



MONASH  
BUSINESS  
SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

**Recursive Estimation in Large Panel Data Models:  
Theory and Practice**

Bin Jiang, Yanrong Yang, Jiti Gao and Cheng Hsiao

February 2017

Working Paper 05/17

# Recursive Estimation in Large Panel Data Models: Theory and Practice

Bin Jiang<sup>1</sup>, Yanrong Yang<sup>2</sup>, Jiti Gao<sup>1\*</sup> and Cheng Hsiao<sup>3†</sup>

<sup>1</sup>Monash University, Australia

<sup>2</sup>The Australian National University, Australia; and <sup>3</sup>University of Southern California, USA

February 22, 2017

## Abstract

Bai (2009) proposes a recursive least-squares estimation method for large panel data models with unobservable interactive fixed effects, but the impact of recursion on the asymptotic properties of the least-squares estimators is not taken into account. In this paper, we extend Bai (2009) by investigating the recursive estimator asymptotically. In general, the asymptotic properties we establish for the recursive estimators largely complement the theory and practice of the recursive least-squares procedure suggested by Bai (2009). In particular, we show that consistency of the recursive estimator depends on three key points, consistency of the initial OLS estimator, the number of recursive steps and the endogeneity arising due to the dependence between regressors and interactive effects. Compared to the theoretical estimator in Bai (2009), such endogeneity affects the convergence rate of recursive least-squares estimators. Finite sample properties of the proposed estimators are investigated using a simulation study.

## 1 Introduction

Recent econometric literature has shown a great deal of interest on panel data regression models with factor structures, especially when both the cross section dimension ( $N$ ) and the length of time periods ( $T$ ) are large. The prevalent use of these models is motivated by the flexibility of factor structures in capturing the unobservable heterogeneity. In fact, a factor structure is often interpreted as the "interactive fixed effects" (e.g., Bai, 2009; Su et al., 2012; Moon and Weidner, 2015). The multiplicative structure allows the time-varying common shocks to have different impact on different individuals. It also covers the traditional additive random and fixed effects models as special cases.

---

\*The first three authors would like to thank the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421.

†*Corresponding author*: Email address: chowderlad@gmail.com

There are no simple transformations to eliminate the multiplicative effects. Bai (2009) advocates to treat both the individual-specific and time-specific effects as unknown constants and proposes the least-squares based recursive approach that involves the use of principal components analysis (PCA) for estimating the factor structure. This approach has the advantage of accommodating the estimation of the regression coefficients and factor components simultaneously, and attracts a large amount of attention (e.g., Su et al., 2012; Bai and Li, 2014; Moon and Weidner, 2015) in the literature. Alternatively, Pesaran (2006) proposes the common correlated effects (CCE) approach which utilizes the weighted cross section average of the dependent variables and regressors to approximate the unobserved factors, and has been widely extended by Chudik et al. (2011), Harding and Lamarche (2011), Kapetanios et al. (2011), Pesaran and Tosetti (2011), Su and Jin (2012), and Huang (2013). In this paper, we focus on the convergence issues of the recursive estimation method proposed by Bai (2009).

A main feature of the approach of Bai (2009) is that the sum of squared residuals function does not possess closed-form least-squares estimators for the slope coefficients and factor components. As a consequence, Bai establishes the asymptotic theory for these least-squares solutions without closed forms. Bai (2009) suggests implementing this approach using the recursive estimation which iteratively alternates the PCA method applied to the least-squares residuals for estimating the factors and the least-squares estimation of the augmented regression with these PCA estimates. Figure 1 shows exactly how the theory and the practice may interact with each other in Bai (2009) by noting that  $\widehat{\beta}_{NT}$  is the theoretical least squares estimator (without closed-form expression) investigated in Bai (2009),  $\widehat{\beta}_{NT}^{(m)}$  is numerically the recursive estimator at the  $m$ -th iteration and  $\beta_0$  is the parameter of interest (which are notations that will be introduced later). Precisely, Bai takes the horizontal route of "double arrow" to establish the asymptotic theory and the vertical route to implement the estimator practically.

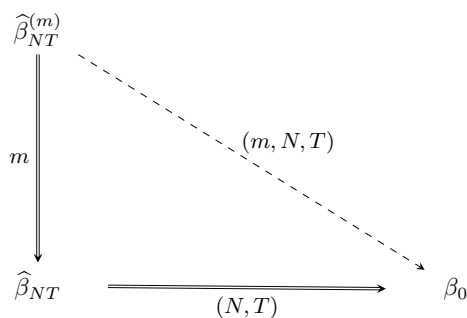


Figure 1: Multiple ways of convergence

In this paper, the first main contribution is that we provide a unified framework accommodating the theory and practice in Bai (2009) by deriving asymptotic properties of the recursive estimator  $\widehat{\beta}_{NT}^{(m)}$  for a fixed number of iterations when  $N$  and  $T$  tend to infinity jointly. Compared to the work of Bai (2009) depicted in Figure 1, we take the direct route of "dashed arrow" to  $\beta_0$  in one go. The asymptotic properties of  $\widehat{\beta}_{NT}^{(m)}$  build up a bridge connecting the numerical estimator applied only in practice straight to the statistical inference about the target parameter.

Under this framework, we are able to complement Bai (2009) in two significant directions. First, note that the validity of the numerical estimator as approximation for the theoretical one relies on

the numerical convergence (which refers to the convergence of sample iterates to the sample fixed point conditioning on observed data with respect to the increase of iterative steps) of the algorithm. If numerical convergence fails, the least-squares based approach would be very difficult to implement. However, it is unclear about what conditions cause this failure and under what circumstances the recursive estimation works in this particular problem. We provide conditions that guarantee the consistency of the recursive estimators for the regression parameter and the common factors. An important lesson we learn is that recursion itself does not secure either the consistency of the recursive estimators or their numerical convergence to the infeasible theoretical estimators. Recursive estimation is only valid when regressors are weakly cross-sectionally and serially dependent and the endogeneity due to the interdependence between the regressors and factor residuals is not strong. Second, as the practical estimator implemented in Bai (2009), exploring statistical properties of the recursive estimator is interesting in its own right. Recently, there are quite a few econometricians who have examined the inference problems for specific EM-type recursive procedures. For example, Dominitz and Sherman (2005) develop general conditions for rates of convergence and convergence in distribution of recursive procedures for estimating finite-dimensional parameters. Pastorello et al. (2003) propose the latent backfitting recursive technique which is particularly designed for estimating structural nonadaptive models involving latent variables such as financial market models like CAPM and Black-Scholes option pricing models. Song et al. (2005) present a maximization by parts (MBP) algorithm to avoid numerical intractability of solving complicated likelihood functions and Frazier (2013) generalizes the MBP algorithm to cover semiparametric models.

The second main contribution of this paper is that we establish the asymptotic orders of recursive estimators which reveal how consistency is related to the number of iterations for a particular data generating process. These asymptotics can be used as guidelines for determining checkable stopping rules based on user-specific numerical accuracy (e.g., Dominitz and Sherman, 2005).

Finally, it is worth mentioning that we employ a generalized  $\alpha$ -mixing framework for panel data (e.g., Su and Jin, 2012; Dong et al., 2015) under which cross-sectional and serial dependence in the idiosyncratic errors can be captured in a concrete and checkable manner. By doing so, we can deliver a better understanding about the high-order moment conditions given in Bai (2009, Assumption C). In addition, we weaken the independence among the regressors, factor components and idiosyncratic errors assumed in the Assumption D of Bai (2009) by replacing it with the **conditional mean independence** which covers some dependent scenarios in particular forms.

The rest of the paper is organized as follows. Section 2 provides the panel data regression model with interactive fixed effects and the recursive estimators we consider. Section 3 discusses the main assumptions. Section 4 analyzes the asymptotic properties of these recursive estimators. A number of finite sample simulation results are present in Section 5. Section 6 concludes this paper. All proofs and technical details are documented in the appendix.

## 2 Model and Recursive Estimation

We revisit the large panel data regression model with interactive fixed effects in Bai (2009)

$$Y_{it} = \mathbf{X}'_{it}\boldsymbol{\beta}_0 + \boldsymbol{\lambda}'_i \mathbf{F}^0_t + \epsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (2.1)$$

where  $\mathbf{X}_{it}$  is a  $p \times 1$  vector of regressors,  $\boldsymbol{\beta}_0$  is a  $p \times 1$  vector of slope coefficients,  $\boldsymbol{\lambda}_i^0$  is a  $r \times 1$  vector of factor loadings and  $\mathbf{F}_t^0$  is a  $r \times 1$  vector of the unobserved common factors;  $\epsilon_{it}$  is the idiosyncratic error.

Alternatively, one can rewrite this model in matrix notation as below

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{F}^0 \boldsymbol{\Lambda}^0 + \boldsymbol{\epsilon}, \quad (2.2)$$

where  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N)$  is a  $T \times N$  matrix with  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})'$  for any  $i = 1, 2, \dots, N$ ;  $\boldsymbol{\Lambda}^0 = (\boldsymbol{\lambda}_1^0, \boldsymbol{\lambda}_2^0, \dots, \boldsymbol{\lambda}_N^0)'$ ;  $\mathbf{F}^0 = (\mathbf{F}_1^0, \mathbf{F}_2^0, \dots, \mathbf{F}_T^0)'$  and  $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots, \boldsymbol{\epsilon}_T)'$  with  $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{Nt})'$  for any  $t = 1, 2, \dots, T$ .

We find it also convenient to work with the following vector form of model (2.1) in some occasions

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_0 + \mathbf{F}^0 \boldsymbol{\lambda}_i^0 + \epsilon_i, \quad i = 1, 2, \dots, N. \quad (2.3)$$

For the identification of the factor structure, a common practice is to impose the normalization restriction

$$\frac{\mathbf{F}^{0'} \mathbf{F}^0}{T} = \mathbf{I}_r, \quad \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \text{ is diagonal}, \quad (2.4)$$

which ensures the identification of factors  $\mathbf{F}^0$  and the factor loadings  $\boldsymbol{\Lambda}^0$  up to an orthogonal matrix.

Bai (2009) proves that minimizing the sum of squared residuals objective function

$$SSR = \frac{1}{NT} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{F} \boldsymbol{\lambda}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{F} \boldsymbol{\lambda}_i) \quad (2.5)$$

with respect to  $(\boldsymbol{\beta}, \mathbf{F}, \boldsymbol{\Lambda}) \in (\mathbb{R}^p, \mathbb{R}^T \times \mathbb{R}^r, \mathbb{R}^N \times \mathbb{R}^r)$  subject to (2.4), yields consistent and asymptotically normally distributed estimator of  $\boldsymbol{\beta}_0$  when  $(N, T) \rightarrow \infty$ . However, simultaneously solving the first-order conditions to derive the least-squares solution  $(\widehat{\boldsymbol{\beta}}_{NT}, \widehat{\mathbf{F}}, \widehat{\boldsymbol{\Lambda}})$  to the minimization problem (2.5) is computationally tedious. Hence, Bai (2009) suggests the following recursive procedure to obtain the least-squares estimator of  $\boldsymbol{\beta}_0$ . It should be noticed that we use  $(\widehat{\boldsymbol{\beta}}_{NT}^{(m)}, \widehat{\mathbf{F}}^{(m)}, \widehat{\boldsymbol{\Lambda}}^{(m)})$  to stand for the recursive estimator at step  $m$  in what follows.

1. Initialize the recursive procedure with the ordinary least-squares estimator by ignoring the factor components

$$\widehat{\boldsymbol{\beta}}_{NT}^{(0)} = \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{Y}_i.$$

2. Extract  $(\widehat{\mathbf{F}}^{(m)}, \widehat{\boldsymbol{\Lambda}}^{(m)})$  from the least-squares residuals using principal components approach. In particular, given the  $m$ -th step  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$ , estimate the  $m$ -th step  $\widehat{\mathbf{F}}^{(m)}$  as the eigenvectors (multiplied by  $\sqrt{T}$  in order to admit the normalization restriction (2.4)) corresponding to the  $r$  largest eigenvalues of the matrix  $\mathbf{S}^{(m)} = \frac{1}{NT} \sum_{i=1}^N \left( \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{NT}^{(m)} \right) \left( \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{NT}^{(m)} \right)'$ , that is,

$$\widehat{\mathbf{F}}^{(m)} \mathbf{V}_{NT}^{(m)} = \mathbf{S}^{(m)} \widehat{\mathbf{F}}^{(m)} \quad (2.6)$$

in which  $\mathbf{V}_{NT}^{(m)}$  is a  $r \times r$  diagonal matrix with diagonal elements being the  $r$  largest eigenvalues of  $\mathbf{S}^{(m)}$ . For each  $i = 1, 2, \dots, N$ , obtain  $\widehat{\boldsymbol{\lambda}}_i^{(m)}$  according to

$$\widehat{\boldsymbol{\lambda}}_i^{(m)} = \frac{1}{T} \widehat{\mathbf{F}}^{(m)'} \left( \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{NT}^{(m)} \right). \quad (2.7)$$

3. Derive  $\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)}$  by either

$$\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} = \left( \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \left( \mathbf{Y}_i - \widehat{\mathbf{F}}^{(m)} \widehat{\boldsymbol{\lambda}}_i^{(m)} \right), \quad (2.8)$$

or

$$\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} = \left( \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{Y}_i \quad (2.9)$$

where  $\mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} = \mathbf{I}_T - \widehat{\mathbf{F}}^{(m)} \left( \widehat{\mathbf{F}}^{(m)'} \widehat{\mathbf{F}}^{(m)} \right)^{-1} \widehat{\mathbf{F}}^{(m)'} = \mathbf{I}_T - \widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{F}}^{(m)'} / T$ .

4. Repeat steps 2 and 3 until numerical convergence.

In this paper, we extend Bai (2009) by examining the asymptotic properties of the recursive estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$ . In fact, this estimator is exactly what Bai implements in practice, so the statistical theory for  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  is important in its own right and is also more relevant compared to  $\widehat{\boldsymbol{\beta}}_{NT}$  when applying the recursive procedure in empirical studies. Intuitively, we can understand the relationship between the theoretical estimator  $\widehat{\boldsymbol{\beta}}_{NT}$  Bai (2009) considers and the recursive estimator noting that the estimation procedure above essentially defines a mapping  $\boldsymbol{\theta} \mapsto M_{NT}(\boldsymbol{\theta})$  from parameter space  $\Theta$  to itself

$$M_{NT}(\boldsymbol{\theta}) = \operatorname{argmin}_{\boldsymbol{\theta}} \left\{ \frac{1}{NT} \sum_{i=1}^N \left( \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta} - \widehat{\mathbf{F}}(\boldsymbol{\theta}) \widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\theta}) \right)' \left( \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta} - \widehat{\mathbf{F}}(\boldsymbol{\theta}) \widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\theta}) \right) \right\}$$

where  $\widehat{\mathbf{F}}(\boldsymbol{\theta})$  and  $\widehat{\boldsymbol{\lambda}}_i^{(m)}(\boldsymbol{\theta})$  are derived from (2.6) and (2.7) with  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  replaced by  $\boldsymbol{\theta}$  for any  $N$  and  $T$ . Moreover, this procedure implies

$$\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} = M_{NT} \left( \widehat{\boldsymbol{\beta}}_{NT}^{(m)} \right)$$

so that  $\widehat{\boldsymbol{\beta}}_{NT}$  is the fixed point of the sample mapping  $M_{NT}(\cdot)$  as the number of iterations  $m$  goes to infinity for any fixed  $(N, T)$  under some regularity conditions.

### 3 Assumptions

In this section, we provide assumptions required for deriving asymptotic properties of the recursive estimation along with necessary explanations.

**Assumption 1** (Regularity Conditions).

Let  $\mathcal{F} = \left\{ \mathbf{F} : \frac{\mathbf{F}'\mathbf{F}}{T} = \mathbf{I}_r \right\}$ . We assume

$$\inf_{\mathbf{F} \in \mathcal{F}} \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\mathbf{F}} \mathbf{X}_i > 0. \quad (3.1)$$

Assumption 1 guarantees the denominator on the right hand side of (2.9) to be non-singular so that the recursive estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  is always well defined for each positive integer  $m$ .

**Assumption 2** (Moment Conditions).

1.  $\mathbb{E} \|\mathbf{F}_t\|^4 < \infty$  and  $\mathbb{E} \|\boldsymbol{\lambda}_i\|^4 < \infty$  for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ . Moreover,  $\frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{N} \xrightarrow{i.p.} \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}} > 0$  for some  $r \times r$  positive definite matrix  $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}$ , as  $N \rightarrow \infty$ . In addition,  $\mathbb{E}(\boldsymbol{\lambda}_i^0 \mathbf{F}_t^0) = 0$  for any  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ .
2.  $\mathbb{E} \|\mathbf{X}_{it}\|^4 < \infty$  for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ . Moreover, as  $(N, T)$  tend to infinity,  $\frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \xrightarrow{i.p.} \boldsymbol{\Sigma}_{\mathbf{x}}$  which is a  $p \times p$  positive matrix.
3.  $\mathbb{E} |\epsilon_{it}|^4 < \infty$  for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$  and  $\mathbb{E}(\epsilon_{it} | \mathbf{X}_{js}, \boldsymbol{\lambda}_j, \mathbf{F}_s) = 0$  for all  $i, j = 1, 2, \dots, N$  and  $t, s = 1, 2, \dots, T$ .

Assumption 2.1 and 2.2 are essentially the same as the Assumption B of Bai (2009) except that we assume the factor components to have zero mean. Compared to the Assumption D of Bai (2009), we impose a weaker restriction on the correlation among the idiosyncratic error components, common factor components and regressors. Rather than assuming independence strictly, we only require that the first moment of the idiosyncratic errors conditional on the common components and regressors is equal to zero. For example, we do not rule out cases  $\epsilon_{it} = g(\mathbf{X}_{js}, \boldsymbol{\lambda}_j, \mathbf{F}_s) \cdot \zeta_{it}$  for any  $j$  and  $s$  where  $g(\cdot)$  is an arbitrary deterministic function and  $\zeta_{it}$  is i.i.d. with zero mean across  $i$  and  $t$ .

