

Market force, ecology, and evolution¹

J. Doyne Farmer
 Prediction Company
 236 Montezuma Ave.
 Santa Fe NM 87501

In financial markets an excess of buying tends to drive prices up, and an excess of selling tends to drive them down. This is called market impact. Based on a simplified model for market making it is possible to derive a unique functional form for market impact. This can be used to formulate a non-equilibrium theory for price formation. Commonly used trading strategies such as value investing and trend following induce characteristic dynamics in the price. Although there is a tendency for self-fulfilling prophecies, this is not always the case; in particular, many value investing strategies fail to make prices reflect values. When there is a diversity of perceived values, nonlinear strategies give rise to excess volatility. Many market phenomena such as trends and temporal correlations in volume and volatility have simple explanations. The theory is both simple and experimentally testable.

Under this theory there is an emphasis on the interrelationships of strategies that makes it natural to regard a market as a financial ecology. A variety of examples show how diversity emerges automatically as new strategies exploit the inefficiencies of old strategies. This results in capital reallocations that evolve on longer timescales, and cause apparent nonstationarities on shorter timescales. The drive toward market efficiency can be studied in the dynamical context of pattern evolution. The evolution of the capital of a strategy is analogous to the evolution of the population of a biological species. Several different arguments suggest that the timescale for market efficiency is years to decades.

Preliminary version. Referencing is incomplete. Comments are greatly appreciated (jdf@predict.com).

1. I would like to thank Shareen Joshi for critical help with the simulations. Sections 3.2-3.5, 4.1.2, and 4.2 are joint work that will be published elsewhere.

1.	Introduction.....	3
2.	Force	7
2.1	Market impact implies market dynamics	7
2.2	Simple market making model	8
2.2.1	Derivation of basic model.....	9
2.2.2	Relation to supply and demand	13
2.2.3	Relationship to market making with limit orders	14
2.2.4	Relation to prior research in market making	14
2.3	Time varying liquidity.....	15
2.3.1	Inventory effect.....	15
2.3.2	Asymmetric markets.....	16
2.4	Dynamics	17
2.5	Profits, losses, and game theory	19
2.6	Market friction	21
2.7	Other market forces.....	22
2.8	Experimental verification.....	23
2.9	Summary and discussion.....	24
3.	Ecology	25
3.1	January effect	25
3.2	Value investors	26
3.2.1	Pure order-based value strategies and equilibrium economics.....	28
3.2.2	Pure position based strategies.....	31
3.2.3	State-dependent threshold value strategies.....	33
3.2.4	Mixed technical and value strategies.....	37
3.2.5	Diverse values and excess volatility	37
3.3	Trends and trend followers.....	39
3.3.1	Information diffusion and rumors.....	40
3.3.2	Trend followers.....	40
3.4	Market drift and inventory effects.....	42
3.5	Value investors and trend followers together	44
3.6	Summary and discussion.....	49
4.	Evolution	51
4.1	Mechanisms of financial evolution	51
4.1.1	Capital reallocation and separation of timescales	51
4.1.2	Long term adjustments in liquidity.....	53
4.2	Competition and diversity in financial ecologies	54
4.2.1	Competition between different classes of strategies.....	54
4.2.2	Competition within a given class of strategies	56
4.3	Efficiency	58
4.3.1	Definition of an efficient market.....	58
4.3.2	Efficiency via complexity	59
4.4	Pattern evolution	60
4.5	Several traders in the same niche	66
4.6	Timescale for efficiency	67
4.7	Evolution toward higher complexity?	69
4.8	Summary and discussion.....	69
5.	Conclusions.....	71

1. Introduction

Modern financial markets are not true auctions. Transactions are made on an ongoing basis. Many agents buy and sell simultaneously, usually at different prices. Prices are often in a state of flux, and sometimes give the subjective impression of being far from equilibrium. While it is possible to model price formation as a repeated series of auctions, this is complicated and obscures the temporal dynamics of the price. It is also not an accurate description of how markets really work.

A principal motivation behind the work presented here is to construct an inherently non-equilibrium theory for price formation, in which dynamics emerge naturally and automatically, and the price at one time is easily understood in terms of the price at a previous time. The idea behind this approach was inspired in part by George Soros' principle of market reflexivity [2]. He states that "Buy and sell decisions are based on expectations about future prices, and future prices, in turn, are contingent on present buy and sell decisions". He argues that as a result, financial transactions and prices are always in flux.

I have conversed with hundreds of financial traders with a diversity of different approaches to making money. In this remarkable assortment of clever and contradictory views, I noticed one precept that all agree on. It can be expressed in the simple statement that "buying tends to drive the price up, and selling tends to drive it down." This can be viewed as a weakened statement of the usual law of supply and demand, but without the assumption of equilibrium. It suggests a dynamical feedback loop in which changes in the price cause trading decisions, that cause changes in the price, that in turn cause trading decisions ... The natural way to describe this is in terms of a dynamical system.

Putting this into mathematical form presents several problems. Since for every buyer there is a seller, does the statement above make any sense? Assuming it does, how does one choose from the infinite set of possible mathematical expressions that are consistent with it? Answering these questions is in part the cause of a four year delay in the publication of this paper¹.

The goal here is not to formulate the most realistic possible market making model, but rather to formulate the simplest model that is also reasonable. The purpose is to make a canonical model around which other models can be viewed as refinements. One of the major reasons for keeping this simple is to allow transparent analysis of complex questions, such as whether prices reflect values, how markets evolve on longer timescales, and whether markets are efficient. Once we understand the answers to these questions using the simplest reasonable model, we can return to address possible modifications under more accurate models. The theory is built in a manner that can be directly connected to market data, and that can be refined as better measurements become available.

Under this theory the price dynamics are determined by the collection of trading strategies that comprise the market. The price dynamics in turn determine whether a given strategy will make profits or losses. Thus, the profits or losses of a given strategy are deter-

1. See footnote on page 18.

mined by the collection of strategies present in the market. A market can therefore be viewed as an ecology of trading strategies, in which the fitness and survival of each strategy are determined by its relationship to other strategies.

To determine whether this theory gives sensible results a series of examples are developed. These examples are based on common trading strategies, such as trend following and value investing. The approach is empirical, similar to that used in population biology. We describe some strategies, with varying levels of realism, and study their effect on the price. We start by studying strategies one at a time and then study their interactions with each other. Because of the simplicity of this theory, several examples are easily worked analytically.

One of the most surprising results concerns value investing. Under the conventional wisdom that trading strategies create self-fulfilling prophecies, one would assume that value investing strategies should necessarily cause the price to revert to value. This is indeed true in some cases. However, there are also very reasonable value investing strategies that do not produce this behavior. Interestingly, as we go from strategies that are behaviorally unrealistic to those that are more realistic, we see that the price dynamics also become more realistic. The desire to avoid transaction costs leads to strategies that entrain price to value in a manner qualitatively similar to that observed in real markets. We show that a collection of linear value investing strategies with a diversity of perceived values are equivalent to a single value strategy with the mean value. However, for nonlinear strategies the situation is quite different: A diversity of values leads to excess volatility in the price, i.e. the price fluctuates more than the value.

We also study trend following strategies. In addition to showing that they cause trends, we see that they also cause oscillations. Value investing strategies tend to induce negative autocorrelations in price returns, and trend following strategies tend to induce positive autocorrelations. In a population containing both trend and value strategies we observe many of the characteristics of a real market, including long tailed price returns (excess kurtosis), autocorrelations in volume, correlations between volume and volatility, and correlations in volatility. Even when the autocorrelation of price returns is zero, implying no linear structure, there are periods in which trading is dominated by value investors, with negative correlations, and periods in which the trading is dominated by trend followers, with positive correlations. Thus there is nonlinear structure that can potentially be exploited to make profits.

Again following an approach commonly used in population genetics, we can study the emergence of diversity in financial ecologies. If the ecology is initially dominated by a particular strategy, other strategies can invade if they are profitable relative to the dominant strategy. For simple examples we can compute the profitability. We see that many different invasions are possible. We argue by example that we can expect a succession of new strategies to emerge, and that as a result financial ecologies spontaneously generate diversity.

Although the perception that economics should be more like biology has been present at least since the work of Alfred Marshall [1], a perusal of historical work with this intent

leaves the impression that this dream has yet to be realized (see for example the references in [38]). Although the foundations of this theory show the influences of physics, they lead to a transparent analogy between finance and biology. A trading strategy is analogous to the phenotype. The capital invested in a particular strategy determines the scale of its buying and selling, and therefore the magnitude of its effect on the price dynamics, and is analogous to the population. Under some assumptions the capital changes slowly in time relative to the price; this can be used to separate the timescales and write down equations that are analogous to the Lotka-Volterra equations. Because statistical averages over shorter timescales depend on the capital, variations in the capital over longer timescales causes apparent nonstationarity.

Under the classic theory of market efficiency, if there are any profitable patterns in the market they should disappear as they are exploited to make a profit. We investigate the case of an isolated pattern. It is possible to compute how an arbitrary pattern will evolve as the capital of the strategy exploiting it is increased. We see that patterns both spread and evolve toward earlier times, depending on the trading style that originally generated them. If a pattern is over-exploited it is pushed forward in time. This can happen either because traders fail to understand their transaction costs, or because many agents attempt to inhabit the same niche, which results in an unfavorable competitive optimum. Some estimates of timescales suggest that the evolution of capital, and hence the approach to efficiency, is measured in timescales of years or decades. This analysis leaves open the question of whether markets are ultimately efficient, but it provides some insight into how efficiency occurs and what it depends on, and places it in a dynamical context.

Although there are many difficulties to doing so in practice, if we knew the collection of strategies, and the capital of each, the approach proposed here could be used to make profits by making predictions. Since this requires knowledge of all other strategies, such a strategy is generally more algorithmically complex than all other previous strategies combined. One can imagine a succession of such strategies, each more complex than the former. Although this is unrealistic, it may be that some crude approximation of such a strategic “arms race” is partially responsible for the trend toward greater complexity in real markets.

This work has similar motivations to other recent studies using artificial markets, such as the Santa Fe Stock Market [4, 5, 15]. These studies have shown interesting results illustrating market dynamics in an evolutionary context. The Santa Fe Stock market is based on a traditional view of the market as a sequence of auctions, each of which involves an iteration of prices until buying and selling are matched. Each agent has a forecasting algorithm; to translate the predictions of the algorithm into trading it is necessary to assume all the agents employ a given utility function. There can be situations in which it is impossible to match buyers and sellers without temporarily freezing the price and incorporating new information. The process is sufficiently complicated that analysis of the results can be difficult. The Santa Fe Stock Market is based on standard principles in finance; the theory presented here offers a new approach that is simpler, allowing many problems to be addressed analytically. Time will tell whether it is more or less realistic.

The paper is divided into three substantive sections. The first, called “Force”, derives the market dynamics used here, outlines the connection to game theory, and present several background results that will be used later. The next section, “Ecology”, studies the price dynamics of strategies both alone and in combination, and tries to demonstrate that this approach gives some sensible results. In the final section, “Evolution”, we study the profitability of strategies, the emergence of diversity, and the evolution of capital through time. This section addresses market efficiency and tries to give some insight about how patterns evolve as they are exploited to make profits.

One of the main aspirations of this work is to provide a more convenient quantitative forum to address the problems raised in the new field of behavioral economics [12]. Behavioral economists have demonstrated that there are many respects in which investors are less than rational. Examples include tendencies toward overconfidence, poor ability to incorporate statistical analysis in decision making, and the intrusion of emotions on rationality. While it seems clear that such behaviors contribute to such market phenomena as excess volatility [7] or large market movements in the absence of news [30], it can be difficult to translate empirically observed behavioral characteristics into utility functions and equilibria. By making it possible to simply observe which market strategies are actually used, and compute market dynamics without such intermediate assumptions, this theory hopes to provide a convenient quantitative framework for behavioral economics¹.

1. Note that with the theory presented here it is still possible to assume utility functions, derive strategies, and then compute dynamics. The difference is that assumptions about utility functions are not necessary if the strategies are already known.

2. Force

In this section we develop a simple dynamical model for financial markets. It is based on the premise that trading has market impact, and changes in price can be regarded (at least in part) as the aggregate of the market impacts of each trade. Based on certain plausible assumptions, it is possible to derive a canonical model for market impact. This model makes several idealizations; the goal is to construct the simplest possible reasonable model, rather than to construct the most accurate model. This section also presents some background results that will be needed later, such as formulas relating to the profitability of strategies.

2.1 Market impact implies market dynamics

Trading has market impact. Buying tends to push the price up and selling tends to push it down. This goes under many names, such as “slippage”, “market friction”, or “price impact”. For trading at large size market friction is the dominant source of transaction costs. It determines an upper bound on the profitability of trading strategies. On average, the larger the order, the larger the market impact.

For every buyer there must also be a seller, so at first glance it is not obvious that this makes any sense. If the buyer in a given transaction drives the price up, why doesn’t the seller on the opposite side drive it down by the same amount? The answer is that traders usually differ in patience, which causes asymmetric price impact. A trader with an urgent need to make a transaction pays a premium to one who can be more patient. This is referred to as the “cost of immediacy”[33]. The premium for getting an immediate transaction is often paid to a *market maker*, who simultaneously offers to buy at a low price, called the *bid*, and sell at a higher price, called the *offer* or the *ask*. The difference between the offer and the bid is called the *spread*. The strategy of the market maker is to buy low and sell high, and do this repeatedly by making many “round-trip” transactions across the spread¹. A pure market maker is a patient trader who does not have a directional view and receives the spread as compensation for providing liquidity. This is contrast to a *directional trader*, who at one time may want to buy, and at some other time may want to sell, but never wants to do both at once. A given market participant may play the role of market maker at some times and that of directional trader at others, but for the results presented here we will assume that the roles are fixed.

Market making happens through diverse institutional structures. For example, in the New York Stock Exchange, for each stock there is a designated specialist who is given special privileges in return for providing liquidity. However, in most markets this niche is simply filled informally, often by competing market makers. Of course, not all trading involves a market maker; if buy and sell orders of similar size are submitted at roughly the same time, they can simply be crossed with no net price impact. Nonetheless, some fraction of the time directional traders find themselves without other directional traders to take

1. A round trip is circuit, i.e. a set of buys and sells that cancel each other out.

the opposite side, and a market maker is involved. Thus on average directional traders experience market impact.

The situation is further complicated because there are many different kinds of orders. The two most common are *market orders*, which are requests to transact immediately at the best available market price, and *limit orders*, which are requests to transact only if this can be done at a given price or better. A market order is always filled, whereas a limit order may go unfilled if the market price never crosses the limit price. A limit price that is close to the current market price has a relatively high probability of being filled, while one far away from the current price has a lower probability of being filled. Limit orders provide a continuum of different levels of patience depending on how close the limit price is to the market. A market maker may be thought of as someone who simultaneously submits limit orders both to buy and to sell, adjusting the limit prices to bracket the current price at which orders are being filled.

The price impact of an order can be measured by comparing prices before an order is placed to those after it is filled. The resulting price shift may depend on factors such as the volume of trading or the identity of the trader. Studies of market microstructure have devoted considerable effort to understanding the “information content” of different types of trades [8]. We will ignore such complications.

A particularly important factor determining price impact is the order size¹. For small orders the price shift is very noisy, with the probability of an upward price shift for a buy order only slightly higher than that of a downward shift for a sell order. As orders become larger the systematic tendency of the price to shift in the direction of the order becomes more apparent.

It is important to distinguish two types of market impact. There is the *direct impact* of each order, which alters the price of that particular transaction in relation to prices observed when the order was placed. There is also *indirect impact*: insofar as each order alters the price, this may alter the placement of subsequent orders. We will now develop a simplified model for the direct market impact. In the section on “Ecology” we will study indirect market impact, as manifested in the price dynamics.

2.2 Simple market making model

In this section we develop a simple model for market making. Complications such as competition between market makers, limit orders, or the information content of an order based on factors other than its size and direction will be neglected. The resulting model will be based on market orders only, and will assume that all trades are made with a single market maker. The strategy is to simplify the description of market making in order to

1. Chan and Lakonishok [10] observed that the identity of the trader is more important than the order size in determining price impact. However, their sample consisted of trades from many different institutions, with heterogeneous trading styles and different levels of patience. The level of patience is clearly very important in determining market impact -- there is a clear trade-off between getting a trade done soon vs. getting it done at a good price. But for a given level of patience, the order size should be the most important factor.

study the complexity caused by directional trading. The purpose is to construct a theory simple enough to answer difficult questions.

2.2.1 Derivation of basic model

Assume that all trading occurs via market orders filled by the market maker. Suppose a directional trader places an order $\omega(t)$ based on the midpoint price $x(t) > 0$. Buy (positive) orders are filled at price $\tilde{x}(t) + s(t)/2$, and sell (negative) orders at price $\tilde{x}(t) - s(t)/2$, where $s(t) > 0$ is the spread. The new midpoint price where the transaction takes place at time t can be written

$$\tilde{x}(t) = \tilde{x}(x(t), \omega(t), S(t)), \quad (\text{Eq 1})$$

where $S(t)$ is the internal state of the market maker, which may depend on past trading history. The *position* is the net holding of the asset, i.e. if the initial holding is zero, since all orders are filled, the position is the sum of all previous orders.

In the spirit of getting the simplest model that gives reasonable market dynamics, we will not attempt to model the temporal behavior of the spread. In fact, unless otherwise stated we will assume $s(t) = 0$. This is unrealistic, but the ratio of the spread to the midpoint price is usually small, and the midpoint price has a much larger affect on the dynamics. The main effect of the spread is on the profits and losses, which may have a significant impact on strategy selection, as discussed in Section 4. Modeling the spread would add complications that are not crucial to the main results.

To develop a canonical model for price setting by the market maker we impose the following conditions. The first four should not be controversial:

1. *The price is always positive.*

$$\tilde{x}(x, \omega, S) > 0 \text{ for } x > 0.$$

2. *The price is always finite.*

3. *\tilde{x} is an increasing function of ω .* This means that price impact is in the direction of the order and increases with order size.

4. *If there are no orders there is no market impact.*

$$\tilde{x}(x, 0, S) = x.$$

The first four conditions are already sufficient to eliminate many possibilities. For example, the relation $\tilde{x} = x(1 + a\omega^b)$, where $a > 0$ and b is a positive odd integer, can be eliminated because for a sufficiently negative ω the price can become negative. Similarly, for the relation $\tilde{x} = x/(1 - a\omega^b)$ the price can become infinite with a single large positive order. This indicates that \tilde{x} must increase at least as fast as e^{ω^a} , i.e as fast as an exponential to a power of the order.

We now add some stronger assumptions.

5. *The price $x(t)$ is a continuously valued variable* (that makes discontinuous jumps immediately when orders are placed). Though the quantization of transaction levels is important for some problems, it is not important for those considered here.
6. *There is no dependence on the internal state S of the market maker.* This is clearly not true for real market makers. As the market maker trades she tends to accumulate a net long (positive) or short (negative) position and makes price adjustments to compensate. Alternatively, a real market maker may act as a directional trader to unload the position, and generate additional market impact. Factors such as recent volume and volatility may have an influence in determining the steepness of the price response to an order of a given size. Nonetheless, including the market makers' state introduces complications that we wish to avoid at this point; this assumption will be re-examined in Section 2.3.1.
7. *It is not possible to make profits by repeatedly trading through a circuit.* A circuit is a sequence of trades that sum to zero, also called a "round-trip". If a circuit does not return the price to its original value, it becomes possible to take a net position and make arbitrarily large profits by manipulating the price by repeatedly executing the circuit. Any market maker who does not prevent this should go out of business very quickly. We show below that this leads to the conclusion that \tilde{x} must satisfy the *additivity* condition

$$\tilde{x}(\tilde{x}(x, \omega_1), \omega_2) = \tilde{x}(x, \omega_1 + \omega_2). \quad (\text{Eq 2})$$

This says that the direct market impact of two orders is the same as that of a single order equal to their sum. It implies that the direct market impact is independent of the order of transactions and is unaffected by order splitting or merging.

8. *The ratio of the prices before and after a transaction is a function of ω alone.*

$$\frac{\tilde{x}}{x} = \phi(\omega). \quad (\text{Eq 3})$$

From (3), ϕ must be an increasing function.

The last two assumptions uniquely determine the market impact function. Plugging equation 3 into equation 2 gives

$$\phi(\omega_1 + \omega_2) = \phi(\omega_1)\phi(\omega_2).$$

This functional equation has the solution

$$\tilde{x} = x e^{\omega/\lambda} \quad \text{or} \quad \log \tilde{x} - \log x = \frac{\omega}{\lambda}. \quad (\text{Eq 4})$$

λ is a scale factor that normalizes the order size, and will be called the *liquidity*. It determines how much the price changes for an order of a given size. The liquidity is measured in the same units as orders, e.g. if the orders are measured in dollars, the liquidity is in dollars. If the liquidity is a billion dollars, an order of a billion dollars will cause the price to

increase by a factor of e . For convenience we will initially assume the liquidity is constant, although we will also consider the case where it varies in time later.

We will now derive the additivity condition for market impact, equation 2, from assumption (7). The simplest circuit is composed of alternately buying and selling, i.e. of executing trades $(\omega, -\omega)$. If this results in a net increase in price then

$$\tilde{x}(\tilde{x}(x, \omega), -\omega) > x.$$

If this is true then it is possible to make arbitrarily large profits by taking a net long (positive) position y , and ratcheting the price upward by alternately buying and selling. Similarly, if buying and selling results in a net decrease in price, arbitrarily large profits are possible by taking a net short (negative) position and ratcheting the price downward. Thus assumption (7) implies that

$$\tilde{x}(\tilde{x}(x, \omega), -\omega) = x.$$

Since by assumption \tilde{x} is an increasing function of ω , for fixed x its inverse \tilde{x}^{-1} exists and we can take the inverse of both sides of the expression above, which gives

$$\tilde{x}(x, \omega) = \tilde{x}^{-1}(x, -\omega). \quad (\text{Eq 5})$$

This says that buying and selling have inverse market impact.

Now consider the more complicated circuit $(\omega_1, \omega_2, -(\omega_1 + \omega_2))$. Under the same reasoning used above, arbitrarily large profits are possible unless

$$\tilde{x}(\tilde{x}(\tilde{x}(x, \omega_1), \omega_2), -(\omega_1 + \omega_2)) = x.$$

Taking the inverse of both sides and using equation 5 gives equation 2. Since any sequence of trades can be decomposed into a series of trades in this form, this implies that the net price change from any circuit is zero, and that the total direct market impact due to any set of orders is invariant under permutations. Bear in mind that this is only true for the direct impact; since orders typically depend on the price, the orders that are actually placed *do* depend very much on the sequence, and thus the indirect market impact may be quite sensitive to the sequence of orders.

The market maker can block profits from trading through a circuit by making the spread large enough. Suppose the market impact satisfies the assumptions above without satisfying equation 2. The spread must be greater than the fractional price change in making a round trip, i.e.

$$s > \frac{\tilde{x}}{x} - 1 = |\phi(\omega_1)\phi(\omega_2)\phi(-(\omega_1 + \omega_2)) - 1|$$

for any (ω_1, ω_2) . This places a lower bound on the spread that depends on order size. Under the additivity assumption this lower bound is zero. Insofar as the spread is not zero the additivity assumption may be violated.

