EPJ manuscript No.
(will be inserted by the editor)

Exact multivariate amplitude distributions for non-stationary Gaussian or algebraic fluctuations of covariances or correlations in correlated financial markets

Juan C. Henao-Londono [a] and Thomas Guhr [b]

Fakultät für Physik, Universität Duisburg-Essen, Lotharstraße 1, 47048 Duisburg, Germany

Received: date / Revised version: date

Abstract Correlated financial markets are perfect examples of highly non-stationary systems. In particular, sample averaged observables of time series as variances and correlation coefficients are continuously fluctuating, directly depending on the time window in which they are evaluated. Thus, models that describe the multivariate amplitude distributions of such systems are of considerable interest. Extending previous works, we apply a methodology, where a set of measured, non-stationary correlation matrices are viewed as an ensemble for which is set up a random matrix model. This ensemble is used to average the stationary multivariate amplitude distributions measured on short time scales and thus obtain for large time scales multivariate amplitude distributions which feature heavy tails. We explicitly use four cases, combining Gaussian and algebraic distributions to compare the distributions with empirical returns distributions using daily data from companies listed in the Standard & Poor's (S&P 500) stock market index in different periods of time. The comparison in the four cases with the empirical data reveals good agreement.

PACS. 89.65.Gh Econophysics – 89.75.-k Complex systems – 05.10.Gg Statistical physics

1 Introduction

A key feature of complex systems is non-stationarity. These systems out of equilibrium still pose fundamental challenges [1,2,3,4]. A wealth of different aspects in complex systems can be traced back to non-stationarity [5,6]

2 Data set

In this study, we analyze stocks from different continuously traded companies in the S&P 500 stock market index in different periods of time. We use daily adjusted closing prices obtained from Yahoo! Finance.

To compare the distributions, we use three different time intervals. The first interval is between the years 1972 and 1992. In this interval we have data from $K = 23$ companies. The second interval is between the years 1992 and 2012. This interval is selected to compare with the results obtained in Ref. [7]. In this interval we have data from $K = 277$ companies. Finally, the third interval is between the years 2012 and 2020. In this interval we have data from $K = 461$ companies. All the companies used in the different intervals are listed in Appendix A.

[a] *e-mail:* juan.henao-londono@uni-due.de
[b] *e-mail:* thomas.guhr@uni-due.de

3 Exact multivariate amplitude distributions

3.1 Key concepts

Consider time series $S_k(t)$, $k = 1, 2, \dots, K$ of stock prices for K companies. The values $S_k(t)$ are taken in fixed time steps Δt. In general, the data contain an exponential increase due to the drift. Thus, to measure the correlations independently of this trend, it is better to use logarithmic differences instead of returns

$$G_k(t) = \ln S_k(t + \Delta t) - \ln S_k(t) = \ln \frac{S_k(t + \Delta t)}{S_k(t)}. \quad (1)$$

Anyway, logarithmic differences and returns almost coincide if the time steps Δt are sufficiently short [8,9]

$$G_k(t) \approx r_k(t) = \frac{S_k(t + \Delta t) - S_k(t)}{S_k(t)}. \quad (2)$$

The returns are well known to have distributions with heavy tails, the smaller Δt, the heavier [7]. Furthermore, the sample standard deviations σ_k, referred to as volatilities, strongly fluctuate for different time windows of the same length T [7,10].

The mean of the logarithmic differences reads [11]

$$\langle G_k(t) \rangle_T = \frac{1}{T} \sum_{t=1}^{T} G_k(t). \quad (3)$$

To compare the different K companies, it is necessary to normalize the time series. The normalized time series are defined by [11,7]

$$M_k(t) = \frac{G_k(t) - \langle G_k(t)\rangle}{\sqrt{\langle G_k^2(t)\rangle_T - \langle G_k(t)\rangle_T^2}}, \tag{4}$$

where

$$\sigma_k = \sqrt{\langle G_k^2(t)\rangle_T - \langle G_k(t)\rangle_T^2} \tag{5}$$

is the volatility of the k company in the time window of length T. These values can be viewed as the elements of a $K \times T$ rectangular matrix M. With these normalizations and rescalings, it can be measured correlations in such a way that all companies and all stocks are treated on equal footing.

