Forecasting stock market returns over multiple time horizons

Kroujiline, Dimitri and Gusev, Maxim and Ushanov, Dmitry and Sharov, Sergey V. and Govorkov, Boris

18 August 2015

Forecasting stock market returns over multiple time horizons

Dimitri Kroujiline*[1], Maxim Gusev[2], Dmitry Ushanov[3], Sergey V. Sharov[4] and Boris Govorkov[2]

Abstract

In this paper we seek to demonstrate the predictability of stock market returns and explain the nature of this return predictability. To this end, we further develop the news-driven analytic model of the stock market derived in Gusev et al. (2015). This enables us to capture market dynamics at various timescales and shed light on mechanisms underlying certain market behaviors such as transitions between bull- and bear markets and the self-similar behavior of price changes. We investigate the model and show that the market is nearly efficient on timescales shorter than one day, adjusting quickly to incoming news, but is inefficient on longer timescales, where news may have a long-lasting nonlinear impact on dynamics attributable to a feedback mechanism acting over these horizons. Using the model, we design the prototypes of algorithmic strategies that utilize news flow, quantified and measured, as the only input to trade on market return forecasts over multiple horizons, from days to months. The backtested results suggest that the return is predictable to the extent that successful trading strategies can be constructed to harness this predictability.

Keywords: stock market dynamics, return predictability, price feedback, market efficiency, news analytics, sentiment evolution, agent-based modeling, Ising, dynamical systems, synchronization, self-similar behavior, regime transitions, news-based strategies, algorithmic trading.

[1] LGT Capital Partners, Pfäffikon, Switzerland, dimitri.kroujiline@lgt.com (*contact author). [2] IBC Quantitative Strategies, Tärnaby, Sweden. [3] Moscow State University, Department of Mechanics and Mathematics, Moscow, Russia. [4] Lobachevsky State University, Advanced School of General & Applied Physics, N. Novgorod, Russia.

Introduction

The integration of news analytics into trading strategies continues to be at the R&D forefront in the investment industry. The efforts have mainly focused on constructing early indicators for a change in investor sentiment to enable a trader to act ahead of the majority of investors. The potential for success is therefore reliant on the speed and precision with which information retrieval and text parsing algorithms process vast amounts of data and recognize, in the incoming flow of news, the events that may move prices substantially.

As prices generally tend to quickly reflect new information, the objective of news-based trading is to produce short-term (intraday) strategies. On short timescales prices may be assumed to react to news linearly. This linearization is helpful because a trade's sign would then depend only on the sign of sentiment assigned to a news item and its size on price sensitivity – each still being a formidable task. As with any other short-term trading strategy, the downside here is a limited capacity and high, turnover-driven costs. Furthermore, should successful strategies eventually be developed – so far results have been mixed to the best of our knowledge – competition in this segment will spur an "arms race" of speed, leading to increasingly fast price reaction to news, exacerbating capacity restrictions and reducing profit margins.[1]

The above-described approach is based on a premise that financial markets need a finite amount of time to digest news. In other words, return prediction using news analytics relies on the market's informational inefficiency in the time interval where prices adjust to new information. Does it then follow that news analytics are necessarily useless for trading over longer horizons?

[1] For example, competition among high frequency trading firms has increased the trade execution speed from roughly 100 milliseconds to 10 microseconds over the last decade.

At first glance, the answer seems to be a firm "yes". Indeed, while it appears reasonable to suppose that a market can "remember", in terms of its reaction, previous events on the order of minutes and hours, this same supposition sounds absurd when considering periods that span days, weeks or months. Yet, there is certain empirical evidence of long-term return predictability.[2]

It might be useful to tackle the problem from a different angle and consider whether there exist any logical possibilities for long-term predictability. Incidentally, such a possibility does exist, provided the mechanism of price formation over long time horizons is different from that involved on short timescales. Let us hypothesize how it may operate.

Efficient markets ensure that new information is manifested in a change in market price shortly following its release. However, this price change can also be an important event on its own that will

[2] There is a large body of research that examines, typically applying regression methods, the predictive power of observable variables, such as the dividend yield and many others; see, for example, Fama and French (1988, 1989), Campbell and Shiller (1988a,b), Baker and Wurgler (2000), Campbell and Thompson (2008), Cochrane (2008). However, the evidence for return prediction remains inconclusive: e.g. Ferson et al. (2003, 2008), Goyal and Welch (2003, 2008). The model of stock market dynamics that we develop here is fundamentally nonlinear, indicating that causal relations among the variables are substantially more complex than regression dependence. It follows that the standard approach to return prediction, based on regression methods, may be ill-suited to capture this predictability; e.g. Novy-Marx (2014) vividly pointed out this limitation by extending stock market predictive regressions to a number of rather implausible variables, such as sunspot activity and planetary motion among others. Gusev et al. (2015) proposed an alternative approach that combines, analogous to weather forecasting, theoretical models with empirical data. The present work applies this same approach to demonstrate that stock market returns can be predicted in an economically significant manner.

draw media response. This can incite subsequent price changes, causing in turn further news releases and so forth. Thus, the original event may trigger a "ripple effect" of the interlinked price changes and news releases, unfolding over an extended time period. It follows that news can have a long-lasting impact in the market via the above-described feedback mechanism,[3] which is absent on short timescales as we will see later.

Thus, from a news-price system with no feedback, likely valid on the order of a day or less, we move toward a mutually-coupled news-price system, operating over longer horizons. Consequently, we must apply a different framework for return prediction. While the short-term prediction requires fast detection of the news releases that may provoke material changes in price, the long-term prediction, on the contrary, can only be based on the regularity of the system's behavior. In other words, we must develop a dynamic model that correctly describes interaction between news and price. Then, provided the model admits non-stochastic solutions, it would be sufficient to know the market position in the news-price reference frame to forecast return by following the market evolution path provided by the model.

Gusev et al. (2015) proposed and investigated, theoretically and empirically, such a model of stock market dynamics. This model describes dynamics in terms of the interaction between prices, opinions and information. It is formulated as an Ising-family agent-based model[4] with two types of

[3] The idea that the observations of price changes may generate a feedback loop that significantly affects market dynamics is not new (see Shiller, 2003). However, its application for return forecasting, which is the subject of the present work, is nontrivial due to nonlinear behaviors induced by it.

[4] A family of models, named after Ising (1925), developed originally to explain the phenomenon of ferromagnetism via the interaction of discrete atomic spins in an external magnetic field and later broadly

interacting agents: investors, who invest or divest according to their opinions, and analysts, who interpret news, form opinions and channel them to investors.[5] To derive the model equations in analytic form and facilitate its study, it was assumed that investors made up a homogeneous group in which any two market participants interacted identically. Despite this and other simplifications, the model reproduced the price path and return distribution of the S&P 500 Index within reasonable tolerance. Based on these results, the authors suggested that stock market returns are predictable, but did not conduct tests of this predictability.

The objective of the present work is to elaborate on the ideas from Gusev et al. (2015) to demonstrate that returns can be predicted over time horizons longer than one day. To accomplish this, we introduce heterogeneity into the model by replacing homogeneous investors with groups of investors that have different investment horizons. This enables us to extract characteristic dynamics on different timescales and produce market forecasts of multiple time horizons, upon which we design the prototypes of trading strategies.

---

applied to study problems in social and economic dynamics (see the reviews by Castellano et al., 2009; Lux, 2009; Slanina, 2014; Sornette, 2014). We take note of two recent works that share common ground with Gusev et al. (2015). First, Franke (2014), referencing Lux's (1995) analytic stock market model, studied a generic sentiment-driven economic model with feedback, which has some features similar to those found in Gusev et al. (2015). Second, Carro et al. (2015) investigated the influence of exogenous information on endogenous sentiment dynamics in the stock market, which is also a central theme in Gusev et al. (2015).

[5] This approach contrasts with that of the established agent-based financial models, where market dynamics are sought to emerge, primarily, through the interaction among agents pursuing different trading strategies, such as the influential work by Lux and Marchesi (1999) among many others.

The paper is organized as follows. Section 1 describes the news-driven market model with homogeneous investors and develops the model with heterogeneous investors. Section 2 studies the heterogeneous model analytically, numerically and empirically. Additionally, Section 2.3 shows that the market is efficient on short timescales. Section 3 designs and backtests the prototypes of trading strategies. Section 4 further discusses the nature of return predictability. Section 5 provides a summary of conclusions.

1. Models

This section introduces the model with homogeneous investors, developed in Gusev et al. (2015), and using it as a starting point, derives the model with heterogeneous investors that we will apply for market forecasting.

1.1 Homogeneous model

The model of stock market dynamics in Gusev et al. (2015) is formulated as a dynamical system governing the evolution of three independent variables: market price p, investor sentiment s and information flow h. It was obtained by defining, based on observed behaviors, interactions among agents at a micro level and applying methods from statistical mechanics to produce dynamic equations for averaged variables at a macro level. Before exploring the equations, it may be helpful to explain the proper context in which sentiment and information are used in the model.

Investor sentiment is defined as a summary view on future market performance, averaged across the investment community, and is determined as the ratio of the number of investors who opine that the market will rise minus the number of investors who opine that the market will fall over the total number of investors. Thus, sentiment s can vary between -1 and 1. By this definition sentiment s encompasses all types of opinions, irrespective of whether an opinion has been formed rationally or irrationally.

Information flow considered in the model as relevant comprises publicly expressed opinions about the direction of anticipated market movement. It is quantified similarly to sentiment as the ratio of the number of news items with positive expectations minus the number of news items with negative expectations over the total number of news items concerning the market. Like sentiment, information h is bounded between -1 and 1. The fact that h can be readily measured allows the model to be empirically verified.[6,7]

---

[6] Gusev et al. (2015) proposed that information h, referred to as "direct information", can effectively influence investors' view. That paper provided a rule-based parsing methodology for measuring h based on marketing research techniques, essentially treating each news item as if it were a "sales pitch" aimed at investors to buy or sell the market. In practice, as news about current and recent market movements can also influence investors, such direct information was included in the measurement of h, along with information concerning anticipated market movement.

[7] Extensive research has been done on empirical measures of sentiment – which include indices based on periodic surveys of investor opinion, various proxies such as trading volume, call vs. put contracts and others, and applications of machine learning and rule-based techniques for parsing financial news and social media content – and their correlation with price movement (e.g. Antweiler and Frank, 2004; Brown and Cliff, 2004; Baker and Wurgler, 2007; Das and Chen, 2007; Tetlock, 2007; Loughran and McDonald, 2011; Lux, 2011; Da et al., 2014). Alternatively, Gusev et al. (2015) modeled empirical sentiment $s(h)$ and price $p(h)$ from measured h, using the homogeneous model described in this section. The present paper adopts this same approach, but applies the heterogeneous model developed in the next section.

