EVALUATION OF PROBABILITY OF MARKET STRATEGY PROFITABILITY USING PCA METHOD

Pavla Kotatkova Stranska – Jana Heckenbergerova

Abstract

Principal component analysis (PCA) is a commonly used and very powerful method for simplification of multivariate data. In the presented contribution, it is, however, utilized in a novel way. The methodology described in the paper provides tables of probabilities of market strategy profitability. Values of the first two principal components are utilized in evaluation of conditional probability. The resulting probability tables can help to detect buying signals for profitable trade opportunities or can provide a warning against a high financial loss.

A time period of two weeks before an actual buying signal is analyzed. After one week of holding, the share is sold and the resultant relative price differences determine the group of profitability (high profit, profit, loss, high loss). For each group, tables of occurrences and conditional probabilities are evaluated using the PCA method. The whole methodology is tested on a case study that takes in account one year long dataset (1.Aug 2014 – 14.Aug 2015) of L-3 Communications Holding shares.

Key words: profitability, conditional probability, principal component analysis

JEL Code: C38, G11, G17

Introduction

Analysis of the financial market is very large topic and one needs a lot of knowledge and experience for its full understanding you. Most of the investors, however, do not attempt to fully understand all the analyses and they prefer to use one or two methods that are considered most effective (Lo et al., 2000), (Neely et al., 1997), (Byun et al., 2015). Each method is based on a different approach and has various prerequisites for successful investing. Nevertheless, all methods have one thing in common; they are trying to understand the current situation in the market and to predict whether the prices or shares on financial market will grow or decline. According to Bauer and Dahlquist (1998), the main motto of the majority of trades is "cheap acquisition of a financial asset or his precious sales". Based on scientific studies Lee et al. (2017), Lynch et al. (2000), Lo et al. (2000) and Neely et al. (1997), all kinds of statistical methods are applied nowadays to build reliable prediction model, e.g. principal component analysis (PCA). So we can conclude that prediction of prices future development is a constantly evolving area.

PCA is a popular method in the area of financial analyses; in particular it is used to define and to quantify the factors that influence changes in the prices of securities or shares (Gemperline, 2006). Utilization of PCA for evaluation of profit probability is, however, novelty in this research field. The prerequisite is classification of profit levels for selected trading strategy so each trading day, consequently, belong to one profit class. Based on data precedent selected trading day, principal components are evaluated. As a result from whole dataset, we are getting conditional probabilities of profit class occurrences based on principal component values. Proposed methodology is similar to method for prediction of wind ramp events, which was presented and validated in (Heckenbergerova, 2014), (Heckenbergerova, 2015) and (Heckenbergerova, 2016).

This paper consist five main sections. After introduction, principal component analysis is briefly described. Second section summarizes methodology that is used in case study presented in the section 3. Major conclusions with direction of future work are provided in the last section.

1 Principal Component Analysis

The basic idea of principal component analysis (PCA) is the reduction of the number of variables by using the so-called the principal components. Each principal component is formed as linear combination of the original variables and it is characterized by covariance of all the variables. Therefore most significant principal components can summarize knowledge about the default variables while preserving all the original information (Lo et al., 2000), (Neely et al., 1997), (Gradojevic and Lento, 2015).

According to Rencher (2002), PCA is based on a covariance matrix of the p variables $\boldsymbol{X}_j$. From its eigenvalues and eigenvectors, new variables $\boldsymbol{Z}_j$ (principal components) are derived that are uncorrelated and sorted according to decreasing variance (eigenvalues).

Let us assume that all data from selected dataset are forming *N*-dimensional random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ and moreover let us consider the time frame of length *n*. Observation matrix $\mathbf{X}$ is then in form:

$$\mathbf{X} = \begin{pmatrix} Y_1 & Y_2 & \cdots & Y_n \\ Y_2 & Y_3 & \cdots & Y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{N=n} & Y_{N=n+1} & & Y_N \end{pmatrix}, \qquad (1)$$

where each row corresponds to one time frame of simulation and each column is a random vector $\mathbf{X}_\mathrm{j}$ of given time index.

In the next step eigenvalues and eigenvectors of covariance matrix var **X** are evaluated using singular value decomposition:

$$\operatorname{var} \mathbf{X} = \mathbf{SVD}' \qquad (2)$$

where **S** is an orthogonal (*n* × *n*)-matrix with columns of singular vectors, **V** is a diagonal (*n* × *n*)-matrix corresponding to eigenvalues of **X** and columns of matrix **XD**′ are the principal components of matrix **X** (Wall et.al, 2003).