Note that the while the additivity condition implies the direct market impact is additive, it does not imply that the average transaction cost is additive. We will refer to a market impact function that increases more slowly than exponential as superadditive, and one that increases more rapidly as subadditive, as illustrated in Figure 1. When ϕ is additive or

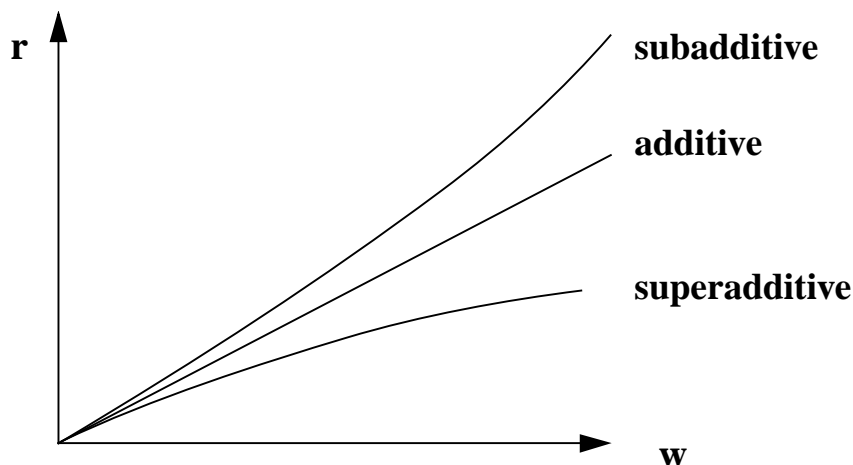


FIGURE 1. A comparison of possible market impact functions. The log-return of the market impact is plotted vs. the order size ω . The function derived here is additive and linear with slope $1/\lambda$. The market impact is super-additive if it increases slower than linearly, which implies that two small orders have more impact than a single large order. The market impact is sub-additive if it increases faster than linearly, which implies two small orders have less impact than a single order.

subadditive, it is generally possible to reduce direct transaction cost by splitting orders. This can be seen from the relation

$$\frac{x\omega_1\phi(\omega_1) + x\omega_2\phi(\omega_1)\phi(\omega_2)}{\omega_1 + \omega_2} \leq x\phi(\omega_1 + \omega_2). \quad (\text{Eq 6})$$

When either $\omega_1 > 0$ and $\omega_2 > 0$, or $\omega_1 < 0$ and $\omega_2 < 0$, the left side is the mean transaction price for order ω_1 followed by order ω_2 . The right side is the transaction price for a single order $\omega_1 + \omega_2$. If ϕ is additive or subadditive this inequality is easy to derive¹. Thus order splitting results in more favorable prices. This is consistent with what one would expect: By being more patient it is possible to reduce direct transaction cost, but only at the cost of delays in execution and a possible increase in *indirect* market impact.

1. Equation 6 can be derived by multiplying by $\omega_1 + \omega_2$ and proving inequalities for the ω_1 and ω_2 terms separately. For the ω_1 term the inequality follows because ϕ is increasing, and for the ω_2 term it follows from the assumption of additivity or subadditivity. For superadditive functions the inequality may go either way depending on values of ω . One may ask whether there is a function that satisfies equation 6 as an equality; however, expanding in a power series shows that the only analytic solutions are $\phi = 0$ and $\phi = 1$, which do not satisfy assumption (7).

Equality holds only if ϕ is additive and either $\omega_1 = 0$ or $\omega_2 = 0$. If ϕ were a superadditive function with the opposite inequality, patience would be punished rather than rewarded.

Because of the assumption of no state dependence, this market impact function cannot accurately model that of a real market maker. However, under the assumptions above it should describe the market impact for a trader who is ignorant of the market maker's state. It provides a reasonable starting point for developing a dynamical model, and is the market impact function that will be used throughout most of what follows.

2.2.2 Relation to supply and demand

Under certain conditions this market impact function is related to supply and demand. Assume an increasing supply function $S(z)$ and a decreasing demand function $D(z)$, where $z = \log x$. At equilibrium these are equal, i.e. $S(z_e) = D(z_e)$. If the demand changes by δD and the supply changes by δS , as shown in Figure 2, then to first order it

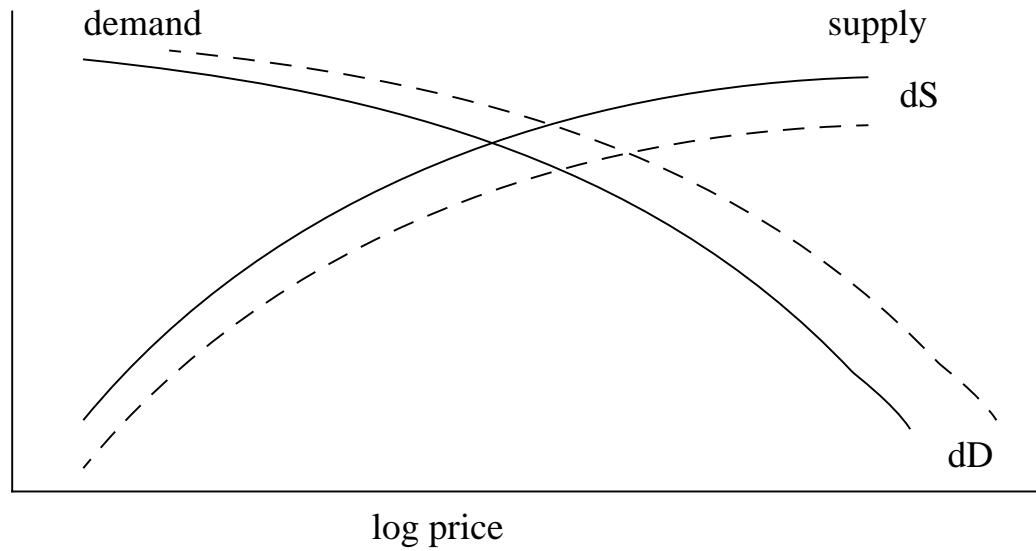


FIGURE 2. A variation in supply and demand leads to a new equilibrium price. This is equal to the market impact only when the liquidity λ satisfies special conditions.

is easy to show that

$$\delta z_e = \frac{-(\delta S - \delta D)}{(S'(z_e) - D'(z_e))}.$$

If we set $\omega = \delta D - \delta S$ and $\lambda(z) = S'(z) - D'(z)$ we see that this is consistent with classical arguments based on supply and demand. However, in the context of this model, changes in supply minus demand are caused by impatient traders, while the liquidity is set by the market maker. They are different agents. Thus in Section 3.2.1 we will argue that the condition for λ is generally not satisfied. Whether this is a reasonable approximation

depends on the whether the liquidity and capital co-evolve over the long term to make the market efficient.

2.2.3 Relationship to market making with limit orders

We can get some insight into the liquidity by comparing this to a market with limit orders as well as market orders. As before, let prices be continuous, and let the density of the dollar value of limit orders at different price levels be $d(x)$. When a market order ω is received, it is crossed with unfilled limit orders that are of opposite sign to the market order, beginning with those closest to the existing price. The resulting shift in the price can be found from the condition

$$\omega = \int_x^{\tilde{x}} d(x') dx'. \quad (\text{Eq 7})$$

A market maker can be thought of as patient trader with no directional view who submits both buy and sell limit orders symmetrically about the current price¹. Since only relative price changes are important, and since the price has to be positive, the density $d(x) = 1/(\lambda x)$ is a natural choice. Substituting this into equation 7 implies equation 4. Even if this density does not hold exactly, it may still be a good approximation when ω is sufficiently small and the limit order density $d(x)$ is sufficiently smooth². In a market where most of the traders use limit orders and only a few use market orders, this shows that the liquidity of market orders is proportional to the volume of limit orders.

2.2.4 Relation to prior research in market making

There is a rich literature on market making, which is reviewed in O'Hara [8]. The model presented here is in the spirit of Demsetz [33], who discussed the cost of immediacy and manner in which market making differs from a traditional Walrasian auction, and empirically investigated the size of the spread in relation to volume. Current thinking classifies contributions to the spread and the dynamics of the midpoint price based on order processing costs, adverse information [34], and inventory effects [16]. Order processing costs are simply the charges incurred for handling transactions; adverse information occurs because directional traders may possess information that market maker makers do not, and that tends to reduce their profits; inventory effects occur because of the market

1. This assumes that as market orders are filled the limit orders removed from the book will be immediately replaced. This is consistent with the limit order traders having no risk aversion and no directional view. If there are time lags in replacing limit orders, filling of a market order will leave a "hole" in the book, so that if it is followed by another market order of the opposite sign, it will have to "jump the hole" to get filled. Such behavior will cause mean reversion of the price and make price changes dependent on previous price changes. This is consistent with results of Huang and Stoll [19], who analyzed tick by tick transaction data from the NYSE and found a negative correlation to past price changes.

2. One potential problem is the correlation between market orders and limit orders. This is particularly a problem for limit orders near the current price. Such orders are placed by the least patient limit order traders, who are most like market order traders. They also have the largest effect on the execution of market orders.

marker's aversion to risk and their desire to keep their inventory (net position) as low as possible.

Under the model derived here the price is manipulated in the direction of the net of incoming orders. This can be motivated both by the desire to deal with adverse information and to reduce inventory. The assumption of no state dependence means that there is no explicit dependence on the inventory. This is clearly an approximation that should be regarded as a simple starting point. Extensions are discussed in the next section.

2.3 Time varying liquidity

In the preceding derivation we made several assumptions. The most important is that that of no state dependence. The resulting solution has constant liquidity and symmetry between buying and selling. In this section we discuss empirical approaches to relaxing these assumptions.

We can allow the liquidity to vary in time by making the ansatz that ϕ depends on $w = \omega/\lambda(t)$ rather than on ω . The market impact is then

$$\tilde{x} = x e^{\omega/\lambda(t)}. \quad (\text{Eq 8})$$

Time variations in liquidity can be driven by factors such as the market maker's inventory, volume, volatility, or asymmetries in the market.

2.3.1 Inventory effect

Without explicit risk aversion the market maker will in some circumstances build up a substantial net position, or *inventory*. For example, if there is temporarily an excess of buyers the market maker will do an excess of selling and accumulate a negative position. Real market makers are typically highly risk averse and attempt to manipulate the price to keep their positions small [16, 17, 18]. If the market maker accumulates a negative position, she will raise the price more than usual to encourage selling; similarly for a large positive position she will lower the price more than usual to encourage buying. Assuming equation 8, this implies a liquidity function of the form

$$\lambda(t) = \bar{\lambda} f(\mathfrak{t}), \quad (\text{Eq 9})$$

where \mathfrak{t} is the market maker's inventory and $\bar{\lambda}$ is the liquidity when the inventory is zero, and f is an increasing function with $f(0) = 1$. As an example, let

$$\lambda(t) = \frac{\bar{\lambda}}{(1 \mp b\mathfrak{t})}. \quad (\text{Eq 10})$$

$b > 0$ is a constant that must be chosen so that $|b\mathfrak{t}| < \bar{\lambda}$ if we want to ensure that the market impact is always an increasing function of ω . The minus sign applies when the order is positive, and the plus sign when the order is negative. This differs from the form used by Huang and Stoll [19], in that I assume that the inventory effect is felt only through changes

in liquidity, whereas they assume that the market maker's inventory causes the price to change with or without trading. We will see later that when there are asymmetries in the market the inventory effect is needed to ensure that the market behaves sensibly. Recent studies of market data make it clear that the inventory effect is important [35].

Note that under any version of the inventory effect the market maker responds asymmetrically to buy and sell orders. This can potentially be exploited to make profits by repeatedly trading through a circuit as discussed under assumption (7) in Section 2.2.1. The market maker can prevent this by raising the spread accordingly. This suggests that if the market impact depends on the inventory then the spread does also. Also, note that it is not possible to make profits this way by manipulating the inventory only by self-trading. That is, if a given trader develops a positive position by buying, the market maker develops a corresponding negative position; under the inventory effect, trading through a circuit drives the price down, causing that trader losses. It is only possible to profit by trading through a circuit if a trader is able to detect that the market maker has accumulated an inventory due to the trading of others. It would be interesting to attempt to derive the proper form of f in equation 9, but this is beyond the scope of this paper.

2.3.2 Asymmetric markets

A market is *symmetric* if there is no *a priori* difference between buy and sell orders. Currency markets are a good example. The American stock market, in contrast, is an asymmetric market. There are two kinds of asymmetries. The first comes from the tendency of the market to go up. The second comes from regulatory restrictions that may make it easier to make a transaction in one direction than in the other. In the derivation so far, to keep things simple we assumed market symmetry.

Regulatory restrictions can be used to alter the liquidities for different types of orders. For example, in the American stock market a sale is defined to be a short sale if the resulting position is net short (negative). Normal sales receive preference over short sales, e.g. short sales must go after normal sales at the same price level. This lowers the liquidity for short sales relative to normal sales. The assumption that profits cannot be made by trading through a circuit places limitations on the relative liquidities for buying, normal sales, and short sales. Since such profits could potentially be made from a long position by alternating buying and normal selling, the market impact for normal sales should be at least as great as that of buying (within the spread) to keep the price from drifting up. Similarly, from a short position this requires alternate buying and short selling; the market impact of a buy should be at least as great as that of a short sale to keep the price from drifting down.

One could argue from a different point of view that if the number of buyers and sellers is roughly equal, the overall liquidity of buying and selling should be equal, and to the extent that short selling is made less liquid, normal selling becomes more liquid. In practice the spreads are probably large enough to absorb differences in market impact and prevent traders from profiting by trading through circuits. If there are differences in the overall liquidity for buy and sell orders, under the ansatz of equation 8 we can expect a discontinuity in the derivative of ϕ at $\omega = 0$. This was observed by Chan and Lakon-

ishok [10], who studied market impact in the American stock market¹. Unless otherwise stated we will assume the market is symmetric.

2.4 Dynamics

This section uses the market impact function derived so far to develop a dynamical system that describes the feedback loop between the placement of orders and changes in the price. Assume that there are N directional traders who place orders $\omega^{(i)}(t)$, where i is an index labeling the trader. The function or algorithm that each trader uses to place orders can be thought of as his *trading strategy*. Strategies may depend on the price, price history, and external information $I(t)$. The external information can be anything that is believed to be relevant to forecasting the price, such as fundamental information for stocks, weather for commodities, or purchasing power parity for currencies. It also can be something as simple as the day of the week or a random number. Since $I(t)$ can be random, the trading strategies can also have random components. $I(t)$ may be trader-specific, in which case we will write it $I^{(i)}(t)$. Trader i in general cannot observe the orders of trader $j \neq i$, although it may be possible for him to infer or partially infer this from the past behavior of the price. There may be several traders using the same strategy, but unless otherwise stated we will assume that the strategies of each trader are distinct. The strategies may be arbitrarily complex, though in some cases it may be useful to decompose them into “sub-strategies”, for example if a given trader uses different sub-strategies at different times. Strategies must be causal, i.e they can only depend on present and past information.

In real markets there is a characteristic time lag Δt from when an order is submitted until its effect on the price is observed. This means that detailed time ordering relationships on timescales faster than this are likely to be violated. This makes it reasonable to synchronize the trading, introducing an explicit lag between when information is observed and when transactions take place. By choosing the units of time appropriately, without loss of generality we can let $\Delta t = 1$. At the beginning of a timestep t , all the traders observe the price x_t , or equivalently, log price z_t . Each trader submits orders $\omega_t^{(i)}$. The market maker applies a market making algorithm, e.g. equation 4, and publishes a new price x_{t+1} . All the orders are filled at this new price.

The price setting of the market maker may be influenced by factors other than order flow. Information from outside the market may be received that indicates the price should be adjusted to another value. Examples of such events are news announcements or perceived arbitrage possibilities with a related market. Such behavior can be taken into account by adding a random term ξ_t to the new price. Alternatively this term can be thought of as “noise trading”, i.e. trading that is not price driven and occurs at random.

The dynamics can be written:

1. For American stock market data Chan and Lakonishok [10] report a discontinuity in market impact between buying and selling. They note that this cannot be accounted for by the tendency of the market to drift upward. An alternate possibility is that this is caused by the short sale restrictions in the American stock market, which also affect the liquidity of normal sales. If this is true, international markets in which short sale restrictions do not apply should not show such a strong asymmetry in liquidity.

$$r_{t+1} = \log x_{t+1} - \log x_t = z_{t+1} - z_t = \frac{1}{\lambda} \sum_{i=1}^N \omega^{(i)}(z_t, z_{t-1}, \dots, I_t) + \xi_t. \quad (\text{Eq 11})$$

The dynamics of equation 11 are very general. Depending on the collection of strategies $\omega^{(i)}$, they can be stable or unstable, and contain fixed points, limit cycles, or chaos. The market is a blank slate on which the collection of trading strategies write the dynamics¹.

The effective timescale Δt for the dynamics above depends on the population of trading strategies. If all the traders closely observe the intraday data stream and place orders very frequently, the timescale for a single iteration can sensibly be as small as a few minutes. Many traders, however, observe the dynamics only on timescales of a day or longer. Thus to model their behavior it is sensible to regard the timescale for the dynamics as a day or more. In reality there will be a mixture of strategies on different timescales and the market dynamics will reflect this.

For some purposes it is convenient to approximate this still further as a continuous time differential equation. On a daily timescale it is typically the case that the log returns are less than a percent, i.e. that $|r_t| < 0.01$; on an intraday timescale the returns are correspondingly smaller. If we assume that price movements and noise terms are a continuous time random process, we can take the limit where the spacing between timesteps goes to zero. In this case ω must be interpreted as a continuous order flow, and λ must be thought of as the rate at which the market maker adjusts the logarithm of the price. Equation 11 can be re-written

$$\frac{d \log x}{dt} = \frac{1}{x(t)} \frac{dx}{dt} = \frac{1}{\lambda} \sum_{i=1}^N \omega^{(i)}(x(t), x(t-\tau_1), \dots, x(t-\tau_m), I(t)) + \xi(t). \quad (\text{Eq 12})$$

where $\tau_1, \dots, \tau_m > 0$ are arbitrary time lags. Strictly speaking the continuous time model is less accurate than the discrete time model, since trading is an inherently discontinuous process, and in the limit as $\Delta t \rightarrow 0$ there will rarely be any orders. However, continuous time models are more convenient for some purposes.

The *position* y is the cumulative sum of the orders,

1. While I conceived of this idea and wrote a preliminary manuscript in 1994, I have waited four years to publish. In part this was because of other commitments, but also because at that time I was unable to give solid arguments for the correct form of the market impact function. Since then two papers have appeared presenting the basic idea of a dynamical system with a non-equilibrium version of supply and demand. One of these, Bouchaud and Rama [28] (January 1998), graciously acknowledges an oral presentation I gave in Paris in June, 1997. Their development in the first half of their paper parallels the one I presented at that time, except that they formulate the market impact in terms of the price rather than its logarithm. In the second half of their paper they present interesting results suggesting that volatility constraints may induce crashes. Another paper by Caldarelli et al. [29] (1997) presents interesting results, but these are based on an apparently *ad hoc* market model. The models in these papers do not satisfy the conditions of Section 2.2.1.

$$y_t^{(i)} = \sum_{j=1}^t \omega_j^{(i)} + y_0^{(i)}. \quad (\text{Eq 13})$$

For convenience we will generally assume $y_0^{(i)} = 0$. Similarly the orders can be written in terms of the positions as

$$\omega_t = y_t - y_{t-1}.$$

One can think about a strategy either in terms of its positions or in terms of its orders. The orders determine the market impact, but the positions determine the profits and losses.

An important property of a strategy is its scale. We will typically write strategies in the form

$$\omega^{(i)} = c^{(i)} \tilde{\omega}^{(i)}, \quad (\text{Eq 14})$$

where $\tilde{\omega}^{(i)}$ is a fixed function and $c^{(i)}$ is a parameter that controls the scale. This implies a similar relation for the position, $y_t^{(i)} = c^{(i)} \tilde{y}_t^{(i)}$, where \tilde{y} is defined as in equation 13 with ω replaced by $\tilde{\omega}$. Thus $c^{(i)}$ determines not just the size of the orders, but also the size of the positions, and is proportional to the capital at risk. For any given level of risk tolerance, it is also proportional to the “funds under management”. We could impose an arbitrary definition, for example, by defining the capital as the time average of the absolute value of the positions, and scaling $\tilde{\omega}$ accordingly. This would complicate things later on, however, since $\tilde{\omega}^{(i)}$ and $\tilde{y}^{(i)}$ generally depend on prices, which depend on other strategies as well. For convenience, we will simply refer to $c^{(i)}$ as the *capital* of the i^{th} strategy, bearing in mind that since two different fixed functions $\tilde{\omega}$ may have different risk levels, this is just a proportionality.

In the studies of ecology in the next section the strategies and their capital will be fixed in any given simulation. In the section on evolution we will investigate the longer term dynamics that occur when the capital is allowed to vary in time, e.g. under reinvestment.

2.5 Profits, losses, and game theory

If we assume that the asset does not make payments such as dividends or coupons and neglect the spread, the profit or loss¹, or the *gain* of the i^{th} strategy is

$$g_t^{(i)} = \frac{\Delta x_t}{x_{t-1}} y_{t-1}^{(i)} \quad (\text{Eq 15})$$

1. This includes both realized and unrealized gains, i.e. it includes the value of positions marked to the current midpoint price. There are many markets, such as currencies and commodities, where there are no payments (such as dividends or coupons). Payments affect profits and losses but do not directly affect market dynamics. Studying how payments alter the population of strategies is an interesting problem that is beyond the scope of this paper.

where $\Delta x_t = x_t - x_{t-1}$ is the change in price, and $y_t^{(i)}$ is the position taken by strategy i at timestep t . $\Delta x_t/x_{t-1}$ is commonly called the *return*, and $r_t = \log x_t - \log x_{t-1}$ is called the *log-return*. When the returns are small $\Delta x_t/x_{t-1} \approx r_t$ and

$$g_t^{(i)} \approx r_t y_{t-1}^{(i)}. \quad (\text{Eq 16})$$

Substituting for r_t from equation 11 gives

$$g_t^{(i)} \approx \left(\frac{1}{\lambda} \sum_{j=1}^N \omega_t^{(j)} + \xi_t \right) y_{t-1}^{(i)}. \quad (\text{Eq 17})$$

If the noise ξ is uncorrelated with the position y , taking time averages gives

$$\langle g^{(i)} \rangle \approx \frac{1}{\lambda} \sum_{j=1}^N \langle \omega_t^{(j)} y_{t-1}^{(i)} \rangle + \mu \langle y^{(i)} \rangle. \quad (\text{Eq 18})$$

$\langle \rangle$ denotes a time average, and $\mu = \langle \xi \rangle$ is the mean value of the noise, often called the *drift term*. If we define the *gain matrix*

$$G_{ij} = \frac{1}{\lambda} \langle \omega_t^{(j)} y_{t-1}^{(i)} \rangle$$

then the mean gain for strategy i can be written

$$\langle g^{(i)} \rangle \approx \sum_{j=1}^N G_{ij} + \mu \langle y^{(i)} \rangle. \quad (\text{Eq 19})$$

The gain matrix G_{ij} gives the approximate amount that strategy i wins or loses due to the price movements induced by strategy j . It is generally asymmetric. The last term corresponds to profits that may be made from the long-term tendency of the market to move up or down.

The dynamics together with the definition of the gain matrix define a game with continuous payoffs and continuous states, and discrete or continuous time (depending on whether we use equation 11 or equation 12). ω is the “move”, and G is the payoff matrix (this is approximate for discrete time dynamics, and exact for continuous dynamics). Letting

$$\tilde{G}_{ij} = \frac{1}{\lambda} \langle \tilde{\omega}_t^{(j)} \tilde{y}_{t-1}^{(i)} \rangle,$$

where $\tilde{\omega}$ and \tilde{y} are scale-independent versions of the strategies and positions as defined in equation 14, we can write the gains in the form

$$\langle g^{(i)} \rangle \approx \sum_{j=1}^N \tilde{G}_{ij} c^{(i)} c^{(j)} + \mu \langle \tilde{y}^{(i)} \rangle c^{(i)}. \quad (\text{Eq 20})$$

This will be useful later when we study the evolution of the market.

