1-2019

On factor models with random missing: EM estimation, inference, and cross validation

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Ke MIAO
Singapore Management University, ke.miao.2015@phdecons.smu.edu.sg

Sainan JIN
Singapore Management University, snjin@smu.edu.sg

SMU ECONOMICS & STATISTICS

On Factor Models with Random Missing:

EM Estimation, Inference, and Cross Validation

Liangjun Su, Ke Miao, Sainan Jin

January 2019

Paper No. 04-2019

On Factor Models with Random Missing: EM Estimation, Inference, and Cross Validation *

Liangjun Su, Ke Miao, and Sainan Jin
School of Economics, Singapore Management University

January 15, 2019

Abstract

We consider the estimation and inference in approximate factor models with random missing values. We show that with the low rank structure of the common component, we can estimate the factors and factor loadings consistently with the missing values replaced by zeros. We establish the asymptotic distributions of the resulting estimators and those based on the EM algorithm. We also propose a cross-validation-based method to determine the number of factors in factor models with or without missing values and justify its consistency. Simulations demonstrate that our cross validation method is robust to fat tails in the error distribution and significantly outperforms some existing popular methods in terms of correct percentage in determining the number of factors. An application to the factor-augmented regression models shows that a proper treatment of the missing values can improve the out-of-sample forecast of some macroeconomic variables.

JEL Classification: C23, C33, C38; C55

Key Words: Cross-validation; Expectation-Maximization (EM) algorithm; Factor models; Matrix completion; Missing at random; Principal component analysis; Singular value decomposition

1 Introduction

Since the seminal work of Geweke (1977), Sargent and Sims (1977), Chamberlain and Rothschild (1983), factor models have been widely used in economics and finance. Some important theoretical contributions include Stock and Watson (1998), Forni et al. (2000), Bai and Ng (2002), Bai (2003),

*Su acknowledges the funding support provided by the Lee Kong Chian Fund for Excellence. Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

Hallin and Liška (2007), Onatski (2009, 2010, 2012), and Ahn and Horenstein (2013), among others. Nevertheless, all these authors assume a balanced panel in their asymptotic analyses.

Empirical data typically contain a variety of irregularities, including occasionally missing observations, unbalanced panel, and mixed frequency (e.g., monthly and quarterly) data. One simple way to handle missing data is to omit the cross-sectional units with missing values; see, e.g., Ludvigson and Ng (2007). But this will result in efficiency loss that can be substantial in some applications. To handle the missing data problem in factor models effectively, two methods have been proposed: the expectation-maximization (EM) algorithm and the Kalman filter (KF). These two methods have been widely used to handle missing data for principal component (PC) estimation with missing data and state space estimation with missing data. The details on how missing data are handled differ a lot in PC and state space applications. For the PC estimation with missing data, Stock and Watson (2002) propose an iterative method based on the EM algorithm that has proved to be easy and effective. Schumacher and Breitung (2008) apply Stock and Watson's methodology to nowcast German gross domestic product (GDP).

The state space framework has been adapted to missing data by either allowing the measurement equation to vary depending on what data are available at a given time point or keeping the dimension of the measurement equation to be the same over time by including a proxy value for the missing observation while adjusting the model parameters so that the Kalman filter places no weights on the missing observation. See Giannone et al. (2008), Mariano and Murasawa (2010), Doz et al. (2011), Jungbacker et al. (2011), Pinheiro et al. (2013), Bańbura and Modugno (2014), and Marcellino and Sivec (2016) for variations on this latter approach. In particular, Giannone et al. (2008) propose a two-step procedure that is able to solve the "ragged edge" problem in an approximate factor model when data are observed at different frequencies. They estimate the model by PC analysis with truncated balanced panel in the first step and update the estimates of factors by the KF with unbalanced panel data in the second step. Doz et al. (2011) show the consistency of the two-step estimators but do not have any asymptotic distributional results. Jungbacker et al. (2011) propose a new state space formulation of the factor model and apply the KF to estimate the underlying parameters with computational efficiency when the observations are missing at random. In view of the fact that it is not straightforward to apply Giannone et al.'s (2008) methodology to mixed frequency datasets with series of different lengths or, in general, to any pattern of missing data, Bańbura and Modugno (2014) propose a modified EM algorithm to allow for an arbitrary pattern of missing data where the KF is incorporated to estimate the factors in the maximization-step. A drawback of their approach is that for large cross-sections, the dimension of the augmented state vector becomes very large, which leads to computational inefficiency. Pinheiro et al. (2013) also propose an EM algorithm to estimate a dynamic factor model for panel data sets with jagged edge without significantly increasing the computation time relative to the balanced panel case. In addition, Foroni and Marcellino (2013) survey methods for handling mixed-frequency data, including dynamic factor models and alternative approaches; Stock and Watson (2016) summarize the advantage and disadvantage of the state space estimation for factor models with missing observations; Athey et al. (2018) develop new methods for estimating causal effects in panel data with missing values based on matrix completion methods.

Despite the popularity of the EM algorithm and the KF method in empirical researches, the asymptotic properties of the resulting estimators have been rarely studied. To the best of our knowledge, there is no formal study of the asymptotic properties for the EM estimators of the factors and factor loadings for the PC estimation with missing observations. For the KF estimators, Doz et al. (2011) prove the consistency but not the asymptotic normality.

In this paper we consider the EM estimation of approximate factor models with missing observations. For simplicity, we focus on the case where the missing occurs at random and remark in the end on the other forms of missing. As Stock and Watson (2016) remark, all the procedures in common use adopt the assumption that the data are missing at random, that is, whether a datum is missing is independent of the latent variables, and the missing-at-random assumption arguably is a reasonable assumption for the main sources of missing data in dynamic factor models in most macroeconomic applications to date. In the case of random missing, we draw support from the literature on matrix completion in computer science. It is well known that the low rank matrix such as the common component matrix in factor models can be recovered in the presence of missing observations when the noise matrix exhibit certain sparsity feature; see Cai et al. (2010), Candès and Plan (2010) and Candès and Li (2011). We show that similar phenomenon occurs when the noise matrix does not have any sparsity feature but lower order spectral norm than the common component matrix. In computation, we can simply replace the missing observations by zeros and conduct the usual PC analysis for a scaled version of the data matrix where the scale is determined by the percentage of observed values in the data. We show that the resulting estimators of factors, factor loadings, and common components are consistent but not asymptotically normal in general. Following the EM algorithm, we replace the missing observations by such initial estimators of the common components and obtain updated PC estimators. This procedure can be iterated until convergence. We show that the final estimators of the factors, factor loadings, and PCs are asymptotically more efficient than the initial estimators. We also characterize the efficiency loss for such EM estimators relative to the PC estimators without missing observations.

In some sense, the pure approximate factor model possesses the "self-fulfilling" property in that one does not need to observe all values in the data matrix in order to estimate the factors, factor loadings and common components and the missing values can be well recovered from the observed data. Such a self-fulfilling property motivates us to propose a novel method to determine the number of pervasive factors in approximate factor models no matter whether the original data contains missing observations or not. Our key insight is that we can draw each observation at random with probability p to construct the pseudo-data matrix with missing values. The original data are then divided into two sets, with one set containing the training observations used for the PC estimation for any prescribed number of factors (say, R) and the other set containing the held-out entries used for the out-of-sample evaluation. Then we can construct a cross-validation (CV) objective function that is indexed by R and choose R to minimize it. We show that this procedure consistently estimates the number of true factors. The finite sample performance of this procedure can be improved via iterations and some design for stability selection (e.g., Meinshausen and Bühlmann (2010)). Monte carlo simulations indicate that our new estimator of the number of factors significantly outperforms some existing popular estimators including those based on either information criterion (Bai and Ng (2002)), or eigenvalue distribution function (Onatski (2010)), or eigenvalue/growth ratio (Ahn and Horenstein (2013)). Moreover, our simulations also demonstrate that our new estimators are robust to fat tails in the error terms.

The paper is organized as follows. Section 2 introduces the EM estimators of factor models with random missing and their asymptotic properties. Section 3 proposes a novel method to determine the number of factors in approximate factor models. In Section 4, we report the Monte Carlo simulation results for our EM estimators of the factors, factor loadings and common components, and compare our method for the determination of the number of factors with the methods of Bai and Ng (2002), Onatski (2010), and Ahn and Horenstein (2013). In Section 5, we apply our method to an empirical application and show that it helps the out-of-sample forecasts based on factor-augmented regressions. Final remarks are contained in Section 6. The proofs of the results in Sections 2 and 3 are respectively relegated to Appendix A (in the main paper) and Appendix B (in the online supplement). The proofs of the technical lemmas in Appendix A along with some additional simulation results can be found in the additional online supplement that is available at http://www.mysmu.edu/faculty/ljsu/Publications/Factor_Missing19.pdf.

NOTATION. For an $m \times n$ real matrix A, we denote its transpose as A', its entrywise L_∞ norm as $\|A\|_\infty$ ($\equiv \max_{i,t} |A_{it}|$), its Frobenius norm as $\|A\|$ ($\equiv [\mathrm{tr}(AA')]^{1/2}$), its spectral norm as $\|A\|_{\mathrm{sp}}$ ($\equiv \sqrt{\mu_1 (A'A)}$) and its Moore-Penrose generalized inverse as A^+, where $\equiv$ means "is defined as" and $\mu_s(\cdot)$ denotes the sth largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. Note that the two norms are equal when A is a vector. We will frequently use the submultiplicative property of these norms and the fact that $\|A\|_{\rm sp} \leq \|A\| \leq \|A\|_{\rm sp} {\rm rank}(A)^{1/2}$. We also use $\mu_{\max}(B)$ and $\mu_{\min}(B)$ to denote the largest and smallest eigenvalues of a symmetric matrix B, respectively. We use $B > 0$ to denote that B is positive definite. Let $P_A \equiv A(A'A)^+ A'$ and $M_A \equiv I_m - P_A$, where I_m denotes an $m \times m$ identity matrix. The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, and plim probability limit. Let $\vee$ and $\wedge$ denote the max and min operators, respectively. E.g., $N \vee T = \max(N, T)$. Let $[N] = \{1, 2, ..., N\}$ and $[T] = \{1, 2, ..., T\}$. We use $(N, T) \to \infty$ to denote that N and T pass to infinity jointly. We let $\delta_{NT} = \sqrt{N} \wedge \sqrt{T}$.

2 Large Dimensional Factor Models with Random Missing

In this section, we consider the PCA estimation of large dimensional models with observations that are missing at random by assuming the true number of factors is known. We will propose a novel cross validation method to determine the number of factors in the next section.

2.1 EM Estimation

We consider the following factor model

$$X_{it} = \lambda_i' F_t + \varepsilon_{it}, \tag{2.1}$$

where $i = 1, ..., N$, $t = 1, \ldots, T$, F_t and λ_i are $R \times 1$ vectors of factors and factor loadings, respectively, and ε_{it} is the idiosyncratic error term. Following the lead of Stock and Watson (2002) and Bai et al. (2015), we study the estimation of the factors and factor loadings when some of the observations, X_{it}, are missing at random. Let $X = (X_1, ..., X_N)$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)$, where $X_i \equiv (X_{i1}, \ldots, X_{iT})'$ and $\varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ for $i = 1, ..., N$. We can write (2.1) in matrix form:

$$X = F\Lambda' + \varepsilon \tag{2.2}$$

where $F = (F_1, ..., F_T)'$ and $\Lambda = (\lambda_1, ..., \lambda_N)'$. We will use $F^0 = \left(F_1^0, ..., F_T^0\right)'$ and $\Lambda^0 = (\lambda_1^0, ..., \lambda_N^0)'$ to denote the true values of F and Λ, respectively. Let $\Omega \subset [N] \times [T]$ be the index set of the observations that are observed. That is,

$$\Omega = \{(i, t) \in [N] \times [T] : X_{it} \text{ is observed}\}.$$

Let G denote a $T \times N$ matrix with (t, i)th element given by $g_{it} = \mathbf{1}\{(i, t) \in \Omega\}$. Under the random missing mechanism, g_{it}'s are independently and identically distributed as Bernoulli(q) with $q \in (0, 1]$ and independent of X, F^0, Λ^0 and ε. So the population missing probability is given by $1 - q \in [0, 1)$. Let $|\Omega|$ denote the cardinality of the set Ω. It is easy to see that $\tilde{q} \equiv |\Omega| / (NT)$ is a $\sqrt{NT}$-consistent estimator of q.

2.1.1 The initial estimates

Let $\tilde{X} = X \circ G$ and $\tilde{X}_{it} = X_{it}g_{it}$, where $\circ$ denotes the Hadamard product. Our key observation is that the common component

$$C^0 \equiv F^0 \Lambda^{0\prime}$$

is a low rank matrix and ε is the noise component. In this case, it is possible to recover C^0 even when a large proportion of elements in the data matrix X are missing at random.

Let $E\left(\frac{1}{q}\tilde{X}|F^0, \Lambda^0\right)$ denote the $T \times N$ matrix with a typical element given by $E\left(\frac{1}{q}\tilde{X}_{it}|F_t^0, \lambda_i^0\right)$. Under the standard condition that $E\left(\varepsilon_{it}|F_t^0, \lambda_i^0\right) = 0$, we can readily verify that $E\left(\frac{1}{q}\tilde{X}|F^0, \Lambda^0\right) = F^0\Lambda^{0\prime}$. This motivates us to estimate F^0 and Λ^0 by minimizing the following least squares objective function

$$\mathcal{L}_{NT}^0 (F, \Lambda) \equiv \frac{1}{NT} \text{tr} \left[\left(\frac{1}{\tilde{q}} \tilde{X} - F\Lambda' \right) \left(\frac{1}{\tilde{q}} \tilde{X} - F\Lambda' \right)' \right] \tag{2.3}$$

under the identification restrictions: $F'F/T = I_R$ and $\Lambda'\Lambda$ is a diagonal matrix. By concentrating out Λ and using the normalization that $F'F/T = I_R$, the above minimization problem is identical to maximizing $\frac{1}{\tilde{q}^2}\text{tr}\left\{F'\tilde{X}\tilde{X}'F\right\}$. The estimated factor matrix, denoted by $\hat{F}^{(0)}$ is $\sqrt{T}$ times the eigenvectors corresponding to the R largest eigenvalues of the $T \times T$ matrix $\frac{1}{NT\tilde{q}^2}\tilde{X}\tilde{X}'$:

$$\frac{1}{NT\tilde{q}^2}\tilde{X}\tilde{X}'\hat{F}^{(0)} = \hat{F}^{(0)}\hat{D}^{(0)}, \tag{2.4}$$

where $\hat{D}^{(0)}$ is an $R \times R$ diagonal matrix consisting of the R largest eigenvalues of $\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'$, arranged in descending order along its diagonal line. Then the estimator of Λ' is given by

$$\hat{\Lambda}^{(0)\prime} = \frac{1}{\tilde{q}} \left(\hat{F}^{(0)\prime}\hat{F}^{(0)} \right)^{-1} \hat{F}^{(0)\prime}\tilde{X} = \frac{1}{T\tilde{q}}\hat{F}^{(0)\prime}\tilde{X}. \tag{2.5}$$

Let $\hat{F}_t^{(0)}$ denote the tth column of $\hat{F}^{(0)\prime}$ and $\hat{\lambda}_i^{(0)}$ the ith column of $\hat{\Lambda}^{(0)\prime}$. We can obtain an initial estimate of the (t, i)th element, C_{it}^0, of C^0 by $\hat{C}_{it}^{(0)} = \hat{\lambda}_i^{(0)\prime}\hat{F}_t^{(0)}$. We will show that the initial estimators $\hat{F}_t^{(0)}, \hat{\lambda}_i^{(0)}$ and $\hat{C}_{it}^{(0)}$ are consistent and follow mixture normal distributions under some standard conditions.

2.1.2 The iterated estimates

Despite the consistency of the initial estimators, they are not asymptotically efficient. To improve the efficiency, we consider iterative estimators. Let $\ell \geq 1$ be an integer. Suppose that we have obtained the estimates $\hat{F}_t^{(\ell-1)}, \hat{\lambda}_i^{(\ell-1)}$ and $\hat{C}_{it}^{(\ell-1)}$. In step ℓ, we can replace the missing values (X_{it}) in the matrix X with the estimated common components $\hat{C}_{it}^{(\ell-1)}$. Define the $T \times N$ matrix $\hat{X}^{(\ell)}$ with its (t,i)th element given by

$$\hat{X}_{it}^{(\ell)}=\begin{cases} X_{it} & \text{if } (i,t)\in\Omega \\ \hat{C}_{it}^{(\ell-1)} & \text{if } (i,t)\in\Omega_{\perp} \end{cases},\ \ell\geq 1,$$

where $\Omega_{\perp}=\{(i,t)\in[N]\times[T]:(i,t)\notin\Omega\}$. Then we can conduct the PC analysis based on $\hat{X}^{(\ell)}$ under the identification restrictions that $F'F/T=I_R$ and $\Lambda'\Lambda$ is a diagonal matrix. The estimated factor matrix, denoted by $\hat{F}^{(\ell)}$, is $\sqrt{T}$ time the eigenvectors corresponding to the R largest eigenvalues of the $T\times T$ matrix $\frac{1}{NT}\hat{X}^{(\ell)}\hat{X}^{(\ell)\prime}$:

$$\frac{1}{NT}\hat{X}^{(\ell)}\hat{X}^{(\ell)\prime}\hat{F}^{(\ell)}=\hat{F}^{(\ell)}\hat{D}^{(\ell)},$$

where $\hat{D}^{(\ell)}$ is a diagonal matrix consisting of the R largest eigenvalues of $\frac{1}{NT}\hat{X}^{(\ell)}\hat{X}^{(\ell)\prime}$ arranged in descending order along it diagonal line. Then the estimator of Λ' is given by

$$\hat{\Lambda}^{(\ell)\prime}=\left(\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}\right)^{-1}\hat{F}^{(\ell)\prime}\hat{X}^{(\ell)}=\frac{1}{T}\hat{F}^{(\ell)\prime}\hat{X}^{(\ell)}.$$

Let $\hat{F}_t^{(\ell)}$ denote the tth column of $\hat{F}^{(\ell)\prime}$ and $\hat{\lambda}_i^{(\ell)}$ the ith column of $\hat{\Lambda}^{(\ell)\prime}$. We obtain the updated estimate of C_{it}^0 by $\hat{C}_{it}^{(\ell)}=\hat{\lambda}_i^{(\ell)\prime}\hat{F}_t^{(\ell)}$. We will study the asymptotic properties of $\hat{F}_t^{(\ell)},\hat{\lambda}_i^{(\ell)}$ and $\hat{C}_{it}^{(\ell)}$, $\ell=1,2,...$, below.

Remark 1 (Connection with Stock and Watson's (2002) EM estimation) Stock and Watson (2002, SW hereafter) propose an EM algorithm to conduct the PC analysis for panel data with missing values. The least squares objective function they consider is given by

$$V(F,\Lambda)=\frac{1}{NT}\mathrm{tr}\left[\left[\left(X-F\Lambda'\right)\circ G\right]\left[\left(X-F\Lambda'\right)\circ G\right]'\right]=\frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{N}\left(X_{it}-\lambda_i'F_t\right)^2 g_{it}.$$

Minimization of $V(F,\Lambda)$ requires iterative methods. SW (2002) motivate the EM algorithm by assuming that ε_{it}'s are independently and identically distributed (i.i.d.) according to $N\left(0,\sigma^2\right)$. They suggest various ways to obtain the initial estimates. For example, when the full dataset contains a subset constituting a balanced panel, they suggest using estimates of the factors from the balanced subset as the starting value $\hat{F}_t^{(0)}$. Given the estimates $\hat{C}_{it}^{(\ell-1)}$ at stage $\ell-1$, our construction of the expectation object $\hat{X}_{it}^{(\ell)}$ is the same as SW's (2002) and so is our ℓth stage estimator. But SW (2002) do not provide any theoretical justification for their EM estimates. With our well-chosen initial estimators, we are able to formally justify the use of EM estimator.

2.2 Asymptotic properties of the initial estimators $\hat{F}_t^{(0)},\hat{\lambda}_i^{(0)}$ and $\hat{C}_{it}^{(0)}$

Let M denote a generic finite positive constant that may vary across lines. We make the following assumptions.

Assumption A.1 (i) $\max_t E\left\|F_t^0\right\|^{4/\gamma_1} \leq M$ for some $\gamma_1 \in (0,1)$ and $T^{-1}F^{0\prime}F^0 \xrightarrow{P} \Sigma_{F^0} > 0$ for some $R \times R$ matrix Σ_{F^0} as $T \to \infty$.

(ii) $\max_i E\left\|\lambda_i^0\right\|^{4/\gamma_2} \leq M$ for some $\gamma_2 \in (0,1)$ and $N^{-1}\Lambda^{0\prime}\Lambda^0 \xrightarrow{P} \Sigma_{\Lambda^0} > 0$ for some $R \times R$ matrix Σ_{Λ^0} as $N \to \infty$.

(iii) $\max_{i,t} E[\left(\lambda_i^{0\prime}F_t^0\right)^4] \leq M$.

(iv) The eigenvalues of $\Sigma_{\Lambda^0}\Sigma_{F^0}$ are distinct from each other.

(v) $N^{-1}\Lambda^{0\prime}\Lambda^0 - \Sigma_{\Lambda_0} = O_P(N^{-1/2})$ and $T^{-1}F^{0\prime}F^0 - \Sigma_{F_0} = O_P(T^{-1/2})$.

Assumption A.2 (i) $E\left(\varepsilon_{it}|\lambda_i^0, F_t^0\right) = 0$, $E\left(\varepsilon_{it}^4\right) \leq M$, and $\|\varepsilon\|_{\text{sp}} = O_P(\max(\sqrt{N}, \sqrt{T}))$.

(ii) $\max_s \sum_{t=1}^T |\gamma_N(s,t)| \leq M$, where $\gamma_N(s,t) = N^{-1}\sum_{i=1}^N |E(\varepsilon_{is}\varepsilon_{it})|$.

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

Assumption A.1 parallels Assumptions A-B in Bai (2003) and Assumption A.2 is analogous to Assumption C in Bai (2003). The major difference is that we require both the factors and factor loadings have finite moments higher than the usual fourth order. Bai and Ng (2002) and Bai (2003) assume finite fourth moments for F_t^0 but require that λ_i^0 be uniformly bounded. Assumption A.1(v) imposes the standard convergence rates for $N^{-1}\Lambda^{0\prime}\Lambda^0$ and $T^{-1}F^{0\prime}F^0$. It implies that $\mu_r(\frac{1}{NT}F^0\Lambda^{0\prime}\Lambda^0F^{0\prime}) - \sigma_r^2 = O_P\left(\delta_{NT}^{-1}\right)$ for $r = 1, ..., R$, where $\sigma_r^2 = \mu_r\left(\Sigma_{\Lambda^0}\Sigma_{F^0}\right)$. Assumption A.2(i) is also assumed in Su and Chen (2013), Lu and Su (2016), and Moon and Weidner (2017). In particular, Moon and Weidner (2017) demonstrate that this condition can be satisfied for various error processes.

The following theorem establishes the mean squared convergence of $\hat{F}_t^{(0)}$. Define

$$\hat{H}^{(0)} = \left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)T^{-1}F^{0\prime}\hat{F}^{(0)}(\hat{D}^{(0)})^{-1},$$

where $\hat{D}^{(0)}$ is asymptotically nonsingular by Lemma A.1.

Theorem 2.1 *Suppose Assumptions A.1 and A.2 hold. Then* $\frac{1}{T}\left\|\hat{F}^{(0)} - F^0\hat{H}^{(0)}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ *where* $\delta_{NT} = \sqrt{N} \wedge \sqrt{T}$.

Theorem 2.1 reports the mean square (MS) convergence rate of $\hat{F}_t^{(0)}$. It implies that we can estimate the space spanned by the columns of F^0 consistently.

To proceed, we assume the following limiting objects exist and are finite:

$$\begin{aligned}
\Gamma_{1g,t}(q) &= \lim_{N\to\infty}\text{Var}\left(\frac{1}{\sqrt{N}q}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}\right),\ \Gamma_{2g,t}(q) = \text{plim}_{N\to\infty}\frac{1-q}{q}\sum_{i=1}^N \lambda_i^0\lambda_i^{0\prime}\left(\lambda_i^{0\prime}F_t^0\right)^2, \\
\Phi_{1g,i}(q) &= \lim_{T\to\infty}\text{Var}\left(\frac{1}{\sqrt{T}q}\sum_{t=1}^T F_t^0\varepsilon_{it}g_{it}\right),\ \Phi_{2g,i}(q) = \text{plim}_{N\to\infty}\frac{1-q}{q}\sum_{t=1}^T F_t^0F_t^{0\prime}\left(\lambda_i^{0\prime}F_t^0\right)^2.
\end{aligned}$$

Let

$$\Gamma_{g,t}(q) = \Gamma_{1g,t}(q) + \Gamma_{2g,t}(q) \text{ and } \Phi_{g,i}(q) = \Phi_{1g,i}(q) + \Phi_{2g,i}(q).$$

Note that $\Gamma_{2g,t}$ and $\Phi_{2g,i}$ and therefore $\Gamma_{g,t}$ and $\Phi_{g,i}$ are generally random objects under our assumptions that allow for random factors and random factor loadings. To study the asymptotic distributions of $\hat{F}_t^{(0)}, \hat{\lambda}_i^{(0)}$ and $\hat{C}_{it}^{(0)}$, we add the following assumptions.

Assumption A.3 (i) Either $\max_{t,s} E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\chi_{i,st}\right\|^4 \leq M$ or $E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\chi_{i,st}\right\|^2 \leq M$, where $\chi_{i,st} = \varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})$.

(ii) $E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\varepsilon_{is}\right\|^2 \leq M$.

(iii) Let $\sigma_{ij,ts} = E(\varepsilon_{it}\varepsilon_{js})$. $\max_t N^{-1}\sum_{i=1}^{N}\sigma_{ii,tt} \leq M$, $\max_{1\leq t\leq T} N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sigma_{ij,tt}| \leq M$, $\max_{1\leq i\leq N} T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|\sigma_{ii,ts}| \leq M$, and $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\sigma_{ij,ts}| \leq M$.

Assumption A.4 (i) $\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it} \xrightarrow{d} N(0, \Gamma_{1g,t})$,

(ii) $\frac{1}{\sqrt{Tq}}\sum_{t=1}^{T}F_t^0\varepsilon_{it}g_{it} \xrightarrow{d} N(0, \Phi_{1g,i})$.

The first part of Assumption A.3(i) strengthens Assumption A.2(iii) and is also assumed in Bai and Ng (2002) and Bai (2003). The latter authors also assume that the second part of A.3(i) holds simultaneously with the first part, which we do not need. In the special case where $E(F_s^0\chi_{i,st}) \neq 0$ for enough (s,t) pairs (e.g., when $E(F_s^0) = 0$ but $E(F_s^0\varepsilon_{it}\varepsilon_{is}) \neq 0$ for all $s > t$), the second part of A.3(i) is not satisfied.

Let $\mathcal{G}_{Ni}^t = \sigma(\{g_{jt}, j \leq i\}, \Lambda^0, F_t^0)$, the minimal sigma-field generated from $\{g_{jt},\ j \leq i\}$ and (Λ^0, F_t^0). Let $\mathcal{G}^t = \sigma\left(\cup_{N=1}^{\infty}\mathcal{G}_{NN}^t\right)$. Analogously, let $\mathcal{G}_{Tt}^i = \sigma\left(\{g_{is}, s \leq t\}, \lambda_i^0, F^0\right)$, $\mathcal{G}^i = \sigma\left(\cup_{T=1}^{\infty}\mathcal{G}_{TT}^i\right)$, and $\mathcal{G}^{it} = \sigma\left(\mathcal{G}^i \cup \mathcal{G}^t\right)$.

The following theorem presents the asymptotic distributions of $\hat{F}_t^{(0)}, \hat{\lambda}_i^{(0)}$ and $\hat{C}_{it}^{(0)}$ based on the notation of stable convergence.

Theorem 2.2 *Suppose Assumptions A.1-A.4 hold. Suppose that* $(T^{1/2} + N^{1/2})\delta_{NT}^{-2} = o(1)$. *Let* $\hat{\Pi}_{tN}^{(0)} = \sqrt{N}(\hat{F}_t^{(0)} - \hat{H}^{(0)\prime}F_t^0)$ *and* $\hat{\Pi}_{iT}^{(0)} = \sqrt{T}(\hat{\lambda}_i^{(0)} - (\hat{H}^{(0)})^{-1}\lambda_i^0)$. *Then as* $(N,T) \to \infty$

(i) $\hat{\Pi}_{tN}^{(0)} = (\hat{D}^{(0)})^{-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\xi_{it} + O_P\left(N^{1/2}\delta_{NT}^{-2}\right) \to N\left(0, D^{-1}Q\Gamma_{g,t}(q)Q'D^{-1}\right)$ $\mathcal{G}^t$*-stably,*

(ii) $\hat{\Pi}_{iT}^{(0)} = \hat{H}^{(0)\prime}\frac{1}{\sqrt{Tq}}\sum_{t=1}^{T}F_t^0\xi_{it} + O_P\left(T^{1/2}\delta_{NT}^{-2}\right) \to N\left(0, (Q')^{-1}\Phi_{g,i}(q)Q^{-1}\right)$ $\mathcal{G}^i$*-stably,*

(iii) $\left(\frac{1}{N}\Sigma_{F,it}^{(0)}(q) + \frac{1}{T}\Sigma_{\Lambda,it}^{(0)}(q)\right)^{-1/2}\left(\hat{C}_{it}^{(0)} - C_{it}^0\right) \xrightarrow{d} N(0,1)$,

where $\xi_{it} = \varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0(g_{it} - q)$, $\Sigma_{F,it}^{(0)}(q) = \lambda_i^{0\prime}\Sigma_{\Lambda^0}^{-1}\Gamma_{g,t}(q)\Sigma_{\Lambda^0}^{-1}\lambda_i^0$ *and* $\Sigma_{\Lambda,it}^{(0)}(q) = F_t^{0\prime}\Sigma_{F^0}^{-1}\Phi_{g,i}(q)\Sigma_{F^0}^{-1}F_t^0$ *signify the contributions of the factor and factor loading estimators to the asymptotic variance of* $\hat{C}_{it}^{(0)}$, *respectively, and D denotes the diagonal matrix consisting of the eigenvalues of $\Sigma_{\Lambda^0}^{1/2}\Sigma_{F^0}\Sigma_{\Lambda^0}^{1/2}$ in descending order with the corresponding eigenvector matrix denoted as Υ such that $\Upsilon'\Upsilon = I_R$ and $Q = D^{1/2}\Upsilon'\Sigma_{\Lambda^0}^{-1/2}$.*

Theorem 2.2 parallels Theorems 1-3 in Bai (2003). Bai (2003) obtains the asymptotic normal distributions for his estimators of factors and factor loadings. In contrast, we show that the sequence$\left\{\hat{\Pi}_{tN}^{(0)},\ N \geq 1\right\}$ converges $\mathcal{G}^t$-stably as $(N,T) \to \infty$ to a mixture normal whose asymptotic variance is random but measurable with respect to certain limit sigma-field, and similarly, the sequence $\left\{\hat{\Pi}_{iT}^{(0)},\ T \geq 1\right\}$ converges $\mathcal{G}^i$-stably as $(N,T) \to \infty$ to a mixture normal whose asymptotic variance is random but measurable with respect to certain limit sigma-field. We refer the reader directly to the Häusler and Luschgy (2015) for stable convergence in general and the stable martingale central limit theorem in particular. To understand the limiting distribution of $\hat{\Pi}_{tN}^{(0)}$ in Theorem 2.2(i), we notice that its influence function depends on ξ_{it} through two terms, $\varepsilon_{it}g_{it}$ and $\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)$. The first term also appears in the influence function for the factor estimators in the absence of random missing at time t (i.e., $g_{it} = 1\ \forall i$) while the second term is introduced by the random missing mechanism. Due to the presence of common factor F_t^0 for all cross-sectional units, $\frac{1}{\sqrt{Nq}}\sum_{i=1}^N \lambda_i^0\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)$ does not have a limiting normal distribution. Instead, it converges to $N\left(0,\Gamma_{2g,t}\right)$ $\mathcal{G}^t$-stably as $N \to \infty$, where $N\left(0,\Gamma_{2g,t}\right)$ can be regarded as a normal random vector with random variance given by $\Gamma_{2g,t}$. In the special case where F_t^0 is nonrandom, the limiting distribution reduces to the usual normal distribution. Similar remarks for $\hat{\Pi}_{iT}^{(0)}$ in Theorem 2.2(ii). Theorem 2.2(iii) only reports the limiting distribution for the normalized common component estimator. One can also follow the analyses of parts (i)-(ii) in the theorem and report the stable limiting distribution of $\delta_{NT}(\hat{C}_{it}^{(0)} - C_{it}^0)$ as $(N,T) \to \infty$.

By Corollary 6.3 in Häusler and Luschgy (2015) and the Cramér-Wold device, we can show that

$$\left[(D^{-1}Q\Gamma_{g,t}Q'D^{-1}\right]^{-1/2}\hat{\Pi}_{tN}^{(0)} \xrightarrow{d} N\left(0,I_R\right) \text{ as } (N,T) \to \infty, \text{ and}$$
$$\left[(Q')^{-1}\Phi_{g,i}Q^{-1}\right]^{-1/2}\hat{\Pi}_{iT}^{(0)} \xrightarrow{d} N\left(0,I_R\right) \text{ as } (N,T) \to \infty.$$

With these results and the result in Theorem 2.2(iii), we could make inference on the factors, factor loadings, and common component. But because these estimates are not the final estimates, we will study the asymptotic properties of the iterated estimators of these objects later on.

2.3 Asymptotic properties of the iterated estimators of the factors and factor loadings

Let $\hat{H}^{(\ell)} = \left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)T^{-1}F^{0\prime}\hat{F}^{(\ell)}(\hat{D}^{(\ell)})^{-1}$. To study the asymptotic properties of $\hat{F}_t^{(\ell)}, \hat{\lambda}_i^{(\ell)}$ and $\hat{C}_{it}^{(\ell)}$, we add the following assumption.

Assumption A.5 (i) $\max_t \left\| \frac{1}{N} \sum_{i=1}^{N} \zeta_{1,it} \right\| = O_P\left((N/\ln N)^{-1/2}\right)$ and $\max_{t,s} \left\| \frac{1}{N} \sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it} g_{is} \right\| = O_P\left((N/\ln N)^{-1/2}\right)$, where $\zeta_{1,it} = \lambda_i^0 \varepsilon_{it} g_{it}$ and $\lambda_i^0 \lambda_i^{0\prime} F_t^0 (g_{it} - q)$.

(ii) $\max_i \left\| \frac{1}{T} \sum_{t=1}^{T} \zeta_{2,it} \right\| = O_P\left((T/\ln T)^{-1/2}\right)$, where $\zeta_{2,it} = F_t^0 \varepsilon_{it} g_{it}$ and $F_t^0 \lambda_i^{0\prime} F_t^0 (g_{it} - q)$.

(iii) $\max_t \left\| \frac{1}{NT} \sum_{i,s} \zeta_{3,its} \right\| = O_P\left(\delta_{NT}^{-2} \ln N\right)$ and $\max_t \left\| \frac{1}{NT} \sum_i \sum_{s=1, s\neq t}^{T} F_s^0 F_s^{0\prime} \lambda_i^0 \lambda_i^{0\prime} (g_{is} - q)(g_{it} - q) \right\| = O_P\left(\delta_{NT}^{-2} \ln N\right)$, where $\zeta_{3,its} = F_s^0 \left[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})\right] g_{it} g_{is}$, $F_s^0 F_s^{0\prime} \lambda_i^0 \varepsilon_{it} g_{it} (g_{is} - q)$ and $\lambda_i^0 F_s^{0\prime} \varepsilon_{is} g_{is} (g_{it} - q)$.

Assumption A.5 imposes some uniform convergence conditions that are similar to those imposed in Su et al. (2015) and Su and Wang (2017). Following these authors, one can verify Assumption A.5 under some primitive conditions on $\left\{\lambda_i^0, F_t^0, \varepsilon_{it}\right\}$.

The following theorem establishes the mean squared convergence of $\hat{F}_t^{(\ell)}$.

Theorem 2.3 *Suppose Assumptions A.1-A.5 hold. Then* $\frac{1}{T} \left\| \hat{F}^{(\ell)} - F^0 \hat{H}^{(\ell)} \right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ *for each* ℓ.

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

Theorem 2.4 *Suppose Assumptions A.1-A.5 hold. Suppose that* $\sqrt{N}(T^{\gamma_1/4}\delta_{NT}^{-2} \ln T + T^{-1+3\gamma_1/4}) = o(1)$ *and* $\sqrt{T}(N^{\gamma_2/4}\delta_{NT}^{-2} \ln N + N^{-1+3\gamma_2/4}) = o(1)$. *Let* $\hat{\Pi}_{tN}^{(\ell)} = \sqrt{N}(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime} F_t^0)$ *and* $\hat{\Pi}_{iT}^{(\ell)} = \sqrt{T}(\hat{\lambda}_i^{(\ell)} - \hat{H}^{(\ell)-1}\lambda_i^0)$. *Then*

(i) $\hat{\Pi}_{tN}^{(\ell)} = D^{-1} Q \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it} + (1-q)\, \hat{\Pi}_{tN}^{(\ell-1)} + o_P(1)$ *uniformly in* t *and* $\hat{\Pi}_{tN}^{(\ell)} \xrightarrow{d} N(0, D^{-1} Q \times \Gamma_{1g,t}(q)\, Q' D^{-1})$ *as* $(\ell, N, T) \to \infty$,

(ii) $\hat{\Pi}_{iT}^{(\ell)} = (Q')^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t^0 \varepsilon_{it} g_{it} + (1-q)\, \hat{\Pi}_{iT}^{(\ell-1)} + o_P(1)$ *uniformly in* i *and* $\hat{\Pi}_{iT} \xrightarrow{d} N(0, (Q')^{-1} \times \Phi_{1g,i}(q)\, Q^{-1})$ *as* $(\ell, N, T) \to \infty$,

(iii) $(\frac{1}{N}\Sigma_{1F,it} + \frac{1}{T}\Sigma_{1\Lambda,it})^{-1/2}(\hat{C}_{it}^{(\ell)} - C_{it}^0) \xrightarrow{d} N(0,1)$ *as* $(\ell, N, T) \to \infty$,

where $\Gamma_{1g,t}, \Phi_{1g,i}$, D *and* Q *are as defined in the last subsection, and* $\Sigma_{1F,it} = \lambda_i^{0\prime} \Sigma_{\Lambda^0}^{-1} \Gamma_{1g,t}(q)\, \Sigma_{\Lambda^0}^{-1} \lambda_i^0$, *and* $\Sigma_{1\Lambda,it} = F_t^{0\prime} \Sigma_{F^0}^{-1} \Phi_{1g,i}(q)\, \Sigma_{F^0}^{-1} F_t^0$ *signify the contribution of the factor and factor loading estimators to the asymptotic variance of* $\hat{C}_{it}^{(\ell)}$ *for large* ℓ*, respectively.*

Remark 2 Noting that $\Gamma_{g,t}(q) = \Gamma_{1g,t}(q) + \Gamma_{2g,t}(q)$ and $\Phi_{g,i}(q) = \Phi_{1g,i}(q) + \Phi_{2g,i}(q)$, a comparison of Theorem 2.4 with Theorem 2.2 indicates that $\hat{F}_t^{(\ell)}$, $\hat{\lambda}_i^{(\ell)}$ and $\hat{C}_{it}^{(\ell)}$ are asymptotically more efficient than $\hat{F}_t^{(0)}$, $\hat{\lambda}_i^{(0)}$ and $\hat{C}_{it}^{(0)}$, respectively. In theory, the distributional results in Theorem 2.4 require $\ell \to \infty$. In practice, ℓ can diverge to infinity at an arbitrarily slow rate. To see this point, we take a close look at the iterative relationship between $\hat{\Pi}_{tN}^{(\ell)}$ and $\hat{\Pi}_{tN}^{(\ell-1)}$. Let $\beta_{F,t} = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it}$. Note that the result in Theorem 2.4(i) implies

$$\hat{\Pi}_{tN}^{(\ell)} = D^{-1} Q \sqrt{N} \beta_{F,t} \sum_{s=0}^{\ell-1} (1-q)^s + (1-q)^\ell\, \hat{\Pi}_{tN}^{(0)} + o_P(1),$$

where the first term is the dominant term and the second term can be made arbitrarily small for sufficiently large ℓ. In practice, we find it is not necessary to iterate too many times so that we can stop the iteration when $(1-q)^{\ell}$ is small enough. For example, we can iterate ℓ^{*} times such that $(1-q)^{\ell^{*}} \asymp \epsilon_{NT}$ for some small positive number ϵ_{NT}. Simulations suggest that $\ell^{*}=\lfloor\ln(\epsilon_{NT})/\ln(1-q)\rfloor$ with $\epsilon_{NT}=0.001$ works very well for all data generating processes under our investigation. Note that $\ell^{*}=3$, 4, and 5 for $q=0.9$, 0.8, and 0.7, respectively. This suggests a small number of iterations is sufficient.

Remark 3 (Comparison with the oracle estimators) We can also compare the asymptotic variances of our EM estimators with those of the oracle estimators that are obtained in the absence of missing values (viz., $q=1$). For example, we consider the factor estimation and use $\hat{F}_{t}^{\text{oracle}}$ to denote the oracle estimator of F_{t}^{0} with the corresponding rotational matrix $\hat{H}^{\text{oracle}}$. It is well known that the asymptotic variance-covariance (Avar) of $\sqrt{N}(\hat{F}_{t}^{\text{oracle}}-\hat{H}^{\text{oracle}\prime}F_{t}^{0})$ is given by $D^{-1}Q\Gamma_{t}^{\text{oracle}}Q'D^{-1}$, where

$$\Gamma_{t}^{\text{oracle}}=\lim_{N\to\infty}\text{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{i}^{0}\varepsilon_{it}\right).$$

In contrast, by the law of iterated expectations

$$\begin{aligned}\Gamma_{1g,t}(q) &= \lim_{N\to\infty}\left\{\text{Var}\left[E\left(\frac{1}{\sqrt{N}q}\sum_{i=1}^{N}\lambda_{i}^{0}\varepsilon_{it}g_{it}|\Lambda^{0},\varepsilon\right)\right]+E\left[\text{Var}\left(\frac{1}{\sqrt{N}q}\sum_{i=1}^{N}\lambda_{i}^{0}\varepsilon_{it}g_{it}|\Lambda^{0},\varepsilon\right)\right]\right\}\\ &= \lim_{N\to\infty}\left\{\text{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{i}^{0}\varepsilon_{it}\right)+\frac{1-q}{q}E\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i}^{0}\lambda_{i}^{0\prime}\varepsilon_{it}^{2}\right)\right\}\\ &\geq \Gamma_{t}^{\text{oracle}}.\end{aligned}$$

The difference, $\Gamma_{1g,t}(q)-\Gamma_{t}^{\text{oracle}}$, given by $\lim_{N\to\infty}\frac{1-q}{q}E\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i}^{0}\lambda_{i}^{0\prime}\varepsilon_{it}^{2}\right)$, reflects the cost of missing $(1-q)$ proportion of observations. The larger proportion of missing observations, the larger value $\Gamma_{1g,t}(q)$ is. In the absence of cross-sectional correlation among $\left\{\lambda_{i}^{0}\varepsilon_{it}\right\}$, it is easy to verify that

$$\Gamma_{1g,t}(q)=\frac{1}{q}\lim_{N\to\infty}E\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i}^{0}\lambda_{i}^{0\prime}\varepsilon_{it}^{2}\right)=\frac{1}{q}\Gamma_{t}^{\text{oracle}}.$$

So q reflects the relative asymptotic efficiency of the EM estimator compared to the oracle estimator. Analogous remarks hold for our EM estimators of the factor loadings.

With the results in Theorem 2.4, we can make inference on the factors, factor loadings, and common component. Below we focus on the inference on the factors due to the widespread use of estimated factors, say, in various factor-augmented regression or forecasting models.

2.4 Inference on the factors

Let $\hat{F}_t$, $\hat{\lambda}_i$, and $\hat{C}_{it}$ denote $\hat{F}_t^{(\ell)}$, $\hat{\lambda}_i^{(\ell)}$, and $\hat{C}_{it}^{(\ell)}$ respectively, when $\ell \to \infty$. To make inference on the factors, we need to estimate the asymptotic variance $V_F \equiv D^{-1} Q \Gamma_{1g,t}(q) Q' D^{-1}$ consistently. By Lemma A.1 in the appendix, we can consistently estimate D by the diagonal matrix $\hat{D} = \hat{D}^{(\infty)}$, that contains the R largest eigenvalues of $(NT)^{-1} \hat{X}^{(\infty)} \hat{X}^{(\infty)\prime}$, arranged in descending order. So the key is to estimate $Q\Gamma_{1g,t}Q'$ consistently.

To estimate $Q\Gamma_{1g,t}(q)Q'$, we consider two cases: (1) $\{\lambda_i^0 \varepsilon_{it} g_{it}\}$ are cross-sectionally uncorrelated; (2) $\{\lambda_i^0 \varepsilon_{it} g_{it}\}$ are cross-sectionally correlated. In Case (1), we have a simplified expression for $\Gamma_{1g,t}(q)$

$$\Gamma_{1g,t}(q) = \lim_{N\to\infty} \frac{1}{Nq^2} \sum_{i=1}^{N} \mathrm{Var}\left(\lambda_i^0 \varepsilon_{it} g_{it}\right) = \lim_{N\to\infty} \frac{1}{Nq^2} \sum_{i=1}^{N} E\left[\lambda_i^0 \lambda_i^{0\prime} (\varepsilon_{it}^g)^2\right],$$

where $\varepsilon_{it}^g = \varepsilon_{it} g_{it}$. Noting that with $\tilde{H} \equiv \hat{H}^{(0)}$, $\tilde{H}^{-1} \xrightarrow{p} Q$ by Lemma A.2(ii) in the appendix, it is easy to show that a consistent estimator of $Q\Gamma_{1g,t}(q)Q'$ is given by

$$\hat{\Gamma}_{1g,t}^{(1)} = \frac{1}{N\tilde{q}^2} \sum_{i=1}^{N} \hat{\lambda}_i \hat{\lambda}_i' (\hat{\varepsilon}_{it}^g)^2,$$

where $\hat{\varepsilon}_{it}^g = (X_{it} - \hat{C}_{it}) g_{it}$.

In Case (2), for simplicity we consider the case where the factor loadings are nonrandom and the process $\{\varepsilon_{it},\ t \geq 1\}$ is covariance stationary. Let $\varepsilon_{\cdot t}^g = \left(\varepsilon_{1t}^g, \varepsilon_{2t}^g, ..., \varepsilon_{Nt}^g\right)'$. Let $\Sigma^g \equiv E(\varepsilon_{\cdot t}^g \varepsilon_{\cdot t}^{g\prime}) = \{\sigma_{ij}^g\}$, which is an $N \times N$ matrix. Then

$$\Gamma_{1g,t}(q) = \lim_{N\to\infty} \frac{1}{Nq^2} \mathrm{Var}\left(\Lambda^{0\prime} \varepsilon_{\cdot t}^g\right) = \lim_{N\to\infty} \frac{1}{Nq^2} \Lambda^{0\prime} \Sigma^g \Lambda^0.$$

Suppose that $\tilde{\Sigma}^g$ is a consistent estimator of Σ^g in the sense $\left\|\hat{\Sigma}^g - \Sigma^g\right\|_{\mathrm{sp}} = o_p(1)$. Then we can readily show that a consistent estimator of $Q\Gamma_{1g,t}Q'$ is given by

$$\hat{\Gamma}_{1g,t}^{(2)} \equiv \frac{1}{N\tilde{q}^2} \hat{\Lambda}' \hat{\Sigma}^g \hat{\Lambda}.$$

Fortunately, a feasible consistent estimator of Σ^g is available as ε_{it}^g can be estimated by $\hat{\varepsilon}_{it}^g$ and there is no need to estimate the error terms corresponding to those missing observations. To see this, define

$$\hat{\sigma}_{ij}^g = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^g \hat{\varepsilon}_{jt}^g \text{ and } \hat{\theta}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \left(\hat{\varepsilon}_{it}^g \hat{\varepsilon}_{jt}^g - \hat{\sigma}_{ij}^g\right)^2.$$

We follow the lead of Fan, Liao and Mincheva (2013, FLM hereafter) and propose to estimate Σ^g by $\hat{\Sigma}^g = \left\{\hat{\sigma}_{ij}^{g,\mathcal{T}}\right\}$, where

$$\hat{\sigma}_{ij}^{g,\mathcal{T}} = \begin{cases} \hat{\sigma}_{ij}^g & \text{if } i = j \\ s_{ij}(\hat{\sigma}_{ij}^g) & \text{if } i \neq j \end{cases},$$

where $s_{ij}(\cdot)$ is the soft thresholding function: $s_{ij}(z) \equiv sgn(z)(|z| - \tau_{ij})_+$, $\tau_{ij} = c_0 \omega_{NT} (\hat{\theta}_{ij})^{1/2}$, $\omega_{NT} = [\max(N^{-1+\gamma_2/2}, T^{-1}\ln T)]^{1/2}$, and c_0 is a positive constant.[1] We will show that $\left\| \hat{\Sigma}^g - \Sigma^g \right\|_{\text{sp}} = o_P(1)$ under some additional conditions.