**Assumption 3** (Mixing Conditions).

1. For each  $i = 1, 2, \dots, N$ , the random sequence  $\{\mathbf{v}_{it} := (\mathbf{X}_{it}', \mathbf{F}_t', \epsilon_{it}) : t \geq 1\}$  is a stationary and  $\alpha$ -mixing process with mixing coefficient  $\alpha_i(\tau)$  and  $\tau > 0$ . Furthermore, there exists a positive coefficient function  $\alpha(\tau)$  such that

$$\sup_i \alpha_i(\tau) \leq \alpha(\tau) \quad \text{and} \quad \sum_{\tau=1}^T \alpha^{4+\eta}(\tau) = O(1)$$

for some  $\eta > 0$ .

2. For cross sections  $i$  and  $j$  where  $i \neq j$ , the random sequence  $\{(\mathbf{v}_{it}, \mathbf{v}_{jt}) : t \geq 1\}$  is also an  $\alpha$ -mixing process with mixing coefficient  $\alpha_{ij}(\tau)$  such that

$$\begin{aligned} (a) \quad & \sum_{i,j=1}^N [\alpha_{ij}(0)]^{\frac{\eta}{4+\eta}} = O(N), \\ (b) \quad & \sum_{i,j=1}^N \sum_{\tau=1}^T \tau \cdot [\alpha_{ij}(\tau)]^{\frac{\eta}{4+\eta}} = O(NT), \\ (c) \quad & \sum_{i,j=1}^N \sum_{\tau=1}^T \tau^3 \cdot [\alpha_{ij}(\tau)]^{\frac{\eta}{4+\eta}} = O(NT^2). \end{aligned}$$

3. For different cross sections  $i, j, k, m$ , the random sequence  $\{(\mathbf{v}_{it}, \mathbf{v}_{jt}, \mathbf{v}_{kt}, \mathbf{v}_{mt}) : t \geq 1\}$  is an  $\alpha$ -mixing process with mixing coefficient  $\alpha_{ijkm}(\tau)$  such that

$$\sum_{i,j,k,m=1}^N \sum_{\tau=1}^T \tau [\alpha_{ijkm}(\tau)]^{\frac{\eta}{4+\eta}} = O(N^2T).$$

Assumption 3 extends the notion of “ $\alpha$ -mixing” in the time series context (e.g., Fan and Yao, 2003; Gao, 2007) to capture both the cross section and temporal dependence exhibiting in large panels. Such an  $\alpha$ -mixing framework (e.g., Chen et al., 2012; Su and Jin, 2012; Dong et al., 2015) can depict complicated cross-sectional and serial correlations present in panel data in a concise manner and provide an alternative to understanding the high-order moment conditions employed by Bai (2009) in his Assumption C. In Assumption 3.2 and 3.3, the  $\alpha$ -mixing process involving different cross sections can model not only serial dependence but also cross-sectional dependence in the idiosyncratic error components.

The weak dependence in both serial and cross section dimensions constructed in Assumption 3 satisfies that characterized by moment inequalities in Assumption C of Bai (2009). One main advantage of the  $\alpha$ -mixing assumption for panel data is the conciseness of utilizing the  $\alpha$ -mixing coefficients to control the levels of serial and cross section dependence simultaneously. For example, we show below how we can use  $\alpha$ -mixing condition to derive moment conditions.

1. Assumption C(ii) in Bai (2009): Based on Assumption 3.2, we utilize Lemma 2 to obtain

$$\left| \frac{1}{NT} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E}(\epsilon_{it}\epsilon_{js}) \right| \leq \frac{1}{NT} \sum_{i,j=1}^N \sum_{t,s=1}^T \alpha_{ij}^{\frac{\delta}{2+\delta}}(|t-s|) \cdot [\mathbb{E}|\epsilon_{it}|^{2+\delta}]^{\frac{1}{2+\delta}} \cdot [\mathbb{E}|\epsilon_{js}|^{2+\delta}]^{\frac{1}{2+\delta}} \leq M.$$

2. Assumption C(iii)(2) in Bai (2009): Based on Assumption 3.3, we can also utilize Lemma 2 to derive

$$\begin{aligned} \left| \frac{1}{TN^2} \sum_{t,s=1}^T \sum_{i,j,k,\ell=1}^N \text{cov}(\epsilon_{it}\epsilon_{jt}, \epsilon_{ks}\epsilon_{\ell s}) \right| & \leq \frac{1}{TN^2} \sum_{t,s=1}^T \sum_{i,j,k,\ell=1}^N \alpha_{ijk\ell}^{\frac{\delta}{2+\delta}}(|t-s|) \\ & \cdot [\mathbb{E}|\epsilon_{it}\epsilon_{jt}|^{2+\delta}]^{\frac{1}{2+\delta}} \cdot [\mathbb{E}|\epsilon_{ks}\epsilon_{\ell s}|^{2+\delta}]^{\frac{1}{2+\delta}} \leq M \end{aligned}$$



## 4 Asymptotic Theory

In this section, we will provide the asymptotic properties of the recursive estimators  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  and  $\widehat{\mathbf{F}}^{(m)}$  respectively. The results we obtain convey two important messages. First, the recursive estimator will be consistent if the initial least squares estimator is consistent. Moreover, recursion will accelerate the convergence rate of the recursive estimator to the true parameter. Second, the recursive estimation procedure will not necessarily yield consistent estimation when the initial estimator is inconsistent. Before proceeding to the main results, we stress that the notation “ $a \asymp b$ ” is employed to denote that  $a = O_P(b)$  and  $b = O_P(a)$  (e.g., Lam et al., 2011).

### 4.1 Asymptotic Properties of $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$

In this subsection, we focus on the recursive estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  through investigating three main issues. The first one is the asymptotic order of the initial OLS estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(0)}$  when  $(N, T) \rightarrow \infty$ . In practice, implementing a recursive algorithm with a good starting point would often achieve fast numerical convergence. The second one is how the current recursive estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)}$  depends on the estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  at the previous step for  $m \geq 0$ . This is also the key aspect of understanding the impact of the recursion on the asymptotic behaviour of  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$ . The final issue is how the order of the estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  is affected by the initial estimator and the data structure.

The asymptotic property of the initial OLS estimator is formulated as below.

**Theorem 1.** *Under Assumptions 2 and 3, we have*

$$\left\| \widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 \right\| = O_P \left( \max \left( \frac{1}{\sqrt{NT}}, \gamma_{NT}^{1/2} \right) \right), \quad (4.1)$$

where

$$\gamma_{NT} = \frac{1}{N^2 T^2} \sum_{i,j=1}^N \mathbb{E} \left( \boldsymbol{\lambda}_i \mathbf{F}' \mathbf{X}_i \mathbf{X}_j' \mathbf{F} \boldsymbol{\lambda}_j \right). \quad (4.2)$$

Theorem 1 implies that the asymptotic order of  $\left\| \widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 \right\|$  depends on the quantity  $\gamma_{NT}$ , which captures the interdependence between the regressors  $\{\mathbf{X}_i : i = 1, 2, \dots, N\}$  and the common factor components  $\mathbf{F}\boldsymbol{\Lambda}'$ . In particular, when the common components are independent of the regressors  $\mathbf{X}$ , the asymptotic order of  $\gamma_{NT}$  is  $\frac{1}{NT}$  so that  $\widehat{\boldsymbol{\beta}}_{NT}^{(0)}$  is  $\sqrt{NT}$ -consistent as expected. However, the initial least squares estimator is not necessarily consistent, since the regressors can ‘heavily’ depend on or driven by the common factors (e.g., Pesaran, 2006). In this case, ignoring the impact of the factor components will result in inconsistent estimation for the regression parameter  $\boldsymbol{\beta}_0$ . For instance, suppose that the regressors follow a factor model such that  $\mathbf{X}_i = \mathbf{F}\mathbf{q}_i + \boldsymbol{\eta}_i$  where the idiosyncratic error  $\boldsymbol{\eta}_i$  is independent of  $(\mathbf{F}, \mathbf{q}_i)$ , then we have that  $\gamma_{NT} \asymp 1$  when  $(N, T)$  tend to infinity. The example below will illustrate that the initial least squares estimator can be also consistent but converge to the true parameter at a rate slower than  $\sqrt{NT}$ .

Assume that the regressors follow the factor structure below

$$X_{it} = d_{NT}\lambda_i f_t + \zeta_{it}, \quad (4.3)$$

where  $\{\zeta_{it} : i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$  are i.i.d and independent of the common factors  $\mathbf{F}$  and factor loadings  $\mathbf{\Lambda}$ .  $d_{NT}$  is a deterministic constant which depends on the values of  $N, T$  and  $d_{NT} \rightarrow 0$  as  $(N, T) \rightarrow \infty$ .

Under the assumptions of Theorem 1, one can easily verify that  $\gamma_{NT} = O(d_{NT}^2)$  so that the convergence rate of  $\widehat{\boldsymbol{\beta}}_{NT}^{(0)}$  is  $O_P\left(\max\left(\frac{1}{\sqrt{NT}}, d_{NT}\right)\right)$ .

The following theorem shows how  $\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)}$  exactly depends on  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  asymptotically for any fixed  $m \geq 0$ .

**Theorem 2.** *Under Assumptions 1–3, for any recursive step  $m \geq 0$ ,*

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} - \boldsymbol{\beta}_0 &\asymp \frac{1}{\sqrt{NT}} + \sum_{k=1}^p \widehat{\mathbf{w}}_{1,k}^{(m)} \left( \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right) \\ &\quad + \sum_{k=1}^p \sum_{\ell=1}^p \widehat{\mathbf{w}}_{2,k\ell}^{(m)} \left( \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right) \left( \beta_{0,\ell} - \widehat{\beta}_{NT,\ell}^{(m)} \right), \end{aligned}$$

where  $\widehat{\mathbf{w}}_{1,k}^{(m)} = O_P(1)$  and  $\widehat{\mathbf{w}}_{2,k\ell}^{(m)} = O_P(1)$  with

$$\begin{aligned} \widehat{\mathbf{w}}_{1,k}^{(m)} &= \frac{1}{NT} \sum_{j=1}^N \mathbf{X}'_j \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \frac{1}{NT} \sum_{i=1}^N \left( \mathbf{x}_{i(k)} \boldsymbol{\lambda}'_i \mathbf{F}^{0'} + \mathbf{F}^0 \boldsymbol{\lambda}'_i \mathbf{x}'_{i(k)} \right) \widehat{\mathbf{F}}^{(m)} \left( \frac{\mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \boldsymbol{\lambda}_j^0, \\ \widehat{\mathbf{w}}_{2,k\ell}^{(m)} &= \frac{1}{NT} \sum_{j=1}^N \mathbf{X}'_j \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \mathbf{x}'_{i(\ell)} \widehat{\mathbf{F}}^{(m)} \left( \frac{\mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \boldsymbol{\lambda}_j^0. \end{aligned}$$

As can be seen from this theorem, in the limit, the deviation of the current estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)}$  from the true parameter  $\boldsymbol{\beta}_0$  is determined by the linear and quadratic forms of the deviation of the estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  at the previous step. It is worth noting that the quadratic term arises due to the fact that deriving  $\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)}$  recursively requires us to handle simultaneously the estimation of the nonlinear factor structure using principal components analysis. Moreover, suppose that  $\widehat{\mathbf{F}}^{(m)}$  converges to  $\mathbf{R}$  in the sense that  $\frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{F}}^{(m)} - \mathbf{R} \right\| = o_P(1)$  with  $\mathbf{R}$  being a  $T \times r$  matrix, then the two coefficients  $\widehat{w}_{1,k}^{(m)}$  and  $\widehat{w}_{2,k\ell}^{(m)}$  will converge to  $w_{1,k}$  and  $w_{2,k\ell}$  respectively with  $\widehat{\mathbf{F}}^{(m)}$  replaced by  $\mathbf{R}$ . We stress that the common factors  $\mathbf{F}$  are only identified up to a rotation so that the recursive estimator  $\widehat{\mathbf{F}}^{(m)}$  does not necessarily converge to  $\mathbf{F}$  and  $\mathbf{R}$  is thus used for its limit instead.

Theorem 2 immediately implies the following result.

**Corollary 1.** *Under Assumptions 1–3, for any recursive step  $m \geq 0$ ,*

$$\begin{aligned} \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} - \boldsymbol{\beta}_0 \right\| &= O_P\left(\frac{1}{\sqrt{NT}}\right) + O_P\left(\left(\max\left(\gamma_{NT}^{1/2}, \frac{1}{\sqrt{NT}}\right)\right) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|\right) \\ &\quad + O_P\left(\eta_{NT}^{1/2} \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2\right), \end{aligned}$$

where  $\gamma_{NT}$  has been defined in (4.2) and

$$\eta_{NT} = \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} \left( \mathbf{X}'_{it} \mathbf{X}_{jt} \mathbf{X}'_{is} \mathbf{X}_{js} \right). \quad (4.4)$$

Notice that  $\gamma_{NT}$  determines the convergence rate of the initial estimator  $\widehat{\beta}_{NT}^{(0)}$  due to Theorem 1. Furthermore, Corollary 1 shows that this quantity also has an influence on the recursive relationship between the numerical estimators at two successive steps. In addition, we find another important quantity  $\eta_{NT}$  which is associated with the quadratic term  $\left\| \widehat{\beta}_{NT}^{(m)} - \beta_0 \right\|^2$  and affects the convergence rate of  $\widehat{\beta}_{NT}^{(m+1)}$  to the true parameter. In fact,  $\eta_{NT}$  can be regarded as some measure of the inter-correlation of the regressors  $\mathbf{X}$  among both cross section and time series dimensions. Particularly, if we assume that the regressors are i.i.d. both cross-sectionally and serially,  $\eta_{NT}$  actually becomes the fourth moment of the regressor  $\mathbf{X}_{it}$  which is commonly known as the ‘‘kurtosis’’ after standardization. The following example will show what  $\eta_{NT}$  entails in a scenario that is a little more complex than the i.i.d. case. For simplicity, we will assume that  $p = 1$  and  $r = 1$  in this example.

Suppose that the regressors follow the spatial moving average model (e.g., Kapoor et al., 2007; Kelejian and Prucha, 2010)

$$X_{it} = \sum_{h=1}^N w_{ih} \xi_{ht} + \zeta_{it}, \quad (4.5)$$

where  $\{\xi_{ht} : h = 1, 2, \dots, N; t = 1, 2, \dots, T\}$  and  $\{\zeta_{it} : i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$  are both independent and identical distributed.  $\{w_{ih} : i, h = 1, 2, \dots, N\}$  are deterministic weighting parameters.

Under the assumptions of Theorem 1, we have

$$\eta_{NT} = \frac{1}{N^2} \sum_{i,j=1}^N \sum_{h_1, h_2=1}^N w_{ih_1} w_{jh_1} w_{ih_2} w_{jh_2} + O \left( \max \left( \frac{1}{N}, \frac{1}{T} \right) \right).$$

In this example, the asymptotic order of  $\eta_{NT}$  is essentially determined by the spatial weights  $w_{ih}$ 's. Notice that for Example (4.3) it is easy to derive that  $\eta_{NT} = O(d_{NT}^4)$ .

According to Corollary 1, we can easily derive the sufficient condition for the recursive estimator  $\widehat{\beta}_{NT}^{(m)}$  to achieve consistency.

**Theorem 3.** *In addition to Assumptions 1–3, we assume that as  $N, T \rightarrow \infty$ ,*

$$\left\| \widehat{\beta}_{NT}^{(0)} - \beta_0 \right\| = o_P(1). \quad (4.6)$$

Then for any step  $m \geq 0$ ,

$$\left\| \widehat{\beta}_{NT}^{(m+1)} - \beta_0 \right\| = O_P \left( \frac{1}{\sqrt{NT}} \right) + O_P \left( \max \left( \gamma_{NT}^{1/2}, \frac{1}{\sqrt{NT}} \right) \right)^m \cdot \left\| \widehat{\beta}_{NT}^{(0)} - \beta_0 \right\|.$$

It follows from Theorem 1 that the condition  $\gamma_{NT} = o_P(1)$  can guarantee (4.6) and then Theorem 3 implies that the recursive estimator  $\widehat{\beta}_{NT}^{(m)}$  is consistent. This is to say that the consistency of the

initial OLS estimator will yield the consistency of the recursive estimation. Although in this case the initial estimator has already been consistent, recursive estimators are still desirable since it will converge faster.

**Theorem 4.** *In addition to Assumptions 1–3, we assume*

$$\begin{aligned} (1) \quad & \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 \right\| = o_P(1); \\ (2) \quad & \widehat{\mathbf{H}}_{(0)}^{-1} - \mathbf{I}_r = o_P(1), \end{aligned} \quad (4.7)$$

where  $\widehat{\mathbf{H}}_{(0)}^{-1} = \widehat{\mathbf{V}}_{NT}^{(0)} \left( \frac{\mathbf{F}^0 \widehat{\mathbf{F}}^{(0)}}{T} \right)^{-1} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1}$  with  $\mathbf{V}_{NT}^{(0)}$  is a diagonal matrix that consists of the first  $r$  largest eigenvalues arranged in decreasing order, of the matrix

$$\frac{1}{NT} \sum_{i=1}^N \left( \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{NT}^{(0)} \right) \left( \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{NT}^{(0)} \right)'. \quad (4.8)$$

Then we have, if the step  $m$  satisfies

$$m \geq \frac{2 \log(\delta) - 2 \log \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 \right\| \right)}{\log \left( \sum_{k=1}^p \|\widehat{\mathbf{w}}_{1,k}\|^2 \right)} - 1, \quad (4.9)$$

then for any positive constant  $\delta > 0$ ,

$$\left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \leq \delta,$$

where

$$\widehat{\mathbf{w}}_{1,k} = \lim_{N,T \rightarrow \infty} \left[ \frac{1}{NT} \sum_{j=1}^N \mathbf{X}'_j \mathbf{M}_{\mathbf{F}^0} \frac{1}{NT} \sum_{i=1}^N \left( \mathbf{x}_{i(k)} \boldsymbol{\lambda}_i^{0'} \mathbf{F}^{0'} + \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \mathbf{x}'_{i(k)} \right) \mathbf{F}^0 \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \boldsymbol{\lambda}_j^0 \right].$$

This theorem provides an approximation for the minimum number of iterations at which the recursive estimator is close to the true value given a particular error tolerance, when the initial least squares estimator is consistent. It is worth noting that the condition (2) required in (4.7) is in fact not restrictive, since to achieve this one just needs to completely identify the factors and factor loadings (cf. Bai and Ng, 2013). In practice, there are circumstances under which the common factors are of primary interest and need to be completely identified for specific economic interpretations. For example, Geweke and Zhou (1996) uniquely determine the factors and factor loadings when analyzing the arbitrage pricing theory in a Bayesian framework. Bai and Ng (2013) point out that examining the sensitivity to risk in arbitrage pricing theory requires complete identification of the factors and factor loadings. In the factor-augmented vector autoregressive (FAVAR) framework of Bernanke et al. (2005), exact identification of the factors are required for investigating impulse responses of multiple macroeconomic indicators due to monetary policy shocks. Bai and Wang (2012) investigate the structural dynamic factor model with complete identification of the factor structure.

