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

Bai, Jushan and Li, Kunpeng

10 January 2012

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

Jushan Bai* and Kunpeng Li†

October 19, 2012

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. Comparison with the principal component method is made. The likelihood-based estimators are more efficient than those of principal component based. Monte Carlo simulations show the method is easy to implement and an application to the U.S. yield curves is considered.

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

*Department of Economics, Columbia University, New York, NY, 10027

†Department of Economics, Tsinghua University, Beijing, China.

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 Λ is an $N \times r$ matrix of factor loadings, r is the number of factors, and Ω is the covariance matrix of the idiosyncratic errors. An approximate factor model does not require Ω to be a diagonal matrix. In fact, there are no restrictions on the elements of Ω 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 Ω are fixed at certain known values, the number of free parameters in Ω alone is as many as that of Σ. Under fixed N, the model is not identifiable because the number of parameters (including those of Λ) exceeds the number of elements of Σ. However, Chamberlain and Rothschild show that the space spanned by the columns of Λ is identifiable from Σ as N goes to infinity under the assumption of an approximate factor model (bounded eigenvalue for Ω). However, Chamberlain and Rothschild do not study the sampling properties of the model because they assume Σ is known, which is equivalent to the case of $T = \infty$. In this paper, we do not assume a known Σ, but T observations on z_t ($t = 1, 2, ..., T$). 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 Σ. 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 method is more efficient than the principal components method. 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 the idiosyncratic variances, and derives 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 consider estimating the factors as functions of these estimated quantities, which is similar to the study of Doz et al. (2011a). 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. All these estimators are more efficient than the principal component estimator. In view of the ML method's predominant position in the statistics literature, it is of theoretical and practical interest to analyze the MLE for the approximate factor models. The analysis of the MLE in this paper is more challenging than the two-step estimators. The difficulty lies in the simultaneous estimation of the loadings and idiosyncratic variances; the estimators are solutions to a large number of nonlinear equations (first order conditions).

It should be 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.

Our analysis of the maximum likelihood estimator is invariably different from the classical literature. In classical factor analysis, a key assumption is that $\sqrt{N}\text{vech}(M_{zz}-\Sigma_{zz})$ has a normal limiting distribution as the number of observations T going to infinity, where M_{zz} is the sample covariance matrix of the data and $\Sigma_{zz} = E(M_{zz})$. This assumption does not hold when the dimension of data, N, also increases to infinity. In our case, the dimension of the matrix M_{zz} expands as N increases. When $N > T$, M_{zz} is not of full rank. Our analysis requires a limiting theory as both N and T go to infinity. While the analysis is more difficult, the final results (e.g., the limiting distributions) are much simpler than classical factor analysis, demonstrating the advantage of high dimensional framework.

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 Factor models

Let N denote the number of variables and T the sample size. For $i = 1, \ldots N$ and $t = 1, \ldots T$, the observation z_{it} is said to have a factor structure if

$$z_{it} = \alpha_i + \lambda_i' f_t + e_{it}, \tag{1}$$

where $f_t = (f_{t1}, f_{t2}, ..., f_{tr})'$ is the factor, and $\lambda_i = (\lambda_{i1}, ..., \lambda_{ir})'$ is the factor loading. 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 α be similarly defined. In matrix form,

$$z_t = \alpha + \Lambda f_t + e_t. \tag{2}$$

Only z_t is observable ($t \leq T$). Let $M_{zz} = \frac{1}{T}\sum_{t=1}^T \dot{z}_t \dot{z}_t'$, the sample variance of the observable data, where $\dot{z}_t = z_t - \frac{1}{T}\sum_{t=1}^T z_t$. Note the division by T in M_{zz} instead of $T - 1$ is for notational simplicity. 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 $M_{ff} = \frac{1}{T}\sum_{t=1}^T \dot{f}_t \dot{f}_t'$, which is the sample variance of f_t (we treat f_t as a sequence of fixed constants, see Assumption A below). Let $\Omega_t = E(e_t e_t')$, which allows for heteroskedasticity over t. In classical factor analysis, Ω_t is assumed to be diagonal. Here Ω_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 Ω_t contains as many free parameters as the number of elements in the sample variance M_{zz}, the number of parameters exceeds the number of estimating equations. So it is difficult to estimate all elements of Ω_t. Let

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

where $\mathrm{dg}(A)$ is a diagonal matrix that sets the off-diagonal elements of A to zero. We are interested in estimating the elements of Φ, 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. Define

$$\Sigma_{zz} = \Lambda M_{ff} \Lambda' + \Phi.$$

Because we restrict Φ to be a diagonal matrix, Σ_{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

$$\ln L = -\frac{1}{2N} \ln |\Sigma_{zz}| - \frac{1}{2N} tr(M_{zz} \Sigma_{zz}^{-1}). \tag{3}$$

Because Σ_{zz} is not the covariance matrix of z_t due to correlations and heteroscedasticities 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 Σ_{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 λ_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 Σ_{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_t 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 of exact factor specifications. We will establish consistency and derive the limiting distributions.

2.1 Assumptions

The following assumptions are needed for our analysis.

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 $M_{ff} = \frac{1}{T}\sum_{t=1}^T \dot{f}_t \dot{f}_t'$ be the sample variance of f_t, where $\dot{f}_t = f_t - T^{-1}\sum_{t=1}^T f_t$. There exists an $\overline{M}_{ff} > 0$ such that $\lim\limits_{T\to\infty} M_{ff} = \overline{M}_{ff}$.

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. As in Breitung and Tenhofen (2011), there is no need to model the dynamic process of f_t, especially when the parameters governing f_t are not of direct interest. The assumption that f_t are fixed constants is consistent with the fixed effects assumption and is also consistent with the idea that they are often the parameters of interest, although we do not directly estimate f_t.

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

Assumption B requires that the columns of Λ be linearly independent. If not, the matrix Q will not be of full rank.

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 $\Phi = \text{dg}\{\frac{1}{T}\sum_{t=1}^T E(e_t e_t')\} = \text{dg}\{\frac{1}{T}\sum_{t=1}^T \Omega_t\}$. So Φ is an $N \times N$ diagonal matrix with the ith element $\phi_i^2 = \frac{1}{T}\sum_{t=1}^T \tau_{ii,t}$ where $\tau_{ii,t}$ is the (i,i) element of Ω_t. We assume $C^{-2} \leq \phi_i^2 \leq C^2$ for all i.

C.3 $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq \tau_{ij}$ for some $\tau_{ij} > 0$ and for all t. In addition, $\sum_{i=1}^N \tau_{ij} \leq C$ for any j.

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

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

Assumption C allows for heteroskedasticities and weak correlations over the cross section and the time dimension, and is more general than 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 ϕ_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 Φ 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 Λ 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 First order conditions and identification restrictions

The first-order conditions of the MLE are (see e.g. Lawley and Maxwell (1971)):

$$\hat{\Lambda}'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = 0 \tag{4}$$

$$\text{diag}(\hat{\Sigma}_{zz}^{-1}) = \text{diag}(\hat{\Sigma}_{zz}^{-1} M_{zz} \hat{\Sigma}_{zz}^{-1}) \tag{5}$$

$$\hat{\Lambda}'\hat{\Sigma}_{zz}^{-1}\hat{\Lambda} = \hat{\Lambda}'\hat{\Sigma}_{zz}^{-1}M_{zz}\hat{\Sigma}_{zz}^{-1}\hat{\Lambda} \tag{6}$$

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 (4) is derived from the partial derivatives with respect to Λ, (5) is with respect to the diagonal elements of Φ, and (6) is with respect to M_{ff}. Equation (6) can be obtained from (4) by post-multiplying $\hat{\Sigma}_{zz}^{-1}\hat{\Lambda}$. So (6) is redundant. This redundancy arises from rotational indeterminacy, a well known fact for factor models. There are r^2 redundant parameters, so we need at least r^2 restrictions in order to uniquely fix the parameters. Rotational indeterminacy can be seen from, for any full rank matrix R, $\Sigma_{zz} = \Lambda M_{ff}\Lambda' + \Phi = \Lambda R'(R'^{-1}M_{ff}R^{-1})R\Lambda' + \Phi$. To fix the indeterminacy, we consider five sets of commonly used restrictions:

IC1: $\Lambda = (I_r, \Lambda_2')'$.

IC2: $\frac{1}{N}\Lambda'\Sigma_{ee}^{-1}\Lambda = I_r$ and $M_{ff} = D$, where D is a diagonal matrix, whose diagonal element are distinct and arranged in descending order.

IC3: $\frac{1}{N}\Lambda'\Sigma_{ee}^{-1}\Lambda = D$ and $M_{ff} = I_r$, where D is a diagonal matrix, whose diagonal element are distinct and arranged in descending order.

IC4: Λ_1 is a lower triangular matrix with all diagonal elements being 1 and $M_{ff} = D$, where Λ_1 is the upper $r \times r$ submatrix of Λ and D is a diagonal matrix.

IC5: Λ_1 is a lower triangular matrix with none of its diagonal element being 0 and $M_{ff} = I_r$, where Λ_1 is the upper $r \times r$ submatrix of Λ.

Under any one of these restrictions, the parameters can be either fully identified or identified up to a column sign change of Λ. More specifically, IC1 and IC4 allow full identification of the model, while IC2, IC3 and IC5 identify Λ up to a column sign change. In practice, IC1, IC4, and IC5 require careful choice of the first r variables (in order to give meaningful interpretations to the loadings and the factors). For more details on the identification conditions, we refer readers to Anderson and Rubin (1956), Lawley and Maxwell (1971), and Bai and Li (2012).

3 Asymptotic properties of the estimators

In this section, we establish consistency, rates of convergence, and the limiting distributions of the MLE.

3.1 Consistency and convergence rate

The challenge of the analysis lies in the infinite number of parameters in the limit, which makes the usual consistency concept not well defined. We tackle the problem by obtaining an average consistency first, and from the average consistency we 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 A.1 in the supplement gives the average consistency:

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

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.)

The average consistency of $\hat{\phi}_i^2$ is obtained by analyzing the properties of the likelihood function. The proof of the first and third results requires the use of identification conditions. The key idea of the proof is to find out the corresponding matrix which plays the same role as the rotation matrix R, and then use the identification conditions to prove that it converges in probability to an identity matrix. This matrix, as shown in Appendix A, is $\Lambda'\hat{\Phi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}$. The proof of $\Lambda'\hat{\Phi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1} \xrightarrow{p} I_r$ is quite different under different sets of the identification conditions. Under IC2, IC3 and IC5, we need to assume that the estimator $\hat{\Lambda}$ has the same column signs as those of Λ in order to have consistency. This restriction will be regarded as part of the identification conditions under IC2, IC3 and IC5.

We now state the rate of convergence.

Theorem 1 (Convergence rates) *Under Assumptions A-D, when $N, T \to \infty$, with any one of the identification conditions, we have*

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\tfrac{1}{\hat{\phi}_i^2}\|\hat{\lambda}_i - \lambda_i\|^2 = O_p(T^{-1}) + O_p(N^{-2}),$$

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\hat{\phi}_i^2 - \phi_i^2)^2 = O_p(T^{-1}) + O_p(N^{-2}),$$

$$\|\hat{M}_{ff} - M_{ff}\|^2 = O_p(T^{-1}) + O_p(N^{-2}).$$

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. Theorem 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.

The principal components estimator has a slower convergence rate. Bai (2003) shows that $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\lambda}_i - R\lambda_i\|^2 = O_p(1/T) + O_p(1/N)$, where R is an $r \times r$ invertible matrix. The principal components method does not take into account heteroskedasticity, there is a bias arising from ignoring the heteroskedasticity.

Remark 1. Part of the ML analysis includes showing that $R = I_r$, that is, the MLE directly estimates λ_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 is not true, 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.

3.2 Asymptotic representation and limiting distribution

Additional assumptions are needed for the asymptotic representations and the limiting distributions of the QMLE.

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

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

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

E.3 the $r \times r$ matrix satisfies $E\Big[\Big\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_i^4}\lambda_i\lambda_i'(e_{it}^2 - \phi_i^2)\Big\|^2\Big] \le C$.

Assumption F [Central Limit Theorem]:

F.1 For each i, as $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t e_{it} \xrightarrow{d} N(0, \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} f_t f_s' \rho_{i,ts})$.

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 $\sigma_i^2 = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E\Big[(e_{it}^2 - \phi_i^2)(e_{is}^2 - \phi_i^2)\Big]$.

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 standard. 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 $\hat{\phi}_i^2$.

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 $\mathcal{P}_t$ and $\mathcal{Q}_t$. Their (g, h)-th elements are defined, respectively, as $(g, h = 1, 2, ..., r)$

$$\mathcal{P}_{gh,t} = \begin{cases} -m_g^{-1} f_{tg}\xi_t'\Lambda_1'^{-1}v_h & \text{if } g \ge h \\ -m_g^{-1} m_h \mathcal{P}_{hg,t} & \text{if } g < h \end{cases}, \qquad \mathcal{Q}_{gh,t} = \begin{cases} -f_{tg}\xi_t'\Lambda_1'^{-1}v_h & \text{if } g > h \\ 0 & \text{if } g = h \\ -\mathcal{Q}_{hg,t} & \text{if } g < h \end{cases}$$

where m_g is the gth diagonal element of M_{ff}; f_{tg} is the gth component of f_t; Λ_1 is the first $r \times r$ block of Λ; and v_h is the hth column of the identity matrix I_r. Matrix $\mathcal{Q}_t$

is skew-symmetric. Now we state the asymptotic representations for the estimated factor loadings.

Theorem 2 (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:*

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

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

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

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

where ξ_t, $\mathcal{P}_t$ and $\mathcal{Q}_t$ are all defined earlier.

Theorem 2 shows that $\hat{\lambda}_j$ is $\sqrt{T}$-consistent for λ_j. Theorem 2 also indicates that different sets of identification conditions lead to different asymptotic representations. Under IC2 and IC3 the asymptotic representations are simpler. The restrictions of IC1, IC4 and IC5 impose restrictions on the first r factor loadings, which in turn put more weights on the first r observations. This explains why the error terms of the first r observations enter into the asymptotic representations (via ξ_t, $\mathcal{P}_t$ and $\mathcal{Q}_t$), leading to more complex representations.

If the factors f_t can be observable, the estimator of λ_j by applying OLS is $\hat{\lambda}_j^{ols} = \Big(\frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})'\Big)^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(z_{tj} - \bar{z}_j)\Big)$ (a time series regression), which will yield the same asymptotic representation as that of IC2 and IC3 (note we assume $\bar{f} = 0$). So the MLE under high dimension amounts to make the unobservable factors observable. It is interesting that we never attempt to estimate the individual f_t, but we achieve the same effects as if the individual f_t were known, an interesting result for high dimensional data.

The limiting distributions of $\hat{\lambda}_i$ follow from the asymptotic representations.

Corollary 1 (Limiting distributions for factor loadings) *Under the same assumptions as Theorem 2, together with Assumption F, we have:*

$$\text{Under IC1,} \quad \sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, (\overline{M}_{ff})^{-1} \Gamma_j^{\lambda} (\overline{M}_{ff})^{-1});$$

$$\text{Under IC2 or IC3,} \quad \sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, (\overline{M}_{ff})^{-1} \Upsilon_j^{\lambda} (\overline{M}_{ff})^{-1});$$

$$\text{Under IC4,} \quad \sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, \Pi_j^{\lambda});$$

$$\text{Under IC5,} \quad \sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, \Psi_j^{\lambda});$$

where $\Gamma_j^{\lambda}, \Upsilon_j^{\lambda}, \Pi_j^{\lambda}, \Psi_j^{\lambda}$ are defined in Table 3, and $\overline{M}_{ff}$ is defined in Assumption A.

From the asymptotic representations of Theorem 2, under each set of the identification conditions, the summation over t only involves f_t and e_{jt}. So Assumption F.1 is sufficient for the limiting results. The superscript λ in the limiting variances signifies the association with the factor loadings. We will use similar matrices with a superscript f when estimating factors f_t in a later section.

Now we state the limiting results for the estimated M_{ff}.

Theorem 3 (Asymptotic representations for $\hat{M}_{ff}$) *Under the assumptions of Theorem 2 and $\sqrt{T}/N \to 0$, we have:*

$$\text{Under IC1,} \quad \sqrt{T}[\text{vech}(\hat{M}_{ff} - M_{ff})] = D_r^+\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\xi_t \otimes f_t + f_t \otimes \xi_t)\Big) + o_p(1);$$

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

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

where D_r^+ is the Moore-Penrose inverse of the duplication matrix D_r.

Note that under IC3 and IC5, $M_{ff} = I_r$ is known, not estimated.

Corollary 2 (Limiting distribution for $\hat{M}_{ff}$) *Under the assumptions of Theorem 3 together with Assumption F, we have:*

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

$$\text{Under IC4,} \quad \sqrt{T}\Big(\text{diag}\{\hat{M}_{ff} - M_{ff}\}\Big) \xrightarrow{d} N\Big(0, 4J_r\Pi^M J_r'\Big);$$

where Γ^M and Π^M are defined in Table 3; J_r is an $r \times r^2$ matrix, which satisfies, for any $r \times r$ matrix M, $\text{diag}\{M\} = J_r\text{vec}(M)$.

Theorem 3 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 $\sqrt{NT}$ 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 3 also shows that, under IC1 and IC4, the asymptotic representation of $\hat{M}_{ff} - M_{ff}$ only involves the error terms $\xi_t = (e_{1t}, e_{2t}, \cdots, e_{rt})'$. 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 4 *Under the assumptions of Theorem 2 and $\sqrt{T}/N \to 0$, irrespective of which set of identification conditions, we have*

$$\sqrt{T}(\hat{\phi}_i^2 - \phi_i^2) = \tfrac{1}{\sqrt{T}} \textstyle\sum_{t=1}^{T} (e_{it}^2 - \phi_i^2) + o_p(1)$$

Corollary 3 *Under the assumptions of Theorem 4 and Assumption F, we have*

$$\sqrt{T}(\hat{\phi}_i^2 - \phi_i^2) \xrightarrow{d} N(0, \sigma_i^2),$$

where σ_i^2 is defined in Assumption F.2.

Theorem 4 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, the estimator ϕ_i^2 gives a consistent estimator for the variance of the process. With heteroskedasticity, the estimator $\hat{\phi}_i^2$ provides an estimate for the average variance.

It is interesting to note that, to estimate ϕ_i^2, there is no need to estimate the residuals e_{it}. Estimating the residuals would require to estimate both λ_i and f_t, as in two-step procedures. If N is fixed, then f_t cannot be consistently estimated (even for exact factor models). This would imply that the idiosyncratic variances cannot be consistently estimated using the residuals. The ML procedure does not estimate f_t $(t = 1, 2, ..., T)$ but only the sample covariance of f_t, thus it is able to provide a consistent estimation of the idiosyncratic variances under fixed N with an exact factor structure. Under large N and T, an exact factor structure is not required.

4 Asymptotic properties for the estimated factors

The factors f_t can be estimated by two different methods. One is the projection formula and the other is the generalized least squares (GLS). They are

$$\text{(Projection formula)} \quad \tilde{f}_t = (\hat{M}_{ff}^{-1} + \hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1}(z_t - \bar{z}) \tag{7}$$

$$\text{(GLS)} \quad \hat{f}_t = (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1}(z_t - \bar{z}) \tag{8}$$

see, e.g., Anderson (2003). 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′ $\quad \sum_{t=1}^{T} \rho_{ts} \le C$, where $\rho_{ts} \ge 0$ is defined in Assumption C.4.

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

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

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

E.6 for all t, $t = 1, 2, \cdots, T$, $E\Big(\Big\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{1}{\phi_i^4}\lambda_i(e_{is}^2-\phi_i^2)e_{it}\Big\|^2\Big) \leq C$.

Assumption F [Central Limit Theorem (continued)]

F.3 for each t, as $N \to \infty$, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it} \xrightarrow{d} N\Big(0, \lim\limits_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2\phi_j^2}\lambda_i\lambda_j'\tau_{ij,t}\Big)$.

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 $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}(\hat{\phi}_i^2-\phi_i^2)\lambda_i e_{it}$, 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 5 (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 $N/T \to \Delta$,

$$\sqrt{N}(\hat{f}_t - f_t) = -\sqrt{\Delta}\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\xi_s f_s'\Big)M_{ff}^{-1}f_t + Q^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it} + o_p(1).$$

Under IC2 or IC3,

$$\sqrt{N}(\hat{f}_t - f_t) = Q^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it} + o_p(1)$$

Under IC4 and $N/T \to \Delta$,

$$\sqrt{N}(\hat{f}_t - f_t) = -\sqrt{\Delta}\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathcal{P}_s'\Big)f_t + Q^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it} + o_p(1).$$

Under IC5 and $N/T \to \Delta$,

$$\sqrt{N}(\hat{f}_t - f_t) = -\sqrt{\Delta}\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathcal{Q}_s'\Big)f_t + Q^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it} + o_p(1).$$

The variables ξ_t, $\mathcal{P}_s$ and $\mathcal{Q}_s$ are defined earlier.

