A Grouped Factor Model

Chen, Pu

Melbourne University

1 October 2010

A Grouped Factor Model

Pu Chen[*]

01.10.2010

Abstract

In this paper we present a grouped factor model that is designed to explore grouped structures in factor models. We develop an econometric theory consisting of a consistent classification rule to assign variables to their respective groups and a class of consistent model selection criteria to determine the number of groups as well as the number of factors in each group. As a result, we propose a procedure to estimate grouped factor models, in which the unknown number of groups, the unknown relationship between variables to their groups as well as the unknown number of factors in each group are statistically determined based on observed data. The procedure can help to estimate common factor that are pervasive across all groups and group-specific factors that are pervasive only in the respective groups. Simulations show that our proposed estimation procedure has satisfactory finite sample properties.

KEYWORDS: Factor Models, Generalized Principal Component Analysis

JEL Classification: C1, C32, E24,

[*] Melbourne University, Alan Gilbert Building, Parkville, Victoria 3010, Australia. Tel.: 0061 424789545, E-Mail: chenpu60@hotmail.com

1 Introduction

Factor models are widely used to summarize common features in large data sets, such that the behavior of a large number of observed variables can be explained by a small number of unobserved variables called factors. This class of models have been successfully applied, for example, in finance to model asset returns known as arbitrage pricing theory (see Ross 1985 for more details), in applied macroeconomics to construct coincident indices to describe business cycles and to forecast macroeconomic variables (see Stock and Watson (2002b) for more details), and in marketing to identify the salient attributes consumers use to evaluate products. Often the large number of variables can be grouped into different categories. For example asset returns consist of asset returns in different industries; macrovariables 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 the data, in particular, for explaining the group-specific behavior of the data. So, for example, industrial indices are used to measure industry specific risks that can in turn explain the asset returns of the respective industries (see Fama Frence 1975 for more details). Regarding forecasting Boivin and Ng (2006) find grouped data can produce better forecasts.

While a factor model is capable of summarizing main characteristics among a large number of variables, it ignores potentially existing group-specific common features, which may be useful in studying the data-generating mechanism. Although there is a vast literature on factor models[1] little is written about factor models with grouped structures. Boivin and Ng (2006) find that in the forecasting of US macrovariables, group-specific factors produce better results. Ludvigson and Ng (2009) analyze the relation between bond excess returns and macro economic variables. They use 8 group-specific factors extracted from 131 variables to explain the bond risk premia. In the context of estimation of group-specific factors and common

[1] see Johnson and Wichern (1992), Stock and Watson (2002a), Connor and Korajzyk (1986), Bai (2003) and Bai and Ng (2002) for more details.

factors the first class of methods deal with situation in which the group-specific factor spaces and the common factor space are the same. Krzanowski (1979) proposes to determine the common factor space by minimizing its angles to the group-specific factor spaces. Flury (1984) and Flury (1987) consider the case in which the group-specific covariance matrices can be orthogonalized by a same matrix. This method is then extended by Schott (1999) to take into account of the situation in which the group-specific factor spaces are only subspaces of the common factor space respectively. He suggests to estimate the common factor space by applying principal components method to the sum of the eigenprojection of each group. Goyal, Perignon, and Villa (2008) apply this method to study the asset returns in NYSE and NASDAQ and find that these two markets share one common factor and each market has one group-specific factor respectively. Heaton and Solo (2009) study a grouped factor model, where the groups are based on non-vanishing cross-sectional correlation among the residuals within a group and they provide an instrumental variable approach to estimate the model.

Common to these papers, not only the number of groups in the factor models is assumed a priori but also the grouping of variables is given a priori rather than estimated from observed data. Our paper contributes to the literature on factor models in that it presents a theory on grouping the variables, determination of the number of groups and estimation of the group-specific factors, such that the grouped structures are determined statistically from data. The paper is structured as follows. Section 2 presents the grouped factor model and discusses its relation to a conventional factor model. Section 3 is on the estimation of grouped factor models, where we present a method of generalized principal component analysis (GPCA) and establish the consistency of the classification of variables based on GPCA. We develop then a class of consistent model selection criteria to determine the number of groups as well the number of factors in each group. Section 4 documents some simulation results on the performance of the estimation procedure for grouped factor models. 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 of variables:

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

Further we assume that the variables in each group are generated from a factor model. For the jth variable of ith group at time t we have

$$\underset{(1\times1)}{X_{i,jt}} = \underset{(1\times k_i)}{\lambda'_{i,j}} \underset{(k_i\times1)}{F_{i,t}} + \underset{(1\times1)}{e_{i,jt}}, \qquad \text{for } j = 1,2,...N_i, t = 1,2,...,T, i = 1,2,...,n, \tag{2.2}$$

where $F_{i,t}$ is a k_i dimensional common factor of the ith group at time t and $\lambda_{i,j}$ is the 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 $X_{i,jt}$.

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

$$\underset{(T\times1)}{X_{i,j}} = \underset{(T\times k_i)}{F_i} \underset{(k_i\times1)}{\lambda_{i,j}} + \underset{(T\times1)}{e_{i,j}}, \qquad \text{for } j = 1,2,...N_i, i = 1,2,...,n, \tag{2.3}$$

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 variables in the ith group. We have

$$X_i = F_i\Lambda_i + E_i, \qquad \text{for } i = 1,2,...,n, \tag{2.4}$$

where

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

ith group over T periods.

- $\sum_{i=1}^{n} N_i = N$ and $N_i/N \rightarrow \alpha_i > \underline{\alpha}$ for all i, where $\underline{\alpha}$ is a positive constant.

We call the model in (2.4) a grouped factor model (GFM). We consider the situation where the number of groups n, the membership relation between variables and groups, and the number of factors in each group are not a priori given. Our objective is to determine the number of groups, the membership relation between variables and groups, and the group-specific factors.

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 a factor model from the literature. The model setup in Bai and Ng (2002) serves well for this purpose. It is general enough for applications using economic time series. Further the technique there fits well to investigate grouped structure in a factor model as we will see later. Therefore we adopt the model assumptions on a factor model in Bai and Ng (2002) in our paper.

2.1 Assumptions

Assumption 2.1

(a) We assume that group-specific factors $F_{i,t}$ are generated from a k dimensional $overall factor$ G_t with $k \leq \sum_{i=1}^{n} k_i$ in the following way:

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

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 dependence of group-specific factors of one group on those of the other groups. If $k < \sum_{i=1}^{n} k_i$, group-specific factors will be linear dependent across groups. For instance, with $n = 3$, $k_1 = 2$ and $k_2 = 2$, $k_3 = 1$ and $k = 3$ we are considering a 3 dimensional factor space consisting of two factor planes and one factor line. These three sets of group-specific factors are not independent from each other and each of them can be represented as a linear combination of the 3 dimensional overall factor G_t. If $k = \sum_{i=1}^{n} k_i$, the overall factor G_t is just the collection of all group-specific factors after some rotations. Assumption 2.1 (b) is made to ensure group-specific factors are not liner dependent within a group. (c) is to make sure that every component of the overall factor G_t is used in generating the group-specific factors. Under Assumption 2.1, the set of N variables X adopt a factor structure with G as the factor:

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

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 an ungrouped factor model with k factors.

Assumption 2.2
$E||G_t||^4 < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} G_t G_t' \xrightarrow{P} \Sigma$ as $T \to \infty$ for some positive definite matrix Σ.

Assumption 2.2 is standard in a factor model. Under Assumption 2.1 and Assumption 2.2 it is easy to see that the group-specific factor F_i also satisfies the requirements of Assumption 2.2, i.e.

(1) $E||F_{i,t}||^4 = E||C_i G_t||^4 < \infty$

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

Assumption 2.3

$\lambda_{i,j} < \lambda < \infty$ 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 2.3 is to make sure that each component of a group-specific factor makes a nontrivial contribution to the variance of the variables in the group.

Proposition 2.4

Under Assumption 2.3 and Assumption 2.1 (b) and (c), the factor loading matrix Λ in the ungrouped model (2.6) satisfies the requirement in Assumption 2.3, 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.

Assumption 2.5

(a) There is no constant $k_j \times k_i$ matrix C such that $C_i = C_j C$, for any $i \neq j$, $i = 1, 2, ..., n$ and $j = 1, 2, ..., n$.

(b) Any pair of loading vectors of 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:

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

Assumption 2.5 is about identification of groups. Assumption 2.5 (a) says no group-specific factors are linear combinations of those of another group. In the case with two factor planes and one factor line, this assumption 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. This assumption is to make sure that groups are identified. Assumption 2.5 (b) says that data points located in the intersection of two groups are events of probability zero. $C_i\lambda_{i,m} \neq C_j\lambda_{j,l}$ implies $C_i\lambda_{i,m} - C_j\lambda_{j,l} \neq 0$. Hence we have $P(G(C_i\lambda_{i,m} - C_j\lambda_{j,l}) = 0) = P(F_i\lambda_{i,m} - F_j\lambda_{j,l} = 0) = 0$. This event can be reformulated as $P(F_i\lambda_{i,m} = F_j\lambda_{j,l}) = 0$. Now $F_i\lambda_{i,m}$ and $F_j\lambda_{j,l}$ represent two points (without errors) in different factor spaces. Assumption 2.5 (b) excludes the situation in which a data point lies in the intersection of the factor spaces of two groups[2].

In order to apply grouped factor models to economic time series, both serial correlation and cross-sectional correlation among idiosyncratic errors are to be considered. We assume that the idiosyncratic errors in the ungrouped model (2.6) satisfy the assumptions made as given in Bai and Ng (2002). Let X_{it} denote the observation of the ith variable at time t in X. 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. $E(\sum_{i=}^{N} e'_{is}e_{it}/N) = (N^{-1}\sum_{i=1}^{N} e_{is}e_{it} = \gamma_N(s,t)), |\gamma_N(s,s)| \leq M$ 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, $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} |\tau_{ij}| < M$

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)

$$E\left(\frac{1}{N}\sum_{j=1}^{N}\left|\left|\frac{1}{\sqrt{T}}G_t^{0\prime}e_{jt}\right|\right|^2\right) \leq M.$$

Note that the group-specific idiosyncratic errors in the grouped factor model are the same as the idiosyncratic errors in the ungrouped factor model. Therefore, the

[2] This is a technical assumption to simplify the presentation of a correct classification of variables. See remarks of Proposition 3.5 for more details.

group-specific factors and the group-specific errors also satisfy the weak dependence Assumption 2.7.

It is to note that under Assumption 2.1 through Assumption 2.7, the ungrouped factor model (2.6) satisfies the assumptions on factor models given in Bai and Ng (2002), and each group in (2.4) also satisfies the assumptions on factor models given in Bai and Ng (2002).

The benefit of studying the grouped factor model (2.4) instead of the ungrouped factor model (2.6) is to obtain group-specific factors, which may be useful for group-specific analysis. If we understand a factor model as a means of condensing information from a large number of variables to a small number of factors, then the grouped factor model (2.4) investigates the structure in the factor space, i.e. which parts of variables are influenced by which factors.

While an ungrouped factor model (2.6) with $k = 3$ says the factor space is a 3 dimensional space spanned by the factor G, a grouped factor model (2.4) with $n = 3$, $k_1 = 2$ $k_2 = 2$, $k_3 = 1$ and $k = 3$ says this 3 dimensional factor space consists of two factor planes and a factor line. Through the grouped factor model we say data points are located around the two planes and the line inside a 3 dimensional factor space instead of just saying that the points are located around a 3 dimensional factor space. In this sense, the grouped factor model provides more detailed information on data structure than the ungrouped factor model.

