{R_g \times 1}^\top)^\top$, $F_t^a = (1, F_t^\top)^\top$ and $\phi_{it} = (\phi_{f,it}^\top, \phi_{g,it}^\top)^\top$. We allow the errors $\{e_{y,t+1}\}$ in the y-equation in (2.2) to be conditionally heteroscedastic, so that it may depend on F_t and time index t. Then, following Cai (2007), we denote $\varepsilon_{y,t+1} = e_{y,t+1}/\sigma_t(F_t)$.

Following the nonparametric literature on TV models (e.g., Robinson (1989), Robinson (1991), Cai (2007), and Su and Wang (2017)), we model β_t and ϕ_{it} as nonrandom functions of t/T:

$$\beta_t = \beta(t/T) = (\beta_0(t/T), \beta_1(t/T), \ldots, \beta_R(t/T))^\top \text{ and } \phi_{it} = \phi_i(t/T),$$

where $\beta_l(\cdot)$ and $\phi_i(\cdot)$ are unknown piecewise smooth functions on $[0, 1]$. An intuitive explanation to this requirement is that the increasingly intensive sampling of data points ensures consistent estimation of $\beta(t/T)$ and $\phi_i(t/T)$ for each i at some fixed point t/T when the sample sizes increase. This intensity assumption is also applied to $\{\varepsilon_{y,t}\}$ by assuming that $\sigma_t(F_t) = \sigma(F_t, t/T)$. Importantly, we assume that $\beta_l(\cdot) \neq 0$ on a nontrivial subset of its support for any $l \leq R_f$ and $\beta_l(\cdot) = 0$ for any $l > R_f$. In other words, we assume that the first R_f factors in F_t are truly relevant but the rest are not. This paper aims to identify the relevant factors in F_t and perform forecasting simultaneously.

2.2 Identification Conditions

It is well known that latent common factors and factor loadings are not separately identifiable even when one assumes that $E(F_tF_t^\top) = \Sigma_F$, which is not time-varying. Noting that

$$\phi_{it}^\top F_t = (H_t^{-1}\phi_{it})^\top H_t^\top F_t$$

for any $R \times R$ nonsingular matrix $H_t \equiv H(t/T)$. Without further restrictions, F_t and ϕ_{it} can only be identified up to certain rotation matrix H_t. As discussed in Cheung (2024), the continuity of $H(\cdot)$ can be violated under certain scenario with $R > 1$, which in turn causes the discontinuity of $H_t^{-1}\phi_{it}$ and subsequently undermines the performance of forecasting. To deal with this problem, by letting $\Phi_t = (\phi_{1t}, \dots, \phi_{Nt})^\top$ and following Cheung (2024), we consider the following set of identification restrictions:

$$\text{(IR) } E(F_tF_t^\top) = \Sigma_F = I_R \text{ and } \Phi_t = (\Phi_{1:R,t}^\top, \Phi_{R+1:N,t}^\top)^\top,$$
$$\text{where } \Phi_{1:R,t} \text{ is a lower triangular matrix with positive diagonal elements.} \tag{2.3}$$

The above identification restrictions correspond to IR2 in Cheung (2024), which helps to pin down the rotation matrix $H(\cdot)$ and improves multi-step-ahead forecasts. Indeed, (IR) assumes that the common factors are weakly stationary with the second moment being the identity matrix, which is much weaker than the restriction that $\frac{1}{T}F'F = I_R$ in Bai and Ng (2013), where I_R is the $R \times R$ identity matrix. Furthermore, (IR) is analogous to a triangular system of simultaneous equations. The choice of the first R variables of x_t and their ordering provide the auxiliary information for identification.

The above identification conditions are imposed on the population model. That is, we assume the data is generated under such conditions perhaps after renormalizing the factors and loadings. In Section 4.2, we will show that under (IR), by imposing specific transformation matrix on the estimators of factors and loadings, the transformed rotation matrix asymptotically converges to the identity matrix I_R. Thus, we may write (2.2) as follows:

$$y_{t+1} = \beta_t^\top \check{F}_t^a + \beta_t^\top (F_t^a - \check{F}_t^a) + e_{y,t+1} = \beta_t^\top \check{F}_t^a + w_{y,t+1} \quad \text{and} \quad x_{it} = \phi_{it}^\top F_t + e_{x,it}, \tag{2.4}$$

where $\check{F}_t^a = (1, \check{F}_t^\top)^\top$, $\check{F}_t$ is an estimator of F_t satisfying (IR), and $w_{y,t+1} = \beta_t^\top (F_t^a - \check{F}_t^a) + e_{y,t+1}$ is the new error term.

3 Estimation Procedures

In this section, we consider a two-stage approach to forecast y_{t+1}. In the first stage, we estimate F_t via the local PCA method of Su and Wang (2017) but using the restriction (IR) in (2.3). In the second stage, we consider variable selection based on the estimated factor by a penalized local linear method with weighted group LASSO proposed in Li et al. (2015) and applied in Chen et al. (2025). Since R can be estimated consistently via the information criterion proposed by Su and Wang (2017), we assume R is known in the following estimation procedure.

3.1 Stage 1: Factor Estimation

At Stage 1, we obtain the estimator of F_t by using the local PCA approach as proposed by Su and Wang (2017). Specifically, let $k_{h_0,tr} = h_0^{-1} K((t-r)/(Th_0))$ where $K(\cdot)$ is a nonnegative kernel function and h_0 is a bandwidth. Define the $T \times N$ matrices $X^{(r)} = (x_1^{(r)}, \ldots, x_N^{(r)})$, where $x_i^{(r)} = (k_{h_0,1r}^{1/2} x_{i1}, \ldots, k_{h_0,Tr}^{1/2} x_{iT})^\top$. Let $F^{(r)} = (k_{h_0,1r}^{1/2} F_1, \ldots, k_{h_0,Tr}^{1/2} F_T)^\top$ and $e_x^{(r)} = (e_{x,1}^{(r)}, \ldots, e_{x,N}^{(r)})$, where $e_{x,i}^{(r)} = (k_{h_0,1r}^{1/2} e_{x,i1}, \ldots, k_{h_0,Tr}^{1/2} e_{x,iT})^\top$. Let $\Phi_t = (\phi_{1t}, \ldots, \phi_{Nt})^\top$. The local PCA estimators can be obtained via the following minimization problem:

$$\min_{F^{(r)}, \Phi_r} tr\left[\left(X^{(r)} - F^{(r)}\Phi_r^\top\right)\left(X^{(r)} - F^{(r)}\Phi_r^\top\right)^\top\right], \tag{3.1}$$
$$\text{s.t.} \quad F^{(r)\top} F^{(r)}/T = I_R \ \text{ and } \ \Phi_r^\top \Phi_r = \text{diagonal matirx}.$$

Then, the estimated factor matrix, denoted by $\tilde{F}^{(r)} = (\tilde{F}_1^{(r)}, \ldots, \tilde{F}_T^{(r)})^\top$, is $\sqrt{T}$ times eigenvectors corresponding to the R largest eigenvalues of the $T \times T$ matrix $X^{(r)} X^{(r)\top}$, arranged in descending order, and $\tilde{\Phi}_r^\top = (\tilde{F}^{(r)} \tilde{F}^{(r)\top})^{-1} \tilde{F}^{(r)\top} X^{(r)} = \tilde{F}^{(r)\top} X^{(r)}/T$ for $r \in [T]$, are the estimators of the corresponding TV factor loadings. We use $\tilde{\phi}_{ir}$ to denote the ith column of $\tilde{\Phi}_r^\top$. Next, based on the consistent estimators $\tilde{\phi}_{it}$, we can obtain the consistent estimator $\tilde{F}_t$ of F_t by considering the following least squares problem:

$$\min_{F_t \in \mathbb{R}^R} \sum_{i=1}^N \left[x_{it} - \tilde{\phi}_{it}^\top F_t \right]^2, \quad t \in [T].$$

Note that $\tilde{F}_t = \left(\sum_{i=1}^N \tilde{\phi}_{it} \tilde{\phi}_{it}^\top \right)^{-1} \left(\sum_{i=1}^N \tilde{\phi}_{it} x_{it} \right)$ for $t \in [T]$. As derived in Su and Wang (2017), $\tilde{F}_t$ and $\tilde{\phi}_{it}$ are consistent estimators of $H_t^\top F_t$ and $H_t^{-1} \phi_{it}$, respectively, for some rotation matrix H_t.

To ensure that the identification restrictions given in (2.3) are satisfied by the estimators, we conduct the QR decomposition of the top $R \times R$ submatrix of $\tilde{\Phi}_t$, namely, $\tilde{\Phi}_{1:R,t}^\top = \tilde{Q}_t^\top \tilde{R}_t$, where $\tilde{R}_t$ is upper triangular with positive diagonal elements and $\tilde{Q}_t \tilde{Q}_t^\top = I_R$. Define the rotated local PCA estimators F_t and Φ_t respectively as

$$\hat{F}_t = \tilde{Q}_t \tilde{F}_t \ \text{ and } \ \hat{\Phi}_t = \tilde{\Phi}_t \tilde{Q}_t^{-1}. \tag{3.2}$$

The transformed rotation matrix becomes $H_t^* = H_t \tilde{Q}_t^\top$. One can verify that

$$\hat{\Phi}_t = \begin{pmatrix} \hat{\Phi}_{1:R,t} \\ \hat{\Phi}_{(R+1):N,t} \end{pmatrix} = \begin{pmatrix} \tilde{\Phi}_{1:R,t} \tilde{Q}_t^{-1} \\ \tilde{\Phi}_{(R+1):N,t} \tilde{Q}_t^{-1} \end{pmatrix} = \begin{pmatrix} \tilde{R}_t^\top \\ \tilde{\Phi}_{(R+1):N,t} \tilde{Q}_t^{-1} \end{pmatrix}.$$

By construction, $\hat{\Phi}_{1:R,t}$ is a lower triangular matrix with positive diagonal elements, satisfying the conditions on Φ_t in (2.3).

Notice that the local PCA method involves a local constant estimation for time-varying factor model, which may suffer from boundary bias problem when estimating a fixed conditional mean object; see, for example, Su and Wang (2017) for detailed discussions on this point. To deal with this issue, we follow Hong and Li (2005) and Li and Racine (2007) to apply the following boundary kernel:

$$k^*_{h_0,tr} = \begin{cases} h_0^{-1}K((t-r)/Th_0)/\int_{-r/(Th_0)}^{1} K(v)dv & \text{if } r \in [1, \lfloor Th_0 \rfloor) \\ h_0^{-1}K((t-r)/Th_0) & \text{if } r \in [\lfloor Th_0 \rfloor, T - \lfloor Th_0 \rfloor] \\ h_0^{-1}K((t-r)/Th_0)/\int_{-1}^{(1-r/T)/h_0} K(v)dv & \text{if } r \in (T - \lfloor Th_0 \rfloor, T]. \end{cases}$$

We replace $k_{h_0,tr}$ in (3.1) with this boundary kernel and use the notation $k^*_{h_0,tr}$ instead of $k_{h_0,tr}$ hereafter. For more about the boundary kernels, the reader is referred to the paper by Gasser and Müller (1979). An alternative way to overcome this difficulty is to use the so-called reflection approach proposed by Hall and Wehrly (1991).

3.2 Stage 2: Factor Selection

At this stage, we conduct factor selection for the prediction of y_{t+1}. We replace the generic estimator $\check{F}_t$ in (2.4) with the estimator $\hat{F}_t$ obtained at the end of Stage 1. Let $z_t = t/T$ and $\beta_t = \beta(z_t) = (\beta_0(z_t), \beta_1(z_t), \ldots, \beta_R(z_t))^\top$. For any given grid point $z_s \in [0,1]$, when z_t is in a neighborhood of z_s, $\beta(z_t)$ can be approximated by a polynomial function as $\beta(z_t) \approx \sum_{r=0}^{\varsigma} \beta^{(r)}(z_s)(z_t - z_s)^r/r!$, where $\approx$ denotes the approximation by ignoring the higher order terms and $\beta^{(r)}(\cdot)$ denotes the rth derivative of $\beta(\cdot)$. Let $\theta_r^0(\cdot) = \beta^{(r)}(\cdot)/r! = (\theta^0_{r,0}(\cdot), \theta^0_{r,1}(\cdot), \ldots, \theta^0_{r,R}(\cdot))^\top$ and $\theta^0_{r,l} = (\theta^0_{r,l}(z_1), \ldots, \theta^0_{r,l}(z_T))^\top$ be the true values. Note that identifying the zero coefficients among $\theta^0_{1,l}$ is equivalent to identifying the indices $l \in 0 \cup [R]$ such that $D_l = 0$, where $D_l = \left\{ \sum_{t_1=1}^T [\theta^0_{0,l}(z_{t_1}) - \frac{1}{T}\sum_{s_1=1}^T \theta^0_{0,l}(z_{s_1})]^2 \right\}^{1/2}$. In practice, we approximate D_l by $\tilde{D}_l = \left\{ \sum_{t_1=1}^T [\tilde{\theta}_{0,l}(z_{t_1}) - \frac{1}{T}\sum_{s_1=1}^T \tilde{\theta}_{0,l}(z_{s_1})]^2 \right\}^{1/2}$, where $\tilde{\theta}_{0,l}(\cdot)$ is the unpenalized local linear estimator of $\theta^0_{0,l}(\cdot)$. Let $\tilde{\theta}_{0,l} = (\tilde{\theta}_{0,l}(z_1), \ldots, \tilde{\theta}_{0,l}(z_T))^\top$ and $\hat{F}^a_t = (1, \hat{F}_t^\top)^\top$.

Denote $\Theta_0 = (\theta_{0,0}, \theta_{0,1}, \ldots, \theta_{0,R}) \in \mathbb{R}^{T\times(R+1)}$ and $\Theta_1 = (\theta_{1,0}, \theta_{1,1}, \ldots, \theta_{1,R}) \in \mathbb{R}^{T\times(R+1)}$.

We propose the following weighted group LASSO estimate:

$$\begin{pmatrix} \hat{\Theta}_0 \\ \hat{\Theta}_1 \end{pmatrix} \equiv \underset{\Theta_0,\Theta_1 \in \mathbb{R}^{T\times(R+1)}}{\operatorname{argmin}} Q_\lambda(\Theta_0,\Theta_1), \tag{3.3}$$

where

$$\begin{aligned} Q_\lambda(\Theta_0,\Theta_1) =& \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{y_{t+1} - \hat{F}_t^{a\top}[\theta_0(z_s)+\theta_1(z_s)(z_t-z_s)]\right\}^2 K_{h_1}(z_t-z_s) \\ &+ \sum_{l=0}^{R} P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)\|\theta_{0,l}\| + \sum_{l=0}^{R} P'_{\lambda_T}(\tilde{D}_l)\|h_1\theta_{1,l}\| \end{aligned}$$

with the penalty terms suggested by Li et al. (2015) and Chen et al. (2025). Here, $K_h(\cdot) = h^{-1}K(\cdot/h)$, $h_1 = h_1(n)$ is a bandwidth sequence, $\lambda_T \in R$ is the tuning parameter, and $P'_{\lambda_T}(\cdot)$ is the derivative of the SCAD penalty function:

$$P'_\lambda(z) = \begin{cases} \lambda & \text{if } z \le \lambda \\ (a\lambda - z)/(a-1) & \text{if } \lambda < z \le a\lambda \\ 0 & \text{if } z > a\lambda. \end{cases} \tag{3.4}$$

Fan and Li (2001) suggests choosing $a = 3.7$. The weights $P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)$ and $P'_{\lambda_T}(\tilde{D}_l)$ for the group LASSO in (3.3) are derived from the local linear approximation to the SCAD penalty function as in Zou and Li (2008). It is worthwhile to note that we use $\tilde{D}_l$ instead of $h_1\tilde{\theta}_{1,l}$ to construct the weight of the second penalty function. Indeed, as suggested in Chen et al. (2025), the estimation of the time-varying coefficients in $\tilde{D}_l$ often exhibits greater stability than that of $h_1\tilde{\theta}_{1,l}$. Below, we write $T\times(R+1)$ matrix $\hat{\Theta}_r$ as

$$\hat{\Theta}_r = (\hat{\theta}_r(z_1),\ldots,\hat{\theta}_r(z_T))^\top \text{ for } r = 0,1,$$

where $\hat{\theta}_r(z) = (\hat{\theta}_{r,0}(z),\hat{\theta}_{r,1}(z),\ldots,\hat{\theta}_{r,R}(z))^\top \in \mathbb{R}^{R+1}$. Finally, denote the final estimator as

$$\hat{\beta}_\lambda(z_s) = (\hat{\beta}_{\lambda,0}(z_s),\hat{\beta}_{\lambda,1}(z_s),\ldots,\hat{\beta}_{\lambda,R}(z_s))^\top = (\hat{\theta}_{0,0}(z_s),\hat{\theta}_{0,1}(z_s),\ldots,\hat{\theta}_{0,R}(z_s))^\top = \hat{\theta}_0(z_s). \tag{3.5}$$

Let $\hat{\theta}_{r,l} = (\hat{\theta}_{r,l}(z_1), \ldots, \hat{\theta}_{r,l}(z_T))^\top \in \mathbb{R}^T$ for $r = 0, 1$ and $\hat{\beta}_{\lambda,l} = \hat{\theta}_{0,l}$.

Remark 3.1. *As in Cheung (2024), the IR condition in Section 2.1 can always be fulfilled by renormalization. To see this, notice that IR assumes that the first $R \times R$ submatrix of Φ_t to be lower triangular. If this condition is not satisfied, we can still write*

$$x_t = \begin{pmatrix} \Phi_{1:R,t} \\ \Phi_{R+1:N,t} \end{pmatrix} F_t + e_{x,t} = \begin{pmatrix} R_t^\top \\ \Phi_{(R+1):N,t} Q_t^\top \end{pmatrix} Q_t F_t + e_{x,t}$$

where $\Phi_{1:R,t}^\top = Q_t^\top R_t$ is the QR decomposition of $\Phi_{1:R,t}$. If $E(F_t F_t^\top) = I_R$, then $E(Q_t^\top F_t F_t^\top Q_t)$ $= Q_t^\top Q_t = I_R$ and all conditions in IR are satisfied. In this case, we will be estimating $\Phi_t Q_t^\top$ and $Q_t F_t$ instead of Φ_t and F_t if we apply the IR rotations. When $E(F_t F_t^\top) = \Sigma \neq I_R$, we need to normalize F_t before the above renormalization:

$$\Phi_t F_t = (\Phi_t \Sigma^{1/2})(\Sigma^{-1/2} F_t) = \Phi_t^* F_t^*.$$

Now we can conduct the QR decomposition on the first $R \times R$ submatrix of Φ_t^* as above and $F_t^* = \Sigma^{-1/2} F_t$ satisfies $E(F_t^* F_t^{*\top}) = I_R$.

4 Limiting Distributions

In this section we first state some basic assumptions and then study the asymptotic properties of the two-stage estimators introduced in the last section.

4.1 Basic Assumptions

To proceed, we introduce some notations. Let $e_{x,t} = (e_{x,1t}, \ldots, e_{x,Nt})^\top$, $\gamma_N(s,t) = N^{-1} E(e_{x,s}^\top e_{x,t})$, $\gamma_{N,F}(s,t) = N^{-1} E(F_s e_{x,s}^\top e_{x,t})$, $\gamma_{N,FF}(s,t) = N^{-1} E(F_s e_{x,s}^\top e_{x,t} F_t^\top)$, and $\zeta_{st} = N^{-1}[e_{x,s}^\top e_{x,t} - E(e_{x,s}^\top e_{x,t})]$. Define

$$\varpi_{NT,1}(r) = \frac{h^{1/2}}{\sqrt{NT}} F^{(r)\top} e_x^{(r)} \phi_r = \frac{h^{1/2}}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} k_{h,tr}^* F_t e_{x,it} \phi_{ir}^\top, \text{ and}$$

$$\varpi_{NT,2}(r,t) = \frac{h^{1/2}}{\sqrt{NT}}[F^{(r)\top}e_x^{(r)}e_{x,t} - E(F^{(r)\top}e_x^{(r)}e_{x,t})]$$
$$= \frac{h^{1/2}}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}k_{h,sr}^{*}[F_s e_{x,is}e_{x,it} - E(F_s e_{x,is}e_{x,it})].$$

Let $C < \infty$ denote a generic constant that may vary across occurrences. We make the following assumptions.

Assumption 4.1. *(i)* $E(e_{x,it}) = 0$ *and* $\max_{i,t} E(e_{x,it}^8) < \infty$.

(ii) $\max_t E\|F_t\|^8 < \infty$ *and* $\Sigma_F \equiv E(F_t F_t^\top) = I_R$.

(iii) $\phi_i(\cdot)$ *is nonrandom and third-order continuously differentiable such that* $\max_{i\in[N]} \sup_{z\in[0,1]} \|\phi_i^{(c)}(z)\| \leq \bar{c}_\phi < \infty$ *for* $c = 0 \cup [3]$. *In addition,* $N^{-1}\Phi_r^\top\Phi_r = \Sigma_{\Phi_r} + O(N^{-1/2})$ *for some* $R\times R$ *positive definite matrix* Σ_{Φ_r} *and for all* $r \in [T]$.

(iv) $\max_t \sum_{s=1}^{T} |Cov(F_{l_1 t}F_{l_2 t}, F_{l_1 s}F_{l_2 s})| \leq C$ *for* $l_1,\ l_2 \in [R]$, *where* $F_{l_1 t}$ *is the* l_1*th element of* F_t.

