Large Dimensional Factor Models with a Multi-Level Factor Structure: Identification, Estimation and Inference

Peng Wang *

November 12, 2008

Abstract

This paper develops an econometric theory for large dimensional factor models with a multi-level factor structure. Such a multi-level feature arises in a wide range of economic applications, such as finance, labor economics and international economics. For example, in labor economics, households can be divided into different income groups, each group facing both economy-wide common risk and group-specific risk. The baseline model is a two-level factor model, where factors are interpreted as unobserved economic shocks and categorized into two types: one is pervasive, affecting all economic sectors; the other is nonpervasive, affecting only a specific economic sector. Under these assumptions, the resulting large dimensional factor model has two features that are different from the usual model: (i) a large number of zero restrictions are imposed on factor loadings; (ii) the number of factors grows with the number of sectors. I provide a minimal set of identifying conditions of these two types of factors, as well as effective estimation methods. The estimators, which are jointly determined by a set of eigenvector problems, are shown to be consistent and have a normal limiting distribution. Finally, I apply the model to investigate different patterns of comovement within real and financial sectors respectively. Empirical results suggest that comovement within each sector is largely sector specific and the pervasive common factors play only a limited role.

Key Words Large dimensional factor models, multi-level factor structure, common factor, sector-specific factor

*I would like to thank the members of my thesis committee Jushan Bai, Thomas Sargent, and Joerg Stoye for their constant guidance and support. I am also grateful to David Backus, Christopher Flinn, Ahu Gemici, Mark Gertler, Stefan Hoderlein, Nazgul Jenish, Boyan Jovanovic, John Leahy, Sydney Ludvigson, Virgiliu Midrigan, James Ramsey, Gianluca Violante, Yi (Daniel) Xu along with the seminar participants at NYU for helpful discussions and comments. All remaining errors are mine. Department of economics, New York University, New York, NY, 10012. Email: peng.wang@nyu.edu

1 Introduction

This paper provides an econometric theory for analyzing large dimensional factor models with a multi-level factor structure. Such a multi-level feature arises naturally in a wide range of economic applications. For example, in labor economics, a panel of households can be divided into several income groups. Each income group faces both the economy-wide common risk and the group-specific risk, being understood as the top-level factor and the sub-level factor respectively. In international economics, the global economy consists of the industrialized economy and the emerging economy, each being understood as an economic group of the world economy. Further more, both groups include a large number of countries, each country being a subsector of either economic group. The global common shocks, or the top-level factors, have impact on all countries, while one group-specific shock, or a level-2 factor, only directly affects one particular economic group. A country-specific shock, or a level-3 factor, only has direct impact on one country.

Factor model itself as a dimension reduction tool has been widely applied in various economic fields. Inferential theory concerning explanatory static factor models of large dimensions has been derived for the computationally simple principal components estimators (Bai, 2003), where the model is estimated under a set of exact identifying restrictions[1]. Currently, neither computationally simple estimators nor inferential theory for large dimensional factor model is available when extra restrictions are present. The multi-level factor model is a special restricted model, which differs from the conventional explanatory factor model in two aspects.

Firstly, the multi-level factor structure implies lots of zero restrictions on the factor loadings. When the number of variables is small, such models are objectives of confirmatory factor analysis, where maximum likelihood estimator (MLE) is proposed and inferential theory is available (see Geweke and Singleton 1981). However, under a large N and large T setup, the dimension of parameters regarding factor loadings is of the order $O(N)$, which makes MLE computationally intensive. Another issue with MLE is that distributional assumptions are often made for both the factors and idiosyncratic terms, and it is still an open question whether misspecification matters for the inference under such a large N and large T setup. In particular, existing identification method requires sector-specific factors are orthogonal to each other. Instead, I provide a minimal set of identifying conditions, which does

[1] For large dimensional dynamic factor models, see Forni, Hallin, Lippi and Reichlin (2000, 2001, 2004, 2005), Forni, Giannone, Lippi and Reichlin (2003), Doz, Giannone and Reichlin (2007), Stock and Watson (1998, 2002a, 2002b, 2005). A nice survey of the literature is given by Bai and Ng (2008).

not impose the orthogonality restrictions between sector-specific factors. Furthermore, such a orthogonality assumption is testable using the inferential theory derived in this paper.

Secondly, the number of factors grows with the number of sectors. When we add more variables into a model by adding more sectors, we are also expanding the factor space because new sectors bring new sector-specific shocks into the model. While in the conventional setup, the number of factors is always a fixed number. The multi-level factor structure allows the number of factors to grow without bound as the number of sectors increases to infinity.

The multi-level factor structure is used to characterize how different shocks affect different range of economic variables. The baseline model considered in this paper is a two-level factor model, consisting of several or many parallel economic sectors. Within each sector, we observe a large number of time series. Factors are interpreted as unobserved economic shocks, and are categorized into two types: one is the pervasive top-level factor, or the common factor, affecting every individual time series across all economic sectors; the other is the nonpervasive sub-level factor, or the sector-specific factor, affecting only one particular sector. Let sector s be a specific subsector. Let x_{it}^s be the i^{th} variable of sector s observed at time t, then the baseline model has the following representation,

$$\begin{aligned} x_{it}^s &= \gamma_i^{s\prime} G_t + \lambda_i^{s\prime} F_t^s + e_{it}^s,\ i = 1, ..., N_s, s = 1, ..., S, \\ G_t &: \ \text{common shock, an } r \times 1 \text{ vector}, \\ F_t^s &: \ \text{shock specific to sector } s, \text{ an } r_s \times 1 \text{ vector}, \\ N_s &: \ \text{number of time series within sector } s, \\ N &= N_1 + ... + N_S : \ \text{total number of time series}, \\ S &: \ \text{number of sectors}, \end{aligned} \tag{1}$$

where the exposure to common shocks and sector-specific shocks for individual i in sector s is captured by γ_i^s and λ_i^s respectively.

We can also write down the above model in a vector form to compare with the conventional factor model,

$$\begin{bmatrix} x_t^1 \\ x_t^2 \\ ... \\ x_t^S \end{bmatrix} = \begin{bmatrix} \Gamma^1 & \Lambda^1 & ... & 0 & 0 \\ \Gamma^2 & 0 & \Lambda^2 & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S & ... & 0 & 0 & \Lambda^S \end{bmatrix} \begin{bmatrix} G_t \\ F_t^1 \\ ... \\ F_t^S \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \\ ... \\ e_t^S \end{bmatrix}. \tag{2}$$

This representation makes clear its difference from conventional factor models: (i) lots of zero restrictions are imposed on factor loadings; (ii) the number of factors grows with the number of sectors. I leave the discussion of identification strategies and efficient estimation methods to the next section.

Recently, Boivin and Ng (2006) use empirical assessment to argue that a large N not necessarily helps the estimation of common factors, due to potential existence of strong cross-sectional correlations. The multi-level factor model deals with this problem from a special angle, using sector-specific factor to capture cross-correlation within one sector not explained by common factors, which in turn helps estimation of the common factors. It is worth mentioning that, when the number of sectors is large, the information criteria in Bai and Ng (2002) is not able to consistently estimate the total number of factors. This is because the sector-specific factors are not pervasive enough to be counted as economy-wide common factors, and the rank condition for factor loadings in Bai and Ng (2002)'s Assumption B is not satisfied when the number of sectors is large. To handle this problem, I provide a two-step procedure to consistently estimate the number of both common factors and sector-specific factors. Then a $\sqrt{N}-$consistent estimator for the common factor is proposed, which is not possible from methods using less observations.[2]

In general, the economic sectors are differentiated into K hierarchical levels, with each level containing many units, any next level sector being a subsector for one of those units. Level-1 sector is assumed to have only one unit, which includes all the time series considered. Any unit in level $k+1$ sector is a subsector of a specific unit within level k sector. To see an example, the whole world is the level-1 sector. Region is the level-2 sector, consisting of two units, the industrialized economy and the emerging economy, each being a subsector of the world. Country is the level-3 sector. US is a unit in the level-3 sector, and is a subsector of the northern economy. Likewise, we may define country's industry as the level-4 sector, consisting of units such as automobile and agriculture industry within one country, so on and so forth.[3] Accordingly, Factors are interpreted as unobserved economic shocks, and are categorized into different levels ordered from 1 to K. For example, a level-2 factor will only directly affect economic variables within one specific level-2 sector. By assumption, level-1

[2] For example, if sector-specific factors are uncorrelated with each other and independent of common factors, then we may consistently estimate common factors using a subsample, consisting of one time series from each subsector. However, the estimated common factors are only $\min(\sqrt{S}, T)-$consistent as proved in Bai (2003).

[3] An important feature of the model in this paper is to allow level-k factors to be correlated across units within level-k sector. This is one feature generally not allowed in the state space approach to small dimensional multi-level factor models. (Kose, et al., 2003)

factors affect all economic variables.

Next, we provide an example to further illustrate economic environments where the multi-level factor structure will present.

Example (Serial correlated factors): *Assume a world with only two countries, home and foreign. Suppose home country's technology shock a_t affects a vector of contemporaneous home country's variables x_t, while only affects the foreign country's variables x_t^* with a lag. And vice versa for the foreign country's technology shock a_t^*. The technology shocks follow an AR(1) process with i.i.d. error terms ε_t and ε_t^**

$$a_t = \rho a_{t-1} + \varepsilon_t \text{ and } a_t^* = \rho^* a_{t-1}^* + \varepsilon_t^*.$$

Assume a linear model for x_t and x_t^* with i.i.d. error terms e_t and e_t^*

$$\begin{aligned} x_t &= \phi_1 a_t + \phi_2 a_{t-1}^* + e_t, \\ x_t^* &= \phi_1^* a_t^* + \phi_2^* a_{t-1} + e_t^*. \end{aligned}$$

Combined with the AR(1) process for technology shocks, the above model can be rewritten as

$$\begin{aligned} x_t &= \phi_1(\rho a_{t-1} + \varepsilon_t) + \phi_2 a_{t-1}^* + e_t = \phi_1 \rho a_{t-1} + \phi_2 a_{t-1}^* + \phi_1 \varepsilon_t + e_t, \\ x_t^* &= \phi_1^*(\rho^* a_{t-1}^* + \varepsilon_t^*) + \phi_2^* a_{t-1} + e_t^* = \phi_2^* a_{t-1} + \phi_1^* \rho^* a_{t-1}^* + \phi_1^* \varepsilon_t^* + e_t^*, \end{aligned}$$

or in the vector form

$$\begin{bmatrix} x_t \\ x_t^* \end{bmatrix} = \begin{bmatrix} \phi_1 \rho & \phi_2 & \phi_1 & 0 \\ \phi_2^* & \phi_1^* \rho^* & 0 & \phi_1^* \end{bmatrix} \begin{bmatrix} a_{t-1} \\ a_{t-1}^* \\ \varepsilon_t \\ \varepsilon_t^* \end{bmatrix} + \begin{bmatrix} e_t \\ e_t^* \end{bmatrix}.$$

In this case, the global factor is defined as $G_t = [a_{t-1}, a_{t-1}^*]'$, while country specific factor for home and foreign countries are defined as $F_t = \varepsilon_t$ and $F_t^* = \varepsilon_t^*$ respectively. Denote the factor loadings as $\Gamma = [\phi_1 \rho, \phi_2]$, $\Gamma^* = [\phi_2^*, \phi_1^* \rho^*]$, $\Lambda = \phi_1$ and $\Lambda^* = \phi_1^*$, then the model can be represented as model (2) with a 2-level factor structure

$$\begin{bmatrix} x_t \\ x_t^* \end{bmatrix} = \begin{bmatrix} \Gamma & \Lambda & 0 \\ \Gamma^* & 0 & \Lambda^* \end{bmatrix} \begin{bmatrix} G_t \\ F_t \\ F_t^* \end{bmatrix} + \begin{bmatrix} e_t \\ e_t^* \end{bmatrix}.$$

By model assumption, G_t is uncorrelated with $[F_t, F_t^*]'$. This is also an example where a special form of dynamic factor model can be reinterpreted as a static factor model with a multi-level factor structure.

2 The Multi-Level Factor Structure and Model Assumptions

I first briefly review the conventional static factor model setup. A static factor model for $\{\{x_{it}\}_{i=1}^N\}_{t=1}^T$ is given by[4]

$$x_{it} = \lambda_i' f_t + e_{it}, \tag{3}$$

where x_{it} is the observation for individual i at time t, factor f_t is assumed to be common to all individuals, and factor loading λ_i is individual i's specific response to the common factor, e_{it} is the idiosyncratic error term, or the part of x_{it} not explained by the common component $\lambda_i' f_t$. The number of factors r, or the dimension of the vector f_t, is assumed to be known for simplicity.

Recall that in our baseline model with a two-level factor structure, x_{it}^s is the i^{th} observation of sector s at time t, which admits the representation given by equation (1):

$$x_{it}^s = \gamma_i^{s\prime} G_t + \lambda_i^{s\prime} F_t^s + e_{it}^s,\ i = 1, ..., N_s, s = 1, ..., S,$$

where G_t is the pervasive factor, which affects all sectors. F_t^s is the nonpervasive factor, which only affects sector s. It is convenient to express the model as $N-$dimensional time series with T observations:

$$x_t^s = \Gamma^s G_t + \Lambda^s F_t^s + e_t^s,\ s = 1, ..., S,\ x_t^s \text{ is } N_s \times 1. \tag{4}$$

Define

$$x_t \equiv \begin{bmatrix} x_t^1 \\ ... \\ x_t^S \end{bmatrix}, \Gamma = \begin{bmatrix} \Gamma^1 \\ ... \\ \Gamma^S \end{bmatrix}, \Lambda^F = \begin{bmatrix} \Lambda^1 & 0 & ... & 0 \\ 0 & \Lambda^2 & 0 & ... \\ ... & ... & ... & ... \\ 0 & ... & 0 & \Lambda^S \end{bmatrix}, F_t = \begin{bmatrix} F_t^1 \\ ... \\ F_t^S \end{bmatrix}, e_t \equiv \begin{bmatrix} e_t^1 \\ ... \\ e_t^S \end{bmatrix}, \tag{5}$$

[4] Or in matrix form, $X = F\Lambda' + E$, with dimension of X, F, Λ being $T \times N$, $T \times r$,and $N \times r$ respectively. Both N and T are assumed to be large and are allowed to increase to infinity.

then the model can be represented by

$$x_t \equiv \begin{bmatrix} x_t^1 \\ ... \\ x_t^S \end{bmatrix} = \Gamma G_t + \Lambda^F F_t + e_t = [\Gamma, \Lambda^F] \begin{bmatrix} G_t \\ F_t \end{bmatrix} + e_t, \tag{6}$$

where Λ^F is block diagonal. If $\Lambda^F F_t$ is known, then Γ and G_t are obtained based on data from all countries using the pure static factor model $x_t - \Lambda^F F_t = \Gamma G_t + e_t$. If ΓG_t is known, Λ^s and F_t^s are obtained using data from sector s from the pure static factor model $x_t^s - \Gamma^s G_t = \Lambda^s F_t^s + e_t^s$ using principal components method. However, we do not directly observe G_t or F_t, which must be jointly inferred from data.

The estimator for $(\gamma_i^s, G_t, \lambda_i^s, F_t^s)$ we considered in this paper is the one which minimizes sum of squared residuals

$$\sum_{t=1}^{T} \sum_{s=1}^{S} \sum_{i=1}^{N_s} \left(x_{it}^s - \gamma_i^{s\prime} G_t - \lambda_i^{s\prime} F_t^s\right)^2 = \sum_{t=1}^{T} \left(x_t - \Gamma G_t - \Lambda^F F_t\right)' \left(x_t - \Gamma G_t - \Lambda^F F_t\right),$$

subject to some identifying assumptions for $(\gamma_i^s, G_t, \lambda_i^s, F_t^s)$, which we provide in the next section. Notice that Λ^F is a block diagonal matrix, which provides a large number of zero restrictions. Moreover, the number of sector-specific factors, or the dimension of F_t, grows with the number of sectors, and thus we must specify the asymptotic behavior of the number of subsectors S when deriving large sample theory for F_t and G_t. In sum, we need to derive a new asymptotic theory for estimators defined above.

Let $||A|| = [tr(A'A)]^{1/2}$ denote the norm of matrix A. The following assumptions are extensions of Bai and Ng (2002) and Bai (2003) to the multi-level factor model, which are needed to prove the consistency and large sample theory for the estimators.

Assumption A (Factors): $E||G_t||^4 \leq M < \infty$, $T^{-1} \sum_{t=1}^{T} G_t G_t' \xrightarrow{p} \Sigma_G$ for some $r \times r$ positive definite matrix Σ_G. $E||F_t^s||^4 \leq M < \infty$, $T^{-1} \sum_{t=1}^{T} F_t^s F_t^{s\prime} \xrightarrow{p} \Sigma_{F^s}$ for some $r_s \times r_s$ positive definite matrix Σ_{F^s}, $s = 1, ..., S$. Define $H_t = [G_t', F_t']'$. When S is fixed, assume that $T^{-1} \sum_{t=1}^{T} H_t H_t' \xrightarrow{p} \Sigma_H$ for some positive definite matrix Σ_H with rank $r + r_1 + ... + r_s$. When $S \to \infty$, assume that $\text{plim}_{(T,S)\to\infty} \frac{\mu_{\max}}{\mu_{\min}} < c$ for some constant $c > 0$, where $\mu_{\max}$ and $\mu_{\min}$ are the largest and smallest eigenvalues of $T^{-1} \sum_{t=1}^{T} H_t H_t'$ respectively.

Assumption B (Factor loadings): $||\gamma_i^s|| \leq \bar{\gamma} < \infty$, $||\lambda_i^s|| \leq \bar{\gamma} < \infty$, $||\Lambda^{s\prime}\Lambda^s / N_s - \Sigma_{\Lambda^s}|| \to 0$ for some $r \times r$ positive definite matrix Σ_{Λ^s}, and $||\Gamma^{s\prime}\Gamma^s / N_s - \Sigma_{\Gamma^s}|| \to 0$ for some $r \times r$ positive definite matrix Σ_{Γ^s}, $s = 1, ..., S$, and $||\Gamma'\Gamma / N - \Sigma_\Gamma|| \to 0$ for some $r \times r$ positive definite matrix $\Sigma_\Gamma = \lim_{S\to\infty} \frac{1}{S}\sum_{s=1}^{S} \Sigma_{\Gamma^s}$. Further, $\text{rank}\left(\begin{bmatrix} \Gamma^s & \Lambda^s \end{bmatrix}\right) = r_s + r$.

Assumption C (Time and Cross-Section Dependence and Heteroskedasticity): There exists a positive constant $M < \infty$ such that for all i, s and t :

1. $E(e_{it}^s) = 0$, $E(e_{it}^s)^8 \leq M$.
2. $E(e_k' e_t/N) = E(\frac{1}{N}\sum_{s=1}^{S}\sum_{i=1}^{N_s} e_{ik}^s e_{it}^s) = \gamma_N(k,t)$, $|\gamma_N(t,t)| \leq M$ for all t, and

$$\frac{1}{T}\sum_{k=1}^{T}\sum_{t=1}^{T} |\gamma_N(k,t)| \leq M.$$

3. $E(e_{it}^{s_1} e_{jt}^{s_2}) = \tau_{ij,t}^{s_1 s_2}$, with $|\tau_{ij,t}^{s_1 s_2}| \leq \tau_{ij}^{s_1 s_2}$ for some $\tau_{ij}^{s_1 s_2} \geq 0$ and for all t. Moreover

$$\frac{1}{N}\sum_{s_1=1}^{S}\sum_{s_2=1}^{S}\sum_{i=1}^{N_{s_1}}\sum_{j=1}^{N_{s_2}} \tau_{ij}^{s_1 s_2} \leq M.$$

4. $E(e_{ik}^{s_1} e_{jt}^{s_2}) = \tau_{ij,kt}^{s_1 s_2}$ and $(NT)^{-1}\sum_{s_1=1}^{S}\sum_{s_2=1}^{S}\sum_{i=1}^{N_{s_1}}\sum_{j=1}^{N_{s_2}}\sum_{k=1}^{T}\sum_{t=1}^{T} |\tau_{ij,kt}^{s_1 s_2}| \leq M$.
5. For every (k,t), $E|N^{-1/2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}[e_{ik}^s e_{it}^s - E(e_{ik}^s e_{it}^s)]|^4 \leq M$.

