Exact multivariate amplitude distributions in correlated financial markets

Juan C. Henao-Londono[a,1], Anton J. Heckens[a,1], Thomas Guhr[a,1]

[a] *Fakultät für Physik, Universität Duisburg-Essen, Lotharstraße 1, Duisburg, 47048, NRW, Germany*

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.

Keywords: Econophysics, Complex systems, Statistical physics, Exact multivariate amplitude distributions, Correlated financial markets

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

*Corresponding author

Email addresses: juan.henao-londono@uni-due.de (Juan C. Henao-Londono), anton.heckens@uni-duisburg-essen.de (Anton J. Heckens), thomas.guhr@uni-due.de (Thomas Guhr)

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.

The paper is organized as follows: in Sect. 2 we present our data set of stocks. We define the key concepts, present the epochs analysis of the stock time series and discuss the method used by T. A. Schmitt et al. [14] in Sect. 3. In Sect. 4 we simulate returns time series to test the rotation and scaling method and check the influence of the epochs in the results. We define the general considerations to use the exact multivariate distributions in Sect. 5. In Sect. 6 we compare the theoretical distributions with the empirical aggregated distribution of returns. Our conclusions follow in Sect. 7.

2. Data set

In this study, we analyze stocks from different continuously traded companies in the S&P 500 stock market index. Almost for the whole work, we use daily adjusted closing prices of 200 companies from January 1, 1990 to December 31, 2020. All the data was obtained from Yahoo! Finance.

In the particular case of the non-stationarity example in Sect. 3.2 we use daily adjusted closing prices of 421 companies for the fourth quarter of 2005 and the first quarter of 2006. We do not use the 421 companies for the other analysis because half of this stocks' data is not available between the time intervals we wanted to do the analysis.

All the companies used in the paper are listed in Appendix A.

3. 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. Finally, in 3.3 we explain the methodology to see how the returns are to a good approximation multivariate Gaussian distributed or multivariate Algebraic distributed.

3.1. Key concepts

Consider time series $S_k(t)$, $k = 1, 2, \ldots, K$ 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

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

Anyway, logarithmic differences and returns almost coincide if the time steps Δt are sufficiently short [32, 33]

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

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]

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

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

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

where

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

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 [14]

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

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 [15, 35, 36, 37, 38]

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

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

I think this is maybe not true. Should we remove this paragraph? 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].

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 change of the $K \times K$ correlation matrix C as a whole in time is an example of non-stationary. Fig. 1 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. The stripes in these correlation matrices indicate the structuring of the market in industrial sectors [13, 14]

In Fig. 1 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. Choice of amplitude in each of the epochs

We use the same methodology from T. A. Schmitt et al. [14] to show that to a good approximation, the returns can be multivariate Gaussian distributed or multivariate algebraic distributed, depending of the epochs window length choice. All of this, if the covariance matrix Σ is fixed within the epochs.

We start from the assumption that the K dimensional vectors $r(t) = (r_1(t), \ldots, r_K(t))$ for a fixed return interval Δt is given by the multivariate Gaussian distribution

$$P_G\left(r|\Sigma_{ep}\right) = \frac{1}{\sqrt{\det 2\pi\Sigma_{ep}}} \exp\left(-\frac{1}{2} r^\dagger \Sigma_{ep}^{-1} r\right). \tag{11}$$

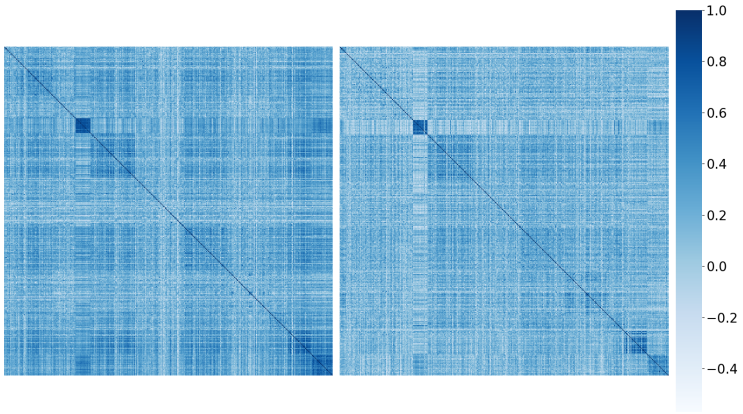

Figure 1: Correlation matrices of $K = 421$ companies for the fourth quarter of 2005 (left) and the first quarter of 2006 (right), the darker, the stronger the correlation. The companies are sorted according to industrial sectors.

To test the assumption, we divide the time series in windows of length T where the sampled covariances can be viewed as constant within these windows.

To carry out the data analysis, whether for the multivariate Gaussian distribution or the multivariate Algebraic distribution, we choose all pairs of returns, normalize the epochs (mean $\mu = 0$ and variance $\sigma^2 = 1$) and create two-component vectors $(r_k(t), r_l(t))$. For each pair, we evaluate the 2×2 sample covariance matrix $\Sigma^{(k,l)}$. Then we diagonalize the covariance matrix as $\Sigma = U\Lambda U^\dagger$, such that $\Sigma^{-1/2} = U\Lambda^{-1/2}U^\dagger$, where U is an orthogonal $K \times K$ matrix and Λ is the diagonal matrix of the eigenvalues Λ_K. We rotate the two-component returns vectors into the eigenbasis of $\Sigma^{(k,l)}$,

$$(\tilde{r}_1(t), \tilde{r}_2(t)) = U(r_k(t), r_l(t)) \tag{12}$$

and normalize the axis with the eigenvalues as

$$\frac{\tilde{r}_1(t)}{\sqrt{\Lambda_1}} \text{ and } \frac{\tilde{r}_2(t)}{\sqrt{\Lambda_2}}. \tag{13}$$

Finally, we aggregate all the components into a single univariate distribution, as all of them are now comparable.

Now we need to find the one dimensional function related to the multivariate Gaussian distribution in Eq. 11 to compare with the empirical single univariate distributions. In the Gaussian case, we know that

$$P_G^{(ij)}\left(r^{(ij)}|\Sigma_{ep}^{(ij)}\right) = \int d[r]_{\neq i,j} P_G\left(r|\Sigma_{ep}\right) \tag{14}$$

$$= \frac{1}{\sqrt{\det 2\pi\Sigma_{ep}^{(ij)}}} \exp\left(-\frac{1}{2} r^{(ij)\dagger} \Sigma_{ep}^{(ij)^{-1}} r^{(ij)}\right) \tag{15}$$

is the corresponding bivariate model where

$$r^{(ij)} = \begin{bmatrix} r_i \\ r_j \end{bmatrix} \text{ and } \Sigma_{ep}^{(ij)} = \begin{bmatrix} \Sigma_{ep,ii} & \Sigma_{ep,ij} \\ \Sigma_{ep,ji} & \Sigma_{ep,jj} \end{bmatrix} \tag{16}$$

are the two-component return vector and the 2×2 covariance matrix. $\Sigma_{ep,ij}$ is just the (ij) element of Σ_{ep}. Diagonalising

$$\Sigma_{ep}^{(ij)} = U^{(ij)} \Lambda_{ep}^{(ij)} U^{(ij)\dagger} \tag{17}$$

with the 2×2 rotation matrix $U^{(ij)}$ and the eigenvalue matrix

$$\Lambda_{ep}^{(ij)} = \text{diag}\left(\Lambda_{ep,1}^{(ij)}, \Lambda_{ep,2}^{(ij)}\right). \tag{18}$$

We rotate into the eigenbasis

$$\bar{r}^{(ij)} = \begin{bmatrix} \bar{r}_1^{(ij)} \\ \bar{r}_2^{(ij)} \end{bmatrix} = U^{(ij)\dagger} r^{(ij)} \tag{19}$$

such that

$$P_G^{(ij)}\left(r^{(ij)}|\Sigma_{ep}^{(ij)}\right) = P_G^{(ij)}\left(\bar{r}^{(ij)}|\Lambda_{ep}^{(ij)}\right) \tag{20}$$

$$= \frac{1}{\sqrt{2\pi\Lambda_{ep,1}^{(ij)}}} \exp\left(-\frac{\bar{r}_1^{(ij)^2}}{2\Lambda_{ep,1}^{(ij)}}\right) \frac{1}{\sqrt{2\pi\Lambda_{ep,2}^{(ij)}}} \exp\left(-\frac{\bar{r}_2^{(ij)^2}}{2\Lambda_{ep,2}^{(ij)}}\right). \tag{21}$$

The univariate rotated distribution is

$$\tilde{P}_G^{(ij)}(\tilde{r}) = \int_{-\infty}^{+\infty} d\bar{r}_2^{(ij)} P_G^{(ij)}\left(\bar{r}^{(ij)}|\Lambda_{ep}^{(ij)}\right)|_{\bar{r}_1^{(ij)}/\sqrt{\Lambda_{ep,1}^{(ij)}}=\tilde{r}} \frac{d\bar{r}_1^{(ij)}}{d\tilde{r}} \tag{22}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\tilde{r}^2\right), \tag{23}$$

which is the standard Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$.

