

Spelta, Alessandro

**Working Paper**

## Stock prices prediction via tensor decomposition and links forecast

Working Paper, No. 41

**Provided in Cooperation with:**

Università Cattolica del Sacro Cuore, Dipartimento di Economia e Finanza (DISCE)

*Suggested Citation:* Spelta, Alessandro (2016) : Stock prices prediction via tensor decomposition and links forecast, Working Paper, No. 41, Università Cattolica del Sacro Cuore, Dipartimento di Economia e Finanza (DISCE), Milano

This Version is available at:

<https://hdl.handle.net/10419/170638>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

UNIVERSITÀ CATTOLICA DEL SACRO CUORE  
**Dipartimento di Economia e Finanza**

**Working Paper Series**

**Stock prices prediction via tensor  
decomposition and links forecast**

*Alessandro Spelta*

**Working Paper n. 41**

**May 2016**



UNIVERSITÀ  
CATTOLICA  
del Sacro Cuore

# Stock prices prediction via tensor decomposition and links forecast

**Alessandro Spelta**

*Università Cattolica del Sacro Cuore*

Working Paper n. 41  
May 2016

Dipartimento di Economia e Finanza  
Università Cattolica del Sacro Cuore  
Largo Gemelli 1 - 20123 Milano – Italy  
tel: +39.02.7234.2976 - fax: +39.02.7234.2781  
e-mail: [dip.economiaefinanza@unicatt.it](mailto:dip.economiaefinanza@unicatt.it)

The Working Paper Series promotes the circulation of research results produced by the members and affiliates of the Dipartimento di Economia e Finanza, with the aim of encouraging their dissemination and discussion. Results may be in a preliminary or advanced stage. The Dipartimento di Economia e Finanza is part of the Dipartimenti e Istituti di Scienze Economiche (DISCE) of the Università Cattolica del Sacro Cuore.

# Stock prices prediction via tensor decomposition and links forecast

A. Spelta<sup>+</sup>

Catholic University and Complexity Lab in Economics, Milan, Italy

May 24, 2016

**Abstract:** Many complex systems display fluctuations between alternative states in correspondence to tipping points. These critical shifts are usually associated with generic empirical phenomena such as strengthening correlations between entities composing the system. In finance, for instance, market crashes are the consequence of herding behaviors that make the units of the system strongly correlated, lowering their distances. Consequently, determining future distances between stocks can be a valuable starting point for predicting market down-turns. This is the scope of the work. It introduces a multi-way procedure for forecasting stock prices by decomposing a distance tensor. This multidimensional method avoids aggregation processes that could lead to the loss of crucial features of the system. The technique is applied to a basket of stocks composing the S&P500 composite index and to the index itself so as to demonstrate its ability to predict the large market shifts that arise in times of turbulence, such as the ongoing financial crisis.

JEL Codes: C02, C63, C63.

Keywords: Stock prices, Correlations, Tensor Decomposition, Forecast.

<sup>+</sup>Department of Economics and Finance, via Necchi 5 - 20123 Milano, Italy. E-mail:

[alessandro.spelta@unicatt.it](mailto:alessandro.spelta@unicatt.it)

**Introduction:** From biochemical reactions (1-4) to financial markets (5-7), passing through medicine (8,9) and ecological systems (10,11), the science of complexity is percolating through our everyday life, emerging as a unifying feature of the world.

The complex system underlying different processes in nature, seems to behave in similar ways (even if the drivers of these behaviors are process-specific), showing abrupt fluctuations from one state to another in a highly irregular way (7).

These shifts usually take place at critical thresholds - the so-called tipping points - and are associated with critical transitions between alternative states (6-11).

In medicine, asthma or epileptic attacks are triggered by spontaneous systemic failures (8-12), abrupt shifts also occur in ocean circulation and climate changes (13) or ecological systems (10,11). Financial markets, being a paradigm of complex systems (6,7), also experience sudden regime shifts where fluctuations are characterized by bumps that create upward and downward trends.

Predicting critical transitions and abrupt changes in complex systems is a difficult task but fortunately some theoretical works (14,15) suggest the existence of generic indicators for critical transitions even when the knowledge of the functioning of the systems is insufficient for building predictive models.

The underlying principle of most of these indicators is a phenomenon known in dynamical systems theory as critical slowing down. This phenomenon occurs in most bifurcation points when the dominant eigenvalue characterizing the rates of change of the system around the equilibrium becomes zero (14).

At these points the system becomes increasingly slow in recovering from small perturbations and the resulting time-series turn out to be highly auto-correlated.

Beside the growing auto-correlations of the state variables of the system, recent works (11,16) have suggested that the critical slowing down phenomenon might, in theory, generate also spatial signals such as an increasing spatial correlation near transitions.

This is due to the fact that the entities composing the system pass from isolated to coordinated behaviors, where a spontaneous order emerges (6,7,16). When the intrinsic dynamics of each entity is weakened, the state of a unit will be strongly dependent on that of its neighbors. As a result, units will become more strongly correlated close to the transition.

In finance, for instance, the formation and collapse of speculative bubbles have been largely considered the consequence of herding behaviors emerging from the broken balance between autonomous behavior and peer influence (7). When the effect of exchanging influence with the rest of the environment dominates, large-scale phenomena occur.

These phenomena have a strong impact on the everyday life of millions of people and have attracted the interest of many researchers. Beside the pioneering work of Mandelbrot (17) that investigates the underlying stochastic process generating rare events, correlation-based measures play a central role also in the study of financial market fluctuations (18-20).

Indeed, in finance, like in many other complex systems, researchers have observed modifications in the correlation structure between stock prices in the face of market turbulence, such as an ongoing crisis (21). While, during expansion and normal periods, financial markets tend toward randomness, in crisis phases their structures are reinforced due to a generalized increase in the level of correlations (21,22).

Although there exists empirical evidence of connections between the strengthening of correlations and crisis episodes in financial markets, most of the existing studies mainly focus on correlations between stock prices (18-20), the resulting distance-based matrices and on their Minimum Spanning Tree representations (21-23), to provide optimal asset allocations and portfolio risk estimates.

This paper, in turn, explicitly addresses the question of inferring the forthcoming dynamic of stock prices through the prediction of future distances between stocks.

This issue amounts to a link prediction problem (24). Given past connections (distances) between stocks, what will be their next period value? If predictions suggest a contraction of the next period distances for instance, then, we could expect a decrease in stock prices because of a strengthening of correlations and a higher likelihood of a crisis episode.

The mainstream class of link prediction methods, are based on the so-called similarity-based algorithms, which are further classified into three categories: local, global and quasi-local depending on the information used (24). Usually all these techniques (24) collapse the temporal data into a single matrix by summing (with or without weights) the records corresponding to the each time period. Then similarity-based measures like the Katz centrality (25) or the singular value decompositions (SVD) are applied to perform links prediction.

This paper instead is the first attempt to use tensor decompositions and multi-way analysis (26,27) to extract complex relationships from stock prices' time series and use them in a link prediction application. This approach prevents the temporal aggregation of the data, avoiding losses of crucial features of the system that can be observed only by holding the original time-varying nature of the records.

Starting from  $N$  time series of stock prices, a rolling window of length  $n_1$  is applied to compute the correlation  $C_{k,l}$  among each pair  $(k, l)$  of stocks. Given these pair-wise correlations, at each time step, a distance matrix with elements

$$d_{kl} = \sqrt{2(1 - C_{k,l})}$$

is created (28).

Once the rolling window has produced  $Z$  distance matrices  $\mathbf{D} \in \mathbb{R}^{N \times N}$ , these matrices are embedded into a 3D-tensor  $\mathcal{D} \in \mathbb{R}^{N \times N \times Z}$  whose generic element  $\delta_{klz}$  represents the distance between stock  $k$  and stock  $l$  at time  $z$ .

The tensor is thus approximated as the outer product of three vectors through the canonical decomposition (29), also known as parallel factorization (30), the so-called CP decomposition. This technique can be regarded as a generalization of SVD to tensors (see Methods).

The decomposition aims at writing the tensor  $\mathcal{D}$  as the outer product of two identical vectors  $\mathbf{v}$ , that contains the *overall spatial dissimilarity* between stocks and a vector  $\mathbf{u}$ , containing the *temporal profile* of the dissimilarities

$$\mathcal{D} \cong \lambda \mathbf{v} \circ \mathbf{v} \circ \mathbf{u}$$

where  $\mathbf{v} \in \mathbb{R}^N$ ,  $\mathbf{u} \in \mathbb{R}^Z$  and  $\lambda = \|\mathbf{v}\| \|\mathbf{v}\| \|\mathbf{u}\|$ .

While a stock with a high (low) overall spatial dissimilarity score has, on average, a different (similar) behavior compared with the one of the rest of the stocks, a period in which a high (low) temporal profile score is registered will be a period in which most of the stocks are highly dissimilar (similar).

The next step consists in generating the matrix of the forecasted distances. Instead of predicting the  $N^2$  possible distances using  $N^2$  data points, with this method one has to predict only the next value of the temporal profile  $\mathbf{u}$  and use it, together with the two fixed overall spatial dissimilarity vectors  $\mathbf{v}$ , to build the matrix of the forecasted distances.

An exponential smoothing, applied to the last  $n_2$  observations of the temporal profile vector  $\mathbf{u}$ , extracts a scalar  $\tau$  representing the presumed value of this vector in the next period. Then the matrix containing the forecast distances of all stock pairs is obtained as a linear combination of the two spatial dissimilarity vectors  $\mathbf{v}$ , the parameter  $\lambda$  and of the forecast  $\tau$  of the temporal profile vector.

In matrix terms;  $\widehat{\mathbf{D}} = \tau \lambda \mathbf{v} \mathbf{v}^T$  or, element-wise  $\widehat{d}_{kl} = \tau \lambda \mathbf{v}_k \mathbf{v}_l$  (where the superscript  $\widehat{\phantom{x}}$  denotes the predicted distance).