The market can be viewed as a pure anticipatory game. The market maker plays the role of the casino. Each player attempts to forecast the aggregate action of the other players and bets accordingly. Players that make accurate forecasts are rewarded and those that make poor forecasts are penalized. The average player tends to lose money to the market maker. However, if a player is good enough, under some circumstances it may be possible to anticipate the other players well enough to overcome the “house edge” and make a profit.

2.6 Market friction

Market friction refers to the fact that uninformed transactions tend to produce losses due to market impact. Because of the tendency for an order to push the market away from it, this is true even when the bid-ask spread is zero. Market friction corresponds to the diagonal elements of the gain matrix, which are generally negative. To see this for convenience assume $\langle y \rangle = 0$. The diagonal elements are

$$G_{ii} = \frac{1}{\lambda} \langle \omega_t^{(i)} y_{t-1}^{(i)} \rangle = \frac{1}{\lambda} \langle (y_t^{(i)} - y_{t-1}^{(i)}) y_{t-1}^{(i)} \rangle = \frac{1}{\lambda} (\rho_y(1) - 1) \sigma_y^2, \quad (\text{Eq 21})$$

where $\rho_y(1)$ is the first autocorrelation and σ_y is the standard deviation of y . Since $\rho_y(1) \leq 1$, the diagonal elements are less than or equal to zero.

For finite transactions market friction is path dependent. As an example, consider the case where there is only one trader ($N = 1$). Assume a starting position $y_0 = 0$, and suppose the trader buys and then immediately sells, e.g. $\omega_1 = \omega$ and $\omega_2 = -\omega$, with $\omega > 0$. In the absence of noise, from equation 15, since $\Delta x_t / x_{t-1} = e^{r_t} - 1$, the gain is

$$g = (e^{-\omega/\lambda} - 1)\omega \approx -\frac{\omega^2}{\lambda} + \frac{\omega^3}{2\lambda^2}.$$

If instead the trader first sells and then buys, the gain is

$$g = -(e^{\omega/\lambda} - 1)\omega \approx -\frac{\omega^2}{\lambda} - \frac{\omega^3}{2\lambda^2}.$$

For finite transactions the loss in the two cases is different.

In fact, it is possible to show that the path dependence of the market friction for finite size trades is generally incompatible with any form of the market impact function that has the properties that the market impact function increases with ω and that two successive trades that sum to zero return the price to its starting value. To prove this, compute the gain

for the sequence $(\omega, -\omega)$ and set it equal to the gain from $(-\omega, \omega)$, where $\omega > 0$, assuming the same final value x_f for the price. Equation 1 and equation 15 imply

$$\frac{(x_f - \tilde{x}(x, \omega, S))\omega}{\tilde{x}(x, \omega, S)} = \frac{-(x_f - \tilde{x}(x, -\omega, S))\omega}{\tilde{x}(x, -\omega, S)}.$$

A little algebra gives

$$(\tilde{x}(x, \omega, S) + \tilde{x}(x, -\omega, S))x_f = 2\tilde{x}(x, \omega, S)\tilde{x}(x, -\omega, S).$$

This equation is symmetric in ω , so the solution must also be symmetric in ω . This is incompatible with the requirement that \tilde{x} is an increasing function of ω . Thus we see that the path dependence of the market friction is a very general property.

2.7 Other market forces

In general the price may change due to events other than receipt of orders. A clear example occurs in markets that have arbitrage relationships, such as the currency futures market in Chicago and the intrabank currency market. Since a futures contract can be converted into the underlying currency, it necessarily maintains a relationship to the exchange rate in the intrabank market. If a large price change occurs in the intrabank market a large change will likely occur in the futures market, even without any transactions taking place, simply because everyone knows this arbitrage is possible. Within a range corresponding to the transaction cost to perform the arbitrage, the futures price will tend to remain close to the intrabank price. From the point of view of someone in the futures pit, the intrabank market appears to exert an outside “force” on the futures price.

Another example is news. If the probable impact of a news item on the price is clear in advance, receipt of news can impact the price even before any transactions take place. In the futures pits a major news event may cause trading to halt while everyone attempts to understand the impact of the new information. The market makers widen their spreads, and trading may eventually resume at a price with a significant gap from the previous price. Again, it is as though the news exerted a “force” on the price that caused it to change.

If we let f denote the aggregate of any such “forces”, the price change between time t_a and time t_b can generally be expressed as an equation of the form

$$\log x(t_b) - \log x(t_a) = \int_{t_a}^{t_b} f(t, x(t), \dots) dt.$$

For example, with the approximations given in Section 2.2 the force caused by an order $\omega(t_i)$ can be written $f(t) = \omega(t_i)\delta(t - t_i)$, where δ may be thought of as an infinitely sharp pulse satisfying $\delta(z) = 0$ for $z \neq 0$ and $\int \delta(z) dz = 1$. Thus the change in price can be called the “market impact”, and the kernel f causing the change in price the “market force”. The force f can be either deterministic or stochastic.

2.8 Experimental verification

One of the advantages of this approach is that the market impact, which is the foundation of the theory, is something that can be measured directly. It is thus possible to determine how closely the canonical market making rule of equation 4 describes real markets, and to modify it as needed. All of the analyses performed in this paper can be repeated, at least numerically, with more general market making rules. One of the goals of measuring the market impact is not just to determine its shape, but also the magnitude of its effect. What fraction of price movements can be accounted for as the aggregate of market dynamics?

Market impact can be measured directly by measuring prices before an order is placed and then measuring them again after a transaction based on the order has occurred. Some attempts to measure this function have already been made [9, 10, 11, 27, 13, 14]. The results so far make it clear that market impact is an increasing function of order size, but are too noisy to determine its functional form. While ultimately it should be possible to measure this accurately, there are several problems in doing so:

- In public data sets the identity of the traders is generally unknown. This means that the level of patience of the two parties in the transaction is unknown.
- Data sets that give orders as well as transaction prices are typically proprietary. Particular trading strategies may use market timing rules that make the results atypical. If proprietary data sets are mixed together, the level of patience and order tactics of different firms or different traders may be quite different.
- Between when a given order is placed and when it is filled, many other orders may be received and filled. As a result the market impact of any particular order looks very noisy. This obscures the basic effect and makes it more difficult to estimate the fraction of price changes that are accounted for by market impact. This can be resolved by using data sets containing the orders and transactions of all market participants.
- A significant fraction of orders are limit orders. These also have market impact.
- Many traders split large orders to reduce transaction costs. Such order splitting can be spread over months [10]. Large orders are the most useful for determining market impact. However, large orders are precisely those that tend to get split, and are most likely to be limit orders rather than market orders.
- The total transaction cost, which includes the bid-ask spread, is easily confused with market impact. It is generally necessary to either compare the prices before an order is placed to those of *other* transactions after that order is filled, or to make adjustments for the bid-ask spread. In many data sets, for example futures, individual transactions can be lumped together and it is difficult to distinguish one's own transactions from those of others. To first approximation the bid-ask spread is a step function; when it is confused with market impact, the result tends to be sigmoidal.

It is beyond the scope of this paper to measure the market impact function, but I hope that the theoretical implications of measuring this function accurately as developed here will stimulate more work in this area. This theory also predicts relationships between volatility and volume on different timescales, but this is a topic for future research.

2.9 Summary and discussion

The main result of this section is to show that under plausible simplifying assumptions it is possible to derive a unique form of the market impact. Furthermore, this can be used to make a dynamical formulation of price formation. Within this framework it is natural to regard the market as a continuous game with continuous payoffs.

This approach has the advantage that the foundation of the theory, the market impact function, can be measured directly. Even if the form of the market impact that I have derived turns out to be wrong, all the results presented here can be easily revisited, at least numerically, with any market impact function.

3. Ecology

Under the model established in the previous section the market dynamics are determined by the collection of trading strategies that comprise the market. We begin by studying a few representative trading strategies individually, and then study them in concert. On its own each strategy creates a characteristic *induced market dynamics*, which provides a good starting point to understand the dynamics when they are together.

The approach taken here is that of an ecologist. We observe and classify the strategies that exist from an empirical point of view, and study their induced dynamics and their interactions with each other. At the highest level we will distinguish three types of strategies:

- Value investing.
- Purely temporal (e.g. the January effect).
- Trend following.

Although there are exceptions, we will show that value investing strategies typically induce negative autocorrelations in the log-returns, and trend following strategies induce positive correlations. Purely temporal strategies do not depend on prices, and are neutral in this respect.

One of the main purposes of this section is to determine whether this theory makes any sense. What causes a market to be stable or unstable? Do prices reflect value? All the strategies have regions of stability and instability; some value strategies help prices reflect values, and some don't. Commonly observed market phenomena such as long tails in the distribution of log-returns, correlated volume and volatility, and oscillations between price and value, are natural consequences that occur for broad classes of strategies.

Throughout this section we will assume that the strategies are fixed throughout a given simulation. Profits are not reinvested. The consequences of reinvestment and other capital reallocations will be studied in the section on evolution.

3.1 January effect

The famous "January effect" is just one example of a situation in which regularities in cash flow causes temporal patterns in the price. Such cash flows may be largely independent of price, for example, if they are driven by external rhythms of events such as taxes or annual bonuses. Because it is so simple the January effect serves as a good first example to illustrate how easy it is to make calculations with this theory.

Suppose there are two groups of traders. In January the first group receives cash and invests it in the market. During the remaining eleven months, they slowly withdraw this investment, in uniform increments every month. The second group of traders has less cash flow constraints or more alternative investments, and exploits the first group by taking up a position in December and holding it for a month, in order to profit from the January rise in price.

To analyze this more quantitatively, let $\text{mod}_x(y) = y/x - \text{floor}(y/x)$, be the modulus function, where $\text{floor}(x)$ is the smallest integer less than x . Dividing the year into twelve trading periods (where trading happens at the end of each month), the two strategies can be expressed as follows.

$$\omega_t^{(1)} = \begin{cases} 1 & \text{mod}_{12}(t) = 1 \\ -1/12 & \text{mod}_{12}(t) \neq 1 \end{cases}$$

$$\omega_t^{(2)} = \begin{cases} \omega & \text{mod}_{12}(t) = 0, \\ -\omega & \text{mod}_{12}(t) = 1 \\ 0 & \text{mod}_{12}(t) \neq 0, 1 \end{cases}$$

where ω is the size of the position held by the second group of traders during January.

This pattern of trading induces movements in the price that are non-zero on average. Using equation 11 the mean log-returns caused by this trading are $(1 - \omega)/\lambda$ in “January” (when $\text{mod}_{12}(t) = 1$), $(\omega - 1/12)/\lambda$ in “December” (when $\text{mod}_{12}(t) = 0$), and $-1/(12\lambda)$ otherwise. If the market has an average upward drift term μ (e.g. caused by another group of investors), with the initial condition $y_0^{(1)} = y_0^{(2)} = 0$, using equation 17 it is straightforward to show that the annual profit for strategy 2 is

$$g_1^{(2)} = \left(\frac{1}{\lambda}(1 - \omega) + \mu \right) \omega.$$

This is positive as long as $\omega < 1 + \mu\lambda$ and reaches its maximum value when $\omega = (1 + \mu\lambda)/2$. The gains for strategy 2 are

$$g_1^{(1)} = \frac{13}{2}\mu - \frac{4}{9\lambda},$$

independent of ω .

This illustrates how constraints on market participants may drive cash flows, which in turn may drive patterns in the markets. We will return in Section 4 to discuss whether such patterns can persist on evolutionary timescales. It also illustrates that it is straightforward to make calculations about price movements using the simple dynamical model of equation 11.

3.2 Value investors

Value investors trade based on an assessment of value in relation to the price. If they think the market is undervalued they tend to buy, and if they think it is overvalued they tend to sell. The assessment of value may involve a complicated analysis of non-financial or “fundamental” data, and different traders may arrive at quite different conclusions as to correct value. Even if we assume the value for each trader is given, there are many possi-

ble strategies for exploiting mispricings. We will begin by considering simple value strategies that are easy to analyze, building up to more complicated but more realistic strategies later.

Markets can be viewed as an organ of society, that performs the function of resource allocation. Markets help set society's goals. If the price of pork bellies go up, people will grow more pigs. One measure of how well society performs this function is the extent to which prices reflect other measures of value. Value is inherently subjective -- different people may have different opinions about what things are worth -- which is what drives trading. Nonetheless, to the extent that people agree about value, prices should track it, at least over the long term. Because this theory makes no assumptions about equilibrium this is by no means given *a priori* -- it will happen only if the strategies active in the market place orders that influence the price to keep it near the value. Evidence from market data strongly suggests that, while the price tends to roughly track value, large deviations are the rule rather than the exception. This is referred to as *excess volatility* [7]. As we will see, this theory provides a natural explanation of excess volatility, and provides some understanding about why it occurs.

Here we will assume that the perceived values are given. We will use the model that they are given by a random process of the form

$$v_{t+1} = v_t + \eta_t, \quad (\text{Eq 22})$$

where for simplicity η_t is a normal, IID noise process with standard deviation σ_η and mean μ_η . Thus the logarithm of the value follows a random walk. We will begin by studying the case where all traders perceive the same value, and return to study the case where they perceive different values in Section 3.2.5.

The natural way to quantify how well the price tracks the value is in terms of the theory of cointegration [20]. A random process whose n^{th} time difference is stationary is integrated of order n , or $I(n)$. For example, in equation 22 the value follows an $I(1)$, or *unit root* random process ("unit" because the coefficient in front of v_t is one). Two variables with unit roots are *cointegrated* if there is a linear combination of them that is stationary, i.e., that is $I(0)$. If the price tracks the value, we would expect the difference between price and value to be cointegrated.

A value strategy can be expressed in terms of the position $y_{t+1} = y(x_t, \dots)$, or in terms of the orders $\omega_{t+1} = \omega(x_t, \dots)$. Of course, this really doesn't matter, since it is easy to convert from one to the other by the relation $\omega_t = y_t - y_{t-1}$. But as we will see, simple strategies that one would naturally think of as value strategies have very different properties when they are formulated in terms of orders vs. positions. We will also see that these simple strategies produce market behavior that is unrealistic in some respects; to get more realistic behavior, we need to consider strategies closer to those actually used by real traders.

3.2.1 Pure order-based value strategies and equilibrium economics

A pure order based value strategy is of the form¹

$$\omega_{t+1}(v_t, z_t) = f(z_t - v_t)$$

where f is a generally decreasing function with $f(0) = 0$, v_t is the logarithm of the perceived present value, and z_t is the logarithm of the price. “Generally decreasing” means that f either decreases or remains constant, and is not constant everywhere. Note that this class of strategies have no state dependence -- they only depend on the *mispricing* $m_t = z_t - v_t$. If we expand in a Taylor’s series and assume the first derivative exists, then to leading order this becomes

$$\omega_{t+1} = -c(z_t - v_t).$$

$c > 0$ is a constant and will be called the *capital*. Under this strategy, as long as the market is undervalued the trader continues to buy and as long as it is over-valued the trader continues to sell. If this is the only strategy used in the market, then from equation 11 the dynamics can be written

$$z_{t+1} = z_t - \alpha(z_t - v_t) + \xi_t, \quad (\text{Eq 23})$$

where $\alpha = c/\lambda$. Writing this in terms of the log-return r_t and the mispricing $m_t = z_t - v_t$ gives

$$r_{t+1} = -\alpha m_t + \xi_t \quad (\text{Eq 24})$$

Since $r_{t+1} = z_{t+1} - z_t = m_{t+1} - m_t + v_{t+1} - v_t$, we can write this entirely in terms of the mispricing, which gives

$$m_{t+1} = (1 - \alpha)m_t - \eta_t + \xi_t. \quad (\text{Eq 25})$$

Since we have assumed in equation 22 that v_t is a unit root process, its time difference $\eta_t = v_{t+1} - v_t$ is stationary, and the second two terms may be regarded as a combined noise term $n_t = -\eta_t + \xi_t$. The mispricing is therefore a stationary random process as long as $0 < \alpha < 2$, and the price reverts to the value, as illustrated in Figure 3.

Multiplying both sides of equation 25 by m_{t+1} , taking averages, and assuming η and ξ are independent shows that the mispricing has variance

$$\sigma_m^2 = \frac{\sigma_\eta^2 + \sigma_\xi^2}{\alpha(2 - \alpha)}, \quad (\text{Eq 26})$$

1. Value strategies could equally well have been defined in terms of value and price rather than their logarithms; since there is a diffeomorphism between $e^{v_t} - x_t$ and $v_t - \log x_t$ this doesn’t matter. However, since the term on the left side of equation 11 is written in terms of the difference of the logarithm of the price, when expanding the right side in a Taylor’s series it is more natural to do so in terms of the logarithm.

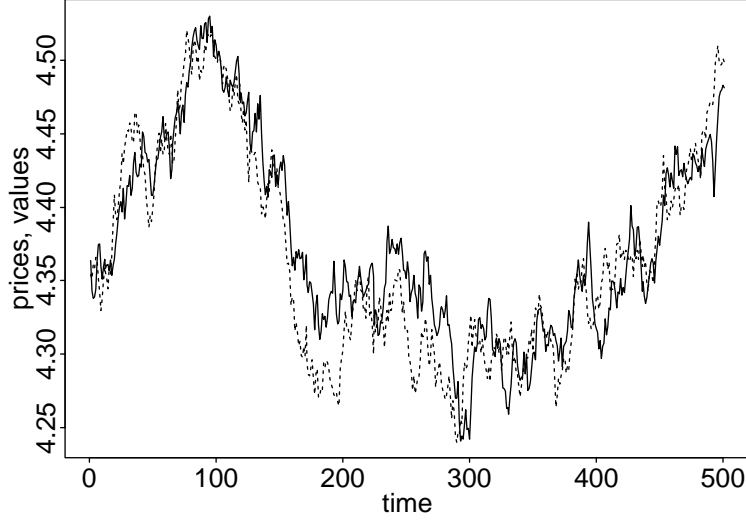


FIGURE 3. The log-price (dashed line) and the log-value (solid line) for the linear order-based value strategy of equation 23 with $\alpha = 0.8$, $\sigma_{\eta} = 0.01$, and $\sigma_{\xi} = 0.01$. The price is co-integrated to value, so even when the value changes according to a random walk the price remains close to the value.

where σ_{η}^2 and σ_{ξ}^2 are the variances of η_t and ξ_t . Without loss of generality we can assume they are both zero mean, and the first autocorrelation of the mispricing is

$$\rho_m(1) = \frac{\langle m_{t+1} m_t \rangle}{\langle m_t^2 \rangle} = 1 - \alpha. \quad (\text{Eq 27})$$

Equation 26 makes it clear that the effect of the external noise ξ_t and the value noise $-\eta_t$ on the mispricing are equivalent.

The basic statistics for the log-returns $r_t = z_t - z_{t-1}$ can be computed similarly. Squaring both sides of equation 24 and averaging gives

$$\sigma_r^2 = \alpha^2 \sigma_m^2 + \sigma_{\xi}^2 = \frac{\alpha \sigma_{\eta}^2 + 2 \sigma_{\xi}^2}{2 - \alpha}. \quad (\text{Eq 28})$$

The first autocorrelation of the log-returns can be found by multiplying equation 24 by r_t , averaging, and making use equation 24 again, which gives

$$\langle r_t r_{t+1} \rangle = \alpha^2 \langle m_t m_{t-1} \rangle - \alpha \langle m_t \xi_{t-1} \rangle.$$

The first term can be rewritten in terms of equations 26 and 27, and in the second m_t can be written in terms of m_{t-1} from equation 25. Some algebra then shows that

$$\rho_r = \frac{\alpha(1 - \alpha) \sigma_{\eta}^2 - \alpha \sigma_{\xi}^2}{\alpha \sigma_{\eta}^2 + 2 \sigma_{\xi}^2}. \quad (\text{Eq 29})$$

This shows that for the log-returns (and hence the price) the effect of the value noise η is not equivalent to that of the external noise ξ . If the process is driven purely by changes in value, i.e. when $\xi = 0$, the first autocorrelation of log-returns is $\rho_r = 1 - \alpha$. However, if it is driven purely by external noise, i.e. if the value is constant, then the first autocorrelation is $\rho_r = -\alpha/2$.

A correspondence to classical equilibrium economics occurs when $\alpha = 1$. From equation 23, when $\alpha = 1$

$$z_{t+1} = v_t + \xi_t.$$

Except for the external noise (which is $I(0)$), the price tracks the value exactly. The sequence of mispricings is uncorrelated, and under changes in value the log-returns are uncorrelated. The variance of the mispricing is equal to the sum of the variance of the changes in value and the variance of the external noise, and there is no “excess volatility”.

This is not surprising, as when $\alpha = 1$ and $\xi = 0$ this can be understood as a classic example of equilibrium supply and demand, as discussed in Section 2.2.2. In particular, let

$$S(z) - D(z) = c(z - v)$$

The value of λ that is consistent with maintaining equilibrium is $\lambda = c$. Changes in value cause shifts in the y -intercept. The system remains at its classic equilibrium, and absent external noise, the price tracks the value exactly.

Unless the parameters are adjusted exactly right to maintain the system at the classic supply-demand equilibrium, the price will not track the value exactly. If $\alpha < 1$ there is a market *under-reaction*. There is a positive autocorrelation in the mispricing, and for pure value-noise, a positive autocorrelation in the log-returns. Similarly, if $\alpha > 1$ there is a market *over-reaction*, which induces negative autocorrelations.

Is it reasonable to expect that the parameters will adjust to move the market to equilibrium? For this to happen either the market maker needs to adjust the liquidity, or the traders need to adjust their capital. In practice both of these will happen. But they are independent agents, and no one has complete knowledge of the universe of strategies and the capital allocated to each. For the system to go to equilibrium requires an adjustment in the capital of the individual strategies and/or in the liquidity of the market maker, which must be driven by the profit-seeking goals of the individual agents, and occurs on longer timescales. This will be studied further in Section 4.

Purely order-based strategies are patently unrealistic from a behavioral point of view because their positions can grow without bound. Orders are placed as long as there is a mispricing, regardless of the position. As a result, the position is strongly path-dependent; the longer the mispricing goes without changing sign, the larger the position becomes. Even when the mispricing goes to zero the trader is left with a position, which decreases only as the mispricing persists with the opposite sign. In fact, since ω_t is proportional to the mispricing, and the mispricing is an $I(0)$ process, it is clear that the position y_t , which is the accumulated sum of the orders through time, is an $I(1)$ process. This means that the

position, and hence the risk, can become unbounded, and the gains are not well defined¹. The tendency of the position to increase without bound is illustrated in Figure 4. Thus, the

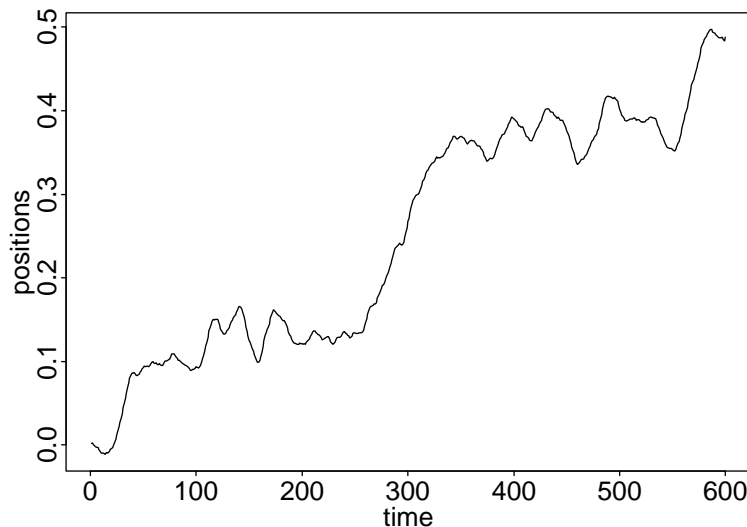


FIGURE 4. The orders (solid line) and the positions (dashed line) for the linear order based value strategy of equation 23. The position can grow arbitrarily large.

simple order-based strategy is unrealistic both because it doesn't correspond to what traders actually do, and because it leads to unbounded risk. Furthermore, making the strategy nonlinear will not fix these problems.

