

Modeling cross correlations within a many-assets market

H. E. Roman,¹ M. Albergante,¹ M. Colombo,¹ F. Croccolo,¹ F. Marini,² and C. Riccardi¹
¹*Dipartimento di Fisica, Università di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*
²*Dipartimento di Fisica, Università di Milano, Via Celoria 16, 20133 Milano, Italy*

(Received 6 December 2005; revised manuscript received 3 February 2006; published 28 March 2006)

A simple model for simulating cross correlations of a many-assets market is discussed. Correlations between assets are initially considered within the context of the well-known one-factor model, in which a driving term common to all stocks is present. The results are compared to those of real market data corresponding to a set of 445 stocks taken from the Standard and Poors 500 index. The model is further extended by introducing a stochastic volatility within each time series using an autoregressive scheme. This artificial market reproduces the empirically observed fat tails in the distribution function of logarithmic price variations and, more important, leads to additional cross correlations between the time series, in better agreement with the real market behavior.

DOI: [10.1103/PhysRevE.73.036129](https://doi.org/10.1103/PhysRevE.73.036129)

PACS number(s): 89.65.Gh, 89.75.-k, 89.75.Hc

I. INTRODUCTION

Very often the underlying mechanisms responsible for the unpredictable behavior of a complex system are not known. Typically, the temporal (or spatial) behavior of the system can be characterized by recording the simultaneous variations of observables associated to each of its single components. The variations are represented in the form of time series, from which one can determine the degree of correlation between them.

In the case of a many-component complex system, an accurate description of such cross correlations requires the use of a large number of parameters. A challenging complex system is a stock market, typically composed of several hundreds or thousands of assets which are necessarily interrelated to each other, either because they belong to the same economic sector, or because of circumstantial factors determined by the global market behavior [1].

In a recent paper, Bonanno *et al.* [2] have analyzed the structure of a many-assets market by looking at the so-called minimal spanning tree (MST). They show that the ubiquitous one-factor model [1] leads to wrong predictions regarding several features typical of real market MST. In this paper, we elaborate such a model further by introducing a stochastic volatility (see, e.g., [3,4]) that, in keeping with the simplicity of the approach, significantly improves its performance. In particular, the variable volatility, introduced within an autoregressive approach and constructed to reproduce the fat tails of the empirical probability distributions, has the interesting merit of inducing additional cross correlations between the assets such that the internal structure of the resulting MST becomes similar to that of the market.

The paper is organized as follows. In Sec. II we analyze real market data and introduce the basic definitions and statistical quantities of interest. In Sec. III, the one-factor model is discussed and the results compared to the Standard and Poors 500 (S&P500) data. In Sec. IV, the additional ingredient of a stochastic volatility, obtained within a one-parameter autoregressive model, is introduced. Limitations of the model are discussed in Sec. V. Finally, in Sec. VI we present our conclusions.

II. REAL MARKET DATA: STATISTICAL ANALYSIS

We consider daily (close) price variations for a set of $N=445$ stocks taken from the Standard and Poors 500 index, spanning a time interval of $T=1600$ trading days [5] (1994–2001). Here, we consider the logarithm of the i th stock price, $\ln \bar{P}_i$, as the working variable, and denote its daily variation at day t as $\Delta \bar{S}_i(t) = \ln[\bar{P}_i(t)/\bar{P}_i(t-1)]$, with $1 \leq i \leq N$. Variables carrying the overline will indicate real market values.

In order to proceed with the analysis, we define the mean market variations $\Delta \bar{S}_0$ as

$$\Delta \bar{S}_0(t) = \frac{1}{N} \sum_{i=1}^N \Delta \bar{S}_i(t), \quad (1)$$

which is playing the role of the market index. Other possible definitions of the index, i.e., by using weighted averages of $\Delta \bar{S}_i(t)$, yield similar quantitative results. Further, we denote the variance (square “volatility”) of stock i as, $\bar{\Sigma}_i^2 = \langle (\Delta \bar{S}_i)^2 \rangle_T - \langle \Delta \bar{S}_i \rangle_T^2$, and the associated index volatility as $\bar{\Sigma}_0$. Here, the averages are evaluated over the whole time series of length T .

The probability distribution function (PDF) $P(g)$ of daily variations $\Delta \bar{S}_i(t)$ is shown in Fig. 1 as a function of the scaled variable $g \equiv (\Delta \bar{S}_i - \langle \Delta \bar{S}_i \rangle_T) / \bar{\Sigma}_i$. To be noted is that $P(g)$ displays long tails for $|g| \gg 1$, decaying as a power-law $P(g) \sim |g|^{-\gamma}$, with $\gamma \approx 5$. Notice also that for the present data, the PDF turns out to be slightly asymmetric [6], with a longer tail for negative values of g . In this work, however, we will disregard such an asymmetry and consider a mean decaying exponent $\gamma \approx 4.75$ at both sides of the distribution in our fitting procedure (see Sec. IV below).

