THE ROYAL INSTITUTE OF TECHNOLOGY
MASTER THESIS

The Market Graph

A study of its characteristics, structure & dynamics

Students:

David Jallo
djallo@kth.se

Daniel Budai
dbudai@kth.se

December 2010

Supervisors:

Associate Professor Harald Lang

Distinguished Professor Panos M. Pardalos

Abstract

In this thesis we have considered three different market graphs; one solely based on stock returns, another one based on stock returns with vertices weighted with a liquidity measure and lastly one based on correlations of volume fluctuations. Research is conducted on two different markets; the Swedish and the American stock market. We want to introduce graph theory as a method for representing the stock market in order to show that one can more fully understand the structural properties and dynamics of the stock market by studying the market graph. We found many signs of increased globalization by studying the clustering coefficient and the correlation distribution. The structure of the market graph is such that it pinpoints specific sectors when the correlation threshold is increased and different sectors are found in the two different markets. For low correlation thresholds we found groups of independent stocks that can be used as diversified portfolios. Furthermore, the dynamics revealed that it is possible to use the daily absolute change in edge density as an indicator for when the market is about to make a downturn. This could be an interesting topic for further studies. We had hoped to get additional results by considering volume correlations, but that did not turn out to be the case. Regardless of that, we think that it would be interesting to study volume based market graphs further.

Sammanfattning

I denna uppsats har vi tittat på tre olika marknadsgrafer; en enbart baserad på avkastning, en baserad på avkastning med likvidviktade noder och slutligen en baserad på volymkorrelationer. Studien är gjord på två olika marknader; den svenska och den amerikanska aktiemarknaden. Vi vill introducera grafteori som ett verktyg för att representera aktiemarknaden och visa att man bättre kan förstå aktiemarknadens strukturerade egenskaper och dynamik genom att studera marknadsgrafen. Vi fann många tecken på en ökad globalisering genom att titta på klusterkoefficienten och korrelationsfördelningen. Marknadsgrafens struktur är så att den lokaliserar specifika sektorer när korrelationstaket ökas och olika sektorer är funna för de två olika marknaderna. För låga korrelationstak fann vi grupper av oberoende aktier som kan användas som diversifierade portföljer. Vidare, avslöjar dynamiken att det är möjligt att använda daglig absolut förändring i bågdensiteten som en indikator för när marknaden är på väg att gå ner. Detta kan vara ett intressant ämne för vidare studier. Vi hade hoppats på att erhålla ytterligare resultat genom att titta på volymkorrelationer men det visade sig att så inte var fallet. Trots det tycker vi att det skulle vara intressant att djupare studera volymbaserade marknadsgrafer.

Acknowledgements

We would like to thank Dr. Panos M. Pardalos and Dr. Petraq Papajorgji at the Center for Applied Optimization, University of Florida for inviting us and giving us the chance to write our thesis abroad. A special thanks to Dr. Pardalos for his insightful comments, interesting discussions and professional guidance throughout the process. We would also like to thank our supervisor Harald Lang at the Institution of Mathematical Statistics, KTH for making this thesis possible.

Table of Contents

1 Introduction

In this chapter we introduce the reader to the subject and present the purpose and the outline of the thesis.

In today's world of seemingly endless information, one often faces challenges of dealing with large sets of data when trying to solve different problems. In many cases, these massive data sets can be represented by large network structures, or graphs. These graphs have specific properties and attributes which, if studied properly, provide us with a lot of information about the applications they portray. Using graphs to represent real world dynamics is common in many different fields such as military systems and technology, ecology, telecommunications, medicine and biotechnology, astrophysics, geographical systems and finance. An example of a data set which can be represented by a graph is telecommunications traffic data. In that graph, the vertices are telephone numbers which are connected through edges if a call has been made from one number to another. Other examples of graph representations are the Internet and the human brain (1).

In this thesis we will concentrate our efforts on a graph representation of the stock market, called the market graph. Since the stock market lacks physical connections between stocks it is not at all obvious how the market can be represented. Nonetheless, a somewhat intuitive representation of the stock market can be based on the correlations of stock price movements. Another approach that we will introduce later on in the report is the correlations of volume fluctuations between stocks. Hence, a market graph can be constructed by letting each stock be represented by a vertex and let two vertices be connected by an edge if the correlation coefficient of the stock pair exceeds a pre-specified threshold. We want to introduce graph theory as a method for representing the stock market in order to show that one can more fully understand the structural properties and dynamics of the stock market by studying the market graph.

Being inspired by the article '*Statistical analysis of financial networks*' (2), we mainly want to construct two market graphs; both of which are based on stock returns but with the difference that one has vertices which are weighted with a liquidity measure. After having presented the theory that will be used throughout this thesis in chapter 2, we will analyze the characteristics of both the unweighted and the weighted graph in chapter 3. In order to get more reliable results we will consider two markets with distinguishable difference in size, namely the Swedish market and the American market. Chapter 4 is about the structure of the market graph where we implement means of data mining using optimization algorithms to find clusters within the graphs. By splitting our data into different time periods we get more dynamics in our research which is presented in chapter 5. This way we can extract information about the dynamics of the market graphs by studying some of their properties and structures. This chapter will also contain an analysis of the market graph based on volume correlations since it has to our knowledge not yet been investigated. We hope that the information obtained from it might help us understand the market from another point of view. We will conclude the thesis in chapter 6 where we will summarize our results and have a brief discussion and also mention how this subject can be studied further.

2 Theoretical study

Important parts from the field of graph theory as well as the mathematical algorithms used are presented in this part of the thesis.

2.1 Basic definitions and notations

A graph $G = (V, E)$ consists of a nonempty *vertex set* V and an *edge set* E. If $e \in E$ is an edge and $v, w \in V$ are different vertices such that $e = \{v, w\}$, then v and w are said to be *adjacent* i.e. two vertices are adjacent if they share a common edge. Similarly, two edges are adjacent if they share a common vertex. The *degree* of a vertex v in a graph G, denoted by k, is the number of edges of G incident with v, each loop counting as two edges. If k is an even number then v is said to be an *even* vertex; if k is odd the vertex is said to be *odd*, and if $k = 0$ then v is called an *isolated* vertex. (3)

For every integer number k one can calculate the number of vertices $n(k)$ with degree equal to k and then get the probability that a vertex has degree k as $P(k) = \frac{n(k)}{n}$, where n is the total number of vertices and the function $P(k)$ is referred to as the *degree distribution* of the graph (4).

Edge density if defined as the number of edges of a graph divided by the total number of possible edges in the graph: $\frac{number\ of\ edges}{n(n-1)/2}$, where n is the number of vertices of the graph G.

2.2 Weighted graph

A *weighted graph* $G = (V, E)$ is a graph in which each vertex v is assigned a nonnegative real number $w(v)$ called the *weight* of v. The *weight of the graph* $\mathcal{G}$, denoted by $w(\mathcal{G})$, is the sum of the weights of all vertices. Weighted graphs are often used when practical problems are modeled with means of graph theory. Throughout this thesis the weights will be represented by the liquidity of the stocks. (3)

2.3 Power-law and scale invariance property

A quantity x obeys a power-law if it is drawn from a probability distribution

$$p(x) \propto x^{-\alpha}$$

or equivalently,

$$\log p(x) \propto -\alpha \log x$$

where α is a constant parameter of the distribution known as the *exponent or scaling parameter* (5).

Power laws have different properties and the main one is perhaps the *scale invariance property*

$$p(cx) \propto c^{-\alpha} x^{-\alpha}$$

where c is a constant. That is, we get a proportional relationship where the original power-law relation is multiplied by the scaling factor $c^{-\alpha}$.

In our case, where y is the number of nodes with degree x, the power-law graph model $P(\alpha, \beta)$ is according to (1) defined as,

$$y = e^{\alpha}/x^{\beta}$$

or equivalently

$$\log y = \alpha - \beta \log x$$

2.4 Clusters, cliques, quasi-cliques and independent sets

Clusters are groups of data such that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Since we in this thesis define vertices as stocks, we consider vertices to be “more similar” the higher the correlation is between them. Often, one distinguishes to which cluster a specific vertex belongs by measuring its distance to the rest of the vertices in the cluster; in this case however, distance can be substituted by correlation. (6)

A clique is a fully connected graph; i.e. a subset of a graph's vertices $C \subseteq V$ such that every two vertices in the subset are connected by an edge (see Figure 2). Considering the market graph, a clique would characterize a group of highly correlated and interrelated stocks such as a specific industry. Moreover, a clique is referred to as being maximum if the graph contains no larger clique and it is called maximal if the clique cannot be extended to a larger clique. We will in this thesis only focus on maximum cliques (MC). Another set of interest is the maximum independent set (MIS). It is defined as a set of vertices no two of which are connected. More formally, it can be depicted as a clique in the complementary graph $\bar{G}$. Since this basically is the complete opposite of a clique, seeing as the vertices are negatively correlated, it is natural to interpret the set as a possible diversified portfolio. (3)

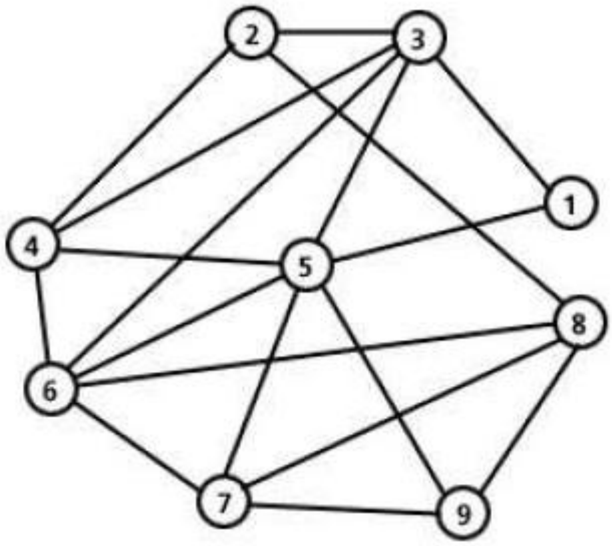

Figure 1. A graph with 9 vertices.

Quasi-cliques are special kinds of clusters that either have a constraint on minimum vertex degree or minimum edge density. Hence, a quasi-clique can be defined in two different ways. With a constraint on minimum vertex degree, a quasi-clique is defined in the following way: Let $S \subseteq V(G),\ |S| = k,$ be the set of vertices of the subgraph G_S we wish to find. Then, the set of vertices S is a γ-quasi-clique $(0 < \gamma \leq 1)\ if\ \forall\ v \in S,$ $deg^{G_S}(v) \geq \lceil \gamma \cdot (k-1) \rceil$ i.e. a sub-graph that satisfies the user-specified minimum vertex degree bound $\lceil \gamma \cdot (k-1) \rceil$. As a special case, a γ-quasi-clique is a fully connected graph, or a clique, when $\gamma = 1$. (7)

If on the other hand one would consider a γ-quasi-clique as being a cluster with a minimum constraint on edge density, the definition would be the same as above but with the constraints that the graph G_S has to be connected and $|E(G_S)| \geq \gamma \binom{k}{2}$, i.e. the number of edges of the graph G_S has to be greater than some number dependent on γ and the number of vertices in the graph (8).

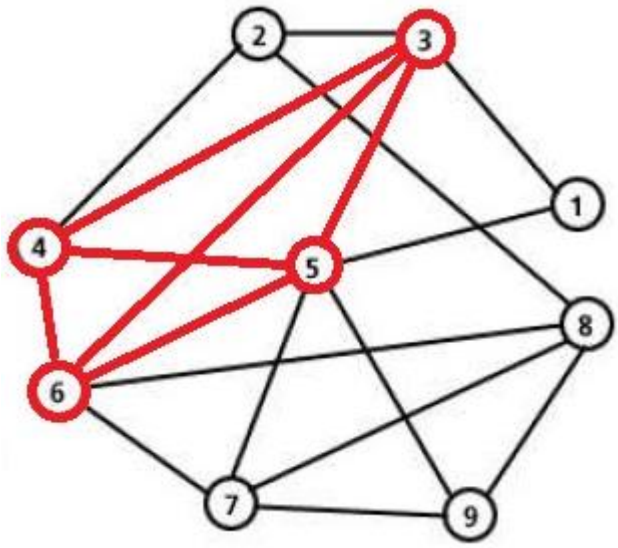

Figure 2. A fully connected sub-graph, or clique, within a graph, highlighted in red.

