Zooming into market states

Desislava Chetalova, Rudi Schäfer and Thomas Guhr

Fakultät für Physik, Universität Duisburg Essen, D-47048 Duisburg, Germany
E-mail: desislava.chetalova@uni-due.de

Received 16 June 2014
Accepted for publication 18 December 2014
Published 21 January 2015

Abstract. We analyze the daily stock data of the Nasdaq Composite index in the 22-year period 1992–2013 and identify market states as clusters of correlation matrices with similar correlation structures. We investigate the stability of the correlation structure of each state by estimating the statistical fluctuations of true correlations due to their non-stationarity. Our study is based on a random matrix approach recently introduced to model the non-stationarity of correlations by an ensemble of random matrices. This approach reduces the complexity of the correlated market to two parameters, one of which measures the fluctuations of the correlations and can be determined directly from the empirical return distributions. This parameter characterizes the stability of the correlation structure of each market state as well as the dynamics of the correlation structure in the whole observation period. We find clear indications for correlation fluctuations within market states. The analysis reveals an intriguing relationship between average correlation and correlation fluctuations. The strongest fluctuations occur during periods of high average correlation which is the case particularly in times of crisis.

Keywords: models of financial markets
1. Introduction

Financial markets are highly complex and continuously evolving systems. To understand their statistical behavior and dynamics the analysis of the correlations between the market constituents is of crucial importance. Much research has thus been focused on the information acquired by the correlations, see e.g. [1, 2].

Recently, correlation matrices were used to identify states of a financial market based on similarities in the correlation structure at different times [3]. The study revealed the existence of several typical market states between which the market jumps back and forth over time. Each market state has a characteristic correlation structure and temporal behavior.

Here, we take a closer look at the statistics of market states. In particular, we investigate the stability of their corresponding correlation structures. We address this question by means of a random matrix approach introduced in [4, 5]. It models the non-stationarity of the correlations by an ensemble of random matrices. This approach not only yields a quantitative description of heavy-tailed return distributions but also reduces the effect of fluctuating correlations to a single parameter measuring the fluctuation strength. Our random matrix approach provides a method to estimate the fluctuations of the true correlations due to non-stationarity [6–12]. This is a crucial difference to another usage of random matrices where the estimation errors which arise due to the finiteness of the time series are modeled [13–18].

We note that our study is based on the definition of market states as clusters of correlation matrices, as first suggested in [3]. The concept of different market states or regimes in which the market operates is, however, not entirely new to the economics literature, see e.g. [19, 20].

doi:10.1088/1742-5468/2015/01/P01029
The paper is organized as follows. In section 2, we identify market states for the Nasdaq Composite stock market in the period 1992–2013 by performing a clustering analysis based on the Partitioning Around Medoids (PAM) algorithm [21] and study their dynamics. We briefly summarize the main features of the random matrix model in section 3. We apply it to study the stability of the correlation structure of each market state in section 4 and the correlation structure dynamics in the whole observation period in section 5. In section 6 we study the fine structure of the return distribution related to the principal components. We conclude our findings in section 7.

2. Market states: identification and dynamics

We begin with identifying the market states as clusters of correlation matrices. We consider $K = 258$ stocks of the Nasdaq Composite index traded in the 22 year period from January 1992 to December 2013\(^1\) which corresponds to 5542 trading days. For each stock $k$ we calculate the return time series

$$r_k(t) = \frac{S_k(t + \Delta t) - S_k(t)}{S_k(t)}, \quad k = 1, \ldots, K,$$

where $S_k(t)$ is the price of the $k$th stock at time $t$ and $\Delta t$ is the return interval. We choose $\Delta t$ to be one trading day and calculate the daily returns for each stock.

