Exploring the Network Dynamics of the Stock Markets

Wei-Fang Niu

Risk Management Institute, National University of Singapore

Henry Horng-Shing Lu

Institute of Statistics, National Chiao Tung University

January, 2011

Abstract

In this paper, a network based dynamic conditional correlation model is proposed. A market network can be easily constructed by digitizing pair wise correlations and the main idea is to introduce the transitivity of the market network. If two stocks are highly correlated to a large number of stocks, they are supposed to be highly correlated. The assertion is examined with samples from S&P 500 in the period from January 1996 to August 2009. Then we propose a network correlation model.

Under this context, the transitivity for the whole network can be also used as a collective correlation index for the stock market. We found that the index is co-integrated with the CBOE VIX. This implies its usage for risk management. Then we conduct a simulation study to reproduce the characteristics of the observed collective phenomenon using the network based dynamic correlation model. These results illustrate the role of clustering effect in financial crisis and affirm the observation that extreme financial risk is endogenous.

Keywords: Dynamic Conditional Correlation Model, CBOE VIX, Endogenous risk

1. Introduction

This paper proposes a network based approach to explore the correlation structure of the stock markets. The idea is that indirect information for pairs of stocks (their respective correlations to other stocks) may be helpful for evaluating their correlations. This enhances the class of correlation models and offers an alternative way to understand nonlinear phenomenon of the financial markets.

Correlations of asset prices have been known to be time-varying for a long time. The Dynamic Conditional Correlation (DCC) model proposed by Engle (2002) and its variants may be the most well known and commonly used in academia and industries for forecasting and filtering correlations. The estimates for correlations in this class of models are based on combinations of certain proxy quantities such as sample covariance and cross products of innovations. No indirect information enters the model explicitly.

For correlation studies, there are stylized facts less commonly recognized than volatility. In the past years, many articles have tried to address the issue of the inter-market correlation increasing during the bears market, for example Campbell et al. (2002). However, recent research by Campbell et al. (2007) attributes the phenomenon to be caused by fat tails. Becker and Schmidt (2009) investigated pairs of stocks and showed that correlations are either constant or increasing in bull markets. It should be noted that the methods in these articles are difficult to be generalized to a large number of assets.

On the other hand, a more interesting question for the stock market could be "is the market highly integrated prior to the crash or does the market crash cause the stocks to be highly integrated?" Some researchers had suggested that large market crashes could be explained by the behavior of large participants, for example Gabaix et al. (2003) and Khandani and Lo(2007). That is, position changes of large participants may trigger a large market crash if all stocks are bound together. Under this context, tight integration of the market should be observed prior to the crash.

The key to addressing these issues is to find a measure for the degree of integration among a large number of stocks. Intuitively, analyzing the correlation matrix will help in gathering certain information. However, as pointed out by Laloux et al. (1999), the correlation matrices calculated with asset price returns are mainly composed of noises. Thus dimension reduction is essential in dealing with the correlations. Araújo and Louçã (2007) proposed a procedure to denoise, determine the number of effective dimensions, and then compute the eigenvalues in the projected space as an index for market integration. They found that unusually high levels of the market structure index generally corresponded to major market crashes. That is, the market can be highly integrated prior to its major crashes.

The eigenvalue approach over the (denoised) correlation matrix provides a macroscopic view of the stock market. It indexes simultaneously the extent and amplitude of correlations for a large amount of stocks. However, it gains less insights of microscopic phenomenon - how stocks interact with each other as they get closer to a market crash.

An alternative approach is to digitize the correlations. By treating each stock as a vertex, an edge is established when the two connecting stocks are significantly correlated. As clustering effects generally exist in the financial markets, for example Bonanno et al. (2004) and Coelho et al. (2007), the network approach provides an easy and efficient way to explore the correlation structure of the market. Emmert-Streib and Dehmer (2010) then proposed to use graph edit distance to quantify structural differences between two networks. They also found that abrupt changes in the measure correspond to market crashes.

In this paper, we investigate the network dynamics under overlapping rolling windows for a sample from the S&P 500 index constituents between 1996 and 2009. As in the emerging studies of social network, the transitivity of the network plays an important role in modeling correlations. A positive relation is found between the number of common neighbors and future correlation on pairs of stocks. In addition, a market-wide correlation measure can be built on this concept. It is also found that this index is co-integrated with the CBOE VIX. This implies correlation and volatility share the same driving forces.

Similar to modeling magnetism with the Ising model, a multivariate time series model which incorporates the transitivity of the network may help explain the endogenously emerged crisis. Simulation studies show the model can reproduce the abrupt changes of the collective correlation index. This implies that the usefulness of this type of model in risk management.

The rest of the study is organized as follows. Section 2 investigates the network dynamics and proposes a multivariate time series model. Market-wide correlations are discussed in Section 3. Section 4 shows simulation results and Section 5 discusses conclusions and future extensions.

2. Market network and correlation modeling

In this study, we use daily returns of a sample from S&P 500 components. Stock prices are collected from Yahoo! Finance. These data consist of 3441 daily prices of 100 companies from January 2, 1996 to August 31, 2009. The 100 samples are randomly drawn from 389 S&P 500 components that have at least 3400 trading days during the period. All returns are adjusted to dividends and splits. Missing values of returns are filled with 0.

2.1 Modeling correlations

Since the invention of multivariate GARCH models, financial econometricians have proposed a lot of models for correlations. The BEKK model (Engle and Kroner, 1995) focuses on the modeling of covariance but also implies dynamic conditional correlations. The earliest dynamic correlation model is the Constant Conditional Correlation model proposed by Bollerslev (1990). The covariance matrix is assumed as

$$\sqrt{H_t}R\sqrt{H_t}\ ,$$

where R is the correlation matrix and H_t is a diagonal variances matrix.

Engle (2002) and Tse and Tsui (2002) proposed similar dynamic correlation models. The DCC model proposed by Engle (2002) can be formulated as below:

$$Q_t = \Psi \circ (11' - A - B) + A \circ \varepsilon_t \varepsilon_t^{'} + B \circ Q_{t-1},$$

$$R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2}\ ,$$

where R_t is the dynamic correlation matrix, ε_t is the standardized innovation, $diag(X)$ represents the matrix composed of the diagonal elements of X and $\circ$ denotes the

Hadamard product.

Cappielo et al. (2006) extended the DCC model by introducing asymmetry in the dynamics and expressed their model in a quadratic form:

$$Q_t = \left(S - A'SA - B'SB - G'VG\right) + A'\varepsilon_t\varepsilon_t^{'}A + B'Q_{t-1}B + G'\xi_t\xi_t^{'}G,$$

where $\xi_t = \min(0, \varepsilon_t)$, S is the sample covariance matrix of ε_t, and V is the sample variance of ξ_t.

McAleer et al. (2008) proposed a generalized autoregressive conditional correlation (GARCC) model which incorporates the prior models. The structural properties of the GARCC model, the analytical forms of the regularity conditions and the asymptotic results are established in their paper.

