NON-STATIONARITY IN FINANCIAL MARKETS: DYNAMICS OF MARKET STATES VERSUS GENERIC FEATURES*

Thomas Guhr

Fakultät für Physik, Universität Duisburg-Essen, Duisburg, Germany

(Received December 4, 2014)

Correlations play an important rôle when estimating risk in the financial markets. This is so from a systemic viewpoint when trying to assess the stability of the markets, but also from a practical one when, *e.g.*, optimizing portfolios. The non-stationarity of the correlations in time poses challenges for the modelling usually not encountered in the more traditional systems of statistical physics. Three recent results are presented and discussed. First it is shown how severely the exclusive look on the correlations can lead to misjudgments of the mutual dependencies. Second, the identification of distinct market states is reported and, third, generic features of return distributions are shown to be well-captured by a random matrix model.

DOI:10.5506/APhysPolB.46.1625
PACS numbers: 89.75.–k, 89.65.Gh, 05.45.–a

1. Introduction

The use of random matrices already has a successful history in econophysics [1, 2] and finance [3–7]. In particular, the spectral density of *individual* large empirical correlation matrices was found to be well-described by a random matrix ansatz. More precisely, the Marchenko–Pastur formula models the bulk of the spectral density, while those large eigenvalues which give direct information about industrial sectors lie outside. These findings had a considerable impact on portfolio optimization. Moreover, they also caught the attention of many statistical physicists who then joined the field of econophysics.

Here, I wish to present three recent results. Random matrices will appear in a new application, conceptually different from the above one. In Sec. 2, I begin with reporting a large-scale data analysis, carried out by Münnix

* Presented at the Conference "Random Matrix Theory: Foundations and Applications", Kraków, Poland, July 1–6, 2014.

and Schäfer [8], which shows, in a rather drastic fashion, that correlations alone do not provide sufficient information. Copulas, a new concept from statistics, are used to yield a much improved understanding of the mutual dependencies. Nevertheless, as shown in Sec. 3, correlations may serve to identify quasi-stationary states [9]. This is remarkable as the markets, in particular the correlations, are non-stationary. Although it might at first sound contradictory, the existence of these distinct states is still compatible with generic features. This is demonstrated in Sec. 4 where a random matrix model is discussed that can cope with the non-stationarity of the correlations to yield universal return distributions [10]. In contrast to the above mentioned application to *individual* correlation matrices, the random matrices used here model the ensemble of correlation matrices that is generated by the fluctuations, *i.e.*, by the non-stationarity. This ensemble exists and is not fictitious. The reason why the latter two findings are not contradictory is given in the conclusions in Sec. 5.

2. Financial correlations and beyond

We consider K stocks with prices $S_k(t)$, $k = 1, \ldots, K$ measured at times $t = 1, \ldots, T$. One is interested in the statistical properties of the returns $R_k(t) = (S_k(t + \Delta t) - S_k(t))/S_k(t)$. These are the dimensionless price changes from time t to time $t + \Delta t$ which depend on the chosen return horizon Δt. From the return time series $R_k(t)$, one calculates the normalized return time series $M_k(t)$ which have zero mean and unit standard deviation. The Pearson correlation coefficient

$$C_{kl} = \frac{1}{T} \sum_{t=1}^{T} M_k(t) M_l(t) \tag{2.1}$$

is used to measure the linear mutual dependence of two stocks. The normalized time series $M_k(t)$ form the rows of the $K \times T$ rectangular data matrix M such that

$$C = \frac{1}{T} M M^\dagger \tag{2.2}$$

is the $K \times K$ correlation matrix. The dagger denotes the transpose.

Even though the Pearson correlation coefficients are successfully used in a large variety of applications, they have severe limitations. The linear correlation coefficient reduces a potentially much more complex statistical dependence to a single number. Strictly speaking, this is only meaningful, if the dependence is multivariate Gaussian, *e.g.*, bivariate

$$f_{kl}(x, y) = \frac{1}{2\pi\sqrt{1 - C_{kl}^2}} \exp\left(-\frac{1}{2} \frac{x^2 - 2C_{kl}xy + y^2}{1 - C_{kl}^2}\right) , \tag{2.3}$$

where we use $x = R_k(t),\ y = R_l(t)$ as a short-hand notation. If the dependence is not of this simple Gaussian form, one has to retrieve better information from the full joint pdf $f_{kl}(x,y)$. This can be done by using copulas as introduced by Sklar [11–13]. Here, I report a large scale analysis of financial data recently carried out by Münnix and Schäfer [8]. To define copulas, we need some basic tools and, in particular, the marginal distribution or probability density of the random variable x,

