Quantitative Finance, 2015
Vol. 15, No. 4, 569–578, http://dx.doi.org/10.1080/14697688.2014.946660

Routledge
Taylor & Francis Group

Partial correlation analysis: applications for financial markets

DROR Y. KENETT*†, XUQING HUANG*†, IRENA VODENSKA†‡, SHLOMO HAVLIN†§ and H. EUGENE STANLEY†

†Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
‡Administrative Sciences Department, Metropolitan College, Boston University, Boston, MA 02215, USA
§Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

(*Received 10 February 2014; accepted 10 July 2014*)

1. Introduction

Understanding the complex nature of financial markets remains a great challenge, especially in light of the most recent financial crisis. Recent studies have investigated large data-sets of financial markets, and have analysed and modelled the static and dynamic behaviour of this very complex system (Fama 1965, Lo and Craig MacKinlay 1990, Campbell *et al.* 1997, Lux and Marchesi 1999, Cont and Bouchaud 2000, Bouchaud and Potters 2003, Sornette 2004, Voit 2005, Eisler and Kertesz 2006, Takayasu 2006, Brock *et al.* 2009, Sinha *et al.* 2010, Abergel *et al.* 2011), suggesting that financial markets exhibit systemic shifts and display non-equilibrium properties.

One prominent feature in financial markets is the presence of observed correlation (positive or negative) between the price movements of different financial assets. The presence of a high degree of cross-correlation between the synchronous time evolution of a set of equity returns is a well-known empirical fact (Markowitz 1952, Campbell *et al.* 1997, Elton *et al.* 2009). The Pearson correlation coefficient (Pearson 1895) provides information about the similarity in the price change behaviour of a given pair of stocks. Much effort has been devoted to extracting meaningful information from the observed correlations in order to gain insights into the underlying structure and dynamics of financial markets (Laloux *et al.* 1999, Plerou *et al.* 1999, Laloux *et al.* 2000, Morck *et al.* 2000, Cizeau *et al.* 2001, Embrechts *et al.* 2002, Forbes and Rigobon 2002, Campbell *et al.* 2008, Campbell *et al.* 2008, Krishan *et al.* 2009, Podobnik *et al.* 2009, Aste *et al.* 2010, Pollet and Wilson 2010, Tumminello *et al.* 2010, Kenett *et al.* 2011a, Huang *et al.* 2013).

A large body of work has dealt with the systemic risks in the financial system in the presence of co-movement of financial

*Corresponding authors. Email: dkenett@bu.edu (D. Y. Kenett), eqing2700@gmail.com (X. Huang)

assets. To understand how risks propagate through the entire system, many studies have focused on understanding the synchronization in financial markets that is especially pronounced during periods of crisis (Haldane and May 2011, Bisias *et al.* 2012). Recent advancements include the CoVaR methodology (Adrian and Brunnermeier 2011), and Granger causality analysis (Granger 1969, Billio *et al.* 2012). These measures focus on the relationship of one variable on a second variable, for a given time period. Finally, much work has been focused on the issue of conditional correlation (Engle 2002) and event conditional correlation (Maugis 2014), and its applications in financial markets. However, a missing dimension of these methodologies is the investigation of many-body interaction between financial assets.

Despite the meaningful information provided by investigating the correlation coefficient, it lacks the capacity to provide information about whether a different stock(s) eventually controls the observed relationship between other stocks. Causality, and more specifically the nature of the correlation relationships between different stocks, is a critical issue to unveil. Thus, challenge is to understand the underlying mechanisms of influence that are present in financial markets. To overcome this issue, the use of the partial correlation coefficient (Baba *et al.* 2004) was recently introduced (Kenett *et al.* 2010).

A partial (or residual) correlation measures how much a given variable, say j, affects the correlations between another pair of variables, say i and k. Thus, in this (i, k) pair, the partial correlation value indicates the correlation remaining between i and k after the correlation between i and j and between k and j have been subtracted. Defined in this way, the difference between the correlations and the partial correlations provides a measure of the *influence* of variable j on the correlation (i, k). Therefore, we define the influence of variable j on variable i, or the dependency of variable i on variable j, as $D(i, j)$, to be the sum of the influence of variable j on the correlations of variable i with all other variables. This methodology has originally been introduced for the study of financial data (Kenett *et al.* 2010, Kenett *et al.* 2012a; 2012b, Maugis 2014), and has been extended and applied to other systems, such as the immune system (Madi *et al.* 2011), and semantic networks (Kenett *et al.* 2011b).

Previous work has focused on how variable j affects variable i, by averaging over all (i, k) pairs, thus quantifying how variable j affects the average correlation of i with all other variables. While this has provided important information that has been both investigated and statistically validated, our goal here is to present a more general and robust method to statistically pick the meaningful relationships without first averaging over all pairs. Unlike the previous work, in which the average influence of j on the correlation of i with all others was calculated, and then statistically validated, here, we first filter for validated links, and then average the influence. In order to achieve this, we expand the original methodology and use statistical validation methods to filter the significant links. This statistically validated selection process reveals significant influence relationships between different financial assets. This new methodology allows us to quantify the influence of different factors (e.g. economic sectors, other markets or macroeconomic factors) have on a given asset. The information generated by this methodology is applicable to such areas as risk management, portfolio optimization and financial contagion, and is valuable to both policy-makers and practitioners.

The rest of this paper is organized as follows: In section 2, we introduce the partial correlation approach to quantify influence between financial assets. We present the new extensions of the methodology, which allows the selection of statistically significant influence links between different assets. In sections 3 and 4, we present two possible applications of the methodology. In section 3, we focus on how the methodology provides new insights into market structure and its stability across time, while in section 4, we present a practical application, which provides information on the how a company is influenced by different economic sectors, and how the sectors interact with each other. Finally, in section 5, we discuss our results and provide additional insights into further applications of this methodology.

2. Quantifying underlying relationships between financial assets

The aim of this paper is to introduce a methodology that sheds new light on the underlying relationships between financial assets. Building on previous work (Kenett *et al.* (2010)), we present a robust and statistically significant approach to extracting hidden underlying relationships in financial systems.

2.1. Data

For the analysis reported in this paper, we use daily adjusted closing stock price time series from four different markets, data provided by the Thomson Reuters Datastream. The markets investigated are the US, the UK, Japan and India, (see table 1 for details, also Kenett *et al.* (2012b)). We only consider stocks that are active from January 2000 until December 2010. Volume data was used to identify and filter illiquid stocks from the sample. Table 1 presents the number of stocks remaining after filtering out the stocks that had no price movement for more than 6% of the 2700 trading days.

