Department of Economics
School of Social Sciences

Cross-Correlation Measures in the High-Frequency Domain

Ovidiu V. Precup [1]
King's College London

Giulia Iori[2]
City University

Department of Economics
Discussion Paper Series
No. 05/04

[1] Department of Mathematics, King's College London, Strand, London, WC2R 2LS, UK. Email: ovidiu.precup@kcl.ac.uk

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

Cross-Correlation Measures in the High-Frequency Domain

Ovidiu V. Precup*
Department of Mathematics, King's College London
Strand, London WC2R 2LS, U.K.
E-mail: ovidiu.precup@kcl.ac.uk

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

October 28, 2005

Abstract

On a high-frequency scale, financial time series are not homogeneous, therefore standard correlation measures can not be directly applied to the raw data. To deal with this problem the time series have to be either homogenized through interpolation or methods that can handle raw non-synchronous time series need to be employed. This paper compares two traditional methods that use interpolation with an alternative method applied directly to the actual time series. The three methods are tested on simulated data and actual trades time series. The temporal evolution of the correlation matrix is revealed through the analysis of the full correlation matrix and of the Minimum Spanning Tree representation. To perform the analysis we implement several measures from the theory of random weighted networks.

Keywords: High-Frequency Correlation, Fourier method, Epps Effect, Minimum Spanning Tree, random networks.

1 Introduction

A robust correlation measure for high-frequency data has direct applications in derivatives pricing, risk management, portfolio optimization and is necessary in the study of market microstructure effects (information aggregation, "Epps effect" [1]).

The conventional method of computing correlations from high-frequency data is the Pearson coefficient after the time series have been synchronized through an interpolation scheme. Recently, Dacorogna et al(2001) proposed a method similar to the Pearson coefficient with a weight factor that depends on the joint volatility of the time series. This method also requires synchronous time series.

Barucci and Renò(2002), Renò(2003) have adapted a Fourier method developed by Malliavin and Mancino(2002) to the computation of FX rates correlations. The Fourier method

*Corresponding author
[1] The correlation between stocks falls when decreasing the time scale, Epps(1979)

can be directly applied to the actual time series to obtain correlation statistics. An alternative method that uses the raw time series is by de Jong and Nijman(1997). It is based on a regression type estimator but it relies on a rather strong assumption of independence between prices and transaction times.

In this paper we compare the two interpolation based methods (Pearson and Co-volatility Weighted) with the Fourier. We show that the Fourier method generates more accurate results than the other two. We use one month (September 2002) of high frequency trades in the member stocks of the S&P100 [2] index to compute the cross-correlation matrix of returns. In the context of this paper, high-frequency data is defined as the raw time series of trades. The time interval between transactions ranges from 0 seconds (several distinct trades recorded at the same time) to 40 minutes. In our opinion one month worth of trades is a sufficient data set in terms of statistical power. We select only a month of data on purpose because this is of higher practical use to a market agent who is interested in the short time evolution of correlation.

The three correlation measures are first introduced and then compared on simulated and actual trades data in section 2. The "Epps effect" becomes apparent during this analysis. The time scale evolution of correlation matrices is investigated trough the network analysis of the full correlation matrix and of its minimum spanning tree (MST) representation in section 3. Section 4 concludes.

2 Correlation Measures

An extension of the standard Pearson correlation measure is proposed in Dacorogna et al(2001) by incorporating a "co-volatility weighting" for the time series. The weight has the role of emphasizing periods where trading has a noticeable effect on asset prices.

The method works as follows. Let X, Y be two asset price time series which have been homogenized and synchronized to a time step Δt, co-volatility weights are given by ω_i and the time length of the trading period is T. We define Δx, Δy as the corresponding log returns series on a time scale Δt and ΔX, ΔY as the log returns on a larger time scale $m\Delta t$. The co-volatility adjusted correlation measure is defined as:

$$\rho(\Delta X, \Delta Y) = \frac{\sum_{i=1}^{T/m\Delta t}(\Delta X_i - \overline{\Delta X})(\Delta Y_i - \overline{\Delta Y})\omega_i}{\sqrt{\sum_{i=1}^{T/m\Delta t}(\Delta X_i - \overline{\Delta X})^2\omega_i \sum_{i=1}^{T/m\Delta t}(\Delta Y_i - \overline{\Delta Y})^2\omega_i}} \tag{1}$$

$$\text{where} \quad \omega_i = \sum_{j=1}^{m}(|\Delta x_{i\cdot m-j} - \overline{\Delta x}_{i\cdot m}| \cdot |\Delta y_{i\cdot m-j} - \overline{\Delta y}_{i\cdot m}|)^\alpha, \tag{2}$$

$$\overline{\Delta x}_{i\cdot m} = \sum_{j=1}^{m}\frac{\Delta x_{i\cdot m-j}}{m}, \quad \Delta X_i = \sum_{j=1}^{m}\Delta x_{i\cdot m-j}, \quad \overline{\Delta X} = \frac{\sum_{i=1}^{T/m\Delta t}\Delta X_i \cdot \omega_i}{\sum_{i=1}^{T/m\Delta t}\omega_i} \tag{3}$$

Setting $\omega_i = 1$ reduces (1) to the standard Pearson coefficient. In this paper as in Dacorogna et al(2001) $\alpha = 0.5$ but this can be varied so that more weight is given to periods where the returns volatility is above average. In Dacorogna et al(2001) $m = 6$, in our analysis it varies from 3 to 480 (the number of time units of Δt in the trading day). This was determined by the choice of $\Delta t = 60$ seconds which was taken as a tradeoff value for the average trading interval pattern. The intention is to avoid extensive imputation towards the end of the trading

[2] data source: NYSE Trades and Quotes (TAQ) database

day when there are few transactions occurring.

The Fourier method is model independent, it 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(2002) so only the main results are given 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:

$$a_k(dp_i) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dp_i(t), \quad b_k(dp_i) = \frac{1}{\pi}\int_0^{2\pi} \sin(kt)dp_i(t), \quad k\geq 1. \tag{4}$$

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 (4) to be in discrete form:

$$\begin{aligned} a_k(dp_i) &= \frac{1}{\pi}\sum_{n=1}^{N}\Big([p_i(t_n)\cos(kt_n) - p_i(t'_n)\cos(kt'_n)] - p_i(t'_n)[\cos(kt_n) - \cos(kt'_n)]\Big), \\ b_k(dp_i) &= \frac{1}{\pi}\sum_{n=1}^{N}\Big([p_i(t_n)\sin(kt_n) - p_i(t'_n)\sin(kt'_n)] - p_i(t'_n)[\sin(kt_n) - \sin(kt'_n)]\Big). \\ \text{where} \quad & t'_n = t_{n-1} \end{aligned} \tag{5}$$

In (5), 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 :

$$a_0(\Sigma_{ij}) = \lim_{\tau\to 0}\frac{\pi\tau}{T}\sum_{k=1}^{T/2\tau}[a_k(dp_i)a_k(dp_j) + b_k(dp_i)b_k(dp_j)]. \tag{6}$$

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 $\rho_{ij} = \hat{\sigma}^2_{ij}/(\hat{\sigma}_{ii}\cdot\hat{\sigma}_{jj})$.

