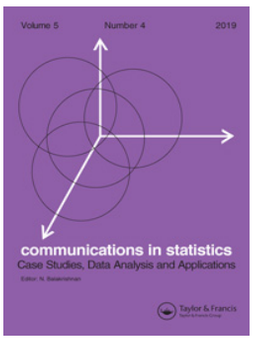

Taylor & Francis
Taylor & Francis Group

Communications in Statistics: Case Studies, Data Analysis and Applications

ISSN: (Print) 2373-7484 (Online) Journal homepage: https://www.tandfonline.com/loi/ucas20

A comparison of two estimation methods for common principal components

Toni Duras

To cite this article: Toni Duras (2019) A comparison of two estimation methods for common principal components, Communications in Statistics: Case Studies, Data Analysis and Applications, 5:4, 366-393, DOI: 10.1080/23737484.2019.1656117

To link to this article: https://doi.org/10.1080/23737484.2019.1656117

Published online: 27 Aug 2019.

COMMUNICATIONS IN STATISTICS: CASE STUDIES, DATA ANALYSIS AND APPLICATIONS
2019, VOL. 5, NO. 4, 366–393
https://doi.org/10.1080/23737484.2019.1656117

Taylor & Francis
Taylor & Francis Group

A comparison of two estimation methods for common principal components

Toni Duras

Department of statistics, Jönköping International Business School, Jönköping, Sweden

ABSTRACT

Common principal components (CPCs) are often estimated using maximum likelihood estimation through an algorithm called the Flury–Gautschi (FG) Algorithm. Krzanowski proposed a simpler estimation method via a principal component analysis of a weighted sum of the sample covariance matrices. These methods are compared for real-world datasets and in a Monte Carlo simulation. The real-world data is used to compare the selection of a common eigenvector model and the estimated coefficients. The simulation study investigates how the accuracy of the methods is affected by autocorrelation, the number of covariance matrices, dimensions, and sample sizes for multivariate normal and chi-square distributed data. The findings in this article support the use of Krzanowski's method in situations where the CPC assumption is appropriate.

ARTICLE HISTORY
Received 6 October 2018
Accepted 8 August 2019

KEYWORDS
Common principal components; identification of common eigenvector models; maximum likelihood estimation; Monte Carlo simulation

1. Introduction

Multivariate analysis deals with making inferences of many, usually intercorrelated, variables simultaneously. Applications include fields such as mechanics, biology, engineering, and economics, to mention a few. The straight-forward and intuitive understanding of univariate problems is partly lost as several variables are involved in the analysis. This difficulty in comprehending the meaning and role of multidimensional quantities becomes even greater when involving groups of such multivariate measurements. To make sense of such quantities, one may utilize relations between variables to impose certain simplifying restrictions on the models. One such relationship is the extension of principal component analysis (PCA) to multiple groups, called common principal components analysis (CPCA). Introduced and developed by Flury (1984, 1988), CPCA assumes that some rotation can diagonalize the covariance matrices simultaneously; that is, they have a common eigenvector structure but may have group-specific eigenvalues. This relationship between covariance matrices is commonly known as the common principal component (CPC)

CONTACT Toni Duras toni.duras@ju.se Department of statistics, Jönköping International Business School, Box 1026, 551 11 Jönköping, Sweden.
Communicated by 12 August 2019

model. In one sample PCA, the eigenvectors are ordered by the rank order of their corresponding eigenvalue. For CPCA, no natural order of the set of eigenvectors exists because the rank order of the eigenvalues may vary across groups. When the CPC model is suitable, more accurate estimations of the covariance matrices are provided than if each covariance matrix is estimated individually.

CPCs are commonly estimated using maximum likelihood estimation (MLE). An algorithm is required to solve the maximum likelihood (ML) equations. The Flury–Gautschi (FG) algorithm was proposed by Flury and Gautschi (1986) and is an iterative numerical algorithm that requires considerable computational power. It loops through all pairs of columns of the eigenvector matrix to complete a single iteration, which makes it ineffective and sometimes infeasible in higher dimensional settings (Browne and McNicholas 2014). The asymptotic distribution of the maximum likelihood estimators was derived by Flury (1986).

Krzanowski (1984) proposed a simpler approximation of the CPCs which requires only the standard computer packages for PCA. Under the CPC setting, any weighted sum of the marginal covariance matrices has the same eigenvectors as the marginal covariance matrices. Thus, the CPC coefficients can be estimated by performing a PCA of any weighted sum (or product) of the sample covariance matrices. The asymptotic behavior of PCA was derived by Girshick (1939), Lawley (1953, 1956), and Anderson (1963). A summary of the applications and properties of PCA is given by Jolliffe (2002).

Flury (1988) introduced a hierarchy of similarities among several covariance matrices. This hierarchy consists of 5 levels, ranging from equality between covariance matrices (level 1) to completely unrelated covariance matrices (level 5). Level 2 is the proportionality of covariance matrices, studied by Pillai, Al-Ani, and Jouris (1969) and Rao (1973), level 3 is the mentioned CPC model, and the fourth level is the partial CPC model, which assumes that only q ($< p$) out of p eigenvectors are common for all groups, denoted as CPC(q) (it is possible to break this level down further into models CPC($p-2$) to CPC(1)). To identify an appropriate common eigenvector model for the data, two goodness-of-fit measurements were proposed by Flury (1988): a log-likelihood ratio (LR) statistic and the Akaike information criterion (AIC). Pepler, Uys, and Nel (2016) compared 8 methods for selecting suitable common eigenvector models, including the LR statistic and AIC. This article considers three model identification methods: the AIC, the Schwarz criterion [also known as the Bayesian information criterion (BIC)], and the LR statistic, to be used to identify an appropriate model for a given dataset.

For a detailed overview concerning CPCA of covariance matrices, see Flury (1988).

This article compares the ML and Krzanowski's estimators on two real-world datasets and in a Monte Carlo simulation study. The two datasets are, (i) annual Swedish municipal level data related to innovations, collected for

11 years, and (ii) the Iris flower data. Based on these datasets, the estimated CPC coefficients and the three model identification statistics are calculated and compared for the two estimation methods. The results obtained by Flury (1984) and Krzanowski (1984) on the Iris data are confirmed. The simulation study examines how the accuracy of the estimation methods is affected by the number of groups, dimensions, sample sizes, and by autocorrelations covariance matrices for multivariate normal data and for chi-square-distributed data (with 2 and 10 degrees of freedom).

2. Preliminaries

Bold capital letters denote matrices (e.g., $\mathbf{A} \in \mathbb{R}^{N\times p}$), bold lower-case letters denote vectors (e.g., $\mathbf{a} \in \mathbb{R}^p$), and normal lower-case letters denote constants (e.g., $c \in \mathbb{R}$). The dimension p corresponds to the number of "variables," N to the number of "observations," and G to the number of groups.

Notation 1. Let $|\mathbf{A}|$ denote the determinant of a square matrix $\mathbf{A}$.

Notation 2. Let $\bar{\mathbf{A}}_G = \frac{1}{G}(\mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_G)$.

Notation 3. Let $\mathbf{A}'$ denote the transpose of $\mathbf{A}$.

Notation 4. The continuous uniform distribution, the matrix normal distribution and the Wishart distribution are denoted as $U \sim U(a,b)$, $\mathbf{X} \sim N_{N\times p}(\mathbf{M}, \Psi, \Sigma)$ and $\mathbf{W} \sim W_p(N, \Sigma)$, respectively. Throughout this article $\Psi = \mathbf{I}_N$ and the notation $\mathbf{X} \sim N_{N\times p}(\mathbf{M}, \Sigma)$ is used instead.

Definition 1 (Orthogonal group). The set of orthogonal $p\times p$ matrices is denoted $\mathbb{O}_p$ and defined as $\mathbb{O}_p = \{\mathbf{O}_{p\times p} : \mathbf{O}'\mathbf{O} = \mathbf{O}\mathbf{O}' = \mathbf{I}_p\}$. $\mathbb{O}_p$ forms a *group*, with its group operation being matrix multiplication.

Definition 2. For any symmetric matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, we can find $\lambda \in \mathbb{R}$ and $\boldsymbol{\pi} \in \mathbb{R}^p$ such that

$$\mathbf{A}\boldsymbol{\pi} = \lambda\boldsymbol{\pi}. \tag{2.1}$$

We say that $\lambda \in \mathbb{R}$ and $\boldsymbol{\pi} \in \mathbb{R}^p$ are a (right) *eigenvalue* and *eigenvector* pair. The eigenvectors are orthogonal and usually normalized to unit length, i.e., $\boldsymbol{\pi}_i'\boldsymbol{\pi}_j = 1$ for $i = j = 1,\ldots,p$ and zero otherwise. Π collects the set of eigenvectors of $\mathbf{A}$ as columns, $\Pi = (\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_p) \in \mathbb{R}^{p\times p}$. Since the columns of Π are orthonormal vectors, the matrix is orthogonal. The diagonal matrix Λ collects the corresponding set of eigenvalues in decreasing order, i.e., $\Lambda = \operatorname{diag}(\lambda_1,\ldots,\lambda_p)$ where $\lambda_1 \geq \cdots \geq \lambda_p$.

Definition 3 (*Spectral decomposition*). Any symmetric matrix $\mathbf{A}$ can be factorized as

$$\mathbf{A} = \Pi\Lambda\Pi' = \sum_{i=1}^{p} \lambda_i \boldsymbol{\pi}_i \boldsymbol{\pi}_i'. \tag{2.2}$$

2.1. Common principal components analysis

CPCA is the generalization of PCA to multiple groups. Let $\mathbf{X}_1, \ldots, \mathbf{X}_G$ be distributed as $N_{N_g \times p}\left(\mathbf{M}_g, \Sigma_g\right)$ for $g = 1, \ldots, G$, where the same set of variables are measured for all groups. In the CPC model, the associated covariance matrices, Σ_g, can simultaneously be diagonalized by the same orthogonal matrix Π, i.e.,

$$\Sigma_g = \Pi \Lambda_g \Pi', \quad g = 1, \ldots, G, \tag{2.3}$$

where Π collect the eigenvectors of Σ_g as columns and Λ_g is a diagonal matrix whose jth element is the eigenvalue corresponding to the eigenvector in the jth column of Π. Consequently, all covariance matrices share the same set of eigenvectors but have group-specific eigenvalues.

3. The hierarchy levels and estimation

Definition 4. The five levels of Flury's (1988) hierarchy of similarities among covariance matrices are defined below. Consider the positive definite covariance matrices $\Sigma_1, \ldots, \Sigma_G$ of dimension $p \times p$:

Level 1: Equality of all Σ_g, i.e., $\Sigma_1 = \cdots = \Sigma_G$.

The principal components can be obtained by treating the groups as a single one, pooling the sample covariance matrices, and performing a PCA. The number of estimated parameters is thus $\frac{1}{2}p(p+1)$.

Level 2: Proportionality of all Σ_g, i.e.,

$$\Sigma_g = \rho_g \Sigma_1, \quad g = 2, \ldots, G, \quad \rho_2, \ldots, \rho_G > 0. \tag{3.1}$$

The number of estimated parameters is $\frac{1}{2}p(p+1) + G - 1$.