The correlations between time series are commonly measured via the Pearson correlation coefficient

$$C_{kl} = \frac{\langle r_k(t) r_l(t) \rangle - \langle r_k(t) \rangle \langle r_l(t) \rangle}{\sqrt{\langle r_k^2(t) \rangle - \langle r_k(t) \rangle^2} \sqrt{\langle r_l^2(t) \rangle - \langle r_l(t) \rangle^2}},$$

where $\langle \ldots \rangle$ denotes the average over a time window yet to be specified. The main object of interest in the following is the $K \times K$ correlation matrix $C$ which contains the correlation coefficients between all pairs of return time series.

The Pearson correlation coefficient is commonly used as a measure of dependence. Sometimes, however, it can be problematic in particular for non-linear dependencies or for non-stationary data. The latter is extremely relevant for financial data since drift and volatilities [22–24] fluctuate considerably in time. Thus, the correlation coefficient averages over time-varying trends and volatilities, which results in an estimation error of the correlations. In order to eliminate this kind of error we employ the method of local normalization [25]. For each return time series $k$ we subtract the local mean and divide by the local standard deviation

$$\hat{r}_k(t) = \frac{r(t) - \langle r(t) \rangle_n}{\sqrt{\langle r^2(t) \rangle_n - \langle r(t) \rangle_n^2}},$$

where $\langle \ldots \rangle_n$ denotes the local average which runs over the $n$ most recent sampling points. For daily data we use $n = 13$ as discussed in [25]. Thus, the local normalization removes the local trends and variable volatilities while preserving the correlations between the time series. An alternative approach would be to use the residuals of a GARCH fit [26], which also yields stationary time series. We choose the local normalization, as it does not require any model assumptions.

\(^1\) NASDAQ Composite data obtained from http://finance.yahoo.com/
Using the locally normalized daily returns we now obtain a set of 131 correlation matrices measured on disjoint two-month intervals of the 22 year observation period. To identify the market states we perform a clustering analysis based on the PAM algorithm where the number of clusters is estimated via the gap statistic [27]. The clustering analysis separates the correlation matrices into six groups based on the similarity of their correlation structures. Each group is associated with a market state. Figure 1 (top) shows the time evolution of the market states.

The market switches back and forth between states: Sometimes it remains in a state for a long time, sometimes it jumps briefly to another state and returns back or evolves further. On longer time scales, the market evolves towards new states, whereas previous states die out. How frequently does the market switch between states? Figure 1 (bottom) shows the number of jumps from one state to another calculated on a one-year sliding window.

After a stable five-year period we observe that the market begins to switch between states. The highest number of jumps per year can be found in the period around 2010. In figure 2 we compare the number of jumps frequency in both halves of the observation period. In the second half of the observation period the number of jumps per year increases. At the same time, the lifetime, i.e. the time the market stays in a certain state before it jumps to another one, decreases. Figure 3 shows the histograms of the lifetime in both halves of the observation period. The first half contains mostly long-lived states. In the second half the frequency of the short-lived states increases considerably, while the frequency of the long-lived states decreases.

To illustrate the different correlation structures of each state, we sort the stocks according to their industry sector and calculate the corresponding average correlation matrices, see figure 4. We indeed recognize different characteristic correlation structures.
State 1 shows an overall weak correlation. In state 2 we have the strongest correlation within the technology sector and between technology and capital goods, whereas in states 3 and 4 the correlation within the finance sector is the strongest. We observe that the average correlation level increases from state to state, reaching its highest value in state 5. In state 6 the average correlation level decreases. The finance sector, however, is still strongly correlated. Further, we note that the health care sector is weakly correlated to the rest of the market in almost all states.

3. Random matrix approach to non-stationary correlations

Before we take a closer look at each market state we briefly summarize the main aspects of our random matrix approach, which gives us a handy tool to estimate the correlation fluctuations due to non-stationarity.

Consider a market consisting of $K$ stocks. We are aiming at deriving a multivariate return distribution for the whole market in the presence of time-varying correlations among the stocks. We begin with assuming that the $K$ component return vector $r(t) = (r_1(t), \ldots, r_K(t))$ at each time $t$ is multivariate normally distributed with a correlation

**Figure 2.** Histograms of the number of jumps between states (a) in the first half 1992–2002 and (b) in the second half 2003–2013 of the observation period.