The correlation coefficient for the stocks k and l is defined as [7]

$$C_{kl} = \langle M_k(t) M_l(t)\rangle_T = \frac{1}{T}\sum_{t=1}^{T} M_k(t) M_l(t), \tag{6}$$

which can be written as

$$C_{kl} = \frac{\langle G_k(t) G_l(t)\rangle_T - \langle G_k(t)\rangle_T \langle G_l(t)\rangle_T}{\sigma_k \sigma_l}. \tag{7}$$

The coefficients C_{kl} are the elements of a $K \times K$ square matrix C, the correlation matrix. The limiting values of these correlation coefficients

$$C_{kl}^{\lim} = \begin{cases} +1 & \text{completely correlated} \\ 0 & \text{completely uncorrelated} \\ -1 & \text{completely anticorrelated} \end{cases}. \tag{8}$$

The time average of C_{kl} can be viewed as the matrix product of the rectangular matrix M $(K \times T)$ with its transpose matrix $M^\dagger$ $(T \times K)$, divided by T. Thus, the correlation matrix can be written in the form

$$C = \frac{1}{T} M M^\dagger. \tag{9}$$

The correlation matrix C is real and symmetric. Using the correlation matrix C is possible to define the covariance matrix [12,13,14,15,11]

$$\Sigma = \sigma C \sigma, \tag{10}$$

where the diagonal matrix σ contains the volatilities σ_k, $k = 1, \ldots, K$.

From the results obtained in [13,7], it does not make a difference the calculation with a covariance or a correlation matrix. Thus, to ease the comparison between works, we will use covariance matrices as in [7].

3.2 General considerations

To compare the K variate distributions with data, the crucial idea is to construct K univariate distributions out of the K variate one which are then overlaid [11]. To decouple the amplitudes, we rotate the vector r into the eigenbasis of the correlation matrix C or the eigenbasis of the covariance matrix Σ [7,11]. More precisely, we use the diagonalization

$$C = U\Lambda U^\dagger \text{ such that } C^{-1/2} = U\Lambda^{-1/2}U^\dagger \tag{11}$$

or

$$\Sigma = U\Lambda U^\dagger \text{ such that } \Sigma^{-1/2} = U\Lambda^{-1/2}U^\dagger, \tag{12}$$

where U is an orthogonal $K \times K$ matrix and Λ is the diagonal matrix of the eigenvalues Λ_k. As they are positive definite, the square roots $\Lambda_k^{1/2}$ are real, we choose them positive. We use the rotated amplitudes

$$\tilde{r} = U^\dagger r \tag{13}$$

as new arguments of the ensemble averaged amplitude distribution.

When analyzing data, K is given, we obtain the matrices C and Σ by using the originally measured amplitudes for sampling over the long time interval. In all cases, the parameter N is a fit parameter, measuring the strenght of the fluctuations. Experience tells, that N sensitevily determines the shape and is best obtained by fitting the whole distribution to the data.

3.3 Four cases distributions

In the Gaussian-Gaussian case with the Markovian situation $D = \mathbb{1}_N$ the distribution reads

$$\langle p\rangle_{GG}^{(k)}(\tilde{r}_k|\Lambda_k, \mathbb{1}_N) = \frac{1}{2^{(N-1)/2}\Gamma(N/2)\sqrt{\pi\Lambda_k/N}} \sqrt{\frac{N\tilde{r}_k^2}{\Lambda_k}}^{(N-1)/2} \mathrm{K}_{(1-N)/2}\left(\sqrt{\frac{N\tilde{r}_k^2}{\Lambda_k}}\right), \tag{14}$$

In the Gaussian-algebraic case with the Markovian situation $D = \mathbb{1}_N$ the distribution reads

$$\langle p\rangle_{GA}^{(k)}(\tilde{r}_k|\Lambda_k, \mathbb{1}_N) = \frac{\Gamma(L-(K+N)/2+1)\,\Gamma(L-(K-1)/2)}{\Gamma(L-(K+N-1)/2)\,\Gamma(N/2)\sqrt{2\pi\Lambda_k M/N}} \mathrm{U}\left(L - \frac{K+N}{2} + 1, \frac{1-N}{2} + 1, \frac{N}{2M}\frac{\tilde{r}_k^2}{\Lambda_k}\right) \tag{15}$$

In the algebraic-Gaussian case with the Markovian situation $D = \mathbb{1}_N$ the distribution reads

$$\langle p\rangle_{AG}^{(k)}(\tilde{r}_k|\Lambda_k, \mathbb{1}_N) = \frac{\Gamma(l-(K-1)/2)\,\Gamma(l-(K-N)/2)}{\Gamma(l-K/2)\,\Gamma(N/2)\sqrt{2\pi\Lambda_k m/N}} \mathrm{U}\left(l - \frac{K-1}{2}, \frac{1-N}{2} + 1, \frac{N}{2m}\frac{\tilde{r}_k^2}{\Lambda_k}\right) \tag{16}$$

