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 500 (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 bunch of different characteristics in complex systems can be tracked down to non-stationarity [1,2]. These systems lack any kind of equilibrium [3,4,5,6]. Financial markets are perfect examples of non-stationarity as they fluctuate considerably in time. In general, the business relations between companies and agents can change due to market expectations. During state of crisis, non-stationarity becomes dramatic [7,8,9,10,11,12,13].

The fluctuation of the correlations induces generic features in financial time series, where we showed that these fluctuations lift the tails of the multivariate amplitude distributions, making them heavy-tailed [14,15].

Our goal is to use the analytical results for the multivariate distributions of amplitudes, measured as time series in correlated, non-stationary financial markets and provide quantitative measures for the degree of non-stationarity in the correlations using the methodology first proposed in [14] and extended in [15]. These amplitudes refer to the stock price changes for the entire market.

We carry out a detailed data analysis that exposes generic features. Then we use a random matrix model to explain them. We show that non-stationarity of the correlations leads to heavy tails in the multivariate return distribution and finally, we use the approach in [14,15] to map a non-invariant situation to an effectively invariant one.

A remarkable feature of the multivariate distributions we use to compare with the financial data, is that eventually, they are of closed form or involve only single integrals. Moreover, they use a low number of free parameters: one measuring how strongly the non-stationary correlations fluctuate, and one or two shape parameters for the tails. All the other parameters can be directly measured from the data [15].

Random matrix models [16,17] fall into two classes: (I) The ensemble is fictitious. It comes into play via an ergodicity argument only. (II) The ensemble really exist and can be identified in the system. The issue of ergodicity does not arise. It is conceptually important that we here use a random matrix model in class (II) which may be seen as a new interpretation of the Wishart model and generalizations thereof for random covariance or correlation matrices [18]. In finance there are numerous random matrix applications [19,20,21,22,23,24,25,26,27,28,29,30,31] including non-Gaussian ensembles. To the best of our knowledge, all of them fall into class (I) and focus on other observables. The distributions we use arrive at rather universal and generic results, supporting view that non-stationarities can lead to universal features.

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

The paper is organized as follows: in Sect. 2 we present our data set of stocks and define the analyzed time intervals. We define the key concepts and the general considerations in Sect. 3. In Sect. 4 we show compare the theoretical distributions with the aggregated distribution of returns. Our conclusions follow in Sect. 5.

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 two different time intervals. The first interval is between the years 1992 and 2012. This interval is selected to compare with the results obtained in Ref. [14] and to extend this previous work with the new proposed distributions. In this interval we have data from $K = 277$ companies. The second interval is between the years 2012 and 2020. We want to explore the stock data for recent years and validate the universal and generic results of the distributions. In this interval we have data from $K = 461$ companies. All the companies used in the different intervals are listed in Appendix A.

3 Exact multivariate amplitude distributions

In Sect. 3.1 we introduce the fundamental quantities used in the distribution definitions. An example of the non-stationarity in financial markets is shown in Sect. 3.2. In Sect. 3.3 we set the theoretical considerations to construct the distributions. In Sect. 3.4 we define the four distribution cases and in 3.5 we show their graphical representation.

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 [32,33]

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

The returns are well known to have distributions with heavy tails, the smaller Δt, the heavier [14]. Furthermore, the sample standard deviations σ_k, referred to as volatilities, strongly fluctuate for different time windows of the same length T [14,34].

The mean of the logarithmic differences reads [15]

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

To compare the different K companies, it is necessary to normalize the time series. The normalized time series are defined by [14,15]

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

where

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

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 [14]

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

which can be written as

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

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

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

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

$$C = \frac{1}{T} M M^\dagger. \quad (9)$$

The correlation matrix C is real and symmetric. Using the correlation matrix C is possible to define the covariance matrix [15,35,36,37,38]

$$\Sigma = \sigma C \sigma, \quad (10)$$

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

From the results obtained in [14,36], 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 [14].