Another category of correlation models is the smooth transition models. Berben and Jansen (2005) investigated the smooth transition of the co-movement in international stock markets and considered the following form for the correlation:

$$\rho_t = \rho_0\left(1 - G\left(s_t;\gamma,c\right)\right) + \rho_1 G\left(s_t;\gamma,c\right),$$

$$G\left(s_t;\gamma,c\right) = \frac{1}{1+\exp(-\gamma(s_t - c))},$$

where ρ_0 and ρ_1 correspond to two extreme states and s_t is the transition variable, endogenous or exogenous. The multivariate version can be found in Silvennoinen and Teräsvirta (2009).

The building blocks used in the DCC model and its variations are generally restricted in sample correlation matrix, cross products of standardized innovations and their variations. The introduction of the transition variable may enable the model to incorporate more useful information, endogenous or exogenous. Here we focus on the information from the correlations themselves.

However, it should be noted that the instability of estimates for correlation coefficients can be higher than expected when the sample size is not sufficiently large. Figure 1 shows the confidence intervals for different values of estimates with correlated normally distributed samples. It is easily seen that the uncertainty changes with respect to the value of the correlation coefficient. Furthermore, the correlation coefficient can be judged to be significantly different from 0 only when the estimate deviates quite far from 0. This implies that the correlation coefficients may not be additive.

[Insert Figure 1 Here]

Therefore, the estimate of the correlation matrix can be much more unreliable because of the many entities. Specifically, as argued in Laloux et al. (1999) from the viewpoint of random matrix theory, the correlation matrices calculated with asset price returns are mainly composed of noises. Thus, while utilizing the information from a correlation matrix, (nonlinear) dimension reduction is essential and necessary.

There are some possible approaches to deal with these difficulties. Principal component or its nonlinear variations provide the most intuitive approach. By treating each time series for one single stock as a point in a high dimension, Araújo and Louçã (2007) proposed to project the points into a lower dimension space through some nonlinear algorithm. The eigenvalues for the points in the lower dimension space can be used as a measure for the integration of the market. Certainly, this idea can be incorporated into the smooth transition model.

On the other hand, Emmert-Streib and Dehmer (2010) proposed a network based approach to address correlations and detect upcoming financial crisis. They built three networks monthly by testing null hypotheses of the correlation coefficient:

(A) H_0: ρ=0 vs H_1: $\rho\neq$0;

(B) H_0: ρ=0 vs H_1: ρ<0;

(C) H_0: ρ=0 vs H_1: ρ>0.

For two networks (graphs) G and G', the graph edit distance is the minimum number of edge-additions and edge-deletions needed to transform G into G'. Denote the reference graph G_t^r as moving average of the graphs up to *t*,

$$G_t^r = \frac{1}{t}\sum_{s=1}^{t} G_s \ .$$

The two authors calculated the graph edit distance between G_t and G_t^r, and found that the peaks of the time series coincided with the recognized stock market crashes.

2.2 Correlations and network topology

Before proceeding to a correlation model, this section presents some observations with the S&P 500 sample. First, daily logarithmic returns for all stocks are calculated. As in Riskmetrics, the covariance matrix up to day *t* is computed by the exponentially weighted moving average method with the decay factor λ=0.94,

$$S_t = \left(\sum_{s=1}^{t} \lambda^{t-s} r_s r_s^{'}\right) \Big/ \left(\sum_{s=1}^{t} \lambda^{t-s}\right). \tag{1}$$

And the correlation matrix is obtained as

$$R_t = \left(diag(S_t)\right)^{-1/2} S_t \left(diag(S_t)\right)^{-1/2} . \tag{2}$$

Figure 2 shows the distribution of correlation coefficients calculated up to the dates January 3 2005, July 15 2007 and October 1, 2008. Looking back from 2011, the first date can be simply viewed as an ordinary day, the second corresponds to the summit of the 2004~2008 cycle, while on the third date the market was in turbulence after Lehman Brothers failed.

[Insert Figure 2 Here]

Even though these correlation coefficients are dependent, their distribution still reveals some useful information. The density function is smoothed with the kernel method. It is easily seen that the second distribution (on July 15 2007) is notably skewed to the left compared to the first (on January 2 2005), but their tails do not differ significantly. And the third distribution (on October 1 2008) is shifted in whole to the right compared to the others.

These observations imply that at least two regimes exist for the correlation structure. One corresponds to the bear market in which most assets are highly correlated. During “normal” days when the market “seems ” to randomly walks up and down, some lightly or moderately correlated pairs may be enhanced while some specific pairs remain at their originally high level.

With this observation and considerations for the instability of estimating correlations, digitization of correlations suggested in Emmert-Streib and Dehmer (2000) would be a reasonable approach if the interest is simply to model the correlation prior to the crashes. Under this context, analysis of a large amount of asset could be treated with the tools from network theory that is used in many areas in recent years.

A simple rule of thumb that has been widely applied in many e-businesses, for example Facebook and amazon.com, could be like this: for two subjects simultaneously related with someone else, they will have higher probability to be related to each other. An analog for the stock market network would be that two stocks highly correlated to one stock tend to be highly correlated. Such arguments, though intuitively reasonable but frequently ignored, can be useful for the modeling of correlations.

To examine this idea, a network is built based on digitizing correlations with a preset threshold ρ_0. This is also equivalent to testing the hypothesis H_0: $\rho_{ij,t}>0$. A stock k is said to be the common neighbor of stock i and j at time t if $\min(\rho_{ik,t}, \rho_{jk,t})>\rho_0$.

For the S&P 500 sample, the correlation matrix is estimated day by day with the EWMA approach (1) and (2), and then the correlation matrix is converted to a binary matrix element by element with a threshold value ρ_0=0.3. In this way, the market network on all stocks is established and the number of common neighbors for each pairs of stocks can be easily computed.

Figure 3 displays the scatter plots for the correlations calculated with returns in next coming 15 and 60 trading days and the number of common neighbors so that the data used for correlations and the number of common neighbors do not overlap. A positive relation between the two variables seems to exist. For further investigations, we perform ordinary linear regressions for every 5 trading days. The time series of t statistics for testing the slope is shown in Figure 4. The result suggests the short term impact for correlation incorporating the number of common neighbors is just flickering around the significance level. But it is significant for medium term forecasts of correlations. This also implies that the nonlinear term does not dominate the short term dynamics of correlations.

[Insert Figure 3 Here]

[Insert Figure 4 Here]

In addition, the number of common neighbors also effectively addresses the clustering effect in the stock market mentioned in Bonanno et al. (2004) and Coelho et al. (2007). To some extent, clustering can be seen as an externally top-down process that searches for the best segmentation of the stock market. However, it is generally hard and unrealistic to divide the market into non-overlapped blocks since the relationships among all stocks are very complex. We can ask the question inversely: what kind of cluster would a pair of stocks fit in. The enumeration on common neighbors actually provides the first order approximation to the concept.

