



Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Financial markets and the phase transition between water and steam

Christof Schmidhuber

Zurich University of Applied Sciences, School of Engineering, Technikumstrasse 9, CH-8401 Winterthur, Switzerland

ARTICLE INFO

Article history:

Received 5 August 2021

Received in revised form 10 December 2021

Available online 10 January 2022

Keywords:

Financial markets

Trends

Renormalization group

Critical dynamics

Fractal dimension

Hurst exponents

ABSTRACT

A lattice gas model of financial markets is presented that has the potential to explain previous empirical observations of the interplay of trends and reversion in the markets, including a Langevin equation for the time evolution of trends. In this model, the shares of an asset correspond to gas molecules that are distributed across a hidden social network of investors. Neighbors in the network tend to align their positions due to herding behavior, corresponding to an attractive force of the gas molecules.

This model is equivalent to the Ising model on this network, with the magnetization in the role of the deviation of the market price of an asset from its long-term value. Moreover, in efficient markets the system should naturally drive itself to its critical temperature, where it undergoes a second-order phase transition. There, it is characterized by long-range correlations and universal critical exponents, in analogy with the phase transition between water and steam.

Applying scalar field theory and the renormalization group, we show that these critical exponents imply predictions for the auto-correlations of financial market returns and for Hurst exponents. For a network topology of R^D , consistency with observation implies a fractal dimension of the network of $D \approx 3$, and a correlation time of at least the length of the economic cycle. However, while this simplest model agrees well with market data on very long and on short time scales, it does not explain the observed market trends on intermediate time horizons from one month to one year.

In a next step, the approach should therefore be extended to other models of critical dynamics, to general network topologies, and to the neighbourhood of the critical point. It allows us to indirectly measure universal properties of the hidden social network of investors from the empirically observable scaling behaviour of financial markets.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction and summary

Trends in financial markets have long been exploited by the tactical trading industry [1], and they have also been well-studied in the literature (see, e.g., [2–10]). Traditional trend-followers are “long” a given market when its current trend is positive, and “short” when the trend is negative. Thus, they suffer draw-downs when trends revert.

In [11], we have carefully measured the interplay of trends and reversion in financial markets with high precision by aggregating in new ways across 30 years of daily futures returns for equity indices, interest rates, currencies and commodities, and across trends over time horizons T ranging from 2 days to 4 years. The key results are shown in Fig. 1. A brief summary of how they were derived is given in the appendix.

E-mail address: christof@schmidhuber.ch.

<https://doi.org/10.1016/j.physa.2022.126873>

0378-4371/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

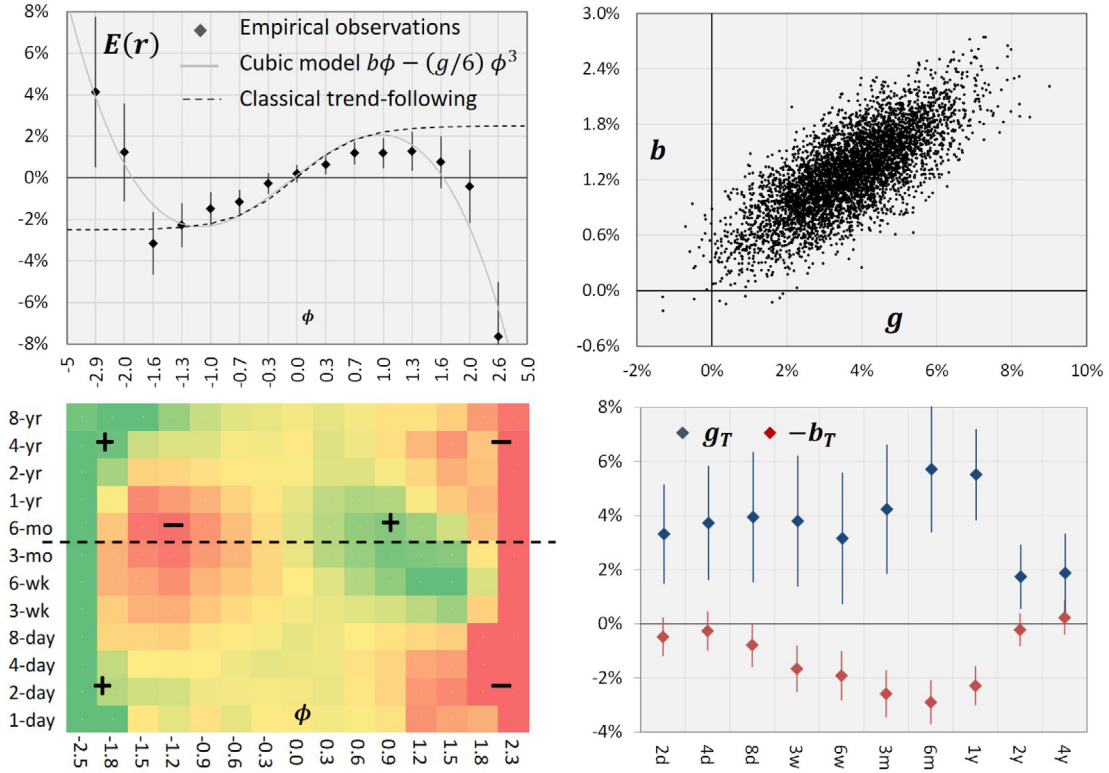


Fig. 1. Upper left (a): The expectation value $E(r)$ of the next day's return of a futures market can be modeled by a cubic polynomial of the trend strength ϕ , whose linear term $b\phi$ represents trend-persistence, and whose cubic term $-(g/6)\phi^3$ represents trend-reversion. Upper right (b): As confirmed by bootstrapping, the regression coefficients b and g corresponding to the linear and cubic terms are statistically highly significant. Lower left (c): A heat map shows how the expectation value of tomorrow's return depends both on today's trend strength ϕ and its time horizon T . (a) can be thought of as a cross-section of (c) along the dashed line. Lower right (d): The coefficients b and g as a function of the time horizon T , corresponding to the dashed line in (c) as it moves from 2 days to 4 years.

We have found that trends tend to revert before they become statistically strongly significant. In other words, by the time a trend has become so obvious that everybody can see it in a price chart, it is already over. This is consistent with the hypothesis that any obvious market inefficiencies are quickly eliminated by investors. More precisely, we have defined the strength ϕ_T of a trend (32) measured over a time horizon T as its statistical significance. We were able to accurately model tomorrow's return $R(t + 1)$ in a market (normalized to have variance 1) by a quartic potential $V_T(\phi_T)$ of today's trend strength $\phi_T(t)$ in that market:

$$R(t + 1) = a - \frac{\partial}{\partial \phi_T} V_T(\phi_T(t)) + \epsilon(t) \quad \text{with} \quad V_T(\phi_T) = -\frac{b_T}{2} \cdot \phi_T^2 + \frac{g_T}{24} \cdot \phi_T^4, \tag{1}$$

where a is a risk premium, ϵ represents random noise, and b_T and g_T are kinetic coefficients (g_T is related to c_T of [11] by $g_T = -6c_T$). From this, we have derived a Langevin equation for the time evolution of the trend strength, which describes a damped motion in the potential $V_T(\phi)$ (see Fig. 2). We have interpreted b_T as the persistence of trends, and g_T as the strength of trend reversion. Within the limits of statistical significance, we have found the following (I thank J.-P. Bouchaud for pointing out that related observations were made in [12]):

- The parameters b_T, g_T are universal in the sense that they are the same for all assets
- While g_T has been fairly stable over time, b_T appears to have vanished over the decades
- As a function of the time scale T , g_T is approximately constant
- b_T depends on the time scale: it peaks at $T \sim 3\text{--}12$ months and decays for longer or shorter horizons. It appears to become negative for $T < 1$ week or $T >$ several years.

These results are reminiscent of physical systems at critical points. Stochastic processes with a quartic potential such as in (1) are well-known to also describe the dynamics of the order parameter of many statistical mechanical systems near second-order phase transitions [13]; see [14] for a review. In this analogy, $V(\phi)$ is the Landau potential as a function of the order parameter. E.g., for a magnet, the order parameter is the overall magnetization.

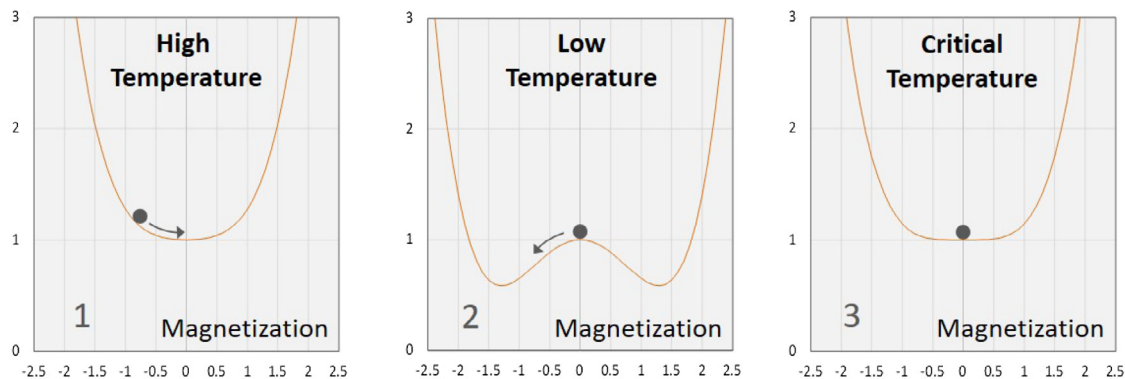
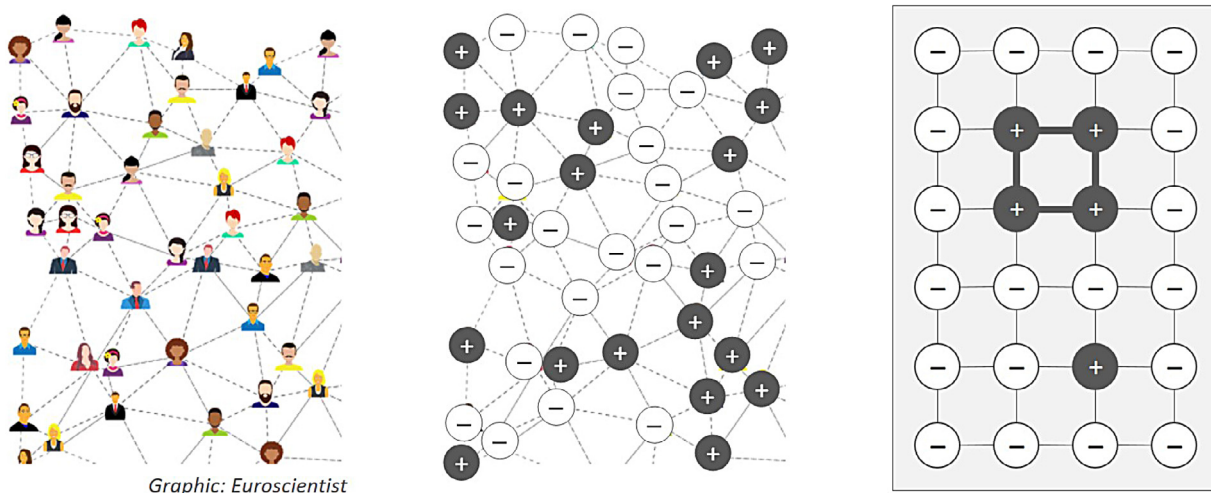


Fig. 2. Landau potential of the Ising model as a function of the magnetization at high temperature (left), low temperature (center), and the critical temperature (right).