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

(vi) $\max_{1\leq s,t\leq T} E|N^{1/2}\zeta_{st}|^4 \leq C$ *and* $\max_{r,t} E\|N^{-1/2}\Phi_r^\top e_{x,t}\|^4 \leq C$.

(vii) $\varpi_{NT,1}(r) = O_p(1)$ *and* $\max_t E\|\varpi_{NT,2}(r,t)\|^2 \leq C$ *for each* r.

(viii) For all $r \in [T]$, *the eigenvalues of the* $R\times R$ *matrix* $\Sigma_{\Phi_r}^{1/2}\Sigma_F\Sigma_{\Phi_r}^{1/2}$ *are distinct.*

Assumption 4.2. *(i)* $\|e_x\|_{sp} = O_p(N^{1/2}+T^{1/2})$ *and* $\max_{t,s} ||N^{-1}\sum_{i=1}^{N}[e_{x,it}e_{x,is} - E(e_{x,it}e_{x,is})]|| = O_p(N^{-1/2}(\ln T)^{1/2})$.

(ii) $\max_r \|N^{-1}\Phi_r^\top\Phi_r - \Delta_{\Phi_r}\| = o(1)$, *and the eigenvalues of* Δ_{Φ_r} *are bounded below from* 0 *and above from infinity uniformly in* r.

(iii) $\max_{i,r}\left\|\frac{1}{T}\sum_{t=1}^{T}k_{h,tr}^{*}F_t e_{x,it}\right\| = O_p((Th_0)^{-1/2}(\ln(NT))^{1/2})$, $\max_r\left\|\frac{1}{T}\sum_{t=1}^{T}k_{h,tr}^{*}(F_tF_t^\top - \Sigma_F)\right\| = O_p(T^{-1/2}(\ln T)^{1/2})$ *and* $\max_r\left\|\frac{1}{T}\sum_{t=1}^{T}k_{h,tr}^{*}(\|F_t\| - E\|F_t\|)\right\| = O_p(T^{-1/2}(\ln T)^{1/2})$.

(iv) $\max_{s,t}\left\|N^{-1}\Phi_s^\top e_{x,t}\right\| = O_p(N^{-1/2}(\ln T)^{1/2})$ *and* $\max_s\left\|N^{-1}\Phi_s^\top e_{x,s}F_s'\right\| = O_p(N^{-1/2}(\ln T)^{1/2})$.

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

(vi) $\max_r \left\| \frac{1}{NT} \sum_{t=1}^{T} \Phi_r^\top e_{x,t}^{(r)} F_t^{(r)} \right\| = \max_r \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} k_{h,tr}^* \phi_{ir} e_{x,it} F_t^\top \right\| = O_p((NTh_0)^{-1/2}$ $(\ln T)^{1/2})$.

Assumption 4.3. *(i) The kernel function $K(\cdot)$ is a symmetric continuously differentiable probability density function (PDF) with compact support $[-1, 1]$. Let $\mu_{ij} = \int v^i K^j(v)dv$ for $i, j = 0, 1, 2$.*

(ii) As $(N, T) \to \infty$, $h_0 \to 0$, $Th_0^7 \to 0$, $Nh_0^6 \to 0$, $Th_0/N \to 0$, $Nh_0/T \to 0$, $Th_0^2 \to \infty$, $Nh_0^2 \to \infty$ and $Th_0/N^{1/2} \to \infty$.

Assumption 4.1 are widely applied in the literature; see, e.g., Bai and Ng (2002), Bai (2003) and Su and Wang (2017). In particular, Assumption 4.1(i) allows $e_{x,it}$ to be both unconditionally heteroscedastic over (i, t) and conditionally heteroscedastic given F_t. Assumption 4.1(ii) corresponds to that in Su and Wang (2017) except we further require $\Sigma_F = I_R$. The time-invariance of Σ_F in Assumption 4.1(ii) is not as restrictive as it appears. For instance, if $E(F_t F_t^\top) = \Sigma_{F,t}$, we could rewrite the common component as

$$\phi_{it}^\top F_t = \left(\Sigma_F^{-1/2} \Sigma_{F,t}^{1/2} \phi_{it} \right)^\top \Sigma_F^{1/2} \Sigma_{F,t}^{-1/2} F_t = \phi_{it}^{*\top} F_t^*,$$

where $\phi_{it}^* = \Sigma_F^{-1/2} \Sigma_{F,t}^{1/2} \phi_{it}$ and $F_t^* = \Sigma_F^{1/2} \Sigma_{F,t}^{-1/2} F_t$ satisfies $E(F_t^* F_t^{*\top}) = \Sigma_F$ by construction. Similarly, if $E(F_t F_t^\top) = \Sigma_F$, we can redefine the factor as $\Sigma_F^{-1/2} F_t$, which will have second moment I_R. Assumption 4.1(iii) mainly imposes some smoothness conditions on $\phi_i(\cdot)$ explicitly. Of course, it is possible to relax the third-order continuous differentiability in 4.1 4.1(iii) by allowing for Hölder continuity as in Cheung (2024). However, this would require more complex notation and could slow the convergence rates of both the common factors and loadings. Assumption 4.1(iv)-(vi) restricts the time and cross-sectional dependence for the idiosyncratic errors and the weak dependence between the factors and errors, which are in the same spirit as Assumptions C.2–C.5, D and E in Bai (2003). Assumption 4.1(vii) is a kernel-weighted version of Assumptions F.1–F.2 in Bai (2003).

Assumption 4.2 parallels Assumption A.4 in Su and Wang (2017) and can be verified under some primitive conditions that are used in the factor literature. For example, Moon and Weidner (2017) demonstrate that the first part of Assumption 4.2(i) can be satisfied for various processes and Su and Wang (2017) verify similar conditions to those in 4.2(iii)-(vi) under some mixing conditions. Assumption 4.2(ii) strengthens 4.1(iii) as we now need $N^{-1}\Phi_r^\top \Phi_r$ and its limit to be well behaved uniformly in r. Assumption 4.2(iii)-(vi) imposes conditions on the uniform probability order of some summation objects. The first part of Assumption 4.2(iii) corresponds to and extends Assumption D in Bai and Ng (2002), because we need the condition to hold uniformity in (i, r). The second and third parts of Assumption 4.2(iii) are also imposed in Su and Wang (2017) and can be verified under some weak data dependence conditions and moment conditions. Similarly, one can also verify the conditions in Assumption 4.2(iv)-(vi). See the online supplementary appendix of Su and Wang (2017) for the verification of 4.2(vi) under some primitive conditions. Assumption 4.3 imposes some regularity conditions on the kernel function and bandwidth, similar to Assumption A.1*(ii) in Su and Wang (2020a).

4.2 Consistency of Transformed Rotation Matrix

In this section, we study the asymptotic properties of Stage 1 estimators.

Theorem 4.1. *Suppose that Assumptions 4.1–4.3 hold. Then, under (IR) in (2.3), we have* $\max_t \|H_t^* - I_R\| = O_p((Th_0/\ln T)^{-1/2} + h_0^2 + N^{-1}\ln T)$.

The above theorem states that the transformed rotation matrix converges to the identity matrix uniformly in t under some conditions. Then the rotated local PCA estimators are consistent with the true values without any rotations. Recall that $H_t^* = H_t \tilde{Q}_t^\top$ and $\hat{F}_t = \tilde{Q}_t \tilde{F}_t$. Noting that

$$\sqrt{N}(\hat{F}_t - F_t) = \sqrt{N}\tilde{Q}_t(\tilde{F}_t - H_t^\top F_t) + \sqrt{N/(Th_0)}\sqrt{Th_0}(H_t^* - I_R)^\top F_t, \tag{4.1}$$

we observe that the second term in the above decomposition reflects the contribution from H_t^*, which is asymptotically nonnegligible and affects the asymptotic variance of $\hat{F}_t$ under our assumptions. Therefore, to derive the asymptotic distribution of $\hat{F}_t$, we also need to consider that of the transformed rotation matrix H_t^*. The following corollary reports the uniform convergence rate of $\hat{F}_t$.

Corollary 4.2. *Suppose that Assumptions 4.1–4.3 hold. Then, under (IR) in (2.3), we have* $\max_t \|\hat{F}_t - F_t\| = O_p((N/\ln T)^{-1/2} + h_0^2) + o_p((Th_0/\ln T)^{-1/2}T^{1/8} + h_0^2 T^{1/8})$.

4.3 Asymptotic Normality

Let

$$H_r = (N^{-1}\Phi_r^\top \Phi_r)(T^{-1}F^{(r)\top}\tilde{F}^{(r)})V_{NT}^{(r)-1}, \tag{4.2}$$

where $V_{NT}^{(r)}$ is an $R\times R$ diagonal matrix containing the first R largest eigenvalues of $(NT)^{-1}X^{(r)} \times X^{(r)\top}$, ordered in descending magnitude along its main diagonal. Let V_r be a diagonal matrix consisting of the eigenvalues of $\Sigma_{\Phi_r}^{1/2}\Sigma_F\Sigma_{\Phi_r}^{1/2}$ in descending order with Υ_r being the corresponding (normalized) eigenvector matrix ($\Upsilon_r^\top\Upsilon_r = I_R$), and $P_r = V_r^{1/2}\Upsilon_r^{-1}\Sigma_{\Phi_r}^{-1/2}$. Let $\tilde{\Sigma}_F^{(r)} \equiv T^{-1}F^{(r)\top}F^{(r)}$. We add the following assumption.

Assumption 4.4. *(i)* $N^{-1/2}\Phi_r^\top e_{x,t} \xrightarrow{d} G_{r,t} \sim N(0,\Gamma_{rt})$ *for each* r, t, *where* $\Gamma_{rt} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N\sum_{j=1}^N \phi_{ir}\phi_{jr}^\top E(e_{x,it}e_{x,jt})$.

(ii) $Z_i^{(r)} \equiv \sqrt{h_0/T}\sum_{s=1}^T k_{h,sr}^* F_s e_{x,is} \xrightarrow{d} \mathbb{Z}_{i,r} \sim N(0,\Omega_{i,r})$, *where*

$$\Omega_{i,r} = \lim_{T\to\infty}\left[\frac{h_0}{T}\sum_{s=1}^T k_{h,sr}^{*2}E(F_sF_s^\top e_{x,is}^2) + \frac{2h_0}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^T k_{h,sr}^*k_{h,tr}^*E(F_sF_t^\top e_{x,is}e_{x,it})\right].$$

(iii) $(Z_1^{(r)\top},\ldots,Z_R^{(r)\top}) \xrightarrow{d} (\mathbb{Z}_{1,r}^\top,\ldots,\mathbb{Z}_{R,r}^\top)$.

(iv) $\hat{\boldsymbol{N}}_r \equiv \sqrt{Th_0}(\tilde{\Sigma}_F^{(r)} - \Sigma_F) \xrightarrow{d} N_r$. *Here,* $vech(N_r) \sim N(0_{R(R+1)/2},\Psi_r)$ *for some positive definite matrix*

$$\Psi_r = \lim_{T\to\infty}\left[\frac{h_0}{T}\sum_{s=1}^T k_{h,sr}^{*2}E(\xi_s\xi_s^\top) + \frac{2h_0}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^T k_{h,sr}^*k_{h,tr}^*E(\xi_s\xi_t^\top)\right],$$

where $\xi_t = vech(F_tF_t^\top - \Sigma_F)$.

Assumption 4.4(i)-(ii) is standard; see, e.g., Su and Wang (2017). Both parts are used to establish the asymptotic normality of our local PCA estimators and can be verified under some primitive conditions. For example, by the central limit theorem (CLT hereafter) for strong mixing processes (e.g., Theorem 5.20 in White (2001)), one can readily verify Assumption 4.4(ii). Assumption 4.4(iii)-(iv) is also made by Cheung (2024). Assumption 4.4(iii) is needed for the derivation of the asymptotic distribution of H_t^* at Stage 1 under (IR). Assumption 4.4(iv) imposes the condition on an asymptotically non-negligible term in the asymptotic expansion of the transformed rotation matrix H_t^*. Since we assume that $E(F_tF_t^\top) = I_R$ instead of $F^\top F/T = I_R$ as in Bai and Ng (2013), the sample second moment of the common factors remains a random variable.

To study the asymptotic distribution of H_t^*, we add some notation. Let $\mathbb{Z}_{1:R,t} = (\mathbb{Z}_{1,t}, \ldots, \mathbb{Z}_{R,t})$, where $\mathbb{Z}_{i,t}$ is defined in Assumption 4.4(ii).

Theorem 4.3. *Suppose that Assumptions 4.1–4.4 hold. Let* $B_\Phi(1:R,t) = [B_\Phi(1,t), ..., B_\Phi(R,t)]^\top$, *where* $B_\Phi(i,t)$ *is defined in (??) in the online appendix. Then under (IR) in (2.3),* $\sqrt{Th_0}(H_t^* - I_R - H_t B_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1}) \xrightarrow{d} \Xi_t$, *where* $\Xi_t = \bar{\Delta}_t - 0.5N_t$ *and for any* $k_1, k_2 \in [R]$,

$$[\bar{\Delta}_t]_{k_1,k_2} = \begin{cases} \left[\mathbb{Z}_{1:R,t}(\Phi_{1:R,t}^\top)^{-1} + 0.5\boldsymbol{N}_t\right]_{k_1,k_2} & \text{if } k_1 > k_2 \\ 0 & \text{if } k_1 = k_2 \\ -(\bar{\Delta}_t)_{k_2,k_1} & \text{if } k_1 < k_2 \end{cases}$$

with $[A]_{k_1,k_2}$ *denoting the* (k_1,k_2)*th element of the matrix* A.

Theorem 4.3 studies the asymptotic distribution of H_t^*. Compared to the results in Cheung (2024), there is an asymptotic bias term $H_t B_\Phi^\top(1:R,t)$ that is of order $O_p(h_0^2)$, similar to the order of δ_t in Theorem 4.1 in Cheung (2024) under his Hölder continuity conditions. This bias is asymptotically vanishing if we impose the additional condition $Th_0^5 \to 0$.

Given the result in Theorem 4.3, we study the asymptotic distributions of $\hat{F}_t$ and $\hat{\phi}_{it}$ in the following theorem.

Theorem 4.4. *Suppose that Assumptions 4.1–4.4 hold and let* Ξ_t *be as defined in Theorem 4.3. Then, under (IR) in (2.3):*

(i) If $N(Th_0)^{-1} \to c_0 < \infty$ and $Nh_0^4 \to 0$, then $\sqrt{N}[\hat{F}_t - F_t] \xrightarrow{d} \Sigma_{\Phi_t}^{-1} G_{t,t} + \sqrt{c_0}\Xi_t^\top F_t$; If $N(Th_0)^{-1} \to \infty$ and $Th_0^5 \to 0$, then $\sqrt{Th_0}(\hat{F}_t - F_t) \xrightarrow{d} \Xi_t^\top F_t$.

(ii) If $Th_0^5 \to 0$, then $\sqrt{Th_0}(\hat{\phi}_{it} - \phi_{it}) \xrightarrow{d} \mathbb{Z}_{i,t} - \Xi_t \phi_{it}$.

Theorem 4.4(i) indicates that one can obtain the usual $\sqrt{N}$-rate of consistency for $\hat{F}_t$ only when $N(Th_0)^{-1} \to c_0 < \infty$. Unfortunately, the latter requirement rules out the case where N and T diverge to infinity at the same rate. On the other hand, if $N(Th_0)^{-1} \to \infty$ (which allows N and T to diverge to infinity at the same rate), the estimation error in H_t^* dominates the asymptotics so that we now only have $\sqrt{Th_0}$-rate of consistency for $\hat{F}_t$. Theorem 4.4(ii) implies that the estimator $\hat{\phi}_{it}$ enjoys the usual $\sqrt{Th_0}$-rate of consistency when we impose an undersmoothing condition ($Th_0^5 \to 0$) to eliminate the asymptotic bias of $\hat{\phi}_{it}$.

4.4 Theoretical Properties of Shrinkage Estimators

To proceed, we add the following assumptions.

Assumption 4.5. *(i) For $l \in 0 \cup [R_f]$, the second order derivative of coefficient $\beta_l(\cdot)$ exists and is continuous. In addition, $\sigma_t(F) = \sigma(F, z_t)$, where $\sigma(\cdot, \cdot)$ is a bounded and continuous function in $\mathbb{R}^R \times [0,1]$. Further, $|\sigma(F, z_t) - \sigma(F, z_s)| \leq M(F)|z_t - z_s|$ for some function $M(\cdot)$.*

(ii) The process $\{(\varepsilon_{y,t}, F_t) : t \geq 1\}$ is a strictly stationary and α-mixing process such that $E(\varepsilon_{y,t+1}|F_t) = 0$. There exists some $\delta > 0$ such that $E\|F_t\|^{2(2+\delta)} < \infty$ and $E|\varepsilon_{y,t} M(F_t)|^{2+\delta} < \infty$, and the mixing coefficient satisfies $\alpha(n) = O(n^{-\tau})$ with $\tau = (2+\delta)(1+\delta)/\delta$.

(iii) As $(N,T) \to \infty$, for some $c_1 \in [0, \infty)$, $Th_1^5 \to c_1$, $T(\ln T)h_1/N \to 0$ and $Th_1h_0^4 = o(T^{-1/4})$.

(iv) As $T \to \infty$, $\sqrt{h_1}\lambda_T \to 0$, $\lim_{T\to\infty} \inf_{\theta \to 0^+} P'_{\lambda_T}(\theta)/\lambda_T > 0$. In addition, $\min_{1 \leq l \leq R_f} \|\theta_{0,l}^0\| \geq (a+1)\lambda_T$ and $\min_{1 \leq l \leq R_f} D_l \geq (a+1)\lambda_T$, where a is defined in the SCAD function.

Assumption 4.5(i) imposes some smoothness conditions on the regression coefficients and conditional variance in the y-equation, which are commonly used in literature; see Robinson (1989), Robinson (1991) and Cai (2007). Assumption 4.5(ii) is a common condition with weakly dependent data and is satisfied for many stationary financial time series such as

ARMA, ARCH and GARCH processes (c.f., Cai (2002)). Assumption 4.5(iii) imposes some conditions on (h_0, h_1), similar to those in Assumption A.2(iii) of Fu et al. (2024), which guarantee that the estimation error $\hat{F}_t - F_t$ at Stage 1 does not contribute to the asymptotic distribution of $\hat{\beta}(\cdot)$. The first part of Assumption 4.5(iv) specifies the conditions on the regularization parameter λ_T. The second part of Assumption 4.5(iv) is also imposed by Chen et al. (2025), which ensures that with probability approaching one, the significant functionals and their derivatives can be detected.

To state the next asymptotic result, we add some notation. Let $\hat{F}^a_{a,t} = (1, \hat{F}_{1t}, \ldots, \hat{F}_{R_f t})^\top \in \mathbb{R}^{R_f+1}$. Similarly, let $F^a_{a,t} = (1, F_{1t}, \ldots, F_{R_f t})^\top \in \mathbb{R}^{R_f+1}$. In addition, we define $\hat{\beta}_{a,\lambda}(z) = (\hat{\beta}_{\lambda,0}(z), \hat{\beta}_{\lambda,1}(z), \ldots, \hat{\beta}_{\lambda,R_f}(z))^\top \in \mathbb{R}^{R_f+1}$ and $\hat{\beta}_{b,\lambda}(z) = (\hat{\beta}_{\lambda,R_f+1}(z), \ldots, \hat{\beta}_{\lambda,R}(z))^\top \in \mathbb{R}^{R-R_f}$. Also, let $\beta_a(z) = (\beta_0(z), \beta_1(z), \ldots, \beta_{R_f}(z))^\top$ and $\beta_b(z) = (\beta_{R_f+1}(z), \ldots, \beta_R(z))^\top$. Moreover, let $\boldsymbol{R}_{a,u}(z_t) = Cov(\sigma(F_t, z_t)\varepsilon_{y,t}F^a_{a,t}, \sigma(F_{t+u}, z_t)\varepsilon_{y,t+u}F^a_{a,t+u})$ and $\Sigma_{a,0}(z_t) = \sum_{u=-\infty}^{\infty} \boldsymbol{R}_{a,u}(z_t)$. Under Assumption 4.5(ii), we know that $\Sigma_{a,0}(z_t)$ exists by Lemma **??** in the online Appendix. Finally, denote $\Sigma^a_F = E(F^a_{a,t}F^{a\top}_{a,t})$ and $\Sigma_{a,\beta}(z_t) = (\Sigma^a_F)^{-1}\Sigma_{a,0}(z_t)(\Sigma^a_F)^{-1}$.

The following theorem establishes the sparsity and asymptotic normality of the penalized estimator in (3.5).

Theorem 4.5. *Suppose that Assumptions 4.1–4.5 hold and $R_g = R - R_f > 0$. Then we have*

(a) (Sparsity) $P(\sup_{z\in[0,1]} \|\hat{\beta}_{b,\lambda}(z)\| = 0) \to 1$.

(b) (Asymptotic Normality)

$$\sqrt{Th_1}\left(\hat{\beta}_{a,\lambda}(z) - \beta_a(z) - \frac{h_1^2\mu_{21}}{2}\beta_a^{(2)}(z) + o_p(h_1^2)\right) \xrightarrow{d} N(0, \mu_{02}\Sigma_{a,\beta}(z)),$$

where μ_{21} and μ_{02} are defined in Assumptions 4.3(i).