To check the validity of our assumption, in Fig. 2 we plot the daily adjusted closing aggregated returns distribution of 200 companies from the S&P 500 dataset, with a $\Delta t = 1d$ and an epoch window length $T = 25d$. For the univariate distribution case we observe a good agreement with the one dimensional Gaussian distribution, as expected.

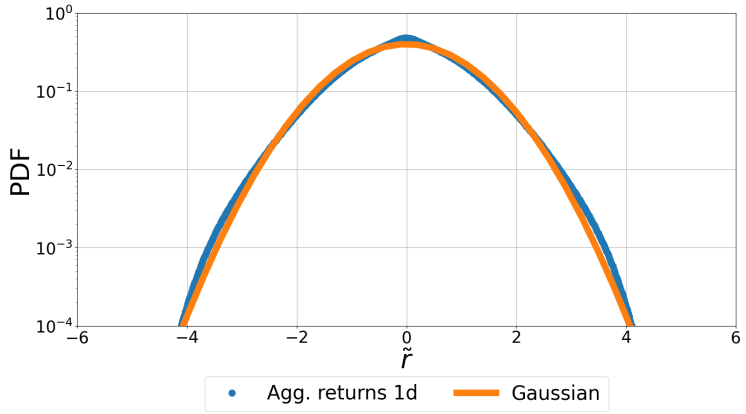

Figure 2: Aggregated distribution of returns ($\tilde{r}$) for fixed covariance of 200 companies selected from the S&P 500 dataset. $\Delta t = 1d$ and epoch window length $T = 25d$.

To model heavy tails, we assume that the K dimensional vectors $r(t) = (r_1(t), \ldots, r_K(t))$ for a fixed return interval Δt is given by the multivariate algebraic distribution

$$P_A\left(r|\Sigma_{ep}\right) = \frac{\alpha_{K1lm}}{\left(1 + \frac{1}{m} r^\dagger \Sigma_{ep}^{-1} r\right)^l} \tag{24}$$

$$\alpha_{K1lm} = \sqrt{\frac{2}{m}}^K \frac{\Gamma(l)}{\Gamma(l - K/2)} \frac{1}{\sqrt{\det 2\pi\Sigma_{ep}}}. \tag{25}$$

To test the new assumption and compare the different time step data, we repeat the same method to rotate and scale the returns to then get the aggregate distribution of returns. Once more, we need to find the corresponding one dimensional functions related to the multivariate algebraic distribution in Eq. 24 to compare with the empirical single univariate distributions.

We know that the algebraic model from Eq. 24 relates to the Gaussian model from Eq. 11 via

$$P_A\left(r|\Sigma_{ep}\right) = \int_0^\infty \chi^2_{2(l-K/2)}\left(z\right)\; P_G\left(r|\frac{m}{2}\Sigma_{ep}\right)dz. \tag{26}$$

We immediately find

$$P_A^{(ij)}\left(r^{(ij)}|\Sigma_{ep}^{(ij)}\right) = \int d\left[r\right]_{\neq i,j}\, P_A\left(r|\Sigma_{ep}\right) \tag{27}$$

$$= \int_0^\infty dz \chi^2_{2(l-K/2)}\left(z\right) \frac{1}{\sqrt{\det 2\pi\frac{m}{2}\Sigma_{ep}^{(ij)}}} \exp\left(-\frac{z}{2m} r^{(ij)\dagger}{\Sigma_{ep}^{(ij)}}^{-1} r^{(ij)}\right) \tag{28}$$

$$= \frac{1}{\sqrt{\det 2\pi\Sigma_{ep}^{(ij)}}} \int_0^\infty dz \chi^2_{2(l-K/2)}\left(z\right) \frac{z}{m} \exp\left(-\frac{z}{2m} r^{(ij)\dagger}{\Sigma_{ep}^{(ij)}}^{-1} r^{(ij)}\right). \tag{29}$$

Using

$$\chi^2_{q+2}\left(z\right) = \frac{1}{q}\chi^2_q\left(z\right) z \tag{30}$$

we have

$$P_A^{(ij)}\left(r^{(ij)}|\Sigma_{ep}^{(ij)}\right) = \frac{2\left(l-K/2\right)}{m\sqrt{\det 2\pi\Sigma_{ep}^{(ij)}}} \int_0^\infty dz \chi^2_{2(l+1-K/2)}\left(z\right) \exp\left(-\frac{z}{2m} r^{(ij)\dagger}{\Sigma_{ep}^{(ij)}}^{-1} r^{(ij)}\right) \tag{31}$$

$$= \frac{2\left(l-K/2\right)}{m\sqrt{\det 2\pi\Sigma_{ep}^{(ij)}}} \frac{1}{2^{l+1-K/2}\Gamma\left(l+1-K/2\right)} \int_0^\infty dz\, z^{l-K/2} \exp\left[-\frac{z}{2}\left(1+\frac{1}{m} r^{(ij)\dagger}{\Sigma_{ep}^{(ij)}}^{-1} r^{(ij)}\right)\right] \tag{32}$$

$$= \frac{2l-K}{m\sqrt{\det 2\pi\Sigma_{ep}^{(ij)}}} \frac{1}{\left(1+\frac{1}{m} r^{(ij)\dagger}{\Sigma_{ep}^{(ij)}}^{-1} r^{(ij)}\right)^{l+1-K/2}}. \tag{33}$$

Rotating into the eigenbasis, we can integrate out one of the variables

$$\tilde{P}_A^{(ij)}\left(\bar{r}_1^{(ij)}|\Lambda_{ep,1}^{(ij)}\right) = \int_{-\infty}^{+\infty} d\bar{r}_2^{(ij)} P_A^{(ij)}\left(\bar{r}^{(ij)}|\Lambda_{ep}^{(ij)}\right). \tag{34}$$

Using Eq. 33, we have

$$\tilde{P}_A^{(ij)}\left(\bar{r}_1^{(ij)}|\Lambda_{ep,1}^{(ij)}\right) = \frac{2\left(l-K/2\right)}{\sqrt{m}\sqrt{2\pi\Lambda_{ep,1}^{(ij)}}} \frac{1}{2^{l+1-K/2}\Gamma\left(l+1-K/2\right)} \int_0^\infty dz\, z^{l-K/2-1/2} \exp\left[-\frac{z}{2}\left(1+\frac{1}{m}\frac{\bar{r}_1^{(ij)^2}}{\Lambda_{ep,1}^{(ij)}}\right)\right] \tag{35}$$

$$= \frac{2\left(l-K/2\right)}{\sqrt{m}\sqrt{2\pi\Lambda_{ep,1}^{(ij)}}} \frac{2^{l+1/2-K/2}\Gamma\left(l+1/2-K/2\right)}{2^{l+1-K/2}\Gamma\left(l+1-K/2\right)} \frac{1}{\left(1+\frac{1}{m}\frac{\bar{r}_1^{(ij)^2}}{\Lambda_{ep,1}^{(ij)}}\right)^{l-(K-1)/2}} \tag{36}$$

$$= \sqrt{\frac{2}{m}} \frac{\Gamma\left(l-\left(K-1\right)/2\right)}{\Gamma\left(l-K/2\right)} \frac{1}{\sqrt{2\pi\Lambda_{ep,1}^{(ij)}}} \frac{1}{\left(1+\frac{1}{m}\frac{\bar{r}_1^{(ij)^2}}{\Lambda_{ep,1}^{(ij)}}\right)^{l-(K-1)/2}}. \tag{37}$$