Graphic: Euroscientist

Fig. 3. Left: the social network of investors. Center: investors that are over/under-weight in a given asset relatively to the CAPM allocation are replaced by spins pointing up/down. Right: excerpt of an Ising model with 5 spins pointing up and energy $E = -4/2 = -2$.

The classical such system is the Ising model [15], which consists of magnetic spins on a square lattice. Spins can point either up or down, and neighboring spins tend to align (Fig. 3, right). The effective potential $V(\phi)$ describes the energy of the system as a function of the magnetization ϕ , i.e., of the average value of the spins (Fig. 2). There is a low-temperature magnetized phase, where the potential has two minima corresponding to spins pointing either up or down, and a high-temperature disordered phase, where the average magnetization at the minimum of the potential is zero. The two phases are separated by a second-order phase transition at some critical temperature. The time evolution of the magnetization follows a similar Langevin equation as we have derived for the trend strength ϕ .

Of course, this could be a coincidence. On the other hand, analogies between financial markets and critical phenomena, such as scaling relations, have long been observed [16–18]. Much of the literature has focused on the scaling of volatilities rather than autocorrelations; see [19] for a summary of empirical results and, e.g., [20] for a general review. Such analogies are plausible, if financial markets are regarded as social networks, whose nodes correspond to individual investors (Fig. 3), who interact with each other in analogy with spins on a lattice, thereby creating the macroscopic phenomena of trends (herding behavior) and reversion (contrarian behavior). To replicate these phenomena, various spin- and agent models have been proposed (see, e.g., [21–23]).

In fact, the idea of using the Ising model and its cousins to model social networks and financial markets has a long history (see [24] for an extensive review). What is new about the results of [11] is that they make these long-standing ideas very specific, in the sense that we can now directly observe such a quartic Landau-like potential in financial markets, and we can empirically measure its coefficients and their scaling behavior. This allows us to make this conjectured relation between financial markets and phase transitions on a social network precise, and to test quantitatively, whether it agrees with financial market data.

The first question is what exactly the microscopic degrees of freedom represent, which sit on the vertices of the network, corresponding to individual investors. If we would identify up/down spins with buy/sell orders, as in the review [24], this would imply that the role of the order parameter (the magnetization) is played by the net order size, and not by the trend strength, as in our empirical observations.

We therefore take a different route. We identify the microscopic degrees of freedom with the relative portfolio weights, by which the investor sitting on a given node over- or underweights a given asset. By “relative”, we mean relative to the market capitalization of the asset, as in the CAPM model. If we represent these relative weights by spins that can point either up (over-weight) or down (under-weight), and if we also assume that neighboring spins tend to align due to herding behavior, then we have all the ingredients that make a phase transition plausible, in analogy with the magnet. Moreover, we will argue that, in efficient markets [25–27], this system is driven to the critical temperature, at which the phase transition becomes second-order. It thus represents a form of self-organized criticality [28].

We will show that this implies that the order parameter $\pi(t)$, integrated over the whole lattice (i.e., the overall magnetization), corresponds to the deviation of the asset price from its long-term value. This deviation thus follows a Langevin equation with potential $V(\pi)$:

$$\dot{\pi}(t) = a - \frac{\partial}{\partial \pi} V(\pi) + \epsilon(t) \quad \text{with} \quad V(\pi) = -\frac{b}{2} \cdot \pi^2 + \frac{g}{24} \cdot \pi^4. \quad (2)$$

We will further show how this can indeed be translated into an equation of the form (1) for the trend strength ϕ_T , with coefficients b_T, g_T that depend on the time scale.

Other models, in which the deviation of the asset price from its long-term mean moves in a polynomial potential have previously been considered [29–32]. There, various potentials were found for shorter time windows, whose coefficients depend on the asset and the time window. By contrast, our quartic potential is measured over the whole 30-year time window, and is universal across assets. Indeed, if we assume that (2) arises from the dimensional reduction of a higher-dimensional system near its critical point, this implies strong constraints on stochastic price processes such as (2) and their scaling behavior: it allows only processes that derive from renormalizable higher-dimensional field theories.

Thus, the network leaves its imprint on the asset price dynamics even after it is integrated over. This allows us to infer the properties of the presumed unobservable social network that makes up financial markets from the observable interplay of trends and reversion. In the case where the network topology is assumed to be that of R^D , we find that the correlation time in financial markets must be at least of the same order of magnitude as the length of the economic cycle. At shorter time horizons, there must be a “scaling regime” with a power-law fall-off of autocorrelations. Consistency with empirical observations, such as Hurst exponents, then implies a (fractal) network dimension of approximately $D \approx 3$.

However, upon closer look at the auto-correlations of market returns, we also find that they cannot be fully explained by a simple Ising model on a network with the topology of R^D . In a next step, this approach should therefore be extended to other models of critical dynamics, as well as to more general network topologies. Candidates for the latter include small-world networks [33], scale-free networks [34], or the Feynman diagrams of large- N field theory [35]. For a recent review, see [36]. To our knowledge, no convincing specific model has emerged as a consensus so far. However, our results provide an empirical basis for accepting or rejecting such candidates: any statistical–mechanical model of financial markets, if accurate, must replicate the interplay of trends and reversion observed in [11].

We hope that this will lead us to a specific and precise statistical model of the presumed social network of investors, thereby lead to a deeper understanding of financial markets, and – along the way – make the powerful tools of field theory accessible to finance.

2. Lattice gas model of financial markets

We begin with a rudimentary lattice gas model of financial markets. It is based on the following highly simplifying assumptions, to be refined later:

- The social network of investors is modeled by a hypercubic lattice of given dimension D . Each node corresponds to an investor.
- There are only two assets in the world: (i) a bond, which can also be thought of as cash, and (ii) a single stock, whose shares represent the molecules of the lattice gas
- Each investor can hold either one bond, or one share of the stock. The fact that an investor can hold at most one share represents a “hard core” of the gas molecules, or in other words, two molecules cannot sit at the same lattice site.
- The long-term equity allocation of the average investor is 50%, i.e., if there are N lattice sites (investors), the average number of shares is $n_0 = N/2$.
- Neighbors tend to align their positions: if your “friend” owns the stock, you tend to also be interested in it. This represents an attractive force between gas molecules.

A sample configuration of how shares may be distributed is shown in Fig. 3 (right). Needless to say, the real network of investors is *not* a hypercubic lattice, and more generally, one may say that this is at best a rough cartoon of any real physical or financial system. However, it is well-known that this primitive lattice gas model can precisely describe real

gases, such as water and steam or carbon dioxide, near their second-order phase transitions. There, all the microscopic details of these systems become irrelevant, resulting in a universal macroscopic behavior. Let us therefore assume that this phenomenon of universality also occurs on social networks. Later, we will allow for the possibility that these universality classes must be refined for nontrivial network topologies. For now, let us quantify the above model.

For a given configuration, we define the occupation numbers $n_i \in \{0, 1\}$ as the number of shares the investor i holds. The total number of shares is $n = \sum_i n_i$. E.g., in Fig. 3 (right), $n = 5$. We also define the energy E of a configuration \vec{n} of occupation numbers as

$$E(n) = -\frac{1}{D} \sum_{\langle ij \rangle} n_i n_j, \tag{3}$$

where $\langle ij \rangle$ denote all pairs of neighbors, i.e., all links on the lattice. The normalization factor $1/D$ accounts for the fact that each node appears in $2D$ links. E.g., in Fig. 3 (right), $E = -4/2 = -2$, as there are four links with one share on both ends. This term represents the attractive force, as shares tend to lie next to each other to minimize the energy.

We now imagine that the occupation numbers fluctuate while the total number of shares is kept fixed. In other words, shares can change owner (i.e., their location on the lattice). In addition, we assume that new shares can be issued or existing shares can be redeemed at a given cost μ , which we set equal to 1. So we add a second term to the energy (3):

$$E(n) = -\frac{1}{D} \sum_{\langle ij \rangle} n_i n_j + \mu \sum_i n_i \quad \text{with } \mu = 1. \tag{4}$$

This term allows us to also let the total number of shares fluctuate, as in a system that is in equilibrium with a larger system. It is a counterweight to the attractive force (if μ was zero, the network would fill up with shares to minimize the energy). μ (or $-\mu$, depending on the convention) is called the ‘‘chemical potential’’ in statistical mechanics.

Where in this model is the time-dependent share price $P(t)$ that corresponds to real market prices? Since we have implicitly assumed that new shares are issued or redeemed at price 1 to satisfy any increase or decrease of demand, the share price is not dynamical. However, there is a dynamical market capitalization, namely the total number of shares $n(t)$, which is equivalent to a dynamical share price.

To make this equivalence precise, suppose that the firm did not issue or repurchase shares. Thus, $n = n_0$ would be fixed, rather than fluctuate with the demand. Instead, there would be a share price $P(t)$ that dynamically adjusts to satisfy the demand. Then the market capitalization would be $n_0 \cdot P(t)$. Putting both equations for the market capitalization together, we identify

$$P(t) = n(t)/n_0 \tag{5}$$

as the dynamical share price, where $n_0 = N/2$ is the long-term value of the firm.

