Financial interaction networks inferred from traded volumes

Hong-Li Zeng$^1$, Rémi Lemoy$^{1,2}$ and Mikko Alava$^1$

1 Department of Applied Physics, Aalto University, Espoo, Finland
2 Department of Information and Computer Science, Aalto University, Espoo, Finland

E-mail: hong.zeng@aalto.fi, remilemoy@gmail.com and mikko.alava@aalto.fi

Received 18 November 2013
Accepted for publication 12 May 2014
Published 11 July 2014

Abstract. In order to use the advanced inference techniques available for Ising models, we transform complex data (real vectors) into binary strings, using local averaging and thresholding. This transformation introduces parameters, which must be varied to characterize the behaviour of the system. The approach is illustrated on financial data, using three inference methods—equilibrium, synchronous and asynchronous inference—to construct functional connections between stocks. We show that the traded volume information is enough to obtain well-known results about financial markets that use, however, presumably richer price information: collective behaviour (‘market mode’) and strong interactions within industry sectors. Synchronous and asynchronous Ising inference methods give results that are coherent with equilibrium ones and are more detailed since the obtained interaction networks are directed.

Keywords: models of financial markets, network reconstruction, statistical inference, kinetic Ising models

ArXiv ePrint: 1311.3871
1. Introduction

Inferring Ising models is an interesting and important field in the study of complex systems. Growing amounts of data are being generated in different domains studying complex interacting systems: biology, economics, finance, social sciences and others. Inverse Ising problems provide very flexible tools for guessing the interaction patterns from the observed data, even for non-equilibrium systems [1–4], by linking pairwise correlations in the data to Ising couplings of the inferred model. An important question, which is studied in the present work, concerns the mapping of these complex data into binary variables. The data studied in inverse Ising problems consists of binary strings, which explains why these models have been preferably applied to systems presenting activity/inactivity patterns, like neural networks [5, 6].

We show here how these tools can be applied to systems that provide more detailed data than activity/inactivity. This study deals with financial data generated by transactions on the New York Stock Exchange (NYSE). The activity of 100 highly traded stocks was recorded over a few years, and each trade is characterized by a time, the volume traded and the price. Here we do not consider the price information and focus on time and volume traded. Although there is a well-studied statistical relationship between the absolute value of price change and the volume traded [7], the link between volume data and the sign of the price change (used in [8] for Ising inference) is not so clear, so that we can consider this information lost. The simplest way to map volume
information to binary variables consists in forgetting about the traded volumes, and considering a trade as an activity, in the spirit of work on neural networks. This is done for instance by Mastromatteo and Marsili [3], a work on which we build. However, in such work, the mapping of the data to binary strings has only one parameter, the length of the window. For large window sizes, one usually has a magnetization going to 1—as at least one trade is quite sure to happen if the window is large enough—and low connected correlations. Then, in order to access phenomena occurring on bigger timescales, we use here a different mapping, which also allows us to consider volume information instead of just activity.

The paper presents the data, mapping and inference methods in section 2, the inference methods in section 3, the couplings obtained in section 4, and describes the inferred financial networks in section 5 before concluding.

2. Data and inference

We focus on the trades between 100 highly traded stocks of the NYSE, for 100 trading days between 01.02.2003 and 30.05.2003. As [3], we only study the $10^4$ central seconds of each day so that the system is not perturbed by the opening and closing periods of the stock exchange. Nonetheless, a Fourier transform of the data still shows a sharp peak at a frequency corresponding to this $10^4$ s long day—a result also observed in [9]. This should correspond to daily patterns of human activity, reflected in the market data, but maybe also to the effect of the boundaries between days.

This work has two particular details in the way it maps multi-dimensional financial trade data to binary strings of activity and inactivity. The first one is the use of a sliding time window, of length $\Delta t$, and shifted by a constant $\Delta s = 1$ s, which is the time resolution of the data. This means that the information contained in two mapped datapoints separated by a time less than $\Delta t$ is partly redundant. In the case of a simple activity/inactivity mapping (neuronal data for instance), it also means that no information from the original data is lost.