The asymptotic representations depend on the identification conditions. Once again, 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 Δ 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 Δ 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 5 shows that the relative ratio between N and T plays a role in efficiency.

From Theorem 5, the limiting distributions can be obtained easily. Under IC1, IC4, and IC5, we assume that the two terms in the representations are asymptotically independent. This is a reasonable assumption since the first term involves the sum of e_{is} over the time dimension for the first r individuals only ($i = 1, 2, ..., r$), whereas the second term involves the sum over the entire cross section for a given period. It is also easy to derive the limiting distribution without the asymptotic independence assumption, and in this case, the covariances of the two terms also enter into the limiting variance.

Corollary 4 (Limiting distributions for the estimated factors) *Under the assumptions of Theorem 5 and Assumption F, we have*

$$\text{under IC1 and } N/T \to \Delta, \quad \sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, \Gamma_t^f);$$
$$\text{under IC2 or IC3,} \quad \sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, \Upsilon_t^f);$$
$$\text{under IC4 and } N/T \to \Delta, \quad \sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, \Pi_t^f);$$
$$\text{under IC5 and } N/T \to \Delta, \quad \sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, \Psi_t^f);$$

where $\Gamma_t^f, \Upsilon_t^f, \Pi_t^f, \Psi_t^f$ *are given in Table 3.*

Note that IC2 and IC3 do not need $N/T \to \Delta$ but only $\sqrt{N}/T \to 0$.

Consider a special case in which e_{it} are uncorrelated over i and homoscedastic over t (still allow cross-section heteroskedasticity and serial correlation), then the limiting distributions under IC2 and IC3 reduce to $\sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, Q^{-1})$ because Υ^f reduces to Q^{-1}. This is the same limiting distribution as the infeasible GLS in the cross-section regression $z_{it} = f_t'\lambda_i + e_{it}$ as if all λ_i and ϕ_i^2 were observable.

Remark 2. Suppose that f_t is a vector autoregressive process such that $\Psi(L)f_t = u_t$, where $\Psi(L)$ is a finite order polynomial of the lag operator L. We point out that modeling the dynamics of f_t will not improve the estimation efficiency. In Appendix F of the supplementary document, we show that $\hat{f}_t$ has the same asymptotic representation as the Kalman-smoother-based estimators $\hat{f}_t^{ks}$ that takes into account the dynamics of f_t. That is, we establish that $\sqrt{N}(\hat{f}_t^{ks} - \hat{f}_t) = o_p(1)$. The estimator $\hat{f}_t^{ks}$ is similar to that of Doz et al. (2011b), although the first step here is based on the QMLE instead of the PC estimates. The asymptotic equivalence implies the limiting distribution for $\hat{f}_t^{ks}$ and also for the estimator of Doz et al. (2011b), who do not study the limiting distribution.

5 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 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. In this section, we first consider a two-step procedure that ignores the dynamics in f_t. We then consider the full maximum likelihood method that jointly estimates the dynamics of f_t and that of e_{it}.

5.1 Ignoring the dynamics in factors

Consider the following model

$$\begin{aligned} z_{it} &= \lambda_i' f_t + e_{it}, \\ e_{it} &= \rho_{i,1} e_{it-1} + \cdots + \rho_{i,p_i} e_{it-p_i} + \epsilon_{it} \end{aligned} \tag{9}$$

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 ϵ_{it} and ϵ_{jt} are independent for $i \neq j$, for simplicity; $E\epsilon_{it} = 0$ and $\text{var}(\epsilon_{it}) = \sigma_{\epsilon i}^2$.

Breitung and Tenhofen (2011) consider a two-step method to estimate model (9). 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.

However, 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 $\tilde{\Lambda}, \tilde{F}, \hat{\rho}_1, \ldots, \hat{\rho}_N, \hat{\Phi}$, are calculated by the following two steps.

1. Apply the QML method to the first equation of (9) 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 $\hat{e}_{it} = z_{it} - \hat{\lambda}_i'\hat{f}_t$. For each i, obtain the estimators $\hat{\rho}_i$ by running the following regression
$$\hat{e}_{it} = \rho_{i,1}\hat{e}_{i,t-1} + \cdots + \rho_{i,p_i}\hat{e}_{i,t-p_i} + error, \quad t = p_i + 1, \ldots, T$$

2. Given $(\hat{\rho}_{i,1}, \hat{\rho}_{i,2}, \ldots, \hat{\rho}_{i,p_i})$ and $\hat{F}$, update the estimator of Λ, denoted by $\tilde{\Lambda}$, by running the regression
$$z_{it} - \hat{\rho}_{i,1}z_{i,t-1} - \cdots - \hat{\rho}_{i,p_i}z_{i,t-p_i} = (\hat{f}_t - \hat{\rho}_{i,1}\hat{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\hat{f}_{t-p_i})'\lambda_i + error, \quad t = p_i + 1, \ldots, T$$
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
$$\frac{1}{\hat{\phi}_i} z_{it} = \left(\frac{1}{\hat{\phi}_i}\tilde{\lambda}_i\right)' f_t + error, \quad i = 1, 2, \ldots, N$$

The iterated ML-GLS can be obtained by iterating the above two steps several times and, for each iteration, $\hat{\Lambda}, \hat{F}$ 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 6 *Under the Assumptions in Appendix E, when $N, T \to \infty$, we have*

$$\hat{\rho}_i \xrightarrow{p} \rho_i, \quad \tilde{\lambda}_i \xrightarrow{p} \lambda_i, \quad \text{for} \quad i = 1, 2, \ldots, N$$
$$\tilde{f}_t \xrightarrow{p} f_t, \quad \text{for} \quad t = 1, 2, \ldots, T$$

Furthermore, with the condition $\sqrt{T}/N \to 0$, for each i,

$$\sqrt{T}(\hat{\rho}_i - \rho_i) = \Big(\frac{1}{T}\sum_{t=p_i+1}^{T} \psi_{it}\psi_{it}'\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=p_i+1}^{T} \psi_{it}\epsilon_{it}\Big) + o_p(1)$$

$$\sqrt{T}(\tilde{\lambda}_i - \lambda_i) = \Big(\frac{1}{T}\sum_{t=p_i+1}^{T} g_{it}g_{it}'\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=p_i+1}^{T} g_{it}\epsilon_{it}\Big) + o_p(1)$$

and with the condition $\sqrt{N}/T \to 0$, for each t,

$$\sqrt{N}(\tilde{f}_t - f_t) = \Big(\frac{1}{N}\sum_{i=1}^{N} \frac{1}{\phi_i^2}\lambda_i\lambda_i'\Big)^{-1}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{1}{\phi_i^2}\lambda_i e_{it}\Big) + o_p(1)$$

where $\psi_{it} = (e_{it-1}, e_{it-2}, \ldots, e_{it-p_i})'$ and $g_{it} = f_t - \rho_{i,1}f_{t-1} - \cdots - \rho_{i,p_i}f_{t-p_i}$.

Here are some intuitions for Theorem 6. Consider the estimation of λ_i. If both f_t and $\rho_{i,1}, \ldots, \rho_{i,p_i}$ are observable, directly applying GLS to the equation

$$z_{it} - \rho_{i,1}z_{i,t-1} - \cdots - \rho_{i,p_i}z_{i,t-p_i} = (f_t - \rho_i f_{t-1} - \cdots - \rho_{i,p_i}f_{t-p_i})'\lambda_i + \epsilon_{it}$$

will give the same limiting distributions as stated in Theorem 6. Thus the ML-GLS estimation amounts to make the unobservable f_t and $\rho_{i,1}, \ldots, \rho_{i,p_i}$ observable asymptotically. Similar results hold for the estimated f_t and ρ_i. From Theorem 6 we can easily obtain the following limiting distributions:

Corollary 5 *Under the assumptions of Theorem 6, if $\sqrt{T}/N \to 0$, for each i,*

$$\sqrt{T}(\hat{\rho}_i - \rho_i) \xrightarrow{d} N\Big(0, \sigma_{\epsilon i}^2\Big[\text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=p_i+1}^{T} \psi_{it}\psi_{it}'\Big]^{-1}\Big),$$

$$\sqrt{T}(\tilde{\lambda}_i - \lambda_i) \xrightarrow{d} N\Big(0, \sigma_{\epsilon i}^2\Big[\text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=p_i+1}^{T} g_{it}g_{it}'\Big]^{-1}\Big).$$

If $\sqrt{N}/T \to 0$ and with Q given in Assumption B, for each t,

$$\sqrt{N}(\tilde{f}_t - f_t) \xrightarrow{d} N(0, Q^{-1}).$$

A consistent estimator for $\sigma_{\epsilon i}^2$ is $\hat{\sigma}_{\epsilon i}^2 = \frac{1}{T-p_i}\sum_{t=p_i+1}^{T}\hat{\epsilon}_{it}^2$, where

$$\hat{\epsilon}_{it} = z_{it} - \hat{\rho}_{i,1}z_{i,t-1} - \cdots - \hat{\rho}_{i,p_i}z_{i,t-p_i} - (\tilde{f}_t - \hat{\rho}_{i,1}\tilde{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\tilde{f}_{t-p_i})'\tilde{\lambda}_i$$

The asymptotic variance of $\sqrt{T}(\tilde{\lambda}_i - \lambda_i)$ can be constructed for finite samples by $\hat{\sigma}_{\epsilon i}^2(\frac{1}{T}\sum_{t=p_i+1}^{T}\tilde{g}_{it}\tilde{g}_{it}')^{-1}$ with $\tilde{g}_{it} = \tilde{f}_t - \hat{\rho}_{i,1}\tilde{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\tilde{f}_{t-p_i}$ and the asymptotic variance of $\sqrt{T}(\hat{\rho}_i - \rho_i)$ can be constructed by $\hat{\sigma}_{\epsilon i}^2(\frac{1}{T}\sum_{t=p_i+1}^{T}\tilde{v}_{it}\tilde{v}_{it}')^{-1}$ with $\tilde{v}_{it} = (\tilde{e}_{it-1}, \tilde{e}_{it-2}, \ldots, \tilde{e}_{it-p_i})'$ and $\tilde{e}_{it} = z_{it} - \tilde{\lambda}_i'\tilde{f}_t$. Matrix Q can be consistently estimated by $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i\tilde{\lambda}_i'$.

5.2 Joint estimation of the dynamics in factors and in errors

Assume that f_t follows a vector autoregressive process of order K:

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

We can still use the foregoing two-step method to obtain the estimators $\tilde{\Lambda}, \tilde{F}, \hat{\Phi}$ and all $\hat{\rho}_i$ $(i = 1, 2, ..., N)$. These estimators have the same limiting distributions as in Section 5.1. To obtain an estimate for Ψ_k $(k \leq K)$, an extra step is taken by regressing $\tilde{f}_t$ on its lags. Let $\hat{\Psi}_k$ denote the resulting estimator. The limiting distribution of $\hat{\Psi}_k$ is the same as the case of known f_t.

The dynamics in the factors and in the idiosyncratic errors 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 (1989) 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. Note that the model here is a special case of the generalized dynamic factor model of Forni et el. (2000); the latter model is estimated by the frequency domain approach. With more structure, the present model allows a full maximum likelihood estimation.

Here we elaborate the ECM (expectation and constrained maximization) algorithm of Meng and Rubin (1993). ECM is a sequential maximization procedure that maximizes the expected complete-data likelihood with respect to a subcomponent of the parameters, and with the remaining components constrained at the previously obtained optimal values. A useful property of the ECM is that it has closed-form solutions when the parameters are appropriately divided into subgroups.

For ease of exposition, we assume that the idiosyncratic errors and the factors are $AR(1)$ processes, namely, $e_{it} = \rho_i e_{it-1} + \epsilon_{it}$ and $f_t = \Psi f_{t-1} + u_t$, $\varepsilon_{it} \sim N(0, \sigma^2_{\epsilon i})$ and $u_t \sim N(0, I_r)$; both errors are iid over t. The procedures can be easily stated for more heterogeneous dynamics. The model can be written as

$$\begin{aligned} z_t - \rho z_{t-1} &= [\Lambda, -\rho\Lambda] \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} + \epsilon_t \\ \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} &= \begin{bmatrix} \Psi & 0 \\ I_r & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \end{bmatrix} \end{aligned} \tag{10}$$

where $\rho = \text{diag}(\rho_1, \ldots, \rho_N)$. Let $\theta = (\Lambda, \rho_1, \ldots, \rho_N, \sigma^2_{\epsilon 1}, \ldots, \sigma^2_{\epsilon N}, \Psi)$ denote the parameters. The complete-data likelihood function is

$$\ln L(\theta) = C - \frac{1}{2N}\sum_{i=1}^{N} \ln \sigma^2_{\epsilon i} - \frac{1}{2NT}\sum_{i=1}^{N} \frac{1}{\sigma^2_{\epsilon i}} \sum_{t=2}^{T} \Big(z_{it} - \rho_i z_{it-1} - \lambda_i' f_t + \rho_i \lambda_i' f_{t-1}\Big)^2$$

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

$$Q(\theta|\theta^*) = C - \frac{1}{2N}\sum_{i=1}^{N} \ln \sigma^2_{\epsilon i} - \frac{1}{2NT}\sum_{i=1}^{N} \frac{1}{\sigma^2_{\epsilon i}} \sum_{t=2}^{T} \Big\{ (z_{it} - \rho_i z_{it-1})^2$$

$$-2(z_{it}-\rho_i z_{it-1})\lambda_i' E(f_t|\theta^*) + 2(z_{it}-\rho_i z_{it-1})\rho_i\lambda_i' E(f_{t-1}|\theta^*) \tag{11}$$
$$+\lambda_i' E(f_t f_t'|\theta^*)\lambda_i + \rho_i^2\lambda_i' E(f_{t-1}f_{t-1}'|\theta^*)\lambda_i - 2\rho_i\lambda_i' E(f_{t-1}f_t'|\theta^*)\lambda_i\Big\}$$

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 $V_{00,t}=E(f_tf_t'|\theta^*,Z)$, $V_{01,t}=E(f_tf_{t-1}'|\theta^*,Z), V_{11,t}=E(f_{t-1}f_{t-1}'|\theta^*,Z)$. In the E-step, we compute these conditional expectations at $\theta^*=\theta^{(k)}$, where $\theta^{(k)}$ denotes the kth iteration of θ in the ECM algorithm. These conditional expectations are computed via the Kalman smoothers in view that system (10) 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 θ in (11). By dividing θ into four subgroups, the ECM of Meng and Rubin (1993) leads to the following updating formulae:

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

$$\rho_i^{(k+1)} = \Big[\sum_{t=2}^T\Big(z_{it-1}^2-2z_{it-1}\lambda_i^{(k+1)\prime}E(f_{t-1}|\theta^{(k)})+\lambda_i^{(k+1)\prime}V_{11,t}\lambda_i^{(k+1)}\Big)\Big]^{-1}$$
$$\times\Big[\sum_{t=2}^T\Big(z_{it}z_{it-1}-z_{it}\lambda_i^{(k+1)\prime}E(f_{t-1}|\theta^{(k)})-z_{it-1}\lambda_i^{(k+1)\prime}E(f_t|\theta^{(k)})+\lambda_i^{(k+1)\prime}V_{01,t}\lambda_i^{(k+1)}\Big)\Big],$$

$$\begin{aligned}(\sigma_{\epsilon i}^{(k+1)})^2 =& \frac{1}{T-1}\sum_{t=2}^T\Big((z_{it}-\rho_i^{(k+1)}z_{it-1})^2-2(z_{it}-\rho_i^{(k+1)}z_{it-1})\lambda_i^{(k+1)\prime}E(f_t|\theta^{(k)})\\
&+2\rho_i^{(k+1)}(z_{it}-\rho_i^{(k+1)}z_{it-1})\lambda_i^{(k+1)\prime}E(f_{t-1}|\theta^{(k)})+\lambda_i^{(k+1)\prime}V_{00,t}\lambda_i^{(k+1)}\\
&-2\rho_i^{(k+1)}\lambda_i^{(k+1)\prime}V_{10,t}\lambda_i^{(k+1)}+(\rho_i^{(k+1)})^2\lambda_i^{(k+1)\prime}V_{00,t}\lambda_i^{(k+1)}\Big),\end{aligned}$$

$$\Psi^{(k+1)} = \Big(\sum_{t=2}^T V_{01,t}\Big)\Big(\sum_{t=2}^T V_{11,t}\Big)^{-1}.$$

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. The estimator will be referred to as ML-EM in the next subsection.

While the computation is straightforward, the statistical analysis of the full maximum likelihood estimators require extensive argument, and is more challenging than the QMLE considered here, largely owing to additional and more complex first order conditions. This issue is being examined by the authors.

6 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 ϵ_t being *i.i.d.* $N(0, \mathcal{T})$; $A(L)$ and $D(L)$ are defined as $A(L) = I_r - \psi I_r L, D(L) = I_N - \rho L$, where $\rho = \text{diag}(\rho_1, \ldots, \rho_N)$ and ψ is a scalar. Matrix $\mathcal{T}$ is $N \times N$ with its (i,j)th element $\tau^{|i-j|}[\phi_i^2 \phi_j^2 (1-\rho_i^2)(1-\rho_j^2)]^{1/2}$. The variance of e_{it}, ϕ_i^2, is generated according to

$$\phi_i^2 = \frac{\beta_i}{1-\beta_i} \frac{1}{1-\psi^2} \lambda_i' \lambda_i \tag{12}$$

where β_i are iid $U[u, 1-u]$ with $u \in [0, 0.5]$. All the elements of Λ 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, β_i is the ratio between the variance of e_{it} and the variance of z_{it}. Since β_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 τ 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 ρ and ψ 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 ρ_i from $U[0, 0.9]$[1].

As a measure of goodness-of-fit, we use the Trace-Ratio (TR) to evaluate how close the estimated values Λ 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.

For comparison, we also compute the PC estimators, PC-GLS estimators and iterative PC-GLS estimators (denoted by PC-ITE below)[2]. These estimators are discussed in Section 5.1. 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.

Table 1 reports 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 Λ. 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

[1]We also consider ρ_i from $U[0.5, 0.9]$ and the simulation results are presented in Appendix G.

[2]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.

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 (say, $N = 150, T = 100$). This is consistent with the theory.

Table 2 reports 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 Λ and 0.819 for F. In contrast, the counterparts in Table 2 are 0.783 for Λ 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 table 2 are 0.944 for Λ 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.

