Discovering and Modeling the Driving Forces of Multidimensional Scales with Application to Financial Data

Dr. Emtair M. Abdalla

Department of Mathematics, Faculty of Science, Sirte University, Libya
email: emtair2001@yahoo.de

Abstract

In this paper, the driving forces are represented by linear and nonlinear principal components. In order to extract such forces, dimension reduction is applied using spectral decomposition technique. The most contributing lower order principal components are retained to represent the driving forces. The modeling of these forces is achieved in comparison to the Gaussian one.

Key words: *Descriptive tools, Gaussian distribution, Spectral decomposition, Dimension reduction, Driving Forces, principal components, modeling characteristics, skewness and kurtosis, Financial Factors.*

1. Objectives and Methodology

The objective of this paper is to discover the descriptive characteristics and driving forces of the multidimensional DAX portfolio of nine financial factors. The study lays the light on the most influencing factors within the portfolio. Several techniques are utilized, among them are descriptive methods, visual aids and some algebra.

2. The Data and Descriptive Tools

The data under consideration consist of 605 points of DAX portfolio[1] comprising of nine financial factors as listed below

[1] The source is the comdirect bank at www.comdirect.de

X_1: the ratio of share holders' equity to total assets

X_2: the ratio of dept to total assets

X_3:the ratio of total liability to shareholders' equity

X_4:the ratio of liquid assets to total assets

X_5:the ratio of net income to shareholders' equity

X_6:the ratio of net income to total assets

X_7:the ratio of net income to sales

X_8:the ratio of sales to total assets

X_9:the ratio of dividends per share to earnings per share

For a summary of descriptive statistics of the factors, see Figure 1.

3. Modeling Characteristics

There is a lot of modeling characteristics that can be consider in modeling the financial factors as well as their driving forces. We restrict ourselves to the Gaussian model for comparison purposes.

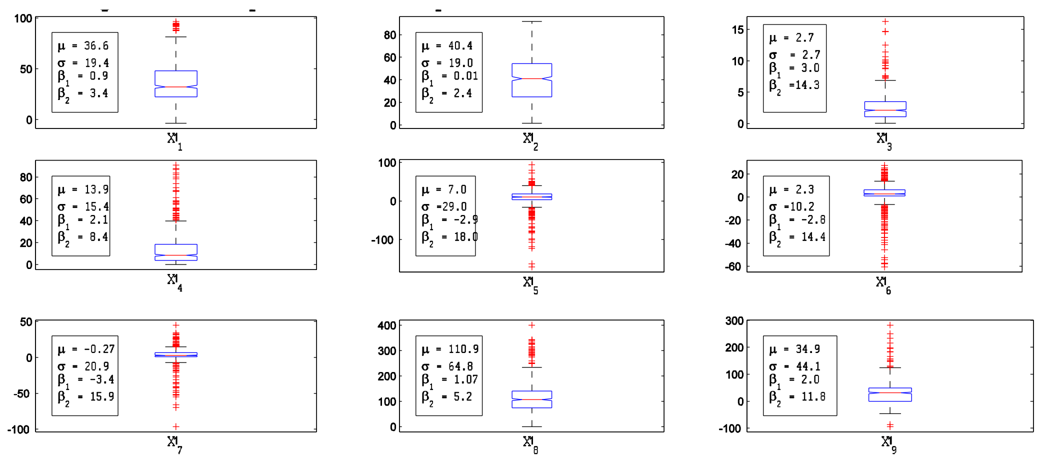

Figure 1 : Box plots of the financial factors together with descriptive characteristics such as the mean μ, the standard deviation σ, the coefficient of skewness β_1 and coefficient of kurtosis β_2.

3.1 Gaussian Distribution

A random variable × is said to be distributed as Gaussian with mean μ and variance σ^2 if and only if its probability density function is defined as

$$f_{\times}(x|\mu,\sigma) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\,exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, x \in \mathbb{R}$$

The distributions is completely defined by its mean and variance, [1].

