Country effects in CEE3 stock market networks: a preliminary study

Výrost, Tomáš

Faculty of Business Economics in Košice, University of Economics in Bratislava

29 December 2012

Country effects in CEE3 stock market networks: a preliminary study

Tomáš Výrost

Abstract

The stock markets in the Czech Republic, Poland and Hungary (CEE3) are studied in the context of stock market networks. A total of 17 shares are followed during the period of 1998 – 2012. The daily returns are used for calculation of rolling correlations of various window lengths. The resulting correlation matrices are then used to construct network models. Minimum spanning trees (MST) are used as a form of abstraction in the graph structure, and their evolution is studied over time. The main objective of the paper is to test whether the individual assets cluster in the MSTs by the country to which they belong or whether the origin is of lesser importance, leading to cross-country links within the MSTs. The latter might hint at increasing integration within CEE3 stock markets. We find that at the beginning of the series, the MSTs exhibited very strong country clustering, which changed in the later 2000s. The country effects do not seem to be synchronized between all markets.

Keywords: stock market networks, minimum spanning trees, stock market integration

JEL Classification: L14, G10

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0666-11.

Introduction

The analysis of networks has increased significantly in the last decade. The topic is studied in many different settings, including transportation problems, biology and economics. Within the field of economics in general, and finance in particular, one of the key areas that uses network models is the study of market structure. Here, networks are represented by graphs in which the vertices represent entities (or agents) and the edges represent the relationships between them.

When used in the context of stock markets, it is natural to assign vertices to individual assets traded on a given exchange and to create edges between those that exhibit some form of association. Because the correlation coefficient represents the most basic quantitative measure of association, stock market networks are often constructed from covariance matrices of stock returns. This conceptually simple approach has some drawbacks. One of these drawbacks is the fact that because we can measure correlation between any series of returns, there is always an edge between any two vertices (assets). Thus, the structure becomes a complete graph: if there are $n \in \mathbb{N}$ vertices, there will be $n(n-1)/2$ edges.

To make the analysis more tractable, there are several ways to reduce the number of edges, preferably keeping only the most important ones. One of the possible approaches is to use some sort of threshold to create an edge. This threshold may be based on statistical significance or some other rule. Another approach that has been used frequently is the use of minimum spanning trees (MSTs, see, e.g., Mantegna, 1999), which form a subgraph of the correlation-based network. Because MSTs are trees, there are exactly $n-1$ edges on n vertices, which is a significant reduction.

Despite their prevalent use, MSTs continue to have some issues that make the results of an analysis of MSTs difficult to interpret.

First, a large number of edges are dropped, possibly discarding a large number of highly correlated pairs of assets. One may object to the principle of the creation of MSTs, in which the number of edges not only drops from $n(n-1)/2$ to $n-1$ but this number actually stays the same regardless of the value of the correlations in the graph (which is worse). The reasoning for keeping an a priori given number of edges seems to reflect a "rule of thumb". Many such rules may be generally good but less appropriate in a particular situation. In some cases, this method may drop a number of large correlations; in other cases, it may retain even small correlations.

Second, MST (as a tree) has a property of connectedness; that is, it represents a connected graph. To maintain this property, some of the edges must to be kept in the MST regardless of their value. If we have a single vertex unrelated to the rest (which are correlated) in an MST, there must be a retained edge connecting it to them.

When we observe the two properties together, we conclude that we not only exclude the majority of edges but also do not keep the $n - 1$ largest correlation coefficients. The MST thus selects the largest correlations provided that they result in a spanning tree.

Despite these difficulties, our paper is based on the use of MSTs as a simple tool for a preliminary analysis of CEE3 stock market networks. The main objective of this paper is to investigate the country effects in the clustering of assets within the CEE3 stock market networks and to observe their development over time.