Going over to the rescaled variable

$$\tilde{r} = \frac{\bar{r}^{(ij)}}{\sqrt{\Lambda^{(ij)}_{ep,1}}} \tag{38}$$

we arrive at

$$\tilde{P}^{(ij)}_{A}(\tilde{r}) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2}{m}} \frac{\Gamma\left(l-(K-1)/2\right)}{\Gamma\left(l-K/2\right)} \frac{1}{\left(1+\frac{1}{m}\tilde{r}^2\right)^{l-(K-1)/2}} \tag{39}$$

which is the algebraic distribution to be fitted to the aggregated returns. In this case we have three parameters: K, l and m. K is the number of companies used and is known from the data. l and m are the shape parameters of the distribution and have to be determined by fitting. In the model, the expectation value $\langle rr^{\dagger}\rangle$ serves as an estimator for the sample covariances. We find

$$\langle rr^{\dagger}\rangle_Y = \beta_Y \Sigma_{ep} \tag{40}$$

with

$$\beta_Y = \begin{cases} 1, & \text{if } Y = G \\ \frac{m}{2l-K-2}, & \text{if } Y = A \end{cases}. \tag{41}$$

Due to its very definition, we have $\beta_G = 1$ for the multivariate Gaussian, but a different value β_A for the algebraic distribution. The relation in the latter case then suggest the useful fixing

$$m = 2l - K - 2 \tag{42}$$

also for finite values of l and m. Only with this choice, Σ_{ep} can be estimated by the sample covariance matrix, otherwise only up to some factor.

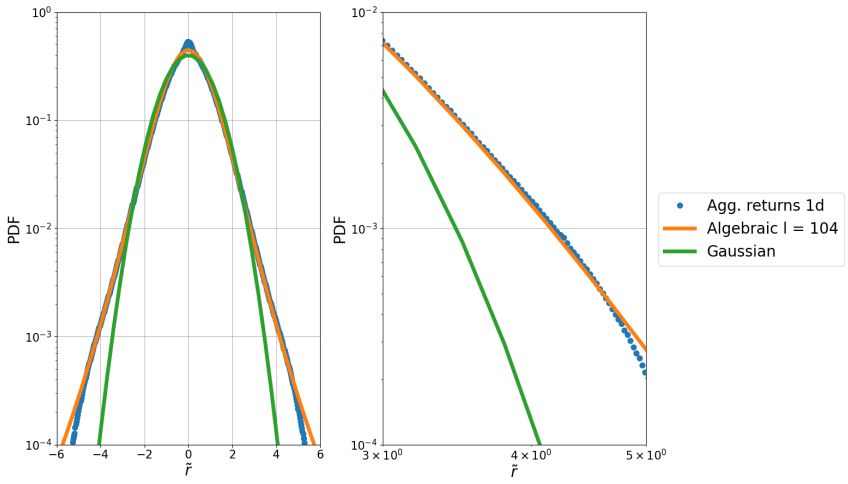

Figure 3: Semilog plot (left) and loglog plot (right) of the aggregated distribution of returns ($\tilde{r}$) for fixed covariance of 200 companies selected from the S&P 500 dataset. $\Delta t = 1d$ and epoch window length $T = 55d$. The empirical data is compared with the best fit of one dimensional algebraic distribution with shape parameter $l = 104$.

Having these details in mind and to check the validity of our assumption, in Fig. 3 we plot the daily adjusted closing aggregated returns distribution of 200 companies from the S&P 500 dataset, with a $\Delta t = 1d$ and an epoch window length $T = 55d$. Using Eq. 42 to compute m, we found that for one day, $l = 104$ fits well with the aggregated returns. We plot the one dimensional Gaussian distribution as a reference.

We know that the aggregated returns have internal Gaussian and algebraic structures depending on the epochs window length. The larger the value of T, the heavier the tails and the smaller the stationary assumption within the epochs. However, during the epochs analysis we found an unusual behavior when we used small epochs window lengths.

Using the methodology described in the work of Schmitt et al. [14], we wanted to find which was the best value of the epochs window length to fit the Gaussian and algebraic case. We used epochs window lengths $T = 10, 25, 40, 55$ days and $\Delta t = 1d$. In Fig. 4 can be seen the difference between epochs window lengths of the aggregated distribution of returns for fixed covariance of different amount of random companies selected from the S&P 500 dataset. The first interesting characteristic, is that with few companies is enough information to obtain the general behavior. This, because we are selecting pairs of stocks from the dataset, such that the order of the selected stocks does not matter and without repetition, i.e. a combination. Thus, even with a small amount of companies, the results are statistically relevant. Then, We choose random numbers of companies within our dataset to show this characteristic. We can see how for $T = 25d$ (top left) there is a good agreement with the Gaussian distribution, and how for values larger than $T = 25d$, as $T = 40d$ (bottom left) and $T = 55d$ (bottom right), the aggregated returns have fat tails. However, for $T = 10d$ (top left), the aggregated returns have a platykurtic behavior. This particular behavior is unusual and makes us wonder if somehow there is an artifact in the methodology we are using, that is affecting the other epochs window lengths as well.

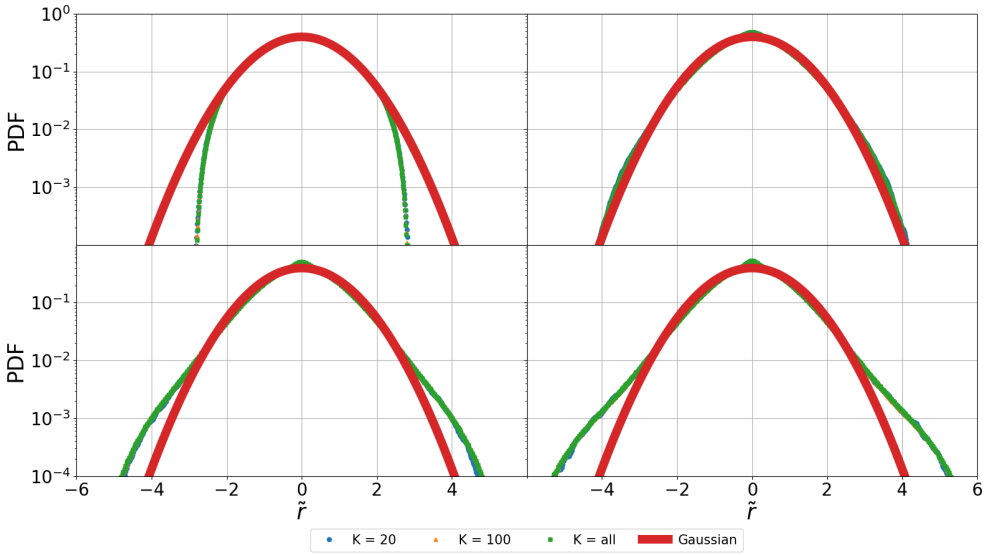

Figure 4: Aggregated distribution of returns ($\tilde{r}$) for fixed covariance of different number of companies selected from the S&P 500 dataset. $\Delta t = 1d$ and different epochs window lengths $T = 10d$ (top left), $T = 25d$ (top right), $T = 40d$ (bottom left) and $T = 55d$ (bottom right).

We consider two different causes for this situation in the small epochs window lengths. First, it could be an ergodicity defect because only the length of the time window of the epochs show an effect. Second, the platykurtic behavior like in local normalization plots [39] could be a detrending method. It is possible that we get this behavior from normalizing each epoch to mean $\mu = 0$ and $\sigma^2 = 1$.

To check what is the real cause of this behavior we will use simulations. With these simulations we know which parameters we are using, and we have a controlled environment. Thus, we can evaluate where is the problem with the method.

4. Returns time series simulations

In Sect. 4.1 we simulate returns time series following a multivariate Gaussian and algebraic distribution. We test the influence of the normalization within the epochs in the ergodicity defect in Sect. 4.2 and propose a solution for this defect in Sect. 4.3. Finally we use empirical data to support our simulation findings in Sect. 4.4.

