Maximum likelihood estimation and inference for approximate factor models of high dimension

Jushan Bai* and Kunpeng Li†

October 16, 2014

Abstract

An approximate factor model of high dimension has two key features. First, the idiosyncratic errors are correlated and heteroskedastic over both the cross-section and time dimensions; the correlations and heteroskedasticities are of unknown forms. Second, the number of variables is comparable or even greater than the sample size. Thus a large number of parameters exist under a high dimensional approximate factor model. Most widely used approaches to estimation are principal component based. This paper considers the maximum likelihood-based estimation of the model. Consistency, rate of convergence, and limiting distributions are obtained under various identification restrictions. Monte Carlo simulations show that the likelihood method is easy to implement and has good finite sample properties.

Key Words: Factor analysis; Approximate factor models; Maximum likelihood; Principal components; Inferential theory

JEL C33, C55

*Department of Economics, Columbia University, New York, NY, 10027
†International School of Economics and Management, Capital University of Economics and Business, Beijing, China.
We thanks the editor and two anonymous referees for their constructive comments. This paper was presented at Brown University, Yale University, and the 8th International Symposium on Econometric Theory and Applications (SETA 2012, Shanghai, China). We thank seminar and conference participates for helpful comments.
1 Introduction

Factor analysis is an essential tool in psychology. It is also fundamental in modern finance theory. The Arbitrage Pricing Theory (APT) of Ross (1976), for example, is built upon a multiple factor model for asset returns. Due to its effectiveness in estimating the co-movement and common shocks from a large number of variables, factor analysis has been used increasingly by economists for policy analysis in a “data rich environment.” (See, for example, Bernanke and Boivin, 2003, Bernanke et al. 2005, and Kose et al. 2003.) The purpose of this paper is to provide an inferential theory for the estimated parameters of high dimensional approximate factor models.

The notion of approximate factor models is proposed by Chamberlain and Rothschild (1983). Let $z_t$ be an $N \times 1$ random vector in period $t$ ($t = 1, 2, \cdots, T$); so $N$ represents the number of variables and $T$ the number of observations. Suppose that the covariance of $z_t$ has a factor structure $\Sigma = \Lambda \Lambda' + \Omega$, where $\Lambda$ is an $N \times r$ matrix of factor loadings, $r$ is the number of factors, and $\Omega$ is the covariance matrix of the idiosyncratic errors. An approximate factor model does not require $\Omega$ to be a diagonal matrix. In fact, there are no restrictions on the elements of $\Omega$ except that its maximum eigenvalue is bounded for all $N$. Thus, the idiosyncratic errors are allowed to be cross sectionally correlated with an unknown form.

Because none of the elements of $\Omega$ are fixed at certain known values, the number of free parameters in $\Omega$ alone is as many as that of $\Sigma$. Under fixed $N$, the model is not identifiable because the number of parameters (including those of $\Lambda$) exceeds the number of elements of $\Sigma$. However, Chamberlain and Rothschild show that the space spanned by the columns of $\Lambda$ is identifiable from $\Sigma$ as $N$ goes to infinity under the assumption of an approximate factor model (bounded eigenvalue for $\Omega$). However, Chamberlain and Rothschild do not study the sampling properties of the model because they assume $\Sigma$ is known, which is equivalent to the case of $T = \infty$. In this paper, we do not assume a known $\Sigma$, but $T$ observations on $z_t$ ($t = 1, 2, \cdots, T$). By
admitting the possibility that the number of variables \( N \) far exceeds the number of observations \( T \) such that \( T/N \) can converge to zero, our inferential theory cannot rely on a known or even a consistently estimable covariance matrix \( \Sigma \). Furthermore, we allow the observations \( z_t \) to be serially correlated and heteroscedastic over time. This setting is more general than the original notion of approximate factor models.

Most theory and applications in the literature are developed around the principal components method, e.g. Bai (2003), Breitung and Tenhofen (2011), Choi (2007), Connor and Korajczyk (1988), Doz et al. (2011b), Fan et al. (2011), Goyal et al. (2009), Inoue and Han (2011), Stock and Watson (2002ab), Wang (2010), among others. The present paper considers the likelihood-based estimation of the model. The likelihood-based method eliminates one source of bias arising from the cross-sectional heteroskedasticity, but the principal component method does not. Our paper is closely related to Doz et al. (2011a), which is also based on the likelihood framework. The latter does not directly study the maximum likelihood estimators; it focuses on estimating functions of the maximum likelihood estimators. More specifically, their paper studies the estimated factor as a function of the estimated loadings and variances, deriving an average consistency of the estimated factors.

The present paper shows that the maximum likelihood estimators (MLE) for the factor loadings and idiosyncratic variances are consistent. We establish individual parameters consistency in addition to average consistency. We further derive the rate of convergence and the limiting distributions. Having obtained the MLE of factor loadings and the idiosyncratic variances, in the second step, we also derive the limiting distribution of the estimated factors. We further estimate the dynamics in the idiosyncratic errors.

Efficient estimation of approximate factor models is also considered by Breitung and Tenhofen (2011) and Choi (2007). These papers propose two-step procedures for efficient estimation and derive the limiting distributions of the estimators. They
also suggest an iterated procedure. The simulation results of Breitung and Tenhofen (2011) show that iterated procedures can substantially improve upon the two-step procedure. This paper applies the maximum likelihood method to the approximate factor models. The analysis of the MLE in this context is more challenging than the two-step estimators. The theoretical difficulty stems from the joint estimation of the loadings and idiosyncratic variances; the estimators are solutions to a large number of nonlinear equations (first order conditions). We point out that the actual computation of MLE is relatively easy. There is no need to solve for the first order conditions, and the EM solutions satisfy the first order conditions.

It is noted that, unlike the usual linear or nonlinear regressions in which heteroskedasticity is often an issue of efficiency rather than consistency, heteroskedasticity in factor models is an issue of consistency, not only of efficiency. To be more specific, under fixed \( N \), if cross-sectional heteroskedasticity exists but is not allowed in the estimation, then the estimated factor loadings are inconsistent. Thus allowing heteroskedasticity is not innocuous as it may seem to be. Simultaneously analyzing the factor loadings and the variances is a demanding task owing to the increased nonlinearity of the estimation problem. Under large \( N \), heteroskedasticity will not affect consistency when ignored, but will still affect biases and efficiency.

Throughout the paper, we use \( \text{dg}(A) \) to denote the diagonal matrix that retains the diagonal elements of \( A \), while \( \text{diag}(A) \) denotes the vector consisting of the diagonal elements of \( A \). The norm of matrix \( A \) is defined as \( \|A\| = [\text{tr}(A'A)]^{1/2} \). The proofs for theoretical results are provided in the supplementary document.

## 2 QMLE for approximating factor models

We consider the following factor model:

\[
z_{it} = \alpha_i + \lambda_i'f_t + e_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T,
\]  

(1)
where $f_t$ and $\lambda_i$ are $r \times 1$ dimension of factors and factor loadings, respectively, and $e_{it}$ are the errors. A main feature of an approximate factor model is that $e_{it}$ are heteroskedastic and correlated over $i$. We shall also allow serial correlation and heteroskedasticity over the time dimension. Let $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_N)'$ be the $N \times r$ matrix of factor loadings and $z_t = (z_{1t}, ..., z_{Nt})'$ be the $N \times 1$ vector of variables. Let $e_t$ and $\alpha$ be similarly defined. We can rewrite (1) as,

$$z_t = \alpha + \Lambda f_t + e_t.$$  

(2)

Let $\Omega_t = E(e_t e'_t)$, which allows for heteroskedasticity over $t$. In classical factor analysis, $\Omega_t$ is assumed to be diagonal. Here $\Omega_t$ is $N \times N$ without the diagonality restriction, except that its maximum eigenvalue is bounded for all $N$. This is the essence of the approximate factor models. Because $\Omega_t$ contains as many free parameters as the number of elements in the sample variance of the observations, the number of parameters exceeds the number of estimating equations. So it is difficult to estimate all elements of $\Omega_t$. Let

$$\Phi = dg\left(\frac{1}{T} \sum_{t=1}^{T} \Omega_t\right)$$

where $dg(A)$ is a diagonal matrix that sets the off-diagonal elements of $A$ to zero. We are interested in estimating the elements of $\Phi$, a diagonal matrix. In the absence of cross-sectional correlation and time series heteroscedasticity, then $\Phi = E(e_t e'_t)$ and this reduces to the setting of classical factor analysis, except that the dimension $N$ is allowed to increase without a bound.

Let $M_{zz} = \frac{1}{T} \sum_{t=1}^{T} \bar{z}_t \bar{z}_t'$, the sample variance of the observable data, where $\bar{z}_t = z_t - \frac{1}{T} \sum_{t=1}^{T} z_t$. Then

$$E(M_{zz}) = \Lambda M_{ff} \Lambda' + \frac{1}{T} \sum_{t=1}^{T} E[(e_t - \bar{e})(e_t - \bar{e})']$$

where $M_{ff} = \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t \hat{f}_t'$, which is the sample variance of $f_t$ (we treat $f_t$ as a sequence of fixed constants, see Assumption A below). Define
\[ \Sigma_{zz} = \Lambda M_{ff} \Lambda' + \Phi. \]

Then \( \Sigma_{zz} \) is an approximation of \( E(M_{zz}) \) because we restrict \( \Phi \) to be diagonal. Thus \( \Sigma_{zz} \) is not the covariance matrix of \( z_t \). Furthermore, \( M_{ff} \) is not the population variance of \( f_t \), but the sample variance. Consider the objective function

\[
\ln L = -\frac{1}{2N} \ln |\Sigma_{zz}| - \frac{1}{2N}tr(M_{zz}\Sigma_{zz}^{-1}).
\]

Because \( \Sigma_{zz} \) is not the covariance matrix of \( z_t \) due to correlations and heteroskedasticities of unknown form in both dimensions, the above is not the likelihood function even under normality of \( e_{it} \). We may regard the objective function as a misspecified likelihood function. This particular form of misspecification is desirable as it coincides with the classical factor analysis under the exact factor structure. In general, we should view (3) as a distance measure between \( M_{zz} \) and \( \Sigma_{zz} \), as in Amemiya, Fuller, and Pantula (1987), and Anderson and Amemiya (1988). One goal of this paper is to show that this likelihood approach is robust to misspecifications under large \( N \) and large \( T \), similar to Doz et al. (2011a). Additionally, although \( f_t \) are fixed constants, we only estimate its sample variance instead of individual \( f_t \). This avoids the incidental parameters problem caused by estimating \( f_t \). In fact, when jointly estimating \( \lambda_i \) and \( f_t \), the likelihood function diverges to infinity for a judicious choice of parameter values (Anderson, 2003, p587). The above likelihood function does not have this problem.

Also note that, when \( N > T \), the sample covariance matrix \( M_{zz} \) is not invertible, but \( \Sigma_{zz} \) is invertible. Thus the likelihood function is well defined even when the number of variables is larger than the number of observations.

The parameters to be estimated are \( \theta = (\Lambda, \Phi, M_{ff}) \). If the variance of \( e_t = (e_{1t}, e_{2t}, \cdots, e_{Nt})' \) is diagonal and the \( e_{it} \) are iid over time, then we have an exact factor model. Estimating an exact factor model is considered by Bai and Li (2012) and they show that MLE is consistent. However in the present context, as indicated
in Assumption C, the true covariance matrix of \( e_t \) may be quite general. But the objective function (3) still regards the error terms as having an exact factor structure. Thus, as in Doz et al. (2011a), the ML method should be regarded as a quasi-ML (QML), and the resulting estimator will be referred to as QMLE. We will use MLE and QMLE interchangeably. We show that the QMLE is robust to departure from exact factor specifications. We will establish consistency and derive the limiting distributions.

2.1 Assumptions for approximating factor models

We make the following assumptions and their meanings are briefly explained below.

**Assumption A [Factors]:** The factors \( f_t \) are a sequence of fixed constants with \( \| f_t \| \leq C \) for all \( t \), where \( C \) is a constant large enough. Let \( M_{ff} = \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t \hat{f}_t' \) be the sample variance of \( f_t \), where \( \hat{f}_t = f_t - T^{-1} \sum_{t=1}^{T} f_t \). There exists an \( M_{ff} > 0 \) such that \( \lim_{T \to \infty} M_{ff} = \tilde{M}_{ff} \).

Although Assumption A assumes \( f_t \) being fixed constants, \( f_t \) can be random variables. In this case, we assume \( f_t \) to be independent of the errors \( e_{is} \) for all \( (i, s) \) and also \( E \| f_t \|^4 \leq C \) instead of \( \| f_t \| \leq C \). Note that \( f_t \) can be a dynamic process with arbitrary dynamics.

**Assumption B [Factor loadings]:** The factor loadings \( \lambda_i \) satisfy \( \| \lambda_i \| \leq C \) for all \( i \). In addition, there exists an \( r \times r \) positive matrix \( Q \) such that \( \lim_{N \to \infty} N^{-1} \Lambda' \Phi^{-1} \Lambda = Q \), where \( \Phi \) is defined earlier.

**Assumption C [Cross-sectional and serial dependence and heteroskedasticity]:**

For a constant \( C \) large enough, not depending on \( N \) and \( T \),

C.1 \( E(e_{it}) = 0, E(e_{it}^8) \leq C \).

C.2 Let \( \Phi = \text{dg} \left\{ \frac{1}{T} \sum_{t=1}^{T} E(e_t e_t') \right\} = \text{dg} \left\{ \frac{1}{T} \sum_{t=1}^{T} \Omega_t \right\} \). So \( \Phi \) is an \( N \times N \) diagonal matrix with the \( i \)th element \( \phi_i^2 = \frac{1}{T} \sum_{t=1}^{T} \tau_{ii,t} \) where \( \tau_{ii,t} \) is the \( (i, i) \) element of \( \Omega_t \). We assume \( C^{-2} \leq \phi_i^2 \leq C^2 \) for all \( i \).
C.3 \( E(e_{it}e_{jt}) = \tau_{\tilde{y},t} \) with \( |\tau_{\tilde{y},t}| \leq \tau_{\tilde{y}} \) for some \( \tau_{\tilde{y}} > 0 \) and for all \( t \). In addition, 
\[ \sum_{i=1}^{N} \tau_{ij} \leq C \text{ for any } j. \]

C.4 \( E(e_{it}e_{is}) = \rho_{i,s} \) with \( |\rho_{i,s}| \leq \rho_{s} \) for some \( \rho_{s} > 0 \) and for all \( i \). In addition, 
\[ \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \rho_{s} \leq C. \]

C.5 for all \( i, j = 1, 2, \cdots, N \), 
\[ E \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left| e_{it}e_{jt} - E(e_{it}e_{jt}) \right|^{4} \right] \leq C \]

Assumption C allows for heteroskedasticities and weak correlations over the cross section and the time dimension, and is more general than the traditional factor analysis. This assumption also introduces notations for correlations and moments to be used in the proof. Assumption C.1 is a standard moment condition. We refer \( \phi_{i}^{2} \) in Assumption C.2 as the time-average variance for individual \( i \). C.2 requires that the time-average variance of \( e_{it} \) be bounded away from below and above. Assumption C.3 aims to control the correlation over the cross section. Assumptions C.4 and C.5 control the magnitude of the correlation of \( e_{it} \) over time.

Assumption D: The diagonal elements of \( \Phi \) are estimated in the compact set \( [C^{-2}, C^{2}] \). Furthermore, \( M_{ff} \) is also restricted in a compact set with all the elements bounded in the interval \( [C^{-1}, C] \), where \( C \) is a constant large enough.

Assumption D requires that part of the variance estimators be estimated in a compact set. Restricting parameters in a compact set is usually made for nonlinear models, e.g., Newey and McFadden (1994), Jenirich (1969), and Wu (1981). The objective function for factor models is highly nonlinear. Nevertheless, no restrictions for \( \Lambda \) are needed. Throughout, we also assume that the number of factors \( r \) is known. When unknown, it can be consistently estimated (e.g., Bai and Ng, 2002).

### 2.2 Identification restrictions

It is well known that the factor model can only be identified up to a rotation. To fix the indeterminacy, we consider five sets of commonly used restrictions:
IC1: \( \Lambda = (I_r, \Lambda_2') \); IC2: \( \frac{1}{N} \Lambda' \Sigma^{-1} \Lambda = I_r \) and \( M_{ff} = D \), where \( D \) is a diagonal matrix, whose diagonal element are distinct and arranged in descending order; IC3: \( \frac{1}{N} \Lambda' \Sigma^{-1} \Lambda = D \) and \( M_{ff} = I_r \), where \( D \) is a diagonal matrix, whose diagonal element are distinct and arranged in descending order; IC4: \( \Lambda_1 \) is a lower triangular matrix with all diagonal elements being 1 and \( M_{ff} = D \), where \( \Lambda_1 \) is the upper \( r \times r \) submatrix of \( \Lambda \) and \( D \) is a diagonal matrix; IC5: \( \Lambda_1 \) is a lower triangular matrix with none of its diagonal element being 0 and \( M_{ff} = I_r \), where \( \Lambda_1 \) is the upper \( r \times r \) submatrix of \( \Lambda \).

IC1 is similar to a measurement error problem: the first variable is related to the first factor plus an error, the second variable is related to the second factor plus an error, and so on. IC2 and IC3 are used by the classical maximum likelihood method. IC4 and IC5 are similar to a recursive simultaneous equations system for the first \( r \) variables. IC1 and IC4 allow for full identification of the model, while IC2, IC3 and IC5 identify \( \Lambda \) up to a column sign change. In practice, IC1, IC4, and IC5 require a careful choice of the first \( r \) variables (in order to give meaningful interpretations to the loadings and the factors). We refer the readers to Anderson and Rubin (1956), Lawley and Maxwell (1971), and Bai and Li (2012) for further details.

Remark 1. The first order conditions of the objective function (3) can be solved by the EM algorithm. Bai and Li (2012) show that the EM solutions satisfy the first order conditions of the log-likelihood function for exact factor models. In the present paper, the misspecified likelihood function is the same as the objective function considered in Bai and Li (2012). So the EM solutions also satisfy the first order conditions.

2.3 Consistency and convergence rate

The infinite number of parameters in the limit makes the usual consistency concept not well defined. We solve the problem by obtaining an average consistency first, and
from the average consistency we further derive individual parameter consistency. Let
\[ \hat{\theta} = (\hat{\lambda}_1, \cdots, \hat{\lambda}_N, \hat{\phi}_1^2, \cdots, \hat{\phi}_N^2, \hat{M}_{ff}) \] be the MLE. Proposition S.1 in the supplement gives the average consistency:
\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \| \hat{\lambda}_i - \lambda_i \|^2 \overset{p}{\to} 0, \quad \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \overset{p}{\to} 0, \quad \hat{M}_{ff} - M_{ff} \overset{p}{\to} 0
\]
where \( \phi_i^2 = \frac{1}{T} \sum_{t=1}^{T} E(e_{it}^2) = \frac{1}{T} \sum_{t=1}^{T} \tau_{ii,t} \). The first result shows that the estimated factor loadings are consistent on average. The second result is interesting. In view of Assumption C, the error term \( e_{it} \) is allowed to have very general cross-section and serial correlations, but the estimator \( \hat{\phi}_i^2 \) has no relation with these correlations, and is estimating the average variance over time for each individual \( i \). In a sense, the cross-section and serial correlations do not contaminate the estimator (these correlations do affect the limiting variance, as is shown in later sections.)

We now state the rate of convergence.

**Proposition 1 (Convergence rates)** Under Assumptions A-D, when \( N, T \to \infty \),
with any one of the identification conditions, we have
\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \| \hat{\lambda}_i - \lambda_i \|^2 = O_p(T^{-1}) + O_p(N^{-2}),
\]
\[
\frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 = O_p(T^{-1}) + O_p(N^{-2}),
\]
\[
\| \hat{M}_{ff} - M_{ff} \|^2 = O_p(T^{-1}) + O_p(N^{-2}).
\]

For exact factor models, the \( O_p(N^{-2}) \) term does not exist. Bai and Li (2012) show that \( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \| \hat{\lambda}_i - \lambda_i \|^2 = O_p(T^{-1}) \). The same is true for \( \hat{\phi}_i^2 \) and \( \hat{M}_{ff} \). Whether \( N \) is fixed or large, the MLE is consistent under exact factor models. Proposition 1 shows that there is a cost associated with the generality of the approximate factor models. That is, under fixed \( N \), the estimated factor loadings will not be consistent for approximate factor models; this should not be surprising. Under large \( N \), the MLE becomes consistent, illustrating the advantage of high dimension data.
Remark 2. Part of the ML analysis includes showing that the rotation matrix (denoted by $R$) is equal to $I_r$, that is, the MLE directly estimates $\lambda_i$ instead of its rotation. This is obtained by assuming that the underlying parameters satisfy the identification restrictions, as in classical factor analysis. If this assumption does not hold, then we will be estimating rotations of the factor loadings. The absence of rotation ($R = I_r$) is more difficult to establish than allowing a rotation. The principal component analysis of Bai (2003) and the two-step estimators of Breitung and Tenhofen (2011) and Choi (2007) do not investigate this rotational properties.

2.4 Limiting distributions

We first derive the asymptotic representations of QMLE, from which the limiting distributions follow easily. Asymptotic representations contain more information than limiting distributions as they show equality up to an $o_p(1)$ term rather than equality in distribution. We need additional assumptions.

Assumption E [moment conditions]: There exists a constant $C$ such that

E.1 $E(e_{it}e_{js}) = \gamma_{ij,ts}$ with $\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\gamma_{ij,ts}| \leq C$.

E.2 for each $j = 1, 2, \cdots, N$, $E\left[\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i} \lambda_i[e_{it}e_{jt} - E(e_{it}e_{jt})]\right\|^2 \right] \leq C$.

E.3 the $r \times r$ matrix satisfies $E\left[\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i} \lambda_i\lambda_i'(e_{it}^2 - \phi_i^2)\right\|^2 \right] \leq C$.

Assumption F [Central Limit Theorem]:

F.1 For each $i$, as $T \to \infty$, $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t e_{it} \xrightarrow{d} N(0, \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} f_t f_{ts} \rho_{i,ts})$.

F.2 For each $i$, as $T \to \infty$, $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (e_{it}^2 - \phi_i^2) \xrightarrow{d} N(0, \sigma_i^2)$, where $\sigma_i^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[ (e_{it}^2 - \phi_i^2)(e_{is}^2 - \phi_i^2) \right]$.

Assumption E.1 controls the magnitude of correlation of $e_{it}$ over the cross section and the time dimensions. Assumptions E.2 and F.1 are conventional. Similar
assumptions are also made in Bai (2003). Assumption E.3 and F.2 are extra due to the estimation of heteroskedasticity, and is used for the limiting distribution $\hat{\phi}_t^2$.

Throughout the paper, let $\xi_t = (e_{1t},...,e_{rt})'$, a vector consisting of the idiosyncratic errors in the first $r$ equations. This vector will appear in the asymptotic representations of the estimators under IC1, IC4, and IC5. In addition, under IC4 and IC5, the asymptotic representations involve two $r \times r$ matrices $P_t$ and $Q_t$. Their $(g, h)$-th elements are defined, respectively, as

$$
P_{gh,t} = \begin{cases} -m_g^{-1} f_{tg} \xi_t' \Lambda_1^{-1} v_h & \text{if } g \geq h \\ -m_g^{-1} m_h P_{hg,t} & \text{if } g < h \end{cases}, \quad Q_{gh,t} = \begin{cases} -f_{tg} \xi_t' \Lambda_1^{-1} v_h & \text{if } g > h \\ 0 & \text{if } g = h \\ -Q_{hg,t} & \text{if } g < h \end{cases}$$

(4)

where $m_g$ is the $g$th diagonal element of $M_{ff}$; $f_{tg}$ is the $gth$ component of $f_t$; $\Lambda_1$ is the first $r \times r$ block of $\Lambda$; and $v_h$ is the $h$th column of the identity matrix $I_r$. Matrix $Q_t$ is skew-symmetric. Now we state the asymptotic representations for the estimated factor loadings.

**Theorem 1 (Asymptotic representations for factor loadings)** Under Assumptions A-E, and $N, T \to \infty$ with $\sqrt{T}/N \to 0$, for each $j = 1, 2, \cdots, N$ under IC2 and IC3, and for $j > r$ under IC1, IC4, and IC5, we have:

- Under IC1,
  $$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = M_{ff}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (f_t e_{jt} - f_t \xi_t' \lambda_j) + o_p(1);$$

- Under IC2 or IC3,
  $$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = M_{ff}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t e_{jt} + o_p(1);$$

- Under IC4,
  $$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (P_t \lambda_j + M_{ff}^{-1} f_t e_{jt}) + o_p(1);$$

- Under IC5,
  $$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (Q_t \lambda_j + f_t e_{jt}) + o_p(1);$$

where $\xi_t$, $P_t$, and $Q_t$ are all defined earlier. Given the above result, together with Assumption F, we have

- Under IC1,
  $$\sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, (M_{ff})^{-1} \Gamma_2^\lambda (M_{ff})^{-1});$$
under IC2 or IC3, \[
\sqrt{T}(\hat{\lambda}_j - \lambda_j) \overset{d}{\to} N(0, (\bar{M}_{ff})^{-1}\Gamma^\lambda_j(\bar{M}_{ff})^{-1});
\]
under IC4, \[
\sqrt{T}(\hat{\lambda}_j - \lambda_j) \overset{d}{\to} N(0, \Pi^\lambda_j);
\]
under IC5, \[
\sqrt{T}(\hat{\lambda}_j - \lambda_j) \overset{d}{\to} N(0, \Psi^\lambda_j);
\]
where \(\Gamma^\lambda_j, \Psi^\lambda_j, \Pi^\lambda_j\) are defined in Table 1, and \(\bar{M}_{ff}\) is defined in Assumption A.

2.5 Estimating the factors

Following Mardia et al. (1979) and Anderson (2003), we use the projection formula and the generalized least squares (GLS) to estimate the factors:

(Projection formula) \[
\tilde{f}_t = (\hat{\Lambda}^{-1} + \hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Phi}^{-1}(z_t - \bar{z})
\] (5)

(GLS) \[
\hat{f}_t = (\hat{\Lambda}^{-1} + \hat{\Phi}^{-1})^{-1}(z_t - \bar{z})
\] (6)

It is easy to show that \(\tilde{f}_t = \hat{f}_t + O_p(N^{-1})\). So the two estimators are asymptotically equivalent. In what follows, we only focus on \(\hat{f}_t\). To analyze the asymptotic properties of \(\hat{f}_t\), we strengthen Assumption C.4 to C.4' below.

Assumption C [continued]: There exists a constant \(C\) large enough such that:

C.4' \[
\sum_{t=1}^T \rho_{ts} \leq C, \text{ where } \rho_{ts} \geq 0 \text{ is defined in Assumption C.4.}
\]

Assumption E [moment conditions (continued)]: There exists a \(C\) such that

E.4 for all \(t, t = 1, 2, \cdots, T\), \(E\left(\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{s=1}^T \frac{1}{\phi_i^t} f_s[e_{it}e_{is} - E(e_{it}e_{is})]\right\|^2 \right) \leq C.\)

E.5 for all \(t, t = 1, 2, \cdots, T\), \(E\left(\left\| \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^T f_s[e_{it}e_{is} - E(e_{it}e_{is})]\right\|^2 \right) \leq C.\)

E.6 for all \(t, t = 1, 2, \cdots, T\), \(E\left(\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{s=1}^T \frac{1}{\phi_i^t} \lambda_i(e_{is}^2 - \phi_i^t)^2 e_{it}\right\|^2 \right) \leq C.\)

Assumption F [Central Limit Theorem (continued)]

F.3 for each \(t\), as \(N \to \infty\), \[
\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{\phi_i^t} \lambda_i e_{it} \overset{d}{\to} N\left(0, \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{N} \frac{1}{\phi_i^t \phi_j^t} \lambda_i \lambda_j \tau_{ij,t} \right).
\]
Most of the preceding assumptions are intuitive and reasonable. They are the counterparts of the assumptions made earlier. For example, Assumption C.4′ corresponds to Assumption C.3; Assumption E.4 corresponds to Assumption E.2; Assumption E.5 corresponds to Assumption C.5, which aims to control the correlation of the cross-product term \( e_{it}e_{is} \) over time. Assumption E.6 is used to bound \( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^4} (\hat{\phi}_i^2 - \phi_i^2) \lambda_{ie_it} \), and insures that it has a fast convergence rate; Assumption F.3 corresponds to Assumption F.1.

The following theorem states the asymptotic representations for \( \hat{f}_t \):

**Theorem 2 (Asymptotic representations for the factors)** Under Assumptions A-E and \( N, T \to \infty \) with \( \sqrt{N}/T \to 0 \), and for \( \Delta \in [0, \infty) \), we have:

under IC1 and \( N/T \to \Delta \),

\[
\sqrt{N}(\hat{f}_t - f_t) = -\sqrt{\Delta}(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \xi_s f'_s) M_{ff}^{-1} f_t + Q^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \lambda_{ie_it} + o_p(1).
\]

under IC2 or IC3,

\[
\sqrt{N}(\hat{f}_t - f_t) = Q^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \lambda_{ie_it} + o_p(1).
\]

under IC4 and \( N/T \to \Delta \),

\[
\sqrt{N}(\hat{f}_t - f_t) = -\sqrt{\Delta}(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \mathcal{P}_s f'_s) f_t + Q^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \lambda_{ie_it} + o_p(1).
\]

under IC5 and \( N/T \to \Delta \),

\[
\sqrt{N}(\hat{f}_t - f_t) = -\sqrt{\Delta}(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \mathcal{Q}_s f'_s) f_t + Q^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \lambda_{ie_it} + o_p(1).
\]

The variables \( \xi_t, \mathcal{P}_s \) and \( \mathcal{Q}_s \) are defined earlier. Given the above result, together with Assumption F, we have

under IC1 and \( N/T \to \Delta \), \( \sqrt{N}(\hat{f}_t - f_t) \overset{d}{\to} N(0, \Gamma_T^f) \);

under IC2 or IC3, \( \sqrt{N}(\hat{f}_t - f_t) \overset{d}{\to} N(0, \Upsilon_T^f) \);

under IC4 and \( N/T \to \Delta \), \( \sqrt{N}(\hat{f}_t - f_t) \overset{d}{\to} N(0, \Pi_T^f) \);

under IC5 and \( N/T \to \Delta \), \( \sqrt{N}(\hat{f}_t - f_t) \overset{d}{\to} N(0, \Psi_T^f) \);

where \( \Gamma_T^f, \Upsilon_T^f, \Pi_T^f, \Psi_T^f \) are given in Table 1.

14
The asymptotic representations depend on the identification conditions. The identification conditions of IC2 and IC3 imply a simpler asymptotic expression. For IC1, IC4 and IC5, there are two terms in the representation. The first term involves partial sums over the time dimension, whereas the second term involves partial sums over the cross-section dimension. If $\Delta$ is large ($N$ is large relative to $T$), then the first term is more important in the determination of the asymptotic variance. This means that the error terms in the time dimension for the first $r$ individuals are the primary source of the variability of $\hat{f}_t - f_t$ [noting $\xi_t = (e_{1t}, ..., e_{rt})$]. If $\Delta$ is small, the error terms over the entire cross section for period $t$ are the primary source of the variability. That is, the second term of the presentation will be more important. If $\Delta \to 0$, the first term drops out. Theorem 2 shows that the relative ratio between $N$ and $T$ plays a role in efficiency.