Our objectives follow a simple idea. CEE3 stock markets are all located in post-communist countries that have experienced significant changes. The stock markets in these countries were established in the early 1990s. As we analyze the stock returns for 1998 – 2012, it is reasonable to assume that early in their development, there had not yet been time for the markets to establish themselves and form strong cross-country bonds. However, as they developed, incentives for greater stock market integration grew stronger. The reasons for this are numerous: all CEE3 countries have joined the European Union, thus making their legislation more uniform; they have dropped the barriers on capital flows; and, in general, they have experienced an increase in the links of economies at the macroeconomic level. Over time, some cross-listed companies have even been added to the Polish market (Hungarian Mol added in 12/2004 and E-Star added in 03/2011).

As noted below, if we consider the stock market networks constructed from CEE3 companies, we might expect them to initially have strong intra-country links and to subsequently have stronger inter-country links. If we model the network as an MST, in which all vertices are connected, this would correspond to a tree in which we would initially see branches made of companies from the same country. If the country effects subsequently weaken, we could expect the edges in a MST to link more companies from different countries.

To verify this hypothesis, the MST represents a suitable network structure despite our earlier cautionary notes. The first problem with MSTs is the sheer volume of edges we drop when constructing an MST. The second problem involves the implications of creating a connected graph. In our setting, it is desirable to force a connected structure on a network. It would be rather farfetched to believe that the markets had no relation to each other; hence, they should be connected, even though the correlations may be small. Keeping only a small number of edges allows us to identify a change in country effects. To create a different MST with inter-country links, the correlations to assets in other markets must become stronger than the correlations to those in a domestic market. Thus, although the problem with the interpretation of MSTs is not solved in general, this structure serves us well in the context of our problem (for other studies on MSTs, see, e.g., Bonanno *et al.*, 2001; Vandewalle *et al.*, 2001; Onnela *et al.*, 2003; Onnela *et al.*, 2004; Tabak *et al.*, 2010; Tse *et al.*, 2010; Lee *et al.,* 2012 and Bonanno *et al.*, 2003).

1. Data and methodology

The data we use for our analysis consist of 17 stocks from three exchanges: PM, CEZ, KB, UNI and O2 are traded at the Prague Stock Exchange (Czech Republic); EGIS, PAE, REG, OTP, MOL and MTK are traded at the Budapest Stock Exchange (Hungary); and BRE, HAND, KGHM, BPOL, ASC and TEL are traded at the Warsaw Stock Exchange (Poland). All stocks are constituents of their national stock indices, PX, BUX and WIG-20.

Although the dataset could be enhanced to include more recent assets, our objective was to study the development of country effects over time, which requires a large time series. Because the new additions to listed companies would cause a shorter common trading window, for the purposes of this preliminary analysis, we use only the stocks that traded between December 1, 1998 and October17, 2012 (thus, the number of observations is $N + 1 = 3622$) [1]. This leaves us with the above-mentioned 17 stocks (we denote this as $M = 17$).

For all series, we first calculate continuous returns[2] as

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

where $r_{t,i}$ are the continuous returns, $P_{t,i}$ are daily prices, $i = 1, 2, \ldots, M$ denotes the asset and $t = 1, 2, \ldots, N$. The use of daily data avoids the problems described by Baumöhl – Lyócsa (2012).

To obtain correlations that can be observed over time, we calculate rolling correlations based on two window sizes: 150 and 500 days. To calculate the rolling correlation coefficient between assets i and j at time t using a window with length $w \in \{150, 500\}$, we use the well-known formula

$$\bar{r}_{t,i} = \frac{1}{w}\sum_{s=0}^{w-1} r_{t-s,i}\,, \qquad i = 1, 2, \ldots, M \qquad (2)$$

[1] To avoid ambiguity, please note that we denote the $N + 1 = 3622$ daily prices as $P_{0,i}, P_{1,i}, \ldots, P_{N,i}$ for $i = 1, 2, \ldots, M$. Also note that when referring to the assets, the index follows the ordering introduced at the beginning of the section: PM, CEZ, KB, UNI, O2, EGIS, PAE, REG, OTP, MOL, MTK, BRE, HAND, KGHM, BPOL, ASC and TEL.