4.1. Simulation of multivariate Gaussian and algebraic distributions

In Sect. 4.1.1 we describe the methodology to simulate multivariate Gaussian distributed returns time series and in Sect. 4.1.2 we describe the methodology to simulate multivariate algebraic distributed returns time series.

4.1.1. Multivariate Gaussian distributions

To simulate the returns time series, we use a method [40] for drawing a random vector x from the N-dimensional multivariate Gaussian distribution with mean vector μ and covariance matrix C. First, we create a correlation matrix C with $c = 1$ on its diagonal and $c = 0.3$ on its non-diagonal entries. Then, we compute the eigenvalues and eigenvectors of the correlation matrix, such that $C = U\Lambda U^{-1}$. We get a z vector whose components are drawn from a independent standard Gaussian distribution. Finally we obtain the returns with the desired distribution as

$$r = \mu + U\Lambda^{1/2}z \tag{43}$$

In our case, the r vector components are drawn from a normal distribution with correlation matrix C and μ vector zero. With these method we want to obtain time series simulating the data matrix G with dimensions $2 \times T$, where T is the window length of the epochs. These returns can be later normalized, rotated and aggregated to compare with the behavior of the results in Sect. 3.3. The goal of this approach is that all simulations should show standard normal distributions.

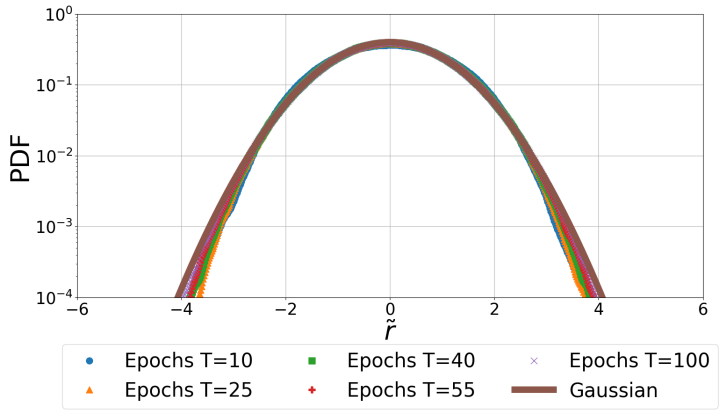

Figure 5: Simulated aggregated rotated and scaled Gaussian distributed returns ($\tilde{r}$) for fixed covariance and $K = 200$ without normalization, neither within the epochs nor for all the time series. $\Delta t = 1$ unit and epochs window lengths $T = 10, 25, 40, 55, 100$ units.

In Fig. 5 we simulate time series for $K = 200$. Each time series is made of 200 epochs to make them comparable to the empirical data. We use epochs window lengths $T = 10, 25, 40, 55$ units to rotate, scale and aggregate without normalizing neither within the epochs nor the full time series. As expected, as we draw the returns from a multivariate Gaussian distribution with correlation matrix C and μ vector zero, all the simulations show standard Gaussian distribution behavior. Thus, this is our reference to check what is introducing the ergodicity effect in the original method for the multivariate Gaussian case.

4.1.2. Multivariate algebraic distributions

To simulate returns time series drawn from multivariate algebraic distributions, we use a similar approach as in Sect. 4.1.1. First, we create a correlation matrix C with $c = 1$ on its diagonal and $c = 0.3$ on its non-diagonal entries. From [41] we know that

$$T = \left(S^{-1/2}\right)^{\dagger} X + M, \tag{44}$$

where T is a vector of length K. Then it is needed to repeat the following steps to generate a data matrix where the columns are the T vectors. X is drawn from a matrix variate normal distribution. Matrix S is a Wishart distributed covariance matrix without normalization and M is a parameter that for this case is zero. With these in mind, we first generate S as a positive semi-definite matrix. To do this, we first create time series of simulated data matrix G with dimension $K \times (n + K - 1)$ where n is the a parameter of the degree of freedom and is connected to the shape parameter l as

$$l = \frac{n + K}{2}, \tag{45}$$

where K is the number of companies. These time series are generated by calculating

$$y = U\Lambda^{1/2}z \tag{46}$$

with a fixed covariance matrix Σ. Vector y is a column vector of G, z is a univariate standard normal distribution vector, U has the eigenvectors of Σ as columns and the diagonal matrix Λ contains the eigenvalues of Σ. Thus, we compute

$$S = GG^{\dagger} \tag{47}$$

Then, we obtain $S^{1/2}$ as

$$S^{1/2} = U_S \Lambda_S^{1/2} U_S^{\dagger}, \tag{48}$$

since

$$\begin{aligned} S &= S^{1/2} S^{1/2} && (49)\\ &= U_S \Lambda_S^{1/2} U_S^{\dagger} U_S \Lambda_S^{1/2} U_S^{\dagger} && (50)\\ &= U_S \Lambda_S U_S^{\dagger} && (51) \end{aligned}$$

We generate X as

$$X = \sqrt{m}z \tag{52}$$

where z is a univariate standard normal distribution vector of length K and m is the variance.

These returns can be later normalized, rotated and aggregated to compare with the behavior of the results in Sect. 3.3. The goal of this approach is that all simulations should show standard algebraic distributions.

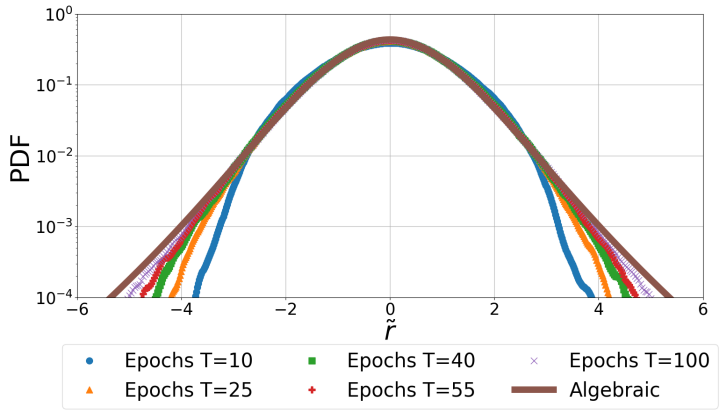

Figure 6: Simulated aggregated rotated and scaled algebraic distributed returns ($\tilde{r}$) for fixed covariance and $K = 200$ without normalization, neither within the epochs nor for all the time series. $\Delta t = 1$ unit and epochs window lengths $T = 10, 25, 40, 55, 100$ units.

In Fig. 6 we simulate time series for $K = 200$. Each time series is made of 200 epochs to make them comparable to the empirical data. We use epochs window lengths $T = 10, 25, 40, 55$ units to rotate, scale and aggregate without normalizing neither within the epochs nor the full time series. We can see an interesting behavior in the algebraic case, where for small epochs window lengths, the tails are similar to the Gaussian distribution, and as the epochs window lengths grow, the simulations reveals good agreement with the algebraic distribution. Thus, this is our reference to check what is introducing the ergodicity effect in the original method for the multivariate algebraic case.

4.2. Normalization within the epochs

Now, to check the normalization within the epochs, we first simulate the returns and then normalize each epoch to mean $\mu = 0$ and variance $\sigma^2 = 1$. Finally we repeat the procedure of rotate, scale and aggregate.

For the Gaussian case with the simulated pair returns time series, we proceed to normalize the epoch, compute the 2×2 sample covariance matrix and diagonalize it. We rotate the two-component returns vectors into the eigenbasis

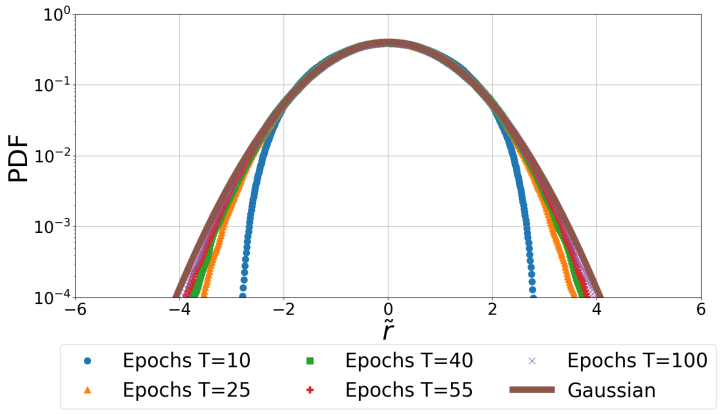

Figure 7: Simulated aggregated rotated and scaled Gaussian distributed returns ($\tilde{r}$) for fixed covariance and $K = 200$ with normalization within the epochs. $\Delta t = 1$ unit and epochs window lengths $T = 10, 25, 40, 55, 100$ units.

of the covariance matrix and normalize the axis with the eigenvalues. Finally we aggregate all the components into a single univariate distribution.

