A Grouped Factor Model

Chen, Pu

Melbourne University

1 October 2010

A Grouped Factor Model

Pu Chen*

11.10.2011

Abstract

In this paper we present a grouped factor model that is designed to explore clustering structures in large factor models. We develop a procedure that will endogenously assign variables to groups, determine the number of groups, and estimate common factors for each group. The grouped factor model provides not only an alternative way to factor rotations in discovering orthogonal and non-orthogonal clusterings in a factor space. It offers also an effective method to explore more general clustering structures in a factor space which are invisible by factor rotations: the factor space can consist of subspaces of various dimensions that may be disjunct, orthogonal, or intersected with any angels. Hence a grouped factor model may provide a more detailed insight into data and thus also more understandable and interpretable factors.

KEYWORDS: Factor Models, Generalized Principal Component Analysis

JEL Classification: C1, C32, E24,

*Melbourne Institute of Technology, 154 Sussex Street, NSW 2000, Australia, E-Mail: pchen@academics.mit.edu.au

1 Introduction

Factor models are widely used to summarize common features in large data sets, such that behaviors of a large number of observed variables can be explained by a small number of unobserved variables called factors. This class of models has been successfully applied, for example, in finance to model asset returns known as arbitrage pricing theory (see Ross (1976) for more details), in applied macroeconomics to construct coincident indices to describe business cycles and to forecast macroeconomic activities (see Stock and Watson (2002) for more details), and in marketing to identify the salient attributes with which consumers evaluate products. Often the large number of variables consist of variables from different groups. For example asset returns consist of asset returns of different industries; macroeconomic variables include usually price variables, real activity measures, interest rates, labour statistics ect; consumers can be classified into different profession groups, income classes, and age groups ect. Group-specific information is useful in understanding data, in particular, in explaining group-specific features in the data. So, for example, industrial indices which 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.). Regarding forecasting Boivin and Ng (2006) find factors extracted from grouped data can produce better forecasts. Ludvigson and Ng (2009) analyze the relationship between bond excess returns and macro economic variables. They use 8 group-pervasive factors extracted from 131 variables to explain the bond risk premia. Goyal, Perignon, and Villa (2008) apply a factor model with two groups to NYSE and NASDAQ data and find that these two markets share one common factor and each market has one group-specific factor respectively. Heaton and Solo (2009) study a grouped factor model in which the groups are characterized by non-vanishing cross-sectional correlation among the residuals within a group.

In most studies using large factor models with groups, grouping of variables is assumed to be known a priori. Often the *a priori* assumptions on groups are based on structural information, such as geographical locations, organizational memberships or demographical characteristics. Although in many cases the non-statistical structural information provides a natural criterion to classify the variables under consideration, this kind of classifications, however, do not necessarily reflect the statistical properties of the variables. Consequently, the statistical inference based on this kind of classifications might be biased or inefficient.

It raises questions: How can we know whether a classification according to certain structural information is statistically adequate? How can we achieve a statistically valid classification of the variables if there are indeed some grouped structures in the variables? How can we determine the number of groups? How can we characterize the grouped structures? and what are the conditions under which we can obtain a valid estimation of the group-pervasive factors? Answering these questions constitutes the contents of this paper. Our paper contributes to the literature on large factor models in that it presents a theory on grouping the variables, determination of the number of groups and estimation of the group-pervasive factors, such that the grouped structures are statistically determined from observed data rather than assumed to be known *a priori*. Our consistent classification rule, consistent model selection criteria and consistent estimation of factors are developed under large cross

sections (N) and large time dimensions (T) without any restriction on the relation between T and N.

The paper is organized as follows. In section 2 we define a grouped factor model and discuss its relation to a conventional factor model. Section 3 deals with estimation of grouped factor models. We establish a consistent classification rule to classify variable into groups based on a method called generalized principal component analysis (GPCA). We present a class of consistent model selection criteria to determine the number of groups as well as the number of factors in each group. Section 4 documents some simulation studies on the performance of the estimation procedure for grouped factor models in finite sample settings. After we demonstrate an empirical application of the grouped factor model in section 5, the last section concludes.

2 The Model

Let X be a $(T \times N)$ matrix collecting the observations of a set of N variables observed over T periods. We assume that this set of variables 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 variables in each group are generated from a factor model. For the jth variable of the ith group at time t we have

$$\underset{(1\times 1)}{X_{i,jt}} = \underset{(1\times k_i)}{\lambda'_{i,j}} \underset{(k_i\times 1)}{F_{i,t}} + \underset{(1\times 1)}{e_{i,jt}}, \qquad \text{for } j = 1, 2, ...N_i, t = 1, 2, ..., T, i = 1, 2, ..., n, \tag{2.2}$$

where $F_{i,t}$ is a k_i-dimensional random factor of the ith group at time t and $\lambda_{i,j}$ is a k_i-dimensional factor loading for the jth variable in the ith group. $e_{i,jt}$ is the idiosyncratic component of $X_{i,jt}$ and $\lambda'_{i,j}F_{i,t}$ is the common component of $X_{i,jt}$. $F_{i,t}$ is called group-pervasive factor of the ith group.

Let $X_{i,j}$ collect the time series observations of $X_{i,jt}$ over T periods. We have

$$\underset{(T\times 1)}{X_{i,j}} = \underset{(T\times k_i)}{F_i} \underset{(k_i\times 1)}{\lambda_{i,j}} + \underset{(T\times 1)}{e_{i,j}}, \qquad \text{for } j = 1, 2, ...N_i, i = 1, 2, ..., n, \tag{2.3}$$

where $X_{i,j} = (X_{i,j1}, X_{i,j1}, ..., X_{i,jT})'$, $F_i = (F_{i,1}, F_{i,2}, ..., F_{i,T})'$, and $e_{i,j} = (e_{i,j1}, e_{i,j1}, ..., e_{i,jT})'$.

Let X_i collect observations of all N_i variables in the ith group. We have

$$\underset{(T\times N_i)}{X_i} = \underset{(T\times k_i)}{F_i} \underset{(k_i\times N_i)}{\Lambda_i} + \underset{(T\times N_i)}{E_i}, \qquad \text{for } i = 1, 2, ..., n, \tag{2.4}$$

where

- $X_i = (X_{i,1}, X_{i,2}, ..., X_{i,N_i})$: $(T \times N_i)$ matrix of observations of N_i variables in the ith group over T periods.
- F_i: $(T \times k_i)$ matrix of unobservable k_i group-pervasive factors of the ith group over T periods.
- $\Lambda_i = (\lambda_{i,1}, \lambda_{i,2}, ..., \lambda_{i,N_i})$: $(k_i \times N_i)$ matrix of unobservable factor loadings of the ith group.

- $E_i = (e_{i,1}, e_{i,2}, ..., e_{i,N_i})$: $(T \times N_i)$ matrix of unobservable idiosyncratic components of the ith group over T periods.
- $\sum_{i=1}^{n} N_i = N$.

We call the model in (2.4) a grouped factor model (GFM).

2.1 Assumptions

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

Assumption 2.1
(a) A group-pervasive factor $F_{i,t}$ is a linear function of a k dimensional random vector G_t with $k \leq \sum_{i=1}^{n} k_i$ in the following way:

$$F_{i,t} = C_i' G_t, \qquad \text{for } i = 1, 2, ..., n, \tag{2.5}$$

where C_i is a $(k \times k_i)$ constant matrix.

(b) $rank(C_i) = k_i$.

(c) $rank(C_1, C_2, ..., C_n) = k$.

Assumption 2.1 (a) is made to allow for possible dependence among group-pervasive factors across groups. If $k < \sum_{i=1}^{n} k_i$, some components of group-pervasive factors must be linearly dependent across groups. For instance, with $n = 3$, $k_1 = 2$ and $k_2 = 2$, $k_3 = 1$ and $k = 3$ we are considering three groups with 2, 2 and 1 group-pervasive factors respectively. These five components of the three group-pervasive factors are not independent from each other. Only three components are independent and they are represented by a three dimensional random vector G_t. Then each group-pervasive factor can be represented as a linear function of G_t. If $k = \sum_{i=1}^{n} k_i$, G_t is just the collection of all group-pervasive factors possibly after some rotations. Assumption 2.1 (b) is made to ensure group-pervasive factors are not linearly dependent within a group. (c) is to make sure that every component of G_t is used in generating the group-pervasive factors. Under Assumption 2.1, X adopts a factor structure with G as the factor:

$$\begin{aligned} X &= \begin{pmatrix} X_1 & X_2 & \dots & X_n \end{pmatrix} \\ &= \begin{pmatrix} F_1\Lambda_1 & F_2\Lambda_2 & \dots & F_n\Lambda_n \end{pmatrix} + \begin{pmatrix} E_1 & E_2 & \dots & E_n \end{pmatrix} \\ &= \begin{pmatrix} GC_1\Lambda_1 & GC_2\Lambda_2 & \dots & GC_n\Lambda_n \end{pmatrix} + \begin{pmatrix} E_1 & E_2 & \dots & E_n \end{pmatrix} \\ &= G \begin{pmatrix} C_1\Lambda_1 & C_2\Lambda_2 & \dots & C_n\Lambda_n \end{pmatrix} + \begin{pmatrix} E_1 & E_2 & \dots & E_n \end{pmatrix}, \end{aligned}$$

where $G = (G_1, G2, ..., G_T)'$ is a $(T \times k)$ matrix collecting the unobservable random vector G_t over T periods. Defining $\Lambda = (C_1\Lambda_1, C_2\Lambda_2, ..., C_n\Lambda_n)$ and $E = (E_1, E_2, ..., E_n)$, we have:

$$\underset{(T\times N)}{X} = \underset{(T\times K)}{G}\underset{(K\times N)}{\Lambda} + \underset{(T\times N)}{E} \tag{2.6}$$

The equation above says that X can be accommodated in a pooled ungrouped factor model with a k-dimensional factor G_t. Hence, G_t is called overall factor and k is referred to dimension of the overall factor space.

In order that each group is identified, the factor space of each group must be different i.e. $F_{i,t} \neq F_{j,t}$ for $i \neq j$ and no factor space of one group is a subspace of that of another group, in other words $F_{i,t}$ must not be a linear function of $F_{j,t}$, i.e. $F_{i,t} \neq C'F_{j,t}$ for any constant matrix C. Because $F_{i,t} = C_i'G_t$ and $F_{j,t} = C_j'G_t$, we will require that $C_i \neq C_jC$ for any constant matrix C. This leads to the following assumption.

Assumption 2.2
(a) C_i and C_j are not linearly dependent, i.e. $C_i \neq C_jC$, for any constant matrix C with $i \neq j$, $i = 1, 2, ..., n$ and $j = 1, 2, ..., n$.

(b) Any pair of factor loadings from two different groups $\lambda_{i,m}$ and $\lambda_{j,l}$ 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: $C_i\lambda_{i,m} \neq C_j\lambda_{j,l}$.

In the case with two factor planes and one factor line, assumption (a) excludes the situation in which the line lies on any one of the two planes and the situation where one plane lies on the other, such that the three group-pervasive factor spaces are distinguished from each other. Assumption 2.2 (b) is a technical assumption in order to simplify our presentation of a correct classification. (b) says that the common components of two variables from differen groups must not be the same. $C_i\lambda_{i,m} \neq C_j\lambda_{j,l}$ implies $F_i\lambda_{i,m} \neq F_j\lambda_{j,l}$. $F_i\lambda_{i,m}$ and $F_j\lambda_{j,l}$ represent two points (without errors) from two groups, respectively. Assumption 2.2 (b) excludes the situation, in which a data point lies in the intersection of the factor spaces of two groups. Otherwise we would be involved in an unfruitful discussion why the data point belongs to one group not the other[1].

Since our objective is to investigate the grouped structure in a factor model not to develop a new asymptotical result for a factor model, we are going to borrow well-established assumptions on factors and loadings as well as on the idiosyncratic components from the literature. The model setup in Bai and Ng (2002) serves well for this purpose. It is general enough for most applications. Further techniques in Bai and Ng (2002) fit well to investigation of a grouped factor model as we will see later. Therefore, we adopt the following assumptions from Bai and Ng (2002) in this paper.

Assumption 2.3
$E||G_t||^4 < \infty$ *and* $\frac{1}{T}\sum_{t=1}^{T} G_tG_t' \xrightarrow{P} \Sigma$ *as* $T \to \infty$ *for some positive definite matrix* Σ.

[1]See remarks of Proposition 3.5 for more details.

Assumption 2.3 is standard in a factor model. Under Assumption 2.1 and Assumption 2.3 it is easy to see that the group-pervasive factor $F_{i,t}$ also satisfies the requirements on factors given in Assumption 2.3, i.e.

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

(2) $\frac{1}{T}\sum_{t=1}^{T} F_{i,t}F_{i,t}' = \frac{1}{T}\sum_{t=1}^{T} C_i'G_tG_t'C_i \xrightarrow{P} C_i'\Sigma C_i$ as $T \to \infty$. Since $rank(C_i) = k_i$, $C_i\Sigma C_i'$ is a positive definite matrix.

Assumption 2.4
$\lambda_{i,j} < \lambda < \infty$ and $||\Lambda_i\Lambda_i'/N_i - D_i|| \to 0$ as $N_i \to \infty$ for some $(k_i \times k_i)$ positive definite matrix D_i, for $i = 1, 2, ..., n$.

Assumption 2.4 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 2.5
Under Assumption 2.4 and Assumption 2.1, the factor loading matrix Λ in the pooled ungrouped model (2.6) satisfies the requirement in Assumption 2.4, i.e. $\lambda_j < \lambda < \infty$ and $||\Lambda\Lambda'/N - D|| \to 0$ as $N \to \infty$ for some $(k \times k)$ positive definite matrix D.

Proof (See Appendix.)

Let X_{it} denote the observation of the ith variable at time t in X and e_{it} be the idiosyncratic component of X_{it}.

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. $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$;
2. $E(\sum_{i=1}^{N} e_{is}'e_{it}/N) = E(N^{-1}\sum_{i=1}^{N} e_{is}e_{it} = \gamma_N(s,t)), |\gamma_N(s,s)| \leq M$ for all s, and $T^{-1}\sum_{t=1}^{T}|\gamma_N(s,t)| \leq M$;
3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $\tau_{ij,t} \leq |\tau_{ij}|$ for some τ_{ij}, and for all t, in addition, $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}| < M$
4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \leq M$,
5. for every (t,s), $E|N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]|^4 \leq M$.

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)

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

Note that the idiosyncratic components in the pooled ungrouped factor model (2.6) are the same as the corresponding idiosyncratic components in the grouped factor model (2.4). Therefore the idiosyncratic errors in the grouped factor model satisfy the requirements in Assumptions 2.6. Since $F_{i,t}$ is a linear function of G_t, the requirement on weak dependence holds also between group-pervasive factors and idiosyncratic errors, i.e. it holds:

$$E\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\left|\left|\frac{1}{\sqrt{T}}F_{i,t}e_{i,jt}\right|\right|^2\right) \leq M \qquad \text{for } i = 1, 2, ...n.$$

Grouped Factor Models v.s. Pooled Ungrouped Factor Models
Comparing the grouped factor model (2.4) with the pooled ungrouped factor model (2.6), we can see that a successfully applied traditional method of orthogonal factor rotation is a special case of the grouped factor model, in which the group-pervasive factors are orthogonal to each other. Through classification of the variables into different groups and estimation of the group-pervasive factors, what we will obtain is a particular set of factors, i.e. F_t, such that different groups of variables have their non-zero factor loadings only at respectively different components of F_t. In this context, estimating a grouped factor model can be seen as a means to find a set of properly rotated factors that can offer a better understanding and interpretation of the data.

Grouped factor models allow, however, more general structures in the overall factor space than orthogonal clustering. The group-pervasive factors can be independent or dependent as well as correlated, i.e. the group-pervasive factor spaces can be disjunct, orthogonal, or intersected with any angels. With a grouped factor model we can investigate more general structures in the overall factor space.

One benefit of studying the grouped factor model (2.4) instead of the pooled ungrouped factor model (2.6) is to obtain group-pervasive factors, which may be useful for group-wise analysis. If we understand a pooled ungrouped factor model as a means to condense information from a large number of N variables to a small number of k overall factors and thus providing an explanation how each variable depends on the overall factors, then the grouped factor model (2.4) explains in detail which parts of variables are influenced by which kind of specific factors.

3 Estimation of GFM

Suppose that we know the number of groups $n \in \mathbb{N}$ as well as the correct grouping $s_n \in S_n$, where $\mathbb{N}$ is the set of natural numbers and S_n is the set of all possible groupings of variables given n. Then the estimation problem can be solved group by group using principal component method that corresponds to the minimization of squares residuals in each group. If the number of groups and the grouping of the variables are unknown, we could try to solve this problem by minimizing over n and s_n as follows.

$$\min_{n\in\mathbb{N}} \min_{s_n\in S_n,\, \Lambda_i, F_i} \sum_{i=1}^{n} ||X_i^{s_n} - F_i\Lambda_i||^2, \tag{3.7}$$

where $X_i^{s_n}$ is the data matrix collecting variables grouped into group i according to the grouping of s_n. The objective function (3.7) expresses clearly main features of the estimation problem of a grouped factor model: we estimate the unknown number of groups, the unknown grouping of variables, the unknown number of factors in each group and the unknown factors in each group. This problem can be seen as a problem of high dimensional clustering in which the cluster centers are subspaces of different unknown dimensions instead of centriods. A pragmatic approach to solve this kind of problems is to iterate between classification and estimation. Well known procedures are $k-means$ algorithms and expectation maximization algorithm. In high dimensional clustering, it is well known that these procedures depends sensi-

tively on starting values[2]. A thorough search over all groupings is NP-hard even in the case of two groups[3]. In this paper we adopt the idea of generalized principal component analysis[4] to estimate the grouped factor model.