A central role in the model is played by the correlation, or covariance, between stock i and the index, defined as

$$\bar{C}_{i,0} = \frac{1}{\bar{\Sigma}_i \bar{\Sigma}_0} (\langle \Delta \bar{S}_i \Delta \bar{S}_0 \rangle_T - \langle \Delta \bar{S}_i \rangle_T \langle \Delta \bar{S}_0 \rangle_T). \quad (2)$$

Results for the present data are reported in Fig. 2, where one can see the rather conspicuous correlations between assets

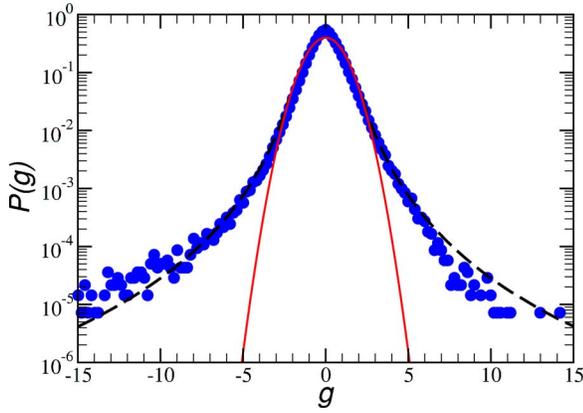


FIG. 1. (Color online) The PDF of scaled logarithmic price variations $P(g)$ versus $g = (\Delta\bar{S}_i - \langle \Delta\bar{S}_i \rangle_T) / \bar{\Sigma}_i$, for the set of 445 stocks taken from the S&P500 index (full circles). The dashed line is a fit with the form $f(x) \sim |x|^{-\gamma}$ for $|x| > 2$, where $\gamma = 4.75$. The Gaussian distribution is represented by the continuous line.

and the index [6], having a mean value of about 0.45.

The corresponding cross correlation between stocks i and j is denoted as $\bar{C}_{i,j}$. The distribution of $\bar{C}_{i,j}$ will be shown later and compared with model predictions in Sec. III and IV.

From $\bar{C}_{i,j}$, one can define a distance $\bar{d}_{i,j}$ between the two stocks according to

$$\bar{d}_{i,j} = \sqrt{2(1 - \bar{C}_{i,j})}. \quad (3)$$

One way of characterizing the resulting internal structure of the market is to evaluate the minimal spanning tree [2]. The latter is a loopless structure having the minimal chemical length, where each node (asset) is connected to its closest one and no isolated parts are present. The resulting MST for the real market can be visualized in Fig. 3. The market tree displays few large clusters of nodes reflecting the stock subdivision in economic sectors [2], and a complex ramified structure. We consider next the problem of modeling the stock market discussed so far.

III. ONE-FACTOR MODEL: CROSS CORRELATIONS BETWEEN ASSETS

We start by modeling the daily variations of the logarithmic price of asset i ΔS_i according to the one-factor model (OFM) [1,2],

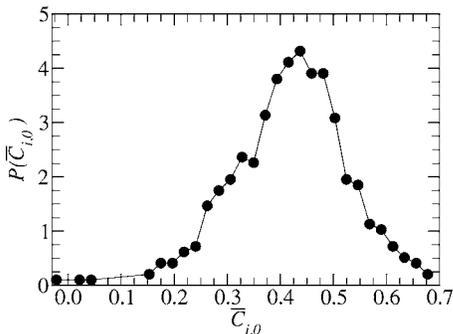


FIG. 2. The distribution function of stock-index correlations $\bar{C}_{i,0}$ for the 445 stocks from the S&P500 index.

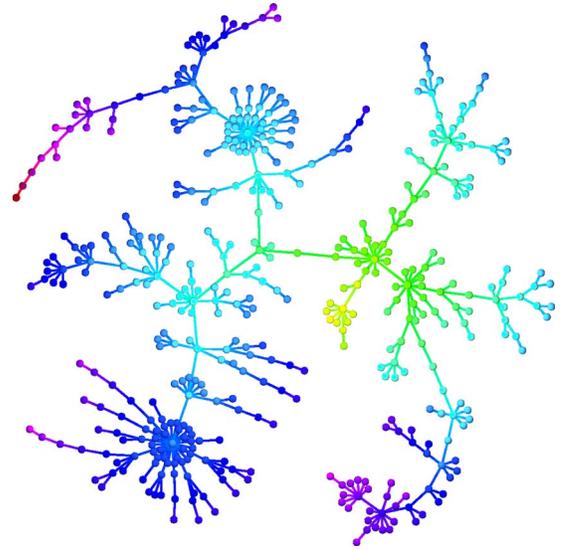


FIG. 3. (Color online) The minimal spanning tree (lines) for a set of 445 stocks (full circles) taken from the S&P500 index. The MST has been constructed using the Whitney's Algorithm 422 [7] and drawn using the Tulip software [8].

$$\Delta S_i(t) = \alpha_i + \beta_i \Delta G(t) + \sigma_i \eta_i(t), \quad (4)$$

where α_i , β_i , and σ_i are parameters, ΔG is the driving stochastic process common to all stocks, and η_i is a random number, drawn from a Gaussian distribution with zero mean and unit variance.

Following the OFM, we identify $\Delta G(t)$ with the real market index variations, which are here obtained from the mean stock variations according to Eq. (1), i.e., we assume that $\Delta G(t) = \Delta \bar{S}_0(t)$ in Eq. (4). Next, the set of three parameters α_i , β_i , and σ_i can be determined for each stock i by imposing the three conditions: $\langle \Delta S_i \rangle_T = \langle \Delta \bar{S}_i \rangle_T$, $\bar{\Sigma}_i = \bar{\Sigma}_i$, and $C_{i,0} = \bar{C}_{i,0}$, containing the first and second moments of $\Delta \bar{S}_i$ and its correlation with the market index. In this way, we expect to reproduce the behavior of the real stock $\Delta \bar{S}_i$ with the OFM counterpart ΔS_i . This fitting procedure then yields (see Appendix A for details)

$$\beta_i = \frac{\bar{\Sigma}_i}{\bar{\Sigma}_0} \bar{C}_{i,0},$$

$$\alpha_i = \langle \Delta \bar{S}_i \rangle_T - \beta_i \langle \Delta \bar{S}_0 \rangle_T,$$

$$\sigma_i = \bar{\Sigma}_i \sqrt{1 - \bar{C}_{i,0}^2}. \quad (5)$$

Using the above expressions, Eq. (4) becomes

$$\Delta S_i(t) = \langle \Delta \bar{S}_i \rangle_T + \bar{\Sigma}_i \bar{C}_{i,0} \Delta A(t) + \bar{\Sigma}_i \sqrt{1 - \bar{C}_{i,0}^2} \eta_i(t), \quad (6)$$

where the driving factor $\Delta A = (\Delta \bar{S}_0 - \langle \Delta \bar{S}_0 \rangle_T) / \bar{\Sigma}_0$, has zero mean and unit variance.