Table 1.1: The Trace Ratio of the seven estimators for estimating Λ with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.847	0.874	0.893	0.916	0.939	0.943	0.947
10	50	0.865	0.896	0.913	0.948	0.966	0.968	0.972
10	100	0.890	0.919	0.932	0.973	0.984	0.984	0.986
20	30	0.760	0.801	0.883	0.899	0.931	0.933	0.936
20	50	0.803	0.845	0.922	0.939	0.961	0.962	0.963
20	100	0.849	0.887	0.949	0.971	0.982	0.982	0.982
50	30	0.753	0.804	0.908	0.890	0.925	0.926	0.927
50	50	0.816	0.866	0.951	0.934	0.957	0.957	0.958
50	100	0.878	0.918	0.974	0.966	0.979	0.979	0.979
100	30	0.798	0.856	0.922	0.890	0.925	0.925	0.925
100	50	0.877	0.922	0.955	0.933	0.956	0.956	0.956
100	100	0.932	0.960	0.978	0.966	0.978	0.978	0.978
150	30	0.824	0.883	0.924	0.890	0.924	0.924	0.925
150	50	0.898	0.939	0.956	0.933	0.956	0.956	0.956
150	100	0.948	0.970	0.978	0.966	0.978	0.978	0.978

Table 1.2: The Trace Ratio of the seven estimators for estimating F with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.655	0.711	0.702	0.819	0.825	0.828	0.836
10	50	0.640	0.703	0.699	0.837	0.843	0.845	0.871
10	100	0.648	0.717	0.718	0.856	0.860	0.860	0.888
20	30	0.633	0.754	0.840	0.909	0.915	0.917	0.926
20	50	0.646	0.780	0.863	0.922	0.926	0.927	0.939
20	100	0.662	0.808	0.879	0.931	0.933	0.933	0.946
50	30	0.715	0.881	0.951	0.966	0.970	0.970	0.974
50	50	0.743	0.915	0.964	0.971	0.973	0.973	0.978
50	100	0.781	0.943	0.969	0.974	0.975	0.975	0.980
100	30	0.820	0.951	0.983	0.984	0.986	0.986	0.988
100	50	0.866	0.978	0.986	0.986	0.987	0.987	0.989
100	100	0.892	0.985	0.988	0.987	0.988	0.988	0.990
150	30	0.871	0.974	0.991	0.989	0.991	0.991	0.992
150	50	0.914	0.989	0.992	0.991	0.992	0.992	0.993
150	100	0.933	0.991	0.992	0.992	0.992	0.992	0.994

Table 2.1: The Trace Ratio of the seven estimators for estimating Λ with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.738	0.749	0.764	0.783	0.801	0.808	0.816
10	50	0.749	0.757	0.764	0.804	0.816	0.819	0.829
10	100	0.762	0.771	0.781	0.828	0.841	0.846	0.853
20	30	0.672	0.690	0.756	0.792	0.828	0.837	0.848
20	50	0.702	0.719	0.796	0.849	0.883	0.890	0.903
20	100	0.736	0.749	0.827	0.898	0.919	0.923	0.936
50	30	0.670	0.706	0.856	0.829	0.881	0.886	0.890
50	50	0.747	0.783	0.913	0.890	0.929	0.931	0.933
50	100	0.813	0.838	0.952	0.940	0.964	0.964	0.966
100	30	0.728	0.783	0.887	0.837	0.888	0.891	0.893
100	50	0.814	0.862	0.934	0.895	0.934	0.935	0.936
100	100	0.888	0.920	0.966	0.943	0.967	0.967	0.968
150	30	0.754	0.814	0.889	0.835	0.887	0.890	0.891
150	50	0.848	0.899	0.937	0.896	0.936	0.937	0.937
150	100	0.915	0.945	0.968	0.944	0.967	0.968	0.968

Table 2.2: The Trace Ratio of the seven estimators for estimating F with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.587	0.615	0.599	0.681	0.681	0.679	0.694
10	50	0.562	0.587	0.562	0.662	0.662	0.658	0.686
10	100	0.550	0.578	0.561	0.669	0.669	0.668	0.698
20	30	0.584	0.653	0.705	0.810	0.814	0.814	0.843
20	50	0.581	0.663	0.730	0.839	0.842	0.841	0.873
20	100	0.578	0.666	0.733	0.854	0.856	0.854	0.887
50	30	0.669	0.805	0.924	0.950	0.955	0.957	0.963
50	50	0.709	0.857	0.946	0.960	0.963	0.963	0.971
50	100	0.732	0.890	0.955	0.966	0.967	0.967	0.974
100	30	0.788	0.914	0.977	0.978	0.981	0.982	0.985
100	50	0.834	0.957	0.984	0.983	0.985	0.985	0.987
100	100	0.868	0.975	0.985	0.985	0.986	0.986	0.989
150	30	0.844	0.953	0.988	0.986	0.988	0.989	0.990
150	50	0.896	0.983	0.990	0.989	0.990	0.990	0.992
150	100	0.920	0.988	0.991	0.991	0.991	0.991	0.993

7 Application

In this section, we estimate the U.S. yield curves by the factor method. The data used here are the U.S. Treasury yields for the period of November 1971 to May 2009, with the same 17 maturities as in Diebold and Li (2006). These maturities are 3, 6, 9, 12, 15, 18, 21, 24, 28, 32, 36, 48, 60, 72, 84, 96, 108 months.

A well-known parametric model for yield curves is that of Nelson-Siegel (see, Diebold et al. (2006) and Nelson and Siegel (1987)):

$$y_t(\tau) = L_t + S_t\Big(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\Big) + C_t\Big(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\Big) + e_{t\tau}. \tag{13}$$

where $y_t(\tau)$ denotes the yield at time t with maturity τ; L_t, S_t and C_t denote the time-varying level, slope, and curvature factors. These factors are interpreted as long-term, short-term and medium term factors; see Diebold and Li (2006) for details.

Specification (13) is parametric because all the factor loadings depend on a single parameter λ. For comparison purpose, we also fit the data to this parametric model. Our estimation is as follows: (i) for a given λ, obtain L_t, S_t and C_t by regressing $y_t(\tau)$ on the known factor loadings; (ii) given the estimated factors $\hat{L}_t, \hat{S}_t$ and $\hat{C}_t$, obtain λ by the nonlinear least squares. Iterating the above two steps until the changes in λ are small. Using this method, the estimate of λ is 0.0606, which is close to 0.0609 in Diebold and Li (2006), who obtain the value by maximizing the loading on the curvature factor at maturity $\tau = 30$ months.

We next relax the restrictions on the factor loadings and consider the following nonparametric specification:

$$y_t(\tau) = L_t + S_t D_{\tau 1} + C_t D_{\tau 2} + e_{t\tau}. \tag{14}$$

Equation (14) is more general than (13). The factor loadings $D_{\tau 1}$ and $D_{\tau 2}$ are not restricted to be parametric. This provides a way of checking wether the parametric specification of (13) is supported by the actual data. To estimate (14), we first estimate L_t by the cross-sectional mean, then apply the QML method to the demeaned data. For economic interpretation, we use identification IC1. More specifically, we rotate the estimated factor loadings $[\hat{D}_{\tau_1}, \hat{D}_{\tau_2}]$ in such a way that the upper 2×2 submatrix is identical to the parametric estimates. Note that fix an $r \times r$ block of the factor loading matrix to any given matrix (not necessarily an identity matrix) is equivalent to IC1. For comparison, we also compute the PC estimate of (14).

The following two figures depict the estimated slope and curvature factors by the three methods. The level factor is not shown since the three methods all estimate the level factor as the sample mean over τ.

Figure 1 shows that the three different methods give similar estimates for the slope factor; the QML method and the parametric method are especially close. Figure 1 also shows that the slope factors are mostly negative over the sample period and they experience dramatic swings during 1990-1995 and 2001-2006.

Figure 2 displays the estimated curvature factor by the three methods. Although the estimates of the curvature factor by the QML and the parametric methods do not match so well as for the slope factor, they are not far apart. As a comparison, the estimates by the PC method show noticeable departures (scaled by 0.5 as in Diebold and Li). Nevertheless, Figure 2 shows that the three methods identify similar turning points for the rise and fall in the curvature factor. Taking together, the nonparametrically estimated yield curves appear to support the Nelson-Siegel parametric model.

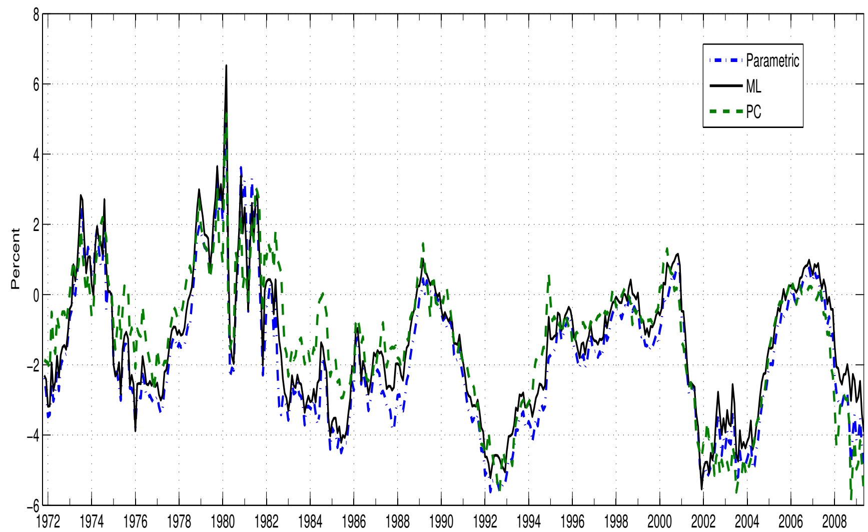

Figure 1: Estimates of the slope factor by three different methods

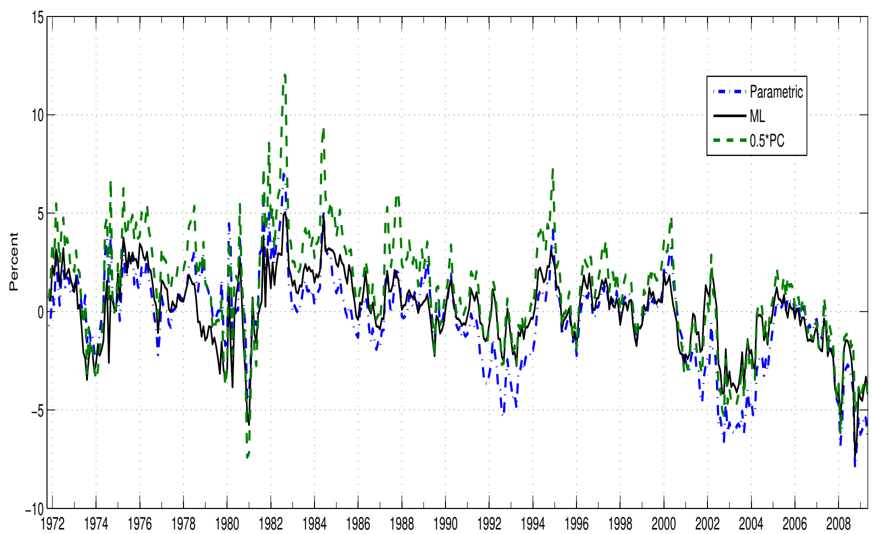

Figure 2: Estimates of the curvature factor by three different methods

Table 3: Definition of symbols in the limiting distributions:

Γ_j^λ	$=$	$\lim\limits_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\left[(\lambda_j'\otimes f_t)E(\xi_t\xi_s')(\lambda_j\otimes f_s') + (\lambda_j'\otimes f_t)E(\xi_t e_{js})f_s' + f_tE(e_{jt}\xi_s')(\lambda_j\otimes f_s') + f_tE(e_{jt}e_{js})f_s'\right]$
Υ_j^λ	$=$	$\lim\limits_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T f_tf_s'\rho_{j,ts}$
Π_j^λ	$=$	$\lim\limits_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\left(E(\mathcal{P}_t\lambda_j\lambda_j'\mathcal{P}_s') + M_{ff}^{-1}f_t\lambda_j'E(e_{jt}\mathcal{P}_s) + E(\mathcal{P}_te_{js})\lambda_jf_s'M_{ff}^{-1} + M_{ff}^{-1}f_tf_s'M_{ff}^{-1}\rho_{j,ts}\right)$
Ψ_j^λ	$=$	$\lim\limits_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\left(E(\mathcal{Q}_t\lambda_j\lambda_j'\mathcal{Q}_s') + f_t\lambda_j'E(e_{jt}\mathcal{Q}_s) + E(\mathcal{Q}_te_{js})\lambda_jf_s' + f_tf_s'\rho_{j,ts}\right)$

Γ^M	$=$	$\lim\limits_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T[f_tf_s'\otimes E(\xi_t\xi_s')]$
Π^M	$=$	$\lim\limits_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T[(\Lambda_1^{-1}E(\xi_t\xi_s')\Lambda_1^{-1'})\otimes(f_tf_s')]$

Γ_t^f	$=$	$\Delta\lim\limits_{T\to\infty}\frac{1}{T}\sum_{s=1}^T\sum_{u=1}^T(f_s'M_{ff}^{-1}f_t)(f_u'M_{ff}^{-1}f_t)E(\xi_s\xi_u') + Q^{-1}\left(\lim\limits_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\frac{1}{\phi_i^2\phi_j^2}\lambda_i\lambda_j'\gamma_{ij,t}\right)Q^{-1}$
Υ_t^f	$=$	$Q^{-1}(\lim\limits_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\frac{1}{\phi_i^2\phi_j^2}\lambda_i\lambda_j'\gamma_{ij,t})Q^{-1}$
Π_t^f	$=$	$\Delta(I_r\otimes f_t')\left(\lim\limits_{T\to\infty}\frac{1}{T}\sum_{s=1}^T\sum_{u=1}^T E[\mathrm{vec}(\mathcal{P}_s)\mathrm{vec}(\mathcal{P}_u)']\right)(I_r\otimes f_t) + Q^{-1}\left(\lim\limits_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\frac{1}{\phi_i^2\phi_j^2}\lambda_i\lambda_j'\gamma_{ij,t}\right)Q^{-1}$
Ψ_t^f	$=$	$\Delta(I_r\otimes f_t')\left(\lim\limits_{T\to\infty}\frac{1}{T}\sum_{s=1}^T\sum_{u=1}^T E[\mathrm{vec}(\mathcal{Q}_s)\mathrm{vec}(\mathcal{Q}_u)']\right)(I_r\otimes f_t) + Q^{-1}\left(\lim\limits_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\frac{1}{\phi_i^2\phi_j^2}\lambda_i\lambda_j'\gamma_{ij,t}\right)Q^{-1}$

8 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.

References

Amemiya Y., W. A. Fuller and S. G. Pantula (1987) The asymptotic distributions of some estimators for a factor analysis model, *Journal of Multivariate Analysis*, **22:1**, 51–64.

Anderson, T. W. (2003) *An Introduction to Multivariate Statistical Analysis*, John Wily & Sons.

Anderson, T. W. and Y. Amemiya (1988) The asymptotic normal distribution of estimators in factor analysis under general conditions, *The Annals of Statistics*, **16:2**, 759–771.

Anderson, T. W. and H. Rubin (1956) Statistical inference in factor analysis, In *Proceedings of the third Berkeley Symposium on mathematical statistics and probability: contributions to the theory of statistics*, University of California Press.

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

Bai, J. and K. Li (2012) Statistical analysis of factor models of high dimension, *The Annals of Statistics*, **40:1**, 436–465.

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

Bernanke, B. S. and J. Boivin (2003) Monetary policy in a data-rich environment, *Journal of Monetary Economics*, **50:3**, 525–546

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

Breitung, J. and J. Tenhofen (2011) GLS estimation of dynamic factor models, *Journal of the American Statistical Association*, 106, 1150-1156.

Chamberlain, G. and M. Rothschild (1983) Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica*, **51:5**, 1281–1304.

Choi, I. (2007) Efficient Estimation of Factor Models, *Econometric theory*, forthcoming.

Connor, G. and R. A. Korajczyk (1988) Risk and return in an equilibrium APT: Application of a new test methodology, *Journal of Financial Economics*, **21:2**, 255–289.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society,* Series B, 39, 1-38.

Diebold, F and C. Li (2006) Forecasting the term structure of government bond yields, *Journal of Econometrics*, **130:2**, 337–364.

Diebold, F, G. Rudebusch and B. Aruoba (2006) The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics*, 131, 309–338.

Doz, C., D. Giannone, and L. Reichlin (2011a), A qausi-maximum likelihood approach for large approximate dynamic factor models, *Review of economics and statistics*, forthcoming.

Doz, C., D. Giannone, and L. Reichlin (2011b), A Two-Step estimator for large approximate dynamic factor models based on Kalman filtering, *Journal of Econometrics*, 164:1, 188–205.

Fan, J., Y. Liao, and M. Mincheva (2011). High Dimensional Covariance Matrix Estimation in Approximate Factor Models. *Annals of Statistics*, 39, 3320-3356.

Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000), The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4): 540-554.

Goyal, A. C. Perignon, C. Villa (2008), How common are common return factors across the NYSE and Nasdaq, *Journal of Financial Economics,* 252-271.

Inoue, A. and X. Han (2011) Tests for Parameter Instability in Dynamic Factor Models. Department of Economics, North Carolina State University.

Jenrich, R. I. (1969) Asymptotic properties of non-linear least squares estimators, *The Annals of Mathematical Statistics*, **40(2)**:633–643.

Jungbacker, B. and S.J. Koopman (2008), Likelihood-based analysis for dynamic factor models, unpublished manuscript, Tinbergen Institute.

Kose, M. A. and C. Otrok, and C.H. Whiteman (2003), International business cycles: World, region, and country-specific factors, *American Economic Review*, **93(4)**: 1216–1239.

Lawley D. N. and A. E. Maxwell (1971), *Factor Analysis as a Statistical Method*, New York: American Elsevier Publishing Company.

Meng, X.L. and D.B. Rubin (1993), Maximum likelihood estimation via the ECM algorithm: a genenral framework, *Biometrika*, **80(2)**, 267–278.

Nelson, C.R. and A.F. Siegel (1987) Parsimonious modeling of yield curves, *Journal of Business*, **60(4)**, 473–489.

Newey, W.K. and D. McFadden (1994) Large sample estimation and hypothesis testing, *Handbook of econometrics*, Vol.4, 2111–2245

Quah, D. and T. Sargent (1992). A dynamic index model for large cross-section. Federal Reserve Bank of Minneapolis, Discussion Paper 77.

Ross, S. A. (1976) The arbitrage theory of capital asset pricing, *Journal of Economic Theory*,**13:3**, 341–360.

Stock, J. H. and M. W. Watson (2002a) Macroeconomic forecasting using diffusion indexes" *Journal of Business and Economic Statistics*, 20, 147-162.

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

Wang, P. (2010) Large dimensional factor models with a multi-level factor structure: identification, estimation, and inference. Department of Economics, HKUST.

Watson, M.W. and R.F. Engle. (1983) Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400, 1983.

Wu, C.F.J. (1981) Asymptotic theory of nonlinear least squares estimation, *The Annals of Statistics*, 501–513.

Wu, C.F.J. (1983) On the convergence properties of the EM algorithm, *The Annals of Statistics*, 95–103.

Supplement: The detailed proofs for the propositions, theorems and corollaries in the main text

Appendix A: Consistency and its proof

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

Proposition A.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*

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\|\hat{\lambda}_i - \lambda_i\|^2 \xrightarrow{p} 0$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2 - \phi_i^2)^2 \xrightarrow{p} 0$$

$$\hat{M}_{ff} - M_{ff} \xrightarrow{p} 0$$

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

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 $i = 1, 2, \cdots, N$,

$$E\Big(\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t e_{it}\Big\|^2\Big) \le C \tag{A.1}$$

$$E\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t e_{it}\Big\|^2\Big) \le C \tag{A.2}$$

Furthermore, under Assumption C.5, we have, by taking $i = j$,

$$E\Big[\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(e_{it}^2 - \phi_i^2)\Big|^2\Big] \le C \tag{A.3}$$

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 A.1 *Let Q be an $r \times r$ matrix satisfying*

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

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 A.1: See Bai and Li (2012). $\square$

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

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

$$(a) \quad \sup_{\theta \in \Theta} \frac{1}{NT} tr\Big[\Lambda^{*\prime}\Sigma_{zz}^{-1}\sum_{t=1}^{T} e_t f_t^{*\prime}\Big] \xrightarrow{p} 0$$

$$(b) \quad \sup_{\theta \in \Theta} \frac{1}{NT} tr\Big[\sum_{t=1}^{T}(e_t e_t' - \Omega_t^*)\Sigma_{zz}^{-1}\Big] \xrightarrow{p} 0$$

$$(c) \quad \sup_{\theta \in \Theta} \frac{1}{N} tr\Big[\bar{e}\bar{e}'\Sigma_{zz}^{-1}\Big] \xrightarrow{p} 0$$

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

Proof of Lemma A.2: Notice that

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

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

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big)^2 = O_p(T^{-1}).$$

The first result follows by (A.2). The second result follows by (A.3). The third result is implied by Assumption C.5.

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

Lemma A.3 *Under Assumptions A-D, for $\theta = (\Lambda, \Phi, M_{ff})$, we have*

$$(a) \quad \sup_{\theta \in \Theta} \frac{1}{N} tr\Big[\frac{1}{T}\sum_{t=1}^{T}\Omega_t^*\Phi^{-1}\Lambda G\Lambda'\Phi^{-1}\Big] = O_p(N^{-1}) = o_p(1)$$

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

Proof of Lemma A.3: Consider (a). The left hand side of (a) can be written as $\frac{1}{N}tr[\Lambda'\Phi^{-1}\frac{1}{T}\sum_{t=1}^{T}\Omega_t^*\Phi^{-1}\Lambda G]$, which, by the definition of Ω_t^*, is equivalent to

$$\frac{1}{N}tr\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2\phi_j^2}H^{1/2}\lambda_i\lambda_j'H^{1/2}\frac{1}{T}\sum_{t=1}^{T}\tau_{ij,t}(H^{1/2}M_{ff}^{-1}H^{1/2}+I_r)^{-1}\Big].$$

Consider the term $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2\phi_j^2}H^{1/2}\lambda_i\lambda_j'H^{1/2}\frac{1}{T}\sum_{t=1}^{T}\tau_{ij,t}$, which is bounded in norm by

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{\phi_i^2}H^{1/2}\lambda_i\Big\|\cdot\Big\|\frac{1}{\phi_j^2}\lambda_j'H^{1/2}\Big\|\cdot\Big|\frac{1}{T}\sum_{t=1}^{T}\tau_{ij,t}\Big|.$$

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

$$C^2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{\phi_i}H^{1/2}\lambda_i\Big\|\cdot\Big\|\frac{1}{\phi_j}\lambda_j'H^{1/2}\Big\|\tau_{ij}.$$

Let $\chi_i=\|\frac{1}{\phi_i}H^{1/2}\lambda_i\|$ and $\chi=(\chi_1,\chi_2,\cdots,\chi_N)'$, the above term is equal to $\frac{1}{N}C^2\chi'\mathcal{T}\chi$ 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 $\mathcal{T}$ is a $N\times N$ matrix consisting of τ_{ij}. So the above term is bounded by $C^2r\frac{1}{N}\tau_{max}$, where τ_{max} is the largest eigenvalue of the matrix $\mathcal{T}$. By Assumption C.3, $\tau_{max}\leq C$. Then (a) follows.

