A statistical physics perspective on criticality in financial markets

Thomas Bury
Service OPERA (CP194/5), Université libre de Bruxelles, Avenue F D Roosevelt 50, B-1050 Brussels, Belgium
E-mail: tbury@ulb.ac.be

Received 5 August 2013
Accepted 10 October 2013
Published 8 November 2013

Online at stacks.iop.org/JSTAT/2013/P11004
doi:10.1088/1742-5468/2013/11/P11004

Abstract. Stock markets are complex systems exhibiting collective phenomena and particular features such as synchronization, fluctuations distributed as power-laws, non-random structures and similarity to neural networks. Such specific properties suggest that markets operate at a very special point. Financial markets are believed to be critical by analogy to physical systems, but little statistically founded evidence has been given. Through a data-based methodology and comparison to simulations inspired by the statistical physics of complex systems, we show that the Dow Jones and index sets are not rigorously critical. However, financial systems are closer to criticality in the crash neighborhood.

Keywords: critical phenomena of socio-economic systems, models of financial markets
A hundred years ago, the Italian economist Pareto introduced the notion of a power-law describing wealth distribution. This is a major concept related to the notion of scale invariance which is widely used in finance and economics (fractional Brownian motion, detrended fluctuation analysis, volatility modeling, etc). Scale invariance is crucial in finance because large absolute returns are power-law distributed [1]. This lack of any characteristic scale is surprising at first glance but finds its foundation in the theory of complex systems. As complex systems compounded of many correlated entities, financial markets exhibit collective behaviors like synchronization or non-random structure, propensity to self-arrange in large correlated structures as highlighted in [2]–[4], large fluctuations [5] and power-laws [1]. Moreover, it has been shown that financial networks share common properties with neural networks [6]. One recovers these features in a class of models belonging to statistical physics, pairwise maximum entropy models which are particularly suited to the capture of collective behaviors. One knows that the market may exhibit some of the former features at a critical state, defined in a precise sense [7], and that maximum entropy models may describe collective behaviors observed in neural networks [8] and in financial markets [4]. It is therefore tempting to think that financial markets are critical [9] (in the statistical physics sense) as seems to be the case for neural networks [10]–[12]. It is not obvious how to empirically validate the presence of a critical state. Criticality was proposed for the approach of log-periodicity [13]. The phenomenological comparison to critical phenomena was made by substituting the
temperature by the time, which becomes therefore the control parameter [14], but this is merely an analogy, log-periodicity should be understood as a dynamical feature rather than a second order phase transition. Indeed, several dynamical mechanisms generate log-periodicity [15]. Criticality was also proposed for agent-based models exhibiting power-laws and volatility clustering at this particular state [16]–[21]. However different rules and models lead to the same qualitative stylized facts. There is still ambiguity since there are non-critical mechanisms which generate stylized facts [22]. Furthermore, detecting criticality is not the same task as modeling complex systems, even if relations obviously exist. A rigorous approach to criticality detection is the inverse (or data-based) approach. A transition between scale dependence and scale invariance is highlighted [23] in this way. Here, we also follow an inverse (starting from the data without initial assumptions) procedure described in [24] and applied in [25]. This procedure is also inspired by statistical physics and provides several statistical tests of criticality.

We find that the considered financial systems are not strictly critical even if some signatures are observed. It is more likely that financial systems do not stay in the same regime and are closer to the criticality when the system gets closer to a crash. The critical scaling parameter (see hereafter) reaches its maximal value in the vicinity of the beginning of the crash. Namely, the response function to a shock (a shock can be a modification of exogenous variables or of the level of stochasticity) has a peak and its position scales with system size towards the operating point (at which the probability distribution is the empirical one) for European market places. The operating point of the Dow Jones is far from the critical one but criticality could be reached if the size of the index is large enough. The distribution of rank of configurations is not a power-law if the system is well sampled and the entropy is not a linear function of the log-likelihood. Moreover, we use a pairwise maximum entropy model [26, 4] to check that the variance of the log-likelihood and the variance of the overlap parameter reach their maxima at $x$-axis coordinates in line with the empirical ones. We compare empirical results to simulations of a multivariate GARCH process and a Monte Carlo Markov chain. They corroborate the empirical findings. Last, we give an interpretation of criticality in financial markets. These findings can be important in portfolio optimization which relies on the market structure (through the correlation matrix, for instance) and to figure out how the market processes information which may eventually lead to a crash.

2. Signatures of criticality

A critical state can be thought of as a state where the system lies at the threshold between order and disorder. If there is no uncertainty, markets are perfectly ordered and thus homogeneous (either positive or negative). In the opposite situation where uncertainty is maximal, markets are completely random and uncorrelated; the probability to observe a positive or negative return is equal to $1/2$ whatever the returns (positive or negative) of other market exchanges. A critical state is halfway between these extreme states, allowing markets which are on the edge of disorder and highly heterogeneous.

Strictly, a critical state can be achieved only for infinite size systems. For finite systems, one will not observe divergences but we will still say critical through the misuse of language and we should compare the empirical results to a finite version of a model which may...
A statistical physics perspective on criticality in financial markets

actually reach criticality (the nearest neighbor Ising model in two dimensions, for instance) as proposed in [25].

Statistical physics provides several tests of criticality. The signatures detailed in [24] will be briefly recalled. First, we define a financial system as a set of stocks (or indices). Relative stock returns are taken as random variables \( r_t \) and can be rewritten as \( r_t = \text{sgn}(r_t)|r_t| \). The signs of stock returns are sometimes considered as uncorrelated and attract less attention. However, correlations may appear in a complicated (non-linear) fashion as synchronization during crises [27]. It is interesting to study orientation (sign of returns) changes since Ising-like models are suited to describe collective behaviors. Moreover, the nature of the relative return sign is more subtle than that of simple independent random variables and can render the particular structure of financial markets [26, 4]. The net orientation is defined as \( m(t) = N^{-1}\sum_i s_{i,t} \); if \( m(t) > 0 \) the market trend is positive for the period \( t \). In order to study orientation changes, we consider a set of \( N \) market indices or \( N \) stocks described by binary variables \( s_i \equiv \text{sgn}(r_t) \) (\( s_i = \pm 1 \) for all \( i = 1, \ldots, N \)). A system configuration will be described by a vector \( s = (s_1, \ldots, s_N) \). The binary variable will be equal to one if the trend of the associated stock is positive and equal to \(-1\) if not. A configuration \((s_1, \ldots, s_N)\) may also be thought of as a binary version of the returns.

One can formally write the probability \( P(s) \) of finding the system in state \( s \) as a Gibbs distribution \( P(s) = Z^{-1} e^{U(s)} \) and without loss of generality set \( Z = 1 \) which leads to the definition of the utility function (or energy \( H = -U \), potential, etc) as the log-likelihood: \( U(s) = \log P(s) \). The rank \( r(s) \) of a given configuration \( s \) is defined as the number of configurations with a higher utility (more frequent) than the value associated to \( s \).

