Journal of the Nigerian Statistical Association Vol. **28**, 2016, 16– 29

On Refined Principal Component Method for Factor Analysis

D. F. Nwosu[a*], S. I. Onyeagu[b], J.I. Mbegbu[c] and V. U. Ekhosuehi[d]

[a]*Department of Mathematics/Statistics, Federal Polytechnic, Nekede,Owerri, Nigeria*; [b]*Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria*; [c,d]*Department of Mathematics, University of Benin, P.M.B. 1154, Benin City, Nigeria*

This paper is centred on the development of a method, known as the refined Principal Component Method (rPCM), for the construction of the underlying relationships between variables. The proposed method settles the perturbing issue in the literature on the initial assumption of exact dependence of variables on the factors in the classical Principal Component Method (PCM). The development of the rPCM is hinged on matrix splitting. Theoretical aspects of eigenvalues and eigenvectors as it relates to symmetric and commutative matrices are carefully applied. Findings reveal that the rPCM generates results as that of the PCM and gives better factor loadings and communalities in terms of the error matrix and the admissible error than that of the PCM.

Keywords: communalities; factor analysis; factor loadings; Principal Component Method; matrix splitting.

1. Introduction

The ways to simplify a complex data set by representing the set of observable variables in terms of a smaller number of underlying (hypothetical or unobservable) variables, known as factors or latent variables, have remained a subject of interest in multivariate analysis. Such a representation is important as it could aid researchers to explore the covariance or correlation relationships among observed and random variables (Choi, 2010). This subject area is a well known aspect of multivariate analysis called factor analysis (Rencher, 2002). A popular estimation technique in factor analysis is the Principal Component Method (Boik, 2013; Qi et al., 2013; Torohti and Friedland, 2009). The Principal Component Method (PCM) employs the idea of dimension reduction in principal component analysis. The idea is that the first few principal components might retain most of the variation in the data (Shen and Huang, 2008). The PCM assumes, at the initial stage, that there is exact dependence of the variables on the factors, and then proceed to estimate the loadings and communalities using either the covariance matrix or the correlation matrix of the observed variables. This assumption is an unlikely outcome (Rencher, 2002). The exact dependence of variables on the factors is an unlikely outcome because dimension reduction is prone to error which is due to the truncation arising from the use of few factor components.

The study is aimed at refining the PCM with a view to relaxing the assumption of the exact dependence of the variables on the factors. This aim is achieved by the use of matrix splitting. We call the method developed in this study the refined Principal Component Method (rPCM). Earlier on, the number of principal components or factors in a factor model was a challenge. However, this challenge has been addressed considerably in the literature (Parmet et al., 2010). The common methods in the literature for adjudging principal components are: the Kaiser-Guttman (KG) rule of one and the 80% threshold. Even so, the KG rule of one and the 80% threshold applies only to the variance component of the covariance matrix. This is a snag in the use of these existing rules as a yardstick for the explanation provided by the estimated factor model. Consequent upon this, this study

*Corresponding author. Email: fedocon2003@gmail.com

utilises the error matrix and the admissible error as a means of adjudging an estimated factor model.

Without loss of generality, the method developed in this study utilises the covariance matrix. However, this matrix may be replaced by the correlation matrix, depending on the choice of the user. A program is written in MATLAB to ease the implementation of the rPCM. The program is given in the Appendix.

2. Methodology

In this section, we develop the rPCM and present the main results. The rPCM is developed using the principle of matrix splitting. It is well known that matrix splitting is a representation of a given matrix as the sum or difference of matrices (Climent and Perea, 1998; Elsner, 1989; Jedrzejec and Woznicki, 2001; Song, 1991). Woznicki (2001) presented useful results for the linear equation system based on the splitting of a non-singular matrix as the difference of two matrices. The research expanded upon in this direction is considered in this paper. Nonetheless, our method is based on the splitting of a non-singular matrix as the sum of two matrices.

2.1 *Preliminaries*

The factor analysis model is expressed as

$$\mathbf{Y} - \boldsymbol{\mu} = \boldsymbol{\beta}\mathbf{F} + \mathbf{e},$$

where $\mathbf{Y}$ is an $n \times 1$ vector of the observed variables, $\boldsymbol{\mu}$ is the mean of $\mathbf{Y}$, $\boldsymbol{\beta}$ is an $n \times m$ matrix of the factor loadings with $m < n$, $\mathbf{F}$ is an $m \times 1$ vector of the underlying constructs or latent variables that generate $\mathbf{Y}$, and $\mathbf{e}$ is an $n \times 1$ vector of the unobserved error, often called the specific factor. It is assumed that: $\mathbb{E}(\mathbf{F}) = \mathbf{0}$, $cov(\mathbf{F}) = \mathbf{I}$, $cov(\mathbf{F}, \mathbf{e}) = \mathbf{0}$, $\mathbb{E}(\mathbf{e}) = \mathbf{0}$ and $cov(\mathbf{e})$ is heteroscedastic with $\psi_i, i = 1, 2, \cdots, n$, as the specific variance and the covariances are zero. That is

$$cov(\mathbf{e}) = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_n \end{pmatrix} = \boldsymbol{\Psi}$$