Assumption D (Weak dependence between factors and idiosyncratic errors):

$$\begin{aligned} E\left(\frac{1}{N_s}\sum_{i=1}^{N_s}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t^s e_{it}^s\right\|^2\right) &\leq M, s = 1, ..., S, \\ E\left(\frac{1}{N}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t e_{it}^s\right\|^2\right) &\leq M. \end{aligned}$$

Assumption A allows factors to be arbitrary stationary autoregressive processes, while the relationship between factors and x_{it}^s is still static. When a finite number of lagged factors also affect x_{it}^s, we can always redefine a new factor as a vector of current and lagged original factors, such that the relationship between the newly defined factor and x_{it}^s is static. For example, if we have the following dynamic factor model

$$x_t^s = \Gamma_1^s G_t + \Gamma_2^s G_{t-1} + \Lambda_1^s F_t^s + \Lambda_2^s F_{t-1}^s + e_t^s.$$

We may redefine a new global factor $\tilde{G}_t = [G_t, G_{t-1}]'$ and new sector-specific factor $\tilde{F}_t^s = [F_t^s, F_{t-1}^s]'$, such that a new static factor model is obtained

$$x_t^s = \Gamma^s \tilde{G}_t + \Lambda^s \tilde{F}_t^s + e_t^s,$$

where the new factor loadings are defined as $\Gamma^s = [\Gamma_1^s, \Gamma_2^s]$ and $\Lambda^s = [\Lambda_1^s, \Lambda_2^s]$. Thus we may focus on the static factor model, while the derived properties still hold for dynamic factor specification with a finite number of lagged factors directly affecting x_{it}^s. The rank condition for H_t rules out the possibility that different factors are perfectly correlated.

Assumption B guarantees that each global factor G_{mt} has a nontrivial contribution to the variance of x_t, $m = 1, ..., r$, while each sector-specific factor F_{jt}^s has a nontrivial contribution to the variance of x_t^s, $j = 1, ..., r_s$. Thus G_t is pervasive to all variables, while F_t^s is only pervasive within sector s. Further, the rank condition, $\text{rank}\left(\begin{bmatrix} \Gamma^s & \Lambda^s \end{bmatrix}\right) = r_s + r$, guarantees enough heterogeneity among individual variables within sector s when responding to both factors. This rank condition is crucial for separate identification of G_t and F_t^s. For example, the following model is not identified without further assumptions,

$$\begin{aligned} x_{it}^1 &= G_t + F_t^1 + e_{it}^1, \;\; i = 1, ..., N_1, \\ x_{jt}^2 &= G_t + F_t^2 + e_{jt}^2, \;\; j = 1, ..., N_2. \end{aligned}$$

Assumption C allows for limited time series and cross section dependence, as well as heteroskedasticities in both the time and cross-section dimensions in the idiosyncratic errors. The cross-section correlation in the idiosyncratic errors allows the model to have an approximate factor structure as in Chamberlain and Rothschild (1983), in contrast to the conventional strict factor model where idiosyncratic errors are uncorrelated across section. Moreover, assumption C is more general than the approximate factor model defined in Chamberlain and Rothschild (1983), because heteroskedasticity in the time dimension is also allowed.

When deriving the large sample theory, I assume that the numbers of factors r, r_s, $s = 1, ..., S$ are fixed and known. When the number of factors is unknown, we may apply a two step procedure to select the number of factors. In step 1, Bai and Ng (2002)'s information criteria is applied to each sector to obtain $\widehat{(r + r_s)}$, $s = 1, ..., S$. Then, combining any two sectors, say $s = 1$ and j, we may use the information criteria again to estimate the dimension of $[G_t', F_t^{1\prime}, F_t^{j\prime}]'$, with the estimator given by $\widehat{r + r_1 + r_j}$ for $j = 2, ..., S$. In the second step, define

$$\hat{r} = \min_{\{j=2,...,S\}} \{\widehat{r + r_i} + \widehat{r + r_j} - \widehat{r + r_i + r_j}\}. \tag{7}$$

Then the resulting $\hat{r}$ is a consistent estimator for the dimension of global factors. And we may consistently estimate r_s by $\hat{r}_s = \widehat{(r + r_s)} - \hat{r}$.

It is worth mentioning that $\widehat{(r + r_1 + r_2)}$ is not necessarily consistent for the true $r + r_1 + r_2$.

The reason is that a factor common to two sectors is not necessarily common to all sectors. One example is the regional effect. The shocks common to the North America region are not necessarily pervasive enough to be counted as global shocks. Thus in a two level sector setup, such regional effects, if not directly affecting countries out of the North America region, should be regarded as factors specific to countries in North America. In fact, we may only prove that $\text{plim}\{(\widehat{r+r_1+r_2}) \leq r+r_1+r_2\} = 1$.

We leave the efficient selection of the number of factors to future research, where $r, r_1, ..., r_S$ are jointly estimated. It is also worth mentioning that the asymptotic theory for factors and factor loadings is not affected if the number of factors is estimated. This is shown in the footnote 5 in Bai (2003). In this case, the following stronger assumption is needed.

Assumption E (Weak Dependence): For all k, t, j, s_2, T and N
1. $\sum_{k=1}^{T} |\gamma_N(k,t)| \leq M < \infty$.
2. $\sum_{s_1=1}^{S} \sum_{i=1}^{N_{s_1}} \tau_{ij}^{s_1 s_2} \leq M < \infty$

This assumption is stronger than assumption C2 and C3, but is still very general.

3 Identification of the Multi-Level Factor Model

The objective of this section is to find restrictions on the model, such that (i) the model is uniquely identified under such normalization, and (ii) common factors and sector-specific factors are separately identified. Notice that (i) can be achieved by imposing extra restrictions on factor loadings only, while the resulting factor estimators are lack of economy explanations. We treat (ii) as an important issue, because it allows us to cast economic meanings and examine the interaction of different factors.

Although the multi-sector factor model imposes a large number of zero restrictions on factor loadings, common factors and sector-specific factors are not separately identified, unless we made further model assumptions about correlations between G_t and F_t. To see a simple example, notice that the data generating process $x_t^s = \Gamma^s G_t + \Lambda^s F_t^s + e_t^s$ is observationally equivalent to $x_t^s = \tilde{\Gamma}^s G_t + \Lambda^s \tilde{F}_t^s + e_t^s$, where $\tilde{\Gamma}^s = \Gamma^s - \Lambda^s R^s$ and $\tilde{F}_t^s = F_t^s + R^s G_t$, R^s being an arbitrary $r_s \times r$ matrix. Thus we make the following model assumptions such that the sector-specific factors is separately identified from common factors.

Assumption F: If factors have zero mean, assume $\sum_{t=1}^{T} G_t F_t^{s\prime} = 0$ for $s = 1, ..., S$. If factors have nonzero mean, assume $\frac{1}{T}\sum_{t=1}^{T} G_t F_t^{s\prime} - [\frac{1}{T}\sum_{t=1}^{T} G_t][\frac{1}{T}\sum_{t=1}^{T} F_t^{s\prime}] = 0$.

The population correspondent of Assumption F is $Cov(G_t, F_t^s) = 0$ for $s = 1, ..., S$. We assume assumption F holds throughout. The above assumptions rule out the possibility that common factors contains information about sector-specific factors.

In general, consider a transformation(rotation) matrix of the following form

$$R = \begin{bmatrix} A & 0 & ... & 0 \\ B_1 & A_1 & ... & ... \\ ... & ... & ... & ... \\ B_S & 0 & ... & A_S \end{bmatrix}$$

for any A, $A_j, j = 1, ..., S$ full rank and B_j conformable. A multi-sector factor model scaled by the rotation matrix R will retain the zero restrictions on factor loadings and still be observationally equilvalent to the original one. For example

$$\begin{bmatrix} \Gamma^1 & \Lambda^1 & 0 & ... & 0 \\ \Gamma^2 & 0 & \Lambda^2 & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S & ... & 0 & 0 & \Lambda^S \end{bmatrix} \begin{bmatrix} A & 0 & ... & 0 \\ B_1 & A_1 & ... & ... \\ ... & ... & ... & ... \\ B_S & 0 & ... & A_S \end{bmatrix} = \begin{bmatrix} \Gamma^1 A + \Lambda^1 B_1 & \Lambda^1 A_1 & 0 & ... & 0 \\ \Gamma^2 A + \Lambda^2 B_2 & 0 & \Lambda^2 A_2 & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S A + \Lambda^S B_S & ... & 0 & 0 & \Lambda^S A_S \end{bmatrix}$$

This implies that we need at least $r^2 + (r_1 + ... + r_S)r + r_1^2 + ... + r_S^2$ more restrictions to make the model identified. If there is no structural restrictions from economic theory to achieve that amount, we need some normalizations to obtain a unique solution for both factor loadings and factors.

Recall that for a pure static factor models of the same dimension as the above one, the number of restrictions needed for identifying the model is

$$(r + r_1 + ... + r_S)^2$$

while in the multi-sector factor model, the number of restrictions implied by zero restrictions on the factor loadings is

$$(S - 1) \cdot (N_1 r_1 + ... + N_S r_S)$$

Notice that $(r + r_1 + ... + r_S)^2 \leq S \cdot (r^2 + r_1^2 + ... + r_S^2)$. Although this upper bound is much smaller than $(S - 1) \cdot (N_1 r_1 + ... + N_S r_S)$, the model still lacks identification.

Definition 1: Within sector identification. *If $\Gamma^s G_t$ is known for $s = 1, ..., S$, Λ^s and F_t^s in the model $x_t^s - \Gamma^s G_t = \Lambda^s F_t^s + e_t^s$ are uniquely identified.*

Definition 2: Between sector identification. *If $\Lambda^s F_t^s$ is given for $s = 1, ..., S$, then Γ and G_t are uniquely identified from the model $x_t^s - \Lambda^s F_t^s = \Gamma^s G_t + e_t^s$.*

Remark: *Between sector identification requires r^2 more restrictions, given the sector-specific common components, while within sector need $r_1^2 + ... + r_S^2$ more given the common components. The orthogonality between common and sector-specific factors impose $(r_1 + ... + r_S)r$ restrictions. We expect these restrictions will uniquely pin down the above rotation matrix as an identify matrix.*

Proposition 1: *Given the rank conditions in Assumption B, the factor loadings for the multi-level factor model (2) are identified up to a linear transformation of the following form,*

$$R^* = \begin{bmatrix} A_{00} & 0 & ... & 0 \\ A_{10} & A_{11} & ... & 0 \\ ... & ... & ... & ... \\ A_{S0} & 0 & ... & A_{SS} \end{bmatrix}$$

where A_{ij} is any $r_i \times r_j$ matrix with $r_0 = r$. It means that the factor loadings in model (6), after being multiplied by the matrix R^*, will preserve the same zero restrictions. Common factors can be identified up to an $r \times r$ transformation, while the sector-specific factors can only be identified as a linear combination of common factors and original sector-specific factors. If we further assume Assumption F holds, then the space spanned by columns of G and the space spanned by columns of F^s are separately identified, $s = 1, ..., S$.

Proof of proposition 1: Assume a rotation matrix R will preserve the same zero restrictions, which means the structural form of the grand factor loading matrix will not change after being multiplied by R,

$$\begin{bmatrix} \Gamma^1 & \Lambda^1 & ... & 0 & 0 \\ \Gamma^2 & 0 & \Lambda^2 & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S & 0 & ... & 0 & \Lambda^S \end{bmatrix} \begin{bmatrix} A_{00} & A_{01} & ... & A_{0S} \\ A_{10} & A_{11} & ... & A_{1S} \\ ... & ... & ... & ... \\ A_{S0} & ... & ... & A_{SS} \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}^1 & \hat{\Lambda}^1 & ... & 0 & 0 \\ \hat{\Gamma}^2 & 0 & \hat{\Lambda}^2 & ... & 0 \\ ... & ... & ... & ... & 0 \\ \hat{\Gamma}^S & 0 & ... & 0 & \hat{\Lambda}^S \end{bmatrix}$$

$$\text{with} \begin{bmatrix} A_{00} & A_{01} & ... & A_{0S} \\ A_{10} & A_{11} & ... & A_{1S} \\ ... & ... & ... & ... \\ A_{S0} & ... & ... & A_{SS} \end{bmatrix} = R$$

Notice that A_{ij} is of dimension $r_i \times r_j$ and $r_0 \equiv r$. First consider the second block column of the transformed grand factor loading matrix, the restrictions implies $(j, 2) - th$ block of rotated loadings are zeros, $j = 2, ..., S$, in particular

$$\begin{aligned}
\Gamma^2 A_{01} + \Lambda^2 A_{21} &= 0, ..., \Gamma^S A_{01} + \Lambda^S A_{S1} = 0 \\
\begin{bmatrix} \Gamma^2 & \Lambda^2 \end{bmatrix} \begin{bmatrix} A_{01} \\ A_{21} \end{bmatrix} &= 0 \text{ implies } \begin{bmatrix} A_{01} \\ A_{21} \end{bmatrix} = 0, \\
\text{Provided that rank}\left(\begin{bmatrix} \Gamma^2 & \Lambda^2 \end{bmatrix} \right) &= r_2 + r \\
\text{Similarly, rank}\left(\begin{bmatrix} \Gamma^s & \Lambda^s \end{bmatrix} \right) &= r_s + r \text{ implies } A_{s1} = 0, s = 2, ..., S
\end{aligned}$$

The rank condition for factor loadings is implied by model assumptions. Likewise, we have $A_{ij} = 0$, for $i \neq j$ and $j \neq 0$. Thus we can pin down the rotation matrix to have the following form

$$R^* = \begin{bmatrix} A_{00} & 0 & ... & 0 \\ A_{10} & A_{11} & ... & 0 \\ ... & ... & ... & ... \\ A_{S0} & 0 & ... & A_{SS} \end{bmatrix}$$

The above transformation implies the rotated factor loadings become

$$\begin{bmatrix} \Gamma^1 & \Lambda^1 & ... & 0 & 0 \\ \Gamma^2 & 0 & \Lambda^2 & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S & 0 & ... & 0 & \Lambda^S \end{bmatrix} \begin{bmatrix} A_{00} & 0 & ... & 0 \\ A_{10} & A_{11} & ... & 0 \\ ... & ... & ... & ... \\ A_{S0} & 0 & ... & A_{SS} \end{bmatrix} = \begin{bmatrix} \Gamma^1 A_{00} + \Lambda^1 A_{10} & \Lambda^1 A_{11} & ... & 0 & 0 \\ ... & 0 & \Lambda^2 A_{22} & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S A_{00} + \Lambda^S A_{S0} & 0 & ... & 0 & \Lambda^S A_{SS} \end{bmatrix}$$

The inverse of the rotation matrix R^* takes a special form as well, which is determine by the following formula

$$\begin{bmatrix} A_{00} & 0 & ... & 0 \\ A_{10} & A_{11} & ... & 0 \\ ... & ... & ... & ... \\ A_{S0} & 0 & ... & A_{SS} \end{bmatrix} \begin{bmatrix} B_{00} & B_{01} & ... & B_{0S} \\ B_{10} & B_{11} & ... & ... \\ ... & ... & ... & ... \\ B_{S0} & B_{S1} & ... & B_{SS} \end{bmatrix} = \begin{bmatrix} I_r & & & \\ & I_{r_1} & & \\ & & ... & \\ & & & I_{r_S} \end{bmatrix}$$

After solving the above equation, we obtain

$$(R^*)^{-1} = \begin{bmatrix} B_{00} & 0 & ... & 0 \\ B_{10} & B_{11} & ... & 0 \\ ... & ... & ... & ... \\ B_{S0} & 0 & ... & B_{SS} \end{bmatrix}$$

with $B_{jj} = A_{jj}^{-1}$. Apply this transformation on factors and we obtain

$$(R^*)^{-1}\begin{bmatrix} G_t \\ F_t^1 \\ ... \\ F_t^S \end{bmatrix} = \begin{bmatrix} B_{00} & 0 & ... & 0 \\ B_{10} & B_{11} & ... & 0 \\ ... & ... & ... & ... \\ B_{S0} & 0 & ... & B_{SS} \end{bmatrix}\begin{bmatrix} G_t \\ F_t^1 \\ ... \\ F_t^S \end{bmatrix} = \begin{bmatrix} B_{00}G_t \\ B_{10}G_t + B_{11}F_t^1 \\ ... \\ B_{S0}G_t + B_{SS}F_t^S \end{bmatrix}$$

We can see that common factor (up to an $r \times r$ rotation) is well identified, while sector-specific factors is mixed with common factors after the rotation. Recall the example that the data generating process $x_t^s = \Gamma^s G_t + \Lambda^s F_t^s + e_t^s$ is observationally equivalent to $x_t^s = \tilde{\Gamma}^s G_t + \tilde{\Lambda}^s \tilde{F}_t^s + e_t^s$, where $\tilde{\Gamma}^s = \Gamma^s - \Lambda^s (B_{ss})^{-1} B_{s0}, \tilde{\Lambda}^s = \Lambda^s (B_{ss})^{-1}$ and $\tilde{F}_t^s = B_{s0} G_t + B_{ss} F_t^s$. Without further assumptions, we can always redefine a new sector-specific factor as a linear combination of common factor and original sector-specific factor, such that the new model is observationally equivalent to the original one.

If we further assume assumption F holds, then $B_{s0} = 0$, $s = 1, ..., S$. The rotation matrix R^* becomes a block diagonal matrix, and then $B_{00}G_t$ and $B_{ss}F_t^s$ are separately identified. Thus the space spanned by columns of G and the space spanned by columns of F^s are separately identified. **Q.E.D.**

Suppose we have estimated both common factors and sector-specific factors, with $\hat{G}_t$ being an estimator for the true common factors up to a rotation while $\tilde{F}_t^s$ estimating a linear combination of rotated true common factors and rotated true sector-specific factors for sector s. After imposing the orthogonality assumption F, we can recover rotated sector-specific factors $\hat{F}_t^s$ based on the following regression

$$\tilde{F}_t^s = \hat{R}_s \hat{G}_t + \hat{F}_t^s$$

where $\hat{R}_s$ is the OLS estimator. If factors have nonzero mean, $\hat{F}_t^s$ cannot be treated just as residuals of a linear regression equation. Since the objective is to estimate $\hat{R}_s$, we can demean the above equation to obtain

$$\tilde{F}_t^s - \mu^s = \hat{R}_s(\hat{G}_t - \mu^G) + \hat{F}_t^s - \mu^F$$

where μ^s, μ^G and μ^F are sample average of corresponding variables. Then $\hat{R}_s = \{\sum_{t=1}^T (\tilde{F}_t^s - \mu^s)(\hat{G}_t - \mu^G)'\}\{\sum_{t=1}^T (\hat{G}_t - \mu^G)(\hat{G}_t - \mu^G)'\}^{-1}$ is consistent, because of the orthogonality assumption between common and sector-specific factors. The resulting $\hat{F}_t^s = \tilde{F}_t^s - \hat{R}_s \hat{G}_t$

will be an estimator for the true sector-specific factor up to a full rank $r_s \times r_s$ matrix transformation.

The original estimated model has the representation $x_t^s = \hat{\Gamma}^s \hat{G}_t + \hat{\Lambda}^s \hat{F}_t^s + \hat{e}_t^s$ which is equivalent to

$$x_t^s = (\hat{\Gamma}^s + \hat{\Lambda}^s \hat{R}_s)\hat{G}_t + \hat{\Lambda}^s(\hat{F}_t^s - \hat{R}_s \hat{G}_t) + \hat{e}_t^s$$

If theory suggests that upper square block of sector-specific factor loadings is identity matrix, we have

$$R^{-1}\begin{bmatrix} G_t \\ F_t^1 \\ ... \\ F_t^S \end{bmatrix} = \begin{bmatrix} B_{00} & 0 & ... & 0 \\ B_{10} & I_{r_1} & ... & 0 \\ ... & ... & ... & ... \\ B_{S,0} & 0 & ... & I_{r_S} \end{bmatrix} \begin{bmatrix} G_t \\ F_t^1 \\ ... \\ F_t^S \end{bmatrix} = \begin{bmatrix} B_{00}G_t \\ B_{10}G_t + F_t^1 \\ ... \\ B_{S,0}G_t + F_t^S \end{bmatrix}$$

Then the projection yields estimates for true sector-specific factors instead of a rotation.