As we can see in Fig. 7, the ergodicity defect clearly appears for an epoch window length $T = 10$. As the epoch window length grows, the ergodicity defect starts to disappear. We could even argue, that with an epoch window length greater or equal to $T = 25$, we already are close enough to the Gaussian distribution. Thus, this ergodicity defect only affects small epochs window lengths. Furthermore, with large epoch windows lengths we can confirm that the ergodicity defect disappear, as it can be seen with $T = 100$

Something similar happens with the algebraic case. We again simulate the returns, normalize each epoch to mean $\mu = 0$ and variance $\sigma^2 = 1$ and finally repeat the procedure of rotate, scale and aggregate. In Fig. 8 can be seen how the ergodicity effect appears again. It seems to have a large effect in epochs window lengths with small values. Thus, we need to find an alternative to solve this problem.

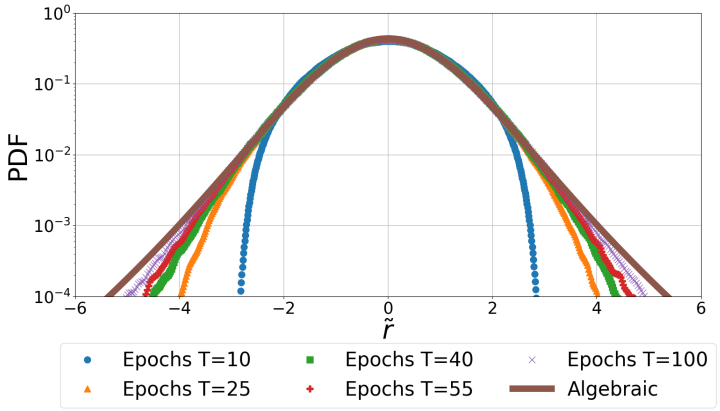

Figure 8: Simulated aggregated rotated and scaled Gaussian distributed returns ($\tilde{r}$) for fixed covariance and $K = 200$ with normalization within the epochs. $\Delta t = 1$ unit and epochs window lengths $T = 10, 25, 40, 55, 100$ units.

4.3. Normalization complete return time series

We already showed that the ergodicity effect is directly related with the normalization of the time series. To try to solve this issue, instead of normalizing the time series within each epoch, we normalize the complete time series and then proceed to rotate, scale and aggregate.

In Fig. 9 and Fig. 10 are shown the Gaussian and algebraic cases with the corresponding simulations. We can notice how in both cases the ergodicity defect disappears. In each epochs window length, the probability density function perfectly fits the Gaussian distribution and the algebraic distribution. It can be noted that the larger the epochs window length, the better the simulated data fit the distribution. However, to accomplish our assumption of stationarity on short time scales, the results for short epochs window lengths have a good agreement with the theoretical distribution.

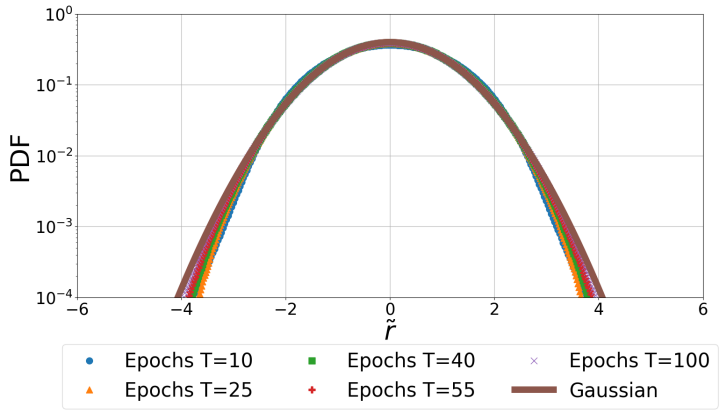

Figure 9: Simulated aggregated rotated and scaled Gaussian distributed returns ($\tilde{r}$) for fixed covariance and $K = 200$ with normalization for the complete time series. $\Delta t = 1$ unit and epochs window lengths $T = 10, 25, 40, 55, 100$ units.

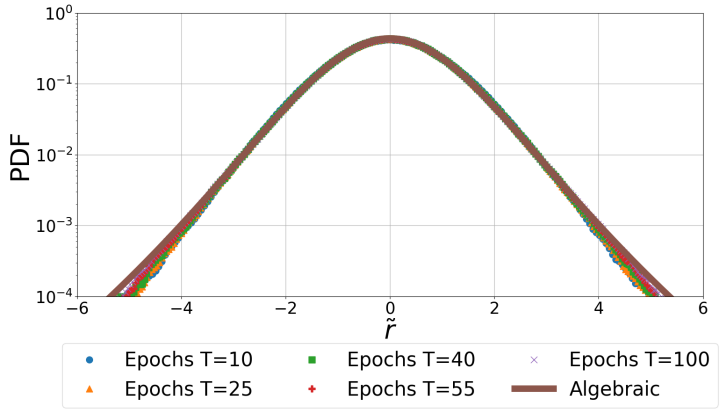

Figure 10: Simulated aggregated rotated and scaled algebraic distributed returns ($\tilde{r}$) for fixed covariance and $K = 200$ with normalization for the complete time series. $\Delta t = 1$ unit and epochs window lengths $T = 10, 25, 40, 55, 100$ units.

4.4. Empirical results

After we found the problem in the method and solved for the simulated time series, it is time to check the solution in the empirical data. We first normalize the complete time series. Then we rotate and scale the returns and finally we aggregate them. With this method, we can see how the ergodicity defect disappears, as we expected from the simulations. Furthermore, we can now see that with a epoch window length of $T = 25$ the returns show small fat tails. Then, we confirm that the aggregated returns have internal structures that we will use to define the exact multivariate amplitude distributions in Sect. 5.

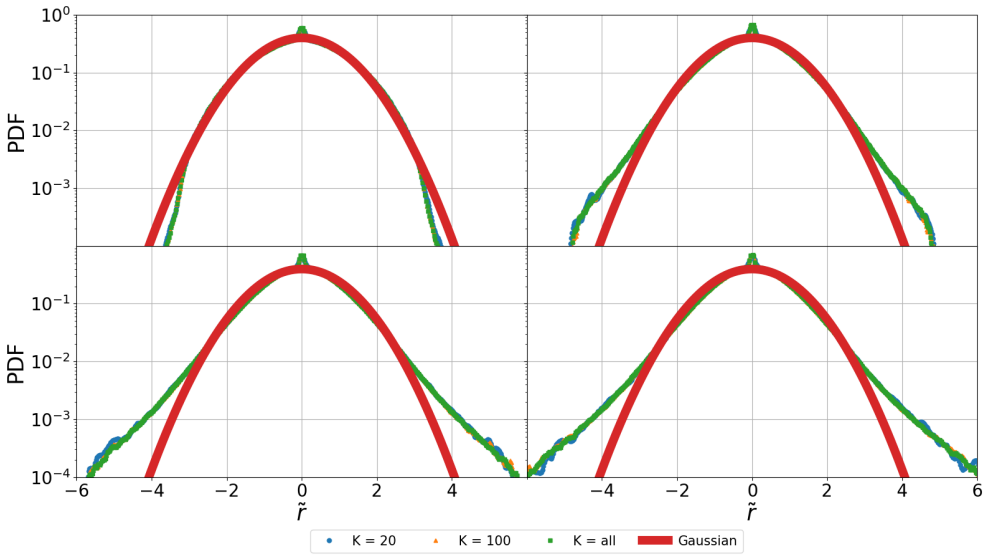

Figure 11: Aggregated distribution of returns ($\tilde{r}$) for fixed covariance of different number of companies selected from the S&P 500 dataset. $\Delta t = 1d$ and different epochs window lengths $T = 10d$ (top left), $T = 25d$ (top right), $T = 40d$ (bottom left) and $T = 55d$ (bottom right).

