Common Factors and Specific Factors

Chen, Pu

Melbourne Insitute of Technology

20 January 2012

Common Factors and Specific Factors

Pu Chen*

20.01.2012

Abstract

In this paper we study factor models for security returns on financial markets, where some pervasive factors are common across all securities and other pervasive factors prevail only within some groups of securities but not in others. This kind of structured factors allow a more nuanced analysis of determinants of the security returns, in particular, they allow to study clustering structures in security returns as well as their determinants. The clustering structure provides a natural way to group the securities and to interpret common factors and group-specific factors. We give conditions under which the common factor space and the group-specific factor spaces can be identified, and propose an effective procedure to estimate the unobservable structure in the factor space. Concretely, the procedure will determine the unknown number of groups, endogenously classify securities into groups, determine the number of common factors across all groups as well as the number of group-specific factors in each group, and estimate the common factors and the group-specific factors. The estimated factor structure will provides a more meaningful interpretation of the estimated factors in practical applications.

KEYWORDS: Factor Models, Generalized Principal Component Analysis, Multiset Canonical Correlation

JEL Classification: C63, G10, G12,

*Melbourne Institute of Technology, E-Mail: chenpu60@hotmail.com

1 Introduction

The arbitrage pricing theory (ATP) of Ross (1976) provides a theoretically founded multifactor model for asset pricing. A key feature of the APT is that the random returns of each asset are assumed to be driven by a linear combination of a small number of common factors and an asset-specific random factor. The asset specific factor called idiosyncratic factor is assumed to be independent across all assets[1]. A series of statistical procedures (see Chapter 6 in Campell, Lo, and Mackenlay (1997)) have been developed to carry out empirical investigations based on APT. A large body of empirical literature on APT documents the success of APT.

APT maintains the possibility that some common factors that are pervasive across all assets, while some other pervasive factors prevail only within a group of assets. Group-specific factors are particularly useful in understanding data of the groups. So, industrial indices that are considered as industry-specific factors are used to measure industry-specific risks that can in turn explain the asset returns in respective industries (See Fama and French (1993) for more details.). In characteristic-based factor models it becomes a common practice to use industry-country factors as group-specific factors (see L. and Rouwenborst (1994)). In statistical factor models treatments of grouped data have a long tradition. Classic methods of factor rotation[2] can be seen as procedures that implicitly seeking for some kinds of grouped structures in factor spaces. By discovering a "simple structure" in the factor space, variables are divided accordingly into groups at the same time. In the literature there are also works that explicitly study grouped data in statistical factor models. Krzanowski (1979) considers the situation in which the group-specific factor spaces and the common factor space are the same and proposed to determine the common factor space by minimizing its angles to the group-specific factor spaces. Flury (1984) and Flury (1987) consider the case in which all group-specific covariance matrices can be orthogonalized by a same matrix. This method is then extended by Schott (1999) to take into account of the situation in which the group-specific factor spaces are only subspaces of the common factor space respectively. He suggests to estimate the common factor space by applying principal components method to the sum of the eigenprojection of each group. Goyal, Perignon, and Villa (2008) apply this method to study the asset returns in NYSE and NASDAQ and find that these two markets share one common factor and each market has one group-specific factor respectively.

In the papers on grouped factor models listed above, the most important model parameters of the grouped structure: the number of groups, the membership relations between groups and variables are assumed to be known *a priori* rather than estimated from observed data. In many cases even the numbers of common factors and group-specific factors are also given *a priori*. An attempt that assumes the grouped structure is unknown and estimates it from observed data is given in Chen (2010). The author applies the method of generalized principle component analysis to classify the variables into their groups and uses an information criterion to determine the number of groups and number of factors in each groups. After the

[1]Cross-sectional independence is required only in exact factor models. In approximate factor models the independence assumption is replaced by an assumption that the idiosyncratic factors are diversifiable.

[2]See for instance Kaiser (1958) and Johnson and Wichern (1992) Chapter 9 for more details on orthogonal factor rotation and Jackson (2005) for oblique factor rotations.

classification of variables into groups group-pervasive factors are estimated group by group using the standard principal component method.

The procedure in Chen (2010) provides however no inference on common factors and group-specific factors, which are among main interests in studying a grouped factor model. In this paper we extend the procedure in Chen (2010) to estimate the common factors and group-specific factors. Concretely, we apply the method of multiset canonical correlation analysis to extract the common factors from the estimated group-pervasive factors resulted from the procedure in Chen (2010). Then we subtract the common component due to common factors from data and apply the principal component method to the data net the common component to obtain an estimate of the group-specific factors. This method works only when we subtract the right common component due to the common factors from the data. For this reason we develop a consistent model selection criterion to determine the number of common factors and the number of group-specific factors in each group. Our paper contributes to the literature on factor analysis in that it provides a coherent method to explore structures in an unobservable factor space. The method will determine the number of groups, endogenously classify variables into groups, determine the number of common factors and the number group-specific factors in each group, and give consistent estimates of the common factor space and the group-specific factor space.

The paper is organized as follows. In section 2 we defined a grouped factor model with explicitly formulated common factors and group-specific factors. In section 3 we present a procedure to estimate common factors and group-specific factors. Section 4 presents simulation studies on the estimation procedure. Last section concludes.

2 The Model

Let X be a $(T \times N)$ matrix collecting observations of a set of N security returns observed over T periods. We assume that this set of securities consists of n groups:

$$\underset{(T\times N)}{X} = (\underset{T\times N_1}{X_1}, \underset{T\times N_2}{X_2}, ..., \underset{T\times N_n}{X_n}), \text{ with } N = \sum_{i}^{n} N_i. \tag{2.1}$$

Further we assume that the return generating process for group i has a factor structure as follows.

$$\underset{(1\times1)}{X_{i,jt}} = \underset{(1\times r_i)}{F_{i,t}}{}' \underset{(r_i\times1)}{\lambda_{i,j}} + \underset{(1\times1)}{e_{i,jt}}, \qquad \text{for } i = 1, ..., n; j = 1, ...N_i; t = 1, ..., T; \tag{2.2}$$

where $X_{i,jt}$ is the return of the j-th security in the i-th group at time t, $F_{i,t}$ is an r_i-vector of unobservable factors in the i-th group at time t, $\lambda_{i,j}$ is an r_i-vector of factor loadings, r_i is the number of factor in the ith group and $e_{i,jt}$ is the idiosyncratic component of $X_{i,jt}$. In matrix form, we have:

$$\underset{(T\times N_i)}{X_i} = \underset{(T\times r_i)}{F_i} \underset{(r_i\times N_i)}{\Lambda_i} + \underset{(T\times N_i)}{E_i}, \qquad \text{for } i = 1, 2, ..., n, \tag{2.3}$$

where

- X_i: $(T \times N_i)$ matrix collecting observations of N_i security returns in the ith group over T periods.

- $F_i = (F_{i,1}, F_{i,2}, ..., F_{i,T})'$: $(T \times r_i)$ matrix containing r_i unobservable group-pervasive factors of the ith group over T periods.

- $\Lambda_i = (\lambda_{i,1}, \lambda_{i,2}, ..., \lambda_{i,N_i})$: $(r_i \times N_i)$ matrix of unobservable group-pervasive factor loadings of the ith group.

- E_i: $(T \times N_i)$ matrix of unobservable idiosyncratic components of the N_i security returns.

- $\sum_{i=1}^{n} N_i = N$.

In order to model common factors and group-specific factors explicitly, we assume that a small number, say r^c, of pervasive factors are common across all groups and the other pervasive factors are only pervasive within some groups. We call the former common factors and the later group-specific factors. Since factors and loadings are identified up to a full rank transformation, we can assume with loss of generality that the common factors and the group-specific factors are uncorrelated in each group. This leads to the following formal assumption.

Assumption 2.1

(a) Group-pervasive factors in each group consist of r^c common factors F_t^c and a set of r_i^s group-specific factors $F_{i,t}^s$.

$$\underset{(r_i \times 1)}{F_{i,t}} = \underset{((r^c + r_i^s) \times 1)}{\begin{pmatrix} F_t^c \\ F_{i,t}^s \end{pmatrix}}, \qquad \textit{for } i = 1, 2, ..., n,$$

with $r^c + r_i^s = r_i$.

(b) The common factors and the group-specific factors are uncorrelated.

$$E(F_t^c, F_{i,t}^{s\,\prime}) = \underset{(r^c \times r_i^s)}{\mathbf{0}}, \qquad \textit{for } i = 1, 2, ..., n. \tag{2.4}$$

We assume that all common factors have been included in F_t^c, so that under Assumption 2.1 (a) and (b) the common factor space is identified. In terms of the common factors and the group-specific factors, the grouped factor model (2.3) can be rewritten as

$$\underset{(1\times 1)}{X_{i,jt}} = \underset{(1\times r^c)}{F_t^{c\prime}} \underset{(r^c\times 1)}{\lambda_{i,j}^c} + \underset{(1\times r_i^s)}{F_{i,t}^{s\,\prime}} \underset{(r_i^s\times 1)}{\lambda_{i,j}^s} + \underset{(1\times 1)}{e_{i,jt}}, \qquad \text{for } i = 1, 2, ..., n, \tag{2.5}$$

where $\lambda_{i,j}^c$ and $\lambda_{i,j}^s$ are the loadings corresponding to the common factors and the group-specific factors with $(\lambda_{i,j}^{c\prime}, \lambda_{i,j}^{s\prime})' = \lambda_{i,j}$. In matrix form we have:

$$\underset{(T\times N_i)}{X_i} = \underset{(T\times r^c)}{F^c} \underset{(r^c\times N_i)}{\Lambda_i^c} + \underset{(T\times r_i^s)}{F_i^s} \underset{(r_i^s\times N_i)}{\Lambda_i^s} + \underset{(T\times N_i)}{E_i}, \qquad \text{for } i = 1, 2, ..., n, \tag{2.6}$$

where $F^c = (F_1^c, F_2^c, ..., F_T^c)'$, $\Lambda_i^c = (\lambda_{i,1}^c, \lambda_{i,2}^c, ..., \lambda_{i,N_i}^c)$, $F_i^s = (F_{i,1}^s, F_{i,2}^s, ..., F_{i,T}^s)'$ and $\Lambda_i^s = (\lambda_{i,1}^s, \lambda_{i,2}^s, ..., \lambda_{i,N_i}^s)$. We called model (2.6) a grouped factor model with common factors.

If group-specific factors are linearly independent across all groups, the union of the group-specific factor spaces will be r^s-dimensional with $r^s = \sum_{i=1}^n r_i^s$. Collecting all group-specific factors together, we have $G_t^s = (F_{1,t}^s{}', F_{2,t}^s{}', ..., F_{n,t}^s{}')'$ and subsequently each group-specific factor $F_{i,t}^s$ can be represented as a linear function of G_t^s. If some components of a group-specific factor are exactly linearly dependent on those of other groups, the dimension of the union of the group-specific factor spaces, say K^s, will be less than $\sum_{i=1}^n r_i^s$. In fact, the dimension of the union will be the number of all linearly independent components of the group-specific factors over all groups. Let a K^s-dimensional vector F_t^s collect all these linearly independent components of the group-specific factors of all groups, then each group-specific factor $F_{i,t}^s$ can be represented as a linear function of F_t^s. Therefore, we make the following assumption.

Assumption 2.2

(a) Group-specific factor $F_{i,t}^s$ is a linear function of a K^s dimensional factor F_t^s with $K^s \leq \sum_{i=1}^n r_i^s$ in the following way:

$$F_{i,t}^s = C_i^{s\prime} F_t^s, \qquad \text{for } i = 1, 2, ..., n. \tag{2.7}$$

where C_i^s is a $(K^s \times r_i^s)$ constant matrix.

(b) $rank(C_i^s) = r_i^s$.

(c) $rank(C_1^s, C_2^s, ..., C_n^s) = K^s$.

While we assume that the common factors and the group-specific factors are uncorrelated, we allow correlations and linear dependence among the group-specific factors across groups. This is included in Assumption 2.2 (a): If $K^s < \sum_{i=1}^n r_i^s$, group-specific factors will be linearly dependent across groups. For instance, with $n = 3$, $r^c = 2$, $r_1^s = 2$ and $r_2^s = 2$, $r_3^s = 1$ and $K^s = 3$ we are considering a grouped factor model with three groups. All three groups share two common factors; each of the three groups has two, two and one group-specific factors respectively; and the five group-specific factors are located in a three dimensional specific factor space, such that these group-specific factors must be linearly dependent across the three groups. Assumption 2.2 (b) is made to ensure that with in a group the group-specific factors are not linearly dependent, such that no components of a group-specific factor are redundant. (c) is to make sure that every component of the factor F_t^s is used in generating the group-specific factors.

In order that each group is identified, no group-specific factor space of one group can be a subspace of that of another group, in other words $F_{i,t}^s$ must not be a linear function of $F_{j,t}^s$: formally $F_{i,t}^s \neq C' F_{j,t}^s$ for any constant matrix C. Because $F_{i,t}^s = C_i^{s\prime} F_t^s$ and $F_{j,t}^s = C_j^{s\prime} F_t^s$, we will require that $C_i^s \neq C_j^s C$ for any constant matrix C. This excludes in particular the possibility that $F_{i,t}^s$ is just a rotation of $F_{j,t}^s$. Further, since groups are characterized through their data points, we want that a data point belongs unambiguously to only one group. Because a data point consists of a systematical common component and an idiosyncratic random component, we require that the common components of data from different groups must be different[3]. This leads to the following assumption.

Assumption 2.3

(a) C_i^s *and* C_j^s *are not linearly dependent, i.e.* $C_i^s \neq C_j^s C$*, for any constant* C *with* $i \neq j$*,* $i = 1, 2, ..., n$ *and* $j = 1, 2, ..., n$*.*

(b) Any pair of loading vectors of different groups $\lambda_{i,m}^s$ *and* $\lambda_{j,l}^s$ *for* $m = 1, 2, ...N_i$*,* $l = 1, 2, ..., N_j$*,* $i = 1, 2, ..., n$*,* $j = 1, 2, ..., n$ *and* $i \neq j$ *satisfy the restriction:*

$C_i^s \lambda_{i,m}^s \neq C_j^s \lambda_{j,l}^s$.

Assumption 2.3 (a) says no group-specific factors are linear combinations of those of another group. In the case with two group-specific factor planes and one group-specific factor line mentioned earlier, this assumption excludes the situation in which the line lies on either of the two planes and the situation where one plane lies on the other, so that the three group-specific factor spaces are different. Assumption 2.3 (b) excludes the situation in which a data point is located around the intersection of the factor spaces of two groups[4], such that the relationship between variables and groups is unambiguously defined.

The common factors F_t^c and the specific factors F_t^s constitute an $(r^c + K^s)$ dimensional overall pervasive factor space. We used an $(r^c + K^s)$ dimensional vector $F_t^p = \begin{pmatrix} F_t^c \\ F_t^s \end{pmatrix}$ to represent the overall pervasive factors[5]. Then, following Assumption 2.2, each group-pervasive factor $F_{i,t}$ can be written as a linear function of F_t^p.

$$F_{i,t} = \begin{pmatrix} F_t^c \\ F_{i,t}^s \end{pmatrix} = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & C_i^{s\prime} \end{pmatrix} \begin{pmatrix} F_t^c \\ F_t^s \end{pmatrix} = C_i' F_t^p, \qquad \text{for } i = 1, 2, ..., n, t = 1, 2, ..., T, \tag{2.8}$$

with $C_i' = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & C_i^{s\prime} \end{pmatrix}$. Denoting $F^p = (F_1^p, F_2^p, ..., F_T^p)'$, we can present equation (2.8) in a matrix form: $F_i = F^p C_i$.

Under Assumptions 2.1 to 2.3, the polled security returns X adopts a factor structure with F^p as the factors:

$$\begin{aligned} X &= \begin{pmatrix} X_1 & X_2 & \dots & X_n \end{pmatrix} \\ &= \begin{pmatrix} F_1\Lambda_1 & F_2\Lambda_2 & \dots & F_n\Lambda_n \end{pmatrix} + \begin{pmatrix} E_1 & E_2 & \dots & E_n \end{pmatrix} \\ &= \begin{pmatrix} F^pC_1\Lambda_1 & F^pC_2\Lambda_2 & \dots & F^pC_n\Lambda_n \end{pmatrix} + \begin{pmatrix} E_1 & E_2 & \dots & E_n \end{pmatrix} \\ &= F^p \begin{pmatrix} C_1\Lambda_1 & C_2\Lambda_2 & \dots & C_n\Lambda_n \end{pmatrix} + \begin{pmatrix} E_1 & E_2 & \dots & E_n \end{pmatrix} \end{aligned}$$

[3]This is a technical assumption to simplify the presentation of a correct classification. Without this assumption we may have situations in which some data points may belong to more that one group. which will complicate a definition of correct classification.

[4]This is a technical assumption to simplify the presentation of a correct classification of variables. However, this assumption is not essential for estimation of the structured factor space. See Chen (2010) for more details.

[5]Note that Assumption 2.1 (b) allows us a simple presentation of the overall pervasive factor as combination of $F_t^p = (F_t^{c'}, F_t^{s'})'$. Without this assumption F_t^p will be a basis of the union space of F_t^c and F_t^s.

Defining $\Lambda = (C_1\Lambda_1, C_2\Lambda_2, ..., C_n\Lambda_n)$ and $E = (E_1, E_2, ..., E_n)$, we have:

$$\underset{(T\times N)}{X} = \underset{(T\times K)}{F^p}\underset{(K\times N)}{\Lambda} + \underset{(T\times N)}{E} \tag{2.9}$$

The equation above says that X can be accommodated in an ungrouped factor model with $K = r^c + K^s$ factors.

Benefits of studying the grouped factor model (2.6) instead of the pooled ungrouped factor model (2.9) are that the grouped factor model provides a more detailed information on the data as well as the data-generating process. In stead of saying that all variables are influenced by K factors, we can say the variables consist of n groups and there are r^c common factors that influence all variables. In addition variables in each group is influenced by additional k_i^s group-specific factors. This detailed group-specific information can be used in association with possible group-specific structural information to provide a better interpretation of the factors.

In order to apply the estimation procedure given in Chen (2010) we adopt the model assumptions on factors and factor loading from Bai and Ng (2002), which is also applied in Chen (2010).

Assumption 2.4
$E||F_t^p||^4 < \infty$ and $E||\frac{1}{\sqrt{T}}\sum_{t=1}^T(F_t^pF_t^{p\prime} - E(F_t^pF_t^{p\prime}))|| < M$ as $T \to \infty$. Denote the positive definite matrix $E(F_t^pF_t^{p\prime})$ by Σ^p.

Assumption 2.5
$||\lambda_{i,j}|| < \lambda < \infty$ and $||\Lambda_i\Lambda_i'/N_i - L_i|| \to 0$ as $N_i \to \infty$ for some $(r_i \times r_i)$ positive definite matrix L_i, for $i = 1, 2, ..., n$.

Let X_{jt} denote the observation of the jth variable at time t in X and e_{jt} be the idiosyncratic component of X_{jt} in the ungrouped model (2.9).

Assumption 2.6 (Time and Cross-Section Dependence and Heteroskedasticity)
There exists a positive constant $M \leq \infty$, such that for all N and T,
1. $E(e_{jt}) = 0$, $E|e_{jt}|^8 \leq M$;
2. $E(\sum_{j=1}^N e_{js}e_{jt}/N) = \gamma_N(s,t), |\gamma_N(s,s)| \leq M$ for all s, and
$T^{-1}\sum_{t=1}^T\sum_{s=1}^T|\gamma_N(s,t)| \leq M$;
3. $E(e_{jt}e_{kt}) = \tau_{jk,t}$ with $|\tau_{jk,t}| \leq |\tau_{ij}|$ for some τ_{jk} and for all t, $N^{-1}\sum_{j=1}^N\sum_{k=1}^N|\tau_{jk}| < M$
4. $E(e_{jt}e_{ks}) = \tau_{jk,ts}$ and $(NT)^{-1}\sum_{j=1}^N\sum_{k=1}^N\sum_{t=1}^T\sum_{s=1}^T|\tau_{jk,ts}| \leq M$,
5. for every (t,s), $E|N^{-1/2}\sum_{i=1}^N[e_{js}e_{jt} - E(e_{js}e_{jt})]|^4 \leq M$.

Further we adopt also the assumption on weak dependence between factors and errors given in Bai and Ng (2002).

Assumption 2.7 (Weak Dependence between Factors and Errors)

$$E\left(\frac{1}{N}\sum_{j=1}^N\left|\left|\frac{1}{\sqrt{T}}F_t^pe_{jt}\right|\right|^2\right) \leq M.$$

Assumption 2.4 is to a certain degree a strong assumption in a factor model. A standard assumption such as in Bai and Ng (2002) requires only law of large number type convergency: $\frac{1}{T}\sum_{t=1}^{T} F_t^p F_t^{p\prime} \xrightarrow{P} \Sigma^p$. We require instead a stronger condition $E||\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(F_t^p F_t^{p\prime} - \Sigma^p)|| < M$. However, for practical application purposes, these two kinds of assumptions make no essential difference.

Under Assumption 2.2 and Assumption 2.4 it is easy to see that the group-pervasive factor $F_{i,t}$ also satisfies the requirements of Assumption 2.4, i.e.

(1) $E||F_{i,t}||^4 = E||C_i' F_t^p||^4 < \infty$

(2) $\frac{1}{T}\sum_{t=1}^{T} F_{i,t} F_{i,t}{}' = \frac{1}{T}\sum_{t=1}^{T} C_i F_t^p F_t^{p\prime} C_i' \xrightarrow{P} C_i \Sigma^p C_i'$ as $T \to \infty$. Since $rank(C_i) = r_i = r^c + r_i^s$, $C_i \Sigma^p C_i'$ is a positive definite matrix.

Assumption 2.5 is to make sure that each component of a group-pervasive factor makes a nontrivial contribution to the variance of the variables in the group.

Proposition 2.8
Under Assumption 2.5 and Assumption 2.2 (d) and (e), the factor loading matrix Λ in the ungrouped model (2.9) satisfies the requirement in Assumption 2.5, i.e. $||\lambda_j|| < \lambda < \infty$ and $||\Lambda\Lambda'/N - L|| \to 0$ as $N \to \infty$ for some $(K \times K)$ positive definite matrix L.

Proposition 2.9
(a) $rank(C_i) = r_i = r^c + r_i^s$.
(b) $rank(C_1, C_2, ..., C_n) = K$.
(c) $K \leq r^c + \sum_{i=1}^{n} r_i^s$.

This proposition follows the assumption that the common factors and the group-specific factors are uncorrelated. Comparing model Assumptions 2.1 through 2.7 with the model assumptions given in Chen (2010), we know that our group factor models with common factors here satisfy all the assumptions on grouped factor models given in Chen (2010). Therefore, group factor models with common factors belong to a special class of grouped factor models with explicitly defined common factors and group-specific factors.

3 Estimation of Factors

In studying grouped factor models with common factors given in (2.6), instead of assuming that the number of groups, the membership relation between securities and their respective groups are known *a priori*, we want to estimate them from observed data. In other words, determination of the number of groups, classification of the securities into their respective groups, and estimation of the common factors and the group-specific factors are objectives in this paper.

3.1 Estimation of Group-pervasive Factors

Since grouped factor models with common factors are a special class of grouped factor models, we can apply the estimation procedure given in Chen (2010) to determine the number of groups, classify the variables in to groups, determine the number of the group-pervasive factors and estimate group-pervasive factors for each group. We restate the estimation procedure briefly as follows.

- Step 1: Estimate the dimension of the overall pervasive factor space K by the PC criterion of Bai and Ng (2002).
- Step 2: Project the $(T \times N)$ data matrix X onto a $(K \times N)$ matrix:

$$\bar{X}^T = \frac{1}{T}\hat{F}^{K\prime}X,$$

  where $\hat{F}^K$ is a principal component estimator of F^p with $\hat{F}^K = \frac{1}{NT}(XX')\sqrt{T}Q$, where Q is a $(T \times K)$ matrix containing K eigenvectors corresponding to the K largest eigenvalues of XX'.
- Step 3: According to a set of chosen model parameters $(n, \{k_i\}_{i=1}^n)$, solve the classification problem for the projected data by polynomial differentiation algorithm with voting scheme
- Step 4: Use the model selection criterion to evaluate alternative choices of models to obtain an optimal model $(\hat{n}, \{\hat{k}_i\}_{i=1}^{\hat{n}})$ and the corresponding classification of variables $\{X_i^{s_n}\}_{i=1}^{\hat{n}}$.($X_i^{s_n}$ represents the securities classified into the ith group.)
- Step 5: Estimate a factor model for each group of data in $\{X_i^{s_n}\}_{i=1}^{\hat{n}}$ by the principal component method to obtain estimates for the respective group-pervasive factors $\hat{F}_i = \frac{1}{N_iT}(X_i^{s_n}X_i^{s_n\prime})\sqrt{T}Q_i$, where Q_i contains the $\hat{k}_i$ eigenvectors corresponding to the $\hat{k}_i$ largest eigenvalues of the matrix $X_i^{s_n}X_i^{s_n\prime}$.

It is shown that this procedure will achieve a consistent classification of the securities into their respective groups. The procedure gives also consistent estimates of group-pervasive factor space for each group.

What we want particularly to focus on in this paper is to estimate the common factors and the group-specific factors in a grouped factor model. A question raises naturally: can we directly derive estimates for the common factors and the group-specific factors from the estimates of the group-pervasive factors? In some circumstances the answer is positive. Let's denote a grouped factor model by the numbers of factors in each group and the dimension of the overall pervasive factor space: $[r_1, r_2, ..., r_n|K]$. For example [2 2 | 3] indicates a grouped factor model with two groups. Each group has two factors and the two pairs of factors are located in a 3 dimensional overall pervasive factor space. For the model [2 2 | 3], using the procedure described above, we will obtain two pairs of group-pervasive factors. Then the first canonical correlation coefficient between these two pairs will be close to one and the canonical variable can be taken as an estimate of the common factor. However, for a model [2 2 2|4] the situation is more complicated. For each pair of the group-pervasive factors we will have a canonical variable. Altogether we have three canonical variables for three different pairs. If the data-generating model does have one common factor for the three groups, then the three canonical variables will be highly correlated. Each one of them can be a good estimate of the common factor and any linear combinations of them are also good estimates of the common factor. But, it may happen that the model [2 2 2|4] does not have any common factor at all.

For instance the factors in the three groups can be: $([F_1F_2][F_2F_3][F_3F_4])$ in which the first two groups share one factor F_2 and group 2 and group 3 share one factor F_3 and the three groups do not have any common factors. In this case none of the three canonical variables can be used as an estimate of the common factor.

Generally the information on the group-pervasive factors estimated group by group are not directly conclusive for the common factor space and the group-specific factor spaces. We need a more detailed study in order to obtain an estimate for the common factors. In the next subsection we will present a procedure to estimate the common factors and the group-specific factors as well.

3.2 Estimation of Common Factors and Group-Specific Factors

Multiset canonical correlation analysis is extends the canonical correlation analysis between two groups to more groups[6]. It targets at finding linear combinations in each group, such that the correlations among these linear combinations are maximized across all groups. Since F_t^c are the common factors among all group-pervasive factors $F_{i,t}$ for $i = 1, 2, ...n$, the first r^c multiset canonical variables across all group-pervasive factors must be the common factors or linear combinations of the common factors.

Let Σ_{ij} denote the covariance matrix between the group-pervasive factors of the ith group and those of the jth group: $\Sigma_{ij} = E(F_{i,t}F'_{j,t})$, the calculation of the multiset canonical correlation coefficients is to solve the following maximization problem:

$$\max_{a_i,a_j} \text{ tr } \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (a'_i \Sigma_{ij} a_j) \right) = \max_{a_i,a_j} \text{ tr } \left(\sum_{i=1}^{n} \sum_{j=1}^{n} E\left(a'_i F_{i,t}, a'_j F_{j,t}\right) \right) \tag{3.10}$$

$$\text{s.t. } a'_i \Sigma_{ii} a_i = I_{r^c}. \qquad \text{for } i = 1, 2, ...n. \tag{3.11}$$

where a_i and a_j are $(r_i \times r^c)$ and $(r_j \times r^c)$ matrices. $a'_i F_{i,t}$ and $a'_j F_{j,t}$ represent r^c linear combinations of the pervasive factors in group i and group j respectively. The motivation of this maximization problem explains itself. The objective function is the sum of pairwise canonical correlation coefficients over all groups. The restrictions in (3.11) is to make sure that the summands in the objective function are properly normalized to be canonical correlation coefficients.

Collecting $F_{i,t}$, a_i, and Σ_{ij} over all groups into $(\sum_{i=1}^n r_i \times 1)$, $(\sum_{i=1}^n r_i \times r^c)$ and $(\sum_{i=1}^n r_i \times \sum_{i=1}^n r_i)$ matrices respectively, we obtain:
$F_t = \begin{pmatrix} F_{1,t} \\ \vdots \\ F_{n,t} \end{pmatrix}$, $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, and $\Sigma = \begin{pmatrix} \Sigma_{11} & \dots & \Sigma_{11} \\ \vdots & \ddots & \vdots \\ \Sigma_{n1} & \dots & \Sigma_{nn} \end{pmatrix}$, so that the maximization problem (3.10) can be written in the following matrix form:

$$\max_{\mathbf{a}} \text{ tr } (\mathbf{a}'\Sigma\mathbf{a}) \tag{3.12}$$

$$\text{s.t. } \mathbf{a} = (a'_1, a'_2, ...a'_n)', \text{ and } a'_i \Sigma_{ii} a_i = I_{r^c}. \qquad \text{for } i = 1, 2, ...n. \tag{3.13}$$

where $\mathbf{a}$ is a $((\sum_{i=1}^n r_i) \times r^c)$ matrix.

[6]See Nielsen (2002) and Hasan (June 2009) for more details on multiset canonical correlation analysis.

Since all group-pervasive factors share r^c common factors, the common factors must be the canonical variables, i.e. we can calculate the common factors as follows: $F_t^c = \frac{1}{n}\sum_{i=1}^n a_i' F_{it} = \frac{1}{n}\mathbf{a}' F_t$, where $\mathbf{a}$ is a solution of the maximization problem (3.12).

In order to overcome the difficulties of local maxima and to make the computation more efficiently, we reformulate the nonlinear optimization problem (3.12) under restriction in (3.13) into an eigenvalue problem through standardizing the group-pervasive factors. Defining $F_{it}^* = \Sigma_{ii}^{-\frac{1}{2}} F_{i,t}$, we have $\Sigma_{ij}^* = E\left(F_{it}^* F_{ji}^{*'}\right) = E\left(\Sigma_{ii}^{-\frac{1}{2}} F_{it} F_{ji}' \Sigma_{jj}^{-\frac{1}{2}}\right) = \Sigma_{ii}^{-\frac{1}{2}} \Sigma_{ij} \Sigma_{jj}^{-\frac{1}{2}}$, and $\Sigma_{ii}^* = Var(F_{it}^*) = I_{k_i}$. Stacking $F_{i,t}^*$ over all groups, we have $F_t^{*'} = (F_{1,t}^{*'}, F_{2,t}^{*'}, ..., F_{n,t}^{*'})$ and $\Sigma^* = E(F_t^* F_t^{*'})$. Then the maximization problem (3.12) can be reformulated as follows:

$$\max_{\mathbf{a}} \ \text{tr}\ (\mathbf{a}'\Sigma^*\mathbf{a}) \tag{3.14}$$

$$\text{s.t. } \mathbf{a}' = (a_1, a_2, ...a_n)', \text{ and } a_i' a_i = I_{r^c}. \qquad \text{for } i = 1, 2, ...n. \tag{3.15}$$

Being multiset canonical variables, the standardized common factors can be calculated as follows: $F_t^{c*} = \frac{1}{n}\sum_{i=1}^n a_i' F_{it}^* = \frac{1}{n}\mathbf{a}' F_t^*$, where $\mathbf{a}$ is the solution of the maximization problem (3.14). The maximization problem (3.14) differs from the following problem:

$$\max_{a} \ \text{tr}\ (\mathbf{a}'\Sigma^*\mathbf{a}) \tag{3.16}$$

$$\text{s.t. } \mathbf{a}'\mathbf{a} = nI_{r^c}, \tag{3.17}$$

only in that the n restrictions in problem (3.15) are replaced by one single restriction on the sum of the n restrictions: $\mathbf{a}'\mathbf{a} = \sum_{i=1}^n a_i' a_i = nI_{r^c}$ in (3.17). It is well known that the maximization problem (3.16) under (3.17) is an eigenvalue problem (See Johansen and Wichern (1992) p. 459): the solution are the eigenvectors of length $\sqrt{n}$ corresponding to the r^c largest eigenvalues. Because the maximization problem in (3.16) relaxes the restrictions in the maximization problem in (3.14), generally the two problems will have different solutions. However, in our case the Σ^* matrix has a particular structure, such that it can be shown that these two problems have identical solutions. Therefore we can solve the problem (3.14) via solving the eigenvalue problem (3.16). We summarize this fact in the following proposition.

