Proofs for “Efficient Estimation of Approximate Factor Models via Regularized Maximum Likelihood”

Jushan Bai
Columbia University
Department of Economics

Yuan Liao
University of Maryland
Department of Mathematics

September 25, 2012

Abstract

This document contains the proofs of all the results in the main paper.

Keywords: High dimensionality, unknown factors, principal components, sparse matrix, conditional sparsity, thresholding, cross-sectional correlation, penalized maximum likelihood, SCAD, adaptive lasso, heteroskedasticity

A Proofs for generic estimators

We need to establish the results for two sets of estimators: the two-step estimator and the joint estimator, whose proofs for consistency share some similarities. Therefore in this section we establish some preliminary results for generic estimators that can be used for both cases. We denote by $(\widehat{\Lambda}, \widehat{\Sigma}_u)$ as a generic estimator for (Λ_0, Σ_{u0}), which can be either $(\widehat{\Lambda}^{(1)}, \widehat{\Sigma}_u^{(1)})$ or $(\widehat{\Lambda}^{(2)}, \widehat{\Sigma}_u^{(2)})$. Define

$$Q_2(\Lambda, \Sigma_u) = \frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda_0 - \Lambda_0'\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}\Lambda_0), \tag{A.1}$$

$$Q_3(\Lambda, \Sigma_u) = \frac{1}{N}\log|\Lambda\Lambda'+\Sigma_u| + \frac{1}{N}\mathrm{tr}(S_y(\Lambda\Lambda'+\Sigma_u)^{-1}) - \frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1}) - \frac{1}{N}\log|\Sigma_u| - Q_2(\Lambda, \Sigma_u). \tag{A.2}$$

Define the set

$$\Xi_\delta = \{(\Lambda, \Sigma_u): \quad \delta^{-1} < \lambda_{\min}(N^{-1}\Lambda'\Lambda) \leq \lambda_{\max}(N^{-1}\Lambda'\Lambda) < \delta, \quad \delta^{-1} < \lambda_{\min}(\Sigma_u) \leq \lambda_{\max}(\Sigma_u) < \delta\}$$

We first present a lemma that will be needed throughout the proof.

Lemma A.1. *(i)* $\max_{i,j\leq r}|\frac{1}{T}\sum_{t=1}^T f_{it}f_{jt}-Ef_{it}f_{jt}|=O_p(\sqrt{1/T})$.
(ii) $\max_{i,j\leq N}|\frac{1}{T}\sum_{t=1}^T u_{it}u_{jt}-Eu_{it}u_{jt}|=O_p(\sqrt{(\log N)/T})$.
(iii) $\max_{i\leq r,j\leq N}|\frac{1}{T}\sum_{t=1}^T f_{it}u_{jt}|=O_p(\sqrt{(\log N)/T})$.

Proof. See Lemmas A.3 and B.1 in Fan, Liao and Mincheva (2011). □

Lemma A.2. *Under Assumption 3.2, for any* $\delta>0$,

$$\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}|Q_3(\Lambda,\Sigma_u)|=O\left(\frac{\log N}{N}+\sqrt{\frac{\log N}{T}}\right).$$

Therefore we can write

$$\begin{aligned}&\frac{1}{N}\log|\Lambda\Lambda'+\Sigma_u|+\frac{1}{N}\mathrm{tr}(S_y(\Lambda\Lambda'+\Sigma_u)^{-1})\\&=\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1})+\frac{1}{N}\log|\Sigma_u|+Q_2(\Lambda,\Sigma_u)+O\left(\frac{\log N}{N}+\sqrt{\frac{\log N}{T}}\right).\end{aligned}\tag{A.3}$$

Proof. First of all, note that $|\Lambda\Lambda'+\Sigma_u|=|\Sigma_u|\times|I_r+\Lambda'\Sigma_u^{-1}\Lambda|$, and $\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}\frac{1}{N}\log|I_r+\Lambda'\Sigma_u^{-1}\Lambda|=O\left(\frac{\log N}{N}\right)$, hence we have

$$\frac{1}{N}\log|\Lambda\Lambda'+\Sigma_u|=\frac{1}{N}\log|\Sigma_u|+O\left(\frac{\log N}{N}\right),\tag{A.4}$$

where $O(\cdot)$ is uniform in Ξ_δ. Equation (A.4) will be used later in the proof.

We now consider the term $N^{-1}\mathrm{tr}(S_y(\Lambda\Lambda'+\Sigma_u)^{-1})$. With the identification condition $\frac{1}{T}\sum_{t=1}^T f_tf_t'=I_r$, $\bar f=0$, and $S_u=\frac{1}{T}\sum_{t=1}^T u_tu_t'$,

$$S_y=\frac{1}{T}\sum_{t=1}^T(y_t-\bar y)(y_t-\bar y)'=\Lambda_0\Lambda_0'+S_u+\Lambda_0\frac{1}{T}\sum_{t=1}^T f_tu_t'+(\Lambda_0\frac{1}{T}\sum_{t=1}^T f_tu_t')'-\bar u\bar u'.$$

By the matrix inversion formula $(\Lambda\Lambda'+\Sigma_u)^{-1}=\Sigma_u^{-1}-\Sigma_u^{-1}\Lambda(I_r+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}$,

$$\frac{1}{N}\mathrm{tr}(S_y(\Lambda\Lambda'+\Sigma_u)^{-1})=\frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda_0)+\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1})-A_1+A_2+A_3-A_4-A_5,\tag{A.5}$$

where $A_1=N^{-1}\mathrm{tr}(\Lambda_0\Lambda_0'\Sigma_u^{-1}\Lambda(I_r+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1})$, $A_2=\frac{1}{N}\mathrm{tr}(\frac{1}{T}\sum_{t=1}^T\Lambda_0f_tu_t'(\Lambda\Lambda'+\Sigma_u)^{-1})$, $A_3=\frac{1}{N}\mathrm{tr}(\frac{1}{T}\sum_{t=1}^T u_tf_t'\Lambda_0'(\Lambda\Lambda'+\Sigma_u)^{-1})$, and $A_4=\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1}\Lambda(I_r+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1})$. Term $A_5=N^{-1}\mathrm{tr}(\bar u\bar u'(\Lambda\Lambda'+\Sigma_u)^{-1})=O_p((\log N)/T)$ uniformly in the parameter space, and hence can be ignored.

Let us look at terms A_1, A_2, A_3 and A_4 subsequently. Note that $\lambda_{\max}(\Sigma_u)$ and $N\lambda_{\min}^{-1}(\Lambda'\Lambda)$ are both bounded from above uniformly in Ξ_δ, we have,

$$\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}\lambda_{\max}[(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}]\leq\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}\frac{\lambda_{\max}(\Sigma_u)}{\lambda_{\min}(\Lambda'\Lambda)}=O(N^{-1}), \tag{A.6}$$

$$\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}\lambda_{\max}[(I_r+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}]\leq\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}\lambda_{\max}[(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}]=O(N^{-1}). \tag{A.7}$$

In addition, $\|\Lambda\|_F=O(\sqrt{N})$, $\lambda_{\max}(\Sigma_u^{-1})=O(1)$ uniformly in Ξ_δ, and $\|\Lambda_0\|_F=O(\sqrt{N})$. Applying the matrix inversion formula yields

$$\begin{aligned}A_1 &= \frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}\Lambda_0)-\frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}(I_r+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}\Lambda_0)\\ &= \frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}\Lambda_0)+O\left(\frac{1}{N}\right), \qquad (A.8)\end{aligned}$$

where $O(\cdot)$ is uniform over $(\Lambda,\Sigma_u)\in\Xi_\delta$. In the second equality above we applied (A.6) and (A.7) and the following inequality:

$$\begin{aligned}&\frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}(I_r+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}\Lambda_0)\\ &\leq\frac{1}{N}\|\Lambda_0'\Sigma_u^{-1}\Lambda\|_F^2\lambda_{\max}[(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}]\lambda_{\max}[(I_r+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}]\\ &\leq O(N^{-3})\|\Lambda_0\|_F^2\|\Lambda\|_F^2\lambda_{\max}(\Sigma_u^{-1})=O(N^{-1}).\end{aligned}$$

By Lemma A.1(iii), and $\lambda_{\max}((\Lambda\Lambda'+\Sigma_u)^{-1})\leq\lambda_{\max}(\Sigma_u^{-1})=O(1)$ uniformly in Ξ_δ,

$$\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}|A_2|\leq\frac{1}{N}\|\Lambda_0'(\Lambda\Lambda'+\Sigma_u)^{-1}\|_F\left\|\frac{1}{T}\sum_{t=1}^T f_tu_t'\right\|_F=O_p(\sqrt{\frac{\log N}{T}}). \tag{A.9}$$

Similarly, $\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}|A_3|=O_p(\sqrt{\frac{\log N}{T}})$. Again by the matrix inversion formula,

$$A_4=\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1})-\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}(I+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}).$$

The second term on the right hand side is of smaller order (uniformly) than the first term, because it has an additional term $(I+\Lambda'\Sigma_u^{-1}\Lambda)^{-1}$, whose maximum eigenvalue is $O(N^{-1})$ uniformly by (A.7). The first term is bounded by (uniformly in Ξ_δ):

$$\frac{c}{N}\|S_u\Sigma_u^{-1}\Lambda\|_FO(N^{-1})\|\Lambda'\Sigma_u^{-1}\|_F\leq O(N^{-1})\lambda_{\max}(S_u)=O(\sqrt{\frac{\log N}{T}}+\frac{1}{N}).$$

Hence $\sup_{(\Lambda,\Sigma_u)\in\Xi_\delta}|A_4| = O(T^{-1/2}(\log N)^{1/2}+N^{-1})$. Results (A.4) and (A.5) then yield

$$
\begin{aligned}
&\frac{1}{N}\log|\Lambda\Lambda'+\Sigma_u|+\frac{1}{N}\mathrm{tr}(S_y(\Lambda\Lambda'+\Sigma_u)^{-1})\\
&=\frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda_0)+\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1})+\frac{1}{N}\log|\Sigma_u|-\frac{1}{N}\mathrm{tr}(\Lambda_0'\Sigma_u^{-1}\Lambda(\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}\Lambda_0)\\
&+O\left(\frac{\log N}{N}+\sqrt{\frac{\log N}{T}}\right)\\
&=\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1})+\frac{1}{N}\log|\Sigma_u|+Q_2(\Lambda,\Sigma_u)+O\left(\frac{\log N}{N}+\sqrt{\frac{\log N}{T}}\right).
\end{aligned}
$$

□

Throughout the proofs, we note that the consistency depends crucially on the consistency of the following quantities:

$$J=(\widehat{\Lambda}-\Lambda_0)'\widehat{\Sigma}_u^{-1}\widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\widehat{\Lambda})^{-1}$$

We state the following lemma for the generic estimators.

Lemma A.3. *(i)* $\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0-(I_r-J)\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\widehat{\Lambda}(I_r-J)'=o_p(N)$
(ii) First order condition: $\widehat{\Lambda}'(\widehat{\Lambda}\widehat{\Lambda}'+\widehat{\Sigma}_u)^{-1}(S_y-\widehat{\Lambda}\widehat{\Lambda}'-\widehat{\Sigma}_u)=0$.

We will prove Lemma A.3 for both $(\widehat{\Lambda}^{(1)},\widehat{\Sigma}_u^{(1)})$ and $(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})$ later when we deal with these two estimators individually.

Lemma A.4. *Suppose Lemma A.3 holds, then*
(i) $\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}(S_y-\widehat{\Lambda}\widehat{\Lambda}'-\widehat{\Sigma}_u)=0$.
(ii) $(J-I_r)'(J-I_r)-I_r=O_p(N^{-1}+T^{-1/2}(\log N)^{1/2})$.

Proof. (i) Using the matrix inverse formula, the same argument of Bai and Li (2012)'s (A.2) implies $\widehat{\Lambda}'(\widehat{\Lambda}\widehat{\Lambda}'+\widehat{\Sigma}_u)^{-1}=(I_r+\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\widehat{\Lambda})^{-1}\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}$. Thus part (i) follows from the first order condition in Lemma A.3.

(ii) Let $H=(\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\widehat{\Lambda})^{-1}$. Part (i) can be equivalently written as $J+J'-J'J+K=0$ where

$$K=J'\frac{1}{T}\sum_{t=1}^T f_tu_t'\widehat{\Sigma}_u^{-1}\widehat{\Lambda}H+H\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\frac{1}{T}\sum_{t=1}^T u_tf_t'J-\frac{1}{T}\sum_{t=1}^T f_tu_t'\widehat{\Sigma}_u^{-1}\widehat{\Lambda}H-H\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\frac{1}{T}\sum_{t=1}^T u_tf_t'$$

$$-H\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}(S_u-\widehat{\Sigma}_u)\widehat{\Sigma}_u^{-1}\widehat{\Lambda}H.$$

Note that for $(\widehat{\Lambda}, \widehat{\Sigma}_u) \in \Xi_\delta$, $H = O_p(N^{-1})$, $J = O_p(1)$ for each element, $\|\widehat{\Sigma}_u^{-1}\| = O_p(1)$, $\|\widehat{\Lambda}\|_F = O_p(\sqrt{N})$, hence

$$\|\frac{1}{T}\sum_{t=1}^T f_t u_t' \widehat{\Sigma}_u^{-1}\widehat{\Lambda}H\|_F \leq O_p(1)\|\frac{1}{NT}\sum_{t=1}^T f_t u_t'\|_F \|\widehat{\Sigma}_u^{-1}\|\|\widehat{\Lambda}\|_F = O_p(\frac{1}{N}\sqrt{\frac{N\log N}{T}}\sqrt{N}) = O_p(\sqrt{\frac{\log N}{T}})$$

Moreover, for the empirical covariance $\|S_u\|^2 \leq 2\sum_{i,j\leq N}(T^{-1}\sum_{t=1}^T u_{it}u_{jt} - \sigma_{u0,ij})^2 + 2\|\Sigma_{u0}\|^2 = O_p(T^{-1}N^2\log N + 1)$ by Lemma A.1, which implies $H\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}S_u\widehat{\Sigma}_u^{-1}\widehat{\Lambda}H = O_p(N^{-1} + T^{-1/2}(\log N)^{1/2})$. Also, $H\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\widehat{\Sigma}_u\widehat{\Sigma}_u^{-1}\widehat{\Lambda}H = H = O_p(N^{-1})$. Therefore $K = O_p(N^{-1} + T^{-1/2}(\log N)^{1/2})$. It then implies (ii).