3.1 An Alternative Representation of GFM

From a geometric point of view we can interpret factor models as follows. Each variable can be seen as a point in a T-dimensional space. We have N such points. While a pooled ungrouped factor model (2.6) says the N sample points are located nearly within a k-dimensional overall factor space spanned by G, a grouped factor model (2.4) says more precisely that the N points are actually located close to n different subspaces within the overall factor space, each of which is spanned by F_i with $i = 1, 2, ..., n$, respectively.

Denote the normalized complementary vectors to factor F_i by $\mathbf{B}_i$, i.e. $\mathbf{B}_i' F_i = 0$ and $\mathbf{B}_i' \mathbf{B}_i = I_{T-k_i}$. Denoting $F_i \Lambda_i$ by $\tilde{X}_i$, we can represent a GFM in the following alternative way:

$$X_i = \tilde{X}_i + E_i, \qquad \text{with} \qquad \mathbf{B}_i' \tilde{X}_i = 0 \qquad \text{for } i = 1, 2, ..., n. \tag{3.8}$$

While in GFM (2.4) the common components $\tilde{X}_i$ in each groups are represented as a linear function of the basis F_i, in equation (3.8) the common components $\tilde{X}_i$ are characterized through the orthogonality to the normal vectors $\mathbf{B}_i$. To estimate the number of groups and the number of factors in each group is equivalent to estimation of the number of the corresponding subspaces and their dimensions.

3.2 Method of Generalized Principal Component Analysis(GPCA)

While principal component analysis can be seen as a problem of estimating a linear subspace of unknown dimension k from N sample points, our problem is to estimate an unknown number of n linear subspaces with unknown dimensions k_i $(i = 1, 2, ...n)$ from N sample points. This is why this method is called *generalized principal component analysis*.

The subspaces in (3.8) can be represented as follows.

$$\prod_{i=1}^{n} ||\mathbf{B}_i' \mathbf{x}|| = 0, \tag{3.9}$$

where $\mathbf{x}$ is a point lying in one of the n subspaces and $||\ \ ||$ is the Euclidian norm in vector spaces. The left hand side of equation (3.9) is in fact a collection of

[2]See Zhang and Xia (2009) and Yedla, Pathakota, and Srinivasa (2010) for more details.

[3]The $k - means$ procedure is NP-hard. See http://en.wikipedia.org/K-means_clustering for more details.

[4]see Vidaly, Ma, and Sastry (2003) for more details.

$m = \prod_{i=1}^{n}(T-k_i)$ equations of homogeneous polynomials of degree n:

$$\prod_{i=1}^{n} ||(\mathbf{B}_i'\mathbf{x})|| = \prod_{i=1}^{n} ||((\mathbf{b}_{i1}, \mathbf{b}_{i2}, ...\mathbf{b}_{i(T-k_i)})'\mathbf{x})|| = \mathbf{0}$$
$$\iff \quad p_n(\mathbf{x}) = (p_{n1}(\mathbf{x}), p_{n2}(\mathbf{x}), ..., p_{nm}(\mathbf{x})) = \mathbf{0}. \tag{3.10}$$

Equation (3.10) says the subspaces can be equivalently presented as the null space of the m homogeneous polynomials of degree n. We demonstrate this fact in the following example.

Example 3.1
For the case $T = 3$, $n = 2$, $k_1 = 1$ and $k_2 = 2$ we are considering a line and a plane as two subspaces in a 3-dimensional space (See Fig.1). We have here $m = \prod_{i=1}^{n}(T-k_i) = 2$. In this case $\mathbf{B}_1$ is a 3×2 matrix and $\mathbf{B}_2$ is a 3×1 vector: $\mathbf{B}_1 = (\mathbf{b}_{11}, \mathbf{b}_{12})$ and $\mathbf{B}_2 = (\mathbf{b}_{21})$.

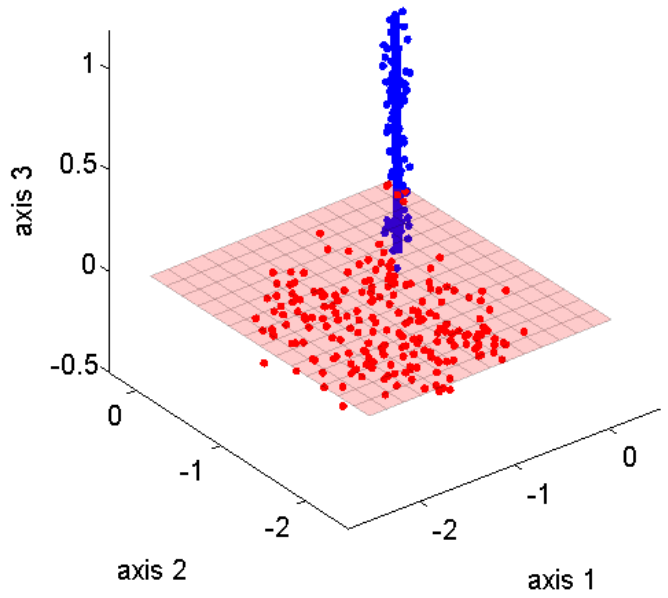

Figure 1: GPCA for $n = 2$, $k_1 = 1$, $k_2 = 2$, $N = 200$, $T = 3$

$$\prod_{i=1}^{2} ||(\mathbf{B}_i'\mathbf{x})|| = \mathbf{0} \iff p_2(\mathbf{x}) = ((\mathbf{b}_{11}'\mathbf{x})(\mathbf{b}_{21}'\mathbf{x}), (\mathbf{b}_{12}'\mathbf{x})(\mathbf{b}_{21}'\mathbf{x})) = 0. \tag{3.11}$$

More concretely, for a line $S_1 = \{\mathbf{x}|x_1 = 0, x_2 = 0\}$ and a plane $S_2 = \{\mathbf{x}|x_3 = 0\}$, we have

$$\mathbf{B}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \textit{and} \quad \mathbf{B}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{3.12}$$

The polynomials representing the two subspaces are:

$$p_2(\mathbf{x}) = ((\mathbf{b}_{11}'\mathbf{x})(\mathbf{b}_{21}'\mathbf{x}), (\mathbf{b}_{12}'\mathbf{x})(\mathbf{b}_{21}'\mathbf{x})) = (x_1x_3, x_2x_3) = \mathbf{0}. \tag{3.13}$$

A useful property of the polynomial representation of the subspaces is that the normal vectors of the subspaces can be obtained by differentiating the polynomials and evaluating the derivatives at one point in the respective subspaces.

For Example 3.1 the differential of $p_2(\mathbf{x})$ is given by:

$$\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}} = (\mathbf{b}_{11}(\mathbf{b}'_{21}\mathbf{x}) + \mathbf{b}_{21}(\mathbf{b}'_{11}\mathbf{x}), \mathbf{b}_{12}(\mathbf{b}'_{21}\mathbf{x}) + \mathbf{b}_{21}(\mathbf{b}'_{12}\mathbf{x}))\,. \tag{3.14}$$

Evaluating the differential at a point $\mathbf{x} \in S_1$ with $(\mathbf{b}_{11}, \mathbf{b}_{12})'\mathbf{x} = 0$, we obtain:

$$\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_1} = (\mathbf{b}_{11}(\mathbf{b}'_{21}\mathbf{x}), \mathbf{b}_{12}(\mathbf{b}'_{21}\mathbf{x}))\,. \tag{3.15}$$

Normalizing the derivative above we obtain:

$$\frac{\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_1}}{||\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_1}||} = (\mathbf{b}_{11}, \mathbf{b}_{12}) = \mathbf{B}_1. \tag{3.16}$$

Similarly, we have

$$\frac{\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_2}}{||\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_2}||} = (\mathbf{b}_{21}, \mathbf{b}_{21}) = \mathbf{B}_2. \tag{3.17}$$

Differentiating $p_n(\mathbf{x})$ to obtain the normal vectors of the subspaces provides one way to solve for the subspaces from the data. Our question is now: how can we obtain the polynomial $p_n(\mathbf{x})$, while the subspaces are still unknown? Since we have N sample points, each lying in one of the n subspaces, we can construct the subspaces from the sample points. Recall that $p_n(\mathbf{x})$ consists of m homogeneous polynomials of degree n in the elements of $\mathbf{x}$ and each such homogeneous polynomial of degree n is a linear combination of the monomials of the form $x_1^{n_1}x_2^{n_2}...x_T^{n_T}$ with $0 \le n_j \le n$ for $j = 1, ..., T$ and $n_1 + n_2 + ... + n_T = n$. Hence, we need only to find m linear combinations of the monomials that assume the value of zero at $\mathbf{x}$s that are points in the n subspaces. To this end, we look again at Example 3.1, where the polynomial representing the subspaces can be formulated as follows.

$$\begin{aligned}
& p_n(\mathbf{x}) = ((\mathbf{b}'_{11}\mathbf{x})(\mathbf{b}'_{21}\mathbf{x}), (\mathbf{b}'_{12}\mathbf{x})(\mathbf{b}'_{21}\mathbf{x})) \\
= \;& ((b_{111}x_1 + b_{112}x_2 + b_{113}x_3)(b_{211}x_1 + b_{212}x_2 + b_{213}x_3), \\
& (b_{121}x_1 + b_{122}x_2 + b_{123}x_3)(b_{211}x_1 + b_{212}x_2 + b_{213}x_3)) \\
= \;& (c_{11}x_1^2 + c_{12}x_1x_2 + c_{13}x_1x_3 + c_{14}x_2^2 + c_{15}x_2x_3 + c_{16}x_3^2, \\
& c_{21}x_1^2 + c_{22}x_1x_2 + c_{23}x_1x_3 + c_{24}x_2^2 + c_{25}x_2x_3 + c_{26}x_3^2) \\
= \;& (\mathbf{c}'_1\nu_2(\mathbf{x}), \mathbf{c}'_2\nu_2(\mathbf{x})) = (\mathbf{c}_1, \mathbf{c}_2)'\nu_2(\mathbf{x})) = 0,
\end{aligned} \tag{3.18}$$

where $\nu_2(\mathbf{x}) = (x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2)'$ is the Veronese map of degree 2, and the coefficients $\mathbf{c}_1$ is related to the normal vectors of the subspaces in the following way: $\mathbf{c}_1 = (c_{11}, c_{12}, c_{13}, c_{14}, c_{15}, c_{16})'$, with $c_{11} = b_{111}b_{211}$, $c_{12} = b_{111}b_{212} + b_{112}b_{211}$, $c_{13} = b_{111}b_{213} + b_{113}b_{211}, c_{14} = b_{112}b_{212}$, $c_{15} = b_{112}b_{213} + b_{113}b_{212}, c_{16} = b_{113}b_{213}$; and $\mathbf{c}_2$ is defined accordingly.

Generally, the Veronese map of degree n is defined as $\nu_n(\mathbf{x}) : \mathbb{R}^T \to \mathbb{R}^{M_n}$ with $M_n = \begin{pmatrix} n+T-1 \\ T-1 \end{pmatrix}$. $\nu_n : (x_1, ..., x_T)' \to (..., \mathbf{x}^I, ...)'$, where $\mathbf{x}^I = x_1^{n_1}x_2^{n_2}...x_T^{n_T}$ with $0 \le n_j \le n$ for $j = 1, ..., T$, and $n_1 + n_2 + ... + n_T = n$.

In Example 3.1 we see that a collection of n subspaces can be described as the set of points satisfying a set of homogeneous polynomials of the form (see equation (3.18)):

$$p(\mathbf{x}) = \mathbf{c}'\nu_n(\mathbf{x}) = 0. \tag{3.19}$$

Since each point in one of the n subspaces satisfies equation (3.19), for N points in the subspaces we will have a linear equation system:

$$L_n(X)\mathbf{c} = \begin{pmatrix} \nu_n(\mathbf{x}^1)' \\ \nu_n(\mathbf{x}^2)' \\ \vdots \\ \nu_n(\mathbf{x}^N)' \end{pmatrix} \mathbf{c} = 0, \tag{3.20}$$

where $L_n(X)$ is an $(N \times M_n)$ matrix. $L_n(X)\mathbf{c} = 0$ suggests that $\mathbf{c}$ can be calculated from the eigenvectors of the null space of $L_n(X)$. Once we have $\mathbf{c}$, we have a representation of the subspaces as $\nu_n(\mathbf{x})'\mathbf{c} = 0$. This suggests further that we can obtain the normal vectors to the subspaces by differentiating $\nu_n(\mathbf{x})'\mathbf{c}$ with respect to $\mathbf{x}$ and evaluating the derivative at points in the respective subspaces. This fact is summarized in Theorem 5 in Vidaly (2003).

Proposition 3.2 *(Polynomial differentiation Theorem 5 in Vidaly (2003))*
For the GPCA problem, if the given sample set X is such that $\dim(null(L_n)) = \dim(I_n)$ *and one generic point y_i is given for each subspace S_i, then we have*

$$S_{i\perp} = span\left\{ \frac{\partial c_n' \nu_n(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}=y_i}, \forall c_n \in null(L_n) \right\}.$$

Here $S_{i\perp}$ represents normal vectors of the subspace S_i, L_n is the data matrix as given in (3.20) and I_n is the ideal of the algebra set $p_n(\mathbf{x}) = 0$ that represents the n subspaces.

Following Proposition 3.2, the determination of the subspaces boils down to evaluating the derivatives of $\nu_n(\mathbf{x})'\mathbf{c}$ at one point in each subspace. For data generated without noises, we only need to find one point in each subspace in order to calculated the normal vectors of the respective subspaces and the classification problem can be solved perfectly. This method is called polynomial differentiation algorithm(PDA) (see Vidal, Ma, and Piazzi (2004) for more details). In the following we demonstrate how PDA works in Example 3.1.

Example 3.1 (continue) *We consider a set of 8 sample points from the two subspaces. The coordinates of the 8 points are collected in a data matrix X. Each column in X is one sample point.*

$$X = \begin{pmatrix} 1 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 \end{pmatrix} \tag{3.21}$$

Obviously, the first four points are located in the subspace of the plane S_2, and the next four points are located in the subspace of the line S_1. The Veronese mapping matrix with $\nu_2(\mathbf{x}) = (x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2)'$ is as follows.

$$L_n(X) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 4 & 4 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 0 & 0 & 16 \end{pmatrix}. \qquad \mathbf{c} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}$$

From $L_n(X)$ we can solve for its null space by singular value decomposition. $\mathbf{c}$ is the matrix containing the two eigenvectors of $Null(L_n(X))$.

The two polynomials that represent the the two subspaces can be obtained in the form of $\nu_n(\mathbf{x})'\mathbf{c} = \mathbf{0}$. So we have

$$\nu_n(\mathbf{x})'\mathbf{c} = (x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2)\mathbf{c} = (x_1x_3, -x_2x_3) = \mathbf{0}.$$

Comparing with equation (3.13), we know $\nu_n(\mathbf{x})'\mathbf{c} = 0$ represents the two subspaces: the line $S_1 = \{\mathbf{x}|x_1 = 0, x_2 = 0\}$ and the plane $S_2 = \{\mathbf{x}|x_3 = 0\}$.

According to Proposition 3.2, the normal vectors of the subspaces can be calculated by evaluating

$$\frac{\partial \nu_n(\mathbf{x})'c}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_1} \\ \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_2} \\ \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_3} \end{pmatrix} = \begin{pmatrix} 2x_1 & x_2 & x_3 & 0 & 0 & 0 \\ 0 & x_1 & 0 & 2x_2 & x_3 & 0 \\ 0 & 0 & x_1 & 0 & x_2 & 2x_3 \end{pmatrix} \mathbf{c} = \begin{pmatrix} x_3 & 0 \\ 0 & -x_3 \\ x_1 & -x2 \end{pmatrix}$$

at one point in the respective subspace. Evaluating the partial derivative at $\mathbf{x}^1$ to $\mathbf{x}^8$, we have:

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.22}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^3} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & -1 \end{pmatrix}, \qquad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^4} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 2 & -2 \end{pmatrix}, \tag{3.23}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^5} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}, \qquad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^6} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \\ 0 & 0 \end{pmatrix}, \tag{3.24}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^7} = \begin{pmatrix} 3 & 0 \\ 0 & -3 \\ 0 & 0 \end{pmatrix} \quad and \quad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^8} = \begin{pmatrix} 4 & 0 \\ 0 & -4 \\ 0 & 0 \end{pmatrix}. \tag{3.25}$$

Note that the rank of $\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^k}$ corresponds to the codimension of the respective subspace and the normal vectors of the respective subspace can be calculated as the principal component of $\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^k}$. For the points $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4$, the principal components of the partial derivatives are identical (0 0 1)'. Therefore these four points

belong to the subspace S_2 defined by the normal vector $\mathbf{B}_2$. The normalized derivatives for points $\mathbf{x}^5, \mathbf{x}^6, \mathbf{x}^7, \mathbf{x}^8$ are identical. Hence these four points belong to the subspace S_1 characterized by the normal vectors $\mathbf{B}_1$.

$$\mathbf{B}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad \mathbf{B}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}. \tag{3.26}$$

3.3 Method of Generalized Principal Component Analysis with Noisy Data

Sofar we know how to solve the classification problem if there is no noise in the data, i.e. $E_i = 0$ in equation (3.8). If $E_i \neq 0$ several problems arise: (1) $L_n(X)$ will have full rank and thus equation system (3.20) has only zero solution. (2) It may happen that no point lies exactly in any one of the subspaces, such that we cannot obtain an accurate inference on the normal vectors. Yang, Rao, Wagner, Ma, and Fossum (2005) propose a PDA with a voting scheme to solve the problem with noisy data.