The following theorem shows a particular scenario in which the recursive estimator is inconsistent

when the regressors and factor components are driven by dynamics in common.

**Theorem 5.** *In addition to Assumptions 1-3, suppose that*

$$\mathbf{x}_{it} = \boldsymbol{\gamma}_i' \mathbf{G}_t, \quad \mathbf{F}_t = \mathbf{G}_t + \mathbf{Q}_t, \quad \boldsymbol{\lambda}_i = \phi \boldsymbol{\gamma}_i, \quad (4.10)$$

where  $p = 1$ ,  $\phi > 0$  is a constant number;  $\mathbf{G}_t : r \times 1$  and  $\mathbf{Q}_t : r \times 1$  are independent random vectors; and  $\boldsymbol{\gamma}_i : r \times 1$  is a random vector. Then as  $(N, T) \rightarrow \infty$ , for any step  $m \geq 1$ ,

$$\widehat{\boldsymbol{\beta}}_{NT}^{(m)} \xrightarrow{i.p.} \boldsymbol{\beta}_0 + \phi. \quad (4.11)$$

In this case, the regressors and common factors are strongly correlated. As a result, identification of the true parameter will be a problem which incurs the inconsistency of recursive estimation.

## 4.2 Asymptotic Properties of $\widehat{\mathbf{F}}^{(m)}$

In this subsection, we present the asymptotic results about the recursive estimator  $\widehat{\mathbf{F}}^{(m)}$  for the common factors.

The following theorem illustrates how the recursive estimator of regressive coefficient influences the recursive estimator of the common factors at the same recursive step.

**Theorem 6.** *Under Assumptions 1-3, we have for any recursive step  $m \geq 1$ ,*

$$\begin{aligned} \frac{1}{\sqrt{T}} \left( \widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{H}}_{(m)}^{-1} - \mathbf{F}^0 \right) &\asymp \delta_{NT}^{-1} + \sum_{k=1}^p w_{1,k}^{(m)} \left( \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right) \\ &\quad + \sum_{k=1}^p \sum_{\ell=1}^p w_{2,k\ell}^{(m)} \left( \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right) \left( \beta_{0,\ell} - \widehat{\beta}_{NT,\ell}^{(m)} \right), \end{aligned} \quad (4.12)$$

where

$$\begin{aligned} \delta_{NT} &= \min\{\sqrt{N}, \sqrt{T}\}, \quad \widehat{\mathbf{H}}_{(m)}^{-1} = \widehat{\mathbf{V}}_{NT}^{(m)} \left( \frac{\mathbf{F}^0 \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1}, \\ w_{2,k\ell}^{(m)} &= \frac{1}{NT^{3/2}} \sum_{i=1}^N \mathbf{x}_{i(k)} \mathbf{x}_{i(\ell)}' \widehat{\mathbf{F}}^{(m)} \left( \frac{\mathbf{F}^0 \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1}, \\ w_{1,k}^{(m)} &= \frac{1}{NT^{3/2}} \sum_{i=1}^N \left( \mathbf{x}_{i(k)} \boldsymbol{\lambda}_i^{0'} \mathbf{F}^{0'} + \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \mathbf{x}_{i(k)}' \right) \widehat{\mathbf{F}}^{(m)} \left( \frac{\mathbf{F}^0 \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1}. \end{aligned}$$

In Theorem 6, we have derived the formula that controls the asymptotic order of the recursive estimator  $\widehat{\mathbf{F}}^{(m)}$ . As can be observed from (4.12), when the recursive estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  is a consistent estimator of  $\boldsymbol{\beta}_0$  at step  $m$ , that is,  $\left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| = o_P(1)$ , the second term on the right side of (4.12) will be the leading term. Otherwise, the third term which captures the quadratic effects may also be the leading term in determining the asymptotic order of  $\frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{H}}_{(m)}^{-1} - \mathbf{F}^0 \right\|$ .

The following corollary is an immediate consequence of Theorem 6.

**Corollary 2.** *Under Assumptions 1–3, for any recursive step  $m \geq 1$ ,*

$$\begin{aligned} \frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{H}}_{(m)}^{-1} - \mathbf{F}^0 \right\| &= O_P(\delta_{NT}^{-1}) + O_P \left( \max \left( \gamma_{NT}^{1/2}, \delta_{NT}^{-1} \right) \right) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \\ &\quad + O_P(\eta_{NT}^{1/2}) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2. \end{aligned} \quad (4.13)$$

According to (4.13), the convergence rate of the recursive estimator  $\widehat{\mathbf{F}}_{NT}^{(m)}$  is affected by the dependence between the regressors and the common factor components characterized by  $\gamma_{NT}$ . Moreover, it is also influenced by the inner-dependence among the elements of the regressors  $\mathbf{X}$  measured by  $\eta_{NT}$ . In fact, compared to Corollary 1, the ingredients that determine the asymptotic order of the recursive estimator  $\widehat{\mathbf{F}}^{(m)}$  are similar to those for  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$ . However, the optimal convergence rate for  $\widehat{\mathbf{F}}^{(m)}$  is  $O_P(\delta_{NT}^{-1})$  which is much slower than  $O_P\left(\frac{1}{\sqrt{NT}}\right)$  that is the best rate for the recursive estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$ , since  $\widehat{\mathbf{F}}^{(m)}$  grows with  $T$ .

When the initial least squares estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(0)}$  is consistent, the asymptotic order of the recursive estimator  $\widehat{\mathbf{F}}^{(m)}$  becomes rather simple and is formulated as below.

**Theorem 7.** *Under the assumptions of Theorem 3, we have for  $m \geq 1$ ,*

$$\frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{H}}_{(m)}^{-1} - \mathbf{F}^0 \right\| = O_P(\delta_{NT}^{-1}) + O_P \left( \left( \max \left( \gamma_{NT}^{1/2}, \delta_{NT}^{-1} \right) \right)^m \right) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 \right\|. \quad (4.14)$$

## 5 Simulation

In this section, we investigate further the recursive least-squares estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(m)}$  through Monte Carlo simulations<sup>1</sup>. In particular, we focus on providing numerical evidence for two major implications of the asymptotic theory that have been established in Section 4. The first implication is that when the initial OLS estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(0)}$  is inconsistent, then the recursive estimation does not necessarily yield consistency either. The second one is that the consistency of the initial OLS estimator will guarantee the recursive estimator to be also consistent, and increasing the number of iterations can push the convergence rate of the recursive estimator towards the optimal rate  $1/\sqrt{NT}$ .

**Example 1.** *Consider the following DGP*

$$\begin{aligned} Y_{it} &= X_{it}\beta_0 + \lambda_i + F_t + \lambda_i F_t + \epsilon_{it} \\ X_{it} &= \lambda_i + F_t + \lambda_i F_t + \eta_{it} \end{aligned}$$

where for each  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ , the factor loadings  $\lambda_i \sim i.i.d.N(1, 6)$ ,  $\eta_{it} \sim i.i.d.N(0, 1)$  and  $\beta_0 = 1$ . The factor  $F_t$  is generated *i.i.d.* from  $N(0, 1)$  and is then normalized to satisfy identification restriction. The error term  $\epsilon_{it}$ , following Chen et al. (2012), is generated by

$$\epsilon_{it} = 0.3\epsilon_{i,t-1} + \varsigma_{it},$$

<sup>1</sup>For the simulation study in each example, we consider 1000 replications.

where  $\varsigma_{it}$  is cross-sectionally dependent but serially independent. Specifically, let  $\varsigma_t = (\varsigma_{1t}, \varsigma_{2t}, \dots, \varsigma_{Nt})'$  and we generate  $\{\varsigma_t | t = 1, 2, \dots, T\}$  as a sequence of i.i.d. Gaussian random vectors with zero mean and covariance matrix  $\Sigma_\varsigma = (\sigma_{ij})_{N \times N}$  the element of which  $\sigma_{ij} = 0.5^{|j-i|}$  for  $1 \leq i, j \leq N$ .

Table 1 summarises the performance of initial OLS estimator and resulting recursive estimators at various iterative steps in this example. We consider Two stopping rules for implementing recursive estimators in practice where  $\text{eps} = 10^{-3}$  and  $10^{-4}$  respectively. The results clearly show that in the presence of the endogeneity the initial OLS estimator is not consistent and recursive estimators are inconsistent either, since when either  $(N, T)$  or the number of iteration increases these estimators will remain distant from the true parameter.

Table 1: The performance of initial OLS estimator versus recursive estimators at various iterative steps in Example 1 with  $\beta_0 = 1$ . We consider recursive estimators using two different stopping rules for achieving numerical convergence in practice where  $\text{eps} = 10^{-3}$  and  $10^{-4}$  respectively.

$T/N$	OLS(0)	1	10	50	100	150	180	200	250	300	500
$\text{eps} = 10^{-3}$											
100/30	1.941	1.941	1.695	1.384	1.364	1.363	1.363	1.363	1.363	1.363	1.362
100/50	1.943	1.943	1.651	1.327	1.308	1.307	1.307	1.307	1.307	1.307	1.307
100/80	1.944	1.944	1.624	1.294	1.275	1.274	1.274	1.274	1.274	1.274	1.274
100/100	1.943	1.943	1.614	1.287	1.269	1.268	1.268	1.268	1.268	1.268	1.268
100/200	1.944	1.944	1.606	1.284	1.268	1.267	1.267	1.267	1.267	1.267	1.267
100/300	1.944	1.944	1.605	1.284	1.268	1.267	1.267	1.267	1.267	1.267	1.267
300/300	1.944	1.944	1.602	1.281	1.266	1.266	1.266	1.266	1.266	1.266	1.266
$\text{eps} = 10^{-4}$											
100/30	1.941	1.941	1.694	1.391	1.371	1.370	1.370	1.370	1.370	1.370	1.370
100/50	1.942	1.942	1.648	1.328	1.309	1.308	1.308	1.308	1.308	1.308	1.308
100/80	1.943	1.943	1.620	1.293	1.275	1.275	1.274	1.274	1.274	1.274	1.274
100/100	1.943	1.943	1.614	1.285	1.267	1.267	1.267	1.267	1.267	1.267	1.267
100/200	1.944	1.944	1.606	1.284	1.267	1.266	1.266	1.266	1.266	1.266	1.266
100/300	1.944	1.944	1.605	1.283	1.267	1.267	1.267	1.266	1.266	1.266	1.266
300/300	1.944	1.944	1.601	1.281	1.266	1.266	1.266	1.266	1.266	1.266	1.266

**Example 2.** We now simulate the example shown in Theorem 5, in which recursive estimation still leads to inconsistency. Precisely, we consider  $Y_{it} = X_{it}\beta_0 + \lambda_i F_t + \epsilon_{it}$  where  $X_{it} = \gamma_i G_t$ ,  $\beta_0 = 2$ ,  $G_t$  and  $\gamma_i$  are i.i.d.  $N(0, 1)$  across  $i$  and  $t$ . The number of factors  $r = 1$  and  $\lambda_i = \phi \gamma_i$  for each  $i$  and  $\phi = 2$ .  $F_t = G_t + Q_t$  where  $Q_t$  are generated i.i.d. following  $N(0, 1)$ .  $F_t$  is then normalized to satisfy identification restriction. The error term  $\epsilon_{it}$  is generated as in Example 1.

As can be seen, the numerical results in Table 2 actually implies that  $\widehat{\beta}_{NT}^{(m)} \xrightarrow{i.p.} \beta_0 + \phi = 4$  which confirms the finding in Theorem 5, noting that recursive estimators take values that are approximately 4 at all steps.

Table 2: The performance of initial OLS estimator and recursive estimators at various iterative steps in Example 2 with  $\beta_0 = 2$ . We consider recursive estimators using two different stopping rules for achieving numerical convergence in practice where  $\text{eps} = 10^{-3}$  and  $10^{-4}$  respectively.

$T/N$	OLS(0)	1	10	50	100	150	180	200	250	300	500
$\text{eps} = 10^{-3}$											
100/30	4.004	4.004	4.004	4.005	4.007	4.008	4.008	4.008	4.009	4.010	4.010
100/50	3.999	3.999	3.999	3.999	4.000	4.000	4.000	4.000	4.000	4.001	4.001
100/80	4.001	4.001	4.001	4.000	3.998	3.997	3.996	3.995	3.994	3.992	3.986
100/100	4.005	4.005	4.004	4.003	4.002	4.001	4.000	4.000	3.999	3.998	3.993
100/200	3.998	3.998	3.998	3.998	3.998	3.997	3.997	3.997	3.997	3.997	3.996
100/300	4.010	4.010	4.010	4.011	4.011	4.012	4.013	4.013	4.014	4.015	4.018
300/300	3.994	3.994	3.994	3.994	3.993	3.993	3.993	3.993	3.993	3.993	3.992
$\text{eps} = 10^{-4}$											
100/30	4.004	4.004	4.004	4.005	4.007	4.008	4.010	4.010	4.012	4.014	4.022
100/50	3.998	3.998	3.998	3.996	3.994	3.993	3.992	3.991	3.990	3.989	3.986
100/80	3.995	3.995	3.995	3.995	3.994	3.994	3.994	3.994	3.993	3.993	3.991
100/100	4.008	4.008	4.008	4.007	4.005	4.003	4.002	4.002	4.000	3.999	3.994
100/200	3.998	3.998	3.998	3.997	3.997	3.996	3.996	3.995	3.995	3.995	3.993
100/300	3.990	3.990	3.990	3.990	3.991	3.991	3.991	3.991	3.991	3.992	3.992
300/300	4.007	4.007	4.007	4.007	4.006	4.006	4.005	4.005	4.005	4.004	4.002

**Example 3.** Consider the following panel data model

$$Y_{it} = X_{it}\beta + \lambda_i F_t + \epsilon_{it}, \quad (5.1)$$

$$X_{it} = \psi_i F_t + \eta_{it}, \quad (5.2)$$

for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ . Notice that (5.1) and (5.2) are the simplified version of the model considered in Pesaran (2006) with observable common shocks being removed. In this model, we have one individual-specific regressor (i.e.  $p = 1$ ) with the corresponding coefficients  $\beta = 1$  and one unobserved common factor ( $r = 1$ ). We assume that the regressors are also driven by the same factor  $F_t$  but multiplied by a different loading term  $\psi_i$  which is non-random. The factor  $F_t$  is generated i.i.d. from  $N(0, 1)$  and the factor loading  $\lambda_i \sim i.i.d.N(1, 1)$ . The idiosyncratic errors in the factor models for the regressors (5.2) are generated i.i.d. as  $\eta_{it} \sim N(0, 0.2^2)$  across  $i$  and  $t$ . We also generate the error term  $\epsilon_{it}$  as in the previous examples.

In this example, we will compare the average number of recursive steps required in simulation studies with the theoretical number obtained by equation (4.9) in Theorem 4 when the initial OLS



estimator is consistent. Simple calculations can yield

$$\begin{aligned}
\eta_{NT} &= \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} (X_{it} X'_{jt} X_{is} X'_{js}) = \frac{1}{N^2 T^2} \sum_{i,j=1}^N \mathbb{E} (X'_i X_j X'_i X_j) \\
&= \mathbb{E} \left( \frac{1}{N} \sum_{i=1}^N \psi_i^2 \right)^2 + 2\mathbb{E}\eta_{it}^2 \left( \frac{1}{N} + \frac{1}{T} + \frac{1}{NT} \right) \mathbb{E} \left( \frac{1}{N} \sum_{i=1}^N \psi_i^2 \right) \\
&+ \frac{N+T-2}{NT} (\mathbb{E}\eta_{it}^2)^2 + \frac{1}{NT} \mathbb{E}\eta_{it}^4 \asymp \left( \frac{1}{N} \sum_{i=1}^N \psi_i^2 \right)^2
\end{aligned} \tag{5.3}$$

and

$$\begin{aligned}
\gamma_{NT} &= \frac{1}{N^2 T^2} \sum_{i,j=1}^N \mathbb{E} (\lambda'_i F' X_i X'_j F \lambda_j) \\
&= \mathbb{E} \left( \frac{1}{N} \sum_{i=1}^N \psi_i \lambda_i \right)^2 + \frac{1}{NT} \mathbb{E}\lambda_i^2 \cdot \mathbb{E}\eta_{it}^2 \asymp \left( \frac{1}{N} \sum_{i=1}^N \psi_i \right)^2.
\end{aligned} \tag{5.4}$$

Notice that the magnitude of  $\eta_{NT}$  in (5.3) depends on the quantity  $\overline{\psi^2} = \frac{1}{N} \sum_{i=1}^N \psi_i^2$ .