3.2.2 Pure position based strategies

Another natural class of value strategies that might give hope for fixing the problems with order based strategies are position based value strategies, which are of the form

$$y_{t+1}(v_t, z_t) = g(z_t - v_t)$$

where as before g is a generally decreasing function with $g(0) = 0$. Expanding in a Taylor series, to first order the position can be approximated as

$$y_{t+1} = -c(z_t - v_t).$$

Since $\omega_t = y_t - y_{t-1}$, the induced dynamics are

1. One might try to modify the order-based strategy so that the positions are bounded, for example by distributing the capital among different traders, each of whom has bounds on his positions. This just creates a new problem, however; instead of the position growing without bound, the number of traders in the market grows without bound. Furthermore, one must postulate an additional mechanism to signal new traders to enter the market.

$$z_{t+1} - z_t = \frac{c}{\lambda}(v_t - v_{t-1} - (z_t - z_{t-1})) + \xi_t.$$

As before c is a positive constant proportional to the trading capital. Letting $r_t = z_t - z_{t-1}$, $\alpha = c/\lambda$, and $\eta_{t-1} = v_t - v_{t-1}$, this can be written

$$\begin{aligned} r_{t+1} &= -\alpha r_t + \alpha \eta_{t-1} + \xi_t \\ z_{t+1} &= z_t + r_{t+1} \end{aligned} \quad (\text{Eq 30})$$

This makes it clear that there are several fundamental differences between this and the order-based strategies of the previous section. The log-return does not depend on the price explicitly -- it depends on the previous return. Furthermore, it only depends on changes in the log-value, rather than the value itself. In addition, the dynamics are second order, i.e. the state depends on both the current price and the previous price. The eigenvalues are $(1, -\alpha)$. Thus when $\alpha \leq 1$ the dynamics are neutrally stable, and when $\alpha > 1$ they are unstable. The first autocorrelation of the log-returns is $\rho_r = -\alpha$, and the variance of the log-returns is

$$\sigma_r^2 = \frac{\alpha^2 \sigma_\eta^2 + \sigma_\xi^2}{1 - \alpha^2}. \quad (\text{Eq 31})$$

Position-based strategies do not generally co-integrate the price to the value. This is already suggested by the fact that to first order there is no explicit dependence on price or value¹. For the linearized example above, the lack of cointegration can be shown explicitly by substituting $m_t = z_t - v_t$ into equation 30, to get

$$\Delta m_{t+1} = -\alpha \Delta m_t - \eta_t + \xi_t,$$

where $\Delta m_t = m_t - m_{t-1}$. When the dynamics are stable ($|\alpha| < 1$), Δm_t is an $I(0)$ process, and m_t is an $I(1)$ process, i.e. the mispricing is a random walk. This seems to be the case even for more general nonlinear position based value strategies². The intuitive reason is that, while the position-based strategy resists increases in the absolute value of the mispricing, once a mispricing occurs, it also resists decreases with equal intensity. Thus, while the negative autocorrelation induced by position-based strategies makes the price not wander from value as quickly as it would otherwise, this is not sufficient to keep the price close to the value. The lack of co-integration is illustrated in Figure 5. Note that since the mispricing is unbounded, and the position is proportional to the mispricing, the position is also unbounded. Numerical simulations indicate that these conclusions are not altered if the market maker makes explicit inventory-based adjustments using equation 10.

1. Nonlinear position-based strategies generally do have explicit dependence on both price and value. However, both numerical experiments and the argument given in the next footnote show that they suffer from the same problems as the linear strategy.

2. In the general nonlinear case the mispricing can be written $\Delta m_{t+1} = 1/\lambda(g(m_{t+1}) - g(m_t)) - \eta_t + \xi_t$. Because g is generally decreasing, this can be written $\Delta m_{t+1} = -c(m_t)\Delta m_t - \eta_t + \xi_t$, where $c(m_t) \geq 0$. It would seem that either this is a stable random process, in which case m_t is $I(1)$, or it is unstable, in which case m_t is also unstable.

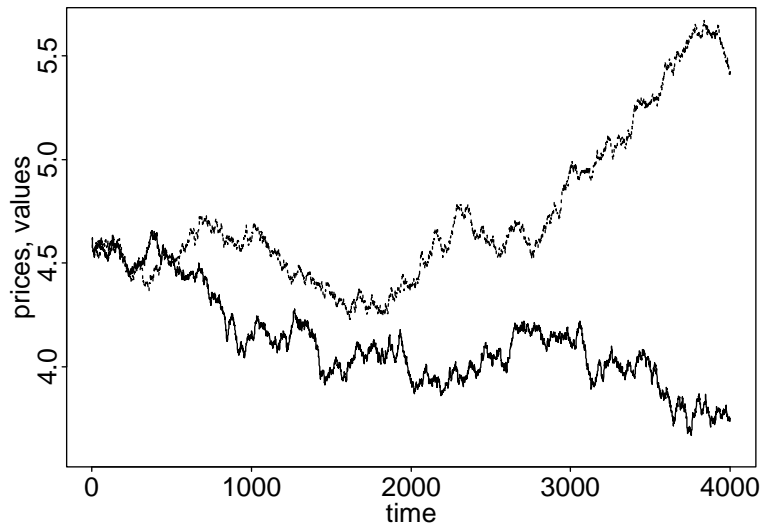


FIGURE 5. The log-price z_t (dashed line) and the log-value v_t (solid line) for the linear position based value strategy of equation 30 with $\alpha = 0.1$. Unlike the order based strategy, the price is not co-integrated to the value, and can wander arbitrarily far away from it.

In contrast to the order-based strategies, which correspond to a behavioral pattern that is simply not followed in the real world, real traders do use position-based value strategies. However, as we see, in a universe consisting *only* of position-based value strategies, there is no cointegration of price and value, leading to unbounded positions. In order to have sensible behavior and bounded risk, position-based value strategies depend on other strategies to provide cointegration between price and value.

3.2.3 State-dependent threshold value strategies

The analysis of the simple value strategies above presents the question of whether there exist strategies that both cointegrate price and value and have bounded positions. In this section we discuss a class of value strategies that are more complicated but nonetheless commonly used, and demonstrate that they satisfy this property.

From the point of view of the individual trader, one of the problems with the position-based value strategies studied in the previous section is that they may incur excess transaction costs. Trades are made every time the mispricing moves up or down, and within a short space of time fluctuations may cause alternating buying and selling with no net change in the mispricing. One common approach to solving this problem is to use state dependent strategies, with different conditions for entering vs. exiting a position. Like the simpler value strategies studied earlier, such strategies are based on the belief that the price will revert to the value. By only entering a position when the mispricing is large, and only exiting when it is small, the trader hopes to profit by only trading when the price movement toward value is large enough to beat transaction costs.

An example of such a strategy, which is both nonlinear and state dependent, can be constructed as follows: Assume that a short position $-c$ is entered when the mispricing exceeds a threshold T and exited when it goes below a threshold τ . Similarly, a long position c is entered when the mispricing drops below a threshold $-T$ and exited when it exceeds $-\tau$. This is illustrated in Figure 6. Since this strategy depends on its own position

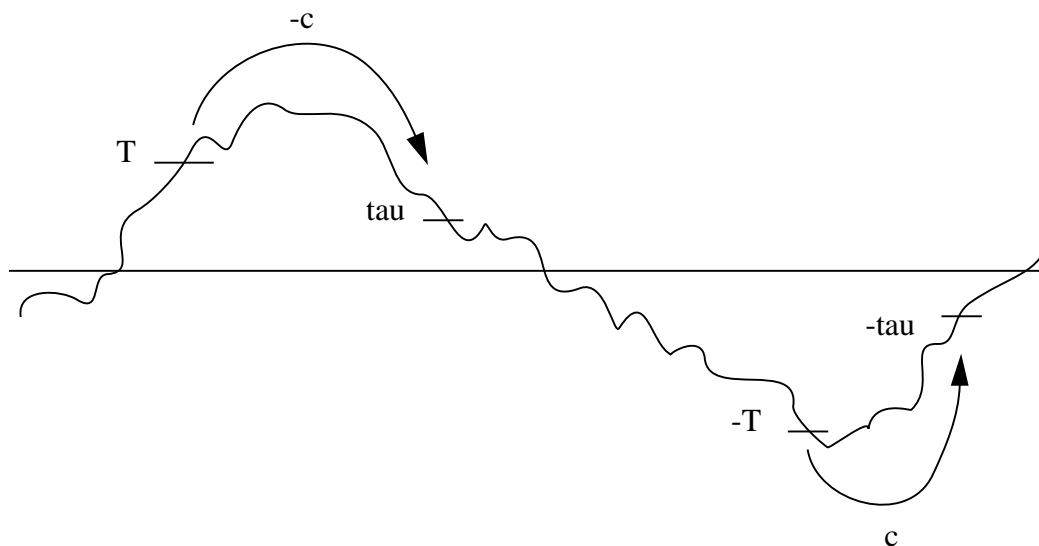


FIGURE 6. Schematic view of a nonlinear, state-dependent value strategy. The trader enters a short position $-c$ when the mispricing exceeds a threshold T , and holds it until the mispricing goes below τ . The reverse is true for long positions.

as well as the mispricing, it is a finite state machine, as shown in Figure 7.

In general different traders will choose different entry and exit thresholds. Let trader i have entry threshold $T^{(i)}$ and exit threshold $\tau^{(i)}$. For the simulations presented here we will assume a uniform distribution of entry thresholds ranging from T_{min} to T_{max} , and a uniform density of exit thresholds ranging from τ_{min} to τ_{max} , with a random pairing of entry and exit thresholds. Values of c are assigned as $c = a(T - \tau)$, where a is a positive constant¹.

It is clear that to correspond to a sensible value strategy the entry threshold should be positive and greater than the exit threshold, i.e. $T > 0$ and $T > \tau$. The choice of the exit threshold τ is not as obvious. Given the transaction cost of entering and exiting positions, to be sure that the full return has been extracted from the position, many traders will take $\tau < 0$. However, some traders may decide to exit their positions before the mispricing is zero, under the theory that once the price is near the value there is little expected return remaining. We can simulate a mixture of the two approaches by making $\tau_{min} < 0$ and

1. This assignment is natural because traders managing more money (with larger c) incur larger transaction costs. The expected gain absent market impact, hence the gain needed to beat transaction costs, is proportional to $T - \tau$.

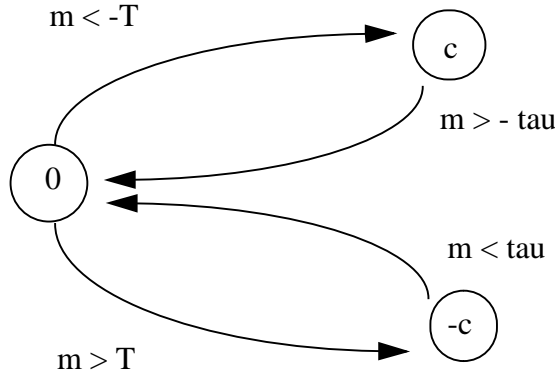


FIGURE 7. The nonlinear state-dependent value strategy represented as a finite-state machine. From a zero position a long-position c is entered when the mispricing m drops below the threshold $-T$. This position is exited when the mispricing exceeds a threshold $-\tau$. Similarly, a short position $-c$ is entered when the mispricing exceeds a threshold T and exited when it drops below a threshold τ .

$\tau_{max} > 0$. However, to be a sensible value strategy, a trader would not exit a position at a mispricing that is further from zero than where the position was entered. τ_{min} should not be *too* negative, so we should have $-T < \tau < T$ and $|\tau_{min}| \leq T_{min}$.

To achieve cointegration of price and value it is clear that $\tau < 0$ is a desirable property. This gives the strongest cointegration, since the price changes induced by trading always have the opposite sign of the mispricing for both entry and exit, so the trading always acts to reduce the mispricing. A simulation of the price dynamics induced by this strategy using $\tau_{max} = 0$ and $\tau_{min} < 0$ is shown in Figure 8. The price and value are cointegrated.

The behavior of the mispricing, and the cointegration of price and value, are quite different for this example than for the simple order-based value strategy studied earlier. This example corresponds much more closely to the behavior in real markets, namely, the mispricing changes sign only infrequently, and cointegration is much weaker.

Figure 9 shows a simulation with the range of exit thresholds chosen so that $\tau_{min} < 0$ but $\tau_{max} > 0$. The price and value are still cointegrated, but weaker than before, as illustrated by the increased amplitude of the mispricing. In addition, there is a tendency for the price to “bounce” as it approaches the value. This is caused by the fact that when the mispricing approaches zero some traders exit their positions, which pushes the price away from the value. The value becomes a “resistance level” for the price, and there is a tendency for the mispricing to cross zero less frequently than it does when $\tau^{(i)} < 0$ for all i . Based on results from numerical experiments it appears that the price and value can be cointegrated as long as $\tau_{min} < 0$. Necessary and sufficient conditions for cointegration deserve further study¹.

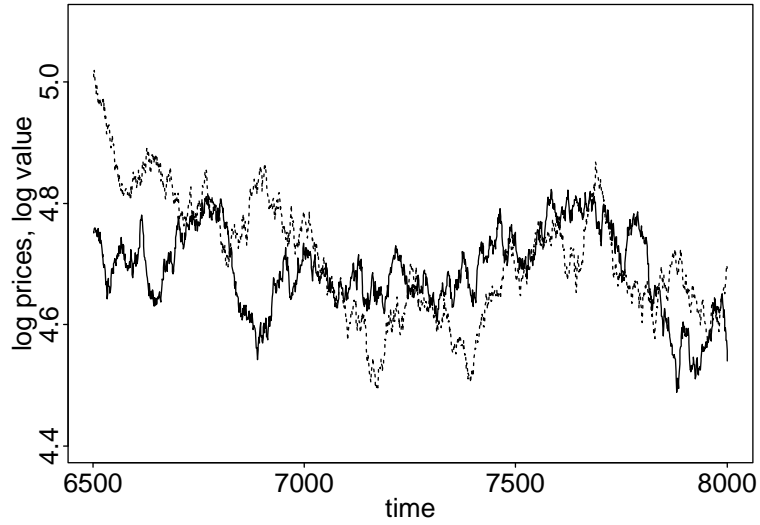


FIGURE 8. The induced price dynamics of a nonlinear state-dependent value strategy with 1000 traders using different thresholds. The log-price is shown as a solid line and the log-value as a dashed line. $\tau_{min} = -0.5$, $\tau_{max} = 0$, $T_{min} = 0.5$, $T_{max} = 6$, $N = 1000$, $a = 0.001$, $\sigma_{\eta} = 0.01$, and $\sigma_{\xi} = 0.01$, and $\lambda = 1$.

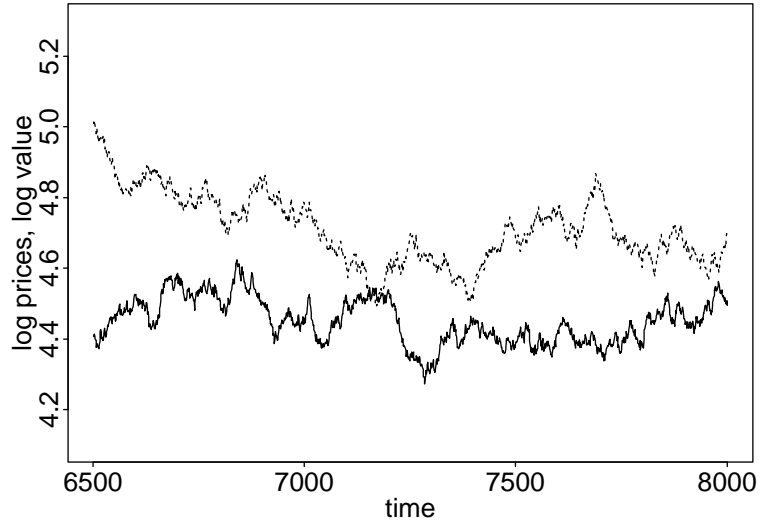


FIGURE 9. Price (solid) and value (dashed) vs. time for the nonlinear state-dependent strategy of Figure 7. The parameters and random number seed are the same as Figure 8, except that $\tau_{min} = -0.5$ and $\tau_{max} = 0.5$

1. Problems can occur in the simulations if the capital $c = a(T - \tau)$ for each strategy is not assigned reasonably. If a is too small the traders may not provide enough restoring force for the mispricing; once all N traders are committed to a long or short position, price and value cease to be cointegrated. If a is too big instabilities can result because the kick provided by a single trader creates oscillations between entry and exit. Nonetheless, between these two extremes there is a large parameter range with reasonable behavior.

This demonstrates that there is at least one class of value strategies that tends to cointegrated price and value. It is interesting that cointegration of price and value should depend on something as apparently indirect as state-dependence induced by the motivation to reduce transaction costs. Also, note that the nature of the cointegration relationship is realistically weak; mispricings can persist for thousands of iterations.

3.2.4 Mixed technical and value strategies

A technical trading strategy is one that bases decisions on past values of the price. Cointegration may also be helped by mixed technical and value strategies. A clear example is a “value strategy with a technical confirmation signal”. A trader may believe in value, but decide to wait until the price goes through a turning point to take a position. The reasoning behind such a strategy is to avoid risk by waiting until the market indicates that the other value traders are starting to enter their positions and push the mispricing down. For instance, consider the preceding strategy, but make the entry condition for a short position of the form $m_t > T$ and $z_{max} - z_t > D$, i.e. the mispricing must exceed a given level and the price must have dropped from an earlier maximum (e.g. the maximum over the last thirty iterations) by at least a certain amount. Such a strategy will aid the cause of cointegration, since it produces a trade with the opposite sign of the mispricing at a point where the traders using the threshold strategy are simply holding their positions. This illustrates how strategies may act in concert to perform a given function with the market, e.g. a pure value strategy may begin a price reversal and a technical strategy may reinforce it.

3.2.5 Diverse values and excess volatility

In general different traders will perceive different values. For strategies that are linear in the value (or the logarithm of value) the induced dynamics will be identical to those of a single strategy with the mean value and the combined value. However, for nonlinear strategies this will not be true -- different perceptions of value can cause excess volatility and create opportunities for trend followers.

For example, consider the simple order based strategy of Section 3.2.1. Suppose there is a group of N different traders each perceiving value $v_t^{(i)}$. The dynamics are

$$z_{t+1} = z_t + \sum_{i=1}^N \frac{c_i}{\lambda} (v_t^{(i)} - z_t).$$

Letting

$$\bar{v}_t = \frac{1}{c} \sum_{i=1}^N c_i v_t^{(i)},$$

where $c = \sum c_i$, this becomes identical to equation 23, except that the logarithm of the value is replaced by the weighted mean of the logarithms of the value computed by each

trader. Or in other words, the market is equivalent to that of a single trader who believes the correct value is the weighted geometric mean of the individual values. A similar relation will be true for any strategy that depends linearly on v_t . Thus, for strategies that depend linearly on log-value, the dynamics are driven solely by the *collective* valuation.

The situation can be quite different when the strategies depend nonlinearly on the value. For the purposes of simulation it is convenient to assume that, although different traders perceive diverse values, they change in tandem. This can be modeled as a simple “base” value process \bar{v}_t that follows equation 22, with a fixed random offset $v^{(i)}$ that is different for each trader. The value perceived by the i^{th} trader at time t is

$$v_t^{(i)} = \bar{v}_t + v^{(i)}.$$

In the simulations the value offsets are assigned uniformly between v_{min} and v_{max} , where $v_{min} = -v_{max}$. The range of perceived values is $2v_{max}$.

We will define the excess volatility as

$$V = \sqrt{\sigma_r^2 / (\sigma_\eta^2 + \sigma_\xi^2)}, \quad (\text{Eq 32})$$

i.e. as the ratio of the volatility of the log-returns to the volatility of the noise terms. Figure 10 and Figure 11 illustrate the effect of a diversity of perceived values using the

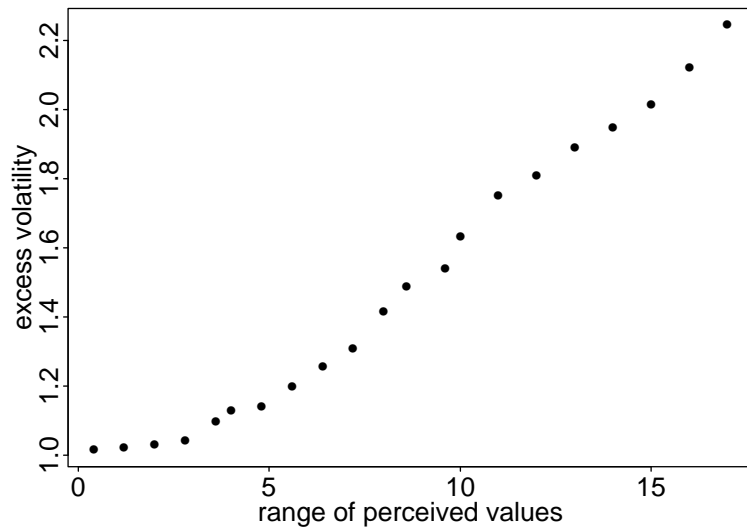


FIGURE 10. Excess volatility as the range of perceived values increases while the capital is fixed at 0.035. See equation 32. The other parameters are the same as those in Figure 8.

threshold value strategy of Section 3.2.3. The mispricing is measured relative to the geometric mean of the log-value. We see that the excess volatility increases with the range of perceived values and the capital. This excess volatility is generated by trading caused by disagreements about value. Extremes of the mispricing drive most traders to take either

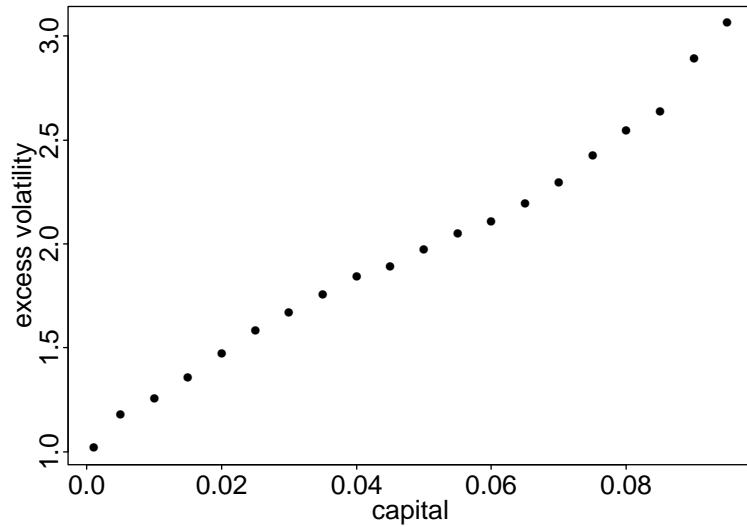


FIGURE 11. Excess volatility, defined as in the text, as the capital varies while the range of perceived values is fixed at 2, i.e. $v_{min} = -1$ and $v_{max} = 1$. Parameters are as in Figure 8.

long or short positions, and causes cointegration, but when the mispricing is close to the value there is “noise trading” that generates excess volatility.