□

Lemma A.5. *Suppose Lemma A.3 holds, then* $J = o_p(1)$.

Proof. By our assumption, both $\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\widehat{\Lambda}$ and $\Lambda_0'\Sigma_{u0}^{-1}\Lambda$ are diagonal. Moreover, the eigenvalues of $N^{-1}\widehat{\Lambda}'\widehat{\Sigma}_u^{-1}\widehat{\Lambda}$ and $N^{-1}\Lambda_0'\Sigma_{u0}^{-1}\Lambda$ are bounded away from zero. Therefore by Lemma A.3(i) and Lemma A.4(ii), there are two diagonal matrices M_1 and M_2 whose eigenvalues are all bounded away from zero, such that

$$(I_r - J)M_1(I_r - J)' = M_2 + o_p(1), \quad (J - I_r)'(J - I_r) = I_r + o_p(1) \tag{A.10}$$

Applying Lemma A.1 of Bai and Li (2012), we have $J = o_p(1)$ and $M_1 = M_2 + o_p(1)$. We also assumed $\widehat{\Lambda}$ and Λ_0 have the same column signs, as a part of identification condition.

□

B Proofs for Section 3

In this section, $(\widehat{\Lambda}, \widehat{\Sigma}_u) = (\widehat{\Lambda}^{(1)}, \widehat{\Sigma}_u^{(1)})$ and $J = (\widehat{\Lambda}^{(1)} - \Lambda_0)(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}$. Throughout Appendix B, we will let $H = (\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}$. For notational simplicity, we let

$$\omega_T = \frac{1}{\sqrt{N}} + \sqrt{\frac{\log N}{T}}.$$

We first cite a result from Fan et al. (2012):

Theorem B.1 (Theorem 3.1 in Fan et al. (2012))**.** *Suppose* $(\log N)^{6/\gamma} = o(T)$ *and* $\sqrt{T} = o(N)$, *then under Assumptions 3.1- 3.5,*

$$\|\widehat{\Sigma}_u^{(1)} - \Sigma_{u0}\| = O_p\left(m_N\omega_T^{1-q}\right) = \|(\widehat{\Sigma}_u^{(1)})^{-1} - \Sigma_{u0}^{-1}\|.$$

Proof. The sufficient conditions of this theorem are satisfied by our assumptions. See Fan et al. (2012). $\square$

We then prove Lemma A.3, which then enables us to apply Lemmas A.4 and A.5. Under Assumptions 3.1- 3.3, there is $\delta > 0$ such that $(\Lambda_0, \Sigma_{u0}) \in \Xi_\delta$ and $(\widehat{\Lambda}^{(1)}, \widehat{\Sigma}_u^{(1)}) \in \Xi_\delta$ with probability approaching one for Ξ_δ in Appendix A.

Lemma B.1. *For $(\widehat{\Lambda}, \widehat{\Sigma}_u) = (\widehat{\Lambda}^{(1)}, \widehat{\Sigma}_u^{(1)})$, Lemma A.3 is satisfied.*

Proof. The first order condition with respect to $\widehat{\Lambda}^{(1)}$ in (ii) is easy to verify, which is the same as that in Bai and Li (2012). We only show part (i).

By definition, $L_1(\widehat{\Lambda}^{(1)}) \leq L_1(\Lambda_0)$. Also the representation defined in Lemma A.2 yields

$$Q_3(\Lambda, \Sigma_u) + Q_2(\Lambda, \Sigma_u) = L_1(\Lambda) - N^{-1}\mathrm{tr}(S_u(\widehat{\Sigma}_u^{(1)})^{-1}) + N^{-1}\log|\widehat{\Sigma}_u^{(1)}|.$$

Thus

$$Q_2(\widehat{\Lambda}^{(1)}, \widehat{\Sigma}_u^{(1)}) + Q_3(\widehat{\Lambda}^{(1)}, \widehat{\Sigma}_u^{(1)}) \leq Q_2(\Lambda_0, \widehat{\Sigma}_u^{(1)}) + Q_3(\Lambda_0, \widehat{\Sigma}_u^{(1)})$$

Note that Q_2 is always nonnegative and $Q_2(\Lambda_0, \widehat{\Sigma}_u^{(1)}) = 0$. Therefore by Lemma A.2, $0 \leq Q_2(\widehat{\Lambda}^{(1)}, \widehat{\Sigma}_u^{(1)}) = o_p(1)$. Moreover, the matrix in the trace operation of Q_2 is semi-positive definite, hence

$$\frac{1}{N}\Lambda_0'(\widehat{\Sigma}_u^{(1)})^{-1}\Lambda_0 - (I_r - J)\frac{1}{N}\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}(I_r - J)' = o_p(1). \tag{B.1}$$

It remains to show that $N^{-1}\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1} - \Sigma_{u0}^{-1})\Lambda_0 = o_p(1)$, which follows immediately from Theorem B.1 and that $m_N\omega_T^{1-q} = o(1)$.

$\square$

B.1 Proof of Theorem 3.1

B.1.1 Consistency for $\widehat{\Lambda}^{(1)}$

The equality (B.1) implies

$$\frac{1}{N}(\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Lambda}^{(1)} - \Lambda_0) - \frac{1}{N}J\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}J' = o_p(1).$$

The second term is bounded by $N^{-1}\|J\|_F^2\|\widehat{\Lambda}^{(1)}\|_F^2\|(\widehat{\Sigma}_u^{(1)})^{-1}\| = O_p(\|J\|_F^2)$. Lemma A.5 then implies the second term is $o_p(1)$, which then implies that the first term is $o_p(1)$. Because $(\widehat{\Sigma}_u^{(1)})^{-1}$ has eigenvalues bounded away from zero asymptotically, we have $N^{-1}\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^2 = o_p(1)$.

B.1.2 Consistency for $\widehat{\lambda}_j^{(1)}$

Lemma A.4 (i) can be equivalently written as: for any $j \leq N$,

$$\widehat{\lambda}_j^{(1)} - \lambda_{0j} = -J'\lambda_{0j} + H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}a_j \tag{B.2}$$

where $\widehat{\Sigma}_{u,j}^{(1)}$ denotes the jth column of $\widehat{\Sigma}_u^{(1)}$, and a_j is an $N \times 1$ vector

$$a_j = \Lambda_0 T^{-1}\sum_{t=1}^{T} f_t u_{jt} + T^{-1}\sum_{t=1}^{T}(u_t u_{jt} - \widehat{\Sigma}_{u,j}^{(1)}) + T^{-1}\sum_{t=1}^{T} u_t f_t'\lambda_{0j} - \bar{u}\bar{u}_j.$$

The consistency of $\max_{j\leq N}\|\widehat{\lambda}_j^{(1)} - \lambda_{0j}\|$ follows from Lemma A.5 and the following Lemma B.2.

Lemma B.2. $\max_{j\leq N}\|H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}a_j\| = O_p(m_N N^{-1/2}\omega_T^{1-q} + T^{-1/2}(\log N)^{1/2})$.

Proof. By Lemma A.1, uniformly in $j \leq N$,

$$H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(\frac{1}{T}\sum_{t=1}^{T} u_t f_t'\lambda_{0j} + \Lambda_0\frac{1}{T}\sum_{t=1}^{T} f_t u_{jt}) = O_p(\frac{\sqrt{N}}{N}(2\sqrt{N}\sqrt{\frac{\log N}{T}})) = O_p(\sqrt{\frac{\log N}{T}}).$$

$$H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(\frac{1}{T}\sum_{t=1}^{T} u_t u_{jt} - \Sigma_{u0,j}) = O_p(\frac{\sqrt{N}}{N}\sqrt{N}\sqrt{\frac{\log N}{T}}) = O_p(\sqrt{\frac{\log N}{T}}).$$

$$H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Sigma}_{u,j}^{(1)} - \Sigma_{u0,j}) = O_p(\frac{\sqrt{N}}{N}m_N\omega_T^{1-q}) = O_p(\frac{m_N}{\sqrt{N}}\omega_T^{1-q}).$$

Finally, $\max_{j\leq N}\|H\widehat{\Lambda}^{(1)}(\widehat{\Sigma}_u^{(1)})^{-1}\bar{u}\bar{u}_j\| = O_p(\log N/T)$. The result then follows from a triangular inequality and that $m_N\omega_T^{1-q} = o(1)$. □

B.2 Proof of Theorem 3.2

B.2.1 Uniform rate for $\widehat{\lambda}_j^{(1)}$

By (B.2), the uniform rate of convergence follows from Lemma B.2 and the following Lemma B.3.

Lemma B.3. $J = O_p(m_N\omega_T^{1-q})$.

Proof. The first order condition in Lemma A.4 (i) is equivalent to:

$$J'J + J' + J + H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}B(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}H = 0 \tag{B.3}$$

where $B=\Lambda_0 T^{-1}\sum_{t=1}^T f_t u_t' + (\Lambda_0 T^{-1}\sum_{t=1}^T f_t u_t')' + S_u - \widehat{\Sigma}_u^{(1)} - \bar{u}\bar{u}'$. We have, $\|\Lambda_0\|_F = O(\sqrt{N})$, $\bar{u}\bar{u}' = O_p(N\log N/T)$, and $\|S_u - \widehat{\Sigma}_u^{(1)}\| \leq \|\widehat{\Sigma}_u^{(1)} - \Sigma_{u0}\| + \|S_u - \Sigma_{u0}\| = O_p(NT^{-1/2}(\log N)^{1/2} + m_N\omega_T^{1-q})$. Therefore $H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}B(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}H = O_p(T^{-1/2}(\log N)^{1/2} + m_N N^{-1}\omega_T^{1-q})$. Since $J = o_p(1)$, $J'J$ can be ignored. It follows from (B.3) that

$$J' + J = O_p(\sqrt{\frac{\log N}{T}} + \frac{m_N \omega_T^{1-q}}{N}). \tag{B.4}$$

Let J_{ij} denote the $(i,j)$the entry of J. It then follows that $J_{ii} = O_p(T^{-1/2}(\log N)^{1/2} + m_N N^{-1}\omega_T^{1-q})$ for all $i \leq r$. It is also not hard to verify that $\sqrt{(\log N)/T} = O(m_N \omega_T^{1-q})$ for any $0 \leq q < 1$ since $m_N \geq 1$.

On the other hand, due to the identification condition, both $\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0$ and $\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}$ are diagonal. Let ndg(M) denote the off-diagonal elements of M. Then ndg$(\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0)$ =ndg$(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}) = 0$ is equivalent to

$$\mathrm{ndg}\{(\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)} + \widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Lambda}^{(1)} - \Lambda_0)\}$$

$$= \mathrm{ndg}\{-\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1} - \Sigma_{u0}^{-1})\Lambda_0 + (\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Lambda}^{(1)} - \Lambda_0)\}$$

Note that if ndg$\{M_1\}$ = ndg$\{M_2\}$ then ndg$\{HM_1H\}$ = ndg$\{HM_2H\}$ for two matrices M_1 and M_2 since H is diagonal. Also, $(\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)}H = J$. The above identification condition implies

$$\mathrm{ndg}\{HJ + J'H\} = \mathrm{ndg}\{-H\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1} - \Sigma_{u0}^{-1})\Lambda_0 H + H(\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Lambda}^{(1)} - \Lambda_0)H\} \tag{B.5}$$

Note that $H\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1} - \Sigma_{u0}^{-1})\Lambda_0 H = O_p(m_N N^{-1}\omega_T^{1-q})$. Let h_{ii} denote the ith diagonal entry of H. Let $X = (\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Lambda}^{(1)} - \Lambda_0)$. Then for $i \neq j$, (B.4) and (B.5) imply that

$$J_{ji} + J_{ij} = O_p(\sqrt{\frac{\log N}{T}} + \frac{m_N \omega_T^{1-q}}{N}),$$

$$h_{ii}J_{ij} + h_{jj}J_{ji} = O_p(\frac{m_N \omega_T^{1-q}}{N}) + h_{ii}h_{jj}X_{ij}.$$

By assumption, with probability one, there is $\delta > 0$ such that $(N\delta)^{-1} < h_{ii} < N^{-1}\delta$, and $h_{ii} \neq h_{jj}$ for $i \neq j$. Moreover, since all the eigenvalues of $\widehat{\Sigma}_u$ are bounded away from zero and infinity, wpa1, $\|\widehat{\Lambda} - \Lambda_0\|_F^2 \geq c\|X\|_F$ for some $c > 0$. Then the above two equations imply that for any $i \neq j$, $J_{ij} = O_p(m_N\omega_T^{1-q}) + O_p(N^{-1})X_{ji}$ (since $\sqrt{\log N/T} = O(m_N\omega_T^{1-q})$). Then

$$\|J\|_F^2 = O_p(m_N^2 \omega_T^{2-2q} + \frac{1}{N^2}\|X\|_F^2). \tag{B.6}$$

Moreover, by Lemma B.2, $\max_{j\le N}\|H\widehat{\Lambda}^{(1)}(\widehat{\Sigma}_u^{(1)})^{-1}a_j\| = O_p(m_N\omega_T^{1-q})$.