Proposition 3.1
Under Assumptions 2.1 to 2.4

(i) Σ^* *has* K *nonzero eigenvalues.*

(ii) The first r^c *largest eigenvalues of* Σ^* *are identical and their value is* n. *The length-*$\sqrt{n}$ *eigenvectors that correspond to the* r^c *largest eigenvalues of* Σ^* *solve the maximization problem (3.16) and they solve the maximization problem (3.14) as well.*

(iii) There exists a particular set of eigenvectors a_{r^c} *(see below) corresponding to the first* r^c *largest eigenvalues of* Σ^**, such that* $F_t^{c*} = \frac{1}{\sqrt{n}} a_{r^c}' F_t^*$.

$$a_{r^c}' = \frac{1}{\sqrt{n}}(\underset{(r^c\times r^c)}{I_{r^c}}, \underset{(r^c\times r_1^s)}{\mathbf{0}}, \underset{(r^c\times r^c)}{I_{r^c}}, \underset{(r^c\times r_2^s)}{\mathbf{0}}, ..., \underset{(r^c\times r^c)}{I_{r^c}}, \underset{(r^c\times r_n^s)}{\mathbf{0}}).$$

(iv) The r^c+1 to k^c ($r^c < k^c \leq K$) largest eigenvalues of Σ^* are nonzero and the corresponding eigenvectors denoted by a^{ps} has the property that $\frac{1}{\sqrt{n}} a^{ps\prime} F_t^*$ spans a subspace of the standardized specific factor space and hence also a subspace of the specific factor space. We denote $\frac{1}{\sqrt{n}} a^{ps\prime} F_t^*$ by F_t^{ps*}.

$$a^{ps\prime} = (\underset{(kr^c \times r^c)}{\mathbf{0}_{r^c}}, \underset{(kr^c \times r_1^s)}{a_1^{ps\prime}}, \underset{(kr^c \times r^c)}{\mathbf{0}_{r^c}}, \underset{(kr^c \times r_2^s)}{a_2^{ps\prime}}, ..., \underset{(kr^c \times r^c)}{\mathbf{0}_{r^c}}, \underset{(kr^c \times r_n^s)}{a_n^{ps\prime}}),$$

with $kr^c = k^c - r^c$.

Comments:

Because the first r^c largest eigenvalues of Σ^* are identical, obviously, any orthogonal transformations of a_{r^c} are still eigenvectors corresponding to the same eigenvalues. The r^c+1 to k^c ($r^c < k^c \leq K$) largest eigenvalues of Σ^* are positive. Whether these eigenvalues are unique or not depends on data generating models, in particular, it depends on the relationship between the group-specific factors across groups. Because our main focus is not on the relations among the group-specific factors, we make an auxiliary assumption on the uniqueness of the r^c+1 to k^c ($r^c < k^c \leq K$) largest eigenvalues of Σ^*, in order to simplify the presentation.

Assumption 3.2
The r^c+1 to k^c ($r^c < k^c \leq K$) largest eigenvalues of Σ^* are unique.

Under this assumption the corresponding eigenvectors a^{ps} are unique if we require that the first non-zero element in each column of a^{ps} is positive.

Proposition 3.1 establishes that we can calculate the standardized common factors and henceforth an estimate of the space of the common factors using the eigenvectors of Σ^*.

$$F_t^{c*} = \frac{1}{\sqrt{n}} a_{r^c}' F_t^*. \tag{3.18}$$

Still, we cannot use (3.18) to estimate F_t^{c*} directly, because we don't know F_t^*. However, through estimation of group-pervasive factors we have an estimate $\hat{F}_{i,t}^*$ for $F_{i,t}^*$ and an estimate $\hat{\Sigma}_{ij}^*$ for Σ_{ij}^* with $\hat{\Sigma}_{ij}^* = \frac{1}{T}\sum_{t=1}^T \hat{F}_{i,t}^* \hat{F}_{j,t}^{*\prime}$. Stacking $\hat{F}_{i,t}^*$ together over all groups, we have $\hat{F}_t^* = (\hat{F}_{1,t}^{*\prime}, \hat{F}_{2,t}^{*\prime}, ..., \hat{F}_{n,t}^{*\prime})'$ and $\hat{\Sigma}^* = \frac{1}{T}\sum_{t=1}^T \hat{F}_t^* \hat{F}_t^{*\prime}$.

Using $\hat{\Sigma}^*$ instead of Σ^*, we can estimate the standardized common factors and hence the space of the common factors, by solving the following maximization problem:

$$\max_{\mathbf{a}} \ \text{tr} \ (\mathbf{a}'\hat{\Sigma}^*\mathbf{a}) \tag{3.19}$$

$$\text{s.t.} \ \mathbf{a}'\mathbf{a} = I_{r^c}. \tag{3.20}$$

Obviously, the solutions of the maximization problem (3.19) are the unit length eigenvectors of $\hat{\Sigma}^*$ that correspond to the r^c largest eigenvalues of $\hat{\Sigma}^*$. Denoting the solution of the maximization problem in (3.19) by $\hat{\mathbf{a}}$, the rescaled canonical variable $\frac{1}{\sqrt{n}}\hat{\mathbf{a}}'\hat{F}_t^*$ serves as an estimate of the standardized common factors. If $\hat{\mathbf{a}}$ and $\hat{F}_t^*$ are consistent estimates of a_{r^c} and F_t^* respectively, $\hat{F}_t^c = \frac{1}{\sqrt{n}}\hat{\mathbf{a}}'\hat{F}_t^*$ will be a consistent estimate of the common factor space. Our discussion sofar depicts a procedure to estimate the common factor space. However, this procedure is based on an assumption that we know the number of the common factors r^c. If r^c needs to be determined, which is one of our objectives in this paper, we can only make a guess k^c for r^c. If $k^c < r^c$ the depicted procedure will lead to an estimate of only a subset of the common factor space; and if $k^c > r^c$ the procedure will lead to an estimate of a factor space containing the common factor space and a subspace of the specific factor space. We summarize this fact in the following theorem.

Theorem 3.3
Let k^c be a guess of the number of the common factors with $0 < k^c < r_i$ and let the $(\sum_{i=1}^n r_i) \times k^c$ matrix $\hat{h}_{k^c}$ be eigenvectors corresponding to the k^c largest eigenvalues of $\hat{\Sigma}^*$ and $\hat{F}_t^c = \frac{1}{\sqrt{n}} \hat{h}'_{k^c} \hat{F}_t^*$. Under Assumptions 2.1 to 2.7, for $1 \leq k^c \leq r^c$, there exists an $(r^c \times k^c)$ matrix B_{rk^c} with $rank(B_{rk^c}) = k^c$, such that

$$\frac{1}{T} \sum_{t=1}^{T} ||\hat{F}_t^c - B'_{rk^c} F_t^c||^2 = O_p(C_{N,T}^{-2}), \tag{3.21}$$

with $C_{N,T} = \min(\sqrt{T}, \sqrt{N})$; and for $r^c < k^c \leq K$, there exists a $(k^c \times k^c)$ full rank matrix B_{k^c}, such that

$$\frac{1}{T} \sum_{t=1}^{T} \left|\left| \hat{F}_t^c - B'_{k^c} \left(\begin{array}{c} F_t^c \\ F_t^{ps} \end{array} \right) \right|\right|^2 = O_p(C_{N,T}^{-2}), \tag{3.22}$$

where $F_t^{ps} = \frac{1}{\sqrt{n}} a^{ps'} F_t^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n a_i^{ps'} \Sigma_i^{s-\frac{1}{2}} F_{i,t}^s$ representing some $(k^c - r^c)$ linear combinations of the group-specific factors of all groups and a^{ps} is the matrix containing eigenvectors corresponding to the r^c+1 to k^c largest eigenvalues of Σ^* defined in Proposition 3.1 (iv).

Theorem 3.3 implies, in particular, that if we know the number of common factors, we can consistently estimate the space spanned by the common factors in the sense that the time average of the squared deviations between the estimated common factors and those lie in the true common factor space vanish as $N, T \to \infty$. Because the common factors F_t^c can only be identified up to a rotation, we can only estimate the space spanned by the common factors.

Now we turn to estimation of the group-specific factors $F_{i,t}^s$. We wish to have observations on $X_i^s = X_i - F^c \Lambda_i^c = F_i^s \Lambda_i^s + E_i$, because we could derive an estimate for the group-specific factors from X_i^s that were generated only by the group-specific factors. X_i^s is unfortunately unobservable. A natural estimate for X_i^s is the residuals of a linear regression of X_i on the estimate of the common factors $\hat{F}_t^c$:

$$X_{i,jt} = \hat{F}_t^{c'} \hat{\lambda}_{i,j}^c + \hat{X}_{i,jt}^{rs}, \qquad i = 1, 2, ...n; j = 1, 2, ..., N_i, \tag{3.23}$$

where $\hat{\lambda}_{i,j}^c$ is the regression coefficient and $\hat{X}_{i,jt}^{rs}$ is the regression residual. In matrix form we have

$$\hat{X}_i^{rs} = X_i - \hat{F}^{c'} \hat{\Lambda}_i^c, \qquad i = 1, 2, ...n. \tag{3.24}$$

The equation above says $\hat{X}_i^{rs}$ can be seen as data net the common components due to the common factors, i.e. they are generated only by the group-specific factors and idiosyncratic components. So we can apply the principal component method to the data of $\hat{X}_i^{rs}$ to obtain estimates of the group-specific factors group by group.

The strategy above works only when our guess k^c is correct, i.e. $k^c = r^c$. Because r^c is unknown, our guess k^c may differ from r^c. For a choice of k^c we may potentially have three cases: (1) $k^c < r^c$, (2) $k^c = r^c$ and (3) $k^c > r^c$. According to Theorem 3.3, for these three cases, $\hat{F}_t^{c'}$ may span different subsets of the pervasive factor space and hence $\hat{F}_t^{c'}\hat{\lambda}_{i,j}^c$ may contain different parts of the common component. Consequently, $\hat{X}_{i,jt}^{rs} = X_{i,jt} - \hat{F}_t^{c'}\hat{\lambda}_{i,j}^c$ may have different influencing factors. For $k^c < r^c$, $\hat{X}_{i,jt}^{rs} = X_{i,jt} - \hat{F}_t^{c'}\hat{\lambda}_{i,j}^c$ will contain more factors than $F_{i,jt}^s$, while for $k^c > r^c$, $\hat{X}_{i,jt}^{rs} = X_{i,jt} - \hat{F}_t^{c'}\hat{\lambda}_{i,j}^c$ may contain fewer factors. Indeed, the number of the factors that influence $\hat{X}_{i,jt}^{rs}$ depends on the choice of k^c. The following proposition states this fact formally.

Proposition 3.4

For a given choice of $0 < k^c < r_i$, to the common factor estimate $\hat{F}_t^c$ based on Theorem 3.3 and the regression residuals $\hat{X}_{i,jt}^{rs}$ based on equation (3.23), there exit population counterparts F_t^{pc} and $X_{i,jt}^{rs}$, respectively, such that

(i) $X_{i,jt}$ can be decomposed as:

$$X_{i,jt} = F_t^{pc'}\lambda_{i,j}^{pc} + X_{i,jt}^{rs}$$

and $X_{i,jt}^{rs}$ is generated by a factor model.

$$X_{i,jt}^{rs} = F_{i,t}^{rs'}\lambda_{i,j}^{rs} + e_{i,jt}.$$

(ii) Both F_t^{pc} and the number of factors generating $X_{i,jt}^{rs}$, denoted by k_i^{rs}, vary with the choice of k^c as follows:

- *for $k^c < r^c$, $F_t^{pc} = B'_{rk^c}F_t^c$, with B_{rk^c} an $(r^c \times k^c)$ matrix defined in Theorem 3.3 and $k_i^{rs} = r^c - k^c + r_i^s$.*
- *for $k^c = r^c$, $F_t^{pc} = B'_{r^c}F_t^c$, $X_{i,jt}^{rs} = X_{i,jt}^s = F_{i,t}^{s'}\lambda_{i,j}^s + e_{i,jt}$ and $k_i^{rs} = r_i^s$.*
- *For $k^c > r^c$, $F_t^{pc} = B'_{k^c}\begin{pmatrix} F_t^c \\ F_t^{ps} \end{pmatrix}$, where B_{k^c} and F_t^{ps} are a $(k^c \times k^c)$ matrix and a $(k^c - r^c)$ random vector defined in Theorem 3.3 respectively, and*
  - * *$k_i^{rs} = r_i^s$, if there is no exact linear dependence between $F_{i,t}^s$ and F_t^{pc}.*
  - * *$k_i^{rs} = r_i^s - k_i^{s*}$, if there exits k_i^{s*} linearly dependent relations between $F_{i,t}^s$ and F_t^{ps}: $F_{i,t}^{s}{}'C = F_t^{ps'}B$, where C is an $(r_i^s \times k_i^{s*})$ matrix with $k_i^{s*} \leq r_i^s$ and $rank(C) = k_i^{s*}$.*

(iii) Let $\tilde{F}_i^s$ and $\hat{F}_i^s$ be principal component estimates of factors based on the data of $(X_i^{rs}X_i^{rs'})$ and on the data of $(\hat{X}_i^{rs}\hat{X}_i^{rs'})$, respectively: $\tilde{F}_i^s = \frac{1}{N_iT}(X_i^{rs}X_i^{rs'})\sqrt{T}\tilde{Q}_{k_i^s}$ with $\tilde{Q}_{k_i^s}$ a $(T \times k_i^s)$ matrix of the eigenvectors corresponding to the k_i^s largest eigenvalues of $(X_i^{rs}X_i^{rs'})$ and $\hat{F}_i^s = \frac{1}{N_iT}(\hat{X}_i^{rs}\hat{X}_i^{rs'})\sqrt{T}\hat{Q}_{k_i^s}$ with $\hat{Q}_{k_i^s}$ is a $(T \times k_i^s)$ matrix of the eigenvectors that correspond to the k_i^s largest eigenvalues of $(\hat{X}_i^{rs}\hat{X}_i^{rs'})$.

Then we have

$$\frac{1}{T}\sum_{t=1}^{T}||\tilde{F}^s_{i,t} - \hat{F}^s_{i,t}||^2 = O_p(C^{-2}_{N,T}).$$

Proposition 3.4 (i) and (ii) imply in particular that for a correct guess of k^c, $X^s_{i,t}$ is generated only by the group-specific factors $F^s_{i,t}$, so that we could obtain an estimate for the group-specific factors from $\frac{X^s_i X^{s\prime}_i}{N_i T}$. (iii) states that the factor estimate based on $\frac{\hat{X}^{rs}_i \hat{X}^{rs\prime}_i}{N_i T}$ converges to the factor estimate based on $\frac{X^{rs}_i X^{rs\prime}_i}{N_i T}$. Therefore, we can use the available $\frac{\hat{X}^{rs}_i \hat{X}^{rs\prime}_i}{N_i T}$ instead of the unobservable $\frac{X^{rs}_i X^{rs\prime}_i}{N_i T}$ to construct an estimate for the group-specific factors. The following theorem states properties of the estimation based on $\frac{\hat{X}^{rs}_i \hat{X}^{rs\prime}_i}{N_i T}$.

Theorem 3.5
For a given choice of (k^c, k^s_i) satisfying $0 < k^c < r_i$ and $1 < k^s_i$, let $\hat{F}^c_t$ be the estimate of the common factors given in Theorem 3.3 and $\hat{X}^{rs}_i$ be residuals from the regression in (3.24) and $\hat{F}^s_i$ be the estimate of k^s_i group-specific factors obtained by applying the asymptotical principal component method to the data set of $\hat{X}^{rs}_i$: $\hat{F}^s_i = \frac{\hat{X}^{rs}_i \hat{X}^{rs\prime}_i}{N_i T}\sqrt{T}\hat{Q}_{k^s_i}$, where $\hat{Q}_{k^s_i}$ is the eigenvectors corresponding to the k^s_i largest eigenvalues of $\frac{\hat{X}^{rs}_i \hat{X}^{rs\prime}_i}{N_i T}$. Under assumption 2.1 through 2.7, there exit a $(k^c \times k^c)$ full rank matrix $\mathcal{H}_{k^c}$ and a $(k^s_i \times k^{rs}_i)$ full rank matrix $\mathcal{H}'_{k^s_i}$ with rank$(\mathcal{H}_{k^s_i}) = \min(k^s_i, k^{rs}_i)$, such that

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t||^2 = O_p(C^{-2}_{N,T}), \tag{3.25}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}^s_{i,t} - \mathcal{H}'_{k^s_i}F^{rs}_{i,t}||^2 = O_p(C^{-2}_{N,T}). \tag{3.26}$$

Combining (3.25) with (3.26), we have

$$\frac{1}{T}\sum_{t=1}^{T}\left|\left|\begin{pmatrix}\hat{F}^c_t\\ \hat{F}^s_{i,t}\end{pmatrix} - \begin{pmatrix}\mathcal{H}'_{k^c} & 0\\ 0 & \mathcal{H}'_{k^s_i}\end{pmatrix}\begin{pmatrix}F^{pc}_t\\ F^{rs}_{i,t}\end{pmatrix}\right|\right|^2 = O_p(C^{-2}_{N,T}). \tag{3.27}$$

Denoting $(\hat{F}^{c\prime}_t, \hat{F}^{s\prime}_t)'$ by $\hat{F}^0_t$, $(F^{pc\prime}_t, F^{rs\prime}_t)'$ by $\hat{F}^0_t$ and diag$(\mathcal{H}'_{k^c}, \mathcal{H}'_{k^s_i})$ by $\mathcal{H}^{0\prime}$ we can rewrite the equation above compactly as follows.

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}^0_t - \mathcal{H}^{0\prime}F^0_t|| = O_p(C^{-2}_{N,T}). \tag{3.28}$$

Corollary 3.6
For $k^c = r^c$ and $k^s_i \geq 1$, there exists an $(r^s_i \times k^s_i)$ matrix $\mathcal{H}_{k^s_i}$ with rank$(\mathcal{H}_{k^s_i}) =$ $\min(r^s_i, k^s_i)$[7], such that

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}^s_{i,t} - \mathcal{H}'_{k^s_i}F^s_{i,t}||^2 = O_p(C^{-2}_{N,T}),\ \textit{for } i = 1, 2, ..., n. \tag{3.29}$$

[7] $\mathcal{H}_{k^s_i}$ matrix corresponds to the H^k matrix as defined in Theorem 1 in Bai and Ng (2002)

Theorem 3.3 and Corollary 3.6 establish that if we know the number of the common factors and the number of group-specific factors in each group, we can consistently estimate the common factor space as well as the respective group-specific factor spaces in the sense that the time average of the squared deviations between the estimated factors and those lie in the respective true factor spaces vanish as $N, T \to \infty$. Now a key question is how can we infer the number of the common factors and the number of the group-specific factors in each group from data. Following the approach developed in Chen (2010), we are going to construct an information criterion to determine the number of common factors and the number of group-specific factors. Generally, an information criterion consists addictively of two terms: the likelihood of a model under consideration plus a penalty term due to the dimensionality of the model. To ensure the consistence of the criterion, the penalty term must depend increasingly on the dimensionality of the model and converge at a slower rate than the likelihood term[8]. For this purpose we use the sum of squared residuals calculated as follows as the likelihood term of a group.

$$V_i(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) = \min_{\Lambda_i^s, F_i^s} \frac{1}{N_i T} \sum_{j=1}^{N_i} \sum_{t=1}^{T} [(X_{i,jt} - \hat{F}_t^{c\prime} \hat{\lambda}_{i,j}^c) - F_{i,t}^{s\prime} \lambda_{i,j}^s]^2 \text{ for } i = 1, 2, ...n.$$

To take into account of the fact that common factors across all groups represent a more restrictive model than a model with the same numbers of group-pervasive factors, we need to put less penalty on the number of common factors than on the average number of group-specific factors.

The following theorem gives a concrete formulation of the penalty term and thus a consistent model selection criterion to determine the number of groups, the number of common factors and the numbers of group-specific factors as well.

Theorem 3.7
Let $X_i^{s_n}$ represent the data classified into ith group by the classification procedure given in Chen (2010) and $(k^c, \{k_i^s\})$ represent a choice of the numbers of the common factors and the group-specific factors in each group. Under Assumptions 2.1 through 2.7 a model selection criterion

$$C(n, k^c, \{k_i^s\}, \{X_i^{s_n}\}) = \sum_{i=1}^{n} \frac{N_i}{N} V_i(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) + \hat{\sigma}^2(\bar{k}^s + \bar{h} + \overline{\alpha} k^c) g(N, T) \quad (3.30)$$

is a consistent model selection criterion for a grouped factor model with common factors and group-specific factors, if the following additional conditions are satisfied:

1. $\lim_{N\to\infty} \frac{N_i}{N} \to \alpha_i$, *where* $\frac{N_i}{N}$ *is the share of variables in the* $i-$*th group,* $\alpha_i > \underline{\alpha} > 0$ *and* $\overline{\alpha} = 1 - \underline{\alpha}$. *It is to note that* $\underline{\alpha}$ *is the lower bound for all candidate models.*

2. $g(N, T) \to +0, \quad C_{N,T}^2 g(N, T) \to \infty \qquad$ *as* $N, T \to \infty$,
where $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

3. *h is a real valued function over* $(0, 1)$ *with the following properties:*
(a) $0 < h(\alpha) < 1$ *for any* $0 \leq \alpha \leq 1$

[8] See Nishi (1984) for more details.

(b) $h(\alpha_i) \geq h(\alpha_j)$ *for any* $0 \leq \alpha_i \leq \alpha_j \leq 1$.
(c) $\sum_l \alpha_l h(\alpha_l) > \sum_j \alpha_j h(\alpha_j)$ *for and* $\{\alpha_j\} \precsim \{\alpha_l\}$.
We use the notation $\{\alpha_j\} \precsim \{\alpha_l\}$ *to present that* $\{\alpha_j\}$ *is a finer partition of the variables than* $\{\alpha_l\}$, *with* $\sum_l \alpha_l = \sum_j \alpha_j = 1$. $\bar{h} = \sum_{i=1}^{n} \alpha_i h(\alpha_i)$ *is the weighted average of* $h(\alpha_i)$ *over all groups.*

Remarks

Compared to the criterion in Chen (2010), the likelihood term remains the weighted average of the sums of squared residuals over all groups. The criterion keeps the the penalty term $\bar{h}$ due to dispersion of the groups. It modify the penalty term due to the average number of factor in that the number of common factors has a smaller penalty than the number of group-specific factors, reflecting the fact that one group-specific factor in each group gives more model flexibility than one common factor over all groups. Condition 1 is to make sure that the proportion of a group will not vanish asymptotically, Condition 2 is to get the right rate of convergence for the penalty term, and Condition 3 is to make sure that the average number of factors is the dominating parameter of the model and the dispersion of groups is a dominated parameter. While comparing two models, we compare first the dominating parameter, only when the dominating parameter are equal we compare the dispersion of the groups in the two models.

A concrete choice of $g(N, T)$ can be:

- $g(N, T) = \frac{N+T}{NT} \log\left(\frac{NT}{N+T}\right)$,

and a concrete choice of $h(N_i/N)$ is:

- $h(\hat{\alpha}_i) = \frac{\frac{\hat{\alpha}_i N+T}{\hat{\alpha}_i NT} \log\left(\frac{\hat{\alpha}_i NT}{\hat{\alpha}_i N+T}\right)}{\frac{\underline{\alpha} N+T}{\underline{\alpha} NT} \log\left(\frac{\underline{\alpha} NT}{\underline{\alpha} N+T}\right)} = \frac{g(\hat{\alpha}_i N, T)}{g(\underline{\alpha} N, T)}$,

where $\hat{\alpha}_i = \frac{N_j}{N}$. This h function is used in our simulation study.

3.3 Summary of the Estimation Procedure

- Step 1: Apply the procedure in Section 3.1 to obtain estimates of $\hat{n}$, $\hat{r}_i$ and $\hat{F}_{i,t}$ for $i = 1, 2, ..., \hat{n}$ and calculate estimates of standardized group-pervasive factors and covariance matrix $\hat{F}^*_{i,t}$, $\hat{F}^*_t$ and $\hat{\Sigma}^*$.
- Step 2: Choose a set of proper model parameters $(\hat{n}, k^c, \{k^s_i\}_{i=1}^{\hat{n}})$.
- Step 3: Calculate $\hat{F}^c = \frac{1}{\sqrt{\hat{n}}} \hat{h}'_{k^c} \hat{F}^*_t$.
- Step 4: Regress X_i on $\hat{F}^c$ to obtain the regression residuals $\hat{X}^{rs}_i = X_i - \hat{F}^c \hat{\Lambda}_i$.
- Step 5: Estimate group-specific factors and the loadings using the data of $\hat{X}^{rs}_i$ for each group $\hat{F}^s_i = \frac{1}{N_i T}(\hat{X}^{rs}_i \hat{X}^{rs\prime}_i)\sqrt{T} Q_i$, where Q_i contains the k^s_i eigenvectors corresponding to the k^s_i largest eigenvalues of the matrix $\hat{X}^{rs}_i \hat{X}^{rs\prime}_i$.
- Step 6: Calculate the model selection criterion values for alternative models
$$C(\hat{n}, k^c, \{k^s_i\}, \{X^{s_n}_i\}) = \sum_{i=1}^{\hat{n}} \frac{N_i}{N} V_i(k^s_i, \hat{F}^s_i | \hat{F}^c, k^c) + \hat{\sigma}^2(\bar{k}^s + \bar{h} + \overline{\alpha} k^c) g(N, T).$$
- Repeat Step 3 to Step 6 for alternative choices of model parameters and select the best model according to the model selection criterion.

4 Simulation Studies and an Application Example

4.1 Simulation Studies

The theoretical results presented in the last section can be used in two ways. If data are already available in grouped form, we do not need the classification step of the procedure in Section 3.3. The procedure can be used to determine the dimension of the overall pervasive factor space K and determine common factors across the groups and group-specific factors. If the existence of a grouped structure is only a working hypothesis and the groups need to be determined, the procedure in Section 3.3 will determine the number of groups, deliver a consistent classification of the variables and estimate the common factors and group-specific factors. Obviously, the performance of the estimation procedure for the former case can only be better than for the later case due the classification uncertainty involved in the later case. Therefore, we focus on a simulation study in the later case. If we have a satisfactory results here, we do not need to bother the performance in the former case.

The simulation study is conducted in order to assess the performance of the proposed estimation procedure in finite sample situations. In particular we want to assess the ability of the model selection criterion in identifying the true model, i.e. the number of groups and the number of common factors and the number of group-specific factors in each group. We use a vector consisting of the number of common factors r^c and the number of group-pervasive factors in each group r_i $i = 1, 2, ..., n$ and the number of common factors r^c to represent a GFM. For example $[322, 1|5]$ represents a GFM with three groups: the overall factor space in 5 dimensional; the number of factors in each group is 3, 2 and 2 respectively; and the number of common factors is one. From the relationship: $r_i^s = r_i - r^c$ we know the number of group-specific factors is 2, 2, and 1 respectively.

The data in the simulation study are generated from the following model:

$$X_{i,jt} = \sum_{l=1}^{r^c} F_{lt}^c \lambda_{i,lj}^c + \sum_{l=1}^{r_i^s} F_{i,lt}^s \lambda_{i,lj}^s + \sqrt{\theta}_i e_{i,jt} \qquad j = 1, 2, ...N_i, i = 1, 2, ...n,$$

where the common factor $F_t^c = (F_{1t}, F_{2t}, ..., F_{r^c t})$ and the group-specific factor $F_{i,t}^s = (F_{i,1t}^s, F_{i,2t}^s, ..., F_{i,r_i^s t}^s)'$ for the ith group are $r^c \times 1$ and $r_i^s \times 1$ vectors of $N(0, 1)$ variables; the factor loadings for the group $\lambda_{i,j} = (\lambda_{i,1j}, \lambda_{i,2j}, ..., \lambda_{i,r_i j})'$ is a $r_i \times 1$ matrix of $N(0, 1)$ variables: and $e_{i,jt} \sim N(0, 1)$. In this setting the common component of $X_{i,jt}$ has variance $r_i = r^c + r_i^s$. The base case under consideration is that the common component has the same variance as the idiosyncratic component, i.e. $\theta_i = r_i$. We consider the cases in which the number of groups in a GFM varies from 2 to 4; the number of variables in each group varies from 30 to 200; and the number of observations varies from 60 to 500. These are plausible data sets for monthly and quarterly macroeconomic variables and financial variables in practical applications.

In each simulation run we compare the value of the model selection criterion of the true model with those of alterative candidate models. The candidate models are chosen in a way that they include both more restrictive models and more general models in order to assess the sharpness of the model selection criterion in identifying the true model from similar model candidates. For a true model [2 2,1—3], [3 1,0—4] and [2 2 2,1—4] are more general models. In our simulation design the true model [2 2,1—3] has one common factor, the dimension of the overall pervasive factor space is three. i.e. there are two factor planes in a three dimensional space. Therefore, the model [3 1,0—4] is a more general model because it contains a three-dimensional subspace and a one-dimensional subspace, and [2 2 2,1—4] is also a more general model because it contains three two-dimensional subspaces. While [2 1,0—3] is a more restrictive model because it contains only one two-dimensional subspace and one one-dimensional subspace in a three dimensional factor space.

The outcomes of the simulation study are summarized in Table 1 to Table 5. The first three columns in Table 1 - Table 5 give numbers of variables and numbers of observations in the respective simulation runs. The numbers under the headers of N_i is the number of variables in a group and N is the total number of variables in the model. T is the number of observations. In the fourth column under the header *Candidates* we list the candidate models under consideration in a panel. The fifth column gives the true data-generating grouped factor models. For all simulation runs in a penal we compare the value of the model selection criterion of each candidate model in the penal with that of the true model and select a model with the minimal criterion value.

For all configurations in the simulation study $T = 80$ and $N_i = 30$ are good enough for a choice of the projection dimension K that is a key parameter in the classification step of the procedure (see Section 3.1). The numbers under the header $UGRP$ report the proportions that the dimension of the overall pervasive factor space K is correctly chosen by PC_{p1} and the $PC_{p1}(K)$ is larger than the minimum value of the model selection criterion of the candidate models in the penal. Since we can view ungrouped factor models as our candidate models, $PC_{p1}(K)$ is larger than the minimum value of the model selection criterion of the candidate models in the penal is a necessary condition that we prefer a grouped factor model rather than the ungrouped factor model. We observe that the proportions reported in this column are very high already for $T = 80$ and $N_i = 30$ and most of them are one or very close to one.

To assess whether the model selection criterion is biased towards grouped factor models in finite sample situation, we also conduct one simulation design with an ungrouped factor model $[4, 0|4]$ as a data generating model and contest it against alternative grouped factor models $[33, 2|4]$, $[333, 2|5]$ and $[3333, 2|6]$ (See Table 5.). In this particular setting, the model selection procedure performs very well in selecting the right model. It demonstrates that the model selection criterion is not biased towards grouped factor models in the considered finite sample situations.

The column under the header $CCLM$ reports the proportion of correctly selected true models among the candidates in 1000 simulation replications under the condition that the projection dimension is chosen correctly. Most of the numbers in the column of $CCLM$ are very close to one, indicating that for the considered configurations the estimation procedure performs well in selecting the correct model from the competing candidates, in many cases already for $T \geq 80$ and $N_i \geq 30$. Since the consistence of the model selection criterion holds under $T \to \infty$ and $N \to \infty$, it is not surprising that in some configurations for $T = 80$ and $N_i = 30$ the proportions of finding the correct models are still low (see Table 1, Table 3, Table 4 and Table 5.) However, we observe that for a given configuration the proportion of correctly identified models approaches to one with increasing T and N_i, for $T = 150$ and

$N_i = 30$ the results are already satisfactory.

The column under the header $MCLV$ gives the average proportion of misclassified variables in respective 1000 simulation runs. A perfect classification would have a zero in this column. Indeed, the numbers in this column are small, implying the classification works well. For those configurations without common factors, the numbers in the column of $MCLV$ are close to zero. For models with common factors, i.e. with intersected factor spaces, some misclassification is inevitable because there are data points lying closely to in the intersection of two factor subspaces. Therefore, we observe that for models with a common factor the proportions of misclassification are higher than for models without common factors. However, with increasing T and N_i the proportion of misclassification decreases, which reflects the consistency of the classification criterion.

$SFFC$[9] and $SFFS$ report the average goodness of fit of the estimated common factors and group-specific factors to the true common factors and the true group-specific factors, respectively. $SFFC$ and $SFFS$ are normalized to be between zero and one. A number close to one implies a good fitting of the estimated factors to the true factors. Because most of the variables are correctly classifies into their groups, the quality of the fit of the estimated factors to the true factors can be expected to be high. We observe the quality of the goodness of fit increases with the the increase of the number of variables and the number of observations. In most cases the numbers in the columns of $SFFC$ and $SFFS$ are above 90%, implying that the quality of the factor estimation is good. It is to note that the degree of misclassification does not have a sever consequence on the goodness of fit. This is because the misclassified data points lie closely to the intersection of two groups and have thus little impact on the goodness of fit of the estimated factors.

[9] $SFFC = \frac{tr(F^{c\prime}\hat{F}^c(\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\prime}F^c)}{tr(F^{c\prime}F^c)}$, $SFFS_i = \frac{tr(F_i^{s\prime}\hat{F}_i^s(\hat{F}_i^{s\prime}\hat{F}_i^s)^{-1}\hat{F}_i^{s\prime}F^s)}{tr(F^{s\prime}F^c)}$ and $SFFS = \frac{1}{n}\sum_i^n SFFS_i$.