(Rencher, 2002). Under these assumptions,

$$cov(\mathbf{Y}) = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Psi}.$$

The task is to estimate the factor loading matrix, $\boldsymbol{\beta}$, and the communalities, $tr(\boldsymbol{\beta}\boldsymbol{\beta}')$. With $\mathbf{A} = cov(\mathbf{Y})$, $\mathbf{M} = \boldsymbol{\beta}\boldsymbol{\beta}'$ and $\mathbf{N} = \boldsymbol{\Psi}$, we have $\mathbf{A} = \mathbf{M}+\mathbf{N}$. The expression $\mathbf{A} = \mathbf{M}+\mathbf{N}$ is the matrix splitting considered in this paper. More specifically, the problem is how to split the matrix $\mathbf{A}$ into the symmetric matrix, $\mathbf{M}$, and the diagonal matrix, $\mathbf{N}$, with non-negative diagonal entries. This problem is non-trivial as the matrix $\mathbf{M}$ should be the product of a rectangular array $\boldsymbol{\beta}$ of order $n \times m$ and its transpose such that the proportion $\frac{tr(\boldsymbol{\beta}\boldsymbol{\beta}')}{tr(\mathbf{A})} > \epsilon$, where ϵ is the threshold for the explained variance.

Let $\mathbf{A} = \mathbf{M} + \mathbf{N}$ be a splitting of $\mathbf{A} \in \mathbb{R}^{n\times n}$, where $\mathbf{A}$ and $\mathbf{M}$ are non-singular matrices. Then

$$\mathbf{A}^{-1} = (\mathbf{M} + \mathbf{N})^{-1} = \left[\mathbf{M}(\mathbf{I} + \mathbf{M}^{-1}\mathbf{N})\right]^{-1}.$$

From the principle of matrix inverse

$$\mathbf{A}^{-1} = (\mathbf{I} + \mathbf{M}^{-1}\mathbf{N})^{-1}\mathbf{M}^{-1}.$$

Thus

$$(\mathbf{I} + \mathbf{M}^{-1}\mathbf{N})\mathbf{A}^{-1} = \mathbf{M}^{-1}.$$

Alternatively $\mathbf{M}^{-1}$ may be obtained from $\mathbf{A}^{-1} = (\mathbf{M} + \mathbf{N})^{-1}$ as follows

$$\mathbf{A}^{-1} = \left[(\mathbf{I} + \mathbf{N}\mathbf{M}^{-1})\mathbf{M}\right]^{-1} = \mathbf{M}^{-1}(\mathbf{I} + \mathbf{N}\mathbf{M}^{-1})^{-1}.$$

So

$$\mathbf{A}^{-1}(\mathbf{I} + \mathbf{N}\mathbf{M}^{-1}) = \mathbf{M}^{-1}$$

It follows that

$$(\mathbf{I} + \mathbf{M}^{-1}\mathbf{N})\mathbf{A}^{-1} = \mathbf{A}^{-1}(\mathbf{I} + \mathbf{N}\mathbf{M}^{-1}) = \mathbf{M}^{-1}.$$

Thus

$$\mathbf{A}^{-1} + \mathbf{M}^{-1}\mathbf{N}\mathbf{A}^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{N}\mathbf{M}^{-1}.$$

Hence

$$\mathbf{M}^{-1}\mathbf{N}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{N}\mathbf{M}^{-1}.$$

2.2 *Determination of the specific variance matrix*

Let the admissible error, denoted as ϑ, be given. Consider the matrix $\mathbf{A}^{-1}\mathbf{N}$. Let $\tau_i, i = 1, 2, \cdots, n$, be the eigenvalues of $\mathbf{A}^{-1}\mathbf{N}$ and γ_i the eigenvalues of the covariance matrix $\mathbf{A}$. Then the determinant of $\mathbf{A}^{-1}\mathbf{N}$ is

$$\left|\mathbf{A}^{-1}\mathbf{N}\right| = \prod_{i=1}^{n} \tau_i.$$

Thus

$$\left|\mathbf{A}^{-1}\right| |\mathbf{N}| = \left(\prod_{i=1}^{n} \gamma_i\right)^{-1} |\mathbf{N}| = \prod_{i=1}^{n} \tau_i.$$

Therefore

$$|\mathbf{N}| = \prod_{i=1}^{n} \gamma_i \tau_i.$$

Since $\mathbf{N}$ is a diagonal matrix, then

$$\mathbf{N} = diag\left(\gamma_1\tau_1, \gamma_2\tau_2, \cdots, \gamma_n\tau_n\right).$$