A power-law \(-\log P(s) = \alpha \log r(s) + \text{Cst}\) is a strong signature of criticality. In this framework, it is possible to obtain these quantities directly from a large enough sample and test the validity of Zipf’s law. Another consequence of Zipf’s law is the linearity of the Shannon entropy [25], which measures the average surprise or average log-likelihood, expressed in terms of a utility function [24]. A weaker signature is the divergence of the variance of the likelihood at the operating point (in the limit of an infinite number of entities). For finite systems, the variance of the likelihood should peak near the operating point [25]. It is interesting to study orientation changes, we consider a set of \( N \) market indices or \( N \) stocks described by binary variables \( s_i \equiv \text{sgn}(r_t) \) (\( s_i = \pm 1 \) for all \( i = 1, \ldots, N \)). A system configuration will be described by a vector \( s = (s_1, \ldots, s_N) \). The binary variable will be equal to one if the trend of the associated stock is positive and equal to \(-1\) if not. A configuration \((s_1, \ldots, s_N)\) may also be thought of as a binary version of the returns.

A statistical physics perspective on criticality in financial markets

statistical point of view, this extremum is the point where the deviation from equiprobability of events is the largest. Operation at this point means that the variance of the log-likelihood reaches its largest value, whereas for equiprobable events the variance of the log-likelihood is equal to zero. A large variance of the log-likelihood also implies a large deviation from its mean value, the entropy, and thus large structural changes. The rescaling parameter \( T \) can be thought of as a randomness measure; changing this parameter leads to a reweighting of the empirical distribution. For \( T > 1 \), the distribution will be flattened and closer to the uniform distribution as illustrated in figure 1. The entropy of the remaining distribution
A statistical physics perspective on criticality in financial markets

Figure 1. Schematic illustration of the rescaling of a bimodal distribution as encountered in the Landau phenomenological theory of phase transitions. The original probability density is illustrated by the full line at an arbitrary temperature $T^*$; the rescaled distributions are illustrated by the dashed line (at $T = 0.25T^*$) and the dotted line (at $T = 10T^*$).

will thus be larger than the original one: the closer to the uniform distribution, the larger the entropy. We note that the relation $T\partial S/\partial T$ is useful when direct sampling of the probability distribution is feasible, and the relation $T^{-2} \left[ \langle U^2 \rangle - \langle U \rangle^2 \right]$ allows estimation through a Monte Carlo simulation even if direct sampling is unfeasible. When direct sampling is feasible, one can estimate the empirical distribution as $P(s) = M^{-1}\sum_{i=1}^{M} \delta_{s_i,s}$, where $M$ is the sample length, compute the scaled distribution $P_T(s)$ for any value of $T$ and then use the relation $T\partial S/\partial T$ for the empirical derivation of the response function.

3. Sampling indices and stock exchanges

We observed the opening and closing prices of eight European indices (AEX, BEL, CAC, DAX, EUROSTOXX, FTSE, IBEX, MIB); the sample length $M$ is 2300 trading days (approximately nine trading years encompassing two global crises, 2002–11 period). We consider European stock exchanges because some issues (debt crisis, etc) are specific to these market places and to ensure the simultaneity of time series. We also observed the stocks of the Dow Jones index during $3 \times 10^4$ trading minutes and at daily sampling from 2002 to 2011. We consider two different time-scales to explore the differences when the correlations decrease. According to the Epps effect [28], we would expect that systems sampled at low frequency (daily sampling) should be closer to criticality than systems sampled at larger frequencies (minute sampling, for instance). Positive returns are set to 1 and negative returns to $-1$. The first sample is ten times larger than the number of possible configurations. Indeed, there are two possible values for each variable $s_i$, thus there are $2^N = 256$ configurations. The second sample is not large enough for a satisfactory probability estimation (and thus a direct estimation of entropy). Since entities may be strongly correlated, it is not obvious to know whether the configurations are well sampled or not. In the case of strongly correlated entities, the relevant region in the configuration space is narrow in comparison to independent entities. If the true configuration distribution is sharply peaked, there are only a few relevant states. In this situation, a small ($M < 2^N$) sample is enough to sample the configuration distribution.
properly. In the opposite case where entities are independent, every configuration has the same statistical weight and the sample size must be large \((M \gg 2^N)\). It is crucial to identify the maximum number of entities one should consider to avoid undersampling of the configuration distribution \(P[s]\) because power-laws occur spontaneously in the undersampling regime [29]. In particular, Zipf’s law is only a genuine feature if \(P[s]\) is well sampled. To assess the maximum number of entities to consider in the analysis, we follow the procedure described in [29]. The limit between proper sampling and undersampling is defined by the coordinates of the maximum of \(H[K]\) in the plane \(\{H[K], H[s]\}\), where \(H[s]\) is the entropy of the empirical configuration frequencies and \(H[K]\) is the entropy of the random variable \(K_i = K(s_i)\) which is the number of times the configuration \(s_i\) is observed in the sample. Beyond this point, \(H[K]\) decreases when \(H[s]\) increases which means that configurations are sampled (approximately) the same number of times. Briefly, given a sample of \(M\) independent configurations \((s_1, \ldots, s_M)\), the empirical distribution of the configurations is \(\hat{p}_s \equiv P[s_i = s] = M^{-1}\sum_{i=1}^{M}\delta_{s_i,s}\). The distribution of the random variable \(K_i\), corresponding to the number of times the configuration \(s_i\) occurs in the sample, is written \(P[K_i = k] = k m_k / M\), where \(m_k = \sum_{s} \delta_{s,M\hat{p}_s}\) is the number of configurations that are sampled exactly \(k\) times in the sample. Their entropies are

\[
H[s] = -\sum_{s} \hat{p}_s \log \hat{p}_s = -\sum_k \frac{k m_k}{M} \log \frac{k}{M}
\]

\[
H[K] = -\sum_k k m_k \log \frac{k m_k}{M} = H[s] - \sum_k \frac{k m_k}{M} \log m_k.
\]

These quantities can be evaluated to obtain the statistical significance of each data set. The points in figure 2 have been obtained by considering increasing system size. Each point is obtained in this way and by averaging over several sets of randomly chosen entities (see the caption). Moreover, the theoretical limit is given by the most informative samples (full lines in figure 2) which are those maximizing \(H[K]\) with respect to \(\{m_k, k > 0\}\) and satisfying the constraints \(H[s] \leq N\), \(\sum_k k m_k = M\) and \(H[K] \leq H[s]\) since the random variable \(K\) is a function of \(s\) (see [29] for the complete discussion and derivation).

The statistical significance is illustrated in figure 2 for each data set and for artificial data simulated by fitting a pairwise maximum entropy model (see appendix D). We simulate also a time series of a Sherrington–Kirkpatrick (SK) spin glass of size \(N = 25\) near criticality. The index set is correctly sampled up to seven indices, the Dow Jones at minute up to eight stocks. Increasing the sample length \(M 15\) times, allows us to consider up to \(N = 11\) entities.

