Business Applications

CLUSTERING IN THE SPACE OF INDEPENDENT COMPONENTS AS A TOOL FOR REVEALING CONTAGION EFFECT ON FINANCE MARKETS

Konstantin Dylko, Mikhail Alexandrov

Abstract: *This paper introduces a new method for analyzing time-series of stock market indexes (Dow Jones, NASDAQ etc) in order to reveal the quantitative level of their interference. In literature, the interference between stock indexes is usually called 'contagion'. Strictly, contagion – is transmission of shocks from one region to another. In order to prevent destructive effects of crisis it's very important for the government to predict them. The proposed method consists of two steps: transformation of time series to the space of independent components and clustering in this space using the principle of stability. The contents of clusters just define companies, countries, and regions related by the effect of contagion. To complete the mentioned steps we use: the IcaLab tool from MatLab package to construct independent components, MajorClust method and DEM index for clustering and testing its quality. MajorClust and DEM were developed in Benno Stein research group. The results of experiments showed the essential advantage of the proposed method over the traditional approach based on formal relations in correlation matrix. Notably, we marked out the groups of market indexes, which were most likely connected via 'contagion'. Such a conclusion was made using external data like geographical position, level of country development, macroeconomic rates, etc.*

Keywords: *independent component analysis, clustering with MajorClust, contagion effect, financial market*

ACM Classification Keywords: *I.2M Miscellaneous*

Introduction

There are many different types of crisis in economy. For example: currency, bank, stock, world crisis and etc. Asian crisis in 1997, crisis in Latin America, Great Depression and, finally, recent world crisis has shown their destructive power [Sakho, 2006; Yeyati, 2008]. Some indicators of financial crisis were created. One of them is effect of contagion. There is no strictly accepted definition of contagion in literature. There are many definitions of this effect, but all of them connect this phenomenon with transmission of shocks between the markets [Allen, 2000; Bradley, 2004; Bradley, 2005; Francel, 1998; Kodres, 2002, Serva, 2006]. Let's give a definition of contagion on which we will base our work.

Effect of contagion – is phenomenon of transmission of shock or crisis which had begun in one country to another country or region. One also says that effect of contagion occurs when there is a strong bound between the markets in shock periods of time.

The contagion is relatively new topic, although many working papers have been issued. Kodres and Pritsker in their paper [Kodres, 2002] determined and analyzed main channels for contagion. Their paper gives a good understanding of issue of contagion. Siu-Ming Cha and Lai-Wan Chan in their work [Siu-Ming Cha, 2000] introduced mathematical methods in factor model analysis in finance and gave number results. The period of high correlation was called shock period [Loretan, 2002]. Bradley and Taqqu, in their papers [Bradley, 2004; Bradley, 2005], proposed another method for contagion definition based on local correlation. The main idea was that markets are more bounded in crisis period than usually.

But, this entire works don't take into account the structure of financial data. It mean that, for example, stock market index could be explain using complex factor model in which only few of factors are responsible for the contagion effect.

In this paper we propose methodology, which takes into account the structure of data. This methodology consists of two steps:

1) Transformation of stock market time-series in the space of independent components. This step gives us a possibility to find factors responsible for the contagion effect and to erase noise. The problem is to define optimal number of components. We give more detailed view on this problem in section 2.
2) Cluster analysis with automatic define of number of clusters in the space of independent components. This allows to avoid the distortion of results, as influence of external factors is decreased significantly. The main problem here is to find the optimal number of clusters as this number is sensitive to the links matrix. We give more detailed view on this problem in section 3.

To show the advantage of methodology proposed, we make experiments with grouping of indexes in two cases: with transformation and without transformation. The results of experiments are shown in section 4. In section 5 we give conclusions.

Independent component analysis

2.1 Procedure description

There is an example in the book [Hyvarinen, 2001] , which explains ICA method. Imagine that you are in a room where three people are speaking simultaneously. You have three microphones, which you hold in different locations. The microfones give you three recorded time signals, which we could denote by $x_1(t), x_2(t), x_3(t)$. Each of this recorded signals emitted by the two speakers, which we denote by $s_1(t), s_2(t), s_3(t)$.