2.5 Liquidity

Most people have an intuitive feeling about what liquidity is but not many can state how it should be mathematically defined. Linguistically, liquidity can be defined as *"the probability that an asset can be converted into an expected amount of value within an expected amount of time"* (9). In the context of this thesis however, a more suitable definition of liquidity is *"the ability to convert shares into cash (and the converse) at the lowest transaction costs"* (10). There is no consensus in the academic community exactly how to mathematically quantify the aforementioned definitions, but two common measures are the bid-ask spread and the turnover rate. The bid-ask spread is simply the difference between the bid price, the price people are willing to sell a specific share for at time t, and the ask price, the price people are willing to buy a specific share for at time t. The second most common measure, the turnover rate, is defined as

$$Turnover\ Rate_t = \frac{number\ of\ shares\ traded_t}{number\ of\ outstanding\ shares_t} \quad (10)$$

The notion is that investors expect a higher rate of return when investing in illiquid assets since the transaction costs for them are higher than for their more liquid counterparts. This price premium is evident in markets in the form of the bid-ask spread where the prices include premiums for immediate buying and immediate selling (11). Therefore, the bid-ask spread can be regarded as the price one pays for liquidity and hence, the lower the spread the more liquid the asset is considered to be. The situation is the converse in the case of the turnover rate, meaning that a higher turnover rate implies a higher liquidity.

There are many articles that analyze different proxies for liquidity and also the relationship between liquidity and stock returns, but unfortunately their results are not conclusive. One of the bigger reasons for that is because the different researchers use different measures, or proxies, in their attempts to quantify liquidity. (11), (12) and (10) have all conducted empirical investigations in the matter and found that the bid-ask spread measure has yielded inconclusive results as a proxy for liquidity while the turnover rate measure, although not as prevalently used, has led to more stable and uniform results. This is especially true for quote-driven markets such as the NYSE, NASDAQ and OMX. In light of that evidence we will in this thesis use the turnover rate as defined above as our proxy for liquidity.

2.6 Maximum weighted clique- and Maximum weighted independent set problem

We use the following formulation for the MWC-problem (13):

$$\text{Max} \sum_{i=1}^{n} \omega_i x_i$$

Subject to:

$$x_i + x_j \leq 1, \forall\, (i,j) \in \bar{E}$$

$$x_i \in \{0,1\}, i = 1, \dots, n$$

where:

$$x_i = \begin{cases} 1, \mathit{if\ node\ i\ belongs\ to\ maximum\ clique} \\ 0\ \ \mathit{otherwise} \end{cases}$$

As a special case, if the graph is unweighted, we set all the weights $\omega_i = 1$. The maximum weighted independent set (MWIS) problem is equivalent to the MWC problem in the complementary graph and solving it will give the independent set with largest weight for a given graph.

2.7 Constructing the market graph

The construction of the market graph is quite simple. We let each vertex represent a stock and for any pair of vertices i and j, an edge is connecting them if the corresponding correlation coefficient $C_{ij} \in [-1\,,1]$, based on the returns of instruments i and j, is greater than or equal to a specified threshold $\theta \in [-1\,,1]$. Now, let $P_t(t)$ denote the price of the instrument i on day t. Then

$$R_i(t) = \ln\left(\frac{P_i(t)}{P_i(t-1)}\right)$$

defines the logarithm of the return of instrument i over the one-day period from $(t-1)$ to t. The correlation coefficient between instruments i and j is calculated as

$$C_{ij} = \frac{E(R_iR_j)-E(R_i)E(R_j)}{\sqrt{Var(R_i)Var(R_j)}},$$

where $E(R_i)$ is the average return of stock i over n days, i.e. $E(R_i) = \frac{1}{n}\sum_{t=1}^{n} R_i(t)$.

2.8 Algorithms

2.8.1 NP-hard

That a problem is NP-hard means, among other things, that it cannot be solved exactly using in polynomial time. All exact algorithms therefore have exponential runtimes which makes the solution process that much more difficult (2). However, algorithms for the maximum clique problem utilize the clique's downward closure property, i.e. the fact that every subset of a clique also is a clique. This piece of information makes it possible to construct efficient algorithms for the maximum clique problem. Unfortunately though, the downward closure property does not hold for finding maximum quasi-cliques (MQC) which means that the algorithms for finding exact solutions are much less efficient.

2.8.2 Heuristic algorithm

In order to get good starting points for the exact solution method used, we implemented a fast heuristic method to get approximate solutions to the MC-problem. A heuristic algorithm will usually not produce the optimal solution. However, a close to optimal solution is often found within a fraction of the time it takes to run an exact algorithm.

The heuristic algorithm used in this thesis is called the Vertex Support Algorithm (14). It is designed to solve the minimum vertex cover problem, which actually is equivalent to solving the MC- or MIS-problem. A vertex cover is defined as a set of vertices such that each edge of the graph is incident to at least one vertex in the set. A minimum vertex cover is a vertex cover of smallest possible size.

The algorithm works in the following way: First we calculate the degree and support of every vertex in the graph. The support of a vertex is defined as the sum of the degrees of its neighbors. The vertex with the largest support is then added to the vertex cover and is subsequently removed from the graph. If two or more vertices have equivalent maximum support we add the one with the largest degree to the vertex cover. This continues iteratively and when no edges between the vertices are left we have found our minimum vertex cover. The MIS or MC, depending on if you look at the graph or its complement, is then the vertices which are not in the minimum vertex cover.

2.8.3 Algorithm for Maximum clique and Maximum weighted clique

To solve the MC-, MIS-, MWC- and MWIS-problem, Pardalos' and Carraghan's exact algorithm was used (15). However, to speed up the algorithm, we implemented a preprocessing procedure which utilizes the results from the heuristic algorithm. Since we know that the exact solution for the MC-problem is larger than or equal to the heuristic result, we can remove all vertices in the graph which has degree smaller than the heuristic MC size since they obviously cannot be a part of the MC. This will significantly reduce the problem size and speed up the calculations. The exact algorithm works in the following way:

We start with one vertex, v_1, and we look for all vertices adjacent to it. When those nodes are found, we look for all nodes adjacent to the ones we just found that were adjacent to v_1. This is done iteratively until we find all cliques containing $v_1$and then we simply pick out the largest clique containing that vertex and save it as our current best clique. Next, we remove $v_1$from the graph and go through the same procedure with the next node, v_2. Since this algorithm uses brute force to find the MC it would not be efficient unless some pruning strategies were implemented. The pruning strategies help to speed up the search in two ways. First of all, every time a new clique is found it is compared to the current best clique in order to find out if it is larger. If it is, we save the new clique as the current best clique and discard the old one. Say that we are about to evaluate a new vertex at some step in the search and our current best clique consists of ten vertices. If the vertex we are about to evaluate only has 9 neighbors or less, we know that it cannot be part of a clique larger than our current best one since the largest clique it can be a part of is of size ten. Therefore we skip that vertex altogether and go on to the next. Furthermore, suppose that we have a graph containing 100 vertices, that we at the moment have searched through 70 of them and that our current best clique is of size 32. Now we only have 30 vertices left to go through, since we remove every evaluated vertex from the graph, and hence there is no possibility that we will find a clique larger than our current best one which is of size 32. Because of that, we will not look through the last 30 vertices and we have found the MC. The weighted counterpart of the algorithm works in the same way except that it prunes based on the weight and not the degree of the remaining vertices. For the interested reader, VB code for the two algorithms can be found on the internet (16).

2.8.4 Quasi-clique algorithm

The exact optimization problem for finding MQC is very difficult, as well as computationally challenging, to solve. The main problem is one concerning memory and computational time. To be able to run the optimization on a single desktop computer one would have to write a very efficient program, where as little memory as possible is needed in every step of the calculation. In addition, the time it would take to find a globally optimal solution is too long because of two reasons; first of all since the problem is more complex than the MC-problem, which is NP-hard, and secondly because it doesn't satisfy the same closure property as the MC-problem does. Therefore, a greedy randomized adaptive search procedure (GRASP) with a local search algorithm will be used instead of an exact algorithm. GRASP is an iterative method that constructs a random solution, i.e. a clique, at each iteration, and then searches for a locally optimal solution in the neighborhood of the created clique (17). This way one cannot know how good QC one obtains. However, this is not a problem since the goal is not to find exact solutions but rather to identify different sectors and find larger independent sets.

In the beginning of the algorithm, a vertex is randomly chosen from a list of vertices that all have degrees greater than some threshold. This vertex will serve as the start of the clique. The next vertex to be added to the clique is chosen based on a similar list wherein all vertices are adjacent to the first chosen vertex whilst their degrees are greater than some new threshold. This procedure is repeated until no more candidates can be found and then we have a found a solution. Now we implement a local search procedure in order to improve the solution. The local search creates a better solution by randomly choosing a vertex from the previously obtained solution, removing it from the clique, and then adding two or more new vertices that are connected to all vertices in the remaining clique. This continues as long as it is possible to find such vertices that, if they are removed, can be replaced by two or more other vertices to improve the solution.

The only difference between our algorithm for MQC and the one explained above for cliques is that the constraint on the vertices one adds is relaxed. Instead of demanding that they are connected to all the other vertices it is sufficient that they are connected to at least $\lceil \gamma \cdot (k-1) \rceil$ of the vertices in the clique. This will furthermore guarantee that the new solution's edge density is at least

$$\gamma \cdot \binom{k}{2}, \quad 0 < \gamma \leq 1, \quad k = \text{number of vertices in the QC.}$$

Since there is a constraint on the degree of each vertex, instead of on the edge density, no undesirable QC with high edge density but including vertices with only one connection to the rest of the QC will be found. Therefore, we are ensured to only find MQC of good quality.

3 Characteristics of the market graph

In this part of the thesis we explore and analyze some traits of the market graph.

3.1 Data

In order to present some characteristics of the market graph and study its dynamics we have used stock returns from October 20, 2008 to October 15 2010. Research will be conducted on two different markets; the Swedish stock market OMX and the American stock market consisting of NASDAQ, AMEX and NYSE. This will give us the possibility to compare the different markets as well as get better and more reliable results. 266 stocks have been collected for the Swedish market for 500 consecutive trading days and 5700 stocks for the American market for 502 consecutive trading days, the two additional days being due to differences in holidays. Although there is a loss in the amount of American securities, since data was not available for some of them, we believe that we have enough data to get reliable and consistent results.

3.2 Clustering coefficient

The clustering coefficient is a probability measure that quantifies the probability that the nodes adjacent to a single node v are connected. In other words, it gives us a measure of how well nodes in a graph tend to cluster together and thus, how well connected the neighborhood of a node is. Let us look at node v_i which is of degree k_i. Then we get the clustering coefficient C_i for that node by taking the ratio of the number of edges E_i that actually exist between its k_i neighbors and the total number of edges $k_i(k_i-1)/2$ that could exist in the neighborhood of v_i , i.e.

$$C_i = \frac{2E_i}{k_i(k_i-1)}\ ,\quad k_i \geq 2.$$

The entire graph's clustering coefficient is simply the mean of the individual clustering coefficients of the nodes that have degree greater than two.