Algorithm 1 Generalized Principal Component Analysis
Given a set of samples $\{x_k\}_{k=1}^N$, $(x_k \in \mathbb{R}^K)$ fit an n linear subspaces model with codimensions $d_1, ..., d_n$:

1: Set *angleTolerance*, let C be the number of distinct codimensions, and obtain D by the Hilbert function constraint.
2: Let $V\{1\}, ..., V\{C\}$ be integer arrays as voting counters and $U\{1\}, ..., U\{C\}$ be matrix arrays for basis candidates.
3: Construct $L_N = [\nu_n(\mathbf{x}^1), ..., \nu_n(\mathbf{x}^N)]$.
4: Form the set of polynomials $p_n(\mathbf{x})$ and compute $Dp_n(\mathbf{x})$.
5: **for all** sample $\mathbf{x}^k$ **do**
6: **for all** $1 \leq i \leq C$ **do**
7: Assume $\mathbf{x}^k$ is from a subspace with the codimension d equal to that of the class i. Find the first d principal components $B \in \mathbb{R}^{K \times d}$ in the matrix $Dp_n(\mathbf{x})|_{\mathbf{x}^k}$.
8: Compare B with all candidates in $U\{i\}$.
9: **if** $\exists j$, $subspaceangle[B, U\{i\}(j)] < angleTolerance$ **then**
10: $V\{i\}(j) = V\{i\}(j) + 1$.
11: Average the principal directions with the new basis B.
12: **else**
13: Add a new entry in $V\{i\}$ and $U\{i\}$.
14: **end if**
15: **end for**
16: **end for**
17: **for all** $1 \leq i \leq C$ **do**
18: m = the number of subspaces in class i.
19: Choose the first m highest votes in V{i} with their corresponding bases in U{i}.
20: Assign corresponding samples into the subspaces, and cancel their votes in the other classes.
21: **end for**
22: Segment the remaining samples based on these bases.

The motivation of PDA with a voting scheme is the following: for a given number of subspaces n and their codimensions $\{d_i\}_{i=1}^n$, the theoretical rank of the data matrix $L_n(X)$ called the *Hilbert function constraint* can be calculated. Then a set of polynomials $p_n(\mathbf{x})$ with coefficients equal to the eigenvectors in the null space of $L_n(X)$ are formed. Through evaluating $Dp_n(\mathbf{x})$ at each data point, a set of vectors normal to the subspace in which the point lies are obtained. The original PDA method relies on one good sample per subspace to classify the data. In the presence of noises, no single sample is reliable. However, through averaging the normal vectors of all samples in one subspace, it will smooth out the random noises. The table above is the algorithm given in Yang et al. (2005)[5]. We demonstrate how the PDA with a voting scheme works for Example 3.1 in the Appendix.

3.4 Classification of Variables

After obtaining a solution $\{\hat{\mathbf{B}}_1, \hat{\mathbf{B}}_2, ...\hat{\mathbf{B}}_n\}$ for the subspaces, a variable $\mathbf{x}^j$ is classified to that subspace to which $\mathbf{x}^j$ has the smallest distance among all subspaces. Given the set of estimated normal vectors $\{\hat{\mathbf{B}}_1, \hat{\mathbf{B}}_2, ...\hat{\mathbf{B}}_n\}$, we can calculate the distance between the j-th variable $\mathbf{x}^j$ and the ith subspace $\hat{\mathbf{B}}_i$ as follows:

$$||\hat{\mathbf{e}}_i^j|| = ||\hat{\mathbf{B}}_i' \mathbf{x}^j||.$$

The rule for classification is the following:

$$||\hat{\mathbf{e}}_i^j|| = \min\{||\hat{\mathbf{e}}_1^j||, ||\hat{\mathbf{e}}_2^j||, ..., ||\hat{\mathbf{e}}_n^j||\} \quad \rightarrow \quad \mathbf{x}^j \Rightarrow S_i, \tag{3.27}$$

where $\mathbf{x}^j \Rightarrow S_i$ means that $\mathbf{x}^j$ is classified to the subspace S_i.

We use $\mathbf{x}^{ji}$ to denote that the j-th variable is generated by the factors of the i-th group and $\mathbf{e}^{ji}$ is the corresponding noise. If

$$||\hat{\mathbf{e}}_i^{ji}|| = \min\{||\hat{\mathbf{e}}_1^{ji}||, ||\hat{\mathbf{e}}_2^{ji}||, ..., ||\hat{\mathbf{e}}_n^{ji}||\} \tag{3.28}$$

holds, then $\mathbf{x}^{ji} \Rightarrow S_i$ follows. This classification is correct. Assumption 2.2 implies that if there is no noise, all data points from one group do not lie in the subspaces of other groups, so that their distances to the subspaces of other groups are always strictly positive. This ensures that the classification according to distance will lead to a unique correct classification. The existence of noises will inevitably result in some errors in the classification despite use of the voting scheme. We show how to achieve a consistent classification in the next subsection.

3.5 Projected Models

In principle, we could obtain an estimate for each subspace by PDA as described in subsection 3.3. However, the usual case of a large factor model is that the number of observations T is large and the number of overall factors k is very small. $\mathbf{B}_i$ is of dimension $T \times (T - k_i)$ and the Veronese mapping matrix is of dimension $N \times \begin{pmatrix} n+T-1 \\ T-1 \end{pmatrix}$, such that the dimension of data involved in the PDA algorithm is very large. Consequently, the algorithm may not be practically executable due

[5]Yang et al. (2005) document good performance of this procedure in data segmentation.

to extremely heavy computational burdens. But, as far as classification of variables is concerned, a large T-dimensional problem ($T >> k$) can be casted into a K-dimensional problem with $T >> K \geq k$ to reduced the dimension of the problem. The reason is that projecting the T dimensional points onto a K dimensional subspace that is not orthogonal to the factor space, the classification is preserved[6] (See Fig.2). Hence, we can first transform the T-dimensional classification problem into a K-dimensional classification problem with $K \geq k$. After solving the classification problem, we can estimate the factors for each group using the original data.

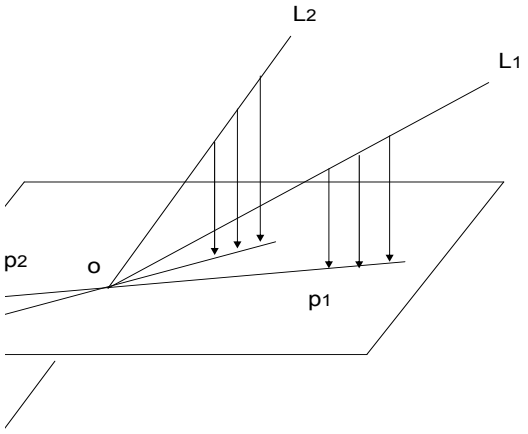

Figure 2: GPCA for $n = 2$, $k_1 = 1$, $k_2 + 1$, $T = 3$ and $K = 2$.

Let Q be the $(T \times K)$ matrix containing the K eigenvectors corresponding to K largest eigenvalues of XX'. $\sqrt{T}Q'$ is a principal component estimate of factor space spanned by G. A rescaled estimate can be calculated as follows:

$$\hat{G}^K = \frac{1}{NT}(XX')\sqrt{T}Q. \tag{3.29}$$

We project the original models (2.6) and (2.4) by premultiply $\frac{\hat{G}^K}{T}$ to both sides of the models and obtain:

$$\frac{1}{T}\hat{G}^{K'}X = \frac{1}{T}\hat{G}^{K'}G\Lambda + \frac{1}{T}\hat{G}^{K'}E \tag{3.30}$$

and

$$\frac{1}{T}\hat{G}^{K'}X_i = \frac{1}{T}\hat{G}^{K'}F_i\Lambda_i + \frac{1}{T}\hat{G}^{K'}E_i \qquad \text{for } i = 1, 2, ..., n. \tag{3.31}$$

Denoting $\frac{1}{T}\hat{G}^{K'}X$, $\frac{1}{T}\hat{G}^{K'}G$, $\frac{1}{T}\hat{G}^{K'}E$, $\frac{1}{T}\hat{G}^{K'}X_i$, $\frac{1}{T}\hat{G}^{K'}F_i$ and $\frac{1}{T}\hat{G}^{K'}E_i$ by $\bar{X}^T$, $\bar{G}^T$ and $\bar{E}^T$, $\bar{X}_i^T$, $\bar{F}_i^T$ and $\bar{E}_i^T$ respectively, we have

$$\underset{(K\times N)}{\bar{X}^T} = \underset{(K\times k)(k\times N)}{\bar{G}^T\ \Lambda} + \underset{(K\times N)}{\bar{E}^T} \tag{3.32}$$

[6]See Proposition 3.3 for more details.

and

$$\underset{(K\times N_i)}{\bar{X}_i^T} = \underset{(K\times k_i)}{\bar{F}_i^T}\underset{(k_i\times N_i)}{\Lambda_i} + \underset{(K\times N_i)}{\bar{E}_i^T} \qquad \text{for } i = 1, 2, ..., n \tag{3.33}$$

or equivalently

$$\bar{X}_i^T = \tilde{\bar{X}}_i^T + \bar{E}_i^T \qquad \text{with} \qquad \bar{\mathbf{B}}_i^{T\prime}\tilde{\bar{X}}_i^T = 0 \qquad \text{for } i = 1, 2, ..., n \tag{3.34}$$

The projected models (3.32) and (3.33) has the following property.

Proposition 3.3
Under Assumption 2.1 to Assumption 2.7, for $K = k$ it holds:

- *(a) $\bar{X}_i^T \xrightarrow{P} \bar{X}_i$ and $\bar{X}^T \xrightarrow{P} \bar{X}$ as $N \to \infty$, $T \to \infty$*
- *(b) $\bar{F}_i^T \xrightarrow{P} \bar{F}_i$ and $\bar{G}^T \xrightarrow{P} \bar{G}$ as $N \to \infty$, $T \to \infty$ and $\bar{F}_i = \bar{G}C_i$.*
- *(c) $\bar{E}_i^T \xrightarrow{P} 0$ and $\bar{E}^T \xrightarrow{P} 0$ as $N \to \infty$, $T \to \infty$*
- *(d) $\bar{F}_i \neq \bar{F}_j$, for $i \neq j$.*
- *(e) $\bar{F}_i$ is not a linear function of $\bar{F}_j$.*
- *(f) $\bar{F}_i\lambda_{i,m} \neq \bar{F}_j\lambda_{j,l}$ for any pair of factor loadings $\lambda_{i,m}$ and $\lambda_{j,l}$ for $m = 1, 2, ...N_i$, $l = 1, 2, ..., N_j$, $i = 1, 2, ..., n$, $j = 1, 2, ..., n$ and $i \neq j$.*

Proof (see Appendix).

Comparing the projected model (3.33) with the original model (2.4), we see that the projected model is also a grouped factor model with the same number of groups. Proposition 3.3 (a) through (c) state that the projected model will converge to a grouped factor model without noises, i.e. all data points will eventually lie directly in the respective factor spaces. (d) through (e) state that the membership relation between variables and their groups remain preserved after projection.

Benefits of a projection from a T dimensional problem onto a K dimensional problem are twofold: (1) it reduces the dimension of the numerical calculation in PDA and thus makes the problem practically solvable. The dimension of $\mathbf{B}_i$ reduces from $\{T \times (T - k_i)\}$ to $\{K \times (K - k_i)\}$. For a case of $T = 200$, $k_i = 4$, $K = 6$, and $n = 5$, the number of variables in $\mathbf{B}_i$ reduces from 195000 to 60. (2) The projection reduces the distance between data points and their subspaces, and thus enables a more precise classification. Eventually it will become a correct classification, as the idiosyncratic errors converge to zero as $T \to \infty$, $N \to \infty$.

Since the classification rule defined in (3.27) depends on the estimated residuals, the results of the classification is stochastic. Therefore, we need to characterize the stochastic property of a classification rule.

Definition 3.4
A classification rule is called consistent if

$$P(||\hat{\mathbf{e}}_i^{ji}|| = \min\{||\hat{\mathbf{e}}_1^j||, ||\hat{\mathbf{e}}_2^j||, ..., ||\hat{\mathbf{e}}_n^j||\}) \to 1 \qquad as \qquad T \to \infty, N \to \infty. \tag{3.35}$$

Proposition 3.5
Given a set of correct model parameters $(n, \{k_i\}_{i=1}^n)$, the classification rule (3.27) based on PDA with a voting scheme applied to the projected model (3.33) with $K = k$ is consistent.

Proof: According to Proposition 3.3 we have $\bar{E}_i^T \xrightarrow{P} 0$, as $T \to \infty, N \to \infty$. It follows $\bar{X}_i^T \xrightarrow{P} \bar{X}_i$, as $T \to \infty, N \to \infty$. For a variable j in $\bar{X}_i^T$ we have $\bar{\mathbf{x}}^{T,ji} \xrightarrow{P} \bar{\mathbf{x}}^{ji}$, as $T \to \infty, N \to \infty$. Let $\{\hat{\bar{\mathbf{B}}}_1, \hat{\bar{\mathbf{B}}}_2, ...\hat{\bar{\mathbf{B}}}_n\}$ be the estimate of the normal vectors of the subspaces using PDA based on the data $\{\bar{X}_i^T\}_{i=1}^n$ and $\{\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2, ...\bar{\mathbf{B}}_n\}$ be the normal vectors of the subspaces calculated with PDA based on the data $\{\bar{X}_i\}_{i=1}^n$. Because $\{\hat{\bar{\mathbf{B}}}_1, \hat{\bar{\mathbf{B}}}_2, ...\hat{\bar{\mathbf{B}}}_n\}$ is a continuous function of $\{\bar{X}_i^T\}_{i=1}^n$ at $\{\bar{X}_i\}_{i=1}^n$, it follows according to Slusky theorem:

$$\{\hat{\bar{\mathbf{B}}}_1, \hat{\bar{\mathbf{B}}}_2, ...\hat{\bar{\mathbf{B}}}_n\} \xrightarrow{P} \{\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2, ...\bar{\mathbf{B}}_n\}, \text{as } T \to \infty, N \to \infty$$

Therefore, we have

$$||\hat{\bar{\mathbf{e}}}_i^{T,ji}|| = ||\hat{\bar{\mathbf{B}}}_i' \bar{\mathbf{x}}^{T,ji}|| \xrightarrow{P} ||\bar{\mathbf{B}}_i' \bar{\mathbf{x}}^{ji}|| = 0 \quad \text{as} \quad T \to \infty, N \to \infty,$$

where $\hat{\bar{\mathbf{e}}}_i^{T,ji}$ is the distance between the data point $\bar{\mathbf{x}}^{T,ji}$ and the estimated ith subspace $\hat{\bar{\mathbf{B}}}_i$ in the projected model (3.33) and $\bar{\mathbf{x}}^{ji}$ is the limit of $\bar{\mathbf{x}}^{T,ji}$ as $T \to \infty, N \to \infty$. The probability limit in the equation above follows from Slusky theorem and the last equality is due to the definition of $\bar{\mathbf{x}}^{ji}$. Next we show that the probability that $\bar{\mathbf{x}}^{T,ji}$ has a strictly positive distance to other factor spaces converges to one.

$$1 = P(||\hat{\bar{e}}_l^{T,ji}|| \geq 0) = P(\{||\hat{\bar{e}}_l^{T,ji}|| > 0\} \cup \{||\hat{\bar{e}}_l^{T,ji}|| = 0\}) = P(||\hat{\bar{e}}_l^{T,ji}|| > 0) + P(||\hat{\bar{e}}_l^{T,ji}|| = 0)$$

From Proposition 3.3 (c) and (f) we have

$$P(||\hat{\bar{e}}_l^{T,ji}|| = 0) \to P(||\bar{e}_l^{ji}|| = 0) = P(\bar{F}_l \lambda_{l,j} = \bar{F}_i \lambda_{i,j}) = 0, \text{ as } T \to \infty, N \to \infty.$$

It follows then

$$P(||\hat{\bar{e}}_l^{T,ji}|| > 0) \to 1 \text{ as } T \to \infty, N \to \infty.$$

Because $||\hat{\bar{e}}_i^{T,ji}|| \xrightarrow{P} 0$ and $P(||\hat{\bar{e}}_l^{T,ji}|| > 0) \xrightarrow{P} 1$ for $k \neq i$, as $T \to \infty, N \to \infty$, we have

$$P(\bar{\mathbf{x}}^{T,ji} \Rightarrow \bar{S}_i) = P(||\hat{\bar{\mathbf{e}}}_i^{T,ji}|| = \min\{||\hat{\bar{\mathbf{e}}}_1^{T,ji}||, ||\hat{\bar{\mathbf{e}}}_2^{T,ji}||, ..., ||\hat{\bar{\mathbf{e}}}_n^{T,ji}||\}) \to 1, \text{ as } T \to \infty, N \to \infty. \tag{3.36}$$

□

Remarks: Assumption 2.2 (b) leads to the results that $P(||\hat{\bar{e}}_l^{ji}|| = 0) \to 0$ for $l \neq i$ and hence the proof of the consistent classification above. This assumption is, however, not essential for conducting a correct inference of the group-pervasive factors. If $P(||\hat{\bar{e}}_l^{ji}|| = 0) > 0$, a significant proportion of data would lie in the intersection of two factor spaces i and l. Because these data lie in the intersection of the two factor spaces, no matter to which one of the two groups they are classified, it will lead to a correct inference of group-pervasive factors. Allowing $P(||\hat{\bar{e}}_l^{ji}|| = 0) > 0$

will nevertheless complicate the definition of a correct classification. In order to avoid this complication and simplify the presentation, we make Assumption 2.2 (c).