Finally, the vector of the forecasted prices is given by the outer product of the current price vector and of the normalized matrix representing the predicted future distances  $\widetilde{\mathbf{D}}$ . The normalization is obtained by



dividing each entry of the predicted distance matrix by the number of stocks in the dataset. In this way the forecast stock price will be equal to the current price multiplied by the average of the predicted distances that relate it to the rest of the stocks. In matrix terms;  $\hat{\mathbf{P}} = \mathbf{P}\tilde{\mathbf{D}}$  or, element-wise,  $\hat{P}_i = P_i \frac{1}{N} \sum \widehat{d}_{i,:} = P_i \sum \widetilde{d}_{i,:}$ .

In accordance with empirical evidence suggesting that distances contract during crisis periods, the predicted price for each stock will be lower than the current one if, on average, the distance between that stock and the rest is decreasing.

When the steps of the moving window exceed the parameter  $Z$ , the tensor is allowed to move in time with each new step, as new data become available. The temporal shift of the tensor permits to compare the forecasts produced by two consecutive decompositions (31). The difference between the values of the two predictions generates a signal whose sign indicates the future direction of the price.

To investigate whether this method is able to correctly identify changes in stock prices a back-test based on a hypothetical investment strategy is implemented (32).

If the sign of the signal for a given stock  $i$  is negative, a short position is taken by selling the stock and buying it back the next trading day. In this case, the cumulative return made on that stock  $R_i$  changes by  $\frac{P_i^t - P_i^{t+1}}{P_i^{t+1}}$ . Otherwise, if the difference is positive, a long position is taken by buying the stock and then selling it the next trading day. The cumulative return in this case changes by  $\frac{P_i^{t+1} - P_i^t}{P_i^t}$ . Notice that profits are only possible if at least some future changes in stock prices are correctly anticipated, in particular regarding large market movements. Figure 1 gives a graphical representation of the technique.

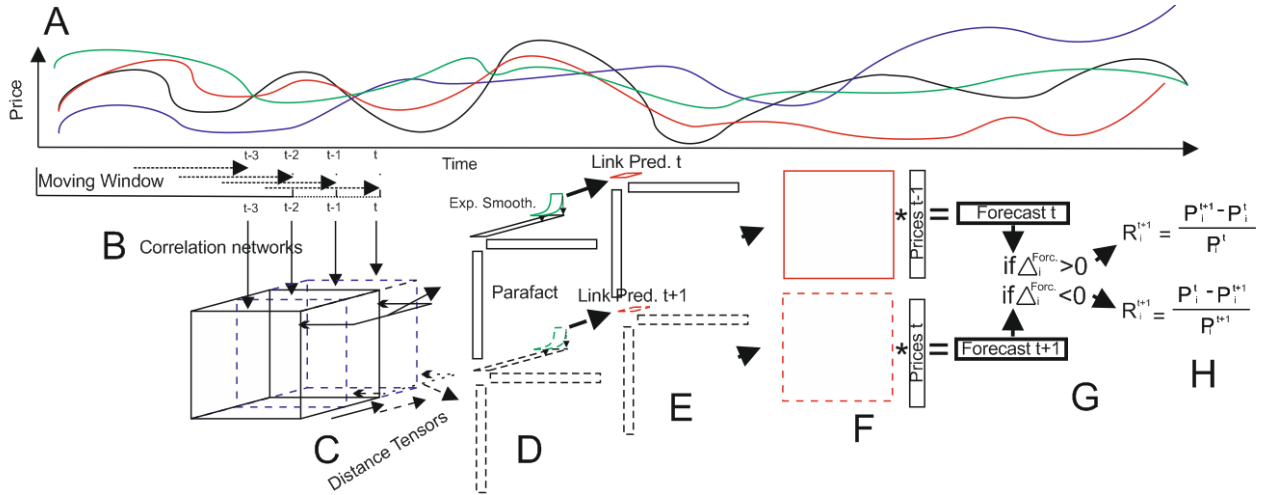


Figure 1: Graphical representation of the method. Starting from stock price time series (A), a rolling window is applied to compute the correlation among each pair of stocks (B). At each time step a distance matrix is created. Once the rolling window has produced  $Z$  distance matrices, those matrices are embedded into a 3D-tensor (C). When the steps of the moving window exceed  $Z$ , the tensor is allowed to move in time with each new step, as new data are available (C - solid line vs. dashed line). This temporal shift of the tensor permits to compare the forecasts produced by two consecutive decompositions and, depending on their difference, to take a long or short investment position. For graphical purposes, the decomposition of two consecutive tensors, the resulting links prediction and price forecasts are drawn using solid and dashed lines. The two consecutive tensors are approximated as the linear combination of three vectors (D) representing spatial ( $\mathbf{v}$ ) and temporal ( $\mathbf{u}$ ) relationships between stocks distances;  $\mathcal{D} \cong \lambda \mathbf{v} \circ \mathbf{v} \circ \mathbf{u}$ . The exponential smoothing (D - green lines) applied to  $\mathbf{u}$  extracts a scalar  $\tau$  representing the forecast of temporal profile for the next period (E - red lines). The forecast of the future distance matrix is obtained as a linear combination of the two spatial dissimilarity vectors, the parameter  $\lambda$  and of the forecast  $\tau$  of the temporal profile vector;  $\hat{\mathcal{D}} = \tau \lambda \mathbf{v} \mathbf{v}^T$  (E and F - red squares). Finally, the prediction of future prices is computed as the outer product of the past price vector and of the normalized matrix representing the predicted distances  $\hat{P}_i = P_i \frac{1}{N} \sum \hat{a}_{i,:}$  (F and G). An investment strategy is proposed to assess the efficiency of the method. If the difference  $\Delta_i^{Forc} = \hat{P}_i^t - \hat{P}_i^{t-1}$  between two consecutive price forecast for a generic stock  $i$  is negative (G) then the stock is sold at price  $P_i^t$  and bought back the next trading day at price  $P_i^{t+1}$ . Otherwise, the stock is bought at price  $P_i^t$  and sold back in  $t+1$  at price  $P_i^{t+1}$ . The returns are calculated as  $\frac{P_i^t - P_i^{t+1}}{P_i^{t+1}}$  or as  $\frac{P_i^{t+1} - P_i^t}{P_i^t}$  depending whether a short or a long position is taken (H).

## Results

The method is applied to the closure price of a basket of 388 stocks composing the S&P500 composite index, traded during 3527 working days, from 1999/08/04 to 2013/08/09. Additionally, a modified version of the method is also employed to forecast the dynamic of the S&P500 composite index for the period ranging from 2004/06/24 to 2013/04/30.

In order to give a rough overview of the market dynamic, Fig. 2(a) shows the average price of the basket of stocks under analysis (a) while Fig. 2(b) displays the S&P500 composite index dynamic. Besides the general increasing trend, Fig. 2(a) shows two broad recession periods, the collapsed of the dot-com bubble occurred in 2002-03, and the global financial recession created by the bursting of the sub-prime mortgage bubble of 2008-09 . Figure 2(b), instead, reports the behavior of the S&P500 composite index for a shorter time length, not displaying the effects of the dot-com bubble.

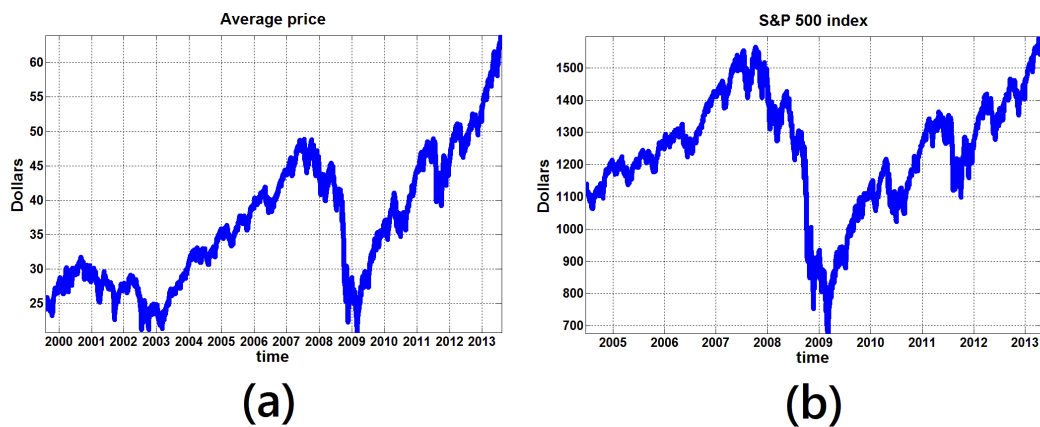


Figure 2 : Average stock price dynamic during the period 1999-2013 (a) together with the S&P500 index dynamic during the period 2004-2013 (b). Starting from the boom phase of the dot-com bubble in 2000, Figure 2(a) also shows the market decline of 2002, while both panels display the recovering period ended in 2007 with the sub-prime mortgage crisis together with the global financial crisis of the 2008-09. The final part on the sample shows the recovering taking place until 2013.

Figure 3 shows the cumulative sum of the returns obtained for each stock together with the average cumulative performance, namely, the mean of the stocks' returns (solid black line). Besides the fact that the investment strategy does not produce positive returns for all the stocks, the values of y-axis, biased in favor of positive quantities, together with the positive average return (black line), that reaches the value of 230% at the end of the sample, confirm the ability of the methodology to produce good predictions.

Figure 3(a) displays the cumulative returns obtained by only taking short while the cumulative returns obtained by only taking long positions are showed in Fig. 3(c). In this way one is able to compare the performance of the methodology in predicting down-turns or up-turns of stock prices. From the average performance (black line) reported in Fig. 3(b) it clearly emerges that deep crashes, the burst of the dot-com

bubble and the 2008 financial crisis, are correctly anticipated. This more than compensates the losses of taking long positions (Fig. 3(c)) during these phases. These simulations are performed using the following parameters:  $n_1 = 15$ ,  $Z = 25$ ,  $n_2 = 7$  and the exponential smoothing parameter is set to be equal to 0.2.