Three scenarios will be considered for two different error tolerance ( $eps = 10^{-6}$  and  $eps = 10^{-8}$ ) used practically. First, we will restrict  $\psi_i$  to be zero for each  $i$  and this leads to the case where the regressors  $X_{it}$  are independent of the factor residuals. Second, we will set  $\psi_i = e^{-i}$  for each  $i$  and under this situation  $0 < \sum_{i=1}^N \psi_i^2 < 0.5$  so that the factors are weak. In fact, the value of  $\sum_{i=1}^N \psi_i^2$  has been used by Lam et al. (2011, p. 905) to reflect the strength of the factors. Finally, we assume that  $\sum_{i=1}^N \psi_i^2 \asymp N^{0.2}$  (i.e.,  $\overline{\psi^2} \asymp N^{-0.8}$  noting that the notation " $a \asymp b$ " denotes that  $a = O_P(b)$  and  $b = O_P(a)$ ). This scenario is accomplished in simulation by assigning the values of  $\psi_i$  appropriately to be 0 or 1 for each  $i$ . In this case, the factors strength is clearly growing with  $N$  and relatively stronger compared to the second case where the factors strength is constant when  $N$  increases. Moreover, it is worth mentioning that the magnitude of  $\gamma_{NT}$  is determined by  $\overline{\psi} = \frac{1}{N} \sum_{i=1}^N \psi_i$  which can be regarded as the average effect of the common shock among individual units. One should note that  $\overline{\psi}^2 \leq \overline{\psi^2}$  always. Therefore, it holds that  $0 \leq \overline{\psi} = \sqrt{\overline{\psi^2}} \leq 1$  for all three scenarios we consider.

Table 3 shows that the average number of recursive steps in practice with the theoretical number obtained by equation (4.9) are very close. It is worth noting that we have also examined circumstances where the strength of the factors is even stronger than cases mentioned above. Specifically, suppose that  $\sum_{i=1}^N \psi_i^2 \asymp N^{1-\delta}$  (i.e.,  $\overline{\psi^2} \asymp N^{-\delta}$ ) where the index  $0 \leq \delta \leq 1$  indicates the factors strength. The situation in which  $\delta = 0$  corresponds to the standard "strong factors" assumption while  $\delta > 0$  implies weak factors (e.g., Lam et al., 2011). In general, we find that the analytical approximation will be less accurate for the number of recursive steps required in practice as  $\delta$  decreases. This is exactly what Theorem 3 implies, that is, when the factors strength increases the convergence rate of the recursive estimator will decrease, and in the extreme case where  $\delta = 0$  so that each factor loading is of exactly constant order the initial OLS is not consistent and the recursive estimation is very unlikely to achieve

consistency either. As a result, the analytical approximation depending on asymptotics becomes less reliable practically when the factors strength rises up.

Table 3: Comparison between recursive steps required practically (No. iter.) and analytically (Iter. Theo.) for choices of  $\psi_i$ . In this example, we consider recursive estimators using two different stopping rules for achieving numerical convergence in practice where  $\text{eps} = 10^{-6}$  and  $10^{-8}$  respectively.

$T/N$	No. iter.	Iter. Theo.	$T/N$	No. iter.	Iter. Theo.
$\psi_i = 0, \text{eps} = 10^{-6}$			$\psi_i = 0, \text{eps} = 10^{-8}$		
100/30	4.80	2.69	100/30	6.18	3.87
100/50	4.34	2.33	100/50	5.72	3.36
100/80	3.97	2.04	100/80	5.22	3.00
100/100	3.94	1.93	100/100	5.05	2.85
100/200	3.82	1.65	100/200	4.88	2.44
100/300	3.73	1.50	100/300	4.79	2.25
$\psi_i = e^{-i}, \text{eps} = 10^{-6}$			$\psi_i = e^{-i}, \text{eps} = 10^{-8}$		
100/30	4.99	2.76	100/30	6.60	3.95
100/50	4.60	2.36	100/50	5.92	3.43
100/80	4.11	2.07	100/80	5.52	3.03
100/100	4.00	1.94	100/100	5.23	2.86
100/200	3.89	1.66	100/200	4.89	2.45
100/300	3.76	1.50	100/300	4.85	2.24
$\psi_i = 1/N^{0.2}, \text{eps} = 10^{-6}$			$\psi_i = 1/N^{0.2}, \text{eps} = 10^{-8}$		
100/30	7.14	2.99	100/30	9.4	4.2
100/50	6.67	2.59	100/50	8.7	3.6
100/80	5.72	2.25	100/80	7.5	3.2
100/100	5.37	2.11	100/100	7.0	3.0
100/200	4.64	1.76	100/200	5.9	2.5
100/300	4.46	1.62	100/300	5.8	2.3

**Example 4.** Consider a DGP with factor components being fixed:  $Y_{it} = X_{it}\beta_0 + \lambda_i F_t + \epsilon_{it}$  where  $X_{it} \sim N(1, 1)$  and  $\beta_0 = 1$ . The number of factors  $r = 2$  and the elements of  $\lambda_i$  and  $F_t$  are generated i.i.d. (only once outside the simulation loop so they are fixed) following  $N(3, 1)$  respectively.  $F_t$  is then normalized to satisfy identification restriction. The error term  $\epsilon_{it}$  is generated following the same equation as in Example 1.

The results in Table 4 show the performance of the initial OLS estimator and recursive estimators at various iterative steps. A general observation is that recursive estimation can yield consistency in this scenario where the factor components are fixed, although the initial OLS estimator is inconsistent. This example implies that in the absence of the endogeneity recursion yields consistency despite the inconsistency of initial OLS estimator. The impression we can have combining this example with preceding ones in this section is that endogeneity between the regressors and the errors largely accounts for the failure in consistency of the recursive estimation.

Table 4: The performance of initial OLS estimator and recursive estimators at various iterative steps in Example 4 with  $\beta_0 = 1$ . We consider recursive estimators using two different stopping rules for achieving numerical convergence in practice where  $\text{eps} = 10^{-3}$  and  $10^{-4}$  respectively.

$T/N$	OLS(0)	1	10	50	100	150	180	200	250	300	500
$\text{eps} = 10^{-3}$											
100/30	-1.027	-1.027	-0.486	0.643	0.960	1.018	1.026	1.029	1.030	1.031	1.031
100/50	-0.869	-0.869	-0.370	0.658	0.943	0.994	1.002	1.004	1.005	1.006	1.006
100/80	-0.901	-0.901	-0.390	0.656	0.942	0.993	1.000	1.002	1.003	1.003	1.004
100/100	-0.879	-0.879	-0.368	0.668	0.947	0.995	1.001	1.003	1.005	1.005	1.005
100/200	-0.992	-0.992	-0.449	0.648	0.941	0.991	0.998	1.000	1.001	1.001	1.001
100/300	-0.929	-0.929	-0.401	0.663	0.945	0.993	0.999	1.001	1.002	1.002	1.002
300/300	-0.867	-0.867	-0.351	0.678	0.947	0.991	0.997	0.999	1.000	1.000	1.000
$\text{eps} = 10^{-4}$											
100/30	-0.774	-0.774	-0.304	0.684	0.962	1.013	1.020	1.022	1.024	1.024	1.024
100/50	-0.963	-0.963	-0.440	0.639	0.938	0.992	0.999	1.001	1.003	1.003	1.004
100/80	-0.867	-0.867	-0.363	0.664	0.942	0.990	0.997	0.999	1.000	1.001	1.001
100/100	-0.889	-0.889	-0.378	0.666	0.947	0.996	1.003	1.005	1.006	1.006	1.007
100/200	-0.955	-0.955	-0.421	0.657	0.944	0.993	0.999	1.001	1.003	1.003	1.003
100/300	-1.014	-1.014	-0.463	0.646	0.940	0.990	0.996	0.998	1.000	1.000	1.000
300/300	-0.954	-0.954	-0.414	0.665	0.945	0.992	0.998	0.999	1.001	1.001	1.001

## 6 Conclusion

This paper considered the recursive estimation approach for large panel data models with interactive fixed effects proposed by Bai (2009). We showed that statistical properties of recursive estimators depend heavily on the properties of the initial OLS estimator and gave conditions for recursive estimators to converge to the true parameters. A sufficient condition for the consistent recursive estimation is

the consistency of initial OLS estimation. As a matter of fact, when the initial OLS estimator is consistent, but with a convergence rate lower than  $\frac{1}{\sqrt{NT}}$ , the statistical properties of recursive estimator,  $\widehat{\beta}^{(m)}$ , is a function of the number of iterations,  $m$ . Moreover, strong dependence between regressors and the common factors in the errors (endogeneity) will lead to inconsistency of both the initial OLS and resulting recursive estimators.

## 7 Appendix: Proofs of Theorems

This section consists of two sets of results. First, we provide proofs of all the theorems stated in the main texts. Second, we include lemmas along with proofs, which are important axillary results for deriving main theorems. Before showing technical details any further, we need to introduce some notations used in what follows. Let  $\mathbf{X}_{it}$ ,  $\mathbf{x}_{i(k)}$  and  $x_{it(k)}$  represent different vectors related to the regressor  $\mathbf{X}$ . Precisely,  $\mathbf{X}_{it}$  is a  $p \times 1$  vector for any  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ ;  $\mathbf{x}_{i(k)}$  is a  $T \times 1$  vector for any  $i = 1, 2, \dots, N$  and  $k = 1, 2, \dots, p$ ; and  $x_{it(k)}$  is a scalar for any  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ ;  $k = 1, 2, \dots, p$ .  $M$  denotes a constant which might vary from line to line.

### 7.1 Proofs of Main Results

#### 7.1.1 Proof of Theorem 1

Theorem 1 provides the asymptotic order of the initial OLS estimator  $\widehat{\beta}_{NT}^{(0)}$  derived by ignoring the common factor components in model (2.1).

*Proof of Theorem 1.* First notice that the initial OLS estimator  $\widehat{\beta}_{NT}^{(0)}$  can be written as

$$\widehat{\beta}_{NT}^{(0)} = \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{xy}}, \quad (7.1)$$

where

$$\mathbf{S}_{\mathbf{xx}} = \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{X}_i, \quad \mathbf{S}_{\mathbf{xy}} = \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{Y}_i.$$

Since the true parameter  $\beta_0$  satisfies model (2.1), we have the relationship between the initial estimator  $\widehat{\beta}_{NT}^{(0)}$  and  $\beta_0$  as follows

$$\widehat{\beta}_{NT}^{(0)} - \beta_0 = \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{xc}} + \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{x}\epsilon}, \quad (7.2)$$

where

$$\mathbf{S}_{\mathbf{xc}} = \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F}^0 \lambda_i^0, \quad \mathbf{S}_{\mathbf{x}\epsilon} = \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \epsilon_i.$$

In order to obtain the order of  $\widehat{\beta}_{NT}^{(0)}$ , it suffices to evaluate the orders of  $\mathbf{S}_{\mathbf{xx}}$ ,  $\mathbf{S}_{\mathbf{xc}}$  and  $\mathbf{S}_{\mathbf{x}\epsilon}$  based on (7.2).

Assumption 2(2) entails that  $\mathbf{S}_{\mathbf{xx}} \xrightarrow{i.p.} \Sigma_{\mathbf{x}}$  as  $N, T$  tend to infinity jointly. Next, we examine  $\mathbf{S}_{\mathbf{xc}}$  and  $\mathbf{S}_{\mathbf{x}\epsilon}$ .

For  $\mathbf{S}_{\mathbf{x}\epsilon}$ ,

$$\begin{aligned}
\mathbb{E} \|\mathbf{S}_{\mathbf{x}\epsilon}\|^2 &= \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E} \left( \epsilon_i' \mathbf{X}_i \mathbf{X}_j' \epsilon_j \right) = \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \sum_{k=1}^p \mathbb{E} (\epsilon_{it} x_{itk} x_{j sk} \epsilon_{js}) \\
&= \frac{1}{N^2 T^2} \sum_{k=1}^p \sum_{i,j=1}^N \sum_{t,s=1}^T [\mathbb{E} (x_{itk} \epsilon_{it} x_{j sk} \epsilon_{js}) - \mathbb{E} (x_{itk} \epsilon_{it}) \cdot \mathbb{E} (x_{j sk} \epsilon_{js})] \\
&\quad + \frac{1}{N^2 T^2} \sum_{k=1}^p \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} (x_{itk} \epsilon_{it}) \cdot \mathbb{E} (x_{j sk} \epsilon_{js}) \\
&\leq \frac{1}{N^2 T^2} \sum_{k=1}^p \sum_{i,j=1}^N \sum_{t,s=1}^T 10 \cdot \alpha_{ij}^{\frac{\delta}{2+\delta}} (|t-s|) \cdot \left[ \mathbb{E} |x_{itk} \epsilon_{it}|^{2+\delta} \right]^{\frac{1}{2+\delta}} \cdot \left[ \mathbb{E} |x_{j sk} \epsilon_{js}|^{2+\delta} \right]^{\frac{1}{2+\delta}} \\
&\quad + \frac{1}{N^2 T^2} \sum_{k=1}^p \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} [x_{itk} \cdot \mathbb{E} (\epsilon_{it} | x_{itk})] \cdot \mathbb{E} [x_{j sk} \cdot \mathbb{E} (\epsilon_{js} | x_{j sk})] \\
&= O \left( \frac{1}{NT} \right), \tag{7.3}
\end{aligned}$$

where the last inequality holds due to Lemma 2 and the facts that

$$\mathbb{E} |x_{j sk} \epsilon_{js}|^{2+\delta} \leq \left( \mathbb{E} |x_{j sk}|^{4+2\delta} \right)^{\frac{1}{2}} \cdot \left( \mathbb{E} |\epsilon_{js}|^{4+2\delta} \right)^{\frac{1}{2}} \leq M \tag{7.4}$$

according to Assumptions 2.2 and 2.3,

$$\mathbb{E} (\epsilon_{js} | x_{j sk}) = 0 \tag{7.5}$$

by Assumption 2.3, and

$$\sum_{t,s=1}^T \sum_{i,j=1}^N \alpha_{ij}^{\frac{\delta}{2+\delta}} (|t-s|) = \sum_{i,j=1}^N \alpha_{ij}^{\frac{\delta}{2+\delta}} (0) + 2 \sum_{i,j=1}^N \sum_{\tau=1}^{T-1} (T-\tau) \alpha_{ij}^{\frac{\delta}{2+\delta}} (\tau) = O(NT)$$

guaranteed by Assumption 3.2.

In view of (7.3), we can obtain

$$\mathbf{S}_{\mathbf{x}\epsilon} = O_P \left( \frac{1}{\sqrt{NT}} \right). \tag{7.6}$$

The order of  $\mathbf{S}_{\mathbf{x}\epsilon}$  depends on the relationship between the regressors  $\{\mathbf{X}_i : i = 1, 2, \dots, N\}$  and the common factor components  $\{(\boldsymbol{\lambda}_i^0, \mathbf{F}^0) : i = 1, 2, \dots, N\}$ . To understand clearly the order of  $\mathbf{S}_{\mathbf{x}\epsilon}$ , we will consider the following two extreme cases.

1.  $\{\mathbf{X}_i : i = 1, 2, \dots, N\}$  and  $\{(\boldsymbol{\lambda}_i^0, \mathbf{F}^0) : i = 1, 2, \dots, N\}$  are independent. In this case, we have

$$\mathbb{E} \mathbf{S}_{\mathbf{x}\epsilon} = \frac{1}{NT} \sum_{i=1}^N \mathbb{E} (\mathbf{X}_i)' \mathbb{E} (\mathbf{F}^0 \boldsymbol{\lambda}_i^0) = 0, \tag{7.7}$$

where the last equality is valid due to Assumption 2.1. Moreover,

$$\begin{aligned}
\mathbb{E} \|\mathbf{S}_{\mathbf{x}\mathbf{c}}\|^2 &= \frac{1}{N^2 T^2} \sum_{i,j=1}^N \mathbb{E} \left( \boldsymbol{\lambda}_i^{0'} \mathbf{F}^{0'} \mathbf{X}_i \mathbf{X}_j' \mathbf{F}^0 \boldsymbol{\lambda}_j^0 \right) \\
&= \frac{1}{N^2 T^2} \sum_{t,s=1}^T \sum_{k=1}^p \mathbb{E} \left( \boldsymbol{\lambda}_i^{0'} \mathbf{F}_t^0 x_{itk} x_{j sk} \mathbf{F}_s^{0'} \boldsymbol{\lambda}_j^0 \right) \\
&= \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \sum_{k=1}^p \mathbb{E} \left( \boldsymbol{\lambda}_i^{0'} \mathbf{F}_t^0 \boldsymbol{\lambda}_j^{0'} \mathbf{F}_s^0 \right) \cdot \mathbb{E}(x_{itk} x_{j sk}) \\
&\leq \frac{M}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \left| \mathbb{E} \left( \boldsymbol{\lambda}_i^{0'} \mathbf{F}_t^0 \boldsymbol{\lambda}_j^{0'} \mathbf{F}_s^0 \right) \right| \\
&\leq \frac{M}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \alpha_{ij}^{\frac{\delta}{2+\delta}} (|t-s|) \left[ \mathbb{E} \left| \boldsymbol{\lambda}_i^{0'} \mathbf{F}_t^0 \right|^{2+\delta} \right]^{\frac{1}{2+\delta}} \cdot \left[ \mathbb{E} \left| \boldsymbol{\lambda}_j^{0'} \mathbf{F}_s^0 \right|^{2+\delta} \right]^{\frac{1}{2+\delta}} \\
&= O \left( \frac{1}{NT} \right)
\end{aligned} \tag{7.8}$$

in which the last inequality is a result from Lemma 2 with Assumption 2.1 and Assumption 3.2.

Thus, it follows from (7.7) and (7.8) that  $\mathbf{S}_{\mathbf{x}\mathbf{c}} = O_P \left( \frac{1}{\sqrt{NT}} \right)$ .

2. The regressors  $\{\mathbf{X}_i : i = 1, 2, \dots, N\}$  and the common factor components  $\{(\boldsymbol{\lambda}_i^0, \mathbf{F}^0) : i = 1, 2, \dots, N\}$  are dependent in the sense that  $\mathbf{X}_i = \mathbf{F} \mathbf{q}_i + \boldsymbol{\eta}_i$ .