The model is described by the differential equations:[8]

$$\dot{p} = a_1\dot{s} + a_2(s - s_*), \tag{1a}$$

$$\tau_s\dot{s} = -s + \tanh(\beta_1 s + \beta_2 h), \tag{1b}$$

$$\tau_h\dot{h} = -h + \tanh(\kappa_1\dot{p} + \kappa_2\xi_t). \tag{1c}$$

The first equation, derived by observing that investors tend to act on their opinions differently over short and long horizons, defines a phenomenological relation between the change in log price and investor sentiment. This equation states that price changes proportionally, first, to the change in sentiment and, second, to the deviation of sentiment from a certain reference level s_*. The former is the main source of short-term price variation, while the latter determines leading behavior over long-term horizons.

The second and third equations were derived together as a single system, using methods from statistical mechanics. The second equation describes the change in sentiment due to the impact of information flow on investors via the term $\beta_2 h$ and the interaction among investors via the term $\beta_1 s$, where τ_s is the characteristic time of sentiment variation and β_1 determines the relative importance of the herding behavior and the random behavior of investors. Information flow acts as

[8] Gusev et al. (2015) (Eq. 13, Fig. 12). The dot denotes the derivative with respect to time. Parameters $a_1, a_2, \beta_1, \beta_2, \kappa_1, \kappa_2, \tau_s, \tau_h$ are positive, while s_* can take any sign. The parameter values, estimated using the empirical data, are provided in Table 1 of that same paper. We note that equation (1b), with h as an exogenous variable, was obtained by Suzuki and Kubo (1968) in the context of a purely statistical mechanics problem.

a force that moves sentiment away from equilibrium. If it were to cease, sentiment would come to rest at a nonzero value for $\beta_1 > 1$ (ordered state) or at zero for $\beta_1 < 1$ (disordered state).

The third equation states that the change in information flow is caused by exogenous news ξ_t and news about price changes $\dot{p}$, with τ_h being the characteristic response time.[9]

Equations (1) form a three-dimensional nonlinear dynamical system. Each point in the phase space (h, s, p) represents a unique market state and each solution $\big(h(t), s(t), p(t)\big)$ represents a phase trajectory of market evolution. This evolution is driven by the flow of exogenous news ξ_t that induces random fluctuations of the phase trajectory and by the feedback mechanism h -> s -> p -> h that generates inertial dynamics, giving rise to deterministic behaviors. It follows that according to this model, market evolution may contain deterministic regimes and thus be potentially predictable.

It is important to note that equations (1) were obtained under a simplifying assumption of the all-to-all interaction pattern among agents.[10] In reality, interaction among investors is hardly so simple. The utilization of more sophisticated patterns of interaction in the Ising-type models is known to cause the emergence of heterogeneous structures – i.e. in the present case, clusters of investors with co-aligned sentiments. Because the size of a cluster determines its reaction time to incoming information, this heterogeneity can generate diverse dynamics involving interactions on many timescales. It follows that model (1) should be regarded as a coarse-grained approximation

[9] We note that this form of the equation neglects direct interaction between the agents, omitting the terms proportional to h and s in the argument of the hyperbolic tangent (Gusev et al., 2015: Eq. 12c).

[10] The all-to-all interaction mode is the leading-order approximation for a general interaction topology in this model, which makes it a sensible first step for studying this problem.

that determines the average investor behavior evolving within a single timeframe τ_s[11]. Thus, although this model provides certain valuable insights into market dynamics, it is unlikely to be sufficiently realistic for market forecasting (especially over periods shorter than τ_s). We must refine this model to improve precision, which is the subject of the next section.

1.2 Heterogeneous model

We wish to replace the above-described framework, where each investor interacts with all other investors with the same strength, by a framework with a more realistic interaction pattern. Selecting such a pattern in the form of rules applicable to individual agents is a hard problem to solve. This is because there exist many plausible choices for interactions at the micro-level and the model's statistical properties will be sensitive to these choices.[12] Additionally, it would be difficult, if at all possible, to derive a closed-form dynamical system for the evolution of macro-level variables based on the interaction patterns more complex than all-to-all.

Instead, it may be more practical to account for investor heterogeneity via a phenomenological approach. As discussed above, we expect that realistic interaction patterns would produce clusters of investors, each characterized by a specific (as a function of size) response time to incoming

[11] Gusev et al. (2015) estimated τ_s to be around 25 business days.

[12] Cont and Bouchaud (2000) have addressed heterogeneity in opinion formation as a topological problem within the framework of percolation theory leading to the emergence of clusters of investors with shared sentiment. We note that following this work, a number of percolation models have been proposed that replicate some of distinctive market behaviors. However, as mentioned above, the results are sensitive to the choice of topology in a model and it is difficult to economically justify any one particular topology choice.

information, which we can assume proportional to the investment time horizon. It is therefore sensible to select the investment horizon as the attribute whereby variability is introduced into the model. Thus, we wish to modify model (1) by populating it with the investors that have various investment horizons.[13]

Next, we must make assumptions on how these investors would interact. Presumably, any organization, whether in the investment industry or elsewhere, should tend to connect best with its peers, owing to shared professional interests. For example, long-term investors, such as pension plans, have little in common with wealth management companies oriented toward mid-term performance and even less so with the day-trading community. Each of these investment industry segments maintains its own professional publications, conferences, seminars, awards and other platforms for discourse that promote networking and interaction. Therefore, we can suppose that interaction within the networks of peers, whom we propose to identify with respect to the investment horizon, is more efficient than across them.

We have thus arrived at a framework where investors with similar horizons form peer networks or groups within which they interact efficiently, but have little interaction externally, at the same time being impacted in equal measure by information flow h. In the limiting case, we can assume, first, that the interaction pattern within each peer group is all-to-all and, second, that there is no

[13] We presume that the average memory time span and the average holding period are proportional to the investment horizon and use these terms interchangeably, depending on the context. We note that the agent-based model developed in a series of publications by Levy, Levy and Solomon (see Levy et al., 2000) includes investors with different memory time spans.

interaction across these groups. This enables us to apply equations (1b,c) to describe the market with N participating peer groups as follows:

$$\tau_i \dot{s}_i = -s_i + \tanh(\beta_1 s_i + \beta_2 h), \quad i = 1, 2, \dots, N, \tag{2a}$$

$$\tau_h \dot{h} = -h + \tanh(\kappa_1 \dot{p} + \kappa_2 \xi_t), \tag{2b}$$

where s_i is the average sentiment of the i-th group, which having been normalized by the total number of investors in the group takes values between -1 and 1, and τ_i is its investment horizon. Note that both the "herding" parameter β_1 and the constant β_2, which determines sensitivity to information flow, are assumed to be in the leading order uniform across the groups.

Next, we must adapt equation (1a). Its derivation in Gusev et al. (2015) was based on the observation that capital flows in the market are caused differently on different timescales, namely by ds at $t \ll \tau_s$ and by $s - s_*$ at $t \gg \tau_s$. Let us consider the i-th network of investors, characterized by the horizon τ_i, on the timescales t where $t \ll \tau_i$. In this regime, ds_i would cause the capital flow $dc_i \sim c_i ds_i$, where c_i is the total capital managed within this network. To estimate c_i, we observe that market liquidity, which implies how much capital can be traded over τ_i without materially affecting market price, effectively imposes a cap on c_i. Assuming for simplicity that c_i is proportional to the average trading volume (as a proxy of liquidity) and noting that the average trading volume scales linearly with time, we arrive at $c_i \sim \tau_i$, $dc_i \sim \tau_i ds_i$ and $\dot{c}_i \sim \tau_i \dot{s}_i$ at $t \ll \tau_i$. Similarly, we obtain that $\dot{c}_i \sim \tau_i (s_i - s_{*i})$ at $t \gg \tau_i$.

We superpose the asymptotic relations $\dot{c}_i \sim \tau_i \dot{s}_i$ and $\dot{c}_i \sim \tau_i (s_i - s_{*i})$ and as the change in market price is determined by the net flow of capital into or out of the market, sum across all i to derive the following equation of price formation:

$$\dot{p} = a \sum \dot{c}_i = a_1 \left(\frac{\sum \tau_i \dot{s}_i}{\sum \tau_i} \right) + a_2 \left(\frac{\sum \tau_i (s_i - s_{*i})}{\sum \tau_i} \right) = a_1 \dot{s} + a_2 (s - s_*), \tag{3}$$

where the constants a_1 and a_2 are positive and the constant s_* can be of any sign. Equation (3) implies that the sentiments s_i of the investor groups with various τ_i and the overall sentiment s in the market are related by the formula:[14]

$$s = \frac{\sum \tau_i s_i}{\sum \tau_i}\,. \tag{4}$$

The heterogeneous market model is then given by the dynamical system:

$$\dot{p} = a_1 \dot{s} + a_2(s - s_*), \tag{5a}$$

$$\tau_i \dot{s}_i = -s_i + \tanh(\beta_1 s_i + \beta_2 h), \quad i = 1, 2, \dots, N, \tag{5b}$$

$$\tau_h \dot{h} = -h + \tanh(\kappa_1 \dot{p} + \kappa_2 \xi_t), \tag{5c}$$

where the aggregate sentiment s is defined by (4).

According to this dynamical system, investor groups with different investment horizons collectively form the aggregate sentiment that determines the market price, which in turn influences the information flow that acts on all groups participating in the market. Although there is no direct

[14] This relation, derived using the scaling property of the average trading volume, should be treated as an average relation applicable under normal market conditions or over extended time periods. In particular, the relation is not expected to hold during spikes in trading activity, such as, for example, those accompanying market crashes. Also, this relation may not be true for groups with very long investment horizons because on the corresponding timescales, effects due to the finite size of the market can affect the assumed linear relation between the investment horizon and the amount of investment capital. Nevertheless, as a first approximation, this relation will prove helpful for gaining insight into market dynamics on the relevant timescales.

interaction among the investor groups, each continues to impact the others by eventually contributing to the common information flow. Thus, the information flow plays a dual role: it is a force that impacts the sentiments of different investor groups and also a link through which these sentiments are mutually coupled.

2. Study of heterogeneous model

This section investigates the model with heterogeneous investors. Section 2.1 offers a preliminary analysis of the main effects expected in this model. Section 2.2 studies the model numerically, using both direct simulations and empirical data. Section 2.3 applies the model to demonstrate that the efficient market regime occurs on short timescales. The relevant technical details are in Appendix A.