Remark: *We achieve identification through assumptions on both factors and factor loadings. For example, we require $G'F^s = 0$ for all $s = 1, ..., S$, which pins down $B_{s,0} = 0$. As a by-product, G_t and F_t^s are separately identified. Adding the assumption that within and between sector identification is achieved, the model will be uniquely identified. For example, we may require that the upper $r \times r$ blocks of Γ^1, Λ^s are I_r and I_{r_s} respectively, $s = 1, ..., S$.*

3.1 Exact Identifying Restrictions

The following identification scheme imposes $r^2 + (r_1 + ... + r_S)r + r_1^2 + ... + r_S^2$ restrictions to make the multi-level factor model exactly identified, namely, the multi-level factor model structure coupled with our extra imposed restrictions uniquely pin down factors and factor loadings as parameters.

type	Summary of restrictions	# of restrictions
1	$\frac{G'G}{T} = I_r$ and $\Gamma'\Gamma$ diagonal	r^2
2	$\frac{F^{s\prime}F^s}{T} = I_{rs}$ and $\Lambda^{s\prime}\Lambda^s$ diagonal, $\forall s$	$r_1^2 + ... + r_S^2$
3	$G'F^s = 0, \forall s$	$(r_1 + ... + r_S)r$

Recall that

$$\begin{bmatrix} \Gamma^1 & \Lambda^1 & ... & 0 & 0 \\ \Gamma^2 & 0 & \Lambda^2 & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S & 0 & ... & 0 & \Lambda^S \end{bmatrix} \begin{bmatrix} A_{00} & 0 & ... & 0 \\ A_{10} & A_{11} & ... & 0 \\ ... & ... & ... & ... \\ A_{S,0} & 0 & ... & A_{S,S} \end{bmatrix} = \begin{bmatrix} \Gamma^1 A_{00} + \Lambda^1 A_{10} & \Lambda^1 A_{11} & ... & 0 & 0 \\ ... & 0 & \Lambda^2 A_{22} & ... & 0 \\ ... & ... & ... & ... & 0 \\ \Gamma^S A_{00} + \Lambda^S A_{S,0} & 0 & ... & 0 & \Lambda^S A_{S,S} \end{bmatrix}$$

The following chart is a summary of role played by each type of restrictions

$$\begin{array}{ccc} \text{type 2} & \Leftrightarrow & A_{ss} = I_{r_s}, s = 1, ..., S \\ \text{type 1 and 3} & \Longleftrightarrow & A_{10} = 0, ..., A_{S0} = 0, \; A_{00} = I_r \end{array}$$

We may see both within group identification and between group identification are necessary conditions for the identification of the model.

Remark: *Restrictions implied by model assumptions only involve zero blocks in the grand factor loading* $[\Gamma, \Lambda^F]$ *and orthogonality between common factors and sector-specific factors. All other restrictions we assume serve the purpose of 1) producing unique solution of the least squares problem, and 2) separately identifying common factors and sector-specific factors. Type 1 and type 2 restrictions are normalizations as in the standard analysis of static factor models. Type 3 restriction is not a normalization, but an indispensable additional model assumption such that the sector-specific factor is well defined instead of being mixed with the common factor.*

The existing identification scheme for small dimension models, such as in Kose, et al (2003), assumes not only sector-specific factors are uncorrelated with common factors, but sector-specific factors are mutually uncorrelated to each other. When the true sector-specific factors show certain degree of correlation, this identification scheme misspecifies the model, and it is still an open question whether the resulting estimators, explained as quasi maximum likelihood (QMLE) estimators, are consistent or not. The exact identifying restrictions assumed by this paper is immune to such misspecification problems, and thus will be able to provide valid information about dynamic properties as well as the correlation feature of the factors.

4 Estimation of the Multi-Level Factor Model

4.1 Maximum Likelihood Estimation

When we have small N and large T, maximum likelihood estimation method can be used. In particular, an EM algorithm can be easily derived following Anderson (1980). If in addition dynamic structure is imposed on factors, we can form a state space model with restrictions on its parameters, and use Kalman filter to compute the likelihood function, with Kalman smoother being used as estimators for factors. The latter approach was applied by Kose, et al (2003) to study global, regional and country-specific shock under an international business cycle context. However, when N is large, MLE involves a large number of parameters, imposing a great burden to the computation of the maximum.

Another alternative method is to apply Geweke and Singleton (1981)'s spectral density estimation, to deal with the restrictions imposed on the model. However, the large sample theory therein is set forth for fixed N and large T. When the number of cross section variables is large, inference should be based on both N and T going to infinity, where although sample covariance matrix is component-wise convergent to population covariance matrix, the overall convergence for the covariance matrix is not defined for the case with large N. When $N > T$, the sample covariance matrix is not a full rank matrix, however the population covariance matrix can always be of full rank.

One challenge for likelihood approach is that explicit dynamic processes and correlation assumptions need to be made for the whole factor vector $[G_t, F_t^1, ..., F_t^S]$, which would be a nontrivial parametrization task. When the number of sectors is large, it is more likely that sector-specific factors will show certain degree of cross correlation, and the correlation might be strong for a group of sector-specific factors.

This paper treat both factor loadings and factors as parameters of interest, and allow certain degrees of correlation between sector-specific factors. We also allow stochastic volatility to present in factors, as long as assumption A and D are satisfied. The asymptotic theory is based on asymptotic expansion for the nonlinear restricted least squares estimator, which is defined in the next section.

4.2 Least Squares Estimation

In the least squares estimation, one chooses $(\gamma_i^s, G_t, \lambda_i^s, F_t^s)$ to minimize the total sum of squared residuals

$$\min \sum_{t=1}^{T} \sum_{s=1}^{S} \sum_{i=1}^{N_s} (x_{it}^s - \Gamma_i^{s\prime} G_t - \Lambda_i^{s\prime} F_t^s)^2 \quad (8)$$

subject to the following three types of restrictions,

1) Within sector identification: $F^{s\prime}F^s/T = I_{r_s}$ and $\Lambda^{s\prime}\Lambda^s$ diagonal.

2) Between sector identification: $G'G/T = I_r$ and $\Gamma'\Gamma$ diagonal.

3) Separating common and sector-specific factors: $G'F = 0$.

The following theorem characterizes least squares solution under the above restrictions.

Theorem 1: *Assume 1) within sector identification, 2) between sector identification, 3) orthogonality between common factors and sector-specific factors. Let* $F = [F^1, F^2, ..., F^S]$, $X^s = [x_1^s, ..., x_T^s]'$. *Define* $A^s = X^s X^{s\prime}$ and $A = A^1 + ... + A^S$. *Assume rank*$(F) = r_1 + ... + r_S$, *then the least squares solution for factors is determined by the following eigenvector problem*

1) $\frac{1}{\sqrt{T}}\hat{G} = r$ eigenvectors for $P_{\hat{F}}A$ corresponding to its largest r eigenvalues,
2) $\frac{1}{\sqrt{T}}\hat{F}^s = r_s$ eigenvectors for $P_{\hat{G}}A^s$ corresponding to its largest r_s eigenvalues,
3) $P_{\hat{F}} = I_T - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'$
4) $P_{\hat{G}} = I_T - \hat{G}(\hat{G}'\hat{G})^{-1}\hat{G}' = I_T - \hat{G}\hat{G}'/T$

The estimators for factor loadings are given by $[\hat{\Gamma}^s, \hat{\Lambda}^s] = X^{s\prime}[\hat{G}, \hat{F}^s]/T$.

Proof of theorem 1: see appendix.

Remark: *The solution is quite intuitive. For example, the entire data sets contain information of global factors, resulting the use of* A, *and common factors are orthogonal to all sector-specific factors, resulting the use of projection matrix* P_F *to eliminate information of sector-specific factors contained in* A. *Likewise, only sector* s *contains information of sector-specific factor* F^s, *resulting the use of only* $A^s = X^s X^{s\prime}$, *and the projection matrix* P_G *removes information regarding common factors from* A^s.

Remark: *Iterative Principal Component Analysis (Iterative PCA): It is easy to prove that an equivalent characterization of the least squares estimators is given by*

1) $[\hat{\Gamma}, \hat{G}]$ is principal components estimator for $z_t = \Gamma G_t + v_t$
2) $[\hat{\Lambda}^s, \hat{F}^s]$ is principal components estimator for $y_t^s = \Lambda^s F_t^s + u_t^s, s = 1, ..., S$
3) $\hat{G}'\hat{F} = 0$

where $z_t = [z_t^{1\prime}, ..., z_t^{S\prime}]'$ *with* $z_t^s = x_t^s - \hat{\Lambda}^s \hat{F}_t^s$*, and* $y_t^s = x_t^s - \hat{\Gamma}^s \hat{G}_t$*. This motivates the following alternative computational algorithm, which is fast and robust to choice of starting values.*

1) Choose an initial estimates for global factors $\hat{G}$ *and corresponding factor loadings* $\hat{\Gamma}$*.*

2) Perform principal component analysis according to $y_t^s = \Lambda^s F_t^s + e_t^s$ *to obtain* $\hat{\Lambda}^s$ *and* $\hat{F}_t^s$ *for all* s*, where* $y_t^s = x_t^s - \hat{\Gamma}^s \hat{G}_t$*.*

3) Perform principal component analysis according to $z_t = \Gamma G_t + u_t$ *to obtain new* $\hat{\Gamma}$ *and* $\hat{G}$*, where* $z_t = [z_t^{1\prime}, ..., z_t^{S\prime}]'$ *and* $z_t^s = x_t^s - \hat{\Lambda}^s \hat{F}_t^s$*.*

4) Iterate between 2) and 3) until some convergence criteria for global factors is met.

The above algorithm only imposes within and between sector identification restrictions, and does not utilize the assumption that $G'F = 0$*. However, the common factors are well identified up to an* $r \times r$ *matrix transformation, then we may obtain estimates for sector-specific factors as* $\tilde{F}^s = P_{\hat{G}} \hat{F}^s$*, where* $P_{\hat{G}} = I_T - \hat{G}\hat{G}'/T$*. Also notice that the iterative PCA has the desired property that each iteration will decrease the objective function, i.e., the total sum of squared residuals. To guarantee that the alternative algorithm converges to the fixed point solution characterized by theorem 1, we need to add the projection step* $\tilde{F}^s = P_{\hat{G}} \hat{F}^s$ *between step3 and step 4.*

Theorem 1 can be readily extended to estimate a factor model with more than two levels of factors. For example, suppose one has a three-level factor model defined as follows

$$\begin{aligned} x_{it}^{sk} &= \gamma_i^{sk\prime} G_t + \lambda_i^{sk\prime} F_t^s + \mu_i^{sk\prime} R_t^k + e_{it}^{sk}, \\ i &= 1, ..., N_{sk}, \;\; s = 1, ..., S, \;\; k = 1, ..., K, \end{aligned}$$

where s is the index for level-2 factors and k is the index for level-3 factors. Using similar notations as Theorem 1, the following Corollary characterizes the least squares estimators for the above three-level factor model. Define the $T \times N_{sk}$ matrix $X^{sk} = \left(x_{it}^{sk}\right)'$.

Corollary 1: *Assume 1) within sector identification, 2) between sector identification, 3) orthogonality between common factors* G_t*, level-2 factors* F_t^s *and level-3 factors* R_t^k*. Let* $F = [F^1, F^2, ..., F^S]$*,* $R = [R^1, R^2, ..., R^K]$*,* $FR = [F, R]$*,* $FG = [F, G]$*,* $RG = [R, G]$ and $X^{sk} = [x_1^{sk}, ..., x_T^{sk}]'$*. Define* $A^{sk} = X^{sk} X^{sk\prime}$*,* $A^s = A^{s1} + ... + A^{sK}$, and $A = A^1 + ... + A^S$*. Assume* FR, FG and RG all have full column rank, *then the least squares solution for factors* *is determined by the following eigenvector problem,*

1) $\frac{1}{\sqrt{T}}\hat{G} \;= r$ eigenvectors for $P_{\widehat{FR}}A$ corresponding to its largest r eigenvalues,
2) $\frac{1}{\sqrt{T}}\hat{F}^s = r_s$ eigenvectors for $P_{\widehat{RG}}A^s$ corresponding to its largest r_s eigenvalues,
3) $\frac{1}{\sqrt{T}}\hat{R}^k = r_k$ eigenvectors for $P_{\widehat{FG}}A^k$ corresponding to its largest r_k eigenvalues,
4) $\widehat{FR} = [\hat{F}, \hat{R}],\; \widehat{RG} = [\hat{R}, \hat{G}],\; \widehat{FG} = [\hat{F}, \hat{G}],$
5) $P_Y = I_T - Y(Y'Y)^{-1}Y'$ is the projection matrix for a matrix Y

The estimators for factor loadings are given by $[\hat{\Gamma}^{sk}, \hat{\Lambda}^{sk}, \hat{\Pi}^{sk}] = X^{sk\prime}[\hat{G}, \hat{F}^s, \hat{R}^k]/T.$

5 Inference

Additional assumptions are needed to derive the large sample theory.

Assumption G (Moments and Central Limit Theorem): There exists a positive constant $M < \infty$ such that for all N, S and T :

1. for each t and s

$$\begin{aligned} E\left\|\frac{1}{\sqrt{N_s T}}\sum_{i=1}^{N_s}\sum_{k=1}^{T} F_k^s\left[e_{ik}^s e_{it}^s - E(e_{ik}^s e_{it}^s)\right]\right\|^2 &\leq M \\ E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{k=1}^{T} G_k\left[e_{ik}^s e_{it}^s - E(e_{ik}^s e_{it}^s)\right]\right\|^2 &\leq M \end{aligned}$$

2. for F and G

$$\begin{aligned} E\left\|\frac{1}{\sqrt{N_s T}}\sum_{i=1}^{N_s}\sum_{t=1}^{T} F_t^s \lambda_i^{s\prime} e_{it}^s\right\|^2 &\leq M \\ E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{k=1}^{T} G_t \gamma_i^{s\prime} e_{it}^s\right\|^2 &\leq M \end{aligned}$$

3. for each t and s, as $N_s \to \infty$ and $N \to \infty$

$$\begin{aligned} \frac{1}{\sqrt{N_s}}\sum_{i=1}^{N_s}\lambda_i^s e_{it}^s &\xrightarrow{d} N(0, \Sigma_t^s) \\ \frac{1}{\sqrt{N}}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\gamma_i^s e_{it}^s &\xrightarrow{d} N(0, \Sigma_t) \end{aligned}$$

where $\Sigma_t^s = \lim_{N_s \to \infty} (1/N_s) \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \lambda_i^s \lambda_j^{s\prime} E(e_{it}^s e_{jt}^s)$ and

$\Sigma_t = \lim_{N \to \infty} (1/N) \sum_{s=1}^{S} \sum_{i=1}^{N_s} \sum_{j=1i}^{N_s} \gamma_i^s \gamma_j^{s\prime} E(e_{it}^s e_{jt}^s)$;

4. for each i, as $T \to \infty$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t^s e_{it}^s \xrightarrow{d} N(0, \Phi_i^s)$$
$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t e_{it}^s \xrightarrow{d} N(0, \Psi_i^s)$$

where $\Phi_i^s = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1j}^{Ts\prime} E(F_t^s F_k^s e_{it}^s e_{jt}^s)$ and $\Psi_i^s = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1j}^{Ts\prime} E(G_t G_t' e_{it}^s e_{jt}^s)$.

Assumption H: The eigenvalues of the $r_s \times r_s$ matrix $(\Sigma_{\Lambda^s} \Sigma_{F^s})$ are distinct. The eigenvalues of the $r \times r$ matrix $(\Sigma_\Gamma \Sigma_G)$ are distinct.

Given the model $x_t^s = \Gamma^s G_t + \Lambda^s F_t^s + E_t^s$, the least squares estimator is not equivalent to the least squares estimator from a transformed representation using projection matrix $P^s \equiv I_M - \Lambda^s (\Lambda^{s\prime} \Lambda^s)^{-1} \Lambda^{s\prime}$

$$P^s x_t^s = P^s \Gamma^s G_t + U_t^s$$

because the latter loses information on F_t^s and the information on the orthogonality between F_t and G_t. Asymptotic theory is based on the fixed point solution for $\{G_t\}$ and $\{F_t\}$ jointly determined by first order conditions.

Remark: *If S is fixed, then Bai's 2003 results can be applied with some modification, taking into account the zero restrictions. The convergence rates for G_t and F_t are at the same level.*

Remark: *If S is large, we may perform sector by sector analysis. Convergence rates for G_t and F_t will be different. We will prove that to make inference on F_t, we can treat G_t as known. The convergence rate is $\sqrt{N}$ for G_t while $\sqrt{N_s}$ for F_t.*

Remark: *Requiring orthogonality between F and G pins down rotation matrix as block diagonal, in order to let the rotated model have same factor loading structure. To separately identify G_t and F_t, two sectors in the sample are enough for the purpose.*

The development of asymptotic theory for the fixed point type estimator is a non-standard one, because the estimation error for G depends on the estimation error for $\{F^1, ..., F^S\}$, and vice versa. Assuming $S \to \infty$, the approach taken by this paper follows three steps.

1) We provide an $\sqrt{N}-$consistent initial estimator for G_t and $\sqrt{T}-$consistent estimator for corresponding factor loadings γ_i^s.

2) We then prove that the estimated F_t^s is $\sqrt{N_s}-$consistent, given $\sqrt{N}-$consistent initial estimator of G_t. The limiting distribution for estimated F_t^s is normal, and invariant to the choice of initial estimator for G_t as long as it is $\sqrt{N}-$consistent.

3) Given the asymptotic expansion of the previous step estimated F_t^s, we are able to derive the asymptotic expansion of estimated G_t as a function of previous step estimated G_t. The fixed point representation for this asymptotic expansion will provide the asymptotic distribution of fixed point estimator for G_t, which is $\sqrt{N}-$normal.

5.1 Inference when $S = 2$

To fix idea, we assume $S = 2$. Let $H = [G, F^1, F^2]$ be $T \times (r + r_1 + r_2)$ and $H^1 = [G, F^1]$, $H^2 = [G, F^2]$. The overall model can be seen as a static factor model with $r + r_1 + r_2$ factors, with $N = N_1 + N_2$ observations at any time t,

$$Y_t = \Lambda H_t + e_t, \text{ where } \Lambda = \begin{bmatrix} \Gamma^1 & \Lambda^1 & 0 \\ \Gamma^2 & 0 & \Lambda^2 \end{bmatrix}$$

Notice that by assumption

$$\Lambda'\Lambda/N = \begin{bmatrix} \Gamma^{1\prime}\Gamma^1 + \Gamma^{2\prime}\Gamma^2 & \Gamma^{1\prime}\Lambda^1 & \Gamma^{2\prime}\Lambda^2 \\ \Lambda^{1\prime}\Gamma^1 & \Lambda^{1\prime}\Lambda^1 & 0 \\ \Lambda^{2\prime}\Gamma^2 & 0 & \Lambda^{2\prime}\Lambda^2 \end{bmatrix} /N \to \Sigma_\Lambda$$

It is straightforward to check that assumptions $A-G$ in Bai (2003) are satisfied, which means the inferential theory developed in Bai (2003) regarding principal components estimator $\hat{\Lambda}$ and $\hat{H}_t$ hold. However, the inferential theory is developed with only $(r+r_1+r_2)^2$ restrictions, while the zero restrictions on the grand factor loading matrix Λ are not considered. This implies that, even with $S = 2$, the multi-level factor model structure impose extra restrictions over the grand static factor model, because $N_2 r_1 + N_1 r_2 >> (r + r_1 + r_2)^2$ for N_1 and N_2 large. Also the principal components estimator $\hat{H}_t$ does not directly translate into $\hat{G}_t$, $\hat{F}_t^1$ and $\hat{F}_t^2$. Thus we need to estimate a rotation matrix for $\hat{H}_t$ such that $\hat{G}_t$, $\hat{F}_t^1$ and $\hat{F}_t^2$ are separated obtained.

5.1.1 A Two-Step Estimator

Before turning to our iterative type estimator, we first present a two-step estimator whose properties are easily studied. For simplicity, we assume $r = r_s$, $N_1 = N_2 = M$. Let a variable with superscript "*" denotes foreign variable and assume

$$\begin{aligned} X_t &= \Gamma G_t + \Lambda F_t + E_t \\ X_t^* &= \Gamma^* G_t + \Lambda^* F_t^* + E_t^* \end{aligned}$$

To separately identify G_t and F_t, F_t^*, we assume $G_t \perp (F_t, F_t^*)$.