We now show that $J = O_p(m_N\omega_T^{1-q})$. Suppose this does not hold, then (B.6) implies $J = O_p(N^{-1}X)$. By the definition

$$X = (\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Lambda}^{(1)} - \Lambda_0),$$

$\|X\|_F = O_p(\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^2)$. Therefore $J = O_p(N^{-1}X)$ yields $\|J\|_F^2 = O_p(N^{-2}\|\widehat{\Lambda} - \Lambda_0\|_F^4)$. The first order condition (B.2) also yields

$$\max_{j\le N}\|\widehat{\lambda}_j^{(1)} - \lambda_{0j}\|^2 = O_p(\|J\|_F^2) = O_p(N^{-2}\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^4),$$

which implies $\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^2 = \sum_{j=1}^N \|\widehat{\lambda}_j^{(1)} - \lambda_{0j}\|^2 = O_p(N^{-1}\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^4)$. Therefore

$$\frac{1}{N^{-1}\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^2} = \frac{\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^2}{N^{-1}\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^4} = O_p(1),$$

which contradicts with the consistency $N^{-1}\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F^2 = o_p(1)$. This concludes the proof. $\square$

Therefore, (B.2) gives $\max_{j\le N}\|\widehat{\lambda}_j^{(1)} - \lambda_{0j}\| = O_p(\|J\|_F) = O_p(m_N\omega_T^{1-q})$. The rate of convergence for $N^{-1/2}\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F$ then follows immediately since it is bounded by $\max_{j\le N}\|\widehat{\lambda}_j^{(1)} - \lambda_{0j}\|$.

B.3 Proof of Theorem 3.3

By the definition of the covariance estimator in the first step, $\widehat{\Sigma}_u^{(1)} = (s_{ij}(R_{ij}))_{N\times N}$, where s_{ij} is a chosen thresholding function. It was shown by Fan et al. (2012, Theorem 2.1) that R_{ij} is the PCA estimator of $T^{-1}\sum_{t=1}^T u_{it}u_{jt}$, that is, $R_{ij} = T^{-1}\sum_{t=1}^T \widehat{u}_{it}^{PCA}\widehat{u}_{jt}^{PCA}$.

Lemma B.4. *For any $\epsilon > 0$, and any constant $M > 0$, for all large enough N, T,*

$$P(|R_{ij}| > M\tau_{ij}, \forall(i,j)\in S_U) > 1-\epsilon.$$

Proof. We have, $|R_{ij}| \ge |\Sigma_{u0,ij}| - |\Sigma_{u0,ij} - R_{ij}|$. Thus for all large enough N, T,

$$\begin{aligned} P(|R_{ij}| > M\tau_{ij}, \forall(i,j)\in S_U) &\ge P(|\Sigma_{u0,ij}| > M\tau_{ij} + |\Sigma_{u0,ij} - R_{ij}|, \forall(i,j)\in S_U) \\ &\ge P(|\Sigma_{u0,ij}|/2 > |\Sigma_{u0,ij} - R_{ij}|, \forall(i,j)\in S_U) > 1-\epsilon, \end{aligned}$$

where in the second and last inequalities we used the assumption that $\omega_T =$

$o(\min_{(i,j)\in S_U}|\Sigma_{u0,ij}|)$ and the fact that $\max_{ij}|\Sigma_{u0,ij}-R_{ij}|=O_p(\omega_T)$.

$\square$

Proof of Theorem 3.3

By Fan et al. (2012), $\max_{i,j}|R_{ij}-\Sigma_{u0,ij}|=O_p(\omega_T)$, which implies for any $\epsilon>0$, there is $C>0$ such that $P(\max_{i,j}|R_{ij}-\Sigma_{u0,ij}|>C\omega_T)<\epsilon/2$. For some universal $M>0$, we set the threshold $\tau_{ij}=M\alpha_{ij}\omega_T$ at entry (i,j), where α_{ij} is a data-dependent value that satisfies, for any $\epsilon>0$, there is $C_1>0$ such that $P(\alpha_{ij}>C_1,\forall i\neq j)>1-\epsilon/2$. Then as long as the constant M in the definition of the threshold is larger than $2C/C_1$,

$$P(\max_{i,j}|R_{ij}-\Sigma_{u0,ij}|>\min_{ij}\tau_{ij}/2)<P(\max_{i,j}|R_{ij}-\Sigma_{u0,ij}|>MC_1\omega_T/2)+\epsilon/2<\epsilon.$$

Note also that if $s_{ij}(R_{ij})\equiv\widehat{\Sigma}_{ij}^{(1)}\neq 0$, then $|R_{ij}|>\tau_{ij}$, by the definition of s_{ij}. This implies,

$$P(\widehat{\Sigma}_{ij}^{(1)}\neq 0,\exists(i,j)\in S_L)\leq P(|R_{ij}|>\tau_{ij},\exists(i,j)\in S_L)\leq P(\max_{(i,j)\in S_L}|R_{ij}|>\min_{ij}\tau_{ij})$$

$$\leq P(\max_{i,j}|R_{ij}-\Sigma_{u0,ij}|+\max_{(i,j)\in S_L}|\Sigma_{u0,ij}|>\min_{ij}\tau_{ij}).$$

Since $\max_{(ij)\in S_L}|\Sigma_{u0,ij}|=o(\omega_T)$ by assumption, for all large T,N

$$P(\widehat{\Sigma}_{ij}^{(1)}\neq 0,\exists(i,j)\in S_L)\leq P(\max_{i,j}|R_{ij}-\Sigma_{u0,ij}|>\min_{ij}\tau_{ij}/2)<\epsilon.$$

On the other hand, for arbitrarily small $\epsilon>0$, $P(\max_{ij}\tau_{ij}\leq K\omega_T)>1-\epsilon/2$ for some $K>0$, which implies $P(|R_{ij}|\geq M\omega_T+K\omega_T,\forall(i,j)\in S_U)\leq P(|R_{ij}|\geq M\omega_T+\tau_{ij},\forall(i,j)\in S_U)+\epsilon/2$. By the definition of s_{ij}, $|s_{ij}(z)-z|\leq\tau_{ij}$ for all $z\in\mathbb{R}$. Therefore $|R_{ij}-\widehat{\Sigma}_{u,ij}^{(1)}|=|R_{ij}-s_{ij}(R_{ij})|\leq\tau_{ij}$, hence for arbitrarily large $M>0$,

$$P(|\widehat{\Sigma}_{u,ij}^{(1)}|>M\omega_T,\forall(i,j)\in S_U)\geq P(|R_{ij}|\geq M\omega_T+|R_{ij}-\widehat{\Sigma}_{u,ij}^{(1)}|,\forall(i,j)\in S_U)$$

$$\geq P(|R_{ij}|\geq(M+K)\omega_T,\forall(i,j)\in S_U)-\epsilon/2\geq 1-\epsilon$$

where the last inequality follows from Lemma B.4.

B.4 Proof of Theorems 3.4 and 3.5

B.4.1 Proof of Theorem 3.4

A simple derivation implies that $\|N^{-1}\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1}-\Sigma_{u0}^{-1})\Lambda_0\|_F\leq N^{-1}\|\Lambda_0\|_F^2\|(\widehat{\Sigma}_u^{(1)})^{-1}-\Sigma_{u0}^{-1}\|=O_p(m_N\omega_T^{1-q})$. This rate is not tight enough for the $\sqrt{T}$-consistency and limiting distribution $\widehat{\lambda}_j^{(1)}$. A more refined rate of $N^{-1}\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1}-\Sigma_{u0}^{-1})\Lambda_0$ depends on the convergence properties of the PCA estimator. We begin by citing some results proved by Fan et al. (2012). Recall that R_{ij} denotes the (i,j)th entry of the orthogonal complement covariance in the sample covariance's spectrum decomposition, and $\widehat{\Sigma}_{u,ij}^{(1)} = s_{ij}(R_{ij})$.

Let $\{\widehat{u}_{it}\}_{i\leq N,t\leq T}$ be the PCA estimates of $\{u_{it}\}_{i\leq N,t\leq T}$. Let $\widehat{\lambda}_j^{PCA}$ and $\widehat{f}_t^{PCA}$ denote the PCA estimators of the factor loadings and factors.

Lemma B.5. *(i) For any i,j, with probability one* $R_{ij} = T^{-1}\sum_{t=1}^T \widehat{u}_{it}\widehat{u}_{jt}$,
(ii) $\max_{i\leq N} T^{-1}\sum_{t=1}^T(\widehat{u}_{it}-u_{it})^2 = O_p(\omega_T^2)$.
(iii) There is a nonsingular matrix $\bar{H}$ such that $T^{-1}\sum_{t=1}^T \|\widehat{f}_t^{PCA} - \bar{H}f_t\|^2 = O_p(T^{-1}+N^{-1})$ *and* $\max_j \|\widehat{\lambda}_j^{PCA} - \bar{H}'^{-1}\lambda_{0j}\| = O_p(\omega_T)$.
(iv) $\max_{i,j\leq N}|R_{ij} - \Sigma_{u0,ij}| = O_p(\omega_T)$.

Proof. See Theorem 2.1 and Lemma C.11 of Fan et al. (2012). □

Lemma B.6. $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N u_{it}\lambda_{0i}'\bar{H}^{-1}(\widehat{f}_t^{PCA} - \bar{H}f_t)\xi_i\xi_i' = O_p(\frac{1}{\sqrt{NT}} + \frac{1}{T} + \frac{1}{N})$.

Proof. By Bai (2003), there are two $r\times r$ matrices $\bar{H}$ and V, $\|V\|_F = O_p(1)$, $\|\bar{H}\|_F = O_p(1)$ such that $\widehat{f}_t^{PCA} - \bar{H}f_t = V(NT)^{-1}\sum_{s=1}^T \widehat{f}_s^{PCA}[u_s'u_t + f_s'\sum_{j=1}^N \lambda_{0j}u_{jt} + f_t'\sum_{j=1}^N\lambda_{0j}u_{js}]$. The desired result then follows from the following Lemma B.7.

□

Lemma B.7. *(i)* $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N u_{it}\lambda_{0i}'\bar{H}^{-1}(NT)^{-1}\sum_{s=1}^T\widehat{f}_s^{PCA}u_s'u_t\xi_i\xi_i' = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T}+\frac{1}{N})$
(ii) $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N u_{it}\lambda_{0i}'\bar{H}^{-1}(NT)^{-1}\sum_{s=1}^T\widehat{f}_s^{PCA}f_s'\sum_{j=1}^N\lambda_{0j}u_{jt}\xi_i\xi_i' = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{N})$
(iii) $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N u_{it}\lambda_{0i}'\bar{H}^{-1}(NT)^{-1}\sum_{s=1}^T\widehat{f}_s^{PCA}f_t'\sum_{j=1}^N\lambda_{0j}u_{js}\xi_i\xi_i' = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T})$.

Proof. (i) We have,

$$\|\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N u_{it}\frac{1}{NT}\sum_{s=1}^T\widehat{f}_s^{PCA'}u_s'u_t\bar{H}^{-1'}\lambda_{0i}\xi_i\xi_i'\| \leq \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{i=1}^N u_{it}\sum_{s=1}^T f_s'\bar{H}'u_s'u_t\bar{H}^{-1'}\lambda_{0i}\xi_i\xi_i'\|$$
$$+\|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{i=1}^N u_{it}\sum_{s=1}^T(\widehat{f}_s^{PCA'} - f_s'\bar{H}')u_s'u_t\bar{H}^{-1'}\lambda_{0i}\xi_i\xi_i'\| = a+b. \tag{B.7}$$

We bound a,b separately. Here a is upper bounded by a_1+a_2, where by Cauchy-Schwarz,

$$\begin{aligned} a_1 &= \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{i=1}^N u_{it}\sum_{s=1}^T f_s'\bar{H}'(u_s'u_t - Eu_s'u_t)\bar{H}^{-1'}\lambda_{0i}\xi_i\xi_i'\| \\ &\leq \max_{i\leq N}\|\lambda_{0i}\xi_i\xi_i'\|(\frac{1}{T}\sum_{t=1}^T u_{it}^2)^{1/2}\|\frac{1}{N}(\frac{1}{T}\sum_{t=1}^T\|\frac{1}{T}\sum_{s=1}^T f_s(u_s'u_t - Eu_s'u_t)\|^2)^{1/2}\end{aligned}$$

$$\leq \quad O_p(1)(\frac{1}{T}\sum_{t=1}^{T}\|\frac{1}{TN}\sum_{s=1}^{T}f_s(u_s'u_t - Eu_s'u_t)\|^2)^{1/2}. \tag{B.8}$$

Note that $E\frac{1}{T}\sum_{t=1}^{T}\|\frac{1}{TN}\sum_{s=1}^{T}f_s(u_s'u_t - Eu_s'u_t)\|^2 = E\|\frac{1}{TN}\sum_{s=1}^{T}f_s(u_s'u_t - Eu_s'u_t)\|^2$, which is $O(T^{-1}N^{-1})$ by Assumption 3.9. Hence $a_1 = O_p((NT)^{-1/2})$.

$$a_2 = \|\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}u_{it}\sum_{s=1}^{T}f_s'\bar{H}'Eu_s'u_t\bar{H}^{-1'}\lambda_{0i}\xi_i\xi_i'\| \leq \max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}|u_{it}|O(1)\frac{1}{TN}\sum_{s=1}^{T}\|f_sEu_s'u_t\| \tag{B.9}$$

Since $\max_{t\leq T}E(T^{-1}N^{-1}\sum_{s=1}^{T}\|f_sEu_s'u_t\|) \leq O(T^{-1})\max_t\sum_{s=1}^{T}|Eu_s'u_t|/N = O(T^{-1})$ by the strong mixing condition (Lemma C.5 of Fan Liao and Mincheva 2012), we have $a_2 = O_p(T^{-1})$. This implies $a = O_p(N^{-1/2}T^{-1/2} + T^{-1})$.