When Λ^0 is random, the above procedure also works under the additional restriction that $\text{Var}(\varepsilon^g_{\cdot t}|\Lambda^0) = \text{Var}(\varepsilon^g_{\cdot t}) = \Sigma^g$. To see this, we notice that by the variance decomposition formula, we have

$$\begin{aligned}
\Gamma_{1g,t}(q) &= \lim_{N\to\infty} \frac{1}{Nq^2} E\left[\text{Var}\left(\Lambda^{0\prime}\varepsilon^g_{\cdot t}|\Lambda^0\right)\right] + \lim_{N\to\infty} \frac{1}{Nq^2} \text{Var}\left(E\left(\Lambda^{0\prime}\varepsilon^g_{\cdot t}|\Lambda^0\right)\right) \\
&= \lim_{N\to\infty} \frac{1}{Nq^2} E\left[\Lambda^{0\prime}\text{Var}\left(\varepsilon^g_{\cdot t}|\Lambda^0\right)\Lambda^0\right] + 0 = \lim_{N\to\infty} \frac{1}{Nq^2} E\left[\Lambda^{0\prime}\Sigma^g\Lambda^0\right].
\end{aligned}$$

$\frac{1}{Nq^2} E\left[\Lambda^{0\prime}\Sigma^g\Lambda^0\right]$ can be estimated in the same procedure as outlined above.

To allow for possible cross-sectional dependence, we recommend using $\hat{\Gamma}^{(2)}_{1g,t}$ and will justify the consistency of this estimator below. To proceed, we add the following assumption.

Assumption A.6 (i) The process $\{\varepsilon^g_{\cdot t},\ t \geq 1\}$ is covariance-stationary with covariance matrix $\Sigma^g = E(\varepsilon^g_{\cdot t}\varepsilon^{g\prime}_{\cdot t}) = \left\{\sigma^g_{ij}\right\}$.

(ii) There exists $\gamma_3 \in [0,1)$ such that $\max_i \sum_j \left|\sigma^g_{ij}\right|^{\gamma_3} \leq M$.

(iii) Let $\omega_{NT} = [\max(N^{-1+\gamma_2/2}, T^{-1}\ln T)]^{1/2}$. $T^{-1/2+\gamma_1/4}(N^{\gamma_2/4} + T^{\gamma_1/4})(\ln T)^{1/2} \to 0$ and $T^{-1+\gamma_1/4}\omega_{NT}^{1-\gamma_3}N^{1/2} \to 0$ as $(N,T) \to \infty$.

Assumption A.6(i) is typically assumed in the literature when there is no missing value. Assumption A.6(ii) strengthens the standard weak cross-sectional dependence condition $\max_i \sum_j \left|\sigma^g_{ij}\right| = O(1)$. It is satisfied if $\varepsilon^g_{\cdot t}$'s satisfy certain m-dependence condition cross-sectionally or the correlation between ε^g_{it} and ε^g_{jt} vanishes sufficiently fast as the "distance" between i and j increases, perhaps after reordering of the data along the cross-sectional dimension. Assumption A.6(iii) imposes further restrictions on the relative magnitude of N and T.

The following theorem reports the consistency of $\hat{D}^{-1}\hat{\Gamma}_{1g,t}\hat{D}^{-1}$.

Theorem 2.5 *Suppose that Assumptions A.1-A.6 hold. Then* $\hat{D}^{-1}\hat{\Gamma}_{1g,t}\hat{D}^{-1} \xrightarrow{p} D^{-1}Q\Gamma_{1g,t}(q)Q'D^{-1}$, *where* $\hat{\Gamma}_{1g,t} = \hat{\Gamma}^{(2)}_{1g,t}$.

Given the above result, we can make inference on the global factors. The procedure is standard and omitted for brevity.

[1] In our simulations and applications, we let $c_0 = 1$. In most situations, when $c_0 = 1$, $\tilde{\Sigma}^g$ is positive definite. Otherwise, we choose c_0 to be the smallest value such that $\tilde{\Sigma}^g$ is positive definite. For details, see FLM's Section 4.

3 Determining the Number of Factors via Cross Validation

In this section, we propose a novel method to determine the number of factors via cross validation (CV). Our method can be used no matter whether there are random missing values in the original data matrix X or not. For notational simplicity, we first focus on the CV method when the original dataset does not have missing value problem and then remark on the case with missing values.

3.1 The cross validation method

Let R denote the generic number of factors with the true value given by R_0. The key insight for our CV method is that one can consistently estimate the common component for the factor models with random missing. Given the $T \times N$ matrix of observations X, we propose to randomly sample elements in X with a fixed probability $p \in (0,1)$ and leave the rest $(1-p)$-proportion of observations as held-out entries for the out-of-sample evaluation.

As before, let $\Omega^* \subset [N] \times [T]$ be the index set of the training entries and $\Omega^*_{\perp}$ the index set of the held-out entries. Define the operator $P_{\Omega^*} : \mathbb{R}^{T\times N} \to \mathbb{R}^{T\times N}$ by

$$(P_{\Omega^*}X)_{ti} = X_{it}g^*_{it} = X_{it}\mathbf{1}\left\{(i,t) \in \Omega^*\right\},$$

where $g^*_{it} = \mathbf{1}\left\{(i,t) \in \Omega^*\right\}$. Let G^* denote a $T \times N$ matrix with (t,i)th element given by g^*_{it}. Now we can regard $P_{\Omega^*}X$ as the $T \times N$ data matrix with missing values replaced by zeros. Given $P_{\Omega^*}X$, we apply the proposed EM algorithm to recover the data via estimating the common component matrix C for any given number of factors.

To proceed, we consider the full singular value decomposition (SVD) for $\frac{1}{p}P_{\Omega^*}X$:

$$\frac{1}{p}P_{\Omega^*}X = \tilde{U}\tilde{\Sigma}\tilde{V}' = \sum_{r=1}^{T\wedge N} \tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r,$$

where $\tilde{U} = (\tilde{u}_1, ..., \tilde{u}_T)$ and $\tilde{V} = (\tilde{v}_1, ..., \tilde{v}_N)$ are respectively the $T \times T$ matrix of left singular vectors and $N \times N$ matrix of right singular vectors of $\frac{1}{p}P_{\Omega^*}X$, and $\tilde{\Sigma}$ is the $T \times N$ diagonal matrix that contains the singular values, $\tilde{\sigma}_1$, $\tilde{\sigma}_2, ..., \tilde{\sigma}_{T\wedge N}$, arranged in descending order along the main diagonal line. Given any $R \le T \wedge N$ and the training entries in $P_{\Omega^*}X$, we can estimate the common component C by the singular value thresholding procedure:

$$\tilde{C}_R = S_H\left(\frac{1}{p}P_{\Omega^*}X, R\right) = \tilde{U}_R\tilde{\Sigma}_R\tilde{V}_R' = \sum_{r=1}^{R} \tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r, \tag{3.1}$$

where $S_H(\cdot, R)$ is the rank-R truncated SVD of $\cdot$, the subscript H stands for hard thresholding, $\tilde{U}_R = (\tilde{u}_1, ..., \tilde{u}_R)$, $\tilde{V}_R = (\tilde{v}_1, ..., \tilde{v}_R)$, and $\tilde{\Sigma}_R$ =diag$(\tilde{\sigma}_1, ..., \tilde{\sigma}_R)$. We can regard $\tilde{C}_R$ as a matrix-completion version of $P_{\Omega^*}X$. Let $\tilde{C}_{R,it}$ denote the (t,i)th element of $\tilde{C}_R$. We propose to choose R to minimize the following CV criterion function

$$\widetilde{CV}(R) = \sum_{(i,t)\in\Omega_\perp^*} \left[X_{it} - \tilde{C}_{R,it}\right]^2. \tag{3.2}$$

Let $\tilde{R} = \arg\min_{0\le R\le R_{\max}} \widetilde{CV}(R)$ where $R_{\max}$ is a fixed integer that is no less than R_0 and $\tilde{C}_{0,it} = 0$ for all (i,t). We will show the consistency of $\tilde{R}$ under some regularity conditions.

Note that the CV function in (3.2) is based on the initial estimator $\tilde{C}_R$ of the common component matrix C^0. As demonstrated in the last subsection, one can update the estimator of C^0 via the EM algorithm and obtain a more efficient estimator of C. It is expected that using such a more efficient estimator would yield better finite sample performance for the choice of the correct number of factors. As before, let $\hat{C}_{R,it}^{(0)} = \tilde{C}_{R,it}$ and $\ell \ge 1$ be an integer. Suppose that we have obtained the estimates $\hat{C}_{R,it}^{(\ell-1)}$. In step ℓ, we can replace the zero elements in $X^* \equiv P_{\Omega^*}X$ with the estimated common components $\hat{C}_{R_{\max},it}^{(\ell-1)}$.[2] Define the $T\times N$ matrix $\hat{X}^{*(\ell)}$ with its (t,i)th element given by

$$\hat{X}_{it}^{*(\ell)} = \begin{cases} X_{it} & \text{if } (i,t)\in\Omega^* \\ \hat{C}_{R_{\max},it}^{(\ell-1)} & \text{if } (i,t)\in\Omega_\perp^* \end{cases}, \ \ell \ge 1, \tag{3.3}$$

where $\Omega_\perp^* = \{(i,t)\in[N]\times[T] : (i,t)\notin\Omega^*\}$. Then we can conduct the singular value thresholding procedure:

$$\hat{C}_R^{(\ell)} = S_H\left(\hat{X}^{*(\ell)}, R\right) = \hat{U}_R^{(\ell)}\hat{\Sigma}_R^{(\ell)}\hat{V}_R^{(\ell)\prime}, \tag{3.4}$$

where $\hat{U}_R^{(\ell)\prime}\hat{U}_R^{(\ell)} = I_R$, $\hat{V}_R^{(\ell)\prime}\hat{V}_R^{(\ell)} = I_R$, and $\hat{\Sigma}_R^{(\ell)}$ is a diagonal matrix that contains the R largest singular values of $\hat{X}^{*(\ell)}$ arranged in descending order along its diagonal line. Following Remark 2, we recommend repeating the above procedure for $\ell = 1, ..., \ell^* \equiv \lfloor \ln(\epsilon_{NT})/\log(p)\rfloor$ where, e.g., $\epsilon_{NT} = 0.001$. Let $\hat{C}_R = \hat{C}_R^{(\ell^*)}$ and $\hat{R} = \arg\min_{0\le R\le R_{\max}} \widehat{CV}(R)$, where

$$\widehat{CV}(R) = \sum_{(i,t)\in\Omega_\perp^*} \left[X_{it} - \hat{C}_{R,it}\right]^2. \tag{3.5}$$

We will show the consistency of $\hat{R}$ under some regularity conditions.

3.2 The consistency of the CV method

Let $\tilde{u}_r$ and $\tilde{v}_r$ denote the rth left and right singular vectors of $\frac{1}{\sqrt{p}}X^*$, respectively, associated with its rth largest singular value. We add one assumption.

[2] We conjecture that one can replace $\hat{C}_{R_{\max},it}^{(\ell-1)}$ by $\hat{C}_{R,it}^{(\ell-1)}$ in which case $\hat{X}_{it}^{*(\ell)}$ becomes

$$\hat{X}_{R,it}^{*(\ell)} = \begin{cases} X_{it} & \text{if } (i,t)\in\Omega^* \\ \hat{C}_{R,it}^{(\ell-1)} & \text{if } (i,t)\in\Omega_\perp^* \end{cases}, \ell \ge 1.$$

But the justification for this method is far more complicated than the proof of Theorem 3.2 below because of the dependence of $\hat{X}_{R,it}^{*(\ell)}$ on R and the inconsistency of $\hat{C}_{R,it}^{(\ell-1)}$ for $R < R_0$.

Assumption A.7. (i) For $r = R_0+1, ..., R_{\max}$, $P(\|\tilde{u}_r\|_\infty \|\tilde{v}_r\|_\infty \leq 1/(c_0\sqrt{(N+T)\log(N+T)})) \to 1$ for some fixed $c_0 < \infty$ as $(N,T) \to \infty$, $\|\tilde{u}_r\|_\infty = o_P(1)$, and $\|\tilde{v}_r\|_\infty = o_P(1)$;

(ii) $\max_{(i,t)\in\Omega_\perp^*} \sum_{(j,s)\in\Omega_\perp^*} |E[\varepsilon_{it}\varepsilon_{js}|P_{\Omega^*}X, \Omega^*]| = o_P(\delta_{NT}^2)$.

Assumption A.7(i) is a high order condition that restricts the spikeness of singular vectors of X. A similar condition is also imposed in Negahban and Wainwright (2012). Since $\|\tilde{u}_r\|_2 = \|\tilde{v}_r\|_2 = 1$, on average each entry of $\tilde{u}_r\tilde{v}_r'$ is of the order $(NT)^{-1/2}$. We require the maximum entry is bounded by the order $((N+T)\log(N+T))^{-1/2}$. We can show that $\tilde{u}_r$ and $\tilde{v}_r$ are asymptotically equal to the $(r-R_0)$th singular vector of $\varsigma^* \equiv \varepsilon \circ G^* + F^0\Lambda^{0\prime} \circ [G^* - E(G^*)]/p$, where each entry has zero mean. As we do not have the explicit form of $\tilde{u}_r$ and $\tilde{v}_r$, it is difficult to show its spikeness. It is well known that for an i.i.d. Gaussian random matrix, the elements of its right and left eigenvectors are uniformly distributed on the unit spheres S^{N-1} and S^{T-1}, respectively. Then Assumption A.7(i) is satisfied in this case. It is expected that the singular vectors of a general random matrix behave similarly. Assumption A.7 (ii) is a higher order condition that requires low degree of correlations among $\{\varepsilon_{it}\}$, conditional on kept-in information. It is satisfied when ε_{it} is i.i.d. and the factors and factor loadings are nonrandom. When we have $|E[\varepsilon_{it}\varepsilon_{js}|P_{\Omega^*}X, \Omega^*]| \leq M\rho^{|t-s|+|j-i|}$ for some $M < \infty$ and $\rho < 1$ perhaps after reordering the data along the cross-sectional direction, the condition is also satisfied.

The next two theorems establish the selection consistency of our CV method based on $\widetilde{CV}(R)$ and $\widehat{CV}(R)$.

Theorem 3.1 *Suppose Assumptions A.1-A.3 hold, and Assumptions A.4-A.5 hold with* $g_{it} \equiv 1$. *Then* $P\left(\tilde{R} < R_0\right) \to 0$ *as* $(N,T) \to \infty$. *If Assumption A.7 also holds, then* $P\left(\tilde{R} > R_0\right) \to 0$ *as* $(N,T) \to \infty$.

Theorem 3.2 *Suppose Assumptions A.1-A.3 hold, and Assumptions A.4-A.5 hold with* $g_{it} \equiv 1$. *Then* $P\left(\hat{R} < R_0\right) \to 0$ *as* $(N,T) \to \infty$. *If Assumption A.7 also holds, then* $P\left(\hat{R} > R_0\right) \to 0$ *as* $(N,T) \to \infty$.

Theorems 3.1 and 3.2 indicate that the CV estimators $\tilde{R}$ and $\hat{R}$ consistently estimate the true number of factors R_0 in large samples when Assumptions A.1-A.5 and A.7 hold. As we show in the proof of Theorem 3.1, the consistency of $\tilde{R}$ is established by demonstrating that

$$\begin{aligned}\widetilde{CV}(R) - \widetilde{CV}(R_0) &= (1-p)\sum_{r=R+1}^{R_0} \sigma_r^2 + O_P\left(\delta_{NT}^{-1}\right) \text{ when } R < R_0, \text{ and} \\ \text{plim}_{(N,T)\to\infty}\delta_{NT}^2\left[\widetilde{CV}(R) - \widetilde{CV}(R_0)\right] &\geq \frac{1-p}{256}(R-R_0)\,c_\sigma > 0 \text{ when } R > R_0,\end{aligned}$$

where c_σ is the lower probability bound of $\delta_{NT}^2 (NT)^{-1} \tilde{\sigma}_r^2$ for $r \in \{R_0+1, ..., R_{\max}\}$. Note that $\tilde{\sigma}_r^2$ diverges to infinity in probability at the rate NT for $r \in \{1, ..., R_0\}$ and $(NT)^{-1}\tilde{\sigma}_r^2$ converges to zero in probability at the rate δ_{NT}^{-2} when $r \in \{R_0+1, ..., R_{\max}\}$. Similar remarks hold true for $\widehat{CV}(R) - \widehat{CV}(R_0)$.

3.3 CV in the presence of random missing

From the proof of Theorem 3.1 we can see that the same result holds with some modifications when the original data matrix X contains random missing values. To see the modifications, we continue to use $\Omega \subset [N] \times [T]$ to denote the index set of the observations that are observed. Let $g_{it} = \mathbf{1}\{(i,t) \in \Omega\}$ and $\tilde{q} \equiv |\Omega| / (NT)$. As before, $P(g_{it} = 1) = q \in (0, 1]$ and g_{it} is independent of X, F^0, Λ^0 and ε. In this case, we consider the SVD for $\frac{1}{p\tilde{q}} P_{\Omega^*} P_\Omega X$:

$$\frac{1}{p\tilde{q}} P_{\Omega^*} P_\Omega X = \tilde{U} \tilde{\Sigma} \tilde{V}',$$

where $\tilde{U}$ is now the $T \times T$ matrix of left singular vectors of $\frac{1}{p\tilde{q}} P_{\Omega^*} P_\Omega X$, $\tilde{V}$ is the $N \times N$ matrix of right singular vector of $\frac{1}{p\tilde{q}} P_{\Omega^*} P_\Omega X$, and $\tilde{\Sigma}_R$ contains the singular values of $\frac{1}{p\tilde{q}} P_{\Omega^*} P_\Omega X$ arranged in descending order along its diagonal line. Then we estimate the common component C by the singular value thresholding procedure:

$$\tilde{C}_R = S_H \left(\frac{1}{p\tilde{q}} P_{\Omega^*} P_\Omega X, R \right) = \tilde{U}_R \tilde{\Sigma}_R \tilde{V}_R', \tag{3.6}$$

where $\tilde{U}_R$, $\tilde{V}_R$, and $\tilde{\Sigma}_R$ are defined as before. Let $\tilde{R} \in \{0, 1, 2, ..., R_{\max}\}$ minimize the following CV function

$$\widetilde{CV}(R) = \sum_{(i,t) \in \Omega_\perp^* \cap \Omega} \left[X_{it} - \tilde{C}_{R,it} \right]^2, \tag{3.7}$$

where $\tilde{C}_{R,it}$ denote the (t,i)th element of $\tilde{C}_R$. Following the proof of Theorem 3.1, we can also show that $P(\tilde{R} = R_0) \to 1$ as $(N, T) \to \infty$ in this case.

As in the last subsection, we can consider iterative estimates of C. Let $\hat{C}_{R,it}^{(0)} = \tilde{C}_{R,it}$. Suppose that we have obtained the estimates $\hat{C}_{R,it}^{(\ell-1)}$. In step ℓ, we can replace the zero elements in $P_{\Omega^*} P_\Omega X$ with the estimated common components $\hat{C}_{R_{\max},it}^{(\ell-1)}$.[3] Define the $T \times N$ matrix $\hat{X}^{*(\ell)}$ with its (t,i)th

[3] We conjecture that one can replace $\hat{C}_{R_{\max},it}^{(\ell-1)}$ by $\hat{C}_{R,it}^{(\ell-1)}$ in which case $\hat{X}_{it}^{*(\ell)}$ becomes

$$\hat{X}_{R,it}^{*(\ell)} = \begin{cases} X_{it} & \text{if } (i,t) \in \Omega^* \cap \Omega \\ \hat{C}_{R,it}^{(\ell-1)} & \text{if } (i,t) \in \Omega^* \cap \Omega_\perp \\ 0 & \text{if } (i,t) \in \Omega_\perp^* \end{cases}, \ell \geq 1.$$

element given by

$$\hat{X}_{it}^{*(\ell)} = \begin{cases} X_{it} & \text{if } (i,t) \in \Omega^* \cap \Omega \\ \hat{C}_{R_{\max},it}^{(\ell-1)} & \text{if } (i,t) \in \Omega^* \cap \Omega_\perp \\ 0 & \text{if } (i,t) \in \Omega_\perp^* \end{cases}, \ \ell \geq 1. \tag{3.8}$$

Note that for observations with $(i,t) \in \Omega_\perp^*$ we do not need to replace them by the iterated estimates $\hat{C}_{R_{\max},it}^{(\ell-1)}$ in step ℓ. Then we can conduct the singular value thresholding procedure:

$$\hat{C}^{(\ell)}(R) = S_H\left(\frac{1}{p}\hat{X}^{*(\ell)}, R\right) = \hat{U}_R^{(\ell)}\hat{\Sigma}_R^{(\ell)}\hat{V}_R^{(\ell)\prime}, \tag{3.9}$$

where $\hat{U}_R^{(\ell)\prime}\hat{U}_R^{(\ell)} = I_R$, $\hat{V}_R^{(\ell)\prime}\hat{V}_R^{(\ell)} = I_R$, and $\hat{\Sigma}_R^{(\ell)}$ is a diagonal matrix that contains the R largest singular values of $\hat{X}^{*(\ell)}$ arranged in descending order along its diagonal line. Following Remark 2, let $\hat{C}_R = \hat{C}_R^{(\ell^*)}$ and $\hat{R} = \arg\min_{0 \leq R \leq R_{\max}} \widehat{CV}(R)$, where

$$\widehat{CV}(R) = \sum_{(i,t) \in \Omega_\perp^* \cap \Omega} \left[X_{it} - \hat{C}_{R,it}\right]^2. \tag{3.10}$$

Following the proof of Theorem 3.2, we can also show that $P(\hat{R} = R_0) \to 1$ as $(N,T) \to \infty$ in this case.

3.4 Averaging CV and stability selection

The CV method in Sections 3.1 and 3.3 is based on a single random draw for the training set of observations. The resulting performance of the CV method can be affected by the quality of such a draw. In practice, we can always average $\widetilde{CV}(R)$ or $\widehat{CV}(R)$ over a large number (say, J) of draws.

Recognizing the notorious difficulty in the estimation of discrete structures, such as in variable selection and cluster analysis, Meinshausen and Bühlmann (2010) introduce stability selection based on subsampling in combination with some selection algorithms. The procedure serves as a general method to reduce noise by repeating the model selection many times over random splits of the data. Our CV procedure can benefit from the stability selection since it relies on random data splits. An additional benefit of stability selection in our context is that it is more robust to the choices of p and J. The algorithm is given below.

Algorithm 1 (The CV procedure)

1. For $(j,k) \in [J] \times [K]$

   (a) Randomly choose a subset of training observations $\Omega \subset [N] \times [T]$ where each observation in X can be chosen with probability p.

(b) Apply the thresholding SVD in (3.1) or (3.6) to obtain $\tilde{C}_R$ or that in (3.4) or (3.9) to obtain $\hat{C}_R$ for $R = 0, 1, ..., R_{\max}$, respectively. Here $\tilde{C}_0$ and $\hat{C}_0$ are $T \times N$ matrices of zeros.

(c) For each $R \in \{0, 1, ..., R_{\max}\}$, calculate the CV value via (3.2) or (3.7) and denote it as $\widetilde{CV}^{(j,k)}(R)$ or that via (3.5) or (3.10) and denote it as $\widehat{CV}^{(j,k)}(R)$.

2. Let $\widetilde{CV}_k(R) = \frac{1}{J}\sum_{j=1}^{J} \widetilde{CV}^{(j,k)}(R)$ and $\widehat{CV}_k(R) = \frac{1}{J}\sum_{j=1}^{J} \widehat{CV}^{(j,k)}(R)$ for $k = 1, ..., K$. Let

$$\tilde{R}_k = \arg\min_{0 \le R \le R_{\max}} \widetilde{CV}_k(R) \text{ and } \hat{R}_k = \arg\min_{0 \le R \le R_{\max}} \widehat{CV}_k(R) \text{ for } k = 1, ..., K.$$

Let $\tilde{R}$ and $\hat{R}$ denote the modes in $\{\tilde{R}_1, ..., \tilde{R}_K\}$ and $\{\hat{R}_1, ..., \hat{R}_K\}$, respectively. $\tilde{R}$ and $\hat{R}$ serve as the estimator of the true number of factors without and with iterations.

We will evaluate the finite sample performance of $\tilde{R}$ and $\hat{R}$ through simulations by setting $K = 10$ and $J = 5$.

4 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performance of our proposed EM estimators and CV method.

4.1 Data generating processes

First, we introduce data generating processes (DGP) for the factors and factor loadings. We generate the factors according to

$$F_t - \mu_f \iota_R = \rho_f (F_{t-1} - \mu_f \iota_R) + v_t, \; t = 1, ..., T$$

where ι_R is an $R \times 1$ vector of ones, μ_f is a scalar, v_t is independent and identically distributed (i.i.d.) from $N(0, (1 - \rho_f^2) I_R)$, and $\rho_f \in (0, 1)$. To avoid the start-up effect, we throw away the first 1000 observations of $\{F_t\}$ and use the next T observations for the estimation below. For the factor loadings, we let λ_{ir}, $i = 1, ..., N$ and $r = 1, ..., R$ be i.i.d. draws from $c_s \cdot N(1, 1)$, where c_s is a constant controlling the signal strength. In addition, F, Λ and ε are mutually independent for all DGPs.

Next, we introduce the generation of the idiosyncratic error terms ε_{it} in DGPs 1–6:

DGP 1. We let $\varepsilon_{it} = \left[0.9 + 0.1(\lambda_i' F_t)^2 / E(\lambda_i' F_t)^2\right] u_{it}$, where u_{it} is i.i.d. from $t(3)$, the student t-distribution with 3 degrees of freedom. In this case, the error term ε_{it} does not have a finite fourth moment, which violates Assumption A.2(i). There is conditional heteroskedasticity but no serial or cross-sectional correlation among ε_{it}'s.

DGP 2. The setting is the same as DGP 1 except that u_{it} is i.i.d. from $t\,(5)$. Now all the assumptions are satisfied but the tail distribution is not sub-Gaussian.

DGP 3. We generate autoregressive ε_{it} via an AR(1) process: $\varepsilon_{it} = \rho_{1\varepsilon}\varepsilon_{i,t-1} + u_{it}$, where u_{it} is i.i.d. $N\,(0,1)$ and $\rho_{1\varepsilon} \in (0,1)$. In simulation, we delete the first 100 observations to avoid the start-up effect.

DGP 4. We stack ε_{it} into a $T \times N$ matrix ε and generate $\varepsilon = UA$, where U is a $T \times N$ random matrix and A is an $N \times N$ random matrix. The (t,i)th entry u_{it} of U is i.i.d. from $N\,(0,1)$ and the matrix A controls the cross-sectional dependence. In particular, we let $A = VDV'$, where V is a random orthonormal matrix, $D =$diag$(d_1, ..., d_N)$ is a diagonal matrix, and V and D are independent. To generate D, we draw N i.i.d. observations $\{d_i\}_{i=1}^N$ from the uniform distribution U[0.5,1.5]. Then we set

$$D = \text{diag}(N^{1/8}d_1, ...,\ N^{1/8}d_{\lfloor 0.1N \rfloor},\ d_{\lfloor 0.1N \rfloor +1}, ...,\ d_N),$$

where $\lfloor \cdot \rfloor$ returns the integer part of $\cdot$. Now, there is strong cross-sectional correlation as we allow the top 10% of the eigenvalues of D to be $O(N^{1/8})$. So the weak dependence conditions on the error terms in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) are not satisfied. We want to examine the performance of different methods in this case.

DGP 5. We generate $\varepsilon_{it} = \rho_{2\varepsilon}\varepsilon_{i-1,t} + u_{it}$, where u_{it} is i.i.d. $N\,(0,1)$. This DGP is similar to DGP 3 except that we now allow the error terms to be cross-sectionally dependent.

DGP 6. We generate $\varepsilon_{it} = u_{it} + \rho_{3\varepsilon}u_{i,t-1} + \rho_{3\varepsilon}u_{i-1,t} + \rho_{3\varepsilon}^2 u_{i-1,t-1}$, where u_{it} is i.i.d. $N\,(0,1)$ and $\rho_{3\varepsilon} \in (-1,1)$. Note that we now allow for both cross-sectional and serial dependence in the error terms.

In all our experiments, we let $\mu_f = 0.6$, $\rho_f = 0.3$ and choose c_s such that signal to noise ratio (SNR) equals 4 for each DGP. Specifically, we define SNR as var$(\lambda_i' F_t)$/var(ε_{it}).

4.2 Simulation results

In this subsection, we present our simulation results in two parts. In the first part, we examine the accuracy of the CV method proposed in section 3, measured by the empirical frequency of correct determination of the number of factors. In the second part, we estimate the model with the true number of factors and report the finite sample performance of the proposed estimator introduced in section 2.

4.2.1 Determining the number of factors

In this part, we use the CV method to determine the number of factors for data with or without random missing observations. For both cases, we let $R_0 = 3$, and $R_{\max} = 5$. In addition, we set $\rho_{1\varepsilon} = 0.6$, $\rho_{2\varepsilon} = 0.6$ and $\rho_{3\varepsilon} = 0.3$ in DGPs 3, 5 and 6, respectively. For each DGP, we consider $N = 50, 100$ and $T = 50, 100$, leading to four combinations of cross-sectional and time series dimensions. To implement the averaging CV and stability selection method in Section 3.4, we set $K = 10$ and $J = 5$. For the case of complete data, we consider two leave-out probabilities: $p = 0.7$, 0.9. For the case of incomplete data, we consider two random missing probabilities: $q = 0.7$, 0.9 and use the leave-out probability $p = 0.9$. The number of replications is 1000 in all cases.

When the original data form a balanced panel, there are existing methods including the growth ratio (GR) and eigenvalue ratio (ER) of Ahn and Horenstein (2013), the edge distribution (ED) of Onatski (2010) and the PC and IC methods of Bai and Ng (2002), among others. We also report the performance of these methods for the purpose of comparison.

TABLE 1 around here.

Table 1 presents the under/over-estimation frequency with complete data. We summarize some important findings from Table 1. First, for DGP 1 with fat-tailed error terms, our CV method tends to outperform all existing methods. Specifically, ED, PC and IC over-estimated more than 300 times out of 1000 for all four combinations of N and T, and GR and ER tend to under-estimate the number of factors. From the performance of these five existing methods, we can hardly observe any pattern of convergence. In contrast, the CV method outperforms these methods by a big margin and shows an obvious pattern of convergence. This indicates the CV method is somewhat robust to error terms with fat tails. Second, for DGP 2 where error terms are well behaved with no serial or cross-sectional dependence, all the methods under investigation show a pattern of convergence, and the CV method with $p = 0.9$ obviously outperforms all the other methods. Third, for DGPs 3–6 where either serial dependence or cross-sectional dependence, or both are present in the error terms, the performance of various methods are similar to that for DGP 2. Among all the methods under study, ER, PC and IC tend to be outperformed by the CV and ED methods. Fourth, in general, the results for the CV method with $p = 0.9$ are better than that with $p = 0.7$. Therefore, we recommend the use of $p = 0.9$ in empirical applications.

When the original data has random missing observations, existing methods such as ED, GR, ER, PC and IC are not directly applicable. We modify the methods in two ways:

(M-1) We replace the missing observations by zeros and obtain the estimators of R based on ED, GR, ER, PC and IC.

(M-2) Following our theoretical analysis in Sections 2-3, we can replace the missing observations by the predicted values to work on the estimated data matrix $\hat{X}$, where

$$\hat{X}_{it} = \begin{cases} X_{it} & \text{if } (i,t) \in \Omega \\ \hat{C}_{R_{\max},it} & \text{if } (i,t) \in \Omega_{\perp} \end{cases}, \quad \ell \geq 1,$$

where $\hat{C}_{R_{\max},it}$ is the EM estimator of C_{it} with $R_{\max}$ factors. For ED, GR and ER, we can find the eigenvalues of $\hat{X}'\hat{X}/(NT)$ and then apply the procedures to these eigenvalues; for PC and IC, we can replace $\hat{\sigma}^2(R)$ in the usual definitions by $\hat{\sigma}^2(R) = \frac{1}{|\Omega|}\sum_{(i,t)\in\Omega}\left[X_{it} - \hat{C}_{R,it}\right]^2$.

TABLE 2 around here.

Table 2 presents the under/over-estimation frequency with incomplete data over 1000 Monte Carlo replications for $q = 0.7$. The case for $q = 0.9$ is reported in Table A1 in the additional online supplement. We consider the three CV methods discussed in Section 3.4, namely, $\widetilde{CV}(R)$, $\widehat{CV}(R)$ with $\hat{C}^{(\ell-1)}_{R_{\max},it}$ used in the ℓth iteration, and $\widehat{CV}(R)$ with $\hat{C}^{(\ell-1)}_{R,it}$ used in the ℓth iteration. As before, we stop the iterations when $\ell = \ell^*$ and denote these three cases as $CV^{(0)}$, $CV^{(\ell^*)}_{R_{\max}}$ and $CV^{(\ell^*)}_{R}$, respectively in Tables 2 and A1, where $CV^{(0)}$ signifies that no iteration is used in the procedure.

We summarize some important findings from Table 2. First, when the proportion of missing observations is large ($q = 0.7$ in Table 2), all the three CV methods yield decreasing percentage of under/over-estimation frequency as either N or T increases, and $CV^{(\ell^*)}_{R}$ and $CV^{(\ell^*)}_{R_{\max}}$ have better finite sample performance than $CV^{(0)}$. Therefore, the iterations to complete some missing observations can help improve the finite performance of the CV method. In general, $CV^{(\ell^*)}_{R}$ and $CV^{(\ell^*)}_{R_{\max}}$ have similar performance with the latter being slightly better. Second, for the other methods, either modification (M-1 or M-2) does not appear promising. The M-1 of ED shows some convergence pattern but the finite sample performance is not as good as either $CV^{(\ell^*)}_{R}$ or $CV^{(\ell^*)}_{R_{\max}}$; the M-2 of ED always over-estimates the number of factors; the M-1 of GR and ER always under-estimates the number of factors, the M-2 of GR and ER is also badly behaved; and both M-1 and M-2 of PC and IC always over-estimate the number of factors. Third, when the proportion of missing observations is small ($q = 0.9$ in Table A1 in the online supplement), the three CV methods all outperform both M-1 and M-2 of existing methods for most cases. Among the other methods, only the M-1 of ED shows a pattern of convergence in all cases.

4.2.2 Estimation of Λ and F

In this subsection, we work on the scenario with random missing observations where $q = 0.7$ and 0.9. We estimate the factors and factor loadings using the method introduced in Section 2 and make inference on factors. For simplicity, we focus on the case where $R_0 = 1$. We set $\rho_{1\varepsilon} = 0.25$, $\rho_{2\varepsilon} = 0.25$ and $\rho_{3\varepsilon} = 0.3$ in DGPs 3, 5 and 6.

TABLE 3 around here.

Tables 3 shows the estimation results for $q = 0.7$ with $\ell = 0$, 5, 20 and ∞, where $\ell = \infty$ corresponds to the final EM estimate. The corresponding results for $q = 0.9$ are reported in Table A2 in the additional online supplement. We also present the results of the oracle estimates for comparison. The first measure of consistency is mean squared error (MSE) of C_{it} and the second is average correlation coefficients between $\{\hat{F}_t\}_{t=1}^T$ and $\{F_t^0\}_{t=1}^T$ which is defined as

$$R^2(\hat{F}) = \frac{\text{trace}(F^{0\prime}\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'F^{0\prime})}{\text{trace}(F^{0\prime}F^{0\prime})}.$$

We summarize some findings from Table 3. First, the MSE becomes smaller and $R^2(\hat{F})$ becomes larger as ℓ increases from zero to 5. But further increases of ℓ does not help much in the reduction of the MSE or the increase of R^2 in general. Second, the EM estimates in the presence of random missing are less efficient than the oracle estimate. This is consistent with Remark 3 in Section 2.3. In fact, despite the presence of serial dependence, or cross-sectional dependence, or both in DGPs 3-6, the MSE of the EM estimator is approximately equal to that of the oracle estimator multiplied by $1/q$ in DGPs 2-6. DGP 1 is an exception because of the violation of the moment conditions on the error terms.

To make inference on F^0, we follow the lead of Bai (2003) and consider the regression model: $F^0 = \hat{F}^{(\ell)}\beta+$ error, where $\ell = 0$ or ℓ^*. Let $\hat{\beta}$ denote the least squares estimator of β. Then the 95% confidence interval for $L'F_t^0$ is

$$[L'\hat{\beta}'\hat{F}_t^{(\ell)} - 1.96\left(L'\hat{\beta}'\hat{\Sigma}_{F_t}^{(\ell)}\hat{\beta}L\right)^{1/2}/\sqrt{N}, L'\hat{\beta}'\hat{F}_t^{(\ell)} + 1.96\left(L'\hat{\beta}'\hat{\Sigma}_{F_t}^{(\ell)}\hat{\beta}L\right)^{1/2}/\sqrt{N}].$$

To estimate the covariance matrix, we consider both the standard covariance matrix estimate based on $\hat{\Gamma}_{1g,t}^{(1)}$ and the robust one based on $\hat{\Gamma}_{1g,t}^{(2)}$ introduced in Section 2.4, which are labeled as "standard" and "robust" in Table 4 below. To obtain $\hat{\Gamma}_{1g,t}^{(2)}$, we need to specify two parameters c_0 and γ_2 : $c_0 = 1$ and $\gamma_2 = 0.5$.

TABLE 4 around here.

Table 4 reports the results of inference on factors when $q = 0.7$ and the corresponding results for $q = 0.9$ are reported in Table A3 in the online supplement. We report both the coverage probability

(CP) and average length (Length) of the 95% confidence intervals when F_t^0 is estimated by the EM method with $\ell = 0$ and ℓ^*. We find some interesting results. First, Table 4 suggests the average length of the EM estimator with no iterations (i.e., $\ell = 0$) is much larger than that with $\ell = \ell^*$. This reflects the efficiency gain from iterations. Second, for DGPs 2–3 where there is no correlation across i for the error terms, both standard and robust covariance estimators provide asymptotically valid inferences. The coverage probabilities are near the nominal coverage probabilities in this case. Third, for DGPs 4 and 6 where there is cross-sectional dependence across i, the coverage probability using standard covariance estimator tends to be smaller than that using robust covariance estimator. This suggests that ignoring the cross-sectional dependence may lead to the underestimation of the standard errors. In general, the confidence intervals constructed using robust covariance estimator have coverage probability near the nominal one. Similar findings hold true for DGPs 1 and 5 that do not satisfy all the assumptions in the paper and are used for robustness check.

5 Empirical Application: Forecasting Macroeconomic Variables

In this section, we show the usefulness of the proposed method by considering factor-augmented regressions to forecast macroeconomic variables. The procedure starts from estimating a set of latent factors using panel data. In practice, some variables have missing observations due to short collection history or lagged publications. A simple and frequently used method to deal with this problem is to delete those units/variables with missing observations to obtain a balanced panel and the PC estimators of latent factors (PC-F). However we may lose some useful information by doing so. To exploit information of predictors with missing observations, we can use the EM estimators to estimate latent factors (EM-F). In our application, we use EM-F or PC-F to forecast macroeconomic variables, respectively. Then we show that EM-F outperforms PC-F in terms of mean squared error.

In particular, we consider the forecasts of U.S. real gross domestic product (RGDP), gross domestic product (GDP), industrial production (IP) and real disposal personal income (RDPI) at 1, 2 and 4 quarter horizons. These four time series are collected from the Federal Reserve Bank website.

5.1 Implementation

We use a panel dataset FRED-QD, which is an unbalanced panel at the quarterly frequency. FRED-QD is a quarterly frequency companion of FRED-MD that is introduced by McCracken and Ng (2016, MN hereafter). The dataset consists of 248 quarterly U.S. indicators from 1959Q1 to 2018Q2. We use 125 time series that are used in Stock and Watson (2002) to estimate the latent factors.

We take 1960Q1 as the start of the sample. Then we lose two periods of observations due to data transformations as in MN and obtain an unbalanced panel with $(T, N) = (236, 125)$. There are 37 variables containing 1594 missing observations in total. Following the lead of MN, we check for outliers in each variable where an outlier is defined as an observation that deviates from the observed sample median by more than 10 times interquantile range. The outliers are removed and treated as missing observations. As a result, the total number of missing observations becomes 1602 ($\hat{q} = 0.946$). All columns of the data matrix X are standardized to have zero mean and unit standard deviation before estimating EM-F. To estimate PC-F, we drop 37 variables with missing observations to obtain a balanced panel with $(T, N) = (236, 88)$. We also standardize the balanced panel before estimating PC-F. We estimate the first factor by PC and EM and use them to do the out-of-sample forecasting.

Next, we consider the forecast based on the following factor-augmented autoregression (FA-AR) models:

$$y_{t+h}^h = \phi_h^{(1)} + \phi_h^{(2)}(L)\hat{F}_t + \phi_h^{(3)}(L)y_t + \varepsilon_{t+h}^h, \ h = 1, 2, \ 4, \tag{5.1}$$

where y_t is one of the four macro-variables (i.e., RGDP, GDP, IP, and RDPI), $\hat{F}_t$ is the estimated factor, $\phi_h^{(1)}$ is the intercept term, L is the lag operator, and $\phi_h^{(2)}(L)$ and $\phi_h^{(3)}(L)$ are finite-order polynomials of the lag operators. For all four variables to be forecasted, we treat them as $I(1)$ series and define the dependent variable as average annualized quarterly growth rate. As an example, for IP, we define

$$y_{t+h}^h = (400/h)\ln(IP_{t+h}/IP_t) \text{ and } y_t = 400\ln(IP_t/IP_{t-1}).$$

All the models are estimated recursively by ordinary least squares (OLS). We use BIC to select the number of autoregressive lags (from 0 to 6) and lags of the first factor (from 1 to 6) in EM-F and PC-F, respectively.

5.2 Forecast results

We consider three out-of-sample periods, namely, 1987Q1-2016Q4, 1997Q1-2016Q4 and 2007Q1-2016Q4. Table 5 reports the mean squared error (MSE) of forecasts using EM-F and its ratio to the MSE associated with autoregression (AR) or FA-AR using PC-F, where the AR model is used with $\hat{F}_t$ absent in (5.1) and the number of lags are also determined by the BIC. Ratios smaller than one are in favor of the method using EM-F. For all the four macroeconomic variables under investigation, the forecasts using EM-F outperforms the forecasts only using autoregression. Therefore, we can conclude that the estimated latent factors contain some predictive power. For Real GDP, IP and RDPI, the forecast using EM-F provides smaller MSE for almost all horizons and periods compared to that using PC-F. For GDP, we can see that the forecasts using EM-F and PC-F have comparable performance. In short, the EM estimation of the factors generally help the out-of-sample forecast of some major macroeconomic variables.

TABLE 5 around here.

6 Conclusion

In this paper we study the asymptotic properties of the EM estimators of factors and factor loadings in an approximate factor model with random missing. Based on the asymptotic results, we also propose a novel cross-validation method to determine the number of factors in factor models with or without random missing observations. Simulations demonstrate the good finite sample performance of the proposed method and empirical applications suggest the usefulness of our method.

The paper can be extended in various directions. First, we only consider random missing and it is possible to extend our method to allow for missing with certain patterns. Second, we focus on a pure approximate factor model and one may consider the extension to the panel data models with multi-factor error structure and random missing values (see, Bai et al. (2015) and Athey et al. (2017)). We are exploring some of these topics in ongoing works.

APPENDIX

A Proofs of the main results in Section 2

In this appendix, we prove the main results in Section 2 by calling upon some technical lemmas whose proofs can be found in the online supplement. For notational simplicity, we will use $\tilde{F}$, $\tilde{\Lambda}$, $\tilde{C}$, $\tilde{D}$, $\tilde{H}$, $\tilde{F}_t, \tilde{\lambda}_i$ and $\tilde{C}_{it}$ to denote $\hat{F}^{(0)}$, $\hat{\Lambda}^{(0)}$, $\hat{C}^{(0)}$, $\hat{D}^{(0)}$, $\hat{H}^{(0)}$, $\hat{F}_t^{(0)}, \hat{\lambda}_i^{(0)}$ and $\hat{C}_{it}^{(0)}$, respectively.

To prove Theorem 2.1, we need the following lemma.

Lemma A.1 *Suppose that Assumptions A.1-A.2 hold. Then* $T^{-1}\tilde{F}' \left(NT\tilde{q}^2\right)^{-1} \tilde{X}\tilde{X}'\tilde{F} = \tilde{D} = D + \delta_{NT}^{-(1-\gamma/2)}$*, where* $\gamma = \gamma_1 \vee \gamma_2$*,* $\tilde{D}$ *is an* $R \times R$ *diagonal matrix consisting of the* R *largest eigenvalues of* $\left(NT\tilde{q}^2\right)^{-1} \tilde{X}\tilde{X}'$*, and* D *is an* $R \times R$ *matrix consisting of the* R *eigenvalues of* $\Sigma_{\Lambda^0}\Sigma_{F^0}$*, arranged in descending order along the diagonal line.*

Proof of Theorem 2.1. From the principal component analysis, we have the identity $\left(NT\tilde{q}^2\right)^{-1} \tilde{X}\tilde{X}'\tilde{F} = \tilde{F}\tilde{D}$. By Lemma A.1 and Assumption A.1, $\tilde{D}$ is asymptotically nonsingular so that we can post-multiply both sides by $\tilde{D}^{-1}$ to obtain $\tilde{F} = \left(NT\tilde{q}^2\right)^{-1} \tilde{X}\tilde{X}'\tilde{F}\tilde{D}^{-1}$. Recall that $\tilde{H} = \left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1} T^{-1}F^{0\prime}\tilde{F}\tilde{D}^{-1}$.