2.3 A network based correlation model

Typical components in the correlation models consist of the time varying correlation matrix, a long term target matrix, and the cross products of unit innovations. Generally the long term target assumes stationarity on correlation but introduces a large amount of parameters to be estimated with the observations above. This term can be obtained through the number of the common neighbors. Then the model would be more parsimonious and can be considered as composed of direct and indirect information for correlation of pairs of stocks.

Thus a network based correlation model can be specified as

$$r_t = \sqrt{H_t} R_t \varepsilon_t , \tag{3}$$

$$Q_t = (1-\alpha_1-\beta_1)\Psi_{t-1} + \alpha_1 \varepsilon_{t-1}\varepsilon_{t-1}^{'} + \beta_1 Q_{t-1} , \tag{4}$$

$$0 < \alpha_1, \beta_1 < 1 ,\ \ 0 < \alpha_1 + \beta_1 < 1 ,$$

$$R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} , \tag{5}$$

$$\Psi_t = h\left(\alpha_0 11' + \beta_0 \Gamma_t \Gamma_t^{'}\right) - diag\left(h\left(\alpha_0 11' + \beta_0 \Gamma_t \Gamma_t^{'}\right)\right) + I ,\ \ \alpha_0, \beta_0 > 0 , \tag{6}$$

$$h(x) = \left(e^x - 1\right) / \left(e^x + 1\right),$$

where r_t is the return vector, H_t is the diagonal matrix composed of variances for each stock, and Γ_t is the matrix for common neighbors. Note that Γ_t can be obtained from Q_t by digitization or other sources.

The following proposition addresses the requirements for the correlation matrix to be positive definite.

***Proposition* 2.1**. The correlation matrix Q_t defined in (2)-(4) is positive definite.

Proof. First note that $\alpha_0 11' + \beta_0 \Gamma_t \Gamma_t^{'}$ is positive definite as α_0, β_0>0 so $h\left(\alpha_0 11' + \beta_0 \Gamma_t \Gamma_t^{'}\right)$ is also positive definite since $h(x)$ is increasing and $h(0)=0$.

As the diagonal elements of $h\left(\alpha_0 11' + \beta_0 \Gamma_t \Gamma_t^{'}\right)$ are always less than 1 so $I - diag\left(h\left(\alpha_0 11' + \beta_0 \Gamma_t \Gamma_t^{'}\right)\right)$ is also positive definite. As shown by Ding and Engle (2001), Q_t is also positive definite as $0<\alpha_1, \beta_1<1$.■

With the assumption of normality for ε_t, the estimation of the network correlation model can be conducted by a two stage strategy. The log-likelihood can be decomposed into a volatility part and a correlation part:

$$l(\theta,\phi;Y) = l_v(\theta;r) + l_c(\theta,\phi;z) ,$$

$$l_v(\theta;r) = -\frac{1}{2}\sum_{t=1}^{T}\left(N \log 2\pi + \log\left|H_t\right| + r_t^{'} H_t^{-1} r_t\right),$$

$$l_c(\theta,\phi;z) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left|R_t\right| + z_t^{'} R_t^{-1} z_t - z_t^{'} z_t\right), \tag{7}$$

where $\phi = \left(\alpha_0, \beta_0, \alpha_1, \beta_1\right)$ and θ is the parameter vector for the volatility part.

Two issues about estimation should be noted. First, the dimension is large and the inversion of the correlation matrix can be difficult to obtain. Alternative approach such as composite marginal likelihood in Engle et al. (2008) can be an option. Second, recall that in Figure 4 the test statistics for β_0 appear to be more significant with future correlations calculated from longer periods. This implies that estimation for β_0 based on short term information, say likelihood estimation like (7), may lead to biases. Empirical results shall be shown in Section 4.

The slope β_0 in (6) plays an important role for the growth of the network. Positive β_0 corresponds to the property that the clusters of stocks will tend to enhance themselves. Thus the evolution of the correlations under the model is self-organized. Combined with the observations in the subsequent section, this implies that the risk can be endogenous.

Assuming a positive value of α_0 implicitly leads to the tendency toward positive values for correlations of all pairs of stocks. Theoretically this can be an intuitive consequence of CAPM. As all stocks are related to the market portfolio, it is reasonable to expect that they tend to positively correlated to each other. Empirically, the distributions in Figure 2 actually indicate that correlations for most pairs of stocks are positive.

Still the model can be also rearranged in the form:

$$Q_t - Q_{t-1} = (1-\alpha_1-\beta_1)\left(\Psi_{t-1} - Q_{t-1}\right) + \alpha_1\left(\varepsilon_{t-1}\varepsilon_{t-1}' - Q_{t-1}\right).$$

The form looks as if Ψ_t plays the role of the target of mean reversion for Q_t. However, as Ψ_t changes over time, the process does not necessarily appear to be mean reverting. In fact, the stationarity or ergodicity of this type of models would require further explorations. A simple property can be shown in the proposition below.

Proposition 2.2. When the number of assets approaches infinity, the condition $Q_t \sim 11'$ and $\Gamma_t = 11'$ is an absorbing state.

At a first glance, the proposition seems to pose some restrictions to the usage of the model. It implies that that the model may behave irrationally when all pairs of stocks are extremely highly correlated. However, such situation is generally seen in the severe bear market. At that moment, it is irrational to expect that the market is in any kind of equilibrium or stationarity. Thus, the question is just when to stop using this model. Thus a measure for the collective correlation of the market is necessary.

3. Collective behavior of the market

In this section a measure for the collective correlation of the whole market is proposed and investigated. This measure is then compared with the CBOE VIX, which is commonly recognized as a forward looking risk metric. As the two time series is found to be co-integrated, volatility and correlation may share the same driving forces. This measure is also used in judging the performance of the correlation models.

3.1 Measuring the collective correlation

As shown earlier, there are always some highly correlated pairs of stocks while others change over time. In addition to counting the number of highly correlated pairs or calculating the eigenvalues of the correlation matrix, a more elaborate method is to measure collective correlations of the market and investigate how many candidates of pairs are actually significantly correlated.

However, which pair of stocks can be considered to be a candidate? Certainly, knowledge such as sectors may reveal some clues. But such static information cannot be sufficient for describing a dynamic system. The rule of thumb described in Section 2.2 can be an alternative choice. After digitizing all the correlations, the whole market can be seen as a network and the transitivity of the network actually provides a good measure for this concept.

The transitivity of a network is defined below.

$$C_{all} = \frac{3 \cdot \text{number of triangles}}{\text{number of connected triples}}$$

Connected triples correspond to those pairs of stocks that are simultaneously highly correlated to another stock, and triangles mean triples of stocks highly correlated to one another. While calculated over the whole network, the denominator exactly represents how many pairs of stocks are supposed to be the candidates that are significantly correlated according to the time series data obtained. The numerator just reflects the realized number of candidates.

It is noted that transitivity not only measures the level of correlation but also reflects the clustering effect. The transitivity would be higher than the proportion of connected edges if the clustering effect actually exists. In other words, a market is highly integrated when the clustering effect is strengthened.

Some more descriptions about the properties of the random network are shown in the Appendix.