Now we bound b. Using Cauchy Schwarz inequality, we have $b \leq b_1 + b_2$ where

$$\begin{aligned}
b_1 &= \|\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}u_{it}\sum_{s=1}^{T}(\widehat{f}_s^{PCA'} - f_s'\bar{H}')(u_s'u_t - Eu_s'u_t)\bar{H}^{-1'}\lambda_{0i}\xi_i\xi_i'\| \\
&\quad \leq O_p(1)\frac{1}{N^2T}\sum_{t=1}^{T}\sum_{i=1}^{N}|u_{it}|\left(\frac{1}{T}\sum_{s=1}^{T}\|\widehat{f}_s^{PCA} - \bar{H}f_s\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}|u_s'u_t - Eu_s'u_t|^2\right)^{1/2} \\
&\leq O_p(\frac{1}{N})O_p(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}})O_p(\sqrt{N}) = O_p(\frac{1}{N} + \frac{1}{\sqrt{NT}}),
\end{aligned} \tag{B.10}$$

where the second inequality follows from $ET^{-1}\sum_{s=1}^{T}|u_s'u_t - Eu_s'u_t|^2 = O(N)$. Using Cauchy-Schwarz inequality, we also obtain

$$\begin{aligned}
b_2 &= \|\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}u_{it}\sum_{s=1}^{T}(\widehat{f}_s^{PCA'} - f_s'\bar{H}')(Eu_s'u_t)\bar{H}^{-1'}\lambda_{0i}\xi_i\xi_i'\| \\
&\leq O_p(1)(\frac{1}{T}\sum_{s=1}^{T}\|\widehat{f}_s^{PCA} - \bar{H}f_s\|^2)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}|Eu_s'u_t/N|^2\right)^{1/2} \\
&= O_p(\frac{1}{\sqrt{NT}} + \frac{1}{T}).
\end{aligned} \tag{B.11}$$

(ii) Let $d_{i,kl}$ be the (k,l)th element of $\xi_i\xi_i'$. Then the (k,l)th element of the object of interest is bounded by $d_1 + d_2$, where, by Cauchy Schwarz inequality,

$$\begin{aligned}
d_1 &= |\frac{1}{(NT)^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N}(u_{it}u_{jt} - Eu_{it}u_{jt})\lambda_{0i}'\bar{H}^{-1}\widehat{f}_s^{PCA}f_s'\lambda_{0j}d_{i,kl}| \\
&\leq O_p(1)(\frac{1}{T}\sum_{s=1}^{T}\|\widehat{f}_s^{PCA}\|^2)^{1/2}(\frac{1}{T}\sum_{s=1}^{T}\|\widehat{f}_s^{PCA}\|^2)^{1/2}\|\frac{1}{N^2T}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(u_{it}u_{jt} - Eu_{it}u_{jt})\lambda_{0i}\lambda_{0j}'d_{i,kl}\|
\end{aligned}$$

$$= O_p(\frac{1}{\sqrt{NT}}). \tag{B.12}$$

The last equality follows from Assumption 3.9. Also, $\sum_{i,j\leq N}|Eu_{it}u_{jt}| = \sum_{(i,j)\in S_U}|\Sigma_{u0,ij}| + \sum_{(i,j)\in S_L}|\Sigma_{u0,ij}| = O(N)$. Thus

$$\begin{aligned} d_2 &= |\frac{1}{(NT)^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{j=1}^N\sum_{i=1}^N(Eu_{it}u_{jt})\lambda_{0i}'\bar{H}^{-1}\widehat{f}_s^{PCA}f_s'\lambda_{0j}d_{i,kl}| \\ &\leq O_p(1)\frac{1}{N^2T}\sum_{s=1}^T\|\widehat{f}_s^{PCA}\|\|f_s\|\sum_{i,j\leq N}|Eu_{it}u_{jt}| = O_p(\frac{1}{N}). \end{aligned} \tag{B.13}$$

(iii) The object of interest is bounded by e_1+e_2, where

$$e_1 = \|\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{j=1}^N\sum_{t=1}^T\sum_{i=1}^N u_{it}\lambda_{0i}'\bar{H}^{-1}(\widehat{f}_s^{PCA}-\bar{H}f_s)f_t'\lambda_{0j}u_{js}\xi_i\xi_i'\| = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T}), \tag{B.14}$$

and we used the fact that $\frac{1}{T}\sum_{t=1}^T\|\widehat{f}_t^{PCA}-\bar{H}f_t\|^2 = O_p(T^{-1}+N^{-1})$ from Lemma B.5, and that $N^{-1}\sum_{i=1}^N\|\frac{1}{T}\sum_{t=1}^T f_tu_{it}\| = O_p(T^{-1/2})$.[1]

$$e_2 = \|\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{j=1}^N\sum_{t=1}^T\sum_{i=1}^N u_{it}\lambda_{0i}'f_sf_t'\lambda_{0j}u_{js}\xi_i\xi_i'\| = O_p(\frac{1}{T}). \tag{B.15}$$

□

Lemma B.8. *For S_U in the partition $\{(i,j): i,j\leq N\} = S_L\cup S_U$,*
(i) $\frac{1}{NT}\sum_{t=1}^T\sum_{i\neq j,(i,j)\in S_U}u_{it}\lambda_{0j}'\bar{H}^{-1}(NT)^{-1}\sum_{s=1}^T\widehat{f}_s^{PCA}u_s'u_t\xi_i\xi_j' = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T}+\frac{1}{N})$
(ii) $\frac{1}{NT}\sum_{t=1}^T\sum_{i\neq j,(i,j)\in S_U}u_{it}\lambda_{0j}'\bar{H}^{-1}(NT)^{-1}\sum_{s=1}^T\widehat{f}_s^{PCA}f_s'\sum_{v=1}^N\lambda_{0v}u_{vt}\xi_i\xi_j' = O_p(\frac{1}{\sqrt{NT}}+\frac{m_N}{N})$
(iii) $\frac{1}{NT}\sum_{t=1}^T\sum_{i\neq j,(i,j)\in S_U}u_{it}\lambda_{0j}'\bar{H}^{-1}(NT)^{-1}\sum_{s=1}^T\widehat{f}_s^{PCA}f_t'\sum_{v=1}^N\lambda_{0v}u_{vs}\xi_i\xi_j' = O_p(\sqrt{\frac{\log N}{NT}}+\frac{\log N}{T})$.

Proof. (i) The term of interest is bounded by $a+b$, where

$$\begin{aligned} a &= \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{(i,j)\in S_U,i\neq j}u_{it}\sum_{s=1}^T f_s'\bar{H}'u_s'u_t\bar{H}^{-1'}\lambda_{0j}\xi_i\xi_j'\|, \\ b &= \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{(i,j)\in S_U,i\neq j}u_{it}\sum_{s=1}^T(\widehat{f}_s^{PCA'}-f_s'\bar{H}')u_s'u_t\bar{H}^{-1'}\lambda_{0j}\xi_i\xi_j'\|. \end{aligned}$$

[1] We have $(N^{-1}\sum_{i=1}^N\|\frac{1}{T}\sum_{t=1}^T f_tu_{it}\|)^2\leq N^{-1}\sum_{i=1}^N\|\frac{1}{T}\sum_{t=1}^T f_tu_{it}\|^2 = N^{-1}\sum_{i=1}^N\sum_{j=1}^r(\frac{1}{T}\sum_{t=1}^T f_tu_{it})^2$, whose expectation is $N^{-1}\sum_{i=1}^N\sum_{j=1}^r\mathrm{var}(\frac{1}{T}\sum_{t=1}^T f_{jt}u_{it})$. Note that $\mathrm{var}(\frac{1}{T}\sum_{t=1}^T f_{jt}u_{it}) = O(T^{-1})$ uniformly in $i\leq N$.

Here a is upper bounded by $a_1 + a_2$, where
$a_1 = \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{(i,j)\in S_U, i\neq j} u_{it}\sum_{s=1}^T f_s'\bar{H}'(u_s'u_t - Eu_s'u_t)\bar{H}^{-1'}\lambda_{0j}\xi_i\xi_j'\|$, and
$a_2 = \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{(i,j)\in S_U, i\neq j} u_{it}\sum_{s=1}^T f_s'\bar{H}'Eu_s'u_t\bar{H}^{-1'}\lambda_{0j}\xi_i\xi_j'\|$. Note that a_1 and a_2 can be bounded in the same way as (B.8) and (B.9). The only difference is that $N^{-1}\sum_{i=1}^N$ is replaced by a double sum $N^{-1}\sum_{(i,j)\in S_U, i\neq j}$. By the assumption, $N^{-1}\sum_{(i,j)\in S_U, i\neq j} 1 = O(1)$. The result of the proof is exactly the same, so is omitted. We conclude that $a = O_p(N^{-1/2}T^{-1/2} + T^{-1})$.

On the other hand, $b \leq b_1 + b_2$ where
$b_1 = \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{(i,j)\in S_U, i\neq j} u_{it}\sum_{s=1}^T(\widehat{f}_s^{PCA'} - f_s'\bar{H}')(u_s'u_t - Eu_s'u_t)\bar{H}^{-1'}\lambda_{0j}\xi_i\xi_j'\|$, and
$b_2 = \|\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{(i,j)\in S_U, i\neq j} u_{it}\sum_{s=1}^T(\widehat{f}_s^{PCA'} - f_s'\bar{H}')(Eu_s'u_t)\bar{H}^{-1'}\lambda_{0j}\xi_i\xi_j'\|$. Using Cauchy-Schwarz inequality and the strong mixing condition, b_1 and b_2 can be also bounded in an exactly the same way of (B.10) and (B.11). We conclude that $b = O_p(N^{-1} + T^{-1} + (NT)^{-1/2})$.

(ii) Let $d_{ij,kl}$ be the (k,l)th element of $\xi_i\xi_j'$. Then the (k,l)th element of the object of interest is bounded by $d_1 + d_2$, where
$d_1 = |\frac{1}{(NT)^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{(i,j)\in S_U, i\neq j}\sum_{v=1}^N(u_{it}u_{vt} - Eu_{it}u_{vt})\lambda_{0j}'\bar{H}^{-1}\widehat{f}_s^{PCA}f_s'\lambda_{0v}d_{ij,kl}|$, and
$d_2 = |\frac{1}{(NT)^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{(i,j)\in S_U, i\neq j}\sum_{v=1}^N(Eu_{it}u_{vt})\lambda_{0j}'\bar{H}^{-1}\widehat{f}_s^{PCA}f_s'\lambda_{0v}d_{ij,kl}|$. Bounding d_1, d_2 is slightly different from (B.12) and (B.13), and we give the detail here. By Cauchy Schwarz inequality,

$$d_1 \leq O_p(1)(\frac{1}{T}\sum_{s=1}^T\|\widehat{f}_s^{PCA}\|^2)^{1/2}(\frac{1}{T}\sum_{s=1}^T\|f_s\|^2)^{1/2}\|\frac{1}{N^2T}\sum_{i\neq j,(i,j)\in S_U}\sum_{t=1}^T\sum_{v=1}^N(u_{it}u_{vt} - Eu_{it}u_{vt})\lambda_{0j}\lambda_{0v}'d_{ij,kl}\|$$

which is $O_p((NT)^{-1/2})$ by Assumption 3.9. On the other hand,
$d_2 \leq O_p(N^{-2})\sum_{i\neq j,(i,j)\in S_U}\sum_{k=1}^N|\Sigma_{u0,ik}|$. Note that $\|\Sigma_{u0}\|_1 = O(m_N)$, where m_N is as defined in Assumption 3.1. Thus $d_2 = O_p(N^{-1}m_N)$.

(iii) The object of interest is bounded by $e_1 + e_2$, where
$e_1 = \|\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{i\neq j,(i,j)\in S_U}\sum_{t=1}^T\sum_{v=1}^N u_{it}\lambda_{0j}'\bar{H}^{-1}(\widehat{f}_s^{PCA} - \bar{H}f_s)f_t'\lambda_{0v}u_{vs}\xi_i\xi_j'\|$,
$e_1 = \|\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{i\neq j,(i,j)\in S_U}\sum_{t=1}^T\sum_{v=1}^N u_{it}\lambda_{0j}'f_sf_t'\lambda_{0v}u_{vs}\xi_i\xi_j'\|$.

Since $\max_{i\leq N}\|T^{-1}\sum_{t=1}^T f_tu_{it}\| = O_p(\sqrt{\log N/T})$, we conclude that $e_1 = O_p(\frac{\sqrt{\log N}}{T} + \frac{\sqrt{\log N}}{\sqrt{NT}})$, and $e_2 = O_p(\frac{\log N}{T})$. □

From Lemma B.8, immediately we have the following result.

Lemma B.9. $\frac{1}{NT}\sum_{t=1}^T\sum_{i\neq j,(i,j)\in S_U} u_{it}\lambda_{0j}'\bar{H}^{-1}(\widehat{f}_t^{PCA} - \bar{H}f_t)\xi_i\xi_j' = O_p(\omega_T^2 + m_N/N)$.

Proof. Note that results (i)(ii)(iii) in Lemma B.8 sum up to $O_p(\omega_T^2 + m_N/N)$. Hence Lemma B.9 follows from the equality $\widehat{f}_t^{PCA} - \bar{H}f_t = V(NT)^{-1}\sum_{s=1}^T\widehat{f}_s^{PCA}[u_s'u_t + f_s'\sum_{j=1}^N\lambda_{0j}u_{jt} + f_t'\sum_{j=1}^N\lambda_{0j}u_{js}]$. □

The following lemma strengthens the results of Bai (2003) when Σ_{u0} is sparse.

Lemma B.10. *For the PCA estimator,*
(i) $N^{-1}\sum_{i=1}^N (R_{ii}-\Sigma_{u0,ii})\xi_i\xi_i' = O_p(\omega_T^2)$.
(ii) $N^{-1}\sum_{i\neq j,(i,j)\in S_U}(R_{ij}-\Sigma_{u0,ij})\xi_i\xi_j' = O_p(\omega_T^2 + m_N/N)$.