5. Exact multivariate amplitude distributions

In Sect. 5.1 we set the theoretical considerations to construct the distributions. In Sect. 5.2 we define the four distribution cases and in Sect. 5.3 we show their graphical representation.

5.1. 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 eigenbasis of the covariance matrix Σ [14, 15]. More precisely, we use the diagonalization

$$\Sigma = U\Lambda U^{\dagger} \text{ such that } \Sigma^{-1/2} = U\Lambda^{-1/2}U^{\dagger}, \tag{53}$$

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

$$\tilde{r} = U^{\dagger} r \tag{54}$$

as new arguments of the ensemble averaged amplitude distribution.

When analyzing data, K is given, we obtain the matrix 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 strength of the fluctuations. Experience tells, that N sensitively determines the shape and is best obtained by fitting the whole distribution to the data. For three cases, the parameters L and l are shape parameters.

5.2. Four cases distributions

As the aim of this paper is to compare the proposed distributions with empirical data, we present the final form of the corresponding distributions. For a detailed and complete explanation of the process to obtain the distributions, we suggest to review the work of Guhr and Schell [15]

[AQUÍ] For the following cases, In the Gaussian-Gaussian case with the Markovian situation $D = \mathbb{1}_N$ the distribution reads

$$\langle p\rangle_{GG}^{(k)}(\tilde{r}_k|\Lambda_k,\mathbb{1}_N) = \frac{1}{2^{(N-1)/2}\Gamma(N/2)\,\sqrt{\pi\Lambda_k/N}} \sqrt{\frac{N\tilde{r}_k^2}{\Lambda_k}}^{(N-1)/2} \mathrm{K}_{(1-N)/2}\left(\sqrt{\frac{N\tilde{r}_k^2}{\Lambda_k}}\right), \tag{55}$$

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

$$\langle p\rangle_{GA}^{(k)}(\tilde{r}_k|\Lambda_k,\mathbb{1}_N) = \frac{\Gamma(L-(K+N)/2+1)\,\Gamma(L-(K-1)/2)}{\Gamma(L-(K+N-1)/2)\,\Gamma(N/2)\,\sqrt{2\pi\Lambda_k M/N}} \mathrm{U}\left(L-\frac{K+N}{2}+1, \frac{1-N}{2}+1, \frac{N}{2M}\frac{\tilde{r}_k^2}{\Lambda_k}\right) \tag{56}$$

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

$$\langle p\rangle_{AG}^{(k)}(\tilde{r}_k|\Lambda_k,\mathbb{1}_N) = \frac{\Gamma(l-(K-1)/2)\,\Gamma(l-(K-N)/2)}{\Gamma(l-K/2)\,\Gamma(N/2)\,\sqrt{2\pi\Lambda_k m/N}} \mathrm{U}\left(l-\frac{K-1}{2}, \frac{1-N}{2}+1, \frac{N}{2m}\frac{\tilde{r}_k^2}{\Lambda_k}\right) \tag{57}$$

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

$$\langle p\rangle_{AA}^{(k)}(\tilde{r}_k|\Lambda_k,\mathbb{1}_N) = \frac{\Gamma(l-(K-1)/2)\,\Gamma(l-(K-N)/2)}{\Gamma(l-K/2)\,\Gamma(L+l-(K-1))\,\sqrt{\pi\Lambda_k Mm/N}} \frac{\Gamma(L-(K+N)/2+1)\,\Gamma(L-(K-1)/2)}{\Gamma(L-(K+N-1)/2)\,\Gamma(N/2)} {}_2\mathrm{F}_1\left(l-\frac{K-1}{2}, L-\frac{K+N}{2}+1, L+l-(K-1), 1-\frac{N}{Mm}\frac{\tilde{r}_k^2}{\Lambda_k}\right) \tag{58}$$

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. 12 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.

5.3. 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 $\langle p\rangle_{YY'}(k)(\tilde{r}_k|D)$ then allows us to normalize the rotated amplitude $\tilde{r}_k$ by the standard deviation

$$\tilde{r} = \frac{\tilde{r}_k}{\sqrt{\Lambda_k}} \tag{59}$$

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

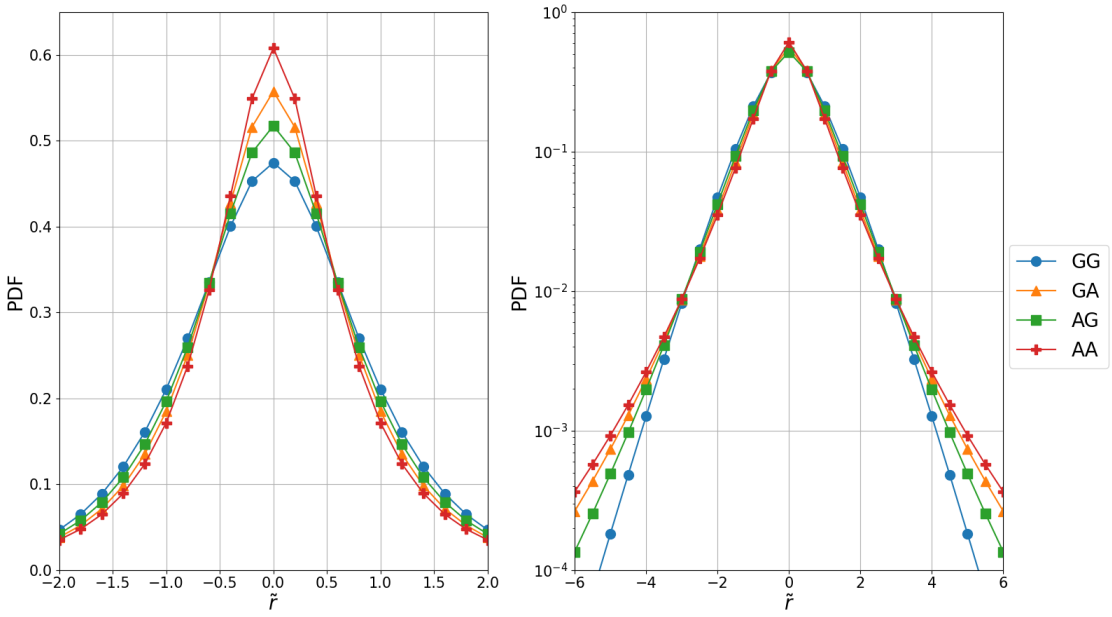

Figure 12: Probability densities $\langle p \rangle^{(k)}_{YY'}$, in the Markovian case versus the rotated amplitudes $\tilde{r}$, normalized to unit standard deviation. The four cases Gaussian-Gaussian, Gaussian-Algebraic, Algebraic-Gaussian and Algebraic-Algebraic are labeled $YY' = GG$, GA, AG and 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 $+8$.

6. Comparison with empirical portfolio returns

6.1. Response functions on trade time scale

6.2. Response functions on physical time scale

7. Conclusion

8. Author contribution statement

TG proposed the research. JCHL carried out the analysis. The idea to simulate the returns was due to AJH. JCHL and AJH tested the simulations. All the authors contributed equally to analyzing the results and writing the paper.