Noting that the (t,i)th element of $\tilde{X}$ is given by $\tilde{X}_{it} = \left(\lambda_i^{0\prime} F_t^0 + \varepsilon_{it}\right) g_{it} = \lambda_i^{0\prime} F_t^0 q + \varepsilon_{it} g_{it} + \lambda_i^{0\prime} F_t^0 \left(g_{it} - q\right),$ we have

$$\begin{aligned}
\tilde{F}_t - \tilde{H}' F_t^0 &= \frac{1}{NT\tilde{q}^2} \tilde{D}^{-1} \sum_{s=1}^T \tilde{F}_s \sum_{i=1}^N \{E\left(\varepsilon_{is}\varepsilon_{it}\right) g_{is} g_{it} + \left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right] g_{is} g_{it} \\
&\quad + F_s^{0\prime} \lambda_i^0 \varepsilon_{it} g_{is} g_{it} + F_t^{0\prime} \lambda_i^0 \varepsilon_{is} g_{is} g_{it} + F_s^{0\prime} \lambda_i^0 \lambda_i^{0\prime} F_t^0 \left(g_{is} - q\right) q \\
&\quad + F_s^{0\prime} \lambda_i^0 \lambda_i^{0\prime} F_t^0 \left(g_{it} - q\right) q + F_s^{0\prime} \lambda_i^0 \lambda_i^{0\prime} F_t^0 \left(g_{is} - q\right)\left(g_{it} - q\right)\} + O_p((NT)^{-1/2}) \\
&\equiv a_{1t} + a_{2t} + \ldots + a_{7t} + O_p((NT)^{-1/2}), \qquad \text{(A.1)}
\end{aligned}$$

where, e.g., $a_{1t} = \frac{1}{NT\tilde{q}^2} \tilde{D}^{-1} \sum_{s=1}^T \tilde{F}_s \sum_{i=1}^N E\left(\varepsilon_{is}\varepsilon_{it}\right) g_{is} g_{it}$ and the first equality used the fact $\tilde{q} - q = O_p((NT)^{-1/2})$. It follows that $T^{-1}\sum_{t=1}^T \left\|\tilde{F}_t - \tilde{H}' F_t^0\right\|^2 \leq 7\sum_{l=1}^7 T^{-1} \sum_{t=1}^T \|a_{lt}\|^2 + O_p((NT)^{-1/2})$ by the Cauchy-Schwarz (CS) inequality. For a_{1t}, we have

$$\begin{aligned}
T^{-1}\sum_{t=1}^T \|a_{1t}\|^2 &\leq \left\|\tilde{D}^{-1}\right\|^2 T^{-1} \sum_{t=1}^T \left\| \frac{1}{T\tilde{q}^2} \sum_{s=1}^T \tilde{F}_s \frac{1}{N} \sum_{i=1}^N E\left(\varepsilon_{is}\varepsilon_{it}\right) g_{is} g_{it} \right\|^2 \\
&\leq \frac{1}{T\tilde{q}^4} \left\|\tilde{D}^{-1}\right\|^2 \frac{1}{T} \sum_{s=1}^T \left\|\tilde{F}_s\right\|^2 \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T \left| \frac{1}{N} \sum_{i=1}^N E\left(\varepsilon_{is}\varepsilon_{it}\right) g_{is} g_{it} \right|^2 \\
&\leq \frac{R}{T\tilde{q}^4} \left\|\tilde{D}^{-1}\right\|^2 \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T \left|\gamma_N\left(s,t\right)\right|^2 = O_P\left(T^{-1}\right),
\end{aligned}$$

where the second inequality follows from the CS inequality and the third inequality follows from the fact that $\frac{1}{T}\sum_{s=1}^T \left\|\tilde{F}_s\right\|^2 = \frac{1}{T}\text{tr}(\tilde{F}'\tilde{F}) = \text{tr}(I_R) = R$ and that $|g_{it}| \leq 1$, and the last equality holds by Assumption A.2. Similarly, for a_{2t}, we have

$$T^{-1}\sum_{t=1}^T \|a_{2t}\|^2 \leq \left\|\tilde{D}^{-1}\right\|^2 T^{-1} \sum_{t=1}^T \left\| \frac{1}{T\tilde{q}^2} \sum_{s=1}^T \tilde{F}_s \zeta_{1g,st} \right\|^2 \leq \frac{R}{T\tilde{q}^4} \left\|\tilde{D}^{-1}\right\|^2 \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T \zeta_{1g,st}^2,$$

where $\zeta_{1g,st} = \frac{1}{N}\sum_{i=1}^N \left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right] g_{is} g_{it}$. Noting that

$$\begin{aligned}
\zeta_{1g,st} &= \frac{1}{N} \sum_{i=1}^N \left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right] \left\{q^2 + (g_{is} - q)q + (g_{it} - q)q + (g_{is} - q)(g_{it} - q)\right\} \\
&\equiv \zeta_{1g,st1} + \zeta_{1g,st2} + \zeta_{1g,st3} + \zeta_{1g,st4},
\end{aligned}$$

where, e.g., $\zeta_{1g,st1} = \frac{1}{N}\sum_{i=1}^N \left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right] q^2$, we have $\zeta_{1g,st}^2 \leq 4\sum_{l=1}^4 \zeta_{1g,stl}^2$. By Assumption A.2,

$$\begin{aligned}
\frac{1}{T}\sum_{s=1}^T \sum_{t=1}^T E\left(\zeta_{1g,st,1}^2\right) &= \frac{q^4}{TN} \sum_{s=1}^T \sum_{t=1}^T E\left[\frac{1}{N^{1/2}} \sum_{i=1}^N \left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right]\right]^2 = O\left(T/N\right), \\
\frac{1}{T}\sum_{s=1}^T \sum_{t=1}^T E\left(\zeta_{1g,st,2}^2\right) &= \frac{q^2}{T} \sum_{s=1}^T \sum_{t=1}^T E\left[\frac{1}{N} \sum_{i=1}^N \left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right]\left(g_{is} - q\right)\right]^2 \\
&= \frac{q^2}{TN^2} \sum_{s=1}^T \sum_{t=1}^T \sum_{i=1}^N E\left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right]^2 = O\left(T/N\right).
\end{aligned}$$

Similarly, we can show that $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}E(\zeta_{1g,st,l}^2)=O\left(T/N\right)$ for $l=3,4$. Then $T^{-1}\sum_{t=1}^{T}\|a_{2t}\|^2=O_P\left(N^{-1}\right)$ by Markov inequality. For a_{3t}, we have

$$T^{-1}\sum_{t=1}^{T}\|a_{3t}\|^2\leq\left\|\tilde{D}^{-1}\right\|^2T^{-1}\sum_{t=1}^{T}\left\|\frac{1}{T\tilde{q}^2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{2g,st}\right\|^2\leq\frac{R}{\tilde{q}^4}\left\|\tilde{D}^{-1}\right\|^2\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{2g,st}^2,$$

where $\zeta_{2g,st}=\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\varepsilon_{it}g_{is}g_{it}$. Using $g_{is}=q+(g_{is}-q)$, we have

$$\begin{aligned}\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{2g,st}^2 &\leq \frac{2}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\varepsilon_{it}g_{it}q\right]^2+\frac{2}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\varepsilon_{it}g_{it}\left(g_{is}-q\right)\right]^2\\ &\equiv 2A_1+2A_2,\ \text{say.}\end{aligned}$$

Noting that $\frac{1}{T}\sum_{t=1}^{T}E\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}\right\|^2=\frac{1}{N^2T}\sum_{t=1}^{T}\sum_{i=1}^{N}E[\left\|\lambda_i^0\right\|^2\varepsilon_{it}^2]E\left(g_{it}^2\right)=O\left(N^{-1}\right)$ under Assumptions A.1(ii) and A.2(i), we have $A_1\leq\frac{1}{T}\sum_{s=1}^{T}\left\|F_s^0\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}\right\|^2=O_P\left(N^{-1}\right)$. Similarly, $A_2=O_P\left(T^{-1}\right)$ by Markov inequality. It follows that $T^{-1}\sum_{t=1}^{T}\|a_{3t}\|^2=O_P\left(N^{-1}+T^{-1}\right)$. Analogously, we can show that $T^{-1}\sum_{t=1}^{T}\|a_{4t}\|^2=O_P\left(N^{-1}+T^{-1}\right)$.

For a_{5t}, we have

$$T^{-1}\sum_{t=1}^{T}\|a_{5t}\|^2\leq\frac{q^2}{\tilde{q}^4}\left\|\tilde{D}^{-1}\right\|^2T^{-1}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\zeta_{3g,st}\right\|^2\leq\frac{Rq^2}{\tilde{q}^4}\left\|\tilde{D}^{-1}\right\|^2\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{3g,st}^2=O_P\left(N^{-1}\right),$$

where $\zeta_{3g,st}=\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\lambda_i^{0\prime}F_t^0\left(g_{is}-q\right)$ and the last equality follows from the Markov inequality and the fact that $\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}E(\zeta_{3g,st}^2)=\frac{q(1-q)}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\left(F_s^{0\prime}\lambda_i^0\lambda_i^{0\prime}F_t^0\right)^2]=O\left(N^{-1}\right)$. Similarly, we can show that $T^{-1}\sum_{t=1}^{T}\|a_{6t}\|^2=O_P\left(N^{-1}\right)$ and $T^{-1}\sum_{t=1}^{T}\|a_{7t}\|^2=O_P\left(N^{-1}\right)$.

In sum, we have shown that $T^{-1}\sum_{t=1}^{T}\left\|\tilde{F}_t-\tilde{H}'F_t^0\right\|^2=O_P\left(N^{-1}+T^{-1}\right)$. ■

To prove Theorem 2.2, we need the following lemma.

Lemma A.2 *Suppose that Assumptions A.1-A.3 hold. Then*

(i) $T^{-1}\tilde{F}'F^0=Q+O_P(\delta_{NT}^{-(1-\gamma/2)})$,
(ii) $\tilde{H}=Q^{-1}+O_P(\delta_{NT}^{-(1-\gamma/2)})$,
(iii) $\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t-\tilde{H}F_t^0)\varepsilon_{it}g_{it}=O_P\left(\delta_{NT}^{-2}\right)$,
(iv) $\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t-\tilde{H}'F_t^0)(\tilde{F}_t-\tilde{H}'F_t^0)'g_{it}=O_P\left(\delta_{NT}^{-2}\right)$,
(v) $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t(\tilde{F}_t-\tilde{H}'F_t^0)'g_{it}=O_P\left(\delta_{NT}^{-2}\right)$,
(vi) $\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t-\tilde{H}'F_t^0)F_t^{0\prime}\tilde{H}(g_{it}-q)=O_P\left(\delta_{NT}^{-2}\right)$,
(vii) $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'(g_{it}-q)=\tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H}(g_{it}-q)+O_P\left(\delta_{NT}^{-2}\right)$,
(viii) $\tilde{H}\tilde{H}'=(\frac{1}{T}F^{0\prime}F^0)^{-1}+O_P\left(\delta_{NT}^{-2}\right)$.

Proof of Theorem 2.2. (i) By the decomposition in (A.1) and Lemma A.1, it suffices to show that $A_{lt}=\tilde{D}a_{lt}=o_P\left(N^{-1/2}\right)$ for $l=1,2,4,5,7$ and $\sqrt{N}\tilde{D}\left(a_{3t}+a_{6t}\right)\xrightarrow{d}N\left(0,Q\Gamma_{g,t}Q'\right)$.

For A_{1t}, we make the following decomposition $A_{1t} = \frac{1}{NT\tilde{q}^2}\sum_{s=1}^T \left(\tilde{F}_s - \tilde{H}'F_s^0\right)\sum_{i=1}^N E\left(\varepsilon_{is}\varepsilon_{it}\right)g_{is}g_{it}$ $+\frac{\tilde{H}'}{\tilde{q}^2}\frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N E\left(\varepsilon_{is}\varepsilon_{it}\right)g_{is}g_{it} \equiv A_{1t,1} + \frac{\tilde{H}'}{\tilde{q}^2}A_{1t,2}$. By the CS inequality and Theorem 2.1,

$$\|A_{1t,1}\| \leq \frac{1}{\tilde{q}^2}\left\{\frac{1}{T}\sum_{s=1}^T \left\|\tilde{F}_s - \tilde{H}'F_s^0\right\|^2\right\}^{1/2}\left\{\frac{1}{T}\sum_{s=1}^T \left\|\frac{1}{N}\sum_{i=1}^N E\left(\varepsilon_{is}\varepsilon_{it}\right)g_{is}g_{it}\right\|^2\right\}^{1/2} = O_P\left(\delta_{NT}^{-1}\right)O_P(T^{-1/2}),$$

where we use the fact that $\frac{1}{T}\sum_{s=1}^T \left\|\frac{1}{N}\sum_{i=1}^N E\left(\varepsilon_{is}\varepsilon_{it}\right)g_{is}g_{it}\right\|^2 \leq \frac{1}{T}\max_t \sum_{s=1}^T |\gamma_N(s,t)|^2 = O\left(T^{-1}\right)$. For $A_{1t,2}$, we have $E\|A_{1t,2}\| \leq \frac{\max_s E\|F_s^0\|}{T}\sum_{s=1}^T |\gamma_N(s,t)| = O\left(T^{-1}\right)$. It follows that $A_{1t,2} = O_P\left(T^{-1}\right)$ and $A_{1t} = O_P\left(\delta_{NT}^{-1}T^{-1/2}\right)$. For A_{2t}, we have

$$A_{2t} = \frac{1}{NT\tilde{q}^2}\sum_{s=1}^T \left(\tilde{F}_s - \tilde{H}'F_s^0\right)\sum_{i=1}^N \chi_{i,st}g_{is}g_{it} + \frac{\tilde{H}'}{\tilde{q}^2}\frac{1}{NT}\sum_{s=1}^T F_s^0\sum_{i=1}^N \chi_{i,st}g_{is}g_{it} \equiv A_{2t,1} + \frac{\tilde{H}'}{\tilde{q}^2}A_{2t,2},$$

where $\chi_{i,st} = \varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)$. As in the analysis of $A_{1t,1}$, we can show that $\|A_{2t,1}\| = O_P\left(\delta_{NT}^{-1}\right)O_P\left(N^{-1/2}\right)$ by the CS inequality and Theorem 2.1. For $A_{2t,2}$, we make the following decomposition

$$A_{2t,2} = \frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \chi_{i,st}\left[q^2 + (g_{is}-q)q + (g_{it}-q)q + (g_{is}-q)(g_{it}-q)\right] \equiv \sum_{l=1}^4 A_{2t,2l}.$$

By straightforward moment calculations, we can show that $E\|A_{2t,2l}\|^2 = O((NT)^{-1})$ under Assumptions A.3(i) and A.1(i) for $l = 1,2,3,4$. Then $A_{2t,2} = O_P((NT)^{-1/2})$. For A_{3t}, we use $g_{is} = q + (g_{is}-q)$ and $\tilde{F}_s = (\tilde{F}_s - \tilde{H}'F_s^0) + \tilde{H}'F_s^0$ to make the following decomposition

$$\begin{aligned} A_{3t} &= \frac{1}{T\tilde{q}^2}\sum_{s=1}^T \tilde{F}_s F_s^{0\prime}\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{is}g_{it} \\ &= \frac{q}{T\tilde{q}^2}\sum_{s=1}^T \tilde{F}_s F_s^{0\prime}\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it} + \frac{1}{T\tilde{q}^2}\sum_{s=1}^T (\tilde{F}_s - \tilde{H}'F_s^0)F_s^{0\prime}\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}(g_{is}-q) \\ &\quad + \frac{\tilde{H}'}{\tilde{q}^2}\left[\frac{1}{T}\sum_{s=1}^T F_s^0F_s^{0\prime}\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}(g_{is}-q)\right] \equiv A_{3t,1} + A_{3t,2} + \frac{\tilde{H}'}{\tilde{q}^2}A_{3t,3}. \end{aligned}$$

By the CS inequality and Theorem 2.1,

$$\|A_{3t,2}\| \leq \frac{1}{\tilde{q}^2}\left\{\frac{1}{T}\sum_{s=1}^T \left\|\tilde{F}_s - \tilde{H}'F_s^0\right\|^2\right\}^{1/2}\left\{\frac{1}{T}\sum_{s=1}^T \left\|F_s^{0\prime}\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}(g_{is}-q)\right\|^2\right\}^{1/2} = O_P(\delta_{NT}^{-1})O_P(N^{-1/2}),$$

where we use the fact that $\frac{1}{T}\sum_{s=1}^T E\left\|F_s^{0\prime}\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}(g_{is}-q)\right\|^2 = O\left(N^{-1}\right)$. For $A_{3t,3}$, it is easy to verify that $E\left(A_{3t,3}\right) = O\left(T^{-1}\right)$ and $E\|A_{3t,3}\|^2 = O((NT)^{-1} + T^{-2})$. Then $A_{3t,3} = O_P(\delta_{NT}^{-1}T^{-1/2})$ and $A_{3t} = \frac{1}{T}\tilde{F}'F^0\frac{1}{Nq}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it} + O_P\left(\delta_{NT}^{-2}\right)$, where we use the fact that $\tilde{q} = q + O_P((NT)^{-1/2})$. For

A_{4t}, we apply $g_{it} = q + (g_{it} - q)$ and $\tilde{F}_s = (\tilde{F}_s - \tilde{H}' F_s^0) + \tilde{H}' F_s^0$ to make the following decomposition

$$\begin{aligned} A_{4t} &= \frac{q}{NT\tilde{q}^2} \sum_{s=1}^{T} \left(\tilde{F}_s - \tilde{H}' F_s^0\right) \sum_{i=1}^{N} \lambda_i^{0\prime} \varepsilon_{is} g_{is} F_t^0 + \frac{1}{NT\tilde{q}^2} \sum_{s=1}^{T} \left(\tilde{F}_s - \tilde{H}' F_s^0\right) \sum_{i=1}^{N} \lambda_i^{0\prime} \varepsilon_{is} g_{is} \left(g_{it} - q\right) F_t^0 \\ &\quad + \frac{q\tilde{H}'}{\tilde{q}^2} \left[\frac{1}{NT} \sum_{s=1}^{T} F_s^0 \sum_{i=1}^{N} \lambda_i^{0\prime} \varepsilon_{is} g_{is}\right] F_t^0 + \frac{\tilde{H}'}{\tilde{q}^2} \left[\frac{1}{NT} \sum_{s=1}^{T} F_s^0 \sum_{i=1}^{N} \lambda_i^{0\prime} \varepsilon_{is} g_{is} \left(g_{it} - p\right)\right] F_t^0 \\ &\equiv A_{4t,1} F_t^0 + A_{4t,2} F_t^0 + \frac{q\tilde{H}'}{\tilde{q}^2} A_{4t,3} F_t^{0\prime} + \frac{\tilde{H}'}{\tilde{q}^2} A_{4t,4} F_t^0. \end{aligned}$$

For $A_{4t,1}$ and $A_{4t,2}$, we can readily use the CS inequality and Theorem 2.1 to show that $A_{4t,1} = O_P\left(\delta_{NT}^{-1} N^{-1/2}\right)$ and $A_{4t,2} = O_P\left(\delta_{NT}^{-2}\right)$. For $A_{4t,3}$ we apply $g_{it} = q + (g_{it} - q)$, the CS inequality, and Assumption A.3(ii) to obtain $E\left\|A_{4t,3}\right\|^2 \leq \frac{2}{N^2T^2} E|| \sum_{s=1}^{T} \sum_{i=1}^{N} F_s^0 \lambda_i^{0\prime} \varepsilon_{is} q||^2 + \frac{2}{N^2T^2} E|| \sum_{s=1}^{T} \sum_{i=1}^{N} F_s^0 \lambda_i^{0\prime} \varepsilon_{is} \times \left(g_{is} - q\right)||^2 = O\left((NT)^{-1}\right) + O\left((NT)^{-1}\right) = O\left((NT)^{-1}\right)$. It follows that $A_{4t,3} = O_P((NT)^{-1/2})$. Similarly, $A_{4t,4} = O_P((NT)^{-1/2})$. Then $A_{4t} = O_P(\delta_{NT}^{-2})$.

For A_{5t}, we use $\tilde{F}_s = (\tilde{F}_s - \tilde{H}' F_s^0) + \tilde{H}' F_s^0$ to obtain

$$\begin{aligned} A_{5t} &= \frac{q}{\tilde{q}^2} \left[\frac{1}{T} \sum_{s=1}^{T} (\tilde{F}_s - \tilde{H}' F_s^0) F_s^{0\prime} \frac{1}{N} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} \left(g_{is} - q\right)\right] F_t^0 + \frac{q\tilde{H}'}{\tilde{q}^2} \left[\frac{1}{NT} \sum_{s=1}^{T} F_s^0 F_s^{0\prime} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} \left(g_{is} - q\right)\right] F_t^0 \\ &\equiv \frac{q}{\tilde{q}^2} A_{5t,1} F_t^0 + \frac{p\tilde{H}'}{\tilde{q}} A_{5t,2} F_t^{0\prime}. \end{aligned}$$

By the CS inequality and Theorem 2.1,

$$\left\|A_{5t,1}\right\| = \left\{\frac{1}{T} \sum_{s=1}^{T} \left\|\tilde{F}_s - \tilde{H}' F_s^0\right\|^2\right\}^{1/2} \left\{\frac{1}{T} \sum_{s=1}^{T} \left\|F_s^{0\prime} \frac{1}{N} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} \left(g_{is} - q\right)\right\|^2\right\}^{1/2} = O_P(\delta_{NT}^{-1} N^{-1/2}),$$

where we use the fact that $\frac{1}{T} \sum_{s=1}^{T} E\left\|F_s^{0\prime} \frac{1}{N} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} \left(g_{is} - q\right)\right\|^2 = O\left(N^{-1}\right)$. Similarly, $E\left\|A_{5t,2}\right\|^2 = \frac{q(1-q)}{(NT)^2} \sum_{s=1}^{T} \sum_{i=1}^{N} E\left\|F_s^0 F_s^{0\prime} \lambda_i^0 \lambda_i^{0\prime}\right\|^2 = O((NT)^{-1})$ under Assumption A.1(iii).Then $A_{5t,2} = O_P((NT)^{-1/2})$ and $A_{5t} = O_P\left(\delta_{NT}^{-2}\right)$. For A_{6t}, we apply the fact that $\tilde{q} = q + O_P((NT)^{-1/2})$ to obtain

$$A_{6t} = \frac{1}{T} \tilde{F}' F^0 \frac{q}{\sqrt{N}\tilde{q}^2} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} F_t^0 \left(g_{it} - q\right) = \frac{1}{T} \tilde{F}' F^0 \frac{1}{\sqrt{N}q} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} F_t^0 \left(g_{it} - q\right) + O_P((NT)^{-1/2}).$$

For A_{7t}, we have $A_{7t} = [\frac{1}{T\tilde{q}^2} \sum_{s=1}^{T} \left(\tilde{F}_s - \tilde{H}' F_s^0\right) F_s^{0\prime} \frac{1}{N} \sum_{i=1}^{N} \alpha_{i,st}] F_t^0 + \frac{\tilde{H}'}{\tilde{q}^2} [\frac{1}{T} \sum_{s=1}^{T} F_s F_s^{0\prime} \frac{1}{N} \sum_{i=1}^{N} \alpha_{i,st}] F_t^0 \equiv A_{7t,1} F_t^0 + \frac{\tilde{H}'}{\tilde{q}^2} A_{7t,2} F_t^{0\prime}$, where $\alpha_{i,st} = \lambda_i^0 \lambda_i^{0\prime} \left(g_{is} - q\right) \left(g_{it} - q\right)$. As in the analysis of A_{5t}, we can readily show that $A_{7t,1} = O_P\left(\delta_{NT}^{-2}\right)$ and $A_{7t,2} = O_P((NT)^{-1/2})$. Then $A_{7t,1} = O_P(\delta_{NT}^{-2})$.

In sum, we have

$$\sqrt{N}(\tilde{F}_t - \tilde{H}' F_t^0) = \tilde{D}^{-1} \frac{1}{T} \tilde{F}' F^0 \frac{1}{\sqrt{N}q} \sum_{i=1}^{N} \lambda_i^0 \left[\varepsilon_{it} g_{it} + \lambda_i^{0\prime} F_t^0 \left(g_{it} - q\right)\right] + O_P(N^{1/2} \delta_{NT}^{-2}). \qquad \text{(A.2)}$$

By Assumption A.4(i), $\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}\xrightarrow{d}N\left(0,\Gamma_{1g,t}\right)$. Let $\omega\in\mathbb{R}^R$ be a nonrandom vector with $\|\omega\|=1$. Let $\varphi_{it}=\frac{1}{\sqrt{Nq}}\omega'\lambda_i^0\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)$ and $\mathcal{G}_{Ni}^t=\sigma\left(\{g_{jt},j\leq i),\Lambda^0,F_t^0\right)$, the sigma-field generated from $\{g_{jt},\ j\leq i\}$ and (Λ^0,F^0). Let $\mathcal{G}^t=\sigma\left(\cup_{N=1}^{\infty}\mathcal{G}_{NN}^t\right)$. By the independence of g_{it} along the i-dimension, we have $E\left(\varphi_{it}|\mathcal{F}_{Nt,i-1,t}\right)=0$ and

$$\sum_{i=1}^{N}E\left(\varphi_{it}^2|\mathcal{G}_{N,i-1}^t\right)=\frac{1-q}{Nq}\sum_{i=1}^{N}\left(\omega'\lambda_i^0\lambda_i^{0\prime}F_t^0\right)^2=\omega'\left(\frac{1-q}{Nq}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}\left(F_t^{0\prime}\lambda_i^0\right)^2\right)\omega\xrightarrow{p}\omega'\Gamma_{2g,t}\omega.$$

Let $\epsilon=\frac{4}{\lambda_2}-4$. Then by Assumption A.1(ii), $\sum_{i=1}^{N}E(|\varphi_{it}|^{2+\epsilon/2}|\mathcal{G}_{N,i-1}^t)\leq\frac{\|F_t^0\|^{2+\epsilon/2}}{N^{\delta/2}}\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^{4+\epsilon}\xrightarrow{p}$ 0, which is sufficient for the conditional Lindeberg condition in Häusler and Luschgy (2015). Then by the stable martingale central limit theorem (e.g., Theorem 6.1 in Häusler and Luschgy (2015)), we have

$$\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)=\sum_{i=1}^{N}\varphi_{it}\to N\left(0,\Gamma_{2g,t}\right)\quad\mathcal{G}^t\text{-stably as }N\to\infty,$$

where $\Gamma_{2g,t}$ is $\mathcal{G}_{\infty}^t$ measurable. Noting that $\text{Cov}(\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it},\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right))=\frac{1}{Nq^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E(\lambda_i^0\lambda_j^{0\prime}\varepsilon_{it}\lambda_j^{0\prime}F_t^0)E\left[g_{it}\left(g_{jt}-q\right)\right]=\frac{1-q}{Nq}\sum_{i=1}^{N}E\left[\lambda_i^0\lambda_i^{0\prime}\varepsilon_{it}\lambda_i^{0\prime}F_t^0\right]=0$ by the i.i.d. of g_{it}, the independence between $\{g_{it}\}$ and $\left\{\Lambda^0,F^0,\varepsilon\right\}$, and Assumption A.2(i), we also have

$$\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it}+\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]\to N\left(0,\Gamma_{1g,t}+\Gamma_{2g,t}\right)\ \mathcal{G}^t\text{-stably as }N\to\infty.$$

Then by Lemmas A.1(i) and A.2(i) and Corollary 6.3 in Häusler and Luschgy (2015), we have

$$\begin{aligned}\sqrt{N}(\tilde{F}_t-\tilde{H}'F_t^0)&=\tilde{D}^{-1}\frac{1}{T}\tilde{F}'F^0\frac{1}{\sqrt{Nq}}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it}+\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]+O_P\left(N^{1/2}\delta_{NT}^{-2}\right)\\&\to N\left(0,D^{-1}Q\left(\Gamma_{1g,t}+\Gamma_{2g,t}\right)Q'D^{-1}\right)\ \mathcal{G}^t\text{-stably as }(N,T)\to\infty.\end{aligned}$$

This completes the proof of (i).

(ii) Noting that $\tilde{\Lambda}'=\frac{1}{T\tilde{q}}\tilde{F}'\tilde{X}$, $\tilde{X}=(F^0\Lambda^{0\prime}+\varepsilon)\circ G$, and $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'=I_R$, we have

$$\begin{aligned}\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^0&=\frac{1}{T\tilde{q}}\sum_{t=1}^{T}\tilde{F}_t\left(\varepsilon_{it}+F_t^{0\prime}\lambda_i^0\right)g_{it}-\tilde{H}^{-1}\lambda_i^0\\&=\frac{1}{T\tilde{q}}\sum_{t=1}^{T}\tilde{F}_t\left\{\varepsilon_{it}+\left[\tilde{F}_t'\tilde{H}^{-1}+\left(F_t^{0\prime}-\tilde{F}_t'\tilde{H}^{-1}\right)\right]\lambda_i^0\right\}g_{it}-\tilde{H}^{-1}\lambda_i^0\\&=\frac{\tilde{H}'}{T\tilde{q}}\sum_{t=1}^{T}F_t^0\varepsilon_{it}g_{it}+\frac{1}{T\tilde{q}}\sum_{t=1}^{T}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\varepsilon_{it}g_{it}+\frac{1}{T\tilde{q}}\sum_{t=1}^{T}\tilde{F}_t\left(\tilde{H}'F_t^0-\tilde{F}_t\right)'\tilde{H}^{-1}\lambda_i^0g_{it}\\&\quad+\frac{1}{T\tilde{q}}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'\tilde{H}^{-1}\lambda_i^0(g_{it}-q)+\frac{q-\tilde{q}}{\tilde{q}}\tilde{H}^{-1}\lambda_i^0\\&\equiv B_{1i}+B_{2i}+B_{3i}+B_{4i}+B_{5i},\ \text{say.}\end{aligned}$$

By Lemma A.2(ii)-(v) and (vii), $\sqrt{T}B_{1i} = \tilde{H}'\frac{1}{\sqrt{T}q}\sum_{t=1}^{T} F_t^0\varepsilon_{it}g_{it} + o_P(1)$ and $\sqrt{T}B_{li} = O_P\left(T^{1/2}\delta_{NT}^{-2}\right) = o_P(1)$ for $l = 2, 3$. By Lemma A.2(ii) and (vii), $\sqrt{T}B_{4i} = \tilde{H}'\frac{1}{\sqrt{T}q}\sum_{t=1}^{T} F_t^0 F_t^{0\prime}\lambda_i^0(g_{it}-q) + O_P\left(T^{1/2}\delta_{NT}^{-2}\right)$. Noting that $\tilde{q} - q = O_P((NT)^{-1/2})$, we have $\sqrt{T}B_{5i} = O_P(N^{-1/2})$. Therefore we have shown that

$$\sqrt{T}\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) = \tilde{H}'\frac{1}{\sqrt{T}q}\sum_{t=1}^{T} F_t^0\left[\varepsilon_{it}g_{it} + F_t^{0\prime}\lambda_i^0(g_{it}-q)\right] + O_P(T^{1/2}\delta_{NT}^{-2}). \quad \text{(A.3)}$$

Recall that $\mathcal{G}_{Tt}^i = \sigma\left(\{g_{is}, s \leq t), \lambda_i^0, F^0\right)$ denotes the sigma-field generated from $\{\{g_{is}, s \leq t)\}$ and (λ_i^0, F^0) and $\mathcal{G}^i = \sigma\left(\cup_{T=1}^{\infty}\mathcal{G}_{TT}^i\right)$. Following the analysis at the end of the proof of part (i), we can show that

$$\sqrt{T}\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) \to N\left(0, (Q')^{-1}\left(\Phi_{1g,t} + \Phi_{2g,t}\right)(Q)^{-1}\right) \quad \mathcal{G}^i\text{-stably as } N \to \infty,$$

where we use Lemma A.2(ii) and the fact $\text{Cov}(\frac{1}{\sqrt{T}q}\sum_{t=1}^{T} F_t^0\varepsilon_{it}g_{it}, \frac{1}{\sqrt{T}q}\sum_{t=1}^{T} F_t^0F_t^{0\prime}\lambda_i^0(g_{it}-q)) = 0$.

(iii) Let $\varsigma_{it} = \varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0(g_{it}-q)$. By the proofs of (i) and (ii),

$$\begin{aligned}
\tilde{C}_{it} - C_{it}^0 &= \lambda_i^{0\prime}(\tilde{H}')^{-1}(\tilde{F}_t - \tilde{H}'F_t^0) + \tilde{F}_t'(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0) \\
&= \lambda_i^{0\prime}(\tilde{H}')^{-1}(\tilde{F}_t - \tilde{H}'F_t^0) + F_t^{0\prime}\tilde{H}(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0) + O_P((NT)^{-1/2}) \\
&= \lambda_i^{0\prime}(\tilde{H}')^{-1}\tilde{D}^{-1}(\frac{1}{T}\tilde{F}'F^0)\frac{1}{Nq}\sum_{i=1}^{N}\lambda_i^0\varsigma_{it} + F_t^{0\prime}\tilde{H}\tilde{H}'\frac{1}{Tq}\sum_{t=1}^{T}F_t^0\varsigma_{it} + O_P\left(\delta_{NT}^{-2}\right) \\
&= \lambda_i^{0\prime}(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\frac{1}{Nq}\sum_{i=1}^{N}\lambda_i^0\varsigma_{it} + F_t^{0\prime}(\frac{1}{T}F^{0\prime}F^0)^{-1}\frac{1}{Tq}\sum_{t=1}^{T}F_t^0\varsigma_{it} + O_P\left(\delta_{NT}^{-2}\right),
\end{aligned}$$

where the second equality follows from the fact that $\tilde{F}_t - \tilde{H}'F_t^0 = O_P\left(N^{-1/2}\right)$ and $\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0 = O_P\left(T^{-1/2}\right)$, the third equality holds by the results in (i) and (ii), and fourth equality holds because $(\tilde{H}')^{-1}\tilde{D}^{-1}\frac{1}{T}\tilde{F}'F^0 = (\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}$ by the definition of $\tilde{H}$ and $\tilde{H}\tilde{H}' = (\frac{1}{T}F^{0\prime}F^0)^{-1} + O_P\left(\delta_{NT}^{-2}\right)$ by Lemma A.2(viii). Following the proof of Theorem 3 in Bai (2003), we can readily show that $(\frac{1}{N}\Sigma_{1it} + \frac{1}{T}\Sigma_{2it})^{-1/2}\left(\tilde{C}_{it} - C_{it}^0\right) \xrightarrow{d} N(0,1)$, where $\Sigma_{1it} = \lambda_i^{0\prime}\Sigma_{\Lambda^0}^{-1}\Gamma_{g,t}\Sigma_{\Lambda^0}^{-1}\lambda_i^0$ and $\Sigma_{2it} = F_t^{0\prime}\Sigma_{F^0}^{-1}\Phi_{g,i}\Sigma_{F^0}^{-1}F_t^0$. ■

To prove Theorems 2.3-2.4, we introduce some notations. Recall that $\hat{H}^{(\ell)} = \left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1} \times T^{-1}F^{0\prime}\hat{F}^{(\ell)}\hat{D}^{(\ell)-1}$. Define

$$\begin{aligned}
\hat{\phi}_{F,t}^{(0)} &= \hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0\frac{1}{Nq}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0(g_{it}-q)\right], \\
\hat{\phi}_{\Lambda,i}^{(0)} &= \hat{H}^{(0)\prime}\frac{1}{Tq}\sum_{t=1}^{T}F_t^0\left[\varepsilon_{it}g_{it} + F_t^{0\prime}\lambda_i^0(g_{it}-q)\right], \\
\hat{\phi}_{F,t}^{(\ell)} &= \hat{D}^{(\ell)-1}\frac{1}{T}\hat{F}^{(\ell)\prime}F^0\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}^{(\ell)} \text{ for } \ell \geq 1, \text{ and} \\
\hat{\phi}_{\Lambda,i}^{(\ell)} &= \hat{H}^{(\ell)\prime}\frac{1}{T}\sum_{t=1}^{T}F_t^0\varepsilon_{it}^{(\ell)} \text{ for } \ell \geq 1,
\end{aligned}$$

where $\varepsilon_{it}^{(\ell)}$ is defined sequentially in (A.6) below, and $\hat{\phi}_{F,t}^{(\ell)}$ and $\hat{\phi}_{\Lambda,i}^{(\ell)}$ denote the leading influence functions of $\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime} F_t^0$ and $\hat{\lambda}_i^{(\ell)} - (\hat{H}^{(\ell)})^{-1}\lambda_i^0$, respectively. Let $\hat{r}_{F,t}^{(\ell)} = \hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime} F_t^0 - \hat{\phi}_{F,t}^{(\ell)}$ and $\hat{r}_{\Lambda,i}^{(\ell)} = \hat{\lambda}_i^{(\ell)} - (\hat{H}^{(\ell)})^{-1}\lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(\ell)}$ where $\ell \geq 0$. Then

$$\hat{\lambda}_i^{(\ell)\prime} \hat{F}_t^{(\ell)} = \left[(\hat{H}^{(\ell)})^{-1}\lambda_i^0 + \hat{\phi}_{\Lambda,i}^{(\ell)} + \hat{r}_{\Lambda,i}^{(\ell)} \right]' \left[\hat{H}^{(\ell)\prime} F_t^0 + \hat{\phi}_{F,t}^{(\ell)} + \hat{r}_{F,t}^{(\ell)} \right] = \lambda_i^{0\prime} F_t^0 + \eta_{it}^{(\ell)}, \tag{A.4}$$

where $\eta_{it}^{(\ell)} = \eta_{1,it}^{(\ell)} + \eta_{2,it}^{(\ell)}$,

$$\begin{aligned} \eta_{1,it}^{(\ell)} &= F_t^{0\prime} \hat{H}^{(\ell)} \hat{\phi}_{\Lambda,i}^{(\ell)} + \lambda_i^{0\prime} (\hat{H}^{(\ell)\prime})^{-1} \hat{\phi}_{F,t}^{(\ell)} + \lambda_i^{0\prime} (\hat{H}^{(\ell)\prime})^{-1} \hat{r}_{F,t}^{(\ell)} + F_t^{0\prime} \hat{H}^{(\ell)\prime} \hat{r}_{\Lambda,i}^{(\ell)}, \text{ and} \\ \eta_{2,it}^{(\ell)} &= \hat{\phi}_{\Lambda,i}^{(\ell)\prime} \hat{\phi}_{F,t}^{(\ell)} + \hat{\phi}_{\Lambda,i}^{(\ell)\prime} \hat{r}_{F,t}^{(\ell)} + \hat{\phi}_{F,t}^{(\ell)\prime} \hat{r}_{\Lambda,i}^{(\ell)} + \hat{r}_{\Lambda,i}^{(\ell)\prime} \hat{r}_{F,t}^{(\ell)}. \end{aligned} \tag{A.5}$$

Let $\bar{g}_{it} = 1 - g_{it}$ and

$$\varepsilon_{it}^{(\ell)} = \varepsilon_{it} g_{it} + \eta_{it}^{(\ell-1)} \bar{g}_{it}, \ \ \ell \geq 1. \tag{A.6}$$

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

$$\hat{X}_{it}^{(\ell)} = \left(\lambda_i^{0\prime} F_t^0 + \varepsilon_{it} \right) g_{it} + \hat{\lambda}_i^{(\ell-1)\prime} \hat{F}_t^{(\ell-1)} \bar{g}_{it} = \left(\lambda_i^{0\prime} F_t^0 + \varepsilon_{it} \right) g_{it} + \left(\lambda_i^{0\prime} F_t^0 + \eta_{it} \right) \bar{g}_{it} = \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}^{(\ell)}. \tag{A.7}$$

This expression will be used repeatedly in the following derivation.

The following three lemmas are used in the proofs of Theorems 2.3 and 2.4. When Lemmas A.3-A.5 hold for $\ell = 1$, Theorems 2.3 and 2.4 also hold for $\ell = 1$. With the results in Lemmas A.3-A.5 and Theorems 2.3 and 2.4 for $\ell = 1$, we can show that they also hold for $\ell = 2$. This procedure is repeated until convergence which requires ℓ to be at most of order $\ln N$.

Lemma A.3 *Suppose that Assumptions A.1-A.5 hold. Then for any $\ell \geq 1$ we have*

(i) $\max_t \left\| \hat{\phi}_{F,t}^{(\ell-1)} \right\| = O_P((N/\ln N)^{-1/2})$ *and* $\max_i \left\| \hat{\phi}_{\Lambda,i}^{(\ell-1)} \right\| = O_P((T/\ln T)^{-1/2})$,

(ii) $\max_t \left\| \hat{r}_{F,t}^{(\ell-1)} \right\| = O_P(T^{\gamma_1/4} \delta_{NT}^{-2} \ln T + T^{-1+3\gamma_1/4})$ *and* $\max_i \left\| \hat{r}_{\Lambda,i}^{(\ell-1)} \right\| = O_P(N^{\gamma_2/4} \delta_{NT}^{-2} \ln N)$,

(iii) $\max_{i,t} \left\| \eta_{1,it}^{(\ell-1)} \right\| = O_P(\delta_{NT}^{-1+\gamma/2} \ln N)$ *and* $\max_{i,t} \left\| \eta_{2,it}^{(\ell-1)} \right\| = O_P\left(\delta_{NT}^{-2} \ln N \right)$,

(iv) $\max_t \left\| \frac{1}{N} \sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(\ell-1)} \varepsilon_{it} g_{it} \right\| = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2} \ln N \right)$, $\left\| \frac{1}{N} \sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| = O_P(T^{-1+\gamma_1/4}$ $+ N^{\gamma_2/4} \delta_{NT}^{-2} \ln N)$, *and* $\max_t \left\| \frac{1}{N} \sum_{i=1}^N \hat{r}_{\Lambda,i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2} \ln N \right)$,

(v) $\max_i \left\| \frac{1}{T} \sum_{t=1}^T \hat{\phi}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2} \ln N + N^{-1+\gamma_2/2} \right)$ *and* $\max_i \left\| \frac{1}{T} \sum_{t=1}^T \hat{r}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2} \ln N \right)$,

(vi) $\max_t \frac{1}{N} \sum_{i=1}^N \left\| \eta_{it}^{(\ell-1)} \right\|^2 = O_P(T^{-1+\gamma_1/2} + N^{-1} \ln N)$ *and* $\max_i \frac{1}{T} \sum_{t=1}^T \left\| \eta_{it}^{(\ell-1)} \right\|^2 = O_P(N^{-1+\gamma_2/2}$ $+ T^{-1} \ln N)$,

(vii) $\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (1 + \left\| F_t^0 \right\|^2)(\eta_{it}^{(\ell-1)})^2 = O_P\left(\delta_{NT}^{-2} \right)$,

(viii) $\frac{1}{NT} \sum_{s=1}^T F_s^0 \sum_{i=1}^N \lambda_i^{0\prime} \eta_{is}^{(\ell-1)} \bar{g}_{is} = O_P\left(\delta_{NT}^{-2} \ln N \right)$,

(ix) $\max_t \left\| \frac{1}{NT} \sum_{s=1}^T F_s^0 \sum_{i=1}^N \eta_{it}^{(\ell-1)} \bar{g}_{it} \varepsilon_{is} g_{is} \right\| = O_P(T^{-1+\gamma_1/4} + (NT/\ln N)^{-1/2})$,

(x) $\max_t \left\| \frac{1}{NT} \sum_{s=1}^T F_s^0 \sum_{i=1}^N \varepsilon_{it} g_{it} \eta_{is}^{(\ell-1)} \bar{g}_{is} \right\| = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2} \ln N \right)$.

Lemma A.4 *Suppose that Assumptions A.1-A.5 hold. Then for any $\ell \geq 1$ we have*

(i) $T^{-1}\hat{F}^{(\ell)\prime}(NT)^{-1}\hat{X}^{(\ell)}\hat{X}^{(\ell)\prime}\hat{F}^{(\ell)} = \hat{D}^{(\ell)} = D + O_P\left(\delta_{NT}^{-1}\ln N\right)$,

(ii) $T^{-1}\hat{F}^{(\ell)\prime}F^0 = Q + O_P\left(\delta_{NT}^{-1}\ln N\right)$,

(iii) $\hat{H}^{(\ell)} = Q^{-1} + O_P\left(\delta_{NT}^{-1}\ln N\right)$,

(iv) $\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0)F_t^{0\prime} = O_P\left(\delta_{NT}^{-2}\right)$,

(v) $\max_i \left\|\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0)\varepsilon_{it}^{(\ell)}\right\| = O_P\left(N^{-1/2+\gamma_2/4}\delta_{NT}^{-1} + \delta_{NT}^{-2}\ln N\right)$.

Lemma A.5 *Suppose that Assumptions A.1-A.5 hold. Then*

(i) $\hat{\phi}_{F,t}^{(\ell)} = D^{-1}Q\beta_{F,t} + (1-q)\hat{\phi}_{F,t}^{(\ell-1)} + O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+\gamma_1/4}\right)$,

(ii) $\hat{\phi}_{\Lambda,i}^{(\ell)} = (Q')^{-1}\beta_{\Lambda,i} + (1-q)\,\hat{\phi}_{\Lambda,i}^{(\ell-1)} + O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N + N^{-1+3\gamma_2/4}\right)$,

where $\beta_{F,t} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}$, *and* $\beta_{\Lambda,i} = \frac{1}{T}\sum_{t=1}^{T}F_t^0\varepsilon_{it}g_{it}$.

The proof of Theorem 2.4 below suggests that $\hat{\phi}_{F,t}^{(\ell)}$ and $\hat{\phi}_{\Lambda,i}^{(\ell)}$ are associated with the leading influence functions of $\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0$ and $\hat{\lambda}_i^{(\ell)} - (\hat{H}^{(\ell)})^{-1}\lambda_i^0$, respectively.

Proof of Theorem 2.3. The proof follows closely from that of Theorem 2.1 and we only outline the main differences. From the identity $\hat{F}^{(\ell)} = (NT)^{-1}\hat{X}^{(\ell)}\hat{X}^{(\ell)\prime}\hat{F}^{(\ell)}\hat{D}^{(\ell)-1}$ where $\hat{D}^{(\ell)}$ is asymptotically nonsingular by Lemma A.4(i), we have by (A.7),

$$\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0 = \frac{1}{NT}\hat{D}^{(\ell)-1}\sum_{s=1}^{T}\hat{F}_s^{(\ell)}\sum_{i=1}^{N}\left\{\varepsilon_{it}^{(\ell)}\varepsilon_{is}^{(\ell)} + F_s^{0\prime}\lambda_i^0\varepsilon_{it}^{(\ell)} + F_t^{0\prime}\lambda_i^0\varepsilon_{is}^{(\ell)}\right\} \equiv \hat{a}_{1t}^{(\ell)} + \hat{a}_{2t}^{(\ell)} + \hat{a}_{3t}^{(\ell)}. \quad \text{(A.8)}$$

Then $T^{-1}\sum_{t=1}^{T}\left\|\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0\right\|^2 \leq 3\sum_{l=1}^{3}T^{-1}\sum_{t=1}^{T}(\hat{a}_{lt}^{(\ell)})^2$ by the CS inequality. For $\hat{a}_{1t}^{(\ell)}$, using $\varepsilon_{it}^{(\ell)} = \varepsilon_{it}g_{it} + \eta_{it}^{(\ell-1)}\bar{g}_{it}$ and the CS inequality, we have

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\left\|\hat{D}^{(\ell)}\hat{a}_{1t}^{(\ell)}\right\|^2 &\leq 4T^{-1}\sum_{t=1}^{T}\Bigg\{\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(\ell)}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}g_{it}\varepsilon_{is}g_{is}\right\|^2 + \left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(\ell)}\frac{1}{N}\sum_{i=1}^{N}\eta_{it}^{(\ell-1)}\bar{g}_{it}\eta_{is}^{(\ell-1)}\bar{g}_{is}\right\|^2 \\
&\quad + \left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(\ell)}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}g_{it}\eta_{is}^{(\ell-1)}\bar{g}_{is}\right\|^2 + \left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(\ell)}\frac{1}{N}\sum_{i=1}^{N}\eta_{it}^{(\ell-1)}\bar{g}_{it}\varepsilon_{is}g_{is}\right\|^2\Bigg\} \\
&\equiv 4(\hat{A}_{1,1} + \hat{A}_{1,2} + \hat{A}_{1,3} + \hat{A}_{1,4}),
\end{aligned}$$

where we suppress the dependence of $\hat{A}_1$'s on ℓ. Following the analyses of $T^{-1}\sum_{t=1}^{T}\|a_{1t}\|^2$ and $T^{-1}\sum_{t=1}^{T}\|a_{2t}\|^2$ in the proof of Theorem 2.1, we can readily show that $\hat{A}_{1,1} = O_P\left(\delta_{NT}^{-2}\right)$. For $\hat{A}_{1,2}$ and $\hat{A}_{1,3}$, we can apply the fact $\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}/T = I_R$, the CS inequality, and Lemma A.3(vii) to obtain

$$\begin{aligned}
\hat{A}_{1,2} &\leq \frac{R}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_{it}^{(\ell-1)}\bar{g}_{it}\eta_{is}^{(\ell-1)}\bar{g}_{is}\right)^2 \leq R\left\{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(\eta_{it}^{(\ell-1)})^2\right\}^2 = O_P\left(\delta_{NT}^{-4}\right), \text{ and} \\
\hat{A}_{1,3} &\leq \frac{R}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}g_{it}\eta_{is}^{(\ell-1)}\bar{g}_{is}\right)^2 \leq \frac{R}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}g_{it}|^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}(\eta_{is}^{(\ell-1)})^2 = O_P\left(\delta_{NT}^{-2}\right).
\end{aligned}$$

Analogously, $\hat{A}_{1,4} = O_P\left(\delta_{NT}^{-2}\right)$. It follows that $\hat{A}_1 = O_P\left(\delta_{NT}^{-2}\right)$.