$$f_k(x) = \int\limits_{-\infty}^{+\infty} f_{kl}(x,y)dy \tag{2.4}$$

and accordingly for y, as well as the cumulative distribution

$$F_k(x) = \int\limits_{-\infty}^{x} f_k(x')dx' \tag{2.5}$$

which is the probability to find the random variable in the interval $(-\infty, x]$. Furthermore, we need the concept of quantiles. Inverting the cumulative distribution, we can say that left of $x = F_k^{-1}(u)$ are u percent of events. This is referred to as u quantile, while u itself is the probability. The joint probability is the double integral

$$F_{kl}(x,y) = \int\limits_{-\infty}^{x} dx' \int\limits_{-\infty}^{y} dy' f_{kl}(x',y')\,, \tag{2.6}$$

which still contains all information about the mutual dependencies.

The idea is now to separate marginal distributions from the statistical dependencies,

$$F_{kl}(x,y) = \mathrm{Cop}_{kl}\left(F_k(x), F_l(y)\right)\,, \tag{2.7}$$

$$\mathrm{Cop}_{kl}(u,v) = F_{kl}\left(F_k^{-1}(u), F_l^{-1}(v)\right)\,, \tag{2.8}$$

$$\mathrm{cop}_{kl}(u,v) = \frac{\partial^2}{\partial u \partial v}\mathrm{Cop}_{kl}(u,v)\,, \tag{2.9}$$

where Cop is the copula and cop the corresponding copula density. The mutual dependencies are measured as functions of the probabilities u and v. This allows one to compare joint probabilities or joint probability densities with different marginal distributions. This is similar to a "moving frame" or to "unfolding" in quantum chaos.

Münnix and Schäfer [8] thoroughly investigated how much the true, *i.e.*, the empirical copula density $\mathrm{cop}_{kl}(u,v)$ differs from the Gaussian copula density $\mathrm{cop}_{kl}^{(G)}(u,v)$ resulting from Eq. (2.3). They defined the distance

$$d(u,v) = \frac{1}{K(K-1)/2} \sum_{k<l} \left(\mathrm{cop}_{kl}(u,v) - \mathrm{cop}_{kl}^{(G)}(u,v) \right) \tag{2.10}$$

as an average over all pairs constructed from a set of K stock returns. They performed their empirical study for the U.S. stock market for $K = 428$ firms from the Standard & Poor's TAQ data set between 2007 and 2010, comprising more than 12 billion transactions. The results are shown in Fig. 1, exhibiting a substantial difference from the Gaussian case. Münnix and Schäfer [8] demonstrate that the bivariate Gaussian assumption drastically underestimates simultaneous extreme events. The structure remains unchanged even down to a return horizon of $\Delta t = 30$ min.

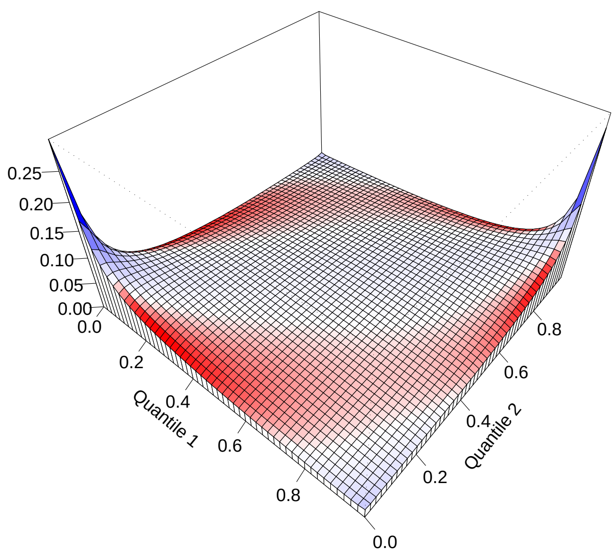

Fig. 1. Distance $d(u,v)$ of copula densities for $\Delta t = 2\,\mathrm{h}$. Adopted from Ref. [8].

3. Market states

The volatilities, *i.e.*, the standard deviations of the stock prices and the returns are non-stationary because they reflect the trading activity. In hectic times, the volatility is higher, and in calm times, lower. The correlations provide detailed structural information about the market, but the Pearson correlation coefficients are non-stationary, too. As the business relations between the firms as well as the traders' market expectations change, there is no reason for them to be constant. Thus, the structures visible in the time correlation matrices also evolve with time. In Ref. [9], we used this fact to identify states of the entire market characteristic for a finite but larger period in time. Importantly, this structural information is in a very good approximation the same when analysed with correlations or, alternatively, with bivariate copulas. As the correlations are computed much faster, we stick to them in the sequel.