We thank Hirdesh K. Pharasi for fruitful discussions. 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] R. K. P. Zia, P. A. Rikvold, Fluctuations and correlations in an individual-based model of biological coevolution, Journal of Physics A: Mathematical and General 37 (19) (2004) 5135–5155. doi:10.1088/0305-4470/37/19/003.
URL https://doi.org/10.1088/0305-4470/37/19/003

[2] R. K. P. Zia, B. Schmittmann, A possible classification of nonequilibrium steady states, Journal of Physics A: Mathematical and General 39 (24) (2006) L407–L413. doi:10.1088/0305-4470/39/24/104.
URL https://doi.org/10.1088/0305-4470/39/24/104

[3] J. B. Gao, Recurrence time statistics for chaotic systems and their applications, Phys. Rev. Lett. 83 (1999) 3178–3181. doi:10.1103/PhysRevLett.83.3178.
URL https://link.aps.org/doi/10.1103/PhysRevLett.83.3178

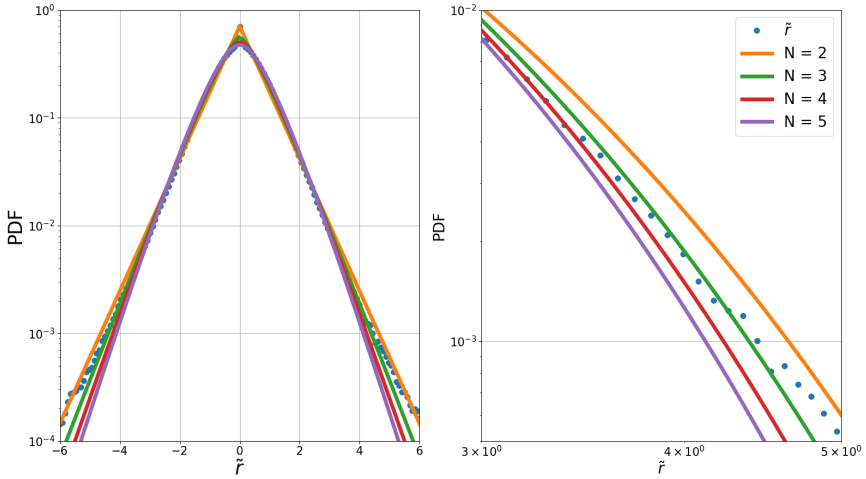

Figure 13:

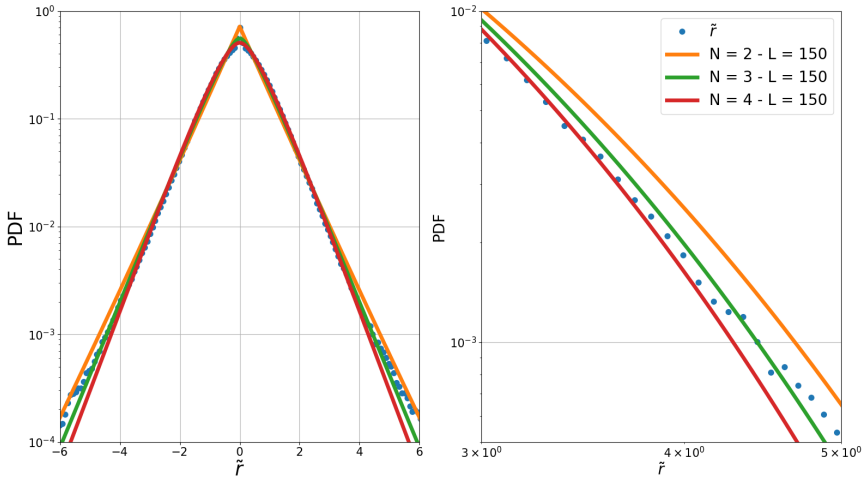

Figure 14:

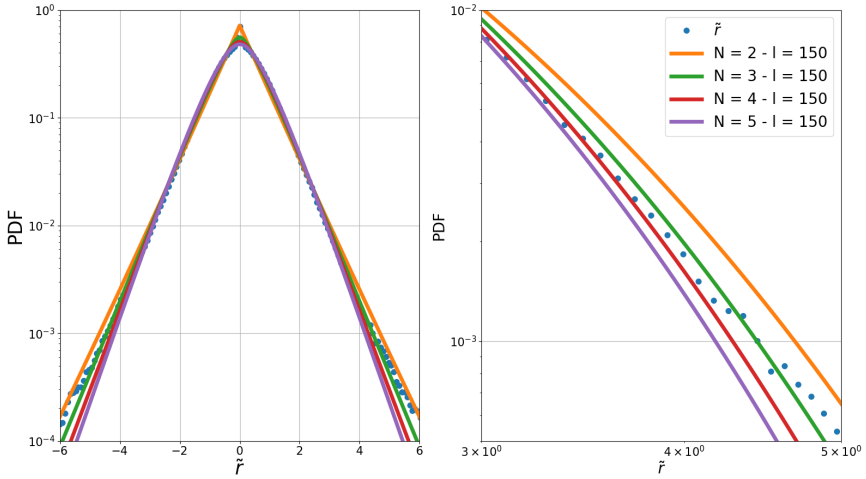

Figure 15:

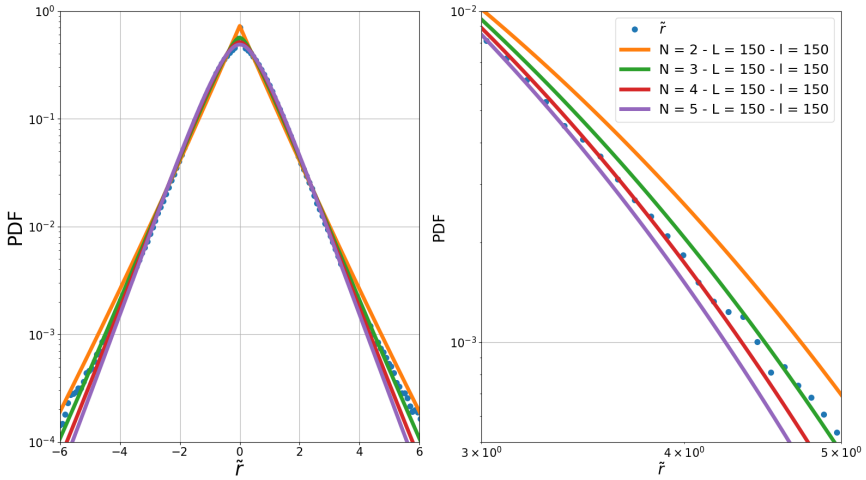

Figure 16:

