Principal Component Analysis

Abstract

This methodological framework analyzes the principal component analysis model for reducing the size of indicators. In particular, in principal component analysis, the size of several correlated indicators is reduced by linear substitution while maintaining the existing dispersion to the maximum extent. This article was prepared in order to develop a practice for combining indicators into a single general indicator while maintaining the characteristics of the indicators in drawing general conclusions about financial changes.

Keywords: determinant, component, covariance matrix, correlation, vector.

Prepared by the Financial Stability Department of the Central Bank of Uzbekistan.

If you have any comments, please send them to udjumanazarov@cbu.uz.

Principal Component Analysis

Principal component analysis (PCA) involves reducing the dimension of multiple indicators with mutual correlation by linear transformation while retaining the maximum possible variance[1].

Several common components may cause changes in a large number of indicators. Also, the relationship between components and indicators is expressed as follows[2]:

$$x_{it} = \lambda_i F_t + e_{it}$$
$$x_{it} = C_{it} + e_{it}$$

Where, x_{it} are indicators, λ_i is the eigenvalue between component and indicator, F is the principal component, e_{it} is the variance attributed to each indicator, C_{it} is the common component of the model, i is the number of indicators, t is time period.

For instance, principal component analysis is performed as follows for three conditional indicators (x_1, x_2, x_3). These indicators are first standardized[3] and then converted to matrix form:

$$A = \begin{vmatrix} X_{11} & X_{21} & X_{31} \\ X_{12} & X_{22} & X_{32} \\ X_{13} & X_{23} & X_{33} \\ X_{14} & X_{24} & X_{34} \\ X_{15} & X_{25} & X_{35} \end{vmatrix}$$

Where, X_{it} is a standardized value of indicator i in period t.

For these standardized indicators, the covariance matrix with square[4] and symmetric matrix[5] roperties is calculated as follows:

---

[1] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning.

[2] Petrovska, M., & Mihajlovska, E. M. (2013). Measures of Financial Stability in Macedonia. Journal of Central Banking Theory and Practice.

[3] The historical mean is subtracted from the indicator value, and then the result is divided by the standard deviation.

[4] A matrix with an equal number of columns and rows.

[5] A matrix with symmetric arrangement of matrix elements relative to the main diagonal.

$$\text{Covariance matrix} = \begin{vmatrix} cov(X_1, X_1) & cov(X_1, X_2) & cov(X_1, X_3) \\ cov(X_2, X_1) & cov(X_2, X_2) & cov(X_2, X_3) \\ cov(X_3, X_1) & cov(X_3, X_2) & cov(X_3, X_3) \end{vmatrix}$$

$$cov(X_i, X_j) = \frac{\sum_{t=1}^{T}((X_{it} - \overline{X_{\iota t}}) \times (X_{jt} - \overline{X_{Jt}}))}{T-1};$$

Where, X_{it} is the standardized value of indicator i in period t, X_{jt} is the standardized value of indicator j in period t, $\overline{X_{\iota t}}$ and $\overline{X_{\iota t}}$ are the corresponding average values of the standardized indicators X_i and X_{jt} during the observation period, T is the number of observations.

In the subsequent step, the eigenvalue and eigenvectors for the covariance matrix are calculated. The eigenvalues and eigenvectors of the covariance matrix have the following relationship:

$$B.v = \lambda.v$$

Where, B is the covariance matrix, v is an eigenvector, λ is an eigenvalue.

If it is possible to choose a number λ, such that any nonzero vector v satisfying the equality $B.v = \lambda.v$, then v is called an eigenvector of a linear transformation. And λ is referred to as the eigenvalue corresponding to the eigenvector of the linear transformation[6].