To this end, we defined a similarity measure to quantitatively estimate how much the correlation matrix $C^{(T)}(t_1)$ extracted at time t_1 with sampling time T backwards differs from $C^{(T)}(t_2)$ extracted from the data at time t_2. The distance of two correlation matrices is

$$\zeta^{(T)}(t_1,t_2) = \frac{1}{K(K-1)/2}\sum_{k<l}\left|C_{kl}^{(T)}(t_1) - C_{kl}^{(T)}(t_2)\right| . \tag{3.1}$$

These distances $\zeta^{(T)}(t_1,t_2)$ yield an array or matrix in the (discretely chosen) points (t_1,t_2), not to be confused with the correlation matrix itself. Empirical results for the U.S. financial market from 1992 to 2010 are shown in Fig. 2. The darker, the more different are the correlation matrices and thus the correlation structure. Stripes emerge which mean that the correlation matrices at all later times differ from the one at a given time.

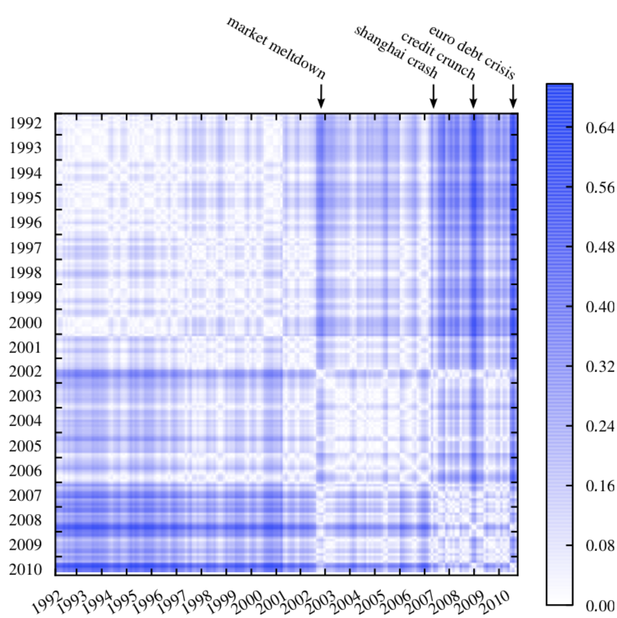

Fig. 2. Distance matrix $\zeta^{(T)}(t_1,t_2)$ for $1992 \le t_1,t_2 \le 2010$. Taken from Ref. [9].

Employing this distance measure, we identified market states by a cluster analysis [9]. All correlation matrices computed with a two-months backward sampling time T from 1992 to 2010 form the set $C^{(T)}(t),\ t = 1992, \ldots, 2010$. We divided it in two disjunct subsets where the distance $\zeta^{(T)}$ from the average within each set is smallest. In the same way, we further divided these two subsets in two subsubsets each, and so on. We stopped when the distances within the sub…subsets became comparable to the distances between the sub…subsets. We refer to these remaining sub…subsets as clusters. The average of the correlation matrices in each cluster grasps the typical correlation structure which we interpret as a quasi-stationary market state. We found eight such market states which are depicted in Fig. 3. The colour coding ranges from dark blue (highest correlation) over white (no correlation) to red (highest anticorrelation). Market state number 8 belongs to a period of crisis in which sheer panic makes the traders act in a highly correlated fashion.

