

A New Attempt to Identify Long-term Precursors for Financial Crisis in the Market Correlation Structures

Anton J. Heckens* and Thomas Guhr†

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

Prediction of events in financial markets is every investor’s dream and, usually, wishful thinking. From a more general, economic and societal viewpoint, the identification of indicators for large events is highly desirable to assess systemic risks. Unfortunately, the very nature of financial markets, particularly the predominantly non-Markovian character as well as non-stationarity, make this challenge a formidable one, leaving little hope for fully fledged answers. Nevertheless, it is called for to collect pieces of evidence in a variety of observables to be assembled like the pieces of a puzzle that eventually might help to catch a glimpse of long-term indicators or precursors for large events — if at all in a statistical sense. Here, we present a new piece for this puzzle. We use the quasi-stationary market states which exist in the time evolution of the correlation structure in financial markets. Recently, we identified such market states relative to the collective motion of the market as a whole. We study their precursor properties in the US stock markets over 16 years, including two crises, the dot-com bubble burst and the pre-phase of the Lehman Brothers crash. We identify certain interesting features and critically discuss their suitability as indicators.

I. INTRODUCTION

Critical events in the financial markets bear, in the age of globalization, ever higher risks for the world’s economic system, which then, in a feedback loop, can cause negative impacts on the financial markets. An early warning system is as desirable as in the case of geologic, seismic and volcanic hazards, but at least as difficult to design. The natural laws under which the latter emerge do not alter, while the fast economic and societal development implies, for the structure and the functionality of the financial system, a new and considerable unpredictability which adds to the risks due to the kind of non-stationarity which has always been present in the markets [1–6].

Musmeci et al. [7] showed that the correlation structure can be used, to some extent, to forecast the volatilities within “persistent” periods, *i.e.* periods of quasi-stationary correlation structures. This ends when transitions between persistent periods take place which is often followed by larger volatility changes. These persistent periods are strongly connected to “market states” or “regimes” in the economics terminology [8–14]. There are different ways to define market states [15]. Here, we follow Refs. [9, 16–21] and identify quasi-stationary markets states in the time evolution of the non-stationary correlation structure. The industrial sectors, which are clearly visible in the correlations [22–29] and covariances [30], and their mutual connections are thereby analyzed in a time resolved fashion. This is accomplished by applying k -means clustering [31–35], a machine learning algorithm, to a set of correlation matrices measured over time in a moving window. The resulting clusters which we identify as market states can be regarded as

quasi-stationary structures in time, such that the individual correlation matrices fluctuate about the cluster centers [16, 18]. The market states emerge, exist for some time and eventually disappear [9, 16–21, 36–40]. Recently, similar methods have been applied in other fields, such as studies of epileptic seizures [41] of freeway traffic [21].

The market states are known to be dominated by the collective motion of the market as a whole [16] which is related to the fact that the corresponding eigenvalue is largest and a measure for the average correlation coefficient [42]. To uncover the dynamics of the correlation structure, particularly due to the industrial sectors, relative to the dominating collective motion, *i.e.* to measure the correlations in the moving frame of the collective motion, we recently put forward a systematic and mathematically well-defined method [20]. We subtract the dyadic matrix belonging to the largest eigenvalue and thereby define two types of reduced-rank correlation matrices whose dynamics are then analyzed with the above k -means clustering algorithm. Reduced-rank correlation matrices of another kind appear in the context of filtering where the smaller eigenvalues are removed as a noise reduction technique [22, 29, 43–45]. Subsequently, from the filtered standard correlation matrices, reduced-rank correlation matrices are calculated in order to obtain well-defined correlation matrices [46–48].

The quasi-stationary market states in the time evolution of these reduced-rank correlation matrices are not directly related to the ones of the standard, *i.e.* not reduced-rank, ones. Surprisingly, they appear to be more stable, sometimes over several years, than the ones for the standard correlation matrices. This observation prompted the present study. Pharasi et al. [49] proposed to identify precursors for critical events in the quasi-stationary market states of the standard correlation matrices, while we here use the reduced-rank ones in view of their higher stability. More precisely, we exploit the separation of two different time scales. The quickly changing

* anton.heckens@uni-due.de

† thomas.guhr@uni-due.de

collective contribution is separated from the more stable one due to the industrial sectors.

It is worth mentioning that the collective market motion is at least more likely to be influenced by exogenous effects than the motion relative to it. Exogenous information can trigger collective volatility outbursts as the traders' reaction to unexpected events [50–53]. By removing these risk contributions, we analyze the dynamics of the endogenous risk in the mutual interaction of the industrial sectors and the corresponding contagious effect for the whole market, potentially leading to market instabilities. Our analysis adds to previous studies of systemic risk [54]. Noteworthy are investigations related to principal components analysis [55–57] and causality measures [56, 58].

Here, we take data into account exclusively from epochs before crises events or within crises periods in order to investigate market state transitions as precursors. We address these two questions: How many epochs before a crisis event or within a crisis period does a “crisis market state” in the correlation structure show up? — Can we identify characteristic precursor signals in the dynamics of the reduced-rank correlation matrices without using cluster methods?

The paper is organized as follows. In Sec. II, we introduce our set of daily data for the analysis. The construction of reduced-rank correlation matrices is briefly sketched in Sec. III. In Sec. IV, we analyze the data and present our results. The conclusions are given in Sec. V.

II. DATA SET

Using data collected by QuoteMedia [59] and provided by Quandl [60] we construct a survivorship-biased portfolio of $K = 250$ US stocks (see Tab. IV in App. A), *i.e.* the selected stocks do not change for the entire investigation period. The investigation period ranges from 02 January, 1997 to 31 December, 2012 in order to concentrate on the two periods, the dotcom-bubble burst and the Lehman Brother pre-crash phase. Our portfolio represents the S&P 500 index since it comprises the 11 Global Industry Classification Standard (GICS) sectors in Tab. I (cf. [61]). Furthermore, we also sorted the stocks within the sectors according to the sub-industry sectors (see Tab. IV in App. A). In Ref. [20], we demonstrated that the market states depend on the choice of the stocks. Nevertheless, our portfolio allows us to make general statements about the correlation structure of the US stock markets because all GICS industry sectors [61] are covered by our portfolio and no industry sector is underrepresented, even not the real estate sector which mainly causes a market state transition in Ref. [20].

From adjusted closing prices $S_i(t)$, we calculate the daily logarithmic returns ($\Delta t = 1$ day)

$$G_i(t) = \ln \frac{S_i(t + \Delta t)}{S_i(t)}, \quad i = 1, \dots, K. \quad (1)$$

TABLE I. Global Industry Classification Standard (GICS).

Abbreviation	Sector	Number of companies
E	Energy	16
M	Materials	14
I	Industrials	43
CD	Consumer Discretionary	24
CST	Consumer Staples	24
HC	Health Care	28
F	Financials	36
RE	Real Estate	8
I	Information Technology	27
CSE	Communication Services	9
U	Utilities	21

We set up a $K \times T_{\text{tot}}$ data matrix for the total investigation period

$$G_{\text{tot}} = \begin{bmatrix} G_1(1) & \dots & G_1(T_{\text{tot}}) \\ \vdots & & \vdots \\ G_i(1) & \dots & G_i(T_{\text{tot}}) \\ \vdots & & \vdots \\ G_K(1) & \dots & G_K(T_{\text{tot}}) \end{bmatrix} \quad (2)$$

with $K = 250$, being the number of stocks and $T_{\text{tot}} = 4026$, being the total number of points in the return time series of a company. We do not use the full data matrix G_{tot} for the market state analysis. In order to analyze the non-stationarity of correlation matrices we select sub-blocks of the data matrix G_{tot} with all $K = 250$ stocks and intervals of $T_{\text{ep}} = 42$ trading days which correspond to 2 trading months. Only in the case of disjoint intervals we refer to these intervals as epochs.

III. REDUCED-RANK MATRICES

In Ref. [20], we introduced the covariance approach and the correlation approach. The covariance approach uses the standard covariance matrix, the correlation approach employs the standard correlation matrix, where the term “standard” refers to the original covariance and correlation matrix as obtained from the measured time series. In Sec. III A, we give a short introduction to the covariance approach, as well as in Sec. III B to the correlation approach.

A. Covariance approach

The starting point is the evaluation of the standard covariance matrix

$$\Sigma = \frac{1}{T} A A^\dagger = \sum_{i=1}^K \kappa_i u_i u_i^\dagger. \quad (3)$$

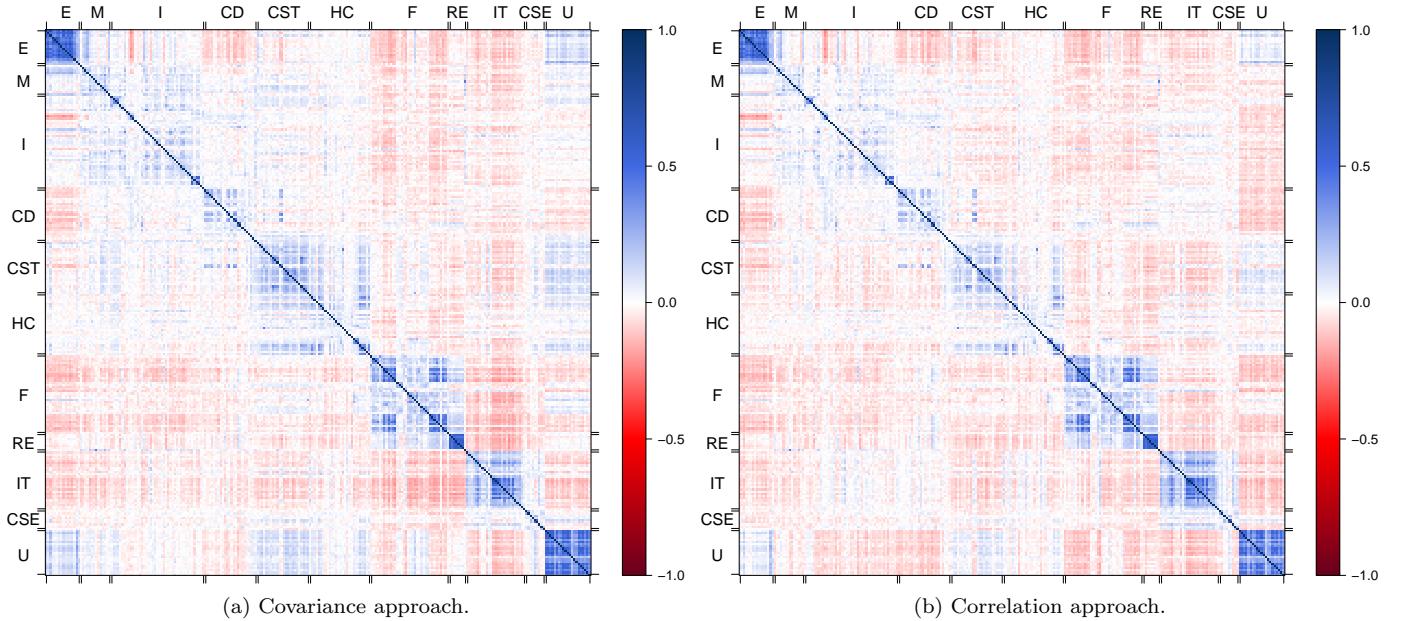


FIG. 1. Reduced-rank correlation matrices of $K = 250$ stocks for (a) the covariance approach and (b) the correlation approach. Both matrices are calculated for the 16 year period from 02 January, 1997 to 31 December, 2012. Capital Letters indicate industrial sectors (see Tab. I) (Data from QuoteMedia via Quandl).

