Country and industry effects in CEE stock market networks: Preliminary results

Vyrost, Tomas

University of Economics, Slovakia

27 July 2015

Country and industry effects in CEE stock market networks: Preliminary results

Working paper, version 27.07.2015

Tomáš Výrost

University of Economics in Bratislava, Slovakia

Abstract

In this working paper, the topic of country vs. industry effects in stock returns is explored. An approach based on stock market network modeling is used to assess both effects. Three different network subgraphs are employed: Minimum Spanning Trees, Planar Maximal Filtered Graphs and Threshold Graphs. By constructing the networks for the whole sample covering 2003 – 2012, significance of country and industry effects are shown both by visual inspection, as well as simulation and fitting of Exponential Random Graph Models. The relative importance of country/industry effects are assessed using the indicators "Relative Country Links" and "Relative Industry Links", in a rolling windows analysis covering the sample period, indicating dominance of country effects.

JEL Classification: G01, L14

Keywords: stock market networks, emerging and frontier markets, portfolio diversification.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0666-11. I would like to thank Eduard Baumöhl and Štefan Lyócsa for their ideas and input regarding this paper, as well as routines for stationarity testing.

Introduction

Since the seminal works of Markowitz (1952), many papers have been written on the topic of portfolio diversification. The exploitation of low correlation for minimizing the risk of a portfolio within the mean-variance frameworks has led to a search for asset classes (and asset groups within these classes) that would offer the best risk-reward ratios. A lengthy debate ensued on the benefits of international and cross-industry diversification. The general idea is simple – as each sector is affected differently by the business cycle, diversification across industries should be beneficial. International diversification should help even further, as there are fewer common factors and thus systematic risk should be lower. This effect however is mitigated by the development on internationalization of markets, globalization and growing market interdependencies (e.g. cross-listings of stocks and the rise of transnational companies). Thus, the puzzle of superiority of industry/international diversification remains.

This paper does not have the ambition to solve the long lasting puzzle. It focuses on the use of stock market network analysis tools to compare the two approaches. The paper analyzes the industry/country effects present in the networks constructed from stock returns of CEE-3 markets (the Czech Republic, Poland and Hungary), together with the neighbouring major stock market of Germany.

1. Related literature

1.1 International vs. industry diversification

The discussion of country/industry effects in stock returns go back as far as 1974, as Lessard (1974) states that the country effects are more important. These findings were supported by Solnik (1974), demonstrating the benefits of international diversification.

More recently, the work of Heston and Rouwenhorst (1995) have marked the beginning of a series of papers on the topic, with ambiguous outcome. Griffin and Karolyi (1998) confirm that little of the variation in country index returns can be explained by their industrial composition. Cavaglia et al. (2000) followed the analysis of Heston and Rouwenhorst, used a different return decomposition structure in their econometric treatment and expressed their view that the preference on international diversification over the industry diversification is not warranted. They described the relationship between the effects as a dynamic one, with a growing trend in favour of industry factors.

Diermeier and Solnik (2001) analyzed the proportion of domestic and foreign sales, as well as currency risk exposure. They found evidence that companies are priced globally, the location of company's headquarters is not a major determinant of stock price, and that foreign stock market exposure is more important than foreign currency exposure. Cavaglia and Moroz (2002) support the notion of related companies creating closer ties, thus aiding stronger industry links in their paper on cross-industry, cross-country allocation. Baca et al. (2002) confirm the rise of industry effects, and express their view that the findings suggest that country-based approaches to global investment management may be losing their effectiveness. In Wang et al. (2003), the authors analyze 7 equity markets and 22 industrial group returns indexes in the period of January 1990 – February 2001. Their results support the dominance of industry effects over country effects since 1999. They also find that country effects tend to show a cyclical trend.

More recently, much of the research focused on a related topic of contagion of markets, which may further reduce the meaningfulness of international diversification. In their notable paper, Forbes and Rigobon (2002) define contagion as the rise in correlation among stock market returns in time of crises, or an external shock in one of the economies. Although the literature on contagion is extensive (e.g. Bekaert et al., 2002; Kearney and Lucey, 2004; Goetzman et al, 2005; Bekaert et al, 2009 and others), we will not pursue this topic in more detail, but rather focus on the dichotomy of industry/country effects in stock returns within the context of stock market networks.

1.2 Stock market networks

Stock market network modeling is an area based on graph theory, studied in discrete mathematics. The seminal paper on this topic is by Mantegna (1999), who analyzed the constituents of Dow Jones Industrial Average and S&P500 during the period 1989 – 1995. This paper introduced several key topics: a way to define a network as a set of vertices (assets) and their relationships (return correlations) forming edges. It also solved a problem of meaningful assignment of edge weights, where (possibly also negative) correlations are transformed into distances. The problem of impracticality of working with complete graphs was solved by the proposal of using a minimum spanning tree (MST) to select a sub-graph retaining the most important edges while retaining connectivity and acyclic properties.

The research on stock market networks that followed was quite extensive, and several improvements and alternative subgraph creation algorithms have been proposed. The approach using MSTs was used e.g. on the US market by Bonanno et al. (2001), who used high-frequency data, and Vandewalle et al. (2001). The analysis of Bonanno et al. (2004) extended the analysis to the stock markets of 24 countries during the period 1988 – 1996. The paper introduced some ideas dealing with stock trading non-synchronicity. The paper by Onella et al. (2002) contributed by analyzing the dynamics of evolution of stock market networks. Their analysis of S&P500 constituents on the sample of 1982 – 2000 demonstrated the rise of correlations between stock returns, which justifies the dynamic approach. This rise is demonstrably also reflected in various network characteristics, which shows the economic meaningfulness of the network approach. They also demonstrated the relation between portfolio diversification and the so-called normalized tree length (which is a network property).

Coelho (2007) used the network analysis on the stocks constituting the FTSE 100 index. They compared the industry structure of FTSE 100 to the clustering induced by the ensuing MST, stating their similarity. Their results were in contrast with a prior analysis by Coronello et al. (2005), who used intraday data, thus providing evidence on the significance of the sampling frequency. Gilmore et al. (2008) on the sample of 21 EU countries demonstrated the central role of the older EU members, such as Germany and France. They confirm the usefulness of using MSTs. Also, the lower linkages of CEE countries suggested diversification potential. Eryiğit and Eryiğit (2009) analyzed 143 stock market indices in 59 countries during the years 1995 – 2008. Apart from the traditional MST, they also used planar maximally filtered graphs (PMFG), introduced by Tumminello et al. (2005). With both approaches leading to similar results, the authors confirm the rise in correlations in time and spatial clustering of stock indices, particularly in the daily data. Similarly, Di Matteo et al. (2010) studied PMFGS, with emphasis on centrality. The results showed the dominant position of the financial sector. More recent studies, spanning also the crisis period include the work of Dias (2012), Sandoval Jr and Franca (2012) and Sandoval Jr (2012).