Finally, in the algebraic-algebraic case with the Markovian situation $D = \mathbb{1}_N$ the distribution reads

$$\begin{aligned}
&\langle p \rangle_{AA}^{(k)} \left(\tilde{r}_k | \Lambda_k, \mathbb{1}_N \right) = \\
&\quad \frac{\Gamma \left(l - (K-1)/2 \right) \Gamma \left(l - (K-N)/2 \right)}{\Gamma \left(l - K/2 \right) \Gamma \left(L + l - (K-1) \right) \sqrt{\pi \Lambda_k M m / N}} \\
&\quad \frac{\Gamma \left(L - (K+N)/2 + 1 \right) \Gamma \left(L - (K-1)/2 \right)}{\Gamma \left(L - (K+N-1)/2 \right) \Gamma \left(N/2 \right)} \\
&\quad {}_2\mathrm{F}_1 \left(l - \frac{K-1}{2}, L - \frac{K+N}{2} + 1, L + l - (K-1), 1 - \right.
\end{aligned} \tag{17}$$

For a visual comparison of the distributions, we plot the GG, GA, AG and AA distributions in the Markovian case in the same figure. In Fig. 1 we consider $K = 100$ positions with shape parameters $L = 55$, $l = 55$, as well as $N = 5$ which is a typical value from an empirical viewpoint. In the top are the probability densities in linear scale and in the bottom are the probability densities in logarithmic scales. From figure, it can be seen the more algebraic, the stronger peaked is the distribution and heavier are the tails.

3.4 Graphical representations

To plot a comparison of the distributions involving algebraic distributions with those in the Gaussian-Gaussian case, we choose values of L and l which ensure the existence of the first matrix moment (?). We notice that the conditions on the existence of the algebraic distributions, i.e., of their normalizations, are slightly weaker. The variances $\langle \tilde{r}_k \rangle_{YY'}(k)$ are simply given by Λ_k. The functional form of all distributions $\langle p \rangle_{YY'}(k) \left(\tilde{r}_k | D \right)$ then allows us to normalize the rotated amplitude $\tilde{r}_k$ by the standard deviation

$$\tilde{r} = \frac{\tilde{r}_k}{\sqrt{\Lambda_k}} \tag{18}$$

such that all K distributions in this variable coincide and the corresponding variances are all given by one.