The $K \times T$ data matrix A contains mean-normalized time series as rows. Additionally, we apply a spectral decomposition to the standard covariance matrix. Eigenvalues are denoted by κ_i and eigenvectors are denoted by u_i . The largest eigenvalue and the corresponding eigenvector can be interpreted as market part of Σ , while the other larger eigenvalues correspond to industrial sectors [16, 22–25, 29, 30, 42, 62]. Now, we subtract the dyadic matrix corresponding to the largest eigenvalue

$$\Sigma_B = \Sigma - \kappa_K u_K u_K^\dagger \quad (4)$$

arriving at the matrix Σ_B as an intermediate quantity. We mention in passing that Σ_B is a well-defined covariance matrix [20]. Using the standard deviations ordered in the diagonal matrix

$$\sigma_B = \text{diag}(\sigma_{B1}, \dots, \sigma_{BK}) , \quad (5)$$

we define the reduced-rank correlation matrix in the covariance approach

$$C_B = (\sigma_B)^{-1} \Sigma_B (\sigma_B)^{-1} . \quad (6)$$

In Fig. 1(a), the correlation matrix of the covariance approach is depicted for the entire 16 year period. It shows positively correlated block-diagonal entries corresponding to eleven industrial sectors and anti-correlations in the inter-sector structure.

B. Correlation approach

Here, the starting point is the calculation of the standard correlation matrix

$$C = \frac{1}{T} M M^\dagger = \sum_{i=1}^K \lambda_i x_i x_i^\dagger , \quad (7)$$

where M is a $K \times T$ data matrix whose rows are normalized to standard deviation one and mean value zero. Eigenvalues are denoted by λ_i and eigenvectors by x_i . Analogously to Eq. (6), by means of the matrix

$$\Sigma_L = C - \lambda_K x_K x_K^\dagger \quad (8)$$

and the diagonal matrix of standard deviations

$$\sigma_L = \text{diag}(\sigma_{L1}, \dots, \sigma_{LK}) , \quad (9)$$

we define the reduced-rank correlation matrix of the correlation approach

$$C_L = (\sigma_L)^{-1} \Sigma_L (\sigma_L)^{-1} . \quad (10)$$

The reduced-rank correlation matrix in the correlation approach is depicted in Fig. 1(b). It looks very similar to the one in the covariance approach for the entire 16 year period in Fig. 1(a). As shown in Ref. [20], around the Lehman Brother crisis, the reduced-rank correlation matrices for both approaches differ very much from each other for a one year time period.

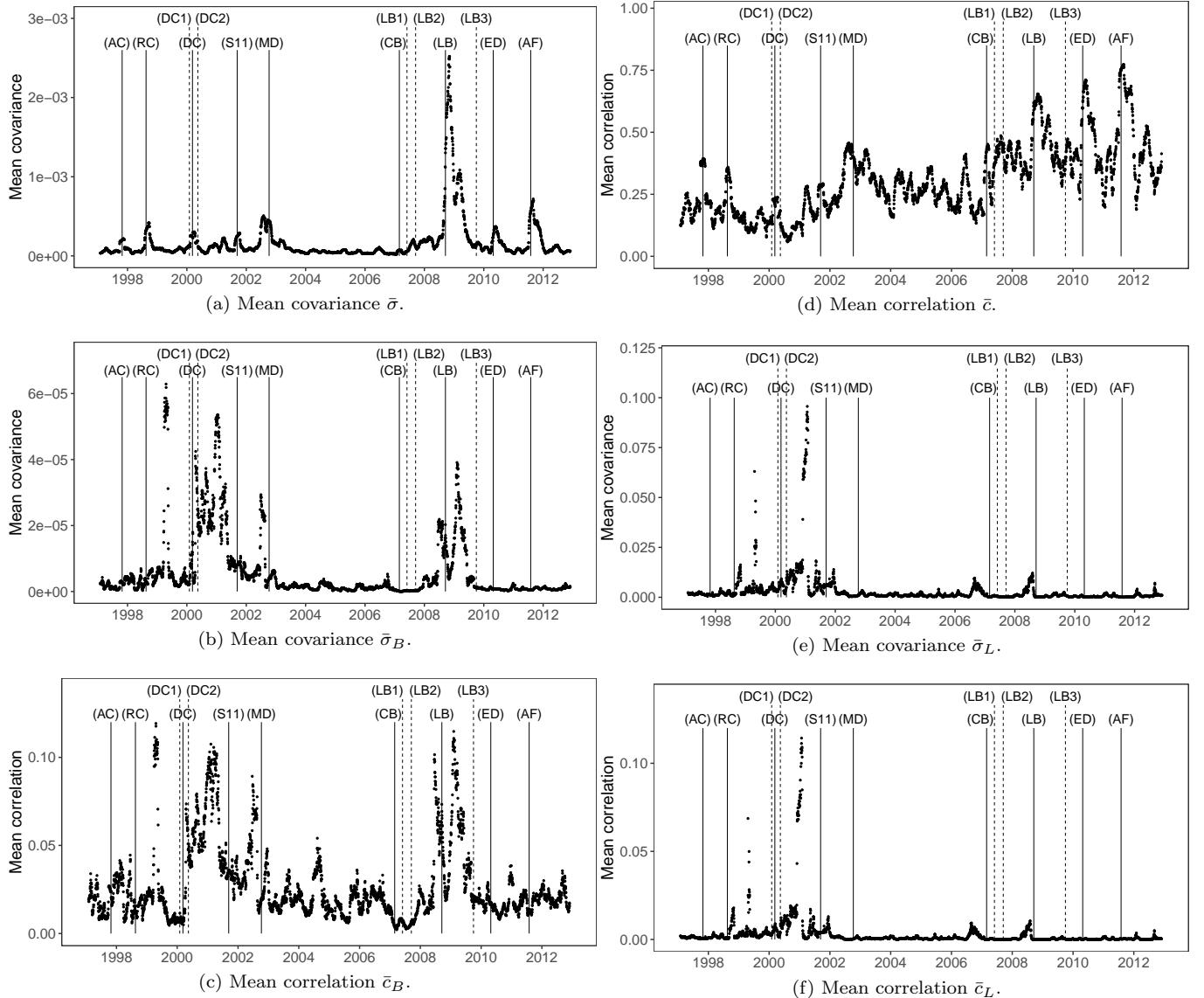


FIG. 2. Comparison of different mean values computed according to Eqs. (11) and (12): (a) mean covariance $\bar{\sigma}$, (b) mean covariance $\bar{\sigma}_B$, (c) mean correlation \bar{c}_B , (d) mean correlation \bar{c} , (e) mean covariance $\bar{\sigma}_L$ and (f) mean correlation \bar{c}_L . The beginning of the dot-com bubble burst is highlighted by historical event (DC) and the Lehman Brother crash is highlighted by (LB). Further label explanations for historical events (lower row) and estimated events (upper row) can be found in Tabs. II and III. Every dot stand for an interval of 42 trading days (Data from QuoteMedia via Quandl).

IV. DATA ANALYSIS AND RESULTS

In Sec. IV A, we briefly recapitulate our concept of market states. In Sec. IV B, we define mean values, distance matrices and averaged distances for the reduced-rank correlation matrices and visualize their temporal evolution. We compare the time series of these mean values and averaged distances with historical and estimated events in Sec. IV C in order to study transitions of market states. Other precursors are identified in Sec. IV D. In Sec. IV E, we investigate market states for the dot-com bubble burst and the pre-phase of the Lehman Brothers crises. Importantly, we do not use post-crisis data.

A. Concept of market states

We identify quasi-stationary structures in the time-dependent correlation matrices. We refer to them as market states. In previous works [9, 16–21, 36–40, 49], time periods of several decades are divided into so-called epochs (see Sec. II). Epochs are usually disjoint, *i.e.* non-overlapping intervals. The fixed length of each epoch is typically one or two trading months. For each epoch, we calculate a correlation matrix, a standard one or a reduced-rank one. We group correlation matrices of the same kind by employing a clustering algorithm, the k -means clustering algorithm [31–35]. We identify these

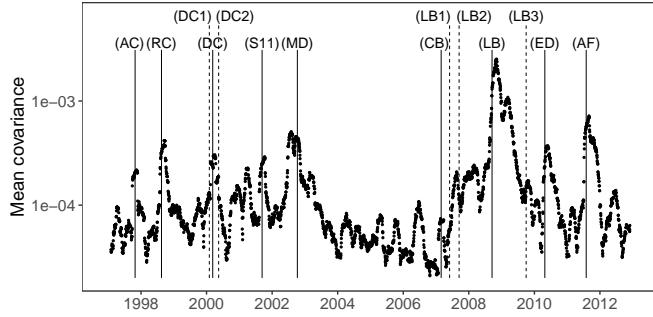


FIG. 3. Mean covariance $\bar{\sigma}$ plotted on a logarithmic scale in order to visualize the increasing and decreasing mean covariances around crisis listed in Tabs II. The other events can be found in Tab. III (Data from QuoteMedia via Quandl).

groups as market states which comprise correlation matrices of a similar structure. We do that for the standard and for the reduced-rank correlation matrices which show striking differences. In particular, we find longer life times in the latter ones. The market operates in a state for a certain time, then jumps to another one and yet another one and also might return to a state. Put differently, the states emerge, exist for some time, reappear and eventually disappear.

B. Definitions of mean values, distance matrices and averaged distances

Since we want to scrutinize what happens at the transitions of the market states we use overlapping 42 trading day to find signals in our data potentially being connected to such transitions. Relevant signals may be found in the mean values of all reduced-rank correlation matrices elements and in the distance matrices derived from the reduced-rank correlation matrices.

We set up $K \times T_{\text{ep}}$ data matrices A and M where $K = 250$ is the number of stocks and $T_{\text{ep}} = 42$ trading days is the length of one interval. The matrices are computed on a 42 trading day sliding window which is shifted forward by one trading day and moved over the 4026 trading days of the return time series. In total, we compute 3984 data matrices A and 3984 data matrices M . According to Sec. III, we calculate the standard covariance matrix Σ , the covariance matrix Σ_B , the reduced-rank correlation matrix C_B in the covariance approach, the standard correlation matrix C , the covariance matrix Σ_L in the covariance approach and the reduced-rank correlation matrix C_L in the correlation approach. The latter matrices are singular and their elements are strongly influenced by noise.

As a next step, the mean value of all matrix elements for every matrix is computed. For the covariance matrix

Σ , we take the average of all covariance matrix elements

$$\bar{\sigma} = \frac{1}{K^2} \sum_{k,l=1}^K \Sigma_{kl}. \quad (11)$$

Correspondingly, we define the mean covariance $\bar{\sigma}_B$ in the covariance approach and the mean covariance $\bar{\sigma}_L$ in the correlation approach. Since we include the diagonal elements in the average in Eq. (11), we analogously define the average for the standard correlation matrix C as

$$\bar{c} = \frac{1}{K^2} \sum_{k,l=1}^K C_{kl}. \quad (12)$$

For the reduced-rank correlation matrix C_B and C_L , the mean correlation is denoted by \bar{c}_B and \bar{c}_L , respectively. Finally, we generate time series of the mean covariances $\bar{\sigma}$, $\bar{\sigma}_B$, $\bar{\sigma}_L$ and the mean correlations \bar{c} , \bar{c}_B , \bar{c}_L . Each of these time series comprises 3984 mean values.

In Figs. 2(a), 2(b) and 2(c), the time evolutions of the mean values $\bar{\sigma}$, $\bar{\sigma}_B$ and \bar{c}_B are displayed on a linear scale. This is compared to the time evolutions of the mean values \bar{c} , $\bar{\sigma}_L$ and \bar{c}_L in Figs. 2(d), 2(e) and 2(f) on a linear scale as well. Each data matrix, correlation matrix and mean value receive a time stamp which corresponds to the center of a 42 trading day interval and which is represented by a black dot in Fig. 2. A logarithmic scale in Fig. 3 for mean covariance $\bar{\sigma}$ facilitates a comparison with its peaks. Both mean covariances $\bar{\sigma}$ and $\bar{\sigma}_B$ are positive for the entire 16 year time period. In contrast to the mean correlation \bar{c} , the mean covariance $\bar{\sigma}$ shows no trend to larger mean values. Although the mean correlations \bar{c}_B and \bar{c}_L can be small in contrast to \bar{c} , this does not imply that these are artifacts due to noise as we effectively take the average of $K(K-1)/2 = 31125$ correlation matrix elements, taking the symmetry and the unities on the diagonals of correlation matrices into account. For the mean covariances, the variances (squared volatilities) are on the diagonals and are time-dependent quantities.