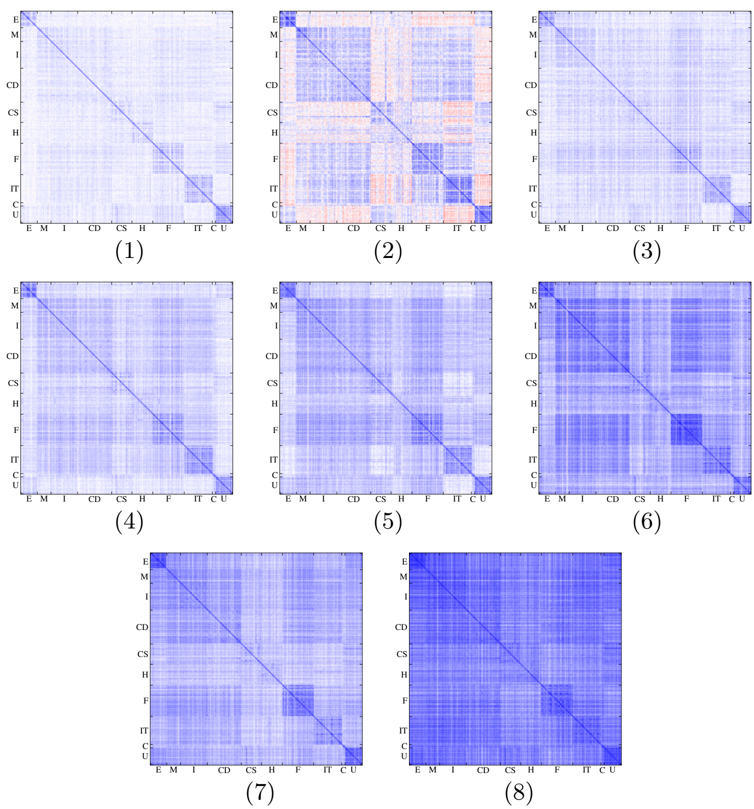

Fig. 3. (Colour online) The eight market states which emerged between 1992 and 2010. Taken from Ref. [9].

These states also provide rich information about the time evolution of the market as demonstrated in Fig. 4. The subsequent formation and disintegration of the market states is clearly visible. The market can return to a state and even jump back and forth between states. Eventually, each state disappears. Old states die out, new ones emerge. Thus, market states have a lifetime and it would be interesting to relate it to other time scales, *e.g.*, to the average time between crashes. Usually, stock market data are analysed "as if they were thrown on the floor", but the identification of the market states yields a tool to study the time evolution of the entire market in a much reduced parameter space.

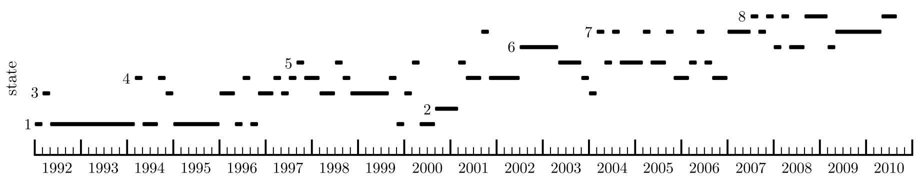

Fig. 4. Time evolution of the market states, numbering as in Fig. 3. Taken from Ref. [9].

4. Generic features

Although Fig. 3 features the averages in the eight clusters, it illustrates the significant non-stationarity of the correlations. In Ref. [10], we showed that the above mentioned set of all correlation matrices may be viewed as an ensemble allowing for ensemble averages of observables. Clearly, the strength of the fluctuations within this ensemble will be carried over to the ensemble averaged observable. A particularly interesting observable is the multivariate distribution of all K returns, ordered in the vector $R = (R_1(t), \ldots, R_K(t))$ at a given time t. We carefully checked that this distribution is well-approximated by the multivariate Gaussian distribution

$$g(R|\Sigma) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2} R^\dagger \Sigma^{-1} R\right) \tag{4.1}$$

under the condition that the backwards sampling time T is short enough to leave the covariances constant. We recall that the $K \times K$ covariance matrix $\Sigma = \sigma C \sigma$ contains the correlation matrix C and the diagonal matrix σ of the volatilities. The empirical multivariate return distribution for the Standard & Poor's data set is depicted in Fig. 5 for $T = 1$ a month after the rotation into the eigenbasis of Σ and subsequent aggregation. As one sees, the Gaussian yields a satisfactory description over several decades.

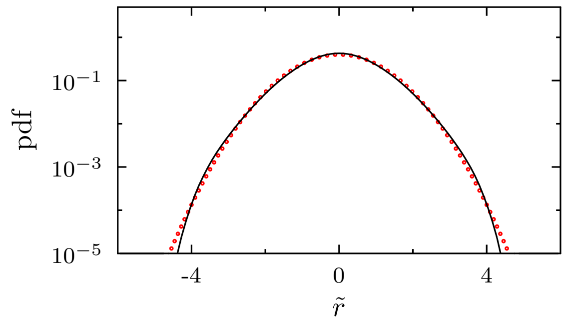

Fig. 5. Aggregated multivariate return distribution (solid line) for short sampling times compared to a Gaussian (circles). Taken from Ref. [10].