[2] For a thorough discussion about stationarity of the series used, see Lyócsa – Baumöhl (2012).

$$\rho_{t,i,j}^{w} = \frac{\sum_{s=0}^{w-1}\left(r_{t-s,i} - \bar{r}_{t,i}\right)\left(r_{t-s,j} - \bar{r}_{t,j}\right)}{\sqrt{\sum_{s=0}^{w-1}\left(r_{t-s,i} - \bar{r}_{t,i}\right)^2}\sqrt{\sum_{s=0}^{w-1}\left(r_{t-s,j} - \bar{r}_{t,j}\right)^2}} \tag{3}$$

Using the previous formula, it would be straightforward to calculate correlation matrices for all t (except the first w). However, the usual approach is not to use correlations directly. When creating graphs, the correlations are usually transformed into distances by the formula

$$d_{t,i,j}^{w} = \sqrt{2(1-\rho_{t,i,j}^{w})} \tag{4}$$

This has the consequence of low correlations being transformed into high $d_{t,i,j}^{w}$ and vice versa. The distances are then stacked into a matrix, which we denote $\mathbf{D}_t^w$. We continue to use Prim's algorithm (see, e.g., Prim, 1957; Kruskal, 1956 and Papadimitrou and Steigliz, 1982) to construct a minimum spanning tree[3] MST_t^w from the distances $\mathbf{D}_t^w$. As shown, we obtain $N - w$ minimum spanning trees for both window sizes. Each MST_t^w has the same set of vertices $V(MST_t^w) = \{1, 2, ..., M\}$ but a varying set of edges $E(MST_t^w)$ specific to each MST.

To ascertain the role of the country effect, we define for any subset of vertices $U \subset V(MST_t^w)$ with $|U| \geq 2$ the measure we call *relative country links* (*RCL*), given by[4]

$$RCL(U, MST_t^w) = \frac{\left|\{(u,v); u,v \in U \wedge (u,v) \in E(MST_t^w) \wedge u < v\}\right|}{|U| - 1} \tag{5}$$

The nominator of the formula for *RCL* gives the number of edges in the subgraph of MST_t^w based on the vertex set U. Because MST_t^w is a tree, the maximum number of edges in the subgraph is $|U| - 1$, that is, the number of vertices less one. However, the subgraph based on U may not be a tree; in that case, there will be less than $|U| - 1$ edges. The *RCL* thus takes a value between 0 and 1 and can be thought of as a percentage of the maximum possible number of edges that exist between vertices U within MST_t^w.

The logic behind the introduction of *RCL* is straightforward. Because we expect the MSTs in the beginning of the series to be clustered by countries, there should be a high number of edges between assets from the same country, and their *RCL* should be high. In contrast, when the country

---

[3] For an alternative approach based on dynamic conditional correlations instead of rolling correlations, see Lyócsa *et al.* (2012).

[4] We denote the number of elements in U as $|U|$. This is the same as the cardinality of U because we deal with finite graphs.

effect is less pronounced, we expect the number of edges within the same vertices to drop, and the *RCL* also drops.

To extend our analysis, we first note that the number of *RCL* indicators to be calculated is quite large; we obtain a new MST for all t in $w, w+1, \ldots, N$. Within each MST, we should calculate *RCL* for three groups of vertices corresponding to assets from the three analyzed markets. This gives us a time-series for each w and country, which describes the magnitude of the country effect and its evolution over time.

However, although this result may provide a visual clue about the country effect, it would be helpful to have a quantitative measure of its significance. For example, the maximum number of edges between Czech stocks is four (because there are five stocks). If, in a given MST, there are three edges between these vertices, how frequently would three edges occur if there is no country effect and the choice of vertices constituting our selected group is, in fact, as good as random?

The question can be reformulated as follows: given the same MST_t^w, what would be the distribution of *RCL* if we selected a random group of five vertices? If having three edges between a random group of vertices is very common, then having three edges between five Czech stocks does not seem significant.