2.1. Preliminary analysis: key effects

We can substitute $\dot{p}$ from (5a) into (5c) to obtain a self-contained dynamical system for s_i and h. When making this substitution, we follow Gusev et al. (2015) and approximate the second term on the right-hand side of (5a), which describes the evolution of price over long-term horizons, by a positive constant that represents the growth rate of the stock market.[15] We obtain the following equations:

$$\tau_i \dot{s}_i = -s_i + \tanh(\beta_1 s_i + \beta_2 h), \quad i = 1, 2, \dots, N, \qquad (6a)$$

[15] First, since $s \sim \sum \tau_i s_i$ and $|s_i| \leq 1$, the term $a_2(s - s_*)$ in (5a) is dominated by the sentiment of long-term investors, that is s_i corresponding to large τ_i. Second, equation (5b) implies that s_i varies by $O(1)$ over τ_i, i.e. the longer the investment horizon, the slower the sentiment variation (which is not unreasonable). Thus, $a_2(s - s_*)$ contributes to price development over the long term, e.g. months and years, which enables us to replace it in the leading order by a constant growth rate.

$$\tau_h \dot{h} = -h + \tanh(\gamma \dot{s} + \delta + \kappa \xi_t), \qquad (6b)$$

where $s = \frac{\sum \tau_i s_i}{\sum \tau_i}$ in accordance with (4), δ is a positive constant proportional to the stock market growth rate, $\gamma = \kappa_1 a_1$ and κ is κ_2 renamed.

Equations (6) define a dynamical system of $N + 1$ mutually-coupled nonlinear equations. As we will see later, this coupling leads to interesting, nontrivial behaviors in the system, such as the emergence of self-sustained oscillations and their synchronization.

To develop further intuition about this system, we express it in the following approximate form (see Appendix A):

$$\tau_i \ddot{s}_i + G(s_i)\dot{s}_i + \frac{dU(s_i)}{ds_i} = F_i^c + F^e, \quad i = 1, 2, \dots, N, \qquad (7)$$

where $U(s_i)$, $G(s_i)$, F_i^c and F^e are given by equations (A3), (A4), (A5) and (A6), respectively.

Equations (7) govern the motion of N particles (oscillators), each representing a network of investors characterized by horizon τ_i, driven by the applied force. Interestingly, as follows from (7), τ_i takes on a meaning of the mass of the i-th particle in the sense that particles with small τ_i ("light" particles) are more sensitive to any force than particles with large τ_i ("heavy" particles). We can say that particles with small τ_i have smaller inertia than particles with large τ_i.

The form of equations (7) allows their interpretation in terms of the particle's motion inside the potential well $U(s_i)$ in the presence of damping $-G\dot{s}_i$, driven by the external force F_i^c stemming from interaction between the particles (via common information flow h) and the external force F^e due to the flow of exogenous news.

As follows from (A3), the shape of the potential well $U(s_i)$ is identical for all particles. Further, it depends only on two parameters, β_1 and δ. Gusev et al. (2015) demonstrated for the case $N = 1$ that

the regime relevant for the stock market corresponds to $\beta_1 = 1.1$ and $\delta = 0.03$, which results in an asymmetric double-well shape of the potential (Figure 1).[16]

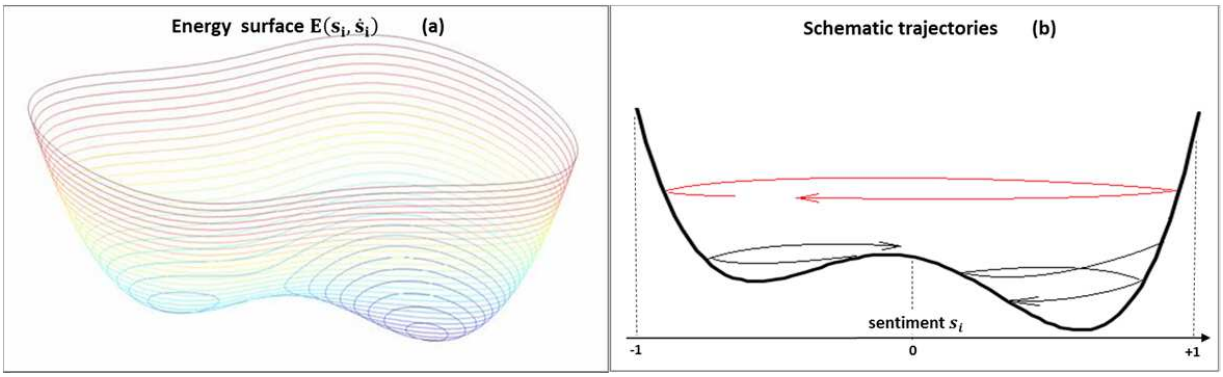


Figure 1: The profiles of the energy surface and the potential well, corresponding to $\beta_1 \gtrsim 1$ and $\delta \ll 1$. (a) The energy surface $E(s_i, \dot{s}_i)$ is shown as a function of s_i and $\dot{s}_i$ in the space $(s_i, \dot{s}_i, E)$. The colors indicate energy levels, from low (blue) to high (red). (b) The potential well $U(s_i)$ is shown as a function of s_i. The equilibrium point at the cusp of the potential is the unstable saddle, while the equilibrium points at its minima can be stable or unstable nodes or stable or unstable foci, depending on the value of feedback strength γ. The well where sentiment is positive is deeper than the well where sentiment is negative. Three typical trajectories are shown schematically in the well.

Figure (1a) depicts a surface corresponding to the kinetic and potential energy of the i-th particle as a function of s_i and $\dot{s}_i$: $E(s_i, \dot{s}_i) = \frac{\tau_i}{2}\dot{s}_i^{\,2} + U(s_i)$. Its cross-section by the plane (E, s_i) gives the shape of the potential well and by the plane $(E, \dot{s}_i)$ has the familiar parabolic profile of the

[16] As β_1 increases, the potential well undergoes a bifurcation from a single-well U-shape to a double-well W-shape at $\beta_1 = 1$. The potential is symmetric for $\delta = 0$; positive δ breaks the symmetry, making the part of the well where sentiment is positive deeper and the part where sentiment is negative more shallow.

kinetic energy. All trajectories lie on this energy surface. Figure (1b) depicts typical trajectories that we will discuss below.

If the impacts of damping ($-G\dot{s}_i$), interaction (F_i^c) and news (F^e) were negligible, a particle would oscillate periodically in response to the restoring force $-\frac{dU}{ds_i}$ along the energy conserving trajectories, given by $E(s_i, \dot{s}_i) = \text{constant}$, on horizontal planes.

Let us consider the impact of damping on a particle's motion. Damping, if not counteracted, causes the particle to lose energy, so that its path spirals down toward either the negative or positive stable equilibrium points located in the minima of $E(s_i, \dot{s}_i)$. Momentarily returning from this analogy to the real world, we can say that the interaction among investors within each peer group, subject to random idiosyncratic influences, compels the group's sentiment toward either a negative or positive equilibrium, where the consensus of opinion will be reached.

Price feedback adds a fascinating twist to this dynamic. It follows from (A4) that the damping coefficient G is a function of the particle's position and is also dependent on several parameters, most notably the price feedback strength γ. Interestingly, G becomes negative in some regions on the $(s_i, \dot{s}_i)$-plane for γ exceeding a certain critical value (equation A7), implying that for sufficiently strong feedback, damping begins to supply energy to the system instead of dissipating it. As a result, for large γ, some or all trajectories may converge to the limit cycle orbit where the supplied and dissipated energies compensate each other. This yields a potentially new state of dynamic equilibrium in which s_i would exhibit self-sustaining, large-amplitude, periodic oscillations above the cusp of the energy surface between negative and positive sentiment values (the red trajectory in Figure 1b).

The critical value of γ is roughly the same for all s_i (equation A9). Therefore, for supercritical γ, the total sentiment will undergo the limit cycle oscillations, giving rise to the permanent regime of

rallies and crashes, which contradicts the observed market behavior. Conversely, Gusev et al. (2015) showed for the case where $N = 1$ that subcritical γ leads to realistic market regimes. We should briefly inspect this case because in the absence of interaction (F_c), the situation of $N > 1$ is qualitatively similar to that of $N = 1$ under the approximation (7).

The case $N = 1$ offers a simple portrait of trajectories as numerical solutions to equations (6) for $\xi_t = 0$ on the (s, h)-plane (Figure 2a). The distinct regimes illustrated schematically in Figure (1b) in the $(s, \dot{s})$-space are clearly visible here.[17] First, there are small-amplitude, decaying oscillations around the positive equilibrium point inside the deep well. Second, there are large-scale trajectories passing above the cusp of the potential, along which the particle can escape from one well into the other. Third, oscillations also occur inside the shallow well, where sentiment is negative, but as the equilibrium point there is unstable, the particle is quickly ejected onto the trajectories leading into the well in which sentiment is positive. Figure (2b), which depicts the empirical sentiment path[18] traced by the US stock market during 1995-2015, confirms the existence of the above-described three types of sentiment motion.

Thus, we can presume that subcritical γ permits realistic trajectories of sentiment evolution and so will apply subcritical values of γ in the numerical and empirical analyses in the next section.

---

[17] The motion on the (s, h)-plane bears resemblance to the motion on the $(s, \dot{s})$-plane because h can be expressed as a function of s and $\dot{s}$ from equation (6a).

[18] Following the methodology from Gusev et al. (2015), we have constructed a time series of daily $h(t)$ from news data retrieved from the DJ/Factiva archive and substituted $h(t)$ into equation (6a) for $N = 1$ to generate a time series of daily $s(t)$ and obtain, after filtering, the empirical sentiment evolution path.

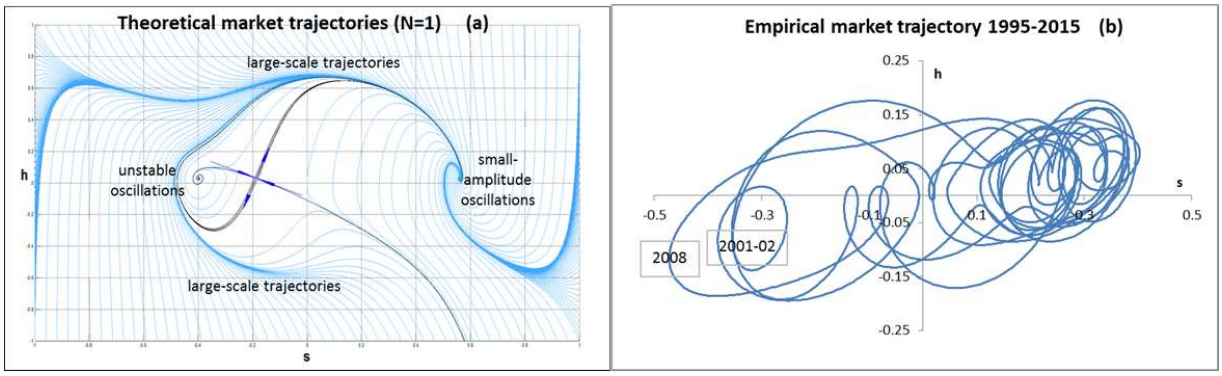


Figure 2: (a) The phase portrait on the (s, h)-plane for the model with $N = 1$, showing an unstable focus in the negative-sentiment well (red asterisk), a stable focus in the positive-sentiment well (green asterisk) and large-scale trajectories crossing the wells (from Gusev et al., 2015). (b) The phase portrait of the empirical market sentiment trajectory (1995-2015). To make this plot, the empirical series of daily $h(t)$ and $s(t)$ have been smoothed by a Fourier filter, removing harmonics with periods less than 100 business days. This path remained predominantly in the positive well, with only two excursions into the negative well during the bear markets of 2001-2002 and 2008.