We can now verify that the average over the N time series in Eq. (6) is consistent with the variations of the market index, i.e.,

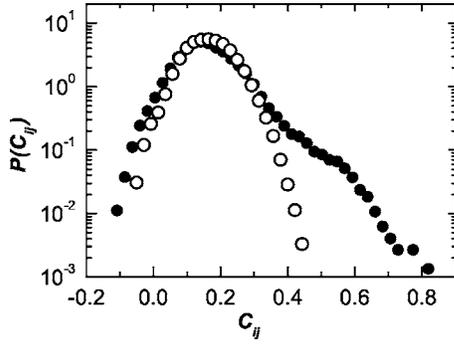


FIG. 4. The PDF of the cross-correlations $C_{i,j}$ between the 445 time series (open circles) obtained using the OFM, Eq. (6). The corresponding results obtained from the 445 stocks of the S&P500 index are shown by the full circles.

$$\frac{1}{N} \sum_{i=1}^N \Delta S_i(t) = \frac{1}{N} \sum_{i=1}^N \langle \Delta \bar{S}_i \rangle_T + \Delta A(t) \frac{1}{N} \sum_{i=1}^N \bar{\Sigma}_i \bar{C}_{i,0}, \quad (7)$$

obtained by neglecting the term $(1/N) \sum_{i=1}^N \sigma_i \eta_i$, which is small in absolute value and at most of $\mathcal{O}(1/\sqrt{N})$. Then, using the definition of $\bar{C}_{i,0}$ from Eq. (2), we find

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \bar{\Sigma}_i \bar{C}_{i,0} &= \frac{1}{\bar{\Sigma}_0} \left(\left\langle \frac{1}{N} \sum_{i=1}^N \Delta \bar{S}_i \Delta \bar{S}_0 \right\rangle_T \right. \\ &\quad \left. - \left\langle \frac{1}{N} \sum_{i=1}^N \Delta \bar{S}_i \right\rangle_T \langle \Delta \bar{S}_0 \rangle_T \right) \\ &= \frac{1}{\bar{\Sigma}_0} (\langle (\Delta \bar{S}_0)^2 \rangle_T - \langle \Delta \bar{S}_0 \rangle_T^2) = \bar{\Sigma}_0, \end{aligned} \quad (8)$$

yielding, together with the definition of ΔA in Eq. (6),

$$\frac{1}{N} \sum_{i=1}^N \Delta S_i(t) = \langle \Delta \bar{S}_0 \rangle_T + \bar{\Sigma}_0 \Delta A(t) = \Delta \bar{S}_0(t). \quad (9)$$

We have numerically verified that Eq. (9) is indeed accurately satisfied.

Let us consider next the internal structure of the OFM artificial market. It is straightforward to show that now the cross correlations $C_{i,j}$ (see Appendix A for details) are given by,

$$C_{i,j} = \bar{C}_{i,0} \bar{C}_{j,0}, \quad i \neq j, \quad (10)$$

indicating that the driving term is the only source of correlation between the i th and j th time series. We have calculated numerically the whole set of values $C_{i,j}$ for the OFM and obtained the corresponding probability distribution function. The latter is compared with the real market results and shown in Fig. 4.

As is apparent from the figure, the OFM fails to describe the long positive tail of the cross correlations distribution of the market. If, for instance, such large cross-correlation events occur for a pair of stocks in which one, or both, of them are weakly correlated to the index, then Eq. (10) underestimates their mutual correlation, yielding a faster decay-

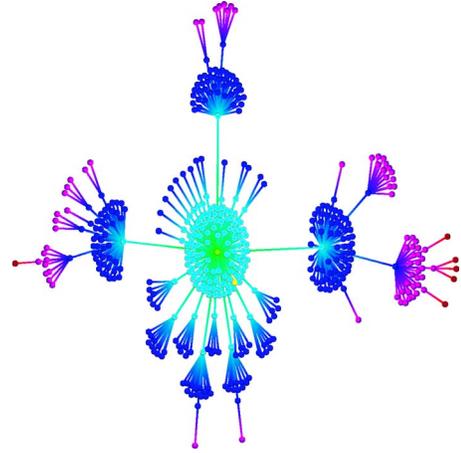


FIG. 5. (Color online) The minimal spanning tree (lines) for a set of 445 time series (full circles) generated using the OFM, Eq. (6).

ing tail. The resulting MST has therefore a low similarity with that of the real market, as can be inspected from Fig. 5. In particular, a central dominating cluster is evident and the ramifications show a higher degree of order as compared to the more random structure of the S&P500 MST.

There is another important drawback of the OFM regarding its PDF of logarithmic price variations. The latter has a Gaussian shape (see, e.g., the continuous line in Fig. 1), i.e., it lacks fat tails. This failure can be repaired in a simple way by introducing a stochastic volatility in the artificial time series, as we discuss in the following.