Table 1: Index constituents for indices of CEE-3 and Germany

Ticker	Company	Country	Sector	NACE	Model specification	LB	LB^2
ERSTE	Erste group bank	CZE	Financial	K	ARIMA(1,1,1)-sGARCH(1,1)	0.120	0.592
PM	Philip morris CR	CZE	Consumer Goods	G	ARIMA(1,1,1)-sGARCH(1,1)	0.833	0.365
CEZ	ČEZ	CZE	Utilities	D, E	ARIMA(1,1,1)-csGARCH(1,1)	0.308	0.504
KB	Komerční banka	CZE	Financial	K	ARIMA(3,1,1)-gjrGARCH(1,1)	0.070	0.141
UNI	Unipetrol	CZE	Basic Materials	B	ARIMA(1,1,1)-eGARCH(1,1)	0.073	0.101
O2	Telefónica CR	CZE	Technology	J	ARIMA(1,1,1)-eGARCH(1,1)	0.246	0.754
EGIS	Egis pharmaceuticals	HUN	Healthcare	Q	ARIMA(1,1,1)-gjrGARCH(1,1)	0.330	0.090
EST	Est media	HUN	Services	I, R, H	ARIMA(1,1,1)-eGARCH(1,1)	0.835	0.635
MOL	MOL	HUN	Basic Materials	B	ARIMA(1,1,1)-eGARCH(1,1)	0.191	0.436
MTK	Magyar telekom	HUN	Technology	J	ARIMA(2,1,1)-csGARCH(1,1)	0.136	0.055
OTP	OTP bank	HUN	Financial	K	ARIMA(1,1,1)-eGARCH(3,1)	0.330	0.228
PAE	PannErgy	HUN	Utilities	D, E	ARIMA(1,1,1)-csGARCH(1,1)	0.278	0.595
REG	Richter Gedeon	HUN	Healthcare	Q	ARIMA(2,1,2)-sGARCH(1,1)	0.216	0.251
SYN	Synergon	HUN	Technology	J	ARIMA(1,1,1)-sGARCH(1,1)	0.109	0.276
KGHM	KGHM	POL	Basic Materials	B	ARIMA(1,1,1)-eGARCH(1,1)	0.322	0.449
PEO	Bank Polska Kasa Opieki	POL	Financial	K	ARIMA(2,1,1)-sGARCH(1,1)	0.168	0.337
PKN	Polski Kon. Naftowy Orlen	POL	Basic Materials	B	ARIMA(1,1,1)-sGARCH(2,2)	0.397	0.133
TPS	Telekomunikacja Polska	POL	Technology	J	ARIMA(1,1,1)-eGARCH(1,1)	0.070	0.065
ACP	Asseco Poland	POL	Technology	J	ARIMA(1,1,1)-sGARCH(1,1)	0.138	0.192
BHW	Bank Handl. w Warszawie	POL	Financial	K	ARIMA(1,1,1)-sGARCH(1,1)	0.457	0.059
BRE	BRE Bank	POL	Financial	K	ARIMA(3,1,5)-eGARCH(1,1)	0.059	0.330
BRS	Boryszew	POL	Basic Materials	B	ARIMA(3,1,1)-gjrGARCH(1,1)	0.067	0.632
ADS	Adidas	DEU	Consumer Goods	G	ARIMA(1,1,1)-sGARCH(1,1)	0.055	0.198
ALV	Allianz	DEU	Financial	K	ARIMA(1,1,1)-gjrGARCH(1,1)	0.520	0.187
BAS	BASF	DEU	Basic Materials	B	ARIMA(1,1,1)-csGARCH(1,1)	0.334	0.144
BMW	Bayerische Motoren Werke	DEU	Consumer Goods	G	ARIMA(1,1,1)-sGARCH(1,1)	0.434	0.209
BAYN	Bayer	DEU	Healthcare	Q	ARIMA(1,1,1)-eGARCH(1,1)	0.335	0.121
BEI	Beiersdorf	DEU	Consumer Goods	G	ARIMA(1,1,1)-sGARCH(1,1)	0.377	0.410
CBK	Commerzbank	DEU	Financial	K	ARIMA(1,1,1)-sGARCH(1,1)	0.107	0.663
CON	Continental	DEU	Consumer Goods	G	ARIMA(1,1,1)-gjrGARCH(3,2)	0.115	0.063
DAI	Daimler	DEU	Consumer Goods	G	ARIMA(1,1,1)-csGARCH(1,1)	0.515	0.084
DBK	Deutsche Bank	DEU	Financial	K	ARIMA(1,1,1)-sGARCH(3,1)	0.188	0.256
DB1	Deutsche Boerse	DEU	Financial	K	ARIMA(1,1,1)-gjrGARCH(1,1)	0.219	0.085
DPW	Deutsche Post	DEU	Services	I, R, H	ARIMA(1,1,1)-sGARCH(1,1)	0.131	0.104
DTE	Deutsche Telekom	DEU	Technology	J	ARIMA(1,1,1)-sGARCH(1,1)	0.504	0.236
EOAN	E.ON	DEU	Utilities	D, E	ARIMA(1,1,1)-sGARCH(1,1)	0.548	0.058
FME	Fresenius Medical Care	DEU	Healthcare	Q	ARIMA(1,1,1)-csGARCH(1,1)	0.111	0.132
FRE	Fresenius SE & Co KGaA	DEU	Healthcare	Q	ARIMA(1,1,1)-sGARCH(1,1)	0.487	0.256
HEI	HEICO Corporation	DEU	Industrial Goods	C	ARIMA(3,1,2)-sGARCH(1,1)	0.105	0.090
HEN3	Henkel AG & Co.	DEU	Consumer Goods	G	ARIMA(1,1,1)-sGARCH(1,1)	0.560	0.202
IFX	Infineon Technologies	DEU	Technology	J	ARIMA(1,1,1)-sGARCH(1,1)	0.366	0.321
SDF	K+S Aktiengesellschaft	DEU	Basic Materials	B	ARIMA(1,1,1)-eGARCH(1,1)	0.067	0.375
LIN	Linde Aktiengesellschaft	DEU	Basic Materials	B	ARIMA(1,1,1)-sGARCH(1,1)	0.200	0.085
LHA	Deutsche Lufthansa	DEU	Services	I, R, H	ARIMA(1,1,1)-sGARCH(2,1)	0.326	0.107
MRK	Merck KGaA	DEU	Healthcare	Q	ARIMA(2,1,2)-csGARCH(1,1)	0.088	0.498
MUV2	Munich RE	DEU	Financial	K	ARIMA(1,1,1)-gjrGARCH(1,1)	0.788	0.087
SAP	SAP	DEU	Technology	J	ARIMA(1,1,1)-sGARCH(1,1)	0.565	0.476
SIE	Siemens Aktiengesellschaft	DEU	Industrial Goods	C	ARIMA(1,1,1)-eGARCH(1,1)	0.422	0.441
TKA	ThyssenKrupp AG	DEU	Basic Materials	B	ARIMA(1,1,1)-sGARCH(1,1)	0.520	0.077
VOW3	Volkswagen	DEU	Consumer Goods	G	ARIMA(1,1,1)-eGARCH(1,1)	0.708	0.214

Notes: LB and LB^2 are the p-values for Ljung-Box test for autocorrelation in model residuals and squared residuals on first 25 lags. GARCH models used are described in more detail in Appendix 1.

2. Data and methodology

The data used in the paper encompasses the major stock market index constituents in CEE-3 markets (the Czech republic, Poland and Hungary) and Germany, with a total of $N = 50$ traded companies. Germany was selected as geographically closest major stock exchange. The CEE-3 countries also have strong real economic ties to Germany.

The sample spans the time frame January, 2003 – December, 2012. This avoids the problematic transition period before 2000, which was characterized by privatizations and market irregularities in the CEE-3 countries. The sample includes a period of market crisis and two recessions. In contrast to many other network studies, the analysis is conducted on individual stock instead of stock market indices. This better corresponds to the idea, that stock market networks should capture the structure of the analyzed markets. This also allows avoiding several potential pitfalls, such as dealing with changes in the definition of market indices (e.g. the Czech PX index replaced the prior PX-D and PX-50 indices in March 2006).

The daily prices were used to create the returns:

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}) \qquad (1)$$

where $r_{i,t}$ is return and $P_{i,t}$ market price at time $t = 1, 2, ...$ for series $i \in \{1, 2, ..., N\}$.