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

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

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

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

$$(a)\quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda^*(\frac{1}{T}\sum_{t=1}^{T}f_t^*e_{jt})=\|N^{1/2}\hat{H}^{1/2}\|\cdot O_p(T^{-1/2}),\ \text{for each } j$$

$$(b)\quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\frac{1}{T}\sum_{t=1}^{T}e_tf_t^{*\prime})=\|N^{1/2}\hat{H}^{1/2}\|\cdot O_p(T^{-1/2})$$

$$(c)\quad \hat{H}\Big(\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)=\|N^{1/2}\hat{H}^{1/2}\|\cdot O_p(T^{-1/2}),\ \text{for each } j$$

$$(d)\quad \hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}\hat{\lambda}_i\hat{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)\hat{H}=\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(T^{-1/2})$$

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

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

$$(a)\quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1})$$

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

$$(c)\quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}_j = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(T^{-1}), \quad \text{for each } j$$

$$(d)\quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi} - \frac{1}{T}\sum_{t=1}^{T}\Omega_t^*)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = \|\hat{H}\| \cdot O_p(1) + \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(N^{-1})$$

Proof of Lemma A.5: Consider (a). The left hand side of (a) is bounded in norm by

$$C^2\|\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\|\hat{H}^{1/2}\hat{\lambda}_j\|^2\Big)\Big(\sum_{i=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}e_{it}\Big)^2\Big)$$

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

$$C^2 r\|N^{1/2}\hat{H}^{1/2}\|^2\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}\right)^2$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2 O_p(T^{-1})$ 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,

$$\sum_{i=1}^{N}\|\frac{1}{\hat{\phi}_i}\hat{H}^{1/2}\hat{\lambda}_i\|\tau_{ij} \le \Big(\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\|\hat{H}^{1/2}\hat{\lambda}_i\|^2\Big)^{1/2}\Big(\sum_{i=1}^{N}\tau_{ij}^2\Big)^{1/2} = \sqrt{r}\Big(\sum_{i=1}^{N}\tau_{ij}^2\Big)^{1/2}.$$

However, $\sum_{i=1}^{N}\tau_{ij}^2 \le C\sum_{i=1}^{N}\tau_{ij} \le C^2$ because $\tau_{ij} \le C$ and $\sum_{i=1}^{N}\tau_{ij} \le C$. Given this result, the above expression is $O(1)$. Then (b) follows.

Consider (c). By (a), it follows that $\|\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\| = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(T^{-1/2})$. So (c) follows by $\bar{e}_j = O_p(T^{-1/2})$ 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 $\|N^{1/2}\hat{H}^{1/2}\|^2$ $O_p(N^{-1})$, similarly as result (a) of Lemma A.3. □

Proof of Proposition A.1: By $z_t = \alpha^* + \Lambda^* f_t^* + e_t$, it follows that

$$
\begin{aligned}
M_{zz} = \Lambda^* M_{ff}^* \Lambda^{*\prime} + \Phi^* + \frac{1}{T}\sum_{t=1}^T \Lambda^* f_t^* e_t' + \frac{1}{T}\sum_{t=1}^T e_t f_t^{*\prime} \Lambda^{*\prime} \\
+ \frac{1}{T}\sum_{t=1}^T (e_t e_t' - \Omega_t^*) + (\frac{1}{T}\sum_{t=1}^T \Omega_t^* - \Phi^*) - \bar{e}\bar{e}'
\end{aligned}
\tag{A.4}
$$

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

$$
\overline{L}(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N} tr[\Sigma_{zz}(\theta^*)\Sigma_{zz}^{-1}]
$$

$$
R_1(\theta) = -\frac{1}{2N} tr\Big[\Big(\frac{1}{T}\sum_{t=1}^T \Lambda^* f_t^* e_t' + \frac{1}{T}\sum_{t=1}^T e_t f_t^{*\prime}\Lambda^{*\prime} + \frac{1}{T}\sum_{t=1}^T (e_t e_t' - \Omega_t^*)\Big)\Sigma_{zz}^{-1}\Big]
$$

$$
R_2(\theta) = -\frac{1}{2N} tr\Big[\Big((\frac{1}{T}\sum_{t=1}^T \Omega_t^* - \Phi^*) - \bar{e}\bar{e}'\Big)\Sigma_{zz}^{-1}\Big]
$$

Then the likelihood function can be written as

$$
L(\theta) = \overline{L}(\theta) + R(\theta)
$$

where $R(\theta) = R_1(\theta) + R_2(\theta)$. Lemma A.2 and Lemma A.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

$$
\frac{1}{N}\sum_{i=1}^N (\hat{\phi}_i^2 - \phi_i^{*2})^2 \xrightarrow{p} 0 \tag{A.5}
$$

$$
\hat{G} = o_p(1); \qquad \hat{H} = o_p(1) \tag{A.6}
$$

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

$$
\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^* - (I_r - A)\Big(\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda}\Big)(I_r - A)' \xrightarrow{p} 0 \tag{A.7}
$$

and

$$
\frac{1}{N}(\hat{\Lambda} - \Lambda^*)'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda^*) - A\Big(\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda}\Big)A' \xrightarrow{p} 0 \tag{A.8}
$$

Now we turn to the first order conditions. The jth column of the first order condition (4) implies

$$
\begin{aligned}
\hat{\lambda}_j - \lambda_j^* = -\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda^*)M_{ff}^*\lambda_j^* - \hat{M}_{ff}^{-1}(\hat{M}_{ff} - M_{ff}^*)\lambda_j^* \\
+\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda^*\frac{1}{T}\sum_{t=1}^T f_t^* e_{jt} + \hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^T e_t f_t^{*\prime}\lambda_j^*
\end{aligned}
$$

$$+\hat{M}_{ff}^{-1}\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]-\hat{M}_{ff}^{-1}\hat{H}\hat{\lambda}_j \tag{A.9}$$

$$+\hat{M}_{ff}^{-1}\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_iE(e_{it}e_{jt})-\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}_j$$

The first order condition for M_{ff} in (5) implies

$$\begin{aligned}\hat{M}_{ff}-M_{ff}^* &= -\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda^*)M_{ff}^*-M_{ff}^*(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \\ &+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda^*)M_{ff}^*(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \\ &+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda^*\frac{1}{T}\sum_{t=1}^{T}f_t^*e_t'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t^{*\prime}\Lambda^{*\prime}\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \\ &+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Omega_t^*)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi}-\frac{1}{T}\sum_{t=1}^{T}\Omega_t^*)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\end{aligned} \tag{A.10}$$

Substituting (A.10) into (A.9), we have

$$\begin{aligned}\hat{\lambda}_j-\lambda_j^* &= \hat{M}_{ff}^{-1}M_{ff}^*(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j^*-\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda^*)M_{ff}^*(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j^* \\ &-\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda^*(\frac{1}{T}\sum_{t=1}^{T}f_t^*e_t')\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j^*-\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\frac{1}{T}\sum_{t=1}^{T}e_tf_t^{*\prime})\Lambda^{*\prime}\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j^* \\ &-\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Omega_t^*)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j^*+\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}'\Phi^{-1}\hat{\Lambda}\hat{H}\lambda_j^* \\ &+\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi}-\frac{1}{T}\sum_{t=1}^{T}\Omega_t^*)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j^*+\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_t^*e_{jt} \\ &+\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t^{*\prime}\lambda_j^*+\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{jt}-E(e_te_{jt})] \\ &-\hat{M}_{ff}^{-1}\hat{H}\hat{\lambda}_j+\hat{M}_{ff}^{-1}\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_iE(e_{it}e_{jt})-\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}_j\end{aligned} \tag{A.11}$$

Consider (A.10). The sixth term on the right of (A.10) can be written as

$$\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t^*-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t^*A$$

where $A=(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}$. The first term of the above is $\|N^{1/2}\hat{H}^{1/2}\|\cdot O_p(T^{-1/2})$ by Lemma A.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 (A.10) is the transpose of the sixth. The last term is governed by Lemma A.5(d). The fourth term is governed by Lemma A.5(a). These results together with (A.6) imply that, in terms of A,

$$\hat{M}_{ff}-M_{ff}^*=-A'M_{ff}^*-M_{ff}^*A+A'M_{ff}^*A-A\cdot\|N^{1/2}\hat{H}^{1/2}\|\cdot O_p(T^{-1/2}) \tag{A.12}$$

$$+\|N^{1/2}\hat{H}^{1/2}\|\cdot O_p(T^{-1/2})+\|N^{1/2}\hat{H}^{1/2}\|^2\cdot[O_p(T^{-1/2})+O_p(N^{-1})]+o_p(1)$$

By the definition of $\hat{H}$, $N\hat{H}=(\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}$. Equation (A.7) implies $(\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}=(I_r-A)'(\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*)^{-1}(I_r-A)+o_p(\|I_r-A\|^2)$. So we have

$$\|N^{1/2}\hat{H}^{1/2}\|^2=tr[N\hat{H}]=tr\Big[(I_r-A)'(\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*)^{-1}(I_r-A)+o_p(\|I_r-A\|^2)\Big]$$

These results imply that matrix A is stochastically bounded. To see this, the left hand side of (A.12) 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 $A=O_p(1)$, and hence $\|N^{1/2}\hat{H}^{1/2}\|=O_p(1)$ by the preceding equation. From this, we have, by (A.12),

$$\hat{M}_{ff}-M_{ff}^*=-A'M_{ff}^*-M_{ff}^*A+A'M_{ff}^*A+o_p(1) \tag{A.13}$$

Next consider (A.11). The last two terms are all $o_p(1)$ by Lemma A.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 $\sqrt{r}C^2\|\hat{H}^{1/2}\|$ by $\sum_{i=1}^N\|\frac{1}{\hat{\phi}_i}$ $\hat{H}^{1/2}\hat{\lambda}_i\|^2=r$. So the third from the last term is $o_p(1)$ by (A.6). The 3rd-10th terms are summarized in Lemmas A.4 and A.5 and they are all $o_p(1)$ due to $\|N^{1/2}\hat{H}^{1/2}\|=O_p(1)$. Thus we can express (A.11) as

$$\hat{\lambda}_j-\lambda_j^*=\hat{M}_{ff}^{-1}M_{ff}^*A\lambda_j^*-\hat{M}_{ff}^{-1}A'M_{ff}^*A\lambda_j^*+o_p(1) \tag{A.14}$$

Results (A.13) and (A.14), together with the identification conditions, imply $A=(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\xrightarrow{p}0$, as is shown by Bai and Li (2012). With $A\xrightarrow{p}0$, equation (A.8) implies $\frac{1}{N}(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda^*)=o_p(1)$, which is the first part of Proposition A.1. Moreover, (A.13) implies that $\hat{M}_{ff}-M_{ff}=o_p(1)$, which is the last part of Proposition A.1. This completes the proof of the proposition. □

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

(a) $\quad \dfrac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda}-\dfrac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*=o_p(1)$

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

(c) $\quad \dfrac{1}{N}(\hat{\Lambda}-\Lambda^*)'\hat{\Phi}^{-1}\hat{\Lambda}=o_p(1)$

Proof of Corollary A.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 (A.7). Result (a) implies that $\hat{H}=O_p(N^{-1})$ since $N^{-1}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*\to Q>0$ 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 $N\hat{H}=(\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*)^{-1}+o_p(1)$ by part (a). □

Appendix 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 B.1 *Under Assumptions A-D,*

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

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

Lemma B.2 *Under Assumptions A-D:*

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

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

$$(c)\quad \Big\|\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}\hat{\lambda}_i\hat{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)\hat{H}\Big\|^2 = O_p\left(T^{-1}\right)$$

Lemma B.3 *Under Assumptions A-D:*

$$(a)\quad \Big\|\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_t\xi_t'\Big\|^2 = O_p(T^{-1})$$

$$(b)\quad \Big\|\hat{H}\Big(\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\frac{1}{T}\sum_{t=1}^{T}[e_{it}\xi_t'-E(e_{it}\xi_t')]\Big)\Big\|^2 = O_p(T^{-1})$$

where $\xi_t' = (e_{1t}, e_{2t}, \cdots e_{rt})$.

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 B.1 given below.

Lemma B.4 *Let* $\mathcal{E} = \left[\frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\phi}_i^2}\|\hat{\lambda}_i - \lambda_i\|^2\right]^{1/2}$. *Under Assumptions A-D, we have*

$$(a) \quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi} - \frac{1}{T}\sum_{t=1}^T \Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = O_p(N^{-1})$$

$$(b) \quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = O_p(T^{-1})$$

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

$$(d) \quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{\xi}' = O_p(T^{-1})$$

$$(e) \quad \hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i E(e_{it}e_{jt}) = O_p(N^{-1}) + \mathcal{E}\cdot O_p(N^{-1/2}) \quad \text{for any } j$$

$$(f) \quad \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}_j = O_p(T^{-1}) \quad \text{for any } j$$

where ξ_t *is defined in Lemma B.3.*

Proof of Lemma B.4: Part (a) is a direct result of Lemma A.5(d) and Corollary A.1(b). Part (b) is a direct result of Lemma A.5(a) and Corollary A.1(b).

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

$$\hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\phi}_i^2}\lambda_i E(e_{it}\xi_t') + \hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\phi}_i^2}(\hat{\lambda}_i - \lambda_i)E(e_{it}\xi_t') = I_1 + I_2 \quad \text{say}$$

Consider I_1, which is bounded in norm by

$$\|\hat{H}\|\Big(\max_{i\le N}\|\frac{1}{\hat{\phi}_i^2}\lambda_i\|\Big)\sum_{i=1}^N\frac{1}{T}\sum_{t=1}^T\|E(e_{it}\xi_t')\| \le C^3\|\hat{H}\|\sum_{i=1}^N\frac{1}{T}\sum_{t=1}^T\|E(e_{it}\xi_t')\|$$

where $\xi_t = (e_{1t}, e_{2t}, \cdots, e_{rt})$. For any $j \le r$, by Assumption C.3, we have

$$\sum_{i=1}^N\frac{1}{T}\sum_{t=1}^T|E(e_{it}e_{jt})| \le \sum_{i=1}^N\frac{1}{T}\sum_{t=1}^T|\tau_{ij,t}| \le \sum_{i=1}^N \tau_{ij} \le C$$

So the term $\sum_{i=1}^N\frac{1}{T}\sum_{t=1}^T\|E(e_{it}\xi_t')\|$ is bounded by $\sqrt{r}C$. Given this result, we have $I_1 = O_p(N^{-1})$ by Corollary A.1(b).

Consider I_2. I_2 is bounded in norm by

$$C\|\hat{H}_N\|\Big(\frac{1}{N}\sum_{i=1}^N\frac{1}{\hat{\phi}_i^2}\|\hat{\lambda}_i - \lambda_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{T}\sum_{t=1}^T E(e_{it}\xi_t')\Big\|^2\Big)^{1/2}$$

Noting $\xi_t = (e_{1t}, e_{2t}, ..., e_{rt})'$. For any $j \le r$,

$$\frac{1}{N}\sum_{i=1}^N\Big(\frac{1}{T}\sum_{t=1}^T E(e_{it}e_{jt})\Big)^2 \le \frac{1}{N}\sum_{i=1}^N\tau_{ij}^2 \le \frac{1}{N}\sup_{i\le N}|\tau_{ii}|\sum_{i=1}^N|\tau_{ij}| \le N^{-1}C$$

Thus, $I_2 = \mathcal{E}\cdot O_p(N^{-1/2})$, and (c) follows.

Part (d) is a direct result of Lemma A.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 B.1 *Under Assumptions A-D, irrespective which set of identification conditions,*

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

where $\mathcal{E}$ is defined in Lemma B.4.

Proof of Proposition B.1: 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 (A.11) are zero. So we have

$$\begin{aligned}M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}&=\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\\&+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\\&+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi}-\frac{1}{T}\sum_{t=1}^{T}\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\\&-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}'\Phi^{-1}\hat{\Lambda}\hat{H}-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_t\xi_t'-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\\&-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t\xi_t'-E(e_t\xi_t')]+\hat{H}-\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_iE(e_{it}\xi_t')+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{\xi}'\end{aligned}\tag{B.1}$$

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 B.1(b) and Corollary A.1(c). The 4th term is $O_p(T^{-1/2})$ by Lemma B.2(c). The 5th and 6th term are $O_p(N^{-1})$ and $O_p(T^{-1})$ by Lemma B.4(a) and (b). The 7th term is $O_p(T^{-1/2})$ by Corollary A.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 B.3(b). The last two terms are $O_p(N^{-1})+O_p(T^{-1})+o_p(\mathcal{E})$ by Lemma B.4(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

$$\begin{aligned}&\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}+\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\\&\qquad=-\frac{1}{N}\Lambda'(\hat{\Phi}^{-1}-\Phi^{-1})\Lambda+\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\end{aligned}\tag{B.2}$$

The first term on right hand side of the above equation is $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2\phi_i^2}(\hat{\phi}_i^2-\phi_i^2)\lambda_i\lambda_i'$, which is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^4\phi_i^4}\|\lambda_i\|^4\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2-\phi_i^2)^2\Big)^{1/2}\le C^6\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2-\phi_i^2)^2\Big)^{1/2}$$

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

$$\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}+\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)'=O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2-\phi_i^2)^2\Big]^{1/2}\Big)+o_p(\mathcal{E}) \quad \text{(B.3)}$$

Consider (A.10). Since both $\hat{M}_{ff}$ and M_{ff} are diagonal matrices, we have

$$\begin{aligned}
&\text{Ndiag}\Big\{\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}+M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big\}\\
&=\text{Ndiag}\Big\{\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\\
&\quad+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \qquad \text{(B.4)}\\
&\quad-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi}-\frac{1}{T}\sum_{t=1}^{T}\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big\}
\end{aligned}$$

where Ndiag denotes the off-diagonal elements. Following the discussion after equation (B.1), the right hand side of the above equation is $O_p(T^{-1/2})+O_p(N^{-1})$. Thus equation (B.4) can be written as

$$\text{Ndiag}\Big\{\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}+M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big\}=O_p(T^{-1/2})+O_p(N^{-1}) \quad \text{(B.5)}$$

Note that under IC2, $\hat{H}=\frac{1}{N}I_r$, thus both (B.3) and (B.5) put restrictions on $\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}$. Equation (B.3) puts $\frac{1}{2}r(r+1)$ restrictions, while (B.5) 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 (B.3) and (B.5) we obtain,

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

Under IC3: The proof of Proposition B.1 under IC3 is quite similar to the case of IC2. The details are omitted; also see, Bai and Li (2012).