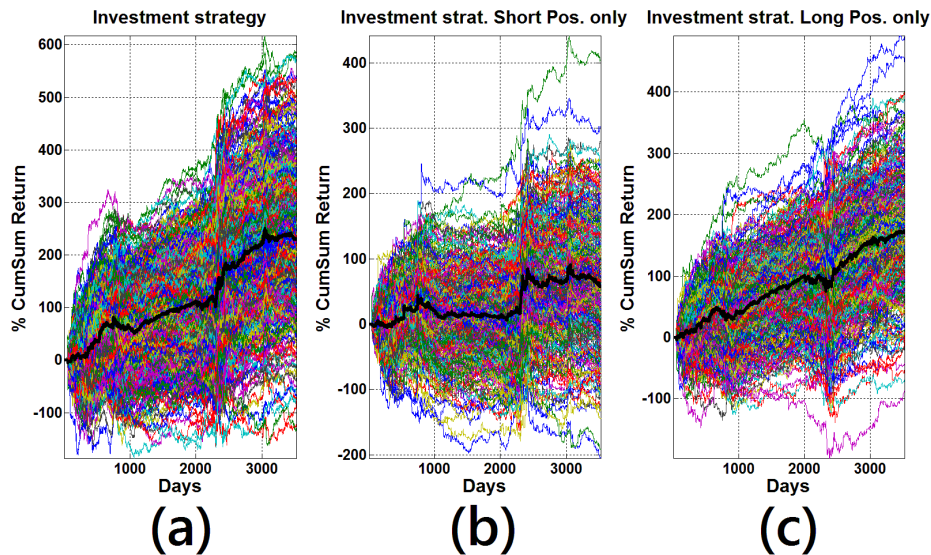


Figure 3: Cumulative sum of returns obtained for each stock together with the average cumulative performance (solid black line). Figure 3(a) shows the performance of the investment strategy. Panels (b)-(c) display the cumulative returns obtained by investing only taking short or long positions respectively. This helps in quantifying how the correct forecast of booms and burst phases affects the overall cumulative returns. The simulations are obtained using the following parameters:  $n_1 = 15$ ,  $Z = 25$ ,  $n_2 = 7$  and the exponential smoothing parameter is set to be equal to 0.2.

While Fig. 3 aims at discovering whether the movements of each stock are correctly anticipated by only looking at the sign of the signal produced by the methodology (positive vs. negative), the next step consists in assessing the quality of the signals. In theory, the larger the absolute difference between two consecutive forecasts, i.e. the larger the absolute value of a signal, the more credible the forecast is.

In order to show this feature, at each time step, the signals are sorted in descending order, based on their absolute values. Figure 4(a) displays the average cumulative sum of the returns associated with different quantiles of the distribution of the signals. In particular the upper blue line is associated with the strongest signal, the green line shows the average cumulative returns produced by the two strongest signals, the red line indicates the average performance of the first forty-five signals. The other lines illustrate the

performance associated with the cumulative sum of signals of gradually lower quantiles. Finally the lowest purple line displays the average cumulative return associated with the whole distribution (and it is equivalent to the black line of Fig. 3).

Also in this case, the methodology is able to correctly predict the largest market movements, especially near deep burst phases as shown in Fig. 4(b). Moreover Fig. 4 points out that stronger signals produce better forecasts, indeed the cumulative returns associate the most robust signal (436%) doubles the average performance associate to all the signals (230%).

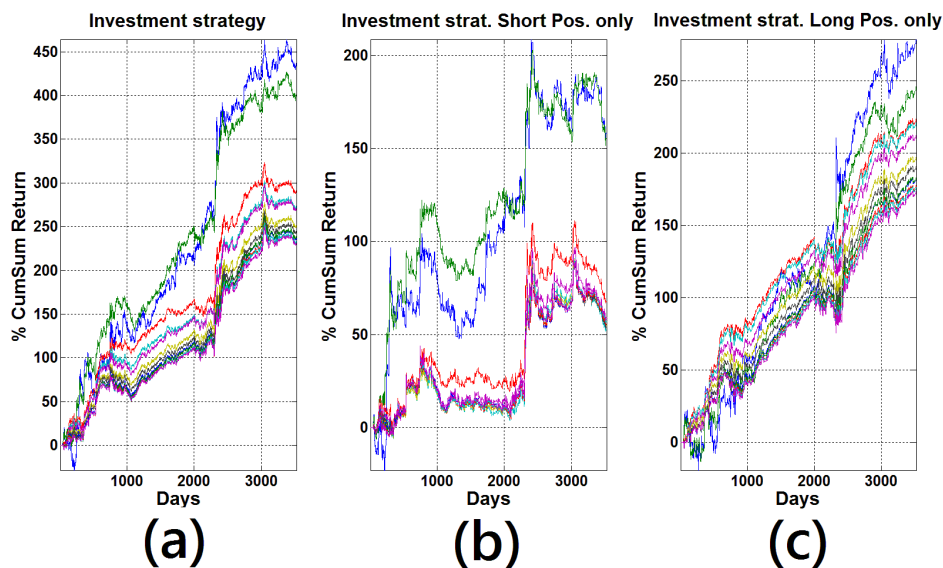


Figure 4: Cumulative sum of the returns associated with different quantiles of the signals' distribution. The goodness of each signal depends on the absolute difference between the two consecutive forecasts that compose the signal. The larger the difference, the better the signal. Figure 4(a) shows the performance of the whole investment strategy. Panels (b)-(c) display the cumulative returns obtained by only taking short or long positions respectively. The signals are sorted according to their absolute values and therefore there is no one-to-one correspondence between each plotted line and a particular stock. For instance, the best signal can regard different stocks in different moments in time. The simulations are obtained using the same parameters of Figure 3.

To further analyze the accuracy of the proposed methodology, the technique is also applied to forecasting the behavior of the S&P500 index as a whole (see Fig. 1(b)). Consequently, the procedure has been slightly modified to produce predictions for the whole composite index and not for each stock constituting the

basket. First, the number of stocks in the dataset is augmented (455 stocks are employed in this exercise), by restricting the temporal observations to the period 2004-2013. Secondly, only the forecasted distances  $\widehat{\mathbf{D}}$  are used and not the forecasted stock prices.

Similarly to Ref. (33) a Multidimensional Scaling Technique, the Principal Coordinates Analysis (34) is applied to  $\widehat{\mathbf{D}}$  with the aim of embedding the data in a space of lower dimensions while retaining the pairwise distances between the points as much as possible. The dimensionality reduction facilitates the classification of high-dimensional data, by mitigating the curse of dimensionality and other undesired properties of high-dimensional spaces.

After having found the centering matrix  $\mathbf{H} = \mathbf{I} - N^{-1}\mathbf{1}\mathbf{1}^T$ , where  $\mathbf{I}$  is the  $N \times N$  identity matrix, and  $\mathbf{1}$  is a vector of  $N$  ones, the eigenvalue and eigenvectors of the matrix  $\mathbf{B} = \mathbf{H} \left( -\frac{1}{2} \widehat{\mathbf{D}}^2 \right) \mathbf{H}$  are found. The coordinates in the lower-dimensional space are recorded in a matrix  $\mathbf{X} = \mathbf{A}_s \mathbf{L}_s^{1/2}$ , where  $\mathbf{A}_s$  contains the eigenvectors corresponding to the  $s$  largest eigenvalues of  $\mathbf{B}$ , and  $\mathbf{L}_s^{1/2}$  contains the square root of the  $s$  largest eigenvalues along the diagonal.

Following Ref. (22) these points are embedded in a space of 6 dimensions ( $s = 6$ ). The 6<sup>th</sup> root of the product of the eigenvalues of  $\mathbf{X}'\mathbf{X}$  defines the volume of the geometrical object composed by the embedded data (33).

The volume is used as a reference for the identification of abnormal periods. The volume expands whenever the cloud of points represents a situation of “business as usual” and the market space is similar to that of a random universe. On the other hand, in critical periods, the volume of the geometric object severely contracts, leading to the emergence of distorted shapes (33).

Now, the new signal is given by the difference of two subsequently predicted volumes  $\Delta^{forc} = \widehat{V}^t - \widehat{V}^{t-1}$ . Whenever this difference is negative the index is sold at price  $P^t$  and bought back the next trading day at

price  $P^{t+1}$ . Otherwise, the index is bought at price  $P^t$  and sold back in  $t+1$  at price  $P^{t+1}$ . The cumulative returns are calculated as  $\frac{P^t - P^{t+1}}{P^{t+1}}$  in the first case, and as  $\frac{P^{t+1} - P^t}{P^t}$  in the second.

Figure 5 shows the cumulative sum of the returns obtained by investing in the S&P500 composite index by following the differences in the predicted volumes. The simulations are obtained using the following parameters:  $n_1 = 7, Z = 30, n_2 = 20$ . As for the investment strategy based on stock price predictions, also in this case, Figure 5(a) shows that the predicted movements of stocks' distances anticipates the market dynamic. Large down-turns are correctly anticipated as indicated by the cumulative returns illustrated in Fig. 5(b) near day 1000 (that corresponds to the initial period of the 2008 financial crisis).

Figure 5(c), on the other hand, suggests that market up-turns, besides providing higher returns, are less severe than bust phases. The sum of the cumulative returns indeed has a smoother increasing behavior compared with the one obtained by correctly predicting market down-turns.