2 Methodology

Methodology of evaluation of conditional probability tables can be summarized in following steps:

1) Dataset (time series of share prices) selection;
2) Market strategy selection (holding period);
3) Sorting of trading days based on profitability (creation of profitability classes);
4) Creation of observation matrix **X** with selected length of time frame prior trade;
5) Evaluation of principal components;
6) Division of principal components ranges to create support rectangles and table axes;
7) Evaluation of occurrence frequencies on each support rectangle from whole dataset and from each profitability class;
8) Evaluation of conditional probability for each support rectangle and creation of conditional probability tables.

Resulting tables show us conditional probability of the future trade profitability based on actual principal component values.

3 Case study

1.1 Data and Market Strategy

Let us take in account the one year long dataset of L-3 Communications Holding shares (1.Aug 2014 – 14.Aug 2015). Our market strategy is selling the share after one week of holding (5 business days). For each trading day relative profitability (*RP*) is evaluated using formula:

$$RP = \frac{(P_{sell} - P_{buy})}{P_{buy}}, \qquad (3)$$

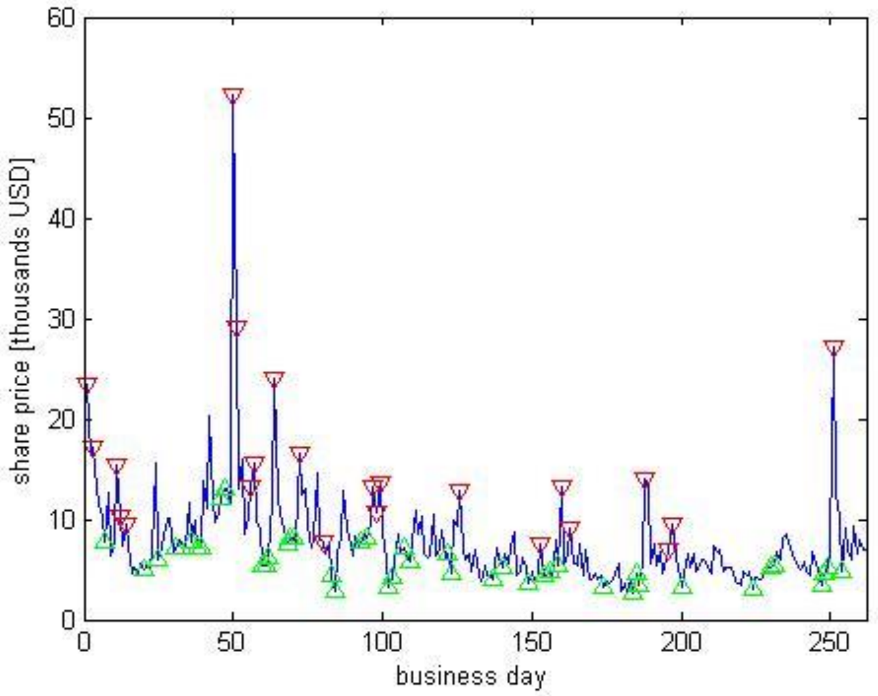
where P_{buy} is buying price for selected trading day and P_{sell} is share price on selling day. Based on these values, data are divided into four groups: HP (high profit) for profitability greater than 0.5; P (profit) for *RP* values between 0 and 0.5; L (loss) for negative profitability between -0.5 and 0; and HL (high loss) for profitability below -0.5 threshold. Table 1 summarizes the size of profitability classes and Figure 1 shows time series of L-3 Communications Holding shares with high profit and high loss trades highlighted. All figures are generated in Matlab using our own scripts.

Tab. 1: Profitability classes

Type of class	Size of class
whole dataset	262
HP (high profit)	43
P (profit)	84
L (loss)	106
HL (high loss)	23

Source: own

Fig. 1: Share prices of L-3 Communications Holding from 1st Aug 2014 till 15th Aug 2015 augmented with high profitable (triangle up) and high loss (triangle down) trades