Under IC4: Consider (A.11). Pre-multiplying $\hat{M}_{ff}$ on both sides, the first r equations can be written as

$$\begin{aligned}
\hat{M}_{ff}(\hat{\Lambda}_1'-\Lambda_1')&=M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda_1'-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda_1'\\
&\quad-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda_1'-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda_1' \qquad \text{(B.6)}\\
&\quad-\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda_1'+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi}-\frac{1}{T}\sum_{t=1}^{T}\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Lambda_1'\\
&\quad+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}'\Phi^{-1}\hat{\Lambda}\hat{H}\Lambda_1'+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_t\xi_t'+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda_1'
\end{aligned}$$

$$+\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t\xi_t' - E(e_t\xi_t')] - \hat{H}\hat{\Lambda}_1' + \hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i E(e_{it}\xi_t') - \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{\xi}'$$

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 A.1(b) and the latter one is of a smaller order term than $\hat{M}_{ff}(\hat{\Lambda}_1' - \Lambda_1')$ by $\hat{M}_{ff} \xrightarrow{p} M_{ff}$. So this term is $O_p(N^{-1})$. Given this result, following the discussion after equation (B.1), 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 (B.6) 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 ${\Lambda'}_1^{-1}$ on each side since the latter matrix is also upper triangular. It follows that

$$\text{nonupper}\Big\{M_{ff}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big\} = O_p(T^{-1/2}) + O_p(N^{-1}) + o_p(\mathcal{E}) \qquad \text{(B.7)}$$

where **nonupper** means lower triangular elements plus diagonal ones. The above equation has $\frac{1}{2}r(r+1)$ restrictions. But equation (B.5), 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 (B.5) and (B.7). Then we obtain

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

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 B.1.□

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

Lemma B.5 *Under Assumptions A-D,*

(a) $\displaystyle \frac{1}{N}\sum_{j=1}^{N}\Big\|\lambda_j'\hat{H}\Big(\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big)\Big\|^2 = O_p(T^{-1})$

(b) $\displaystyle \frac{1}{N}\sum_{j=1}^{N}\Big\|\lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda)\frac{1}{T}\sum_{t=1}^{T}f_t e_{jt}\Big\|^2 = o_p(T^{-1})$

(c) $\displaystyle \frac{1}{N}\sum_{j=1}^{N}\Big\|\lambda_j'\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i E(e_{it}e_{jt})\Big\|^2 = O_p(N^{-2}) + \frac{1}{N}O_p(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\Big\|\hat{\lambda}_i - \lambda_i\Big\|^2)$

(d) $\displaystyle \frac{1}{N}\sum_{j=1}^{N}\Big\|\lambda_j'\hat{H}\hat{\Phi}^{-1}\bar{e}\bar{e}_j\Big\|^2 = O_p(T^{-2})$

Proof of Lemma B.5: 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 $\|\lambda_j'\hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i E(e_{it}e_{jt})\|$ is bounded by

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

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

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

By the boundedness of $\lambda_i, \hat{\phi}_i^2$, the first term of the above is bounded by $2C^8\|\hat{H}\|^2\frac{1}{N}\sum_{j=1}^N$ $(\sum_{i=1}^N\tau_{ij})^2$. So the first term is $O_p(N^{-2})$ by $\sum_{i=1}^N\tau_{ij}<C$ for all j. By

$$\Big\|\lambda_j'\hat{H}\sum_{i=1}^N\frac{1}{\hat{\phi}_i^2}(\hat{\lambda}_i-\lambda_i)\frac{1}{T}\sum_{t=1}^T\tau_{ij,t}\Big\|^2\le C^4\|\hat{H}_N\|^2\Big(\frac{1}{N^2}\sum_{i=1}^N\frac{1}{\hat{\phi}_i^2}\Big\|\hat{\lambda}_i-\lambda_i\Big\|^2\Big)\Big(\sum_{i=1}^N\tau_{ij}^2\Big)$$

because $\tau_{ij,t}\le\tau_{ij}$ for all t. Note that $\sum_{i=1}^N\tau_{ij}^2$ is bounded, thus (c) follows.

Consider (d). The left hand side of (d) is equal to $\|\hat{H}\hat{\Phi}^{-1}\bar{e}\|^2\frac{1}{N}\sum_{j=1}^N\|\lambda_j\bar{e}_j\|^2$. Since $\|\hat{H}\hat{\Phi}^{-1}\bar{e}\|^2=O_p(T^{-1})$ by Lemma B.4(c) and $\frac{1}{N}\sum_{j=1}^N\|\lambda_j\bar{e}_j\|^2=O_p(T^{-1})$, (d) follows. □

Proof of Theorem 1: We begin with the first order condition on diag$\{\Phi\}$. By the same method in deducing (A.9) and (A.10), we have

$$\hat{\phi}_j^2-\phi_j^2=\frac{1}{T}\sum_{t=1}^T(e_{jt}^2-\phi_j^2)-(\hat{\lambda}_j-\lambda_j)'\hat{M}_{ff}(\hat{\lambda}_j-\lambda_j) \tag{B.8}$$

$$+\lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j+2\lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^T f_te_t'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_j$$

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

$$-\lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Phi}-\frac{1}{T}\sum_{t=1}^T\Omega_t)\hat{\Phi}^{-1}\hat{\Lambda}'\hat{H}\lambda_j-2\lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^T e_tf_t'\lambda_j$$

$$+2\lambda_j'\hat{H}\hat{\lambda}_j-2\lambda_j'\hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_iE(e_{it}e_{jt})-2\lambda_j'\hat{H}\Big(\sum_{i=1}^N\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\frac{1}{T}\sum_{t=1}^T[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)$$

$$+2\lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^T f_te_{jt}+2\lambda_j'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}_j=a_{1,j}+a_{2,j}+\cdots+a_{13,j}\quad\text{say}$$

By the Cauchy-Schwarz inequality, we have

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

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 (A.3). The second term is bounded by $\|\hat{M}_{ff}\| \cdot \frac{1}{N}\sum_{j=1}^{N}\|\hat{\lambda}_j-\lambda_j\|^4$. Using (A.11), 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 $O_p(T^{-2})+O_p(N^{-4})+\frac{1}{N}O_p(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\|\hat{\lambda}_i-\lambda_i\|^2)$. Consider the 3rd term, which is bounded in norm by

$$\Big\|(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big\|^4\Big\|M_{ff}\Big\|^2\frac{1}{N}\sum_{j=1}^{N}\|\lambda_j\|^4$$

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

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

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 $C^4\|\hat{H}\|^2\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_j^2}\|\hat{\lambda}_j-\lambda_j\|^2$. However, $\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_j^2}\|\hat{\lambda}_j-\lambda_j\|^2=o_p(1)$, 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 summarized in Lemma B.5. So we have

$$\frac{1}{N}\sum_{j=1}^{N}(\hat{\phi}_j^2-\phi_j^2)^2 = O_p(T^{-1})+O_p(N^{-2})+o_p\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\Big\|\hat{\lambda}_i-\lambda_i\Big\|^2\Big) \tag{B.9}$$

We next derive bounds involving $\|\hat{\lambda}_j-\lambda_j\|^2$ and $\hat{M}_{ff}-M_{ff}$. Consider (A.11). There are 13 terms on the left hand side of (A.11). We use $b_{1j}, b_{2,j}, \cdots, b_{13,j}$ to denote them. By the Cauchy-Schwarz inequality, $\|b_{1,j}+b_{2,j}+\cdots+b_{13,j}\|^2 \le 13(\|b_{1j}\|^2+\|b_{2,j}\|^2+\cdots+\|b_{13,j}\|^2)$. By this inequality, and noticing $C^{-2}\le\hat{\phi}_j^2\le C^2$, we have

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

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 A.1 and $\frac{1}{N}\sum_{j=1}^{N}\|\lambda_j\|^2=O(1)$ by Assumption B. By Proposition B.1 and negelcing the smaller order term of $\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_j^2}\|\hat{\lambda}_j-\lambda_j\|^2$, we have

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

The 2nd term $\frac{1}{N}\sum_{j=1}^{N}\|b_{2,j}\|^2$ is dominated by the first and is negligible. By Lemmas B.1 and B.2, the 3rd-10th terms are $O_p(T^{-1})+O_p(N^{-2})+O_p(\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2-\phi_i^2)^2)$. 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 B.5. So we have

$$\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_j^2}\Big\|\hat{\lambda}_j-\lambda_j\Big\|^2=O_p(T^{-1})+O_p(N^{-2})+O_p\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2-\phi_i^2)^2\Big) \tag{B.10}$$

Similarly, using Lemmas B.1, B.2 and B.5, we deduce

$$\begin{aligned}\|\hat{M}_{ff}-M_{ff}\|^2=O_p(T^{-1})+O_p(N^{-2})+O_p\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2-\phi_i^2)^2\Big)\\ +o_p\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\Big\|\hat{\lambda}_i-\lambda_i\Big\|^2\Big)\end{aligned} \tag{B.11}$$

Substituting (B.10) into (B.9), 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 (B.10) and (B.11), we obtain the two remaining results of Theorem 1. This completes the proof of Theorem 1. □

Corollary B.1 *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 B.1 is a direct result of Proposition B.1 and Theorem 1.

Appendix C: Proof for the asymptotic representations

Given Assumption E.1, we have

$$E\Big(\Big\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_i^2}\lambda_i f_t' e_{it}\Big\|^2\Big)\le C. \tag{C.1}$$

We need the following lemmas to derive the limiting distributions.

Lemma C.1 *Under Assumptions A-E,*

$$(a)\quad \Big\|\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\Big\| = O_p(T^{-1/2}) + O_p(N^{-1})$$

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

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

$$(d)\quad \Big\|\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}\hat{\lambda}_i\hat{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{ij,t}\Big)\hat{H}\Big\| = O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2})$$

$$(e)\quad \Big\|\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Big\| = \|\hat{H}_N\|\Big\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_i^2}\lambda_i f_t' e_{it}\Big\| + O_p(T^{-1}) + O_p(N^{-1}T^{-1/2}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

where $\varepsilon_{ij,t} = e_{it}e_{jt} - E(e_{it}e_{jt})$ *for notational simplicity.*

Proof of Lemma C.1: Part (a) is implied by Lemma B.1(a) and Theorem 1. It is also implied by Corollary B.1.

Consider (b). The left-hand side of (b) is bounded by

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

$$+\|\hat{H}_N\|\Big\|\frac{1}{NT}\sum_{i=1}^{N}\Big(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2}\Big)\lambda_i\sum_{t=1}^{T}[e_{it}\xi_t' - E(e_{it}\xi_t')]\Big\|$$

$$+\|\hat{H}_N\|\Big\|\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\Big(\hat{\lambda}_i-\lambda_i\Big)\sum_{t=1}^{T}[e_{it}\xi_t' - E(e_{it}\xi_t')]\Big\|$$

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

$$\|\hat{H}_N\|\Big(\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}\lambda_i[e_{it}\xi_t' - E(e_{it}\xi_t')]\Big\|^2\Big)^{1/2}$$

which is further bounded by

$$C^5\|\hat{H}_N\|\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^2-\phi_i^2)^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[e_{it}\xi_t' - E(e_{it}\xi_t')]\Big\|^2\Big)^{1/2}$$

which is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by $E(\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[e_{it}\xi_t' - E(e_{it}\xi_t')]\|^2) \le C$ 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

$$\Big\|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2\phi_j^2}\lambda_i\lambda_j'\sum_{t=1}^{T}\varepsilon_{ij,t}\Big\| + \Big\|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{\hat{\phi}_i^2\phi_j^2} - \frac{1}{\phi_i^2\phi_j^2}\Big)\lambda_i\lambda_j'\sum_{t=1}^{T}\varepsilon_{ij,t}\Big\|$$

$$+\Big\|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2} - \frac{1}{\hat{\phi}_i^2\phi_j^2}\Big)\lambda_i\lambda_j'\sum_{t=1}^{T}\varepsilon_{ij,t}\Big\|$$

$$+\Big\|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}(\hat{\lambda}_i - \lambda_i)\lambda_j'\sum_{t=1}^{T}\varepsilon_{ij,t}\Big\|$$

$$+\Big\|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}\hat{\lambda}_i(\hat{\lambda}_j - \lambda_j)'\sum_{t=1}^{T}\varepsilon_{ij,t}\Big\|$$

The first term is bounded in norm by

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

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

$$\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{\hat{\phi}_i^2 - \phi_i^2}{\hat{\phi}_i^2\phi_i^2\phi_j^2}\lambda_i\lambda_j'\Big\|^2\Big)^{1/2}\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{ij,t}\Big)^2\Big)^{1/2}.$$

The above term is further bounded by

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

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

Using the similar arguments, by (C.1), (e) can be proved and the details are omitted. □

Lemma C.2 *Under Assumptions A-E,*

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\phi}_i^2 - \phi_i^2}{\phi_i^4}\lambda_i\lambda_i' = O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Proof of Lemma C.2: Using (B.8), the expression $\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\phi}_i^2 - \phi_i^2}{\phi_i^4}\lambda_i\lambda_i'$ can be expanded into 13 terms. We consider them one by one. The first term is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_i^4}(e_{it}^2 - \phi_i^2)\lambda_i\lambda_i'$$

which is $O_p(N^{-1/2}T^{-1/2})$ by Assumption E.3. The second term is equal to

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

which is bounded in norm by $C^8\|\hat{M}_{ff}\|\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\|\hat{\lambda}_i-\lambda_i\|^2$, which is $O_p(T^{-1})+O_p(N^{-2})$ by Proposition A.1 and Theorem 1.

Consider the third term, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i\lambda_i'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_i\lambda_i'$$

The above term is bounded in norm by $\|M_{ff}\|\cdot\|(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\|^2\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^4$, which is $O_p(T^{-1})+O_p(N^{-2})$ by Corollary B.1. The 4th-8th terms can be proved to be $O_p(N^{-1})+O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ similarly as the third term.

The 9th term is $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_i'\hat{H}\hat{\lambda}_i\lambda_i\lambda_i'$, which is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_i'\hat{H}\lambda_i\lambda_i\lambda_i'+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_i'\hat{H}(\hat{\lambda}_i-\lambda_i)\lambda_i\lambda_i'=c_1+c_2$$

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

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

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

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

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\lambda_i\lambda_i'(\lambda_i'\hat{H}\lambda_j)}{\phi_i^4\hat{\phi}_j^2}\frac{1}{T}\sum_{t=1}^{T}\tau_{ij,t}+\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^4\hat{\phi}_j^2}\lambda_i\lambda_i'[\lambda_i'\hat{H}(\hat{\lambda}_j-\lambda_j)]\frac{1}{T}\sum_{t=1}^{T}\tau_{ij,t}$$

We use c_3 and c_4 to denote the above two terms. Notice $|\frac{1}{T}\sum_{t=1}^{T}\tau_{ij,t}|\le\tau_{ij}$. By the boundedness of $\lambda_i,\phi_i^2,\hat{\phi}_i^2$, 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

$$C\|\hat{H}\|\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^8}\|\lambda_i\|^6\tau_{ij}\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_j^2}\|\hat{\lambda}_j-\lambda_j\|^2\tau_{ij}\Big)^{1/2}$$

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 $\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^4}\lambda_i\lambda_i'(\lambda_i'\hat{H}\sum_{j=1}^N \frac{1}{\hat{\phi}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t})$, where $\epsilon_{ij,t} = e_{it}e_{jt} - E(e_{it}e_{jt})$. This term can be written as

$$\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \frac{1}{\phi_i^4\hat{\phi}_j^2}\lambda_i\lambda_i'[\lambda_i'\hat{H}(\hat{\lambda}_j-\lambda_j)]\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t}$$

$$-\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \frac{\hat{\phi}_j^2-\phi_j^2}{\phi_i^4\phi_j^2\hat{\phi}_j^2}\lambda_i\lambda_i'(\lambda_i'\hat{H}\lambda_j)\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t}$$

$$+\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \frac{1}{\phi_i^4\phi_j^2}\lambda_i\lambda_i'(\lambda_i'\hat{H}\lambda_j)\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t} = c_5 - c_6 + c_7$$

The term c_5 is bounded in norm by

$$C\|\hat{H}_N\|\Big(\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^8}\|\lambda_i\|^6\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^N \frac{1}{\hat{\phi}_j^2}\|\hat{\lambda}_j-\lambda_j\|^2\Big)^{1/2}\Big[\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t}\Big)^2\Big]^{1/2}$$

which is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Theorem 1 and Assumption C.5. By the boundedness of $\phi_i^2, \hat{\phi}_i^2$ and λ_i, the term c_6 is bounded in norm by

$$C^{12}\|\hat{H}_N\|\Big(\frac{1}{N}\sum_{j=1}^N(\hat{\phi}_j^2-\phi_j^2)^2\Big)^{1/2}\Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N(\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t})^2\Big)^{1/2}$$

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

$$\|\hat{H}_N\|\Big(\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^8}\|\lambda_i\|^6\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{N}\sum_{j=1}^N \frac{1}{\phi_j^2}\lambda_j\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t}\Big\|^2\Big)^{1/2}$$

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 $\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^4}\lambda_i\lambda_i'(\lambda_i'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^T f_te_{it})$, which is an $r\times r$ matrix. We consider its (g,h) $(g,h=1,2,\cdots,r)$ entry, which is equal to

$$tr\Big[\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\phi_i^4}\lambda_{ig}\lambda_{ih}f_t\lambda_i'e_{it}\Big]$$

Since

$$E(\Big\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\phi_i^4}\lambda_{ig}\lambda_{ih}f_t\lambda_i'e_{it}\Big\|^2)$$

$$= tr\Big[\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T \frac{1}{\phi_i^4\phi_j^4}\lambda_{ig}\lambda_{ih}\lambda_{jg}\lambda_{jh}f_t\lambda_i'\lambda_jf_s'\gamma_{ij,ts}\Big]$$

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

by Assumption E.1. So $tr[\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_i^4}\lambda_{ig}\lambda_{ih}f_t\lambda_i'e_{it}] = O_p(N^{-1/2}T^{-1})$ $+O_p(N^{-3/2}T^{-1/2})$ by Corollary B.1. 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 $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_i\lambda_i'(\lambda_i'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}_i)$. We denote the lth element of $\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}$ by δ_l temporarily. Notice that Lemma A.5 (a) indicates $\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}$ 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

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{r}\frac{1}{\phi_i^4}\lambda_{ig}\lambda_{il}\delta_l\lambda_{ih}\bar{e}_i = \sum_{l=1}^{r}\delta_l\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_{ig}\lambda_{il}\lambda_{ih}\bar{e}_i$$

Consider the term $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_{ig}\lambda_{il}\lambda_{ih}\bar{e}_i$, which is equal to $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_i^4}\lambda_{ig}\lambda_{il}\lambda_{ih}e_{it}$ 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 C.2. □

Proof of Theorems 2–4: The limiting distributions depend on the identification conditions, and we derive the limits under each of identification conditions.