2.2. Stock raw and partial correlation

To study the similarity between stock price changes, we calculate the time series of the daily log return, given by

$$r_i(t) = log[P_i(t)/P_i(t-1)], \quad (1)$$

where $P_i(t)$ is the daily adjusted closing price of stock i at day t. The stock raw correlations are calculated using the Pearson correlation coefficient (Pearson 1895)

$$\rho(i, j) = \frac{\langle r(i) - \langle r(i)\rangle\rangle \ \cdot \ \langle r(j) - \langle r(j)\rangle\rangle}{\sigma(i) \ \cdot \ \sigma(j)}, \quad (2)$$

where $\langle\rangle$ represents average over all days, and $\sigma(i)$ denotes the standard deviation.

In some cases, a strong correlation not necessarily means strong direct relation between two variables. For example, two stocks in the same market can be influenced by common macroeconomic factors or investor psychological factors. To study the direct correlation of the performance of these two stocks, we need to remove the common driving factors, which

Table 1. Summary of data sample.

Market	Stocks used	Index used	# before	# filtered
US	S&P 500	S&P 500	500	403
UK	FTSE 350	FTSE 350	356	116
Japan	Nikkei 500	Nikkei 500	500	315
India	BSE 200	BSE 100	193	126

are represented by the market index. Partial correlation quantifies the correlation between two variables, e.g. stocks returns, when conditioned on one or several other mediating variables (Baba *et al.* 2004, Shapira *et al.* 2009, Kenett *et al.* 2010). Specifically, let X, Y be two stock return time series and M be the index. The partial correlation, $\rho(X, Y : M)$, between variables X and Y conditioned on variable M is the Pearson correlation coefficient between the residuals of X and Y that are uncorrelated with M. To obtain these residuals of X and Y, they are both regressed on M. The partial correlation coefficient can be expressed in terms of the Pearson correlation coefficients as

$$\rho(X, Y : M) \equiv \frac{\rho(X, Y) - \rho(X, M)\rho(Y, M)}{\sqrt{\left[1 - \rho^2(X, M)\right]\left[1 - \rho^2(Y, M)\right]}}. \quad (3)$$

In figure 1(a), we plot the correlation vs. partial correlation (using the index as the conditioning variable) between stocks that belong to the S&P 500 index. The figure shows that all points are below the diagonal straight line, which means the influence from the index to the correlation between any pair of stocks is always positive. Furthermore, when two stocks X and Y both have business relation with a third common stock Z, their prices can be both affected by the performance of the third stock, thus showing similar price movements even after removing the effect of the index. By removing the influence from the third company, we can see the importance of the role that the third stock acts in the correlation of two stocks. The partial correlation coefficient between X and Y conditioned on both M and Z is

$$\rho(X, Y : M, Z) \equiv \frac{\rho(X, Y : M) - \rho(X, Z : M)\rho(Y, Z : M)}{\sqrt{\left[1 - \rho^2(X, Z : M)\right]\left[1 - \rho^2(Y, Z : M)\right]}}. \quad (4)$$

In order to quantify the influence of stock Z on the pair of X and Y, we focus on the *Influence* quantity

$$d(X, Y : Z) \equiv \rho(X, Y : M) - \rho(X, Y : M, Z). \quad (5)$$

This quantity is large when a significant fraction of the partial correlation $\rho(X, Y : M)$ can be explained in terms of Z. In previous research, Kenett *et al.* defined this quantity by $d^*(X, Y : Z) \equiv \rho(X, Y) - \rho(X, Y : Z)$, which holds for general cases (Kenett *et al.* 2010). However, for the stock market case specifically, the fraction of $\rho(X, Y)$ that can be explained by Z contains two parts, index influence and stocks Z influence, because stock Z can also contains information of the index. Usually, the influence from the index prevails and exceeds the influence from any other individual stock. For example, when X and Y are competitors and partners, respectively, to stock Z, performances of X and Y should have negative correlation because of Z. In this case, the influence from Z to the correlation between stocks X and Y should be negative. However, because of the dominant correlation between these two stocks and the index, the $d^*(X, Y : Z)$ is still positive. Thus, we suggest to remove the influence of the market before studying the influence of a stock Z on a pair of stocks. In the scatter plot of the partial correlation conditioned on the index vs. the partial correlation conditioned on both index and an individual stock (figure 1(b)), the points distribute at both sides of the diagonal line, meaning a significant fraction of $d(X, Y : Z)$ is negative.

The average influence $d(X : Z)$ of stock Z on the correlations between stock X and all the other stocks in the system is defined as

$$d(X : Z) \equiv \langle d(X, Y : Z)\rangle_{Y \neq X}. \quad (6)$$

It is important to note that $d(X : Z)$ approximates the net influence from stock Z to stock X, excluding the influence from the index.

2.3. *Test of statistical significance*

In a system of size N, there exists $N(N-1)(N-2)/2$ partial correlation interactions, $d(X, Y : Z)$, when all possible stock combinations are considered. To simplify the description of the system, only the non-trivial interactions with a certain significance level are selected. To identify the significance of partial correlation, we provide two methods: (1) Fisher's transformation-based approach; and (2) empirical-based approach.

2.3.1. Fisher transformation statistical significance test. We first introduce the Fisher's transformation method. According to ref. (Fisher 1915), when X and Y follow a bi-variate normal distribution and $X(t), Y(t)$ pairs to form the correlation are independent for $t = 1 \ldots n$, a transformation of the Pearson correlation

$$z(\rho) = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) = \operatorname{artanh}(\rho) \quad (7)$$

approximately follows normal distribution $\mathcal{N}(\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \frac{1}{\sqrt{N-3}})$, where r is the population correlation coefficient and N is the sample size. The Fisher transformation holds when ρ is not too large and N is not too small. Furthermore, the Fisher's z-transform of the partial correlation coefficients approximately follows (Fisher 1924)

$$\begin{cases} z\left(\rho(X, Y : M)\right) \sim \mathcal{N}\left(\frac{1}{2}\ln\left(\frac{1+r(X,Y:M)}{1-r(X,Y:M)}\right), \frac{1}{\sqrt{N-3}}\right) \\ z\left(\rho(X, Y : M, Z)\right) \sim \mathcal{N}\left(\frac{1}{2}\ln\left(\frac{1+r(X,Y:M,Z)}{1-r(X,Y:M,Z)}\right), \frac{1}{\sqrt{N-3}}\right) \end{cases} \quad (8)$$