Table 1: Definition of symbols in the limiting distributions:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Gamma_j^\delta$</td>
<td>$\lim_{T \to \infty} \frac{1}{T} \sum_{s,t=1}^T [(X_j^s \otimes f_t)E(\xi_t \xi'_s)(\lambda_j \otimes f'_s) + (X_j^s \otimes f_t)E(\xi_t e_js) f'<em>s + f_t E(e</em>{jt} \xi'_s)(\lambda_j \otimes f'<em>s) + f_t E(e</em>{jt} e_js) f'_s]$</td>
</tr>
<tr>
<td>$\Upsilon_j^\delta$</td>
<td>$\lim_{T \to \infty} \frac{1}{T} \sum_{s,t=1}^T f_t f'_s \rho_j ts$</td>
</tr>
<tr>
<td>$\Pi_j^\delta$</td>
<td>$\lim_{T \to \infty} \frac{1}{T} \sum_{s,t=1}^T [E(P_i \lambda_j \lambda_j P'<em>s) + M</em>{ij}^{-1} f_i \lambda_j \rho_j e_{jt} P_s + E(P_i e_{jt} P_s) \lambda_j \rho_j^{-1} M_{ij}^{-1} f_i f'_s \rho_j ts]$</td>
</tr>
<tr>
<td>$\Psi_j^\delta$</td>
<td>$\lim_{T \to \infty} \frac{1}{T} \sum_{s,t=1}^T [E(Q_i \lambda_j \lambda_j Q'<em>s) + f_t \lambda_j E(e</em>{jt} Q_s) + E(Q_i e_{jt} Q_s) \lambda_j f'_s + f_t f'_s \rho_j ts]$</td>
</tr>
<tr>
<td>$\Gamma_i^\delta$</td>
<td>$\Delta \lim_{N \to \infty} \frac{1}{N} \sum_{i,j=1}^N \frac{1}{\sigma_i^2 \sigma_j^2} \lambda_i \lambda_j \gamma_{ij,t} Q^{-1}$</td>
</tr>
<tr>
<td>$\Upsilon_i^\delta$</td>
<td>$Q^{-1} \left( \lim_{N \to \infty} \frac{1}{N} \sum_{i,j=1}^N \frac{1}{\sigma_i^2 \sigma_j^2} \lambda_i \lambda_j \gamma_{ij,t} Q^{-1} \right)$</td>
</tr>
<tr>
<td>$\Pi_i^\delta$</td>
<td>$\Delta (I_r \otimes f'<em>t) \left( \lim</em>{T \to \infty} \frac{1}{T} \sum_{s,u=1}^T E[\text{vec}(P_{s,u}) \text{vec}(P_{s,u}')] (I_r \otimes f_t) + Q^{-1} \left( \lim_{N \to \infty} \frac{1}{N} \sum_{i,j=1}^N \frac{1}{\sigma_i^2 \sigma_j^2} \lambda_i \lambda_j \gamma_{ij,t} Q^{-1} \right) Q^{-1} \right)$</td>
</tr>
<tr>
<td>$\Psi_i^\delta$</td>
<td>$\Delta (I_r \otimes f'<em>t) \left( \lim</em>{T \to \infty} \frac{1}{T} \sum_{s,u=1}^T E[\text{vec}(Q_{s,u}) \text{vec}(Q_{s,u}')] (I_r \otimes f_t) + Q^{-1} \left( \lim_{N \to \infty} \frac{1}{N} \sum_{i,j=1}^N \frac{1}{\sigma_i^2 \sigma_j^2} \lambda_i \lambda_j \gamma_{ij,t} Q^{-1} \right) Q^{-1} \right)$</td>
</tr>
</tbody>
</table>

Note: $P_t$ and $Q_t$ are defined in (4).
3 Dynamics in $f_t$ and the Kalman smoother

If the dynamics in $f_t$ is explicitly modeled by a vector autoregressive (VAR) process, the Kalman smoother is an alternative method to estimate $f_t$. This section presents the asymptotic results for the Kalman-smoother-based estimator (KSE), when the smoother is evaluated at the QML estimator. We further derives its limiting distribution. We show that KSE has the same limiting distribution as the GLS of the previous section.

Consider the following dynamics of the factors characterized by a VAR($K$):

$$f_t = \Psi_1 f_{t-1} + \Psi_2 f_{t-2} + \cdots + \Psi_K f_{t-K} + u_t.$$  \hfill (7)

We rewrite model (2) as $Z = \Lambda F' + E$, where $F = (f_1, f_2, \ldots, f_T)'$, $Z = (z_1, z_2, \ldots, z_T)$ and $E = (e_1, e_2, \ldots, e_T)$. Both $Z$ and $E$ are $N \times T$. Let

$$Z = \text{vec}(Z), \quad F = \text{vec}(F'), \quad E = \text{vec}(E),$$

where $Z$ is $NT \times 1$, $F$ is $rT \times 1$ and $E$ is $NT \times 1$. Then we have

$$Z = (I_T \otimes \Lambda)F + E.$$  \hfill (8)

The following analysis will assume normality for $F$ and $E$ and they are independent. If we interpret the conditional expectation as a linear population projection, normality is in fact not needed. Assume

$$\mathcal{F} \sim N(0, \Sigma_{\mathcal{F}}), \quad \mathcal{E} \sim N(0, I_T \otimes \Phi)$$

where $\Sigma_{\mathcal{F}} = \text{var}(\mathcal{F})$, an $rT \times rT$ matrix. The above assumption implies that $e_{it}$ are uncorrelated over time. If $\Psi$ is further assumed to be diagonal, then $e_{it}$ is also uncorrelated over $i$. We have, from (8)

$$\begin{bmatrix} \mathcal{F} \\ \mathcal{Z} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathcal{F}} & \Sigma_{\mathcal{F}}(I_T \otimes \Lambda') \\ (I_T \otimes \Lambda)\Sigma_{\mathcal{F}} & (I_T \otimes \Lambda)\Sigma_{\mathcal{F}}(I_T \otimes \Lambda') + I_T \otimes \Phi \end{bmatrix} \right),$$

16
Thus the best prediction for $F$ given $(Z, \Lambda, \Phi, \Sigma_F)$, denoted by $E(F|Z)$, is

$$E(F|Z) = \Sigma_F(I_T \otimes \Lambda') \left[ (I_T \otimes \Lambda)\Sigma_F(I_T \otimes \Lambda') + I_T \otimes \Phi \right]^{-1} Z$$

$$= \left[ \Sigma_F^{-1} + I_T \otimes (\Lambda'\Phi^{-1}\Lambda) \right]^{-1} \left[ I_T \otimes (\Lambda'\Phi^{-1}) \right] Z,$$

where the second equality is due to the Woodbury identity. Equation (9) is the Kalman smoother for the factors, which serve as the basis for the KSE.

Because the parameters $\Lambda, \Phi, \Sigma_F$ are unknown we replace them with their corresponding QMLE. More specifically, we first apply the QML method to obtain $\hat{\Lambda}, \hat{\Phi}, \hat{F}$, where $\hat{F} = Z'\hat{\Phi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}$ given in (6), then obtain $\hat{\Sigma}_F$ by the standard vector time series regression based on $\hat{f}_t$ and (7). For example, if $f_t = \Psi f_{t-1} + u_t$, $\hat{\Sigma}_F$ contains elements $\hat{\sigma}_u^2 \hat{\Psi}_t$ with $t, s = 1, 2, \ldots, T$. Given $\hat{\Sigma}_F, \hat{\Lambda}$ and $\hat{\Phi}$, the KSE for the entire vector is

$$\hat{F} = \left[ \hat{\Sigma}_F^{-1} + I_T \otimes (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda}) \right]^{-1} \left[ I_T \otimes (\hat{\Lambda}'\hat{\Phi}^{-1}) \right] Z.$$

This implies that the KSE for $f_t$, denoted by $\hat{f}_t^{ks}$ is

$$\hat{f}_t^{ks} = (v'_t \otimes I_r) \left[ \hat{\Sigma}_F^{-1} + I_T \otimes (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda}) \right]^{-1} \left[ I_T \otimes (\hat{\Lambda}'\hat{\Phi}^{-1}) \right] Z \tag{10}$$

where $v_t$ is the $t$-th column of the $T \times T$ identity matrix. To analyze the above estimator, we assume $f_t$ in (7) is stationary.

**Assumption A':** The factor $f_t$ admits the VAR representation (7), where $u_t$ is a mean-zero i.i.d process with $E(||u_t||^4) \leq C$ for some constant $C$ large enough. Furthermore, the roots of the polynomial $\Psi(L) = I_r - \Psi_1 L - \cdots - \Psi_K L^K = 0$ are all outside the unit circle.

Now we state the asymptotic results on $\hat{f}_t^{ks}$.

**Theorem 3 (asymptotic equivalence between $\hat{f}_t^{ks}$ and $\hat{f}_t$)** Under Assumptions A', B-E, when $N, T \to 0, T/N^3 \to 0$, we have $\sqrt{N}(\hat{f}_t^{ks} - \hat{f}_t) = o_p(1)$, where $\hat{f}_t$ is the GLS estimator in (6).

Theorem 3 shows that the difference between the KSE, which takes into account of
the dynamics in factors, and the projection-based estimator, which only uses the con-temporaneous relations between the factors and the observables, are asymptotically negligible. Thus, modeling the dynamics of factors will not improve the asymptotic efficiency under large $N$, though there will be efficiency gain under small $N$. Doz et al. (2011b) also consider the Kalman smoother method. They derive the rate of convergence of the KSE when evaluated at the principal components estimator. The results here also include the limiting distributions.

4 Modeling the dynamics in the errors $e_{it}$

So far we have assumed that the serial correlation in $e_{it}$ is of an unknown form. If we are willing to assume that $e_{it}$ is an autoregressive process, then this should be modeled and the factor loadings can be more efficiently estimated. The dynamic coefficients in $e_{it}$ can also be consistently estimated. Consider the following model

$$z_{it} = \alpha_i + \lambda_i'f_t + e_{it},$$

$$e_{it} = \rho_{i,1}e_{it-1} + \cdots + \rho_{i,p_i}e_{it-p_i} + \epsilon_{it}$$ (11)

so $e_{it}$ follows an $AR(p_i)$ process with the lag orders $p_i$ depending on $i$. Let $\rho_i(L) = 1 - \rho_{i,1}L - \cdots - \rho_{i,p_i}L^{p_i}$. The $e_{it}$ process can be rewritten as $\rho_i(L)e_{it} = \epsilon_{it}$. We assume that $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$ is an $i.i.d$ process over $t$. In what follows, we assume $\epsilon_{it}$ and $\epsilon_{jt}$ are independent for $i \neq j$, for simplicity; $E\epsilon_{it} = 0$ and $\text{var}(\epsilon_{it}) = \sigma^2_{\epsilon_i}$. Breitung and Tenhofen (2011) consider a two-step method to estimate model (11). In the first step, they use PC method to obtain the estimates of the factors and factor loadings, and based on the residuals, they calculate the estimates of the variance of $e_{it}$ and the coefficients $(\rho_{i,1}, \rho_{i,2}, \ldots, \rho_{i,p_i})$. In the second step, by taking into account the heteroscedasticity and autocorrelation of $e_{it}$, they use GLS to improve the estimates of the factors and factor loadings. They call the procedure PC-GLS. Iterating this procedure several times leads to, what they call, iterated PC-GLS. Their simulation
shows that the iterated PC-GLS has better finite sample properties.

When the sample size is small or moderate, especially when heteroscedasticity of the cross section is strong, the PC method gives poor estimates for the variance of $e_{it}$ and the coefficients $\rho_i = (\rho_{i,1}, \rho_{i,2}, \ldots, \rho_{i,p_i})$, which lead to unsatisfactory performance of the PC-GLS and the iterated PC-GLS. Motivated by this concern, we propose two estimators, ML-GLS and iterated ML-GLS estimators. The ML-GLS estimators, which include $\hat{\Lambda}, \hat{F}, \hat{\rho}_1, \ldots, \hat{\rho}_N, \hat{\Phi}$, are calculated by the following two steps:

1. Apply the QML method to the first equation of (11) to obtain the QMLE $\hat{\Lambda}$ and $\hat{\Phi}$. Then calculate $\hat{F} = Z'\hat{\Phi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}$ and the residuals $\hat{e}_{it} = z_{it} - \hat{\Lambda}_i^t\hat{f}_t$.

   For each $i$, obtain the estimators $\hat{\rho}_i$ by running the following regression
   \[
   \hat{e}_{it} = \rho_{i,1}\hat{e}_{i,t-1} + \cdots + \rho_{i,p_i}\hat{e}_{i,t-p_i} + \text{error}, \quad t = p_i + 1, \ldots, T
   \]

2. Given $(\hat{\rho}_{i,1}, \hat{\rho}_{i,2}, \ldots, \hat{\rho}_{i,p_i})$ and $\hat{F}$, update the estimator of $\Lambda$, denoted by $\tilde{\Lambda}$, by running the regression
   \[
   z_{it} - \hat{\rho}_{i,1}z_{i,t-1} - \cdots - \hat{\rho}_{i,p_i}z_{i,t-p_i} = (\hat{f}_t - \hat{\rho}_{i,1}\hat{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\hat{f}_{t-p_i})\lambda_i + \text{error}, \quad t = p_i + 1, \ldots, T
   \]

   Given $\hat{\Phi} = \text{diag}(\hat{\phi}_1^2, \ldots, \hat{\phi}_N^2)$ and $\tilde{\Lambda}$, update the estimator of $F$, denoted by $\tilde{F}$, by running the regression
   \[
   \frac{1}{\hat{\phi}_i}z_{it} = \left(\frac{1}{\hat{\phi}_i}\tilde{\lambda}_i\right)'f_t + \text{error}, \quad i = 1, 2, \ldots, N
   \]

The iterated ML-GLS can be obtained by iterating the above two steps several times and, for each iteration, $\tilde{\Lambda}, \tilde{F}$ are replaced with the estimators of the previous iteration.

The asymptotic properties of ML-GLS now can be formally analyzed given the asymptotic properties of the QMLE in the previous two sections. We state the results in the following theorem.

**Theorem 4** Under the Assumptions in Supplement E, when $N,T \to \infty$, together with the identification condition IC3, we have, for each $i$ and $t$

\[
\hat{\rho}_i \overset{p}{\to} \rho_i, \quad \tilde{\lambda}_i \overset{p}{\to} \lambda_i, \quad \tilde{f}_t \overset{p}{\to} f_t
\]
Furthermore, with the condition $\sqrt{T}/N \to 0$, for each $i$,  
\[
\sqrt{T}(\hat{\rho}_i - \rho_i) = \left( \frac{1}{T} \sum_{t=p_i+1}^T \psi_{it}\psi_{it}' \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=p_i+1}^T \psi_{it}\epsilon_{it} \right) + o_p(1)
\]
\[
\sqrt{T}(\tilde{\lambda}_i - \lambda_i) = \left( \frac{1}{T} \sum_{t=p_i+1}^T g_{it}g_{it}' \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=p_i+1}^T g_{it}\epsilon_{it} \right) + o_p(1)
\]
and with the condition $\sqrt{N}/T \to 0$, for each $t$,  
\[
\sqrt{N}(\hat{f}_t - f_t) = \left( \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma_i^2} \lambda_i \lambda_i' \right)^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{\phi_i} \lambda_i \epsilon_{it} \right) + o_p(1)
\]
where $\psi_{it} = (\epsilon_{it-1}, \epsilon_{it-2}, \ldots, \epsilon_{it-p_i})'$ and $g_{it} = f_t - \rho_{i,1} f_{t-1} - \cdots - \rho_{i,p_i} f_{t-p_i}$. Thus, if $\sqrt{T}/N \to 0$, for each $i$,  
\[
\sqrt{T}(\hat{\rho}_i - \rho_i) \xrightarrow{d} N\left(0, \sigma_i^2 \left[ \operatorname{plim}_{T \to \infty} \frac{1}{T} \sum_{t=p_i+1}^T \psi_{it}\psi_{it}' \right]^{-1} \right),
\]
\[
\sqrt{T}(\tilde{\lambda}_i - \lambda_i) \xrightarrow{d} N\left(0, \sigma_i^2 \left[ \operatorname{plim}_{T \to \infty} \frac{1}{T} \sum_{t=p_i+1}^T g_{it}g_{it}' \right]^{-1} \right).
\]
If $\sqrt{N}/T \to 0$, then $\sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, Q^{-1})$, with $Q$ given in Assumption B.

As for the estimation of the asymptotic variance, a consistent estimator for $\sigma_i^2$ is  
\[
\hat{\sigma}_i^2 = \frac{1}{1-p_i} \sum_{t=p_i+1}^T \hat{\epsilon}_{it}^2, \text{ where}
\]
\[
\hat{\epsilon}_{it} = z_{it} - \hat{\rho}_{i,1} z_{i,t-1} - \cdots - \hat{\rho}_{i,p_i} z_{i,t-p_i} - (\hat{f}_t - \hat{\rho}_{i,1} \hat{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i} \hat{f}_{t-p_i})' \hat{\lambda}_i.
\]
The asymptotic variance of $\sqrt{T}(\tilde{\lambda}_i - \lambda_i)$ can be estimated by $\hat{\sigma}_i^2 \left( \frac{1}{T} \sum_{t=p_i+1}^T \tilde{g}_{it}\tilde{g}_{it}' \right)^{-1}$ with $\tilde{g}_{it} = \tilde{f}_i - \hat{\rho}_{i,1} \tilde{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i} \tilde{f}_{t-p_i}$, and the asymptotic variance of $\sqrt{T}(\hat{\rho}_i - \rho_i)$ is estimated by $\hat{\sigma}_i^2 \left( \frac{1}{T} \sum_{t=p_i+1}^T \tilde{v}_{it}\tilde{v}_{it}' \right)^{-1}$ with $\tilde{v}_{it} = (\tilde{\epsilon}_{it-1}, \tilde{\epsilon}_{it-2}, \ldots, \tilde{\epsilon}_{it-p_i})'$ and $\tilde{\epsilon}_{it} = z_{it} - \hat{\lambda}_i' \tilde{f}_i$. Matrix $Q$ can be consistently estimated by $\frac{1}{N} \sum_{i=1}^N \frac{1}{\phi_i} \hat{\lambda}_i \hat{\lambda}_i'$.

## 5 Modeling the dynamics in $f_t$ and $e_{it}$

We have considered estimating the dynamics in $f_t$ and in $e_{it}$ in separation. Now we consider estimating both dynamics simultaneously. The model is  
\[
z_{it} = \alpha_i + \lambda_i' f_t + e_{it}
\]
\[
f_t = \Psi_1 f_{t-1} + \Psi_2 f_{t-2} + \cdots + \Psi_K f_{t-K} + u_t.
\]
\[
e_{it} = \rho_{i,1} e_{it-1} + \cdots + \rho_{i,p_i} e_{it-p_i} + \epsilon_{it}, \quad \text{(12)}
\]
Stock and Watson (2005) study the estimation of similar models. A further more general model is considered by Forni et al. (2000) with estimation by the frequency domain approach.

5.1 A three-step procedure

The estimation procedure consists of

1. use the QML method to obtain \( \hat{\Lambda}, \hat{F} \) and \( \hat{\Phi} \);

2. use the method in Section 4 to obtain \( \tilde{\Lambda}, \tilde{F} \) and \( (\tilde{\rho}_{i,1}, \ldots, \tilde{\rho}_{i,p_i}) \) \( (i = 1, 2, \ldots, N) \);

3. regress \( \tilde{f}_t \) on its lags to obtain \( \tilde{\Psi}_1, \ldots, \tilde{\Psi}_K \).

Only the last step is new. The last step does not update other parameters. This is reasonable because modeling the dynamics in \( f_t \) does not improve the efficiency of estimated factor loadings and \( \rho \). The limiting distributions for \( \hat{\Lambda}, \hat{F} \) and \( \hat{\rho} = (\hat{\rho}_{i,1}, \ldots, \hat{\rho}_{i,p_i}) \) for \( i = 1, 2, \ldots, N \) are already given in Section 4. It can be shown that the limiting distribution for \( (\tilde{\Psi}_1, \ldots, \tilde{\Psi}_K) \) is the same as if \( f_t \) were observable if \( \sqrt{T/N} \to 0 \).

5.2 Simultaneous estimation

Model (12) can also be jointly estimated by the full maximum likelihood method, which can be implemented by the EM algorithm of Dempster et al. (1977). Based on the work of Watson and Engle (1983) and Wu (1983), Quah and Sargent (1992) explain the feasibility of the EM algorithm for high dimensional data. Jungbacker and Koopman (2008) propose a transformation that aims to reduce the dimensionality of the computation. Reis and Watson (2010) estimate a similar model to study the price changes in consumption goods in the United States.

Joint estimation is obtained by putting the model in the state space form. For simplicity, assume that the idiosyncratic errors and the factors are \( AR(1) \) such that 
\[
e_{it} = \rho_t e_{i,t-1} + \epsilon_{it} \quad \text{and} \quad f_t = \Psi f_{t-1} + u_t.
\]

Then the model can be written as
\[ z_t - \rho z_{t-1} = [\Lambda, -\rho \Lambda] S_t + \epsilon_t \]
\[ S_t = \begin{bmatrix} \Psi & 0 \\ I_r & 0 \end{bmatrix} S_{t-1} + \begin{bmatrix} u_t \\ 0 \end{bmatrix} \]  
\hfill (13)

where \( S_t = (f_t', f_{t-1}')' \) and \( \rho = \text{diag}(\rho_1, \ldots, \rho_N) \). The above is in a standard state-space form with implied parameter restrictions. Supplement F contains the updating formulas for the EM algorithm. In our simulation study, the resulting estimator is called ML-EM. This type of estimator has been considered by Reis and Watson (2010), and readers are referred to their paper for details.

The limiting distribution for the jointly estimated \((\Lambda, \rho_1, \ldots, \rho_N)\) is conjectured to be the same as the two-step estimator in Section 4. This follows from an intuitive explanation and from the Monte Carlos simulations. The performance of the full MLE is superior than that of the two-step estimator, but inferior than infeasible estimator when \( f_t \) is treated as known. Further, the two-step estimator has the same limiting distribution as if \( f_t \) were observable. The estimated state variable \( f_t \) should also have the same limiting distribution as in Section 4. However, to rigorously justify these limiting distributions does not appear to be trivial. We leave this as a future research topic.

### 6 Simulation results

We conduct Monte Carlo simulations to examine the finite sample properties of QMLE, ML-GLS, iterated ML-GLS and ML-EM estimators. We also compute the PC, PC-GLS and iterated PC-GLS for comparison. Overall, simulations show that the estimators in the ML class outperform those in the PC class. The better performance of the ML estimators becomes less pronounced when \( N \) is large. Standard QMLE performs better than the standard PC even for large \( N \). Supplement G contains the details.
7 Conclusion

This paper develops an inferential theory for the likelihood-based estimators of approximate factor models under high dimension. The idiosyncratic errors in the model exhibit heteroscedasticity and correlations of unknown forms over the cross sections and over the time dimension. Various identification conditions are considered. We show that the likelihood based estimators are consistent; we also derive the rates of convergence and the limiting distributions. Monte Carlo simulations show that the likelihood method is easy to implement and the ML-type estimators are more efficient than the PC-type estimators when the number of cross-section is small relative to the number of observations.

Appendix: Proof of Theorem 3

To prove Theorem 3, we need additional results. Let 
\[ \hat{G} = \left( \hat{\Sigma}^{-1} + I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda}) \right)^{-1}. \]
Hereafter, we use \( \|M\|_2 \) to denote the operator norm of matrix \( M \), i.e., \( \|M\|_2 = \inf \{ C, \|Mv\| \leq C\|v\| \text{ for all } v \} \). We also use \( \lambda_{max}(M) \) to denote the largest eigenvalue of the matrix \( M \). It is well known that \( \|M\|_2^2 = \lambda_{max}(M'M) \). Throughout the appendix, \( (\hat{\Lambda}, \hat{F}, \hat{\Phi}) \) denote the QMLE estimation of Section 2. The following lemma will be used in our derivation.

**Lemma A.1** Under Assumptions A’ and B-E,

\[
\begin{align*}
(a) & \quad \frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda} - \frac{1}{N} \Lambda' \Phi^{-1} \Lambda = o_p(1) \\
(b) & \quad A = (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} = O_p(N^{-1}) + O_p(T^{-1})
\end{align*}
\]

For the proof of Lemma A.1, see Corollaries S.1 and S.4.

**Lemma A.2** Under Assumptions A’ and B-E,

\[
\begin{align*}
(a) & \quad \| (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1} \|_2 = O_p(N^{-1/2}) \\
(b) & \quad \| (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Phi}^{-1} - (\Lambda' \Phi^{-1} \Lambda)^{-1} \Lambda' \Phi^{-1} \|_2 = O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1/2})
\end{align*}
\]
Proof of Lemma A.2: Consider (a). For notational simplicity, we use \( \hat{H} = (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \) and \( H = (\Lambda' \Phi^{-1} \Lambda)^{-1} \). Notice \( \hat{\Phi}^{-1} \leq C^2 I_N \), thus

\[
\| \hat{H} \Lambda' \hat{\Phi}^{-1} \|^2 = \lambda_{\text{max}} \left( \hat{H} \Lambda' \hat{\Phi}^{-2} \Lambda \hat{H} \right) \leq C^2 \lambda_{\text{max}} \left( \hat{H} \Lambda' \hat{\Phi}^{-1} \Lambda \hat{H} \right) = C^2 \lambda_{\text{max}}(\hat{H}) = O_p(N^{-1}).
\]

Consider (b). The left hand side is equal to \( \| \hat{H} \Lambda' \hat{\Phi}^{-1} - H \Lambda' \Phi^{-1} \|_2 \), which is further bounded by

\[
\| \hat{H} \Lambda' \hat{\Phi}^{-1} - H \Lambda' \Phi^{-1} \|_2 \leq \| (\hat{H} - H) \Lambda' \hat{\Phi}^{-1} \|_2 + \| H(\Lambda' \hat{\Phi}^{-1} - \Lambda' \Phi^{-1}) \|_2 \tag{A.1}
\]

\[
\leq \| \hat{H} - H \|_2 \cdot \| \Lambda' \hat{\Phi}^{-1} \|_2 + \| H \|_2 \cdot \| (\hat{\Lambda} - \Lambda) \Lambda' \Phi^{-1} \|_2 + \| H \|_2 \cdot \| \Lambda'(\hat{\Phi}^{-1} - \Phi^{-1}) \|_2.
\]

Consider the first term. Notice \( \| \hat{H} - H \|_2 = \| \hat{H}(\Lambda' \Phi^{-1} \Lambda - \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda}) H \|_2 \leq \| \hat{H} \|_2 \cdot \| \hat{\Lambda}' \hat{\Phi}^{-1} \Lambda - \Lambda' \Phi^{-1} \Lambda \|_2 \cdot \| H \|_2 \), where the first equality uses the definitions of \( \hat{H} \) and \( H \). Notice

\[
\frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda} - \frac{1}{N} \Lambda' \Phi^{-1} \Lambda = \frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} + \frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) - \frac{1}{N} (\hat{\Lambda} - \Lambda)' \Phi^{-1} (\hat{\Lambda} - \Lambda) + \frac{1}{N} \Lambda'(\hat{\Phi}^{-1} - \Phi^{-1}) \Lambda
\]

Corollary S.2 in the supplement implies \( \frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} = O_p(N^{-1}) + O_p(T^{-1/2}) \). Following the discussion below (S.19) of the supplement, \( \frac{1}{N} \Lambda'(\hat{\Phi}^{-1} - \Phi^{-1}) \Lambda = O_p(T^{-1/2}) + O_p(N^{-1}) \). Given these results, together with Proposition 1, we have \( \frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda} - \frac{1}{N} \Lambda' \Phi^{-1} \Lambda = O_p(N^{-1}) + O_p(T^{-1/2}) \). So \( \| \hat{H} - H \|_2 = O_p(N^{-2}) + O_p(N^{-1} T^{-1/2}) \). However, \( \| \hat{\Lambda}' \hat{\Phi}^{-1} \|^2 = \lambda_{\text{max}}(\hat{\Lambda}' \hat{\Phi}^{-2} \hat{\Lambda}) \leq C^2 \lambda_{\text{max}}(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda}) = O_p(N) \). This implies \( \| \hat{H} - H \|_2 \cdot \| \hat{\Lambda}' \hat{\Phi}^{-1} \|_2 = O_p(N^{-3/2}) + O_p(N^{-1/2} T^{-1/2}) \). Consider the second term of (A.1). Notice

\[
\frac{1}{N} \| (\hat{\Lambda} - \Lambda)' \Phi^{-1} \|^2 = \lambda_{\text{max}} \left( \frac{1}{N} (\hat{\Lambda} - \Lambda)' \Phi^{-2} (\hat{\Lambda} - \Lambda) \right) \leq C^2 \lambda_{\text{max}} \left( \frac{1}{N} (\hat{\Lambda} - \Lambda)' \Phi^{-1} (\hat{\Lambda} - \Lambda) \right)
\]

So the second term is \( O_p(N^{-3/2}) + O_p(N^{-1/2} T^{-1/2}) \) by Proposition 1 and \( \| H \|_2 = O(N^{-1}) \). Consider the last term of (A.1). Notice

\[
\frac{1}{N} \lambda'(\hat{\Phi}^{-1} - \Phi^{-1})^2 = \lambda_{\text{max}} \left( \frac{1}{N} \Lambda'(\hat{\Phi}^{-1} - \Phi^{-1})^2 \Lambda \right)
\]

24
The expression in the parentheses is equal to \( \frac{1}{N} \sum_{i=1}^{N} \frac{(\hat{\phi}_i^2 - \phi_i^2)^2}{\hat{\phi}_i^2 \phi_i^2} \lambda_i \gamma_i \), which is bounded by \( C^{10} \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \), and thus is \( O_p(N^{-2}) + O_p(T^{-1}) \) by Proposition 1. So the third term of (A.1) is \( O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1/2}) \). These results imply (b). □

Lemma A.3 Under Assumptions A’ and B-E,

(a) \( \| \hat{\mathcal{G}} \|_2 = O_p(N^{-1}) \), \( \| \hat{\mathcal{S}}_\mathcal{F} \|_2 = O_p(1) \), \( \| \hat{\mathcal{S}}^{-1}_\mathcal{F} \|_2 = O_p(1) \);
(b) \( \| \Sigma_{\mathcal{F}}^{-1} - \hat{\Sigma}_{\mathcal{F}}^{-1} \|_2 = O_p(N^{-1}) + O_p(T^{-1/2}) \)

Lemma A.3 is proved by Doz et al. (2011b). □

Proof of Theorem 3: Using \((A + B)^{-1} = B^{-1} - (A + B)^{-1}AB^{-1} \), we have

\[
\hat{\mathcal{G}} \equiv \left( \hat{\mathcal{S}}_{\mathcal{F}}^{-1} + I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda}) \right)^{-1} = I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} - \hat{\mathcal{G}} \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \left( I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \right)
\]

(A.2)

So we have

\[
\hat{f}_t^{ks} = \hat{f}_t + (v_t' \otimes I_r) \hat{\mathcal{G}} \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \left[ I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda}) \right]^{-1} \left[ I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1}) \right] \mathcal{Z}.
\]

(A.3)

where \( \hat{f}_t \) is the GLS estimator considered in Subsection 2.5. We analyze the second expression above. From \( \mathcal{Z} = (I_T \otimes \Lambda) \mathcal{F} + \mathcal{E} \), we have

\[
(v_t' \otimes I_r) \hat{\mathcal{G}} \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \left[ I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda}) \right]^{-1} \left[ I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1}) \right] \mathcal{Z} = (v_t' \otimes I_r) \hat{\mathcal{G}} \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \left[ I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}] \right] \mathcal{F}
\]

\[
+ (v_t' \otimes I_r) \hat{\mathcal{G}} \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \left( I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}] \right) \mathcal{E} = IG_1 + IG_2, \quad \text{say}
\]

To take into account of the many zeros in \( v_t' \otimes I_r \), we split \( IG_1 \) into

\[
(v_t' \otimes I_r) [I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \left( I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}] \right) \mathcal{F}
\]

\[
- (v_t' \otimes I_r) \hat{\mathcal{G}} \hat{\mathcal{S}}_{\mathcal{F}}^{-1} [I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \left( I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}] \right) \mathcal{F} = IG_3 - IG_4
\]

By \( \|AB\|_2 \leq \|A\|_2 \|B\|_2 \), \( IG_4 \) is bounded by

\[
\|IG_4\| \leq \| (v_t' \otimes I_r) \|_2 \cdot \| \hat{\mathcal{G}} \|_2 \cdot \| \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \|_2 \cdot \| [I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \|_2 \times \| \hat{\mathcal{S}}_{\mathcal{F}}^{-1} \|_2 \cdot \| I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}] \|_2 \cdot \| \mathcal{F} \|,
\]

25
which is $O_p(T^{1/2}N^{-2})$ by Lemmas A.1 and A.3. Now consider $IG_3$, which is 
\[ [v_t' \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \Sigma_{\mathcal{F}}^{-1} \mathcal{F} + [v_t' \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] (\hat{\Sigma}_{\mathcal{F}}^{-1} - \Sigma_{\mathcal{F}}^{-1}) \mathcal{F} - [v_t' \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \hat{\Sigma}_{\mathcal{F}}^{-1} (I_T \otimes A') \mathcal{F} \]

The second term of the above expression is $O_p(T^{1/2}N^{-2}) + O_p(N^{-1})$ and the third term is $O_p(T^{1/2}N^{-2}) + O_p(N^{-1})$ by Lemmas A.1 and A.3. Consider the first term. Notice $\text{var}(\Sigma_{\mathcal{F}}^{-1} \mathcal{F}) = \Sigma_{\mathcal{F}}^{-1}$, so each element of $\Sigma_{\mathcal{F}}^{-1} \mathcal{F}$ is $O_p(1)$. By the definition of $v_t$ and $(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} = O_p(N^{-1})$, the first term is $O_p(N^{-1})$. So $IG_3 = O_p(N^{-1}) + O_p(T^{1/2}N^{-2})$. Given the results on $IG_3$ and $IG_4$, we have $IG_1 = O_p(N^{-1}) + O_p(T^{1/2}N^{-2})$.