We may think of the market as a machine whose function is to keep the price near the correct value. If the machine were perfectly efficient price and value would track exactly. The inverse of the excess volatility is one possible measure of efficiency. By any measure, this market is a machine whose efficiency is less than one.

3.3 Trends and trend followers

A *trend* occurs when successive price movements are positively correlated. This may be episodic, i.e. the market may trend during one period of time and display negative correlations during other periods of time. While the existence of trends has caused considerable debate in the literature, it is clear that trend following is a commonly used strategy [13]. Furthermore, there are several possible causes of trends that make this behavior rational:

- To minimize transaction costs, large positions are usually acquired gradually. A single institution may take weeks or even months to take a position. Such positions are sometimes a significant fraction of the market share, i.e. large enough that the transition into such a position may have a significant impact on the price [10].
- A stop is an order to flatten an existing position depending on the price, e.g. “sell all my stock if the price drops below \$50”. If stops are placed at a range of different price levels, as each stop is hit it induces a price change in the same direction, which may in turn cause the next stop to be hit, creating a chain reaction.

- Market under-reaction by value investors can cause positive correlations in the price.
- Information may diffuse into the market gradually.
- Inventory effects when a market maker acquires a net long or short position may cause correlated price movements.
- Induced price dynamics by trend followers generates trends, creating a self-fulfilling prophesy.

We have already seen an example of market under-reaction. In this section we will discuss information diffusion and self-fulfilling prophesies. The causes listed above may reinforce each other, e.g. if there is a market under-reaction and trend followers exploit it the trend becomes even stronger.

3.3.1 Information diffusion and rumors

Gradual information diffusion may cause trends. For example, if information is transmitted via rumors it may enter the market slowly. This can be a particularly strong effect if there is feedback between the rumor and trading. Once a trader has already taken his position, it is to his advantage to encourage others to do the same. Someone with a “hot tip” may buy, then encourage others to buy, etc., causing a buying wave that generates a trend in the price. For example, assume that the rate at which such a rumor spreads is proportional to the amount of buying that it generates. Each unit of excess buying generates $c(t)$ units on the next, i.e. $\omega_t = c(t)\omega_{t-1}$. Such information will lose its value over time; for example, suppose it degrades linearly with time until it becomes worthless. Let $c(t) = c_0(1 - kt)$ for $0 < t < 1/k$ and $c(t) = 0$ for $t > 1/k$. Solving for ω_t , from equation 11 we have

$$\langle z_t - z_0 \rangle = \frac{\omega_0 c_0^t}{\lambda} \prod_{i=0}^{t-1} (1 - ik)^i, \quad (\text{Eq 33})$$

for $0 < t < 1/k$. This is a classic sigmoidal growth pattern.

3.3.2 Trend followers

Trend followers are investors who invest based on the belief that markets tend to trend. When they perceive an upward trend they buy, and when they perceive a downward trend, they sell. To be more specific, a trading strategy is trend following on timescale θ if the net position y_t has a positive correlation with past price movements on timescale θ . Assuming for convenience that both y and the log-returns are zero mean, y_t is a trend following strategy if

$$\langle y_{t+1}(z_t - z_{t-\theta}) \rangle > 0.$$

Note that a given strategy may be trend following on several different timescales, and may be trend following on some timescales but not on others.

We have defined trend following strategies in terms of their position. As for value strategies, one could define order-based trend following strategies. In fact the position-based value strategy can be regarded as an order-based trend strategy that also depends on changes in value. This is evident in equation 30; this strategy only depends on r_t and noise terms. Order-based trend strategies have the same problem as their value counterparts that their positions are unbounded, and will not be considered further.

An example of a simple linear trend following strategy is

$$y_{t+1} = c(z_t - z_{t-\theta}).$$

From equation 11, the dynamics induced by this strategy are

$$\begin{aligned} r_{t+1} &= \frac{c}{\lambda}(y_{t+1} - y_t) + \xi_t \\ &= \alpha(r_t - r_{t-\theta}) + \xi_t \end{aligned} \tag{Eq 34}$$

where $\alpha = c/\lambda$. These dynamics tend to induce trends, as illustrated in Figure 12.

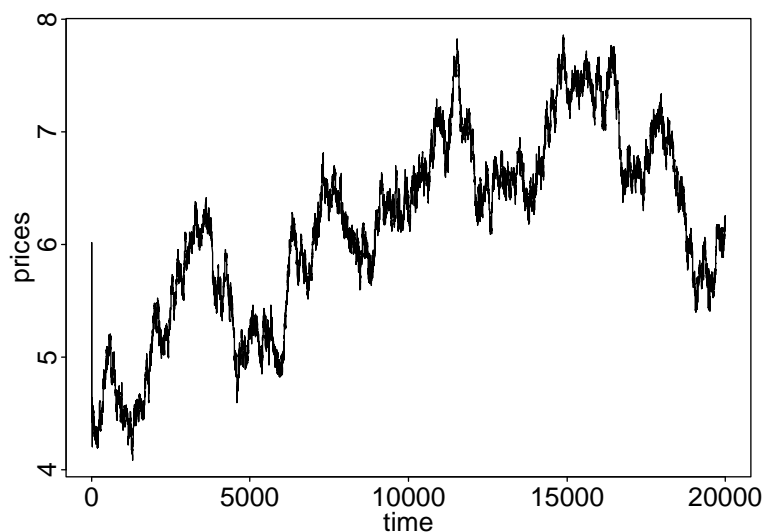


FIGURE 12. Log price vs. time for trend followers with $\alpha = 0.2$ and $\theta = 10$ in equation 34. This illustrates how trend followers tend to induce trends.

The eigenvalues are

$$\varepsilon_{\pm} = \frac{\alpha(1 - \alpha) \pm \sqrt{5 - 2\alpha + \alpha^2}}{2}.$$

Both eigenvalues have absolute value less than one, and the dynamics are stable when $\alpha < 1$. The autocorrelation function can be solved by multiplying by r_{t-n} and averaging, which gives $\rho_{n+1} = \alpha(\rho_n - \rho_{n-\theta})$. Doing this for $n = 0, \dots, \theta - 1$ gives a system of θ linear equations that can be solved for the first θ values of the autocorrelation function;

the remainder can be found by iterating these relations for higher values of n . Doing this for several different values of θ illustrates the basic pattern for ρ_n : The first autocorrelation ρ_1 is positive and of order α ; the remaining autocorrelation coefficients oscillate and reach a minimum at $n = \theta$; as n increases further, they continue to decay, while oscillating between positive and negative values with a period of 2θ . An example is shown for $\theta = 10$. The oscillation in the autocorrelation function corresponds to a tendency to induce oscillations in the price. It is a side-effect of formulating the strategy in terms of the position. There are many different types of trend-following strategies defined in the technical trading literature [21]. They all share the property of inducing trends; the extent to which they also induce oscillations depends on the strategy.

It is instructive to compare the simplicity and power of this analysis to that of DeLong et al. [6], who used equilibrium methods to point out that trend followers could create self-fulfilling prophecies, and that given that others do this, trend following becomes rational.

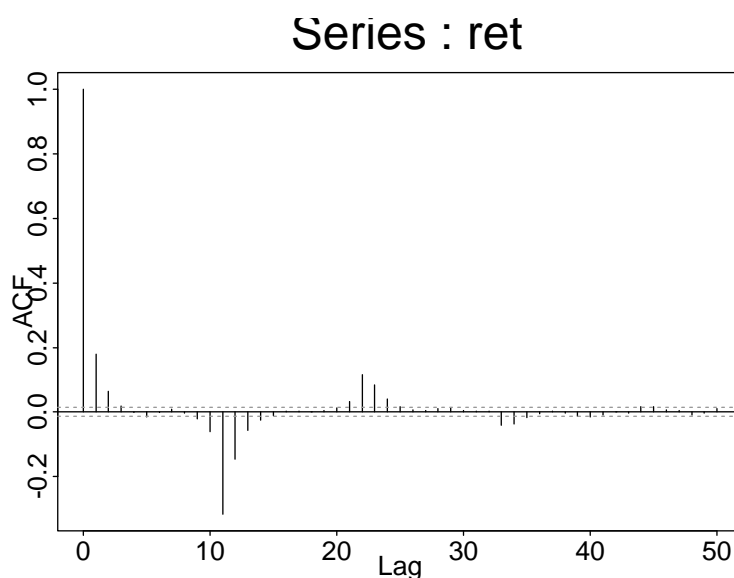


FIGURE 13. The autocorrelation function for equation 34 with $\alpha = 0.2$ and $\theta = 10$. In addition to inducing trends, position based strategies tend to induce oscillations.

Like value investors, trend followers often use thresholds to reduce transaction costs. Given a trend indicator $I(z_t, \dots, z_{t-\theta})$, a nonlinear trend strategy can be defined as a finite state machine, as shown in Figure 14. This will be used later.

3.4 Market drift and inventory effects

As discussed already some markets, such as the stock market, have a tendency to drift in one direction. If we assume this is driven by a systematic drift in the underlying perceived values, we can model this by adding a constant drift term to the dynamics of the

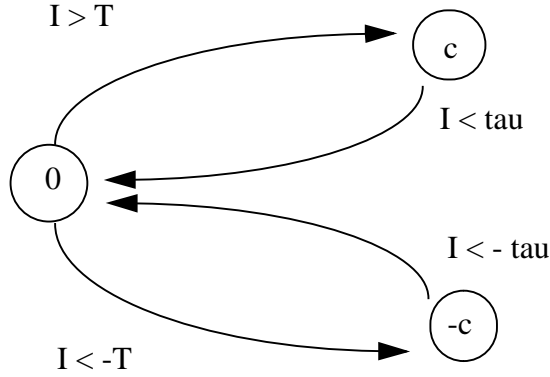


FIGURE 14. A nonlinear state-dependent trend strategy represented as a finite-state machine.

value in equation 22. In this case, using a population of threshold value investors, the price remains cointegrated to the value as one would hope; even though there is no explicit drift term in the price, it locks on to the drift in value.

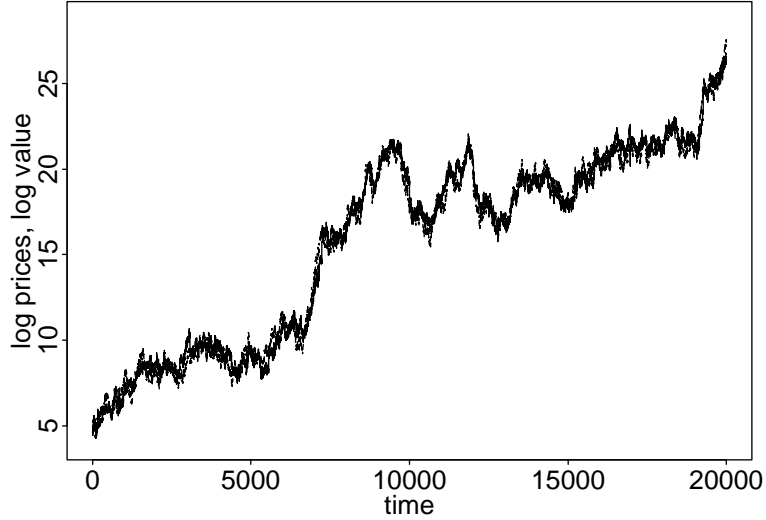


FIGURE 15. The induced price dynamics when a trend term of magnitude 0.001 is added to the value process. There are 1000 traders using threshold-value strategies as in Figure 8. The parameters are $c = 0.01$, $\sigma_{\xi} = 0.01$, $\sigma_{\eta} = 0.01$, $b = 0.3$, $v_{min} = -1.2$, and $v_{max} = 1.2$.

However, the resulting dynamics using the simple market making rule of equation 4 are unrealistic in at least one respect. When there is a positive drift term in value, with the passage of time the market maker tends to accumulate a net short position. Furthermore, this position appears to grow without bound.

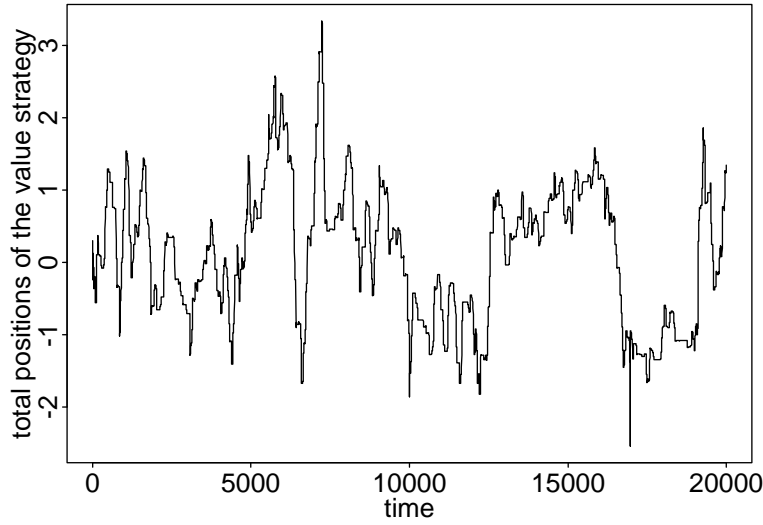


FIGURE 16. The positions for the simulation described above.

This situation is one in which the inventory effect appears to be necessary to get realistic behavior. In numerical simulations with a positive drift term in the value, by using a sufficiently large value of b in equation 10 it was possible to prevent the market maker's position from growing. Having a diversity of values also helps, although this does not seem to be sufficient to prevent growth in the market maker's position on its own. This is illustrated in Figure 15 and Figure 16.

3.5 Value investors and trend followers together

So far we have investigated homogeneous ecologies consisting of either single strategies or closely related groups of strategies. The price dynamics of homogeneous strategies tend to be unrealistic because they have linear structure in the price. For value investing strategies the autocorrelation of the log-returns is typically negative, and for trend following strategies it is positive. For real price series, in contrast, the autocorrelation tends to be very close to zero [22]. One simple way to achieve this is to combine value investors and trend followers in the proper ratio so that the linear structure disappears. To do this we use the threshold based value and trend strategies of Figure 7 and Figure 14. We begin by assigning the same thresholds to both the trend followers and value investors, and adjust the capital of the trend followers by trial and error so that the autocorrelation of the log-returns is close to zero¹. This means the trading volume of each group is roughly matched, and there is no significant linear temporal structure in the price. There is significant non-

1. The parameters for the simulation are $N_{value} = N_{trend} = 1200$, $\sigma_{\xi} = 0.35$, $T_{trend}^{min} = T_{value}^{min} = 0.2$, $T_{trend}^{max} = T_{value}^{max} = 4$, $\tau_{trend}^{min} = \tau_{value}^{min} = -0.2$, $\tau_{trend}^{max} = \tau_{value}^{max} = 0$, $a_{value} = 2.5 \times 10^{-3}$, $a_{trend} = 2.6 \times 10^{-3}$, $v_{min} = -2$, and $v_{max} = 2$, $\theta_{min} = 1$, $\theta_{max} = 100$, $b = 0.2$, and $\lambda = 1$. The mean of the perceived values as a function of time were imposed externally to match the American stock market, as described on page 47.

linear structure, however, as illustrated in Figure 17. This shows the smoothed volume¹ of

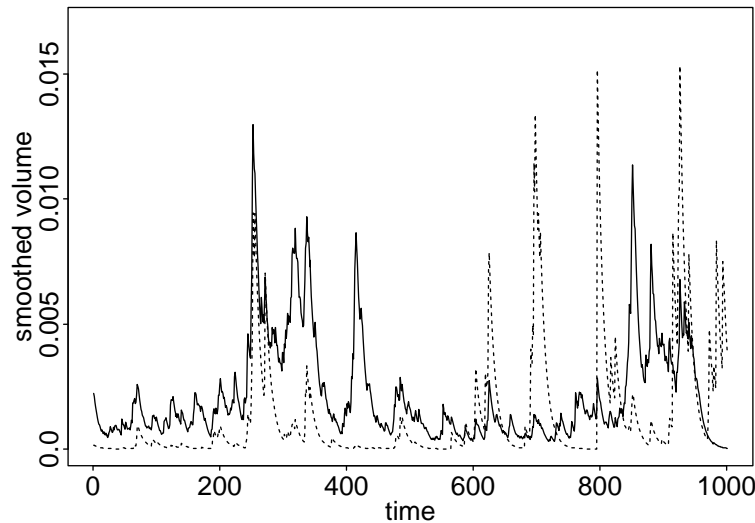


FIGURE 17. Smoothed trading volume of value investors (solid line) and trend followers (dashed line). The two groups become active at different times; when the value investors dominate the log-returns have a negative autocorrelation, and when the trend followers dominate there is a positive autocorrelation. Even though there is no linear temporal structure, there is strong nonlinear structure. Parameters are as described in the text; this is only a short portion of the total simulation.

value investors and trend followers as a function of time. It makes it clear that the two groups of traders become active at different times. Since the trend followers induce positive autocorrelations and the value investors negative autocorrelations, for a trader who understands the underlying dynamics there is predictable nonlinear structure². Statistical analyses of the volume and prices display many of the characteristic properties of real financial timeseries, as illustrated in Figure 18. The log-returns are more long-tailed than those of a normal distribution, i.e. there is a higher density of values at the extremes and in the center with a deficit in between. This also evident in the size of the fourth moment. The excess kurtosis

$$k = \frac{\langle (r_t - \bar{r}_t)^4 \rangle}{\sigma_r^4} - 3,$$

1. The smoothed volume is computed as $\bar{V}_t = \beta \bar{V}_{t-1} + (1 - \beta)V_t$, where V_t is the volume and $\beta = 0.9$.

2. The nonlinear structure can be exploited by any trader that knows the underlying generating process. It also is possible to extract the nonlinear structure directly from the time series, but due to statistical estimation problems this may not be easy. The forecasting accuracy depends strongly on how well the model matches the true dynamics. This deserves further investigation.

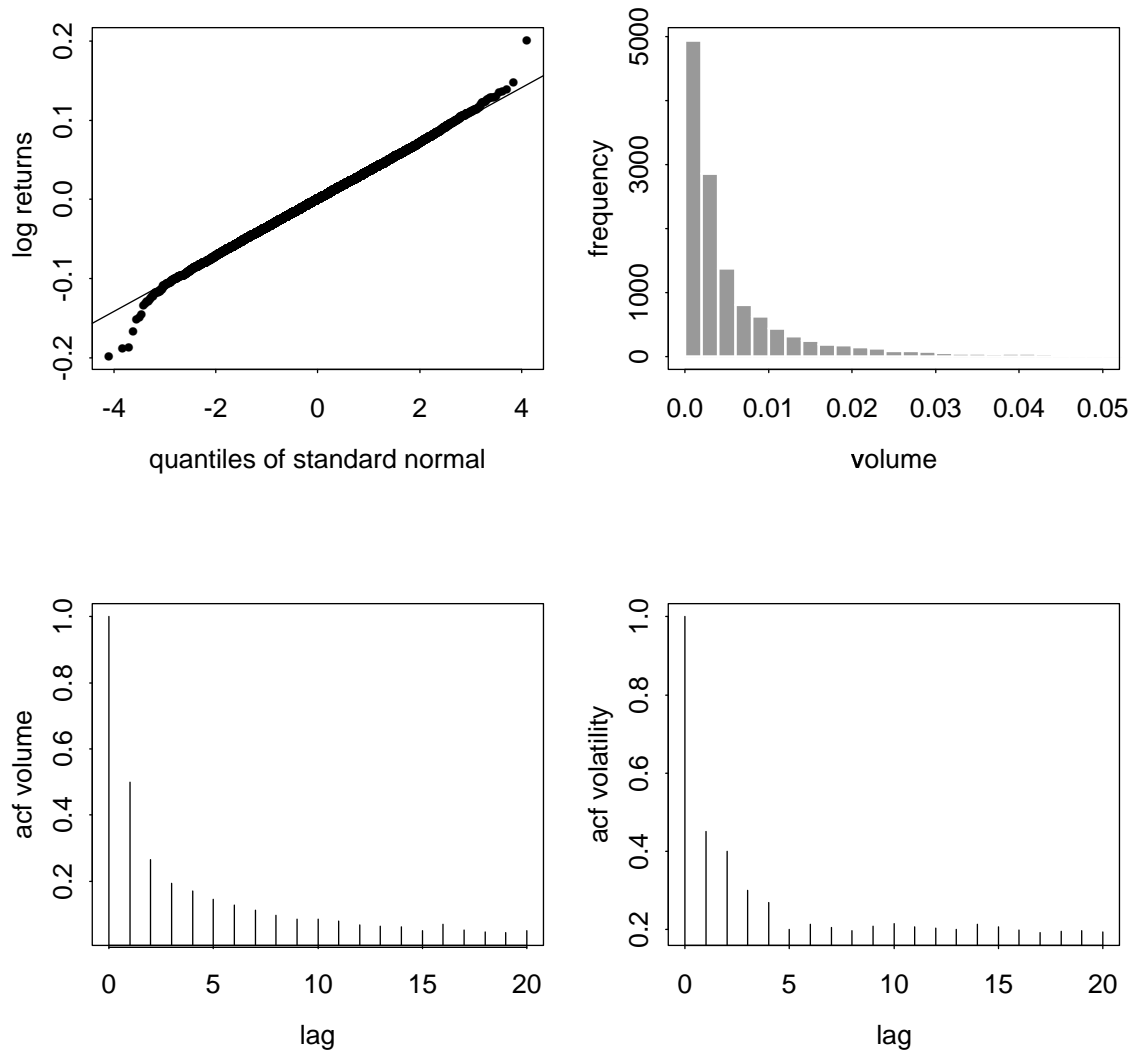


FIGURE 18. An illustration that an ecology of threshold based value investors and trend followers shows statistical properties that are typical of real financial time series. The upper left panel is a “q-q” plot, giving the ratio of the quantiles of the cumulative probability distribution for the log-returns to those of a normal distribution. If the distribution were normal this would be a straight line; since it is “long tailed” the slope is flatter in the middle and steeper at the extremes. The upper right panel shows a histogram of the volume. It is heavily positively skewed. The lower left panel shows the autocorrelation of the volume, and the lower right panel shows the autocorrelation of the volatility. These vary based on parameters, but long tails and temporal autocorrelation of volume and volatility are typical.

is roughly $k \approx 9$, in contrast to the expected value $k = 0$ for a normal distribution. The histogram of volumes is peaked near zero with a heavy positive skew¹. The volume and

1. For the threshold strategies used here there is a fraction of iterations with no trading at all. This is no longer the case when linear strategies are included, which also results in a more realistic distribution of trading volumes.

volatility both have strong positive autocorrelations. Thus if there is higher than usual volume or volatility on a given day, there tends to be higher than usual volume or volatility on subsequent days. The intensity of the long-tails and correlations vary somewhat as the parameters are changed or strategies are altered, for example if linear trend followers are substituted for threshold trend followers. However, the basic properties of long tails and autocorrelated volume and volatility are robust¹.

The existence of long tails and autocorrelations in volume and volatility have been a topic of debate from a theoretical point of view. The dynamical formulation presented here offers a simple explanation. For a broad class of strategies, under the dynamics given in equation 11, a larger than average change in price at one time will drive larger than average trading volume at the next time. To the extent that the trading is unbalanced, this will again drive a larger than average change in price. When this occurs temporal correlations in volume and volatility are to be expected. This is clearly true of trend following strategies. The results of this section make it clear that nonlinear trend and value strategies do not just cancel each other out. Thus it is possible to have a large autocorrelation in volatility at the same time that there is zero autocorrelation in the log returns. Note that temporal variations in volatility imply that the distribution of log-returns can be regarded as a superposition of normal distributions with different standard deviations. Such a distribution is generally long-tailed. This hypothesis deserves more quantitative study [31]. This explanation seems more natural and straightforward than many of the other alternatives.