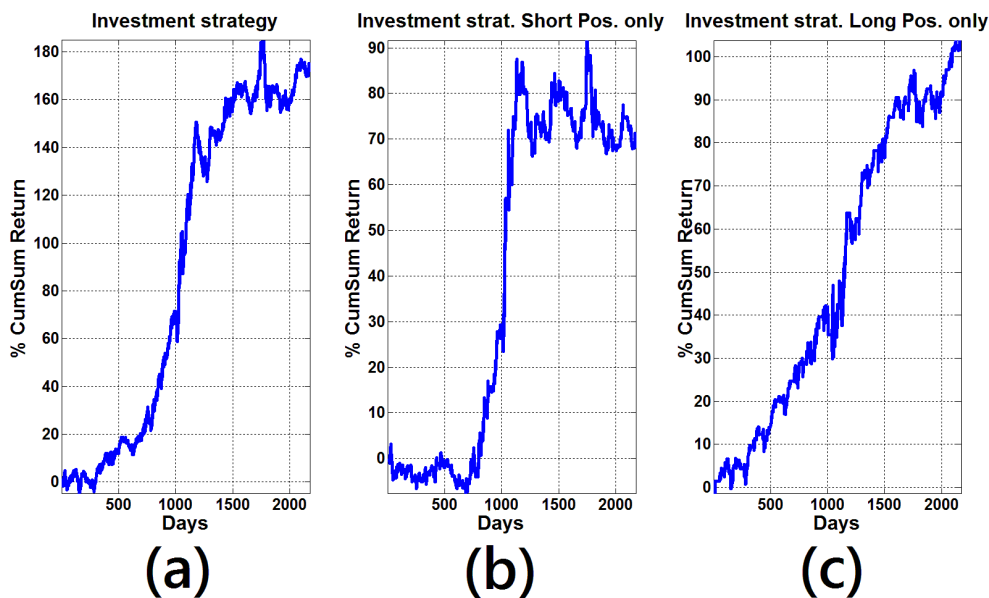


Figure 5: Cumulative sum of returns obtained for the S&P500 composite index. Figure 5(a) shows the performance of the investment strategy. Panels (b)-(c) display the cumulative returns obtained by investing only taking short or long positions respectively. This helps in quantifying how the correct forecasts of booms and burst phases affects the overall cumulative returns. The simulations are obtained using the following parameters:  $n_1 = 7, Z = 30, n_2 = 20$ .

Finally, Fig. 6 provides a robustness analysis. Since the parameter space is huge, the length of the tensor is kept fixed together with the parameter of the exponential smoothing while  $n_1$  and  $n_2$  take different values along the simulations. Each sub plot of Fig. 6 shows the cumulative returns obtained at the end of the time sample for different parameter values. In particular, the first row represents the end-of-sample cumulative returns obtained by averaging the cumulative performance of the method while forecasting the dynamic of the 388 stocks of the first dataset (see also Fig. 2). In particular Fig. 6(a) refers to the composite investment strategy, encompassing both long and short positions. Panels 6(b)-(c), instead, differentiate between short and long positions respectively. The central row of Fig. 6 shows the returns obtained by following only the best signal (as emphasized also in Fig. 4), for the whole investment strategy (d) and for short (e) and long (f) positions respectively. The last row, on the other hand, provides the same results but looking at the performance obtained by the application of the modified method to the S&P500 composite index.

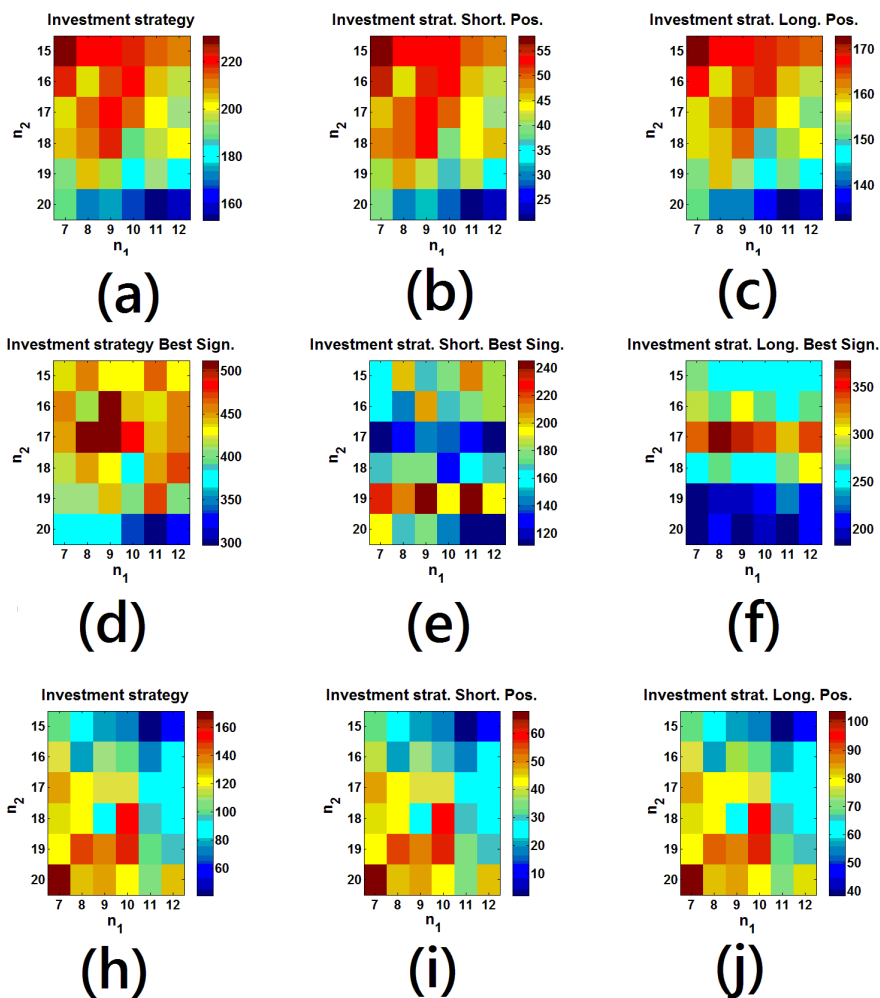




Figure 6: Cumulative returns obtained at the end of the time sample for different values of  $n_1$  and  $n_2$ . Panel 6(a) represents the end-of-sample cumulative performance obtained by averaging the returns obtained by forecasting the dynamic of the 388 stocks of the first dataset. Panel 6(b) displays the end-of-sample returns by investing using only short positions and figure 6(c) encompasses the results for the long positions investment strategy. Panel 6(d) shows the results for the investment strategy obtained by looking only at the best signal produced by the method together with the results for short (panel 6(e)) and long (panel 6(f)) investment strategy only. Finally, the last row provides the end-of-sample cumulative returns obtained by applying the modified method to the S&P500 composite index, for the whole investment strategy (panel 6(h)) and for short (panel 6(i)) and long (panel 6(j)) investment positions.

**Discussion:** The findings obtained by the application of this methodology have important consequences for the understanding of financial systems. Indeed, as pointed out by the recent financial crisis, financial systems are increasingly built on interdependencies and relationships that are difficult to predict and control. This work proposes a new dynamical approach to financial systems and stresses the systemic importance of empirical signs that can be used to extend the knowledge of financial markets and complex systems in general.

Predicting abrupt market down-turns, as a matter of fact, facilitates the drafting of policies that can reduce the severity of financial crises, by decreasing the risk of global collapses of financial services by making economic networks more robust.

The results suggest that tensor decompositions and multi-way analysis can effectively extract complex relationships from stock prices' time series opening new insights into large-scale collective decision-making.

## Methods

### Tensor decomposition

Tensor decompositions and multi-way analysis can be naturally employed to represent the time-varying distance matrices as a single mathematical object, a three-way tensor, and approximate this tensor as a product of vectors by extracting the most relevant spatial and temporal factors (29,30).

Uncovering the spatial ( $\mathbf{v}$ ) and the temporal profile ( $\mathbf{u}$ ) vectors that contain the overall dissimilarities between stocks and the related activity pattern requires the identification and the extraction of lower-dimensional features. This can be achieved by means of the so-called canonical CP decomposition in three dimensions.

The decomposition aims at writing the tensor  $\mathcal{D}$  as the outer product of two identical vectors  $\mathbf{v}$ , that contains the *overall spatial dissimilarities* between stocks' time series and a vector  $\mathbf{u}$ , containing the *temporal profile* of the dissimilarities:

$$\mathcal{D} \cong \lambda \mathbf{v} \circ \mathbf{v} \circ \mathbf{u}$$

where  $\mathbf{v} \in \mathbb{R}^N$ ,  $\mathbf{u} \in \mathbb{R}^Z$  and  $\lambda = \|\mathbf{v}\| \|\mathbf{v}\| \|\mathbf{u}\|$ .

Such an approximation of the tensor  $\mathcal{D}$  is equivalent to minimizing the Frobenius norm of the difference between  $\mathcal{D}$  and  $\lambda \mathbf{v} \circ \mathbf{v} \circ \mathbf{u}$ . Solving this problem amounts at finding the rank-1 tensors that best approximate the  $\mathcal{D}$

$$\min_{\mathbf{v}, \mathbf{u}} \|\mathcal{D} - \lambda \mathbf{v} \circ \mathbf{v} \circ \mathbf{u}\|_F^2 \quad (1)$$

The 3-dimensional problem is divided into 3 sub-problems by unfolding the tensor  $\mathcal{D}$ . This means reordering the elements of a tensor into a matrix.

The mode-3 unfolding of a tensor  $\mathcal{D}$  is denoted by  $\mathbf{D}_{(q)}$  and arranges the mode- $q$  fibers to be the columns of the resulting matrix. Tensor elements  $(k, l, z)$  maps to matrix element  $(i_Q, j)$ , where  $i_Q = \{k, l, z\}$  and  $j = 1 + \sum_{r=1, r \neq q}^Z (i_r - 1) J_r$  with  $J_r = \prod_{m=1, m \neq q}^{r-1} I_r$ .

The three resulting matrices have respectively a size of  $N \times NZ$ ,  $N \times NZ$  and  $Z \times N^2$ . In this way problem (1) is equivalent to minimizing the difference between each of the modes and their respective approximation in terms of factors. Problem (1) is thus converted into three problems

$$\begin{aligned}
& \min_{\mathbf{v} > \mathbf{0}} \|\mathbf{D}_{(1)} - \lambda \mathbf{v}(\mathbf{u} \odot \mathbf{v})^T\|_F^2 \\
& \min_{\mathbf{v} > \mathbf{0}} \|\mathbf{D}_{(2)} - \lambda \mathbf{v}(\mathbf{u} \odot \mathbf{v})^T\|_F^2 \quad (2) \\
& \min_{\mathbf{u} > \mathbf{0}} \|\mathbf{D}_{(3)} - \lambda \mathbf{u}(\mathbf{v} \odot \mathbf{v})^T\|_F^2
\end{aligned}$$

where  $\odot$  denotes the Khatri-Rao product, namely the column-wise Kronecker product. Since distances are always non negative, a non-negative tensor factorization method is employed to solve (2) because it greatly simplifies the interpretation of the resulting decomposition. The Block Coordinate Descent Method for Regularized Multiconvex Optimization (34) and the Matlab Tensor Toolbox (35) are used to solve (2).