In the first step, we conduct sector-by-sector principal component analysis to extract sectorial factors $\hat{H}_t, \hat{H}_t^*$, which have the following property

$$\begin{aligned} \hat{H}_t &= \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} G_t \\ F_t \end{pmatrix} + u_t, \\ \hat{H}_t^* &= \begin{pmatrix} R_{11}^* & R_{12}^* \\ R_{21}^* & R_{22}^* \end{pmatrix} \begin{pmatrix} G_t \\ F_t^* \end{pmatrix} + u_t^* \end{aligned}$$

where $u_t, u_t^* = O_p(\frac{1}{\sqrt{M}})$ if assuming $\frac{M}{T^2} \to 0$ according to Theorem 1 in Bai (2003), R_{ij} and R_{ij}^* are $r \times r$ matrices for all i, j. Moreover, $\sqrt{M} u_t$ and $\sqrt{M} u_t^*$ have normal limiting distributions and are asymptotically independent.

In the second step, we provide consistent estimator for the rotation matrix based on our identifying assumptions. Then the two step estimator for G_t, F_t and F_t^* are obtained by multiplying $\hat{H}_t$ and $\hat{H}_t^*$ using the estimated rotation matrix. The resulting estimators have the same convergence rate as $\hat{H}_t$ and $\hat{H}_t^*$.

Proposition 2: Assuming 1) $R_{11} = R_{12} = R_{12}^* = I_r$, 2) $\sum_{t=1}^T G_t F_t' = \sum_{t=1}^T G_t F_t^{*\prime} = 0$, then all the remaining parameters $\{R_{21}, R_{22}, R_{11}^*, R_{21}^*, R_{22}^*, \Sigma_G, \Sigma_F, \Sigma_{F^*}\}$ are uniquely identified.

Remark: We are interested in estimating a rotation of G_t, F_t and F_t^* respectively, which leaves $r^2 + r_1^2 + r_2^2 = 3r^2$ degrees of freedom. Rewrite the above equations as

$$\begin{aligned} \hat{H}_t &= \begin{pmatrix} I_r & I_r \\ R_{21} R_{11}^{-1} & R_{22} R_{12}^{-1} \end{pmatrix} \begin{pmatrix} R_{11} G_t \\ R_{12} F_t \end{pmatrix} + u_t, \\ \hat{H}_t^* &= \begin{pmatrix} R_{11}^* R_{11}^{-1} & I_r \\ R_{21}^* R_{11}^{-1} & R_{22}^* (R_{12}^*)^{-1} \end{pmatrix} \begin{pmatrix} R_{11} G_t \\ R_{12}^* F_t^* \end{pmatrix} + u_t^* \end{aligned}$$

Because $R_{11}G_t$ can be treated the same as true global factors G_t as our objective of interest, to save notation, we replace $R_{11}G_t$ by G_t and similarly for $R_{12}F_t$ and $R_{12}^*F_t^*$. Redefine the coefficient matrix to obtain the following equivalent representation

$$\hat{H}_t = \begin{pmatrix} I_r & I_r \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} G_t \\ F_t \end{pmatrix} + u_t,$$
$$\hat{H}_t^* = \begin{pmatrix} A_{11}^* & I_r \\ A_{21}^* & A_{22}^* \end{pmatrix} \begin{pmatrix} G_t \\ F_t^* \end{pmatrix} + u_t^*$$

or more specifically

$$\begin{aligned} \hat{H}_{1t} &= G_t + F_t + u_{1t}, \\ \hat{H}_{2t} &= A_{21}G_t + A_{22}F_t + u_{2t}, \\ \hat{H}_{1t}^* &= A_{11}^*G_t + F_t^* + u_{1t}^* \\ \hat{H}_{2t}^* &= A_{21}^*G_t + A_{22}^*F_t^* + u_{2t}^* \end{aligned}$$

or in matrix notation

$$\tilde{H}_t \equiv \begin{pmatrix} \hat{H}_t \\ \hat{H}_t^* \end{pmatrix} = \begin{pmatrix} I_r & I_r & 0 \\ A_{21} & A_{22} & 0 \\ A_{11}^* & 0 & I_r \\ A_{21}^* & 0 & A_{22}^* \end{pmatrix} \begin{pmatrix} G_t \\ F_t \\ F_t^* \end{pmatrix} + O_p(\frac{1}{\sqrt{M}})$$

The above proposition states that, coupled with the assumption that $G'F = 0, G'F^* = 0$, the above model is identified.

Proof for Proposition 2: see appendix.

The above analysis implies that quasi maximum likelihood analysis of the above system yields unique solution. Assume that

$$(G, F, F^*)' \cdot (G, F, F^*)/T \rightarrow_p \begin{pmatrix} V_G & 0 & 0 \\ 0 & V_F & V' \\ 0 & V & V_{F^*} \end{pmatrix}$$

where the RHS is the covariance matrix for factors.

In the second step, we can perform confirmatory factor analysis on the above restricted linear system. Using quasi maximum likelihood estimation or minimum distance estimator (least squares estimator)[5], we will obtain at least $\sqrt{T}-$consistent coefficient estimator A and A^*.[6]

The 2-step estimator for common factors and sector-specific factors are given by

$$\begin{pmatrix} \hat{G}_t \\ \hat{F}_t \\ \hat{F}_t^* \end{pmatrix} = (\hat{B}'\hat{B})^{-1}\hat{B}' \begin{pmatrix} \hat{H}_t \\ \hat{H}_t^* \end{pmatrix}$$

where $B = \begin{pmatrix} I_r & I_r & 0 \\ A_{21} & A_{22} & 0 \\ A_{11}^* & 0 & I_r \\ A_{21}^* & 0 & A_{22}^* \end{pmatrix}$.

Proposition 3. The two-step estimators for G_t, F_t and F_t^* are $\sqrt{M}-$consistent.
Proof: see appendix.

5.2 Inference when $S \to \infty$

We first provide a candidate estimator for common factors, which is $\sqrt{N}-$consistent.

To fix idea, we assume throughout the rest of the paper $N_1 = ... = N_S = M$, and $r = r_1 = ... = r_S$, and thus $N = M \cdot S$. Notice that using pairwise sectorial data, we can separately identify common factors and sector-specific factors, and obtain $\sqrt{M}-$ consistency of G_t. We have

$$g_t^s \equiv \sqrt{M}(\hat{G}_t^s - H^s G_t) = O_p(1), s = 1, 2, , , S/2 \tag{9}$$

where any two pairs contain different four sectors. Assuming i.i.d. e_t^s across $s = 1, 2, ...S$, then $\{g_t^s\}$ are asymptotically independent across s, and we can apply central limit theorem on g_t^s such that

$$\frac{g_t^1 + ... + g_t^{S/2}}{\sqrt{S/2}} = O_p(1)$$

[5]The sample covariance matrix provides $2r(4r+1)$ restrictions, while we have $6r^2+3r(r+1)/2$ parameters regarding $\{A_{21}, A_{22}, A_{11}^*, A_{21}^*, A_{22}^*, \Sigma_G, \Sigma_F, \Sigma_{F^*}\}$, which is strictly less than the first number for all $r \geq 1$.

[6]The convergence rate might even be $\sqrt{MT}$ given that $u_t, u_t^* = O_p(\frac{1}{\sqrt{M}})$, and we might indeed obtain $(\sqrt{M}\hat{A} - \sqrt{M}A) = O_p(\frac{1}{\sqrt{T}})$. The latter argument needs further justification although not essential for deriving the large sample properties.

If we define $\hat{G}_t = \frac{\hat{G}_t^1 + ... + \hat{G}_t^{S/2}}{S/2}$, then

$$\begin{aligned}
\frac{S/2 \cdot \sqrt{M}(\hat{G}_t - HG_t)}{\sqrt{S/2}} &= O_p(1), \text{ or} \\
\hat{G}_t - HG_t &= O_p(\frac{1}{\sqrt{MS}})
\end{aligned}$$

where $H = \frac{2}{S}\sum_{s=1}^{S/2} H^s$.

The above procedure provides a $\sqrt{N}-$consistent estimator $\hat{G}_t$ for the common factor. This would have important implications when we derive limiting distribution of estimators for sector-specific factors, which is proved to have a slower convergence rate $\sqrt{M}$.

5.3 Asymptotic Expansion for Iterative Principal Components Estimator

Given a static factor model $x_t = \Lambda F_t + e_t$, $t = 1, ..., T$, the principal components estimator yields the following identity as in Bai and Ng (2002), and equation (A.1) in Bai (2003)

$$\begin{aligned}
\hat{F}_t - H'F_t &= V_{NT}^{-1}\{\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\delta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st}\} \quad (10) \\
\delta_{st} &= \frac{e_s'e_t}{N} - \gamma_N(s,t) \\
\gamma_N(s,t) &= E(\frac{e_s'e_t}{N}) \\
\eta_{st} &= F_s'\Lambda'e_t/N \\
\xi_{st} &= F_t'\Lambda'e_s/N = e_s'\Lambda F_t/N
\end{aligned}$$

where by definition $\frac{1}{NT}XX'\hat{F} = \hat{F}V_{NT}$ or $\frac{1}{NT}XX'\hat{F}V_{NT}^{-1} = \hat{F}$, and $H = (\Lambda'\Lambda/N)(F'\hat{F}/T)V_{NT}^{-1}$. V_{NT} is a diagonal matrix consisting of the first r eigenvalues of $\frac{XX'}{NT}$ in decreasing order.

5.3.1 Sector-Specific Factors

Let $\hat{G}_t$ be the candidate estimator with the property that $\sqrt{N}(\hat{G}_t - H'G_t) = O_p(1)$, and let $\hat{\Gamma}^s$ be the estimator from pairwise principal components estimation, which is a rotation of principal component estimator, we can rewrite the model $x_t^s = \Gamma^s G_t + \Lambda^s F_t^s + e_t^s$ as

$$y_t^s = \Lambda^s F_t^s + u_t^s$$

where $y_t^s = x_t^s - \hat{\Gamma}^s \hat{G}_t$ and $u_t^s = \Gamma^s G_t - \hat{\Gamma}^s \hat{G}_t + e_t^s$. Assume $S \to \infty$, then $\sqrt{M}(\hat{G}_t - H'G_t) = o_p(1)$. Using the identity for principal components estimator $\hat{F}_t^s$, we may prove the following

$$\begin{aligned}
\hat{F}_t^s - H^{s\prime} F_t^s &= A^s \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s u_{it}^s + o_p(1) \\
A^s &= plim V_{MT}^{-1} \frac{1}{T} \sum_{k=1}^{T} (\hat{F}_k^s F_k^{s\prime}) \\
\frac{1}{M} \sum_{i=1}^{M} \lambda_i^s u_{it}^s &= \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s (\Gamma_i^{s\prime} G_t - \hat{\Gamma}_i^{s\prime} \hat{G}_t + e_{it}^s) \\
&= \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s e_{it}^s + \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s (H^{-1} \Gamma_i^s - \hat{\Gamma}_i^s)' \hat{G}_t \\
&\quad + \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s \Gamma_i^{s\prime} H'^{-1} (H' G_t - \hat{G}_t)
\end{aligned}$$

where the last equality comes from he fact that $\Gamma_i^{s\prime} G_t - \hat{\Gamma}_i^{s\prime} \hat{G}_t = (H^{-1}\Gamma_i^s)' H' G_t - (H^{-1}\Gamma_i^s)' \hat{G}_t + (H^{-1}\Gamma_i^s)' \hat{G}_t - \hat{\Gamma}_i^{s\prime} \hat{G}_t = (H^{-1}\Gamma_i^s)'(H'G_t - \hat{G}_t) + (H^{-1}\Gamma_i^s - \hat{\Gamma}_i^s)' \hat{G}_t$.

The second term is $o_p(1)$. From Bai (2005), we have $\frac{1}{M} \sum_{i=1}^{M} \lambda_i^s (H^{-1}\Gamma_i^s - \hat{\Gamma}_i^s)' = O_p(\frac{1}{\min\{M,T\}})$, then $\frac{1}{\sqrt{M}} \sum_{i=1}^{M} \lambda_i^s (H^{-1}\Gamma_i^s - \hat{\Gamma}_i^s)' = O_p(\frac{\sqrt{M}}{\min\{M,T\}}) = o_p(1)$ if either $M \le T$ or $\frac{\sqrt{M}}{T} = o_p(1)$.

The third term is $o_p(1)$. By assumption $\frac{1}{M} \sum_{i=1}^{M} \lambda_i^s \Gamma_i^{s\prime} = O_p(1)$. And $\sqrt{M}(H'G_t - \hat{G}_t) = o_p(1)$, given $S \to \infty$.

In sum

$$\frac{1}{\sqrt{M}} \sum_{i=1}^{M} \lambda_i^s u_{it}^s = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \lambda_i^s e_{it}^s + o_p(1), \text{ given } \frac{\sqrt{M}}{T} \to 0.$$

Theorem 2: *Under assumptions $A - H$, assume $\frac{\sqrt{M}}{T} \to 0$ and $\sqrt{M}(\hat{G}_t - H'G_t) = o_p(1)$, then we have*

$$\begin{aligned}
\sqrt{M}(\hat{F}_t^s - H^{s\prime} F_t^s) &= (V_{MT}^s)^{-1} \left(\frac{\hat{F}^{s\prime} F^s}{T} \right) \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \lambda_i^s e_{it}^s + o_p(1) \qquad (11) \\
&= A^s \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \lambda_i^s e_{it}^s + o_p(1) \xrightarrow{d} N(0, A^s \Phi_t^s A^{s\prime})
\end{aligned}$$

where $A^s = plim V_{MT}^{-1} \frac{1}{T}\sum_{k=1}^{T}(\hat{F}_k^s F_k^{s\prime})$, V_{MT}^s *is a diagonal matrix consisting of the first* r *eigenvalues of* $\frac{Y^s Y^{s\prime}}{MT}$ *in decreasing order,* Φ_t^s *is defined in assumption* F3.

Remark: The candidate estimator for G_t is provided in the previous section using disjoint pairwise principal components estimation. A key result is that the convergence rate for global factor is $\sqrt{N}$, and thus when $S \to \infty$, $\sqrt{M}(H'G_t - \hat{G}_t) = o_p(1)$.

Proof of theorem 2: see appendix.

Define the asymptotic covariance matrix as
$\Sigma_t^s \equiv A^s \Sigma_t A^{s\prime} = \text{plim}(V_{MT}^s)^{-1} \frac{\hat{F}^{s\prime} F^s}{T}\left(\frac{1}{M}\sum_{i=1}^{M}(\sigma_{it}^s)^2 \lambda_i^s \lambda_i^{s\prime}\right)\frac{F^{s\prime}\hat{F}^s}{T}(V_{MT}^s)^{-1}$, and let $\hat{e}_{it}^s = x_{it}^s - \gamma_i^{s\prime}\hat{G}_t - \lambda_i^{s\prime}\hat{F}_t^s$, then a consistent estimator of the covariance matrix is given by

$$\hat{\Sigma}_t^s = (V_{MT}^s)^{-1}\frac{\hat{F}^{s\prime}\hat{F}^s}{T}\left(\frac{1}{M}\sum_{i=1}^{M}(\hat{e}_{it}^s)^2 \hat{\lambda}_i^s \hat{\lambda}_i^{s\prime}\right)\frac{\hat{F}^{s\prime}\hat{F}^s}{T}(V_{MT}^s)^{-1} \tag{12}$$

where V_{MT}^s *is a diagonal matrix consisting of the first* r *eigenvalues of* $\frac{Y^s Y^{s\prime}}{MT}$ *in decreasing order.*

5.3.2 Common Factors

Rewrite the data generating process as

$$\begin{aligned} z_t^s &= \Gamma^s G_t + v_t^s, \text{ where} \\ z_t^s &= x_t^s - \hat{\Lambda}^s \hat{F}_t^s \\ v_t^s &= \Lambda^s F_t^s - \hat{\Lambda}^s \hat{F}_t^s + e_t^s \end{aligned}$$

Recall the representation of the identity for principal components estimator

$$\begin{aligned} V_{NT}(\hat{G}_t - H'G_t) &= \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s v_s' v_t + \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s \eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s \xi_{st} \\ \eta_{st} &= G_s'\Gamma' v_t / N \\ \xi_{st} &= G_t'\Gamma' v_s / N = v_s'\Gamma G_t / N \end{aligned}$$

where V_{NT} *is a diagonal matrix consisting of the first* r *eigenvalues of* $\frac{ZZ'}{NT}$ *in decreasing order.* The following theorem can be proved based on the above identity.

Theorem 3: *Under assumptions $A-H$, assume $\frac{\sqrt{N}}{T} \to 0$, we have*

$$\sqrt{N}(\hat{G}_t - H'G_t) - \sqrt{N}\mu_t \xrightarrow{d} N(0, \Sigma_t^G) \tag{13}$$

where $\mu_t = O(\frac{1}{\sqrt{N}}\frac{S}{M})$ is the bias correction term. The bias correction term can be ignored if $\frac{S}{M} \to 0$.

To prove theorem 3, we need the following lemmas.

Lemma 1: $I1 = \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s v_s' v_t = O_p(\frac{1}{\sqrt{N}}\sqrt{\frac{S}{M}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N}) = O_p(\frac{1}{\sqrt{N}}\sqrt{\frac{S}{M}}) + O_p(\frac{1}{\min\{N,T\}})$.

Lemma 2: $I2 = \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s\eta_{st} = O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{N}})O_p(\frac{\sqrt{N}}{\min(M,T)})$

Lemma 3: $I3 = \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s\xi_{st} = O_p(\frac{1}{\sqrt{N}}\frac{1}{\sqrt{T}})$.

Proof for Lemma 1–3 and Theorem 3: see appendix for details.

5.3.3 Factor Loadings

The factor loadings measure individual variables' heterogeneous response to both common factors and sector-specific factors. If assuming $\frac{\sqrt{T}}{N_s} \to 0$, the convergence rate of the estimators of factor loadings will be $\sqrt{T}$, which is the same as Bai (2003). The following corollary summarizes the asymptotic distribution of factor loadings. Define the $(r+r_s)\times 1$ vector $\mu_i^s = [\gamma_i^{s\prime}, \lambda_i^{s\prime}]'$. Define the $(r+r_s)\times(r+r_s)$ rotation matrix $\bar{H} = \begin{bmatrix} H & 0 \\ 0 & H^s \end{bmatrix}$, where H and H^s are defined in Theorem 2 and Theorem 3.

Corollary 2: *Under assumptions $A-H$, assume $\frac{\sqrt{T}}{N_s} \to 0$, then we have*

$$\sqrt{T}(\hat{\mu}_i^s - \bar{H}^{-1}\mu_i^s) \xrightarrow{d} N(0, \Sigma_i^s)$$

where Σ_i^s is defined in Theorem 2 of Bai (2003).

The above corollary is a direct application of Theorem 2 in Bai (2003).

5.3.4 Estimating the Covariance Matrices

According to theorem 2, $\sqrt{M}(\hat{F}_t^s - H^{s\prime}F_t^s) \to^d N(0, \Sigma_t^s)$. The following estimator can be used in our Monte Carlo experiments.

$$\begin{aligned}
\hat{\Sigma}_t^s &= (V_{MT}^s)^{-1}\frac{\hat{F}^{s\prime}\hat{F}^s}{T}\left(\frac{1}{M}\sum_{i=1}^{M}(\hat{e}_{it}^s)^2\hat{\lambda}_i^s\hat{\lambda}_i^{s\prime}\right)\frac{\hat{F}^{s\prime}\hat{F}^s}{T}(V_{MT}^s)^{-1} \qquad (14)\\
&= (V_{MT}^s)^{-1}\frac{\hat{F}^{s\prime}\hat{F}^s}{T}\frac{1}{M}\hat{\Lambda}^{s\prime}diag\{(\hat{e}_{1t}^s)^2, (\hat{e}_{2t}^s)^2, ..., (\hat{e}_{Mt}^s)^2\}\hat{\Lambda}^s\frac{\hat{F}^{s\prime}\hat{F}^s}{T}(V_{MT}^s)^{-1}\\
&= \frac{1}{M}(V_{MT}^s)^{-1}\hat{\Lambda}^{s\prime}diag\{(\hat{e}_{1t}^s)^2, (\hat{e}_{2t}^s)^2, ..., (\hat{e}_{Mt}^s)^2\}\hat{\Lambda}^s(V_{MT}^s)^{-1}
\end{aligned}$$

where V_{MT}^s is a diagonal matrix with largest r eigenvalues of $\frac{Y^sY^{s\prime}}{MT}$ on the diagonal, $Y^s = X^s - \hat{G}\hat{\Gamma}^{s\prime}$, and the normalization $\frac{\hat{F}^{s\prime}\hat{F}^s}{T} = I_{r_s}$ is applied. The estimator for error term is $\hat{e}_{it}^s = x_{it}^s - \hat{\gamma}_i^{s\prime}\hat{G}_t - \hat{\lambda}_i^{s\prime}\hat{F}_t^s$.