3.2 Correlation and volatility

The same procedure in Section 2.2 for the estimation of the covariance and correlation matrix can be used for the construction of the market network and the transitivity of the network can be easily calculated. Figure 5 shows the trend chart for the collective correlation constructed at different thresholds 0.25, 0.3 and 0.35. As the three time series do not differ significantly, setting the threshold at 0.3 seems to be a reasonable and robust choice.

[Insert Figure 5 Here]

The most obvious feature of the collective correlation index may be the frequently seen jumps. This generally corresponds to correlations increasing during the market plunges. A coupling-decoupling process in a business cycle can be easily observed along the trend of the index. The average levels of the collective correlation index subsequent to the NASDAQ bubble seem to be higher than that prior to the bubble. This phenomenon probably indicates changes in trading behaviors.

For further investigation of the property of the index, CBOE VIX serves as a good reference. This nonparametric volatility index has been viewed as one of the most important indicator which reflects the participants' expected market condition in the future.

Figure 6 shows the trend chart for the collective correlation index and the CBOE VIX. Generally speaking, the two time series have similar trends and, most importantly, they have very similar timing and patterns for jumps. Table 1 lists the testing result of the Johansen procedure on co-integration. The significant result for the co-integration test of the two series shows that they could be driven mainly by the same risk factors.

[Insert Figure 6 Here]

[Insert Table 1 Here]

In addition, in the middle of 2007 both the two index had rose from their bottoms before the subprime crisis. This is essentially compliant with the results by Becker and Schimdt (2009).

4. Application to risk management

This section demonstrates the usage of the proposed network correlation model through simulation studies. Since high collective correlation coincides with the possible forthcoming market turbulence and thus the restriction of the model, the most natural way is to simulate the collective correlation with the model until it exceeds some threshold. The distribution of the first passage time can be obtained through the simulations and used for planning of hedges and allocations. The simulations proceed as follows.

First the time series of unit innovations for each stock is filtered by a simple GARCH process, in which parameters are estimated every year with returns in the past two years. Then the estimates for parameters in the correlation model are obtained by the composite marginal likelihood constructed by contiguous pairs. Table 2 lists the parameter estimates for the samples from several periods.

[Insert Table 2 Here]

The estimation results are basically compliant with the previous results, for example Engle et al. (2009), regarding high levels of auto-regression coefficients. For the very strong auto-regression effect, the short term forecast under the two types of models will not deviate from each other very significantly. However, as time elapses, the transitivity effect will start to dominate the correlation matrix and cause an abrupt rise. It is noted that, except the periods 2000-2001, the estimates for β_0 are generally below 0.01. This implies that a pairs of stocks tend to be significantly correlated if they have more than 20 common neighbors.

For further investigations on the correlation models, we then conduct more simulation experiments. Figure 7 demonstrates several typical paths of the collective correlation generated by the DCC model and the network correlation model with different values of β_0. The initial condition for the correlation matrix is set as on January 2, 2007. It is clear that the DCC model leads to a constant level for the collective correlation index. However, as β_0 is small as 0.01, similar behaviors can be observed for the network correlation model. Only with larger values of β_0, say 0.02 and 0.04, the collective correlation index reproduced with the network correlation model appear to have an upward trend or an abrupt increase. Note that the paths seem to be less stochastic while β_0=0.04. This implies that only some appropriate values of β_0 correspond to the observed collective correlations.

[Insert Figure 7 Here]

This feature of the network correlation model can be applied in risk management. The intuition is that higher collective correlations generally correspond to higher volatilities and thus higher probabilities for future turbulence. Thus by evaluating the distribution of the first passage times over a prescribed threshold may help portfolio managers to arrive at their allocation and hedge strategies.

Figure 8 shows the distributions of the first passage times over the threshold 0.7 for collective correlations with initial conditions set on January 2, 2007. The values of β_0 are set as 0.025, 0.03, 0.035, 0.04 and 0.045. The distribution corresponding to β_0=0.03 (solid line) is to the left of other distributions, including the one with β_0=0.025 (dashed). Again this means that the choice can be a little restrictive and a proper level may be roughly 0.03, which is in fact quite larger than estimated by maximum likelihood. However, all these first passage time distributions point to the assertion that the collective correlation index may rise over the threshold 0.7 after about 100 trading days on average.

[Insert Figure 8 Here]

5. Conclusions

This study proposes to include indirect information for the modeling of the correlation of a pair of stocks. The idea about indirect information is implemented by digitizing the correlations and counting the number of common neighbors. With a sample from S&P 500, a positive relation between the number of common neighbors and future correlation is found. We propose a network correlation to extend the DCC model.

The same idea is also used for the construction of a collective correlation index. The index is coherent with the CBOE VIX in terms of dynamics. This also suggests that a multivariate time series model, especially for a large number of stocks, can be used.

As for the technical issues, it may not be necessary to digitize all the correlations but adding some kind of weights with respective to different values of correlation coefficients is essentially important in consideration of the instability of correlation estimates. Within the network setting, some possibilities are worth further exploring. For example, constructing the network by checking the existence of the tail dependence for a pair of stocks can be an alternative approach.

The model in this study incorporates ideas from statistical mechanics so as to produce nonlinear and even critical phenomenon similar to phase transition. It also illustrates how extreme risks may evolve through interactions among a large number of assets. In other words, it would be a rational conjecture that certain extreme financial risks could be endogenous, although the true mechanism is still not well understood.

The model and the concept proposed in this study should not be simply viewed as an alternative dynamic conditional correlation model. It aims to address the issue of understanding the collective dynamics of a large number of stocks. In this context, it is also an alternative approach to the systematic risk addressed in Brownlees and Engle (2010) and Acharya et al. (2010), which has become more and more important since the subprime crisis.

Appendix. Random network

A network or a graph *G* is composed of an ordered pair (*V*, *E*), where *V* is a set of vertices and *E* is a set of edges that join pairs of vertices. A graph can be either directed or undirected. In this paper, only undirected graphs are considered.

In practical applications, each vertex represents a subject or a random variable and the edge reveal certain dependency relationships between each pair of vertices. The network approach thus provides an easy way to visualize the interrelation structure of a large-scale complex system.

Two vertices are neighbors if they share a common edge. The degree of a vertex is the number of edges it has. The set of vertices that a vertex can reach through edges is the component it belongs to. A geodesic path for a pair of vertices is the shortest path between them. The diameter of a graph is its longest geodesic path.

An interesting property for a network is transitivity or clustering. Consider a network and three vertices *a*, *b* and *c*. If *a* and *b* are connected and *a* and *c* are connected, then it is well expected that *b* and *c* are connected to each other. So a clustering coefficient can be

$$C_{all} = \frac{3 \cdot \text{number of triangles}}{\text{number of connected triples}},$$

where the connected triple is the combination of three vertices and at least two edges among them.