The second characteristic of our mapping is more important, and allows us to consider also the volume information related to a trade. This is motivated by the fact that the distribution of traded volumes in the data is very broad, as illustrated by the top panel of figure 1, where the distribution looks roughly like a power law, and spreads over more than three decades (see also [9] for distributions of trading volume rates). Our goal is to capture this important volume information through our mapping.

To this end, we consider for each stock $i$ the sum $V_i(t, \Delta t)$ of the volumes traded in the time window of length $\Delta t$ beginning at time $t$, and compare it to a given volume threshold $V_i^{\text{th}} = \chi V_i^{\text{av}} \Delta t$, where $V_i^{\text{av}}$ is the average volume (over the whole time series) of the considered stock traded per second$^3$, and $\chi$ a parameter governing our volume threshold:

\[
s_i(t) = \begin{cases} 
1 & \text{if } V_i(t, \Delta t) \geq V_i^{\text{th}} \\
-1 & \text{if } V_i(t, \Delta t) < V_i^{\text{th}}
\end{cases}
\]  

Note that $V_i^{\text{av}}$ is related to the average of the traded volume histogram presented on the top panel of figure 1 only through the average frequency of trades, which we do not study here.

\[\text{doi:10.1088/1742-5468/2014/07/P07008}\]
This mapping, illustrated on the bottom panel of figure 1, ensures that volume information (and not only trade/no trade information) is taken into account in the binary strings obtained.

For instance with $\chi = 1$, there will be a signal if the total volume traded in the considered time window exceeds the average. In addition, it preserves the possibility of using only trade/no trade information: if $\chi \to 0$ (concretely, if $\chi$ is small enough that $V_i^{th}$ is for every stock $i$ smaller than the minimal traded volume, which is 100 in the data), we obtain the usual trade/no trade pattern used for instance by Mastromatteo and Marsili [3].

Using this mapping, parameters $\Delta t$ and $\chi$ control the time and volume scales, which will be explored by the inference. However, in order to completely characterize the behaviour of the system, the inference should be done for all possible values of these parameters.
\(\Delta t\) and \(\chi\), giving different results each time. The goal of this study is then to find values of the parameters that yield inferred couplings containing interesting information.

We note that this volume-threshold mapping is also interesting for high frequency sampling, compared to a price-based mapping that associates a +1 state with increasing price and a −1 state with decreasing price [8]. Indeed, with high frequency data, there can be no trade in a given time step (see figure 1), in which case the price-based mapping is ambiguous, while our volume-threshold mapping is clearly defined. Note also that instead of using the average volume to obtain a volume scale for the threshold, it is possible to use the median one, which is less sensitive to very high values. It is also possible to apply a threshold based on quantiles of the distribution of volumes traded in a window of length \(\Delta t\) for each stock, which gives results that are consistent with the ones presented in this work, and has the advantage of guaranteeing that all stocks have a non-trivial activity at the considered volume scale. This corresponds to having a threshold on volume rates, a quantity studied in [9].

3. Inference methods

With the mapping described in the previous section, it is natural to define local magnetizations \(m_i\) and connected correlations \(C_{ij}(\tau)\) as follows:

\[m_i = \langle s_i(t) \rangle_t\quad\text{and}\quad C_{ij}(\tau) = \langle s_i(t+\tau)s_j(t) \rangle_t - m_im_j\]

where \(\langle . \rangle_t\) denotes time-averaging over the whole data length. Several inference methods have been developed recently for Ising models. We focus here on mean-field approximation, using three different versions. In each case, we want to infer a fully connected Ising model, composed of \(N = 100\) binary spins \(s_i \in \{-1, 1\}, i = 1, ..., N\), where each pair \((s_i, s_j)\) of spins is linked by a coupling \(J_{ij}\), and each spin \(s_i\) is subject to a local external field \(h_i\). Let us first present the inference formula for coupling strengths, which differs for each method:

- **Equilibrium inference**, which focuses on equal time correlations [10], supposes that the state of the system at each time step is drawn independently from the same distribution, and that \(J_{ij} = J_{ji}\) (symmetric couplings) for all pairs:

\[J_{ij}^{eq} = \frac{\delta ij}{1-m_i^2} - C(0)^{-1}\]