**Figure 3.** Histograms of the lifetime in months (a) in the first half 1992–2002 and (b) in the second half 2003–2012 of the observation period.
Figure 4. (a)–(f) Average correlation matrices for each market state. (g) Overall average correlation matrix. Industry sectors legend: BI, basic industries; CG, capital goods; CD, consumer durables; CN, consumer non-durables; CS, consumer services; E, energy; F, finance; HC, health care; M, miscellaneous; PU, public utilities; T, technology; TR, transportation.
matrix $C_t$. Careful data analysis performed in [4] verified this assumption for daily returns on short time horizons where the correlation matrix can be viewed as fixed. This is not a contradiction to the heavy-tailed distributions for the individual returns on longer time horizons. Moreover, the very fat tails of the univariate distributions are suppressed in the multivariate distribution of all returns. The reason for that are the correlations. Later on, we look at aggregated distributions, obtained by rotating the returns into the eigenbasis of the covariance matrix. Hence, the aggregated distributions are distributions of linear combinations of returns. Such effects are well known in statistics to reduce tails.

On long time horizons, however, non-stationarity is clearly present. We take it into account by replacing the correlation matrix at each time $t$ by a random matrix

$$C_t \rightarrow WW^\dagger,$$

where the dagger denotes the transpose. The model matrix $W$ is a rectangular $K \times N$ real random matrix, where $N$ formally represents the length of $K$ model time series. It is drawn from a Gaussian distribution with the probability density function (pdf)

$$w(W|C, N) = \frac{1}{\sqrt{\det C_N^{KN}}} \exp \left( -\frac{N}{2} \text{tr} W^\dagger C^{-1} W \right),$$

where $C$ is the empirical average correlation matrix, computed over the whole sample, not to be confused with the matrices $C_t$ introduced above. This sample may be the whole observation period, a given market state or some arbitrary time window. Hence, we model the non-stationary correlation matrices by an ensemble of random Wishart matrices $WW^\dagger$ which fluctuate around the average correlation matrix $C$. We note that in section 4 $C$ will refer to the respective correlation matrix of each market state, which we denote as $C^{(i)}$, $i = 1, \ldots, 6$, whereas in section 5 it will represent the average correlation matrix in a sliding 500-trading-day window. We further point out that in our approach the Wishart ensemble models fluctuations in the actual correlation structure and not the estimation error of large correlation matrices as commonly used. The variance of the Wishart ensemble is determined by the empirical correlation matrix $C$ scaled with the parameter $N$

$$\text{var} \left([WW^\dagger]_{kl}\right) = \frac{1 + C_{kl}^2}{N},$$

where $C_{kl}$ is the $kl$th element of $C$. Thus, the parameter $N$ is directly related to the fluctuation strength of the correlations in the respective sample. The larger $N$, the smaller the fluctuations around $C$ become, eventually vanishing in the limit $N \to \infty$. Averaging over the random matrix ensemble we derive a correlation averaged multivariate normal distribution for the returns of a correlated financial market. The average return pdf reads

$$\langle g \rangle (r|\Sigma, N) = \frac{\sqrt{2}^{1-N} \sqrt{N^K}}{\Gamma(N/2) \sqrt{2\pi \det(\Sigma)}} \frac{\mathcal{K}_{\frac{\nu}{2}} \left( \sqrt{N r^\dagger \Sigma^{-1} r} \right)}{\sqrt{N r^\dagger \Sigma^{-1} r} \frac{\mathcal{K}_{\frac{\nu}{2}}}{2}},$$

where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order $\nu$. It depends only on the empirical covariance matrix $\Sigma = \sigma C \sigma$, evaluated over the respective sample, where $\sigma$ is the diagonal matrix of the volatilities $\sigma_k$. The free parameter $N$ can be determined