3 Estimation of GFM

Suppose that we know the number of groups $n \in \mathbb{N}$ as well as the correct grouping $s_n \in S$, where $\mathbb{N}$ is the set of natural number and S_n is the set of all possible grouping of variables given n. Then the estimation problem can be solved group by group using principal component method that corresponds to the minimization of the the squares residuals in each group. If the number of groups and the grouping of the variables are unknown, we may 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 the main feature 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 groups and the unknown factors the respective groups. This problem can be seen as a problem of high dimensional clustering in which the cluster centers are subspaces of different dimensions instead of centriods. Intuitively, we could apply the standard method to estimate the factors and the corresponding loadings for each group, if we knew the membership relation between variables and groups. On the other hand, if we knew the factors in each group we could use the factor models to classify the variables to each group using some minimum-distance criteria. So, we are in the dilemma of chicken or the egg. 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 sensitively on starting values[3]. A thorough search over all groupings is NP-hard even in the case of two groups[4]. In this paper we adopt the idea of generalized principal component analysis[5] 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 an ungrouped factor model (2.6) says the N sample points are locates close

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

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

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

to a k dimensional factor space spanned by G, a grouped factor model (2.4) says the k dimensional factor space consists of n different subspaces spanned by F_i with dimension k_i for $i = 1, 2, ..., n$, respectively, and the N sample points are located close to these n subspaces. The magnitude of the idiosyncratic components E_i measure how close the points in X_i are located to their subspace.

Denote the 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) each subspace is represented by the basis F_i spanning the subspace, in equation (3.8) the subspace is represented by its normal vectors $\mathbf{B}_i$. For a point $\tilde{\mathbf{x}}^j$ lying in one of the n subspaces we have:

$$\prod_{i=1}^{n} ||\mathbf{B}_i' \tilde{\mathbf{x}}^j|| = 0 \qquad \text{for } j = 1, 2, ...N, \tag{3.9}$$

where || || is the Euclidian norm in vector spaces. Equations (3.8) and (3.9) are an alternative representation of the grouped factor model (2.4). To estimate the number of groups and the number of factors in each group is equivalent to estimate the number of subspaces and their dimensions. The estimation of a grouped factor model involves two tasks: classification of the data into n groups and estimation of the subspace of each group. The difficulty of the problem lies in solving the classification and estimation problem simultaneously.

3.2 Method of Generalized Principal Component Analysis(GPCA)

Principal component analysis can be seen as a problem of estimating a linear subspace of unknown dimension k from N sample points. Generalized principal component analysis refers to a problem of estimating an unknown number n of linear subspaces with unknown dimensions k_i (i =1,2,...n) from N sample points.

As discussed in the last subsection, a variable $\mathbf{x}$ lying in one of the subspaces must satisfy the following equation:

$$\prod_{i=1}^{n} ||(\mathbf{B}_i'\mathbf{x})|| = 0. \tag{3.10}$$

The factors of the product on the left hand side of equation (3.10) can be reformulated as a collection of $m = \prod_{i=1}^{n}(T - k_i)$ equations of homogeneous polynomials of degree n:

$$\begin{aligned} & \prod_{i=1}^{n} ||(\mathbf{B}_i'\mathbf{x})|| = \prod_{i=1}^{n} ||((\mathbf{b}_{i1}, \mathbf{b}_{i2}, ...\mathbf{b}_{i(T-k_i)})'\mathbf{x})|| = 0 \\ \Longleftrightarrow \quad & p_n(\mathbf{x}) = (p_{n1}(\mathbf{x}), p_{n2}(\mathbf{x}), ..., p_{nm}(\mathbf{x})) = 0. \end{aligned} \tag{3.11}$$

In other words 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 $\mathbf{B}_2 = (\mathbf{b}_{21})$.

$$\prod_{i=1}^{2} ||(\mathbf{B}_i'\mathbf{x})|| = 0 \Longleftrightarrow p_2(\mathbf{x}) = ((\mathbf{b}_{11}'\mathbf{x})(\mathbf{b}_{21}'\mathbf{x}), (\mathbf{b}_{12}'\mathbf{x})(\mathbf{b}_{21}'\mathbf{x})) = 0. \tag{3.12}$$

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

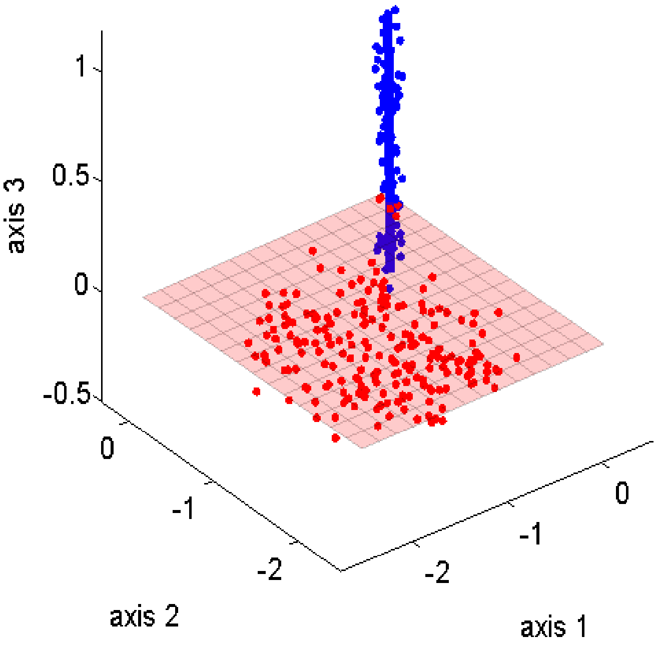

Figure 1: GPCA for $n = 2$, $k_1 = 1$, $k_2 = 2$, $N = 200$, $T = 3$

we have

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \textit{and} \quad B_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{3.13}$$

The polynomials representing the two subspaces are:

$$p_2(\mathbf{x}) = ((\mathbf{b}'_{11}\mathbf{x})(\mathbf{b}'_{21}\mathbf{x}), (\mathbf{b}'_{12}\mathbf{x})(\mathbf{b}'_{21}\mathbf{x})) = (x_1x_3, x_2x_3) = 0. \tag{3.14}$$

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:

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

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

$$\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_1} = (\mathbf{b}_{11}(\mathbf{b}'_{21}\mathbf{x}), \mathbf{b}_{12}(\mathbf{b}'_{21}\mathbf{x}))\,. \tag{3.16}$$

Normalizing the derivative above we obtain:

$$\frac{\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_1}}{||\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_1}||} = (\mathbf{b}_{11}, \mathbf{b}_{12}) = \mathbf{B}_1. \tag{3.17}$$

Similarly, we have

$$\frac{\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_2}}{||\frac{\partial p_2(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}\in S_2}||} = (\mathbf{b}_{21}, \mathbf{b}_{21}) = \mathbf{B}_2. \tag{3.18}$$

The calculation above shows that the subspaces can be represented as the null space of $p_n(\mathbf{x})$, and the normal vectors of the subspaces can be obtained from the derivative of $p_n(\mathbf{x})$ with respect to $\mathbf{x}$ evaluated at one point in the respective subspaces. This fact holds for a general subspace arrangement, as stated later in Proposition 3.2.

Differentiating $p_n(\mathbf{x})$ to obtain the normal vectors of the subspaces provides one way to solve for the subspaces from the data. The question is now how can we obtain the polynomial $p_n(\mathbf{x})$, when the normal vectors of the subspaces are unknown? 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 $0 \leq n_j \leq n$ 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.

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

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 $\nu_n(\mathbf{x}) : \mathbb{R}^T \rightarrow \mathbb{R}^{M_n}$ with $M_n = \begin{pmatrix} n+T-1 \\ T-1 \end{pmatrix}$. $\nu_n : (x_1, ..., x_T)' \rightarrow (..., \mathbf{x}^I, ...)'$, where $\mathbf{x}^I = x_1^{n_1}x_2^{n_2}...x_T^{n_T}$ with $0 \leq n_j \leq n$ 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 3.14):

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

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

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

$L_n(\mathbf{X})$ is an $N \times M_n$ matrix. $L_n(\mathbf{X})\mathbf{c} = 0$ implies that coefficient $\mathbf{c}$ can be calculated from the eigenvectors of the null space of $L_n(\mathbf{X})$. Once we have $\mathbf{c}$, we have a representation of the subspaces $\nu_n(\mathbf{x})'\mathbf{c} = 0$. This suggests 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 out 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 sub-*

spaces. 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, 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:

$$L_n(X) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 4 & 4 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 0 & 0 & 16 \end{pmatrix} \tag{3.22}$$

From $L_n(\mathbf{X})$ we can solve for its null space by singular value decomposition. We obtain two eigenvectors of $Null(L_n(\mathbf{X}))$:

$$\mathbf{c} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}. \tag{3.23}$$

As stated in Proposition 3.2, the two polynomials $p_n(\mathbf{x})$ that represent the subspaces can be obtained in the form of $\nu_n(\mathbf{x})'\mathbf{c} = 0$ with coefficients equal to the eigenvectors in the null space of $L_n(\mathbf{X})$. In this example we have

$$\nu_n(\mathbf{x})'\mathbf{c} = (x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2)\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix} = (x_1x_3, -x_2x_3) = 0.$$

Comparing with equation (3.14), we know $p_n(\mathbf{x}) = \nu_n(\mathbf{x})'\mathbf{c} = 0$ represents the two subspaces the line $S_1 = \{\mathbf{x}|x_1 = 0, x_2 = 0\}$ and the plane $S_2 = \{\mathbf{x}|x_3 = 0\}$.

Since we have data for $\nu_n(\mathbf{x})$, we can calculate $v_n(\mathbf{x})'\mathbf{c}$ and $\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}$. According to Proposition 3.2, the normal vectors of the subspaces can be calculated by evaluating

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

at one point in the respective subspace. For the three components of the derivative above we have:

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_1} = \frac{\partial \nu_n(\mathbf{x})'}{\partial x_1}\mathbf{c} = (2x_1, x_2, x_3, 0, 0, 0)\mathbf{c} = (x_3, 0)$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_2} = \frac{\partial \nu_n(\mathbf{x})'}{\partial x_2}\mathbf{c} = (0, x_1, 0, 2x_2, x_3, 0)\mathbf{c} = (0, -x_3)$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_3} = \frac{\partial \nu_n(\mathbf{x})'}{\partial x_3}\mathbf{c} = (0, 0, x_1, 0, x_2, 2x_3)\mathbf{c} = (x_1, -x_2).$$

Evaluating the three components of the partial derivative at the 8 sample points is to replace (x_1, x_2, x_3) in the three formulas above by the corresponding coordinates of the 8 sample points, i.e. the numbers in the data matrix X in (3.21). We obtain *then the following three matrices:*

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_1}|_X = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 2 & 0 \\ 3 & 0 \\ 4 & 0 \end{pmatrix}, \quad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_2}|_X = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & -2 \\ 0 & -3 \\ 0 & -4 \end{pmatrix}, \quad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_3}|_X = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & -1 \\ 2 & -2 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{3.24}$$

The three components of the partial derivative $\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}$ evaluated at each sample point are given in the corresponding rows in the three matrices above. Collecting the partial derivatives evaluated at $\mathbf{x}^1$ to $\mathbf{x}^8$, we have:

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.25}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^3} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & -1 \end{pmatrix}, \qquad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^4} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 2 & -2 \end{pmatrix}, \tag{3.26}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^5} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}, \qquad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^6} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \\ 0 & 0 \end{pmatrix}, \tag{3.27}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^7} = \begin{pmatrix} 3 & 0 \\ 0 & -3 \\ 0 & 0 \end{pmatrix} \quad \textit{and} \quad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^8} = \begin{pmatrix} 4 & 0 \\ 0 & -4 \\ 0 & 0 \end{pmatrix}. \tag{3.28}$$

Notice 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[6] *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 component is (0 0 1)'. Therefore these four points belong to the subspace S_2 defined by the normal vector*

$$\mathbf{B}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{3.29}$$

The normalized derivative for points $\mathbf{x}^5, \mathbf{x}^6, \mathbf{x}^7, \mathbf{x}^8$ is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix} \tag{3.30}$$

Hence these four points belong to the subspace S_1 characterized by the normal vectors:

$$\mathbf{B}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}. \tag{3.31}$$

This example confirms that we need only to evaluate the derivative at one point in each subspace to obtain the normal vectors. Sofar we know how to solve the classification problem when 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) $L_n(\mathbf{X})$ will be of full rank and thus

[6] This property of the derivative $\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}$ can be used to determine the dimension of the subspace. Unfortunately, this rank condition holds only when the data are noiseless. However, it is still possible to use this relation to determine the dimension of the subspace by thresholding the singular values. See Vidal et al. (2004) for more details.

equation system (3.20) has only zero solution. (2) It may happen that no point lies really in any subspace, such that we can not obtain an accurate inference on the normal vectors. In the following subsection we present a procedure of PDA with a voting scheme in order to overcome these problems.