A qualitative observation is that if entities are highly correlated (low stochasticity), almost all observed configurations (words) should be such that the mean orientation \(m(t) = N^{-1}\sum_{i} s_{i,t}\) is non-zero. One expects an \(H[s]\) significantly lower than the theoretical upper bound \(N\) and \(H[K] \approx H[s]\) since few different configurations are observed. On the other hand, nearly independent entities do not favor any value of \(m(t)\) and the configuration distribution \(P[s]\) should be approximately uniform. \(H[s]\) should be close to \(\min(N, \log_2 M)\) for large system sizes and \(H[K]\) should be small since each configuration is observed approximately the same number of times. From pairwise maximum entropy (maxent) models [30], one knows that criticality is a regime where no net orientation is observed but where the fluctuations are the largest. We would expect that the sampling
A statistical physics perspective on criticality in financial markets

Figure 2. The statistical significance of data sets. The configuration distribution $P[s]$ is correctly sampled in the left part of the plane $\{H[K], H[s]\}$, delimited by the dashed line. The full line stands for the theoretical relation, $H[K]$ as a function of $H[s]$. The dots stand for the empirical values for each data set, as the system size increases (from left to right in the plane $\{H[K], H[s]\}$). The right-bottom panel illustrates the results for an SK spin glass of size $N = 25$ near criticality. For the European index set, each point is calculated by averaging over $(8N)$. For the Dow Jones (daily and minute samplings), each point is calculated by averaging over 100 sets of $N$ randomly chosen stocks.

of a truly critical regime should return a situation halfway between both former extreme cases, as illustrated in figure 2 for the SK-model.

After fitting a pairwise maxent model, we record artificial data for the Dow Jones varying the stochasticity by one third smaller and larger than the actual one. The results are illustrated in figure 3. It seems that the Dow Jones (minute sampling) is rather disordered; we will check this in detail hereafter.

4. Results

In the following, we check whether the signatures of criticality are observed in the considered data sets. The variance of the log-likelihood is illustrated in figure 4.

doi:10.1088/1742-5468/2013/11/P11004
Figure 3. The statistical significance of the Dow Jones data set for different levels of stochasticity. The full line stands for the theoretical relation, $H[K]$ as a function of $H[s]$. The dots stand for artificial data generated with a pairwise maxent model fitted to the Dow Jones data with one third larger stochasticity than the actual one. The squares illustrate artificial data with the same level of stochasticity as the actual one and the pentagons illustrate data generated with one third lower stochasticity.

Figure 4. Variance of the log-likelihood for the European index set (left) and for the Dow Jones at minute sampling (right) versus the rescaling parameter. The peak moves from left to right when we consider larger sets. For the European set, we plot the variance for sizes $N = 2, 4, 5, 8$. The dashed curve is a Monte Carlo simulation (appendix D) for $N = 8$. For the Dow Jones, we consider $N = 2, 4, 6, 8, 10, 12$; the last two values are not statistically significant. These curves have been obtained by direct sampling of the probability (and entropy) and by using the relation $T\partial S/\partial T$, see section 2 for details.

We can observe that the peak position scales with the system size, moving from left to right towards the operating point $T = 1$, and that the maximum value of the variance becomes larger when the number of entities increases. For a given and fixed size, one expects a larger value of the critical scaling parameter for sets (of $N$ randomly chosen entities) with larger mean correlation coefficients [30, 26]. We consider 100 sets of $N = 6$ randomly chosen entities for the Dow Jones (daily and minute samplings) and for the S&P 100. The results illustrated in figure 5 suggest a roughly linear relation between the
A statistical physics perspective on criticality in financial markets

Figure 5. The critical scaling parameter ($x$-axis coordinate of the maximum of the response function $R_U(T)$) versus the mean correlation coefficient of the considered set of $N = 6$ randomly chosen entities. The results are illustrated for the Dow Jones at minute (squares, left panel) and daily samplings (triangles, center panel) and for the S&P100 (circles, right panel). The size $N = 6$ is chosen consistently with the later analysis of statistical significance.

Figure 6. The value of the scaling parameter at which the response function $R_U$ reaches its maximum value versus the number of entities $N$. Mean values $\bar{T}$ and error bars (one standard deviation on $\bar{T}$) are computed over $\binom{8}{N}$ samples for the European index set. The full line stands for a power fit and the dashed line stands for an exponential fit on the first seven values. The response function is calculated using the relation (1).

critical scaling parameter $T_{\text{max}}$ and the mean correlation coefficient. Any further results will thus be averaged over several sets for each considered size.

To formalize the relation $T_{\text{max}} = T_{\text{max}}(N)$, we compute the value of the scaling parameter at which the response function $R_U$ reaches its maximum value for different sets of $N$ randomly chosen entities. The results are illustrated in figure 6 for the European index set and in figure 7 for the Dow Jones (daily and minute samplings).

The power and exponential fits return asymptotic critical scaling parameters equal to 1.38 and 0.92 respectively.
Figure 7. The value of the scaling parameter at which the response function $R_U$ reaches its maximum value versus the number of entities $N$. Mean values are computed over 100 sets of $N$ randomly chosen stocks for the Dow Jones at daily (triangles) and minute (squares) samplings.

An exponential fit, on size up to $N = 8$, of the DJ (min) returns an asymptotical critical parameter equal to 0.70, and equal to 0.74 if we fit up to $N = 12$ (but the latter value is not trustworthy since the system is undersampled for $N > 8$). An exponential fit, on size up to $N = 6$, of the DJ (daily) returns an asymptotical critical parameter equal to 0.71, and equal to 0.72 if we fit up to $N = 10$ (but the latter value is not trustworthy since the system is undersampled for $N > 6$). Furthermore, even in the undersampled regime, we observe an increase of the critical scaling parameter.

The larger correlations measured when the size ($N$) increases may be a spurious effect due to the consideration of a particular time interval. One can perform the same study by changing the size and scaling sample length simultaneously and considering different time-windows. For the set of European indices, we choose the sample length $L(N) = 2^{N+3}$ such that $L(8) \approx L_{\text{max}} = 2300$ and we average the results on five different time-windows. The results are illustrated in figure 8. Each point (square) falls into the confidence interval of the constant size results except the last one ($N = 6$). Larger correlations for increasing size is thus a genuine feature.

As no inference method has been used, we would expect that the Kullback–Leibler divergence (KLD) $D_{\text{KL}}(P_{\text{crit}} \parallel P_{\text{emp}})$ between the critical distribution $P_{T=T_{\text{max}}}(s)$ (such that the maximum value of $R_U$ is reached at $T_{\text{max}}$) and the empirical distribution $P_{\text{emp}}(s)$ should be of the same magnitude order as for a truly critical system operating at $T_{\text{crit}} + \Delta T$. The relative deviation $\Delta T/T_{\text{crit}}$ and $(T_{\text{op}} - T_{\text{max}})/T_{\text{max}}$ are equal (by definition $T_{\text{op}} = 1$). Following [25], a reasonable benchmark is the two-dimensional square lattice nearest neighbor Ising model with periodic boundaries of size $N = 9$. The response function $R_U$ reaches its maximum value at $T_{\text{crit}} = 2.40$. We compute the exact distribution $P_{\text{crit}}$ and the KLD with the scaled distribution $P_{\text{scaled}} = P_{\text{crit}}^{1/(1+x)}$ where $x = (T_{\text{crit}} - T)/T_{\text{crit}}$. We found $T_{\text{max}} = 0.88$ and $D_{\text{KL}}(P_{\text{crit}} \parallel P_{\text{emp}}) = 0.070$ for empirical data (European indices). The results for the Ising model are illustrated in figure 9.
Figure 8. The value of the scaling parameter at which the response function \( R_U \) reaches its maximum value versus the number of entities \( N \). Mean values and error bars (one standard deviation) are computed over \( \binom{8}{N} \) samples for the European index set (circles). The squares illustrate results for the same sets with scaled sample length \( L(N) = 2^{N+3} \) and averaged over five different time-windows.