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t)$$

$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)$$

Or we could also write it as matrices: X= AS."

The task is to find S using only X. It means that using ICA method we find how to S depends on X.

We make some suggestions in ICA. First of all let $s_1(t), s_2(t), s_3(t)$ be independent. It means that $s_i(t)$ has no information about $s_j(t)$, while $i \neq j$. Secondly, we assume that $s_1(t), s_2(t), s_3(t)$ have a distribution different from normal.

Here first suggestion allows to build algorithm with isolated account of components $s_i(t)$, and the second one allows to filter white noise.

The algorithm is based on the central limit theorem, which denotes that with certain conditions sum of independetly distributed random quantities is seeking normal distribution as we increase number of summands.

So we seek combination that makes distribution of independent variables as much as possible unlike to normal distribution

We use algorithm called JADE. It consists of two steps:

- Decorrelation. We diagonalize correlation matrix.
- Turn. We find such a turn matrix that outgoing components are as much as possible independend.

This method allows finding hidden structure of stock returns.

2.2 Preprocessing

Usually, time-series is non-stationary. In this case ICA method works incorrectly. We take first differences of time-series to avoid non-stationarity.

Also we make such procedures as:

Norming time-series on the interval [-1;1] on the basis of it standard deviation

Centring on the basis of its mathematical expectation.

Back et al in their work [Back, 1997] used this preprocessing and procedure JADE to analyze data from Tokyo stock exchange. Independend components were divided on two categories: higher than threshold and lower. First category is responsible for the sinificant changes in conomy which take place not often. And the secon category is responsible for the noise. This method gives a fresh look on the understanding of structure of stock returns.

Clustering

3.1 MajorClust description

For a moment, there are dozens of methods and their modifications in cluster analysis, which can satisfy practically all necessities of users. The most popular ones are K-means, oriented on the structures of spherical form, and Nearest Neighbor (NN), oriented on the extended structures of chain form [Hartigan, 1975]. In our work, we use the MajorClust method, developed by Stein et al. [Stein, 1999], which has the following advantages over the mentioned two:

- MajorClust distributes objects to clusters in such a way that the similarity of an object to the assigned cluster exceeds its similarity to any other cluster. This natural criterion provides the grouping of objects, which better corresponds to the users' intuitive representation. Neither *K*-means nor NN methods possess such optimization property: they do not evaluate the similarity between clusters.
- MajorClust determines the number of clusters automatically and in all cases tends to reduce this number. *K*-means requires the number of cluster to be given, and NN does not determine this number at all: cutting of the dendrite is performed by the user.

MajorClust has been successfully used with various data sets and demonstrated very good results especially in Text Processing [Stein, 2003a; Alexandrov, 2005]. The main disadvantage of MajorClust is its runtime. However, in case of sparse matrix of relations this disadvantage is not essential. In our problem, we are faced just with this case because of weak similarity between the majority of objects. These weak connections are eliminated that leads to the mentioned matrix.

3.2 Cluster validity

For testing cluster validity we use so-called DEM, density expected measure, which reflects the contrast of object distribution in the space object parameters. This index was firstly introduced in [Stein, 2003b]. It is easy calculated by the formula

$$\lambda = \Sigma\ (n_i / n)\ (w_i / n_i^{\theta})\ , \quad n^{\theta}=w$$

where: n_i and n are the number of objects in i-th cluster and total number of objects,w_i and w are sum of closeness in i-th cluster and totally in matrix of closeness. Higher values of λ mean better clustering. Stein *et al*.showed that DEM correlates well with expert opinion [Stein, 2003b]. Thus DEM can be an indicator of quality.

The main criterion for choosing the best number of clusters is the criterion of stability. The procedure consists in the following: 1) one eliminates weak links in the matrix of closeness on a low level, 2) clustering and calculation of DEM 3) these steps are repeated on higher level of closeness

Naturally, the higher level of threshold the more number of clusters we have. The sharp change of this number is a point of stopping.

Experiments

4.1 Source data and tools

In experiment we analyzed 13 stock indexes from different countries for the period of 2005-2010 year. We were expecting to find effect of contagion between USA indexes and EU indexes. Other indexes could be marginals.