3.3 Method of Generalized Principal Component Analysis with Noisy Data

3.3.1 PDA with Voting Scheme

Yang, Rao, Wagner, Ma, and Fossum (2005) propose the PDA with a voting scheme solve the problem with noisy data. Their motivation 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 $L_n(\mathbf{X})$ 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 $L_n(\mathbf{X})$ 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 following is an algorithm given in Yang et al. (2005):

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 $d_1, ..., d_n$:

```
Set $angleTolerance$, let $C$ be the number of distinct codimensions,
  and obtain $D$ by the Hilbert function constraint.
Let $V\{1\}, ..., V\{C\}$ be integer arrays as voting counters and $U\{1\}, ..., U\{C\}$
  be matrix arrays for basis candidates.
Construct $L_N = [\nu_n(\mathbf{x}^1), ..., \nu_n(\mathbf{x}^N)]$.
Form the set of polynomials $p_n(\mathbf{x})$ and compute $Dp_n(\mathbf{x})$.
for all sample $\mathbf{x}^k$ do
    for all $1 \leq i \leq C$ do
        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 $Dp_n(\mathbf{x})|_{\mathbf{x}^k}$.
        Compare $B$ with all candidates in $U\{i\}$.
        if $\exists j$, $subspaceangle[B, U\{i\}(j)] < angleTolerance$ then
            $V\{i\}(j) = V\{i\}(j) + 1$.
            Average the principal directions with the new basis $B$.
        else
            Add a new entry in $V\{i\}$ and $U\{i\}$.
        end if
    end for
end for
for all $1 \leq i \leq C$ do
    m = the number of subspaces in class i.
    Choose the first m highest votes in V{i} with their corresponding bases in U{i}.
    Assign corresponding samples into the subspaces, and cancel their votes
      in the other classes.
end for
Segment the remaining samples based on these bases.
```

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

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:

$$||\hat{\mathbf{e}}_i^j|| = \min\{||\hat{\mathbf{e}}_1^j||, ||\hat{\mathbf{e}}_2^j||, ..., ||\hat{\mathbf{e}}_n^j||\} \quad \rightarrow \quad \mathbf{x}^j \Rightarrow S_i, \tag{3.32}$$

where $\mathbf{x}^j \Rightarrow S_i$ means that $\mathbf{x}^j$ is classified to the subspace S_i. Equation (3.32) says, if the distance between a point $\mathbf{x}^j$ and the subspace i, i.e. $||\hat{\mathbf{e}}_i^j||$, is the smallest among all subspaces, $\mathbf{x}^j$ will be classified to the subspace S_i.

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

$$||\hat{\mathbf{e}}_i^{ji}|| = \min\{||\hat{\mathbf{e}}_1^j||, ||\hat{\mathbf{e}}_2^j||, ..., ||\hat{\mathbf{e}}_n^j||\} \tag{3.33}$$

holds, then $\mathbf{x}^{ji} \Rightarrow S_i$ follows. This classification is correct. Assumption 2.5 (a) and (b) implies that if there is no noise, with an exception of probability zero 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 result in some errors in the classification.

We show how to solve this problem in the next subsection.

3.5 Projected Model

In principle, we could obtain an estimate for each subspace by PDA as described in subsection 3.3.1. However, the usual case of a dynamic factor model is that the number of observations T is large and the number of 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 to extreme heavy computational burden. 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[7] (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.

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 matrix space spanned by G. A rescaled/rotated estimate can be calculated as follows:

$$\hat{G}^K = \frac{1}{NT}(XX')\sqrt{T}Q, \tag{3.34}$$

where (XX') is cross-product of the data matrix. We project the original models (2.6) and (2.4) and by premultiply $\frac{\hat{G}^K}{T}$ to both sides of the models and obtain:

$$\frac{1}{T}\hat{G}^{K'}X = \frac{1}{T}\hat{G}^{K'}G\Lambda + \frac{1}{T}\hat{G}^{K'}E \tag{3.35}$$

[7] See Vidaly et al. (2003) for more details.

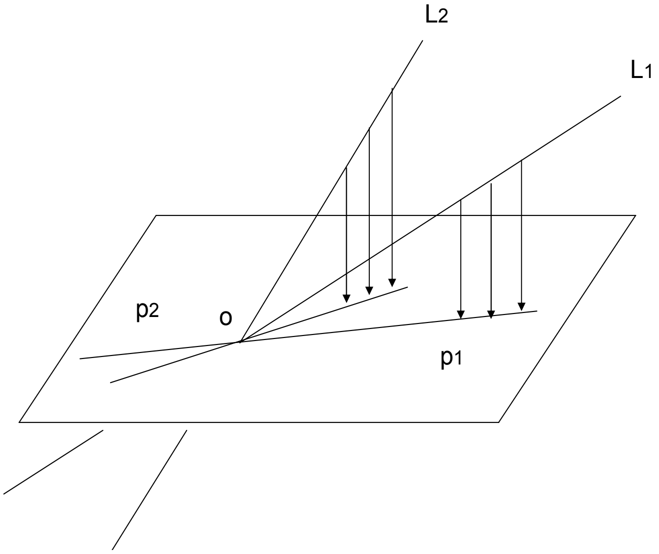

Figure 2: GPCA for $n = 2$, $k = 1$ and $T = 3$.

and

$$\frac{1}{T}\hat{G}^{K\prime}X_i = \frac{1}{T}\hat{G}^{K\prime}F_i\Lambda_i + \frac{1}{T}\hat{G}^{K\prime}E_i \qquad \text{for } i = 1, 2, ..., n. \tag{3.36}$$

Equation (3.36) defines again a grouped factor model with K observations. Denoting $\frac{1}{T}\hat{G}^{K\prime}X$, $\frac{1}{T}\hat{G}^{K\prime}G^o$, $\frac{1}{T}\hat{G}^{K\prime}E$, $\frac{1}{T}\hat{G}^{K\prime}X_i$, $\frac{1}{T}\hat{G}^{K\prime}F_i$ and $\frac{1}{T}\hat{G}^{K\prime}E_i$ 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

$$\bar{X}^T = \bar{G}^T\Lambda + \bar{E}^T \tag{3.37}$$

and

$$\bar{X}_i^T = \bar{F}_i^T\Lambda_i + \bar{E}_i^T \qquad \text{for } i = 1, 2, ..., n \tag{3.38}$$

or equivalently

$$\bar{X}_i^T = \tilde{\bar{X}}_i^T + \bar{E}_i^T \qquad \text{with} \qquad \bar{\mathbf{B}}_i^{T\prime}\tilde{\bar{X}}_i^T = 0 \qquad \text{for } i = 1, 2, ..., n \tag{3.39}$$

The projected models (3.37) and (3.38) has the following property.

Proposition 3.3

Under Assumption 2.1 to Assumption 2.7, 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* $N \to \infty$, $T \to \infty$
- *(c)* $\bar{E}_i^T \xrightarrow{P} 0$ *and* $\bar{E}^T \xrightarrow{P} 0$ *as* $N \to \infty$, $T \to \infty$
- *(d)* $P(\bar{F}_i = \bar{F}_j C) = 0$, *where* C *is some constant matrix.*
- *(e)* $P(\bar{F}_i \lambda_{i,m} = \bar{F}_j \lambda_{j,l}) = 0$ *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).

Proposition 3.3 (a) through (c) say that the projected model will converge to a grouped factor model without noises, i.e. all data points lie directly in the respective factor spaces. (d) and (e) say that the groups in the projected model are identified and the projection will not change the membership relation between variables and groups.

The 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 lead to a correct classification, as the idiosyncratic errors converge zero for $T \to \infty$, $N \to \infty$.

Since the classification rule defined in (3.32) 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

$$P(||\hat{\mathbf{e}}_i^{ji}|| = \min\{||\hat{\mathbf{e}}_1^j||, ||\hat{\mathbf{e}}_2^j||, ..., ||\hat{\mathbf{e}}_n^j||\}) \to 1 \qquad as \qquad T \to \infty, N \to \infty. \quad (3.40)$$

Proposition 3.5

The classification rule (3.32) based on the PDA with a voting scheme applied to the projected model (3.38) 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 $T \to \infty, N \to \infty$. As $\{\hat{\bar{\mathbf{B}}}_1, \hat{\bar{\mathbf{B}}}_2, ...\hat{\bar{\mathbf{B}}}_n\}$ 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:

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

where $\{\hat{\bar{\mathbf{B}}}_1, \hat{\bar{\mathbf{B}}}_2, ...\hat{\bar{\mathbf{B}}}_n\}$ is the estimate of subspaces using PDA based on the data $\{\bar{X}_i^T\}_{i=1}^n$ and $\{\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2, ...\bar{\mathbf{B}}_n\}$ is the subspaces calculated with PDA based on the data $\{\bar{X}_i\}_{i=1}^n$. Therefore, we have

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

where $\hat{\bar{\mathbf{e}}}_i^{ji}$ 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 (3.38) 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 $\bar{\mathbf{x}}^{ji}$ has a strick positive distance to other factor spaces converges to one.

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

From Proposition 3.3 (e) we have

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

The last probability convergence follows from Proposition 3.3 (e) and the probability equality is due to the fact that $||\bar{e}_k^{ji}|| = 0$ implies the point $\bar{\mathbf{x}}^{ji}$ lies in the intersection of the factor spaces of group i and group k, and hence there exist $\lambda_{k,l}$ and $\lambda_{i,m}$ such that $\bar{\mathbf{x}}^{ji} = \bar{F}_k\lambda_{k,l} = \bar{F}_i\lambda_{i,m}$.

It follows then

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

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

$$P(\bar{\mathbf{x}}^{T,ji} \Rightarrow \bar{S}_i) = P(||\hat{\bar{\mathbf{e}}}_i^{ji}|| = \min\{||\hat{\bar{\mathbf{e}}}_1^{j}||, ||\hat{\bar{\mathbf{e}}}_2^{j}||, ..., ||\hat{\bar{\mathbf{e}}}_n^{j}||\}) \to 1, \text{ as } T \to \infty, N \to \infty. \tag{3.41}$$

□

Remarks: Assumption 2.5 (b) leads to the results that $P(||\hat{\bar{e}}_k^{ji}|| = 0) \to 0$ and hence the proof of the consistent classification above. This assumption is not essential for conducting a correct inference of the group-specific factors. If $P(||\hat{\bar{e}}_k^{ji}|| = 0) > 0$, a significant proportion of data will lie in the intersection of two factor spaces. 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-specific factors. Allowing $P(||\hat{\bar{e}}_k^{ji}|| = 0) > 0$ will however complicate the definition of correct classification. In order to avoid this complication and simplify the presentation, we make the assumption Assumption 2.5 (b).