Under IC1: By equation (B.1), Lemma C.1, and Theorem 1, we have

$$\begin{aligned} M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} = & -\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\xi_t'\Big) \\ & + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \end{aligned} \tag{C.2}$$

Substituting the above result into (A.11) and using the results of Lemmas B.4 and C.1, we have

$$\begin{aligned} \hat{\lambda}_j - \lambda_j = -\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_t\xi_t'\lambda_j + \hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_{jt} \\ + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \end{aligned} \tag{C.3}$$

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

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = -M_{ff}^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\xi_t'\Big)\lambda_j + M_{ff}^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_te_{jt}\Big) + o_p(1) \tag{C.4}$$

By Assumption F.1, it follows

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N\Big(0, (\overline{M}_{ff})^{-1}\Gamma_j^{\lambda}(\overline{M}_{ff})^{-1}\Big)$$

For the limiting distribution of $\hat{\phi}_i^2 - \phi_i^2$, consider equation (B.8). By Lemmas B.1, B.2 and C.1, equation (B.8) 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 (C.3) 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 (A.11) and Lemmas B.4 and C.1 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 C.1(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

$$\hat{\phi}_i^2 - \phi_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \phi_i^2) + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \tag{C.5}$$

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 (A.10). By Lemmas B.4 and C.1, (A.10) implies that

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

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

Using (C.2) and noticing $\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda \xrightarrow{p} I_r$, 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

$$\sqrt{T}\text{vech}(\hat{M}_{ff} - M_{ff}) = D_r^+\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_t \otimes f_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t \otimes \xi_t\Big) + o_p(1) \tag{C.6}$$

where D_r^+ denotes the Moose-Penrose inverse of the r-order duplication matrix D_r. By Assumption F.1, it follows ,

$$\sqrt{T}\text{vech}(\hat{M}_{ff} - M_{ff}) \xrightarrow{d} N\Big(0, 4D_r^+\Gamma^M D_r^{+'}\Big)$$

Under IC2: Consider equation (B.2). 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

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

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 A.1. Thus we can rewrite (B.2) 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 C.2, we can further write it as

$$
\begin{aligned}
&\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}+\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda) \\
&\qquad = O_p(N^{-1/2}T^{-1/2})+O_p(N^{-1})+O_p(T^{-1})
\end{aligned} \tag{C.7}
$$

Both $\hat{M}_{ff}$ and M_{ff} are diagonal matrices. By (A.10) and Lemmas B.4 and C.1, we have

$$
\begin{aligned}
&\text{Ndiag}\Big\{\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}+M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big\} \\
&\qquad = \text{Ndiag}\{\zeta\}+O_p(N^{-1})+O_p(T^{-1})
\end{aligned} \tag{C.8}
$$

where ζ is defined as $\zeta=\frac{1}{NT}\sum_{t=1}^{T}f_te_t'\Phi^{-1}\Lambda+\Lambda'\Phi^{-1}\frac{1}{NT}\sum_{t=1}^{T}e_tf_t'$ and 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)

$$
\begin{aligned}
&\text{Ndiag}\Big\{\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}+M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\frac{1}{N}\Big\} \\
&\qquad = O_p(N^{-1/2}T^{-1/2})+O_p(N^{-1})+O_p(T^{-1})
\end{aligned} \tag{C.9}
$$

Equation (C.7) 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 (C.9) 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 (C.7) and (C.9). We have

$$
\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\equiv(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}=O_p(N^{-1/2}T^{-1/2})+O_p(N^{-1})+O_p(T^{-1}) \tag{C.10}
$$

Given this result, it follows, by (A.10) and Lemma C.1,

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

Next, consider the right hand side of (A.11). The first term is $O_p(N^{-1/2}T^{-1/2})+O_p(N^{-1})+O_p(T^{-1})$ by (C.10) and Proposition A.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 B.4 and C.1. So it follows

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

Since $\hat{M}_{ff}^{-1}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda \xrightarrow{p} M_{ff}^{-1}$ by Proposition A.1 and Corollary A.1(c), it follows, under the condition $\sqrt{T}/N \to 0$,

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = M_{ff}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t e_{jt} + o_p(1) \tag{C.11}$$

So we have

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N\Big(0, (\overline{M}_{ff})^{-1}\Upsilon_j^{\lambda}(\overline{M}_{ff})^{-1}\Big)$$

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 (B.6). By Lemmas B.4, C.1 and Corollary B.1, the right hand side of (B.6), 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 A.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 (B.6), 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

$$\text{nonupper}\Big\{M_{ff}A_4 + \frac{1}{T}\sum_{t=1}^{T} f_t\xi_t'{\Lambda'}_1^{-1}\Big\} = O_p(N^{-1}) + O_p(T^{-1}) \tag{C.12}$$

where nonupper denotes the elements on and below the diagonal. Since under IC4 both $\hat{M}_{ff}$ and M_{ff} are diagonal matrices, equation (C.8) holds. The right hand side of (C.8) is $O_p(N^{-1}) + O_p(T^{-1})$. Rewrite (C.8) in terms of A_4,

$$\text{nondiag}\Big\{A_4'M_{ff} + M_{ff}A_4\Big\} = O_p(N^{-1}) + O_p(T^{-1}) \tag{C.13}$$

By solving the system of equations (C.12) and (C.13), we have

$$(A_4)_{gh} = \left\{\begin{array}{ll} -T^{-1}\sum_{t=1}^{T} m_g^{-1} f_{tg} d_{ht} + O_p(N^{-1}) + O_p(T^{-1}) & \text{if } g \geq h \\ -m_g^{-1}m_h(A_4)_{hg} + O_p(N^{-1}) + O_p(T^{-1}) & \text{if } g < h \end{array}\right. \tag{C.14}$$

where $d_{ht} = \xi_t'{\Lambda'}_1^{-1}v_h$, v_h is the hth column of an $r \times r$ identity matrix, f_{th} is hth component of f_t. That is,

$$A_4 = \mathcal{P}_t + O_p(N^{-1}) + O_p(T^{-1}) \tag{C.15}$$

where $\mathcal{P}_t$ is defined in the main body of the text.

Consider (A.11). By Lemmas B.4 and C.1, (A.11) can be simplified as

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

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

$$\sqrt{T}(\hat{\lambda}_j - \lambda_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\mathcal{P}_t \lambda_j + M_{ff}^{-1} f_t e_{jt}) + o_p(1)$$

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 B.4 and C.1, equation (A.10) 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

$$\operatorname{diag}\{\hat{M}_{ff} - M_{ff}\} = -2\operatorname{diag}\{M_{ff} A_4\} + O_p(N^{-1}) + O_p(N^{-1/2} T^{-1/2}) + O_p(T^{-1})$$
$$= 2\operatorname{diag}\Big\{ \frac{1}{T} \sum_{t=1}^{T} f_t \xi_t' \Lambda_1'^{-1} \Big\} + O_p(N^{-1}) + O_p(N^{-1/2} T^{-1/2}) + O_p(T^{-1})$$

where the second equality follows from (C.12). By Assumption F.1, we have, under the condition $\sqrt{T}/N \to 0$,

$$\sqrt{T}\operatorname{diag}\{\hat{M}_{ff} - M_{ff}\} \xrightarrow{d} N\Big(0, 4 J_r \Pi^M J_r'\Big)$$

where J_r is defined as $\operatorname{diag}\{M\} = J_r \operatorname{vec}(M)$ 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 (C.15),

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

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

This completes the proof of Theorems 2–4. □

Corollary C.1 *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 C.1: Under IC2, Corollary C.1 is immediately obtained by (C.10). Under IC3, an analogous result to (C.10) can still be derived. So Corollary C.1 holds. □

Appendix D: Proof of results for estimated factors

Lemma D.1 *Under Assumptions A-E, we have*

$$(a)\quad \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}(\hat{\lambda}_i-\lambda_i)e_{it}=O_p(N^{-3/2})+O_p(T^{-1})+O_p(N^{-1/2}T^{-1/2})$$

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

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

Proof of Lemma D.1: Consider (a). Substituting (A.11) into (a), the left hand side can be expanded into an expression with 13 terms. The 1st term is equal to

$$\hat{M}_{ff}^{-1}M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it}+\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})\lambda_i e_{it}\Big)$$

The term $\hat{M}_{ff}^{-1}M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}=O_p(T^{-1/2})+O_p(N^{-1})$ by Proposition A.1 and Corollary B.1. 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 $E(\|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it}\|^2)=O(N^{-1})$ by Assumption C.3. The term $\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})\lambda_i e_{it}$ is bounded in norm by

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

which is $O_p(T^{-1/2})+O_p(N^{-1})$ by Theorem 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 $\hat{M}_{ff}^{-1}\hat{H}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\ e_{it}$, is of a smaller order term than $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i e_{it}$. So it is negligible. We remain to check the 8th, 10th, 12th, and 13th terms. The 8th term is $\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{1}{\hat{\phi}_i^2}f_s\ e_{it}e_{is}$, which is equivalent to

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

The second expression $\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{1}{\hat{\phi}_i^2}f_s\rho_{i,ts}$ is bounded by $C^3\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\ \rho_{ts}\le C^4T^{-1}$ by $\sum_{s=1}^{T}\rho_{ts}\le C$ by Assumption C.4′. The first expression can be written as

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{1}{\phi_i^2}f_s[e_{is}e_{it}-E(e_{is}e_{it})]+\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})f_s[e_{is}e_{it}-E(e_{is}e_{it})]$$

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

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

which is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Assumption E.5 and Theorem 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

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

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

$$\begin{aligned}
&\hat{M}_{ff}^{-1}\hat{H}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}(\hat{\lambda}_j-\lambda_j)e_{it}\frac{1}{T}\sum_{s=1}^{T}\epsilon_{ij,s}\\
&+\hat{M}_{ff}^{-1}\hat{H}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}(\frac{1}{\hat{\phi}_j^2}-\frac{1}{\phi_j^2})\frac{1}{\hat{\phi}_i^2}\lambda_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\epsilon_{ij,s}\\
&+\hat{M}_{ff}^{-1}\hat{H}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_j^2}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})\lambda_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\epsilon_{ij,s}\\
&+\hat{M}_{ff}^{-1}\hat{H}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2\phi_j^2}\lambda_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\epsilon_{ij,s}
\end{aligned}\tag{D.1}$$

The first expression is bounded in norm by

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

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

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

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

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

which is further bounded by

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

The last factor of the above expression is bounded by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}e_{it}^4\Big)^{1/4}\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\big(\frac{1}{T}\sum_{s=1}^{T}\epsilon_{ij,s})^4\Big)^{1/2}$$

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

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

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

The 12th term is equal to

$$\hat{M}_{ff}^{-1}\hat{H}_N\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}\hat{\lambda}_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s}$$

The above term can be split into

$$\begin{aligned}
&\hat{M}_{ff}^{-1}\hat{H}_N\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}(\hat{\lambda}_j-\lambda_j)e_{it}\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s}\\
&+\hat{M}_{ff}^{-1}\hat{H}_N\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_j^2}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})\lambda_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s}\\
&+\hat{M}_{ff}^{-1}\hat{H}_N\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2}(\frac{1}{\hat{\phi}_j^2}-\frac{1}{\phi_j^2})\lambda_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s}\\
&+\hat{M}_{ff}^{-1}\hat{H}_N\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2\phi_j^2}\lambda_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s}
\end{aligned}\tag{D.2}$$

The first expression of the above is bounded in norm by

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

Since

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s})^2 \le \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij}^2 \le (\sup_{i,j\le N}\tau_{ij})\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij} = O(N^{-1}) \tag{D.3}$$

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

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

which is $O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2})$ by

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}e_{it}^2(\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s})^2 \le \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}e_{it}^2\tau_{ij}^2 \le (\sup_{i,j\le N}\tau_{ij})\frac{1}{N^2}\sum_{i=1}^{N}e_{it}^2\sum_{j=1}^{N}\tau_{ij}$$

$$\leq C(\sup_{i,j\leq N}\tau_{ij})\frac{1}{N^2}\sum_{i=1}^{N}e_{it}^2=O_p(N^{-1})$$

The third expression is bounded in norm by

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

which is $O_p(N^{-1/2}T^{-1/2})+O_p(N^{-3/2})$ by Theorem 1 and (D.3). Consider the last expression. Since

$$E\Big(\Big\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\phi_i^2\phi_j^2}\lambda_j e_{it}\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s}\Big\|^2\Big)$$

$$=\frac{1}{N^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\frac{1}{\phi_i^2\phi_j^2\phi_m^2\phi_n^2}\lambda_j'\lambda_n E(e_{it}e_{mt})(\frac{1}{T}\sum_{s=1}^{T}\tau_{ij,s})(\frac{1}{T}\sum_{s=1}^{T}\tau_{mn,s})$$

$$\leq C^{10}\frac{1}{N^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\tau_{im}\tau_{ij}\tau_{mn}\leq C^{11}\frac{1}{N^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{N}\tau_{im}\tau_{ij} \tag{D.4}$$

$$\leq C^{12}\frac{1}{N^4}\sum_{i=1}^{N}\sum_{m=1}^{N}\tau_{im}\leq C^{13}N^{-3}$$

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

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

Consider (b), which can be written as

$$-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}(\hat{\phi}_i^2-\phi_i^2)\lambda_i e_{it}-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2\phi_i^4}(\hat{\phi}_i^2-\phi_i^2)^2\lambda_i e_{it} \tag{D.5}$$

Using (B.8), 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

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

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

$$C^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\lambda}_i-\lambda_i\|^2 e_{it}^2\Big)^{1/2}$$

Using the argument following (B.9) 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

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\Big(\lambda_i'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\lambda_i\Big)\lambda_i e_{it}.$$

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

$$tr\Big[\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)M_{ff}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_i\lambda_i'\lambda_{ik}e_{it}\Big]$$

The term $\frac{1}{N}\sum_{i=1}^{N}\lambda_i\lambda_i'\lambda_{ik}e_{it}$ is $O_p(N^{-1/2})$ due to $E\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\lambda_i'\lambda_{ik}e_{it}\|^2=O(N^{-1})$ 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 C.1(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 $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_i e_{it}(\lambda_i'\hat{H}\hat{\lambda}_i)$. Its kth element can be written as

$$tr\Big[\hat{H}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_{ik}\lambda_i\lambda_i'e_{it}+\hat{H}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^4}\lambda_{ik}(\hat{\lambda}_i-\lambda_i)\lambda_i'e_{it}\Big]$$

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

$$C^7\|\hat{H}\|\Big(\frac{1}{N}\sum_{i=1}^{N}e_{it}^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\|\hat{\lambda}_i-\lambda_i\|^2\Big)^{1/2}$$

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

The 10th term is equal to

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

Its kth element can be written as

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

The above expression is equal to

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

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

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

Using the argument in analyzing (D.2), 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

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

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

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

which can be written as

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

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

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

Using argument in analyzing (D.1), 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

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^4}\lambda_i e_{it}(\lambda_i'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{s=1}^T f_s e_{is})$$

Its kth element is

$$tr\Big[\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T \frac{1}{\phi_i^4}\lambda_{ik}f_s\lambda_i'e_{it}e_{is}\Big]$$

Since the term $\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T \frac{1}{\phi_i^4}\lambda_{ik}f_s\lambda_i'e_{it}e_{is}$ is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^8}\|\lambda_{ik}\lambda_i\|^2 e_{it}^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N \|\frac{1}{T}\sum_{s=1}^T f_s e_{is}\|^2\Big)^{1/2}$$

which is $O_p(T^{-1/2})$ by (A.2). Thus the kth element of the 12th term is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ by Corollary B.1. So the 12th term is $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$.

The 13th term is equal to $\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^4}\lambda_i e_{it}(\lambda_i'\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\bar{e}_i)$. Its kth element is

$$tr\Big[\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\bar{e}\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^4}\lambda_{ik}e_{it}\bar{e}_i\lambda_i'\Big]$$

The term $\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^4}\lambda_{ik}e_{it}\bar{e}_i\lambda_i'$ is bounded in norm by

$$C^6\Big(\frac{1}{N}\sum_{i=1}^N e_{it}^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N \bar{e}_i^2\Big)^{1/2}$$

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 B.4(b). So the last term is $O_p(T^{-1})$.

Summarizing results, we have

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{\phi_i^4}(\hat{\phi}_i^2-\phi_i^2)\lambda_i e_{it} = O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-3/2}).$$

Next, consider the second term of (D.5), i.e. $\frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\phi}_i^2\phi_i^4}(\hat{\phi}_i^2-\phi_i^2)^2\lambda_i e_{it}$, which can be written as

$$\frac{1}{N}\sum_{i=1}^N \Big(\frac{\hat{\phi}_i^2-\phi_i^2}{\hat{\phi}_i^2\phi_i^4}\Big)\Big((\hat{\phi}_i^2-\phi_i^2)\lambda_i e_{it}\Big)$$

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

$$\Big(\frac{1}{N}\sum_{i=1}^N \big(\frac{\hat{\phi}_i^2-\phi_i^2}{\hat{\phi}_i^2\phi_i^4}\big)^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N \Big\|(\hat{\phi}_i^2-\phi_i^2)\lambda_i e_{it}\Big\|^2\Big)^{1/2}$$

which is further bounded by

$$C^7\Big(\frac{1}{N}\sum_{i=1}^N (\hat{\phi}_i^2-\phi_i^2)^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N (\hat{\phi}_i^2-\phi_i^2)^2 e_{it}^2\Big)^{1/2}$$

The first factor of the above expression is $O_p(T^{-1/2}) + O_p(N^{-1})$. Using the argument following (B.8) 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

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\phi}_i^2\phi_i^4}(\hat{\phi}_i^2-\phi_i^2)^2\lambda_i e_{it} = O_p(T^{-1}) + O_p(N^{-2}),$$

completing the proof of (b).

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

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}(\hat{\lambda}_i-\lambda_i)\bar{e}_i+\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})\lambda_i\bar{e}_i+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i\bar{e}_i=c_1+c_2+c_3$$

Term c_1 is bounded in norm by

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

which is $O_p(T^{-1})+O_p(N^{-1}T^{-1/2})$ by Theorem 1 and $\bar{e}_i=O_p(T^{-1/2})$.

Term c_2 is bounded in norm by

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

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 $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\phi_i^2}\lambda_i e_{it}$, which is $O_p((NT)^{-1/2})$. Thus $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\bar{e}_i=O_p(T^{-1})+O_p(N^{-1/2}T^{-1/2})$.

Nest, we consider $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i e_{it}$, which can be written as

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

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

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

This yields (c). $\square$

Given Lemma D.1, the proof of Theorem 5 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 D.1

Proposition D.1 *Assume that Assumptions A-E hold. Under each of IC1, IC4, and IC5, we have*

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

and under IC2 or IC3, we have

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

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 $I_r - \Lambda'\hat{\Phi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}$ under different identification restrictions. As pointed out in the discussion preceding Theorem 1 the matrix $\Lambda'\hat{\Phi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}$ plays the same role as the rotation matrix and its asymptotic property depends on the identification conditions.

The principal components estimator uses similar identifications as IC2 and IC3, but the convergence rate is $\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t - R_1 f_t\|^2 = O_p(T^{-1}) + O_p(N^{-1})$, see Bai (2003), where R_1 is an invertible matrix. So the rate is slower than the likelihood method. The primary reason is that the principal components method ignore the heteroskedasticity and there is a bias of order $O(1/T)$.

The simulation results of Doz et al. (2011a) show that the likelihood method performs better than the principal components method, Corollary 4 and Proposition D.1 provide a theoretical justification.

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 Λ, 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 Theorem 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 Λ.

Appendix E: Assumptions and proofs for Section 5

The following assumptions are needed to derive the limiting results in Theorem 6. 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 λ_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 $\Phi = \mathrm{diag}(\phi_1^2, \ldots, \phi_N^2)$ 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 ϵ_{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, ϵ_{it} is independent over i and *i.i.d.* over t for each given i. The fourth moment of ϵ_{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.

Assumption 5D (Identification conditions): To fix the rotational indeterminacy, we impose $N^{-1}\Lambda'\Phi^{-1}\Lambda$ to be a diagonal matrix with distinct diagonal elements (arranged in decreasing order) and $\frac{1}{T}\sum_{t=1}^T f_t f_t' = I_r$.

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

$$
\begin{aligned}
\psi_{it} &= (e_{it-1}, e_{it-2}, \ldots, e_{it-p_i})', && \text{accordingly} \quad \hat{\psi}_{it} = (\hat{e}_{it-1}, \hat{e}_{it-2}, \ldots, \hat{e}_{it-p_i})' \\
\rho_i &= (\rho_{i,1}, \rho_{i,2}, \ldots, \rho_{i,p_i})', && \text{accordingly} \quad \hat{\rho}_i = (\hat{\rho}_{i,1}, \hat{\rho}_{i,2}, \ldots, \hat{\rho}_{i,p_i})' \\
g_{it} &= f_t - \rho_{i,1} f_{t-1} - \cdots - \rho_{i,p_i} f_{t-p_i}, && \text{accordingly} \quad \hat{g}_{it} = \hat{f}_t - \hat{\rho}_{i,1}\hat{f}_{t-1} - \cdots - \hat{\rho}_{i,p_i}\hat{f}_{t-p_i} \\
\widehat{\Delta f}_{t-p} &= \hat{f}_{t-p} - f_{t-p}, && \text{for} \quad p = 0, 1, \ldots, p_i \\
\widehat{\Delta\lambda}_i &= \hat{\lambda}_i - \lambda_i, && \\
\widehat{\Delta\rho}_{i,p} &= \hat{\rho}_{i,p} - \rho_{i,p} && \text{for} \quad p = 1, \ldots, p_i
\end{aligned}
$$

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 C.1 holds. The following two lemmas are useful.