For $\hat{a}_{2t}^{(\ell)}$, we have

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\left\|\hat{D}^{(\ell)}\hat{a}_{2t}^{(\ell)}\right\|^2 &= T^{-1}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(\ell)}\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\varepsilon_{it}^{(\ell)}\right\|^2 \leq \frac{R}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\varepsilon_{it}^{(\ell)}\right)^2 \\ &\leq \frac{2R}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\varepsilon_{it}g_{it}\right)^2 + \frac{2R}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}F_s^{0\prime}\lambda_i^0\eta_{it}^{(\ell-1)}\bar{g}_{it}\right)^2. \end{aligned}$$

By the analysis of $T^{-1}\sum_{t=1}^{T}\|a_{3t}\|^2$ in the proof of Theorem 2.1, the first term is $O_P(\delta_{NT}^{-2})$. For the second term, by the CS inequality and Lemma A.3(vii) it is bounded above by $\frac{2R}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|F_s^{0\prime}\lambda_i^0\right\|^2 \times\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\eta_{it}^{(\ell-1)})^2 = O_P(\delta_{NT}^{-2})$. Then $T^{-1}\sum_{t=1}^{T}\left\|\hat{a}_{2t}^{(\ell)}\right\|^2 = O_P(\delta_{NT}^{-2})$. Analogously, we can show that $T^{-1}\sum_{t=1}^{T}\left\|\hat{D}^{(\ell)}\hat{a}_{3t}^{(\ell)}\right\|^2 = O_P(\delta_{NT}^{-2})$. In sum, we have shown that $T^{-1}\sum_{t=1}^{T}\left\|\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0\right\|^2 = O_P(\delta_{NT}^{-2})$. ■

Proof of Theorem 2.4. (i) Let $\hat{A}_{lt}^{(\ell)} = \hat{D}^{(\ell)}\hat{a}_{lt}^{(\ell)}$ for $l = 1,2,3$. By the decomposition in (A.8) and Lemma A.4(i), we will bound $\hat{A}_{lt}^{(\ell)}$ for $l = 1,3$ and find the leading influence function for $\hat{A}_{2t}^{(\ell)}$. For $\hat{A}_{1t}^{(\ell)}$, we use $\hat{F}_s^{(\ell)} = (\hat{F}_s^{(\ell)} - \hat{H}^{(\ell)\prime}F_s^0) + \hat{H}^{(\ell)\prime}F_s^0$ to make the decomposition

$$\hat{A}_{1t}^{(\ell)} = \frac{1}{NT}\sum_{s=1}^{T}(\hat{F}_s^{(\ell)} - \hat{H}^{(\ell)\prime}F_s^0)\sum_{i=1}^{N}\varepsilon_{it}^{(\ell)}\varepsilon_{is}^{(\ell)} + \hat{H}^{(\ell)\prime}\frac{1}{NT}\sum_{s=1}^{T}F_s^0\sum_{i=1}^{N}\varepsilon_{it}^{(\ell)}\varepsilon_{is}^{(\ell)} \equiv \hat{A}_{1t,1}^{(\ell)} + \hat{H}^{(\ell)\prime}\hat{A}_{1t,2}^{(\ell)}.$$

It is easy to show that $\hat{A}_{1t,1}^{(\ell)}$ is of smaller order than $\hat{A}_{1t,2}^{(\ell)}$. We focus on the study of $\hat{A}_{1t,2}^{(\ell)}$. By (A.6), we have $\hat{A}_{1t,2}^{(\ell)} = \frac{1}{NT}\sum_{s=1}^{T}F_s^0\sum_{i=1}^{N}(\varepsilon_{it}\varepsilon_{is}g_{it}g_{is} + \eta_{it}^{(\ell-1)}\eta_{is}^{(\ell-1)}\bar{g}_{it}\bar{g}_{is} + \varepsilon_{it}g_{it}\eta_{is}^{(\ell-1)}\bar{g}_{is} + \eta_{it}^{(\ell-1)}\bar{g}_{it}\varepsilon_{is}g_{is}) \equiv \sum_{l=1}^{4}\hat{A}_{1t,2l}^{(\ell)}$. By the analysis of $A_{1t,2}$ and $A_{2t,2}$ in the proof of Theorem 2.2(i), $\max_t\left\|\hat{A}_{1t,21}^{(\ell)}\right\| = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$. By Lemma A.3(vi)-(vii) and the CS inequality,

$$\begin{aligned} \max_t\left\|\hat{A}_{1t,22}^{(\ell)}\right\| &\leq \left\{\max_t\frac{1}{N}\sum_{i=1}^{N}(\eta_{it}^{(\ell-1)})^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\left\|F_s^0\right\|^2(\eta_{is}^{(\ell-1)})^2\bar{g}_{is}\right\}^{1/2} \\ &= O_P\left(\delta_{NT}^{-1}(T^{-1/2+\gamma_1/4} + (N/\ln N)^{-1/2})\right). \end{aligned}$$

By Lemma A.3(ix)-(x), $\hat{A}_{1t,23}^{(\ell)} + \hat{A}_{1t,24}^{(\ell)} = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$. Thus $\hat{A}_{1t,2}^{(\ell)} = O_P\left(T^{-1+\gamma/4} + \delta_{NT}^{-2}\ln N\right)$ and $\hat{A}_{1t}^{(\ell)} = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$.

For $\hat{A}_{3t}^{(\ell)}$, we apply (A.6) and $\hat{F}_s^{(\ell)} = (\hat{F}_s^{(\ell)} - \hat{H}^{(\ell)\prime}F_s^0) + \hat{H}^{(\ell)\prime}F_s^0$ to make the decomposition

$$\begin{aligned} \hat{A}_{3t}^{(\ell)} &= \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^{(\ell)}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{is}'g_{is}F_t^{0\prime} + \frac{1}{NT}\sum_{s=1}^{T}(\hat{F}_s^{(\ell)} - \hat{H}^{(\ell)\prime}F_s^0)\sum_{i=1}^{N}\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}F_t^0 \\ &\quad + \hat{H}^{(\ell)\prime}[\frac{1}{NT}\sum_{s=1}^{T}F_s^0\sum_{i=1}^{N}\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}]F_t^0 \equiv \left(\hat{A}_{3,1}^{(\ell)} + \hat{A}_{3,2}^{(\ell)} + \hat{H}^{(\ell)\prime}\hat{A}_{3,3}^{(\ell)}\right)F_t^0. \end{aligned}$$

Following the analysis of $A_{4t,1}$ and $A_{4t,2}$ in the proof of Theorem 2.2(i), we can show that $\hat{A}_{3,1}^{(\ell)} = O_P\left(\delta_{NT}^{-2}\right)$. For $\hat{A}_{3,2}^{(\ell)}$, we have by the CS inequality, Theorem 2.3 and Lemma A.3(vii)-(vii)

$$\begin{aligned}\left\|\hat{A}_{3,2}^{(\ell)}\right\| &\leq \frac{1}{T^{1/2}}\left\|\hat{F}^{(\ell)} - F^0\hat{H}^{(\ell)}\right\|\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}\right\|^2\right\}^{1/2} \\ &\leq O_P\left(\delta_{NT}^{-1}\right)\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\frac{1}{TN}\sum_{s=1}^{T}\sum_{i=1}^{N}(\eta_{is}^{(\ell-1)})^2\bar{g}_{is}\right\}^{1/2} = O_P\left(\delta_{NT}^{-2}\right),\end{aligned}$$

and $\left\|\hat{A}_{3,3}^{(\ell)}\right\| \leq \left\|\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}\right\| = O_P\left(\delta_{NT}^{-2}\right)$. It follows that $\max_t\left\|\hat{A}_{3t}^{(\ell)}\right\| = \max_t\left\|F_t^0\right\| \times O_P(\delta_{NT}^{-2}) = O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\right)$.

It follows that

$$\begin{aligned}\hat{\Pi}_{tN}^{(\ell)} &\equiv \sqrt{N}(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0) = \sqrt{N}\hat{\phi}_{F,t}^{(\ell)} + O_P(\sqrt{N}(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})) \\ &= \sqrt{N}[\hat{D}^{(\ell-1)}]^{-1}\frac{1}{T}\hat{F}^{(\ell-1)\prime}F^0\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}^{(\ell)}\bar{g}_{it} + O_P(\sqrt{N}(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})) \\ &= D^{-1}Q\sqrt{N}\beta_{F,t} + (1-q)\sqrt{N}\hat{\phi}_{F,t}^{(\ell-1)} + O_P(\sqrt{N}(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})),\end{aligned}$$

where the remainder term $O_P(\sqrt{N}(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}))$ holds uniformly in t. This, in conjunction with Lemma A.5(i), implies that

$$\begin{aligned}\hat{\Pi}_{tN}^{(\ell)} &= D^{-1}Q\sqrt{N}\beta_{F,t} + (1-q)\hat{\Pi}_{tN}^{(\ell-1)} + O_P(\sqrt{N}(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})) \\ &= D^{-1}Q\sqrt{N}\beta_{F,t}\sum_{s=0}^{\ell-1}(1-q)^s + (1-q)^\ell\,\hat{\Pi}_{tN}^{(0)} + o_P(1) \\ &\xrightarrow{d} N\left(0, D^{-1}Q\Gamma_{1g,t}(q)\,Q'D^{-1}\right)\ \text{ as }\ (N,T,\ell)\to\infty.\end{aligned}$$

(ii) Let $\varepsilon^{(\ell)}$ be the $T\times N$ matrix with (t,i)th element given by $\varepsilon_{it}^{(\ell)}$. Noting that $\hat{\Lambda}^{(\ell)\prime} = \frac{1}{T}\hat{F}^{(\ell)\prime}\hat{X}^{(\ell)}$, $\hat{X}^{(\ell)} = F^0\Lambda^{0\prime} + \varepsilon^{(\ell)}$, and $\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t^{(\ell)}\hat{F}_t^{(\ell)\prime} = I_R$, we have

$$\begin{aligned}\hat{\lambda}_i^{(\ell)} - \hat{H}^{(\ell)-1}\lambda_i^0 &= \frac{1}{T\tilde{q}}\sum_{t=1}^{T}\hat{F}_t^{(\ell)}\left\{\varepsilon_{it}^{(\ell)} + \left[\hat{F}_t^{(\ell)\prime}\hat{H}^{(\ell)-1} + (F_t^{0\prime} - \hat{F}_t^{(\ell)\prime}\hat{H}^{(\ell)-1})\right]\lambda_i^0\right\} - \hat{H}^{(\ell)-1}\lambda_i^0 \\ &= \frac{\hat{H}^{(\ell)\prime}}{T}\sum_{t=1}^{T}F_t^0\varepsilon_{it}^{(\ell)} + \frac{1}{T}\sum_{t=1}^{T}(F_t^{0\prime} - \hat{F}_t^{(\ell)\prime}\hat{H}^{(\ell)-1})\varepsilon_{it}^{(\ell)} + \frac{1}{T}\sum_{t=1}^{T}\hat{F}_t^{(\ell)}(\hat{H}^{(\ell)\prime}F_t^0 - \hat{F}_t^{(\ell)})'\hat{H}^{(\ell)-1}\lambda_i^0 \\ &\equiv \hat{B}_{1i}^{(\ell)} + \hat{B}_{2i}^{(\ell)} + \hat{B}_{3i}^{(\ell)}.\end{aligned}$$

By Lemma A.4(iv)-(v), we have $\max_i\left\|\hat{B}_{2i}^{(\ell)}\right\| = O_P\left(N^{-1/2+\gamma_2/4}\delta_{NT}^{-1} + \delta_{NT}^{-2}\ln N\right)$ and $\max_i\left\|\hat{B}_{3i}^{(\ell)}\right\| = \max_i\left\|\lambda_i^0\right\|O_P\left(\delta_{NT}^{-2}\right) = O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\right)$. It follows that

$$\begin{aligned}\hat{\Pi}_{iT}^{(\ell)} &\equiv \sqrt{T}\left(\hat{\lambda}_i^{(\ell)} - \hat{H}^{(\ell)-1}\lambda_i^0\right) = \sqrt{T}\hat{B}_{1i}^{(\ell)} + O_P(\sqrt{T}(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N + N^{-1+3\gamma_2/4})) \\ &= (Q')^{-1}\sqrt{T}\beta_{\Lambda,i} + (1-q)\sqrt{T}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + o_P(1),\end{aligned}$$

where the remainder term $O_P(\sqrt{T}(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N + N^{-1+3\gamma_2/4}))$ holds uniformly in i. This, in conjunction with Lemma A.5(ii), implies that

$$
\begin{aligned}
\hat{\Pi}_{iT}^{(\ell)} &= (Q')^{-1}\sqrt{T}\beta_{\Lambda,i} + (1-q)\hat{\Pi}_{iT}^{(\ell-1)} + O_P(\sqrt{T}N^{\gamma_2/4}\delta_{NT}^{-2}\ln N) \\
&= (Q')^{-1}\sqrt{T}\beta_{\Lambda,i}\sum_{s=0}^{\ell-1}(1-q)^s + (1-q)^\ell\,\hat{\Pi}_{iT}^{(0)} + O_P(\sqrt{T}(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N + N^{-1+3\gamma_2/4})) \\
&\xrightarrow{d} N\left(0, (Q')^{-1}\Phi_{1g,i}(q)\,Q^{-1}\right) \text{ as } (N,T,\ell)\to\infty.
\end{aligned}
$$

(iii) By the proof of (i) and (ii) and as in the proof of Theorem 2.2(iii), we have

$$
\begin{aligned}
\hat{C}_{it}^{(\ell)} - C_{it}^0 &= \hat{\lambda}_i^{(\ell)\prime}\hat{F}_t^{(\ell)} - \lambda_i^{0\prime}F_t^0 = \lambda_i^{0\prime}(\hat{H}^{(\ell)\prime})^{-1}(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0) + \hat{F}_t^{(\ell)\prime}(\hat{\lambda}_i^{(\ell)} - \hat{H}^{(\ell)-1}\lambda_i^0) \\
&= \frac{1}{\sqrt{N}}\lambda_i^{0\prime}(\hat{H}^{(\ell)\prime})^{-1}\sqrt{N}(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0) + \frac{1}{\sqrt{T}}F_t^{0\prime}\hat{H}^{(\ell)}\sqrt{T}(\hat{\lambda}_i^{(\ell)} - \hat{H}^{(\ell)-1}\lambda_i^0) \\
&\quad + O_P((N^{\gamma_2/4} + T^{\gamma_1/4})(NT)^{-1/2}) \\
&= \frac{1}{\sqrt{N}}\lambda_i^{0\prime}(\hat{H}^{(\ell)\prime})^{-1}\hat{\Pi}_{tN}^{(\ell)} + \frac{1}{\sqrt{T}}F_t^{0\prime}\hat{H}^{(\ell)}\hat{\Pi}_{iT}^{(\ell)} + o_P(1).
\end{aligned}
$$

Then we have $(\frac{1}{N}\Sigma_{1F,it} + \frac{1}{T}\Sigma_{1\Lambda,it})^{-1/2}(\hat{C}_{it}^{(\ell)} - C_{it}^0) \xrightarrow{d} N(0,1)$ as $(N,T,\ell)\to\infty$, where $\Sigma_{1F,it} = \lambda_i^{0\prime}\Sigma_{\Lambda^0}^{-1}\Gamma_{1g,t}(q)\,\Sigma_{\Lambda^0}^{-1}\lambda_i^0$ and $\Sigma_{1\Lambda,it} = F_t^{0\prime}\Sigma_{F^0}^{-1}\Phi_{1g,i}(q)\,\Sigma_{F^0}^{-1}F_t^0$. ■

To prove Theorem 2.5, we need the following lemma.

Lemma A.6 *Suppose that Assumptions A.1-A.6 hold. Then*

(i) $\max_i \frac{1}{T}\sum_{t=1}^T |\hat{\varepsilon}_{it} - \varepsilon_{it}|^2 = O_P(N^{-1+\gamma_2/2} + T^{-1}\ln T)$,

(ii) $\max_{i,t}|\hat{\varepsilon}_{it} - \varepsilon_{it}| = O_P\left((T^{-1/2+\gamma_1/4} + N^{-1/2+\gamma_2/4})(\ln T)^{1/2}\right) = o_P(1)$,

(iii) $\left\|\hat{\Sigma}^g - \Sigma^g\right\|_{sp} = o_P(1)$.

Proof of Theorem 2.5. To show $\hat{D}^{-1}\hat{\Gamma}_{1g,t}^{(2)}\hat{D}^{-1} \xrightarrow{p} D^{-1}Q\Gamma_{1g,t}(q)\,Q'D^{-1}$, it suffices to show that (i) $\hat{D}^{-1} \xrightarrow{p} D^{-1}$ and (ii) $\hat{\Gamma}_{1g,t}^{(2)} \xrightarrow{p} Q\Gamma_{1g,t}Q'$. (i) holds by Lemma A.4(i) and positive definiteness of D. To show (ii), we recall that $\hat{\Gamma}_{1g,t}^{(2)} = \frac{1}{N\tilde{q}^2}\hat{\Lambda}'\hat{\Sigma}^g\hat{\Lambda}$ and $\Gamma_{1g,t}(q) = \lim_{N\to\infty}\Gamma_{1g,tN}(q)$, where $\Gamma_{1g,tN}(q) = \frac{1}{Nq^2}\Lambda^{0\prime}\Sigma^g\Lambda^0$. Then by the triangle inequality, we have

$$
\begin{aligned}
\left\|\hat{\Gamma}_{1g,t}^{(2)} - Q\Gamma_{1g,t}Q'\right\|_{\text{sp}} \le{}& \frac{1}{N\tilde{q}^2}\left\|\hat{\Lambda}'\hat{\Sigma}^g\hat{\Lambda} - Q\Lambda^{0\prime}\Sigma^g\Lambda^0Q'\right\|_{\text{sp}} + \frac{1}{N}\left\|Q\Lambda^{0\prime}\Sigma^g\Lambda^0Q'\right\|_{\text{sp}}\left(\frac{1}{q^2} - \frac{1}{\tilde{q}^2}\right) \\
&+ \left\|Q\left(\Gamma_{1g,tN}(q) - \Gamma_{1g,t}(q)\right)Q'\right\|_{\text{sp}}.
\end{aligned}
$$

The last term on the right hand side (rhs) of the last expression is $o(1)$ and the second term is $O_P((NT)^{-1/2})$ by noting that $\tilde{q} - q = O_P((NT)^{-1/2})$. For the first term, we have

$$
\left\|\hat{\Lambda}'\hat{\Sigma}^g\hat{\Lambda} - Q\Lambda^{0\prime}\Sigma^g\Lambda^0Q'\right\|_{\text{sp}} \le \left\|[\hat{\Lambda} - \Lambda^0Q']'\hat{\Sigma}^g\hat{\Lambda}\right\|_{\text{sp}} + \left\|Q\Lambda^{0\prime}(\hat{\Sigma}^g - \Sigma^g)\hat{\Lambda}\right\|_{\text{sp}} + \left\|Q\Lambda^{0\prime}\Sigma^g[\hat{\Lambda} - \Lambda^0Q']\right\|_{\text{sp}}.
$$

It is standard to show $\frac{1}{N}\left\|\hat{\Lambda}-\Lambda^0 Q\right\|^2 \leq \frac{1}{N}\left\|\hat{\Lambda}-\Lambda^0 \hat{H}^{(\ell)-1}\right\|^2+\frac{1}{N}\left\|\Lambda^0(\hat{H}^{(\ell)-1}-Q)\right\|^2=o_P(1)$ by using the expression of $\hat{\lambda}_i-\hat{H}^{(\ell)-1}\lambda_i^0$ in the proof of Theorem 2.4(ii) and Lemma A.4(iii). In addition, $\left\|\hat{\Sigma}^g\right\|_{\mathrm{sp}} \leq \left\|\Sigma^g\right\|_{\mathrm{sp}}+\left\|\hat{\Sigma}^g-\Sigma^g\right\|_{\mathrm{sp}}=O\left(1\right)+o_P\left(1\right)=O_P\left(1\right)$ by Lemma A.6. It follows that

$$\begin{aligned}
\frac{1}{N}\left\|[\hat{\Lambda}-\Lambda^0 Q']'\hat{\Sigma}^g\hat{\Lambda}\right\|_{\mathrm{sp}} &\leq \left\|\hat{\Sigma}^g\right\|_{\mathrm{sp}}\frac{1}{N^{1/2}}\left\|\hat{\Lambda}\right\|_{\mathrm{sp}}\frac{1}{N^{1/2}}\left\|\hat{\Lambda}-\Lambda^0 Q\right\|_{\mathrm{sp}} \\
&\leq O_P\left(1\right)\frac{1}{N^{1/2}}\left\{\left\|\hat{\Lambda}-\Lambda^0(\hat{H}^{(\ell)\prime})^{-1}\right\|_{\mathrm{sp}}+\left\|\Lambda^0\left[(\hat{H}^{(\ell)\prime})^{-1}-Q'\right]\right\|_{\mathrm{sp}}\right\}=o_P\left(1\right).
\end{aligned}$$

Similarly, by Lemma A.6, we have $\frac{1}{N}||Q\Lambda^{0\prime}(\hat{\Sigma}^g-\Sigma^g)\hat{\Lambda}||_{\mathrm{sp}} \leq \|Q\|_{\mathrm{sp}}\frac{1}{N^{1/2}}\left\|\Lambda^0\right\|_{\mathrm{sp}}\frac{1}{N^{1/2}}||\hat{\Lambda}||_{\mathrm{sp}}||\hat{\Sigma}^g-\Sigma^g||_{\mathrm{sp}}=o_P\left(1\right)$. By the same token, we have $\frac{1}{N}||Q\Lambda^{0\prime}\Sigma^g[\hat{\Lambda}-\Lambda^0 Q']||_{\mathrm{sp}}=o_p\left(1\right)$. It follows that $\frac{1}{N}||\hat{\Lambda}'\hat{\Sigma}^g\hat{\Lambda}-Q\Lambda^{0\prime}\Sigma^g\Lambda^0 Q'||_{\mathrm{sp}}=o_P\left(1\right)$. ■

REFERENCES

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203-1227.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., Khosravi, K., 2018. Matrix completion methods for causal panel data models. Working Paper, Graduate School of Business, Stanford University.

Bai, J., 2003. Inferential theory for factors models of large dimensions. Econometrica 71, 135-173.

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

Bai, J., Liao, Y., Yang, J., 2015. Unbalanced panel data models with interactive effects. In B.H. Baltagi (eds), The Oxford Handbook of Panel Data, pp., 149-170.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business & Economic Statistics 25, 52-60.

Bańbura, M., Modugno, M., 2014. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. Journal of Applied Econometrics 29, 133-160.

Bernstein, D. S., 2005. Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, Princeton.

Cai, J-F., Candès, E.J., Shen, Z., 2010. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization 20, 1956-1982.

Candès, E.J., Li, X., 2011. Robust principal component analysis. Journal of the ACM 58(3), 11:1-37.

Candès, E.J.,Plan, Y., 2010. Matrix completion with noise. Proceedings of the IEEE 98(6), 925-936.

Doz, C., Giannone, D., Reichlin, L., 2011. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. Journal of Econometrics 164, 188-205.

Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B 75, 603-680.

Foroni, C., Marcellino, M.G., 2013. A Survey of Econometric Methods for Mixed-Frequency Data. SSRN Electronic Journal.

Foroni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Review of Economics and Statistics 82, 540-554.

Geweke, J.F., 1977. The dynamic factor analysis of economic time series models. In D. Aigner and A. Goldberger (eds.), Latent Variables in Socioeconomic Models, pp. 365-383, North-Holland, Amsterdam.

Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics 55, 665-676.

Hallin, M., Liśka, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102, 603-617.

Häusler, E., Luschgy, H., 2015. Stable Convergence and Stable Limit Theorems. Springer, New York.

Jungbacker, B., Koopman, S., Wel, M.V.D., 2011. Maximum likelihood estimation for dynamic factor models with missing data. Journal of Economic Dynamics and Control 35, 1358-1368.

Ludvigson, S., Ng, S., 2007. The empirical risk-return relation: a factor analysis approach. Journal of Financial Economics 83,171-222.

Kapetanios, G., 2010. A testing procedure for determining the number of factors in approximate factor models with large datasets. Journal of Business & Economic Statistics 28, 397-409.

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

Marcellino, M., Sivec, V., 2016. Monetary, fiscal, and oil shocks: evidence based on mixed frequency structural FAVARs. Journal of Econometrics 193, 335-348.

Mariano, R.S., Murasawa, Y., 2010. A coincident index, common Factors, and monthly real GDP. Oxford Bulletin of Economics and Statistics 72, 27-46.

Meinshausen, N., Bühlmann, 2010. Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72, 417-473.

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

Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447-1479.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economics and Statistics 92, 1004-1016.

Onatski, A., 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. Journal of Econometrics 168, 244-258.

Pinheiro, M., Rua, A., Dias, F., 2012. Dynamic Factor Models with Jagged Edge Panel Data: Taking on Board the Dynamics of the Idiosyncratic Components. Oxford Bulletin of Economics and Statistics 75, 80-102.

Sargent, T.J., Sims, C., 1977. Business cycle modelling without pretending to have too much a-priori economic theory. In C. Sims (ed.) New Methods in Business Cycle Research, pp. 45-109. Federal Reserve Bank of Minneapolis, Minneapolis.

Schumacher, C., Breitung, J., 2008. Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. International Journal of Forecasting 24, 386-398.

Stock, J.H., Watson, M.W., 1998. Diffusion indexes. Working paper 6702, National Bureau of Economic Research.

Table 1: Under/Over-estimation frequency with complete data

			Cross-validation		ED	GR	ER	PC	IC
DGP	N	T	p=0.7	p=0.9					
1	50	50	460/14	62/43	0/393	530/37	864/10	0/685	0/445
	50	100	121/47	7/54	0/360	381/49	791/11	0/497	0/377
	100	50	166/27	7/42	0/336	350/39	769/15	0/518	0/364
	100	100	10/48	0/30	0/374	212/49	643/13	0/452	0/363
2	50	50	440/4	51/2	0/113	416/3	816/0	0/246	0/99
	50	100	104/0	1/1	0/75	194/1	676/0	0/106	0/53
	100	50	166/1	6/1	0/79	218/2	694/0	0/118	0/52
	100	100	12/0	0/0	0/60	39/0	409/0	0/54	0/32
3	50	50	308/0	12/2	0/72	566/0	910/0	0/995	0/840
	50	100	40/0	0/0	0/3	198/0	779/0	0/473	0/46
	100	50	88/0	0/5	0/17	424/0	884/0	0/1000	0/929
	100	100	1/0	0/0	0/0	42/0	589/0	0/518	0/63
4	50	50	361/0	19/5	2/175	655/0	931/0	0/685	0/280
	50	100	67/0	0/18	0/236	525/0	897/0	0/679	0/380
	100	50	108/0	1/0	0/40	418/0	861/0	0/272	0/43
	100	100	1/0	0/13	0/45	184/0	766/0	0/416	0/124
5	50	50	360/0	15/1	1/92	631/0	938/0	0/1000	0/836
	50	100	57/0	0/3	0/20	451/0	895/0	0/1000	0/928
	100	50	91/0	0/0	0/3	223/0	782/0	0/433	0/35
	100	100	0/0	0/0	0/0	47/0	576/0	0/465	0/63
6	50	50	322/0	18/0	0/1	282/0	780/0	0/69	0/0
	50	100	46/0	0/0	0/0	84/0	603/0	0/0	0/0
	100	50	89/0	0/0	0/0	89/0	583/0	0/0	0/0
	100	100	1/0	0/0	0/0	2/0	216/0	0/0	0/0

Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20, 147-162.

Stock, J., Watson, M., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. Handbook of Macroeconomics 415–525.

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

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

Su, L., Jin, S., Zhang, Y., 2015. Specification test for panel data models with interactive fixed effects. Journal of Econometrics 186, 222-244.

Table 2 : Under/Over-estimation frequency with missing data (q=0.7)

			Cross-validation			ED		GR		ER		PC		IC	
DGP	N	T	$CV^{(0)}$	$CV_R^{(l^*)}$	$CV_{R_{\max}}^{(l^*)}$	M-1	M-2	M-1	-2l	M-1	M-2l	M-1	-2l	M-1	M-2
1	50	50	628/19	424/21	371/17	446/183	0/1000	991/2	153/843	1000/0	658/342	0/1000	0/1000	0/996	0/1000
	50	100	282/51	94/50	71/26	287/127	0/1000	993/0	152/842	998/0	686/311	0/999	0/1000	0/978	0/1000
	100	50	368/37	173/42	140/27	355/122	0/1000	991/2	186/808	999/0	698/300	0/998	0/1000	0/979	0/1000
	100	100	42/58	6/49	2/24	78/108	0/1000	987/1	150/834	1000/0	701/296	0/990	0/1000	0/941	0/1000
2	50	50	642/6	445/0	356/1	500/125	0/1000	995/0	285/709	999/0	856/138	0/1000	0/1000	0/981	0/1000
	50	100	298/1	103/1	55/0	302/48	0/1000	989/0	370/568	999/0	932/48	0/985	0/1000	0/740	0/1000
	100	50	360/2	152/0	90/0	378/94	0/1000	987/0	354/600	998/0	915/59	0/987	0/1000	0/744	0/1000
	100	100	35/0	6/0	1/0	70/41	0/1000	987/0	402/370	998/0	929/17	0/784	0/1000	0/455	0/1000
3	50	50	528/0	279/0	209/0	437/60	0/1000	992/0	535/380	999/0	964/12	0/1000	0/1000	0/1000	0/1000
	50	100	176/0	27/0	12/0	209/15	0/1000	981/0	432/55	998/0	903/0	0/1000	0/1000	0/957	0/1000
	100	50	215/0	61/0	33/0	233/44	0/1000	990/0	553/102	997/0	938/0	0/1000	0/1000	0/1000	0/1000
	100	100	5/0	0/0	0/0	22/10	0/1000	972/0	94/0	997/0	739/0	0/1000	0/1000	0/956	0/1000
4	50	50	560/0	313/0	262/0	494/49	0/1000	990/0	562/409	998/0	961/22	0/999	0/1000	0/988	0/1000
	50	100	212/0	53/0	29/0	240/10	0/1000	990/0	627/167	999/0	959/2	0/997	0/1000	0/903	0/1000
	100	50	298/0	94/0	55/0	282/29	0/1000	989/0	624/120	997/0	944/1	0/984	0/1000	0/733	0/1000
	100	100	7/0	1/0	0/0	26/6	0/1000	985/0	306/5	998/0	899/0	0/951	0/1000	0/721	0/1000
5	50	50	567/1	305/0	230/0	483/66	0/1000	992/0	599/336	998/0	970/9	0/1000	0/1000	0/1000	0/1000
	50	100	186/0	39/0	23/0	199/13	0/1000	989/0	599/85	1000/0	951/0	0/1000	0/1000	0/1000	0/1000
	100	50	259/0	60/0	33/0	230/40	0/1000	978/0	458/72	997/0	886/1	0/999	0/1000	0/957	0/1000
	100	100	13/0	0/0	0/0	21/6	0/1000	981/0	119/0	999/0	766/0	0/1000	0/1000	0/962	0/1000
6	50	50	539/1	290/0	214/0	425/66	0/1000	990/0	523/338	997/0	960/2	0/993	0/1000	0/910	0/1000
	50	100	175/0	35/0	17/0	189/23	0/1000	990/0	327/32	998/0	884/0	0/930	0/1000	0/178	0/1000
	100	50	244/0	48/0	26/0	232/46	0/1000	982/0	316/40	995/0	868/0	0/915	0/1000	0/171	0/1000
	100	100	11/0	0/0	0/0	13/13	0/1000	972/0	29/0	997/0	559/0	0/83	0/1000	0/1	0/1000

Table 3: MSE and $R^2(\hat{F})$ with missing observations (q=0.7)

			MSE					$R^2(\hat{F})$				
			oracle	iterated estimate				oracle	iterated estimate			
DGP	N	T		ℓ=0	ℓ=5	ℓ=20	$\ell=\infty$		ℓ=0	ℓ=5	ℓ=20	$\ell=\infty$
1	50	50	0.460	2.103	0.766	0.807	0.886	0.964	0.856	0.936	0.941	0.940
	50	100	0.367	1.484	0.546	0.578	0.617	0.967	0.876	0.946	0.948	0.947
	100	50	0.423	1.659	0.604	0.636	0.687	0.978	0.913	0.965	0.967	0.967
	100	100	0.221	0.890	0.332	0.355	0.376	0.982	0.935	0.973	0.973	0.973
2	50	50	0.352	1.907	0.616	0.588	0.594	0.971	0.863	0.947	0.953	0.953
	50	100	0.259	1.280	0.406	0.405	0.406	0.972	0.885	0.957	0.958	0.958
	100	50	0.258	1.333	0.408	0.405	0.405	0.986	0.925	0.978	0.979	0.979
	100	100	0.172	0.785	0.258	0.260	0.260	0.986	0.943	0.979	0.980	0.980
3	50	50	0.403	1.703	0.562	0.555	0.555	0.975	0.886	0.961	0.963	0.963
	50	100	0.266	1.127	0.373	0.375	0.375	0.976	0.901	0.964	0.964	0.964
	100	50	0.328	1.250	0.432	0.431	0.431	0.987	0.938	0.981	0.981	0.981
	100	100	0.198	0.743	0.263	0.264	0.264	0.988	0.950	0.983	0.983	0.983
4	50	50	0.350	1.749	0.562	0.551	0.551	0.970	0.873	0.951	0.954	0.954
	50	100	0.261	1.160	0.395	0.397	0.397	0.970	0.894	0.956	0.956	0.956
	100	50	0.262	1.220	0.399	0.400	0.399	0.985	0.934	0.977	0.977	0.977
	100	100	0.173	0.739	0.257	0.258	0.258	0.985	0.946	0.978	0.978	0.978
5	50	50	0.386	1.704	0.554	0.542	0.542	0.970	0.878	0.955	0.957	0.957
	50	100	0.316	1.183	0.420	0.422	0.422	0.970	0.894	0.958	0.958	0.959
	100	50	0.260	1.193	0.370	0.370	0.369	0.985	0.935	0.979	0.979	0.979
	100	100	0.190	0.731	0.256	0.257	0.257	0.985	0.947	0.980	0.980	0.980
6	50	50	0.322	1.627	0.492	0.483	0.483	0.976	0.886	0.961	0.963	0.963
	50	100	0.239	1.106	0.347	0.348	0.348	0.976	0.900	0.964	0.964	0.964
	100	50	0.244	1.168	0.353	0.354	0.353	0.988	0.939	0.982	0.982	0.982
	100	100	0.161	0.703	0.226	0.227	0.227	0.988	0.950	0.983	0.983	0.983

Table 4: Coverage probability and average length of the 95% confidence intervals (q=0.7)

			Oracle				ℓ=0				$\ell=\ell^*$			
			standard		robust		standard		robust		standard		robust	
DGP	N	T	CP	Length	CP	Length	CP	Length	CP	Length	CP	Length	CP	Length
1	50	50	0.926	0.514	0.947	0.551	0.919	0.874	0.943	0.947	0.906	0.568	0.935	0.645
	50	100	0.919	0.529	0.930	0.562	0.912	0.859	0.934	0.928	0.881	0.595	0.920	0.656
	100	50	0.926	0.365	0.940	0.400	0.910	0.641	0.943	0.685	0.937	0.439	0.938	0.476
	100	100	0.940	0.374	0.943	0.403	0.936	0.650	0.948	0.684	0.932	0.438	0.940	0.478
2	50	50	0.918	0.537	0.932	0.550	0.919	0.892	0.936	0.936	0.909	0.619	0.929	0.642
	50	100	0.922	0.538	0.924	0.557	0.924	0.912	0.943	0.950	0.896	0.625	0.926	0.653
	100	50	0.943	0.388	0.946	0.395	0.941	0.645	0.956	0.673	0.935	0.453	0.943	0.467
	100	100	0.938	0.390	0.936	0.401	0.926	0.655	0.943	0.678	0.932	0.460	0.944	0.474
3	50	50	0.926	0.550	0.936	0.557	0.902	0.922	0.930	0.945	0.894	0.646	0.923	0.653
	50	100	0.932	0.565	0.938	0.567	0.921	0.909	0.940	0.931	0.921	0.658	0.927	0.666
	100	50	0.930	0.400	0.937	0.398	0.934	0.680	0.944	0.688	0.906	0.462	0.923	0.472
	100	100	0.925	0.403	0.933	0.404	0.931	0.660	0.942	0.667	0.922	0.472	0.943	0.478
4	50	50	0.917	0.601	0.937	0.607	0.928	0.972	0.937	0.995	0.896	0.697	0.923	0.710
	50	100	0.928	0.607	0.943	0.614	0.917	0.948	0.937	0.969	0.909	0.709	0.933	0.719
	100	50	0.927	0.440	0.928	0.436	0.926	0.704	0.944	0.712	0.935	0.517	0.941	0.519
	100	100	0.932	0.445	0.943	0.447	0.914	0.703	0.926	0.712	0.918	0.520	0.930	0.529
5	50	50	0.891	0.322	0.908	0.327	0.918	0.749	0.946	0.778	0.875	0.379	0.899	0.386
	50	100	0.896	0.323	0.901	0.328	0.900	0.732	0.927	0.754	0.892	0.381	0.912	0.388
	100	50	0.885	0.233	0.885	0.233	0.921	0.542	0.950	0.561	0.894	0.276	0.910	0.277
	100	100	0.904	0.234	0.905	0.236	0.928	0.546	0.944	0.555	0.884	0.276	0.905	0.281
6	50	50	0.897	0.320	0.911	0.325	0.924	0.737	0.939	0.767	0.891	0.377	0.909	0.384
	50	100	0.875	0.325	0.896	0.330	0.917	0.734	0.940	0.752	0.894	0.384	0.907	0.390
	100	50	0.913	0.233	0.917	0.233	0.928	0.524	0.951	0.545	0.907	0.275	0.920	0.276
	100	100	0.908	0.236	0.913	0.236	0.926	0.532	0.939	0.543	0.898	0.277	0.924	0.280

Table 5: Results of forecasts

		Real GDP			GDP			IP			RDPI		
		MSE	ratio		MSE	ratio		MSE	ratio		MSE	ratio	
period	horizon		AR	PC-F		AR	PC-F		AR	PC-F		AR	PC-F
1987 ∽ 2016	h=1	4.571	0.923	0.985	6.665	0.921	1.004	11.488	0.911	0.929	11.896	0.958	0.988
	h=2	2.986	0.853	0.968	5.349	0.874	1.003	13.091	0.896	0.922	4.505	0.888	0.985
	h=4	2.683	0.948	0.927	5.727	0.940	0.996	13.489	0.969	0.994	2.565	0.841	0.989
1997 ∽ 2016	h=1	4.734	0.870	1.009	6.745	0.892	1.000	12.131	0.853	0.896	14.982	0.957	0.987
	h=2	3.246	0.813	0.957	5.531	0.851	0.998	15.583	0.875	0.918	5.085	0.856	0.995
	h=4	3.020	0.924	0.955	5.924	0.916	0.997	16.964	0.948	0.983	2.832	0.809	0.983
2007 ∽ 2016	h=1	5.049	0.746	0.982	8.170	0.794	0.984	16.818	0.805	0.862	20.446	0.941	0.982
	h=2	4.247	0.749	0.922	7.167	0.801	1.004	23.777	0.851	0.886	6.565	0.785	0.985
	h=4	4.445	0.901	0.950	8.145	0.923	1.011	26.810	0.904	0.936	4.047	0.777	0.973

Online Supplement for
"Inference for Approximate Factor Models: Random Missing and Cross Validation"

Liangjun Su, Ke Miao, and Sainan Jin
School of Economics, Singapore Management University

This online supplement contains the proofs of the theorems in Section 3.

B Proofs of the Main Results in Section 3

To prove Theorem 3.1, we need to introduce some notations and lemmas. Note that the true number of factors is assumed to be R_0 but the working model is given by

$$X = F\left(R\right)\Lambda\left(R\right)' + \varepsilon\left(R\right),$$

where we make the dependence of F and Λ on the assumed number of factors (R) explicit and $\varepsilon\left(R\right) \equiv X - F\left(R\right)\Lambda\left(R\right)'$. As in Bai and Ng (2017), we want to establish the connection between the usual principal component (PC) estimators of the factors and factor loadings and the SVD estimators.

Let $X^* = P_{\Omega^*}X$. Noting that $\tilde{C}_R = S_H(\frac{1}{p}P_{\Omega^*}X, R) = \tilde{U}_R\tilde{\Sigma}_R\tilde{V}_R'$, $\tilde{U}_R$ and $\tilde{V}_R$ are respectively the eigenvector matrices of $\frac{1}{p^2}X^*X^{*\prime}$ and $\frac{1}{p^2}X^{*\prime}X^*$ associated with their R largest eigenvalues, and the diagonal elements of $\tilde{\Sigma}_R^2$ are the R largest eigenvalues of $\frac{1}{p^2}X^*X^{*\prime}$. Let $\tilde{F}^R$ and $\tilde{\Lambda}^R$ denote the conventional principal component (PC) estimators of $F\left(R\right)$ and $\Lambda\left(R\right)$ under the normalization restrictions that $T^{-1}F\left(R\right)'F\left(R\right) = I_R$ and $\Lambda\left(R\right)'\Lambda\left(R\right) =$diagonal matrix. It is well known that $\tilde{F}^R$ is given by $\sqrt{T}$ times the normalized eigenvector matrix of $\frac{1}{p^2}X^*X^{*\prime}$ associated with its R largest eigenvalues and $\tilde{\Lambda}^{R\prime} = (\tilde{F}^{R\prime}\tilde{F}^R)^{-1}\tilde{F}^{R\prime}\frac{1}{p}X^* = \tilde{F}^{R\prime}\frac{1}{Tp}X^*$. This indicates that

$$\tilde{F}^R = \sqrt{T}\tilde{U}_R. \tag{B.1}$$

In addition, we consider the full SVD of $\frac{1}{p}X^*$: $\frac{1}{p}X^* = \tilde{U}\tilde{\Sigma}\tilde{V}' = \sum_{r=1}^{T\wedge N}\tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r$. Then $\frac{1}{p}X^{*\prime}\tilde{U} = \tilde{V}\tilde{\Sigma}'\tilde{U}'\tilde{U} = \tilde{V}\tilde{\Sigma}'$. This implies that

$$\tilde{V}_R\tilde{\Sigma}_R = \frac{1}{p}X^{*\prime}\tilde{U}_R = \frac{\sqrt{T}}{Tp}X^{*\prime}\tilde{F}^R = \sqrt{T}\tilde{\Lambda}^R. \tag{B.2}$$

(B.1) says that $\tilde{U}_R$ is a scaled version of $\tilde{F}^R$ and (B.2) says that each column of $\tilde{V}_R$ is a scaled version of the corresponding column of $\tilde{\Lambda}^R$. It is easy to see that

$$\tilde{U}_R\tilde{\Sigma}_R\tilde{V}_R' = \tilde{F}^R\tilde{\Lambda}^{R\prime}. \tag{B.3}$$

That is, both the SVD and the PCA yield the same estimates of the common component once R is given. Following the lead of Bai and Ng (2002), we consider a rotational version of $\tilde{F}^R$: $\check{F}^R =$

$\left(NTp^2\right)^{-1} X^* X^{*\prime} \tilde{F}^R$. Let $\breve{H}_{1R} = (N^{-1}\Lambda^{0\prime}\Lambda^0)(T^{-1}F^{0\prime}\tilde{F}^R)$. The properties of $\breve{F}^R$ can be established along the lines of proofs in Bai and Ng (2002) and those in the proof of Theorem 2.1 in the presence of random missing values.

Alternatively, we can consider the PC estimation under the normalization restrictions that $N^{-1}\Lambda(R)' \Lambda(R) = I_R$ and $F(R)' F(R)$ =diagonal matrix. Let $\bar{F}^R$ and $\bar{\Lambda}^R$ denote the conventional PC estimators of $F(R)$ and $\Lambda(R)$ in this case. Then following the above arguments, we can show that

$$\bar{\Lambda}^R = \sqrt{N}\tilde{V}_R,\ \tilde{U}_R\tilde{\Sigma}_R = \sqrt{N}\bar{F}^R,\ \text{and } \tilde{U}_R\tilde{\Sigma}_R\tilde{V}_R' = \bar{F}^R\bar{\Lambda}^{R\prime}. \tag{B.4}$$

Following the lead of Bai and Ng (2002), we consider a rotational version of $\bar{\Lambda}_R : \breve{\Lambda}^R = \left(NTp^2\right)^{-1} X^{*\prime} X^* \bar{\Lambda}^R$. Let $\breve{H}_{2R} = (T^{-1}F^{0\prime}F^0)(N^{-1}\Lambda^{0\prime}\bar{\Lambda}^R)$.

Finally, let $\tilde{D}_R$ denote the $R \times R$ diagonal matrix that contains the R largest eigenvalues of $(NTp^2)^{-1}X^*X^{*\prime}$ arranged in descending order along its diagonal line. Note that $\tilde{D}_R = (NT)^{-1}\tilde{\Sigma}_R^2$.

Recall that $\bar{g}_{it}^* = \mathbf{1}\{(i,t) \in \Omega_\perp^*\}$ and $g_{it}^* = \mathbf{1}\{(i,t) \in \Omega^*\}$. Let $\bar{G}^*$ be the $T \times N$ matrix with (t,i)th element given by $\bar{g}_{it}^*$. Define G^* analogously. Let e_{rR} denote the rth column of the $R \times R$ identity matrix I_R. Similarly, e_{rN} and e_{rT} denote the rth column of I_N and I_T, respectively. Note that $\tilde{u}_r \equiv \tilde{U}_R e_{rR}$ and $\tilde{v}_r \equiv \tilde{V}_R e_{rR}$, $r = 1, ..., R$, denote the rth column of $\tilde{U}_R$ and $\tilde{V}_R$, respectively. In addition, $\tilde{C}_R = \sum_{r=1}^R \tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r$.

The proof of Theorem 3.1 needs the following three lemmas and two theorems, whose proofs are given after we finish the proofs of Theorems 3.1 and 3.2.

Lemma B.1 *Suppose that all the conditions but Assumption A.7 in Theorem 3.1 hold. Then*
(i) $\frac{1}{T}\left\|\sqrt{T}\tilde{U}_R\tilde{D}_R - F^0\breve{H}_{1R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$,
(ii) $\frac{1}{N}\left\|\sqrt{N}\tilde{V}_R\tilde{D}_R - \Lambda^0\breve{H}_{2R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$.

Lemma B.2 *Let* $\breve{\sigma}_r = (NT)^{-1/2}\tilde{\sigma}_r$. *Let* σ_r^2 *denote the* r*th largest eigenvalue of* $\Sigma_{F^0}\Sigma_{\Lambda^0}$ *for* $r = 1, ..., R_0$. *Suppose that all the conditions but Assumption A.7 in Theorem 3.1 hold. Then*
(i) $\breve{\sigma}_r^2 = \sigma_r^2 + O_P\left(\delta_{NT}^{-1}\right)$ *for* $r = 1, ..., R_0$,
(ii) $\breve{\sigma}_{R_0+r}^2 = O_P\left(\delta_{NT}^{-2}\right)$ *for* $r \geq 1$,
(iii) $\delta_{NT}^2\breve{\sigma}_{R_0+r}^2 \geq c_\sigma + o_P(1)$ *for some positive constant* c_σ *and any* $r \geq 1$ *with* $R_0 + r \leq R$.

Lemma B.3 *Let* $\tilde{u}_r$ *and* $\tilde{v}_r$ *be the* r*th left and right singular vector of* $\frac{1}{p}X^*$. *Suppose that all the conditions but Assumption A.7 in Theorem 3.1 hold. Then for* $r = R_0+1, ..., R_{\max}$, *we have* $\tilde{u}_r'F^0 = O_P(\delta_{NT}^{-1})$ *and* $\tilde{v}_r'\Lambda^0 = O_P(\delta_{NT}^{-1})$.

To proceed, we define some notations. For a real matrix Γ, recall that $\|\Gamma\|$ and $\|\Gamma\|_\infty$ denote its Frobenius norm and entrywise L_∞ norm, respectively. We use $\|\Gamma\|_*$ to denote the nuclear norm of

Γ, which is defined as the summation of the singular values of Γ. For a nonzero matrix $\Gamma \in \mathbb{R}^{T\times N}$, we define two measures to control its spikeness and rank. First, we define the spikeness ratio as

$$\alpha_{sp}(\Gamma) \equiv \frac{\sqrt{NT}\,\|\Gamma\|_\infty}{\|\Gamma\|},$$

which satisfies $1 \leq \alpha_{sp}(\Gamma) \leq \sqrt{NT}$. The lower bound can be reached when all the entries of Γ are the same, and the upper bound can be reached when there is only one nonzero entry in Γ. Next, we define a tractable measure of how close Γ is to a low-rank matrix via the ratio

$$\beta_{ra}(\Gamma) \equiv \frac{\|\Gamma\|_*}{\|\Gamma\|}.$$

Note that $1 \leq \beta_{ra}(\Gamma) \leq \delta_{NT} \equiv \sqrt{N} \wedge \sqrt{T}$. Let $d = (N+T)/2$. Define the constraint set

$$\mathcal{C}_{NT}(c_0) \equiv \left\{ \Gamma \in \mathbb{R}^{N\times T}, \Gamma \neq 0 \mid \alpha_{sp}(\Gamma)\beta_{ra}(\Gamma) \leq \frac{1}{c_0}\sqrt{\frac{NT}{d\log d}} \right\}, \tag{B.5}$$

where c_0 is a universal constant. For a low rank matrix $\Gamma \in \mathcal{C}_{NT}(c_0)$, the constraint requires it to be not very spiky.