We now turn to measuring distances between correlation matrices. As we use k -means clustering in our market state analysis which employs the Euclidean metric [31–35], we introduce a distance matrix – as defined for the standard correlation matrix in Refs. [9, 16] – by calculating the pairwise distance between two reduced-rank correlation matrices in the covariance approach

$$\zeta_B^{\text{Eucl}}(t_a, t_b) = \frac{\sqrt{\sum_{i,j} (C_{Bij}(t_a) - C_{Bij}(t_b))^2}}{K}. \quad (13)$$

Correspondingly, in the correlation approach, we define the Euclidean distance $\zeta_L^{\text{Eucl}}(t_a, t_b)$. Rows and columns are labeled by the indices $t_a = 1, \dots, 3984$ and $t_b = 1, \dots, 3984$. Thus, we obtain distance matrices of dimension 3984×3984 . Due to the normalization with K in Eq. (13) we may compare distance matrices for another selection of stocks.

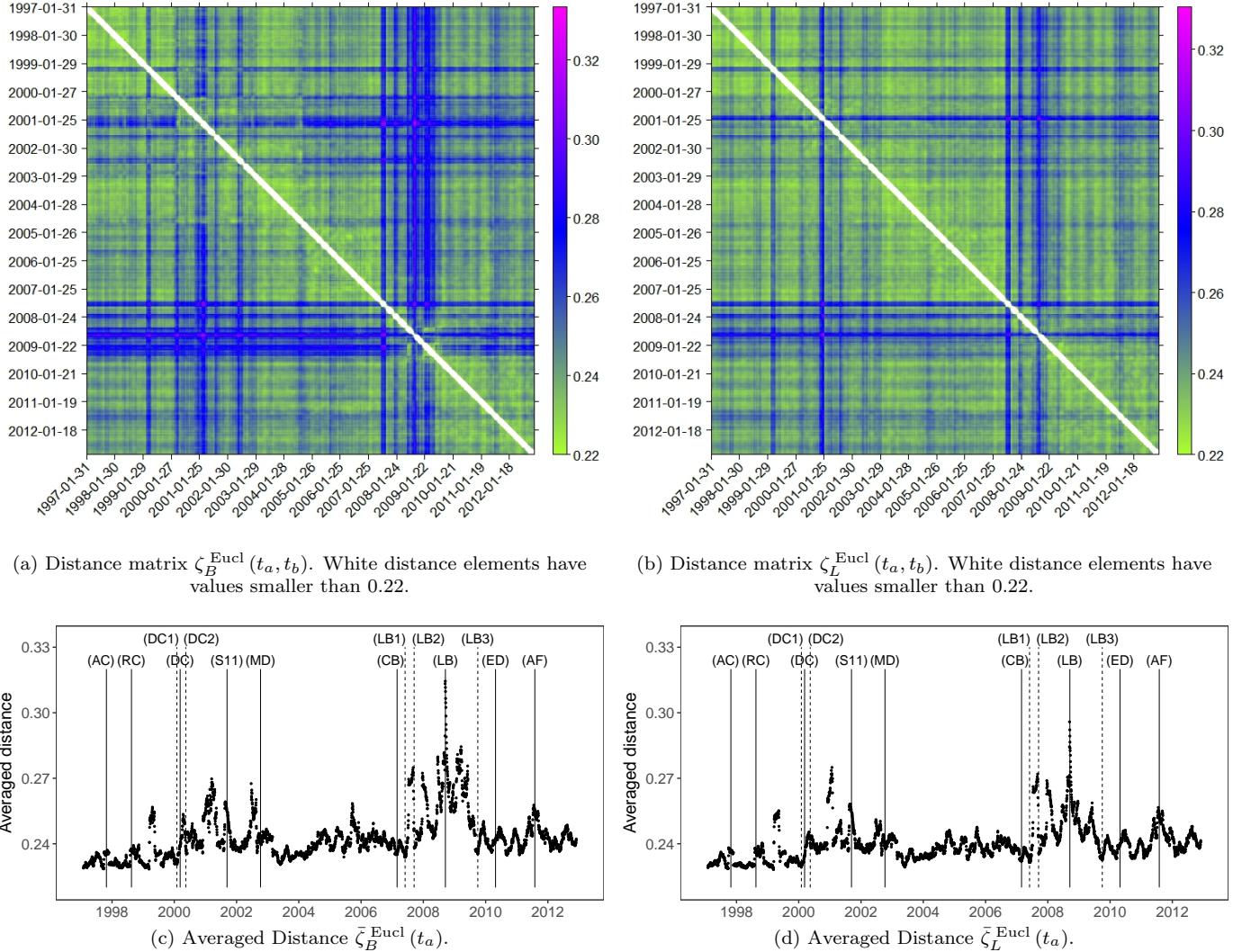


FIG. 4. Distance matrices (a) for the covariance approach $\zeta_B^{\text{Eucl}}(t_a, t_b)$ and (b) for the correlation approach $\zeta_L^{\text{Eucl}}(t_a, t_b)$ and averaged distances (c) for the covariance approach $\bar{\zeta}_B^{\text{Eucl}}(t_a)$ and (d) for the correlation approach $\bar{\zeta}_L^{\text{Eucl}}(t_a)$ calculated from the reduced-rank correlation matrices (see Sec. IV B). For the averaged distances, values smaller than 0.22 are excluded. Events in brackets are listed in Tabs. II and III (Data from QuoteMedia via Quandl).

In Figs. 4(a) and 4(b) the distance matrices calculated according to Eq. (13) are displayed. The larger the distance, the more dissimilar are two reduced-rank correlation matrices. The distance matrices derived from the standard correlation matrices show their highest values for financial crashes [9]. We observe a completely different dynamics for the distance matrices $\zeta_B^{\text{Eucl}}(t_a, t_b)$ and $\zeta_L^{\text{Eucl}}(t_a, t_b)$. Quasi-stationary periods of distances with values of approximately 0.24 are followed by less stable periods of larger distances with values larger than 0.26. The distance matrix $\zeta_B^{\text{Eucl}}(t_a, t_b)$ for the covariance approach shows a much more pronounced structure than that for the correlation approach $\zeta_L^{\text{Eucl}}(t_a, t_b)$.

We introduce a new time series — the *averaged distance* — to analyze the market state transitions in more

detail. Due to the characteristic stripes in Figs. 4(a) and 4(b), it is beneficial to take the average of the rows of a distance matrix since these stripes might contain information on possible market state transitions. Using the distance matrix $\zeta_B^{\text{Eucl}}(t_a, t_b)$ in Eq. (13), we calculate the averaged distance as

$$\bar{\zeta}_B^{\text{Eucl}}(t_a) = \frac{1}{t_c} \sum_{t'_b=1}^{t_c} \zeta_B^{\text{Eucl}}(t_a, t'_b). \quad (14)$$

keeping the row index t_a fixed and summing up to a specific column index t_c . The column index $t'_b = 1$ corresponds to the trading date 1997-01-31 and $t_c = 484$ to the trading date 1998-12-31. Analogously, we define the averaged distance $\bar{\zeta}_L^{\text{Eucl}}(t_a)$ for the correlation approach.

The averaged distances are depicted in Figs. 4(c) for the covariance approach and in Fig. 4(d) for the correlation approach.

In Figs. 4(a) and 4(b), we cut off distance matrix elements with values smaller than 0.22. Thereby, we exclude distances between overlapping reduced-rank correlation matrices in Eq. (14). By including values smaller than 0.22 we would calculate systematically lower values for the averaged distance in Eq. (14) from $t'_b = 1$ to t_c compared to the time period from t_c to $t_d = 3984$. We would create an artificial jump at t_c .

C. Historical and estimated events

We wish to compare the time evolution of the mean values and averaged distances with events which are on the one hand historical crisis events listed in Tab. II and on the other hand events in Tab. III estimated from the averaged distances in Sec. IV B. Those estimated events are connected to market state transitions and are highlighted as dashed lines in Figs. 2, 3 and 4.

The mean covariance $\bar{\sigma}$ in Fig. 2(a) shows its highest peaks for the Lehman Brother crash (LB) whereas the mean correlation \bar{c} in Fig. 2(d) has its largest value for the August 2011 stock market fall (AF). We want to emphasize that the trading day when the NASDAQ Composite stock market index peaked is labeled by (DC) [63]. Shortly after this event, the dot-com bubble bursted.

Our new analysis corroborates our results in Ref. [20]: We separate exogenous effects appearing in the mean correlation \bar{c} as crisis events (cf. Tab. II) from the endogenous parts remaining in the mean correlations \bar{c}_B and \bar{c}_L . The mean correlations \bar{c}_B and \bar{c}_L show a non-stationary behavior with sometimes even high values, especially in the case of the covariance approach (cf. Figs. 2(c) and 2(f)). We additionally calculated the mean covariances $\bar{\sigma}_B$ and $\bar{\sigma}_C$ which both show a separation from the historical events in their temporal behavior as well (cf. Figs. 2(b) and 2(e)).

The averaged distances in Figs. 4(c) and 4(d) facilitate the identification of signals in the distance matrix which are connected to market state transitions. The market state transitions occur for the dot-com bubble burst roughly between (DC1) and (DC2) and for the pre-phase of the Lehman Brother crash between (LB1) and (LB2) (cf. Tab. III). The duration of the market state transition period in the vicinity of historical event (DC) is 73 trading days and prior to (LB) 74 trading days. The market state transition for the Lehman Brother crisis market state appears in mid-2007 as it was observed in Ref. [55] for the absolute changes of the largest eigenvalues of the standard correlation matrices. The presumable cause was the freezing of the Interbank market [64]. This is a precursor signal exclusively for (LB).

The dashed lines (LB1) and (LB3) around (LB) show a high agreement with the recession period of December 2007 - June 2009 [65]. Information on such recession

periods is provided by the National Bureau of Economic Research (NBER) for the US economy. Apparently, we are able to detect connections to this recession period using reduced-rank correlation matrices on US stock markets. It is possible that the observables in our study are influenced by the recession period from March 2001 - November 2001 as well.

Additionally, Figs. 4(c) and 4(d), also make visible the historical events (AC), (RC), (S11) and (AF) and potentially (ED) in the averaged distances. The difference to the dot-com bubble burst and the Lehman Brother crisis is that these four to five crises are not accompanied by larger outburst in the mean covariances $\bar{\sigma}_B$, $\bar{\sigma}_L$ and the mean correlations \bar{c}_B and \bar{c}_L in their direct vicinity (see Fig. 2). In spite of removing to a large extent exogenous contributions in the mean covariances $\bar{\sigma}_B$ and $\bar{\sigma}_L$ and mean correlations \bar{c}_B and \bar{c}_L we find that the averaged distances still contain exogenous information of several financial crises.

TABLE II. Historical events taken from [66].

Label	Crisis	Date (Year-Month-Day)
(AC)	Asian financial crisis	1997-10-27
(RC)	Russian financial crisis	1998-08-17
(DC)	Dot-com bubble (before burst)	2000-03-10
(S11)	September 11th	2001-09-11
(MD)	Stock market downturn of 2002	2002-10-09
(CB)	Chinese stock bubble	2007-02-27
(LB)	Lehman Brothers crash	2008-09-16
(ED)	European debt crisis	2010-04-27
(AF)	August 2011 stock markets fall	2011-08-01

TABLE III. Dates for events estimated from the averaged distances introduced in Sec. IV B. First to fourth label highlight dates for begin and end of market state transitions (Data from QuoteMedia via Quandl).

Label	Description	Date (Year-Month-Day)
(DC1)	Start of market state transition for dot-com bubble burst	2000-02-01
(DC2)	End of market state transition for dot-com bubble burst	2000-05-15
(LB1)	Start of market state transition for Lehman Brothers crisis	2007-06-01
(LB2)	End of market state transition for Lehman Brothers crisis	2007-09-15
(LB3)	End of Lehman Brothers crisis period	2009-10-01

D. Other long-term precursors

The periods of the dot-com bubble burst between (DC1) and (DC2) and the pre-phase of the Lehman

Brother crash between (LB1) and (LB2) start with very low values of \bar{c}_B accompanied by an increasing mean covariance $\bar{\sigma}$ for (LB), also depicted on a logarithmic scale in Fig. 3. These very low values of \bar{c}_B coincide with market state transitions. In the case of the dot-com bubble burst a sudden mean correlation outburst in \bar{c}_B appears at (DC1). For the Lehman pre-phase, however, the outburst of \bar{c}_B does not happen at (LB1) but shortly before (LB). The low values for the mean correlation \bar{c}_B are precursor signals for (DC) and (LB).