[Insert Figure A1 Here]

For a specific vertex *k*, the coefficient can be also defined as

$$C_k = \frac{\text{number of triangles connected to vertex } k}{\text{number of connected triples centerd on vertex } k}.$$

A random network has edges that are randomly produced. The simplest random network is the Erdős–Rényi model, in which the presence of an edge is independent of the presence of any other edge with a constant probability *p*. Thus, the probability for any vertex to have a degree *k* is

$$P(X = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{z^k e^{-z}}{k!},$$

where $z=(N\text{-}1)p$ and the approximation relies on large *N*.

Studies for random networks with arbitrary degree distributions can be found in Newman et al. (2001), and a short review for properties of graphs can found in Newman (2003).

A famous example for a network in the real world is the small world model proposed by Watts and Strogatz (1998) and Watts (1999). The model can be built from a regular lattice by randomly rewiring a fraction of the edges. As shown by Watts and Strogatz through simulations, the rewired network has a low average distance between vertices and high transitivity.

Reference

V. Acharya, , L. H. Pedersen, T. Philippon, and M. P. Richardson, 2010. Measuring systemic risk. FRB of Cleveland Working Paper No. 10-02.

T. Araúo and F. Loucã, 2007. The geometry of crashes. A measure of the dynamics of stock market crises. *Quantitative Finance* **7**, 63-74.

C. Becker and W. M. Schimdt, 2009, State-dependent dependencies: A continuous-time dynamics for correlations.

R. Berben and W. J. Jansen, 2005, Comovement in international equity markets: A sectoral view. *Journal of International Money and Finance* **24**, 832-857.

T. Bollerslev, 1990. Modelling the coherence in short-run nominal exchangerates: a multivariate generalized ARCH approach. *Review of Economic Studies* 72, 498–505.

G. Bonanno, G. Caldarelli, F. Lillo, S. Miccich`e, N. Vandewalle, and R.N. Mantegna, 2004. Networks of equities in financial markets. *European Physical Journal B* 38, 363-371.

C. T. Brownlees and R. Engle, 2010. Volatility, correlation and tails for systemic risk measurement. SSRN working paper.

R. A. Campbell, K. Koedijk and P. Kofman, 2002. Increased correlation in bear markets. *Financial Analyst* **58**, 87-94.

R. A. Campbell, C. Forbesc, K. Koedijk and P. Kofman, 2007. Increasing correlations or just fat tails? *Journal of Empirical Finance* **15**, 287-309.

L. Cappiello, R. Engle and K. Shepard, 2006. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* **4**, 537-572.

Z. Ding and R. Engle, 2001. Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers* **29**, 157–184.

F. Emmert-Streib and M. Dehmer, 2010, Identifying critical financial networks of the DJIA: Toward a network-based index. *Complex*, **16**, 24-33.

R. Engle, K.F. Kroner, 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.

R. Engle, and A. Patton. 2001, What good is a volatility model? *Quantitative Finance* **1**, 2, 237-245.

R. Engle, 2002. Dynamic Conditional Correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* **20**, 339-350.

R. Engle and B. Kelly, 2009, Dynamic equicorrelation. *NYU working paper* No. FIN-08-038.

R. Engle, N. Shephard and K. Shephard, 2008, Fitting vast dimensional time-varying covariance models. *NYU working paper* No. FIN-08-009.

X. Gabaix, P. Gopikrishnan, , V. Plerou, and H. E. Stanley, 2003. A theory of power-law distributions in financial market fluctuations, *Nature* **423**, 267-270.

P. J. F. Groenen and P. H. Franses, 2000. Visualizing time-varying correlations across stock markets. *Journal of Empirical Finance,* 7, 155–172.

A. Johansen, O. Ledoit, and D. Sornette, 2000. Crashes as critical point. *International Journal of Theoretical and Applied Finance*, **3**, 219-255.

A. E. Khandani and A. W. Lo, 2011. What happened to the quants in August 2007? Evidence from factors and transactions data. *Journal of Financial Markets* **14**, 1–46.

M. McAleer, F. Chan, S. Hoti and O. Lieberman, 2008. Generalized autoregressive conditional correlation. *Econometric Theory* 24, 1554-1583.

M. E. J. Newman, S. H. Strogatz and D. J. Watts, 2001. Random graph with arbitrary distributions and their applications. *Physical Review E* 64, 026118.

M. E. J. Newman, 2003. The structure and function of complex networks. *SIAM review* 45, 167-256.

T. Okimoto, 2009. New Evidence of Asymmetric Dependence Structures in International Equity Markets. *Journal of Financial and Quantitative Analysis* **43**, 787-815.

L. Ramchand and R. Susmel, 1998. Volatility and cross correlation across major stock markets. *Journal of Empirical Finance* **5**, 397-416.

RiskMetrics^{TM}, Technical Document, 4th ed., 1996 (J.P. Morgan).

A. Silvennoinen and T. Teräsvirta, 2009. Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH Model. *Journal of Financial Econometrics* **7**, 373-411.

Y. K. Tse and A. K. C. Tsui, 2002. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics* **20**, 351-362.

M. Tumminello, T. Di Matteo, T. Aste, and R.N. Mantegna, 2007. Correlation based networks of equity returns sampled at different time horizons. *European Physical Journal B* 55, 209-217.

Table 1. Co-integration test by the Johansen Procedure for the collective correlation index and the CBOE index.

	Test statistic	10%	5%	1%
r<=1	9.53	6.50	8.18	11.65
r=0	19.61	12.91	14.90	19.19

Table 2. Parameter estimates of the network correlation model with data from several periods. Numbers in the brackets are asymptotic standard deviations.

Data period	α_0	β_0	α_1	β_1
1999-2000	0.1792 (0.0084)	0.0063 (1.4697)	0.9798 (0.0276)	0.0072 (0.0003)
2000-2001	0.3186 (0.0019)	0.0346 (0.8196)	0.9730 (0.0043)	0.0126 (0.0001)
2001-2002	0.7355 (0.0354)	0.0029 (0.0001)	0.9854 (0.0299)	0.0109 (0.0005)
2002-2003	0.6126 (0.0522)	0.0070 (0.0009)	0.9855 (0.0257)	0.0099 (0.0005)
2003-2004	0.4796 (0.0089)	0.0044 (0.0001)	0.9801 (0.0174)	0.0094 (0.0003)
2004-2005	0.4892 (0.0308)	0.0049 (0.0006)	0.9851 (0.0154)	0.0060 (0.0003)
2005-2006	0.4341 (0.0122)	0.0061 (0.0004)	0.9789 (0.0170)	0.0069 (0.0005)
2006-2007	0.5369 (0.0750)	0.0041 (0.0002)	0.9855 (0.0142)	0.0066 (0.0013)

Figure 1. Confidence interval for different sample correlation coefficients. The dash line and dotted line correspond to 90% and 95% of confidence levels respectively. The sample size is set as 25.