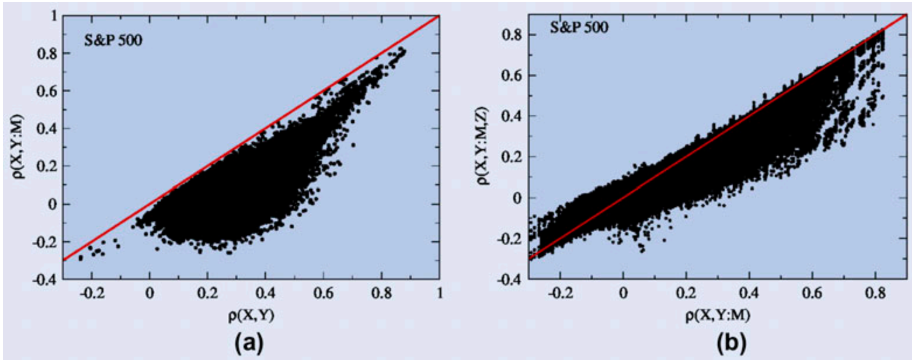

Figure 1. (a) Scatter plot of correlation vs. partial correlation conditioned on index. It is possible to observe that all points are below the diagonal straight line, which means the influence from the index to the correlation between any pair of stocks is always positive. (b) Scatter plot of partial correlation conditioned on index vs. partial correlation conditioned on both index and a third stock. It is possible to observe that the points distribute at both sides of the diagonal line, meaning a significant fraction of $d(X, Y : Z)$ is negative (mainly for low $\rho(X, Y : M)$ values). In both figures, X, Y, Z represent return time series of stocks, M represents return time series of S&P 500 index. The red curve is the diagonal straight line.

which leads to

$$z(\rho(X, Y : M)) - z(\rho(X, Y : M, Z)) \sim \mathcal{N}\left(\frac{1}{2}\ln\left(\frac{1 + r(X, Y : M)}{1 - r(X, Y : M)}\right) - \frac{1}{2}\ln\left(\frac{1 + r(X, Y : M, Z)}{1 - r(X, Y : M, Z)}\right), \sqrt{\frac{2}{N - 3}}\right). \quad (9)$$

When $\rho(X, Y : M)$ is significantly different from $\rho(X, Y : M, Z)$, $d(X, Y : Z)$ is significantly different from zero. Thus, the Student's t-test is used to determine if $\rho(X, Y : M)$ and $\rho(X, Y : M, Z)$ are different. Since N is large, the degree of freedom is also large, where the t-test is approximately the same as the Z-test (Chou 1975). In the next section, we propose a complementary empirical statistical significance test.

2.3.2. Empirical statistical significance test. Next, we introduce the empirical time series shuffling statistical significance method. While the Fisher transformation is easier to implement and faster to calculate, it requires some assumptions regarding the underlying distributions of the investigated time series.

For each of the sample time series, we perform the empirical significance test by first shuffling the sequence of the returns. The shuffling process destroys correlations between each pairs of stocks returns, and between stock returns and benchmark return, i.e. $\rho(X, Y)$ and $\rho(X, M)$ should be zero, which leads to $d(X, Y : Z)$ equal to zero. By plotting the distribution of $\hat{d}(X, Y : Z)$, we can find the thresholds of different significant levels for $d(X, Y : Z)$. In figure 2, we shuffle the return time series of 403 S&P500 stocks, and the index, and plot the distribution of $\hat{d}(X, Y : Z)$ (solid black curve). As a comparison, we plot the Gaussian distribution with same average and standard deviation value in brown colour. From the comparison, it is possible to observe that the empirical distribution of $\hat{d}(X, Y : Z)$ has fat tails, and significantly deviates from the case of a gaussian distribution observed for random or shuffled data. The dashed lines represent the positions of one-tailed 1, 5 and 10% significance levels or two-tailed 2, 10 and 20% significance levels. We suggest to make use of the two-tailed test, due to the fact that the significant negative influence is also important. The red curve in figure 2 presents the distribution of empirical $d(X, Y : Z)$. For the case of the S&P500 companies example presented in figure 2, significance level of 2% corresponds to a $z - value$ (equation 9) $z > 1.6449$; confidence level of 10% corresponds to $z > 1.2816$; and confidence level of 20% corresponds to $z > 0.8416$ (see table 2).

While the two methods are interchangeable in respect to the resulting significance levels, there are additional benefits in using the empirical approach. The empirical shuffling method conserves the distribution of returns without requiring ad-hoc assumptions on the underlying distribution of the data. Furthermore, the empirical approach can also be expanded to meet stricter requirements. For example, if the short-term time series structure is required to be conserved, the whole time series can be divided into segments with a given length and we can than shuffle the segments without changing the sequence of time series within the segments. The empirical approach requires randomizing time series, calculating N^3 partial correlation coefficients and making a distribution out of these coefficients, which for a large number of variables requires a huge amount of calculation power for real-time real-data analysis. In practice, making use of the empirical approach is preferable; however, if the empirical is not feasible, then the Fisher transformation can be used as an estimation method.

The empirical statistical significance test allows a selection of significant influence relationships between the investigated financial assets. In previous work (Kenett *et al.* (2010)), this was achieved by using different network-based approaches, which than further allowed to investigate the nature of these relationships. Below, we propose two new applications of this methodology, using the empirically statistically significant values of $d(X, Y : Z)$. The significance level used throughout the paper is the $z > 1.6449$, corresponding to two-tailed 2%.