Since the group membership relation remains preserved after a projection from a T dimensional space onto a K dimensional space. The classification of variables obtained in the projected model (3.38) is a consistent classification of the variables in the original model.

$$P(\mathbf{x}^{ji} \Rightarrow S_i) = P(\bar{\mathbf{x}}^{T,ji} \Rightarrow \bar{S}_i) \xrightarrow{P} 1, \text{ as } \; T \to \infty, N \to \infty. \tag{3.42}$$

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

As shown in the previous subsection, an estimate of the subspaces by the PDA can be obtained when the number of subspaces and their dimensions are given, i.e. when we know the number of groups and the number of factor in each group, we can estimate the group-pervasive factors. However, in practical applications the number of the subspaces and their dimensions are often unknown. It raises naturally a question how we can decide the number of the subspaces and their dimensions. We consider again the following grouped factor model:

$$X_i = F_i \Lambda_i + E_i \qquad i = 1, 2, ...n. \tag{3.43}$$

Given the number of subspaces and their respective dimensions $(n, \{k_i\}_{i=1}^n)$, we can classify the variables into n groups, using the classification methods discussed in the previous subsection. For group i ($i = 1, 2, ...n$), we denote the T observations of the N_i variables which are classified into this group by X_i^s. Given the classified variables $(X_1^s, X_2^s, ..., X_n^s)$, we can estimate the group-pervasive factors group by group using principal component method: $\hat{F}_i = \sqrt{T}Q$, where Q contains the k_i eigenvectors corresponding to the largest k_i eigenvalues of the matrix $X_i^s X_i^{s\prime}$. The factor loading estimate is given accordingly

$$\hat{\Lambda}_i = \hat{F}_i' X_i^s / T.$$

Denote the mean squared residuals by of the ith group by:
$V_i(k_i, \hat{F}_i, N_i) = \frac{1}{N_i T} \sum_{j=1}^{N_i} \sum_{t=1}^{T} (X_{i,jt}^s - \hat{\lambda}_{i,j} \hat{F}_{i,t})^2$. The asymptotic principal component

estimate of factors $\hat{F}_i$ is the solution of the following minimization problem:

$$V_i(k_i, \hat{F}_i, N_i) = \min_{\Lambda_i, F_i} \frac{1}{N_i T} \sum_{j=1}^{N_i} \sum_{t=1}^{T} (X^s_{i,jt} - \lambda_{i,j} F_{i,t})^2, \tag{3.44}$$

where $\Lambda_i = (\lambda_{i,1}, \lambda_{i,2}, ..., \lambda_{i,N_i})$ and $F_i = (F_{i,1}, F_{i,2}, ..., F_{i,T})'$. If a correct classification were known, the information criterion developed in Bai and Ng (2002) could be used to determine the number of factors k_i group by group. However, for a grouped factor model as a whole, the situation is more complex. In fact we are dealing with an unknown number of different factor models simultaneously. In other words we have to determine the number of groups as well as the the number of factors for each group at same time. We denote these key parameters of a grouped factor model by $(n, \{k_i\}_{i=1}^n)$, where n is the number of groups in the model and k_i $(i = 1, 2, ..., n)$ is the number of group-pervasive factors of the ith group. A model selection criterion $C(n, \{k_i\}_{i=1}^n, \{X_i^s\})$ that is a scalar function of data, model parameters and the classification of the variables measures the goodness of fit of the model to the data.

Definition 3.6

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

$$P\{C(n^o, \{k_i^o\}_{i=1}^n, \{X_i^s\}) < C(n', \{k_i'\}_{i=1}^{n'}, \{X_i^u\})\} \to 1 \qquad \textit{for } T, N \to \infty. \tag{3.45}$$

Here n^o is the number of groups in the true model and $\{k_i^o\}_{i=1}^n$ are the numbers of group-pervasive factors of respective groups in the true model. $(n', \{k_i'\}_{i=1}^{n'})$ represents an alternative model.

Because we are considering the asymptotical property of a model selection criterion for the number of groups and number of factor in each group, 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. We denote this lower bound by $\underline{\alpha}$. The model selection criterion is formulated as follows.

Proposition 3.7

Under Assumption 2.1 to Assumption 2.7 of a grouped factor model (2.4),

$$PC(n, \{k_i\}_{i=1}^n, \{X_i^s\}) = \sum_{i=1}^n \frac{N_i}{N} V_i(k_i, \hat{F}^{k_i}, N_i) + \hat{\sigma}^2 \left(\sum_{i=1}^n \frac{N_i}{N} \left(k_i + h(N_i/N)\right) \right) g(N,T) \tag{3.46}$$

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* $0 \leq \alpha \leq 1$
   (b) $h(\alpha_i) \geq h(\alpha_j)$ *for any* $0 \leq \alpha_i \leq \alpha_j \leq 1$.
   (c) $\sum_l \alpha_l h(\alpha_l) > \sum_j \alpha_j h(\alpha_j)$ *for and* $\{\alpha_j\} \precsim \{\alpha_l\}$.
   We use the notation $\{\alpha_j\} \precsim \{\alpha_l\}$ *to present that* $\{\alpha_j\}$ *is a finer partition of the variables than* $\{\alpha_l\}$, *with* $\sum_l \alpha_l = \sum_j \alpha_j = 1$.

Proof (See Appendix).

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

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

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 mean of number of factors over all groups and $\bar{h}$ is the weighted mean 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 PC criteria developed in Bai and Ng (2002) 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, Condition 2 is to get the right rate of convergence for the penalty term, and Condition 3 is to make sure that the average number of factors is the dominating parameter of the model and the dispersion of groups is a dominated parameter. While comparing two models, we compare first the dominating parameter, only when the dominating parameter are equal we compare the dispersion of the groups in the two models.

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

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

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

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

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

3.7 Estimation Procedure for a Grouped Factor Model

- Step 1: Estimate K by the PC criterion of Bai and Ng (2002).

- Step 2: Project the $T \times N$ data matrix X onto a $K \times N$ matrix:

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

  where $\hat{G}^K$ is defined in (3.34).

- Step 3: According to a chosen model $(n, \{k_i\}_{i=1}^n)$, solve for the corresponding the subspaces $(\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2, ..., \bar{\mathbf{B}}_n)$ of the projected model (3.39) by polynomial differentiation algorithm with voting scheme and classify the variables according to rule (3.32).

- Step 4: Use the model selection criterion to evaluate alternative choices of models to obtain an optimal model and the corresponding classification of variables $\{X_i^s\}_{i=1}^n$.

- Step 5: Estimate a factor model for each group of data in $\{X_i^s\}_{i=1}^n$ by the standard principal component method to obtain estimates for the respective group-pervasive factors $\hat{F}_i = \sqrt{T}Q_i$ and factor loadings $\hat{\Lambda}_i = \hat{F}_i^{\prime} X_i^s/T$, where Q_i contains the k_i eigenvectors corresponding to the k_i largest eigenvalues of the matrix $X_i^s X_i^{s\prime}$.

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 factor space for each group.

Proof

As $T \to \infty$ and $N \to \infty$, according to Proposition 3.5 the classification of the variables are correct. Given a correct classification, for each group of variables the assumptions on a factor model in Bai and Ng (2002) are satisfied. Therefore, Theorem 1 in Bai and Ng (2002) can be applied to each group of data, i.e. the principal component method provides a consistent estimate of the factor space for each group.

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 a vector consisting of the number of factors in each group k_i $i = 1, 2, ..., n$ and the number of the factor in the ungrouped model K to represent a GFM. For example [321|5] represents a GFM with three groups, the ungrouped 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 the 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 pervasive 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 $F_{i,t} = (F_{i,1t}, F_{i,2t}, ..., F_{i,k_it})'$ for the ith group is a $k_i \times 1$ matrix 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 $k_i \times 1$ matrix of $N(0, 1)$ variables: and $e_{i,jt} \sim N(0, 1)$. In this setting the common component of $X_{i,jt}$ has variance 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 the cases in which the number of groups in a GFM varies from 2 to 4; the number of variables in each group varies from 30 to 60; 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 and 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 investigate the sharpness of the model selection criterion in identifying the true model from similar model candidates. For a true model [2 2|3], [3 1] and [2 2 2] are more general models. Because in our simulation design the true model [2 2|3] has one common factor, the total number of factors in the data set is three. We are considering here two factor planes in a three dimensional ambient space. Therefore, the model [3 1] is a more general model because it contains a three-dimensional subspace and a one-dimensional subspace, and [2 2 2] is also a more general model 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 ambient factor space.

The outcomes of the simulation study are summarized in Table 1 to Table 4. The first and the second columns in these tables give the numbers of observations and the numbers of variables in the respective simulation settings. The numbers in a pair of brackets in the second column are the numbers of variables in the respective groups. The third 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 ambient space which is the number behind the bar in the square bracket. For candidate models we do not give the dimensions of the ambient 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 ambient space and (2) select the correct model from the candidates, we report the performance with respect to choosing the correct projection dimensions and the performance with respect to choosing the correct models from the competing candidates after the projection.

It is to note that determining the projection dimension can be seen as a problem of comparing ungrouped models models with grouped models. $UGRP$ reports the performance of the model selection criterion in this respect. A number in the column of $UGRP$ is the proportion that the correct projection dimension, i.e. the dimension of the pervasive factor space, is chosen by the PC criterion of Bai and Ng (2002) in the pooled data and at least one grouped factor model is chosen over the correct ungrouped factor model in the respective 1000 simulation runs. Since our data generating models are all grouped factor models, for good performance of the selection criterion we expect the numbers 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 for a correct determination of the projection dimension, i.e. the proportions of finding the right projection dimension are very high: all numbers in this column are one. This result is consistent with the simulation result given in Bai and Ng (2002).

The column under the header $CCLM$ reports the proportion of correctly identified models among the candidates in 1000 simulation replications under the condition that the projection dimension is chosen correctly. Most of the numbers in the column of $CCLM$ are very close to one, indicating that for the considered configurations the estimation procedure performs well in 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 still low (see the first and the third panels in Table 2, the third panel in Table 3, and the second and the third panels in Table 4.) 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. Indeed the numbers in the column of $MCLV$ are all under 10 percent, indicating that the classification of variables works very well. It is to note that as far as the subspaces are intersected, there is always some proportion of misclassification of variables, though this proportion is low.

$SFF0$[8] reports the average goodness of fit of the estimated factors for 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 most of the variables are correctly classifies into their groups, the goodness of fit of the estimated factors to the true factors is comparable to the goodness of fit in ungrouped factor models. Indeed in most cases the numbers in the column of $SFF0$ are over 90%.

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

Table 1: Estimation of grouped factor models

T	N	Model and Candidates	CCLM	SFF0	MCLV	UGRP
		[11|2]				
80	(30 30)	[111] [1 1]	1.00	0.96	0.03	1.00
150	(30 30)	[111] [1 1]	1.00	0.97	0.02	1.00
300	(30 30)	[111] [1 1]	1.00	0.96	0.02	1.00
500	(30 30)	[111] [1 1]	1.00	0.97	0.02	1.00
80	(60 60)	[111] [1 1]	1.00	0.98	0.02	1.00
150	(60 60)	[111] [1 1]	1.00	0.98	0.02	1.00
300	(60 60)	[111] [1 1]	1.00	0.98	0.02	1.00
500	(60 60)	[111] [1 1]	1.00	0.98	0.01	1.00
		[21|3]				
80	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	0.99	0.95	0.05	1.00
150	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	0.99	0.95	0.05	1.00
300	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	0.99	0.95	0.02	1.00
500	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.95	0.02	1.00
80	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.97	0.05	1.00
150	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.98	0.05	1.00
300	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.98	0.02	1.00
500	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 1]	1.00	0.98	0.02	1.00
		[22|3]				
80	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	0.99	0.93	0.03	1.00
150	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	1.00	0.93	0.03	1.00
300	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	1.00	0.93	0.02	1.00
500	(30 30)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	1.00	0.93	0.02	1.00
80	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	1.00	0.97	0.02	1.00
150	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	1.00	0.97	0.02	1.00
300	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	1.00	0.97	0.02	1.00
500	(60 60)	[2 2] [2 1] [1 1] [1 1 1] [2 2 2]	1.00	0.97	0.01	1.00
		[32|4]				
80	(30 30)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	0.98	0.91	0.08	1.00
150	(30 30)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.92	0.06	1.00
300	(30 30)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.92	0.04	1.00
500	(30 30)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.91	0.03	1.00
80	(60 60)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.96	0.07	1.00
150	(60 60)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.96	0.05	1.00
300	(60 60)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.96	0.04	1.00
500	(60 60)	[3 2] [3 1] [2 1] [3 3] [3 2 1]	1.00	0.96	0.03	1.00

Notes: Table 1 reports the results of estimation a GFM in 1000 Monte Carlo simulations. The first column gives number of observations. The second column gives the numbers of variables in the respective data-generating grouped factor model. The third columns gives the data-generating grouped factor models and the candidate models, over which the model selection procedure was applied. $CCLM$ gives the proportion of the correctly identified true models over 1000 runs. $SFF0$ 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 in all variables over 1000 runs. $UGRP$ gives the proportion of correctly identified projection spaces in 1000 runs.