The indicators on the right-hand side of the equation are moved to the left-hand side, resulting in equation $(B - \lambda).v = 0$. Taking into consideration the fact that the eigenvector v is a nonzero, the values of λ are determined by calculating the following determinant equation. Additionally, λ is multiplied by the unit matrix:

$$det(B - \lambda.I) = 0$$

$$det\left(\begin{pmatrix} cov(X_1, X_1) & cov(X_1, X_2) & cov(X_1, X_3) \\ cov(X_2, X_1) & cov(X_2, X_2) & cov(X_2, X_3) \\ cov(X_3, X_1) & cov(X_3, X_2) & cov(X_3, X_3) \end{pmatrix} - \lambda \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = 0$$

$$det\begin{pmatrix} cov(X_1, X_1) - \lambda & cov(X_1, X_2) & cov(X_1, X_3) \\ cov(X_2, X_1) & cov(X_2, X_2) - \lambda & cov(X_2, X_3) \\ cov(X_3, X_1) & cov(X_3, X_2) & cov(X_3, X_3) - \lambda \end{pmatrix} = 0$$

[6] Kuttler, K. (2022). Linear Algebra, Theory and Applications.

The determinant above is used to calculate the eigenvalues[7], and the corresponding eigenvectors are then calculated. In this case, equality is used to express the relationship between the eigenvalues and eigenvectors of the covariance matrix:

$$B.v = \lambda.v$$

$$\begin{pmatrix} cov(X_1,X_1) & cov(X_1,X_2) & cov(X_1,X_3) \\ cov(X_2,X_1) & cov(X_2,X_2) & cov(X_2,X_3) \\ cov(X_3,X_1) & cov(X_3,X_2) & cov(X_3,X_3) \end{pmatrix} \times \begin{pmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{pmatrix} = \lambda_i \times \begin{pmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{pmatrix}$$

$$\begin{cases} cov(X_1,X_1) \times v_{1i} + cov(X_1,X_2) \times v_{2i} + cov(X_1,X_3) \times v_{3i} = \lambda_i \times v_{1i} \\ cov(X_2,X_1) \times v_{1i} + cov(X_2,X_2) \times v_{2i} + cov(X_2,X_3) \times v_{3i} = \lambda_i \times v_{2i} \\ cov(X_3,X_1) \times v_{1i} + cov(X_3,X_2) \times v_{2i} + cov(X_3,X_3) \times v_{3i} = \lambda_i \times v_{3i} \end{cases}$$

Through the system of equations given above, the values of eigenvectors corresponding to each eigenvalue are determined. The values of the determined eigenvectors are arranged in the matrix in descending order. Specifically, the eigenvector corresponding to the largest eigenvalue is positioned in the first column of the matrix. This ordering signifies that the first column represents the largest variance in the initial indicator data.

The initial standardized indicators are linearly transformed by multiplying them with the defined eigenvectors:

$$\begin{pmatrix} X_{11} & X_{21} & X_{31} \\ X_{12} & X_{22} & X_{32} \\ X_{13} & X_{23} & X_{33} \\ X_{14} & X_{24} & X_{34} \\ X_{15} & X_{25} & X_{35} \end{pmatrix} \times \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \\ F_{41} & F_{42} & F_{43} \\ F_{51} & F_{52} & F_{53} \end{pmatrix}$$

$$\begin{cases} F_{11} = X_{11} \times v_{11} + X_{21} \times v_{21} + X_{31} \times v_{31} \\ F_{12} = X_{11} \times v_{12} + X_{21} \times v_{22} + X_{31} \times v_{32} \\ F_{13} = X_{11} \times v_{13} + X_{21} \times v_{23} + X_{31} \times v_{33} \\ \qquad \qquad \dots \\ F_{53} = X_{51} \times v_{13} + X_{52} \times v_{23} + X_{53} \times v_{33} \end{cases}$$

Where, X are the standardized indicators, v are the eigenvectors, F are the principal components.

[7] The value of λ is defined by setting the value of the determinant to zero. In this case, since the determinant is 3X3, a cubic equation with respect to λ is formed, and λ can have 3 values ($\lambda = \lambda_i;\ i = 1,2,3$).

The set of values obtained after performing the above linear transformation is referred to as the principal components. The principal component in the first column of the set represents the largest variance associated with the indicator data.