Level 3: CPC model

$$\Sigma_g = \Pi \Lambda_g \Pi', \quad g = 1, \ldots, G. \tag{3.2}$$

The number of estimated parameters for the CPC model is $\frac{1}{2}p(p-1)$ for the orthogonal matrix plus Gp for the eigenvalues.

Level 4: Partial common principal component model (CPC(q))

$$\Sigma_g = \Pi^{(g)} \Lambda_g \Pi^{(g)'}, \quad g = 1, \ldots, G, \tag{3.3}$$

where $\Pi^{(g)} = \left(\Pi_c : \Pi_s^{(g)}\right)$, Π_c is the common $p \times q$ part across all groups and $\Pi_s^{(g)}$ is the $p \times (p-q)$ part specific to group g. The number of estimated parameters for the CPC(q) model is: Gp parameters for the eigenvalues, $p(p-1)/2 - (p-q)(p-q-1)/2$ parameters for the common eigenvectors Π_c, and $G(p-q)(p-q-1)/2$ for the specific eigenvectors $\Pi_s^{(g)}$. The total number of estimated parameters is therefore $\frac{1}{2}p(p-1) + Gp + \frac{1}{2}(G-1)(p-q)(p-q-1)$.

Level 5: Unrelated $\Sigma_1, \ldots, \Sigma_G$

The number of estimated parameters is $\frac{1}{2}Gp(p+1)$ and the covariance matrices are analyzed separately.

3.1. Maximum likelihood estimation

Assume G independent matrices $\mathbf{X}_1, \ldots, \mathbf{X}_G$ to be normally distributed as $N_{N_g \times p}\left(\mathbf{M}_g, \Sigma_g\right)$, where the same p variables have been measured across all groups for sample sizes $N_g = n_g + 1$, $\mathbf{M}_g = \mathbf{1}_{N_g}\boldsymbol{\mu}_g'$, $\boldsymbol{\mu}_g = (\mu_{g1}, \ldots, \mu_{gp})'$ and $\mathbf{1}_{N_g}$ is a $N_g \times 1$ vector of ones. Let $\mathbf{S}_g$ denote the usual unbiased covariance estimator

$$\mathbf{S}_g = \frac{1}{n_g}\mathbf{X}_g'\mathbf{Q}_g\mathbf{X}_g = \frac{1}{n_g}\left(\mathbf{X}_g - \mathbf{1}_{N_g}\bar{\mathbf{x}}_g'\right)'\left(\mathbf{X}_g - \mathbf{1}_{N_g}\bar{\mathbf{x}}_g'\right), \tag{3.4}$$

where $\mathbf{Q}_g = \mathbf{I}_{N_g} - \frac{1}{N_g}\mathbf{1}_{N_g}\mathbf{1}_{N_g}'$ and $\bar{\mathbf{x}}_g = \frac{1}{N_g}\mathbf{1}_{N_g}'\mathbf{X}_g$, further $n_g\mathbf{S}_g \sim W_p(n_g, \Sigma_g)$. The likelihood function of $\Sigma_1, \ldots, \Sigma_G$ given $\mathbf{S}_1, \ldots, \mathbf{S}_G$ is

$$L = L\left(\Sigma_1, \ldots, \Sigma_G | \mathbf{S}_1, \ldots, \mathbf{S}_G\right) = C \times \prod_{g=1}^{G} \exp\left[\operatorname{tr}\left(-\frac{n_g}{2}\Sigma_g^{-1}\mathbf{S}_g\right)\right] |\Sigma_g|^{-n_g/2}, \tag{3.5}$$

where C is a constant that does not depend on any Σ_g. The likelihood function attains its maximum at the same point as the log-likelihood function. Thus for convenience reasons the function $-2\log L$ is minimized.

Level 2:
If the proportionality assumption holds, by imposing the orthogonal restriction on the eigenvectors and letting $\rho_1 = 1$, the likelihood equations are (Flury 1988; Pepler 2014)

$$\rho_g = \frac{1}{p}\sum_{j=1}^{p}\frac{\boldsymbol{\pi}_j'\mathbf{S}_g\boldsymbol{\pi}_j}{\lambda_j}, \quad g = 2, \ldots, G, \tag{3.6}$$

$$\lambda_j = \frac{1}{n}\sum_{g=1}^{G}\frac{n_g}{\rho_g}\boldsymbol{\pi}_j'\mathbf{S}_g\boldsymbol{\pi}_j, \quad j = 1, \ldots, p, \tag{3.7}$$

where $n = n_1 + \cdots + n_G$, and

$$\left(\frac{1}{\lambda_i} - \frac{1}{\lambda_j}\right)\boldsymbol{\pi}_j'\left(\sum_{g=1}^{G}\frac{n_g}{\rho_g}\mathbf{S}_g\right)\boldsymbol{\pi}_i = 0, \quad i \neq j. \tag{3.8}$$

The likelihood equations can be solved by the iterative algorithm proposed by Flury (1988), outlined in Appendix A.1.

Level 3:
By assuming that the CPC assumption holds and imposing the orthogonal

restriction on the eigenvectors, the minimization problem can be simplified to one that involves solving the following system of equations (Flury 1984)

$$\boldsymbol{\pi}_i' \left(\sum_{g=1}^{G} n_g \frac{\lambda_{gi} - \lambda_{gj}}{\lambda_{gi}\lambda_{gj}} \mathbf{S}_g \right) \boldsymbol{\pi}_j = 0, \quad i, j = 1, \ldots, p;\ i \neq j, \tag{3.9}$$

where $\lambda_{gj} = \boldsymbol{\pi}_j' \mathbf{S}_g \boldsymbol{\pi}_j$.

Flury and Gautschi (1986) developed an algorithm known as the FG-algorithm to solve the equation system. The ML solution is denoted as $\hat{\Pi} = (\hat{\boldsymbol{\pi}}_1, \ldots, \hat{\boldsymbol{\pi}}_p)$ and $\hat{\lambda}_{gj}$, and let $\hat{\Lambda}_g = \operatorname{diag}\left(\hat{\Pi}' \mathbf{S}_g \hat{\Pi}\right) = \operatorname{diag}\left(\hat{\lambda}_{g1}, \ldots, \hat{\lambda}_{gp}\right)$ and $\hat{\Sigma}_g = \hat{\Pi} \hat{\Lambda}_g \hat{\Pi}'$, for $g = 1, \ldots, G$ and $j = 1, \ldots, p$.

In PCA (the special case of $G = 1$) the columns of Π are ordered according to the rank order of its eigenvalues. For the CPC model, no self-evident order exists. To be able to discuss the first, second, or other principal components, some arrangement is needed. For the ML estimation, Flury's (1986) suggestion of using the rank order of the first group is adopted, i.e.,

$$\boldsymbol{\pi}_1' \Sigma_1 \boldsymbol{\pi}_1 > \cdots > \boldsymbol{\pi}_p' \Sigma_1 \boldsymbol{\pi}_p. \tag{3.10}$$

Level 4:
The CPC model risks being rejected due to non-common components which are discarded in the succeeding analysis, even if the components of interest are common throughout all groups. Thus, the partial CPC model could be selected if some of the components with low variance can be abandoned. Obtaining its exact ML estimates are substantially more complicated than in the CPC model, no simple method is available. However, given the ML estimates of the CPC model, approximative estimates can easily be obtained by the following procedure (Flury 1988):

Step 1: Compute $\hat{\Pi} = \left(\hat{\Pi}_1 : \hat{\Pi}_2\right) \in \mathbb{O}_p$ through the CPC model, where $\hat{\Pi}_1$ contains the q columns common across the groups. Some reordering of the columns of $\hat{\Pi}$ might be required to obtain the common components as the first q columns.

Step 2: Diagonalize the matrix $\hat{\Pi}_2' \mathbf{S}_g \hat{\Pi}_2$, i.e., find $\mathbf{Q}_g \in \mathbb{O}_{p-q}$ such that $\mathbf{Q}_g' \hat{\Pi}_2' \mathbf{S}_g \hat{\Pi}_2 \mathbf{Q}_g$ is diagonal, and let $\hat{\Pi}_2^{(g)} = \hat{\Pi}_2 \mathbf{Q}_g$ (for all $g = 1, \ldots, G$).

Step 3: The approximate ML estimated partial CPCs are then

$$\hat{\hat{\Pi}}^{(g)} = \left(\hat{\Pi}_1 : \hat{\Pi}_2^{(g)}\right), \tag{3.11}$$

$$\hat{\hat{\Lambda}}_g = \operatorname{diag}\left(\hat{\hat{\Pi}}^{(g)'} \mathbf{S}_g \hat{\hat{\Pi}}^{(g)}\right) = \operatorname{diag}\left(\hat{\hat{\lambda}}_{g1}, \ldots, \hat{\hat{\lambda}}_{gp}\right), \tag{3.12}$$

$$\hat{\hat{\Sigma}}_g = \hat{\hat{\Pi}}^{(g)} \hat{\hat{\Lambda}}_g \hat{\hat{\Pi}}^{(g)'}, \quad g = 1, \ldots, G. \tag{3.13}$$

3.2. Krzanowski's estimation of CPC

Krzanowski (1984) proposed a simple method for obtaining estimates of Π and Λ_g which are available in any standard computer package for PCA. Under the CPC assumption

$$\Pi'\bar{\Sigma}\Pi = G^{-1}\left(\Pi'\Sigma_1\Pi + \cdots + \Pi'\Sigma_G\Pi\right) = G^{-1}\left(\Lambda_1 + \cdots + \Lambda_G\right) = \bar{\Lambda}, \tag{3.14}$$

where $\bar{\Lambda}$ and Π are the collection of eigenvalues and eigenvectors of $\bar{\Sigma}$. Estimates of Π can thus be obtained from a PCA of the sample counterpart $\bar{\mathbf{S}}$, i.e., $\bar{\mathbf{S}} = \mathbf{P}\bar{\mathbf{D}}\mathbf{P}'$, and estimates of Λ_g are established as $\mathbf{D}_g = \operatorname{diag}\left(\mathbf{P}'\mathbf{S}_g\mathbf{P}\right)$. Note that any weighted sum (or product) of the marginal covariance matrices will have the same eigenvectors. For this estimator, it is natural to rank order the eigenvectors according to the eigenvalues of the arithmetic mean of the sample covariance matrices, i.e.,

$$\mathbf{p}_1'\bar{\mathbf{S}}\mathbf{p}_1 > \cdots > \mathbf{p}_p'\bar{\mathbf{S}}\mathbf{p}_p. \tag{3.15}$$

To obtain partial CPC estimates using this method, substitute the ML estimated $\hat{\Pi}$ with $\mathbf{P}$ and follow the procedure given above.

Clearly, this is not an exact solution to the ML maximization problem; thus, it will always produce lower valued likelihood functions than the ML estimates.