3. Market structure and its stability

High correlation between two stocks at a given time does not necessarily guarantee high correlation in the future, because the

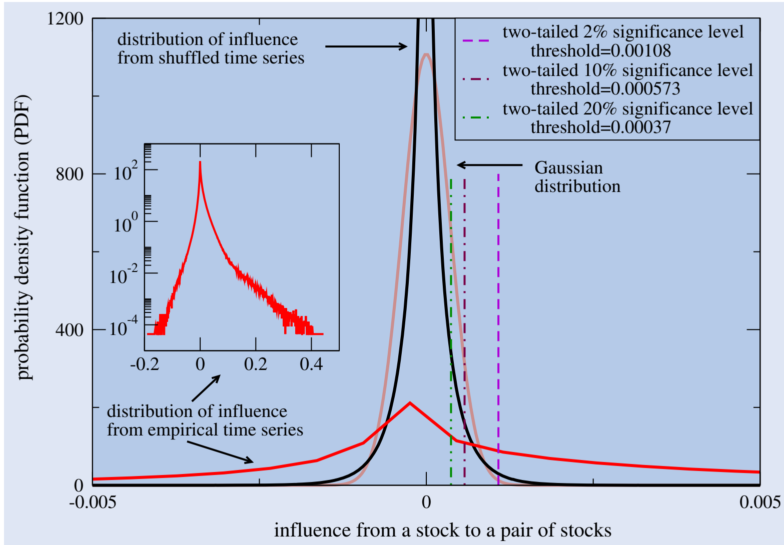

Figure 2. Empirical statistical significance test of influence between financial assets. We present the distribution of influence from one stock to pair of stocks, for the case study of 403 S&P500 companies, for a period of 11 years. The black curve represents the distribution of the influences, $\hat{d}(X, Y : Z)$, calculated from the shuffled time series. The dashed lines denotes the positions of one-tailed 1, 5 and 10% significance levels or two-tailed 2, 10 and 20% significance levels. The distribution of $\hat{d}(X, Y : Z)$ has an average value $9.9 \cdot 10^{-8}$, standard deviation $\sigma = 3.6 \cdot 10^{-4}$, skewness $9.5 \cdot 10^{-7}$ and kurtosis 9.83. The brown curve represents the Gaussian distribution with the same average value and standard deviation as the black curve. The red curve is the distribution of influences, $d(X, Y : Z)$, calculated from the empirical stock return time series. We suggest to make use of the two-tailed test, due to the fact that the significant negative influence values are also important. For the two-tailed test, significance level of 2% corresponds to $z > 1.6449$; confidence level of 10% corresponds to $z > 1.2816$; and confidence level of 20% corresponds to $z > 0.8416$. The inset shows the full empirical distribution.

Table 2. Summary of two-tail significance thresholds.

Significance level (%)	Threshold
1	0.00152
2	0.00108
5	0.00081
10	0.00057
20	0.00037

behaviour of stocks in financial markets is extremely dynamic. In certain markets, companies change their strategies faster than in the other markets, which can be uncovered by the partial correlation analysis of the behaviour of their stocks. If the companies tend to keep their past strategies, then the level of partial correlation between two companies' stocks tends to be stable. In markets where companies switch their strategies more quickly, two companies which had similar behaviour in the previous year might have quite different behaviour in the next year. In such markets, partial correlation between stocks should be more volatile.

We apply the partial correlation influence analysis to study the stability of the market structure. Specifically, we define the average influence $d(X)$ of stock X on all the other stocks in the market as

$$d(X) \equiv \langle d(X : Z) \rangle , \tag{10}$$

where $\langle\rangle$ is the average over all Z stocks. We rank the stocks by their $d(X)$ values, which we consider as a representation of the structure of the market. By dividing the 11-year period into 44 quarterly periods, we can compare similarity of the market structures (ranking of stocks) in different years. Kendall τ rank correlation coefficient (Kendall 1938) is applied to measure the similarity of the orderings for different periods. Let $(x_1(t), x_1(t')), (x_2(t), x_2(t')), \ldots, (x_n(t), x_n(t'))$ be a set of rankings of the variables X for different periods t and t', respectively. Any pair of observations $(x_i(t), x_i(t'))$ and $(x_j(t), x_j(t'))$ are said to be concordant if both $x_i(t) > x_j(t)$ and $x_i(t') > x_j(t')$ or if both $x_i(t) < x_j(t)$ and $x_i(t') < x_j(t')$. Otherwise, they are said to be discordant. The Kendall τ coefficient is defined as

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\frac{1}{2}n(n-1)}, \tag{11}$$

where if two rankings are the same, τ is one, if two rankings are independent, τ is zero, and if two rankings are discordant, τ equals minus one.

In figure 3, we present the Kendall τ coefficient for each different quarter pairs for the four investigated markets. Generally speaking, each market shows that the longer the time interval, the smaller the rank correlation coefficient, meaning lower similarity between the market structures for the two quarters which are compared. Comparing the rank correlations for the four markets, we find that S&P 500, FTSE 350 and Nikkei 500 stocks show strong market stability patterns, while Indian BES 200 stocks almost do not demonstrate any stable patterns. This can be understood by considering that developed markets tend to keep their market structure longer than fast developing markets. Furthermore, it is possible to observe that for the US market, there were structural changes in the market following the 'dot com' crisis of 2000 and the 'credit

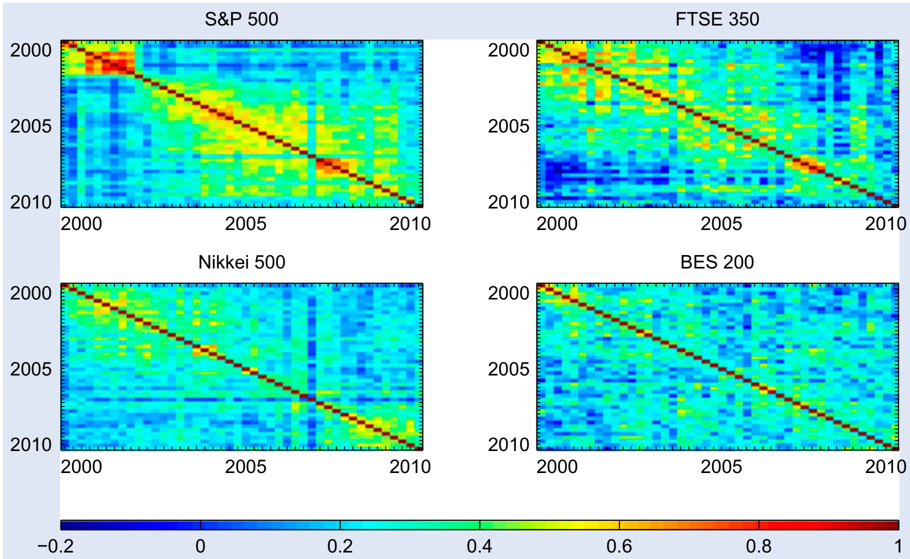

Figure 3. Kendall τ correlation coefficient of stock rankings in different quarters. The total investigated period of 11 years is divided into 44 quarters. For each quarter, we calculate the average influence of each stock, and rank the stocks according to their influence. We then use the Kendal τ correlation to quantify the similarity between every possible quarter pair, for the four investigated markets. The colour code is used to present the strength of the correlation, ranging from blue for negative rank correlation to red for positive rank correlation.

crunch' crisis of 2008. These can be identified in figure 3 by the red rectangle in the upper left corner for the former (Q4 of 2000 till Q4 of 2001), and the red rectangle in the bottom right corner for the latter (Q4 of 2007 till Q4 of 2008). These rectangles present a strong similarity in the structure during the two crises, followed by consecutive quarters with low values of three rank correlations, representing the change in structure. Studying the other markets, it is also possible to observe the structural changes resulting from the 2008 financial crisis in the UK, but not in the structure of Japan or India.