IV. STOCHASTIC VOLATILITY: FAT TAILS AND MINIMAL SPANNING TREE

A stochastic volatility associated to $\Delta S_i(t)$, Eq. (6), can be introduced by defining a new variable in the form

$$\Delta S'_i(t) = \Sigma'_i(t) \frac{\Delta S_i(t)}{\bar{\Sigma}_i}, \quad (11)$$

where the variable volatility $\Sigma'_i(t)$ obeys a one-parameter Auto-Regressive Conditional Heteroskedasticity (ARCH) model [9] of the type

$$\Sigma_i'^2(t) = a_i + b[\Delta S'_i(t-1)]^2. \quad (12)$$

As is well known, the resulting PDF of the logarithmic variations $\Delta S'_i$ displays power-law decaying tails, with an exponent γ that depends on b [10,11]. In this work, we use the value $b=0.6$ for all artificial time series, which yields $\gamma \approx 4.75$, as dictated by the market PDF shown in Fig. 1. Using the result $\langle \Sigma_i'^2 \rangle = a_i / (1-b)$, the parameter a_i can be determined by the condition $\langle \Sigma_i'^2 \rangle = \bar{\Sigma}_i^2$, yielding

$$a_i = (1-b) \bar{\Sigma}_i^2. \quad (13)$$

Stochastic volatility models have been considered extensively in the literature (see, e.g., [4,9,12,13], and refs. therein). A particular stochastic volatility model has been considered recently within the context of modeling a market index [14].

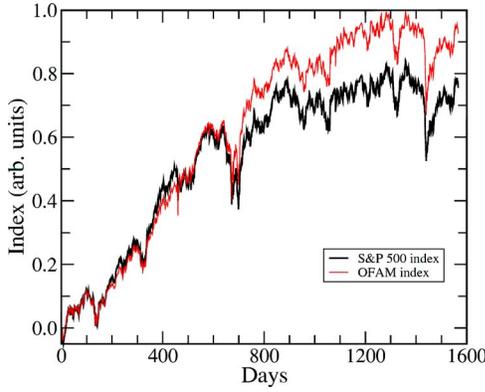


FIG. 6. (Color online) Comparison between the driving index $G(t)$ (thick black line) and the OFAM result $I(t)$ (thin line) versus trading days, obtained from the partial sums $G(t) = \sum_{t'=1}^t \Delta \bar{S}_0(t')$ and $I(t) = \sum_{t'=1}^t \Delta S'_i(t')$, respectively.

In the present model, that we denote as the one-factor ARCH model (OFAM), the driving factor ΔA used in Eq. (6) can itself be taken as a one-parameter ARCH variable with zero mean and unit variance, i.e., with $a=1-b$. We have verified that this choice yields similar quantitative results as in the case in which ΔA is taken from the real market data. In the following, we use the same driving term as in Eq. (6) to be consistent with the previous results.

We show below that the average over the N time series in Eq. (11), here denoted as $\Delta I(t)$, still behaves as the variations of the market index. Let us first evaluate the former analytically by neglecting, for simplicity, the term proportional to η_i in Eq. (11) [cf. also Eq. (7)], yielding the approximate expression

$$\Delta I(t) \cong \frac{1}{N} \sum_{i=1}^N \frac{\Sigma'_i(t)}{\bar{\Sigma}_i} \langle \Delta \bar{S}_i \rangle_T + \Delta A(t) \frac{1}{N} \sum_{i=1}^N \Sigma'_i(t) \bar{C}_{i,0}. \quad (14)$$

Note that Eq. (14) reduces to Eq. (9) in the case of no stochastic volatility $\Sigma'_i(t) = \bar{\Sigma}_i$, i.e., when $b=0$ in Eq. (12). The comparison between the full numerical results for the average value $I(t) = \sum_{t'=1}^t \Delta S'_i(t')$ and the real market one, $G(t) = \sum_{t'=1}^t \Delta \bar{S}_0(t')$, is shown in Fig. 6. Clearly, $I(t) \approx G(t)$ over the whole time scale considered.

Let us consider now the behavior of the correlations. The presence of a variable volatility $\Sigma'_i(t)$ yields a new correlation $C'_{i,0}$ between asset $i (\neq 0)$ and the driving term ΔA , which reads (see the Appendix B for details)

$$C'_{i,0} = \frac{1}{\bar{\Sigma}_i^2} \langle \Delta \bar{S}_i \rangle_T \langle \Sigma'_i \Delta A \rangle_T + \frac{\bar{C}_{i,0}}{\bar{\Sigma}_i} \langle \Sigma'_i (\Delta A)^2 \rangle_T, \quad (15)$$

which reduces to Eq. (2), i.e., $C'_{i,0} = \bar{C}_{i,0}$, when $\Sigma'_i(t) = \bar{\Sigma}_i$. Although the values $C'_{i,0} \neq \bar{C}_{i,0}$, the corresponding distribution function is similar to the real market one (Fig. 2), i.e., $P(C'_{i,0}) \approx P(\bar{C}_{i,0})$.

Similarly, the cross correlations between assets $C'_{i,j}$ with $i \neq j$, are modified with respect to $C_{i,j}$ by the presence of the stochastic volatility, yielding (see Appendix B for details),

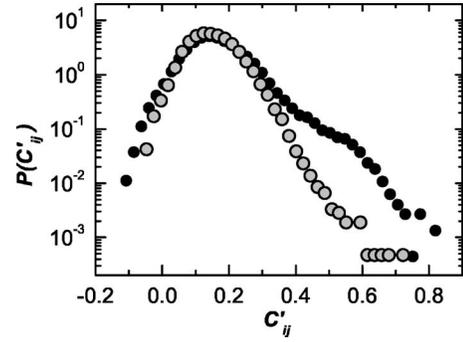


FIG. 7. The PDF of the cross correlations $C'_{i,j} (i \neq j)$ between the 445 time series (grey circles) obtained using the one-factor ARCH model, Eq. (11). The results are compared with those from the S&P500 market (full circles).