The sudden decreases in the mean covariances $\bar{\sigma}_B$ and $\bar{\sigma}_L$ and in the mean correlations \bar{c}_B and \bar{c}_L at (LB) indicate a connection to the Lehman Brother crash whereas the mean covariance $\bar{\sigma}$ as a measure for the entire portfolio risk and the mean correlation \bar{c} have peaks. The financial sector crisis spread over to the entire market, leading to a market-wide crash. This can be interpreted as a spill-over effect. Therefore, $\bar{\sigma}_B$, $\bar{\sigma}_L$, \bar{c}_B and \bar{c}_L show features of potential measures for systemic risk. We do not observe such a spill-over effect for the dot-com bubble. The largest peaks in the averaged distances in Figs. 4(c) and 4(d) at (LB) coincide with lower mean correlations \bar{c}_B and \bar{c}_L . This spill-over effect is a precursor signal exclusively for (LB).

Comparing the averaged distances in Figs. 4(c) and 4(d) and the time evolution of the mean correlations \bar{c}_B and \bar{c}_L in Figs. 2(c) and 2(f) we see that changes in the averaged distances occur prior to the changes in the mean correlation at a market state transition to the crisis period for (LB). In the pre-phase of the Lehman Brother crisis, the mid-2007 peak discussed in Sec. IV C does not coincide with a larger change in mean correlations. The spill-over effect appears later. In the case of the dot-com bubble burst, we observe for the correlation approach between (DC1) and (DC2) that a change in the averaged distance is accompanied by an outburst of the mean correlation \bar{c}_B .

E. Market state transitions as long-term precursors of financial crises

In Ref. [20], we divided a 15 year time period into epochs (time periods of equal length) with $T_{\text{ep}} = 42$ trading days and set up a $K \times T_{\text{ep}}$ data matrix A and M with $K = 262$ and $T_{\text{ep}} = 42$ for each epoch. Then we calculated the reduced-rank correlation matrices for these epochs and clustered all reduced-rank correlation matrices by the k -means algorithm [31–35], resulting in different market states.

Here, we follow a different method which we want to explain in the case of the Lehman Brother crash (LB) by means of Figs. 5(a)–5(e) for the covariance approach. We divide a time period of several years before the Lehman Brother crash (LB) into epochs with $T_{\text{ep}} = 42$ trading days. Figs. 5(a)–5(e) have in common that the starting epoch of the time period is fixed. In Fig. 5(a), the last epoch is seven epochs before (LB) and in Fig. 5(e)

one epoch before (LB). For each epoch, we calculate a reduced-rank correlation matrix of dimension 250×250 and cluster all reduced-rank correlation matrices employing the k -means clustering algorithm for $k = 2$. We refer to Figs. 5(a)–5(e) as “snapshots”. All snapshots together for one approach form a sequence of snapshots.

For the first snapshot in Fig. 5(a), we divide the time period from 2004-01-15 to 2007-07-18 into 21 epochs. The last epoch is in the vicinity of where we observe a small value of the mean value \bar{c}_B (see Sec. IV D), *i.e.* between (LB1) and (LB2) (see Tab. III). In Fig. 5(b), we divide the time period from 2004-01-15 to 2007-09-17 into 22 epochs. We cluster 22 reduced-rank correlation matrices, *i.e.* additionally, we cluster one more reduced-rank correlation matrix corresponding to an epoch of 42 trading days. In contrast to Fig. 5(a), we observe in Fig. 5(b) a drastic change in the market state evolution. For the first 21 epochs, all reduced-rank correlation matrices are clustered into the first market state. The second market state consists of one reduced-rank correlation matrix located between (LB1) and (LB2). In Fig. 5(c), we consider 25 epochs, *i.e.* three more reduced-rank correlation matrices compared to Fig. 5(b). The two remaining reduced-rank correlation matrices in Figs. 5(d) and 5(e) are assigned to the second market state as well, revealing a certain stability in the market states.

The sequence of snapshots for the correlation approach in Figs. 5(f)–5(j) looks very similar compared to the one for the covariance approach in Figs. 5(a)–5(e). For both approaches, there are three epochs between the last epoch in Figs. 5(c) and 5(h) and event (LB) which is half of a trading year. We are able to detect such a structural change without using post-crash data, thereby demonstrating precursor properties of the reduced-rank correlation matrices.

We obtain very similar results for the covariance and the correlation approach which is supported by the plots in Fig. 6. Here, we show so-called *typical* market states [9, 20] as element-wise average of all reduced-rank correlation matrices of a single market state for the second market state and for different snapshots. Corresponding typical second market states for the covariance and the correlation approach are very similar.

For dot-com bubble burst in Fig. 7, we are not able to identify a market state transition before (DC). Major correlation structure changes appear afterwards. The clustered period starts for all snapshots at 1997-01-13. Estimated events (DC1) and (DC2) (see Tab. III) highlight the market state transition and small values of mean value \bar{c}_B (cf. Sec. IV D). Instead, we take into account epochs after (DC) and analyze for which snapshots the market state transitions can be detected. In contrast to the Lehman Brother crash, the temporal evolutions of the typical market states within the snapshots for the covariance approach differ qualitatively from those of the correlation approach in the case of the dot-com bubble burst. It is intriguing that we are able to detect a first market state transition after two epochs in Fig. 7(b) which

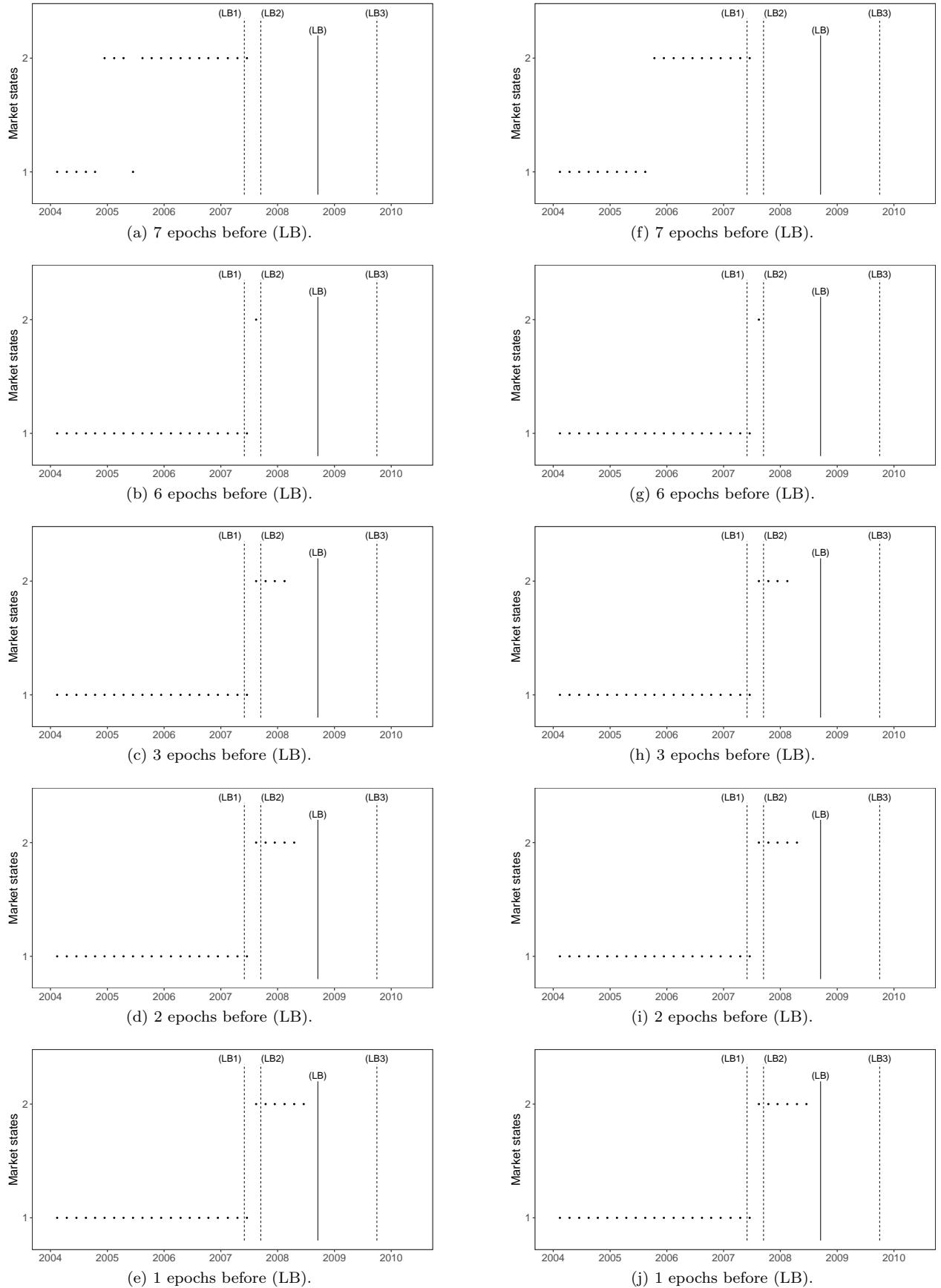


FIG. 5. Snapshots of the market state analysis for the reduced-rank correlation matrices of the (a)-(e) covariance approach and (f)-(j) correlation approach compared to estimated events around the Lehman Brother crash (LB) (for lower row, see Tab. II). For all snapshots, the first epoch is fixed. The clustered period starts for all snapshots at 2004-01-15. The estimated events in the upper row (LB1), (LB2) and (LB3) can be found in Tab. III. Every dot stands for an epoch of 42 trading days. From subplots (a) to (j), the number of the remaining epochs before (LB) is specified (Data from QuoteMedia via Quandl).

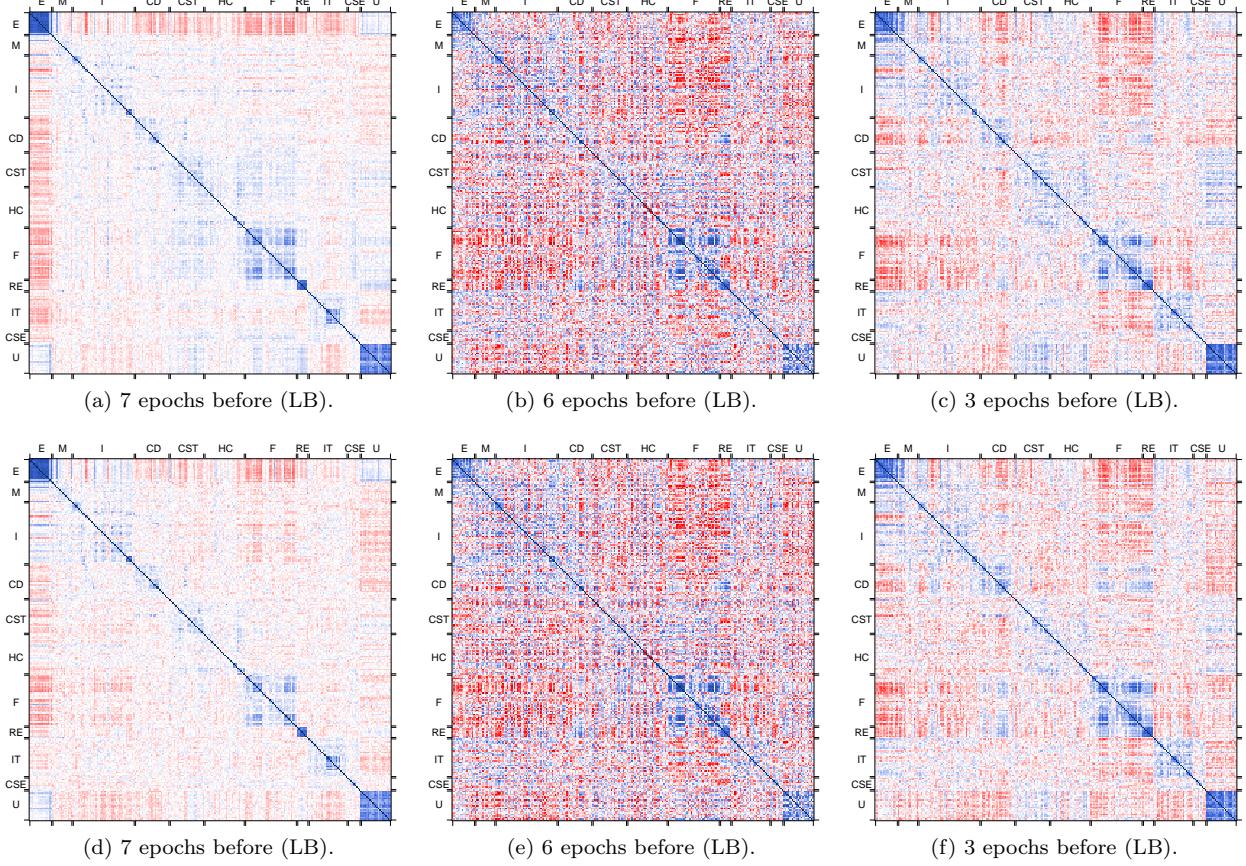


FIG. 6. Typical second market states ($K = 250$ stocks) corresponding to the snapshots before the Lehman Brother crash marked by historical event (LB) (cf. Fig. 5) for the reduced-rank correlation matrices of the (a)-(c) covariance approach and (d)-(f) correlation approach. From subplots (a) to (f), the number of the remaining epochs before (LB) is specified. A typical market state is obtained by taking the element-wise average of the correlation matrices of the respective market state. Capital Letters indicate industrial sectors (see Tab. I) (Data from QuoteMedia via Quandl).