As is well-known, the lattice gas model is equivalent to the Ising model. Replacing the occupation numbers n_i by spins s_i , the energy (4) simply becomes

$$E(n) = -\frac{1}{D} \sum_{\langle ij \rangle} s_i s_j \quad \text{with } s_i = n_i - 1/2 \in \{+1/2, -1/2\}, \tag{6}$$

up to an irrelevant constant that depends only on the total number of lattice sites. The spins represent the over- or under-weight of a given investor in the stock (setting $\mu \neq 1$ in (4) would correspond to introducing an external magnetic field in the Ising model; here, we focus on the case $\mu = 1$). From (5), (6), the total magnetization

$$M(t) = \sum_i s_i = n(t) - \frac{N}{2} = \frac{N}{2}(P(t) - 1) \tag{7}$$

of the Ising model is proportional to the deviation of the share price from its long-term value.

To summarize, we have re-interpreted the well-known Ising model in terms of financial markets. The order parameter (the overall magnetization) plays the role of the deviation of the share price from its long-term equilibrium value as defined, e.g., in the CAPM model. In the above, we have assumed a long-term market weight of 50% for the asset; this can be generalized to an arbitrary market weight by assigning different probabilities p and $(1 - p)$ to the occupation numbers 0 and 1.

For k assets, the spins have k components. E.g., for $k = 3$, they can point in three different directions, corresponding to relative weights of a portfolio of three assets. It will be interesting to empirically measure whether there is an $O(k)$ symmetry for uncorrelated assets, in which case the Ising model would be replaced by a $O(k)$ vector model, such as a Heisenberg model for $k = 3$.

As usual in statistical mechanics, we now consider the ensemble of all possible spin configuration \vec{s} , weighted with relative probability $G(\vec{s})/Z$, where

$$G(\vec{s}) = \exp\{-\frac{1}{kT}E(\vec{s})\} \quad \text{and} \quad Z = \sum_{\vec{s}} G(\vec{s}) \tag{8}$$

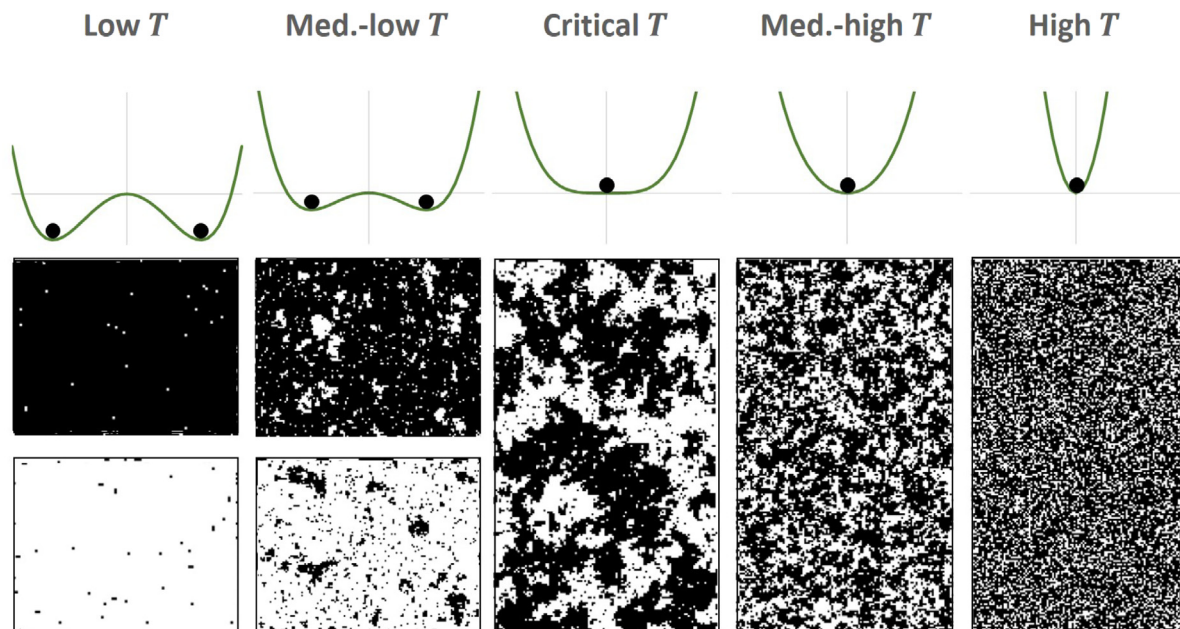


Fig. 4. Computer simulation of the lattice gas/Ising model. The upper graphs show the Landau potential. The lower graphs show typical spin configurations. If interpreted as assets in the network of investors, black dots represent shares, and white dots represent cash.

where G is the “Gibbs factor”, and the “partition function” Z acts as a normalization factor. The parameter T (which is unrelated to the time horizon) controls how important the statistical fluctuations are. In physics, T is interpreted as the temperature. T is multiplied by Boltzmann’s constant k to give it the dimension of an energy. Below, we will relate T to the persistence of trends b . Given (8), the expectation value of a spin s_i is

$$\langle s_i \rangle = \frac{1}{Z} \sum_{\vec{s}} s_i \exp\left\{-\frac{1}{kT} E(\vec{s})\right\}.$$

Fig. 4 shows images from a computer simulation of this model. Black dots represent shares, white dots represent cash. There is a high-temperature phase, corresponding to a strongly positive quadratic potential term, where all spin configurations are equally likely according to (8), and the spins are thus randomly distributed. On the other hand, at low temperature, the energy (6) is much more important than the statistical fluctuations in (8). So the lowest energy configurations dominate: all spins point either up or down, corresponding to the two minima of the potential with a strongly negative quadratic term.

As the temperature rises, statistical fluctuations become more important in (8) and bubbles – so-called Peierls droplets – start to form. Likewise, if the temperature is reduced starting from the high-temperature phase, spins start to feel the aligning force (6) and clusters start to form. Both the clusters and the bubbles have a typical size, the correlation length ξ . As we approach the critical point, where the quadratic term in the potential vanishes and the second-order phase transition occurs, the correlation length diverges: structures of all sizes can be seen, and the system is scale invariant. At and above the critical temperature, the distinction between the two ground states disappears: one cannot distinguish whether one observes clusters of shares in a sea of cash, or clusters of cash in a sea of shares.

Let us now argue that financial markets are driven towards this critical point. First, suppose that the persistence of trends b , which corresponds to the quadratic term of the Landau potential, was significantly larger than zero, as in the low-temperature phase. Then trend-following would be so profitable that large amounts of capital would flow into this strategy. This would drive b back to zero. On the other hand, if b was significantly smaller than zero, markets would revert strongly and reliably. As a result, mean reversion strategies (such as statistical arbitrage) would be so successful that a large amount of capital would flow into them. This would again drive b back to zero.

The important conclusion is that b should be close to its critical value zero in efficient markets. In fact, a key empirical observation of [11] was that b , which was already small (of the order of a few percent) several decades ago, has further decreased since then. It now appears to be so close to zero that even a diversified portfolio of pure trends can no longer be expected to yield significant excess return. This can indeed be explained by the fact that hundreds of billions, if not trillions of dollars, have flown into trend-following, thereby eliminating much of this market inefficiency. It gives insight into how markets dynamically *become* efficient over time. The precise dynamical feedback mechanism that drives b to 0 will be studied in separate work.

Returning to the lattice gas, as we approach the critical temperature T_c from below, the expectation value $\langle M(T) \rangle$ of the magnetization (7) goes to zero. If one measures how exactly it goes to zero, one finds a power law. Likewise, the correlation length ξ diverges as a power:

$$M(T) \sim |\hat{t}|^\beta, \quad \xi(T) \sim |\hat{t}|^{-\nu} \quad \text{with } \hat{t} \equiv (T - T_c)/T \quad (9)$$

with so-called “critical coefficients” β, ν that can be measured or computed. E.g., in $D = 3$ dimensions one finds $\beta = 0.33$ and $\nu = 0.63$. The crucial property of β, ν is that they are “universal”. E.g., our primitive lattice gas model in $D = 3$ dimensions yields exactly the same values β, ν as are observed for the second-order phase transitions of water and steam, of CO_2 , and of all other systems in the same universality class. In fact, β and ν depend only on 2 parameters: on the dimension D of the lattice (but not on whether it is a square, triangular, or similar lattice), and on the symmetries. Here, the symmetry is Z_2 , corresponding to flipping all spins.

A potential caveat is that this only applies to networks with the trivial topology of D -dimensional Euclidean space. For more general network topologies, it is conceivable that the universality classes are characterized by additional parameters, such as the exponent γ of the so-called “degree distribution” $p(k) \sim k^{-\gamma}$ of nodes with k neighbors in the case of scale-free networks [34]. It will be important to investigate this in the future.

In either case, even though our lattice gas model of financial markets may seem very far from reality on a microscopic scale, the phenomenon of universality may allow it to precisely model the macroscopic behavior of real markets in the vicinity of the phase transition. In particular, the universal coefficients β and ν in (9) also determine how the system behaves under a rescaling in space and time. E.g., at the critical point, the covariance of spins at different points \vec{x}, \vec{y} scales with their distance like

$$\langle s(\vec{x})s(\vec{y}) \rangle \sim |\vec{x} - \vec{y}|^{2-D-\eta} \quad \text{with } \eta = 2\frac{\beta}{\nu} + 2 - D. \quad (10)$$

In the next section, we will see how η can also be used to predict the auto-correlations of financial market returns at different points in time. Those can be compared with auto-correlations in real markets. This establishes a direct link between observable properties of asset prices, and the critical exponents of the presumed social network that makes up financial markets but cannot be observed directly.