$$C'_{i,j} \cong C(\Sigma'_i \Delta A, \Sigma'_j \Delta A) \bar{C}_{i,0} \bar{C}_{j,0} + C(\Sigma'_i, \Sigma'_j) \frac{\langle \Delta \bar{S}_i \rangle_T \langle \Delta \bar{S}_j \rangle_T}{\bar{\Sigma}_i \bar{\Sigma}_j} + C(\Sigma'_j, \Sigma'_i \Delta A) \bar{C}_{i,0} \frac{\langle \Delta \bar{S}_j \rangle_T}{\bar{\Sigma}_j} + C(\Sigma'_i, \Sigma'_j \Delta A) \bar{C}_{j,0} \frac{\langle \Delta \bar{S}_i \rangle_T}{\bar{\Sigma}_i}, \quad (16)$$

where $C(X_i, Y_j) = (\langle X_i Y_j \rangle_T - \langle X_i \rangle_T \langle Y_j \rangle_T) / (\bar{\Sigma}_i \bar{\Sigma}_j)$. One can see that in the case of no stochastic volatility, i.e., when $\Sigma'_i = \bar{\Sigma}_i$ and $\Sigma'_j = \bar{\Sigma}_j$, then $C'_{i,j} = \bar{C}_{i,0} \bar{C}_{j,0} = C_{i,j}$ as in Eq. (10).

The results for the PDF corresponding to $C'_{i,j}$, $P(C'_{i,j})$, obtained numerically from the artificial time series are reported in Fig. 7. As is apparent from the figure, $P(C')$ develops a much longer tail for $C' > 0$ than its OFM counterpart shown in Fig. 4, in better agreement with the market data. Thus, the presence of a stochastic volatility induces additional cross correlations among the time series with respect to the OFM, Eq. (10).

The question now is whether such additional correlations produce a satisfactory internal metric. To answer this question, we have calculated the corresponding MST for the series generated using the OFAM. The results shown in Fig. 8 seem to confirm this expectation when compared to those from the real market, Fig. 3.

A more quantitative test consists in calculating the degree of nodes for the tree. The degree of a node represents the number of connections (also called edges) impinging on it. As one can see from Fig. 9, the distribution functions of the degree of nodes for the market and for the OFAM virtually coincide, yielding further support to our model.

V. LIMITATIONS OF THE MODEL

The previous features indicate the suitability of a stochastic volatility, in addition to the basic asset-asset correlations introduced within the OFM, to describe global aspects of the market. However, some discrepancies emerge when analyzing the results in more detail.

One way of assessing the accuracy of the previous models is to study the MST when the mean stock variation (or index) is also included in the set of assets as the $N+1$ time series.

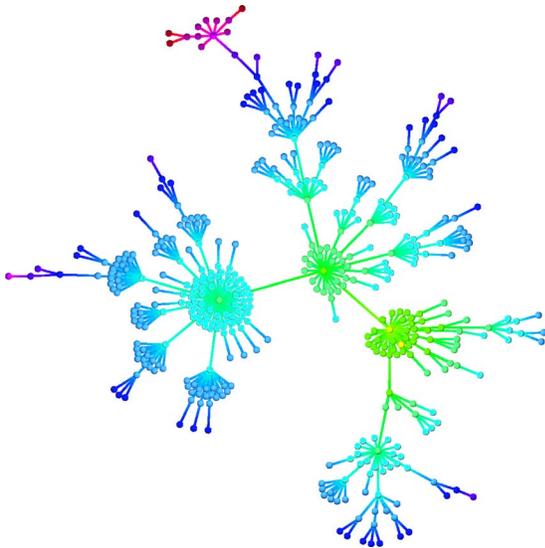


FIG. 8. (Color online) The minimal spanning tree (lines) for a set of 445 time series (full circles) generated using the one-factor model with stochastic volatility, Eq. (11).

The corresponding tree for the S&P500 market is shown in Fig. 10. Now the tree displays a clear one-center shape due to the presence of the index, although the resulting structure still displays a rather significant ramification.

The results for the OFM and OFAM are reported in Figs. 11 and 12, respectively. It is clear that the OFAM is a bit superior to the OFM (the latter showing essentially a single-cluster tree), but still inappropriate as compared to the real market behavior. We have implemented different variants of the OFAM to reproduce this market feature (Fig. 10), without obtaining any significant improvement. This problem therefore remains open.

There is another aspect of the model which deserves some attention. This regards volatility-volatility correlations. It is well known, for instance, that for real markets the volatility autocorrelation function has a long-time memory, decaying as a power-law [15] with quite small exponents, typically in the range 0.1–0.3. Here, we wish to show this market behavior and compare it with model predictions. For this purpose, we have considered the mean autocorrelation function of the

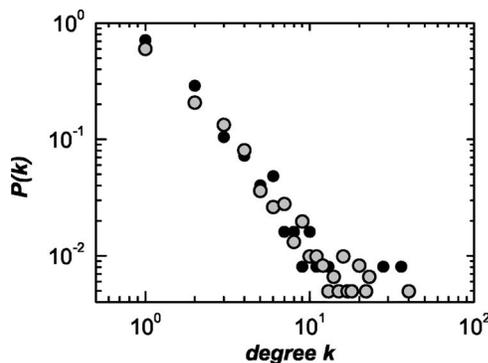


FIG. 9. The distribution function of the degree of nodes $P(k)$ versus degree k for the minimal spanning trees corresponding to the real market (full circles) and to the OFAM one (grey circles).