Table 2: Estimation of grouped factor models

T	N	Model and Candidates	CCLM	SFF0	MCLV	UGRP
		[33|5]				
80	(30 30)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	0.77	0.89	0.02	1.00
150	(30 30)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	0.97	0.90	0.01	1.00
300	(30 30)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	0.98	0.90	0.01	1.00
500	(30 30)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	1.00	0.90	0.00	1.00
80	(60 60)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	0.99	0.95	0.01	1.00
150	(60 60)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	1.00	0.95	0.01	1.00
300	(60 60)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	1.00	0.95	0.01	1.00
500	(60 60)	[1 1 1] [2 2] [3 2 1] [3 3 1] [3 3 2] [3 3 3] [3 3]	1.00	0.95	0.00	1.00
		[31|4]				
80	(30 30)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	0.95	0.93	0.07	1.00
150	(30 30)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	0.99	0.93	0.06	1.00
300	(30 30)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	1.00	0.93	0.04	1.00
500	(30 30)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	1.00	0.93	0.03	1.00
80	(60 60)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	1.00	0.97	0.07	1.00
150	(60 60)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	1.00	0.97	0.05	1.00
300	(60 60)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	1.00	0.97	0.04	1.00
500	(60 60)	[3 1] [2 1] [2 2] [3 2] [3 2 1] [3 1 1]	1.00	0.97	0.03	1.00
		[311|5]				
80	(30 30 30)	[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	0.68	0.94	0.09	1.00
150	(30 30 30)	[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	0.84	0.94	0.06	1.00
300	(30 30 30)	[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	0.95	0.94	0.04	1.00
500	(60 60 60)	[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	0.99	0.94	0.03	1.00
80	(60 60 60)	[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	0.99	0.97	0.09	1.00
150	(60 60 60)	[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	1.00	0.97	0.06	1.00
300	(60 60 60)	[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	1.00	0.97	0.04	1.00
500		[3 2] [3 1 1 1] [3 1] [3 2 1] [3 2 2] [3 1 1]	1.00	0.97	0.04	1.00
		[111|3]				
80	(30 30 30)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.96	0.03	1.00
150	(30 30 30)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.97	0.03	1.00
300	(30 30 30)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.97	0.02	1.00
500	(30 30 30)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.97	0.02	1.00
80	(60 60 60)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.02	1.00
150	(60 60 60)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.03	1.00
300	(60 60 60)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.03	1.00
500	(60 60 60)	[1 1 1] [2 1] [1 1] [2 2] [2 2 1]	1.00	0.98	0.01	1.00

Notes: Table 2 reports the results of estimation a GFM in 1000 Monte Carlo simulations. The first column gives number of observations. The second column gives the numbers of variables in the respective data-generating grouped factor model. The third columns gives the data-generating grouped factor models and the candidate models, over which the model selection procedure was applied. $CCLM$ gives the proportion of the correctly identified true models over 1000 runs. $SFF0$ 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 in all variables over 1000 runs. $UGRP$ gives the proportion of correctly identified projection spaces in 1000 runs.

Table 3: Estimation of grouped factor models

T	N	Model and Candidates	CCLM	SFF0	MCLV	UGRP
		[211|4]				
80	(30 30 30)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	0.91	0.93	0.04	1.00
150	(30 30 30)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	0.99	0.93	0.03	1.00
300	(30 30 30)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.93	0.03	1.00
500	(60 60 60)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.93	0.02	1.00
80	(60 60 60)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.97	0.03	1.00
150	(60 60 60)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.97	0.02	1.00
300	(60 60 60)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.97	0.02	1.00
500	(60 60 60)	[2 1 1] [2 1] [2 2] [3 1] [2 2 1] [2 2 2] [3 1 1]	1.00	0.97	0.02	1.00
		[222|4]				
80	(30 30 30)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	0.91	0.93	0.04	1.00
150	(30 30 30)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	0.99	0.93	0.03	1.00
300	(30 30 30)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	1.00	0.93	0.03	1.00
500	(30 30 30)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	1.00	0.93	0.02	1.00
80	(60 60 60)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	1.00	0.97	0.03	1.00
150	(60 60 60)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	1.00	0.97	0.02	1.00
300	(60 60 60)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	1.00	0.97	0.02	1.00
500	(60 60 60)	[2 2 2] [3 2] [3 2 1] [3 2 2] [3 1 1] [2 2 2 2]	1.00	0.97	0.02	1.00
		[322|5]				
80	(30 30 30)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.61	0.88	0.10	1.00
150	(30 30 30)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.94	0.92	0.07	1.00
300	(30 30 30)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.96	0.92	0.05	1.00
500	(30 30 30)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.99	0.92	0.04	1.00
80	(60 60 60)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.93	0.96	0.08	1.00
150	(60 60 60)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	0.99	0.96	0.06	1.00
300	(60 60 60)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.96	0.04	1.00
500	(60 60 60)	[3 2 2] [4 3] [4 2] [3 3 2] [3 3 1] [3 1 1] [4 2 2]	1.00	0.96	0.04	1.00
		[2222|5]				
80	(30 30 30 30)	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	0.80	0.92	0.04	1.00
150	(30 30 30 30)	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	0.96	0.93	0.04	1.00
300	(30 30 30 30)	[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	0.03	1.00
500	(30 30 30 30)	[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	0.03	1.00
80	(60 60 60 60)	[2 2 2 2] [3 3] [4 2] [3 2 2 2] [2 2 2 1] [2 2 2 2 1]	0.93	0.96	0.03	1.00
150	(60 60 60 60)	[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	0.03	1.00
300	(60 60 60 60)	[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	0.02	1.00
500	(60 60 60 60)	[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	0.02	1.00

Notes: Table 3 reports the results of estimation a GFM in 1000 Monte Carlo simulations. The first column gives number of observations. The second column gives the numbers of variables in the respective data-generating grouped factor model. The third columns gives the data-generating grouped factor models and the candidate models, over which the model selection procedure was applied. $CCLM$ gives the proportion of the correctly identified true models over 1000 runs. $SFF0$ 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 in all variables over 1000 runs. $UGRP$ gives the proportion of correctly identified projection spaces in 1000 runs.

Table 4: Estimation of grouped factor models

T	N	Model and Candidates	CCLM	SFF0	MCLV	UGRP
		[2211|5]				
80	(30 30 30 30)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.86	0.94	0.06	1.00
150	(30 30 30 30)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.96	0.95	0.04	1.00
300	(30 30 30 30)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.99	0.95	0.03	1.00
500	(30 30 30 30)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.99	0.95	0.02	1.00
80	(60 60 60 60)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	0.97	0.97	0.04	1.00
150	(60 60 60 60)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.97	0.03	1.00
300	(60 60 60 60)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.97	0.03	1.00
500	(60 60 60 60)	[2 2 1 1] [3 1 1] [3 2] [2 2 1] [2 1 1 1]	1.00	0.97	0.02	1.00
		[3211|6]				
80	(30 30 30 30)	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.61	0.93	0.09	1.00
150	(30 30 30 30)	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.83	0.94	0.06	1.00
300	(30 30 30 30)	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.92	0.94	0.05	1.00
500	(30 30 30 30)	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.99	0.94	0.04	1.00
80	(60 60 60 60)	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.96	0.97	0.08	1.00
150	(60 60 60 60)	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	0.97	0.97	0.06	1.00
300	(60 60 60 60)	[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
500	(60 60 60 60)	[3 2 1 1] [4 2 2] [4 1 1] [4 3 1 1] [2 2 1 1]	1.00	0.97	0.03	1.00
		[3221|6]				
80	(30 30 30 30)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.67	0.92	0.09	1.00
150	(30 30 30 30)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.79	0.93	0.06	1.00
300	(30 30 30 30)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.91	0.93	0.05	1.00
500	(30 30 30 30)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.96	0.93	0.04	1.00
80	(60 60 60 60)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.87	0.96	0.08	1.00
150	(60 60 60 60)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.96	0.97	0.06	1.00
300	(60 60 60 60)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	0.99	0.97	0.05	1.00
500	(60 60 60 60)	[3 2 2 1] [4 3] [4 2 1] [4 1 1] [3 2 2 2] [3 3 1]	1.00	0.97	0.03	1.00

Notes: Table 4 reports the results of estimation a GFM in 1000 Monte Carlo simulations. The first column gives number of observations. The second column gives the numbers of variables in the respective data-generating grouped factor model. The third columns gives the data-generating grouped factor models and the candidate models, over which the model selection procedure was applied. $CCLM$ gives the proportion of the correctly identified true models over 1000 runs. $SFF0$ 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 in all variables over 1000 runs. $UGRP$ gives the proportion of correctly identified projection spaces in 1000 runs.

4.2 An Empirical Application

In this subsection we apply the GFM to stock returns in the Australian Stock Exchange. The data used in this exercise are stock returns of companies included in ASX200. ASX200 is one of the most important share index in Australia Stock Exchange. It accounts for roughly 85% of the market capitalization of all stocks listed in Australia Stock Exchange. The data set consists of monthly returns of shares included in ASX200 from 2004 to 2009. All together there are 168 variables and each of them contains 77 observations[9]. 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 whole data set. 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 spaces. We decide not to include subspace configurations with more that 4 groups because in those cases it is highly probable that some group will contain less than 30 variables such that the model selection criterion would become unreliable. The estimation results for the considered configurations are summarizes in Table 5.

Table 5: Estimation of Grouped Dynamic Factor Models for ASX200

No.	Model	PC	No.	Model	PC
1	[1]	0.00571	19	[1 1 1]	0.00518
2	[2]	0.00527	20	[2 1 1]	0.00537
3	[3]	0.00524	21	[2 2 1]	0.00508
4	[4]	0.00526	22	[2 2 2]	0.00508
5	[5]	0.00530	23	[1 1 1 1]	0.00527
6	[6]	0.00536	24	[2 1 1 1]	0.00524
7	[1 1]	0.00526	25	[2 2 1 1]	0.00508
8	[2 1]	0.00511	26	[2 2 2 1]	0.00522
9	[2 2]	0.00508	27	[2 2 2 2]	0.00508

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.

In Table 5 we see that 5 model candidates [2 2] [2 2 1] [2 2 2] [2 2 1 1] and [2 2 2

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

2] have the same lowest criterion values. However, three models: [2 2 1] [2 2 1 1] and [2 2 2 2] all contain groups with less that 20 variables, such that the criterion values in these cases are not very reliable. Hence we will not consider these three model as proper models for the data. The two models [2 2] and [2 2 2] are competing. Since [2 2] has a simpler structure than [2 2 2], we take [2 2] is the most suitable grouped factor model for the data. This implies that we understand that the 168 shares consist of 2 groups each of which are driven by two factors (See Fig. 3).

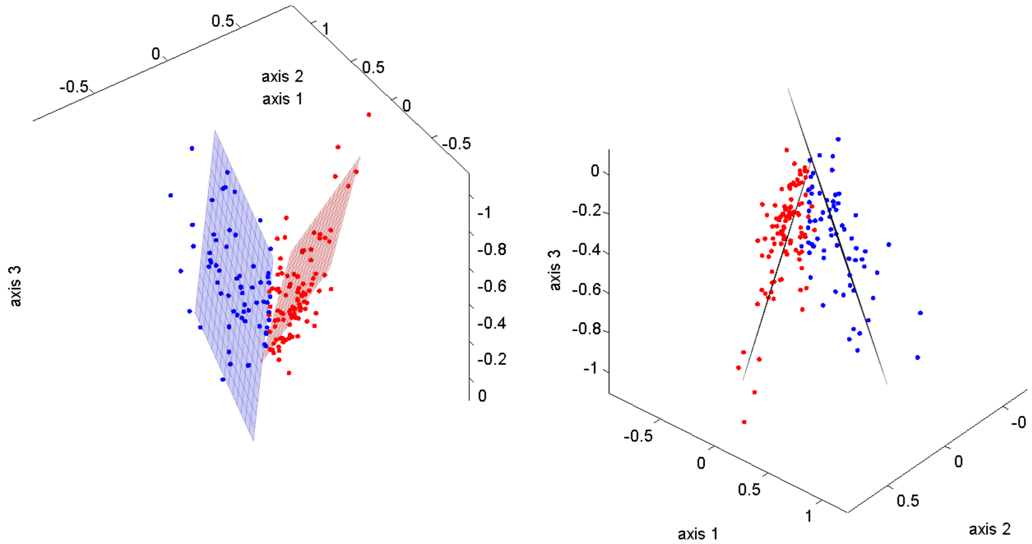

Figure 3: ASX200 shares in two groups in the projected model

The grouping of the 168 variables are given in Table 8. It is to note the GFM classifies almost all companies in resource sectors including mining, energy, and exploration into the second groups. Among the 54 companies in the second group there are only five companies (See (*) in Table 8, 9, 10.) that are not in the ming and energy sectors. The first group contains 114 companies among which only four companies (See (*) in Table 8, 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-specific 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

In this paper we present grouped factor models to investigate the structure in the factor space of a factor model. We propose a procedure to estimate the grouped factor models. The main feature of the procedure is that it can identify grouped structures, classify variables into groups and estimate group-pervasive factors based on the observed data. More concretely, if data are generated from a factor model without any grouped structure, the procedure will return a conventional ungrouped factor model with correctly identified number of factors. If data are generated from a grouped factor model, the procedure will output the number of groups and the number of factors in each group, a classification of the variables into the groups and group-pervasive factors. In this sense, our model generalizes the framework of the conventional factor models, such that it can be used to assess grouped structure in the data and estimate the group-pervasive factors, which may be useful for understanding the behavior of the data.