When calculating the clustering coefficient for different values of the correlation θ, we found that the clustering coefficient is higher for large and positive θ in comparison to small and negative θ in the complementary graph, where the clustering coefficient turned out to be very close to 0. We suggest that this is a sign of globalization, meaning that more and more stocks are dependent on each other and that the market movements are less random. For instance, with $\theta = 0.7$, the Swedish market has almost the same edge density in the original graph as in the complementary graph with $\theta = -0.08$. However, the corresponding values of the clustering coefficients are $C = 0.74$ and $C = 0.0045$. This result is analogous in the American market graph where $C = 0.76$ for $\theta = 0.6$ and $C = 0.02$ for $\theta = -0.1$. Consequentially, we expect to find significantly larger MC than MIS in both market graphs. This result corresponds to the findings in (2). The fact that the clustering coefficient is much higher than the edge density is a typical characteristic for power-law graphs.

3.3 Correlation distribution

As one of the characteristics of the market graph, the correlation distribution provides information about how the stocks are correlated to one another, thus telling us what type of market structure we are dealing with. Figure 3 shows the correlation distribution of the Swedish stock market where the red curve is a normal distribution fitted to the data. Obviously, the correlations between stocks at OMXS are not normally distributed. The data lacks symmetry and the heavy tail on the right will not be encompassed by a normal curve. Moreover, the mean value μ is 0.158 and the standard deviation σ is 0.13. With that in mind and the fact that the correlation of most stocks are greater than zero, i.e. the stock prices tend to move in the same direction, we get yet another indication that the modern stock market is affected by globalization.

Figure 4 is the corresponding plot for the American stock market with $\mu = 0.183$ and $\sigma = 0.157$. In contrast to OMXS; both tails of the correlation distribution of the American stocks are almost entirely covered by the fitted normal distribution. However, the shape of the correlation distribution and the shape of the normal distribution do not match. Thus, a normal fit is not appropriate. Similar to the Swedish market, the American stocks also mainly exhibit positive correlations which further corroborate the theory of increased globalization.

By studying the evolution of the correlation distribution of the American market graph over time, one can show that it remains stable. Consequently, the degree distribution will remain stable over time and a plot can be approximated by a straight line (in a logarithmic scale), which means that it can represent the power-law distribution. For a more stringent analysis see (4). We will later on in the report study the evolution of the Swedish market graph over different time periods.

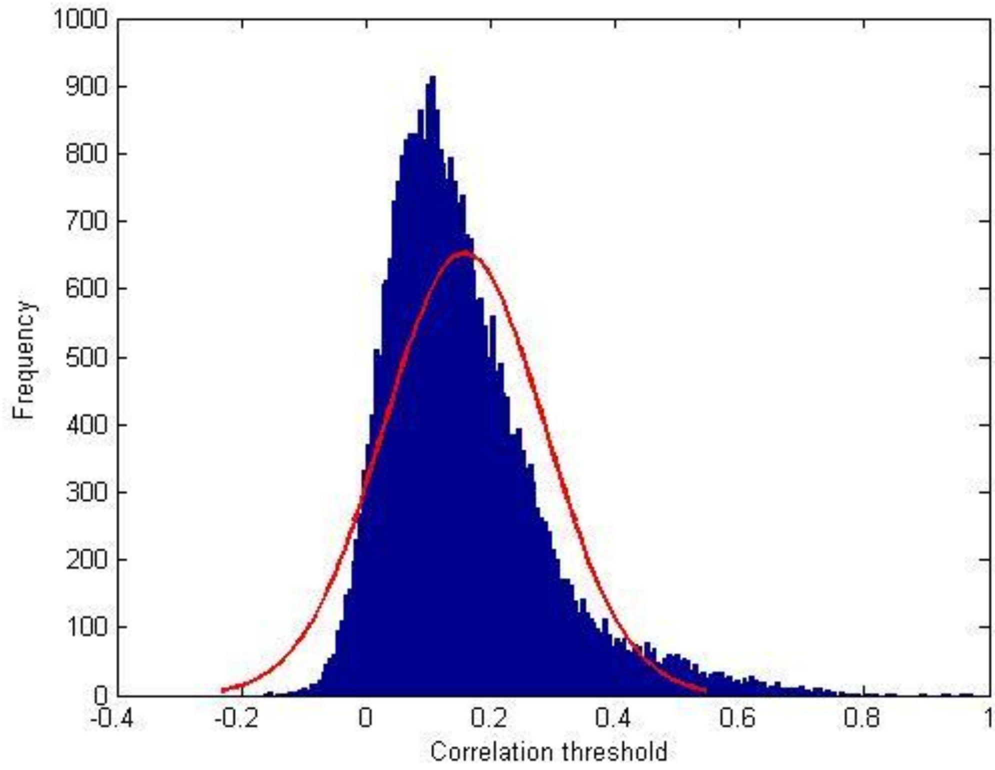

Figure 3. Distribution of correlation coefficients in the Swedish stock market with a fitted normal distribution.

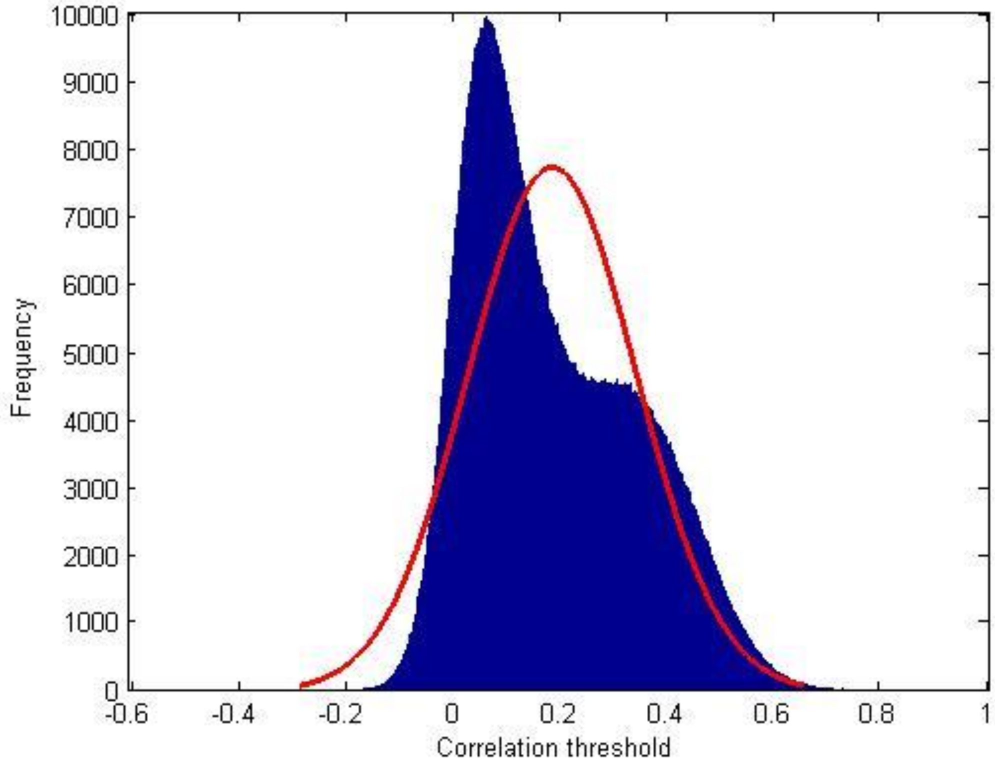

Figure 4. Distribution of correlation coefficients in the American stock market with a fitted normal distribution.

3.4 Edge density

Edge density is a ratio obtained by dividing the numbers of edges in a graph with the maximum possible number of edges, $n(n-1)/2$, where n is the number of vertices in the graph. Changing the value of the correlation threshold will affect the edge density, and by doing so one can construct market graphs with different degrees of correlation between stocks. This can be used to alter sizes of cliques and independent sets in a graph. Figure 5 shows the edge density for the Swedish stock market for different values of the correlation threshold. It is clear that the edge density decreases with increasing threshold values. This result is not surprising since we expect to find fewer stocks that behave similarly as we increase the correlation threshold. Also, higher edge density is linked to lower correlation between stocks, which is in line with the notion that a portfolio with a larger amount of stocks is better diversified. Figure 6 shows the edge density of the American stock market and it is easy to see that it almost has the exact same shape as the Swedish. One can expect similar shapes for any stock market in the world.

In (4) the authors studied the edge density and its change during different consecutive time periods. By setting the value of the correlation threshold to 0.5, they made sure that they got edges that corresponded to those stocks which were significantly correlated with each other. It turned out that the edge density was approximately 8.5 times higher in the last period than the first which, according to the authors, was an indication of the increasing globalization of the modern stock market.

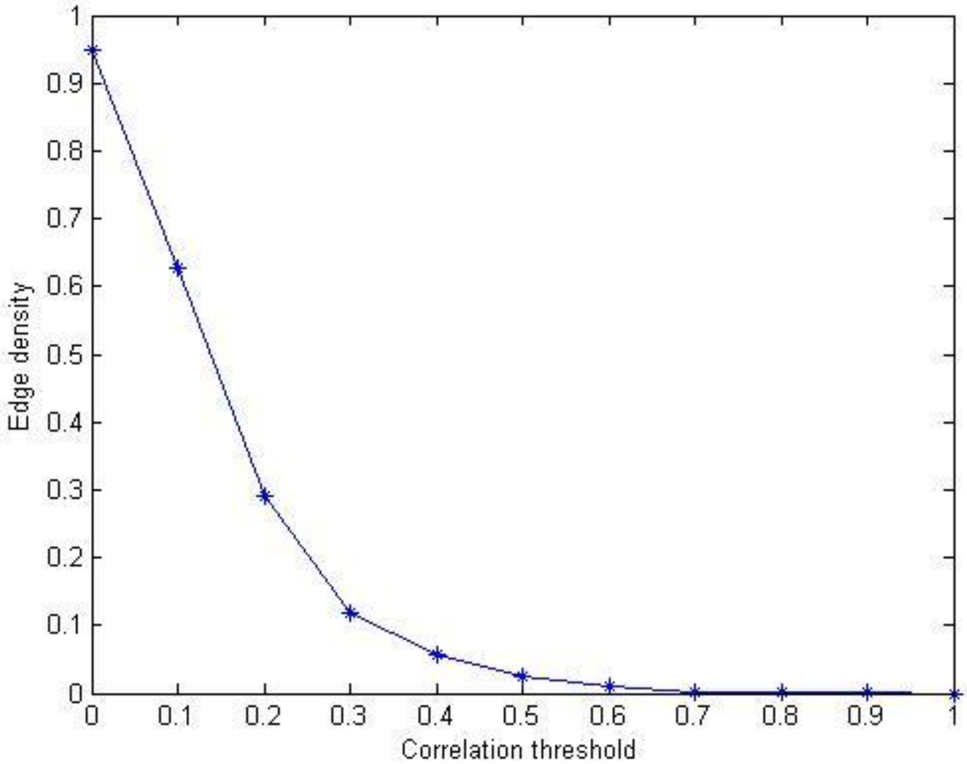

Figure 5. Edge density of the Swedish market graph for different values of the correlation threshold.