Proof. (i) $N^{-1}\sum_{i=1}^N (R_{ii}-\Sigma_{u0,ii})\xi_i\xi_i' = \sum_{i=1}^N (R_{ii}-S_{u,ii})\xi_i\xi_i'/N + \sum_{i=1}^N (S_{u,ii}-\Sigma_{u0,ii})\xi_i\xi_i'/N$. By Assumption 3.9, $\sum_{i=1}^N (S_{u,ii}-\Sigma_{u0,ii})\xi_i\xi_i'/N = \sum_{i=1}^N\sum_{t=1}^T (u_{it}^2-\Sigma_{u0,ii})\xi_i\xi_i'/(NT) = O_p(1/\sqrt{NT})$. On the other hand, $\frac{1}{N}\sum_{i=1}^N (R_{ii}-S_{u,ii})\xi_i\xi_i'$ is equal to

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\widehat{u}_{it}^2-u_{it}^2)\xi_i\xi_i' = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\widehat{u}_{it}-u_{it})^2\xi_i\xi_i' + \frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T u_{it}(\widehat{u}_{it}-u_{it})\xi_i\xi_i'.$$

The first term on the right hand side is $O_p(\omega_T^2)$. We now work on the second term. By Bai (2003), there is a nonsingular matrix $\bar{H}$ such that

$$\widehat{u}_{jt}-u_{jt} = \lambda_{0j}'\bar{H}^{-1}(\widehat{f}_t^{PCA}-\bar{H}f_t) + (\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j})'(\widehat{f}_t^{PCA}-\bar{H}f_t) + (\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j})\bar{H}f_t. \tag{B.16}$$

By Lemma B.6 $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N u_{it}\lambda_{0i}'\bar{H}^{-1}(\widehat{f}_t^{PCA}-\bar{H}f_t)\xi_i\xi_i' = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T}+\frac{1}{N})$. In addition, for each element $d_{i,kl}$ of $\xi_i\xi_i'$,

$$\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T u_{jt}(\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j})\bar{H}f_t d_{j,kl} \leq \frac{1}{N}\sum_{j=1}^N \|d_{j,kl}\frac{1}{T}\sum_{t=1}^T u_{jt}f_t'\bar{H}'\| \max_j \|\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j}\|,$$

which is $O_p(\omega_T\sqrt{\frac{\log N}{T}})$. Also,

$$\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T u_{jt}(\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j})'(\widehat{f}_t^{PCA}-\bar{H}f_t)d_{j,kl} = \frac{1}{T}\sum_{t=1}^T(\widehat{f}_t^{PCA}-\bar{H}f_t)'\frac{1}{N}\sum_{j=1}^N u_{jt}(\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j})d_{j,kl}$$

$$\leq \left(\frac{1}{T}\sum_{t=1}^T\|\widehat{f}_t^{PCA}-\bar{H}f_t\|^2\max_j\|\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j}\|^2\frac{1}{T}\sum_{t=1}^T[\frac{1}{N}\sum_{j=1}^N|u_{jt}d_{j,kl}|]^2\right)^{1/2} = O_p(\frac{\omega_T}{\sqrt{T}}+\frac{\omega_T}{\sqrt{N}}).$$

(ii) Since $R_{ij}=T^{-1}\sum_{t=1}^T\widehat{u}_{it}\widehat{u}_{jt}$, the term of interest equals

$$\frac{2}{N}\sum_{i\neq j,(i,j)\in S_U}\frac{1}{T}\sum_{t=1}^T u_{it}(\widehat{u}_{jt}-u_{jt})\xi_i\xi_j' + \frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it}-u_{it})(\widehat{u}_{jt}-u_{jt})\xi_i\xi_j'$$

$$+\frac{1}{NT}\sum_{i\neq j,(i,j)\in S_U}\sum_{t=1}^T(u_{it}u_{jt}-\Sigma_{u0,ij})\xi_i\xi_j'.$$

By Assumption 3.9, the third term is $O_p((NT)^{-1/2})$. By the assumption that $\sum_{i\neq j,(i,j)\in S_U} 1 = O(N)$ and Cauchy Schwarz inequality, the second term is $O_p(\omega_T^2)$. We now work out the first term. Again we use the equality $\widehat{u}_{jt}-u_{jt} = \lambda_{0j}'\bar{H}^{-1}(\widehat{f}_t^{PCA}-\bar{H}f_t)+(\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j})'(\widehat{f}_t^{PCA}-\bar{H}f_t)+(\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j})'\bar{H}f_t$. Lemma B.9 gives

$$\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}\frac{1}{T}\sum_{t=1}^{T}u_{jt}\lambda_{0i}'\bar{H}^{-1}(\widehat{f}_t^{PCA}-\bar{H}f_t)\xi_i\xi_j' = O_p(\omega_T^2+\frac{m_N}{N}).$$

On the other hand, $\frac{2}{N}\sum_{i\neq j,(i,j)\in S_U}\frac{1}{T}\sum_{t=1}^{T}u_{jt}(\widehat{\lambda}_i^{PCA}-\bar{H}^{-1'}\lambda_{0i})'\bar{H}f_t\xi_i\xi_j'$ is bounded by,

$$\max_{i\leq N}\|\xi_i\|^2\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}\|\frac{1}{T}\sum_{t=1}^{T}u_{jt}f_t'\bar{H}'\|\max_j\|\widehat{\lambda}_j^{PCA}-\bar{H}'^{-1}\lambda_{0j}\| = O_p(\omega_T\sqrt{\frac{\log N}{T}}),$$

since $\max_{i\leq N}\|\xi_i\| = O(1)$. Also, $\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}\frac{1}{T}\sum_{t=1}^{T}u_{jt}(\widehat{\lambda}_i^{PCA}-\bar{H}^{-1'}\lambda_{0i})'(\widehat{f}_t^{PCA}-\bar{H}f_t)\xi_i\xi_j'$ is bounded by

$$O(1)\max_{i\leq N}\|\hat{b}_i-H^{-1'}\lambda_{0i}\|\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t-\bar{H}f_t\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}[\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}|d_{ij,kl}u_{it}|]^2\right)^{1/2}$$

which is $O_p(\frac{\omega_T}{\sqrt{T}}+\frac{\omega_T}{\sqrt{N}})$.

□

Proof of Theorem 3.4 $N^{-1}\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1}-\Sigma_{u0}^{-1})\Lambda_0 = O_p(\omega_T^{2-2q}m_N^2)$

Proof. By the triangular inequality, the left-hand-side is bounded by

$$\frac{1}{N}\|\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1}-\Sigma_{u0}^{-1})(\Sigma_{u0}-\widehat{\Sigma}_u^{(1)})\Sigma_{u0}^{-1}\Lambda_0\|_F+\frac{1}{N}\|\Lambda_0'\Sigma_{u0}^{-1}(\Sigma_{u0}-\widehat{\Sigma}_u^{(1)})\Sigma_{u0}^{-1}\Lambda_0\|_F.$$

The first term is $O_p(\omega_T^{2-2q}m_N^2)$. We now bound the second term, which is

$$\begin{aligned}\frac{1}{N}\Xi(\widehat{\Sigma}_u^{(1)}-\Sigma_{u0})\Xi' &= \frac{1}{N}\sum_{i=1}^{N}(R_{ii}-\Sigma_{u0,ii})\xi_i\xi_i'+\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}(\widehat{\Sigma}_{u,ij}^{(1)}-\Sigma_{u0,ij})\xi_i\xi_j'\\ &\quad+\frac{1}{N}\sum_{(i,j)\in S_L}(\widehat{\Sigma}_{u,ij}^{(1)}-\Sigma_{u0,ij})\xi_i\xi_j',\end{aligned}$$

where $\Xi=\Lambda_0'\Sigma_{u0}^{-1}$. The first term on the right hand side is $O_p(\omega_T^2)$ by Lemma B.10. The third term is dominated by, $O(N^{-1})(\sum_{S_L}|\Sigma_{u0,ij}|+\sum_{S_L}|\widehat{\Sigma}_{u,ij}^{(1)}|) = O(N^{-1})+O(N^{-1})\sum_{S_L}|\widehat{\Sigma}_{u,ij}^{(1)}|$. By Theorem 3.3, for any $\epsilon>0$ and any $M>0$, $P(\frac{1}{N}\sum_{(i,j)\in S_L}|\widehat{\Sigma}_{u,ij}^{(1)}|>M\omega_T^2)\leq P(\exists(i,j)\in$

$S_L, \widehat{\Sigma}^{(1)}_{u,ij} \neq 0) \leq \epsilon$. This implies the third term is $O_p(\omega_T^2)$. The second term equals

$$\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}(\widehat{\Sigma}^{(1)}_{u,ij} - R_{ij})\xi_i\xi_j' + \frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}(R_{ij} - \Sigma_{u0,ij})\xi_i\xi_j'.$$

By Lemma B.10 (ii), $N^{-1}\sum_{i\neq j,(i,j)\in S_U}(R_{ij} - \Sigma_{u0,ij})\xi_i\xi_j' = O_p(\omega_T^2 + m_N/N)$. On the other hand, recall that $|s_{ij}(z) - z| \leq a\tau_{ij}^2$ when $|z| > b\tau_{ij}$ (Section 3.1),

$$\begin{aligned}&\|\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}(\widehat{\Sigma}^{(1)}_{u,ij} - R_{ij})\xi_i\xi_j'\| = \|\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U,|R_{ij}|>b\tau_{ij}}(s_{ij}(R_{ij}) - R_{ij})\xi_i\xi_j' \\ &+\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U,|R_{ij}|>b\tau_{ij}}(s_{ij}(R_{ij}) - R_{ij})\xi_i\xi_j'\| \leq O_p(\omega_T^2) + \|\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U,|R_{ij}|\leq b\tau_{ij}}(s_{ij}(R_{ij}) - R_{ij})\xi_i\xi_j'\|.\end{aligned}$$

Write $v = \|N^{-1}\sum_{i\neq j,(i,j)\in S_U,|R_{ij}|\leq b\tau_{ij}}(s_{ij}(R_{ij}) - R_{ij})\xi_i\xi_j'\|$, then for any $C > 0$, and $\epsilon > 0$, Lemma B.4 implies $P(v > M\omega_T^2) \leq P(\exists(i,j) \in S_U, |R_{ij}| \leq b\tau_{ij}) < \epsilon$, which yields $v = O_p(\omega_T^2)$. Therefore $\frac{1}{N}\sum_{i\neq j,(i,j)\in S_U}(\widehat{\Sigma}^{(1)}_{u,ij} - \Sigma_{u0,ij})\xi_i\xi_j' = O_p(\omega_T^2)$. This implies $N^{-1}\Lambda_0'((\widehat{\Sigma}^{(1)}_u)^{-1} - \Sigma_{u0}^{-1})\Lambda_0 = O_p(\omega_T^2 + \omega_T^{2-2q}m_N^2 + m_N/N) = O_p(\omega_T^{2-2q}m_N^2)$.

□

B.4.2 Convergence rate for J

We now improve the rate in Lemma B.3.

Lemma B.11. *(i)* $H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}^{(1)}_u)^{-1}[\Lambda_0\frac{1}{T}\sum_{t=1}^T f_tu_t' + (\Lambda_0\frac{1}{T}\sum_{t=1}^T f_tu_t')'](\widehat{\Sigma}^{(1)}_u)^{-1}\widehat{\Lambda}^{(1)}H = O_p(m_NT^{-1/2}(\log N)^{1/2}\omega_T^{1-q})$.
(ii) $H\widehat{\Lambda}^{(1)\prime}\widehat{\Sigma}_u^{-1}(S_u - \widehat{\Sigma}^{(1)}_u)(\widehat{\Sigma}^{(1)}_u)^{-1}\widehat{\Lambda}^{(1)}H = O_p(m_N\omega_T^{1-q}T^{-1/2}(\log N)^{1/2} + m_N\omega_T^{1-q}N^{-1})$.

Proof. (i) By Theorem 3.2,

$$\|\widehat{\Lambda}^{(1)} - \Lambda_0\|_F = O_p(\sqrt{N}m_N\omega_T^{1-q}) = \|\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}^{(1)}_u)^{-1} - \Lambda_0'\Sigma_{u0}^{-1}\|_F. \tag{B.17}$$

Therefore the RHS of part (i) equals

$$H\Lambda_0'\Sigma_{u0}^{-1}\left[\Lambda_0\frac{1}{T}\sum_{t=1}^T f_tu_t' + (\Lambda_0\frac{1}{T}\sum_{t=1}^T f_tu_t')'\right]\Sigma_{u0}^{-1}\Lambda_0H + O_p(m_N\sqrt{\frac{\log N}{T}}\omega_T^{1-q}). \tag{B.18}$$

Now it follows from Assumption 3.10 that

$$\frac{1}{NT}\sum_{t=1}^T f_tu_t'\Sigma_{u0}^{-1}\Lambda_0 = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N f_tu_{it}\xi_i' = O_p(\frac{1}{\sqrt{NT}}) = O_p(m_N\sqrt{\frac{\log N}{T}}\omega_T^{1-q}),$$

which then yields the desired result.

(ii) Recall that $\|S_u - \widehat{\Sigma}_u^{(1)}\| = O_p(NT^{-1/2}(\log N)^{1/2} + m_N\omega_T^{1-q})$ and that $\|\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1} - \Lambda_0'\Sigma_{u0}^{-1}\|_F = O_p(\sqrt{N}m_N\omega_T^{1-q})$. By Theorem B.1, the RHS of (ii) equals

$$H\Lambda_0'\Sigma_{u0}^{-1}(S_u - \Sigma_{u0})\Sigma_{u0}^{-1}\Lambda_0 H + O_p(m_N\omega_T^{1-q}\sqrt{\frac{\log N}{T}} + \frac{m_N\omega_T^{1-q}}{N}).$$

By Assumption 3.10 (note that $H = O_p(N^{-1})$),

$$H\Lambda_0'\Sigma_{u0}^{-1}(S_u - \Sigma_{u0})\Sigma_{u0}^{-1}\Lambda_0 H = \frac{1}{T}H\sum_{i\leq N, j\leq N}\sum_{t\leq T}(u_{it}u_{jt} - Eu_{it}u_{jt})\xi_i\xi_j'H = O_p(\frac{1}{\sqrt{NT}}).$$

$\square$

Lemma B.12. $J = O_p(m_N^2\omega_T^{2-2q})$.

Proof. By (B.3) and Lemma B.11, ignoring the smaller order $J'J$, we have

$$J + J' = O_p(m_N\omega_T^{1-q}\sqrt{\frac{\log N}{T}} + \frac{m_N\omega_T^{1-q}}{N}).$$

This implies that $J_{ii} = O_p(m_N\omega_T^{1-q}(T^{-1/2}(\log N)^{1/2} + N^{-1}))$.