Lemma E.1 *Under Assumptions 5A-5D, we have*

$$(a) \quad \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T \widehat{\Delta f}_{t-p}\widehat{\Delta f}'_{t-q} = O_p(N^{-1}) + O_p(T^{-1}), \quad \textit{for } p, q = 0, 1, \ldots, p_i$$

$$(b) \quad \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T f_{t-p}\widehat{\Delta f}'_{t-q} = O_p(N^{-1}) + O_p(T^{-1}), \quad \textit{for } p, q = 0, 1, \ldots, p_i$$

$$(c) \quad \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T \hat{f}_{t-p}\widehat{\Delta f}'_{t-q} = O_p(N^{-1}) + O_p(T^{-1}), \quad \textit{for } p, q = 0, 1, \ldots, p_i$$

Proof of Lemma E.1: Consider (a).

$$
\begin{aligned}
\widehat{\Delta f}_{t-p} \equiv \hat{f}_{t-p} - f_{t-p} &= -(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda)f_{t-p} + (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1}e_{t-p} \\
&= -A'f_{t-p} + \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}e_{t-p}
\end{aligned}
\tag{E.1}
$$

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

$$
\begin{aligned}
&A'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T f_{t-p}f'_{t-q}\Big)A + \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T (e_{t-p}e'_{t-q})\hat{\Phi}^{-1}\hat{\Lambda}\hat{H} \\
&\quad - \Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T \hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}e_{t-p}f'_{t-q}\Big)A - A'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T f_{t-p}e'_{t-q}\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big)
\end{aligned}
\tag{E.2}
$$

The first term of (E.2) is $O_p(N^{-2}) + O_p(T^{-2})$ by $\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^T f_{t-p}f'_{t-q} = O_p(1)$ and

Corollary C.1. The second term is equal to

$$\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}\hat{\lambda}_i\hat{\lambda}_j'\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}[e_{it-p}e_{jt-q}-E(e_{it-p}e_{jt-q})]\Big)\hat{H}$$
$$+\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\phi}_i^2\hat{\phi}_j^2}\hat{\lambda}_i\hat{\lambda}_j'\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}E(e_{it-p}e_{jt-q})\Big)\hat{H}$$

The first expression can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ similarly as Lemma C.1(d). The second expression is equal to $\hat{H}[\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^4}\hat{\lambda}_i\hat{\lambda}_i'\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}E(e_{it-p}e_{it-q})]\hat{H}$ by the assumption of cross-sectional independence, which is further bounded by

$$C^2\|\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\Big\|\hat{H}^{1/2}\frac{\hat{\lambda}_i}{\hat{\phi}_i}\Big\|^2\Big)\cdot\sup_i|E(e_{it-p}e_{it-q})|=O_p(N^{-1})$$

So the second term of (E.2) is $O_p(N^{-1})+O_p(T^{-1})$. Term $\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}e_{t-p}f_{t-q}'$ is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$, which can be shown similarly as Lemma C.1(e), so the third term of (E.2) is $O_p(N^{-3/2}T^{-1/2})+O_p(T^{-2})$. The last term of (E.2) 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 (E.1), the left hand side of (b) is equal to

$$-\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}f_{t-p}f_{t-q}'\Big)A+\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}f_{t-p}e_{t-q}'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}$$

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

Consider (c). Notice

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}\widehat{\Delta f}_{t-q}'=\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\widehat{\Delta f}_{t-q}'+\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}f_{t-p}\widehat{\Delta f}_{t-q}'.$$

So (c) follows immediately by (a) and (b). □

Lemma E.2 *Under Assumptions 5A-5D,*

(a) $\displaystyle\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-q}\epsilon_{it}=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1}),\quad \textit{for } q=1,2,\dots,p_i$

(b) $\displaystyle\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-q}\epsilon_{it}=O_p(T^{-1/2}),\quad \textit{for } q=1,2,\dots,p_i$

(c) $\displaystyle\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}e_{it-q}=O_p(N^{-1})+O_p(T^{-1}),\quad \textit{for } p,q=0,1,2,\dots,p_i$

(d) $\displaystyle\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}e_{it-q}=O_p(T^{-1/2}),\quad \textit{for } p,q=0,1,2,\dots,p_i$

Proof of Lemma E.2: Consider (a). By (E.1), the left hand side of (a) is

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

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 C.1. The second term is equal to $\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\hat{\lambda}_i\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} e_{it-q}\epsilon_{it}$. Notice $E(e_{it-q}\epsilon_{it}) = 0$, thus this term can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ similarly as Lemma C.1(c). Given these results, we have (a).

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

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-q}\epsilon_{it} + \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} f_{t-q}\epsilon_{it}$$

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 (E.1), the left hand side of (c) is equal to

$$-A'\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} f_{t-p}e_{it-q} + H\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\Lambda}'\hat{\Phi}^{-1}e_{t-p}e_{it-q}$$

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

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

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

Consider (d). Notice

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}e_{it-q} = \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}e_{it-q} + \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} f_{t-p}e_{it-q}$$

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

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

Lemma E.3 *Under Assumptions 5A-5D,*

$$(a)\quad \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{e}_{it-p}\epsilon_{it} = \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}e_{it-p}\epsilon_{it} + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}),\ \textit{for}\ p=1,\ldots,p_i$$

$$(b)\quad \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{e}_{it-p}\widehat{\Delta f}'_{t-q} = O_p(N^{-1}) + O_p(T^{-1}),\ \textit{for}\ p,q=0,1,\ldots,p_i$$

$$(c)\quad \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{e}_{it-p}\hat{f}'_{t-q} = O_p(N^{-1}) + O_p(T^{-1/2}),\ \textit{for}\ p,q=0,1,\ldots,p_i$$

$$(d)\quad \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{e}_{it-p}\hat{e}_{it-q} = \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}e_{it-p}e_{it-q} + O_p(N^{-1}) + O_p(T^{-1}),\ \textit{for}\ p,q=1,\ldots,p_i$$

Proof of Lemma E.3 : Consider (a). By

$$\hat{e}_{it-p} = e_{it-p} - \lambda_i'(\hat{f}_{t-p} - f_{t-p}) - (\hat{\lambda}_i - \lambda_i)'\hat{f}_{t-p} = e_{it-p} - \lambda_i'\widehat{\Delta f}_{t-p} - \widehat{\Delta\lambda}_i{}'\hat{f}_{t-p}, \quad \text{(E.3)}$$

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

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}e_{it-p}\epsilon_{it} - \lambda_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\epsilon_{it}\Big) - \widehat{\Delta\lambda}_i{}'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}\epsilon_{it}\Big)$$

The second term of the above expression is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma E.2(a). The third term is $O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$ by Lemma E.2(b) and $\widehat{\Delta\lambda}_i = O_p(N^{-1}) + O_p(T^{-1/2})$. Given these results, (a) follows.

Consider (b). By (E.3), the left hand side of (b) is equal to

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}e_{it-p}\widehat{\Delta f}'_{t-q} - \lambda_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\widehat{\Delta f}'_{t-q}\Big) - \widehat{\Delta\lambda}_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}\widehat{\Delta f}'_{t-q}\Big)$$

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

Consider (c). By (E.3), the left hand side of (c) is equal to

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}e_{it-p}\hat{f}'_{t-q} - \lambda_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\hat{f}'_{t-q}\Big) - \widehat{\Delta\lambda}_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}\hat{f}'_{t-q}\Big)$$

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

Consider (d). By (E.3), the left hand side of (d) is equal to

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}e_{it-p}e_{it-q} - \lambda_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}e_{it-q}\Big) - \widehat{\Delta\lambda}_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}e_{it-q}\Big)$$

$$-\lambda_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-q}e_{it-p}\Big)+\lambda_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\widehat{\Delta f}_{t-q}'\Big)\lambda_i'+\widehat{\Delta\lambda}_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}\widehat{\Delta f}_{t-q}'\Big)\lambda_i$$

$$-\widehat{\Delta\lambda}_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-q}e_{it-p}\Big)+\lambda_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\hat{f}_{t-q}'\Big)\widehat{\Delta\lambda}_i+\widehat{\Delta\lambda}_i'\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{f}_{t-p}\hat{f}_{t-q}'\Big)\widehat{\Delta\lambda}_i$$

The second and fourth terms are $O_p(N^{-1})+O_p(T^{-1})$ by Lemma E.2(c). The third and seventh terms are both $O_p(N^{-1}T^{-1/2})+O_p(T^{-1})$ by Lemma E.2(d) and $\widehat{\Delta\lambda}_i = O_p(N^{-1})+O_p(T^{-1/2})$. Using the results in Lemma E.1, the remaining terms except the first one are $O_p(N^{-1})+O_p(T^{-1})$. These results imply (d). □

Proof of Theorem 6: Recall that the estimator $\hat{\rho}_i$ is obtained by running the regression

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

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

$$\hat{\rho}_i=\Big(\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{\psi}_{it}'\Big)^{-1}\Big(\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{e}_{it}\Big)$$

Then it follows

$$\hat{\rho}_i-\rho_i=\Big(\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{\psi}_{it}'\Big)^{-1}\Big(\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}(\hat{e}_{it}-\rho_{i,1}\hat{e}_{it-1}-\cdots-\rho_{i,p_i}\hat{e}_{it-p_i})\Big)$$

By (E.3) and $\epsilon_{it}=e_{it}-\rho_{i,1}e_{it-1}-\cdots-\rho_{i,p_i}e_{it-p_i}$, we have

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

So we have

$$\hat{\rho}_i-\rho_i=\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{\psi}_{it}'\Big)^{-1}\Big[\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\epsilon_{it}\Big)-\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\widehat{\Delta f}_t'\Big)\lambda_i \quad \text{(E.4)}$$

$$-\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{f}_t'\Big)\widehat{\Delta\lambda}_i+\sum_{j=1}^{p_i}\rho_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\widehat{\Delta f}_{t-j}'\Big)\lambda_i+\sum_{j=1}^{p_i}\rho_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{f}_{t-j}'\Big)\widehat{\Delta\lambda}_i\Big]$$

By Lemma E.3(a),

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

By Lemma E.3(b),

$$-\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\widehat{\Delta f}_t'\Big)\lambda_i+\sum_{j=1}^{p_i}\rho_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\widehat{\Delta f}_{t-j}'\Big)\lambda_i=O_p(N^{-1})+O_p(T^{-1})$$

By Lemma E.3(c),

$$-\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{f}_t'\Big)\widehat{\Delta\lambda}_i+\sum_{j=1}^{p_i}\rho_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{f}_{t-j}'\Big)\widehat{\Delta\lambda}_i$$
$$=[O_p(N^{-1})+O_p(T^{-1/2})][O_p(N^{-1})+O_p(T^{-1/2})]=O_p(N^{-2})+O_p(T^{-1})$$

By Lemma E.3(d),

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{\psi}_{it}\hat{\psi}_{it}'=\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\psi_{it}\psi_{it}'+O_p(N^{-1})+O_p(T^{-1})$$

Then it follows

$$\hat{\rho}_i-\rho_i=\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\psi_{it}\psi_{it}'\Big)^{-1}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\psi_{it}\epsilon_{it}\Big)+O_p(N^{-1})+O_p(T^{-1}) \tag{E.5}$$

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

$$\sqrt{T-p_i}(\hat{\rho}_i-\rho_i)=\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\psi_{it}\psi_{it}'\Big)^{-1}\Big(\frac{1}{\sqrt{T-p_i}}\sum_{t=\bar{p}_i}^{T}\psi_{it}\epsilon_{it}\Big)+o_p(1) \tag{E.6}$$

By the martingale difference central limiting theorem,

$$\sqrt{T-p_i}(\hat{\rho}_i-\rho_i)\xrightarrow{d}N\Big(0,\sigma_{\epsilon i}^2\Big[\text{plim}_{T\to\infty}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\psi_{it}\psi_{it}'\Big]^{-1}\Big)$$

This completes the proof of the $\hat{\rho}_i$ part of Theorem 6. □

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

Lemma E.4 *Under Assumptions 5A-5D,*

(a) $\displaystyle\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\widehat{\Delta f}_{t-q}'=O_p(N^{-1})+O_p(T^{-1}),$ *for* $q=0,1,\dots,p_i$

(b) $\displaystyle\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}e_{it-p}=O_p(N^{-1})+O_p(T^{-1/2}),$ *for* $p=1,\dots,p_i$

(c) $\displaystyle\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\epsilon_{it}=\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}\epsilon_{it}+O_p(N^{-1}T^{-1/2})+O_p(T^{-1})$

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

Proof of Lemma E.4: Consider (a). By $\hat{\rho}_{i,j}\hat{f}_{t-j}=\hat{\rho}_{i,j}\widehat{\Delta f}_{t-j}+\hat{\rho}_{i,j}f_{t-j}=\hat{\rho}_{i,j}\widehat{\Delta f}_{t-j}+\widehat{\Delta\rho}_{i,j}f_{t-j}+\rho_{i,j}f_{t-j}$, we have

$$\hat{g}_{it}=g_{it}-\sum_{j=1}^{p_i}\widehat{\Delta\rho}_{i,j}f_{t-j}-\widehat{\Delta f}_t-\sum_{j=1}^{p_i}\hat{\rho}_{i,j}\widehat{\Delta f}_{t-j} \tag{E.7}$$

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

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}\widehat{\Delta f}'_{t-q} - \sum_{p=1}^{p_i}\widehat{\Delta\rho}_{i,p}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} f_{t-p}\widehat{\Delta f}'_{t-q}$$

$$-\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_t\widehat{\Delta f}'_{t-q} - \sum_{p=1}^{p_i}\hat{\rho}_{i,j}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\widehat{\Delta f}'_{t-q}$$

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

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

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}e_{it-p} - \sum_{q=1}^{p_i}\widehat{\Delta\rho}_{i,q}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} f_{t-q}e_{it-p}$$

$$-\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_t e_{it-p} - \sum_{q=1}^{p_i}\hat{\rho}_{i,q}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-q}e_{it-p}$$

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 $\widehat{\Delta\rho}_{i,q} = O_p(N^{-1}) + O_p(T^{-1/2})$. The third and fourth terms are both $O_p(N^{-1}) + O_p(T^{-1})$ by Lemma E.2(c) and $\hat{\rho}_{i,q} \xrightarrow{p} \rho_{i,q}$. This proves (b).

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

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

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

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 E.2(a) and $\hat{\rho}_{i,q} \xrightarrow{p} \rho_{i,q}$. Thus (c) follows.

Consider (d). Let $\hat{\rho}_{i,0} \equiv 1$. Then equation (E.7) can be written as

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

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

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

$$-\sum_{q=1}^{p_i}\widehat{\Delta\rho}_{i,q}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}f'_{t-q} + \sum_{p=1}^{p_i}\sum_{q=1}^{p_i}\widehat{\Delta\rho}_{i,p}\widehat{\Delta\rho}_{i,q}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} f_{t-p}f'_{t-q}$$

$$+\sum_{p=0}^{p_i}\sum_{q=1}^{p_i}\hat{\rho}_{i,p}\widehat{\Delta\rho}_{i,q}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}f'_{t-q}-\sum_{q=0}^{p_i}\hat{\rho}_{i,q}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}\widehat{\Delta f}'_{t-q}\Big)$$

$$+\sum_{p=1}^{p_i}\sum_{q=0}^{p_i}\widehat{\Delta\rho}_{i,p}\hat{\rho}_{i,q}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}f_{t-p}\widehat{\Delta f}'_{t-q}\Big)+\sum_{p=0}^{p_i}\sum_{q=0}^{p_i}\hat{\rho}_{i,p}\hat{\rho}_{i,q}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\widehat{\Delta f}_{t-p}\widehat{\Delta f}'_{t-q}\Big)$$

The second and fourth terms are both $O_p(N^{-1})+O_p(T^{-1/2})$ by $\widehat{\Delta\rho}_{i,p}=O_p(N^{-1})+O_p(T^{-1/2})$ and $\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}f_{t-p}g'_{it}=O_p(1)$. The third and seventh terms are both $O_p(N^{-1})+O_p(T^{-1})$ by Lemma E.1(b) and $\hat{\rho}_{i,j}\xrightarrow{p}\rho_{i,j}$. 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 E.1(b) and $\widehat{\Delta\rho}_{i,p}=O_p(N^{-1})+O_p(T^{-1/2})$. The fifth term is $O_p(N^{-2})+O_p(T^{-1})$ by $\widehat{\Delta\rho}_{i,p}=O_p(N^{-1})+O_p(T^{-1/2})$ and $\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}f_{t-p}f'_{t-q}=O_p(1)$. The last term is $O_p(N^{-1})+O_p(T^{-1})$ by Lemma E.1(a) and $\hat{\rho}_{i,j}\xrightarrow{p}\rho_{i,j}$. Summarizing all the results, we have (d). □

Proof of Theorem 6 (continued): Recall that the estimator $\tilde{\lambda}_i$ is obtained by running the regression

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

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

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

Rewrite $\tilde{\lambda}_i$ as

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

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

$$\tilde{\lambda}_i-\lambda_i=\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\hat{g}'_{it}\Big)^{-1}\Big[\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\widehat{\Delta f}'_t\Big)-\sum_{j=1}^{p_i}\hat{\rho}_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\widehat{\Delta f}'_{t-j}\Big)\Big]\lambda_i$$

$$+\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\hat{g}'_{it}\Big)^{-1}\Big[\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\epsilon_{it}\Big)-\sum_{j=1}^{p_i}\widehat{\Delta\rho}_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}e_{it-j}\Big)\Big]\quad\text{(E.8)}$$

By Lemma E.4(a),

$$\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\widehat{\Delta f}'_t\Big)-\sum_{j=1}^{p_i}\hat{\rho}_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\widehat{\Delta f}'_{t-j}\Big)=O_p(N^{-1})+O_p(T^{-1}).$$

By Lemma E.4(b) and (E.5)

$$\sum_{j=1}^{p_i}\widehat{\Delta\rho}_{i,j}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}e_{it-j}\Big)=O_p(N^{-2})+O_p(T^{-1}).$$

By Lemma E.4(c),

$$\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}\hat{g}_{it}\epsilon_{it} = \frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}\epsilon_{it} + O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}).$$

By Lemma E.4(d),

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

Then it follows

$$\tilde{\lambda}_i - \lambda_i = \Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}g_{it}'\Big)^{-1}\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}\epsilon_{it}\Big) + O_p(N^{-1}) + O_p(T^{-1}) \tag{E.9}$$

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

$$\sqrt{T-p_i}(\tilde{\lambda}_i - \lambda_i) = \Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}g_{it}'\Big)^{-1}\Big(\frac{1}{\sqrt{T-p_i}}\sum_{t=\bar{p}_i}^{T} g_{it}\epsilon_{it}\Big) + o_p(1) \tag{E.10}$$

By the cental limiting theorem,

$$\sqrt{T-p_i}(\tilde{\lambda}_i - \lambda_i) \xrightarrow{d} N\Big(0, \sigma_{\epsilon i}^2\Big[\text{plim}_{T\to\infty}\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}g_{it}'\Big]^{-1}\Big)$$

We proceed to consider the limiting results on $\tilde{f}_t$. Recall that $\tilde{f}_t$ is obtained by the regression

$$\frac{1}{\hat{\phi}_i} z_{it} = \frac{1}{\hat{\phi}_i}\tilde{\lambda}_i' f_t + error$$

So we have

$$\tilde{f}_t = \Big(\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i\tilde{\lambda}_i'\Big)^{-1}\Big(\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i z_{it}\Big) = \Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i\tilde{\lambda}_i'\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i z_{it}\Big)$$

By $z_{it} = \lambda_i' f_t + e_{it}$, we have

$$\tilde{f}_t - f_t = -\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i\tilde{\lambda}_i'\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i(\tilde{\lambda}_i - \lambda_i)'\Big) f_t + \Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i\tilde{\lambda}_i'\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i e_{it}\Big) \tag{E.11}$$

Given (E.9), together with the boundedness of $\hat{\phi}_i^2$, it follows

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\|\tilde{\lambda}_i - \lambda_i\|^2 = O_p(N^{-2}) + O_p(T^{-1}) \tag{E.12}$$

Consider the expression $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i(\tilde{\lambda}_i - \lambda_i)'$. By (E.9), the expression is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}'\epsilon_{it}\Big)\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T} g_{it}g_{it}'\Big)^{-1} + O_p(N^{-1}) + O_p(T^{-1})$$

The first term of the above expression is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}(\tilde{\lambda}_i-\lambda_i)\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}'\epsilon_{it}\Big)\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}g_{it}'\Big)^{-1}$$

$$+\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\hat{\phi}_i^2}-\frac{1}{\phi_i^2})\lambda_i\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}'\epsilon_{it}\Big)\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}g_{it}'\Big)^{-1}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}'\epsilon_{it}\Big)\Big(\frac{1}{T-p_i}\sum_{t=\bar{p}_i}^{T}g_{it}g_{it}'\Big)^{-1}$$

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 C.1(b). The last term is $O_p(N^{-1/2}T^{-1/2})$. So $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i(\tilde{\lambda}_i-\lambda_i)'=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. By the similar argument, we have

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

Given this result, notice $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\phi}_i^2}\tilde{\lambda}_i\tilde{\lambda}_i'=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i\lambda_i'+o_p(1)$, we have

$$\tilde{f}_t-f_t=\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i\lambda_i'\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it}\Big)+O_p(N^{-1})+O_p(T^{-1})$$

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

$$\sqrt{N}(\tilde{f}_t-f_t)=\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i\lambda_i'\Big)^{-1}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\phi_i^2}\lambda_i e_{it}\Big)+o_p(1)$$

This completes the proof of Theorem 6. $\square$

Appendix F: The asymptotic equivalence between the GLS estimators and the Kalman-soomther-based estimators

Equation (7) can be viewed as the Kalman smoother in the absence of dynamics in the factors. However, when the dynamics of factors are explicitly modeled, intuitively, the Kalman smoother should be a preferred method in the estimation. In this appendix, we analyze the Kalman-smoother-based method. We present two results. First, we prove that when f_t is a vector autoregressive process as in Remark 2, modeling and estimating the dynamic process f_t will not improve the asymptotic efficiency of $\hat{f}_t$. A similar point is observed by Breitung and Tenhofen (2011). Second, we deliver the limiting distributions of the Kalman-smoother-based estimators. Doz et al. (2011b) also consider the Kalman-smoother-based estimators. They consider the rate of convergence of the estimators. Our results imply the limiting distributions of the Kalman-smoother-based estimators.