Table 1: Estimation of grouped factor models

N_i	N	T	Candidates	DGP Model	CCLM	SFFC	SFFS	MCLV	UGRP
30	60	80	[2 2, 1]	[2 2, 0]	0.99	0.91		0.03	1.00
30	60	150	[2 1, 0]		1.00	0.91		0.01	1.00
30	60	300	[3 1, 0]		1.00	0.92		0.01	1.00
30	60	500	[3 2, 0]		1.00	0.92		0.00	1.00
60	120	80	[3 2, 1]		1.00	0.95		0.02	1.00
60	120	150	[3 2, 0]		1.00	0.95		0.01	1.00
60	120	300	[2 2 2, 0]		1.00	0.96		0.00	1.00
60	120	500	[2 2, 0]		1.00	0.96		0.00	1.00
200	400	80			1.00	0.97		0.02	1.00
200	400	150			1.00	0.98		0.01	1.00
200	400	300			1.00	0.98		0.01	1.00
200	400	500			1.00	0.99		0.00	1.00
30	60	80	[2 2, 0]	[2 2, 1]	0.94	0.90	0.94	0.10	1.00
30	60	150	[2 1, 0]		1.00	0.91	0.95	0.08	1.00
30	60	300	[3 2, 1]		1.00	0.91	0.95	0.05	1.00
30	60	500	[2 2 2, 0]		1.00	0.92	0.96	0.04	1.00
60	120	80	[2 2 2, 1]		1.00	0.93	0.97	0.10	1.00
60	120	150	[2 2, 1]		1.00	0.95	0.97	0.08	1.00
60	120	300			1.00	0.95	0.98	0.05	1.00
60	120	500			1.00	0.95	0.98	0.04	1.00
200	400	80			1.00	0.97	0.98	0.10	1.00
200	400	150			1.00	0.98	0.99	0.08	1.00
200	400	300			1.00	0.98	0.99	0.05	1.00
200	400	500			1.00	0.99	0.99	0.02	1.00
30	60	80	[3 2, 0]	[3 2, 1]	0.81	0.87	0.93	0.11	0.99
30	60	150	[3 3, 1]		0.95	0.89	0.94	0.08	0.99
30	60	300	[2 2, 0]		1.00	0.89	0.95	0.04	1.00
30	60	500	[3 3, 0]		1.00	0.90	0.95	0.05	1.00
60	120	80	[3 2 2, 0]		1.00	0.92	0.96	0.11	1.00
60	120	150	[3 2 2, 1]		1.00	0.94	0.97	0.07	1.00
60	120	300	[3 2, 1]		1.00	0.94	0.97	0.06	1.00
60	120	500			1.00	0.95	0.97	0.05	1.00
200	400	80			1.00	0.96	0.97	0.10	1.00
200	400	150			1.00	0.97	0.99	0.07	1.00
200	400	300			1.00	0.98	0.99	0.05	1.00
200	400	500			1.00	0.98	0.99	0.03	1.00

Notes: Table 1-Table 5 report the results of estimation of a GFM in 1000 Monte Carlo simulation runs. N_i is the number of variables in a group. N is the total number of variables in a model. T is the number of observations. In the column of $Candidates$ we list all grouped factor models under consideration. $CCLM$ is the proportion of correctly identified models. $SFFC$ and $SFFS$ are the average goodness of fit of the estimated common factors and group-specific factors to the true factors respectively. $MCLV$ gives the average proportion of misclassified variables in all variables over 1000 runs. $UGRP$ gives the proportion of correctly identified projection spaces in 1000 runs.

Table 2: Estimation of grouped factor models

N_i	N	T	Candidates	DGP Model	CCLM	SFFC	SFFS	MCLV	UGRP
30	60	80	[3 2, 0]	[3 3, 1]	0.99	0.85	0.92	0.05	0.98
30	60	150	[3 3, 0]		0.98	0.87	0.93	0.03	0.98
30	60	300	[2 2 2, 0]		1.00	0.88	0.93	0.01	1.00
30	60	500	[2 2 2, 1]		1.00	0.88	0.93	0.01	1.00
60	120	80	[4 2, 0]		1.00	0.91	0.95	0.04	1.00
60	120	150	[3 2, 1]		1.00	0.93	0.96	0.02	1.00
60	120	300	[3 3 3, 0]		1.00	0.94	0.97	0.01	1.00
60	120	500	[3 3 3, 1]		1.00	0.94	0.97	0.01	1.00
200	400	80	[3 3, 1]		1.00	0.96	0.98	0.04	1.00
200	400	150			1.00	0.97	0.99	0.02	1.00
200	400	300			1.00	0.98	0.99	0.01	1.00
200	400	500			1.00	0.98	0.99	0.00	1.00
30	60	80	[3 2, 0]	[3 3, 0]	1.00	0.87		0.02	0.81
30	60	150	[3 2, 1]		1.00	0.88		0.01	0.90
30	60	300	[3 3, 1]		1.00	0.88		0.00	0.90
30	60	500	[3 2, 1]		1.00	0.88		0.00	1.00
60	120	80	[3 3, 2]		1.00	0.93		0.02	1.00
60	120	150	[3 3 3, 1]		1.00	0.94		0.00	1.00
60	120	300	[3 3 2, 1]		1.00	0.94		0.00	1.00
60	120	500	[3 3 3, 0]		1.00	0.95		0.00	1.00
200	400	80	[3 3, 0]		1.00	0.97		0.01	1.00
200	400	150			1.00	0.97		0.01	1.00
200	400	300			1.00	0.98		0.00	1.00
200	400	500			1.00	0.98		0.00	1.00
30	90	80	[2 2 2 2, 0]	[2 2 2, 0]	1.00	0.90		0.04	0.88
30	90	150	[2 2 2, 1]		1.00	0.91		0.02	0.95
30	90	300	[3 3, 1]		1.00	0.91		0.01	1.00
30	90	500	[3 2 2, 0]		1.00	0.92		0.01	1.00
60	180	80	[2 2 1, 0]		1.00	0.95		0.04	1.00
60	180	150	[3 3, 0]		1.00	0.95		0.02	1.00
60	180	300	[2 2 2, 0]		1.00	0.96		0.01	1.00
60	180	500			1.00	0.96		0.01	1.00
200	600	80			1.00	0.97		0.04	1.00
200	600	150			1.00	0.98		0.02	1.00
200	600	300			1.00	0.99		0.01	1.00
200	600	500			1.00	0.99		0.00	1.00

Table 3: Estimation of grouped factor models

N_i	N	T	Candidates	DGP Model	CCLM	SFFC	SFFS	MCLV	UGRP
30	90	80	[2 2 2 2, 0]	[2 2 2, 1]	0.93	0.89	0.96	0.16	0.98
30	90	150	[3 2 2, 1]		0.99	0.91	0.96	0.11	1.00
30	90	300	[3 3, 1]		1.00	0.90	0.97	0.07	1.00
30	90	500	[3 2 2, 0]		1.00	0.92	0.97	0.06	1.00
60	180	80	[2 2 1, 0]		1.00	0.93	0.97	0.15	1.00
60	180	150	[3 3, 0]		1.00	0.95	0.98	0.10	1.00
60	180	300	[2 2 2, 1]		1.00	0.96	0.98	0.07	1.00
60	180	500			1.00	0.96	0.98	0.06	1.00
200	600	80			1.00	0.96	0.98	0.14	1.00
200	600	150			1.00	0.98	0.99	0.11	1.00
200	600	300			1.00	0.98	0.99	0.08	1.00
200	600	500			1.00	0.99	0.99	0.04	1.00
30	90	80	[3 3 2, 1]	[3 2 2, 1]	0.72	0.87	0.95	0.18	0.72
30	90	150	[4 4, 1]		0.98	0.89	0.96	0.11	0.87
30	90	300	[4 3, 1]		1.00	0.90	0.96	0.08	0.92
30	90	500	[3 2 2 2, 1]		1.00	0.91	0.96	0.06	0.96
60	180	80	[3 2 2, 0]		0.95	0.93	0.97	0.15	0.99
60	180	150	[3 2 2, 1]		1.00	0.94	0.98	0.11	1.00
60	180	300			1.00	0.95	0.98	0.08	1.00
60	180	500			1.00	0.95	0.98	0.06	1.00
200	600	80			1.00	0.96	0.98	0.14	1.00
200	600	150			1.00	0.97	0.99	0.11	1.00
200	600	300			1.00	0.98	0.99	0.07	1.00
200	600	500			1.00	0.98	0.99	0.04	1.00
30	90	80	[3 2 2 1, 0]	[3 2 1, 0]	0.99	0.90		0.12	0.85
30	90	150	[3 3 3, 1]		1.00	0.91		0.10	0.97
30	90	300	[3 3 2, 1]		1.00	0.92		0.06	1.00
30	90	500	[3 2 2, 0]		1.00	0.91		0.05	1.00
60	180	80	[4 4, 1]		1.00	0.94		0.10	0.99
60	180	150	[3 3, 0]		1.00	0.95		0.08	1.00
60	180	300	[3 2 1, 0]		1.00	0.96		0.05	1.00
60	180	500			1.00	0.96		0.04	1.00
200	600	80			1.00	0.98		0.11	1.00
200	600	150			1.00	0.98		0.07	1.00
200	600	300			1.00	0.99		0.05	1.00
200	600	500			1.00	0.99		0.03	1.00

Table 4: Estimation of grouped factor models

N_i	N	T	Candidates	DGP Model	CCLM	SFFC	SFFS	MCLV	UGRP
30	90	80	[3 2 2 2, 0]	[3 2 2, 0]	0.99	0.89		0.05	0.97
30	90	150	[3 3 3, 1]		0.99	0.90		0.03	1.00
30	90	300	[3 3 2, 1]		1.00	0.90		0.02	1.00
30	90	500	[3 2, 0]		1.00	0.91		0.01	0.90
60	180	80	[3 2 1, 0]		1.00	0.94		0.05	1.00
60	180	150	[4 3, 1]		1.00	0.95		0.02	1.00
60	180	300	[4 3, 0]		1.00	0.95		0.01	1.00
60	180	500	[3 2 2, 0]		1.00	0.95		0.01	1.00
200	600	80			1.00	0.97		0.04	1.00
200	600	150			1.00	0.98		0.02	1.00
200	600	300			1.00	0.98		0.01	1.00
200	600	500			1.00	0.99		0.00	1.00
30	90	80	[3 2 2 2, 0]	[3 3 2, 1]	0.73	0.87	0.95	0.12	0.61
30	90	150	[3 3 3, 1]		0.93	0.88	0.96	0.08	0.81
30	90	300	[3 2 2, 1]		1.00	0.89	0.96	0.06	0.84
30	90	500	[3 3 1, 0]		1.00	0.89	0.96	0.04	0.94
60	180	80	[4 3, 0]		0.96	0.92	0.97	0.12	0.90
60	180	150	[4 3, 1]		1.00	0.94	0.97	0.08	1.00
60	180	300	[3 2 2, 0]		1.00	0.94	0.98	0.05	1.00
60	180	500	[3 3 2, 1]		1.00	0.95	0.98	0.05	1.00
200	600	80			0.99	0.96	0.98	0.11	1.00
200	600	150			1.00	0.97	0.99	0.08	1.00
200	600	300			1.00	0.98	0.99	0.05	1.00
200	600	500			1.00	0.98	0.99	0.02	1.00
30	90	80	[3 3 3, 2]	[3 3 3, 1]	0.71	0.85	0.93	0.08	0.98
30	90	150	[5 3, 1]		0.97	0.87	0.95	0.04	0.99
30	90	300	[5 3, 0]		1.00	0.87	0.95	0.02	0.90
30	90	500	[3 3 2, 0]		1.00	0.88	0.96	0.02	0.90
60	180	80	[3 3 3 3, 0]		0.96	0.92	0.96	0.06	0.99
60	180	150	[3 3 3 3, 1]		0.99	0.93	0.97	0.04	1.00
60	180	300	[3 3 3 2, 1]		1.00	0.94	0.98	0.02	1.00
60	180	500	[3 3 3, 1]		1.00	0.94	0.98	0.01	1.00
200	600	80			1.00	0.96	0.98	0.06	1.00
200	600	150			1.00	0.97	0.99	0.03	1.00
200	600	300			1.00	0.98	0.99	0.02	1.00
200	600	500			1.00	0.98	0.99	0.00	1.00

Table 5: Estimation of grouped factor models

N_i	N	T	Candidates	DGP Model	CCLM	SFFC	SFFS	MCLV	UGRP
30	90	80	[3 3 3, 1]	[3 3 3, 2]	0.73	0.83	0.93	0.25	0.80
30	90	150	[5 3, 1]		0.95	0.87	0.95	0.15	0.91
30	90	300	[4 1, 0]		1.00	0.88	0.95	0.10	1.00
30	90	500	[3 3 2, 0]		1.00	0.88	0.96	0.07	1.00
60	180	80	[3 3 3 3, 0]		0.94	0.90	0.96	0.20	0.99
60	180	150	[3 3 3 3, 1]		1.00	0.91	0.97	0.14	1.00
60	180	300	[3 3 3 2, 1]		1.00	0.93	0.98	0.09	1.00
60	180	500	[3 3 3, 2]		1.00	0.94	0.98	0.07	1.00
200	600	80			1.00	0.94	0.98	0.18	1.00
200	600	150			1.00	0.96	0.99	0.14	1.00
200	600	300			1.00	0.98	0.99	0.09	1.00
200	600	500			1.00	0.98	0.99	0.05	1.00
30	90	80	[5 3, 1]	[3 3 3, 0]	0.84	0.87		0.03	0.85
30	90	150	[4 1, 0]		1.00	0.88		0.01	0.89
30	90	300	[3 3 3, 2]		1.00	0.88		0.01	0.80
30	90	500	[3 3 3, 1]		1.00	0.88		0.00	1.00
60	180	80	[3 3 2, 0]		1.00	0.93		0.02	0.98
60	180	150	[3 3 3 3, 0]		1.00	0.94		0.01	1.00
60	180	300	[3 3 3 3, 1]		1.00	0.94		0.00	1.00
60	180	500	[3 3 3 2, 1]		1.00	0.94		0.00	1.00
200	600	80	[3 3 3, 0]		1.00	0.97		0.02	1.00
200	600	150			1.00	0.98		0.01	1.00
200	600	300			1.00	0.98		0.00	1.00
200	600	500			1.00	0.98		0.00	1.00
30	90	80	[3 3, 2]	[4,0]	1.00	0.83			0.79
30	90	150	[3 3 3, 2]		1.00	0.84			1.00
30	90	300	[3 3 3 3, 2]		1.00	0.85			1.00
30	90	500	[4, 0]		1.00	0.85			1.00
60	180	80			1.00	0.91			1.00
60	180	150			1.00	0.92			1.00
60	180	300			1.00	0.92			1.00
60	180	500			1.00	0.92			1.00
200	600	80			1.00	0.97			1.00
200	600	150			1.00	0.97			1.00
200	600	300			1.00	0.97			1.00
200	600	500			1.00	0.98			1.00

4.2 A Demonstrative Empirical Example

In this subsection we apply the grouped factor models with common and group-specific factors to stock returns in the Australian Stock Exchange. The data used in this exercise are stock returns of companies included in ASX200. ASX200 is one of the most important share index in Australia Stock Exchange. It accounts for about 85% of the market capitalization of all stocks listed in Australia Stock Exchange. The data set consists of monthly returns of shares included in ASX200 from 2004 to 2009. All together there are 168 variables and each of them contains 77 observations[10]. A full name list of the shares is given in Table 7 in the appendix. We transform the data so that each series has mean zero. Using the PC_{1p} criterion of Bai and Ng (2002) we identify that the dimension of the overall pervasive factor space is 3. After choosing $K = 3$ we investigate 15 potential candidate models. These 15 candidate models include all possible group configurations up to 4 groups within a three dimensional overall pervasive factor space. We decide not to include group configurations with more that 4 groups because in those cases it is highly probable that some group will contain less than 30 variables such that the model selection criterion would become very unreliable. The estimation results for the considered models are summarizes in Table 6.

Table 6: Estimation of Grouped Dynamic Factor Models for ASX200

No.	Model	PC	No.	Model	PC
1	[1]	0.0057	12	[2 1 1, 0]	0.0053
2	[2]	0.0053	13	[2 2 1, 0]	0.0051
3	[3]	0.0052	14	[2 2 2, 1]	0.0051
4	[4]	0.0053	15	[2 2 2, 0]	0.0052
5	[5]	0.0053	16	[1 1 1 1, 0]	0.0051
6	[6]	0.0054	17	[2 1 1 1, 0]	0.0053
7	[1 1, 0]	0.0052	18	[2 2 1 1, 0]	0.0055
8	[2 1, 0]	0.0052	19	[2 2 2 1, 0]	0.0052
9	[2 2, 1]	0.0050	20	[2 2 2 2, 1]	0.0052
10	[2 2, 0]	0.0052	21	[2 2 2 2, 0]	0.0053
11	[1 1 1, 0]	0.0052			

Notes: The columns under the header *PC* report the values of the model selection criterion for the corresponding models.

In Table 6 we see that [2 2, 1] obtains the lowest criterion value. We conclude, therefore, [2 2, 1] is the most suitable model for the data. The estimation procedure separated the 168 shares into two group (See Fig.2): the first group contains 115 shares and the second group contains 53 shares. Detailed information on grouping of the shares is given in Table 7. The grouping of shares depicts an industrial clustering in returns: the second group contains with few exceptions exclusively companies in resource sectors including mining, energy, and exploration, while the first group contains companies from other industries. Among the 53 companies in the smaller group there are only 6 companies (See (*) in Table 9.) that are not in the ming

[10]Due to missing data in the investigation periods we include only 168 shares in steady of 200 shares in the study.

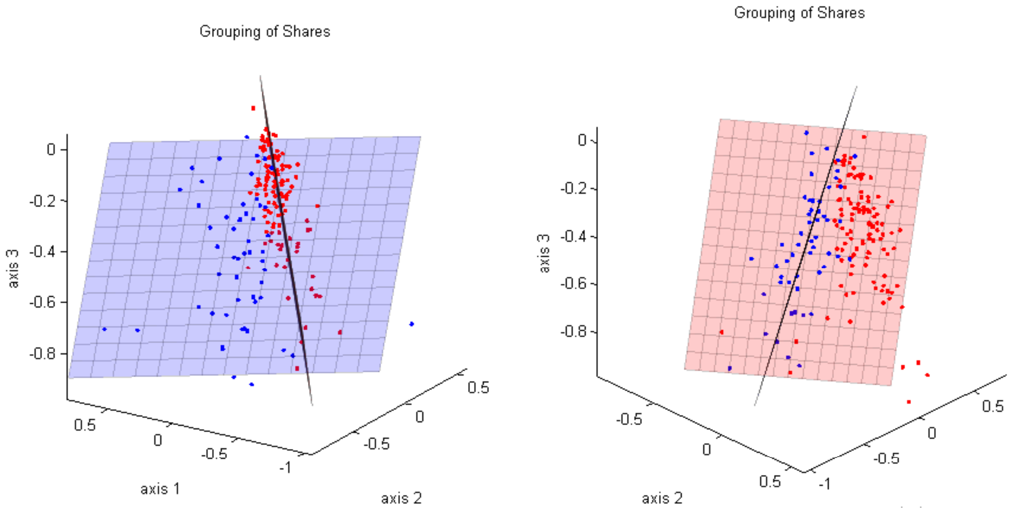

Figure 1: ASX200 shares in two groups in the factor space

and energy sectors. The first group contains 115 companies, among which only 6 companies (See (*) in Table 8 and 9.) are in the mining and energy sectors. This industrial clustering allows us to interpret the estimation results in the following way: we call the second group the resource group and the first group the non-resource group. The common factor of the two groups drives shares in both groups. It reflects the overall economic and financial situation in Australia. The group-specific factor of the resource group can be called resource-factor. It reflects the special business conditions in the resource sector. The group-specific factor of the non-resource group drives the shares in the non-resource sectors. The estimated common factor and the group-specific factors of the two groups are given in Fig. 2. In the upper left penal we

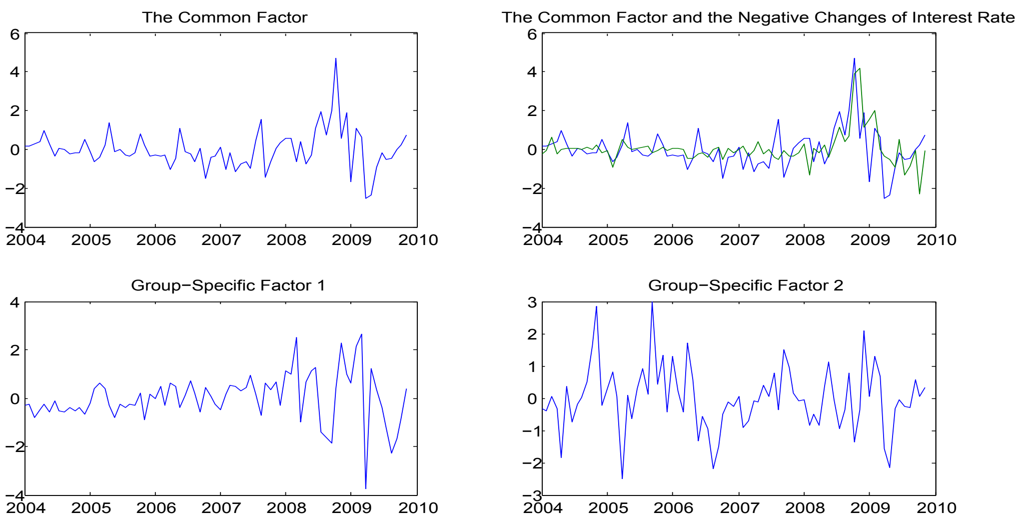

Figure 2: Common Factors and Group-specific Factors

observe that the common factor shows large volatilities during 2008-2009, reflecting the high uncertainty during the global financial crisis in the periods in 2008 and 2009. In the upper right panel we plot the common factor together with re-scaled negative changes in market interest rate represented by the rate of 180 days bank bills. We observe a strong comovement between the common factor and the changes in the interest rate, which is itself an important macroeconomic and financial market indicator. In fact they have a significant correlation of -0.46. The lower panel in Fig. 2 shows the group-specific factor of the two groups respectively. While in the non-resource group the second systematic risk is still predominantly related to the global financial crisis during the 2008 to 2009 periods, the group-specific systematic risk in the resource group presents special dynamics in the resource sector with little affection from the global financial crisis. This reflected the fact that the resource industry was one of the driving force in the Australia economy from 2004 to 2010. Even 2009, the non-rural commodity export expanded by 2 per cent and thus made a positive contribution to helping Australia avoid a possible economic recession.

The structural difference in these two groups is manifested also in their earning performance. The mean returns in the two groups differ astonishingly large from each other, despite that in the estimation we used exclusively demeaned data. While the mean monthly return of the non-resource group is 0.00676, this value is 0.02196 in the resource group, which is more than three times higher than that in the non-resource group. Corresponding to the higher mean return, the mean volatility of the returns measured by standard deviation is also higher in the resource group. The value of the mean volatility is 0.10729 in the resource group, which is one and a half times higher than the value of 0.06942 in the non-resource group. This confirms the old wisdom in financial markets: higher return higher risk. The good performance of the shares in the resource industry reflected the so called resource boom in Australia during the periods from 2004 to 2009.

Since the main focus of this paper is to present a method to estimate grouped factor models, we will leave further detailed empirical investigations on Australian stock market resulted from the estimated example above to another paper by concluding at this place that structured factor spaces exist in empirical data. Grouped factor model provides an effective way to uncover the unknown structure in the factor space. The resulting information from the estimation: the grouping of the variables, the common factors and the group-specific factors provide new insights into the data, which can be then used further for a more detailed and more nuanced analysis.

5 Concluding Remarks

In this paper we provide a method to estimate grouped factor models, in particular to estimate common factors as well as group-specific factors in the grouped factors models. The results provided in this paper can be used in two ways. If data are already available in grouped form, which is most of the cases in the literature on factor analysis of grouped data, we can skip Step 1 in the procedure given in Section 3.3. The consistency of the model selection criterion (3.30) remains valid under an a priori correct classification. Therefore the procedure will consistently determine the number of common factors that are pervasive across all groups and the number of group-specific factors in each group, it will also give consistent estimation of the common factor space as well as consistent estimation of the group-specific factor space for each group. In this context this paper contributes to the literature on statistics by providing a criterion to determine the number of multiset (perfect) canonical correlations among more than two groups, which has been an unsolved problem until now.

If the presence grouped structure is only a working hypothesis that has to be determined, the procedure given in Section 3.3, will consistently determine the number of groups and classify variables into respective groups. It will also give consistent estimation of the common factor space as well as consistent estimation of the group-specific factor space for each group. Furthermore, even when an a priori classification is available, our procedure can be used to assess the statistical adequacy of the a priori classification by comparing the value of the model selection criterion under a priori classification with that under endogenous classification.

Simulation results in Section 4 documents satisfactory performance of the proposed procedure in application-relevant possible data sets. As demonstrated by the simulation study in Section 4, if data are generated from a factor model without any grouped structure, the procedure will return a conventional ungrouped factor model with correctly identified number of factors. If data are generated from a grouped factor model, the procedure will output the number of groups and the number of factors in each group, a classification of the variables into the groups and group-pervasive factors. In this sense, our model generalizes the framework of the conventional factor models, such that it can be used to assess grouped structure in the data and estimate the group-pervasive factors, which may be useful for understanding the behavior of the data.

We set up the grouped factor models as approximate factor models and allow certain serial and cross-sectional correlation in the idiosyncratic errors. This approximate factor structure is more suitable for application to economic and financial data. The results can also be apply to strict factor models with uncorrelated idiosyncratic errors. Simulation study shows that our procedure has a satisfactory finite sample property in empirical relevant situations.

In an empirical example we demonstrate that grouped structures exist indeed in empirical data: the stock returns from 2004 to 2009 in the Australian stock exchange consists of two groups: one $resource$-group and one $non - resource$ group. Fitting a grouped factor model to the data, we identify three factors: one common factor that influenced all share returns, one resource factor that influenced only the shares in the resource group and one non-resource factors that impacted the shares in the non-resource groups. The obtained grouped information can be used then to conduct further more detailed and nuanced analysis.

6 Appendix

6.1 Proofs

Proof of **Proposition 2.8**

Because Λ_i and C_i are bounded and $\Lambda = (C_1\Lambda_1, C_2\Lambda_2, ..., C_n\Lambda_n)$, Λ is bounded.

$$\underset{(K\times K)}{\frac{\Lambda\Lambda'}{N}} = \sum_{i=1}^{n} \frac{N_i}{N} \underset{(K\times r_i)}{C_i} \underset{(r_i\times r_i)}{\frac{\Lambda_i\Lambda_i'}{N_i}} \underset{(r_i\times K)}{C_i}{}' \tag{6.31}$$

Let $\mathbf{b}$ be a $K \times 1$ nonzero vector. To show that $\frac{\Lambda\Lambda'}{N}$ converges to a positive definite matrix we need to show $\mathbf{b}'\frac{\Lambda\Lambda'}{N}\mathbf{b} > 0$ when N is large enough.

$$\underset{(1\times K)}{\mathbf{b}'} \underset{(K\times K)}{\frac{\Lambda\Lambda'}{N}} \underset{(K\times 1)}{\mathbf{b}} = \sum_{i=1}^{n} \frac{N_i}{N} \underset{(1\times K_i)}{\mathbf{b}'C_i} \underset{(K_i\times K_i)}{\frac{\Lambda_i\Lambda_i'}{N_i}} \underset{(K_i\times 1)}{C_i'\mathbf{b}} \tag{6.32}$$

Because $\frac{\Lambda_i\Lambda_i'}{N_i}$ converges to a positive definite matrix according to Assumption 2.5, the summands on the right hand side of the equation above are all nonnegative. In order to show the sum is strictly positive we need to show at least one summand is strictly positive.

If $C_i'\mathbf{b} = 0$ for all $i = 1, 2, ..., n$, it would imply that all column vectors in $(C_1, C_2, ..., C_n)$ are orthogonal to $\mathbf{b}$. This contradicts to the assumption that $rank(C_1, C_2, ..., C_n) = K$. Therefore, for some $i \in \{1, 2, ..., n\}$ we have $C_i'\mathbf{b} \neq 0$. Because $\frac{\Lambda_i'\Lambda_i}{N_i}$ converges to a positive definite matrix, we have $\mathbf{b}'C_i\frac{\Lambda_i'\Lambda_i}{N_i}C_i'\mathbf{b} > 0$ for $C_i'\mathbf{b} \neq 0$ and N large enough. Further we have $\frac{N_i}{N} \to \alpha_i > 0$. Therefore, the summand $\frac{N_i}{N}\mathbf{b}'C_i\frac{\Lambda_i'\Lambda_i}{N_i}C_i'\mathbf{b}$ is strictly positive. It follows the sum in equation (6.32) is strictly positive.

□

Lemma 6.1

Let $F_{it}^{*'} = (F_t^{c*'}, F_{it}^{s*'})$ denote the standardized group-pervasive factors of group i and let $F_t^{*'} = (F_{1t}^{*'}, F_{2t}^{*'}, ..., F_{nt}^{*'})$ collect all standardized group-pervasive factors over n groups. Under Assumptions 2.1 to 2.4, we have:

(i) The covariance matrix $\Sigma^* = E(F_t^* F_t^{*'})$ has K positive eigenvalues.

(ii) The r^c largest eigenvalues of Σ^* are identically of value n.

(iii) a_{r^c} as defined in (6.33) is an eigenvector matrix corresponding to the r^c largest eigenvalues of Σ^*.

$$\begin{aligned} a_{r^c}' &= \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 & \dots & 0 & 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & 0 & 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots \\ 0 & \frac{1}{\sqrt{n}} & \ddots & \vdots & 0 & \dots & 0 & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & 0 & \dots & 0 & 0 & \frac{1}{\sqrt{n}} & \dots \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \dots & \vdots & \vdots & \ddots & \dots \\ 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & 0 & 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & 0 & 0 & 0 & \dots \end{pmatrix} \\ &= \frac{1}{\sqrt{n}}(\underset{(r^c\times r^c)}{I}, \underset{(r^c\times r_1^s)}{\mathbf{0}}, \underset{(r^c\times r^c)}{I}, \underset{(r^c\times r_2^s)}{\mathbf{0}}, ..., \underset{(r^c\times r^c)}{I}, \underset{(r^c\times r_n^s)}{\mathbf{0}}). \end{aligned} \tag{6.33}$$

(iv) $F_t^{c*} = \frac{1}{\sqrt{n}}(a'_{r^c} F_t^*)$

(v) *Let a^K denote the matrix of the eigenvectors corresponding to the r^c+1 to K largest eigenvalues of Σ^*. Under the additional Assumption 3.2, each eigenvector in a^K is unique up to a sign. If we normalize the first non-zero element of the eigenvectors to be positive, a^K is unique. a^K matrix has the following structure.*

$$a^{K'} = \frac{1}{\sqrt{n}}(\underset{(K^s\times r^c)}{\mathbf{0}}, \underset{(K^s\times r_1^s)}{a_1^K}, \underset{(K^s\times r^c)}{\mathbf{0}}, \underset{(K^s\times r_2^s)}{a_2^K}, ..., \underset{(K^s\times r^c)}{\mathbf{0}}, \underset{(K^s\times r_n^s)}{a_n^K}).$$

(vi) *Let h_{k^c} be an eigenvector matrix corresponding to the k^c largest eigenvalues of Σ^*. We have:*

- *for $k^c = r^c$, there exists an $(r^c \times r^c)$ orthogonal matrix H_{r^c}, such that*

$$h_{k^c} = a_{r^c} H_{r^c}$$
$$\frac{1}{\sqrt{n}}(h'_{k^c} F_t^*) = H'_{r^c} F_t^{c*}.$$

- *for $k^c < r^c$, the H_{r^c} matrix can be decomposed: $H_{r^c} = (\underset{(r^c\times k^c)}{H_{rk^c}}, \underset{(r^c\times(r^c-k^c))}{H_{rk^c\perp}})$, and we have*

$$h_{k^c} = a_{r^c} H_{rk^c}$$
$$\frac{1}{\sqrt{n}}(h'_{k^c} F_t^*) = H'_{rk^c} F_t^{c*},$$

where H_{rk^c} is the first k^c columns of the $(r^c \times r^c)$ orthogonal matrix H_{r^c}.

- *for $r^c < k^c \leq K$, we have a $(k^c \times k^c)$ H_{k^c} with $H_{k^c} = \begin{pmatrix} H_{r^c} & 0 \\ 0 & S_{k^c-r^c} \end{pmatrix}$, where $S_{k^c-r^c}$ is a $(k^c-r^c)\times(k^c-r^c)$ diagonal matrix with either $+1$ or -1 on the diagonal, such that*

$$h_{k^c} = (a_{r^c}, a^{ps}) H_{k^c} = a_{k^c} H_{k^c}$$
$$\frac{1}{\sqrt{n}}(h'_{k^c} F_t^*) = H'_{k^c} \begin{pmatrix} F_t^{c*} \\ F_t^{ps*} \end{pmatrix},$$

where $F_t^{ps*} = \frac{1}{\sqrt{n}} a^{ps'} F_t^*$ is a linear combination of $(F_{1,t}^{s*}, F_{2,t}^{s*}, ..., F_{n,t}^{s*})$ and a^{ps} is the first $(k^c - r^c)$ columns of a^K and $a_{k^c} = (a_{r^c}, a^{ps})$ is the first k^c columns of a^K.