Since the membership relations between variables and their groups remain preserved after the projection from a T dimensional space onto a K dimensional space. The classification of variables obtained in the projected model (3.33) is a consistent classification of the variables in the original model.

$$P(\mathbf{x}^{ji} \Rightarrow S_i) = P(\bar{\mathbf{x}}^{T,ji} \Rightarrow \bar{S}_i) \to 1, \text{ as } T \to \infty, N \to \infty. \tag{3.37}$$

3.6 Determination of the number of groups and the number of factors in each group

Given a set of key parameters of a grouped factor model $(n, \{k_i\}_{i=1}^n)$, we can classify N variables into n groups by GPCA method. For group i we denote the T observations of $N_i^{s_n}$ variables that are classified into this group by $X_i^{s_n}$, where s_n denotes this particular grouping of the variables. If the given parameters $(n, \{k_i\}_{i=1}^n)$ are correct, the classification will be asymptotically correct and we can estimate, group by group, the group-pervasive factors using the standard principal component method, which is equivalent to solving of the following minimization problem:

$$V_i(k_i, \hat{F}_i, N_i^{s_n}) = \min_{\Lambda_i, F_i} \frac{1}{N_i^{s_n} T} \sum_{j=1}^{N_i^{s_n}} \sum_{t=1}^{T} (X_{i,jt}^{s_n} - \lambda_{i,j} F_{i,t})^2, \tag{3.38}$$

where $\Lambda_i = (\lambda_{i,1}, \lambda_{i,2}, ..., \lambda_{i,N_i^{s_n}})$ and $F_i = (F_{i,1}, F_{i,2}, ..., F_{i,T})'$.

A question is now how can we know whether this set of parameters $(n, \{k_i\}_{i=1}^n)$ are correct or not? One insight of Bai and Ng (2002) is that the number of factors in a group can be determined through minimizing an information criterion that consists of mean squared errors of the model and a properly scaled penalty term:

$$\hat{k}_i = \text{argmin}_{0<k_i\leq k} \left(V_i(k_i, \hat{F}_i, N_i^{s_n}) + \hat{\sigma}_i k_i g(N_i^{s_n}, T) \right),$$

where $g(N_i^{s_n}, T)$ is a scaling function[7].

Since we have more than one group, we need to extend the mean squared errors as well as the penalty terms over all groups. In this way we can construct a model selection criterion to determine the number of groups and the number of factors in each group. A model selection criterion, $C(n, \{k_i\}_{i=1}^n, \{X_i^{s_n}\})$, is a scalar function of data, model parameters and the classification of the variables, which measures the goodness of fit of the model to the data.

Definition 3.6
A model selection criterion $C(n, \{k_i\}_{i=1}^n, \{X_i^{s_n}\})$ is called consistent if it satisfies the following condition:

$$P\{C(n^o, \{k_i^o\}_{i=1}^n, \{X_i^s\}) < C(n', \{k_i'\}_{i=1}^{n'}, \{X_i^u\})\} \to 1 \qquad \text{as } T, N \to \infty. \tag{3.39}$$

Here $(n^o \ \{k_i^o\}_{i=1}^n)$ are parameters of the true model, and $\{X_i^s\}$ is the corresponding classification based on GPCA; $(n', \{k_i'\}_{i=1}^{n'})$ are parameters of an alternative model and $\{X_i^u\}$ is the corresponding classification using GPCA.

[7]See Bai and Ng (2002) for more details.

Proposition 3.7
Under Assumption 2.1 to Assumption 2.7,

$$PC(n,\{k_i\}_{i=1}^n,\{X_i^{s_n}\}) = \sum_{i=1}^{n}\frac{N_i}{N}V_i(k_i,\hat{F}^{k_i},N_i)+\hat{\sigma}^2\left(\sum_{i=1}^{n}\frac{N_i}{N}\left(k_i+h(N_i/N)\right)\right)g(N,T) \tag{3.40}$$

is a consistent model selection criterion if the following conditions are satisfied:

1. $\lim_{N\to\infty}\frac{N_i}{N}\to\alpha_i>\underline{\alpha}$, *where* $\frac{N_i}{N}$ *is the share of variables in the ith group. 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. *(a)* $0<h(\alpha)<1$ *for any* $0\le\alpha\le 1$
   (b) $h(\alpha_i)\ge h(\alpha_j)$ *for any* $0\le\alpha_i\le\alpha_j\le 1$.
   (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* $\sum_l\alpha_l=\sum_j\alpha_j=1$.

The model selection criterion can be reformulated in the following more compact form:

$$PC(n,\{k_i\},\{X_i^{s_n}\}) = \bar{V}(\{k_i\},\{\hat{\alpha}_i\})+\hat{\sigma}^2(\bar{k}+\bar{h})g(N,T)$$

where $\hat{\sigma}^2$ is a consistent estimate of $(NT)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{t=1}^{T}E(e_{i,jt})^2$, $\bar{k}$ is the weighted average number of factors over all groups and $\bar{h}$ is the weighted average of the penalty function $h(\hat{\alpha}_i)$ over all groups.

Remarks

In this formulation it is clear that $\bar{k}$ is the penalty due to the average number of factors and $\bar{h}$ is the penalty due to dispersion of groups. Compared to the PC criterion in Bai and Ng (2002), obviously this model selection criterion is a variant of weighted average of PC criteria over all groups with an additional penalty on the dispersion of groups in a model. Condition 1 is to make sure that the proportion of a group will not vanish asymptotically. Because we are considering the asymptotical property of the model selection criterion, the proportion of a group in a candidate model should not be vanishing. Hence we assume that for all candidate models, there exists a constant lower bound for the ratio of the number of variables in a group to the total number of variables in a model. 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 parameters, only when the dominating parameters 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.

3.7 Estimation Procedure for a Grouped Factor Model

- Step 1: Estimate k by the PC criterion of Bai and Ng (2002) using pooled data.
- Step 2: Project the $(T \times N)$ pooled data matrix X onto a $(k \times N)$ matrix:

$$\bar{X}^T = \frac{1}{T}\hat{G}^{k'}X,$$

  where $\hat{G}^k$ is defined in (3.29).

- Step 3: According to a chosen model $(n, \{k_i\}_{i=1}^n)$, solve for the corresponding subspaces $(\hat{\bar{\mathbf{B}}}_1, \hat{\bar{\mathbf{B}}}_2, ..., \hat{\bar{\mathbf{B}}}_n)$ of the projected model (3.34) by PDA with voting scheme and classify the variables according to rule (3.27).
- Step 4: Use the model selection criterion to evaluate alternative models to obtain an optimal model and the corresponding classification of variables $\{X_i^{s_n}\}_{i=1}^n$.
- Step 5: Estimate a factor model for each group of data in $\{X_i^{s_n}\}_{i=1}^n$ by the standard principal component method to obtain estimates for the respective group-pervasive factors $\hat{F}_i = \frac{1}{N_i^{s_n}T}(X_i^{s_n}X_i^{s_n\prime})\sqrt{T}Q_i$, where Q_i contains the k_i eigenvectors corresponding to the k_i largest eigenvalues of the matrix $(X_i^{s_n}X_i^{s_n\prime})$.

The procedure above will give a consistent classification of the variables as well as consistent estimates of the group-pervasive factor spaces.

Proposition 3.8
Under Assumption 2.1 to Assumption 2.7 and the three conditions given in Proposition 3.7, the procedure described above will provide a consistent estimate of the group-pervasive factor space for each group, i.e.

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t} - H_i^{k_i'}F_{i,t}||^2 \xrightarrow{P} 0, \text{ as } T \to \infty, N \to \infty, \text{ for } i = 1, 2, ...n, \qquad (3.41)$$

where $\hat{F}_{i,t}$ is the estimate of the group-pervasive factor of the ith group and $H_i^{k_i'}$ is a $(k_i \times k_i)$ full rank matrix.

4 Simulation Studies and an Application Example

4.1 Simulation Studies

In this section we document results of our simulation study. The simulation study is conducted in order to assess the performance of the proposed estimation procedure in finite sample cases. 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 group-pervasive factors in each group. We use the number of factors in each group k_i $i = 1, 2, ..., n$ and the dimension of the overall factors space k to

represent a GFM. For example [321|5] represents a GFM with three groups, the overall factor space is 5-dimensional and the number of factors in each group is 3, 2 and 1 respectively. To take into account that different group-pervasive factors may be correlated and hence may have common factors, our data generating process is designed in a way that there exists one common factor in all groups except the groups with only one factor. According to this setting, in the model [321|5] there exists one common factor in the first and the second groups and hence the overall factor space is 5-dimensional.

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

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

where the factor $F_{i,t} = (F_{i,1t}, F_{i,2t}, ..., F_{i,k_it})'$ of the ith group is a $(k_i \times 1)$ vector of $N(0,1)$ variables; the factor loadings for the group $\lambda_{i,j} = (\lambda_{i,1j}, \lambda_{i,2j}, ..., \lambda_{i,k_ij})'$ is a $(k_i \times 1)$ vector 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 k_i. The base case under consideration is that the common component has the same variance as the idiosyncratic component, i.e. $\theta_i = k_i$. We consider 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 80 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 competing model candidates. For a true model [2 2|3], [3 1] and [2 2 2] are more general models. The true model [2 2|3] consists of two group-pervasive factor planes within a 3-dimensional overall factor space. The model [3 1] is more general because it contains a three-dimensional subspace and a one-dimensional subspace, and [2 2 2] is also more general because it contains three two-dimensional subspaces. But, [2 1] is a more restrictive model because it contains only one two-dimensional subspace and one one-dimensional subspace in a three dimensional overall factor space.

The outcomes of the simulation study are summarized in Table 1 to Table 5. The first three columns in these tables give numbers of variables in each group, total numbers of variables and numbers of observations in the respective simulation settings. The fourth column gives the true data-generating grouped factor models and the candidate models under consideration. The integers in a pair of square brackets give the numbers of factors in the respective groups of a grouped factor model. For a data-generating model we give also the dimension of the overall factor space which is the number behind the bar in the square bracket. For candidate models we do not give the dimensions of the overall factor spaces, because they will be determined in the estimation procedure. Since the estimation procedure consists of two steps: (1) projection of the data onto a k dimensional overall factor space and (2) select the correct model from the candidates, we report the performance with respect to choosing the correct projection dimensions and with respect to choosing the correct models from the competing candidates.

Determination of the projection dimension can be seen as a problem of comparing pooled ungrouped models with grouped models. The column under the header of $UGRP$ reports the performance of the model selection criterion in this respect. A number in the column of $UGRP$ gives the proportion that the correct projection dimension is chosen and at least one grouped factor model is preferred over the correctly chosen ungrouped factor model in the respective 1000 simulation runs. Since our data generating models are all grouped factor models, for a good performance of the selection criterion we expect the numbers in this column to be close to one. The numbers in the column of $UGRP$ show that the model selection criterion works well in determining the right dimension of the projection space. For all configurations in the simulation $T = 80$ and $N_i = 30$ are enough to obtain the correct projection dimensions, i.e. the proportions of finding the right projection dimension are very high: almost all numbers in this column are one and a few numbers below one are also close to one[8].

The column under the header $CCLM$ reports the proportion of correctly chosen 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 close to one, indicating that for the considered configurations the estimation procedure performs well in identifying 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 relatively low: in 5 cases the numbers are below 90% but still over 80%. 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 $N_i = 60$ the results are already satisfactory.

The column under the header $MCLV$ gives the average proportion of misclassified variables in respective 1000 simulation runs. If the classification works well, the numbers in this column should be close to zero. Most of the numbers in the column of $MCLV$ are under 10 percent, indicating a good performance of the classification procedure. We observe that if the group-pervasive factor spaces are intersected, the share of misclassification tends to be higher. This is because as long as the group-pervasive factor spaces are intersected, data points lying close to the intersection of the group-pervasive factor spaces will lead to higher proportion of misclassification. However, because these data points are close to both group-pervasive factor spaces, this misclassification has little negative impact on estimation of group-pervasive factors.

$SFF0$[9] reports the average goodness of fit of the estimated factors to the true factors in 1000 simulation runs. $SFF0$ is 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 variable classification works well, we expect also a good performance in factor estimation. Indeed in most cases the numbers in the column of $SFF0$ are over 90% and with increasing N and T, the numbers are approaching one.

[8] This result is consistent with the simulation result given in Bai and Ng (2002).

[9] $SFF0 = \frac{tr(F^{0\prime}\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'F^0)}{tr(F^{0\prime}F^0)}$

Table 1: Estimation of grouped factor models

N_i	N	T	Model and Candidates	CCLM	SFF0	MCLV	UGRP
			[11|2]				
30	60	80	[111] [1 1]	0.92	0.97	0.07	1.00
30	60	150	[111] [1 1]	0.97	0.97	0.05	1.00
30	60	300	[111] [1 1]	1.00	0.97	0.03	1.00
30	60	500	[111] [1 1]	1.00	0.96	0.03	1.00
60	120	80	[111] [1 1]	0.94	0.98	0.07	1.00
60	120	150	[111] [1 1]	0.97	0.98	0.05	1.00
60	120	300	[111] [1 1]	1.00	0.98	0.04	1.00
60	120	500	[111] [1 1]	1.00	0.98	0.03	1.00
200	400	80	[111] [1 1]	0.94	0.99	0.07	1.00
200	400	150	[111] [1 1]	0.99	0.99	0.05	1.00
200	400	300	[111] [1 1]	1.00	0.99	0.04	1.00
200	400	500	[111] [1 1]	1.00	0.99	0.03	1.00
			[21|3]				
30	60	80	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	0.98	0.94	0.06	1.00
30	60	150	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	0.99	0.95	0.04	1.00
30	60	300	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.96	0.03	1.00
30	60	500	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.95	0.03	1.00
60	120	80	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	0.99	0.98	0.04	1.00
60	120	150	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.98	0.03	1.00
60	120	300	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.97	0.03	1.00
60	120	500	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.98	0.02	1.00
200	400	80	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.97	0.08	1.00
200	400	150	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.97	0.09	1.00
200	400	300	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.98	0.06	1.00
200	400	500	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.99	0.04	1.00
			[22|3]				
30	60	80	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	0.96	0.91	0.09	1.00
30	60	150	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.91	0.08	1.00
30	60	300	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.92	0.05	1.00
30	60	500	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.91	0.04	1.00
60	120	80	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	0.94	0.95	0.09	1.00
60	120	150	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	0.99	0.95	0.07	1.00
60	120	300	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.95	0.06	1.00
60	120	500	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.96	0.05	1.00
200	400	80	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.97	0.11	1.00
200	400	150	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.97	0.09	1.00
200	400	300	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.98	0.06	1.00
200	400	500	[2 2 1] [2 1] [1 1] [1 1 1] [2 2]	1.00	0.99	0.04	1.00

Notes: Table 1 reports the results of 1000 Monte Carlo runs of estimation of GFMs. The first three columns give numbers of observations and numbers of variables in the respective simulation runs. The fourth columns gives the true model and the candidate models. $CCLM$ gives the proportion of the correctly identified true models. $SFF0$ is the average goodness of fit of the estimated group-pervasive factors to the true group-pervasive factors over all groups. $MCLV$ gives the average proportion of misclassified variables. $UGRP$ gives the proportion of correctly identified projection spaces.

Table 2: Estimation of grouped factor models

N	N_i	T	Model and Candidates	CCLM	SFF0	MCLV	UGRP
			[32|4]				
30	60	80	[3 2] [3 1] [2 1] [3 3] [3 2 1]	0.98	0.91	0.09	1.00
30	60	150	[3 2] [3 1] [2 1] [3 3] [3 2 1]	0.99	0.92	0.07	1.00
30	60	300	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.92	0.06	1.00
30	60	500	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.94	0.05	1.00
60	120	80	[3 2] [3 1] [2 1] [3 3] [3 2 1]	0.98	0.95	0.08	1.00
60	120	150	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.97	0.08	1.00
60	120	300	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.97	0.06	1.00
60	120	500	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.98	0.04	1.00
200	400	80	[3 2] [3 1] [2 1] [3 3] [3 2 1]	0.99	0.99	0.09	1.00
200	400	150	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.99	0.09	1.00
200	400	300	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.98	0.06	1.00
200	400	500	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.99	0.04	1.00
			[33|5]				
30	60	80	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	0.99	0.90	0.05	0.97
30	60	150	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	0.99	0.90	0.02	0.98
30	60	300	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.90	0.01	1.00
30	60	500	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.90	0.01	1.00
60	120	80	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.95	0.04	1.00
60	120	150	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.95	0.02	1.00
60	120	300	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.95	0.01	1.00
60	120	500	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.98	0.04	1.00
200	400	80	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.98	0.04	1.00
200	400	150	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.98	0.03	1.00
200	400	300	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.98	0.02	1.00
200	400	500	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3]	1.00	0.98	0.01	1.00
			[31|4]				
30	60	80	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	0.85	0.93	0.07	0.99
30	60	150	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	0.88	0.93	0.05	1.00
30	60	300	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	0.99	0.93	0.04	1.00
30	60	500	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	1.00	0.93	0.03	1.00
60	120	80	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	0.99	0.97	0.07	1.00
60	120	150	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	0.99	0.95	0.05	1.00
60	120	300	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	1.00	0.96	0.04	1.00
60	120	500	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	1.00	0.95	0.03	1.00
200	400	80	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	1.00	0.98	0.07	1.00
200	400	150	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	1.00	0.98	0.05	1.00
200	400	300	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	1.00	0.98	0.03	1.00
200	400	500	[2 1] [2 2] [3 2 1] [3 1 1] [3 1]	1.00	0.99	0.03	1.00

Notes: Table 2 reports the results of 1000 Monte Carlo runs of estimation of GFMs. The first three columns give numbers of observations and numbers of variables in the respective simulation runs. The fourth columns gives the true model and the candidate models. $CCLM$ gives the proportion of the correctly identified true models. $SFF0$ is the average goodness of fit of the estimated group-pervasive factors to the true group-pervasive factors over all groups. $MCLV$ gives the average proportion of misclassified variables. $UGRP$ gives the proportion of correctly identified projection spaces.