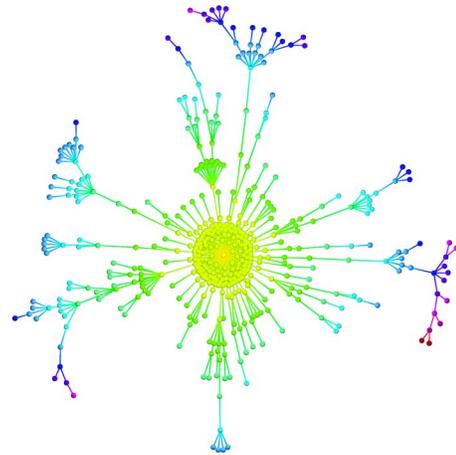


FIG. 10. (Color online) The minimal spanning tree for the set of 445 stocks plus a 446th series corresponding to their average values (“index”), obtained for the S&P500 market.

absolute returns $|\Delta \bar{S}_i(t)|$ defined as the average over the whole data set as

$$C_V(\tau) = \frac{1}{N} \sum_{i=1}^N C_V^{(i)}(\tau), \quad (17)$$

where

$$C_V^{(i)}(\tau) = \frac{1}{\Sigma_V^2(i)} (\langle |\Delta \bar{S}_i(t)| |\Delta \bar{S}_i(t + \tau)| \rangle_T - \langle |\Delta \bar{S}_i| \rangle_T^2) \quad (18)$$

is the volatility autocorrelation function of series i , and $\Sigma_V^2(i)$ is the corresponding variance. The function $C_V(\tau)$ is displayed in Fig. 13, suggesting that for the real market $C_V(\tau)$ has indeed a long power-law tail, here decaying with an exponent of about 0.2.

The OFAM results, on the other hand, show an exponential decay with a characteristic time scale of about 1.8 days. (We have checked that the slower decay displayed by the OFAM results at large τ 's is due to the lack of sufficient statistics. A more involved simulation yields an exponential

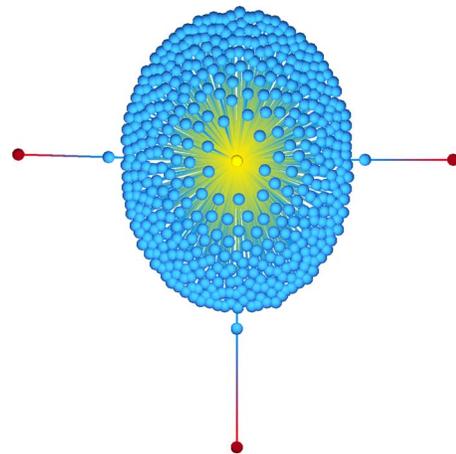


FIG. 11. (Color online) Same as in Fig. 10 for the OFM market.

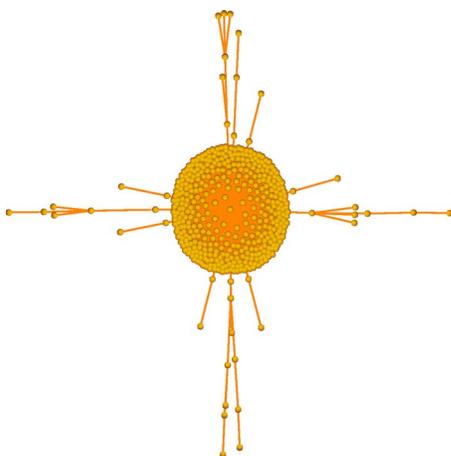


FIG. 12. (Color online) Same as in Fig. 10 for the OFAM market.

decay up to $\tau \approx 10$ with a time decay of about 1.5 days.) It is known that such exponential decay is typical of ARCH models [15]. Thus, the question remains of how a long-time memory can be implemented within our scheme to reproduce the empirical behavior. This goal goes beyond the aim of the present work and will be considered elsewhere.

VI. CONCLUSIONS

We have modeled the internal structure of a stock market by introducing correlations between each asset (time series) and a driving stock or index within the context of the one-factor model. The further introduction of a stochastic volatility for each time series, using a single ARCH parameter [9,10], considerably improves the performance of the model as reflected by the statistics of the corresponding minimal spanning tree. Indeed, the one-factor model with ARCH volatility introduced here leads, in addition to a power-law decaying PDF of logarithmic returns, to further cross correlations between stocks in closer agreement with real market behavior.

We have also generated a fully artificial market by taking the OFM parameters from the empirical PDF's, $P(\bar{C}_{i,0})$, $P(\bar{\Sigma}_i)$, and $P(\langle \Delta \bar{S}_i \rangle_T)$, instead of using the actual values $\langle \Delta S_i \rangle_T = \langle \Delta \bar{S}_i \rangle_T$, $\Sigma_i = \bar{\Sigma}_i$, and $C_{i,0} = \bar{C}_{i,0}$, for each artificial time series as in Eq. (5), and the driving series from an ARCH model with the parameter $b=0.6$. The results are quantitatively similar to those presented here.

Discrepancies between the model and the observed market behavior, however, still remain to be understood and solved. These concern the underestimation of stock-stock cross correlations at the positive tail of the distribution. Indeed, the real market PDF displays a rather large skewness and fat tail which is difficult to simulate within the simple context of the OFAM. Another important issue regards the lack of long-range memory in volatility-volatility correlations within the OFAM which needs to be addressed in the future. Despite these drawbacks, we may conclude that the OFAM can serve as a “minimal” model of a many-stock

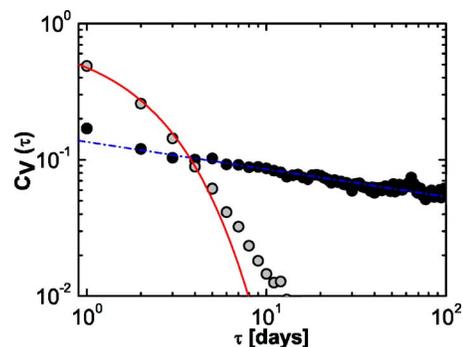


FIG. 13. (Color online) Mean autocorrelation function for the absolute value of returns (volatility) for the real market (full circles) and for the OFAM (grey circles). The dashed-dotted line is a power-law fit to the real market data of the form $y=0.135\tau^{-0.2}$, while the continuous line is the exponential fit $y=0.825 \exp(-\tau/1.8)$ to OFAM data. The values of $C_V(\tau)$ for the OFM remain below 10^{-2} and are not shown in this figure.

market, which reproduces satisfactorily some of its structural features quantified by the minimal spanning tree.