Figure 9. The Kullback–Leibler divergence between the critical and scaled distributions for the two-dimensional square lattice nearest neighbor Ising model with \( N = 9 \) (light gray circles) and for the set of eight European indices (squares).

For both systems, the results are similar. Furthermore, we simulated (see appendix D) artificial binary returns with a Monte Carlo Markov chain \((1 \times 10^4 \) equilibration steps and \(2.3 \times 10^3 \) recorded configurations for \( N = 8 \)) using a pairwise maximum entropy model fitted on the data. We obtained an absolute net orientation \( |\tilde{m}| = 0.812 \pm 0.010 \) (one standard deviation). The empirical value is \( \langle |m| \rangle = 0.726 \), not included in the confidence interval, but near a critical state, a slight change in inferred parameters may lead to a significant change of observables estimated by simulations [31]. To quantify the effect of a
small reconstruction error on the estimated observable, we inferred Lagrange parameters with a regularized pseudo-maximum likelihood (see appendix C for details) and we shifted the parameters slightly such that $\Delta = 0.015$, consistently with [32]. The reconstruction error is $\Delta = \sqrt{N} \langle (J_{ij} - J^{\text{true}}_{ij})^2 \rangle^{1/2}$ and quantifies the ratio between the root mean square error of the reconstruction and a canonical standard deviation. We obtained 10% relative deviation between the two estimations of $|m|$. The empirical and critical values of $|m|$ are thus similar.

The European market places seem to operate near the point corresponding to the maximum of the variance of the log-likelihood while for the Dow Jones, the critical scaling parameter seems to be far away from the operating point $T_{op} = 1$ in the range of considered sizes. In figure 18, we extend this plot for larger sizes by simulating artificial data (see hereafter). This may be explained by larger correlation coefficients between stock exchanges than between the stocks of the Dow Jones, as illustrated in figure 10, and by the Epps effect (decreasing correlation magnitude with decreasing time-scale) [28].

Another observation is that the so-called critical exponent of the variance is equal to zero for each curve illustrated in the left panel of figure 4 in agreement with the mean-field value of the Ising model at the critical temperature. The critical exponent can be obtained by taking the limit $\lim_{\epsilon \to 0} \log R_U(\epsilon) / \log \epsilon$ where $\epsilon = (T - T_{\text{max}})/T_{\text{max}}$ and $T_{\text{max}}$ is such that $R_U(T)$ reaches its maximum at this point [33].

We also study the distribution of the configuration rank. In order to know whether we should reject or not Zipf’s law, we perform a modified version (discrete power-law with a natural upper bound due to the finite number of configurations) of the statistical test described in [34], see appendix B for details. If the $p$-value is smaller than 0.05, the power-law hypothesis is ruled out, and for a $p$-value close to one, we can consider it as a good distribution candidate (without guarantee that it is the correct distribution). The empirical rank distribution is illustrated in figure 11.

Test results for different sets are reported in table 1. The considered size should not exceed $N = 8$ for empirical data to have a good estimate of the distribution $P[s]$ by direct sampling. As expected, the power-law test outcomes depend on the system size. For the Dow Jones, the power-law is rejected when the system is properly sampled, whereas in the undersampling regime the power-law is not rejected. As detailed in [29], the power-law is the most informative distribution when the distribution $P[s]$ is undersampled.
Figure 11. From left to right: empirical relative frequencies of configurations versus the configuration rank of the observed time series for a set of seven randomly chosen stocks of the Dow Jones (min), artificial rank distribution for a real power-law and empirical relative frequencies for a set of 12 randomly chosen stocks of the Dow Jones. The fit (dashed line) is obtained with the maximum likelihood estimator.

Table 1. Statistical test of the power-law hypothesis for sets of $N$ randomly chosen stocks of the Dow Jones ($3 \times 10^4$ points at minute sampling). We report the maximum likelihood estimator $\hat{\alpha}$ of the power-law exponent $\alpha$ and its standard deviation $\sigma_\alpha$, the Kolmogorov–Smirnov statistic ($D$) and the $p$-value. One does not reject the power-law hypothesis if the $p$-value is larger than 0.05.

| # of stocks | $\hat{\alpha}$ | $\sigma_\alpha$ | $D$  | $p$-val |
|-------------|-----------------|-----------------|------|---------|
| 6           | 0.0119          | 0.00            |      |         |
| 7           | 0.0117          | 0.00            |      |         |
| 8           | 0.0194          | 0.00            |      |         |
| 9           | 0.6654          | 0.0038          | 0.0147| 0.10    |
| 10          | 0.6584          | 0.0035          | 0.0164| 0.36    |
| 11          | 0.7192          | 0.0027          | 0.0210| 0.87    |
| 12          | 0.7441          | 0.0025          | 0.0292| 0.96    |
| 13          | 0.7699          | 0.0024          | 0.0290| 0.98    |

The maximum likelihood estimator (MLE) of the exponent is derived by maximization of the log-likelihood

$$\ln L(\alpha) = -\alpha \sum_{i=1}^{N} \ln x_i - N \ln \left( \sum_{x=1}^{x_{\text{max}}} x^{-\alpha} \right)$$

where $x_{\text{max}}$ is the upper bound. The standard deviation of this MLE is obtained by taking the expansion of the likelihood up to second order (Gaussian approximation). It reads

$$\sigma_{\alpha,\text{MLE}} = \frac{1}{\sqrt{N \left[ \zeta''(x_{\text{max}}, \alpha_{\text{MLE}})/\zeta(x_{\text{max}}, \alpha_{\text{MLE}}) - ( \zeta'(x_{\text{max}}, \alpha_{\text{MLE}})/\zeta(x_{\text{max}}, \alpha_{\text{MLE}}) )^2 \right]}}$$

where $\zeta(x_{\text{max}}, \alpha) = \sum_{x=1}^{x_{\text{max}}} x^{-\alpha}$ and the prime stands for the derivative with respect to $\alpha$.

The empirical probability density function (pdf) of this estimator for $N = 13$ and $10^4$ tests and its Gaussian approximation are illustrated in figure 12.

doi:10.1088/1742-5468/2013/11/P11004
As a complement to the later analyses, we study the linearity of the entropy expressed as a function of the utility. Zipf’s law induces a linear relation between the entropy and the log-likelihood [24]. Strict linearity can be achieved at a single value of the utility (as for the 2D nearest neighbor Ising model) or for any value of the entropy if the distribution of the rank is a power-law [24]. The expansion of the entropy around the mean utility $U$ is written (where $U$ is the notation for $\langle U \rangle$)

$$S(U) \simeq S(U) - \frac{1}{T}(U - U) + \frac{1}{2T^2 R_U} (U - U)^2.$$  \hspace{1cm} (6)

For ranks distributed following a power-law, the quadratic and higher order terms are sub-intensive; the entropy should be a linear function of the utility [25].