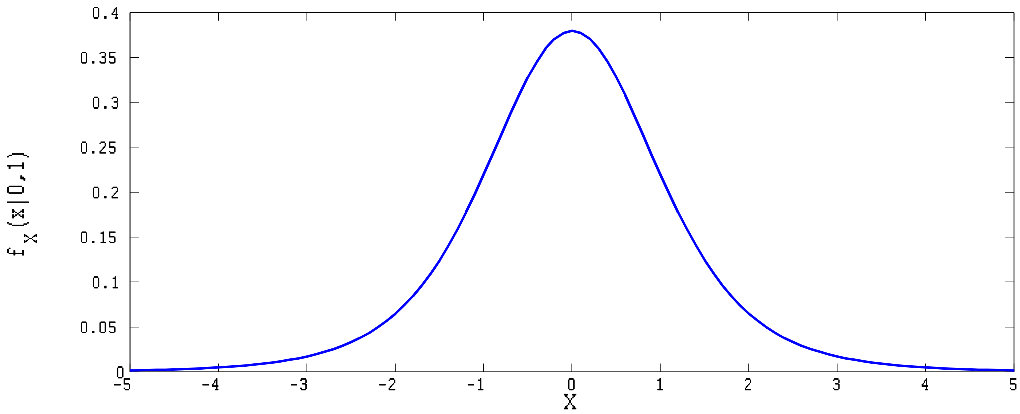

Figure 2 : The density function of Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

and is considered to be the best known and most important with its amazing characteristics of symmetry (=0) and mesokurtosis (=3), see Figure 2.

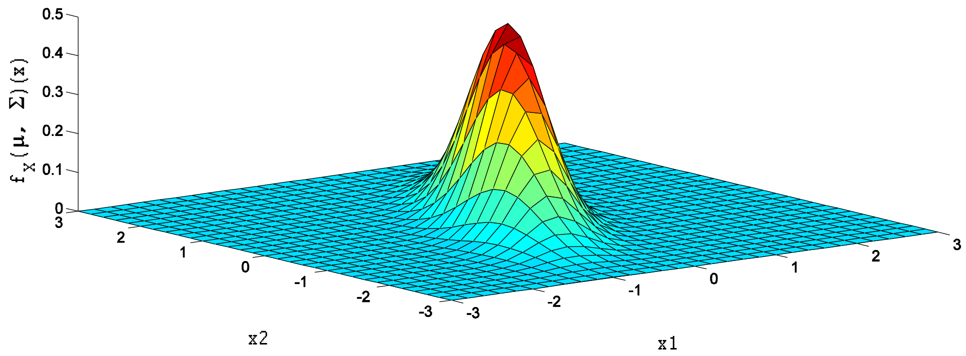

Figure 3: The density function of a bivariate Gaussian distribution with mean vector $\mu = [2\ 5]^t$ and dispersion matrix $\Sigma = \begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.75 \end{bmatrix}$

A multivariate Gaussian for a p-dimensional vector $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_p]^t$ with mean vector $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]'$ and a dispersion matrix $\Sigma = [\sigma_{ij}]_{i,j \le p}$ is given by

$$f_X(\boldsymbol{x}|\boldsymbol{\mu},\Sigma) = \frac{1}{\sqrt{2\,\pi^p\,|\Sigma|}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\,\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{p}}$$

[2] . See Figure 3.

3.2 Diagnostics

For the diagnostic of Gaussianity of the underlying factors, we apply quick indicative tools namely, the quantile-quantile plot, qq-plot, in its univariate and multivariate versions.

3.2.1 Univariate qq-Plot

The qq-plot is constructed by plotting the empirical quantile of the original data against the theoretical quantiles of the Gaussian distribution, namely the set of pairs

$$\left(\hat{F}^{-1}\left(\frac{i}{n+1}\right), \hat{\Phi}^{-1}\left(\frac{i}{n+1}\right)\right), i = 1,2,\dots,n.$$

See Figure 4.