[4] R. Hegger, H. Kantz, L. Matassini, T. Schreiber, Coping with nonstationarity by overembedding, Phys. Rev. Lett. 84 (2000) 4092–4095. doi:10.1103/PhysRevLett.84.4092.
URL https://link.aps.org/doi/10.1103/PhysRevLett.84.4092
[5] P. Bernaola-Galván, P. C. Ivanov, L. A. N. Amaral, H. E. Stanley, Scale invariance in the nonstationarity of human heart rate, Phys. Rev. Lett. 87 (2001) 168105. doi:10.1103/PhysRevLett.87.168105.
URL https://link.aps.org/doi/10.1103/PhysRevLett.87.168105
[6] C. Rieke, K. Sternickel, R. G. Andrzejak, C. E. Elger, P. David, K. Lehnertz, Measuring nonstationarity by analyzing the loss of recurrence in dynamical systems, Phys. Rev. Lett. 88 (2002) 244102. doi:10.1103/PhysRevLett.88.244102.
URL https://link.aps.org/doi/10.1103/PhysRevLett.88.244102
[7] G. Bekaert, C. R. Harvey, Time-varying world market integration, The Journal of Finance 50 (2) (1995) 403–444.
URL http://www.jstor.org/stable/2329414
[8] F. Longin, B. Solnik, Is the correlation in international equity returns constant: 1960–1990?, Journal of International Money and Finance 14 (1) (1995) 3–26. doi:https://doi.org/10.1016/0261-5606(94)00001-H.
URL https://www.sciencedirect.com/science/article/pii/026156069400001H
[9] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, A. Kanto, Dynamics of market correlations: Taxonomy and portfolio analysis, Phys. Rev. E 68 (2003) 056110. doi:10.1103/PhysRevE.68.056110.
URL https://link.aps.org/doi/10.1103/PhysRevE.68.056110
[10] Y. Zhang, G. H. T. Lee, J. C. Wong, J. L. Kok, M. Prusty, S. A. Cheong, Will the us economy recover in 2010? a minimal spanning tree study, Physica A: Statistical Mechanics and its Applications 390 (11) (2011) 2020–2050. doi:https://doi.org/10.1016/j.physa.2011.01.020.
URL https://www.sciencedirect.com/science/article/pii/S0378437111000847
[11] D.-M. Song, M. Tumminello, W.-X. Zhou, R. N. Mantegna, Evolution of worldwide stock markets, correlation structure, and correlation-based graphs, Phys. Rev. E 84 (2011) 026108. doi:10.1103/PhysRevE.84.026108.
URL https://link.aps.org/doi/10.1103/PhysRevE.84.026108
[12] L. Sandoval, I. D. P. Franca, Correlation of financial markets in times of crisis, Physica A: Statistical Mechanics and its Applications 391 (1) (2012) 187–208. doi:https://doi.org/10.1016/j.physa.2011.07.023.
URL https://www.sciencedirect.com/science/article/pii/S037843711100570X
[13] M. C. Münnix, T. Shimada, R. Schäfer, F. Leyvraz, T. H. Selgman, T. Guhr, H. E. Stanley, Identifying states of a financial market, Scientific Reports 2 (2012). doi:10.1038/srep00644.
URL https://doi.org/10.1038/srep00644
[14] T. A. Schmitt, D. Chetalova, R. Schäfer, T. Guhr, Non-stationarity in financial time series: Generic features and tail behavior, EPL (Europhysics Letters) 103 (5) (2013) 58003. doi:10.1209/0295-5075/103/58003.
URL https://doi.org/10.1209/0295-5075/103/58003
[15] T. Guhr, A. 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) (2021) 125002. doi:10.1088/1751-8121/abe3c8.
URL https://doi.org/10.1088/1751-8121/abe3c8
[16] T. Guhr, A. Müller–Groeling, H. A. Weidenmüller, Random-matrix theories in quantum physics: common concepts, Physics Reports 299 (4) (1998) 189–425. doi:https://doi.org/10.1016/S0370-1573(97)00088-4.
URL https://www.sciencedirect.com/science/article/pii/S0370157397000884
[17] M. L. Mehta, Random Matrices, Elsevier, 2004.
[18] J. Wishart, The generalised product moment distribution in samples from a normal multivariate population, Biometrika 20A (1/2) (1928) 32–52.
URL http://www.jstor.org/stable/2331939
[19] J.-P. Bouchaud, M. Potters, Thery of Financial Risks, from statistical physics to risk management, Cambridge University Press, 2000.
[20] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (1999) 1467–1470. doi:10.1103/PhysRevLett.83.1467.
URL https://link.aps.org/doi/10.1103/PhysRevLett.83.1467
[21] L. Laloux, P. Cizeau, M. Potters, J.-P. Bouchaud, Random matrix theory and financial correlations, International Journal of Theoretical and Applied Finance 03 (03) (2000) 391–397. arXiv:https://doi.org/10.1142/S0219024900000255, doi:10.1142/S0219024900000255.
URL https://doi.org/10.1142/S0219024900000255
[22] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Phys. Rev. Lett. 83 (1999) 1471–1474. doi:10.1103/PhysRevLett.83.1471.
URL https://link.aps.org/doi/10.1103/PhysRevLett.83.1471
[23] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, H. E. Stanley, Random matrix approach to cross correlations in financial data, Phys. Rev. E 65 (2002) 066126. doi:10.1103/PhysRevE.65.066126.
URL https://link.aps.org/doi/10.1103/PhysRevE.65.066126
[24] S. Pafka, I. Kondor, Estimated correlation matrices and portfolio optimization, Physica A: Statistical Mechanics and its Applications 343 (2004) 623–634. doi:https://doi.org/10.1016/j.physa.2004.05.079.
URL https://www.sciencedirect.com/science/article/pii/S0378437104007447
[25] M. Potters, J.-P. Bouchaud, L. Laloux, Financial applications of random matrix theory: Old laces and new pieces, Acta Physica Polonica Series B B35 (2005) 2767–2784.
URL http://inis.iaea.org/search/search.aspx?orig_q=RN:36100087
[26] S. Drozdz, J. Kwapien, P. Oswiecimka, Empirics versus rmt in financial cross-correlations, Acta Physica Polonica Series B B39 (2008) 4027–4039.

URL http://inis.iaea.org/search/search.aspx?orig_q=RN:39016403
[27] J. Kwapień, S. Drożdż, P. Oświecimka, The bulk of the stock market correlation matrix is not pure noise, Physica A: Statistical Mechanics and its Applications 359 (2006) 589–606. doi:https://doi.org/10.1016/j.physa.2005.05.090.
URL https://www.sciencedirect.com/science/article/pii/S0378437105005716
[28] G. Biroli, J.-P. B. M. Potters, The student ensemble of correlation matrices: Eigenvalue spectrum and and kullback-leibler entropy, Acta Physica Polonica Series B B39 (2008) 4009–4026.
URL http://inis.iaea.org/search/search.aspx?orig_q=RN:39016402
[29] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, I. Zahed, Free lévy matrices and financial correlations, Physica A: Statistical Mechanics and its Applications 343 (2004) 694–700. doi:https://doi.org/10.1016/j.physa.2004.05.049.
URL https://www.sciencedirect.com/science/article/pii/S0378437104007186
[30] G. Akemann, P. Vivo, Power law deformation of wishart-laguerre ensembles of random matrices, J. Stat. Mech. 0809 (2008) P09002. arXiv:0806.1861, doi:10.1088/1742-5468/2008/09/P09002.
[31] Z. Burda, A. Jarosz, M. A. Nowak, J. Jurkiewicz, G. Papp, I. Zahed, Applying free random variables to random matrix analysis of financial data. part i: The gaussian case, Quantitative Finance 11 (7) (2011) 1103–1124. arXiv:https://doi.org/10.1080/14697688.2010.484025, doi:10.1080/14697688.2010.484025.
URL https://doi.org/10.1080/14697688.2010.484025
[32] J.-P. Bouchaud, The subtle nature of financial random walks, Chaos: An Interdisciplinary Journal of Nonlinear Science 15 (2) (2005) 026104. arXiv:https://doi.org/10.1063/1.1889265, doi:10.1063/1.1889265.
URL https://doi.org/10.1063/1.1889265
[33] A. Chakraborti, I. Toke, M. Patriarca, F. Abergel, Econophysics: Empirical facts and agent-based models, arXiv.org, Quantitative Finance Papers (09 2009).
[34] W. G. Schwert, Why does stock market volatility change over time?, The Journal of Finance 44 (5) (1989) 1115–1153. doi:https://doi.org/10.1111/j.1540-6261.1989.tb02647.x.
[35] T. A. Schmitt, D. Chetalova, R. Schäfer, T. Guhr, Credit risk and the instability of the financial system: An ensemble approach, EPL (Europhysics Letters) 105 (3) (2014) 38004. doi:10.1209/0295-5075/105/38004.
URL https://doi.org/10.1209/0295-5075/105/38004
[36] D. Chetalova, T. A. Schmitt, R. Schäfer, T. Guhr, Portfolio return distributions: Sample statistics with stochastic correlations, International Journal of Theoretical and Applied Finance 18 (02) (2015) 1550012. arXiv:https://doi.org/10.1142/S0219024915500120, doi:10.1142/S0219024915500120.
URL https://doi.org/10.1142/S0219024915500120
[37] T. Schmitt, R. Schäfer, T. Guhr, Credit risk: Taking fluctuating asset correlations into account, Journal of Credit Risk 11 (3) (2015).
URL https://ssrn.com/abstract=2795524
[38] F. Meudt, M. Theissen, R. Schäfer, T. Guhr, Constructing analytically tractable ensembles of stochastic covariances with an application to financial data, Journal of Statistical Mechanics: Theory and Experiment 2015 (11) (2015) P11025. doi:10.1088/1742-5468/2015/11/p11025.
URL https://doi.org/10.1088/1742-5468/2015/11/p11025
[39] R. Schäfer, T. Guhr, Local normalization: Uncovering correlations in non-stationary financial time series, Physica A-statistical Mechanics and Its Applications 389 (2010) 3856–3865. doi:10.1016/j.physa.2010.05.030.
[40] J. E. Gentle, Computational Statistics, Springer-Verlag New York, 2009. doi:10.1007/978-0-387-98144-4.
[41] A. Gupta, D. Nagar, Matrix Variate Distributions, Monographs and Surveys in Pure and Applied Mathematics, Taylor & Francis, 1999.
URL https://books.google.de/books?id=PQOYnT7P1loC