2.1 Methods Compared

We tested the Fourier method on simulated bivariate GARCH(1,1) processes in a similar setting to that in Renò(2003). The GARCH(1,1) model is the following:

$$\begin{aligned} p_1(t+1) &= p_1(t) + \sigma_1(t)\epsilon_1\sqrt{\Delta t}, \\ p_2(t+1) &= p_2(t) + \sigma_2(t)(\rho\epsilon_1 + \epsilon_2\sqrt{1-\rho^2})\sqrt{\Delta t} \\ \sigma_1^2(t+1) &= \sigma_1^2(t) + \lambda_1[\omega_1 - \sigma_1^2(t)]\Delta t + \sigma_1^2(t)\epsilon_3\sqrt{2\lambda_1\omega_1\Delta t} \\ \sigma_2^2(t+1) &= \sigma_2^2(t) + \lambda_2[\omega_2 - \sigma_2^2(t)]\Delta t + \sigma_2^2(t)\epsilon_4\sqrt{2\lambda_2\omega_2\Delta t}. \end{aligned} \tag{7}$$

The GARCH parameters are:
$\theta_1 = 0.035,\ \omega_1 = 0.636,\ \lambda_1 = 0.296,\ \theta_2 = 0.054,\ \omega_2 = 0.476,\ \lambda_2 = 0.480\ \epsilon_{1..4} \sim i.i.d.\ N(0,1),\ \rho = Corr(log\ ret\ p_1,\ log\ ret\ p_2),\ \Delta t = 1/86400.$

The model was run 1000 times for a 1 day trading period length (86400 seconds) with a time step of 1 second. The time interval between trades in S&P100 equities approximately follows an exponential distribution with rate parameter β (mean) in the range 1 (very liquid stock) to 67 (least liquid stock) seconds. In Figure 1 the histogram of inter-transaction times for a low liquidity stock (Black & Decker) is plotted. The near straight slope (on a log-linear scale) of the plot indicates that the underlying distribution of transaction times is exponential.

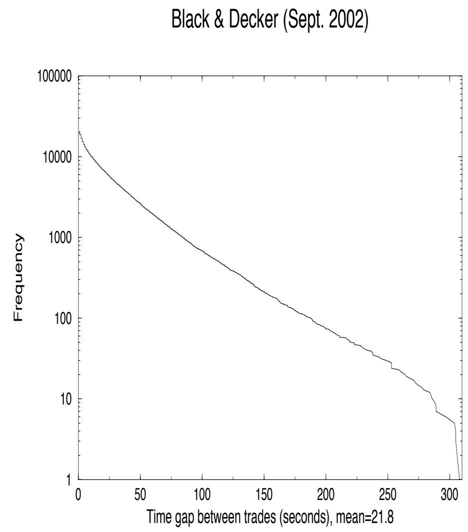

Figure 1:

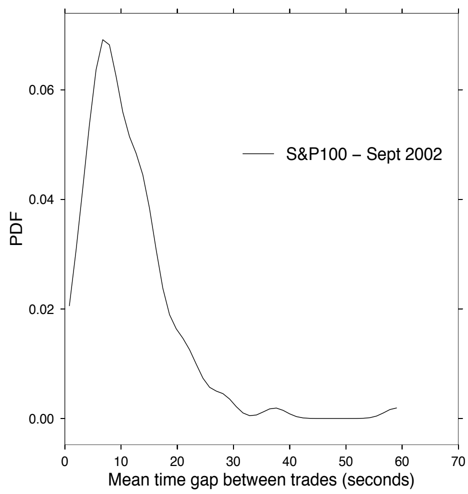

Figure 2:

Figure 2 plots the probability density function (PDF) of intertrade times of all stocks in the S&P100 for September 2002. Half the values are below 9.2 seconds and three quarters are below 14 seconds. This indicates that most of the stocks have medium or high liquidity. The most and the least liquid stocks are Microsoft Corp. and Allegheny Technologies respectively.

We sampled the generated GARCH process using the exponential distribution and varied β so as to resemble actual trading patterns. Figure 3 is a plot of the simulation results with an induced correlation $\rho = -0.70$.

The mean exponential rates (β) for the generated series are indicated by the numbers in the plot legend. For example "Synch 20" are two synchronous time series with $\beta = 20$. $P1 - 5, P2 - 15$ are asynchronous series with $\beta_1 = 5$ and $\beta_2 = 15$ respectively. A $\beta < 5$ corresponds to a high liquidity stock whilst a $\beta > 15$ is associated with a low liquidity one. The Full GARCH line represents the actual simulation data without exponential sampling and is to be taken as a benchmark. When the time series are synchronously sampled ("Synch 20"), the correlation structure mimics the benchmark perfectly. This means that the Fourier method works very well on synchronous series with random gaps. For non-synchronous series, the correlation spectra on very short time scales (less than 10 minutes) are dependent on the exponential rates. The higher the mean rate of a series, the faster it deviates from the induced correlation level at short time scales. This is irrespective of how low the exponential rate of the corresponding series happens to be. $P1-3, P2-5$ are series with low rates and their correlation

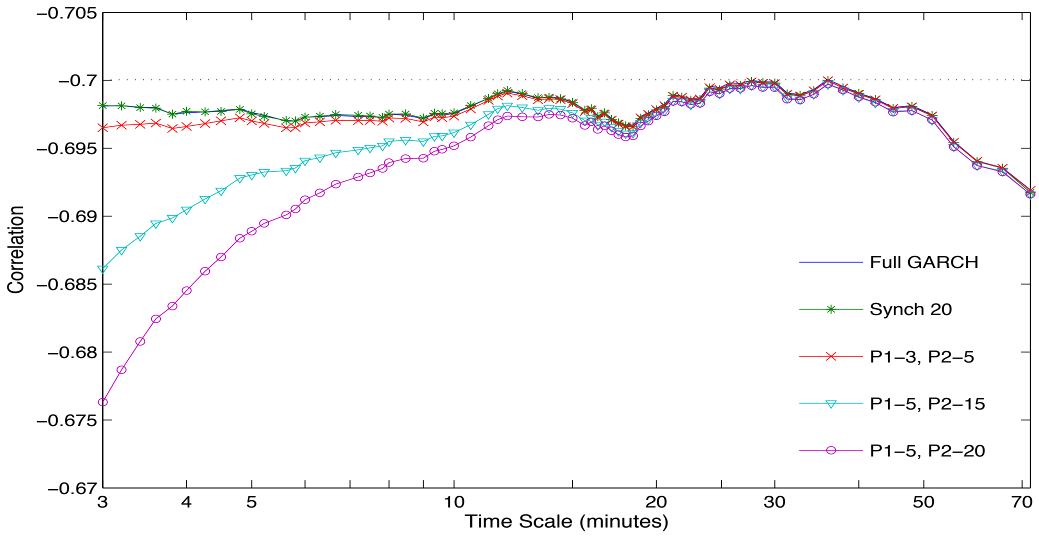

Figure 3: Simulated bivariate GARCH(1,1) - Fourier method

spectrum closely follows the benchmark on all time scales. The series $P1-5, P2-15$ and $P1-5, P2-20$ display noticeable deviations from the benchmark on time scales under 10 minutes. An explanation for the fast decay is that for exponentially distributed intertrade times the proportion and magnitude of large positive deviations from the mean increase with the mean itself. Therefore series with larger mean intertrade times will be further out of synch with each other, thus reducing their correlations at high frequencies. On time scales greater than 10 minutes all the series converge very fast towards the benchmark. The simultaneous decay of all the correlation spectra from the induced level (-0.7) on time scales above 50 minutes is due to fewer data points available for estimation.