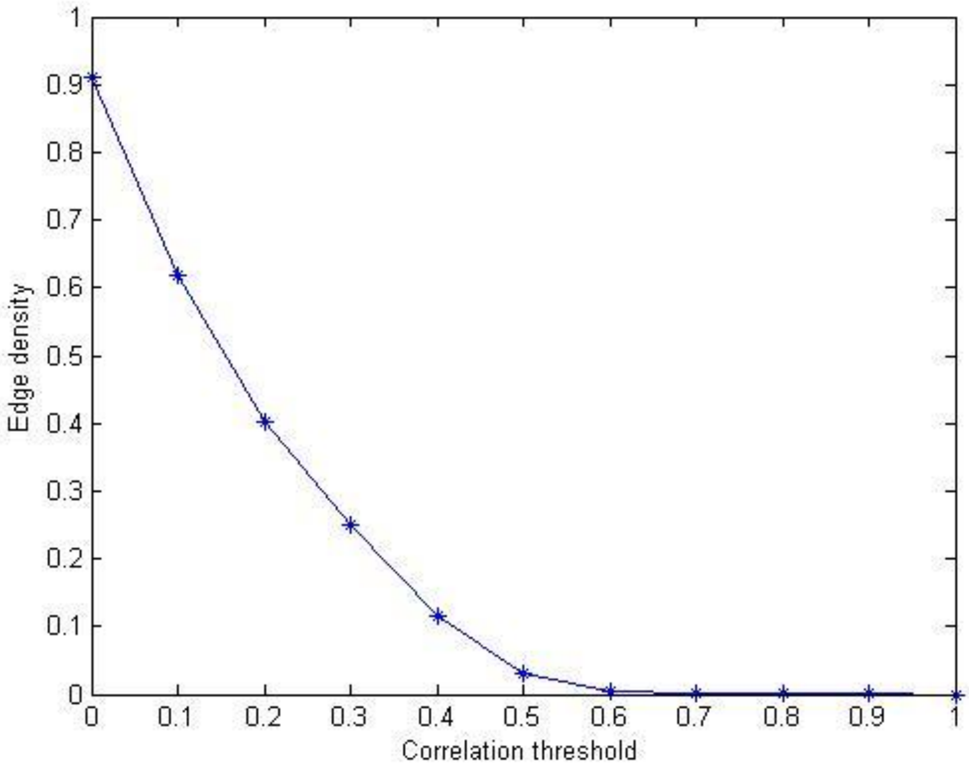

Figure 6. Edge density of the American market graph for different values of the correlation threshold.

4 Structure of the market graph

In this chapter we utilize the earlier described algorithms in order to find differences and similarities between stocks in the graphs.

4.1 Maximum clique and Maximum independent set

As we already have mentioned, the MC is the largest cluster in which all nodes are connected to every other node, thus making it a complete graph. The MIS is the corresponding complete graph in the complementary market graph. Keeping in mind that we are looking at return correlations between stocks, the MC will represent the maximum number of stocks whose price fluctuations exhibit similar behavior. Correspondingly, the MIS represents the maximum set of stocks whose price returns are the most uncorrelated, and thus constitutes the largest diversified portfolio.

In the previous chapter we mentioned that it is easier to find a MC in the original graph than a MIS in the complementary graph. By looking in Appendix I and Appendix II, one quickly realizes that this is also the case. The MIS for $\theta \in [-0.05\,, 0.05]$ are smaller than the MC for $\theta \in [0.2\,, 0.7]$. Moreover, the MIS size becomes even smaller for $\theta < -0.05$ which is consistent with the results we got from Figure 3 and Figure 4, namely that globalization has a strong affect on the market. The fact that the MIS are small, and hence contain too few stocks to choose from when considering building a portfolio, we are led to search for alternate methods which can assist us in finding good, diversified portfolios. This method will be explored in chapter 4.3.

Comparing Appendix I to Appendix II, we clearly see that the Swedish market yields smaller MIS than the American market. One way to decrease unsystematic risk is to hold a portfolio consisting of many uncorrelated stocks. Therefore, it is favorable to invest in the MIS in markets of larger size.

At $\theta = 0.6$ for the Swedish market graph, the MC includes stocks from the industrial and the material sectors. With higher values of θ, one can expect to get a clique consisting of stocks from only one sector. However, the result is still satisfying since the industrial and the material sector are highly correlated. This is quite obvious since the manufacturing industry depends on the companies supplying their materials in order to function and conduct business properly. Another interesting observation is that three financial companies appear in the same MC. This is due to the large positions that the financial companies have in the other companies from the same clique. Consequently at $\theta = 0.7$, the MC will neither include the financial company INDU nor will it include the industrial companies SAND and VOLV, the two of which INDU has large positions in. The same behavior is found in the American market. According to Appendix B, the MC at $\theta = 0.85$ only consists of companies from the basic materials industry, more specifically silver and gold companies. Unlike the Swedish market, these companies do not emerge in other cliques and must therefore be very strongly correlated which is something one can anticipate from both silver and gold securities.

An interesting observation is that as θ decreases, the algorithm either adds stocks from sectors already existing in the MC or stocks belonging to companies that in some other way are highly dependent on the ones already in the MC. An interesting difference between the two markets is that the cliques in the Swedish market are based on some of the biggest companies while the cliques in the American market are built strictly around specific sectors.

4.2 Maximum weighted clique and Maximum weighted independent set

By adding weights to the stocks we get solutions to the MWC- and MWIS-problem which not only considers the price fluctuations between the stocks, but also the liquidity. This will provide us with information about the stock market from another perspective that we can compare to the unweighted case.

The cliques in the weighted case behave very similarly to the unweighted ones except for a few notable differences. Instead of pinpointing gold and silver companies, the algorithm for the MWC-problem generates cliques consisting of market indices for $\theta \in [0.75, 0.85]$ in the American market. This means that not only are the indices highly correlated with each other, but they are also highly liquid. The Swedish market on the other hand is unaltered since we only included stocks from OMX. All stocks included in the Swedish MC and MWC are, unsurprisingly, from OMX Large Cap. Thus, they are the most correlated and most liquid stocks at the same time. We furthermore found that the industrial and the material sectors appear in the weighted as well as in the unweighted case and we therefore draw the conclusion that these stocks are the most correlated as well as the most liquid. Moreover, it is preferable to choose diversified portfolios from the weighted graphs since their liquidity risk is lower, even though their sizes are a bit smaller than the diversified portfolios in the unweighted case (see Appendix III and Appendix IV).

4.3 Maximum quasi-clique and Maximum quasi-independent set

Since the MIS for $\theta = 0$ are small in both markets, we look for MQIS in order to find larger independent sets. By calculating MQIS, we reduce the requirements in the sense that we no longer demand complete graphs. Stocks can be a part of MQIS even though they are not connected to all other stocks in the same MQIS. Thus, we can expect larger MQIS for the price of less diversification. For instance, in the American graph at $\theta = 0$ and $\gamma = 0.6$, a MQIS consisting of 21 stocks, i.e. about 60 % larger than the corresponding MIS, is found (see Appendix V). Also, a MQIS in the Swedish market graph at $\theta = 0.05$ and $\gamma = 0.7$ will generate a quasi clique consisting of 33 stocks, which is a significantly larger diversified portfolio than in the earlier cases (see Appendix VI). However, each stock within the MQIS only needs degree $0.7 * (33 - 1) = 23$ and not 32 in order to be accepted as a part of the QIS. Investing in such a portfolio would be riskier since the information about how the stocks are correlated i.e. exactly how diversified the portfolio really is, is somewhat incomplete. The MQC does however provide us with useful information which, if utilized appropriately, increases our chances of building a large, well diversified portfolio.

5 Dynamics of the market graph

Here we study how the interaction between stocks in the market change as time goes by.

5.1 Data

Earlier we studied and analyzed static market graphs but will now shift our focus to how some of the previously studied features change over time. The hope is to learn more about how those features evolve and what, if anything, that says about the market as time goes by. The data used for this part of the study is the same as the Swedish data used earlier but with a longer time span, namely between April 20, 2008 and October 15, 2010. We then split our data into four equally large periods which resulted in four data sets, each consisting of 155 observations of daily returns. Out of those sets of data we constructed four market graphs and computed their correlation distribution and edge density. This can be seen in Figure 7 and in Figure 8.

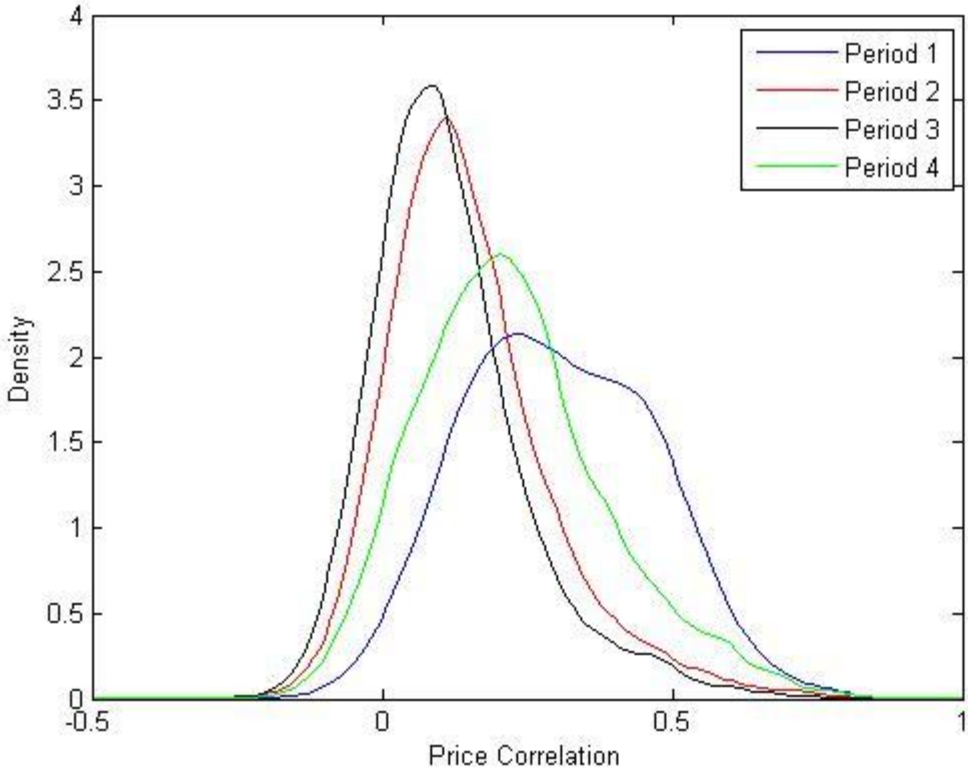

Figure 7. Price correlation density for the Swedish stock market for different time periods.

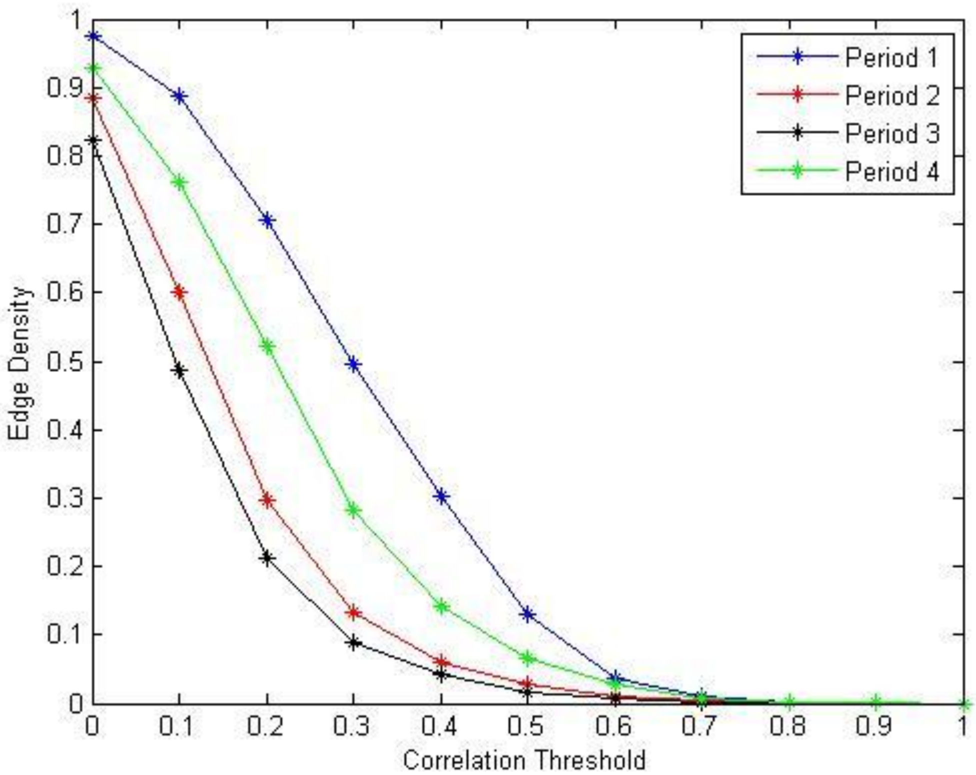

Figure 8. Edge density for the Swedish stock market for different time periods.