doi:10.1088/1742-5468/2015/01/P01029 7
by fitting to the data. Thus, we can assess the correlation fluctuations, which provide information about the stability of the correlation structure in the considered sample, directly from the empirical data. For the comparison with empirical returns we rotate the return vector \( r \) into the eigenbasis of the covariance matrix and normalize its components with the eigenvalues. Integrating out all but one component of the rotated vector, which we call \( \tilde{\eta} \), leads to

\[
\langle g \rangle (\tilde{\eta} | N) = \sqrt{2^{1-N} N} \sqrt{\frac{1}{\pi} \Gamma(N/2)} \sqrt{N \tilde{\eta}^2}^{N-1} \mathcal{K}_{N-\frac{1}{2}} \left( \sqrt{N \tilde{\eta}^2} \right). \tag{8}
\]

Distributions of this kind are often referred to as \( \mathcal{K} \)-distributions. We illustrate this pdf for different values of \( N \) in figure 5. It has exponential tails, which become more and more dominant the smaller \( N \). The smaller \( N \) the stronger the fluctuations around \( C \). For large \( N \) the pdf approaches the normal distribution. The limit corresponds to a stationary case with no fluctuations around \( C \).

We note that in the calculation of the average return distribution we treated the volatility \( \sigma \) as fixed. However, we can alternatively model the full covariance matrix by an ensemble of random Wishart matrices, which leads to exactly the same result (7). In this case, \( N \) can also be interpreted as fluctuation strength around the average covariance matrix \( \Sigma = \sigma C \sigma \). In the following we use the interpretation of \( N \) as fluctuation strength around \( C \), which we justify in section 5.

4. Stability of the correlation structure for each market state

In section 2 we used a clustering analysis to group our original set of 131 correlation matrices, calculated on two-month time intervals of the 22 year observation period, into six distinct groups based on the similarity of the correlation structure. Each of these six groups, which we identify as different market states, is characterized by its average correlation matrix \( C^{(i)} \). In this section, we study the stability of the correlation structure for each state. In particular, we address the question: Are the correlations of a given
state stationary or do they fluctuate around the respective average state correlation matrix \( C^{(i)} \), and if so, how strongly? This question, however, cannot be answered by looking at the empirical correlation matrices. Since they are calculated on very short time intervals of two months, they contain a considerable amount of noise. The clustering is not so sensitive to the noise level, since it is based on a distance measure which averages over the noise. This noise, however, competes with actual fluctuations of the correlations. Thus it prevents us from studying the stability of the correlation structure directly. In order to assess the actual fluctuations of correlations due to non-stationarity, we use the random matrix model introduced in section 3. Instead of assuming a constant correlation matrix for each market state, we assume a Wishart ensemble of correlation matrices which allows for fluctuations around the average state correlation matrix \( C^{(i)} \). Further, we assume conditional normality, i.e. return vectors following a multivariate normal distribution conditioned on a fixed correlation or covariance matrix. Not only is this assumption common in the literature (e.g. in GARCH models and stochastic volatility models), but it is also justified by empirical data, as discussed in section 3. The two extreme assumptions of stationary versus non-stationary correlations lead to different sample statistics for the multivariate returns observed within a given market state. In case of stationary correlations, we would expect a normal distribution. In case of non-stationary correlations, we expect a \( K \)-distribution of the form (8) instead. Once the average state correlation matrix is fixed we may use the parameter \( N^{(i)} \) to fully characterize the fluctuations. By fitting to the empirical return distributions we obtain a measure \( N^{(i)} \) for these fluctuations. In a further step, we compress the information contained in each \( C^{(i)} \) into a single number, namely an average correlation coefficient \( c^{(i)} \). This allows us to study the relationship between fluctuations and average market correlation.