In order not to introduce spurious effects into the analysis, univariate ARMA-GARCH models have been fitted for all series. Table 1 gives details on all stocks from the respective markets, along with the ARMA-GARCH model specifications. The ARMA part is traditional,

$$\left(1 - \rho(L)\right)\left(1 - L\right)r_{i,t} = \left(1 + \alpha(L)\right)\varepsilon_{i,t} \qquad (2)$$

where $\varepsilon_{i,t}$ is the error term. The feasible GARCH specifications are listed in Appendix 1. The model fitting strategy was to fit ARMA-GARCH models which remove all autocorrelation from residuals and their squares, and then choose the most parsimonious model by the Bayesian information criterion (BIC).

All series were checked for stationarity (for the results of unit-root testing, see Appendix 2). The ARMA-GARCH filtering was used in order to remove all information from the series that can be explained by prior returns. When working with the standardized residuals, all other identified effects are thus unambiguously a manifestation of the relationship between series and are not induced by autocorrelation within a single series. The calculated standardized residuals are then used to construct the stock market networks.

A network is a graph G, defined by the set of vertices $V(G)$, corresponding to the traded companies, and set of edges $E(G) = \{\{u, v\}; u \neq v, u, v \in V(G)\}$. In this paper, we consider only correlation based networks, the edges are therefore undirected. However, it is useful for the edges to be weighted. The edge weights reflects the relationships of stock returns, and are given by the formula

$$c_{ij} = \sqrt{2(1-\rho_{ij})} \tag{3}$$

where c_{ij} is the edge weight for the edge connecting vertices $i,j \in V(G)$ and ρ_{ij} is the Pearson correlation coefficient between stock returns of stocks i and j.

As correlations are defined for all pairs of return series, it is theoretically possible to use them to create a complete graph on $N = 50$ vertices, having $N(N-1)/2 = 1225$ edges. The analysis of this large number of edges is not only impractical, it is also not very useful, as we are retaining many (possibly non-significant) relationships.

The literature defines several ways a suitable subgraph may be selected. In this paper, we will use three approaches:

1. Minimum spanning trees (MST) defined by Mantegna (1999). The strategy is to select a subgraph, a so-called spanning tree, with minimal overall edge weights. A spanning tree is a connected acyclic subgraph – there exists a path between any two vertices, and there are no circles. The requirement for minimal sum of edge weights means, that given the stated conditions, the subgraphs contains the highest correlations possible. Less technically, the graph retains the most important relationships under the conditions of connectedness and acyclicity. An MST has $N-1$ edges.
2. Planar maximally filtered graph (PMFG) by Tumminello et al. (2005). These subgraphs replace the condition of MST, which requires no circles to be present with a condition of planarity, which requires that the graph may be embedded in an Euclidean plane without edges intersecting. This raises the number of edges to $3N-6$, and allow for richer structures to be preserved, such as cliques of the order 4. However, the economic reasoning behind requiring planarity is unclear.
3. Threshold graphs (THR), e.g. Tse et al. (2010). Here the subgraph is created by comparing edge weights (or their transformations) to a pre-specified threshold, and retaining only those edges satisfying the threshold condition. These graphs pose no limitations on the structure of the network (unlike MST and PMFG). The threshold is usually chosen with respect to he size, or significance of the correlation coefficient between stock returns.

In this paper we analyze all three kinds of subgraphs. Apart from creating the networks, it is also interesting to construct a model, which would explain the presence/absence of edges. Particularly, it would be interesting to see how the country and industry affiliation relate to the presence of edges between individual stocks.

A framework that allows incorporating such exogenous factors into the modeling of edges is the Exponential random graph model (ERGM), as defined in the seminal work of Wasserman and Pattison (1996). Here the existence of edges and other networks structures is modeled by a logit-type model, which may (in simple cases) be modeled by maximum-likelihood estimation, or by Markov chain Monte Carlo simulations. More formally, an ERGM focuses on the probability

$$P(g = G \mid \boldsymbol{\theta}) = \frac{\exp\left(\boldsymbol{\theta}^T s(G)\right)}{c(\boldsymbol{\theta})} \tag{4}$$

where G is the constructed stock market network, g is a randomly created graph, $\boldsymbol{\theta}$ is a vector of parameters and $s(G)$ is a vector of graph characteristics, which might be node, edge and structure related (such as number of edges, vertex degrees, number of cliques etc.).

The use of ERGM opens interesting options with respect to the modeling of the network – since the network encompasses both stocks from different countries, as well as different industries, it should allow for the estimation of both the country and industry effects. Thus, it should be possible to assess whether there are country/industry effects that explain the structure and strength of the relationships between stock returns of CEE-3 countries and Germany.

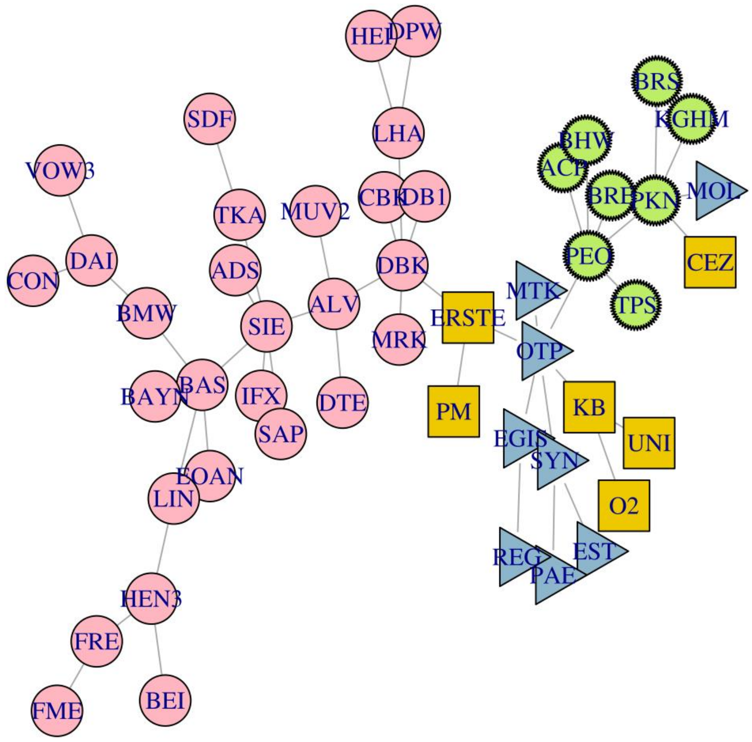


Figure 1: Minimum Spanning Tree (MST) for the stock returns from CEE-3 and Germany

Note: German stocks are color-coded pink, Poland is green, Hungary is blue and Czech stocks are yellow.

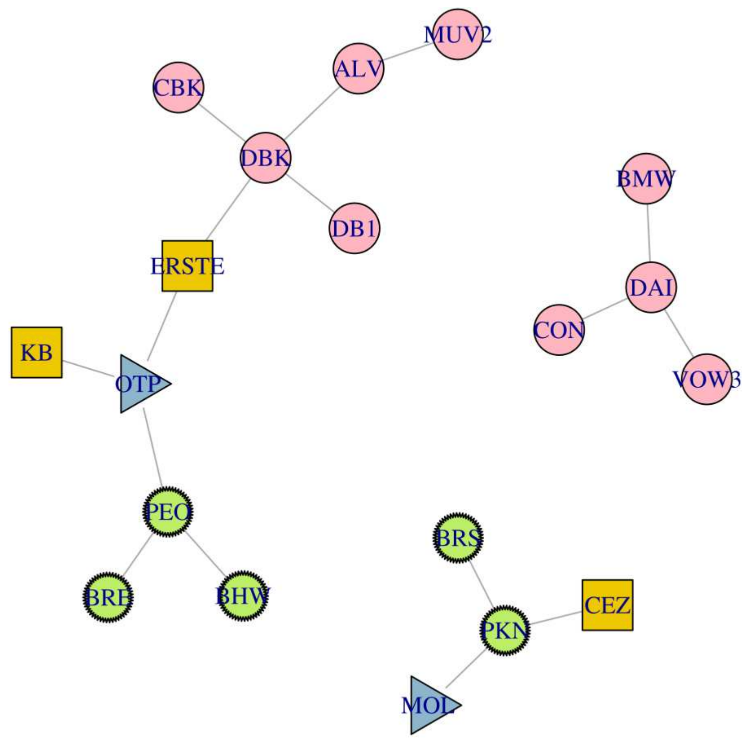


Figure 2: Selected subgraphs of the MST

Note: German stocks are color-coded pink, Poland is green, Hungary is blue and Czech stocks are yellow.