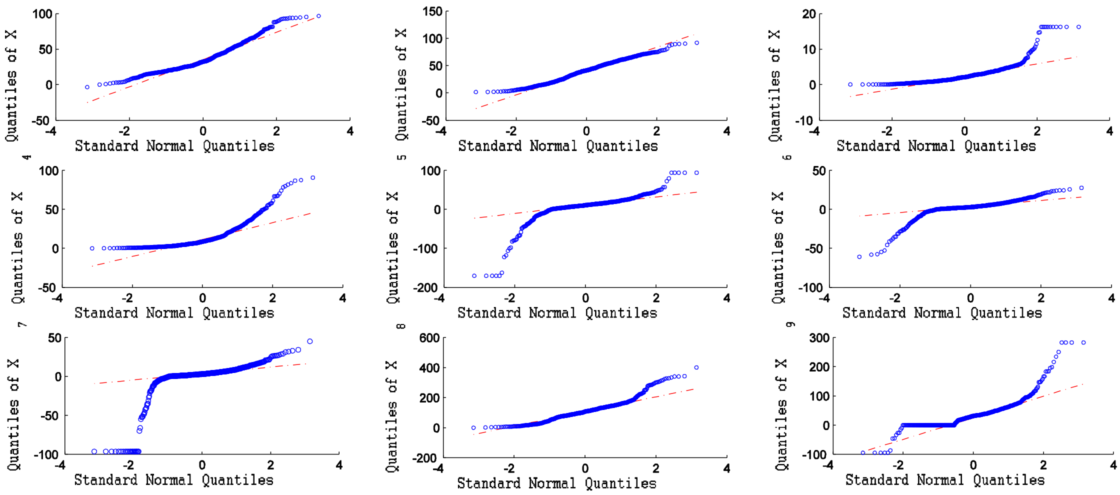

Figure 4: The qq-plots of each of the financial factors, the assumption of Gaussianity is violated by almost all factors except probably X_2.

Under the Gaussianity assumption, the univariate version of the Gaussian qq-plot is associated with the population correlation test

$$H_0\!:\ \mathrm{P} > \mathrm{P}_0\ vs\ \mathrm{P} \leq \mathrm{P}_0$$

that gives an indicative of the associateship between the empirical quantiles and the Gaussian quantiles. In order to guarantee a strong associateship the hypothesized value P_0 should be positive and large enough. The test is based, originally, on the test statistic ρ. Due to the complicated sampling distribution of the test statistic, it is replaced with a modified Gaussian alternative for which under the null hypothesis [1],

$$\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \sim N\left(\frac{1}{2}ln\left(\frac{1+R}{1-R}\right), \frac{1}{n+3}\right)$$

3.2.2 Multivariate qq-Plot

With analogy to the univariate version, the multivariate version the qq-plot is obtained by plotting the pairs

$$\left(D^{-1}\left(\frac{i}{n+1}\right), \hat{\chi}^{-1}\left(\frac{i}{n+1}\right)\right), i = 1,2, \dots, n$$

of the quantiles of the squared Mahalanobis[2] distance D against the quantiles of chi-square distribution χ ,[3], see Figure 5. For details of related tests we refer to [6].