After having initialized  $\mathbf{v}$  and  $\mathbf{u}$  as vectors of ones, the three components are updated according with the following rules:

$$\mathbf{v}^k = \max(0, \check{\mathbf{v}}^{k-1} - \mathbf{G}_v^{k-1}/\mathbf{L}_v^{k-1})$$

where  $\check{\mathbf{v}}^{k-1} = \mathbf{v}^{k-1} + w_v^{k-1}(\mathbf{v}^{k-1} - \mathbf{v}^{k-1})$ ,  $w_v^{k-1} = \min\left(\frac{t_{k-1}-1}{t_k}, \theta \sqrt{\frac{\mathbf{L}_v^{k-1}}{\mathbf{L}_v^{k-2}}}\right)$  with  $t_0 = 1$  and  $t_k = 1/2\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$ ;  $\mathbf{L}_v^{k-1} = |(\mathbf{B}_v^{k-1})^T(\mathbf{B}_v^{k-1})|$  with  $\mathbf{B}_v^{k-1} = \mathbf{u} \odot \mathbf{v}$  and  $\mathbf{G}_v^{k-1} = (\check{\mathbf{v}}^{k-1}(\mathbf{B}_v^{k-1})^T - \mathbf{D}_{(1)})\mathbf{B}_v^{k-1}$

$$\mathbf{v}^k = \max(0, \check{\mathbf{v}}^{k-1} - \mathbf{G}_v^{k-1}/\mathbf{L}_v^{k-1})$$

where  $\check{\mathbf{v}}^{k-1} = \mathbf{v}^{k-1} + w_v^{k-1}(\mathbf{v}^{k-1} - \mathbf{v}^{k-1})$ ,  $w_v^{k-1} = \min\left(\frac{t_{k-1}-1}{t_k}, \theta \sqrt{\frac{\mathbf{L}_v^{k-1}}{\mathbf{L}_v^{k-2}}}\right)$  with  $t_0 = 1$  and  $t_k = 1/2\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$ ;  $\mathbf{L}_v^{k-1} = |(\mathbf{B}_v^{k-1})^T(\mathbf{B}_v^{k-1})|$  with  $\mathbf{B}_v^{k-1} = \mathbf{u} \odot \mathbf{v}$  and  $\mathbf{G}_v^{k-1} = (\check{\mathbf{v}}^{k-1}(\mathbf{B}_v^{k-1})^T - \mathbf{D}_{(2)})\mathbf{B}_v^{k-1}$

$$\mathbf{u}^k = \max(0, \check{\mathbf{u}}^{k-1} - \mathbf{G}_u^{k-1}/\mathbf{L}_u^{k-1})$$

where  $\check{\mathbf{u}}^{k-1} = \mathbf{u}^{k-1} + w_{\mathbf{u}}^{k-1}(\mathbf{u}^{k-1} - \mathbf{u}^{k-1})$ ,  $w_{\mathbf{u}}^{k-1} = \min\left(\frac{t_{k-1}-1}{t_k}, \theta \sqrt{\frac{L_{\mathbf{u}}^{k-1}}{L_{\mathbf{u}}^{k-2}}}\right)$  with  $t_0 = 1$  and  $t_k = 1/2\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$ ;  $L_{\mathbf{u}}^{k-1} = |(\mathbf{B}_{\mathbf{u}}^{k-1})^T(\mathbf{B}_{\mathbf{u}}^{k-1})|$  with  $\mathbf{B}_{\mathbf{u}}^{k-1} = \mathbf{v} \odot \mathbf{v}$  and  $\mathbf{G}_{\mathbf{u}}^{k-1} = (\check{\mathbf{u}}^{k-1}(\mathbf{B}_{\mathbf{u}}^{k-1})^T - \mathbf{D}_{(3)})\mathbf{B}_{\mathbf{v}}^{k-1}$

Notice that, since the tensor is composed of symmetric matrices, the first and the second sub-problems of equation (2) are identical and therefore one only needs an updating rule for  $\mathbf{v}$ . Nevertheless I prefer to show the decomposition for generic (directed) tensors.

Similarly to the TOPHITS algorithm (36), the overall spatial dissimilarity score of a generic stock  $i$  is found as a function of the scores of the rest of the stocks weighted by the product of the distances connecting them to stock  $i$ , and of the temporal profile score of the period in which the distances are observed. The temporal profile score attached to a period, on the other hand, is a weighted sum of the distances recorded in that period. Where each distance is weighted by the product of the spatial dissimilarity score of the stocks connected by such distance.

In this way, the spatial dissimilarity vectors retain also elements representing the temporal evolution of the distances and only the "next step" value of the temporal profile vector has to be inferred from past data. This is a perspective not available when computing link predictions using matrix-based approaches.

A temporal link prediction, naturally follows from the decomposition and can be used to infer future distances between stocks, and, on the basis of these forecasts, to predict future prices.

**Competing interests statement:** The author declares no competing financial interests.

**Author contributions:** Alessandro Spelta provided the codes, the computer simulations, the figures and wrote the paper.

## References

1. Bischofs, I. B., Hug, J. A., Liu, A. W., Wolf, D. M., Arkin, A. P., Complexity in bacterial cell–cell communication: Quorum signal integration and subpopulation signaling in the *Bacillus subtilis* phosphorelay. *Proceedings of the National Academy of Sciences*, **106**, 6459-6464, (2009).
2. Anchang, B., Sadeh, M. J., Jacob, J., Tresch, A., Vlad, M. O., Oefner, P. J., Spang, R., Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. *Proceedings of the National Academy of Sciences*, **106**, 6447-6452, (2009).
3. Litvin, O., Causton, H. C., Chen, B. J., Pe'Er, D., Modularity and interactions in the genetics of gene expression. *Proceedings of the National Academy of Sciences*, **106**, 6441-6446, (2009).
4. Savageau, M. A., Coelho, P. M., Fasani, R. A., Tolla, D. A., & Salvador, A., Phenotypes and tolerances in the design space of biochemical systems. *Proceedings of the National Academy of Sciences*, **106**, 6435-6440, (2009).
5. Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., White, D. R., Economic networks: The new challenges. *science*, **325**, 422, (2009).
6. Moon, H., Lu, T. C., Network Catastrophe: Self-Organized Patterns Reveal both the Instability and the Structure of Complex Networks. *Scientific reports*, **5**, (2015).
7. Preis, T., Schneider, J. J., Stanley, H. E., Switching processes in financial markets. *Proceedings of the National Academy of Sciences*, **108**, 7674-7678, (2011).
8. Venegas, J. G., Winkler, T., Musch, G., Melo, M. F. V., Layfield, D., Tgavalekos, N., Harris, R. S., Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature*, **434**, 777-782, (2005).
9. Liu, R., Yu, X., Liu, X., Xu, D., Aihara, K., Chen, L., Identifying critical transitions of complex diseases based on a single sample. *Bioinformatics*, **30**, 1579-1586, (2014).
10. Rietkerk, M., Dekker, S. C., de Ruiter, P. C., van de Koppel, J., Self-organized patchiness and catastrophic shifts in ecosystems. *Science*, **305**, 1926-1929, (2004).

11. Kéfi, S., Guttal, V., Brock, W. A., Carpenter, S. R., Ellison, A. M., Livina, V. N., Dakos, V., Early warning signals of ecological transitions: methods for spatial patterns. *PloS one*, **9**, e92097, (2014).
12. Martinerie, J., Adam, C., Le Van Quyen, M., Baulac, M., Clemenceau, S., Renault, B., Varela, F. J., Epileptic seizures can be anticipated by non-linear analysis. *Nature medicine*, **4**, 1173-1176, (1998).
13. Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., Held, H., Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, **105**, 14308-14312, (2008).
14. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Sugihara, G., Early-warning signals for critical transitions. *Nature*, **461**, 53-59, (2009).
15. Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Scheffer, M., Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PloS one*, **7**, e41010, (2012).
16. Dakos, V., van Nes, E. H., Donangelo, R., Fort, H., Scheffer, M., Spatial correlation as leading indicator of catastrophic shifts. *Theoretical Ecology*, **3**, 163-174, (2010).
17. Mandelbrot, B. B., *The variation of certain speculative prices* (Springer New York, 1997).
18. Shen, J., Zheng, B., Cross-correlation in financial dynamics. *EPL (Europhysics Letters)*, **86**, 48005, (2009).
19. Utsugi, A., Ino, K., & Oshikawa, M., Random matrix theory analysis of cross correlations in financial markets. *Physical Review E*, **70**, 026110, (2004).
20. Preis, T., Kenett, D. Y., Stanley, H. E., Helbing, D., Ben-Jacob, E., Quantifying the behavior of stock correlations under market stress. *Scientific reports*, **2**, (2012).
21. Onnela, J. P., Chakraborti, A., Kaski, K., Kertiész, J., Dynamic asset trees and portfolio analysis. *The European Physical Journal B-Condensed Matter and Complex Systems*, **30**, 285-288, (2002).
22. Araújo, T., Louçã, F., The geometry of crashes. A measure of the dynamics of stock market crises. *Quantitative Finance*, **7**, 63-74, (2007).

23. Tumminello, M. I. C. H. E. L. E., Di Matteo, T., Aste, T., Mantegna, R. N., Correlation based networks of equity returns sampled at different time horizons. *The European Physical Journal B*, **55**, 209-217, (2007).
24. Lü, L., Zhou, T., Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, **390**, 1150-1170, (2011).
25. Katz, E., Blumler, J. G., Gurevitch, M., Uses and gratifications research. *The Public Opinion Quarterly*, **37**, 509-523, (1973).
26. Dunlavy, D. M., Kolda, T. G., Acar, E., Temporal link prediction using matrix and tensor factorizations. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, **5**, 10, (2011).
27. Gao, S., Denoyer, L., Gallinari, P., Link pattern prediction with tensor decomposition in multi-relational networks. In *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on* (pp. 333-340). IEEE, (2011, April).
28. Mantegna, R. N., & Stanley, H. E., *Introduction to econophysics: correlations and complexity in finance* (Cambridge university press, 1999).
29. Carroll, J. D., & Chang, J. J., Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, **35**, 283-319, (1970).
30. Harshman, R. A., Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. (1970).
31. The price forecasts are not compared with their current values due to a bias given by the fact that, when the correlation is zero, the distance takes a value of  $\sqrt{2}$  artificially incrementing the price of the stocks.
32. Preis, T., Moat, H. S., & Stanley, H. E., Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, **3**, (2013).
33. Araújo, T., Spelta, A., Structural changes in cross-border liabilities: a multidimensional approach. *Physica A: Statistical Mechanics and its Applications*, **394**, 277-287, (2014).
34. Seber, G. A., *Multivariate observations*, (John Wiley & Sons, 2009).