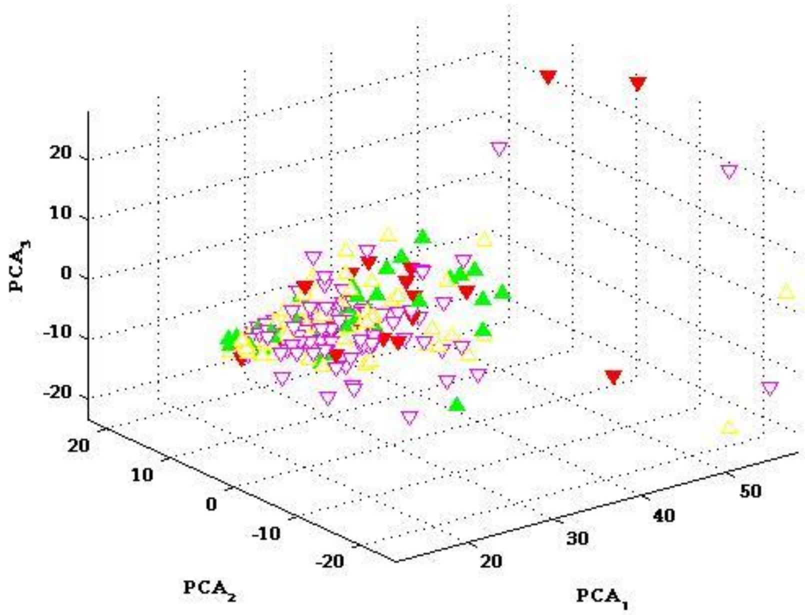
Source: own

1.2 Principal Component Analysis

Let us assume that all data from selected dataset are forming *N*-dimensional random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ and moreover let us consider the time frame of length $n = 10,$ corresponding to two weeks horizon before buying signal (10 business days). Observation matrix $\mathbf{X}$, based on Eq.(1), has then 10 columns and 252 rows.

In the next step eigenvalues and eigenvectors of covariance matrix var **X** are evaluated using singular value decomposition. As a result we are getting principal components of observation matrix **X**. First three principal components (PCA1, PCA2 and PCA3) are illustrated in Figure 2, where color and shape of triangles signalize different profitability classes (green-HP, yellow-P, magenta-L, red-HL). Components PCA1 and PCA2 can explain together 39% respectively 52% of the total variability of data, therefore they are chosen as significant ones.