- **Synchronous inference** is suitable for non-equilibrium inference, and considers also time-lagged correlations with a time lag \(\tau\) in addition to equal time correlations [1]:

\[J_{ij}^{syn} = \frac{1}{1-m_i^2} (C(\tau) C(0)^{-1})_{ij}\]

- **Asynchronous inference**, which also models non-equilibrium processes, uses the derivative of the time-lagged correlations \(\dot{C}_{ij}(\tau)\) [11]:

\[J_{ij}^{asy} = \frac{1}{1-m_i^2} \left\{ \frac{dC(\tau)}{d\tau} \bigg|_{\tau=0} \cdot C(0)^{-1} \right\}_{ij}\]
And, each time, the inference formula for the fields is the same:

\[ h_i = \text{arctanh } m_i - \sum_{j \neq i} J_{ij} m_j. \]

The two last methods do not suppose that couplings are symmetric, but the diagonal elements \( J_{ii}, i = 1, \ldots, N \), are supposed to be zero. Synchronous inference assumes that all spins update in parallel at discrete time instances, while asynchronous inference does not have such an assumption: the update times themselves are stochastic variables. The probability that a spin \( s_i \) will be flipped and become \( -s_i \) when updated depends on the local field \( H_i = \sum_j J_{ij} s_j + h_i \) applied to this spin. The asynchronous method is supposedly more powerful, as it monitors the decay in time of all pair correlations, and thus the time derivative \( \dot{C}_ij(\tau) \) appears in the formula.

It can be remarked that, in addition to the mapping parameters \( \Delta t \) and \( \chi \), the synchronous and asynchronous inference methods have a parameter \( \tau \), which is in these respective cases the considered time lag of correlations and the timescale on which the derivative of correlations \( dC(\tau) / d\tau \big|_{\tau=0} \) is computed. For the asynchronous case, this timescale does not appear explicitly in the formula, but arises when the derivative is computed from the data.

For the quality of the inference, [1] (respectively [11]) shows the evolution, with the length of the synthetic data coming from a synchronous (respectively asynchronous) Sherrington–Kirkpatrick model, of the mean square error on the interaction strengths. With a microscopic scale of 1 s, our data has \( L = 10^6 \) time steps, which guarantees a low error. This is discussed further in appendix A.

4. Results

Our mapping with a volume threshold allows us to obtain information for larger timescales at which trade/no trade patterns are inefficient. Indeed, for the trade/no trade mapping, local magnetizations go to one when the window size increases. Then the connected correlations and the inferred couplings tend to zero. This is not the case with our volume-threshold mapping, if the volume-threshold parameter \( \chi \) is well chosen. This is illustrated by the top panel of figure 2, which shows that the average absolute value of the equilibrium couplings \( \langle |J_{eq}| \rangle \), where \( \langle \cdot \rangle \) denotes the average over all pairs of stocks, increases with the window size. This phenomenon, consistent with the Epps effect [12], happens for different values of the volume-threshold parameter \( \chi \). The evolution of the slope is, however, not monotonous with \( \chi \). We study the absolute value of the couplings, because large negative values, which can counterbalance positive ones in an average calculated without using absolute values, might also give interesting information about the system.

Actually, for larger values of the threshold parameter, \( \chi = 2 \) for instance, large negative values of couplings are an important contribution to the high value of \( \langle |J_{eq}| \rangle \). We can note also that the inferred interactions are much bigger than \( L^{-1/2} = 10^{-3} \), which would be observed for independent spins, with data spanning over \( L = 10^6 \) s.

doi:10.1088/1742-5468/2014/07/P07008
Figure 2. Upper panel: $\langle |J_{eq}| \rangle$ versus various time bin lengths $\Delta t$, for different values of $\chi$. Bottom panel: $\langle |J_{syn}| \rangle$ versus $\tau$ for various time bin lengths $\Delta t$, and $\chi = 0.5$. With $\Delta t \lesssim 200$ s, $\langle |J_{syn}| \rangle$ has a maximum at $\Delta t = \tau$, while for $\Delta t > 200$ s, $\langle |J_{syn}| \rangle$ tends to form a plateau.