In practice, an initial guess should be made for the τ_i's before $\mathbf{N}$ can be computed. Let $\mathbf{N} = \mathbf{N}_0$ be the specific variance matrix computed with the initial guess for the τ_i's. When $\tau_i = 0\ \forall i$, then our proposed method (the rPCM) reduces to the PCM as $\mathbf{N} = \mathbf{N}_0 = \mathbf{0}$. Suppose $\tau_i \neq 0$ for at least one i. Then the factor loading vector, β, is estimated with $\mathbf{N}_0$ as the starting specific variance matrix (see Subsection 2.3). With $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, the augmented specific variance matrix, $\tilde{\mathbf{N}}$, is obtained as

$$\tilde{\mathbf{N}} = diag\left(\mathbf{A} - \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right).$$

As mentioned earlier, this study utilise the error matrix and the admissible error as a means of adjudging the estimated factor model. This is because the inferences from the KG rule of one and the 80% threshold may be misleading as these rules apply only to the variance component of the covariance matrix (see Section 3). Let

$$\mathbf{H} = \mathbf{A} - \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'$$

be the error matrix. We define the sum of squared error (the approximation error hereafter), which is due to the use of the m eigenvectors of the principal component to construct $\hat{\boldsymbol{\beta}}$ as

$$AE = (\mathbf{H} \bullet \mathbf{H})\,\mathbf{1}.$$

where the operation $\bullet$ is used to indicate that the matrix product is done element-wise. The notation $\mathbf{1}$ is used to denote an $n \times 1$ vector of ones. If any entry in $\tilde{\mathbf{N}}$ is negative or $AE \geq \vartheta\mathbf{1}$, then new values should be assigned to the eigenvalues τ_i's and $\hat{\boldsymbol{\beta}}$ should be recomputed. This step is repeated until an admissible solution, wherein the entries in $\tilde{\mathbf{N}}$ are all nonnegative and $AE < \vartheta\mathbf{1}$, is obtained.

2.3 *Estimation of the factor loadings*

Suppose the number of principal components or factors, m, in the underlying data generating model is known. Then consider the matrix $\mathbf{M}^{-1}\mathbf{N}$. Let λ_i, $i = 1, 2, \cdots, n$, be the eigenvalues of $\mathbf{M}^{-1}\mathbf{N}$. Since $\mathbf{M}^{-1}\mathbf{N}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{N}\mathbf{M}^{-1}$, the matrices $\mathbf{M}^{-1}\mathbf{N}$ and $\mathbf{A}^{-1}\mathbf{N}$ are commutative and their eigenvectors are the same. Let $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n)$ be the eigenvector with $\mathbf{e}_1$ corresponding to τ_1, $\mathbf{e}_2$ corresponding to τ_2, and so on. Then $\mathbf{M}^{-1}\mathbf{N}\mathbf{E} = \lambda\mathbf{E}$ and $\mathbf{A}^{-1}\mathbf{N}\mathbf{E} = \tau\mathbf{E}$. Since $\mathbf{M}^{-1} = (\mathbf{I} + \mathbf{M}^{-1}\mathbf{N})\mathbf{A}^{-1}$, then $(\mathbf{I} + \mathbf{M}^{-1}\mathbf{N})\mathbf{A}^{-1}\mathbf{N}\mathbf{E} = \lambda\mathbf{E}$. Thus

$$(\mathbf{I} + \mathbf{M}^{-1}\mathbf{N})\tau\mathbf{E} = \lambda\mathbf{E}$$

so that

$$\tau\mathbf{E} + \tau\mathbf{M}^{-1}\mathbf{N}\mathbf{E} = \lambda\mathbf{E}.$$

Therefore

$$(\tau + \tau\lambda)\mathbf{E} = \lambda\mathbf{E}.$$

Hence

$$\lambda = \frac{\tau}{1 - \tau}.$$

This last expression indicates that any initial guess for τ should be different from one. Let $\mathbf{D} = diag\,(\lambda_1, \lambda_2, \cdots, \lambda_n)$. Then, by the eigenvalue matrix decomposition,

$$\mathbf{M}^{-1}\mathbf{N} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1}.$$

Thus

$$\mathbf{M}^{-1}(\mathbf{A} - \mathbf{M}) = \mathbf{E}\mathbf{D}\mathbf{E}^{-1}.$$

It follows that

$$\mathbf{M}^{-1}\mathbf{A} = \mathbf{I} + \mathbf{E}\mathbf{D}\mathbf{E}^{-1}$$

so that

$$\mathbf{M} = \mathbf{A}(\mathbf{I} + \mathbf{E}\mathbf{D}\mathbf{E}^{-1})^{-1}.$$