Fig. 2: Three principal components for L-3 Communications Holding dataset

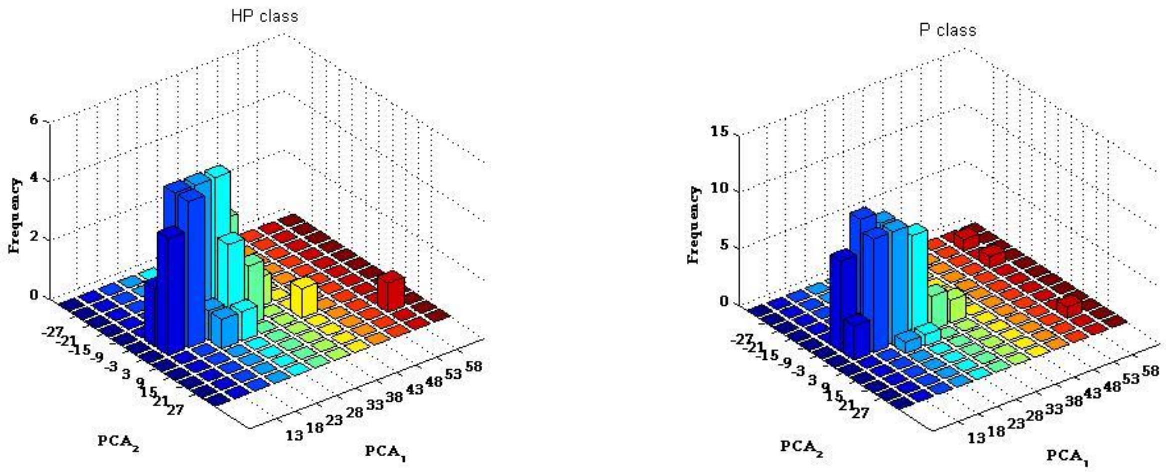


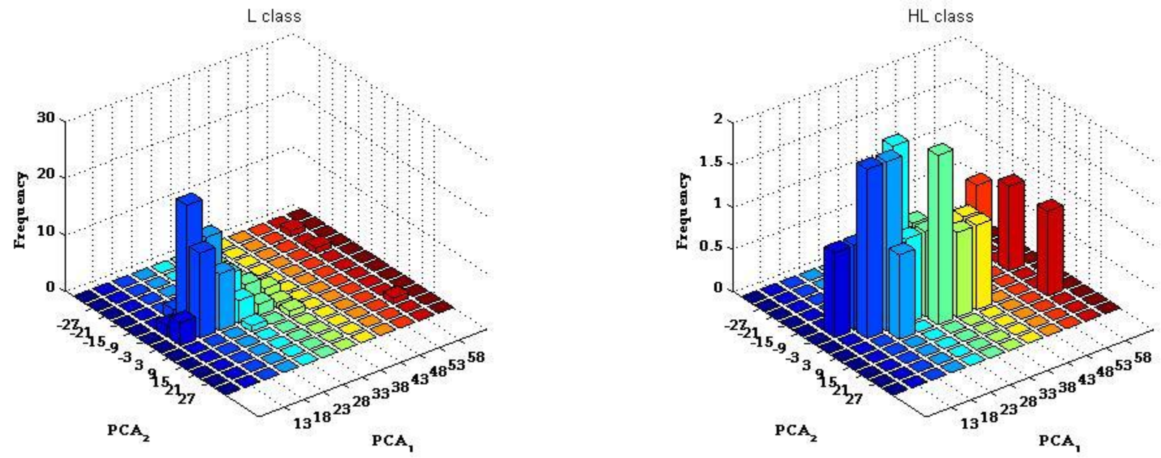
Source: own

1.3 2D Tables of Trade Profit Conditional Probabilities

Let us divide the PCA axes into 12 parts so that the whole graph area is covered by 2D blocks (rectangles). Frequency tables for HP, P, L and HL events are created by the number of occurrences in the corresponding 2D block. They are captured in following figure 3.

Fig. 3: Frequencies of profit classes occurrence on selected 2D blocks.



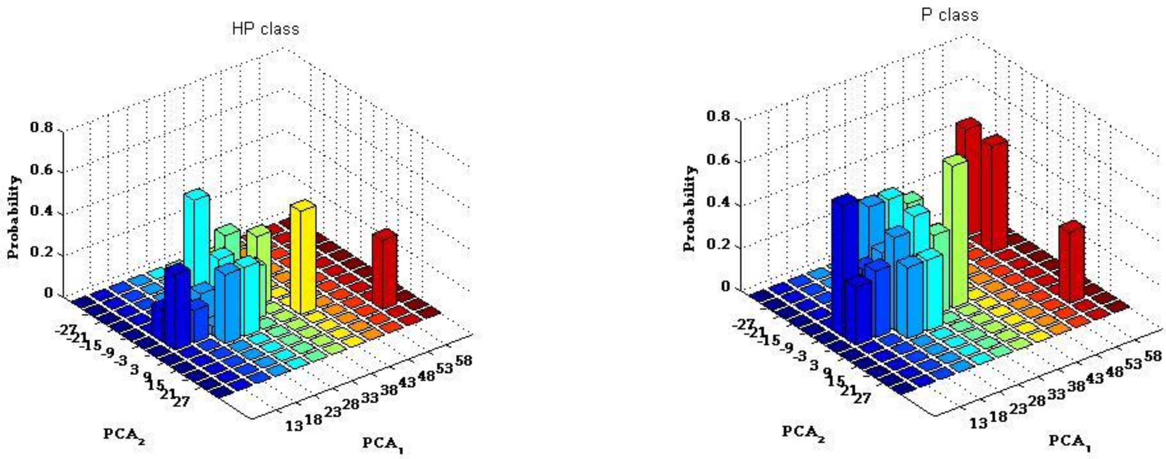


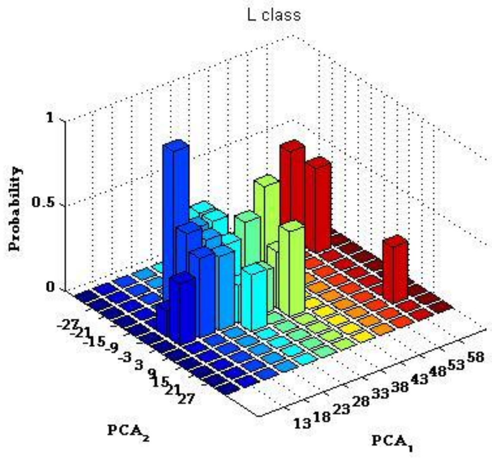
Source: own

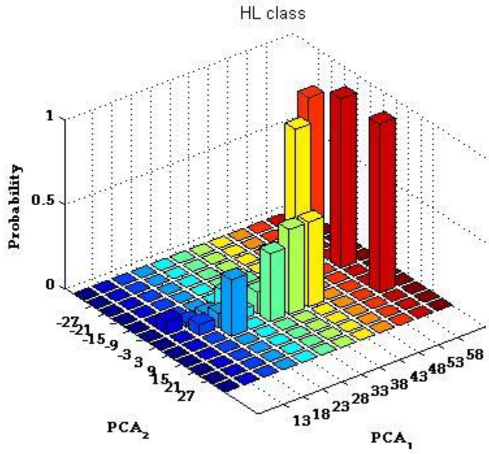
Conditional probability value in selected rectangle for any profitability class is defined as fraction of profitability class frequency and all data frequency in that block. Figure 4 shows column graphs of evaluated conditional probabilities.