Proof of (i):

Under Assumptions 2.1, 2.2 and 2.4 (i) we have $E(F_{i,t}F'_{i,t}) = C'_i \Sigma^p C_i = \Sigma_{ii}$ and $E(F_{i,t}^* F_{i,t}^{*'}) = \Sigma_{ii}^{-\frac{1}{2}} C'_i \Sigma^p C_i \Sigma_{ii}^{-\frac{1}{2}}$. It follows

$$E(F_t^* F_t^{*'}) = \begin{pmatrix} \Sigma_{11}^{-\frac{1}{2}} C'_1 \\ \Sigma_{22}^{-\frac{1}{2}} C'_2 \\ \vdots \\ \Sigma_{nn}^{-\frac{1}{2}} C'_n \end{pmatrix} \underset{(K\times K)}{\Sigma^p} \begin{pmatrix} C_1 \Sigma_{11}^{-\frac{1}{2}} & C_2 \Sigma_{22}^{-\frac{1}{2}} & \dots & C_n \Sigma_{nn}^{-\frac{1}{2}} \end{pmatrix} = \Sigma^*.$$

From the equation above it is straightforward to verify that the rank of $\Sigma^* = E(F_t^{c*}F_t^{c*'})$ is K, which is also the rank of Σ^p. Therefore the covariance matrix Σ^* has K nonzero positive eigenvalues. Note that $K = r^c + K^s$.

Proof of (ii):

According to Assumption 2.1 that the common factors are not correlated with the group-specific factors. Therefore, the covariance matrix of F_t^* has the following structure.

$$
\begin{aligned}
\Sigma^* = E(F_t^* F_t^{*'}) &= E\begin{pmatrix} F_{1t}^*F_{1t}^{*'} & F_{1t}^*F_{nt}^{*'} & \dots & F_{1t}^*F_{nt}^{*'} \\ F_{2t}^*F_{1t}^{*'} & F_{2t}^*F_{2t}^{*'} & \dots & F_{2t}^*F_{nt}^{*'} \\ \vdots & \ddots & \dots & \vdots \\ F_{nt}^*F_{1t}^{*'} & F_{nt}^*F_{2t}^{*'} & \dots & F_{nt}^*F_{nt}^{*'} \end{pmatrix} \\
&= \begin{pmatrix} I_{r^c} & 0 & I_{r^c} & 0 & \dots & I_{r^c} & 0 \\ 0 & \Sigma_{11}^{s*} & 0 & \Sigma_{12}^{s*} & \dots & 0 & \Sigma_{1n}^{s*} \\ I_{r^c} & 0 & I_{r^c} & 0 & \dots & I_{r^c} & 0 \\ 0 & \Sigma_{21}^{s*} & 0 & \Sigma_{22}^{s*} & \dots & 0 & \Sigma_{2n}^{s*} \\ \vdots & \vdots & \ddots & \ddots & \dots & \vdots & \vdots \\ I_{r^c} & 0 & I_{r^c} & 0 & \dots & I_{r^c} & 0 \\ 0 & \Sigma_{n1}^{s*} & 0 & \Sigma_{n2}^{s*} & \dots & 0 & \Sigma_{nn}^{s*} \end{pmatrix}
\end{aligned}
$$

Since rearranging columns and rows will not change the eigenvalues of a matrix, we can pull the columns and the rows corresponding to the common factors to the upper left of Σ^* and obtain the following block diagonal matrix:

$$
\Sigma^\diamond = \begin{pmatrix} I_{r^c} & I_{r^c} & \dots & I_{r^c} & 0 & \dots & 0 \\ I_{r^c} & I_{r^c} & \dots & I_{r^c} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ I_{r^c} & I_{r^c} & \dots & I_{r^c} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \Sigma_{11}^{s*} & \dots & \Sigma_{1n}^{s*} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \Sigma_{n1}^{s*} & \dots & \Sigma_{nn}^{s*} \end{pmatrix} = \begin{pmatrix} \Sigma^{c*} & 0 \\ 0 & \Sigma^{s*} \end{pmatrix}. \tag{6.34}
$$

Σ^* and $\Sigma^\diamond$ have the same eigenvalues. Since $\Sigma^\diamond$ is a block diagonal matrix, the eigenvalues of Σ^{c*} and those of Σ^{s*} are the eigenvalues of Σ^*. It is straightforward to verify that Σ^{c*} has r^c identical positive eigenvalues n. To prove (ii) we need to show that the eigenvalues of Σ^{s*} are strictly less than n. Now, for any none zero vectors a_i and a_j we have:

$$
\frac{a_i'\Sigma_{ij}^{s*}a_j}{\sqrt{a_i'\Sigma_{ii}^{s*}a_i}\sqrt{a_j'\Sigma_{jj}^{s*}a_j}} = \frac{a_i'\Sigma_{ij}^{s*}a_j}{\sqrt{a_i'a_i}\sqrt{a_j'a_j}} \leq 1. \tag{6.35}
$$

The left hand side of the inequality above can be interpreted as the correlation coefficient between $a_i'F_{it}^{s*}$ and $a_j'F_{jt}^{s*}$. The equality holds only when they are perfect correlated. From (6.35) we have:

$$
a_i'\Sigma_{ij}^{s*}a_j \leq \sqrt{a_i'a_i}\sqrt{a_j'a_j}.
$$

Summing up over i and j we have

$$
\sum_{i=1}^{n}\sum_{j=1}^{n} a_i'\Sigma_{ij}^{s*}a_j \leq \sum_{i=1}^{n}\sum_{j=1}^{n} \sqrt{a_i'a_i}\sqrt{a_j'a_j}.
$$

The equality holds only when the equality holds in (6.35) for every combination of i and j. This implies however that there are still common factors among the n groups and it is henceforth contradictory to the assumption that there is no common factors among the group-specific factors F_i^{s*} over all groups. Therefore we have

$$a'\Sigma^{s*}a = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i'\Sigma_{ij}^{s*}a_j < \sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{a_i'a_i}\sqrt{a_j'a_j} \leq \sum_{i=1}^{n}\sum_{j=1}^{n} a_i'a_i = n\sum_{i=1}^{n} a_i'a_i = na'a.$$

Because the inequality above holds for all none zero vector a. In particular for a being the eigenvector corresponding to the largest eigenvalues of Σ^{s*} denoted by λ_{max}^{s*}, we have

$$a'\Sigma^{s*}a = \lambda_{max}^{s*}a'a < na'a.$$

We proved that $\lambda_{max}^{s*} < n$.

Proof of (iii):

It is straightforward to verify that the $(nr^c \times r^c)$ matrix b_{r^c} given below is an eigenvector matrix corresponding to the r^c non zeros eigenvalues of Σ^{c*}.

$$b_{r^c}' = \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots \\ 0 & \frac{1}{\sqrt{n}} & 0 & 0 & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & 0 & \frac{1}{\sqrt{n}} & \dots \\ 0 & \ddots & \ddots & & \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots \\ 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & 0 & \frac{1}{\sqrt{n}} & 0 & \dots & \dots \end{pmatrix} = \frac{1}{\sqrt{n}}(\underbrace{I_{r^c}, I_{r^c}, ..., I_{r^c}}_{\text{n groups}}). \tag{6.36}$$

Note that Σ^{c*} is the covariance matrix of $\begin{pmatrix} F_t^{c*} \\ \vdots \\ F_t^{c*} \end{pmatrix}$ and we have

$$b_{r^c}' \begin{pmatrix} F_t^{c*} \\ \vdots \\ F_t^{c*} \end{pmatrix} = \frac{1}{\sqrt{n}}(I_{r^c}, ..., I_{r^c}) \begin{pmatrix} F_t^{c*} \\ \vdots \\ F_t^{c*} \end{pmatrix} = \sqrt{n}F_t^{c*}. \tag{6.37}$$

It is to note that the none zero elements on a row of b_{r^c}' pick out the same component of the common factor across n groups. Since $\Sigma^{\diamond}$ is a block diagonal matrix, the eigenvalues of Σ^{c*} and those of Σ^{s*} are the eigenvalues of $\Sigma^{\diamond}$, and the eigenvectors of Σ^{c*} and Σ^{s*}, extended by a zero block, are the corresponding eigenvectors of $\Sigma^{\diamond}$.

Therefore, $(b_{r^c}', \mathbf{0}')'$ is an eigenvector matrix of $\Sigma^{\diamond}$ corresponding to the eigenvalues of n. Now $\Sigma^{\diamond}$ is the covariance matrix of F_t^* after pulling the common factors to the beginning of the vector. So in the original order of F_t^* the corresponding eigenvector matrix of Σ^* is

$$a_{r^c}' = \frac{1}{\sqrt{n}}(I_{r^c}, \mathbf{0}_{r_1^s}, I_{r^c}, \mathbf{0}_{r_2^s}, ..., I_{r^c}, \mathbf{0}_{r_n^s}), \tag{6.38}$$

where $\mathbf{0}_{r_i^s}$ denotes an $(r^c \times r_i^s)$ zero matrix.

Proof of (iv):

$$\frac{1}{\sqrt{n}}(a_{r^c}'F_t^*) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_i'F_{it}^* = \frac{1}{n}\sum_{i=1}^{n}(I_{r^c}, \mathbf{0}_{r_i^s}) \begin{pmatrix} F_t^{c*} \\ F_{i,t}^{s*} \end{pmatrix} = F_t^{c*}. \tag{6.39}$$

Proof of (v):

Under Assumption 3.2 the r^c+1 to K eigenvalues of Σ^* are unique. This implies the K^s positive eigenvalues of Σ^{s*} are unique. Let $b^{K'} = (\underset{(K^s\times r_1^s)}{a_1^{K'}}, \underset{(K^s\times r_2^s)}{a_2^{K'}}, ..., \underset{(K^s\times r_n^s)}{a_n^{K'}})$ be the matrix of the K^s eigenvectors corresponding to the positive eigenvalues of Σ^{s*}. Then $(\mathbf{0}', b^{K'})'$ is the eigenvector matrix of $\Sigma^\diamond$ corresponding to the K^s eigenvalues less than n. $\Sigma^\diamond$ is the covariance matrix of F_t^* after resorting the group-specific factors the rear. So in the original order of F_t^* the corresponding eigenvector matrix is

$$a^{K'} = (\underset{(K^s\times r^c)}{\mathbf{0}}, \underset{(K^s\times r_1^s)}{a_1^{K'}}, \underset{(K^s\times r^c)}{\mathbf{0}}, \underset{(K^s\times r_2^s)}{a_2^{K'}}, ..., \underset{(K^s\times r^c)}{\mathbf{0}}, \underset{(K^s\times r_n^s)}{a_n^{K'}}). \tag{6.40}$$

Proof of (vi)

For the case of $k^c = r^c$, since h_{r^c} and a_{r^c} are both eigenvector matrices corresponding to the r^r nonzero eigenvalue of Σ^*, we have

$$\Sigma^* = h_{r^c}(nI_{r^c})h'_{r^c} = a_{r^c}(nI_{r^c})a'_{r^c}.$$

Postmultiply the equation above with h_{r^c} we have

$$h_{r^c} = h_{r^c}h'_{r^c}h_{r^c} = a_{r^c}a'_{r^c}h_{r^c}.$$

Let $H_{r^c} = a'_{r^c}h_{r^c}$. We have $H'_{r^c}H_{r^c} = (a'_{r^c}h_{r^c})'(a'_{r^c}h_{r^c}) = I_{r^c}$. Therefore, for any h_{r^c} there exists an orthogonal matrix H_{r^c} such that $h_{r^c} = a_{r^c}H_{r^c}$. It follows

$$\frac{1}{\sqrt{n}}h'_{r^c}F_t^* = \frac{1}{\sqrt{n}}(a_{r^c}H_{r^c})'F_t^* = H'_{r^c}\frac{1}{\sqrt{n}}a'_{r^c}F_t^* = H'_{r^c}F_t^{c*}. \tag{6.41}$$

In the last step we have used the result in (6.39).

For $k^c < r^c$, h_{k^c} is the first k^c columns of the eigenvector matrix h_{r^c} matrix. Decomposing H_{r^c} into first k^c columns and the rest: $H_{r^c} = (H_{rk^c}, H_{rk^c\perp})$, we have $h_{k^c} = a_{r^c}H_{rk^c}$ and

$$\frac{1}{\sqrt{n}}(h'_{k^c}F_t^*) = H'_{rk^c}F_t^{c*}.$$

For the case $r^c < k^c \le K$, we have $h_{k^c} = (h_{r^c}, h_{r^c+1:k^c})$, where h_{r^c} represents first r^c eigenvectors and $h_{r^c+1:k^c}$ represents the eigenvectors corresponding to r^c+1 through k^c largest eigenvalues of Σ^*. For h_{r^c} we have $h_{r^c} = a_{r^c}H_{r^c}$. According to Assumption 3.2, the eigenvectors that corresponds to r^c+1 through k^c largest eigenvalues of Σ^* are unique up to a sign of the vectors. Therefor we have $h_{r^c+1:k^c} = a^{ps}S_{k^c-r^c}$, where $S_{k^c-r^c}$ is a (k^c-r^c) diagonal matrix with either $+1$ or -1 on the diagonal. Hence we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}(h'_{k^c}F_t^*) &= \frac{1}{\sqrt{n}}\begin{pmatrix} h'_{r^c} \\ h'_{r^c+1:k^c} \end{pmatrix} F_t^* \\
&= \frac{1}{\sqrt{n}}\begin{pmatrix} H'_{r^c}a'_{r^c} \\ S'_{k^c-r^c}a^{ps'} \end{pmatrix} F_t^* \\
&= \begin{pmatrix} H'_{r^c} & 0 \\ 0 & S_{k^c-r^c} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{n}}a'_{r^c} \\ \frac{1}{\sqrt{n}}a^{ps'} \end{pmatrix} F_t^* \\
&= H'_{k^c}\begin{pmatrix} F_t^{c*} \\ F_t^{ps*} \end{pmatrix}.
\end{aligned}$$

Because a^{ps} contains the first $(k^c - r^c)$ columns of a^K, following equation (6.40) we have

$$a^{ps'} = (\underset{(kr^c \times r^c)}{\mathbf{0}_{r^c}}, \underset{(kr^c \times r_1^s)}{a_1^{ps'}}, \underset{(kr^c \times r^c)}{\mathbf{0}_{r^c}}, \underset{(kr^c \times r_2^s)}{a_2^{ps'}}, ..., \underset{(kr^c \times r^c)}{\mathbf{0}_{r^c}}, \underset{(kr^c \times r_n^s)}{a_n^{ps'}}), \tag{6.42}$$

with $kr^c = k^c - r^c$.

$$\begin{aligned}
F_t^{ps*} &= \frac{1}{\sqrt{n}} a^{ps'} F_t^* \\
&= \frac{1}{\sqrt{n}} (\mathbf{0}_{r^c}, a_1^{ps'}, \mathbf{0}_{r^c}, a_2^{ps'}, ..., \mathbf{0}_{r^c}, a_n^{ps'}) \begin{pmatrix} F_{1,t}^* \\ F_{2,t}^* \\ \vdots \\ F_{n,t}^* \end{pmatrix} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{0}_{r^c}, a_i^{ps'}) \begin{pmatrix} F_t^{c*} \\ F_{i,t}^{s*} \end{pmatrix} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_i^{ps'} F_{i,t}^{s*}
\end{aligned}$$

Therefore, $F_t^{ps*} = \frac{1}{\sqrt{n}} a^{ps'} F_t^*$ is a linear combination of the standardized group-specific factors only. □

Proof of Proposition 3.1

Proof of (i): See Lemma 6.1 (i).

Proof of (ii): By Lemma 6.1 (ii) we know that the first r^c largest eigenvalues of Σ^* are n. Let $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, ...\mathbf{a}_{r^c})$ be a $(\sum_{i=1}^n r_i \times r^c)$ matrix. For the objective function of the maximization problem (3.16) it holds

$$\text{tr}\,(\mathbf{a}'\Sigma^*\mathbf{a}) = \sum_{k=1}^{r^c} \mathbf{a}_k' \Sigma^* \mathbf{a}_k \leq \sum_{k=1}^{r^c} \lambda_{max} \mathbf{a}_k' \mathbf{a}_k,$$

where $\mathbf{a}_k$ is the kth column of $\mathbf{a}$, λ_{max} is the largest eigenvalue of Σ^*. The equality holds when $\mathbf{a}_k$ is the eigenvector corresponding to the eigenvalue λ_{max}. Because Σ^* has r^c identical largest eigenvalues n, the solutions are eigenvectors that correspond to the r^c largest eigenvalues of Σ^*. Now we will show that the length $\sqrt{n}$ eigenvectors satisfy the restrictions given in (3.15). Following Lemma 6.1 (iii) and (vi) any set of eigenvectors corresponding to the r^c largest eigenvalues of Σ^* can be written as an orthogonal transformation of a_{r^c}. This implies for $\mathbf{a}$ that is a set of length $\sqrt{n}$ eigenvectors corresponding to the r^c largest eigenvalues of Σ^* we have:

$$\mathbf{a}' = \sqrt{n} H_{r^c}' a_{r^c}' = H_{r^c}'(I_{r^c}, \mathbf{0}_{r_1^s}, I_{r^c}, \mathbf{0}_{r_2^s}, ..., I_{r^c}, \mathbf{0}_{r_n^s}) = (\underbrace{H_{r^c}', \mathbf{0}_{r_1^s}}_{a_1'}, \underbrace{H_{r^c}', \mathbf{0}_{r_2^s}}_{a_2'}, ..., \underbrace{H_{r^c}', \mathbf{0}_{r_n^s}}_{a_n'}),$$

where H_{r^c} is an $(r^c \times r^c)$ orthogonal matrix. It is straightforward to verify that $\mathbf{a} = \sqrt{n} a_{r^c}$ satisfies the restrictions in (3.15), i.e. $a_i' a_i = (H_{r^c}', 0) \begin{pmatrix} H_{r^c} \\ 0 \end{pmatrix} = I_{r^c}$ for $i = 1, 2, ..., n$.

Proof of (iii) is given by Lemma 6.1 (iii) and (iv).

Proof of (iv) is given by Lemma 6.1 (v) and (vi).

□

Next we present four lemmata which are needed in the proof of Theorem 3.3.

Lemma 6.2

Let A and $\hat{A}$ be both positive semidefinite matrices of same dimension with $||\hat{A} - A||^2 = O_p(C_{N,T}^{-2})$. Let $\hat{\lambda}_i$ and $\hat{P}_i$ be the i-th nonzero eigenvalue and the corresponding eigenvector of $\hat{A}$ and λ_i and P_i be the ith nonzero eigenvalue and the corresponding eigenvectors of A. Then we have

(i)

$$||\hat{\lambda}_i - \lambda_i||^2 = O_p(C_{N,T}^{-2}) \tag{6.43}$$

$$||sin(\theta(P_i, \hat{P}_i))||^2 = O_p(C_{N,T}^{-2}), \tag{6.44}$$

where $\theta(P_i, \hat{P}_i)$ denotes the canonical angles between the subspace spanned by P_i and that spanned by $\hat{P}_i$.

(ii) If the nonzero eigenvalues are different, we have

$$||P_i - \hat{P}_i||^2 = O_p(C_{N,T}^{-2}), \tag{6.45}$$

(iii) If the multiplicity of λ_i is r^c, then there exists an $(r^c \times r^c)$ orthogonal matrix H_{r^c} such that,

$$||\hat{P}_i - P_i H_{r^c}||^2 = O_p(C_{N,T}^{-2}). \tag{6.46}$$

(iv) If the multiplicity of λ_i is r^c, then for any give eigenvectors $\hat{P}_i$ of $\hat{A}$ there exists a particular set of eigenvectors P_i^* of A such that,

$$||\hat{P}_i - P_i^*||^2 = O_p(C_{N,T}^{-2}). \tag{6.47}$$

Proof of (i) Applying Theorem 5.2 and Theorem 5.3 in Truhar (2000) and replacing δA in the theorems by $O_p(C_{N,T}^{-2})$ we obtain the results in (6.43) and (6.44).

Proof of (ii)

$$\begin{aligned}
P_i'\hat{P}_i &= (P_i - \hat{P}_i + \hat{P}_i)'(\hat{P}_i - P_i + P_i) \\
&= -(P_i - \hat{P}_i)'(P_i - \hat{P}_i) + (P_i - \hat{P}_i)'P_i + \hat{P}_i'(\hat{P}_i - P_i) + \hat{P}_i'P_i \\
&= -||P_i - \hat{P}_i||^2 + P_i'P_i - \hat{P}_i'P_i - \hat{P}_i'P_i + \hat{P}_i'\hat{P}_i + \hat{P}_i'P_i \\
&= -||P_i - \hat{P}_i||^2 + 1 - \hat{P}_i'P_i + 1
\end{aligned}$$

$$||P_i - \hat{P}_i||^2 = 2 - 2cos(\theta) = 2(1 - cos(\theta)) = 2\big(1 - \sqrt{1 - sin^2(\theta)}\big) = O_p(C_{N,T}^{-2}).$$

Proof of (iii)

$||\sin(\theta(P_i, \hat{P}_i))||^2 = O_p(C_{N,T}^{-2})$ implies that there exist r^c-pairs of independent linear combinations of P_i and $\hat{P}$ such that the angle between each pair satisfies $||\sin(\theta_j(P_i d_j, \hat{P}_i b_j))||^2 = O_p(C_{N,T}^{-2})$ for $j = 1, 2, ..., r^c$. Let $Q_j = P_i d_j$ and $\hat{Q}_j = \hat{P}_i b_j$ be normalized to have unit length. We have $Q_j'Q_j = d_j'P_i'P_i d_j = d_j' I_{k^c} d_j = 1$.

$$||Q_j - \hat{Q}_j||^2 = 2 - 2\hat{Q}_j'Q_j = 2(1 - cos(\theta_j)) = O_p(C_{N,T}^{-2}). \tag{6.48}$$

Put all r^c linear combinations together we have

$$Q = (Q_1, ..., Q_{k^c}) = P_i(d_1, d_2, ...d_{k^c}) = P_i D$$

and

$$\hat{Q} = (\hat{Q}_1, ..., \hat{Q}_{k^c}) = \hat{P}_i(a_1, a_2, ...a_{k^c}) = \hat{P}_i B$$

$$\begin{aligned} ||Q - \hat{Q}||^2 &\leq \sum_{j=1}^{k^c} ||Q_j - \hat{Q}_j||^2 \\ &= \sum_{j=1}^{k^c} 2 - 2\hat{Q}_j' Q_j \\ &= \sum_{j=1}^{k^c} 2\big(1 - \sqrt{1 - sin(\theta_j)^2}\big) = O_p(C_{N,T}^{-2}) \end{aligned}$$

$$||\hat{P}_i - P_i D B^{-1}||^2 = ||(P_i D - \hat{P} B) B^{-1}||^2 \leq ||Q - \hat{Q}||^2 ||B^{-1}||^2 = O_p(C_{N,T}^{-2})$$

Because D and B are orthogonal matrices, DB^{-1} is an orthogonal matrix. Let $H_{r^c} = DB^{-1}$ (iii) is proved.

Proof of (iv):

Let $P_i^* = P_i H_{r^c}$. Obviously P_i^* is a set of eigenvectors of A. Using the proof of (iii) above, we obtain (iv).

□

Lemma 6.3

(i) $||\frac{XX'}{NT}|| = O_p(1)$

(ii) $||\frac{X}{\sqrt{NT}}|| = O_p(1)$

(iii) $||\frac{EE'}{NT}|| = O_p(C_{N,T}^{-2})$

(iv) $||\frac{E\Lambda'}{\sqrt{NT}}|| = O_p(1)$

(v) $\tilde{\Sigma}^p - \Sigma^p = O_p(C_{N,T}^{-1})$ *, with* $\tilde{\Sigma}^p = \frac{1}{T}\sum_{t=1}^T F_t^p F_t^{p'}$

(vi) $\tilde{\Sigma}_i^{*\frac{1}{2}} - I_{r_i} = O_p(C_{N,T}^{-1})$ *, with* $\tilde{\Sigma}_i^* = \frac{1}{T}\sum_{t=1}^T \Sigma_{ii}^{-\frac{1}{2}} F_{i,t} F_{i,t}' \Sigma_{ii}^{-\frac{1}{2}}$

Proof of (i)

Because the ungrouped model (2.9) satisfies the assumptions in Bai and Ng (2002) we can apply Lemma 1 (iii) in Bai and Ng (2002) to X and obtain

$$E\left(\Big|\Big|\frac{XX'}{NT}\Big|\Big|^2\right) = E\left(T^{-2}\sum_{t=1}^T\sum_{\tau=1}^T\Big(\frac{1}{N}\sum_{j=1}^N X_{i,jt}X_{i,j\tau}\Big)^2\right) < M.$$

This implies $||\frac{XX'}{NT}|| = O_p(1)$ and $\frac{XX'}{NT} = O_p(1)$. It follows tr $\left(\frac{XX'}{NT}\right) = O_p(1)$

$$E\left(\Big|\Big|\frac{X}{\sqrt{NT}}\Big|\Big|^2\right) = E\left(\text{ tr }\frac{XX'}{NT}\right) < M.$$

This implies $||\frac{X}{\sqrt{NT}}|| = O_p(1)$.

Proof of (iii)

$$
\begin{aligned}
E\Big|\Big|\frac{EE'}{NT}\Big|\Big|^2 &= \Big(T^{-2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}E\Big(\frac{1}{N}\sum_{j=1}^{N}e_{jt}e_{j\tau}\Big)^2\Big) \\
&= T^{-2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\gamma(t,s)^2 \\
&= T^{-1}\Big(T^{-1}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\gamma(t,s)^2\Big) \\
&\leq T^{-1}M = O_p(C_{N,T}^{-2}).
\end{aligned}
$$

Here we use the fact the $T^{-1}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\gamma(t,s)^2$ is bounded because $T^{-1}\sum_{t=1}^{T}\sum_{\tau=1}^{T}|\gamma(t,s)|$ is bounded by Assumption 2.5 (3). It follows $||\frac{EE'}{NT}|| = O_p(C_{N,T}^{-2})$.

Proof of (iv)

Using Lemma 1 (ii) in Bai and Ng (2002) we have

$$
E\Big|\Big|\frac{E\Lambda'}{\sqrt{NT}}\Big|\Big|^2 = E\Big(T^{-1}\sum_{t=1}^{T}||\frac{1}{\sqrt{N}}e_t'\Lambda||^2\Big) \leq M.
$$

This implies $||\frac{E\Lambda'}{\sqrt{NT}}|| = O_p(1)$

Proof of (v)

According to Assumption 2.4 we have

$$
T^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[F_t^p F_t^{p'} - \Sigma^p] = T^{-\frac{1}{2}}O_p(1) = O_p(C_{N,T}^{-1}).
$$

Proof of (vi)

From proof of (v) it follows

$$
\frac{1}{T}\sum_{t=1}^{T}[F_{i,t}F_{i,t}' - \Sigma_{ii}] = O_p(C_{N,T}^{-1}),
$$

and

$$
\tilde{\Sigma}_i^* - I_{r_i} = \frac{1}{T}\sum_{t=1}^{T}[\Sigma_{ii}^{-\frac{1}{2}}F_{i,t}F_{i,t}'\Sigma_{ii}^{-\frac{1}{2}} - I_{r_i}] = O_p(C_{N,T}^{-1}),
$$

Let $A = \tilde{\Sigma}_i^* - I_{r_i}$ and $Q(\tilde{\Sigma}_i^*)D(\tilde{\Sigma}_i^*)Q(\tilde{\Sigma}_i^*)'$ be the eigendecomposition of $\tilde{\Sigma}_i^*$. We have

$$
\begin{aligned}
\tilde{\Sigma}_i^* - I_{r_i} &= Q(\tilde{\Sigma}_i^*)D(\tilde{\Sigma}_i^*)Q(\tilde{\Sigma}_i^*)' - Q(\tilde{\Sigma}_i^*)Q(\tilde{\Sigma}_i^*)' \\
&= Q(\tilde{\Sigma}_i^*)(D(\tilde{\Sigma}_i^*)^{\frac{2}{2}} - I)Q(\tilde{\Sigma}_i^*)' \\
&= Q(\tilde{\Sigma}_i^*)(D(\tilde{\Sigma}_i^*)^{\frac{1}{2}} - I)(D(\tilde{\Sigma}_i^*)^{\frac{1}{2}} + I)Q(\tilde{\Sigma}_i^*)'
\end{aligned}
$$

It follows

$$
||(D(\tilde{\Sigma}_i^*)^{\frac{1}{2}} - I)|| \leq ||Q(\tilde{\Sigma}_i^*)'||||A||||Q(\tilde{\Sigma}_i^*)||||(D(\tilde{\Sigma}_i^*)^{\frac{1}{2}} + I)^{-1}|| = O_p(C_{N,T}^{-1})
$$

Then we have

$$||\tilde{\Sigma}_i^{*\frac{1}{2}} - I_{r_i}|| \leq ||Q(\tilde{\Sigma}_i^*)(D(\tilde{\Sigma}_i^*)^{\frac{1}{2}} - I)Q(\tilde{\Sigma}_i^*)|| = O_p(C_{N,T}^{-1}).$$

□

Lemma 6.4

Let $\hat{F}_i^* = \sqrt{T}Q_{r_i}$, where Q_{r_i} is the eigenvector corresponding to the r_i largest eigenvalues of the $(T \times T)$ matrix $\frac{X_i X_i'}{N_i T}$. Under Assumptions 2.1 through 2.7, there exists an $(r_i \times r_i)$ matrices $\mathcal{H}_i$ with rank$(\mathcal{H}_i) = r_i$ for $i = 1, 2, ..., n$, and henceforth a $(\sum_{i=1}^n r_i \times \sum_{i=1}^n r_i)$ block diagonal matrix $\mathcal{H} = diag(\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_n)$, such that we have

(i) $\frac{1}{T}\sum_{t=1}^T ||\hat{F}_{i,t}^* - \mathcal{H}_i' F_{i,t}^*||^2 = O_p(C_{N,T}^{-2})$

(ii) $\frac{1}{T}\sum_{t=1}^T ||\hat{F}_t^* - \mathcal{H}' F_t^*||^2 = O_p(C_{N,T}^{-2})$, *with* $F_t^{*'} = (F_{1,t}^{*'}, F_{2,t}^{*'}, ..., F_{n,t}^{*'})$ *and* $\hat{F}_t^{*'} = (\hat{F}_{1,t}^{*'}, \hat{F}_{2,t}^{*'}, ..., \hat{F}_{n,t}^{*'})$,

(iii) There exist an orthogonal matrix H, such that $||\mathcal{H} - H||^2 = O_p(C_{N,T}^{-2})$.

Proof of (i):

Let Q_{r_i} be the eigenvectors corresponding to the r_i largest eigenvalues of $\frac{(X_i X_i')}{N_i T}$. We define a principal component estimate of the group-pervasive factor $\hat{F}_i$ as follows.

$$\hat{F}_i = \left(\frac{X_i X_i'}{N_i T}\right)\sqrt{T}Q_{r_i} = \sqrt{T}Q_{r_i}D_{r_i}, \tag{6.49}$$

where D_{r_i} is a diagonal matrix of the r_i largest eigenvalues of $\left(\frac{X_i X_i'}{N_i T}\right)$. Then we have $D_{r_i}\hat{F}_{i,t}^* = \hat{F}_{i,t}$. Applying Theorem 1 in Bai and Ng (2002) to the data of the ith group, there exits an $(r_i \times r_i)$ matrix $\tilde{\mathcal{H}}_i$, such that

$$\frac{1}{T}\sum_{t=1}^T ||\hat{F}_{i,t} - \tilde{\mathcal{H}}_i' F_{i,t}||^2 = O_p(C_{N,T}^{-2}), \tag{6.50}$$

with $\tilde{\mathcal{H}}_i' = \frac{\hat{F}_i^{*'} F_i}{T}\frac{\Lambda_i \Lambda_i'}{N_i}$.