Remark 4.6. *Denote $\hat{F}^{a*}_{a,t} = (\hat{F}^{a\top}_{a,t}, ((z_t - z)/h_1)\hat{F}^{a\top}_{a,t})^\top$ and define the oracle estimator (i.e., the unpenalized estimator obtained under the true model) as*

$$\begin{aligned}\hat{\beta}^{oracle}(z) = \; & [I_{R_f+1}, \mathbf{0}_{(R_f+1)\times(R_f+1)}]\left(T^{-1}\sum_{t=1}^{T} \hat{F}^{a*}_{a,t}\hat{F}^{a*\top}_{a,t}K_{h_1}(z_t - z)\right)^{-1} \\ & \times\left(T^{-1}\sum_{t=1}^{T} \hat{F}^{a*}_{a,t}y_{t+1}K_{h_1}(z_t - z)\right).\end{aligned}$$

By using arguments similar to those in Wang and Xia (2009), one can show that under some regularity conditions, $\sup_{z\in[0,1]} \|\hat{\beta}_{a,\lambda}(z) - \hat{\beta}^{oracle}(z)\| = o_p((Th_1)^{-1/2})$. *This indicates that* $\hat{\beta}_{a,\lambda}(z)$ *shares the same asymptotic distribution as the oracle estimator* $\hat{\beta}^{oracle}(z)$.

5 Computational Implementations

In this section, we provide the computational algorithm for estimation at Stage 2 and discuss how to select the tuning parameter λ_T in our computing.

5.1 Computational Algorithm for Stage 2

Following Li et al. (2015) and Chen et al. (2021), we introduce an iterative procedure to compute the penalized local linear estimates of the functional coefficients proposed at Stage 2 in Section 3. This algorithm can be viewed as a nonparametric extension of the coordinate descent algorithm, which is commonly applied to successively minimize a function along the coordinate directions.

Define $\tilde{\mathbf{\Omega}}_{Tl}(\ell) = diag\{\tilde{\Omega}_{Tl,1}(\ell), \ldots, \tilde{\Omega}_{Tl,T}(\ell)\}$ for $l \in 0 \cup [R]$ and $\ell = 0, 2$, where $\tilde{\Omega}_{Tl,s}(\ell) = \frac{2}{T}\sum_{t=1}^{T} \hat{F}_{lt}^2[(z_t - z_s)/h_1]^\ell K_{h_1}(z_t - z_s)$, with $\hat{F}_{0t} \equiv 1$. Now, the procedure goes as follows:

1. Find the initial estimates of $\theta_{0,l}^0(\cdot)$ and $\theta_{1,l}^0(\cdot)$, which are denoted by

$$\hat{\theta}_{0,l}^{(0)} = (\hat{\theta}_{0,l}^{(0)}(z_1), \ldots, \hat{\theta}_{0,l}^{(0)}(z_T))^\top \text{ and } \hat{\theta}_{1,l}^{(0)} = (\hat{\theta}_{1,l}^{(0)}(z_1), \ldots, \hat{\theta}_{1,l}^{(0)}(z_T))^\top$$

   respectively. These initial estimates can be obtained by using the conventional (non-penalized) local linear estimation method.

2. Let $\hat{\theta}_{0,l}^{(m)}(\cdot)$ and $\hat{\theta}_{1,l}^{(m)}(\cdot)$ be the estimates after the mth iteration. We next update the $(l+1)$th functional coefficient starting from $l = 0$. Let

$$\hat{\boldsymbol{\theta}}_{0,-l}^{(m)}(z_s) = (\hat{\theta}_{0,0}^{(m+1)}(z_s), \ldots, \hat{\theta}_{0,l-1}^{(m+1)}(z_s), 0, \hat{\theta}_{0,l+1}^{(m+1)}(z_s), \ldots, \hat{\theta}_{0,R}^{(m+1)}(z_s))^\top,$$
$$\hat{\boldsymbol{\theta}}_{1}^{(m)}(z_s) = (\hat{\theta}_{1,0}^{(m)}(z_s), \ldots, \hat{\theta}_{1,R}^{(m)}(z_s))^\top,$$

$$
\begin{aligned}
&\hat{y}^{(m)}_{t+1,-l} = y_{t+1} - \hat{F}^{a\top}_t[\hat{\boldsymbol{\theta}}^{(m)}_{0,-l}(z_s) + \hat{\boldsymbol{\theta}}^{(m)}_1(z_s)(z_t - z_s)], \text{ and} \\
&\tilde{\boldsymbol{E}}_{Tl} = (\tilde{E}_{Tl,1}, \ldots, \tilde{E}_{Tl,T})^\top,
\end{aligned}
$$

where $\tilde{E}_{Tl,s} = \frac{2}{T}\sum_{t=1}^T \hat{F}_{lt}\hat{y}^{(m)}_{t+1,-l}K_{h_1}(z_t - z_s)$. If $\|\tilde{\boldsymbol{E}}_{Tl}\| < P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)$, we update $\hat{\theta}^{(m+1)}_{0,l} = 0_{T\times 1}$; otherwise, let

$$
\hat{\theta}^{(m+1)}_{0,l} = \left(\tilde{\boldsymbol{\Omega}}_{Tl}(0) + P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)I_T/c_l\right)^{-1}\tilde{\boldsymbol{E}}_{Tl}, \tag{5.1}
$$

where $c_l = \|\hat{\theta}^{(m)}_{0,l}\|$ if $\|\hat{\theta}^{(m)}_{0,l}\| \neq 0$, and $c_l = \max_{k\neq l}\|\hat{\theta}^{(m)}_{0,k}\|$ if $\|\hat{\theta}^{(m)}_{0,l}\| = 0$.

3. Update the derivative of the $(l+1)$th functional coefficient starting from $l = 0$. Let

$$
\begin{aligned}
&\hat{\boldsymbol{\theta}}^{(m+1)}_0(z_s) = (\hat{\theta}^{(m+1)}_{0,0}(z_s), \ldots, \hat{\theta}^{(m+1)}_{0,R}(z_s))^\top, \\
&\hat{\boldsymbol{\theta}}^{(m)}_{1,-l}(z_s) = (\hat{\theta}^{(m+1)}_{1,0}(z_s), \ldots, \hat{\theta}^{(m+1)}_{1,l-1}(z_s), 0, \hat{\theta}^{(m)}_{1,l+1}(z_s), \ldots, \hat{\theta}^{(m)}_{1,R}(z_s))^\top, \\
&\check{y}^{(m)}_{t+1,-l} = y_{t+1} - \hat{F}^{a\top}_t[\hat{\boldsymbol{\theta}}^{(m+1)}_0(z_s) + \hat{\boldsymbol{\theta}}^{(m)}_{1,-l}(z_s)(z_t - z_s)], \text{ and} \\
&\check{\boldsymbol{E}}_{Tl} = (\check{E}_{Tl,1}, \ldots, \check{E}_{Tl,T})^\top,
\end{aligned}
$$

where $\check{E}_{Tl,s} = \frac{2}{T}\sum_{t=1}^T \hat{F}_{lt}\check{y}^{(m)}_{t+1,-l}[(z_t - z_s)/h_1]K_{h_1}(z_t - z_s)$. If $\|\check{\boldsymbol{E}}_{Tl}\| < P'_{\lambda_T}(\tilde{D}_l)$, we update $\hat{\theta}^{(m+1)}_{1,l} = 0_{T\times 1}$; otherwise, let

$$
h_1\hat{\theta}^{(m+1)}_{1,l} = \left(\tilde{\boldsymbol{\Omega}}_{Tl}(2) + P'_{\lambda_T}(\tilde{D}_l)I_T/d_l\right)^{-1}\check{\boldsymbol{E}}_{Tl}, \tag{5.2}
$$

where $d_l = \|h_1\hat{\theta}^{(m)}_{1,l}\|$ if $\|\hat{\theta}^{(m)}_{1,l}\| \neq 0$, and $d_l = \max_{k\neq l}\|h_1\hat{\theta}^{(m)}_{1,k}\|$ if $\|\hat{\theta}^{(m)}_{1,l}\| = 0$.

4. Repeat Steps 2 and 3 until certain convergence criterion is achieved.

5.2 Tuning Parameter Selection

Following Chen et al. (2025), we can use the following generalized information criterion (GIC) to select λ_T at Stage 2. The GIC is introduced by Fan and Tang (2013) in the context of high-dimensional penalized likelihood estimation. Since our model involves unknown

TV coefficients, the GIC needs to be modified as in Li et al. (2015). Specifically, Cheng et al. (2009) argues that each unknown functional coefficients would amount to $36/(35h_1)$ unknown constant parameters when the Epanechnikov kernel is used. Hence, we define the GIC objective function as

$$\mathrm{GIC}(\lambda_T) = \ln\left[\frac{1}{T}\sum_{t=1}^{T}\left\{y_{t+1} - \hat{F}_t^{a\top}\hat{\beta}_{\lambda_T}(z_t)\right\}^2\right] + \frac{\gamma_{T,R}}{T}\cdot\frac{36R_f(\lambda_T)}{35h_1}, \tag{5.3}$$

where $\gamma_{T,R}$ is a function of T and R, $\hat{\beta}_{\lambda_T}(z_t)$ is the TV weighted group LASSO estimate at Stage 2 using the tuning parameter λ_T, and $R_f(\lambda_T)$ is the number of selected TV coefficients using λ_T. We choose $\gamma_{T,R} = \gamma\ln(\ln(T))\ln(36R/35h_1)$ with $\gamma \in (0,1]$. We determine the tuning parameter by minimizing $\mathrm{GIC}(\lambda_T)$ defined in (5.3) with respect to λ_T. See Chen et al. (2025) for more discussions about the $\mathrm{GIC}(\lambda_T)$ criterion.

6 Monte Carlo Simulation Studies

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performance of our estimators. For each simulation study, we assume that the factor loadings are generated such that the identification restriction (IR) in (2.3) is satisfied. We compare the performance of our method with that of Fu et al. (2024). To facilitate the comparison, we ignore the intercept term in each data generating process (DGP) in our simulations with loss of generality.

6.1 Data Generating Process

We generate the data with $R = R_f + R_g = 4$ common factors:

$$x_{it} = \phi_{it}^\top F_t + e_{x,it} \text{ and } y_{t+1} = \beta_t^\top F_t + e_{y,t+1},$$

where for simplicity we do not include a TV intercept term in the y-equation. Here, $F_t = (f_t^\top, g_t^\top)^\top$, $\beta_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t})^\top$ and $\phi_{it} = (\phi_{1it}, \phi_{2it}, \phi_{3it}, \phi_{4it})^\top$, where

$$\phi_{1it} = G(10t/T; 2, 5i/N + 2) + 3.5,\ \phi_{2it} = \mu_i + G(10t/T; 0.1, (2,4,6,8)^\top),\ \mu_i \sim \text{IID } N(3.5, 1),$$

$$\phi_{3it} = \begin{cases} \phi_{i0} + 0.5b_0 & \text{for } 0.6T < t \leq 0.8T \\ \phi_{i0} - 0.5b_0 & \text{for } 0.2T < t \leq 0.4T \\ \phi_{i0} & \text{otherwise} \end{cases},\ b_0 = 2,\ \phi_{i0} \sim \text{IID } N(4, 1),$$

$$\phi_{4it} = \begin{cases} \phi_{i0} & \text{for } t = 1, 2, \ldots, T/2 \\ \phi_{i0} + b_1 & \text{for } t = T/2 + 1, \ldots, T \end{cases},\ b_1 = 1,\ \phi_{i0} \sim \text{IID } N(4, 1),$$

$G(z; \kappa, \gamma) = \{1 + \exp[-\kappa_0 \prod_{l=1}^{p}(z - \gamma_l)]\}^{-1}$ denotes the Logistic function with tuning parameter κ_0 and location parameter $\gamma = (\gamma_1, \ldots, \gamma_p)^\top$. For $l \in [4]$, we generate f_{lt} independently as an AR(1) process: $f_{lt} = 0.5 f_{l,t-1} + u_{lt}$, where $u_{lt} \sim$ IID $N(0, 1)$. We consider the following setups for β_t and the error terms $e_{x,it}$ and $e_{y,t+1}$:

DGP 1: ($R_f = 1$, serial correlation + cross sectional dependence)

$f_t = f_{1t}$, $g_t = (g_{1t}, g_{2t}, g_{3t})^\top$, $g_{lt} = 0.5 g_{l,t-1} + v_{lt}$ for $l \in [3]$, where $v_t \equiv (v_{1t},\ v_{2t},\ v_{3t})^\top \sim$ IID $N(0, \Sigma_{g3})$, where $\Sigma_{g3} = diag\{2.25\sigma_{f1}^2, 3.25\sigma_{f1}^2, 3.75\sigma_{f1}^2\}$ with $\sigma_{f1}^2 =$Var(f_{1t}), and u_{1t} and v_t are mutually independent. $e_{x,t} = (e_{1t}, \ldots, e_{Nt})^\top \sim$ IID $N(0_N, \Sigma_e)$, where $\Sigma_e = (c_{ij})_{i,j\in[N]}$ with $c_{ij} = 0.5^{|i-j|}$, $\beta_{1,t} = \cos(t/T)$, $\beta_{2,t} = \beta_{3,t} = \beta_{4,t} = 0$, $e_{y,t} = 0.5 e_{y,t-1} + u_{y,t}$, where $u_{y,t} \sim$ IID $N(0, 4^{-2})$.

DGP 2: ($R_f = 2$, serial correlation + cross sectional dependence)

$f_t \equiv (f_{1t},\ f_{2t})^\top$, $g_t = (g_{1t}, g_{2t})^\top$, $g_{lt} = 0.5 g_{l,t-1} + v_{lt}$ for $l \in [2]$, where $v_t \equiv (v_{1t}, v_{2t})^\top \sim$ IID $N(0, \Sigma_{g2})$, where $\Sigma_{g2} = diag\{4.25\sigma_{f1}^2, 4.75\sigma_{f1}^2\}$, with $\sigma_{f1}^2 =$Var(f_{1t}), $u_t \equiv (u_{1t}, u_{2t})^\top$ and v_t are mutually independent. $e_{x,t} = (e_{1t}, \ldots, e_{Nt})^\top \sim$ IID $N(0_N, \Sigma_e)$, where $\Sigma_e = (c_{ij})_{i,j\in[N]}$ with $c_{ij} = 0.5^{|i-j|}$, $\beta_{1,t} = \cos(t/T)$, $\beta_{2,t} = \sin(t/T)$, $\beta_{3,t} = \beta_{4,t} = 0$, $e_{y,t} = 0.5 e_{y,t-1} + u_{y,t}$, where $u_{y,t} \sim$ IID $N(0, 4^{-2})$.

DGP 3: ($R_f = 3$, serial correlation + cross sectional dependence)

$f_t = (f_{1t}, f_{2t}, f_{3t})^\top$, $g_{1t} = 0.5g_{1,t-1} + v_{1t}$, where $v_{1t} \sim$ IID $N(0, \Sigma_{g1})$, $\Sigma_{g1} = 8.75\sigma_{f1}^2$, and σ_{f1}^2 =Var(f_{1t}); $u_t \equiv (u_{1t},\ u_{2t},\ u_{3t})^\top$ and v_{1t} are mutually independent. $e_{x,t} = (e_{1t}, \ldots, e_{Nt})^\top \sim$ IID $N(0_N, \Sigma_e)$, where $\Sigma_e = (c_{ij})_{i,j\in[N]}$ with $c_{ij} = 0.5^{|i-j|}$, $\beta_{1,t} = \cos(t/T)$, $\beta_{2,t} = \sin(t/T)$, $\beta_{3,t} = \ln(1 + t/T)$, $\beta_{4,t} = 0$, $e_{y,t} = 0.5e_{y,t-1} + u_{y,t}$, where $u_{y,t} \sim$ IID $N(0, 4^{-2})$.

DGP 4: ($R_f = 4$, serial correlation + cross sectional dependence)

$f_t = (f_{1t}, f_{2t}, f_{3t}, f_{4t})^\top$, $e_{x,t} = (e_{1t}, \ldots, e_{Nt})^\top \sim$ IID $N(0_N, \Sigma_e)$, where $\Sigma_e = (c_{ij})_{i,j\in[N]}$ with $c_{ij} = 0.5^{|i-j|}$, $\beta_{1,t} = \cos(t/T)$, $\beta_{2,t} = \sin(t/T)$, $\beta_{3,t} = \ln(1+t/T)$, $\beta_{4,t} = 4(t/T)(1-t/T)$, $e_{y,t} = 0.5e_{y,t-1} + u_{y,t}$, where $u_{y,t} \sim$ IID $N(0, 4^{-2})$.

DGPs 5-8: (Serial correlation + cross sectional dependence + heteroskedasticity)

All settings are the same as that in DGPs 1–4, respectively, except that for each setting, $e_{y,t} = \sigma_{y,t}\varepsilon_{y,t}$, where $\sigma_{y,t}^2 = \sum_{l=1}^3 f_{l,t-1}^2$ and $\varepsilon_{y,t} = 0.5\varepsilon_{y,t-1} + v_{y,t}$ with $v_{y,t} \sim$ IID $N(0, 4^{-2})$.

6.2 Performance of Variable Selection

For the experiment of variable selection, we simulate 500 data sets with N, $T = 100$, 200 or 500. To conduct the local PCA approach, we use the Epanechnikov kernel and Silverman's rule of thumb (RoT) to set the bandwidth as $h_0 = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$. To determine the number of factors, we consider the information criteria suggested by Su and Wang (2017). In particular, let $\hat{F}_t(R)$ and $\hat{\phi}_{it}(R)$ denote the local PCA estimators of the factors and factor loadings by assuming R factors in the model and using the following normalization rule:

$$N^{-1}\Phi_r^\top \Phi_r = I_R \text{ and } T^{-1}F^{(r)\top}F^{(r)} \text{ is a diagonal matrix with descending diagonal elements,}$$

where we make the dependence of the $R \times 1$ vectors $\hat{F}_t(R)$ and $\hat{\phi}_{it}(R)$ on R explicit. Let $\hat{\Phi}_r^{(R)} = (\hat{\phi}_{1r}(R), \ldots, \hat{\phi}_{Nr}(R))^\top$ and $\breve{\Phi}_r^{(R)} = (NT)^{-1}X^{(r)\top}X^{(r)}\hat{\Phi}_r^{(R)}$ for $r \in [T]$. Let $\breve{\phi}_{ir}(R)$

denote the transpose of the ith row of $\breve{\Phi}_r^{(R)}$. Define

$$V(R, \{\breve{\Phi}_r^{(R)}\}) = \min_{\breve{F}=(\breve{F}_1^\top,\ldots,\breve{F}_T^\top)^\top} \frac{1}{NT} \sum_{i=1}^{N} \sum_{r=1}^{T} \left[x_{ir} - \breve{F}_r^\top \breve{\phi}_{ir}(R) \right]^2.$$

Then, we use the following BIC-type information criterion to determine R:

$$\mathrm{IC}(R) = \ln\left\{V(R, \{\breve{\Phi}_r^{(R)}\})\right\} + R \cdot \frac{N + Th_0}{NTh_0} \ln(C_{NT}^2), \tag{6.1}$$

where $C_{NT} = \min\{\sqrt{Th_0}, \sqrt{N}\}$. Let $\hat{R} = \arg\min_R IC(R)$.[1] As is commonly done in the empirical research, each series of $\{x_i\}_{i=1}^N$ for implementing the local PCA method is demeaned and standardized to have unit variance.

To apply our method, we first fit an unpenalized functional coefficient estimate based on the local linear method, for which the Epanechnikov kernel is used and the optimal bandwidth is selected via a 4-fold cross-validation criterion proposed in Cai et al. (2000). Then, the same bandwidth is used as h_1 for the TV LASSO, where the optimal tuning parameter λ_T is determined by minimizing the GIC criterion in (5.3). Finally, the number of factors $\hat{R}$ is also determined by the criterion in (6.1).

We classify the model selection results into three distinct scenarios: an underfitted model (missing at least one relevant predictor), an overfitted model (containing irrelevant predictors but omitting none), and a correctly fitted model (identical to the true model). The percentages of experiments resulting in each scenario are presented in Table 1. The results clearly indicate that the probability of selecting the true model increases steadily with larger N and T, approaching 100% quickly.

[1]Note that in Su and Wang (2017), they derive the selection consistency of the BIC criterion in (6.1) and compare the performance of (6.1) to that of criterions for the factor model with constant loadings, the four criteria of Bai and Ng (2002), the eigenvalue ratio (ER) and growth ratio (GR) of Ahn and Horenstein (2013), and the sequential testing procedure of Onatski (2010). Since all datasets in our simulation study are generated from the TV factor models, we prefer to using (6.1) to determine the number of factors.

Table 1: Proportion of the experiments with correctly, under and over fitted regression models

(T, N)	DGP	Under fitted	Correctly fitted	Over fitted	DGP	Under fitted	Correctly fitted	Over fitted
(100,100)	1	0.000	0.972	0.028	5	0.000	0.958	0.042
(100,200)		0.000	0.986	0.014		0.000	0.966	0.034
(200,100)		0.000	0.994	0.006		0.000	0.972	0.028
(200,200)		0.000	0.990	0.010		0.000	0.982	0.018
(500,500)		0.000	1.000	0.000		0.000	1.000	0.000
(100,100)	2	0.000	0.996	0.004	6	0.016	0.936	0.048
(100,200)		0.012	0.984	0.004		0.018	0.920	0.062
(200,100)		0.000	1.000	0.000		0.000	0.974	0.026
(200,200)		0.000	1.000	0.000		0.000	0.986	0.014
(500,500)		0.000	1.000	0.000		0.000	1.000	0.000
(100,100)	3	0.012	0.978	0.010	7	0.040	0.934	0.026
(100,200)		0.012	0.988	0.000		0.032	0.940	0.028
(200,100)		0.002	0.998	0.000		0.000	0.988	0.012
(200,200)		0.000	1.000	0.000		0.002	0.982	0.016
(500,500)		0.000	1.000	0.000		0.000	1.000	0.000
(100,100)	4	0.000	1.000	0.000	8	0.002	0.998	0.000
(100,200)		0.000	1.000	0.000		0.004	0.996	0.000
(200,100)		0.000	1.000	0.000		0.000	1.000	0.000
(200,200)		0.000	1.000	0.000		0.000	1.000	0.000
(500,500)		0.000	1.000	0.000		0.000	1.000	0.000

6.3 Performance of Forecasting

In this subsection, we study the finite sample performance of one-step ahead prediction for the target variable y_{t+1}. We also compare our forecasting results (namely, LCS) with that obtained from the approach proposed by Fu et al. (2024), abbreviated with FSW below, with the local constant method in their second step being replaced by the local linear approach. For each method, we simulate 500 data sets with N, $T = 100$, 200.