Next, we examine the influence of the stochastic force F^e generated by the flow of exogenous news (equation A6). This force acts to thrust a particle randomly from one trajectory to another, occasionally forcing it into a region which can lend the particle new dynamics (e.g. from the vicinity of the equilibrium points to the large-scale trajectories that traverse the well and vice versa). Thus, exogenous news flow plays a key role in market dynamics, being a random external force that may, from time to time, trigger changes in market regimes. Note that the asymmetry of the energy surface implies that a stronger force is needed to move a particle onto a path crossing from the deep (positive) well to the shallow (negative) well.

Additionally, owing to their lower inertia, "light" particles with small τ_i react more strongly to F^e than "heavy" particles with large τ_i. As a result, "light" particles can be expected to appear frequently on large-scale trajectories high on the energy surface, while "heavy" particles are likely to spend most of their time on small-scale trajectories around the equilibrium points at its bottom

(Figure 1b).[19] This situation is relevant for the stock market, since a greater volatility in sentiment is expected from short-term investors than from long-term investors.

Finally, there are effects due to the force F_i^c exerted on the i-th particle by the other particles (equation A5). Its action can be viewed through the prism of constraints imposed on the motion by the relations between each pair of particles in terms of their mutual positions and velocities that restrict the degrees of freedom of the motion.

We write down equations for these constraints by observing that according to equation (6a), at any time all particles must share the same h when moving along their paths on the energy surface. Thus, we invert (6a) to express h as a function of s_i and $\dot{s}_i$, and obtain an equation for the constraint f_{ij} between the i-th and j-th particles:[20]

$$f_{ij} = f\left(s_i, \dot{s}_i; s_j, \dot{s}_j\right) = \frac{\operatorname{arctanh}(s_i + \tau_i \dot{s}_i) - \beta_1 s_i}{\operatorname{arctanh}\left(s_j + \tau_j \dot{s}_j\right) - \beta_1 s_j} = 1, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, N. \tag{8}$$

Equations (8) determine the relations between the sentiments and the rates of change in the sentiments of different investor groups due to mutual influences exerted by these groups on each

[19] This analysis is relevant for the particles with $\tau_i \gtrsim \tau_h$. In Section 2.3 we will show that the "ultra-light" particles with $\tau_i \ll \tau_h$ possess no intrinsic dynamics, adjusting instead to the dynamics of "heavier" particles.

[20] We note that these equations constitute $\frac{(N-1)N}{2}$ nontrivial first integrals of motion, out of which $N-1$ are independent. The independent first integrals reduce the degrees of freedom of system (7) from $2N$ to $N+1$, which matches the number of equations in the dynamical system (6).

other. These relations drive synchronization patterns, discussed in the next section, resulting in self-similar behaviors, as well as other effects, in the market.[21]

Acting together, the above-described forces can generate diverse and complex dynamics. For example, "light" particles may in response to negative news migrate from higher orbits in the well to orbits in the vicinity of the negative sentiment equilibrium at the well's bottom. According to (8), this change in the dynamic of "light" particles will require that "heavy" particles adapt their motion to synchronize frequencies and amplitudes. Should this dynamic persist, "heavy" particles, which make a major contribution to total sentiment (4), may cross from the positive well into the negative well, tipping overall sentiment in the market from positive to negative and, as a result, pressure market price downward. We will encounter this scenario of a bear market transition in numerical simulations and empirical analysis in the next section.

At this point, we wish to remind the reader of the main purpose of this section, that is, to develop a conceptual understanding of the dynamics in model (6) prior to submitting it to the brute force of numerical simulations. We have therefore severely truncated this model to isolate the forces acting on a particle (i.e. an investor group) in equation (7) and explored the dynamic stemming from each force separately. The intuition developed here will aid in untangling the dynamics obtained in the next section, with the caveat that these interpretations remain inexact because a specific dynamic of a particle, strictly speaking, can neither be completely attributed to one particular force, nor considered in isolation from other particles.

---

[21] Synchronization is ubiquitous among the behaviors of interacting nonlinear oscillators: e.g. Pikovsky et al. (2001) provide an in-depth treatment of various synchronization effects in coupled oscillator systems.

2.2. Numerical and empirical analyses

In this section, we investigate system (6) numerically for $\tau_i \gtrsim \tau_h$, while the case where $\tau_i \ll \tau_h$ will be treated in the next section. Following Gusev et al. (2015), we use the estimate $\tau_h \sim 1$ day, which is consistent with the behavior of the autocorrelation of $h(t)$ showing a fast decay of "memory" effects on the order of 1-3 business days.

We proceed, first, by considering only two groups of investors, with $\tau_1 = 1$ business day and $\tau_{15} = 15$ business days, to illustrate the effects discussed in the previous section. In terms of oscillator dynamics, the groups with τ_1 and τ_{15} behave, respectively, as "light" and "heavy" particles on the energy surface.

Figure 3 depicts one simulation spanning 700 business days. The "light" particle undergoes a large-scale motion high in the potential well, covering the distance between extreme negative and extreme positive sentiment values in a 1-2 week timeframe. However, this particle can also get caught in a small-scale motion around the equilibrium points at the well's bottom, sometimes staying there for extended periods of time before it can escape (e.g. the intervals 120-170 days and 250-290 days in the positive well and the intervals 290-440 days and 610-690 days in the negative well). As discussed in the previous section, these transitions between large- and small-scale motions are triggered by the stochastic force exerted by exogenous news flow $\kappa\xi_t$.[22]

[22] Exogenous news flow $\kappa\xi_t$ is modeled on daily intervals as normally-distributed white noise with zero mean and unit variance. However, we have chosen $\kappa\xi_t$ to have a small positive intraday autocorrelation on the assumption that news events are positively correlated on intraday time intervals; the autocorrelation is zero over the intervals of one day or longer.

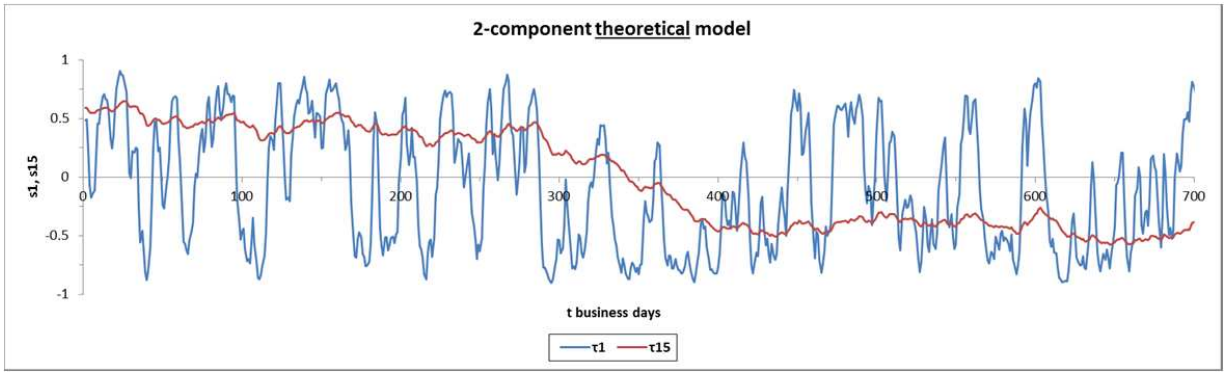


Figure 3: Sentiment evolution in the two-component theoretical model with $\tau_i = 1,\ 15$. Other parameters: $\beta_1 = 1.1, \beta_2 = 1.0, \delta = 0.02, \tau_h = 1, \gamma = 5$.

The "heavy" particle stays on the orbits near the bottom of the positive well during the first 300 days. Its motion is correlated with the motion of the "light" particle, such that its path shifts toward negative or positive values when the "light" particle is in the negative or positive sentiment region, respectively. Equation (8) attributes this behavior to the synchronization of the particles' dynamics. As a result, by observing the motion of one particle, we can deduce the motion of the other. For example, when the "light" particle remains sufficiently long as the solitary particle inside a well, the "heavy" particle will move from the well where it resides into the well in which the "light" particle is residing. Indeed, we observe that the "heavy" particle follows the "light" particle into the negative well in the interval 300-400 days. Visually, it appears as if the "light" particle pulls the "heavy" particle across the wells. As we will see below, this is the basic characteristic of the cascade mechanism governing regime transitions between bull- and bear markets.

Let us now discuss the results of simulations in a more realistic model that consists of nine investor groups with $\tau_i = 1,\ 2,\ 3,\ 4,\ 11,\ 15,\ 19,\ 24,\ 28$, which will be applied to design trading strategies in Section 3. Figure 4 shows sentiment evolution for four groups with $\tau_i = 1,\ 3,\ 11,\ 19$. The synchronicity among these groups is evident. For example, in the interval 750-800 days we can observe how the move of the group with $\tau_i = 1$ from the negative to positive well causes a similar move of the group with $\tau_i = 3$, followed by the group with $\tau_i = 4$ and then the rest of the groups,

each with progressively smaller amplitude. Thus, it appears that regime transitions occur as the cascades propagating from "light" investors with small τ_i toward "heavy" investors with large τ_i.[23]

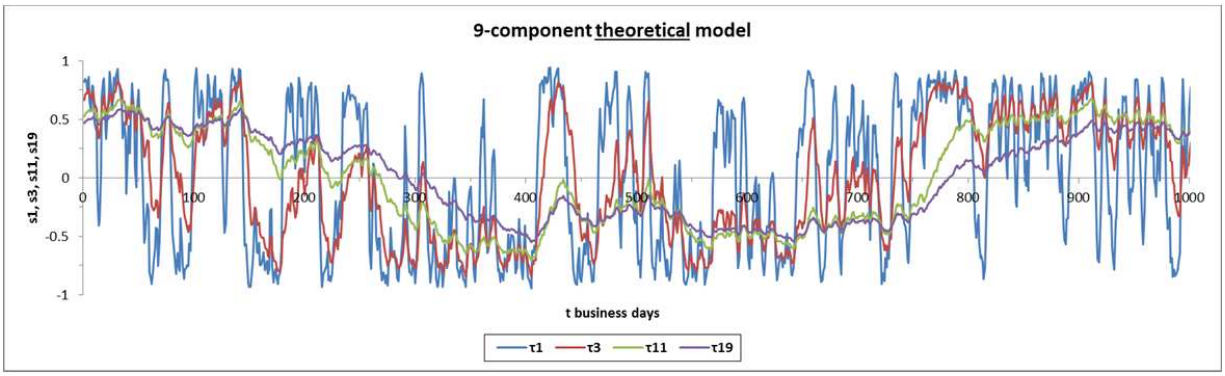


Figure 4: Sentiment evolution in the nine-component theoretical model with $\tau_i = 1, 2, 3, 4, 11, 15, 19, 24, 28$. Other parameters: $\beta_1 = 1.1, \beta_2 = 1.0, \delta = 0.02, \tau_h = 1, \gamma = 10$.