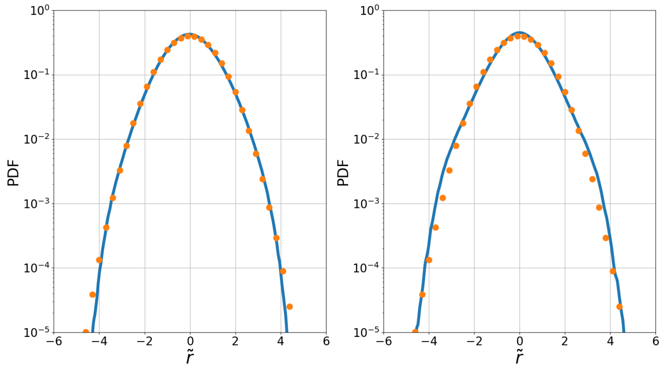

Figure 1.

3.2 Non-stationarity

In general, in non-stationary complex systems crucial parameters or distributions of observables change in an erratic, unpredictable way over time. In financial markets, the $K\times K$ correlation matrix C as a whole changes in time is an example of non-stationary. Fig. 2 shows two large correlation matrices of logarithmic differences of stock prices for companies in the Standard & Poor's 500 (S&P 500) stock market index ordered according to the Global Industry Classification Standard (GICS). The time series were measured in successive quarters.

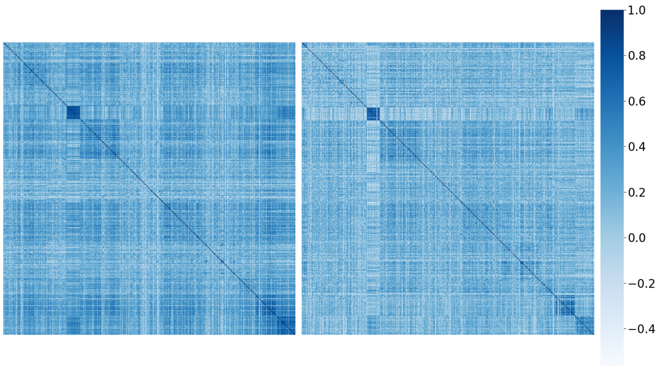

Figure 2. Correlation matrices of $K = 421$ companies for the fourth quarter of 2005 (left) and the first quarter of 2006 (right), the darker, the stronger the correlation. The companies are sorted according to industrial sectors.

In Fig. 2 it can be seen that the matrices look different for the successive quarters, because the business relations between the companies and the market expectations of the traders change in time. Despite the difference, the coarse structure remains similar, indicating some stability of the industrial sectors.

3.3 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 [15]. 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 [14,15]. 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.

3.4 Four cases distributions

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

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

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

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

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

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

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

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

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. 3 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.

3.5 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.