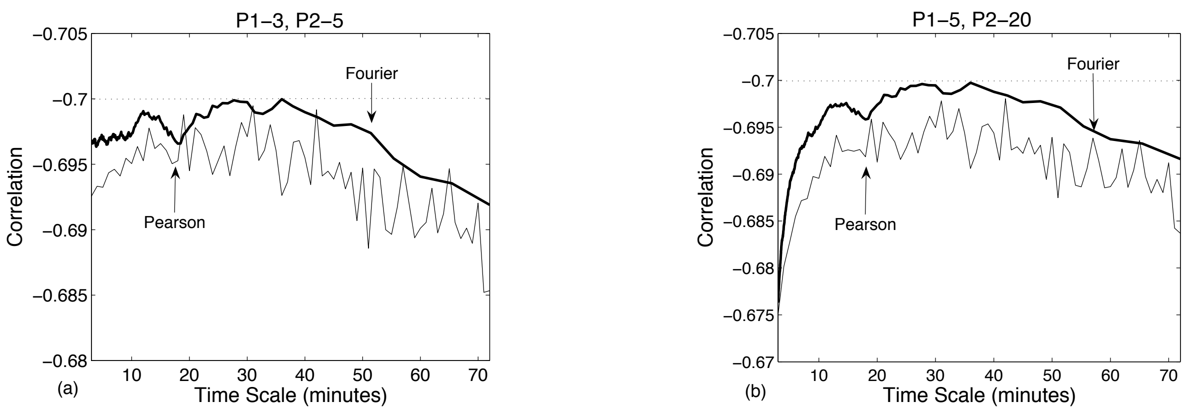

Figure 4: Simulated bivariate GARCH(1,1) - Fourier and Pearson methods

Besides successfully testing the Fourier method, the simulated model also provides an insight into how non-synchronicity in transactions affects the correlation spectra on very short time scales.

In Figure 4 we compare the results given by the Fourier and Pearson methods applied to the simulated GARCH series. The co-volatility weighted method generates correlation spectra very similar to those of the Pearson method and have not been plotted in order to improve the clarity of Figure 4. Both parts of Figure 4 show the Pearson method generating correlation estimates inferior to the Fourier method in terms of magnitude and smoothness. On time scales under 10 minutes the Pearson correlation estimates are significantly less volatile than those on higher time scales. In Figure 4.b where the second series has a large (20 seconds) mean exponential rate, the Pearson method displays a decay in the correlation estimates on a very short time scale similar to that produced by the Fourier method.

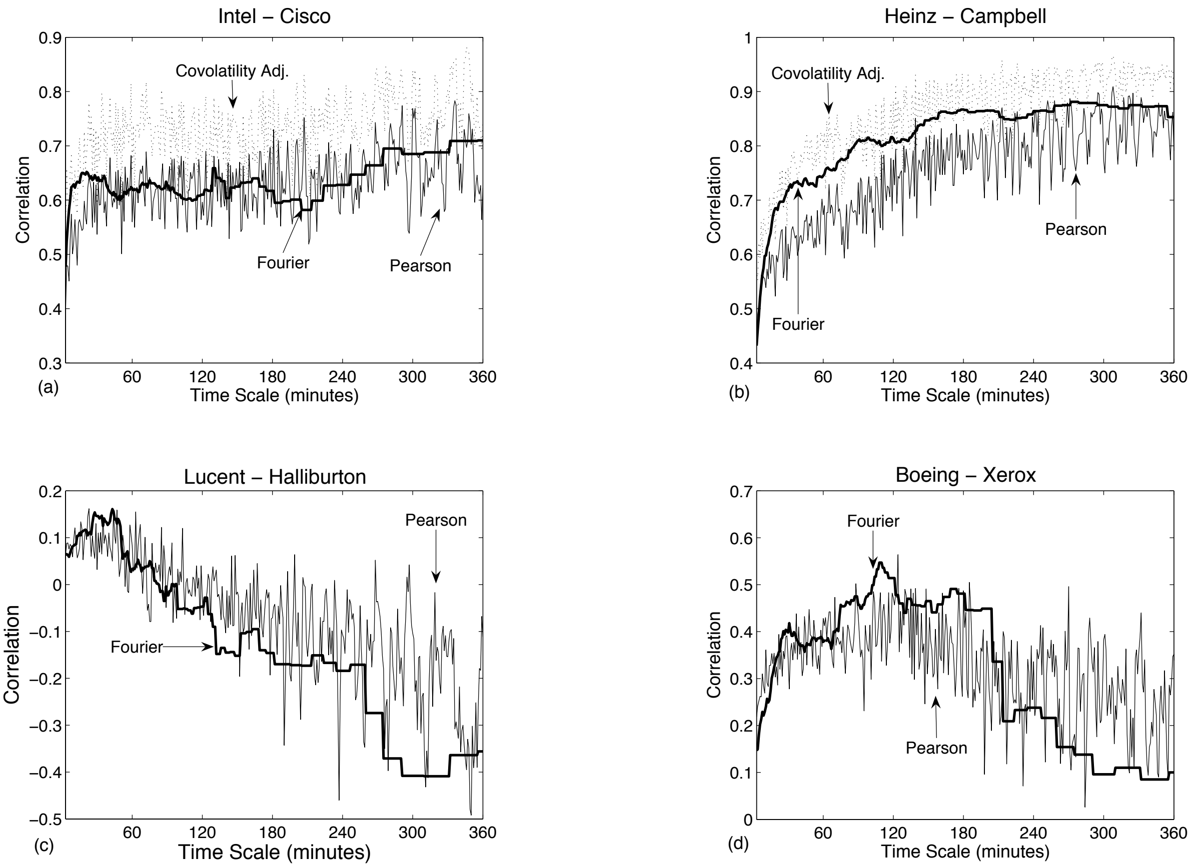

Figure 5: Example correlation spectra of stocks

Figure 5 is an example of correlation spectra for four pairs of stocks. In the first two plots all three methods of computing the correlation are shown whilst in the following two only the results of the Fourier and Pearson methods are presented. The Co-volatility Adjusted method produces results very similar to the Pearson method for the latter two plots. In all cases the Fourier correlation method provides a much smoother spectrum than the other two methods which use interpolation. The Fourier correlation trails a moving average of either the Co-volatility Adjusted or Pearson coefficient depending on stock and time scale. The variation in the correlation coefficient computed with interpolation based methods can be very large (from 0.5 to 0.1 for Boeing-Xerox) between consecutive time scales. The rise in the Pearson/Co-volatility Adjusted correlation coefficient variation with time scale could be attributed to loss of statistical power. A one month data set contains on average 22 trading days with approximately 8 hours of market activity [3]. Analysis with interpolation methods on time scales larger than 4 hours will therefore employ fewer than 44 data points and fewer than