3. Empirical results and discussion

Figure 1 shows the calculated MST networks for the ARMA-GARCH filtered standardized residuals of stock returns for the whole sample period. Even after brief consideration it is clear that the network is strongly clustered by country, which is particularly true of Germany, Poland, and Hungary, with slight irregularities for the Czech Republic[1].

The MST also has subgraphs that are economically interesting. The articulation that connects all German stock to the CEE-3 stock is DBK (Deutsche Bank). It is itself connected to other German financial stocks, namely Commerzbank, Deutsche Boerse and Allianz, which is connected to Munich RE, creating a strong cluster of German financial companies.

The aforementioned DBK is connected to the Czech ERSTE Bank, which is connected to Hungarian OTP Bank, which in turn connect to two other banks – Czech Komerční banka (KB), but also Polish PEO (Bank Polska Kasa Opieki). The financial cluster is completed by adding BRE (BRE Bank, currently mBank) and BHW (Bank Handl. w Warszawie).

The financial cluster is very notable for two reasons: first, all the banks in the sample turn out as connected. This result is obtained after filtering the series with ARMA-GARCH, and then again by the algorithm creating the MST, which retains only 49 out of 1225 edges. Even then, the MST links all the banks together. This seems a rather strong evidence for clustering by industry. The second reason is, that the banks form the stocks which connect the individual country clusters – as explained before, all countries tend to create national cluster. But in all cases, these clusters are interlinked to other country cluster by stock from the financial sector, confirming its importance.

Figure 2 also shows other interesting clusters. For example, Daimler AG (DAI), Bayerische Motoren Werke (BMW), Volkswagen (VOW3) and Continental AG (CON) present a cluster of three carmakers and a company delivering components and tires to the car industry. The last selected cluster contains Polish Kon. Naftowy Orlen (PKN), Czech ČEZ (CEZ), Hungarian MOL (MOL) and Polish Boryszew (BRS), which are all oil and energy related companies.

These results clearly indicate that even though the filtering of the data might seem rather extensive, the results have reasonable economic interpretation. Industry and country clustering is also evident.

[1] Visualizations for PMFG and THR networks are shown in Appendix 3 and 4, due to their higher complexity given by the larger number of edges.

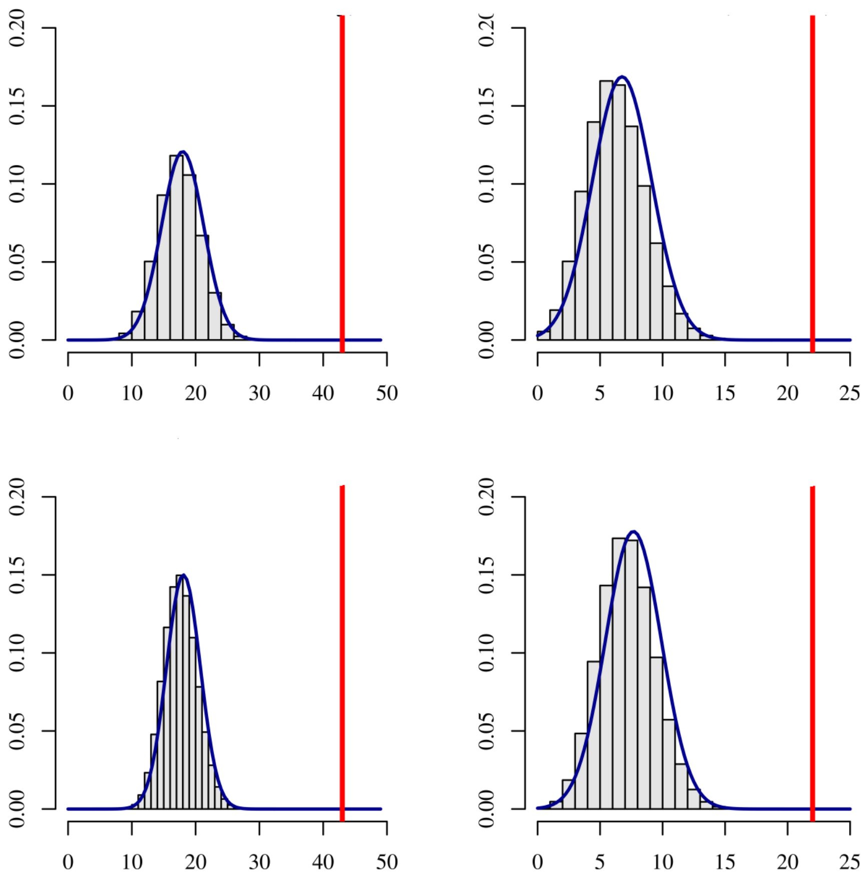


Figure 3: Simulations of random graphs and their relation to the MST

Note: The figure shows the distribution for the number of intra-country (left) and intra-industry (right) edges, obtained in Erdős and Rényi (1960; top), as well as Viger and Latapy (2005; bottom) simulations. The red lines represent the number of edges in the empirical MST.

To test this more explicitly, we note that there are 43 out of 49 edges connecting vertices from the same country, and 22 edges connecting vertices from the same industry. To see, how likely a result like this would be, if the networks were created at random, two simulations have been performed. The first was the famous Erdős and Rényi model (Erdős and Rényi, 1960). This model generates random graphs on a selected number of vertices (here, $N = 50$) and given number of edges (here, 49).

Although this may be considered a classical model, it has some disadvantages. First, the structure created in the simulation might necessarily not be a tree – while the empirical network is a MST. Also, the importance and connectivity of vertices might differ. Thus, another simulation was performed, which retains the degree sequence in all iterations (Viger and Latapy, 2005). By keeping the degree sequence constant, it follows that all generated random networks are trees, and thus precisely follow the structure of the empirical network.

Table 2: ERGM for subgraphs MST, PMFG and THR

	MST			PMFG			THR		
	Coef.	**Std. err.**		**Coef.**	**Std. err.**		**Coef.**	**Std. err.**	
Edges	-4.607	0.518	***	-3.192	0.259	***	-0.659	0.081	***
Country	2.806	0.461	***	2.349	0.241	***	2.331	0.153	***
Industry	1.958	0.327	***	1.431	0.230	***	0.647	0.190	***
Degree 1	2.715	0.574	***						
Degree 2	0.527	0.617							
Degree 3				3.230	0.562	***			
Degree 4				2.137	0.549	***			

*Note: *, **, and *** denote significance at the 10%, 5%, and 1% significance level, respectively.*

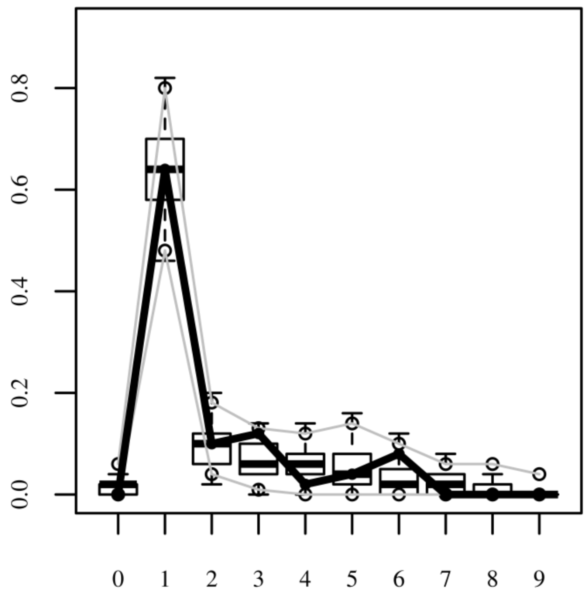


Figure 4: Relative frequency for MST ERGM models by vertex degree

Note: The vertical axis depicts relative frequency. The boxplots describe the simulations created by the specified model. The thick line shows the vertex degrees of the empirical MST.