Moreover, since $H(\widehat{\Lambda}^{(1)} - \Lambda_0)'(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Lambda}^{(1)} - \Lambda_0)H = O_p(N^{-1}m_N^2\omega_T^{2-2q})$, (B.5) and Theorem 3.4 imply $\mathrm{ndg}\{HJ + J'H\} = O_p(N^{-1}m_N^2\omega_T^{2-2q})$. Therefore for $i \neq j$, $J_{ij} = O_p(m_N^2\omega_T^{2-2q} + m_N\omega_T^{1-q}(T^{-1/2}(\log N)^{1/2} + N^{-1})) = O_p(m_N^2\omega_T^{2-2q})$. The desired result follows immediately. $\square$

B.4.3 Improved rate for $\widehat{\lambda}_j^{(1)}$

Lemma B.13. *(i)* $H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}T^{-1}\sum_{t=1}^T(u_tu_{jt} - \widehat{\Sigma}_{u,j}^{(1)}) = O_p(m_N\omega_T^{2-q} + m_N^2\omega_T^{2-2q}N^{-1/2})$.
(ii) $H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}T^{-1}\sum_{t=1}^T u_tf_t'\lambda_{0j} = O_p(m_N\omega_T^{1-q}T^{-1/2}(\log N)^{1/2})$.

Proof. (i) We have, $\|\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1} - \Lambda_0'\Sigma_{u0}^{-1}\|_F = O_p(\sqrt{N}m_N\omega_T^{1-q})$. Hence $H(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1} - \Lambda_0'\Sigma_{u0}^{-1})T^{-1}\sum_{t=1}^T(u_tu_{jt} - \widehat{\Sigma}_{u,j}^{(1)}) = O_p(m_N\omega_T^{1-q}T^{-1/2}(\log N)^{1/2} + N^{-1/2}m_N^2\omega_T^{2-2-q})$. Hence part (i) equals

$$H\Lambda_0'\Sigma_{u0}^{-1}T^{-1}\sum_{t=1}^T(u_tu_{jt} - E(u_tu_{jt})) + O_p(m_N\omega_T^{2-q} + \frac{m_N^2\omega_T^{2-2q}}{\sqrt{N}})$$

where $O_p(\cdot)$ is uniform in $j \leq N$. By Assumption 3.10, for each $j \leq N$,

$$H\Lambda_0'\Sigma_{u0}^{-1}T^{-1}\sum_{t=1}^T(u_tu_{jt} - E(u_tu_{jt})) = H\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T\xi_i(u_{it}u_{jt} - E(u_tu_{jt})) = O_p((NT)^{-1/2}).$$

(ii) We have $\|H(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}-\Lambda_0'\Sigma_{u0}^{-1})T^{-1}\sum_{t=1}^T u_tf_t'\|_F=O_p(m_N\omega_T^{1-q}T^{-1/2}(\log N)^{1/2})$. Hence (ii) equals

$$H\Lambda_0'\Sigma_{u0}^{-1}\frac{1}{T}\sum_{t=1}^T u_tf_t'\lambda_{0j}+O_p(m_N\omega_T^{1-q}\sqrt{\frac{\log N}{T}}).$$

By Assumption 3.10, the first term equals $NH(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\xi_iu_{it}f_t'\lambda_{0j}=O_p((NT)^{-1/2})$, which yields the desired result. □

Lemma B.14. *For each fixed $j\le N$,*

$$\widehat{\lambda}_j^{(1)}-\lambda_{0j}=H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\Lambda_0\frac{1}{T}\sum_{t=1}^T f_tu_{jt}+O_p(m_N^2\omega_T^{2-2q}).$$

Proof. Note that those two terms in Lemma B.13 (i) (ii) are dominated by $O_p(m_N^2\omega_T^{2-2q})$. Therefore, the desired expansion follows from the first order condition (B.2) and Lemma B.12. □

B.4.4 Proof of Theorem 3.5

By Lemma B.14, and (B.17)

$$\begin{aligned}\widehat{\lambda}_j^{(1)}-\lambda_{0j} &= H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(\Lambda_0-\widehat{\Lambda}^{(1)})\frac{1}{T}\sum_{t=1}^T f_tu_{jt}+\frac{1}{T}\sum_{t=1}^T f_tu_{jt}+O_p(m_N^2\omega_T^{2-2q})\\ &= \frac{1}{T}\sum_{t=1}^T f_tu_{jt}+O_p(m_N^2\omega_T^{2-2q}+m_N\omega_T^{1-q}\sqrt{\frac{\log N}{T}})=\frac{1}{T}\sum_{t=1}^T f_tu_{jt}+O_p(m_N^2\omega_T^{2-2q}).\end{aligned}$$

By the assumption that $m_N^2\omega_T^{2-2q}=o(T^{-1/2})$, we have $\sqrt{T}(\widehat{\lambda}_j^{(1)}-\lambda_{0j})=T^{-1/2}\sum_{t=1}^T f_tu_{jt}+o_p(1)$. The limiting distribution follows since

$$T^{-1/2}\sum_{t=1}^T f_tu_{jt}\to^d N_r(0,E(u_{jt}f_tf_t')).$$

B.5 Proof of Theorem 3.6

For any $t\le T$, $y_t-\bar{y}=\Lambda_0f_t+u_t-\bar{u}$. Hence

$$\widehat{f}_t^{(1)}-f_t=-J'f_t+(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(u_t-\bar{u}).\tag{B.19}$$

B.5.1 Convergence rate

Since both f_t and u_t have exponential tails, using Bonferroni's method we have, $\max_t \|f_t\| = O_p((\log T)^{1/r_2})$ and $\max_t \|u_t\| = O_p(\sqrt{N}(\log T)^{1/r_1})$. Thus by Lemma B.12, $\max_{t\leq T}\|J'f_t\| = O_p(m_N^2\omega_T^{2-2q}(\log T)^{1/r_2})$. The term with $\bar{u}$ in (B.19) is of smaller order hence is negligible. Also $\|(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1} - \Lambda_0'\Sigma_{u0}^{-1})\|_F = O_p(N^{-1/2}m_N\omega_T^{1-q})$, where we used $\|\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1} - \Lambda_0'\Sigma_{u0}^{-1}\|_F = O_p(\sqrt{N}m_N\omega_T^{1-q})$. Hence

$$\max_{t\leq T}(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(u_t-\bar{u}) = O_p(\frac{1}{N})\Lambda_0'\Sigma_{u0}^{-1}u_t$$

$$+O_p(m_N^2\omega_T^{2-2q}(\log T)^{1/r_2}+m_N\omega_T^{1-q}(\log T)^{1/r_1}) = O_p(\frac{1}{N})\Lambda_0'\Sigma_{u0}^{-1}u_t+O_p(m_N\omega_T^{1-q}(\log T)^{1/r_1+1/r_2}).$$

Finally, because $E(\frac{1}{N}\Lambda_0'\Sigma_{u0}^{-1}u_tu_t'\Sigma_{u0}^{-1}\Lambda_0) = \frac{1}{N}\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0$, whose eigenvalues are bounded. Hence $\frac{1}{\sqrt{N}}\Lambda_0'\Sigma_{u0}^{-1}u_t = O_p(1)$. Also, $O_p(N^{-1/2})$ is of smaller order than $O_p(m_N\omega_T^{1-q}(\log T)^{1/r_1+1/r_2+1})$. This implies

$$\|\widehat{f}_t^{(1)} - f_t\| = O_p(m_N\omega_T^{1-q}(\log T)^{1/r_1+1/r_2+1}).$$

The above proof also shows that the rate can be made uniform if $\max_{t\leq T}\|\frac{1}{\sqrt{N}}\Lambda_0'\Sigma_{u0}^{-1}u_t\| = O_p(\log T)$.

B.5.2 Asymptotic normality

Recall that $\Xi = \Lambda_0'\Sigma_{u0}^{-1}$ and $\beta_t = \Sigma_{u0}^{-1}u_t$.

Lemma B.15. *For any fixed $t \leq T$, $N^{-1/2}(\widehat{\Lambda}^{(1)} - \Lambda_0)'\Sigma_{u0}^{-1}u_t = o_p(1)$.*

Proof. We expand $\widehat{\Lambda}^{(1)} - \Lambda_0$ using the first order condition

$$(\widehat{\Lambda}^{(1)} - \Lambda_0)' = J\Lambda_0' + H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}[\Lambda_0\frac{1}{T}\sum_{s=1}^T f_su_s' + \frac{1}{T}\sum_{s=1}^T u_sf_s'\Lambda_0' + S_u - \widehat{\Sigma}_u^{(1)}] \qquad \text{(B.20)}$$

and investigate each term separately. First of all, since $J = O_p(m_N^2\omega_T^{2-2q})$, and by assumption that $\Lambda_0'\Sigma_{u0}^{-1}u_t = \sum_{i=1}^N \xi_iu_{it} = O_p(\sqrt{N})$, we have $N^{-1/2}J\Lambda_0'\Sigma_{u0}^{-1}u_t = O_p(m_N^2\omega_T^{2-2q})$. Second, by the assumption that $(TN)^{-1/2}\sum_{s=1}^T f_su_s'\Sigma_{u0}^{-1}u_t = O_p(1)$, we have

$$\frac{1}{\sqrt{N}}H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\Lambda_0\frac{1}{T}\sum_{s=1}^T f_su_s'\Sigma_{u0}^{-1}u_t = O_p(\frac{1}{\sqrt{T}}).$$

Third, $N^{-1/2}H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\frac{1}{T}\sum_{s=1}^T u_sf_s'\Lambda_0'\Sigma_{u0}^{-1}u_t = O_p(\sqrt{\log N/T})$. Moreover,

$N^{-1/2}H(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}-\Lambda_0'\Sigma_{u0}^{-1})(S_u-\Sigma_{u0})\Sigma_{u0}^{-1}u_t = O_p(m_N\omega_T^{1-q}\sqrt{N\log N/T}) = o_p(1)$. Therefore, by the assumption that $(NT\sqrt{N})^{-1}\sum_{i=1}^N\sum_{s=1}^T \xi_i(u_{is}u_s'-Eu_{is}u_s')\beta_t = o_p(1)$, we have,

$$\frac{1}{\sqrt{N}}H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(S_u-\Sigma_{u0})\Sigma_{u0}^{-1}u_t = \frac{1}{\sqrt{N}}H\Lambda_0'\Sigma_{u0}^{-1}(S_u-\Sigma_{u0})\Sigma_{u0}^{-1}u_t + o_p(1)$$
$$= \frac{1}{T\sqrt{N}}H\sum_{i=1}^N\sum_{j=1}^N\sum_{s=1}^T \xi_i(u_{is}u_{js}-Eu_{is}u_{js})\beta_{jt} = o_p(1).$$

Finally, $N^{-1/2}H\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}(\widehat{\Sigma}_u^{(1)}-\Sigma_{u0})\Sigma_{u0}^{-1}u_t = O_p(\frac{1}{\sqrt{N}}m_N\omega_N^{1-q})$. □

Lemma B.16. *For any fixed $t\le T$, $N^{-1/2}\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1}-\Sigma_{u0}^{-1})u_t = o_p(1)$*

Proof. We note that, $N^{-1/2}\Lambda_0'((\widehat{\Sigma}_u^{(1)})^{-1}-\Sigma_{u0}^{-1})u_t = N^{-1/2}\Xi(\widehat{\Sigma}_u^{(1)}-\Sigma_{u0})\beta_t + O_p(\sqrt{N}m_N^2\omega_T^{2-2q})$. On the other hand,

$$\begin{aligned}\frac{1}{\sqrt{N}}\Xi(\widehat{\Sigma}_u^{(1)}-\Sigma_{u0})\beta_t &= \frac{1}{\sqrt{N}}\sum_{i=1}^N(R_{ii}-\Sigma_{u0,ii})\xi_i\beta_{it} + \frac{1}{\sqrt{N}}\sum_{i\neq j,(i,j)\in S_U}(\widehat{\Sigma}_{u,ij}^{(1)}-\Sigma_{u0,ij})\xi_i\beta_{jt}\\ &\quad+\frac{1}{\sqrt{N}}\sum_{(i,j)\in S_L}(\widehat{\Sigma}_{u,ij}^{(1)}-\Sigma_{u0,ij})\xi_i\beta_{jt}.\end{aligned}$$

The result of the proof is very similar to that of Lemmas B.10 and Theorem l3.1, based on the expansion (B.10) and Theorem 3.3, hence is omitted. □

Proof of asymptotic normality

We now fix t, then Lemma B.12 gives $J'f_t = O_p(m_N^2\omega_T^{2-2q})$. Hence $\sqrt{N}J'f_t$ is negligible as $\sqrt{N}m_N^2\omega_T^{2-2q} = o(1)$. Moreover, $(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\bar{u}$ is of smaller order of $(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}u_t$, hence is negligible. Next,

$$\sqrt{N}(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)})^{-1}\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}u_t = \sqrt{N}(\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0)^{-1}\Lambda_0'\Sigma_{u0}^{-1}u_t$$

$$+O_p(N^{-1/2})(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}-\Lambda_0'\Sigma_{u0}^{-1})u_t + O_p(m_N\omega_T^{1-q})$$

where we used $(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}\widehat{\Lambda}^{(1)\prime})^{-1}-(\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0)^{-1} = O_p(N^{-1}m_N\omega_T^{1-q})$. By Lemmas B.15 and B.16, $N^{-1/2}(\widehat{\Lambda}^{(1)\prime}(\widehat{\Sigma}_u^{(1)})^{-1}-\Lambda_0'\Sigma_{u0}^{-1})u_t = o_p(1)$. This implies, for each fixed t,

$$\begin{aligned}\sqrt{N}(\widehat{f}_t^{(1)}-f_t) &= \sqrt{N}(\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0)^{-1}\Lambda_0'\Sigma_{u0}^{-1}u_t + O_p(\sqrt{N}m_N^2\omega_T^{2-2q}+m_N\omega_T^{1-q})\\ &= \sqrt{N}(\Lambda_0'\Sigma_{u0}^{-1}\Lambda_0)^{-1}\Lambda_0'\Sigma_{u0}^{-1}u_t + o_p(1).\end{aligned}$$