We made preprocessing for all data.

In the first experiment we used independend components as it was described in chapter 2.1. Constructing of independend components was made with the help of ICALAB software, which is a part if well-known package MatLab [www.bsp.brain.riken.jp]. In the second experiment transformation into the space of independend component wasn't produced. Clusterisation with MajorClust and calculation of DEM were made on the basis of library procedures of cluster analysis, made in Autonomus University of Barcelona on C++.

4.2 Experiment without independent components

In the first experiment we made clusterization of initial data correlation matrix after preprocessing but without ICA implementation. In the Table 1 one can see the results. The data in columns represents number of cluster relative to the index in the row. Therefore, we have here the following 7 clusters on the best threshold, which is equal 0.4: (RTS) , (CAC, FTCE) , (AMEX, DAX,DOW, NASDAC, S&P), (FTSE Bursa), (Heng), (IPC), (NIKKEI)

Table 1. Results of clustering (experiment without independent components)

	Threshold	**0,3**	**0,35**	**0,4**
	Number of clusters	**4**	**4**	**7**
	DEM	**1,27**	**1,25**	**1,49**
Indexes of financial markets		No. of cluster	No. of cluster	No. of cluster
RTS		1	1	1
AMEX HONG KONG		2	2	2
CAC		2	2	3
DAX		2	2	2
DOW Jones Industrial		2	2	2
FTSE Bursa Malaysia		2	2	4

FTSE		2	2	3
Heng seng index		2	2	5
IPC (MMX)		3	3	6
NASDAQ		2	2	2
NIKKEI		4	4	7
S&P 500 INDEX RTH		2	2	2
STI		2	2	4

4.3 Experiment with independent components

In the second experiment we made the same thing but using ICA method. We have Found the correlation matrix for the data after cleaning with ICA from the noise. The results are represented in the Table 2.

The data in columns represents number of cluster relative to the index in the row. Therefore, we have here the following 6 clusters on the best threshold, which is equal 0.4:

(RTS) , (CAC, DAX, FTSE, STI), (AMEX, FTSE Bursa), (Heng, IPC), (NASDAQ, S&P), (NIKKEI)

Table 2. Results of clustering (experiment with independent components)

	Threshold	**0,3**	**0,35**	**0,4**
	Number of clusters	**3**	**4**	**6**
	DEM	**1,15**	**1,36**	**1,45**
Indexes of financial markets		No. of cluster	No. of cluster	No. of cluster
RTS		1	1	1
AMEX HONG KONG		2	3	3
CAC		2	2	2
DAX		2	2	2
DOW Jones Industrial		3	3	5
FTSE Bursa Malaysia		3	2	3
FTSE		2	2	2
Heng seng index		2	4	4
IPC (MMX)		2	4	4
NASDAQ		3	3	5
NIKKEI		2	2	6
S&P 500 INDEX RTH		3	3	5
STI		2	2	2

As we can see the results of this table are much better then the previous. There are clear clusters of USA indexes and EU indexes. Other indexes are marginal. It means that our methodology give better resuls than usual methodology whose results are reflected in the Table 1.

Conclusions

Results of research:

- We have proposed a new method for constructing of indicator of effect of contagion. This indicator is a modified coefficient of correlation. We remove noise from the time-series using ICA method which hides effect of contagion.
- We have proposed methodology of selection of groups of indexes interconnected with effect of contagion. This methodology consists of joint implementation of clustering methood MajorClust and criterion of stability of grouping DEM. MajorClust allows automatic chose of number of clusters. DEM allows to define transition point to sherp desc of number of clusters while cuttong of weak bounds between indexes.
- We have made experiment with 13 stock market indexes on the period from 2005 to 2010 year. The 13 indexes included Russian stock index RTS.
- Experiments show that proposed methodology gives better results than traditional indexes. Clustering of indexes shows expected groups (USA, EU, others). The attempt to find groups of indexes using usual correlation matrix didn't give the expected results.

Implementation of obtained results:

- Proposed indicator could be used to predict and prevent crises, while government anti-crisis measures
- Investors can also use this methodology to diversify their portfolios.