To verify that this is the case, the usual approach would be to perform a simulation. Because we are dealing with a fairly small graph of seventeen vertices, the number of all possible combinations of picking five (for the Czech Republic) and six (for Hungary and Poland) vertices is given by

$$\binom{17}{5} = 6\,188 \qquad \binom{17}{6} = 12\,376 \tag{6}$$

Because these numbers are not high, we can proceed by directly evaluating all combinations. In fact, the task is not easy because we must do this for all MSTs (there are approximately 3600 MSTs for each window size).

2. Results and discussion

As described in the previous section, we begin by calculating the rolling correlations, which are later transformed into distances upon which we construct the minimum spanning trees. A visual examination (Figure 1) of individual MSTs shows that our initial hypothesis of the presence of country effects seems valid: the MSTs in the beginning of our sample show perfect clustering by countries. However, if we look at the more recent graphs, there is some variation in the way the individual stocks are linked, including several inter-country edges.

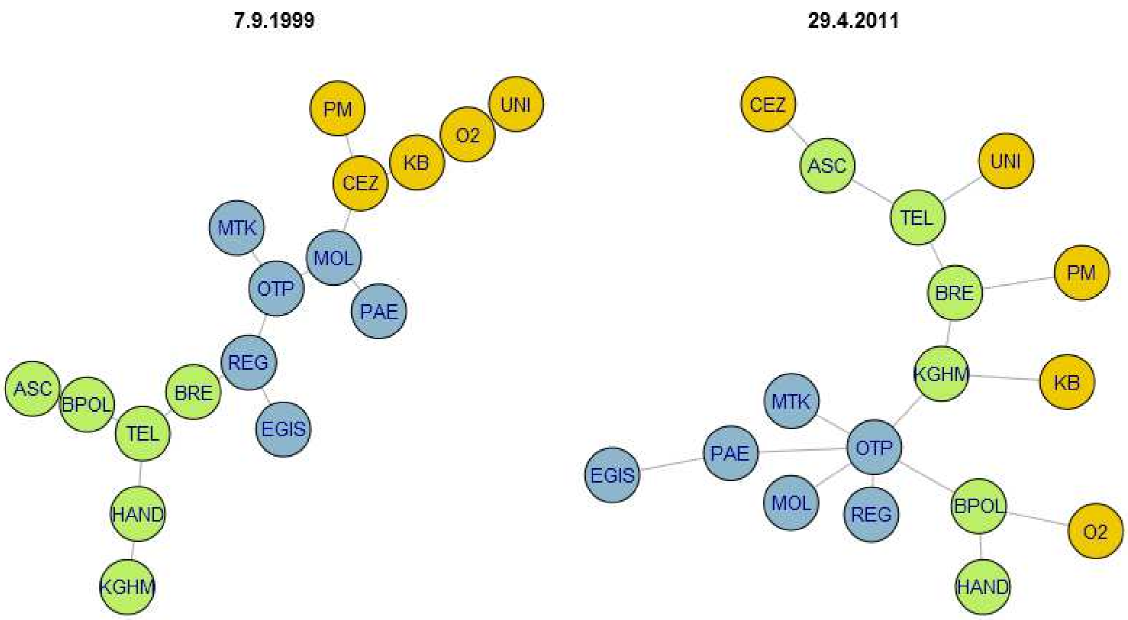

Figure 1: Sample MSTs in 1999 (left) and 2011 (right)

These preliminary thoughts can be extended by capturing the evolution of *RCL* we have defined earlier. Recall that a high level of *RCL* corresponds to a high country clustering. As shown in Figure 2, the early observations overwhelmingly contain high *RCL*, supporting country-specific clustering. We also see that the high *RCL*s are quite persistent and hold almost until 2002 (occasionally missing one edge). However, there are significant deviations after this date. The Hungarian and Polish stocks seem to form fewer links in the period of 2004 – 2008, and the Czech stocks do so after 2010. The correlation coefficients between the *RCL*s of the Czech Republic and Hungary are -0.097, the Czech Republic and Poland are 0.16, and Poland and Hungary are 0.27.

The volatility of the *RCL,* as shown in Figure 2, can be explained in two ways. First, we see that the *RCL* seems to change up and down quite rapidly. This behavior can be explained by the relative magnitude of the correlation coefficient of assets within and between respective countries.