[3] The NYSE trading hours are 09:30 - 16:00 but the TAQ database records trades outside these times as well

22 wave harmonics for computing the Fourier correlation. By comparison, on a 3 minutes time scale there are 3520 data points available for the Pearson/Co-volatility Adjusted methods and 1760 wave harmonics for the Fourier. Our interest is in the very high frequency regime though (under 3 hours) and the selected data sets provide sufficient statistics on this time scale. The lower frequency results have been used for comparison with the high frequency domain.

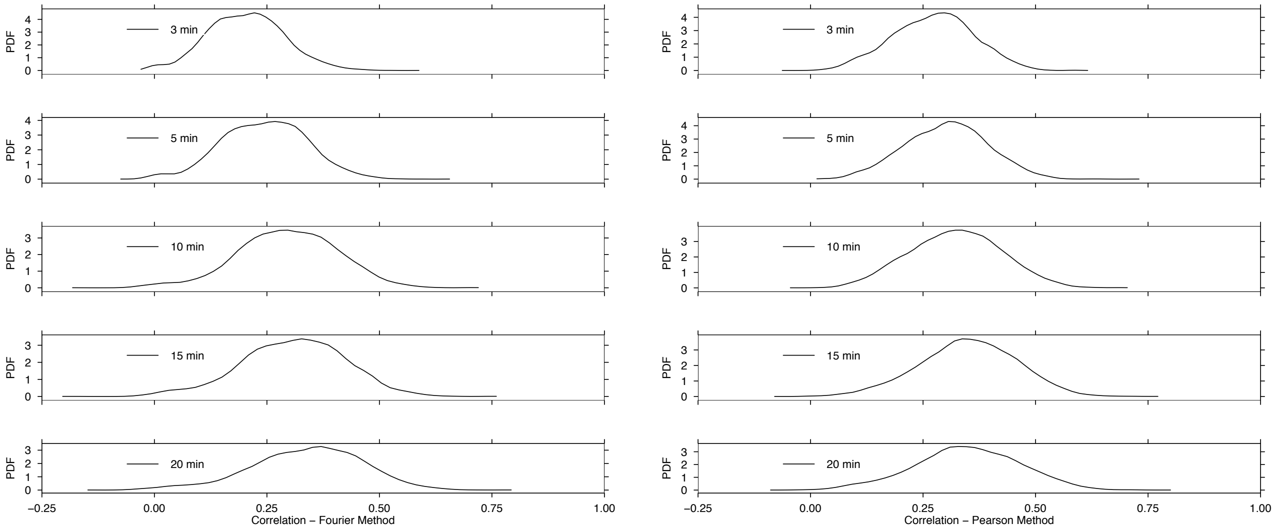

Figure 6: Correlation probability density functions at short time scales

The Fourier method results resemble a step function on time scales greater than 2 hours due to conversion from the frequency to the time domain. The wave harmonics (k in equations (5) and (6)) are integers which determine the magnitude of $T/2\tau$, where T is the total time length of the series and τ is the time scale of analysis. When k is large the difference in time scale for consecutive k values is well under a minute. As k decreases, the time scale difference increases to over a minute which leads to the step function.

The plots in Figure 5 exemplify four different types of correlation evolution with time scale. Plots a) and b) show the correlation structure of two intra-sector pairs of stocks. The correlation between Intel and Cisco (highly liquid, β_{intc}=1, β_{csco}=1.2) reaches a stable level after approximately 15 minutes whilst for Heinz and Campbell (lower liquidity, β_{hnz}=14, β_{cpb}=23) it takes about 2 hours to stabilize. Plots c) and d) are inter-sector examples. In Figure 5.c, on a time scale smaller than an hour Lucent (high liquidity, β_{lu}=3.7) and Halliburton (average liquidity, β_{hal}=11.3) have little correlation and this turns into anti-correlation when the time scale increases. Boeing (β_{ba}=8.4) and Xerox (β_{xrx}=15.9) in Figure 5.d start as poorly correlated on a 3 minutes time scale, the correlation coefficient then rises steadily to 0.55 on a 2 hour time scale and falls back to a less significant level (0.2) at 4 hours.

The Epps effect can also be detected in the Figure 5 plots and displays one of the properties described in Lundin and Dacorogna(1999), namely that the time scale at which it can be observed is shorter for highly traded assets. The correlation decay observed in Figure 3 at short time scales due to non-synchronicity in trades can not account for the correlation structure that develops over 2 hours in Figures 5. b) and d). Thus, the Epps effect present in the correlation structure of illiquid stocks can not be explained by non-synchronicity in transactions but is an actual market microstructure phenomenon related to the information aggregation and price formation processes. Further studies are required to understand the contributing factors of the Epps effect and the role of synchronicity in transactions.

The results of the two methods are also compared in Figure 6 and Table 1 provides the corresponding summary statistics. The source data are correlation matrices of S&P100 (Sept. 2002) stock returns computed at time scales ranging from 3 to 20 minutes. It is expected that the average correlation level increases with the time scale as predicted by the "Epps effect".

The Fourier correlation measure generates probability density functions (PDF) with characteristics distinguishable from those of the Pearson. The mean and standard deviation of correlation are unambiguously increasing with the time scale for both methods with one exception. On a 20 minutes time scale, the PDF for the Pearson method is left shifted relative to the 15 minutes case whilst the Fourier one does not display this anomalous behaviour. The reason for the left shift is the high level of noise captured in the Pearson correlation estimates as already shown in Figure 5. The left tails are generally heavier (negative skewness) for the Fourier method. The tails of the PDFs grow asymmetrically with the time scale. The positive correlation values develop faster relative to the negative ones.

Method	Time Scale	Min	Max	Mean	Std Dev	Skew	Excess Kurtosis
Fourier	3	-0.030	0.588	0.204	0.085	0.121	0.044
	5	-0.075	0.656	0.245	0.095	-0.043	-0.033
	10	-0.182	0.720	0.296	0.111	-0.175	0.145
	15	-0.204	0.760	0.309	0.116	-0.183	0.121
	20	-0.148	0.793	0.339	0.124	-0.311	0.209
Pearson	3	-0.063	0.616	0.273	0.090	-0.061	-0.222
	5	0.014	0.730	0.302	0.092	-0.012	-0.131
	10	-0.045	0.704	0.322	0.102	0.002	-0.234
	15	-0.080	0.772	0.350	0.108	-0.126	0.059
	20	-0.089	0.800	0.348	0.117	-0.099	-0.039

Table 1: Short time scale correlation statistics - time scale in minutes

3 Network Analysis

The correlation matrix can be represented as a network of vertices (stocks) and weighted links (correlations). The *degree* of a vertex in the network is defined 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 points i and j. We use the degree measure for the analysis of the MST and the strength for analysis of the overall correlation matrices.