We check this property for several sets of seven randomly chosen stocks of the Dow Jones Index. We compute the average entropy–utility relation $S(-U)$ for 100 sets of seven randomly chosen stocks; the results are illustrated in figure 13.

We measured the relative non-linearity [35]; the typical value is 0.053 (equal to zero if the function is exactly linear). The typical value of the slope is 0.71. We also simulate $5 \times 10^5$ artificial returns with multivariate GARCH(2, 2) and pairwise maxent processes fitted on the data. The entropy dependence on the log-likelihood is illustrated in figure 14. The relative non-linearities are 0.032 and 0.035, and the slopes are equal to 0.77 and 0.59 respectively. For larger sample sizes the entropy is not linear either; however, in a restricted utility range ([0.3, 0.4], about 10% of the possible values of the utility, for instance) the entropy is almost linear (as measured by the relative non-linearity).

As suggested by Zipf’s law checking, the entropy is not a linear function of the log-likelihood. However, we cannot reject the possibility of linearity in a restricted range or zero curvature in a single point as for the 2D nearest neighbor Ising model.

Last, as the returns are believed to be non-stationary with volatility clustering (often modeled by a GARCH process), we study the evolution of the critical rescaling parameter $T_{\text{max}}$ (at which the variance of the log-likelihood reaches its maximum value). As expected, for fixed size, $T_{\text{max}}$ increases during the growth period and reaches its maximal value in...
Figure 13. Left: Shannon entropy versus the opposite of the log-likelihood for several sets of seven stocks (randomly chosen) of the Dow Jones Index (min). Right: the average entropy–utility relation $S(-U)$ for 100 sets of seven randomly chosen stocks. The dashed lines are the best linear fits with slope equal to 0.71 and 0.68 respectively.

Figure 14. Shannon entropy versus minus log-likelihood for 100 sets of nine randomly chosen stocks. Artificial returns are simulated with a multivariate GARCH(2, 2) process (light line) and with a pairwise maxent model (bold line). The dashed lines are linear fits on a restricted range.

5. Link to maximum entropy models

In the following, we use an inference procedure to check whether the existence of a critical state is supported. One can show that the pairwise maximum entropy model (maxent) is a consistent statistical model when the aim is to study collective behaviors rather than to give a precise model of the market [26, 4]. Rather than making specific assumptions on the underlying dynamics, we build a model which is consistent with the recorded data in the vicinity of the crash beginning (when the fluctuations are the largest), as illustrated in figure 15, and gets closer to $T_{op}$. 

doi:10.1088/1742-5468/2013/11/P11004
and the observed structure. This maxent model is directly linked to the former discussion since spin glasses and neural networks are also represented by pairwise maxent models which actually exhibit critical states. In this framework the configuration distribution \(P[s]\) is rewritten as a Gibbs distribution

\[
p_2(s) = Z^{-1} \exp \left( \frac{1}{2} \sum_{i,j} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i \right) \equiv \frac{e^{U(s)}}{Z}
\]

where \(J_{ij}\) and \(h_i\) are Lagrange multipliers (chosen to retrieve the first and second empirical moments). They can be thought of as a measure of the pairwise mutual and individual influences. Another well known application of the pairwise maxent model is the characterization of a neural network structure [8] where the operating point seems to be a critical one [24, 10]. One can show that this model is able to generate correlation matrices with non-Gaussian eigenvalues [26], as observed in real financial time series [36], but also scale-free asset trees, order–disorder periods [4]. This pairwise model gives more insights into the possibility of a critical operating point. The rescaling of the Gibbs distribution is then viewed as a rescaling of all the parameters by a common factor \(T^{-1}\). This rescaling is an investigation of a slice of the parameter space which corresponds to a stochasticity variation. A small value of \(T\) favors co-movements and a large value favors randomness. In this work, Lagrange multipliers are estimated with a regularized pseudo-maximum likelihood [32], see appendix C for a short description. We note that close to \(T = 1\), many models are distinguishable and a slight change in parameters may lead to a significant change of the measured observables. One should compare artificial and empirical results.

First, we simulate artificial data with the estimated Lagrange parameters from the real time series. The Monte Carlo Markov chain (MCMC) is defined as follows. A randomly chosen orientation is flipped if the conditional flipping probability \(p(s_{i,t} = -s_{i,t-1}|s_{-i,t})\) is larger than a realization of a uniform law on the interval \([0,1]\), where \(s_{-i,t}\) is the configuration excluding the \(i\)th entity. A configuration is recorded each \(N\) flipping

Figure 15. The critical rescaling parameter \(T_{\text{max}}\) for six European indices (black curve, left ordinate) and the normalized sum of indices (light blue curve, right ordinate). The critical rescaling parameter is empirically estimated on a sliding window of \(2^{N+2}\) trading days translated by one trading day each step.
Figure 16. First row: the empirical probability density function (pdf) is illustrated for several data sets. Second row: comparison of the empirical probability mass function (pmf) of the net orientation to the artificial distributions resulting from simulations. Third row: ten values of the stochasticity level $T$ (in the range $[0.8, 2]$, blue to red respectively) are used to check whether the pdf can go continuously from unimodal to bimodal; the results are compared to a 2D nearest neighbor Ising model without individual biases. The pdf and the pmf are estimated on $5 \times 10^5$ Monte Carlo steps.

attempts, which defines a Monte Carlo step (MCS) [37]. See appendix D for details. The result of the procedure applied to these artificial data is illustrated in figure 4 by the dashed curve ($1 \times 10^4$ equilibration MCSs and $1 \times 10^5$ recorded MCSs). This is consistent with the empirical variance; both peaks (blue and dashed curves) are located at the same value of the $T$-parameter. If the Lagrange parameters $\{J_{ij}\}$ are positive, the orientation distribution should be unimodal for large values of $T$ and bimodal for small values of $T$. As a qualitative test, we check whether the empirical distributions are unimodal or bimodal and whether they can become bimodal if we change the stochasticity level $T$; an order–disorder transition is then possible. As illustrated in the first row of figure 16, the empirical distribution of the index set is bimodal whereas the distributions of stock sets are unimodal as expected from the former empirical analyses. The second row of figure 16 illustrates the difference between the empirical orientation distribution and the simulated ones $P_m(T)$ at different stochasticity levels. The index set is a rather ordered system; the
A statistical physics perspective on criticality in financial markets

Figure 17. Variances of the overlap parameter (dashed lines) and the log-likelihood (full lines) for the eight index set, the Dow Jones (daily and minute samplings) and the SP100. Each point is computed over $5 \times 10^5$ MCSs after an equilibration period of $5 \times 10^4$ MCSs. The coordinate of the maximum is pinned (the coordinate on the left stands for the variance of the log-likelihood).

probability mass peaks at the extreme values $-1, 1$ of the net orientation. A disordered state exists for a high level of stochasticity ($T = 2$). The third row of figure 16 illustrates the continuous deformation of the probability density function for a stochasticity varying from a low level (blue) to a high level (red). This deformation is compared to that of the 2D nearest neighbor Ising model of the corresponding size without individual biases.