(see Appendix A.). In Appendix B., we present the corresponding values of the average magnetization and connected correlations.

The same analysis performed on the synchronous couplings shown on the bottom panel of figure 2 shows that for small values of the window size $\Delta t$, the average absolute value of the couplings has a maximum for the correlation time lag $\tau = \Delta t$. This corresponds to the smallest time lag that has no redundancy between the two considered time windows. For higher values of the window size, big values of the time lag give a maximal plateau.

In order to find values of the parameters introduced by the mapping and the inference that yield interesting values of the inferred couplings, a first rough approach is
Financial interaction networks inferred from traded volumes

Figure 3. Histograms $N(J)$ of inferred couplings. Upper panel has four subplots. Upper left: histogram of $N(J_{\text{eq}})$ with different time bins. Upper right: $N(J_{\text{syn}})$, using $\tau = \Delta t$. Bottom left $N(J_{\text{asyn}})$ and bottom right $N(J_{\text{syn}})$ with different values of $\tau$. Bottom panel: couplings obtained by the three inference methods. $J_{\text{syn}}$ and $J_{\text{asyn}}$ are rescaled to have the same standard deviation as $N(J_{\text{eq}})$. For the three versions, $\chi = 0.5$ and $\Delta t = 200$ s, and for synchronous inference $\tau = \Delta t$.

to consider that couplings contain interesting information if they are big in absolute value: this indicates there is a strong interaction between stocks. A more refined, but also laborious, method consists in looking at the distributions of couplings for different values of the parameters. This is done for figure 3. For asynchronous inference, the derivative of the time-lagged correlations $C_y(\tau)$ (see Section 3) is computed through a linear fitting of this function $C_y(\tau)$ using four points: $C(0)$, $C(\Delta t/5)$, $C(2\Delta t/5)$ and $C(3\Delta t/5)$. This explains why the histogram of $J_{\text{asyn}}$ becomes sharper as $\Delta t$ is increased as shown on the upper panel of figure 3, as this parameter is in the denominator of the derivative.
The bottom panel of figure 3 shows that the three inference methods give similar distributions of couplings. For comparison, the distributions are rescaled on the bottom panel so as to have the same standard deviation. The upper panel shows how these distributions change with the parameters. For small timescales, they have a strictly positive mean and some high positive values stand out. For higher timescales, the distributions are more centred around zero, but they keep an asymmetry and more big positive values than negative ones stand out. This prevalence of positive couplings can intuitively be linked with the market mode phenomenon.

Indeed, a spectral analysis of the inferred interaction matrices, illustrated in appendix C along with a spectral analysis of the correlation matrix, yields the well-known fact (at least for correlations of stock prices [8], [13–15]) that a large eigenvalue appears, corresponding to a collective activity of all stocks. This phenomenon is usually named the market mode when correlations are considered. However, we did not manage to give an interpretation to the eigenvectors corresponding to the next eigenvalues in order of magnitude, as they do not seem to correspond to specific industrial or activity sectors for instance.

With increasing values of $\Delta t$, the histograms of $J_{eq}$ (and $J_{syn}$ with $\tau = \Delta t$) become broader, which is consistent with the increasing values of $\langle |J_{syn}| \rangle$ observed on the upper panel of figure 2 (and with the lower panel of figure 2). The last figure in the upper panel of figure 3 shows that the histogram of $J_{syn}$ does not change much with $\tau$ for high values of this parameter, which for $\langle |J_{syn}| \rangle$ is indicated also on the bottom panel of figure 2.

Appendix D describes a more precise measurement $Q_{JJ'}$ of the similarity between interaction matrices $J$ and $J'$ inferred by different methods. Figure 4 (see also figure D1 in the appendix) shows that the inferred coupling matrices for the different methods are much more similar than matrices whose elements are drawn at random from the same distribution, which also indicates the coherence of the different inference methods. Indeed, figure 4 displays high similarities between couplings obtained from equilibrium, and synchronous and asynchronous inference. Synchronous and asynchronous inference give especially close results, while equilibrium inference gives couplings that
Financial interaction networks inferred from traded volumes

10doi:10.1088/1742-5468/2014/07/P07008

J. Stat. Mech. (2014) P07008

differ more from the other two methods. Note that the synchronous and asynchronous methods infer directed networks, whereas equilibrium inference infers an undirected one. All similarities increase when \( \Delta t \) decreases, which is also consistent with the Epps effect and the fact that the system becomes less interacting for small timescales.