Table 3: Estimation of grouped factor models

N	N_i	T	Model and Candidates	CCLM	SFF0	MCLV	UGRP
			[311|5]				
30	90	80	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	0.89	0.93	0.11	1.00
30	90	150	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	0.98	0.94	0.08	1.00
30	90	300	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	0.99	0.95	0.05	1.00
30	90	500	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	1.00	0.94	0.04	1.00
60	180	80	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	0.90	0.97	0.10	1.00
60	180	150	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	0.98	0.97	0.07	1.00
60	180	300	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	0.99	0.98	0.05	1.00
60	180	500	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	1.00	0.96	0.04	1.00
200	400	80	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	0.96	0.97	0.07	1.00
200	400	150	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	1.00	0.97	0.05	1.00
200	400	300	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	1.00	0.98	0.05	1.00
200	400	500	[2 1 1] [3 1] [3 2] [3 1 1 1] [3 1 1]	1.00	0.99	0.03	1.00
			[111|3]				
30	90	80	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.96	0.10	1.00
30	90	150	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.97	0.07	1.00
30	90	300	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.97	0.05	1.00
30	90	500	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.97	0.04	1.00
60	180	80	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.10	1.00
60	180	150	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.07	1.00
60	180	300	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.05	1.00
60	180	500	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.02	1.00
200	400	80	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.99	0.10	1.00
200	400	150	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.99	0.07	1.00
200	400	300	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.99	0.05	1.00
200	400	500	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.99	0.04	1.00
			[211|4]				
30	90	80	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	0.99	0.96	0.06	1.00
30	90	150	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	0.99	0.96	0.03	1.00
30	90	300	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.95	0.04	1.00
30	90	500	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.95	0.02	1.00
60	180	80	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.98	0.06	1.00
60	180	150	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.98	0.05	1.00
60	180	300	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.98	0.02	1.00
60	180	500	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.98	0.03	1.00
200	400	80	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.98	0.13	1.00
200	400	150	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.98	0.10	1.00
200	400	300	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.99	0.07	1.00
200	400	500	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.99	0.04	1.00

Notes: Table 3 reports the results of 1000 Monte Carlo runs of estimation of GFMs. The first three columns give numbers of observations and numbers of variables in the respective simulation runs. The fourth columns gives the true model and the candidate models. $CCLM$ gives the proportion of the correctly identified true models. $SFF0$ is the average goodness of fit of the estimated group-pervasive factors to the true group-pervasive factors over all groups. $MCLV$ gives the average proportion of misclassified variables. $UGRP$ gives the proportion of correctly identified projection spaces.

Table 4: Estimation of grouped factor models

N	N_i	T	Model and Candidates	CCLM	SFF0	MCLV	UGRP
			[222|4]				
30	90	80	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	0.99	0.92	0.17	1.00
30	90	150	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.93	0.13	1.00
30	90	300	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.93	0.08	1.00
30	90	500	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.93	0.06	1.00
60	180	80	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.96	0.16	1.00
60	180	150	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.97	0.11	1.00
60	180	300	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.97	0.08	1.00
60	180	500	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.97	0.06	1.00
200	400	80	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.99	0.14	1.00
200	400	150	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.99	0.11	1.00
200	400	300	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.99	0.08	1.00
200	400	500	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 1 1]	1.00	0.99	0.06	1.00
			[322|5]				
30	90	80	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.92	0.91	0.16	0.97
30	90	150	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.96	0.92	0.11	1.00
30	90	300	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.92	0.07	1.00
30	90	500	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.93	0.06	1.00
60	180	80	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.99	0.98	0.13	1.00
60	180	150	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.96	0.11	1.00
60	180	300	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.96	0.07	1.00
60	180	500	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.96	0.05	1.00
200	400	80	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.99	0.12	1.00
200	400	150	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.99	0.09	1.00
200	400	300	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.99	0.08	1.00
200	400	500	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.99	0.05	1.00
			[2222|5]				
30	120	80	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	0.88	0.92	0.20	0.97
30	120	150	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	0.97	0.92	0.13	0.99
30	120	300	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.93	0.11	1.00
30	120	500	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.93	0.07	1.00
60	240	80	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	0.98	0.95	0.18	1.00
60	240	150	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.97	0.15	1.00
60	240	300	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.96	0.10	1.00
60	240	500	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.96	0.08	1.00
200	800	80	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.98	0.17	1.00
200	800	150	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.99	0.12	1.00
200	800	300	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.99	0.09	1.00
200	800	500	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	1.00	0.99	0.07	1.00

Notes: Table 4 reports the results of 1000 Monte Carlo runs of estimation of GFMs. The first three columns give numbers of observations and numbers of variables in the respective simulation runs. The fourth columns gives the true model and the candidate models. $CCLM$ gives the proportion of the correctly identified true models. $SFF0$ is the average goodness of fit of the estimated group-pervasive factors to the true group-pervasive factors over all groups. $MCLV$ gives the average proportion of misclassified variables. $UGRP$ gives the proportion of correctly identified projection spaces.

Table 5: Estimation of grouped factor models

N_i	N	T	Model and Candidates	CCLM	SFF0	MCLV	UGRP
			[2211|5]				
30	120	80	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.90	0.94	0.14	1.00
30	120	150	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.93	0.95	0.09	1.00
30	120	300	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.99	0.95	0.05	1.00
30	120	500	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.98	0.02	1.00
60	240	80	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.98	0.97	0.13	1.00
60	240	150	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.97	0.09	1.00
60	240	300	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.98	0.03	1.00
60	240	500	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.97	0.02	1.00
200	800	80	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.99	0.93	0.13	1.00
200	800	150	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.98	0.10	1.00
200	800	300	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.99	0.03	1.00
200	800	500	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.99	0.03	1.00
			[3211|6]				
30	120	80	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.85	0.93	0.14	0.98
30	120	150	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.95	0.94	0.10	1.00
30	120	300	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.94	0.07	1.00
30	120	500	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.94	0.05	1.00
60	240	80	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.94	0.97	0.13	1.00
60	240	150	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.99	0.97	0.09	1.00
60	240	300	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.97	0.07	1.00
60	240	500	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.97	0.04	1.00
200	800	80	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.99	0.12	1.00
200	800	150	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.99	0.09	1.00
200	800	300	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.99	0.07	1.00
200	800	500	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.99	0.05	1.00
			[3221|6]				
30	120	80	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.80	0.92	0.15	1.00
30	120	150	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.86	0.93	0.11	1.00
30	120	300	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.99	0.93	0.07	1.00
30	120	500	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.93	0.06	1.00
60	240	80	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.95	0.96	0.15	1.00
60	240	150	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.99	0.96	0.11	1.00
60	240	300	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.97	0.07	1.00
60	240	500	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.97	0.05	1.00
200	800	80	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.99	0.14	1.00
200	800	150	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.99	0.10	1.00
200	800	300	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.99	0.07	1.00
200	800	500	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.99	0.05	1.00

Notes: Table 5 reports the results of 1000 Monte Carlo runs of estimation of GFMs. The first three columns give numbers of observations and numbers of variables in the respective simulation runs. The fourth columns gives the true model and the candidate models. $CCLM$ gives the proportion of the correctly identified true models. $SFF0$ is the average goodness of fit of the estimated group-pervasive factors to the true group-pervasive factors over all groups. $MCLV$ gives the average proportion of misclassified variables. $UGRP$ gives the proportion of correctly identified projection spaces.

4.2 An Empirical Application

In this subsection we apply the GFM to stock returns in the Australian Stock Exchange. Our purpose of this exercise is to demonstrate that grouped structures do exist in empirical data and our estimation procedure is capable of uncovering them. The data used in this exercise are stock returns of companies included in ASX200. ASX200 is one of the most important share index in the Australian Stock Exchange. It accounts for roughly 85% of the market capitalization of all stocks listed in the Australian 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 the appendix. We transform the data so that each series has mean zero. Using the PC criterion of Bai and Ng (2002) we identify that there are three factors in the data set (see Table 6). After choosing $k = 3$ we investigate 18 potential candidate models. These 18 candidate models include all possible subspace configurations up to 4 groups within a three dimensional overall factor space. We exclude models with more that 4 groups because in these cases it is highly probable that some group will contain less than 30 variables such that the model selection criterion would become unreliable[11]. The estimation results for the considered configurations are summarized in Table 6.

Table 6: Estimation of Grouped Dynamic Factor Models for ASX200

No.	Model	PC	(N_i)	No.	Model	PC	(N_i)
1	[1]	0.005727	(168)	10	[1 1 1]	0.005151	(101 46 21)
2	[2]	0.005298	(168)	11	[2 1 1]	0.005312	(126 38 4)
3	[3]	0.005270	(168)	12	[2 2 1]	0.005078	(115 43 10)
4	[4]	0.005288	(168)	13	[2 2 2]	0.005052	(95 47 26)
5	[5]	0.005339	(168)	14	[1 1 1 1]	0.005173	(97 50 18 3)
6	[6]	0.005399	(168)	15	[2 1 1 1]	0.005161	((93 32 32 11)
7	[1 1]	0.005282	(100 68)	16	[2 2 1 1]	0.004946	(61 50 42 10)
8	[2 1]	0.005108	(93 75)	17	[2 2 2 1]	0.005054	(91 38 38 2)
9	[2 2]	0.005086	(109 59)	18	[2 2 2 2]	0.004855	(85 50 17 16)

Notes: We use numbers in a pair of squared brackets to represent a model. [2 2] represents a model with two groups and each with two factors. The column PC reports the values the model selection criterion for the corresponding models. The column under the header (N_i) gives the numbers of variables classified into the respective groups.

In Table 6 there are 9 models with four or three groups. Common to these 9 models, each model has at least one group that contains less than 30 variables. Since numbers of variables in each group are crucial for the reliability of the estimation procedure and a number smaller than 30 is too low to achieve a reliable estimation, we regard the estimation of these 9 models to be unreliable. Therefore, we will focus only on models with two groups and ungrouped models. According to the values of the model selection criterion, we conclude that [2 2] 1s the most suitable model

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

[11]We have estimated all model configurations with four groups within a three dimensional overall factor space. Indeed in all these models there are at least one group with less than 20 variables.

for the data set. This implies that we understand that the 168 shares consist of 2 groups each of which are driven by two factors respectively (See Fig. 3).

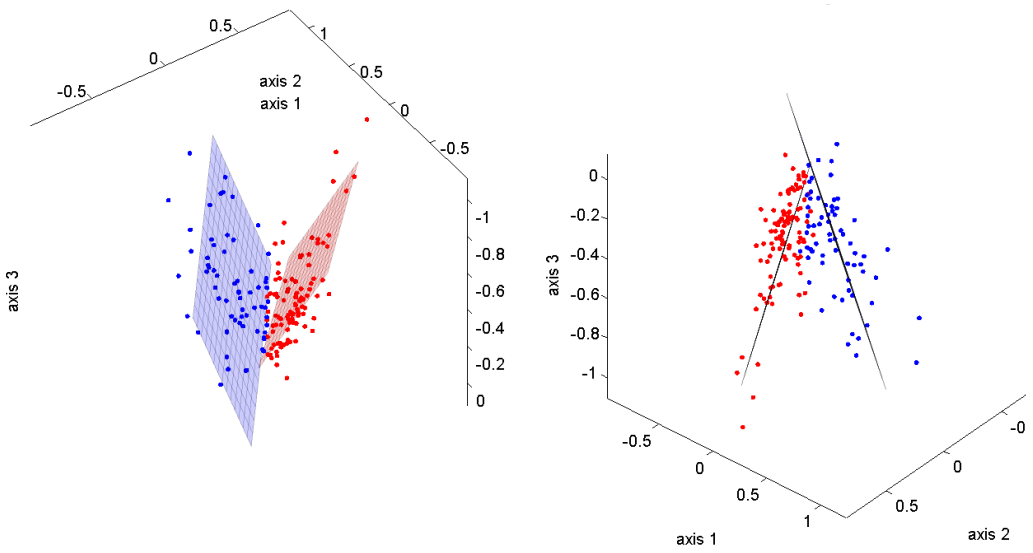

Figure 3: ASX200 shares in two groups in the projected model

The grouping of the 168 variables are given in Table 9. Interestingly, almost all companies in resource sectors including mining, energy, and exploration are classified into the second groups, while other company are classified into the first group. Among the 59 companies in the second group there are only seven companies (See (*) in Table 10.) that are not in the ming and energy sectors. The first group contains 109 companies among which only seven companies (See (*) in Table 9, 10.) are in the mining and energy sectors. This grouped structure allows us to identify the factor that lies in the intersection of the two group-factor-spaces as the common factor. Further we can identify two orthogonal factors that are both orthogonal to the common factors but lie in the two subspaces respectively as group-pervasive factors. Through this identification we can say, the returns in the resource group are driven by a resource-specific factor and the common factor, while the returns in the non-resource group are driven by the common factor and a nonresource-specific factor.

5 Concluding Remarks

The grouped factor model presented in this paper provides a means to explore potentially existing grouped structures in a large factor model. The proposed estimation procedure will consistently determine the number of groups, endogenously classify variables into groups and provide consistent estimates of the group-pervasive factor space for each group. Thus we offer a method to verify whether uses of non-statistical structural information in classification of variables to conduct grouped factor analysis are statistically adequate or not. The grouped factor model is suitable for analysis of a general configuration of grouped structures: the group-pervasive factors can be disjunctive, orthogonal or intersected with any angles. In particular, it is applicable

for cases in which group-pervasive factors are orthogonal. In these cases, our estimation procedure can be seen as an efficient method to find a set properly rotated factors that allow a better understanding and interpretation of the data. Grouped factor models allow correlations and dependence between group-pervasive factors across groups, our estimation procedure is also applicable for a non-orthogonal factor rotation. More importantly, grouped factor models can be used to assess grouped structures that are otherwise invisible by a factor rotation procedure.

We set up the grouped factor models as approximate factor models which allow certain serial and cross-sectional correlation in the idiosyncratic errors. Therefore they are suitable for applications to economic data and time series data. Simulation study shows that our procedure has good finite sample properties. In an application example we have demonstrated 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 *nonresource* group. Based on the grouped structure we can identify one of the three factors as the common factor, one as the *resource*-specific factor and one as the *nonresource*-specific factor.

In studying factor models with grouped structures, one often asked question is what are the common factors over all groups and what are the group-specific factors (See Goyal et al. (2008), Flury (1984), Flury (1987) and Schott (1999) for more details.) The latter ones defines in fact the groups. Although our grouped factor models do allow the existence common factors and group-specific factors, the proposed estimation procedure, however, does not provide a direct inference on the common factors and the group-specific factors. To integrate this issue into this paper would be a natural choice. Our study sofar (See Chen (2011a)) shows that estimation of common factors and group-specific factors is not a trivial issue. Integrating this issue into the current paper would overstretch this already lengthy paper to an even unacceptable length. Interested readers are referred to Chen (2011a) in which a grouped factor model with common factors and group-specific factors are defined and a procedure is proposed to estimate the common factors as well as group-specific factors. One genuine innovation of this paper is to project the pooled T dimensional data set onto a lower k dimensional data set to achieve the consistent classification. Therefore choice of k is crucial for the proposed procedure. Currently we use the PC criterion of Bai and Ng (2002) to determined k - the dimension of the union of the group-pervasive factor spaces. However, the presence of grouped structures, in particular, the presence of uneven groups tends to deteriorate the performance of the PC criterion (See Boivin and Ng (2006) for more discussions.[12]). Therefore, an improved procedure for the determination of k is an issue which deserves a further investigation. Interested readers are referred to Chen (2011b) for more detained discussions.

[12]The presence of uneven groups can be seen as problems of oversampling and correlation between idiosyncratic components.

6 Appendix

6.1 Example of PDA with a Voting Schema for noisy data

Example 3.1 (continue) *We consider here a set of 8 sample points with noises. The coordinates of the 8 points are collected in a data matrix X. Each row in X' is one sample point.*

$$X' = \begin{pmatrix} 1.0725 & 0.0607 & 0.0943 \\ 0.0603 & 1.0801 & 0.0460 \\ 1.0245 & 1.0977 & 0.0694 \\ 2.0909 & 2.0205 & 0.0854 \\ 0.0493 & 0.0667 & 1.0687 \\ 0.0653 & 0.0385 & 2.0011 \\ 0.0575 & 0.0383 & 3.0351 \\ 0.0857 & 0.0213 & 4.0375 \end{pmatrix} \tag{6.42}$$

Obviously, the first four points are located closely to the subspace of the plane S_2, the next four points are located closely to the subspace of line S_1. The data matrix of the Veronese mapping $\nu_2(\mathbf{x}) = (x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2)'$ is:

$$L_n(\mathbf{X}) = \begin{pmatrix} 1.1588 & 0.0042 & 0.0399 & 0.0000 & 0.0001 & 0.0014 \\ 0.0025 & 0.0522 & 0.0035 & 1.0816 & 0.0716 & 0.0047 \\ 1.0142 & 1.1073 & 0.0196 & 1.2090 & 0.0214 & 0.0004 \\ 4.0604 & 4.1306 & 0.1878 & 4.2020 & 0.1911 & 0.0087 \\ 0.0056 & 0.0017 & 0.0790 & 0.0005 & 0.0235 & 1.1091 \\ 0.0012 & 0.0022 & 0.0702 & 0.0043 & 0.1346 & 4.2418 \\ 0.0097 & 0.0083 & 0.3004 & 0.0072 & 0.2581 & 9.3041 \\ 0.0092 & 0.0076 & 0.3866 & 0.0063 & 0.3210 & 16.2398 \end{pmatrix} \tag{6.43}$$

Since we have noisy data, $L_n(X)$ is of full rank. However, we know that if we had noiseless data the rank of $Null(L_n(X))$ would be two, which is given by the Hilbert function constraint[13]*. We choose the two eigenvectors corresponding to the two smallest singular values as the basis of the nullspace of $L_n(\mathbf{X})$.*

$$\mathbf{c} = \begin{pmatrix} 0.0412 & 0.0782 \\ -0.0286 & -0.0477 \\ -0.4290 & -0.8970 \\ 0.0446 & -0.0123 \\ -0.9007 & 0.4320 \\ 0.0161 & 0.0157 \end{pmatrix}. \tag{6.44}$$

After obtaining $\mathbf{c}$, we can calculate $\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}$ at each sample point. For the three components of the partial derivative, we have:

$$\frac{\partial \nu_n(\mathbf{x})'c}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_1} \\ \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_2} \\ \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_3} \end{pmatrix} = \begin{pmatrix} x_3 & 0 \\ 0 & -x_3 \\ x_1 & -x2 \end{pmatrix}$$

[13] See Yang et al. (2005) for more details.