We set up the grouped factor models as approximate factor models which allow certain serial and cross-sectional correlation in the idiosyncratic errors. Therefore they are suitable for applications to economic data. Simulation study shows that our procedure has good finite sample properties. In an application example we shows 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.

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.*

$$X' = \begin{pmatrix} 1.0725 & 0.0607 & 0.0943 \\ 0.0603 & 1.0801 & 0.0460 \\ 1.0245 & 1.0977 & 0.0694 \\ 2.0909 & 2.0205 & 0.0854 \\ 0.0493 & 0.0667 & 1.0687 \\ 0.0653 & 0.0385 & 2.0011 \\ 0.0575 & 0.0383 & 3.0351 \\ 0.0857 & 0.0213 & 4.0375 \end{pmatrix} \tag{6.47}$$

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:

$$L_n(\mathbf{X}) = \begin{pmatrix} 1.1588 & 0.0042 & 0.0399 & 0.0000 & 0.0001 & 0.0014 \\ 0.0025 & 0.0522 & 0.0035 & 1.0816 & 0.0716 & 0.0047 \\ 1.0142 & 1.1073 & 0.0196 & 1.2090 & 0.0214 & 0.0004 \\ 4.0604 & 4.1306 & 0.1878 & 4.2020 & 0.1911 & 0.0087 \\ 0.0056 & 0.0017 & 0.0790 & 0.0005 & 0.0235 & 1.1091 \\ 0.0012 & 0.0022 & 0.0702 & 0.0043 & 0.1346 & 4.2418 \\ 0.0097 & 0.0083 & 0.3004 & 0.0072 & 0.2581 & 9.3041 \\ 0.0092 & 0.0076 & 0.3866 & 0.0063 & 0.3210 & 16.2398 \end{pmatrix} \tag{6.48}$$

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[10]*. We choose the two eigenvectors corresponding to the two smallest singular values as the basis of the nullspace of $L_n(\mathbf{X})$.*

$$\mathbf{c} = \begin{pmatrix} 0.0412 & 0.0782 \\ -0.0286 & -0.0477 \\ -0.4290 & -0.8970 \\ 0.0446 & -0.0123 \\ -0.9007 & 0.4320 \\ 0.0161 & 0.0157 \end{pmatrix}. \tag{6.49}$$

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:*

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_1} = \frac{\partial \nu_n(\mathbf{x})'}{\partial x_1}\mathbf{c} = (2x_1, x_2, x_3, 0, 0, 0)\mathbf{c}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_2} = \frac{\partial \nu_n(\mathbf{x})'}{\partial x_2}\mathbf{c} = (0, x_1, 0, 2x_2, x_3, 0)\mathbf{c}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_3} = \frac{\partial \nu_n(\mathbf{x})'}{\partial x_3}\mathbf{c} = (0, 0, x_1, 0, x_2, 2x_3)\mathbf{c}$$

Evaluating the three components of the partial derivative at the eight sample points is to replace (x_1, x_2, x_3) in the three formulas above by the corresponding numbers in the data matrix X. We obtain the following three matrices:

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

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_1}|_X = \begin{pmatrix} 0.0461 & 0.0802 \\ -0.0457 & -0.0833 \\ 0.0232 & 0.0456 \\ 0.0777 & 0.1540 \\ -0.4563 & -0.9541 \\ -0.8542 & -1.7866 \\ -1.2985 & -2.7153 \\ -1.7257 & -3.6092 \end{pmatrix}, \quad \frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_2}|_X = \begin{pmatrix} -0.1102 & -0.0119 \\ 0.0532 & -0.0096 \\ 0.0061 & -0.0459 \\ 0.0434 & -0.1125 \\ -0.9580 & 0.4577 \\ -1.8007 & 0.8605 \\ -2.7318 & 1.3075 \\ -3.6370 & 1.7397 \end{pmatrix}, \tag{6.50}$$

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial x_3}|_X = \begin{pmatrix} -0.5117 & -0.9328 \\ -0.9972 & 0.4140 \\ -1.4259 & -0.4425 \\ -2.7140 & -0.9999 \\ -0.0468 & 0.0182 \\ 0.0018 & 0.0209 \\ 0.0387 & 0.0603 \\ 0.0742 & 0.0592 \end{pmatrix}. \tag{6.51}$$

The three components of the partial derivative $\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}$ evaluated at each sample point are given in the corresponding row respectively in the three matrices above. So, the partial derivative evaluated at $\mathbf{x}^1$ is:

$$\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1} = \begin{pmatrix} 0.0461 & 0.0802 \\ -0.1102 & -0.0096 \\ -0.5117 & -0.9328 \end{pmatrix}. \tag{6.52}$$

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 (6.52) we have the following principal components:*

$$\frac{\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1}}{||\frac{\partial \nu_n(\mathbf{x})'\mathbf{c}}{\partial \mathbf{x}}|_{\mathbf{x}^1}||} = \begin{pmatrix} -0.02 & -0.09 \\ 0.99 & 0.06 \\ -0.06 & 0.99 \end{pmatrix}. \tag{6.53}$$

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 6 and Table 7 in a simplified form.*

We consider first the choice of normal vectors for the subspace of dimension one. Table 6 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 6. 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 6: 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 7.

Table 7 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.*

From the voting procedure in Table 6 and Table 7, the estimates of the two subspaces are:

$$\hat{B}_1 = \begin{pmatrix} 0.9993 & -0.0131 \\ -0.0132 & -0.9992 \\ -0.0135 & -0.0095 \end{pmatrix} \quad \textit{and} \quad \hat{B}_2 = \begin{pmatrix} -0.0361 \\ 0.0039 \\ 0.998 \end{pmatrix}. \tag{6.54}$$

Compared with equations (3.29) and (3.31), these two estimates of the normal vectors are very close to the true normal vectors.

6.2 Proofs

Proof of **Proposition 2.4**

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.

$$\underset{(k\times k)}{\frac{\Lambda\Lambda'}{N}} = \sum_{i=1}^{n} \frac{N_i}{N} \underset{(k\times k_i)}{C_i} \underset{(k_i\times k_i)}{\frac{\Lambda_i\Lambda_i'}{N_i}} \underset{(k_i\times k)}{C_i}{}' \tag{6.55}$$

Let $\mathbf{b}$ be a $k \times 1$ nonzero vector. To show that $\frac{\Lambda\Lambda'}{N}$ converges to a positive definite

Table 7: 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.

matrix we need to show $\mathbf{b}'\frac{\Lambda\Lambda'}{N}\mathbf{b} > 0$ when N is large enough.

$$\underset{(1\times k)}{\mathbf{b}'}\underset{(k\times k)}{\frac{\Lambda\Lambda'}{N}}\underset{(k\times 1)}{\mathbf{b}} = \sum_{i=1}^{n}\frac{N_i}{N}\underset{(1\times k_i)}{\mathbf{b}'C_i}\underset{(k_i\times k_i)}{\frac{\Lambda_i\Lambda_i'}{N_i}}\underset{(k_i\times 1)}{C_i'\mathbf{b}} \tag{6.56}$$

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 (6.56) 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 $\hat{F}_t$ 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.

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

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

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 (i, j) element of the first term, we have by Cauchy-Schwarz inequality:

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

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

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

From Assumption 2.6, we have:

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

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:

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

To show $\underset{T,N\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T} H^{K\prime}G_t^o E_t = 0$, we need only to show $\underset{T,N\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T} G_t^o E_t = 0$. 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 $\underset{T,N\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T} G_t^o e_{it} = 0$. In matrix form we have

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

This proves (c) in Proposition 3.3.

To prove (b) we have

$$\begin{aligned}
\bar{F}_i^T = \frac{1}{T}\hat{G}^{K'}F_i = \left(\frac{1}{T}\hat{G}^{K'}G^o\right)C_i &= \left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K'}G_t^o + H^{K'}G_t^o)G_t^{o'}\right)C_i \\
= \quad & \left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K'}G_t^o)G_t^{o'}\right)C_i + \left(\frac{1}{T}\sum_{t=1}^{T}(H^{K'}G_t^o)G_t^{o'}\right)C_i \\
= \quad & \left(\frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t^K - H^{K'}G_t^o)G_t^{o'}\right)C_i + H^{K'}\left(\frac{1}{T}\sum_{t=1}^{T}G_t^oG_t^{o'}\right)C_i \\
\xrightarrow{P} \quad & 0 + H^{K'}\Sigma C_i \neq 0.
\end{aligned}$$

The limit in the last row above is because of

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

and

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

From the existence of the limit of (b) and (c) follows the existence of the limit of (a).

From Assumption 2.5 (a), we have $C_i - C_jC \neq 0$. Since $\hat{G}^K$ is a principal component estimate of the overall factor. From $\frac{\hat{G}^{K'}G^o}{T} \xrightarrow{P} \bar{G}$ will not converge to zero. It follows

$$P\left(\bar{G}(C_i - C_jC) = 0\right) = 0.$$

Reformulate the equation above we have

$$P\left(\bar{G}(C_i - C_jC) = 0\right) = P\left(\bar{G}C_i = \bar{G}C_jC\right) = P\left(\bar{F}_i = \bar{F}_jC\right) = 0.$$

This proves (d).

From Assumption 2.5 (b) we have $C_i\lambda_{i,m} \neq C_j\lambda_{j,l}$. It follows

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

□

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

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

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 finer new partition of the variables. In each group of the intersection partition, all variables belong to only one group in the true model and they belong to also only 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 never underestimates the number of factors in its groups, and $k_i' = k_i^o$ for all groups in the intersection partition.

- **Case 3:** The alternative model never underestimates the number of factors in its groups and :$k'_i \geq k^o_i$ for all i and $k'_i > k^o_i$ 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.

- $V(k^o_i, \hat{F}^{k^o_i}, N^I_i)$: the mean squared residuals calculated from the true model with the a priori true classification .

- $V(k^o_i, F^{k^o_i}, N^I_i)$: the mean squared residuals calculated using k^o_i population factors.

- $V(k^o_l, F^{k^o_l}, N^I_i)$: the mean squared residuals calculated using population factors in the lth group of the alternative model.

- $V(k^o_i, \hat{F}^{k^o_i}_{N^I_i}, N^I_i)$: 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^o_i.