For this scenario, we have

$$\begin{aligned}
\mathbf{S}_{\mathbf{x}\mathbf{c}} &= \frac{1}{NT} \sum_{i=1}^N (\mathbf{F}^0 \mathbf{q}_i + \boldsymbol{\eta}_i)' \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \\
&= \frac{1}{NT} \sum_{i=1}^N \mathbf{q}_i' \mathbf{F}^{0'} \mathbf{F}^0 \boldsymbol{\lambda}_i^0 + \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\eta}_i' \mathbf{F}^0 \boldsymbol{\lambda}_i^0.
\end{aligned} \tag{7.9}$$

It follows from the independence between  $\boldsymbol{\eta}_i$  and  $\mathbf{F}^0 \boldsymbol{\lambda}_i^0$  that

$$\mathbb{E} \left( \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\eta}_i' \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \right) = \frac{1}{NT} \sum_{i=1}^N \mathbb{E}(\boldsymbol{\eta}_i)' \mathbb{E}(\mathbf{F}^0 \boldsymbol{\lambda}_i^0) = 0. \tag{7.10}$$

Moreover,

$$\begin{aligned}
& \mathbb{E} \left\| \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \right\|^2 = \left| \frac{1}{N^2 T^2} \sum_{i,j=1}^N \mathbb{E} \left( \boldsymbol{\lambda}_i^{0'} \mathbf{F}^{0'} \boldsymbol{\eta}_i \boldsymbol{\eta}'_j \mathbf{F}^0 \boldsymbol{\lambda}_j^0 \right) \right| \\
&= \left| \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{k_1, k_2=1}^r \sum_{t, s=1}^T \mathbb{E} \left( \lambda_{ik_1}^0 F_{tk_1} \eta_{it} \eta_{js} F_{sk_2} \lambda_{jk_2}^0 \right) \right| \\
&= \left| \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{k_1, k_2=1}^r \sum_{t, s=1}^T \mathbb{E} \left( \lambda_{ik_1}^0 F_{tk_1}^0 \lambda_{jk_2}^0 F_{sk_2}^0 \right) \cdot \mathbb{E} \left( \eta_{it} \eta_{js} \right) \right| \\
&\leq \frac{M}{N^2 T^2} \sum_{i,j=1}^N \sum_{t, s=1}^T \alpha_{ij}^{\frac{\delta}{4+\delta}} (|t-s|) \cdot \left[ \mathbb{E} \left( \boldsymbol{\lambda}_i^{0'} \mathbf{F}_t^0 \right)^2 \right]^{\frac{1}{2}} \cdot \left[ \mathbb{E} \left( \boldsymbol{\lambda}_j^{0'} \mathbf{F}_s^{0'} \right)^{2+\delta} \right]^{\frac{1}{2+\delta}} = O \left( \frac{1}{NT} \right),
\end{aligned} \tag{7.11}$$

where the last equality holds due to Assumption 2.1 and Assumption 3.2. Based on (7.10) and (7.11), we have

$$\frac{1}{NT} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \xrightarrow{i.p.} 0, \quad \text{as } N, T \rightarrow \infty. \tag{7.12}$$

Next, we show that

$$\frac{1}{NT} \sum_{i=1}^N \mathbf{q}'_i \mathbf{F}^{0'} \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \xrightarrow{i.p.} \mathbb{E} \left( \mathbf{q}_i \right)' \cdot \mathbb{E} \left( \boldsymbol{\lambda}_i^0 \right). \tag{7.13}$$

In fact,

$$\begin{aligned}
& \mathbb{E} \left[ \frac{1}{NT} \sum_{i=1}^N \mathbf{q}'_i \mathbf{F}^{0'} \mathbf{F}^0 \boldsymbol{\lambda}_i^0 - \mathbb{E} \left( \frac{1}{NT} \sum_{i=1}^N \mathbf{q}'_i \mathbf{F}^{0'} \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \right) \right]^2 \\
&= \frac{1}{N^2} \sum_{i,j=1}^N \mathbb{E} \left( \mathbf{q}'_i \boldsymbol{\lambda}_i^0 \boldsymbol{\lambda}_j^{0'} \mathbf{q}_j \right) - \frac{1}{N^2} \sum_{i,j=1}^N \mathbb{E} \left( \mathbf{q}_i \right)' \cdot \mathbb{E} \left( \boldsymbol{\lambda}_i^0 \right) \cdot \mathbb{E} \left( \mathbf{q}_j \right)' \cdot \mathbb{E} \left( \boldsymbol{\lambda}_j^0 \right) \\
&= \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left( \mathbf{q}'_i \boldsymbol{\Sigma}_\Lambda \mathbf{q}_i \right) - \frac{1}{N^2} \sum_{i=1}^N \left[ \mathbb{E} \left( \mathbf{q}_i \right)' \mathbb{E} \left( \boldsymbol{\lambda}_i \right) \right]^2 = O \left( \frac{1}{N} \right),
\end{aligned} \tag{7.14}$$

where the first equality uses  $\frac{1}{T} \mathbf{F}^{0'} \mathbf{F}^0 = \mathbf{I}_r$  and the last equality utilizes the assumption that  $\mathbf{q}_i$  are independent of  $\boldsymbol{\lambda}_i$  and they are both independent across different  $i = 1, 2, \dots, N$ .

Then, due to (7.9), (7.12) and (7.13), it holds

$$\mathbf{S}_{\mathbf{x}\mathbf{c}} \xrightarrow{i.p.} \mathbb{E} \left( \mathbf{q}'_i \right) \mathbb{E} \left( \boldsymbol{\lambda}_i \right) \asymp 1. \tag{7.15}$$

From (7.2) and (7.6), we have

$$\left| \widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 \right| \asymp \left( \max \left( \mathbf{S}_{\mathbf{x}\mathbf{c}}, \frac{1}{\sqrt{NT}} \right) \right). \tag{7.16}$$

By the definition of  $\gamma_{NT}$  in (4.2),

$$\gamma_{NT} = \mathbb{E} \|\mathbf{S}_{\mathbf{x}\mathbf{c}}\|^2. \quad (7.17)$$

So

$$\left| \widehat{\beta}_{NT}^{(0)} - \beta_0 \right| = O_P \left( \max \left( \gamma_{NT}^{1/2}, \frac{1}{\sqrt{NT}} \right) \right). \quad (7.18)$$

□

Since the recursive estimator  $\widehat{\beta}_{NT}^{(m+1)}$  is derived based on  $(\widehat{\Lambda}^{(m)}, \widehat{\mathbf{F}}^{(m)})$ , we will provide the proof of Theorem 6 ahead of Theorem 2.

*Proof of Theorem 6.* With the estimator  $\widehat{\beta}_{NT}^{(m)}$  at the  $m$ -th step, we obtain the estimator  $(\widehat{\mathbf{F}}^{(m)}, \widehat{\Lambda}^{(m)})$  at the  $m$ -th step by

$$\left[ \frac{1}{NT} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \widehat{\beta}_{NT}^{(m)}) (\mathbf{Y}_i - \mathbf{X}_i \widehat{\beta}_{NT}^{(m)})' \right] \widehat{\mathbf{F}}^{(m)} = \widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{V}}_{NT}^{(m)} \quad (7.19)$$

and

$$\widehat{\lambda}_i^{(m)} = \frac{1}{T} \widehat{\mathbf{F}}^{(m)'} (\mathbf{Y}_i - \mathbf{X}_i \widehat{\beta}_{NT}^{(m)}).$$

Substituting  $\mathbf{Y}_i$  with the expression  $\mathbf{Y}_i = \mathbf{X}_i \beta_0 + \mathbf{F}^0 \lambda_i^0 + \epsilon_i$ , we can get

$$\widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{V}}^{(m)} = \sum_{h=1}^9 \mathbf{I}_h^{(m)}, \quad (7.20)$$

where

$$\mathbf{I}_1^{(m)} = \frac{1}{NT} \sum_{i=1}^N \sum_{k=1}^p \sum_{w=1}^p \mathbf{x}_{i(k)} \mathbf{x}_{i(w)}' \widehat{\mathbf{F}}^{(m)} (\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}) (\beta_{0h} - \widehat{\beta}_{NT,w}^{(m)}),$$

$$\mathbf{I}_2^{(m)} = \frac{1}{NT} \sum_{i=1}^N \sum_{k=1}^p \mathbf{x}_{i(k)} \lambda_i^0 \mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)} (\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}),$$

$$\mathbf{I}_3^{(m)} = \frac{1}{NT} \sum_{i=1}^N \sum_{k=1}^p \mathbf{x}_{i(k)} \epsilon_i' \widehat{\mathbf{F}}^{(m)} (\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}),$$

$$\mathbf{I}_4^{(m)} = \frac{1}{NT} \sum_{i=1}^N \sum_{k=1}^p \mathbf{F}^0 \lambda_i^0 \mathbf{x}_{i(k)}' \widehat{\mathbf{F}}^{(m)} (\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}),$$

$$\mathbf{I}_5^{(m)} = \frac{1}{NT} \sum_{i=1}^N \sum_{k=1}^p \epsilon_i \mathbf{x}_{i(k)}' \widehat{\mathbf{F}}^{(m)} (\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}),$$

$$\mathbf{I}_6^{(m)} = \frac{1}{NT} \sum_{i=1}^N \mathbf{F}^0 \lambda_i^0 \epsilon_i' \widehat{\mathbf{F}}^{(m)}, \quad \mathbf{I}_7^{(m)} = \frac{1}{NT} \sum_{i=1}^N \epsilon_i \lambda_i^0 \mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)},$$

$$\mathbf{I}_8^{(m)} = \frac{1}{NT} \sum_{i=1}^N \epsilon_i \epsilon_i' \widehat{\mathbf{F}}^{(m)}, \quad \mathbf{I}_9^{(m)} = \frac{1}{NT} \sum_{i=1}^N \mathbf{F}^0 \lambda_i^0 \lambda_i^0 \mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)}.$$



Subtracting the term  $\mathbf{I}_9^{(m)}$  on both sides of (7.20) and multiplying  $\left(\frac{\mathbf{F}^{0'}\widehat{\mathbf{F}}^{(m)}}{T}\right)^{-1}\left(\frac{\boldsymbol{\Lambda}^{0'}\boldsymbol{\Lambda}^0}{N}\right)^{-1}$ , we have

$$\widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1} - \mathbf{F}^0 = \left(\sum_{h=1}^8 \mathbf{I}_h^{(m)}\right) \left(\frac{\mathbf{F}^{0'}\widehat{\mathbf{F}}^{(m)}}{T}\right)^{-1} \left(\frac{\boldsymbol{\Lambda}^{0'}\boldsymbol{\Lambda}^0}{N}\right)^{-1}, \quad (7.21)$$

where

$$\mathbf{H}_{(m)}^{-1} = \widehat{\mathbf{V}}_{NT}^{(m)} \left(\frac{\mathbf{F}^{0'}\widehat{\mathbf{F}}^{(m)}}{T}\right)^{-1} \left(\frac{\boldsymbol{\Lambda}^{0'}\boldsymbol{\Lambda}^0}{N}\right)^{-1}. \quad (7.22)$$

Next, we investigate the eight terms  $\frac{1}{\sqrt{T}}\|\mathbf{I}_h^{(m)}\|$ ,  $h = 1, 2, \dots, 8$ . In fact, we will show that the terms  $\frac{1}{\sqrt{T}}\mathbf{I}_d$ ,  $d = 3, 5, 6, 7, 8$  are of order  $\frac{1}{\sqrt{N}}$ .

First,

$$\frac{1}{\sqrt{T}}\|\mathbf{I}_3^{(m)}\| \leq \sum_{k=1}^p \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \boldsymbol{\epsilon}'_i \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot |\beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)}| = O_P\left(\frac{1}{\sqrt{N}}\right),$$

where the last equality holds noting (7.49) of Lemma 2 and that

$$\left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| = \sqrt{\text{tr} \left( \frac{\widehat{\mathbf{F}}^{(m)'}\widehat{\mathbf{F}}^{(m)}}{T} \right)} = \sqrt{\text{tr}(\mathbf{I}_r)} = \sqrt{r}. \quad (7.23)$$

Similarly, we will have the following results. For  $\mathbf{I}_5$ ,

$$\frac{1}{\sqrt{T}}\|\mathbf{I}_5^{(m)}\| \leq \sum_{k=1}^p \left\| \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i \mathbf{x}'_{i(k)} \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot |\beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)}| = O_P\left(\frac{1}{\sqrt{N}}\right),$$

where we obtain the last equality using (7.49) of Lemma 2 and (7.23).

For  $\mathbf{I}_6^{(m)}$ ,

$$\frac{1}{\sqrt{T}}\|\mathbf{I}_6^{(m)}\| \leq \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \boldsymbol{\epsilon}_i \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| = O_P\left(\frac{1}{\sqrt{N}}\right)$$

noting (7.50) of Lemma 2 and (7.23).

For  $\mathbf{I}_7^{(m)}$ ,

$$\frac{1}{\sqrt{T}}\|\mathbf{I}_7^{(m)}\| \leq \left\| \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i \boldsymbol{\lambda}_i^{0'} \mathbf{F}^{0'} \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| = O_P\left(\frac{1}{\sqrt{N}}\right),$$

where the last equality uses (7.50) of Lemma 2 and the restriction (7.23).

For  $\mathbf{I}_8^{(m)}$ ,

$$\frac{1}{\sqrt{T}}\|\mathbf{I}_8^{(m)}\| \leq \left\| \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}'_i \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| = O_P\left(\frac{1}{\sqrt{N}}\right),$$

where the last equality comes from (7.48) of Lemma 2 and (7.23).

Combining these results and (7.21), we obtain that the leading terms of  $\frac{1}{\sqrt{T}}\left(\widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1} - \mathbf{F}^0\right)$  are  $\frac{1}{\sqrt{T}}\mathbf{I}_1^{(m)}$ ,

$\frac{1}{\sqrt{T}}\mathbf{I}_2^{(m)}$  and  $\frac{1}{\sqrt{T}}\mathbf{I}_4^{(m)}$ . So (4.12) in Theorem 6 can be derived.  $\square$

*Proof of Theorem 2.* With the estimator  $(\widehat{\Lambda}^{(m)}, \widehat{\mathbf{F}}^{(m)})$  at the  $m$ -th step, we can obtain the estimator  $\widehat{\beta}_{NT}^{(m+1)}$  at the next step by

$$\widehat{\beta}_{NT}^{(m+1)} = \left( \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{X}_i \right)^{-1} \left( \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{Y}_i \right).$$

Because  $\mathbf{Y}_i = \mathbf{X}_i \beta_0 + \mathbf{F}^0 \lambda_i^0 + \epsilon_i$ ,

$$\begin{aligned} \widehat{\beta}_{NT}^{(m+1)} - \beta_0 &= \left( \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{X}_i \right)^{-1} \left( \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{F}^0 \lambda_i^0 \right) \\ &\quad + \left( \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{X}_i \right)^{-1} \left( \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \epsilon_i \right), \end{aligned}$$

that is,

$$\left( \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{X}_i \right) (\widehat{\beta}_{NT} - \beta_0) = \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{F}^0 \lambda_i^0 + \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \epsilon_i.$$

Next, we examine the term  $\frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{F}^0 \lambda_i^0$ .

$$\begin{aligned} \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{F}^0 \lambda_i^0 &= \frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \left( \mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)} \widehat{\mathbf{H}}_{(m)}^{-1} \right) \lambda_i^0 \\ &= -\frac{1}{NT} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \left( \sum_{h=1}^8 \mathbf{I}_h^{(m)} \right) \left( \frac{\mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \left( \frac{\Lambda^{0'} \Lambda^0}{N} \right)^{-1} \lambda_i^0 =: \sum_{h=1}^8 \mathbf{J}_h^{(m)}. \end{aligned}$$

Then we will look at the asymptotic orders of the eight terms  $\mathbf{J}_h^{(m)}$ ,  $h = 1, 2, \dots, 8$ . In fact, we will show that

$$\begin{aligned} \mathbf{J}_3^{(m)} &= o_P(1) \cdot (\widehat{\beta}_{NT}^{(m)} - \beta_0), \quad \mathbf{J}_5^{(m)} = o_P(1) \cdot (\widehat{\beta}_{NT}^{(m)} - \beta_0), \\ \mathbf{J}_6^{(m)} &= o_P(1) \cdot (\widehat{\beta}_{NT}^{(m)} - \beta_0) + o_P\left(\frac{1}{\sqrt{NT}}\right), \quad \mathbf{J}_7^{(m)} = O_P\left(\frac{1}{\sqrt{NT}}\right), \\ \mathbf{J}_8^{(m)} &= o_P(1) \cdot (\widehat{\beta}_{NT}^{(m)} - \beta_0) + O_P\left(\frac{1}{\sqrt{NT}}\right) + O_P\left(\frac{1}{T}\right). \end{aligned} \tag{7.24}$$

For simplicity, we define

$$\mathbf{G}^{(m)} = \left( \frac{\mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \left( \frac{\Lambda^{0'} \Lambda^0}{N} \right)^{-1}. \tag{7.25}$$

For  $\mathbf{J}_3^{(m)}$ ,

$$\begin{aligned}
\left\| \mathbf{J}_3^{(m)} \right\| &= \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \frac{1}{NT} \sum_{j=1}^N \sum_{k=1}^p \mathbf{x}_{j(k)} \boldsymbol{\epsilon}'_j \widehat{\mathbf{F}}^{(m)} (\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}) \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \right\| \\
&\leq \sum_{k=1}^p \frac{1}{N\sqrt{T}} \left( \sum_{i=1}^N \|\mathbf{X}'_i\|^2 \right)^{1/2} \cdot \left( \sum_{i=1}^N \|\boldsymbol{\lambda}_i^0\|^2 \right)^{1/2} \cdot \|\mathbf{M}_{\widehat{\mathbf{F}}^{(m)}}\| \\
&\quad \cdot \left\| \frac{1}{NT} \sum_{j=1}^N \mathbf{x}_{j(k)} \boldsymbol{\epsilon}'_j \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot |\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}| \\
&= O_P \left( \frac{1}{\sqrt{N}} \right) \cdot \sum_{k=1}^p |\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}| \leq O_P \left( \frac{1}{\sqrt{N}} \right) \cdot \|\widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0\|,
\end{aligned}$$

where the last equality is due to (7.49) in Lemma 2 and the fact that

$$\mathbb{E} \left( \sum_{i=1}^N \|\mathbf{X}'_i\|^2 \right) = \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^p \mathbb{E} \left( X_{it(k)}^2 \right) = O(NT) \tag{7.26}$$

and

$$\mathbb{E} \left( \sum_{i=1}^N \|\boldsymbol{\lambda}_i^0\|^2 \right) = \sum_{i=1}^N \sum_{\ell=1}^r \mathbb{E} \left[ (\lambda_{i\ell}^0)^2 \right] = O(N) \tag{7.27}$$

according to Assumption 2.