- 35.** Xu, Y., Yin, W., A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences*, **6**, 1758-1789, (2013).
- 36.** Bader, B. W., Kolda, T. G., Matlab tensor toolbox version 2.5. *Available online, January, 7*, (2012).
- 37.** Kolda, T., & Bader, B., The TOPHITS model for higher-order web link analysis. In *Workshop on link analysis, counterterrorism and security*, **7**, 26-29, (2006, April)



## Supplementary Information

### Stock price forecast of the supplementary dataset

This section presents the application of the tensor forecast technique to the dataset used to predict the movements of the S&P500 composite index. In the main text the distances between 455 stocks computed over the period 2004/06/24 to 2013/04/30 have been employed to forecast the S&P500 index dynamics along with the Principal Coordinates Analysis and the resulting volume of the cloud of points embedded in 6<sup>th</sup> dimension space.

Those data are analyzed here with the aim of forecasting the movements of each stock price and to assess the quality of the signal produced by the methodology.

Similarly to the results obtained for the first dataset, Fig. S1(a) shows the performance of the investment strategy in terms of cumulative returns obtained for each of the 455 stocks along with the average performance (black line). Also in this case, the technique does not guarantee positive returns for all the stocks but the positive average performance indicates that this is the case for most of the stocks. Moreover the decreasing trend associated with the 2008 financial crisis is correctly anticipated, as depicted by Fig. S1(b), while the long run positive trend results in the increasing cumulative returns associated with long investment positions (Fig. S1(b)).

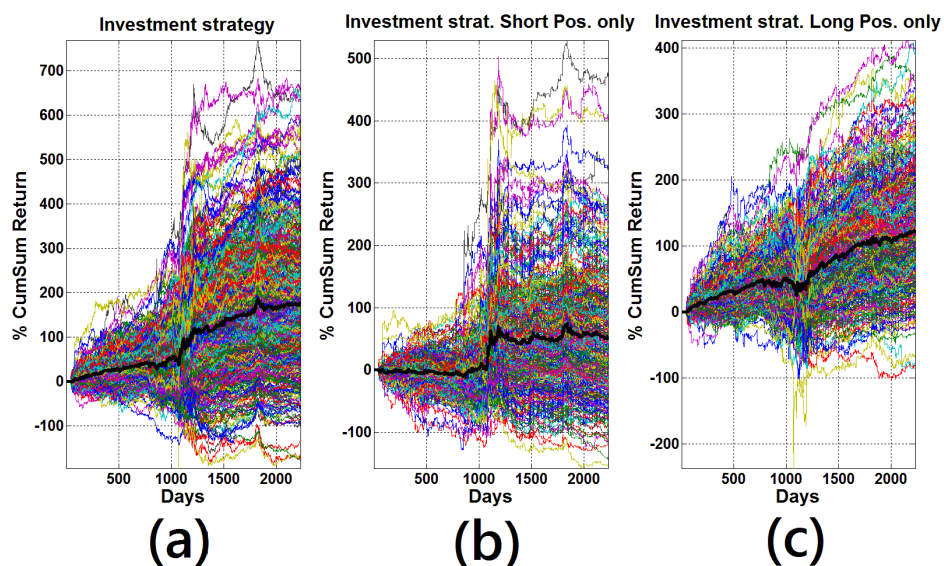


Figure S1: Cumulative sum of returns obtained for each stock together with the average cumulative performance (solid black line). Figure S1(a) shows the performance of the investment strategy. Panels (b)-(c) display the cumulative returns obtained only taking short or long positions respectively. This helps in quantifying how the correct forecast of booms and burst phases affects the overall cumulative returns. The simulations are obtained using the following parameters:  $n_1 = 15$ ,  $Z = 25$ ,  $n_2 = 7$  and the exponential smoothing parameter is set to be equal to 0.2.

In order to verify the hypothesis that larger absolute differences between two consecutive forecasts are associated with better signals as regards the accuracy of the prediction, the signals are sorted in descending order and Fig. S2 displays the average cumulative sum of the returns associated with different quantiles of the distribution of the signals.

Figure S2(a) stresses that stronger signals produce better forecasts, indeed the upper blue line associated with the strongest signal yields the highest cumulative returns (385%), the green line that encompasses the average cumulative returns produced by the two strongest signals comes second, representing a cumulative return of 281%.

The other lines illustrate the decreasing performances associated with the cumulative sum of signals of gradually lower quantiles of the distribution. This feature holds for both short and long investment positions as shown by Figs. S3(b) and (c) respectively.

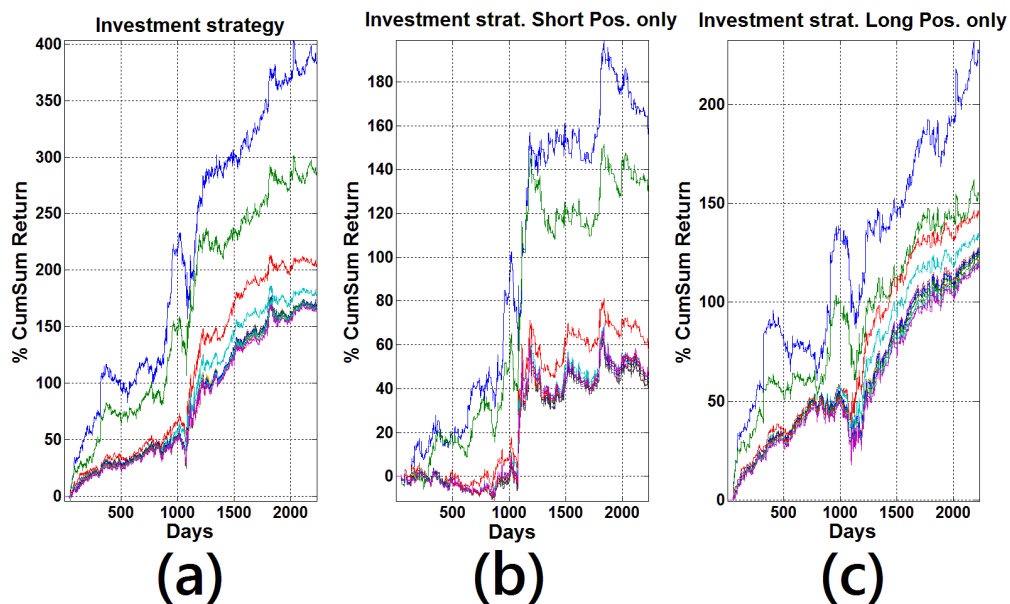


Figure S2: Cumulative sum of the returns associated with different quantiles of the signals' distribution. The quality of each signal depends on the absolute difference between the two consecutive forecasts that compose the signal. The larger the difference the better the signal. While Figure S2(a) shows the performance of the whole investment strategy, panels (b)-(c) display the cumulative returns obtained by only taking short or long positions respectively. The simulations are obtained using the following parameters:  $n_1 = 15$ ,  $Z = 25$ ,  $n_2 = 7$  and the exponential smoothing parameter is set to be equal to 0.2.

Finally Fig. S3 shows the sensitivity of the methodology to changes of the parameters  $n_1$  and  $n_2$  by indicating the cumulative returns obtained at the end of the time sample. In particular Fig. S3(a) shows the average performance of the method in forecasting the dynamic of the 455 stocks, while Fig. S3(b) and (c) refer only to the performance of short and long investment positions. Figure S3 in the bottom panels shows the returns obtained by looking at the best signal, for the whole investment strategy (d) and for short (e) and long (f) positions.

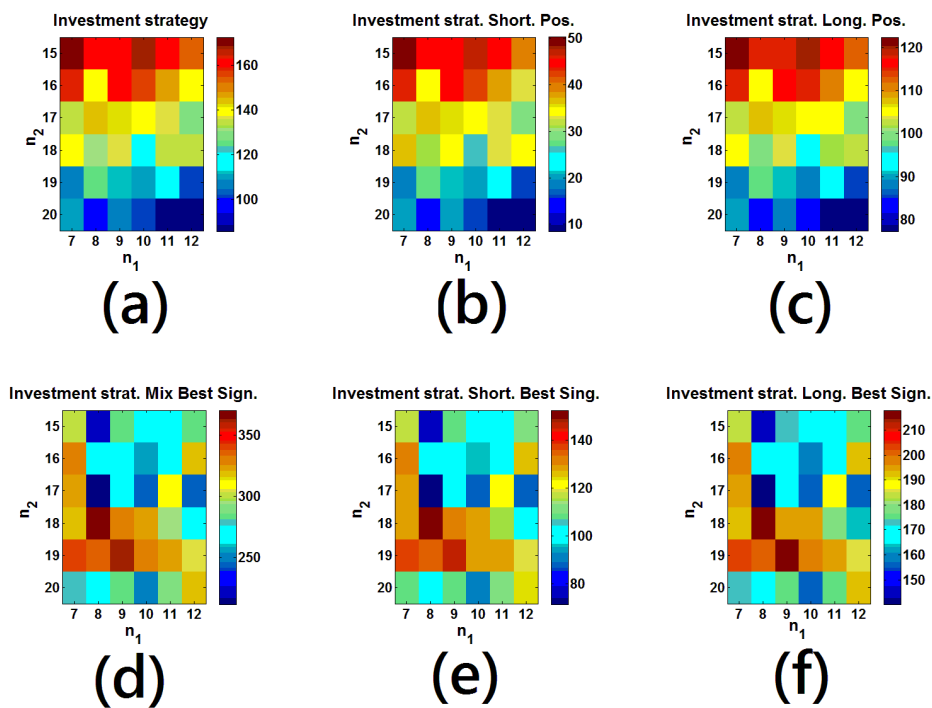


Figure S3: Cumulative returns obtained at the end of the time sample for different values of  $n_1$  and  $n_2$ . Panel S3(a) represents the end-of-sample cumulative performance obtained by averaging the returns obtained in forecasting the dynamic of the 455 stocks of the second dataset. Panel (b) displays the end-of-sample returns obtained by investing using only short positions and (c) encompasses the results for the long positions investment strategies. Figure S3(d) shows the results for the investment strategy

obtained by looking only at the best signal produced by the method together with the results for short (e) and long (f) investment strategies only.