Another pattern discernable in this figure is that the particles form two groups with distinct dynamics: the first group with $\tau_i = 1, 3$ that follows a typical "light" particle dynamic and the second group with $\tau_i = 11, 19$ that behaves like a typical "heavy" particle. This separation implies a relatively sharp transition between the two dynamics as a function of the investment horizon τ_i. Therefore, for a qualitative analysis, it seems justified to approximate interaction in the market as the interaction between two types of participants: volatile short-term investors who are sensitive to incoming information and relatively-static long-term investors whose views on the market are firmly established. It is interesting that both groups are vital to market dynamics since the short-

[23] In this paper we do not consider the effect of a slowly varying β_1 on regime transitions, which was studied in detail by Gusev et al. (2015). These authors showed that β_1 slightly increased during bull markets and slightly decreased during bear markets (while remaining above unity), affecting the shape of the potential well and, correspondingly, the probability of regime transitions.

term investors are sufficiently "nimble" to initiate a change in market regime, while the long-term investors are sufficiently "massive" to actually effect the change.[24]

At this juncture, we can test the model with empirical data for the US stock market between 1995-2015, using the model with the nine components, studied above, to obtain daily $s_i(t)$ from measured daily $h(t)$. Figure 5 shows the results for the period 2005-2009, chosen to highlight the empirical behavior of sentiment during transition to- and from the bear market regime. These results are visually similar to the results of numerical simulation. We particularly note the cascade mechanism of the market regime transitions and the distinct patterns in the behavior of the short-term and long-term investor groups, thus corroborating the main features of the model dynamics empirically.

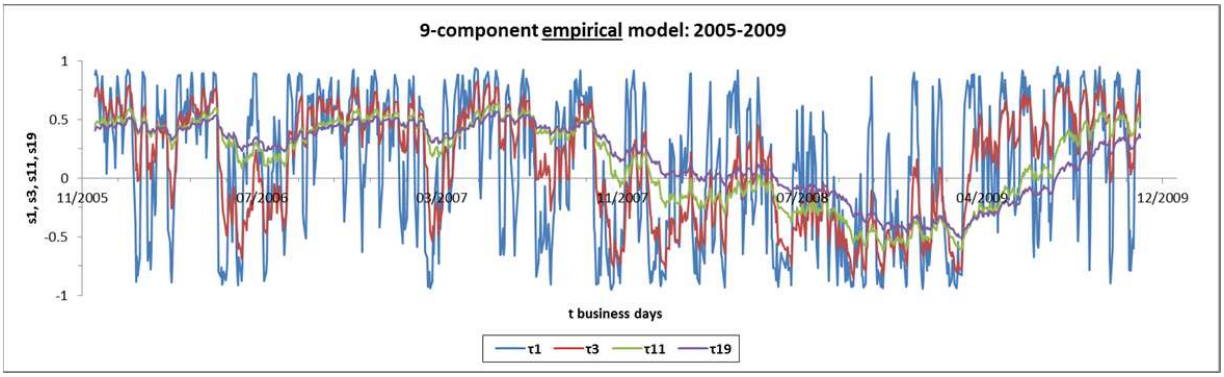


Figure 5: Sentiment evolution in the nine-component empirical model with $\tau_i = 1, 2, 3, 4, 11, 15, 19, 24, 28$. Other parameters: $\beta_1 = 1.1, \beta_2 = 1.0$.

[24] Incidentally, there appear certain parallels between the behavior of these two investor groups and the two types of investors ubiquitous in the market modeling literature, namely systematic traders (also called chartists or noise traders), on the one hand, and fundamental or informed traders, on the other hand.

2.3. Efficient market regime

In this section we show that in the leading order the dynamic of short-term investors decouples from the dynamics of investor groups with longer horizons, where short-term investors are defined as traders operating on timescales much shorter than $\tau_h \sim 1$ day. In particular, we will see that investment processes on these timescales are not involved in the feedback mechanism, but instead cause market price to adjust quickly to new information, contributing to market efficiency.

Let us consider a two-component [25] system (6) such that $\tau_1 \ll \tau_h \lesssim \tau_2$:

$$\tau_1 \dot{s}_1 = -s_1 + \tanh(\beta_1 s_1 + \beta_2 h), \quad (9a)$$

$$\tau_2 \dot{s}_2 = -s_2 + \tanh(\beta_1 s_2 + \beta_2 h), \quad (9b)$$

$$\tau_h \dot{h} = -h + \tanh(\bar{\gamma}(\tau_1 \dot{s}_1 + \tau_2 \dot{s}_2) + \delta + \kappa \xi_t), \quad (9c)$$

where $\bar{\gamma} = \frac{\gamma}{\tau_1 + \tau_2}$.

We first examine this system on timescales $\sim \tau_1$. It follows from (9a) that s_1 can change by $O(1)$ over τ_1. Similarly, it follows from (9b,c) that h and s_2 change respectively by $O\left(\frac{\tau_1}{\tau_h}\right)$ and $O\left(\frac{\tau_1}{\tau_2}\right)$ over τ_1. Such a slow variation in h and s_2 can in the first order be neglected, leading to

$$\tau_1 \dot{s}_1 = -s_1 + \tanh(\beta_1 s_1 + \beta_2 h), \quad (10a)$$

$$\dot{s}_2 = 0, \quad (10b)$$

$$\dot{h} = 0. \quad (10c)$$

[25] This result holds for model (6) with $N \geq 2$. For simplicity, we show its derivation in the case $N = 2$.

According to these equations, both h and s_2 remain approximately constant on timescales $\sim\tau_1$, over which s_1 converges from any initial position toward its equilibrium state ($\dot{s}_1 = 0$), given by

$$s_1 = \tanh(\beta s_1 + \beta_1 h). \qquad (11a)$$

Next, we study system (9) on timescales $\sim\tau_h$ or longer. As viewed on these timescales, s_1 almost instantaneously ($\sim\tau_1 \ll \tau_h$) converges to the position of equilibrium (11a). Consequently, sentiment s_1 behaves as if it were in a state of permanent equilibrium, while h and s_2 evolve according to

$$\tau_2 \dot{s}_2 = -s_2 + \tanh(\beta_1 s_2 + \beta_2 h), \qquad (11b)$$

$$\tau_h \dot{h} = -h + \tanh(\bar{\gamma}\tau_2 \dot{s}_2 + \delta + \kappa\xi_t). \qquad (11c)$$

Accordingly, system dynamics on these timescales are determined by equations (11b,c), whereas s_1 merely follows any changes in h by moving along the equilibrium solution (11a), which is called the isocline.[26,27]

---

[26] Strictly speaking, although s_1 spends most of its time ($\sim\tau_h$) on the isocline, where its velocity is close to zero, it can also leave the isocline and move briefly ($\sim\tau_1$) along a trajectory in its vicinity (Figure 6). Therefore, $\dot{s}_1$ is nearly zero at all times, except for brief moments when the trajectory departs the isocline, so that the average contribution in (9c) due to s_1 is small as compared to s_2 and can be neglected.

[27] This approximation does not work in a system with $N = 1$, as there exist $\bar{\gamma}$ for which the term $\bar{\gamma}\tau_1\dot{s}_1$ in (9c) cannot be neglected. As a result, for large $\bar{\gamma}$ the coupling between s_1 and h can be strong enough to cause a limit cycle dynamic. The situation is different in systems with $N \geq 2$, which model market dynamics with a greater precision. There, $\tau_1\dot{s}_1$ can in average be neglected in comparison with $\tau_i\dot{s}_i$ ($\tau_i \gtrsim \tau_h$) in (9c), so that the motion of s_1 is completely determined in this case by the dynamics between h and s_i ($i \neq 1$).

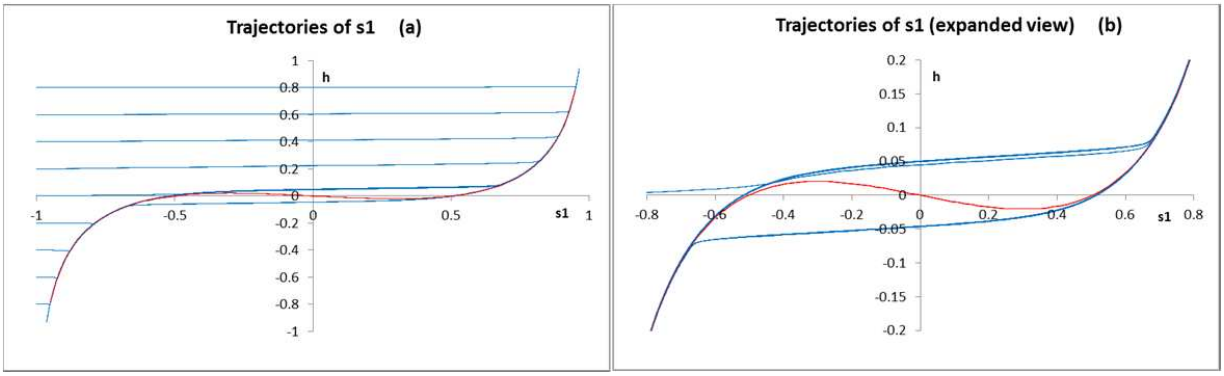


Figure 6: The isocline (red) and trajectories of s_1 (blue) for different initial conditions in system (9) with $\tau_1 = 0.01$ and $\tau_2 = 25$. Sentiment s_1 falls on the isocline along the approximately horizontal lines: the motion occurring so fast that h has little time to change. Sentiment s_1 continues to move along the isocline, following slowly evolving h. The segment of the isocline between its extrema is unstable, which causes sentiment to vacillate between the isocline's left and right branches. The overall motion consists of slow passages along the isocline and fast jumps between its branches, determined solely by the dynamics between h and s_2.