The Minimum Spanning Tree (MST) is a tool borrowed from Graph Analysis that has been used - Mantegna and Stanley(2000), Bonanno et al(2000,2001), Onnela et al(2003) - to extract economic information from the network representation of a correlation matrix. Mantegna and Stanley(2000) provide an extensive account of the properties of MSTs and their applications to securities correlation matrices. A good algorithm for implementing the MST is that by Kruskal in Cormen et al(2001). One of the appealing features of the MST is that it enables a clear and direct visualization of the important links in a network. This also facilitates the analysis of the underlying structure in time.

In previous studies - Bonanno et al(2000) - on high-frequency data covering 4 years up to the end of 1998 but on lower frequency resolutions (shortest time scale was 20 minutes) there is evident clustering of stocks in the MST according to economic sectors. We first want to check whether this feature is present at a much shorter time scale in a single month. Subsequently we discuss the robustness of the MST analysis.

Figures 7 are examples of minimum spanning trees with vertices colour coded according to the different economic sectors of the stocks. The table in the Appendix is an indexed list of the component stocks of the S&P100 grouped according to sector classification[4]. The stock indices from this table have been used to plot the MSTs in Figures 7 and Table 2 provides the colour legend.

Colour	Sector
turquoise	Technology
red	Financial
yellow	Basic Material
grey	Capital Good
rose	Conglomerates
violet	Energy
brown	Services
green	Transport
orange	Utilities
light blue	Health Care
darkgreen	Non Cyclical Consumer Goods
blue	Cyclical Consumer Goods

Table 2: Colour Codes for the MST figures

The MST on 3 minutes time scale are centralized graphs with several vertices that aggregate a large number of connections. At longer time scales the graph structure is a lot more dispersed with no obvious central vertex. The variation in MST centralization with time scale has also been observed by Bonanno et al. (2000).

[4]The classification is the one provided by http://biz.yahoo.com/p/s_conameu.html.

Figure 7 at 120 minutes shows a number of well defined clusters for the Technology, Financial, Non Cyclical Consumer Goods and Utilities sectors. The MST derived from the Pearson correlations at the same time scale does not have the sectors well defined.

Method	Index	Stock
	97	Wal-Mart Stores
	13	Boise Cascade Corp
Fourier	66	Microsoft Corp
	14	Black & Decker
	97	Wal-Mart Stores
Pearson	11	Bank of America
	91	US Bancorp
	27	Du Pont

Table 3: Vertex with average highest degree in the MST across all time scales

The MST representation of the correlation matrix has a shortcoming. The algorithm discards a large amount of data in arriving at a solution and this makes it very sensitive to the structure of the correlation matrix. Table 3 lists the actual stocks corresponding to the four most frequently most connected vertices of the MST. In previous work by Bonanno et al [3] General Electric was reported as the most central stock. We do not recover this result for the month of data we have analyzed.

Figure 8 shows the evolution of the degree of the most connected stock with the time scale. At time scale above 30 minutes the degree stabilize around 6 while at shorter time scale the tree structure is more centralized.

Figure 9 provides a different perspective on the correlation matrix.The CDF plots of the strenght and normalized strength [5] at different time scales are very similar. This indicates that while correlation increases (or decreases for the negative ones) the distributions of the strengths collapse on each other when rescaled and do not reveal any significative change of structure, as observed in the MST.

Table 4 shows that the variation in the correlation structure is substantially less than that from the MST. Three of the strongest correlated stocks (Wal-Mart Stores, US Bancorp ad Bank One Corp) are identified by both methods. Nonetheless figure 10 shows that the The Fourier correlation is overall more stable than the Pearson, and the most strongly connected stock are always the same two. The top four most connected vertices are largely different from the equivalently ranked strongest vertices. This may be due to a property of the MST algorithm which ignores some strong links (in order to avoid loops in the graph) in favour of weaker ones.

Method	Index	Stock
Fourier	97	Wal-Mart Stores
	91	US Bancorp
	9	American Express
	72	Bank One Corp
Pearson	97	Wal-Mart Stores
	91	US Bancorp
	11	Bank of America
	72	Bank One Corp

Table 4: Vertex with highest average strength across all time scales

In Figures 11 we analyze the time evolution of the MST and the overall correlation matrices. We define the distance between correlation matric as $\sum_{i,j} |c_{ij}(t) - c_{ij}(t_0)|$, and plot them as a function of the time scale. We selected 120 (2 hours) as the time scale t_0 against which to compare matrices. On a 2 hours time scale the correlation levels of stocks have stabilized so it makes a suitable reference scale. Figure 11 reveals that while the Fourier MSTs and correlations matrices are closer to the reference matrix at time scale close to the reference one, Pearson correlations, because of the high level of noise, are almost at a constant distance from the reference one. This is the strongest indication of the better performance of the Fourier method in our dataset.

In Figure 12 we analyze the weighted clustering coefficient at different time scales defined as

$$C_i^w = \frac{2}{N(N-1)} \sum_{j,h} \tilde{c}_{ij}(\tau)\tilde{c}_{ih}(\tau)\tilde{c}_{jh}(\tau), \tag{8}$$

where $\tilde{c}_{ij}(\tau) = c_{ij}/\bar{c}(\tau)$ and $\bar{c}(\tau) = \frac{1}{N(N-1)} \sum_{i,j} c_{ij}(\tau)$

The Fourier method clearly shows how that the clustering coefficient increases as time scales decreases below 20 minutes and then increases again. The high clustering at short time scale is an indication that correlation initially develop mainly intra-sectors and as time goes also inter-sector correlation become important.

[5]The strength of each vertex has been divided by the correlation matrix average at that time scale

Method	Index	Stock
Fourier	97	Wal-Mart Stores
	91	US Bancorp
	9	American Express
	72	Bank One Corp
Pearson	97	Wal-Mart Stores
	91	US Bancorp
	11	Bank of America
	72	Bank One Corp

Table 5: Most clustered vertex across all time scales

4 Conclusions

From the analysis carried out it can be inferred that the Fourier method of computing the correlation matrix from high-frequency data is better than the alternatives in terms of generating smooth, robust estimates in small sample data.

The MST representation of the correlation matrix exhibits similar characteristics to those found in previous studies. The graph is centralized on a very short time scale and becomes more dispersed on longer time scales. There is evident clustering of stocks along their economic sectors affiliation. However, the MST is structurally unstable between consecutive time scales due to its sensitivity to the noise in the correlation matrix. The analysis of the entire correlation matrix provides a more accurate picture of the structural evolution of correlations over time.