The fitted maxent models allow an order–disorder transition which justifies their use in the criticality check. As mentioned in [31] such models are prone to accumulate in the vicinity of the critical point $T = 1$ but are also highly distinguishable in this neighborhood. Accordingly, we check whether they return a $T_{\text{max}}$ in line with the empirical results. One can estimate the variance $R_Q$ of the overlap parameter $q = N^{-1} \sum_i s_i^{(1)} s_i^{(2)}$ and the variance of the log-likelihood. The overlap parameter measures the correlation between the configurations of two identical systems denoted by the superscripts (1) and (2). The variances $R_U$ and $R_Q$ are known to peak at the critical value of the rescaling parameter [30]. If the operating point is indeed critical, we should find the peak near the value $T = 1$. The results are illustrated in figure 17.

doi:10.1088/1742-5468/2013/11/P11004
We note that the peaks are indeed located near the empirical values. For the index set, the relative difference between the empirical and simulated $T_{\text{max}}$ is equal to 2%, slightly underestimated. For the Dow Jones (min), the relative difference is equal to 6%, slightly overestimated, and for the Dow Jones (daily), $T_{\text{max}}$ is overestimated by 14%. The first two fitted models are consistent with the data and lead to the same conclusion: the index set is close to criticality ($1 - T_{\text{max}} \leq 10\%$) and the Dow Jones is far from criticality ($1 - T_{\text{max}} \geq 25\%$). The larger deviation between empirical and simulated values for the Dow Jones (daily) may be due to inference errors in the Lagrange parameter estimation. The ratio $M/N$ (sample length over the number of entities) is too small, ten times smaller than for the Dow Jones (min). Consequently, one may expect the same relative error for the critical scaling parameter of the SP100 index.

Since the simulations are consistent with the empirical results, we simulate data to complete figure 6 for sizes larger than $N = 8$. We simulate a binary sample of length $5 \times 10^6$ with the previous MCMC and also artificial returns with a multivariate GARCH process, known to capture clustering volatility and have a fatter tail than the Gaussian one. We obtain consistent results with the empirical ones. The critical value of the rescaling parameter $T_{\text{crit}}$ is illustrated in figure 18. The critical value increases with size but is still far from $T = 1$.

We note that in 2010, 12,807 companies (excluding investment funds) were listed in stock exchanges (see www.world-exchanges.org/). There is thus no obvious reason to consider the limit $N \to \infty$. The market place system is significantly closer to criticality despite its small size. This may due to information aggregation of an index about the subjacent stocks [38]. A set of indices may operate as a system of larger size.

Furthermore, one can show that the financial network exhibits small-world organization [6], and one knows that the Ising model on a complex network, among others, is a small-world one only at the critical temperature [10].

Figure 18. The value of the $T$-parameter at which the response function $R_U$ reaches its maximum value versus the number of entities. The squares illustrate the real data, the pentagons stand for a multivariate GARCH(2, 2) process and the dots for the MCMC.
6. Discussion

Stock markets are embedded in a non-uniform background. They should therefore be heterogeneous and go through regular periods interspersed with surprising events. In a complex economic background, reactiveness is an expected behavior. In the case of the Fukushima nuclear accident or the 2008 subprime crisis for instance, the market response was clear and prompt. All stocks fell quickly in an organized fashion. This behavior can help to secure the profit made or prevent excessive losses if the situation gets even worse. Then, when the situation seems stabilized, or stock prices have fallen so dramatically that stocks become cheap and attractive, the market goes up again in an ordered fashion. These large bearish–bullish movements of stock prices are encountered at any time-scale [39]. During such phases, the market exhibits large correlated structures and an ordered state [40, 27, 4] corresponding to an increase of the correlation strength. Such dramatic events impact the market globally (all economic sectors). On the other hand, some events (like the end of a state subsidy for eco-friendly goods, nuclear energy, etc) have an impact on a single or few economic sectors. Criticality is then thought of as a competition between global effects inducing homogeneity and local effects inducing heterogeneity in trades.

We have seen that the Shannon entropy has an inflexion point near the operating point $T = 1$ for the European index set. We deduce that the number of micro-states increases (or decreases) drastically following a variation of the stochasticity. The entropy is related to the logarithm of the averaged number of micro-states and we can obtain this quantity by a simple integration of $R_u(T)/T$. We observe that the largest slope is approximately at the actual operating point $T = 1$ far from the saturation zones (where the slope is close to zero). In the neighborhood of the operating point, the logarithm of the number of micro-states is almost linear with a large slope, thus a variation of Lagrange parameters will induce a drastic (in an exponential fashion) change in the micro-structure. This shows that the market network has a great structural malleability. The entropy also measures the degree of statistical dependence between stocks. If stocks did not influence each other, the system would be considered as a random one which implies small covariances and low reactiveness. Thus the entropy would reach its largest value. In the opposite case, if stock correlations are maximal (implying again low reactiveness), there would not be any uncertainty anymore, the whole market state $s$ would be predictable on knowing an individual state $s_i$ and the entropy would be zero. Therefore, if the slope of the entropy reaches its maximum value at the operating point, this means that the market is on the edge. Any variation can tip the market towards either a random (disordered, with independent trades) or a highly interactive (ordered, synchronized trades) state. We expect thus a large predictability by exploiting instantaneous information: using the system configuration with the $i$th entity $s_{-i}$ removed, one should be able to predict the state of this entity $s_i$ with high accuracy. This will be the subject of another article.

Last, the fact that the European index set is closer to criticality than the Dow Jones may follow from information aggregation [38]. A set of indices is a weighted average of stock prices. Considering the stocks as the fundamental hubs of the financial network, the indices represent super-hubs acting as systems of significantly larger size. The typical relative cluster size is also larger in the index set where each cluster contains roughly
A statistical physics perspective on criticality in financial markets

Figure 19. Illustration of the clusters of each data set. The clusters of the index set (left) return a partition of the European economy. The clustering of the Dow Jones (right) returns the different economic sectors (technologies, distribution, aircraft industry, TV broadcasting, finance, chemical and industrial companies, telecoms, consumer goods, health care, oil).

30% of the total number of entities, as illustrated in figure 19 [4]. For the Dow Jones, the cluster size is about 10% of the index size. Correlated structures have thus a larger relative size in the index set which may give the right balance between co-movements and fluctuations.

From the data analysis and simulations, we saw that the European market places seem to operate at a point where the variance of the log-likelihood is close to its largest value. An exponential empirical fit returns $T_{op} = 0.92$ as an asymptotical value (thus maximum) for European indices and $T_{op} = 0.70$ for the Dow Jones at minute sampling. The entropy is not a linear function of the log-likelihood. The estimation of $T_{op}$ with simulated data returns a value close to one but this value is suspected of being overestimated by about 15%. For the Dow Jones, large simulated samples $M = 5 \times 10^6$ (using parameters obtained by fitting real data) return a consistent value $T_{op} \approx 0.65$. Moreover, financial systems are closer to criticality close to a crash beginning meaning large fluctuation and large deviation from the uniform distribution of the configurations. This evolution also suggests a process of self-organization. The market is a highly adaptive system. By self-organization, the market reacts strongly to a change or unexpected events and by itself does not consider all possible events as equiprobable. However, through the data analysis, the stock exchange system is not exactly critical and the Dow Jones seems to be far from criticality. Furthermore, financial systems do not stay in the same regime, and get closer to criticality just before a crisis. This is an interesting finding since in such models, large avalanches occur more probably close to criticality [41].

doi:10.1088/1742-5468/2013/11/P11004
Acknowledgments

I would like to thank B De Rock and P Emplit for their helpful comments and discussions. This work was undertaken with financial support from the Solvay Brussels School of Economics and Management.