However, the available data set comprises 20 years, from 1992 to 2012. What is the multivariate distribution for all the returns measured in this long period in view of the fact that the covariances are non-stationary? We began with assuming that the ensemble of the correlation matrices can be modelled by a Wishart ensemble of random matrices $\sigma X X^\dagger \sigma$ where the $K \times N$ rectangular and real matrix X is drawn from the Gaussian distribution

$$w(X|C_0) = \frac{1}{\det^{N/2}(2\pi C_0)} \exp\left(-\frac{N}{2} \mathrm{tr}\, X^\dagger C_0^{-1} X\right) . \tag{4.2}$$

Here, C_0 is the empirically found mean correlation matrix in the entire period from 1992 to 2012. The Gaussian (4.2) ensures that C_0 is the resulting ensemble average, $\langle X X^\dagger \rangle = C_0$. The matrix X is the model data matrix. One of its dimensions, K, was dictated to us as we want to analyse returns of K firms. The other dimension, N, is a free parameter which can be interpreted as the length of the model time series. A first understanding of its meaning within the model is achieved by realizing that it is also contained in the exponent in Eq. (4.2) as an inverse variance. To obtain the multivariate distribution for the entire period from 1992 to 2012, we perform the ensemble average

$$\langle g \rangle (R|C_0, \sigma, N) = \int g\left(R|\sigma X X^\dagger \sigma\right) w(X|C_0) d[X] \tag{4.3}$$

over the empirically confirmed multivariate Gaussian distribution (4.2) for very short periods of time. The matrix integral can be computed in a closed form and yields

$$\langle g\rangle(R|C_0,\sigma,N) = \frac{1}{2^{N/2+1}\Gamma(N/2)\sqrt{\det(2\pi\Sigma_0/N)}} \times \frac{\mathcal{K}_{(K-N)/2}\left(\sqrt{NR^\dagger\Sigma_0^{-1}R}\right)}{\sqrt{NR^\dagger\Sigma_0^{-1}R}^{(K-N)/2}}\,, \tag{4.4}$$

where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind and of the order of ν. Due to the invariances of the Wishart ensemble (4.2), the result only depends on the bilinear form $R^\dagger\Sigma_0^{-1}R$ with $\Sigma_0 = \sigma C_0 \sigma$ being the mean covariance matrix.

We compared the result (4.4) with the return data from 1992 to 2012, again by rotating in the eigenbasis, now of Σ_0, and subsequent aggregation. We used daily and monthly returns. In both cases, the distribution clearly has much heavier tails than a Gaussian, but the tails for the daily data are even heavier than those for the monthly data. This simply reflects the heaviness for the *individual* return distribution which are the heavier, the smaller the return horizon Δt. The important observation, however, is that our analytical result (4.4) can grasp that for the *multivariate* return distribution over some decades. Here, the parameter N comes into play again. It crucially determines the width of the analytical result (4.4). We also showed that there is roughly a linear dependence between N and Δt. Thus, we find that the smaller N, the stronger are the correlations and the heavier are the tails. In Fig. 6, a fit gives $N = 5$ for the daily and $N = 14$ for the monthly data. The message is that the fluctuation of the correlations lift the Gaussian tails in the distribution (4.1). Furthermore, the ensemble approach drastically reduces the parameters needed to describe the mul-

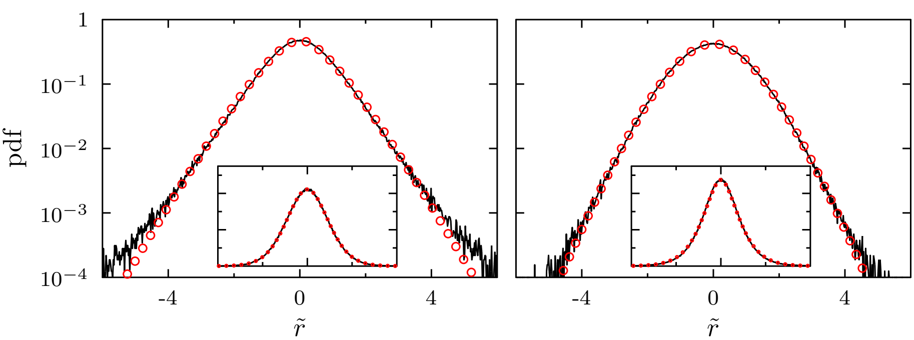

Fig. 6. Aggregated multivariate return distribution (solid line) for the 20 year sampling time from 1992 to 2012 compared to Eq. (4.4) (circles), daily (left) and monthly (right) returns. Taken from Ref. [10].

tivariate return distribution to the empirical mean correlation matrix, the empirical volatilities and the width parameter N which is the only free one. This implies that we have identified generic features of return distributions resulting from non-stationary correlations.