5.2 Correlation distribution and edge density

From Figure 7 and Figure 8 we instantly see that the four periods are quite different. Even though periods 1 and 3 appear to be fairly similar it is obvious that they significantly differ from the other two periods. Most importantly, the right tails in periods 1 and 4 (see Figure 7) are greater than for the other two periods and that has an impact on the edge density as well (see Figure 8). This indicates that one would find larger cliques and smaller independent sets if one was looking in period 1 instead of in any other period. Furthermore, we learn that a single market graph constructed using combined data from all the four periods, as was done earlier, is considerably different from the four market graphs constructed with the periodically divided data. Since the market and its structure constantly changes it is important to visualize and keep in mind what impact that can have on the final results. Lastly, this provides more evidence strengthening the hypothesis that negative returns tend to correlate more than positive returns. This is apparent since the first period in our data consists of the last six months of the downturn in the recent financial crisis and that period is also by far the most positively correlated and has the highest edge density of all four measured periods.

5.3 Evolution of the market graph over time

Now that we have seen how much the market graph can vary depending on the choice of time period, we think it would be interesting to make a more in-depth study of how its characteristics evolve during our two and a half years of data. In order to do so we create market graphs for all 100-day and 20-day periods in our data, i.e. one graph for day 1-100, one for 2-101, one for 3-102 etc. and we analyze how their properties change over time.

We begin by considering the evolution of the market graph with a correlation threshold of 0.5 and 100-day intervals. We calculate the mean correlation coefficient, edge density, clique number and clustering coefficient for each period and compared them to the OMXSPI, which is representative of our data. This is done in order to find out if we can acquire new knowledge about the market.

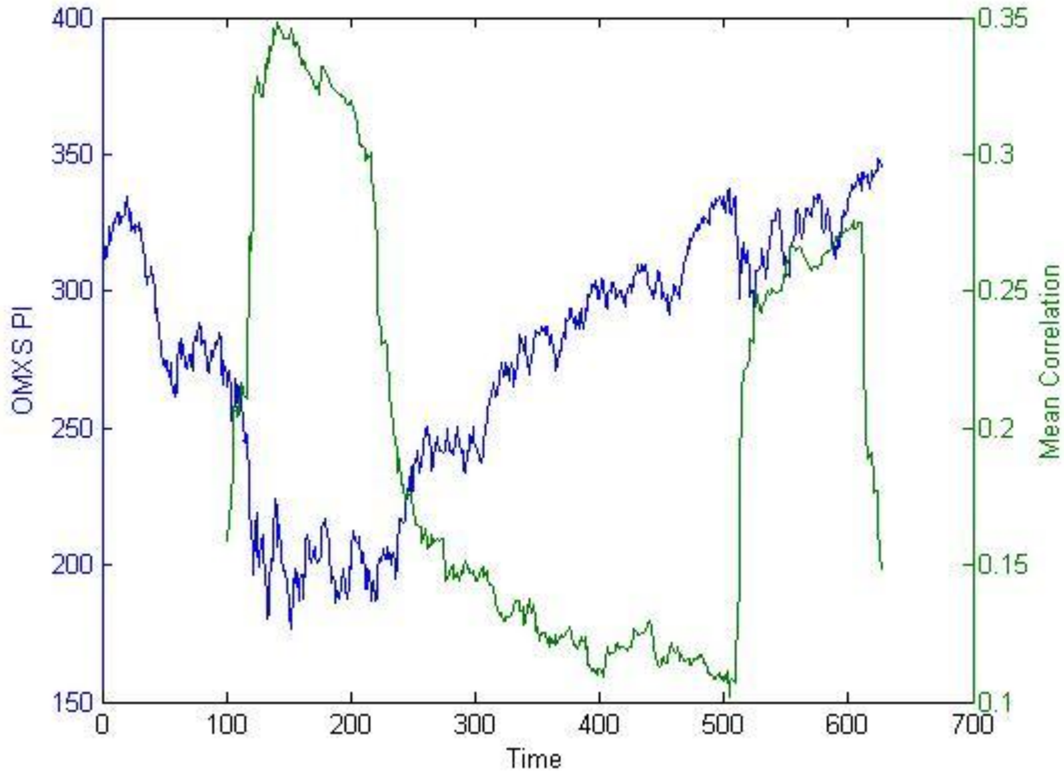

Figure 9. Mean correlation in the Swedish market graph plotted vs. the OMXSPI for continuous 100-day periods.

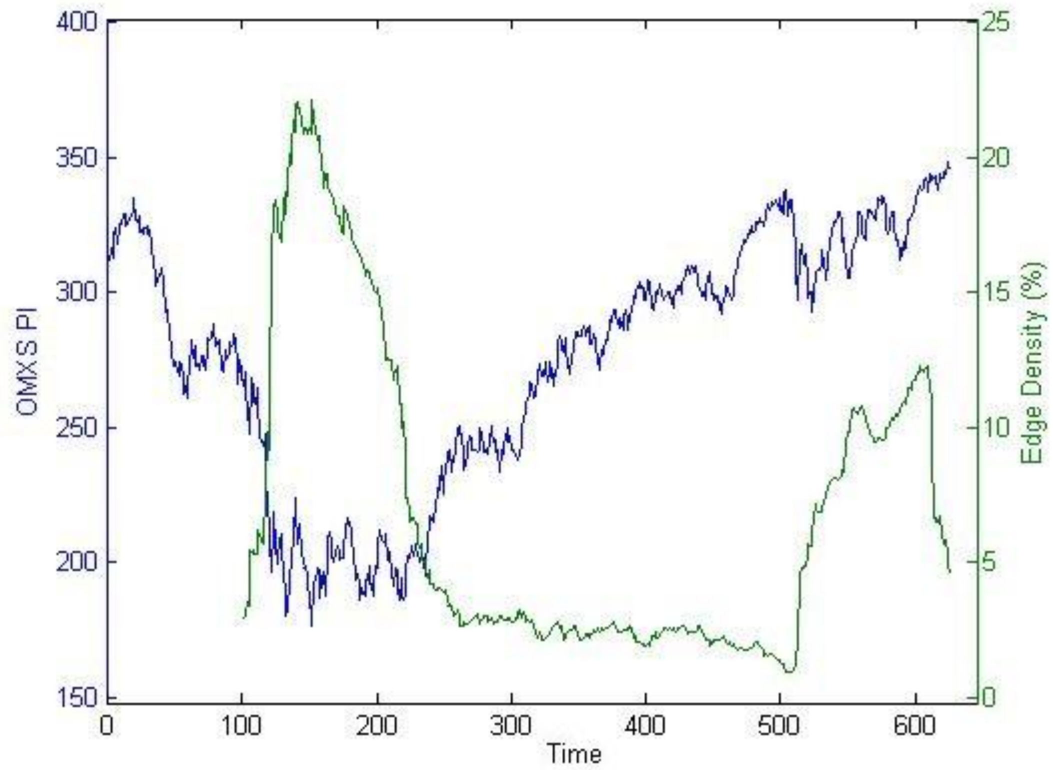

Figure 10. Edge density of the Swedish market graph plotted vs. the OMXSPI for continuous 100-day periods.

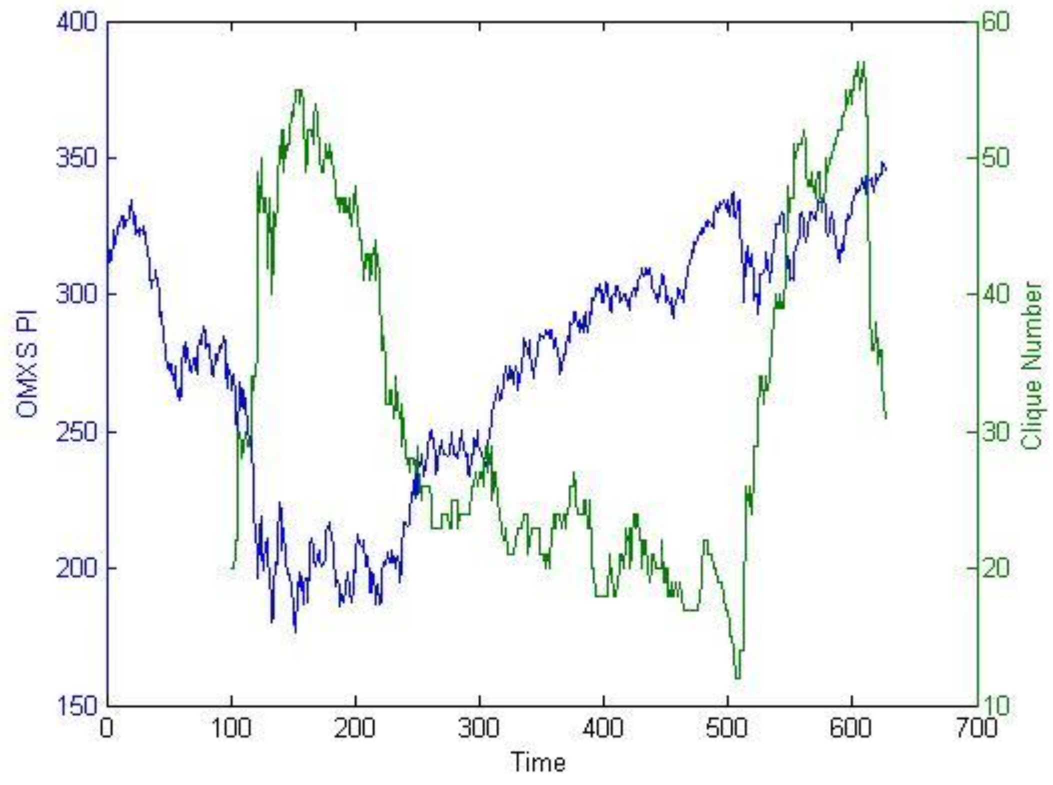

Figure 11. Clique number generated from the Swedish market graph plotted vs. OMXSPI for continuous 100-day periods.

The first observation we make is that the green curves in Figure 9, Figure 10 and Figure 11 look quite similar whilst being negatively correlated with the OMXSPI. We find that this is true since the edge density and the mean correlation have a correlation of about 0.96 with each other and –0.5 with the index. It is also interesting to see that the clique number follows the pattern of the mean correlation and the edge density. This is because the clustering coefficient never drops below 0.65 for the entire period, and as explained earlier, a high clustering coefficient leads to graphs with denser clusters since new edges tend to be added to already dense areas of the graph. However, even though we find that the edge density and mean correlation is strongly negatively correlated with the market, we cannot really use the information from Figure 9, Figure 10 and Figure 11 for anything useful since the 100-day period is far too long in order for us to be able to detect swift changes in market movements. We will therefore divert our attention to the shorter time period we have intended to study, which is the 20-day period with correlation threshold 0.2. The reason we chose to lower the correlation threshold in the 20-day period case is because we wanted a higher edge density in order to get more observations and thus better results. Still, 20-day correlations are not entirely reliable because of the small number of observations but if one wants to be able to catch quick market movements using correlation we believe that the only way is to shorten the time period.