Appendix A. Practical recipe

(i) Binarize the returns.
(ii) Test the statistical significance and determine the corresponding maximum size $N$.
   (a) Compute the empirical distribution of configurations $\hat{p}_s$.
   (b) Compute $m_k$ (the number of configurations sampled exactly $k$ times) and the empirical distribution of $K_i$ (the number of times the configuration $s_i$ is observed in the sample).
   (c) Deduce their entropies $H_s$ and $H_K$.
   (d) Locate the maximum of the relation $H_s$ versus $H_K$.
(iii) Get the response function $R_U$ and find its maximum. Repeat for several ($\sim 100$) sets of $N$ randomly chosen entities. For each set:
   (a) Compute the empirical distribution of configurations $P(s)$.
   (b) Rescale the empirical distribution as $P_T(s) = P(s)^{1/T}/\sum \{s\} P(s)^{1/T}$.
   (c) Compute $R_U = T(\partial S/\partial T)$ where $S(T) = -\sum \{s\} P_T(s) \log P_T(s)$.
   (d) Store the coordinates of its maximum.
(iv) Compare to a finite size version of a truly critical system.
   (a) Compute the relative difference $x = (T_{\text{op}} - T_{\text{max}})/T_{\text{op}}$ where $T_{\text{op}} = 1$.
   (b) Compute the Kullback–Leibler divergence (KLD) between $P_{T=T_{\text{max}}}(s)$ and $P_{\text{emp}}(s)$.
   (c) Compare the latter KLD value to the KLD between $P_{T=T_{\text{crit}}}(s)$ and $P_{T=(1+x)T_{\text{crit}}}(s)$ for the 2D nearest neighbor Ising model.
(v) Perform a statistical test of Zipf’s law as described in appendix B.
(vi) Check the linearity of the relation $S(U)$ versus $U$ where $U = \log P(s)$.
(vii) Compare the empirical results to simulations.
   (a) Infer the Lagrange parameters (see appendix C).
   (b) Simulate data using a Monte Carlo Markov chain (see appendix D).
   (c) Check whether an order–disorder transition is allowed by computing the orientation distribution and by varying the scaling parameter $T$.
   (d) Compute the variance of the log-likelihood and of the overlap parameter $q = N^{-1}\sum_i s_1^{(1)} s_2^{(2)}$ (two copies denoted by the superscripts, linked with the covariance of the utility function $U$). Compare empirical and simulated results for a common size. A large difference ($>10\%$) between the simulated value of $T_{\text{max}}$ and the asymptotical value returned by fitting the empirical relation $T_{\text{max}}(N)$ may unveil difficulties in the inference of Lagrange parameters [31] and therefore a poor fitting.

doi:10.1088/1742-5468/2013/11/P11004
Appendix B. Discrete power-law

A statistical test for a power-law is given in [34]. We adapt this test to a discrete power-law with a natural upper bound. Before considering the discrete case, we note that if the distribution \( p(x) \sim x^{-\beta} \) has a finite upper bound \( x_{\text{max}} \), then the cumulative distribution function (CDF) will not be a straight line in a log–log plot because

\[
P[X \geq x] = \text{Cst} \int_x^{x_{\text{max}}} y^{-\beta} \, dy = \frac{\text{Cst}}{1 - \beta} [x_{\text{max}}^{-1 - \beta} - x^{-1 - \beta}] \tag{B.1}
\]

where the constant normalizes the distribution to 1 and \( \beta > 1 \). Taking the logarithm of both sides, it becomes

\[
\log P[X \geq x] = \log (x_{\text{max}}^{1 + \beta} - x^{1 + \beta}) + \log \frac{\text{Cst}}{\beta - 1}. \tag{B.2}
\]

The statistical test proposed in [34] consists of the following scheme.

(i) Determine the best fit of the power-law to the data using a maximum likelihood estimator.

(ii) Calculate the Kolmogorov–Smirnov (KS) statistics for the goodness-of-fit. The KS statistics is the maximum absolute value between the empirical CDF and the CDF of the estimated power-law.

(iii) Generate a large number (~1000) of synthetic data sets.

(iv) Calculate the \( p \)-value as the fraction of the KS statistics for the synthetic data sets whose values exceed the KS statistics of the real data.

(v) If the \( p \)-value is sufficiently small (~0.05), the power-law is ruled out.

The MLE of a discrete power-law with a natural cut-off \( x_{\text{max}} \) is derived from the first order condition for the log-likelihood based on \( N \) observations

\[
L(\beta) = \ln L(\beta) = -\beta \sum_{i=1}^{N} \ln x_i - N \ln \left( \sum_{x=1}^{x_{\text{max}}} x^{-\beta} \right). \tag{B.3}
\]

Taking the derivative with respect to \( \beta \) leads to the MLE \( \beta_{\text{MLE}} \) satisfying

\[
\frac{1}{N} \sum_{i=1}^{N} \ln x_i = \frac{\sum_{x=1}^{x_{\text{max}}} x^{-\beta_{\text{MLE}}} \ln x_{\text{max}}}{\sum_{x=1}^{x_{\text{max}}} x^{-\beta_{\text{MLE}}}}. \tag{B.4}
\]

The standard deviation of \( \beta_{\text{MLE}} \) is obtained by taking the expansion of the likelihood around \( \beta_{\text{MLE}} \)

\[
\mathcal{L}(\beta) = \mathcal{L}(\beta_{\text{MLE}}) + \frac{1}{2!} \frac{\partial^2 \mathcal{L}(\beta)}{\partial \beta^2} \bigg|_{\beta_{\text{MLE}}} (\beta - \beta_{\text{MLE}})^2. \tag{B.5}
\]

Identifying the terms to the Gaussian approximation \(-\ln(\sigma \sqrt{2\pi}) - \frac{1}{2} ((x - \beta)/\sigma)^2\), we have

\[
\sigma_{\beta_{\text{MLE}}} = \frac{1}{\sqrt{N \left[ \zeta''(x_{\text{max}}, \beta_{\text{MLE}})/\zeta(x_{\text{max}}, \beta_{\text{MLE}}) - (\zeta'(x_{\text{max}}, \beta_{\text{MLE}})/\zeta(x_{\text{max}}, \beta_{\text{MLE}}))^2 \right]}} \tag{B.6}
\]

where \( \zeta(x_{\text{max}}, \beta) = \sum_{x=1}^{x_{\text{max}}} x^{-\beta} \) and the prime stands for the derivative with respect to \( \beta \).
Synthetic data distributed as a discrete power-law with a finite upper bound are generated as follows. One generates a realization $u$ of a uniform random variable $U$ in $[0, 1]$, and one calculates $\sum_{x=1}^{x_{\text{max}}} x^{-\beta}$ and the cumulative sum $\sum_{y=1}^{y_{\text{max}}} y^{-\beta}$. The smallest integer $x$ such that $\sum_{y=1}^{y_{\text{max}}} y^{-\beta} \geq u \sum_{x=1}^{x_{\text{max}}} x^{-\beta}$ is stored. This process is repeated to generate a sample of desired length.