If the correlation coefficients of two assets from the same country are close to those of a correlation with an asset from a different country, then a single observation may tip the scales in favor of one or the other. If the correlation coefficients are close, then even minor shifts from one day to another may induce a slightly different structure of the MST.

The other possible explanation lies in the calculation of correlation coefficients. It is reasonable to assume that the volatility of the *RCL* is inversely proportional to the length of the windows used (w). Because averaging reduces the variance, the behavior of the *RCL* may be related to a particular choice of w. Rather than simply calculating one window size w = 150, we also add a calculation based on w = 500. The resulting *RCL*s are shown in Figure 3.

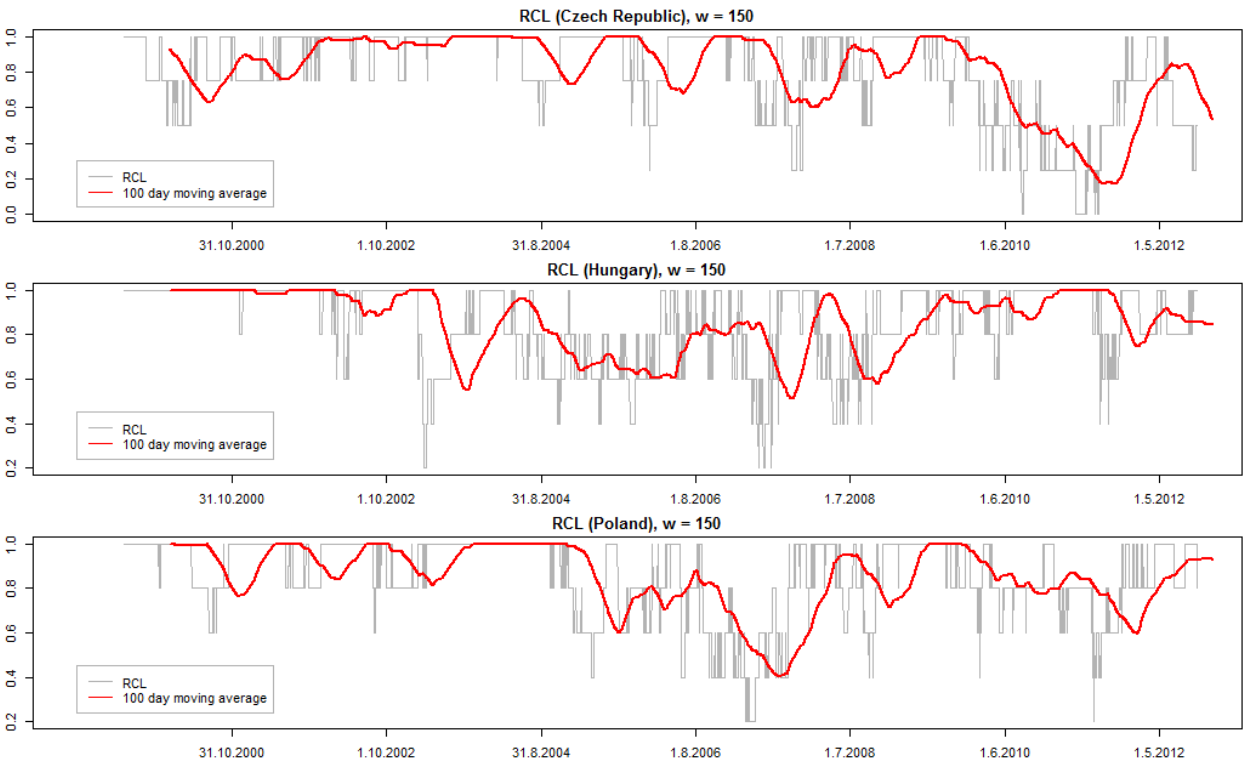

Figure 2: *RCL* for the smaller window size, $w = 150$

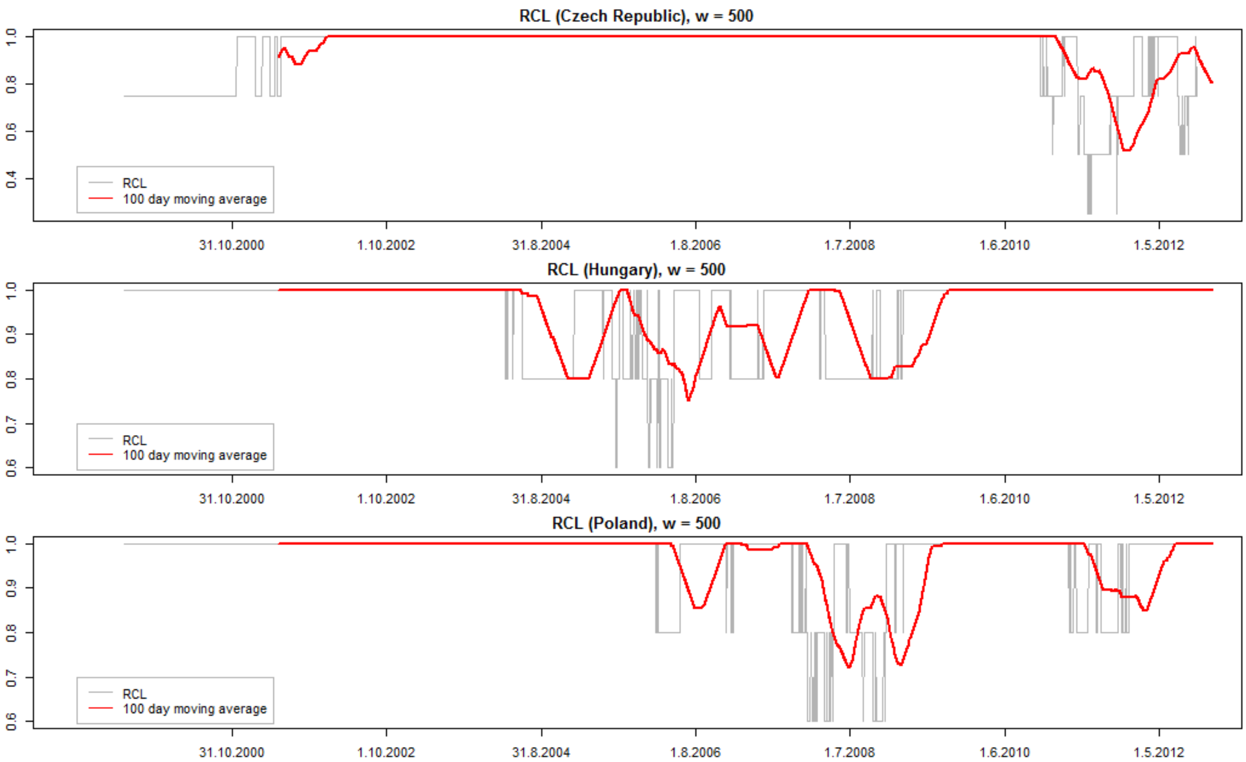

Figure 3: *RCL* for the larger window size, $w = 500$

We see that the resulting pattern is very similar, and the volatility of the *RCL* is reduced. However, the decrease in its values is still present in the data. Because a window of 500 observations is quite long and should not be affected by short-term irregularities, we may conclude that there is substantial evidence for a change in the magnitude and overall effect of country-specific factors. As in the previous case, we observe that periods with lower inter-country links intersect for Hungary and Poland, but less so for the Czech Republic (there is similar development with Poland during the last part of the sample range).

The last step in our analysis is the estimation of the rarity of this clustering of vertices. To assess this, we explore all combinations of vertices for all MSTs in our sample. This allows us to construct an empirical distribution function of the number of edges within randomly selected vertices. Under the null hypothesis of random clustering, we can estimate the quantile of the number of vertices present for a single country in a specific MST.

Figure 4 shows the quantiles based on such empirical distribution functions. It can be observed that the number of links between vertices from the Czech Republic is rather high initially but declines after 2011. In some cases, the drop is rather spectacular because in some instances, the number of within-country links is zero.