Recall that $z_t = t/T$, given the observations (y_{t+1}, z_t, x_t) with $t \in [T-1]$ and (z_T, x_T), the one-step ahead prediction of y_{t+1} is as follows:

$$\hat{y}_{T+1|T} = \hat{\beta}_{T-1}^{\top}(z_T)\hat{f}_T,$$

where $\hat{\beta}_{T-1}^{\top}(\cdot)$ is the nonparametric estimation of $\beta_{T-1}(z_T)$ using the sample $\{y_{t+1}, z_t, x_t\}_{t\in[T-1]}$ and $z_T = 1$. For both the methods of Fu et al. (2024) and ours, we first use the data $\{x_t : t \in [T]\}$ to obtain the estimates $\{\hat{f}_1, \ldots, \hat{f}_{T-1}, \hat{f}_T\}$, and then conduct nonparametric estimation based on the sample $(y_t, z_t, \hat{f}_t)$ with $t \in [T]$ to obtain $\hat{y}_{T+1|T}$, which is similar to the procedure in Li et al. (2020).

The forecasting performance is evaluated by dividing the simulated sample into two segments: an in-sample period containing the first 90% of observations for model estimation, and an out-of-sample period comprising the remaining 10% for assessing prediction accuracy. This assessment is based on the following mean squared forecast errors (MSFE):

$$\mathrm{MSFE} = \frac{1}{\lfloor 0.1T \rfloor} \sum_{t=T-\lfloor 0.1T \rfloor+1}^{T} (\hat{y}_{t+1|t} - y_{t+1})^2.$$

We report the median and standard deviation (in parentheses) of 500 MSFE values in Table 2. The results show that our estimator outperforms the benchmark from Fu et al. (2024) in all scenarios. A key finding is that forecast accuracy improves with the time series length (T), as evidenced by the declining median and standard deviation of the MSFE. This aligns with the theoretical principle that the convergence rate is driven by T, not the cross-sectional size (N). Therefore, performance metrics understandably show little variation with increases in N.

Table 2: Comparision of the median and standard deviation (in parentheses) of 500 MSFE values of the our approach (LCS) with that of FSW

(T, N)	DGP	LCS (LASSO)	FSW (TV-FAVAR)	DGP	LCS (LASSO)	FSW (TV-FAVAR)
(100,100)	1	0.201 (0.138)	0.274 (0.197)	5	0.428 (0.367)	0.739 (0.625)
(100,200)		0.197 (0.137)	0.243 (0.176)		0.389 (0.367)	0.685 (0.549)
(200,100)		0.157 (0.065)	0.181 (0.088)		0.368 (0.212)	0.512 (0.298)
(200,200)		0.145 (0.063)	0.175 (0.076)		0.376 (0.228)	0.547 (0.327)
(100,100)	2	0.170 (0.146)	0.289 (0.298)	6	0.417 (0.432)	0.757 (0.761)
(100,200)		0.191 (0.138)	0.285 (0.253)		0.426 (0.393)	0.778 (0.678)
(200,100)		0.137 (0.071)	0.189 (0.095)		0.377 (0.203)	0.551 (0.283)
(200,200)		0.138 (0.064)	0.192 (0.085)		0.358 (0.205)	0.501 (0.300)
(100,100)	3	0.278 (0.227)	0.339 (0.340)	7	0.512 (0.435)	0.774 (0.680)
(100,200)		0.276 (0.208)	0.340 (0.278)		0.527 (0.500)	0.755 (0.690)
(200,100)		0.172 (0.079)	0.197 (0.108)		0.423 (0.242)	0.579 (0.297)
(200,200)		0.177 (0.092)	0.205 (0.118)		0.417 (0.248)	0.517 (0.297)
(100,100)	4	0.324 (0.235)	0.337 (0.289)	8	0.612 (0.489)	0.802 (0.609)
(100,200)		0.348 (0.280)	0.366 (0.353)		0.609 (0.506)	0.776 (0.621)
(200,100)		0.195 (0.099)	0.197 (0.109)		0.457 (0.256)	0.560 (0.305)
(200,200)		0.200 (0.095)	0.206 (0.104)		0.475 (0.260)	0.573 (0.303)

7 Revisiting U.S. Macroeconomic Predictions

This section applies our approach to do predictions using the FRED-QD quarterly macroeconomic dataset that was originally developed by McCracken and Ng (2016) maintained by the Federal Reserve Bank of St. Louis. We use data for the period of 1959:03–2025:03, transforming each series according to its prescribed transformation code (T-code) as detailed in the appendix of the source publication and summarized in Table 3. The first two quarters are excluded due to missing values induced by differencing, resulting in a final dataset of $N = 207$ series with $T = 262$ observations per series. Each series was standardized to have zero mean and unit variance prior to analysis.

Our main objective is to compare the forecasting performance of our TV-LASSO-based method with that of the TV-factor augmented forecasting approach proposed in Fu et al.

(2024). The estimated predictive regression model is

$$\hat{y}^{\tau}_{T+\tau|T} = \hat{\beta}^{\top}_{T-\tau}(z_T)\hat{f}_T, \tag{7.1}$$

where $\hat{\beta}_{T-\tau}(\cdot)$ is the nonparametric estimation of $\beta_{T-\tau}(z_T)$ using the sample $\{y_{t+\tau}, z_t, x_t\}_{t\in[T-\tau]}$. Similar to Stock and Watson (2009) and Cheung (2024), let x_t be the original series, the transformed series y_t and the τ-month ahead variable $y^{\tau}_{t+\tau}$ are defined according to Table 3. The procedure of estimating $\hat{y}^{\tau}_{T+\tau|T}$ is similar to that in Section 6. That is, a TV factor model was estimated with data from 1959:03 through 2022:12-τ. Then, the prediction model is used to forecast $y^{\tau}_{2022:12}$. The whole estimation procedure is repeated with data from 1959:03 through 2023:03-τ to forecast $y^{\tau}_{2023:03}$, and so on.

We first determine the appropriate number of common factors for the local PCA approach. The maximum number of common factors is set to be 8 in this empirical study. Other pre-settings such as the kernel and bandwidth are the same as those in the simulation section. Then, we use the whole data from 1959:03 through 2025:03 to select the number of factors, which is $\hat{R} = 5$. As is observed in Cheung (2024), the number of factors vary across different time periods in this dataset. Therefore, our method provides a unified framework to consistently select the factors that are relevant for predicting at different time spans.

After determining the number of factors, we estimate the common factors by local PCA with the kernel and bandwidth h_0 chosen in the same way as those in the simulation study. In addition, h_1 is selected using the cross-validation criterion discussed in Section 6 and the tuning parameter λ_T is determined by the GIC criterion in (5.3). The identification restriction (IR) is discussed in Section 2.2 and the ordering of the data matters in the set of restriction. Since $\hat{R} = 5$, we need to fix the order of the first five series. To this end, we rearrange the data such that the first six series are (a) PAYEMS (All Employees: Total nonfarm); (b) INDPRO (IP Index); (c) GS5 (5-Year Treasury Constant Maturity Rate); (d) PCEPI (Personal Consumption Expenditures: Chain-type Price Index); (e) TOTRESNS (Total Reserves of Depository Institutions). Our target variables of interest include eight series from four categories; see Table 4 for details.

To evaluate the forecasting performance, we compute the mean squared forecast errors (MSFE) as follows:

$$\text{MSFE}(\tau) = \frac{1}{T_{2022:12}} \sum_{t=T-T_{2022:12}+1}^{T} (\hat{y}^{\tau}_{t+\tau|t} - y^{\tau}_{t+\tau})^2.$$

Table 5 reports the MSFEs for τ-step-ahead forecasts, where $\tau \in [4]$. Our penalized TV-FA model outperforms the TV-FA model of Fu et al. (2024) in 29 of the 32 cases (90.625%), with superior performance being particularly pronounced at larger forecast horizons ($\tau = 3, 4$), where it achieves uniform dominance. We conjecture that this occurs because the prediction error in equation (7.1) increases with τ, making the role of regularization more critical. These results suggest that irrelevant factors become increasingly prevalent in the prediction model at longer horizons, which can be effectively mitigated by our penalized approach.

Table 3: Transformations of variables.

T-code	Transformation (y_t)	τ-month ahead variable ($y^{\tau}_{t+\tau}$)
1	$y_t = x_t$	$y^{\tau}_{t+\tau} = x_{t+\tau}$
2	$y_t = \Delta x_t$	$y^{\tau}_{t+\tau} = x_{t+\tau} - x_t$
3	$y_t = \Delta^2 x_t$	$y^{\tau}_{t+\tau} = \tau^{-1} \sum_{j=0}^{\tau-1} \Delta x_{t+\tau-j} - \Delta x_t$
4	$y_t = \ln(x_t)$	$y^{\tau}_{t+\tau} = \ln(x_{t+\tau})$
5	$y_t = \Delta \ln(x_t)$	$y^{\tau}_{t+\tau} = \ln(x_{t+\tau}) - \ln(x_t)$
6	$y_t = \Delta^2 \ln(x_t)$	$y^{\tau}_{t+\tau} = \tau^{-1} \sum_{j=0}^{\tau-1} \Delta \ln(x_{t+\tau-j}) - \Delta \ln(x_t)$
7	$y_t = \Delta(x_t/x_{t-1} - 1)$	$y^{\tau}_{t+\tau} = \tau^{-1} \sum_{j=0}^{\tau-1} x_{t+\tau-j}/x_{t+\tau-j-1} - x_t/x_{t-1}$

8 Conclusion

This paper introduces a time-varying factor-augmented (TV-FA) forecasting model that accommodates smoothly evolving factor loadings and allows the number of factors relevant for forecasting to be smaller than the number driving the entire predictor set. We propose a two-stage procedure that performs nonparametric estimation and variable selection simultaneously, establishing the asymptotic properties of the estimators within a large N, large

Table 4: Variables of interest for forecasting.

Series		T-code
(a)	Output and income	
	INDPRO, IP Index	5
	DPIC96: Real Disposable Personal Income	5
(b)	Labor market	
	HWIx, Help-Wanted Index	1
	UNRATE, Civilian Unemployment Rate	2
(c)	Consumption, orders and inventories	
	PCECC96: Real Personal Consumption Expenditures	5
	ISRATIOx: Total Business: Inventories to Sales Ratio	2
(d)	Prices	
	WPSFD49207: PPI by Commodity for Final Demand: Finished Goods	6
	CUSR0000SA0L2: CPI: All items less shelter	6

Table 5: The comparison of our forecasting result with that of Fu et al. (2024)

Variables	τ	LCS (LASSO)	FSW (TV-FAVAR)	Variables	LCS (LASSO)	FSW (TV-FAVAR)
INDPRO	1	**0.138**	0.292	PCECC96	**0.251**	0.684
	2	**0.176**	0.363		**0.202**	0.855
	3	**0.202**	0.269		**0.193**	1.201
	4	**0.237**	0.274		**0.190**	1.230
DPIC96	1	**1.755**	3.102	ISRATIOx	**0.301**	0.402
	2	**3.161**	7.031		**0.900**	1.795
	3	**3.036**	4.929		**0.838**	1.926
	4	**2.023**	4.488		**0.896**	1.889
HWIx	1	8.575	**7.135**	WPSFD49207	**3.210**	3.465
	2	7.448	**6.578**		**2.698**	3.593
	3	**7.352**	8.746		**3.058**	3.971
	4	**6.259**	7.722		**3.385**	4.133
UNRATE	1	**2.469**	2.728	CUSR0000SA0L2	**0.850**	0.924
	2	1.158	**1.011**		**1.338**	1.569
	3	**2.202**	2.886		**1.871**	2.034
	4	**2.277**	2.627		**2.081**	2.355

Note: Bold entries highlight better performance in each case.

T framework. Monte Carlo simulations demonstrate the strong performance of our estimators and forecasts against competing methods. Finally, an empirical application to a U.S. macroeconomic dataset shows that our penalized TV-FA model outperforms conventional TV-FAVAR models as in Fu et al. (2024) in predicting key macroeconomic time series.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J., Li, K., and Lu, L. (2016). Estimation and inference of FAVAR models. *Journal of Business & Economic Statistics*, 34(4):620–641.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

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

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Baltagi, B. H., Kao, C., and Wang, F. (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics*, 220(2):349–365.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly jJournal of Economics*, 120(1):387–422.

Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1):71–84.

Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory*, 18(1):169–192.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

Chen, J., Li, D., Li, Y.-N., and Linton, O. (2025). Estimating time-varying networks for high-dimensional time series. *Journal of Econometrics*, 249(Part C):105941.

Chen, J., Li, D., Wei, L., and Zhang, W. (2021). Nonparametric homogeneity pursuit in functional-coefficient models. *Journal of Nonparametric Statistics*, 33(3-4):387–416.

Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, 180(1):30–48.

Cheng, M.-Y., Zhang, W., and Chen, L.-H. (2009). Statistical estimation in generalized multiparameter likelihood models. *Journal of the American Statistical Association*, 104(487):1179–1191.

Cheung, Y. L. (2024). Identification of time-varying factor models. *Journal of Business & Economic Statistics*, 42(1):76–94.

Corradi, V. and Swanson, N. R. (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics*, 182(1):100–118.

Dendramis, Y., Kapetanios, G., and Marcellino, M. (2023). Time varying three pass regression filter. Technical report, CEPR Discussion Papers, No. 18547. CEPR Press, Paris & London.

Eickmeier, S., Lemke, W., and Marcellino, M. (2015). Classical time varying factor-augmented vector auto-regressive models—estimation, forecasting and structural analysis. *Journal of the Royal Statistical Society Series A*, 178(3):493–533.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Fan, Y. and Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society Series B*, 75(3):531–552.

Fu, Z., Hong, Y., and 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., and Wang, X. (2024). Estimation and inference on time-varying FAVAR models. *Journal of Business & Economic Statistics*, 42(2):533–547.

Gasser, T. and Müller, H. (1979). *Kernel Estimation of Regression Functions*. Lecture Notes in Mathematics No. 757. Springer, Berlin.

Hall, P. and Wehrly, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association*, 86(415):665–672.

Han, X. and Inoue, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory*, 31(5):1117–1152.

Hong, Y. and Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies*, 18(1):37–84.

Kelly, B. and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2):294–316.

Li, D., Ke, Y., and Zhang, W. (2015). Model selection and structure specification in ultra-high dimensional generalized semi-varying coefficient models. *Annals of Statistics*, 43(6):2676–2705.

Li, D., Tosasukul, J., and Zhang, W. (2020). Nonlinear factor-augmented predictive regression models with functional coefficients. *Journal of Time Series Analysis*, 41(3):367–386.

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

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

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

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92(4):1004–1016.

Pelger, M. and Xiong, R. (2022). State-varying factor models of large dimensions. *Journal of Business & Economic Statistics*, 40(3):1315–1333.

Peng, B., Su, L., and Yan, Y. (2025). A robust residual-based test for structural changes in factor models. *Journal of Econometrics*, 251:106042.

Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*, pages 253–264. Springer, Berlin.

Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structural Change: Analysis and Forecasting*, pages 179–190. Springer, Berlin.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. In *The Methodology and Practice of Econometrics. A Festschrift in Honou of David F. Hendry*, pages 205–261. Oxford University Press Oxford.

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

Su, L. and Wang, X. (2020a). Corrigendum to "on time-varying factor models: Estimation and testing" [j. econometrics 198 (2017) 84-101].

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

Wang, H. and Xia, Y. (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association*, 104(486):747–757.

Wang, X., Jin, S., Li, Y., Qian, J., and Su, L. (2025). On time-varying panel data models with time-varying interactive fixed effects. *Journal of Econometrics*, 249(Part B):105960.

White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press, New York.

Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of statistics*, 36(4):1509.

Online Supplement for **"Time-varying Factor-augmented Forecasting Models with Variable Selection"**

This online supplement is comprised of two sections. Section A contains the proofs of the main results in the paper, which call upon some technical lemmas. Section B contains the proofs of technical lemmas in Section A.

A Proofs of the main results in the paper

Let H_r, $V_{NT}^{(r)}$, V_r, Υ_r and P_r be as defined in the beginning of Section 4.3 Let $\Phi_r^{(c)} = (\phi_1^{(c)}(r/T), \ldots, \phi_N^{(c)}(r/T))^\top$, where $\phi_i^{(c)}(\cdot)$ denotes the cth order derivative of $\phi_i(\cdot)$ for $c \in [2]$. Recall that $k^*_{h_0,tr} = h_0^{-1} K_r^*((t-r)/(Th_0))$ and K_r^* is the boundary kernel defined in the main text and used in Su and Wang (2017). Let

$$B_t^{(r)} = \left[\bar{C}_1^{(r)} \left(\frac{t-r}{T}\right) + \bar{C}_2^{(r)} \mu_{21} h_0^2 + \bar{C}_3^{(r)} \left(\frac{t-r}{T}\right)^2\right] F_t^{(r)},$$

where

$$\begin{aligned}
\bar{C}_1^{(r)} &= V_{NT}^{(r)-1} H_r^\top \Sigma_F (\Phi_r^\top \Phi_r^{(1)}/N), \\
\bar{C}_2^{(r)} &= V_{NT}^{(r)-1} \left[\frac{1}{2} H_r^\top \Sigma_F (\Phi_r^{(2)\top} \Phi_r/N) + \bar{C}_1^{(r)} \Sigma_F (\Phi_r^\top \Phi_r^{(1)}/N)\right], \text{ and} \\
\bar{C}_3^{(r)} &= \frac{1}{2} V_{NT}^{(r)-1} H_r^\top \Sigma_F (\Phi_r^\top \Phi_r^{(2)}/N).
\end{aligned}$$

Let $B^{(r)} = (B_1^{(r)}, \ldots, B_T^{(r)})^\top$. Finally, let $\Delta_{it}^{(r)} = d_i(t,r)^\top F_t^{(r)}$ and $\Delta_i^{(r)} = (\Delta_{i1}^{(r)}, \ldots, \Delta_{iT}^{(r)})^\top$, where $d_i(t,r) = \phi_i\left(\frac{t}{T}\right) - \phi_i\left(\frac{r}{T}\right)$.

Noting that by $T^{-1}\tilde{F}^{(r)\top}\tilde{F}^{(r)} = I_R$, we make the following decomposition:

$$\begin{aligned}
\tilde{\phi}_{ir} &= T^{-1}\tilde{F}^{(r)\top} x_i^{(r)} = T^{-1}\tilde{F}^{(r)\top} F^{(r)} \phi_{ir} + T^{-1}\tilde{F}^{(r)\top} e_{x,i}^{(r)} + T^{-1}\tilde{F}^{(r)\top} \Delta_i^{(r)} \\
&= H_r^{-1}\phi_{ir} + T^{-1} H_r^\top F^{(r)\top} e_{x,i}^{(r)} + T^{-1}\left(\tilde{F}^{(r)} - F^{(r)} H_r - B^{(r)}\right)^\top e_{x,i}^{(r)} \\
&\quad - T^{-1}\tilde{F}^{(r)\top}\left(\tilde{F}^{(r)} - F^{(r)} H_r - B^{(r)}\right) H_r^{-1}\phi_{ir} + T^{-1}\left(\tilde{F}^{(r)} - F^{(r)} H_r - B^{(r)}\right)^\top \Delta_i^{(r)}
\end{aligned}$$

$$
\begin{aligned}
&+T^{-1}H_r^\top F^{(r)\top}\Delta_i^{(r)}+T^{-1}B^{(r)\top}e_{x,i}^{(r)}-T^{-1}\tilde{F}^{(r)\top}B^{(r)}H_r^{-1}\phi_{ir}+T^{-1}B^{(r)\top}\Delta_i^{(r)}\\
\equiv\;& H_r^{-1}\phi_{ir}+\sum_{m=1}^{8}D_m(i,r).
\end{aligned}\tag{A.1}
$$

Recall that $\max_t=\max_{t\in[T]}$ and $\max_{i,t}=\max_{i\in[N],t\in[T]}$. The following four lemmas are needed in the proofs of Theorem 4.1, Corollary 4.2 and Theorem 4.3.

Lemma A.1. *Suppose that Assumptions 4.1–4.3 hold. Then*

(i) $\max_{i,t}\left\|\tilde{\phi}_{it}-H_t^{-1}\phi_{it}\right\|=O_p((Th_0/\ln T)^{-1/2}+h_0^2)$;

(ii) $\max_t\left\|\tilde{F}_t-H_t^\top F_t\right\|=O_p((N/\ln T)^{-1/2}+h_0^2)$;

(iii) $\max_{i,t}\left\|\tilde{\phi}_{it}^\top\tilde{F}_t-\phi_{it}^\top F_t\right\|=o_p((Th_0/\ln T)^{-1/2}T^{1/8}+h_0^2T^{1/8})$.