It is also interesting to look at the inferred local fields in addition to the couplings. For instance, the magnitudes of the local external field part \( h_i \) and the interaction part \( \sum_j J_{ij} m_j \) of the average local field \( \langle H_i \rangle_t = h_i + \sum_j J_{ij} m_j \) can be studied. We find that the patterns change a lot with the mapping and inference parameters \( \Delta t \), \( \chi \) and \( \tau \). For equilibrium inference, with values of \( \chi \) smaller than 1, the local field part tends to be negative, which means that the system is pushed by the external fields towards negative magnetization (corresponding to having no trade activity on the corresponding time and volume scales), and the interaction part positive, while for values of \( \chi \) bigger than 1, this pattern is reversed. This rule is especially true for big values of \( \Delta t \). As the interaction strengths \( J_{ij} \) are rather positive, this must be related to the fact that magnetization tends to be positive when \( \chi \) is small, and negative when \( \chi \) is large, as illustrated in appendix B.

5. Financial networks

Although the spectral analysis does not distinguish industrial sectors, these can still be found in the inferred couplings, over a wide range of parameters, as figure 5 shows. We focus only on the largest couplings, which can easily be explained as being due to the closely related activities of the stocks considered. The left panel of figure 5 shows that with equilibrium inference, more than half the stocks in the data can be displayed on

**Figure 5.** Inferred financial networks, showing only the largest interaction strengths (proportional to the width of links and arrows). Colors are indicative, and chosen by a modularity-based community detection algorithm [16]. Parameters: \( \chi = 0.5 \) and \( \Delta t = 100 \) s. Left panel: equilibrium inference. Right panel: synchronous inference, with \( \tau = 20 \) s.
a network where almost all links have simple economic interpretations. The following couplings in order of magnitude (not displayed on the figure) can also in some cases be given such simple explanations, but this is generally not true for small positive couplings, or negative ones.

The network shown on the left panel of figure 5 presents different communities, which are, most of the time, determined by a common industrial activity. Some of the links are very easy to explain (and often quite robust) using the proximity of activities. For instance, the pairs FNM and FRE (Fannie Mae and Freddie Mac, active in home loans and mortgages), UNP and BNI (Union Pacific Corporation and Burlington Northern Santa Fe Corporation, railroads), BLS and SBC (BellSouth and SBC Communications, two telecommunications companies now merged in AT&T), NCC and PNC (National City Corp and PNC Financial Services, now merged), HD and LOW (The Home Depot and Lowe’s, both retailers of home improvement products), DOW and DD (Dow and DuPont, chemical companies), MRK and PFE (Merck & Co. and Pfizer, pharmaceutical companies), and KO and PEP (The Coca-Cola Company and PepsiCo, beverages).

These two last companies display strong links with the medical sector at different scales of volume and time, such as KO with MDT (Medtronic) and JNJ (Johnson & Johnson). The medical sector is itself linked to the pharmaceutical sector with links between PFE, MRK, LLY (Lilly), BMY (Bristol-Myers Squibb) and SGP (Schering–Plough). Telecommunications (BLS, SBC) are linked to electric power company DUK (Duke Energy), itself linked to electric utilities such as SO (Southern Company).

GE (General Electrics) is for a large range of parameters a very central node, which is consistent with its very diversified activities. On the left panel of figure 5, the relation between PG (Procter & Gamble) and WMT (Walmart), both retailers of consumer goods, is mediated by GE.

The banking sector, with WFC (Wells Fargo), BAC (Bank of America), MER (Merrill Lynch), BSC (Bear Stearns) and LEH (Lehman Brothers), has a structure resembling a chain. It is linked to the electronic technology sector: IBM, TYC (Tyco International), CA (CA Technologies, software company), TXN (Texas Instruments), ADI (Analog Devices) and MOT (Motorola, telecommunications).