We obtain the return time series for each market state in the following way: We take the original daily return time series \( r(t) \) and divide it into a sequence of disjoint two-month intervals. We merge all intervals belonging to a given state according to the cluster analysis described in section 2. We note that the return time series for the six market states differ in length. For the comparison with the model we rotate the return vector for each state into the eigenbasis of the state covariance matrix \( \Sigma^{(i)} \) and normalize its components with the eigenvalues. We aggregate all components into a single histogram and compare it with the average return distribution (8). Figure 6 shows the results for each market state \((a)-(f)\) and for the whole 22 year observation period \((g)\). For the latter, we rotate the original return time series \( r(t) \) into the eigenbasis of the covariance matrix \( \Sigma \) calculated over the 22 year observation period.

It is already obvious from the heavy-tailed empirical return distributions that the assumption of stationary correlations within a market state has to be dismissed. Instead, we observe a clear indication for fluctuations around each average state correlation matrix \( C^{(i)} \). For the whole observation period we find a much smaller \( N \). In this case, the fluctuations are stronger than for the single states. The parameter \( N^{(i)} \) is estimated by the maximum likelihood method and depicted for each state together with the average correlation \( c^{(i)} \) in figure 7\((a)\). We obtain \( c^{(i)} \) by averaging over the off-diagonal correlation coefficients \( C^{(i)}_{kl}, \ k \neq l \) of the average correlation matrix for a given state. We observe that the states 1 and 2, which cover the period 1992 to roughly 2002, are rather stable. We find
Zooming into market states

Figure 6. (a)–(f) Histograms of the rotated and rescaled returns for each market state and (g) for the whole observation period 1992–2013 compared with the average return distribution (8).

low average correlation with high $N$ values, i.e. weak fluctuations. In states 3 and 4 the fluctuations increase. While the $N$ values for both states are equal, the average correlation is rising. In state 5, first appearing during the crisis in 2008, the fluctuations increase further. It is the most unstable state with the smallest $N$ value and the highest average correlation. In state 6 the fluctuations and the average correlation decrease, the market stabilizes. To examine the relationship between average correlation and fluctuations we look at the scatter plot between $c^{(i)}$ and $N^{(i)}$, see figure 7(b). We observe a clear decreasing trend, i.e. a negative correlation.

doi:10.1088/1742-5468/2015/01/P01029
5. Dynamics of the correlation structure

To further investigate the relationship between fluctuations and average correlation we now take a closer look at the variation of the correlation structure over time. To this end, we examine the parameter $N$ and the average correlation $c$ computed on a sliding window of 500 trading days shifted by 21 trading days, see figure 8. As in the previous section, the parameter $N$ for each time window is estimated by fitting to the aggregated distribution of the rotated and scaled returns, where for the rotation we use the average covariance matrix in the given time window. The parameter $c$ is obtained by averaging over the off-diagonal elements of the average correlation matrix for the corresponding time window. We recognize four distinct regimes: The first regime covers the period 1/1992 to 9/1996, which mainly corresponds to the stable market state 1, see figure 1. Here we find the lowest average correlation and the highest $N$ value. While the average correlation is relatively stable in this period, the $N$ value shows a clear decreasing trend indicating increasing fluctuations. The second regime covers the period 10/1996 to 9/2006, which corresponds to the stable states 1 and 2 and the more unstable states 3 and 4. While the average correlation in this period is steadily growing, the $N$ value is mostly stable. Compared to the first regime we find smaller $N$ values because of the transitions between the different market states. The third regime, beginning 10/2006 to 5/2009, covers mostly the period before and during the financial crisis in 2008 and corresponds to the unstable states 4 and 5. We observe a sharp increase in the average correlation, which is over two times larger compared with the first regime. Indeed, in times of market instabilities collective behavior is induced which results in larger correlations. Here we find the smallest $N$ values, i.e. the strongest fluctuations. The last regime, beginning 6/2009, covers the rest of the observation period and corresponds to the unstable states 4 and 5 and the stable state 6. The fluctuations decrease slightly. The average correlation increases at first even further.
compared to the previous regime but decreases again after 2010. The market stabilizes after the crisis.