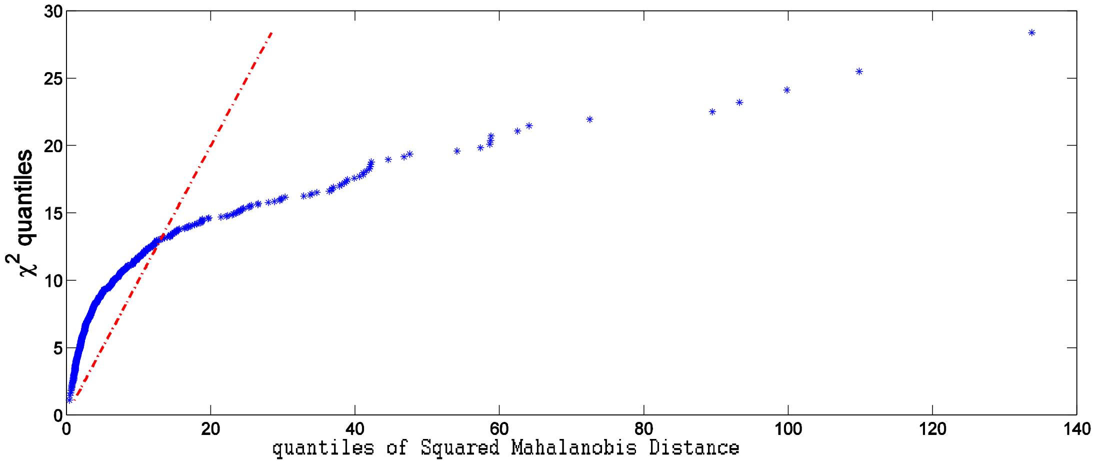

Figure 5: The multivariate qq-plot of the DAX portfolio, the assumption of multivariate Gaussianity is not tenable.

The modeling characteristics of the univariate factors as well as the multivariate portfolio are shown on Figure 4 and Figure 5 .As the points do not lie on a straight line, the Gaussianity assumption is violated which seem to be a characteristic of the underlying financial data.

[2] Referred to the Indian scientist and applied statistician Prasanta, M., Mahalanobis (June 29, 1893-June 28, 1972).

4. Spectral Decomposition

Spectral decomposition is achieved by decomposing the positive nonsingular matrix Σ into two matrices, namely, a diagonal one $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, and an orthogonal one $\Gamma = [\Gamma_j]_{j \leq p}$ as follows.

$$\Sigma = \Gamma\Lambda\Gamma^{-1}$$

The procedure has to be achieved in two steps, firstly, find the unknown eigenvalues $\lambda_i, i \leq p$ by solving the p-degree characteristic equation

$$|\Sigma - \lambda\, \mathrm{I}_p| = 0.$$

Secondly, solve the equation

$$|\Sigma - \lambda\, \mathrm{I}_p|\Gamma_j = 0$$

to find the unknown eigenvectors $\Gamma_j = [\gamma_{ij}]_{i \leq p}, j = 1,2, \dots, p$ corresponding to the positive eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$, [4].

5. The Driving Forces

The first step needed for discovering the driving forces and hence to extract the ones with most variation is to reduce the multidimensionality of the underlying portfolio to a lower dimensional coordinate system using the above procedure of spectral decomposition to the underlying covariance matrix.

5.1 Extracting the Linear Forces

The techniques is to apply spectral decomposition to the covariance of the centered data matrix $\boldsymbol{X} = [\times_1, \times_2, \dots, \times_p]^t$.

$$\Sigma = \frac{1}{N}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t$$

Then project $\boldsymbol{X}$ on the reduced q-dimensional coordinate system $\Gamma^{(q)}, q \ll p$ corresponding to the few λ's with higher values.

$$\mathbf{Y} = \mathbf{X}^t \Gamma^{(q)}, q \ll p$$

This reduces and interprets the dispersion matrix using a few linear combinations called linear principal components, $Y = \left[Y_1, Y_2, \dots, Y_q\right]^t, q \leq p$ with uncorrelated components,

$$cor\left(Y_i, Y_j\right)_{i \neq j} = 0; \;\; i, j = 1,2, \dots, q \; .$$

The desired principal component are obtained by excluding the components with lower contribution to the total variation, [9], [11].