Consider $IG_2$, by (A.2), which is equal to
\[
(v_t' \otimes I_r)[I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \Sigma_{\mathcal{F}}^{-1} (I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}]) \mathcal{E}
\]
\[ + (v_t' \otimes I_r)[I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] (\hat{\Sigma}_{\mathcal{F}}^{-1} - \Sigma_{\mathcal{F}}^{-1}) (I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}]) \mathcal{E}
\]
\[ - (v_t' \otimes I_r) \hat{\Sigma}_{\mathcal{F}}^{-1} (I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}) (I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}]) \mathcal{E} = IG_5 + IG_6 - IG_7
\]

However,
\[
\|IG_6\| \leq \|v_t' \otimes I_r\|_2 \cdot \|I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}\|_2 \cdot \|\hat{\Sigma}_{\mathcal{F}}^{-1} - \Sigma_{\mathcal{F}}^{-1}\|_2 \cdot \|I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}]\|_2 \cdot \|\mathcal{E}\|
\]
which is $O_p(N^{-3/2}) + O_p(N^{-5/2}T^{1/2})$ by Lemmas A.1-A.3. Similarly,
\[
\|IG_7\| \leq \|v_t' \otimes I_r\|_2 \cdot \|\hat{\Sigma}_{\mathcal{F}}^{-1}\|_2 \cdot \|\|I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}\|_2 \cdot \|I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1}]\|_2 \cdot \|\mathcal{E}\|
\]
which is $O_p(N^{-5/2}T^{1/2})$. Now consider $IG_5$, which can be written as
\[
(v_t' \otimes I_r)[I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \Sigma_{\mathcal{F}}^{-1} (I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1} - (\Lambda' \Phi^{-1} \Lambda)^{-1} \Lambda' \Phi^{-1}]) \mathcal{E}
\]
\[ + (v_t' \otimes I_r)[I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \Sigma_{\mathcal{F}}^{-1} (I_T \otimes [(\Lambda' \Phi^{-1} \Lambda)^{-1} \Lambda' \Phi^{-1}]) \mathcal{E}
\]

The first term is bounded in norm by
\[
\|v_t' \otimes I_r\|_2 \cdot \|I_T \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}\|_2 \cdot \|\Sigma_{\mathcal{F}}^{-1}\|_2 \cdot \|I_T \otimes [(\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1} - (\Lambda' \Phi^{-1} \Lambda)^{-1} \Lambda' \Phi^{-1}]\|_2 \cdot \|\mathcal{E}\|
\]
which is $O_p(N^{-5/2}T^{1/2}) + O_p(N^{-3/2})$ by Lemma A.2(b). The second term is equal to
\[ [v_t' \otimes (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}] \mathcal{A}, \text{ where } \mathcal{A} = \Sigma_{\mathcal{F}}^{-1} (I_T \otimes [(\Lambda' \Phi^{-1} \Lambda)^{-1} \Lambda' \Phi^{-1}]) \mathcal{E}. \text{ It is easy to show }
\[ E(\mathcal{A}\mathcal{A}^t) = \Sigma_F^{-1}(I_T \otimes H)\Sigma_F^{-1} \]

Notice \( \lambda_{\text{max}}(\Sigma_F^{-1}(I_T \otimes H)\Sigma_F^{-1}) = O(N^{-1}). \) So we have \([\nu_t' \otimes (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}]\mathcal{A} = O_p(N^{-3/2}). \) It follows that \( IG_5 = O_p(N^{-3/2}) + O_p(N^{-5/2}T^{1/2}). \) The results on \( IG_5, IG_6 \) and \( IG_7 \) lead to \( IG_2 = O_p(N^{-3/2}) + O_p(N^{-5/2}T^{1/2}). \) Summarizing the results on \( IG_1 \) and \( IG_2, \) we have

\[ \hat{f}^{ks}_t = \hat{f}_t + O_p(N^{-1}) + O_p(T^{1/2}N^{-2}). \]

This proves Theorem 3. □

References


Online supplement to “Maximum likelihood estimation and inference for approximate factor models of high dimension”

This supplement contains the omitted technical proofs and simulation results. We first give three first order conditions of the MLE (see, e.g., Lawley and Maxwell (1971) and Bai and Li (2012)):

\[ \hat{\Lambda}'\hat{\Sigma}_{zz}^{-1}z = 0 \]  \hspace{1cm} (S.1)

\[ \text{diag}(\hat{\Sigma}_{zz}^{-1}) = \text{diag}(\hat{\Sigma}_{zz}^{-1}M_{zz}\hat{\Sigma}_{zz}^{-1}) \]  \hspace{1cm} (S.2)

\[ \hat{\Lambda}'\hat{\Sigma}_{zz}^{-1}\hat{\Lambda} = \hat{\Lambda}'\hat{\Sigma}_{zz}^{-1}M_{zz}\hat{\Sigma}_{zz}^{-1}\hat{\Lambda} \]  \hspace{1cm} (S.3)

where \( \hat{\Lambda}, \hat{M}_{ff}, \) and \( \hat{\Phi} \) denote the MLE and \( \hat{\Sigma}_{zz} = \hat{\Lambda}\hat{M}_{ff}\hat{\Lambda}' + \hat{\Phi} \). Condition (S.1) is derived from the partial derivatives with respect to \( \Lambda \), (S.2) is with respect to the diagonal elements of \( \Phi \), and (S.3) is with respect to \( M_{ff} \). Equation (S.3) can be obtained from (S.1) by post-multiplying \( \hat{\Sigma}_{zz}^{-1}\hat{\Lambda} \). So (S.3) is redundant. This redundancy arises from rotational indeterminacy, a well known fact for factor models.

Supplement A: Consistency and its proof

We start with an average consistency stated in the following proposition.

Proposition S.1 (Average consistency) Let \( \hat{\theta} \) be the solution by maximizing (3), where \( \hat{\theta} = (\hat{\lambda}_1, \cdots, \hat{\lambda}_N, \hat{\phi}_1^2, \cdots, \hat{\phi}_N^2, \hat{M}_{ff}) \). Under Assumptions A-D, when \( N,T \to \infty \), with any one of the identification conditions, we have

\[ \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \|\hat{\lambda}_i - \lambda_i\|^2 \overset{p}{\to} 0 \]
\[
\frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \xrightarrow{p} 0
\]
\[
\hat{M}_{ff} - M_{ff} \xrightarrow{p} 0
\]

where \( \phi_i^2 = \frac{1}{T} \sum_{t=1}^{T} E(\epsilon_{it}^2) = \frac{1}{T} \sum_{t=1}^{T} \tau_{ii,t} \).

To prove the proposition, we introduce some preliminary results and notations. Throughout, we define \( H = (\Lambda'\Phi^{-1}\Lambda)^{-1} \) and \( G = (M_{ff}^{-1} + \Lambda'\Phi^{-1}\Lambda)^{-1} \). Matrix algebra shows \( H = G(I - M_{ff}^{-1}G)^{-1} \). Let \( \hat{H} \) denote the estimated version, i.e., \( \hat{H} = (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1} \). Let \( \hat{G} \) be defined similarly. We also put \( H_N = N \cdot H \) and \( G_N = N \cdot G \). We first state several moment inequalities implied by the assumptions in the main text. These results will be used in the following proof.

Under Assumptions A and C.4, we have, for all \( i = 1, 2, \cdots, N \),
\[
E \left( \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t e_{it} \right\|^2 \right) \leq C \tag{S.4}
\]
\[
E \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t e_{it} \right\| \right) \leq C \tag{S.5}
\]
Furthermore, under Assumption C.5, we have, by taking \( i = j \),
\[
E \left[ \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\epsilon_{it}^2 - \phi_i^2) \right\| \right] \leq C \tag{S.6}
\]

To prove consistency, we need to distinguish three sets of parameters: the true parameters, the estimator, and the arguments of the likelihood function (input variables). We use a superscript "\*'" to denote the true parameters such that \( \theta^* = (\Lambda^*, \Phi^*, M_{ff}^*) \). Parameters without the superscript "\*'" denote the arguments of the likelihood function such that \( \theta = (\Lambda, \Phi, M_{ff}) \). The estimator is denoted by \( \hat{\theta} = (\hat{\Lambda}, \hat{\Phi}, \hat{M}_{ff}) \). Once consistency is established, we will remove the superscript "\*'" from the true parameters.
Lemma S.1 Let $Q$ be an $r \times r$ matrix satisfying

$$QQ' = I, \quad \text{and} \quad Q'VQ = D$$

where $V$ is a diagonal matrix with strictly positive and distinct elements, arranged in decreasing order, and $D$ is also diagonal. Then $Q$ must be a diagonal matrix with elements either $-1$ or $1$ and $V = D$.

Proof of Lemma S.1: See Bai and Li (2012). □

Let $\theta = (\Lambda, \Phi, M_{ff})$ and let $\Theta$ denote the parameter space such that $\Phi$ and $M_{ff}$ satisfy Assumption D.

Lemma S.2 Under Assumptions A-D, we have

$$\begin{align*}
(a) \quad & \sup_{\theta \in \Theta} \frac{1}{NT} \text{tr} \left[ \Lambda' \Sigma_{xx}^{-1} \sum_{t=1}^{T} e_t f_t' \right] \overset{p}{\to} 0 \\
(b) \quad & \sup_{\theta \in \Theta} \frac{1}{NT} \text{tr} \left[ \sum_{t=1}^{T} (e_t f_t' - \Omega_t') \Sigma_{xx}^{-1} \right] \overset{p}{\to} 0 \\
(c) \quad & \sup_{\theta \in \Theta} \frac{1}{N} \text{tr} \left[ \bar{e} \bar{e}' \Sigma_{xx}^{-1} \right] \overset{p}{\to} 0
\end{align*}$$

where $\theta^*$ is the true parameter, and $\Sigma_{xx} = \Lambda M_{ff} \Lambda' + \Phi$, depending on $\theta = (\Lambda, \Phi, M_{ff})$, and $\Omega_t^* = E(e_t e_t')$.

Proof of Lemma S.2: Notice that

$$\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} f_t' e_{it} \right)^2 = O_p(T^{-1}),$$

$$\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \phi_t^*)^2 \right)^2 = O_p(T^{-1}),$$

$$\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} (e_{it} e_{jt} - E(e_{it} e_{jt})) \right)^2 = O_p(T^{-1}).$$
The first result follows by (S.5). The second result follows by (S.6). The third result is implied by Assumption C.5.

Given the above three results, Lemma S.2 can be proved similarly as Lemma A.2 of Bai and Li (2012). □

**Lemma S.3** Under Assumptions A-D, for \( \theta = (\Lambda, \Phi, M) \), we have

\[
(a) \quad \sup_{\theta \in \Theta} \frac{1}{N} tr \left[ \frac{1}{T} \sum_{t=1}^{T} \Omega_t^* \Phi^{-1} \Lambda G \Phi^{-1} \right] = O_p(N^{-1}) = o_p(1)
\]

\[
(b) \quad \sup_{\theta \in \Theta} \frac{1}{N} tr \left[ \left( \frac{1}{T} \sum_{t=1}^{T} \Omega_t^* - \Phi^* \right) \Sigma^{-1} \right] = O_p(N^{-1}) = o_p(1)
\]

**Proof of Lemma S.3:** Consider (a). The left hand side of (a) can be written as

\[
\frac{1}{N} tr[A^T \frac{1}{T} \sum_{t=1}^{T} \Omega_t^* \Phi^{-1} \Lambda G],
\]

which, by the definition of \( \Omega_t^* \), is equivalent to

\[
\frac{1}{N} tr \left[ \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} H^{1/2} \lambda_i \lambda_j^T H^{1/2} \frac{1}{T} \sum_{t=1}^{T} \tau_{ij,t} (H^{1/2} M^{-1} H^{1/2} + I_r)^{-1} \right].
\]

Consider the term \( \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} H^{1/2} \lambda_i \lambda_j^T H^{1/2} \frac{1}{T} \sum_{t=1}^{T} \tau_{ij,t} \), which is bounded in norm by

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| \frac{1}{\phi_i^2} H^{1/2} \lambda_i \right\| \cdot \left\| \frac{1}{\phi_j^2} \lambda_j^T H^{1/2} \right\| \cdot \left\| \frac{1}{T} \sum_{t=1}^{T} \tau_{ij,t} \right\|.
\]

By the boundedness of \( \phi_i^2 \) and \( |\tau_{ij,t}| \leq \tau_{ij} \), the above term is bounded by

\[
C^2 \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| \frac{1}{\phi_i} H^{1/2} \lambda_i \right\| \cdot \left\| \frac{1}{\phi_j} \lambda_j^T H^{1/2} \right\| \tau_{ij}.
\]

Let \( \chi_i = \left\| \frac{1}{\phi_i} H^{1/2} \lambda_i \right\| \) and \( \chi = (\chi_1, \chi_2, \cdots, \chi_N)' \), the above term is equal to \( \frac{1}{N} C^2 \chi' T \chi \) with \( \|\chi\|^2 = \sum_{i=1}^{N} \chi_i^2 = \sum_{i=1}^{N} \|\frac{1}{\phi_i} H^{1/2} \lambda_i\|^2 = r \), where \( T \) is a \( N \times N \) matrix consisting of \( \tau_{ij} \). So the above term is bounded by \( C^2 r \frac{1}{N} \tau_{\text{max}} \), where \( \tau_{\text{max}} \) is the largest eigenvalue of the matrix \( T \). By Assumption C.3, \( \tau_{\text{max}} \leq C \). Then (a) follows.
Lemma S.4 Under Assumptions A-D, we have

\[ \sup_{\theta \in \Theta} \text{tr} \left[ \frac{1}{N} \left( \frac{1}{T} \sum_{t=1}^{T} \Omega_t^* - \Phi^* \right) (\Phi^{-1} - \Phi^{-1} \Lambda G \Lambda' \Phi^{-1}) \right]. \]

The term \( \text{tr} \left[ \frac{1}{N} \left( \frac{1}{T} \sum_{t=1}^{T} \Omega_t^* - \Phi^* \right) \Phi^{-1} \right] = 0 \) because the diagonal elements of \( \frac{1}{T} \sum_{t=1}^{T} \Omega_t^* - \Phi^* \) are all zero and \( \Phi \) is a diagonal matrix. The term \( \text{tr} \left[ \frac{1}{N} \left( \frac{1}{T} \sum_{t=1}^{T} \Omega_t^* \Phi^{-1} \Lambda G \Lambda' \Phi^{-1} \right) \right] = o_p(1) \) has already been proved by (a). It remains to prove \( \text{tr} \left[ \frac{1}{N} \Phi^* \Phi^{-1} \Lambda G \Lambda' \Phi^{-1} \right] = o_p(1) \) uniformly on \( \Theta \). Since the matrix \( \Phi^* \Phi^{-1} \) is bounded by \( C^4 I_N \), the term \( \text{tr} \left[ \frac{1}{N} \Phi^* \Phi^{-1} \Lambda G \Lambda' \Phi^{-1} \right] \) is bounded by \( C^4 \frac{1}{N} \text{tr} [\Lambda' \Phi^{-1} G] \). By the definition of \( G \), (b) follows. □

**Lemma S.4** Under Assumptions A-D, we have

(a) \( \hat{H} \hat{\Lambda} \Phi^{-1} \Lambda^* \left( \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t^* e_{jt} \right) = \| N^{1/2} \hat{H}^{1/2} \| \cdot O_p(T^{-1/2}) \), for each \( j \)

(b) \( \hat{H} \hat{\Lambda} \Phi^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} e_t f_t^* e_{jt} \right) = \| N^{1/2} \hat{H}^{1/2} \| \cdot O_p(T^{-1/2}) \)

(c) \( \hat{H} \left( \frac{1}{T} \sum_{t=1}^{T} e_t e_{jt} - E(e_t e_{jt}) \right) = \| N^{1/2} \hat{H}^{1/2} \| \cdot O_p(T^{-1/2}) \), for each \( j \)

(d) \( \hat{H} \left( \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \hat{\lambda}_i \sum_{t=1}^{T} e_{it} e_{jt} - E(e_{it} e_{jt}) \right) \hat{H} = \| N^{1/2} \hat{H}^{1/2} \|^2 \cdot O_p(T^{-1/2}) \)

**Proof of Lemma S.4:** This lemma can be proved similarly as Lemma A.3 in Bai and Li (2012). □

**Lemma S.5** Under Assumptions A-D, we have

(a) \( \hat{H} \hat{\Lambda} \Phi^{-1} \Phi^{-1} \hat{\lambda} \hat{H} = \| N^{1/2} \hat{H}^{1/2} \|^2 \cdot O_p(T^{-1}) \)

(b) \( \hat{H} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\hat{\phi}_i^2} \hat{\lambda}_i E(e_{it} e_{jt}) = \| \hat{H}^{1/2} \| \cdot O_p(1), \) for each \( j \)

(c) \( \hat{H} \hat{\Lambda} \Phi^{-1} \hat{\lambda} \hat{e}_{jt} = \| N^{1/2} \hat{H}^{1/2} \| \cdot O_p(T^{-1}), \) for each \( j \)

(d) \( \hat{H} \hat{\Lambda} \Phi^{-1} \hat{\phi} \hat{e}_{jt} = \| \hat{H} \| \cdot O_p(1) + \| N^{1/2} \hat{H}^{1/2} \|^2 \cdot O_p(N^{-1}) \)
**Proof of Lemma S.5:** Consider (a). The left hand side of (a) is bounded in norm by
\[
C^2 \| \hat{H}^{1/2} \|^2 \left( \sum_{i=1}^N \frac{1}{\hat{\phi}_i^2} \| \hat{H}^{1/2} \hat{\lambda}_j \|^2 \right) \left( \sum_{i=1}^N \left( \frac{1}{T} \sum_{t=1}^T e_{it} \right)^2 \right)
\]
Since \( \sum_{i=1}^N \frac{1}{\hat{\phi}_i^2} \| \hat{H}^{1/2} \hat{\lambda}_j \|^2 = r \), the above term is bounded by
\[
C^2 r \| \hat{H}^{1/2} \|^2 \left( \sum_{i=1}^N \left( \frac{1}{T} \sum_{t=1}^T e_{it} \right)^2 \right)
\]
which is \( \| N^{1/2} \hat{H}^{1/2} \| O_p(T^{-1}) \) because \( T^{-1} \sum_{t=1}^T e_{it} = O_p(T^{-1/2}) \).

Consider (b). The left hand side of (b) is equal to \( \hat{H} \sum_{i=1}^N \frac{1}{\hat{\phi}_i^2} \hat{\lambda}_i \frac{1}{T} \sum_{t=1}^T \tau_{ij, t} \), which is bounded in norm by
\[
C \| \hat{H}^{1/2} \| \cdot \sum_{i=1}^N \| \frac{1}{\hat{\phi}_i} \hat{H}^{1/2} \hat{\lambda}_i \| \tau_{ij}.
\]
By the Cauchy-Schwarz inequality,
\[
\sum_{i=1}^N \| \frac{1}{\hat{\phi}_i} \hat{H}^{1/2} \hat{\lambda}_i \| \tau_{ij} \leq \left( \sum_{i=1}^N \frac{1}{\hat{\phi}_i^2} \| \hat{H}^{1/2} \hat{\lambda}_i \|^2 \right)^{1/2} \left( \sum_{i=1}^N \tau_{ij}^2 \right)^{1/2} = \sqrt{T} \left( \sum_{i=1}^N \tau_{ij}^2 \right)^{1/2}.
\]
However, \( \sum_{i=1}^N \tau_{ij}^2 \leq C \sum_{i=1}^N \tau_{ij} \leq C^2 \) because \( \tau_{ij} \leq C \) and \( \sum_{i=1}^N \tau_{ij} \leq C \). Given this result, the above expression is \( O(1) \). Then (b) follows.

Consider (c). By (a), it follows that \( \| \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \tilde{e} \| = \| N^{1/2} \hat{H}^{1/2} \| \cdot O_p(T^{-1/2}) \). So (c) follows by \( \tilde{e}_j = O_p(T^{-1/2}) \) due to Assumption C.4.

Consider (d). The left hand side of (d) is equal to
\[
\hat{H} - \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \frac{1}{T} \sum_{t=1}^T \Omega_t^* \hat{\Phi}^{-1} \hat{\Lambda} \hat{H}.
\]
The first term is \( \| \hat{H} \| \cdot O_p(1) \). The second term can be proved to be \( \| N^{1/2} \hat{H}^{1/2} \|^2 \cdot O_p(N^{-1}) \), similarly as result (a) of Lemma S.3. \( \square \)
Proof of Proposition S.1: By $z_t = \alpha^* + \Lambda^* f_t^* + e_t$, it follows that

\[
M_{zz} = \Lambda^* M_{ff}^* \Lambda^* + \Phi^* + \frac{1}{T} \sum_{t=1}^{T} \Lambda^* f_t^* e_t' + \frac{1}{T} \sum_{t=1}^{T} e_t f_t'' \Lambda^*
+ \frac{1}{T} \sum_{t=1}^{T} (e_t e_t' - \Omega_t^*) + \left(\frac{1}{T} \sum_{t=1}^{T} \Omega_t^* - \Phi^*\right) - \bar{e}e'
\]

(S.7)

Let $\Sigma_{zz}(\theta^*) = \Lambda^* M_{ff}^* \Lambda^* + \Phi^*$. Furthermore, we define

\[
\bar{\mathcal{L}}(\theta) = -\frac{1}{2N} \ln |\Sigma_{zz}| - \frac{1}{2N} tr[\Sigma_{zz}(\theta^*) \Sigma_{zz}^{-1}] \\
R_1(\theta) = -\frac{1}{2N} tr\left[\left(\frac{1}{T} \sum_{t=1}^{T} \Lambda^* f_t^* e_t' + \frac{1}{T} \sum_{t=1}^{T} e_t f_t'' \Lambda^* + \frac{1}{T} \sum_{t=1}^{T} (e_t e_t' - \Omega_t^*)\right) \Sigma_{zz}^{-1}\right] \\
R_2(\theta) = -\frac{1}{2N} tr\left[\left(\frac{1}{T} \sum_{t=1}^{T} \Omega_t^* - \Phi^*\right) - \bar{e}e'\right] \Sigma_{zz}^{-1}
\]

Then the likelihood function can be written as

\[
L(\theta) = \bar{\mathcal{L}}(\theta) + R(\theta)
\]

where $R(\theta) = R_1(\theta) + R_2(\theta)$. Lemma S.2 and Lemma S.3 imply that $\sup_\theta |R_1(\theta)| = o_p(1)$ and $\sup_\theta |R_2(\theta)| = o_p(1)$. Thus $\sup_{\theta \in \Theta} |R(\theta)| = o_p(1)$. So the present objective function has the same properties as that of Proposition 5.1 in Bai and Li (2012). Using their arguments, we have

\[
\frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \overset{p}{\rightarrow} 0
\]

(S.8)

\[
\hat{G} = o_p(1); \quad \hat{H} = o_p(1)
\]

(S.9)

In addition, let $A = (\hat{\Lambda} - \Lambda^*)' \hat{\Phi}^{-1} \hat{\Lambda}(\hat{\Phi}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1}$, then

\[
\frac{1}{N} \Lambda^* \Phi^{-1} \Lambda^* - (I_r - A) \left(\frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda}\right) (I_r - A)' \overset{p}{\rightarrow} 0
\]

(S.10)
and
\[
\frac{1}{N}(\hat{\Lambda} - \Lambda^*)'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda^*) - A \left( \frac{1}{N} \hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda} \right) A' \xrightarrow{\text{ } T\rightarrow\infty} 0 \quad (S.11)
\]

Now we turn to the first order conditions. The \(j\)th column of the first order condition (S.1) implies
\[
\hat{\lambda}_j - \lambda_j^* = -\hat{M}^{-1}_{ff} \hat{H} \hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda^*)M^*_j \lambda_j^* - \hat{M}^{-1}_{ff} (\hat{M} - M^*_f) \lambda_j^*
\]
\[
+ \hat{M}^{-1}_{ff} \hat{H} \Lambda'\hat{\Phi}^{-1} \Lambda^* \frac{1}{T} \sum_{t=1}^{T} f^*_t e_{jt} + \hat{M}^{-1}_{ff} \hat{H} \Lambda'\hat{\Phi}^{-1} \frac{1}{T} \sum_{t=1}^{T} e_t f^*_t \lambda_j^*
\]
\[
+ \hat{M}^{-1}_{ff} \hat{H} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \hat{\lambda}_i \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - \hat{E}(e_{it} e_{jt})] - \hat{M}^{-1}_{ff} \hat{H} \hat{\lambda}_j
\]
\[
+ \hat{M}^{-1}_{ff} \hat{H} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^2} \hat{\lambda}_i E(e_{it} e_{jt}) - \hat{M}^{-1}_{ff} \hat{H} \hat{\Lambda}'\hat{\Phi}^{-1} \bar{e}\bar{e}
\]
(S.12)

The first order condition for \(M_{ff}\) in (S.2) implies
\[
\dot{M}_{ff} - M^*_f = -\dot{H} \dot{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda^*)M^*_f - M^*_f (\hat{\Lambda} - \Lambda^*)'\hat{\Phi}^{-1}\dot{\Lambda}\dot{H}
\]
\[
+ \hat{H} \Lambda'\hat{\Phi}^{-1}(\Lambda - \Lambda^*)M^*_f (\Lambda - \Lambda^*)'\hat{\Phi}^{-1}\dot{\Lambda}\dot{H} - \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} \bar{e}\bar{e}'\dot{\Phi}^{-1}\dot{\Lambda}\dot{H}
\]
\[
+ \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} \Lambda^* \frac{1}{T} \sum_{t=1}^{T} f^*_t e_{jt} \dot{\Phi}^{-1} \dot{\Lambda}\dot{H} + \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} \frac{1}{T} \sum_{t=1}^{T} e_t f^*_t \Lambda^* \dot{\Phi}^{-1} \dot{\Lambda}\dot{H}
\]
\[
+ \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} \frac{1}{T} \sum_{t=1}^{T} (e_t e_{jt}' - \Omega^*_{jt}) \dot{\Phi}^{-1} \dot{\Lambda}\dot{H} - \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} (\dot{\Phi} - \frac{1}{T} \sum_{t=1}^{T} \Omega^*_{jt}) \dot{\Phi}^{-1} \dot{\Lambda}\dot{H}
\]
(S.13)

Substituting (S.13) into (S.12), we have
\[
\hat{\lambda}_j - \lambda_j^* = \hat{M}^{-1}_{ff} M^*_f (\hat{\Lambda} - \Lambda^*)'\hat{\Phi}^{-1} \dot{\Lambda}\dot{H}\lambda_j^* - \hat{M}^{-1}_{ff} \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda^*)M^*_f (\Lambda - \Lambda^*)'\hat{\Phi}^{-1} \dot{\Lambda}\dot{H}\lambda_j^*
\]
\[
- \hat{M}^{-1}_{ff} \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} \Lambda^* (\frac{1}{T} \sum_{t=1}^{T} f^*_t e_{jt}' \dot{\Phi}^{-1} \dot{\Lambda}\dot{H}\lambda_j^* - \hat{M}^{-1}_{ff} \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} (\frac{1}{T} \sum_{t=1}^{T} e_t f^*_t) \Lambda^* \dot{\Phi}^{-1} \dot{\Lambda}\dot{H}\lambda_j^*)
\]
\[
- \hat{M}^{-1}_{ff} \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} \frac{1}{T} \sum_{t=1}^{T} (e_t e_{jt}' - \Omega^*_{jt}) \dot{\Phi}^{-1} \dot{\Lambda}\dot{H}\lambda_j^* + \hat{M}^{-1}_{ff} \hat{H} \dot{\Lambda}'\hat{\Phi}^{-1} \bar{e}\bar{e}' \dot{\Phi}^{-1} \dot{\Lambda}\dot{H}\lambda_j^*
\]
\[ +\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}\hat{\Phi}^{-1}(\hat{\Phi} - \frac{1}{T}\sum_{t=1}^{T}\Omega_{tt}^{*})\hat{\Phi}^{-1}\hat{\Lambda}\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f^{*}_{t}e_{jt} + \hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f^{*}_{t}e_{jt} \]

\[ +\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_{t}f^{*}_{t}\lambda^{*}_{j} + \hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_{t}e_{jt} - E(e_{t}e_{jt})] \]

\[ -\hat{M}_{ff}^{-1}\hat{H}\hat{\lambda}_{j} + \hat{M}_{ff}^{-1}\hat{H}\frac{1}{T}\sum_{t=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_{tt}^{2}}\lambda_{t}E(e_{it}e_{jt}) - \hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}\hat{\Phi}^{-1}\bar{e}\bar{e}_{j} \]  
(S.14)