The necessity for a simulation stems from the Cayley formula (Aigner and Ziegler, 2010), which states that the number of trees in $N = 50$ vertices equals $N^{N-2} = 50^{48}$, which is unfeasible. Figure 3 shows the simulations results, which clearly indicates the significance of both the country and industry effects.

Another way to formally test the importance of both effects is the calculation of the ERGM. Table 2 gives the results of ERGM models. The explanatory variables contain the number of edges, country and industry factors. In case of MST and PMFG, structural parameters given by the frequency of given vertex degrees were also included. The specific degrees have been chosen by the Akaike information criterion (AIC).

The results in Table 2 are again very reasonable. As all network structures have relatively few edges compared to the complete graph (the number of edges increases from MST, PMFG to THR), the coefficient by the number of edges is negative. The coefficients for Country and Industry factors are positive – hence, industry and country factor both matter, and their effect is positive.

To conclude the analysis of both effects, we have to take into account the maximum potential total number of edges that may correspond to intra-industry and intra-country links. As the number of countries and industries is not the same, moreover, the distribution between groups is not the same; the analysis conducted so far does not make the two effects comparable. To make a reasonable comparison, we introduce two measures, called RCL (Relative Country Links) and RIL (Relative Industry Links).

To define these measures, we first define the set of indices

$$IC = \{1,2,3,4\} \tag{5}$$

$$II = \{1,2,...,8\} \tag{6}$$

The values of IC (indices of countries), namely 1, 2, 3, 4 represent the Czech Republic, Germany, Hungary and Poland (in that order). The values of *II* (indices of industries), e.g. 1, 2, ..., 8 represent Energy, Financial services, Industrial goods, Services, Consumer goods, Technology, Basic materials and Healthcare (in that order).

Futher, set nk_l for $l \in IC$ the number of links from country l. Similarly, set no_t the number of stocks from individual industries.

Lastly, define $ZK(i, j, l, G)$, which for network G and any pair of vertices $i, j \in V(G)$ and country index $l \in IC$ has value 1 in case both vertices correspond to stock from country l, otherwise its value is 0. Similarly, function $ZO(i, j, t, G)$ is set to 1 if both stock corresponding to vertices i and j in network G belong to the same industry $t \in II$, and 0 otherwise.

The function *RCL* characterizes the number of edges from the same country within a network. For a given network (MST) we define $RCL_l^{MST}(G)$ for fixed $l \in IC$:

$$RCL_l^{MST}(G) = \frac{\sum_{i=1}^{V(G)-1} \sum_{j=i+1}^{V(g)} ZK(i,j,l,G)}{nk_l - 1} \tag{7}$$

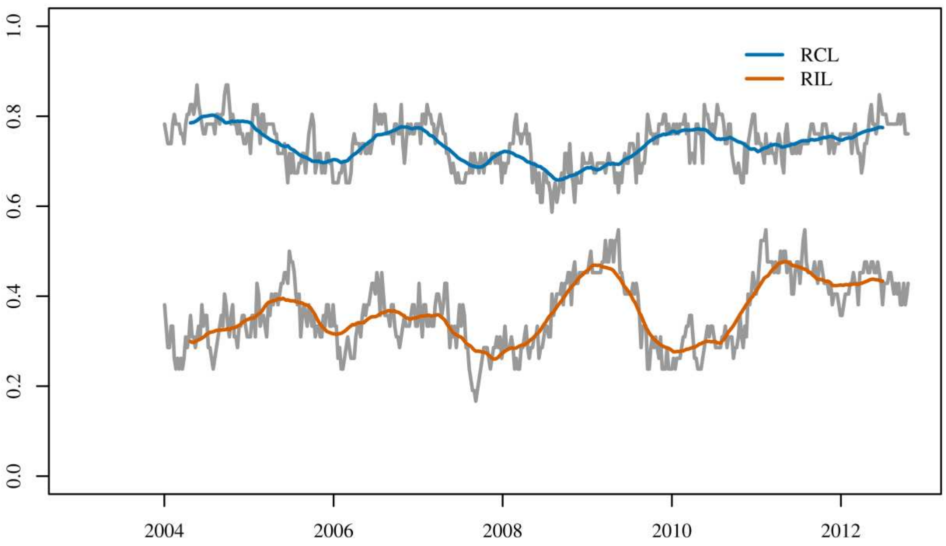


Figure 5: Evolution of *RCL* a *RIL* for rolling MST during 2003 – 2012
Note: The colored lines correspond to 32-week centered moving average.

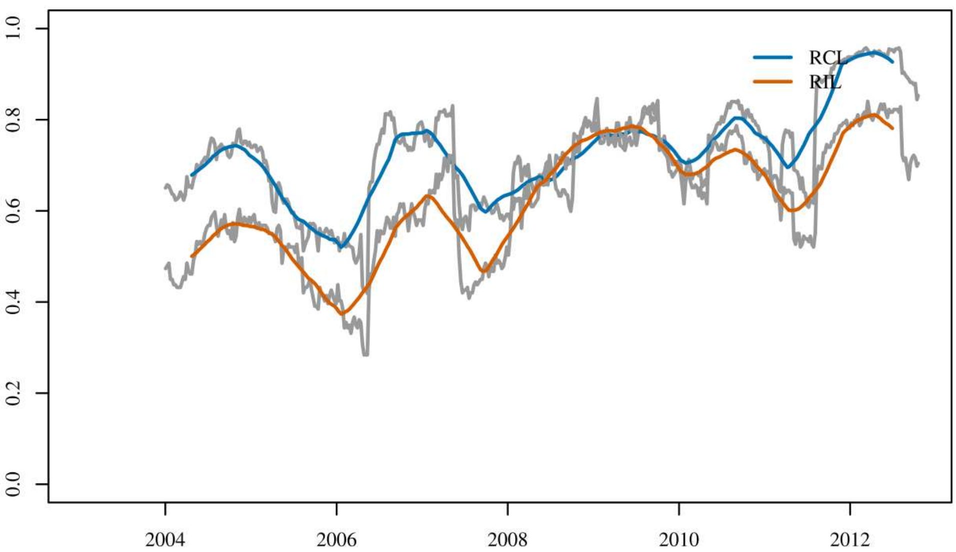


Figure 6: Evolution of *RCL* a *RIL* for rolling THR during 2003 – 2012
Note: The colored lines correspond to 32-week centered moving average.

For a network as a whole we set

$$RCL^{MST}(G) = \frac{\sum_{l=1}^{\max(IC)} \sum_{i=1}^{V(G)-1} \sum_{j=i+1}^{V(g)} ZK(i,j,l,G)}{\sum_{l=1}^{\max(IC)} \left(nk_l - 1\right)} \tag{8}$$

The previous equation follows the idea, that for every group of stocks (partitioned by the country of origin) of nk_l vertices there may be a maximum of $nk_l - 1$ edges. As we are considering a MST, the subgraph induced by the vertices belonging to the country l has to be a tree – and the maximum number of edges is thus nk_l – 1. $RCL_l^{MST}(G)$ may therefore be interpreted as a ratio of the empirical and theoretically possible number of edges. $RCL^{MST}(G)$ is not just a summation of $RCL_l^{MST}(G)$, in order to keep the interpretation of *RCL* as a percentage.