5.2 Extracting the Nonlinear Forces

A nonlinear form of the principal components analysis is the kernel principal components. The method applied for extracting the nonlinear driving forces is the Kernel principal components technique, it is applied in three steps; firstly, using one of the known kernel functions, say Gaussian kernel, project the portfolio data matrix $\boldsymbol{X}$ into non-linear orthogonal matrix $\Phi(\boldsymbol{X})$ in a high dimensional functional space, called Hilbert space $\mathcal{H}$.

$$\Phi : \mathbb{R}^p \longrightarrow \mathcal{H}$$

Secondly, apply spectral decomposition to the covariance matrix of the centered kernel matrix $\Phi(\boldsymbol{X})$.

$$\Sigma = \frac{1}{N} \left(\Phi(\boldsymbol{X}) - \boldsymbol{\mu}_{\boldsymbol{\Phi}}\right)\left(\Phi(\boldsymbol{X}) - \boldsymbol{\mu}_{\boldsymbol{\Phi}}\right)^t,$$

Thirdly, project Φ into a reduced q-dimensional orthogonal space which ends to the desired kernel components, [5], [7].

The components that have most contribution to the total variation are represented by the lower order principal components. For details about principal component retention we refer to [8].

It is shown from Table 1. that the first three linear principal components can express more than 86% of the total variation and hence can be retained as a lower dimensional coordinate system.

Table 1. Percent of Variation explained by the first k principal components

Type of kernel	k				
	1	2	3	4	5
Linear	51.9757	74.4147	86.4100	94.9952	97.0736
Polynomial	39.8324	62.0023	78.8899	87.7216	92.7442
Gaussian	13.7894	27.5788	41.3682	55.1576	68.8334

The nonlinear structure also can be represented by the first three kernel components using a polynomial kernel of degree two, these components explain about 79% of the total variation and can be used as a lower dimensional coordinate system for the nonlinear structure of the underlying portfolio. Therefore, the nine dimensional coordinate system of the original factors can be reduced to three-dimensional coordinate system of principal components by projecting the data matrix onto the retained eigen-direction of the largest variance.

$$Y = \mathrm{X}^{\mathrm{t}}\Gamma^{(\mathrm{q})} = X^t \begin{pmatrix} -0.05 & 0.08 & 0.01 \\ 0.04 & -0.05 & -0.10 \\ -0.01 & -0.01 & -0.00 \\ -0.06 & 0.04 & -0.01 \\ 0.13 & -0.17 & 0.80 \\ 0.06 & -0.05 & 0.25 \\ 0.10 & -0.13 & 0.46 \\ 0.96 & 0.26 & -0.11 \\ 0.21 & -0.94 & -0.25 \end{pmatrix}$$

$$= \begin{pmatrix} -0.05x_1 & 0.08x_1 & 0.01x_1 \\ 0.04x_2 & -0.05x_2 & -0.10\,x_2 \\ -0.01x_3 & -0.01x_3 & -0.00\,x_3 \\ -0.06x_4 & 0.04x_4 & -0.01x_4 \\ 0.13x_5 & -0.17x_5 & 0.80x_5 \\ 0.06x_6 & -0.05x_6 & 0.25x_6 \\ 0.10x_7 & -0.13x_7 & 0.46x_7 \\ 0.96x_8 & 0.26x_8 & -0.11x_8 \\ 0.21x_9 & -0.94x_9 & -0.25x_9 \end{pmatrix}$$

The most influencing variable in the first linear principal component is x_8 whereas for the second and third components the most influencing variables are x_9 and x_5 respectively. A qq-plot of the univariate and multivariate principal components is given on Figure 6(a) and Figure

6 (b). A violation of the Gaussianity assumption as well as skewness appear in both the univariate and multivariate cases.