Theorem 3 proves $\sqrt{N}(\hat{G}_t - H'G_t) \to^d N(0, \Sigma_t)$. We first define

$$\begin{aligned}
H^s &= (\Lambda^{s\prime}\Lambda^s/M)(F^{s\prime}\hat{F}^s/T)(V_{MT}^s)^{-1},\ H = (\Gamma'\Gamma/N)(G'\hat{G}/T)V_{NT}^{-1},\ \bar{H} = blockdiag\{H^1, ..., H^S\}\\
A^s &= plim(V_{MT}^s)^{-1}\frac{\hat{F}^sF^s}{T},\ A = blockdiag\{A^1, ..., A^S\}
\end{aligned}$$

where the operator is defined as $blockdiag(C^1, C^2) = \begin{pmatrix} C^1 & 0 \\ 0 & C^2 \end{pmatrix}$, for arbitrary matrices C^1, C^2 not necessarily having the same dimension. Then let

$$\begin{aligned}
\Omega &= \left(\frac{1}{T}\sum_{k=1}^{T}\hat{G}_kG_k'\right)\left\{\frac{1}{S}\sum_{s=1}^{S}\frac{\Gamma^{s\prime}\Lambda^s}{M}(H^{s\prime})^{-1}A^s\frac{\Lambda^{s\prime}\Gamma^s}{M}H'^{-1}\right\}\\
&= \frac{1}{MNT}\hat{G}'\left(G\Gamma'\right)\left(\Lambda(\bar{H}')^{-1}\right)\left(A\Lambda'\right)\left(\Gamma H'^{-1}\right)\\
Q &= (V_{NT} - \Omega)^{-1}\left(\frac{\hat{G}'G}{T}\right)\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\Gamma^{s\prime}\left(I_M - \frac{1}{M}\Lambda^s(H^{s\prime})^{-1}A^s\Lambda^{s\prime}\right)\\
&= (V_{NT} - \Omega)^{-1}\left(\frac{\hat{G}'G}{T}\right)\frac{1}{\sqrt{N}}\Gamma'\left(I_N - \frac{1}{M}\Lambda(\bar{H}')^{-1}A\Lambda'\right)
\end{aligned}$$

Because H' is the rotation matrix for G, and thus H'^{-1} is the rotation matrix for Γ, which implies we can use $\hat{\Gamma}$ to estimate $\Gamma H'^{-1}$. Likewise, we can use $\hat{\Lambda}$ to estimate $\Lambda(\bar{H}')^{-1}$. Because factors and factor loadings appear in $A\Lambda'$ as products, we can use $(V_{MT}^s)^{-1}\frac{1}{T}\hat{F}^{s\prime}\hat{F}^s\hat{\Lambda}^{s\prime} =$

$(V_{MT}^s)^{-1}\hat{\Lambda}^{s\prime}$ to estimate $A^s\Lambda^{s\prime}$, which is the s^{th} block diagonal term of $A\Lambda'$, $s = 1, ..., S$. Similarly, we can use $\hat{G}\hat{\Gamma}'$ to estimate true $G\Gamma'$. In sum, let $\hat{A}^s = (V_{MT}^s)^{-1}\frac{1}{T}\hat{F}^s\hat{F}^s$, then the estimator for Ω is given by

$$\begin{aligned}\hat{\Omega} &= \frac{1}{MNT}\hat{G}'\left(\hat{G}\hat{\Gamma}'\right)\hat{\Lambda}\left(\hat{A}\hat{\Lambda}'\right)\hat{\Gamma} = \frac{1}{MN}\hat{\Gamma}'\hat{\Lambda}\hat{A}\hat{\Lambda}'\hat{\Gamma} \\ &= \frac{1}{MN}\hat{\Gamma}' \cdot blockdiag\{\hat{\Lambda}^s(V_{MT}^s)^{-1}\hat{\Lambda}^{s\prime}, s = 1, ..., S\} \cdot \hat{\Gamma}\end{aligned}$$

Using the same argument, we may estimate Q using

$$\begin{aligned}\hat{Q} &= \frac{1}{\sqrt{N}}(V_{NT} - \hat{\Omega})^{-1}\left(\frac{\hat{G}'\hat{G}}{T}\right)\hat{\Gamma}'\left(I_N - \frac{1}{M}\hat{\Lambda}\hat{A}\hat{\Lambda}'\right) \\ &= \frac{1}{\sqrt{N}}(V_{NT} - \hat{\Omega})^{-1}\hat{\Gamma}'\left(I_N - \frac{1}{M}\hat{\Lambda}\hat{A}\hat{\Lambda}'\right) \\ &= \frac{1}{\sqrt{N}}(V_{NT} - \hat{\Omega})^{-1}\hat{\Gamma}'\left(I_N - \frac{1}{M}blockdiag\{\hat{\Lambda}^s(V_{MT}^s)^{-1}\hat{\Lambda}^{s\prime}, s = 1, ..., S\}\right)\end{aligned}$$

where the normalization $\left(\frac{\hat{G}'\hat{G}}{T}\right) = I_r$ is applied.

Then we can use the following estimator for the covariance matrix

$$\hat{\Sigma}_t = \hat{Q}diag\{(\hat{e}_{1t}^1)^2, (\hat{e}_{2t}^1)^2, ..., (\hat{e}_{Mt}^S)^2\}\hat{Q}' \tag{15}$$

6 Monte Carlo Studies of the Least Squares Estimator

We evaluate the estimators by projecting them onto the true ones. The goodness of fit of common factors and their loadings, sector-specific factors and their loadings, as well as fit of common components are reported. Consider the iterative principal components (IPC hereafter) method with projection in the last step. Given common components regarding common factors, the sum of squared residuals is minimized by principal components estimators for sector-specific factors and loadings sector by sector. Likewise, given common components regarding sector-specific factors, the objective function is minimized by principal components estimators for common factors and loadings. Thus each step of iteration will decrease the objective function. The solution is characterized by first order conditions regarding all the model parameters. The fixed point solution is the least squares solution. Simulations suggest that this algorithm is robust to the choice of initial values, given enough number of iterations.

6.1 Robustness of the IPC Algorithm and Consistency

In this simulation design, we assume the number of sectors $S = 2$, number of periods $T = 200$, number of variables within one sector $M = 200$, and number of factors $r = r_s = 2$. Let $\mathbb{N}(m, n)$ denotes an $m \times n$ matrix with elements being *i.i.d.* standard normal. Then we simulate model (1) as follows.

Common factors:	$G = 2 + 1.5 \cdot \mathbb{N}(T, r)^2$
Common factor loadings:	$\Gamma = 0.5 + \mathbb{N}(N, r)$
Sector-specific factors:	$F^s = 2 + 2 \cdot \mathbb{N}(T, r)^2, \text{for } s = 1, 2$
Sector-specific factor loadings:	$\Lambda^s = 0.5 + \mathbb{N}(M, r), \text{for } s = 1, 2$
Idiosyncratic error terms:	$E = 2 \cdot \mathbb{N}(T, N)$

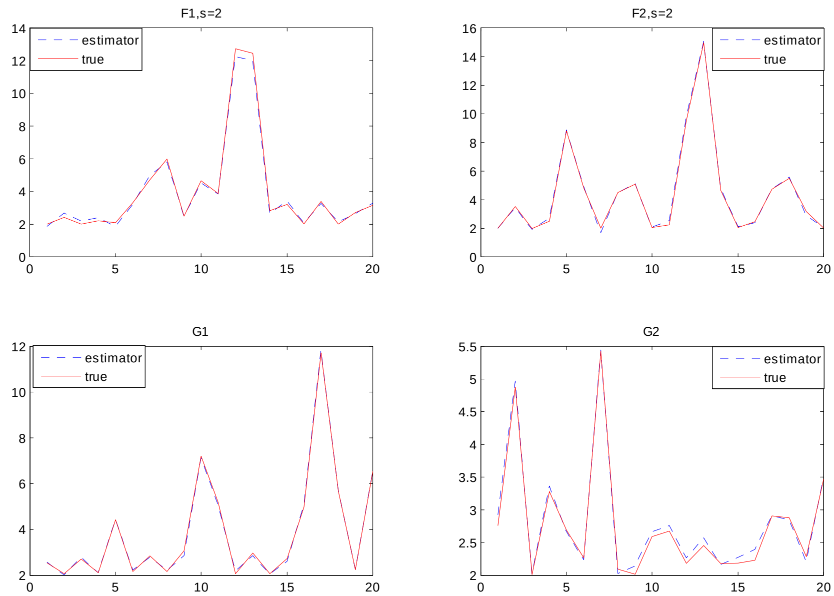
When evaluate the performance, we first project true factors on the estimated one to find the rotation matrix. Then we use the inverse of the same rotation matrix to rotate factor loadings.

Let G_proj be the rotated estimated common factors and let $G\_fit = trace(G\_proj' \cdot G\_proj)/trace(G\_true' \cdot G\_true)$, which is a measure of the fit of estimated factors. Similarly define F_fit, LG_fit, LF_fit, where LG and LF denote factor loadings for common factors and sector-specific factors respectively. Our default choice of initial values for G are chosen to be the first r principal components of data matrix X. The following table shows that principal components estimators for factors and factor loadings are not consistent. Instead, the iterative principal components estimators are consistent, where the common factors and sector specific factors are separately identified.

# of iterations =	1	5	80
G_fit	0.84598	0.86992	0.99939
LG_fit	287.4504	230.6089	0.99508
F_fit	0.76044	0.82007	0.9980
LF_fit	135.9995	821.1267	1.03

In the following figure, we plot the time series of projected estimators against the true ones.

We can see that the estimators accord with the true ones very well.



The dashed line are the estimators for factors projected onto the true factors. To make the graph clear, we only show the estimation results from $t = 1$ to 20. The title "F1,s=2" means the graph is for the first element of sector-specific factor of sector 2, F_{1t}^2. The title "G2" means the graph is for the second element of common factor, G_{2t}.

To check the robustness of the iterative principal components algorithm to the choice of initial values, we reestimate the above model, using the same data but with arbitrary random initial values to start the iteration. The results suggest that the fixed point solution, which is approximated by enough number of iterations, is consistent.

# of iterations =	200	1000
G_fit	0.30166	0.99948
LG_fit	2.2437	0.98362
F_fit	6.0591	0.99794
LF_fit	4.7297	1.04

For the same model, we modify sector-specific factors according to $F^s = 12 + 2 \cdot \mathbb{N}(T, r)^2$. The outcome still performs very well. For a typical experiment we obtain $G\_fit =$

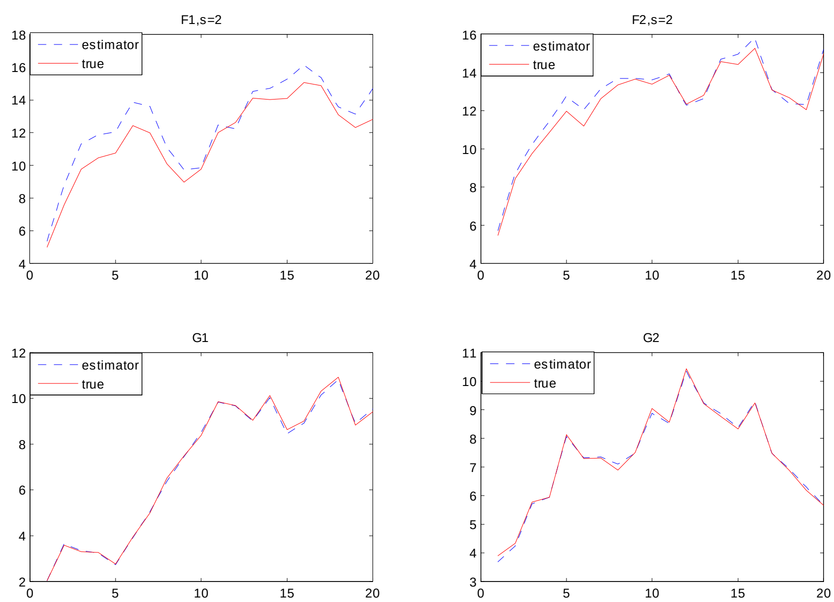
$0.99813, LG\_fit = 1.1779, F\_fit = 0.99969, LF\_fit = 0.99018$. This suggests that relative magnitude of common factors and sector-specific factors would not affect the estimation results.

Now we add dynamics in factors. Let $\rho_F = 0.7, \rho_G = 0.8$. Then we generate the model as follows.

Global factors: $G(t,:) = 2 + \rho_G \cdot G(t-1,:) + \mathbb{N}(1,r)$.

Country factors: $F^s(t,:) = 4 + \rho_F \cdot F^s(t-1,:) + \mathbb{N}(1,r)$.

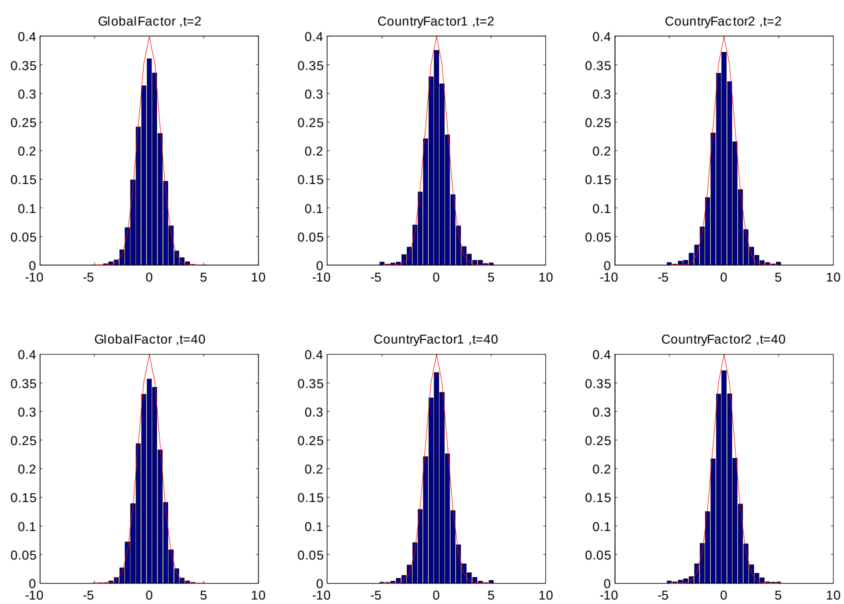
After iterating over the first order conditions 100 times, we obtain $G\_fit = 0.9999, LG\_fit = 1.0924, F\_fit = 0.99882, LF\_fit = 0.96947$. Plot of projected estimators accord with the true ones with great precision.



Another observation is that common factors are more precisely estimated than sector-specific factors. This is consistent with our theory, which states that the convergence rate of common factors is generally faster than that of sector-specific factors.

6.2 Finite Sample Performance of the Asymptotic Theory for Factors

In this section, we choose the sample size to be $T = 30$, $M = 40$, $S = 5$ and $r = r_s = 1$. We generate data according to $x_{it}^s = \gamma_i^{s\prime} G_t + \lambda_i^{s\prime} F_t^s + e_{it}^s$, where $\gamma_i^s, G_t, \lambda_i^s, F_t^s$ and e_{it}^s are i.i.d. $N(0,1)$ for all i, t and s. We then run various Monte Carlo experiments to evaluate finite sample performance of our asymptotic theory. For the given model, we make 2000 independent simulations. For each of the 2000 samples, we estimate factors F^s $(s = 1, ..., 10)$ and G, and their asymptotic covariance matrices according to theorem 2 and 3. Then the estimated asymptotic covariance matrices are used to normalize the difference of estimated factor and rotation of true factors. If our asymptotic theory provides a nice approximation in such a finite sample, then the standard normal density should resemble the resulting histogram for each element of $\sqrt{N}\left(\hat{\Sigma}_t\right)^{-1/2}(\hat{G}_t - H' G_t)$ and $\sqrt{M}\left(\hat{\Sigma}_t^s\right)^{-1/2}(\hat{F}_t^s - H^{s\prime} F_t^s)$, $s = 1, ..., S$, where $\hat{\Sigma}_t$ and $\hat{\Sigma}_t^s$ are estimated asymptotic covariance matrices for factors. The following figure justifies that our asymptotic theory performs nicely in such a finite sample.



As we can see, each histogram well accords with the standard normal density function. This suggests that our theory provides a nice approximation in the finite sample.

7 Comovement in Real and Financial Sectors

In this section, I carry an empirical study about different patterns of comovement within real and financial sectors, using a 2-level factor model. Suppose we have a large vector of data for each sector, x_t^R and x_t^F, where the superscript R and F denote real sector and financial sector respectively. Using a 2-level factor model, we are able to decompose the shock to any economic variable into three components, namely, the common component, sector-specific component, and the idiosyncratic component. The model for x_t^R and x_t^F is given by

$$x_t^s = \Gamma^s G_t + \Lambda^s F_t^s + e_t^s, \ \ s = R, F, t = 1, ..., T$$

where the $N_s \times 1$ observed vector x_t^s is affected by factor common to both sectors G_t, factor common within the s sector F_t^s and idiosyncratic shock e_t^s. Under the orthogonality conditions between different components, we may decompose the sample covariance of x_t^s into three parts

$$\begin{aligned} \frac{1}{T} X^s X^{s\prime} &\approx \hat{\Gamma}^s \frac{\hat{G}'\hat{G}}{T} \hat{\Gamma}^{s\prime} + \hat{\Lambda}^s \frac{\hat{F}^{s\prime}\hat{F}^s}{T} \hat{\Lambda}^{s\prime} + \frac{1}{T}\hat{E}^s \hat{E}^{s\prime}, \ \ s = R, F \\ &= \hat{\Gamma}^s \hat{\Gamma}^{s\prime} + \hat{\Lambda}^s \hat{\Lambda}^{s\prime} + \frac{1}{T}\hat{E}^s \hat{E}^{s\prime}, \ \text{by normalization of factors} \end{aligned}$$

where X^s is the $T \times N$ data matrix for sector s, $\hat{G}$ is the estimated common factor, $\hat{F}^s$ is the estimated sector specific common factor, and $\hat{E}^s$ is the estimated $T \times N$ idiosyncratic error term matrix defined as $\hat{E}^s = X^s - \hat{G}\hat{\Gamma}^{s\prime} - \hat{F}^s \hat{\Lambda}^{s\prime}$. Then we are able to analyze how different types of factors explain the variation observed in the data.

7.1 Data and Empirical Results

For real sector data, we use Boivin, Giannoni and Mihov (2007)'s monthly BBE dataset, covering 353 months from 1976 Feb to 2005 Jun. We only use those series transformed by log difference, which amounts to 234 series of real sector data, covering categories including industrial production, employment, personal consumption expenditure, etc. The following chart provides the summary statistics of the standard deviations of all 234 series.

Summary for standard deviations of real sector time series (in %)

mean	std	min	max	median	75% percentile
2.5955	5.6968	0.1656	75.60	1.4588	2.3595

We remove those series with standard deviation greater than 1.5%, which contains very noisy information about factors so far as monthly growth rates are considered. Finally, we have 120 series for the real sector.

For financial sector data, I adopt the 100 portfolios data sets constructed by Fama and French[7], using the same time span as the real sector data and removing four portfolios due to missing observations. I also add the Dow Jones Industrial Average, S&P Composite, and S&P Industrials. This amounts to 99 series for the financial sector. All the series are demeaned. Before doing factor analysis, I multiply the data sets of the real sector with a constant such that they share similar magnitudes as the data sets of the financial sector.

As a benchmark, we select the number of common factors to be $r = 3$, the number of sector specific factors to be $r_R = 9, r_F = 5$. The following chart reports the variance decomposition exercise.