The simulations above differ from those presented previously in an important respect. Rather than randomly generating values using equation 22, the values were imposed externally. This was done in an attempt to make a qualitative comparison to a real price series. Though the details may differ, all of the properties above are also observed, in many cases more strongly, using randomly generated values.

As our point of comparison we use annual prices and dividends for the S&P index² from 1889 to 1984. Both series are adjusted for inflation. We use the dividends as a crude measure of value. We somewhat arbitrarily assume that the simulations are on a daily timescale and expand the dividend series to allow this. This is done by linearly interpolating 250 surrogates between each annual value of the logarithm of the dividends; 250 is chosen because it is roughly the number of trading days in a year. Thus the log-value series used as an input to the simulation contains a total of $250 \times 95 = 23,750$ numbers that vary linearly except for a discontinuous change in the derivative every 250 values. As mentioned above, the main criterion for choosing the parameters of the simulation was to adjust the capital of the trend followers to ensure that the autocorrelation of the price is zero; a secondary criterion was to adjust the external noise to match the volatility in the mispricing. The real series of American prices and values are shown in Figure 19 and the

1. The inclusion of trend followers in the mix is important; the autocorrelations in volatility are much weaker for a population of pure value investors. But all the simulations that have the autocorrelation of log-returns near zero show these properties, albeit to varying degree. Understanding the dependence on the mixture of trading strategies is an interesting topic for further research.

2. We would like Robert Shiller for making these data available on his web site. See Campbell and Shiller [36] and references [3, 7].

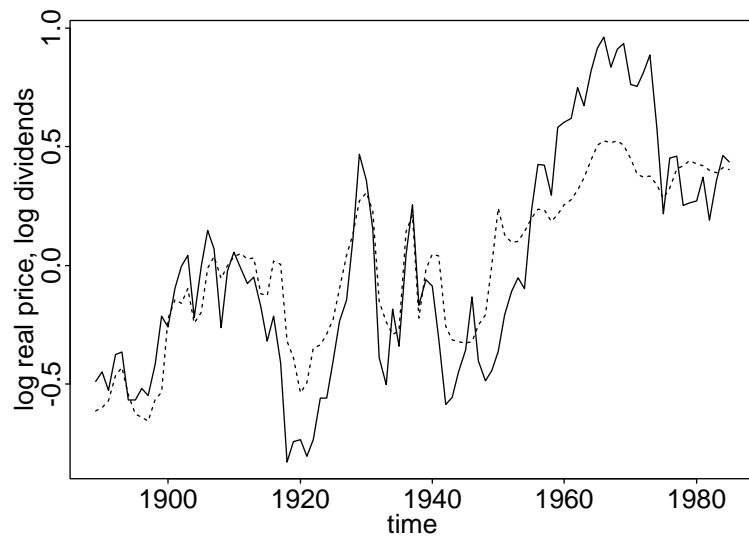


FIGURE 19. Inflation-adjusted annual prices (solid) and dividends for the S&P index of American stock prices.

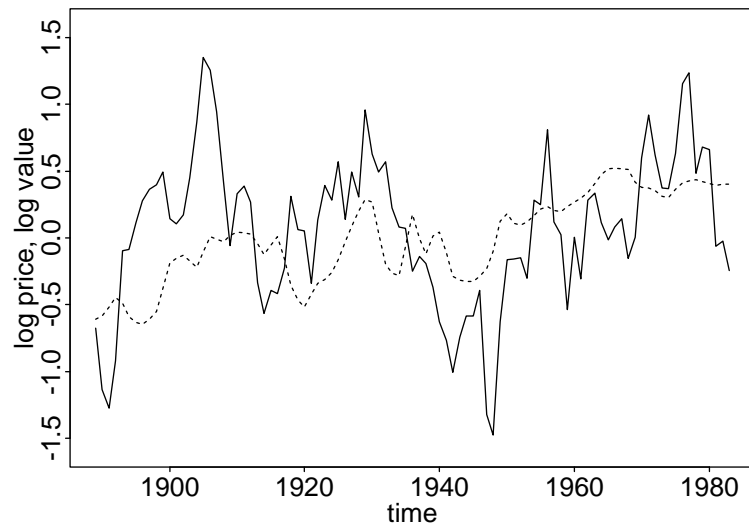


FIGURE 20. The price for a simulation consisting of value investors and trend followers using linearly interpolated dividend series from Figure 19 as inputs. The price was averaged over periods of 250 iterations to simulate the reduction of a series of this many trading days to annual data. There was some limited adjustment of parameters, as described in the text, but no attempt was made to match initial conditions. There is qualitative agreement in that the price fluctuates around the value in a similar manner

simulation results are shown in Figure 20. While these differ in detail, there is a certain qualitative correspondence. In both series the price fluctuates around value, and mispricing persist for periods that are sometimes measured in decades.

This comparison should obviously not be taken too seriously. No real attempt has been made to fit the parameters of the simulation. Furthermore, no attempt was made to match initial conditions. To do this it is necessary to initialize the states of the trend followers and value investors, which for state-dependent strategies is not trivial. Furthermore, the detailed price series generated by the simulation depends on the realization of the random process for the external noise. While I believe that all these problems can be solved, using models of this type for forecasting is beyond the scope of this paper [31].

The collection of strategies used in the American stock market is certainly far more complex than the simple examples used in the simulation. Nonetheless, the results in this section demonstrate an encouraging capability to reproduce qualitative features of the market, both in the long-term dynamics of the mispricing and in the temporal statistical properties of daily data.

3.6 Summary and discussion

The examples worked in this section demonstrate that the dynamics developed here is capable of producing sensible results. The strategies studied in this section are only a small subsample of those actually used in real markets. There are a large number of possible variations, combinations, and alternatives. The work here is just a first step and only begins to explore the complexity of real strategies. For example, the range of possible technical strategies goes far beyond the simple trend strategy explored here [21]. Most traders do not follow simple formulas. Trading has emotional components, such as fear and greed. While this may be difficult to take into account analytically, in some cases it may be possible. For example, for value strategies a fear factor can be incorporated by adding an enhanced willingness to sell on signs of high volatility when the market is strongly overvalued; this may give rise to crashes. It is beyond the scope of this paper to explore a broader range of strategies and their corresponding dynamical behaviors. We have demonstrated that even for the simple strategies studied here, observed market phenomena such as correlated volume and volatility and long tails in price returns emerge naturally, indeed are difficult to avoid.

Based on the nonequilibrium theory developed in the previous section, there is no *a priori* reason to assume cointegration between price and value. This depends on the collection of strategies that comprise the market. This depends on human behavior as reflected in the choice of trading strategies. Market impact may be only one of several market forces that influence price formation¹. A key question for future research will be to understand the degree to which different forces contribute to market dynamics. This theory makes it clear that only certain trading strategies are useful in causing cointegration; insofar as markets are cointegrated, we should observe these strategies in use. It is encouraging that the results in this section are qualitatively realistic, in that price and value track each other, but only weakly, with deviations on one side or the other for long periods of time. There is some evidence in the literature that price and value are more closely cointe-

grated in some markets than others; this model provides a good context for further study of this question¹.

Another interesting feature of the results in this section is that the dynamics can become unstable. In each case the dynamics become unstable if the capital of a given strategy exceeds a threshold. This raises the possibility that real markets may become unstable. To the extent that real markets are stable, it suggests that the relevant parameters, which depend on liquidity and capital, evolve to ensure this. This is studied in more detail in the next section.

In biology, ecology can be defined as “the study of the interrelationships of organisms with their environment and each other”. In an analogy to biology, an individual agent can be thought of as an organism and a strategy as the phenotype of a species. In this section we have demonstrated that this approach to market dynamics focuses on the interrelationships between strategies and naturally fosters an ecological point of view. The diversity of views generate an ecology of different strategies, each causing different effects that contribute to the overall dynamics. In the next section we will argue that the emergence of a diversity of complex strategies is natural in financial ecologies.

1. If the direct market impact of the order flow due to value strategies does not provide a mechanism to cointegrate price and value, the main alternative appears to be an information process in which market makers receive information that may come from sources other than orders, and adjust prices to keep them near value. An example might be the “force of arbitrage” alluded to earlier: Prices may change, even without trading, simply because everyone knows that arbitrage is possible. While I believe that such effects make a contribution to price dynamics, particularly in illiquid markets, I think it would be disturbing if there were not mechanisms that cointegrate price and value directly through trading.

1. The series of price and value for stock markets studied by Campbell and Shiller [36] appears to be cointegrated. For currency markets, in contrast, the evidence for cointegration of exchange rates and purchasing price parity is controversial, with evidence both for and against [37]. This is naturally predicted from this theory. Holding stocks or bonds produce ongoing revenues, while currencies do not. Thus one would expect a higher ratio of value investors, and stronger cointegration for stocks and bonds. It may be possible to predict the ratio of value investors to technical traders based on the relative importance of the revenue stream.

4. Evolution

In the discussion so far we have assumed that the capital of each trading strategy is fixed. In reality the capital varies as profits are reinvested, strategies change in popularity, and new strategies are discovered. In addition, market makers adjust the liquidity and the spread. All of these factors alter the financial ecology and change its dynamics. Adjustments in capital and liquidity are an important component of *market evolution*. This includes the emergence of new strategies (when the capital changes from zero to a finite value). Market evolution occurs on longer timescales, causing nonstationarity in the day to day market dynamics. There is feedback between the two timescales: The day to day dynamics determine profits, which affect capital reallocations on evolutionary timescales, which in turn affect the day to day dynamics.

Under the classic theory of market efficiency, new strategies should appear and capital should rapidly adjust to exploit any opportunities for profit making, in such a way that “abnormal profits” are impossible. Market efficiency, if it occurs, is an outcome of market evolution. This theory provides a convenient dynamical framework in which to investigate this question.

4.1 Mechanisms of financial evolution

4.1.1 Capital reallocation and separation of timescales

The capital determines the influence of each strategy on the dynamics. The capital of a strategy sets the scale of its influence on the ecology, and is analogous to the population of a species in biology. Ultimately decisions about capital allocation are entirely in the hands of human beings, and like most decision-making processes are difficult to model. Nonetheless, there are regularities in how capital is allocated. Three factors that influence this are:

- *Reinvestment of earnings.* The profits or losses of a strategy are added to existing capital.
- *Attracting capital because people believe a strategy is profitable.* Funds are organized that pool money from different investors, and capital is allocated by individuals or within organizations.
- *Restricting capital due to capacity limitations.* With excessive capital profits will decrease due to transaction costs. Competent traders attempt to understand this and maximize their profits by limiting capital accordingly. Funds close, and occasionally capital is even returned.

The reinvestment of earnings is straightforward to model. Let a be the fraction of profits that are reinvested, where $0 \leq a \leq 1$. The rate of change of the capital $c^{(i)}$ is

$$\Delta c_t^{(i)} = c_t^{(i)} - c_{t-1}^{(i)} = a g_t^{(i)}.$$

If reinvestment is sufficiently slow then the capital will vary slowly in comparison to single time period returns. As long as a is sufficiently small it is reasonable to approximate the gains using equation 20, which gives

$$\Delta c_t^{(i)} \approx a \left(\sum_{j=1}^N \tilde{G}_{ij} c^{(i)} c^{(j)} + \mu \langle \tilde{y}^{(i)} \rangle c^{(i)} \right). \quad (\text{Eq 35})$$

\tilde{G} and $\langle \tilde{y}^{(i)} \rangle$ generally depend on c . Equation 35 gives an approximation for the *replicator dynamics* of financial strategies. It is similar to the Lotka-Volterra equations, and makes precise the analogy to predator-prey systems and population biology [26].

Because it depends on human behavior the process of attracting capital is more difficult to model. The rate at which money flows in or out of a fund depends on many factors, such as advertising and salesmanship. Capital allocations inside an investment bank may depend on internal politics, regulatory restrictions, or taxes. Fads may dictate an overall preference for value strategies vs. trend strategies based on fluctuations in cultural mythology. Some people are more risk averse than others. Statistical fluctuations may cause profits purely by chance, and people sometimes make decisions based on statistically insignificant results. There are clearly many factors that influence capital allocation, such as the opportunities for investing in other assets. Nonetheless, it seems reasonable to assume that capital tends to flow into strategies that people believe will be profitable, and that this belief has at least some correlation with actual profitability¹. In this case we can use equation 35 as a model, with the alteration that it becomes possible that $a > 1$. An obvious and important extension would be to assume that capital reallocations are relative to the mean performance of all the strategies in the market, including those in other asset classes. Other enhancements such as the extension to multiple asset classes are clearly possible, but are beyond the scope of this paper².

As the capital of a strategy increases so does market friction. The capital eventually reaches a level where profits are at a maximum and the strategy has reached its capacity. If the traders using a given strategy understand this, once this level is reached they will cease to increase the capital of the strategy. When this occurs equation 35 becomes inoperative. Capital adjustments from then on only depend on changes in the profitability of the strategy, i.e. improvements in its capacity. This effect can be modeled in terms of a stopping condition. When this condition is met, the capital ceases to be modeled by equation 35, and from then on is based on optimizing profits for each trader. This introduces dependen-

1. An important exception are trading strategies that are motivated by reduction of risk or consumption rather than profit. An example is a farmer who buys a futures contract to lock in a price, or an agent who buys a good to mark it up and distribute to consumers. Such strategies are fundamental and can be viewed as drivers for more speculative strategies that seek to make profits.

2. Note that with the assumption of separation of timescales the rate of reinvestment is based on the expected profitability, which is a given number, rather than on past or perceived profitability, which are subject to statistical fluctuations and human error. Prospective investors often consider other measures, such as trailing return/risk ratio, or backtests based on historical data.

cies on the number of traders using a given strategy, which are discussed in more detail in Section 4.5.

Because of statistical fluctuations and estimation problems it is often difficult to distinguish between profitable and unprofitable strategies. Understanding transaction costs is also challenging, as evidenced by the fact that the market impact function is still not well described in the published literature. Thus we can expect that many investors will follow strategies that are not profitable, and many traders will fail to understand their market impacts, ramping the capital of their strategies until they cease to be profitable. This will be studied in Section 4.4

4.1.2 Long term adjustments in liquidity

Although the main focus of this paper is on trading strategies, evolutionary changes in the dynamics also occur due to the adjustment of liquidity or the spread by the market maker. From the point of view of the dynamics, increasing the liquidity is equivalent to increasing the capital of all the strategies. An example of how the profits of the market maker depend on the liquidity is shown in Figure 21.

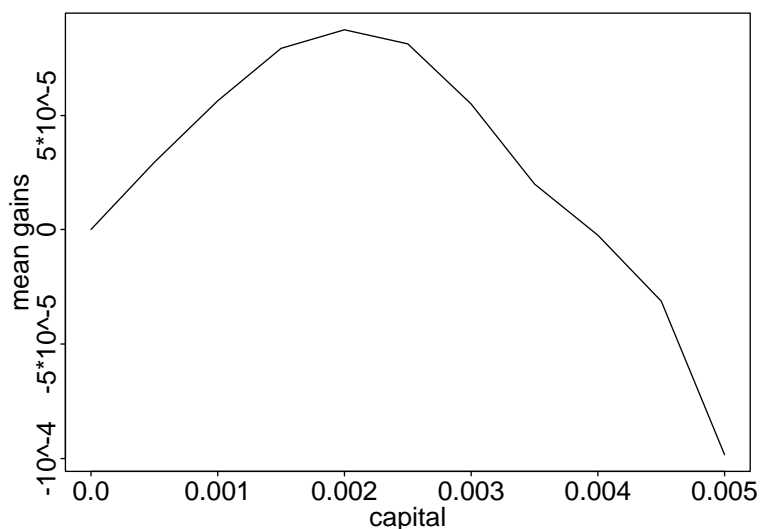


FIGURE 21. Mean profits of value investors as a group vs. the capital of individual traders. There are 1000 traders with entry thresholds uniformly distributed between 0.3 and 7 and exit thresholds uniformly distributed between -0.3 and 0 . $\sigma_{\xi} = \sigma_{\eta} = 0.02$, $v_{min} = -1$, $v_{max} = 1$, $b = 0.2$, and $\bar{\lambda} = 1$.

In a market consisting only of value investors all using the threshold-based value strategy of Section 3.2.3, they make a profit as a group when the capital is sufficiently low, and take losses as a group when the capital is sufficiently high. Because the position of the market maker is the negative of the position of the value investors as a group, the profits of

the market maker are also the negative of their profits. Similarly, the liquidity is inversely proportional to the capital. Thus this figure demonstrates that the market maker profits when the liquidity is low, and takes losses when the liquidity is high. This makes it clear that market makers have an incentive to keep the liquidity low to make profits. However, market making profits are proportional to volume, and with more realistic strategies, if the liquidity is too low we would expect the volume to decrease. This suggests that the market maker will lower the liquidity to make a profit, but not make it too low. The market maker can also make adjustments in the spread; this affects profits and losses but does not directly affect the dynamics of the midpoint of the price. Competition will tend to drive the market maker profits near the threshold. This makes it clear that liquidity and capital co-evolve. Given the demonstration in equation 21 that market friction generally causes losses, it may seem surprising that there are conditions where value investors can make a profit as a group. My conjecture is that this is because they collectively alter the dynamics; even though the diagonal terms in the gain matrix of any given trader are still negative, the off-diagonal terms are such that they can make a profit. This deserves further investigation.

4.2 Competition and diversity in financial ecologies

How diverse are financial ecologies? If there were a single optimal strategy, then we should expect to find only that strategy in the market. However, it is clear that financial markets are extremely diverse. In practice many different strategies are used [13]. In this section we demonstrate why such diversity comes about.

4.2.1 Competition between different classes of strategies

When do different strategies co-exist? To investigate this question we use a standard technique from population genetics. Assume a pre-existing set of strategies. If we introduce a new strategy does it make a profit? If it does, then according to equation 35 it will *invade* the population. Such calculations are easy to perform because we can use the approximation that the capital of the invading strategy is small, which means that it has a negligible effect on the price dynamics. The ability of a strategy to invade a population does not necessarily mean that the diversity increases over the long term, but it does imply that it will increase in the short term. (Over the long term it may co-exist with the other strategies, or it may drive some of them extinct, which might even cause the new strategy to become extinct as well).

For instance, consider the position based strategies of Section 3.2.2 in a market that is dominated by traders using the order based strategies of Section 3.2.1. Consider the case that the position-based trader perceives a different value than the consensus of the order based traders, so the position is

$$y_{t+1}^{(p)} = -c(z_t - \tilde{v}_t) = -cm_t + c\delta v \quad (\text{Eq 36})$$

where $\tilde{v}_t = v_t + \delta v$ is the value perceived by the position-based trader and v_t is the consensus value of the order based traders. For convenience the difference δv is assumed to be constant in time.

Substituting equations 24 and 36 into equation 16, to first order in c the gains of the position-based trader are approximately

$$g_{t+1} \approx r_{t+1} y_t^{(p)} \approx (-\alpha m_t + \xi_t)(-c m_{t-1} + c \delta v). \quad (\text{Eq 37})$$

Taking averages implies that

$$\langle g_{t+1} \rangle \approx c \alpha \langle m_t m_{t-1} \rangle = c \alpha \rho_m(1) \sigma_m^2 = \frac{c(1-\alpha)(\sigma_\eta^2 + \sigma_\xi^2)}{(2-\alpha)}.$$

Thus the position-based strategy is generally able to invade the order-based strategy when $\alpha < 1$. Similarly, by flipping the sign (so that it is an “anti-value” strategy), its opposite can invade the order-based strategy when $\alpha > 1$.

The mean gains do not depend on whether the position-based trader’s estimate of value matches the consensus estimate. However, it is clear from equation 37 that failing to match the value, i.e. $\delta v \neq 0$, increases the risk. Since traders are generally risk averse, this creates an incentive to match one’s perceived value to the consensus estimate. Note that it is irrelevant whether the consensus perceived value is actually correct¹.

A similar calculation is possible for the simple trend following strategy of Section 3.3.2 invading order or position-based value strategies. The results are summarized in Table 1. There is no result given for a position-based value strategy invading itself, since the position and thus the gains are non-stationary.

From Table 1 we see that in the appropriate parameter ranges both the simple position-based value strategy and the trend strategy with $\theta = 1$ are able to invade the order-based strategy. This depends both on α (the ratio of capital to liquidity) and on the relative importance of changes in value and external noise as drivers of the stochastic part of the dynamics. Not surprisingly, neither of them are able to invade when $\alpha = 1$. However, when $\alpha \neq 1$, if these strategies cannot invade, then the “anti-strategy” with the reverse sign can.

Since the position-based value strategy induces negative autocorrelations in the returns, it may be surprising at first glance that the trend strategy is able to invade it. The reason has to do with the time-lags inherent in taking a position and taking profits: It takes two timesteps to observe a market movement, take a position, and profit from it. For the simple trend strategy with $\theta = 1$ invading another strategy with autocorrelation ρ_r and standard deviation σ_r , the mean gains are generally

$$\langle g_{t+1} \rangle \approx (\langle r_{t+1} y_t \rangle = c \langle r_{t+1} r_{t-1} \rangle = c \rho_r(2) \sigma_r^2).$$

1. This conclusion may change in markets where there are payments, e.g. dividends.

	position-based value strategy	trend following strategy with $\tau = 1$
order-based	$c\alpha\rho_m(1)\sigma_m^2$ $= \frac{c(1-\alpha)(\sigma_\eta^2 + \sigma_\xi^2)}{(2-\alpha)}$	$c(1-\alpha)\rho_r(1)\sigma_r^2$ $= c \frac{(1-\alpha)(\alpha(1-\alpha)\sigma_\eta^2 - \alpha\sigma_\xi^2)}{(2-\alpha)}$
position-based	--	$-c\alpha\rho_r(1)\sigma_r^2$ $= c\alpha^2 \left(\frac{\alpha^2\sigma_\eta^2 + \sigma_\xi^2}{1-\alpha^2} \right)$

TABLE 1. Mean gain to first order in c when the strategy listed in the columns invades the strategy listed in the rows. In each case c is the capital of the invading strategy, and α is the ratio of the capital of the strategy being invaded to the liquidity. When the gain is positive it implies that the strategy will be able to invade.

Thus, because of the lag needed to take up a position and obey causality the gains of this strategy depend on the second autocorrelation.

These calculations are easily extended to second order¹ in c . To do this assume that the log-return is the sum of the log-returns induced by each of the two strategies. For example, when the position-based strategy invades the order-based strategy the result is

$$\langle g_i \rangle \approx \frac{c(1-\alpha-c/\lambda)(\sigma_\eta^2 + \sigma_\xi^2)}{(2-\alpha)}.$$

This expression is quadratic in c . The added term comes about because of the self-interaction of the invading strategy. The profits grow linearly when c is small and reach a maximum at $c = (1-\alpha)/2$. This behavior is generic -- as a function of capital, the gains of a profitable strategy will go through a quadratic maximum and then decline. Simulation results illustrating this result are given in Figure 22

4.2.2 Competition within a given class of strategies

We can also ask whether one member of a family of strategies can invade other members of the same family. For example, consider the family of simple trend following strategies of Section 3.3.2. Can a trend strategy having delay parameter θ_2 invade another with delay parameter θ_1 ? The position of the invading strategy is

1. These simple strategies are linear, so it is possible to solve for the gains exactly, even when they are combined. Nonetheless, the easy calculation done here is all that is needed to illustrate the point.