4. Identification of common eigenvector models

Comparing the ML functions under various models is one way of assessing which model better fits the data. The LR test for comparing model A (the null hypothesis) to a hierarchically lower model B is

$$\chi^2\left(A|B\right) = \sum_{g=1}^{G} n_g \log \frac{\left|\hat{\Sigma}_g^{(A)}\right|}{\left|\hat{\Sigma}_g^{(B)}\right|}, \tag{4.1}$$

where the estimates $\hat{\Sigma}_g^{(A)}$ refer to model A and $\hat{\Sigma}_g^{(B)}$ to model B. χ^2 is chi-square distributed with degrees of freedom equal to the difference in dimensionality between model A and model B. The covariance matrices are assumed to be independent if an overall test of equality is rejected. The hypotheses for such test are

$$\begin{aligned} &H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_G, \\ &H_1 : \text{Unrelated covariance matrices}, \end{aligned} \tag{4.2}$$

and the LR statistics is (Flury 1988)

$$\chi^2_{\text{total}} = \chi^2(\text{equality} \mid \text{unrelated}) = \sum_{g=1}^{G} n_g \log \frac{\left|\mathbf{S}_p\right|}{\left|\mathbf{S}_g\right|}, \tag{4.3}$$

Table 1. Decomposition of the log-likelihood ratio statistic into partial log-likelihood ratio statistics.

Higher model	Lower model	Degrees of freedom
Equality	Proportionality	$G-1$
Proportionality	CPC	$(p-1)(G-1)$
CPC	CPC(q)	$\frac{1}{2}(G-1)(p-q)(p-q-1)$
CPC(q)	Unrelated	$\frac{1}{2}(G-1)(2pq-q^2-q)$

where $\mathbf{S}_p = \frac{1}{n}\sum_{g=1}^{G} n_g \mathbf{S}_g$ is the pooled covariance matrix. Flury (1988) showed that (4.3) can be decomposed into partial LR statistics, following the hierarchy levels, as

$$\begin{aligned}\chi^2_{\text{total}} = {} & \chi^2(\text{equality} \mid \text{proportionality}) \\ & + \chi^2(\text{proportionality} \mid \text{CPC}) \\ & + \chi^2(\text{CPC} \mid \text{CPC}(q)) \\ & + \chi^2(\text{CPC}(q) \mid \text{unrelated}).\end{aligned} \tag{4.4}$$

The decomposition is summarized in Table 1.

He further mentions that if one lets the data decide which model to use, hypothesis testing might not be appropriate. Instead, he suggested using the partial LR statistics as descriptive and comparing their relative sizes after dividing them with the associated degrees of freedom. The value closest to one indicates the most appropriate model. Note that the statistic for unrelated covariance matrices cannot be calculated explicitly, which is a major shortcoming.

To be able to compare all the possible models and to avoid the multiple testing problem, two alternative methods are also proposed, namely, AIC and BIC.

Proposed by Akaike (1973), AIC is a criterion used for selecting models for a given set of data, essentially measuring the quality of each model for the given dataset. AIC is defined as follows:

$$\begin{aligned}\text{AIC}^* = {} & -2\log(\text{maximum of log-likelihood function}) \\ & + 2(\text{number of estimated parameters}).\end{aligned} \tag{4.5}$$

The measurement has a close relationship with the ML estimation and penalizes models based on the number of estimated parameters. Unlike hypothesis testing, the AIC is not intended to find a "true" model; it compares the fit of competing models and the model that attains the lowest value is considered to best fit the data.

Flury (1988) proposed a modified version of the AIC. Assume r competing models, let $k_1 < k_2 < \cdots < k_r$ denote the number of estimated parameters, and $L_1 \leq L_2 \leq \ldots \leq L_r$ the maxima of the likelihood functions for the r models. The modified AIC for model i is then

$$\text{AIC}(i) = -2\log\frac{L_i}{L_r} + 2\,(k_i - k_1)\,. \tag{4.6}$$

The model with the lowest AIC is selected.

Based on Bayesian argumentation Schwarz (1978) proposed an alternative to the AIC, known as the Schwarz criterion or the BIC, which is defined as

$$\text{BIC}^* = -2\log L + k\log N, \tag{4.7}$$

where N is the number of observations and k is the number of estimated parameters. The model with the lowest BIC is selected. Notice that BIC penalizes higher-dimensional models more than AIC does (when $N \geq 8$).

Modifying BIC in the same way as Flury did for the AIC measure, the BIC calculation becomes

$$\text{BIC}(i) = -2\log\frac{L_i}{L_r} + (k_i - k_1)\log N. \tag{4.8}$$

5. Simulation study

A Monte Carlo simulation study was conducted to compare the performance of the two estimation methods. Groups of $G = 2$, 4, 10, and 50 are considered, with sample sizes of $N = 50$, 100, 500, 1,000, and 10,000 collected for $p = 5$, 10, 20, and 50 variables simulated from multivariate normal and chi-square distributions (with 2 and 10 degrees of freedom). Furthermore, autocorrelation between groups is implemented through an AR(1) process where the autocorrelation parameter is set to $\phi = -0.9$, 0, and 0.9. Because the AR(1) process is a summation of variables, it is generally at least approximately normally distributed. To obtain autocorrelated data from a chi-square distribution, let

$$\mathbf{Y}_t = \phi\mathbf{Y}_{t-1} + \boldsymbol{\delta}_t, \tag{5.1}$$

where $\mathbf{Y}_t = \left\{Y_{tij}\right\}_{\substack{i=1,\ldots,N\\ j=1,\ldots,p}}$, $\boldsymbol{\delta}_t \overset{\text{iid}}{\sim} N\left(\mathbf{0}, \mathbf{I}_p\right)$ and $|\phi| < 1$. Further, let $\tilde{\mathbf{Y}}_t = \left(\tilde{\mathbf{y}}_{t1}, \ldots, \tilde{\mathbf{y}}_{tp}\right) = \mathbf{Y}_t/\sqrt{\sigma_Y^2}$ where $\sigma_Y^2 = \text{Var}\left(Y_{tij}\right) = 1/\left(1-\phi^2\right)$. Thus, $\tilde{Y}_{tij}^2 \sim \chi_{(1)}^2$ independently of ϕ with autocorrelation $\rho_{\tilde{Y}_{tij}^2}(k) = \phi^{2k}$, and the sum of r independent $\chi_{(1)}^2$ variables are $\chi_{(r)}^2$-distributed (Holgersson 2005). To obtain normally distributed data, do not square $\tilde{Y}_{tij}$. To start up the process, 40 iterations are executed.

For group g, let $\mathbf{X}_g = \left(\mathbf{x}_{g1}, \ldots, \mathbf{x}_{gp}\right)$ denote the data matrix obtained from the AR(1) process above where $t = g$. When the independent observation vectors $\mathbf{x}_{gj}$ are $\chi_{(r)}^2$-distributed, they are normalized to have unit variance, i.e.,

$$\mathbf{z}_{gj} = \frac{\mathbf{x}_{gj}}{\sqrt{2r}}, \quad g = 1, \ldots, G; \quad j = 1, \ldots, p. \tag{5.2}$$

Then, let $\mathbf{W}_g = \mathbf{Z}_g\Sigma_g^{1/2}$ where $\mathbf{Z}_g = \left(\mathbf{z}_{g1}, \ldots, \mathbf{z}_{gp}\right)$, $\Sigma_g^{1/2} = \Pi\Lambda_g^{1/2}\Pi'$ and $\Lambda_g^{1/2} = \text{diag}\left(\sqrt{\lambda}_{g1}, \ldots, \sqrt{\lambda}_{gp}\right)$. For the normally distributed data, the normalization process (5.2) is skipped and $\mathbf{W}_g = \mathbf{X}_g\Sigma_g^{1/2}$. For a sufficiently

large sample size the covariance matrix of $\mathbf{W}_g$ is approximately Σ_g (Pepler, Uys, and Nel 2016).

The underlying common eigenvectors (Π) are obtained from $\Omega = \mathbf{A}'\mathbf{A}/8$ where $\mathbf{A} : 8 \times p$ is simulated from a multivariate normal distribution with a zero mean and with an identity covariance matrix. The group-specific eigenvalues (Λ_g) are sorted in decreasing magnitude and simulated from $(0.5 + U)^2$ where $U \sim U(0, 1)$. Pepler (2014) showed that increased separation of the eigenvalues per group translates to an improved performance in identifying a suitable common eigenvector structure. It also likely increases the accuracy of the two estimation methods in this simulation study because it simplifies the ordering of the eigenvectors. The performance of the estimation methods is measured using the p-normalized Frobenius norm

$$\|\hat{\Theta} - \Theta\|_{F_p} = \frac{1}{p}\sqrt{\text{tr}(\hat{\Theta} - \Theta)'(\hat{\Theta} - \Theta)} \tag{5.3}$$

for matrices and the Euclidean vector norm

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| = \sqrt{\left(\hat{\theta}_1 - \theta_1\right)^2 + \cdots + \left(\hat{\theta}_p - \theta_p\right)^2} \tag{5.4}$$

for vectors. The average Euclidean vector norm and the average p-normalized Frobenius norm across groups are denoted by $\|\cdot\|^{\text{AVG}}$ $\|\cdot\|_{F_p}^{\text{AVG}}$, respectively.

For the results presented below Flury's suggestion of ordering the eigenvectors according to the eigenvalues of the first group is adopted.

Each result is based on 5,000 simulations, and the benchmark parameter values are $G = 4$, $N = 100$, $p = 10$, and $\phi = 0$. The results for normally distributed data are summarized in Table 2, and the results for the chi-square distributions are summarized in Tables 3 and 4.

First the estimation of the eigenvector matrix and the group specific eigenvalues is considered. Generally, both estimation methods perform best with multivariate normal data and worst with the chi-square data with 2 degrees of freedom. However, the performance differences between normal and chi-square data with 10 degrees of freedom are mostly very small, not surprising as the chi square distribution becomes more symmetric as the degrees of freedom increases . Krzanowski's estimation clearly outperforms the ML estimation for all distributions and for all simulated parameter value combinations. Moreover, the accuracy of both methods improves with increasing sample sizes and groups, and deteriorates with increasing dimensions. However, by the results it seems like the overall precision of the estimation of Π increases when p increases. This result is contradictory on an intuitively basis and goes against general results found in multivariate applications. This could possibly be a result of the normalization constant for the Frobenius norm being selected as $\frac{1}{p}$. The autocorrelation between groups has a negative effect on the accuracy of Krzanowski's estimation; the ML estimator of Π is not notably affected by autocorrelation, however, the precision of the estimated eigenvalues are negatively affected. Krzanowski's

Table 2. Simulation results for normally distributed data.

