Department of Economics
School of Social Sciences

Weighted Network Analysis of High Frequency Cross-Correlation Measures

Giulia Iori[1]

Department of Economics, City University

Ovidiu V. Precup[2]

Department of Mathematics, London School of Economics

Department of Economics
Discussion Paper Series
No. 06/10

---

[1] Department of Economics, City University, Northampton Square, London, EC1V 0HB, UK. Email: g.iori@city.ac.uk

[2] Department of Mathematics, London School of Economics, Houghton Street, London, WC2A 2AE, U.K. Email: E-mail: O.V.Precup@lse.ac.uk

Weighted Network Analysis of High Frequency Cross-Correlation Measures

Giulia Iori*
Department of Economics, City University
Northampton Square London, EC1V 0HB, U.K.
E-mail: g.iori@city.ac.uk

Ovidiu V. Precup
Department of Mathematics, London School of Economics
Houghton Street, London, WC2A 2AE, U.K.
E-mail: O.V.Precup@lse.ac.uk

October 25, 2006

Abstract

In this paper we implement a Fourier method to estimate high frequency correlation matrices from small data sets. The Fourier estimates are shown to be considerably less noisy than the standard Pearson correlation measure and thus capable of detecting subtle changes in correlation matrices with just a month of data. The evolution of correlation at different time scales is analysed from the full correlation matrix and its Minimum Spanning Tree representation. The analysis is performed by implementing measures from the theory of random weighted networks.

Keywords: High-Frequency Correlation, Fourier method, random weighted networks.

1 Introduction

Robust correlation measures are important for derivatives pricing, risk management, portfolio optimisation and for understanding market microstructure effects.

The conventional method of computing correlation is the Pearson coefficient. This method requires homogeneous time series and in order to apply it to high frequency data, the time series need to be homogenised and synchronised first through an interpolation scheme.

An alternative, non parametric approach has been suggested in [1] where the variance-covariance matrix estimator of a multivariate process is computed via Fourier analysis. Previous applications of the method can be found in [2, 3, 4, 5, 6, 7].

In this paper we compare the performance of the Pearson and Fourier methods by computing returns cross-correlation matrices at different time scales using one month (September 2002) of high frequency trades in the member stocks of the S&P100[1] index.

The selected stocks are grouped into twelve different industry sectors[2]: Technology (16

*Corresponding author
[1]data source: NYSE Trades and Quotes (TAQ) database
[2]according to the classification provided by finance.yahoo.com

stocks), Basic Materials (7 stocks), Financial (13 stocks), Capital Goods (3 stocks), Conglomerates (5 stocks), Energy (4 stocks), Services (16 stocks), Transport (4 stocks), Utilities (7 stocks), Health Care (10 stocks), Non-Cyclical Consumer Goods (11 stocks), Cyclical Consumer Goods (4 stocks).

The estimation of intra-day correlations over short periods of time (eg. a month) is of high practical value for day trading and hedging purposes. In fact, such estimates are more sensitive to short timescale economic changes than correlation measures obtained from averaging over several months. Thus, we choose to investigate a month of tick-by-tick data aiming to compare the quality of the information that can be derived by applying each of the two methods on limited statistics. The Fourier estimates reproduce the structural changes on filtered correlation matrices observed in previous studies [8, 9, 10, 11, 12, 13, 14, 15, 16] with much larger data sets. Moreover, we show that the Fourier estimates are sufficiently accurate to reveal further structural changes in the full, unfiltered, correlation matrices.

2 Fourier Correlation Measure

The Fourier method is model independent, produces very accurate, smooth estimates and handles the time series in their original form without imputation or discarding of data. A rigorous proof of the method is given in the original paper by Malliavin and Mancino[1] and only the main results are summarized below.