To further quantitatively study the market stability, we plot the correlation coefficient of two rankings against the time interval of these two rankings (figure 4). By averaging the correlation coefficients for each time interval, we can study how correlation coefficients decay as time evolves. We find the decay of the τ rank correlation coefficients follow an approximate exponential process, $\tau = \tau_0 e^{-t/\lambda}$, as shown in the insets in figure 4. Parameter τ_0 describes the consistency of the rankings between two consecutive quarters. The larger τ_0 is, the more consistent two consecutive ranking are. The λ parameter describes the characteristic time after which the correlation coefficient decays. Larger λ values mean longer persistence period, and thus, describe the change in influence ranking across time. These two parameters together describe the stability of the markets. For the investigated markets, we obtain the following values: US—$\tau_0 = 0.28, \lambda = 16.2$; UK—$\tau_0 = 0.22, \lambda = 19.8$; Japan—$\tau_0 = 0.2, \lambda = 18.8$; and India—$\tau_0 = 0.17, \lambda = 42.9$. As can also be observed in figure 3, τ_0 has the largest value for the S&P500 case, and smalls value for the BES200 case; however, the persistence in India is largest (as represented by the values of λ). We observe that the results for the Indian market differ from the other three markets. This is possibly related to the differences observed between developing and developed markets.

Put together, these analyses provide new insights into the dynamics of financial markets. Using the τ_0 and λ parameters, can help in monitoring structural changes in the market, and their persistence. Thus, this methodology presents a unique tool for regulators and policy-makers to monitor the stability and robustness of financial markets.

4. Quantifying the influence of economic sectors

As our society becomes more and more integrated, production activities from different industries depend upon and influence each other. Categorizing a company into only one industrial sector, cannot reflect its whole performance and associated risk. Many listed companies in the stock market belong to conglomerates, conducting their business in different industry sectors; hence, these companies' performance will naturally be influenced by multiple industries. Even if a company only conducts its business in one sector, its performance can still be influence by other sectors because of the division of labour in modern society. For example, Alcoa Inc. as the world's third largest producer of aluminium is listed in the materials sector in NYSE. However, the production of Alcoa Inc. requires dedicated supply of energy, e.g. Alcoa accounts for 15% of State of Victoria's annual electricity consumption in Australia. Thus, their performance is also heavily influenced by and contributes to the performance of the energy sector. In this section, we present an application of the partial correlation methodology to study the multiple-sector influence on stocks. We use the sector classification from the Global Industry Classification Standard (GICS).

To study the influence on a stock X from different sectors, we first calculate the influence $d(X : Z)$ (equation 6) from all other stocks Z. The analysis in this section is performed for the

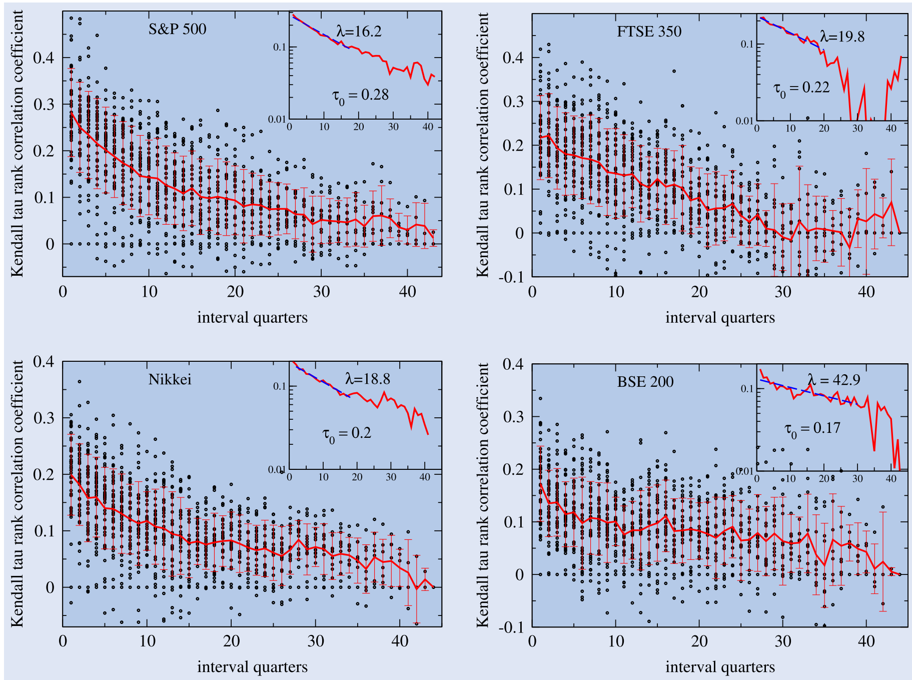

Figure 4. Kendall tau correlation coefficient between rankings vs. time interval between rankings. By averaging the correlation coefficients for each time interval, we can study how correlation coefficients decay as time evolves. We find the decay of the τ rank correlation coefficients follow an approximate exponential process, $\tau = \tau_0 e^{-t/\lambda}$. Parameter τ_0 describes the consistency of the rankings between two consecutive quarters. The larger τ_0 is, the more consistent two consecutive rankings are. The λ describes the speed that the correlation coefficient decays as time evolves. Larger λ means longer decay period and smaller decay speed. These two parameters together describe the stability of the markets. For the investigated markets, we obtain the following values: US—$\tau_0 = 0.28, \lambda = 16.2$; UK—$\tau_0 = 0.22, \lambda = 19.8$; Japan—$\tau_0 = 0.2, \lambda = 18.8$; and India—$\tau_0 = 0.17, \lambda = 42.9$.

entire investigated time period. Next, we calculate the average influence by sector, in which we use the sector categorization information of other stocks, as follows

$$d_X^S = \frac{1}{N_S} \sum_{Z_S=1}^{N_S} d(X : Z_S), \tag{12}$$

where X represents the investigated stock, S represents a given sector, N_S is the number of stocks in sector S and Z_S represents the stocks in sector S. The average influence d_X^S reflects the level of influence that stock X receives from sector S. After we normalize the average influence, we can attribute stock X's performance to sectors' performances with coefficients

$$\beta_X^S = \frac{d_X^S}{\sum_S d_X^S}. \tag{13}$$

In figure 5, we present an example of four typical stocks to show the pie picture of β_X^S. We can see from the figure that in the case of Alcoa Inc., we observe significant influence from the energy, materials and industrials sector. In the case of Franklin Templeton Investments, we find that the largest influence is from the financials sector. In the case of GE, we find that the main influence stems from the materials, utilities and financial sector. Finally, studying the example of Apple, we find that there is a more homogeneous division of the influence between the different sectors. This could possibly indicate that out of these four companies, Apple is the most diverse in its activities, being influenced almost uniformly by different sectors of the economy.