	$\|\hat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}_1\|$	$\|\mathbf{p}_1 - \boldsymbol{\pi}_1\|$	$\|\hat{\boldsymbol{\pi}}_p - \boldsymbol{\pi}_p\|$	$\|\mathbf{p}_p - \boldsymbol{\pi}_p\|$	$\|\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\|_{F_p}$	$\|\mathbf{P} - \boldsymbol{\Pi}\|_{F_p}$	$\|\hat{\boldsymbol{\Lambda}}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\mathbf{D}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\hat{\boldsymbol{\Lambda}}_g - \text{diag}(\hat{\boldsymbol{\Lambda}}_g)\|_{F_p}^{\text{AVG}}$	$\|\mathbf{D}_g - \text{diag}(\mathbf{D}_g)\|_{F_p}^{\text{AVG}}$
$G = 2$	0.7954	0.6447	0.5939	0.4182	0.3263	0.2622	0.5407	0.5124	0.0628	0.0735
$G = 4$	0.7444	0.4928	0.5293	0.2698	0.3197	0.1983	0.4983	0.4688	0.0825	0.0898
$G = 10$	0.6799	0.3070	0.4183	0.1460	0.3069	0.1139	0.4576	0.4400	0.0937	0.0972
$G = 50$	0.5672	0.1230	0.3611	0.0609	0.2961	0.0422	0.4306	0.4263	0.0996	0.1003
$N = 50$	0.8952	0.6383	0.7241	0.3893	0.3531	0.2536	0.7229	0.6644	0.1139	0.1264
$N = 100$	0.7444	0.4928	0.5293	0.2698	0.3197	0.1983	0.4983	0.4688	0.0825	0.0898
$N = 500$	0.4599	0.2403	0.2390	0.1157	0.2265	0.0940	0.2204	0.2156	0.0385	0.0407
$N = 1{,}000$	0.3498	0.1713	0.1714	0.0819	0.1848	0.0654	0.1576	0.1557	0.0274	0.0288
$N = 10{,}000$	0.1176	0.0512	0.0606	0.0250	0.0774	0.0196	0.0510	0.0509	0.0088	0.0091
$p = 5$	0.4560	0.2522	0.3187	0.1610	0.2930	0.1353	0.3354	0.3300	0.0763	0.0831
$p = 10$	0.7444	0.4928	0.5293	0.2698	0.3197	0.1983	0.4983	0.4688	0.0825	0.0898
$p = 20$	1.0261	0.7825	0.8313	0.4655	0.2738	0.2194	0.7644	0.6654	0.0839	0.0921
$p = 50$	1.2560	1.1161	1.2385	0.8507	0.1910	0.1808	1.6694	1.3378	0.0853	0.0931
$\phi = -0.9$	0.7366	0.7102	0.5406	0.4717	0.3070	0.2814	0.5559	0.5641	0.0529	0.0559
$\phi = 0$	0.7444	0.4928	0.5293	0.2698	0.3197	0.1983	0.4983	0.4688	0.0825	0.0898
$\phi = 0.9$	0.7301	0.7023	0.5413	0.4613	0.3081	0.2818	0.5516	0.5599	0.0528	0.0557

The benchmark parameter values are $G = 4, N = 100, p = 10, \phi = 0$, and the eigenvectors of the MLE are ordered by the eigenvalues of the first group.
$\hat{\boldsymbol{\Lambda}}_g^* = (\hat{\lambda}_{g1}, \ldots, \hat{\lambda}_{gp})$ and $\mathbf{D}_g^* = (d_{g1}, \ldots, d_{gp})$, where $\hat{\lambda}_{g1}$ and d_{gi} are the ith estimated eigenvalue for group g using MLE and Krzanowski's method, respectively.

Table 3. Simulation results for chi-square distributed data with 10 degrees of freedom.

	$\|\hat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}_1\|$	$\|\mathbf{p}_1 - \boldsymbol{\pi}_1\|$	$\|\hat{\boldsymbol{\pi}}_p - \boldsymbol{\pi}_p\|$	$\|\mathbf{p}_p - \boldsymbol{\pi}_p\|$	$\|\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\|_{F_p}$	$\|\mathbf{P} - \boldsymbol{\Pi}\|_{F_p}$	$\|\hat{\boldsymbol{\Lambda}}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\mathbf{D}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\hat{\boldsymbol{\Lambda}}_g - \text{diag}(\hat{\boldsymbol{\Lambda}}_g)\|_{F_p}^{\text{AVG}}$	$\|\mathbf{D}_g - \text{diag}(\mathbf{D}_g)\|_{F_p}^{\text{AVG}}$
$G = 2$	0.8034	0.6562	0.6082	0.4330	0.3309	0.2670	0.5864	0.5570	0.0655	0.0766
$G = 4$	0.7745	0.5151	0.5479	0.2806	0.3261	0.2042	0.5345	0.5021	0.0859	0.0940
$G = 10$	0.7100	0.3211	0.4341	0.1537	0.3137	0.1206	0.4936	0.4722	0.0979	0.1019
$G = 50$	0.6185	0.1289	0.3590	0.0636	0.3025	0.0445	0.4630	0.4576	0.1042	0.1050
$N = 50$	0.9055	0.6548	0.7445	0.4022	0.3575	0.2585	0.7737	0.7107	0.1180	0.1315
$N = 100$	0.7745	0.5151	0.5479	0.2806	0.3261	0.2042	0.5345	0.5021	0.0859	0.0940
$N = 500$	0.4664	0.2528	0.2534	0.1221	0.2313	0.0985	0.2364	0.2311	0.0402	0.0425
$N = 1{,}000$	0.3527	0.1760	0.1799	0.0847	0.1884	0.0680	0.1684	0.1662	0.0287	0.0302
$N = 10{,}000$	0.1235	0.0550	0.0561	0.0265	0.0784	0.0206	0.0544	0.0543	0.0092	0.0096
$p = 5$	0.4840	0.2712	0.3391	0.1717	0.3073	0.1474	0.3755	0.3683	0.0819	0.0896
$p = 10$	0.7745	0.5151	0.5479	0.2806	0.3261	0.2042	0.5345	0.5021	0.0859	0.0940
$p = 20$	1.0308	0.7991	0.8382	0.4727	0.2748	0.2219	0.8038	0.7010	0.0859	0.0945
$p = 50$	1.2630	1.1215	1.2349	0.8619	0.1912	0.1811	1.7166	1.3741	0.0862	0.0942
$\phi = -0.9$	0.7657	0.7000	0.5551	0.4350	0.3150	0.2733	0.5783	0.5787	0.0649	0.0691
$\phi = 0$	0.7745	0.5151	0.5479	0.2806	0.3261	0.2042	0.5345	0.5021	0.0859	0.0940
$\phi = 0.9$	0.7643	0.6894	0.5461	0.4447	0.3143	0.2736	0.5806	0.5813	0.0647	0.0689

The benchmark parameter values are $G = 4$, $N = 100$, $p = 10$, $\phi = 0$, and the eigenvectors of the MLE are ordered by the eigenvalues of the first group. $\hat{\boldsymbol{\Lambda}}_g^* = (\hat{\lambda}_{g1}, \ldots, \hat{\lambda}_{gp})$ and $\mathbf{D}_g^* = (d_{g1}, \ldots, d_{gp})$, where $\hat{\lambda}_{g1}$ and d_{gi} are the ith estimated eigenvalue for group g using MLE and Krzanowski's method, respectively.

Table 4. Simulation results for chi-square distributed data with 2 degrees of freedom.

	$\|\hat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}_1\|$	$\|\mathbf{p}_1 - \boldsymbol{\pi}_1\|$	$\|\hat{\boldsymbol{\pi}}_p - \boldsymbol{\pi}_p\|$	$\|\mathbf{p}_p - \boldsymbol{\pi}_p\|$	$\|\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\|_{F_p}$	$\|\mathbf{P} - \boldsymbol{\Pi}\|_{F_p}$	$\|\hat{\boldsymbol{\Lambda}}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\mathbf{D}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\hat{\boldsymbol{\Lambda}}_g - \text{diag}(\hat{\boldsymbol{\Lambda}}_g)\|_{F_p}^{\text{AVG}}$	$\|\mathbf{D}_g - \text{diag}(\mathbf{D}_g)\|_{F_p}^{\text{AVG}}$
$G = 2$	0.8700	0.7139	0.6732	0.4851	0.3437	0.2860	0.7298	0.6923	0.0729	0.0871
$G = 4$	0.8493	0.5719	0.6093	0.3210	0.3416	0.2261	0.6745	0.6286	0.0971	0.1078
$G = 10$	0.8010	0.3797	0.5126	0.1789	0.3336	0.1416	0.6224	0.5853	0.1116	0.1176
$G = 50$	0.7160	0.1506	0.4160	0.0743	0.3260	0.0521	0.5836	0.5627	0.1195	0.1214
$N = 50$	0.9686	0.7111	0.8014	0.4511	0.3675	0.2796	0.9841	0.9030	0.1318	0.1490
$N = 100$	0.8493	0.5719	0.6093	0.3210	0.3416	0.2261	0.6745	0.6286	0.0971	0.1078
$N = 500$	0.5230	0.2924	0.2920	0.1422	0.2525	0.1145	0.2927	0.2845	0.0465	0.0495
$N = 1{,}000$	0.4051	0.2065	0.2164	0.1000	0.2082	0.0807	0.2081	0.2048	0.0333	0.0352
$N = 10{,}000$	0.1383	0.0638	0.0710	0.0305	0.0887	0.0241	0.0667	0.0665	0.0107	0.0111
$p = 5$	0.5680	0.3367	0.4146	0.2157	0.3503	0.1807	0.5072	0.4930	0.0978	0.1092
$p = 10$	0.8493	0.5719	0.6093	0.3210	0.3416	0.2261	0.6745	0.6286	0.0971	0.1078
$p = 20$	1.0752	0.8323	0.8692	0.5158	0.2784	0.2307	0.9505	0.8239	0.0926	0.1029
$p = 50$	1.2751	1.1319	1.2425	0.8856	0.1914	0.1820	1.8893	1.4998	0.0892	0.0980
$\phi = -0.9$	0.8207	0.7669	0.6068	0.4881	0.3284	0.2938	0.7276	0.7242	0.0716	0.0762
$\phi = 0$	0.8493	0.5719	0.6093	0.3210	0.3416	0.2261	0.6745	0.6286	0.0971	0.1078
$\phi = 0.9$	0.8167	0.7590	0.6049	0.4881	0.3280	0.2931	0.7297	0.7267	0.0714	0.0760

The benchmark parameter values are $G = 4, N = 100, p = 10, \phi = 0$, and the eigenvectors of the MLE are ordered by the eigenvalues of the first group.
$\hat{\boldsymbol{\Lambda}}_g^* = (\hat{\lambda}_{g1}, \ldots, \hat{\lambda}_{gp})$ and $\mathbf{D}_g^* = (d_{g1}, \ldots, d_{gp})$, where $\hat{\lambda}_{g1}$ and d_{gi} are the ith estimated eigenvalue for group g using MLE and Krzanowski's method, respectively.

estimator still performs better than the ML estimator for all parameter combinations.

The simulation study also investigates how well the covariance matrices are simultaneously diagonalized. No surprising results are found: the diagonalization using both methods are working better with increasing sample sizes and worse with increasing number of groups and dimensions. In the presence of autocorrelation the diagonalization works better and it can be argued that when autocorrelation is present the covariance matrices are more similar to each other. There is no big difference between the two estimation methods but ML consistently diagonalizes the covariance matrices slightly better.