So, the partial derivative evaluated at $\mathbf{x}^1$ *is:*

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1} = \begin{pmatrix} 0.0461 & 0.0802 \\ -0.1102 & -0.0096 \\ -0.5117 & -0.9328 \end{pmatrix}. \tag{6.45}$$

The partial derivatives evaluated at all sample points are then normalized to be orthogonal and have a unit length. This is done by calculating the principal components of the derivatives using singular value decomposition. For the derivative evaluated at $\mathbf{x}^1$ *given in (6.45) we have the following principal components:*

$$\frac{\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1}}{||\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1}||} = \begin{pmatrix} -0.02 & -0.09 \\ 0.99 & 0.06 \\ -0.06 & 0.99 \end{pmatrix}. \tag{6.46}$$

We give votes to candidates of normal vectors of the subspaces in the following way (see also Algorithm 1). If a normalized derivative at a point $\mathbf{x}^k$ *is similar to a candidate of the normal vectors, this candidate will have one more vote, otherwise the normalized derivative becomes itself a new candidate. The voting procedure is demonstrated in Table 7 and Table 8 in a simplified form.*

We consider first the choice of normal vectors for the subspace of dimension one. Table 7 reports the voting results for different candidates of the normal vectors. The second column collects the normalized partial derivatives evaluated at the corresponding sample points which are given in the first column of Table 7. We start with the row of $\mathbf{x}^1$*. In the third column, * represents that the normalized derivative at the same row is chosen as a candidate. The header* $U\{2\}\{1\}$ *says this is the first candidate for the subspaces with codimension 2. The numbers in this column measure the angels between the candidate and the corresponding partial derivatives at respective rows. A number close to zero means the corresponding angle is small, and a number close to* $\pi/2$ *means the angle is large. In the third column no number is close to zero. Therefore the vote for* $U\{2\}\{1\}$ *is only one. This is given in the fourth column under the header* V*. Now we look at the second row, i.e. the second sample point* $\mathbf{x}^2$*. Since the normalized derivative at* $\mathbf{x}^2$ *has a direction that is not close to the direction of the first candidate* $U\{2\}\{1\}$*, it becomes itself the second candidate under the header* $U\{2\}\{2\}$*. This is symbolized by * in the fifth column and the row of* $\mathbf{x}^2$*. The numbers in the fifth column are not close to zero. This implies that the derivative of* $Dp_n(\mathbf{x})$ *evaluated at other sample points do not have the similar direction as* $U\{2\}\{2\}$*. Hence the vote for the second candidate is also only one, which is given in the sixth column under the header* V*. Similarly,* $DP_n(\mathbf{x})|_{X_3}$ *becomes a new candidate that is given in the seventh column under the header* $U\{2\}\{3\}$*. From the numbers in the seventh column we can see that only* $DP_n(\mathbf{x})|_{\mathbf{x}^4}$ *has a similar direction as* $U\{2\}\{3\}$*. Therefore,* $U\{2\}\{3\}$ *has two votes and* $DP_n(\mathbf{x})|_{\mathbf{x}^4}$ *does not become a new candidate.* $DP_n(\mathbf{x})|_{\mathbf{x}^5}$ *does not have similar directions as the exiting candidates, it becomes the fourth candidate for the normal vectors, which is given in the ninth column under the header* $U\{2\}\{4\}$*. The numbers in the ninth column show that the derivative* $DP_n(\mathbf{x})$ *at* $\mathbf{x}^6$*,* $\mathbf{x}^7$ *and* $\mathbf{x}^8$ *have directions very close to that of* $U\{2\}\{4\}$*. Therefore it has four votes, which are given in the tenth column. Now the fourth candidate has the most votes. The average of* $DP_n(\mathbf{x})$ *at* $\mathbf{x}^5$*,* $\mathbf{x}^6$*,* $\mathbf{x}^7$ *and* $\mathbf{x}^8$ *is the estimate of the normal vectors for the subspace of dimension one and these four sample points are classified to this subspace.*

Table 7: Voting and Choice of Candidates for the Normal Vectors for the Subspace with $k_1 = 1$

Sample	$\frac{Dp_n(x)}{\|\|Dp_n(x)\|\|}$	U{2}{1}	V	U{2}{2}	V	U{2}{3}	V	U{2}{4}	V
$\mathbf{x}^1$	$\begin{bmatrix} -0.02 & -0.09 \\ 0.99 & 0.06 \\ -0.06 & 0.99 \end{bmatrix}$	*	1	0.9789		0.37		0.99	
$\mathbf{x}^2$	$\begin{bmatrix} 0.99 & 0.01 \\ -0.12 & -0.05 \\ -0.02 & 0.99 \end{bmatrix}$	0.97		*	1	0.48		0.99	
$\mathbf{x}^3$	$\begin{bmatrix} 0.63 & -0.02 \\ -0.78 & -0.01 \\ 0.02 & 0.99 \end{bmatrix}$	0.37		0.48		*	2	0.99	
$\mathbf{x}^4$	$\begin{bmatrix} 0.70 & -0.04 \\ -0.72 & -0.00 \\ 0.03 & 0.99 \end{bmatrix}$	0.46		0.39		0.01		0.99	
$\mathbf{x}^5$	$\begin{bmatrix} -0.99 & -0.06 \\ 0.06 & -0.99 \\ 0.00 & -0.05 \end{bmatrix}$	0.99		0.99		0.99		*	4
$\mathbf{x}^6$	$\begin{bmatrix} -0.99 & -0.01 \\ 0.01 & -0.99 \\ 0.01 & -0.00 \end{bmatrix}$	0.99		0.99		0.99		0.002	
$\mathbf{x}^7$	$\begin{bmatrix} 0.99 & -0.02 \\ 0.02 & 0.99 \\ -0.02 & -0.00 \end{bmatrix}$	0.98		0.99		0.99		0.003	
$\mathbf{x}^8$	$\begin{bmatrix} 0.99 & 0.00 \\ -0.00 & 0.99 \\ -0.02 & -0.01 \end{bmatrix}$	0.98		0.99		0.99		0.004	

Notes: The first column gives the sample points from $\mathbf{x}^1$ to $\mathbf{x}^8$. The second column collects the normalized derivatives $Dp_n(\mathbf{x})$ evaluated at corresponding sample points. Third and the fourth column collect the results of evaluation of the first candidate of the normal vectors for the subspace. The number under headers $U\{i\}\{j\}$ are the measures of the angles between the candidate and the respective derivatives at the corresponding rows. The integers under the headers V are the numbers of votes for the corresponding candidate at the same row.

After determining the subspace with $k_i = 1$, we turn to determination of the subspace with $k_i = 2$. The presence of noises makes $Dp_n(\mathbf{x})$ usually a full rank matrix. However, for noiseless cases the rank of $Dp_n(\mathbf{x})$ evaluated at points located in the subspace with $k_i = 2$ is one. Hence, we evaluate only the first principal component of $Dp_n(\mathbf{x})$. The results are collected in the second column of Table 8.

*Table 8 reports the voting results for the candidates of the normal vector for the subspace of dimension two. The second column collects the first principal component of normalized derivatives evaluated at the corresponding sample points. In the third column, * represents that the normalized derivative at the same row is chosen as a candidate. The header $U\{1\}\{1\}$ says that this is the first candidate for the subspace with codimension one. The numbers in this column measure the angels between the candidate and the derivatives at the respective rows. A number close to zero means*

the corresponding angle is small, and a number close to $\pi/2$ means the angle is large. In the third column three numbers are close to zero. Therefore, $U\{1\}\{1\}$ has 4 votes. This is given in the fourth column under the header V. Since the points X_5, X_6, X_7 and X_8 are already classified to the other subspace. $U\{1\}\{1\}$ is the candidate with most votes. Averaging the first principal components for the derivatives at $\mathbf{x}^1$, $\mathbf{x}^2$, $\mathbf{x}^3$ and $\mathbf{x}^4$ gives an estimate for the normal vector of the subspace. These four points are assigned to this subspace accordingly.

Table 8: Voting and Choices of Candidates of the Normal Vectors for the Subspace with $k_i = 2$

Sample	$\frac{Dp(x)}{\|\|Dp(x)\|\|}$	U{1}{1}	V
$\mathbf{x}^1$	$\begin{bmatrix} -0.09 \\ 0.06 \\ 0.99 \end{bmatrix}$	*	4
$\mathbf{x}^2$	$\begin{bmatrix} 0.01 \\ -0.05 \\ 0.99 \end{bmatrix}$	0.0209	
$\mathbf{x}^3$	$\begin{bmatrix} -0.02 \\ -0.01 \\ 0.99 \end{bmatrix}$	0.0069	
$\mathbf{x}^4$	$\begin{bmatrix} -0.04 \\ -0.00 \\ 0.99 \end{bmatrix}$	0.0055	
$\mathbf{x}^5$	$\begin{bmatrix} -0.06 \\ -0.99 \\ -0.05 \end{bmatrix}$	0.9897	
$\mathbf{x}^6$	$\begin{bmatrix} -0.01 \\ -0.99 \\ -0.00 \end{bmatrix}$	0.9961	
$\mathbf{x}^7$	$\begin{bmatrix} -0.02 \\ 0.99 \\ -0.00 \end{bmatrix}$	0.9965	
$\mathbf{x}^8$	$\begin{bmatrix} 0.00 \\ 0.99 \\ -0.01 \end{bmatrix}$	0.9976	

Notes: The second column collect the first principal component of derivative $Dp_n(\mathbf{x})$ evaluated at corresponding sample points. The numbers under the header $U\{1\}\{1\}$ are measures of the angles between the candidate and the corresponding derivatives at the respective rows. The integer 4 under the header V is the number of votes for the candidate normal vector at the same row.

From the voting procedure in Table 7 and Table 8, the estimates of the two subspaces are:

$$\hat{\mathbf{B}}_1 = \begin{pmatrix} 0.9993 & -0.0131 \\ -0.0132 & -0.9992 \\ -0.0135 & -0.0095 \end{pmatrix} \quad \textit{and} \quad \hat{\mathbf{B}}_2 = \begin{pmatrix} -0.0361 \\ 0.0039 \\ 0.998 \end{pmatrix}. \tag{6.47}$$

Compared with equations (3.26), these two estimates of the normal vectors are very close to the true normal vectors.

6.2 Proofs

Proof of **Proposition 2.5**

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.

$$\underset{(k\times k)}{\frac{\Lambda\Lambda'}{N}} = \sum_{i=1}^{n} \frac{N_i}{N} \underset{(k\times k_i)}{C_i} \underset{(k_i\times k_i)}{\frac{\Lambda_i\Lambda_i'}{N_i}} \underset{(k_i\times k)}{C_i}' \tag{6.48}$$

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.

$$\underset{(1\times k)}{\mathbf{b}'} \underset{(k\times k)}{\frac{\Lambda\Lambda'}{N}} \underset{(k\times 1)}{\mathbf{b}} = \sum_{i=1}^{n} \frac{N_i}{N} \underset{(1\times k_i)}{\mathbf{b}'C_i} \underset{(k_i\times k_i)}{\frac{\Lambda_i\Lambda_i'}{N_i}} \underset{(k_i\times 1)}{C_i'\mathbf{b}} \tag{6.49}$$

Because $\frac{\Lambda_i\Lambda_i'}{N_i}$ converges to a positive definite matrix, 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 (6.49) is strictly positive.

□

Proof of **Proposition 3.3**

Since both the ungrouped factor model (2.6) and each group in the grouped factor model (2.4) satisfy the assumptions on a factor model in Bai and Ng (2002). We will extensively applied the results in Bai and Ng (2002) in our proofs. In the following $\xrightarrow{P}$ denotes the probability limit as $T, N \to \infty$.

To prove (c) we need only to show $\frac{1}{T}\hat{G}^{K\prime}E \xrightarrow{P} 0$. Since $\hat{G}_t^K$ corresponds to the factor estimator in Theorem 1 in Bai and Ng (2002), we can directly apply the result of Theorem 1 (in Bai and Ng (2002) p.213) in our proof.

$$\begin{aligned}
\frac{\hat{G}^{K\prime}E}{T} &= \frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K E_t) = \frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K\prime}G_t^o + H^{K\prime}G_t^o)E_t \\
&= \frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K\prime}G_t^o)E_t + \frac{1}{T}\sum_{t=1}^{T} H^{K\prime}G_t^o E_t
\end{aligned}$$

G_t^o and H^K are the true factor and the rotation matrix as defined in Theorem 1 in Bai and Ng (2002). We need to show the two terms in the last equation above converge to zero in probability. For the (l, m) element of the first term, we have by Cauchy-Schwarz inequality:

$$\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_{lt}^K - {H^{K\prime}}_l G_t^o)e_{mt}\right)^2 \leq \frac{1}{T}\sum_{t=1}^{T}(\hat{G}_{lt}^K - {H^{K\prime}}_l G_t^o)^2 \frac{1}{T}\sum_{t=1}^{T} e_{mt}^2$$

According to Theorem 1 in Bai and Ng (2002), we have $\frac{1}{T}\sum_{t=1}^{T}||\hat{G}_t^K - H^{K\prime}G_t^o||^2 \xrightarrow{P} 0$. It follows then

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_{lt}^K - H^{K\prime}{}_l G_t^o)^2 \xrightarrow{P} 0.$$

From Assumption 2.6, we have:

$$\frac{1}{T}\sum_{t=1}^{T} e_{lt}^2 < M_1,$$

where M_1 is a positive constant.

Using Slutsky theorem, it follows then

$$\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_{it}^K - H^{K\prime}{}_i G_t^o)e_{jt}\right)^2 \leq \left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_{it}^K - H^{K\prime}{}_i G_t^o)^2 \frac{1}{T}\sum_{t=1}^{T} e_{jt}^2\right) \xrightarrow{P} 0$$

In the matrix form we have:

$$\operatorname*{plim}_{T,N\to\infty}\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K\prime}G_t^o)E_t = 0.$$

To show $\operatorname*{plim}_{T,N\to\infty}\frac{1}{T}\sum_{t=1}^{T} H^{K\prime}G_t^o E_t = 0$, we need only to show $\operatorname*{plim}_{T,N\to\infty}\frac{1}{T}\sum_{t=1}^{T} G_t^o E_t = 0$. According to Assumption 2.7, we have

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\left|\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t^o e_{it}\right|\right|^2\right) = \frac{1}{N}\sum_{i=1}^{N} E\left|\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t^o e_{it}\right|\right|^2 \leq M$$

It follows then

$$E||\frac{1}{T}\sum_{t=1}^{T} G_t^o e_{it}||^2 \xrightarrow{P} 0,$$

otherwise the inequality above will not hold. This implies $\operatorname*{plim}_{T,N\to\infty}\frac{1}{T}\sum_{t=1}^{T} G_t^o e_{it} = 0$. In matrix form we have

$$\operatorname*{plim}_{T,N\to\infty}\frac{1}{T}\sum_{t=1}^{T} G_t^o E_t = 0.$$

This proves (c) in Proposition 3.3.

To prove (b) we have

$$\begin{aligned}
\bar{F}_i^T = \frac{1}{T}\hat{G}^{K\prime}F_i &= \left(\frac{1}{T}\hat{G}^{K\prime}G^o\right)C_i = \left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K\prime}G_t^o + H^{K\prime}G_t^o)G_t^{o\prime}\right)C_i \\
= &\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K\prime}G_t^o)G_t^{o\prime}\right)C_i + \left(\frac{1}{T}\sum_{t=1}^{T}(H^{K\prime}G_t^o)G_t^{o\prime}\right)C_i \\
= &\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K\prime}G_t^o)G_t^{o\prime}\right)C_i + H^{K\prime}\left(\frac{1}{T}\sum_{t=1}^{T} G_t^o G_t^{o\prime}\right)C_i \\
\xrightarrow{P} &\quad 0 + H^{K\prime}\Sigma C_i \neq 0.
\end{aligned}$$

The limit in the last row above is because of

$$\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_{it}^{K}-H^{K'}{}_{i}G_{t}^{o})G_{jt}^{o}\right)^{2}\leq\frac{1}{T}\sum_{t=1}^{T}\left|\left|\hat{G}_{it}-H^{K'}{}_{i}G_{t}^{o}\right|\right|^{2}\frac{1}{T}\sum_{t=1}^{T}\left|\left|G_{jt}^{0}\right|\right|^{2}\xrightarrow{P}0,$$

and

$$\frac{1}{T}\sum_{t=1}^{T}G_{t}^{o}G_{t}^{o'}\xrightarrow{P}\Sigma.$$

Hence we have $\bar{F}=H^{K'}\Sigma C_i\neq 0$. In the proof above if we use $C_i=I_k$, we have

$$\bar{G}^{T}=\frac{1}{T}\hat{G}^{K'}G\xrightarrow{P}\bar{G}=0+H^{K'}\Sigma\neq 0.$$

So we have $\bar{G}=H^{K'}\Sigma\neq 0$. It follows $\bar{F}_i=\bar{G}C_i$. This proves (b). From the existence of the limit of (b) and (c) follows the existence of the limit of (a).

(d) follows (b) and $C_i\neq C_j$.

According to Assumption 2.2 we have $C_i\lambda_{i,m}\neq C_j\lambda_{j,l}$ for any loadings of group i and group j.

$$\bar{F}_i\lambda_{i,m}-\bar{F}_j\lambda_{j,l}=\bar{G}(C_i\lambda_{i,m}-C_j\lambda_{j,l})\neq 0.$$

This proves f.