The following two theorems are needed to show that the probability of overselecting the number of factors is approaching zero.

Theorem B.4 *Let G be a $T\times N$ random matrix with all entries i.i.d. from the Bernoulli distribution with parameter $p \in (0,1)$. There are universal constants $c_0, c_1, c_2,$ and c_3 such that*

$$\left\| \frac{1}{\sqrt{p}}\Gamma \circ G \right\| \geq \frac{1}{8}\|\Gamma\| \left\{ 1 - \frac{c_3 \alpha_{sp}(\Gamma)}{\sqrt{NT}} \right\} \qquad \text{for all } \Gamma \in \mathcal{C}_{NT}(c_0)$$

with probability greater than $1 - c_1\exp(-c_2 NT\log d/d)$.

Theorem B.5 *Let G be a $T\times N$ random matrix with all entries i.i.d. from the Bernoulli distribution with parameter $p \in (0,1)$. Then*

$$\sup_{\Gamma\in\mathcal{C}_{1NT}} \|\Gamma \circ [G - E(G)]\|_{sp} = O_P\left(c_{1NT} + c_{2NT} + c_{3NT}\sqrt{(N+T)\log\log(N+T)} + 1/\log(N+T) \right),$$

where $\mathcal{C}_{1NT} \equiv \mathcal{C}_{1NT}(c_{1NT}, c_{2NT}, c_{3NT}) \equiv \{\Gamma \in \mathbb{R}^{N\times T}, \mid \Gamma = UV',\ U \in \mathbb{R}^T$ and $V \in \mathbb{R}^N$ are vectors such that $\|U\| = \|V\| = 1,\ \|U\|_\infty \leq c_{1NT},\ \|V\|_\infty \leq c_{2NT},\ \|U\|_\infty \|\|V\|_\infty\| \leq c_{3NT}\}$.

Proof of Theorem 3.1. Noting that $X = C^0 + \varepsilon$, we make the following decomposition

$$\begin{aligned}
\widetilde{CV}(R) &= \frac{1}{NT}\left\|(X - \tilde{C}_R)\circ \bar{G}^*\right\|^2 \\
&= \frac{1}{NT}\left\|(C^0 - \tilde{C}_R)\circ \bar{G}^*\right\|^2 + \frac{1}{NT}\left\|\varepsilon\circ\bar{G}^*\right\|^2 + \frac{2}{NT}\mathrm{tr}\left\{\left[(C^0 - \tilde{C}_R)\circ\bar{G}^*\right]\left(\varepsilon\circ\bar{G}^*\right)'\right\} \\
&\equiv \widetilde{CV}_1(R) + \widetilde{CV}_2 + 2\widetilde{CV}_3(R),
\end{aligned}$$

where $\widetilde{CV}_2$ does not depend on R. Then we have

$$\widetilde{CV}(R) - \widetilde{CV}(R_0) = \left[\widetilde{CV}_1(R) - \widetilde{CV}_1(R_0)\right] + 2\left[\widetilde{CV}_3(R) - \widetilde{CV}_3(R_0)\right]. \tag{B.6}$$

It is sufficient to study the asymptotic properties of $\widetilde{CV}_1(R) - \widetilde{CV}_1(R_0)$ and $\widetilde{CV}_3(R) - \widetilde{CV}_3(R_0)$ under the under-fitted and over-fitted cases, respectively.

We first study the under-fitted case where $R < R_0$. Noting that $\|A\|^2 - \|B\|^2 = \text{tr}(A'A - B'B) = \text{tr}\{(A-B)'(A-B)\} + 2\text{tr}\big((A-B)'B\big)$, we have

$$\begin{aligned}
\widetilde{CV}_1(R) - \widetilde{CV}_1(R_0) &= \frac{1}{NT}\left\|(\tilde{C}_R - C^0)\circ\bar{G}^*\right\|^2 - \frac{1}{NT}\left\|(\tilde{C}_{R_0} - C^0)\circ\bar{G}^*\right\|^2 \\
&= \frac{1}{NT}\left\|(\tilde{C}_R - \tilde{C}_{R_0})\circ\bar{G}^*\right\|^2 + \frac{2}{NT}\text{tr}\left\{\left[(\tilde{C}_R - \tilde{C}_{R_0})\circ\bar{G}^*\right]'\left[(\tilde{C}_{R_0} - C^0)\circ\bar{G}^*\right]\right\} \\
&\equiv \widetilde{CV}_{11}(R) + 2CV_{12}(R).
\end{aligned} \tag{B.7}$$

Noting that $\tilde{C}_{R_0} - \tilde{C}_R = \sum_{r=R+1}^{R_0}\tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r$, $\tilde{u}_r = \tilde{U}_{R_0}e_{rR_0}$, $\tilde{v}_r = \tilde{V}_{R_0}e_{rR_0}$, and $\breve{\sigma}_r = (NT)^{-1/2}\tilde{\sigma}_r$, we have

$$\begin{aligned}
\widetilde{CV}_{11}(R) &= \frac{1}{NT}\left\|\left(\sum_{r=R+1}^{R_0}\tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r\right)\circ\bar{G}^*\right\|^2 \\
&= \frac{1}{NT}\left\|\left(\sum_{r=R+1}^{R_0}\tilde{U}_{R_0}e_{rR_0}e_{rR_0}'\tilde{V}_{R_0}'\tilde{\sigma}_r\right)\circ\bar{G}^*\right\|^2 \\
&= \frac{1}{NT}\left\|\left(\sum_{r=R+1}^{R_0}\left(\sqrt{N}\tilde{U}_{R_0}\tilde{D}_{R_0}\right)\tilde{D}_{R_0}^{-1}e_{rR_0}e_{rR_0}'\tilde{D}_{R_0}^{-1}(\sqrt{T}\tilde{V}_{R_0}\tilde{D}_{R_0})'\breve{\sigma}_r\right)\circ\bar{G}^*\right\|^2.
\end{aligned} \tag{B.8}$$

Let $\varsigma_{1R} = \sqrt{N}\tilde{U}_R\tilde{D}_R - F^0\breve{H}_{1R}$ and $\varsigma_{2R} = \sqrt{N}\tilde{V}_R\tilde{D}_R - \Lambda^0\breve{H}_{2R}$. Then $\sqrt{N}\tilde{U}_{R_0}\tilde{D}_{R_0} = F^0\breve{H}_{1R_0} + \varsigma_{1R_0}$

and $\sqrt{N}\tilde{V}_{R_0}\tilde{D}_{R_0} = \Lambda^0 \breve{H}_{2R_0} + \varsigma_{2R_0}$. It is easy to apply Lemma B.1 to show that

$$
\begin{aligned}
&\widetilde{CV}_{11}(R) \\
=\ & \frac{1}{NT}\left\|\left(\sum_{r=R+1}^{R_0}\left(F^0\breve{H}_{1R_0}+\varsigma_{1R_0}\right)\tilde{A}_{rR_0}(\Lambda^0\breve{H}_{2R_0}+\varsigma_{2R_0})'\breve{\sigma}_r\right)\circ\bar{G}^*\right\|^2 \\
=\ & \frac{1}{NT}\left\|\left(\sum_{r=R+1}^{R_0}F^0\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}'\Lambda^{0\prime}\breve{\sigma}_r\right)\circ\bar{G}^*\right\|^2+O_P\left(\delta_{NT}^{-1}\right) \\
=\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{r=R+1}^{R_0}e_{iT}'F^0\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}'\Lambda^{0\prime}e_{iN}\breve{\sigma}_{R_0}\right)^2\bar{g}_{it}^*+O_P\left(\delta_{NT}^{-1}\right) \\
=\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=R+1}^{R_0}\sum_{l=R+1}^{R_0}\operatorname{tr}\left\{\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}'\Lambda^{0\prime}e_{iN}e_{iN}'\Lambda^0\breve{H}_{2R_0}\tilde{A}_{lR_0}'\breve{H}_{1R_0}'F^0e_{tT}e_{tT}'F^{0\prime}\right\}\breve{\sigma}_r\breve{\sigma}_l\bar{g}_{it}^* \\
& +O_P\left(\delta_{NT}^{-1}\right) \\
=\ & \sum_{r=R+1}^{R_0}\sum_{l=R+1}^{R_0}[\operatorname{vec}(\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}')]'\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[(\Lambda^{0\prime}e_{iN}e_{iN}'\Lambda^0)\otimes(F^{0\prime}e_{tT}e_{tT}'F^{0\prime})]\bar{g}_{it}^*\right\}\breve{\sigma}_r\breve{\sigma}_l \\
& \times\operatorname{vec}(\breve{H}_{1R_0}\tilde{A}_{lR_0}\breve{H}_{2R_0}')+O_P\left(\delta_{NT}^{-1}\right) \qquad \text{(B.9)}
\end{aligned}
$$

where $\tilde{A}_{rR} = \tilde{D}_R^{-1}e_{rR}e_{rR}'\tilde{D}_R^{-1}$, $\bar{g}_{it}^* = \mathbf{1}\{(i,t)\in\Omega_\perp^*\}$, and the last equality follows from the fact that $\operatorname{tr}(A_1A_2A_3A_4) = [\operatorname{vec}(A_1)]'(A_2\otimes A_4')\operatorname{vec}(A_3')$ and the Fubini theorem. Now using $\bar{g}_{it}^* = (1-p) + [\bar{g}_{it}^* - (1-p)]$ and the fact that $\bar{g}_{it}^*$ are i.i.d. and independent of $\left(\Lambda^{0\prime}, F^{0\prime}\right)$, we can readily show that

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[(\Lambda^0e_{iN}'e_{iN}'\Lambda^{0\prime})\otimes(F^0e_{tT}e_{tT}'F^{0\prime})]\bar{g}_{it}^* &= \frac{1-p}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[(\Lambda^0e_{iN}'e_{iN}'\Lambda^{0\prime})\otimes(F^0e_{tT}e_{tT}'F^{0\prime})] \\
&\quad +O_P((NT)^{-1/2}).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\widetilde{CV}_{11}(R) &= (1-p)\sum_{r=R+1}^{R_0}\sum_{l=R+1}^{R_0}[\operatorname{vec}(\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}')]'\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[(\Lambda^{0\prime}e_{iN}e_{iN}'\Lambda^{0\prime})\otimes(F^{0\prime}e_{tT}e_{tT}'F^{0\prime})]\right\} \\
&\quad \times\operatorname{vec}(\breve{H}_{1R_0}\tilde{A}_{lR_0}\breve{H}_{2R_0}')\breve{\sigma}_r\breve{\sigma}_l+O_P\left(\delta_{NT}^{-1}\right) \\
&= \frac{1-p}{NT}\left\|\sum_{r=R+1}^{R_0}\tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r\right\|^2+O_P\left(\delta_{NT}^{-1}\right) = (1-p)\sum_{r=R+1}^{R_0}(NT)^{-1}\tilde{\sigma}_r^2+O_P\left(\delta_{NT}^{-1}\right) \\
&= (1-p)\sum_{r=R+1}^{R_0}\sigma_r^2+O_P\left(\delta_{NT}^{-1}\right),
\end{aligned}
$$

where the second equality is obtained by reversing the operations in (B.9) and (B.8), the third equality holds by the fact that $\tilde{U}_R'\tilde{U}_R = I_R$ and $\tilde{V}_R'\tilde{V}_R = I_R$, and the fourth equality follows because $(NT)^{-1}\tilde{\sigma}_r^2 = \sigma_r^2 + O_P(\delta_{NT}^{-1})$ for $r \le R_0$ by Lemma B.2(i).

Following the proof of Theorem 2.4, we can show that $\frac{1}{NT}\left\|(C^0-\tilde{C}_{R_0})\circ\bar{G}^*\right\|^2\leq\frac{1}{NT}\left\|C^0-\tilde{C}_{R_0}\right\|^2=O_P\left(\delta_{NT}^{-2}\right)$. Then by the Chebyshev inequality,

$$\begin{aligned}\left|\widetilde{CV}_{12}(R)\right| &= \frac{1}{NT}\left|\operatorname{tr}\left\{\left[\left(\tilde{C}_R-\tilde{C}_{R_0}\right)\circ\bar{G}^*\right]'\left[\left(C^0-\tilde{C}_{R_0}\right)\circ\bar{G}^*\right]\right\}\right| \\ &\leq \left\{\frac{1}{NT}\left\|\left(\tilde{C}_R-\tilde{C}_{R_0}\right)\circ\bar{G}^*\right\|^2\right\}^{1/2}\left\{\frac{1}{NT}\left\|\left(C^0-\tilde{C}_{R_0}\right)\circ\bar{G}^*\right\|^2\right\}^{1/2} \\ &= O_P(1)\,O_P\left(\delta_{NT}^{-1}\right)=O_P\left(\delta_{NT}^{-1}\right).\end{aligned}$$

It follows that $\widetilde{CV}_1(R)-\widetilde{CV}_1(R_0)=(1-p)\sum_{r=R+1}^{R_0}\sigma_r^2+O_P\left(\delta_{NT}^{-1}\right).$

Next, $\widetilde{CV}_3(R)-\widetilde{CV}_3(R_0)=\frac{1}{NT}\operatorname{tr}\{[(\tilde{C}_{R_0}-\tilde{C}_R)\circ\bar{G}^*]\left(\varepsilon\circ\bar{G}^*\right)'\}$. Noting that $\frac{1}{NT}\left\|\varepsilon\circ\bar{G}^*\right\|^2\leq\frac{1}{NT}\left\|\varepsilon\right\|^2=O_P(1)$, we can readily apply Lemma B.1 and follow the analysis of $\widetilde{CV}_{11}(R)$ to show that

$$\begin{aligned}&\widetilde{CV}_3(R)-\widetilde{CV}_3(R_0)\\ &= \frac{1}{NT}\operatorname{tr}\left\{\left(\sum_{r=R+1}^{R_0}\left[\left(F^0\breve{H}_{1R_0}+\varsigma_{1R_0}\right)\tilde{A}_{rR_0}(\Lambda^0\breve{H}_{2R_0}+\varsigma_{2R_0})'\breve{\sigma}_r\right]\circ\bar{G}^*\right)\left(\varepsilon\circ\bar{G}^*\right)'\right\}\\ &= \frac{1}{NT}\operatorname{tr}\left\{\left(\sum_{r=R+1}^{R_0}\left[F^0\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}'\Lambda^{0\prime}\breve{\sigma}_r\right]\circ\bar{G}^*\right)\left(\varepsilon\circ\bar{G}^*\right)'\right\}+O_P\left(\delta_{NT}^{-1}\right)\\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=R+1}^{R_0}\operatorname{tr}\left(e_{iT}'F^0\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}'\Lambda^{0\prime}e_{iN}\right)\breve{\sigma}_r\varepsilon_{it}\bar{g}_{it}^*+O_P\left(\delta_{NT}^{-1}\right)\\ &= \sum_{r=R+1}^{R_0}\operatorname{tr}\left(\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}'\,\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda^{0\prime}e_{iN}e_{iT}'F^0\varepsilon_{it}\bar{g}_{it}^*\right)\breve{\sigma}_r+O_P\left(\delta_{NT}^{-1}\right)\\ &= (1-p)\sum_{r=R+1}^{R_0}\operatorname{tr}\left(\breve{H}_{1R_0}\tilde{A}_{rR_0}\breve{H}_{2R_0}'\,\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda^{0\prime}e_{iN}e_{iT}'F^0\varepsilon_{it}\right)\breve{\sigma}_r+O_P\left(\delta_{NT}^{-1}\right)\\ &= O_P((NT)^{-1/2})+O_P\left(\delta_{NT}^{-1}\right)=O_P\left(\delta_{NT}^{-1}\right),\end{aligned}$$

where the last line follows from the fact that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda^{0\prime}e_{iN}e_{iT}'F^0\varepsilon_{it}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i^0F_t^{0\prime}\varepsilon_{it}=O_P((NT)^{-1/2})$.

In sum, we have shown that when $R<R_0$, $\widetilde{CV}(R)-\widetilde{CV}(R_0)=(1-p)\sum_{r=R+1}^{R_0}\sigma_r^2+O_P\left(\delta_{NT}^{-1}\right).$ This implies that $P(\tilde{R}<R_0)\to 0$ as $(N,T)\to\infty$.

Now, we study the overfitted case where $R>R_0$. We continue to use the decompositions in (B.6) and (B.7). We first study $\widetilde{CV}_{11}(R)$. When $R>R_0$, $\tilde{D}_R^{-1}\neq O_P(1)$ and thus $\tilde{A}_{rR}\neq O_P(1)$. This implies that we cannot use similar arguments as used in the case where $R<R_0$. In addition, $\tilde{C}_R-\tilde{C}_{R_0}$ is not independent of $\bar{G}^*$, which further complicates the analysis. To tackle the problem, we call upon Assumption A.7. Let $\tilde{\Gamma}_R\equiv\tilde{C}_R-\tilde{C}_{R_0}$. By Assumption A.7(i), we have $\left\|\tilde{\Gamma}_R\right\|_\infty\leq$

$\sum_{r=R_0+1}^{R}\tilde{\sigma}_r/(c_0\sqrt{(N+T)\log(N+T)})$ with probability approaching 1 (w.p.a.1). In addition, by the definitions of Frobenius and nuclear norms, $\left\|\tilde{\Gamma}_R\right\|=(\sum_{r=R_0+1}^{R}\tilde{\sigma}_r^2)^{1/2}$ and $\left\|\tilde{\Gamma}_R\right\|_*=\sum_{r=R_0+1}^{R}\tilde{\sigma}_r$. By the Jensen inequality and the fact that $R\leq R_{\max}$,

$$\frac{\sqrt{NT}\left\|\tilde{\Gamma}_R\right\|_\infty\left\|\tilde{\Gamma}_R\right\|_*}{\left\|\tilde{\Gamma}_R\right\|^2}\leq\frac{R_{\max}-R_0}{c_0}\sqrt{\frac{NT}{(N+T)\log(N+T)}}\leq\frac{1}{\tilde{c}_0}\sqrt{\frac{NT}{d_{NT}\log d_{NT}}},$$

where $d_{NT}=\frac{1}{2}(N+T)$ and $\tilde{c}_0=\sqrt{2}c_0/(R_{\max}-R_0)$. Therefore, $\tilde{\Gamma}_R\in\mathcal{C}_{NT}(\tilde{c}_0)$ w.p.a.1. Then we can apply Theorem B.4 and the fact that $\left\|\tilde{\Gamma}_R\right\|_\infty/\left\|\tilde{\Gamma}_R\right\|=o_P(1)$ to obtain that

$$\left\|\frac{1}{\sqrt{1-p}}\tilde{\Gamma}_R\circ\bar{G}^*\right\|\geq\frac{1}{16}\left\|\tilde{\Gamma}_R\right\|\text{ w.p.a.1.}$$

It follows that $\widetilde{CV}_{11}(R)=\frac{1}{NT}\left\|\tilde{\Gamma}_R\circ\bar{G}^*\right\|^2\geq\frac{1-p}{256}\frac{1}{NT}\left\|\tilde{\Gamma}_R\right\|^2=\frac{1-p}{256}\sum_{r=R_0+1}^{R}\breve{\sigma}_r^2$ w.p.a.1, where $\breve{\sigma}_r^2=O_P\left(\delta_{NT}^{-2}\right)$ for $r=R_0+1,...,R_{\max}$ by Lemma B.2(ii). Then by Lemma B.2(iii) we have $\text{plim}_{(N,T)\to\infty}\delta_{NT}^2\widetilde{CV}_{11}(R)\geq(R-R_0)\frac{1-p}{256}c_\sigma>0$.

Next, we study $\widetilde{CV}_{12}(R)$. Noting that $\tilde{\Gamma}_R=\tilde{C}_R-\tilde{C}_{R_0}=\sum_{r=R_0+1}^{R}\tilde{u}_r\tilde{v}_r'\tilde{\sigma}_r$, we have

$$\begin{aligned}\widetilde{CV}_{12}(R)&=\frac{1}{NT}\text{tr}\left\{(\tilde{\Gamma}_R\circ\bar{G}^*)\left[(\tilde{C}_{R_0}-C^0)\circ\bar{G}^*\right]'\right\}=\sum_{r=R_0+1}^{R}\frac{\tilde{\sigma}_r}{NT}\text{tr}\left\{(\tilde{u}_r\tilde{v}_r')'\left[(\tilde{C}_{R_0}-C^0)\circ\bar{G}^*\right]\right\}\\&\equiv\sum_{r=R_0+1}^{R}CV_{12r}.\end{aligned}$$

In addition,

$$\frac{1}{\sqrt{NT}}\text{tr}\left\{(\tilde{u}_r\tilde{v}_r')'(\tilde{C}_{R_0}-C^0)\right\}=\frac{-1}{\sqrt{NT}}\text{tr}\left\{(\tilde{u}_r\tilde{v}_r')'C^0\right\}=\frac{-1}{\sqrt{NT}}\text{tr}\left\{\tilde{u}_r'F^0\Lambda^{0\prime}\tilde{v}_r\right\}=O_P(\delta_{NT}^{-4}),$$

where the first equality holds by the orthogonality between $\tilde{u}_r$ and $\tilde{C}_{R_0}$ for $r>R_0$ and the third equality holds by Lemma B.3. It follows that

$$\begin{aligned}CV_{12r}&=\frac{\tilde{\sigma}_r}{NT}\text{tr}\left\{(\tilde{u}_r\tilde{v}_r')'\left[(\tilde{C}_{R_0}-C^0)\circ\bar{G}^*\right]\right\}\\&=-\frac{\tilde{\sigma}_r}{NT}\text{tr}\left\{(\tilde{u}_r\tilde{v}_r')'\left[(\tilde{C}_{R_0}-C^0)\circ(G^*-E(G^*))\right]\right\}+O_P(\delta_{NT}^{-4})\\&\equiv\overline{CV}_{12r}+O_P(\delta_{NT}^{-4}).\end{aligned}$$

Note that

$$\begin{aligned}\left|\overline{CV}_{12r}\right|&=\frac{\tilde{\sigma}_r}{\sqrt{NT}}\frac{1}{\sqrt{NT}}\left|\text{tr}\left\{(\tilde{C}_{R_0}-C^0)\left[(\tilde{u}_r\tilde{v}_r')\circ(G^*-E(G^*))\right]\right\}\right|\\&\leq O_P\left(\delta_{NT}^{-1}\right)\frac{1}{\sqrt{NT}}\left\|\tilde{C}_{R_0}-C^0\right\|_*\left\|(\tilde{u}_r\tilde{v}_r')\circ(G^*-E(G^*))\right\|_{\text{sp}}\\&\leq O_P\left(\delta_{NT}^{-2}\right)\sup_{\Gamma\in\mathcal{C}_{1NT}(c_{1NT},c_{2NT},c_{3NT})}\left\|\Gamma\circ(G^*-E(G^*))\right\|_{\text{sp}}\\&=o_P\left(\delta_{NT}^{-2}\right),\end{aligned}$$

where first inequality follows the fact that $\frac{\tilde{\sigma}_r}{\sqrt{NT}} = O_P\left(\delta_{NT}^{-1}\right)$ and $|\text{tr}\,(AB)| \leq \|A\|_* \|B\|_{\text{sp}}$, the second inequality follows because

$$\frac{1}{\sqrt{NT}} \left\| \tilde{C}_{R_0} - C^0 \right\|_* \leq \frac{\sqrt{2R_0}}{\sqrt{NT}} \left\| \tilde{C}_{R_0} - C^0 \right\| = O_P\left(\delta_{NT}^{-1}\right)$$

and the last equality holds by Theorem B.5 with $c_{1NT} = o(1)$, $c_{2NT} = o(1)$ and $c_{3NT} = 1/\sqrt{(N+T)\log(N+T)}$. Then we have $\widetilde{CV}_{12}(R) = o_P\left(\delta_{NT}^{-2}\right)$.

Now, we study $\widetilde{CV}_3(R) - \widetilde{CV}_3(R_0)$.

$$\begin{aligned}
\widetilde{CV}_3(R) - \widetilde{CV}_3(R_0) &= \frac{1}{NT}\text{tr}\left\{\left[\left(\tilde{C}_{R_0} - \tilde{C}_R\right) \circ \bar{G}^*\right] (\varepsilon \circ \bar{G}^*)'\right\} = -\sum_{r=R_0+1}^{R} \frac{\tilde{\sigma}_r}{NT}\text{tr}\left\{(\tilde{u}_r \tilde{v}_r')' \left(\varepsilon \circ \bar{G}^*\right)\right\} \\
&= -\sum_{r=R_0+1}^{R} \frac{\tilde{\sigma}_r}{NT} \tilde{u}_r' \left(\varepsilon \circ \bar{G}^*\right) \tilde{v}_r = -\sum_{r=R_0+1}^{R} \frac{\tilde{\sigma}_r}{NT} \sum_{i,t} \tilde{u}_{tr} \tilde{v}_{ir} \varepsilon_{it} (1 - g_{it}^*) \\
&\equiv -\sum_{r=R_0+1}^{R} CV_{3r}.
\end{aligned}$$

where $\tilde{u}_{tr}$ and $\tilde{v}_{ir}$ denote the tth and ith entries of $\tilde{u}_r$ and $\tilde{v}_r$, respectively. Noting that $\tilde{\sigma}_r^2/(NT) = O_P(\delta_{NT}^{-2})$, we have

$$\begin{aligned}
E_{\mathcal{D}_{NT}}\left[CV_{3r}^2\right] &= \frac{\tilde{\sigma}_r^2}{NT} \frac{1}{NT} \sum_{(i,t)\in\Omega_\perp^*} \sum_{(j,s)\in\Omega_\perp^*} \tilde{u}_{tr} \tilde{v}_{ir} \tilde{u}_{sr} \tilde{v}_{jr} E_{\mathcal{D}_{NT}}(\varepsilon_{it}\varepsilon_{js}) \\
&\leq O_P(\delta_{NT}^{-2}) \frac{1}{2NT} \sum_{(i,t)\in\Omega_\perp^*} \sum_{(j,s)\in\Omega_\perp^*} (\tilde{u}_{tr}^2 \tilde{v}_{ir}^2 + \tilde{u}_{sr}^2 \tilde{v}_{jr}^2) \left|E_{\mathcal{D}_{NT}}(\varepsilon_{it}\varepsilon_{js})\right| \\
&= O_P(\delta_{NT}^{-2}) \frac{1}{NT} \sum_{(i,t)\in\Omega_\perp^*} \tilde{u}_{tr}^2 \tilde{v}_{ir}^2 \sum_{(j,s)\in\Omega_\perp^*} \left|E_{\mathcal{D}_{NT}}(\varepsilon_{it}\varepsilon_{js})\right| \\
&\leq O_P(\delta_{NT}^{-2}) \frac{1}{NT} \max_{(i,t)\in\Omega_\perp^*} \sum_{(j,s)\in\Omega_\perp^*} \left|E_{\mathcal{D}_{NT}}(\varepsilon_{it}\varepsilon_{js})\right| = o_p(\delta_{NT}^{-4}),
\end{aligned}$$

where $E_{\mathcal{D}_{NT}}(\cdot) = E\left(\cdot | P_{\Omega^*}X, \Omega^*\right)$, the first inequality holds by the Cauchy-Schwarz inequality, the second inequality holds by the fact that $\sum_{(i,t)\in\Omega_\perp^*} \tilde{u}_{tr}^2 \tilde{v}_{ir}^2 \leq \|\tilde{u}_r\|^2 \|\tilde{v}_r\|^2 = 1$, and the last equality holds by Assumption 7(ii). Hence, $CV_{3r} = o_p(\delta_{NT}^{-2})$ for each $r \in (R_0, R]$ and $\widetilde{CV}_3(R) - \widetilde{CV}_3(R_0) = o_p(\delta_{NT}^{-2})$. It follows that

$$\text{plim}_{(N,T)\to\infty} \delta_{NT}^2 \left[\widetilde{CV}(R) - \widetilde{CV}(R_0)\right] \geq \frac{(R - R_0)(1-p)}{256} c_\sigma > 0 \text{ for any } R > R_0.$$

This implies that $P(\tilde{R} > R_0) \to 0$ as $(N,T) \to \infty$. This completes the proof of the theorem. ■

Proof of Theorem 3.2. The proof is essentially the same as that of Theorem 3.1 given the results in Theorem 2.4. Here, we only outline the major differences. Let $\hat{X}^* = \hat{X}^{*(\ell^*)}$. Noting that $\hat{C}_R = S_H(\hat{X}^*, R) = \hat{U}_R \hat{\Sigma}_R \hat{V}_R'$, $\hat{U}_R$ and $\hat{V}_R$ are respectively the eigenvector matrices of $\hat{X}^* \hat{X}^{*\prime}$ and

$\hat{X}^{*\prime}\hat{X}^{*}$ associated with their R largest eigenvalues, and the diagonal elements of $\hat{\Sigma}_R^2$ are the R largest eigenvalues of $\hat{X}^{*}\hat{X}^{*\prime}$. Let $\hat{F}^R$ and $\hat{\Lambda}^R$ denote the conventional principal component (PC) estimators of $F(R)$ and $\Lambda(R)$ based on $\hat{X}^{*}$ under the normalization restrictions that $T^{-1}F(R)'F(R) = I_R$ and $\Lambda(R)'\Lambda(R)$ =diagonal matrix. Let $\ddot{F}^R$ and $\ddot{\Lambda}^R$ denote the conventional PC estimators of $F(R)$ and $\Lambda(R)$ based on $\hat{X}^{*}$ under the normalization restrictions that $N^{-1}\Lambda(R)'\Lambda(R) = I_R$ and $F(R)'F(R)$ =diagonal matrix. Define

$$\ddot{H}_{1R} = (N^{-1}\Lambda^{0\prime}\Lambda^0)(T^{-1}\Lambda^{0\prime}\ddot{F}^R) \text{ and } \ddot{H}_{2R} = (T^{-1}F^{0\prime}F^0)(N^{-1}F^{0\prime}\ddot{\Lambda}^R).$$

Let $\hat{D}_R$ denote the $R\times R$ diagonal matrix that contains the R largest eigenvalues of $(NT)^{-1}\hat{X}^{*}\hat{X}^{*\prime}$ arranged in descending order along its diagonal line. Note that $\hat{D}_R = (NT)^{-1}\hat{\Sigma}_R^2$.

Following the proof of Theorem 2.4, we can show that $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\|\hat{C}_{R_{\max},it}^{(\ell^*-1)} - C_{it}^0\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. With this result, we can show that the results analogous to those in Lemmas B.1-B.2 hold:

(i) $\frac{1}{T}\left\|\sqrt{T}\hat{U}_R\hat{D}_R - F^0\ddot{H}_{1R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$,

(ii) $\frac{1}{N}\left\|\sqrt{N}\hat{V}_R\hat{D}_R - \Lambda^0\ddot{H}_{2R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$,

(iii) $\ddot{\sigma}_r^2 = \sigma_r^2 + O_P\left(\delta_{NT}^{-1}\right)$ for $r = 1, ..., R_0$,

(iv) $\ddot{\sigma}_{R_0+r}^2 = O_P\left(\delta_{NT}^{-2}\right)$ for $r \geq 1$,

(v) $\delta_{NT}^2\ddot{\sigma}_{R_0+r}^2 \geq c_\sigma + o_P(1)$ for some positive constant c_σ and any $r \geq 1$ with $R_0 + r \leq R$.

where $\ddot{\sigma}_r = (NT)^{-1/2}\hat{\sigma}_r$.

Noting that $X = C^0 + \varepsilon$, we make the following decomposition

$$\begin{aligned}\widehat{CV}(R) &= \frac{1}{NT}\left\|(X - \hat{C}_R)\circ\bar{G}^*\right\|^2 \\ &= \frac{1}{NT}\left\|(C^0 - \hat{C}_R)\circ\bar{G}^*\right\|^2 + \frac{1}{NT}\left\|\varepsilon\circ\bar{G}^*\right\|^2 + \frac{2}{NT}\text{tr}\left\{\left[(C^0 - \hat{C}_R)\circ\bar{G}^*\right]\left(\varepsilon\circ\bar{G}^*\right)'\right\} \\ &\equiv \widehat{CV}_1(R) + \widehat{CV}_2 + 2\widehat{CV}_3(R).\end{aligned}$$

Then we have $\widehat{CV}(R) - \widehat{CV}(R_0) = [\widehat{CV}_1(R) - \widehat{CV}_1(R_0)] + 2[\widehat{CV}_3(R) - \widehat{CV}_3(R_0)]$. When $R < R_0$, we can follow the proof of Theorem 3.1 and apply the above results in (i)-(iii) to show that

$$\widehat{CV}_1(R) - \widehat{CV}_1(R_0) = (1-p)\sum_{r=R+1}^{R_0}\sigma_r^2 + O_P\left(\delta_{NT}^{-1}\right) \text{ and } \widehat{CV}_3(R) - \widehat{CV}_3(R_0) = O_P\left(\delta_{NT}^{-1}\right).$$

Then $\widehat{CV}(R) - \widehat{CV}(R_0) = (1-p)\sum_{r=R+1}^{R_0}\sigma_r^2 + O_P\left(\delta_{NT}^{-1}\right)$ and $P(\hat{R} < R_0) \to 0$ as $(N,T) \to \infty$.

Similarly, when $R > R_0$, we can follow the proof of Theorem 3.1 and apply the above results in (i)-(ii) and (iv)-(v) and analogous results to those in Theorems B.4-B.5 to show that

$$\widehat{CV}_1(R) - \widehat{CV}_1(R_0) \geq \frac{(1-p)}{256}\sum_{r=R_0+1}^{R}\ddot{\sigma}_r^2 + O_P\left(\delta_{NT}^{-3}\right) \text{ and } \widehat{CV}_3(R) - \widehat{CV}_3(R_0) = o_P\left(\delta_{NT}^{-2}\right).$$

Then $\text{plim}_{(N,T)\to\infty}\delta_{NT}^2[\widehat{CV}(R) - \widehat{CV}(R_0)] \geq \frac{(R-R_0)(1-p)}{256}c_\sigma > 0$ and $P(\hat{R} > R_0) \to 0$ as $(N,T) \to \infty$. This completes the proof of the theorem. ■

Proof of Lemma B.1. (i) Following the proof of Theorem 1 in Bai and Ng (2002) and that of Theorem 2.1, we can readily show that

$$\frac{1}{T}\left\|\breve{F}^R - F^0\breve{H}_{1R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right). \tag{B.10}$$

Recall that $\tilde{D}_R$ denotes the $R\times R$ diagonal matrix that contains the R largest eigenvalues of $(NTp^2)^{-1}X^*X^{*\prime}$ arranged in descending order along its diagonal line. Then $(NTp^2)^{-1}X^*X^{*\prime}\tilde{U}_R = \tilde{U}_R\tilde{D}_R$. This, along with the definition that $\breve{F}^R = \left(NTp^2\right)^{-1}X^*X^{*\prime}\tilde{F}^R$ and the fact that $\tilde{F}^R = \sqrt{T}\tilde{U}_R$, implies that

$$\breve{F}^R = \sqrt{T}\left(NTp^2\right)^{-1}X^*X^{*\prime}\tilde{U}_R = \sqrt{T}\tilde{U}_R\tilde{D}_R.$$

Then by (B.10), we have $\frac{1}{T}\left\|\sqrt{T}\tilde{U}_R\tilde{D}_R - F^0\breve{H}_{1R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$.

(ii) Following the proof of Theorem 1 in Bai and Ng (2002) and that of Theorem 2.1, we can readily show that $\frac{1}{N}\left\|\breve{\Lambda}^R - \Lambda^0\breve{H}_{2R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. Noting that $(NTp^2)^{-1}X^{*\prime}X^*\tilde{V}_R = \tilde{V}_R\tilde{D}_R$ and $\bar{\Lambda}^R = \sqrt{N}\tilde{V}_R$, we have

$$\breve{\Lambda}^R = \left(NTp^2\right)^{-1}X^*X^{*\prime}\bar{\Lambda}^R = \sqrt{N}\left(NTp^2\right)^{-1}X^*X^{*\prime}\tilde{V}_R = \sqrt{N}\tilde{V}_R\tilde{D}_R.$$

It follows that $\frac{1}{N}\left\|\sqrt{N}\tilde{V}_R\tilde{D}_R - \Lambda^0\breve{H}_{2R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. ■

Proof of Lemma B.2. (i) Note that $\breve{\sigma}_r^2 = (NT)^{-1}\tilde{\sigma}_r^2$ denotes the rth largest eigenvalue of $(NTp^2)^{-1}X^*X^{*\prime}$. In view of that $X^* = X\circ G^* = \left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G^*$, we have

$$\begin{aligned}
(NTp^2)^{-1}X^*X^{*\prime} &= \frac{1}{NTp^2}\left[\left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G^*\right]\left[\left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G^*\right]' \\
&= \frac{1}{NTp^2}\left[\left(F^0\Lambda^{0\prime}\right)\circ G^*\right]\left[\left(F^0\Lambda^{0\prime}\right)\circ G^*\right]' + \frac{1}{NT}(\varepsilon\circ G^*)(\varepsilon\circ G^*)' \\
&\quad + \frac{1}{NT}\left[\left(F^0\Lambda^{0\prime}\right)\circ G^*\right](\varepsilon\circ G^*)' + \frac{1}{NT}(\varepsilon\circ G^*)\left[\left(F^0\Lambda^{0\prime}\right)\circ G^*\right]' \\
&\equiv IV_1 + IV_2 + IV_3 + IV_4.
\end{aligned}$$

As in the proof of Lemma A.1 and using Lemma B.9 below, it is easy to show that $\|\varepsilon\circ G^*\|_{\text{sp}} \le p\|\varepsilon\|_{\text{sp}} + \|\varepsilon\circ[G^* - p\mathbf{1}_{T\times N}]\|_{\text{sp}} = O_P(\sqrt{N}+\sqrt{T})$. Then

$$\begin{aligned}
\|IV_2\|_{\text{sp}} &\le \frac{1}{NTp^2}\|\varepsilon\circ G^*\|_{\text{sp}}^2 = O_P\left(\delta_{NT}^{-2}\right), \text{ and} \\
\|IV_3\|_{\text{sp}} &= \|IV_4\|_{\text{sp}} \le \frac{1}{p^2\sqrt{NT}}\left\|F^0\Lambda^{0\prime}\right\|\frac{1}{\sqrt{NT}}\|\varepsilon\circ G^*\|_{\text{sp}} = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

For IV_1, we use $G^* = p\mathbf{1}_{T\times N} + (G^* - p\mathbf{1}_{T\times N})$ and make the following decomposition,

$$\begin{aligned}
IV_1 &= \frac{1}{NT}F^0\Lambda^{0\prime}\Lambda^0F^{0\prime} + \frac{1}{NTp^2}\left[\left(F^0\Lambda^{0\prime}\right)\circ(G^* - p\mathbf{1}_{T\times N})\right]\left[\left(F^0\Lambda^{0\prime}\right)\circ(G^* - p\mathbf{1}_{T\times N})\right]' \\
&\quad + \frac{1}{NTp}\left(F^0\Lambda^{0\prime}\right)\left[\left(F^0\Lambda^{0\prime}\right)\circ(G^* - p\mathbf{1}_{T\times N})\right]' + \frac{1}{NTp}\left[\left(F^0\Lambda^{0\prime}\right)\circ(G^* - p\mathbf{1}_{T\times N})\right]\Lambda^0F^{0\prime} \\
&\equiv IV_{1,1} + IV_{1,2} + IV_{1,3} + IV_{1,4}.
\end{aligned}$$

Using Lemma B.9 and following the analysis of $\left\| \left(F^0 \Lambda^{0\prime} \right) \circ G \right\|$ in the proof of Lemma A.1, it is easy to show that $\left\| \left(F^0 \Lambda^{0\prime} \right) \circ (G^* - p\mathbf{1}_{T\times N}) \right\|_{\text{sp}} = O_P(\sqrt{N} + \sqrt{T})$, with which we can show that

$$\|IV_{1,2}\| = O_P\left(\delta_{NT}^{-2}\right) \text{ and } \|IV_{1,3}\| = \|IV_{1,4}\| = O_P\left(\delta_{NT}^{-1}\right).$$

Then by the Weyl's and triangular inequalities, we have

$$\left| \breve{\sigma}_r^2 - \mu_r \left(\frac{1}{NT} F^0 \Lambda^{0\prime} \Lambda^0 F^{0\prime} \right) \right| \leq \|IV_2 + IV_3 + IV_4 + IV_{1,2} + IV_{1,3} + IV_{1,4}\|_{\text{sp}} = O_P\left(\delta_{NT}^{-1}\right).$$

In addition, $\mu_r \left(\frac{1}{NT} F^0 \Lambda^{0\prime} \Lambda^0 F^{0\prime} \right) - \sigma_r^2 = O_P\left(\delta_{NT}^{-1}\right)$ under Assumption A.1(v). It follows that $\left| \breve{\sigma}_r^2 - \sigma_r^2 \right| = O_P\left(\delta_{NT}^{-1}\right)$.

(ii) Let $\varepsilon^* = \frac{1}{p} \varepsilon \circ G^*$, $C^* = \frac{1}{p} (F^0 \Lambda^{0\prime}) \circ [G^* - p\mathbf{1}_{T\times N}]$ and $\varsigma^* = C^* + \varepsilon^*$. Then

$$\frac{1}{p} X^* = \frac{1}{p} X \circ G^* = \frac{1}{p} \left(F^0 \Lambda^{0\prime} + \varepsilon \right) \circ G^* = F^0 \Lambda^{0\prime} + \varsigma^*.$$

Let $P_{\Lambda^0} = \Lambda^0 (\Lambda^{0\prime} \Lambda^0)^{-1} \Lambda^{0\prime}$ and $Q_{\Lambda^0} = I_N - P_{\Lambda^0}$. Let $F^* = F^0 + \varsigma^* \Lambda^0 (\Lambda^{0\prime} \Lambda^0)^{-1}$. Then

$$\frac{1}{NTp^2} X^* X^{*\prime} = \frac{1}{NT} F^{*\prime} \Lambda^{0\prime} \Lambda^0 F^* + \frac{1}{NT} \varsigma^* Q_{\Lambda^0} \varsigma^{*\prime}.$$

It follows that for any $r \geq 1$

$$\begin{aligned} \breve{\sigma}_{R_0+r}^2 &= \mu_{R_0+r} \left(\frac{1}{NTp^2} X^* X^{*\prime} \right) \leq \mu_{R_0+1} \left(\frac{1}{NT} F^{*\prime} \Lambda^{0\prime} \Lambda^0 F^* \right) + \mu_r \left(\frac{1}{NT} \varsigma^* Q_{\Lambda^0} \varsigma^{*\prime} \right) \\ &= \mu_r \left(\frac{1}{NT} \varsigma^* Q_{\Lambda^0} \varsigma^{*\prime} \right), \end{aligned}$$

where we use the fact that $\text{rank}(F^{*\prime} \Lambda^{0\prime} \Lambda^0 F^*) \leq R_0$. Using Lemma B.9, we can readily show that $\|\varsigma^*\|_{\text{sp}} = O_P(\sqrt{N} + \sqrt{T})$. Then

$$\mu_r \left(\frac{1}{NT} \varsigma^* Q_{\Lambda^0} \varsigma^{*\prime} \right) \leq \mu_r \left(\frac{1}{NT} \varsigma^* \varsigma^{*\prime} \right) \leq \frac{1}{NT} \|\varsigma^*\|_{\text{sp}}^2 = O_P\left(\delta_{NT}^{-2}\right).$$

It follows that $\breve{\sigma}_{R_0+r}^2 = O_P\left(\delta_{NT}^{-2}\right)$ for any $r \geq 1$.

(iii) To determine the lower probability bound for $\breve{\sigma}_{R_0+r}^2$, we notice that

$$\begin{aligned} \mu_{2R_0+r} \left(\frac{1}{NT} \varsigma^* \varsigma^{*\prime} \right) &\leq \mu_{R_0+r} \left(\frac{1}{NT} \varsigma^* Q_{\Lambda^0} \varsigma^{*\prime} \right) + \mu_{R_0+1} \left(\frac{1}{NT} \varsigma^* P_{\Lambda^0} \varsigma^{*\prime} \right) \\ &= \mu_{R_0+r} \left(\frac{1}{NT} \varsigma^* Q_{\Lambda^0} \varsigma^{*\prime} \right) \leq \mu_{R_0+r} \left(\frac{1}{NTp^2} X^* X^{*\prime} \right) = \breve{\sigma}_{R_0+r}^2. \end{aligned}$$

Without loss of generality we assume that $T \leq N$ and consider two cases: (1) T and N pass to infinity at the same rate (viz., $T \asymp N$), and (2) $T = o(N)$. In Case (1), we can follow the proof of Lemma A.9 in Ahn and Horenstein (2013) to show that $\delta_{NT}^2 \mu_{2R_0+r} \left(\frac{1}{NT} \varsigma^* \varsigma^{*\prime} \right)$ is bounded from below by a positive constant. In Case (2), we can consider the principal submatrix of ς^* and show that

$\delta_{NT}^2\mu_{2R_0+r}\left(\frac{1}{NT}\varsigma^*\varsigma^{*\prime}\right)$ is also bounded from below by a positive constant. It follows that $\delta_{NT}^2\breve{\sigma}_{R_0+r}^2$ is bounded in probability from below by a positive constant, say c_σ, as $(N,T)\to\infty$. ■

Proof of Lemma B.3. Let $r\geq R_0+1$. Recall from the proof of Theorem 3.1 that $\tilde{F}=\tilde{F}^{R_0}$ and $\tilde{H}=\tilde{H}_{R_0}$. Note that

$$\left\|\frac{\tilde{u}_r'F^0}{\sqrt{T}}\right\|=\left\|\frac{\tilde{u}_r'F^0\tilde{H}}{\sqrt{T}}\tilde{H}^{-1}\right\|=\left\|\frac{\tilde{u}_r'(F^0\tilde{H}-\tilde{F})}{\sqrt{T}}\tilde{H}^{-1}\right\|\leq\left\|\tilde{H}^{-1}\right\|\|\tilde{u}_r\|\left\|\frac{F^0\tilde{H}-\tilde{F}}{\sqrt{T}}\right\|=O_P(\delta_{NT}^{-1}),$$

where the second inequality is by orthogonality between $\tilde{u}_r$ and $\tilde{F}=\tilde{F}^{R_0}$ for $r>R_0$. Analogously, we can show that $\frac{\tilde{v}_r'\Lambda^0}{\sqrt{N}}=O_P(\delta_{NT}^{-1})$. In the following, we aim at improving the probability order to show that $\tilde{v}_r'\Lambda^0=O_P(\delta_{NT}^{-1})$ and $\tilde{u}_r'F^0=O_P(\delta_{NT}^{-1})$

By the definition of singular value decomposition (SVD), we can write $\frac{1}{p}X^*=\sum_{k=1}^{N\wedge T}\tilde{u}_k\tilde{v}_k'\tilde{\sigma}_k$. Recall that $\varsigma^*\equiv\varepsilon\circ G^*+F^0\Lambda^{0\prime}\circ[G^*-E(G^*)]/p$, $\frac{1}{p}X^*=F^0\Lambda^{0\prime}+\varsigma^*$, and $\tilde{u}_r$ denotes the rth eigenvector of $\frac{1}{p^2}X^*X^{*\prime}$ that is associated with its rth largest eigenvalue. If follows that

$$\left(\frac{F^0\Lambda^{0\prime}\Lambda^0F^{0\prime}}{NT}+\frac{F^0\Lambda^{0\prime}\varsigma^{*\prime}}{NT}+\frac{\varsigma^*\Lambda^0F^{0\prime}}{NT}+\frac{\varsigma^*\varsigma^{*\prime}}{NT}\right)\tilde{u}_r=\tilde{u}_r\frac{\tilde{\sigma}_r^2}{NT}.$$

Premultiplying both sides of the above equation by $F^{0\prime}/\sqrt{T}$, we have

$$\frac{F^{0\prime}F^0}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^{0\prime}\tilde{u}_r}{\sqrt{T}}+\frac{F^{0\prime}F^0}{T}\frac{\Lambda^{0\prime}\varsigma^{*\prime}\tilde{u}_r}{N\sqrt{T}}=O_P(\delta_{NT}^{-2}),$$

where we used the fact that $\frac{F^{0\prime}e\Lambda^0}{\sqrt{NT}}=O_P(1)$, $\frac{\tilde{\sigma}_r^2}{NT}=\breve{\sigma}_r^2=O_P(\delta_{NT}^{-2})$ for $r>R_0$, $\|\tilde{u}_r\|=1$, $\frac{1}{\sqrt{T}}\left\|F^{0\prime}\tilde{u}_r\right\|=O_P(\delta_{NT}^{-1})$ and $\left\|\varsigma^*/\sqrt{NT}\right\|_{\text{sp}}=O_P(\delta_{NT}^{-1})$. Premultiplying both sides of the above equation by $\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}$, we have

$$O_P(\delta_{NT}^{-2})=\frac{\Lambda^{0\prime}}{\sqrt{N}}\frac{\Lambda^0F^{0\prime}+\varsigma^{*\prime}}{\sqrt{NT}}\tilde{u}_r=\frac{\Lambda^{0\prime}}{\sqrt{N}}\frac{\frac{1}{p}X^{*\prime}\tilde{u}_r}{\sqrt{NT}}=\frac{\tilde{\sigma}_r}{\sqrt{NT}}\Lambda^{0\prime}\tilde{v}_r,$$

where the second equality follows from the decomposition $\frac{1}{p}X^*=F^0\Lambda^{0\prime}+\varsigma^*$ and the third one holds by the fact that $\frac{1}{p}X^{*\prime}\tilde{u}_r=\tilde{\sigma}_r\tilde{v}_r$. It follows that $\Lambda^{0\prime}\tilde{v}_r=O_P(\delta_{NT}^{-1})$ as $\frac{\delta_{NT}^{-}\tilde{\sigma}_r}{\sqrt{NT}}=\delta_{NT}^{-}\breve{\sigma}_r$ is bounded away from zero by Lemma B.2(iii). A symmetric argument gives that $\tilde{u}_r'F^0=O_P(\delta_{NT}^{-1})$. ■

Proof of Theorem B.4. The proof follows closely from that of Theorem 1 in Negahban and Wainwright (2012). It suffices to show the probability of the event

$$\mathbf{E}_{NT}\equiv\left\{\exists\,\Gamma\in\mathcal{C}_{NT}(c_0)\mid\left|\left\|\frac{1}{\sqrt{p}}\Gamma\circ G\right\|-\|\Gamma\|\right|>\frac{7}{8}\|\Gamma\|+\frac{c_3\|\Gamma\|_\infty}{8}\right\}$$

is bounded by $c_1\exp(-c_2d\log d)$. Note that the claimed result holds for $c\Gamma$ too if it holds for Γ. In addition, since $\mathcal{C}_{NT}(c_0)$ is invariant to the rescaling of Γ, without loss of generality, we can prove the result by assuming that $\|\Gamma\|_\infty = \frac{1}{d}$. For any $\Gamma \in \mathcal{C}_{NT}(c_0)$ with $\|\Gamma\|_\infty = \frac{1}{d}$ and $\|\Gamma\| \leq D$, we have $\|\Gamma\|_* \leq \rho(D)$, where $\rho(D) \equiv \frac{D^2\sqrt{d}}{c_0\sqrt{\log d}}$ by the definition of $\mathcal{C}_{NT}(c_0)$. For each radius $D > 0$, consider the set

$$\mathcal{B}(D) \equiv \left\{\Gamma \in \mathcal{C}_{NT}(c_0) \mid \|\Gamma\|_\infty = \frac{1}{d}, \|\Gamma\| \leq D, \|\Gamma\|_* \leq \rho(D)\right\},$$

and the associated event

$$\mathbf{E}_{NT,D} \equiv \left\{\exists\, \Gamma \in \mathcal{B}(D) \mid \left|\left\|\frac{1}{\sqrt{p}}\Gamma \circ G\right\| - \|\Gamma\|\right| \geq \frac{3}{4}D + \frac{c_3}{8d}\right\}.$$

Lemma B.6 below shows that it suffices to obtain the upper bound for the probability of the event $\mathbf{E}_{NT,D}$ for each fixed $D > 0$. In the second step, we show the probability of $\mathbf{E}_{NT,D}$ is bounded by $c_1 \exp(-c_2 D^2 NT)$ for some universal constants (c_1, c_2).