- $V(k'_i, \hat{F}^{k'_l}_{N^I_i}, N^I_i)$: 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 $\{N_j\}^n_{j=1}$, $\{N_l\}^{n'}_{l=1}$ and $\{N^I_i\}^{n^I}_{i=1}$ denote the a priori true classification, an alternative classification and the intersection partition, respectively.*

$$\sum_{j=1}^{n} \frac{N_j}{N} V(k^o_j, \hat{F}^{k^o_i}, N_j) = \sum_{i=0}^{n^I} \frac{N^I_i}{N} V(k^o_i, \hat{F}^{k^o_i}, N^I_i)$$

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

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

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

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 intersection group belong to only one group of the true model, we denote this group by j. Let k_j^o be the number of true factors in group j. We have $k_j^o = k_i^o$. Following equation (10) in Lemma 4 of Bai and Ng (2002) we have

$$V(k_i^o, \hat{F}^{k_i^o}, N_j) - V(k_i^o, F^{k_i^o}, N_j) = O_p(C_{N,T}^{-2}), \tag{6.57}$$

where N_j is the number of variables in jth group of the true model. The difference on the left hand side of the equation above can be written as follows:

$$\begin{aligned}
& V(k_i^o, \hat{F}^{k_i^o}, N_j) - V(k_i^o, F^{k_i^o}, N_j) \\
= & \frac{1}{N_j T}\left(\frac{N_i^I}{N_i^I}\sum_{i=1}^{N_i^I}\sum_{t=1}^{T}(X_{it} - \lambda_i \hat{F}_t^{k_i^o})^2 + \frac{N_j - N_i^I}{N_j - N_i^I}\sum_{i=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{it} - \lambda_i \hat{F}_t^{k_i^o})^2\right) \\
& -\frac{1}{N_j T}\left(\frac{N_i^I}{N_i^I}\sum_{i=1}^{N_i^I}\sum_{t=1}^{T}(X_{it} - \lambda_i F_t^{k_i^o})^2 + \frac{N_j - N_i^I}{N_j - N_i^I}\sum_{i=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{it} - \lambda_i F_t^{k_i^o})^2\right) \\
= & \frac{N_i^I}{N_j}\left(\frac{1}{N_i^I T}\sum_{i=1}^{N_i^I}\sum_{t=1}^{T}(X_{it} - \lambda_i \hat{F}_t^{k_i^o})^2 - \frac{1}{N_i^I T}\sum_{i=1}^{N_i^I}\sum_{t=1}^{T}(X_{it} - \lambda_i F_t^{k_i^o})^2\right) \\
& \frac{N_j - N_i^I}{N_j}\left(\frac{1}{(N_j - N_i^I)T}\sum_{i=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{it} - \lambda_i \hat{F}_t^{k_i^o})^2 - \frac{1}{(N_j - N_i^I)T}\sum_{i=N_i^I+1}^{N_j}\sum_{t=1}^{T}(X_{it} - \lambda_i F_t^{k_i^o})^2\right) \\
= & \underbrace{\frac{N_i^I}{N_j}\left(V(k_i^o, \hat{F}^{k_i^o}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I)\right)}_{\text{A}} \\
& + \underbrace{\frac{N_j - N_i^I}{N_j}\left(V(k_i^o, \hat{F}^{k_i^o}, N_j - N_i^I) - V(k_i^o, F^{k_i^o}, N_j - N_i^I)\right)}_{\text{B}} \\
\leq & \ 0.
\end{aligned}$$

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

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

and

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

These two inequalities are because the estimated factors based on the data in the intersection group N_i^I is the solution of minimization of the mean squared residuals of the group. Applying relation (6.57) to the data in the intersection partition N_i^I and $N_j - N_i^I$ 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 $A + B = O_p(C_{N,T}^{-2})$. 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* $k_l' \geq k_l^o$*,*

$$V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) = O_p(C_{N,T}^{-2}). \tag{6.58}$$

Proof

Since the variables in the ith group of the intersection partition belong to only one group of the true model and they belong to also only one group of the alternative model, we denote these two groups by j and l respectively. 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. Since all variables in the ith group of the intersection partition belong to group l of the alternative model, we have $k_i' = k_l'$. Then, 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 (6.58) into four differences:

$$
\begin{aligned}
& V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I) \qquad (6.59) \\
= \;& V(k_i', \hat{F}^{k_i'}, N_i^I) - V(k_l^o, F^{k_l^o}, N_i^I) \\
& + V(k_l^o, F^{k_l^o}, N_i^I) - V(k_l^o, \hat{F}^{k_l^o}_{N_i^I}, N_i^I) \\
& + V(k_l^o, \hat{F}^{k_l^o}_{N_i^I}, N_i^I) - V(k_i^o, F^{k_i^o}, N_i^I) \\
& + V(k_i^o, F^{k_i^o}, N_i^I) - V(k_i^o, \hat{F}^{k_i^o}, N_i^I)
\end{aligned}
$$

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

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

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

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

For the third difference we have $k_l^o \geq k_i^o$ where k_i^o is the true number of factors in the ith group of the intersection 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 different 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 $V(k_i', \hat{F}_{N_i^I}^{k_i'}, N_i^I)$ is the solution of the minimization problem (3.44) using the data in group i. 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 intersection 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,

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

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. PC^o means that the calculation of the model selection criterion value is based on the a priori true classification but not on PDA, while PC means generally that the model selection criterion value is calculated based on the 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 intersection 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 intersection partition. We calculate again the difference in the penalty factors.

$$\begin{aligned}
\phi = & \sum_{l=1}^{n'} \frac{N'_l}{N} k'_l + \sum_{l=1}^{n'} \frac{N'_l}{N} h(\hat{\alpha}'_l) - \sum_{j=1}^{n} \frac{N_j}{N} k^o_j - \sum_{j=1}^{n} \frac{N_j}{N} h(\hat{\alpha}^o_j) \\
= & \sum_{i=1}^{n^I} \frac{N^I_i}{N} k'_i + \sum_{i=1}^{n^I} \frac{N^I_i}{N} h(\hat{\alpha}'_i) - \sum_{i=1}^{n^I} \frac{N^I_i}{N} k^o_i - \sum_{i=1}^{n^I} \frac{N^I_i}{N} h(\hat{\alpha}^o_i) \\
= & \sum_{i=1}^{n^I} \frac{N^I_i}{N} (k'_i - k^0_i) + \sum_{i=1}^{n^I} \frac{N^I_i}{N} (h(\hat{\alpha}'_i) - h(\hat{\alpha}^o_i)) \\
= & \sum_{k'_i > k^o_i} \frac{N^I_i}{N} (k'_i - k^0_i) + \sum_{k'_i > k^o_i} \frac{N^I_i}{N} (h(\hat{\alpha}'_i) - h(\hat{\alpha}^o_i)) + \sum_{k'_i = k^o_i} \frac{N^I_i}{N} (h(\hat{\alpha}'_i) - h(\hat{\alpha}^o_i)) \\
\geq & \sum_{k'_i > k^o_i} \frac{N^I_i}{N} + \sum_{k'_i > k^o_i} \frac{N^I_i}{N} h(\hat{\alpha}'_i) - \sum_{k'_i > k^o_i} \frac{N^I_i}{N} h(\hat{\alpha}^o_i) + \sum_{k'_i = k^o_i} \frac{N^I_i}{N} (h(\hat{\alpha}'_i) - h(\hat{\alpha}^o_i)) \\
= & \sum_{k'_i > k^o_i} \frac{N^I_i}{N} (1 - h(\hat{\alpha}^o_i)) + \sum_{k'_i > k^o_i} \frac{N^I_i}{N} h(\hat{\alpha}'_i) + \sum_{k'_i = k^o_i} \frac{N^I_i}{N} (h(\hat{\alpha}'_i) - h(\hat{\alpha}^o_i)) \\
> & \; 0
\end{aligned}$$

The first two terms are positive because of condition 3(a) for h function. For the case of $k'_i = k^o_i$ we must have $\hat{\alpha}'_i < \hat{\alpha}^o_i$, because $\hat{\alpha}'_i > \hat{\alpha}^o_i$ 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^o_i. This contradicts the assumption of $k'_i = k^o_i$. Therefore the third term is nonnegative according to condition 3(b). Hence, we always have $\phi > 0$.

$$\begin{aligned}
& P(PC^o(n, \{k^o_j\}, \{X_j\}) > PC(n', \{k'_l\}, \{X^s_l\})) \\
= & \; P\left\{ \sum_{j=1}^{n} \frac{N_j}{N} V(k^o_j, \hat{F}^o_j, N_j) - \sum_{l}^{n'} \frac{N_l}{N} V(k'_l, \hat{F}'_l, N_l) > \phi g(N,T) \right\} \\
= & \; P\left\{ \sum_{i=1}^{n^I} \frac{N^I_i}{N} V(k^o_i, \hat{F}^o_i, N^I_i) - \sum_{i}^{n^I} \frac{N^I_i}{N} V(k'_i, \hat{F}'_i, N^I_i) > \phi g(N,T) \right\} \\
= & \; P\left\{ \sum_{i=1}^{n^I} \frac{N^I_i}{N} [V(k^o_i, \hat{F}^o_i, N^I_i) - V(k'_i, \hat{F}'_i, N^I_i)] > \phi g(N,T) \right\}
\end{aligned}$$

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.

$$P(PC^o(n, \{k_j^o\}, \{X_j\}) < PC(n', \{k_l'\}, \{X_l^s\})) \to 1. \tag{6.60}$$

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 (6.60), i.e. we need to prove

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

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

Because the PDA with the voting scheme is consistent we have

$$P\left(PC(n,\{k_j^o\},\{X_j^s\}) - PC^o(n,\{k_j^o\},\{X_i\}=0\right) = P\left(\{X_i^s\}=\{X_i\}\right) \xrightarrow{P} 1 \quad (6.61)$$

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

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

This proves Proposition 3.7.

□

6.3 Variable List for the Empirical Example

Table 8: List of Variables and Classification

Group	No.	Name	code
1	1	AUSTRALIAN AGRICULTURAL - TOT RETURN IND	A:AACX(RI)
1	2	ADELAIDE BRIGHTON - TOT RETURN IND	A:ABCX(RI)
1	3	ABACUS PROPERTY GROUP - TOT RETURN IND	A:ABPX(RI)
1	4	AGL ENERGY - TOTRETURN IND	A:AGKX(RI)(*)
1	5	AUSTRALIAN INFR.FUND - TOT RETURN IND	A:AIXX(RI)
1	7	ARISTOCRAT LEISURE - TOT RETURN IND	A:ALLX(RI)
1	8	ALESCO - TOT RETURN IND	A:ALSX(RI)
1	9	AUSTRALAND PR.GP. - TOT RETURN IND	A:ALZX(RI)
1	10	AMCOR - TOT RETURN IND	A:AMCX(RI)
1	11	AMP - TOT RETURNIND	A:AMPX(RI)
1	12	ANSELL - TOT RETURN IND	A:ANNX(RI)
1	13	AUS.AND NZ.BANKING GP. - TOT RETURN IND	A:ANZX(RI)
1	15	APA GROUP - TOT RETURN IND	A:APAX(RI)
1	16	APN NEWS & MEDIA- TOT RETURN IND	A:APNX(RI)
1	19	ASX - TOT RETURNIND	A:ASXX(RI)
1	20	AUSTAR UNITED COMMS. - TOT RETURN IND	A:AUNX(RI)
1	22	AWB - TOT RETURNIND	A:AWBX(RI)
1	23	ALUMINA - TOT RETURN IND	A:AWCX(RI)
1	25	AXA ASIA PACIFICHDG. - TOT RETURN IND	A:AXAX(RI)
1	26	BILLABONG INTERNATIONAL - TOT RETURN IND	A:BBGX(RI)
1	27	BENDIGO & ADELAIDE BANK - TOT RETURN IND	A:BENX(RI)
1	29	BORAL - TOT RETURN IND	A:BLDX(RI)
1	30	BANK OF QLND. - TOT RETURN IND	A:BOQX(RI)
1	32	BLUESCOPE STEEL - TOT RETURN IND	A:BSLX(RI)
1	33	BUNNINGS WHSE.PR.TST. - TOT RETURN IND	A:BWPX(RI)
1	34	BRAMBLES - TOT RETURN IND	A:BXBX(RI)
1	35	CABCHARGE AUSTRALIA - TOT RETURN IND	A:CABX(RI)
1	36	COMMONWEALTH BK.OF AUS. - TOT RETURN IND	A:CBAX(RI)
1	37	COCA-COLA AMATIL- TOT RETURN IND	A:CCLX(RI)
1	40	CFS RETAIL PR.TST. - TOT RETURN IND	A:CFXX(RI)
1	41	CHALLENGER FINL.SVS.GP. - TOT RETURN IND	A:CGFX(RI)
1	42	CONSOLIDATED MEDIA HDG. - TOT RETURN IND	A:CMJX(RI)
1	43	COCHLEAR - TOT RETURN IND	A:COHX(RI)
1	44	COMMONWEALTH PR.OFFE.FD. - TOT RETURN IND	A:CPAX(RI)
1	45	COMPUTERSHARE - TOT RETURN IND	A:CPUX(RI)
1	46	CRANE GROUP - TOT RETURN IND	A:CRGX(RI)
1	47	CSL - TOT RETURNIND	A:CSLX(RI)
1	48	CSR - TOT RETURNIND	A:CSRX(RI)
1	49	CALTEX AUSTRALIA- TOT RETURN IND	A:CTXX(RI)(*)
1	51	CORPORATE EXPRESS AUS. - TOT RETURN IND	A:CXPX(RI)
1	52	DAVID JONES - TOT RETURN IND	A:DJSX(RI)
1	54	DOWNER EDI - TOTRETURN IND	A:DOWX(RI)
1	55	DEXUS PROPERTY GROUP - TOT RETURN IND	A:DXSX(RI)
1	56	ELDERS - TOT RETURN IND	A:ELDX(RI)
1	57	ENVESTRA - TOT RETURN IND	A:ENVX(RI)
1	63	FOSTER'S GROUP -TOT RETURN IND	A:FGLX(RI)
1	64	FKP PROPERTY GROUP - TOT RETURN IND	A:FKPX(RI)
1	65	FLIGHT CENTRE - TOT RETURN IND	A:FLTX(RI)
1	67	FLEETWOOD - TOT RETURN IND	A:FWDX(RI)
1	68	FAIRFAX MEDIA - TOT RETURN IND	A:FXJX(RI)
1	70	GOODMAN GROUP - TOT RETURN IND	A:GMGX(RI)
1	71	GRAINCORP - TOT RETURN IND	A:GNCX(RI)
1	72	GUNNS - TOT RETURN IND	A:GNSX(RI)
1	73	GPT GROUP - TOT RETURN IND	A:GPTX(RI)
1	74	GUD HOLDINGS - TOT RETURN IND	A:GUDX(RI)
1	75	GWA INTERNATIONAL - TOT RETURN IND	A:GWTX(RI)
1	76	HENDERSON GROUP CDI. - TOT RETURN IND	A:HGGX(RI)
1	77	HILLS INDUSTRIES- TOT RETURN IND	A:HILX(RI)
1	78	HEALTHSCOPE - TOT RETURN IND	A:HSPX(RI)
1	79	HARVEY NORMAN HOLDINGS - TOT RETURN IND	A:HVNX(RI)
1	80	INSURANCE AUS.GROUP - TOT RETURN IND	A:IAGX(RI)