For  $\mathbf{J}_5^{(m)}$ ,

$$\begin{aligned}
\mathbf{J}_5^{(m)} &= \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \frac{1}{NT} \sum_{j=1}^N \sum_{k=1}^p \boldsymbol{\epsilon}_j \mathbf{x}'_{j(k)} \widehat{\mathbf{F}}^{(m)} (\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}) \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \right\| \\
&\leq \sum_{k=1}^p \frac{1}{N\sqrt{T}} \left( \sum_{i=1}^N \|\mathbf{X}'_i\|^2 \right)^{1/2} \cdot \left( \sum_{i=1}^N \|\boldsymbol{\lambda}_i^0\|^2 \right)^{1/2} \cdot \|\mathbf{M}_{\widehat{\mathbf{F}}^{(m)}}\| \\
&\quad \cdot \left\| \frac{1}{NT} \sum_{j=1}^N \boldsymbol{\epsilon}_j \mathbf{x}'_{j(k)} \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot |\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}| \\
&= O_P \left( \frac{1}{\sqrt{N}} \right) \cdot \sum_{k=1}^p |\beta_{0k} - \widehat{\beta}_{NT,k}^{(m)}| \leq O_P \left( \frac{1}{\sqrt{N}} \right) \cdot \|\widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0\|,
\end{aligned}$$

where we can obtain the last equality based on (7.49) in Lemma 2, (7.26) and (7.27).

For  $\mathbf{J}_6^{(m)}$ ,

$$\begin{aligned}
\left\| \mathbf{J}_6^{(m)} \right\| &= \left\| \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \mathbf{F}^0 \boldsymbol{\lambda}_j^0 \boldsymbol{\epsilon}_j \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \right\| \\
&= \left\| \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \left( \mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)} \mathbf{H}_{(m)}^{-1} \right) \boldsymbol{\lambda}_j^0 \boldsymbol{\epsilon}_j \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \right\| \\
&\leq \left( \frac{1}{NT} \sum_{i=1}^N \left\| \mathbf{X}'_i \right\|^2 \right)^{\frac{1}{2}} \cdot \left( \frac{1}{N} \sum_{i=1}^N \left\| \boldsymbol{\lambda}_i^0 \right\|^2 \right)^{\frac{1}{2}} \cdot \left\| \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \right\| \cdot \frac{1}{\sqrt{T}} \left\| \mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)} \mathbf{H}_{(m)}^{-1} \right\| \\
&\quad \cdot \left\| \frac{1}{N\sqrt{T}} \sum_{j=1}^N \boldsymbol{\lambda}_j^0 \boldsymbol{\epsilon}_j \right\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot \left\| \mathbf{G}^{(m)} \right\| \\
&= O_P \left( \frac{1}{\sqrt{N}} \right) \cdot \left[ O_P \left( \frac{1}{\sqrt{N}} \right) + O_P \left( \max(\gamma_{NT}^{\frac{1}{2}}, \frac{1}{\sqrt{N}}) \right) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \right. \\
&\quad \left. + O_P \left( \eta_{NT}^{\frac{1}{2}} \right) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2 \right] \\
&= O_P \left( \frac{1}{N} \right) + o_P(1) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|,
\end{aligned}$$

where the last second equality is valid because of Theorem 6, (7.49) in Lemma 2, (7.26) and (7.27).

For  $\mathbf{J}_7^{(m)}$ ,

$$\begin{aligned}
\mathbf{J}_7^{(m)} &= -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \boldsymbol{\epsilon}_j a_{ij}^{(m)} \\
&= -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}'_i \mathbf{M}_{\mathbf{F}^0} \boldsymbol{\epsilon}_j a_{ij}^{(m)} + \frac{1}{N^2 T^2} \sum_{i=1}^N \mathbf{X}'_i \left( \mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \right) \boldsymbol{\epsilon}_j a_{ij}^{(m)} \\
&= O_P \left( \frac{1}{\sqrt{NT}} \right) + o_P(1) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|,
\end{aligned}$$

where the last equality is the consequence of Lemma 3 and the fact that

$$\begin{aligned}
&\left| -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}'_i \mathbf{M}_{\mathbf{F}^0} \boldsymbol{\epsilon}_j a_{ij} \right| = \left| \frac{1}{N^2 T} \sum_{j=1}^N \sum_{i=1}^N \boldsymbol{\lambda}_i^{0'} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \boldsymbol{\lambda}_j^0 \mathbf{X}'_i \mathbf{M}_{\mathbf{F}^0} \boldsymbol{\epsilon}_j \right| \\
&= \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T \mathbf{g}_{jt} \boldsymbol{\epsilon}_{jt} = O_P \left( \frac{1}{\sqrt{NT}} \right) \tag{7.28}
\end{aligned}$$

with  $g_{jt} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\lambda}_i^{0'} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \boldsymbol{\lambda}_j^0 \mathbf{X}'_i \mathbf{M}_{\mathbf{F}^0} \mathbf{e}_t$  and  $a_{ij} = \boldsymbol{\lambda}_j^{0'} \mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 = T \boldsymbol{\lambda}_j^{0'} \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \boldsymbol{\lambda}_i^0$ .

For  $\mathbf{J}_8^{(m)}$ ,

$$\begin{aligned}
\mathbf{J}_8^{(m)} &= -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \\
&= -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 + \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \boldsymbol{\Omega}_j \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \\
&=: \mathbf{A}_1 + \mathbf{A}_2.
\end{aligned}$$

We now focus on the order of the first term  $\mathbf{A}_1$ . Rewrite it as

$$\mathbf{A}_1 = \mathbf{A}_{1,1} + \mathbf{A}_{1,2},$$

where

$$\begin{aligned}
\mathbf{A}_{1,1} &= -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}_i' \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0, \\
\mathbf{A}_{1,2} &= \frac{1}{N} \sum_{i=1}^N \left( \frac{\mathbf{X}_i' \widehat{\mathbf{F}}^{(m)}}{T} \right) \frac{1}{N T^2} \sum_{j=1}^N \widehat{\mathbf{F}}^{(m)'} \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \widehat{\mathbf{F}}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0.
\end{aligned}$$

For  $\mathbf{A}_{1,1}$ , we have

$$\begin{aligned}
\mathbf{A}_{1,1} &= -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}_i' \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \widehat{\mathbf{F}}^0 \mathbf{H}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \\
&\quad - \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{X}_i' \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \left( \widehat{\mathbf{F}}^{(m)} - \widehat{\mathbf{F}}^0 \right) \mathbf{H}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \\
&=: \mathbf{A}_{1,1}^{(1)} + \mathbf{A}_{1,1}^{(2)}.
\end{aligned}$$

It is easy to derive the order of  $\mathbf{A}_{1,1}^{(1)}$  from Lemma 4 noting that

$$\begin{aligned}
\mathbf{A}_{1,1}^{(1)} &= -\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathbf{X}_{it}' \left[ \boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{js}' - \mathbb{E}(\boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{js}') \right] \mathbf{F}_s^{0'} \mathbf{H}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \\
&= \frac{1}{T \sqrt{N}} \cdot \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{\sqrt{N}} \sum_{j=1}^N \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbf{X}_{it}' \left( \boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{js}' - \mathbb{E}(\boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{js}') \right) \mathbf{F}_s^{0'} \right] \mathbf{H}^{(m)} \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \\
&= O_P \left( \frac{1}{T \sqrt{N}} \right).
\end{aligned}$$

It follows from the Cauchy inequality that the term  $\mathbf{A}_{1,1}^{(2)}$  can be written as

$$\begin{aligned}
\|\mathbf{A}_{1,1}^{(2)}\| &= \left\| \frac{1}{\sqrt{NT}} \cdot \frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{s=1}^T \mathbf{a}_s \left( \widehat{\mathbf{F}}_s^{(m)} - \mathbf{F}_s^0 \mathbf{H}_{(m)} \right)' \mathbf{G}^{(m)} \boldsymbol{\lambda}_i \right\| \\
&\leq \frac{1}{\sqrt{NT}} \cdot \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{T} \sum_{s=1}^T \mathbf{a}_s \left( \widehat{\mathbf{F}}_s^{(m)} - \mathbf{F}_s^0 \mathbf{H}_{(m)} \right)' \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \right\| \\
&\leq \frac{1}{\sqrt{NT}} \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{T} \sum_{s=1}^T \|\mathbf{a}_s\|^2 \right)^{1/2} \cdot \left( \frac{1}{T} \sum_{s=1}^T \left\| \widehat{\mathbf{F}}_s^{(m)} - \mathbf{F}_s^0 \mathbf{H}_{(m)} \right\|^2 \right)^{1/2} \left\| \mathbf{G}^{(m)} \right\| \cdot \|\boldsymbol{\lambda}_i^0\| \\
&= \frac{1}{\sqrt{NT}} \left[ O_P \left( \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right) + O_P \left( \delta_{NT}^{-1} \right) \right].
\end{aligned}$$

In summary, the order of  $\mathbf{A}_{1,1}$  is

$$\|\mathbf{A}_{1,1}\| = O_P \left( \frac{1}{T\sqrt{N}} \right) + \frac{1}{\sqrt{NT}} \left[ O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \right) + O_P \left( \delta_{NT}^{-1} \right) \right]. \quad (7.29)$$

Next, we consider the term  $\mathbf{A}_{1,2}$ . Notice first that

$$\begin{aligned}
\|\mathbf{A}_{1,2}\| &\leq \frac{1}{N} \sum_{i=1}^N \left\| \frac{\mathbf{X}_i \widehat{\mathbf{F}}^{(m)}}{T} \right\| \cdot \left\| \mathbf{G}^{(m)} \boldsymbol{\lambda}_i^0 \right\| \cdot \left\| \frac{1}{NT^2} \sum_{j=1}^N \widehat{\mathbf{F}}^{(m)'} \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \widehat{\mathbf{F}}^{(m)} \right\| \\
&= O_P(1) \cdot \left\| \frac{1}{NT^2} \sum_{j=1}^N \widehat{\mathbf{F}}' \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \widehat{\mathbf{F}}^{(m)} \right\|.
\end{aligned}$$

Moreover, we have

$$\frac{1}{NT^2} \sum_{j=1}^N \widehat{\mathbf{F}}^{(m)'} \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \widehat{\mathbf{F}}^{(m)} =: \mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_3 + \mathbf{B}_4$$

where

$$\begin{aligned}
\mathbf{B}_1 &= \frac{1}{NT^2} \sum_{j=1}^N \mathbf{H}_{(m)} \mathbf{F}^{0'} \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \mathbf{F}^0 \mathbf{H}_{(m)} \\
\mathbf{B}_2 &= \frac{1}{NT^2} \sum_{j=1}^N \mathbf{H}_{(m)} \mathbf{F}^{0'} \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \left( \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right) \\
\mathbf{B}_3 &= \frac{1}{NT^2} \sum_{j=1}^N \left( \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right)' \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \mathbf{F}^0 \mathbf{H}_{(m)} \\
\mathbf{B}_4 &= \frac{1}{NT^2} \sum_{j=1}^N \left( \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right)' \left( \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' - \boldsymbol{\Omega}_j \right) \left( \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right).
\end{aligned}$$

Applying Lemma 4, we will have

$$\mathbf{B}_1 = \frac{1}{NT^2} \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathbf{H}_{(m)} \mathbf{F}_s^0 \mathbf{F}_t^{0'} \left[ \boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{js} - \mathbb{E} \left( \boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{js} \right) \right] \cdot \mathbf{H}_{(m)} = O_P \left( \frac{1}{T\sqrt{N}} \right).$$

For  $\mathbf{B}_2$ ,

$$\begin{aligned}
\|\mathbf{B}_2\| &= \left\| \frac{1}{\sqrt{NT}} \cdot \frac{1}{T} \sum_{s=1}^T \mathbf{H}_{(m)} \mathbf{C}_s \left( \widehat{\mathbf{F}}_s^{(m)} - \mathbf{H}'_{(m)} \mathbf{F}_s^0 \right) \right\| \\
&\leq \|\mathbf{H}_{(m)}\| \cdot \frac{1}{\sqrt{NT}} \left( \frac{1}{T} \sum_{s=1}^T \|\mathbf{C}_s\|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{s=1}^T \left\| \widehat{\mathbf{F}}_s - \mathbf{H}'_{(m)} \mathbf{F}_s^0 \right\|^2 \right)^{1/2} \\
&= \frac{1}{\sqrt{NT}} \left( O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \right) + O_P(\delta_{NT}^{-1}) \right).
\end{aligned}$$

Since  $\mathbf{B}_3$  is the transpose of  $\mathbf{B}_2$ , it has the same order as  $\mathbf{B}_2$ .

For  $\mathbf{B}_4$ ,

$$\begin{aligned}
\|\mathbf{B}_4\| &= \left\| \frac{1}{\sqrt{N}} \cdot \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left( \widehat{\mathbf{F}}_t^{(m)} - \mathbf{H}'_{(m)} \mathbf{F}_t^0 \right) \left( \widehat{\mathbf{F}}_s^{(m)} - \mathbf{H}'_{(m)} \mathbf{F}_s^0 \right)' \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N [\epsilon_{jt} \epsilon_{js} - \mathbb{E}(\epsilon_{jt} \epsilon_{js})] \right\| \\
&\leq \frac{1}{\sqrt{N}} \left( \frac{1}{T} \sum_{t=1}^T \left\| \widehat{\mathbf{F}}_t^{(m)} - \mathbf{H}'_{(m)} \mathbf{F}_t^0 \right\|^2 \right)^{1/2} \left( \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N (\epsilon_{jt} \epsilon_{js} - \mathbb{E}(\epsilon_{jt} \epsilon_{js})) \right)^2 \right)^{1/2} \\
&= \frac{1}{\sqrt{N}} O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2 \right) + \frac{1}{\sqrt{N}} O_P(\delta_{NT}^{-2}).
\end{aligned}$$

In summary, the order of  $\mathbf{A}_{1,2}$  is

$$\begin{aligned}
\|\mathbf{A}_{1,2}\| &= O_P \left( \frac{1}{T\sqrt{N}} \right) + \frac{1}{\sqrt{NT}} \left[ O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \right) + O_P(\delta_{NT}^{-1}) \right] \\
&\quad + \frac{1}{\sqrt{N}} \left[ O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2 \right) + O_P(\delta_{NT}^{-2}) \right].
\end{aligned} \tag{7.30}$$

Let us return to the term  $\mathbf{A}_2$ .

$$\|\mathbf{A}_2\| \leq \frac{r}{N^2 T} \sum_{i=1}^N \sum_{j=1}^N \left\| \frac{\mathbf{X}'_i}{\sqrt{T}} \right\| \cdot \|\boldsymbol{\Omega}_j\| \cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot \|\mathbf{G}^{(m)}\| \cdot \|\boldsymbol{\lambda}_i\| = O_P \left( \frac{1}{T} \right). \tag{7.31}$$

According to (7.29), (7.30) and (7.31), we can derive the order of  $\mathbf{J}_8^{(m)}$ .

□

*Proof of Corollary 1.* It follows from Theorem 2 that

$$\begin{aligned}
\left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} - \boldsymbol{\beta}_0 \right\| &\leq \left( \sum_{k=1}^p \left\| \widehat{w}_{1,k}^{(m)} \right\|^2 \right)^{1/2} \cdot \left( \sum_{k=1}^p \left| \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right|^2 \right)^{1/2} \\
&+ \left( \sum_{k=1}^p \sum_{\ell=1}^p \left\| \widehat{w}_{2,k\ell}^{(m)} \right\|^2 \right)^{1/2} \cdot \left( \sum_{k=1}^p \sum_{\ell=1}^p \left| \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right|^2 \cdot \left| \beta_{0,\ell} - \widehat{\beta}_{NT,\ell}^{(m)} \right|^2 \right)^{1/2} \\
&= \left( \sum_{k=1}^p \left\| \widehat{w}_{1,k}^{(m)} \right\|^2 \right)^{1/2} \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| + \left( \sum_{k=1}^p \sum_{\ell=1}^p \left\| \widehat{w}_{2,k\ell}^{(m)} \right\|^2 \right)^{1/2} \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2
\end{aligned} \tag{7.32}$$

This is actually sufficient for us to derive the orders of  $\widehat{w}_{1,k}^{(m)}$  and  $\widehat{w}_{2,k\ell}^{(m)}$ . For  $\widehat{w}_{1,k}^{(m)}$ ,

$$\begin{aligned}
\left\| \widehat{w}_{1,k}^{(m)} \right\| &\leq \frac{1}{N} \sum_{j=1}^N \left\| \frac{\mathbf{X}'_j}{\sqrt{T}} \right\| \cdot \left\| \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \right\| \cdot \left\| \frac{1}{NT} \sum_{i=1}^N \left( \mathbf{x}_{i(k)} \boldsymbol{\lambda}'_i \mathbf{F}^{0'} + \mathbf{F}^0 \boldsymbol{\lambda}_i \mathbf{x}'_{i(k)} \right) \right\| \\
&\cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot \left\| \left( \frac{\mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \right\| \cdot \left\| \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \right\| \cdot \left\| \boldsymbol{\lambda}_j^0 \right\| \\
&= O_P \left( \left\| \frac{1}{NT} \sum_{i=1}^N \left( \mathbf{x}_{i(k)} \boldsymbol{\lambda}'_i \mathbf{F}^{0'} + \mathbf{F}^0 \boldsymbol{\lambda}_i \mathbf{x}'_{i(k)} \right) \right\| \right) \\
&= O_P \left( \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \boldsymbol{\lambda}'_i \mathbf{F}^{0'} \right\| \right).
\end{aligned}$$

Then,

$$\sum_{k=1}^p \left\| \widehat{w}_{1,k}^{(m)} \right\|^2 = O_P \left( \sum_{k=1}^p \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \boldsymbol{\lambda}'_i \mathbf{F}^{0'} \right\|^2 \right) = O_P(\gamma_{NT}), \quad (7.33)$$

where the last equality holds due to the fact that

$$\mathbb{E} \left( \sum_{k=1}^p \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \boldsymbol{\lambda}'_i \mathbf{F}^{0'} \right\|^2 \right) = \mathbb{E} \left( \frac{1}{N^2 T^2} \sum_{i,j=1}^N \boldsymbol{\lambda}'_i \mathbf{F}^{0'} \mathbf{X}_i \mathbf{X}'_j \mathbf{F}^{0'} \boldsymbol{\lambda}_j \right) =: \gamma_{NT}.$$