Now, define

$$Z_{NT}(D) \equiv \sup_{\Gamma \in \overline{\mathcal{B}}(D)} \left|\left\|\frac{1}{\sqrt{p}}\Gamma \circ G\right\| - \|\Gamma\|\right|,$$

where $\overline{\mathcal{B}}(D) \equiv \left\{\Gamma \in \mathcal{C}_{NT}(c_0) \mid \|\Gamma\|_\infty \leq \frac{1}{d}, \|\Gamma\| \leq D, \|\Gamma\|_* \leq \rho(D)\right\}$. It suffices to show that there are universal constants (c_1, c_2, c_3) such that

$$P\left[Z_{NT}(D) \geq \frac{3}{4}D + \frac{c_3}{8d}\right] \leq c_1 \exp(-c_2 D^2 NT) \text{ for each fixed } D > 0.$$

In order to prove the above result, we begin with a discretization argument. Let $\Gamma^1, \ldots, \Gamma^{N(\delta)}$ be a δ-covering of $\overline{\mathcal{B}}(D)$ in Frobenius norm. By definition, for any $\Gamma \in \overline{\mathcal{B}}(D)$, there exists some $k \in [N(\delta)]$ such that $\left\|\Gamma - \Gamma^k\right\| \leq \delta$. Let $\Delta \equiv \Gamma - \Gamma^k$. Then by the repeated use of the triangle inequality,

$$\begin{aligned}
\left\|\frac{1}{\sqrt{p}}\Gamma \circ G\right\| - \|\Gamma\| &= \left\|\frac{1}{\sqrt{p}}(\Gamma^k + \Delta) \circ G\right\| - \left\|\Gamma^k + \Delta\right\| \\
&\leq \left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\| + \left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| + \|\Delta\| \\
&\leq \left|\left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| + \left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| + \delta.
\end{aligned}$$

A symmetric argument gives the lower bound and establishes that this inequality holds for the absolute value of the difference:

$$\left|\left\|\frac{1}{\sqrt{p}}\Gamma \circ G\right\| - \|\Gamma\|\right| \leq \left|\left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| + \left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| + \delta.$$

Because both Γ and Γ^k belong to $\overline{\mathcal{B}}(D)$, we have that $\|\Delta\|_* \leq 2\rho(D)$ and $\|\Delta\|_\infty \leq 2/d$. Consequently, we have

$$Z_{NT}(D) \leq \delta + \max_{k \in [N(\delta)]} \left|\left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| + \sup_{\Delta \in \mathcal{D}(D,\delta)} \left\|\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\|\right\|,$$

where $\mathcal{D}(D,\delta) \equiv \left\{\Gamma \in \mathcal{C}_{NT}(c_0) \mid \|\Gamma\|_\infty \leq \frac{2}{d},\ \|\Gamma\| \leq \delta,\ \|\Gamma\|_* \leq 2\rho(D)\right\}$. Then by Lemmas B.7-B.8 below with the choice of $\delta = D/8$, we have

$$Z_{NT}(D) \leq \frac{D}{8} + \left(\frac{D}{8} + \frac{24}{d\sqrt{p}}\right) + \frac{D}{2} = \frac{3D}{4} + \frac{c_3}{8d},$$

with probability larger than $1 - c_1 \exp(-c_2 D^2 NT)$ by choosing large enough c_3. ■

The proof of Theorem B.4 relies on the following three lemmas whose proofs are given at the end of this section.

Lemma B.6 *Suppose that there are universal constants (c_1, c_2) such that*

$$P(\mathbf{E}_{NT,D}) \leq c_1 \exp(-c_2 D^2 NT)$$

for each fixed $D > 0$. Then there is a universal constant c_2' such that

$$P(\mathbf{E}_{NT}) \leq c_1 \frac{\exp(-c_2' NT \log d/d)}{1 - \exp(-c_2' NT \log d/d)}.$$

Lemma B.7 *As long as $d \geq 10$, we have*

$$\max_{k \in [N(D/8)]} \left| \left\| \frac{1}{\sqrt{p}} \Gamma^k \circ G \right\| - \left\| \Gamma^k \right\| \right| \leq \frac{D}{8} + \frac{24}{d\sqrt{p}}$$

with probability greater than $1 - 4\exp(-cd^2 \cdot D^2)$ for some constant $c > 0$.

Lemma B.8 $\sup_{\Delta \in \mathcal{D}(D,\delta)} \left\| \frac{1}{\sqrt{p}} \Delta \circ G \right\| \leq \frac{D}{2}$ *with probability at least $1 - 2\exp(-\frac{pd^2 D^2}{512})$.*

To prove Theorem B.5, we need the following lemma.

Lemma B.9 *Let $Z = \{Z_{it}\}$ be a $T \times N$ matrix such that Z_{it} are independent across (i,t), $E(Z_{it}) = 0$, and $\max_{i,t} |Z_{it}| \leq c_c < \infty$ with probability 1. Then there exists constants M_1 and M_2 such that for any $t \geq 0$*

$$P\left(\|Z\|_{sp} \geq M_2(c_a \vee c_b) + t\right) \leq (N \wedge T) \exp\left(\frac{-t^2}{M_1 c_c^2}\right),$$

where $c_a = \max_i \sqrt{\sum_{t=1}^T E\left(Z_{it}^2\right)}$ and $c_b = \max_t \sqrt{\sum_{i=1}^N E\left(Z_{it}^2\right)}$.

Proof. See Proposition 13 of Klopp (2015). ■

Proof of Theorem B.5. On the set $\mathcal{C}_{1NT}$, we define the metric $d(\cdot,\cdot)$ by the Frobenius norm, i.e., $d(\Gamma_1, \Gamma_2) \equiv \|\Gamma_1 - \Gamma_2\|$. For $\Gamma_1 = U_1 V_1'$, $\Gamma_2 = U_2 V_2' \in \mathcal{C}_{1NT}$, we have

$$\begin{aligned} \|\Gamma_1 - \Gamma_2\|^2 &= \sum_{i=1}^N \sum_{t=1}^T (U_{1t} V_{1i} - U_{2t} V_{2i})^2 = \sum_{i=1}^N \sum_{t=1}^T [(U_{1t} - U_{2t}) V_{1i} + U_{2t}(V_{1i} - V_{2i})]^2 \\ &\leq 2 \sum_{i=1}^N V_{1i}^2 \sum_{t=1}^T (U_{1t} - U_{2t})^2 + 2 \sum_{i=1}^N (V_{1i} - V_{2i})^2 \sum_{t=1}^T {U_{2t}}^2 \\ &= 2(\|U_1 - U_2\|^2 + \|V_1 - V_2\|^2), \end{aligned}$$

where the inequality holds by the fact $(a+b)^2 \leq 2(a^2+b^2)$ and the last equality is due to the fact $\|U_2\| = \|V_1\| = 1$. Let $\{U_l\}$ and $\{V_m\}$ be the minimum $\varepsilon/2$-nets of unit sphere in $\mathbb{R}^T$ and $\mathbb{R}^N$, respectively. Then for all $\Gamma = UV'$, there exists a pair (l, m) such that

$$\left\|\Gamma - U_l V_m'\right\|^2 \leq 2(\|U - U_l\|^2 + \|V - V_m\|^2) \leq \varepsilon^2.$$

Hence, $\{U_l\} \times \{V_m\}$ is an ε-net of $\mathcal{C}_{1NT}$. The covering number $\mathcal{N}(\mathcal{C}_{1NT}, d, \varepsilon)$ can be bounded by $\mathcal{N}(B_2^N, \|\cdot\|, \varepsilon/2) \times \mathcal{N}(B_2^T, \|\cdot\|, \varepsilon/2)$, where B_2^N denotes the unit ball in R^N space. By Corollary 4.2.13 of Vershynin (2018), we have $\mathcal{N}(\mathcal{C}_{1NT}, d, \varepsilon) \leq (6/\varepsilon)^{N+T}$. Let $\varepsilon_{NT} = 1/\log(N+T)$ and fix the minimum ε_{NT}-net $\{\Gamma_1, ..., \Gamma_K\}$ where $K \leq (6/\varepsilon_{NT})^{N+T}$. We have

$$\begin{aligned}
\sup_{\Gamma \in \mathcal{C}_{1NT}} \|\Gamma \circ [G - E(G)]\|_{\text{sp}} &\leq \max_{k \in \{1,...,K\}} \sup_{d(\Gamma, \Gamma_k) \leq \varepsilon_{NT}} \|\Gamma \circ [G - E(G)]\|_{\text{sp}} \\
&\leq \max_{k \in \{1,...,K\}} \left\{ \|\Gamma_k \circ [G - E(G)]\|_{\text{sp}} + \sup_{d(\Gamma, \Gamma_k) \leq \varepsilon_{NT}} \|(\Gamma - \Gamma_k) \circ [G - E(G)]\|_{\text{sp}} \right\} \\
&\leq \max_{k \in \{1,...,K\}} \|\Gamma_k \circ [G - E(G)]\|_{\text{sp}} + \max_{k \in \{1,...,K\}} \sup_{d(\Gamma, \Gamma_k) \leq \varepsilon_{NT}} \|\Gamma - \Gamma_k\| \\
&\leq \max_{k \in \{1,...,K\}} \|\Gamma_k \circ [G - E(G)]\|_{\text{sp}} + \varepsilon_{NT},
\end{aligned}$$

where the second inequality holds by the triangle inequality, the third inequality is due to the fact that $\|A\|_{\text{sp}} \leq \|A\|$ and every element of $G - E(G)$ is bounded by 1. For each k, we have $\Gamma_k = U_k V_k'$ for some unit vectors U_k and V_k. Let $Z^{(k)} \equiv \Gamma_k \circ [G - E(G)]$ and denote its (t, i)th entry as $Z_{it}^{(k)}$. By the definition of $\mathcal{C}_{1NT}$ and the fact that $G - E(G)$ has bounded i.i.d. entries, we can show

$$\begin{aligned}
\max\nolimits_{i,t} \left|Z_{it}^{(k)}\right| &\leq \|U_k\|_\infty \|V_k\|_\infty \leq c_{3NT}, \\
\max\nolimits_i \left(\sum_{t=1}^T E[(Z_{it}^{(k)})^2]\right)^{1/2} &\leq \|V_k\|_\infty \leq c_{2NT}, \text{ and } \max\nolimits_t \left(\sum_{i=1}^N E[(Z_{it}^{(k)})^2]\right)^{1/2} \leq \|U_k\|_\infty \leq c_{1NT}.
\end{aligned}$$

By Lemma B.9, there are some universal constants M_1 and M_2 such that

$$P\left(\left\|Z^{(k)}\right\|_{\text{sp}} \geq M_2(c_{1NT} \vee c_{2NT}) + t\right) \leq (N \wedge T) \exp\left(-\frac{t^2}{M_1 c_{3NT}^2}\right).$$

Letting $t = K M_1^{1/2} c_{3NT} \sqrt{(N+T) \log\log(N+T)}$ and noting that $K = (6\log(N+T))^{N+T}$, we have

$$\begin{aligned}
& P\left(\max_{k \in \{1,...,K\}} \left\|Z^{(k)}\right\|_{\text{sp}} \geq M_2(c_{1NT} \vee c_{2NT}) + t\right) \\
\leq\ & (6\log(N+T))^{N+T} (N \wedge T) \exp\left(-K^2 (N+T) \log\log(N+T)\right) \\
=\ & \exp\left(-(K^2 - 1)(N+T) \log\log(N+T) + \log(N \wedge T) + (N+T)\log 6\right) \\
\leq\ & \exp\left(-(N+T) \log\log(N+T)\right),
\end{aligned}$$

as long as $(K^2 - 3)\log\log(N+T) \geq \log 6$ and $\log(N+T) > (N \wedge T)^{1/(N+T)}$. Hence we have shown that

$$\max_{k \in \{1,...,K\}} \left\|Z^{(k)}\right\|_{\text{sp}} = O_P(c_{1NT} + c_{2NT} + c_{3NT} \sqrt{(N+T) \log\log(N+T)}).$$

To sum up, we have

$$\sup_{\Gamma\in\mathcal{C}_{1NT}} \|\Gamma\circ[G-E(G)]\| = O_P(c_{1NT}+c_{2NT}+c_{3NT}\sqrt{(N+T)\log\log(N+T)}+1/\log(N+T)).$$

■

Proof of Lemma B.6. For all $\Gamma\in\mathcal{C}_{NT}(c_0)$ with $\|\Gamma\|_\infty=\frac{1}{d}$, we have

$$\|\Gamma\|^2 \geq c_0\|\Gamma\|_*\sqrt{\frac{\log d}{d}} \geq c_0\|\Gamma\|\sqrt{\frac{\log d}{d}},$$

which implies that $\|\Gamma\|\geq\mu\equiv c_0\sqrt{\frac{\log d}{d}}$. Accordingly, recalling the definition (B.5), it suffices to restrict our attention to the sets $\mathcal{B}(D)$ with $D\geq\mu$. For $l=1,2,\ldots$ and $\alpha=7/6$, define the sets

$$\mathbb{S}_l\equiv\{\Gamma\in\mathcal{C}_{NT}(c_0)\mid \|\Gamma\|_\infty=\frac{1}{d},\ \|\Gamma\|\in[\alpha^{l-1}\mu,\alpha^l\mu],\text{ and }\|\Gamma\|_*\leq\rho(\alpha^l\mu)\}.$$

Now, if the event $\mathbf{E}_{NT}$ holds for some matrix Γ, then $\Gamma\in\mathbb{S}_l\subset\overline{\mathcal{B}}(\alpha^l\mu)$ for some l and

$$\left|\left\|\frac{1}{\sqrt{p}}\Gamma\circ G\right\|-\|\Gamma\|\right| > \frac{7}{8}\|\Gamma\|+\frac{c_3\|\Gamma\|_\infty}{8} \geq \frac{7}{8}\alpha^{l-1}\mu+\frac{c_3\|\Gamma\|_\infty}{8} = \frac{3}{4}\alpha^l\mu+\frac{c_3}{8d},$$

where the equality holds by the fact that $\alpha=7/6$ and $\|\Gamma\|_\infty=\frac{1}{d}$. Thus, $\mathbf{E}_{NT,\alpha^l\mu}$ occurs for some l. It follows that $\mathbf{E}_{NT}\subset\cup_{l=1}^\infty\mathbf{E}_{NT,\alpha^l\mu}$. By the union bound and the fact that $\alpha^{2l}\geq 2c^*l$ for some $c^*>0$ and all $l\geq1$, we have

$$\begin{aligned} P(\mathbf{E}_{NT}) &\leq \sum_{l=1}^\infty P(\mathbf{E}_{NT,\alpha^l\mu}) \leq c_1\sum_{l=1}^\infty\exp(-c_2\alpha^{2l}\mu^2NT) \leq c_1\sum_{l=1}^\infty\exp(-2c^*c_2\mu^2NTl) \\ &= c_1\sum_{l=1}^\infty\left[\exp(-2c^*c_2\mu^2NT)\right]^l = c_1\frac{\exp(-c_2'NT\mu^2)}{1-\exp(-c_2'NT\mu^2)}, \end{aligned}$$

where the second inequality follows from the hypothesis on $P(\mathbf{E}_{NT,D})$ and $c_2'=2c^*c_2$.

Since $NT\mu^2=\frac{NT}{d}\log d$, the claim follows. ■

Proof of Lemma B.7. We first consider a fixed Γ and establish the exponential tail bound. Then we bound the covering number $N(D/8)$ and use the union bound to establish the result.

By the definition of Frobenius norm, we observe that for any $T\times N$ matrix A with typical element A_{it}, we have

$$\|A\| = \left[\sum_{i,t}(A_{it})^2\right]^{1/2} = \left[\sum_{i,t}(A_{it}z_{it})^2\right]^{1/2} = \sup_{\|U\|=1}\sum_{i,t}u_{it}A_{it}z_{it}$$

where z_{it}'s are i.i.d. Rademacher variables. Then

$$\left\|\frac{1}{\sqrt{p}}\Gamma\circ G\right\| = \left[\frac{1}{p}\sum_{i,t}(\Gamma_{it}g_{it})^2\right]^{1/2} = \sup_{\|U\|=1}\left(\sum_{i,t}u_{it}Y_{it}\right) \equiv Z_{NT},$$

where $Y_{it} \equiv \frac{1}{\sqrt{p}} z_{it} \Gamma_{it} g_{it}$, and z_{it}'s are i.i.d. Rademacher variables that are independent of $\{g_{it}\}$. Note that each Y_{it} is zero-mean, and bounded by $\frac{1}{\sqrt{p}d}$. By Corollary 4.8 in Ledoux (2001), we conclude that

$$P\left(|Z_{NT} - E(Z_{NT})| \geq \delta + \frac{8\sqrt{\pi}}{d\sqrt{p}}\right) \leq 4\exp(-\frac{p\delta^2 d^2}{8}), \text{ and } E(Z_{NT}^2) - [E(Z_{NT})]^2 \leq \frac{64}{pd^2}.$$

It follows that $\left|E(Z_{NT}) - \sqrt{E(Z_{NT}^2)}\right| \leq \frac{8}{\sqrt{p}d}$. With the above results and the fact that $E(Z_{NT}^2) = \|\Gamma\|^2$, we can conclude that

$$P\left(\left|\left\|\frac{1}{p}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| \geq \frac{D}{8} + \frac{24}{\sqrt{p}d}\right) \leq 4\exp(-\frac{pD^2 d^2}{512}).$$

The upper bound of covering number $N(\delta)$ can be bounded similarly as in the proof of Lemma 4 in Negahban and Wainwright (2012). Then we have that

$$\log N(\delta) \leq 36(\rho(D)/\delta)^2 d, \text{ where } \rho(D) \equiv \frac{D^2\sqrt{d}}{c_0\sqrt{\log d}}.$$

Combining the tail bound with the union bound, we obtain

$$P\left(\max_{k\in[N(D/8)]}\left|\left\|\frac{1}{p}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| > \frac{D}{8} + \frac{24}{\sqrt{p}d}\right) \leq 4\exp(-\frac{pD^2d^2}{512} + 36(\rho(D)/\delta)^2 d).$$

Choosing the constant c_0 sufficiently large, we have the desired result. ■

Proof of Lemma B.8. Our goal is to bound the function

$$f(G) \equiv \sup_{\Delta\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\|,$$

where we recall that $\mathcal{D}(D,\delta) \equiv \left\{\Gamma \in \mathcal{C}_{NT}(c_0) | \, \|\Gamma\|_\infty \leq \frac{2}{d}, \|\Gamma\| \leq \delta, \|\Gamma\|_* \leq 2\rho(D)\right\}$.

(i) Our approach is to show concentration of G around its expectation $E\left[f(G)\right]$, and then upper bound the expectation. For any independent copy $\tilde{G}$ of G, we have

$$\begin{aligned} f(G) - f(\tilde{G}) &= \sup_{\Delta\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| - \sup_{\tilde{\Delta}\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\tilde{\Delta} \circ \tilde{G}\right\| \\ &\leq \sup_{\Delta\in\mathcal{D}(D,\delta)}\left[\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| - \left\|\frac{1}{\sqrt{p}}\Delta \circ \tilde{G}\right\|\right] \\ &\leq \sup_{\Delta\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\Delta \circ \left(G - \tilde{G}\right)\right\| \\ &\leq \frac{2}{\sqrt{p}d}\left\|G - \tilde{G}\right\|, \end{aligned}$$

where the last inequality is by the fact $G - \tilde{G}$ has entries bounded by 1 and $\|\Delta\|_\infty \leq \frac{2}{d}$. Therefore, by the bounded differences variant of the Azuma-Hoeffding inequality (Ledoux (2001, p.17)), we have

$$P\left\{|f(G) - E[f(G)]| \geq t\right\} \leq 2\exp(-\frac{pd^2t^2}{8}).$$

Setting $t = \frac{D}{8}$, we have $P\left\{|f(G) - E[f(G)]| \geq \frac{D}{8}\right\} \leq 2\exp(-\frac{pd^2D^2}{512})$.

(ii) Next we bound the expectation. First applying Jensen's inequality, we have

$$\begin{aligned}
(E[f(G)])^2 &\leq E[f^2(G)] = E\left(\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}\Delta_{it}^2\frac{g_{it}}{p}\right) \\
&= E\left\{\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}\left[\Delta_{it}^2\frac{g_{it}}{p} - E\left(\Delta_{it}^2\frac{g_{it}}{p}\right)\right] + \|\Delta\|^2\right\} \\
&\leq E\left\{\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}\left[\Delta_{it}^2\frac{g_{it}}{p} - E\left(\Delta_{it}^2\frac{g_{it}}{p}\right)\right]\right\} + \delta^2,
\end{aligned}$$

where we have used the fact that $\sum_{i,t}E\left(\Delta_{it}^2\frac{g_{it}}{p}\right) = \|\Delta\|^2 \leq \delta^2$. By a standard Rademacher symmetrization argument, we can show

$$E[f^2(G)] \leq 2E\left[\sup_{\Delta\in\mathcal{D}(D,\delta)}\frac{1}{NT}\sum_{i,t}\left(NT\Delta_{it}^2\frac{g_{it}}{p}\xi_{it}\right)\right] + \delta^2,$$

where ξ_{it}s are i.i.d. Rademacher variables. Since $\left|NT\Delta_{it}^2\frac{g_{it}}{p}\xi_{it}\right| \leq \frac{4NT}{d^2}$ for all (i,t), the Ledoux-Talagrand contraction inequality (e.g., Ledoux and Talagrand (1991, p.112)) implies that

$$E[f^2(G)] \leq \frac{32\sqrt{NT}}{d^2\sqrt{p}}E\left[\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}(\Delta_{it}g_{it}\xi_{it})\right] + \delta^2.$$

By the inequality that $|\mathrm{tr}(AB)| \leq \|A\|_* \|B\|_{\mathrm{sp}}$, we have $\left|\sum_{i,t}(\Delta_{it}g_{it}\xi_{it})\right| \leq \|\Delta\|_1 \|G\circ\xi\|_{\mathrm{sp}}$. It follows that

$$E[f^2(G)] \leq \frac{32\sqrt{NT}}{d^2\sqrt{p}}\rho(D)E\|G\circ\xi\|_{\mathrm{sp}} + \delta^2,$$

where we used the fact that $\|\Delta\|_* \leq \rho(D)$. Noting that $G\circ\xi$ is a random matrix with bounded i.i.d. zero-mean entries, we have $E\|G\circ\xi\|_{\mathrm{sp}} \leq \sqrt{d\log d}$; see, e.g. Theorem 4.4.5 of Vershynin (2018). Hence, we have

$$E[f(G)] \leq \sqrt{E[f^2(G)]} \leq \left(\frac{32\sqrt{NT}}{c_0 d\sqrt{p}}D^2 + \delta^2\right)^{1/2} \leq \frac{7}{16}D,$$

by choosing a large enough c_0 and noting that $d = (N+T)/2 \geq \sqrt{NT}$.

Combining the results of part (i)-(ii), we have the result desired. ■

REFERENCES

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203-1227.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Bai, J., Ng, S., 2017. Principal components and regularized estimation of factor models. Working Paper, Dept. of Economics, Columbia University.

Klopp, O., 2015. Matrix completion by singular value thresholding: sharp bounds. Electronic Journal of Statistics 9, 2348-2369.

Ledoux, M., 2001. The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs Number 89. American Mathematical Society.

Ledoux, M. and Talagrand, M., 2013. Probability in Banach Spaces: Isoperimetry and Processes. Springer Science & Business Media.

Negahban, S. and Wainwright, M.J., 2012. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. Journal of Machine Learning Research 13, 1665-1697.

Vershynin, R., 2018. High-dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, New York.

Additional Online Supplement for
"Inference for Approximate Factor Models: Random Missing and Cross Validation"

Liangjun Su, Ke Miao, and Sainan Jin
School of Economics, Singapore Management University

This additional online supplement consists of two parts. Section C contains the proofs of the technical lemmas in Appendix A. Section D provides some additional simulation results.

C Proof of the Technical Lemmas in Appendix A

Proof of Lemma A.1. From the principal component analysis (PCA), we have the identity $\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F}=\tilde{F}\tilde{D}$. Pre-multiplying both sides by $T^{-1}\tilde{F}'$ and using the normalization $T^{-1}\tilde{F}'\tilde{F}=I_R$ yield $T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F}=\tilde{D}$. The lemma follows provided $D=\text{plim}\tilde{D}$, which we show below.

Noting that $\tilde{X}=(F^0\Lambda^{0\prime}+\varepsilon)\circ G$, we have

$$\begin{aligned}
\tilde{D} &= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F}\\
&= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime}+\varepsilon)\circ G\right]\left[(F^0\Lambda^{0\prime}+\varepsilon)\circ G\right]'\tilde{F}\\
&= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ G\right]\left[(F^0\Lambda^{0\prime})\circ G\right]'\tilde{F}+T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}(\varepsilon\circ G)(\varepsilon\circ G)'\tilde{F}\\
&\quad +T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ G\right](\varepsilon\circ G)'\tilde{F}+T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}(\varepsilon\circ G)'\left[(F^0\Lambda^{0\prime})\circ G\right]'\tilde{F}\\
&\equiv D_{NT,1}+D_{NT,2}+D_{NT,3}+D_{NT,4},\ \text{say.}
\end{aligned}$$

We first study $D_{NT,1}$. Noting that $E(G)=q\mathbf{1}_{T\times N}$ with $\mathbf{1}_{T\times N}$ being a $T\times N$ matrix of ones, we make the following decomposition

$$\begin{aligned}
D_{NT,1} &= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ G\right]\left[(F^0\Lambda^{0\prime})\circ G\right]'\tilde{F}\\
&= \frac{q^2}{\tilde{q}^2}\frac{\tilde{F}'F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\tilde{F}}{T}+T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\tilde{F}\\
&\quad +\frac{q}{\tilde{q}^2}T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right](F^0\Lambda^{0\prime})\tilde{F}+\frac{q}{\tilde{q}^2}T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}(F^0\Lambda^{0\prime})\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\tilde{F}\\
&\equiv D_{NT,11}+D_{NT,12}+D_{NT,13}+D_{NT,14}
\end{aligned}$$

where $\tilde{G}=G-E(G)$. By the i.i.d. property of g_{it}, we can readily show that $\tilde{q}/q-1=O_P((NT)^{-1/2})$. By Lemma A.3(ii) in Bai (2003), $\frac{\tilde{F}'F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\tilde{F}}{T}\xrightarrow{p}D$. This result can be strengthened to $||\frac{\tilde{F}'F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\tilde{F}}{T}-D||=O_P(\delta_{NT}^{-1})$ under our assumptions. Then $\|D_{NT,11}-D\|=O_P(\delta_{NT}^{-1})$.

For $D_{NT,12}$, we have

$$\begin{aligned}\|D_{NT,12}\|_{\text{sp}} &= \left(NT\tilde{q}^2\right)^{-1}\lambda_{\max}\left(T^{-1}\tilde{F}'\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\tilde{F}\right)\\ &\le \text{tr}\left(T^{-1}\tilde{F}'\tilde{F}\right)\left(NT\tilde{q}^2\right)^{-1}\lambda_{\max}\left(\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\right)\\ &= R\left(NT\tilde{q}^2\right)^{-1}\left\|(F^0\Lambda^{0\prime})\circ\tilde{G}\right\|_{\text{sp}}^2\end{aligned}$$

where the last equality follows from the fact that tr$(T^{-1}\tilde{F}'\tilde{F})$ =tr$(I_R) = R$. Let $c_{\lambda,F} = \max_{i,t}\left|\lambda_i^{0\prime}F_t^0\right|$ and $Z = [(F^0\Lambda^{0\prime})\circ\tilde{G}]/c_{\lambda,F}$. Let Z_{it} denote a typical element of Z : $Z_{it} = \lambda_i^{0\prime}F_t^0\,(g_{it}-q)\,/c_{\lambda,F}$. By construction, $\max_{i,t}|Z_{it}| \le 1$. We want to apply Lemma B.9 by conditioning on $\mathcal{F} = \sigma\left\{F^0,\Lambda^0\right\}$, the sigma-field generated by F^0 and Λ^0. By straightforward moment calculations

$$\begin{aligned}c_1 &\equiv \max_i\sqrt{\sum_{t=1}^T E\left(Z_{it}^2|\mathcal{F}\right)} = \max_i\sqrt{\sum_{t=1}^T\frac{\left(\lambda_i^{0\prime}F_t^0\right)^2}{c_{\lambda,F}^2}E\left(g_{it}-q\right)^2}\\ &= \frac{\sqrt{q\,(1-q)}}{c_{\lambda,F}}\max_i\sqrt{\lambda_i^{0\prime}F^{0\prime}F^0\lambda_i^0} \le \frac{c_{\lambda,N}\left\|F^{0\prime}F^0\right\|^{1/2}}{c_{\lambda,F}},\end{aligned}$$

and

$$\begin{aligned}c_2 &\equiv \max_t\sqrt{\sum_{i=1}^N E\left(Z_{it}^2|\mathcal{F}\right)} = \max_t\sqrt{\sum_{i=1}^N\frac{\left(\lambda_i^{0\prime}F_t^0\right)^2}{c_{\lambda,F}^2}E\left(g_{it}-q\right)^2}\\ &= \frac{\sqrt{q\,(1-q)}}{c_{\lambda,F}}\max_t\sqrt{F_t^{0\prime}\Lambda^{0\prime}\Lambda^0F_t^0} \le \frac{c_{F,T}\left\|\Lambda^{0\prime}\Lambda^0\right\|^{1/2}}{c_{\lambda,F}},\end{aligned}$$

where $c_{\lambda,N} = \max_i\left\|\lambda_i^0\right\|$ and $c_{F,T} = \max_t\left\|F_t^0\right\|$. It follows that

$$\left\|(F^0\Lambda^{0\prime})\circ\tilde{G}\right\|_{\text{sp}} = O_P\left(\max\left\{c_{\lambda,N}\left\|F^{0\prime}F^0\right\|^{1/2}, c_{F,T}\left\|\Lambda^{0\prime}\Lambda^0\right\|^{1/2}, c_{\lambda,F}\log(N\vee T)\right\}\right).$$

This result, in conjunction with the fact $\left\|F^{0\prime}F^0\right\| = O_P(T)$, $\left\|\Lambda^{0\prime}\Lambda^0\right\| = O_P(N)$, $c_{\lambda,N} = O_P\left(N^{\gamma_1/4)}\right)$, $c_{F,T} = O_P\left(T^{\gamma_2/4)}\right)$, and $c_{\lambda,F} = O_P\left((NT)^{1/4}\right)$ under our moment conditions on λ_i^0 and F_t^0 in Assumption A.1, implies that

$$\begin{aligned}\|D_{NT,12}\| &\le \sqrt{R}\,\|D_{NT,12}\|_{\text{sp}} = O_P\left(\frac{1}{NT}\max\left\{c_{\lambda,N}^2T, c_{F,T}^2N, c_{\lambda,F}^2\left[\log(N\vee T)\right]^2\right\}\right)\\ &\le O_P\left\{\frac{1}{NT}\max\left\{N^{\gamma_1/2}T, T^{\gamma_2/2}N, (NT)^{1/2}\left[\log(N\vee T)\right]^2\right\}\right\} = O_P(\delta_{NT}^{-(2-\gamma)})\end{aligned}$$

where $\gamma = \gamma_1\vee\gamma_2$. Then $\|D_{NT,13}\| = \|D_{NT,14}\| \le \{\|D_{NT,11}\|\,\|D_{NT,12}\|\}^{1/2} = O_P(\delta_{NT}^{-(1-\gamma/2)})$ by the matrix version of Cauchy-Schwarz (CS) inequality. Therefore we have $\|D_{NT,1}-D\| = O_P(\delta_{NT}^{-(1-\gamma/2)})$.

Noting that $D_{NT,2}$ is positive semidefinite (p.s.d.), we have

$$\begin{aligned}\|D_{NT,2}\|_{\text{sp}} &\le \left(NT\tilde{q}^2\right)^{-1}\text{tr}\left(T^{-1}\tilde{F}'\left(\varepsilon\circ G\right)\left(\varepsilon\circ G\right)'\tilde{F}\right) \le \text{tr}\left(T^{-1}\tilde{F}'\tilde{F}\right)\left(NT\tilde{q}^2\right)^{-1}\lambda_{\max}\left(\left(\varepsilon\circ G\right)\left(\varepsilon\circ G\right)'\right)\\ &= R\left(NT\tilde{q}^2\right)^{-1}\|\varepsilon\circ G\|_{\text{sp}}^2,\end{aligned}$$

where the first inequality follows from the fact that $\|A\|_{\rm sp} = \lambda_{\max}(A) \leq {\rm tr}(A)$ for any p.s.d. symmetric matrix A, the second inequality follows because ${\rm tr}(A'BA) \leq {\rm tr}(A'A)\,\lambda_{\max}(B)$ for any symmetric p.s.d. matrix B and conformable matrix A, the equality follows because ${\rm tr}(T^{-1}\tilde{F}'\tilde{F}) = {\rm tr}(I_R) = R$. Note that

$$\|\varepsilon \circ G\|_{\rm sp} \leq \left\|\varepsilon \circ \tilde{G}\right\|_{\rm sp} + \|\varepsilon \circ E(G)\|_{\rm sp} = \left\|\varepsilon \circ \tilde{G}\right\|_{\rm sp} + q\,\|\varepsilon\|_{\rm sp}.$$

By Assumption A.2(i), $\|\varepsilon\|_{\rm sp} = O_P(\sqrt{N}+\sqrt{T})$. As in the analysis of $\left\|(F^0\Lambda^{0\prime}) \circ \tilde{G}\right\|_{\rm sp}$, we can readily apply Lemma B.1 by conditioning on ε to obtain with high probability

$$\left\|\varepsilon \circ \tilde{G}\right\|_{\rm sp} = O_P\left(\max\left\{\sqrt{N}, \sqrt{T}, \max_{i,t}|\varepsilon_{it}|\log(N\vee T)\right\}\right) \leq O_P\left(\sqrt{N}+\sqrt{T}+(NT)^{1/4}\log(N\vee T)\right).$$

It follows that $\|D_{NT,2}\| \leq \sqrt{R}\,\|D_{NT,2}\|_{\rm sp} \leq (NT)^{-1}\,O_P\left(N+T+(NT)^{1/2}\left[\log(N\vee T)\right]^2\right) = o_P(\delta_{NT}^{-(2-\gamma)})$ and $\|D_{NT,2}\| \leq \sqrt{R}\,\|D_{NT,2}\|_{\rm sp} = o_P(\delta_{NT}^{-(2-\gamma)})$ and $\|D_{NT,3}\| = \|D_{NT,4}\| \leq \{\|D_{NT,1}\|\,\|D_{NT,2}\|\}^{1/2} = o_P(\delta_{NT}^{-(1-\gamma/2)})$ by the CS inequality.

In sum, we have $\left\|\tilde{D} - D\right\| = O_P(\delta_{NT}^{-(1-\gamma/2)})$. ■

Proof of Lemma A.2. (i) From the method of PCA, we have

$$\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F} = \tilde{F}\tilde{D}. \tag{C.1}$$

Using $\tilde{X} = \left(F^0\Lambda^{0\prime} + \varepsilon\right)\circ G$ and $G = E(G) + \tilde{G} = q\mathbf{1}_{T\times N} + \tilde{G}$, we make the following decomposition

$$\begin{aligned}
&\tilde{X}\tilde{X}' \\
=\;& \left[\left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G\right]\left[\left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G\right]' \\
=\;& \left[(F^0\Lambda^{0\prime})\circ G\right]\left[(F^0\Lambda^{0\prime})\circ G\right]' + (\varepsilon\circ G)(\varepsilon\circ G)' + \left[(F^0\Lambda^{0\prime})\circ G\right](\varepsilon\circ G)' + (\varepsilon\circ G)'\left[(F^0\Lambda^{0\prime})\circ G\right]' \\
=\;& q^2F^0\Lambda^{0\prime}\Lambda^0F^{0\prime} + d_{NT},
\end{aligned} \tag{C.2}$$

where

$$\begin{aligned}
d_{NT} =\;& \left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]' + q(F^0\Lambda^{0\prime})\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]' + q\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\Lambda^0F^{0\prime} \\
&+ (\varepsilon\circ G)(\varepsilon\circ G)' + \left[(F^0\Lambda^{0\prime})\circ G\right](\varepsilon\circ G)' + (\varepsilon\circ G)'\left[(F^0\Lambda^{0\prime})\circ G\right]'.
\end{aligned}$$

Premultiplying both sides of (C.1) by $\left(\frac{1}{N}\Lambda^{0\prime}\Lambda^0\right)^{1/2}\frac{1}{T}F^{0\prime}$ and plugging (C.2) yield

$$\frac{q^2}{\tilde{q}^2}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}}{T}\right) + \bar{d}_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\tilde{F}}{T}\right)\tilde{D}, \tag{C.3}$$

where $\bar{d}_{NT} = \frac{1}{\tilde{q}^2}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\frac{1}{T}F^{0\prime}d_{NT}\tilde{F}$. Following the analysis of D_{NT}'s in the proof of Lemma A.1, we can readily show that $\left\|\bar{d}_{NT}\right\| = O_P(\delta_{NT}^{-(1-\gamma/2)})$. Letting

$$B_{NT} = \frac{q^2}{\tilde{q}^2}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2} \text{ and } R_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\tilde{F}}{T}\right),$$

we can write (C.3) as follows: $[B_{NT}+\bar{d}_{NT}R_{NT}^{-1}]R_{NT}=R_{NT}\tilde{D}$. Hence, each column of R_{NT} is non-standardized eigenvector of the matrix $B_{NT}+\bar{d}_{NT}R_{NT}^{-1}$. Let $\breve{D}_{NT}$ be a diagonal matrix consisting of the diagonal elements of $R_{NT}^{'}R_{NT}$. Denote the standardized eigenvector $\Upsilon_{NT}=R_{NT}\breve{D}_{NT}^{-1/2}$. Hence, we have $[B_{NT}+\bar{d}_{NT}R_{NT}^{-1}]\Upsilon_{NT}=\Upsilon_{NT}\tilde{D}^{-1}$. That is, $\tilde{D}$ contains the eigenvalues of $B_{NT}+\bar{d}_{NT}R_{NT}^{-1}$ with the corresponding normalized eigenvectors contained in Υ_{NT}. It is trivial to show that with high probability

$$\left\|B_{NT}+\bar{d}_{NT}R_{NT}^{-1}-B\right\|=O_P(\delta_{NT}^{-(1-\gamma/2)}), \tag{C.4}$$

where B denotes the probability of B_{NT}, i.e., $B=\Sigma_{\Lambda^0}^{1/2}\Sigma_{F^0}\Sigma_{\Lambda^0}^{1/2}$.

Let Υ denote the probability limit of Υ_{NT}. Note that $\Upsilon'=\Upsilon^{-1}$ by normalization. By (C.4) and the eigenvector perturbation theory that requires distinctness of eigenvalues (see, e.g., Steward and Sun (1990, Ch. V), and Allez and Bouchaud (2013)), $\|\Upsilon_{NT}-\Upsilon\|=O_P(\delta_{NT}^{-(1-\gamma/2)})$ by (C.4) and Assumption A.1(iv). This, in conjunction with the definition of R_{NT}, implies that

$$\frac{F^{0\prime}\tilde{F}}{T}=\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1/2}R_{NT}=\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1/2}\Upsilon_{NT}\breve{D}_{NT}^{1/2}$$

satisfies $\left\|\frac{F^{0\prime}\tilde{F}}{T}-\Sigma_{\Lambda^0}^{-1/2}\Upsilon D^{1/2}\right\|=O_P(\delta_{NT}^{-(1-\gamma/2)})$. The result follows by noticing that $Q'=\Sigma_{\Lambda^0}^{-1/2}\Upsilon D^{1/2}$.

(ii) By Lemma A.1, (i) and Assumption A.1(ii) , we have

$$\begin{aligned}\tilde{H} &= \left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\left(T^{-1}F^{0\prime}\tilde{F}\right)\tilde{D}^{-1}=\Sigma_{\Lambda^0}(\Sigma_{\Lambda^0}^{-1/2}\Upsilon D^{1/2})D^{-1}+O_P(\delta_{NT}^{-(1-\gamma/2)})\\ &= \Sigma_{\Lambda^0}^{1/2}\Upsilon D^{-1/2}+O_P(\delta_{NT}^{-(1-\gamma/2)})=Q^{-1}+O_P(\delta_{NT}^{-(1-\gamma/2)}).\end{aligned}$$

(iii) The proof follows closely that of Lemma B.1 in Bai (2003) and Theorem 2.1 and thus omitted. The major difference is that we now use the decomposition in (A.1) and the fact that g_{it} are i.i.d. Bernoulli(q) and independent of F^0, Λ^0 and ε.

(iv) The proof is analogous to that of Theorem 2.1 and thus omitted.

(v) The claim follows from (iv) provided that we can show that $\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t-\tilde{H}'F_t^0)F_t^{0\prime}g_{it}=O_P\left(\delta_{NT}^{-2}\right)$. The proof of the latter result follows closely that of Theorem 2.1 (or Lemma B.2 in Bai (2003)) and thus omitted.

(vi) By (v), the claim follows provided that $\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t-\tilde{H}'F_t^0)F_t^{0\prime}=O_P\left(\delta_{NT}^{-2}\right)$. We can prove the latter result by using analogous arguments as used in the proof of Theorem 2.1 and Lemma B.2 in Bai (2003).

(vii) Using $\tilde{F}_t=(\tilde{F}_t-\tilde{H}'F_t^0)+\tilde{H}'F_t^0$, we make the following decomposition

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'(g_{it}-q) &= \tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H}(g_{it}-q)+\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t-\tilde{H}'F_t^0)(\tilde{F}_t-\tilde{H}'F_t^0)'(g_{it}-q)\\ &\quad+\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t-\tilde{H}'F_t^0)F_t^{0\prime}\tilde{H}'(g_{it}-q)+\tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0(\tilde{F}_t-\tilde{H}'F_t^0)'(g_{it}-q)\\ &\equiv d_{1t}+d_{2t}+d_{3t}+d_{4t}.\end{aligned}$$

By Theorem 2.1 and Lemma A.2(iv), $d_{2t} = O_P\left(\delta_{NT}^{-2}\right)$. By Lemma A.2(vi)-(vii), $d_{3t} = O_P\left(\delta_{NT}^{-2}\right)$ and $d_{4t} = O_P\left(\delta_{NT}^{-2}\right)$. Then $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'(g_{it}-q) = \tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H}(g_{it}-q) + O_P\left(\delta_{NT}^{-2}\right)$.

(viii) As in (vii), we can also show that $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t' = \tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H} + O_P\left(\delta_{NT}^{-2}\right)$. This, in conjunction with the fact that $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t' = I_R$, implies that

$$\tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H} = I_R + O_P\left(\delta_{NT}^{-2}\right).$$

Premultiplying and postmultiplying both sides by $(\tilde{H}')^{-1}$ and $\tilde{H}^{-1}$ in order yields $\frac{1}{T}F^{0\prime}F^0 = (\tilde{H}\tilde{H}')^{-1} + O_P\left(\delta_{NT}^{-2}\right)$. It follows that $\tilde{H}\tilde{H}' = (\frac{1}{T}F^{0\prime}F^0)^{-1} + O_P\left(\delta_{NT}^{-2}\right)$. ■

A Cautionary Note. We can prove Lemmas A.3-A.5 for $\ell = 1$ based on the results in Theorems 2.1-2.2. When these lemmas hold for $\ell = 1$, Theorems 2.3-2.4 also hold for $\ell = 1$. With the results in Lemmas A.3-A.5 and Theorems 2.3-2.4 for $\ell = 1$, we can prove them to hold for $\ell = 2$. This procedure is repeated until convergence. Since the verification of Lemma A.3 for $\ell = 1$ is different from the general case with $\ell \geq 2$, we first prove it for $\ell = 1$ in detail and then prove it for $\ell \geq 2$ after we prove Lemmas A.4-A.5.

Proof of Lemma A.3 ($\ell = 1$)**.** (i) Noting that $\hat{\phi}_{F,t}^{(0)} = \hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0\frac{1}{Nq}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]$, $\max_t\left\|\hat{\phi}_{F,t}^{(0)}\right\| \leq O_P(1)\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]\right\| = O_P\left((N/\ln N)^{-1/2}\right)$ by Lemmas A.1-A.2 and Assumption A.5(i). Similarly, $\max_i\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\| \leq O_P(1)\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]\right\|$ $= O_P\left((T/\ln T)^{-1/2}\right)$ by Lemmas A.1-A.2 and Assumption A.5(ii).