The method works as follows. Let $S_i(t)$ be the price of asset i at time t and $p_i(t) = \ln S_i(t)$. The physical time interval of the asset price series is re-scaled to $[0, 2\pi]$. The variance/covariance matrix Σ_{ij} of log returns is derived from its Fourier coefficient $a_0(\Sigma_{ij})$ which is obtained from the Fourier coefficients of dp_i:

$$a_k(dp_i) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dp_i(t), \quad b_k(dp_i) = \frac{1}{\pi}\int_0^{2\pi} \sin(kt)dp_i(t), \quad {\scriptstyle k\geq 1.} \tag{1}$$

In practice, the coefficients are computed through integration by parts. As $p_i(t)$ is not observed continuously but given by unevenly spaced tick-by-tick observations of trades prices, the actual implementation requires the integrals in (1) to be in discrete form:

$$\begin{aligned} a_k(dp_i) &= \frac{1}{\pi}\sum_{n=1}^{N}\Big([p_i(t_n)\cos(kt_n) - p_i(t'_n)\cos(kt'_n)] - p_i(t'_n)[\cos(kt_n) - \cos(kt'_n)]\Big), \\ b_k(dp_i) &= \frac{1}{\pi}\sum_{n=1}^{N}\Big([p_i(t_n)\sin(kt_n) - p_i(t'_n)\sin(kt'_n)] - p_i(t'_n)[\sin(kt_n) - \sin(kt'_n)]\Big). \\ \text{where} \quad & t'_n = t_{n-1} \end{aligned} \tag{2}$$

In (2), N corresponds to the number of trades in the re-scaled interval and we set the price $p_i(t) = p_i(t_{n-1})$ to compute the integrals between two consecutive trading times $[t_{n-1},\ t_n]$.

The Fourier coefficient of the pointwise variance/covariance matrix Σ_{ij} is :

$$a_0(\Sigma_{ij}) = \lim_{\tau\to 0}\frac{\pi\tau}{T}\sum_{k=1}^{T/2\tau}[a_k(dp_i)a_k(dp_j) + b_k(dp_i)b_k(dp_j)]. \tag{3}$$

The highest wave harmonic ($T/2\tau$) that can be analysed is determined by the lower bound of τ (time gap between two consecutive trades) which is 1 second for all S&P100 price series.

The integrated value of Σ_{ij} over the time window is defined as $\hat{\sigma}^2_{ij} = 2\pi a_0(\Sigma_{ij})$ which leads to the Fourier correlation matrix $\rho_{ij} = \hat{\sigma}^2_{ij}/(\hat{\sigma}_{ii} \cdot \hat{\sigma}_{jj})$.

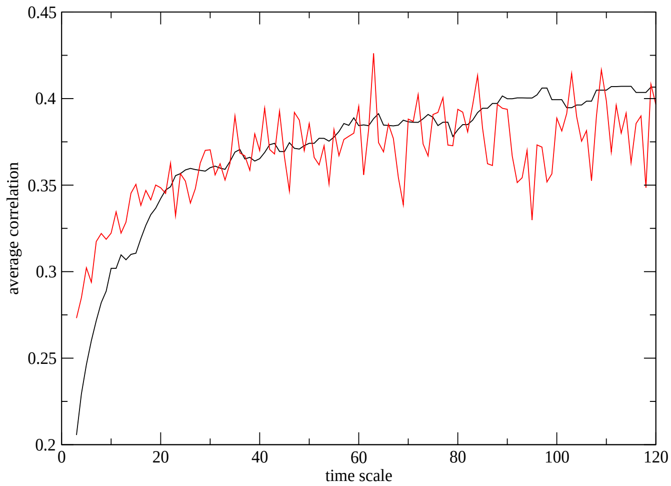

Figure 1: The average correlation across all stocks increases with time scale (Fourier (black) and Pearson (red)). This is indicative of the Epps affect being present.

3 Network Analysis

The correlation matrix can be represented as a network of vertices (stocks) and weighted links (correlations). Following [17, 18] we define the *degree* of a vertex in the network as $k_i = \sum_{j \in \mathcal{V}(i)} 1_{ij}$ where the sum runs over the set $\mathcal{V}(i)$ of neighbours of i and 1_{ij} is an indicator function for whether there is a connection between i and j. The *strength* of a vertex is defined as $s_i = \sum_{j \in \mathcal{V}(i)} c_{ij}$ where c_{ij} is the correlation between vertices (stocks) i and j. We use the degree k_i as a measure of stock centrality for MSTs and the strength s_i as a measure of stock centrality in the overall correlation matrices. For the weighted clustering coefficient we use the definition suggested in [19],

$$C_i^w = \frac{\sum_{j,h} c_{ij} c_{ih} c_{jh}}{\sum_{j,h} c_{ij} c_{ih}}. \tag{4}$$

This definition reduces to the standard clustering coefficient in the binary case and retains the property $0 \leq C_i^W \leq 1$.