Suppose $\mathbf{M}$ has eigenvalues v_1, $v_2, \cdots$, v_n with the corresponding eigenvectors $\mathbf{c}_1$, $\mathbf{c}_2, \cdots$, $\mathbf{c}_n$. Then the n eigenvectors of $\mathbf{M}$ may be normalised and inserted as columns of a matrix $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n)$ such that $\mathbf{C}\mathbf{C}' = \mathbf{I}$. This is possible because the eigenvectors of a symmetric matrix are mutually orthogonal (Rencher, 2002). With $\mathbf{V} = diag\,(v_1, v_2, \cdots, v_n)$, we express the matrix $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{C}\mathbf{V}\mathbf{C}^{-1} = \mathbf{C}\mathbf{V}\mathbf{C}^{-1}\mathbf{C}\mathbf{C}' = \mathbf{C}\mathbf{V}\mathbf{C}' = \mathbf{C}\mathbf{V}^{1/2}\mathbf{V}^{1/2}\mathbf{C}' = (\mathbf{C}\mathbf{V}^{1/2})(\mathbf{C}\mathbf{V}^{1/2})',$$

where $\mathbf{V}^{1/2} = diag\left(\sqrt{v_1}, \sqrt{v_2}, \cdots, \sqrt{v_n}\right)$. Thus $\boldsymbol{\beta}\boldsymbol{\beta} = (\mathbf{C}\mathbf{V}^{1/2})(\mathbf{C}\mathbf{V}^{1/2})'$. But $\mathbf{C}\mathbf{V}^{1/2}$ is a matrix of order $n \times n$, whereas we seek a matrix $\boldsymbol{\beta}$ of order $n \times m$ with $n > m$. For this reason, we define $\mathbf{V}^{1/2}_{n\times m} = diag\left(\sqrt{v_1}, \sqrt{v_2}, \cdots, \sqrt{v_m}\right)$, wherein $v_1 > v_2 > \cdots > v_m$, and its corresponding eigenvectors as $\mathbf{C}_{n\times m} = (\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_m)$. Hence

$$\hat{\boldsymbol{\beta}} = \mathbf{C}_{n\times m}\mathbf{V}^{1/2}_{n\times m} = \left(\sqrt{v_1}\mathbf{c}_1, \sqrt{v_2}\mathbf{c}_2, \cdots, \sqrt{v_m}\mathbf{c}_m\right).$$

2.4 *The factor scores*

Let $\mathbf{f}_j = (f_{j1}, f_{j2}, \cdots, f_{jm})'$, $j = 1, 2, \cdots, N$, be the jth vector of factor scores with N being the size of the jth vector of the observed variable. We adopt the centred regression method discussed in Rencher (2002) to estimate the factor scores as

$$\hat{\mathbf{F}} = \mathbf{Y}_c\mathbf{A}^{-1}\hat{\boldsymbol{\beta}},$$

where

$$\mathbf{Y}_c = \begin{pmatrix} (\mathbf{y}_1 - \bar{\mathbf{y}})' \\ (\mathbf{y}_2 - \bar{\mathbf{y}})' \\ \vdots \\ (\mathbf{y}_n - \bar{\mathbf{y}})' \end{pmatrix}$$

is an $N \times n$ matrix of the deviations from the mean.

3. Numerical Illustration

To demonstrate the utility of the rPCM vis-a-vis the PCM, we consider some examples in the literature (Choi, 2010; Rencher, 2002). All computations in this section are done in the MATLAB environment.[1]

Example 1: Choi (2010) showed that the covariance matrix

$$\begin{bmatrix} 7 & 4 & 15 & 6 \\ 4 & 8 & 12 & -3 \\ 15 & 12 & 50 & 18 \\ 6 & -3 & 18 & 43 \end{bmatrix}$$

can be exactly split into two terms as

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 6 \\ -4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & -4 \\ 2 & 1 & 6 & 5 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

The proportion of variance explained by this exact splitting is 0.8889. This means that the splitting presented by Choi (2010) explains about 88.9% of the variation in the covariance matrix.

Suppose we set $\vartheta = 0.1$. Then, using the PCM, the factor loadings matrix is obtained as

$$\boldsymbol{\beta}_{PCM} = \begin{bmatrix} 0.8060 & 2.0721 \\ 1.9881 & 1.1677 \\ 2.6800 & 6.5037 \\ -4.5731 & 4.6842 \end{bmatrix}$$

the proportion of variance explained is 0.95, the error matrix is

$$\mathbf{H}_{PCM} = \begin{bmatrix} 0 & -0.0222 & -0.6367 & -0.0203 \\ -0.0222 & 0 & -0.9228 & 0.6220 \\ -0.6367 & -0.9228 & 0 & -0.2091 \\ -0.0203 & 0.6220 & -0.2091 & 0 \end{bmatrix}$$

and the approximation error vector is

$$AE_{PCM} = \begin{bmatrix} 0.4063 \\ 1.2390 \\ 1.3007 \\ 0.4311 \end{bmatrix}.$$

Although the proportion of variance explained by the PCM is higher than that of the exact split by Choi (2010), the approximation error of the PCM is higher than that of Choi (2010). Moreover, all entries in AE_{PCM} exceed the admissible error, $\vartheta = 0.1$. This is an indication that the KG rule of one should be taken with a pitch of salt.

[1]It is important to mention here that results from the MATLAB program should be looked upon carefully as they may be inaccurate if any of the eigenvalues within the number of principal components of interest are negative.