The relationship between average correlation and fluctuations for the two-year time window is depicted in figure 9(a). We observe an overall negative correlation between $c$ and $N$. Moreover, the data corresponding to the four regimes cluster into different regions: a stable region (regime I) characterized by low average correlation and weak fluctuations, which are typical for calmer periods; an unstable region (regime III) characterized by high average correlation and strong fluctuations, typical for crisis periods; and an intermediate region (regime II and IV) characterized by varying average correlation and more moderate fluctuations.

Finally, we examine the dependence between average volatility $\sigma$ and fluctuations for the two-year time window, shown in figure 9(b). Again, we find clustering into regions as observed before: a stable region with weak fluctuations and nearly constant volatility $\sigma \approx 0.03$; an unstable region with strong fluctuations and high volatility; and an intermediate region. In this case, we do not recognize a clear trend, $\sigma$ and $N$ show no clear dependence. This justifies our interpretation of $N$ as correlation rather than covariance fluctuations.

6. Fine structure of the return distribution

When studying the distribution of returns one can look at different objects. Here, we are interested in the return distribution of the whole market as well as in the return
distributions of the single states in which the market operates. Thus, we look at the aggregated distribution of returns. It is obtained by rotating the return vector \( r \) into the eigenbasis of the covariance matrix, \( \eta = U r \). The matrix \( U \) is an orthogonal matrix which diagonalizes the covariance matrix as \( \Sigma = U^\dagger \Lambda U \), where \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K) \) is the diagonal matrix of the eigenvalues, ordered in descending order. The components of \( \eta \) are then scalar products of the eigenvectors of the covariance matrix with the return vector \( \eta_k = U_k \cdot r \). Normalizing these components by the empirical eigenvalues \( \tilde{\eta}_k = \eta_k / \sqrt{\lambda_k} \), allows us to view all of them on equal footing and to aggregate them into one histogram which then captures the properties of the multivariate return distribution as a whole. The distribution for the entire market, i.e. including all eigenvalues, is by no means artificial. It plays an important role for example in credit risk [28] and portfolio optimization [5].

On the other hand, one can look at the distributions of the single components \( \tilde{\eta}_k \), which represent the principal components of the original data. While in our model all components have the same distribution (8), in the data we find significant deviations for the largest 20 and the smallest 4 eigenvalues. The remaining components have similar statistics and are consistent with the aggregated distribution of all returns. Figure 10 shows the distributions of the normalized components \( \tilde{\eta}_k \) averaged over 43 components.

**Figure 9.** Scatter plots (a) \( c \) versus \( N \) and (b) \( \sigma \) versus \( N \), for the time window of 500 trading days.

**Figure 10.** Distributions of the normalized components \( \tilde{\eta}_k \) averaged over 43 components compared to the aggregated distribution of all returns, plotted (a) linearly and (b) logarithmically.
We observe that the average distribution of the first 43 components, which belong to the largest eigenvalues, has different statistics compared to the rest. The associated eigenvectors have different structures and interpretations. The first component $\hat{\eta}_1$ belongs to the largest eigenvalue $\lambda_1$, which describes the whole market. The other large eigenvalues correspond to eigenvectors that have only a subset of components different from zero. These eigenvectors roughly correspond to market sectors. Furthermore, the smallest eigenvalues are most sensitive to measurement noise. A histogram of the eigenvalues of the covariance matrix $\Sigma$, evaluated over the whole observation period, is depicted in figure 11.

Still, the different statistics of some of the principal components is not an obstacle to aggregate all components together when looking at the multivariate distribution as a whole. No model can capture all aspects of reality. Although we find an overall good agreement with the aggregated distribution of all returns, the fine structure cannot be fully captured by our model.