Let $\mathcal{H}_i'$ denote the $(r_i \times r_i)$ matrix $D_{r_i}^{-1}\tilde{\mathcal{H}}_i'\Sigma_{ii}^{\frac{1}{2}}$

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^T ||\hat{F}_{i,t}^* - \mathcal{H}_i' F_{i,t}^*||^2 \\
= & \frac{1}{T}\sum_{t=1}^T ||D_{r_i}^{-1}D_{r_i}(\hat{F}_{i,t}^* - D_{r_i}^{-1}\tilde{\mathcal{H}}_i'\Sigma_{ii}^{\frac{1}{2}}F_{i,t}^*)||^2 \\
= & \frac{1}{T}\sum_{t=1}^T ||D_{r_i}^{-1}(D_{r_i}\hat{F}_{i,t}^* - \tilde{\mathcal{H}}_i'\Sigma_{ii}^{\frac{1}{2}}F_{i,t}^*)||^2 \\
\leq & \frac{1}{T}\sum_{t=1}^T ||D_{r_i}^{-1}||^2 ||\hat{F}_{i,t} - \tilde{\mathcal{H}}_i' F_t||^2 \\
= & ||D_{r_i}^{-1}||^2 \frac{1}{T}\sum_{t=1}^T ||\hat{F}_{i,t} - \tilde{\mathcal{H}}_i' F_t||^2 \\
\leq & O_p(1)O_p(C_{N,T}^{-2}) = O_p(C_{N,T}^{-2})
\end{aligned}$$

Proof of (ii):

Since the probability limit above holds for all groups, stacking all groups together we have

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^* - \mathcal{H}'F_t^*||^2 = O_p(C_{N,T}^{-2}).$$

Proof of (iii):

Let A be a positive semi definite matrix with different positive eigenvalues and $Q(A)$ and $D(A)$ denote the eigenvector and the eigenvalues of matrix A respectively. According to Lemma 6.2, we have

$$\begin{aligned}
Q\left(\frac{X_iX_i'}{N_iT}\right) &= Q\left(\frac{F_i\Lambda_i\Lambda_i'F_i'}{N_iT} + \frac{E_i\Lambda_i'F_i'}{N_iT} + \frac{F_i\Lambda_i\Lambda_i'E_i}{N_iT} + \frac{E_iE_i}{N_iT}\right) \\
&= Q\left(\frac{F_i\Lambda_i\Lambda_i'F_i'}{N_iT}\right) + O_p(C_{N,T}^{-1}) \\
&= Q\left(\frac{F_i^*}{\sqrt{T}}\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}\frac{F_i^{*'}}{\sqrt{T}}\right) + O_p(C_{N,T}^{-1}) \\
&= Q\left(\frac{F_i^*}{\sqrt{T}}\tilde{\Sigma}_i^{*-\frac{1}{2}}\tilde{\Sigma}_i^{*\frac{1}{2}}\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}\tilde{\Sigma}_i^{*\frac{1}{2}}\tilde{\Sigma}_i^{*-\frac{1}{2}}\frac{F_i^{*'}}{\sqrt{T}}\right) + O_p(C_{N,T}^{-1}) \\
&= \frac{F_i^*}{\sqrt{T}}\tilde{\Sigma}_i^{*-\frac{1}{2}}Q\left((\tilde{\Sigma}_i^{*\frac{1}{2}} - I + I)\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}(\tilde{\Sigma}_i^{*\frac{1}{2}} - I + I)\right) + O_p(C_{N,T}^{-1}) \\
&= \frac{F_i^*}{\sqrt{T}}\tilde{\Sigma}_i^{*-\frac{1}{2}}Q\left(\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}\right) + O_p(1)O_p(C_{N,T}^{-1}) + O_p(C_{N,T}^{-1}) \\
&= \frac{F_i^*}{\sqrt{T}}\tilde{\Sigma}_i^{*-\frac{1}{2}}P_i^* + O_p(C_{N,T}^{-1}),
\end{aligned}$$

where P_i^* is the eigenvector matrix of $\left(\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}\right)$ and $\tilde{\Sigma}_i^* = \frac{F_i^{*'}F_i^*}{T}$. From the fifth row to sixth row we have used Lemma 6.3 (vi). In the first row of the equation above we have used

$$\begin{aligned}
&\left|\left|\frac{E_i\Lambda_i'F_i'}{N_iT} + \frac{F_i\Lambda E_i'}{N_iT} + \frac{E_iE_i}{N_iT}\right|\right| \\
= &\left|\left|\frac{E_i\Lambda_i'}{\sqrt{NT}}\frac{F_i'}{\sqrt{T}}\frac{1}{\sqrt{T}}\right|\right| + \left|\left|\frac{F_i'}{\sqrt{T}}\frac{\Lambda_iE_i'}{\sqrt{NT}}\frac{1}{\sqrt{T}}\right|\right| + \left|\left|\frac{E_iE_i}{N_iT}\right|\right| \\
\leq &\left|\left|\frac{E_i\Lambda_i'}{\sqrt{NT}}\right|\right|\left|\left|\frac{F_i'}{\sqrt{T}}\right|\right|\frac{1}{\sqrt{T}} + \left|\left|\frac{F_i'}{\sqrt{T}}\right|\right|\left|\left|\frac{\Lambda_iE_i'}{\sqrt{NT}}\right|\right|\frac{1}{\sqrt{T}} + \left|\left|\frac{E_iE_i}{N_iT}\right|\right| = O_p(C_{N,T}^{-1})
\end{aligned}$$

The equation above is based on Lemma 6.3 (ii) (iii) and (iv). Similarly we can calculate the eigenvalues of $\left(\frac{X_i X_i'}{N_i T}\right)$.

$$
\begin{aligned}
D\left(\frac{X_i X_i'}{N_i T}\right) &= D\left(\frac{F_i\Lambda_i\Lambda_i'F_i'}{N_iT} + \frac{E_i\Lambda_i'F_i'}{N_iT} + \frac{F_i\Lambda_i\Lambda_i'E_i}{N_iT} + \frac{E_iE_i}{N_iT}\right) \\
&= D\left(\frac{F_i\Lambda_i\Lambda_i'F_i'}{N_iT}\right) + O_p(C_{N,T}^{-1}) \\
&= D\left(\frac{F_i^*}{\sqrt{T}}\tilde{\Sigma}_i^{*-\frac{1}{2}}\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}\tilde{\Sigma}_i^{*-\frac{1}{2}}\frac{F_i^{*'}}{\sqrt{T}}\right) + O_p(1)O_p(C_{N,T}^{-1}) + O_p(C_{N,T}^{-1}) \\
&= D\left(\frac{F_i^*}{\sqrt{T}}\tilde{\Sigma}_i^{*-\frac{1}{2}}P_i^*D_{r_i}^*P_i^{*'}\tilde{\Sigma}_i^{*-\frac{1}{2}}\frac{F_i^{*'}}{\sqrt{T}}\right) + O_p(C_{N,T}^{-1}) = D_{r_i}^* + O_p(C_{N,T}^{-1}).
\end{aligned}
$$

where $D_{r_i}^*$ is the diagonal matrix the of the eigenvalues of $\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}$, i.e. $\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N} = P_i^*D_{r_i}^*P_i^{*'}$. In the last two rows of the equations above we have used the fact the $P_i^*D_i^*P_i^{*'}$ is the eigen-decomposition of $\frac{\Sigma_i^{\frac{1}{2}}\Lambda_i\Lambda_i'\Sigma_i^{\frac{1}{2}}}{N}$ and $\frac{F_i^*}{\sqrt{T}}\tilde{\Sigma}_i^{*-\frac{1}{2}}P_i^*D_i^*P_i^{*'}\tilde{\Sigma}_i^{*-\frac{1}{2}}\frac{F_i^{*'}}{\sqrt{T}}$ is itself in the eigen-decomposition form. Inserting the equation above into the expression of $\tilde{\mathcal{H}}_i$, we have

$$
\begin{aligned}
\tilde{\mathcal{H}}_i &= \frac{\hat{F}_i^{*'}F_i}{T}\frac{\Lambda_i\Lambda_i'}{N_i} \\
&= \frac{Q_{r_i}'F_i}{\sqrt{T}}\frac{\Lambda_i\Lambda_i'}{N} \\
&= \frac{P_i^{*'}\tilde{\Sigma}_i^{*-\frac{1}{2}}F_i^{*'}F_i}{T}\frac{\Lambda_i\Lambda_i'}{N} + O_p(C_{N,T}^{-1})\frac{F_i}{\sqrt{T}}\frac{\Lambda_i\Lambda_i'}{N} \\
&= P_i^{*'}\tilde{\Sigma}_i^{*-\frac{1}{2}}\frac{F_i^{*'}F_i^*}{T}\Sigma_i^{\frac{1}{2}}\frac{\Lambda_i\Lambda_i'}{N}\Sigma_i^{\frac{1}{2}}\Sigma_i^{-\frac{1}{2}} + O_p(C_{N,T}^{-1})O_p(1) \\
&= P_i^{*'}(\tilde{\Sigma}_i^{*\frac{1}{2}} - I + I)\Sigma_i^{\frac{1}{2}}\frac{\Lambda_i\Lambda_i'}{N}\Sigma_i^{\frac{1}{2}}\Sigma_i^{-\frac{1}{2}} + O_p(C_{N,T}^{-1})O_p(1) \\
&= D_{r_i}^*P_i^{*'}\Sigma_i^{-\frac{1}{2}} + O_p(C_{N,T}^{-1})O_p(1),
\end{aligned}
$$

By the definition of $\mathcal{H}_i = D_{r_i}^{-1}\tilde{\mathcal{H}}_i\Sigma_{ii}^{\frac{1}{2}}$ we have

$$
\mathcal{H}_i' = D_{r_i}^{-1}D_{r_i}^*P_i^{*'}\Sigma_i^{-\frac{1}{2}}\Sigma_i^{\frac{1}{2}} + O_p(C_{N,T}^{-1}) = P_i^{*'} + O_p(C_{N,T}^{-1})
$$

In the previous equation we have used the result $D_{r_i} = D_{r_i}^* + O_p(C_{N,T}^{-1})$ shown above. Because P_i^* is the eigenvector matrix of a positive definite matrix, it is orthogonal. Define $H = \text{diag}(P_1^*, P_2^*, ..., P_n^*)$, we have

$$
||\mathcal{H} - H|| = ||\text{diag}(\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_n) - \text{diag}(P_1^*, P_2^*, ..., P_n^*)|| < \sum_{i=1}^{n}||\mathcal{H}_i - P_i^*|| = O_p(C_{N,T}^{-1}).
$$

This proves (iii). □

Lemma 6.5

Let $\hat{\Sigma}_{ij}^* = \frac{1}{T}\sum_{t=1}^T \hat{F}_{i,t}^*\hat{F}_{j,t}^{*'}$ *and* $\Sigma_{ij}^{\mathcal{H}*} = \mathcal{H}_i'E(F_{i,t}^*F_{j,t}^{*'})\mathcal{H}_j = \mathcal{H}_i\Sigma_{ij}^*\mathcal{H}_j'$, *where* $\hat{F}_{i,t}^*$ *is the* *estimate of the standardized group-pervasive factors of the group i satisfying: $\frac{1}{T}\sum_{t=1}^{T}||\hat{F}^*_{i,t}-\mathcal{H}'_iF^*_{i,t}||^2 = O_p(C_{N,T}^{-2})$ and $\mathcal{H}_i$ is an $(r_i\times r_i)$ rotation matrix as defined in Lemma 6.4. Further, let $\Sigma^* = E(F^*_tF^{*\prime}_t)$, $\hat{\Sigma}^* = \frac{1}{T}\sum_{t=1}^{T}\hat{F}^*_t\hat{F}^{*\prime}_t$, and $\Sigma^{\mathcal{H}*} = \mathcal{H}'\Sigma^*\mathcal{H}$, where $\mathcal{H}$ is the $(\sum_{i=1}^{n}r_i\times\sum_{i=1}^{n}r_i)$ matrix defined in Lemma 6.4. Then we have*

(i) $\left|\left|\hat{\Sigma}^*_{ij}-\Sigma^{\mathcal{H}*}_{ij}\right|\right|^2 = O_p(C_{N,T}^{-2})$, *and*

(ii) $\left|\left|\hat{\Sigma}^*-\Sigma^{\mathcal{H}*}\right|\right|^2 = O_p(C_{N,T}^{-2})$.

(iii) Let $\hat{h}_{k^c}$ be eigenvectors corresponding to the k^c largest eigenvalues of $\hat{\Sigma}^*$. Then there exist an $(r^c\times r^c)$ orthogonal matrix H_{r^c} and a $(\sum_{i=1}^{n}r_i\times\sum_{i=1}^{n}r_i)$ orthogonal matrix H, such that

- *for* $k^c = r^c$, $||\hat{h}_{r^c}-H'a_{r^c}H_{r^c}||^2 = O_p(C_{N,T}^{-2})$,
- *for* $k^c < r^c$, $||\hat{h}_{k^c}-H'a_{r^c}H_{rk^c}||^2 = O_p(C_{N,T}^{-2})$,
- *for* $r^c < k^c \leq K$, $\left|\left|\hat{h}_{k^c}-H'a_{k^c}H_{k^c}\right|\right|^2 = O_p(C_{N,T}^{-2})$.

where H_{rk^c} is the first k^c columns of H_{r^c} and $S_{k^c-r^c}$ is a diagonal matrix with either +1 or -1 on the diagonal.

Proof of (i):

$$
\begin{aligned}
& \parallel \hat{\Sigma}^*_{ij}-\mathcal{H}'_i\Sigma^*_{ij}\mathcal{H}_j \parallel \\
= \; & \parallel \frac{1}{T}\sum_{t=1}^{T}\hat{F}^*_{i,t}\hat{F}^{*\prime}_{j,t}-\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}'_iF^*_{i,t}F^{*\prime}_{j,t}\mathcal{H}_j+\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}'_iF^*_{i,t}F^{*\prime}_{j,t}\mathcal{H}_j-\mathcal{H}'_i\Sigma^*_{ij}\mathcal{H}_j \parallel \\
\leq \; & \parallel \frac{1}{T}\sum_{t=1}^{T}(\hat{F}^*_{i,t}-\mathcal{H}'_iF^*_{i,t}+\mathcal{H}'_iF^*_{i,t})(\hat{F}^*_{j,t}-\mathcal{H}'_jF^*_{j,t}+\mathcal{H}'_jF^*_{j,t})'-\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}'_iF^*_{i,t}F^{*\prime}_{j,t}\mathcal{H}_j \parallel \\
& +\parallel \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}'_iF^*_{i,t}F^{*\prime}_{j,t}\mathcal{H}_j-\mathcal{H}'_i\Sigma^*_{ij}\mathcal{H}_j \parallel \\
= \; & \parallel \frac{1}{T}\sum_{t=1}^{T}(\hat{F}^*_{i,t}-\mathcal{H}'_iF^*_{i,t})\hat{F}^{*\prime}_{j,t}+\mathcal{H}'_iF^*_{i,t}(\hat{F}_{j,t}-\mathcal{H}'_jF^*_{j,t})' \parallel + \parallel \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}'_iF^*_{i,t}F^{*\prime}_{j,t}\mathcal{H}_j-\mathcal{H}'_i\Sigma_{ij}\mathcal{H}_j \parallel \\
\leq \; & \parallel \frac{1}{T}\sum_{t=1}^{T}(\hat{F}^*_{i,t}-\mathcal{H}'_iF^*_{i,t})\hat{F}^{*\prime}_{j,t} \parallel + \parallel \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}'_iF^*_{i,t}(\hat{F}^*_{j,t}-\mathcal{H}'_jF^*_{j,t})' \parallel \\
& +\parallel \frac{1}{\sqrt{T}}\mathcal{H}'_i\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F^*_{i,t}F^{*\prime}_{j,t}-E(F^*_{i,t}F^{*\prime}_{j,t})\right)\mathcal{H}_j \parallel \\
\leq \; & \left(\frac{1}{T}\sum_{t=1}^{T}\parallel \hat{F}^*_{i,t}-\mathcal{H}'_iF^*_{i,t}\parallel^2\frac{1}{T}\sum_{t=1}^{T}\parallel \hat{F}^*_{j,t}\parallel^2\right)^{\frac{1}{2}}+\left(\frac{1}{T}\sum_{t=1}^{T}\parallel \mathcal{H}'_iF^*_{i,t}\parallel^2\frac{1}{T}\sum_{t=1}^{T}\parallel \hat{F}_{j,t}-\mathcal{H}'_jF^*_{j,t}\parallel^2\right)^{\frac{1}{2}} \\
& +\frac{1}{T}\parallel \mathcal{H}'_i\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(F^*_{i,t}F^{*\prime}_{j,t}-E(F^*_{i,t}F^{*\prime}_{j,t})\right)\right)\mathcal{H}_j \parallel \\
= \; & O_p(C_{N,T}^{-1})O_p(1)+O_p(C_{N,T}^{-1})O_p(1)+O_p(C_{N,T}^{-1})O_p(1)
\end{aligned}
$$

It follows $\left|\left|\hat{\Sigma}^*_{ij} - \Sigma^{\mathcal{H}*}_{ij}\right|\right|^2 = O_p(C^{-2}_{N,T})$. This proves (i). Collecting $\hat{\Sigma}^*_{ij}$ and $\Sigma^{\mathcal{H}*}_{ij}$ over i and j into Σ^* and $\Sigma^{\mathcal{H}*}$ matrices respectively, we obtain $\left|\left|\hat{\Sigma}^* - \Sigma^{\mathcal{H}*}\right|\right|^2 = O_p(C^{-2}_{N,T})$. (ii) is proved.

Proof of (iii): Let $\hat{h}_{k^c}$ be a matrix containing k^c eigenvectors corresponding to the k^c largest eigenvalues of $\hat{\Sigma}^*$. By (ii) and Lemma 6.2, there exists an eigenvector matrix $h^{\mathcal{H}}_{k^c}$ corresponding to the k^c largest eigenvalues of $\Sigma^{\mathcal{H}*}$, such that $||\hat{h}_{k^c} - h^{\mathcal{H}}_{k^c}|| = O_p(C^{-1}_{N,T})$. From Lemma 6.4 (iii) we have $||\mathcal{H}'\Sigma^*\mathcal{H} - H'\Sigma^* H|| = O_p(C^{-1}_{N,T})$. Then there exists an eigenvector matrix $h^{H}_{k^c}$ corresponding to the k^c largest eigenvalues of $H'\Sigma^* H$, such that $||h^{\mathcal{H}}_{k^c} - h^{H}_{k^c}|| = O_p(C^{-1}_{N,T})$. Because H is an orthogonal matrix, $H'\Sigma^* H = H' h_K D^*_K h'_K H$ is an eigen-decomposition of $H'\Sigma^* H$, where $h_K D^*_K h_K$ is the eigen-decomposition of Σ^*. Therefore we have $h^{H}_{k^c} = H' h_{k^c}$ with h_{k^c} the eigenvector matrix corresponding to the k^c largest eigenvalues of Σ^*. It follows

$$\begin{aligned}
||\hat{h}_{k^c} - H' h_{k^c}|| &= ||\hat{h}_{k^c} - h^{\mathcal{H}}_{k^c} + h^{\mathcal{H}}_{k^c} - h^{H}_{k^c} + h^{H}_{k^c} - H' h_{k^c}|| \\
&\leq ||\hat{h}_{k^c} - h^{\mathcal{H}}_{k^c}|| + ||h^{\mathcal{H}}_{k^c} - h^{H}_{k^c}|| = O_p(C^{-1}_{N,T}).
\end{aligned}$$

Now for the given $\hat{h}_{k^c}$ above, following Lemma 6.1 (vi), there exists an $(r^c \times r^c)$ orthogonal matrix H_{r^c}, and a diagonal matrix $S_{k^s - r^c}$, such that

- for $k^c = r^c$, $||\hat{h}_{r^c} - H' a_{r^c} H_{r^c}||^2 = O_p(C^{-2}_{N,T})$,
- for $k^c < r^c$, $||\hat{h}_{k^c} - H' a_{r^c} H_{rk^c}||^2 = O_p(C^{-2}_{N,T})$,
- for $r^c < k^c \leq K$, $||\hat{h}_{k^c} - H' a_{k^c} H_{k^c}||^2 = O_p(C^{-2}_{N,T})$, with $H_{k^c} = \mathrm{diag}(H_{r^c}, S_{k^s - r^c})$.

This proved (iii). □

Proof of Theorem 3.3

Let $\hat{h}_{k^c}$ be the eigenvectors corresponding to the k^c largest eigenvalues of $\hat{\Sigma}^*$.

For $r^c < k^c \leq K$, let $B'_{k^c} = H'_{k^c}\begin{pmatrix} \Sigma^{c-\frac{1}{2}} & 0 \\ 0 & I \end{pmatrix}$, we have

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\left|\left|\hat{F}^c_t - B'_{k^c}\begin{pmatrix} F^c_t \\ F^{ps}_t \end{pmatrix}\right|\right|^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{n}}(\hat{h}_{k^c} - H' a_{k^c} H_{k^c} + H' a_{k^c} H_{k^c})'(\hat{F}^*_t - H' F^*_t + H' F^*_t) - H'_{k^c}\begin{pmatrix} F^{c*}_t \\ F^{ps*}_t \end{pmatrix}\right|\right|^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{n}}(\hat{h}_{k^c} - H' a_{k^c} H_{k^c})'\hat{F}^*_t + \frac{1}{\sqrt{n}} H'_{k^c} a'_{k^c} H(\hat{F}^*_t - H' F^*_t) + \frac{1}{\sqrt{n}} H'_{k^c} a'_{k^c} F^*_t - H'_{k^c}\begin{pmatrix} F^{c*}_t \\ F^{ps*}_t \end{pmatrix}\right|\right|^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{n}}(\hat{h}_{k^c} - H' a_{k^c} H_{k^c})'\hat{F}^*_t + \frac{1}{\sqrt{n}} H'_{k^c} a'_{k^c} H(\hat{F}^*_t - H' F^*_t)\right|\right|^2 \\
\leq & \frac{1}{n}\frac{1}{T}\sum_{t=1}^{T}||(\hat{h}_{k^c} - H' a_{k^c} H_{k^c})'||^2||\hat{F}^*_t||^2 + \frac{1}{n}\frac{1}{T}\sum_{t=1}^{T}||H'_{k^c} a'_{k^c} H||^2||(\hat{F}^*_t - H' F^*_t)||^2 \\
& + \frac{2}{n}\frac{1}{T}\sum_{t=1}^{T}||(\hat{h}_{k^c} - H' a_{k^c} H_{k^c})'\hat{F}^*_t||||H'_{k^c} a'_{k^c} H(\hat{F}^*_t - H' F^*_t)||
\end{aligned}$$

$$
\begin{aligned}
\leq & \frac{1}{n}||(\hat{h}_{k^c} - H'a_{k^c}H_{k^c})'||^2 \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^*||^2 + \frac{1}{n}||H'_{k^c}a'_{k^c}H||^2 \frac{1}{T}\sum_{t=1}^{T}||(\hat{F}_t^* - H'F_t^*)||^2 \\
& + \frac{2}{n}\left(||(\hat{h}_{k^c} - H'a_{k^c}H_{k^c})'||^2 \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^*||^2 ||H'_{k^c}a'_{k^c}H||^2 \frac{1}{T}\sum_{t=1}^{T}||(\hat{F}_t^* - H'F_t^*)||^2\right)^{\frac{1}{2}} \\
= & O_p(C_{N,T}^{-2})O_p(1) + O_p(1)O_p(C_{N,T}^{-2}) + 2(O_p(C_{N,T}^{-2})O_p(C_{N,T}^{-2}))^{\frac{1}{2}} = O_p(C_{N,T}^{-2})
\end{aligned}
$$

At the third equality we have used the relationship: $\frac{1}{\sqrt{n}}a'_{k^c}F_t^* = \begin{pmatrix} F_t^{c*} \\ F_t^{ps*} \end{pmatrix}$ according to Lemma 6.1. The probability limit at the last row follows from Lemma 6.5 $||(\hat{h}_{k^c} - H'a_{k^c}H_{k^c})'||^2 = O_p(C_{N,T}^{-2})$, Lemma 6.4: $\frac{1}{T}\sum_{t=1}^{T}||(\hat{F}_t^* - H'F_t^*)||^2 = O_p(C_{N,T}^{-2})$ as well as the Cauchy-Schwarz inequality.

For the cases of $k^c \leq r^c$, let $B'_{rk^c} = H'_{rk^c}\Sigma^{c-\frac{1}{2}}$, then we have

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\left|\left|\hat{F}_t^c - B'_{rk^c}F_t^c\right|\right|^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{n}}(\hat{h}_{k^c} - H'a_{r^c}H_{rk^c} + H'a_{r^c}H_{rk^c})'(\hat{F}_t^* - H'F_t^* + H'F_t^*) - H'_{rk^c}F_t^{c*}\right|\right|^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{n}}(\hat{h}_{k^c} - H'a_{r^c}H_{rk^c})'\hat{F}_t^* + \frac{1}{\sqrt{n}}H'_{rk^c}a'_{k^c}H(\hat{F}_t^* - H'F_t^*) + \frac{1}{\sqrt{n}}H'_{rk^c}a'_{r^c}F_t^* - H'_{rk^c}F_t^{c*}\right|\right|^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{n}}(\hat{h}_{k^c} - H'a_{r^c}H_{rk^c})'\hat{F}_t^* + \frac{1}{\sqrt{n}}H'_{rk^c}a'_{r^c}H(\hat{F}_t^* - H'F_t^*)\right|\right|^2 \\
\leq & \frac{1}{n}\frac{1}{T}\sum_{t=1}^{T}||(\hat{h}_{k^c} - H'a_{r^c}H_{rk^c})'||^2||\hat{F}_t^*||^2 + \frac{1}{n}\frac{1}{T}\sum_{t=1}^{T}||H'_{rk^c}a'_{r^c}H||^2||(\hat{F}_t^* - H'F_t^*)||^2 \\
& + \frac{2}{n}\frac{1}{T}\sum_{t=1}^{T}||(\hat{h}_{k^c} - H'a_{r^c}H_{rk^c})'\hat{F}_t^*||||H'_{rk^c}a'_{r^c}H(\hat{F}_t^* - H'F_t^*)|| \\
\leq & \frac{1}{n}||(\hat{h}_{k^c} - H'a_{r^c}H_{rk^c})'||^2 \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^*||^2 + \frac{1}{n}||H'_{rk^c}a'_{r^c}H||^2 \frac{1}{T}\sum_{t=1}^{T}||(\hat{F}_t^* - H'F_t^*)||^2 \\
& + \frac{2}{n}\left(||(\hat{h}_{k^c} - H'a_{r^c}H_{rk^c})'||^2 \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^*||^2||H'_{rk^c}a'_{r^c}H||^2 \frac{1}{T}\sum_{t=1}^{T}||(\hat{F}_t^* - H'F_t^*)||^2\right)^{\frac{1}{2}} \\
= & O_p(C_{N,T}^{-2})O_p(1) + O_p(1)O_p(C_{N,T}^{-2}) + 2(O_p(C_{N,T}^{-2})O_p(C_{N,T}^{-2}))^{\frac{1}{2}} = O_p(C_{N,T}^{-2})
\end{aligned}
$$

□

Proof of **Proposition 3.4**

Proof of (i) and (ii):

Following Theorem 3.3 for $k^c \leq r^c$, there exits an $(r^c \times k^c)$ matrix B_{rk^c} such that $\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^c - B'_{rk^c}F_t^c||^2 = O_p(C_{N,T}^{-2})$. We rotate the the common factors into the direction of B_{rk^c} and $B_{rk^c\perp}$, where $B_{rk^c\perp}$ is an $(r^c \times (r^c - k^c))$ matrix orthogonal to B_{k^c}, such that $B_{r^c} = (B_{rk^c}, B_{rk^c\perp})$ constitutes a full rank matrix.

$$
\begin{aligned}
X_{i,jt} &= F_t^{c'}(B_{rk^c}, B_{rk^c\perp})(B_{rk^c}, B_{rk^c\perp})^{-1}\lambda_{i,j}^c + F_{i,t}^{s'}\lambda_{i,j}^s + e_{i,jt} \\
&= ((B'_{rk^c}F_t^c)', (B'_{rk^c\perp}F_t^c)')(B_{rk^c}, B_{rk^c\perp})^{-1}\lambda_{i,j}^c + F_{i,t}^{s'}\lambda_{i,j}^s + e_{i,jt} \\
&= ((B'_{rk^c}F_t^c)', (B'_{rk^c\perp}F_t^c)')\lambda_{i,j}^* + F_{i,t}^{s'}\lambda_{i,j}^s + e_{i,jt} \\
&= ((B'_{rk^c}F_t^c)', (B'_{rk^c\perp}F_t^c)')\begin{pmatrix} \lambda_{i,j,rk^c}^* \\ \lambda_{i,j,rk^c\perp}^* \end{pmatrix} + F_{i,t}^{s'}\lambda_{i,j}^s + e_{i,jt} \\
&= (B'_{rk^c}F_t^c)'\lambda_{i,j,rk^c}^* + (B'_{rk^c\perp}F_t^c)'\lambda_{i,j,rk^c\perp}^* + F_{i,t}^{s'}\lambda_{i,j}^s + e_{i,jt}
\end{aligned}
$$

with $\lambda_{i,j}^* = (B_{rk^c}, B_{rk^c\perp})^{-1}\lambda_{i,j}^c$ and $\begin{pmatrix} \lambda_{i,j,rk^c}^* \\ \lambda_{i,j,rk^c\perp}^* \end{pmatrix} = \lambda_{i,j}^*$ is a decomposition of $\lambda_{i,j}^*$ into the first k^c elements and the rest $r^c - k^c$ elements. We define for $k^c < r^c$

$$
F_t^{pc} = B'_{rk^c}F_t^c\ ,\ F_t^{rs} = \begin{pmatrix} B'_{rk^c\perp}F_t^c \\ F_{i,t}^s \end{pmatrix},\ \lambda_{i,j}^{pc} = \lambda_{i,j,rk^c}^* \text{ and } \lambda_{i,j}^{rs} = \begin{pmatrix} \lambda_{i,j,rk^c\perp}^* \\ \lambda_{i,j}^s \end{pmatrix}, \tag{6.51}
$$

then we have

$$
X_{i,jt} = F_t^{pc'}\lambda_{i,j}^{pc} + F_{i,t}^{rs'}\lambda_{i,j}^{rs} + e_{i,jt}.
$$

Let $X_{i,jt}^{rs} = X_{i,jt} - F_t^{pc'}\lambda_{i,j}^{pc}$. Obviously, $X_{i,jt}^{rs}$ is generate by a factor model:

$$
X_{i,jt}^{rs} = F_{i,t}^{rs'}\lambda_{i,j}^{rs} + e_{i,jt}. \tag{6.52}
$$

Because $B'_{rk^c\perp}F_t^c$ and $F_{i,t}^s$ are uncorrelated, the number of factors represented by F_t^{rs} is $(r^c - k^c + r_i^s)$. It is to note that although equation (6.52) presents a factor model, some factors represented by $B'_{rk^c\perp}F_t^c$ do not satisfy our model assumptions on factors. In fact $B_{rk^c\perp}$ changes with the direction of the factor estimate $\hat{F}^c$ and has no convergency. Therefore, some standard results on factor estimation cannot be directly applied to model (6.52) for the case of $k^c < r^c$.

For the case of $k^c = r^c$, B_{rk^c} becomes the $(r^c \times r^c)$ squared full rank matrix B_{r^c} and $B_{rk^c\perp}$ is empty. Therefore, we can define $F_t^{pc} = B'_{r^c}F_t^c$ and equation (6.52) becomes

$$
X_{i,jt}^{rs} = X_{i,jt}^s = F_{i,t}^{s'}\lambda_{i,j}^s + e_{i,jt}. \tag{6.53}
$$

In this case the number of factors generating X_i^s is r_i^s. Obviously the factor model in (6.53) satisfies our Assumptions 2.4 and 2.7 on factors.

For the case of $k^c > r^c$, following Proposition 3.1 we know that $\hat{F}_t^c$ will span the common factor space and a subspace of the specific factor space represented by F_t^{ps}. Hence, we want to decompose $F_{i,t}^s$ into two components: one is orthogonal to F_t^{ps} and the other one is linearly perfect correlated with F_t^{ps}, in order to find the population counterpart of $\hat{F}_t^c$. According to the definition of F^{ps} we have

$$
F^{ps} = \sum_i^n a_i^{ps'}F_{i,t}^s = \sum_i^n a_i^{ps'}C_i^{s'}F_t^s = (\sum_i^n a_i^{ps'}C_i^{s'})F_t^s = C^{ps'}F^s, \tag{6.54}
$$

with $C^{ps'} = \sum_i^n a_i^{ps'} C_i^{s'}$. So we have $E(F_{i,t}^s F_t^{ps'}) = C_i^{s\prime}\Sigma^s C^{ps}$, with $\Sigma^s = E(F_t^s F_t^{s'})$, $E(F_t^{ps} F_t^{ps'}) = C^{ps\prime}\Sigma^s C^{ps}$. Now we decompose $F_{i,t}^s$ as follows:

$$
\begin{aligned}
F_{i,t}^s &= F_{i,t}^s - E(F_{i,t}^s F_t^{ps'})(E(F_i^{ps} F_t^{ps'}))^{-1} F_t^{ps} + E(F_{i,t}^s F_t^{ps'})(E(F_i^{ps} F_t^{ps'}))^{-1} F_t^{ps} \\
&= C_i^{s\prime} F_t^s - C_i^{s\prime}\Sigma^s C^{ps}(C^{ps\prime}\Sigma^s C^{ps}))^{-1} C^{ps\prime} F_t^s + E(F_{i,t}^s F_t^{ps'})(E(F_i^{ps} F_t^{ps'}))^{-1} F_t^{ps} \\
&= C_i^{s\prime}\Sigma^{s\frac{1}{2}} F_t^{s*} - C_i^{s\prime}\Sigma^{s\frac{1}{2}}\Sigma^{s\frac{1}{2}} C^{ps}(C^{ps\prime}\Sigma^{s\frac{1}{2}}\Sigma^{s\frac{1}{2}} C^{ps}))^{-1} C^{ps\prime}\Sigma^{s\frac{1}{2}} F_t^{s*} \\
&\quad + E(F_{i,t}^s F_t^{ps'})(E(F_i^{ps} F_t^{ps'}))^{-1} F_t^{ps} \\
&= B_i^{s\prime} F_t^{s*} - (B_i^{s'} B^{ps})(B^{ps'} B^{ps})^{-1} B^{ps'} F_t^{s*} + (B_i^{s'} B^{ps})(B^{ps'} B^{ps})^{-1} B^{ps'} F_t^{s*} \\
&= B_i^{s\prime} M_{B^{ps}} F_t^{s*} + B_i^{s'} P_{B^{ps}} F_t^{s*} \\
&= F_{i,t}^{rs} + B_i^{s\prime} B^{ps}(B^{ps'} B^{ps})^{-1} F_t^{ps},
\end{aligned}
$$

with $B_i^s = \Sigma^{s\frac{1}{2}} C_i^s$, $B^{ps} = \Sigma^{s\frac{1}{2}} C^{ps}$ and $F_{i,t}^{rs} = B_i^{s\prime} M_{B^{ps}} F_t^{s*}$. Note that $F_{i,t}^{rs}$ is uncorrelated with F_t^{ps} by construction. Inserting the decomposition above into the model, the common components of returns in group i can be separated into three parts:

$$
\begin{aligned}
X_{i,jt} &= F_t^{c'}\lambda_{i,j}^c + (B_i^{s\prime} B^{ps}(B^{ps'} B^{ps})^{-1} F_t^{ps} + F_{i,t}^{rs})'\lambda_{i,j}^s + e_{i,jt} \\
&= F_t^{c'}\lambda_{i,j}^c + F_t^{ps'}(B^{ps'} B^{ps})^{-1} B^{ps'} B_i^s \lambda_{i,j}^s + F_{i,t}^{rs\prime}\lambda_{i,j}^s + e_{i,jt} \\
&= (F_t^{c\prime}, F_t^{ps\prime}) \begin{pmatrix} \lambda_{i,j}^c \\ \lambda_{i,j}^{ps} \end{pmatrix} + F_{i,t}^{rs\prime}\lambda_{i,j}^s + E_i,
\end{aligned}
$$

with $\lambda_{i,j}^{ps} = (B^{ps'} B^{ps})^{-1} B^{ps'} B_i^s \lambda_{i,j}^s$. We define for $k^c > r^c$,

$$
F_t^{pc'} = (F_t^{c\prime}, F_t^{ps\prime}) \ , \ \lambda_{i,j}^{pc} = \begin{pmatrix} \lambda_{i,j}^c \\ \lambda_{i,j}^{ps} \end{pmatrix} \text{ and } F_{i,t}^{rs} = B_i^{s\prime} M_{B^{ps}} F_t^{s*} \ , \ \lambda_{i,j}^{rs} = \lambda_{i,j}^s. \tag{6.55}
$$

In terms of F_t^{pc} and F_t^{rs}, we have

$$
X_{i,jt} = F_t^{pc'}\lambda_{i,j}^{pc} + F_{i,t}^{rs'}\lambda_{i,j}^{rs} + e_{i,jt}.
$$

Again we have

$$
X_{i,jt}^{rs} = X_{i,jt} - F_t^{pc'}\lambda_{i,j}^{pc} = F_{i,t}^{rs'}\lambda_{i,j}^{rs} + e_{i,jt} \text{ with } F_t^{rs} = C^{rs\prime} F_t^s \tag{6.56}
$$

where $C^{rs} = B_i^{s\prime} M_{B^{ps}} \Sigma^{s-\frac{1}{2}}$ is an $(r_i^s \times K^s)$ constant matrix. Obviously. The factor model in (6.56) satisfies our Assumptions 2.4 and 2.7 on factors. The number of factors represented by F_t^{rs} depends on the rank of C^{rs}.