Finally, we perform a validation test on the partial correlation analysis results, investigating whether the result of multi-sector influence on stocks is plausible. To this end, we first rank all the stocks in the S&P500 data-set by their fraction of influence (β_X^S) from the financials sector. We then investigate what are the economic sectors influencing these stocks, according to the rank. We find that the top stocks in the ranking according to our partial correlation analysis are dominantly classified into the financials sector. We repeat this analysis for all other economic sectors. Indeed, all other sectors show that our analysis is in agreement with the GICS sector classification. To quantitatively show this agreement, we calculate the correct prediction rate. According to the GICS, we find the total numbers of stocks (N_S) in all sectors. From the ranking of stocks according to the influence from a given sector, we select the N_S top stocks. We then calculate the fraction of these top stocks that are classified by GICS into that certain sector as the correct prediction rate. If the partial correlation analysis prediction is in total agreement with the formal classification, then this correct prediction rate should be 1. If the prediction corresponds to the case of random picking, this correct prediction rate should be $\frac{N_S}{N}$, where N is the total number of stocks. In figure 6(c), we show that the partial correlation analysis of sector keeps a high correct prediction rate for all sectors, except the telecommunications sector, which could be related to the small number of telecommunication stocks that are part of the S&P500 index. An alternative interpretation to these results is that the financials and energy sectors are both highly

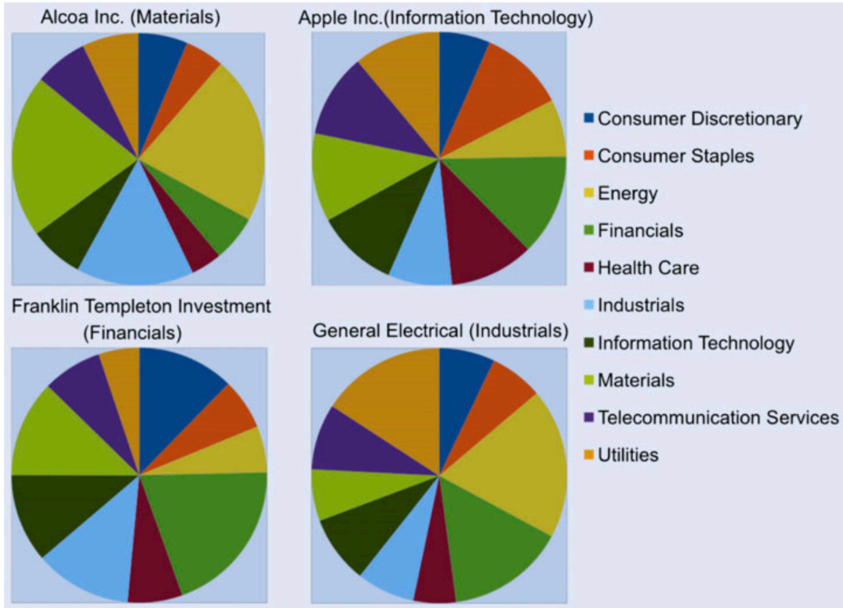

Figure 5. Fraction of influence from each sector to example stocks, Alcoa Inc., Apple Inc., Franklin Templeton Investment and General Electrical. we present an example of four typical stocks to show the pie picture of β_X^S. We can see from the figure that in the case of Alcoa Inc., we observe significant influence from the energy, materials and industrials sector. In the case of Franklin Templeton Investments, we find that the largest influence is from the financials sector. In the case of GE, we find that the main influence stems from the materials, utilities and financial sector. Finally, studying the example of Apple, we find that there is a more homogeneous division of the influence between the different sectors. This could possibly indicate that out of these four companies, Apple is the most diverse, as its business is affected by different sectors of the economy.

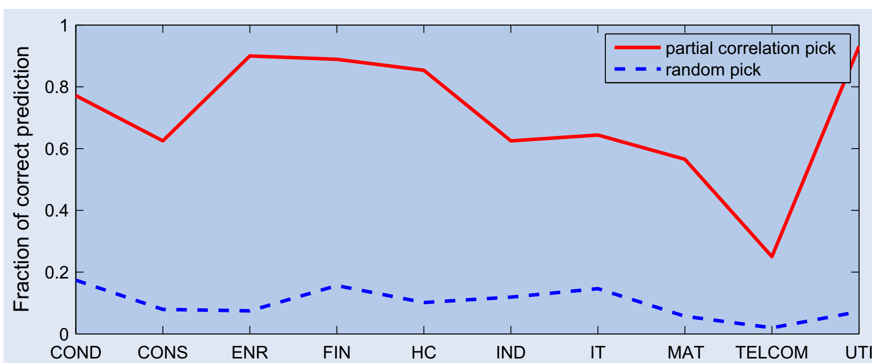

Figure 6. Partial correlation test of the extent of sectorial influence. To this end, we first rank all the stocks in S&P 500 by their fraction of influence (β_X^S). We plot the fraction of true prediction of stocks' sector from partial correlation analysis (red solid curve). The blue dashed curve is for the case of random picking strategy, for the purpose of comparison.

cohesive sectors of economic activity. This means that the activity in these economic sectors is mainly contained inside each sector, with little influence of external economic sectors. Other sectors, such as Industrials and Materials, have stronger dependencies on other sectors, and are strongly influenced other sectors' activities. Repeating this analysis for shorter periods of time, or combining the analysis with a moving window approach, could provide meaningful insights into the degree of interdependencies between the different economic sectors over time.