Consider (S.13). The sixth term on the right of (S.13) can be written as

\[ \hat{H}\hat{\Lambda}\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_{t}f^{*}_{t} - \hat{H}\hat{\Lambda}\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_{t}f^{*}_{t}A \]

where \( A = (\hat{\Lambda} - \Lambda^{*})\hat{\Phi}^{-1}\Lambda\hat{H} \). The first term of the above is \( \|N^{1/2}\hat{H}^{1/2}\| \cdot O_{p}(T^{-1/2}) \) by Lemma S.4(b) and the second term is \( A \cdot \|N^{1/2}\hat{H}^{1/2}\| \cdot O_{p}(T^{-1/2}) \). The fifth term of (S.13) is the transpose of the sixth. The last term is governed by Lemma S.5(d). The fourth term is governed by Lemma S.5(a). These results together with (S.9) imply that, in terms of \( A \),

\[ \hat{M}_{ff} - M_{ff}^{*} = -A'M_{ff}^{*} - M_{ff}^{*}A + A'M_{ff}^{*}A - A \cdot \|N^{1/2}\hat{H}^{1/2}\| \cdot O_{p}(T^{-1/2}) \]  
(S.15)

\[ +\|N^{1/2}\hat{H}^{1/2}\| \cdot O_{p}(T^{-1/2}) + \|N^{1/2}\hat{H}^{1/2}\|^{2} \cdot [O_{p}(T^{-1/2}) + O_{p}(N^{-1})] + o_{p}(1) \]

By the definition of \( \hat{H} \), \( N\hat{H} = (\frac{1}{N}\hat{\Lambda}\hat{\Phi}^{-1}\Lambda)^{-1} \). Equation (S.10) implies \( (\frac{1}{N}\hat{\Lambda}\hat{\Phi}^{-1}\Lambda)^{-1} = (I_{r} - A)'(\frac{1}{N}\Lambda^{*}\Phi^{*}\Lambda)^{-1}(I_{r} - A) + o_{p}(\|I_{r} - A\|^{2}) \). So we have

\[ \|N^{1/2}\hat{H}^{1/2}\|^{2} = tr[N\hat{H}] = tr[(I_{r} - A)'(\frac{1}{N}\Lambda^{*}\Phi^{*}\Lambda)^{-1}(I_{r} - A) + o_{p}(\|I_{r} - A\|^{2})] \]

These results imply that matrix \( A \) is stochastically bounded. To see this, the left hand side of (S.15) is stochastically bounded by Assumption D. If \( A \) is not stochastically bounded, the right hand side is dominated by \( A'M_{ff}^{*}A \), which will be unbounded since \( M_{ff}^{*} \) is positive definite. Thus a contradiction is obtained. It follows that
\[ A = O_p(1), \text{ and hence } \| N^{1/2} \hat{H}^{1/2} \| = O_p(1) \text{ by the preceding equation. From this, we have, by (S.15),} \]

\[ \hat{M}_{ff} - M_{ff} = -A'M_{ff}^* - M_{ff}^*A + A'M_{ff}^*A + o_p(1) \]  

(S.16)

Next consider (S.14). The last two terms are all \( o_p(1) \) by Lemma S.5 and \( \| N^{1/2} \hat{H}^{1/2} \| = O_p(1) \). The third from the last term can be written as \( \hat{\phi}_i \hat{M}_{ff}^{-1} \hat{H}^{1/2} \hat{H}^{1/2} \frac{1}{\hat{\phi}_i} \hat{\lambda}_i \), which is bounded in norm by \( C^2 \| \hat{H}^{1/2} \| \cdot \| \frac{1}{\hat{\phi}_i} \hat{H}^{1/2} \hat{\lambda}_i \| \) due to the boundedness of \( \hat{\phi}_i \) and \( \hat{M}_{ff} \). This term is further bounded by \( \sqrt{r} C^2 \| \hat{H}^{1/2} \| \sum_{i=1}^N \| \frac{1}{\hat{\phi}_i} \hat{H}^{1/2} \hat{\lambda}_i \|^2 = r \). So the third from the last term is \( o_p(1) \) by (S.9). The 3rd-10th terms are summarized in Lemmas S.4 and S.5 and they are all \( o_p(1) \) due to \( \| N^{1/2} \hat{H}^{1/2} \| = O_p(1) \). Thus we can express (S.14) as

\[ \hat{\lambda}_j - \lambda_j^* = \hat{M}_{ff}^{-1} M_{ff}^* A \lambda_j^* - \hat{M}_{ff}^{-1} A'M_{ff}^*A \lambda_j^* + o_p(1) \]  

(S.17)

Results (S.16) and (S.17), together with the identification conditions, imply \( A = (\hat{\Lambda} - \Lambda^*)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{\Phi} \rightarrow 0 \), as is shown by Bai and Li (2012). With \( A \rightarrow 0 \), equation (S.11) implies \( \frac{1}{N}(\hat{\Lambda} - \Lambda^*)' \hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda^*) = o_p(1) \), which is the first part of Proposition S.1. Moreover, (S.16) implies that \( \hat{M}_{ff} - M_{ff} = o_p(1) \), which is the last part of Proposition S.1. This completes the proof of the proposition. \( \square \)

**Corollary S.1** *Under Assumptions A-D, irrespective which set of identification conditions, we have*

(a) \[ \frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda} - \frac{1}{N} \Lambda^* \Phi^*^{-1} \Lambda^* = o_p(1) \]

(b) \[ \hat{H} = O_p(N^{-1}), \hat{H}_N = O_p(1), \hat{G} = O_p(N^{-1}), \hat{G}_N = O_p(1) \]

(c) \[ \frac{1}{N} (\hat{\Lambda} - \Lambda^*)' \hat{\Phi}^{-1} \hat{\Lambda} = o_p(1) \]
Proof of Corollary S.1: Irrespective which identification conditions, we have $A = (\hat{\Lambda} - \Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = o_p(1)$. Part (a) follows from (S.10). Result (a) implies that $\hat{H} = O_p(N^{-1})$ since $N^{-1}\Lambda^*\Phi^* - 1\Lambda^* \rightarrow Q > 0$ by Assumption B. It follows $\hat{H}_N = N \cdot \hat{H} = o_p(1)$. The claims on $\hat{G}$ follows from the relationship between $\hat{G}$ and $\hat{H}$. Part (c) follows from $A = o_p(1)$ and $N\hat{H}$ has a positive limit since $N\hat{H} = (\frac{1}{N}\Lambda^*\Phi^* - 1\Lambda^*)^{-1} + o_p(1)$ by part (a). □

Supplement B: Proof of the convergence rate

Having established consistency, we drop the superscript “*” from the true parameters for notational simplicity (there is no need to carry them). Any element without a hat denotes the true element from the model. We focus on the aspects that call for different analysis from the exact factor models in previous literature.

Lemma S.6 Under Assumptions A-D,

(a) $\|\hat{H}\Lambda'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda)\| = O_p\left(\left[\frac{1}{N}\sum_{i=1}^{N} \frac{1}{\phi_i^2}\|\hat{\lambda}_i - \lambda_i\|^2\right]^{1/2}\right)$

(b) $\left\| \hat{H}\Lambda'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t f_t' \right\| = O_p(T^{-1/2})$

Lemma S.7 Under Assumptions A-D:

(a) $\frac{1}{N}\sum_{j=1}^{N} \|\hat{H}\Lambda'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_{jt}\|^2 = O_p(T^{-1})$

(b) $\frac{1}{N}\sum_{j=1}^{N} \|\hat{H}\left(\sum_{i=1}^{N} \frac{1}{\phi_i^2}\hat{\lambda}_i\frac{1}{T}\sum_{t=1}^{T} [e_{it}e_{jt} - E(e_{it}e_{jt})]\right)\|^2 = O_p(T^{-1})$

(c) $\left\| \hat{H}\left(\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2}\phi_j^2\hat{\lambda}_i\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T} [e_{it}e_{jt} - E(e_{it}e_{jt})]\right)\hat{H}\right\|^2 = O_p\left(T^{-1}\right)$
Lemma S.8 Under Assumptions A-D:

\[(a) \left\| \hat{H} \hat{\Lambda}^{-1} \hat{\Phi}^{-1} \frac{1}{T} \sum_{t=1}^{T} f_t \xi'_t \right\|^2 = O_p(T^{-1}) \]

\[(b) \left\| \hat{H} \left( \sum_{i=1}^{N} \frac{1}{\phi'_i} \hat{\lambda}_i \sum_{t=1}^{T} [e_{it} \xi'_t - E(e_{it} \xi'_t)] \right) \right\|^2 = O_p(T^{-1}) \]

where \( \xi'_t = (e_{1t}, e_{2t}, \ldots, e_{rt}) \).

The above three lemmas can be proved similarly as Lemmas B1, B2, and B3 of Bai and Li (2012). So the detailed proofs are omitted. We need an additional lemma to establish Proposition S.2 given below.

Lemma S.9 Let \( \mathcal{E} = \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi'_i} \| \hat{\lambda}_i - \lambda_i \|^2 \right]^{1/2} \). Under Assumptions A-D, we have

\[(a) \hat{H} \hat{\lambda}' \hat{\Phi}^{-1} (\hat{\Phi} - \frac{1}{T} \sum_{t=1}^{T} \Omega_t) \hat{\Phi}^{-1} \hat{\lambda} \hat{H} = O_p(N^{-1}) \]

\[(b) \hat{H} \hat{\lambda}' \hat{\Phi}^{-1} \hat{\Phi}^{-1} \hat{\lambda} \hat{H} = O_p(T^{-1}) \]

\[(c) \hat{H} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi'_i} \hat{\lambda}_i E(e_{it} \xi'_t) = O_p(N^{-1}) + \mathcal{E} \cdot O_p(N^{-1/2}) \]

\[(d) \hat{H} \hat{\lambda}' \hat{\Phi}^{-1} \hat{\lambda} \mathcal{E} = O_p(T^{-1}) \]

\[(e) \hat{H} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi'_i} \hat{\lambda}_i E(e_{it}e_{jt}) = O_p(N^{-1}) + \mathcal{E} \cdot O_p(N^{-1/2}) \text{ for any } j \]

\[(f) \hat{H} \hat{\lambda}' \hat{\Phi}^{-1} \hat{\Phi}^{-1} \hat{\lambda} \mathcal{E}_j = O_p(T^{-1}) \text{ for any } j \]

where \( \xi_t \) is defined in Lemma S.8.

Proof of Lemma S.9: Part (a) is a direct result of Lemma S.5(d) and Corollary S.1(b). Part (b) is a direct result of Lemma S.5(a) and Corollary S.1(b).

Consider (c). The left hand side of (c) can be written as

\[ \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi'_i} \hat{\lambda}_i E(e_{it} \xi'_t) + \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi'_i} (\hat{\lambda}_i - \lambda_i) E(e_{it} \xi'_t) = I_1 + I_2 \text{ say} \]
Consider $I_1$, which is bounded in norm by
\[ \| \hat{H} \| \left( \max_{i \leq N} \| \frac{1}{\hat{\phi}_i^2} \lambda_i \| \right) \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \| E(e_{it} \xi_t^i) \| \leq C^3 \| \hat{H} \| \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \| E(e_{it} \xi_t^i) \| \]
where $\xi_t = (e_{1t}, e_{2t}, \ldots, e_{rt})$. For any $j \leq r$, by Assumption C.3, we have
\[ \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} |E(e_{it} e_{jt})| \leq \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} |\tau_{ij,t}| \leq \sum_{i=1}^{N} \tau_{ij} \leq C \]
So the term $\sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \| E(e_{it} \xi_t^i) \|$ is bounded by $\sqrt{T}C$. Given this result, we have $I_1 = O_p(N^{-1})$ by Corollary S.1(b).

Consider $I_2$. $I_2$ is bounded in norm by
\[ C \| \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \| \hat{\lambda}_i - \lambda_i \|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} E(e_{it} \xi_t^i) \right\|^2 \right)^{1/2} \]
Noting $\xi_t = (e_{1t}, e_{2t}, \ldots, e_{rt})'$. For any $j \leq r$,
\[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} E(e_{it} e_{jt}) \right)^2 \leq \frac{1}{N} \sum_{i=1}^{N} \tau_{ij}^2 \leq \frac{1}{N} \sup_{i \leq N} |\tau_{ii}| \sum_{i=1}^{N} |\tau_{ij}| \leq N^{-1}C \]
Thus, $I_2 = \mathcal{E} \cdot O_p(N^{-1/2})$, and (c) follows.

Part (d) is a direct result of Lemma S.5(c) and $\|N^{1/2} \hat{H}^{1/2}\| = O_p(1)$.

The proofs of (e) and (f) are contained in the proofs of (c) and (d). □

**Proposition S.2** Under Assumptions A-D, irrespective which set of identification conditions,
\[ M_{ff}(\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} = O_p(T^{-1/2}) + O_p(N^{-1}) + O_p\left( \left[ \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right]^{1/2} \right) + o_p(\mathcal{E}) \]
where $\mathcal{E}$ is defined in Lemma S.9.
Proof of Proposition S.2: The proof depends on the identification restrictions, so we consider each set of identification conditions separately.

Under IC1: The left hand side of the first \( r \) equations in (S.14) are zero. So we have

\[
M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda)M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \\
+ \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} + \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \\
+ \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} - \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi} - \frac{1}{T}\sum_{t=1}^{T}\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \\
- \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\epsilon}e'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} - \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t \xi_t' - \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t' \\(S.18) \\
- \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t \xi_t' - E(e_t \xi_t')] + \hat{H} - \hat{H}\frac{1}{T}\sum_{t=1}^{N}\sum_{i=1}^{T} \frac{1}{\phi_i^2} \lambda_i E(e_u \xi_t') + \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\epsilon} \xi'
\]

Consider the right hand side of the above equation. The first term is of a smaller order term than \((\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\) and hence negligible. The 2nd, 3rd and 8th terms are \( O_p(T^{-1/2}) \) by Lemma S.6(b) and Corollary S.1(c). The 4th term is \( O_p(T^{-1/2}) \) by Lemma S.7(c). The 5th and 6th term are \( O_p(N^{-1}) \) and \( O_p(T^{-1}) \) by Lemma S.9(a) and (b). The 7th term is \( O_p(T^{-1/2}) \) by Corollary S.1(c) and the fact \( E\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t \xi_t\|^2 < \infty \). The 9th term is \( O_p(T^{-1/2}) \) by Lemma S.8(b). The last two terms are \( O_p(N^{-1}) + O_p(T^{-1}) + o_p(\mathcal{E}) \) by Lemma S.9(c) and (d). Given these results, we have

\[
M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = O_p(T^{-1/2}) + O_p(N^{-1}) + o_p(\mathcal{E})
\]

Under IC2: From the identification condition \( \frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda} = \frac{1}{N}\Lambda'\Phi^{-1}\Lambda = I_r \), by adding and subtracting terms, we have the identity

\[
\frac{1}{N}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda} + \frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda) \\
= -\frac{1}{N}\Lambda'\Phi^{-1}(\Phi^{-1} - \Phi^{-1})\Lambda + \frac{1}{N}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda) \\(S.19) 
\]
The first term on right hand side of the above equation is \( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i \phi_i^*} (\hat{\phi}_i^2 - \phi_i^2) \lambda_i \lambda_i' \), which is bounded in norm by

\[
\left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i \phi_i^*} \|\lambda_i\|^4 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right)^{1/2} \leq C^6 \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right)^{1/2} 
\]

From this and noticing \( \frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) = o_p(E) \), we have

\[
\frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} + \frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda)' = O_p \left( \left[ \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right]^{1/2} \right) + o_p(E) \quad (S.20)
\]

Consider (S.13). Since both \( \hat{M}_{ff} \) and \( M_{ff} \) are diagonal matrices, we have

\[
\text{Ndiag} \left\{ \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff} + M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \right\}
\]

\[
= \text{Ndiag} \left\{ \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} + \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda} \frac{1}{T} \sum_{t=1}^{T} f_t e_t' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} 
+ \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda} \frac{1}{T} \sum_{t=1}^{T} (e_t e_t' - \Omega_t) \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \right. 
\]

\[
- \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Phi} - \frac{1}{T} \sum_{t=1}^{T} \Omega_t) \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} - \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Phi} \hat{\Phi}^{-1} \hat{\Phi} \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \right\} \quad (S.21)
\]

where \( \text{Ndiag} \) denotes the off-diagonal elements. Following the discussion after equation (S.18), the right hand side of the above equation is \( O_p(T^{-1/2}) + O_p(N^{-1}) \). Thus equation (S.21) can be written as

\[
\text{Ndiag} \left\{ \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff} + M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \right\} = O_p(T^{-1/2}) + O_p(N^{-1}) \quad (S.22)
\]

Note that under IC2, \( \hat{H} = \frac{1}{N} I_r \), thus both (S.20) and (S.22) put restrictions on \( \frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \). Equation (S.20) puts \( \frac{1}{2} r(r+1) \) restrictions, while (S.22) puts \( \frac{1}{2} r(r-1) \) restrictions. So the \( r \times r \) matrix \( \frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \) can be uniquely determined. By
solving the system of equations of (S.20) and (S.22) we obtain,

\[ M_{ff}(\hat{\Lambda} - \Lambda)^{'}\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = O_p(T^{-1/2}) + O_p(N^{-1}) + O_p \left( \left[ \frac{1}{N} \sum_{i=1}^{N} (\phi_i^2 - \phi_i^2)^2 \right]^{1/2} \right) + o_p(\mathcal{E}) \]

**Under IC3:** The proof of Proposition S.2 under IC3 is quite similar to the case of IC2. The details are omitted; also see, Bai and Li (2012).

**Under IC4:** Consider (S.14). Pre-multiplying \( \hat{M}_{ff} \) on both sides, the first \( r \) equations can be written as

\[
\hat{M}_{ff}(\hat{\Lambda}'_1 - \Lambda'_1) = M_{ff}(\hat{\Lambda} - \Lambda)^{'}\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1 - \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda)M_{ff}(\hat{\Lambda} - \Lambda)^{'}\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1
\]

\[
- \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda \frac{1}{T} \sum_{t=1}^{T} f_t e'_t \hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1 - \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda \frac{1}{T} \sum_{t=1}^{T} e_t f_t' \Lambda' \hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1
\]

\[
- \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda \frac{1}{T} \sum_{t=1}^{T} (e_t e'_t - \Omega_t) \hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1 + \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi} - \frac{1}{T} \sum_{t=1}^{T} \Omega_t) \hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1
\]

\[
+ \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e} \bar{e}' \hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1 + \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda \frac{1}{T} \sum_{t=1}^{T} f_t e'_t + \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda \frac{1}{T} \sum_{t=1}^{T} e_t f_t' \Lambda'_1
\]

\[
+ \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T} \sum_{t=1}^{T} [e_t e'_t - E(e_t e'_t)] - \hat{H}\hat{\Lambda}' + \hat{H} \frac{1}{T} \sum_{t=1}^{T} \sum_{t=1}^{T} \frac{1}{\phi_t^2} \hat{\lambda}_t E(e_t e'_t) - \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e} \bar{e}'
\]

Consider the third to last term. It can be split into \( \hat{H}\Lambda'_1 \) and \( \hat{H}(\hat{\Lambda}'_1 - \Lambda'_1) \). The former term is \( O_p(N^{-1}) \) by Corollary S.1(b) and the latter one is of a smaller order term than \( \hat{M}_{ff}(\hat{\Lambda}'_1 - \Lambda'_1) \) by \( \hat{M}_{ff} \overset{p}{\to} M_{ff} \). So this term is \( O_p(N^{-1}) \). Given this result, following the discussion after equation (S.18), the right hand side of the above equation, except the first term, is \( O_p(T^{-1/2}) + O_p(N^{-1}) + o_p(\mathcal{E}) \). Thus, we can rewrite (S.23) as

\[
\hat{M}_{ff}(\hat{\Lambda}'_1 - \Lambda'_1) = M_{ff}(\hat{\Lambda} - \Lambda)^{'}\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda'_1 + O_p(T^{-1/2}) + O_p(N^{-1}) + o_p(\mathcal{E})
\]

However, by the identification restrictions, the left hand side matrix is upper triangular and has zero diagonal elements, so its elements on and below the diagonal are...
all zero. This is still true after multiplying $\Lambda_i^{-1}$ on each side since the latter matrix is also upper triangular. It follows that

$$\text{nonupper}\left\{M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\right\} = O_p(T^{-1/2}) + O_p(N^{-1}) + o_p(\mathcal{E}) \quad (S.24)$$

where \text{nonupper} means lower triangular elements plus diagonal ones. The above equation has $\frac{1}{2}r(r + 1)$ restrictions. But equation (S.22), which holds since IC4 also requires that $\hat{M}_{ff}$ and $M_{ff}$ be diagonal matrices, gives another $\frac{1}{2}r(r - 1)$ restrictions. So the matrix $M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}$ can be uniquely determined by solving (S.22) and (S.24). Then we obtain

$$M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = O_p(T^{-1/2}) + O_p(N^{-1}) + o_p(\mathcal{E})$$

\textbf{Under IC5:} The above result still holds under IC5. The derivation is similar to IC4 and hence omitted.

Summarizing all the results, we obtain Proposition S.2. □

In order to prove Theorem 1, we need the following lemma.

\textbf{Lemma S.10} Under Assumptions A-D,

\begin{align*}
(a) & \quad \frac{1}{N} \sum_{j=1}^{N} \left\| \lambda_j'\hat{H}\left(\sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})]\right)\right\|^2 = O_p(T^{-1}) \\
(b) & \quad \frac{1}{N} \sum_{j=1}^{N} \left\| \lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda)\frac{1}{T} \sum_{t=1}^{T} f_t e_{jt}\right\|^2 = o_p(T^{-1}) \\
(c) & \quad \frac{1}{N} \sum_{j=1}^{N} \left\| \lambda_j'\hat{H}\frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^2} \lambda_i E(e_{it} e_{jt})\right\|^2 = O_p(N^{-2}) + \frac{1}{N} O_p\left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2}\left\|\lambda_i - \lambda_i\right\|^2\right) \\
(d) & \quad \frac{1}{N} \sum_{j=1}^{N} \left\| \lambda_j'\hat{H}\hat{\Phi}^{-1}\hat{e}^j\right\|^2 = O_p(T^{-2})
\end{align*}

\textbf{Proof of Lemma S.10:} The proofs of (a) and (b) are similar to those of Lemma B.4 in Bai and Li (2012) and hence omitted.
Consider (c). The term $\|\lambda_j' \hat{H} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \frac{1}{\phi_i} \hat{\lambda}_t E(e_{it} e_{jt})\|$ is bounded by

$$\left\| \lambda_j' \hat{H} \sum_{i=1}^N \frac{1}{\phi_i} \lambda_i \frac{1}{T} \sum_{t=1}^T \tau_{ij,t} \right\| + \left\| \lambda_j' \hat{H} \sum_{i=1}^N \frac{1}{\phi_i^2} (\hat{\lambda}_i - \lambda_i) \frac{1}{T} \sum_{t=1}^T \tau_{ij,t} \right\|$$

So the left hand side of (c) is bounded by

$$2 \frac{1}{N} \sum_{j=1}^N \left( \left\| \lambda_j' \hat{H} \sum_{i=1}^N \frac{1}{\phi_i} \lambda_i \frac{1}{T} \sum_{t=1}^T \tau_{ij,t} \right\|^2 + \left\| \lambda_j' \hat{H} \sum_{i=1}^N \frac{1}{\phi_i^2} (\hat{\lambda}_i - \lambda_i) \frac{1}{T} \sum_{t=1}^T \tau_{ij,t} \right\|^2 \right)$$

By the boundedness of $\lambda_i, \hat{\phi}_i^2$, the first term of the above is bounded by $2C^6 \| \hat{H} \|_2 \frac{1}{N} \sum_{j=1}^N (\sum_{i=1}^N \tau_{ij})^2$. So the first term is $O_p(N^{-2})$ by $\sum_{i=1}^N \tau_{ij} < C$ for all $j$.

By the boundedness of $\lambda_i, \hat{\phi}_i^2$, the first term of the above is bounded by $2C^6 \| \hat{H} \|_2 \frac{1}{N} \sum_{j=1}^N (\sum_{i=1}^N \tau_{ij})^2$. So the first term is $O_p(N^{-2})$ by $\sum_{i=1}^N \tau_{ij} < C$ for all $j$.

By the same method in deducing (S.12) and (S.13), we have

$$\hat{\phi}_j^2 - \phi_j^2 = \frac{1}{T} \sum_{t=1}^T (e_{jt}^2 - \phi_j^2) - (\hat{\lambda}_j - \lambda_j)' \hat{M}_{ff}(\hat{\lambda}_j - \lambda_j) \quad \text{(S.25)}$$

Consider (d). The left hand side of (d) is equal to $\| \hat{H} \hat{\Phi}^{-1} \hat{e} \|_2^2 \frac{1}{N} \sum_{j=1}^N \| \lambda_j \hat{e}_j \|^2$. Since $\| \hat{H} \hat{\Phi}^{-1} \hat{e} \|^2 = O_p(T^{-1})$ by Lemma S.9(c) and $\frac{1}{N} \sum_{j=1}^N \| \lambda_j \hat{e}_j \|^2 = O_p(T^{-1})$, (d) follows.

**Proof of Theorem 1**: We begin with the first order condition on $\text{diag}\{ \Phi \}$.

By the same method in deducing (S.12) and (S.13), we have

$$\hat{\phi}_j^2 - \phi_j^2 = \frac{1}{T} \sum_{t=1}^T (e_{jt}^2 - \phi_j^2) - (\hat{\lambda}_j - \lambda_j)' \hat{M}_{ff}(\hat{\lambda}_j - \lambda_j) \quad \text{(S.25)}$$

$$+ \lambda_j' \hat{H} \hat{\lambda} \hat{\Phi}^{-1} (\hat{\lambda} - \lambda) \hat{M}_{ff}(\hat{\lambda} - \lambda)' \hat{H} \hat{\lambda} + 2 \lambda_j' \hat{H} \hat{\lambda} \hat{\Phi}^{-1} \hat{\lambda} \frac{1}{T} \sum_{t=1}^T f_t e_t' \hat{\Phi}^{-1} \hat{\lambda} \hat{H} e_t$$

$$+ \lambda_j' \hat{H} \left( \sum_{i=1}^N \sum_{j=1}^N \frac{1}{\phi_i^2} \frac{1}{\phi_j^2} \frac{1}{\phi_i} \hat{\lambda}_i \hat{\lambda}_j \frac{1}{T} \sum_{t=1}^T [e_{it} e_{jt} - E(e_{it} e_{jt})] \hat{H} \lambda_j - \lambda_j' \hat{H} \hat{\Phi}^{-1} \hat{e} \hat{e}' \hat{\Phi}^{-1} \hat{\lambda}' \hat{H} \lambda_j \right)$$

$$- \lambda_j' \hat{H} \hat{\lambda} \frac{1}{T} \sum_{t=1}^T \Omega_t \hat{\Phi}^{-1} \hat{\lambda}' \hat{H} \lambda_j - 2 \lambda_j' \hat{H} \hat{\lambda} \hat{\Phi}^{-1} \hat{\lambda}' \hat{H} \lambda_j$$

$$- \lambda_j' \hat{H} \hat{\Phi}^{-1} (\hat{\Phi} - \frac{1}{T} \sum_{t=1}^T \Omega_t) \hat{\Phi}^{-1} \hat{\lambda}' \hat{H} \lambda_j - 2 \lambda_j' \hat{H} \hat{\lambda} \hat{\Phi}^{-1} \hat{\lambda}' \hat{H} \lambda_j$$
+2\lambda_j' \hat{H} \hat{\lambda}_j - 2\lambda_j' \hat{H} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^2 \lambda_i} E(e_{it} e_{jt}) - 2\lambda_j' \hat{H} \left( \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})] \right)

+2\lambda_j' \hat{H} \lambda' \Phi^{-1}(\hat{\Lambda} - \Lambda) \frac{1}{T} \sum_{t=1}^{T} f_t e_{jt} + 2\lambda_j' \hat{H} \lambda' \Phi^{-1} \tilde{e} e_j = a_{1,j} + a_{2,j} + \cdots + a_{13,j} \quad \text{say}

By the Cauchy-Schwarz inequality, we have

$$\frac{1}{N} \sum_{j=1}^{N} (\hat{\phi}_j^2 - \phi_j^2)^2 \leq \frac{1}{N} \sum_{j=1}^{N} \|a_{1,j} + \cdots + a_{13,j}\|^2 \leq 13 \frac{1}{N} \sum_{j=1}^{N} (\|a_{1,j}\|^2 + \cdots + \|a_{13,j}\|^2)$$

The first term is \(\frac{1}{N} \sum_{j=1}^{N} [\frac{1}{T} \sum_{t=1}^{T} (e_{jt}^2 - \phi_j^2)^2] = O_p(T^{-1})\) by (S.6). The second term is bounded by \(\|\hat{M}_{ff}\| \cdot \frac{1}{N} \sum_{j=1}^{N} \|\hat{\lambda}_j - \lambda_j\|^4\). Using (S.14), this term, by neglecting the smaller order term of \(\frac{1}{N} \sum_{j=1}^{N} (\hat{\phi}_j^2 - \phi_j^2)^2\), is bounded by \(O_p(T^{-2}) + O_p(N^{-4}) + \frac{1}{N} O_p(\frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi_j^4} \|\hat{\lambda}_i - \lambda_i\|^2)\). Consider the 3rd term, which is bounded in norm by

$$\|\lambda' \Phi^{-1} \hat{\lambda} \hat{H}\|^2 \|\hat{M}_{ff}\|^2 \frac{1}{N} \sum_{j=1}^{N} \|\lambda_j\|^4$$

By Proposition S.2, the 3rd term is \(O_p(T^{-2}) + O_p(N^{-4}) + O_p(\|\frac{1}{N} \sum_{j=1}^{N} (\hat{\phi}_j^2 - \phi_j^2)^2\|)^2 + o_p(\mathcal{E}^2)\). The 4th term can be proved to be \(O_p(T^{-1})\) similarly as the 3rd term due to Lemma S.6(b) and Corollary S.1(c). The 5th term is \(O_p(T^{-1})\) due to Lemma S.7(c). The 6th term is \(O_p(T^{-2})\) due to Lemma S.9(b). The 7th term is \(O_p(N^{-2})\) due to Lemma S.9(a). The 8th term is \(O_p(T^{-1})\) due to Lemma S.6(b). Consider the 9th term. Because \(\lambda_j' \hat{H} \lambda_j = \lambda_j' \hat{H} \lambda_j + \lambda_j' \hat{H} (\hat{\lambda}_j - \lambda_j)\), the 9th term is bounded in norm by

$$\frac{1}{N} \sum_{j=1}^{N} \|a_{9,j}\|^2 \leq 2 \left( \frac{1}{N} \sum_{j=1}^{N} \|\lambda_j' \hat{H} \lambda_j\|^2 + \frac{1}{N} \sum_{j=1}^{N} \|\lambda_j' \hat{H} (\hat{\lambda}_j - \lambda_j)\|^2 \right)$$

The first term is \(O_p(N^{-2})\) by \(\hat{H} = O_p(N^{-1})\). The second term is bounded by \(C^2 \|\hat{H}\|^2 \frac{1}{N} \sum_{j=1}^{N} \|\hat{\lambda}_j - \lambda_j\|^2\), which is further bounded by \(C^4 \|\hat{H}\|^2 \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi_j^4} \|\hat{\lambda}_j - \lambda_j\|^2\).