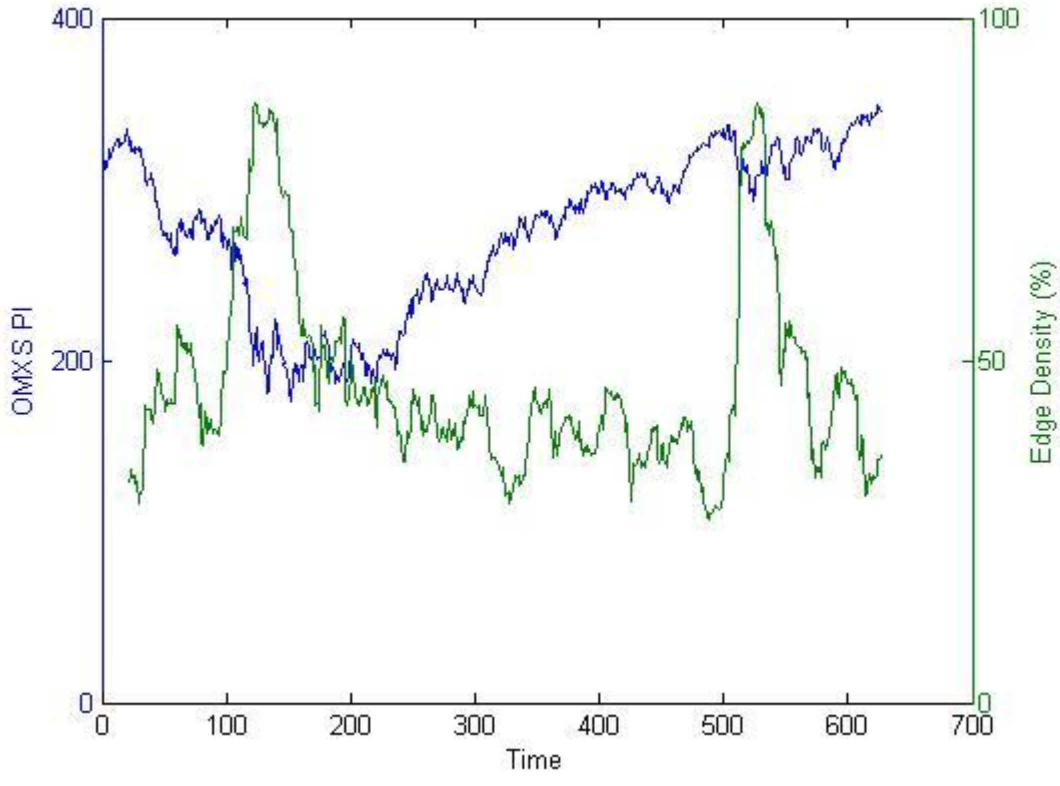

Figure 12. Edge density of the Swedish market graph plotted vs. the OMXSPI for continuous 20-day periods.

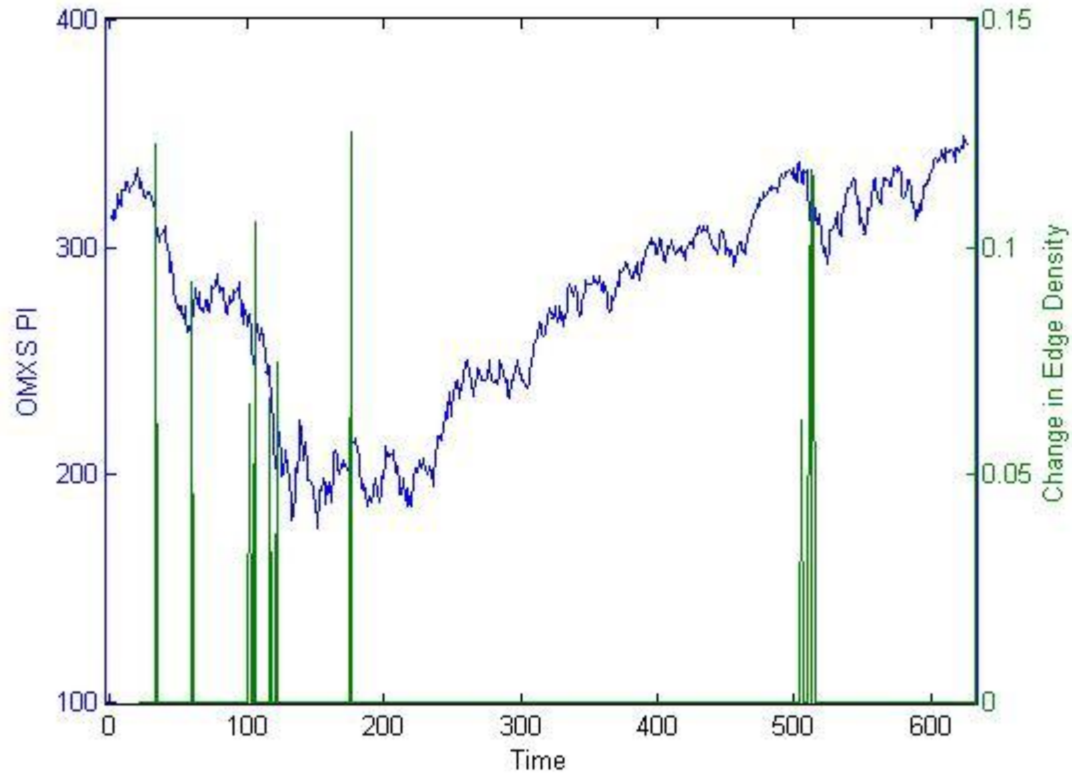

Figure 13. Change in edge density larger than 0.05 of the Swedish market graph plotted vs. the OMXSPI for continuous 20-day periods.

Comparing the 20- day edge density (see Figure 12) with the analogous one for the 100-day period (see Figure 10) we find that the latter is much more volatile. It is however difficult to draw any more conclusions by only studying Figure 10 which is why we calculated the days on which the edge density increased by more than 5 percentage points (see Figure 13). Interestingly, with only two exceptions, it seems as if the edge density only increases by 5 percentage points or more when the market is about to make a sharp downturn. This implies that there is a possibility to use the daily absolute change in edge density as an indicator for when the market is about to go down.

When significant upward or downward jumps occur in the market it is natural to expect that, just as in the case with return correlations, the correlation between different assets' trading volume increase at the same time. To test this hypothesis we construct a market graph using volume correlations instead of price correlations to see if we get similar results. In contrast to the results for the return based market graph, Figure 14 and Figure 15 clearly indicate that the volume correlation and the edge density for the different periods are very similar to one another. Moreover, we can see in Figure 16 that the peaks of the absolute change in edge density do not pinpoint any distinct downturns in the market index in the same way they do for the price based market graphs.

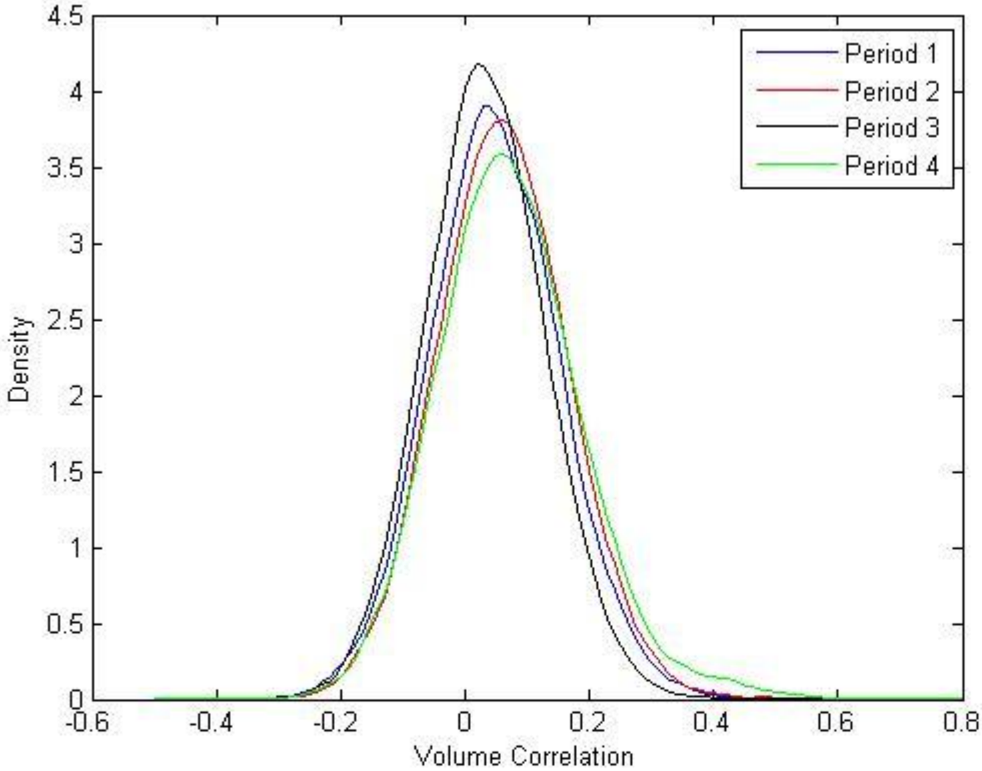

Figure 14. Volume correlation density in the Swedish market graph for different time periods.

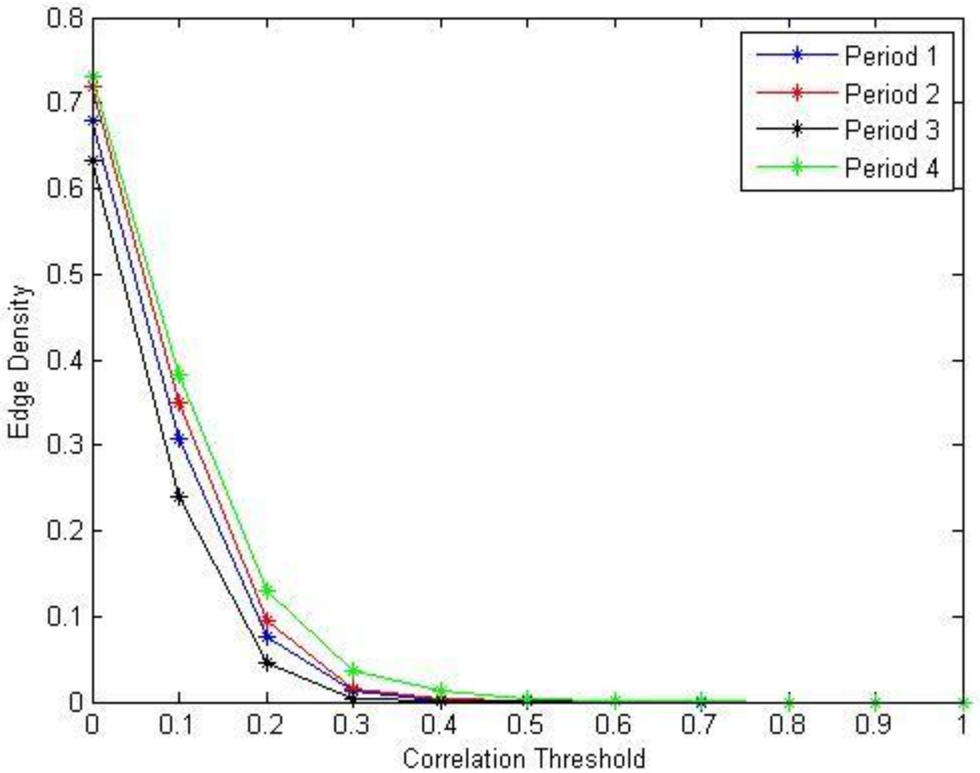

Figure 15. Edge density for the Swedish market graph based on volume correlations for different time periods.