For our analysis we only consider the positive elements of the correlation matrices. Even with this choice the correlation matrix is almost fully connected.

When analysing intra-day data, the choice of time scale on which to measure correlations becomes crucial. In Figure 1 we plot the average correlation at different time scales, from three minutes to two hours. The average correlation increases with the time scale, a result known as the Epps effect [20] [3]. Not only does the average correlation increase with time scale, but it is also accompanied by a structural change in the correlation matrix as shown in [8, 9, 10, 11, 12, 13] and more recently, using the Planar Maximally Filtered Graph method

[3]It has been argued [4, 6] that the Epps effect may be determined not purely by economic factors but also by data asynchronicity, particularly when correlations are measured between less liquid stocks.

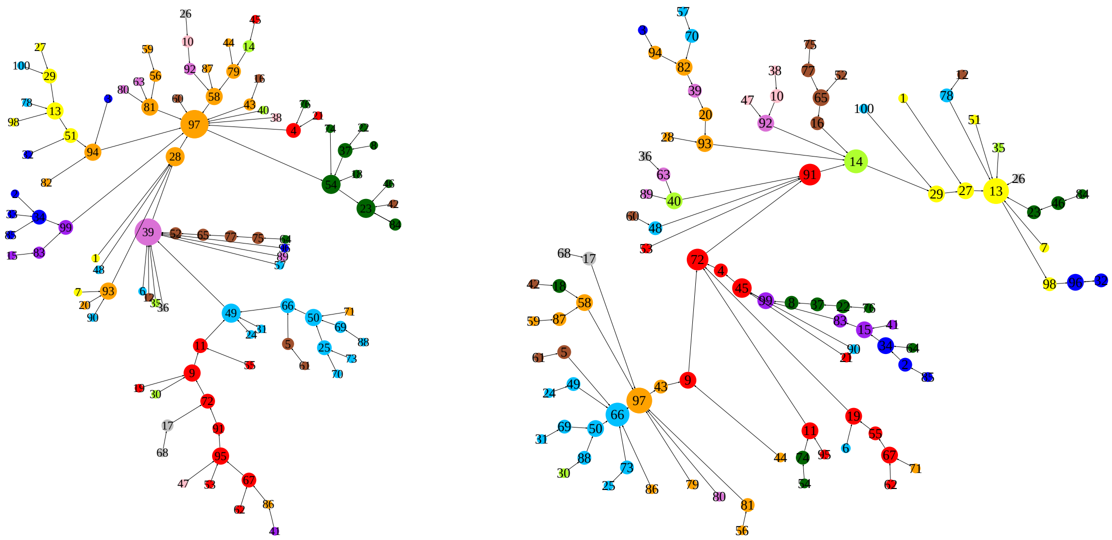

Figure 2: MST obtained with Fourier at 10 minutes (left) and 90 minutes (right). WMT is node 97 and GE is node 39. The size of the dots reprenting the different socks is proportional to the number of links. The color code of the different industrial sectors is given in table 3.