In usual cases it is r_i^s, i.e. $X_{i,jt}^{rs}$ is driven by r_i^s factors. In some special cases in which there exist some exact linear dependence between F_t^{ps} and $F_{i,t}^s$, the rank of C^{rs} will be reduced and consequently the number of factors driving $X_{i,jt}^s$ will also be reduced.

If there exits r_i^{s*} linearly dependent relations between F_t^{ps} and $F_{i,t}^s$, we have $F_{i,t}^{s'} C = F_t^{ps\prime} B$ where C is an $(r_i^s \times r_i^{s*})$ with $rank(C) = r_i^{s*}$. The coefficient matrix B can be calculated as follows.

$$
\begin{aligned}
& F_{i,t}^{s'} C = F_t^{ps'} B \\
= \ & F_t^{ps'}(E(F_t^{ps} F_t^{ps'}))^{-1} E(F_t^{ps} F_t^{s'} C) \\
= \ & F_t^{ps'}(B^{ps\prime} B^{ps})^{-1} B^{ps\prime} B_i^s C
\end{aligned}
$$

Corresponding to C we can find a complementary $(r_i^s \times (r_i^s - r_i^{s*}))$ matrix $C_\perp$ such that $C'C_\perp = 0$ and $(C, C\perp)$ is a full rank matrix. Then we can decompose the group-specific factor as follows

$$\begin{aligned}
X_{i,jt} &= F_t^{c\prime}\lambda_{i,j}^c + F_{i,t}^{s\,\prime}\lambda_{i,j}^s + e_{i,jt} \\
&= F_t^{c\prime}\lambda_{i,j}^c + F_{i,t}^{s\,\prime}(C, C_\perp)(C, C_\perp)^{-1}\lambda_{i,j}^s + e_{i,jt} \\
&= F_t^{c\prime}\lambda_{i,j}^c + (F_{i,t}^{s\,\prime}C, F_{i,t}^{s\,\prime}C_\perp)\lambda_{i,j}^{s*} + e_{i,jt} \\
&= F_t^{c\prime}\lambda_{i,j}^c + (F_{i,t}^{s\,\prime}C, F_{i,t}^{s\,\prime}C_\perp)\begin{pmatrix}\lambda_{i,jC}^{s*} \\ \lambda_{i,jC_\perp}^{s*}\end{pmatrix} + e_{i,jt} \\
&= F_t^{c\prime}\lambda_{i,j}^c + F_{i,t}^{s\,\prime}C\lambda_{i,jC}^{s*} + F_{i,t}^{s\,\prime}C_\perp\lambda_{i,jC_\perp}^{s*} + e_{i,jt} \\
&= F_t^{c\prime}\lambda_{i,j}^c + F_t^{ps\prime}(B^{ps\prime}B^{ps})^{-1}B^{ps\prime}B_i^sC\lambda_{i,jC}^{s*} + F_{i,t}^{s\,\prime}C_\perp\lambda_{i,jC_\perp}^{s*} + e_{i,jt},
\end{aligned}$$

where $\lambda_{i,j}^{s*} = (C, C_\perp)^{-1}\lambda_{i,j}^s$ and $\begin{pmatrix}\lambda_{i,jC}^{s*} \\ \lambda_{i,jC_\perp}^{s*}\end{pmatrix} = \lambda_{i,j}^{s*}$ is a decomposition of the r_i^s-vector $\lambda_{i,j}^{s*}$ into an r_i^{s*}-vector and an $(r_i^s - r_i^{s*})$-vector to be conform with the dimensions in $(C, C_\perp)$.

In order to be consistently with the notation

$$X_{i,jt} = F_t^{pc\prime}\lambda_{i,j}^{pc} + F_{i,t}^{rs\prime}\lambda_{i,j}^{rs} + e_{i,jt}$$

we define in this case $F_t^{pc\prime} = (F_t^{c\prime}, F_t^{ps\prime})$, $\lambda_{i,j}^{pc} = \begin{pmatrix}\lambda_{i,j}^c \\ \lambda_{i,j}^{ps}\end{pmatrix}$, $\lambda_{i,j}^{ps} = (B^{ps\prime}B^{ps})^{-1}B^{ps\prime}B_i^sC\lambda_{i,jC}^{s*}$, $F_t^{rs\prime} = F_{i,t}^{s\prime}C_\perp$ and $\lambda_{i,j}^{rs} = \lambda_{i,jC_\perp}^{s*}$. Again we have

$$X_{i,jt}^{rs} = X_{i,jt} - F_t^{pc\prime}\lambda_{i,j}^{pc} = F_{i,t}^{rs\prime}\lambda_{i,j}^{rs} + e_{i,jt}.$$

In this case X_i^{rs} can be seen as being generated by factors represented through $F_t^{rs} = C_\perp F_{i,t}^s$ and the number of factors is $r_i^s - r_i^{s*}$.

Proof of (iii) see Lemma 6.7. □

Lemma 6.6

For a choice of k^c with $0 < k^c < r_i$, for $i = 1, 2, ..., n$, let

- $\hat{F}^c$ *be the estimate of the common factors given in Theorem 3.3,*
- $\hat{X}_i^{rs} = X_i - \hat{F}^c\hat{\Lambda}_i^c$ *represent the residuals of the regression of X_i on $\hat{F}^c$,*
- $\hat{\hat{X}}_i^{rs} = X_i - \hat{\hat{F}}^c\hat{\hat{\Lambda}}_i^c$ *represent the residuals of the regression of X_i on $\hat{\hat{F}}^c$, with $\hat{\hat{F}}^c = P_{\hat{F}_i}\hat{F}^c$ representing a projection of $\hat{F}^c$ on the estimate of the group-pervasive factor $\hat{F}_i$.*
- $\check{X}_i^{rs} = X_i - F^{pc}\check{\Lambda}_i^{pc}$ *represent the residuals of the regression of X_i on F^{pc} for the given choice of k^c.*
- $X_i^{rs} = X_i - F^{pc}\Lambda_i^{pc}$ *represent the part of common component in X_i after subtraction of $F^{pc}\Lambda_i^{pc}$.*
- $\hat{W} = \frac{1}{N_iT}\hat{X}_i^{rs}\hat{X}_i^{rs\prime} = \frac{1}{N_iT}(X_i - \hat{F}^c\hat{\Lambda}_i^c)(X_i - \hat{F}^c\hat{\Lambda}_i^c)'$,

- $\hat{\hat{W}} = \frac{1}{N_iT}\hat{\hat{X}}_i^{rs}\hat{\hat{X}}_i^{rs\prime} = \frac{1}{N_iT}(X_i - \hat{\hat{F}}^c\hat{\hat{\Lambda}}_i^c)(X_i - \hat{\hat{F}}^c\hat{\hat{\Lambda}}_i^c)'$,
- $\check{W} = \frac{1}{N_iT}\check{X}_i^{rs}\check{X}_i^{rs\prime} = \frac{1}{N_iT}(X_i - F^{pc}\check{\Lambda}_i^{pc})(X_i - F^{pc}\check{\Lambda}_i^{pc})'$, *and*
- $W = \frac{1}{N_iT}X_i^{rs}X_i^{rs\prime} = \frac{1}{N_iT}(X_i - F^{pc}\Lambda_i^{pc})(X_i - F^{pc}\Lambda_i^{pc})'$.

Then we have

(i) $||\hat{W} - W||^2 = O_p(C_{N,T}^{-2})$,

(ii) $\frac{||\hat{F}^c - \hat{\hat{F}}^c||}{\sqrt{T}} = O_p(C_{N,T}^{-1})$ *for* $k^c \leq r^c$.

(iii) $||\hat{\hat{W}} - \hat{W}||^2 = O_p(C_{N,T}^{-2})$ *for* $k^c \leq r^c$.

Let $M_{F^{pc}} = I - F^{pc}(F^{pc\prime}F^{pc})^{-1}F^{pc\prime}$ and $M_{\hat{F}^c} = (I - \hat{F}^c(\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\prime})$. Then we have $\check{W} = \frac{1}{N_iT}\check{X}_i^{rs}\check{X}_i^{rs\prime} = \frac{1}{N_iT}(X_i - F^{pc}\check{\lambda}_i^{pc})(X_i - F^{pc}\check{\lambda}_i^{pc})' = M_{F^{pc}}\frac{(X_iX_i')}{N_iT}M_{F^{pc}}$ and $\hat{W} = \frac{1}{N_iT}\hat{X}_i^{rs}\hat{X}_i^{rs\prime} = \frac{1}{N_iT}(X_i - \hat{F}_c\hat{\lambda}_i^c)(X_i - \hat{F}_c\hat{\lambda}_i^c)' = M_{\hat{F}^c}\frac{(X_iX_i')}{N_iT}M_{\hat{F}^c}$. Let $\mathcal{H}_{k^c}$ be the matrix defined in Theorem 3.5. We have $M_{F^{pc}\mathcal{H}_{k^c}} = I - F^{pc}\mathcal{H}_{k^c}\mathcal{H}_{k^c}^{-1}(F^{pc\prime}F^{pc})^{-1}\mathcal{H}_{k^c}^{\prime -1}\mathcal{H}_{k^c}'F^{pc\prime}$

Proof of (i):

$$\begin{aligned}
M_{\hat{F}^c} - M_{F^{pc}} &= M_{\hat{F}^c} - M_{F^{pc}\mathcal{H}_{k^c}} \\
&= (I - \hat{F}^c(\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\prime}) - (I - F^{pc}\mathcal{H}_{k^c}(\mathcal{H}_{k^c}'F^{pc\prime}F^{pc}\mathcal{H}_{k^c})^{-1}\mathcal{H}_{k^c}'F^{pc\prime}) \\
&= T^{-1}F^{pc}\mathcal{H}_{k^c}\left(\mathcal{H}_{k^c}'\frac{F^{pc\prime}F^{pc}}{T}\mathcal{H}_{k^c}\right)^{-1}\mathcal{H}_{k^c}'F^{pc\prime} - T^{-1}\hat{F}^c\left(\frac{\hat{F}^{c\prime}\hat{F}^c}{T}\right)^{-1}\hat{F}^{c\prime} \\
&= T^{-1}(\hat{F}^c\mathcal{H}_{k^c} - F^{pc})D_{\mathcal{H}_{k^c}}^{-1}(\hat{F}^{c\prime}\mathcal{H}_{k^c} - F^{pc\prime}) \\
&\quad -T^{-1}(\hat{F}^c\mathcal{H}_{k^c} - F^{pc})D_{\mathcal{H}_{k^c}}^{-1}F^{pc\prime} \\
&\quad +T^{-1}F^{pc}\mathcal{H}_{k^c}D_{\mathcal{H}_{k^c}}^{-1}(\hat{F}^c - F^{pc\prime}\mathcal{H}_{k^c})' + T^{-1}F^{pc}\mathcal{H}_{k^c}(\hat{D}^{-1} - D_{\mathcal{H}_{k^c}}^{-1})\mathcal{H}_{k^c}'F^{pc\prime} \\
&= a + b + c + d
\end{aligned}$$

$$\begin{aligned}
||a|| &= T^{-1}||(\hat{F}^c - F^{pc}\mathcal{H}_{k^c})D_{\mathcal{H}_{k^c}}^{-1}(\hat{F}^c - F^{pc}\mathcal{H}_{k^c})'|| \\
&= T^{-1}\sqrt{\sum_{t=}^{T}\sum_{\tau=1}^{T}\left((\hat{F}_t^c - \mathcal{H}_{k^c}'F_t^{pc})'D_{\mathcal{H}_{k^c}}^{-1}(\hat{F}_\tau^c - \mathcal{H}_{k^c}'F_\tau^{pc})\right)^2} \\
&= T^{-1}\sqrt{\sum_{t=}^{T}\sum_{\tau=1}^{T}||(\hat{F}_t^c - \mathcal{H}_{k^c}'F_t^{pc})'D_{\mathcal{H}_{k^c}}^{-1}(\hat{F}_\tau^c - \mathcal{H}_{k^c}'F_\tau^{pc})||^2} \\
&\leq T^{-1}\sqrt{\sum_{t=1}^{T}\sum_{\tau=1}^{T}||\hat{F}_t^c - \mathcal{H}_{k^c}'F_t^{pc}||^2||D_{\mathcal{H}_{k^c}}^{-1}||^2||\hat{F}_\tau^c - \mathcal{H}_{k^c}'F_\tau^{pc}||^2} \\
&\leq \left(||D_{\mathcal{H}_{k^c}}^{-1}||^2\frac{1}{T}\sum_{t=}^{T}||(\hat{F}_t^c - \mathcal{H}_{k^c}'F_t^{pc})'||^2\frac{1}{T}\sum_{\tau=1}^{T}||\hat{F}_\tau^c - \mathcal{H}_{k^c}'F_\tau^{pc}||^2\right)^{\frac{1}{2}} \\
&= O_p(C_{N,T}^{-2})
\end{aligned}$$

$$
\begin{aligned}
||b|| &= T^{-1}(\hat{F}^c - F^{pc}\mathcal{H}_{k^c})D^{-1}_{\mathcal{H}_{k^c}}\mathcal{H}'_{k^c}F^{pc'} \\
&= T^{-1}\sqrt{\sum_{t=}^{T}\sum_{\tau=1}^{T}||(\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t)'D^{-1}_{\mathcal{H}_{k^c}}(\mathcal{H}'_{k^c}F^{pc}_\tau)||^2} \\
&\leq T^{-1}\sqrt{\sum_{t=1}^{T}\sum_{\tau=1}^{T}||\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t||^2||D^{-1}_{\mathcal{H}_{k^c}}||^2||\mathcal{H}'_{k^c}F^c_\tau||^2} \\
&= O_p(C^{-1}_{N,T})
\end{aligned}
$$

Similarly we have $||c|| = O_p(C^{-1}_{N,T})$.

$$
\begin{aligned}
||d|| &= T^{-1}||F^{pc}(D^{-1}_{\mathcal{H}_{k^c}} - \hat{D}^{-1})F^{pc'}|| \\
&= T^{-1}\sqrt{\sum_{t=}^{T}\sum_{\tau=1}^{T}||(\mathcal{H}'_{k^c}F^{pc}_t)'(D^{-1}_{\mathcal{H}_{k^c}} - \hat{D}^{-1})(\mathcal{H}'_{k^c}F^{pc}_\tau)||^2} \\
&= \sqrt{||(D^{-1}_{\mathcal{H}_{k^c}} - \hat{D}^{-1})||^2\frac{1}{T}\sum_{t=}^{T}||(\mathcal{H}'_{k^c}F^{pc}_t)||^2\frac{1}{T}\sum_{\tau=1}^{T}||(\mathcal{H}'_{k^c}F^{pc}_\tau)||^2} \\
&= O_p(C^{-1}_{N,T})
\end{aligned}
$$

In the probability limit on the last row we have used the following result:

$$
\begin{aligned}
&(\hat{D} - D_{\mathcal{H}_{k^c}}) \\
=\ &\frac{\hat{F}^{c'}_i\hat{F}^c}{T} - \frac{\mathcal{H}'_{k^c}F^{pc'}F^{pc}\mathcal{H}_{k^c}}{T} \\
=\ &T^{-1}\sum_{t=1}^{T}[\hat{F}^c_t\hat{F}^{c'}_t - \mathcal{H}'_{k^c}F^{pc}_tF^{pc'_t}\mathcal{H}_{k^c}] \\
=\ &T^{-1}\sum_{i=1}^{T}(\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t)(\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t)' + T^{-1}\sum_{i=1}^{T}(\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t)F^{pc'}_t\mathcal{H}_{k^c}) \\
&+T^{-1}\sum_{i=1}^{T}\mathcal{H}'_{k^c}F^{pc}_t(\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t)'
\end{aligned}
$$

$$
\begin{aligned}
&||(\hat{D} - D_{\mathcal{H}_{k^c}})|| \\
\leq\ &T^{-1}\sum_{i=1}^{T}||\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t||^2 + \left(T^{-1}\sum_{i=1}^{T}||\hat{F}^c_t - \mathcal{H}'_{k^c}F^{pc}_t||^2\right)^{-\frac{1}{2}} + \left(T^{-1}\sum_{i=1}^{T}||\mathcal{H}'_{k^c}F^{pc}_t||^2\right)^{\frac{1}{2}} \\
=\ &O_p(C^{-2}_{N,T}) + O_p(C^{-1}_{N,T})
\end{aligned}
$$

$$
||(\hat{D}^{-1} - D^{-1}_{\mathcal{H}_{k^c}}|| = ||\hat{D}^{-1}(D_{\mathcal{H}_{k^c}} - \hat{D})D^{-1}_{\mathcal{H}_{k^c}}|| \leq O_p(1)O_p(C^{-1}_{N,T})O_p(1) = O_p(C^{-1}_{N,T})
$$

So we have $||M_{\hat{F}^c} - M_{F^{pc}}|| \leq ||a|| + ||b|| + ||c|| + ||d|| = O_p(C^{-1}_{N,T})$. It follows

$$
||M_{\hat{F}^c} - M_{F^{pc}}||^2 = O_p(C^{-2}_{N,T}). \tag{6.57}
$$

$$
\begin{aligned}
& ||\hat{W}-\check{W}|| \\
= & \Big|\Big| M_{F^{pc}} \frac{(X_i X_i')}{N_i T} M_{F^{pc}} - M_{\hat{F}^c} \frac{(X_i X_i')}{N_i T} M_{\hat{F}^c} \Big|\Big| \\
= & \Big|\Big| (M_{F^{pc}} - M_{\hat{F}^c} + M_{\hat{F}^c}) \frac{(X_i X_i')}{N_i T} (M_{F^{pc}} - M_{\hat{F}^c} + M_{\hat{F}^c}) - M_{\hat{F}^c} \frac{(X_i X_i')}{N_i T} M_{\hat{F}^c} \Big|\Big| \\
= & \Big|\Big| (M_{F^{pc}} - M_{\hat{F}^c}) \frac{(X_i X_i')}{N_i T} (M_{F^{pc}}) + (M_{\hat{F}^c}) \frac{(X_i X_i')}{N_i T} (M_{F^{pc}} - M_{\hat{F}^c}) \Big|\Big| \\
\leq & \Big|\Big| (M_{F^{pc}} - M_{\hat{F}^c}) \frac{(X_i X_i')}{N_i T} (M_{F^{pc}}) \Big|\Big| + \Big|\Big| (M_{\hat{F}^c}) \frac{(X_i X_i')}{N_i T} (M_{F^{pc}} - M_{\hat{F}^c}) \Big|\Big| \\
\leq & ||M_{F^{pc}} - M_{\hat{F}^c}|| \Big|\Big| \frac{(X_i X_i')}{N_i T} \Big|\Big| ||(M_{F^{pc}})|| + ||M_{\hat{F}^c}|| \Big|\Big| \frac{(X_i X_i')}{N_i T} \Big|\Big| ||(M_{F^{pc}} - M_{\hat{F}^c})|| \\
= & O_p(C_{N,T}^{-1}) O_p(1) O_p(1) + O_p(1) O_p(1) O_p(C_{N,T}^{-1}) = O_p(C_{N,T}^{-1}).
\end{aligned}
$$

In the last row we use $||\frac{(X_i X_i')}{N_i T}|| = O_p(1)$, $||M_{F^{pc}}|| = O_p(1)$ and $||(M_{F^{pc}} - M_{\hat{F}^c})|| = O_p(C_{N,T}^{-1})$. So we proved $||\hat{W}-\check{W}||^2 = O_p(C_{N,T}^{-2})$.

$$
\begin{aligned}
\frac{1}{N_i T} ||\check{X}^{rs} - X^{rs}||^2 &= \frac{1}{N_i T} ||F^{pc}\Lambda_i^{pc} - F^{pc}\check{\Lambda}_i||^2 \\
&= \frac{1}{N_i T} ||F^{pc}\Lambda_i^{pc} - F^{pc}(F^{pc'}F^{pc})^{-1}F^{pc'}X_i||^2 \\
&= \frac{1}{N_i T} ||F^{pc}\Lambda_i^{pc} - F^{pc}(F^{pc'}F^{pc})^{-1}F^{pc'}(F^{pc}\Lambda_i^{pc} + F_i^s\Lambda_i^s + E_i)||^2 \\
&= \frac{1}{N_i T} ||F^{pc}(F^{pc'}F^{pc})^{-1}F^{pc'}(F_i^s\Lambda_i^s + E_i)||^2 \\
&= \frac{1}{N_i T} ||F^{pc}(F^{pc'}F^{pc})^{-1}F^{pc'}E_i||^2 \\
&= \frac{1}{N_i T} \Big|\Big| \frac{1}{\sqrt{T}} F^{pc} \Big(\frac{F^{pc'}F^{pc}}{T}\Big)^{-1} \frac{F^{pc'}E_i}{\sqrt{T}} \Big|\Big|^2 \\
&= \frac{1}{N_i T} \sum_{t=1}^{T} \sum_{j=1}^{N_i} \Big| \frac{1}{\sqrt{T}} F_t^{pc'} \Big(\frac{1}{T}\sum_{\tau=1}^{T} F_\tau^{pc}F_\tau^{pc'}\Big)^{-1} \frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} F_\tau^{pc} e_{i,j\tau} \Big|^2 \\
&\leq \frac{1}{T^2} \sum_{t=1}^{T} ||F_t^{pc}||^2 \Big|\Big| \Big(\frac{1}{T}\sum_{\tau=1}^{T} F_\tau^{pc}F_\tau^{pc'}\Big)^{-1} \Big|\Big|^2 \frac{1}{N_i} \sum_{j=1}^{N_i} \Big|\Big| \frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} F_\tau^{pc} e_{i,j\tau} \Big|\Big|^2 \\
&\leq \frac{1}{T} O_p(1) O_p(1) M = O_p(C_{N,T}^{-2}).
\end{aligned}
$$

$$
\begin{aligned}
& ||\check{W} - W|| \\
= & \Big|\Big| \frac{\check{X}_i^{rs}\check{X}_i^{rs'}}{N_i T} - \frac{(X_i^{rs} - \check{X}_i^{rs} + \check{X}_i^{rs})(X_i^{s'} - \check{X}_i^{rs'} + \check{X}_i^{rs'})}{N_i T} \Big|\Big| \\
= & \Big|\Big| - \frac{(X_i^{rs} - \check{X}_i^{rs})X_i^{rs'} + \check{X}_i^{rs}(X_i^{s'} - \check{X}_i^{rs'})}{N_i T} \Big|\Big| \\
\leq & \Big|\Big| \frac{(X_i^{rs} - \check{X}_i^{rs})X_i^{rs'}}{N_i T} \Big|\Big| + \Big|\Big| \frac{\check{X}_i^{rs}(X_i^{rs'} - \check{X}_i^{rs'})}{N_i T} \Big|\Big| \\
= & O_p(C_{N,T}^{-1}).
\end{aligned}
$$

The last follows from: $\left|\left|\frac{(X_i^{rs}-\check{X}_i^{rs})}{\sqrt{N_iT}}\right|\right| = O_p(C_{N,T}^{-1})$ and $\left|\left|\frac{X_i^{rs\prime}}{\sqrt{N_iT}}\right|\right| = O_p(1)$.

$$||\hat{W} - W|| = ||\hat{W} - \check{W} + \check{W} - W|| \leq ||\hat{W} - \check{W}|| + ||\check{W} - W|| = O_p(C_{N,T}^{-1}).$$

It follows $||\hat{W} - W||^2 = O_p(C_{N,T}^{-2})$.

Proof of (ii):

For $k^c \leq r^c$ we have

$$\begin{aligned}
||(\hat{F}^c - \hat{\hat{F}}^c)/\sqrt{T}|| &= ||M_{\hat{F}_i}\hat{F}^c/\sqrt{T}|| \\
&= ||(M_{\hat{F}_i} - M_{F_i})\hat{F}^c/\sqrt{T} + M_{F_i}(\hat{F}^c - F^c B_{rk^c})/\sqrt{T} + M_{F_i}F^c B_{rk^c}/\sqrt{T}|| \\
&\leq ||(M_{\hat{F}_i} - M_{F_i})||||\hat{F}^c/\sqrt{T}|| + ||M_{F_i}||||(\hat{F}^c - F^c B_{rk^c})/\sqrt{T}|| \\
&= O_p(C_{N,T}^{-1})O_p(1) + O_p(1)O_p(C_{N,T}^{-1})
\end{aligned}$$

Using the same technique as the proof of equation (6.57) we can show $||(M_{\hat{F}_i} - M_{F_i})|| = O_p(C_{N,T}^{-1})$. $||(\hat{F}^c - F^c B_{rk^c})/\sqrt{T}||O_p(C_{N,T}^{-1})$ is based on Theorem 3.3. $||\hat{F}^c/\sqrt{T}|| = O_p(1)$ is according to Assumption 2.5 and Theorem 3.3.

Proof of (iii)

$$\begin{aligned}
&||\hat{W} - \hat{\hat{W}}|| \\
= \; & \left|\left|M_{\hat{F}^c}\frac{(X_iX_i')}{N_iT}M_{\hat{F}^c} - M_{\hat{\hat{F}}^c}\frac{(X_iX_i')}{N_iT}M_{\hat{\hat{F}}^c}\right|\right| \\
= \; & \left|\left|(M_{\hat{F}^c} - M_{\hat{\hat{F}}^c} + M_{\hat{\hat{F}}^c})\frac{(X_iX_i')}{N_iT}(M_{\hat{F}^c} - M_{\hat{\hat{F}}^c} + M_{\hat{\hat{F}}^c}) + M_{\hat{\hat{F}}^c}\frac{(X_iX_i')}{N_iT}M_{\hat{\hat{F}}^c}\right|\right| \\
= \; & \left|\left|(M_{\hat{F}^c} - M_{\hat{\hat{F}}^c})\frac{(X_iX_i')}{N_iT}M_{\hat{F}^c} + M_{\hat{F}^c}\frac{(X_iX_i')}{N_iT}(M_{\hat{F}^c} - M_{\hat{\hat{F}}^c})\right|\right| \\
\leq \; & ||M_{\hat{F}^c} - M_{\hat{\hat{F}}^c}||||\frac{(X_iX_i')}{N_iT}||||M_{\hat{F}^c}|| + ||M_{\hat{F}^c}||||\frac{(X_iX_i')}{N_iT}||||M_{\hat{F}^c} - M_{\hat{\hat{F}}^c}|| \\
= \; & O_p(C_{N,T}^{-1})O_p(1)O_p(1) + O_p(1)O_p(1)O_p(C_{N,T}^{-1}) = O_p(C_{N,T}^{-1}).
\end{aligned}$$

In the probability convergence above we have used $||M_{\hat{F}^c} - M_{\hat{\hat{F}}^c}|| = O_p(C_{N,T}^{-1})$ that can be proved by going through the the same steps in the proof of equation (6.57).
□

Lemma 6.7

For a given choice of (k^c, k_i^s) with $k^c < r_i$ and $k_i^s > 1$, let $\tilde{F}^s = \frac{1}{N_iT}(X_i^{rs}X_i^{rs\prime})\sqrt{T}\tilde{Q}_{k_i^s}$, where $\tilde{Q}_{k_i^s}$ is a $(T \times k_i^s)$ matrix of the eigenvectors corresponding to the k_i^s largest eigenvalues of $(X_i^{rs}X_i^{rs\prime})$. $\hat{F}^s = \frac{1}{N_iT}(\hat{X}_i^{rs}\hat{X}_i^{rs\prime})\sqrt{T}\hat{Q}_{k_i^s}$, where $\hat{Q}_{k_i^s}$ is a $(T \times k_i^s)$ matrix of the eigenvectors that correspond to the k_i^s largest eigenvalues of $(\hat{X}_i^{rs}\hat{X}_i^{rs\prime})$.

Then we have

$$\frac{1}{T}\sum_{t=1}^{T}||\tilde{F}_t^s - \hat{F}_t^s||^2 = O_p(C_{N,T}^{-2}).$$

Proof: Let $\tilde{D}_{k_i^s}$ and $\hat{D}_{k_i^s}$ be the diagonal matrices of the k_i^s largest eigenvalues of

$\frac{X_i^s X_i^{s\prime}}{TN_i}$ and $\frac{\hat{X}_i^{rs}\hat{X}_i^{s\prime}}{TN_i}$ respectively.

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}||\tilde{F}_{i,t}^s - \hat{F}_{i,t}^s||^2 \\
= & \frac{1}{T}||\tilde{F}_i^s - \hat{F}_i^s||^2 \\
= & \frac{1}{T}\Big|\Big|\frac{X_i^{rs}X_i^{rs\prime}}{TN_i}\sqrt{T}\tilde{Q}_{k_i^s} - \frac{\hat{X}_i^{rs}\hat{X}_i^{rs\prime}}{TN_i}\sqrt{T}\hat{Q}_{k_i^s}\Big|\Big|^2 \\
= & \Big|\Big|\tilde{Q}_{k_i^s}\tilde{D}_{k_i^s} - \hat{Q}_{k_i^s}\hat{D}_{k_i^s}\Big|\Big|^2 \\
= & \Big|\Big|(\tilde{Q}_{k_i^s} - \hat{Q}_{k_i^s} + \hat{Q}_{k_i^s})(\tilde{D}_{k_i^s} - \hat{D}_{k_i^s} + \hat{D}_{k_i^s}) - \hat{Q}_{k_i^s}\hat{D}_{k_i^s}\Big|\Big|^2 \\
= & \Big|\Big|(\tilde{Q}_{k_i^s} - \hat{Q}_{k_i^s})\tilde{D}_{k_i^s} + \hat{Q}_{k_i^s}(\tilde{D}_{k_i^s} - \hat{D}_{k^c})\Big|\Big|^2 \\
\leq & \Big|\Big|(\tilde{Q}_{k_i^s} - \hat{Q}_{k_i^s})\Big|\Big|^2||\tilde{D}_{k_i^s}||^2 + ||\hat{Q}_{k_i^s}||^2\Big|\Big|(\tilde{D}_{k_i^s} - \hat{D}_{k_i^s})\Big|\Big|^2 \\
& +2||\hat{Q}_{k_i^s}||\Big|\Big|(\tilde{D}_{k_i^s} - \hat{D}_{k_i^s})\Big|\Big|\Big|\Big|(\tilde{Q}_{k_i^s} - \hat{Q}_{k_i^s})\Big|\Big|||\tilde{D}_{k_i^s}|| \\
= & O_p(C_{N,T}^{-2})O_p(1) + O_p(1)O_p(C_{N,T}^{-2}) + 2O_p(C_{N,T}^{-1})O_p(1)O_p(1)O_p(C_{N,T}^{-1}) = O_p(C_{N,T}^{-2}).
\end{aligned}$$

The probability limit in the last step is according to Lemma 6.2 for the case the eigenvalues of the matrix $\frac{X_i^{rs}X_i^{rs\prime}}{TN_i}$ are all unique. For the case of multiple eigenvalues. According to Lemma 6.2 (iv), for a given set of eigenvectors $\hat{Q}_{k_i^s}$ of $\hat{W}$ there exists a particular set of eigenvectors $\tilde{Q}_{k_i^s}$ of $\tilde{W}$, such that the probability limit in the last step of the equation holds.