The asymptotic normality then follows from the fact that

$$N^{-1/2}\Lambda_0'\Sigma_{u0}^{-1}u_t=\frac{1}{\sqrt{N}}\sum_{i=1}^N\xi_iu_{it}\to^d N(0,Q).$$

C Proofs of Section 4

C.1 Proof of Theorem 4.1

Define

$$\begin{aligned}Q_1(\Sigma_u) &= \frac{1}{N}\log|\Sigma_u|+\frac{1}{N}\mathrm{tr}(S_u\Sigma_u^{-1})+\frac{\mu_T}{N}\sum_{i\neq j}w_{ij}|\Sigma_{u,ij}| \\ &\quad -\frac{1}{N}\log|\Sigma_{u0}|-\frac{1}{N}\mathrm{tr}(S_u\Sigma_{u0}^{-1})-\frac{\mu_T}{N}\sum_{i\neq j}w_{ij}|\Sigma_{u0,ij}|,\end{aligned}$$

Let $L_c(\Lambda,\Sigma_u)=L_2(\Lambda,\Sigma_u)-N^{-1}\log|\Sigma_{u0}|-N^{-1}\mathrm{tr}(S_u\Sigma_{u0}^{-1})-N^{-1}\mu_T\sum_{i\neq j}w_{ij}|\Sigma_{u0,ij}|$. Then the minimizer of L_c is the same as that of L_2. This implies $L_c(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})\leq L_c(\Lambda_0,\Sigma_{u0})$. Recall the definitions of $Q_2(\Lambda,\Sigma_u)$ and $Q_3(\Lambda,\Sigma_u)$. Then

$$L_c(\Lambda,\Sigma_u)=Q_1(\Sigma_u)+Q_2(\Lambda,\Sigma_u)+Q_3(\Lambda,\Sigma_u).$$

Lemma C.1. *There is a nonnegative stochastic sequence $0\leq d_T=O_p(N^{-1}\log N+T^{-1/2}(\log N)^{1/2})$ such that $Q_1(\widehat{\Sigma}_u^{(2)})\leq d_T$ with probability one.*

Proof. We have $Q_2(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})\geq 0$. In addition, $Q_2(\Lambda_0,\Sigma_{u0})=Q_1(\Sigma_{u0})=0$. Hence

$$\begin{aligned}Q_1(\widehat{\Sigma}_u^{(2)}) &= L_c(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})-Q_2(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})-Q_3(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)}) \\ &\leq L_c(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})-Q_3(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})\leq L_c(\Lambda_0,\Sigma_{u0})-Q_3(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)}) \\ &= Q_3(\Lambda_0,\Sigma_{u0})-Q_3(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)}).\end{aligned}$$

By the definition of $\Theta_\lambda\times\Gamma$, there is $\delta>0$ such that $\Theta_\lambda\times\Gamma\subset\Xi_\delta$. The result then holds for $d_T=|Q_3(\Lambda_0,\Sigma_{u0})|+|Q_3(\widehat{\Lambda}^{(2)},\widehat{\Sigma}_u^{(2)})|$ by Lemma A.2.

□

Throughout, let (recall that $D=\sum_{i\neq j,(i,j)\in S_U}1$.)

$$\Delta=(\widehat{\Sigma}_u^{(2)})^{-1}-\Sigma_{u0}^{-1},\quad K_T=\sum_{(i,j)\in S_L}|\Sigma_{u0,ij}|.$$

Lemma C.2. *For all large enough T and N,*

$$\begin{aligned} NQ_1(\widehat{\Sigma}_u^{(2)}) &\geq \frac{1}{2}\mu_T \min_{(i,j)\in S_L} w_{ij} \sum_{(i,j)\in S_L} |\widehat{\Sigma}_{u,ij} - \Sigma_{u0,ij}| + c\|\Delta\|_F^2 - 2\mu_T \max_{(i,j)\in S_L} w_{ij} K_T \\ &\quad - \left(O_p(\sqrt{\frac{\log N}{T}})\sqrt{N+D} + \mu_T \max_{i\neq j,(i,j)\in S_U} w_{ij}\sqrt{D} \right) \|\Delta\|_F. \end{aligned}$$

Proof. Let $\Omega_0 = \Sigma_{u0}^{-1}$, $\widehat{\Omega} = (\widehat{\Sigma}_u^{(2)})^{-1}$. For any Σ_u, let $\Omega = \Sigma_u^{-1}$. Define a function $f(t) = -\log|\Omega_0 + t\Delta| + \mathrm{tr}(S_u(\Omega_0 + t\Delta))$, $t \geq 0$. Then $-\log|\widehat{\Omega}| + \mathrm{tr}(S_u\widehat{\Omega}) = f(1)$; $-\log|\Omega_0| + \mathrm{tr}(S_u\Omega_0) = f(0)$; and

$$NQ_1(\widehat{\Sigma}_u^{(2)}) = f(1) - f(0) + \mu_T \sum_{i\neq j} w_{ij}|\widehat{\Sigma}_{u,ij}| - \mu_T \sum_{i\neq j} w_{ij}|\Sigma_{u0,ij}| \qquad \text{(C.1)}$$

By the integral remainder Taylor expansion, $f(1) - f(0) = f'(0) + \int_0^1 (1-t)f''(t)dt$. We now calculate $f'(0)$ and $f''(t)$. Using the matrix differentiation formula, we have, $f'(t) = \mathrm{tr}(S_u\Delta) - \mathrm{tr}((\Omega_0 + t\Delta)^{-1}\Delta)$, which implies,

$$\begin{aligned} f'(0) &= \mathrm{tr}((S_u - \Sigma_{u0})(\widehat{\Omega} - \Omega_0)) = \mathrm{tr}(\Omega_0(S_u - \Sigma_{u0})\widehat{\Omega}(\Sigma_{u0} - \widehat{\Sigma}_u^{(2)})) \\ &= \sum_{ij} (\Omega_0(S_u - \Sigma_{u0})\widehat{\Omega})_{ij}(\Sigma_{u0} - \widehat{\Sigma}_u^{(2)})_{ij}. \end{aligned}$$

Note that both $\|\Omega_0\|_1$ and $\|\widehat{\Omega}\|_1$ are bounded from above for $\Sigma_{u0}, \widehat{\Sigma}_u^{(2)} \in \Gamma$. By Lemma A.1(ii), $\max_{ij}|(\Omega_0(S_u - \Sigma_{u0})\widehat{\Omega})_{ij}| \leq \max_{ij}|(S_u - \Sigma_{u0})_{ij}|\|\Omega_0\|_1\|\widehat{\Omega}\|_1 = O_p(\sqrt{\log N/T})$. Therefore, $|f'(0)| = O_p(\sqrt{\log N/T})\sum_{ij}|\Sigma_{u0,ij} - \widehat{\Sigma}_{u,ij}|$. In addition,

$$f''(t) = \mathrm{tr}((\Omega_0 + t\Delta)^{-1}\Delta(\Omega_0 + t\Delta)^{-1}\Delta) = vec(\Delta)(\Omega_0 + t\Delta)^{-1} \otimes (\Omega_0 + t\Delta)^{-1}vec(\Delta),$$

where vec dentoes the vectorization operator and $\otimes$ denotes the Kronecker product. Since both $(\widehat{\Lambda}^{(2)}, \widehat{\Sigma}_u^{(2)})$ and (Λ_0, Σ_{u0}) are inside $\Theta_\lambda \times \Gamma$, $\sup_{0\leq t\leq 1}\lambda_{\max}(t(\widehat{\Sigma}_u^{(2)})^{-1} + (1-t)\Sigma_{u0}^{-1})$ is bounded from above, which then implies $\inf_{0\leq t\leq 1}\lambda_{\min}[(\Omega_0 + t\Delta)^{-1}] = \inf_{0\leq t\leq 1}\lambda_{\max}^{-1}(t(\widehat{\Sigma}_u^{(2)})^{-1} + (1-t)\Sigma_{u0}^{-1})$ is bounded below by a positive constant c. Hence $\inf_{0\leq t\leq 1} f''(t) \geq c\|\Delta\|_F^2$. From (C.1) and $f(1) - f(0) \geq -|f'(0)| + c\|\Delta\|_F^2$, we have

$$\begin{aligned} NQ_1(\widehat{\Sigma}_u^{(2)}) &\geq \mu_T \sum_{i\neq j} w_{ij}|\widehat{\Sigma}_{u,ij}| - \mu_T \sum_{i\neq j} w_{ij}|\Sigma_{u0,ij}| + c\|\Delta\|_F^2 - O_p(\sqrt{\frac{\log N}{T}})\sum_{ij}|\Sigma_{u0,ij} - \widehat{\Sigma}_{u,ij}| \\ &= \mu_T \sum_{(i,j)\in S_L} w_{ij}|\widehat{\Sigma}_{u,ij}| + \mu_T \sum_{i\neq j,(i,j)\in S_U} w_{ij}|\widehat{\Sigma}_{u,ij}| - \mu_T \sum_{i\neq j} w_{ij}|\Sigma_{u0,ij}| + c\|\Delta\|_F^2 \end{aligned}$$

$$-O_p(\sqrt{\frac{\log N}{T}})\sum_{\Sigma_{u0,ij}\in S_U}|\Sigma_{u0,ij}-\widehat{\Sigma}_{u,ij}|-O_p(\sqrt{\frac{\log N}{T}})\sum_{(i,j)\in S_L}|\Sigma_{u0,ij}-\widehat{\Sigma}_{u,ij}|.$$

Since $|\widehat{\Sigma}_{u,ij}|\geq|\widehat{\Sigma}_{u,ij}-\Sigma_{u0,ij}|-|\Sigma_{u0,ij}|$, and $\sum_{i\neq j}w_{ij}|\Sigma_{u0,ij}|=\sum_{i\neq j,(i,j)\in S_U}w_{ij}|\Sigma_{u0,ij}|+\sum_{(i,j)\in S_L}w_{ij}|\Sigma_{u0,ij}|$. It follows that

$$\begin{aligned}
NQ_1(\widehat{\Sigma}_u^{(2)}) \geq\ & \mu_T\sum_{(i,j)\in S_L}w_{ij}|\widehat{\Sigma}_{u,ij}-\Sigma_{u0,ij}|-O_p(\sqrt{\frac{\log N}{T}})\sum_{(i,j)\in S_L}|\Sigma_{u0,ij}-\widehat{\Sigma}_{u,ij}|+c\|\Delta\|_F^2\\
&-\mu_T\sum_{(i,j)\in S_L}w_{ij}|\Sigma_{u0,ij}|-O_p(\sqrt{\frac{\log N}{T}})\sum_{\Sigma_{u0,ij}\in S_U}|\Sigma_{u0,ij}-\widehat{\Sigma}_{u,ij}|\\
&-\mu_T\sum_{i\neq j,(i,j)\in S_U}w_{ij}[|\Sigma_{u0,ij}|-|\widehat{\Sigma}_{u,ij}|]-\mu_T\sum_{(i,j)\in S_L}w_{ij}|\Sigma_{u0,ij}|\\
\geq\ & (\mu_T\min_{(i,j)\in S_L}w_{ij}-O_p(\sqrt{\frac{\log N}{T}}))\sum_{(i,j)\in S_L}|\widehat{\Sigma}_{u,ij}-\Sigma_{u0,ij}|+c\|\Delta\|_F^2\\
&-2\mu_T\sum_{(i,j)\in S_L}w_{ij}|\Sigma_{u0,ij}|-O_p(\sqrt{\frac{\log N}{T}})\sum_{\Sigma_{u0,ij}\in S_U}|\Sigma_{u0,ij}-\widehat{\Sigma}_{u,ij}|\\
&-\mu_T\max_{i\neq j,(i,j)\in S_U}w_{ij}\sum_{i\neq j,(i,j)\in S_U}|\Sigma_{u0,ij}-\widehat{\Sigma}_{u,ij}|\\
\geq\ & \frac{1}{2}\mu_T\min_{(i,j)\in S_L}w_{ij}\sum_{(i,j)\in S_L}|\widehat{\Sigma}_{u,ij}-\Sigma_{u0,ij}|+c\|\Delta\|_F^2-2\mu_T\max_{(i,j)\in S_L}w_{ij}K_T\\
&-O_p(\sqrt{\frac{\log N}{T}})\sqrt{N+D}\|\Delta\|_F-\mu_T\max_{i\neq j,(i,j)\in S_U}w_{ij}\|\Delta\|_F\sqrt{D},
\end{aligned}$$

which implies the desired result. □

Lemma C.3.

$$\begin{aligned}
\frac{1}{N}\|\Sigma_u-\widehat{\Sigma}_u^{(2)}\|_F^2 =\ & O_p\left(\frac{1}{N}\left(\mu_T\max_{(i,j)\in S_L}w_{ij}K_T+\log N+\mu_T^2\max_{i\neq j,(i,j)\in S_U}w_{ij}^2D\right)\right)\\
&+O_p(\frac{D\log N}{NT}+\sqrt{\frac{\log N}{T}}).
\end{aligned}$$

Proof. Lemma C.2 implies

$$NQ_1(\widehat{\Sigma}_u^{(2)})\geq c\|\Delta\|_F^2-2\mu_T\max_{(i,j)\in S_L}w_{ij}K_T-\left(O_p(\sqrt{\frac{\log N}{T}})\sqrt{N+D}+\mu_T\max_{i\neq j,(i,j)\in S_U}w_{ij}\sqrt{D}\right)\|\Delta\|_F.$$

Lemma C.1 gives $NQ_1(\widehat{\Sigma}_u^{(2)}) \leq O_p(\log N + N\sqrt{\log N/T})$. Hence we have

$$
\begin{aligned}
\|\Delta\|_F^2 &= O_p((\sqrt{\frac{(N+D)\log N}{T}} + \mu_T \max_{i\neq j,(i,j)\in S_U} w_{ij}\sqrt{D})^2) \\
&\quad + O_p(\mu_T \max_{(i,j)\in S_L} w_{ij}K_T + \log N + N\sqrt{\log N/T}) \\
&= O_p(\frac{(N+D)\log N}{T} + \mu_T^2 \max_{i\neq j,(i,j)\in S_U} w_{ij}^2 D + \mu_T \max_{(i,j)\in S_L} w_{ij}K_T + \log N + N\sqrt{\log N/T}) \\
&= O_p(\frac{D\log N}{T} + \mu_T^2 \max_{i\neq j,(i,j)\in S_U} w_{ij}^2 D + \mu_T \max_{(i,j)\in S_L} w_{ij}K_T + \log N + N\sqrt{\log N/T}).
\end{aligned}
$$

Note that $\Sigma_{u0} - \widehat{\Sigma}_u^{(2)} = \widehat{\Sigma}_u^{(2)}\Delta\Sigma_{u0}$. Hence the desired result follows from $\|\widehat{\Sigma}_u^{(2)}\| < M$ wp1 and $\|\Sigma_{u0}\| < M$.