APPENDIX A: OFM

The present version of the OFM is given by

$$\Delta S_i(t) = \alpha_i + \beta_i \Delta \bar{S}_0(t) + \sigma_i \eta_i(t), \quad (\text{A1})$$

where $\Delta \bar{S}_0(t)$ is the mean stock variation of the real market, Eq. (1), α_i , β_i , and σ_i are constant parameters, while the random numbers η_i obey $\langle \eta_i \rangle = 0$ and $\langle \eta_i^2 \rangle = 1$. In the following, we calculate the first two moments of Eq. (A1) and its cross correlation with $\Delta \bar{S}_0(t)$. We find,

$$\langle \Delta S_i \rangle_T = \alpha_i + \beta_i \langle \Delta \bar{S}_0 \rangle_T \quad (\text{A2})$$

and

$$\langle (\Delta S_i)^2 \rangle_T = \alpha_i^2 + \beta_i^2 \langle (\Delta \bar{S}_0)^2 \rangle_T + 2\alpha_i \beta_i \langle \Delta \bar{S}_0 \rangle_T + \sigma_i^2 \quad (\text{A3})$$

from which we obtain,

$$\Sigma_i^2 = \langle (\Delta S_i)^2 \rangle_T - \langle \Delta S_i \rangle_T^2 = \beta_i^2 \bar{\Sigma}_0^2 + \sigma_i^2, \quad (\text{A4})$$

where $\bar{\Sigma}_0^2 = \langle (\Delta \bar{S}_0)^2 \rangle_T - \langle \Delta \bar{S}_0 \rangle_T^2$. The cross correlation (covariance) of $\Delta S_i(t)$ with the driving term $\Delta \bar{S}_0(t)$ is, by definition (cf. Eq. (2)) given by

$$C_{i,0} = \frac{1}{\Sigma_i \bar{\Sigma}_0} (\langle \Delta S_i \Delta \bar{S}_0 \rangle_T - \langle \Delta S_i \rangle_T \langle \Delta \bar{S}_0 \rangle_T), \quad (\text{A5})$$

where

$$\langle \Delta S_i \Delta \bar{S}_0 \rangle_T = \alpha_i \langle \Delta \bar{S}_0 \rangle_T + \beta_i \langle (\Delta \bar{S}_0)^2 \rangle_T, \quad (\text{A6})$$

yielding

$$C_{i,0} = \beta_i \frac{\bar{\Sigma}_0}{\Sigma_i}. \quad (\text{A7})$$

Now, imposing the three conditions: $\langle \Delta S_i \rangle_T = \langle \Delta \bar{S}_i \rangle_T$, $\Sigma_i = \bar{\Sigma}_i$ and $C_{i,0} = \bar{C}_{i,0}$, we arrive at the expressions quoted in Eq. (5),

$$\beta_i = \frac{\bar{\Sigma}_i}{\bar{\Sigma}_0} \bar{C}_{i,0},$$

$$\alpha_i = \langle \Delta \bar{S}_i \rangle_T - \beta_i \langle \Delta \bar{S}_0 \rangle_T, \quad (5)$$

$$\sigma_i = \bar{\Sigma}_i \sqrt{1 - \bar{C}_{i,0}^2}.$$

Let us consider next the cross correlation between the series i and j , which is given by,

$$C_{i,j} = \frac{1}{\bar{\Sigma}_i \bar{\Sigma}_j} (\langle \Delta S_i \Delta S_j \rangle_T - \langle \Delta S_i \rangle_T \langle \Delta S_j \rangle_T). \quad (A8)$$

Using Eqs. (A1) and (A2), and taking into account that $\langle \eta_i \eta_j \rangle = 0$ for $i \neq j$, we find

$$C_{i,j} = \frac{1}{\bar{\Sigma}_i \bar{\Sigma}_j} \beta_i \beta_j \bar{\Sigma}_0^2, \quad (A9)$$

which, according to Eq. (A7) yields the expression quoted in Eq. (10)

$$C_{i,j} = \bar{C}_{i,0} \bar{C}_{j,0}, \quad i \neq j. \quad (10)$$

APPENDIX B: OFAM

The OFAM is constructed by starting from the built-in correlations within the OFM according to

$$\Delta S_i(t) = \langle \Delta \bar{S}_i \rangle_T + \bar{\Sigma}_i \bar{C}_{i,0} \Delta A(t) + \bar{\Sigma}_i \sqrt{1 - \bar{C}_{i,0}^2} \eta_i(t), \quad (6)$$

where both ΔA and η_i have zero mean and unit variance, and used within the stochastic volatility scheme,

$$\Delta S'_i(t) = \Sigma'_i(t) \frac{\Delta S_i(t)}{\bar{\Sigma}_i}, \quad (11)$$

where the variable volatility $\Sigma'_i(t)$ obeys a one-parameter ARCH model [9] of the type

$$\Sigma_i'^2(t) = a_i + b[\Delta S'_i(t-1)]^2. \quad (12)$$

Let us evaluate next the correlations between $\Delta S'_i(t)$ and the driving factor $\Delta A(t)$. These are given by

$$C'_{i,0} = \frac{1}{\bar{\Sigma}_i} \langle \Delta S'_i \Delta A \rangle_T, \quad (B1)$$

which, using Eq. (11), becomes

$$C'_{i,0} = \frac{1}{\bar{\Sigma}_i^2} (\langle \Delta \bar{S}_i \rangle_T \langle \Sigma'_i \Delta A \rangle_T + \bar{\Sigma}_i \bar{C}_{i,0} \langle \Sigma'_i (\Delta A)^2 \rangle_T) \quad (B2)$$

as in Eq. (15). Finally, the cross correlations between series i and j are now given by

$$C'_{i,j} = \frac{1}{\bar{\Sigma}_i \bar{\Sigma}_j} (\langle \Delta S'_i \Delta S'_j \rangle_T - \langle \Delta S'_i \rangle_T \langle \Delta S'_j \rangle_T). \quad (B3)$$