Corollary 6.8
For a given choice of (k^c, k_i^s) with $k^c \leq r^c$ and $k_i^s > 1$, let $\hat{\hat{F}}^s = \frac{1}{N_iT}(\hat{\hat{X}}_i^{rs}\hat{\hat{X}}_i^{rs\prime})\sqrt{T}\hat{\hat{Q}}_{k_i^s}$, where $\hat{\hat{Q}}_{k_i^s}$ is a $(T \times k_i^s)$ matrix of the eigenvectors corresponding to the k_i^s largest eigenvalues of $(\hat{\hat{X}}_i^{rs}\hat{\hat{X}}_i^{rs\prime})$. $\hat{F}^s = \frac{1}{N_iT}(\hat{X}_i^{rs}\hat{X}_i^{rs\prime})\sqrt{T}\hat{Q}_{k_i^s}$, where $\hat{Q}_{k_i^s}$ is a $(T\times k_i^s)$ matrix of the eigenvectors that correspond to the k_i^s largest eigenvalues of $(\hat{X}_i^{rs}\hat{X}_i^{rs\prime})$.

Then we have

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{\hat{F}}_t^s - \hat{F}_t^s||^2 = O_p(C_{N,T}^{-2}).$$

Proof:

Going through the same steps as in the proof of Lemma 6.7, we obtain the result. □

Lemma 6.9 *For a choice (k^c, k_i^s) with $k^c < r^c$ and $k^c + k_i^s = k_i$, let $\hat{F}_i$ and $\hat{F}^c$ be the estimates of the group-pervasive factors and the common factors defined in Lemma 6.4 and Theorem 3.3 respectively. Let $\hat{\hat{F}}^c$ and $\hat{\hat{F}}^s$ be the estimates for the common factors and the group-specific factors defined in Lemma 6.6. Then there exists a $(k_i \times k_i)$ full rank matrix matrix B such that*

(i) $\hat{F}_i = (\hat{\hat{F}}^c, \hat{\hat{F}}^s)B$.

(ii) $V(\hat{F}_i, k_i) = V(\hat{\hat{F}}_i^s, k_i^s|\hat{\hat{F}}^c, k^c)$.

Proof

Because $\hat{\hat{F}}^c = P_{\hat{F}_i}\hat{F}^c$ is a linear combination of $\hat{F}_i$ we have $\hat{\hat{F}}^c = \hat{F}_i A_1$, where A_1 is a $(k_i \times k^c)$ matrix.

$\hat{F}_i$ as a principal component estimate for the group-pervasive factors is a solution of the following minimization problem

$$(\hat{F}_i, \hat{\Lambda}_i) = \underset{F_i,\Lambda_i}{\operatorname{argmin}}(N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}(X_{i,jt} - F'_{i,t}\lambda_{i,j})^2 \tag{6.58}$$

Because the solution of identified up to a full rank rotation, the rotated factor estimate $\hat{F}_i A = \hat{F}_i(A_1, A_{1\perp})$ is also a solution of the problem, where $A_{1\perp}$ is a $(k_i \times k_i - k^c)$ matrix and A has full rank and $\hat{\Lambda}_{i,1}$ and $\hat{\Lambda}_{i,1\perp}$ are the corresponding loadings.

$$\begin{aligned}
(\hat{F}_i A_1, \hat{F}_i A_{1\perp}, \hat{\Lambda}_{i,1}, \hat{\Lambda}_{i,1\perp}) &= \underset{F_{i,1},F_{i,2},\Lambda_{i,1},\Lambda_{i,2}}{\operatorname{argmin}}(N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}(X_{i,jt} - F'_{i,1t}\lambda_{i,1j} - F'_{i,2t}\lambda_{i,2j})^2 \\
(\hat{F}_i A_{1\perp}, \hat{\Lambda}_{i,1\perp}) &= \underset{F_{i,2},\Lambda_{i,2}}{\operatorname{argmin}}(N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}(X_{i,jt} - \hat{\hat{F}}_t^{c\prime}\hat{\hat{\lambda}}_{i,j}^c - F_{i,2t}\lambda_{i,2j})^2
\end{aligned}$$

In the second equality above we have used the fact $\hat{\hat{F}}^c = \hat{F}_i A_1$, and $\hat{\hat{\Lambda}}_i^c = \hat{\Lambda}_{i,1}$ is a solution of the maximization problem. Because $\hat{\hat{F}}^s$ is also a solution of the minimization problem in the last row:

$$(\hat{\hat{F}}^s, \hat{\hat{\Lambda}}_i^s) = \underset{F_{i,2},\Lambda_{i,2}}{\operatorname{argmin}}(N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}((X_{i,jt} - \hat{\hat{F}}_t^{c\prime}\hat{\hat{\lambda}}_{i,j}^c) - F_{i,2t}\lambda_{i,2j})^2, \tag{6.59}$$

we have

$$\hat{F}_i A_{1\perp} = \hat{\hat{F}}^s C, \tag{6.60}$$

with C a $(k_i - k^c) \times (k_i - k^c)$ full rank matrix. Then we have

$$\hat{F}_i = \hat{F}_i A A^{-1}(\hat{\hat{F}}^c, \hat{\hat{F}}^s C)A^{-1} = (\hat{\hat{F}}^c, \hat{\hat{F}}^s)\begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix}A^{-1} = (\hat{\hat{F}}^c, \hat{\hat{F}}^s)B, \tag{6.61}$$

with $B = \begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix}A^{-1}$. Comparing equation (6.58) with equation (6.59) we have for $k_i = k^c + k_i^s$

$$\begin{aligned}
V(\hat{F}_i, k_i) &= \min_{F_i,\Lambda_i}(N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}(X_{i,jt} - F'_{i,t}\lambda_{i,j})^2 \\
&= (N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}(X_{i,jt} - \hat{F}'_{i,t}\hat{\lambda}_{i,j})^2 \\
&= \min_{F_{i,1},F_{i,2},\Lambda_{i,1},\Lambda_{i,2}}(N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}(X_{i,jt} - F'_{i,1t}\lambda_{i,1j} - F'_{i,2t}\lambda_{i,2j})^2 \\
&= (N_i T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N_i}(X_{i,jt} - \hat{\hat{F}}_t^{\prime c}\hat{\hat{\lambda}}_{i,j}^c - \hat{\hat{F}}_{i,t}^{s\prime}\hat{\hat{\lambda}}_{i,j}^s)^2 && (6.62) \\
&= V(\hat{\hat{F}}_i^s, k_i^s | \hat{\hat{F}}^c, k^c) && (6.63)
\end{aligned}$$

$\square$

Proof of Theorem 3.5

Restating equations (3.21) and (3.22) of Theorem 3.3 in terms of F_t^{pc} defined in Proposition 3.4 we have for $k^c \leq r^c$

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^c - B'_{rk^c}F_t^c||^2 = \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^c - I'_{k^c}F_t^{pc}||^2 = O_p(C_{N,T}^{-2}), \tag{6.64}$$

and for $r^c < k^c \leq r_i$, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left|\left|\hat{F}_t^c - B'_{k^c}\left(\begin{array}{c} F_t^c \\ F_t^{ps}\end{array}\right)\right|\right|^2 = \frac{1}{T}\sum_{t=1}^{T}\left|\left|\hat{F}_t^c - B'_{k^c}F_t^{pc}\right|\right|^2 = O_p(C_{N,T}^{-2}). \tag{6.65}$$

Define

$$\mathcal{H}_{k^c} = \left\{ \begin{array}{ll} I_{k^c} & \text{for } k^c \leq r^c \\ B_{k^c} & \text{for } r^c < k^c < r_i \end{array}\right.$$

and combine the two cases above we proved

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_t^c - \mathcal{H}'_{k^c}F_t^{pc}||^2 = O_p(C_{N,T}^{-2}). \tag{6.66}$$

According to Proposition 3.4 (i) and (iv), $X_{i,jt}^{rs}$ can be seen as generated by a factor model with $F_{i,t}^{rs}$ as the factor. For the case of $k^c > r^c$, equation (6.56) implies that the factor model generating $X_{i,jt}^{rs}$ satisfies Assumptions 2.4 to 2.7. Therefore, we can apply Theorem 1 of Bai and Ng (2002) to the model: there exists a $(k_i^{rs} \times k_i^s)$ matrix $\mathcal{H}_{k_i^s}$ with $rank(\mathcal{H}_{k_i^s}) = \min(k_i^s, k_i^{rs})$ such that $\frac{1}{T}\sum_{t=1}^{T}||\tilde{F}_{i,t}^s - \mathcal{H}'_{k_i^s}F_{i,t}^{rs}||^2 = O_p(C_{N,T}^{-2})$, where $\tilde{F}_i^s = \frac{1}{N_iT}(X_i^{rs}X_i^{rs\prime})\sqrt{T}\tilde{Q}_{k_i^s}$ is a factor estimate based on data of X_i^{rs}.

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}^s - \mathcal{H}'_{k_i^s}F_{i,t}^{rs}||^2 & (6.67)\\
= \; & \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}^s - \tilde{F}_{i,t}^s + \tilde{F}_{i,t}^s - \mathcal{H}'_{k_i^s}F_{i,t}^{rs}||^2 \\
\leq \; & \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}^s - \tilde{F}_{i,t}^s||^2 + \frac{1}{T}\sum_{t=1}^{T}||\tilde{F}_{i,t}^s - \mathcal{H}'_{k_i^s}F_{i,t}^{rs}||^2 + \frac{2}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}^s - \tilde{F}_{i,t}^s||||\tilde{F}_{i,t}^s - \mathcal{H}'_{k_i^s}F_{i,t}^{rs}|| \\
\leq \; & \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}^s - \tilde{F}_{i,t}^s||^2 + \frac{1}{T}\sum_{t=1}^{T}||\tilde{F}_{i,t}^s - \mathcal{H}'_{k_i^s}F_{i,t}^{rs}||^2 \\
& +2\left(\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}^s - \tilde{F}_{i,t}^s||^2\right)^{-\frac{1}{2}}\left(\frac{1}{T}\sum_{t=1}^{T}||\tilde{F}_{i,t}^s - \mathcal{H}'_{k_i^s}F_{i,t}^{rs}||^2\right)^{-\frac{1}{2}} \\
= \; & O_p(C_{N,T}^{-2}).
\end{aligned}$$

In the probability convergence is according to Lemma 6.7.

For the case of $k^c \leq r^c$, and $k_i^s + k^c = k_i$ combining Lemma 6.6 (ii) with Corollary 6.8 we have $\frac{1}{T}\sum_{t=1}^{T}\left|\left|\left(\begin{array}{c}\hat{F}_t^c \\ \hat{F}_{i,t}^s\end{array}\right) - \left(\begin{array}{c}\hat{\hat{F}}_t^c \\ \hat{\hat{F}}_{i,t}^s\end{array}\right)\right|\right|^2 = O_p(C_{N,T}^{-2})$.

By Lemma 6.9 we have $\begin{pmatrix} \hat{\hat{F}}_t^c \\ \hat{\hat{F}}_{i,t}^s \end{pmatrix} = B^{-1'}\hat{F}_{i,t}$ with B a $(k_i \times k_i)$ full rank matrix. Note that $\hat{F}_i$ is the estimate of the group-pervasive factor F_i satisfying: $\frac{1}{T}\sum_{t=1}^T ||\hat{F}_{i,t} - \tilde{\mathcal{H}}_i' F_{i,t}||^2 = O_p(C_{N,T}^{-2})$.

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^T \Big|\Big| \begin{pmatrix} \hat{F}_t^c \\ \hat{F}_{i,t}^s \end{pmatrix} - B^{-1'}\tilde{\mathcal{H}}_i' F_{i,t} \Big|\Big|^2 \\
= & \frac{1}{T}\sum_{t=1}^T \Big|\Big| \begin{pmatrix} \hat{F}_t^c \\ \hat{F}_{i,t}^s \end{pmatrix} - \begin{pmatrix} \hat{\hat{F}}_t^c \\ \hat{\hat{F}}_{i,t}^s \end{pmatrix} + \begin{pmatrix} \hat{\hat{F}}_t^c \\ \hat{\hat{F}}_{i,t}^s \end{pmatrix} - B^{-1'}\hat{F}_{i,t} + B^{-1'}\hat{F}_{i,t} - B^{-1'}\tilde{\mathcal{H}}_i' F_{i,t} \Big|\Big|^2 \\
\leq & \frac{1}{T}\sum_{t=1}^T \Big|\Big| \begin{pmatrix} \hat{F}_t^c \\ \hat{F}_{i,t}^s \end{pmatrix} - \begin{pmatrix} \hat{\hat{F}}_t^c \\ \hat{\hat{F}}_{i,t}^s \end{pmatrix} \Big|\Big|^2 + \frac{1}{T}\sum_{t=1}^T ||B^{-1'}\hat{F}_{i.t} - B^{-1'}\tilde{\mathcal{H}}_i' F_{i,t}||^2 \\
= & O_p(C_{N,T}^{-2}) + O_p(C_{N,T}^{-2})
\end{aligned}$$

According to the definition of F_t^{pc} and F_t^{rs} in equation (6.51), we can decompose $F_{i,t}$ into

$$F_{i,t} = \begin{pmatrix} B_{r^c}^{-1'} & 0 \\ 0 & I_{r_i^s} \end{pmatrix} \begin{pmatrix} B_{r^c}' F_t^c \\ F_{i,t}^s \end{pmatrix} = \begin{pmatrix} B_{r^c}^{-1'} & 0 \\ 0 & I_{r_i^s} \end{pmatrix} \begin{pmatrix} B_{rk^c}' F_t^c \\ B_{rk^c\perp}' F_t^c \\ F_{i,t}^s \end{pmatrix} = \begin{pmatrix} B_{r^c}^{-1'} & 0 \\ 0 & I_{r_i^s} \end{pmatrix} \begin{pmatrix} F_t^{pc} \\ F_t^{rs} \end{pmatrix}$$

Let $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ be the product matrix of $B^{-1}\tilde{\mathcal{H}}_i' \begin{pmatrix} B_{r^c}^{-1'} & 0 \\ 0 & I_{r_i^s} \end{pmatrix}$ decomposed conformable to $\begin{pmatrix} F_t^{pc} \\ F_t^{rs} \end{pmatrix}$, we have

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^T \Big|\Big| \begin{pmatrix} \hat{F}_t^c \\ \hat{F}_{i,t}^s \end{pmatrix} - \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} F_t^{pc} \\ F_t^{rs} \end{pmatrix} \Big|\Big|^2 \\
= & \frac{1}{T}\sum_{t=1}^T \Big|\Big| \begin{pmatrix} \hat{F}_t^c \\ \hat{F}_{i,t}^s \end{pmatrix} - \begin{pmatrix} A_{11}F_t^{pc} + A_{12}F_t^{rs} \\ A_{21}F_t^{pc} + A_{22}F_t^{rs} \end{pmatrix} \Big|\Big|^2 = O_p(C_{N,T}^{-2})
\end{aligned}$$

Since $\frac{1}{T}\sum_{t=1}^T ||\hat{F}_t^c - F_t^{pc}|| = O_p(C_{N,T}^{-2})$, we have $A_{11} = I_{k^c}$ and $A_{12} = 0$. Since $\hat{F}_t^c$ and $\hat{F}_{i,t}^s$ are orthogonal $\frac{1}{T}\sum_{t=1}^T \hat{F}_{i,t}^s \hat{F}_t^{c'} = 0$, we have $A_{21}\frac{1}{T}\sum_{t=1}^T F_t^{pc}F_t^{pc'} + A_{22}\frac{1}{T}\sum_{t=1}^T F_t^{rs}F_t^{pc'} = O_p(C_{N,T}^{-2})$. Because $\frac{1}{T}\sum_{t=1}^T F_t^{rs}F_t^{pc'} \xrightarrow{P} 0$ and $\frac{1}{T}\sum_{t=1}^T F_t^{pc}F_t^{pc'} \xrightarrow{P} \Sigma^c$ we must have $A_{21} = 0$. Denoting the $k_i^s \times k_i^{rs}$ matrix A_{22} by $\mathcal{H}_{k_i^s}'$ we have

$$\frac{1}{T}\sum_{t=1}^T ||\hat{F}_{i,t}^s - \mathcal{H}_{k_i^s}' F_{i,t}^{rs}||^2 = O_p(C_{N,T}^{-2}). \tag{6.68}$$

Combining equations (6.67) (6.68) and (6.66) and using $\mathcal{H}_i^{0'} = \mathrm{diag}(\mathcal{H}_{k^c}', \mathcal{H}_{k_i^s}')$ we have

$$\frac{1}{T}\sum_{t=1}^T ||\hat{F}_{i,t}^0 - \mathcal{H}_i^{0'} F_{i,t}^0|| = O_p(C_{N,T}^{-2}).$$

□

Proof of Corollary 3.6

For the choice $k^c = r^c$ it holds $F^{rs}_{i,t} = F^s_{i,t}$. Hence we have

$$\frac{1}{T}\sum_{t=1}^{T} ||\hat{F}^s_{i,t} - \mathcal{H}'_{k^s_i} F^s_{i,t}||^2 = O_p(C^{-2}_{N,T}),$$

where $\mathcal{H}_{k^s_i}$ is now an $(r^s_i \times k^s_i)$ matrix with $rank(\mathcal{H}^s_i) = \min(k^s_i, r^s_i)$.

$\square$

In order to prove Theorem 3.7 we prove first the following two lemmata.

Lemma 6.10
For a given choice of (k^c, k^s_i) with $k^c < r_i$ and $k^s_i > 1$, let $V(k^s_i, \hat{F}^s_i|\hat{F}^c, \{\hat{\Lambda}^c_i\}, k^c)$ and $V(k^s_i, \tilde{F}^s_i|F^c, k^c)$ denote the mean squared errors of the principal component estimation based on $\hat{X}^{rs}_i$ and X^{rs}_i respectively:

$$V(k^s_i, \hat{F}^s_i|\hat{F}^c, k^c) = \min_{\Lambda^s_i,, F^s_i} \frac{1}{N_i T}\sum_{j=1}^{N_i}\sum_{t=1}^{T}((X_{i,jt} - \hat{F}^{c'}_t \hat{\lambda}^c_{i,j}) - F^{s'}_{i,t}\lambda^s_{i,j})^2, \tag{6.69}$$

$$V(k^s_i, \tilde{F}^s_i|F^c, k^c) = \min_{\Lambda^s_i,, F^s_i} \frac{1}{N_i T}\sum_{j=1}^{N_i}\sum_{t=1}^{T}((X_{i,jt} - F^{pc'}_t \lambda^{pc}_{i,j}) - F^{s'}_{i,t}\lambda^s_{i,j})^2, \tag{6.70}$$

Given a correct classification of the variables, the mean squared errors of the principal component estimations based on $\hat{X}^{rs}$ and X^{rs} differ no more than $O_p(C^{-1}_{N,T})$:

$$V(k^s_i, \tilde{F}^s_i|F^c, k^c) - V(k^s_i, \hat{F}^s_i|\hat{F}^c, k^c) = O_p(C^{-1}_{N,T}). \tag{6.71}$$

Proof:

$$\begin{aligned}
& \frac{1}{TN_i}||M_{\tilde{F}^s_i} X^{rs}_i||^2 \\
= & \frac{1}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s} + M_{\hat{F}^s_i})(X^{rs}_i - \hat{X}^{rs}_i + \hat{X}^{rs}_i)||^2 \\
= & \frac{1}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s_i})X^{rs}_i + M_{\hat{F}^s_i}(X^{rs}_i - \hat{X}^{rs}_i) + M_{\hat{F}^s_i}\hat{X}^{rs}_i||^2 \\
\leq & \frac{1}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s_i})X^{rs}_i||^2 + \frac{1}{TN_i}||M_{\hat{F}^s_i}(X^{rs}_i - \hat{X}^{rs}_i)||^2 + \frac{1}{TN_i}||M_{\hat{F}^s_i}\hat{X}^{rs}_i||^2 \\
& + \frac{2}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s_i})X^{rs}_i||||M_{\hat{F}^s_i}(X^{rs}_i - \hat{X}^{rs}_i)|| + \frac{2}{TN_i}||M_{\hat{F}^s_i}(X^{rs}_i - \hat{X}^{rs}_i)||||M_{\hat{F}^s_i}\hat{X}^{rs}_i|| \\
& + \frac{2}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s_i})X^{rs}_i||||M_{\hat{F}^s_i}\hat{X}^{rs}_i||
\end{aligned}$$

So we have

$$\begin{aligned}
& V(k^s_i, \tilde{F}^s_i|F^c, k^c) - V(k^s_i, \hat{F}^s_i|F^c, k^c) \\
= & \frac{1}{TN_i}||M_{\tilde{F}^s_i} X^{rs}_i||^2 - \frac{1}{TN_i}||M_{\hat{F}^s_i}\hat{X}^{rs}_i||^2 \\
\leq & \frac{1}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s_i})X^{rs}_i||^2 + \frac{1}{TN_i}||M_{\hat{F}^s_i}(X^{rs}_i - \hat{X}^{rs}_i)||^2 + \frac{2}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s_i})X^{rs}_i||||M_{\hat{F}^s_i}\hat{X}^{rs}_i)|| \\
& + \frac{2}{TN_i}||(M_{\tilde{F}^s_i} - M_{\hat{F}^s_i})X^{rs}_i||||M_{\hat{F}^s_i}(X^{rs}_i - \hat{X}^{rs}_i)|| + \frac{2}{TN_i}||M_{\hat{F}^s_i}(X^{rs}_i - \hat{X}^{rs}_i)||||M_{\hat{F}^s_i}\hat{X}^{rs}_i)|| \\
= & a + b + c + d + e
\end{aligned}$$

$$a = \frac{1}{TN_i}||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})X_i^{rs}||^2 \leq ||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})||^2 \Big|\Big| \frac{X_i^{rs}}{\sqrt{N_i T}} \Big|\Big|^2 = O_p(C_{N,T}^{-2})O_p(1) = O_p(C_{N,T}^{-2})$$

$$b = \frac{1}{N_i T}||M_{\hat{F}_i^s}(X_i^{rs} - \hat{X}_i^{rs})||^2 \leq ||M_{\hat{F}_i^s}||^2 \frac{1}{N_i T}||X_i^{rs} - \hat{X}_i^{rs}||^2 = O_p(C_{N,T}^{-2}).$$

$$\begin{aligned} c &= \frac{2}{TN_i}||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})X_i^{rs}|| ||M_{\hat{F}_i^s}\hat{X}_i^{rs})|| \\ &\leq 2\Big(\frac{1}{TN_i}||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})X_i^{rs}||^2 ||M_{\hat{F}_i^s}||^2 \frac{1}{TN_i}||\hat{X}_i^{rs})||^2\Big)^{\frac{1}{2}} \\ &= O_p(C_{N,T}^{-1})O_p(1) = O_p(C_{N,T}^{-1}). \end{aligned}$$

$$\begin{aligned} d &= \frac{2}{TN_i}||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})X_i^{rs}|| ||M_{\hat{F}_i^s}(X_i^{rs} - \hat{X}_i^{rs})|| \\ &\leq 2\Big(\frac{1}{TN_i}||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})X_i^{rs}||^2 \frac{1}{TN_i}||M_{\hat{F}_i^s}X_i^{rs} - \hat{X}_i^{rs}||^2\Big)^{\frac{1}{2}} \\ &= O_p(C_{N,T}^{-2}). \end{aligned}$$

$$\begin{aligned} e &= \frac{2}{TN_i}||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})X_i^{rs}|| ||M_{\hat{F}_i^s}(\hat{X}_i^{rs})|| \\ &\leq 2\Big(\frac{1}{TN_i}||(M_{\tilde{F}_i^s} - M_{\hat{F}_i^s})X_i^{rs}||^2 \frac{1}{TN_i}||M_{\hat{F}_i^s}X_i^{rs}||^2\Big)^{\frac{1}{2}} \\ &= O_p(C_{N,T}^{-1}). \end{aligned}$$

Putting the results together we have

$$V(k_i, \tilde{F}_i^s | F^c, k^c) - V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) \leq O_p(C_{N,T}^{-1}). \tag{6.72}$$

In the same way, we can show

$$V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) - V(k_i, \tilde{F}_i^s | F^c, k^c) \leq O_p(C_{N,T}^{-1}). \tag{6.73}$$

Therefore we have

$$V(k_i, \tilde{F}_i^s | F^c, k^c) - V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) = O_p(C_{N,T}^{-1}). \tag{6.74}$$

It follows for the weighted sum of the squared residuals we have

$$\sum_{i=1}^{n} \frac{N_i}{N} V(k_i, \tilde{F}_i^s | F^c, k^c) - \sum_{i=1}^{n} \frac{N_i}{N} V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) = O_p(C_{N,T}^{-1}). \tag{6.75}$$

Using the same arguments as given in the proof above we obtain the following result.

Corollary 6.11 *Given a correct classification of the variables, for a choice of* (k^c, k_i^s) *with* $k^c \leq r^c$ *and* $k_i^s > 1$*, the mean squared errors of the principal component estimations based on* $\hat{\hat{X}}_i^{rs}$ *and* $\hat{X}_i^{rs}$ *differ no more than* $O_p(C_{N,T}^{-1})$*:*

$$V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) - V(k_i^s, \hat{\hat{F}}_i^s | \hat{\hat{F}}^c, k^c) = O_p(C_{N,T}^{-1}). \tag{6.76}$$

Lemma 6.12
For a given choice of (k^c, k_i^s) with $k^c < k_i$ and $k_i^s > 1$, let the k^c-vector $\hat{F}_t^c$ be the estimate of the common factors given in Theorem 3.3 and $\hat{F}^s$ be the estimate of the group-specific factor given in Theorem 3.5. We have

(i) $\hat{F}^{c\prime}\hat{F}^s = 0$

(ii) $M_{\hat{F}^s}M_{\hat{F}^c} = M_{(\hat{F}^c,\hat{F}^s)}$

with $M_{\hat{F}^s} = I - \hat{F}^s(\hat{F}^{s\prime}\hat{F}^s)^{-1}\hat{F}^{s\prime}$, $M_{\hat{F}^c} = I - \hat{F}^c(\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\prime}$
$M_{(\hat{F}^c,\hat{F}^s)} = I - (\hat{F}^c,\hat{F}^s)((\hat{F}^c,\hat{F}^s)'(\hat{F}^c,\hat{F}^s))^{-1}(\hat{F}^c,\hat{F}^s)'$,

Proof of (i) According to the definition of $\hat{F}_i^s$ we have:

$$\hat{F}^{c\prime}\hat{F}_i^s = \hat{F}^{c\prime}\frac{\hat{X}_i^{rs}\hat{X}_i^{rs\prime}}{N_iT}\sqrt{T}Q_{k_i^s}(\hat{X}_i^{rs}\hat{X}_i^{rs\prime}) = \hat{F}^{c\prime}\frac{M_{\hat{F}^c}X_iX_i'M_{\hat{F}^c}}{N_iT}\sqrt{T}Q_{k_i^s}(\hat{X}_i^{rs}\hat{X}_i^{rs\prime}) = 0,$$

where $Q_{k_i^s}(\hat{X}_i^{rs}\hat{X}_i^{rs\prime})$ represents the k_i^s eigenvectors corresponding to the k_i^s largest eigenvalues of $(\hat{X}_i^{rs}\hat{X}_i^{rs\prime})$. Using the result $\hat{F}^{c\prime}\hat{F}^s = 0$ it is straight forward to verify $M_{\hat{F}^s}M_{\hat{F}^c} = M_{(\hat{F}^c,\hat{F}^s)}$. This proved (ii).
□

Proof of Theorem 3.7

We prove this theorem by showing first the following:

(i) Given a choice of $(k^c, \{k_i^s\})$ with $0 < k^c < r_i$ and $k_i^s \geq k_i^{rs}$, it holds
$$V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c) - V(r^c, r_i^s, F^c, F_i^s) = O_p(C_{N,T}^{-2}), \qquad (6.77)$$
with $V(r^c, r_i^s, F^c, F_i^s) = \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}'M_{(F^c,F_i^s)}e_{i,j}$ and $e_{i,j} = (e_{i,j1}, e_{i,j2}, ...e_{i,jT})'$.

(ii) Given a choice of $(k^c, \{k_i^s\})$ with $0 < k^c < r_i$ and $k_i^s < k_i^{rs}$ for some group i, then
$V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c) - V(r^c, r_i^s, F^c, F_i^s)$ has a positive lower bound.

(iii) $(\bar{h} + \bar{k}^s + \overline{\alpha}k^c) < (\bar{h} + \bar{r}^s + \overline{\alpha}r^c)$ implies that at least for one group, say group i, it holds $k_i^s < k_i^{rs}$.

(iv) For a give choice of $(k^c, \{k_i^s\})$ with $0 < k^c < r_i$ and $k_i^s \geq k_i^{rs}$ for $i = 1, 2, ..., n$, it holds $(\bar{h} + \bar{k}^s + \overline{\alpha}k^c) \geq (\bar{h} + \bar{r}^s + \overline{\alpha}r^c)$ and the equal sign holds only when $k^c = r^c$ and $k_i^s = r_i^s$ for $i = 1, 2, ..., n$.