5 Acknowledgements

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

References

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

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

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

[4] Bonanno G, Lillo F, Mantegna R N (2001) Levels of complexity in financial markets, cond-mat/0104369.

[5] Bonanno G, Caldarelli G, Lillo, Mantegna R N (2002) Topology of correlation based minimal spanning trees in real and model markets, cond-mat/0211546.

[6] Bonanno G, Caldarelli G, Lillo F, Miccichè S, Vandewalle N, Mantegna R N (2004) Networks of equities in financial markets, The European Physical Journal B 38, Special Issue: Applications of Networks 363-371.

[7] Cormen T H, Leiserson C E, Rivest R L, Stein C (2001) Introduction to Algorithms (second edition), MIT Press, Cambridge Massachusetts.

[8] Dacorogna M, Gencay R, Müler U A, Olsen R B, Pictet O V (2001) An Introduction to High-Frequency Finance, Academic Press.

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

[10] de Jong F, Nijman T (1997) High frequency analysis of lead-lag relationships between financial markets, Journal of Empirical Finance 4, 259-277.

[11] Lundin M, Dacorogna M (1999) Correlation of high-frequency financial time series in Financial Markets Tick by Tick, (Lequeux P Editor), Wiley & Sons.

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

[13] Mantegna R N, Stanley H E (2000) An Introduction to Econophysics, Correlations and Complexity in Finance, CUP, Cambridge.

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

[15] Onnela J-P, Chakraborti A, Kaski K, Kertsz J, Kanto A (2003) Dynamics of market correlations: Taxonomy and portfolio analysis, Phys. Rev. E 68.

[16] Onnela J-P, Chakraborti A, Kaski K, Kertsz J (2002) Dynamic asset trees and portfolio analysis, The European Physical Journal B 30, 285-288.

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

[18] Onnela J-P, Kaski K, Kertsz J (2004) Clustering and information in correlation based financial networks, The European Physical Journal B Vol. 38 No. 2 Special Issue: Applications of Networks p. 353.

[19] Onnela J-P, Saramki J, Kertsz J, Kaski K (2004) Intensity and coherence of motifs in weighted complex networks, cond-mat/0408629.

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

6 Appendix

Table 6: S&P100 - grouped by sector

Index No.	Symbol	Company Name	Sector
6	AOL	AOL Time Warner	
24	CSC	Computer Sciences Corp.	
25	CSCO	Cisco Systems Inc.	
31	EMC	EMC Corp.	
48	HPQ	Hewlett-Packard Co.	
49	IBM	IBM	Technology
50	INTC	Intel Corp.	
57	LU	Lucent Technologies Inc.	
66	MSFT	Microsoft Corp.	
69	NSM	National Semiconductor	
70	NT	Nortel Networks	
73	ORCL	Oracle Corp.	
78	ROK	Rockwell Automation	
88	TXN	Texas Instruments	
90	UIS	Unisys Corp.	
100	XRX	Xerox Corp	
4	AIG	American International Group	
9	AXP	American Express	
11	BAC	Bank of America	
19	C	Citigroup Inc.	
21	CI	CIGNA Corp.	
45	HIG	Hartford Financial Services	
53	JPM	JP Morgan Chase	
55	LEH	Lehman Brothers Holdings Inc.	Financial
62	MER	Merrill Lynch	
67	MWD	Morgan Stanley Dean Witter	
72	ONE	Bank One Corp.	
91	USB	US Bancorp	
95	WFC	Wells Fargo & Co.	
1	AA	Alcoa Inc.	
7	ATI	Allegheny Technologies Inc.	Basic
13	BCC	Boise Cascade Corp.	Material
27	DD	Du Pont	
29	DOW	Dow Chemical Co.	
51	IP	International Paper Co.	
98	WY	Weyerhaeuser Co.	
10	BA	Boeing Co.	Capital
38	GD	General Dynamics Corp.	Good
47	HON	Honeywell International	
39	GE	General Electric	
63	MMM	3M Company	
80	RTN	Raytheon Co.	Conglomerate
89	TYC	Tyco International Ltd.	
92	UTX	United Technologies	
15	BHI	Baker Hughes Inc.	
41	HAL	Halliburton Company	
83	SLB	Schlumberger Ltd.	Energy
99	XOM	Exxon Mobil Corp.	
15	MCD	McDonalds Corp.	
20	CCU	Clear Channel Communications	
28	DIS	Walt Disney Co.	
43	HD	Home Depot Inc.	
44	HET	Harrah's Entertainment	
56	LTD	Limited Brands	Service

Continued on next page

Table 6 – continued from previous page

Index No.	Symbol	Company Name	Sector
58	MAY	May Dept. Stores	
59	NXTL	Nextel Communications Inc.	
79	RSH	RadioShack Corp.	
81	S	Sears Roebuck	
82	SBC	SBC Communications	
86	T	AT&T Corp.	
87	TOY	Toys "R" Us Inc.	
93	VIA	Viacom Inc.	
94	VZ	Verizon Communications Inc.	
97	WMT	Wal-Mart Stores	
17	BNI	Burlington Northern Santa Fe Corp.	Transport
26	DAL	Delta Airlines	
36	FDX	Fedex Corp.	
68	NSC	Norfolk Southern Corp.	
2	AEP	American Electric Power	Utilities
3	AES	AES Corp.	
32	EP	EL Paso Corp.	
33	ETR	Entergy Corp.	
34	EXC	Exelon Corp.	
85	SO	Southern Co.	
96	WMB	Williams Companies	
5	AMGN	Amgen Inc.	Health Care
12	BAX	Baxter International Inc.	
16	BMY	Bristol-Myers Squib	
42	HCA	HCA Inc.	
52	JNJ	Johnson & Johnson	
60	MDT	Medtronic Inc.	
61	MEDI	MedImmune Inc.	
65	MRK	Merck & Co.	
75	PFE	Pfizer Inc.	
77	PHA	Pharmacia Corp.	
8	AVP	Avon Products Inc.	Non-Cyclical Consumer Good
18	BUD	Anheuser-Busch Co.	
22	CL	Colgate-Palmolive Co.	
23	CPB	Campbell Soup Co.	
37	G	Gillette Co.	
46	HNZ	H.J. Heinz Co.	
54	KO	Coca-Cola Co.	
64	MO	Philip Morris Co.	
74	PEP	PepsiCo Inc.	
76	PG	Procter & Gamble	
84	SLE	Sara Lee Corp.	
14	BDK	Black & Decker Co.	Cyclical Consumer Goods
30	EK	Eastman Kodak Co.	
35	F	Ford Motor Co.	
40	GM	General Motors Corp.	