However, \(\frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi_j^4} \|\hat{\lambda}_j - \lambda_j\|^2 = o_p(1)\), so the second term is dominated by the first one. Given these results, the 9th term is \(O_p(N^{-2})\). The 10-13th terms are summa-
ized in Lemma S.10. So we have

$$\frac{1}{N} \sum_{j=1}^{N} (\hat{\phi}^2_j - \phi^2_j)^2 = O_p(T^{-1}) + O_p(N^{-2}) + o_p\left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi^2_i} \|\hat{\lambda}_i - \lambda_i\|^2\right)$$  \hspace{1cm} (S.26)

We next derive bounds involving $\|\hat{\lambda}_j - \lambda_j\|^2$ and $\hat{M}_{ff} - M_{ff}$. Consider (S.14). There are 13 terms on the left hand side of (S.14). We use $b_{1,j}, b_{2,j}, \ldots, b_{13,j}$ to denote them. By the Cauchy-Schwarz inequality, $\|b_{1,j} + b_{2,j} + \cdots + b_{13,j}\|^2 \leq 13(\|b_{1,j}\|^2 + \|b_{2,j}\|^2 + \cdots + \|b_{13,j}\|^2)$. By this inequality, and noticing $C^{-2} \leq \hat{\phi}^2_j \leq C^2$, we have

$$\frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi^2_j} \|\hat{\lambda}_j - \lambda_j\|^2 \leq C^2 \frac{1}{N} \sum_{j=1}^{N} \|\hat{\lambda}_j - \lambda_j\|^2 \leq 13C^2 \frac{1}{N} \sum_{j=1}^{N} (\|b_{1,j}\|^2 + \cdots + \|b_{13,j}\|^2)$$

The 1st term $\frac{1}{N} \sum_{j=1}^{N} \|b_{1,j}\|^2$ is bounded by $\|\hat{M}_{ff}^{-1}\|^2 \|M_{ff}(\hat{\Lambda} - \Lambda)\| \hat{\Phi}^{-1} \hat{\Lambda} \hat{H}\|^2 \frac{1}{N} \sum_{j=1}^{N} \|\lambda_j\|^2$. Notice $\|\hat{M}_{ff}^{-1}\|^2 = O_p(1)$ by Proposition S.1 and $\frac{1}{N} \sum_{j=1}^{N} \|\lambda_j\|^2 = O(1)$ by Assumption B. By Proposition S.2 and neglecting the smaller order term of $\frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi^2_j} \|\hat{\lambda}_j - \lambda_j\|^2$, we have

$$\frac{1}{N} \sum_{j=1}^{N} \|b_{1,j}\|^2 = O_p(T^{-1}) + O_p(N^{-2}) + O_p\left(\frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}^2_i - \phi^2_i)^2\right)$$

The 2nd term $\frac{1}{N} \sum_{j=1}^{N} \|b_{2,j}\|^2$ is dominated by the first and is negligible. By Lemmas S.6 and S.7, the 3rd-10th terms are $O_p(T^{-1}) + O_p(N^{-2}) + O_p\left(\frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}^2_i - \phi^2_i)^2\right)$. The 10th term can be proved to be $O_p(N^{-2})$ similarly as $\frac{1}{N} \sum_{j=1}^{N} \|a_{9,j}\|^2$. The last two terms are summarized in Lemma S.10. So we have

$$\frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi^2_j} \|\hat{\lambda}_j - \lambda_j\|^2 = O_p(T^{-1}) + O_p(N^{-2}) + O_p\left(\frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}^2_i - \phi^2_i)^2\right)$$  \hspace{1cm} (S.27)
Similarly, using Lemmas S.6, S.7 and S.10, we deduce

\[ \| \hat{M}_{ff} - M_{ff} \|^2 = O_p(T^{-1}) + O_p(N^{-2}) + O_p \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right) \]

\[ + o_p \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \| \hat{\lambda}_i - \lambda_i \|^2 \right) \]

(S.28)

Substituting (S.27) into (S.26), we obtain

\[ \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 = O_p(T^{-1}) + O_p(N^{-2}) \]

Substituting \( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 = O_p(T^{-1}) + O_p(N^{-2}) \) into (S.27) and (S.28), we obtain the two remaining results of Theorem 1. This completes the proof of Theorem 1.

□

**Corollary S.2** Under Assumptions A-D, irrespective which set of identification conditions,

\[ M_{ff}(\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} = O_p(T^{-1/2}) + O_p(N^{-1}) \]

Corollary S.2 is a direct result of Proposition S.2 and Proposition 1.

**Supplement C: Proof for the asymptotic representations**

Now we first state the limiting results for the estimated \( M_{ff} \) and \( \hat{\phi}_i^2 \).

**Theorem S.1 (Asymptotic representations for \( \hat{M}_{ff} \))** Under the assumptions of Theorem 1 and \( \sqrt{T}/N \to 0 \), we have:

Under IC1, \( \sqrt{T} \text{vech}(\hat{M}_{ff} - M_{ff}) = D_p^+ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\xi_t \otimes f_t + f_t \otimes \xi_t) \right) + o_p(1); \)

Under IC2, \( \text{diag}\{ \hat{M}_{ff} - M_{ff} \} = O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1}); \)

Under IC4, \( \sqrt{T} \text{diag}\{ \hat{M}_{ff} - M_{ff} \} = 2 \text{diag}\left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t \xi_t' \Lambda_1^{-1} f_t \right\} + o_p(1); \)

where \( D_p^+ \) is the Moore-Penrose inverse of the duplication matrix \( D_p \).
Note that under IC3 and IC5, $M_{ff} = I_r$ is known, not estimated.

**Corollary S.3 (Limiting distribution for $\hat{M}_{ff}$)** Under the assumptions of Theorem S.1 together with Assumption F, we have:

Under IC1, \( \sqrt{T} \left( \text{vech}(\hat{M}_{ff} - M_{ff}) \right) \xrightarrow{d} N\left(0, 4D_r^+ \Gamma^M D_r^{+\prime}\right) \);

Under IC4, \( \sqrt{T} \left( \text{diag}\{\hat{M}_{ff} - M_{ff}\} \right) \xrightarrow{d} N\left(0, 4J_r \Pi^M J_r^{\prime}\right) \);

where $J_r$ is an $r \times r^2$ matrix, which satisfies, for any $r \times r$ matrix $M$, \( \text{diag}\{M\} = J_r \text{vec}(M) \), and $\Gamma^M = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} [f_t f_s' \otimes E(\xi_t \xi_s')]$ and $\Pi^M = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} [(\Lambda^{-1} E(\xi_t \xi_s') \Lambda^{-1\prime}) \otimes (f_t f_s')]$.

Theorem S.1 only gives the asymptotic representations for $\hat{M}_{ff}$ under IC1 and IC4. Under IC2, it states that $\hat{M}_{ff} - M_{ff}$ is of $O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1})$. The terms $O_p(T^{-1})$ and $O_p(N^{-1})$ include some bias terms in the magnitude of $O(T^{-1})$ and $O(N^{-1})$. If some higher order moments assumptions are made, we can extract the biases from $O_p(N^{-1}) + O_p(T^{-1})$ and the remaining term will have a limiting normal distribution with a $\sqrt{N\bar{T}}$ convergence rate. We do not pursue this here, partly because this exercise requires additional assumptions and the derivation is lengthy, and partly because knowing the order of $\hat{M}_{ff} - M_{ff}$ is sufficient. For example, for the limiting distribution of $\hat{f}_t - f_t$, we only need to know the order of $\hat{M}_{ff} - M_{ff}$. In addition, under IC2, the convergence rate is already faster than under IC1 and IC4.

Theorem S.1 also shows that, under IC1 and IC4, the asymptotic representation of $\hat{M}_{ff} - M_{ff}$ only involves the error terms $\xi_t = (e_{1t}, e_{2t}, \ldots, e_{rt})'$. The underlying reason is that the restrictions IC1 and IC4 only involve the first $r$ equations and IC2 involves the entire cross sections. This is also the underlying reason for the faster convergence rate of $\hat{M}_{ff}$ under IC2.

**Theorem S.2** Under the assumptions of Theorem 1 and $\sqrt{T}/N \to 0$, irrespective of which set of identification conditions, we have
Given the above result, together with Assumption F, we have

\[ \sqrt{T}(\hat{\phi}_i^2 - \phi_i^2) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (e_{it}^2 - \phi_i^2) + o_p(1) \]

where \( \sigma_i^2 \) is defined in Assumption F.2.

Theorem S.2 shows that \( \hat{\phi}_i^2 \) is \( \sqrt{T} \)-consistent for \( \phi_i^2 = \frac{1}{T} \sum_{t=1}^{T} E(e_{it}^2) \). If the error \( e_{it} \) is stationary over \( t \), \( \hat{\phi}_i^2 \) gives a consistent estimator for \( \phi_i^2 = E(e_{it}^2) \). With heteroskedasticity, the estimator \( \hat{\phi}_i^2 \) provides an estimate for the average variance.

In addition, we note that given Assumption E.1,

\[ E \left( \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\hat{\phi}_i^2} \lambda_i f_t e_{it} \right\|^2 \right) \leq C. \]  \( (S.29) \)

We need the following lemmas to derive the limiting distributions.

**Lemma S.11** Under Assumptions A-E,

(a) \( \left\| \hat{H} \hat{\lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) \right\| = O_p(T^{-1/2}) + O_p(N^{-1}) \)

(b) \( \left\| \hat{H} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \hat{\lambda}_i \left( \frac{1}{T} \sum_{t=1}^{T} [e_{it} \xi'_t - E(e_{it} \xi'_t)] \right) \right\| = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \)

(c) \( \left\| \hat{H} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \hat{\lambda}_i \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{\hat{\lambda},t} \right) \right\| = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \)

(d) \( \left\| \hat{H} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\hat{\phi}_i^2 \hat{\phi}_j^2} \hat{\lambda}_i \hat{\lambda}_j \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{\hat{\lambda},t} \right) \hat{H} \right\| = O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) \)

(e) \( \left\| \hat{H} \hat{\lambda}' \hat{\Phi}^{-1} \frac{1}{T} \sum_{t=1}^{T} e_{it} f'_t \right\| = \left\| \hat{H} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\hat{\phi}_i^2} \lambda_i f_t e_{it} \right\| + O_p(T^{-1}) \)

\[ + O_p(N^{-1/2}T^{-1/2}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \]

where \( \varepsilon_{\hat{\lambda},t} = e_{it} \varepsilon_{jt} - E(e_{it} \varepsilon_{jt}) \) for notational simplicity.

**Proof of Lemma S.11:** Part (a) is implied by Lemma S.6(a) and Theorem 1. It is also implied by Corollary S.2.
Consider (b). The left-hand side of (b) is bounded by

\[ \| \hat{H}_N \| \left\| \frac{1}{NT} \sum_{i=1}^{N} \left( \frac{1}{\phi_i^2} - \frac{1}{\phi_i^2} \right) \lambda_i \sum_{t=1}^{T} [e_{it} \xi'_t - E(e_{it} \xi'_t)] \right\| \]

\[ + \| \hat{H}_N \| \left| \frac{1}{NT} \sum_{i=1}^{N} \frac{1}{\phi_i^2} (\lambda_i - \lambda_i) \sum_{t=1}^{T} [e_{it} \xi'_t - E(e_{it} \xi'_t)] \right| \]

The first expression is \( O_p(N^{-1/2} T^{-1/2}) \) by Assumption E.2. Using the Cauchy-Schwarz inequality, the second term is bounded by

\[ \| \hat{H}_N \| \left| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\phi_i^2} - \frac{1}{\phi_i^2} \right)^2 \right|^{1/2} \left( \| \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} \lambda_i [e_{it} \xi'_t - E(e_{it} \xi'_t)] \right\|^{2} \right)^{1/2} \]

which is further bounded by

\[ C^5 \| \hat{H}_N \| \left| \frac{1}{N} \sum_{i=1}^{N} \left( \phi_i^2 - \phi_i^2 \right)^2 \right|^{1/2} \left( \| \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} \lambda_i [e_{it} \xi'_t - E(e_{it} \xi'_t)] \right\|^{2} \right)^{1/2} \]

which is \( O_p(T^{-1}) + O_p(N^{-1} T^{-1/2}) \) by \( E(\parallel \frac{1}{\sqrt{T}} \sum_{i=1}^{T} [e_{it} \xi'_t - E(e_{it} \xi'_t)] \parallel^2) \leq C \) for all \( i \).

The third term can be proved to be \( O_p(T^{-1}) + O_p(N^{-1} T^{-1/2}) \) similarly as the second.

This proves (b).

The proof of (c) is similar to that of (b) and hence omitted.

Consider (d). Note \( \hat{H} = H_N \cdot N^{-1} \) and \( \| \hat{H}_N \| = O_p(1) \). Adding and subtracting terms and ignoring \( \| H_N \|^2 \), (d) is bounded by

\[ \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2} \phi_j^2 \lambda_i \lambda_j \sum_{t=1}^{T} \varepsilon_{ij,t} \right\| + \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \sum_{t=1}^{T} \varepsilon_{ij,t} \right\| 

+ \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{\phi_i^2} \phi_j^2 - \frac{1}{\phi_i^2} \phi_j^2 \right) \lambda_i \lambda_j \sum_{t=1}^{T} \varepsilon_{ij,t} \right\| 

+ \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \phi_i^2 - \phi_j^2 \right) \lambda_i \lambda_j \sum_{t=1}^{T} \varepsilon_{ij,t} \right\| 

+ \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \phi_i^2 - \phi_j^2 \right) \lambda_i \lambda_j \sum_{t=1}^{T} \varepsilon_{ij,t} \right\| 

24
\[ + \left\| \frac{1}{N^{2T}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^{2j}} \hat{\lambda}_i (\hat{\lambda}_i - \lambda_i) \sum_{t=1}^{T} \varepsilon_{ij,t} \right\| \]
\[ + \left\| \frac{1}{N^{2T}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^{2j}} \hat{\lambda}_j (\hat{\lambda}_j - \lambda_j) \sum_{t=1}^{T} \varepsilon_{ij,t} \right\| \]

The first term is bounded in norm by

\[ \left( \frac{1}{N} \sum_{i=1}^{N} \| \frac{1}{\phi_i^{2j}} \lambda_i \|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \| \frac{1}{NT} \sum_{j=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_j^{2t}} \lambda_j \varepsilon_{ij,t} \|^2 \right)^{1/2} \]

which is \( O_p(N^{-1/2}T^{-1/2}) \) by Assumption E.2. The second term is bounded by

\[ \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| \frac{\hat{\phi}_i^{2j} - \phi_i^{2j}}{\hat{\phi}_i^{2j} \phi_j^{2j}} \lambda_i \lambda_j \right\|^2 \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{ij,t} \right)^2 \right)^{1/2} . \]

The above term is further bounded by

\[ C^8 \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\hat{\phi}_i^{2j} - \phi_i^{2j}}{\phi_i^{2j}} \lambda_i \lambda_j \right)^2 \right]^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{ij,t} \right)^2 \right)^{1/2} \]

which is \( O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) \). The remaining terms are all \( O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) \) by similar arguments. This proves (d).

Using the similar arguments, by (S.29), (e) can be proved and the details are omitted. □

**Lemma S.12** Under Assumptions A-E,

\[ \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\phi}_i^{2j} - \phi_i^{2j}}{\phi_i^{2j}} \lambda_i \lambda'_i = O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \]

**Proof of Lemma S.12:** Using (S.25), the expression \( \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\phi}_i^{2j} - \phi_i^{2j}}{\phi_i^{2j}} \lambda_i \lambda'_i \) can be expanded into 13 terms. We consider them one by one. The first term is equal to

\[ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^{2j}} (e_{it}^2 - \phi_i^2) \lambda_i \lambda'_i \]
which is $O_p(N^{-1/2}T^{-1/2})$ by Assumption E.3. The second term is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} (\hat{\lambda}_i - \lambda_i)' \hat{M}_{ff}(\hat{\lambda}_i - \lambda_i) \lambda_i'$$

which is bounded in norm by $C^2 \|\hat{M}_{ff}\| \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \|\hat{\lambda}_i - \lambda_i\|^2$, which is $O_p(T^{-1}) + O_p(N^{-2})$ by Proposition S.1 and Proposition 1.

Consider the third term, which is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \lambda_i \lambda_i' \hat{H} \lambda_i \lambda_i'$$

The above term is bounded in norm by $\|M_{ff}\| \cdot \|((\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{H} \lambda_i \lambda_i') = c_1 + c_2$

The term $c_1$ is bounded in norm by $\|\hat{H}\| \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \|\lambda_i\|^4$, which is $O_p(N^{-1})$ by Corollary S.1(b). The term $c_2$ is bounded in norm by

$$C \|\hat{H}\| \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^8} \|\lambda_i\|^6 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \|\hat{\lambda}_i - \lambda_i\|^2 \right)^{1/2}$$

which is of a smaller order term than $\|\hat{H}\|$. So the 9th term is $O_p(N^{-1})$.

The 10th term is $\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} \lambda_i \lambda_i' (\hat{\lambda}_i' \hat{H} \lambda_j) \frac{1}{T} \sum_{t=1}^{T} \tau_{ij,t}$, which is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} \lambda_i \lambda_i' [\lambda_i' \hat{H} (\hat{\lambda}_j - \lambda_j)] \frac{1}{T} \sum_{t=1}^{T} \tau_{ij,t}$$
We use \(c_3\) and \(c_4\) to denote the above two terms. Notice \(\frac{1}{T} \sum_{t=1}^{T} \tau_{ij,t} \leq \tau_{ij}\). By the boundedness of \(\lambda_i, \phi_i^2, \phi_i^2\), term \(c_3\) is bounded in norm by

\[
C^{10} \| \hat{H} \| \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \tau_{ij}
\]

which is \(O_p(N^{-1})\) by Assumption C.3 and \(\| \hat{H} \| = O_p(N^{-1})\). Consider \(c_4\), which is bounded in norm by

\[
C \| \hat{H} \| \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2} \| \lambda_i \|^6 \tau_{ij} \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \| \hat{\lambda}_j - \lambda_j \|^2 \tau_{ij} \right)^{1/2}
\]

The above is easily shown to be \(o_p(\| \hat{H} \|) = o_p(1/N)\) because the middle factor is \(O(1)\) and last factor is \(o_p(1)\). Thus, the 10th term is \(O_p(N^{-1})\).

The 11th term is \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \lambda_i' (\lambda_i' \hat{H} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \hat{\lambda}_j \frac{1}{T} \sum_{t=1}^{T} \epsilon_{ij,t})\), where \(\epsilon_{ij,t} = e_{it}e_{jt} - E(e_{it}e_{jt})\). This term can be written as

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2} \lambda_i \lambda_i' [\lambda_i' \hat{H}(\hat{\lambda}_j - \lambda_j)] \frac{1}{T} \sum_{t=1}^{T} \epsilon_{ij,t}
\]

\[- \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\phi_j^2 - \phi_j^2}{\phi_i^2 \phi_j^2} \lambda_i \lambda_i' (\lambda_i' \hat{H} \lambda_j) \frac{1}{T} \sum_{t=1}^{T} \epsilon_{ij,t}
\]

\[+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} \lambda_i \lambda_i' (\lambda_i' \hat{H} \lambda_j) \frac{1}{T} \sum_{t=1}^{T} \epsilon_{ij,t} = c_5 - c_6 + c_7
\]

The term \(c_5\) is bounded in norm by

\[
C \| \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \| \lambda_i \|^6 \right)^{1/2} \left( \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \| \hat{\lambda}_j - \lambda_j \|^2 \right)^{1/2} \left[ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \epsilon_{ij,t} \right)^2 \right]^{1/2}
\]
which is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Proposition 1 and Assumption C.5. By the boundedness of $\phi_i^2, \hat{\phi}_i^2$ and $\lambda_i$, the term $c_6$ is bounded in norm by

$$C^{12} \| \hat{N} \| \left( \frac{1}{N} \sum_{j=1}^{N} (\hat{\phi}_j^2 - \phi_j^2)^2 \right)^{1/2} \left( \frac{1}{N} \sum_{j=1}^{N} \sum_{j=1}^{T} (\sum_{t=1}^{T} \epsilon_{j,t})^2 \right)^{1/2}$$

which is also $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Proposition 1 and Assumption C.5. The term $c_7$ is bounded in norm by

$$\| \hat{N} \| \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \| \lambda_i \|_6^6 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{i=1}^{T} \frac{1}{\phi_i^4} \lambda_i \frac{1}{T} \sum_{t=1}^{T} \epsilon_{j,t} \right)^{1/2}$$

which is $O_p(N^{-1/2}T^{-1/2})$ by Assumption E.2. So the 11th term is $O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2})$.

The 12th term is $\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_i (\lambda_i^T \hat{H} \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda)^T \sum_{t=1}^{T} \epsilon_{t}^T \epsilon_t)$, which is an $r \times r$ matrix. We consider its $(g, h)$ $(g, h = 1, 2, \cdots, r)$ entry, which is equal to

$$\text{tr} \left[ \hat{H} \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^4} \lambda_i \phi_i \epsilon_{it} \right]$$

Since

$$E(\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^4} \lambda_i \phi_i \epsilon_{it} \|)^2 = \text{tr} \left[ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{\phi_i^4} \phi_j^4 \lambda_i \phi_i \lambda_j \phi_j \epsilon_{it} \epsilon_{st} \right]$$

$$\leq C^{16} \text{tr} \left[ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\gamma_{ij,ts}| \right] = O(N^{-1}T^{-1})$$

by Assumption E.1. So $\text{tr} \left[ \hat{H} \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^4} \lambda_i \phi_i \epsilon_{it} \right] = O_p(N^{-1/2}T^{-1}) + O_p(N^{-3/2}T^{-1/2})$ by Corollary S.2. This implies that the 12th term is $O_p(N^{-1/2}T^{-1}) + O_p(N^{-3/2}T^{-1/2})$.

The 13th term is $\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_i (\lambda_i^T \hat{H} \hat{\Phi}^{-1} \tilde{e} \tilde{e}_i)$. We denote the $l$th element of $\hat{H} \hat{\Phi}^{-1} \tilde{e}$ by $\delta_l$ temporarily. Notice that Lemma S.5 (a) indicates $\hat{H} \hat{\Phi}^{-1} \tilde{e}$ is $O_p(T^{-1/2})$. That is, $\delta_l = O_p(T^{-1/2})$ for all $l = 1, 2, \cdots, r$. The 13th term is an $r \times r$ matrix,
whose \((g,h)\) element \((g,h = 1,2,\cdots,r)\) is equal to

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{r} \frac{1}{\phi_i^l} \lambda_{g_l} \lambda_{u_l} \lambda_{h_i} \epsilon_l = \sum_{l=1}^{r} \frac{\delta_l}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^l} \lambda_{g_l} \lambda_{u_l} \lambda_{h_i} \epsilon_l
\]

Consider the term \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^l} \lambda_{g_l} \lambda_{u_l} \lambda_{h_i} \epsilon_l\), which is equal to \(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\phi_i^l} \lambda_{g_l} \lambda_{u_l} \lambda_{h_i} \epsilon_{it}\) and can be easily shown to be \(O_p(N^{-1/2}T^{-1/2})\). So the 13th term is \(O_p(N^{-1/2}T^{-1})\).

Summarizing all the results, we obtain Lemma S.12. □

**Proof of Theorems 1–S.2:** The limiting distributions depend on the identification conditions, and we derive the limits under each of identification conditions.

**Under IC1:** By equation (S.18), Lemma S.11, and Theorem 1, we have

\[
M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \Lambda \hat{H} = - \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \Lambda \left( \frac{1}{T} \sum_{t=1}^{T} f_t \xi_t' \right) + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})
\]

(S.30)

Substituting the above result into (S.14) and using the results of Lemmas S.9 and S.11, we have

\[
\hat{\lambda}_j - \lambda_j = - M_{ff}^{-1} \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \Lambda \frac{1}{T} \sum_{t=1}^{T} f_t \xi_t' \lambda_j + \hat{M}_{ff}^{-1} \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \Lambda \frac{1}{T} \sum_{t=1}^{T} f_t e_{jt} + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})
\]

(S.31)

Since \(\hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \Lambda = I_r - A' \rightarrow I_r\) and \(\hat{M}_{ff}^{-1} \rightarrow M_{ff}^{-1}\) by Proposition S.1, it follows, under the condition \(\sqrt{T}/N \rightarrow 0\),

\[
\sqrt{T}(\hat{\lambda}_j - \lambda_j) = - M_{ff}^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t \xi_t' \right) \lambda_j + M_{ff}^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t e_{jt} \right) + o_p(1)
\]

(S.32)

By Assumption F.1, it follows

\[
\sqrt{T}(\hat{\lambda}_j - \lambda_j) \overset{d}{\rightarrow} N \left( 0, (M_{ff})^{-1} \Gamma_j^\lambda (M_{ff})^{-1} \right)
\]
For the limiting distribution of $\hat{\phi}_i^2 - \phi_i^2$, consider equation (S.25). By Lemmas S.6, S.7 and S.11, equation (S.25) reduces to

$$\hat{\phi}_i^2 - \phi_i^2 = \frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \phi_i^2) - (\hat{\lambda}_j - \lambda_j)' \hat{M}_{ff} (\hat{\lambda}_i - \lambda_i)$$

$$+ O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Although equation (S.31) implies that $\hat{\lambda}_j - \lambda_j$ is $O_p(T^{-1/2}) + O_p(N^{-1})$, we avoid using this result since its derivation depends on the identification conditions. Here is a different argument that holds under all identification conditions. Equation (S.14) and Lemmas S.9 and S.11 imply that

$$\hat{\lambda}_j - \lambda_j = \hat{M}_{ff}^{-1} M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \lambda_j + O_p(T^{-1/2}) + O_p(N^{-1})$$

But Lemma S.11(a) implies that the first term of the above is also $O_p(T^{-1/2}) + O_p(N^{-1})$. It follows $\hat{\lambda}_j - \lambda_j = O_p(T^{-1/2}) + O_p(N^{-1})$, from which we obtain

$$\hat{\phi}_i^2 - \phi_i^2 = \frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \phi_i^2) + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \tag{S.33}$$

By Assumption F.2, it follows, under the condition $\sqrt{T/N} \to 0$,

$$\sqrt{T}(\hat{\phi}_j^2 - \phi_j^2) \xrightarrow{d} N(0, \sigma_j^2)$$

The above derivation shows that the limiting distribution applies to all five sets of identification conditions.

For the limiting distribution of $\hat{M}_{ff} - M_{ff}$, consider equation (S.13). By Lemmas S.9 and S.11, (S.13) implies that

$$\hat{M}_{ff} - M_{ff} = -\hat{H} \hat{\Lambda} \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff} - M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H}$$
Using (S.30) and noticing $\hat{H}\hat{\Lambda}^{-1}\hat{\Phi}^{-1}\hat{\Phi} I_r$, we have

$$\hat{M}_{ff} - M_{ff} = \frac{1}{T} \sum_{t=1}^{T} f_t \xi_t' + \frac{1}{T} \sum_{t=1}^{T} \xi_t f_t' + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Since $\hat{M}_{ff}$ and $M_{ff}$ are both symmetric matrices, under the condition $\sqrt{T}/N \to 0$, the above result can be further written as

$$\sqrt{T} \text{vech}(\hat{M}_{ff} - M_{ff}) = D_r^+ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_t \otimes f_t + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t \otimes \xi_t \right) + o_p(1) \quad (S.34)$$

where $D_r^+$ denotes the Moore-Penrose inverse of the $r$-order duplication matrix $D_r$.

By Assumption F.1, it follows,

$$\sqrt{T} \text{vech}(\hat{M}_{ff} - M_{ff}) \overset{d}{\to} N \left( 0, 4D_r^+ \Gamma^M D_r^{++} \right)$$

Under IC2: Consider equation (S.19). The term $\frac{1}{N} \sum_{i=1}^{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda)$ is $O_p(T^{-1}) + O_p(N^{-2})$ by Theorem 1. The term $\frac{1}{N} \Lambda'(\hat{\Phi}^{-1} - \Phi^{-1}) \Lambda$ can be written as

$$\frac{1}{N} \Lambda'(\hat{\Phi}^{-1} - \Phi^{-1}) \Lambda = - \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\phi}_i^2 - \phi_i^2}{\phi_i^4} \lambda_i' \lambda_i = - \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\phi}_i^2 - \phi_i^2}{\phi_i^4} \lambda_i' \lambda_i + \frac{1}{N} \sum_{i=1}^{N} \frac{(\hat{\phi}_i^2 - \phi_i^2)^2}{\phi_i^4} \lambda_i' \lambda_i$$

The last term is bounded in norm by $C^8 \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2$ and hence $O_p(T^{-1}) + O_p(N^{-2})$ by Theorem S.1. Thus we can rewrite (S.19) as

$$\frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} + \frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda) = \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\phi}_i^2 - \phi_i^2}{\phi_i^4} \lambda_i' \lambda_i + O_p(T^{-1}) + O_p(N^{-2})$$
By Lemma S.12, we can further write it as

\[
\frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} + \frac{1}{N} \hat{\Lambda} \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) = O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1}) \tag{S.35}
\]

Both \(\hat{M}_{ff}\) and \(M_{ff}\) are diagonal matrices. By (S.13) and Lemmas S.9 and S.11, we have

\[
\text{Ndiag}\left\{ \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff} + M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \right\} = \text{Ndiag}\{\zeta\} + O_p(N^{-1}) + O_p(T^{-1}) \tag{S.36}
\]

where \(\zeta\) is defined as \(\zeta = \frac{1}{NT} \sum_{t=1}^{T} e_t f_t' \Phi^{-1} \Lambda + \Lambda_0' \Phi^{-1} \frac{1}{NT} \sum_{t=1}^{T} e_t f_t\) and \(\text{Ndiag}(A)\) means the off-diagonal elements of \(A\). Since \(\zeta = O_p(N^{-1/2}T^{-1/2})\), we have (notice \(\hat{H} = \frac{1}{N} I_r\) under IC2)

\[
\text{Ndiag}\left\{ \frac{1}{N} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff} + M_{ff} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \frac{1}{N} \right\} = O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1}) \tag{S.37}
\]

Equation (S.35) puts \(\frac{1}{2} r(r+1)\) restrictions (instead of \(r^2\) due to symmetry) on \(\frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda}\), and equation (S.37) puts \(\frac{1}{2} r(r-1)\) restrictions. So the \(r \times r\) matrix \(\frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda}\) can be uniquely determined by solving (S.35) and (S.37). We have

\[
\frac{1}{N} (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \equiv (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} = O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1}) \tag{S.38}
\]

Given this result, it follows, by (S.13) and Lemma S.11,

\[
\hat{M}_{ff} - M_{ff} = O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}).
\]
Next, consider the right hand side of (S.14). The first term is $O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1})$ by (S.38) and Proposition S.1. The other terms except the 8th are all $O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1})$ due to the results of Lemmas S.9 and S.11. So it follows

$$\hat{\lambda}_j - \lambda_j = \hat{M}_{ff}^{-1} \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \Lambda \left( \frac{1}{T} \sum_{t=1}^{T} f_t e_{jt} \right) + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Since $\hat{M}_{ff}^{-1} \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \Lambda \overset{p}{\to} M_{ff}^{-1}$ by Proposition S.1 and Corollary S.1(c), it follows, under the condition $\sqrt{T}/N \to 0$,

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = M_{ff}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t e_{jt} + o_p(1) \quad (S.39)$$

So we have

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) \overset{d}{\to} N \left( 0, (M_{ff})^{-1} \gamma_2 \Lambda (M_{ff})^{-1} \right)$$

**Under IC3:** The matrix $M_{ff}$ is known, thus not estimated. The derivation of $\hat{\lambda}_j - \lambda_j$ is quite similar to IC2 and hence omitted.