□

Now we turn to proof of Proposition 3.7. We have the model selection criterion as follows:

$$PC(n,\{k_i\},\{X_i^{s_n}\})=\sum_{i=1}^{n}\frac{N_i}{N}V_i(k_i,\hat{F}^{k_i},N_i)+\sum_{i=1}^{n}\frac{N_i}{N}\left(k_i+h(\alpha_i)\right)g(N,T)$$

In order to prove this Proposition we compare first the value of the model selection criterion of a true model under a priori true classification with that of an alternative model with a classification determined by PDA procedure. Then we show that the model selection criterion of the true model under the true classification is asymptotically equivalent to the model selection criterion of the true model under the classification determined by PDA procedure.

Since we are considering the asymptotical property of the selection criterion, we assume that in both the a priori correctly classified model and the alternative model each group contains infinitely many variables. The a priori correctly classified model and the alternative model make two different partitions of the variables in n and n' groups respectively. The intersection of these two partitions constitutes a new finer partition of the variables called intersected partition. In each group of the intersected partition, all variables belong to only one group in the true model and they belong to also only to one group in the alternative model. We index the groups in the intersection partition by i. Let k_i^o be the number of the factors of the true model for the variables in group i of the intersection partition and k_i' the estimated number of factors based on the alternative model for the same variables. We can differ three cases:

- **Case 1:** The alternative model underestimates the number of factors in some of its groups. This leads to $k_i' < k_i^o$ for some groups in the intersection partition.

- **Case 2:** The alternative model does not underestimate the number of factors in its groups, and $k_i' = k_i^o$ holds for all groups in the intersection partition.

- **Case 3:** The alternative model does not underestimate the number of factors in its groups and :$k_i' \geq k_i^o$ for all i and $k_i' > k_i^o$ for some groups in the intersection partition.

Let N_i^I be the number of variables in the ith group of the intersection partition. We define several mean squared residuals for the ith group of the intersection partition calculated according to different choices of factors as follows. (Note that the mean squared residuals here are defined in the same way as in Bai and Ng (2002) on page 214.)

- $V(k_i', \hat{F}^{k_i'}, N_i^I)$: the mean squared residuals calculated from the estimated alternative model.

- $V(k_i^o, \hat{F}^{k_i^o}, N_i^I)$: the mean squared residuals calculated from the estimated true model with the a priori true classification .

- $V(k_i^o, F^{k_i^o}, N_i^I)$: the mean squared residuals calculated using k_i^o population factors.

- $V(k_l^o, F^{k_l^o}, N_i^I)$: the mean squared residuals calculated using population factors in the lth group of the alternative model.

- $V(k_i^o, \hat{F}^{k_i^o}_{N_i^I}, N_i^I)$: the mean squared residuals calculated with the estimated factors using only data in the intersection group N_i^I, where the used number of factors is k_i^o.

- $V(k_i', \hat{F}^{k_i'}_{N_i^I}, N_i^I)$: the mean squared residuals calculated with the estimated factors using only data in the intersection group N_i^I, where the used number of factors is k_i'.

Lemma 6.1 *Let $\{N_j\}_{j=1}^n$, $\{N_l^s\}_{l=1}^{n'}$ and $\{N_i^I\}_{i=1}^{n^I}$ denote the indices of the a priori true classification of the true model, the classification using GPCA based on an alternative model and the intersected partition, respectively. It holds*

$$\sum_{j=1}^{n} \frac{N_j}{N} V(k_j^o, \hat{F}^{k_j^o}, N_j) = \sum_{i=0}^{n^I} \frac{N_i^I}{N} V(k_i^o, \hat{F}^{k_i^o}, N_i^I)$$

$$\sum_{j=1}^{n} \frac{N_j}{N} V(k_j^o, F^{k_j^o}, N_j) = \sum_{i=0}^{n^I} \frac{N_i^I}{N} V(k_i^o, F^{k_i^o}, N_i^I)$$

$$\sum_{l=1}^{n'} \frac{N_l^s}{N} V(k_l', \hat{F}^{k_l'}, N_l^s) = \sum_{i=0}^{n^I} \frac{N_i^I}{N} V(k_i', \hat{F}^{k_i'}, N_i^I)$$

$$\sum_{l=1}^{n'} \frac{N_l^s}{N} V(k_l^o, F^{k_l^o}, N_l^s) = \sum_{i=0}^{n^I} \frac{N_i^I}{N} V(k_i^o, F^{k_i^o}, N_i^I)$$

Proof: The above equalities say that the total mean equals the weighted group means. Let $\{z_k\}_{k=1}^N$ be a series with N elements. Suppose that the series is divided into n groups and each group has N_j elements respectively. According to this grouping the element can have two indices: $\{z_{ij}\}$ with $i = 1, 2, ...N_j$ and $j = 1, 2, ..., n$. Now we want to calculate the mean of the series.

$$\bar{z} = \frac{1}{N} \sum_{j=1}^{n} \sum_{i=1}^{N_j} z_{ij} = \sum_{j=1}^{n} \frac{N_j}{N} \frac{1}{N_j} \sum_{i=1}^{N_j} z_{ij} = \sum_{j=1}^{n} \frac{N_j}{N} \bar{z}_j$$

suppose that we have now a different grouping of the series with n^I groups. We have similarly:

$$\bar{z} = \frac{1}{N} \sum_{i=1}^{n^I} \sum_{k=1}^{N_i} z_{ki} = \sum_{i=1}^{n^I} \frac{N_i}{N} \frac{1}{N_i} \sum_{k=1}^{N_i} z_{ki} = \sum_{i=1}^{n^I} \frac{N_i}{N} \bar{z}_i$$

It follows

$$\sum_{j=1}^{n} \frac{N_j}{N} \bar{z}_j = \sum_{i=1}^{n^I} \frac{N_i}{N} \bar{z}_i.$$

Replacing $\bar{z}_j$ and $\bar{z}_l$ in the equation above by $V(k_j^o, \hat{F}^{k_i^o}, N_j)$ and $V(k_i^o, \hat{F}^{k_i^o}, N_i^I)$, we prove the first equality of Lemma 6.1. The other three equalities can be proved in the same way.

Lemma 6.2

$$V(k_i^o, \hat{F}^{k_i^o}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I) = O_p(C_{N,T}^{-2})$$

Proof

The variables in the ith group of the intersected partition belong to only one group of the true model, say group j. Let k_j^o denote the number of true factors in this group. We have $k_j^o = k_i^o$. Since group j with N_j genuine variables of the group satisfy the assumptions on a factor model in Bai and Ng (2002), according to equation (10) in Lemma 4 of Bai and Ng (2002) we have

$$V(k_j^o, \hat{F}^{k_j^o}, N_j) - V(k_j^o, F^{k_j^o}, N_j) = O_p(C_{N,T}^{-2}). \tag{6.50}$$

The difference on the left hand side of the equation above can be written as follows:

$$
\begin{aligned}
& V(k_i^o, \hat{F}^{k_i^o}, N_j) - V(k_i^o, F^{k_i^o}, N_j) \\
= \; & \frac{1}{N_j T}\left(\frac{N_i^I}{N_i^I}\sum_{m=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}'_{k_i^o,m}\hat{F}_t^{k_i^o})^2 + \frac{N_j - N_i^I}{N_j - N_i^I}\sum_{m=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}'_{k_i^o,m}\hat{F}_t^{k_i^o})^2\right) \\
& -\frac{1}{N_j T}\left(\frac{N_i^I}{N_i^I}\sum_{m=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}'_{k_i^o,m}F_t^{k_i^o})^2 + \frac{N_j - N_i^I}{N_j - N_i^I}\sum_{m=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}'_{k_i^o,m}F_t^{k_i^o})^2\right) \\
= \; & \frac{N_i^I}{N_j}\left(\frac{1}{N_i^I T}\sum_{m=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}'_{k_i^o,m}\hat{F}_t^{k_i^o})^2 - \frac{1}{N_i^I T}\sum_{m=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}'_{k_i^o,m}F_t^{k_i^o})^2\right) + \frac{N_j - N_i^I}{N_j} \times \\
& \left(\frac{1}{(N_j - N_i^I)T}\sum_{m=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}'_{k_i^o,m}\hat{F}_t^{k_i^o})^2 - \frac{1}{(N_j - N_i^I)T}\sum_{m=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}'_{k_i^o,m}F_t^{k_i^o})^2\right) \\
= \; & \underbrace{\frac{N_i^I}{N_j}\left(V(k_i^o, \hat{F}^{k_i^o}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I)\right)}_{\text{A}} \\
& + \underbrace{\frac{N_j - N_i^I}{N_j}\left(V(k_i^o, \hat{F}^{k_i^o}, N_j - N_i^I) - V(k_i^o, F^{k_i^o}, N_j - N_i^I)\right)}_{\text{B}} \\
\leq \; & 0.
\end{aligned}
$$

The last inequality is because the the estimated factors minimize the mean squared errors in group j. If we use only data of the N_i^I variables in group i of the intersected partition to estimate factors we have:

$$\underbrace{\frac{N_i^I}{N_j}(V(k_i^o, \hat{F}^{k_i^o}_{Ni^I}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I))}_{\underline{\text{A}}} \leq A$$

and similarly if we use the data of the rest $N_j - N_i^I$ variables in group j of the true model to estimate the factors, we have

$$\underbrace{\frac{N_j - N_i^I}{N_j}(V(k_i^o, \hat{F}^{k_i^o}_{N_j - Ni^I}, N_j - N_i^I) - V(k_i^o, F^{k_i^o}, N_j - N_i^I))}_{\underline{\text{B}}} \leq B.$$

These two inequalities are because the estimated factors minimize of the mean squared residuals in the respective cases. Applying relation (6.50) to the data of N_i^I variables and to the data of $N_j - N_i^I$ variables respectively, under the conditions $\frac{N_i^I}{N_j} \to \eta > 0$ and $\frac{N_j - N_i^I}{N_j} \to 1 - \eta > 0$, we have

$$\underline{A} = O_p(C_{NT}^{-2}) \qquad \text{and} \qquad \underline{B} = O_p(C_{NT}^{-2}).$$

Because $\underline{A} + \underline{B} \leq A + B \leq 0$ and $\underline{A} + \underline{B} = O_p(C_{N,T}^{-2})$ we have $A + B = O_p(C_{N,T}^{-2})$. Since A and B are of same order, we have $A = O_p(C_{N,T}^{-2})$. This proves

$$V(k_i^o, \hat{F}^{k_i^o}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I) = O_p(C_{N,T}^{-2}).$$

□

Lemma 6.3 *For $k_l' \geq k_l^o$,*

$$V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) = O_p(C_{N,T}^{-2}). \tag{6.51}$$

Proof

Since the variables in the ith group of the intersected partition belong to only one group of the true model, say group j, and they belong to also only one group of the alternative model, say group l. Let k_j^o be the number of true factors in group j of the true model and let k_l^o be the number of true factors in group l of the alternative model. So it follows under the condition of Lemma 6.3: $k_i' = k_l' \geq k_l^o \geq k_i^o$.

We reformulate the difference in the left hand side of equation (6.51) into four differences:

$$\begin{aligned} & V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) \\ = \; & V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_l^o, F^{k_l^o}, N_i^I) \\ & + V(k_l^o, F^{k_l^o}, N_i^I) - V(k_l^o, \hat{F}^{k_l^o}_{N_i^I}, N_i^I) \\ & + V(k_l^o, \hat{F}^{k_l^o}_{N_i^I}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I) \\ & + V(k_i^o, F^{k_i^o}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) \end{aligned} \tag{6.52}$$

Now we look at the four differences above in turn. For the first difference we have:

$$\begin{aligned} & V(k_i', \hat{F}^{k_i'}, N_l) - V(k_l^o, F^{k_l^o}, N_l) \\ = \; & V(k_l', \hat{F}^{k_l'}, N_l) - V(k_l^o, F^{k_l^o}, N_l) \\ = \; & \frac{1}{N_l T}\left(\frac{N_i^I}{N_i^I}\sum_{m=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}_{k_l',m}\hat{F}_t^{k_l'})^2 + \frac{N_l - N_i^I}{N_l - N_i^I}\sum_{m=N_i^I+1}^{N_l}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}_{k_l',m}\hat{F}_t^{k_l'})^2\right) \\ & - \frac{1}{N_l T}\left(\frac{N_i^I}{N_i^I}\sum_{m=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}_{k_l^o,m}F_t^{k_l^o})^2 + \frac{N_l - N_i^I}{N_l - N_i^I}\sum_{m=N_i^I+1}^{N_l}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}_{k_l^o,m}F_t^{k_l^o})^2\right) \\ = \; & \frac{N_i}{N_l}\left(\frac{1}{N_i^I T}\sum_{i=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}_{k_l',m}\hat{F}_t^{k_l'})^2 - \frac{1}{N_i^I T}\sum_{m=1}^{N_i^I}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}_{k_l^o,m}F_t^{k_l^o})^2\right) + \frac{N_l - N_i^I}{N_l} \times \\ & \left(\frac{1}{(N_l - N_i^I)T}\sum_{m=N_i^I+1}^{N_l}\sum_{t=1}^{T}(X_{mt} - \hat{\lambda}_{k_l',m}\hat{F}_t^{k_l'})^2 - \frac{1}{(N_l - N_i^I)T}\sum_{m=N_i^I+1}^{N_l}\sum_{t=1}^{T}(X_{mt} - \tilde{\lambda}_{k_l^o,m}F_t^{k_l^o})^2\right) \\ = \; & \frac{N_i}{N_l}\left(V(k_l', \hat{F}^{k_l'}, N_i^I) - V(k_l^o, F^{k_l^o}, N_i^I)\right) \\ + \; & \frac{N_l - N_i^I}{N_l}\left(V(k_l', \hat{F}^{k_l'}, N_l - N_i^I) - V(k_l^o, F^{k_l^o}, N_l - N_i^I)\right) \\ \leq \; & 0. \end{aligned}$$

Applying the same argument as in the proof of Lemma 6.2, we have:

$$V(k_i', \hat{F}^{k_l'}, N_i^I) - V(k_l^o, F^{k_l^o}, N_i^I) = V(k_l', \hat{F}^{k_l'}, N_i^I) - V(k_l^o, F^{k_l^o}, N_i^I) = O_p(C_{NT}^{-2}).$$

For the second difference, using equation (10) in Bai (2003) on page 217, we have

$$V(k_l^o, F^{k_l^o}, N_i^I) - V(k_l^o, \hat{F}^{k_l^o}_{N_i^I}, N_i^I) = O_p(C_{NT}^{-2}).$$

For the third difference we have $k_l^o \geq k_i^o$ where k_i^o is the number of true factors in the ith group of the intersected partition. Using equation (10) in Bai (2003) on page 217, we have

$$V(k_l^o, \hat{F}_{N_i^I}^{k_l^o}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I) = O_p(C_{NT}^{-2}).$$

The fourth difference is not slower than $O_p(C_{N,T}^{-2})$ by Lemma 6.2. Hence We have proved:

$$V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) = O_p(C_{N,T}^{-2}).$$

□

Lemma 6.4 *For $k_i' < k_i^o$,*

$$V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I)$$

has a positive limit.

Proof

$$\begin{aligned}
& V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) \\
\geq\ & V(k_i', \hat{F}_{N_i^I}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) \\
=\ & V(k_i', \hat{F}_{N_i^I}^{k_i'}, N_i^I) - V(k_i', F^{k_i^o} H^{k_i'}, N_i^I) \\
& + V(k_i', F^{k_i^o} H^{k_i'}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I) \\
& + V(k_i^o, F^{k_i^o}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I)
\end{aligned}$$

The first inequality is due to the fact that $\hat{F}_{N_i^I}^{k_i'}$ minimizes the mean squared errors of the estimated factor model for the ith group of the intersected partition with N_i^I variables. Following Lemma 2 and Lemma 3 in Bai and Ng (2002), the first term in the right hand side of the equation is $O_p(C_{N,T}^{-1})$, the second term has a positive limit, and the third term is not slower than $O_p(C_{N,T}^{-2})$ by Lemma 6.2. Hence, $V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I)$ has a positive limit.
□

Proof of Proposition 3.7.
Now we prove Proposition 3.7 in the three possible cases listed before.

Case 1 The alternative model underestimates the number of factors in some of its groups. This leads to $k_i' < k_i^o$ for some groups in the intersected partition.

According to Lemma 6.1 the difference of mean squared residuals between the alternative model and the true model with correct classification can be calculated as follows:

$$\begin{aligned}
& \sum_{l=1}^{n'} \frac{N_l'}{N} V(k_l', \hat{F}^{k_l'}, N_l') - \sum_{j=1}^{n} \frac{N_j}{N} V(k_j^o, \hat{F}^{k_j^o}, N_j) \\
=\ & \sum_{k_i' \geq k_i^o} \frac{N_i^I}{N} (V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I)) + \sum_{k_i' < k_i^o} \frac{N_i^I}{N} (V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I)) \\
=\ & O_p(C_{N,T}^{-2}) + \sum_{k_i' < k_i^o} \frac{N_i^I}{N} [V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I)]
\end{aligned}$$

The first limit in the last row above is by Lemma 6.3. Each summand in the second term has a positive limit by Lemma 6.4. Hence, the left hand side of the equation above also has a positive limit. The difference of the penalties can be calculated as follows:

$$(\bar{k}' - \bar{k}^o + \bar{h}(\{\hat{\alpha}_i'\}) - \bar{h}(\{\hat{\alpha}_i\}))g(N,T).$$

Since $\bar{k}' - \bar{k}^o + \bar{h}(\{\hat{\alpha}_i'\}) - \bar{h}(\{\hat{\alpha}_i\})$ is bounded by condition 3(a), we have

$$(\bar{k}' - \bar{k}^o + \bar{h}(\{\hat{\alpha}_i'\}) - \bar{h}(\{\hat{\alpha}_i\}))g(N,T) \to 0 \qquad \text{as } N,T \to \infty.$$

Therefore,

$$\begin{aligned} & P\{PC(n',\{k_l'\},\{X_l^s\}) > PC^o(n,\{k_j^o\},\{X_j\}) \\ = \; & P\left\{\sum_{l=1}^{n'} \frac{N_l'}{N} V(k_l', \hat{F}^{k_l'}, N_l') - \sum_{j=1}^{n} \frac{N_j}{N} V(k_j^o, \hat{F}^{k_j^o}, N_j^o) > (\bar{k}' - \bar{k}^o + \bar{h}(\{\hat{\alpha}_i'\}) - \bar{h}(\{\hat{\alpha}_i\}))g(N,T)\right\} \\ & \xrightarrow{P} 1, \end{aligned}$$

where we use $PC^o(n,\{k_j^o\},\{X_i\})$ to denote that this model selection value is calculated based on the a priori true classification in the true model and $PC(n',\{k_l'\},\{X_l^s\})$ denotes that the calculation of the model selection criterion value is based on classification using the PDA procedure. The limit in probability in the equation above follows from the fact that the left hand side of the inequality above has a positive limit and the right hand side converges to zero.