3. Field theory of financial markets

In the limit of infinitely many lattice sites with lattice spacing $a \rightarrow 0$, the lattice becomes a continuous manifold of dimension D . Since networks generally have fractal dimensions, we formally allow D to be any positive real number. The index i becomes a D -dimensional vector \vec{x} , and the spins $s_i(t)$ become a field $\phi(\vec{x}, t)$ that measures the local magnetization, corresponding to the excess demand for an asset in a given neighborhood of the social network of investors. Let us first consider the system at a fixed point $t = 0$ in time. One replaces the partition function Z in (8) near the critical temperature $T \sim T_c$ by a path integral

$$Z = \int D\phi(\vec{x}) \exp\{-S[\phi(\vec{x})]\}$$

with action $S[\phi]$ and Landau potential $V(\phi)$:

$$S[\phi(\vec{x})] = \int d^D\vec{x} \left\{ \frac{1}{2} (\nabla\phi)^2 + V(\phi) \right\} \quad \text{with } V(\phi) = \frac{1}{2} r\phi^2 + \frac{1}{24} g\phi^4. \quad (11)$$

Here, the factor $1/kT$ in (8) has been absorbed such that ϕ has a standard kinetic term. The temperature then re-appears in the potential: r is proportional to the deviation $(T - T_c)$ from the critical temperature. For a general introduction to field theory and critical phenomena, see, e.g., [37,38]. Here, we merely list some results that we need in the following:

- The parameters $r(\lambda)$ and $g(\lambda)$ in (11) “flow” under so-called renormalization group transformations [39], in which all distances are rescaled by a factor λ , and all objects with a scaling dimension are rescaled accordingly (see Fig. 5).
- In particular, for dimensions $D \geq 4$, $g(\lambda)$ flows towards zero at large scales, so the quartic term of the potential becomes weaker and weaker. For $D > 4$, the macroscopic scaling behavior of the system is the same as for a free theory (i.e., $g = 0$).
- For dimensions $D < 4$, $g(\lambda)$ flows to a nontrivial fixed point value g^* at large scales (Fig. 5). The coefficients of all other possible terms in the action, such as $\phi^6, (\nabla\phi)^2, \dots$, flow to zero. This explains why systems that are very different on a microscopic scale have the same macroscopic behavior with the same critical coefficients.
- At this fixed point, the field ϕ scales as $\phi \rightarrow \lambda^{\eta/2}\phi$ with the so-called “anomalous dimension” $\eta = 2(\beta/\nu) + 2 - D$. η can be computed perturbatively in g^* via Feynman diagrams [37,38]. Table 1 lists estimated values of η for various dimensions D [37,40].

Thus, the mere fact that we empirically observe a quartic potential in (1) may already be a first indication that the fractal dimension of the social network of investors is less than 4.

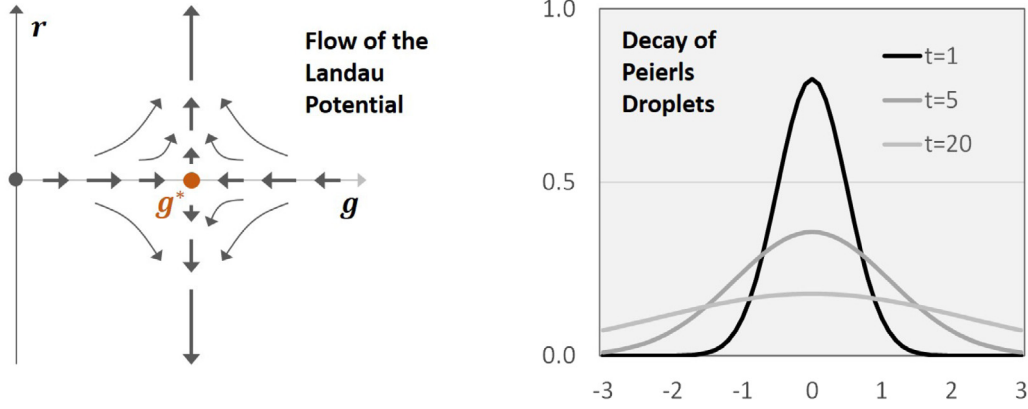


Fig. 5. Left: Renormalization Group Flow in ϕ^4 theory. At large scales, the coefficient of the quartic term flows to a fixed point value g^* . The coefficient of the quadratic term must normally be fine-tuned to zero to reach the critical point. In our model, it is attracted to zero in efficient markets. Right: time evolution of Peierls droplets of positive magnetization.

Table 1
Critical coefficients vs. network dimension D .

Dimension	4	3.5	3	2.5	2	1.5
η	0.00	0.002	0.036	0.106	0.250	0.523
z	2.000	2.001	2.024	2.071	2.167	2.352
$\kappa \equiv (2 - \eta)/z$	1.000	0.998	0.970	0.915	0.808	0.628

The above points refer to static critical phenomena. Let us also discuss the time evolution of the system near the critical point, when it is slightly out of equilibrium. For a general introduction to this field of Critical Dynamics, see [14]. In this paper, we focus on the simplest case – the so-called purely dissipative “Model A” [13].

By emerging and decaying, Peierls droplets of positive or negative magnetization create trends of the overall magnetization. In our model, where the magnetization represents the deviation of an asset price from its value, this is the origin of the price trends in financial markets. The droplets represent groups of investors that collectively over- or underweigh an asset. The diffusion of the field $\phi(\vec{x}, t)$ in time is described by a Langevin equation [14,37]:

$$\frac{\partial}{\partial t} \phi(\vec{x}, t) = \frac{\Omega}{2} \left(\nabla^2 \phi(\vec{x}, t) - r \phi(\vec{x}, t) - \frac{1}{6} g \phi(\vec{x}, t)^3 \right) + \text{random noise}, \tag{12}$$

where Ω is a diffusion coefficient and the variance of the noise is Ω . By a rescaling of time, let us set $\Omega = 1$.

Fig. 5 (right) shows the diffusion of a droplet of size L in the case of one dimension, corresponding to a local magnetization with Gaussian distribution. For $g = 0$, it decays over a time span $\Delta t \sim L^2$. Correspondingly, the spatial correlation length ξ and the correlation time τ are related by $\tau \sim \xi^z$. For nonzero g , there is a correction [13]:

$$\Delta t \sim L^z, \quad \tau \sim \xi^z,$$

where z is a new critical exponent that depends on D and the symmetry group $O(N)$. Its estimated value for various dimensions and $N = 1$ (the Ising model) is shown in Table 1. For a recent calculation, see [41]. For the range of fractal dimensions D that will turn out to be relevant for us, $2 \leq D \leq 3.5$, one may approximate $z = 2 + c \cdot \eta$ [13] with $c \approx 2/3$.

In the following, we focus on the spatially constant mode of the field ϕ , i.e., its integral over the whole network, corresponding to the overall magnetization:

$$\pi(t) = \int d^D x \phi(x, t).$$

In the previous section, $\pi(t)$ has been identified with the deviation of the share price from its long-term value. Since this “value” is not observable, we cannot compare $\pi(t)$ directly with experiment. Fortunately, the unknown “value” drops out of the market returns $R(t)$, which are proportional to the price change $\dot{\pi}(t)$. In the classical action (11), ϕ is normalized to have standard kinetic term, so $\dot{\pi}$ has variance 1. On the other hand, the daily returns in our empirical analysis were also normalized to have variance 1. Thus, we directly identify

$$R(t) \equiv \dot{\pi}(t) = \frac{d}{dt} \pi(t).$$

This immediately yields a first important prediction of ϕ^4 theory for the auto-correlation of market returns that can be compared with observation:

$$\langle \dot{\pi}(t) \dot{\pi}(t') \rangle \sim -\ddot{\Delta}(t - t') \text{ with } \Delta(t) \equiv \langle \pi(0) \pi(t) \rangle, \tag{13}$$



Fig. 6. Left: autocorrelations of market returns (white dots), aggregated by the trend strength $\phi_\omega = \phi_{2/T}$. For intermediate time horizons, they turn positive, corresponding to the regime where trend-following works best. Right: the variance of the trend strength (white dots) as a function of the time scale falls off at short scales, rises at intermediate scales, and seems to fall off again at long scales. As a cross-check, similar analyses were performed with the simpler version of the trend strength $\tilde{\phi}$ (black dots) for non-overlapping time windows.

where $\Delta(t)$ is the so-called two-point function or “propagator” of the theory. Due to the huge amount of noise in financial markets, this autocorrelation of daily returns is very difficult to measure empirically. For this reason, in our empirical observations [11] we have aggregated the recent returns of a given market over a time horizon $T = 2/\omega$ in the form of the strength ϕ_ω (32) of the recent trend in that market. Its continuum definition in terms of $\dot{\pi}$ is:

$$\phi_\omega(t) = 2\omega^{3/2} \cdot \int_{-\infty}^{t-} d\zeta (t - \zeta) e^{-\omega(t-\zeta)} \dot{\pi}(\zeta) \tag{14}$$

Here, $t-$ indicates that the point t is excluded from the integration, i.e., the return at time t is predicted only from returns before time t . As explained in the appendix, the weight function $ne^{-\omega n}$, by which the return from $n \sim (t - \zeta)$ days ago is weighted in (14), peaks at $n = 1/\omega = T/2$, and the average lookback period is $\langle n \rangle = T$. The pre-factor $2\omega^{3/2}$ ensures that the variance of the trend strength is approximately one, so ϕ_ω measures the statistical significance (t-statistics) of the trend. The aggregate auto-correlation can then be written as a (variant of the) Laplace transform of the propagator:

$$\langle \phi_\omega(t), \dot{\pi}(t) \rangle = \langle \phi_\omega(0), \dot{\pi}(0) \rangle = -2\omega^{3/2} \int_{0+}^{\infty} d\zeta \zeta \cdot e^{-\omega\zeta} \ddot{\Delta}(\zeta). \tag{15}$$

Fig. 6 (left) shows the empirically observed two-point function $\langle \phi_\omega, \dot{\pi} \rangle$ on a logarithmic scale, with time horizon $T = 2/\omega = 2^k$ days. It shows a curious behavior: at intra-month scales, the (very small) autocorrelation is negative. It turns positive on scales from one month to about 1 year – the time horizons, at which trend-following has been most profitable. At time scales of several years, it turns negative again. The correct field theory of financial markets must reproduce this propagator.

As an additional cross check, we have repeated the analysis with a simpler version $\tilde{\phi}_T$ of the trend strength (14), proportional to the equally weighted average of the daily log returns in a time window of length T (this corresponds to the step function in Fig. 9 of the appendix):

$$\tilde{\phi}_T(t) = T^{-1/2} \int_{-\infty}^{t-} d\zeta \dot{\pi}(\zeta) = T^{-1/2} [\pi(t) - \pi(t - T)]. \tag{16}$$

To measure its auto-correlations empirically, we have slightly extended our history of futures returns from 30 years to 2^{13} trading days ≈ 31.5 years, and we have constructed 2^{13-k} non-overlapping time windows of length $T = 2^k$ for each k . Fig. 6 (left) shows the correlation of this new trend strength (black dots) between neighbouring time windows. Within estimation errors, these empirical results are in line with the empirical autocorrelations (15). The theoretical prediction for this correlation depends on the propagator:

$$\langle \tilde{\phi}_T(t) \tilde{\phi}_T(t - T) \rangle = -\frac{1}{T} [\Delta(0) - 2\Delta(T) + \Delta(2T)].$$

Much of the literature on scaling in financial markets is concerned with the scaling of the variance, or more generally, of the q th moments of returns over the time period T [42,43]:

$$K_q(T) = \langle |\pi(T) - \pi(0)|^q \rangle / \langle |\pi(0)|^q \rangle \Rightarrow K_2(T) = 2 - 2\Delta(T)/\Delta(0). \tag{17}$$

With a limited amount of data, these moments are easier to measure empirically than the auto-correlations of returns. Here we focus on $q = 2$. From (13), (14) and (16), and after some algebra, we obtain formulas for the variance of our trend strengths:

$$\langle \phi_\omega(0) \phi_\omega(0) \rangle = -2\omega^3 \int_{0+}^{\infty} du e^{-\omega u} \int_0^u dv v \dot{\Delta}(v), \quad (18)$$

$$\langle \tilde{\phi}_T(0) \tilde{\phi}_T(0) \rangle = \frac{2}{T} [\Delta(0) - \Delta(T)] \quad \text{with } T = 2/\omega. \quad (19)$$

Note that these trend strengths were normalized to have variance 1, if the small auto-correlations of financial market returns are neglected. (18) and (19) thus describe the precise effect of the auto-correlations on this normalization, as predicted by field theory.