However, with $\tau_1 = 0.95$, $\tau_2 = 0.9$, $\tau_3 = 0.155$, $\tau_4 = 0$, the proportion of explained variance is 0.9062, the factor loadings matrix is

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 2.0227 & -0.9465 \\ 1.0707 & -1.9534 \\ 6.0703 & -2.8418 \\ 4.9696 & 4.2774 \end{bmatrix},$$

the error matrix is

$$\mathbf{H}_{rPCM} = \begin{bmatrix} 0 & -0.0147 & 0.0318 & -0.0032 \\ -0.0147 & 0 & -0.0509 & 0.0344 \\ 0.0318 & -0.0509 & 0 & -0.0112 \\ -0.0032 & 0.0344 & -0.0112 & 0 \end{bmatrix},$$

and the approximation error vector is

$$AE_{rPCM} = \begin{bmatrix} 0.0012 \\ 0.0040 \\ 0.0037 \\ 0.0013 \end{bmatrix}.$$

Clearly the rPCM outperform the PCM as $AE_{rPCM} < 0.1\mathbf{1} < AE_{PCM}$. Nonetheless, the variance explained by the PCM is closer to one than that of the rPCM. All the same, the choice of eigenvalues, $\tau_1 = 0.95$, $\tau_2 = 0.9$, $\tau_3 = 0.155$, $\tau_4 = 0$, yields factor loadings that closely approximate the covariance matrix.

Example 2: Brown, Williams, and Barlow (1984) as cited by Rencher (2002) had reported the perception rating by an individual using five adjectives on a 9-point semantic differential scale for seven people. The adjectives were kind, intelligent, happy, likeable and just. Rencher (2002) computed the correlation matrix for the perception data as

$$\begin{bmatrix} 1.000 & 0.296 & 0.881 & 0.995 & 0.545 \\ 0.296 & 1.000 & -0.022 & 0.326 & 0.837 \\ 0.881 & -0.022 & 1.000 & 0.867 & 0.130 \\ 0.995 & 0.326 & 0.867 & 1.000 & 0.544 \\ 0.545 & 0.837 & 0.130 & 0.544 & 1.000 \end{bmatrix}.$$

Rencher (2002) opined that the five adjectives may be explained by two factors as the adjectives within the sets $\{1, 3, 4\}$ and $\{2, 5\}$ are highly correlated. The factor loadings and communalities were estimated using the PCM and the correlation matrix was approximated as

$$\begin{bmatrix} 1.000 & 0.296 & 0.881 & 0.995 & 0.545 \\ 0.296 & 1.000 & -0.022 & 0.326 & 0.837 \\ 0.881 & -0.022 & 1.000 & 0.867 & 0.130 \\ 0.995 & 0.326 & 0.867 & 1.000 & 0.544 \\ 0.545 & 0.837 & 0.130 & 0.544 & 1.000 \end{bmatrix} \cong$$

$$\begin{bmatrix} 0.969 & -0.231 \\ 0.519 & 0.807 \\ 0.785 & -0.587 \\ 0.971 & -0.210 \\ 0.704 & 0.667 \end{bmatrix} \begin{bmatrix} 0.969 & 0.519 & 0.785 & 0.971 & 0.704 \\ -0.231 & 0.807 & -0.587 & -0.210 & 0.667 \end{bmatrix}$$

$$+\begin{bmatrix} 0.007 & 0 & 0 & 0 & 0 \\ 0 & 0.079 & 0 & 0 & 0 \\ 0 & 0 & 0.040 & 0 & 0 \\ 0 & 0 & 0 & 0.013 & 0 \\ 0 & 0 & 0 & 0 & 0.060 \end{bmatrix}.$$

The proportion of variance that was accounted for was 0.96. We compute the error matrix and the approximation error vector as

$$\mathbf{H}_{PCM} = \begin{bmatrix} 0 & -0.0215 & -0.0157 & 0.0057 & 0.0163 \\ -0.0215 & 0 & 0.0446 & -0.0087 & -0.0665 \\ -0.0157 & 0.0446 & 0 & -0.0183 & -0.0303 \\ 0.0057 & -0.0087 & -0.0183 & 0 & 0.0006 \\ 0.0163 & -0.0665 & -0.0303 & 0.0006 & 0 \end{bmatrix}$$

and

$$AE_{PCM} = \begin{bmatrix} 0.0010 \\ 0.0070 \\ 0.0035 \\ 0.0004 \\ 0.0056 \end{bmatrix},$$

respectively. With $\tau_1 = 0$, $\tau_2 = 1.2$, $\tau_3 = 0$, $\tau_4 = 0$, $\tau_5 = 0$, the correlation matrix is approximated as

$$\begin{bmatrix} 1.000 & 0.296 & 0.881 & 0.995 & 0.545 \\ 0.296 & 1.000 & -0.022 & 0.326 & 0.837 \\ 0.881 & -0.022 & 1.000 & 0.867 & 0.130 \\ 0.995 & 0.326 & 0.867 & 1.000 & 0.544 \\ 0.545 & 0.837 & 0.130 & 0.544 & 1.000 \end{bmatrix} \cong$$