(ii) By the decomposition in (A.1),

$$\hat{r}_{F,t}^{(0)} = \hat{F}_t^{(0)} - \hat{H}^{(0)\prime}F_t^0 - \hat{\phi}_{F,t}^{(0)} = a_{1t} + a_{2t} + a_{4t} + a_{5t} + a_{7t} + (a_{3t} + a_{6t} - \hat{\phi}_{F,t}^{(0)}).$$

Following the proof of Theorem 2.2(i) and using Assumption A.5 and the fact that $\max_t\left\|F_t^0\right\| = O_P\left(T^{\gamma_1/4}\right)$, it is easy to show that

$$\begin{aligned}
\max_t\|a_{1t}\| &= O_P\left(T^{-1/2}\delta_{NT}^{-1} + T^{-1+\gamma_1/4}\right), \ \max_t\|a_{2t}\| = O_P\left(\delta_{NT}^{-2}\ln N\right), \\
\max_t\|a_{lt}\| &= O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\right) \text{ for } l = 4, 5, \\
\max_t\|a_{7t}\| &= O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}\right),
\end{aligned}$$

and $\max_t\left\|a_{3t} + a_{6t} - \hat{\phi}_{F,t}^{(0)}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$. It follows that $\max_t\left\|\hat{r}_{F,t}^{(0)}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})$. For $\hat{r}_{\Lambda,i}^{(0)}$, we have

$$\hat{r}_{\Lambda,i}^{(0)} = \hat{\lambda}_i^{(0)} - (\hat{H}^{(0)})^{-1}\lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(0)} = B_{2i} + B_{3i} + B_{5i} + (B_{1i} + B_{4i} - \hat{\phi}_{\Lambda,i}^{(0)}),$$

where B_{2i}'s are defined in the proof of Theorem 2.2(ii). Following the proof of Theorem 2.2(ii) and using the fact that $\max_i\left\|\lambda_i^0\right\| = O_P\left(N^{\gamma_2/4}\right)$, $\max_i\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^2 = O_P(1)$, and $\tilde{q} - q = O_P((NT)^{-1/2})$

we have by Theorem 2.1 and Lemma A.2

$$\max_i \|B_{2i}\| = O_P\left(\delta_{NT}^{-2}\ln N\right),\ \ \max_i \|B_{3i}\| = O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right),\ \ \max_i \|B_{5i}\| = O_P\left((NT)^{-1/2}N^{\gamma_2/4}\right),$$

and $\max_i \left\|B_{1i}+B_{4i}-\hat{\phi}_{\Lambda,i}^{(0)}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$. It follows that $\max_i \left\|\hat{r}_{\Lambda,i}^{(0)}\right\| = O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right)$.

(iii) By (i) and the fact that $\max_t \left\|F_t^0\right\| = O_P\left(T^{\gamma_1/4}\right)$ and $\max_i \left\|\lambda_i^0\right\| = O_P\left(N^{\gamma_2/4}\right)$, we have

$$\begin{aligned}
& \max_{i,t}\left\|\eta_{1,it}^{(0)}\right\| \\
= & \max_{i,t}\left\|F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)} + F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}\right\| \\
\leq & \left\|\hat{H}^{(0)}\right\|\max_t\left\|F_t^0\right\|\left\{\max_i\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\| + \hat{r}_{\Lambda,i}^{(\ell)}\right\} + \left\|(\hat{H}^{(0)\prime})^{-1}\right\|\max_i\left\|\lambda_i^0\right\|\left\{\max_t\left\|\hat{\phi}_{F,t}^{(0)}\right\| + \max_t\left\|\hat{r}_{F,t}^{(\ell)}\right\|\right\} \\
= & O_P(T^{\gamma_1/4}((T/\ln T)^{-1/2} + N^{\gamma_2/4}\delta_{NT}^{-2}\ln N)) + O_P(T^{\gamma_1/4}((T/\ln T)^{-1/2} + T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}) \\
= & O_P(\delta_{NT}^{-1+\gamma/2}\ln N).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\max_{i,t}\left\|\eta_{2,it}^{(0)}\right\| = & \max_{i,t}\left\|\hat{\phi}_{\Lambda,i}^{(0)\prime}\hat{\phi}_{F,t}^{(0)} + \hat{\phi}_{\Lambda,i}^{(0)\prime}\hat{r}_{F,t}^{(0)} + \hat{\phi}_{F,t}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)} + \hat{r}_{\Lambda,i}^{(0)\prime}\hat{r}_{F,t}^{(0)}\right\| \\
\leq & O_P\left((N/\ln N)^{-1/2}(T/\ln T)^{-1/2}\right) + O_P(T/\ln T)^{-1/2}\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}\right)) \\
& + O_P\left((N/\ln N)^{-1/2}N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right) + O_P\left(\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}\right)\right) \\
= & O_P(\delta_{NT}^{-2}\ln N).
\end{aligned}$$

(iv) Note that

$$\begin{aligned}
& [\hat{H}^{(0)\prime}]^{-1}\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it} \\
= & \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\left[\varepsilon_{is}g_{is} + \lambda_i^{0\prime}F_s^0\left(g_{is}-q\right)\right]\varepsilon_{it}g_{it} \\
= & \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0 E(\varepsilon_{is}\varepsilon_{it})g_{is}g_{it} + \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\left[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})\right]g_{is}g_{it} \\
& + \frac{q}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0F_s^{0\prime}\lambda_i^0\left(g_{is}-q\right)\varepsilon_{it}g_{it} + \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0F_s^{0\prime}\lambda_i^0\left(g_{is}-q\right)\varepsilon_{it}\left(g_{it}-q\right) \\
= & O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N + \delta_{NT}^{-2}\ln N + \delta_{NT}^{-2}\ln N\right) = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right).
\end{aligned}$$

Then $\max_t \left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it}\right\| = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$.

Observe that $\left\|\hat{H}^{(0)}\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\| = \left\|\hat{H}^{(0)}\hat{H}^{(0)\prime}\right\|\left\|\frac{1}{NTq}\sum_{i=1}^N\sum_{s=1}^T \lambda_i^0F_s^{0\prime}\left[\varepsilon_{is}g_{is} + F_s^{0\prime}\lambda_i^0(g_{is}-q)\right]\bar{g}_{it}\right\|$.

Using $\bar{g}_{it} = (1-q) - (g_{it} - q)$, we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\lambda_i^0 F_s^{0\prime}\left[\varepsilon_{is}g_{is} + F_s^{0\prime}\lambda_i^0(g_{is}-q)\right]\bar{g}_{it} \\
= &\frac{1-q}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\lambda_i^0 F_s^{0\prime}\varepsilon_{is}g_{is} + \frac{1-q}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\lambda_i^0 F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q) \\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\lambda_i^0 F_s^{0\prime}\varepsilon_{is}g_{is}(g_{it}-q) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\lambda_i^0 F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q)(g_{it}-q).
\end{aligned}$$

It is easy to show that the first two terms are $O_P(\delta_{NT}^{-2})$ by Chebyshev inequality. The third term is $O_P(\delta_{NT}^{-2}\ln N)$ by Assumption A.3(iii). For the fourth term, we have

$$\begin{aligned}
&\max_t\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\lambda_i^0 F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q)(g_{it}-q)\right\| \\
= &\max_t\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\lambda_i^0 F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q)(g_{it}-q)\right\| + \frac{1}{NT}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\max_t\left\|F_t^0\right\|^2 \\
= &O_P(\delta_{NT}^{-2}\ln N) + O_P(T^{-1+\gamma_1/2}).
\end{aligned}$$

Then $\left\|\hat{H}^{(0)}\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\| = O_P\left(T^{-1+\gamma_1/2} + \delta_{NT}^{-2}\ln N\right)$.

Noting that $\hat{r}_{\Lambda,i}^{(0)} = \hat{\lambda}_i^{(0)} - (\hat{H}^{(0)})^{-1}\lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(0)} = B_{2i} + B_{3i} + B_{5i} + (B_{1i} + B_{4i} - \hat{\phi}_{\Lambda,i}^{(0)})$, we have

$$\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{r}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\| \leq \max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\left[B_{2i} + B_{3i} + B_{5i} + (B_{1i} + B_{4i} - \hat{\phi}_{\Lambda,i}^{(0)})\right]\lambda_i^{0\prime}\bar{g}_{it}\right\|,$$

where B_{2i}'s are defined in the proof of Theorem 2.2(ii). Using $\bar{g}_{it} = (1-q) + (g_{it} - q)$,

$$\begin{aligned}
\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}B_{2i}\lambda_i^{0\prime}\bar{g}_{it}\right\| &= \frac{\left\|\hat{H}^{(0)}\right\|}{\tilde{q}}\max_t\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}g_{is}\lambda_i^{0\prime}\bar{g}_{it}\right\| \\
&\leq O_P(1)\left\{\left\|\frac{1-q}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}g_{is}\lambda_i^{0\prime}\right\| + \max_t\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}g_{is}\lambda_i^{0\prime}(g_{it}-q)\right\|\right\} \\
&= O_P((NT)^{-1/2} + (NT)^{-1/2}\ln N).
\end{aligned}$$

In addition,

$$\begin{aligned}
\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}B_{3i}\lambda_i^{0\prime}\bar{g}_{it}\right\| &= \max_t\frac{1}{NT\tilde{q}}\left\|\sum_{i=1}^{N}\sum_{s=1}^{T}\hat{F}_s^{(0)}\left(\hat{H}^{(0)\prime}F_s^0 - \hat{F}_s^{(0)}\right)'(\hat{H}^{(0)\prime})^{-1}\lambda_i^0 g_{is}\lambda_i^{0\prime}\bar{g}_{it}\right\| \\
&\leq O_P(1)\max_i\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(0)}\left(\hat{H}^{(0)\prime}F_s^0 - \hat{F}_s^{(0)}\right)'g_{is}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right),
\end{aligned}$$

and $\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}B_{5i}\lambda_i^{0\prime}\bar{g}_{it}\right\| \leq \frac{1}{\tilde{q}}\left|q-\tilde{q}\right|\left\|[\hat{H}^{(0)\prime}]^{-1}\right\|\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2 = O_P((NT)^{-1/2})$. Lastly, noting that the difference lies between $B_{1i} + B_{4i}$ and $\hat{\phi}_{\Lambda,i}^{(0)}$ is controlled by $\left|\tilde{q} - q\right|$, we can readily show that

$\max_t \left\| \frac{1}{N}\sum_{i=1}^{N}(B_{1i}+B_{4i}-\hat{\phi}_{\Lambda,i}^{(0)})\lambda_i^{0\prime}\bar{g}_{it}\right\| = O_P((NT)^{-1/2})$. In sum, we have $\max_t \left\|\hat{H}^{(0)\prime}\frac{1}{N}\sum_{i=1}^{N}\hat{r}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\|$ $= O_P\left(\delta_{NT}^{-2}\ln N\right)$.

(v) Noting that $\hat{\phi}_{F,t}^{(0)} = \hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0\frac{1}{Nq}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it}+\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]$, we have

$$\begin{aligned}
\max_i \left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\phi}_{F,t}^{(0)}F_t^{0\prime}\bar{g}_{it}\right\| &\leq O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\lambda_j^0\left[\varepsilon_{jt}g_{jt}+\lambda_j^{0\prime}F_t^0\left(g_{jt}-q\right)\right]F_t^{0\prime}\bar{g}_{it}\right\| \\
&\leq O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\lambda_j^0F_t^{0\prime}\varepsilon_{jt}g_{jt}\bar{g}_{it}\right\| \\
&\quad +O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\lambda_j^0F_t^{0\prime}\lambda_j^{0\prime}F_t^0\left(g_{jt}-q\right)\bar{g}_{it}\right\| \\
&= O_P\left(\delta_{NT}^{-2}\ln N\right)+O_P(N^{-1+\gamma_2/2}).
\end{aligned}$$

Analogously, by the decomposition in (A.1) we have $\frac{1}{T}\sum_{t=1}^{T}\hat{r}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it} = \frac{1}{T}\sum_{t=1}^{T}[a_{1t}+a_{2t}+a_{4t}+a_{5t}+a_{7t}+(a_{3t}+a_{6t}-\hat{\phi}_{F,t}^{(0)})]F_t^{0\prime}\bar{g}_{it}$. Following the proof of Theorem 2.2(i) and using Assumption A.5 and the fact that $\max_i\left\|\lambda_i^0\right\| = O_P\left(N^{\gamma_2/4}\right)$, it is easy to show that $\max_i ||\frac{1}{T}\sum_{t=1}^{T}a_{1t}F_t^{0\prime}\bar{g}_{it}|| = O_P\left(T^{-1/2}\delta_{NT}^{-1}+T^{-1}\right)$, $\max_i ||\frac{1}{T}\sum_{t=1}^{T}a_{2t}F_t^{0\prime}\bar{g}_{it}|| \leq \max_t\|a_{2t}\|\,O_P(1) = O_P\left(\delta_{NT}^{-2}\ln N\right)$, $\max_i ||\frac{1}{T}\sum_{t=1}^{T}a_{lt}F_t^{0\prime}\bar{g}_{it}|| = O_P\left(\delta_{NT}^{-2}\ln N\right)$ for $l = 4,5$, $\max_i ||\frac{1}{T}\sum_{t=1}^{T}a_{7t}F_t^{0\prime}\bar{g}_{it}|| = O_P\left(\delta_{NT}^{-2}\ln T+T^{-1}\right)$, and $\max_i ||\frac{1}{T}\sum_{t=1}^{T}[a_{3t}+a_{6t}-\hat{\phi}_{F,t}^{(0)}]F_t^{0\prime}\bar{g}_{it}|| = O_P\left(\delta_{NT}^{-2}\ln N\right)$. It follows that $\max_i ||\frac{1}{T}\sum_{t=1}^{T}\hat{r}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}||$ $= O_P\left(\delta_{NT}^{-2}\ln N\right)$.

(vi) Note that $\frac{1}{N}\sum_{i=1}^{N}\left\|\eta_{it}^{(0)}\right\|^2 \leq \frac{2}{N}\sum_{i=1}^{N}\left\|\eta_{1,it}^{(0)}\right\|^2+\frac{2}{N}\sum_{i=1}^{N}\left\|\eta_{2,it}^{(0)}\right\|^2$, where the second term is bounded above by $O_P(\delta_{NT}^{-4}(\ln N)^2)$ by (iii). For the first term, we have

$$\begin{aligned}
\max_t\frac{1}{N}\sum_{i=1}^{N}\left\|\eta_{1,it}^{(0)}\right\|^2 &\leq \max_t\frac{1}{N}\sum_{i=1}^{N}\left\|F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}+\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)}+\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)}+F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}\right\|^2 \\
&\leq 4\left\|\hat{H}^{(0)}\right\|\max_t\left\|F_t^0\right\|^2\frac{1}{N}\sum_{i=1}^{N}(\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2+\left\|\hat{r}_{\Lambda,i}^{(0)}\right\|^2) \\
&\quad +4\left\|[\hat{H}^{(0)\prime}]^{-1}\right\|\left\{\max_t\left\|\hat{\phi}_{F,t}^{(0)}\right\|^2+\max_t\left\|\hat{r}_{F,t}^{(0)}\right\|^2\right\}\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2 \\
&= O_P(T^{-1+\gamma_1/2}+N^{-1}\ln N).
\end{aligned}$$

It follows that $\frac{1}{N}\sum_{i=1}^{N}\left\|\eta_{it}^{(0)}\right\|^2 = O_P(T^{-1+\gamma_1/2}+N^{-1}\ln N)$. Similarly, we can show that $\max_t\frac{1}{T}\sum_{t=1}^{T}||\eta_{it}^{(0)}||^2$ $= O_P(N^{-1+\gamma_2/2}+T^{-1}\ln N)$.

(vii) Let $\kappa_t = 1+\left\|F_t^0\right\|^2$. It suffices to show that $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{l,it}^{(0)})^2 = O_P\left(\delta_{NT}^{-2}\right)$ for $l = 1,2$. By (iii), $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{2,it}^{(0)})^2 \leq \max_{i,t}||\eta_{2,it}^{(0)}||^2\frac{1}{T}\sum_{t=1}^{T}\kappa_t = O_P(\delta_{NT}^{-4}(\ln N)^2)$. In addition, $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{1,it}^{(0)})^2 \leq \frac{4}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t\{||F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}||^2+||\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)}||^2+||\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)}||^2+$

$||F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}||^2\} \equiv 4\left\{J_{1,1}+J_{1,2}+J_{1,3}+J_{1,4}\right\}.$ For $J_{1,1}$, we have

$$J_{1,1} \leq \left\|\hat{H}^{(0)}\right\|^2 \frac{1}{T}\sum_{t=1}^T \kappa_t \left\|F_t^0\right\|^2 \frac{1}{N}\sum_{i=1}^N \left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2 = O_P\left(T^{-1}\right),$$

as we can readily show that $\frac{1}{N}\sum_{i=1}^N \left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2 = O_P\left(T^{-1}\right)$. For $J_{1,2}$, noting that $(\hat{H}^{(0)\prime})^{-1}\hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0 = (\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}$ and $\frac{1}{N}\Lambda^{0\prime}\Lambda^0 - \Sigma_{\Lambda^0} = O\left(N^{-1/2}\right)$, we have

$$\begin{aligned} J_{1,2} &= \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \kappa_t \left\|\lambda_i^{0\prime}(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\frac{1}{Nq}\sum_{j=1}^N \lambda_j^0\left[\varepsilon_{jt}g_{ij} + \lambda_j^{0\prime}F_t^0\left(g_{jt}-q\right)\right]\right\|^2 \\ &\leq O_P(1)\frac{1}{T}\sum_{t=1}^T \kappa_t\left\|\frac{1}{N}\sum_{i=1}^N \lambda_i^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]\right\|^2 = O_P\left(\delta_{NT}^{-2}\right). \end{aligned}$$

Similarly, we can show that $J_{1,l} = O_P\left(\delta_{NT}^{-2}\right)$ for $l=3,4$. Then $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \kappa_t(\eta_{1,it}^{(0)})^2 = O_P\left(\delta_{NT}^{-2}\right)$.

(viii) Note that $\frac{1}{NT}\sum_{s=1}^T\sum_{i=1}^N F_s^0\lambda_i^{0\prime}\eta_{is}^{(0)}\bar{g}_{is} = \sum_{l=1}^2 \frac{1}{NT}\sum_{s=1}^T\sum_{i=1}^N F_s^0\lambda_i^{0\prime}\eta_{l,is}^{(0)}\bar{g}_{is} \equiv \sum_{l=1}^2 J_{2,l}$. For $J_{2,2}$, we can use the uniform bound in (iii) and show that $J_{2,l} = O_P\left(\delta_{NT}^{-2}\ln N\right)$. For $J_{2,1}$, we make the following decomposition

$$\begin{aligned} J_{2,1} &= \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N F_t^0\lambda_i^{0\prime}\left(F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)} + F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}\right)\bar{g}_{it} \\ &\equiv \sum_{a=1}^4 J_{2,1a}. \end{aligned}$$

Let λ_{il}^0 and F_{sl}^0 denote the lth element of λ_i^0 and F_s^0, respectively. Let $J_{2,1a}(l,r)$ denote the (l,r)th element of $J_{2,1a}$ for $a=1,2$. Noting that $\bar{g}_{is} = (1-q)+(q-g_{is})$, we have

$$\begin{aligned} \left\|J_{2,11}(l,r)\right\| &= \left\|\frac{1}{NT}\sum_{t=1}^T F_{tr}^0F_t^{0\prime}\hat{H}^{(0)}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\bar{g}_{it}\lambda_{il}^0\right\| \\ &\leq \left\|\frac{1}{NT}\sum_{t=1}^T F_{tr}^0F_t^{0\prime}\hat{H}^{(0)}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}(1-q)\lambda_{il}^0\right\| + \left\|\frac{1}{NT}\sum_{t=1}^T F_{tr}^0F_t^{0\prime}\hat{H}^{(0)}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}(g_{it}-q)\lambda_{il}^0\right\| \\ &\equiv J_{2,1a}(l,r,1) + J_{2,1a}(l,r,2). \end{aligned}$$

For $J_{2,1a}(l,r,1)$, we have

$$J_{2,1a}(l,r,1) \leq (1-q)\left\|\hat{H}^{(0)}\right\|\frac{1}{T}\sum_{t=1}^T\left\|F_t^0\right\|^2\left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\lambda_{il}^0\right\| = O_P(1)\left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\lambda_{il}^0\right\| = O_P(\delta_{NT}^{-2}),$$

where we use the fact that $\left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\lambda_{il}^0\right\| \leq \left\|\hat{H}^{(0)}\right\|\left\|\frac{1}{N}\sum_{i=1}^N \frac{1}{Tq}\sum_{t=1}^T F_t^0\left[\varepsilon_{it}g_{it} + F_t^{0\prime}\lambda_i^0(g_{it}-q)\right]\lambda_{il}^0\right\|$

$= O_P(\delta_{NT}^{-2})$ by Chebyshev inequality. For $J_{2,1a}(l,r,2)$, we have

$$\begin{aligned} J_{2,1a}(l,r,2) &= \left\| \frac{1}{N}\sum_{i=1}^{N} \lambda_{il}^{0} \hat{\phi}_{\Lambda,i}^{(0)\prime} \hat{H}^{(0)\prime} \left[\frac{1}{T}\sum_{t=1}^{T} F_t^0 F_{tr}^0 (g_{it}-q) \right] \right\| \\ &\le \left\| \hat{H}^{(0)} \right\| \left\{ \frac{1}{N}\sum_{i=1}^{N} \left\| \lambda_i^0 \right\|^2 \left\| \hat{\phi}_{\Lambda,i}^{(0)} \right\|^2 \right\}^{1/2} \left\{ \frac{1}{N}\sum_{i=1}^{N} \left\| \frac{1}{T}\sum_{t=1}^{T} F_t^0 F_{tr}^0 (g_{it}-q) \right\|^2 \right\}^{1/2} \\ &= O_P\left(\delta_{NT}^{-1}\right) O_P(T^{-1/2}), \end{aligned}$$

as we can show that $\frac{1}{N}\sum_{i=1}^{N} \left\|\lambda_i^0\right\|^2 \left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2 = O_P(\delta_{NT}^{-2})$ and $\frac{1}{N}\sum_{i=1}^{N} E\left\|\frac{1}{T}\sum_{t=1}^{T} F_t^0 F_{tr}^0 (g_{it}-q)\right\|^2 = O(T^{-1})$. Then $J_{2,11} = O_P(\delta_{NT}^{-2})$. Similarly,

$$\begin{aligned} \left\| J_{2,12}(l,r) \right\| &= \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} \lambda_{ir}^{0} \lambda_i^{0\prime} (\hat{H}^{(0)\prime})^{-1} \hat{\phi}_{F,t}^{(0)} F_{tl}^0 \bar{g}_{it} \right\| \\ &\le \left\| \frac{1-q}{N}\sum_{i=1}^{N} \lambda_{ir}^{0} \lambda_i^{0\prime} (\hat{H}^{(0)\prime})^{-1} \frac{1}{T}\sum_{t=1}^{T} \hat{\phi}_{F,t}^{(0)} F_{tl}^0 \right\| + \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} \lambda_{ir}^{0} \lambda_i^{0\prime} (\hat{H}^{(0)\prime})^{-1} \hat{\phi}_{F,t}^{(0)} F_{tl}^0 (g_{it}-q) \right\| \\ &\equiv J_{2,12}(l,r,1) + J_{2,12}(l,r,2). \end{aligned}$$

Noting that

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \hat{\phi}_{F,t}^{(0)} F_{tl}^0 &= \hat{D}^{(0)-1} \frac{1}{T} \hat{F}^{(0)\prime} F^0 \frac{1}{T}\sum_{t=1}^{T} \frac{1}{Nq}\sum_{i=1}^{N} \lambda_i^0 \left[\varepsilon_{it} g_{it} + \lambda_i^{0\prime} F_t^0 (g_{it}-q) \right] F_{tl}^0 \\ &= O_P(1) \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} \lambda_i^0 \left[\varepsilon_{it} g_{it} + \lambda_i^{0\prime} F_t^0 (g_{it}-q) \right] F_{tl}^0 = O_P\left(\delta_{NT}^{-2}\right), \end{aligned}$$

$\left\| J_{2,12}(l,r,1) \right\| = O_P\left(\delta_{NT}^{-2}\right)$. For $J_{2,12}(l,r,2)$, we have

$$\begin{aligned} J_{2,12}(l,r,2) &= \left\| \frac{1}{T}\sum_{s=1}^{T} F_{sl}^0 \hat{\phi}_{F,s}^{(0)\prime} (\hat{H}^{(0)})^{-1} \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \lambda_{ir}^0 (g_{is}-q) \right\| \\ &\le \left\| (\hat{H}^{(0)})^{-1} \right\| \left\{ \frac{1}{T}\sum_{s=1}^{T} \left\| F_s^0 \right\|^2 \left\| \hat{\phi}_{F,s}^{(0)} \right\|^2 \right\}^{1/2} \left\{ \frac{1}{T}\sum_{s=1}^{T} \left\| \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \lambda_{ir}^0 (g_{is}-q) \right\|^2 \right\}^{1/2} \\ &= O_P\left(\delta_{NT}^{-1}\right) O_P(N^{-1/2}). \end{aligned}$$

So $J_{2,2} = O_P\left(\delta_{NT}^{-2}\right)$. Similarly, we can show that $J_{2,l} = O_P\left(\delta_{NT}^{-2}\right)$ for $l = 3, 4$. Then $\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N} F_s^0 \lambda_i^{0\prime} \eta_{is} \bar{g}_{is}$ $= O_P\left(\delta_{NT}^{-2}\right)$.

(ix) By (vi) and the fact that the fact that $\frac{1}{N}\sum_{i=1}^N E\left\|\frac{1}{T}\sum_{s=1}^T F_s^0\varepsilon_{is}g_{is}\right\|^2 = O(T^{-1})$

$$\begin{aligned}
\max_t \left\|\frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \eta_{it}^{(0)}\bar{g}_{it}\varepsilon_{is}g_{is}\right\| &= \max_t \left\|\frac{1}{N}\sum_{i=1}^N \eta_{it}^{(0)}\bar{g}_{it}\left(\frac{1}{T}\sum_{s=1}^T F_s^0\varepsilon_{is}g_{is}\right)\right\| \\
&\leq \left\{\max_t \frac{1}{N}\sum_{i=1}^N (\eta_{it}^{(0)})^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{s=1}^T F_s^0\varepsilon_{is}g_{is}\right\|^2\right\}^{1/2} \\
&= O_P\left(T^{-1/2+\gamma_1/4} + (N/\ln N)^{-1/2}\right) O_P(T^{-1/2}) \\
&= O_P\left(T^{-1+\gamma_1/4} + (NT/\ln N)^{-1/2}\right).
\end{aligned}$$

(x) Note that $\frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \varepsilon_{it}g_{it}\eta_{is}^{(0)}\bar{g}_{is} = \sum_{l=1}^2 \frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \varepsilon_{it}g_{it}\eta_{l,is}^{(0)}\bar{g}_{is} \equiv \sum_{l=1}^2 J_{3,lt}$.We can readily bound $J_{3,2t}$ by $O_P(\delta_{NT}^{-2}\ln N)$ by using the uniform bound for $\eta_{2,is}^{(0)}$ in (iii). For $J_{3,1t}$, we have

$$\begin{aligned}
J_{3,1t} &= \frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \varepsilon_{it}g_{it}\left[F_s^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,s}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,s}^{(0)} + F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}\right]\bar{g}_{is} \\
&\equiv J_{3,1t}(1) + J_{3,1t}(2) + J_{3,1t}(3) + J_{3,1t}(4).
\end{aligned}$$

Using $\bar{g}_{is} = (1-q) + (q - g_{is})$, the fact that $F_s^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}$ is a scalar and $\max_t \frac{1}{N}\sum_{i=1}^N \varepsilon_{it}^2 = O_P(1)$, and (iv), we have

$$\begin{aligned}
&\max_t J_{3,1t}(1) \\
&= \max_t \left\|\frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \varepsilon_{it}g_{it}F_s^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}\bar{g}_{is}\right\| \\
&\leq \max_t \left\|\frac{1-q}{T}\sum_{s=1}^T F_s^0 F_s^{0\prime}\hat{H}^{(0)}\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it}\right\| + \max_t \left\|\frac{1}{N}\sum_{i=1}^N \varepsilon_{it}g_{it}\hat{\phi}_{\Lambda,i}^{(0)\prime}\hat{H}^{(0)\prime}\frac{1}{T}\sum_{s=1}^T F_s^0 F_s^{0\prime}(g_{is}-q)\right\| \\
&\leq O_P(1)\left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it}\right\| + \max_i\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|\left\{\max_t \frac{1}{N}\sum_{i=1}^N \varepsilon_{it}^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{s=1}^T F_s^0 F_s^{0\prime}(g_{is}-q)\right\|^2\right\}^{1/2} \\
&= O_P(T^{-1+\gamma_1/4} + N^{-1}\ln N) + O_P((T/\ln T)^{-1/2})O_P(1)\,O_P(T^{-1/2}) = O_P(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N).
\end{aligned}$$

For $J_{3,1t}(2)$, we have by (i) and the fact that $\max_s\left\|\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}\bar{g}_{is}\right\| = O_P(N/\ln N)^{-1/2})$,

$$\begin{aligned}
\max_t J_{3,1t}(2) &= \max_t \left\|\frac{1}{T}\sum_{s=1}^T F_s^0\hat{\phi}_{F,s}^{(0)\prime}[\hat{H}^{(0)}]^{-1}\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}\bar{g}_{is}\right\| \\
&\leq \left\|[\hat{H}^{(0)}]^{-1}\right\|\max_t\left\|\hat{\phi}_{F,t}^{(0)}\right\|\max_s\left\|\frac{1}{N}\sum_{i=1}^N \lambda_i^0\varepsilon_{it}g_{it}\bar{g}_{is}\right\|\frac{1}{T}\sum_{s=1}^T\left\|F_s^0\right\| \\
&= O_P((N/\ln N)^{-1/2})O_P(N/\ln N)^{-1/2}) = O_P\left(\delta_{NT}^{-2}\ln N\right).
\end{aligned}$$

Similarly, we can show that $J_{3,1t}(l) = O_P(\delta_{NT}^{-2}\ln N)$ for $l = 3, 4$. Then $\max_t \left\| \frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \varepsilon_{it} g_{it} \eta_{is}^{(0)} \bar{g}_{is} \right\|$ $= O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$. ■

Proof of Lemma A.4 ($\ell = 1$ and $\ell > 1$)**.** (i) From the PCA, we have the identity $(NT)^{-1}\hat{X}^{(\ell)}\hat{X}^{(\ell)\prime}\hat{F}^{(\ell)}$ $= \hat{F}^{(\ell)}\hat{D}^{(\ell)}$. Pre-multiplying both sides by $T^{-1}\hat{F}^{(\ell)\prime}$ and using the normalization $T^{-1}\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)} = I_R$ yield $T^{-1}\hat{F}^{(\ell)\prime}(NT)^{-1}\hat{X}\hat{X}^{\prime}\hat{F}^{(\ell)} = \hat{D}^{(\ell)}$. Let $\varepsilon^{(\ell)}$ be the $T \times N$ matrix with (t, i)th element given by $\varepsilon_{it}^{(\ell-1)} = \varepsilon_{it} g_{it} + \eta_{it}^{(\ell-1)}\bar{g}_{it}$. Noting that $\hat{X}^{(\ell)} = F^0\Lambda^{0\prime} + \varepsilon_{it}^{(\ell)}$, we have

$$\begin{aligned}
\hat{D}^{(\ell)} &= T^{-1}\hat{F}^{(\ell)\prime}(NT)^{-1}(F^0\Lambda^{0\prime} + \varepsilon^{(\ell)})(F^0\Lambda^{0\prime} + \varepsilon^{(\ell)})^{\prime}\hat{F}^{(\ell)} \\
&= T^{-1}\hat{F}^{(\ell)\prime}(NT)^{-1}\left\{F^0\Lambda^{0\prime}\Lambda^0 F^{0\prime} + \varepsilon^{(\ell)}\varepsilon^{(\ell)\prime} + F^0\Lambda^{0\prime}\varepsilon^{(\ell)\prime} + \varepsilon^{(\ell)}\Lambda^0 F^{0\prime}\right\}\hat{F}^{(\ell)} \\
&\equiv \hat{D}_1^{(\ell)} + \hat{D}_2^{(\ell)} + \hat{D}_3^{(\ell)} + \hat{D}_4^{(\ell)}.
\end{aligned}$$

The result follows if we show that (1) $\hat{D}_1^{(\ell)} = D + O_P(\delta_{NT}^{-1}\ln N)$ and (2) $\hat{D}_l^{(\ell)} = O_P(\delta_{NT}^{-1}\ln N)$ for $l = 2, 3, 4$. Following the proof of Lemma A.1(i) in Su and Wang (2017), $\hat{D}_1^{(\ell)} = \frac{\hat{F}^{(\ell)\prime}F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\hat{F}^{(\ell)}}{T} = D + O_P\left(\delta_{NT}^{-1}\right)$. Noting that $\varepsilon^{(\ell)} = \varepsilon \circ G + \eta^{(\ell-1)} \circ \bar{G}$ where $\bar{G} = \mathbf{1}_{T\times N} - G$ and $\eta^{(\ell-1)}$ has (t, i)th element given by $\eta_{it}^{(\ell-1)}$,

$$\begin{aligned}
\left\|\hat{D}_2^{(\ell)}\right\|_{\text{sp}} &= (NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}(\varepsilon \circ G + \eta^{(\ell-1)} \circ \bar{G})(\varepsilon \circ G + \eta^{(\ell-1)} \circ \bar{G})^{\prime}\hat{F}^{(\ell)}\right) \\
&\leq 2(NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}(\varepsilon \circ G)(\varepsilon \circ G)^{\prime}\hat{F}^{(\ell)}\right) \\
&\quad + 2(NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}(\eta^{(\ell-1)} \circ \bar{G})(\eta^{(\ell-1)} \circ \bar{G})^{\prime}\hat{F}^{(\ell)}\right).
\end{aligned}$$

Following the analysis of $D_{NT,2}$ in the proof of Lemma A.1, we can show that the first term is $O_P(\delta_{NT}^{-2}[\log(N \vee T)]^2)$. For the second term, it suffices to use Lemma A.3(vii) to obtain the following rough probability bound

$$2(NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}\eta^{(\ell-1)}\eta^{(\ell-1)\prime}\hat{F}^{(\ell)}\right) \leq 2T^{-1}\left\|\hat{F}^{(\ell)}\right\|^2(NT)^{-1}\left\|\eta^{(\ell-1)}\right\|^2 = O_P(\delta_{NT}^{-2}).$$

It follows that $\left\|\hat{D}_2\right\| \leq R^{1/2}\left\|\hat{D}_2\right\|_{\text{sp}} = O_P(\delta_{NT}^{-2}(\ln N)^2)$. By the CS inequality, $\left\|\hat{D}_3^{(\ell)}\right\| = \left\|\hat{D}_4^{(\ell)}\right\| \leq \left\{\left\|\hat{D}_1^{(\ell)}\right\|\left\|\hat{D}_2^{(\ell)}\right\|\right\}^{1/2} = O_P(\delta_{NT}^{-1}\ln N))$. In sum, we have $\hat{D}^{(\ell)} = D + O_P(\delta_{NT}^{-1}\ln N)$.

(ii) The proof is analogous to that of Lemma A.2(i) with obvious modifications.

(iii) The proof is analogous to that of Lemma A.2(ii) with obvious modifications.

(iv) The proof follows from that of Lemma B.3 in Bai (2003).

(v) Note that $\frac{1}{T}\sum_{t=1}^T(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0)\varepsilon_{it}^{(\ell)} = \frac{1}{T}\sum_{t=1}^T(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0)\varepsilon_{it}g_{it} + \frac{1}{T}\sum_{t=1}^T(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0)\eta_{it}^{(\ell)}\bar{g}_{it}$. Following the proof of Lemma A.7(v) in Su and Wang (2017), we can show that the first term is $O_P\left(\delta_{NT}^{-2}\ln N\right)$ uniformly in t. By Theorem 2.3 and Lemma A.3(vi), the Frobenius norm of the second term is bounded above by above by$\{\frac{1}{T}||\hat{F}^{(\ell)} - F^0\hat{H}^{(\ell)}||^2\}^{1/2}\{\max_i \frac{1}{T}\sum_{t=1}^T(\eta_{it}^{(\ell)})^2\}^{1/2} = \delta_{NT}^{-1}O_P(N^{-1/2+\gamma_2/4} + (T/\ln N)^{-1/2})$. It follows that $\frac{1}{T}\sum_{t=1}^T(\hat{F}_t^{(\ell)} - \hat{H}^{(\ell)\prime}F_t^0)\varepsilon_{it}^{(\ell)} = O_P(N^{-1/2+\gamma_2/4}\delta_{NT}^{-1} + \delta_{NT}^{-2}\ln N)$. ■

Proof of Lemma A.5 ($\ell = 1$ and $\ell > 1$). (i) Note that $\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}^{(\ell)} = \beta_{F,t} + \frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\eta_{it}^{(\ell-1)}\bar{g}_{it}$ where

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\eta_{it}^{(\ell-1)}\bar{g}_{it} = \sum_{l=1}^{2}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\eta_{l,it}^{(\ell-1)}\bar{g}_{it} \equiv \sum_{l=1}^{2}K_{1,lt}.$$

By Lemma A.3(iii), $\max_t\|K_{1,2t}\| \leq \max_{i,t}\left\|\eta_{2,it}^{(\ell-1)}\right\|\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\| = O_P(\delta_{NT}^{-2}\ln N)$. For $K_{1,1t}$, we make the following decomposition

$$\begin{aligned}K_{1,1t} &= \frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\left[F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)} + F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\right]\bar{g}_{it}\\ &\equiv K_{1,1t}(1) + K_{1,1t}(2) + K_{1,1t}(3) + K_{1,1t}(4).\end{aligned}$$

For $K_{1,1t}(1)$, we apply Lemma A.3(iv) to obtain $\max_t\|K_{1,1t}(1)\| = O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln N\right)$. For $K_{1,1t}(2)$, we have uniformly in t

$$\begin{aligned}K_{1,1t}(2) &= \frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)}\bar{g}_{it}\\ &= (1-q)\frac{1}{N}\Lambda^{0\prime}\Lambda^0(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)} + \left(\frac{1}{N}\sum_{i=1}^{N}(g_{it}-q)\lambda_i^0\lambda_i^{0\prime}\right)(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)}\\ &= (1-q)\left[[\hat{D}^{(\ell-1)}]^{-1}\frac{1}{T}\hat{F}^{(\ell-1)\prime}F^0\right]^{-1}\hat{\phi}_{F,t}^{(\ell-1)} + O_P(N^{-1}\ln N),\end{aligned}$$

where the second equality follows from the use of $\bar{g}_{it} = (1-q) - (g_{it}-q)$, the third equality holds by (i), the fact that $\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}(g_{it}-q)\right\| = O_P\left(N^{-1/2}\ln N\right)$, and the definition of $\hat{H}^{(\ell-1)}$. In addition, we have by (ii) and (iv)

$$\begin{aligned}\max_t\|K_{1,1t}(3)\| &= \max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}\bar{g}_{it}\right\|\\ &\leq \max_t\left\|\hat{r}_{F,t}^{(\ell-1)}\right\|O_P\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\lambda_i^{0\prime}\right\|\right) = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})\end{aligned}$$

and

$$\begin{aligned}\max_t\|K_{1,1t}(4)\| &= \max_t\left\|F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\frac{1}{N}\sum_{i=1}^{N}\hat{r}_{\Lambda,i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\|\\ &\leq \max_t\left\|F_t^{0\prime}\right\|\max_t\left\|\hat{H}^{(\ell-1)\prime}\frac{1}{N}\sum_{i=1}^{N}\hat{r}_{\Lambda,i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln N).\end{aligned}$$

It follows that uniformly in t, $\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}^{(\ell)} = \beta_{F,t} + (1-q)\left[[\hat{D}^{(\ell-1)}]^{-1}\frac{1}{T}\hat{F}^{(\ell-1)\prime}F^0\right]^{-1}\hat{\phi}_{F,t}^{(\ell-1)} + O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})$ and

$$\hat{\phi}_{F,t}^{(\ell)} = [\hat{D}^{(\ell)}]^{-1}\frac{1}{T}\hat{F}^{(\ell)\prime}F^0\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}^{(\ell)} = D^{-1}Q\beta_{F,t} + (1-q)\hat{\phi}_{F,t}^{(\ell-1)} + O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_2/4}).$$

(ii) Note that $\frac{1}{T}\sum_{t=1}^{T}F_t^0\varepsilon_{it}^{(\ell)}=\beta_{\Lambda,i}+\frac{1}{T}\sum_{t=1}^{T}F_t^0\eta_{it}^{(\ell-1)}\bar{g}_{it}$ where $\frac{1}{T}\sum_{t=1}^{T}F_t^0\eta_{it}^{(\ell-1)}\bar{g}_{it}=\sum_{l=1}^{2}\frac{1}{T}\sum_{t=1}^{T}F_t^0\eta_{l,it}^{(\ell-1)}\bar{g}_{it}\equiv\sum_{l=1}^{2}K_{2i,l}$. By Lemma A.3(iii), we can show that that $\max_i\|K_{2i,2}\|=O_P\left(\delta_{NT}^{-2}\ln N\right)$. Using the decomposition $\bar{g}_{it}=(1-q)+(g_{it}-q)$ and Lemma A.2, we can readily show that

$$\begin{aligned}K_{2i,1} &= \frac{1}{T}\sum_{t=1}^{T}F_t^0\left[F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)}+\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}+F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\right]\bar{g}_{it}\\ &\equiv K_{2i,1}(1)+K_{2i,1}(2)+K_{2i,1}(3)+K_{2i,1}(4).\end{aligned}$$

For $K_{2i,1}(1)$, we have that uniformly in i,

$$\begin{aligned}K_{2i,1}(1) &= \frac{1-q}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}(g_{it}-q)\\ &= (1-q)\frac{1}{T}F^{0\prime}F^0\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+O_P((NT)^{-1/2}\ln N)\\ &= (1-q)[\hat{H}^{(\ell-1)\prime}]^{-1}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+O_P((NT)^{-1/2}\ln N),\end{aligned}$$

where the second equality follows from the fact that

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}(g_{it}-q)\right\| &\leq O_P\left(\max_i\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|\right)\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}(g_{it}-q)\right\|\\ &= O_P((N/\ln N)^{-1/2})O_P((T/\ln T)^{-1/2})\end{aligned}$$

and the last equality follows because $\frac{1}{T}F^{0\prime}F^0=[\hat{H}^{(\ell-1)}\hat{H}^{(\ell-1)\prime}]^{-1}+O_P(\delta_{NT}^{-2})$. By Lemma A.3(v) and (ii)

$$\begin{aligned}\max_i\|K_{2i,1}(2)\| &= \max_i\left\|\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{\phi}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\|\\ &\leq O_P\left(\max_i\left\|\lambda_i^0\right\|\right)\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\phi}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\|=N^{\gamma_2/4}O_P\left(\delta_{NT}^{-2}\ln N+N^{-1+\gamma_2/2}\right),\\ \max_i\|K_{2i,1}(3)\| &= \max_i\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}\bar{g}_{it}\right\|\\ &\leq O_P\left(\max_i\left\|\lambda_i^0\right\|\right)\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{r}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\|=N^{\gamma_2/4}O_P\left(\delta_{NT}^{-2}\ln N\right),\end{aligned}$$

and

$$\begin{aligned}\max_i\|K_{2i,1}(4)\| &= \max_i\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\bar{g}_{it}\right\|\\ &\leq O_P\left(\max_i\left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\|\right)\frac{1}{T}\sum_{t=1}^{T}\left\|F_t^0\right\|^2=O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N).\end{aligned}$$

It follows that uniformly in i, $\frac{1}{T}\sum_{t=1}^{T} F_t^0 \varepsilon_{it}^{(\ell)} = \beta_{\Lambda,i} + (1-q)[\hat{H}^{(\ell-1)\prime}]^{-1}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right)$ and

$$\hat{\phi}_{\Lambda,i}^{(\ell)} = \hat{H}^{(\ell-1)\prime}\frac{1}{T}\sum_{t=1}^{T} F_t^0 \varepsilon_{it}^{(\ell)} = (Q')^{-1}\beta_{\Lambda,i} + (1-q)\hat{\phi}_{\Lambda,i}^{(\ell-1)} + O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N + N^{-1+3\gamma_2/4}). \blacksquare$$

Proof of Lemma A.3 ($\ell \geq 2$). The proof relies on the fact that Lemmas A.3-A.5 and Theorems 2.3-2.4 hold for $\ell - 1$.

(i) By Lemma A.5(i)-(ii),

$$\begin{aligned}
\max_t \left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\| &= \max_t \left\|D^{-1}Q\beta_{F,t} + (1-q)\hat{\phi}_{F,t}^{(\ell-2)} + O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})\right\| \\
&\leq \left\|D^{-1}Q\right\| \max_t \left\|\beta_{F,t}\right\| + (1-q)\max_t \left\|\hat{\phi}_{F,t}^{(\ell-2)}\right\| + O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}) \\
&= O_P((N/\ln N)^{-1/2}) + O_P((N/\ln N)^{-1/2}) + o_P((N \vee T)^{-1/2}) = O_P((N/\ln N)^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
\max_i \left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\| &= \max_i \left\|(Q')^{-1}\beta_{\Lambda,i} + (1-q)\hat{\phi}_{\Lambda,i}^{(\ell-2)} + O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N)\right\| \\
&\leq \left\|(Q')^{-1}\right\| \max_i \left\|\beta_{\Lambda,i}\right\| + (1-q)\max_i \left\|\hat{\phi}_{\Lambda,i}^{(\ell-2)}\right\| + O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N) \\
&= O_P((T/\ln T)^{-1/2}) + O_P((T/\ln T)^{-1/2}) + o_P((N \vee T)^{-1/2}) = O_P((T/\ln T)^{-1/2}).
\end{aligned}$$

(ii) By the decomposition in (A.8), $\hat{r}_{F,t}^{(\ell-1)} = \hat{F}_t^{(\ell-1)} - \hat{H}^{(\ell-1)\prime}F_t^{\ell-1} - \hat{\phi}_{F,t}^{(\ell-1)} = \hat{a}_{1t}^{(\ell-1)} + \hat{a}_{3t}^{(\ell-1)} + (\hat{a}_{2t}^{(\ell-1)} - \hat{\phi}_{F,t}^{(\ell-1)})$. Following the proof of Theorem 2.4(i) and using Assumption A.5 and the fact that $\max_t \left\|F_t^0\right\| = O_P(T^{\gamma_1/4})$, it is easy to show that

$$\max_t \left\|\hat{a}_{1t}^{(\ell-1)}\right\| = O_P(T^{-1/2}\delta_{NT}^{-1} + T^{-1+\gamma_1/4}), \quad \max_t \left\|\hat{a}_{3t}^{(\ell-1)}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}),$$

and $\max_t \left\|\hat{a}_{2t}^{(\ell-1)} - \hat{\phi}_{F,t}^{(\ell-1)}\right\| = O_P(\delta_{NT}^{-2}\ln N)$. It follows that $\max_t \left\|\hat{r}_{F,t}^{(\ell-1)}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})$. For $\hat{r}_{\Lambda,i}^{(\ell-1)}$, we have

$$\hat{r}_{\Lambda,i}^{(\ell-1)} = \hat{\lambda}_i^{(\ell-1)} - (\hat{H}^{(\ell-1)})^{-1}\lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(\ell-1)} = \hat{B}_{2i}^{(\ell-1)} + \hat{B}_{3i}^{(\ell-1)} + (\hat{B}_{1i}^{(\ell-1)} - \hat{\phi}_{\Lambda,i}^{(\ell-1)}),$$

where $\hat{B}_{li}^{(\ell)}$'s are defined in the proof of Theorem 2.4(ii). Following the proof of Theorem 2.4(ii) and using the fact that $\max_i \left\|\lambda_i^0\right\| = O_P(N^{\gamma_2/4})$, $\max_i \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 = O_P(1)$, and $\tilde{q} - q = O_P((NT)^{-1/2})$ we have by Lemma A.4

$$\max_i \left\|\hat{B}_{2i}^{(\ell-1)}\right\| = O_P(N^{-1/2+\gamma_2/4}\delta_{NT}^{-1} + \delta_{NT}^{-2}\ln N), \quad \max_i \left\|\hat{B}_{3i}^{(\ell-1)}\right\| = O_P(N^{\gamma_2/4}\delta_{NT}^{-2}),$$

and $\max_i \left\|\hat{B}_{1i}^{(\ell-1)} - \hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\| = O_P(\delta_{NT}^{-2}\ln N)$. It follows that $\max_i \left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\| = O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N)$.

(iii) The proof is similar to the $\ell = 1$ case by replacing the superscript 0 by $\ell - 1$ throughout the proof.