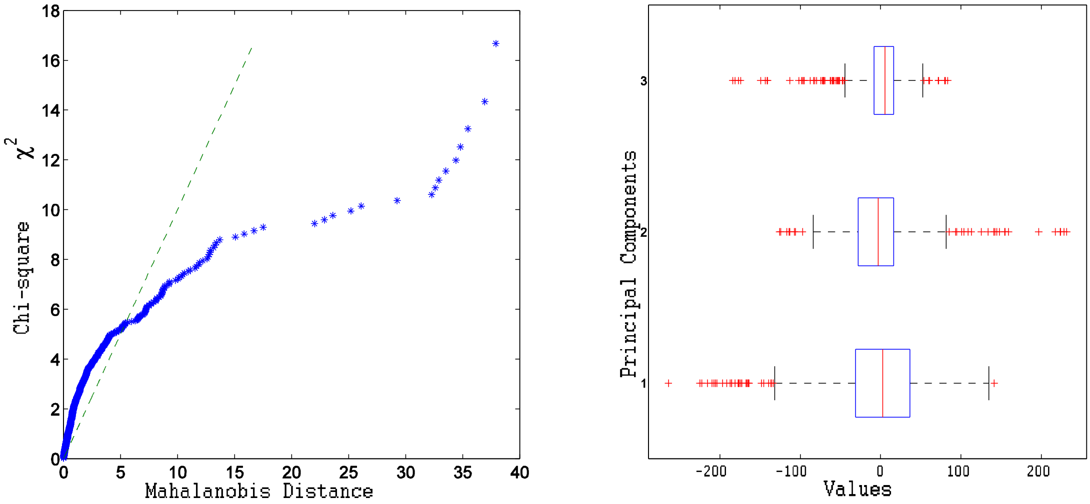

Figure 6 : (a) The univariate qq-plot of single principal components and (b) The multivariate qq-plot of the first three principal components.

6. Conclusion

Two types of variation involved in the multidimensional data are considered, linear and nonlinear. On this basis, two techniques have been applied, the linear principal components to extract the linear forces and the kernel components for extracting the nonlinear ones. In our data, the linear variation may be represented by the first three linear principal components which express more than 86% of the total variation, whereas the first three kernel principal components express about 79% of the total variation and may be retained to represent the portfolio.

From the above modeling procedures, results agree with the violation of the Gaussian assumption by all original financial factors and underlying principal components. The modeling characteristics show that skewness and platy-kurtosis a common characteristic of the driving forces. Hence, the conclusion from the above study is that the modeling characteristics of the driving forces are away from the Gaussian assumptions and are among the asymmetric and non-mesokurtic modeling environment. The most influencing factors in the total variation are; x_8, the ratio of sales to total assets, x_9, the ratio of dividends per share to earnings per share, and x_5, the ratio of net income to shareholders' equity.

7. References

[1] Freund, J. (1992). Mathematical Statistics. Fifth Edition, Prentice-Hall, New Jersy.

[2] Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme values with Applications to Insurance, Finance, Hydrology and other Fields* . Birkhaeuser, Berlin.

[3] Johnson, N., and Wichern, D. (1988). *Applied Multivariate Statistical Analysis*. Second Edition, Printice-Hall Interntional, New Jersy.

[4] Timm, N. H. (2002). *Applied Multivariate Analysis*. Springer-Verlag, Berlin.

[5] Franc, V. and Hlavac, V. (2004). Statistical Pattern Recognition Toolbox for Matlab, User's Guide. Technical University Prague.

[6] Mardia, K. V. (1970). Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, 57(3):519-530.

[7] Rosipal, R., Trejo, J. and Cichock, A. (2000). *Kernel Principal Components Regression with EM Approach to Nonlinear Principal Components Extraction. Technical Report*. University of Paisley, Scotland.

[8] Cangelsoni, R. and Goriely, A. (2007). Component retention in Principal component Analysis with Application to cDNA microarray data. *Biology Direct*, 2:2, BioMed Central.

[9] Falk, M., Marohn, F., and Tewes, B. (2002). *Foundation of Statistical Analysis and Applications with SAS*. First Edition, Birkhaeuser, Berlin.

[11] Hastie, T., Tibshirani, R. and Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Third Edition, Springer, Berlin.

[12] Seber, G. A. F. (2004). *Multivariate Observations*. John Wiley & Sons, New York.