Now we turn to the cases when an alternative model overestimates the number of factors.

Case 2 The alternative model does not underestimate the number of factors in its groups, and $k_i' = k_i^o$ for all groups in the intersected partition.

This can only happen when the alternative model separates a group in the true model into more than one groups. Without loss of generality, we consider the case in which the true model is an un-grouped model and the alternative model contains more than one groups. Let the number of the true factors be k^o. We have $k_l' = k^o$. The difference in the penalty factors can be calculated as follows:

$$\sum_{l=1}^{n'} \hat{\alpha}_i \bar{k}_l' - k^o + \sum_{l=1}^{n'} \hat{\alpha}_l h(\hat{\alpha}_l, N, T) - h(1,N,T) = \sum_{l=1}^{n'} \hat{\alpha}_l h(\hat{\alpha}_l, N, T) - h(1,N,T) > 0$$

The last inequality is due to condition 3(c).

$$\begin{aligned} & P(PC^o(1,k^o,X) > PC(n',\{k_l'\},\{X_l^s\}) \\ = \; & P\left\{V(k^o,\hat{F}^o,N) - \sum_{l}^{n'} \frac{N_l'}{N} V(k_l', \hat{F}^{k_l'}, N_l) > \left(\sum_{l=1}^{n'} \hat{\alpha}_l h(\hat{\alpha}_l, N, T) - h(1,N,T)\right) g(N,T)\right\} \\ = \; & P\left\{\sum_{i}^{n_i^I} \frac{N_i^I}{N} V(k_i^o, \hat{F}^o, N_i^I) - \sum_{i}^{n^I} \frac{N_i^I}{N} V(k_i', \hat{F}^{k_i'}, N_i^I) > \left(\sum_{l=1}^{n'} \hat{\alpha}_l h(\hat{\alpha}_l, N, T) - h(1,N,T)\right) g(N,T)\right\} \\ = \; & P\left\{\sum_{i}^{n_i^I} \frac{N_i^I}{N} [V(k_i^o, \hat{F}^o, N_i^I) - V(k_i', \hat{F}^{k_i'}, N_i^I)] > \left(\sum_{l=1}^{n'} \hat{\alpha}_l h(\hat{\alpha}_l, N, T) - h(1,N,T)\right) g(N,T)\right\} \end{aligned}$$

Now the term on the right hand side of the inequality is positive and converges at a slower rate than $C_{N,T}^{-2}$ to zero, and we have $\sum_i^{n_i^I} \frac{N_i^I}{N}[V(k_i^o, \hat{F}^o, N_i^I) - V(k_i', \hat{F}^{k_i'}, N_i^I)] = O_p(C_{NT}^{-2})$ by Lemma 6.3. Hence,

$$P(PC^o(1, k^o, X) > PC(n', \{k_l'\}, \{X_l^s\})) \to 0.$$

This implies

$$P(PC^o(1, k^o, X) < PC(n', \{k_l'\}, \{X_l^s\})) \to 1.$$

Case 3 The alternative model never underestimates the number of factors in its groups and :$k_i' \geq k_i^o$ for all i and $k_i' > k_i^o$ for some groups in the intersected partition. We calculate again the difference in the penalty factors.

$$\begin{aligned}
\phi = & \sum_{l=1}^{n'} \frac{N_l'}{N} k_l' + \sum_{l=1}^{n'} \frac{N_l'}{N} h(\hat{\alpha}_l') - \sum_{j=1}^{n} \frac{N_j}{N} k_j^o - \sum_{j=1}^{n} \frac{N_j}{N} h(\hat{\alpha}_j^o) \\
= & \sum_{i=1}^{n^I} \frac{N_i^I}{N} k_i' + \sum_{i=1}^{n^I} \frac{N_i^I}{N} h(\hat{\alpha}_i') - \sum_{i=1}^{n^I} \frac{N_i^I}{N} k_i^o - \sum_{i=1}^{n^I} \frac{N_i^I}{N} h(\hat{\alpha}_i^o) \\
= & \sum_{i=1}^{n^I} \frac{N_i^I}{N} (k_i' - k_i^0) + \sum_{i=1}^{n^I} \frac{N_i^I}{N} (h(\hat{\alpha}_i') - h(\hat{\alpha}_i^o)) \\
= & \sum_{k_i' > k_i^o} \frac{N_i^I}{N} (k_i' - k_i^0) + \sum_{k_i' > k_i^o} \frac{N_i^I}{N} (h(\hat{\alpha}_i') - h(\hat{\alpha}_i^o)) + \sum_{k_i' = k_i^o} \frac{N_i^I}{N} (h(\hat{\alpha}_i') - h(\hat{\alpha}_i^o)) \\
\geq & \sum_{k_i' > k_i^o} \frac{N_i^I}{N} + \sum_{k_i' > k_i^o} \frac{N_i^I}{N} h(\hat{\alpha}_i') - \sum_{k_i' > k_i^o} \frac{N_i^I}{N} h(\hat{\alpha}_i^o) + \sum_{k_i' = k_i^o} \frac{N_i^I}{N} (h(\hat{\alpha}_i') - h(\hat{\alpha}_i^o)) \\
= & \sum_{k_i' > k_i^o} \frac{N_i^I}{N} (1 - h(\hat{\alpha}_i^o)) + \sum_{k_i' > k_i^o} \frac{N_i^I}{N} h(\hat{\alpha}_i') + \sum_{k_i' = k_i^o} \frac{N_i^I}{N} (h(\hat{\alpha}_i') - h(\hat{\alpha}_i^o)) \\
> & 0
\end{aligned}$$

The first two terms are positive because of condition 3(a) for h function. For the case of $k_i' = k_i^o$ we must have $\hat{\alpha}_i' < \hat{\alpha}_i^o$, because $\hat{\alpha}_i' > \hat{\alpha}_i^o$ would imply that group l of the alternative model contains more variables than group j of the true model, and hence the number of true factors in group l would be larger than k_i^o. This contradicts the assumption of $k_i' = k_i^o$. Therefore the third term is nonnegative according to condition 3(b). Hence, we always have $\phi > 0$.

$$\begin{aligned}
& P(PC^o(n, \{k_j^o\}, \{X_j\}) > PC(n', \{k_l'\}, \{X_l^s\})) \\
= & P\left\{ \sum_{j=1}^{n} \frac{N_j}{N} V(k_j^o, \hat{F}_j^o, N_j) - \sum_{l}^{n'} \frac{N_l}{N} V(k_l', \hat{F}_l', N_l) > \phi g(N,T) \right\} \\
= & P\left\{ \sum_{i=1}^{n^I} \frac{N_i^I}{N} V(k_i^o, \hat{F}_i^o, N_i^I) - \sum_{i}^{n^I} \frac{N_i^I}{N} V(k_i', \hat{F}_i', N_i^I) > \phi g(N,T) \right\} \\
= & P\left\{ \sum_{i=1}^{n^I} \frac{N_i^I}{N} [V(k_i^o, \hat{F}_i^o, N_i^I) - V(k_i', \hat{F}_i', N_i^I)] > \phi g(N,T) \right\}
\end{aligned}$$

Now the term on the right hand side of the inequality is positive and converges at a slower rate than $C_{N,T}^{-2}$ to zero, and we have $\sum_{i=1}^{n^I} \frac{N_i^I}{N}[V(k_i^o, \hat{F}_i^o, N_i^I) - V(k_i', \hat{F}_i', N_i^I)] = O_p(C_{NT}^{-2})$ by Lemma 6.3. Hence,

$$P(PC^o(n, \{k_i^o\}, \{X_j\}) > PC(n', \{k_i'\}, \{X_l^s\})) \to 0.$$

This implies

$$P(PC^o(n, \{k_i^o\}, \{X_j\}) < PC(n', \{k_i'\}, \{X_l^s\})) \to 1.$$

So far we have shown for all three possible cases the following probability convergence holds.

$$P(PC^o(n, \{k_j^o\}, \{X_j\}) < PC(n', \{k_l'\}, \{X_l^s\})) \to 1. \tag{6.53}$$

Since the true classification is usually unknown in practical applications, we need to replace the true classification by the classification using the PDA procedure and we need to prove that the model selection criterion of the true model using the PDA procedure has the same property as given in (6.53), i.e. we need to prove

$$P(PC(n, \{k_j^o\}, \{X_j^s\}) < PC(n', \{k_l'\}, \{X_l^s\})) \xrightarrow{P} 1 \qquad \text{as } T, N \to \infty.$$

$$\begin{aligned} & \underbrace{PC(n, \{k_j^o\}, \{X_j^s\}) - PC(n', \{k_l'\}, \{X_l^s\})}_{\text{A}} \\ = \; & \underbrace{PC(n, \{k_j^o\}, \{X_j^s\}) - PC^o(n, \{k_j^o\}, \{X_j\})}_{\text{B}} \\ + \; & \underbrace{PC^o(n, \{k_j^o\}, \{X_j\}) - PC(n', \{k_l'\}, \{X_l^s\})}_{\text{C}} \end{aligned}$$

Because the PDA with the voting scheme is consistent we have

$$P\left[PC(n, \{k_j^o\}, \{X_j^s\}) - PC^o(n, \{k_j^o\}, \{X_j\}) = 0\right] = P\left(\{X_j^s\} = \{X_j\}\right) \to 1 \tag{6.54}$$

Because $\operatorname{plim}_{T,N\to\infty} B = 0$, $\operatorname{plim}_{T,N\to\infty} C < 0$ and $A = B + C$, we have

$$\operatorname*{plim}_{T,N\to\infty} A = \operatorname*{plim}_{T,N\to\infty} B + \operatorname*{plim}_{T,N\to\infty} C < 0.$$

This means

$$P(PC(n, \{k_j^o\}_{j=1}^n, \{X_j^s\}) < PC(n', \{k_l'\}_{l=1}^{n'}, \{X_l^s\})) \to 1 \qquad \text{as } T, N \to \infty.$$

This proves Proposition 3.7.

□

Proof of **Proposition 3.8**

Let $\hat{F}_{i,t} = \hat{F}_{i,t}(X_i^s)$ denotes the factor estimate calculated with the data classified into the ith group and $\hat{F}_{i,t}(X_i)$ denote a factor estimate calculated with the genuine data of the ith group. Let $H_i^{k_i'}$ be the $H^{k'}$ matrix defined in Theorem 1 in Bai and Ng (2002).

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t} - H_i^{k_i'}F_{i,t}||^2 \\ \leq\ & \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i^s) - \hat{F}_{i,t}(X_i)||^2 + \frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i) - H_i^{k_i'}F_{i,t}||^2 \\ & +2\left(\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i^s) - \hat{F}_{i,t}(X_i)||^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i) - H_i^{k_i'}F_{i,t}||^2\right)^{\frac{1}{2}} \end{aligned}$$

Following Proposition 3.5, we have

$$P\left(\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i^s) - \hat{F}_{i,t}(X_i)||^2 = 0\right) = P\left(\{X_j^s\} = \{X_j\}\right) \to 1, \text{ as } T \to \infty, N \to \infty. \tag{6.55}$$

This implies

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i^s) - \hat{F}_{i,t}(X_i)||^2 \xrightarrow{P} 0. \tag{6.56}$$

Since the data of the ith group satisfy the assumptions of the factor model in Bai and Ng (2002), we can apply Theorem 1 in Bai and Ng (2002) and have

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i) - H_i^{k_i'}F_{i,t}||^2 = O_p(C_{N,T}^{-2}), \tag{6.57}$$

which implies

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t}(X_i) - H_i^{k_i'}F_{i,t}||^2 \xrightarrow{P} 0. \tag{6.58}$$

Hence we have

$$\frac{1}{T}\sum_{t=1}^{T}||\hat{F}_{i,t} - H_i^{k_i'}F_{i,t}||^2 \xrightarrow{P} 0. \tag{6.59}$$

□.

6.3 Variable List for the Empirical Example

Table 9: List of Variables and Classification

Group	No.	Name	code
1	2	ADELAIDE BRIGHTON -	A:ABCX(RI)
1	3	ABACUS PROPERTY GROUP -	A:ABPX(RI)
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	9	AUSTRALAND PR.GP. -	A:ALZX(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	32	BLUESCOPE STEEL -	A:BSLX(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	46	CRANE GROUP -	A:CRGX(RI)
1	47	CSL - TOT RETURNIND	A:CSLX(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	54	DOWNER EDI -	A:DOWX(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	64	FKP PROPERTY GROUP -	A:FKPX(RI)
1	65	FLIGHT CENTRE -	A:FLTX(RI)
1	67	FLEETWOOD -	A:FWDX(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	74	GUD HOLDINGS -	A:GUDX(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	84	ILUKA RESOURCES -	A:ILUX(RI)(*)
1	85	ING OFFICE FUND -	A:IOFX(RI)
1	87	IRESS MARKET TECH. -	A:IREX(RI)
1	88	ISOFT GROUP -	A:ISFX(RI)
1	90	JB HI-FI -	A:JBHX(RI)
1	91	JAMES HARDIE INDS.CDI. -	A:JHXX(RI)
1	93	LEIGHTON HOLDINGS -	A:LEIX(RI)
1	95	LEND LEASE GROUP-	A:LLCX(RI)
1	97	MACMAHON HOLDINGS -	A:MAHX(RI)
1	98	MAP GROUP -	A:MAPX(RI)
1	101	MACQUARIE COUNTRY.TRUST -	A:MCWX(RI)
1	102	MIRVAC GROUP -	A:MGRX(RI)
1	104	MACQUARIE INFR.GROUP -	A:MIGX(RI)
1	107	MONADELPHOUS GROUP -	A:MNDX(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	115	NUFARM -	A:NUFX(RI)
1	116	NEWS CORP.CDI.'B' (ASX) -	A:NWSX(RI)
1	122	ONESTEEL -	A:OSTX(RI)(*)
1	126	PRIME INFRASTRUCTURE GP. -	A:PIHX(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)

Table 10: List of Variables and Classification(Cont.)

Group	No.	Name	code
1	133	QBE INSURANCE GROUP -	A:QBEX(RI)
1	134	RAMSAY HEALTH CARE -	A:RHCX(RI)
1	137	RESMED CDI -	A:RMDX(RI)
1	141	SEVEN NETWORK -	A:SEVX(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	147	SMS MAN.& TECH. -	A:SMXX(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	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)
2	1	AUSTRALIAN AGRICULTURAL -	A:AACX(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	23	ALUMINA -	A:AWCX(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	39	CENTENNIAL COAL -	A:CEYX(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	71	GRAINCORP -	A:GNCX(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	96	LYNAS -	A:LYCX(RI)
2	99	MACARTHUR COAL -	A:MCCX(RI)
2	100	MINCOR RESOURCES-	A:MCRX(RI)
2	103	MOUNT GIBSON IRON -	A:MGXX(RI)
2	105	MEDUSA MINING -	A:MMLX(RI)
2	106	MURCHISON METALS-	A:MMXX(RI)
2	109	MOLOPO ENERGY -	A:MPOX(RI)
2	111	MINARA RESOURCES-	A:MREX(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	152	TABCORP HOLDINGS-	A:TAHX(RI)(*)
2	156	TELSTRA -	A:TLSX(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. (2003). Inference on factor models of large dimensions. *Econometrica*, 71:135–172.

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

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132:169–194.

Chen, P. (2011a). Common factors and group-specific factors. *Mimeo, Melbourne Institute of Technology*.

— (2011b). Estimation of grouped factor models. *Mimeo, Melbourne Institute of Technology*.

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.

Heaton, C. and Solo, V. (2009). Grouped variable approximate factor analysis. *15th International Conference:Computing in Economics and Finance*.

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

Ludvigson, S. C. and Ng, S. (2009). A factor analysis of bond risk premia. *NBER Working Paper No. 15188*.

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.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Jounal of Business and Economic Statistics*, 20:147–162.

Vidal, R., Ma, Y., and Piazzi, J. (2004). A new gpca algorithm for clustering subspaces by fitting, differentiating and dividing polynomials. *CVPR*, page 510.

Vidaly, R. (2003). Generalized principal component analysis (gpca): an algebraic geometric approach to subspace clustering and motion segmentation. *A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy*.

Vidaly, R., Ma, Y., and Sastry, S. (2003). Generalized principal component analysis (gpca). *Uncertainty in Artificial Intelligence*, pages 255–268.

Yang, A. Y., Rao, S., Wagner, A., Ma, Y., and Fossum, R. M. (2005). Hilbert functions and applications to the estimation of subspace arrangements. *ICCV*.

Yedla, M., Pathakota, S. R., and Srinivasa, T. M. (2010). Enhancing k-means clustering algorithm with improved initial center. *International Journal of Computer Science and Information Technologies*, 1 (2):121–125.

Zhang, C. and Xia, S. (2009). K-means clustering algorithm with improved initial center. *Proceedings of Second International Workshop on Knowledge Discovery and Data Mining*, pages 790–792.