Proof of (i)

$V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c)$ can be seen as the mean squared errors of the regression of $\hat{X}_i^{rs}$ on $\hat{F}_i^s$. Because $\hat{X}_i^{rs}$ are themselves the residuals of a linear regression of X_i on $\hat{F}^c$, and $\hat{F}^c$ are orthogonal to $\hat{F}_i^s$ according to Lemma 6.12, we can view the $V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c)$ as the residuals of a linear regression of X_i on $\hat{F}^c$ and $\hat{F}_i^s$:

$$\begin{aligned}
V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c) &= \frac{1}{N_iT}\,\mathrm{tr}\,(\hat{X}_i^{rs\prime}M_{\hat{F}_i^s}\hat{X}_i^{rs}) \\
&= \frac{1}{N_iT}\,\mathrm{tr}\,(X_i'M_{\hat{F}^c}M_{\hat{F}_i^s}M_{\hat{F}^c}X_i) \\
&= \frac{1}{N_iT}\,\mathrm{tr}\,(X_i'M_{(\hat{F}^s,\hat{F}^c)}X_i)
\end{aligned}$$

According to Proposition 3.4 for $k_i^s > k_i^{rs}$, $X_i = F^c\Lambda_i^c + F_i^s\Lambda_i^s + E_i$ can be decomposed according to a choice of k^c as follows.

$$\begin{aligned}
X_i &= F^c\Lambda_i^c + F_i^s\Lambda_i^s + E_i \\
&= F^{pc}\Lambda_i^{pc} + F_i^{rs}\Lambda_i^{rs} + E_i \\
&= F^{pc}\mathcal{H}_{k^c}\mathcal{H}_{k^c}^{+}\Lambda_i^{pc} + F_i^{rs}\mathcal{H}_{k_i^s}\mathcal{H}_{k_i^s}^{+}\Lambda_i^{rs} + E_i \\
&= \hat{F}^c\mathcal{H}_{k^c}^{+}\Lambda_i^{pc} + \hat{F}_i^s\mathcal{H}_{k_i^s}^{+}\Lambda_i^{rs} + E_i - (\hat{F}^c - F^{pc}\mathcal{H}_{k^c})\mathcal{H}_{k^c}^{+}\Lambda_i^{pc} - (\hat{F}_i^s - F_i^{rs}\mathcal{H}_{k_i^s})\mathcal{H}_{k_i^s}^{s+}\Lambda_i^{rs} \\
&= \hat{F}^c\mathcal{H}_{k^c}^{+}\Lambda_i^{pc} + \hat{F}_i^s\mathcal{H}_{k_i^s}^{+}\Lambda_i^{rs} + U_i \\
&= \hat{F}_i^{0'}\mathcal{H}_i^{0+}\Lambda_i^0 + U_i
\end{aligned}$$

with $U_i = E_i - (\hat{F}_i^0 - F_i^0\mathcal{H}_i^0)\mathcal{H}_i^{0+}\Lambda_i^0$, $\hat{F}_i^0 = (\hat{F}^c, \hat{F}_i^s)$, $\mathcal{H}_i^0 = diag(\mathcal{H}_{k^c}, \mathcal{H}_{k_i^s})$, $\mathcal{H}_i^{0+}$ is the (generalized) inverses of $\mathcal{H}_i^0$, $F_i^0 = (F^{pc}, F_i^{rs})$, $\Lambda_i^0 = \begin{pmatrix} \Lambda_i^{pc} \\ \Lambda_i^{rs} \end{pmatrix}$. By the similar technique used in the proof of Lemma 4 in Bai and Ng (2002) we can prove the result in (6.77). Concretely, we have

$$\begin{aligned}
& V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c) \\
= & \frac{1}{N_iT}\sum_{j=1}^{N_i} u_{i,j}' M_{\hat{F}_i^0} u_{i,j} \\
= & \frac{1}{N_iT}\sum_{j=1}^{N_i} [e_{i,j} - (\hat{F}_i^{0'} - F_i^0\mathcal{H}_i^0)\mathcal{H}_i^{0+}\Lambda_{i,j}^0]' M_{\hat{F}_i^0}[e_{i,j} - (\hat{F}_i^{0'} - F_i^0\mathcal{H}_i^0)\mathcal{H}_i^{0+}\Lambda_{i,j}^0] \\
= & \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' M_{\hat{F}_i^0} e_{i,j} - 2\frac{1}{N_iT}\sum_{j=1}^{N_i}\Lambda_{i,j}^{0'}\mathcal{H}_i^{0+'}(\hat{F}_i^{0'} - F_i^0\mathcal{H}_i^0)' M_{\hat{F}_i^{0'}} e_{i,j} \\
& + \frac{1}{N_iT}\sum_{j=1}^{N_i}\Lambda_{i,j}^{0'}\mathcal{H}_i^{0+'}(\hat{F}_i^{0'} - F_i^0\mathcal{H}_i^0)' M_{\hat{F}_i^0}(\hat{F}_i^{0'} - F_i^0\mathcal{H}_i^0)\mathcal{H}_i^{0+}\Lambda_{i,j}^0 \\
= & \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' M_{\hat{F}_i^0} e_{i,j} + b + c,
\end{aligned}$$

For the term c because $(I - M_{\hat{F}_i^0})$ is positive semi definite, $x'M_{\hat{F}_i^0}x \leq x'x$, we have

$$\begin{aligned}
c &\leq \frac{1}{N_iT}\sum_{j=1}^{N_i}\Lambda_{i,j}^{0'}\mathcal{H}_i^{0+'}(\hat{F}_i^{0'} - F_i^0\mathcal{H}_i^0)'(\hat{F}_i^{0'} - F_i^0\mathcal{H}_i^0)\mathcal{H}_i^{0+}\Lambda_{i,j}^0 \\
&\leq \left(\frac{1}{T}\sum_{t=1}^{T}\left|\left|\hat{F}_{i,t}^0 - \mathcal{H}_i^{0'}F_{i,t}^0\right|\right|^2\right)\left(N_i^{-1}\sum_{j=1}^{N_i}||\mathcal{H}_i^{0+}||^2||\Lambda_{i,j}^0||^2\right) \\
&= O_p(C_{N,T}^{-2})
\end{aligned}$$

by Theorem 3.5. For the term b we use that $|\mathrm{tr}(A)| \leq r||A||$ for any $(r \times r)$ matrix

A. Thus

$$
\begin{aligned}
b &= -2T^{-1}\,\mathrm{tr}\left(\mathcal{H}_i^{0+}(\hat{F}_i^{0\prime}-F_i^0\mathcal{H}_i^0)' M_{\hat{F}_i^{0\prime}}\left(N_i^{-1}\sum_{j=1}^{N_i} e_{i,j}\Lambda_{i,j}^0\right)\right)\\
&\leq -2r_i\|\mathcal{H}_i^{0+}\|\left\|\frac{\hat{F}_i^{0\prime}-F_i^0\mathcal{H}_i^0}{\sqrt{T}}\right\|\left\|\frac{1}{\sqrt{T}N_i}\sum_{j=1}^{N_i} e_{i,j}\Lambda_{i,j}^0\right\|\\
&\leq -2r_i\|\mathcal{H}_i^{0+}\|\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_{i,t}^0-\mathcal{H}_i^{0\prime}F_{i,t}^0\right\|^2\right)^{\frac{1}{2}}\frac{1}{\sqrt{N_i}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i} e_{i,jt}\Lambda_{i,j}^0\right\|^2\right)^{\frac{1}{2}}\\
&= O_p(C_{N,T}^{-1})\frac{1}{\sqrt{N_i}} = O_p(C_{N,T}^{-2})
\end{aligned}
$$

by Theorem 3.5 and Lemma 6.3 (iv). It follows

$$
V(k_i^s,\hat{F}_i^s|\hat{F}^c,k^c) = \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' M_{\hat{F}_i^0} e_{i,j} + O_p(C_{N,T}^{-2}).
$$

Using

$$
V(k^c,r_i^s,F^c,F_i^s) = \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' M_{(F^c,F_i^s)} e_{i,j},
$$

we have

$$
\begin{aligned}
& V(k_i^s,\hat{F}_i^s|\hat{F}^c,k^c) - V(k^c,r_i^s,F^c,F_i^s)\\
=\;& \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' M_{\hat{F}_i^0} e_{i,j} - \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' M_{F_i} e_{i,j} + O_p(C_{N,T}^{-2})\\
=\;& \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' P_{F_i} e_{i,j} - \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' P_{\hat{F}_i^0} e_{i,j} + O_p(C_{N,T}^{-2})
\end{aligned}
$$

For the first term in the last row above we have

$$
\begin{aligned}
& \frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}' P_{F_i} e_{i,j}\\
\leq\;& \|((F^c,F_i^s)'(F^c,F_i^s)/T)^{-1}\|\frac{1}{N_iT}\sum_{j=1}^{N_i} e_{i,j}'(F^c,F_i^s)(F^c,F_i^s)' e_{i,j}\\
\leq\;& O_p(1)\frac{1}{N_iT}\sum_{j=1}^{N_i}\|T^{-\frac{1}{2}}\sum_{t=1}^{T}\begin{pmatrix} F_t^c\\ F_{i,t}^s\end{pmatrix} e_{i,jt}\|^2 = O_p(T^{-1}) = O_p(C_{N,T}^{-2}).
\end{aligned}
$$

For the second term we have

$$
\begin{aligned}
& \frac{1}{N_iT}\sum_{j=1}^{N_i} e'_{i,j}P_{\hat{F}_i^0}e_{i,j} \\
\leq\; & ||(\hat{F}_i^{0\prime}\hat{F}_i^{0\prime}/T)^{-1}||\frac{1}{N_iT^2}\sum_{j=1}^{N_i} e'_{i,j}\hat{F}_i^{0\prime}\hat{F}_i^{0\prime}e_{i,j} \\
\leq\; & ||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\frac{1}{N_i}\sum_{j=1}^{N_i}\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{i,t}^0 e_{i,jt}\Big|\Big|^2 \\
=\; & ||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\frac{1}{N_i}\sum_{j=1}^{N_i}\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_{i,t}^0-\mathcal{H}_i^{0\prime}F_{i,t}^0)e_{i,jt}+\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_i^{0\prime}F_{i,t}^0 e_{i,jt}\Big|\Big|^2 \\
\leq\; & ||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\frac{1}{N_i}\sum_{j=1}^{N_i}\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_{i,t}^0-\mathcal{H}_i^{0\prime}F_{i,t}^0)e_{i,jt}\Big|\Big|^2 \\
& +||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\frac{1}{N_i}\sum_{j=1}^{N_i}\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_i^{0\prime}F_{i,t}^0 e_{i,jt}\Big|\Big|^2 \\
& +2||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_{i,t}^0-\mathcal{H}_i^{0\prime}F_{i,t}^0)e_{i,jt}\Big|\Big|^2\right)^{\frac{1}{2}}\cdot\left(\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_i^{0\prime}F_{i,t}^0 e_{i,jt}\Big|\Big|^2\right)^{\frac{1}{2}} \\
\leq\; & ||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{t=1}^{T}\Big|\Big|\hat{F}_{i,t}^0-\mathcal{H}_i^{0\prime}F_{i,t}^0\Big|\Big|^2\frac{1}{T}\sum_{t=1}^{T}||e_{i,jt}||^2 \\
& +||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\frac{1}{T}\frac{1}{N_i}\sum_{j=1}^{N_i}\Big|\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{H}_i^{0\prime}F_{i,t}^0 e_{i,jt}\Big|\Big|^2 \\
& +2||(\hat{F}_i^{0\prime}\hat{F}_i^{0}/T)^{-1}||\left(\frac{1}{T}\sum_{t=1}^{T}\Big|\Big|\hat{F}^0{}_{i,t}-\mathcal{H}_i^{0\prime}F_{i,t}^0\Big|\Big|^2\right)^{\frac{1}{2}}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{t=1}^{T}||e_{i,jt}||^2\right)^{\frac{1}{2}} \\
& \times\left(\frac{1}{T}\frac{1}{N_i}\sum_{j=1}^{N_i}\Big|\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{H}_i^{0\prime}F_{i,t}^0 e_{i,jt}\Big|\Big|^2\right)^{\frac{1}{2}} \\
=\; & O_p(1)O_p(C_{N,T}^{-2})O_p(1)+O_p(1)O_p(T^{-1})O_p(1)+O_p(1)O_p(C_{N,T}^{-1})O_p(1)O_p(T^{-1/2})O_p(1) \\
=\; & O_p(C_{N,T}^{-2})
\end{aligned}
$$

Hence we have for $k_i^s \geq k_i^{rs}$

$$V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c) - V(k^c, r_i^s, F^c, F_i^s) = O_p(C_{N,T}^{-2}).$$

Proof of (ii) One key insight from Bai and Ng (2002) (See Lemma 2 and Lemma 3 there.) is that in a factor model the difference between the mean squared errors of a factor estimation using principal component method and those of the true model is bounded by a positive number from below, if the estimated number of factors is less than the number of true factors. For $k^c > r^c$, applying this result to the model $X_i^{rs} = F^{rs}\Lambda^{rs} + E_i$, we know

$$V(k_i^s, \tilde{F}_i^s|F^{pc}, \Lambda_i^{pc}, k^c) - V(k_i^{rs}, F_i^{rs}|F^{pc}, \Lambda_i^{pc}, k^c). \tag{6.78}$$

is bounded from below by a positive number for the case of $k_i^s < k_i^{rs}$. According to Lemma 6.10 we have

$$V(k_i^s, \tilde{F}_i^s | F^{pc}, k^c) - V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) = O_p(C_{N,T}^{-1}). \tag{6.79}$$

If follows

$$V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) - V(k_i^{rs}, F_i^{rs} | F^{pc}, \Lambda_i^{pc}, k^c). \tag{6.80}$$

is bounded by a positive number. Further we have

$$\begin{aligned}
& V(r^c, r_i^s, F^c, F_i^s) - V(k_i^{rs}, F_i^{rs} | F^{pc}, \Lambda_i^{pc}, k^c) \\
= & \frac{1}{N_i T} \sum_{j=1}^{N_i} e_{i,j}' M_{F_i} e_{i,j} - \frac{1}{N_i T} \sum_{j=1}^{N_i} e_{i,j}' M_{F_i^0} e_{i,j} \\
& = \frac{1}{N_i T} \sum_{j=1}^{N_i} e_{i,j}' P_{F_i^0} e_{i,j} - \frac{1}{N_i T} \sum_{j=1}^{N_i} e_{i,j}' P_{F_i} e_{i,j} \\
\leq & ||((F_i^0)'(F_i^0)/T)^{-1}|| \frac{1}{N_i T} \sum_{j=1}^{N_i} e_{i,j}' (F_i^0)(F_i^0)' e_{i,j} - \frac{1}{N_i T} \sum_{j=1}^{N_i} e_{i,j}' P_{F_i} e_{i,j} \\
\leq & ||((F_i^0)'(F_i^0)/T)^{-1}|| \frac{1}{N_i T} \sum_{j=1}^{N_i} ||T^{-\frac{1}{2}} \sum_{t=1}^{T} F_{i,t}^0 e_{i,jt}||^2 - \frac{1}{N_i T} \sum_{j=1}^{N_i} e_{i,j}' P_{F_i} e_{i,j} \\
= & O_p(T^{-1}) - O_p(T^{-1}) = O_p(C_{N,T}^{-2})
\end{aligned}$$

If follows

$$V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) - V(r^c, r_i^s, F^c, F_i^s). \tag{6.81}$$

is bounded by a positive number from below.

For the case of $k^c \leq r^c$ and $k_i^s < k_i^{rs}$, we have $k_i = k^c + k_i^s < r^c + (r^c - k^c) + r_i^s = r_i$. It follows $V(\hat{F}_i, k_i) - V(F_i, r_i) = V(\hat{F}_i, k_i) - V(r^c, r_i^s, F^c, F_i^s)$ is bounded from below by a positive number. Because

$$\begin{aligned}
& V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) - V(r^c, r_i^s, F^c, F_i^s) \\
= & V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) - V(k_i^s, \hat{\hat{F}}_i^s | \hat{\hat{F}}^c, k^c) + V(k_i^s, \hat{\hat{F}}_i^s | \hat{\hat{F}}^c, k^c) - V(\hat{F}_i, k_i) \\
& + V(\hat{F}_i, k_i) - V(r^c, r_i^s, F^c, F_i^s) \\
= & O_p(C_{N,T}^{-1}) + V(\hat{F}_i, k_i) - V(r^c, r_i^s, F^c, F_i^s)
\end{aligned}$$

Therefor, for $k^c \leq r^c$ and $k_i^s < k_i^{rs}$, the difference $V(k_i^s, \hat{F}_i^s | \hat{F}^c, k^c) - V(r^c, r_i^s, F^c, F_i^s)$ is bounded from below by a positive number. □

Proof of (iii)

We prove (iii) by showing that $k_i^s > k_i^{rs}$ for all i implies $\bar{k}^s + \overline{\alpha} k^c \geq \bar{r}^s + \overline{\alpha} r^c$. Since $k_i^s > k_i^{rs}$ for all i implies $k_i^s + k^c > r^c + r_i^s$ for all i, we have

$$\bar{k}^s + k^c \geq \bar{r}^s + r^c. \tag{6.82}$$

For cases with $k^c < r^c$, we have $(\overline{\alpha} - 1)k^c > (\overline{\alpha} - 1)r^c$. Combining this inequality with (6.82) we obtain:

$$\bar{k}^s + \overline{\alpha} k^c > \bar{r}^s + \overline{\alpha} r^c.$$

For cases with $k^c > r^c$, we look at each component in F^{pc}. Since F^{pc} does not contain any common factors across all groups, each component in F^{pc} will be linearly independent from the group-specific factors of at least one group, say groups i. Corresponding to this component of F^{pc}, we will have $k_i^s + k^c \geq r^c + r_i^s + 1$. Consequently due to this component in F^{pc} the average number of factors of the model $[k^c, \{k_j^s\}]$ will be greater than the average number of factors of the true model by at least α_i. For all $k^c - r^c$ components of F^{pc}, the average number of factors of the model $[k^c, \{k_j^s\}]$ will be greater than the average number of factors of the true model by at least $\min_i(\{\alpha_i\})(k^c - r^c)$, i.e.

$$\bar{k}^s + k^c - (\bar{r}^s + r^c) \geq \min_i(\{\alpha_i\})(k^c - r^c).$$

Rearranging the inequality we obtain:

$$\bar{k}^s + (1 - \min_i(\{\alpha_i\}))k^c \geq \bar{r}^s + (1 - \min_i(\{\alpha_i\}))r^c.$$

Because $k^c > r^c$ and $\underline{\alpha} > (1 - \min_i(\{\alpha_i\}))$, it follows from the inequality above:

$$\bar{k}^s + \overline{\alpha}k^c > \bar{r}^s + \overline{\alpha}r^c.$$

Therefore, if $(\bar{h} + \bar{k}^s + \overline{\alpha}k^c) < (\bar{h} + \bar{r}^s + \overline{\alpha}r^c)$, there must be at least one group for which it holds $k_i^s < k_i^{rs}$.

□

Proof of (iv)

From the proof of (iii) above we know that if $k_i^s \geq k_i^{rs}$ and $k^c > r^c$, we will have $(\bar{k}^s + \bar{h} + \overline{\alpha}k^c) > (\bar{r}^s + \bar{h} + \overline{\alpha}r^c)$.

So we need only to consider the case of $k^c \leq r^c$. From $\bar{k}^s + \overline{\alpha}k^c = \bar{r}^s + \overline{\alpha}r^c$, we have

$$\sum \frac{N_i}{N}(k_i^s + k^c - r^c - r_i^s) + (1 - \overline{\alpha})(r^c - k^c) = 0$$

Because $(k_i^s + k^c - r^c - r_i^s) \geq 0$ and $r^c - k^c \geq 0$ the equality above holds only when $k^c = r^c$ and $k_i^s = r_i^s$ for $i = 1, 2, ..., n$.

□

Proof of Theorem 3.7

Now we compare the values of model selection criterion of the true model with a candidate model with $k^c \neq r^c$ or $k_i^s \neq r_i^s$ for some i under the assumption that the classification of variable is correct.

$$\begin{aligned} & P\left(C(n, k^c, \{k_i^s\}, \{X_i\}) < C(n, r^c, \{r_i^s\}, \{X_i\})\right) \qquad (6.83) \\ = \; & P\left(\sum_{i=1}^{n}(V(k_i^s, \hat{F}_i^s|\hat{F}^c, k^c) - V(r_i^s, \hat{F}_i^s|\hat{F}^c, r^c)) < ((\bar{r}^s + \overline{\alpha}r^c) - (\bar{k}^s + \overline{\alpha}k^c))g(N,T)\right) \end{aligned}$$

If $k_i^s < k_i^{rs}$ for some group i, according to (ii) the left hand side of the probability inequality (6.83) will be positive with a non-zero lower bound. The right hand side will converge to zero. Hence we have

$$P(C(n, k^c, \{k_i^s\}, \{X_i\}) < C(n, r^c, \{r_i^s\}, \{X_i\})) \to 0$$

If $k_i^s \geq k_i^{rs}$ and $k^c \neq r^c$ and $k_i^s \neq r_i^s$ for $i = 1, 2, ..., n$, according to (i) the left hand side of the probability inequality (6.83) will be $O_p(C_{N,T}^{-2})$. The right hand side will be negative according to (iv) and converges to zero at a slower rate than $O_p(C_{N,T}^{-2})$. Hence we have

$$P\left(C(n,k^c,\{k_i^s\},\{X_i\}) < C(n,r^c,\{r_i^s\},\{X_i\})\right) \to 0$$

Sofar we have proved equation (6.83) under the condition that the classification of variables is correct:

$$\begin{aligned} & P(C(n,k^c,\{k_i^s\},\{X_i^{s_n}\}) < C(n,r^c,\{r_i^s\},\{X_i^{s_n}\})|CC) \\ = \ & P\left(C(n,k^c,\{k_i^s\},\{X_i\}) < C(n,r^c,\{r_i^s\},\{X_i\})\right) \to 0 \quad \text{as } T\to\infty, N\to\infty, \end{aligned}$$

where CC represent the event of a correct classification. According to Chen (2010) we have $P(CC) \to 1 \ as \ T \to \infty, N \to \infty$.

$$\begin{aligned} & P\Big(C(n',k^c,\{k_i^s\},\{X_i^{s_{n'}}\}) < C(n,r^c,\{r_i^s\},\{X_i^{s_n}\})\Big) \\ = \ & P\Big(C(n',k^c,\{k_i^s\},\{X_i^{s_{n'}}\}) < C(n,r^c,\{r_i^s\},\{X_i^{s_n}\}), CC\Big) \\ & +P\Big(C(n',k^c,\{k_i^s\},\{X_i^{s_{n'}}\}) < C(n,r^c,\{r_i^s\},\{X_i^{s_n}\}), \overline{CC}\Big) \\ = \ & P\Big(C(n,k^c,\{k_i^s\},\{X_i^{s_n}\}) < C(n,r^c,\{r_i^s\},\{X_i^{s_n}\})|CC\Big)P(CC) \\ & +P\Big(C(n',k^c,\{k_i^s\},\{X_i^{s_{n'}}\}) < C(n,r^c,\{r_i^s\},\{X_i^{s_n}\})|\overline{CC}\Big)P(\overline{CC}) \\ \to \ & 0 \end{aligned}$$

The convergence is due to $P\Big(C(n,k^c,\{k_i^s\},\{X_i^{s_n}\}) < C(n,r^c,\{r_i^s\},\{X_i^{s_n}\})|CC\Big) \to 0$ and $P(\overline{CC}) = 1 - P(CC) \to 0$.

□

6.2 Variable List of the Empirical Example

Table 7: List of Variables and Classification

Group	No.	Name	code
1	4	AGL ENERGY -	A:AGKX(RI)(*)
1	5	AUSTRALIAN INFR.FUND -	A:AIXX(RI)
1	7	ARISTOCRAT LEISURE -	A:ALLX(RI)
1	8	ALESCO -	A:ALSX(RI)
1	10	AMCOR -	A:AMCX(RI)
1	11	AMP - TOT RETURNIND	A:AMPX(RI)
1	12	ANSELL -	A:ANNX(RI)
1	13	AUS.AND NZ.BANKING GP. -	A:ANZX(RI)
1	15	APA GROUP -	A:APAX(RI)(*)
1	16	APN NEWS & MEDIA-	A:APNX(RI)
1	19	ASX - TOT RETURNIND	A:ASXX(RI)
1	20	AUSTAR UNITED COMMS. -	A:AUNX(RI)
1	22	AWB - TOT RETURNIND	A:AWBX(RI)
1	25	AXA ASIA PACIFICHDG. -	A:AXAX(RI)
1	26	BILLABONG INTERNATIONAL -	A:BBGX(RI)
1	27	BENDIGO & ADELAIDE BANK -	A:BENX(RI)
1	29	BORAL -	A:BLDX(RI)
1	30	BANK OF QLND. -	A:BOQX(RI)
1	33	BUNNINGS WHSE.PR.TST. -	A:BWPX(RI)
1	34	BRAMBLES -	A:BXBX(RI)
1	35	CABCHARGE AUSTRALIA -	A:CABX(RI)
1	36	COMMONWEALTH BK.OF AUS. -	A:CBAX(RI)
1	37	COCA-COLA AMATIL-	A:CCLX(RI)
1	40	CFS RETAIL PR.TST. -	A:CFXX(RI)
1	41	CHALLENGER FINL.SVS.GP. -	A:CGFX(RI)
1	42	CONSOLIDATED MEDIA HDG. -	A:CMJX(RI)
1	43	COCHLEAR -	A:COHX(RI)
1	44	COMMONWEALTH PR.OFFE.FD. -	A:CPAX(RI)
1	45	COMPUTERSHARE -	A:CPUX(RI)
1	48	CSR - TOT RETURNIND	A:CSRX(RI)
1	49	CALTEX AUSTRALIA-	A:CTXX(RI)(*)
1	51	CORPORATE EXPRESS AUS. -	A:CXPX(RI)
1	52	DAVID JONES -	A:DJSX(RI)
1	55	DEXUS PROPERTY GROUP -	A:DXSX(RI)
1	56	ELDERS -	A:ELDX(RI)
1	57	ENVESTRA -	A:ENVX(RI)
1	63	FOSTER'S GROUP -	A:FGLX(RI)
1	65	FLIGHT CENTRE -	A:FLTX(RI)
1	68	FAIRFAX MEDIA -	A:FXJX(RI)
1	70	GOODMAN GROUP -	A:GMGX(RI)
1	72	GUNNS -	A:GNSX(RI)
1	73	GPT GROUP -	A:GPTX(RI)
1	75	GWA INTERNATIONAL -	A:GWTX(RI)
1	76	HENDERSON GROUP CDI. -	A:HGGX(RI)
1	77	HILLS INDUSTRIES-	A:HILX(RI)
1	78	HEALTHSCOPE -	A:HSPX(RI)
1	79	HARVEY NORMAN HOLDINGS -	A:HVNX(RI)
1	80	INSURANCE AUS.GROUP -	A:IAGX(RI)
1	81	IOOF HOLDINGS -	A:IFLX(RI)
1	83	ING INDL.FUND -	A:IIFX(RI)
1	85	ING OFFICE FUND -	A:IOFX(RI)
1	88	ISOFT GROUP -	A:ISFX(RI)
1	91	JAMES HARDIE INDS.CDI. -	A:JHXX(RI)
1	95	LEND LEASE GROUP-	A:LLCX(RI)
1	98	MAP GROUP -	A:MAPX(RI)
1	101	MACQUARIE COUNTRY.TRUST -	A:MCWX(RI)

Table 8: List of Variables and Classification(Cont.)

Group	No.	Name	code
1	102	MIRVAC GROUP -	A:MGRX(RI)
1	104	MACQUARIE INFR.GROUP -	A:MIGX(RI)
1	108	MACQUARIE OFFICETRUST -	A:MOFX(RI)
1	110	MACQUARIE GROUP -	A:MQGX(RI)(*)
1	112	METCASH -	A:MTSX(RI)
1	113	NATIONAL AUS.BANK -	A:NABX(RI)
1	116	NEWS CORP.CDI.'B' (ASX) -	A:NWSX(RI)
1	129	PERPETUAL -	A:PPTX(RI)
1	130	PAPERLINX -	A:PPXX(RI)
1	131	PRIMARY HEALTH CARE -	A:PRYX(RI)
1	132	QANTAS AIRWAYS -	A:QANX(RI)
1	133	QBE INSURANCE GROUP -	A:QBEX(RI)
1	134	RAMSAY HEALTH CARE -	A:RHCX(RI)
1	137	RESMED CDI -	A:RMDX(RI)
1	143	STOCKLAND -	A:SGPX(RI)
1	144	SINGAPORE TELECOM CDI. (ASX) -	A:SGTX(RI)
1	145	SONIC HEALTHCARE-	A:SHLX(RI)
1	146	SIGMA PHARMS. -	A:SIPX(RI)
1	148	SPOTLESS GROUP -	A:SPTX(RI)
1	151	SUNCORP-METWAY -	A:SUNX(RI)
1	153	TRANSURBAN GROUP-	A:TCLX(RI)
1	154	TELECOM CORP.NZ.(ASX) -	A:TELX(RI)
1	155	TEN NETWORK HOLDINGS -	A:TENX(RI)
1	156	TELSTRA -	A:TLSX(RI)
1	157	TOLL HOLDINGS -	A:TOLX(RI)
1	158	TRANSFIELD SERVICES -	A:TSEX(RI)
1	159	UGL - TOT RETURNIND	A:UGLX(RI)
1	160	VIRGIN BLUE HOLDINGS -	A:VBAX(RI)
1	161	WEST AUST.NWSP.HDG. -	A:WANX(RI)
1	162	WESTPAC BANKING -	A:WBCX(RI)
1	163	WESTFIELD GROUP -	A:WDCX(RI)
1	166	WOOLWORTHS -	A:WOWX(RI)
1	1	AUSTRALIAN AGRICULTURAL -	A:AACX(RI)
1	2	ADELAIDE BRIGHTON -	A:ABCX(RI)
1	3	ABACUS PROPERTY GROUP -	A:ABPX(RI)
1	9	AUSTRALAND PR.GP. -	A:ALZX(RI)
1	23	ALUMINA -	A:AWCX(RI)
1	32	BLUESCOPE STEEL -	A:BSLX(RI)
1	39	CENTENNIAL COAL -	A:CEYX(RI)
1	46	CRANE GROUP -	A:CRGX(RI)
1	54	DOWNER EDI -	A:DOWX(RI)
1	64	FKP PROPERTY GROUP -	A:FKPX(RI)
1	67	FLEETWOOD -	A:FWDX(RI)
1	71	GRAINCORP -	A:GNCX(RI)
1	74	GUD HOLDINGS -	A:GUDX(RI)
1	87	IRESS MARKET TECH. -	A:IREX(RI)
1	90	JB HI-FI -	A:JBHX(RI)
1	93	LEIGHTON HOLDINGS -	A:LEIX(RI)
1	96	LYNAS -	A:LYCX(RI)
1	97	MACMAHON HOLDINGS -	A:MAHX(RI)
1	99	MACARTHUR COAL -	A:MCCX(RI)
1	106	MURCHISON METALS-	A:MMXX(RI)
1	107	MONADELPHOUS GROUP -	A:MNDX(RI)
1	111	MINARA RESOURCES-	A:MREX(RI)
1	115	NUFARM -	A:NUFX(RI)
1	122	ONESTEEL -	A:OSTX(RI)(*)
1	126	PRIME INFRASTRUCTURE GP. -	A:PIHX(RI)
1	147	SMS MAN.& TECH. -	A:SMXX(RI)(*)
1	152	TABCORP HOLDINGS-	A:TAHX(RI)

Table 9: List of Variables and Classification (Cont.)

Group	No.	Name	code
2	47	CSL - TOT RETURNIND	A:CSLX(RI)(*)
2	84	ILUKA RESOURCES -	A:ILUX(RI)
2	141	SEVEN NETWORK -	A:SEVX(RI)(*)
2	6	AJ LUCAS GROUP -	A:AJLX(RI)
2	14	ARROW ENERGY -	A:AOEX(RI)
2	17	AQUILA RESOURCES-	A:AQAX(RI)
2	18	AQUARIUS PLATINUM (ASX) -	A:AQPX(RI)
2	21	AVOCA RESOURCES -	A:AVOX(RI)
2	24	AWE - TOT RETURNIND	A:AWEX(RI)
2	28	BHP BILLITON -	A:BHPX(RI)
2	31	BEACH ENERGY -	A:BPTX(RI)
2	38	CUDECO -	A:CDUX(RI)
2	50	CARNARVON PETROLEUM -	A:CVNX(RI)
2	53	DOMINION MINING -	A:DOMX(RI)
2	58	EQUINOX MINERALSCDI. -	A:EQNX(RI)
2	59	ENERGY RES.OF AUS. -	A:ERAX(RI)
2	60	EASTERN STAR GAS-	A:ESGX(RI)
2	61	ENERGY WORLD -	A:EWCX(RI)
2	62	EXTRACT RESOURCES -	A:EXTX(RI)
2	66	FORTESCUE METALSGP. -	A:FMGX(RI)
2	69	GINDALBIE METALS-	A:GBGX(RI)
2	82	INDEPENDENCE GROUP -	A:IGOX(RI)
2	86	INCITEC PIVOT -	A:IPLX(RI)(*)
2	89	INVOCARE -	A:IVCX(RI)(*)
2	92	KINGSGATE CONSOLIDATED -	A:KCNX(RI)
2	94	LIHIR GOLD -	A:LGLX(RI)
2	100	MINCOR RESOURCES-	A:MCRX(RI)
2	103	MOUNT GIBSON IRON -	A:MGXX(RI)
2	105	MEDUSA MINING -	A:MMLX(RI)
2	109	MOLOPO ENERGY -	A:MPOX(RI)
2	114	NEWCREST MINING -	A:NCMX(RI)
2	117	NEXUS ENERGY -	A:NXSX(RI)
2	118	OM HOLDINGS -	A:OMHX(RI)
2	119	ORIGIN ENERGY (EX BORAL) -	A:ORGX(RI)
2	120	ORICA -	A:ORIX(RI)(*)
2	121	OIL SEARCH -	A:OSHX(RI)
2	123	OZ MINERALS -	A:OZLX(RI)
2	124	PANORAMIC RESOURCES -	A:PANX(RI)
2	125	PALADIN ENERGY -	A:PDNX(RI)
2	127	PLATINUM AUSTRALIA -	A:PLAX(RI)
2	128	PANAUST -	A:PNAX(RI)
2	135	RIO TINTO -	A:RIOX(RI)
2	136	RIVERSDALE MINING -	A:RIVX(RI)
2	138	ROC OIL COMPANY -	A:ROCX(RI)
2	139	ST BARBARA -	A:SBMX(RI)
2	140	SUNDANCE RESOURCES -	A:SDLX(RI)
2	142	SIMS METAL MANAGEMENT -	A:SGMX(RI)
2	149	STRAITS RESOURCES -	A:SRLX(RI)
2	150	SANTOS -	A:STOX(RI)
2	164	WESFARMERS -	A:WESX(RI)(*)
2	165	WORLEYPARSONS -	A:WORX(RI)
2	167	WOODSIDE PETROLEUM -	A:WPLX(RI)
2	168	WESTERN AREAS -	A:WSAX(RI)

References

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

Campell, J. Y., Lo, A. W., and Mackenlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, 1st edition.

Chen, P. (2010). A grouped factor model. *Mimeo, Melbourne University, Faculty of Business and Economics*.

Fama, E. and French, K. (1993). Common risk in the returns on stocks and bonds. *Jouranl of Financial Economics*, 33:3–56.

Flury, B. (1984). Common principal components in groups. *Journal of the American Statistical Association*, 79:892–898.

— (1987). Two generalizations of the common principal component model. *Biometrika*, 62:59–69.

Goyal, A., Perignon, C., and Villa, C. (2008). How common are common return factors across nyse and nasdaq? *Journal of Financial Economics*, 90:252–271.

Hasan, M. A. (June 2009). Stable algorithm for multiset canonical correlations analysis. *2009 American Control Conference*, page 1280.

Jackson, J. E. (2005). Promax rotation. *Encyclopedia of Biostatistics*, Second Edition.

Johansen, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall International, 3rd edition.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall International, 3rd edition.

Kaiser, H. F. (1958). The varimax ctiterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200.

Krzanowski, W. J. (1979). Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74:703–707.

L., H. S. and Rouwenborst, K. G. (1994). Does industry structure explain the benifits on international diversification? *Journal of Financial Economics*, 36:3–7.

Nielsen, A. A. (2002). Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. *IEEE Transactions On Image Processing*, 11 No. 3:293–305.

Nishi, R. (1984). Asymptotical properties of criteria for selection of variables in multiple regression. *Ann. Statist.*, 12:758–765.

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

Schott, J. (1999). Partial common principal component subspaces. *Biometrika*, 86:899–908.

Truhar, N. (2000). Relative perturbation theory for matrix spectral decompositions. *UNIVERSITY OF ZAGREB DEPARTMENT OF MATHEMATICS*.