In the following, we wish to estimate $C'_{i,j}$ by resorting to few approximations. Since for $i \neq j$ the terms containing the random numbers η do play a minor role, we can write

$$\Delta S'_i(t) \cong \frac{\Sigma'_i(t)}{\bar{\Sigma}_i} \langle \Delta \bar{S}_i \rangle_T + \bar{C}_{i,0} \Sigma'_i(t) \Delta A(t), \quad (B4)$$

and a similar expression for $\Delta S'_j(t)$. The corresponding average values then read

$$\langle \Delta S'_i \rangle_T \cong \frac{\langle \Sigma'_i \rangle_T}{\bar{\Sigma}_i} \langle \Delta \bar{S}_i \rangle_T + \bar{C}_{i,0} \langle \Sigma'_i \Delta A \rangle_T, \quad (B5)$$

and the cross-correlation term then becomes

$$\begin{aligned} \langle \Delta S'_i \Delta S'_j \rangle_T &\cong \langle \Sigma'_i \Sigma'_j \rangle_T \frac{\langle \Delta \bar{S}_i \rangle_T \langle \Delta \bar{S}_j \rangle_T}{\bar{\Sigma}_i \bar{\Sigma}_j} + \bar{C}_{j,0} \langle \Sigma'_i \Sigma'_j \Delta A \rangle_T \frac{\langle \Delta \bar{S}_i \rangle_T}{\bar{\Sigma}_i} \\ &+ \bar{C}_{i,0} \langle \Sigma'_j \Sigma'_i \Delta A \rangle_T \frac{\langle \Delta \bar{S}_j \rangle_T}{\bar{\Sigma}_j} \\ &+ \bar{C}_{i,0} \bar{C}_{j,0} \langle \Sigma'_i \Delta A \Sigma'_j \Delta A \rangle_T. \end{aligned} \quad (B6)$$

According to the above expressions, we find

$$\begin{aligned} &\langle \Delta S'_i \Delta S'_j \rangle_T - \langle \Delta S'_i \rangle_T \langle \Delta S'_j \rangle_T \\ &= [\langle \Sigma'_i \Sigma'_j \rangle_T - \langle \Sigma'_i \rangle_T \langle \Sigma'_j \rangle_T] \frac{\langle \Delta \bar{S}_i \rangle_T \langle \Delta \bar{S}_j \rangle_T}{\bar{\Sigma}_i \bar{\Sigma}_j} \\ &+ [\langle \Sigma'_i \Delta A \Sigma'_j \Delta A \rangle_T - \langle \Sigma'_i \Delta A \rangle_T \langle \Sigma'_j \Delta A \rangle_T] \bar{C}_{i,0} \bar{C}_{j,0} \\ &+ [\langle \Sigma'_j \Sigma'_i \Delta A \rangle_T - \langle \Sigma'_j \rangle_T \langle \Sigma'_i \Delta A \rangle_T] \bar{C}_{i,0} \frac{\langle \Delta \bar{S}_j \rangle_T}{\bar{\Sigma}_j} \\ &+ [\langle \Sigma'_i \Sigma'_j \Delta A \rangle_T - \langle \Sigma'_i \rangle_T \langle \Sigma'_j \Delta A \rangle_T] \bar{C}_{j,0} \frac{\langle \Delta \bar{S}_i \rangle_T}{\bar{\Sigma}_i}. \end{aligned} \quad (B7)$$

Using this expression into Eq. (B3), we recover Eq. (16).

- [1] Y. J. Campbell, A. W. Lo, and A. C. Mackinlay, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, 1997).
 [2] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, Phys. Rev. E **68**, 046130 (2003).
 [3] E. Ghysels, A. C. Harvey, and E. Renault, in *Statistical Methods in Finance*, Handbook of Statistics, Vol. 14, edited by G.

- S. Maddala and C. R. Rao (North Holland, Amsterdam, 1996).
 [4] J.-P. Fouque, G. Papanicolaou, and K. R. Sircar, *Derivatives in Financial Markets with Stochastic Volatility* (Cambridge University Press, Cambridge, 2000).
 [5] Yahoo Finance, <http://finance.yahoo.com>
 [6] M. Albergante, *Modellizzazione dei Mercati Finanziari* (Tesi di Laurea Triennale, Università di Milano-Bicocca, 2004); M.

- Colombo, *Analisi Statistica di Serie Temporalì con Applicazioni nel Campo Finanziario* (Tesi di Laurea Triennale, Università di Milano-Bicocca, 2004).
- [7] V. K. M. Whitney, *Commun. ACM* **15**, 273 (1972).
- [8] Auber David, <http://www.tulip-software.com>
- [9] R. Engle, *ARCH, Selected Readings* (Oxford University, Oxford, 1995).
- [10] H. E. Roman and M. Porto, *Phys. Rev. E* **63**, 036128(R) (2001).
- [11] Ch. Dose, M. Porto, and H. E. Roman, *Phys. Rev. E* **67**, 067103 (2003).
- [12] E. M. Stein and J. C. Stein, *Rev. Financ. Stud.* **4**, 727 (1991).
- [13] J. Perelló and J. Masoliver, *Phys. Rev. E* **67**, 037102 (2003).
- [14] A. A. Drăgulescu and V. M. Yakovenko, *Quant. Finance* **2**, 443 (2002).
- [15] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics* (Cambridge University Press, Cambridge, 2000).