**Under IC4:** Consider (S.23). By Lemmas S.9, S.11 and Corollary S.2, the right hand side of (S.23), except for the 1st and 8th terms, is $O_p(N^{-1}) + O_p(T^{-1})$. The 8th term is $\frac{1}{T} \sum_{t=1}^{T} f_t \xi'_t + O_p(T^{-1/2})$ by Corollary S.1(c). Thus by letting $A_4 = (\hat{\Lambda} - \Lambda) \hat{\Phi}^{-1} \hat{\Lambda} \hat{H}$ and multiplying $\Lambda'_1^{-1}$ on each side of (S.23), we obtain

$$\hat{M}_{ff}(\hat{\Lambda}'_1 - \Lambda'_1) \Lambda'_1^{-1} = M_{ff}A_4 + \frac{1}{T} \sum_{t=1}^{T} f_t \xi'_t \Lambda'_1^{-1} + O_p(N^{-1}) + O_p(T^{-1})$$

However, by the identification conditions, the left hand side is an upper triangular matrix, so its elements on and below the diagonal are all zeros, it follows that

$$\text{nonupper} \left\{ M_{ff}A_4 + \frac{1}{T} \sum_{t=1}^{T} f_t \xi'_t \Lambda'_1^{-1} \right\} = O_p(N^{-1}) + O_p(T^{-1}) \quad (S.40)$$
where nonupper denotes the elements on and below the diagonal. Since under IC4 both \( \hat{M}_{ff} \) and \( M_{ff} \) are diagonal matrices, equation (S.36) holds. The right hand side of (S.36) is \( O_p(N^{-1}) + O_p(T^{-1}) \). Rewrite (S.36) in terms of \( A_4 \),

\[
\text{nondiag}\left\{ A_4'M_{ff} + M_{ff}A_4 \right\} = O_p(N^{-1}) + O_p(T^{-1}) \tag{S.41}
\]

By solving the system of equations (S.40) and (S.41), we have

\[
(A_4)_{gh} = \begin{cases} 
-T^{-1} \sum_{t=1}^{T} m_g^{-1} f_g y d_{ht} + O_p(N^{-1}) + O_p(T^{-1}) & \text{if } g \geq h \\
-m_g^{-1} m_h(A_4)_{hg} + O_p(N^{-1}) + O_p(T^{-1}) & \text{if } g < h 
\end{cases} \tag{S.42}
\]

where \( d_{ht} = \xi_t A_1^{-1} v_h \), \( v_h \) is the \( h \)th column of an \( r \times r \) identity matrix, \( f_g \) is \( h \)th component of \( f_t \). That is,

\[
A_4 = P_t + O_p(N^{-1}) + O_p(T^{-1}) \tag{S.43}
\]

where \( P_t \) is defined in the main body of the text.

Consider (S.14). By Lemmas S.9 and S.11, (S.14) can be simplified as

\[
\hat{\lambda}_j - \lambda_j = \hat{M}_{ff}^{-1} M_{ff} A_4 \lambda_j + \hat{M}_{ff}^{-1} \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \Lambda \left( \frac{1}{T} \sum_{t=1}^{T} f_t e_{jt} \right) \\
+ O_p(N^{-1/2} T^{-1/2}) + O_p(N^{-1}) + O_p(T^{-1})
\]

Since \( \hat{M}_{ff}^{-1} \overset{p}{\rightarrow} M_{ff}^{-1} \) by Proposition S.1 and \( \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \Lambda \overset{p}{\rightarrow} I_r \) by Corollary S.1(c), we have, under the condition \( \sqrt{T}/N \rightarrow 0 \),

\[
\sqrt{T}(\hat{\lambda}_j - \lambda_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (P_t \lambda_j + M_{ff}^{-1} f_t e_{jt}) + o_p(1)
\]
By Assumption F.1, it follows, under the condition $\sqrt{T/N} \to 0$,

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, \Pi_j^\lambda)$$

It remains to derive the limiting distribution of $\hat{M}_{ff} - M_{ff}$. By Lemmas S.9 and S.11, equation (S.13) can be simplified, in terms of $A_4$, as

$$\hat{M}_{ff} - M_{ff} = -A_4'M_{ff} - M_{ff}A_4 + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Since both $\hat{M}_{ff}$ and $M_{ff}$ are diagonal matrices, we have

$$\text{diag}\{\hat{M}_{ff} - M_{ff}\} = -2\text{diag}\{M_{ff}A_4\} + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$= 2\text{diag}\left\{\frac{1}{T} \sum_{t=1}^{T} f_t \xi_t'N^{-1}_1\right\} + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

where the second equality follows from (S.40). By Assumption F.1, we have, under the condition $\sqrt{T}/\sqrt{N} \to 0$,

$$\sqrt{T}\text{diag}\{\hat{M}_{ff} - M_{ff}\} \xrightarrow{d} N\left(0, 4\text{J}_r\Pi_{M}^M\text{J}_r'\right)$$

where $\text{J}_r$ is defined as $\text{diag}\{M\} = \text{J}_r\text{vec}(M)$ for any $r \times r$ matrix $M$.

**Under IC5:** The derivation of limiting distribution of $\hat{\lambda}_j - \lambda_j$ is similar to IC4. The main difference is that for $A_5 = (\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}$, the solution by solving a system equations is, analogous to (S.43),

$$A_5 = \mathcal{Q}_t + O_p(N^{-1}) + O_p(T^{-1})$$

where $\mathcal{Q}_t$ is defined in the main body of the text. The details are omitted.

This completes the proof of Theorems 1–S.2. □
Corollary S.4  Assume that Assumptions A-E hold. Under either IC2 or IC3,

\[(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = O_p(N^{-1}) + O_p(T^{-1})\]

Proof of Corollary S.4: Under IC2, Corollary S.4 is immediately obtained by (S.38). Under IC3, an analogous result to (S.38) can still be derived. So Corollary S.4 holds. □

Supplement D: Proof of results for estimated factors

Lemma S.13  Under Assumptions A-E, we have

\[
\begin{align*}
(a) & \quad \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tilde{\phi}_i^2} (\hat{\lambda}_i - \lambda_i)e_{it} = O_p(N^{-3/2}) + O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) \\
(b) & \quad \frac{1}{N} \sum_{i=1}^{N} (\frac{1}{\hat{\phi}_i^2} - \frac{1}{\phi_i^2}) \lambda_i e_{it} = O_p(N^{-3/2}) + O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) \\
(c) & \quad \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \hat{\lambda}_i (e_{it} - \bar{e}_i) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it} + O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2})
\end{align*}
\]

Proof of Lemma S.13: Consider (a). Substituting (S.14) into (a), the left hand side can be expanded into an expression with 13 terms. The 1st term is equal to

\[
M_{ff}^{-1}M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it} + \frac{1}{N} \sum_{i=1}^{N} (\frac{1}{\hat{\phi}_i^2} - \frac{1}{\phi_i^2}) \lambda_i e_{it} \right)
\]

The term \(M_{ff}^{-1}M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = O_p(T^{-1/2}) + O_p(N^{-1})\) by Proposition S.1 and Corollary S.2. The term \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it} = O_p(N^{-1/2})\) due to \(E(\|\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it}\|^2) = \)
\[ O(N^{-1}) \text{ by Assumption C.3. The term } \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\phi_i^2} - \frac{1}{\phi_i^2} \right) \lambda_i e_{it} \text{ is bounded in norm by} \]

\[
C^4 \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2) \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \|\lambda_i e_{it}\|^2 \right)^{1/2}
\]

which is \( O_p(T^{-1/2}) + O_p(N^{-1}) \) by Proposition 1 and Assumption C.3. Given this result, we have the 1st term is \( O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2}) \). The 2nd-7th and 9th terms can be proved to be \( O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2}) \) similarly as the 1st one. The 11th term, which is \( \hat{M}_{ij}^{-1} \hat{H} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \hat{\lambda}_i e_{it} \), is of a smaller order term than \( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \hat{\lambda}_i e_{it} \). So it is negligible. We remain to check the 8th, 10th, 12th, and 13th terms. The 8th term is \( \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^2} f_s e_{is} e_{it} \), which is equivalent to

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^2} f_s [e_{is} e_{it} - E(e_{is} e_{it})] + \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^2} f_s \rho_{is,ts}
\]

The second expression \( \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^2} f_s \rho_{is,ts} \) is bounded by \( C^3 \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \rho_{is} \leq C^4 T^{-1} \) by \( \sum_{s=1}^{T} \rho_{is} \leq C \) by Assumption C.4’. The first expression can be written as

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^2} f_s [e_{is} e_{it} - E(e_{is} e_{it})] + \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^2} f_s \rho_{is,ts}
\]

The first expression \( \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^2} f_s [e_{is} e_{it} - E(e_{is} e_{it})] \) is \( O_p(N^{-1/2}T^{-1/2}) \) by Assumption E.4 and the second expression is bounded in norm by

\[
C^4 \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2) \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{s=1}^{T} f_s e_{is} - E(e_{is} e_{it}) \right\|^2 \right)^{1/2}
\]

which is \( O_p(T^{-1}) + O_p(N^{-1}T^{-1/2}) \) by Assumption E.5 and Proposition 1. So the 8th term is \( O_p(T^{-1}) + O_p(N^{-1}T^{-1/2}) \).

Consider the 10th term, which is equal to

\[
\hat{M}_{ij}^{-1} \hat{H} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \hat{\lambda}_j \frac{1}{T} \sum_{s=1}^{T} [e_{js} e_{is} - E(e_{js} e_{is})] \frac{1}{\phi_i^2} e_{it}
\]

37
We use $\epsilon_{ij,s} = e_i e_{js} - E(e_i e_{js})$ temporarily. The above term is equal to

$$
\hat{M}_{ff}^{-1} \hat{H}_N \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} (\hat{\lambda}_j - \lambda_j) e_{it} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s}
$$

$$
+ \hat{M}_{ff}^{-1} \hat{H}_N \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \frac{1}{\phi_i^2} \lambda_j e_{it} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s}
$$

$$
+ \hat{M}_{ff}^{-1} \hat{H}_N \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2} \lambda_j e_{it} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s}
$$

The first expression is bounded in norm by

$$
C^3 \| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} e_{it}^2 \right)^{1/2} \left( \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \| \hat{\lambda}_j - \lambda_j \| \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s}^2 \right)^{1/2}
$$

which is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Proposition 1 and Assumption C.5. The second expression is bounded in norm by

$$
C^7 \| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} e_{it}^2 \right)^{1/2} \left( \frac{1}{N} \sum_{j=1}^{N} (\hat{\phi}_j^2 - \phi_j^2)^2 \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s}^2 \right)^{1/2}
$$

which is also $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Proposition 1 and Assumption C.5. The third expression is bounded in norm by

$$
\| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N^2} \sum_{i=1}^{N} \frac{(\hat{\phi}_i^2 - \phi_i^2)^2}{\phi_i^4 \phi_i^4} \sum_{j=1}^{N} \| \lambda_j \|^2 \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{it}^2 \left( \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s} \right)^2 \right)^{1/2}
$$

which is further bounded by

$$
C^{10} \| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{it}^2 \left( \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s} \right)^2 \right)^{1/2}
$$

38
The last factor of the above expression is bounded by

\[ \left( \frac{1}{N} \sum_{i=1}^{N} e_{it}^4 \right)^{1/4} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s} \right)^4 \right)^{1/2} \]

which is \( O_p(N^{-1/2}T^{-1/2}) \) by Assumption C.1 and C.5. So the third expression is \( O_p(T^{-1}) + O_p(N^{-1}T^{-1/2}) \). The last expression is bounded in norm by

\[ \| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{it}^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N^2} \sum_{j=1}^{N} \sum_{s=1}^{T} \lambda_{ij} \epsilon_{ij,s} \right)^2 \right)^{1/2} \]

which is \( O_p(N^{-1/2}T^{-1/2}) \) by Assumption E.2. Given all the results, it follows that the 10th term is \( O_p(T^{-1}) + O_p(N^{-1}T^{-1/2}) \).

The 12th term is equal to

\[ \hat{M}_{ff}^{-1} \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} \hat{\lambda}_j \epsilon_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \]

The above term can be split into

\[ \hat{M}_{ff}^{-1} \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} (\hat{\lambda}_j - \lambda_j) \epsilon_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \]

\[ + \hat{M}_{ff}^{-1} \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} (\hat{\lambda}_j - \lambda_j) \epsilon_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \]

\[ + \hat{M}_{ff}^{-1} \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} \lambda_j \epsilon_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \]

The first expression of the above is bounded in norm by

\[ C^3 \| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{it}^2 \right)^{1/2} \left( \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \| \hat{\lambda}_j - \lambda_j \|^2 \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right)^2 \right)^{1/2} \]
Since

\[
\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right)^2 \leq \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tau_{ij}^2 \leq \left( \sup_{i,j \leq N} \tau_{ij} \right) \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tau_{ij} = O(N^{-1}) \quad (S.46)
\]

by \( \sup_{i,j \leq N} \tau_{ij} \leq \sup_{i \leq N} \sum_{j=1}^{N} \tau_{ij} \leq C \), we have the first term is \( O_{p}(N^{-1/2}T^{-1/2}) + O_{p}(N^{-3/2}) \) by Proposition 1. The second expression is bounded in norm by

\[
C^7 \| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \hat{\phi}_i^2)^2 \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{it}^2 \left( \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right)^2 \right)^{1/2}
\]

which is \( O_{p}(N^{-1/2}T^{-1/2}) + O_{p}(N^{-3/2}) \) by

\[
\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{it}^2 \left( \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right)^2 \leq \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{it}^2 \tau_{ij}^2 \leq \left( \sup_{i,j \leq N} \tau_{ij} \right) \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{it}^2 \sum_{j=1}^{N} \tau_{ij}
\]

\[\leq C \left( \sup_{i,j \leq N} \tau_{ij} \right) \frac{1}{N^2} \sum_{i=1}^{N} e_{it}^2 = O_{p}(N^{-1})\]

The third expression is bounded in norm by

\[
C^7 \| \hat{M}_{ff}^{-1} \hat{H}_N \| \left( \frac{1}{N} \sum_{i=1}^{N} e_{it}^2 \right)^{1/2} \left( \frac{1}{N} \sum_{j=1}^{N} (\hat{\phi}_j^2 - \hat{\phi}_j^2)^2 \right)^{1/2} \left( \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right)^2 \right)^{1/2}
\]

which is \( O_{p}(N^{-1/2}T^{-1/2}) + O_{p}(N^{-3/2}) \) by Proposition 1 and (S.46). Consider the last expression. Since

\[
E \left( \left\| \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^2 \phi_j^2} \lambda_j \lambda_m E(e_{it} e_{mt}) \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right\|^2 \right)
\]

\[
= \frac{1}{N^4} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \frac{1}{\phi_i^2 \phi_j^2 \phi_m^2 \phi_n^2} \lambda_j \lambda_m E(e_{it} e_{mt}) \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \frac{1}{T} \sum_{s=1}^{T} \tau_{mn,s}
\]

\[\leq C_{10} \frac{1}{N^4} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \tau_{ij}^2 \tau_{mn} \leq C_{11} \frac{1}{N^4} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} \tau_{im} \tau_{mj} \quad (S.47)
\]

\[\leq C_{12} \frac{1}{N^4} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} \tau_{im} \leq C_{13} N^{-3}\]
by Assumption C.3. So the last expression is \( O_p(N^{-3/2}) \). Summing the four expressions gives the 12th term is \( O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2}) \). The 13th term is \( O_p(T^{-1}) \) which can be easily verified.

Summarizing results, we have

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\phi_i^2} (\hat{\lambda}_i - \lambda_i) e_{it} \right) = O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}).
\]

Consider (b), which can be written as

\[
- \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} (\hat{\phi}_i^2 - \phi_i^2) \lambda_i e_{it} \quad - \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2 \phi_i^4} (\hat{\phi}_i^2 - \phi_i^2)^2 \lambda_i e_{it} \quad \text{(S.48)}
\]

Using (S.25), the term \( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} (\hat{\phi}_i^2 - \phi_i^2) \lambda_i e_{it} \) can be expanded into a 13-term expression. The first term is

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i^4} \lambda_i (e_{is}^2 - \phi_i^2) e_{it}
\]

which is \( O_p(N^{-1/2}T^{-1/2}) \) by Assumption E.6. The second term is

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} (\hat{\lambda}_i - \lambda_i) \hat{M}_{ff} (\hat{\lambda}_i - \lambda_i) \lambda_i e_{it}
\]

The above expression is bounded in norm by

\[
C^4 \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\lambda}_i - \lambda_i \|^2 \right)^{1/2} \| \hat{M}_{ff} \| \left( \frac{1}{N} \sum_{i=1}^{N} \| (\hat{\lambda}_i - \lambda_i) \lambda_i e_{it} \|^2 \right)^{1/2}
\]

The first factor is \( O_p(T^{-1/2}) + O_p(N^{-1}) \) The last factor is bounded by

\[
C^2 \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\lambda}_i - \lambda_i \|^2 e_{it}^2 \right)^{1/2}
\]
Using the argument following (S.26) on \(\|\hat{\lambda}_i - \lambda_i\|^2\), the above is also \(O_p(T^{-1/2}) + O_p(N^{-1})\). Given these two results, the second term is \(O_p(T^{-1}) + O_p(N^{-2})\).

The third term is

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \left( \lambda_i' \hat{\Lambda} \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff}(\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{\lambda}_i \right) \lambda_i e_{it}.
\]

Its \(k\)th element \((k = 1, 2, \cdots, r)\) can be written as

\[
\text{tr} \left[ \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) M_{ff}(\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{\lambda}_i \right] \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_i' \lambda_k e_{it} \right)
\]

The term \(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' \lambda_k e_{it}\) is \(O_p(N^{-1/2})\) due to \(E\left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' \lambda_k e_{it}\right)^2 = O(N^{-1})\) by Assumption C.3. So the third term is \(O_p(N^{-1/2}T^{-1}) + O_p(N^{-5/2})\) in view of Lemma S.11(a). The 4th-8th terms can be proved to be \(O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(N^{-3/2})\) similarly as the third term.

The 9th term is equal to \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_i' e_{it} \left( \lambda_i' \hat{\Lambda} \right)\). Its \(k\)th element can be written as

\[
\text{tr} \left[ \hat{H} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_k \lambda_i' e_{it} + \hat{H} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_k (\hat{\lambda}_i - \lambda_i) \lambda_i' e_{it} \right]
\]

The first expression is \(O_p(N^{-3/2})\) by Assumption C.3 and \(\|\hat{H}\| = O_p(N^{-1})\). The second expression inside the trace operator is bounded in norm by

\[
O_p \left( \left( \frac{1}{N} \sum_{i=1}^{N} \phi_i^2 \lambda_i e_{it}^2 \right)^{1/2} \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \|\hat{\lambda}_i - \lambda_i\|^2 \right)^{1/2}
\]

which is \(O_p(N^{-1}T^{-1/2}) + O_p(N^{-2})\) by \(\|\hat{H}\| = O_p(N^{-1})\) [Corollary S.1(a)] and Proposition 1. So the 9th term is \(O_p(N^{-3/2}) + O_p(N^{-1}T^{-1/2})\).

The 10th term is equal to

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_i' \hat{H} \sum_{j=1}^{T} \sum_{s=1}^{T} \frac{1}{\phi_j^2} \hat{\lambda}_j E(e_{js} e_{it}) \lambda_i' e_{it}
\]
Its \(k\)th element can be written as

\[
tr \left[ \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} \hat{\lambda}_j \lambda_i' \lambda_{ik} e_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right]
\]

The above expression is equal to

\[
tr \left[ \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} (\hat{\lambda}_j - \lambda_j) \lambda_i' \lambda_{ik} e_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right]
\]

\[
+ tr \left[ \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} \lambda_j \lambda_i' \lambda_{ik} e_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right]
\]

\[
tr \left[ \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} \hat{\lambda}_j \lambda_i' \lambda_{ik} e_{it} \frac{1}{T} \sum_{s=1}^{T} \tau_{ij,s} \right]
\]

Using the argument in analyzing (S.45), each of the first two expressions is \(O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2})\) and the third expression is \(O_p(N^{-3/2})\). So the 10th term is \(O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2})\).

The 11th term is equal to

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^4} \lambda_i' \hat{H} \sum_{j=1}^{N} \frac{1}{\phi_j^2} \hat{\lambda}_j \lambda_i e_{it} \frac{1}{T} \sum_{s=1}^{T} [e_{js} e_{is} - E(e_{js} e_{is})] \lambda_i e_{it}
\]

We use \(\epsilon_{ij,s} = e_{is} e_{js} - E(e_{is} e_{js})\) temporarily. Its \(k\)th element is

\[
tr \left[ \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} \hat{\lambda}_j \lambda_i' \lambda_{ik} e_{it} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s} \right]
\]

which can be written as

\[
tr \left[ \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} (\hat{\lambda}_j - \lambda_j) \lambda_i' \lambda_{ik} e_{it} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s} \right]
\]

\[
tr \left[ \hat{H}_N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i^4 \phi_j^2} \lambda_j \lambda_i' \lambda_{ik} e_{it} \frac{1}{T} \sum_{s=1}^{T} \epsilon_{ij,s} \right]
\]
\[
tr \left[ \hat{H} N \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\phi_i \phi_j} \lambda_j \lambda_i \epsilon_{it} \right] \frac{1}{T} \sum_{s=1}^{T} \epsilon_{i,s}
\]

Using argument in analyzing (S.44), each of the first two expressions is \(O_p(T^{-1}) + O_p((NT)^{-1/2})\) and the third expression is \(O_p((NT)^{-1/2})\). So the 11th term is \(O_p(T^{-1}) + O_p((NT)^{-1/2})\).

The 12th term is equal to
\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i} \lambda_i \epsilon_{it} (\lambda_i' \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \hat{\Lambda} - \Lambda) \frac{1}{T} \sum_{s=1}^{T} f_s e_{is}
\]

Its \(k\)th element is
\[
tr \left[ \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \hat{\Lambda} - \Lambda \right] \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i} \lambda_i f_s \lambda_i' \epsilon_{it} e_{is}
\]

Since the term \(\frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \frac{1}{\phi_i} \lambda_i f_s \lambda_i' \epsilon_{it} e_{is}\) is bounded in norm by
\[
\left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i} \|\lambda_i\|^2 \epsilon_{it}^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{s=1}^{T} f_s e_{is}^2 \right)^{1/2}
\]

which is \(O_p(T^{-1/2})\) by (S.5). Thus the \(k\)th element of the 12th term is \(O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})\) by Corollary S.2. So the 12th term is \(O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})\).

The 13th term is equal to \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i} \lambda_i \epsilon_{it} (\lambda_i' \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \bar{e} \bar{e}_i)\). Its \(k\)th element is
\[
tr \left[ \hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \bar{e} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i} \lambda_i \epsilon_{it} \bar{e}_i \lambda_i' \right]
\]

The term \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i} \lambda_i \epsilon_{it} \bar{e}_i \lambda_i'\) is bounded in norm by
\[
C^6 \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{it}^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} e_{it}^2 \right)^{1/2}
\]

which is \(O_p(T^{-1/2})\) by Assumption C.4. However, the term \(\hat{H} \hat{\Lambda} \hat{\Phi}^{-1} \bar{e}\) is \(O_p(T^{-1/2})\) by Lemma S.9(b). So the last term is \(O_p(T^{-1})\).
Summarizing results, we have

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} (\hat{\phi}_i^2 - \phi_i^2) \lambda_i e_{it} = O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2}).
\]

Next, consider the second term of (S.48), i.e. \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} (\hat{\phi}_i^2 - \phi_i^2)^2 \lambda_i e_{it}\), which can be written as

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{\hat{\phi}_i^2 - \phi_i^2}{\phi_i^2} \right) \left( (\hat{\phi}_i^2 - \phi_i^2) \lambda_i e_{it} \right)
\]

By the Cauchy-Schwarz inequality, the above term is bounded in norm by

\[
\left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\hat{\phi}_i^2 - \phi_i^2}{\phi_i^2} \right)^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left| (\hat{\phi}_i^2 - \phi_i^2) \lambda_i e_{it} \right|^2 \right)^{1/2}
\]

which is further bounded by

\[
C^2 \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 e_{it}^2 \right)^{1/2}
\]

The first factor of the above expression is \(O_p(T^{-1/2}) + O_p(N^{-1})\). Using the argument following (S.25) on \(\|\hat{\phi}_i - \phi_i\|^2\), the second factor of the above expression is also \(O_p(T^{-1/2}) + O_p(N^{-1})\). This yields that

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} (\hat{\phi}_i^2 - \phi_i^2)^2 \lambda_i e_{it} = O_p(T^{-1}) + O_p(N^{-2}),
\]

completing the proof of (b).

Consider (c). The term \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \hat{\lambda}_i \hat{e}_i\) can be written as

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \hat{\lambda}_i \hat{e}_i + \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\phi_i^2} - \frac{1}{\phi_i^2} \right) \lambda_i e_i + \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \bar{e}_i = c_1 + c_2 + c_3
\]
Term $c_1$ is bounded in norm by

$$C \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \| \hat{\lambda}_i - \lambda_i \|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \bar{e}_i^2 \right)^{1/2}$$

which is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Proposition 1 and $ar{e}_i = O_p(T^{-1/2})$.

Term $c_2$ is bounded in norm by

$$C^5 \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{\phi}_i^2 - \phi_i^2)^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \bar{e}_i^2 \right)^{1/2}$$

which is also $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by the same argument as $c_1$.

Term $c_3$ is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} \hat{\lambda}_i \bar{e}_i$$

which can be written as

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} (\hat{\lambda}_i - \lambda_i) e_{it} + \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\phi}_i^2} (\frac{1}{\hat{\phi}_i^2} - \frac{1}{\phi_i^2}) \lambda_i e_{it} + \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it}$$

By parts (a) and (b) of this lemma, we have

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \hat{\lambda}_i e_{it} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it} + O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2})$$

This yields (c). □

Given Lemma S.13, the proof of Theorem 2 is the same as those of Proposition 6.1 and Theorem 6.1 of Bai and Li (2012). The details are omitted here.

The following average consistency result for the estimated factors is due to Lemma S.13

**Proposition S.3** Assume that Assumptions A-E hold. Under each of IC1, IC4, and IC5, we have

$$\frac{1}{T} \sum_{t=1}^{T} \| \hat{f}_t - f_t \|^2 = O_p \left( \frac{1}{N} \right) + O_p \left( \frac{1}{T} \right),$$

46
and under IC2 or IC3, we have

\[
\frac{1}{T} \sum_{t=1}^{T} \| \hat{f}_t - f_t \|^2 = O_p\left( \frac{1}{N} \right) + O_p\left( \frac{1}{T^2} \right).
\]

**Remark:** The different convergence rates for \( \frac{1}{T} \sum_{t=1}^{T} \| \hat{f}_t - f_t \|^2 \) are due to the different convergence rates of \( I_r - \Lambda' \hat{\Phi}^{-1} \hat{\Lambda} (\Lambda' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \) under different identification restrictions. As pointed out in the discussion preceding Proposition 1 the matrix \( \Lambda' \hat{\Phi}^{-1} \hat{\Lambda} (\Lambda' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \) plays the same role as the rotation matrix and its asymptotic property depends on the identification conditions.

Furthermore, if one is more interested in the factor process \( f_t \), it can be directly estimated by the maximum likelihood method. Putting the model in the form \( z_i = \delta + F \lambda_i + e_i \), where \( F = (f_1, ..., f_T)' \), and \( z_i \) is \( T \times 1 \) (instead of \( N \times 1 \)). In this setup, we avoid estimating \( \Lambda \), but only the sample variance of the factor loadings. And we would have

\[
\frac{1}{T} \sum_{t=1}^{T} \| \hat{f}_t - f_t \|^2 = O_p(N^{-1}) + O_p(T^{-2})
\]

under all identification conditions, an analogous result to Proposition 1 by switching the role of \( N \) and \( T \). Directly estimating \( f_t \) is preferred when \( T \) is small relative to \( N \). This is because the number of parameters in \( F \) is smaller than in \( \Lambda \).

**Supplement E: Assumptions and proofs for Section 5**

The following assumptions are needed to derive the limiting results in Theorem 4. In what follows, \( C \) is a generic constant large enough.

**Assumption 5A:** Assumption A is satisfied when \( f_t \) are fixed constants. When \( f_t \) is a random process, \( f_t \) admits a wold representation \( f_t = u_t + C_1 u_{t-1} + C_2 u_{t-2} + \ldots \) such that \( \sum_{i=1}^{\infty} \| C_i \| < \infty \) and \( u_t \) is an i.i.d process with \( E\| u_t \|^{4} < \infty \).
Assumption 5B: The factor loadings $\lambda_i$ satisfy $\|\lambda_i\| \leq C$ for all $i$. In addition, there exists an $r \times r$ positive matrix $Q$ such that $\lim_{N \to \infty} N^{-1} \Lambda' \Phi^{-1} \Lambda = Q$, where $\Phi = \text{diag}(\phi_1^2, \ldots, \phi_N^2)$ with $\phi_i^2 = E(e_{it}^2)$.

Assumption 5C: The idiosyncratic error terms $e_{it}$ satisfy

1. The lags $p_i$ are bounded by some $p_{\max}$ for all $i$;
2. The roots of the polynomial $\rho_i(L) = 1 - \rho_{i,1}L - \cdots - \rho_{i,p_i}L^{p_i}$ are outside the unit circle for all $i$ (uniformly bounded away from 1 in norm);
3. The variance of the innovation $\epsilon_{it}$, denoted by $\sigma_{\epsilon i}^2$, is bounded from above and below, i.e., $C^{-2} \leq \sigma_{\epsilon i}^2 \leq C^2$ for all $i$. Furthermore, $\epsilon_{it}$ is independent over $i$ and i.i.d. over $t$ for each given $i$. The fourth moment of $\epsilon_{it}$ is bounded for each $i$, i.e., $E(\epsilon_{it}^4) \leq C$.

These assumptions imply that $\phi_i^2 = E(e_{it})^2$ is bounded above and away from zero.