It can be expected that the ordering of the estimated CPCs is influencing the simulation result, thus it is interesting and important to investigate how the estimators performs if they are ordered according to the same rule. The results obtained when the ML estimator too is ordered by the eigenvalues of $\bar{\mathbf{S}}$ are presented in the Appendix (Tables A.10–A.12). This reordering of the ML estimator increased its accuracy and it seems preferable to order the eigenvectors based on the eigenvalues of the arithmetic mean sample covariance matrix rather than the eigenvalues of the first group. The conclusions of this simulation study are the same as above except that the autocorrelation now have a clear negative effect on the ML estimation of the eigenvalues and eigenvectors. One important observation is that for large sample sizes (roughly $N > 500$) the ML estimate of Π performs better than Krzanowski's estimate, most likely due to a slow convergence rate of the ML method. Although, the difference is small and both methods work well.

6. Empirical analysis

CPCA is conducted for two datasets. The first dataset contains Swedish municipality level innovation data collected for 18 variables over 11 years, and the other dataset is one of the most used dataset in multivariate statistics, namely the Iris flower dataset, consisting of four variables collected for three different species of Iris flowers. The Iris data is not the most dynamic one but has yet become almost classic in demonstrating multivariate methods and is readably available for any researcher (through standard software or the literature). It is include here for the purpose of comparison and supplementing of Flury (1984) and Krzanowski (1984) who used this dataset as an example of a practical application of CPCA.

Each year in the innovation data is treated as a group and the CPCA is conducted across years. In the Iris data, each species constitutes a group: thus, the CPCA is conducted between the different species of Iris flowers.

The two datasets serve the purpose of comparing the (dis)similarity of the two estimation methods. Although, this could be addressed with simulations, there

is always a pedagogical dimension in using real data since the data structure cannot be adapted in any sense.

6.1. Data and variables

The innovation dataset consists of annual Swedish municipality level data related to innovations, collected from 2000 to 2010. The dataset were used by Holgersson and Kekezi (2017) as they developed a multi-dimensional innovation index, able to reflect innovation enablers and outputs. PCA and CPCA are methods based on correlation or covariance matrices which in turn are sensitive to outliers. The municipality of Stockholm was removed from the analysis since the city is of substantially larger size than other Swedish cities. All other municipalities are included. For the years 2000–2002 a total of 289 municipalities existed in Sweden; however, since 2003, there have been 290 municipalities. Normality was assessed via Q-Q plots and histograms. The variables are presented in Table 5. All the variables except for *average age* and *related variety*, were log-transformed to ensure at least approximately normal marginal distributions.

Since the dataset consist of yearly data there is some time dependency present. The matter of time dependence is highly interesting, although, no theoretical results are (to my knowledge) available. The nature of the CPC model lies in the simultaneous diagonalization, and hence the possible time dependence should occur in the eigenvalues, leaving the eigenvector space unaffected. Incorporation of eigenvalue time dependence into the CPC model is no doubt a challenging problem and overlooked in this empirical example. However, time dependency is considered in the simulation study in the previous section.
The Iris flower dataset was collected by Edgar Anderson and first published in Anderson (1935). Fisher (1936) used it in his article *the use of multiple measurements in taxonomic problem* as an example of linear discriminant analysis and it has since then been used by several other authors.

The dataset consists of four variables; (1) sepal length, (2) sepal width, (3) petal length, and (4) petal width, and includes 50 samples from each of three Iris species (Iris Setosa, Iris Virginica, and Iris Versicolor).

6.2. Common principal component analysis

For the analysis of the innovation data, each year of municipal data constitutes a group, and the analysis is conducted across 11 years (2000–2010). The model selection procedure is based on the AIC, BIC, and LR statistics and presented in Table 6, where the value marked with a star (*) indicate the best fit.

The CPC(q) models evaluated for this dataset are based on the correlation matrix of the CPCs (described in Sec. A.2 in the Appendix) for each group. For the CPC assumption to be valid, the correlation matrices must be close

Table 5. List of the innovation related variables included in the analysis.

	Variables
Human capital	Average age of the population
	Number of people with compulsory education <9 years
	Number of people with compulsory education 9–10 years
	Number of people with high school education <2 years
	Number of people with high school education 3 years
Tax levels	State property tax (thousands of SEK)
Externalities	Diversity[a]
	Related variety[b]
	Herfindahl-Hirschman index[c]
	Firm density[d]
Firm activities	Mean establishment size of firms[e]
	Export (SEK)
Firm performance	Average social security contributions (thousands of SEK)
	Average wage costs (thousands of SEK)
	Average net sales (thousands of SEK)
	Average gross investment (thousands of SEK)
Accessibility	Number of commuting people
Regional performance	Gross regional product (millions of SEK)

The following descriptions of the variables are taken from Holgersson and Kekezi (2017):
[a] *Diversity* is measured by the total number of different firms on a five-level NACE industry code.
[b] *Related variety* is calculated the same way as in Frenken, Van Oort, and Verburg (2007). It is the weighted sum of five-digit levels within each two-digit group which measures variety within sectors that comes from regional knowledge spillovers. The underlying idea is that innovation comes from the recombination of existing knowledge. Frenken, Van Oort, and Verburg (2007) argue that this is the best estimate of Jacobs externalities.
[c] The *Herfindahl-Hirschman index* is a measure of concentration. It ranges from many small firms (value close to zero) to a single monopolistic producer (value of one). A high value means that employment is concentrated in only a few sectors while a lower one shows an even distribution of employees between sectors and hence more diversification (Essletzbichler 2007).
[d] *Firm density* is calculated as the number of firms divided with the total number of people living in the municipality.
[e] *Mean establishment size of firm* is a measure of the proportion between the number of employees in the municipality and the number of firms.

to the identity matrix. The different CPC(q) models are based on stepwise deeming one of the CPCs with the highest correlations to be non-common, until all components with correlations exceeding an absolute value of 0.3 are considered non-common throughout all groups. The component of the two correlated components that accounts for the least amount of variation is omitted. The absolute correlation value of 0.3 is arbitrary chosen but is argued not to be high from a practical point of view, e.g., Flury (1988, 99) stated the following regarding a highest CPC correlation of −0.23: "*This correlation is almost negligible from a practical point of view, indicating that the CPC model fits the data well.*" In the CPC(14) model, all components with correlations higher than an absolute values of 0.3 are removed, thus the components in the CPC(14) model all have neglectable correlations indicating that the model indeed fits the data well and the more complicated CPC(q) models (i.e., $q < 14$) is therefore not included.

The model selected by the goodness-of-fit measures is the same regardless of which estimation technique is used. The LR statistic selects the CPC model as the best fit, AIC selects the partial CPC model with 14 common components (CPC(14)) as the best fit and BIC selects the proportional model as the best fit.

Table 6. Model selection for the innovation data ($G = 11, p = 18$).

Model		χ^2	df	$\frac{\chi^2}{df}$	AIC[a]	BIC[a]
Higher	Lower					
MLE CPC						
Equality	Proportionality	100.22	10	10.02	3,395.36	3,395.36
Proportionality	CPC	522.44	170	3.07	3,315.14	3,375.77∗
CPC	CPC(16)	12.12	10	1.21∗	3,132.70	4,224.10
CPC(16)	CPC(15)	27.35	20	1.37	3,140.57	4,292.61
CPC(15)	CPC(14)	90.34	30	3.01	3,153.22	4,426.53
CPC(14)	Unrelated	2,642.88	1,470	1.80	3,122.88∗	4,578.09
Unrelated	–				3,420	27,576.75
Equality	Unrelated	3,395.36	1,710			
Krzanowski's CPC approximation						
Equality	Proportionality	100.22	10	10.02	3,395.36	3,395.36
Proportionality	CPC	468.28	170	2.75	3,315.14	3,375.77∗
CPC	CPC(16)	11.76	10	1.18∗	3,186.86	4,278.27
CPC(16)	CPC(15)	27.46	20	1.37	3,195.10	4,347.14
CPC(15)	CPC(14)	94.98	30	3.17	3,207.64	4,480.95
CPC(14)	Unrelated	2,692.66	1,470	1.83	3,172.66∗	4,627.87
Unrelated	–				3,420	27,576.75
Equality	Unrelated	3,395.36	1,710			

The group-specific portions of the CPC(q) models were selected based on correlations (described in Sec. A.2 in the Appendix) between the CPCs (see Table A.1 in the Appendix). CPC(16) omitted components 15 and 17, CPC(15) omitted components 9, 15, and 17, and, CPC(14) omitted components 9, 13, 15, and 17.
[a] AIC and BIC values for the higher model.

Tables A.5 and A.6 in the Appendix list the ML estimated CPC coefficients (ordered according to (3.10)) as well as the difference between the two estimation techniques. The difference in estimated eigenvalues is almost nonexistent and presented in Table A.7. To present the differences more clearly, the eigenvalue differences are multiplied by 100, and only integers are shown. The two estimation techniques produce very similar results.

The Iris dataset analysis is conducted across the three different Iris flowers species. The model selection procedure is summarized in Table 7. For both estimation methods, the AIC selects the unrelated model and BIC selects the CPC(2) model as the best fit. The LR statistic selects the CPC(2) model as the best fit for ML estimation and the proportionality model for Krzanowski's estimation method. However, for Krzanowski's method, the LR statistic for the proportionality model (value 4.29) and the CPC(2) model (value 4.70) are close to each other. Note that the unrelated model is not testable with the LR statistic, which possibly overlooks the best-fitted model indicated by the AIC.

The ML-estimated coefficients are presented in Table A.8 as well as the differences ($\times 100$) in coefficient estimation. The estimated eigenvalues are presented in Table A.9 in the Appendix. Some non-trivial differences in the CPC estimation can be seen, probably because the Iris data is not suitable for the CPC model as can be seen from the correlation coefficients of the CPCs presented in Tables A.1 and A.2 in the Appendix.

Table 7. Model selection for the Iris data ($G = 3, p = 4$).

Model		χ^2	df	$\frac{\chi^2}{df}$	AIC[a]	BIC[a]
Higher	Lower					
MLE CPC						
Equality	Proportionality	34.34	2	17.17	146.66	146.66
Proportionality	CPC	48.41	6	8.07	116.32	122.34
CPC	CPC(2)	39.53	2	19.77	79.91	104.00
CPC(2)	Unrelated	24.38	10	2.44∗	44.38	74.48∗
Unrelated	–				40.00∗	200.43
Equality	Unrelated	146.66	20			
Krzanowski's CPC approximation						
Equality	Proportionality	34.34	2	17.17	146.66	146.66
Proportionality	CPC	25.71	6	4.29∗	116.32	122.34
CPC	CPC(2)	39.58	2	19.79	102.61	126.69
CPC(2)	Unrelated	47.03	10	4.70	67.03	97.14∗
Unrelated	–				40.00∗	200.43
Equality	Unrelated	146.66	20			

[a] AIC and BIC values for the higher model.

Table 8. The effectiveness of the simultaneous diagonalization for the two datasets measured using the Frobenius norm.