is much earlier than the six epochs after (DC) for the correlation approach in Fig. 7(g).

We aim to compare the mean correlation \bar{c}_B with a time series mimicking the NASDAQ index [67]. We construct this time series by taking the average of the adjusted daily closing prices belonging to the 27 IT companies (cf. Fig. I). In Fig. 8, this self-constructed index is displayed (cf. Tab. I and Tab. IV in App. A). The end date of the second epoch after (DC) is 2000-07-11 and highlighted by the label (EP2); the end date of the fourth epoch after (DC) is 2000-11-07 and highlighted by the label (EP4). Event (EP2) lies before a major drop of the dot-com bubble burst in Fig. 8. It is interesting to compare this observation with the temporal evolution of the mean values in Sec. IV B. In Fig. 2(c) the mean correlation \bar{c}_B remains at a relatively high level whereas the mean covariance $\bar{\sigma}$ is decreasing in Fig. 2(a) and Fig. 3 on a logarithmic scale. The mean correlation \bar{c}_B seems to indicate that there is still an endogenous risk in the IT sector which finally results in the market drop. This is another result which adds to the observation of Sec. IV D that the mean correlation \bar{c}_B is a potential measure for

systemic risk.

We can also observe in the correlation structure of the typical second market states in Fig. 9 that both market state transitions are different for the two approaches. The anti-correlations from the IT sector with the other ten sectors as in Fig. 9(c) also appear as a feature of the reduced-rank correlation matrices in the covariance approach around the Lehman Brother crash according to Ref. [20]. Based on this observation and the historical events during the time period, the second market state in the time period of the dot-com bubble burst can be referred to as “crisis state” as well.

V. CONCLUSION

We studied the dynamics of reduced-rank correlation matrices in the covariance and the correlation approach and found long-term precursor properties. The dynamics of the correlation structure was analyzed in two different ways. On the one hand we looked at the market states of the reduced-rank correlation matrices, on the other

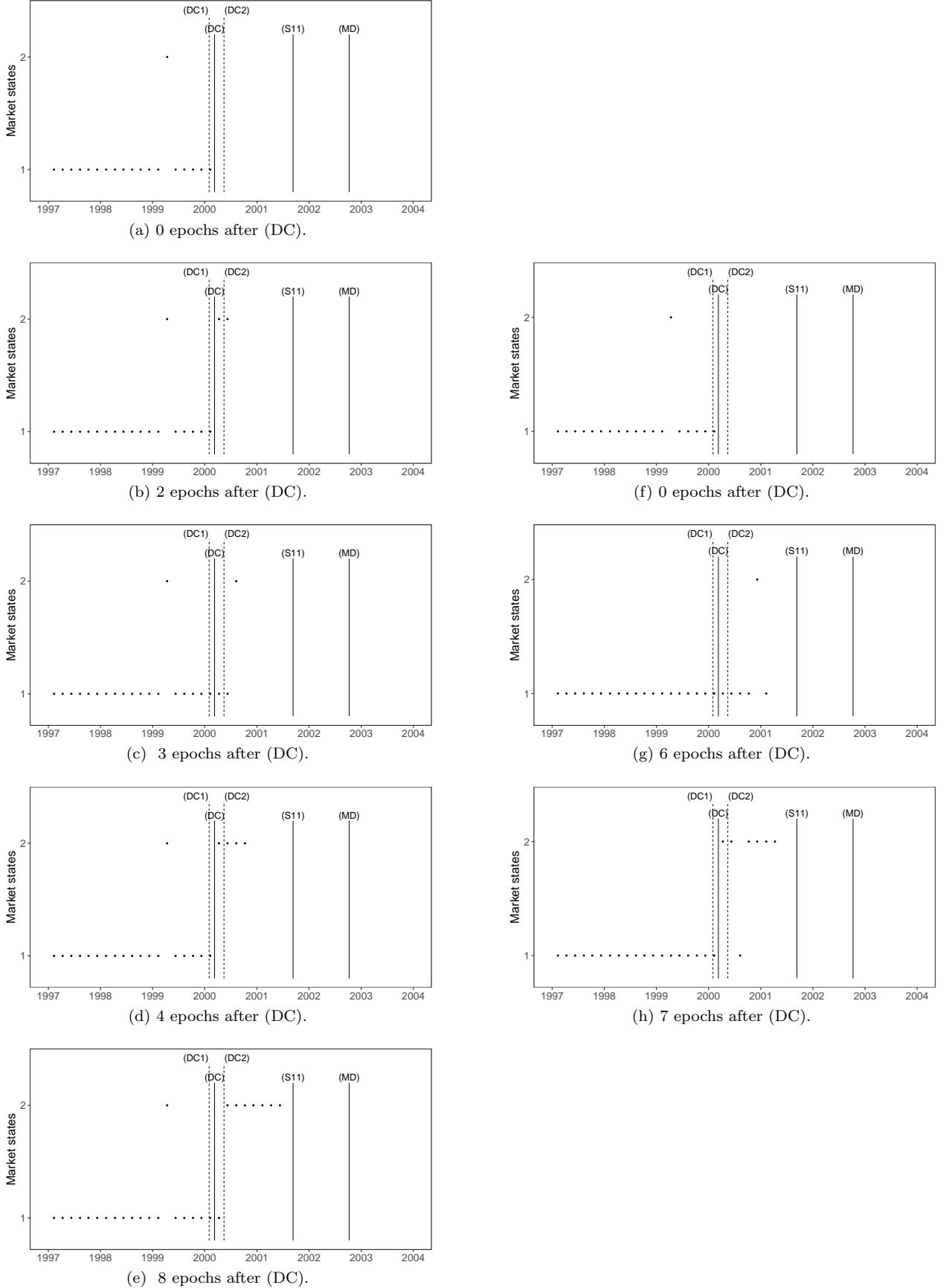


FIG. 7. Snapshots of the market state analysis for the reduced-rank correlation matrices of the (a)-(e) covariance approach and (f)-(h) correlation approach compared to estimated events at the beginning of the dot-com bubble burst marked by historical event (DC) (for lower row, see Tab. II). For all snapshots, the first epoch is fixed. The clustered period starts for all snapshots at 1997-01-13. The estimated events in the upper row (DC1) and (DC2) can be found in Tab. III. Every dot stands for an epoch of 42 trading days. From subplots (a) to (h), the number of the epochs after (DC) is specified (Data from QuoteMedia via Quandl).

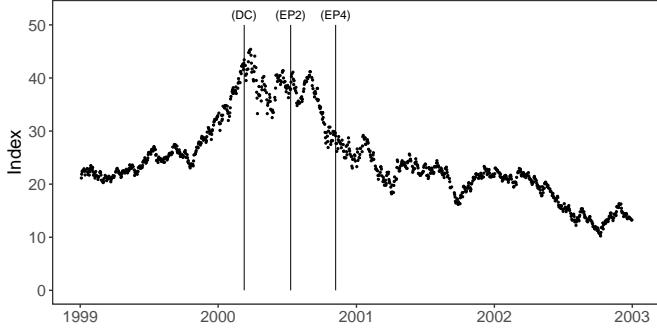


FIG. 8. Self-constructed index as average of the adjusted daily closing prices (see Sec. II) of the 27 IT companies (see Tab. I and Tab. IV in App. A). The dates for the end of the epochs are for (EP2) 2000-07-11 (two epochs after (DC)) and for (EP4) 2000-11-07 (four epochs after (DC)) (Data from Quote-Media via Quandl).

hand we were able to relate the market state transitions to sometimes large peaks or drastic changes in the mean values of covariance and correlation approach and in the corresponding averaged distances.

We have introduced a new method for analyzing market states. With our technique of snapshots for the market state analysis, we can follow the market state transitions by adding reduced-rank correlation matrices of new epochs to the market state analysis, thereby following the trajectories of the reduced-rank correlation matrices before or within crises periods. Our market state analysis is able to detect market state transitions belonging to the Lehman Brother crash occurred. We exclusively used pre-crisis data. The snapshots for the covariance and correlation approach look very much alike. For the dot-com bubble burst both approaches reveal differences concerning the market state dynamics.

We identified structures in the averaged distances which coincide with market state transitions in our cluster analysis. The mid-2007 event (freezing of the Interbank market) is the very first precursor signal starting the Lehman Brother crisis state. The burst of the dot-com bubble marks the beginning of a crisis state as well.

By comparing the mean correlations in the covariance and correlation approach with historical events we saw that the mean correlations (especially in the covariance approach) reflect the dynamics of an endogenous risks for both crises. Points of minimum correlations in the covariance approach indicate the beginning of two crises periods. From these points onwards, the endogenous risk builds up. Thus it is conceivable that a market state transition takes place since a new economical period begins in terms of the endogenous risk dynamics.

Our new market state method relates precursors of different kinds. In the market state analysis, transitions into a crises market state are connected to low mean correlations in the covariance approach and sudden changes in the averaged distances.

Furthermore, we found a period around the Lehman

Brother crash coinciding with a recession. Usually, recession periods are calculated with the gross domestic product (GDP) and techniques such as the Hodrick-Prescott filter [68–70]. Analogously, by subtracting the dyadic matrix corresponding to the largest eigenvalue, we also separated the quickly changing market motion from the more stable sectoral one.

For the Lehman Brothers crisis state, changes in the averaged distances build up prior to the changes in the mean correlation at a market state transition. The interpretation for this phenomenon might be that economical changes like the freezing of the Interbank market [64] are first visible which is supported by our identification of a “recession market state”. The endogenous risk increases as some kind of “economical tension” in the market which builds up and finally dissipates. This risk is potentially contagious for the entire market. In the case of the Lehman Brother crash it might be viewed as a spill-over effect. All our observations lead to the conclusion that the mean correlations for the reduced-rank correlation matrices in both approaches describe to some extent the fragility of a portfolio being exposed to a potentially larger risk initialized by an industrial sector. In the covariance approach, we observed anti-correlations between the IT-sector and the other ten sectors during the dot-com bubble burst. Such anti-correlations are also visible in the Lehman Brother crash pre-phase between the financial sector and the other ten sectors in Ref. [20]. Therefore, the above mentioned endogenous risks, visible in the mean correlations in covariance and correlation approach, are potential measures for systemic risk.