Fig. 6 (right) shows the empirical results. They mirror the curious behavior found for the propagator: the variance of the trend strength as a function of the time scale decreases at short scales, increases at intermediate scales, then may decrease again. The correct field theory description of financial markets must reproduce this empirically observed variance.

A further important test comes from the Langevin Eq. (12). For the spatially constant mode $\pi(t)$, which we have identified with the asset price minus its value, it implies:

$$\dot{\pi}(t) = -\frac{r}{2} \cdot \pi(t) - \frac{g}{12} \cdot \pi(t)^3 + \eta(t), \quad (20)$$

where η represents Gaussian noise. Again, π is not directly observable because of the unknown “value” of an asset. However, this “value” cancels out of the trend strength. To compare with observation, we must translate (20) into an equation involving only the trend strength. Generally, the coefficients will then depend on the time horizon $T = 2/\omega$:

$$\dot{\pi}(t) = \beta_\omega \cdot \phi_\omega(t) + \gamma_\omega \cdot \phi_\omega(t)^3 + \epsilon_\omega(t) \quad (21)$$

The multiple regression coefficients $\beta_\omega, \gamma_\omega$ can be found by inverting the covariance matrix:

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \langle \phi^2 \rangle & \langle \phi^4 \rangle \\ \langle \phi^4 \rangle & \langle \phi^6 \rangle \end{pmatrix}^{-1} \begin{pmatrix} \langle \phi, \dot{\pi} \rangle \\ \langle \phi^3, \dot{\pi} \rangle \end{pmatrix} = \begin{pmatrix} \langle \phi^2 \rangle & \langle \phi^4 \rangle \\ \langle \phi^4 \rangle & \langle \phi^6 \rangle \end{pmatrix}^{-1} \begin{pmatrix} -\frac{r}{2} \langle \phi, \pi \rangle - \frac{g}{12} \langle \phi, \pi^3 \rangle \\ -\frac{r}{2} \langle \phi^3, \pi \rangle - \frac{g}{12} \langle \phi^3, \pi^3 \rangle \end{pmatrix},$$

where all fields sit at $t = 0$ (we have suppressed the index ω). This reduces the problem to the (tedious but straightforward) computation of various correlation functions in ϕ^4 theory, which is left for future work.

The correct field theory must replicate the empirically observed coefficients b_T, g_T of (1), as shown in Fig. 1d. These results can be regarded as a refinement of the results for the propagator plotted in Fig. 6 (left): inserting (20) into the expectation value $\langle \dot{\pi}(0) \dot{\pi}(t) \rangle$ decomposes the propagator into a component that comes from the persistence of trends r , and another component that comes from the strength of reversion g . In particular, this should allow us to directly measure the fixed point value g^* , once we have found the right model. As g^* depends on the (possibly fractal) dimension D of the presumed underlying social network of investors, this will amount to a direct empirical estimate of D .

4. Illustration: D -dimensional lattice

In the previous section, we have derived three key requirements – yielding the correct propagator, the correct variance of the trend strength, and the correct Langevin equation – that a field theory of financial markets must satisfy in order to agree with observation.

In this section, we illustrate how to apply these criteria in practice at the simplest example, to be generalized in future work: the purely dissipative model A on a hypercubic lattice as discussed in Section 3. We do *not* necessarily expect this model to closely match observation. One of the reasons is that the social network of investors is surely much more connected than a hypercubic lattice, where the shortest path between two sites is on average very large. However, this example demonstrates how to connect the rather abstract concepts of quantum field theory with the practical observation of financial markets in order to derive two key properties of the presumed underlying social network of investors: its fractal dimension D , and its correlation length ξ . To this end, we discuss two distinct regimes:

- the “scaling regime” at small time scales T (i.e., $\omega = 2/T \rightarrow \infty$). Here, the renormalization group dictates a power-law fall-off of all correlation functions. Matching with observation should yield an estimate of the anomalous dimension η , and thereby of the dimension D of the network.
- the “exponential regime” at large time scales (i.e., $\omega \rightarrow 0$). Here, all correlations fall off exponentially. Matching with observation should yield an estimate of the correlation time τ of market prices, and thereby of the correlation length $\xi \sim \tau^{-z}$ in the network.

These two regimes also appear in more general models of critical dynamics, albeit with different scaling exponents. The precise interpolation between the two regimes at intermediate time scales should hold the key to identifying the correct model, beyond “model A” on a hypercubic lattice. This is left for future work. In this paper, we only use the fact that the scaling regime must approach the exponential regime, as the dimension approaches the critical dimension $D = 4$.

Section 4.1 discusses the propagator of the model. Section 4.2 turns to the scaling of variances and compares the theoretical predictions with the observed variances and Hurst exponents. Section 4.3 discusses the results.

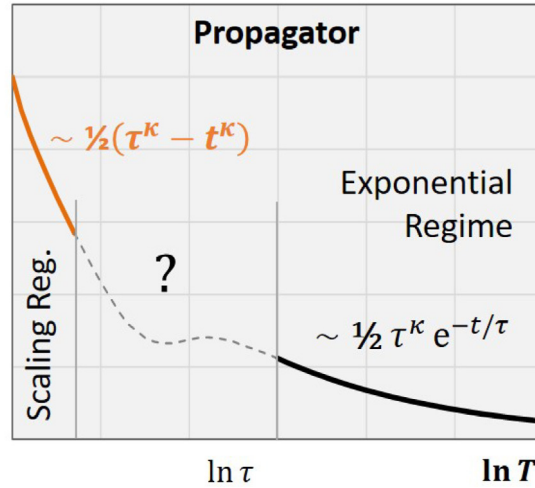


Fig. 7. Left: the propagator $\Delta(t)$ of the spatially constant mode falls off like a constant minus a power at small time scales, and exponentially at time scales larger than the correlation time. The unknown behavior of the correct model in the intermediate regime is indicated by a question mark.

4.1. The propagator

We consider “model A” on a Euclidean manifold of dimension D (representing the network of investors) and set the variance Ω of the random noise to 1. For the free theory ($g = 0$, corresponding to $D > 4$), the propagator of the spatially constant mode $\pi(t)$ is [14]:

$$G_d(p, t) = \frac{1}{p^2 + r} \exp\left\{-\frac{p^2 + r}{2} |t|\right\} \Rightarrow \Delta(t) = \frac{1}{2} \cdot \xi^2 \exp\{-|t|/\xi^2\},$$

where ξ is the correlation length, which is equal to $\sqrt{2/r}$ for the free theory with $g = 0$. For $g \neq 0$ and $D < 4$, the renormalization group provides the following scaling relation for the exact two-point function in the critical domain, slightly above the critical temperature (where ξ in (9) is finite) [14]:

$$\Delta(t) = \langle \pi(0)\pi(t) \rangle = \xi^{2-\eta} \cdot g(t/\xi^z),$$

with some function $g(x)$ that falls off exponentially at large x . Let us first consider the scaling regime. Consistency with the static limit $t \rightarrow 0$ (where $\Delta(0) \sim \xi^{2-\eta}$ from the spatial integral over \vec{x}), and with the limit $D \rightarrow 4$ (where $\kappa \rightarrow 1$ as in Table 1) implies:

$$\Delta(t) = \frac{1}{2} \cdot \tau^\kappa - \frac{1}{2} \cdot |t|^\kappa \quad \text{for } t \ll \tau \quad \text{with } \tau \equiv \xi^z, \quad \kappa \equiv \frac{2-\eta}{z} < 1 \tag{22}$$

On the other hand, in the exponential regime, consistency with the limit $D \rightarrow 4$ implies

$$\Delta(t) = \frac{1}{2} \cdot \tau^\kappa \exp\{-|t|/\tau\} \quad \text{for } t \gg \tau \tag{23}$$

In the low-temperature phase, these relations hold for the connected two-point function $\langle \pi(0)\pi(t) \rangle - \langle \pi(0) \rangle^2$ instead. In both cases, we have for $t > 0$:

$$\dot{\Delta}(t) = \begin{cases} -\frac{1}{2} \tau^{\kappa-1} \cdot \exp\{-t/\tau\} \\ -\frac{1}{2} \kappa \cdot t^{\kappa-1} \end{cases}, \quad \ddot{\Delta}(t) = \begin{cases} \frac{1}{2} \tau^{\kappa-2} \cdot \exp\{-t/\tau\} \\ \frac{1}{2} \kappa(1-\kappa) \cdot t^{\kappa-2} \end{cases} \quad \text{for } t \gg \tau \\ \text{for } t \ll \tau \tag{24}$$

Note that this holds both above and below the critical temperature: the propagator in these cases differs only by the expectation value of π , which is a constant that drops out of $\ddot{\Delta}$.

Fig. 7 schematically shows this propagator. The singularity of the orange line is the imprint that the invisible higher-dimensional social network of investors leaves on the stochastic process $\pi(t)$, even after it is integrated out. As $D \rightarrow 4$ and $\kappa \rightarrow 1$, the propagator in the scaling regime approaches that of the exponential regime, and the singularity disappears.

The propagator (22), (23) predicts that auto-correlations $-\ddot{\Delta}$ (13) of market returns are purely negative. Thus, it does not explain the positive autocorrelations that we observe in Fig. 6 (left) on intermediate scales from one month to one year. These empirical results suggest that the true propagator Δ has negative curvature in part of the intermediate regime, which creates the phenomenon of market trends. Again, this should be a key clue for finding the true model in the future.