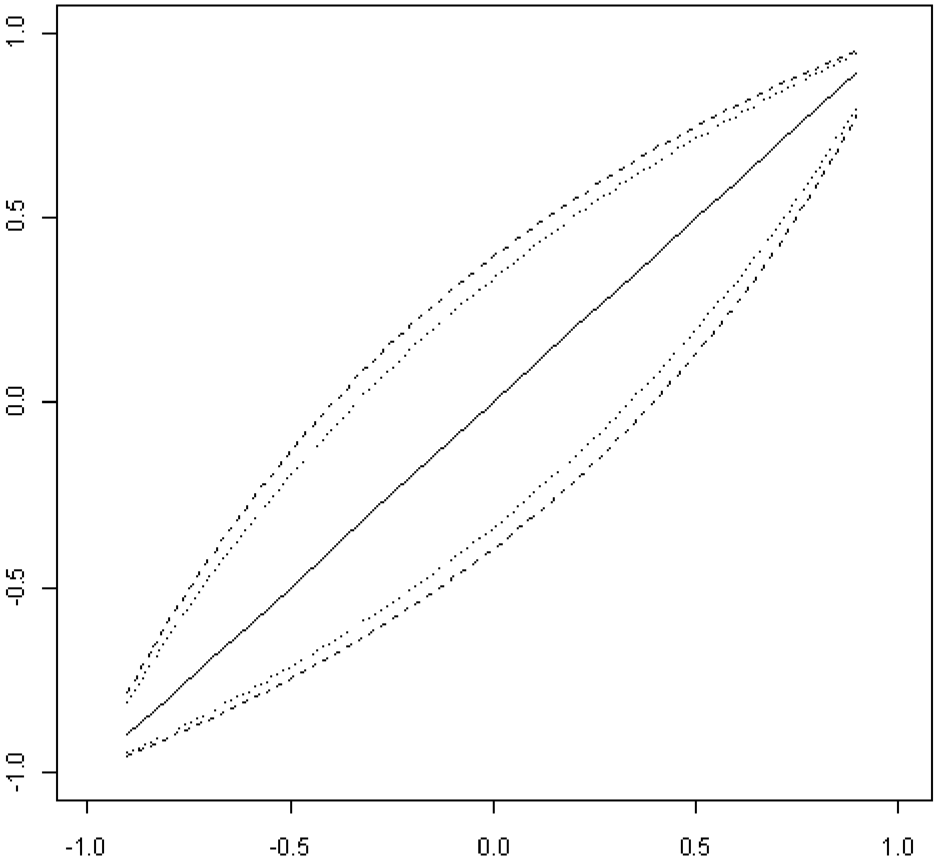


Figure 2. Distribution of correlation coefficients in different market conditions. The first date (January 3, 2005) can be viewed as an ordinary day, the second (July 16, 2007) corresponds to the summit of the 2004~2008 cycle, while on the third date (October 1, 2008) the market was in extreme turbulence after Lehman Brothers failed.

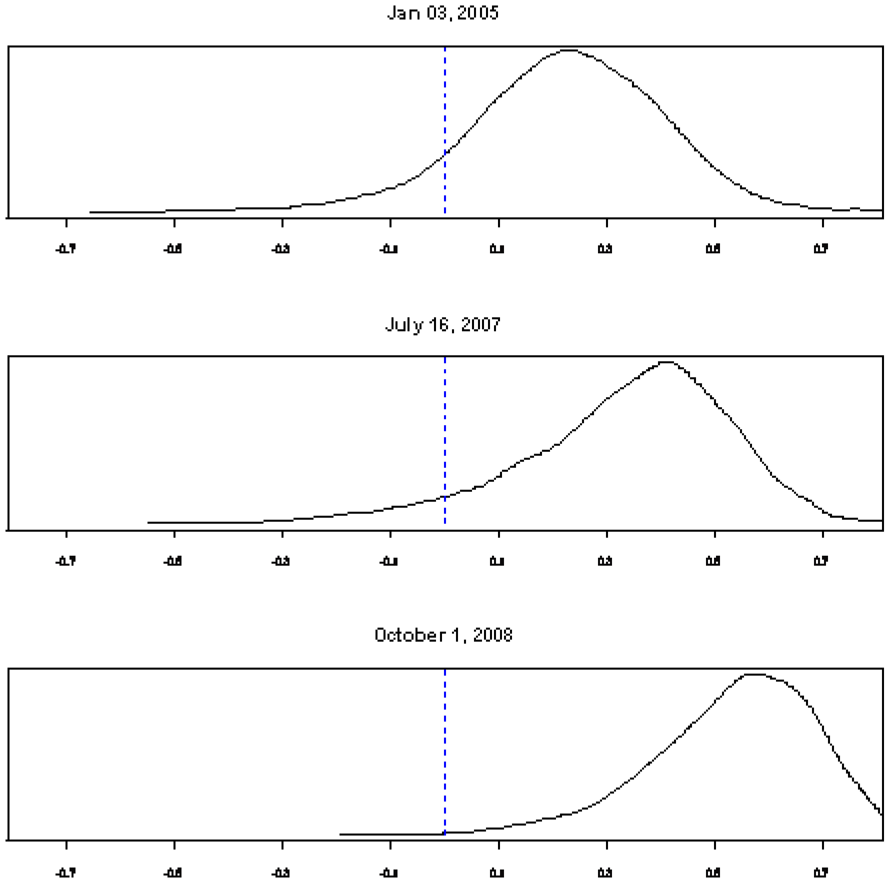


Figure 3. Future correlations versus the number of common neighbors. Left column corresponds to correlations in next 15 days and right column to correlations in 60 days. The correlation coefficients are transformed to $(-\infty,\infty)$ with $h(x)=\log\left((1+x)/(1-x)\right)$.