Thus, we have shown that sentiment s_1, which develops on timescales $\tau_1 \ll 1$ day, decouples from the system's dynamics and does not participate in the sentiment-information feedback; instead, sentiment s_1 resides in a state of approximate equilibrium,[28] adjusting instantaneously to changes in information flow h and so driving corresponding changes in market price. We therefore conclude that market efficiency persists on timescales much less than one day. Further, we can

[28] This analytical result, which follows from equations (10) and (11), has been verified by direct numerical simulations. We also note that equations (10) and (11) can be obtained by rescaling system (9) (using a dimensionless time variable) and inspecting the leading-order balance on relevant timescales; we have chosen an informal derivation above for the sake of preserving the readability of this section.

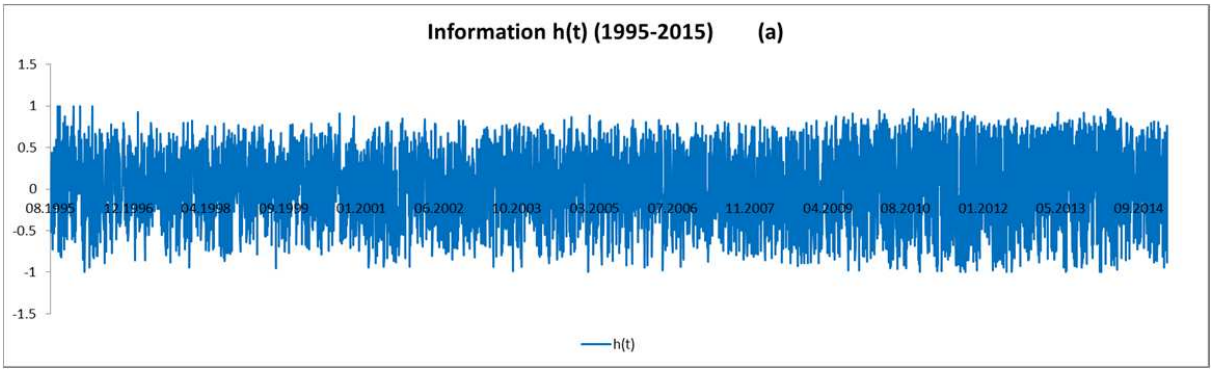
reasonably conjecture that intraday investment processes, generally, take place in a quasi-efficient market regime due to weak feedback, gradually giving way to the dynamics of mutually coupled information and sentiment over horizons longer than one day, which we study in this paper.

3. Trading strategies based on the market return forecasts

The previous section concluded that the stock market is efficient on short timescales $\tau_1 \ll \tau_h \sim 1$ day. This conclusion also sheds light on the mechanics of (intraday) news-based trading. According to model (5), analysts take exogenous news flow ξ_t as an input to generate and propagate information h in time $\sim\tau_h$. Once h is released, the short-term traders will move market price by $\Delta p \sim \Delta s_1$ in time $\sim\tau_1$, i.e. with a practically instantaneous effect. It follows that the objective of news-based trading is to capture this Δp by estimating h from ξ_t before h has been released.

Model (5) states that this same release of information h will cause further changes in price, namely due to Δs_2, Δs_3, Δs_4 and so forth, that will be unfolding over days, weeks and months. In this section, we aim to verify this statement by designing and testing algorithmic trading strategies that can capture these longer-term impacts.

Thus, we apply model (5) to develop return forecasts, upon which we build the prototypes of trading strategies and backtest them against the S&P 500 Index. For this purpose, we make use of the empirical time series of daily $h(t)$, $s(t)$ and $p(t)$ for the period 1995-2015 (Figure 7). We note that the correlation between the daily model prices and the daily index log prices is around 95%.



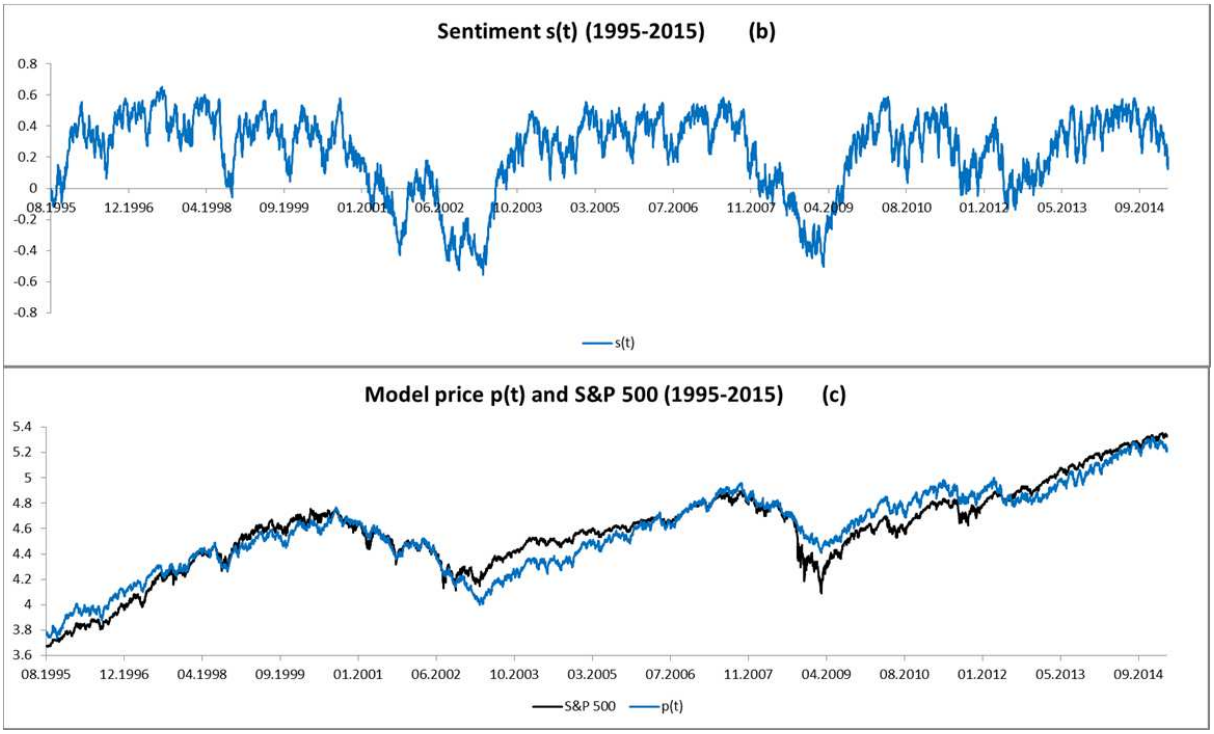


Figure 7: Daily time series of information $h(t)$, sentiment $s(t)$ and price $p(t)$ from 1995 to 2015, where $s(t)$ and $p(t)$ have been obtained from measured $h(t)$, using the nine-component model (5b) with $\tau_i = 1, 2, 3, 4, 11, 15, 19, 24, 28$ and $\beta_1 = 1.1, \beta_2 = 1.0$, and the price formation equation (5a) with $a_1 = 0.368, a_2 = 0.003, s_* = 0.126$ and the integration constant equal to 3.790. As mentioned earlier, $h(t)$ has been extracted from news data in the DJ/Factiva archive, applying the methodology in Gusev et al. (2015). SPDR S&P 500 ETF is taken as an investable proxy of the S&P 500 Index in (c).

The approach to constructing strategies is as follows. In accordance with (5a) and (4), price changes are determined by the sentiments of investor groups with different investment horizons, which contribute to the formation of aggregate market sentiment on different timescales. Thus, if we extract the characteristic sentiment dynamic pertaining to each group, we can forecast price over multiple time horizons and implement trades based upon these forecasts.

We therefore wish to capture the characteristic dynamic of each investor group, while taking into account the influence of the other groups. Then, we apply this dynamic to extrapolate the future

market position from its current position, given by empirically obtained h and $s(h)$, and so generate return forecast over the relevant horizon. In terms of the motion in the $(N+1)$-dimensional phase space (h, s), this means investigating the characteristic behavior of the phase trajectory $\big(h(t), s(t)\big)$ projected on the (h, s_i)-plane, subject to constraints imposed on s_i by s_j $(j \neq i)$.

In practice, the nine-component model (5) with $\tau_i = 1, 2, 3, 4, 11, 15, 19, 24, 28$ has been applied to produce return forecasts over time horizons corresponding to the characteristic timescales in the model. These forecasts can form the basis of a number of trading strategies, four of which, with different holding periods, are presented here. Specifically, we show one strategy based on the shortest forecast, two strategies based on different combinations of the equally-weighted forecasts and the last strategy based on the longest forecast. In the backtest results, the average holding periods of these strategies have been, respectively, around 9, 12, 25 and 45 business days (Table I).

Each strategy generates daily a buy-, sell- or hold signal on the SPDR S&P 500 ETF (Bloomberg ticker: SPY), an exchange-traded fund tracking the S&P 500 Index, such that today's trading instruction is applied to the next day's opening price. The signal has no price input: it is based solely on the forecast derived from news. We emphasize that these strategies are merely crude prototypes, designed not for actual trading but to verify return predictability; as such, these strategies do not include position sizing and risk management.

These strategies have been backtested over the period 1995-2015. Since the strategies require an in-sample period of roughly 2000 business days for parameter value selection, the backtested performance is reported for the out-sample period 2003-2015. We note that the strategies are not sensitive to the location of the in-sample period in the backtest interval and that in-sample and out-sample performance statistics are similar. The backtested results are compared with those of two benchmarks: a passive, long-only investment in SPDR ETF and an active, long-short strategy that

combines a 5-day reversal and 250-day momentum strategies applied to SPDR ETF over this same period.

Figure (8) and Tables I and II show the cumulative returns, performance statistics and cross-correlations, respectively.

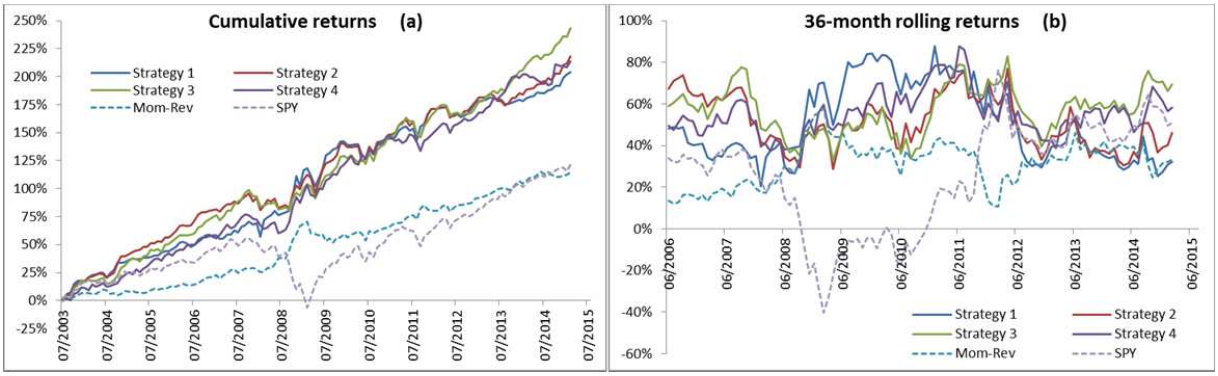


Figure 8: Performance graphs of four news-based strategies backtested against SPDR S&P 500 ETF ("SPY") and a simple price-based momentum-reversal strategy ("Mom-Rev") during the out-sample period 2003-2015. The invested capital, as a base for daily P&L accruals in active strategies, was subject to a requirement, applied on the in-sample, that each strategy be on average 100%-invested during the periods for which trading signal was nonzero. No transaction costs were applied. No risk-free returns and no funding costs were accrued on the under- and overinvested days, respectively. (a) Cumulative returns. (b) 3-year rolling returns.