(iv) By Assumption A.5 and Lemma A.3(x) below

$$\begin{aligned}
\left(\hat{H}^{(\ell-1)\prime}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\varepsilon_{it}g_{it} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\left[\varepsilon_{is}g_{is}+\eta_{is}^{(\ell-1)}\bar{g}_{is}\right]\varepsilon_{it}g_{it} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0E\left(\varepsilon_{is}\varepsilon_{it}\right)g_{is}g_{it}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\left[\varepsilon_{is}\varepsilon_{it}-E(\varepsilon_{is}\varepsilon_{it})\right]g_{is}g_{it} \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\eta_{is}^{(\ell-1)}\bar{g}_{is}\varepsilon_{it}g_{it} \\
&= O_P(T^{-1+\gamma_1/4})+O_P\left(\delta_{NT}^{-2}\ln N\right)+O_P(T^{-1+\gamma_1/4}+\delta_{NT}^{-2}\ln N).
\end{aligned}$$

Then $\max_i\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\varepsilon_{it}g_{it}\right\|=O_P\left(T^{-1+\gamma_1/4}+\delta_{NT}^{-2}\ln N\right).$

Note that $\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\|\leq\left\|\hat{H}^{(\ell-1)\prime}\right\|\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\|$. Using $\bar{g}_{it}=(1-q)-(g_{it}-q)$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\left[\varepsilon_{is}g_{is}+\eta_{is}^{(\ell-1)}\bar{g}_{is}\right]\bar{g}_{it} \\
&= \frac{1-q}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\varepsilon_{is}g_{is}+\frac{1-q}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is} \\
&\quad -\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\varepsilon_{is}g_{is}(g_{it}-q)-\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}(g_{it}-q).
\end{aligned}$$

It is easy to show that the first term is $O_P\left(\delta_{NT}^{-2}\right)$ by Chebyshev inequality. The second term is $O_P(\delta_{NT}^{-2})$ by Lemma A.3(viii) below. The third term is $O_P(\delta_{NT}^{-2}\ln N)$ by Assumption A.5(iii). By Lemma A.3(iii),

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}(g_{it}-q)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\eta_{1,is}^{(\ell-1)}\bar{g}_{is}(g_{it}-q)+O_P(\delta_{NT}^{-2}\ln N)$$

uniformly in t. Now we make the following decomposition

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\eta_{1,is}^{(\ell-1)}\bar{g}_{is}(g_{it}-q) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}[F_s^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,s}^{(\ell-1)} \\
&\quad +\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,s}^{(\ell-1)}+F_s^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}]\bar{g}_{is}(g_{it}-q) \\
&\equiv II_{1t}+II_{2t}+II_{3t}+II_{4t}.
\end{aligned}$$

For II_{3t} and II_{4t}, we apply Lemma A.3(ii) to obtain the rough bound

$$\begin{aligned}
\max_t \|II_{3t}\| &\leq \max_t \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{s=1}^T F_s^0 \lambda_i^{0\prime} \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{r}_{F,s}^{(\ell-1)} \bar{g}_{is}(g_{it}-q) \right\| \\
&\leq O_P(1) \max_s \left\| \hat{r}_{F,s}^{(\ell-1)} \right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2} \ln T + T^{-1+3\gamma_1/4}) \text{ and} \\
\max_t \|II_{4t}\| &= \max_t \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{s=1}^T F_s^0 \lambda_i^{0\prime} \left[F_s^{0\prime} \hat{H}^{(\ell-1)\prime} \hat{r}_{\Lambda,i}^{(\ell-1)} \right] \bar{g}_{is}(g_{it}-q) \right\| \\
&\leq O_P(1) \max_i \left\| \hat{r}_{\Lambda,i}^{(\ell-1)} \right\| = O_P(N^{\gamma_2/4}\delta_{NT}^{-2} \ln N).
\end{aligned}$$

By Lemma A.5(i), we have

$$\begin{aligned}
\max_t \|II_{1t}\| &= \max_t \left\| \frac{1}{NT} \sum_{s=1}^T \left[\sum_{i=1}^N F_s^0 \lambda_i^{0\prime} F_s^{0\prime} \hat{H}^{(\ell-1)} \hat{\phi}_{\Lambda,i}^{(\ell-1)} \right] \bar{g}_{is}(g_{it}-q) \right\| \\
&\leq \max_t \left\| \frac{1}{NT} \sum_{s=1}^T \sum_{i=1}^N F_s^0 \mathrm{tr} \left[\lambda_i^{0\prime} F_s^{0\prime} \hat{H}^{(\ell-1)} (Q')^{-1} \beta_{\Lambda,i} \bar{g}_{is}(g_{it}-q) \right] \right\| \\
&\quad + \max_t \left\| \frac{1}{NT} \sum_{s=1}^T \sum_{i=1}^N F_s^0 \mathrm{tr} \left[\lambda_i^{0\prime} F_s^{0\prime} \hat{H}^{(\ell-1)} (1-q) \hat{\phi}_{\Lambda,i}^{(\ell-2)} \bar{g}_{is}(g_{it}-q) \right] \right\| + O_P(N^{\gamma_2/4}\delta_{NT}^{-2} \ln N) \\
&\leq O_P(1) \max_{s,t} \left\| \frac{1}{N} \sum_{i=1}^N \beta_{\Lambda,i} \lambda_i^{0\prime} \bar{g}_{is}(g_{it}-q) \right\| + O_P(1) \max_t \left\| \frac{1}{N} \sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(\ell-2)} \lambda_i^{0\prime} \bar{g}_{is}(g_{it}-q) \right\| \\
&\quad + O_P(N^{\gamma_2/4}\delta_{NT}^{-2} \ln N) \\
&= O_P(\delta_{NT}^{-2} \ln N) + O_P(T^{-1+\gamma_1/2} + \delta_{NT}^{-2} \ln N) + O_P(N^{\gamma_2/4}\delta_{NT}^{-2} \ln N) \\
&= O_P(T^{-1+\gamma_1/2} + N^{\gamma_2/4}\delta_{NT}^{-2} \ln N)
\end{aligned}$$

Similarly, using Lemma A.5(ii), we can show that

$$\max_t \|II_{2t}\| = \max_t \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{s=1}^T F_s^0 \lambda_i^{0\prime} \left[\lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{\phi}_{F,s}^{(\ell-1)} \right] \bar{g}_{is}(g_{it}-q) \right\| = O_P(T^{-1+\gamma_1/2} + \delta_{NT}^{-2} \ln N).$$

Noting that $\hat{r}_{\Lambda,i}^{(\ell-1)} = \hat{\lambda}_i^{(\ell-1)} - (\hat{H}^{(\ell-1)})^{-1} \lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(\ell-1)} = \hat{B}_{2i}^{(\ell-1)} + \hat{B}_{3i}^{(\ell-1)}$, we have $\max_t \left\| \frac{1}{N} \sum_{i=1}^N \hat{r}_{\Lambda,i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\|$ $\leq \max_t \left\| \frac{1}{N} \sum_{i=1}^N [\hat{B}_{2i}^{(\ell-1)} + \hat{B}_{3i}^{(\ell-1)}] \lambda_i^{0\prime} \bar{g}_{it} \right\|$, where $\hat{B}_{li}^{(\ell-1)}$'s are defined in the proof of Theorem 2.4(ii). By Lemma A.4(iv)-(v), we have

$$\begin{aligned}
\max_t \left\| \frac{1}{N} \sum_{i=1}^N \hat{B}_{2i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| &= \max_t \frac{1}{NT} \left\| (\hat{H}^{(\ell-1)})^{-1} \sum_{i=1}^N \sum_{s=1}^T (\hat{H}^{(\ell-1)\prime} F_s^0 - \hat{F}_s^{(\ell-1)}) \varepsilon_{is}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| \\
&\leq O_P(1) \max_s \left\| \frac{1}{T} \sum_{s=1}^T (\hat{H}^{(\ell-1)\prime} F_s^0 - \hat{F}_s^{(\ell-1)}) \varepsilon_{is}^{(\ell-1)} \right\| = O_P\left(\delta_{NT}^{-2} \ln N \right),
\end{aligned}$$

and

$$\max_t \left\| \frac{1}{N}\sum_{i=1}^{N} \hat{B}_{3i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| = \max_t \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T} \hat{F}_s^{(\ell-1)} (\hat{H}^{(\ell-1)\prime} F_s^0 - \hat{F}_s^{(\ell-1)})' (\hat{H}^{(\ell-1)\prime})^{-1} \lambda_i^{0\prime} \bar{g}_{it} \right\|$$
$$\leq O_P(1) \left\| \frac{1}{N}\sum_{s=1}^{T} \hat{F}_s^{(\ell-1)} (\hat{H}^{(\ell-1)\prime} F_s^0 - \hat{F}_s^{(\ell-1)})' \right\| = O_P(\delta_{NT}^{-2}).$$

In sum, we have $\max_t \left\| \hat{H}^{(\ell-1)\prime} \frac{1}{N}\sum_{i=1}^{N} \hat{r}_{\Lambda,i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2} \ln N\right)$.

(v) By the definition of $\hat{\phi}_{F,t}^{(\ell-1)}$ and $\varepsilon_{jt}^{(\ell-1)}$, we have

$$\max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{\phi}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\|$$
$$\leq O_P(1) \max_i \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 \varepsilon_{jt}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\|$$
$$\leq O_P(1) \max_i \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \varepsilon_{jt} g_{jt} \bar{g}_{it} \right\| + O_P(1) \max_i \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \eta_{jt}^{(\ell-2)} \bar{g}_{jt} \bar{g}_{it} \right\|.$$

We can show that the first term is $O_P(\delta_{NT}^{-2} \ln N)$ by applying Assumption A.5(iii). For the second term, we have by Lemma A.3(viii) and (iii)

$$\max_i \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \eta_{jt}^{(\ell-2)} \bar{g}_{jt} \bar{g}_{it} \right\|$$
$$\leq \left\| \frac{q-1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \eta_{jt}^{(\ell-2)} \bar{g}_{jt} \right\| + \max_i \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \eta_{jt}^{(\ell-2)} \bar{g}_{jt} (g_{it}-q) \right\|$$
$$= O_P\left(\delta_{NT}^{-2}\right) + \max_i \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \eta_{jt}^{(\ell-2)} \bar{g}_{jt} (g_{it}-q) \right\|$$
$$= \max_i \left\| \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \eta_{1,jt}^{(\ell-2)} \bar{g}_{jt} (g_{it}-q) \right\| + O_P\left(\delta_{NT}^{-2} \ln N\right).$$

Noting that $\eta_{1,it}^{(\ell)} = F_t^{0\prime} \hat{H}^{(\ell)} \hat{\phi}_{\Lambda,i}^{(\ell)} + \lambda_i^{0\prime} (\hat{H}^{(\ell)\prime})^{-1} \hat{\phi}_{F,t}^{(\ell)} + \lambda_i^{0\prime} (\hat{H}^{(\ell)\prime})^{-1} \hat{r}_{F,t}^{(\ell)} + F_t^{0\prime} \hat{H}^{(\ell)\prime} \hat{r}_{\Lambda,i}^{(\ell)}$, we have

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} \eta_{1,jt}^{(\ell-2)} \bar{g}_{jt} (g_{it}-q) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} \lambda_j^0 F_t^{0\prime} [F_t^{0\prime} \hat{H}^{(\ell-2)} \hat{\phi}_{\Lambda,j}^{(\ell-2)} + \lambda_j^{0\prime} (\hat{H}^{(\ell-2)\prime})^{-1} \hat{\phi}_{F,t}^{(\ell-2)}$$
$$+ \lambda_j^{0\prime} (\hat{H}^{(\ell-2)\prime})^{-1} \hat{r}_{F,t}^{(\ell-2)} + F_t^{0\prime} \hat{H}^{(\ell-2)\prime} \hat{r}_{\Lambda,j}^{(\ell-2)}] \bar{g}_{jt} (g_{it}-q)$$
$$\equiv III_{1i} + III_{2i} + III_{3i} + III_{4i}.$$

For the first term, we have

$$\begin{aligned}\max_i \|III_{1i}\| &= \max_i \left\| \frac{1}{N}\sum_{j=1}^{N} \lambda_j^0 \hat{\phi}_{\Lambda,j}^{(\ell-2)\prime} \hat{H}^{(\ell-2)\prime} \frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \bar{g}_{jt}(g_{it}-q) \right\| \\ &\le \max_j \left\| \hat{\phi}_{\Lambda,j}^{(\ell-2)} \right\| \max_j \left\| \frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \bar{g}_{jt}(g_{it}-q) \right\| \\ &= O_P((T/\ln T)^{-1/2}) O_P((T/\ln T)^{-1/2} + T^{-1+\gamma_1/2}) = O_P\left(\delta_{NT}^{-2}\ln N\right).\end{aligned}$$

Similarly, we can show that $\max_i \|III_{4i}\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$ and $\max_i \|III_{4i}\| = O_P\left(\delta_{NT}^{-2}\ln N + N^{-1+\gamma_2/2}\right)$ for $l = 2, 3$. Then $\max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{\phi}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2}\ln N + N^{-1+\gamma_2/2}\right)$.

Noting that $\hat{r}_{F,t}^{(\ell-1)} = \hat{F}_t^{(\ell-1)} - \hat{H}^{(\ell-1)\prime} F_t^0 - \hat{\phi}_{F,t}^{(\ell-1)} = \hat{a}_{1t}^{(\ell-1)} + \hat{a}_{3t}^{(\ell-1)}$ by (A.8), we have

$$\left\| \frac{1}{T}\sum_{t=1}^{T} \hat{r}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| \le \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{a}_{1t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| + \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{a}_{3t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\|.$$

Note that

$$\begin{aligned}\max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{a}_{1t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| &\le O_P(1) \max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \frac{1}{NT}\sum_{s=1}^{T} \hat{F}_s^{(\ell-1)} \sum_{j=1}^{N} \varepsilon_{jt}^{(\ell-1)} \varepsilon_{js}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| \\ &= O_P(1) \max_i \left\| \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N} F_s^0 \varepsilon_{jt}^{(\ell-1)} \varepsilon_{js}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) \\ &\le O_P(1) \max_i \frac{1}{N}\sum_{j=1}^{N} \left\| \frac{1}{T}\sum_{t=1}^{T} F_t^0 \varepsilon_{jt}^{(\ell-1)} \bar{g}_{it} \right\| \left\| \frac{1}{T}\sum_{s=1}^{T} F_s^0 \varepsilon_{js}^{(\ell-1)} \right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) \\ &\le O_P(1) \frac{1}{N}\sum_{j=1}^{N} \left\| \frac{1}{T}\sum_{s=1}^{T} F_s^0 \varepsilon_{js}^{(\ell-1)} \right\|^2 + O_P\left(\delta_{NT}^{-2}\ln N\right).\end{aligned}$$

Using the decomposition $\varepsilon_{jt}^{(\ell-1)} = \varepsilon_{jt} g_{jt} + \eta_{jt}^{(\ell-2)} \bar{g}_{jt}$ and Assumption A.5, we can show that $\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N} \varepsilon_{jt}^{(\ell-1)} \varepsilon_{js}^{(\ell-1)} F_s^0 F_t^{0\prime} \bar{g}_{it} = O_P\left(\delta_{NT}^{-2}\ln N\right)$. Then $\max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{a}_{1t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$. Similarly,

$$\begin{aligned}\max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{a}_{3t}^{(\ell)} F_t^{0\prime} \bar{g}_{it} \right\| &\le O_P(1) \max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \left[\frac{1}{NT}\sum_{s=1}^{T} \hat{F}_s^{(\ell)} \sum_{j=1}^{N} \lambda_j^{0\prime} F_t^0 \varepsilon_{js}^{(\ell)} \right] F_t^{0\prime} \bar{g}_{it} \right\| \\ &= O_P(1) \max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \left(\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} F_s^0 \varepsilon_{js}^{(\ell)} \lambda_j^{0\prime} \right) F_t^0 F_t^{0\prime} \bar{g}_{it} \right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) \\ &= O_P(1) \left\| \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N} F_j^0 \varepsilon_{is}^{(\ell)} \lambda_j^{0\prime} \right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) = O_P\left(\delta_{NT}^{-2}\ln N\right).\end{aligned}$$

It follows that $\max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{r}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$.

(vi) Note that $\frac{1}{N}\sum_{i=1}^{N}\left\|\eta_{it}^{(\ell-1)}\right\|^2 \leq \frac{2}{N}\sum_{i=1}^{N}\left\|\eta_{1,it}^{(\ell-1)}\right\|^2 + \frac{2}{N}\sum_{i=1}^{N}\left\|\eta_{2,it}^{(\ell-1)}\right\|^2$, where the second term is bounded above by $O_P(\delta_{NT}^{-4}(\ln N)^2)$ by (iii). For the first term, we have

$$\begin{aligned}
&\max_t \frac{1}{N}\sum_{i=1}^{N}\left\|\eta_{1,it}^{(\ell-1)}\right\|^2 \\
\leq\ & \max_t \frac{1}{N}\sum_{i=1}^{N}\left\|F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)} + F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\right\|^2 \\
\leq\ & 4\left\|\hat{H}^{(\ell-1)}\right\|\max_t\left\|F_t^0\right\|^2\frac{1}{N}\sum_{i=1}^{N}(\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2 + \left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\|^2) \\
&+4\left\|(\hat{H}^{(\ell-1)\prime})^{-1}\right\|\left\{\max_t\left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\|^2 + \max_t\left\|\hat{r}_{F,t}^{(\ell-1)}\right\|^2\right\}\frac{1}{N}\sum_{i=1}^{N}\max_i\left\|\lambda_i^0\right\|^2 \\
=\ & O_P(T^{-1+\gamma_1/2} + N^{-1}\ln N).
\end{aligned}$$

It follows that $\frac{1}{N}\sum_{i=1}^{N}\left\|\eta_{it}^{(\ell-1)}\right\|^2 = O_P(T^{-1+\gamma_1/2}+N^{-1}\ln N)$. Similarly, we can show that $\max_t \frac{1}{T}\sum_{t=1}^{T}\left\|\eta_{it}^{(\ell-1)}\right\|^2 = O_P(N^{-1+\gamma_2/2} + T^{-1}\ln N)$.

(vii) Recall that $\kappa_t = 1 + \left\|F_t^0\right\|^2$. By the CS inequality and (iii), $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{it}^{(\ell-1)})^2 \leq \frac{2}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{1,it}^{(\ell-1)})^2 + O_P(\delta_{NT}^{-4}(\ln N)^2)$. Using $\eta_{1,it}^{(\ell)} = F_t^{0\prime}\hat{H}^{(\ell)}\hat{\phi}_{\Lambda,i}^{(\ell)} + \lambda_i^{0\prime}(\hat{H}^{(\ell)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell)} + \lambda_i^{0\prime}(\hat{H}^{(\ell)\prime})^{-1}\hat{r}_{F,t}^{(\ell)} + F_t^{0\prime}\hat{H}^{(\ell)\prime}\hat{r}_{\Lambda,i}^{(\ell)} = \sum_{l=1}^{4}\eta_{1,it}^{(\ell)}(l)$, we have

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{1,it}^{(\ell-1)})^2 \leq 4\sum_{l=1}^{4}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t\left[\eta_{1,it}^{(\ell-1)}(l)\right]^2 \equiv 4\sum_{l=1}^{4}II_{2,l}.$$

Noting that $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2 = O_P\left(T^{-1}\right)$, we can readily show $II_{2,1} \leq \left\|\hat{H}^{(\ell-1)}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|F_t^0\right\|^4\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2 = O_P\left(T^{-1}\right)$. By Lemma A.5(i)

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\|^2 &\leq \frac{2}{T}\sum_{t=1}^{T}\kappa_t\left\|D^{-1}Q\beta_{F,t} + (1-q)\,\hat{\phi}_{F,t}^{(\ell-2)}\right\|^2 + O_P\left(T^{\gamma_1/2}\delta_{NT}^{-4}(\ln T)^2 + T^{-2+3\gamma_1/2}\right) \\
&\leq O_P(1)\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\beta_{F,t}\right\|^2 + O_P(1)\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\hat{\phi}_{F,t}^{(\ell-2)}\right\|^2 + O_P\left(\delta_{NT}^{-2}\right) \\
&= O_P\left(\delta_{NT}^{-2}\right),
\end{aligned}$$

we have $II_{2,2} \leq \left\|(\hat{H}^{(\ell-1)\prime})^{-1}\right\|^2\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^{0\prime}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. Similarly, we have

$$II_{2,3} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t\left|\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}\right|^2 \leq O_P(1)\max_t\left\|\hat{r}_{F,t}^{(\ell-1)}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\kappa_t\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^{0\prime}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right),$$

and

$$II_{2,4} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t\left|F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\right|^2 \leq O_P(1)\max_i\left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|F_t^0\right\|^2 = O_P\left(\delta_{NT}^{-2}\right).$$

It follows that $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{it}^{(\ell-1)})^2 = O_P\left(\delta_{NT}^{-2}\right).$

(viii) Note that $\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is} = \sum_{l=1}^{2}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\eta_{l,is}^{(\ell-1)}\bar{g}_{is} \equiv \sum_{l=1}^{2}II_{3,l}$. We only show $II_{3,1} = O_P\left(\delta_{NT}^{-2}\right)$ as the other term is of smaller order. Note that

$$\begin{aligned} II_{3,1} &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i^{0\prime}[F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)} \\ &\quad + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)} + F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}]\bar{g}_{it} \\ &\equiv II_{3,1}(1) + II_{3,1}(2) + II_{3,1}(3) + II_{3,1}(4). \end{aligned}$$

Let λ_{il}^0 and F_{sl}^0 denote the lth element of λ_i^0 and F_s^0, respectively. Let $II_{3,1lr}(\cdot)$ denote the (l,r)th element of $B_{3,1}(\cdot)$. Noting that $\bar{g}_{is} = (1-q) + (q - g_{is})$, we have

$$\begin{aligned} & \left\|II_{3,1lr}(1)\right\| \\ = & \left\|\frac{1}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(\ell-1)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\bar{g}_{it}\lambda_{il}^0\right\| \\ \leq & \left\|\frac{1-q}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(\ell-1)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\lambda_{il}^0\right\| + \left\|\frac{1}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(\ell-1)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}(g_{it}-q)\lambda_{il}^0\right\| \\ \equiv & II_{3,1lr}(1,1) + II_{3,1lr}(1,2). \end{aligned}$$

For $II_{3,1lr}(1,1)$, we have

$$\begin{aligned} II_{3,1lr}(1,1) &\leq O_P(1)\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\lambda_{il}^0\right\| = O_P(1)\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{H}^{(\ell-1)\prime}[\beta_{\Lambda,i} + (1-q)\hat{\phi}_{\Lambda,i}^{(\ell-2)}]\lambda_{il}^0\right\| \\ &\leq O_P(1)\left\{\frac{1}{N}\sum_{i=1}^{N}\beta_{\Lambda,i}\lambda_{il}^0 + (1-q)\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-2)}\lambda_{il}^0\right\} = O_P(\delta_{NT}^{-2}). \end{aligned}$$

For $II_{3,1lr}(1,2)$, we have

$$\begin{aligned} II_{3,1lr}(1,2) &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{il}^0\hat{\phi}_{\Lambda,i}^{(\ell-1)\prime}\hat{H}^{(\ell-1)\prime}\left[\frac{1}{T}\sum_{t=1}^{T}F_t^0F_{tr}^0(g_{it}-q)\right]\right\| \\ &\leq \left\|\hat{H}^{(\ell-1)}\right\|\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0F_{tr}^0(g_{it}-q)\right\|^2\right\}^{1/2} \\ &= O_P\left(\delta_{NT}^{-1}\right)O_P(T^{-1/2}) \end{aligned}$$

as we can show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ and $\frac{1}{N}\sum_{i=1}^{N}E\left\|\frac{1}{T}\sum_{s=1}^{T}F_s^0F_{sr}^0(g_{is}-q)\right\|^2 =$

$O\left(T^{-1}\right)$. Then $II_{3,1}(1) = O_P\left(\delta_{NT}^{-2}\right)$. Similarly,

$$\begin{aligned}
& \|II_{3,1rl}(2)\| \\
= & \left\| \frac{1}{NT} \sum_{s=1}^{T} \sum_{i=1}^{N} \lambda_{ir}^{0} \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{\phi}_{F,s}^{(\ell-1)} F_{sl}^{0} \bar{g}_{is} \right\| \\
\leq & \left\| \frac{1-q}{N} \sum_{i=1}^{N} \lambda_{ir}^{0} \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \frac{1}{T} \sum_{s=1}^{T} \hat{\phi}_{F,s}^{(\ell-1)} F_{sl}^{0} \right\| + \left\| \frac{1}{NT} \sum_{s=1}^{T} \sum_{i=1}^{N} \lambda_{ir}^{0} \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{\phi}_{F,s}^{(\ell-1)} F_{sl}^{0} (g_{is} - q) \right\| \\
\equiv & II_{3,1rl}(2,1) + II_{3,1rl}(2,2).
\end{aligned}$$

For $II_{3,1rl}(2,1)$, we have $II_{3,1rl}(2,1) \leq O_P(1) \left\| \frac{1}{T} \sum_{s=1}^{T} \hat{\phi}_{F,s}^{(\ell-1)} F_{sl}^{0} \right\| = O_P\left(\delta_{NT}^{-2}\right)$ as we can show that $\left\| \frac{1}{T} \sum_{s=1}^{T} \hat{\phi}_{F,s}^{(\ell-1)} F_{sl}^{0} \right\| = O_P\left(\delta_{NT}^{-2}\right)$ by following the analysis of $\left\| \frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_{\Lambda,i}^{(\ell-1)} \lambda_{il}^{0} \right\|$. For $II_{3,1rl}(2,2)$, we have

$$\begin{aligned}
II_{3,1rl}(2,2) &= \left\| \frac{1}{T} \sum_{s=1}^{T} F_{sl}^{0} \hat{\phi}_{F,s}^{(\ell-1)\prime} (\hat{H}^{(\ell-1)})^{-1} \frac{1}{N} \sum_{i=1}^{N} \lambda_i^{0} \lambda_{ir}^{0} (g_{is} - q) \right\| \\
&\leq \left\| (\hat{H}^{(\ell-1)})^{-1} \right\| \left\{ \frac{1}{T} \sum_{s=1}^{T} \left\| F_s^{0} \right\|^2 \left\| \hat{\phi}_{F,s}^{(\ell-1)} \right\|^2 \right\}^{1/2} \left\{ \frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{N} \sum_{i=1}^{N} \lambda_i^{0} \lambda_{ir}^{0} (g_{is} - q) \right\|^2 \right\}^{1/2} \\
&= O_P\left(\delta_{NT}^{-1}\right) O_P(N^{-1/2}).
\end{aligned}$$

So $II_{3,1}(2) = O_P\left(\delta_{NT}^{-2}\right)$. Analogously, we can show that $II_{3,1}(l) = O_P\left(\delta_{NT}^{-2}\right)$ for $l = 3, 4$. Then $II_{3,1} = O_P\left(\delta_{NT}^{-2}\right)$.

(ix) By (vi) and the fact that $\frac{1}{N} \sum_{i=1}^{N} E\left(\frac{1}{T} \sum_{s=1}^{T} F_s^{0} \varepsilon_{is} g_{is}\right)^2 = O\left(T^{-1}\right)$,

$$\begin{aligned}
\max_t \left\| \frac{1}{NT} \sum_{s=1}^{T} F_s^{0} \sum_{i=1}^{N} \eta_{it}^{(\ell-1)} \bar{g}_{it} \varepsilon_{is} g_{is} \right\| &= \left\| \frac{1}{N} \sum_{i=1}^{N} \eta_{it}^{(\ell-1)} \bar{g}_{it} \left(\frac{1}{T} \sum_{s=1}^{T} F_s^{0} \varepsilon_{is} g_{is} \right) \right\| \\
&\leq \left\{ \max_t \frac{1}{N} \sum_{i=1}^{N} (\eta_{it}^{(\ell-1)})^2 \right\}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{s=1}^{T} F_s^{0} \varepsilon_{is} g_{is} \right)^2 \right\}^{1/2} \\
&= O_P\left((T^{-1/2+\gamma_1/4} + (N/\ln N)^{-1/2} \right) O_P\left(T^{-1/2} \right) \\
&= O_P\left((T^{-1+\gamma_1/4} + (NT/\ln N)^{-1/2} \right).
\end{aligned}$$

(x) Note that $\frac{1}{NT} \sum_{s=1}^{T} F_s^{0} \sum_{i=1}^{N} \varepsilon_{it} g_{it} \eta_{is}^{(\ell-1)} \bar{g}_{is} \leq \sum_{l=1}^{2} \frac{1}{NT} \sum_{s=1}^{T} F_s^{0} \sum_{i=1}^{N} \varepsilon_{it} g_{it} \eta_{l,is}^{(\ell-1)} \bar{g}_{is} \equiv \sum_{l=1}^{2} II_{4,lt}$. One can bound $II_{4,2t}$ by $O_P\left(\delta_{NT}^{-2} \ln N\right)$ by using the uniform bound for $\eta_{2,is}^{(\ell-1)}$ in (iii). For $II_{4,1t}$, we have

$$\begin{aligned}
II_{4,1t} &= \frac{1}{NT} \sum_{s=1}^{T} F_s^{0} \sum_{i=1}^{N} \varepsilon_{it} g_{it} [F_t^{0\prime} \hat{H}^{(\ell-1)} \hat{\phi}_{\Lambda,i}^{(\ell-1)} + \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{\phi}_{F,t}^{(\ell-1)} \\
&\quad + \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{r}_{F,t}^{(\ell-1)} + F_t^{0\prime} \hat{H}^{(\ell-1)\prime} \hat{r}_{\Lambda,i}^{(\ell-1)}] \bar{g}_{is} \\
&\equiv II_{4,1t}(1) + II_{4,1t}(2) + II_{4,1t}(3) + II_{4,1t}(4).
\end{aligned}$$

For $II_{4,1t}(1)$, by (i), (iv) and the fact $\frac{1}{N}\sum_{i=1}^{N} E\left\|\frac{1}{T}\sum_{s=1}^{T} F_s^0 F_s^{0\prime}(g_{is}-q)\right\|^2 = O\left(T^{-1}\right)$, we have

$$
\begin{aligned}
\max_t II_{4,1t}(1) &\leq \max_t \left\| \frac{1-q}{T}\sum_{s=1}^{T} F_s^0 F_s^{0\prime} \hat{H}^{(\ell-1)} \frac{1}{N}\sum_{i=1}^{N} \hat{\phi}_{\Lambda,i}^{(\ell-1)} \varepsilon_{it} g_{it} \right\| \\
&\quad + \max_t \left\| \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{it} g_{it} \hat{\phi}_{\Lambda,i}^{(\ell-1)\prime} \hat{H}^{(\ell-1)\prime} \frac{1}{T}\sum_{s=1}^{T} F_s^0 F_s^{0\prime}(g_{is}-q) \right\| \\
&\leq O_P(1) \max_t \left\| \frac{1}{N}\sum_{i=1}^{N} \hat{\phi}_{\Lambda,i}^{(\ell-1)} \varepsilon_{it} g_{it} \right\| \\
&\quad + \max_i \left\| \hat{\phi}_{\Lambda,i}^{(\ell-1)} \right\| \left\{ \max_t \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{it}^2 \right\}^{1/2} \left\{ \frac{1}{N}\sum_{i=1}^{N} \left\| \frac{1}{T}\sum_{s=1}^{T} F_s^0 F_s^{0\prime}(g_{is}-q) \right\|^2 \right\}^{1/2} \\
&= O_P(T^{-1+\gamma_1/2} + \delta_{NT}^{-2}\ln N) + O_P((T/\ln T)^{-1/2}) O_P(T^{-1/2}) = O_P(T^{-1+\gamma_1/2} + \delta_{NT}^{-2}\ln N).
\end{aligned}
$$

For $II_{3,1t}(2)$, we have

$$
\begin{aligned}
\max_t II_{4,1t}(2) &= \left\| \frac{1}{T}\sum_{s=1}^{T} F_s^0 \hat{\phi}_{F,s}^{(\ell-1)\prime} [\hat{H}^{(\ell-1)}]^{-1} \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it} \bar{g}_{is} \right\| \\
&\leq \left\| \frac{1-q}{T}\sum_{s=1}^{T} F_s^0 \hat{\phi}_{F,s}^{(\ell-1)\prime} [\hat{H}^{(\ell-1)}]^{-1} \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it} \right\| \\
&\quad + \left\| \frac{1}{T}\sum_{s=1}^{T} F_s^0 \hat{\phi}_{F,s}^{(\ell-1)\prime} [\hat{H}^{(\ell-1)}]^{-1} \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it}(g_{is}-q) \right\| \\
&\leq O_P(1) \max_s \left\| \hat{\phi}_{F,s}^{(\ell-1)} \right\| \max_s \left\| \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it} \bar{g}_{is} \right\| \\
&\quad + O_P(1) \left\{ \frac{1}{T}\sum_{s=1}^{T} \left\| F_s^0 \hat{\phi}_{F,s}^{(\ell-1)\prime} \right\|^2 \right\}^{1/2} \left\{ \frac{1}{T}\sum_{s=1}^{T} \left\| \frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} g_{it}(g_{is}-q) \right\|^2 \right\}^{1/2} \\
&= O_P((N/\ln N)^{-1/2}) O_P((N/\ln N)^{-1/2}) + O_P((N/\ln N)^{-1/2}) O_P(\delta_{NT}^{-1}) = O_P(\delta_{NT}^{-2}\ln N).
\end{aligned}
$$

Analogously, we can show that $\max_t \|II_{4,1t}(l)\| = O_P(\delta_{NT}^{-2}\ln N)$ for $l = 3, 4$. Then $\max_t \|II_{3,1t}\| = O_P\left(T^{-1+\gamma_1/2} + \delta_{NT}^{-2}\ln N\right)$. ■

Proof of Lemma A.6. (i) $\max_i \frac{1}{T}\sum_{t=1}^{T} |\hat{\varepsilon}_{it} - \varepsilon_{it}|^2 = O_P\left(\underline{m}^{-1}\ln T\right)$. Noting that $\hat{\varepsilon}_{it} - \varepsilon_{it} = \hat{\lambda}_i^{\prime}\hat{F}_t^{(0)} - \lambda_i^{0\prime} F_t^{(0)} = (\hat{\lambda}_i^{(0)} - \hat{H}^{-1}\lambda_i^0)^{\prime}\hat{F}_t + \lambda_i^{0\prime}\{[\hat{H}^{\prime}]^{-1}\hat{F}_t - F_t^0\}$, we have

$$
\begin{aligned}
\max_i \frac{1}{T}\sum_{t=1}^{T} |\hat{\varepsilon}_{it} - \varepsilon_{it}|^2 &= \max_i \frac{1}{T}\sum_{t=1}^{T} \left| (\hat{\lambda}_i - \hat{H}^{-1}\lambda_i^0)^{\prime}\hat{F}_t + \lambda_i^{0\prime}\{(\hat{H}^{\prime})^{-1}\hat{F}_t - F_t^0\} \right|^2 \\
&\leq 2R \max_i \left\| \hat{\lambda}_i - \hat{H}^{-1}\lambda_i^0 \right\|^2 + 2\max_i \|\lambda_i^0\|^2 \frac{1}{T}\sum_{t=1}^{T} \left\| \hat{F}_t - \hat{H}^{\prime} F_t^0 \right\|^2 \left\| (\hat{H}^{\prime})^{-1} \right\|^2 \\
&= O_P(T^{-1}\ln T) + O_P(N^{\gamma_2/2}) O_P(N^{-1}) = O_P(N^{-1+\gamma_2/2} + T^{-1}\ln T).
\end{aligned}
$$

(ii) Note that

$$\begin{aligned}
\max_{i,t}|\hat{\varepsilon}_{it}-\varepsilon_{it}| &\leq \max_{i,t}\left|(\hat{\lambda}_i-\hat{H}^{-1}\lambda_i^0)'\hat{F}_t\right|+\max_{i,t}\left|\lambda_i^{0\prime}\{(\hat{H}')^{-1}\hat{F}_t-F_t^0\}\right| \\
&\leq \max_i\left\|\hat{\lambda}_i-\hat{H}^{-1}\lambda_i^0\right\|\max_t\left\|\hat{F}_t\right\|+\max_i\|\lambda_i^0\|\max_t\left\|\hat{F}_t-\hat{H}'F_t^0\right\|\left\|(\hat{H}')^{-1}\right\| \\
&= O_P(T^{-1/2}(\ln T)^{1/2})O_P(T^{\gamma_1/4})+O_P(N^{\gamma_2/4})O_P((N/\ln T)^{1/2}) \\
&= O_P\left((T^{-1/2+\gamma_1/4}+N^{-1/2+\gamma_2/4})(\ln T)^{1/2}\right)=o_P(1),
\end{aligned}$$

where we use the fact that $\max_t\left\|\hat{F}_t\right\|\leq\max_t\left\|\hat{F}_t-\hat{H}'F_t^{(0)}\right\|+\max_t\left\|\hat{H}'F_t^{(0)}\right\|=O_P(T^{\gamma_1/4})$.

(iii) This follows from (i) and (ii) and Theorem 5 in Fan, Liao, and Mincheva (2013). ■

D Some Additional Simulation Results

In this appendix we report some additional simulation results that are associated with the case $q=0.9$, i.e., only 10% observations are missing at random. Tables A1–A3 correspond to Tables 2–4 in the main text.

The results in Table A1 are comparable with those in Table 2. When the proportion of missing observations is smaller ($q=0.9$ here), the three CV methods perform slightly better than the case with a larger proportion of missing observations. In addition, they continue to outperform both M-1 and M-2 of existing methods for most cases. Among the other methods, only the M-1 of ED shows a pattern of convergence in all cases.

The results in Table A2 are comparable with those in Table 3. As expected, the MSE decreases and the R^2 increases as N or T increases; the MSEs in the case of $q=0.9$ are smaller than those for $q=0.7$; the R^2 are slightly larger in the case of $q=0.9$ than in the case of $q=0.7$. Similarly, the results in Table A3 are analogous to those in Table 4.

Table A1: Under/Over-estimation frequency with missing data (q=0.9)

			Cross-validation			ED		GR		ER		PC		IC	
DGPN		T	$CV^{(0)}$	$CV_R^{(l^*)}$	$CV_{R_{\max}}^{(l^*)}$	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2	M-1	M-2
1	50	50	189/42	127/47	124/39	20/343	0/981	849/14	681/221	957/9	923/53	0/941	0/1000	0/742	0/982
	50	100	27/51	15/50	14/45	2/244	0/970	779/16	566/299	959/6	898/69	0/775	0/989	0/578	0/953
	100	50	44/43	21/46	19/36	2/257	0/979	738/15	552/314	945/4	907/60	0/793	0/993	0/613	0/967
	100	100	4/32	2/32	1/25	0/247	0/966	593/34	497/323	899/12	874/76	0/676	0/967	0/574	0/942
2	50	50	161/1	106/1	103/2	17/145	0/934	834/1	682/56	962/1	930/2	0/621	0/991	0/253	0/824
	50	100	16/1	5/2	3/2	0/77	0/834	748/0	522/49	949/0	886/1	0/271	0/866	0/116	0/647
	100	50	37/2	16/0	16/1	3/117	0/869	734/1	576/36	939/0	870/4	0/323	0/888	0/125	0/726
	100	100	0/0	0/0	0/0	0/52	0/690	516/1	306/29	865/0	722/1	0/139	0/647	0/73	0/523
3	50	50	62/0	32/0	28/0	11/46	0/870	826/0	628/0	965/0	919/0	0/1000	0/1000	0/964	0/999
	50	100	2/0	0/0	0/0	0/0	0/438	638/0	260/0	934/0	827/0	0/813	0/999	0/160	0/719
	100	50	2/0	0/0	1/0	0/11	0/656	716/0	443/0	934/0	888/0	0/1000	0/1000	0/990	0/1000
	100	100	0/0	0/0	0/0	0/1	0/146	344/0	52/0	854/0	650/0	0/818	0/999	0/228	0/845
4	50	50	104/0	55/0	50/0	26/39	0/863	866/0	709/0	973/0	950/0	0/888	0/994	0/461	0/818
	50	100	7/0	1/4	1/5	2/31	0/715	801/0	574/0	967/0	920/0	0/794	0/949	0/499	0/758
	100	50	12/0	4/1	5/1	0/11	0/546	743/0	442/0	946/0	885/0	0/491	0/884	0/96	0/413
	100	100	0/0	0/0	0/0	0/0	0/480	535/0	205/0	913/0	805/0	0/588	0/883	0/204	0/613
5	50	50	72/0	31/0	28/0	11/35	0/886	835/0	641/0	966/0	934/0	0/1000	0/1000	0/955	0/1000
	50	100	2/0	1/0	0/1	1/4	0/658	720/0	451/0	961/0	910/0	0/1000	0/1000	0/990	0/1000
	100	50	3/0	2/0	2/0	0/13	0/462	661/0	266/0	918/0	818/0	0/813	0/998	0/164	0/737
	100	100	0/0	0/0	0/0	0/0	0/144	356/0	44/0	852/0	657/0	0/806	0/999	0/184	0/851
6	50	50	85/0	36/0	34/0	4/27	0/624	751/0	345/0	930/0	841/0	0/478	0/982	0/14	0/330
	50	100	4/0	1/0	1/0	0/9	0/180	575/0	104/0	912/0	684/0	0/5	0/312	0/0	0/6
	100	50	8/0	0/0	0/0	0/17	0/199	568/0	116/0	896/0	682/0	0/6	0/305	0/0	0/6
	100	100	0/0	0/0	0/0	0/2	0/16	232/0	3/0	760/0	336/0	0/0	0/0	0/0	0/0

Table A2: MSE and $R^2(\hat{F})$ with missing observations (q=0.9)													
			MSE					$R^2(\hat{F})$					
			oracle	iterated estimate				oracle	iterated estimate				
DGP	N	T		ℓ=0	ℓ=5	ℓ=20	$\ell=\infty$		ℓ=0	ℓ=5	ℓ=20	$\ell=\infty$
1	50	50	0.460	0.794	0.524	0.564	0.604	0.964	0.941	0.958	0.958	0.957
	50	100	0.367	0.590	0.396	0.414	0.425	0.967	0.946	0.962	0.962	0.962
	100	50	0.423	0.658	0.446	0.472	0.499	0.978	0.967	0.976	0.976	0.976
	100	100	0.221	0.374	0.248	0.258	0.264	0.982	0.971	0.980	0.980	0.980
2	50	50	0.352	0.654	0.403	0.407	0.408	0.971	0.949	0.967	0.967	0.967
	50	100	0.259	0.467	0.292	0.295	0.295	0.972	0.953	0.969	0.969	0.969
	100	50	0.258	0.473	0.291	0.293	0.293	0.986	0.976	0.984	0.984	0.984
	100	100	0.172	0.307	0.192	0.193	0.193	0.986	0.977	0.985	0.985	0.985
3	50	50	0.403	0.658	0.436	0.438	0.438	0.975	0.957	0.972	0.972	0.972
	50	100	0.266	0.453	0.291	0.293	0.293	0.976	0.959	0.973	0.973	0.973
	100	50	0.328	0.522	0.352	0.353	0.353	0.987	0.978	0.986	0.986	0.986
	100	100	0.198	0.323	0.214	0.215	0.215	0.988	0.979	0.987	0.987	0.987
4	50	50	0.350	0.621	0.394	0.397	0.396	0.970	0.951	0.966	0.966	0.966
	50	100	0.261	0.455	0.292	0.294	0.294	0.970	0.953	0.967	0.967	0.967
	100	50	0.262	0.463	0.294	0.295	0.295	0.985	0.975	0.983	0.983	0.983
	100	100	0.173	0.304	0.194	0.195	0.195	0.985	0.976	0.984	0.984	0.984
5	50	50	0.386	0.645	0.420	0.423	0.423	0.970	0.951	0.966	0.967	0.967
	50	100	0.316	0.501	0.339	0.341	0.341	0.970	0.952	0.967	0.967	0.967
	100	50	0.260	0.454	0.286	0.287	0.287	0.985	0.976	0.984	0.984	0.984
	100	100	0.190	0.314	0.206	0.207	0.207	0.985	0.977	0.984	0.984	0.984
6	50	50	0.322	0.580	0.358	0.360	0.360	0.976	0.958	0.972	0.973	0.973
	50	100	0.239	0.428	0.265	0.266	0.266	0.976	0.958	0.973	0.973	0.973
	100	50	0.244	0.438	0.270	0.271	0.271	0.988	0.979	0.986	0.987	0.987
	100	100	0.161	0.285	0.177	0.177	0.177	0.988	0.979	0.987	0.987	0.987

Table A3: Coverage probability and average length of the 95% confidence intervals (q=0.9)															
			Oracle				ℓ=0				$\ell=\ell^*$				
			standard		robust		standard		robust		standard		robust		
DGP	N	T	CP	Length	CP	Length	CP	Length	CP	Length	CP	Length	CP	Length	
1	50	50	0.926	0.514	0.947	0.551	0.943	0.626	0.943	0.673	0.908	0.516	0.945	0.578	
	50	100	0.919	0.529	0.930	0.562	0.936	0.629	0.942	0.680	0.920	0.537	0.930	0.588	
	100	50	0.926	0.365	0.940	0.400	0.936	0.459	0.947	0.488	0.940	0.394	0.942	0.421	
	100	100	0.940	0.374	0.943	0.403	0.946	0.478	0.944	0.490	0.952	0.391	0.945	0.424	
2	50	50	0.918	0.537	0.932	0.550	0.927	0.652	0.940	0.671	0.913	0.557	0.942	0.577	
	50	100	0.922	0.538	0.924	0.557	0.920	0.660	0.942	0.682	0.921	0.564	0.939	0.585	
	100	50	0.943	0.388	0.946	0.395	0.943	0.468	0.952	0.481	0.943	0.409	0.950	0.416	
	100	100	0.938	0.390	0.936	0.401	0.932	0.479	0.951	0.490	0.930	0.411	0.936	0.422	
3	50	50	0.926	0.550	0.936	0.557	0.920	0.671	0.937	0.678	0.914	0.582	0.935	0.585	
	50	100	0.932	0.565	0.938	0.567	0.942	0.677	0.947	0.684	0.948	0.589	0.950	0.596	
	100	50	0.930	0.400	0.937	0.398	0.942	0.490	0.946	0.488	0.928	0.416	0.932	0.419	
	100	100	0.925	0.403	0.933	0.404	0.939	0.487	0.943	0.489	0.943	0.422	0.947	0.425	
4	50	50	0.917	0.601	0.937	0.607	0.907	0.718	0.926	0.726	0.906	0.630	0.929	0.638	
	50	100	0.928	0.607	0.943	0.614	0.931	0.716	0.939	0.725	0.919	0.636	0.926	0.645	
	100	50	0.927	0.440	0.928	0.436	0.935	0.524	0.938	0.522	0.933	0.461	0.946	0.460	
	100	100	0.932	0.445	0.943	0.447	0.930	0.527	0.938	0.531	0.934	0.464	0.942	0.471	
5	50	50	0.891	0.322	0.908	0.327	0.920	0.475	0.932	0.481	0.869	0.340	0.887	0.344	
	50	100	0.896	0.323	0.901	0.328	0.916	0.475	0.926	0.480	0.910	0.340	0.923	0.346	
	100	50	0.885	0.233	0.885	0.233	0.923	0.342	0.926	0.346	0.905	0.246	0.897	0.245	
	100	100	0.904	0.234	0.905	0.236	0.922	0.348	0.933	0.350	0.893	0.247	0.903	0.249	
6	50	50	0.897	0.320	0.911	0.325	0.923	0.473	0.929	0.478	0.882	0.339	0.898	0.342	
	50	100	0.875	0.325	0.896	0.330	0.927	0.475	0.930	0.477	0.894	0.343	0.908	0.347	
	100	50	0.913	0.233	0.917	0.233	0.925	0.335	0.931	0.339	0.907	0.245	0.916	0.245	
	100	100	0.908	0.236	0.913	0.236	0.929	0.343	0.932	0.346	0.918	0.247	0.923	0.249	

REFERENCES

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203-1227.

Allez, R., Bouchaud, J.-P., 2013. Eigenvector dynamics: general theory and some applications. Physical Review E, 86(4):046202.

Bai, J., 2003. Inferential theory for factors models of large dimensions. Econometrica 71, 135-173.

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

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B 75, 603-680.

Negahban, S. and Wainwright, M.J., 2012. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. Journal of Machine Learning Research, 13, 1665-1697.

Stewart, G. W., Sun, J.-G., 1990. Matrix Perturbation Theory. Academic Press, Boston.

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