$$\begin{bmatrix} 0.9707 & 0.2261 \\ 0.5116 & -0.7939 \\ 0.7865 & 0.5866 \\ 0.9718 & 0.2054 \\ 0.7026 & -0.6750 \end{bmatrix} \begin{bmatrix} 0.9707 & 0.5116 & 0.7865 & 0.9718 & 0.7026 \\ 0.2261 & -0.7939 & 0.5866 & 0.2054 & -0.6750 \end{bmatrix}$$

$$+\begin{bmatrix} 0.0067 & 0 & 0 & 0 & 0 \\ 0 & 0.1081 & 0 & 0 & 0 \\ 0 & 0 & 0.0374 & 0 & 0 \\ 0 & 0 & 0 & 0.0134 & 0 \\ 0 & 0 & 0 & 0 & 0.0507 \end{bmatrix}.$$

The proportion of explained variance is 0.9567. The error matrix and the approximation error vector are obtained respectively as

$$\mathbf{H}_{rPCM} = \begin{bmatrix} 0 & -0.0211 & -0.0150 & 0.0053 & 0.0156 \\ -0.0211 & 0 & 0.0413 & -0.0081 & -0.0583 \\ -0.0150 & 0.0413 & 0 & -0.0178 & -0.0266 \\ 0.0053 & -0.0081 & -0.0178 & 0 & -0.0001 \\ 0.0156 & -0.0583 & -0.0266 & -0.0001 & 0 \end{bmatrix}$$

and

$$AE_{rPCM} = \begin{bmatrix} 0.0009 \\ 0.0056 \\ 0.0030 \\ 0.0004 \\ 0.0044 \end{bmatrix}.$$

Again $AE_{rPCM} < AE_{PCM}$, even though that the variance explained by the PCM is closer to one than that of the rPCM. The results from the approximate error vector show that it is possible to make an initial guess on the eigenvalues so as to obtain factor loadings that closely approximate the correlation matrix.

Example 3: The table below shows seven observations collected for five components.

Observation	$\mathbf{y}_1$	$\mathbf{y}_2$	$\mathbf{y}_3$	$\mathbf{y}_4$	$\mathbf{y}_5$
1	1	5	5	1	1
2	8	9	7	9	8
3	9	8	9	9	8
4	9	9	9	9	9
5	1	9	1	1	9
6	9	7	7	9	9
7	9	7	9	9	7

The covariance matrix is obtained as

$$\mathbf{A} = \begin{bmatrix} 14.619 & 1.691 & 9.857 & 14.857 & 5.976 \\ 1.691 & 2.238 & -0.095 & 1.905 & 3.595 \\ 9.857 & -0.095 & 8.571 & 9.905 & 1.095 \\ 14.857 & 1.905 & 9.905 & 15.238 & 6.095 \\ 5.976 & 3.595 & 1.095 & 6.095 & 8.238 \end{bmatrix}.$$

Using the PCM, the factor loadings are obtained as

$$\boldsymbol{\beta}_{PCM} = \begin{bmatrix} -0.1598 & 3.8149 \\ 1.2237 & 0.5298 \\ -1.3526 & 2.5455 \\ -0.1150 & 3.8869 \\ 2.3270 & 1.6552 \end{bmatrix},$$

the proportion of variance explained is 0.98, the error matrix is

$$\mathbf{H}_{PCM} = \begin{bmatrix} 0 & -0.1350 & -0.0697 & 0.0104 & 0.0337 \\ -0.1350 & 0 & 0.2114 & -0.0137 & -0.1290 \\ -0.0697 & 0.2114 & 0 & -0.1449 & 0.0295 \\ 0.0104 & -0.0137 & -0.1449 & 0 & -0.0706 \\ 0.0337 & -0.1290 & 0.0295 & -0.0706 & 0 \end{bmatrix}$$

and the approximation error vector is

$$AE_{PCM} = \begin{bmatrix} 0.0243 \\ 0.0797 \\ 0.0714 \\ 0.0263 \\ 0.0236 \end{bmatrix}.$$

With the choice of eigenvalues, $\tau_1 = 0.8$, $\tau_2 = 0.1$, $\tau_3 = 0.05$, $\tau_4 = 0.01$, $\tau_5 = 0.001$, we obtain

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} -3.8118 & -0.1611 \\ -0.5294 & 1.2222 \\ -2.5420 & -1.3497 \\ -3.8783 & -0.1138 \\ -1.6528 & 2.3217 \end{bmatrix},$$

the proportion of variance explained as 0.98, the error matrix is

$$\mathbf{H}_{rPCM} = \begin{bmatrix} 0 & -0.1306 & -0.0498 & 0.0555 & 0.0500 \\ -0.1306 & 0 & 0.2086 & -0.0093 & -0.1175 \\ -0.0498 & 0.2086 & 0 & -0.1074 & 0.0273 \\ 0.0555 & -0.0093 & -0.1074 & 0 & -0.0506 \\ 0.0500 & -0.1175 & 0.0273 & -0.0506 & 0 \end{bmatrix}$$

and the approximation error vector is

$$AE_{rPCM} = \begin{bmatrix} 0.0251 \\ 0.0745 \\ 0.0583 \\ 0.0173 \\ 0.0196 \end{bmatrix}.$$