Lemma A.2. *Suppose that Assumptions 4.1–4.3 hold. Then*

(i) $\sqrt{Th_0}D_m(i,r)=o_p(1)$, *for* $m=2,3,4,6$;

(ii) $\sqrt{Th_0}D_5(i,r)=\sqrt{Th_0}\frac{\mu_{21}h_0^2}{2}P_r^{-1\top}\Sigma_F\phi_{ir}^{(2)}+o_p(1)$;

(iii) $\sqrt{Th_0}D_7(i,r)=-\sqrt{Th_0}\mu_{21}h_0^2\big[H_r^\top\Sigma_F\bar{C}_2^{(r)}+H_r^\top\Sigma_F\bar{C}_3^{(r)}+\bar{C}_1^{(r)}\Sigma_F\bar{C}_1^{(r)\top}\big]H_r^{-1}\phi_{ir}+o_p(1)$;

(iv) $\sqrt{Th_0}D_8(i,r)=\sqrt{Th_0}\mu_{21}h_0^2\bar{C}_1^{(r)}\Sigma_F\phi_{ir}^{(1)}+o_p(1)$.

Lemma A.3. *Suppose that Assumptions 4.1–4.3 hold. Then*

(i) $N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t=P_t\Sigma_{\Phi_t}P_t^\top+o_p(1)$;

(ii) $\sum_{i=1}^N\left[\hat{\phi}_{it}-H_t^{-1}\phi_{it}\right]e_{x,it}=o_p(N^{-1/2})$;

(iii) $\frac{1}{\sqrt{N}}\sum_{i=1}^N\hat{\phi}_{it}\left[\hat{\phi}_{it}-H_t^{-1}\phi_{it}\right]^\top H_t^\top F_t=\sqrt{N}\check{H}_t^\top F_t+o_p(1)$, *where*

$$
\begin{aligned}
\check{H}_t^\top=&h_0^2\mu_{21}H_t^{-1}\Bigg\{\frac{1}{2}\frac{\Phi_t^\top\Phi_t^{(2)}}{N}\Sigma_FH_t+\frac{\Phi_t^\top\Phi_t^{(1)}}{N}\Sigma_F\bar{C}_1^{(t)\top}-\frac{\Phi_t^\top\Phi_t}{N}(H_t^{-1})^\top\\
&\times\left[\bar{C}_1^{(t)}\Sigma_F\bar{C}_1^{(t)\top}+\left(\bar{C}_2^{(t)}+\bar{C}_3^{(t)}\right)\Sigma_FH_t\right]\Bigg\}H_t^\top.
\end{aligned}
$$

Lemma A.4. *Suppose that Assumptions 4.1–4.3 hold. Let* $C_{NT}=\min\{\sqrt{Th_0},\sqrt{N},h_0^{-2}\}$,

$$
B_\Phi(i,r)=\left[\bar{C}_1^{(r)}\Sigma_F\phi_{ir}^{(1)}+\frac{1}{2}P_r^{-1\top}\Sigma_F\phi_{ir}^{(2)}-(H_r^\top\Sigma_F\bar{C}_2^{(r)}+H_r^\top\Sigma_F\bar{C}_3^{(r)}+\bar{C}_1^{(r)}\Sigma_F\bar{C}_1^{(r)\top})H_r^{-1}\phi_{ir}\right]\mu_{21}h_0^2,\tag{A.2}
$$

and $B_\Phi^{(r)}=(B_\Phi(1,r),\ldots,B_\Phi(N,r))^\top$. *Then*

(i) $\max_r T^{-1}\left\|\tilde{F}^{(r)}-F^{(r)}H_r-B^{(r)}\right\|^2=O_p((Th_0)^{-1}+N^{-1}\ln T)+o_p(h_0^4)$;

(ii) $\max_r\left\|T^{-1}\left(\tilde{F}^{(r)}-F^{(r)}H_r-B^{(r)}\right)^\top F^{(r)}H_r\right\|=O_p((Th_0)^{-1}+N^{-1}\ln T)+o_p(h_0^4)$;

(iii) $\max_r \left\|T^{-1}\tilde{F}^{(r)\top}\left(\tilde{F}^{(r)} - F^{(r)}H_r - B^{(r)}\right)\right\| = O_p((Th_0)^{-1} + N^{-1}\ln T) + o_p(h_0^4)$;

(iv) $\max_r \left\|T^{-1}\left(\tilde{F}^{(r)} - F^{(r)}H_r - B^{(r)}\right)^\top e_{x,i}^{(r)}\right\| = O_p((Th_0)^{-1} + N^{-1}\ln T) + o_p(h_0^4)$;

(v) $\max_r \left\|T^{-1}\left(\tilde{F}^{(r)} - F^{(r)}H_r - B^{(r)}\right)^\top \Delta_i^{(r)}\right\| = O_p((Th_0)^{-1} + N^{-1}\ln T) + o_p(h_0^4)$;

(vi) $\max_r \left\|T^{-1}H_r^\top F^{(r)\top}\Delta_i^{(r)}\right\| = O_p((Th_0)^{-1} + (Th_0/\ln T)^{-1/2}h_0 + h_0^2)$;

(vii) $\max_r \left\|T^{-1}B^{(r)\top}F^{(r)}H_r\right\| = O_p((Th_0/\ln T)^{-1/2}h_0)$;

(viii) $\max_r \left\|T^{-1}B^{(r)\top}e_{x,i}^{(r)}\right\| = O_p((Th_0/\ln T)^{-1/2}h_0)$;

(ix) $\max_r \left\|T^{-1}\tilde{F}^{(r)\top}B^{(r)}H_r^{-1}\phi_{ir}\right\| = O_p((Th_0)^{-1} + (Th_0/\ln T)^{-1/2}h_0 + h_0^2 + N^{-1}h_0\ln T)$;

(x) $\max_r \left\|T^{-1}\tilde{F}^{(r)\top}B^{(r)}\right\| = O_p((Th_0)^{-1} + (Th_0/\ln T)^{-1/2}h_0 + h_0^2 + N^{-1}h_0\ln T)$;

(xi) $\max_r \left\|T^{-1}B^{(r)\top}\Delta_i^{(r)}\right\| = O_p(h_0^2)$;

(xii) $\max_r \left\|N^{-1}\tilde{\Phi}_r^\top\tilde{\Phi}_r - P_r\Sigma_{\Phi_r}P_r^\top\right\| = o_p(1)$;

(xiii) $\max_r N^{-1}\left\|\tilde{\Phi}_r - \Phi_r(H_r^{-1})^\top - B_\Phi^{(r)}\right\|^2 = O_p(C_{NT}^{-2}\ln T)$;

(xiv) $\max_r N^{-1}\left\|\left(\tilde{\Phi}_r - \Phi_r(H_r^{-1})^\top - B_\Phi^{(r)}\right)^\top \Phi_r(H_r^{-1})^\top\right\| = O_p(C_{NT}^{-2}\ln T)$;

(xv) $\max_r N^{-1}\left\|\left(\tilde{\Phi}_r - \Phi_r(H_r^{-1})^\top - B_\Phi^{(r)}\right)^\top e_{x,r}\right\| = O_p(C_{NT}^{-2}\ln T)$.

Lemma A.1 studies the uniform convergence of $\tilde{\phi}_{it}$, $\tilde{F}_t$ and $\tilde{\phi}_{it}^\top\tilde{F}_t$, which is mainly needed in the derivation of the convergence of H_t^* in Theorem 4.1.

Proof of Theorem 4.1.

First, we observe that

$$\begin{aligned}
& \max_r \|T^{-1}H_r^\top F^{(r)\top}F^{(r)}H_r - I_R\| \\
= \; & \max_r \|T^{-1}[(H_r^\top F^{(r)\top} - \tilde{F}^{(r)\top} - B^{(r)\top}) + \tilde{F}^{(r)\top} + B^{(r)\top}]F^{(r)}H_r - I_R\| \\
\leq \; & \max_r \|T^{-1}(H_r^\top F^{(r)\top} - \tilde{F}^{(r)\top} - B^{(r)\top})F^{(r)}H_r\| + \max_r \left\|T^{-1}B^{(r)\top}F^{(r)}H_r\right\| \\
& + \max_r \|T^{-1}\tilde{F}^{(r)\top}F^{(r)}H_r - I_R\| \\
= \; & \max_r \|T^{-1}(H_r^\top F^{(r)\top} - \tilde{F}^{(r)\top} - B^{(r)\top})F^{(r)}H_r\| + \max_r \left\|T^{-1}B^{(r)\top}F^{(r)}H_r\right\| \\
& + \max_r \|T^{-1}\tilde{F}^{(r)\top}[(F^{(r)}H_r - \tilde{F}^{(r)} - B^{(r)}) + B^{(r)} + \tilde{F}^{(r)}] - I_R\| \\
\leq \; & \max_r \|T^{-1}(F^{(r)}H_r - \tilde{F}^{(r)} - B^{(r)})^\top F^{(r)}H_r\| + \max_r \|T^{-1}\tilde{F}^{(r)\top}(F^{(r)}H_r - \tilde{F}^{(r)} - B^{(r)})\| \\
& + \max_r \left\|T^{-1}B^{(r)\top}F^{(r)}H_r\right\| + \max_r \left\|T^{-1}\tilde{F}^{(r)\top}B^{(r)}\right\|
\end{aligned}$$

$$= \quad O_p((Th_0)^{-1} + N^{-1}\ln T + (Th_0/\ln T)^{-1/2}h_0 + h_0^2), \tag{A.3}$$

where the first inequality holds by the triangle inequality, the second inequality holds by the fact that $T^{-1}\tilde{F}^{(r)\top}\tilde{F}^{(r)} = I_R$ and the triangle inequality, and the last equality holds by Lemma A.4(ii), (iii), (vii) and (x).

Second, by the error analysis in Riemann sum approximation of a definite integral and Assumptions 4.1(ii) and 4.3, we have $T^{-1}\sum_{t=1}^{T} E(F_t^{(r)}F_t^{(r)\top}) = \frac{1}{Th_0}\sum_{t=1}^{T} K_r^*(\frac{t-r}{Th_0})E(F_tF_t^\top) = \Sigma_F + O(\frac{1}{Th_0})$ uniformly in $r/T \in (0,1]$. Let ι_m be an $R \times 1$ vector with one in its mth position and zero elsewhere. For any m, $n \in [R]$, we have by Assumptions 4.1(iv) and 4.3,

$$\begin{aligned} Var\left(\frac{1}{T}\sum_{t=1}^{T}\iota_m^\top F_t^{(r)}F_t^{(r)\top}\iota_n\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} k_{h,tr}^* k_{h,sr}^* Cov(F_{t,m}F_{t,n}, F_{s,m}F_{s,n}) \\ &\leq T^{-1}\max_t k_{h,tr}^* \frac{1}{T}\sum_{s=1}^{T} k_{h,sr}^* \max_s \sum_{t=1}^{T} |Cov(F_{t,m}F_{t,n}, F_{s,m}F_{s,n})| \\ &= T^{-1}O(h_0^{-1})O(1)O(1) = O((Th_0)^{-1}). \end{aligned}$$

This, in conjunction with (IR) in (2.3), implies that $T^{-1}F^{(r)\top}F^{(r)} - I_R = O_p((Th_0)^{-1/2})$. Also note that by Assumption 4.2(iii),

$$\begin{aligned} \max_r \left\|T^{-1}F^{(r)\top}F^{(r)} - I_R\right\| &= \max_r \left\|T^{-1}\sum_t k_{h,tr}^*(F_tF_t^\top - I_R) + T^{-1}\sum_t k_{h,tr}^* I_R - I_R\right\| \\ &= O((Th_0/\ln T)^{-1/2}) + O((Th_0)^{-1}) = O((Th_0/\ln T)^{-1/2}). \end{aligned} \tag{A.4}$$

Then, by the triangle inequality, (A.3) and (A.4)

$$\begin{aligned} \max_r \|H_r^\top H_r - I_R\| &\leq \max_r \left\|H_r^\top H_r - H_r^\top T^{-1}F^{(r)\top}F^{(r)}H_r\right\| + \max_r \left\|H_r^\top T^{-1}F^{(r)\top}F^{(r)}H_r - I_R\right\| \\ &= O_p((Th_0)^{-1/2} + N^{-1}\ln T + h_0^2). \end{aligned} \tag{A.5}$$

Thus, H_r is an orthogonal matrix up to a term that is of order $O_p((Th_0)^{-1/2} + N^{-1}\ln T + h_0^2)$, and so is $H_r\tilde{Q}_r^\top$ since $\tilde{Q}_r$ is orthogonal by construction. Pre- and post-multiplying $H_r^\top H_r$ by $\tilde{Q}_r$ and $\tilde{Q}_r^\top$ respectively and using (A.5) yields

$$\max_r \|\tilde{Q}_r H_r^\top H_r \tilde{Q}_r^\top - I_R\| = O_p((Th_0)^{-1/2} + N^{-1}\ln T + h_0^2).$$

Now, by the definition of $\tilde{\Phi}_{1:R,r}^\top$ and Lemma A.1(i), we have that

$$\max_r \|\tilde{Q}_r^\top \tilde{R}_r - H_r^{-1}\Phi_{1:R,r}^\top\| = \max_r \|\tilde{\Phi}_{1:R,r}^\top - H_r^{-1}\Phi_{1:R,r}^\top\| = O_p((Th_0/\ln T)^{-1/2} + h_0^2),$$

where $\tilde{R}_r$ is an upper triangular matrix with positive diagonal elements. Since $\Phi_{1:R,r}^\top$ is also an upper triangular matrix and H_r^{-1} is uniformly an orthogonal matrix up to a $O_p((Th_0)^{-1/2}+N^{-1}\ln T+h_0^2)$ term by (A.5), by the uniqueness of the QR decomposition, we have

$$\max_r \|\tilde{Q}_r^\top - H_r^{-1}\| = O_p((Th_0/\ln T)^{-1/2} + N^{-1}\ln T + h_0^2).$$

This, in conjunction with the definition of H_r^* after equation (3.3) in Section 3.1, implies that

$$\max_r \|H_r^* - I_R\| = \max_r \|H_r\tilde{Q}_r^\top - I_R\| = O_p((Th_0/\ln T)^{-1/2} + N^{-1}\ln T + h_0^2). \blacksquare \tag{A.6}$$

Proof of Corollary 4.2.

First, by equation (3.3) in the main text, we make the following decomposition:

$$\hat{F}_t - F_t = \tilde{Q}_t(\tilde{F}_t - H_t^\top F_t) + (\tilde{Q}_t H_t^\top - I_R)F_t. \tag{A.7}$$

Noting that $x_{it} = \tilde{\phi}_{it}^\top H_t^\top F_t + e_{x,it} + (H_t^{-1}\phi_{it} - \tilde{\phi}_{it})^\top H_t^\top F_t$, we have

$$\begin{aligned}
\tilde{F}_t - H_t^\top F_t &= \tilde{S}_{\phi,t}^{-1}\left(\frac{1}{N}\sum_{i=1}^N \tilde{\phi}_{it}x_{it}\right) - H_t^\top F_t \\
&= \tilde{S}_{\phi,t}^{-1}\left\{H_t^{-1}\frac{1}{N}\sum_{i=1}^N \phi_{it}e_{x,it} + \frac{1}{N}\sum_{i=1}^N(\tilde{\phi}_{it} - H_t^{-1}\phi_{it})e_{x,it} - \frac{1}{N}\sum_{i=1}^N \tilde{\phi}_{it}(\tilde{\phi}_{it} - H_t^{-1}\phi_{it})^\top H_t^\top F_t\right\} \\
&\equiv A_1(t) + A_2(t) + A_3(t), \qquad (A.8)
\end{aligned}$$

where $\tilde{S}_{\phi,t} \equiv N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t \equiv N^{-1}\sum_{i=1}^N \tilde{\phi}_{it}\tilde{\phi}_{it}^\top$. By the definition of $B_\Phi^{(r)}$ and Lemma A.4(xii)-(xv), we have $\max_t \|A_2(t)\| + \max_t \|A_3(t)\| = O_p(h_0^2 + C_{NT}^{-2}\ln T)$. Then

$$\max_t \|\tilde{F}_t - H_t^\top F_t\| = \max_t \left\|(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}H_t^{-1}\frac{1}{N}\sum_{i=1}^N \phi_{it}e_{x,it}\right\| + O_p(h_0^2 + C_{NT}^{-2}\ln T).$$

By the fact that $\tilde{Q}_t\tilde{Q}_t^\top = I_R$ and the result in Theorem 4.1, we have

$$\begin{aligned}
\max_t \|\tilde{Q}_t - H_t\| &\leq \max_t \|\tilde{Q}_t H_t^{-1} - I_R\| \max_t \|H_t\| = \max_t \|H_t\| \max_t \|(\tilde{Q}_t^\top)^{-1}H_t^{-1} - I_R\| \\
&\lesssim \max_t \|(H_t\tilde{Q}_t^\top)^{-1} - I_R\| = \max_t \|H_t^{*-1} - I_R\| = o_p(1),
\end{aligned}$$

where $a \lesssim b$ denotes that $a/b = O_p(1)$. Then, by Assumption 4.2(iv), Lemma A.3(i), Theorem 4.1

and (A.3),

$$
\begin{aligned}
\max_t \|\hat{F}_t - F_t\| &\le \max_t \|\tilde{Q}_t\| \left\{ \max_t \left\| (N^{-1}\tilde{\Phi}_t^\top \tilde{\Phi}_t)^{-1} \frac{1}{N} \sum_{i=1}^N \phi_{it} e_{x,it} \right\| + O_p(h_0^2 + C_{NT}^{-2} \ln T) \right\} \\
&\quad + \max_t \|(H_t^* - I_R)^\top F_t\| \\
&\lesssim \max_t \left\| \frac{1}{N} \sum_{i=1}^N \phi_{it} e_{x,it} \right\| + \max_t \|H_t^* - I_R\| \max_t \|F_t\| + O_p(h_0^2 + C_{NT}^{-2} \ln T) \\
&= O_p((N/\ln T)^{-1/2} + h_0^2) + O_p((Th_0/\ln T)^{-1/2} + h_0^2) o_p(T^{1/8}), \\
&= O_p((N/\ln T)^{-1/2} + h_0^2) + o_p((Th_0/\ln T)^{-1/2} T^{1/8} + h_0^2 T^{1/8}),
\end{aligned}
$$

where the second last equality follows from the fact that $\max_t \|F_t\| = o_p(T^{1/8})$ under Assumption 4.1(ii). ■

Proof of Theorem 4.3.

Let $\tilde{\Sigma}_F^{(r)} \equiv T^{-1} F^{(r)\top} F^{(r)}$. Then, using the same deduction as in (A.3) and by the fact that $\tilde{Q}_r$ is orthogonal,

$$
\max_r \|(\tilde{\Sigma}_F^{(r)1/2} H_r \tilde{Q}_r^\top)^\top (\tilde{\Sigma}_F^{(r)1/2} H_r \tilde{Q}_r^\top) - I_R\| = O_p((Th_0)^{-1} + N^{-1} \ln T + (Th_0/\ln T)^{-1/2} h_0 + h_0^2). \tag{A.9}
$$

Now, let $\delta_r \equiv \tilde{\Sigma}_F^{(r)1/2} H_r \tilde{Q}_r^\top - I_R$. Then

$$
\max_r \|(\delta_r + I_R)^\top (\delta_r + I_R) - I_R\| = O_p((Th_0)^{-1} + N^{-1} \ln T + (Th_0/\ln T)^{-1/2} h_0 + h_0^2).
$$

Therefore, $\max_r \|\delta_r^\top \delta_r + \delta_r + \delta_r^\top\| = O_p((Th_0)^{-1} + N^{-1} \ln T + (Th_0/\ln T)^{-1/2} h_0 + h_0^2)$. Noting that

$$
\begin{aligned}
\max_r \|\delta_r\| &= \max_r \left\| H_r \tilde{Q}_r^\top - I_R + \left(\tilde{\Sigma}_F^{(r)1/2} - I_R \right) H_r \tilde{Q}_r^\top \right\| \\
&\lesssim \max_r \left\| H_r \tilde{Q}_r^\top - I_R \right\| + \max_r \left\| \tilde{\Sigma}_F^{(r)1/2} - I_R \right\| = O_p((Th_0/\ln T)^{-1/2} + N^{-1} \ln T + h_0^2)
\end{aligned}
$$

by (A.4) and (A.6), we have

$$
\begin{aligned}
\max_r \|\delta_r^\top \delta_r\| &\le \max_r \|\delta_r\|^2 = O_p \left\{ \left((Th_0/\ln T)^{-1/2} + N^{-1} \ln T + h_0^2 \right)^2 \right\} \\
&= O_p((Th_0/\ln T)^{-1} + N^{-2} (\ln T)^2 + h_0^4).
\end{aligned}
$$

Then by the triangle inequality,

$$
\max_r \|\delta_r + \delta_r^\top\| \le \max_r \|\delta_r^\top \delta_r\| + \max_r \|\delta_r^\top \delta_r + \delta_r + \delta_r^\top\| = O_p([(Th_0)^{-1} + N^{-1}] \ln T + h_0^2),
$$

which means that δ_r is skew-symmetric,[1] up to an asymptotically negligible term.