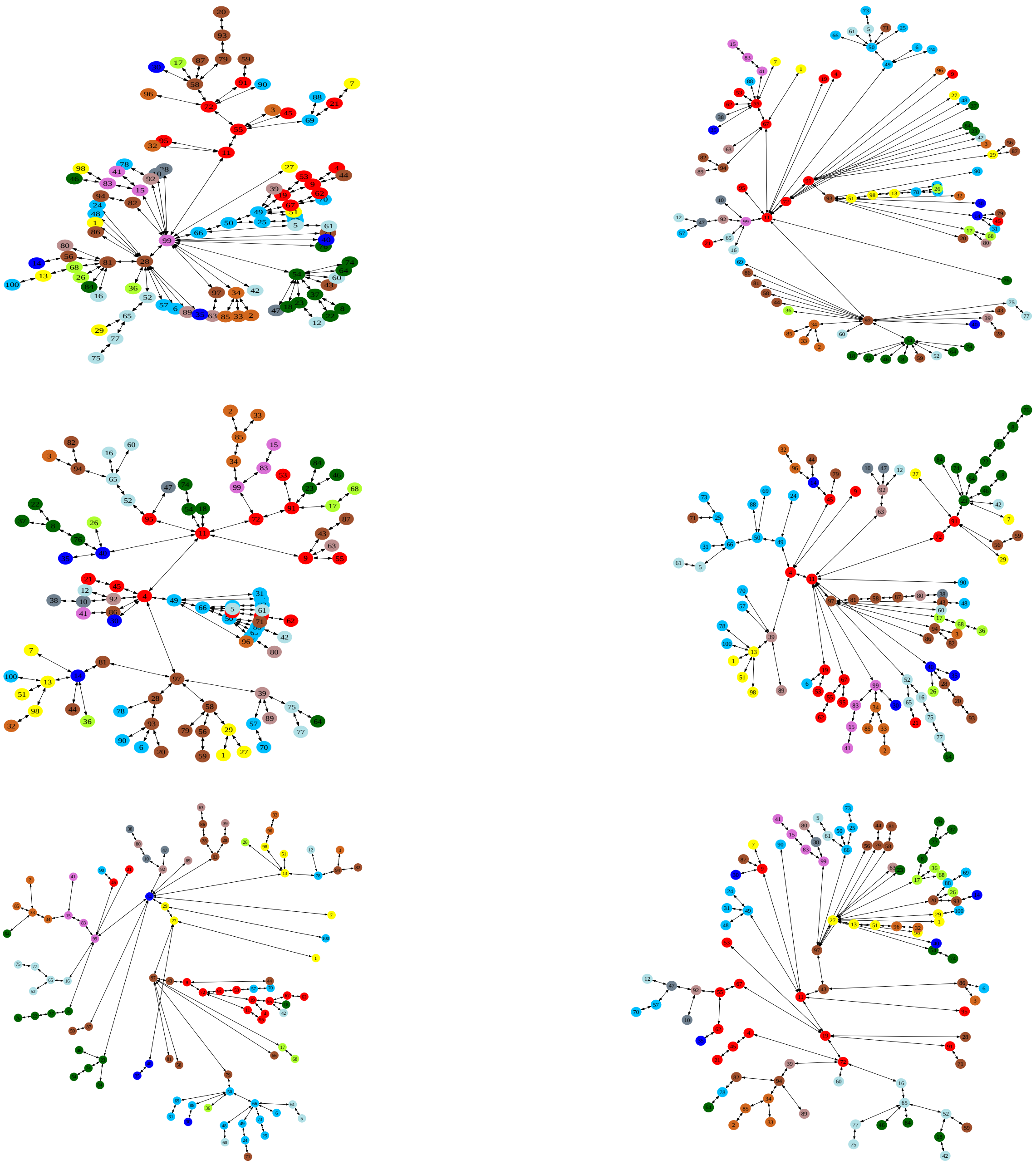

Figure 7: MST - 3,30,120 minutes, left - Fourier, right - Pearson

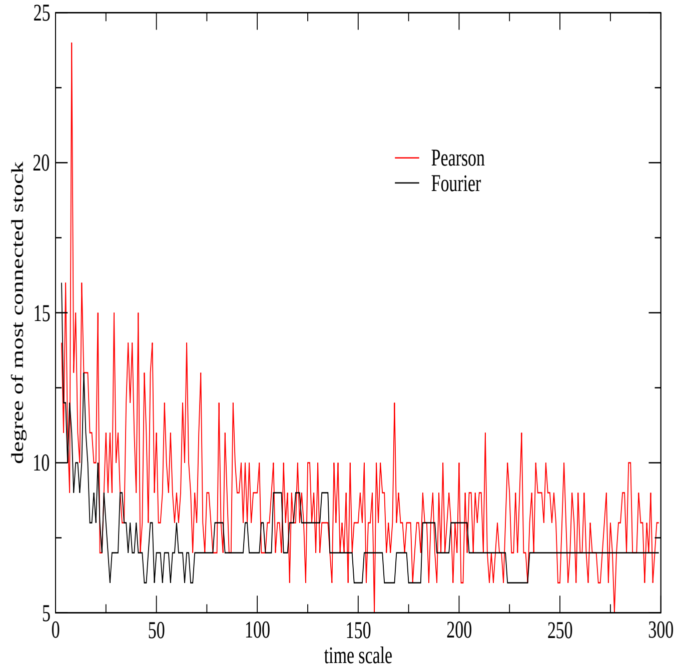

Figure 8: Comparison of the degree of the most connected stock in the MST for Fourier (black) and Pearson (red)