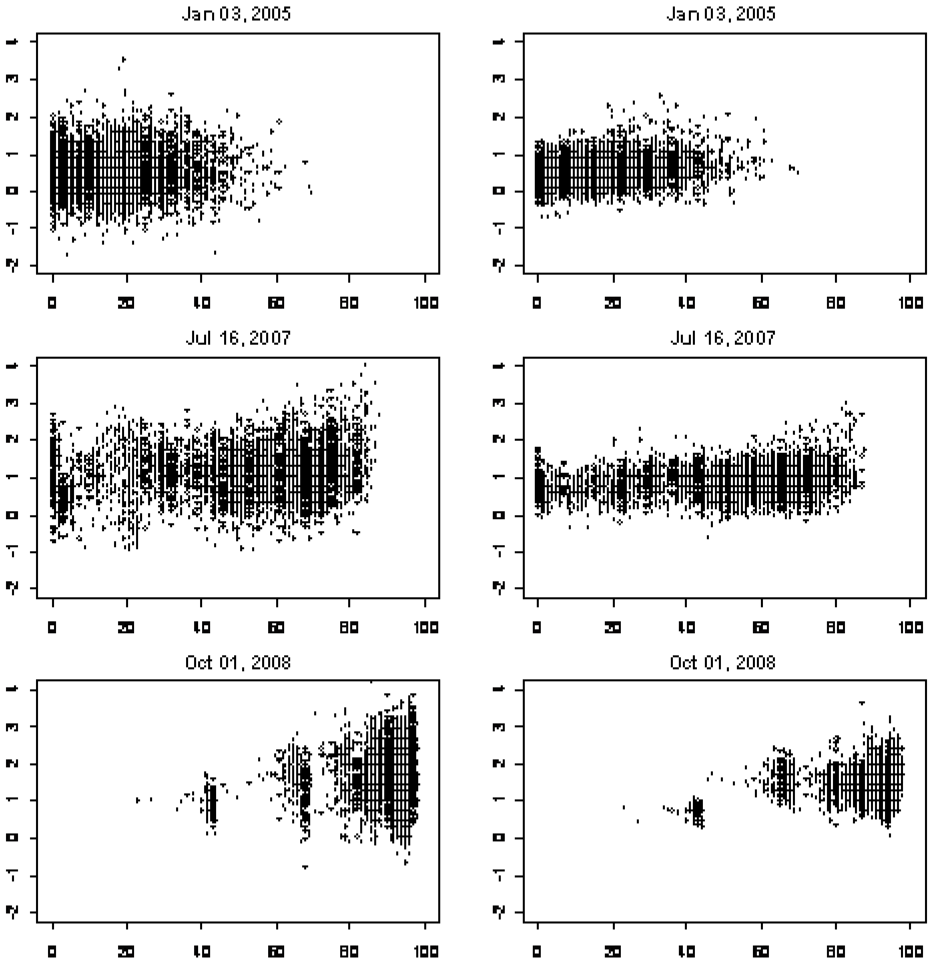
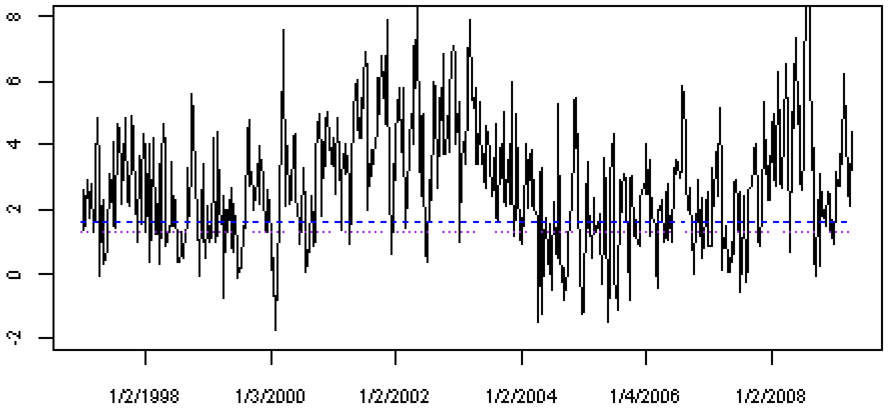


Figure 4. Significance of the slope. The correlation coefficients are transformed to $(-\infty, \infty)$ with $h(x) = \log((1+x)/(1-x))$ and then regressed with the number of the common neighbors every 5 days.



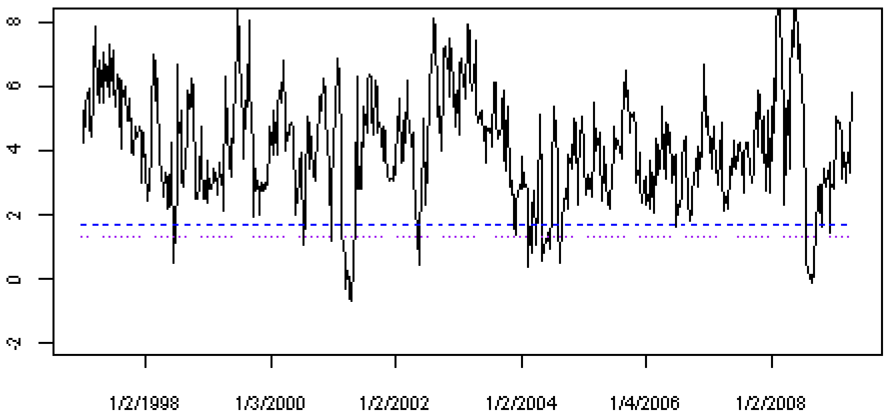


Figure 5. Collective correlation index at different threshold: 0.35 (dotted), 0.3 (solid) and 0.25 (dashed).

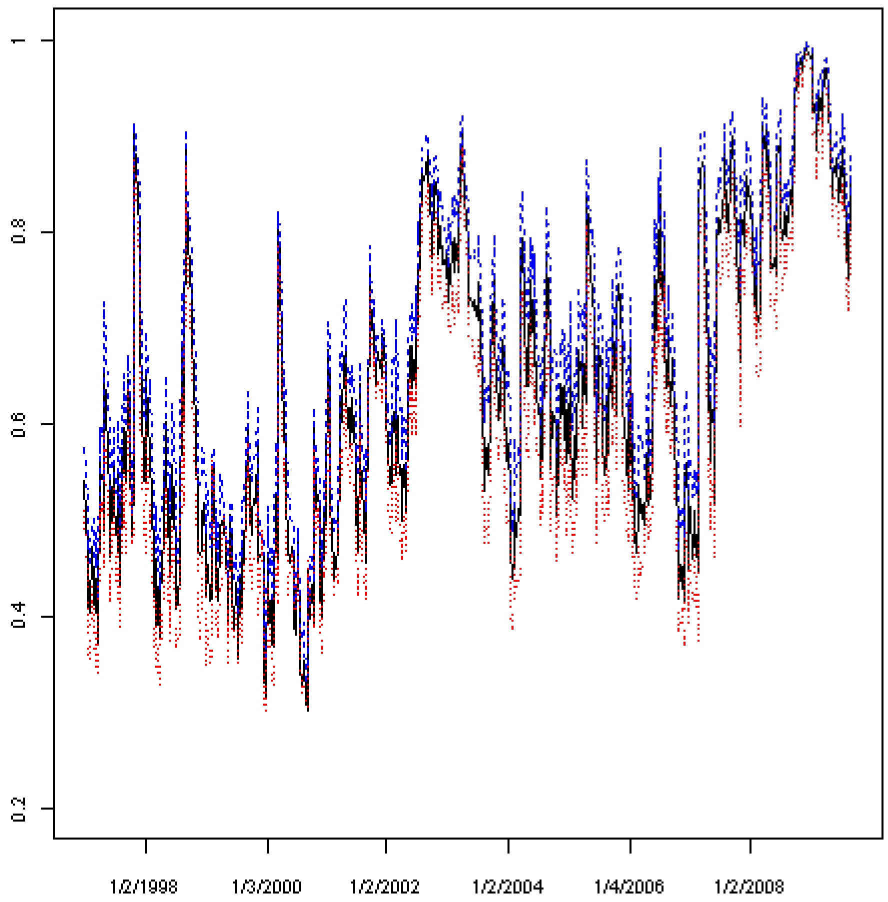


Figure 6. The trend chart for the collective correlations index (dashed), CBOE VIX (solid) and S&P 500 index (dotted) from January 1997 to August 2009.