For  $\widehat{w}_{2,k\ell}^{(m)}$ ,

$$\begin{aligned}
\left\| \widehat{w}_{2,k\ell}^{(m)} \right\| &\leq \frac{1}{N} \sum_{j=1}^N \left\| \frac{\mathbf{X}'_j}{\sqrt{T}} \right\| \cdot \left\| \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} \right\| \cdot \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \mathbf{x}'_{i(\ell)} \right\| \\
&\cdot \left\| \frac{\widehat{\mathbf{F}}^{(m)}}{\sqrt{T}} \right\| \cdot \left\| \left( \frac{\mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)}}{T} \right)^{-1} \right\| \cdot \left\| \left( \frac{\boldsymbol{\Lambda}^{0'} \boldsymbol{\Lambda}^0}{N} \right)^{-1} \right\| \cdot \left\| \boldsymbol{\lambda}_j^0 \right\| \\
&= O_P \left( \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \mathbf{x}'_{i(\ell)} \right\| \right).
\end{aligned}$$

Then,

$$\sum_{k=1}^p \sum_{\ell=1}^p \left\| \widehat{w}_{2,k\ell}^{(m)} \right\|^2 = O_P \left( \sum_{k=1}^p \sum_{\ell=1}^p \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \mathbf{x}'_{i(\ell)} \right\|^2 \right) = O_P(\eta_{NT}), \quad (7.34)$$

where the last equality depends on the fact that

$$\mathbb{E} \left( \sum_{k=1}^p \sum_{\ell=1}^p \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \mathbf{x}'_{i(\ell)} \right\|^2 \right) = \mathbb{E} \left( \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbf{X}'_{it} \mathbf{X}_{jt} \mathbf{X}'_{is} \mathbf{X}_{js} \right) =: \eta_{NT}.$$

This proof is completed with (7.32), (7.33) and (7.34).  $\square$



*Proof of Theorem 3.* It follows from Theorem 1 that when  $\gamma_{NT} = o_P(1)$  the second term on the right hand side of Corollary 1 has higher order than the third term. Thus, the recursive procedure entails

$$\left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} - \boldsymbol{\beta}_0 \right\| = O_P \left( \frac{1}{\sqrt{NT}} \right) + O_P \left( \max \left( \gamma_{NT}^{1/2}, \frac{1}{\sqrt{NT}} \right)^m \right) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 \right\|.$$

□

*Proof of Theorem 4.* As the initial OLS estimator  $\widehat{\boldsymbol{\beta}}_{NT}^{(0)}$  is a consistent estimator, Theorem 2 implies that the leading term on the right hand side of Theorem 2 is the second term, i.e.,

$$\widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} - \boldsymbol{\beta}_0 = (1 + o_P(1)) \left( \sum_{k=1}^p \widehat{w}_{1,k}^{(m)} \left( \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right) \right). \quad (7.35)$$

Applying the Cauchy-Schwartz inequality, we have

$$\begin{aligned} \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m+1)} - \boldsymbol{\beta}_0 \right\| &\leq \left( \sum_{k=1}^p \left[ \widehat{w}_{1,k}^{(m)} \right]^2 \right)^{1/2} \cdot \left( \sum_{k=1}^p \left( \beta_{0,k} - \widehat{\beta}_{NT,k}^{(m)} \right)^2 \right)^{1/2} \\ &= \left( \sum_{k=1}^p \left[ \widehat{w}_{1,k}^{(m)} \right]^2 \right)^{1/2} \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|. \end{aligned} \quad (7.36)$$

Then (4.9) can be obtained from

$$\left( \sum_{k=1}^p \left[ \widehat{w}_{1,k}^{(m)} \right]^2 \right)^{1/2} \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| < \delta. \quad (7.37)$$

□

*Proof of Theorem 5.* For the initial OLS estimator, we have

$$\widehat{\boldsymbol{\beta}}_{NT}^{(0)} - \boldsymbol{\beta}_0 = \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{xc}} + \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{x}\varepsilon}. \quad (7.38)$$

Under this model,

$$\begin{aligned} \mathbf{S}_{\mathbf{xc}} &= \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F}^0 \boldsymbol{\lambda}_i = \frac{1}{NT} \sum_{i,t=1}^{N,T} \boldsymbol{\gamma}'_i \mathbf{G}_t (\mathbf{G}_t + \mathbf{Q}_t)' \phi \boldsymbol{\gamma}_i \\ &= \frac{1}{NT} \sum_{i,t=1}^{N,T} \phi \boldsymbol{\gamma}'_i \mathbf{G}_t \mathbf{G}'_t \boldsymbol{\gamma}_i + \frac{1}{NT} \sum_{i,t=1}^{N,T} \phi \boldsymbol{\gamma}'_i \mathbf{G}_t \mathbf{Q}'_t \boldsymbol{\gamma}_i. \end{aligned} \quad (7.39)$$

According to the assumption that  $\boldsymbol{\gamma}_i, \mathbf{G}_t, \mathbf{Q}_t, \forall i = 1, 2, \dots, N; t = 1, 2, \dots, T$  are independent, we have

$$\begin{aligned} \mathbb{E} \left\| \frac{1}{NT} \sum_{i,t=1}^{N,T} \phi \boldsymbol{\gamma}'_i \mathbf{G}_t \mathbf{Q}'_t \boldsymbol{\gamma}_i \right\|^2 &= \frac{1}{N^2 T^2} \sum_{i_1, i_2=1}^N \sum_{t_1, t_2=1}^T \mathbb{E} \left( \boldsymbol{\gamma}'_{i_1} \mathbf{G}_{t_1} \mathbf{Q}'_{t_1} \boldsymbol{\gamma}_{i_1} \boldsymbol{\gamma}'_{i_2} \mathbf{G}_{t_2} \mathbf{Q}'_{t_2} \boldsymbol{\gamma}_{i_2} \right) \\ &= O_P \left( \frac{1}{NT} \right). \end{aligned} \quad (7.40)$$

Noticing (7.40), Theorem 1 and

$$\mathbf{S}_{\mathbf{xx}} = \frac{1}{NT} \sum_{i,t=1}^{N,T} \boldsymbol{\gamma}'_i \mathbf{G}_t \mathbf{G}'_t \boldsymbol{\gamma}_i, \quad (7.41)$$

we can obtain

$$\widehat{\boldsymbol{\beta}}_{NT}^{(0)} \xrightarrow{i.p.} \boldsymbol{\beta}_0 + \phi. \quad (7.42)$$

As a result,

$$\begin{aligned} y_{it} - \mathbf{x}'_{it} \widehat{\boldsymbol{\beta}}_{NT}^{(0)} &= \mathbf{x}'_{it} (\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_{NT}^{(0)}) + \boldsymbol{\lambda}'_i \mathbf{F}_t + \varepsilon_{it} \\ &= -\phi \mathbf{x}'_{it} + \boldsymbol{\lambda}'_i \mathbf{F}_t + \varepsilon_{it} + o_P(1) \\ &= \phi \boldsymbol{\gamma}'_i \mathbf{Q}_t + \varepsilon_{it} + o_P(1). \end{aligned} \quad (7.43)$$

It then follows from the principal components analysis that estimators for the factor structure satisfy

$$\left| \widehat{\boldsymbol{\lambda}}_i^{(1)'} \widehat{\mathbf{F}}_t^{(1)} - \boldsymbol{\gamma}'_i \mathbf{Q}_t \right| \xrightarrow{i.p.} 0. \quad (7.44)$$

Given the estimator  $(\widehat{\boldsymbol{\lambda}}_i^{(1)}, \widehat{\mathbf{F}}_t^{(1)})$ , we can derive the estimator  $\widehat{\boldsymbol{\beta}}^{(1)}$  applying the least-squares method to

$$y_{it} - \widehat{\boldsymbol{\lambda}}_i^{(1)'} \widehat{\mathbf{F}}_t^{(1)} = \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it} \quad (7.45)$$

It follows immediately from (7.44) and (7.45) that

$$\widehat{\boldsymbol{\beta}}_{NT}^{(1)} - \boldsymbol{\beta}_0 \xrightarrow{i.p.} \phi. \quad (7.46)$$

For this recursive procedure, we have that for any step  $m \geq 1$ ,

$$\widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \xrightarrow{i.p.} \phi. \quad (7.47)$$

□

*Proof of Corollary 2.* The proof is exactly the same as that of Corollary 1.

□

*Proof of Theorem 7.* The proof is totally the same as Theorem 3.

□

## 7.2 Some useful lemmas

**Lemma 1.** [Lemma A.1 of Gao (2007)] Suppose that  $\{M_m^n : -\infty < m \leq n < +\infty\}$  are the  $\sigma$ -fields generated by a stationary and  $\alpha$ -mixing process  $\{\xi_i\}_{-\infty}^{+\infty}$  with the mixing coefficient  $\alpha(i)$ . For some positive integers  $m$ , let  $\eta \in M_{s_i}^{t_i}$  where  $s_1 < t_1 < s_2 < t_2 < \dots < s_m < t_m$  and assume that  $t_i - s_i > \tau$  for all  $i$  and some  $\tau > 0$ . Assume further that, for some  $p_i > 1$ ,

$$\mathbb{E}|\eta_i|^{p_i} < +\infty,$$

for which

$$Q := \sum_{i=1}^{\ell} \frac{1}{p_i} < 1.$$

Then

$$\left| \mathbb{E} \left( \prod_{i=1}^{\ell} \eta_i \right) - \prod_{i=1}^{\ell} \mathbb{E}(\eta_i) \right| < 10(\ell - 1)\alpha(\tau)^{1-Q} \prod_{i=1}^{\ell} (\mathbb{E} |\eta_i|^{p_i})^{\frac{1}{p_i}}.$$

**Lemma 2.** Under Assumptions 2 and 3, we have the following estimations

1.

$$\left\| \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \right\| = O_P \left( \frac{1}{\sqrt{N}} \right); \quad (7.48)$$

2.

$$\left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \boldsymbol{\epsilon}_i' \right\| = O_P \left( \frac{1}{\sqrt{N}} \right); \quad (7.49)$$

3.

$$\left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \boldsymbol{\epsilon}_i' \right\| = O_P \left( \frac{1}{\sqrt{N}} \right). \quad (7.50)$$

*Proof of Lemma 1.* For (7.48), we calculate the second moment of this term as follows.

$$\begin{aligned} & \mathbb{E} \left\| \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \right\|^2 = \mathbb{E} \left[ \frac{1}{N^2 T^2} \sum_{i,j=1}^N \text{tr} \left( \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j' \right) \right] \\ &= \mathbb{E} \left[ \frac{1}{N^2 T^2} \sum_{i,j=1}^N \left( \boldsymbol{\epsilon}_i' \boldsymbol{\epsilon}_j \right)^2 \right] = \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} (\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{is} \boldsymbol{\epsilon}_{js}) \\ &= \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T [\mathbb{E} (\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{jt} \boldsymbol{\epsilon}_{is} \boldsymbol{\epsilon}_{js}) - \mathbb{E} (\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{jt}) \mathbb{E} (\boldsymbol{\epsilon}_{is} \boldsymbol{\epsilon}_{js})] + \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} (\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{jt}) \mathbb{E} (\boldsymbol{\epsilon}_{is} \boldsymbol{\epsilon}_{js}) \\ &\leq \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \alpha_{ij}^{\frac{\eta}{\eta+2}} (|t-s|) \left( \mathbb{E} (\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{jt})^{\eta+2} \right)^{\frac{1}{\eta+2}} \left( \mathbb{E} (\boldsymbol{\epsilon}_{is} \boldsymbol{\epsilon}_{js})^{\eta+2} \right)^{\frac{1}{\eta+2}} \\ &\quad + \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \alpha_{ij}^{\frac{\eta}{\eta+2}} (0) \left( \mathbb{E} (\boldsymbol{\epsilon}_{it}^{\eta+2}) \right)^{\frac{1}{\eta+2}} \left( \mathbb{E} (\boldsymbol{\epsilon}_{jt}^{\eta+2}) \right)^{\frac{1}{\eta+2}} |\mathbb{E} (\boldsymbol{\epsilon}_{is} \boldsymbol{\epsilon}_{js})| \\ &= O \left( \frac{1}{NT} \right) + O \left( \frac{1}{N} \right) = O \left( \frac{1}{N} \right), \end{aligned}$$

where the last inequality is valid due to Lemma 1 with Assumptions 2 and 3.

(7.49) can be obtained noting that

$$\begin{aligned}
& \mathbb{E} \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{x}_{i(k)} \boldsymbol{\epsilon}'_i \right\|^2 = \frac{1}{N^2 T^2} \sum_{i,j=1}^N \mathbb{E} \text{tr} \left( \mathbf{x}_{i(k)} \boldsymbol{\epsilon}'_i \boldsymbol{\epsilon}'_j \mathbf{x}'_{j(k)} \right) \\
&= \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \left[ \mathbb{E} (X_{is(k)} \epsilon_{it} X_{js(k)} \epsilon_{jt}) - \mathbb{E} (X_{is(k)} \epsilon_{it}) \mathbb{E} (X_{js(k)} \epsilon_{jt}) \right] \\
&\quad + \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} (X_{is(k)} \epsilon_{it}) \mathbb{E} (X_{js(k)} \epsilon_{jt}) \\
&\leq \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \alpha_{ij}^{\frac{\eta}{\eta+2}}(0) \left[ \mathbb{E} (X_{is(k)} \epsilon_{it})^{\eta+2} \right]^{\frac{1}{\eta+2}} \left[ \mathbb{E} (X_{js(k)} \epsilon_{jt})^{\eta+2} \right]^{\frac{1}{\eta+2}} \\
&\quad + \frac{1}{N^2 T^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbb{E} (X_{is(k)} \mathbb{E} (\epsilon_{it} | X_{is(k)})) \mathbb{E} (X_{js(k)} \mathbb{E} (\epsilon_{jt} | X_{js(k)})) = O \left( \frac{1}{N} \right),
\end{aligned}$$

where the last inequality holds based on Lemma 1 and Assumptions 2 and 3.

Finally, we have (7.50) noting that

$$\begin{aligned}
& \mathbb{E} \left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \boldsymbol{\epsilon}'_i \right\|^2 = \frac{1}{N^2 T^2} \sum_{i,j=1}^N \mathbb{E} \left[ \text{tr} \left( \mathbf{F}^0 \boldsymbol{\lambda}_i^0 \boldsymbol{\epsilon}'_i \boldsymbol{\epsilon}'_j \boldsymbol{\lambda}_j^0 \mathbf{F}^{0'} \right) \right] \\
&= \frac{1}{N^2 T} \sum_{i,j=1}^N \mathbb{E} \left( \boldsymbol{\epsilon}'_i \boldsymbol{\epsilon}_j \boldsymbol{\lambda}_j^0 \boldsymbol{\lambda}_i^0 \right) = \frac{1}{N^2 T} \sum_{\ell=1}^r \sum_{i,j=1}^N \sum_{t=1}^T \mathbb{E} (\epsilon_{it} \lambda_{i\ell}^0 \epsilon_{jt} \lambda_{j\ell}^0) \\
&\leq \frac{1}{N^2 T} \sum_{\ell=1}^r \sum_{i,j=1}^N \sum_{t=1}^T \left| \mathbb{E} (\epsilon_{it} \lambda_{i\ell}^0 \epsilon_{jt} \lambda_{j\ell}^0) - \mathbb{E} (\epsilon_{it} \lambda_{i\ell}^0) \mathbb{E} (\epsilon_{jt} \lambda_{j\ell}^0) \right| \\
&\quad + \frac{1}{N^2 T} \sum_{\ell=1}^r \sum_{i,j=1}^N \sum_{t=1}^T \left| \mathbb{E} (\epsilon_{it} \lambda_{i\ell}^0) \mathbb{E} (\epsilon_{jt} \lambda_{j\ell}^0) \right| \\
&\leq \frac{1}{N^2 T} \sum_{\ell=1}^r \sum_{i,j=1}^N \sum_{t=1}^T \alpha_{ij}^{\frac{\eta}{\eta+2}}(0) \left[ \mathbb{E} (\epsilon_{it} \lambda_{i\ell}^0)^{\eta+2} \right]^{\frac{1}{\eta+2}} \left[ \mathbb{E} (\epsilon_{jt} \lambda_{j\ell}^0)^{\eta+2} \right]^{\frac{1}{\eta+2}} \\
&\quad + \frac{1}{N^2 T} \sum_{\ell=1}^r \sum_{i,j=1}^N \sum_{t=1}^T \left| \mathbb{E} (\epsilon_{it} \mathbb{E} (\epsilon_{it} | \lambda_{i\ell}^0)) \mathbb{E} (\epsilon_{jt} \mathbb{E} (\epsilon_{jt} | \lambda_{j\ell}^0)) \right| \\
&= O \left( \frac{1}{N} \right),
\end{aligned}$$

where the last inequality is a result of Lemma 1 and Assumptions 2 and 3. □

**Lemma 3.** *Under Assumptions 1–3, we have the following results.*

1.

$$\begin{aligned}
\frac{1}{T} \mathbf{F}^{0'} \left( \widehat{\mathbf{F}}^{(m)} \mathbf{H}_{(m)}^{-1} - \mathbf{F}^0 \right) &= O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \right) \\
&\quad + O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2 \right) + O_P \left( \delta_{NT}^{-2} \right) \tag{7.51}
\end{aligned}$$

2.

$$\mathbf{H}_{(m)}\mathbf{H}'_{(m)} = O_P \left( \sum_{\ell=1}^4 \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^\ell \right) + O_P(\delta_{NT}^{-2}). \quad (7.52)$$

3.

$$\left\| \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} - \mathbf{M}_{\mathbf{F}^0} \right\| = O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \right) + O_P(\delta_{NT}^{-2}). \quad (7.53)$$

*Proof.* 1. From the expansion (7.21), we have

$$\left\| \frac{1}{T} \mathbf{F}^{0'} \left( \widehat{\mathbf{F}}^{(m)} \mathbf{H}_{(m)}^{-1} - \mathbf{F}^0 \right) \right\| = \sum_{h=1}^8 \frac{1}{T} \mathbf{F}^{0'} \mathbf{I}_h^{(m)} \mathbf{G}^{(m)} =: \sum_{h=1}^8 \mathbf{R}_h, \quad (7.54)$$

where  $\mathbf{G}^{(m)}$  and  $\mathbf{H}_{(m)}^{-1}$  are defined in (7.25) and (7.22) respectively.