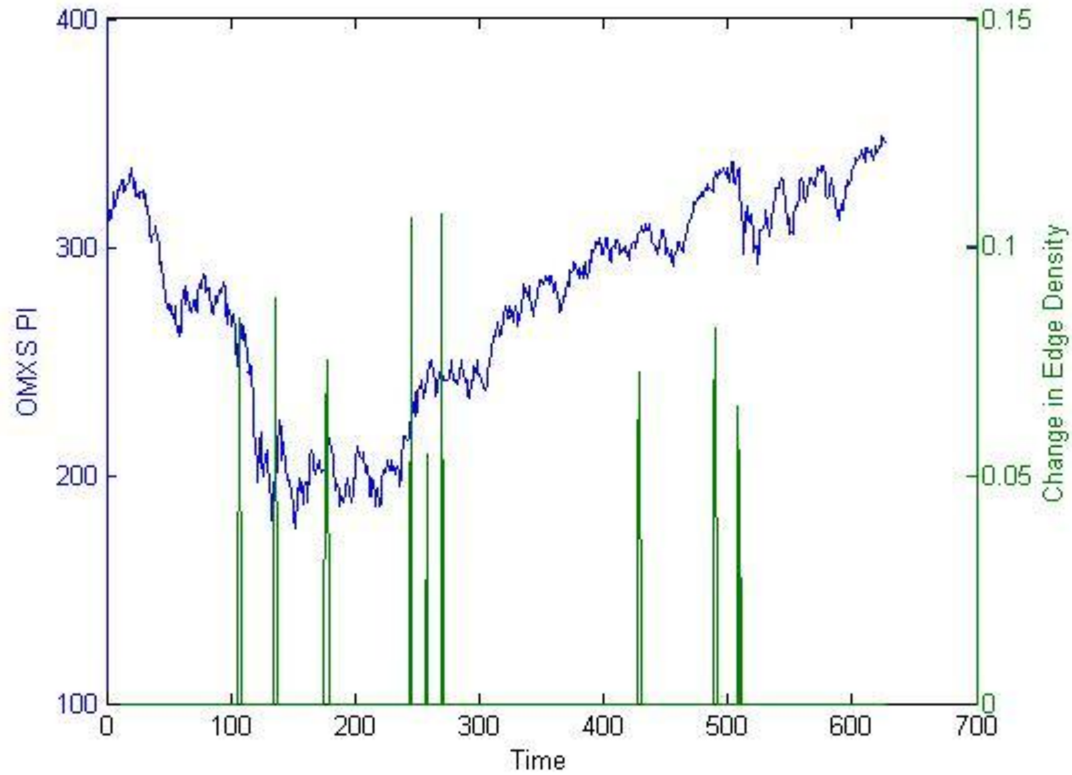

Figure 16. Change in edge density larger than 0.05 of the Swedish market graph based on volume correlations plotted vs. the OMXSPI.

6 Conclusion

The results are summarized and topics for future research are proposed.

In this thesis we have considered three different market graphs; one solely based on stock returns, another one based on stock returns with vertices weighted with a liquidity measure and lastly one based on correlations of volume fluctuations. Research was conducted on two different markets; the Swedish stock market OMX and the American stock market consisting of NASDAQ, AMEX and NYSE.

We found that the clustering coefficient, in both market graphs, was higher for large positive correlations in comparison to small and negative correlations in the complementary graphs. This implies that the MC we found were larger than the MIS which is an effect of globalization. Further, the correlation distributions turned out to lack symmetry and have heavy tails to the right, i.e. the correlations of most stocks are greater than zero. This is yet another sign of the increased globalization making it harder to find diversified portfolios with time.

Solving the MC-problem, we managed to pinpoint specific sectors for higher values of the correlations threshold. For the Swedish market we ended up with the industrial and the material sector, two industries that are highly dependent on each other. The basic material industry, more specifically silver and gold companies, was pinpointed for the American market graph. When we decreased the correlation threshold we found that the algorithm mainly added stocks from the same sector. One of the differences between the two markets was that the cliques in the Swedish market were based on some of the biggest companies while the cliques in the American market were built strictly around specific sectors. Also, in both market graphs, the MIS we found were significantly smaller than the MC.

The cliques for the weighted case behaved very similarly to the unweighted except for a few notable differences. Instead of pinpointing gold and silver companies, the algorithm for the MWC-problem generated cliques consisting of market indices in the American market, telling us that the indices are highly correlated at the same time as they are very liquid. The Swedish market on the other hand turned out to be unaltered since we only included stocks and no indices from OMX. If one ought to invest in a diversified portfolio, it is clearly preferable to do so from the weighted graphs due to their lower liquidity risk. In order to increase the size of the diversified portfolio we also calculated MQIS which gave us independent sets consisting of a larger number of stocks. The price we had to pay was that such a portfolio would be riskier since the information about how the stocks are correlated, i.e. how well diversified the portfolio really is, is somewhat incomplete.

When we split the data into four equally long time periods we found that the price correlations and edge density were quite different. We discovered that one would find larger cliques and smaller independent sets by looking in period 1 instead of in any other period. Furthermore, we learnt that a single market graph constructed using combined data from all four periods is considerably different from the four market graphs constructed with the periodically divided data. This provided more evidence strengthening the hypothesis that negative returns tend to correlate more than positive returns.

The 20-day edge density presented a quite interesting behavior. It seems as if it increases by 5 percentage points or more every time the market is about to make a sharp downturn. This implies that there is a possibility to use the daily absolute change in edge density as an indicator for when the market is about to go down. Obviously this has to be studied further but it is nonetheless an interesting result which maybe even can become a tool in predicting market declines.

Using volume correlations instead of price correlations did not add any new results. Unlike return correlations, the volume correlation distribution and its edge density are very similar to each other for the different time periods. Moreover, we found that the peaks of the change in edge density do not pinpoint any distinct declines in the market index. We had hoped to get additional results by considering volume correlations, but that did not turn out to be the case. Regardless of that, we think that it would be interesting to study volume based market graphs further and perhaps try to find interesting properties which do not exist in the price based graph.

Bibliography

1. **Boginski, Vladimir, Butenko, Sergiy and Pardalos, Panos M.** On structural properties of the market graph. [book auth.] Anna Nagurney. *Innovations in financial and economic networks.* Northampton : Edward Elgar Publishing Inc., 2003, pp. 29-45.

2. *Statistical analysis of financial networks.* **Boginski, Vladimir, Butenko, Sergiy and Pardalos, Panos M.** 2004, Computational statistics & data analysis, Vol. 48, pp. 431-443.

3. **Bondy, Adrian and Murty, U.S.R.** *Graph Theory.* Emeryville : Springer, 2008.

4. *Mining market data: A network approach.* **Boginski, Vladimir, Sergiy, Butenko and Pardalos, Panos M.** 2006, Computers & Operations Research, pp. 3171-3184.

5. *A brief history of generative models for power law and lognormal distributions.* **Mitzenmacher, Michael.** 2004, Internet Mathematics Vol. 1, No. 2, pp. 226-251.

6. **Han, Jiawei and Kamber, Micheline.** *Data mining: concepts and techniques.* San Francisco : Morgan Kaufmann Publishers, 2001.

7. *Coherent Closed Quasi-Clique Discovery from Large Dense Graph Databases.* **Zeng, Zhiping, et al.** Philadelphia, PA : ACM, 2006. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 797-802.

8. *Massive Quasi-Clique Detection.* **Abello, James, Resende, Mauricio G.C. and Sudarsky, Sandra.** [ed.] Sergio Rajsbaum. New York : Springer-Verlag Berlin Heidelberg, 2002. Latin 2002: theoretical informatics. pp. 598-612.

9. **Mainelli, Michael.** Liquidity: Finance in motion or evaporation? *Gresham College.* [Online] Sep 05, 2007. [Cited: Nov 03, 2010.] http://www.gresham.ac.uk/event.asp?PageId=45&EventId=640.

10. *Liquidity and stock returns: Evidence from a pure order-driven market using a new liquidity proxy.* **Marshall, R. Ben.** 2006, International review of Financial Analysis 15, pp. 21-38.

11. *Liquidity and Stock Returns.* **Amihud, Yakov and Mendelson, Haim.** 1986, Financial Analysts Journal Vol. 42 No. 3, pp. 43-48.

12. *Liquidity and stock returns: An alternative test.* **Datar, T. Vinay, Naik, Y. Narayan and Robert, Radcliffe.** 1998, Journal of Financial Markets 1, pp. 203-219.

13. **Bomze, Budinich I.M., Pardalos, Panos M. and Pelilio, M.** The maximum clique problem. [book auth.] Dingzhu Du and Panos M. Pardalos. *Handbook of combinatorial optimixation: Supplement, Volume 1.* Dordrecht : Kluwer Academic Publishers, 1999, pp. 1-74.

14. *A Simple Algorithm to Optimize Maximum Independent Set.* **Balaji, S, Swaminathan, V and Kannan, K.** 1, 2010, Advanced Modeling and Optimization, Vol. 12, pp. 107-118.

15. *An exact algorithm for the maximum clique problem.* **Carraghan, Randy and Pardalos, Panos M.** 6, 1990, Operations Research Letters, Vol. 9, pp. 375-382.

16. **Kumlander, Deniss.** [Online] 2007. [Cited: Dec 15, 2010.] http://www.kumlander.eu/graph/.

17. *On maximum clique problems in very large graphs.* **Abello, J, Pardalos, Panos M and Resende, M G.C.** Providence : American Mathematical Society, 1999. External Memory Algorithms: Dimacs Workshop External Memory and Visualization. pp. 119-130.

18. **Luenberger, David G.** *Investment science.* New York : Oxford university press, 19989.

19. *A random graph model for power law graphs.* **Aiello, William, Chung, Fan and Lu, Linyuan.** 2001, Experiment. Math. Volume 10, Issue 1, pp. 53-66.

Appendix

I – Cliques and independent sets in the Swedish market

Correlation threshold θ	Number of stocks	Stocks
– 0.05	3	ACOM ICTA-B SAS
0	5	ACAN-B DGC NOVE RROS SAS
0.05	14	ARTI-B BALD-B CEVI FEEL GVKO-B KARO MSON-B NOTE NSP-B ORTI-B PSI-SEK RROS SAEK WAFV-B
0.2	76	AAK ABB ALFA ALIV-SDB AOIL-SDB ASSA-B ATCO-A ATCO-B AZA BBTO-B BEF-SDB BEGR BINV BOL CAST ECEX ELUX-B ERIC-A ERIC-B FABG GETI-B HEXA-B HOGA-B HOLM-B HUSQ-A HUSQ-B IJ INDU-A INDU-C INVE-A INVE-B JM KINV-B KLED KLOV KNOW LIAB LUMI-SDB LUND-B LUPE MEDA-A MIC-SDB MTG-B NCC-A NCC-B NDA-SEK NISC-B NOBI ORES ORI-SDB PEAB-B RATO-B SAAB-B SAND SCA-A SCA-B SCV-A SCV-B SEB-A SEB-C SECU-B SHB-A SHB-B SKA-B SKF-A SKF-B SSAB-A SSAB-B STE-R SWED-A TEL2-B TLSN WIHL VNIL-SDB VOLV-A VOLV-B
0.3	54	ABB ALFA ALIV-SDB AOIL-SDB ASSA-B ATCO-A ATCO-B BEGR BOL CAST ECEX ELUX-B FABG GETI-B HEXA-B HOLM-B HUSQ-A HUSQ-B IJ INDU-A INDU-C INVE-A INVE-B JM KINV-B KLED LIAB LUND-B LUPE MTG-B NCC-B NDA-SEK ORI-SDB PEAB-B RATO-B SAND SCA-B SCV-A SCV-B SHB-A SHB-B SKA-B SKF-A SKF-B SSAB-A SSAB-B STE-R SWED-A TEL2-B TLSN WIHL VNIL-SDB VOLV-A VOLV-B
0.4	38	ABB ALFA AOIL-SDB ASSA-B ATCO-A ATCO-B BEGR BOL ECEX ELUX-B HEXA-B INDU-A INDU-C INVE-A INVE-B JM KINV-B LUPE MTG-B NCC-B NDA-SEK PEAB-B RATO-B SAND SCA-B SCV-B SHB-A SHB-B SKA-B SKF-B SSAB-A SSAB-B SWED-A TEL2-B TLSN VNIL-SDB VOLV-A VOLV-B
0.5	25	ABB ALFA ASSA-B ATCO-A ATCO-B BOL ELUX-B INDU-A INDU-C INVE-A INVE-B JM KINV-B LUPE MTG-B NCC-B NDA-SEK SAND SCV-B SKA-B SKF-B SSAB-A SSAB-B TEL2-B VOLV-B
0.6	15	ABB ALFA ATCO-A ATCO-B BOL INDU-C INVE.-A INVE-B KINV-B SAND SKF-B SSAB-A SSAB-B VOLV-A VOLV-B
0.7	8	ALFA ATCO-A ATCO-B INVE-A INVE-B SKF-B SSAB-A SSAB-B