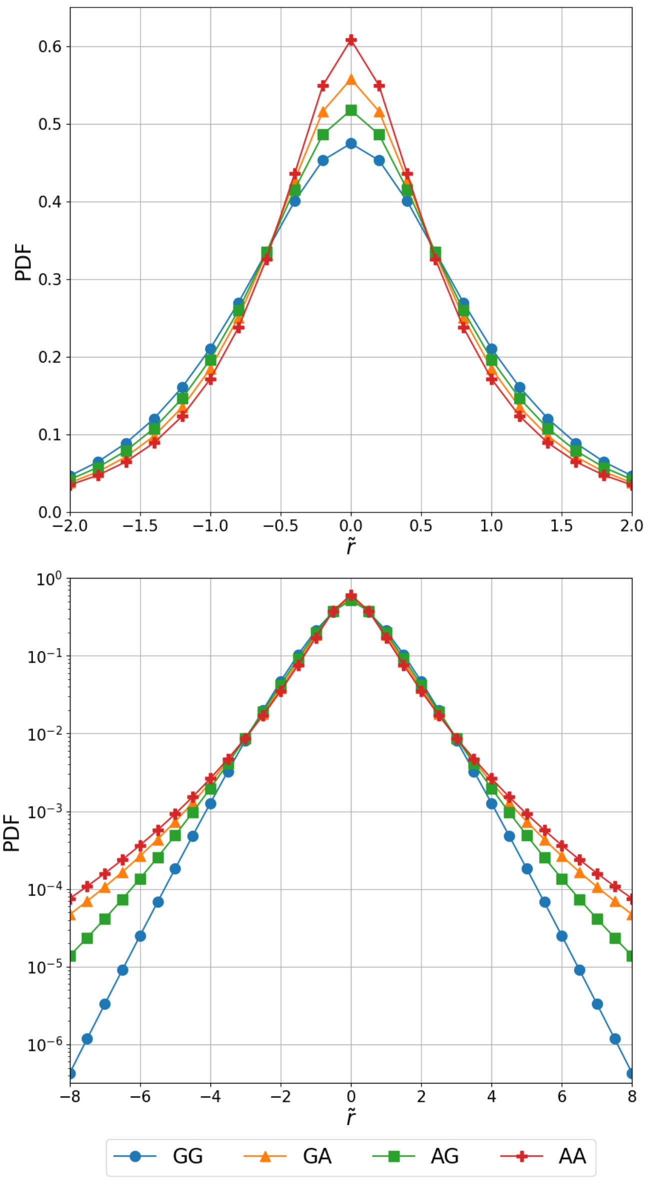

Figure 3. Probability densities $\langle p \rangle_{YY'}^{(k)}$, in the Markovian case versus the rotated amplitudes $\tilde{r}$, normalized to unit standard deviation. The four cases Gaussian-Gaussian, Gaussian-Algebraic, Algebraic-Gaussian and Algebraic-Algebraic are labeled $YY' = GG$, GA, AG and 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)

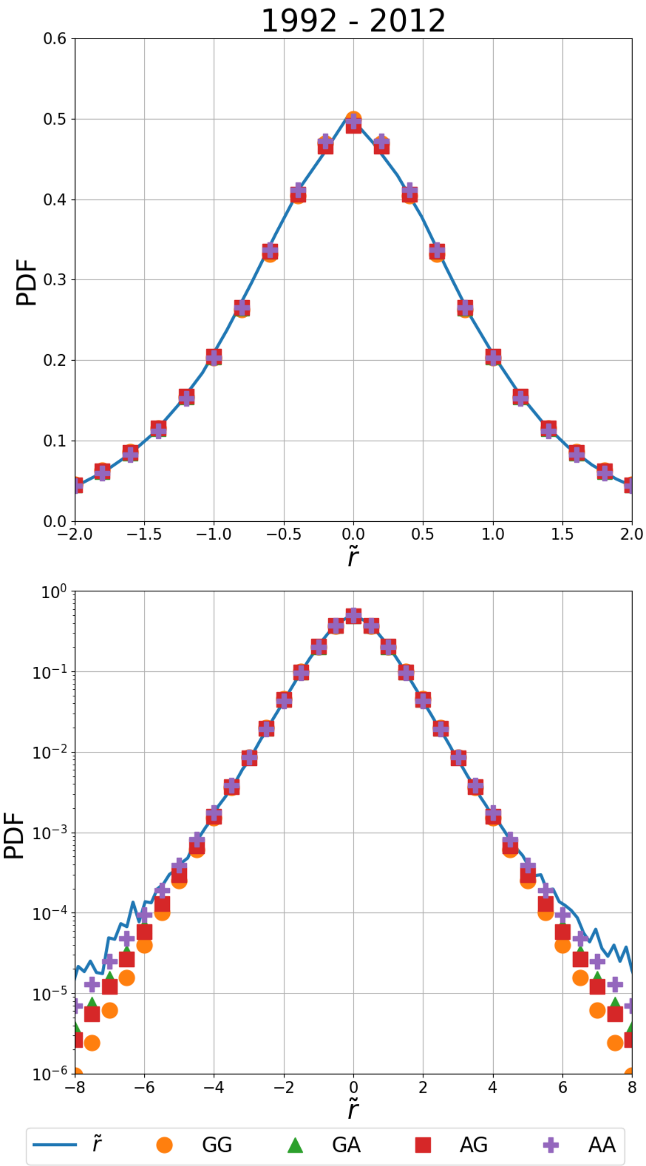

Figure 4.