	Real Sector (x_{it}^R)			Financial Sector (x_{jt}^F)		
	$\gamma_i^{R\prime} G_t$	$\lambda_i^{R\prime} F_t^R$	e_{it}^R	$\gamma_j^{F\prime} G_t$	$\lambda_j^F F_t^F$	e_{jt}^F
Disaggregated Series						
Total	0.1729	0.3593	0.4677	0.0099	0.8495	0.1406
Average	0.1232	0.3002	0.5766	0.0103	0.8363	0.1534
Median	0.0585	0.2728	0.6143	0.0093	0.8464	0.1437
Minimum	0.0014	0.0251	0.0758	0.0015	0.6716	0.0451
Maximum	0.5291	0.8743	0.9697	0.0240	0.9477	0.3173
Std.	0.1435	0.1996	0.2474	0.0057	0.0648	0.0629
Aggregated Series						
Industrial Production	0.0712	0.7418	0.1869			
Personal Income	0.0374	0.1484	0.8141			
Nonfarm Employment	0.0593	0.4894	0.4513			
PCE	0.1145	0.2100	0.6754			
Dow Industrial Avg.				0.0397	0.5240	0.4363
S&P Composite				0.0397	0.5703	0.3900
S&P Industrials				0.0411	0.5674	0.3914

The above table shows that a vast majority of variation within real sector and stock markets

[7] I use the value-weighted return data. Data source: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

are due to different sources or factors. The level-1 factors, which are common to both sectors, only contribute to 17.29% of the total variation observed in real sector, and 0.99% of the total variation in stock market. For the real sector, 9 sector specific factors account for around the same amount of variations as idiosyncratic shock, while for the financial sector, 84.95% of the total variation is explained by 5 sector specific factors.

The sample covariance between sector specific factors is given by the following 5×9 matrix

$$\frac{\hat{F}^{F\prime}\hat{F}^{R}}{T} = \begin{pmatrix} -0.110 & 0.029 & 0.097 & 0.066 & 0.155 & -0.037 & -0.002 & 0.022 & 0.001 \\ -0.023 & -0.024 & -0.082 & -0.124 & -0.071 & -0.051 & -0.036 & -0.013 & -0.030 \\ -0.096 & 0.057 & 0.049 & -0.063 & -0.012 & 0.021 & -0.035 & -0.042 & 0.045 \\ -0.006 & -0.008 & 0.010 & -0.011 & 0.048 & 0.021 & 0.052 & 0.073 & 0.012 \\ 0.050 & 0.079 & -0.008 & -0.061 & 0.052 & -0.088 & 0.054 & -0.028 & -0.080 \end{pmatrix}$$

Coupled with the normalization that $\frac{\hat{F}^{F\prime}\hat{F}^{F}}{T} = I_5$ and $\frac{\hat{F}^{R\prime}\hat{F}^{R}}{T} = I_9$, the above matrix reveals only slightly correlations between sector specific factors.

Let $||A|| = [tr(A'A)]^{1/2}$ denote the norm of matrix A. The estimated factors are show to be orthogonal to idiosyncratic terms

$\lvert\lvert G'E^R/T \rvert\rvert$	$\lvert\lvert F^{R\prime}E^R/T \rvert\rvert$	$\lvert\lvert G'E^F/T \rvert\rvert$	$\lvert\lvert F^{F\prime}E^F/T \rvert\rvert$
$4.9 \cdot 10^{-29}$	$8.5 \cdot 10^{-6}$	$6.9 \cdot 10^{-30}$	$3.3 \cdot 10^{-9}$

Next, we compare our estimated factors with Fama-French benchmark 3 factors, denoted by $H_{T\times 3}$. We normalize H such that $H'H/T = I_3$. Sample covariance matrices are given by

$$\frac{H'G}{T} = \begin{bmatrix} 0.0275 & 0.0068 & -0.0990 \\ 0.1170 & -0.0144 & 0.0400 \\ -0.0684 & 0.0157 & 0.0044 \end{bmatrix}$$

$$\frac{H'F^R}{T} = \begin{bmatrix} -0.074 & 0.041 & 0.083 & 0.128 & 0.148 & -0.020 & 0.002 & 0.024 & -0.003 \\ -0.111 & 0.012 & 0.037 & -0.089 & 0.014 & -0.031 & -0.034 & 0.003 & 0.020 \\ -0.026 & 0.014 & 0.071 & 0.052 & -0.054 & 0.076 & -0.010 & -0.029 & 0.049 \end{bmatrix}$$

$$\frac{H'F^F}{T} = \begin{bmatrix} 0.8604 & -0.3010 & -0.2494 & -0.1817 & -0.0099 \\ 0.4305 & 0.5550 & 0.5401 & 0.2592 & 0.0121 \\ -0.1134 & -0.5242 & 0.5567 & -0.1636 & -0.1068 \end{bmatrix}$$

Two direct observations are in order. First, the Fama-French factors are only weekly correlated with our estimated common factors and factors specific to the real sector, which implies that the factors constructed by Fama and French are not able to explain variations in the real sector. Second, the Fama-French factors are strongly correlated with the first three factors specific to the financial sector.

In the next exercise, we first regress the Fama-French factor H on $\hat{F}^F$

$$H_t = R^H \hat{F}_t^F + U_t^H$$

and the resulting $R-$square is $R^2 = trace(T \cdot \hat{R}^{H\prime} \hat{R}^H)/trace(H'H) = 0.8048$, with $||\hat{F}^{F\prime}\hat{U}^H/T|| = 10^{-16}$. Then we regress H on both $\hat{G}$ and $\hat{F}^R$

$$H_t = R^1 \hat{G}_t + R^2 \hat{F}_t^R + U_t$$

and the resulting $R-$square is $R^2 = 0.0432$, with $||[\hat{G}, \hat{F}^R]'\hat{U}/T|| = 0.0379$. This exercise also suggests that, for the periods between 1976 and 2005, the Fama-French factors are largely specific to the financial sector or stock market, and have very limited explaining power for variations in the real sector.

8 Concluding Remarks

This paper develops a computationally simple estimation method, namely the iterative principal components method, to analyze large dimensional factor models with a multi-level factor structure. We treat common factors, sector-specific factors and factor loadings as parameters. Thus this method is nonparametric since we do not need to specify the dynamic process of the factors. The estimators explicitly take into account such a multi-level structure, which is not considered by the conventional principal components estimators. I prove that the estimators are consistent and have normal limiting distributions under very mild conditions. The estimators of common factors have a faster convergence rate than the estimators of sector-specific factors. The proposed estimation algorithm is easily implemented in practice and is computationally efficient. A two step procedure is proposed to select the number of both common factors and sector-specific factors. Monte Carlo experiments show that the iterative principal components estimators have nice finite sample performance. Such a model is then applied to investigate different patterns of comovement within real and financial sectors for the US economy. Empirical results suggest that the comovement within each sector is largely sector specific and the economy-wide common factors play only a limited role. The new method can also be readily applied to address a wide variety of issues in macroeconomics, international economics, labor economics and finance, where the multi-level structure is likely to present.

Future research agenda includes determining the number of both common factors and sector-specific factors based on their different convergence rate. A new theory is needed for a modified information criteria similar to Bai and Ng (2002)'s. Such a modification is necessary because common factors and sector-specific factors have different degree of pervasiveness. It is also interesting to develop a new large random matrix approach similar to the one proposed by Onatski (2006) to study the large dimensional factor models with a multi-level factor structure.

Another line of research would be empirical applications of the method developed in this paper. For example, in international economics, it is interesting to estimate both global common shocks and country-specific shocks, and then investigate how different types of shocks affect a country's monetary policy. The empirical results in this paper also suggests that one should be cautious when trying to extract global common factors from only the financial data, because those factors might ignore important information for the real sectors. If we want to study international business cycles using data from both real sector and financial sector, the multi-level factor model suggests a model of the following representation. Let x_{it}^s denote the time-t observation of the $i-th$ variable within country s, then

$$\begin{aligned}
x_{it}^s &= \gamma_i^{s\prime} G_t + \gamma_{Ri}^{s\prime} GR_t \cdot I_{is} + \gamma_{Fi}^{s\prime} GF_t \cdot (1 - I_{is}) + \lambda_i^{s\prime} F_t^s + \lambda_{Ri}^{s\prime} FR_t^s \cdot I_{is} + \lambda_{Fi}^{s\prime} FF_t^s \cdot (1 - I_{is}) + e_{it}^s, \\
i &= 1, ..., N_s, s = 1, ..., S, \\
G_t &: \text{ global common shock,} \\
GR_t &: \text{ global common shock specific to the real sector,} \\
GF_t &: \text{ global common shock specific to the financial sector,} \\
F_t^s &: \text{ shock specific to country } s, \text{ common to both real and financial sectors,} \\
FR_t^s &: \text{ shock specific to the real sector in country } s, \\
FF_t^s &: \text{ shock specific to the financial sector in country } s, \\
N &= N_1 + ... + N_S : \text{ total number of time series,} \\
S &: \text{ number of countries.}
\end{aligned}$$

where the indicator $I_{is} = 0$ if x_{it}^s is a financial variable, and $I_{is} = 1$ if x_{it}^s is a real variable.

Such a model divides the world economy into four levels of sectors. The top level is the global economy. The second level includes world real sectors and world financial sectors. The third level includes different countries. The fourth level consists of country-specific real sectors and financial sectors. Such a four-level factor model is easily estimated using the iterative principal components methods similar to Corollary 1. An inferential theory for the estimated factors similar to Theorem 2 and Theorem 3 can also be derived using similar asymptotic expansion method given in the appendix. This would be an important extension and generalization of the theory presented in this paper.

9 Appendix

Proof of theorem 1: Let $H^s = [G, F^s]$ be $T \times (r + r_s)$, then by 1) and 3), $H^{s\prime}H^s/T = I_{r+r_s}$. For each sector we have $x_t^s = [\Gamma^s, \Lambda^s]H_t^s + e_t^s$, thus for a least squares objective function, given H^s, we can solve for the optimal loadings as function of H^s. In particular

$$[\Gamma^s, \Lambda^s] = X^{s\prime}H^s/T$$

The objective function after concentrate out loadings becomes

$$\begin{aligned}
\max_{H^s} trace\left(\sum_{s=1}^{S} H^{s\prime}A^sH^s\right) &= trace\left(G'AG + \sum_{s=1}^{S} F^{s\prime}A^sF^s\right), \\
\text{with } A^s &= X^sX^{s\prime}, A = A^1 + ... + A^S. \\
s.t. F^{s\prime}F^s/T &= I_r, G'G/T = I_r, G'F = 0 \\
& \quad \Lambda^{s\prime}\Lambda^s \text{ and } \Gamma'\Gamma \text{diagonal}
\end{aligned}$$

Form the Lagrangian

$$\begin{aligned}
L = \sum_{i=1}^{r} G_i'AG_i + \sum_{s=1}^{S}\sum_{i=1}^{r_s} F_i^{s\prime}A^sF_i^s \quad & -\sum_{i=1}^{r}(G_i'G_i - T)a_i - \sum_{j>i}\sum_{i=1}^{r-1} G_i'G_jb_{ij} \\
& -\sum_{s=1}^{S}\sum_{i=1}^{r_s}(F_i^{s\prime}F_i^s - T)a_i^s - \sum_{s=1}^{S}\sum_{j>i}\sum_{i=1}^{r_s-1} F_i^{s\prime}F_j^sb_{ij}^s \\
& -\sum_{s=1}^{S}\sum_{j\geq i}\sum_{i=1}^{r} G_i'F_j^sc_{ij}^s
\end{aligned}$$

F.O.C. w.r.t. F_i^s, (given G)

$$0 = 2A^s F_i^s - 2F_i^s a_i^s - \sum_{j \neq i} F_j^s b_{ij}^s - \sum_j G_j c_{ij}^s$$

1) Left multiply $F_i^{s\prime}$ to obtain $a_i^s = F_i^{s\prime} A^s F_i^s / T$.

2) Left multiply $F_j^{s\prime}$ to obtain $b_{ij}^s = 2F_j^{s\prime} A^s F_i^s / T$.

3) Left multiply G_j' to obtain $c_{ij}^s = 2G_j' A^s F_i^s / T$. The implied F.O.C. becomes

$$\begin{aligned} 0 &= 2A^s F_i^s - 2F_i^s F_i^{s\prime} A^s F_i^s / T - 2\sum_{j \neq i} F_j^s F_j^{s\prime} A^s F_i^s / T - 2\sum_j G_j G_j' A^s F_i^s / T, \text{ or} \\ F_i^s (F_i^{s\prime} A^s F_i^s / T) &= \left(I - \sum_{j \neq i} F_j^s F_j^{s\prime} / T - \sum_j G_j G_j' / T \right) A^s F_i^s. \end{aligned}$$

Let $P_G = I - GG'/T = I - \sum_j G_j G_j'/T$. Then we may prove that F_i^s can be solved as eigenvectors of the matrix $P_G A^s$. To see this suppose

$$F_i^s \mu_i = P_G A^s F_i^s, i = 1, ..., r_s$$

Notice that $\mu_i = F_i^{s\prime} P_G A^s F_i^s / T = F_i^{s\prime} P_G A^s F_i^s / T$ because $F_i^{s\prime} G_j = 0$.

Moreover

$$\left(I - \sum_{j \neq i} F_j^s F_j^{s\prime} / T - \sum_j G_j G_j' / T \right) A^s F_i^s = P_G A^s F_i^s$$

because $F_j^{s\prime} A^s F_i^s = F_j^{s\prime} A^s P_G A^s F_i^s / \mu_i = (F_j^{s\prime} A^s P_G)(P_G A^s F_i^s / \mu_i) = \mu_j F_j^{s\prime} F_i = 0$ for $i \neq j$.

Notice that we use the assumption that $G'G/T = I$, and thus $P_G = I - GG'/T = P_G = I - G(G'G)^{-1}G'$ is the projection matrix with $P_G P_G = P_G$.

F.O.C. w.r.t. G_i, (given F)

$$0 = 2AG_i - 2G_i a_i - \sum_{j \neq i} G_j b_{ij} - \sum_s \sum_j F_j^s c_{ij}^s$$

Use the same method to solve for the Lagrangian multiplier to obtain

$$\begin{aligned} a_i &= G_i' A G_i / T \\ b_{ij} &= 2G_j' A G_i / T \\ c_{ij}^s &= 2F_j^{s\prime} A G_i / T (\text{if assuming } F_j^{s_1\prime} F_i^{s_2} = 0) \end{aligned}$$

The implied F.O.C. becomes

$$G_i(G_i'AG_i/T) = \left(I - \sum_{j\neq i} G_jG_j'/T - \sum_s \sum_j F_j^s F_j^{s\prime}/T\right) AG_i.$$

If we assume $F_j^{s_1\prime} F_i^{s_2} = 0$, then the following solution will do the job:

$$\begin{aligned} G_i\mu_i &= P_F AG_i, \text{ with} \\ P_F &= I - \sum_s \sum_j F_j^s F_j^{s\prime}/T, \; P_F \text{ is projection matrix because } F_j^{s_1\prime} F_i^{s_2} = 0 \end{aligned}$$

If $F_j^{s_1\prime} F_i^{s_2} \neq 0$, we left multiply F.O.C. by $F_j^{s\prime}$ to obtain

$$F_j^{s\prime} AG_i = F_j^{s\prime} \sum_{m\neq s} \sum_k F_k^m c_{ik}^m + Tc_{ij}^s$$

Then after some algebra we can solve for G and F by defining $P_F = I - F(F'F)^{-1}F'$ and

G consists of r eigenvectors for $P_F A$ with respect to its largest r eigenvalues.

F^s consists of r_s eigenvectors for $P_G A^s$ with respect to its largest r_s eigenvalues.

Q.E.D.

Remark: Given the sum of squared residuals objective function, given Γ and G, the objective function and the restrictions are the same as the one resulting in principal components estimators for Λ^s and F^s, and vice versa for Γ and G. This is the base for iterative principal components algorithm.

Proof of the proposition 2: Consider any rotation matrix such that

$$\begin{pmatrix} I_r & I_r & 0 \\ * & * & 0 \\ * & 0 & I_r \\ * & 0 & * \end{pmatrix} \begin{pmatrix} R1 & R2 & R3 \\ R4 & R5 & R6 \\ R7 & R8 & R9 \end{pmatrix} = \begin{pmatrix} I_r & I_r & 0 \\ * & * & 0 \\ * & 0 & I_r \\ * & 0 & * \end{pmatrix}$$

The zero restrictions imply $R3 = R6 = R2 = R8 = 0$, and other restrictions imply $R9 = R5 = I_r$, $R1 + R4 = I_r$, which pins down the equation to

$$\begin{pmatrix} I_r & I_r & 0 \\ * & * & 0 \\ * & 0 & I_r \\ * & 0 & * \end{pmatrix} \begin{pmatrix} R1 & 0 & 0 \\ I_r - R1 & I_r & 0 \\ R7 & 0 & I_r \end{pmatrix} = \begin{pmatrix} I_r & I_r & 0 \\ * & * & 0 \\ * & 0 & I_r \\ * & 0 & * \end{pmatrix}$$

Now look at the inverse of the implied rotation matrix

$$\begin{pmatrix} R1 & 0 & 0 \\ I_r - R1 & I_r & 0 \\ R7 & 0 & I_r \end{pmatrix} \begin{pmatrix} R1^{-1} & 0 & 0 \\ I_r - R1^{-1} & I_r & 0 \\ -R7 \cdot R1^{-1} & 0 & I_r \end{pmatrix} = \begin{pmatrix} I_r & 0 & 0 \\ 0 & I_r & 0 \\ 0 & 0 & I_r \end{pmatrix},$$

or $\text{inv}\begin{pmatrix} R1 & 0 & 0 \\ I_r - R1 & I_r & 0 \\ R7 & 0 & I_r \end{pmatrix} = \begin{pmatrix} R1^{-1} & 0 & 0 \\ I_r - R1^{-1} & I_r & 0 \\ -R7 \cdot R1^{-1} & 0 & I_r \end{pmatrix}$ which implies rotated factors of the following form

$\begin{pmatrix} R1^{-1} & 0 & 0 \\ I_r - R1^{-1} & I_r & 0 \\ -R7 \cdot R1^{-1} & 0 & I_r \end{pmatrix} \begin{pmatrix} G_t \\ F_t \\ F_t^* \end{pmatrix} = \begin{pmatrix} R1^{-1}G_t = g_t \\ (I_r - R1^{-1})G_t + F_t = f_t \\ -R7 \cdot R1^{-1}G_t + F_t^* = f_t^* \end{pmatrix}$. The restrictions $G'F = 0, G'F^* = 0, \sum_{t=1}^{T} g_t f_t' = 0$ and $\sum_{t=1}^{T} g_t f_t^{*\prime} = 0$ imply that $R1 = I_r$ and $R7 = 0$.