Dataset	$\|\hat{\mathbf{\Lambda}}_g - \text{diag}(\hat{\mathbf{\Lambda}}_g)\|_{Fp}^{AVG}$	$\|\mathbf{D}_g - \text{diag}(\mathbf{D}_g)\|_{Fp}^{AVG}$
Municipality data	0.0416	0.0409
Iris data	0.6236	0.7321

The effectiveness of the simultaneous diagonalization for the two datasets are measured using the Frobenius norm and displayed in Table 8. The methods diagonalize the municipality data equally well but the ML estimation diagonalize the Iris data better than Krzanowski's method. The correlation coefficients among CPCs for the Iris data are larger for Krzanowski's estimates, possibly indicating that the ML estimation performs better for this particular dataset.

7. Summary

The aim of this article was to compare two different estimation methods of the CPC model, namely the ML estimation (Flury 1984; Flury and Gautschi 1986; Flury 1988) and Krzanowski's estimation method (Krzanowski 1984). The methods were compared for two real-world datasets and in a Monte Carlo simulation study. The two real-world datasets were; (i) annual Swedish municipality level data related to innovations, collected from 2000 to 2010, and (ii) the Iris flower dataset. AIC, BIC, and LR statistics were used to facilitate the model selection procedure for both methods. The Monte Carlo simulation study investigated the performance of the methods under different combinations of the number of groups, sample sizes, and variables from multivariate normal and chi-square distributions, with and without the presence of autocorrelation.

The performance was evaluated by a p-normalized Frobenius norm and the Euclidean norm.

In the Monte Carlo study, both estimation methods performed best with normally distributed data, and they deteriorated as the data became more asymmetric. Expectedly, the accuracies of both methods improved with increasing sample sizes and groups, and worsened with increasing dimensions. Autocorrelation between the groups had a negative effect on the estimators. The ordering of the eigenvectors affected the result in the simulation study to a large extent. When the ML-estimated eigenvectors too were ordered by the magnitude of the eigenvalues of the arithmetic mean sample covariance matrix rather than according to the eigenvalues of the first group, the accuracy heavily increased. This result indicates that it is preferable to order the eigenvectors according to the eigenvalues of the arithmetic mean sample covariance matrix rather than those of the first group suggested by Flury. Krzanowski's estimator of the eigenvectors and eigenvalues outperformed the ML estimator for all parameter combinations when the sample size was smaller than roughly 500 (with $p = 10$). For larger sample sizes (keeping $p = 10$) the ML estimator preformed better, indicating a slow convergence of the ML estimator. However, the difference between the two methods for large sample sizes were small and both methods performed well. The ML estimator did a slightly better job of simultaneously diagonalizing the covariance matrices, although, the difference was very minor.

Both methods produced very similar results for the Swedish innovation data; the estimated coefficients, eigenvalues, and selection criteria only exhibited trivial differences. For the Iris data, AIC and BIC indicated the same model as the best fit regardless of the estimation method, while the LR statistic indicated different models depending on the estimation procedure. The estimated coefficients displayed nontrivial differences, possibly due to the data not having a clear CPC structure across groups.

To conclude, Krzanowski's estimation method is simple, readily available, requires significantly less computational power, and outperformed the ML estimation method in all simulations when the sample size was lower. Even if the ML estimator performed better in high sample sizes the difference was small. The simplicity and intuition of Krzanowski's estimation method, the computational power required for ML and the findings in this article support and promote the use of Krzanowski's estimation method for CPC models over ML in situations where the CPC assumption is appropriate.

References

Akaike, H. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." In *2nd International Symposium on Information Theory*, 267–81.

Anderson, E. 1935. "The Irises of the Gaspe Peninsula." *Bulletin of American Iris Society* 59:2–5.

Anderson, T. W. 1963. "Asymptotic Theory for Principal Component Analysis." *The Annals of Mathematical Statistics* 34 (1):122–48.

Browne, R. P., and P. D. McNicholas. 2014. "Orthogonal Stiefel Manifold Optimization for Eigen-Decomposed Covariance Parameter Estimation in Mixture Models." *Statistics and Computing* 24 (2):203–10.
Essletzbichler, J. 2007. "Diversity, Stability and Regional Growth in the United States, 1975–2002." In *Applied Evolutionary Economics and Economic Geography*, edited by Koen Frenken, 203. Cheltenham: Edward Elgar.
Fisher, R. A. 1936. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Human Genetics* 7 (2):179–88.
Flury, B. N. 1984. "Common Principal Components in *k* Groups." *Journal of the American Statistical Association* 79 (388):892–8.
Flury, B. N. 1986. "Asymptotic Theory for Common Principal Component Analysis." *The Annals of Statistics* 14 (2):418–30.
Flury, B. N. 1988. *Common Principal Components & Related Multivariate Models*. New York: Wiley.
Flury, B. N., and W. Gautschi. 1986. "An Algorithm for Simultaneous Orthogonal Transformation of Several Positive Definite Symmetric Matrices to Nearly Diagonal Form." *SIAM Journal on Scientific and Statistical Computing* 7 (1):169–84.
Frenken, K., F. Van Oort, and T. Verburg. 2007. "Related Variety, Unrelated Variety and Regional Economic Growth." *Regional Studies* 41 (5):685–97.
Girshick, M. 1939. "On the Sampling Theory of Roots of Determinantal Equations." *The Annals of Mathematical Statistics* 10 (3):203–24.
Holgersson, T. 2005. "Simulation of Non-Normal Autocorrelated Variables." *Journal of Modern Applied Statistical Methods* 5 (2):15.
Holgersson, T., and O. Kekezi. 2017. "Towards a Multivariate Innovation Index." *Economics of Innovation and New Technology* 27 (3):254–72.
Jolliffe, I. 2002. *Principal Component Analysis*. Hoboken, NJ: Wiley Online Library.
Krzanowski, W. 1984. "Principal Component Analysis in the Presence of Group Structure." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 33 (2):164–8.
Lawley, D. 1953. "A Modified Method of Estimation in Factor Analysis and Some Large Sample Results." In *Uppsala Symposium on Psychological Factor Analysis*, Vol. 17, 35–42. Taylor & Francis.
Lawley, D. 1956. "Tests of Significance for the Latent Roots of Covariance and Correlation Matrices." *Biometrika* 43 (1/2):128–36.
Pepler, P. T. 2014. *The Identification and Application of Common Principal Components*. PhD thesis, Stellenbosch University, Stellenbosch.
Pepler, P. T., D. Uys, and D. Nel. 2016. "A Comparison of Some Methods for the Selection of a Common Eigenvector Model for the Covariance Matrices of Two Groups." *Communications in Statistics—Simulation and Computation* 45 (8):2917–36.
Pillai, K., S. Al-Ani, and G. Jouris. 1969. "On the Distributions of the Ratios of the Roots of a Covariance Matrix and Wilks' Criterion for Tests of Three Hypotheses." *The Annals of Mathematical Statistics* 40 (6):2033–40.
Rao, C. R. 1973. *Linear Statistical Inference and Its Applications*, Vol. 2. New York: Wiley.
Schwarz, G. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6 (2):461–4.

A. Appendix

A.1. The algorithm for proportional covariance matrices

The following iterative algorithm was proposed by Flury (1988) to solve the likelihood equations for proportional covariance matrices.

Step 1: Define $\boldsymbol{\rho}^0 = (1, \rho_2, \ldots, \rho_G)'$ as an initial approximation and let $r_g = \frac{n_g}{n}$, where $n = n_1 + \cdots + n_G$, in this article $\rho_g = 1$ for $g = 1, \ldots, G$, is used.

Step 2: Put

$$\mathbf{S} \leftarrow \sum_{g=1}^{G} \frac{r_g \mathbf{S}_g}{\rho_g}, \tag{A.1}$$

$$\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_p \leftarrow \text{normalized eigenvectors of } \mathbf{S}, \tag{A.2}$$

$$a_{gj} \leftarrow \boldsymbol{\pi}_j' \mathbf{S}_g \boldsymbol{\pi}_j, \quad g = 1, \ldots, G, \quad j = 1, \ldots p. \tag{A.3}$$

Step 3: Put

$$\lambda_j \leftarrow \sum_{g=1}^{G} \frac{r_g a_{gj}}{\rho_g}, \quad j = 1, \ldots, p. \tag{A.4}$$

Step 4: Put

$$\rho_g^* \leftarrow \frac{1}{p} \sum_{j=1}^{p} \frac{a_{gj}}{\lambda_g}, \quad g = 1, \ldots, G, \tag{A.5}$$

and

$$\rho_g \leftarrow \frac{\rho_g^*}{\rho_1^*} \quad g = 1, \ldots, G. \tag{A.6}$$

Step 5: Put

$$\boldsymbol{\rho}^t \leftarrow (1, \rho_2, \ldots, \rho_G)'. \tag{A.7}$$

Steps 2–5 are repeated until $\|\boldsymbol{\rho}^t - \boldsymbol{\rho}^{t-1}\|_* < \epsilon$ for some predetermined vector norm, where $\boldsymbol{\rho}^t$ is the obtained $\boldsymbol{\rho}$ for the tth iteration. When convergence is reached, the current values of ρ_g, λ_j, and $\boldsymbol{\pi}_j$ are the ML estimates. Flury (1988) used the absolute value of the largest element as the vector norm with $\epsilon = 10^{-4}$. In this study, the same vector norm is used and the value of ϵ is set to 10^{-6}.

A.2. Sample covariances and correlations of sample CPCs

The sample common principal components is defined as

$$\mathbf{Y}_g = \mathbf{X}_g \hat{\Pi}, \quad g = 1, \ldots, G, \tag{A.8}$$

with the associated sample covariance matrix $\mathbf{F}_g$ defined as

$$\mathbf{F}_g = \hat{\Pi}' \mathbf{S}_g \hat{\Pi}, \quad g = 1, \ldots, G. \tag{A.9}$$

At times it is more beneficial to observe the correlation matrices,

$$\mathbf{R}_g = \hat{\Lambda}_g^{-1/2} \mathbf{F}_g \hat{\Lambda}_g^{-1/2}, \quad g = 1, \ldots G, \tag{A.10}$$

where $\hat{\Lambda}_g = \text{diag}\left(\mathbf{F}_g\right)$. For the CPC assumption to hold, $\mathbf{R}_g$ must be close to $\mathbf{I}_p$ (Flury 1984).

For the innovation data, the correlations above an absolute value of 0.3 for both estimation methods are presented in Tables A.1 and A.2, respectively. The correlation matrix for the Iris flower data is presented in its entirety in Tables A.3 and A.4.

A.3. Tables

Table A.1. Correlations with absolute values higher than 0.3 for the innovation data using the ML estimated CPC model.

Year	Components	Correlation
2000	(1,15)	0.32
2001	(1,15)	0.35
2001	(1,9)	−0.32
2001	(4,13)	−0.32
2009	(1,17)	0.41
2010	(1,17)	0.45

Table A.2. Correlations with absolute values higher than 0.3 for the innovation data using the Krzanowski's estimated CPC model.