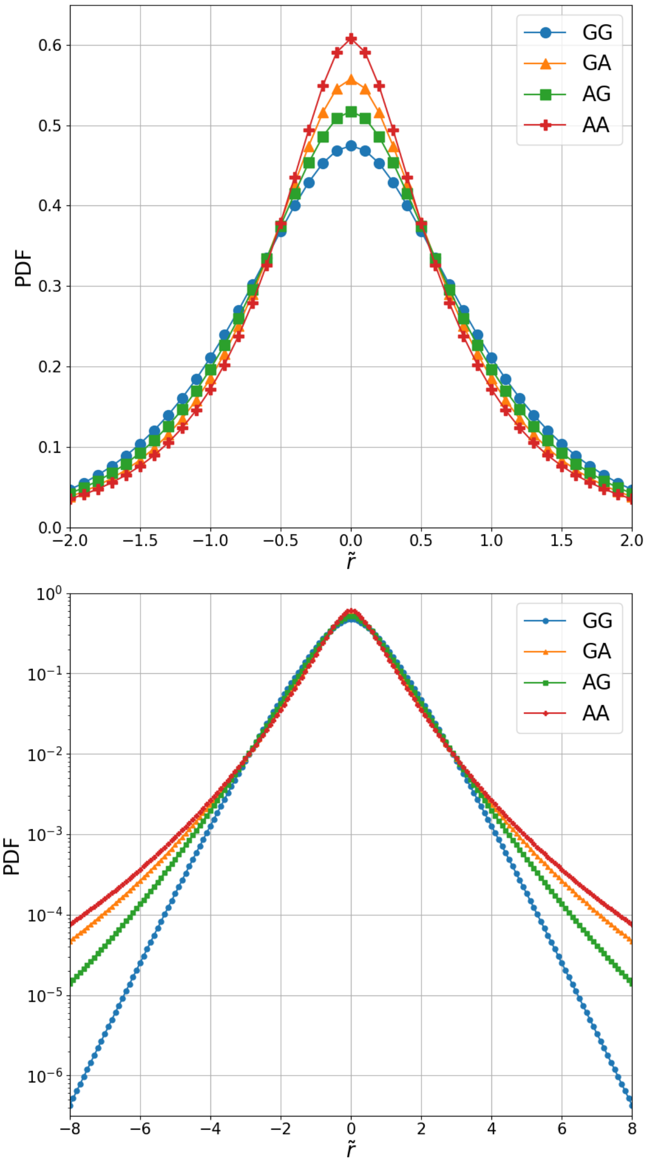

Figure 1. Probability densities $\langle p \rangle_{YY'}^{(k)}$, in the Markovian case versus the rotated amplitudes $\tilde{r}$, normalized to unit standard deviation. The four cases Gaussian-Gaussian, Gaussian-Algebraic, Algebraic-Gaussian and Algebraic-Algebraic are labeled $YY' = GG$, GA, AG and AA, respectively. Number of positions $K = 100$, shape parameters $L = 55$ and $l = 55$, strength parameter for fluctuations of correlations $N = 5$. (Top) Linear scale, abscissa between -2 and $+2$ and (bottom) logarithmic scale, abscissa between -8 and $+8$.

4 Comparison with empirical portfolio returns

4.1 Response functions on trade time scale

4.2 Response functions on physical time scale

5 Conclusion

6 Author contribution statement

TG proposed the research. JCHL carried out the analysis. All the authors contributed equally to analyzing the results and writing the paper.

One of us (JCHL) acknowledges financial support from the German Academic Exchange Service (DAAD) with the program “Research Grants - Doctoral Programmes in Germany” (Funding programme 57381412)

Appendix A S&P 500 stocks used to compare the distributions

References

1. J. B. Gao. Recurrence time statistics for chaotic systems and their applications. *Phys. Rev. Lett.*, 83:3178–3181, Oct 1999.
2. Rainer Hegger, Holger Kantz, Lorenzo Matassini, and Thomas Schreiber. Coping with nonstationarity by overembedding. *Phys. Rev. Lett.*, 84:4092–4095, May 2000.
3. Pedro Bernaola-Galván, Plamen Ch. Ivanov, Luís A. Nunes Amaral, and H. Eugene Stanley. Scale invariance in the nonstationarity of human heart rate. *Phys. Rev. Lett.*, 87:168105, Oct 2001.
4. Christoph Rieke, Karsten Sternickel, Ralph G. Andrzejak, Christian E. Elger, Peter David, and Klaus Lehnertz. Measuring nonstationarity by analyzing the loss of recurrence in dynamical systems. *Phys. Rev. Lett.*, 88:244102, May 2002.
5. R K P Zia and Per Arne Rikvold. Fluctuations and correlations in an individual-based model of biological coevolution. *Journal of Physics A: Mathematical and General*, 37(19):5135–5155, apr 2004.
6. R K P Zia and B Schmittmann. A possible classification of nonequilibrium steady states. *Journal of Physics A: Mathematical and General*, 39(24):L407–L413, may 2006.
7. Thilo A. Schmitt, Desislava Chetalova, Rudi Schäfer, and Thomas Guhr. Non-stationarity in financial time series: Generic features and tail behavior. *EPL (Europhysics Letters)*, 103(5):58003, sep 2013.
8. Jean-Philippe Bouchaud. The subtle nature of financial random walks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026104, 2005.
9. Anirban Chakraborti, Ioane Toke, Marco Patriarca, and Frédéric Abergel. Econophysics: Empirical facts and agent-based models. *arXiv.org, Quantitative Finance Papers*, 09 2009.
10. William G. Schwert. Why does stock market volatility change over time? *The Journal of Finance*, 44(5):1115–1153.
11. Thomas Guhr and Andreas Schell. Exact multivariate amplitude distributions for non-stationary gaussian or algebraic fluctuations of covariances or correlations. *Journal of Physics A: Mathematical and Theoretical*, 54(12):125002, mar 2021.
12. Thilo A. Schmitt, Desislava Chetalova, Rudi Schäfer, and Thomas Guhr. Credit risk and the instability of the financial system: An ensemble approach. *EPL (Europhysics Letters)*, 105(3):38004, feb 2014.
13. Desislava Chetalova, Thilo A. Schmitt, Rudi Schäfer, and Thomas Guhr. Portfolio return distributions: Sample statistics with stochastic correlations. *International Journal of Theoretical and Applied Finance*, 18(02):1550012, 2015.
14. Thilo Schmitt, Rudi Schäfer, and Thomas Guhr. Credit risk: Taking fluctuating asset correlations into account. *Journal of Credit Risk*, 11(3), 2015.
15. Frederik Meudt, Martin Theissen, Rudi Schäfer, and Thomas Guhr. Constructing analytically tractable ensembles of stochastic covariances with an application to financial data. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(11):P11025, nov 2015.