Consider the following specification of the dynamics of the factors:

$$f_t = \Psi_1 f_{t-1} + \Psi_2 f_{t-2} + \cdots + \Psi_K f_{t-K} + u_t. \tag{F.1}$$

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 $\mathcal{Z} = \text{vec}(Z), \mathcal{F} = \text{vec}(F'), \mathcal{E} = \text{vec}(E)$. Then we have

$$\mathcal{Z} = (I_T \otimes \Lambda)\mathcal{F} + \mathcal{E} \tag{F.2}$$

Throughout the section, the normality assumption is maintained. However, if we interpret the conditional expectation as a linear population projection, normality is not needed. Let $\Sigma_{\mathcal{F}} = \text{var}(\mathcal{F})$. Given this assumption, by (F.2), we have

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

Thus the best prediction for $\mathcal{F}$ given $(\mathcal{Z}, \Lambda, \Phi, \Sigma_{\mathcal{F}})$, denoted by $E(\mathcal{F}|\mathcal{Z})$, is

$$\begin{aligned} E(\mathcal{F}|\mathcal{Z}) &= \Sigma_{\mathcal{F}}(I_T \otimes \Lambda')\Big[(I_T \otimes \Lambda)\Sigma_{\mathcal{F}}(I_T \otimes \Lambda') + I_T \otimes \Phi\Big]^{-1}\mathcal{Z} \\ &= \Big(\Sigma_{\mathcal{F}}^{-1} + I_T \otimes (\Lambda'\Phi^{-1}\Lambda)\Big)^{-1}\Big(I_T \otimes (\Lambda'\Phi^{-1})\Big)\mathcal{Z} \end{aligned} \tag{F.3}$$

where the second equality uses the Woodbury identity. Equation (F.3) is the Kalman smoother for the factors, which serve as the basis in the estimation of the factors.

To be consistent with the preceding analysis, we continue to allow e_t to be correlated and heteroskedastic over both the cross section and time dimensions. The true conditional expectation in (F.3) will not have a diagonal Φ, but nothing prevents us from evaluating the conditional expectation at a diagonal Φ. That is, the Kalman smoother is computed as if e_t were *i.i.d* over the time dimension and were uncorrelated over the cross sections.

Because the parameters $\Lambda, \Phi, \Sigma_{\mathcal{F}}$ 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 (8), then obtain $\hat{\Sigma}_{\mathcal{F}}$ by the standard vector time series regression based on $\hat{f}_t$ and (F.1). Given $\hat{\Sigma}_{\mathcal{F}}, \hat{\Lambda}, \hat{\Phi}$, the Kalman-smoother-based estimator for f_t, denoted by $\hat{f}_t^{ks}$, is

$$\hat{f}_t^{ks} = (v_t' \otimes I_r)\Big(\hat{\Sigma}_{\mathcal{F}}^{-1} + I_T \otimes (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})\Big)^{-1}\Big(I_T \otimes (\hat{\Lambda}'\hat{\Phi}^{-1})\Big)\mathcal{Z} \tag{F.4}$$

where v_t is the t-th column of the $T \times T$ identity matrix.

For dynamic factors, we make the following assumption:

Assumption A′: The factor f_t admits the VAR representation (F.1), where u_t is a mean-zero *i.i.d* process with $E(\|u_t\|^4) \le C$ 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 F.1 (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 (8).

Theorem F.1 implies that modeling the dynamics of factors will not improve the asymptotic efficiency under large N, though there will be efficiency gain under small N. The difference between the Kalman-smoother-based estimators, which take into account of the dynamics of factors , and the projection-based estimators, which only make use of the contemporaneous relations between the factors and the observables, are asymptotically negligible.

To prove the theorem, we need additional results. Let

$$\hat{\mathcal{G}} = [\hat{\Sigma}_{\mathcal{F}}^{-1} + I_T \otimes (\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})]^{-1}.$$

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 $\|M\|_2^2 = \lambda_{max}(M'M)$. The following lemma will be used in our derivation.

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

(a) $\|(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1}\|_2 = O_p(N^{-1/2})$

(b) $\left\|(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1} - (\Lambda'\Phi^{-1}\Lambda)^{-1}\Lambda'\Phi^{-1}\right\|_2 = O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1/2})$

Proof of Lemma F.1: 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

$$\|\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\|_2^2 = \lambda_{max}\Big[\hat{H}\hat{\Lambda}'\hat{\Phi}^{-2}\hat{\Lambda}\hat{H}\Big] \leq C^2\lambda_{max}\Big[\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda}\hat{H}\Big] = C^2\lambda_{max}(\hat{H}) = O_p(N^{-1}).$$

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

$$\|\hat{H}\hat{\Lambda}'\hat{\Phi}^{-1} - H\Lambda'\Phi^{-1}\|_2 \leq \|(\hat{H} - H)\hat{\Lambda}'\hat{\Phi}^{-1}\|_2 + \|H(\hat{\Lambda}'\hat{\Phi}^{-1} - \Lambda'\Phi^{-1})\|_2$$

$$\leq \|\hat{H} - H\|_2 \cdot \|\hat{\Lambda}'\hat{\Phi}^{-1}\|_2 + \|H\|_2 \cdot \|(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\|_2 + \|H\|_2 \cdot \|\Lambda'(\hat{\Phi}^{-1} - \Phi^{-1})\|_2 \quad \text{(F.5)}$$

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

$$\begin{aligned}\frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda} - \frac{1}{N}\Lambda'\Phi^{-1}\Lambda = {} & \frac{1}{N}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}\hat{\Lambda} + \frac{1}{N}\hat{\Lambda}'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda) \\ & - \frac{1}{N}(\hat{\Lambda} - \Lambda)'\hat{\Phi}^{-1}(\hat{\Lambda} - \Lambda) + \frac{1}{N}\Lambda'(\hat{\Phi}^{-1} - \Phi^{-1})\Lambda\end{aligned}$$

Corollary B.1 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 (B.2), $\frac{1}{N}\Lambda'(\hat{\Phi}^{-1}-\Phi^{-1})\Lambda = O_p(T^{-1/2}) + O_p(N^{-1})$. Given these results, together with Theorem 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, $\|\hat{\Lambda}'\hat{\Phi}^{-1}\|_2^2 = \lambda_{max}(\hat{\Lambda}'\hat{\Phi}^{-2}\hat{\Lambda}) \leq C^2\lambda_{max}(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda}) = O_p(N)$. 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 (F.5). Notice

$$\frac{1}{N}\|(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}\|_2^2 = \lambda_{max}\Big(\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-2}(\hat{\Lambda}-\Lambda)\Big) \leq C^2\lambda_{max}\Big(\frac{1}{N}(\hat{\Lambda}-\Lambda)'\hat{\Phi}^{-1}(\hat{\Lambda}-\Lambda)\Big)$$

So the second term is $O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1/2})$ by Theorem 1 and $\|H\|_2 = O(N^{-1})$. Consider the last term of (F.5). Notice

$$\frac{1}{N}\|\Lambda'(\hat{\Phi}^{-1}-\Phi^{-1})\|_2^2 = \lambda_{max}\Big(\frac{1}{N}\Lambda'(\hat{\Phi}^{-1}-\Phi^{-1})^2\Lambda\Big)$$

The expression in the parentheses is equal to $\frac{1}{N}\sum_{i=1}^N \frac{(\hat{\phi}_i^2-\phi_i^2)^2}{\hat{\phi}_i^4\phi_i^4}\lambda_i\lambda_i'$, 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 Theorem 1. So the third term of (F.5) is $O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1/2})$. These results imply (b). □

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

(a) $\|\hat{\mathcal{G}}\|_2 = O_p(N^{-1}), \quad \|\hat{\Sigma}_{\mathcal{F}}\|_2 = O_p(1), \quad \|\hat{\Sigma}_{\mathcal{F}}^{-1}\|_2 = O_p(1),$

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

Lemma F.2 is proved by Doz et al. (2011). □

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

$$\hat{\mathcal{G}} \equiv \Big(\hat{\Sigma}_{\mathcal{F}}^{-1} + I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})\Big)^{-1} = I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1} - \hat{\mathcal{G}}\hat{\Sigma}_{\mathcal{F}}^{-1}\Big(I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\Big) \quad \text{(F.6)}$$

So we have

$$\hat{f}_t^{smo} = \hat{f}_t + (v_t'\otimes I_r)\hat{\mathcal{G}}\hat{\Sigma}_{\mathcal{F}}^{-1}\Big[I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})\Big]^{-1}\Big[I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1})\Big]\mathcal{Z}. \quad \text{(F.7)}$$

where $\hat{f}_t$ is the GLS estimator considered in Subsection 4.1. We analyze the second expression above. From $\mathcal{Z} = (I_T\otimes\Lambda)\mathcal{F} + \mathcal{E}$, we have

$$(v_t'\otimes I_r)\hat{\mathcal{G}}\hat{\Sigma}_{\mathcal{F}}^{-1}\Big[I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})\Big]^{-1}\Big[I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1})\Big]\mathcal{Z}$$

$$= (v_t'\otimes I_r)\hat{\mathcal{G}}\hat{\Sigma}_{\mathcal{F}}^{-1}\Big(I_T\otimes[(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1}\Lambda]\Big)\mathcal{F}$$

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

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

$$(v_t' \otimes I_r)[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\hat\Sigma_{\mathcal{F}}^{-1}\Big(I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}\Lambda]\Big)\mathcal{F}$$

$$-(v_t' \otimes I_r)\hat{\mathcal{G}}\hat\Sigma_{\mathcal{F}}^{-1}[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\hat\Sigma_{\mathcal{F}}^{-1}\Big(I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}\Lambda]\Big)\mathcal{F} = IG_3 - IG_4$$

By $\|AB\|_2 \le \|A\|_2\|B\|_2$, IG_4 is bounded by

$$\begin{aligned}\|IG_4\| \le & \|(v_t' \otimes I_r)\|_2 \cdot \|\hat{\mathcal{G}}\|_2 \cdot \|\hat\Sigma_{\mathcal{F}}^{-1}\|_2 \cdot \|[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\|_2 \\ & \times \|\hat\Sigma_{\mathcal{F}}^{-1}\|_2 \cdot \Big\|I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}\Lambda]\Big\|_2 \cdot \|\mathcal{F}\|,\end{aligned}$$

which is $O_p(T^{1/2}N^{-2})$ by Lemma F.2. Now consider IG_3, which is equal to

$$[v_t' \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\Sigma_{\mathcal{F}}^{-1}\mathcal{F} + [v_t' \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}](\hat\Sigma_{\mathcal{F}}^{-1} - \Sigma_{\mathcal{F}}^{-1})\mathcal{F} - [v_t' \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\hat\Sigma_{\mathcal{F}}^{-1}(I_T \otimes A')\mathcal{F}$$

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 Lemma F.2. Consider the first term. Notice $\mathrm{var}(\Sigma_{\mathcal{F}}^{-1}\mathcal{F}) = \Sigma_{\mathcal{F}}^{-1}$, 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 (F.6), which is equal to

$$(v_t' \otimes I_r)[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\Sigma_{\mathcal{F}}^{-1}\Big(I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}]\Big)\mathcal{E}$$

$$+(v_t' \otimes I_r)[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}](\hat\Sigma_{\mathcal{F}}^{-1} - \Sigma_{\mathcal{F}}^{-1})\Big(I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}]\Big)\mathcal{E}$$

$$-(v_t' \otimes I_r)\hat{\mathcal{G}}\hat\Sigma_{\mathcal{F}}^{-1}[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\Big(I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}]\Big)\mathcal{E} = IG_5 + IG_6 - IG_7$$

However,

$$\|IG_6\| \le \|v_t' \otimes I_r\|_2 \cdot \|I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\|_2 \cdot \|(\hat\Sigma_{\mathcal{F}}^{-1} - \Sigma_{\mathcal{F}}^{-1})\|_2 \cdot \Big\|I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}]\Big\|_2 \cdot \|\mathcal{E}\|$$

which is $O_p(N^{-3/2}) + O_p(N^{-5/2}T^{1/2})$ by Lemma F.2. Similarly,

$$\|IG_7\| \le \|v_t' \otimes I_r\|_2 \cdot \|\hat{\mathcal{G}}\|_2 \cdot \|\hat\Sigma_{\mathcal{F}}^{-1}\|_2 \cdot \|I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\|_2 \cdot \Big\|I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1}]\Big\|_2 \cdot \|\mathcal{E}\|$$

which is $O_p(N^{-5/2}T^{1/2})$. Now consider IG_5, which can be written as

$$(v_t' \otimes I_r)[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\Sigma_{\mathcal{F}}^{-1}\Big(I_T \otimes [(\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}\hat\Lambda'\hat\Phi^{-1} - (\Lambda'\Phi^{-1}\Lambda)^{-1}\Lambda'\Phi^{-1}]\Big)\mathcal{E}$$

$$+(v_t' \otimes I_r)[I_T \otimes (\hat\Lambda'\hat\Phi^{-1}\hat\Lambda)^{-1}]\Sigma_{\mathcal{F}}^{-1}\Big(I_T \otimes [(\Lambda'\Phi^{-1}\Lambda)^{-1}\Lambda'\Phi^{-1}]\Big)\mathcal{E}$$

The first term is bounded in norm by

$$\|v_t'\otimes I_r\|_2\cdot\|I_T\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\|_2\cdot\|\Sigma_{\mathcal{F}}^{-1}\|_2\left\|I_T\otimes[(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Phi}^{-1}-(\Lambda'\Phi^{-1}\Lambda)^{-1}\Lambda'\Phi^{-1}]\right\|\cdot\|\mathcal{E}\|$$

which is $O_p(N^{-5/2}T^{1/2})+O_p(N^{-3/2})$ by Lemma F.1(b). The second term is equal to $[v_t'\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}]\mathcal{A}$, where $\mathcal{A}=\Sigma_{\mathcal{F}}^{-1}\Big(I_T\otimes[(\Lambda'\Phi^{-1}\Lambda)^{-1}\Lambda'\Phi^{-1}]\Big)\mathcal{E}$. It is easy to show

$$E(\mathcal{A}\mathcal{A}')=\Sigma_{\mathcal{F}}^{-1}(I_T\otimes H)\Sigma_{\mathcal{F}}^{-1}$$

Notice $\lambda_{max}(\Sigma_{\mathcal{F}}^{-1}(I_T\otimes H)\Sigma_{\mathcal{F}}^{-1})=O(N^{-1})$. So we have $[v_t'\otimes(\hat{\Lambda}'\hat{\Phi}^{-1}\hat{\Lambda})^{-1}]\mathcal{A}=O_p(N^{-3/2})$. 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

$$\hat{f}_t^{smo}=\hat{f}_t+O_p(N^{-1})+O_p(T^{1/2}N^{-2}).$$

This proves Theorem F.1. □

Appendix G: Additional simulation results

Here we allow ρ_i in the error process $e_{it}=\rho_i e_{it-1}+\epsilon_{it}$ to be drawn from the uniform distribution $U[0.5,0.9]$. Tables 4 and 5 report the simulation results, which suggest similar conclusions as for Tables 1 and 2.

Table 4.1: The Trace Ratio of the seven estimators for estimating Λ.
with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0.5, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.811	0.872	0.916	0.873	0.933	0.950	0.963
10	50	0.843	0.907	0.941	0.924	0.970	0.978	0.983
10	100	0.866	0.931	0.956	0.959	0.986	0.989	0.991
20	30	0.722	0.808	0.916	0.860	0.938	0.954	0.963
20	50	0.771	0.860	0.947	0.925	0.975	0.978	0.980
20	100	0.806	0.895	0.962	0.963	0.989	0.989	0.990
50	30	0.680	0.791	0.943	0.873	0.952	0.956	0.958
50	50	0.769	0.877	0.969	0.928	0.975	0.976	0.976
50	100	0.858	0.942	0.986	0.964	0.988	0.988	0.988
100	30	0.703	0.826	0.953	0.875	0.954	0.957	0.958
100	50	0.817	0.926	0.974	0.927	0.975	0.975	0.975
100	100	0.915	0.975	0.987	0.964	0.988	0.988	0.988
150	30	0.712	0.842	0.954	0.874	0.953	0.956	0.957
150	50	0.844	0.945	0.975	0.927	0.975	0.975	0.975
150	100	0.930	0.982	0.988	0.964	0.988	0.988	0.988

Table 4.2: The Trace Ratio of the seven estimators for estimating F.
with $u = 0.1, \tau = 0, \psi = 0$ and $\rho_i \sim U[0.5, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.615	0.675	0.691	0.760	0.774	0.785	0.805
10	50	0.617	0.689	0.705	0.800	0.817	0.824	0.859
10	100	0.616	0.692	0.705	0.831	0.843	0.847	0.881
20	30	0.599	0.727	0.839	0.859	0.877	0.889	0.907
20	50	0.611	0.758	0.866	0.904	0.916	0.919	0.932
20	100	0.614	0.775	0.876	0.923	0.930	0.930	0.942
50	30	0.648	0.827	0.954	0.951	0.962	0.965	0.967
50	50	0.700	0.895	0.964	0.966	0.971	0.971	0.974
50	100	0.760	0.941	0.971	0.972	0.974	0.974	0.977
100	30	0.722	0.894	0.984	0.975	0.983	0.984	0.984
100	50	0.804	0.958	0.986	0.983	0.986	0.986	0.987
100	100	0.874	0.983	0.987	0.986	0.987	0.987	0.988
150	30	0.751	0.913	0.990	0.981	0.988	0.989	0.989
150	50	0.855	0.977	0.991	0.988	0.991	0.991	0.991
150	100	0.914	0.990	0.992	0.991	0.992	0.992	0.992

Table 5.1: The Trace Ratio of the seven estimators for estimating Λ.
with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0.5, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.702	0.740	0.778	0.737	0.791	0.813	0.844
10	50	0.734	0.775	0.813	0.784	0.838	0.856	0.883
10	100	0.741	0.782	0.819	0.800	0.850	0.863	0.893
20	30	0.628	0.687	0.797	0.730	0.820	0.855	0.894
20	50	0.664	0.729	0.839	0.798	0.880	0.905	0.934
20	100	0.708	0.771	0.876	0.865	0.929	0.943	0.960
50	30	0.617	0.709	0.890	0.772	0.884	0.910	0.921
50	50	0.682	0.777	0.938	0.855	0.944	0.952	0.954
50	100	0.763	0.849	0.964	0.921	0.973	0.975	0.976
100	30	0.634	0.755	0.914	0.786	0.902	0.919	0.922
100	50	0.744	0.864	0.954	0.866	0.951	0.954	0.955
100	100	0.849	0.934	0.977	0.929	0.977	0.977	0.978
150	30	0.645	0.776	0.918	0.788	0.907	0.921	0.923
150	50	0.770	0.897	0.955	0.866	0.952	0.955	0.955
150	100	0.884	0.960	0.977	0.930	0.977	0.977	0.978

Table 5.2: The Trace Ratio of the seven estimators for estimating F.
with $u = 0.1, \tau = 0.7, \psi = 0.5$ and $\rho_i \sim U[0.5, 0.9]$

		PC Class			ML Class			
N	T	PC	PC-GLS	PC-ITE	QMLE	ML-GLS	ML-ITE	ML-EM
10	30	0.569	0.601	0.597	0.655	0.660	0.664	0.659
10	50	0.557	0.593	0.589	0.658	0.666	0.670	0.672
10	100	0.538	0.579	0.578	0.652	0.661	0.664	0.678
20	30	0.574	0.655	0.735	0.779	0.792	0.805	0.824
20	50	0.561	0.655	0.748	0.798	0.813	0.824	0.859
20	100	0.562	0.669	0.767	0.828	0.844	0.851	0.885
50	30	0.651	0.792	0.933	0.920	0.935	0.947	0.956
50	50	0.662	0.827	0.949	0.947	0.958	0.961	0.964
50	100	0.700	0.875	0.954	0.961	0.965	0.966	0.969
100	30	0.722	0.873	0.979	0.964	0.976	0.980	0.981
100	50	0.783	0.940	0.984	0.978	0.983	0.984	0.985
100	100	0.840	0.971	0.986	0.984	0.986	0.986	0.987
150	30	0.751	0.897	0.987	0.975	0.985	0.987	0.988
150	50	0.838	0.967	0.990	0.986	0.990	0.990	0.991
150	100	0.903	0.988	0.991	0.990	0.991	0.991	0.992