The defence and aerospace sector, with GD (General Dynamics), NOC (Northrop Grumman) and BA (Boeing), is linked to engines and machinery with CAT (Caterpillar Inc.) and DE (John Deere), and more strangely, to packaged food with CAG (ConAgra Foods), SYY (Sysco) and K (Kellogg Company).

Some non-intuitive links appear on different time and volume scales and with different inference methods: KO’s links with the medical sector have been mentioned. DD is linked to PG, and GD is linked here to OMC (Omnicom Group, advertising and marketing), and sometimes to UNP or HDI (Harley–Davidson, motorcycles manufacturer).

The right panel of figure 5 presents the results of synchronous inference for the same conditions. It shows that the results of equilibrium and synchronous inference are consistent, and that synchronous inference provides additional information, as it infers an undirected network (all this is also true for asynchronous inference). For instance, GE is clearly a node that influences others and is not strongly influenced itself at this level of interaction, and the financial sector is a directed chain. Stocks whose tickers are not indicated in the text are described in appendix E.
In summary, what do these results tell us about financial markets? First, they are consistent with the results regarding the market mode [13, 14]: most of the interaction strengths found are positive, which indicates that the financial market has a clear collective behaviour, even when only trade and volume information is considered. Stocks tend to be traded or not traded at the same time.

In addition, the strongest inferred interactions can be easily understood by similarities in the industrial activities of the considered stocks. This means that financial activity—*a priori* irrespective of whether there is a bullish or bearish market, as no price information is considered—tends to concentrate on a certain activity sector at a certain time. For price dynamics this phenomenon is well known [8, 15, 17], but it is surprising that it appears also on traded volumes.

Actually, these two phenomena (collective behaviour and strong interactions within industrial sectors) and models such as [18] also support the idea that financial markets rely a lot on imitation. This is a fundamental social behaviour, which is still very much at work in the 21st century society, as illustrated for instance by fashion trends, important news subjects or even popular research themes. A complementary, more exogenous interpretation, is that financial markets are influenced by the same external factors, whose effect is then reflected in the data. As a consequence of this simple behaviour, assessing the predictive power of such inferred models, at least on short timescales, is an interesting perspective for work.

6. Conclusion and perspectives

Using a mapping of complex data to binary strings, we infer Ising models for financial data, based only on traded volume information. We recover standard results (obtained using price information), and the inferred networks give a consistent picture of the financial markets. Synchronous and asynchronous inference give a directed network in agreement with the undirected equilibrium one. An obvious first perspective is then to combine this volume information with price information in a more involved mapping (which could be inspired from the relationship between price and volume [7]), to obtain new insights about the interactions on the stock market. Indeed, price is usually more studied, because it corresponds essentially to the economic value of the considered stock on the market.

More immediate perspectives would be to study the influence of the timescale parameter $\tau$ of synchronous and asynchronous inference, and also to study close mappings using only volume information. Indeed, instead of using the average volume traded per unit of time as a reference, other variables characterizing the stock could be used, in order to obtain different information, for instance an indicator of the economic size of a company, such as revenue, operating income or total assets.

Other immediate possibilities would be to introduce a regularization ($L^1$ for instance), which would be clearer than applying a threshold to determine which interactions are meaningful. Mapping using a threshold based on quantiles of the distribution of volumes traded in a fixed time window for each stock would be an interesting way to overcome a limitation of our method, which is that for big scales of both volume and time, some stocks may have a $-1$ average magnetization (meaning that no traded volume
exceeded the threshold over the whole period), which prevents us from inverting the correlation matrix and making an inference. Finally, in order to test the results of different inference methods (including more complex ones like [19] or [20]), the predictive power of the inferred models on different time and volume scales, assessed for instance by measuring the distance between data generated by the model and real data from the period following that for which the inference was made, could be a valuable indicator of their worth. With additional price information, this predictive power can be expected to be greater. It is also an interesting perspective from a financial engineering point of view, which has different goals.