7. Conclusion

To achieve a better understanding of the financial market, the concept of market states as clusters of correlation matrices was introduced in [3]. Here, we take a closer look at the statistics of market states studying the Nasdaq Composite market over a period of 22 years. For this purpose, we use a recently developed random matrix approach, which models the non-stationarity of true correlations by a random matrix ensemble. Alongside with a heavy-tailed distribution for the stock returns, the approach provides a method to study the correlation structure by estimating the fluctuation strength of correlations in a given time period directly from the empirical return distributions.

Our study provides a better understanding of the market state dynamics as well as of the stability of the corresponding correlation structure. Despite the non-stationarity of the market we find a set of quite stable states in which the market operates. We discuss their statistical properties and study the dynamics of the correlation structure in the whole observation period using a sliding window analysis. We find four distinct
regimes with different statistical behavior. The analysis reveals a remarkable relationship between average correlation and fluctuations. Strong fluctuations most likely occur during periods of high average correlation. Unstable periods are thus characterized not only by larger correlations and high volatilities but also by strong correlation fluctuations. Furthermore, we study the relationship between fluctuations and average volatility. In this case, we do not find a clear trend, volatility and fluctuations are mostly independent of each other. In addition, we also study the fine structure of the return distribution related to the principal components. Although we find an overall good agreement with the aggregated distribution of all returns, the fine structure cannot be fully captured by our model.

Another, conceptual aspect of the present study should be mentioned. At first sight, the following two results might appear contradictory: when studying the entire time interval of 22 years from 1992 to 2013, we identify, on the one hand, a small number of distinct states in which the market operates, but on the other hand, we claim that the return distribution for this entire time interval can be modeled by a random matrix ansatz [4]. The simultaneous existence of few distinct states and of an ensemble of homogenously distributed correlation matrices in the random matrix ensemble might seem incompatible. Importantly, these two features can coexist. The present study may be viewed as a refined resolution of the really existing (random) matrix ensemble in terms of a superposition of sub-ensembles around the distinct states. Certainly, one might come up with statistical observables that can make this fine structure of the ensemble visible—as we do in the present study. However, the plain return distribution itself is a highly relevant quantity, see e.g. [28]. To study it, the data are aggregated, i.e. represented in the eigenbasis of the mean covariance matrix for the entire time interval. This involves a rotation of the return vector with an orthogonal matrix and thus further randomization which is just an additional averaging over the sub-ensembles. This is why our random matrix ensemble works inspite of the existence of distinct market states.

Such effects are quite common in random matrix models and one of the reasons for their remarkable robustness. Wigner’s original random matrix ansatz [29] based on a rotation invariant and homogenous ensemble was heavily criticized by many nuclear physicists. They argued that no realistic Hamilton matrix of a nucleus, calculated in some basis, will look like a random matrix, because it will contain many strict zeros due to selection rules. Thus, there was a blatant disagreement with the idea of a rotation invariant ensemble. This then led to the embedded random matrix ensembles [30] which correctly incorporate the selection rules. Nevertheless, the statistical observables of particular interest are indistinguishable for Wigner’s original ansatz and for the embedded random matrix ensembles.

References

[1] Mantegna R N and Stanley H E 1999 An Introduction to Econophysics: Correlations and Complexity in Finance (Cambridge: Cambridge University Press)
[2] Bouchaud J P and Potters M 1999 Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management (Cambridge: Cambridge University Press)
[3] Münnix M C, Shimada T, Schäfer R, Leyvraz F, Seligman T H, Guhr T and Stanley H E 2012 Identifying states of a financial market Sci. Rep. 2 644

doi:10.1088/1742-5468/2015/01/P01029
Zooming into market states

[4] Schmitt T A, Chetalova D, Schäfer R and Guhr T 2013 Non-stationarity in financial time series: generic features and tail behavior Europhys. Lett. 103 58003

[5] Chetalova D, Schmitt T A, Schäfer R and Guhr T 2013 Portfolio return distributions: sample statistics with non-stationary correlations Int. J. Theor. Appl. Finance accepted for publication (arXiv:1308.3961)