### Percentage of correct prediction

Complementary to the cumulative returns obtained from the application of the technique to different dataset, this section presents the percentage of correct predictions achieved by the methodology within the different parameter space. For different quantile of the returns distribution, Fig. S4 shows the average percentage of correct predictions obtained in forecasting the movements of the 388 stocks employed in the first dataset as long as  $n_1$  and  $n_2$  vary. Although Fig. S4(a) informs that the technique is able to correctly predict at least the 50% of price changes, this probability increases for higher quantiles of the distribution (see Figs. S4(b) and (c)), reaching values near 60% for the 99<sup>th</sup> quantile (see Fig. S4(d)). Although this quantile encompasses only few price changes, the returns obtained by correctly forecasting these changes range from 18 to 66% of total returns. This means that large market movements are correctly anticipated with a higher likelihood.

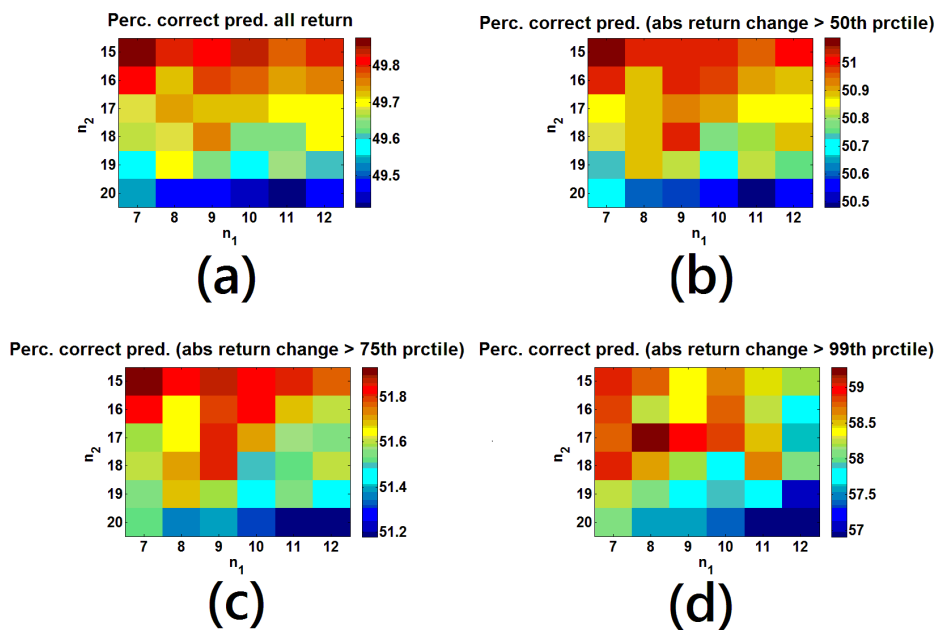


Figure S4: Average percentage of correct predictions for the 388 stocks in the first dataset as long as parameters as  $n_1$  and  $n_2$  changes and for different quantiles of the returns distribution. In particular, Figure S4(a) shows the percentage of correct

predictions for the distribution of all returns. Despite the fact that the percentage is lower than 50, it does not consider the relative impact of changes of different sizes. Figures S4(b), (c) and (d) take into account these imbalances showing the percentage of correct predictions for increasing quantiles of the returns' distribution, namely the 50<sup>th</sup> the 75<sup>th</sup> and the 99<sup>th</sup> respectively.

The next step is to investigate whether stronger signals give rise to better forecasts in terms of the percentage of correct predictions. Figure S5 shows the percentage of correct predictions produced by the best signal for different quantiles of the distribution of returns. By comparing Figs. S4 and S5 it clearly emerges that the strongest signals outperform the average investment strategy, going beyond the threshold of 50% for the whole distribution and reaching a percentage of 65 for the highest quantile.

The same analysis is performed for the second dataset encompassing 455 stock price time series. Also in this case the likelihood of having a correct prediction is higher for higher quartile of the returns' distribution and, moreover, the best signal provides better forecasts compared with the average percentage of correct predictions obtained by forecasting the movements of the 455 stocks. Additionally, Figures S6 and S7 show that, in general, better predictions are obtained for this dataset compared with the previous one. This result seems to indicate that a more complete dataset of the basket of stocks composing the S&P500 is useful for analyzing the dynamic of this index because it includes additional relationships between previously discarded stock prices.

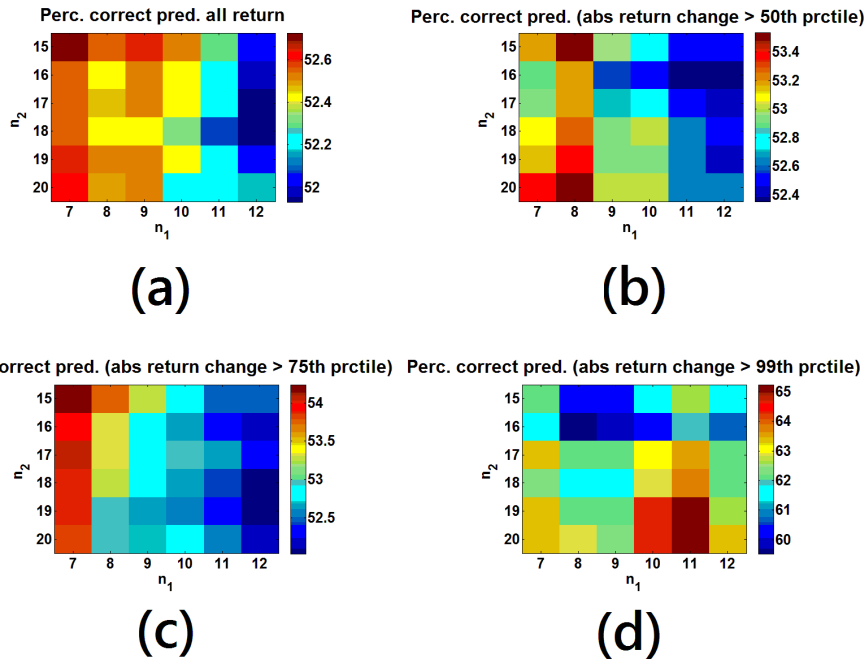


Figure S5: Percentage of correct predictions obtained by looking only at the strongest signal in the first dataset for different values of  $n_1$  and  $n_2$  and for different quantiles of the distribution of returns. In particular, Figure S5(a) shows the percentage of correct predictions for the whole distribution while Figure S5 (b), (c) and (d) display the percentage of correct predictions for increasing quantiles of the returns' distribution, the 50<sup>th</sup> the 75<sup>th</sup> and the 99<sup>th</sup> respectively.

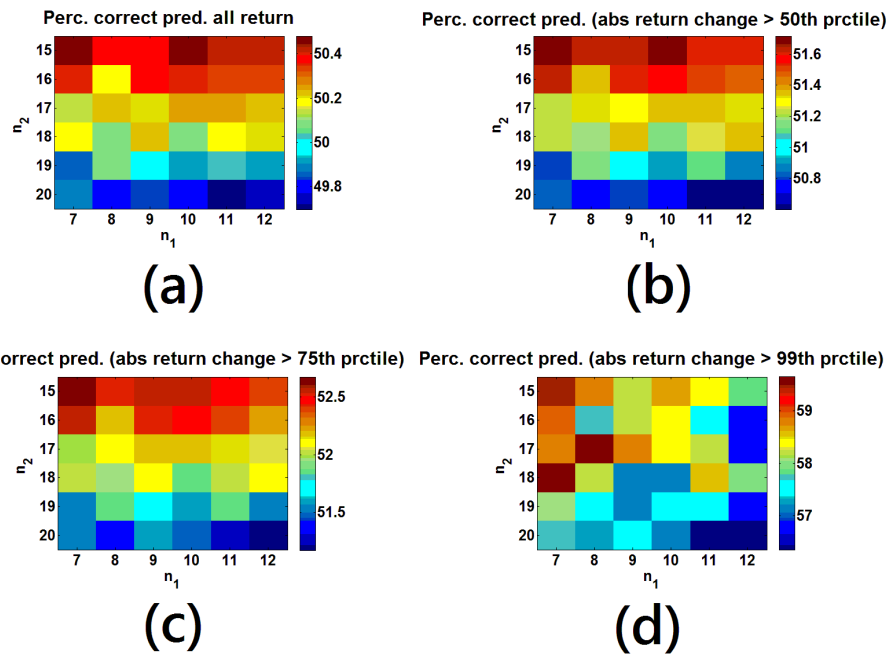


Figure S6: Percentage of correct predictions for the 455 stocks in the second dataset as long as  $n_1$  and  $n_2$  changes and for different quantiles of the distribution of returns. In particular, Figure S6(a) shows the percentage of correct predictions for the

whole distribution and panels (b), (c) and (d) display this percentage for increasing quantiles of the distribution of returns, namely the 50<sup>th</sup>, the 75<sup>th</sup> and the 99<sup>th</sup> respectively.