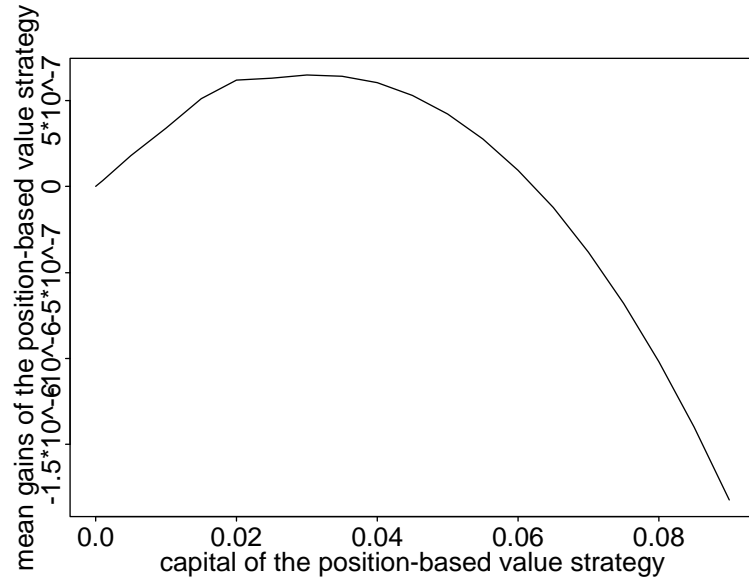


FIGURE 22. The gains to the position-based value strategy as it invades the order-based value strategy. The capital of the order-based strategy is fixed at 0.9. The noise terms for price and value have standard deviation 0.01. Data was averaged over 20,000 time steps.

$$y_t = c(z_{t-1} - z_{t-\theta_2-1}) = c \sum_{i=1}^{\theta_2} r_{t-i}.$$

To first order the dynamics are given by equation 34 with $\theta = \theta_1$. The mean gains are

$$\begin{aligned} \langle g_{t+1} \rangle &\approx \left(\langle r_{t+1} y_t \rangle = c\alpha \sum_{i=1}^{\theta_2} (\langle r_t r_{t-i} \rangle - \langle r_{t-\tau_1} r_{t-i} \rangle) \right) \\ &\approx c\alpha\sigma^2 \sum_{i=1}^{\theta_2} (\rho(i) - \rho(\theta_1 - i)) \end{aligned}$$

ρ and σ are the autocorrelation and standard deviation of the dynamics with $\theta = \theta_1$. The autocorrelation function is symmetric, i.e. $\rho(n) = \rho(-n)$. Furthermore, $\rho(0) = 1$ and $\rho(1) > \rho(n)$ for $n > 1$. To simplify the notation, let $k = c\alpha\sigma^2$. Enumerating a few examples demonstrates that short term trend followers are able to invade longer term trend followers. (Note that taking the spread into account may alter this conclusion, as the spread favors longer term strategies).

θ_1	θ_2	mean gain	θ_2 invades θ_1 ?
1	1	$k(\rho(1) - 1) < 0$	no
2	1	$k(\rho(1) - \rho(1)) = 0$	no
$n > 2$	1	$k(\rho(1) - \rho(n)) > 0$	yes
1	$m > 1$	$k(\rho(m) - 1) < 0$	no

TABLE 2. Can a trend following strategy with timescale $\theta = \theta_2$ invade another with timescale $\theta = \theta_1$? The third column shows the mean gain under the approximation that the capital of the invader is small. When the gain is positive the invader makes profits, and will invade. This shows that short term trend followers can profit from longer term trend followers.

For a new strategy to exploit an old strategy, an obvious way to take advantage of temporal structure is through technical trading, i.e. making the strategy dependent on past price values. The examples worked in this section suggest that the natural path toward market efficiency is through increased diversification.

4.3 Efficiency

4.3.1 Definition of an efficient market

The basic idea of the theory of efficient markets is that the act of exploiting patterns to make a profit alters the market and causes the original patterns to disappear [22, 23]. Efficiency is essentially an evolutionary question, i.e., we can only hope that efficiency will happen as a result of the introduction of new strategies and the readjustment of capital and liquidity. In their introductory textbook, Sharpe et al. define market efficiency as follows [25]:

“A market is efficient with respect to a particular set of information if it is impossible to make abnormal profits by using this set of information to formulate buying and selling decisions.”

The caveat about “abnormal profits” deals with situations where profits are normal, for example, the tendency of the stock market to rise, or the ability of market makers to profit by taking the spread. Abnormal profits correspond to making risk-adjusted profits in excess of a broad market index, or a market maker who makes profits above and beyond what is needed to pay employees, cover costs, and make a living. Clearly, there is room for interpretation in what is considered “abnormal”. Within the context of this model we will define “normal” profits as those of the market maker, or those of a buy-and-hold strategy driven by a positive drift term. With these two exceptions, a market is efficient if no one makes profits on average, i.e. if

$$\langle g_t^{(i)} \rangle \leq 0.$$

for every i (not including the market maker). Note that since the gains are $g_t \approx r_t y_{t-1}$, the ability to make a profit implies a *pattern* π_t , defined as the expectation of the return conditioned on information available two timesteps earlier, i.e.

$$\pi_t = \langle r_t | z_{t-2}, \dots, I_{t-2} \rangle \neq 0. \quad (\text{Eq 38})$$

The lag of two takes into account the fact that the gains depend on the position at the previous timestep, and the position depends on information from the timestep before that. The information set is the history of past prices and possibly other external information I , e.g. which might be used to assess value. The information set does not normally include a record of the past trades of the other players. For the purposes of this paper we are generally neglecting the spread. This means that for any pattern there is a profitable strategy that can exploit it. In fact, spread implies there is a threshold below which a pattern cannot be exploited; once a pattern is below this threshold it can be considered irrelevant.

Efficiency hinges on whether patterns in prices exist, and if they do exist, whether they persist when they are exploited to make profits. The logic driving early arguments for market efficiency is represented in the following statement by Cootner [24] in 1964.

“If any substantial group of buyers thought that prices were too low, their buying would force up the prices. The reverse would be true for sellers... the only price changes that would occur are those that result from new information. Since there is no reason to expect that information to be non-random in appearance, the period-to-period price changes of a stock should be random movements, statistically independent of one another.”

Cootner’s starting hypothesis that buying forces up prices is the basis for the theory developed in Section 2. The understanding of market efficiency has evolved considerably since 1964, but his statement makes it clear why the theory presented here provides a simple context in which to investigate it.

It is useful to distinguish between principle and practice. If there exist no profitable strategies, then we will say that the market is *efficient in principle*. If there exist profitable strategies, but such strategies are unlikely to be found by any reasonable algorithm, then we will say that it is *efficient in practice*. Efficiency in principle has the advantage that it is clear-cut and easy to study. Efficiency in practice is more relevant, but suffers from the vagueness of concepts like “unlikely” and “reasonable”. We will mainly investigate efficiency in principle, only making a few speculations about efficiency in practice.

4.3.2 Efficiency via complexity

A market is inefficient if there are predictable trade imbalances creating mean price movements that are larger than the spread. When a market is composed of enough independent strategies, under certain assumptions trade imbalances will become relatively smaller according to the law of large numbers. This depends on the complexity and degree

of independence of the strategy. In this section we construct an example of a market consisting of independent strategies. In the limit that the number of strategies is sufficiently large, they tend to cancel each other's market impact, creating an efficient market.

Consider a set of randomly chosen “binary” technical trading strategies that depend on the signs of the M previous returns. The space of possible inputs can be represented as a bit string of length M . A given strategy can be constructed as a look-up table by randomly assigning a buy or a sell order $\pm c$ to each of the 2^M possible inputs. (There are $2^{2^{M+1}}$ distinct strategies that can be constructed in this way; note that this number gets huge quickly as M grows large.) Suppose we choose a set of N strategies in this way. How efficient is the resulting market?

When N is large, from equation 11 the deterministic component of volatility (the standard deviation of the deterministic part of the log-returns) is roughly $(c/\lambda)\sqrt{2N/\pi}$ and the volume is cN , so the ratio of volatility to volume is

$$\frac{\text{volatility}}{\text{volume}} \approx \frac{1}{\lambda} \sqrt{\frac{2}{\pi N}}. \quad (\text{Eq 39})$$

If we assume that a fixed fraction of the volume is random “noise trading”, the relative size of the deterministic price movements decreases as $1/\sqrt{N}$. As the number of diverse trading strategies increases the market becomes relatively more efficient.

This example illustrates that for a market to be efficient the strategies must cover their space of inputs uniformly. Each possible state of the market must generate a balanced volume of buy and sell orders. In contrast, if the input conditions for the strategies are clustered, there will be bursts of net buying or selling activity. This implies not only clustered volatility and volume, but also potentially exploitable patterns. For the market to become efficient the population of strategies must evolve so that their inputs are evenly distributed throughout the space of possibilities.

It is clear that a complex space of strategies makes a market more efficient in practice, even if it is not efficient in principle. Suppose you wish to fit a nonlinear timeseries model to predict future price movements for the example above. The goodness of fit will depend on the signal to noise ratio, which as shown in equation 39 decreases with N . In addition, since the deterministic structure in this example is random by construction, the fit will suffer the classic “curse of dimensionality”, and the number of data points needed to get a fit with a given error level will increase exponentially with M . In general, it is worth noting that nonlinear estimation problems require a considerable degree of skill; we can expect that some traders will perform this task better than others.

4.4 Pattern evolution

Do patterns in the market disappear once they are discovered? To profit from a pattern requires trading that otherwise would not have occurred. The market impact of this trading alters prices, which in turn alters the original pattern. In this section we illustrate this for

the simple case of a temporally isolated pattern, i.e. one that is concentrated at a particular time. An isolated pattern is of the form

$$\Pi = (\dots, 0, \pi_{t+1}, 0, \dots),$$

where $\pi_{t+1} = \langle r_{t+1} | z_{t-1}, \dots, I_{t-1} \rangle \neq 0$. This pattern might be generated by a trader or group of traders who buy or sell contingent on a particular event, e.g. at a particular time of year, or in response to a particular mispricing or trend level. If the pattern is recurrent we can think of this as a time average; alternatively, it is perhaps more useful to consider an ensemble average of the form

$$\langle r_{t+1} | z_{t-1}, \dots, I_{t-1} \rangle = \int r_{t+1} P(r_{t+1} | z_{t-1}, \dots, I_{t-1}) dr_{t+1},$$

where $P(r_{t+1} | z_{t-1}, \dots, I_{t-1})$ is the conditional probability density of r_{t+1} given prices and other information available at time $t-1$. The differential dr_{t+1} may be complicated because it depends on the noise (ξ_{t-1}, ξ_t) and possibly (η_{t-1}, η_t) (or in a more general context other random information that might alter the trading at times t and $t-1$). Rather than doing this calculation exactly, we will just make an approximation.

To understand the evolution of the pattern we need to state what originally caused it. To simplify the notation, let

$$\Omega(z_t, \dots) = \sum_i \omega^{(i)}(z_t, \dots).$$

Assume the original pattern is caused by a net order imbalance,

$$\pi_{t+1} = \langle r_{t+1} \rangle = \frac{1}{\lambda} \langle \Omega(z_t, \dots) \rangle.$$

The assumptions $\pi_t = 0$ and $\pi_{t+2} = 0$ imply that

$$\langle \Omega(z_{t-1}, \dots) \rangle = 0$$

$$\langle \Omega(z_{t+1}, \dots) \rangle = 0$$

A new trader can profit from this pattern by taking up a position c at time t of the same sign as π_{t+1} . Assuming his position is initially zero, to enter this position he needs to make a trade c , and to exit he needs to make a trade $-c$. Under the simple canonical market model of equation 11, assuming there are no other nearby patterns it is natural to enter the position at time t and exit at time $t+1$, as this minimizes risk. The new trading only alters prices only for times t or greater.

We can compare the new pattern, including the new trading, to the original pattern. Quantities involving the new trading will be denoted by “ \sim ”. The evolved pattern

$$\tilde{\Pi} = (\dots, \tilde{\pi}_t, \tilde{\pi}_{t+1}, \dots)$$

is

$$\begin{aligned}
\tilde{\pi}_{t-1} &= \pi_{t-1} \\
\tilde{\pi}_t &= \langle \tilde{r}_t | z_{t-2}, \dots \rangle = \frac{c}{\tilde{\lambda}} q \\
\tilde{\pi}_{t+1} &= \langle \tilde{r}_{t+1} | z_{t-1}, \dots \rangle = \frac{1}{\tilde{\lambda}} (\langle \Omega(\tilde{z}_t, \dots) \rangle - c) . \\
\tilde{\pi}_{t+2} &= \langle \tilde{r}_{t+2} | z_t, \dots \rangle = \frac{1}{\tilde{\lambda}} (\langle \Omega(\tilde{z}_{t+1}, \tilde{z}_t, \dots) \rangle) \\
\tilde{\pi}_{t+3} &= \dots
\end{aligned}$$

The first equality simply states that any pattern at times less than t is unaffected by trading at times t or greater. To a trader who knows with certainty that the trade c at time t will happen the new pattern $\tilde{\pi}_t = c/\lambda$; however, in general this may not be known. The factor $0 \leq q \leq 1$ takes into account the fact that the information available about this trade depends on the information available about the original pattern and the extent to which the trade c might be telegraphed by something else.

To simplify matters, in computing the evolved prices it is convenient to assume the same sequence of noise fluctuations with and without the new trades. This is in the spirit of comparing what would have happened with the new trades to what would have happened without them. The log-price can be computed by summing the log-returns, making use of the fact that the price is unaltered at time $t-1$.

$$\begin{aligned}
\tilde{z}_{t-1} &= z_{t-1} \\
\tilde{z}_t &= z_t + \frac{c}{\tilde{\lambda}} \\
\tilde{z}_{t+1} &= z_{t+1} + \frac{1}{\tilde{\lambda}} (\Omega(\tilde{z}_t, \dots) - \Omega(z_t, \dots))
\end{aligned} \tag{Eq 40}$$

Note that at time $t+1$ the direct market impact of the new trades c and $-c$ cancels out, but there is indirect market impact as reflected in a possible change in the net of the orders, which can alter the price.

If Ω is a smooth function whose derivatives exist then providing c is small enough we can approximate $\Omega(\tilde{z}_t, \dots)$ using Taylor's theorem.

$$\begin{aligned}
\tilde{\pi}_{t+1} &= \langle \tilde{r}_{t+1} \rangle \approx \frac{1}{\tilde{\lambda}} \left(\langle \Omega(z_t, \dots) \rangle + \left\langle \frac{\partial \Omega}{\partial z_t} \delta z_t \right\rangle - c \right) \\
\tilde{\pi}_{t+2} &= \langle \tilde{r}_{t+2} \rangle \approx \frac{1}{\tilde{\lambda}} \left(\langle \Omega(z_{t+1}, z_t, \dots) \rangle + \left\langle \frac{\partial \Omega}{\partial z_{t+1}} \delta z_{t+1} \right\rangle + \left\langle \frac{\partial \Omega}{\partial z_t} \delta z_t \right\rangle \right)
\end{aligned} \tag{Eq 41}$$

where the derivatives are evaluated at the original prices, e.g. (z_{t+1}, z_t, \dots) , and $\delta z_t = \tilde{z}_t - z_t$. By assumption

$$\begin{aligned}\pi_{t+1} &= \langle \Omega(z_t, \dots) \rangle / \lambda \\ \pi_{t+2} &= \langle \Omega(z_{t+1}, z_t, \dots) \rangle / \lambda = 0\end{aligned}$$

Furthermore, from equation 40, $\delta z_t = c/\lambda$ and using Taylor's theorem again,

$$\delta z_{t+1} = \frac{1}{\lambda} (\Omega(\tilde{z}_t, \dots) - \Omega(z_t, \dots)) \approx \frac{\partial \Omega}{\partial z_t} \delta z_t.$$

To simplify the notation let

$$\gamma_t^j = \left\langle \frac{\partial z_{t+1}}{\partial z_{t-j}} \right\rangle = \frac{1}{\lambda} \left\langle \frac{\partial \Omega}{\partial z_{t-j}}(z_t, z_{t-1}, \dots) \right\rangle.$$

Collecting these relations together and substituting into equation 41 makes it possible to get a simple estimate for the evolved patterns.

$$\begin{aligned}\tilde{\pi}_t &= \frac{c}{\lambda} q \\ \tilde{\pi}_{t+1} &\approx \pi_{t+1} - (1 - \gamma_t^0) \frac{c}{\lambda}, \\ \tilde{\pi}_{t+2} &\approx (\gamma_{t+1}^0 \gamma_t^0 + \gamma_{t+1}^1) \frac{c}{\lambda}\end{aligned} \tag{Eq 42}$$

where the last expression also requires the further approximation that

$$\left\langle \frac{\partial \Omega}{\partial z_{t+1}} \frac{\partial \Omega}{\partial z_t} \right\rangle \approx \gamma_{t+1}^0 \gamma_t^0.$$

Similar expressions are possible for $\tilde{\pi}_{t+3}$, π_{t+4} , etc., but as long as $|\gamma_t^j| \ll 1$ the disturbance to the original pattern diminishes with increasing time.

The quantity γ_t^j describes the sensitivity of the price at one time to changes in the price at an earlier time. We will call it the *price sensitivity*. Value investing strategies tend to have negative price sensitivity $\gamma^0 < 0$, and trend following strategies tend to have positive price sensitivity $\gamma^0 > 0$. In the example above, providing $\gamma_t^0 < 1$, $\tilde{\pi}_{t+1}$ will diminish¹. It is also the case that the price dynamics are linearly unstable when $\gamma_t^0 > 1$. Thus, providing the dynamics are linearly stable, as a pattern that is entirely concentrated at one time is exploited, it will evolve into one that is smoothed out in time, i.e. $|\tilde{\pi}_t| \geq 0$ and $|\tilde{\pi}_{t+1}| < |\pi_{t+1}|$. The extent to which the pattern evolves depends on the capital used to exploit it. We can naturally assume that the capital will be increased in an attempt to make more profits. There are two natural possibilities to consider:

1. The fact that the pattern diminishes at time $t+1$ when $\gamma_t^0 < 1$ can be seen by examining $\delta \pi_{t+1}$, and recalling that π_{t+1} and c are of the same sign.

- The capital is increased until the gains are maximized.
- The capital is increased until the gains go to zero.

The first case assumes that the trader exploiting the pattern understands his transaction costs, and stops increasing the capital when the profits are maximized. The second case is what would occur if profits are simply blindly re-invested.

The mean gains from exploiting the pattern are

$$\bar{g}(c) \approx \tilde{\pi}_{t+1} c \approx \left(\pi_{t+1} - (1 - \gamma) \frac{c}{\lambda} \right) c.$$

Assuming $\gamma < 1$, the gains as a function of c are approximately an inverted parabola with maximum determined by $\partial \bar{g} / \partial c = 0$. The maximum occurs when

$$c = \frac{\lambda \pi_{t+1}}{2(1 - \gamma_t^0)}.$$

The mean gains at the maximum are approximately

$$\bar{g}_{max} = \frac{\lambda \pi_{t+1}^2}{4(1 - \gamma_t^0)},$$

and the evolved pattern is approximately

$$\begin{aligned} \tilde{\pi}_t &= \frac{q \pi_{t+1}}{2(1 - \gamma_t^0)} \\ \tilde{\pi}_{t+1} &= \frac{\pi_{t+1}}{2} \\ \tilde{\pi}_{t+2} &= \frac{(\gamma_{t+1}^0 \gamma_t^0 + \gamma_{t+1}^1) \pi_{t+1}}{2(1 - \gamma_t^0)} \end{aligned}.$$

In contrast, if the trader simply keeps reinvesting and increasing the capital until the gains go to zero, the pattern will evolve until it is approximately

$$\begin{aligned} \tilde{\pi}_t &= \frac{q \pi_{t+1}}{(1 - \gamma_t^0)} \\ \tilde{\pi}_{t+1} &= 0 \\ \tilde{\pi}_{t+2} &= \frac{(\gamma_{t+1}^0 \gamma_t^0 + \gamma_{t+1}^1) \pi_{t+1}}{(1 - \gamma_t^0)} \end{aligned}.$$

This is illustrated in Figure 23. We see that as the capital is increased the pattern evolves earlier in time. At the point where the gains are a maximum, assuming the dynamics are stable, the pattern at time $t + 1$ is half its previous size. If the trader over-capitalizes the strategy so that the gains go all the way to zero, the pattern is entirely shifted to the previous timestep.

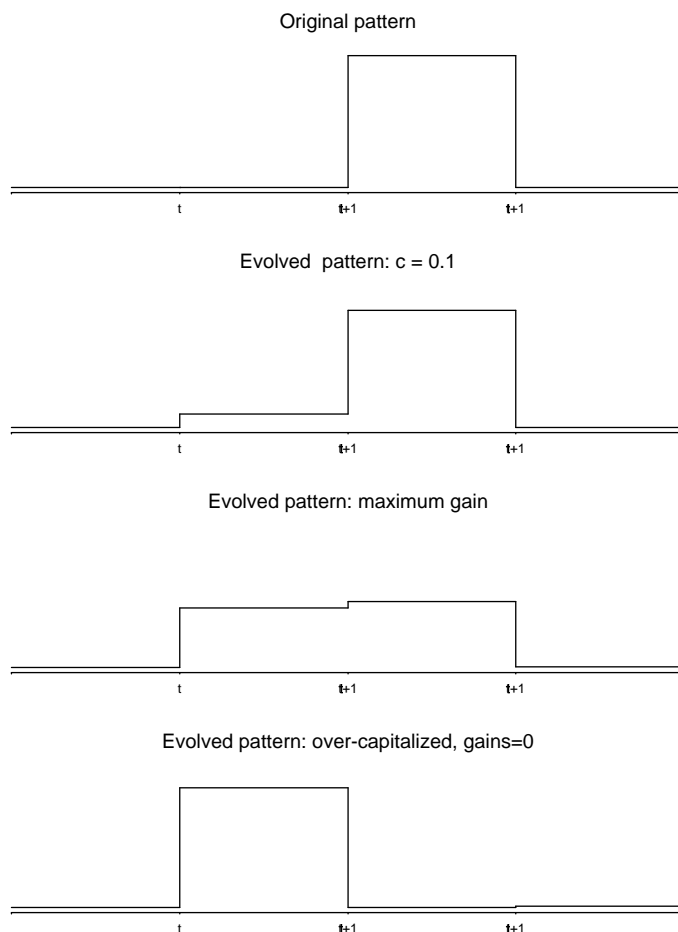


FIGURE 23. The evolution of an isolated pattern as it is exploited with increasing capital. The price sensitivities are $\gamma_t^0 = \gamma_{t+1}^0 = -0.1$ and $\gamma_{t+1}^1 = 0$ throughout, and $q = 1$. As the capital is increased to $c = 0.1$, the pattern is diminished at time $t + 1$ and enhanced at time t . As c is increased this trend continues. The gains are maximized at $c \approx 0.45$, and the pattern is spread between t and $t + 1$. If the strategy is over-capitalized to the point that the gains go to zero, the original pattern is entirely shifted to the previous timestep. It is typically diminished in size depending on q and the price sensitivity.

Figure 24 shows the effect of the price sensitivity in the case where the gains are maximized. If the new trader adjusts his trades to maximize his gains, the evolved pattern at time $t + 1$ is half as big as it was before, independent of the price sensitivities. However, the size of the new pattern at time t and time $t + 2$ both depend on the price sensitivity. If

the price sensitivity is zero, $\tilde{\pi}_t$ is also half the size of the original pattern when $q = 1$, but it is greater than half if the price sensitivity is positive, and less than half if it is negative.