Similarly, we may define $RIL_l^{MST}(G)$:

$$RIL_l^{MST}(G) = \frac{\sum_{i=1}^{V(G)-1} \sum_{j=i+1}^{V(g)} ZO(i,j,t,G)}{no_t - 1} \quad (9)$$

$$RIL^{MST}(G) = \frac{\sum_{t=1}^{\max(II)} \sum_{i=1}^{V(G)-1} \sum_{j=i+1}^{V(g)} ZO(i,j,t,G)}{\sum_{t=1}^{\max(II)} \left(no_t - 1\right)} \quad (10)$$

The calculations for MFG and THR networks can be found in Appendix 5.

Figure 5 and 6 depict a rolling window analysis of *RCL* and *RIL* for the cases of MST a THR (results for PMFG are nearly identical to MST). The rolling analysis was conducted on the sample period of the years 2003 – 2012, the window length was one year (52 weeks). As can be seen, the difference for country/industry effects is quite dramatic for the case of MST. Empirically, country effects clearly dominate industry effects. The picture is less clear for the case of THR, where the effects are similar. As the main difference between MST and THR is mostly in the number of edges they retain (THR sometimes retains as much as half of the edges in the complete graph), we may conclude that the difference between country and industry effects is stronger when considering the most important relationships, as defined by MST. These differences tend to “average out”, as we include higher number of (potentially) less relevant link into the analysis.

4. Concluding remarks

In this working paper we explored a previously heavily researched topic of comparison of country and industry effects for portfolio diversification. Even as we do not construct stock portfolios per se, we use an alternative methodology based on stock market networks to compare these effects.

First, we use the whole sample to construct MST, PMFG and THR networks. By analyzing particularly the MST, we identify interesting relationships, providing evidence for both country and industry clustering, with the finance sector dominating the inter-country relationships. Second, the apparent clustering identified by visual inspection is shown to be significant and non-random, as shown by the results of Erdős – Rényi (1960), as well as Viger – Latapy (2005) simulations. Third, the result is also confirmed by an ERGM model, where country and industry level factors are shown to significantly contribute to the way the networks are constructed. Fourth, we define the *RIL* and *RCL* indicators in order to reasonably compare the effects of industry/country linkage. By conducting a rolling window analysis we demonstrate the differences, with country factors dominating in case of MST.

References

[1] AIGNER, M. – ZIEGLER, G.M. 2010. *Proofs from the Book*. 4th edition. Berlin: Springer, 2010. 274 p. ISBN 978-3-642-00856-6.

[2] BACA, S.P. – GARBE, B.L. – WEISS, R.A. 2000. The rise of sector effects in major equity markets. In: *Financial Analysts Journal*. ISSN 0015-198X, 2000, p. 34 – 40.

[3] BEKAERT, G. – HARVEY, C.R. – LUMSDAINE, R.L. 2002. Dating the integration of world equity markets. In: *Journal of Financial Economics*. ISSN 0304-405X, 2002, vol. 65, no. 2, p. 203 – 247.

[4] BEKAERT, G. – HODRICK, R.J. – ZHANG, X. 2009. International stock return comovements. In: *The Journal of Finance*. ISSN 0022-1082, 2009, vol. 64, no. 6, p. 2591 – 2626.

[5] BOLLERSLEV, T. 1986. Generalized autoregressive conditional heteroskedasticity. In: *Journal of econometrics*. ISSN 0304-4076, 1986, vol. 31, no. 3, p. 307 – 327.

[6] BONANNO, G. – CALDARELLI, G. – LILLO, F. – MICCICHÈ, S. – VANDEWALLE, N. – MANTEGNA, R.N. 2004. Networks of equities in financial

markets. In: *The European Physical Journal B-Condensed Matter and Complex Systems*. ISSN 1434-6028, 2004, vol. 38, no. 2, p. 363 – 371.

[7] BONANNO, G. – LILLO, F. – MANTEGNA, R.N. 2001. High-frequency cross-correlation in a set of stocks. In: *Quantitative Finance*. ISSN 1469-7688, 2001, vol. 1, no. 1, p. 96 – 104.

[8] CAVAGLIA, S. – BRIGHTMAN, C. – AKED, M. 2000. The increasing importance of industry factors. In: *Financial Analysts Journal*. ISSN 0015-198X, 2000, p. 41 – 54.

[9] COELHO, R. – GILMORE, C.G. – LUCEY, B. – RICHMOND, P. – HUTZLER, S. 2007. The evolution of interdependence in world equity markets—Evidence from minimum spanning trees. In: *Physica A: Statistical Mechanics and its Applications*. ISSN 0378-4371, 2007, vol. 376, p. 455 – 466.

[10] CORONNELLO, C. – TUMMINELLO, M. – LILLO, F. – MICCICHE, S. – MANTEGNA, R.N. 2005. Sector Identification in a Set of Stock Return Time Series Traded at the London Stock Exchange. In: *Acta Physica Polonica B*. ISSN 0587-4254, 2005, vol. 36, p. 2653.

[11] DI MATTEO, T. – POZZI, F. – ASTE, T. 2010. The use of dynamical networks to detect the hierarchical organization of financial market sectors. In: *The European Physical Journal B*. ISSN 1434-6028, 2010, vol. 73, no. 1, p. 3 – 11.

[12] DIAS, J. 2012. Sovereign debt crisis in the European Union: A minimum spanning tree approach. In: *Physica A: Statistical Mechanics and its Applications*. ISSN 0378-4371, 2012, vol. 391, no. 5, p. 2046 – 2055.

[13] DIERMEIER, J. – SOLNIK, B. 2001. Global pricing of equity. In: *Financial Analysts Journal*. ISSN 0015-198X, 2001, p. 37 – 47.

[14] ERDÖS, P. – RÉNYI, A. 1960. On the evolution of random graphs. In: *Publ. Math. Inst. Hungar. Acad. Sci,* 1960, vol. 5, p. 17 – 61.

[15] ERYIĞIT, M. – ERYIĞIT, R. 2009. Network structure of cross-correlations among the world market indices. In: *Physica A: Statistical Mechanics and its Applications*. ISSN 0378-4371, 2009, vol. 388, no. 17, p. 3551 – 3562.

[16] FORBES, K.J. – RIGOBON, R. 2002. No contagion, only interdependence: measuring stock market comovements. In: *The Journal of Finance*. ISSN 0022-1082, 2002, vol. 57, no. 5, p. 2223 – 2261.

[17] GILMORE, C.G. – LUCEY, B.M. – BOSCIA, M. 2008. An ever-closer union? Examining the evolution of linkages of European equity markets via minimum

spanning trees. In: *Physica A: Statistical Mechanics and its Applications*. ISSN 0378-4371, 2008, vol. 387, no. 25, p. 6319 – 6329.

[18] GLOSTEN, L.R. – JAGANNATHAN, R. – RUNKLE, D.E. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. In: *The journal of finance*. ISSN 0022-1082, 1993, vol. 48, no. 5, p. 1779 – 1801.

[19] GOETZMANN, W.N. – LINGFENG, L. – ROUWENHORST, K.G. 2005. Long-Term Global Market Correlations. In: *Journal of Business*. ISSN 0021-9398, 2005, vol. 78, no. 1, p. 1 – 38.

[20] GRIFFIN, J.M. – ANDREW KAROLYI, G. 1998. Another look at the role of the industrial structure of markets for international diversification strategies. In: *Journal of financial economics*. ISSN 0304-405X, 1998, vol. 50, no. 3, p. 351 – 373.

[21] HESTON, S.L. – ROUWENHORST, K.G. 1995. Industry and country effects in international stock returns. In: *The Journal of Portfolio Management*. ISSN 0095-4918, 1995, vol. 21, no. 3, p. 53 – 58.