□

Lemma C.4. $N^{-1}\sum_{(i,j)\in S_L}|\widehat{\Sigma}_{u,ij} - \Sigma_{u0,ij}| = o_p(1)$.

Proof. Lemma C.2 implies

$$
\begin{aligned}
&\frac{1}{2}\mu_T \min_{(i,j)\in S_L} w_{ij} \sum_{(i,j)\in S_L} |\widehat{\Sigma}_{u,ij} - \Sigma_{u0,ij}| \leq NQ_1(\widehat{\Sigma}_u^{(2)}) + 2\mu_T \max_{(i,j)\in S_L} w_{ij}K_T \\
&+ \left(O_p(\sqrt{\frac{\log N}{T}})\sqrt{N+D} + \mu_T \max_{i\neq j,(i,j)\in S_U} w_{ij}\sqrt{D}\right)\|\Delta\|_F.
\end{aligned}
$$

We have $NQ_1(\widehat{\Sigma}_u^{(2)}) \leq O_p(\log N + N\sqrt{\log N/T})$. By Lemma C.3,

$$
\begin{aligned}
\|\Delta\|_F &= O_p(\sqrt{\frac{D\log N}{T}} + \mu_T \max_{i\neq j,(i,j)\in S_U} w_{ij}\sqrt{D}) \\
&\quad + O_p(\sqrt{\mu_T \max_{(i,j)\in S_L} w_{ij}K_T} + \sqrt{\log N} + \sqrt{N}(\frac{\log N}{T})^{1/4}).
\end{aligned}
$$

which implies the desired result under Assumption 4.2.

□

Lemma C.5. $N^{-1}\Lambda_0'((\widehat{\Sigma}_u^{(2)})^{-1} - \Sigma_{u0}^{-1})\Lambda_0 = o_p(1)$.

Proof. Let $\Delta_1 = \widehat{\Sigma}_u^{(2)} - \Sigma_{u0}$, $\Xi = \Lambda_0'\Sigma_{u0}^{-1} = (\xi_1, ..., \xi_N)$, and $\widehat{V} = (\widehat{\Sigma}_u^{(2)})^{-1}\Lambda_0$. Since the l_1 norms of $(\widehat{\Sigma}_u^{(2)})^{-1}$ and Σ_{u0}^{-1} are bounded away from infinity, we have, $\sup_{i\leq N}\|\widehat{V}_i\| = O_p(1)$ and $\sup_{i\leq N}\|\xi_i\| = O(1)$. Then

$$
\frac{1}{N}\Lambda_0'(\Sigma_{u0}^{-1} - (\widehat{\Sigma}_u^{(2)})^{-1})\Lambda_0 = \frac{1}{N}\Xi\Delta_1\widehat{V} = \frac{1}{N}\sum_{(i,j)\in S_L}\xi_i\widehat{V}_j'\Delta_{1,ij} + \frac{1}{N}\sum_{\Sigma_{u0,ij}\in S_U}\xi_i\widehat{V}_j'\Delta_{1,ij}
$$

$$\leq O_p(\frac{1}{N}) \sum_{(i,j)\in S_L} |\Delta_{1,ij}| + O_p(\frac{1}{N}) \sum_{\Sigma_{u0,ij}\in S_U} |\Delta_{1,ij}|.$$

The first term on the right hand side is $o_p(1)$ by Lemma C.4, and the second is bounded by $N^{-1}\|\widehat{\Sigma}_u^{(2)} - \Sigma_{u0}\|\sqrt{N+D}$ (using Cauchy-Schwarz inequality), which is also $o_p(1)$ by Lemma C.3 and Assumption 4.2.

□

Lemma C.6. *For $(\widehat{\Lambda}, \widehat{\Sigma}) = (\widehat{\Lambda}^{(2)}, \widehat{\Sigma}_u^{(2)})$, Lemma A.3 is satisfied.*

Proof. We first show part (i) of Lemma A.3. Since $L_c(\widehat{\Lambda}^{(2)}, \widehat{\Sigma}_u^{(2)}) \leq L_c(\Lambda_0, \Sigma_{u0})$, and $Q_1(\Sigma_{u0}) = Q_2(\Lambda_0, \Sigma_{u0}) = 0$, there is a nonnegative sequence $d_n = O_p(N^{-1}\log N + T^{-1/2}(\log N)^{1/2})$ such that $Q_1(\widehat{\Sigma}_u^{(2)}) + Q_2(\widehat{\Lambda}^{(2)}, \widehat{\Sigma}_u^{(2)}) \leq d_n$. Lemma C.2 then implies $0 \leq Q_2(\widehat{\Sigma}_u^{(2)}, \widehat{\Lambda}^{(2)}) = o_p(1)$. On the other hand,

$$Q_2(\widehat{\Sigma}_u^{(2)}, \widehat{\Lambda}^{(2)}) = \frac{1}{N}\mathrm{tr}\left[\Lambda_0'(\widehat{\Sigma}_u^{(2)})^{-1}\Lambda_0 - \Lambda_0'(\widehat{\Sigma}_u^{(2)})^{-1}\widehat{\Lambda}^{(2)}(\widehat{\Lambda}'(\widehat{\Sigma}_u^{(2)})^{-1}\widehat{\Lambda}^{(2)})^{-1}\widehat{\Lambda}'(\widehat{\Sigma}_u^{(2)})^{-1}\Lambda_0\right].$$

The matrix in the bracket is semi-positive definite. Hence

$$\frac{1}{N}\Lambda_0'(\widehat{\Sigma}_u^{(2)})^{-1}\Lambda_0 - (I_r - J)\frac{1}{N}\widehat{\Lambda}'(\widehat{\Sigma}_u^{(2)})^{-1}\widehat{\Lambda}^{(2)}(I_r - J)' = o_p(1). \tag{C.2}$$

Finally, the desired result follows from Lemma C.5.

The first order condition in part (ii) is easy to derive and is the same as that in Bai and Li (2012).

□

Proof of Theorem 4.1

$N^{-1}\|\widehat{\Sigma}_u^{(2)} - \Sigma_{u0}\|_F^2 = o_p(1)$ follows from Lemma C.3 and Assumption 4.2. On the other hand, equation (C.2) also implies

$$\frac{1}{N}(\widehat{\Lambda}^{(2)} - \Lambda_0)'\widehat{\Sigma}_u^{-1}(\widehat{\Lambda}^{(2)} - \Lambda_0) - J\frac{1}{N}H^{-1}J' = o_p(1).$$

By Lemma A.5, $N^{-1}JH^{-1}J' = o_p(1)$. Hence $N^{-1}(\widehat{\Lambda}^{(2)} - \Lambda_0)'\widehat{\Sigma}_u^{-1}(\widehat{\Lambda} - \Lambda_0) = o_p(1)$, which implies the consistency $N^{-1}\|\widehat{\Lambda} - \Lambda_0\|^2 = o_p(1)$ because the eigenvalues of $\widehat{\Sigma}_u^{-1}$ are bounded away from zero. Q.E.D.

To prove the consistency of $\widehat{f}_t^{(2)}$, we note that the expansion (B.19) still holds for $\widehat{f}_t^{(2)}$. Since $J = o_p(1)$ by Lemma A.5, and $\bar{u}$ is of smaller order than u_t for each fixed t. Hence $\widehat{f}_t^{(2)} - f_t = O_p(N^{-1})\widehat{\Lambda}^{(2)\prime}(\widehat{\Sigma}_u^{(2)})^{-1}u_t + o_p(1)$. Moreover, since $\|(\widehat{\Sigma}_u^{(2)})^{-1}\|$ and $\|\widehat{\Sigma}_u^{(2)}\|$ are both $O_p(1)$ and $\|\widehat{\Lambda}^{(2)}\|_F = O_p(\sqrt{N})$ by the restriction of the parameter space $\Theta_\lambda \times \Gamma$, we have

$N^{-1}\|\widehat{\Lambda}^{(2)\prime}(\widehat{\Sigma}_u^{(2)})^{-1}-\Lambda_0\Sigma_{u0}^{-1}\|_F = O_p(N^{-1/2}\|\widehat{\Lambda}^{(2)}-\Lambda_0\|_F + N^{-1/2}\|\widehat{\Sigma}_u^{(2)}-\Sigma_{u0}\|_F)$, which is $o_p(1)$ as proved above. Therefore, since $N^{-1}\Lambda_0'\Sigma_{u0}^{-1}u_t = N^{-1}\sum_{i=1}^N \xi_i u_{it} = O_p(N^{-1/2})$,

$$\widehat{f}_t^{(2)} - f_t = O_p(N^{-1})\Lambda_0'\Sigma_{u0}^{-1}u_t + o_p(1) = o_p(1).$$

C.2 Proof of Theorem 4.2

We now verify Assumption 4.2 for the Adaptive Lasso.

Lemma C.7. *For adaptive lasso,*
(i) $\min_{i\neq j,(i,j)\in S_U}|\Sigma_{u0,ij}|^\gamma \max_{i\neq j,(i,j)\in S_U} w_{ij} = O_p(1)$.
(ii) $\delta_T^\gamma \max_{(i,j)\in S_L} w_{ij} = O_p(1)$,
(iii) $\omega_T^{-\gamma}(\min_{(i,j)\in S_L} w_{ij})^{-1} = O_p(1)$ *(recall that* $\omega_T = N^{-1/2}+T^{-1/2}(\log N)$*).*

Proof. By Lemma B.5 $\max_{i\leq N,j\leq N}|\widehat{\Sigma}^*_{u,ij}-\Sigma_{u0,ij}| = O_p(\tau)$. Given this result and the assumption that $\min_{(i,j)\in S_U}|\Sigma_{u0,ij}| \gg \omega_T$, we have result (i). For any $(i,j)\in S_L$, the following inequality holds: $\delta_T^{-\gamma} \leq w_{ij}^{-1} \leq (|\Sigma_{u0,ij}| + |\Sigma_{u0,ij}-\widehat{\Sigma}_{u,ij}| + \delta_T)^\gamma$, which then implies results (ii) and (iii), due to the assumptions that $\delta_T = o(\omega_T)$, and $\Sigma_{u0,ij} = O(\omega_T)$. □

Proof of Assumption 4.2 for Adaptive Lasso

It follows from the previous lemma that $\alpha_T = O_p(\omega_T^\gamma(\min_{i\neq j,\Sigma_{u0,ij}\in S_U}|\Sigma_{u0,ij}|)^{-\gamma}) = o_p(1)$, and $\beta_T = O_p((\omega_T/\delta_T)^\gamma)$. By the assumption that $D = O(N)$,

$$\zeta = \min\left\{\sqrt{\frac{T}{\log N}}\frac{N}{D}, \left(\frac{T}{\log N}\right)^{1/4}\sqrt{\frac{N}{D}}, \frac{N}{\sqrt{D\log N}}\right\} \gg \min\left\{\left(\frac{T}{\log N}\right)^{1/4}, \sqrt{\frac{N}{\log N}}\right\}.$$

Hence $\alpha_T = O_p(\zeta)$. This together with the lower bound assumption on δ_T yields Assumption 4.2 (i).

For part (ii), note that $\alpha_T = o_p(1)$ implies that with probability approaching one,

$$\min\{N, \frac{N^2}{D}, \frac{N^2}{D}\alpha_T^{-2}\} = N, \quad \min\{\frac{N}{D}, \sqrt{\frac{N}{D}}, \frac{N}{D}\alpha_T^{-1}\} = \sqrt{\frac{N}{D}}.$$

By Lemma C.7(ii), (recall that $K_T = \sum_{(i,j)\in S_L}|\Sigma_{u0,ij}|$) and the lower bound $\delta_T \gg \omega_T(K_T/N)^{1/\gamma}$, $\mu_T\max_{(i,j)\in S_L} w_{ij}K_T = O_p(\mu_T\delta_T^{-\gamma}K_T) = o_p(N)$.

By Lemma C.7(i) and the assumptions that $D = O(N)$ and $\min_{i\neq j,(i,j)\in S_U}|\Sigma_{u0,ij}| \gg \omega_T$, we have $\mu_T\max_{i\neq j,(i,j)\in S_U} w_{ij} = O_p\mu_T(\min_{i\neq j,(i,j)\in S_U}|\Sigma_{u0,ij}|^\gamma)^{-1} = o_p(\sqrt{N/D})$, due to the upper bound on $\mu_T = o(\omega_T^\gamma)$. Finally, by Lemma C.7(iii) and the assumption that $\mu_T \gg \omega_T^{1+\gamma}$, we have $\mu_T\min_{(i,j)\in S_L} w_{ij} \gg \omega_T$.

Proof of Assumption 4.2 for SCAD

Since $\mu_T / \min_{i \neq j, (i,j) \in S_U} |R_{ij}| = o_p(1)$ and $\max_{(i,j) \in S_L} |R_{ij}| = o_p(\mu_T)$, it is easy to verify that with probability approaching one, $\max_{i \neq j, (i,j) \in S_U} w_{ij} = 0$, $\min_{(i,j) \in S_L} w_{ij} = \max_{(i,j) \in S_L} w_{ij} = \mu_T$. Hence $\alpha_T = 0$ and $\beta_T = 1$. This immediately implies the desired result.