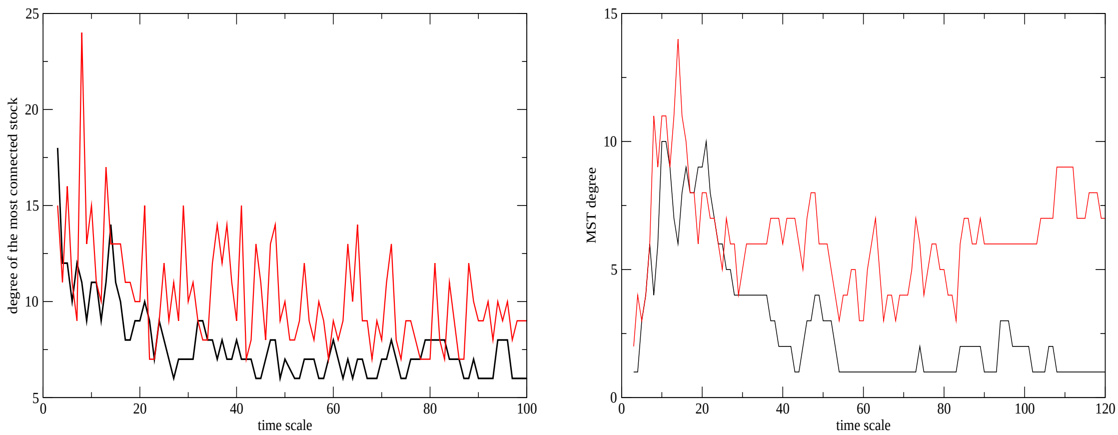

Figure 3: (Left) Degree of the most connected stock in the MST for Fourier (black) and Pearson (red). (Right) Degree of GE (black) and WMT (red) as a function of the time scale (Fourier).

in [14, 15, 16]. The above mentioned studies demonstrate that the shape of the MST changes substantially with the time scale. On very short time scales the MSTs are centralised graphs with a few vertices that collect a large number of connections. On longer time scales the graph structure becomes significantly more dispersed with no obvious hubs. We recover this result as shown in figure 2. On the left we plot the MST, obtained with Fourier, at the scales of 10 minutes and on the right at time scales of 90 minutes.

Method	average degree	Stock
	7.36	Wal-Mart Stores
Fourier	5.94	General Electric
	5.43	Boise Cascade
	4.65	Americal Express
	4.57	Intel Corp
	10.36	Wal-Mart Stores
Pearson	8.40	US Bancorp
	6.89	Bank of America
	5.25	Exxon Mobil Corp
	5.07	Intel Corp.

Table 1: Vertices with the highest average degree in the MST on time scales shorter than 30 minutes.

To quantify the structural change of the MST in figure 3 (left) we plot the evolution, up to a two hours time scales, of the degree of the most connected vertex in the MST derived from both Pearson(red) and Fourier(black) correlation estimates. We notice that while the Pearson estimate gives very noisy results on this small data set (also visible in Figure 1), the Fourier estimator provides much more consistent results across different time scales. In Table 1 we list the five most connected stocks in the MST generated with the two methods. In [15] General Electric (GE) and Wal-Mart (WMT) are reported as the most connected stocks in 2002. When averaging up to time scales of 30 minutes, we find WMT to be the stock with the highest average degree for September 2002. This result is not affected by the method used to estimate correlations. Nonetheless, the Fourier and Pearson methods results differ when it comes to identifying the remaining most connected stocks in the MST. For example while Fourier identifies GE to be the second most connected stock, Pearson ranks GE as the eighth most connected one. Figure 3 (right) shows the evolution of the maximum degree for WMT and GE obtained from the Fourier MST matrix. For both stocks the degree rises quickly and remains high at time scales between 10 and 20 minutes. Nonetheless at time scales longer than 20 minutes, and up to two hours, WMT is the most connected stock in the MST.

To analyze the full correlation matrix we implement the measures of strengths and clustering defined above. While some analysis in this direction has been performed in previous studies, this was based on filtered correlation matrix (either planary filtered graphs [14, 15, 16] or graphs constructed by including only the strongest $N-1$ links, with N the number of stocks [11]).

In Figure 4 (left) we plot the evolution, across time scales, of the normalised strength of the most connected vertex in the full correlation matrix calculated both with Pearson (red) and Fourier (black). The normalised strength, at any time scale τ is defined as $\tilde{s}_i(\tau) = s_i(\tau)/\hat{c}(\tau)$, where $\hat{c}(\tau)$ is the scale τ total correlation. Without this normalisation the strength would trivially increase with time scale as a result of the Epps effect. By normalising we can quantify the way the most correlated stock is central to the network, in terms of proportional contribution to the total correlation. We notice again (Figure 4 (left)) that the Pearson estimator is very noisy while the Fourier estimator is significantly smoother. The Fourier estimator also indicates a rise in the most correlated stocks relative strengths, at time scales shorter than 20 minutes, analogous to the increasing degree of the most connected stock in the MST on short time scales.