Table I: Statistics based on monthly returns

	mean (%, p.a.)	volatility (%, p.a.)	mean / volatility	alpha	beta	holding period (business days)
Strategy 1	17.5	14.9	1.17	1.471	0.011	8.6
Strategy 2	18.7	15.1	1.24	1.299	0.302	12.0
Strategy 3	20.8	13.7	1.52	1.456	0.324	25.3
Strategy 4	18.3	14.0	1.30	1.453	0.084	44.6
Mom-Rev	9.8	9.4	1.04	0.857	-0.034	3.8
SPY	10.4	14.3	0.73	0.000	1.000	n/a

Table II: Correlations (%) based on monthly returns

	Strategy 1	Strategy 2	Strategy 3	Strategy 4	Mom-Rev	SPY
Strategy 1	100	79	62	35	30	1
Strategy 2	79	100	82	45	23	29
Strategy 3	62	82	100	65	25	34
Strategy 4	35	45	65	100	13	9
Mom-Rev	30	23	25	13	100	-5
SPY	1	29	34	9	-5	100

The pro-forma returns of all four strategies have exceeded the returns of the equity index and the priced-based momentum-reversal strategy, on absolute- and risk-adjusted bases, and have also exhibited a relatively low correlation with these benchmarks on the 12-year out-sample period. Note that the lengths of the average holding periods of the news-based strategies are substantially longer than that of the active benchmark. These results point toward return predictability and indicate that the model has, at least partially, captured this predictability.

4. Discussion

The starting point for this paper was the stock market model in Gusev et al. (2015). The model consists of analysts[29], who extract relevant information from price changes and exogenous news, and investors, who apply this information to trade. The interaction mode among the agents is assumed to be all-to-all to derive the model in analytic form as a dynamical system governing the evolution of the mutually-coupled endogenous "macroscopic" variables: market price (p), investor sentiment (s) and information (h) supplied by analysts (equations 1).

[29] This term is applied in a collective sense, comprising financial analysts, newspaper journalists, market commentators, finance bloggers and other participants who communicate their market views through mass media.

The above assumption on the interaction topology in the model is instrumental for identifying the basic mechanisms that drive market dynamics; however it also makes the model insufficiently fine-grained for testing return predictability. An introduction of a more complex topology is not a straightforward task, mainly for the lack of obvious choices and because of the sensitivity of model's properties to the topology, but also because an unnecessary complexity may rather impede understanding than help, which is a general problem with selecting interaction patterns in agent-based models.

We sidestepped this problem by arranging investors in peer networks, according to investment horizons, on the assumption that interaction among peers is strongest, obtaining in the limiting case the all-to-all interaction within each peer network and zero outside interaction (Section 1.2). Note that despite the absence of interaction across the networks, each still can impact the others by contributing to the common information flow that affects all networks in equal measure. This phenomenological approach has enabled us to derive a market model with N investor networks or groups, which is still simple enough to be expressed in analytic form and yet sufficiently realistic to be applied for return prediction (equations 5).

In particular, this model demonstrates that with respect to processing information the market behaves efficiently on intraday timescales (Section 2.3) and inefficiently on timescales longer than one day (Section 2.2). The model equations reveal that it is price feedback that enforces market inefficiency by coupling the endogenous variables. We have shown that feedback is negligible on the intraday scale, but is important over longer time horizons, where it contributes to leading-order dynamics.

The situation where a system exhibits different behaviors on different scales is not unusual in nature. Fluid dynamics provides an instructive example. In fluids, inertia plays a central role on large scales, while viscous damping is dominant on small scales.[30] As a result, the large-scale dynamics and the small-scale dynamics are fundamentally different: inertia induces a nonlinear endogenous dynamic at macro scales, whereas at micro scales inertia is so small that velocities in a fluid's flow can adjust immediately to exogenous changes, leading to a state of adiabatic equilibrium.

The above example presents a useful analogy for market dynamics as the market also evolves on many (time)scales, driven by participants with various investment horizons. Indeed, equation (7) states that the investment horizon τ_i is analogous to the mass of the i-th investor group's sentiment in the context of sentiment dynamics. Accordingly, the contribution of inertia to dynamics on short timescales is negligible because the mass of the sentiment of relevant investor groups is very small: these investors react so fast as to move prices almost instantaneously in response to new information, leading to an (adiabatic) equilibrium regime on these timescales (Section 2.3). On the contrary, inertia cannot be neglected on longer timescales, which results in effective interaction between investors and analysts in the model, yielding complex dynamics characterized by nonlinear feedback (Section 2.2).

The intraday market efficiency does not imply the lack of short-term trading opportunities. In fact, whereas price adjusts instantaneously, information h is released by analysts on average on the scale $\tau_h \sim 1$ day. This delay, which can likely be attributed to information processing (e.g. gathering,

[30] For example, we swim by using water's resistance to create momentum; however this strategy would fail if we were the size of bacteria: for microorganisms, water appears as viscous as honey for humans, forcing them to evolve unique propulsion techniques, such as corkscrew-like locomotion mechanisms among others.

aggregation, analysis, editing) and distribution frequency, creates a window of opportunity for intraday trading between the occurrence of a news event (e.g. the release of an earnings report) and its reflection in h.

This short-term price reaction to information released by analysts is incomplete because the overall market sentiment also includes the sentiments of investor groups with longer investment horizons, which can influence the mid- and long-term price evolution. Equations (6) imply that a change in information will cause changes in sentiment on many different timescales and that these changes will in turn cause changes in information – creating a feedback loop. This complex multi-scale interplay between information and sentiment is the generator of the variety in market behavior, including self-similar variation patterns briefly explained as a synchronization effect in Sections 2.1 and 2.2.

In addition, the market is subject to the impact of exogenous news flow that acts as an external stochastic driving force (equations 5). However, as discussed above, on long timescales the market acquires inertia and with it a resistance to change in direction. As a result, market behavior can be predictable in situations where inertia outweighs noise. A test of this predictability has been carried out in Section 3.

Our approach to return prediction is based on principles similar to those of weather forecasting, i.e. of combining theoretical models and empirical measurements. We have obtained the empirical time series of information and sentiment and applied model (5) to forecast market price and develop the prototypes of trading strategies. The backtested results, compared to passive and active benchmarks, suggest that market forecasting on the above-described principles functions with a precision sufficient for the development of successful trading strategies operating over horizons ranging from days to months.

In this paper, we have sought to develop a market model that is sufficiently sophisticated to both replicate past performance and predict future returns, while being tractable to highlight some of the mechanisms underlying market dynamics. We fully realize that the range of processes occurring in the market is substantially broader than those captured by this model. For example, we have not included fundamental traders (who apply financial analysis) and systematic traders (who use price data). We note, however, that the analysts in the model perform analogous functions, so that, in the first order, the impacts due to these two types of investors are taken into account. In any case, this is work in progress: modeling of economic or natural systems must proceed from simple to complex as the grasp of underlying mechanisms improves. As such, the objective of this paper has been to improve the current understanding of market dynamics and thus provide a basis for further modeling efforts.

5. Conclusion

In this paper we have theoretically and empirically investigated stock market return predictability on various time horizons. In particular, we introduced a news-driven model with heterogeneous investors and, using this model, developed and backtested the prototypes of trading strategies. In the course of this study we have reached the following conclusions:

1. There exist two characteristic timescales of stock market dynamics. Over time horizons, shorter than one day, the market behaves efficiently with respect to processing information. On time horizons longer than one day, the market becomes inefficient.
2. This informational inefficiency is caused by a feedback loop, which acts on timescales longer than one day, interconnecting information, opinion and price and so resulting in fundamentally nonlinear overall dynamics.
3. On these timescales, the relevant model for market dynamics is a dynamical system governing the evolution of mutually-coupled information, opinion and price, driven by exogenous news.

4. According to this model, the sentiments of investor groups with different investment horizons collectively form aggregate investor opinion that determines a price dynamic, which in turn influences information flow acting on all groups participating in the market.
5. This common information flow provides a link through which the sentiments of investor groups are mutually coupled. As such, information induces self-similar dynamics among investor groups on different timescales through synchronization, leading to complex self-similar patterns observable in market behavior.
6. These investor groups form two classes characterized by distinct dynamics. The first class contains investors with horizons less than one week. Their average sentiments are typically volatile, oscillating between negative and positive values in the timeframe from roughly one week to one month. The second class consists of investors with horizons exceeding one week. Their average sentiments primarily undergo small-amplitude oscillations around either a positive or negative equilibrium, where the consensus of opinion is reached.
7. The regime change between bull- and bear markets takes place when investors with long investment horizons transit from one sentiment equilibrium to the other. This transition occurs as a cascade, whereby investors with longer horizons follow, one-by-one, investors with shorter horizons.
8. Lastly, the backtested results of strategy prototypes, designed by combining theoretical models (dynamical systems) with empirical observations (news data) for trading over time horizons that range from days to months, indicate that the stock market dynamics are in fact predictable. The objective of our future research is to test this predictability in actual trading.

Acknowledgments

We are grateful to LGT Capital Partners for partially funding this project and to Dow Jones & Company for providing access to the Factiva.com news archive. We would like to thank a number of people who provided their help over the course of this work: John Orthwein for editing this paper

and contributing ideas on its readability; Ed King and Maxim Zhilyaev for helping with data retrieval and management; George Coplit, Matthias Feiler, Roger Hilty, Christian Jung and Francesco Valenti for useful comments on backtest setup and robustness; and Vitaliy Godlevsky for discussions regarding market behaviors on short timescales and other helpful comments. As always, any remaining errors are ours alone.