After studying the amount of influence that stocks receive from different sectors, we find that some sectors tend to influence the same stocks concurrently. We thus study the Pearson correlation of influences from two sectors to the same stocks, i.e. $\rho(d^{S_i}, d^{S_j})$, where d^{S_i} represents the vector variable of influence from sector i to all stocks. Applying this definition of sector correlation to the S&P 500 data results in values that are presented in the first panel of figure 7. We find that in the S&P 500 index, the pairs of industrials sector and consumer discretionary sector, materials sector and industrials sector and the communications sector and the technology sector are very close to each other, in terms of their influence. Whenever a stock is highly influenced by one of these sectors, the other in the pair also tends to be influential to this stock. We also notice some dark blue areas, e.g. the correlation between the utilities sector and the consumer discretionary sector,

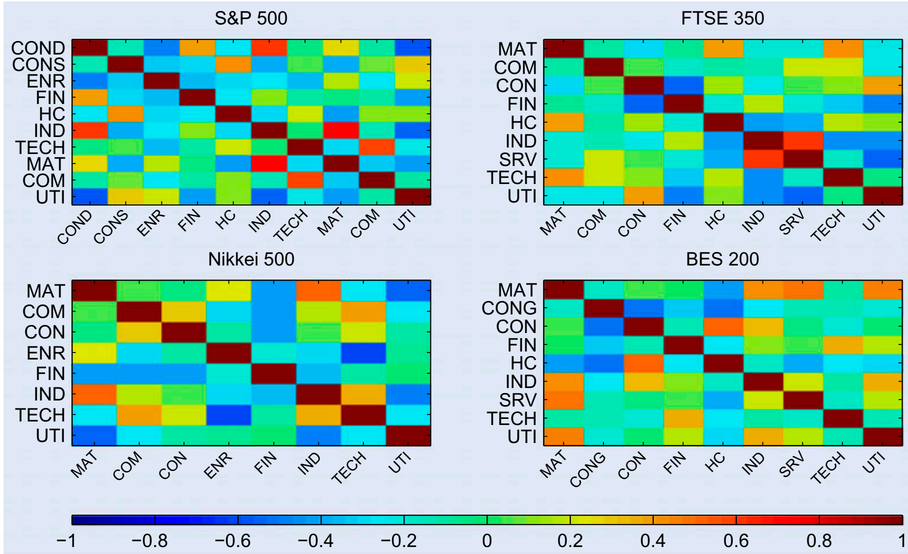

Figure 7. The correlation of influences from two sectors to the same stocks, i.e. $\rho(d^{S_i}, d^{S_j})$, where d^{S_i} represents the vector variable of influence from sector i to all stocks. We represent the correlation using a heat map of closeness between sectors in different markets, using a colour code to represent the strength of the correlation, ranging from blue for full anti correlation, to red, for complete correlation. We observe that for the S&P 500 data, the pairs of industrials sector and consumer discretionary sector, materials sector and industrials sector and the communications sector and the technology sector are very close to each other, in terms of their influence, and an anti-correlation between the utilities sector and the consumer discretionary sector. For the case of the other three investigated markets—the UK FTSE 350 index, the Japanese Nikkei 500 index and the India BES 200 index—we find that all show high correlation between materials sector and industrials sector.

which means when a stocks is highly influence by one of them, the other tends to be of little influence to the stock. We also study the UK FTSE 350 index, the Japanese Nikkei 500 index and the India BES 200 index. They all commonly show high correlation between materials sector and industrials sector.

5. Summary

This work presents a more general, statistically robust framework of the dependency network methodology introduced by Kenett *et al.* (2010). Using the dependency network methodology, we apply the partial correlation analysis to uncover dependency and influence relationship between the different companies in the investigated sample. Here, we present a new statistically robust approach to filtering the extracted influence relationships, by either using a theoretical or an empirical approach. The influence method introduced in this study is generic and scalable, making it highly accessible to both policy-makers and practitioners.

We present two possible applications of this methodology. First, we study the stability of financial market structure and show that developed markets such as the US, UK and Japan exhibit higher degree of market stability compared to developing countries such as India. Second, we show that one stock can be influenced by different sectors outside of its primary sector classification. This provides a new tool for the classification of economic sectors of activities.

While financial analysts are usually specialized in one industry sector, a broader perspective of equity research is required to grasp the insights of stock performance expectations. While the method was demonstrated using equity data, it is generic and can be applied to other asset types, and cross-asset relationships. The presented methodology provides new information on the interaction between different assets, and different economic sectors. Such information is valuable not only for investors and their practitioners, but also for regulators and policy-makers.

Acknowledgements

We wish to thank ONR (Grant N00014-09-1-0380, Grant N00014-12-1-0548), DTRA (Grant HDTRA-1-10-1- 0014, Grant HDTRA-1-09-1-0035), NSF (Grant CMMI 1125290), the European MULTIPLEX (EU-FET project 317532), CONGAS (Grant FP7-ICT-2011-8-317672), FET Open Project FOC 255987 and FOC-INCO 297149, and LINC (no. 289447 funded by the ECs Marie- Curie ITN program) projects, DFG, the Next Generation Infrastructure (Bsik), Bi-national US-Israel Science Foundation (BSF) and the Israel Science Foundation for financial support.

References

Abergel, F., Chakrabarti, B., Chakraborti, A. and Mitra, M., *Econophysics of Order-driven Markets*, 2011 (Springer: Milan).

Adrian, T. and Brunnermeier, M.K., CoVaR. Technical report, National Bureau of Economic Research, 2011.

Aste, T., Shaw, W. and Di Matteo, T., Correlation structure and dynamics in volatile markets. *New J. Phys.*, 2010, **12**, 085009.

Baba, K., Shibata, R. and Sibuya, M., Partial correlation and conditional correlation as measures of conditional independence. *Aust. NZ. J. Stat.*, 2004, **46**, 657–664.

Billio, M., Getmansky, M., Lo, A. and Pelizzon, L., Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J. Financ. Econ.*, 2012, **104**, 535–559.

Bisias, D., Flood, M., Lo, A.W. and Valavanis, S., A survey of systemic risk analytics. *Annu. Rev. Financ. Econ*., 2012, **4,** 255–296.

Bouchaud, J.P. and Potters, M., *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, 2003 (Cambridge University Press: Cambridge).

Brock, W., Hommes, C. and Wagener, F., More hedging instruments may destabilize markets. *J. Econ. Dyn. Control*, 2009, **33**, 1912–1928.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C., *The Econometrics of Financial Markets*, Vol. 1, 1997 (Princeton University Press: Princeton, NJ).