Apart from the first entry in the approximation error vector, the rPCM gives a better approximation of the covariance matrix, $\mathbf{A}$, than the PCM. This is also the conclusion when the matrices $\mathbf{H}_{PCM}$ and $\mathbf{H}_{rPCM}$ are compared. Clearly that of the $\mathbf{H}_{rPCM}$ largely contains smaller entries than the $\mathbf{H}_{PCM}$.

Example 4: Consider the correlation matrix $\mathbf{R}$ defined in terms of the submatrices:

$$\mathbf{R}_{11} = \begin{bmatrix} 1.00 & 0.56 & 0.22 & 0.10 & 0.20 \\ 0.56 & 1.00 & -0.09 & 0.13 & 0.20 \\ 0.22 & -0.09 & 1.00 & 0.16 & 0.70 \\ 0.10 & 0.13 & 0.16 & 1.00 & 0.49 \\ 0.20 & 0.20 & 0.70 & 0.49 & 1.00 \end{bmatrix},$$

$$\mathbf{R}_{12} = \begin{bmatrix} -0.04 & 0.13 & 0.03 & -0.07 & 0.09 \\ -0.17 & 0.17 & 0.24 & 0.16 & 0.06 \\ -0.31 & -0.45 & -0.34 & -0.11 & 0.68 \\ -0.03 & -0.16 & 0.01 & 0.42 & 0.37 \\ -0.32 & -0.34 & -0.19 & 0.30 & 0.87 \end{bmatrix},$$

and

$$\mathbf{R}_{22} = \begin{bmatrix} 1.00 & -0.42 & -0.57 & -0.11 & -0.26 \\ -0.42 & 1.00 & 0.82 & 0.23 & -0.30 \\ -0.57 & 0.82 & 1.00 & 0.45 & -0.17 \\ -0.11 & 0.23 & 0.45 & 1.00 & 0.29 \\ -0.26 & -0.30 & -0.17 & 0.29 & 1.00 \end{bmatrix},$$

as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{12} & \mathbf{R}_{22} \end{bmatrix}.$$

The eigenvalues of $\mathbf{R}$ are computed as 3.1721, 2.5597, 1.4352, 1.2725, 0.5368, 0.4726, 0.2591, 0.1276, 0.0962 and 0.0681. The first four eigenvalues are larger than the remaining six eigenvalues. This indicates that we can extract four factors from the ten variables in $\mathbf{R}$. With the PCM, the factor loadings are obtained as

$$\boldsymbol{\beta}_{PCM} = \begin{bmatrix} 0.1072 & -0.8314 & 0.2894 & 0.2166 \\ 0.2913 & -0.6527 & 0.5102 & 0.1050 \\ -0.4317 & -0.0806 & -0.1035 & 0.8246 \\ 0.5570 & 0.2095 & 0.2577 & 0.5140 \\ -0.0126 & 0.0415 & 0.2130 & 0.9360 \\ 0.6274 & -0.0917 & -0.6699 & -0.1827 \\ -0.1654 & 0.0060 & 0.7268 & -0.5474 \\ -0.1136 & 0.1686 & 0.8621 & -0.3852 \\ 0.4875 & 0.4513 & 0.5897 & 0.1772 \\ -0.0886 & 0.1586 & 0.1587 & 0.8838 \end{bmatrix}.$$

The factor loadings represent 84.4% of the explained variance. The approximation error vector is

$$AE_{PCM} = \begin{bmatrix} 0.0466 \\ 0.0552 \\ 0.0165 \\ 0.0673 \\ 0.0071 \\ 0.0297 \\ 0.0227 \\ 0.0042 \\ 0.0440 \\ 0.0245 \end{bmatrix}.$$

Taking $\tau_1 = 0.85$, $\tau_2 = 0.6$, $\tau_3 = 0.5$, $\tau_4 = 0.1$, $\tau_5 = 0.009$, $\tau_6 = 0.02$, $\tau_7 = 0.01$, $\tau_8 = 0.006$,

$\tau_9 = 0.004$, $\tau_{10} = 0.001$, we obtain

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} -0.2105 & -0.2779 & -0.8103 & 0.1506 \\ -0.1009 & -0.4938 & -0.6314 & 0.3162 \\ -0.8111 & 0.0959 & -0.1007 & -0.4122 \\ -0.5125 & -0.2584 & 0.2267 & 0.5380 \\ -0.9273 & -0.2153 & 0.0316 & -0.0174 \\ 0.1792 & 0.6733 & -0.0498 & 0.6295 \\ 0.5513 & -0.7236 & -0.0102 & -0.1560 \\ 0.3896 & -0.8621 & 0.1537 & -0.1134 \\ -0.1769 & -0.5925 & 0.4703 & 0.4699 \\ -0.8872 & -0.1654 & 0.1494 & -0.1007 \end{bmatrix}.$$