Acknowledgments

We thank Matteo Marsili for providing the data, and acknowledge interesting discussions with Erik Aurell, Matteo Marsili, Iacopo Mastromatteo, Alexander Mozeika and Onur Dikmen, as well as helpful suggestions by the editor and two anonymous referees. This work was supported by funding from the Finnish graduate school for Computational Science (FICS) and the Centre of Excellence program of the Academy of Finland, for the COMP and COIN Centres. We acknowledge the computational resources provided by the Aalto Science-IT project.

Appendix A. Quality of the inference

We estimate the performance of an inference method in two ways. First, we apply the method to infinite temperature (random) data. Then we check that the inferred couplings have a distribution with zero mean and standard deviation scaling as $1/\sqrt{L}$, where $L$ is the number of time steps. With $N = 20$ and $N = 100$, the results do not seem to depend on the number $N$ of spins.

The second method consists in running an asynchronous Ising model with Glauber dynamics, using the couplings and fields obtained from the equilibrium inference on our financial data with temperature $T = 1$ and the same data length $L = 10^6$. This provides us with synthetic data for the equilibrium inference. The agreement between both sets of couplings is very good, as illustrated by figure A1, where the mean square error is $1.6 \times 10^{-6}$.

Appendix B. Correlations

Figure B1 shows that the connected correlations exhibit non-trivial behaviour as a function of the mapping parameters $\chi$ and $\Delta t$. For low values of the time window length $\Delta t$, correlations are small at any volume scale $\chi$, which can be linked to the fact that the average magnetization tends to $-1$. When $\chi$ is either small or big, correlations are small for high values of $\Delta t$, as the magnetization tends to 1 or $-1$ respectively.

There is more complex behaviour for intermediate values of $\chi$ (and $\Delta t$): connected correlations can either have a maximum for a given value of $\Delta t$, or simply increase with $\Delta t$, which is consistent with the Epps effect. The standard deviation across stocks of the (time averaged) connected correlations behaves with changes in the parameters in a similar way to the average.

doi:10.1088/1742-5468/2014/07/P07008
Figure A1. Comparison between couplings $J_{eq}$ obtained from the equilibrium inference with the financial data (with $\chi = 0.5$ and $\Delta t = 50$ s) and the couplings $J'_{eq}$ obtained from the equilibrium inference with synthetic data using $J_{eq}$. The line shows equality. The synchronous and asynchronous inference methods give similarly low reconstruction errors for the same test (running a correspondingly synchronous or asynchronous Ising model with Glauber dynamics).

Figure B1. Evolution with mapping parameters $\chi$ and $\Delta t$ for the connected correlations, their standard deviations (across stocks), and the magnetization, averaged over the whole dataset (all stocks or pairs of stocks, and whole time period).
Appendix C. Eigenvalues

Figure C1 shows that our mapping preserves the ‘market mode’ phenomenon, as the highest eigenvalue of the connected correlation matrix has a much higher value than the others, and corresponds to a collective eigenmode. The interaction matrix also has a largest eigenvalue, but much less pronounced than for the correlation matrix.

Appendix D. Similarities

We want to study more precisely the closeness between couplings obtained with different methods. The distributions presented in figure 3 give a first idea, but we want to assess how similar individual couplings $J_{ij}$ and $J'_{ij}$ are. To this end, we introduce for two coupling matrices $J_{ij}$ and $J'_{ij}$, a similarity measure $Q_{J,J'}$ given by

$$Q_{J,J'} = \frac{\sum_{i,j} J_{ij} J'_{ij}}{\sum_{i,j} \max(J_{ij}, J'_{ij})^2}$$

Figure D1. Similarity $Q_{J,J'}$ between two random matrices whose elements are drawn independently from the same normal distribution. This distribution has mean 0 in each case, and standard deviation 1, 2, ..., 10 for successive points from left to right.
Financial interaction networks inferred from traded volumes

This function considers coupling matrices as vectors containing all their elements and compares them element by element to give a global similarity measure. It takes real values between 1 (when $J_{ij} = J'_{ij}$ for all $i$ and $j$) and $-1$ (when $J_{ij} = -J'_{ij}$ for all $i$ and $j$), and values close to zero indicate uncorrelated couplings.