Campbell, R., Forbes, C., Koedijk, K. and Kofman, P., Increasing correlations or just fat tails? *J. Emp. Finance*, 2008, **15**, 287–309.

Chou, Y., *Statistical Analysis: With Business and Economic Applications*, 1975 (Holt, Rinehart and Winston: New York).

Cizeau, P., Potters, M. and Bouchaud, J.P., Correlation structure of extreme stock returns. *Quant. Finance*, 2001, **1**, 217–222.

Cont, R. and Bouchaud, J., Herd behavior and aggregate fluctuations in financial markets. *Macroecon. Dyn.*, 2000, **4**, 170–196.

Eisler, Z. and Kertesz, J., Scaling theory of temporal correlations and size-dependent fluctuations in the traded value of stocks. *Phys. Rev. E*, 2006, **73**, 046109.

Elton, E.J., Gruber, M.J., Brown, S.J. and Goetzmann, W.N., *Modern Portfolio Theory and Investment Analysis*, 2009 (John Wiley & Sons).

Embrechts, P. and McNeil, A., Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, edited by M.A.H. Dempster, pp. 176–223, 2002 (Cambridge University Press: Cambridge).

Engle, R., Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Stat.*, 2002, **20**, 339–350.

Fama, E.F., The behavior of stock-market prices. *J. Bus.*, 1965, **38**, 34–105.

Fisher, R., Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 1915, **10**, 507–521.

Fisher, R.A., The distribution of the partial correlation coefficient. *Metron*, 1924, **3**, 329–332.

Forbes, K.J. and Rigobon, R., No contagion, only interdependence: Measuring stock market comovements. *J. Finance*, 2002, **57**, 2223–2261.

Granger, C., Investigating causal relations by econometric models and cross-spectral methods. *Econ: J. Econ. Soc.*, 1969, **37**, 424–438.

Haldane, A.G. and May, R.M., Systemic risk in banking ecosystems. *Nature*, 2011, **469**, 351–355.

Huang, X., Vodenska, I., Havlin, S. and Stanley, H.E., Cascading failures in Bi-partite graphs: Model for systemic risk propagation. *Scientif. Rep.*, 2013, **3**, 01219.

Kendall, M.G., A new measure of rank correlation. *Biometrika*, 1938, **30**, 81–93.

Kenett, D.Y., Preis, T., Gur-Gershgoren, G. and Ben-Jacob, E., Dependency network and node influence: Application to the study of Financial Markets. *Int. J. Bifurcat. Chaos*, 2012a, **22**, 1250181.

Kenett, D.Y., Raddant, M., Zatlavi, L., Lux, T. and Ben-Jacob, E., Correlations in the global financial village. *Int. J. Modern Phys. Conf. Ser.* 2012b, **16**, 13–28.

Kenett, D.Y., Shapira, Y., Madi, A., Bransburg-Zabary, S., Gur-Gershgoren, G. and Ben-Jacob, E., Index cohesive force analysis reveals that the US market became prone to systemic collapses since 2002. *PloS one*, 2011a, **6**, e19378.

Kenett, D.Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R.N. and Ben-Jacob, E., Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLoS one*, 2010, **5**, e15032.

Kenett, Y.N., Kenett, D.Y., Ben-Jacob, E. and Faust, M., Global and local features of semantic networks: Evidence from the Hebrew Mental Lexicon. *PloS one*, 2011b, **6**, e23912.

Krishan, C.N.V., Petkova, R. and Ritchken, P., Correlation risk. *J. Emp. Finance*, 2009, **16**, 353–367.

Laloux, L., Cizeau, P., Bouchaud, J.P. and Potters, M., Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 1999, **83**, 1467–1470.

Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J.P., Random matrix theory and financial correlations. *Int. J. Theor. Appl. Finance*, 2000, **3**, 391–398.

Lo, A.W. and Craig MacKinlay, A., An econometric analysis of nonsynchronous trading. *J. Econ.* 1990, **45**, 181–211.

Lux, T. and Marchesi, M., Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 1999, **397**, 498–500.

Madi, A., Kenett, D.Y., Bransburg-Zabary, S., Merbl, Y., Quintana, F., Boccaletti, S., Tauber, A., Cohen, I. and Ben-Jacob, E., Analyses of antigen dependency networks unveil immune system reorganization between birth and adulthood. *Chaos: Interdis. J. Nonlinear Sci.*, 2011, **21**, 016109–016109.

Markowitz, H., Portfolio selection. *J. Finance*, 1952, **7**, 77–91.

Maugis, P., Event conditional correlation: Or how non-linear linear dependence can be. 2014, preprint arXiv:1401.1130.

Morck, R., Yeung, B. and Yu, W., The information content of stock markets: Why do emerging markets have synchronous stock price movements? *J. Financ. Econ.*, 2000, **58**, 215–260.

Pearson, K., Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philos. Trans. Roy. Soc. London. A*, 1895, **186**, 343–414.

Plerou, V., Gopikrishnan, P., Rosenow, B., Nunes Amaral, L. and Stanley, H.E., Universal and nonuniversal properties of cross correlations in financial time series. *Phys. Rev. Lett.*, 1999, **83**, 1471–1474.

Podobnik, B., Horvatic, D., Petersen, A. and Stanley, H., Cross-correlations between volume change and price change. *Proc. Nat. Acad. Sci. US Am.*, 2009, **106**, 22079–22084.

Pollet, J. and Wilson, M., Average correlation and stock market returns. *J. Financ. Econ.*, 2010, **96**, 364–380.

Shapira, Y., Kenett, D.Y. and Ben-Jacob, E., The index cohesive effect on stock market correlations. *Eur. Phys. J. B*, 2009, **72**, 657–669.

Sinha, S., Chatterjee, A., Chakraborti, A. and Chakrabarti, B., *Econophysics: An Introduction*, 2010 (Wiley-VCH: Berlin).

Sornette, D., *Why Stock Markets Crash: Critical Events in Complex Financial Systems*, 2004 (Princeton University Press: Princeton, NJ).

Takayasu, H., *Practical Fruits of Econophysics: Proceedings of the Third Nikkei Econophysics Symposium*, 2006 (Springer: Tokyo).

Tumminello, M., Lillo, F. and Mantegna, R.N., Correlation, hierarchies, and networks in financial markets. *J. Econ. Behav. Org.*, 2010, **75**, 40–58.

Voit, J., *The Statistical Mechanics of Financial Markets*, 2005 (Springer: Berlin).