[6] Bekaert G and Harvey C R 1995 Time-varying world market integration J. Finance 50 403–44

[7] Longin F M and Solnik B 1995 Is the correlation in international equity returns constant: 1960–1990? J. Int. Money Finance 14 3–26

[8] Drozdz S, Kwapien J, Grimmer F, Ruf F and Speth J 2001 Quantifying the dynamics of financial correlations Physica A 299 144–53

[9] Raffaelli G and Marsili M 2006 Dynamic instability in a phenomenological model of correlated assets J. Stat. Mech. L08001

[10] Fenn D. J, Porter M. A, Williams S, McDonald M, Johnson N F and Jones N S 2011 Temporal evolution of financial-market correlations Phys. Rev. E 84 026109

[11] Livan G, Inoue J and Scalas E 2012 On the non-stationarity of financial time series: impact on optimal portfolio selection J. Stat. Mech. P07025

[12] Preis T, Kenett D Y, Stanley H E, Helbing D and Eshel B-J 2012 Quantifying the behavior of stock correlations under market stress Sci. Rep. 2 752

[13] Laloux L, Cizeau P, Bouchaud J P and Potters M 1999 Noise dressing of financial correlation matrices Phys. Rev. Lett. 83 1467–70

[14] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N and Stanley H E 1999 Universal and nonuniversal properties of cross correlations in financial time series Phys. Rev. Lett. 83 1471–4

[15] Potters M, Bouchaud J P and Laloux L 2005 Financial applications of random matrix theory: old laces and new pieces Acta Phys. Pol. B 36 2767–84

[16] Burda Z, Görlich A T and Waclaw B 2006 Spectral properties of empirical covariance matrices for data with power-law tails Phys. Rev. E 74 041129

[17] Abul-Magd A Y, Akemann G and Vivo P 2009 Superstatistical generalizations of Wishart–Laguerre ensembles of random matrices J. Phys. A: Math. Theor. 42 175207

[18] Burda Z, Jarosz A, Jurkiewicz J, Nowak M A, Papp G and Zahed I 2011 Applying free random variables to random matrix analysis of financial data. Part I: a Gaussian case Quant. Finance 11 1103

[19] Schaller H and Norden S V 1997 Regime switching in stock market returns Appl. Financial Econ. 7 177–91

[20] Marsili M 2002 Dissecting financial markets: sectors and states Quant. Finance 2 297–302

[21] Kaufman L and Rousseeuw P J 1990 Finding Groups in Data: an Introduction to Cluster Analysis (New York: Wiley)

[22] Black F 1976 Studies of stock price volatility changes Meetings of the American Statistical Association, Business and Economical Statistics Section (Washington, DC: American Statistical Association) pp 177–81

[23] Christie A A 1982 The stochastic behavior of common stock variances: value, leverage and interest rate effects J. Financial Econ. 10 407–32

[24] Schwartz G W 1989 Why does stock market volatility change over time? J. Finance 44 1115–53

[25] Schäfer R and Guhr T 2010 Local normalization: uncovering correlations in non-stationary financial time series Physica A 389 3856

[26] Bollerslev T 1986 Generalized autoregressive conditional heteroskedasticity J. Econom. 31 307–27

[27] Tibshirani R, Walther G and Hastie T 2001 Estimating the number of clusters in a data set via the gap statistic J. R. Stat. Soc. B 63 411–23

[28] Schmitt T A, Chetalova D, Schäfer R and Guhr T 2014 Credit risk and the instability of the financial system: an ensemble approach Europhys. Lett. 105 38004

[29] Wigner E P 1967 Random matrices in physics SIAM Rev. 9 1–23

[30] Brody T A, Flores J, French J B, Mello P A, Pandey A and Wong S S M 1981 Random-matrix physics: spectrum and strength fluctuations Rev. Mod. Phys. 53 385–480

doi:10.1088/1742-5468/2015/01/P01029