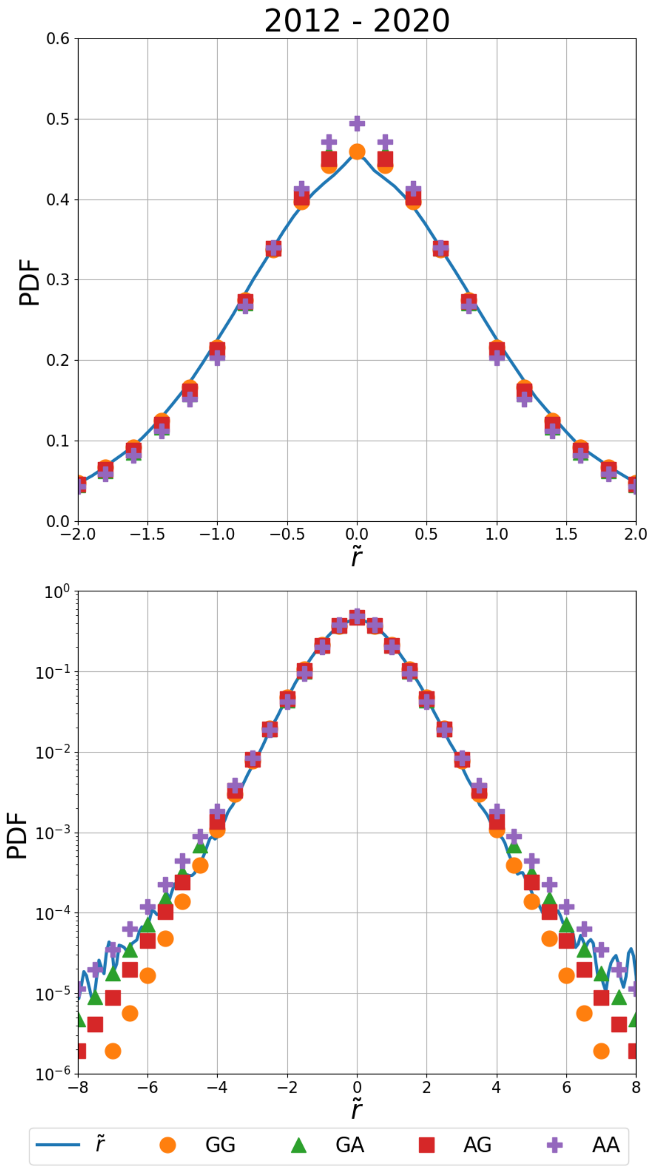

Figure 5.

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

References

1. 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.
2. 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.
3. J. B. Gao. Recurrence time statistics for chaotic systems and their applications. *Phys. Rev. Lett.*, 83:3178–3181, Oct 1999.
4. Rainer Hegger, Holger Kantz, Lorenzo Matassini, and Thomas Schreiber. Coping with nonstationarity by overembedding. *Phys. Rev. Lett.*, 84:4092–4095, May 2000.
5. 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.
6. 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.
7. Geert Bekaert and Campbell R. Harvey. Time-varying world market integration. *The Journal of Finance*, 50(2):403–444, 1995.
8. François Longin and Bruno Solnik. Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance*, 14(1):3–26, 1995.
9. J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto. Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E*, 68:056110, Nov 2003.