Year	Components	Correlation
2000	(1,15)	0.33
2001	(1,15)	0.37
2001	(1,9)	−0.32
2001	(4,13)	−0.31
2009	(1,17)	0.30
2009	(1,15)	−0.31
2010	(1,17)	0.34
2010	(1,15)	−0.35

Table A.3. Correlation matrix for the Iris flower data using the ML estimated CPC model.

Component	Flower type	Correlation			
1	Versicolor	1			
2		0.18	1		
3		−0.07	−0.06	1	
4		0.10	0.07	−0.14	1
1	Virginica	1			
2		−0.08	1		
3		0.11	−0.27	1	
4		−0.15	0.39	−0.35	1
1	Setosa	1			
2		−0.09	1		
3		−0.74	0.01	1	
4		−0.05	−0.25	0.12	1

Table A.4. Correlation matrix for the Iris flower data using the Krzanowski's estimation CPC model.

Component	Flower type	Correlation			
1	Versicolor	1			
2		0.2	1		
3		0.07	−0.08	1	
4		0.16	−0.29	0.19	1
1	Virginica	1			
2		−0.01	1		
3		0.3	−0.25	1	
4		−0.02	0.23	−0.29	1
1	Setosa	1			
2		−0.47	1		
3		−0.71	0.43	1	
4		−0.18	−0.15	0.36	1

Table A.5. The first nine ML-estimated CPC coefficients for the innovation data, consisting of 18 variables and 11 groups (years 2006–2010).

	1	2	3	4	5	6	7	8	9
1. Average age	0.576	0.732	−0.270	0.000	0.235	0.014	0.025	−0.012	0.033
	(0)	(0)	(1)	(0)	(0)	(0)	(0)	(1)	(−1)
2. Compulsory education <9 years	−0.184	0.194	−0.167	0.105	−0.175	0.036	−0.398	−0.405	−0.397
	(0)	(0)	(0)	(0)	(0)	(−5)	(2)	(−3)	(0)
3. Compulsory education 9–10 years	−0.241	0.108	−0.190	0.098	−0.036	0.007	−0.190	−0.159	−0.073
	(0)	(0)	(0)	(0)	(0)	(−3)	(1)	(−1)	(0)
4. High school education <2 years	−0.216	0.137	−0.216	0.112	−0.134	0.058	−0.260	−0.083	0.041
	(0)	(0)	(0)	(0)	(0)	(−3)	(−1)	(−2)	(1)
5. High school education 3 years	−0.264	0.120	−0.220	0.125	−0.033	0.027	−0.121	0.081	0.077
	(0)	(0)	(1)	(0)	(0)	(−1)	(−1)	(0)	(0)
6. Property tax	−0.297	0.063	−0.291	0.051	0.241	0.047	0.386	0.421	−0.590
	(0)	(0)	(1)	(1)	(0)	(5)	(−3)	(3)	(−3)
7. Diversity	−0.105	0.070	−0.120	0.041	−0.041	0.008	−0.010	0.094	−0.007
	(0)	(0)	(0)	(0)	(0)	(0)	(−1)	(−1)	(2)
8. Related variety	−0.099	0.040	−0.205	0.015	−0.100	−0.001	0.027	0.220	0.108
	(0)	(0)	(0)	(0)	(0)	(1)	(−2)	(−2)	(7)
9. Herfindahl-Hirschman index	0.094	−0.095	0.135	−0.081	0.313	0.381	−0.657	0.515	−0.050
	(0)	(0)	(0)	(−1)	(1)	(−9)	(−9)	(−4)	(0)
10. Firm density	0.023	−0.001	−0.063	−0.035	−0.038	0.033	0.172	0.115	0.031
	(0)	(0)	(0)	(0)	(0)	(3)	(−1)	(0)	(3)
11. Mean establishment size of firms	−0.042	0.085	0.119	0.127	−0.081	−0.088	−0.158	0.045	0.153
	(0)	(0)	(0)	(0)	(0)	(−2)	(1)	(0)	(−3)
12. Export	−0.399	0.458	0.358	−0.705	−0.045	0.020	0.036	0.016	0.023
	(0)	(−1)	(−1)	(−1)	(0)	(0)	(−1)	(0)	(0)
13. Avg. social security contributions	−0.087	0.156	0.353	0.290	0.150	−0.187	0.002	−0.018	−0.118
	(0)	(0)	(1)	(−1)	(−1)	(0)	(2)	(0)	(5)
14. Average wage costs	−0.070	0.143	0.311	0.288	0.127	−0.253	−0.037	0.070	−0.059
	(0)	(0)	(1)	(0)	(−1)	(0)	(3)	(−1)	(0)
15. Average net sales	−0.080	0.151	0.332	0.306	0.170	−0.265	−0.032	0.066	−0.080
	(0)	(0)	(1)	(0)	(−1)	(0)	(3)	(−1)	(−1)
16. Average gross investment	−0.051	0.144	0.291	0.313	−0.081	0.815	0.278	−0.171	0.003
	(0)	(0)	(1)	(−2)	(3)	(5)	(−9)	(1)	(0)
17. Commuting	−0.283	−0.105	−0.168	−0.038	0.783	0.058	0.042	−0.366	0.313
	(0)	(0)	(0)	(0)	(1)	(−2)	(1)	(1)	(0)
18. Gross regional product	−0.280	0.215	−0.135	0.255	−0.168	−0.018	0.068	0.331	0.565
	(0)	(0)	(1)	(0)	(0)	(2)	(−1)	(2)	(−3)

The difference in coefficient estimation (×100) between ML and Krzanowski's estimation is displayed in parentheses.

Table A.6. The last six ML-estimated CPC coefficients for the innovation data, consisting of 18 variables and 11 groups (years 2000–2010).

	10	11	12	13	14	15	16	17	18
1. Average age	−0.017	0.031	−0.016	0.008	−0.022	−0.016	−0.003	−0.002	0.004
	(0)	(0)	(0)	(0)	(0)	(0)	(1)	(0)	(0)
2. Compulsory education <9 years	0.027	−0.339	0.188	−0.391	0.243	−0.012	−0.106	−0.053	0.004
	(4)	(−1)	(−2)	(0)	(−2)	(−1)	(−7)	(0)	(0)
3. Compulsory education 9–10 years	0.014	−0.050	−0.007	0.285	−0.830	−0.162	−0.063	−0.101	0.048
	(0)	(0)	(0)	(0)	(−5)	(−2)	(21)	(4)	(−1)
4. High school education <2 years	−0.013	0.259	−0.210	0.378	0.172	0.562	−0.011	0.410	−0.102
	(−2)	(0)	(1)	(0)	(1)	(11)	(−4)	(−12)	(2)
5. High school education 3 years	−0.021	0.215	−0.098	0.298	0.392	−0.714	0.107	−0.056	−0.003
	(−1)	(0)	(1)	(−1)	(3)	(−4)	(−11)	(17)	(0)
6. Property tax	−0.225	0.058	−0.028	−0.128	−0.033	0.072	−0.034	0.087	−0.010
	(7)	(0)	(0)	(0)	(−1)	(2)	(1)	(−1)	(0)
7. Diversity	0.215	−0.059	0.088	0.025	0.016	0.245	0.826	−0.382	0.106
	(1)	(0)	(−1)	(1)	(22)	(−9)	(2)	(−9)	(−1)
8. Related variety	0.767	0.329	0.165	−0.295	−0.067	0.001	−0.243	0.026	−0.027
	(1)	(0)	(−4)	(0)	(−7)	(1)	(1)	(1)	(0)
9. Herfindahl-Hirschman index	0.017	−0.114	−0.010	0.000	−0.024	0.007	−0.033	−0.015	0.008
	(2)	(0)	(0)	(0)	(−1)	(0)	(0)	(0)	(0)
10. Firm density	0.287	−0.629	0.149	0.214	−0.006	−0.156	0.170	0.565	−0.170
	(1)	(0)	(−2)	(0)	(4)	(14)	(2)	(6)	(2)
11. Mean establishment size of firms	−0.276	0.289	0.084	−0.503	−0.231	−0.169	0.382	0.471	−0.164
	(−1)	(−1)	(2)	(−1)	(9)	(12)	(8)	(6)	(2)
12. Export	0.012	0.018	0.016	0.016	0.004	0.002	−0.001	0.001	0.001
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
13. Avg. social security contributions	0.289	−0.144	−0.748	−0.141	−0.020	−0.043	0.039	0.008	−0.012
	(−3)	(1)	(−2)	(1)	(1)	(0)	(1)	(1)	(0)
14. Average wage costs	0.036	0.047	0.354	0.150	0.056	0.029	−0.048	0.200	0.713
	(3)	(0)	(0)	(0)	(−2)	(5)	(−1)	(−2)	(1)
15. Average net sales	0.021	0.049	0.387	0.205	0.039	0.070	−0.069	−0.202	−0.639
	(3)	(0)	(0)	(−1)	(−1)	(−5)	(−2)	(0)	(−1)
16. Average gross investment	0.038	0.087	0.077	0.008	−0.018	−0.014	0.011	0.000	0.004
	(0)	(0)	(−1)	(0)	(0)	(0)	(1)	(0)	(0)
17. Commuting	0.050	−0.025	0.065	−0.117	0.046	0.041	0.030	0.060	−0.002
	(−4)	(0)	(−1)	(0)	(1)	(2)	(−1)	(−1)	(0)
18. Gross regional product	−0.254	−0.357	−0.032	−0.180	0.019	0.143	−0.214	−0.190	0.049
	(−5)	(0)	(3)	(0)	(−5)	(−4)	(−1)	(−4)	(0)

The difference in coefficient estimation (×100) between ML and Krzanowski's estimation is displayed in parentheses.

Table A.7. Estimated eigenvalue difference for the innovation data using the CPC model.

	Eigenvalue difference (×100)										
Eigenvalue	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
1	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	1	1	0	0
4	0	−2	0	0	0	0	0	0	0	−1	0
5	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	0	0	0	−1	0
8	0	0	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0

Table A.8. ML estimated CPC for the Iris data.

	Coefficients of common principal components			
	1	2	3	4
Sepal length	0.737	−0.647	−0.164	0.108
	(0)	(−1)	(−22)	(−12)
Sepal width	0.247	0.466	−0.835	−0.161
	(−7)	(28)	(4)	(16)
Petal length	0.605	0.500	0.522	−0.334
	(3)	(−8)	(6)	(2)
Petal width	0.175	0.338	0.063	0.922
	(2)	(−14)	(22)	(7)

Table A.9. Estimated eigenvalue difference for the Iris flower data using the CPC model.

	Coefficients (×100)		
Eigenvalue	Versicolor	Virginica	Setosa
1	5	77	−156
2	38	55	−61
3	−26	−254	253
4	−17	122	−36

Table A.10. Simulation results for normally distributed data.