5. Conclusions

Financial correlation provides important information about the markets, but it was previously too little appreciated that the strongly non-Gaussian, heavy-tailed distributions observed for the returns of a single stock render it questionable to solely rely on the linear Pearson correlation coefficient when trying to quantitatively assess the mutual dependencies between different stocks. Copulas provide the full picture, demonstrating how drastically the correlations underestimate simultaneous large events. Having this caveat in mind, we may proceeded with the correlations to gather overall structural information.

The claimed coexistence of distinct market states on the one hand and of generic features on the other hand might sound contradictory. When Wigner put forward the Random Matrix Theory in the nuclear physics context, he was heavily criticized. It was rightfully pointed out that the selection rules lead to many zero matrix elements in the nuclear Hamiltonian when represented in a shell model basis, say. Nevertheless, the embedded ensembles of random matrices (see, *e.g.*, Ref. [14]) which take those selection rules into account yield the same *local* fluctuations as Wigner's original model. This is so because of the unfolding that removes the dependence on the level density. As there are no scales competing with the mean level spacing, no deviations from the standard random matrix fluctuations are seen.

A similar line of reasoning solves the above puzzle. The truly existing, empirically found ensemble of correlation matrices has a fine structure revealing itself in the market states which may be viewed as distinct attractors in the set of the correlation matrices. The considered *multivariate* distributions mix returns in linear combinations when rotated in the eigenbasis of the covariance matrix. This implies some self-averaging that washes out the information about the fine structure and the plain Gaussian Wishart model yields a good description of the data. Here, it is of crucial importance that the ensemble averaged return distribution that we found is, by no means, an artificial object, it is of direct relevance, *e.g.*, for portfolio optimization [15] and credit risk [16].

Finally, we mention the conceptual difference of our ensemble approach to the previous application of Random Matrix Theory in econophysics [3, 4]. In the latter studies, an ergodicity argument was implicitly used when comparing the eigenvalue density of a *single*, but large empirical correlation matrix to that of a Wishart *ensemble*. Put differently, the ensemble was fictitious. In our study, we identified a really existing ensemble that models the non-stationarity. The issue of ergodicity does not arise.

REFERENCES

[1] R.N. Mantegna, H.E. Stanley, *Introduction to Econophysics*, Cambridge University Press, Cambridge 1999.

[2] J.P. Bouchaud, M. Potters, *Theory of Financial Risk and Derivative Pricing*, 2nd ed., Cambridge University Press, Cambridge 2003.

[3] L. Laloux, P. Cizeau, J.P. Bouchaud, M. Potters, *Phys. Rev. Lett.* **83**, 1467 (1999).

[4] V. Plerou *et al.*, *Phys. Rev. Lett.* **83**, 1471 (1999).

[5] S. Pafka, I. Kondor, *Physica A* **319**, 487 (2003).

[6] T. Guhr, B. Kälber, *J. Phys. A* **36**, 3009 (2003).

[7] Z. Burda *et al.*, *Physica A* **343**, 694 (2004).

[8] M.C. Münnix, R. Schäfer, *Physica A* **390**, 4251 (2011).

[9] M.C. Münnix *et al.*, *Sci. Rep.* **2**, 644 (2012).

[10] T.A. Schmitt, D. Chetalova, R. Schäfer, T. Guhr, *Europhys. Lett.* **103**, 58003 (2013).

[11] A. Sklar, *Publications de l'Institut de Statistique de L'Université de Paris* **8**, 229 (1959).

[12] A. Sklar, *Kybernetika* **9**, 449 (1973).

[13] A. Sklar, *Random Variables, Distribution Functions, and Copulas — a Personal Look Backward and Forward*, in: L. Rüschendorf, B. Schweizer, M. Taylor (Eds.), *Distributions With Fixed Marginals & Related Topics*, Lecture Notes — Monograph Series, **28**, Hayward 1996.

[14] T.A. Brody *et al.*, *Rev. Mod. Phys.* **53**, 385 (1981).

[15] D. Chetalova, T.A. Schmitt, R. Schäfer, T. Guhr, `arXiv:1308.3961 [q-fin.ST]`, accepted for publication in the *Int. J. Theor. Appl. Finance*.

[16] T.A. Schmitt, D. Chetalova, R. Schäfer, T. Guhr, *Europhys. Lett.* **105**, 38004 (2014).