Now, by (A.1) and Lemma A.2, $\sum_{m=2}^{8} D_m(i,r) = B_\Phi(i,r) + o_p((Th_0)^{-1/2})$, where $B_\Phi(i,r)$ is defined in Lemma A.4. Therefore, by the proof of Theorem 2.5(i) in Su and Wang (2020), we have uniformly in $(i,r) \in [N] \times [T]$,

$$\tilde{\phi}_{ir} - H_r^{-1}\phi_{ir} = T^{-1}H_r^\top F^{(r)\top} e_{x,i}^{(r)} + B_\Phi(i,r) + o_p((Th_0)^{-1/2}) \ .$$

Then uniformly in $r \in [T]$,

$$\tilde{\Phi}_{1:R,r}^\top - H_r^{-1}\Phi_{1:R,r}^\top = H_r^\top T^{-1}\sum_{s=1}^{T} k_{h,sr}^* F_s e_{x,(1:R)s}^\top + B_\Phi^\top(1:R,r) + o_p((Th_0)^{-1/2}), \tag{A.10}$$

where $e_{x,(1:R)s}^\top = (e_{x,1s}, ...e_{x,Rs})$, and

$$\begin{aligned} B_\Phi^\top(1:R,r) =& \Bigg[\bar{C}_1^{(r)}\Sigma_F \Phi_{1:R,r}^{(1)\top} + \frac{1}{2} Q_r^{-1\top}\Sigma_F\Phi_{1:R,r}^{(2)\top} \\ &- \left(H_r^\top \Sigma_F \bar{C}_2^{(r)} + H_r^\top\Sigma_F\bar{C}_3^{(r)} + \bar{C}_1^{(r)}\Sigma_F\bar{C}_1^{(r)\top}\right) H_r^{-1}\Phi_{1:R,r}^\top \Bigg]\mu_{21}h_0^2. \end{aligned}$$

Similar to the derivation of (A.3),

$$H_r^\top = H_r^\top(\tilde{\Sigma}_F^{(r)}H_r)(\tilde{\Sigma}_F^{(r)}H_r)^{-1} = (\tilde{\Sigma}_F^{(r)}H_r)^{-1} + O_p([(Th_0)^{-1} + N^{-1}]\ln T + h_0^2).$$

Premultiplying H_r on both sides of (A.10) yields

$$H_r\tilde{\Phi}_{1:R,r}^\top - \Phi_{1:R,r}^\top = \tilde{\Sigma}_F^{(r)-1}T^{-1}\sum_{s=1}^{T} k_{h,sr}^* F_s e_{x,(1:R)s}^\top + H_r B_\Phi^\top(1:R,r) + o_p((Th_0)^{-1/2}). \tag{A.11}$$

Now, by the QR decomposition of $\tilde{\Phi}_{1:R,r}^\top$,

$$\tilde{\Sigma}_F^{(r)1/2}H_r\tilde{\Phi}_{1:R,r}^\top = \tilde{\Sigma}_F^{(r)1/2}H_r\tilde{Q}_r^\top\tilde{R}_r = (\tilde{\Sigma}_F^{(r)1/2}H_r\tilde{Q}_r^\top - I_R)\tilde{R}_r + \tilde{R}_r = \delta_r\tilde{R}_r + \tilde{R}_r.$$

Subtracting $\Phi_{1:R,r}^\top$ from both ends of the last equation and solving for δ_r yields

$$\delta_r = (\Phi_{1:R,r}^\top - \tilde{R}_r)\tilde{R}_r^{-1} + (\tilde{\Sigma}_F^{(r)1/2}H_r\tilde{\Phi}_{1:R,r}^\top - \Phi_{1:R,r}^\top)\tilde{R}_r^{-1}.$$

Similar to the argument in Bai and Ng (2013), since both $\Phi_{1:R,r}^\top$ and $\tilde{R}_r$ are upper triangular, the below diagonal elements of δ_r are determined by those of the second term on the right hand side of the last equation. Moreover, since δ_r is skew-symmetric up to an asymptotically negligible term,

[1] A matrix A is skew-symmetric (also known as anti-symmetric) if $A + A^\top = 0$. So the diagonal elements of a skew-symmetric matrix are zero, and $A_{ij} = -A_{ji}$

the elements of δ_r above the diagonal are also given, that is,

$$[\delta_r]_{k,l} = \begin{cases} \left[\left(\tilde{\Sigma}_F^{(r)1/2} H_r \tilde{\Phi}_{1:R,r}^\top - \Phi_{1:R,r}^\top\right) \tilde{R}_r^{-1}\right]_{k,l} & \text{if } k>l \\ O_p([(Th_0)^{-1} + N^{-1}] \ln T + h_0^2) & \text{if } k=l \\ -[\delta_r]_{l,k} + O_p([(Th_0)^{-1} + N^{-1}] \ln T + h_0^2) & \text{if } k<l. \end{cases} \tag{A.12}$$

Therefore, we only need to consider the first case with $k > l$. Noting that by (A.11)

$$\begin{aligned} & \left(\tilde{\Sigma}_F^{(r)1/2} H_r \tilde{\Phi}_{1:R,r}^\top - \Phi_{1:R,r}^\top\right) (\Phi_{1:R,r}^\top)^{-1} \\ = \; & \tilde{\Sigma}_F^{(r)1/2} \left(H_r \tilde{\Phi}_{1:R,r}^\top - \Phi_{1:R,r}^\top\right) (\Phi_{1:R,r}^\top)^{-1} + (\tilde{\Sigma}_F^{(r)1/2} - I_R) \\ = \; & \tilde{\Sigma}_F^{(r)-1/2} \left(T^{-1} \sum_{s=1}^T k_{h,sr}^* F_s e_{x,(1:R)s}^\top\right) (\Phi_{1:R,r}^\top)^{-1} + \tilde{\Sigma}_F^{(r)1/2} H_r B_\Phi^\top(1:R,r)(\Phi_{1:R,r}^\top)^{-1} + (\tilde{\Sigma}_F^{(r)1/2} - I_R) \\ & + o_p((Th_0)^{-1/2}). \end{aligned} \tag{A.13}$$

Note that by the Binomial theorem (see, e.g., Fact 10.11.24 in Bernstein (2005)),

$$\begin{aligned} \tilde{\Sigma}_F^{(r)1/2} &= \left(I_R - (I_R - \tilde{\Sigma}_F^{(r)})\right)^{1/2} = I_R + \frac{1}{2}(\tilde{\Sigma}_F^{(r)} - I_R) + o_p((Th_0)^{-1/2}), \\ \tilde{\Sigma}_F^{(r)-1/2} &= \left(I_R - (I_R - \tilde{\Sigma}_F^{(r)})\right)^{-1/2} = I_R - \frac{1}{2}(\tilde{\Sigma}_F^{(r)} - I_R) + o_p((Th_0)^{-1/2}). \end{aligned} \tag{A.14}$$

In addition, by the uniqueness of QR decomposition, $\tilde{R}_r \xrightarrow{p} \Phi_{1:R,r}^\top$. Consequently, we have that by Assumption 4.4(iii) and (iv), (A.13), the second equation in (A.14), and the fact that $\tilde{\Sigma}_F^{(r)} - I_R = o_p(1)$ under (IR) in (2.3),

$$\begin{aligned} & \sqrt{Th_0}\tilde{\Sigma}_F^{(r)-1/2} \left(\tilde{\Sigma}_F^{(r)1/2} H_r \tilde{\Phi}_{1:R,r}^\top - \Phi_{1:R,r}^\top\right) \tilde{R}_r^{-1} - \sqrt{Th_0} H_r B_\Phi^\top(1:R,r)(\Phi_{1:R,r}^\top)^{-1} \\ = \; & \tilde{\Sigma}_F^{(r)-1} \left(\sqrt{h_0/T} \sum_{s=1}^T k_{h,sr}^* F_s e_{x,(1:R)s}^\top\right) \left[(\Phi_{1:R,r}^\top)^{-1} + o_p(1)\right] + \tilde{\Sigma}_F^{(r)-1/2} \sqrt{Th_0}(\tilde{\Sigma}_F^{(r)1/2} - I_R) + o_p(1) \\ = \; & (I_R + o_p(1)) \left(\sqrt{h_0/T} \sum_{s=1}^T k_{h,sr}^* F_s e_{x,(1:R)s}^\top\right) (\Phi_{1:R,r}^\top)^{-1} - \sqrt{Th_0}(\tilde{\Sigma}_F^{(r)-1/2} - I_R) + o_p(1) \\ = \; & \left(\sqrt{h_0/T} \sum_{s=1}^T k_{h,sr}^* F_s e_{x,(1:R)s}^\top\right) (\Phi_{1:R,r}^\top)^{-1} + \frac{1}{2}\sqrt{Th_0}(\tilde{\Sigma}_F^{(r)} - I_R) + o_p(1) \\ \xrightarrow{d} \; & \mathbb{Z}_{1:R,r}(\Phi_{1:R,r}^\top)^{-1} + \frac{1}{2}\boldsymbol{N}_r. \end{aligned}$$

Now, by the definition of δ_r, $H_r \tilde{Q}_r^\top = \tilde{\Sigma}_F^{(r)-1/2}(\delta_r + I_R)$. Then by (A.14) and Assumption 4.4(iv), we have that under IR,

$$
\begin{aligned}
&\sqrt{Th_0}(H_r^* - I_R - H_r B_\Phi^\top(1:R,r)(\Phi_{1:R,r}^\top)^{-1}) \\
=&\sqrt{Th_0}\left[H_r\tilde{Q}_r^\top - I_R - H_r B_\Phi^\top(1:R,r)(\Phi_{1:R,r}^\top)^{-1}\right] \\
=&\sqrt{Th_0}\tilde{\Sigma}_F^{(r)-1/2}\delta_r - \sqrt{Th_0}H_r B_\Phi^\top(1:R,r)(\Phi_{1:R,r}^\top)^{-1} + \sqrt{Th_0}(\tilde{\Sigma}_F^{(r)-1/2} - I_R) \\
=&\sqrt{Th_0}\tilde{\Sigma}_F^{(r)-1/2}\delta_r - \sqrt{Th_0}H_r B_\Phi^\top(1:R,r)(\Phi_{1:R,r}^\top)^{-1} - \frac{1}{2}\sqrt{Th_0}(\tilde{\Sigma}_F^{(r)} - I_R) + o_p(1) \\
\xrightarrow{d}& \bar{\Delta}_r - \frac{1}{2}\boldsymbol{N}_r.
\end{aligned}
$$

where $\bar{\Delta}_r$ is defined in the statement of Theorem 4.3. ■

Proof of Theorem 4.4.

(i) By equation (3.3) in the main text, we have

$$\hat{F}_t - F_t = \tilde{Q}_t(\tilde{F}_t - H_t^\top F_t) + (\tilde{Q}_t H_t^\top - I_R)F_t.$$

Note that $\tilde{F}_t - H_t^\top F_t = A_1(t) + A_2(t) + A_3(t)$ by (A.8). By Lemma A.3, $\sqrt{N}A_2(t) = o_p(1)$ and $\sqrt{N}A_3(t) = (N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}\sqrt{N}\check{H}_t^\top F_t + o_p(1)$, where $\check{H}_t^\top$ is defined in the statement of Lemma A.3. It follows that

$$\sqrt{N}(\tilde{F}_t - H_t^\top F_t) = (N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}H_t^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \phi_{it}e_{x,it} - (N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}\sqrt{N}\check{H}_t^\top F_t + o_p(1).$$

Now, by the fact that $\tilde{Q}_t$ is an orthogonal matrix and Theorem 4.1, we have that

$$\tilde{Q}_t = \tilde{Q}_t H_t^{-1}H_t = (\tilde{Q}_t^\top)^{-1}H_t^{-1}H_t = (H_t\tilde{Q}_t^\top)^{-1}H_t = H_t^{*-1}H_t = H_t + o_p(1). \tag{A.15}$$

It follows by Lemma A.6 (ii) in Su and Wang (2020) and Lemma A.3(i) that

$$
\begin{aligned}
&\sqrt{N}(\hat{F}_t - F_t) \\
=\;& \tilde{Q}_t\sqrt{N}(\tilde{F}_t - H_t^\top F_t) + \sqrt{N}(\tilde{Q}_t H_t^\top - I_R)F_t \\
=\;& \tilde{Q}_t(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}H_t^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \phi_{it}e_{x,it} - \tilde{Q}_t(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}\sqrt{N}\check{H}_t^\top F_t + \sqrt{N}(\tilde{Q}_t H_t^\top - I_R)F_t + o_p(1) \\
=\;& H_t(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}H_t^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \phi_{it}e_{x,it}\{1 + o_p(1)\} - H_t(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}\sqrt{N}\check{H}_t^\top F_t\{1 + o_p(1)\} \\
&+\sqrt{N/(Th_0)}\sqrt{Th_0}(H_t^* - I_R)^\top F_t + o_p(1) \\
=\;& P_t^{-1}(P_t\Sigma_{\Phi_t}P_t^\top)^{-1}P_t\frac{1}{\sqrt{N}}\sum_{i=1}^N \phi_{it}e_{x,it}\{1 + o_p(1)\} - P_t^{-1}(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}\sqrt{N}\check{H}_t^\top F_t\{1 + o_p(1)\}
\end{aligned}
$$

$$
\begin{aligned}
&+\sqrt{N/(Th_0)}\sqrt{Th_0}(H_t^* - I_R)^\top F_t + o_p(1) \\
=\;& (P_t^\top P_t)^{-1}\Sigma_{\Phi_t}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi_{it}e_{x,it}\{1+o_p(1)\} - P_t^{-1}(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}\sqrt{N}\check{H}_t^\top F_t\{1+o_p(1)\} \\
&+\sqrt{N/(Th_0)}\sqrt{Th_0}(H_t^* - I_R)^\top F_t + o_p(1).
\end{aligned}
$$

Again, by (A.5), H_t is an orthogonal matrix up to a term that is of order $O_p((Th_0)^{-1/2} + N^{-1}\ln T + h_0^2)$. Then, by Lemma A.6(ii) in Su and Wang (2020), it is easy to show that P_t is also an orthogonal matrix up to a term that is of order $O_p((Th_0)^{-1/2} + N^{-1}\ln T + h_0^2)$. Thus,

$$
\begin{aligned}
\sqrt{N}(\hat{F}_t - F_t) \;=\;& \{I_R + o_p(1)\}\,\Sigma_{\Phi_t}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi_{it}e_{x,it} - P_t^{-1}(N^{-1}\tilde{\Phi}_t^\top\tilde{\Phi}_t)^{-1}\sqrt{N}\check{H}_t^\top F_t \\
&+\sqrt{N/(Th_0)}\sqrt{Th_0}(H_t^* - I_R - H_tB_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1})^\top F_t \\
&+\sqrt{N}H_tB_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1} + o_p(1).
\end{aligned}
$$

In addition, under the condition that $Nh_0^4 \to 0$ and by the definition of $\check{H}_t$, the bias term $\sqrt{N}\check{H}_t^\top F_t$ is asymptotically vanishing and $\sqrt{N}H_tB_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1} = O(\sqrt{N}h_0^2) = o(1)$. Then, the first part of Theorem 4.4 follows from Assumption 4.4 and Theorem 4.3. On the other hand, if $N(Th_0)^{-1} \to \infty$ and $Th_0^5 \to 0$, then, by Lemma A.1(ii), we have

$$
\begin{aligned}
\sqrt{Th_0}(\hat{F}_t - F_t) \;=\;& \tilde{Q}_t\sqrt{Th_0}(\tilde{F}_t - H_t^\top F_t) + \sqrt{Th_0}(\tilde{Q}_tH_t^\top - I_R)F_t \\
=\;& \sqrt{Th_0}(H_t^* - I_R - H_tB_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1})^\top F_t \\
&+\sqrt{Th_0}H_tB_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1} + o_p(1).
\end{aligned}
$$

Therefore, the second part of Theorem 4.4 follows from Theorem 4.3 and the fact that $\sqrt{Th_0}H_tB_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1} = O(\sqrt{Th_0}h_0^2) = O(Th_0^5) = o(1)$.

(ii) By using arguments as used in the proof of Theorem 2.2 in Su and Wang (2020), we can show that

$$
\tilde{\phi}_{it} - H_t^{-1}\phi_{it} = H_t^\top T^{-1}\sum_{s=1}^{T}k_{h,st}^*F_se_{x,is} + B_\Phi(i,t) + o_p((Th_0)^{-1/2}).
$$

Then, by the condition $Th_0^5 \to 0$, the definition of $B_\Phi(i,t)$ and (A.15), we have

$$
\begin{aligned}
\sqrt{Th_0}(\hat{\phi}_{it} - \phi_{it}) \;=\;& (\tilde{Q}_t^\top)^{-1}\sqrt{Th_0}(\tilde{\phi}_{it} - H_t^{-1}\phi_{it}) + \sqrt{Th_0}\left[(H_t\tilde{Q}_t^\top)^{-1} - I_R\right]\phi_{it} \\
=\;& (\tilde{Q}_t^\top)^{-1}H_t^\top\sqrt{h_0/T}\sum_{s=1}^{T}k_{h,st}^*F_se_{x,is} + \sqrt{Th_0}(H_t^{*-1} - I_R)\phi_{it} + o_p(1) \\
=\;& (H_t\tilde{Q}_t^{-1})^\top\sqrt{h_0/T}\sum_{s=1}^{T}k_{h,st}^*F_se_{x,is} - \sqrt{Th_0}(H_t^* - I_R)(H_t^{*-1} - I_R + I_R)\phi_{it} + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= \{I_R + o_p(1)\} \sqrt{h_0/T} \sum_{s=1}^{T} k_{h,st}^* F_s e_{x,is} - \sqrt{Th_0}(H_t^* - I_R)\{I_R + o_p(1)\}\phi_{it} + o_p(1) \\
&= \sqrt{h_0/T} \sum_{s=1}^{T} k_{h,st}^* F_s e_{x,is} - \sqrt{Th_0}(H_t^* - I_R - H_t B_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1})\phi_{it} \\
&\quad -\sqrt{Th_0} H_t B_\Phi^\top(1:R,t)(\Phi_{1:R,t}^\top)^{-1}\phi_{it} + o_p(1) \\
&\xrightarrow{d} \mathbb{Z}_{i,t} - \Xi_t \phi_{it}.
\end{aligned}
$$

This completes the proof of (ii). ■

Lemma A.5. *Suppose that Assumptions 4.1–4.3 and 4.5 hold. Let $\{(\varepsilon_{y,t}, F_t) : t \geq 1\}$ be a strictly stationary and α-mixing process with mixing coefficient as in Assumption 4.5(ii). Then*

$$
\sup_{z\in[0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T} \left[F_t F_t^\top - E(F_t F_t^\top) \right] K_{h_1}(z_t - z) \right\| = O_p\left(\left(\frac{\ln T}{Th_1} \right)^{1/2} \right).
$$

Lemma A.6. *Let $r_u(z_s)_{(k,l)}$ denote the (k,l)th element of $\boldsymbol{R}_u(z_s)$, where*

$$
\boldsymbol{R}_u(z_s) = Cov(\sigma(F_t, z_s)\varepsilon_{y,t} F_t^a, \sigma(F_{t+u}, z_s)\varepsilon_{y,t+u} F_{t+u}^a).
$$

Suppose that Assumptions 4.1–4.3 and 4.5 hold, then, $\sum_{u=-\infty}^{\infty} |r_u(z_s)_{(k,l)}| < \infty$.

Lemma A.7. *Suppose that Assumptions 4.1–4.3 and 4.5 hold. Denote*

$$
\begin{aligned}
\hat{e}_0(z_s) &= T^{-1/2} h_1^{1/2} \sum_{t=1}^{T} \hat{F}_t^a \left\{ \hat{F}_t^{a\top}[\beta(z_t) - \theta_0^0(z_s) - \theta_1^0(z_s)(z_t - z_s)] + e_{y,t+1} \right\} K_{h_1}(z_t - z_s), \text{ and} \\
\hat{e}_1(z_s) &= T^{-1/2} h_1^{1/2} \sum_{t=1}^{T} \hat{F}_t^a \left\{ \hat{F}_t^{a\top}[\beta(z_t) - \theta_0^0(z_s) - \theta_1^0(z_s)(z_t - z_s)] + e_{y,t+1} \right\} \left(\frac{z_t - z_s}{h_1} \right) K_{h_1}(z_t - z_s),
\end{aligned}
$$

where $\hat{F}_t^a$, $\theta_0^0(\cdot)$ and $\theta_1^0(\cdot)$ are defined in Section 3.2 in the main text. Let $\hat{e}(z_s) = (\hat{e}_0(z_s)^\top, \hat{e}_1(z_s)^\top)^\top$ and $\hat{e} = (\hat{e}(z_1), \ldots, \hat{e}(z_T))^\top \in \mathbb{R}^{T\times 2(R+1)}$. Then, $T^{-1}\|\hat{e}\|^2 = O_p(1)$.

Lemma A.8. *Suppose that Assumptions 4.1–4.3 and 4.5 hold. Then,*

$$
\begin{aligned}
T^{-1} \sum_{t=1}^{T} \|\hat{\beta}_\lambda(z_t) - \beta(z_t)\|^2 &= O_p((Th_1)^{-1}) \text{ and} \\
T^{-1} \sum_{t=1}^{T} \|\hat{\theta}_1(z_t) - \theta_1^0(z_t)\|^2 &= O_p((Th_1)^{-1} h_1^{-2}),
\end{aligned}
$$

where $\hat{\theta}_1(z_t)$ is defined in Section 3.2 in the main text.

Lemma A.9. *Suppose that Assumptions 4.1–4.3 and 4.5 hold. Then,* $P\left(\|\hat{\theta}_{0,l}\| = 0\right) \to 1$ *and* $P\left(\|h_1\hat{\theta}_{1,l}\| = 0\right) \to 1$ *for any* $R_f < l \le R$.