Appendix A: Approximation of the dynamical system

Here we express the dynamical system (6) as a system of forced, coupled, nonlinear oscillators.[31] We differentiate equation (6a) with respect to time and use equation (6b) to obtain

$$\begin{aligned}\tau_i \ddot{s}_i &= \Phi(s_i, \dot{s}_i, \dot{s}, \xi_t) \\ &= -\dot{s}_i + (1 - (s_i + \tau_i \dot{s}_i)^2)\left(\beta_1 \dot{s}_i + \frac{\beta_1}{\tau_h} s_i - \frac{1}{\tau_h} \operatorname{arctanh}(s_i + \tau_i \dot{s}_i)\right) \\ &+ (1 - (s_i + \tau_i \dot{s}_i)^2)\frac{\beta_2}{\tau_h} \tanh(\gamma \dot{s} + \delta + \kappa \xi_t), \quad i = 1, 2, \dots, N, \end{aligned} \tag{A1}$$

where $s = \frac{\sum \tau_i s_i}{\sum \tau_i}$ in accordance with (4).

These equations govern the motion of N oscillators, that is N particles with the coordinates s_i and the velocities $\dot{s}_i$, subjected to the force $\Phi(s_i, \dot{s}_i, \dot{s}, \xi_t)$. Note that τ_i is analogous to the mass of the i-th particle in the sense that the impact of a force on the particles with small τ_i ("light" particles) is

[31] We follow the steps of a similar derivation for $N = 1$ in Gusev et al. (2015) (Appendix C). That same appendix provides a detailed analysis of the phase portrait geometry, including the bifurcations of equilibrium points and the formation of a stable limit cycle.

greater than on the particles with large τ_i ("heavy" particles). In other words, "light" particles have small inertia and "heavy" particles have large inertia.

The first two terms in $\Phi(s_i, \dot{s}_i, \dot{s}, \xi_t)$ contain the restoring and damping force components responsible for autonomous dynamics. The third term describes the force originating from the i-th sentiment component feedback ($\sim\gamma\tau_i\dot{s}_i$ and $\sim\delta$) and the external forces exerted by the other particles ($\sim\gamma\sum_{j\neq i}\tau_j\dot{s}_j$) and by the flow of exogenous news ($\sim\kappa\xi_t$) in the argument of the hyperbolic tangent. Being dependent on position and velocity, these forces vary along a particle's trajectory.

For illustration purposes, we expand $\Phi(s_i, \dot{s}_i, \dot{s}, \xi_t)$ into a truncated Taylor series to separate the above-mentioned force components and write equation (A1) in a "canonical" form:

$$\tau_i\ddot{s}_i + G(s_i)\dot{s}_i + \frac{dU(s_i)}{ds_i} = F_i^c + F^e, \quad i = 1, 2, \dots, N. \tag{A2}$$

In this equation, $U(s_i)$ has the meaning of a potential and is given with the precision up to a constant by

$$U(s_i) = \frac{1}{\tau_h}\left(\frac{\beta_1 - \frac{2}{3}}{4}{s_i}^4 - \frac{\beta_1 - 1}{2}{s_i}^2 - \beta_2\delta s_i\right); \tag{A3}$$

$G(s_i)$ has the meaning of a damping coefficient and is given by

$$G(s_i) = \left(1 - \beta_1 - \beta_2\bar{\gamma}\frac{\tau_i}{\tau_h} + \frac{\tau_i}{\tau_h}\right) + 2\beta_2\frac{\tau_i}{\tau_h}\delta s_i + \left(\beta_1 + \beta_2\bar{\gamma}\frac{\tau_i}{\tau_h} + 2(\beta_1 - 1)\frac{\tau_i}{\tau_h}\right){s_i}^2; \tag{A4}$$

where

$$\bar{\gamma} = \frac{\gamma}{\sum\tau_i};$$

F_i^c has the meaning of an external force exerted by the other particles and is given by

$$F_i^c = \frac{\beta_2 \bar{\gamma}}{\tau_h} \sum_{j \neq i} \tau_j \dot{s}_j \, ; \qquad (A5)$$

and F^e has the meaning of an external force due to the flow of exogenous news and is given by

$$F^e = \frac{\beta_2}{\tau_h} \kappa \xi_t. \qquad (A6)$$

As such, equation (A2) describes the motion of a particle inside an asymmetric W-shaped potential well (A3) for $\beta_1 > 1$ (see Figure 1) in the presence of nonlinear damping (A4), driven by the forces generated through interaction between particles (A5) and through the impact of exogenous news (A6). Note that the feedback force in (A1), proportional to $\bar{\gamma} \tau_i \dot{s}_i$ and δ, has been incorporated into the potential force (only the component $\sim \delta$) and into the damping force on the left-hand side of (A2).

To obtain equations (A2)-(A6), we have truncated the Taylor series of $\Phi(s_i, \dot{s}_i, \dot{s}, \xi_t)$ at terms above cubic in s_i, linear in $\dot{s}_i$ and linear in δ and have kept only the leading terms in the expressions for the forces F_i^c and F^e. Consequently, these equations are, strictly speaking, only valid in the region where $|s_i| \ll 1$ and $|\dot{s}_i| \ll 1$. However, we expect that the formula for the potential $U(s_i)$, which does not contain the heavily truncated terms $\sim \dot{s}_i$, holds reasonably well for all sentiment values ($|s_i| \leq 1$) within the relevant range of parameter values, namely $\beta_1 \sim 1, \beta_2 \sim 1$ and $\delta \ll 1$.

As follows from (A4), the damping coefficient $G(s_i)$ is negative if

$$\bar{\gamma} > \bar{\gamma}_c(s_i, \tau_i) = \frac{\left(1 - \beta_1 + \frac{\tau_i}{\tau_h}\right) + 2\beta_2 \frac{\tau_i}{\tau_h} \delta s_i + \left(\beta_1 + 2(\beta_1 - 1)\frac{\tau_i}{\tau_h}\right) {s_i}^2}{\beta_2 \frac{\tau_i}{\tau_h} (1 - {s_i}^2)}. \qquad (A7)$$

Condition (A7) means that for sufficiently large $\bar{\gamma}$ there are regions where energy in the system is amplified (negative damping), pointing toward the possibility of a limit cycle. Because this condition

has been derived for $|s_i| \ll 1$ and $\delta \ll 1$, we can in the leading order neglect the terms $\sim \delta s_i$ and $\sim {s_i}^2$ to obtain

$$\gamma > \gamma_c(\tau_i) = \frac{1 - (\beta_1 - 1)\frac{\tau_h}{\tau_i}}{\beta_2} \sum \tau_i . \qquad (A8)$$

Since $\beta_1 \sim 1$ (we use $\beta_1 = 1.1$), the second term in the numerator in (A8) is much smaller than unity for particles with $\tau_i \geq \tau_h$ (we set $\tau_h = 1$ day) and can be neglected. This means that γ_c has approximately the same value for all investors with investment horizons equal to or longer than one day, given by

$$\gamma_c = \frac{1}{\beta_2} \sum \tau_i . \qquad (A9)$$

References

Antweiler, W., Frank, M. Z. 2004. Is all that talk just noise? The information content of internet stock message boards. Journal of Finance 59, 1259-1294.

Baker, M., Wurgler, J. 2007. Investor sentiment in the stock market. Journal of Economic Perspectives 21, 129-151.

Baker, M., Wurgler, J. 2000. The equity share in new issues and aggregate stock returns. Journal of Finance 55, 2219-2257.

Brown, G. W., Cliff, M. T. 2004. Investor sentiment and the near-term stock market. Journal of Empirical Finance 11, 1-27.

Campbell, J. Y., Thompson, S. B. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1455–1508.

Campbell, J. Y., Shiller, R. J. 1988a. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661-676.

Campbell, J. Y., Shiller, R. J. 1988b. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195-227.

Carro, A., Toral, R., San Miguel, M. 2015. Markets, Herding and Response to External Information. PLoS ONE 10(7): e0133287 (28pp).

Castellano, C., Fortunato, S., Loreto. V. 2009. Statistical physics of social dynamics. Reviews of Modern Physics 81, 591-646.

Cochrane, J. H. 2008. The dog that did not bark: A defense of return predictability. Review of Financial Studies 21, 1533-1575.

Cont, R., Bouchaud, J. P. 2000. Herd behavior and aggregate fluctuations in financial markets. Macroeconomic Dynamics 4, 170-196.

Da, Z., Engelberg, J., Gao, P. 2014. The sum of all FEARS: investor sentiment and asset prices. Review of Financial Studies 28, 1-32.

Das, S. R., Chen, M. Y. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the Web. Management Science 53, 1375-1388.

Fama, E. F., French, K. R. 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23-49.

Fama, E. F., French, K. R. 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3-25.

Ferson, W., Sarkissian, S., Simin, T. 2008. Asset pricing models with conditional alphas and betas: The effects of data snooping and spurious regression. Journal of Financial and Quantitative Analysis 43, 331-354.

Ferson, W. E., Sarkissian, S., Simin, T. T. 2003. Spurious regressions in financial economics? Journal of Finance, 58, 1393-1413.

Franke, R. 2014. Aggregate sentiment dynamics: A canonical modelling approach and its pleasant nonlinearities. Structural Change and Economic Dynamics. 31, 64-72.

Goyal, A., Welch, I. 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455-1508.

Goyal, A., Welch, I. 2003. The myth of predictability: Does the dividend yield forecast the equity premium? Management Science 49, 639-654.

Gusev, M., Kroujiline D., Govorkov B., Sharov S. V., Ushanov D., Zhilyaev, M. 2015. Predictable markets? A news-driven model of the stock market. Algorithmic Finance 4, 5-51.

Ising, E. 1925. Beitrag zur Theorie des Ferromagnetismus. Zeitschrift für Physik 31, 253-258.

Levy, H., Levy, M., Solomon, S. 2000. Microscopic simulation of financial markets: From investor behavior to market phenomena. Orlando: Academic Press.

Loughran, T., McDonald, B. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66, 35-65,

Lux, T. 2011. Sentiment dynamics and stock returns: The case of the German stock market. Empirical Economics 41, 663-679.

Lux, T. 2009. Stochastic behavioral asset-pricing models and the stylized facts. In: Handbook of financial markets: dynamics and evolution. Amsterdam: North-Holland (pp. 161-216).

Lux, T., Marchesi, M. 1999. Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397, 498-500.

Lux, T. 1995. Herd behaviour, bubbles and crashes. Economic Journal, 105, 881-896.

Novy-Marx, R. 2014. Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. Journal of Financial Economics 112, 137-146.

Pikovsky, A., Rosenblum, M., Kurths, J. 2001. Synchronization – A universal concept in nonlinear sciences. New York: Cambridge University Press.

Shiller, R. J. 2003. From efficient markets theory to behavioral finance. Journal of Economic Perspectives 17, 83-104.

Slanina, F. 2014. Essentials of econophysics modelling. New York: Oxford University Press.

Sornette, D. 2014. Physics and financial economics (1776-2013): Puzzles, Ising and agent-based models. Reports on Progress in Physics 77, 062001 (28pp).

Suzuki, M., Kubo, R. 1968. Dynamics of the Ising model near the critical point. Journal of the Physical Society of Japan 24, 51-60.

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. Journal of Finance 62, 1139-1168.