In comparison with the similarity measures obtained between pairs of different methods on figure 4, we show on figure D1 that this similarity measure takes small values (smaller than 0.02 in absolute value) when all elements of the vectors $J_{ij}$ and $J'_{ij}$ are drawn independently at random from the same Gaussian distribution of mean 0, for different values of the standard deviation of this distribution. Negative values are even obtained sometimes, as can be expected from the formula defining $Q_{JJ'}$. However, figure 4 displays much closer similarities between the different inference methods.

Appendix E. Description of stocks

| Ticker | Name                     | Activity                           |
|--------|--------------------------|------------------------------------|
| AXP    | American Express         | Financial services                 |
| BSX    | Boston Scientific        | Medical devices                    |
| CAH    | Cardinal Health          | Pharmaceutical and medical products |
| CI     | Cigna                    | Health care management             |
| KRB    | MBNA                     | Banking                            |
| SLE    | Sara Lee Corporation     | Consumer-goods                     |
| WLP    | WellPoint                | Health care management             |

References

[1] Roudi Y and Hertz J 2011 Mean field theory for nonequilibrium network reconstruction Phys. Rev. Lett. 106 048702
[2] Mézard M and Sakellariou J 2011 Exact mean-field inference in asymmetric kinetic ising systems J. Stat. Mech. L07001
[3] Mastromatteo I and Marsili M 2011 On the criticality of inferred models J. Stat. Mech. P10012
[4] Tyrcha J, Roudi Y, Marsili M and Hertz J 2013 The effect of nonstationarity on models inferred from neural data J. Stat. Mech. P03005
[5] Schneidman E, Berry M J, Segev R and Bialek W 2006 Weak pairwise correlations imply strongly correlated network states in a neural population Nature 440 1007–12
[6] Shlens J, Field G D, Gauthier J L, Grivich M I, Petrusca D, Sher A, Litke A M and Chichilnisky E J 2006 The structure of multi-neuron firing patterns in primate retina J. Neurosci. 26 8254–66
[7] Karpoff J M 1987 The relation between price changes and trading volume: a survey J. Financial Quant. Anal. 22 109–26
[8] Bury T 2013 Statistical pairwise interaction model of stock market Eur. Phys. J. B 86 89
[9] Mathiesen J, Angheluta L, Ahlgren P T H and Jensen M H 2013 Excitable human dynamics driven by extrinsic events in massive communities Proc. Natl Acad. Sci. 110 17259–62
[10] Kappen H J and Rodriguez F B 1998 Efficient learning in Boltzmann machines using linear response theory Neural Comput. 10 P1137–56
[11] Zeng H-L, Aurell E, Alava M and Mahmoudi H 2011 Network inference using asynchronously updated kinetic ising model Phys. Rev. E 83 041135
[12] Epps T W 1979 Comovements in stock prices in the very short run J. Am. Stat. Assoc. 74 291–8
[13] Bouchaud H L and Potters M 2003 Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management (Cambridge: Cambridge University Press)

doi:10.1088/1742-5468/2014/07/P07008
[14] Mantegna R N and Stanley E H 1999 An Introduction to Econophysics: Correlations and Complexity in Finance (Cambridge: Cambridge University Press)

[15] Biely C and Thurner S 2008 Random matrix ensembles of time-lagged correlation matrices: derivation of eigenvalue spectra and analysis of financial time-series Quant. Finance 8 705–22

[16] Blondel V D, Guillaume J L, Lambiotte R and Lefebvre E 2008 Fast unfolding of communities in large networks J. Stat. Mech. P10008

[17] Kullmann L, Kertész J and Kaski K 2002 Time-dependent cross-correlations between different stock returns: a directed network of influence Phys. Rev. E 66 026125

[18] Cont R and Bouchaud J P 2000 Herd behavior and aggregate fluctuations in financial markets Macroecon. Dyn. 4 170–96

[19] Nguyen H C and Berg J 2012 Mean-field theory for the inverse Ising problem at low temperatures Phys. Rev. Lett. 109 050602

[20] Aurell E and Ekeberg M 2012 Inverse Ising inference using all the data Phys. Rev. Lett. 108 090201