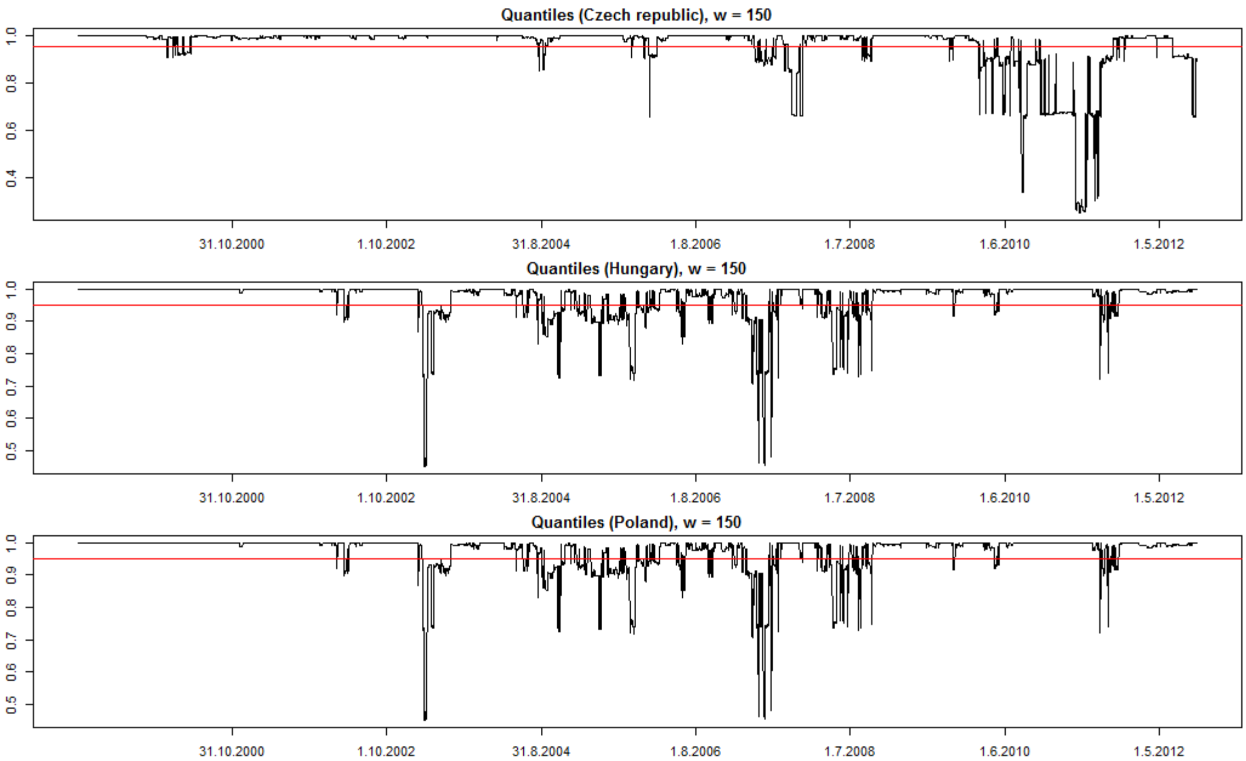

Figure 4: Quantiles of the number of links, w = 150

3. Conclusion

In this paper, we present a brief analysis of the prevalence of country effects in the stock market networks of CEE3 countries. We approach the problem by constructing correlation-based minimum spanning trees and subsequently evaluate the degree to which the MSTs retain edges between vertices corresponding to stocks traded in the same country.

We have shown that at the beginning of the series, the clustering is high, and the stocks traded in the same country form branches in the MSTs. However, we see a marked difference in the way stocks are linked within the MSTs during the middle (Hungary and Poland) and the end of the sample (Czech Republic), in some cases dropping from the largest possible number of edges to zero.