II – Cliques and independent sets in the American market

Correlation threshold θ	Number of stocks	Stocks
– 0.05	6	BNC NEFOI HMNA MEDQ SNFCA VSCP
0	12	ALLB AMTC ARCW CO DD-PA GJJ IMS QADI RGCO SSE UNAM WBNK
0.05	35	AERL ANX BDCO BDL CALL CFBK CO EDCI EDS FFDF GAI GJK GJL GLOI GSLA INV JCDA KGJI LSBI NBXH NFEC NFSB NPBCO OGXI PDEX PSBH RDIB ROIAK RPTP SKH SPRO UBOH ULCM WWIN ZANE
0.65	57	ACC AIV AKR AMB ARE AVB BFS BRE BXP CLI CPT DCT DEI DLR EGP ELS EPR EQR ESS EXR FRT FSP HCN HCP HIW HME HR HST IRC JLL KIM KRC LRY MAA NHP NNN O OFC OHI PCH PCL PKY PPS PRFZ PSA REG RYN SKT SNH SPG SSS TCO UDR VNO WRE WRI VTR
0.7	41	ACC AMB ARE AVB BFS BRE BXP CLI CPT DCT DEI DLR ELS EPR EQR FRT HCN HCP HIW HME HR KIM KRC LRY MAA NNN O OFC OHI PCH PSA REG SNH SPG SSS TCO UDR VNO WRE WRI VTR
0.75	31	AVB BRE BXP CLI CPT DCT ELS EQR FRT HCN HCP HIW HME HR KIM LRY MAA NHP NNN O OHI PCH PSA REG RYN SPG TCO UDR VNO WRE WRI
0.8	16	BRE BXP CLI CPT ELS EQR FRT HCP HIW LRY NNN O PSA REG SPG VNO
0.85	5	ABX AEM AUY GG KGC

III – Weighted cliques and independent sets in the Swedish market

Correlation threshold θ	Number of stocks	Stocks
– 0.05	2	ENRO ORTI-A
0	2	HEBA-B LUMI-SDB
0.05	12	ARTI-B BALD-B DORO ENRO HQ LUXO-SDB MSC MULQ ORTI-A RROS RTIM SAS
0.2	76	AAK ABB ALFA ALIV-SDB AOIL-SDB ASSA-B ATCO-A ATCO-B AZA BBTO-B BEF-SDB BEGR BOL CAST ECEX ELUX-B ERIC-A ERIC-B FABG GETI-B HEXA-B HOGA-B HOLM-B HUSQ-A HUSQ-B IJ INDU-A INDU-C INVE-A INVE-B JM KINV-B KLED KLOV KNOW LIAB LUMI-SDB LUND-B LUPE MEDA-A MIC-SDB MTG-B NCC-A NCC-B NDA-SEK NISC-B NOBI ORES ORI-SDB PEAB-B RATO-B SAAB-B SAND SCA-A SCA-B SCV-A SCV-B SEB-A SEB-C SECU-B SHB-A SHB-B SKA-B SKF-A SKF-B SSAB-A SSAB-B STE-R SWED-A TEL2-B TLSN WIHL VNIL-SDB VOLV-A VOLV-B
0.3	52	ABB ALFA ALIV-SDB AOIL-SDB ASSA-B ATCO-A ATCO-B BEGR BOL ECEX ELUX-B FABG GETI-B HEXA-B HOLM-B HUSQ-A HUSQ-B IJ INDU-A INDU-C INVE-A INVE-B JM KINV-B KLED LIAB LUMI-SDB LUPE MTG-B NCC-B NDA-SEK ORI-SDB PEAB-B RATO-B SAND SCA-B SCV-A SCV-B SECU-B SHB-A SHB-B SKA-B SKF-A SKF-B SSAB-A SSAB-B SWED-A TEL2-B TLSN VNIL-SDB VOLV-A VOLV-B
0.4	34	ABB ALFA AOIL-SDB ASSA-B ATCO-A ATCO-B BEGR BOL ECEX ELUX-B HEXA-B INDU-A INDU-C INVE-A INVE-B JM KINV-B LUMI-SDB LUPE MTG-B NCC-B NDA-SEK PEAB-B RATO-B SAND SCA-B SKA-B SKF-B SSAB-A SSAB-B SWED-A TEL2-B VNIL-SDB VOLV-A VOLV-B
0.5	25	ABB ALFA ASSA-B ATCO-A ATCO-B BOL ELUX-B INDU-A INDU-C INVE-A INVE-B JM KINV-B LUPE MTG-B NCC-B NDA-SEK SAND SCV-B SKA-B SKF-B SSAB-A SSAB-B TEL2-B VOLV-B
0.6	15	ABB ALFA ATCO-A ATCO-B BOL INDU-C INVE.-A INVE-B KINV-B SAND SKF-B SSAB-A SSAB-B VOLV-A VOLV-B
0.7	8	ALFA ATCO-A ATCO-B INVE-B SAND SKF-B SSAB-A SSAB-B

IV – Weighted cliques and independent sets in the American market

Correlation threshold θ	Number of stocks	Stocks
– 0.05	4	CLRO REE SCKT TORM
0	9	ALRN CBIN FCAP MTSL OPTC PKT RITT SCKT TORM
0.05	25	AMIE BTC BWOW CLSN CNYD COBK CZFC DJSP EONC GJI ISRL KENT KRY KSW LEO LONG LSBI NMRX RITT SAVB TORM TRNS TZF USATP ZAGG
0.65	43	AA ACI ACWX ADRE AKS APA ATW BTU BUCY CAM CNQ CNX COP DRQ ECA FCX HAL JOYG MEE MRO MUR NBL NBR NE NOV OII OIS OXY PBR PDE PRFZ PTEN QQQQ RDC SCCO SLB SU TLM UNT VALE WFT WLT
0.7	17	ACI AKS ATI BTU BUCY CLF CNX FCX JOYG MEE NUE QQQQ SCCO STLD VALE WLT X
0.75	6	ADRE ONEQ PRFZ QQEW QQQQ QTEC
0.80	4	ONEQ QQEW QQQQ QTEC
0.85	2	QQQQ QTEC

V – Quasi-cliques and independent sets in the Swedish market

Correlation threshold θ	Degree threshold γ	Number of stocks	Stocks
0.05	0.5	59	ACAN-B AERO-B ARTI-B ATEL AZN BALD-B BTS-B CATE CEVI DAG DGC DIOS DORO DUNI DV ELEC ELGR-B ELUX-A FEEL GVKO-B HAV-B HQ ICTA-B ITAB-B KABE-B KARO LJGR-B LUXO-SDB MOBY MSC-B MSON-A MSON-B MTG-A MTRO-SDB-A MTRO-SDB-B MULQ NAXS NCAS NOTE NOVE OEM-B ORTI-A ORTI-B PHON PREC PROB PSI-SEK RROS RTIM-B SAEK SAGA-PREF SAS SOBI TILG TRAC-B WAFV-B VITR VRG-B XANO-B
0.05	0.6	46	AERO-B ARTI-B AZN BALD-B BTS-B CATE CEVI DAG DGC DORO ELEC ELGR-B FEEL GVKO-B HAV-B HEBA-B HMS HQ ICTA-B LAMM-B LUXO-SDB MOBY MODL MSC-B MSON-A MSON-B MTG-A MTRO-SDB-A MULQ NAXS NCAS-B NSP-B ORTI-A ORTI-B PHON PROB PSI-SEK RROS RTIM-B SAEK SAGA-PREF SAS SOBI TILG WAFV-B VITR
0.05	0.7	33	ARTI-B BALD-B CATE CEVI DORO ELGR-B FEEL GVKO-B HMS HQ ICTA-B KARO MSC-B MSON-B MTG-A MTRO-SDB-A MULQ NOTE NSP-B ORTI-A ORTI-B PHON PREC PROB PSI-SEK RROS RTIM-B SAEK SAS SOBI TILG TRAC-B WAFV-B
0.05	0.8	26	ARTI-B BALD-B CEVI DORO FEEL GVKO-B HQ ICTA-B KARO LUXO-SDB MSC-B MSON-B MULQ NOTE NSP-B ORTI-A ORTI-B PHON PSI-SEK RROS RTIM-B SAEK SAS SOBI TILG WAFV-B
0.05	0.9	18	ARTI-B BALD-B DORO ENRO ICTA-B KARO

			LUXO-SDB MSC-B MULQ NOTE ORTI-A ORTI-B PHON PREC RROS RTIM-B SAEK SAS
0	0.5	14	DGC HEBA-B HOLM-A ICTA-B MOBY MSC-B MSON-B NSP-B ORTI-A ORTI-B PSI-SEK RROS SAS TRAC-B
0	0.6	8	ARTI-B CEVI HAV-B MSON-B ORTI-A SAS SOBI WAFV-B
0	0.7	8	ACAN-B DGC MSON-B ORTI-A ORTI-B PSI-SEK RROS SAS

Correlation threshold θ	Degree threshold γ	Number of stocks	Stocks
0.6	0.5	23	ABB ALFA ASSA-B ATCO-A ATCO-B BOL ELUX-B INDU-A INDU-C INVE-A INVE-B KINV-B NCC-B RATO-B SAND SCV-B SHB-A SKA-B SKF-B SSAB-A SSAB-B VOLV-A VOLV-B
0.6	0.6	19	ABB ALFA ATCO-A ATCO-B BOL INDU-A INDU-C INVE-A INVE-B KINV-B NDA-SEK SAND SHB-A SKA-B SKF-B SSAB-A SSAB-B VOLV-A VOLV-B
0.6	0.7	19	ABB ALFA ATCO-A ATCO-B BOL INDU-A INVE-A INVE-B KINV-B NCC-B RATO-B SAND SCV-B SKA-B SKF-B SSAB-A SSAB-B VOLV-A VOLV-B
0.7	0.5	11	ALFA ATCO-B INDU-C INVE-A INVE-B KINV-B SAND SKA-B SKF-B SSAB-A SSAB-B

VI – Quasi-cliques and independent sets in the American market

Correlation threshold θ	Degree threshold γ	Number of stocks	Stocks
0	0.6	21	BDL BTI CART CIZN CYCCP EEI EMCF FFDF HAVNP IVA KENT KGJI MYF PBHC PCBS PFIN RDIB RIVR SGRP TORM TRCI
0	0.7	14	ADTN BDL CLRO CWBC EOSPN GIA IVA KGJI LPTH NRB SGRP TRCI VMEDW WWIN
0.8	0.5	27	BRE BXP CLI CPT DCT ELS EQR FRT HCN HCP HIW HME HR LRY MAA NNN O PCH PSA REG SNH SPG TCO UDR VNO WRE WRI
0.8	0.6	24	BRE BXP CLI CPT ELS EQR FRT HCN HCP HIW HME HR LRY MAA NNN O PCH PSA REG SNH SPG TCO UDR VNO
0.8	0.7	21	BRE BXP CLI CPT EQR FRT HCP HIW HME LRY MAA NNN O PCH PSA REG SPG TCO UDR VNO WRI
0.8	0.8	19	BRE BXP CLI CPT ELS EQR FRT HCP HIW HME LRY MAA NNN O PCH PSA REG SPG VNO