Denote $\hat{F}^{a*}_{a,t} = (\hat{F}^{a\top}_{a,t}, ((z_t - z_s)/h_1)\hat{F}^{a\top}_{a,t})^\top$, where $\hat{F}^{a}_{a,t}$ is defined in Section 4.4 in the main text and define $\hat{\theta}_a(z) = (\hat{\theta}_{a,0}(z)^\top, h_1\hat{\theta}_{a,1}(z)^\top)^\top$, where $\hat{\theta}_{a,0}(z) = (\hat{\theta}_{0,0}(z), \hat{\theta}_{0,1}(z), \ldots, \hat{\theta}_{0,R_f}(z))^\top$ and $\hat{\theta}_{a,1}(z) = (\hat{\theta}_{1,0}(z), \hat{\theta}_{1,1}(z), \ldots, \hat{\theta}_{1,R_f}(z))^\top$. Similarly, define $\hat{\theta}_b(z) = (\hat{\theta}_{b,0}(z)^\top, h_1\hat{\theta}_{b,1}(z)^\top)^\top$, where $\hat{\theta}_{b,0}(z) = (\hat{\theta}_{0,(R_f+1)}(z), \ldots, \hat{\theta}_{0,R}(z))^\top$ and $\hat{\theta}_{b,1}(z) = (\hat{\theta}_{1,(R_f+1)}(z), \ldots, \hat{\theta}_{1,R}(z))^\top$. Finally, recall that $\theta^0_{r,l} = (\theta^0_{r,l}(z_1), \ldots, \theta^0_{r,l}(z_T))^\top \in \mathbb{R}^T$ is the vector of the true values for $r = 0, 1$.

Proof of Theorem 4.5.

(a) Recall that $\hat{F}_{0t} \equiv 1$, let $\hat{F}_{a,lt}$ be the $(l+1)$th element of $\hat{F}^a_t$ for $l \in 0 \cup [R_f]$ and $\hat{F}_{b,lt}$ be the $(l+1)$th element of $\hat{F}^a_t$ for $l > R_f$. By Lemma A.9 and Hunter and Li (2005), we know that as $m \to \infty$, $\|\hat{\theta}^{(m)}_{0,l}\| \xrightarrow{p} \|\hat{\theta}_{0,l}\|$ for every $0 \le l \le R$, where $\hat{\theta}^{(m)}_{0,l}$ is defined in Section 5.1 and $\hat{\theta}_{0,l}$ is defined in (3.4) in the main text. Then, we must have $\|\hat{\theta}^{(m)}_{0,l}\|$ converge to a positive value in probability for every $l \le R_f$, while $\|\hat{\theta}^{(m)}_{0,l}\| \xrightarrow{p} 0$ for every $l > R_f$. To this end, let $B^{(m)}_l = P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)/c_l$. If $0 \le l \le R_f$, we define $B^{(m)}_l \equiv B^{(m)}_{a,l}$; otherwise, let $B^{(m)}_l \equiv B^{(m)}_{b,l}$. By the definition of $B^{(m)}_l$, $B^{(m)}_{a,l}$ must converge to a finite value in probability, whereas $B^{(m)}_{b,l}$ must diverge to infinity in probability, as $m \to \infty$.

Now, we begin to study the convergence of $\hat{\theta}^{(m+1)}_{0,l}(z_s)$ in Section 5.1 of the main text. By (5.1) in the paper, we have $\hat{\theta}^{(m+1)}_{0,l}(z_s) = \left\{\tilde{\Omega}_{Tl,s}(0) + B^{(m)}_l\right\}^{-1}\tilde{E}_{Tl,s}$, and recall that

$$\begin{aligned}
\tilde{\Omega}_{Tl,s}(0) &= \frac{2}{T}\sum_{t=1}^{T}\hat{F}^2_{lt}K_{h_1}(z_t - z_s) \text{ and} \\
\tilde{E}_{Tl,s} &= \frac{2}{T}\sum_{t=1}^{T}\hat{F}_{lt}\hat{y}^{(m)}_{t+1,-l}K_{h_1}(z_t - z_s),
\end{aligned}$$

where $\hat{y}^{(m)}_{t+1,-l}$ is defined in Section 5.1. Let $\hat{\theta}^{(m+1)}_{a,0,l}(z_s) = \left\{\tilde{\Omega}_{a,Tl,s}(0) + B^{(m)}_{a,l}\right\}^{-1}\tilde{E}_{a,Tl,s}$ and $\hat{\theta}^{(m+1)}_{b,0,l}(z_s) = \left\{\tilde{\Omega}_{b,Tl,s}(0) + B^{(m)}_{b,l}\right\}^{-1}\tilde{E}_{b,Tl,s}$ where

$$\begin{aligned}
\tilde{\Omega}_{a,Tl,s}(0) &= \frac{2}{T}\sum_{t=1}^{T}\hat{F}^2_{a,lt}K_{h_1}(z_t - z_s),\ \tilde{E}_{a,Tl,s} = \frac{2}{T}\sum_{t=1}^{T}\hat{F}_{a,lt}\hat{y}^{(m)}_{t+1,-l}K_{h_1}(z_t - z_s), \ \text{ for } \ l \in 0 \cup [R_f] \\
\tilde{\Omega}_{b,Tl,s}(0) &= \frac{2}{T}\sum_{t=1}^{T}\hat{F}^2_{b,lt}K_{h_1}(z_t - z_s), \text{ and } \tilde{E}_{b,Tl,s} = \frac{2}{T}\sum_{t=1}^{T}\hat{F}_{b,lt}\hat{y}^{(m)}_{t+1,-l}K_{h_1}(z_t - z_s), \ \text{ for } \ l > R_f.
\end{aligned}$$

By Corollary 4.2, $\tilde{\Omega}_{\ell,Tl,s}(0) = \mathcal{M}_{\ell,l}(z_s) + o_p(1)$ and $\tilde{E}_{\ell,Tl,s} = \mathcal{N}_{\ell,l}(z_s) + o_p(1)$ for $\ell \in \{a, b\}$, uniformly on $[0,1]$, where $\mathcal{M}_{\ell,l}(z_s) = 2\sum_{t=1}^{T}F^2_{\ell,lt}K_{h_1}(z_t - z_s)/T$ and $\mathcal{N}_{\ell,l}(z_s) = 2\sum_{t=1}^{T}F_{\ell,lt}\hat{y}^{(m)}_{t+1,-l}K_{h_1}(z_t -$

$z_s)/T$. Then, we only need to consider the convergence of $\left\{\mathcal{M}_{a,l}(z_s)+B_{a,l}^{(m)}\right\}^{-1}\mathcal{N}_{a,l}(z_s)$ and $\left\{\mathcal{M}_{b,l}(z_s)+B_{b,l}^{(m)}\right\}^{-1}\mathcal{N}_{b}(z_s)$ and the rest of the proof is the same as that of Theorem 1 in Wang and Xia (2009).

(b) By Lemma A.9, we know immediately that $\hat{\theta}_b(z)=0$ with probability approaching one (w.p.a.1). Consequently, we know that $\hat{\theta}_a(z)$ must be the solution to the following normal equation:

$$\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{a,t}^{a*}\Big(y_{t+1}-\hat{F}_{a,t}^{a*\top}\hat{\theta}_a(z)\Big)K_{h_1}(z_s-z)+\sum_{l=0}^{R_f}\begin{pmatrix}P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)\hat{\theta}_{0,l}/\|\hat{\theta}_{0,l}\|\\ P'_{\lambda_T}(\tilde{D}_l)h_1\hat{\theta}_{1,l}/\|h_1\hat{\theta}_{1,l}\|\end{pmatrix}=0.$$

By the standard result of local linear smoothing, we can show that $\max_{1\le l\le R_f}\|\tilde{\theta}_{0,l}-\theta_{0,l}^0\|=O_p(h_1^{-1/2})$. Then, by Assumption 4.5(iv), it can be shown that

$$\min_{1\le l\le R_f}\|\tilde{\theta}_{0,l}\|\ge\min_{1\le l\le R_f}\|\theta_{0,l}^0\|-\max_{1\le l\le R_f}\|\tilde{\theta}_{0,l}-\theta_{0,l}^0\|\ge a\lambda_T \text{ w.p.a.1},$$

with probability approaching one, which together with the SCAD structure, indicates that the penalty term $P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)$ is asymptotically negligible. Similarly, $P'_{\lambda_T}(\tilde{D}_l)=0$ when $\tilde{D}_l\neq 0$ and when T is large. Thus, we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{a,t}^{a*}\Big(y_{t+1}-\hat{F}_{a,t}^{a*\top}\hat{\theta}_a(z)\Big)K_{h_1}(z_s-z)=0,$$

which implies that

$$\hat{\theta}_a(z)=\left\{\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{a,t}^{a*}\hat{F}_{a,t}^{a*\top}K_{h_1}(z_t-z)\right\}^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{a,t}^{a*}y_{t+1}K_{h_1}(z_t-z).$$

Then,

$$\begin{aligned}\hat{\beta}_{a,\lambda}(z)&=[I_{R_f+1},\mathbf{0}_{(R_f+1)\times(R_f+1)}]\hat{\theta}_a(z)\\&=[I_{R_f+1},\mathbf{0}_{(R_f+1)\times(R_f+1)}]\Big(T^{-1}\sum_{t=1}^{T}\hat{F}_{a,t}^{a*}\hat{F}_{a,t}^{a*\top}K_{h_1}(z_t-z)\Big)^{-1}\Big(T^{-1}\sum_{t=1}^{T}\hat{F}_{a,t}^{a*}y_{t+1}K_{h_1}(z_t-z)\Big).\end{aligned}$$

Therefore, by Lemma A.6 and Assumptions 4.5, the rest of the proof follows directly from that of Theorem 2 in Cai (2007). ■

B Proofs of the technical lemmas in Section A

Proof of Lemma A.1-A.3. The proofs are completely analogous to that of Theorem 2.5, Lemma A.4 and Lemma A.5 in Su and Wang (2020), respectively, and thus omitted. ■

Proof of Lemma A.4. We only prove (vii) and (x), because the proof of the rest of the lemma is analogous to that of Lemma A.6 in Su and Wang (2020).

(vii) Note that $T^{-1}B^{(r)\top}F^{(r)}H^{(r)} = T^{-1}\sum_{t=1}^{T} B_t^{(r)}F_t^{(r)\top}H^{(r)} \leq T^{-1}\sum_{t=1}^{T}[\bar{C}_1^{(r)}(\frac{t-r}{T})+\bar{C}_2^{(r)}\mu_{21}h_0^2+\bar{C}_3^{(r)}(\frac{t-r}{T})^2]F_t^{(r)}F_t^{(r)\top}H^{(r)} \equiv \sum_{c=1}^{3} C_{c,r}$. For $C_{1,r} = T^{-1}\sum_{t=1}^{T}[\bar{C}_1^{(r)}(\frac{t-r}{T})]k_{h_0,tr}^{*}F_tF_t^{\top}H_r$. Then, by the structure of the boundary kernel, we have

$$\begin{aligned}
\max_r \|C_{1,r}\| &\leq \max_r \|\bar{C}_1^{(r)}\| \max_r \|H_r\| \max_r \left\| T^{-1}\sum_{t=1}^{T}\left(\frac{t-r}{T}\right)k_{h_0,tr}^{*}F_tF_t^{\top}\right\| \\
&\lesssim \max_r \left\| T^{-1}\sum_{t=1}^{T}\left(\frac{t-r}{T}\right)k_{h_0,tr}^{*}F_tF_t^{\top}\right\| \\
&\leq \max_r \left\| T^{-1}\sum_{t=1}^{T}\left(\frac{t-r}{T}\right)k_{h_0,tr}^{*}\Sigma_F\right\| + \max_r \left\| T^{-1}\sum_{t=1}^{T}\left(\frac{t-r}{T}\right)k_{h_0,tr}^{*}(F_tF_t^{\top}-\Sigma_F)\right\| \\
&= O_p((Th_0)^{-1}) + O_p((Th_0/\ln T)^{-1/2}h_0)
\end{aligned}$$

by the triangle inequality and the uniform approximation of Riemann summation to a definite integral. Similarly, one can show that $\max_r \|C_{c,r}\| = O_p((Th_0/\ln T)^{-1/2}h_0^2)$ for $c = 2, 3$. If follows that $\max_r \|T^{-1}B^{(r)\top}F^{(r)}H^{(r)}\| = O_p((Th_0/\ln T)^{-1/2}h_0)$.

(x) Notice that

$$\begin{aligned}
\left\| T^{-1}\tilde{F}^{(r)\top}B^{(r)}\right\| &\leq \left\| T^{-1}(\tilde{F}^{(r)} - F^{(r)}H^{(r)} - B^{(r)})^{\top}B^{(r)}\right\| + \left\| T^{-1}H^{(r)\top}F^{(r)\top}B^{(r)}\right\| \\
&\quad + \left\| T^{-1}B^{(r)\top}B^{(r)}\right\| \equiv \sum_{c=1}^{3} D_{c,r}.
\end{aligned}$$

First, following the proof of Lemma A.6(ix) in Su and Wang (2020), we can show that $\max_r D_{1,r} = \max_r \left\| T^{-1}(\tilde{F}^{(r)} - F^{(r)}H^{(r)} - B^{(r)})^{\top}B^{(r)}\right\| = O_p([(Th_0)^{-1}+N^{-1}\ln T]h_0)+o_p(h_0^4)$. Second, similar to the proof of Lemma A.6(vii) in Su and Wang (2020), we have $\max_r D_{2,r} = \max_r \left\| T^{-1}H^{(r)\top}F^{(r)\top}B^{(r)}\right\|$ $= O_p((Th_0)^{-1} + (Th_0/\ln T)^{-1/2}h_0 + h_0^2)$ and $\max_r D_{3,r} = \max_r \left\| T^{-1}B^{(r)\top}B^{(r)}\right\| = O_p(h_0^2)$. Thus, $\max_r \left\| T^{-1}\tilde{F}^{(r)\top}B^{(r)}\right\| = O_p((Th_0)^{-1} + (Th_0/\ln T)^{-1/2}h_0 + h_0^2 + N^{-1}h_0 \ln T)$. ■

Proof of Lemma A.5. The proof is analogous to that of Theorem 2 in Hansen (2008) and thus omitted. ■

Proof of Lemma A.6. The proof follows from that of Lemma 1 in Cai (2007) and thus omitted. ■

Proof of Lemma A.7. Let $\hat{e}_0 = (\hat{e}_0(z_1), \ldots, \hat{e}_0(z_T))^\top$ and $\hat{e}_1 = (\hat{e}_1(z_1), \ldots, \hat{e}_1(z_T))^\top$. Noting that

$$T^{-1}\|\hat{e}\|^2 = T^{-1}\sum_{s=1}^{T}\|\hat{e}(z_s)\|^2 = T^{-1}\sum_{s=1}^{T}(\|\hat{e}_0(z_s)\|^2 + \|\hat{e}_1(z_s)\|^2),$$

it suffices to prove that $T^{-1}\|\hat{e}_\ell\|^2 = O_p(1)$ for $\ell = 0$ and 1. Below we focus on the proof for the case of $\ell = 0$ as the proof is analogous for $\ell = 1$.

Noting that $\hat{F}_t^a\hat{F}_t^{a\top} = F_t^aF_t^{a\top} + F_t^a(\hat{F}_t^a - F_t^a)^\top + (\hat{F}_t^a - F_t^a)^\top F_t^{a\top} + (\hat{F}_t^a - F_t^a)(\hat{F}_t^a - F_t^a)^\top$ and $\hat{F}_t^a = F_t^a + (\hat{F}_t^a - F_t^a)$, we have

$$\begin{aligned}
T^{-1}\|\hat{e}_0\|^2 &= T^{-1}\sum_{s=1}^{T}\|\hat{e}_0(z_s)\|^2 \\
&\lesssim T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}\hat{F}_t^a\hat{F}_t^{a\top}\varsigma_{ts}K_{h_1,ts}\right\|^2 + T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}\hat{F}_t^a e_{y,t+1}K_{h_1,ts}\right\|^2 \\
&\lesssim T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}F_t^aF_t^{a\top}\varsigma_{ts}K_{h_1,ts}\right\|^2 + T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}F_t^a e_{y,t+1}K_{h_1,ts}\right\|^2 \\
&\quad + T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}(\hat{F}_t^a - F_t^a)e_{y,t+1}K_{h_1,ts}\right\|^2 + T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}F_t^a(\hat{F}_t^a - F_t^a)^\top\varsigma_{ts}K_{h_1,ts}\right\|^2 \\
&\quad + T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}(\hat{F}_t^a - F_t^a)F_t^{a\top}\varsigma_{ts}K_{h_1,ts}\right\|^2 + T^{-2}h_1\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}(\hat{F}_t^a - F_t^a)(\hat{F}_t^a - F_t^a)^\top\varsigma_{ts}K_{h_1,ts}\right\|^2 \\
&\equiv A + B + D_1 + D_2 + D_3 + D_4,
\end{aligned}$$

where $\varsigma_{ts} \equiv \beta(z_t) - \theta_0^0(z_s) - \theta_1^0(z_s)(z_t - z_s)$ and $K_{h_1,ts} \equiv K_{h_1}(z_t - z_s)$.

First, we consider A. Notice that

$$\begin{aligned}
A &= T^{-2}h_1\sum_{s=1}^{T}\sum_{t\neq u}\varsigma_{ts}^\top F_t^aF_t^{a\top}F_u^aF_u^{a\top}\varsigma_{us}K_{h_1,ts}K_{h_1,us} + T^{-2}h_1\sum_{s=1}^{T}\sum_{t}\varsigma_{ts}^\top F_t^aF_t^{a\top}F_t^aF_t^{a\top}\varsigma_{ts}K_{h_1,ts}^2 \\
&= T^{-2}h_1\sum_{s=1}^{T}\sum_{t\neq u}\varsigma_{ts}^\top F_t^aF_t^{a\top}F_u^aF_u^{a\top}\varsigma_{us}K_{h_1,ts}K_{h_1,us} + T^{-2}h_1\sum_{1\leq t\neq s\leq T}\varsigma_{ts}^\top F_t^aF_t^{a\top}F_t^aF_t^{a\top}\varsigma_{ts}K_{h_1,ts}^2 \\
&\quad + T^{-2}h_1\sum_{t=1}^{T}\varsigma_{tt}^\top F_t^aF_t^{a\top}F_t^aF_t^{a\top}\varsigma_{tt}K_{h_1,tt}^2 \\
&\equiv A_1 + A_2 + A_2.
\end{aligned}$$

Since $A \geq 0$, it suffices to prove $A = O_p(1)$ by showing that $E(A_l) = O(1)$ for $l \in [3]$. For A_1, note that z_t is non-random. Let $z_{th} = (z_t - z_s)/h_1$. By Taylor expansion of $\beta(z_t)$ at z_s, we have

$$\varsigma_{ts} \equiv \beta(z_t) - \theta_0^0(z_s) - \theta_1^0(z_s)(z_t - z_s) = \frac{1}{2}\beta^{(2)}(z_s + \zeta_{ts}h_1 z_{th})z_{th}^2 h_1^2$$

for some $\zeta_{ts} \in (0,1)$, where $\beta^{(2)}(\cdot)$ denotes the second order derivative of $\beta(\cdot)$. Then

$$\begin{aligned}
|E(A_1)| &= \left| T^{-2}h_1 \sum_{s=1}^{T}\sum_{t\neq u} \varsigma_{ts}^{\top} E\left[F_t^a F_t^{a\top} F_u^a F_u^{a\top}\right] \varsigma_{us} K_{h_1,ts}K_{h_1,us} \right| \\
&= \left| \frac{1}{4}T^{-2}h_1^5 \sum_{s=1}^{T}\sum_{t\neq u} \left[\beta^{(2)}(z_s + \zeta_{ts}h_1 z_{th})\right]^{\intercal} E(F_t^a F_t^{a\top} F_u^a F_u^{a\top}) \right. \\
&\quad \left. \times \beta^{(2)}(z_s + \zeta_{us}h_1 z_{uh}) z_{th}^2 z_{uh}^2 K_{h_1,ts}K_{h_1,us} \right| \\
&\lesssim Th_1^5 \left(\frac{1}{T}\sum_{s=1}^{T} z_{th}^2 K_{h_1,ts} \right)^2 = O(Th_1^5) = O(1).
\end{aligned}$$

where the last equality holds by Assumptions 4.5(iii). As for A_2, notice that by the Riemann sum approximation of an integral,

$$\begin{aligned}
|E(A_2)| &= \left| T^{-2}h_1 \sum_{1\le t\neq s\le T} \varsigma_{ts}^{\top} E\left[F_t^a F_t^{a\top} F_t^a F_t^{a\top}\right] \varsigma_{ts} K_{h_1,ts}^2 \right| \\
&= \left| \frac{1}{2}T^{-2}h_1^5 \sum_{1\le t\neq s\le T} \left[\beta^{(2)}(z_s + \zeta_{ts}h_1 z_{th})\right]^{\top} E(F_t^a F_t^{a\top} F_t^a F_t^{a\top}) \beta^{(2)}(z_s + \zeta_{ts}h_1 z_{th}) z_{th}^4 K_{h_1,ts}^2 \right| \\
&\lesssim h_1^5 \frac{1}{T^2} \sum_{1\le t\neq s\le T} z_{th}^4 K_{h_1,ts}^2 \lesssim h_1^4 \int_0^1 u^4 K^2(u)du = O(h_1^4) = o(1).
\end{aligned}$$

In addition, it is trivial to show that $E(A_3) = O(T^{-1}h_1^3) = o(1)$. Therefore, $E(A) = O(1)$, implying that $A = O_p(1)$ by the Markov inequality.