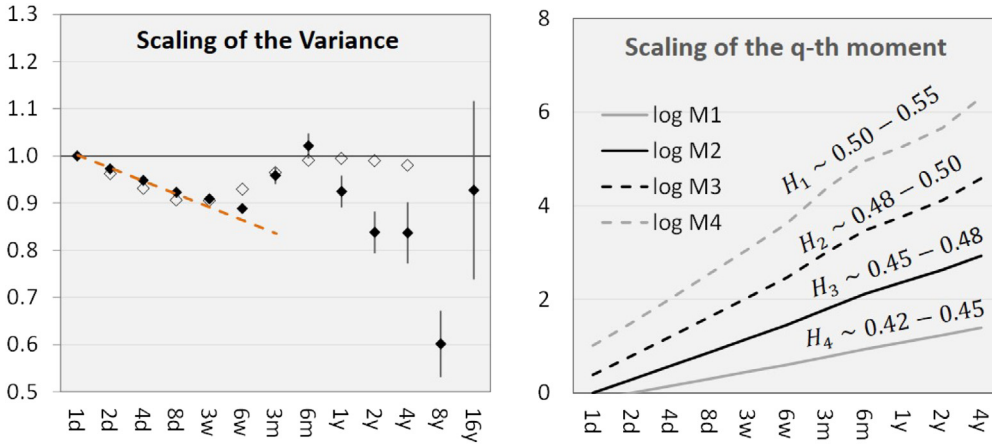


Fig. 8. Left: Fit of the scaling behavior of the variance in the scaling regime (dashed orange line) by the field theory prediction with critical coefficient $\kappa \approx 0.96$, corresponding to a fractal dimension of the presumed underlying network of $D \approx 3$. The correlation time, where the variance falls off rapidly, must be at least as long as the economic cycle. Right: Scaling of the q th moment M_q for $q = \{1, 2, 3, 4\}$ as a function of the time scale T in our dataset of liquid daily futures returns. The slopes of the lines are the Hurst coefficients H_q .

4.2. Scaling of the variance

For the time being, though, let us nevertheless proceed and discuss the scaling regime and the exponential regime, where our model is in line with observations. To this end, it is more convenient to consider the scaling of the variances (18) and (19) of the trend strengths instead of the propagator. Fig. 8 (left) extends the empirically estimated variance of fig. 6 (right) to 16-year windows. From (18), (19) and (24), we obtain in the scaling regime $T = 2^k = 2/\omega \ll \tau$ for $\kappa \approx 1$:

$$\langle \phi_\omega^2 \rangle \sim \omega^{1-\kappa} = e^{(1-\kappa)\ln \omega} \approx 1 - (1 - \kappa) \ln 2 \cdot k \tag{25}$$

$$\langle \tilde{\phi}_T^2 \rangle = T^{\kappa-1} = e^{(\kappa-1)\ln T} \approx 1 - (1 - \kappa) \ln 2 \cdot k \tag{26}$$

Thus, κ can be read off from the slope $(1 - \kappa) \cdot \ln 2$ of the variance as a function of k (orange line in Fig. 8, left). We measure by regression:

$$\kappa = 0.96 \pm 0.01 \quad \Rightarrow \quad D = 2.9 \pm 0.1. \tag{27}$$

Here, we have interpolated the formally non-integer dimension D from Table 1 based on the approximation $z = 2 + \eta(D) \cdot 2/3$. We propose to identify D with the fractal dimension of the network; although there are different definitions of this fractal dimension, they coincide for homogeneous networks such as ours, where all nodes have the same number of links (for other network topologies, this is generally not the case, though [44]).

On the other hand, in the exponential regime $T = 2/\omega \gg \tau$ we get from (18), (19) and (24):

$$\langle \phi_\omega^2 \rangle = \frac{\omega^2}{(\omega + 1/\tau)^2} \cdot \tau^{\kappa-1}, \quad \langle \tilde{\phi}_T^2 \rangle = \frac{\tau}{T} (1 - e^{-T/\tau}) \cdot \tau^{\kappa-1} \tag{28}$$

In both cases, field theory predicts the variance of trends to fall off rapidly as T becomes larger than the correlation time. There is some indication from Fig. 8 (left) that this happens at $T \sim 8$ years, but we do not have enough data to confirm this with significance, and we cannot even rule out that the correlation time is infinite. All we can say is that the correlation time must be larger than 4 – 8 years. Interestingly, this lower bound is of the same order of magnitude as the length of the economic cycle. A more precise statement will require an analysis of many decades, if not centuries of historical data.

Let us also comment on Hurst exponents. We use the definition of the q -th generalized Hurst exponent H_q in terms of the scaling of the q th moment of π (see (17)):

$$M_q(T) \equiv \langle |\pi(T) - \pi(0)|^q \rangle \sim \langle |\pi(0)|^q \rangle \cdot T^{qH_q} \quad \text{with } q \in \mathbb{R}.$$

The process π is called “mono-scaling”, if all H_q are identical, and “multi-scaling”, if H_q depends on q . There is ample evidence that financial market returns are “multi-scaling”, not “mono-scaling” (see [19] and references therein, and [20] for a general review of scaling properties, including in financial markets).

Our data set also supports this. Fig. 8 (right) shows the empirical measurement of $M_q(T)$ for T ranging from 1 day to 4 years for our 30-year history of daily futures returns across asset classes (see the appendix). From the slopes of the lines, we can estimate the generalized Hurst exponents. The results are consistent with those for liquid markets in [19] (our dataset only covers liquid futures). Let us remark that our Hurst exponents are not quite the same for all horizons, though. E.g., H_2 seems to increase from 0.48 at short scales to 0.5 at scales of several years.

What does field theory predict for the Hurst exponents? We focus on even, integer $q = 2n$. In the scaling regime, we get from (16) and (26) and the renormalization group [37]:

$$M_2(T) = T \langle \tilde{\phi}_T^2 \rangle \sim T^\kappa$$

$$M_{2n}(T) = T^n \langle \tilde{\phi}_T^{2n} \rangle \sim T^{n\kappa} \quad \text{with } n \in \{1, 2, 3, \dots\}$$

Thus, for infinite correlation time, field theory predicts mono-scaling with Hurst exponent

$$H = \frac{\kappa}{2} \quad \text{with } \kappa = \frac{2 - \eta}{z} \quad (29)$$

independently of n . Thus, the Hurst exponent is directly related to the anomalous scaling dimension η of the deviation π of the asset price from its value, which in turn is related to the dimension D of the underlying network of investors. E.g., for $D = (2, 3, 4)$, field theory predicts $H \approx (0.40, 0.485, 0.50)$. However, any finite correlation time will destroy this mono-scaling. How precisely the H_{2n} depend on n in this case, and whether this can explain the observed multi-scaling behaviour of financial markets, must be addressed in future work.

4.3. Discussion

To summarize this section, we have tried out the simplest model of critical dynamics ("model A") on the simplest network topology (a D -dimensional hypercubic lattice) at the critical point (where $g = g^*$ is at its fixed point) in order to model financial markets as a social network of investors. Although these assumptions were radical simplifications that seem unrealistic, the predictions of this model are not all that far from the reality of financial markets.

At small time scales, it predicts a power-law fall-off of autocorrelations as a function of the time lag, and a scaling behavior of volatilities that is reminiscent of the one that has been observed in financial markets. In our approach, the origin of the associated scaling laws and of Hurst exponents that differ from $1/2$ is the same as in second-order phase transitions: they reflect anomalous dimensions that are well-known from quantum field theory. We have demonstrated how this allows us to infer the (fractal) dimension of the presumed underlying social network of investors. It is encouraging that the estimated fractal dimension $D \approx 3$ is in line with reported fractal dimensions of other social networks, the majority of which lie between 2 and 4 [44].

Moreover, it is highly nontrivial that the small typical values of the anomalous dimension η have just the right magnitude to explain the observed small negative 1-day market autocorrelations of a few percent. We take this as a sign that we are on the right track by modeling financial markets in analogy with critical phenomena. At the same time, we have seen that "model A" does not describe the intermediate regime at horizons from a few weeks to a few years well. We may need a more complex model, or a more complex network topology, or we may see non-universal properties of the network (see the concluding remarks).

At large time scales, our model predicts an exponential fall-off of all auto-correlations. This should be verified by studying historical market data on a scale of decades and centuries. Of course, the question is then whether markets in previous centuries can already have been efficient enough to be driven to the critical point by the mechanism described in Section 2, and whether the structure and dimension of the network of investors has been stable in time.

5. Conclusion and outlook

Motivated by the previous empirical observations [11] on trends and reversion in financial markets, in this paper we have proposed a statistical–mechanical model of the markets that has the potential to explain these observations. The key clue is that financial market returns can accurately be described by the stochastic process (1) with a quartic potential. Such a quartic potential typically occurs in certain physical models near second-order phase transitions, such as the Ising model or the lattice gas model.

This has led us to propose a lattice gas model of financial markets. The lattice represents the social network of investors, who correspond to its vertices. The microscopic degrees of freedom sitting at these vertices represent the investors' asset allocations. They can either be modeled by gas molecules that correspond to the shares of an asset, or – equivalently – by spins that denote whether the investor is over- or underweight in the asset, relatively to its CAPM weight. For N different assets, the spins have N components.

We have identified the order parameter, corresponding to the magnetization in the spin model, as the deviation of the asset price from its long-term value. If we assume that neighboring spins tend to align, because investors tend to follow their "friends" in the network, the interaction of the spins naturally creates the macroscopic phenomena of trends and reversion of the asset price.

In addition, we have argued that efficient markets drive themselves to the critical point, at which a second-order phase transition occurs. This assumes that investors quickly exploit and thereby eliminate any obvious market inefficiencies either by momentum strategies, such as trend-following, or by reversion strategies, such as statistical arbitrage. This argument is supported by the observation that the persistence of trends, corresponding to the quadratic term in the quartic potential, has gradually disappeared over the decades [11].

There is a wide variety of stochastic processes such as (1) that have been studied in finance [45]. Different assets are modeled by different processes, which are often complex and require the calibration of many parameters to fit the historical data. What is currently missing is a unified theory that determines from first principles, which stochastic process with which parameters should be used to model a given asset. Our approach suggests such a principle: these processes must arise from the dimensional reduction of a higher-dimensional system near its critical point, at which it is described by a renormalizable (self-consistent) quantum field theory. This key criterium of renormalizability leaves us with only a small set of universal field theories, and therefore, allowed stochastic processes for financial markets.