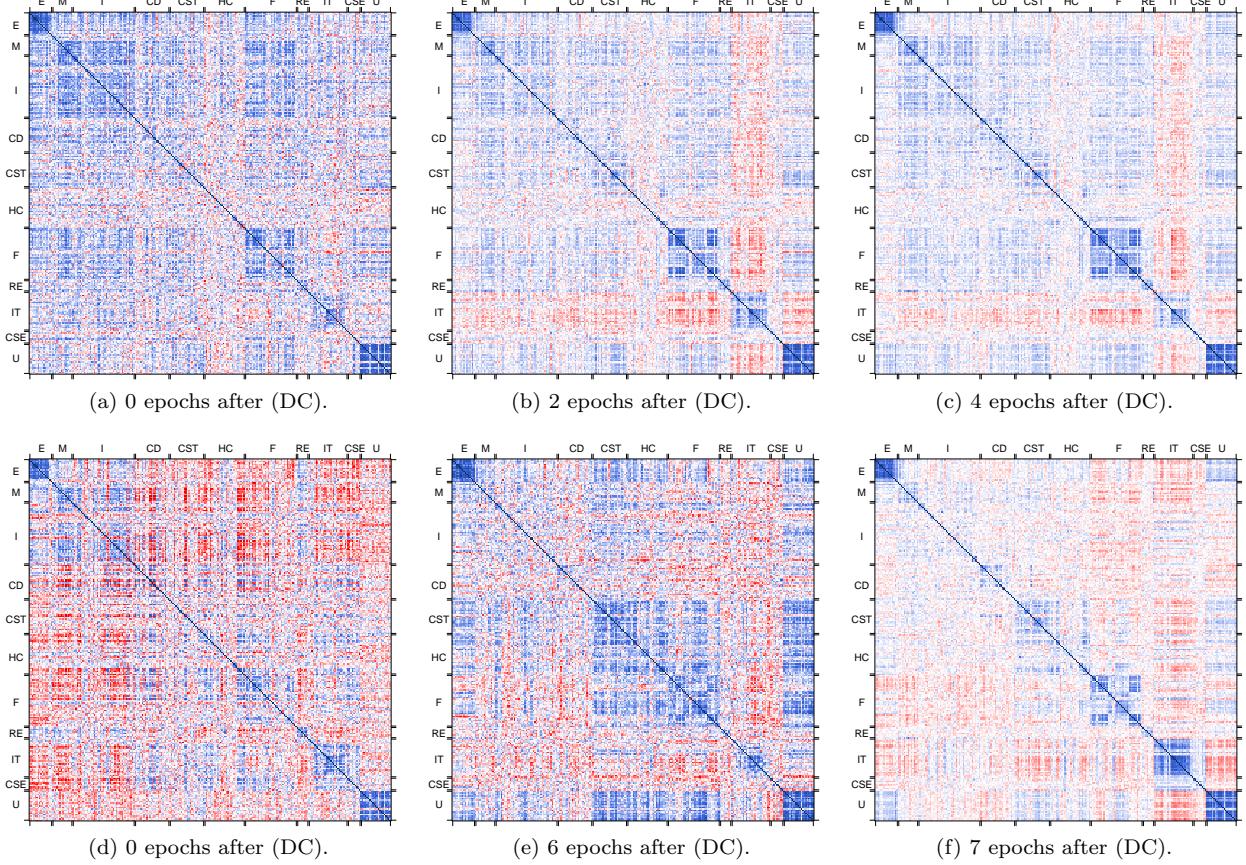


FIG. 9. Typical second market states ($K = 250$ stocks) corresponding to the snapshots of the dot-com bubble burst marked by historical event (DC) (cf. Fig. 5) for the reduced-rank correlation matrices of the (a)-(c) covariance approach and (d)-(f) correlation approach. A typical market state is obtained by taking the element-wise average of the correlation matrices of the respective market state. From subplots (a) to (f), the number of the epochs after (DC) is specified. Capital Letters indicate industrial sectors (see Tab. I) (Data from QuoteMedia via Quandl).

-
- [1] B. B. Mandelbrot, The variation of certain speculative prices, in *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E* (Springer New York, New York, NY, 1997) pp. 371–418.
- [2] G. W. Schwert, Why does stock market volatility change over time?, *The journal of finance* **44**, 1115 (1989).
- [3] F. Longin and B. Solnik, Is the correlation in international equity returns constant: 1960–1990?, *Journal of International Money and Finance* **14**, 3 (1995).
- [4] R. N. Mantegna and H. E. Stanley, *Introduction to econophysics: correlations and complexity in finance* (Cambridge university press, 1999).
- [5] J.-P. Bouchaud and M. Potters, *Theory of financial risk and derivative pricing: from statistical physics to risk management* (Cambridge University Press, 2003).
- [6] R. Kutner, M. Ausloos, D. Grech, T. Di Matteo, C. Schinckus, and H. Eugene Stanley, Econophysics and sociophysics: Their milestones & challenges, *Physica A: Statistical Mechanics and its Applications* **516**, 240 (2019).
- [7] N. Musmeci, T. Aste, and T. Di Matteo, Interplay between past market correlation structure changes and future volatility outbursts, *Scientific reports* **6**, 36320 (2016).
- [8] J. Y. Campbell, A. W. Lo, and A. C. MacKinlay, *The Econometrics of Financial Markets* (Princeton Univers. Press, 1997).
- [9] M. C. Münnix, T. Shimada, R. Schäfer, F. Leyvraz, T. H. Seligman, T. Guhr, and H. E. Stanley, Identifying states of a financial market, *Scientific Reports* **2**, 644 (2012).
- [10] J. D. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**, 357 (1989).
- [11] J. D. Hamilton, Analysis of time series subject to changes in regime, *Journal of Econometrics* **45**, 39 (1990).
- [12] H. Schaller and S. V. Norden, Regime switching in stock market returns, *Applied Financial Economics* **7**, 177 (1997).
- [13] M. Marsili, Dissecting financial markets: sectors and states, *Quantitative Finance* **2**, 297 (2002).
- [14] J. Jurczyk, T. Rehberg, A. Eckrot, and I. Morgenstern, Measuring critical transitions in financial markets, *Scientific reports* **7**, 1 (2017).
- [15] G. Marti, F. Nielsen, M. Bińkowski, and P. Donnat, A review of two decades of correlations, hierarchies, networks and clustering in financial markets (2020), arXiv:1703.00485 [q-fin.ST].
- [16] Y. Stepanov, P. Rinn, T. Guhr, J. Peinke, and R. Schäfer, Stability and hierarchy of quasi-stationary states: financial markets as an example, *Journal of Statistical Mechanics: Theory and Experiment* **2015**, P08011 (2015).
- [17] P. Rinn, Y. Stepanov, J. Peinke, T. Guhr, and R. Schäfer, Dynamics of quasi-stationary systems: Finance as an example, *EPL (Europhysics Letters)* **110**, 68003 (2015).
- [18] D. Chetalova, R. Schäfer, and T. Guhr, Zooming into market states, *Journal of Statistical Mechanics: Theory and Experiment* **2015**, P01029 (2015).
- [19] D. Chetalova, M. Wollschläger, and R. Schäfer, Dependence structure of market states, *Journal of Statistical Mechanics: Theory and Experiment* **2015**, P08012 (2015).
- [20] A. J. Heckens, S. M. Krause, and T. Guhr, Uncovering the dynamics of correlation structures relative to the collective market motion, *Journal of Statistical Mechanics: Theory and Experiment* **2020**, 103402 (2020).
- [21] S. Wang, S. Gartzke, M. Schreckenberg, and T. Guhr, Quasi-stationary states in temporal correlations for traffic systems: Cologne orbital motorway as an example, *Journal of Statistical Mechanics: Theory and Experiment* **2020**, 103404 (2020).
- [22] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Noise dressing of financial correlation matrices, *Phys. Rev. Lett.* **83**, 1467 (1999).
- [23] J. D. Noh, Model for correlations in stock markets, *Phys. Rev. E* **61**, 5981 (2000).
- [24] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley, Quantifying and interpreting collective behavior in financial markets, *Phys. Rev. E* **64**, 035106 (2001).
- [25] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, Random matrix approach to cross correlations in financial data, *Phys. Rev. E* **65**, 066126 (2002).
- [26] C. Borghesi, M. Marsili, and S. Miccichè, Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode, *Phys. Rev. E* **76**, 026104 (2007).
- [27] D. Y. Kenett, Y. Shapira, and E. Ben-Jacob, RMT assessments of the market latent information embedded in the stocks' raw, normalized, and partial correlations, *Journal of Probability and Statistics* **2009**, 10.1155/2009/249370 (2009).
- [28] Y. Shapira, D. Y. Kenett, and E. Ben-Jacob, The index cohesive effect on stock market correlations, *The European Physical Journal B* **72**, 657 (2009).
- [29] M. MacMahon and D. Garlaschelli, Community detection for correlation matrices, *Phys. Rev. X* **5**, 021006 (2015).
- [30] M. Benzaquen, I. Mastromatteo, Z. Eisler, and J.-P. Bouchaud, Dissecting cross-impact on stock markets: an empirical analysis, *Journal of Statistical Mechanics: Theory and Experiment* **2017**, 023406 (2017).
- [31] H. Steinhaus, Sur la division des corp matériels en parties, *Bull. Acad. Polon. Sci.* **IV**, 801 (1956).
- [32] G. H. Ball and D. J. Hall, *ISODATA, a novel method of data analysis and pattern classification*, Tech. Rep. (Technical report NTIS AD 699616. Standford Research Institute, Standford, CA., 1965).
- [33] J. MacQueen, Some methods for classification and analysis of multivariate observations, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (Univerity of California Press, 1967) pp. 281–297.
- [34] S. P. Lloyd, Least squares quantization in pcm, *IEEE Transactions on Information Theory* **28**, 129 (1982).
- [35] A. K. Jain, Data clustering: 50 years beyond k-means, *Pattern Recognition Letters* **31**, 651 (2010), award winning papers from the 19th International Conference on Pattern Recognition (ICPR).
- [36] J. Papenbrock and P. Schwendner, Handling risk-on/risk-off dynamics with correlation regimes and correlation networks, *Financial Markets and Portfolio Management* **29**, 125 (2015).
- [37] L. Qiu, C. Gu, Q. Xiao, H. Yang, and G. Wu, State net-