The results indicate that the relationships within the CEE3 stock returns are not constant, at least within the employed modeling framework. As we have noted previously, calculating minimum spanning trees may pose difficulties in the interpretation of the results obtained because they impose a fairly rigid structure in a graph. However, in our case, the choice has quite an interesting application: by forcing a large reduction of edges from an originally complete graph, it is interesting to see which edges persist. Because the weights of the edges are essentially correlations (more precisely, monotonic transformations of correlations), the growth in the number of inter-country links is indicative of weaker country effects in the data.

The analysis presented in this paper can be considered only a preliminary study for a number of reasons. The first reason for the preliminary status of this paper is the lack of a clear justification for the behavior of the MSTs. There is no clear reason for the drop in within-country links. One might speculate that the behavior of the stock markets can be connected to the European sovereign-debt crisis or similar processes, but the paper provides no evidence for a particular reason for the correlations. We do not pursue this point because of the sample size issue with regard to the number of stocks analyzed.

This brings us to the second reason for considering the results preliminary. The analysis is performed on long-term data, which, on one hand, allow for the examination of the evolution of country effects. On other hand, these data pose a severe restriction on the number of assets that are traded on CEE3 exchanges during the entire 14-year period. This poses some limitations to the analysis. For example, for the Czech Republic, we have five vertices with at most four links in all MSTs. Because the number of edges in a random subgraph on five vertices is limited to five possible numbers of edges (zero to four), we obtain only a very rough estimate of the quantile of the number of links. This could be improved by including more stocks in the sample, which would have the consequence of having to use shorter time intervals. This should not pose a problem because, as we show in the paper, the MSTs retain full branches of country-specific clusters for periods at the beginning, and most of the dynamics occur during the more recent years in the sample.

References

[1] Baumöhl, E. – Lyócsa, Š. 2012. Constructing weekly returns based on daily stock market data: A puzzle for empirical research? MPRA Working Paper 43431, University Library of Munich, Germany.

[2] Bonanno, G. – Caldarelli, G. – Lillo, F. – Mantegna, R. N. 2003. Topology of correlation based minimal spanning trees in real and model markets. *Physical Review E*, 68(4), 046130.

[3] Bonanno, G. – Lillo, F. – Mantegna, R. N. 2001. High-frequency cross-correlation in a set of stocks. *Quantitative Finance*, 1(1), 96–104.

[4] Kruskal, J. B. 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1), 48–50.

[5] Lee, J. – Youn, J. – Chang, W. 2012. Intraday volatility and network topological properties in the Korean stock market. *Physica A*, 391(4), 1354–1360.

[6] Lyócsa, Š. – Baumöhl, E. 2012. Testing the covariance stationarity of CEE stocks. MPRA Working Paper 43432, University Library of Munich, Germany.

[7] Lyócsa, Š. – Výrost, T. – Baumöhl, E. 2012. Stock market networks: The dynamic conditional correlation approach. *Physica A*, 391(16), 4147–4158.

[8] Mantegna, R. N. 1999. Hierarchical structure in financial markets. *The European Physical Journal B*, 11, 193–197.

[9] Onnela, J. P. – Chakraborti, A. – Kaski, K. – Kertész J. 2003. Dynamic asset trees and black Monday. *Physica A*, 324(1), 247–252.

[10] Onnela, J.P. – Kaski, K. – Kertész, J. 2004. Clustering and information in correlation based financial networks. *The European Physical Journal B,* 38, 353–362.

[11] Papadimitrou, C.H. – Steigliz, K. 1982. *Combinatorial Optimization: Algorithms and Complexity.* Prentice-Hall, Englewood Cliffs, NJ.

[12] Prim, R. C. 1957. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36, 1389–1401.

[13] Tabak, B. M. – Serra, T. R. – Cajueiro, D. O. 2010. Topological properties of stock market networks: the case of Brazil. *Physica A*, 389(16), 3240–3249.

[14] Tse, Ch.K. – Liu, J. – Lau, F. C. M. 2010. A network perspective of the stock market. *Journal of Empirical Finance*, 17(4), 659–667.

[15] Vandewalle, N. – Brisbois, F. – Tordoir, X. 2001. Non-random topology of stock markets, *Quantitative Finance*. 1(3), 372–374.