For ease of reference, we list the symbols used in the following proofs.

$$
\psi_{it} = (e_{it-1}, e_{it-2}, \ldots, e_{it-p_i})', \quad \text{accordingly } \hat{\psi}_{it} = (\hat{e}_{it-1}, \hat{e}_{it-2}, \ldots, \hat{e}_{it-p_i})'
$$

$$
\rho_i = (\rho_{i,1}, \rho_{i,2}, \ldots, \rho_{i,p_i})', \quad \text{accordingly } \hat{\rho}_i = (\hat{\rho}_{i,1}, \hat{\rho}_{i,2}, \ldots, \hat{\rho}_{i,p_i})'
$$

$$
g_{it} = f_t - \rho_{i,1}f_{t-1} - \cdots - \rho_{i,p_i}f_{t-p_i}, \quad \text{accordingly } \hat{g}_{it} = \hat{f}_t - \hat{\rho}_{i,1}\hat{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\hat{f}_{t-p_i}
$$

$$
\Delta f_{t-p} = \hat{f}_{t-p} - f_{t-p}, \quad \text{for } p = 0, 1, \ldots, p_i
$$

$$
\Delta \lambda_i = \hat{\lambda}_i - \lambda_i,
$$

$$
\Delta \rho_{i,p} = \hat{\rho}_{i,p} - \rho_{i,p} \quad \text{for } p = 1, \ldots, p_i.
$$

we use $\bar{p}_i$ to denote $p_i + 1$ for notational simplicity. Since the identification conditions (Assumption 5D) employed in the present setting is IC3, Corollary S.4 holds. The following two lemmas are useful.
Lemma S.14 Under Assumptions 5A-5D, we have

(a) \[ \frac{1}{T - p_i} \sum_{t=p_i}^T \hat{f}_{t-p} f'_{t-q} = O_p(N^{-1}) + O_p(T^{-1}) \quad \text{for } p, q = 0, 1, \ldots, p_i \]

(b) \[ \frac{1}{T - p_i} \sum_{t=p_i}^T f_{t-p} \hat{f}'_{t-q} = O_p(N^{-1}) + O_p(T^{-1}) \quad \text{for } p, q = 0, 1, \ldots, p_i \]

(c) \[ \frac{1}{T - p_i} \sum_{t=p_i}^T \hat{f}_{t-p} \hat{f}'_{t-q} = O_p(N^{-1}) + O_p(T^{-1}) \quad \text{for } p, q = 0, 1, \ldots, p_i \]

Proof of Lemma S.14: Consider (a).

\[ \hat{f}_{t-p} - f_{t-p} = - (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1} (\hat{\Lambda} - \Lambda) f_{t-p} + (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Phi}^{-1} e_{t-p} \]

\[ = - A' f_{t-p} + \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} e_{t-p} \]

(S.49)

where \( \hat{H} = (\hat{\Lambda}' \hat{\Phi}^{-1} \hat{\Lambda})^{-1} \) and \( A = (\hat{\Lambda} - \Lambda)' \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \). The left hand side of (a) equals

\[ A' \left( \frac{1}{T - p_i} \sum_{t=p_i}^T f_{t-p} f'_{t-q} \right) A + \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} \frac{1}{T - p_i} \sum_{t=p_i}^T (e_{t-p} e'_{t-q}) \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \]

\[ - \left( \frac{1}{T - p_i} \sum_{t=p_i}^T \hat{H} \hat{\Lambda}' \hat{\Phi}^{-1} e_{t-p} f'_{t-q} \right) A - A' \left( \frac{1}{T - p_i} \sum_{t=p_i}^T f_{t-p} f'_{t-q} \hat{\Phi}^{-1} \hat{\Lambda} \hat{H} \right) \]

(S.50)

The first term of (S.50) is \( O_p(N^{-2}) + O_p(T^{-2}) \) by \( \frac{1}{T - p_i} \sum_{t=p_i}^T f_{t-p} f'_{t-q} = O_p(1) \) and Corollary S.4. The second term is equal to

\[ \hat{H} \left( \sum_{i=1}^N \sum_{j=1}^N \frac{1}{\hat{\phi}^2_{i,j}} \hat{\lambda}_i \hat{\lambda}_j \frac{1}{T - p_i} \sum_{t=p_i}^T [e_{it-p} e_{jt-q} - E(e_{it-p} e_{jt-q})] \right) \hat{H} \]

\[ + \hat{H} \left( \sum_{i=1}^N \sum_{j=1}^N \frac{1}{\hat{\phi}^2_{i,j}} \hat{\lambda}_i \hat{\lambda}_j \frac{1}{T - p_i} \sum_{t=p_i}^T E(e_{it-p} e_{jt-q}) \right) \hat{H} \]

The first expression can be proved to be \( O_p(N^{-1/2} T^{-1/2}) + O_p(T^{-1}) \) similarly as Lemma S.11(d). The second expression is equal to \( \hat{H} \left[ \sum_{i=1}^N \frac{1}{\hat{\phi}^2_{i}} \hat{\lambda}_i \hat{\lambda}_i \frac{1}{T - p_i} \sum_{t=p_i}^T E(e_{it-p} e_{it-q}) \right] \hat{H} \)
by the assumption of cross-sectional independence, which is further bounded by

\[ O^2 \| \hat{H}^{1/2} \|^2 \left( \sum_{i=1}^{N} \left\| \frac{\hat{X}_i}{\phi_i} \right\| ^2 \right) \cdot \sup_i |E(e_{it-p}e_{it-q})| = O_p(N^{-1}) \]

So the second term of (S.50) is \( O_p(N^{-1}) + O_p(T^{-1}) \). Term \( \frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{H} \hat{\Phi}^{-1} e_{it-p} f'_{t-q} \)

is \( O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \), which can be shown similarly as Lemma S.11(e), so

the third term of (S.50) is \( O_p(N^{-3/2}T^{-1/2}) + O_p(T^{-2}) \). The last term of (S.50) is also

\( O_p(N^{-3/2}T^{-1/2}) + O_p(T^{-2}) \) by similar arguments.

Summarizing results, we obtain (a).

Consider (b). By (S.49), the left hand side of (b) is equal to

\[- \left( \frac{1}{T-p_i} \sum_{t=t_i}^{T} f_{t-p} f'_{t-q} \right) A + \frac{1}{T-p_i} \sum_{t=t_i}^{T} f_{t-p} f'_{t-q} \hat{H} \hat{\Phi}^{-1} \hat{H} \]

The first term is \( O_p(N^{-1}) + O_p(T^{-1}) \) by Corollary S.4. The second term can be proved to be \( O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \) similarly as Lemma S.11(d). Then (b) follows.

Consider (c). Notice

\[
\frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{f}_{t-p} \hat{f}'_{t-q} = \frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{f}_{t-p} \hat{f}'_{t-q} + \frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{f}_{t-p} \hat{f}'_{t-q}.
\]

So (c) follows immediately by (a) and (b). \( \square \)

**Lemma S.15**  Under Assumptions 5A-5D,

(a) \( \frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{f}'_{t-q} e_{it} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \), for \( q = 1, 2, \ldots, p_i \)

(b) \( \frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{f}'_{t-q} e_{it} = O_p(T^{-1/2}) \), for \( q = 1, 2, \ldots, p_i \)

(c) \( \frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{f}'_{t-p} e_{it-q} = O_p(N^{-1}) + O_p(T^{-1}) \), for \( p, q = 0, 1, 2, \ldots, p_i \)

(d) \( \frac{1}{T-p_i} \sum_{t=t_i}^{T} \hat{f}'_{t-p} e_{it-q} = O_p(T^{-1/2}) \), for \( p, q = 0, 1, 2, \ldots, p_i \)

50
Proof of Lemma S.15: Consider (a). By (S.49), the left hand side of (a) is

\[-A' \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T f_{t-q} \epsilon_{it} + H \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T \hat{\lambda}' \hat{\Phi}^{-1} \epsilon_{t-q} \epsilon_{it}\]

The first term is \(O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2})\) by \(A = O_p(N^{-1}) + O_p(T^{-1})\) as in Corollary S.4. The second term is equal to \(\hat{H} \sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2} \hat{\lambda}_i \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T e_{it-q} \epsilon_{it}\). Notice \(E(\epsilon_{it-q} \epsilon_{it}) = 0\), thus this term can be proved to be \(O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})\) similarly as Lemma S.11(c). Given these results, we have (a).

Consider (b). The left hand side of (b) is equal to

\[\frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T \Delta f_{t-q} \epsilon_{it} + \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T f_{t-q} \epsilon_{it}\]

The first term is \(O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})\) as in (a). The second term is \(O_p(T^{-1/2})\). These results imply (b).

Consider (c). By (S.49), the left hand side of (c) is equal to

\[-A' \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T f_{t-p} e_{it-q} + H \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T \hat{\lambda}' \hat{\Phi}^{-1} e_{t-p} \epsilon_{it-q}\]

The first expression is \(O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2})\) by Corollary S.4. The second expression can be split into

\[\frac{1}{T - p_i} \hat{H} \sum_{j=1}^{N} \sum_{t = \hat{p}_i}^T \frac{1}{\hat{\sigma}_j^2} \hat{\lambda}_j [e_{jt-p} e_{it-q} - E(e_{jt-p} e_{it-q})] + \frac{1}{\hat{\sigma}_i^2} \hat{H} \hat{\lambda}_i \left( \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T E(e_{it-p} e_{it-q}) \right)\]

The first term can be proved to be \(O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})\) similarly as Lemma S.11(c). The second term is \(O_p(N^{-1})\) by \(\hat{\phi}_i^2 \xrightarrow{p} \phi_i^2, \hat{\lambda}_i \xrightarrow{p} \lambda_i, \frac{1}{T - p_i} \sum_{t = \hat{p}_i}^T E(e_{it-p} e_{it-q}) = O(1)\) and \(\hat{H} = O_p(N^{-1})\). Given these results, (c) follows.
Consider (d). Notice

\[
\frac{1}{T - p_i} \sum_{t = p_i}^{T} \hat{e}_{t-p} = \frac{1}{T - p_i} \sum_{t = p_i}^{T} e_{it} + O_p(T^{-1/2}) + O_p(T^{-1}), \quad \text{for } p = 1, \ldots, p_i
\]

The second term of the right hand side is \(O_p(T^{-1/2})\). Then (d) follows by (c). \(\square\)

The following lemma is useful in deriving the asymptotic representation of \(\hat{\rho}_i - \rho_i\).

Lemma S.16 **Under Assumptions 5A-5D,**

(a) \[
\frac{1}{T - p_i} \sum_{t = p_i}^{T} \hat{e}_{t-p} \epsilon_{it-p} = \frac{1}{T - p_i} \sum_{t = p_i}^{T} e_{it-p} \epsilon_{it} + O_p(T^{-1/2}) + O_p(T^{-1}), \quad \text{for } p = 1, \ldots, p_i
\]

(b) \[
\frac{1}{T - p_i} \sum_{t = p_i}^{T} \hat{e}_{t-p} \hat{\Delta} f'_{t-q} = O_p(T^{-1/2}) + O_p(T^{-1}), \quad \text{for } p, q = 0, 1, \ldots, p_i
\]

(c) \[
\frac{1}{T - p_i} \sum_{t = p_i}^{T} \hat{e}_{t-p} \hat{f}'_{t-q} = O_p(T^{-1/2}) + O_p(T^{-1}), \quad \text{for } p, q = 0, 1, \ldots, p_i
\]

(d) \[
\frac{1}{T - p_i} \sum_{t = p_i}^{T} \hat{e}_{t-p} \hat{e}_{it-q} = \frac{1}{T - p_i} \sum_{t = p_i}^{T} e_{it-p} \epsilon_{it-q} + O_p(T^{-1/2}) + O_p(T^{-1}), \quad \text{for } p, q = 1, \ldots, p_i
\]

**Proof of Lemma S.16**: Consider (a). By

\[
\hat{e}_{it-p} = e_{it-p} - \lambda_i'(\hat{f}_{t-p} - f_{t-p}) - (\hat{\lambda}_i - \lambda_i)' \hat{f}_{t-p} = e_{it-p} - \lambda_i' \hat{\Delta} f_{t-p} - \hat{\Delta} \lambda_i' \hat{f}_{t-p}, \quad \text{(S.51)}
\]

we have the left hand side of (a) is equal to

\[
\frac{1}{T - p_i} \sum_{t = p_i}^{T} e_{it-p} \epsilon_{it} - \lambda_i' \left( \frac{1}{T - p_i} \sum_{t = p_i}^{T} \hat{\Delta} f_{t-p} \epsilon_{it} \right) - \hat{\Delta} \lambda_i' \left( \frac{1}{T - p_i} \sum_{t = p_i}^{T} \hat{f}_{t-p} \epsilon_{it} \right)
\]

The second term of the above expression is \(O_p(T^{-1/2}) + O_p(T^{-1})\) by Lemma S.15(a). The third term is \(O_p(T^{-1/2}) + O_p(T^{-3/2})\) by Lemma S.15(b) and \(\hat{\Delta} \lambda_i = O_p(T^{-1}) + O_p(T^{-1/2})\). Given these results, (a) follows.
Consider (b). By (S.51), the left hand side of (b) is equal to

\[
\frac{1}{T - p_i} \sum_{t=p_i}^{T} e_{it-p} \Delta f'_{t-q} - \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \Delta f_{t-p} \Delta f'_{t-q} \right) - \Delta \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{f}_{t-p} \Delta f'_{t-q} \right)
\]

The first term of the above expression is \(O_p(N^{-1}) + O_p(T^{-1})\) by Lemma S.15(c). The second term is \(O_p(N^{-1}) + O_p(T^{-1})\) by Lemma S.14(a) and the third term is \(O_p(N^{-2}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2})\) by Lemma S.14(c) and \(\Delta \lambda_i = O_p(N^{-1}) + O_p(T^{-1/2})\). Then (b) follows.

Consider (c). By (S.51), the left hand side of (c) is equal to

\[
\frac{1}{T - p_i} \sum_{t=p_i}^{T} e_{it-p} \hat{f}'_{t-q} - \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \Delta f_{t-p} \hat{f}'_{t-q} \right) - \Delta \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{f}_{t-p} \hat{f}'_{t-q} \right)
\]

The first term is \(O_p(T^{-1/2})\) by Lemma S.15(d). The second term is \(O_p(N^{-1}) + O_p(T^{-1})\) by Lemma S.14(c). The third term is \(O_p(N^{-1}) + O_p(T^{-1/2})\) by \(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{f}_{t-p} \hat{f}'_{t-q} = O_p(1)\), which is the result of Lemma S.14(b) and (c). Then (c) follows.

Consider (d). By (S.51), the left hand side of (d) is equal to

\[
\frac{1}{T - p_i} \sum_{t=p_i}^{T} e_{it-p} e_{it-q} - \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \Delta f_{t-p} e_{it-q} \right) - \Delta \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{f}_{t-p} e_{it-q} \right)
\]

\[-\lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \Delta f_{t-q} e_{it-p} \right) + \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \Delta f_{t-p} \Delta f'_{t-q} \right) \lambda_i + \Delta \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{f}_{t-p} \Delta f'_{t-q} \right) \lambda_i
\]

\[-\Delta \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{f}_{t-q} e_{it-p} \right) + \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \Delta f_{t-p} \hat{f}'_{t-q} \right) \Delta \lambda_i + \Delta \lambda_i' \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{f}_{t-p} \hat{f}'_{t-q} \right) \Delta \lambda_i
\]

The second and fourth terms are \(O_p(N^{-1}) + O_p(T^{-1})\) by Lemma S.15(c). The third and seventh terms are both \(O_p(N^{-1}T^{-1/2}) + O_p(T^{-1})\) by Lemma S.15(d) and \(\Delta \lambda_i = O_p(N^{-1}) + O_p(T^{-1/2})\). Using the results in Lemma S.14, the remaining terms except the first one are \(O_p(N^{-1}) + O_p(T^{-1})\). These results imply (d). \(\square\)

**Proof of Theorem 4**: Recall that the estimator \(\hat{\rho}_i\) is obtained by running the
By Lemma S.16(a),

\[ \hat{e}_{it} = \rho_{i,1} \hat{e}_{it-1} + \cdots + \rho_{i,p_i} \hat{e}_{it-p_i} + \text{error}, \quad \text{for } t = p_i + 1, \ldots, T \]

where \( \hat{e}_{it} = z_{it} - \hat{\lambda}_i \hat{f}_t \). So we have

\[ \hat{\rho}_i = \left( \sum_{t=p_i} T \hat{\psi}_{it} \hat{\psi}_{it}' \right)^{-1} \left( \sum_{t=p_i} T \hat{\psi}_{it} \hat{e}_{it} \right) \]

Then it follows

\[ \hat{\rho}_i - \rho_i = \left( \sum_{t=p_i} T \hat{\psi}_{it} \hat{\psi}_{it}' \right)^{-1} \left( \sum_{t=p_i} T \hat{\psi}_{it} (\hat{e}_{it} - \rho_{i,1} \hat{e}_{it-1} - \cdots - \rho_{i,p_i} \hat{e}_{it-p_i}) \right) \]

By (S.51) and \( \epsilon_{it} = \epsilon_{it} - \rho_{i,1} \epsilon_{it-1} - \cdots - \rho_{i,p_i} \epsilon_{it-p_i} \), we have

\[ \hat{e}_{it} - \rho_{i,1} \hat{e}_{it-1} - \cdots - \rho_{i,p_i} \hat{e}_{it-p_i} = \epsilon_{it} - \hat{\lambda}'_i \left[ \sum_{j=1}^{p_i} \rho_{i,j} \hat{f}_{t-j} \right] - \Delta \hat{\lambda}_i \left[ \hat{f}_t - \sum_{j=1}^{p_i} \rho_{i,j} \hat{f}_{t-j} \right] \]

So we have

\[ \hat{\rho}_i - \rho_i = \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \hat{\psi}_{it}' \right)^{-1} \left[ \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \epsilon_{it} \right) - \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \hat{f}_t \right) \lambda_i \right] \quad \text{(S.52)} \]

\[ - \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \hat{f}_t \right) \Delta \lambda_i + \sum_{j=1}^{p_i} \rho_{i,j} \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \hat{f}_{t-j} \right) \lambda_i + \sum_{j=1}^{p_i} \rho_{i,j} \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \hat{f}_{t-j}' \right) \Delta \lambda_i \]

By Lemma S.16(a),

\[ \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \epsilon_{it} = \frac{1}{T - p_i} \sum_{t=p_i} T \psi_{it} \epsilon_{it} + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \]

By Lemma S.16(b),

\[ - \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \hat{f}_t \right) \lambda_i + \sum_{j=1}^{p_i} \rho_{i,j} \left( \frac{1}{T - p_i} \sum_{t=p_i} T \hat{\psi}_{it} \hat{f}_{t-j} \right) \lambda_i = O_p(N^{-1}) + O_p(T^{-1}) \]
By Lemma S.16(c),

\[-\left(\frac{1}{T-p_i}\sum_{t=p_i}^{T} \hat{\psi}_{it}\hat{f}_{it}'\right)\Delta \lambda_i + \sum_{j=1}^{p_i} \rho_{i,j} \left(\frac{1}{T-p_i}\sum_{t=p_i}^{T} \hat{\psi}_{it}\hat{f}_{t-j}'\right)\Delta \lambda_i\]

\[= \left[O_p(N^{-1}) + O_p(T^{-1/2})\right] \left[O_p(N^{-1}) + O_p(T^{-1/2})\right] = O_p(N^{-2}) + O_p(T^{-1})\]

By Lemma S.16(d),

\[\frac{1}{T-p_i}\sum_{t=p_i}^{T} \hat{\psi}_{it}\hat{f}_{it}' = \frac{1}{T-p_i}\sum_{t=p_i}^{T} \psi_{it}\psi_{it}' + O_p(N^{-1}) + O_p(T^{-1})\]

Then it follows

\[\hat{\rho}_i - \rho_i = \left(\frac{1}{T-p_i}\sum_{t=p_i}^{T} \psi_{it}\psi_{it}'\right)^{-1} \left(\frac{1}{T-p_i}\sum_{t=p_i}^{T} \psi_{it}\epsilon_{it}\right) + O_p(N^{-1}) + O_p(T^{-1}) \quad (S.53)\]

Given the above results, we have, under the condition \(\sqrt{T}/N \to 0\),

\[\sqrt{T-p_i}(\hat{\rho}_i - \rho_i) = \left(\frac{1}{T-p_i}\sum_{t=p_i}^{T} \psi_{it}\psi_{it}'\right)^{-1} \left(\frac{1}{\sqrt{T-p_i}}\sum_{t=p_i}^{T} \psi_{it}\epsilon_{it}\right) + o_p(1) \quad (S.54)\]

By the martingale difference central limiting theorem,

\[\sqrt{T-p_i}(\hat{\rho}_i - \rho_i) \overset{d}{\to} N\left(0, \sigma_{\epsilon_i}^2 \left[\text{plim}_{T \to \infty} \frac{1}{T-p_i}\sum_{t=p_i}^{T} \psi_{it}\psi_{it}'\right]^{-1}\right)\]

This completes the proof of the \(\hat{\rho}_i\) part of Theorem 4. □

The following lemma is useful to derive the asymptotic representation of \(\tilde{\lambda}_i - \lambda_i\).

**Lemma S.17** Under Assumptions 5A-5D,

(a) \[\frac{1}{T-p_i}\sum_{t=p_i}^{T} \hat{g}_{it} \Delta f_{t-q} = O_p(N^{-1}) + O_p(T^{-1}), \text{ for } q = 0, 1, \ldots, p_i\]

(b) \[\frac{1}{T-p_i}\sum_{t=p_i}^{T} \hat{g}_{it} \epsilon_{it-p} = O_p(N^{-1}) + O_p(T^{-1/2}), \text{ for } p = 1, \ldots, p_i\]
(c) \[ \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T \hat{g}_{it} \epsilon_{it} = \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T g_{it} \epsilon_{it} + O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) \]

(d) \[ \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T \hat{g}_{it} \hat{g}_i' = \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T g_{it} \hat{g}_i' + O_p(N^{-1}) + O_p(T^{-1/2}) \]

**Proof of Lemma S.17:** Consider (a). By \( \hat{\rho}_{i,j} \hat{f}_{t-j} = \hat{\rho}_{i,j} \hat{f}_{t-j} + \rho_{i,j} f_{t-j} \), we have

\[ \hat{g}_{it} = g_{it} - \sum_{j=1}^{p_i} \Delta \hat{\rho}_{i,j} f_{t-j} - \Delta \hat{f}_{t} - \sum_{j=1}^{p_i} \hat{\rho}_{i,j} \Delta \hat{f}_{t-j} \]  \hspace{1cm} (S.55)

Thus, the left hand side of (a) is equal to

\[ \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T g_{it} \Delta \hat{f}_{t-q} - \sum_{p=1}^{p_i} \hat{\Delta \rho}_{i,p} T - p_i \sum_{t=\bar{p}_i}^T f_{t-p} \Delta \hat{f}_{t-q} \]

\[ - \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T \Delta \hat{f}_{t} \Delta \hat{f}_{t-q} - \sum_{p=1}^{p_i} \hat{\rho}_{i,j} \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T \Delta \hat{f}_{t-p} \Delta \hat{f}_{t-q} \]

The first and second terms are both \( O_p(N^{-1}) + O_p(T^{-1}) \) by the definition of \( g_{it} \), \( \hat{\rho}_{i,j} \rightarrow \rho_{i,j} \pandas 0 \) and Lemma S.14(b). The third and fourth terms are also \( O_p(N^{-1}) + O_p(T^{-1}) \) by \( \hat{\rho}_{i,j} \rightarrow \rho_{i,j} \pandas 0 \) and Lemma S.14(a). This proves (a).

Consider (b). The left hand side of (b) is equal to

\[ \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T g_{it} \epsilon_{it-p} - \sum_{q=1}^{p_i} \Delta \hat{\rho}_{i,q} T - p_i \sum_{t=\bar{p}_i}^T f_{t-q} \epsilon_{it-p} \]

\[ - \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T \Delta \hat{f}_{t} \epsilon_{it-p} - \sum_{q=1}^{p_i} \hat{\rho}_{i,q} \frac{1}{T-p_i} \sum_{t=\bar{p}_i}^T \Delta \hat{f}_{t-q} \epsilon_{it-p} \]

The first term is \( O_p(T^{-1/2}) \). The second term is \( O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) \) by \( \Delta \hat{\rho}_{i,q} = O_p(N^{-1}) + O_p(T^{-1/2}) \). The third and fourth terms are both \( O_p(N^{-1}) + O_p(T^{-1}) \) by Lemma S.15(c) and \( \hat{\rho}_{i,q} \pandas \rho_{i,q} \). This proves (b).
Consider (c). The left hand side of (c) is equal to

\[
\frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} \epsilon_{it} - \sum_{q=1}^{p_i} \Delta \hat{\rho}_{i,q} \frac{1}{T-p_i} \sum_{t=p_i}^{T} f_{t-q} \epsilon_{it} - \frac{1}{T-p_i} \sum_{t=p_i}^{T} \Delta \hat{f}_{t} \epsilon_{it} - \sum_{q=1}^{p_i} \hat{\rho}_{i,q} \frac{1}{T-p_i} \sum_{t=p_i}^{T} \Delta \hat{f}_{t-q} \epsilon_{it}
\]

The second term is \(O_p(N^{-1}T^{-1/2}) + O_p(T^{-1})\). The third and fourth terms are both \(O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})\) by Lemma S.15(a) and \(\hat{\rho}_{i,q} \rightarrow \rho_{i,q}\). Thus (c) follows.

Consider (d). Let \(\hat{\rho}_{i,0} \equiv 1\). Then equation (S.55) can be written as

\[
\hat{g}_{it} = g_{it} - \sum_{j=1}^{p_i} \Delta \hat{\rho}_{i,j} \hat{f}_{t-j} - \sum_{j=0}^{p_i} \hat{\rho}_{i,j} \Delta \hat{f}_{t-j}
\]

The left hand side of (d) can be written as

\[
\frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} g_{it}' - \sum_{p=1}^{p_i} \Delta \hat{\rho}_{i,p} \left( \frac{1}{T-p_i} \sum_{t=p_i}^{T} f_{t-p} g_{it}' \right) - \sum_{p=0}^{p_i} \hat{\rho}_{i,p} \left( \frac{1}{T-p_i} \sum_{t=p_i}^{T} \Delta \hat{f}_{t-p} g_{it}' \right)
\]

\[
- \sum_{q=1}^{p_i} \Delta \hat{\rho}_{i,q} \frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} f_{t-q}' + \sum_{p=1}^{p_i} \sum_{q=1}^{p_i} \Delta \hat{\rho}_{i,p} \Delta \hat{\rho}_{i,q} \frac{1}{T-p_i} \sum_{t=p_i}^{T} f_{t-p} f_{t-q}'
\]

\[
+ \sum_{p=0}^{p_i} \sum_{q=1}^{p_i} \hat{\rho}_{i,p} \Delta \hat{\rho}_{i,q} \frac{1}{T-p_i} \sum_{t=p_i}^{T} \Delta \hat{f}_{t-p} f_{t-q}'
\]

\[
- \sum_{q=0}^{p_i} \hat{\rho}_{i,q} \left( \frac{1}{T-p_i} \sum_{t=p_i}^{T} f_{t-p} \Delta \hat{f}_{t-q}' \right) + \sum_{p=0}^{p_i} \sum_{q=0}^{p_i} \hat{\rho}_{i,p} \hat{\rho}_{i,q} \left( \frac{1}{T-p_i} \sum_{t=p_i}^{T} \Delta \hat{f}_{t-p} \Delta \hat{f}_{t-q}' \right)
\]

The second and fourth terms are both \(O_p(N^{-1}) + O_p(T^{-1/2})\) by \(\Delta \hat{\rho}_{i,p} = O_p(N^{-1}) + O_p(T^{-1/2})\) and \(\frac{1}{T-p_i} \sum_{t=p_i}^{T} f_{t-p} g_{it}' = O_p(1)\). The third and seventh terms are both \(O_p(N^{-1}) + O_p(T^{-1})\) by Lemma S.14(b) and \(\hat{\rho}_{i,j} \rightarrow \rho_{i,j}\). The sixth and eighth terms are both \(O_p(N^{-2}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2})\) by Lemma S.14(b) and \(\Delta \hat{\rho}_{i,p} = O_p(N^{-1}) + O_p(T^{-1/2})\). The fifth term is \(O_p(N^{-2}) + O_p(T^{-1})\) by \(\Delta \hat{\rho}_{i,p} = O_p(N^{-1}) + O_p(T^{-1/2})\) and \(\frac{1}{T-p_i} \sum_{t=p_i}^{T} f_{t-p} f_{t-q}' = O_p(1)\). The last term is \(O_p(N^{-1}) + O_p(T^{-1})\) by Lemma

57
By Lemma S.17(a), $\hat{\rho}_{i,j} \xrightarrow{p} \rho_{i,j}$. Summarizing all the results, we have (d). □

**Proof of Theorem 4 (continued):** Recall that the estimator $\tilde{\lambda}_i$ is obtained by running the regression

$$z_{it} - \hat{\rho}_{i,1}z_{it-1} - \cdots - \hat{\rho}_{i,p_i}z_{it-p_i} = (\hat{f}_t - \hat{\rho}_{i,1}\hat{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\hat{f}_{t-p_i})'\lambda_i + \text{error}, \quad \text{for } t = p_i+1, \ldots, T$$

Notice that $\hat{g}_{it} = \hat{f}_t - \hat{\rho}_{i,1}\hat{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\hat{f}_{t-p_i}$, so we have

$$\tilde{\lambda}_i = \left(\sum_{t=p_i}^{T} \hat{g}_{it}\hat{g}_{it}'\right)^{-1} \left(\sum_{t=p_i}^{T} \hat{g}_{it}(z_{it} - \hat{\rho}_{i,1}z_{it-1} - \cdots - \hat{\rho}_{i,p_i}z_{it-p_i})\right)$$

Rewrite $\tilde{\lambda}_i$ as

$$\tilde{\lambda}_i - \lambda_i = \left(\sum_{t=p_i}^{T} \hat{g}_{it}\hat{g}_{it}'\right)^{-1} \left(\sum_{t=p_i}^{T} \hat{g}_{it}(z_{it} - \hat{\rho}_{i,1}z_{it-1} - \cdots - \hat{\rho}_{i,p_i}z_{it-p_i} - \hat{g}_{it}'\lambda_i)\right)$$

From $z_{it} = \lambda'_i f_t + e_{it}$ and the definition of $\hat{g}_{it}$, we have

$$\tilde{\lambda}_i - \lambda_i = \left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}\hat{g}_{it}'\right)^{-1} \left[\left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}\Delta f_t'\right) - \sum_{j=1}^{p_i} \hat{\rho}_{i,j} \left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}\Delta f_{t-j}'\right)\right] \lambda_i$$

$$+ \left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}\hat{g}_{it}'\right)^{-1} \left[\left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}e_{it}\right) - \sum_{j=1}^{p_i} \Delta \rho_{i,j} \left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}e_{it-j}\right)\right] \tag{S.56}$$

By Lemma S.17(a),

$$\left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}\Delta f_t'\right) - \sum_{j=1}^{p_i} \hat{\rho}_{i,j} \left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}\Delta f_{t-j}'\right) = O_p(N^{-1}) + O_p(T^{-1}).$$

By Lemma S.17(b) and (S.53)

$$\sum_{j=1}^{p_i} \Delta \rho_{i,j} \left(\frac{1}{T - p_i} \sum_{t=p_i}^{T} \hat{g}_{it}e_{it-j}\right) = O_p(N^{-2}) + O_p(T^{-1}).$$
By Lemma S.17(c),
\[
\frac{1}{T-p_i} \sum_{t=p_i}^{T} \hat{g}_{it} \epsilon_{it} = \frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} \epsilon_{it} + O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}).
\]

By Lemma S.17(d),
\[
\frac{1}{T-p_i} \sum_{t=p_i}^{T} \hat{g}_{it} \hat{g}'_{it} = \frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} \hat{g}'_{it} + O_p(N^{-1}) + O_p(T^{-1/2}).
\]

Then it follows
\[
\tilde{\lambda}_i - \lambda_i = \left( \frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} \hat{g}'_{it} \right)^{-1} \left( \frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} \epsilon_{it} \right) + O_p(N^{-1}) + O_p(T^{-1}) \quad (S.57)
\]

Given the above results, we have, under the condition $\sqrt{T}/N \to 0$,
\[
\sqrt{T-p_i}(\tilde{\lambda}_i - \lambda_i) = \left( \frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} \hat{g}'_{it} \right)^{-1} \left( \frac{1}{\sqrt{T-p_i}} \sum_{t=p_i}^{T} g_{it} \epsilon_{it} \right) + o_p(1) \quad (S.58)
\]

By the cental limiting theorem,
\[
\sqrt{T-p_i}(\tilde{\lambda}_i - \lambda_i) \overset{d}{\to} N \left( 0, \sigma_{\epsilon}^2 \left( \lim_{T \to \infty} \frac{1}{T-p_i} \sum_{t=p_i}^{T} g_{it} \epsilon_{it} \right)^{-1} \right)
\]