- work approach to characteristics of financial crises, *Physica A: Statistical Mechanics and its Applications* **492**, 1120 (2018).
- [38] H. K. Pharasi, E. Seligman, and T. H. Seligman, Market states: A new understanding (2020), arXiv:2003.07058 [q-fin.CP].
- [39] H. K. Pharasi, E. Seligman, S. Sadhukhan, and T. H. Seligman, Dynamics of market states and risk assessment (2020), arXiv:2011.05984 [q-fin.ST].
- [40] H. K. Pharasi, S. Sadhukhan, P. Majari, A. Chakraborti, and T. H. Seligman, Dynamics of the market states in the space of correlation matrices with applications to financial markets (2021), arXiv:2107.05663 [q-fin.ST].
- [41] T. Rings, M. Mazarei, A. Akhshi, C. Geier, M. R. R. Tabar, and K. Lehnertz, Traceability and dynamical resistance of precursor of extreme events, *Scientific reports* **9**, 1 (2019).
- [42] D.-M. Song, M. Tumminello, W.-X. Zhou, and R. N. Mantegna, Evolution of worldwide stock markets, correlation structure, and correlation-based graphs, *Phys. Rev. E* **84**, 026108 (2011).
- [43] O. Alter, P. O. Brown, and D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling, *Proceedings of the National Academy of Sciences* **97**, 10101 (2000), <https://www.pnas.org/content/97/18/10101.full.pdf>.
- [44] D.-H. Kim and H. Jeong, Systematic analysis of group identification in stock markets, *Phys. Rev. E* **72**, 046133 (2005).
- [45] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, Random matrix theory and financial correlations, *International Journal of Theoretical and Applied Finance* **3**, 391 (2000).
- [46] M. Miceli and G. Susinno, Ultrametricity in fund of funds diversification, *Physica A: Statistical Mechanics and its Applications* **344**, 95 (2004), applications of Physics in Financial Analysis 4 (APFA4).
- [47] M. Tumminello, F. Lillo, and R. N. Mantegna, Kullback–Leibler distance as a measure of the information filtered from multivariate data, *Phys. Rev. E* **76**, 031123 (2007).
- [48] M. Tumminello, F. Lillo, and R. N. Mantegna, Shrinkage and spectral filtering of correlation matrices: a comparison via the Kullback–Leibler distance, *Acta Physica Polonica B* **38**, 4079 (2007).
- [49] H. K. Pharasi, K. Sharma, R. Chatterjee, A. Chakraborti, F. Leyvraz, and T. H. Seligman, Identifying long-term precursors of financial market crashes using correlation patterns, *New Journal of Physics* **20**, 103041 (2018).
- [50] D. Sornette, Y. Malevergne, and J.-F. Muzy, What causes crashes?, *Risk Magazine* **16**, 67 (2003).
- [51] J.-P. Bouchaud, The endogenous dynamics of markets: price impact and feedback loops (2010), arXiv:1009.2928 [q-fin.ST].
- [52] J. Danielsson, H. S. Shin, and J.-P. Zigrand, Endogenous extreme events and the dual role of prices, *Annual Review of Economics* **4**, 111 (2012).
- [53] J.-P. Bouchaud, The endogenous dynamics of markets: A complex system point of view, *Procedia Computer Science* **7**, 22 (2011), proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11).
- [54] D. Bisias, M. Flood, A. W. Lo, and S. Valavanis, A survey of systemic risk analytics, *Annual Review of Financial Economics* **4**, 255 (2012), <https://doi.org/10.1146/annurev-financial-110311-101754>.
- [55] Z. Zheng, B. Podobnik, L. Feng, and B. Li, Changes in cross-correlations as an indicator for systemic risk, *Scientific reports* **2**, 1 (2012).
- [56] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, *Journal of Financial Economics* **104**, 535 (2012), market Institutions, Financial Market Risks and Financial Crisis.
- [57] M. Kritzman, Y. Li, S. Page, and R. Rigobon, Principal components as a measure of systemic risk, *The Journal of Portfolio Management* **37**, 112 (2011).
- [58] S. Begušić, Z. Kostanjčar, D. Kovač, H. E. Stanley, and B. Podobnik, Information feedback in temporal networks as a predictor of market crashes, *Complexity* **2018**, <https://doi.org/10.1155/2018/2834680> (2018).
- [59] QuoteMedia, Stock market data and financial research solutions, <https://www.quotemedia.com/>, (Online; accessed 18-September-2020).
- [60] Quandl, <https://www.quandl.com/> (Online; accessed 18-September-2020), the source for financial, economic, and alternative datasets, serving investment professionals.
- [61] Wikipedia contributors, Global industry classification standard — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=Global_Industry_Classification_Standard&oldid=996081362 (2020), [Online; accessed 10-January-2021].
- [62] M. Potters and J.-P. Bouchaud, *A First Course in Random Matrix Theory: for Physicists, Engineers and Data Scientists* (Cambridge University Press, 2020).
- [63] Wikipedia contributors, Dot-com bubble — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=Dot-com_bubble&oldid=997314485 (2020), [Online; accessed 5-January-2021].
- [64] Investopedia, The 2007–2008 financial crisis in review, <https://www.investopedia.com/articles/economics/09/financial-crisis-review.asp#axzz2Djve2G8Q>, (Online; accessed 19-February-2021).
- [65] Wikipedia contributors, List of recessions in the United States — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=List_of_recessions_in_the_United_States&oldid=994913408 (2020), [Online; accessed 5-January-2021].
- [66] Wikipedia contributors, List of stock market crashes and bear markets — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=List_of_stock_market_crashes_and_bear_markets&oldid=995044267 (2020), [Online; accessed 5-January-2021].
- [67] W. Commons, File:dead cat bounce nasdaq_ixic_dot-com bubble.png — wikimedia commons, the free media repository, https://commons.wikimedia.org/w/index.php?title=File:Dead_cat_bounce_NASDAQ_ixic_dot-com_bubble.png&oldid=511050763 (2020), [Online; accessed 15-January-2021].
- [68] R. J. Hodrick and E. C. Prescott, Postwar us business cycles: an empirical investigation, *Journal of Money, credit, and Banking* , 1 (1997).
- [69] J. D. Hamilton, Why you should never use the hodrick–prescott filter, *The Review of Economics and Statistics* **100**, 831 (2018).
- [70] R. J. Hodrick, *An Exploration of Trend-Cycle Decompo*

- sition Methodologies in Simulated Data*, Working Paper 26750 (National Bureau of Economic Research, 2020).
- [71] Wikipedia contributors, List of s&p 500 companies — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=List_of_S%26P_500_companies&oldid=978246445 (2020), [Online; accessed 6-January-2021].

Appendix A: List of selected stocks

TABLE IV: Overview of the 250 selected companies of the S&P 500 index (cf. [71]).

Number	Symbol	Security	Sector	Sub-Industry
1	CVX	Chevron Corp.	Energy	Integrated Oil & Gas
2	HES	Hess Corporation	Energy	Integrated Oil & Gas
3	XOM	Exxon Mobil Corp.	Energy	Integrated Oil & Gas
4	BKR	Baker Hughes Co	Energy	Oil & Gas Equipment & Services
5	HAL	Halliburton Co.	Energy	Oil & Gas Equipment & Services
6	SLB	Schlumberger Ltd.	Energy	Oil & Gas Equipment & Services
7	APA	Apache Corporation	Energy	Oil & Gas Exploration & Production
8	COG	Cabot Oil & Gas	Energy	Oil & Gas Exploration & Production
9	COP	ConocoPhillips	Energy	Oil & Gas Exploration & Production
10	EOG	EOG Resources	Energy	Oil & Gas Exploration & Production
11	MRO	Marathon Oil Corp.	Energy	Oil & Gas Exploration & Production
12	NBL	Noble Energy	Energy	Oil & Gas Exploration & Production
13	OXY	Occidental Petroleum	Energy	Oil & Gas Exploration & Production
14	VLO	Valero Energy	Energy	Oil & Gas Refining & Marketing
15	OKE	ONEOK	Energy	Oil & Gas Storage & Transportation
16	WMB	Williams Companies	Energy	Oil & Gas Storage & Transportation
17	VMC	Vulcan Materials	Materials	Construction Materials
18	FMC	FMC Corporation	Materials	Fertilizers & Agricultural Chemicals
19	MOS	The Mosaic Company	Materials	Fertilizers & Agricultural Chemicals
20	NEM	Newmont Corporation	Materials	Gold
21	APD	Air Products & Chemicals Inc	Materials	Industrial Gases
22	BLL	Ball Corp	Materials	Metal & Glass Containers
23	AVY	Avery Dennison Corp	Materials	Paper Packaging
24	IP	International Paper	Materials	Paper Packaging
25	SEE	Sealed Air	Materials	Paper Packaging
26	ECL	Ecolab Inc.	Materials	Specialty Chemicals
27	IFF	International Flavors & Fragrances	Materials	Specialty Chemicals
28	PPG	PPG Industries	Materials	Specialty Chemicals
29	SHW	Sherwin-Williams	Materials	Specialty Chemicals
30	NUE	Nucor Corp.	Materials	Steel
31	BA	Boeing Company	Industrials	Aerospace & Defense
32	GD	General Dynamics	Industrials	Aerospace & Defense
33	LMT	Lockheed Martin Corp.	Industrials	Aerospace & Defense
34	NOC	Northrop Grumman	Industrials	Aerospace & Defense
35	RTX	Raytheon Technologies	Industrials	Aerospace & Defense
36	TXT	Textron Inc.	Industrials	Aerospace & Defense
37	DE	Deere & Co.	Industrials	Agricultural & Farm Machinery
38	EXPD	Expeditors	Industrials	Air Freight & Logistics
39	FDX	FedEx Corporation	Industrials	Air Freight & Logistics
40	ALK	Alaska Air Group Inc	Industrials	Airlines
41	UVL	Southwest Airlines	Industrials	Airlines

Continuation: Overview of the 250 selected companies of the S&P 500 index (cf. [71]).

Number	Symbol	Security	Sector	Sub-Industry
42	AOS	A.O. Smith Corp	Industrials	Building Products
43	FAST	Fastenal Co	Industrials	Building Products
44	JCI	JCI Johnson Controls International	Industrials	Building Products
45	MAS	Masco Corp.	Industrials	Building Products
46	J	Jacobs Engineering Group	Industrials	Construction & Engineering
47	CAT	Caterpillar Inc.	Industrials	Construction Machinery & Heavy Trucks
48	PCAR	PACCAR Inc.	Industrials	Construction Machinery & Heavy Trucks
49	CTAS	Cintas Corporation	Industrials	Diversified Support Services
50	AME	AMETEK Inc.	Industrials	Electrical Components & Equipment
51	EMR	Emerson Electric Company	Industrials	Electrical Components & Equipment
52	ETN	Eaton Corporation	Industrials	Electrical Components & Equipment
53	ROK	Rockwell Automation Inc.	Industrials	Electrical Components & Equipment
54	ROL	Rollins Inc.	Industrials	Environmental & Facilities Services
55	GE	General Electric	Industrials	Industrial Conglomerates
56	HON	Honeywell Int'l Inc.	Industrials	Industrial Conglomerates
57	MMM	3M Company	Industrials	Industrial Conglomerates
58	CMI	Cummins Inc.	Industrials	Industrial Machinery
59	DOV	Dover Corporation	Industrials	Industrial Machinery
60	FLS	Flowserve Corporation	Industrials	Industrial Machinery
61	GWW	Grainger (W.W.) Inc.	Industrials	Industrial Machinery
62	IEX	INDEX Corporation	Industrials	Industrial Machinery
63	ITW	Illinois Tool Works	Industrials	Industrial Machinery
64	PH	Parker-Hannifin	Industrials	Industrial Machinery
65	PNR	Pentair plc	Industrials	Industrial Machinery
66	SNA	Snap-on	Industrials	Industrial Machinery
67	SWK	Stanley Black & Decker	Industrials	Industrial Machinery
68	CSX	CSX Corp.	Industrials	Railroads
69	KSU	Kansas City Southern	Industrials	Railroads
70	NSC	Norfolk Southern Corp.	Industrials	Railroads
71	UNP	Union Pacific Corp	Industrials	Railroads
72	EFX	Equifax Inc.	Industrials	Research & Consulting Services
73	JBHT	J. B. Hunt Transport Services	Industrials	Trucking
74	GPS	Gap Inc.	Consumer Discretionary	Apparel Retail
75	LB	L Brands Inc.	Consumer Discretionary	Apparel Retail
76	ROST	Ross Stores	Consumer Discretionary	Apparel Retail
77	TJX	TJX Companies Inc.	Consumer Discretionary	Apparel Retail
78	NKE	Nike, Inc.	Consumer Discretionary	Apparel, Accessories & Luxury Goods
79	PVH	PVH Corp.	Consumer Discretionary	Apparel, Accessories & Luxury Goods
80	TIF	Tiffany & Co.	Consumer Discretionary	Apparel, Accessories & Luxury Goods
81	VFC	VF Corporation	Consumer Discretionary	Apparel, Accessories & Luxury Goods
82	F	Ford Motor Company	Consumer Discretionary	Automobile Manufacturers
83	MGM	MGM Resorts International	Consumer Discretionary	Casinos & Gaming
84	BBY	Best Buy Co. Inc.	Consumer Discretionary	Computer & Electronics Retail
85	TGT	Target Corp.	Consumer Discretionary	General Merchandise Stores
86	LEG	Leggett & Platt	Consumer Discretionary	Home Furnishings
87	HD	Home Depot	Consumer Discretionary	Home Improvement Retail
88	LOW	Lowe's Cos.	Consumer Discretionary	Home Improvement Retail

Continuation: Overview of the 250 selected companies of the S&P 500 index (cf. [71]).