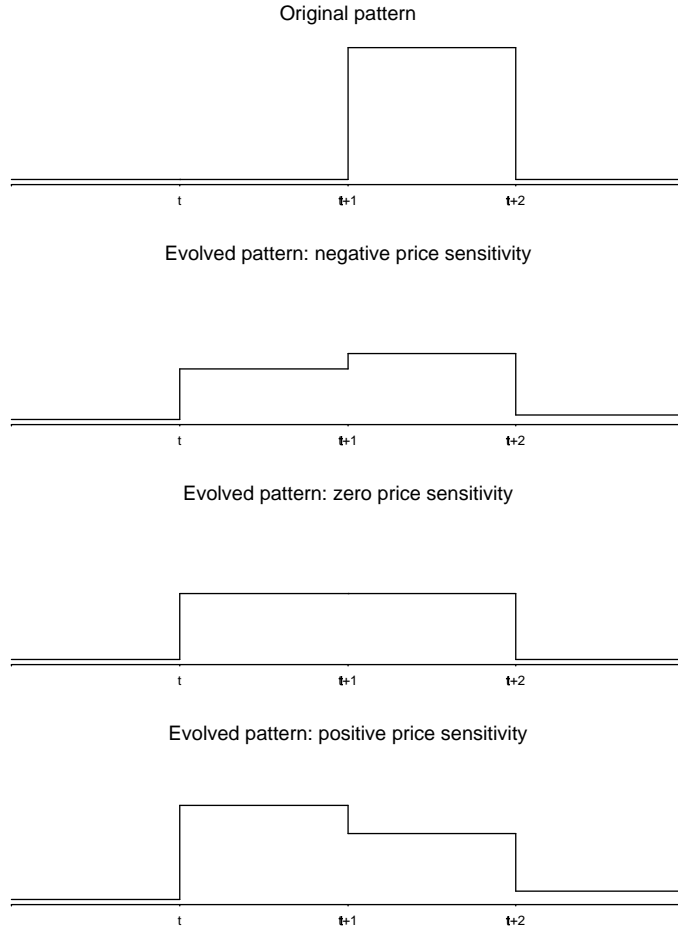


FIGURE 24. The effect of the price sensitivity on pattern evolution. Assume the trader adjusts his capital to maximize his gains. The pattern at time $t + 1$ evolves to half its original size, independent of the price sensitivity. When $q = 1$ at time t it is less than, equal to, or greater than half the size of the original, depending on whether the price sensitivity is negative, zero, or positive; since typically $q < 1$ it is diminished accordingly.

4.5 Several traders in the same niche

Suppose more than one trader discover a pattern. If they all maximize their own profits, what is the effect on the original pattern? If there are a total of N traders, each of whom make trades $(c_i, -c_i)$, from equation 42 the mean gain to trader i is

$$\bar{g}_i = \frac{1}{\bar{\lambda}} \left(\pi_{t+1} - \frac{(1 - \gamma_t^0)}{\lambda} \sum_{j=1}^N c_j \right) c_i.$$

Given that all the others make trades $(c_j, -c_j)$, the maximum gain for trader i occurs when

$$\frac{\partial \bar{g}_i}{\partial c_i} = 0 = \pi_{t+1} - \frac{(1 - \gamma_t^0)}{\lambda} \left(\sum_{j=1}^N c_j + c_i \right).$$

Applying this to each trader $i = 1, \dots, N$, the solution to the resulting set of equations is

$$c_i = \frac{\lambda \pi_{t+1}}{(N+1)(1 - \gamma_t^0)}.$$

The new pattern at time $t+1$ is

$$\tilde{\pi}_{t+1} = \frac{\pi_{t+1}}{N},$$

and the resulting mean gain for each trader is

$$g = \frac{\lambda \pi_{t+1}^2}{N(N+1)(1 - \gamma_t^0)}.$$

Thus as N grows larger the original pattern rapidly disappears. This is in marked contrast to what would occur if the agents were to cooperate, and limit their trading so that

$$c_i = \frac{\lambda \pi_{t+1}}{2N(1 - \gamma_t^0)}.$$

In this case their individual gains are

$$g = \frac{\lambda \pi_{t+1}^2}{4N(1 - \gamma_t^0)}.$$

They are clearly much better off when they cooperate.

This is a classic example of a competitive vs. a cooperative optimum. This makes it clear why traders are so secretive about what they do -- profits diminish rapidly as others discover the same niche. As niches become inhabited by many agents the strategy becomes overcapitalized and the original pattern disappears. As shown in the previous section, however, overcapitalization creates a new, earlier pattern, which is diminished providing $q/(1 - \gamma_t^0) \leq 1$.

4.6 Timescale for efficiency

One of the key questions about market efficiency is the timescale on which it occurs (if it occurs). This depends on the characteristic time for major capital reallocations. While

there are many uncertainties, by examining the factors for market evolution discussed in Section 4.1.1 it is possible to at least make order of magnitude estimates of the timescale.

Pure reinvestment is the simplest case. Annual returns from the stock market are on the order of 10-15% per year, and an average of 25% over many years is considered exceptional. With a 25% annual return under reinvestment it takes roughly 10 years to increase the funds under management by an order of magnitude. Naturally the capital adjustment needed to optimize profits from any given starting point depends on the size of that starting point; nonetheless, it seems reasonable to expect that adjustments in capital of orders of magnitude are generally needed. We can expect that, even considering full reinvestment ($a = 1$ in equation 35), the timescale to optimize profits is measured in decades.

The rate at which capital can be attracted from outside sources is obviously much more variable and difficult to analyze. In principle capital can be raised instantaneously to increase funds to the level where profits are maximized. In practice this is highly unusual. In fund management the rule of thumb is that a five year track record is needed to attract serious money. Once such a track record is obtained it typically takes many years to reach capacity. In investment banks capital may be allocated more rapidly. However, even in banks new strategies are usually tested for several years at lower levels of capital.

As shown in the previous section, the other factor that has significant impact on the capital allocations to a strategy is the number of traders pursuing a given strategy. As the number of independent agents becomes large and each of them achieves the optimal capital level, the strategy as a whole becomes only marginally profitable. How quickly will this happen?

There are two main sources for the creation of new strategies: Information diffusion and independent discovery. These interact, in that independent discovery may be stimulated by information diffusion. For some types of trading strategies, such as options pricing, there is a considerable overlap with academia. There are published papers, conferences, and other interactions that make information diffusion a strong effect. For other types of trading strategies, however, there is a great deal of secrecy, and information diffusion about successful strategies is slow at best. In this case, one of the principal factors driving information diffusion is the migration of employees from firm to firm. Since employees typically spend at least several years at a given firm -- it usually takes this long just to learn the business properly -- once again, the timescale is measured in years to decades. Independent discovery is obviously more difficult to evaluate; my own impression based on anecdotal evidence, is that to discover a nontrivial strategy that is not already known requires either a large scale effort consuming many years, or a great deal of luck.

Another factor that should be noted is the time required to verify the profitability of a strategy at a statistically significant confidence level. One way to evaluate this is in terms of the return/risk ratio. This is often computed based on the ratio of the annual return to the annual standard deviation of returns, called the Sharpe ratio. Assuming the gains are normally distributed, consider a strategy whose true Sharpe ratio is S . The expected statistical significance after trading for t years is $S\sqrt{t}$. A Sharpe ratio of one is generally con-

sidered quite good. In this case four years is typical to achieve a statistically significant track record. (This is the rational basis behind the five year rule of thumb mentioned earlier). The task of selecting between different strategies makes this much worse, because of the “millionth monkey” effect, i.e. the odds that one of the strategies will perform well purely at random.

For all of the factors discussed, the arguments above suggest that in any given niche the evolution to market efficiency typically requires years to decades. It is hard to imagine how it can take place in less than a year, and 5-10 is more likely. I believe that if someone were to study the discovery of the Black-Scholes option pricing formula and the evolution of the profit margins through time, it would provide a good illustration that these timescales are roughly correct.

4.7 Evolution toward higher complexity?

In the twentieth century it is evident that markets have become more complex. This is true of the number of assets, the number of transactions, the timescale on which they operate, and the sophistication of the strategies used for trading. It is a challenge to understand this from a theoretical point of view.

Strategies can make profits by anticipating other strategies. Thus one expects that as successive strategies are added the standards go up -- strategies have to be better and better just to stay even. This suggests that complexity also goes up. One scenario through which this can occur is suggested by equation 11. If the set of strategies in the market and the capital allocated to each of them can be estimated, it is possible to forecast the expected price. A trading strategy can be constructed using dynamical programming [27]. If the market indeed follows equation 11, and the strategies are indeed accurately known, this is an optimal strategy. Since it involves simulating all other strategies in the market, the algorithmic complexity of this strategy is equal to that of all other strategies in the market combined. Once such a strategy enters the market and adjusts its capital to optimize profits, it creates an opportunity for a newer strategy, which knows about all other strategies, including the new strategy. In this succession each strategy is more complex than all previous strategies.

The scenario above is unrealistic -- given the secrecy of traders, no one knows all their strategies exactly. However, it may occur in an approximate sense. For a strategy to make profits it must have some ability to anticipate the order flow of other strategies. It must therefore contain some of the complexity of the other strategies in the market. It will be interesting to study this in simulations that incorporate the generation of new and more complex strategies.

4.8 Summary and discussion

The evolution of markets can be thought of in terms of adjustments in the capital of strategies, which is analogous to the population of biological species. By separating timescales this can be modeled by equations that are analogous to the Lotka-Volterra equations.

Although it is difficult to model the dynamics of the capital, which ultimately involve human decision making, under the assumption that people are somewhat intelligent and are motivated to make profits it is possible to make an approximate model. This will be explored further in a future paper [31].

One of the main points of this section is that diversity is a natural outcome of the drive of individual traders to make profits. This is demonstrated through a variety of worked examples. Diversity comes about because market inefficiencies are multi-faceted, and there is a diversity of possible strategies that can exploit any given market inefficiency. Technical strategies are a good example. The generation of diversity deserves further study, e.g. in simulations that do not suffer from the limitations of the analytic calculations presented here.

While this paper does not provide a final answer concerning market efficiency, it does suggest a dynamical context in which to address the question. A preview of what is possible is evident in the calculation of the evolution of an isolated pattern. This can be extended to a continuous setting. A variety of different order of magnitude arguments suggest that the timescale for market efficiency is measured in years or decades. If new patterns are generated on an ongoing basis, this suggests that there is time to exploit them before they disappear. As in biological evolution, fitness and survival are moving targets.

5. Conclusions

The premise explored here is that impatient trading impacts the price, and price formation can be understood as the aggregate of the impacts of all the trading in the market. The main result of the first section is that, based on plausible simplifying assumptions, it is possible to derive a unique market impact function. This can be viewed as a nonequilibrium generalization of supply and demand. Alternatively, it comes from the fundamental assumption of market friction. Market friction is shown to be path dependent. Price formation can be modeled as a dynamical system that depends on trading strategies. Within this framework it is natural to regard the market as a continuous game with a continuous payoff matrix. The goal is to make profits by anticipating the moves of other players. The market can be viewed as a casino in which the market maker plays the role of the house. I believe this framework is a quantitative expression of the mental model that many traders use to think about markets.

The development here only treats market orders, allowing only one level of patience; it would be very interesting to extend this model to include other types of orders, such as limit orders, which allow different levels of patience. In addition, it would be interesting to extend the theory by generalizing the market impact function to take the market maker's state properly into account.

This approach can be contrasted to other recent efforts, such as the Santa Fe Stock Market [4, 5, 15]. The focus on market dynamics and the understanding of the role of different strategies are similar, but the underlying assumptions are quite different. The SFI Stock Market assumes a common utility function for all investors, and assumes that an auction occurs at every timestep. This sometimes causes problems. For example, when the market doesn't clear because of a lack of buyers or sellers, they have to shut down the market and update the strategies. The approach proposed here, in contrast, guarantees an orderly market as long as the dynamics are stable. It is also much simpler and offers a clear game theoretic context in which the interactions between strategies are easy to understand. Further study is needed to determine to what extent these two approaches coincide, and which is more realistic.

The section on ecology explores the consequences of the postulated dynamics from an empirical point of view. Since there is no assumption of equilibrium, it is not obvious *a priori* that the price dynamics are sensible. Several basic strategies are investigated, first one at a time and then in combination. These are chosen because they are simple examples with varying levels of realism. On its own each of these strategies induces characteristic dynamics in the price. Two typical classes are value investing strategies, which usually induce negative correlations, and trend following strategies, which induce positive correlations. Conventional wisdom says that trading strategies cause self-fulfilling prophecies. For many value investing strategies this is not the case. While some value investing strategies support cointegration of price and value, many popular ones do not. In contrast, it appears that trend following strategies always tend to create trends. However, there can also be side-effects, such as price oscillations. This is due to the fact that the market impact is caused by orders, but the trading strategy is formulated in terms of positions.

It is a truism that markets reflect the consensus view. However, depending on the non-linearity of the strategies, the dynamics under a plurality of different views may be quite different than the dynamics with a single view. When all the trading strategies depend linearly on the logarithm of value the market behaves just as it would for a single strategy based on the mean. But when the strategies depend nonlinearly on the logarithm of value, simulations show that the situation is quite different. The additional trading generated by disagreements about value leads to excess volatility in the price. If we view the market as a machine that has the function of making prices track value, unless all participants have the same perception of value, the market performs this function inefficiently. The market reflects the consensus view, but it behaves differently than it would under a single view.

The worked examples and simulation results demonstrate that whether or not the price behaves sensibly depends on the trading strategies in the market. While each strategy produces stable price dynamics for a range of parameter values, if the ratio of capital to liquidity exceeds a threshold, in every example studied the market becomes unstable. Fortunately, at the extremes there are incentives to prevent this. If the capital of a strategy becomes too large excessive transaction costs make it unprofitable. Liquidity is driven up by competition and the dependence of market making profits on trading volume. Liquidity and capital co-evolve. Nonetheless, it remains an open question whether this guarantees market stability. Competition between market makers within this model deserves further study.

This model suggests that the diversification of strategies is natural in financial markets. Diversification is driven by the quest to make profits. Patterns in order flow can be exploited by the creation of new strategies. Through many worked examples we show that there are many situations in which a new strategy can invade pre-existing strategies; the suggestion is that diversity tends to increase. Technical components of the strategies play an important part. As a new strategy invades and increases its capital, its profits tend to rise to a maximum and then decline as it makes losses due to market friction. Under certain assumptions the dynamics of capital can be modeled in terms of equations similar to the Lotka-Volterra equations of population biology. Changes in capital are driven by variations in price, but occur on much longer timescales; from a short term point of view this may cause apparent nonstationarities. There are many questions about the generation of diversity and the evolutionary dynamics of the capital and liquidity that deserve further investigation, such as whether or not this model supports evolutionarily stable strategies [26]. The importance of diversity has been greatly emphasized in biology; it deserves more study in economics.

By placing an emphasis on the interrelationships between strategies, this approach fosters the view of a market as a financial ecology¹. The environment for each trading strategy consists of the market maker and the other strategies. The success or failure of a given strategy depends on the collection of other strategies in the market. The strategies studied

1. The ecological view may give insight into regulatory policy; for example, there have recently been proposals to limit speculation. There is a sense in which speculators play a role in financial ecologies that is analogous to that of carnivores. Suppressing them might have dangerous and unintended effects, just as it has had in biological ecosystems.

here form only a tiny subsample of those in real markets. More work is needed to thoroughly classify real strategies, and to investigate the interaction of strategies using a richer palette than that studied here. The strategies studied here are all simple enough that their deterministic dynamics are also simple; a more complex set of strategies may display richer dynamics, such as chaotic behavior.

In biology there has been debate as to whether the proper level of selection is the genome or the individual organism. In the context of this model, we see that most questions are naturally answered by thinking in terms of strategies (analogous to the genome level). However, as demonstrated in Section 4.5, for some purposes we must argue at the level of individual agents (which are analogous to organisms). Both levels of selection are useful depending on the context.

Even though the trading strategies studied here are simple, commonly observed market phenomena such as time correlations in volume and volatility and long tails seem difficult to avoid. The explanation is natural: Large price movements generate more than average trading, which tends to generate additional large price movements. Long tails come about because high volatility is episodic. This deserves further research; with this approach it should be possible to get a more quantitative understanding of these phenomena. By studying more complex strategies that better incorporate the human emotions of fear and greed, interesting behavior should appear.

While this paper gives no final answer concerning market efficiency, it does suggest a method to address this question in a dynamical context. The rough calculation of the evolution of an isolated pattern in Section 4.4 gives a preview of what is possible. This can be extended to more general patterns, taking the change in the information set properly into account. These computations suggest that market efficiency is inherently similar to the increase of entropy. The drive to efficiency occurs through an increase in diversity, which almost by definition involves an increase in complexity. Thus self-organization and the second law are interwoven in much the same way they are in nature. Given its simplicity, this model may eventually help to clarify this in a more general context.

A variety of different order of magnitude arguments as discussed in Section 4.6 suggest that the timescale for market efficiency is measured in years to decades. If new patterns are generated on an ongoing basis, this should allow time to exploit them on the road to efficiency. Efficiency is seen as a question that is naturally discussed in evolutionary terms.

This paper presents a new approach to understanding the dynamics of financial markets. It only begins to explore the consequences. It has the desirable features of being simple, experimentally testable, and extensible. The foundation of the theory is the market impact function, which can be measured directly based on appropriate data. The theory must be true on some level. That is, there is clearly market impact, it is clearly an increasing function of order size, and it is clear that it is felt in the price. The important questions are the precise form of the market impact and the magnitude of its role in price formation. I have made a rather specific proposal for the form of this function, but even if this turns out to be wrong, it is straightforward to revisit all the results presented here, at least numeri-

cally, with any empirically measured function. Once this foundation is clarified, by studying the trading strategies used in real markets and the process of generating them and allocating capital, it should be possible to understand the ecology and evolution of markets in quantitative detail.

Acknowledgments

Shareen Joshi's help with simulations in the final months of this project, and her support by the Santa Fe Institute, have been invaluable. I would also like to thank the employees of Prediction Company, and in particular Norman Packard, for creating the context that drove the intellectual development of this theory, and for the mini-sabbatical during which this paper was completed. In addition I would like to thank Seth Lloyd and Dan Freedman for valuable discussions.

References

1. Alfred Marshall, *Principles of Economics*, eighth edition (1920) Porcupine Press, Philadelphia.
2. George Soros, *The Alchemy of Finance* (1987) John Wiley and Sons, New York.
3. John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay, *The Econometrics of Financial Markets* (1997) Princeton University Press.
4. W.Brian Arthur, John .H. Holland, Blake LeBaron, Richard Palmer, and P. Tayler, "Asset pricing under endogenous expectations in an artificial stock market", in W. B. Arthur, D. Lane, and S.N. Durlauf (eds.) *The Economy as an Evolving, Complex Sysetm II* (1997) Addison-Wesley.
5. Blake LeBaron, W. Brian Arthur, and Richard Palmer, "Time series properties of an artificial stock market" (1998) Santa Fe Institute working paper.
6. J.Bradford DeLong Andrei Schleifer, Lawrence H. Summers, and Robert J. Waldmann, "Positive feedback and destabilizing rational speculation, *Journal of Finance*, 45 (1990) 379-395.
7. Robert J. Shiller, *Market Volatility* (1997) MIT Press.
8. Maureen O'Hara, *Market Microstructure Theory* (1995) Blackwell Publishers.
9. Louis K.C. Chan and Josef Lakonishok, "Institutional trades and intraday stock price behavior", *Journal of Financial Economics* 33 (1993) 173-199.
10. Louis K.C. Chan and Josef Lakonishok, "The behavior of stock prices around institutional trades", *Journal of Finance*, 50 (1995) 1147-1174.
11. Jerry A. Hausman and Andrew W. Lo, "An ordered probit analysis of transaction stock prices", *Journal of Financial Economics*, 31 (1992) 319-379.
12. Werner F.M. deBondt and Richard H. Thaler, "Financial decision making in markets and firms: A behavioral perspective", in R. Jarrow et al. (eds.) *Handbooks in OR & MS*, , Vol. 9 (1995) Elsevier Science.
13. Donald B. Keim and Ananth Madhavan, "Anatomy of the trading process: Empirical evidence on the behavior of institutional traders", *Journal of Financial Economics* 37 (1995) 371-398.
14. Donald B. Keim and Ananth Madhavan, "The upstairs market for large block transactions: Analysis and measurement of price effects", *Review of Financial Studies* 9 (1999) 1-37..
15. Shareen Joshi, Jeffrey Parker, and Mark A. Bedau, "Technical trading creates a prisoners dilemma: results from an agent-based model", Santa Fe Institute working paper.
16. Hans Stoll, "The supply of dealer services in securities markets", *Journal of Finance* 33 (1978) 1133 - 1151.
17. Hans R. Stoll, "Inferring the components of the bid-ask spread: Theory and empirical tests", *Journal of Finance* 44 (1989) 115-134.

18. Thomas Ho and Hans Stoll, "The dynamics of dealer markets under competition", *Journal of Finance* 38 (1983) 1053-1074.
19. Roger D. Huang and Hans R. Stoll, "Market microstructure and stock return predictions", *Review of Financial Studies* 7 (1994) 179-213.
20. R. Engle and C. Granger, "Cointegration and error correction: Representation, estimation, and testing", *Econometrica* 55 (1987) 251-276.
21. *Encyclopedia of Technical Analysis*.
22. Eugene Fama, "Efficient capital markets: A review of theory and empirical work", *Journal of Finance* 25 (1970) 383 - 417.
23. B. Malkiel, "Efficient Market Hypothesis", in *New Palgrave Dictionary of Money and Finance*, P. Newman, M. Milgate and J. Eatwell, MacMillan.
24. Paul Cootner (ed.), *The Random Character of Stock Market Prices* (1964) MIT Press, Cambridge.
25. William F. Sharpe, Gordon J. Alexander, and Jeffery W. Bailey, *Investments* (1995) Prentice Hall, p. 106.
26. Joseph Hofbauer and Karl Sigmund, *Evolutionary Games and Replicator Dynamics* (1998) Cambridge University Press.
27. D. Bertismas and Andrew Lo, "Optimal control of execution costs", *Journal of Financial Markets* 1 (1998) 1- 50.
28. J.P. Bouchaud and R. Cont, "A Langevin approach to stock market fluctuations and crashes", cond-mat/9801279.
29. G. Caldarelli, M. Marsili, and Y.C. Zhang, "A prototype model of stock exchange", *Europhysics Letters* 40 (1997) 479.
30. Lawrence H. Summers, "Does the stock market rationally reflect fundamental values", *Journal of Finance* 41 (1986) 591 - 600.
31. J. Doyne Farmer and Shareen Joshi, work in progress.
32. R. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation", *Econometrica* 50 (1982) 987 - 1008.
33. H. Demsetz, "The cost of transacting", *Quarterly journal of economics* 82 (1968) 33-53.
34. L. Glosten and P. Milgrom, "Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics* 13 (1985) 71-100.
35. Oliver Hansch, Narayan Y. Naik, and S. Viswanathan, "Do inventories matter in dealership markets: Evidence from the London stock exchange", *Journal of Finance* 53 (1998) 1623 - 1656.
36. J. Campbell and R. Shiller, "The dividend-price ratio and expectations of future dividends and discount factors", *Review of Financial Studies* 1 (1988) 195-227.
37. Reference on cointegration between purchasing power parity and currency prices.

38. Matthias Ruth, “Evolutionary economics at the crossroads of biology and physics”,
Journal of Social and Evolutionary Systems 19 (1996) 125 - 144.