[22] HIGGINS, M.L. – BERA, A.K. 1992. A class of nonlinear ARCH models. In: *International Economic Review* 1992, p. 137 – 158.

[23] HOBIJN, B. – FRANSES, P.H. – OOMS, M. 2004. Generalizations of the KPSS-test for stationarity. In: *Statistica Neerlandica*. ISSN 0039-0402, 2004, vol. 58, no. 4, p. 483 – 502.

[24] KEARNEY, C. – LUCEY, B.M. 2004. International equity market integration: Theory, evidence and implications. In: *International Review of Financial Analysis*. ISSN 1057-5219, 2004, vol. 13, no. 5, p. 571 – 583.

[25] LESSARD, D.R. 1974. World, national, and industry factors in equity returns. In: *The Journal of Finance*. ISSN 0022-1082, 1974, vol. 29, no. 2, p. 379 – 391.

[26] MANTEGNA, R.N. 1999. Hierarchical structure in financial markets. In: *The European Physical Journal B-Condensed Matter and Complex Systems*. ISSN 1434-6028, 1999, vol. 11, no. 1, p. 193 – 197.

[27] MARKOWITZ, H. 1952. Portfolio selection. In: *The journal of finance*. ISSN 0022-1082, 1952, vol. 7, no. 1, p. 77 – 91.

[28] ONNELA, J.P. – CHAKRABORTI, A. – KASKI, K. – KERTESZ, J. – KANTO, A. 2003. Asset trees and asset graphs in financial markets. In: *Physica Scripta*. ISSN 0031-8949, 2003, vol. 2003, no. T106, p. 48.

[29] PANTULA, S.G. 1986. Comment. In: *Econometric Reviews* 1986, vol. 5, no. 1, p. 71 – 74.

[30] SANDOVAL JR, L. 2012. Pruning a minimum spanning tree. In: *Physica A: Statistical Mechanics and its Applications*. ISSN 0378-4371, 2012, vol. 391, no. 8, p. 2678 – 2711.

[31] SANDOVAL, JR, L. – FRANCA, I.D.P. 2012. Correlation of financial markets in times of crisis. In: *Physica A: Statistical Mechanics and its Applications*. ISSN 0378-4371, 2012, vol. 391, no. 1, p. 187 – 208.

[32] SCHWERT, G.W. 1990. Stock volatility and the crash of'87. In: *Review of financial Studies* 1990, vol. 3, no. 1, p. 77 – 102.

[33] SOLNIK, B.H. 1974. Why not diversify internationally rather than domestically? In: *Financial analysts journal*. ISSN 0015-198X, 1974, p. 48 – 54.

[34] SUL, D. – PHILLIPS, P.C.B. – CHOI, C.-Y. 2005. Prewhitening Bias in HAC Estimation. In: *Oxford Bulletin of Economics and Statistics*. ISSN 1468-0084, 2005, vol. 67, no. 4, p. 517 – 546.

[35] TSE, C.K. – LIU, J. – LAU, F. 2010. A network perspective of the stock market. In: *Journal of Empirical Finance*. ISSN 0927-5398, 2010, vol. 17, no. 4, p. 659 – 667.

[36] TUMMINELLO, M. – ASTE, T. – DI MATTEO, T. – MANTEGNA, R.N. 2005. A tool for filtering information in complex systems. In: *Proceedings of the National Academy of Sciences of the United States of America*. ISSN 1091-6490, 2005, vol. 102, no. 30, p. 10421 – 10426.

[37] VANDEWALLE, N. – BRISBOIS, F. – TORDOIR, X. 2001. Non-random topology of stock markets. In: *Quantitative Finance*. ISSN 1469-7688, 2001, vol. 1, no. 3, p. 372.

[38] VIGER, F. – LATAPY, M. Efficient and simple generation of random simple connected graphs with prescribed degree sequence. In: *Computing and Combinatorics*. [s.l.]: Springer, 2005, p. 440 – 449.

[39] WANG, C.J. – LEE, C.H. – HUANG, B.N. 2003. An analysis of industry and country effects in global stock returns: evidence from Asian countries and the US. In: *The Quarterly Review of Economics and Finance*. ISSN 1062-9769, 2003, vol. 43, no. 3, p. 560 – 577.

[40] WASSERMAN, S. – PATTISON, P. 1996. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and *p**. In: *Psychometrika.* ISSN: 0033-3123, 1996, vol. 61, no. 3, p. 401 – 425.

[41] ZAKOIAN, J.-M. 1994. Threshold heteroskedastic models. In: *Journal of Economic Dynamics and control* 1994, vol. 18, no. 5, p. 931 – 955.

Appendix

Appendix 1: Specifications of the fitted GARCH models

sGARCH (Bollerslev, 1986)	$\sigma_t^2 = \left(\omega + \sum_{j=1}^{m}\zeta_j v_{jt}\right) + \sum_{j=1}^{q}\alpha_j\varepsilon_{t-j}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2,$
eGARCH (Nelson, 1991)	$\log_e(\sigma_t^2) = \left(\omega + \sum_{j=1}^{m}\zeta_j v_{jt}\right) + \sum_{j=1}^{q}\left(\alpha_j z_{t-j} + \gamma_j\left(\lvert z_{t-j}\rvert - E\lvert z_{t-j}\rvert\right)\right) + \sum_{j=1}^{p}\beta_j\log_e\left(\sigma_{t-j}^2\right)$ $E\lvert z_t\rvert = \int_{-\infty}^{\infty}\lvert z\rvert f(z,0,1,...)dz$
gjrGARCH (Glosten, 1993)	$\sigma_t^2 = \left(\omega + \sum_{j=1}^{m}\zeta_j v_{jt}\right) + \sum_{j=1}^{q}\left(\alpha_j\varepsilon_{t-j}^2 + \gamma_j I_{t-j}\varepsilon_{t-j}^2\right) + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2,$ $I_{t-j} = \begin{cases}0, & \varepsilon_{t-j} > 0\\ 1, & \varepsilon_{t-j} \le 0\end{cases}$
APARCH (Ding, 1993)	$\sigma_t^\delta = \left(\omega + \sum_{j=1}^{m}\zeta_j v_{jt}\right) + \sum_{j=1}^{q}\alpha_j\left(\lvert\varepsilon_{t-j}\rvert - \gamma_j\varepsilon_{t-j}\right)^\delta + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^\delta$ • sGARCH (Bollerslev, 1986): $\delta = 2, \gamma_j = 0$. • avGARCH (Schwert, 1990): $\delta = 1, \gamma_j = 0$. • gjrGARCH (Glosten et al., 1993): $\delta = 2$. • tGARCH (Zakoian, 1994): $\delta = 1$. • Nonlinear ARCH (Higgins, 1992): $\gamma_j = 0, \beta_j = 0$. • Log ARCH (Geweke, 1986; Pantula, 1986): $\delta \to 0$.
fGARCH (Hentschel, 1995)	$\sigma_t^\lambda = \left(\omega + \sum_{j=1}^{m}\zeta_j v_{jt}\right) + \sum_{j=1}^{q}\alpha_j\sigma_{t-1}^\lambda\left(\lvert z_{t-j} - \eta_{2j}\rvert - \eta_{1j}\left(z_{t-j} - \eta_{2j}\right)\right)^\delta + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^\lambda$ • sGARCH (Bollerslev, 1986): $\lambda = \delta = 2, \eta_{1j} = \eta_{2j} = 0$. • avGARCH (Schwert, 1990): $\lambda = \delta = 1, \lvert\eta_{1j}\rvert \le 1$. • gjrGARCH (Glosten, 1993): $\lambda = \delta = 2, \eta_{2j} = 0$. • tGARCH (Zakoian, 1994): $\lambda = \delta = 1, \eta_{2j} = 0, \lvert\eta_{1j}\rvert \le 1$. • nGARCH (Higgins, 1992): $\delta = \lambda, \eta_{1j} = \eta_{2j} = 0$. • naGARCH (Engle, 1993): $\delta = \lambda = 2, \eta_{1j} = 0$. • APARCH (Ding et al., 1993): $\delta = \lambda, \eta_{2j} = 0, \lvert\eta_{1j}\rvert \le 1$. • ALLGARCH (Hentschel, 1995): $\delta = \lambda$.
csGARCH (Lee, 1999)	$\sigma_t^2 = q_t + \sum_{j=1}^{q}\alpha_j\left(\varepsilon_{t-j}^2 - q_{t-j}\right) + \sum_{j=1}^{p}\beta_j\left(\sigma_{t-j}^2 - q_{t-j}\right)$ $q_t = \omega + \rho q_{t-1} + \phi\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right)$