The stocks that contribute the most, on average, to the total correlation on time scales shorter than 30 minutes, are again WMT and GE, as shown in Table 2. We also show, in Figure 4 (right), that while GE is central to the correlation matrix only at short time scales,

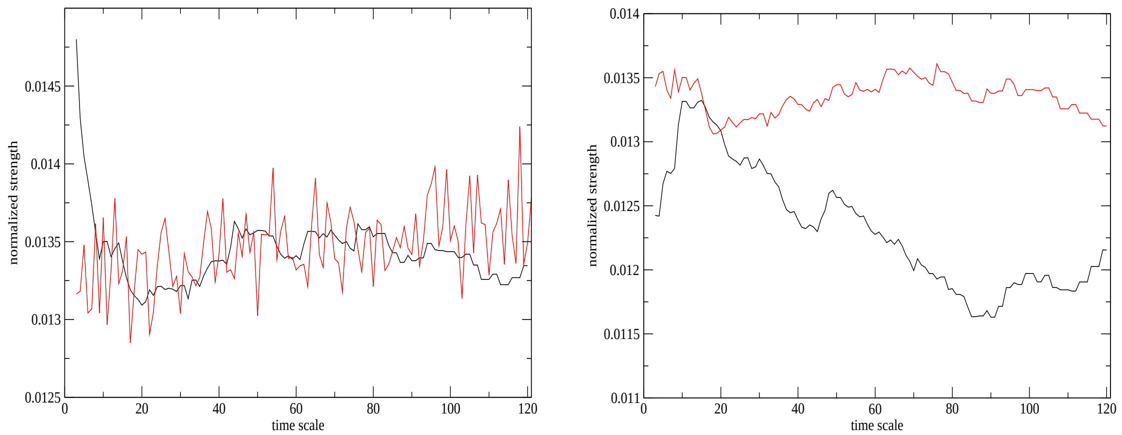

Figure 4: (Left) Normalised strength of the most correlated stock determined with the Fourier (black) and Pearson (red) methods. (Right) Fourier normalised strength of GE (black) and WMT (red) as a function of time scale (Fourier).

WMT is the most central stock at all time scales up to two hours.

Method	Normalised Strength	Stock
	0.0133	Wal-Mart Stores
Fourier	0.0130	General Electric
	0.0129	Bank of America
	0.0128	IBM
	0.0128	Walt Disney Co.
	0.0132	Wal-Mart Stores
	0.0130	US Bancorp
Pearson	0.0128	Bank of America
	0.0127	American International Group
	0.0126	General Electric

Table 2: Vertex with highest average normalised strength at time scales shorter than 30 minutes

In Figure 5 (left) we plot the evolution, across time scales, of the the relative weighted clustering coefficient of the most clustered stock in the full correlation matrix calculated both with Pearson (red) and Fourier (black). The relative weighted clustering coefficient is defined as $\tilde{C}_i^w(\tau) = \frac{C_i^w(\tau)}{\bar{C}^w(\tau)}$, where $\bar{C}^w(\tau)$ is the scale τ average clustering coefficient. The normalisation is also necessary in this case as the clustering coefficient, defined in eq. 1, would trivially increase with the time scale purely as a consequence of the correlation rise. Again the Fourier method provides smooth results which reveal that the relative clustering coefficient of the most clustered stock increases as the time scale falls below 20 minutes. Here, instead of identifying the stock with the highest clustering coefficient, we shift the analysis to the industrial sectors. In Table 3 we report the average of the intra-sector relative strength and intra-sector relative clustering on time scales shorter than 30 minutes. A relative clustering (strength) larger than one implies that that intra-sector clustering (strength) is larger than the average clustering (strength) in the network. We first note that for some sectors there is a significant difference between intra-sector strength and clustering revealing that not all the stocks in that sector are highly correlated with each other. This effect is particularly evident for the Cyclical Consumer Goods and the Capital Goods sectors.

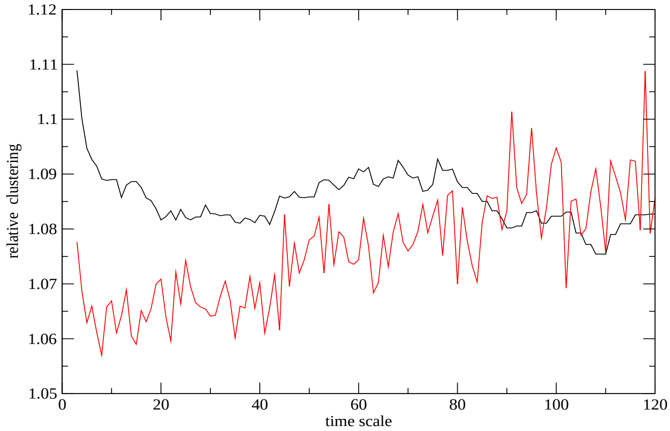

Figure 5: Relative clustering coefficient of the highest cluster coefficient stocks - Fourier (black) and Pearson (red).