Number	Symbol	Security	Sector	Sub-Industry
89	LEN	Lennar Corp.	Consumer Discretionary	Homebuilding
90	PHM	PulteGroup	Consumer Discretionary	Homebuilding
91	CCL	Carnival Corp.	Consumer Discretionary	Hotels, Resorts & Cruise Lines
92	WHR	Whirlpool Corp.	Consumer Discretionary	Household Appliances
93	NWL	Newell Brands	Consumer Discretionary	Housewares & Specialties
94	HAS	Hasbro Inc.	Consumer Discretionary	Leisure Products
95	MCD	McDonald's Corp.	Consumer Discretionary	Restaurants
96	HRB	H&R Block	Consumer Discretionary	Specialized Consumer Services
97	GPC	Genuine Parts	Consumer Discretionary	Specialty Stores
98	ADM	Archer-Daniels-Midland Co	Consumer Staples	Agricultural Products
99	TAP	Molson Coors Beverage Company	Consumer Staples	Brewers
100	BF.B	Brown-Forman Corp.	Consumer Staples	Distillers & Vintners
101	WBA	Walgreens Boots Alliance	Consumer Staples	Drug Retail
102	SYY	Sysco Corp.	Consumer Staples	Food Distributors
103	KR	Kroger Co.	Consumer Staples	Food Retail
104	CHD	Church & Dwight	Consumer Staples	Household Products
105	CL	Colgate-Palmolive	Consumer Staples	Household Products
106	CLX	The Clorox Company	Consumer Staples	Household Products
107	KMB	Kimberly-Clark	Consumer Staples	Household Products
108	COST	Costco Wholesale Corp.	Consumer Staples	Hypermarkets & Super Centers
109	WMT	Walmart	Consumer Staples	Hypermarkets & Super Centers
110	CAG	Conagra Brands	Consumer Staples	Packaged Foods & Meats
111	CPB	Campbell Soup	Consumer Staples	Packaged Foods & Meats
112	GIS	General Mills	Consumer Staples	Packaged Foods & Meats
113	HRL	Hormel Foods Corp.	Consumer Staples	Packaged Foods & Meats
114	HSY	The Hershey Company	Consumer Staples	Packaged Foods & Meats
115	K	Kellogg Co.	Consumer Staples	Packaged Foods & Meats
116	MKC	McCormick & Co.	Consumer Staples	Packaged Foods & Meats
117	TSN	Tyson Foods	Consumer Staples	Packaged Foods & Meats
118	PG	Procter & Gamble	Consumer Staples	Personal Products
119	KO	Coca-Cola Company	Consumer Staples	Soft Drinks
120	PEP	PepsiCo Inc.	Consumer Staples	Soft Drinks
121	MO	Altria Group Inc	Consumer Staples	Tobacco
122	AMGN	Amgen Inc.	Health Care	Biotechnology
123	BMY	Bristol-Myers Squibb	Health Care	Health Care Distributors
124	CAH	Cardinal Health Inc.	Health Care	Health Care Distributors
125	ABMD	ABIOMED Inc	Health Care	Health Care Equipment
126	ABT	Abbott Laboratories	Health Care	Health Care Equipment
127	BAX	Baxter International Inc.	Health Care	Health Care Equipment
128	BDX	Becton Dickinson	Health Care	Health Care Equipment
129	DHR	Danaher Corp.	Health Care	Health Care Equipment
130	HOLX	Hologic	Health Care	Health Care Equipment
131	MDT	Medtronic plc	Health Care	Health Care Equipment
132	PKI	PerkinElmer	Health Care	Health Care Equipment
133	SYK	Stryker Corp.	Health Care	Health Care Equipment
134	TFX	Teleflex	Health Care	Health Care Equipment
135	VAR	Varian Medical Systems	Health Care	Health Care Equipment
136	UHS	Universal Health Services	Health Care	Health Care Facilities
137	CVS	CVS Health	Health Care	Health Care Services
138	WST	West Pharmaceutical Services	Health Care	Health Care Supplies
139	CERN	Cerner	Health Care	Health Care Technology
140	BIO	Bio-Rad Laboratories	Health Care	Life Sciences Tools & Services
141	TMO	Thermo Fisher Scientific	Health Care	Life Sciences Tools & Services
142	CI	CIGNA Corp.	Health Care	Managed Health Care

Continuation: Overview of the 250 selected companies of the S&P 500 index (cf. [71]).

Number	Symbol	Security	Sector	Sub-Industry
143	HUM	Humana Inc.	Health Care	Managed Health Care
144	UNH	United Health Group Inc.	Health Care	Managed Health Care
145	JNJ	Johnson & Johnson	Health Care	Pharmaceuticals
146	LLY	Lilly (Eli) & Co.	Health Care	Pharmaceuticals
147	MRK	Merck & Co.	Health Care	Pharmaceuticals
148	MYL	Mylan N.V.	Health Care	Pharmaceuticals
149	PFE	Pfizer Inc.	Health Care	Pharmaceuticals
150	BEN	Franklin Resources	Financials	Asset Management & Custody Banks
151	BK	The Bank of New York Mellon	Financials	Asset Management & Custody Banks
152	NTRS	Northern Trust Corp.	Financials	Asset Management & Custody Banks
153	STT	State Street Corp.	Financials	Asset Management & Custody Banks
154	TROW	T. Rowe Price Group	Financials	Asset Management & Custody Banks
155	AXP	American Express Co	Financials	Consumer Finance
156	BAC	Bank of America Corp	Financials	Diversified Banks
157	C	Citigroup Inc.	Financials	Diversified Banks
158	CMA	Comerica Inc.	Financials	Diversified Banks
159	JPM	JPMorgan Chase & Co.	Financials	Diversified Banks
160	USB	U.S. Bancorp	Financials	Diversified Banks
161	WFC	Wells Fargo	Financials	Diversified Banks
162	AJG	Arthur J. Gallagher & Co.	Financials	Insurance Brokers
163	AON	Aon plc	Financials	Insurance Brokers
164	MMC	Marsh & McLennan	Financials	Insurance Brokers
165	RJF	Raymond James Financial Inc.	Financials	Investment Banking & Brokerage
166	SCHW	Charles Schwab Corporation	Financials	Investment Banking & Brokerage
167	AFL	AFLAC Inc	Financials	Life & Health Insurance
168	GL	Globe Life Inc.	Financials	Life & Health Insurance
169	UNM	Unum Group	Financials	Life & Health Insurance
170	L	Loews Corp.	Financials	Multi-line Insurance
171	LNC	Lincoln National	Financials	Multi-line Insurance
172	AIG	American International Group	Financials	Property & Casualty Insurance
173	CINF	Cincinnati Financial	Financials	Property & Casualty Insurance
174	PGR	Progressive Corp.	Financials	Property & Casualty Insurance
175	TRV	The Travelers Companies Inc.	Financials	Property & Casualty Insurance
176	WRB	W. R. Berkley Corporation	Financials	Property & Casualty Insurance
177	FITB	Fifth Third Bancorp	Financials	Regional Banks
178	HBAN	Huntington Bancshares	Financials	Regional Banks
179	KEY	KeyCorp	Financials	Regional Banks
180	PNC	PNC Financial Services	Financials	Regional Banks
181	RF	Regions Financial Corp.	Financials	Regional Banks
182	SIVB	SVB Financial	Financials	Regional Banks
183	TFC	Truist Financial	Financials	Regional Banks
184	ZION	Zions Bancorp	Financials	Regional Banks
185	PBCT	People's United Financial	Financials	Thrifts & Mortgage Finance
186	PEAK	Healthpeak Properties	Real Estate	Health Care REITs
187	HST	Host Hotels & Resorts	Real Estate	Hotel & Resort REITs
188	DRE	Duke Realty Corp	Real Estate	Industrial REITs
189	VNO	Vornado Realty Trust	Real Estate	Office REITs
190	UDR	UDR, Inc.	Real Estate	Residential REITs

Continuation: Overview of the 250 selected companies of the S&P 500 index (cf. [71]).

Number	Symbol	Security	Sector	Sub-Industry
191	FRT	Federal Realty Investment Trust	Real Estate	Retail REITs
192	PSA	Public Storage	Real Estate	Specialized REITs
193	WY	Weyerhaeuser	Real Estate	Specialized REITs
194	ADBE	Adobe Inc.	Information Technology	Application Software
195	ADSK	Autodesk Inc.	Information Technology	Application Software
196	CDNS	Cadence Design Systems	Information Technology	Application Software
197	NLOK	NortonLifeLock	Information Technology	Application Software
198	ORCL	Oracle Corp.	Information Technology	Application Software
199	CSCO	Cisco Systems	Information Technology	Communications Equipment
200	MSI	Motorola Solutions Inc.	Information Technology	Communications Equipment
201	FISV	Fiserv Inc	Information Technology	Data Processing & Outsourced Services
202	PAYX	Paychex Inc.	Information Technology	Data Processing & Outsourced Services
203	GLW	Corning Inc.	Information Technology	Electronic Components
204	ADP	Automatic Data Processing	Information Technology	Internet Services & Infrastructure
205	IBM	International Business Machines	Information Technology	IT Consulting & Other Services
206	AMAT	Applied Materials Inc.	Information Technology	Semiconductor Equipment
207	KLAC	KLA Corporation	Information Technology	Semiconductor Equipment
208	LRCX	Lam Research	Information Technology	Semiconductor Equipment
209	ADI	Analog Devices, Inc.	Information Technology	Semiconductors
210	AMD	Advanced Micro Devices Inc	Information Technology	Semiconductors
211	INTC	Intel Corp.	Information Technology	Semiconductors
212	MU	Micron Technology	Information Technology	Semiconductors
213	MXIM	Maxim Integrated Products Inc	Information Technology	Semiconductors
214	SWKS	Skyworks Solutions	Information Technology	Semiconductors
215	TXN	Texas Instruments	Information Technology	Semiconductors
216	MSFT	Microsoft Corp.	Information Technology	Systems Software
217	AAPL	Apple Inc.	Information Technology	Technology Hardware, Storage & Peripherals
218	HPQ	HP Inc.	Information Technology	Technology Hardware, Storage & Peripherals
219	WDC	Western Digital	Information Technology	Technology Hardware, Storage & Peripherals
220	XRX	Xerox	Information Technology	Technology Hardware, Storage & Peripherals
221	IPG	Interpublic Group	Communication Services	Advertising
222	OMC	Omnicom Group	Communication Services	Advertising
223	CTL	CenturyLink Inc	Communication Services	Alternative Carriers
224	CMCSA	Comcast Corp.	Communication Services	Cable & Satellite
225	T	AT&T Inc.	Communication Services	Integrated Telecommunication Services
226	VZ	Verizon Communications	Communication Services	Integrated Telecommunication Services
227	EA	Electronic Arts	Communication Services	Interactive Home Entertainment
228	DIS	The Walt Disney Company	Communication Services	Movies & Entertainment
229	FOX	Fox Corporation (Class B)	Communication Services	Movies & Entertainment
230	AEP	American Electric Power	Utilities	Electric Utilities
231	D	Dominion Energy	Utilities	Electric Utilities
232	DUK	Duke Energy	Utilities	Electric Utilities
233	ED	Consolidated Edison	Utilities	Electric Utilities
234	EIX	Edison Int'l	Utilities	Electric Utilities
235	ETR	Entergy Corp.	Utilities	Electric Utilities
236	EVRG	Energy	Utilities	Electric Utilities
237	LNT	Alliant Energy Corp	Utilities	Electric Utilities

Continuation: Overview of the 250 selected companies of the S&P 500 index (cf. [71]).

Number	Symbol	Security	Sector	Sub-Industry
238	PEG	Public Service Enterprise Group (PSEG)	Utilities	Electric Utilities
239	PPL	PPL Corp.	Utilities	Electric Utilities
240	SO	Southern Company	Utilities	Electric Utilities
241	WEC	WEC Energy Group	Utilities	Electric Utilities
242	ATO	Atmos Energy	Utilities	Gas Utilities
243	CMS	CMS Energy	Utilities	Multi-Utilities
244	CNP	CenterPoint Energy	Utilities	Multi-Utilities
245	DTE	DTE Energy Co.	Utilities	Multi-Utilities
246	EXC	Exelon Corp.	Utilities	Multi-Utilities
247	NEE	NextEra Energy	Utilities	Multi-Utilities
248	NI	NiSource Inc.	Utilities	Multi-Utilities
249	PNW	Pinnacle West Capital	Utilities	Multi-Utilities
250	XEL	Xcel Energy Inc	Utilities	Multi-Utilities