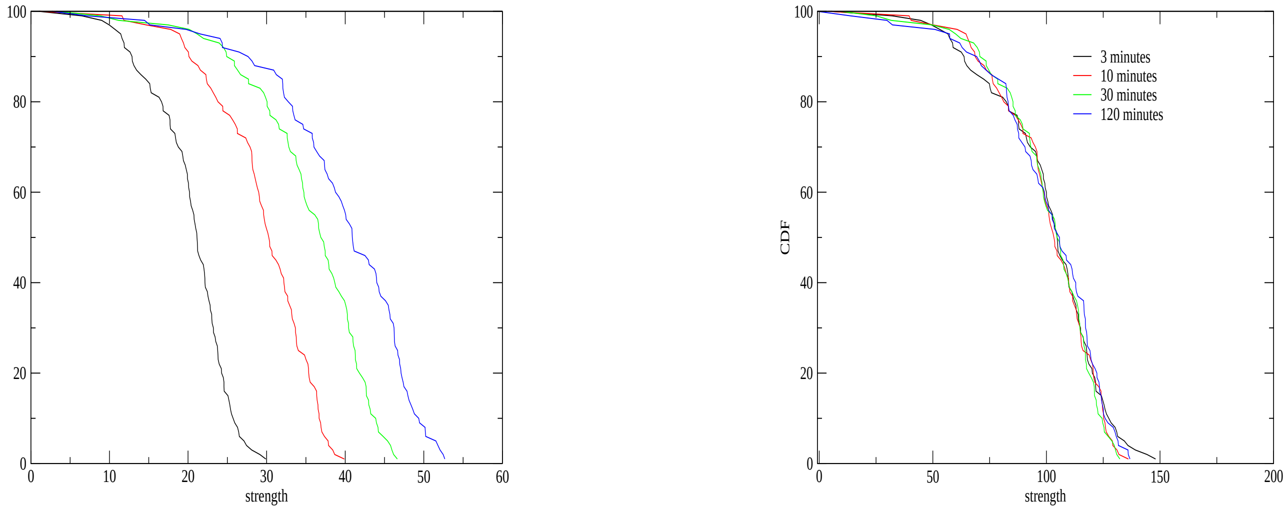

Figure 9: Strength (left) and normalized strength (right) distribution

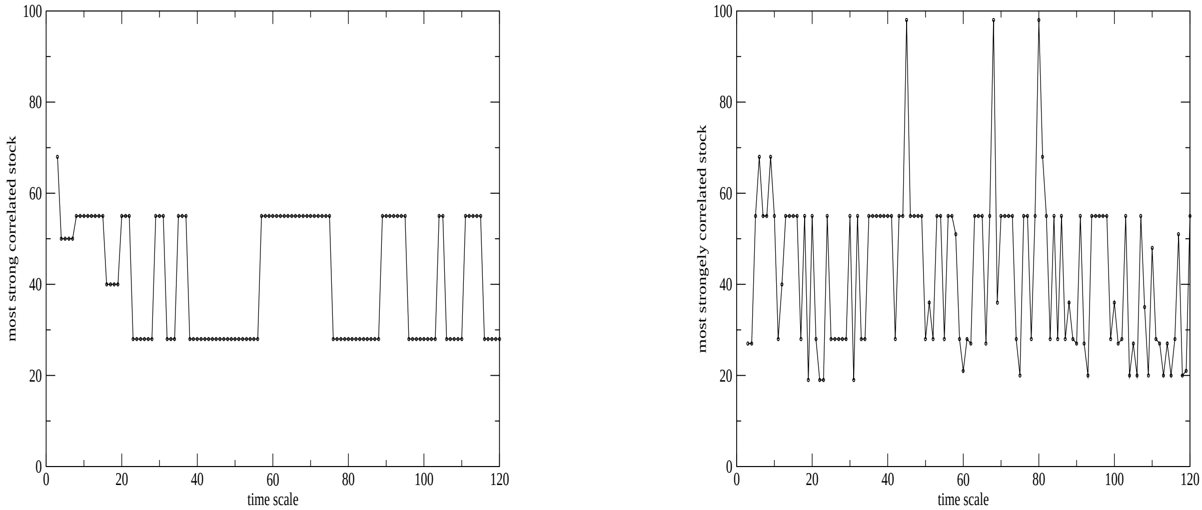

Figure 10: Most strongly correlated stock, left - Fourier, right - Pearson

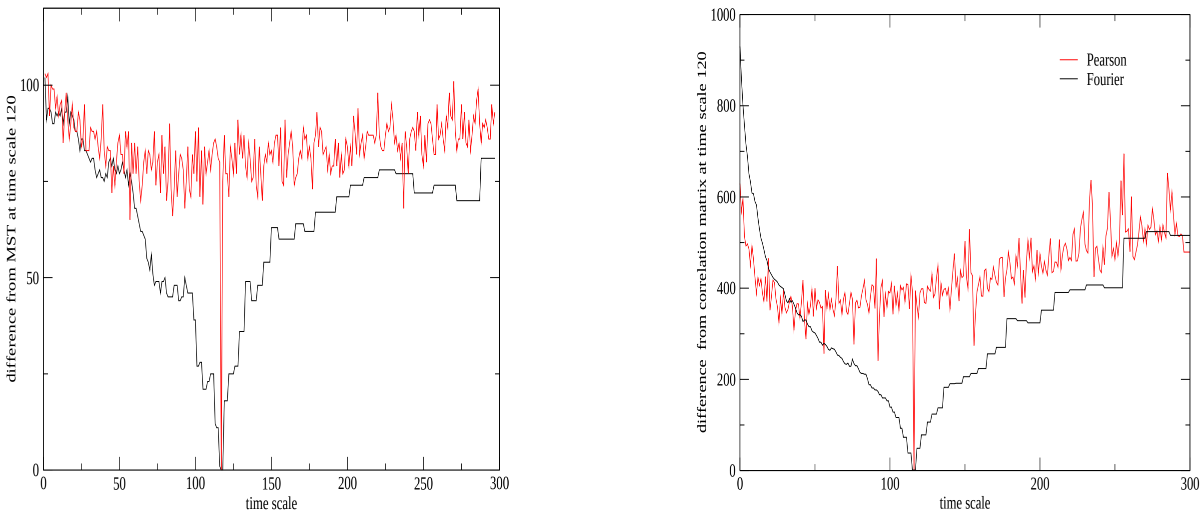

Figure 11: Differences from 2 hours time scale $\sum_{i,j} |c_{ij}(t) - c_{ij}(120)|$: left - MST, right - correlation matrices, Fourier - black, Pearson - red

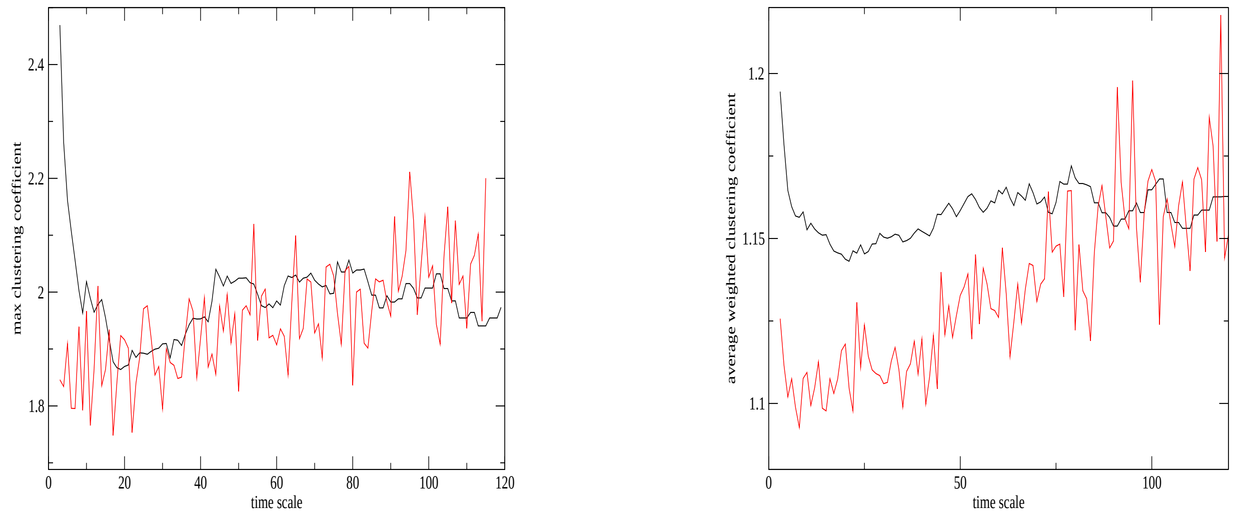

Figure 12: Maximum (left) and average (right) weighted clustering coefficient for full network with Fourier (black) Pearson (red)