We proceed to consider the limiting results on $\tilde{f}_t$. Recall that $\tilde{f}_t$ is obtained by the regression
\[
\frac{1}{\phi_i} z_{it} = \frac{1}{\phi_i} \tilde{\lambda}_i \tilde{f}_t + \text{error}
\]

So we have
\[
\tilde{f}_t = \left( \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i \tilde{\lambda}'_i \right)^{-1} \left( \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i z_{it} \right) = \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i \tilde{\lambda}'_i \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i z_{it} \right)
\]
By $z_{it} = \lambda_i f_t + e_{it}$, we have

$$\tilde{f}_t - f_t = -\left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i \tilde{\lambda}_i' \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i (\tilde{\lambda}_i - \lambda_i)' \right) f_t + \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i \tilde{\lambda}_i' \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i e_{it} \right)$$

(S.59)

Given (S.57), together with the boundedness of $\phi_i^2$, it follows

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \|\tilde{\lambda}_i - \lambda_i\|^2 = O_p(N^{-2}) + O_p(T^{-1})$$

(S.60)

Consider the expression $\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i (\tilde{\lambda}_i - \lambda_i)'$. By (S.57), the expression is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} e_{it} \right) \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} g_{it}' \right)^{-1} + O_p(N^{-1}) + O_p(T^{-1})$$

The first term of the above expression is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} (\tilde{\lambda}_i - \lambda_i) \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} e_{it} \right) \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} g_{it}' \right)^{-1}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \frac{1}{\phi_i^2} \lambda_i \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} e_{it} \right) \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} g_{it}' \right)^{-1}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} e_{it} \right) \left( \frac{1}{T - p_i} \sum_{t=p_i}^{T} g_{it} g_{it}' \right)^{-1}$$

The first two terms can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ in the same way with Lemma S.11(b). The last term is $O_p(N^{-1/2}T^{-1/2})$. So $\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i (\tilde{\lambda}_i - \lambda_i)' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. By the similar argument, we have

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \tilde{\lambda}_i e_{it} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it} + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$
Given this result, notice \( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \chi'_i = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \chi'_i + o_p(1) \), we have

\[
\hat{f}_t - f_t = \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \chi'_i \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it} \right) + O_p(N^{-1}) + O_p(T^{-1})
\]

Then it follows that under \( \sqrt{N}/T \to 0 \),

\[
\sqrt{N}(\hat{f}_t - f_t) = \left( \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i \chi'_i \right)^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\phi_i^2} \lambda_i e_{it} \right) + o_p(1)
\]

This completes the proof of Theorem 4. \( \square \)

**Supplement F: EM algorithm on joint estimation when the dynamics of \( e_{it} \) and \( f_t \) exist**

In this section, we give the updating formulas of the EM algorithm for the simple case considered in Section 5.2. The complete-data likelihood function is

\[
\ln L(\theta) = C - \frac{1}{2N} \sum_{i=1}^{N} \ln \sigma_{\epsilon i}^2 - \frac{1}{2NT} \sum_{i=1}^{N} \frac{1}{\sigma_{\epsilon i}^2} \sum_{t=2}^{T} \left( z_{it} - \rho_i z_{it-1} - \lambda'_i f_t + \rho_i \lambda'_i f_{t-1} \right)^2
\]

Here the marginal likelihood for \( f_t \) (to estimate \( \Psi \)) is omitted for simplicity. The expected complete-data likelihood, conditional on the data and \( \theta^* \), is

\[
Q(\theta|\theta^*) = C - \frac{1}{2N} \sum_{i=1}^{N} \ln \sigma_{\epsilon i}^2 - \frac{1}{2N} \sum_{i=1}^{N} \frac{1}{\sigma_{\epsilon i}^2} \sum_{t=2}^{T} \left( z_{it} - \rho_i z_{it-1} \right)^2
-2(z_{it} - \rho_i z_{it-1})\lambda'_i E(f_t|\theta^*) + 2(z_{it} - \rho_i z_{it-1})\rho_i \lambda'_i E(f_{t-1}|\theta^*)
+\lambda'_i E(f_t f'_t|\theta^*) \lambda_i + \rho_i^2 \lambda'_i E(f_{t-1} f'_{t-1}|\theta^*) \lambda_i - 2\rho_i \lambda'_i E(f_{t-1} f'_{t-1}|\theta^*) \lambda_i
\]

(S.61)

where we omit the data matrix \( Z \) from the conditional expectations so that \( E(f_t|\theta^*) \) denotes \( E(f_t|Z, \theta^*) \), etc. Define \( V_{00,t} = E(f_t f'_t|\theta^*, Z) \), \( V_{01,t} = E(f_t f'_{t-1}|\theta^*, Z) \), \( V_{11,t} = \).
Putting together, we obtain
\( E(f_{t-1}f'_{t-1}|\theta^*, Z) \). In the E-step, we compute these conditional expectations at \( \theta^* = \theta^{(k)} \), where \( \theta^{(k)} \) denotes the \( k \)th iteration of \( \theta \) in the ECM algorithm. These conditional expectations are computed via the Kalman smoothers in view that system (13) is a standard state space model with the first equation being the measurement equation and the second being the transition equation. In the constrained M-step, we take derivatives with respect to \( \theta \) in (S.61). By dividing \( \theta \) into four subgroups, the ECM of Meng and Rubin (1993) leads to the following updating formulae:

\[
\lambda_i^{(k+1)} = \left[ \sum_{t=2}^{T} \left( V_{00,t} - \rho_i^{(k)} V_{01,t} - \rho_i^{(k)} V'_{01,t} + (\rho_i^{(k)})^2 V_{11,t} \right) \right]^{-1} \\
\times \left[ \sum_{t=2}^{T} \left( E(f_t|\theta^{(k)}) - \rho_i E(f_{t-1}|\theta^{(k)}) \right) \left( z_{it} - \rho_i^{(k)} z_{i(t-1)} \right) \right],
\]

\[
\rho_i^{(k+1)} = \left[ \sum_{t=2}^{T} \left( z_{it-1}^2 - 2z_{it-1}\lambda_i^{(k+1)'E(f_{t-1}|\theta^{(k)})} + \lambda_i^{(k+1)'V_{11,t}\lambda_i^{(k+1)}} \right) \right]^{-1} \\
\times \left[ \sum_{t=2}^{T} \left( z_{it}^2 - z_{it}\lambda_i^{(k+1)'E(f_{t-1}|\theta^{(k)})} - z_{it-1}\lambda_i^{(k+1)'E(f_{t-1}|\theta^{(k)})} + \lambda_i^{(k+1)'V_{11,t}\lambda_i^{(k+1)}} \right) \right],
\]

\[
(\sigma_{ei}^{(k+1)})^2 = \frac{1}{T-1} \sum_{t=2}^{T} \left( (z_{it} - \rho_i^{(k+1)} z_{i(t-1)})^2 - 2(z_{it} - \rho_i^{(k+1)} z_{i(t-1)})\lambda_i^{(k+1)'E(f_{t-1}|\theta^{(k)})} \right. \\
+ 2\rho_i^{(k+1)}(z_{it} - \rho_i^{(k+1)} z_{i(t-1)})\lambda_i^{(k+1)'E(f_{t-1}|\theta^{(k)})} + \lambda_i^{(k+1)'V_{00,t}\lambda_i^{(k+1)}}) \\
\left. - 2\rho_i^{(k+1)}\lambda_i^{(k+1)'V_{10,t}\lambda_i^{(k+1)}} + (\rho_i^{(k+1)})^2\lambda_i^{(k+1)'V_{00,t}\lambda_i^{(k+1)}} \right),
\]

\[
\Psi^{(k+1)} = \left( \sum_{t=2}^{T} V_{01,t} \right) \left( \sum_{t=2}^{T} V_{11,t} \right)^{-1}.
\]

The last expression \( \Psi^{(k+1)} \) is obtained from the (omitted) marginal likelihood for \( f_t \). Putting together, we obtain \( \theta^{(k+1)} \). The iteration continues until convergence. For the initial values, we can use the two-step estimators stated in Section 4.1.
Supplement G: Finite sample properties

This section uses Monte Carlo simulations to evaluate the finite sample properties of QMLE, ML-GLS, iterated ML-GLS (denoted by ML-ITE below) and ML-EM estimators (all discussed in Section 5). The data are generated according to

\[ z_{it} = \lambda_i' f_t + e_{it} \]

where \( A(L)f_t = u_t \) with \( u_t \) being i.i.d. \( N(0, I_r) \) and \( D(L)e_t = \epsilon_t \) with \( \epsilon_t \) being i.i.d. \( N(0, \mathcal{T}) \); \( A(L) \) and \( D(L) \) are defined as \( A(L) = I_r - \psi L, D(L) = I_N - \rho L \), where \( \rho = \text{diag}(\rho_1, \ldots, \rho_N) \) and \( \psi \) is a scalar. Matrix \( \mathcal{T} \) is \( N \times N \) with its \((i, j)\)th element

\[ \tau_{ij} = \phi_i^2 \phi_j^2 (1 - \rho_i^2)(1 - \rho_j^2) \]

The variance of \( e_{it}, \phi_i^2 \), is generated according to

\[ \phi_i^2 = \frac{\beta_i}{1 - \beta_i} \frac{1}{1 - \psi^2} \lambda_i' \lambda_i \]  

(S.62)

where \( \beta_i \) are iid \( U[u, 1 - u] \) with \( u \in [0, 0.5] \). All the elements of \( \Lambda \) are iid \( N(0, 1) \). The number of factors is \( r = 2 \) (assumed known). The data generating process is similar to those of Breitung and Tenhofen (2011) and Doz et al. (2011a).

In this DGP, \( \beta_i \) is the ratio between the variance of \( e_{it} \) and the variance of \( z_{it} \). Since \( \beta_i \) is from \( U[u, 1 - u] \), the parameter \( u \) has a close relation with the heteroscedasticity over the cross section. A small \( u \) tends to give more heteroscedasticities. The value \( \tau \) is the correlation between two adjacent units of the cross section. It thus controls the cross section correlations. This correlation decreases exponentially as the distance of two units increases. So the limited cross-sectional correlation required in Assumption C is satisfied. The parameters \( \rho \) and \( \psi \) are used to control the autocorrelations of the idiosyncratic errors and the factors. To evaluate the effect of autocorrelation of \( e_{it} \) on the estimation, we generate \( \rho_i \) from \( U[0, 0.9] \) and \( U[0.5, 0.9] \).

As a measure of goodness-of-fit, we use the Trace-Ratio (TR) to evaluate how
close the estimated values $\Lambda$ and $F$ to their true values. Taking $F$ as an example, the TR is defined as $TR(F) = \text{tr}[(F'\hat{F})(\hat{F}'\hat{F})^{-1}(\hat{F}'F)]/\text{tr}[F'F]$. The measure is a generalized squared correlation coefficient in multivariate analysis and is invariant to the rotation.

For comparison, we also compute the PC estimators, PC-GLS estimators and iterative PC-GLS estimators (denoted by PC-ITE below). These estimators are discussed in Section 5. Of these seven estimators, PC, PC-GLS and PC-ITE belong to the PC class, while QMLE, ML-GLS, ML-ITE and ML-EM belong to the ML class. Reported results are based on 1000 repetitions.

Tables G1-G2 report the trace ratios for the seven estimators under the setting $u = 0.1, \psi = 0, \tau = 0$ and $\rho_i \sim U[0, 0.9]$. The estimators in the ML class outperform the counterpart in the PC class. Consider the estimation of $\Lambda$. In the PC class, the best estimator is that of PC-ITE. However, when $N$ is small such as $N = 10$ or $20$, its performance, which is expected to be superior to QMLE because it takes into account of serial correlation of $e_{it}$, is still dominated by QMLE. The reason is due to the imprecise estimation of the error term by the PC method. So the gain from estimating the serial correlations in the next step is limited. However, if the first step is conducted by the ML method, the performance is substantially improved, which is reflected in the ML-GLS column. As for the estimation of $F$, the advantage of the ML-class of estimators over those in the PC class is even more pronounced. Even the QMLE can perform better than PC-ITE. The former ignore the serial correlations in $e_{it}$, while the latter estimates serial correlation in $e_{it}$. This is especially true for small or moderate $N$ ($N \leq 50$). Of the seven estimators, ML-EM performs the best in all combinations of $N$ and $T$. This is due to the benefit of the simultaneous estimation of all parameters. All estimators, except for PC, perform comparably under large $N$.

\textsuperscript{1}To calculate PC-ITE and ML-ITE, we limit the number of iterations to 5. As pointed out by Breitung and Tenhofen and also confirmed in our simulation, increasing the number of iterations does not noticeably improve the performance.
(say, \(N = 150, T = 100\)). This is consistent with the theory.

Tables G3-G4 report the trace ratios when there exist cross-sectional correlations in \(e_{it}\) and autocorrelations in \(f_t\). In this setting, all the seven estimators have misspecification problem because they do not take into consideration of the cross-sectional correlations in \(e_{it}\). The performance of all estimators deteriorates to some extent. For example, when \(N = 10, T = 30\), the TR values of the QMLE in Table 1 are 0.916 for \(\Lambda\) and 0.819 for \(F\). In contrast, the counterparts in Tables G3 and G4 are 0.783 for \(\Lambda\) and 0.681 for \(F\). However, when the sample size becomes large, the performance of the estimators improves substantially. When \(N = 150, T = 100\), the TR values of the QMLE in tables G3-G4 are 0.944 for \(\Lambda\) and 0.991 for \(F\). This result confirms the theory that the QMLE are robust under misspecification. Also, the estimators in the ML class still outperform those in the PC class, especially when the sample size is small or moderate.

Tables G5-G8 report the simulation results under \(\rho_i \sim U[0.5, 0.9]\). Overall, the simulations show the same result that the ML-type estimators outperform the PC-type estimators in all the sample sizes.
Table G1: The Trace Ratio of the seven estimators for estimating $\Lambda$

with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0, 0.9]$

<table>
<thead>
<tr>
<th></th>
<th>PC Class</th>
<th></th>
<th>ML Class</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PC</td>
<td>PC-GLS</td>
<td>PC-ITE</td>
</tr>
<tr>
<td>$N$</td>
<td>$T$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>0.847</td>
<td>0.874</td>
<td>0.893</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.865</td>
<td>0.896</td>
<td>0.913</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.890</td>
<td>0.919</td>
<td>0.932</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.760</td>
<td>0.801</td>
<td>0.883</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.803</td>
<td>0.845</td>
<td>0.922</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.849</td>
<td>0.887</td>
<td>0.949</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.753</td>
<td>0.804</td>
<td>0.908</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.816</td>
<td>0.866</td>
<td>0.951</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.878</td>
<td>0.918</td>
<td>0.974</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.798</td>
<td>0.856</td>
<td>0.922</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.877</td>
<td>0.922</td>
<td>0.955</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.932</td>
<td>0.960</td>
<td>0.978</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.824</td>
<td>0.883</td>
<td>0.924</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.898</td>
<td>0.939</td>
<td>0.956</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.948</td>
<td>0.970</td>
<td>0.978</td>
</tr>
</tbody>
</table>
Table G2: The Trace Ratio of the seven estimators for estimating $F$
with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0, 0.9]$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>PC</th>
<th>PC-GLS</th>
<th>PC-ITE</th>
<th>QMLE</th>
<th>ML-GLS</th>
<th>ML-ITE</th>
<th>ML-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>0.655</td>
<td>0.711</td>
<td>0.702</td>
<td>0.819</td>
<td>0.825</td>
<td>0.828</td>
<td>0.836</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.640</td>
<td>0.703</td>
<td>0.699</td>
<td>0.837</td>
<td>0.843</td>
<td>0.845</td>
<td>0.871</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.648</td>
<td>0.717</td>
<td>0.718</td>
<td>0.856</td>
<td>0.860</td>
<td>0.860</td>
<td>0.888</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.633</td>
<td>0.754</td>
<td>0.840</td>
<td>0.909</td>
<td>0.915</td>
<td>0.917</td>
<td>0.926</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.646</td>
<td>0.780</td>
<td>0.863</td>
<td>0.922</td>
<td>0.926</td>
<td>0.927</td>
<td>0.939</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.662</td>
<td>0.808</td>
<td>0.879</td>
<td>0.931</td>
<td>0.933</td>
<td>0.933</td>
<td>0.946</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.715</td>
<td>0.881</td>
<td>0.951</td>
<td>0.966</td>
<td>0.970</td>
<td>0.970</td>
<td>0.974</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.743</td>
<td>0.915</td>
<td>0.964</td>
<td>0.971</td>
<td>0.973</td>
<td>0.973</td>
<td>0.978</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.781</td>
<td>0.943</td>
<td>0.969</td>
<td>0.974</td>
<td>0.975</td>
<td>0.975</td>
<td>0.980</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.820</td>
<td>0.951</td>
<td>0.983</td>
<td>0.984</td>
<td>0.986</td>
<td>0.986</td>
<td>0.988</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.866</td>
<td>0.978</td>
<td>0.986</td>
<td>0.986</td>
<td>0.987</td>
<td>0.987</td>
<td>0.989</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.892</td>
<td>0.985</td>
<td>0.988</td>
<td>0.987</td>
<td>0.988</td>
<td>0.988</td>
<td>0.990</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.871</td>
<td>0.974</td>
<td>0.991</td>
<td>0.989</td>
<td>0.991</td>
<td>0.991</td>
<td>0.992</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.914</td>
<td>0.989</td>
<td>0.992</td>
<td>0.991</td>
<td>0.992</td>
<td>0.992</td>
<td>0.993</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.933</td>
<td>0.991</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
<td>0.994</td>
</tr>
</tbody>
</table>
Table G3: The Trace Ratio of the seven estimators for estimating $\Lambda$

with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0, 0.9]$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>PC</th>
<th>PC-GLS</th>
<th>PC-ITE</th>
<th>QMLE</th>
<th>ML-GLS</th>
<th>ML-ITE</th>
<th>ML-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>0.738</td>
<td>0.749</td>
<td>0.764</td>
<td>0.783</td>
<td>0.801</td>
<td>0.808</td>
<td>0.816</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.749</td>
<td>0.757</td>
<td>0.764</td>
<td>0.804</td>
<td>0.816</td>
<td>0.819</td>
<td>0.829</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.762</td>
<td>0.771</td>
<td>0.781</td>
<td>0.828</td>
<td>0.841</td>
<td>0.846</td>
<td>0.853</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.672</td>
<td>0.690</td>
<td>0.756</td>
<td>0.792</td>
<td>0.828</td>
<td>0.837</td>
<td>0.848</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.702</td>
<td>0.719</td>
<td>0.796</td>
<td>0.849</td>
<td>0.883</td>
<td>0.890</td>
<td>0.903</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.736</td>
<td>0.749</td>
<td>0.827</td>
<td>0.898</td>
<td>0.919</td>
<td>0.923</td>
<td>0.936</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.670</td>
<td>0.706</td>
<td>0.856</td>
<td>0.829</td>
<td>0.881</td>
<td>0.886</td>
<td>0.890</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.747</td>
<td>0.783</td>
<td>0.913</td>
<td>0.890</td>
<td>0.929</td>
<td>0.931</td>
<td>0.933</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.813</td>
<td>0.838</td>
<td>0.952</td>
<td>0.940</td>
<td>0.964</td>
<td>0.964</td>
<td>0.966</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.728</td>
<td>0.783</td>
<td>0.887</td>
<td>0.837</td>
<td>0.888</td>
<td>0.891</td>
<td>0.893</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.814</td>
<td>0.862</td>
<td>0.934</td>
<td>0.895</td>
<td>0.934</td>
<td>0.935</td>
<td>0.936</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.888</td>
<td>0.920</td>
<td>0.966</td>
<td>0.943</td>
<td>0.967</td>
<td>0.967</td>
<td>0.968</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.754</td>
<td>0.814</td>
<td>0.889</td>
<td>0.835</td>
<td>0.887</td>
<td>0.890</td>
<td>0.891</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.848</td>
<td>0.899</td>
<td>0.937</td>
<td>0.896</td>
<td>0.936</td>
<td>0.937</td>
<td>0.937</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.915</td>
<td>0.945</td>
<td>0.968</td>
<td>0.944</td>
<td>0.967</td>
<td>0.968</td>
<td>0.968</td>
</tr>
</tbody>
</table>
Table G4: The Trace Ratio of the seven estimators for estimating $F$ with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0, 0.9]$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>PC</th>
<th>PC-GLS</th>
<th>PC-ITE</th>
<th>QMLE</th>
<th>ML-GLS</th>
<th>ML-ITE</th>
<th>ML-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>0.587</td>
<td>0.615</td>
<td>0.599</td>
<td>0.681</td>
<td>0.681</td>
<td>0.679</td>
<td>0.694</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.562</td>
<td>0.587</td>
<td>0.562</td>
<td>0.662</td>
<td>0.662</td>
<td>0.658</td>
<td>0.686</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.550</td>
<td>0.578</td>
<td>0.561</td>
<td>0.669</td>
<td>0.669</td>
<td>0.668</td>
<td>0.698</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.584</td>
<td>0.653</td>
<td>0.705</td>
<td>0.810</td>
<td>0.814</td>
<td>0.814</td>
<td>0.843</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.581</td>
<td>0.663</td>
<td>0.730</td>
<td>0.839</td>
<td>0.842</td>
<td>0.841</td>
<td>0.873</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.578</td>
<td>0.666</td>
<td>0.733</td>
<td>0.854</td>
<td>0.856</td>
<td>0.854</td>
<td>0.887</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.669</td>
<td>0.805</td>
<td>0.924</td>
<td>0.950</td>
<td>0.955</td>
<td>0.957</td>
<td>0.963</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.709</td>
<td>0.857</td>
<td>0.946</td>
<td>0.960</td>
<td>0.963</td>
<td>0.963</td>
<td>0.971</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.732</td>
<td>0.890</td>
<td>0.955</td>
<td>0.966</td>
<td>0.967</td>
<td>0.967</td>
<td>0.974</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.788</td>
<td>0.914</td>
<td>0.977</td>
<td>0.978</td>
<td>0.981</td>
<td>0.982</td>
<td>0.985</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.834</td>
<td>0.957</td>
<td>0.984</td>
<td>0.983</td>
<td>0.985</td>
<td>0.985</td>
<td>0.987</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.868</td>
<td>0.975</td>
<td>0.985</td>
<td>0.985</td>
<td>0.986</td>
<td>0.986</td>
<td>0.989</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.844</td>
<td>0.953</td>
<td>0.988</td>
<td>0.986</td>
<td>0.988</td>
<td>0.989</td>
<td>0.990</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.896</td>
<td>0.983</td>
<td>0.990</td>
<td>0.989</td>
<td>0.990</td>
<td>0.990</td>
<td>0.992</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.920</td>
<td>0.988</td>
<td>0.991</td>
<td>0.991</td>
<td>0.991</td>
<td>0.991</td>
<td>0.993</td>
</tr>
</tbody>
</table>
Table G5: The Trace Ratio of the seven estimators for estimating $\Lambda$. with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0.5, 0.9]$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>PC Class</th>
<th>ML Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PC</td>
<td>PC-GLS</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>0.811</td>
<td>0.872</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.843</td>
<td>0.907</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.866</td>
<td>0.931</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.722</td>
<td>0.808</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.771</td>
<td>0.860</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.806</td>
<td>0.895</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.680</td>
<td>0.791</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.769</td>
<td>0.877</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.858</td>
<td>0.942</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.703</td>
<td>0.826</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.817</td>
<td>0.926</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.915</td>
<td>0.975</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.712</td>
<td>0.842</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.844</td>
<td>0.945</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.930</td>
<td>0.982</td>
</tr>
</tbody>
</table>
Table G6: The Trace Ratio of the seven estimators for estimating $F$.

with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0.5, 0.9]$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>PC</th>
<th>PC-GLS</th>
<th>PC-ITE</th>
<th>QMLE</th>
<th>ML-GLS</th>
<th>ML-ITE</th>
<th>ML-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>0.615</td>
<td>0.675</td>
<td>0.691</td>
<td>0.760</td>
<td>0.774</td>
<td>0.785</td>
<td>0.805</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.617</td>
<td>0.689</td>
<td>0.705</td>
<td>0.800</td>
<td>0.817</td>
<td>0.824</td>
<td>0.859</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.616</td>
<td>0.692</td>
<td>0.705</td>
<td>0.831</td>
<td>0.843</td>
<td>0.847</td>
<td>0.881</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.599</td>
<td>0.727</td>
<td>0.839</td>
<td>0.859</td>
<td>0.877</td>
<td>0.889</td>
<td>0.907</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.611</td>
<td>0.758</td>
<td>0.866</td>
<td>0.904</td>
<td>0.916</td>
<td>0.919</td>
<td>0.932</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.614</td>
<td>0.775</td>
<td>0.876</td>
<td>0.923</td>
<td>0.930</td>
<td>0.930</td>
<td>0.942</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.648</td>
<td>0.827</td>
<td>0.954</td>
<td>0.951</td>
<td>0.962</td>
<td>0.965</td>
<td>0.967</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.700</td>
<td>0.895</td>
<td>0.964</td>
<td>0.966</td>
<td>0.971</td>
<td>0.971</td>
<td>0.974</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.760</td>
<td>0.941</td>
<td>0.971</td>
<td>0.972</td>
<td>0.974</td>
<td>0.974</td>
<td>0.977</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.722</td>
<td>0.894</td>
<td>0.984</td>
<td>0.975</td>
<td>0.983</td>
<td>0.984</td>
<td>0.984</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.804</td>
<td>0.958</td>
<td>0.986</td>
<td>0.983</td>
<td>0.986</td>
<td>0.986</td>
<td>0.987</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.874</td>
<td>0.983</td>
<td>0.987</td>
<td>0.986</td>
<td>0.987</td>
<td>0.987</td>
<td>0.988</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.751</td>
<td>0.913</td>
<td>0.990</td>
<td>0.981</td>
<td>0.988</td>
<td>0.989</td>
<td>0.989</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.855</td>
<td>0.977</td>
<td>0.991</td>
<td>0.988</td>
<td>0.991</td>
<td>0.991</td>
<td>0.991</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.914</td>
<td>0.990</td>
<td>0.992</td>
<td>0.991</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
</tr>
</tbody>
</table>
Table G7: The Trace Ratio of the seven estimators for estimating $\Lambda$. 

with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0.5, 0.9]$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>PC</th>
<th>PC-GLS</th>
<th>PC-ITE</th>
<th>QMLE</th>
<th>ML-GLS</th>
<th>ML-ITE</th>
<th>ML-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>0.702</td>
<td>0.740</td>
<td>0.778</td>
<td>0.737</td>
<td>0.791</td>
<td>0.813</td>
<td>0.844</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.734</td>
<td>0.775</td>
<td>0.813</td>
<td>0.784</td>
<td>0.838</td>
<td>0.856</td>
<td>0.883</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.741</td>
<td>0.782</td>
<td>0.819</td>
<td>0.800</td>
<td>0.850</td>
<td>0.863</td>
<td>0.893</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.628</td>
<td>0.687</td>
<td>0.797</td>
<td>0.730</td>
<td>0.820</td>
<td>0.855</td>
<td>0.894</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.664</td>
<td>0.729</td>
<td>0.839</td>
<td>0.798</td>
<td>0.880</td>
<td>0.905</td>
<td>0.934</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.708</td>
<td>0.771</td>
<td>0.876</td>
<td>0.865</td>
<td>0.929</td>
<td>0.943</td>
<td>0.960</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.617</td>
<td>0.709</td>
<td>0.890</td>
<td>0.772</td>
<td>0.884</td>
<td>0.910</td>
<td>0.921</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.682</td>
<td>0.777</td>
<td>0.938</td>
<td>0.855</td>
<td>0.944</td>
<td>0.952</td>
<td>0.954</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.763</td>
<td>0.849</td>
<td>0.964</td>
<td>0.921</td>
<td>0.973</td>
<td>0.975</td>
<td>0.976</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.634</td>
<td>0.755</td>
<td>0.914</td>
<td>0.786</td>
<td>0.902</td>
<td>0.919</td>
<td>0.922</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.744</td>
<td>0.864</td>
<td>0.954</td>
<td>0.866</td>
<td>0.951</td>
<td>0.954</td>
<td>0.955</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.849</td>
<td>0.934</td>
<td>0.977</td>
<td>0.929</td>
<td>0.977</td>
<td>0.977</td>
<td>0.978</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.645</td>
<td>0.776</td>
<td>0.918</td>
<td>0.788</td>
<td>0.907</td>
<td>0.921</td>
<td>0.923</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.770</td>
<td>0.897</td>
<td>0.955</td>
<td>0.866</td>
<td>0.952</td>
<td>0.955</td>
<td>0.955</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.884</td>
<td>0.960</td>
<td>0.977</td>
<td>0.930</td>
<td>0.977</td>
<td>0.977</td>
<td>0.978</td>
</tr>
</tbody>
</table>
Table G8: The Trace Ratio of the seven estimators for estimating $F$.
with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0.5, 0.9]$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>PC Class</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>QMLE</th>
<th>ML-GLS</th>
<th>ML-ITE</th>
<th>ML-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>$T$</td>
<td>PC</td>
<td>PC-GLS</td>
<td>PC-ITE</td>
<td>QMLE</td>
<td>ML-GLS</td>
<td>ML-ITE</td>
<td>ML-EM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>0.569</td>
<td>0.601</td>
<td>0.597</td>
<td>0.655</td>
<td>0.660</td>
<td>0.664</td>
<td>0.659</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.557</td>
<td>0.593</td>
<td>0.589</td>
<td>0.658</td>
<td>0.666</td>
<td>0.670</td>
<td>0.672</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.538</td>
<td>0.579</td>
<td>0.578</td>
<td>0.652</td>
<td>0.661</td>
<td>0.664</td>
<td>0.678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.574</td>
<td>0.655</td>
<td>0.735</td>
<td>0.779</td>
<td>0.792</td>
<td>0.805</td>
<td>0.824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.561</td>
<td>0.655</td>
<td>0.748</td>
<td>0.798</td>
<td>0.813</td>
<td>0.824</td>
<td>0.859</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.562</td>
<td>0.669</td>
<td>0.767</td>
<td>0.828</td>
<td>0.844</td>
<td>0.851</td>
<td>0.885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>0.651</td>
<td>0.792</td>
<td>0.933</td>
<td>0.920</td>
<td>0.935</td>
<td>0.947</td>
<td>0.956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.662</td>
<td>0.827</td>
<td>0.949</td>
<td>0.947</td>
<td>0.958</td>
<td>0.961</td>
<td>0.964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.700</td>
<td>0.875</td>
<td>0.954</td>
<td>0.961</td>
<td>0.965</td>
<td>0.966</td>
<td>0.969</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.722</td>
<td>0.873</td>
<td>0.979</td>
<td>0.964</td>
<td>0.976</td>
<td>0.980</td>
<td>0.981</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.783</td>
<td>0.940</td>
<td>0.984</td>
<td>0.978</td>
<td>0.983</td>
<td>0.984</td>
<td>0.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.840</td>
<td>0.971</td>
<td>0.986</td>
<td>0.984</td>
<td>0.986</td>
<td>0.986</td>
<td>0.987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>0.751</td>
<td>0.897</td>
<td>0.987</td>
<td>0.975</td>
<td>0.985</td>
<td>0.987</td>
<td>0.988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>50</td>
<td>0.838</td>
<td>0.967</td>
<td>0.990</td>
<td>0.986</td>
<td>0.990</td>
<td>0.990</td>
<td>0.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>0.903</td>
<td>0.988</td>
<td>0.991</td>
<td>0.990</td>
<td>0.991</td>
<td>0.991</td>
<td>0.992</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>