	$\|\hat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}_1\|$	$\|\mathbf{p}_1 - \boldsymbol{\pi}_1\|$	$\|\hat{\boldsymbol{\pi}}_p - \boldsymbol{\pi}_p\|$	$\|\mathbf{p}_p - \boldsymbol{\pi}_p\|$	$\|\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\|_{F_p}$	$\|\mathbf{P} - \boldsymbol{\Pi}\|_{F_p}$	$\|\hat{\boldsymbol{\Lambda}}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\mathbf{D}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\hat{\boldsymbol{\Lambda}}_g - \text{diag}(\hat{\boldsymbol{\Lambda}}_g)\|_{F_p}^{\text{AVG}}$	$\|\mathbf{D}_g - \text{diag}(\mathbf{D}_g)\|_{F_p}^{\text{AVG}}$
$G = 2$	0.6695	0.6447	0.4371	0.4182	0.2747	0.2622	0.5407	0.5124	0.0628	0.0735
$G = 4$	0.5289	0.4928	0.2827	0.2698	0.2188	0.1983	0.4983	0.4688	0.0825	0.0898
$G = 10$	0.3365	0.3070	0.1445	0.1460	0.1288	0.1139	0.4576	0.4400	0.0937	0.0972
$G = 50$	0.1287	0.1230	0.0561	0.0609	0.0418	0.0422	0.4306	0.4263	0.0996	0.1003
$N = 50$	0.6894	0.6383	0.4317	0.3893	0.2775	0.2536	0.7229	0.6644	0.1139	0.1264
$N = 100$	0.5289	0.4928	0.2827	0.2698	0.2188	0.1983	0.4983	0.4688	0.0825	0.0898
$N = 500$	0.2368	0.2403	0.1072	0.1157	0.0957	0.0940	0.2204	0.2156	0.0385	0.0407
$N = 1{,}000$	0.1622	0.1713	0.0729	0.0819	0.0622	0.0654	0.1576	0.1557	0.0274	0.0288
$N = 10{,}000$	0.0427	0.0512	0.0215	0.0250	0.0162	0.0196	0.0510	0.0509	0.0088	0.0091
$p = 5$	0.2474	0.2522	0.1502	0.1610	0.1354	0.1353	0.3354	0.3300	0.0763	0.0831
$p = 10$	0.5289	0.4928	0.2827	0.2698	0.2188	0.1983	0.4983	0.4688	0.0825	0.0898
$p = 20$	0.8512	0.7825	0.5263	0.4655	0.2366	0.2194	0.7644	0.6654	0.0839	0.0921
$p = 50$	1.1728	1.1161	0.9453	0.8507	0.1854	0.1808	1.6694	1.3378	0.0853	0.0931
$\phi = -0.9$	0.6854	0.7102	0.4425	0.4717	0.2792	0.2814	0.5559	0.5641	0.0529	0.0559
$\phi = 0$	0.5289	0.4928	0.2827	0.2698	0.2188	0.1983	0.4983	0.4688	0.0825	0.0898
$\phi = 0.9$	0.6789	0.7023	0.4360	0.4613	0.2802	0.2818	0.5516	0.5599	0.0528	0.0557

The benchmark parameter values are $G = 4, N = 100, p = 10, \phi = 0$, and the eigenvectors of the MLE are ordered by the eigenvalues of $\bar{\mathbf{S}}$.
$\hat{\boldsymbol{\Lambda}}_g^* = (\hat{\lambda}_{g1}, \ldots, \hat{\lambda}_{gp})$ and $\mathbf{D}_g^* = (d_{g1}, \ldots, d_{gp})$, where $\hat{\lambda}_{g1}$ and d_{gi} are the ith estimated eigenvalue for group g using MLE and Krzanowski's method, respectively.

Table A.11. Simulation results for chi-square distributed data with 10 degrees of freedom.

	$\|\hat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}_1\|$	$\|\mathbf{p}_1 - \boldsymbol{\pi}_1\|$	$\|\hat{\boldsymbol{\pi}}_p - \boldsymbol{\pi}_p\|$	$\|\mathbf{p}_p - \boldsymbol{\pi}_p\|$	$\|\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\|_{F_p}$	$\|\mathbf{P} - \boldsymbol{\Pi}\|_{F_p}$	$\|\hat{\boldsymbol{\Lambda}}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\mathbf{D}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\hat{\boldsymbol{\Lambda}}_g - \text{diag}(\hat{\boldsymbol{\Lambda}}_g)\|_{F_p}^{\text{AVG}}$	$\|\mathbf{D}_g - \text{diag}(\mathbf{D}_g)\|_{F_p}^{\text{AVG}}$
$G = 2$	0.6843	0.6562	0.4569	0.4330	0.2806	0.2670	0.5864	0.5570	0.0655	0.0766
$G = 4$	0.5535	0.5151	0.2981	0.2806	0.2254	0.2042	0.5345	0.5021	0.0859	0.0940
$G = 10$	0.3652	0.3211	0.1550	0.1537	0.1394	0.1206	0.4936	0.4722	0.0979	0.1019
$G = 50$	0.1412	0.1289	0.0582	0.0636	0.0462	0.0445	0.4630	0.4576	0.1042	0.1050
$N = 50$	0.7086	0.6548	0.4481	0.4022	0.2840	0.2585	0.7737	0.7107	0.1180	0.1315
$N = 100$	0.5535	0.5151	0.2981	0.2806	0.2254	0.2042	0.5345	0.5021	0.0859	0.0940
$N = 500$	0.2507	0.2528	0.1122	0.1221	0.1010	0.0985	0.2364	0.2311	0.0402	0.0425
$N = 1{,}000$	0.1684	0.1760	0.0765	0.0847	0.0660	0.0680	0.1684	0.1662	0.0287	0.0302
$N = 10{,}000$	0.0452	0.0550	0.0226	0.0265	0.0168	0.0206	0.0544	0.0543	0.0092	0.0096
$p = 5$	0.2713	0.2712	0.1667	0.1717	0.1522	0.1474	0.3755	0.3683	0.0819	0.0896
$p = 10$	0.5535	0.5151	0.2981	0.2806	0.2254	0.2042	0.5345	0.5021	0.0859	0.0940
$p = 20$	0.8633	0.7991	0.5323	0.4727	0.2388	0.2219	0.8038	0.7010	0.0859	0.0945
$p = 50$	1.1768	1.1215	0.9585	0.8619	0.1857	0.1811	1.7166	1.3741	0.0862	0.0942
$\phi = -0.9$	0.6831	0.7000	0.4130	0.4350	0.2744	0.2733	0.5783	0.5787	0.0649	0.0691
$\phi = 0$	0.5535	0.5151	0.2981	0.2806	0.2254	0.2042	0.5345	0.5021	0.0859	0.0940
$\phi = 0.9$	0.6784	0.6894	0.4254	0.4447	0.2746	0.2736	0.5806	0.5813	0.0647	0.0689

The benchmark parameter values are $G = 4, N = 100, p = 10, \phi = 0$, and the eigenvectors of the MLE are ordered by the eigenvalues of $\bar{\mathbf{S}}$.
$\hat{\boldsymbol{\Lambda}}_g^* = (\hat{\lambda}_{g1}, \ldots, \hat{\lambda}_{gp})$ and $\mathbf{D}_g^* = (d_{g1}, \ldots, d_{gp})$, where $\hat{\lambda}_{g1}$ and d_{gi} are the ith estimated eigenvalue for group g using MLE and Krzanowski's method, respectively.

Table A.12. Simulation results for chi-square distributed data with 2 degrees of freedom.

	$\|\hat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}_1\|$	$\|\mathbf{p}_1 - \boldsymbol{\pi}_1\|$	$\|\hat{\boldsymbol{\pi}}_p - \boldsymbol{\pi}_p\|$	$\|\mathbf{p}_p - \boldsymbol{\pi}_p\|$	$\|\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\|_{F_p}$	$\|\mathbf{P} - \boldsymbol{\Pi}\|_{F_p}$	$\|\hat{\boldsymbol{\Lambda}}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\mathbf{D}_g^* - \boldsymbol{\Lambda}_g^*\|^{\text{AVG}}$	$\|\hat{\boldsymbol{\Lambda}}_g - \text{diag}(\hat{\boldsymbol{\Lambda}}_g)\|_{F_p}^{\text{AVG}}$	$\|\mathbf{D}_g - \text{diag}(\mathbf{D}_g)\|_{F_p}^{\text{AVG}}$
$G = 2$	0.7509	0.7139	0.5144	0.4851	0.3008	0.2860	0.7298	0.6923	0.0729	0.0871
$G = 4$	0.6364	0.5719	0.3529	0.3210	0.2536	0.2261	0.6745	0.6286	0.0971	0.1078
$G = 10$	0.4534	0.3797	0.1893	0.1789	0.1756	0.1416	0.6224	0.5853	0.1116	0.1176
$G = 50$	0.2250	0.1506	0.0776	0.0743	0.0773	0.0521	0.5836	0.5627	0.1195	0.1214
$N = 50$	0.7796	0.7111	0.5193	0.4511	0.3085	0.2796	0.9841	0.9030	0.1318	0.1490
$N = 100$	0.6364	0.5719	0.3529	0.3210	0.2536	0.2261	0.6745	0.6286	0.0971	0.1078
$N = 500$	0.3005	0.2924	0.1335	0.1422	0.1214	0.1145	0.2927	0.2845	0.0465	0.0495
$N = 1{,}000$	0.2028	0.2065	0.0904	0.1000	0.0814	0.0807	0.2081	0.2048	0.0333	0.0352
$N = 10{,}000$	0.0540	0.0638	0.0268	0.0305	0.0201	0.0241	0.0667	0.0665	0.0107	0.0111
$p = 5$	0.3573	0.3367	0.2165	0.2157	0.2004	0.1807	0.5072	0.4930	0.0978	0.1092
$p = 10$	0.6364	0.5719	0.3529	0.3210	0.2536	0.2261	0.6745	0.6286	0.0971	0.1078
$p = 20$	0.9110	0.8323	0.5758	0.5158	0.2474	0.2307	0.9505	0.8239	0.0926	0.1029
$p = 50$	1.1878	1.1319	0.9613	0.8856	0.1862	0.1820	1.8893	1.4998	0.0892	0.0980
$\phi = -0.9$	0.7662	0.7669	0.4765	0.4881	0.2973	0.2938	0.7276	0.7242	0.0716	0.0762
$\phi = 0$	0.6364	0.5719	0.3529	0.3210	0.2536	0.2261	0.6745	0.6286	0.0971	0.1078
$\phi = 0.9$	0.7532	0.7590	0.4751	0.4881	0.2972	0.2931	0.7297	0.7267	0.0714	0.0760

The benchmark parameter values are $G = 4, N = 100, p = 10, \phi = 0$, and the eigenvectors of the MLE are ordered by the eigenvalues of $\bar{\mathbf{S}}$.
$\hat{\boldsymbol{\Lambda}}_g^* = (\hat{\lambda}_{g1}, \ldots, \hat{\lambda}_{gp})$ and $\mathbf{D}_g^* = (d_{g1}, \ldots, d_{gp})$, where $\hat{\lambda}_{g1}$ and d_{gi} are the ith estimated eigenvalue for group g using MLE and Krzanowski's method, respectively.