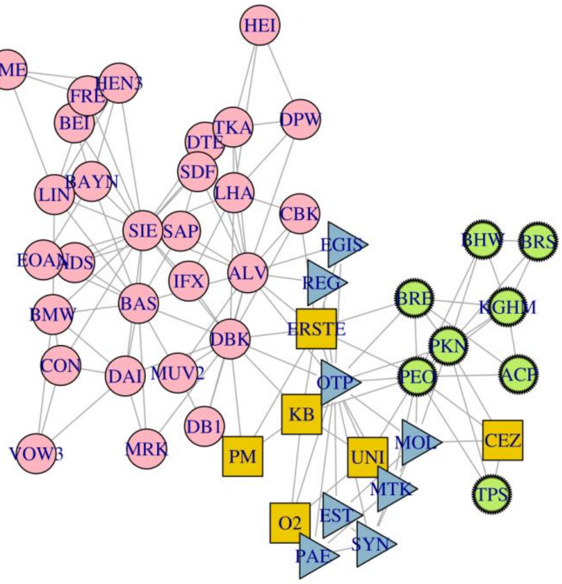
Note: v_{jt} are exogenous regressors and ε_{t-j} are random deviates from the selected probability distribution (Normal, Student or GED).

Appendix 2:Unit root testing results

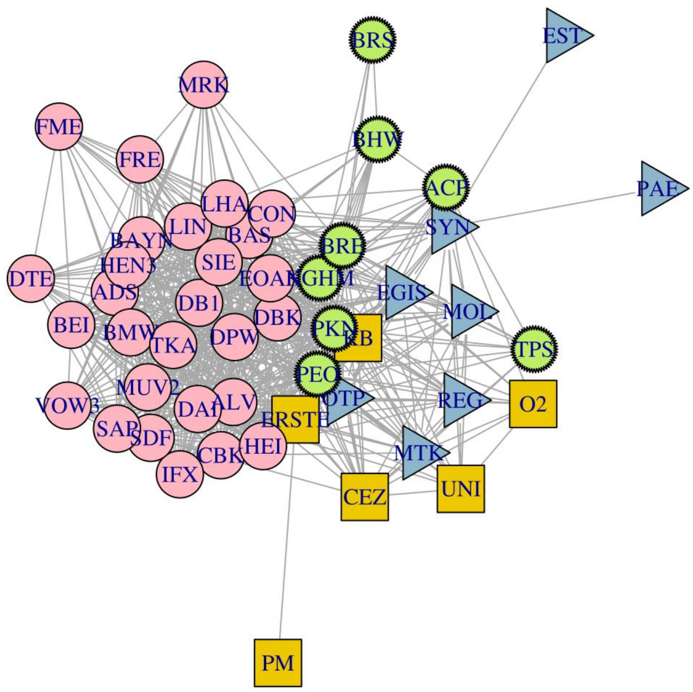
	BW	Hobijn		Sul			BW	Hobijn		Sul	
ERSTE	6	0.289		0.239		BMW	5	0.080		0.081	
PM	6	0.132		0.128		BAYN	5	0.062		0.061	
CEZ	8	0.298		0.268		BEI	4	0.133		0.129	
KB	7	0.045		0.045		CBK	8	0.397	*	0.393	*
UNI	7	0.218		0.222		CON	8	0.143		0.149	
O2	8	0.230		0.238		DAI	6	0.077		0.075	
EGIS	7	0.079		0.076		DBK	7	0.198		0.196	
EST	5	0.172		0.161		DB1	6	0.232		0.148	
MOL	5	0.174		0.165		DPW	4	0.186		0.177	
MTK	6	0.104		0.108		DTE	7	0.061		0.058	
OTP	7	0.179		0.174		EOAN	8	0.353	*	0.351	*
PAE	6	0.225		0.193		FME	12	0.102		0.101	
REG	3	0.142		0.155		FRE	2	0.201		0.180	
SYN	4	0.206		0.182		HEI	9	0.245		0.183	
KGHM	8	0.055		0.056		HEN3	5	0.243		0.241	
PEO	5	0.153		0.158		IFX	6	0.080		0.076	
PKN	7	0.164		0.153		SDF	4	0.076		0.069	
TPS	12	0.241		0.240		LIN	6	0.085		0.085	
ACP	6	0.279		0.289		LHA	6	0.144		0.142	
BHW	3	0.070		0.071		MRK	8	0.145		0.113	
BRE	5	0.105		0.105		MUV2	7	0.068		0.069	
BRS	9	0.127		0.091		SAP	5	0.094		0.092	
ADS	11	0.157		0.159		SIE	3	0.072		0.066	
ALV	5	0.097		0.098		TKA	8	0.141		0.138	
BAS	7	0.052		0.052		VOW3	6	0.073		0.071	

Note: Column BW denotes the bandwidth parameter in the estimate of covariance matrices. The columns denoted "Hobijn" give the test statistics for the test defined in Hobijn et al. (1994). Critical values for 10%, 5% and 1% significance are 0.348, 0.460 and 0.754. The columns denoted "Sul" give the test statistic for the test defined in Sul et al. (2005). Critical values for 10%, 5% and 1% significance are 0.347, 0.463 and 0.739. No statistics are significant at 5%.

Appendix 3: Planar maximally filtered graph (PMFG)



Appendix 4: Threshold graph (THR)



Appendix 5: Definitions of *RCL* and *RIL* for graphs PMFG and THR

$$RCL_l^{PMFG}(G) = \frac{\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZK(i,j,l,G)}{3nk_l - 6} \qquad RCL^{PMFG}(G) = \frac{\sum_{l=1}^{\max(IC)}\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZK(i,j,l,G)}{\sum_{l=1}^{\max(IC)}\left(3nk_l - 6\right)}$$

$$RIL_t^{PMFG}(G) = \frac{\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZO(i,j,t,G)}{3no_t - 6} \qquad RIL^{PMFG}(G) = \frac{\sum_{t=1}^{\max(II)}\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZO(i,j,t,G)}{\sum_{t=1}^{\max(II)}\left(3no_t - 6\right)}$$

$$RCL_l^{THR}(G) = \frac{\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZK(i,j,l,G)}{nk_l(nk_l - 1)/2} \qquad RCL^{THR}(G) = \frac{\sum_{l=1}^{\max(IC)}\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZK(i,j,l,G)}{\sum_{l=1}^{\max(IC)}\left(nk_l(nk_l - 1)/2\right)}$$

$$RIL_t^{THR}(G) = \frac{\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZO(i,j,t,G)}{no_t(no_t - 1)/2} \qquad RIL^{THR}(G) = \frac{\sum_{t=1}^{\max(II)}\sum_{i=1}^{V(G)-1}\sum_{j=i+1}^{V(g)} ZO(i,j,t,G)}{\sum_{t=1}^{\max(II)}\left(no_t(no_t - 1)/2\right)}$$