The most clustered sector at all time scales, up to two hours, is the Financial one. At short time scales this is followed by Services, Technology, Energy, and Non-Cyclical Consumer Goods. The study in [15] uses the same sector classification but a different selection of stocks (100 highly capitalised stocks instead of the member stocks of the S&P100) and finds that the Financial and the Energy sectors have the highest intra-sectors cluestering, on planary filtered graphs, on a daily time scale. This is in agreement with our results, even though we only include 13 stocks in the Financial sector while in [15] 24 stocks are selected.

Sector	size	color	intra-sector strength	intra-sector clustering
Technology	16	turqoise	1.13	1.02
Basic Material	7	yellow	1.04	0.97
Financial	13	red	1.33	1.28
Capital Good	3	pink	0.86	0.41
Conglomerate	5	magenta	1.01	0.87
Energy	4	purple	0.97	1.02
Services	16	orange	1.28	1.12
Transport	4	grey	0.99	0.74
Utilities	7	blue	0.84	0.80
Health Care	10	brown	1.07	0.94
Non-Cyclical Consumer Good	11	dark green	1.08	1.02
Cyclical Consumer Good	4	green yellow	1.10	0.66

Table 3: Intra-sector relative strength and clustering coefficients at time scales shorter than 30 minutes.

Nonetheless our analysis leads to clearly different results for the Service sector (which includes Wal-Mart) which [15] report as being poorly intra-connected. A possible reason for this disagreement could be that the composition of this sector is very different in the two studies (in [15] this sector is composed of 20 stocks while in our study only 7 stocks are

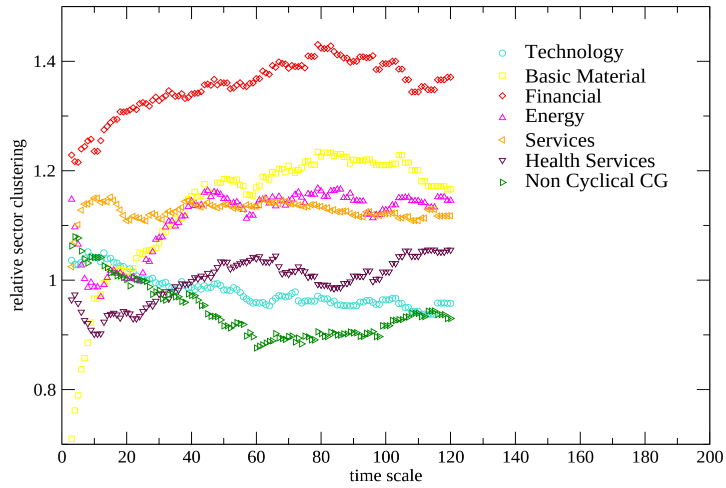

Figure 6: Relative clustering coefficient of the five most clustered industrial sectors.

present). Another possible reason may be the difference in time scale at which the correlations are measured. In Figure 6 we plot the relative clustering coefficient for the 7 most clustered sectors as a function of the time scale. We can see that the ranking of sectors in terms of their relative clustering coefficients changes considerably over time, and in particular the Services sector, which is the second most clustered at short time scale, becomes only the fourth most clustered on two hours time scales. It may well be that that the relative clustering of this sector decreases even further on daily time scales.

The high clustering coefficient of some sectors is reflected in the MST. For example, the MST at 10 minutes in figure 2 (left), identifies very clearly the clusters associated to the Financial (red), Services (orange), Non-cyclical consumer good (green), and Technology (turquoise) sectors. On the contrary at 90 minutes both the Service and Non-Cyclical Consumer Good clusters are broken while in addition to the the Financial and technology, also the Basic Material sector (the second most clustered sector at this time scale) is perfectly identified.