**Appendix C. Regularized pseudo-maximum likelihood**

The rPML method is a powerful method for estimation of the Lagrange parameters of the pairwise maximum entropy model when common maximum likelihood is untractable [32]. This method can be thought of as an autologistic regression in order to predict binary outcomes. The main idea is to factorize the distribution and to consider only conditional probabilities. For an N-dimensional sample of length $M$, the objective function to maximize is

$$\text{PL}(\theta) = \frac{1}{M} \sum_{t=1}^{M} \sum_{i=1}^{N} \log P(s_{i,t} | s_{-i,t}; \theta)$$  \hspace{1cm} (C.1)

where $s_{-i}$ stands for the configuration excluding the $i$th entity and the conditional probabilities of the pairwise maximum entropy model are

$$p(s_{i,t} | s_{-i,t}; \theta) = \frac{1}{2} \left[ 1 + s_{i,t} \tanh \left( \sum_{j \neq i} J_{ij} s_{j,t} + h_i \right) \right].$$  \hspace{1cm} (C.2)

A regularization term is added to the PL function to prevent overfitting which is a negative multiple of the $l_2$-norm of the parameters to be estimated, for instance. The regularized PL (rPL) objective function is thus $\text{PL}(\theta) - \lambda \| \theta \|_2$ with $\lambda > 0$.

**Appendix D. Monte Carlo Markov chain simulations**

To perform the simulation, we should describe how the Gibbs distribution can be reached as the equilibrium distribution of a given Markov process. A way to reach a Gibbs distribution

$$p_2(s) = Z^{-1} \exp \left( \frac{1}{2} \sum_{i,j}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i \right) \equiv \frac{e^{-H(s)}}{Z}$$  \hspace{1cm} (D.1)

is given by the following dynamics (the so-called Glauber dynamics [37]). Namely, one takes an entity $i$ chosen randomly and the attempt to flip the associated binary variable $s_i$ is performed with a rate depending on an exponential weight, the other orientations remaining fixed. We define the reversal operator $F_i$ such that $F_i(s_1, \ldots, s_i, \ldots, s_N) = (s_1, \ldots, -s_i, \ldots, s_N)$. This asynchronous updating involves that two consecutive configurations only differ by a single reversal. To find the exponential rate, we consider the evolution of the probability mass function (pmf) for this dynamics.
which is given by the master equation
\[
\frac{d}{dt} p(s; t) = \sum_{i=1}^{N} \left\{ \omega(s_i | -s_i) p(F_i s; t) - \omega(-s_i | s_i) p(s; t) \right\}
\] (D.2)

where \(\omega(s_i | -s_i)\) is the transition rate from configuration \(F_i s\) to configuration \(s\). It is derived from the transition probability \(P[s_{i,t+\tau} = -s_{i,t} | s_{i,t}, s_{-i,t}] \equiv W(-s_i | s_i) = \omega(-s_i | s_i) \tau + o(\tau)\).

The master equation states that the variation of the pmf is equal to the inward probability flow minus the outgoing probability flow [42]. At equilibrium, this dynamics should lead to the Gibbs distribution (D.1). A sufficient condition to reach equilibrium is
\[
\omega(s_i | -s_i)p_2(F_i s) - \omega(-s_i | s_i)p_2(s) = 0
\] (D.3)

As we are only interested in the equilibrium PMF and not in how one reaches it, we can choose any transition rates satisfying (D.3). A convenient choice for simulation (discrete time) is to take the transition probability
\[
W(-s_i | s_i) = \frac{1}{2} \left[ 1 - s_i \tanh \left( \sum_j J_{ij} s_j + h_i \right) \right].
\] (D.4)

Simulations are performed using the following scheme.

Algorithm
(i) Choose an entity uniformly at random.
(ii) Compute the transition probability (D.4).
(iii) Generate a uniform random number \(x \in [0, 1]\); if \(W(-s_i | s_i) > x\), accept the reversal.
(iv) Parameterize time such that a Monte Carlo step (MCS) corresponds to \(N\) reversal attempts.
(v) Wait for equilibration.
(vi) Store the desired statistics.

A more detailed discussion (equilibration time, proper definition of statistics, etc) can be found in [43].

References

[1] Cont R, Empirical properties of asset returns: stylized facts and statistical issues, 2001 Quant. Finance 1 223
[2] Mantegna R N, Hierarchical structure in financial markets, 1999 Eur. Phys. J. B 11 193
[3] Onnela J P, Chakraborti A, Kaski K, Kertész J and Kanto A, Dynamics of market correlations: taxonomy and portfolio analysis, 2003 Phys. Rev. E 68 056110
[4] Bury T, Market structure explained by pairwise interactions, 2013 Physica A 392 1375
[5] Gabaix X, Gopikrishnan P, Plerou V and Stanley H E, A theory of power-law distributions in financial market fluctuations, 2003 Nature 423 267
[6] Vértes P E, Nicol R M, Chapman S, Watkins N and Bullmore E T, Frontiers: topological isomorphisms of human brain and financial market networks, 2011 Front. Syst. Neurosci. 5 75
[7] Huang K, 1963 Statistical Mechanics (New York: Wiley)
[8] Schneidman E, Berry M J, Segev R and Bialek W, Weak pairwise correlations imply strongly correlated network states in a neural population, 2006 Nature 440 1007
[9] Sornette D, Critical market crashes, 2003 Phys. Rep. 378 1
[10] Fraiman D, Balenzuela P, Foss J and Chialvo D R, Ising-like dynamics in large-scale functional brain networks, 2009 Phys. Rev. E 79 061922

doi:10.1088/1742-5468/2013/11/P11004
A statistical physics perspective on criticality in financial markets

[11] Eguiluz V M, Chialvo D R, Cecchi G A, Baliki M and Apkarian A V, Scale-free brain functional networks, 2005 Phys. Rev. Lett. 94 18102

[12] Tagliazucchi E, Balenzuela P, Fraiman D and Chialvo D R, Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis, 2012 Front. Physiol. 3 15

[13] Sornette D, Johansen A and Bouchaud J P, Stock market crashes, precursors and replicas, 1996 J. Physique I 6 167

[14] Vandewalle N, Boveroux P, Minguet A and Ausloos M, The crash of october 1987 seen as a phase transition: amplitude and universality, 1998 Physica A 255 201

[15] Newman M E J, Power laws, pareto distributions and Zipf's law, 2005 Contem. Phys. 46 323

[16] Savit R, Manuca R and Riolo R, Adaptive competition, market efficiency, and phase transitions, 1999 Phys. Rev. Lett. 82 2203

[17] Marsili M, Mastromatteo I and Roudi Y, On sampling and modeling complex systems, 2013 J. Stat. Mech. P01003

[18] Clauset A, Shalizi C R and Newman M E J, Power-law distributions in empirical data, 2009 SIAM Rev. 51 661

[19] Dal'Maso Peron T K and Rodrigues F A, Collective behavior in financial markets, 2011 Europhys. Lett. 96 48004