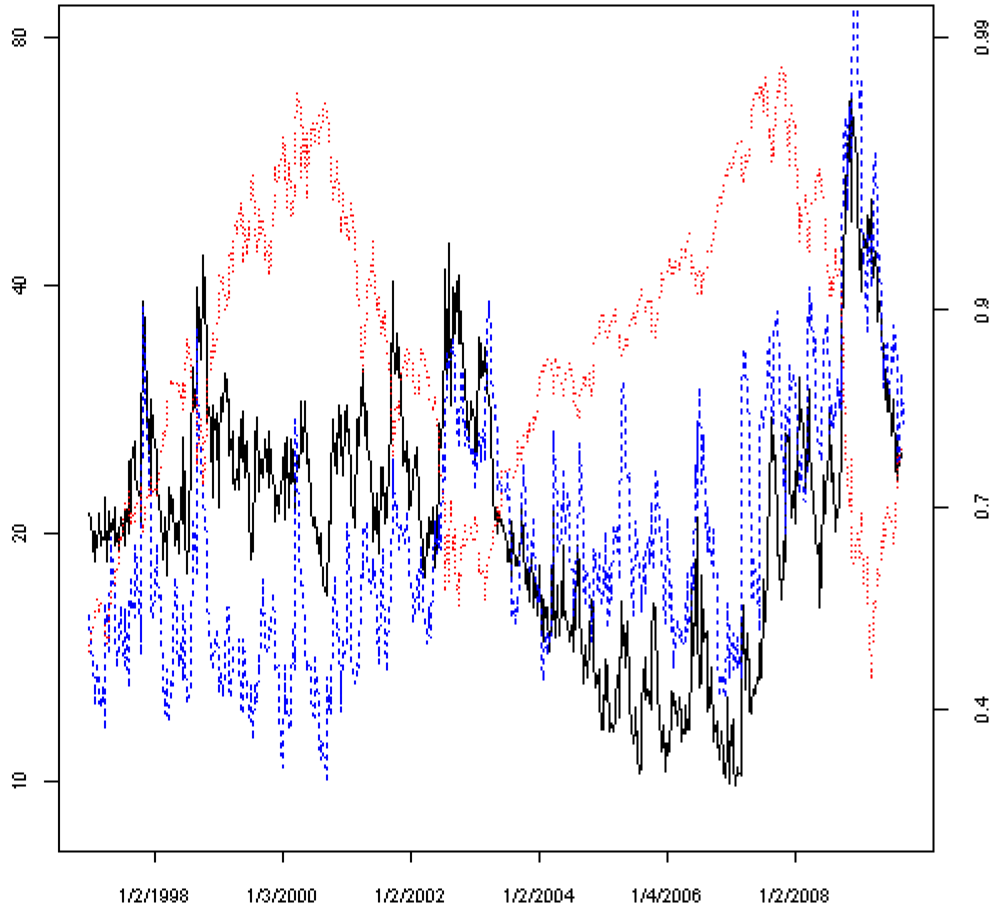


Figure 7. Typical paths of the collective correlations simulated by the DCC model and the network correlation model with different values of β_0. The initial condition is set as of January 2, 2007. The other parameters are estimated with the data in the preceding two years respectively.

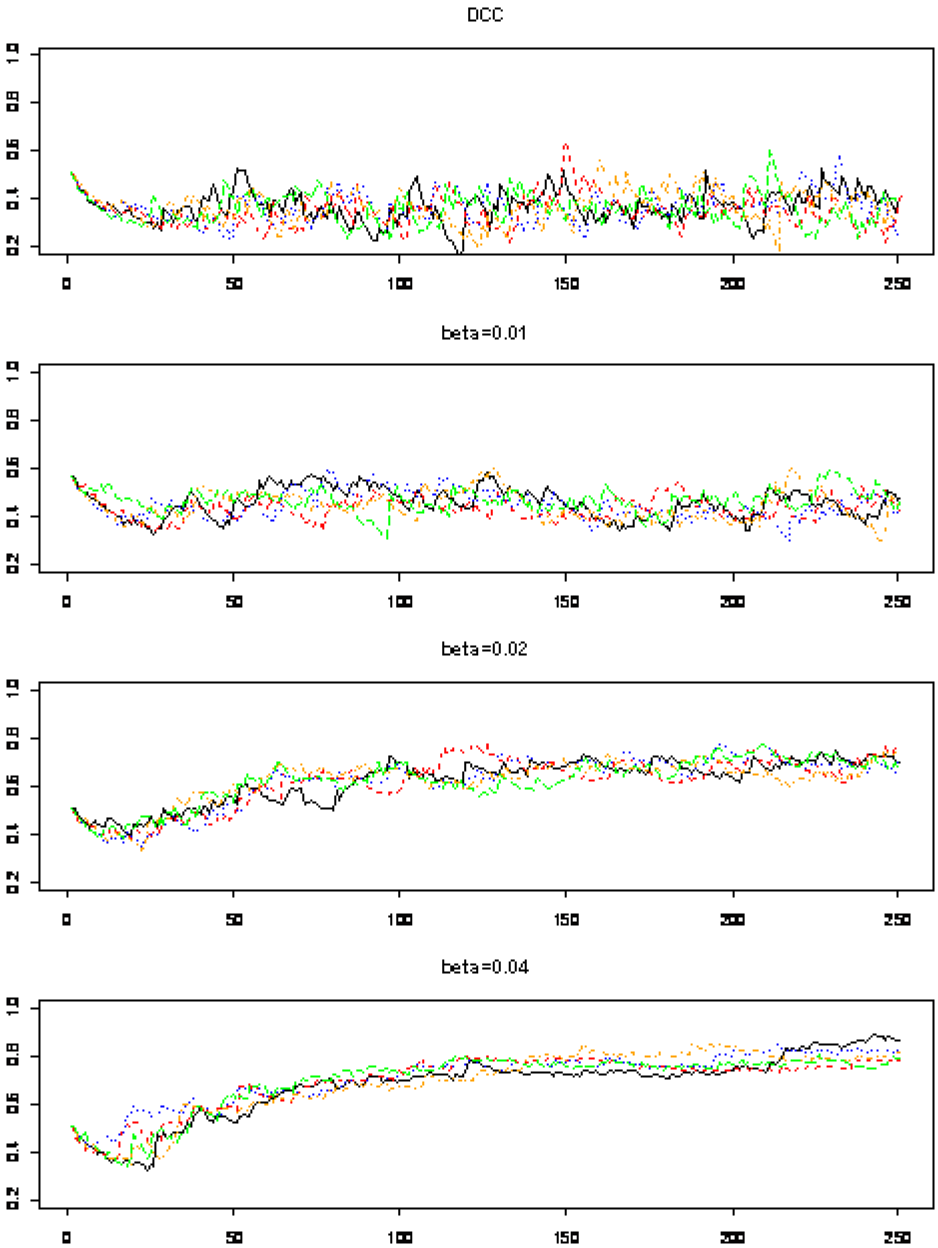


Figure 8. Distribution of the first passage time with respective to different values of β_0: 0.25 (dash), 0.3 (solid), 0.35 (dotted), 0.4(dot-dashed) and 0.45 (long dashed). The initial condition is set as of January 2, 2007.