More precisely, at this critical point, the system is characterized by a divergent correlation length and by critical coefficients, which are universal in the sense that – in the case of a trivial network topology – they depend only on two parameters: the fractal dimension D of the network, and the symmetry group $O(N)$. We have used the methodology of field theory and the renormalization group to relate these critical exponents to observable autocorrelations and Hurst exponents of financial market returns. In particular, the second Hurst exponent follows from the anomalous scaling dimension η and the dynamic critical exponent z (29):

$$H_2 = \frac{\kappa}{2} \quad \text{with} \quad \kappa = \frac{2 - \eta}{z}.$$

This allows us to measure κ , and from its value infer the fractal dimension of the presumed underlying social network of investors.

As an illustrative example, we have applied this procedure to the simplest case: “Model A” on a network with the topology of D -dimensional Euclidean space. Consistency with observation then yields the following “predictions” for the fractal dimension D of the network and for the correlation time τ in financial markets (27):

$$D \approx 3, \quad \tau \geq \text{length of the economic cycle.}$$

This can in principle explain the scaling behaviour observed in financial markets at short and very long time scales, and the small magnitude of market auto-correlations. Encouragingly, D is within the range of fractal dimensions typically found in social networks [44].

On the other hand, by comparing the two-point function of this model with the observed auto-correlations of financial market returns at intermediate time scales, we have also seen that the universal properties of model A on a network with Euclidean topology do not explain the trends observed in real markets over time horizons from 1 month to 1 year. This leaves us with the task of refining the model, such that it can explain the observed interplay of trends and reversion also on intermediate time scales. Some natural refinements are:

1. *Beyond “Model A”*: one may try to model the asset allocations of investors by other degrees of freedom than spins, e.g., by a binary fluid in the case of two assets, leading to different universality classes. More generally, the models “A–J” of critical dynamics and their generalizations provide a rich laboratory of systems [13,14]. The universal properties of some of them indeed include two-point functions that switch between positive and negative auto-correlations, as observed in Fig. 6 (left).
2. *Nontrivial network topologies*: we do expect the social network of investors to be much more connected than a hypercubic lattice, for which the average distance between two points is very long. Candidates for social networks include scale-free networks [34], which are characterized by a “degree distribution” $p(k) \sim k^{-\gamma}$ of nodes with k neighbors. It is currently not clear, which of these networks can be part of renormalizable field theories (except for the planar diagrams of large- N field theory [35], on which the renormalization group flow has been studied [46]). An important next step is thus to classify their universality classes. Apart from the dimension D and the symmetry group $O(N)$, they may also depend on parameters such as γ .
3. *Approach to the critical point*: it is also conceivable that the presumed underlying social network of investors is not exactly at the fixed point $g = g^*$. In this case, we may be able to understand the oscillations of the two-point function in Fig. 6 (left) as non-universal phenomena that arise during the *approach* to the critical point.

In either case, it will be fascinating to search for the right candidate in analogy with the example of Section 4, by using the different models of critical dynamics to predict the scaling behavior of market returns, and by then ruling out those models that are inconsistent with empirical observation. Further work in this direction is underway.

CRediT authorship contribution statement

Christof Schmidhuber: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I would like to thank Uwe Täuber for very helpful remarks and explanations on dynamic scaling, and Wolfgang Breyman for encouragement, valuable discussions, and for pointing out the relation to self-organized criticality. I also thank Didier Sornette for an inspiring conversation and Michel Dacorogna for comments on the manuscript. This research is supported by grant no. CRSK-2 190659 from the Swiss National Science Foundation.

Appendix. Brief review of empirical results

In [11], we have empirically observed the interplay of trends and reversion in financial markets. The analysis is based on historical daily log-returns of a set of 24 futures contracts [47]. This set is diversified across four asset classes (equity indices, interest rates, currencies, commodities), three regions (Americas, Europe, Asia) and three commodity sectors (energy, metals, agriculture). We used futures returns, instead of the underlying market returns, because futures returns are guaranteed to be marked-to-market daily. This is a brief summary of the approach and results. For detailed explanations and references, we refer to [11].

A.1. Trend definition

For each of the 24 markets i , we consider 30 years of daily price data $P_i(t)$ from 1990 to 2019. We define normalized daily log-returns $R_i(t)$:

$$R_i(t) = \frac{r_i(t)}{\sigma_i}, \quad r_i(t) = \ln \frac{P_i(t)}{P_i(t-1)}, \quad \sigma_i^2 = \text{var}(r_i), \quad \mu_i = \text{mean}(r_i), \quad (30)$$

where the long-term daily risk premium μ_i and the long-term daily standard deviation σ_i of a market i are measured over the whole 30-year period. For each market and on each day, we compute 10 different trend strengths on 10 different time scales T_k with

$$T_k = 2^k \text{ business days with } k \in \{1, 2, 3, \dots, 10\}$$

This represents time horizons of approximately 2 days to 4 years. At a given point in time, a market may well be in an up-trend on one scale, but on a down-trend on another scale. For a given horizon T , we define the trend strength $\phi_{i,T}(t)$ of a given market i at the close of trading of day $t \in Z$ as a weighted average of past daily log-returns of that market (i.e., on or before day t), in excess of the long-term risk premium in that market:

$$\phi_{i,T}(t) = \sum_{n=0}^{\infty} w_T(n) \cdot \hat{R}_i(t-n) \quad \text{with} \quad \hat{R}_i(t-n) = R_i(t-n) - \frac{\mu_i}{\sigma_i},$$

where $w_T(n)$ is a weight function for the time scale T (to avoid any biases, in the out-of-sample cross-validation of our results, the risk premia were estimated only from the training samples, excluding the validation samples).

The weight function $w_T(n)$ is always normalized such that the trend strength $\phi_{i,T}$ has standard deviation 1. Assuming that market returns on different days are independent from each other (which is true to high accuracy), this implies:

$$\sum_{n=0}^{\infty} w_T^2(n) = 1.$$

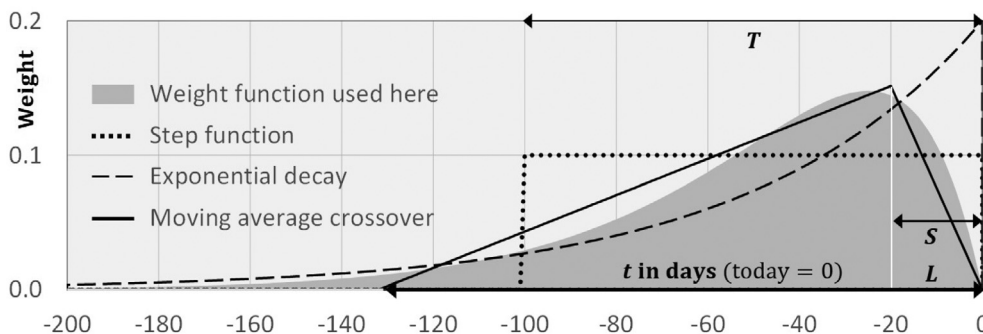


Fig. 9. Our trend strength is defined as a weighted sum of past log-returns. The gray area shows the weight function used in this paper, compared with three standard alternatives. All four weight functions shown here have the same average lookback period.

Table 2
Regression with linear and cubic terms.

Coefficient	Value	Error	t-statistics
a	1.33%	$\pm 0.41\%$	3.3
b	1.29%	$\pm 0.43\%$	3.0
c	-0.62 %	$\pm 0.23\%$	2.7
R-squared	Single time scales	Aggregated across time scales	
R^2	1.31 bp	4.91 bp	
R_{adj}^2	1.03 bp	3.98 bp	

With this normalization, $\phi_{i,T}$ can be regarded as the statistical significance of the trend. E.g., $\phi_{i,T} = 2$ represents a highly significant up-trend, while $\phi_{i,T} = -0.5$ represents a weakly significant down-trend. This normalization makes all trend strengths comparable with each other, and thus allows us to aggregate across different markets and time scales.

The simplest weight function would be a *step function* (Fig. 1, dotted line). In this case, the trend strength $\phi_{i,T}$ is just proportional to the average log-return over the past T days. A wedge-like weight function (Fig. 1, solid line) would correspond to a *moving log-price average crossover*. The advantages and disadvantages of different weight functions are discussed in [11]. For our study, we used two such definitions of the trend strength ψ_T and ϕ_T :

$$\psi_T : \tilde{w}_T(n) = M_T e^{-2n/T} \quad \text{with normalization} \quad M_T = \sqrt{1 - e^{-4/T}}, \quad (31)$$

$$\phi_T : w_T(n) = N_T \cdot (n + 1) \cdot \exp\left(-\frac{2n}{T}\right) \quad \text{with} \quad N_T = \frac{(1 - e^{-4/T})^2}{\sqrt{1 - e^{-8/T}}}. \quad (32)$$

The ‘‘average lookback period’’ of these trend strengths, i.e., the expectation value $E[n + 1]$ of the number of days we look back (where ‘‘today’’, i.e. $n = 0$, counts as a 1-day lookback), is $T/2$ for ψ and T for ϕ . Actually, for a given horizon T , the precise definition of the trend strengths does not matter much – they all yield quite similar results in our regression analysis, as long as the weight function rises gradually, decays gradually, and the average lookback period is the same. However, (31), (32) were best-suited for our study, because they

- have only a single free parameter, the horizon T (to avoid overfitting historical data)
- can be computed recursively (which yields a Langevin equation for continuous time)

(6) also has the advantage of reducing trading costs. It was originally introduced by the author in 2008 at Sydex Capital Management, and has been used to replicate Managed Futures indices as part of a UCITS fund from 2010–2014.

A.2. Prediction of next-day return

The results shown in Fig. 1 are aggregated across 30 years of daily returns for each market, across all 24 markets, and across different time scales. This aggregation was necessary in order to achieve results that are statistically highly significant. In particular, for shorter time windows, or without aggregating across markets, the observed patterns cannot be detected reliably. This may explain why they have not been reported already a decade ago.

Guided by the patterns observed in the graphs in Fig. 1, we have performed a nonlinear regression of the next-day normalized log-return $R_i(t + 1)$ (30) against a polynomial of the current trend strength $\phi_{i,k}(t)$ (32) across all markets i and time scales $T = 2^k$:

$$R_i(t + 1) = a + b \cdot \phi_{i,k}(t) + c \cdot \phi_{i,k}^3(t) + \epsilon_i(t + 1), \quad (33)$$

where ϵ represents random noise, and a measures the average normalized risk premium μ_i/σ_i across all assets. These risk premia have not been the focus of [11]. Instead, we have concentrated on determining the coefficients b, c , which measure how the expected return of an asset varies in time. We interpret b as the persistence of trends, and c as the strength of trend reversion. d and the higher-order terms were not statistically significant. The results of the overall regression analysis are shown in Table 2.