Next, we examine the orders of the eight terms  $\mathbf{R}_h$ ,  $h = 1, 2, \dots, 8$  in details. Applying similar arguments in the proofs of Theorem 2, we obtain

$$\begin{aligned} \mathbf{R}_1 &= O_P(1) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2, \quad \mathbf{R}_2, \mathbf{R}_4 = O_P(1) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|, \\ \mathbf{R}_3, \mathbf{R}_5 &= o_P(1) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|, \quad \mathbf{R}_6, \mathbf{R}_7 = O_P \left( \frac{1}{\sqrt{NT}} \right). \end{aligned} \quad (7.55)$$

For the term  $\mathbf{R}_8$ , it follows from Theorem 2 that

$$\begin{aligned} \|\mathbf{R}_8\| &= \left\| \frac{1}{NT^2} \sum_{i=1}^N \mathbf{F}^{0'} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \widehat{\mathbf{F}}^{(m)} \right\| \\ &\leq \left\| \frac{1}{NT^2} \sum_{i=1}^N \mathbf{F}^{0'} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \left( \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right) \right\| + \left\| \frac{1}{NT^2} \sum_{i=1}^N \mathbf{F}^{0'} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \mathbf{F}^0 \mathbf{H}_{(m)} \right\| \\ &\leq \left\| \frac{\mathbf{F}^{0'}}{\sqrt{T}} \right\| \cdot \left\| \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \right\| \cdot \frac{1}{\sqrt{T}} \left\| \mathbf{F}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right\| + \left\| \frac{1}{NT^2} \sum_{i=1}^N \mathbf{F}^{0'} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \mathbf{F}^0 \right\| \cdot \|\mathbf{H}_{(m)}\| \\ &= O_P \left( \frac{1}{\sqrt{N}} \right) \cdot \left[ O_P \left( \frac{1}{\sqrt{N}} \right) + O_P \left( \gamma_{NT}^{\frac{1}{2}} \right) \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| + O_P \left( \eta_{NT}^{\frac{1}{2}} \right) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2 \right] \\ &\quad + O_P \left( \frac{1}{\sqrt{NT}} \right) \\ &= o_P(1) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| + O_P(1) \cdot \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2 + O_P(\delta_{NT}^2), \end{aligned} \quad (7.56)$$

where the last second equality uses the fact that

$$\begin{aligned}
& \mathbb{E} \left( \frac{1}{NT^2} \sum_{i=1}^N \mathbf{F}^{0'} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \mathbf{F}^0 \right) = \frac{1}{NT^2} \sum_{i=1}^N \sum_{t,s=1}^T \mathbb{E} \left( \mathbf{F}_t^{0'} \boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{is} \mathbf{F}_s^0 \right) \\
& \leq \frac{1}{NT^2} \sum_{i=1}^N \sum_{t,s=1}^T \left| \mathbb{E} \left( \mathbf{F}_t^{0'} \boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{is} \mathbf{F}_s^0 \right) - \mathbb{E} \left( \mathbf{F}_t^{0'} \boldsymbol{\epsilon}_{it} \right) \mathbb{E} \left( \mathbf{F}_s^0 \boldsymbol{\epsilon}_{is} \right) \right| \\
& \quad + \frac{1}{NT^2} \sum_{i=1}^N \sum_{t,s=1}^T \left| \mathbb{E} \left( \mathbf{F}_t^{0'} \boldsymbol{\epsilon}_{it} \right) \mathbb{E} \left( \mathbf{F}_s^0 \boldsymbol{\epsilon}_{is} \right) \right| \\
& \leq \frac{1}{NT^2} \sum_{i=1}^N \sum_{t,s=1}^T \alpha_i^{\frac{\eta}{\eta+2}} (|t-s|) \cdot \left\| \mathbb{E} \left( \mathbf{F}_t^0 \boldsymbol{\epsilon}_{it} \right)^{\eta+2} \right\|^{\frac{2}{\eta+2}} \cdot \left\| \mathbb{E} \left( \mathbf{F}_s^0 \boldsymbol{\epsilon}_{is} \right)^{\eta+2} \right\|^{\frac{2}{\eta+2}} \\
& = O_P \left( \frac{1}{T} \right)
\end{aligned}$$

whose last equality is ensured by Assumption 2, i.e.,  $\mathbb{E} \left( \mathbf{F}_t^0 \boldsymbol{\epsilon}_{it} \right) = \mathbb{E} \left( \mathbf{F}_t^0 \mathbb{E} \left( \boldsymbol{\epsilon}_{it} | \mathbf{F}_t^0 \right) \right) = \mathbf{0}_r$ . Then, (7.51) can be obtained straight from (7.54), (7.55) and (7.58).

2. Multiplying  $\mathbf{H}'_{(m)}$  and  $\mathbf{H}_{(m)}$  on the left and right sides of (7.51) respectively, we get

$$\begin{aligned}
& \frac{1}{T} \mathbf{H}'_{(m)} \mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)} - \mathbf{H}'_{(m)} \mathbf{H}_{(m)} \\
& = O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\| \right) + O_P \left( \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^2 \right) + O_P \left( \delta_{NT}^2 \right). \tag{7.57}
\end{aligned}$$

Moreover,

$$\begin{aligned}
& \left\| \mathbf{I}_r - \frac{1}{T} \mathbf{H}'_{(m)} \mathbf{F}^{0'} \widehat{\mathbf{F}}^{(m)} \right\| = \left\| \frac{1}{T} \widehat{\mathbf{F}}^{(m)'} \left( \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right) \right\| \\
& \leq \frac{1}{T} \left\| \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right\|^2 + \frac{1}{T} \left\| \mathbf{H}_{(m)} \right\| \cdot \left\| \mathbf{F}^{0'} \left( \widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0 \mathbf{H}_{(m)} \right) \right\| \\
& = O_P \left( \sum_{\ell=1}^4 \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^\ell \right) + O_P \left( \delta_{NT}^2 \right). \tag{7.58}
\end{aligned}$$

It follows from (7.57) and (7.58) that

$$\mathbf{I}_r - \mathbf{H}'_{(m)} \mathbf{H}_{(m)} = O_P \left( \sum_{\ell=1}^4 \left\| \widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0 \right\|^\ell \right) + O_P \left( \delta_{NT}^2 \right). \tag{7.59}$$

Hence (7.52) is obtained as desired.

3. Since

$$\left\| \mathbf{M}_{\widehat{\mathbf{F}}^{(m)}} - \mathbf{M}_{\mathbf{F}^0} \right\|^2 = \left\| \mathbf{P}_{\widehat{\mathbf{F}}^{(m)}} - \mathbf{P}_{\mathbf{F}^0} \right\|^2 = \text{tr} \left( \mathbf{P}_{\widehat{\mathbf{F}}^{(m)}} - \mathbf{P}_{\mathbf{F}^0} \right)^2 = 2 \text{tr} \left( \mathbf{I}_r - \frac{1}{T} \widehat{\mathbf{F}}^{(m)'} \mathbf{P}_{\mathbf{F}^0} \widehat{\mathbf{F}}^{(m)} \right),$$

it suffices to examine the term  $\mathbf{I}_r - \frac{1}{T} \widehat{\mathbf{F}}^{(m)'} \mathbf{P}_{\mathbf{F}^0} \widehat{\mathbf{F}}^{(m)}$ .

By pre-multiplying  $\mathbf{I}_r - \widehat{\mathbf{F}}^{(m)'} \mathbf{P}_{\mathbf{F}^0} \widehat{\mathbf{F}}^{(m)} / T$  with  $\widehat{\mathbf{F}}^{(m)'} \mathbf{F}^0 / T$ , and rewriting  $\widehat{\mathbf{F}}^{(m)'} \mathbf{F}^0 / T = \widehat{\mathbf{F}}^{(m)'} (\mathbf{F}^0 -$

$\widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})/T + \mathbf{H}_{(m)}$ , we can obtain what follows.

$$\begin{aligned}
& \left(\frac{\mathbf{F}^{0'}\mathbf{F}^0}{T}\right)^{-1} - \frac{\mathbf{F}^{0'}\widehat{\mathbf{F}}^{(m)}\widehat{\mathbf{F}}^{(m)'}\mathbf{F}^0}{T} \\
&= \left(\frac{\mathbf{F}^{0'}\mathbf{F}^0}{T}\right)^{-1} - \left[\frac{\widehat{\mathbf{F}}^{(m)}(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T} + \mathbf{H}_{(m)}\right]' \times \left[\frac{\widehat{\mathbf{F}}^{(m)}(\mathbf{F}^0 - \mathbf{H}_{(m)}\widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T} + \mathbf{H}_{(m)}\right] \\
&= \left(\frac{\mathbf{F}^{0'}\mathbf{F}^0}{T}\right)^{-1} - \left[\frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'\widehat{\mathbf{F}}^{(m)}\widehat{\mathbf{F}}^{(m)'}(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T^2}\right. \\
&\quad \left. + \frac{\mathbf{H}'_{(m)}\widehat{\mathbf{F}}^{(m)'}(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T} + \frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'\widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}}{T} + \mathbf{H}'_{(m)}\mathbf{H}_{(m)}\right] \\
&= \left(\frac{\mathbf{F}^{0'}\mathbf{F}^0}{T}\right)^{-1} - \left[\frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'(\widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0\mathbf{H}_{(m)})(\widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0\mathbf{H}_{(m)})'(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T^2}\right. \\
&\quad \left. + \frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'(\widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0\mathbf{H}_{(m)})\mathbf{H}'_{(m)}\mathbf{F}^{0'}(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T^2}\right. \\
&\quad \left. + \frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'\mathbf{F}^0\mathbf{H}_{(m)}(\widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0\mathbf{H}_{(m)})'(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T^2}\right. \\
&\quad \left. + \frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'\mathbf{F}^0\mathbf{H}_{(m)}\mathbf{H}'_{(m)}\mathbf{F}^{0'}(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T^2}\right. \\
&\quad \left. + \frac{\mathbf{H}'_{(m)}(\widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0\mathbf{H}_{(m)})'(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T} + \frac{\mathbf{H}'_{(m)}\mathbf{H}'_{(m)}\mathbf{F}^{0'}(\mathbf{F}^0\widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})}{T}\right. \\
&\quad \left. + \frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'(\widehat{\mathbf{F}}^{(m)} - \mathbf{F}^0\mathbf{H}_{(m)})\mathbf{H}_{(m)}}{T} + \frac{(\mathbf{F}^0 - \widehat{\mathbf{F}}^{(m)}\mathbf{H}_{(m)}^{-1})'\mathbf{F}^0\mathbf{H}_{(m)}\mathbf{H}_{(m)}}{T}\right. \\
&\quad \left. + \left(\frac{\mathbf{F}^{0'}\mathbf{F}^0}{T}\right)^{-1} + O_p(\|\widehat{\boldsymbol{\beta}}_{NT}^{(m)} - \boldsymbol{\beta}_0\|) + O_p(\delta_{NT}^{-2})\right] \\
&= O_p(\|\widehat{\boldsymbol{\beta}}_{NT} - \boldsymbol{\beta}_0\|) + O_p(\delta_{NT}^{-2}) + o_p(1)
\end{aligned}$$

where we used (7.51). □

**Lemma 4.** *Under Assumptions 1-3, we have the following results.*

$$\begin{aligned}
(1) \quad & \mathbb{E} \left\| \frac{1}{N^{1/2}} \sum_{j=1}^N \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbf{F}_s^0 \mathbf{F}_t^{0'} (\epsilon_{jt} \epsilon_{js} - \mathbb{E}(\epsilon_{jt} \epsilon_{js})) \right\|^2 \leq M \\
(2) \quad & \mathbb{E} \left\| \frac{1}{\sqrt{N}} \sum_{j=1}^N \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbf{X}_{it} [\epsilon_{jt} \epsilon_{js} - \mathbb{E}(\epsilon_{jt} \epsilon_{js})] F_{hs}^0 \right\|^2 \leq M, \tag{7.60}
\end{aligned}$$

for any  $i = 1, 2, \dots, N$  and  $h = 1, 2, \dots, r$ . Notice here that  $M$  is a constant which may vary from line to line.

*Proof.*

$$\begin{aligned}
& \mathbb{E} \left\| \frac{1}{N^{1/2}} \sum_{j=1}^N \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbf{F}_s \mathbf{F}'_t (\epsilon_{jt} \epsilon_{js} - \mathbb{E}(\epsilon_{jt} \epsilon_{js})) \right\|^2 \\
&= \mathbb{E} \frac{1}{NT^2} \sum_{j_1, j_2=1}^N \sum_{t_1, t_2=1}^T \sum_{s_1, s_2=1}^T \text{tr} \mathbf{F}_{s_1}^0 \mathbf{F}_{t_1}^{0'} \mathbf{F}_{t_2}^0 \mathbf{F}_{s_2}^{0'} [\epsilon_{j_1 t_1} \epsilon_{j_1 s_1} - \mathbb{E}(\epsilon_{j_1 t_1} \epsilon_{j_1 s_1})] \cdot [\epsilon_{j_2 t_2} \epsilon_{j_2 s_2} - \mathbb{E}(\epsilon_{j_2 t_2} \epsilon_{j_2 s_2})] \\
&= \frac{1}{NT^2} \sum_{j_1, j_2=1}^N \sum_{t_1, t_2=1}^T \sum_{s_1, s_2=1}^T \mathbb{E} \left( \mathbf{F}_{t_1}^{0'} \mathbf{F}_{t_2}^0 \mathbf{F}_{s_2}^{0'} \mathbf{F}_{s_1}^0 \right) \mathbb{E} \left( [\epsilon_{j_1 t_1} \epsilon_{j_1 s_1} - \mathbb{E}(\epsilon_{j_1 t_1} \epsilon_{j_1 s_1})] \cdot [\epsilon_{j_2 t_2} \epsilon_{j_2 s_2} - \mathbb{E}(\epsilon_{j_2 t_2} \epsilon_{j_2 s_2})] \right) \\
&\leq \frac{M}{NT^2} \sum_{j_1, j_2=1}^N \sum_{t_1, t_2=1}^T \sum_{s_1, s_2=1}^T \left| \mathbb{E} \left( [\epsilon_{j_1 t_1} \epsilon_{j_1 s_1} - \mathbb{E}(\epsilon_{j_1 t_1} \epsilon_{j_1 s_1})] \cdot [\epsilon_{j_2 t_2} \epsilon_{j_2 s_2} - \mathbb{E}(\epsilon_{j_2 t_2} \epsilon_{j_2 s_2})] \right) \right| \\
&\leq \frac{M}{NT^2} \sum_{j_1, j_2=1}^N \sum_{t_1, t_2=1}^T \sum_{s_1, s_2=1}^T \alpha_{j_1 j_2}^{\frac{\eta}{\eta+2}} (\min(|t_2 - t_1|, |t_2 - s_1|, |s_2 - t_1|, |s_2 - s_1|)) \\
&\quad \cdot \left[ \mathbb{E} |\epsilon_{j_1 t_1} \epsilon_{j_1 s_1} - \mathbb{E}(\epsilon_{j_1 t_1} \epsilon_{j_1 s_1})|^{\eta+2} \right]^{\frac{1}{\eta+2}} \cdot \left[ \mathbb{E} |\epsilon_{j_2 t_2} \epsilon_{j_2 s_2} - \mathbb{E}(\epsilon_{j_2 t_2} \epsilon_{j_2 s_2})|^{\eta+2} \right]^{\frac{1}{\eta+2}} \\
&= \frac{M}{NT^2} \cdot O_P(NT^2) \leq M,
\end{aligned}$$

where the last equality results from Assumption 3(2) and Assumption 2.

We can prove equation (2) similarly as what we did for equation (1).  $\square$

## References

- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.
- Bai, J. and Li, K. (2014). Theory and methods of panel data models with interactive effects. *The Annals of Statistics*, 42(1):142–170.
- Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29.
- Bai, J. and Wang, P. (2012). Identification and estimation of dynamic factor models.
- Bernanke, B. S., Boivin, J., and Eliasziw, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, 120(1):387–422.
- Chen, J., Gao, J., and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85.
- Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 14(1):C45–C90.
- Dominitz, J. and Sherman, R. P. (2005). Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory*, 21(04):838–863.
- Dong, C., Gao, J., and Peng, B. (2015). Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics*, 188(1):301–312.



- Fan, J. and Yao, Q. (2003). Nonlinear time series: Nonparametric and parametric methods.
- Frazier, D. T. (2013). Maximization by parts in semiparametric models.
- Gao, J. (2007). *Nonlinear time series: semi- and non-parametric methods*. Chapman & Hall/CRC Press.
- Geweke, J. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, 9(2):557–587.
- Harding, M. and Lamarche, C. (2011). Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. *Economics Letters*, 111(3):197–199.
- Huang, X. (2013). Nonparametric estimation in large panels with cross-sectional dependence. *Econometric Reviews*, 32(5-6):754–777.
- Kapetanios, G., Pesaran, M. H., and Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics*, 160(2):326–348.
- Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of econometrics*, 140(1):97–130.
- Kelejian, H. H. and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1):53–67.
- Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.
- Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.
- Pastorello, S., Patilea, V., and Renault, E. (2003). Iterative and recursive estimation in structural nonadaptive models. *Journal of Business & Economic Statistics*, 21(4):449–509.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.
- Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.
- Song, P. X.-K., Fan, Y., and Kalbfleisch, J. D. (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association*, 100(472):1145–1158.
- Su, L. and Jin, S. (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics*, 169(1):34–47.
- Su, L., Jin, S., and Zhang, Y. (2012). Specification test for panel data models with interactive fixed effects. *Journal of Econometrics*.