10. Yiting Zhang, Gladys Hui Ting Lee, Jian Cheng Wong, Jun Liang Kok, Manamohan Prusty, and Siew Ann Cheong. Will the us economy recover in 2010? a minimal spanning tree study. *Physica A: Statistical Mechanics and its Applications*, 390(11):2020–2050, 2011.
11. Dong-Ming Song, Michele Tumminello, Wei-Xing Zhou, and Rosario N. Mantegna. Evolution of worldwide stock markets, correlation structure, and correlation-based graphs. *Phys. Rev. E*, 84:026108, Aug 2011.
12. Leonidas Sandoval and Italo De Paula Franca. Correlation of financial markets in times of crisis. *Physica A: Statistical Mechanics and its Applications*, 391(1):187–208, 2012.
13. Michael C. Münnix, Takashi Shimada, Rudi Schäfer, Francois Leyvraz, Thomas H. Selgman, Thomas Guhr, and H. Eugene Stanley. Identifying states of a financial market. *Scientific Reports*, 2, 2012.
14. 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.
15. 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.
16. Thomas Guhr, Axel Müller–Groeling, and Hans A. Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Physics Reports*, 299(4):189–425, 1998.
17. Madan Lal Mehta. *Random Matrices*. Elsevier, 2004.
18. John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52, 1928.
19. Jean-Philippe Bouchaud and Marc Potters. *Thery of Financial Risks, from statistical physics to risk management*. Cambridge University Press, 2000.
20. Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 83:1467–1470, Aug 1999.
21. Laurent Laloux, Pierre Cizeau, Marc Potters, and Jean-Philippe Bouchaud. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 03(03):391–397, 2000.
22. Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, and H. Eugene Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Phys. Rev. Lett.*, 83:1471–1474, Aug 1999.
23. Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, Thomas Guhr, and H. Eugene Stanley. Random matrix approach to cross correlations in financial data. *Phys. Rev. E*, 65:066126, Jun 2002.
24. Szilárd Pafka and Imre Kondor. Estimated correlation matrices and portfolio optimization. *Physica A: Statistical Mechanics and its Applications*, 343:623–634, 2004.
25. M. Potters, J.-P. Bouchaud, and L. Laloux. Financial applications of random matrix theory: Old laces and new pieces. *Acta Physica Polonica Series B*, B35:2767–2784, 2005.
26. S. Drozdz, J. Kwapien, and P. Oswiecimka. Empirics versus rmt in financial cross-correlations. *Acta Physica Polonica Series B*, B39:4027–4039, 2008.
27. J. Kwapień, S. Drożdż, and P. Oświecimka. The bulk of the stock market correlation matrix is not pure noise. *Physica A: Statistical Mechanics and its Applications*, 359:589–606, 2006.
28. G. Biroli and J.-P. Bouchaud M. Potters. The student ensemble of correlation matrices: Eigenvalue spectrum and and kullback-leibler entropy. *Acta Physica Polonica Series B*, B39:4009–4026, 2008.
29. Zdzisław Burda, Jerzy Jurkiewicz, Maciej A. Nowak, Gabor Papp, and Ismail Zahed. Free lévy matrices and financial correlations. *Physica A: Statistical Mechanics and its Applications*, 343:694–700, 2004.
30. G. Akemann and P. Vivo. Power law deformation of wishart-laguerre ensembles of random matrices. *J. Stat. Mech.*, 0809:P09002, 2008.
31. Zdzisław Burda, Andrzej Jarosz, Maciej A. Nowak, Jerzy Jurkiewicz, Gábor Papp, and Ismail Zahed. Applying free random variables to random matrix analysis of financial data. part i: The gaussian case. *Quantitative Finance*, 11(7):1103–1124, 2011.
32. Jean-Philippe Bouchaud. The subtle nature of financial random walks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026104, 2005.
33. Anirban Chakraborti, Ioane Toke, Marco Patriarca, and Frédéric Abergel. Econophysics: Empirical facts and agent-based models. *arXiv.org, Quantitative Finance Papers*, 09 2009.
34. William G. Schwert. Why does stock market volatility change over time? *The Journal of Finance*, 44(5):1115–1153.
35. 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.
36. 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.
37. Thilo Schmitt, Rudi Schäfer, and Thomas Guhr. Credit risk: Taking fluctuating asset correlations into account. *Journal of Credit Risk*, 11(3), 2015.
38. 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.