The coefficients a, b, c and the R-squared’s are very small, but they *must* be so by nature: if the patterns were much stronger, traders would immediately see and exploit them and achieve stellar Sharpe ratios. In fact, it is shown in [11] that the estimated coefficients are just strong enough to explain the Sharpe ratios of order 1 that the Managed Futures (‘‘CTA’’) industry has historically achieved.

[11] also reports further analyzes, in which b and c are allowed to have a simple polynomial dependence on the log of the trend’s time horizon $k \equiv \log_2 T$:

$$R_i(t + 1) = \alpha_i + b(k) \cdot \phi_{i,k}(t) + c(k) \cdot \phi_{i,k}^3(t) + \epsilon_i(t + 1). \quad (34)$$

Within the limits of statistical significance, we find that b and c are universal, i.e., the same for all assets. While the strength of reversion c seems to be approximately constant, we find that the persistence of trends b depends on the time

horizon of the trend. Within the range of time scales considered here, and averaged over the past 30 years, it can be approximated by a parabolic function of the log of the time scale:

$$c \sim -0.6\% \quad , \quad b(k) \sim 2.0\% \cdot \left(1 - \frac{(k - k_0)^2}{\Delta k^2}\right) \quad \text{with } k_0 \sim 6 \quad , \quad \Delta k \sim 5. \quad (35)$$

This implies that trends may only be stable if the log of the time horizon is within the range $k_0 \pm \Delta k$, corresponding to time scales from a few days to several years. The parameters k_0 and Δk are also universal. By bootstrapping and cross-validation, we have found that all four parameters in (35) are statistically highly significant out-of-sample.

A.3. Statistical analysis

Given the small values of the coefficients, and the huge amount of noise in financial market data, we had to pay great attention to a careful analysis of the statistical significance of the results. In particular, since market returns cannot be assumed to be independent, identically distributed normal variables, we cannot trust the usual estimates of the t-statistics, adjusted R-squared's, and other statistics. Instead, the test statistics shown in Table 2 have been measured empirically as follows:

- The standard errors of the coefficients and their t-statistics were computed by bootstrapping, using 5000 bootstrapping samples. Regression on each new sample of days yielded the distribution of the 5000 regression coefficients b and c shown in Fig. 1 (upper right), from which the estimation errors in Table 2 were read off. These actual errors are 3–4 times as large as the errors that standard regression tools reported. Nevertheless, the regression coefficients are still statistically highly significant.
- The adjusted R-squared was computed by 15-fold cross validation. The resulting out-of-sample R-squared is reported as R_{adj}^2 in Table 2. The actual out-of-sample correction $R^2 - R_{adj}^2$ is about 3 times as big as what standard regression tools reported. Nevertheless, the remaining R_{adj}^2 is still highly significant despite being very small, and in fact agrees with the one that is achieved in real trading.
- Table 2 also reports R^2 and R_{adj}^2 “aggregated across time scales”. Those are based on using the equally-weighted mean of the 10 trend strengths on each day to predict the next-day return for each market. I.e., we combine the 10 different trend factors into a single one, which has a higher predictive power than each single factor by itself.

The regression results of Table 2 and further regression analyzes in [11] confirm and quantify our conclusions from Fig. 1. In particular, we see that the values of b and c – although very small – are statistically highly significant, despite the fact that market returns are neither normally distributed, nor independent, nor identically distributed.

References

- [1] For a popular review, see, e.g., M. Covel, *The Complete TurtleTrader: The Legend, the Lessons, the Results*, Collins, 2007.
- [2] D.M. Cutler, J.M. Poterba, L.H. Summers, Speculative dynamics, *Rev. Econom. Stud.* 58 (3) (1991) 529–546.
- [3] W.L. Silber, Technical trading: when it works and when it doesn't, *J. Derivatives* 1 (3) (1994) 39–44.
- [4] W. Fung, D.A. Hsieh, The risk in hedge fund strategies: Theory and evidence from trend followers, *Rev. Financ. Stud.* 14 (2) (2001) 313–341.
- [5] L. Jaeger, C. Wagner, Factor modeling and benchmarking of hedge funds: can passive investments in hedge fund strategies deliver? *J. Alternat. Invest.* 8 (3) (2005) 9–36.
- [6] J. Miffre, G. Rallis, Momentum strategies in commodity futures markets, *J. Bank. Financ.* 31 (6) (2007) 1863–1886.
- [7] T.J. Moskowitz, Y.H. Ooi, L.H. Pedersen, Time series momentum, *J. Financ. Econ.* 104 (2) (2012) 228–250.
- [8] Y. Lemprière, C. Deremble, P. Seager, M. Potters, J.-P. Bouchaud, Two centuries of trend following, *J. Invest. Strateg.* 3 (3) (2014) 41–61.
- [9] B. Hurst, Y.H. Ooi, L.H. Pedersen, A century of evidence on trend-following investing, *J. Portfolio Manage.* 44 (1) (2017) 15–29.
- [10] J. Baz, N. Granger, C.R. Harvey, N. Le Roux, S. Rattray, Dissecting investment strategies in the cross section and time series, 2015, Available at SSRN 2695101.
- [11] Christof Schmidhuber, Trends, reversion, and critical phenomena in financial markets, *Physica A* 566 (2021) 125642.
- [12] Jean-Philippe Bouchaud, et al., Black was right: Price is within a factor 2 of value, 2017, Available at SSRN 3070850.
- [13] P.C. Hohenberg, B.I. Halperin, Theory of dynamic critical phenomena, *Rev. Modern Phys.* 49 (3) (1977) 435.
- [14] Uwe C. Täuber, *Critical Dynamics: A Field Theory Approach To Equilibrium and Non-Equilibrium Scaling Behavior*, Cambridge University Press, 2014.
- [15] E. Ising, Beitrag zur theorie des ferromagnetismus, *Z. Phys. Band* 31 (1925).
- [16] Ulrich A. Müller, et al., Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis, *J. Bank. Finance* 14 (6) (1990).
- [17] Ghashghaie, Shoaie, et al., Turbulent cascades in foreign exchange markets, *Nature* 381 (6585) (1996) 767–770.
- [18] R.N. Mantegna, H.E. Stanley, *Introduction To Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, 1999.
- [19] Tiziana Di Matteo, Multi-scaling in finance, *Quant. Finance* 7 (1) (2007) 21–36.
- [20] Kwapien, Jarosław, Stanisław. Drozd, Physical approach to complex systems, *Phys. Rep.* 515 (3–4) (2012) 115–226.
- [21] T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, *Nature* 297 (1999) 498–500.
- [22] J.D. Farmer, S. Joshi, The price dynamics of common trading strategies, *J. Econ. Behav. Organ.* 49 (2) (2002) 149–171.
- [23] J.P. Bouchaud, R. Cont, A langevin approach to stock market fluctuations and crashes, *Eur. Phys. J. B* 6 (4) (1998).
- [24] D. Sornette, *Physics and financial economics (1776–2014): puzzles, Ising and agent-based models*, *Rep. Progr. Phys.* 77 (6) (2014) 062001.
- [25] Benoit Mandelbrot, The variation of certain speculative prices, *J. Bus.* (ISSN: 0021-9398) 36 (4) (1963) 394, <http://dx.doi.org/10.1086/294632>.
- [26] Samuelson Paul, Proof that properly anticipated prices fluctuate randomly, *Indust. Manage. Rev.* 6 (1965) 41–49.
- [27] Fama Eugene, Efficient capital markets: A review of theory and empirical work, *J. Finance* 25 (2) (1970) 383, <http://dx.doi.org/10.2307/2325486>. JSTOR2325486.

- [28] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality, *Phys. Rev. A* 38 (1) (1988) 364.
- [29] K. Ide, D. Sornette, Oscillatory finite-time singularities in finance, population and rupture, *Physica A* 307 (1–2) (2002) 63–106.
- [30] M. Takayasu, T. Mizuno, H. Takayasu, Potential force observed in market dynamics, *Physica A* 370 (1) (2006) 91–97.
- [31] K. Watanabe, H. Takayasu, M. Takayasu, Random walker in temporally deforming higher-order potential forces observed in a financial crisis, *Phys. Rev. E* 80 (5) (2009) 056110.
- [32] Halperin, Igor, Matthew. Dixon, Quantum equilibrium-disequilibrium: Asset price dynamics, symmetry breaking, and defaults as dissipative instantons, *Physica A* 537 (2020) 122187.
- [33] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* 393 (1998).
- [34] A.L. Barabási, R. Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512.
- [35] E. Brézin, C. Itzykson, G. Parisi, J.B. Zuber, Planar diagrams, in: *The Large N Expansion in Quantum Field Theory and Statistical Physics*, 1993, pp. 567–583.
- [36] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, G. Caldarelli, The statistical physics of real-world networks, *Nat. Rev. Phys.* 1 (1) (2019).
- [37] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, Oxford U. Press, 1989.
- [38] Hagen Kleinert, Schulte-Frohlinde Verena, *Critical Properties of Phi4-Theories*, World Scientific, 2001.
- [39] K.G. Wilson, The renormalization group and the ϵ -expansion, *Phys. Rep.* 12 (1974).
- [40] Yu Holovatch, Critical exponents of ising-like systems in general dimensions, *Theoret. Math. Phys.* 96 (3) (1993) 1099–1109.
- [41] Hasenbusch Martin, Dynamic critical exponent z of the three-dimensional Ising universality class: Monte Carlo simulations of the improved Blume–Capel model, *Phys. Rev. E* 101 (2) (2020) 022126.
- [42] Barabási Albert-László, Tamás Vicsek, Multifractality of self-affine fractals, *Phys. Rev. A* 44 (4) (1991) 2730.
- [43] Benoit B. Mandelbrot, *Fractals and Scaling in Finance*, Springer, New York, 1997.
- [44] Song Chaoming, Shlomo Havlin, Hernan A. Makse, Self-similarity of complex networks, *Nature* 433 (7024) (2005) 392–395.
- [45] For a review of stochastic processes in Finance, see, e.g., Steven E. Shreve, *Stochastic Calculus for Finance I and II*, Springer-Verlag New York, 2004.
- [46] Christof Schmidhuber, On water, steam, and string theory, *Amer. J. Phys.* 65 (11) (1997) 1042–1050, and references therein.
- [47] Christof Schmidhuber, Data for: Trends, reversion, and critical phenomena in financial markets, Mendeley Data V1 (2020) <http://dx.doi.org/10.17632/v73nzd7rt.1>.