Now, we consider B. Note that

$$\begin{aligned}
B &= T^{-2}h_1 \sum_{s=1}^{T}\sum_{t,u} F_t^{a\top} F_u^a e_{y,t+1} e_{y,u+1} K_{h_1,ts}K_{h_1,us} \\
&= T^{-2}h_1 \sum_{s=1}^{T}\sum_{t\neq u} F_t^{a\top} F_u^a e_{y,t+1} e_{y,u+1} K_{h_1,ts}K_{h_1,us} + T^{-2}h_1 \sum_{s=1}^{T}\sum_{t=1}^{T} F_t^{a\top} F_u^a e_{y,t+1}^2 K_{h_1,ts}^2 \\
&= T^{-2}h_1 \sum_{s\neq t, s\neq u, u\neq t} F_t^{a\top} F_u^a e_{y,t+1} e_{y,u+1} K_{h_1,ts}K_{h_1,us}
\end{aligned}$$

$$
\begin{aligned}
&+ 2T^{-2}h_1 \sum_{t\neq u} F_u^{a\top} F_t^a e_{y,t+1} e_{y,u+1} K_{h_1,tt} K_{h_1,ut} + T^{-2}h_1 \sum_{s=1}^{T}\sum_{t=1}^{T} F_t^{a\top} F_t^a e_{y,t+1}^2 K_{h_1,ts}^2 \\
&\equiv B_1 + B_2 + B_3.
\end{aligned}
$$

As in the analysis of A, we prove that $B = O_p(1)$ by showing that $E(B_l) = O(1)$ for $l \in [3]$.

We first study $E(B_1)$. By the Davydov inequality (see, e.g., Corollary A.2 of Hall and Heyde (2014)), one has

$$
|\mathrm{Cov}(F_{l1}^a e_{y,2}, F_{l(u+1)}^a e_{y,(u+2)})| \le C\alpha^{\delta/(\delta+2)}(u)(E|F_{l1}^a e_{y,2}|^{\delta+2})^{1/(2+\delta)} \lesssim \alpha^{\delta/(\delta+2)}(u) \text{ for all } l, u. \quad \text{(B.1)}
$$

Noting that $K_{h_1,ts}K_{h_1,us} \le \frac{1}{2}(K_{h_1,ts}^2 + K_{h_1,us}^2)$ by the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
|E(B_1)| &= \left| T^{-2}h_1 \sum_{s\neq t, s\neq u, u\neq t} E\left[F_t^{a\top} F_u^a e_{y,t+1} e_{y,u+1}\right] K_{h_1,ts} K_{h_1,us} \right| \\
&\le T^{-2}h_1 \sum_{s\neq t, s\neq u, u\neq t} \left| E\left[F_t^{a\top} F_u^a e_{y,t+1} e_{y,u+1}\right]\right| K_{h_1,ts}^2 \lesssim \left(T^{-2}h_1 \sum_{s\neq t} K_{h_1,ts}^2\right) \sum_{\tau=1}^{\infty} \alpha^{\delta/(2+\delta)}(\tau) \\
&= O(1)O(1) = O(1).
\end{aligned}
$$

For B_2 and B_3, we have

$$
\begin{aligned}
|E(B_2)| &= \left| 2T^{-2}h_1 \sum_{t\neq u} E\left(F_u^{a\top} F_t^a e_{y,t+1} e_{y,u+1}\right) K_{h_1,tt} K_{h_1,ut} \right| \lesssim T^{-2} \sum_{t\neq u} K_{h_1,ut} = O(1), \\
|E(B_3)| &= \left| T^{-2}h_1 \sum_{s=1}^{T}\sum_{t=1}^{T} E\left(F_t^{a\top} F_t^a e_{y,t+1}^2\right) K_{h_1,ts}^2 \right| \lesssim T^{-2}h_1 \sum_{s=1}^{T}\sum_{t=1}^{T} K_{h_1,ts}^2 = O(1).
\end{aligned}
$$

In sum, we have shown that $E(B) = O(1)$.

Using the result in Corollary 4.2, we can show that $D_1 = o_p(B) = o_p(1)$ and $D_\ell = o_p(A) = o_p(1)$ for $\ell = 2, 3, 4$. Then the proof of the lemma is completed. ■

Proof of Lemma A.8

For an arbitrary matrix $A = (a_{kl})$, let $\|A\| = \{\sum_{k,l} a_{kl}^2\}^{1/2}$. Let $\boldsymbol{U} = (v_1^\top, w_1^\top, v_2^\top, w_2^\top, \ldots, v_T^\top, w_T^\top)^\top$, where $v_t = (v_{0t}, v_{1t}, \ldots, v_{Rt})^\top$ and $w_t = (w_{0t}, w_{1t}, \ldots, w_{Rt})^\top$. Define

$$
\mathcal{B}^*(C) = \left\{ \boldsymbol{U} : T^{-1}\sum_{t=1}^{T}(\|v_t\|^2 + \|w_t\|^2) = T^{-1}\|\boldsymbol{V}\|^2 + T^{-1}\|\boldsymbol{W}\|^2 = C \right\},
$$

where C is a positive constant which can be sufficiently large, $\boldsymbol{V} = (v_1^\top, \ldots, v_T^\top)^\top$, and $\boldsymbol{W} =$

$(w_1^\top, \ldots, w_T^\top)^\top$. Let

$$\Theta_\ell^0 = (\theta_{\ell,0}^0, \theta_{\ell,1}^0, \ldots, \theta_{\ell,R}^0) \text{ where } \theta_{\ell,l}^0 = (\theta_{\ell,0}^0(z_1), \ldots, \theta_{\ell,R}^0(z_T))^\top \text{ for } \ell = 0, 1.$$

Thus, it suffices to show that for any small probability $\epsilon > 0$, we can always find a large constant $C > 0$, such that

$$\liminf_T P\left(\inf_{\boldsymbol{U}\in\mathcal{B}^*(C)} Q_\lambda\left(\Theta_0^0 + (Th_1)^{-1/2}\boldsymbol{V}, \Theta_1^0 + (Th_1)^{-1/2}\boldsymbol{W}/h_1\right) > Q_\lambda\left(\Theta_0^0, \Theta_1^0\right)\right) = 1 - \epsilon, \quad \text{(B.2)}$$

where $Q_\lambda(\cdot, \cdot)$ is defined in (3.4) in the main text. Observe that

$$h_1\left\{Q_\lambda\left(\Theta_0^0 + (Th_1)^{-1/2}\boldsymbol{V}, \Theta_1^0 + (Th_1)^{-1/2}\boldsymbol{W}/h_1\right) - Q_\lambda\left(\Theta_0^0, \Theta_1^0\right)\right\} = \Pi_1 + \Pi_2 + \Pi_3, \quad \text{(B.3)}$$

where

$$\begin{aligned}
\Pi_1 &= \frac{h_1}{T}\sum_{t=1}^T\sum_{s=1}^T\left\{y_{t+1} - \hat{F}_t^{a\top}\left[\theta_0^0(z_s) + (Th_1)^{-1/2}v_t + (h_1\theta_1^0(z_s) + (Th_1)^{-1/2}w_t)\frac{z_t - z_s}{h_1}\right]\right\}^2 K_{h_1,ts} \\
&\quad - \frac{h_1}{T}\sum_{t=1}^T\sum_{s=1}^T\left\{y_{t+1} - \hat{F}_t^{a\top}[\theta_0^0(z_s) + h_1\theta_1^0(z_s)\frac{z_t - z_s}{h_1}]\right\}^2 K_{h_1,ts}, \\
\Pi_2 &= h_1\sum_{l=0}^R P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|)\left(\|\theta_{0,l}^0 + (Th_1)^{-1/2}\boldsymbol{v}_l\| - \|\theta_{0,l}^0\|\right), \text{ and} \\
\Pi_3 &= h_1\sum_{l=0}^R P'_{\lambda_T}(\tilde{D}_l)\left(\|h_1\theta_{1,l}^0 + (Th_1)^{-1/2}\boldsymbol{w}_l\| - \|h_1\theta_{1,l}^0\|\right).
\end{aligned}$$

Here, $\theta_r^0(z) = (\theta_{r,0}^0(z), \theta_{r,1}^0(z), \ldots, \theta_{r,R}^0(z))^\top \in \mathbb{R}^{R+1}$ for $r = 0, 1$, $\boldsymbol{v}_l = (v_{l1}, \ldots, v_{lT})^\top$ and $\boldsymbol{w}_l = (w_{l1}, \ldots, w_{lT})^\top$. By simple algebraic calculation, we have

$$\Pi_1 = -2T^{-1}\sum_{s=1}^T(v_s^\top, w_s^\top)\hat{e}(z_s) + h_1(Th_1)^{-1}\sum_{s=1}^T(v_s^\top, w_s^\top)\hat{\Sigma}^*(z_s)(v_s^\top, w_s^\top)^\top, \quad \text{(B.4)}$$

where $\hat{\Sigma}^*(z_s) = T^{-1}\sum_{t=1}\hat{F}_t^{a*}\hat{F}_t^{a*\top}K_{h_1,ts}$ with $\hat{F}_t^{a*} = (\hat{F}_t^{a\top}, ((z_t - z)/h_1)\hat{F}_t^{a\top})^\top$. Similar to the derivation as in the proof of Theorem 4.1 and by Corollary 4.2, we can show that $\hat{\Sigma}^*(z_s) = T^{-1}\sum_{t=1}\hat{F}_t^{a*}\hat{F}_t^{a*\top}K_{h_1,ts} = \Sigma_F^*(z_s) + O_p((N/\ln T)^{-1/2} + h_0^2) + o_p((Th_0/\ln T)^{-1/2}T^{1/8})$ uniformly in s, where

$$\Sigma_F^*(z_s) = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & \mu_{21} \end{pmatrix} \otimes \Sigma_F^* & \text{if } z_s \in (0,1), \\ \begin{pmatrix} \mu_{01,c} & \mu_{11,c} \\ \mu_{11,c} & \mu_{21,c} \end{pmatrix} \otimes \Sigma_F^* & \text{if } z_s = ch_1, \end{cases}$$

for some constant $c \in (0,1)$. A similar result holds for the right boundary point $z_s = 1 - ch_1$. Here, μ_{ij} is defined in Assumption 4.3(i), $\mu_{ij,c} = \int_{-c}^{1} v^i K^j(v)dv$ and $\Sigma_F^* = E(F_t^a F_t^{a\top})$ is a $(R+1)\times(R+1)$ positive definite matrix by Assumption 4.5(ii), which implies that $\Sigma_F^*(z_s)$ is a $2\,(R+1)\times 2\,(R+1)$ positive definite matrix uniformly in s. Then, by Assumption 4.1(ii) and 4.5(ii),

$$\begin{aligned} \mu_{\min}(\hat{\Sigma}^*(z_s)) &= \min_{\|x\|=1} \left\{ x^\top \Sigma_F^*(z_s) x + x^\top \left(\hat{\Sigma}^*(z_s) - \Sigma_F^*(z_s) \right) x \right\} \geq \mu_{\min}(\Sigma_F^*(z_s)) - \|\hat{\Sigma}^*(z_s) - \Sigma_F^*(z_s)\| \\ &= \mu_{\min}(\Sigma_F^*(z_s)) - (O_p((N/\ln T)^{-1/2} + h_0^2) + o_p((Th_0/\ln T)^{-1/2} T^{1/8})) \\ &\geq \mu_{\min}(\Sigma_F^*(z_s))/2 \ \text{ w.a.p.1,} \end{aligned} \tag{B.5}$$

where $\mu_{\min}(A)$ denotes the minimal eigenvalue of a positive definite matrix A. Meanwhile, by the definition of $\mathcal{B}^*(C)$, Lemma A.7 and the Cauchy-Schwartz inequality, we can prove that for some large positive constant C_1,

$$T^{-1} \sum_{s=1}^{T} (v_s^\top, w_s^\top) \hat{e}(z_s) \leq (T^{-1}\|\boldsymbol{U}\|^2)^{1/2} (T^{-1}\|\hat{e}\|^2)^{1/2} \leq CC_1 \ \text{ w.a.p.1.} \tag{B.6}$$

Combining (B.4)-(B.6) and by choosing sufficiently large C, we have w.a.p.1,

$$\begin{aligned} \Pi_1 &\geq T^{-1} \sum_{s=1}^{T} \left\| (v_s^\top, w_s^\top) \right\|^2 \min_s \mu_{\min} \left(\hat{\Sigma}^*(z_s) \right) - 2T^{-1} \sum_{s=1}^{T} (v_s^\top, w_s^\top) \hat{e}(z_s) \\ &\geq \frac{1}{2} C^2 \min_s \mu_{\min}(\Sigma_F^*(z_s)) T^{-1} \|\boldsymbol{U}\|^2 - 2CC_1 > \frac{C^2 \min_s \mu_{\min}(\Sigma_F^*(z_s))}{4} T^{-1} \|\boldsymbol{U}\|^2. \end{aligned} \tag{B.7}$$

On the other hand, by applying triangle inequality,

$$\Pi_2 + \Pi_3 \geq -T^{-1/2} h_1^{1/2} \sum_{l=0}^{R_f} P'_{\lambda_T}(\|\tilde{\theta}_{0,l}\|) \|\boldsymbol{v}_l\| - T^{-1/2} h_1^{1/2} \sum_{l=0}^{R_f} P'_{\lambda_T}(\tilde{D}_l) \|\boldsymbol{w}_l\|.$$

Similar to the proof of Theorem 4.5(b), we have that $P(\min_{1\leq l\leq R_f} \|\tilde{\theta}_{0,l}\| \geq a\lambda_T) \to 1$ and $P(\min_{1\leq l\leq R_f} \tilde{D}_l \geq a\lambda_T) \to 1$, which implies that $\Pi_2 + \Pi_3 = 0$ w.a.p.1. Therefore, (B.3) is dominated by Π_1 and the proof is completed. ■

Proof of Lemma A.9

We only show that $P(\|\hat{\theta}_{0,l}\| = 0) \to 1$ for any $R_f < l \le R$, as the proof of that $P(\|h_1 \hat{\theta}_{1,l}\| = 0) \to 1$ is analogous. It suffices to prove that $P(\|\hat{\theta}_{0,l}\| = 0) \to 1$ with $l = R$. The proofs for the case of $R_f < l < R$ are similar. If the claim is not true (i.e., $\|\hat{\theta}_{0,R}\| \neq 0$), then $(\hat{\Theta}_0, \hat{\Theta}_1)$ must be the solution of the following normal equation

$$0 = T \cdot \left. \frac{\partial Q_\lambda(\Theta_0, \Theta_1)}{\partial \theta_{0,R}} \right|_{(\Theta_0,\Theta_1)=(\hat{\Theta}_0,\hat{\Theta}_1)} = J_1 + J_2 + J_3, \tag{B.8}$$

where (Θ_0^0, Θ_1^0) is the true value of (Θ_0, Θ_1), $J_2 = T \cdot \frac{P'_{\lambda_T}(\|\tilde{\theta}_{0,R}\|)\theta_{0,R}}{\|\theta_{0,R}\|}$ and $J_3 = T \cdot \frac{P'_{\lambda_T}(\tilde{D}_l) h_1 \theta_{1,R}}{\|h_1 \theta_{1,R}\|}$, J_1 is a T-dimensional vector with its sth component given by $J_{1s} = -2\sum_{t=1}^T \hat{F}_{Rt} \Big\{ y_{t+1} - \hat{F}_t^{a\top}[\hat{\beta}_\lambda(z_s) + \hat{\theta}_1(z_s)(z_t - z_s)] \Big\} K_{h_1}(z_t - z_s)$. Notice that

$$\begin{aligned}
J_{1s} \;=\; & -2\sum_{t=1}^T F_{Rt} \Big\{ y_{t+1} - F_t^{a\top}[\hat{\beta}_\lambda(z_s) + \hat{\theta}(z_s)(z_t - z_s)] \Big\} K_{h_1}(z_t - z_s) \\
& +2\sum_{t=1}^T F_{Rt} \Big\{ (\hat{F}_t^a - F_t^a)^\top[\hat{\beta}_\lambda(z_s) + \hat{\theta}(z_s)(z_t - z_s)] \Big\} K_{h_1}(z_t - z_s) \\
& -2\sum_{t=1}^T (\hat{F}_{Rt} - F_{Rt}) \Big\{ y_{t+1} - F_t^{a\top}[\hat{\beta}_\lambda(z_s) + \hat{\theta}(z_s)(z_t - z_s)] \Big\} K_{h_1}(z_t - z_s) \\
& +2\sum_{t=1}^T (\hat{F}_{Rt} - F_{Rt}) \Big\{ (\hat{F}_t^a - F_t^a)^\top[\hat{\beta}_\lambda(z_s) + \hat{\theta}(z_s)(z_t - z_s)] \Big\} K_{h_1}(z_t - z_s) \\
\equiv\; & J_{11s} + J_{12s} + J_{13s} + J_{14s}.
\end{aligned}$$

By Corollary 4.2 and similar to the proof of Lemma A.7, $J_{1\ell s} = o_p(J_{11s})$ for $\ell = 2, 3, 4$. Therefore, we only need to bound J_{11s}. Indeed,

$$\begin{aligned}
J_{11s} =& -\sum_{t=1}^T F_{Rt} e_{y,t+1} K_{h_1}(z_t - z_s) - \sum_{t=1}^T F_{Rt} F_t^{a\top} \{\beta(z_t) - \theta_0^0(z_s) - \theta_1^0(z_s)(z_t - z_s)\} K_{h_1}(z_t - z_s) \\
& - \sum_{t=1}^T F_{Rt} F_t^{a\top} \{\theta_0^0(z_s) - \hat{\beta}_\lambda(z_s)\} K_{h_1}(z_t - z_s) \\
& - \sum_{t=1}^T F_{Rt} F_t^{a\top} \{[\theta_1^0(z_s) - \hat{\theta}_1(z_s)](z_t - z_s)\} K_{h_1}(z_t - z_s) \\
\equiv& J_{11s,1} + J_{11s,2} + J_{11s,3} + J_{11s,4}.
\end{aligned}$$

From the proof of Lemma A.7, we have $\{\sum_{s=1}^T J_{11s,1}^2\}^{1/2} = O_p(Th_1^{-1/2})$ and $\{\sum_{s=1}^T J_{11s,2}^2\}^{1/2} = O_p(Th_1^{-1/2})$. Following from Lemma A.5 and the proof of Theorem 4.1, we have $\frac{1}{T}\sum_{t=1}^T F_{Rt} F_t^\top \times K_{h_1}(z_t - z_s) = O_p(1)$, uniformly for all s. Finally, by Lemma A.8 and the boundedness of

$(z_t - z_s)/h_1$, we have

$$\left(\sum_{s=1}^{T} J_{11s,3}^2\right)^{1/2} \leq \left(\sum_{s=1}^{T} \left\|\theta_0^0(z_s) - \hat{\beta}_\lambda(z_s)\right\|^2 \max_s \left\|\sum_{t=1}^{T} F_{Rt} F_t^\top K_{h_1}(z_t - z_s)\right\|^2\right)^{1/2}$$
$$= \left(TO_p((Th_1)^{-1}) O_p(T^2)\right)^{1/2} = O_p(Th_1^{-1/2}).$$

and

$$\left(\sum_{s=1}^{T} J_{11s,4}^2\right)^{1/2} \leq \left(\sum_{s=1}^{T} \left\|[\theta_1^0(z_s) - \hat{\theta}_1(z_s)](z_t - z_s)\right\|^2 \max_s \left\|\sum_{t=1}^{T} F_{Rt} F_t^\top K_{h_1}(z_t - z_s)\right\|^2\right)^{1/2}$$
$$= \left(TO_p((Th_1)^{-1}) O_p(T^2)\right)^{1/2} = O_p(Th_1^{-1/2}).$$

Then,

$$\|J_1\| = \sqrt{\sum_{s=1}^{T} J_{1s}^2} = O_p(Th_1^{-1/2}).$$

On the other hand, $\|J_2\| = TP'_{\lambda_T}(\|\tilde{\theta}_{0,R}\|) = \frac{P'_{\lambda_T}(\|\tilde{\theta}_{0,R}\|)}{\lambda_T} h_1^{1/2} \lambda_T T h_1^{-1/2}$ and $\|J_3\| = \frac{P'_{\lambda_T}(\tilde{D}_l)}{\lambda_T} h_1^{1/2} \lambda_T$ $\cdot Th_1^{-1/2}$. Since $\frac{P'_{\lambda_T}(\|\tilde{\theta}_{0,R}\|)}{\lambda_T} > 0$, $\frac{P'_{\lambda_T}(\tilde{D}_l)}{\lambda_T} > 0$ and $h_1^{1/2}\lambda_T \to 0$ by Assumption 4.5(iv), we have $P(\|J_2\| + \|J_3\| < \|J_1\|) \to 1$, which is contradict with (B.8). This implies that $\hat{\theta}_{0,l}$ must be located at the place where the objective function $Q_\lambda(\Theta_0, \Theta_1)$ is not differentiable. Since the only place where $Q_\lambda(\Theta_0, \Theta_1)$ is not differentiable for $\theta_{0,R}$ is the origin, we know immediately that $P(\|\hat{\theta}_{0,l}\| = 0) \to 1$ when $l = R$. This completes the proof. ■

References

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of econometrics*, 176(1):18–29.

Bernstein, D. S. (2005). *Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)*. Princeton University Press. Google-Books-ID: x7isojLkDTcC.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Hall, P. and Heyde, C. C. (2014). *Martingale limit theory and its application*. Academic press.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Hunter, D. R. and Li, R. (2005). Variable selection using mm algorithms. *Annals of statistics*, 33(4):1617.

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

Su, L. and Wang, X. (2020). Corrigendum to "on time-varying factor models: Estimation and testing" [j. econometrics 198 (2017) 84-101].

Wang, H. and Xia, Y. (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association*, 104(486):747–757.