We can see that for high value of first principal components PCA1, there is significant probability (close to one) that buying now and selling in a week will generate high loss. We can also conclude, that low values of PCA1 (less than 30) shows very low risk of high loss.

Fig. 3: Conditional probabilities of trade profitability given by the 2D block of principal components







Source: own

Conclusion

This paper presents a principal component analysis of share prices time series to evaluate probability of selected market strategy profit. Illustrated methodology uses share prices of L-3 Communications Holding from 1.Aug 2014 till 14.Aug 2015. Market strategy is selling the share after one week of holding. Two weeks horizon before actual buying signal is analyzed to see whether trade profitability is predictable.

Results show that low values of first principal component are not in favor with trade with high loss. In contrary, high values of PCA1 will result in high loss with high probability.

In our further research, validation of presented methodology will be performed with more and current datasets.

References

BAUER, R. J. & DAHLQUIST J. R. (1998). *Technical Market Indicators: Analysis and Performance*. Wiley Trading. ISBN 978-0-471-19721-8.

BYUN, H. W., BAEK, S., AHN, J.J., OH, K. J., KIM, T.Y. (2015). Using a principal component analysis for multi-currencies trading in the foreign Exchange market. *Ingelligent Data Analysis.* 19(3), pp. 683-697. DOI 10.3233/IDA-150738.

HECKENBERGEROVA J., MUSILEK P., JANATA M., WANG J. (2015). Analysis of Wind Ramp Events Using PCA of 3D Tables of their Conditional Probabilities. *16th International Scientific Conference on Electric Power Engineering-EPE 2015*. Kouty nad Desnou, pp. 63-66. ISBN 978-146736788-2.

GRADOJEVIC, N., LENTO, C. (2015). Multiscale analysis of foreign Exchange order flows and technical trading profitability. *Economic Modelling.* 47(1), pp. 156-165. ISSN 0264-9993. DOI 10.1016/j.econmod.2015.02.028.

HECKENBERGEROVA J., MUSILEK P., MAREK J., RODWAY J. (2014). Principal Component Analysis for Evaluation of Wind Ramp Event Probability. *Electrical Power and Energy Conference*-(*EPEC 2014*). Canada: Calgary, pp 208 – 212. ISBN 978-147996038-5.

HECKENBERGEROVA, J., MUSILEK, P., JANATA, M. (2016). Sensitivity Analysis of PCA method for Wind Ramp event Detection. *16th International Conference on Environment and Electrical Engineering (EEEIC).*Italy: Florence. doi:10.1109/eeeic.2016.7555892.

LEE, S. ENKE, D., KIM, Y. (2017). A relative value trading system based on a correlation and rough set analysis for the foreign exchange futures market. *Engineering Applications of Artificial Inteligence.* 61(1). ISSN 0952-1976. DOI 10.1016/j.engappai.2017.02.014

LO, A. W., MAMAYSKY, H. & WANG, J. (2000). *Foundation of Technical Analysis: computational algorithms, statistical inference, and empirical implementation*. National Bureau of Economic Research.

LYNCH, D. F., KELLER, S. B., OZMENT, J. (2000). The effects of logistics Capabilities and Strategy on firm performance. *Journal of Business Logistics*. pp. 47-68.

NEELY, CH., WELLER, P. & DITTMAR, R. (1997). Is Technical Analysis in the Foreign Exchange Market Profitable?: A Genetic Programming Approach. *Journal of Financial and Quantitative Analysis.* 4(32), pp. 405-426.

RENCHER, A. C. (2002). *Methods of Multivariate Analysis*. New York: John Wiley & Sons, Inc. ISBN 0-471-41889-7. DOI: 10.1002/0471271357.ch12

GEMPERLINE, P. J. (2006). Practical Guide To Chemometrics, New York: CRC Press. ISBN 978-1-57444-783-5. DOI 10.1201/9781420018301.ch4

WALL, M., RECHTSTEINER, A. & ROCHA, L. (2003). *Singular value decomposition and Principal component analysis*. Kluwer: Norwell, MA.

Contact

Pavla Kotatkova Stranska, Ing. Ph.D.

Masaryk Institute of Advanced Studies, MIAS School of Business

Department of Economic studies

Czech Technical University in Prague

Kolejní 2637/2A, 160 00 Praha 6

e-mail: pavla.kotatkova.stranska@cvut.cz

Jana Heckenbergerova, Mgr. Ph.D.

Faculty of Economics and Adminitration, Institute of Mathematics and Quantitative Methods,

University of Pardubice, Studentska 95, 532 10 Pardubice, Czech Republic

email: jana.heckenbergerova@upce.cz