These factor loadings explain 83.2% of the variance and the associated approximation error vector is

$$AE_{rPCM} = \begin{bmatrix} 0.0371 \\ 0.0455 \\ 0.0138 \\ 0.0621 \\ 0.0048 \\ 0.0277 \\ 0.0213 \\ 0.0035 \\ 0.0401 \\ 0.0241 \end{bmatrix}.$$

In this example, $AE_{rPCM} < AE_{rPCM}$, which indicates that the rPCM has a smaller sum of squared error than the PCM. This means that the rPCM approximates the correlation matrix, $\mathbf{R}$, better than the PCM.

4. Conclusion

The PCM have found application in the reduction of dimensionality of multivariate observations using either the covariance matrix or the correlation matrix. However, this method has been criticised owing to the assumption that there is exact dependence of the variables on the factors. We show that this assumption may be relaxed by developing the method, known as the refined Principal Component Method (rPCM). The theoretical underpinnings for the rPCM are hinged on matrix splitting and the eigenvalues and eigenvectors as it relates to symmetric and commutative matrices. The use of our new method indicated that it is possible to make an initial guess on the eigenvalues (other than zero) to obtain factor loadings that closely approximate the covariance (or correlation) matrix. Thus the rPCM is more flexible than the PCM. Nonetheless, more research is needed in order to investigate how unique initial eigenvalues can be determined.

References

Boik, R. J. (2013). Model-based principal components of correlation matrices. *Journal of Multivariate Analysis*, 116: 310 – 331.

Choi, J. H. (2010). Penalized Maximum Likelihood Factor Analysis. *PhD Thesis*, University of Minnesota.

Climent, J.-J. and Perea, C. (1998). Some comparison theorems for weak nonnegative splittings of bounded operators. *Linear Algebra and its Applications*, 275/276: 77–106.

Elsner, L. (1989). Comparisons of weak regular splitting and multisplitting methods. *Numerische Mathematik* 56(2-3): 283 – 289.

Jedrzejec, H. A. and Woznicki, Z. I. (2001). On properties of some matrix splitting. *Electronic Journal of Linear Algebra* 8: 47–52.

Parmet, Y., Schechtman, E. and Sherman, M. (2010). Factor analysis revisited – How many factors are there? *Communications in Statistics – Simulation and Computation*, 39: 1893 – 1908.

Qi, X., Luo, R. and Zhao, H. (2013). Sparse principal component analysis by choice of norm. *Journal of Multivariate Analysis*, 114: 127 – 160.

Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2nd Ed.). John Wiley & Sons Inc., New York.

Shen, H. and Huang, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, 99: 1015 – 1034.

Song, Y. Z. (1991). Comparisons of nonnegative splittings of matrices. *Linear Algebra and its Applications*, 154/156: 433–455.

Torokhti, A. and Friedland, S. (2009). Towards theory of generic principal component analysis. *Journal of Multivariate Analysis*, 100: 661 – 669.

Woznicki, Z. (2001). Matrix splitting principles. *International Journal of Mathematics and Mathematical Sciences*, 28(5): 251 – 284.

Appendix

The MATLAB Source Codes

```
function rPCM(Y,p) Y = Y(:,:); %Observe variable over time in matrix form.
J=input('Enter the assumed eigenvalues for the inverse covariance-specific variance product');
K=input('Enter the choice of symmetric matrix: 1 for correlation matrix & 2 for covariance matrix');

[r,s] =size(Y);

if K==1;
  A=corrcoef(Y),
  Yc=(Y-ones(r,1)* mean(Y))./(ones(r,1)* var(Y));
else
  A=cov(Y),
  Yc=(Y-ones(r,1)* mean(Y));
end

[m,n] =size(A);
```

```
T=diag([J]);
[E0,V0]=eig(A);
N0=V0.* T;

L=inv(A)* N0;
[E1,V1]=eig(L);

D=diag([diag(V1./(eye(n)-V1))]);

M=A* inv(eye(n)+E1* D* inv(E1));

[E2,V2]=eig(M),

if V2(1,1)>V2(2,2)>V2(3,3)
   C=E2(:, [1:p]);
   D=V2([1:p], [1:p]);
else
   C=E2(:, [n-p+1:n]);
   D=V2([n-p+1:n], [n-p+1:n]);
end

B=C* sqrt(D),
N=diag([diag(A-B* B')]),

CovY=B* B'+N,
Proportion=sum(diag(B* B'))/sum(diag(A)),

ErrorMatrix=A-CovY,
Er=sum(ErrorMatrix.^ 2,2),

F=Yc* inv(A)* B,
```