4 Conclusions

The analysis carried out in this paper provides further evidence that the Fourier method of computing the correlation matrix from high-frequency data is better than the traditional Pearson alternative in terms of generating smooth, robust estimates from small sample data sets.

The Fourier MST representation of the correlation matrix exhibits similar characteristics to those found in previous studies. The graph is centralised on a very short time scale and becomes more dispersed on longer time scales. The analysis of the entire correlation matrix provides additional evidence of the structural changes that affect the correlation matrix at different time scales. As a result of our analysis we find that Wal-Mart Stores and General Electric are the two most central stocks both in the MST and in the full correlation network on time scales shorter than 20 minutes. Furthermore, Wal-Mart has the highest centrality score at all time scales up to two hours. At aggregate level we have identified the Financial, Energy and Services sectors as the most intra-connected at short time scales with the Financial the most intra-connected at all time scales up to two hours.

5 Acknowledgements

We are very grateful to Roberto Renò, Vanessa Mattiussi, Anirban Chakraborti and Rosario Mantegna for stimulating discussions.

References

[1] Malliavin P, Mancino M (2002) Fourier series method for measurement of multivariate volatilities, Finance & Stochastics 6(1), 49-61.

[2] Barucci E, Renò R (2002) On measuring volatility and the GARCH forecasting performance, Journal of International Financial Markets, Institutions and Money, 12 182-200.

[3] Barucci E, Renò R (2002) On measuring volatility of diffusion processes with high frequency data, Economics Letters 74, 3(2) 371-378.

[4] Renò R (2003) A closer look at the Epps effect, International Journal of Theoretical and Applied Finance, 6 (1), 87-102.

[5] Precup O V, Iori G, (2004) A Comparison of High-Frequency Cross-Correlation Measures, Physica A, Vol 344/1-2, 252-256.

[6] Precup O V, Iori G, (2007) Cross-correlation in the high-frequency domain, European Journal of Finance (forthcoming).

[7] Mattiussi V, Iori G, (2007) "Currency futures volatility during the 1997 East Asian crisis: an application of Fourier analysis", Chapter 5 in B.A. Goss (ed), Debt, Risk and Liquidity in Futures Markets, London and New York: Routledge (forthcoming).

[8] Bonanno G, Lillo F, Mantegna R N, (2001) High-frequency Cross-correlation in a Set of Stocks, Quantitative Finance 1, 96-104.

[9] Bonanno G, Caldarelli G, Lillo, Mantegna R N, (2002) Topology of correlation based minimal spanning trees in real and model markets, Physical Review E 68, 046130.

[10] Bonanno G, Caldarelli G, Lillo F, Miccichè S, Vandewalle N, Mantegna R N (2004) Networks of equities in financial markets, The Eur. Phys. J. B 38, 363-371.

[11] Onnela J-P, Chakraborti A, Kaski K, Kertesz J, Kanto A (2003) Asset trees and asset graphs in financial markets, Physica Scripta Online Vol. T106, 48.

[12] Onnela J-P, Chakraborti A, Kaski K, Kertesz J (2003) Dynamic asset trees and Black Monday, Physica A 324, 247-252.

[13] Onnela J-P, Kaski K, Kertesz J (2004) Clustering and information in correlation based financial networks, Eur. Phys. J. B Vol. 38 No. 2, p.353.

[14] Tumminello M, Aste T, Di Matteo T, Mantegna R N, Proc. Natl. Acad. Sci. USA 102, no. 30,10421-10426 (2005) .

[15] Coronnello C, Tumminello M, Lillo F, Miccich S, Mantegna R N, Economic sector identification in a set of stocks traded at the New York Stock Exchange: a comparative analysis.

[16] Tumminello M, Di Matteo T, Aste T, Mantegna R N, Correlation based networks of equity returns sampled at different time horizons.

[17] Barrat A, Barthélemy M, Vespignani A, (2004), Weighted evolving networks: coupling topology and weights dynamics, Phys. Rev. Lett. **92** 228701.

[18] Newman M E J, (2004), *Analysis of weighted networks* , Phys. Rev. E 70, 05613.

[19] Kalna G, Higham D J, Clustering Coefficients for Weighted Networks

[20] Epps T (1979) Comovements in stock prices in the very short run, Journal of the American Statistical Association 74, 291-298.