Bibliography

[Alexandrov, 2005] Alexandrov, M., Sanchis E., Rosso, P., (2005): Cluster Analysis of Railway Directory Inquire Dialogs. In: "Text, Speech, Dialog", Springer, LNCS, 3685, pp. 385-392

[Allen, 2000] Allen, F., Gal, D. Financial contagion. The journal of political economy, Vol. 105, № 1, p 1-33, 2008

[Back, 1997] Back, A. A first application of independent component analysis to extracting structure from stock returns. International journal of neural systems, Vol.8, №5, 1997

[Breadley, 2004] Breadley, B., Taqqu, M. Framework for analyzing spatial contagion between financial markets. Financial letters 2004, 2(6), 8-15, 2004

[Breadley, 2005a] Breadley, B., Taqqu, M. How to estimate apatial contagion between financial markets. Financial letters 2005, 3(1), 64-76, 2005

[Breadley, 2005b] Breadley, B., Taqqu, M. Empirical evidence on spatial contagion between financial markets. Financial letters 2005, 3(1), 77-86, 2005

[Frankel, 1998] Frankel, J., Schmukler, S. Crisis, Contagion, and Country Funds: Effects on East Asia and Latin America. Cambridge University Press , 1998

[Hartigan, 1975] Hartigan, Wiley (1975). J.: Clustering Algorithms.

[Hyvarinen, 2001] Hyvarinen, A., Karhunen, J., Oja, E. Independent Component Analysis» John Wiley&Sons, 2001

[Kodres, 2002] Kodres, L.,E., Pritsker, M. A rational expectation model of financial contagion. The journal of finance, 2002

[Loretan, 2000] Loretan, M., English, W. Evaluating Changes in Correlations During Periods of High Market Volatility, 2000

[Stein, 1999] Stein, B., Niggemann, O. (1999): On the Nature of Structure and its Identification. In: "Graph-Theoretic Concepts in Computer Science", Springer, LNCS, vol. 1665, pp. 122-134

[Stein, 2003a] Stein,B., Eissen, S.M. Automatic Document Categorization: Interpreting the Performance of Clustering Algorithms. In: Proceedings of the 26th German Conference on Artificial Intelligence "Advanced in Artificial Intelligence" (KI-2003), Springer, LNCS, vol. 2821, pp. 254-266, 2003

[Stein, 2003b] Stein, B., Eissen, S.M., Wissbrock, F.. On Cluster Validity and the Information Need of Users. In: Proceedings of the 3rd IASTED International Conference on Artificial Intelligence and Applications (AIA'03), Benalmadena, Spain, Acta Press, pp.216-221, 2003

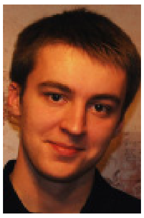
[Sakho, 2006] Sakho, Y.,S. Contagion and Firms, 2003' Internationalization in Latin America: Evidence from Mexico, Brazil, and Chile. World Bank Policy Research Working, Paper 4076, 2006

[Serwa, 2006] Serwa, D., Bohl, M. Financial Contagion Vulnerability and Resistance: A Comparison of European Capital Markets. European University Viadrina Francfurt, 2006

[Siu-Ming Cha, 2000] Siu-Ming Cha, Lai-Wan Chan. Applying independent component analysis to factor model in finance.Springer, 538-544, 2000

[Yeyati, 2008] Yeyati, E., Schmukler, S., Van Horen, N. Crisis, Capital Controls and Financial Integration. World Bank Policy Research Working Paper 4770, 2008

Authors' Information



Konstantin Dylko – *Bachelor, Academy of national economy and civil service under the President of Russia; Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia; Moscow Institute of Physics and Technology; Institutskii per. 9, Dolgoprudny, Moscow Region, 141700, Russia;*

e-mail: Kdylko@gmail.com

Major Fields of Scientific Research: data mining, stock markets, econometrics

Mikhail Alexandrov – *Professor, Academy of national economy and civil service under the President of Russia; Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia; fLexSem Research Group, Autonomous University of Barcelona, Bellaterra (Barcelona), 08193, Spain; e-mail: MAlexandrov@mail.ru*

Major Fields of Scientific Research: data mining, text mining, mathematical modeling