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

Group	No.	Name	code
1	81	IOOF HOLDINGS - TOT RETURN IND	A:IFLX(RI)
1	83	ING INDL.FUND - TOT RETURN IND	A:IIFX(RI)
1	84	ILUKA RESOURCES - TOT RETURN IND	A:ILUX(RI)(*)
1	85	ING OFFICE FUND - TOT RETURN IND	A:IOFX(RI)
1	87	IRESS MARKET TECH. - TOT RETURN IND	A:IREX(RI)
1	88	ISOFT GROUP - TOT RETURN IND	A:ISFX(RI)
1	90	JB HI-FI - TOT RETURN IND	A:JBHX(RI)
1	91	JAMES HARDIE INDS.CDI. - TOT RETURN IND	A:JHXX(RI)
1	93	LEIGHTON HOLDINGS - TOT RETURN IND	A:LEIX(RI)
1	95	LEND LEASE GROUP- TOT RETURN IND	A:LLCX(RI)
1	97	MACMAHON HOLDINGS - TOT RETURN IND	A:MAHX(RI)
1	98	MAP GROUP - TOT RETURN IND	A:MAPX(RI)
1	101	MACQUARIE COUNTRY.TRUST - TOT RETURN IND	A:MCWX(RI)
1	102	MIRVAC GROUP - TOT RETURN IND	A:MGRX(RI)
1	104	MACQUARIE INFR.GROUP - TOT RETURN IND	A:MIGX(RI)
1	107	MONADELPHOUS GROUP - TOT RETURN IND	A:MNDX(RI)
1	108	MACQUARIE OFFICETRUST - TOT RETURN IND	A:MOFX(RI)
1	110	MACQUARIE GROUP - TOT RETURN IND	A:MQGX(RI)(*)
1	112	METCASH - TOT RETURN IND	A:MTSX(RI)
1	113	NATIONAL AUS.BANK - TOT RETURN IND	A:NABX(RI)
1	115	NUFARM - TOT RETURN IND	A:NUFX(RI)
1	116	NEWS CORP.CDI.'B' (ASX) - TOT RETURN IND	A:NWSX(RI)
1	120	ORICA - TOT RETURN IND	A:ORIX(RI)
1	122	ONESTEEL - TOT RETURN IND	A:OSTX(RI)
1	126	PRIME INFRASTRUCTURE GP. - TOT RETURN IND	A:PIHX(RI)
1	129	PERPETUAL - TOT RETURN IND	A:PPTX(RI)
1	130	PAPERLINX - TOT RETURN IND	A:PPXX(RI)
1	131	PRIMARY HEALTH CARE - TOT RETURN IND	A:PRYX(RI)
1	132	QANTAS AIRWAYS -TOT RETURN IND	A:QANX(RI)
1	133	QBE INSURANCE GROUP - TOT RETURN IND	A:QBEX(RI)
1	134	RAMSAY HEALTH CARE - TOT RETURN IND	A:RHCX(RI)
1	137	RESMED CDI - TOTRETURN IND	A:RMDX(RI)
1	141	SEVEN NETWORK - TOT RETURN IND	A:SEVX(RI)
1	143	STOCKLAND - TOT RETURN IND	A:SGPX(RI)
1	144	SINGAPORE TELECOM CDI. (ASX) - TOT RETURN IND	A:SGTX(RI)
1	145	SONIC HEALTHCARE- TOT RETURN IND	A:SHLX(RI)
1	146	SIGMA PHARMS. - TOT RETURN IND	A:SIPX(RI)
1	147	SMS MAN.& TECH. - TOT RETURN IND	A:SMXX(RI)
1	148	SPOTLESS GROUP -TOT RETURN IND	A:SPTX(RI)
1	151	SUNCORP-METWAY -TOT RETURN IND	A:SUNX(RI)
1	152	TABCORP HOLDINGS- TOT RETURN IND	A:TAHX(RI)
1	153	TRANSURBAN GROUP- TOT RETURN IND	A:TCLX(RI)
1	154	TELECOM CORP.NZ.(ASX) - TOT RETURN IND	A:TELX(RI)
1	155	TEN NETWORK HOLDINGS - TOT RETURN IND	A:TENX(RI)
1	157	TOLL HOLDINGS - TOT RETURN IND	A:TOLX(RI)
1	158	TRANSFIELD SERVICES - TOT RETURN IND	A:TSEX(RI)
1	159	UGL - TOT RETURNIND	A:UGLX(RI)
1	160	VIRGIN BLUE HOLDINGS - TOT RETURN IND	A:VBAX(RI)
1	161	WEST AUST.NWSP.HDG. - TOT RETURN IND	A:WANX(RI)
1	162	WESTPAC BANKING - TOT RETURN IND	A:WBCX(RI)
1	163	WESTFIELD GROUP - TOT RETURN IND	A:WDCX(RI)
1	165	WORLEYPARSONS - TOT RETURN IND	A:WORX(RI)
1	166	WOOLWORTHS - TOTRETURN IND	A:WOWX(RI)

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

Group	No.	Name	code
2	6	AJ LUCAS GROUP -TOT RETURN IND	A:AJLX(RI)
2	14	ARROW ENERGY - TOT RETURN IND	A:AOEX(RI)
2	17	AQUILA RESOURCES- TOT RETURN IND	A:AQAX(RI)
2	18	AQUARIUS PLATINUM (ASX) - TOT RETURN IND	A:AQPX(RI)
2	21	AVOCA RESOURCES - TOT RETURN IND	A:AVOX(RI)
2	24	AWE - TOT RETURNIND	A:AWEX(RI)
2	28	BHP BILLITON - TOT RETURN IND	A:BHPX(RI)
2	31	BEACH ENERGY - TOT RETURN IND	A:BPTX(RI)
2	38	CUDECO - TOT RETURN IND	A:CDUX(RI)(*)
2	39	CENTENNIAL COAL - TOT RETURN IND	A:CEYX(RI)
2	50	CARNARVON PETROLEUM - TOT RETURN IND	A:CVNX(RI)
2	53	DOMINION MINING - TOT RETURN IND	A:DOMX(RI)
2	58	EQUINOX MINERALSCDI. - TOT RETURN IND	A:EQNX(RI)
2	59	ENERGY RES.OF AUS. - TOT RETURN IND	A:ERAX(RI)
2	60	EASTERN STAR GAS- TOT RETURN IND	A:ESGX(RI)
2	61	ENERGY WORLD - TOT RETURN IND	A:EWCX(RI)
2	62	EXTRACT RESOURCES - TOT RETURN IND	A:EXTX(RI)
2	66	FORTESCUE METALSGP. - TOT RETURN IND	A:FMGX(RI)
2	69	GINDALBIE METALS- TOT RETURN IND	A:GBGX(RI)
2	82	INDEPENDENCE GROUP - TOT RETURN IND	A:IGOX(RI)
2	86	INCITEC PIVOT - TOT RETURN IND	A:IPLX(RI)(*)
2	89	INVOCARE - TOT RETURN IND	A:IVCX(RI)(*)
2	92	KINGSGATE CONSOLIDATED - TOT RETURN IND	A:KCNX(RI)
2	94	LIHIR GOLD - TOTRETURN IND	A:LGLX(RI)
2	96	LYNAS - TOT RETURN IND	A:LYCX(RI)
2	99	MACARTHUR COAL -TOT RETURN IND	A:MCCX(RI)
2	100	MINCOR RESOURCES- TOT RETURN IND	A:MCRX(RI)
2	103	MOUNT GIBSON IRON - TOT RETURN IND	A:MGXX(RI)
2	105	MEDUSA MINING - TOT RETURN IND	A:MMLX(RI)
2	106	MURCHISON METALS- TOT RETURN IND	A:MMXX(RI)
2	109	MOLOPO ENERGY - TOT RETURN IND	A:MPOX(RI)
2	111	MINARA RESOURCES- TOT RETURN IND	A:MREX(RI)
2	114	NEWCREST MINING - TOT RETURN IND	A:NCMX(RI)
2	117	NEXUS ENERGY - TOT RETURN IND	A:NXSX(RI)
2	118	OM HOLDINGS - TOT RETURN IND	A:OMHX(RI)
2	119	ORIGIN ENERGY (EX BORAL) - TOT RETURN IND	A:ORGX(RI)
2	121	OIL SEARCH - TOTRETURN IND	A:OSHX(RI)
2	123	OZ MINERALS - TOT RETURN IND	A:OZLX(RI)
2	124	PANORAMIC RESOURCES - TOT RETURN IND	A:PANX(RI)
2	125	PALADIN ENERGY -TOT RETURN IND	A:PDNX(RI)
2	127	PLATINUM AUSTRALIA - TOT RETURN IND	A:PLAX(RI)
2	128	PANAUST - TOT RETURN IND	A:PNAX(RI)
2	135	RIO TINTO - TOT RETURN IND	A:RIOX(RI)
2	136	RIVERSDALE MINING - TOT RETURN IND	A:RIVX(RI)
2	138	ROC OIL COMPANY - TOT RETURN IND	A:ROCX(RI)
2	139	ST BARBARA - TOTRETURN IND	A:SBMX(RI)
2	140	SUNDANCE RESOURCES - TOT RETURN IND	A:SDLX(RI)
2	142	SIMS METAL MANAGEMENT - TOT RETURN IND	A:SGMX(RI)
2	149	STRAITS RESOURCES - TOT RETURN IND	A:SRLX(RI)
2	150	SANTOS - TOT RETURN IND	A:STOX(RI)
2	156	TELSTRA - TOT RETURN IND	A:TLSX(RI)(*)
2	164	WESFARMERS - TOTRETURN IND	A:WESX(RI)(*)
2	167	WOODSIDE PETROLEUM - TOT RETURN IND	A:WPLX(RI)
2	168	WESTERN AREAS - TOT RETURN IND	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.

Connor, G. and Korajzyk, R. (1986). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48:1263–1291.

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*.

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

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*.

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

STOCK, J. H. AND WATSON, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.

— (2002b). 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.