The above analysis implies that maximum likelihood analysis or minimum distance estimation of the above system yields unique solution. **Q.E.D.**

Proof of Proposition 3: First notice that

$$\begin{aligned} \begin{pmatrix} \hat{G}_t \\ \hat{F}_t \\ \hat{F}_t^* \end{pmatrix} &= (\hat{B}'\hat{B})^{-1}\hat{B}'(B \begin{pmatrix} G_t \\ F_t \\ F_t^* \end{pmatrix} + \begin{pmatrix} u_t \\ u_t^* \end{pmatrix}) \\ &= \begin{pmatrix} G_t \\ F_t \\ F_t^* \end{pmatrix} + (\hat{B}'\hat{B})^{-1}\hat{B}'(B - \hat{B}) \begin{pmatrix} G_t \\ F_t \\ F_t^* \end{pmatrix} + O_p(\frac{1}{\sqrt{M}}) \\ &= \begin{pmatrix} G_t \\ F_t \\ F_t^* \end{pmatrix} + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\sqrt{M}}) \end{aligned}$$

which proved the $\sqrt{M}-$ consistency of $\begin{pmatrix} \hat{G}_t \\ \hat{F}_t \\ \hat{F}_t^* \end{pmatrix}$ given that $\frac{M}{T} \to c < \infty$.Then we proved that $\sqrt{M}\left(\hat{G}_t - G_t\right) = O_p(1)$, $\sqrt{M}\left(\hat{F}_t - F_t\right) = O_p(1)$, and $\sqrt{M}\left(\hat{F}_t^* - F_t^*\right) = O_p(1)$. Notice that here G_t, F_t and F_t^* are still rotations of the original true G_t, F_t and F_t^* respectively, the above representations ignore the rotation matrix just to save notation. **Q.E.D.**

Proof of theorem 2: First notice that

$$\hat{F}_t^s - H^{s\prime} F_t^s = J1 + J2 + J3$$

where $J1 = (V_{MT}^s)^{-1} \frac{1}{T} \sum_{k=1}^T \hat{F}_k^s \frac{u_k^{s\prime} u_t^s}{M}$, $J2 = (V_{MT}^s)^{-1} \left(\frac{\hat{F}^{s\prime} F^s}{T}\right) \frac{1}{M} \sum_{i=1}^M \lambda_i^s u_{it}^s$,
and $J3 = (V_{MT}^s)^{-1} \frac{1}{T} \sum_{k=1}^T \hat{F}_k^s \frac{u_k^{s\prime} \Lambda^s F_t^s}{M}$.
Considering $J2$, we have proved that

$$\sqrt{M} J2 = (V_{MT}^s)^{-1} \left(\frac{\hat{F}^{s\prime} F^s}{T}\right) \frac{1}{\sqrt{M}} \sum_{i=1}^M \lambda_i^s e_{it}^s + o_p(1)$$

Considering $J1$ and using the restriction that $\sum_{k=1}^T \hat{F}_k^s \hat{G}_k' = 0$, we have

$$\begin{aligned}
\sqrt{M} J1 &= (V_{MT}^s)^{-1} \frac{1}{\sqrt{MT}} \sum_{k=1}^T \hat{F}_k^s \left(\Gamma^s G_k - \hat{\Gamma}^s \hat{G}_k + e_k^s\right)' \left(\Gamma^s G_t - \hat{\Gamma}^s \hat{G}_t + e_t^s\right) \\
&= (V_{MT}^s)^{-1} \frac{1}{\sqrt{MT}} \sum_{k=1}^T \hat{F}_k^s G_k' \Gamma^{s\prime} \left(\Gamma^s G_t - \hat{\Gamma}^s \hat{G}_t + e_t^s\right) \\
&\quad + (V_{MT}^s)^{-1} \frac{1}{\sqrt{MT}} \sum_{k=1}^T \hat{F}_k^s e_k^{s\prime} \left(\Gamma^s G_t - \hat{\Gamma}^s \hat{G}_t + e_t^s\right)
\end{aligned}$$

The first term

$$\begin{aligned}
& (V_{MT}^s)^{-1} \frac{1}{\sqrt{MT}} \sum_{k=1}^T \hat{F}_k^s G_k' \Gamma^{s\prime} \left(\Gamma^s G_t - \hat{\Gamma}^s \hat{G}_t + e_t^s\right) \\
= \; & (V_{MT}^s)^{-1} \frac{1}{\sqrt{MT}} \sum_{k=1}^T \hat{F}_k^s G_k' \Gamma^{s\prime} \Gamma^s H'^{-1} (H' G_t - \hat{G}_t) \\
& + (V_{MT}^s)^{-1} \frac{1}{\sqrt{MT}} \sum_{k=1}^T \hat{F}_k^s G_k' \Gamma^{s\prime} (\Gamma^s H'^{-1} - \hat{\Gamma}^s) \hat{G}_t \\
& + (V_{MT}^s)^{-1} \frac{1}{\sqrt{MT}} \sum_{k=1}^T \hat{F}_k^s G_k' \Gamma^{s\prime} e_t^s
\end{aligned}$$

$$
\begin{aligned}
&= (V_{MT}^s)^{-1}\left(\frac{1}{T}\sum_{k=1}^{T}\hat{F}_k^s G_k'\right)\frac{\Gamma^{s\prime}\Gamma^s}{M}H'^{-1}\sqrt{M}(H'G_t-\hat{G}_t) \\
&\quad +(V_{MT}^s)^{-1}\left(\frac{1}{T}\sum_{k=1}^{T}\hat{F}_k^s G_k'\right)\frac{1}{\sqrt{M}}\Gamma^{s\prime}(\Gamma^s H'^{-1}-\hat{\Gamma}^s)\hat{G}_t \\
&\quad +(V_{MT}^s)^{-1}\left(\frac{1}{T}\sum_{k=1}^{T}\hat{F}_k^s G_k'\right)\frac{1}{\sqrt{M}}\Gamma^{s\prime}e_t^s \\
&= o_p(1)
\end{aligned}
$$

To see the last equality, notice that

$$
\begin{aligned}
\sqrt{M}(H'G_t-\hat{G}_t) &= o_p(1) \\
\frac{1}{\sqrt{M}}\Gamma^{s\prime}(\Gamma^s H'^{-1}-\hat{\Gamma}^s) &= \frac{\sqrt{M}}{\min\{M,T\}} = o_p(1) \text{ given } \frac{\sqrt{M}}{T}\to 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{T}\sum_{k=1}^{T}\hat{F}_k^s G_k' \\
&= \frac{1}{T}\sum_{k=1}^{T}\hat{F}_k^s \hat{G}_k'(H)^{-1}+\frac{1}{T}\sum_{k=1}^{T}\hat{F}_k^s\left(G_k-(H')^{-1}\hat{G}_k\right) \\
&= \frac{1}{T}\sum_{k=1}^{T}\hat{F}_k^s\left(H'G_k-\hat{G}_k\right)'(H)^{-1}=o_p(1).
\end{aligned}
$$

$\frac{1}{\sqrt{M}}\Gamma^{s\prime}e_t^s=O_p(1)$ and $\frac{\Gamma^{s\prime}\Gamma^s}{M}=O_p(1)$ are by assumption.

The second term

$$
\begin{aligned}
&(V_{MT}^s)^{-1}\frac{1}{\sqrt{M}T}\sum_{k=1}^{T}\hat{F}_k^s e_k^{s\prime}\left(\Gamma^s G_t-\hat{\Gamma}^s\hat{G}_t+e_t^s\right) \\
&= (V_{MT}^s)^{-1}\frac{1}{\sqrt{M}T}\sum_{k=1}^{T}\hat{F}_k^s e_k^{s\prime}e_t^s \\
&\quad +(V_{MT}^s)^{-1}\frac{1}{\sqrt{M}T}\sum_{k=1}^{T}\hat{F}_k^s e_k^{s\prime}\left(\Gamma^s H'^{-1}(H'G_t-\hat{G}_t)+(\Gamma^s H'^{-1}-\hat{\Gamma}^s)\hat{G}_t\right) \\
&= o_p(1)
\end{aligned}
$$

where $\frac{1}{\sqrt{MT}}\sum_{k=1}^{T}\hat{F}_k^s e_k^{s\prime}e_t^s=o_p(1)$ comes from Bai (2003). $\frac{1}{\sqrt{MT}}\sum_{k=1}^{T}\hat{F}_k^s e_k^{s\prime}\Gamma^s H'^{-1}(H'G_t-$

$\hat{G}_t) = \left(\frac{1}{MT}\sum_{k=1}^T \hat{F}_k^s e_k^{s\prime}\Gamma^s\right) H'^{-1}\sqrt{M}(H'G_t - \hat{G}_t) = o_p(1)$. And $\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s e_k^{s\prime}(\Gamma^s H'^{-1} - \hat{\Gamma}^s)\hat{G}_t = \frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s e_k^{s\prime}(\Gamma^s H'^{-1} - \hat{\Gamma}^s)\right)\hat{G}_t = o_p(1)$.

Now considering $J3$,

$$
\begin{aligned}
\sqrt{M}J3 &= (V_{MT}^s)^{-1}\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s u_k^{s\prime}\Lambda^s F_t^s \\
&= (V_{MT}^s)^{-1}\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s(\Gamma^s G_k - \hat{\Gamma}^s\hat{G}_k + e_k^s)'\Lambda^s F_t^s
\end{aligned}
$$

Notice that $\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s e_k^{s\prime}\Lambda^s F_t^s = o_p(1)$ has been established by Bai (2003). The other term

$$
\begin{aligned}
& (V_{MT}^s)^{-1}\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s(\Gamma^s G_k - \hat{\Gamma}^s\hat{G}_k)'\Lambda^s F_t^s \\
= \; & (V_{MT}^s)^{-1}\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s\left(\Gamma^s H'^{-1}(H'G_k - \hat{G}_k) + (\Gamma^s H'^{-1} - \hat{\Gamma}^s)\hat{G}_k\right)'\Lambda^s F_t^s \\
= \; & (V_{MT}^s)^{-1}\frac{1}{\sqrt{MT}}\sum_{k=1}^T \hat{F}_k^s(H'G_k - \hat{G}_k)'H^{-1}\Gamma^{s\prime}\Lambda^s F_t^s \\
= \; & (V_{MT}^s)^{-1}\frac{1}{T}\sum_{k=1}^T \hat{F}_k^s\sqrt{M}(H'G_k - \hat{G}_k)'H^{-1}\frac{\Gamma^{s\prime}\Lambda^s}{M}F_t^s = o_p(1)
\end{aligned}
$$

So we have proved $\sqrt{M}J3 = o_p(1)$.

Combine all the above evidence, we proved that

$$
\begin{aligned}
\sqrt{M}(\hat{F}_t^s - H^{s\prime}F_t^s) &= \sqrt{M}J2 + o_p(1) \\
&= (V_{MT}^s)^{-1}\left(\frac{\hat{F}^{s\prime}F^s}{T}\right)\frac{1}{\sqrt{M}}\sum_{i=1}^M \lambda_i^s e_{it}^s + o_p(1) \\
&= A^s\frac{1}{\sqrt{M}}\sum_{i=1}^M \lambda_i^s e_{it}^s + o_p(1) \xrightarrow{d} N(0, A^s\Sigma_t A^{s\prime})
\end{aligned}
$$

where A^s is defined in the theorem. **Q.E.D.**

9.1 Proof of theorem 3

First we prove Lemma 1–3.

Lemma 1:

$$\begin{aligned} I1 &= \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s v_s' v_t = O_p(\frac{1}{\sqrt{N}}\sqrt{\frac{S}{M}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N}) \\ &= O_p(\frac{1}{\sqrt{N}}\sqrt{\frac{S}{M}}) + O_p(\frac{1}{\min\{N,T\}}). \end{aligned}$$

Proof of Lemma 1: Substitute $v_t^s = \Lambda^s F_t^s - \hat{\Lambda}^s \hat{F}_t^s + e_t^s$ into $I1$ to obtain

$$\begin{aligned} I1 &= \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s v_s' v_t = \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s(\Delta_s + e_s)'(\Delta_t + e_t) \\ &= \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s(\Delta_s + e_s)'(\Delta_t + e_t) \\ &= \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s(\Delta_s'\Delta_t + \Delta_s' e_t + e_s'\Delta_t + e_s' e_t) \end{aligned}$$

where the $N \times 1$ vector $\Delta_t = [\Lambda^j F_t^j - \hat{\Lambda}^j \hat{F}_t^j,\ j = 1, ..., S]'$.

Consider the first term

$$\begin{aligned} & \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s \Delta_s' \Delta_t \\ =\ & \frac{1}{NT}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s[F_s^{j\prime}\Lambda^{j\prime} - \hat{F}_s^{j\prime}\hat{\Lambda}^{j\prime}][\Lambda^j F_t^j - \hat{\Lambda}^j \hat{F}_t^j] \\ =\ & \frac{1}{NT}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s F_s^{j\prime}\Lambda^{j\prime}[\Lambda^j F_t^j - \hat{\Lambda}^j \hat{F}_t^j] \\ =\ & \frac{1}{N}\sum_{j=1}^{S}\frac{\sum_{s=1}^{T}\hat{G}_s F_s^{j\prime}}{T}\Lambda^{j\prime}[\Lambda^j F_t^j - \hat{\Lambda}^j \hat{F}_t^j] \\ =\ & \frac{1}{N}\sum_{j=1}^{S}[\frac{1}{T}\sum_{s=1}^{T}\hat{G}_s F_s^{j\prime}]\{\Lambda^{j\prime}[(\Lambda^j H^{j\prime -1} - \hat{\Lambda}^j)\hat{F}_t^j + \Lambda^j H^{j\prime -1}(H^{j\prime}F_t^j - \hat{F}_t^j)]\} \end{aligned}$$

where the second equality is based on the identification restriction $\sum_{s=1}^{T}\hat{G}_s\hat{F}_s^{j\prime} = 0$. Also we have

$$\begin{aligned}
\frac{1}{T}\sum_{s=1}^{T}\hat{G}_s F_s^{j\prime} &= \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s \hat{F}_s^{j\prime}(H^j)^{-1} + \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s (F_s^{j\prime} - \hat{F}_s^{j\prime}(H^j)^{-1}) \\
&= \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s (F_s^{j\prime} - \hat{F}_s^{j\prime}(H^j)^{-1}) = o_p(1)
\end{aligned}$$

Then consider the second term

$$\begin{aligned}
\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s \Delta_s' e_t &= \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s \sum_{j=1}^{S}\Delta_s^{j\prime} e_t^j \\
&= \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s \sum_{j=1}^{S} e_t^{j\prime}[(\Lambda^j H^{j\prime -1} - \hat{\Lambda}^j)\hat{F}_t^j + \Lambda^j H^{j\prime -1}(H^{j\prime}F_t^j - \hat{F}_t^j)]
\end{aligned}$$

where the second term inside the second summation

$$e_t^{j\prime}\Lambda^j H^{j\prime -1}(H^{j\prime}F_t^j - \hat{F}_t^j) = \{\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{it}^j\}' H^{j\prime -1} A^j \{\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{is}^j + o_p(1)\} = O_p(1)$$

The first term of the above equation

$$\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{S}\hat{G}_s e_t^{j\prime}(\Lambda^j H^{j\prime -1} - \hat{\Lambda}^j)\hat{F}_t^j = \frac{1}{NT}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s \hat{F}_s^{j\prime}(\Lambda^j H^{j\prime -1} - \hat{\Lambda}^j)' e_t^j = 0$$

by the identifying assumption $\sum_{s=1}^{T}\hat{G}_s\hat{F}_s^{j\prime} = 0$. In sum, the second term in $I1$ becomes

$$\begin{aligned}
\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s \Delta_s' e_t &= \frac{1}{NT}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s e_t^{j\prime}\Lambda^j H^{j\prime -1}(H^{j\prime}F_t^j - \hat{F}_t^j) \\
&= -\frac{1}{NT}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s[\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{is}^j + o_p(1)]' A^{j\prime}(H^j)^{-1}[\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{it}^j] \\
&= -\frac{1}{\sqrt{N}}\frac{1}{\sqrt{NT}}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s[\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{is}^j + o_p(1)]' A^{j\prime}(H^j)^{-1}[\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{it}^j]
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{\sqrt{N}}\frac{\sqrt{S}}{\sqrt{M}}\frac{1}{ST}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s[\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{is}^j + o_p(1)]' A^{j\prime}(H^j)^{-1}[\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{it}^j] \\
&= O_p(\frac{1}{\sqrt{N}}\sqrt{\frac{S}{M}}).
\end{aligned}
$$

Now consider the third term, $\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s e_s' \Delta_t$. Similarly $e_s'\Delta_t = \sum_{j=1}^{S}\Delta_t^{j\prime} e_s^j$ and

$$
\Delta_t^{j\prime} e_s^j = e_s^{j\prime}[\Lambda^j F_t^j - \hat{\Lambda}^j \hat{F}_t^j] = e_s^{j\prime}[(\Lambda^j H^{j\prime-1} - \hat{\Lambda}^j)\hat{F}_t^j + \Lambda^j H^{j\prime-1}(H^{j\prime}F_t^j - \hat{F}_t^j)]
$$

Still the second term

$$
e_s^{j\prime}\Lambda^j H^{j\prime-1}(H^{j\prime}F_t^j - \hat{F}_t^j) = -\{\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{is}^j\}' H^{j\prime-1} A^j \{\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{it}^j + o_p(1)\}
$$

. The first term $e_s^{j\prime}(\Lambda^j H^{j\prime-1} - \hat{\Lambda}^j)\hat{F}_t^j = o_p(1)$, which is dominated in magnitude by the second term. In sum

$$
\begin{aligned}
\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s e_s' \Delta_t &= -\frac{1}{NT}\sum_{j=1}^{S}\sum_{s=1}^{T}\hat{G}_s\{\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{is}^j\}' H^{j\prime-1} A^j \{\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{it}^j + o_p(1)\} \\
&= O_p(\frac{1}{\sqrt{N}}\sqrt{\frac{S}{M}}).
\end{aligned}
$$

Finally, consider the fourth term. The component $\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s e_s' e_t$ can be split into two terms

$$
\begin{aligned}
\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s e_s' e_t &= \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s \gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s \delta_{st}, \text{ where} \\
\delta_{st} &= \frac{e_s' e_t}{N} - \gamma_N(s,t) \text{ and } \gamma_N(s,t) = E(\frac{e_s' e_t}{N})
\end{aligned}
$$

Notice that

$$
\begin{aligned}
||\frac{1}{T}\sum_{s=1}^{T}\hat{G}_s \gamma_N(s,t)||^2 &\leq \left(\frac{1}{T}\sum_{s=1}^{T}||\hat{G}_s||^2\right)\left(\frac{1}{T}\sum_{s=1}^{T}\gamma_N^2(s,t)\right) \\
&= O_p(1)\cdot O_p(\frac{1}{T}) = O_p(\frac{1}{T})
\end{aligned}
$$

by assumption. And

$$||\frac{1}{T}\sum_{s=1}^{T}\hat{G}_s\delta_{st}||^2 \leq \left(\frac{1}{T}\sum_{s=1}^{T}||\hat{G}_s||^2\right)\left(\frac{1}{T}\sum_{s=1}^{T}\delta_{st}^2\right) = O_p(\frac{1}{N})$$

which comes from the observation that that $\left(\sum_{s=1}^{T}\delta_{st}^2\right)^2 \leq T\left(\sum_{s=1}^{T}\delta_{st}^4\right) \leq T\frac{T}{N^2}M$, thus $\frac{1}{T}\sum_{s=1}^{T}\delta_{st}^2 \leq \frac{1}{N}M^{1/2} = O_p(\frac{1}{N})$. **Q.E.D.**

Lemma 2:

$$I2 = \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s\eta_{st} = O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{N}})O_p(\frac{\sqrt{N}}{\min(M,T)})$$

Proof of Lemma 2:

$$\begin{aligned} I2 &= \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s\eta_{st} = \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_sG_s'\Gamma'v_t = \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_sG_s'\Gamma'[\Delta_t + e_t] \\ &= [\frac{1}{T}\sum_{s=1}^{T}\hat{G}_sG_s'] \cdot \frac{1}{N}\Gamma'\Delta_t + [\frac{1}{T}\sum_{s=1}^{T}\hat{G}_sG_s'] \cdot \frac{1}{N}\Gamma'e_t \end{aligned}$$

The second term

$$I22 = \frac{1}{\sqrt{N}} \cdot [\frac{1}{T}\sum_{s=1}^{T}\hat{G}_sG_s'] \cdot \frac{1}{\sqrt{N}}\Gamma'e_t = O_p(\frac{1}{\sqrt{N}})$$

determines the limiting distribution. The first term

$$\begin{aligned} I21 &= [\frac{1}{T}\sum_{s=1}^{T}\hat{G}_sG_s'] \cdot \frac{1}{N}\sum_{j=1}^{S}\Gamma^{j\prime}[\Lambda^jF_t^j - \hat{\Lambda}^j\hat{F}_t^j] \\ &= [\frac{1}{T}\sum_{s=1}^{T}\hat{G}_sG_s'] \cdot \frac{1}{N}\sum_{j=1}^{S}\Gamma^{j\prime}[(\Lambda^jH^{j\prime-1} - \hat{\Lambda}^j)\hat{F}_t^j + \Lambda^jH^{j\prime-1}(H^{j\prime}F_t^j - \hat{F}_t^j)] \end{aligned}$$

The term $[\frac{1}{T}\sum_{s=1}^{T}\hat{G}_sG_s']$ is $O_p(1)$, so the first term of $I21$ is $O_p(1)\cdot\frac{1}{S}\sum_{j=1}^{S}\frac{1}{M}\Gamma^{j\prime}(\Lambda^jH^{j\prime-1} - \hat{\Lambda}^j)\hat{F}_t^j = O_p(\frac{1}{\min(M,T)})$ by Lemma A.10 of Bai (2006).