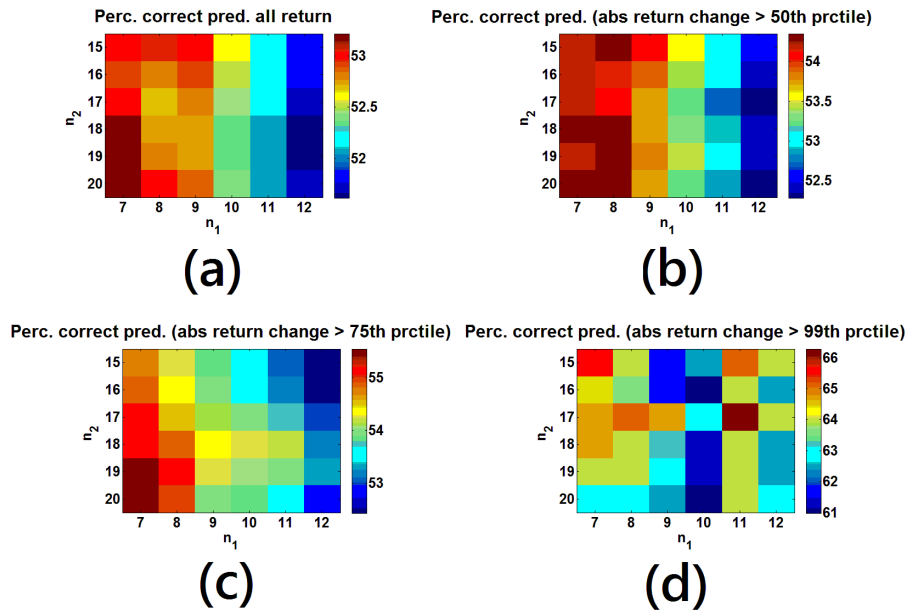


Figure S7: Percentage of correct predictions produced by the best signal in the second dataset as  $n_1$  and  $n_2$  changes and for different quantiles of the distribution of returns. Figure S7(a) shows the percentage of correct predictions for the whole distribution and Figure S7(b), (c) and (d) display this percentage for increasing quantiles of the distribution of returns, namely the 50<sup>th</sup>, the 75<sup>th</sup> and the 99<sup>th</sup> respectively.

Finally, for the sake of completeness, the percentage of correct predictions regarding the S&P500 composite index is also shown in Fig. S8. As for the previous cases, the likelihood of having a correct prediction is higher for higher quartiles of the distribution of returns but not for the whole parameter space. Indeed, it is true that for  $n_2 = 20$  the number of correct predictions is almost 59% but for lower values of  $n_2 = 20$  and high values of  $n_1$  the fraction of correct predictions is very low.

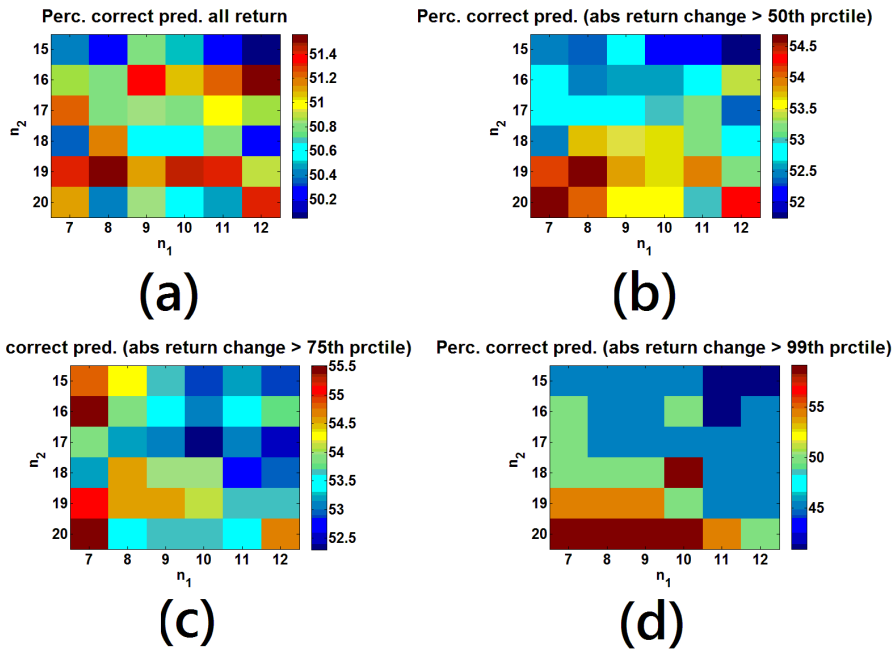


Figure S8: Percentage of correct predictions for the S&P500 composite index as long as  $n_1$  and  $n_2$  changes and for different quantiles of the distribution of returns. In particular, Figure S8(a) shows the percentage of correct predictions for the whole distribution and panels (b), (c) and (d) display this percentage for increasing quantiles of the returns' distribution, namely the 50<sup>th</sup>, the 75<sup>th</sup> and the 99<sup>th</sup> respectively.



1. L. Colombo, H. Dawid, *Strategic Location Choice under Dynamic Oligopolistic Competition and Spillovers*, Novembre 2013.
2. M. Bordignon, M. Gamalerio, G. Turati, *Decentralization, Vertical Fiscal Imbalance, and Political Selection*, Novembre 2013.
3. M. Guerini, *Is the Friedman Rule Stabilizing? Some Unpleasant Results in a Heterogeneous Expectations Framework*, Novembre 2013.
4. E. Brenna, C. Di Novi, *Is caring for elderly parents detrimental to women's mental health? The influence of the European North-South gradient*, Novembre 2013.
5. F. Sobbrío, *Citizen-Editors' Endogenous Information Acquisition and News Accuracy*, Novembre 2013.
6. P. Bingley, L. Cappellari, *Correlation of Brothers Earnings and Intergenerational Transmission*, Novembre 2013.
7. T. Assenza, W. A. Brock, C. H. Hommes, *Animal Spirits, Heterogeneous Expectations and the Emergence of Booms and Busts*, Dicembre 2013.
8. D. Parisi, *Is There Room for 'Fear' as a Human Passion in the Work by Adam Smith?*, Gennaio 2014.
9. E. Brenna, F. Spandonaro, *Does federalism induce patients' mobility across regions? Evidence from the Italian experience*, Febbraio 2014.
10. A. Monticini, F. Ravazzolo, *Forecasting the intraday market price of money*, Febbraio 2014.
11. Tiziana Assenza, Jakob Grazzini, Cars Hommes, Domenico Massaro, *PQ Strategies in Monopolistic Competition: Some Insights from the Lab*, Marzo 2014.
12. R. Davidson, A. Monticini, *Heteroskedasticity-and-Autocorrelation-Consistent Bootstrapping*, Marzo 2014.
13. C. Lucifora, S. Moriconi, *Policy Myopia and Labour Market Institutions*, Giugno 2014.
14. N. Pecora, A. Spelta, *Shareholding Network in the Euro Area Banking Market*, Giugno 2014.
15. G. Mazzolini, *The economic consequences of accidents at work*, Giugno 2014.
16. M. Ambrosanio, P. Balduzzi, M. Bordignon, *Economic crisis and fiscal federalism in Italy*, Settembre 2014.
17. P. Bingley, L. Cappellari, K. Tatsiramos, *Family, Community and Long-Term Earnings Inequality*, Ottobre 2014.
18. S. Frazzoni, M. L. Mancusi, Z. Rotondi, M. Sobrero, A. Vezzulli, *Innovation and export in SMEs: the role of relationship banking*, Novembre 2014.
19. H. Gnutzmann, *Price Discrimination in Asymmetric Industries: Implications for Competition and Welfare*, Novembre 2014.
20. A. Baglioni, A. Boitani, M. Bordignon, *Labor mobility and fiscal policy in a currency union*, Novembre 2014.
21. C. Nielsen, *Rational Overconfidence and Social Security*, Dicembre 2014.
22. M. Kurz, M. Motolese, G. Piccillo, H. Wu, *Monetary Policy with Diverse Private Expectations*, Febbraio 2015.
23. S. Piccolo, P. Tedeschi, G. Ursino, *How Limiting Deceptive Practices Harms Consumers*, Maggio 2015.
24. A.K.S. Chand, S. Currarini, G. Ursino, *Cheap Talk with Correlated Signals*, Maggio 2015.

25. S. Piccolo, P. Tedeschi, G. Ursino, *Deceptive Advertising with Rational Buyers*, Giugno 2015.
26. S. Piccolo, E. Tarantino, G. Ursino, *The Value of Transparency in Multidivisional Firms*, Giugno 2015.
27. G. Ursino, *Supply Chain Control: a Theory of Vertical Integration*, Giugno 2015.
28. I. Aldasoro, D. Delli Gatti, E. Faia, *Bank Networks: Contagion, Systemic Risk and Prudential Policy*, Luglio 2015.
29. S. Moriconi, G. Peri, *Country-Specific Preferences and Employment Rates in Europe*, Settembre 2015.
30. R. Crinò, L. Ogliari, *Financial Frictions, Product Quality, and International Trade*, Settembre 2015.
31. J. Grazzini, A. Spelta, *An empirical analysis of the global input-output network and its evolution*, Ottobre 2015.
32. L. Cappellari, A. Di Paolo, *Bilingual Schooling and Earnings: Evidence from a Language-in-Education Reform*, Novembre 2015.
33. A. Litina, S. Moriconi, S. Zanjaj, *The Cultural Transmission of Environmental Preferences: Evidence from International Migration*, Novembre 2015.
34. S. Moriconi, P. M. Picard, S. Zanjaj, *Commodity Taxation and Regulatory Competition*, Novembre 2015.
35. M. Bordignon, V. Grembi, S. Piazza, *Who do you blame in local finance? An analysis of municipal financing in Italy*, dicembre 2015.
36. A. Spelta, *A unified view of systemic risk: detecting SIFIs and forecasting the financial cycle via EWSs*, gennaio 2016.
37. N. Pecora, A. Spelta, *Discovering SIFIs in interbank communities*, febbraio 2016.
38. M. Botta, L. Colombo, *Macroeconomic and Institutional Determinants of Capital Structure Decisions*, aprile 2016.
39. A. Gamba, G. Immordino, S. Piccolo, *Organized Crime and the Bright Side of Subversion of Law*, maggio 2016.
40. L. Corno, N. Hildebrandt, A. Voena, *Weather Shocks, Age of Marriage and the Direction of Marriage Payments*, maggio 2016.
41. A. Spelta, *Stock prices prediction via tensor decomposition and links forecast*, maggio 2016.