The second term of $I21$ is $O_p(1) \cdot \frac{1}{N}\sum_{j=1}^{S}\Gamma^{j\prime}\Lambda^j H^{j\prime -1}(H^{j\prime}F_t^j - \hat{F}_t^j)$, where

$$\begin{aligned} H^{j\prime}F_t^j - \hat{F}_t^j &= -A^j\frac{1}{M}\sum_{i=1}^{M}\lambda_i^j e_{it}^j - A^j\frac{1}{M}\sum_{i=1}^{M}\lambda_i^j(H^{-1}\Gamma_i^j - \hat{\Gamma}_i^j)'\hat{G}_t \\ &\quad -A^j\frac{1}{M}\sum_{i=1}^{M}\lambda_i^j\Gamma_i^{j\prime}H'^{-1}(H'G_t - \hat{G}_t) + o_p(1). \end{aligned}$$

Substitute into $\frac{1}{N}\sum_{j=1}^{S}\Gamma^{j\prime}\Lambda^j H^{j\prime -1}(H^{j\prime}F_t^j - \hat{F}_t^j) = \frac{1}{N}\sum_{j=1}^{S}\sum_{i=1}^{M}\Gamma_i^j\Lambda_i^{j\prime}H^{j\prime -1}(H^{j\prime}F_t^j - \hat{F}_t^j)$ and analyze each term:

Firstly, we have

$$\begin{aligned} & -\frac{1}{N}\sum_{j=1}^{S}\left(\frac{1}{M}\Gamma^{j\prime}\Lambda^j H^{j\prime -1}A^j\Lambda^{j\prime}e_t^j\right) \\ = \quad & -\frac{1}{\sqrt{N}}\frac{1}{\sqrt{S}}\sum_{j=1}^{S}\left(\left(\frac{1}{M}\Gamma^{j\prime}\Lambda^j\right)H^{j\prime -1}\frac{1}{\sqrt{M}}A^j\Lambda^{j\prime}e_t^j\right) = O_p(\frac{1}{\sqrt{N}}). \end{aligned}$$

Secondly, we have

$$-\frac{1}{S}\sum_{j=1}^{S}\left(\frac{1}{M}\Gamma^{j\prime}\Lambda^j\right)H^{j\prime -1}A^j\frac{1}{M}\sum_{k=1}^{M}\lambda_k^j(H^{-1}\Gamma_k^j - \hat{\Gamma}_k^j)'\hat{G}_t = O_p(\frac{1}{\min(M,T)})$$

by Lemma A.10 of Bai (2006).

Lastly, we have

$$\begin{aligned} & -\frac{1}{S}\sum_{j=1}^{S}\left(\frac{1}{M}\Gamma^{j\prime}\Lambda^j\right)H^{j\prime -1}A^j\left(\frac{1}{M}\Lambda^{j\prime}\Gamma^j\right)H'^{-1}(H'G_t - \hat{G}_t) \\ = \quad & [\frac{1}{S}\sum_{j=1}^{S}\left(\frac{1}{M}\Gamma^{j\prime}\Lambda^j\right)H^{j\prime -1}A^j\left(\frac{1}{M}\Lambda^{j\prime}\Gamma^j\right)H'^{-1}](\hat{G}_t - H'G_t) = O_p(\frac{1}{\sqrt{N}}) \end{aligned}$$

which will be moved to the LHS to solve for the equilibrium fixed point representation. **Q.E.D.**

Lemma 3: $I3 = \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s\xi_{st} = O_p(\frac{1}{\sqrt{N}}\frac{1}{\sqrt{T}})$.

Proof of lemma 3:

$$
\begin{aligned}
I3 &= \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s\xi_{st} = \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s v_s'\Gamma G_t/N \\
&= \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s(\Delta_s+e_s)'\Gamma G_t/N = \frac{1}{T}\sum_{s=1}^{T}\hat{G}_s(\Delta_s+e_s)'\Gamma G_t/N.
\end{aligned}
$$

The term

$$
\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s(\Delta_s)'\Gamma G_t = \frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s\left(\sum_{j=1}^{S}[(\Lambda^j H^{j\prime-1}-\hat{\Lambda}^j)\hat{F}_s^j+\Lambda^j H^{j\prime-1}(H^{j\prime}F_s^j-\hat{F}_s^j)]'\Gamma^j\right)G_t
$$

in which the first part

$$
\frac{1}{NT}\sum_{j=1}^{S}\left(\sum_{s=1}^{T}\hat{G}_s\hat{F}_s^{j\prime}\right)\left[(\Lambda^j H^{j\prime-1}-\hat{\Lambda}^j)'\Gamma^j\right]G_t = 0
$$

by the identifying assumption $\sum_{s=1}^{T}\hat{G}_s\hat{F}_s^{j\prime}=0$ for all j, and the second part

$$
\begin{aligned}
&\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s\left(\sum_{j=1}^{S}[\Lambda^j H^{j\prime-1}(H^{j\prime}F_t^j-\hat{F}_t^j)]'\Gamma^j\right)G_t \\
=&\ \frac{1}{\sqrt{M}ST}\sum_{s=1}^{T}\hat{G}_s\left(\sum_{j=1}^{S}\sqrt{M}(H^{j\prime}F_t^j-\hat{F}_t^j)'(H^j)^{-1}\frac{1}{M}\Lambda^{j\prime}\Gamma^j\right)G_t \\
=&\ -\frac{1}{\sqrt{M}ST}\sum_{s=1}^{T}\hat{G}_s\left(\sum_{j=1}^{S}(\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\Lambda_i^j e_{is}^j+o_p(1))'(H^j)^{-1}\frac{1}{M}\Lambda^{j\prime}\Gamma^j\right)G_t \\
=&\ -\frac{1}{\sqrt{N}}\frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{S}\sum_{i=1}^{M}\hat{G}_s e_{is}^j\Lambda_i^{j\prime}(H^j)^{-1}\frac{1}{M}\Lambda^{j\prime}\Gamma^j\right)G_t \\
=&\ O_p(\frac{1}{\sqrt{N}}\frac{1}{\sqrt{T}}).
\end{aligned}
$$

The term $\frac{1}{NT}\sum_{s=1}^{T}\hat{G}_s e_s{}'\Gamma G_t = \frac{1}{NT}\sum_{s=1}^{T}(\hat{G}_s - H'G_s)e_s{}'\Gamma G_t + \frac{1}{NT}\sum_{s=1}^{T}H'G_s e_s{}'\Gamma G_t = O_p(1/\sqrt{NT})$. **Q.E.D.**

Proof of Theorem 3: Recall the data generating process can be represented as

$$\begin{aligned} z_t^s &= \Gamma^s G_t + v_t^s, \text{ where} \\ z_t^s &= x_t^s - \hat{\Lambda}^s \hat{F}_t^s \\ v_t^s &= \Lambda^s F_t^s - \hat{\Lambda}^s \hat{F}_t^s + e_t^s \end{aligned}$$

From Lemma 1–3, as $\sqrt{N}/T \to 0$ and $S/M \to 0$, we have

$$\begin{aligned} \sqrt{N} V_{NT}(\hat{G}_t - H'G_t) &= \frac{1}{\sqrt{N}T} \sum_{k=1}^{T} \hat{G}_k G_k' \Gamma' v_t + o_p(1) \\ &= \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{k=1}^{T} \hat{G}_k G_k' \right) \sum_{s=1}^{S} \sum_{i=1}^{M} \gamma_i^s v_{it}^s + o_p(1) \\ &= \left(\frac{1}{T} \sum_{k=1}^{T} \hat{G}_k G_k' \right) \frac{1}{\sqrt{N}} \sum_{s=1}^{S} \sum_{i=1}^{M} \gamma_i^s e_{it}^s \\ &\quad + \left(\frac{1}{T} \sum_{k=1}^{T} \hat{G}_k G_k' \right) \frac{1}{\sqrt{N}} \sum_{s=1}^{S} \sum_{i=1}^{M} \gamma_i^s ((H^s)^{-1} \lambda_i^s - \hat{\lambda}_i^s)' \hat{F}_t^s \\ &\quad + \left(\frac{1}{T} \sum_{k=1}^{T} \hat{G}_k G_k' \right) \frac{1}{\sqrt{N}} \sum_{s=1}^{S} \sum_{i=1}^{M} \gamma_i^s \lambda_i^{s\prime} (H^{s\prime})^{-1} (H^{s\prime} F_t^s - \hat{F}_t^s) + o_p(1) \end{aligned}$$

The term $\left(\frac{1}{T} \sum_{k=1}^{T} \hat{G}_k G_k' \right) \frac{1}{\sqrt{N}} \sum_{s=1}^{S} \sum_{i=1}^{M} \gamma_i^s e_{it}^s$ is $O_p(1)$, which partly determines the limiting distribution. From the proof of Lemma 2, we know that

$$\begin{aligned} & [\frac{1}{T} \sum_{s=1}^{T} \hat{G}_s G_s'] \cdot \frac{1}{\sqrt{N}} \sum_{s=1}^{S} \sum_{i=1}^{M} \gamma_i^s ((H^s)^{-1} \lambda_i^s - \hat{\lambda}_i^s)' \hat{F}_t^s \\ = \;& O_p(\frac{\sqrt{N}}{\min(M,T)}) = o_p(1), \text{ assuming } \frac{S}{M} \to 0. \end{aligned}$$

From the proof of theorem 2, we have

$$\begin{aligned} H^{s\prime} F_t^s - \hat{F}_t^s &= -A^s \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s e_{it}^s - A^s \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s (H^{-1} \gamma_i^s - \hat{\gamma}_i^s)' \hat{G}_t \\ &\quad - A^s \frac{1}{M} \sum_{i=1}^{M} \lambda_i^s \gamma_i^{s\prime} H'^{-1} (H' G_t - \hat{G}_t) + O_p(\frac{1}{\sqrt{M} \min\{\sqrt{M}, \sqrt{T}\}}) \end{aligned}$$

Substitute into $\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\sum_{i=1}^{M}\gamma_i^s\lambda_i^{s\prime}(H^{s\prime})^{-1}(H^{s\prime}F_t^s-\hat{F}_t^s)$ and analyze each term:

Firstly, we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\sum_{i=1}^{M}\gamma_i^s\lambda_i^{s\prime}(H^{s\prime})^{-1}(-A^s\frac{1}{M}\sum_{i=1}^{M}\lambda_i^s e_{it}^s)\\
=\;& -\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\sum_{j=1}^{M}\frac{\Gamma^{s\prime}\Lambda^s}{M}(H^{s\prime})^{-1}A^s\lambda_j^s e_{jt}^s=O_p(1)
\end{aligned}$$

which partly determines the limiting distribution.

Secondly, we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\sum_{i=1}^{M}\gamma_i^s\lambda_i^{s\prime}(H^{s\prime})^{-1}(-A^s\frac{1}{M}\sum_{i=1}^{M}\lambda_i^s(H^{-1}\gamma_i^s-\hat{\gamma}_i^s)'\hat{G}_t)\\
=\;& -\sqrt{N}\frac{1}{S}\sum_{s=1}^{S}\frac{\Gamma^{s\prime}\Lambda^s}{M}(H^{s\prime})^{-1}A^s\frac{1}{M}\sum_{i=1}^{M}\lambda_i^s(H^{-1}\gamma_i^s-\hat{\gamma}_i^s)'\hat{G}_t\\
=\;& O_p(\frac{\sqrt{N}}{\min(M,T)})=o_p(1),\ \text{assuming } \frac{S}{M}\to 0.
\end{aligned}$$

Lastly, we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\sum_{i=1}^{M}\gamma_i^s\lambda_i^{s\prime}(H^{s\prime})^{-1}(-A^s\frac{1}{M}\sum_{i=1}^{M}\lambda_i^s\gamma_i^{s\prime}H'^{-1}(H'G_t-\hat{G}_t))\\
=\;& \left\{\frac{1}{S}\sum_{s=1}^{S}\frac{\Gamma^{s\prime}\Lambda^s}{M}(H^{s\prime})^{-1}A^s\frac{\Lambda^{s\prime}\Gamma^s}{M}H'^{-1}\right\}\left\{\sqrt{N}(\hat{G}_t-H'G_t)\right\}\\
=\;& O_p(1)
\end{aligned}$$

The asymptotic equilibrium representation for $\sqrt{N}(\hat{G}_t-H'G_t)$ is given by

$$\begin{aligned}
&\sqrt{N}(\hat{G}_t-H'G_t)\\
=\;& (V_{NT}-\Omega)^{-1}\cdot\left\{\left(\frac{1}{T}\sum_{k=1}^{T}\hat{G}_kG_k'\right)\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\sum_{i=1}^{M}\left(\gamma_i^s-\frac{\Gamma^{s\prime}\Lambda^s}{M}(H^{s\prime})^{-1}A^s\lambda_i^s\right)e_{it}^s\right\}+o_p(1)\\
=\;& (V_{NT}-\Omega)^{-1}\cdot\left\{\left(\frac{1}{T}\sum_{k=1}^{T}\hat{G}_kG_k'\right)\frac{1}{\sqrt{N}}\sum_{s=1}^{S}\Gamma^{s\prime}\left(I_M-\frac{1}{M}\Lambda^s(H^{s\prime})^{-1}A^s\Lambda^{s\prime}\right)e_t^s\right\}+o_p(1)\\
\to\;& N(0,\Sigma_t)
\end{aligned}$$

where

$$
\begin{aligned}
\Omega &= \left(\frac{1}{T}\sum_{k=1}^{T}\hat{G}_k G_k'\right)\left\{\frac{1}{S}\sum_{s=1}^{S}\frac{\Gamma^{s\prime}\Lambda^s}{M}(H^{s\prime})^{-1}A^s\frac{\Lambda^{s\prime}\Gamma^s}{M}H^{\prime -1}\right\} \\
A^s &= plim V_{MT}^{-1}\frac{1}{T}\sum_{k=1}^{T}(\hat{F}_k^s F_k^{s\prime})
\end{aligned}
$$

with V_{MT}^s *being a diagonal matrix consisting of the first* r *eigenvalues of* $\frac{Y^sY^{s\prime}}{MT}$ *in decreasing order.* **Q.E.D.**

References

[1] Amengual, D. and M. Watson, 2007, "Consistent estimation of the number of dynamic factors in large N and T panel," Journal of Business and Economic Statistics 25(1), 91–96.

[2] Anderson, H. and F. Vahid, 2007, "Forecasting the volatility of Australian stock returns: Do common factors help?" Journal of the American Statistical Association 25(1), 75–90.

[3] Anderson, T. W., 1984, *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

[4] Andrews, D. W. K., 2005, "Cross-section regression with common shocks," Econometrica, 73, 1551–1585.

[5] Bai, Jushan, 2003, "Inferential theory for factor models of large dimensions," Econometrica 71(1), 135–172.

[6] ——— 2004, "Estimating cross-section common stochastic trends in non-stationary panel data," Journal of Econometrics 122, 137–183.

[7] ——— 2006, "Panel data models with interactive fixed effects," Department of Economics, New York University, Unpublished Manuscript.

[8] Bai, Jushan and J. L. Carrion-i-Silvestre, 2004, "Structural changes, common stochastic trends, and unit roots in panel data," Unpublished Manuscript.

[9] ——— 2005, "Testing panel cointegration with unobservable dynamic common factors," Department of Economics, New York University, Unpublished Manuscript.

[10] Bai, Jushan and Serena Ng, 2002, "Determining the number of factors in approximate factor models," Econometrica, 70(1), 191-221.

[11] ——— 2006a, "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions," Econometrica, 74(4), 1133-1150.

[12] ——— 2006b, "Evaluating latent and observed factors in macroeconomics and finance," Journal of Econometrics, 113(1-2), 507-537.

[13] ——— 2007, "Determining the number of primitive shocks," Journal of business and Economic Statistics, 25(1), 52-60.

[14] ——— 2008, "Large dimensional factor analysis," Foundations and Trends in Econometrics, Vol. 3, No. 2, p89–163.

[15] Bernanke, B. and J. Boivin, 2003, "Monetary policy in a data rich environment," Journal of Monetary Economics, 50(3), 525-546.

[16] Bernanke, B., J. Boivin, and P. Eliasz, 2005, "Measuring monetary policy: a factor augmented vector autoregressive (FAVAR) approach," Quarterly Journal of Economics, 120(1).

[17] Boivin, J. and S. Ng, 2006, "Are more data always better for factor analysis?" Journal of Econometrics, 132, 169-194.

[18] Boivin, J. and M. Giannoni, 2006, "DSGE models in a data-rich environment," Unpublished Manuscript.

[19] Boivin, J., M. Giannoni and I. Mihov, 2007, "Sticky prices and monetary policy: evidence from disaggregated U.S. data," forthcoming in The American Economic Review.

[20] Chamberlain, G. and M. Rothschild, 1983, "Arbitrage, factor structure and mean-variance analysis in large asset markets," Econometrica, 51, 1281–1304.

[21] Doz, C., D. Giannone, and L. Reichlin, 2007, "A quasi-maximum likelihood approach for large approximate dynamic factor models," European Central Bank Working Paper Series 674.

[22] Forni, M., D. Giannone, M. Lippi, and L. Reichlin, 2003, "Opening the black box: Identifying shocks and propagation mechanisms in VAR and factor models," Unpublished Manuscript.

[23] 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.

[24] ——— 2001, "Do financial variables help in forecasting inflation and real activity in the euro area," Unpublished Manuscript.

[25] ——— 2004, "The generalized factor model: consistency and rates," Journal of Econometrics 119, 231–255.

[26] ——— 2005, "The generalized dynamic factor model, one sided estimation and forecasting," Journal of the American Statistical Association 100, 830–840.

[27] Geweke, John F. and Kenneth J. Singleton, 1981, "Maximum likelihood 'Confirmatory' Factor Analysis of Economic Time Series," International Economic Review, Vol. 22, No. 1, p37-54.

[28] Kose, M. Ayhan, Chris Otrok and Charles H. Whiteman, 2003, "International business cycles: world, region and country specific factors," American Economic Review, Vol. 93, No. 4, p1216-1239.

[29] Ludvigson, S. and S. Ng, 2005, "Macro factors in bond risk premia," NBER Working Paper 11703.

[30] ——— 2007, "The empirical risk return relation: A factor analysis approach," Journal of Financial Economics 83, 171–222.

[31] Marcellino, M., J. H. Stock, and M. Watson, 2003, "Macroeconomic forecasting in the Euro area: country specific versus Euro wide information," European Economic Review 47, 1–18.

[32] Moon, R. and B. Perron, 2004, "Testing for a unit root in panels with dynamic factors," Journal of Econometrics 122(1), 81–126.

[33] Moon, R., B. Perron, and P. Phillips, 2007, "Incidental trends and the power of panel unit root tests," Journal of Econometrics 141(2), 416–459.

[34] Onatski, A., 2005, "Determining the number of factors from empirical distribution of eigenvalues," Department of Economics, Columbia University, Discussion Paper 0405-19.

[35] ——— 2006a, "Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak," Department of Economics, Columbia University, Unpublished Manuscript.

[36] ——— 2006b, "A formal statistical test for the number of factors in approximate factor models," Department of Economics, Columbia University, Unpublished Manuscript.

[37] Phillips, P. C. B. and D. Sul, 2003, "Dynamic panel estimation and homogeneity testing under cross-section dependence," Econometrics Journal 6(1), 217–259.

[38] Quah, D. and T. Sargent, 1992, "A dynamic index model for large cross sections," Federal Reserve Bank of Minneapolis, Discussion Paper 77.

[39] Sargent, T. and C. Sims, 1977, "Business cycle modelling without pretending to have too much a priori economic theory," In: C. Sims (ed.): New Methods in Business Cycle Research. Minneapolis: Federal Reserve Bank of Minneapolis.

[40] Stock, J. H. and M. W. Watson, 1988, "Testing for common trends," Journal of the American Statistical Association 83, 1097–1107.

[41] ——— 1998, "Diffusion Indexes," NBER Working Paper 6702.

[42] ——— 2002a, "Forecasting using principal components from a large number of predictors," Journal of the American Statistical Association 97, 1167–1179.

[43] ——— 2002b, "Macroeconomic forecasting using diffusion indexes," Journal of Business and Economic Statistics 20(2), 147–162.

[44] ——— 2005, "Implications of dynamic factor models for VAR analysis," NBER Working Paper 11467.

[45] ——— 2006, "Forecasting with many predictors," Handbook of Economic Forecasting. North Holland: Elsevier.

[46] Watson, M. and R. Engle, 1983, "Alternative algorithms for the estimation of dynamic factor, MIMIC, and varying coefficient regression models," Journal of Econometrics 23, 385–400.