Abstract
In this thesis, it will be investigated whether a model of a single financial asset that combines two approaches from agent based modeling reproduces some of the stylized facts observed in financial markets. More specifically, a dynamical stochastic model is constructed that explicitly takes into account social interactions and evolutionary learning. A special case of the model and the model’s equilibrium distribution are studied analytically, while it is estimated on stock market data and investigated using simulations. It is found that the model exhibits most of the stylized facts. In addition, a tentative estimation of the model shows that more advanced methods are needed for a reliable result.
Contents

1 Introduction 2

2 Literature 3

3 Two modeling frameworks 5
   3.1 Generalized Ant Process ........................................... 5
      3.1.1 FP equation ................................................. 7
      3.1.2 Formal derivation ............................................ 10
      3.1.3 SDE ....................................................... 11
   3.2 BH model .......................................................... 12
      3.2.1 Market equation .............................................. 13
      3.2.2 Evolutionary component ...................................... 14

4 The model 15
   4.1 A special case ...................................................... 15
      4.1.1 Equilibrium distribution ...................................... 16
      4.1.2 SDE ........................................................ 16
      4.1.3 Moments and stylized facts ................................... 16
   4.2 General model ....................................................... 22
      4.2.1 Continuous dynamics .......................................... 23
   4.3 Discrete dynamics ................................................... 25
   4.4 Equilibrium distribution .......................................... 27

5 Estimation 32
   5.1 MLE ............................................................... 33
   5.2 Data .............................................................. 35
   5.3 Results ............................................................ 36
   5.4 Estimation-based simulation ....................................... 38

6 Conclusion 40
1 Introduction

Financial markets have been the subject of research for many decades. In a sense they are the ultimate example of a complex system: a system that consists of many interacting agents and shows aggregate behavior that is qualitatively different from and cannot be explained solely by the individual behavior of its agents. In addition, financial markets might be viewed as adaptive since they fundamentally change over time; the volatility smile in option prices, for instance, was only observed after the crash of 1987.

Over the years, stock price movements have been modeled in different ways, both within and outside mainstream economics. Much work has been done in the fields of modern portfolio theory and the Black-Scholes theory of derivative valuation, the cornerstones of conventional asset pricing. In addition, Benoit Mandelbrot, an outsider to economics, observed early that changes in asset prices did not follow a normal distribution, in contrast with assumptions sometimes made in portfolio theory [Mandelbrot, 1963b]. More specifically, the following stylized facts have been observed in financial markets: (1) little autocorrelations in returns, (2) fat tails in the return distribution, (3) volatility clustering (i.e. autocorrelation in absolute returns) and (4) bubbles and crashes [Bouchaud, 2002]. Note that the last feature is about the asymmetry between increases and decreases of the stock price.

In this thesis, it will be investigated whether a model of a single financial asset that combines two approaches from agent based modeling (namely a heterogeneous agent model and a stochastic population model) reproduces the aforementioned stylized facts well. From the perspective of the recent financial crisis it is interesting to study in what way social interactions have an impact on asset prices, since people are more connected today through social networks and other technology (increasing the likelihood of herding behavior). In addition, having a better understanding of what causes these stylized facts is important for policy makers since they can exert considerable influence through the regulation of markets. Finally, obtaining a stochastic process for the evolution of asset prices might be useful theoretically for the rational pricing of options beyond Geometric Brownian Motion (GBM).

More specifically, a model is constructed and analyzed that explicitly takes into account social interactions and evolutionary learning, but is difficult to study analytically. A simple model is obtained as a special case, which can be studied analytically, and the general model will be studied in an equilibrium setting. Secondly, the general model will be estimated using stock market data, which can then be used for out-of-sample predictions. Finally, incorporating the estimated parameters, a simulation study is performed to investigate the stylized facts numerically.

The rest of the thesis is outlined as follows. In Section 2 the literature on rationality and agent based models in financial markets will be reviewed. Then, Section 3 will introduce and go into the details of the two modeling frameworks; new theoretical results are obtained with respect to the stochastic model in Subsection 3.2. The general model, as well as its analysis, is discussed...
in Section 4. In all subsections the results are new as far as we know, expect for a part of the results on the moments and autocorrelations in Subsection 4.1. In Section 5, the data will be discussed and the model will be estimated, after which simulation results follow. Finally, Section 6 concludes.

2 Literature

The model framework used in this thesis tries to integrate (at least) two modeling paradigms. Because of this, it is useful to give an overview of these models. Moreover, since the role of expectation formation and learning is important in many of these models, the origin of these concepts within economics is discussed first.

The discussion on rationality goes back as far as an exchange of letters between Walras and Poincaré [Wagener, 2014]. In the late 1950’s and early 1960’s both Simon’s concept of Bounded Rationality and Muth’s idea of rational expectations originated from joint work [Holt et al., 1960] on inventory problems. According to Muth, if agents’ expectations are only weakly correlated, their individual variation will average out in the aggregate by the law of large numbers [Muth, 1961]. Bounded rationality is the observation that agents have limited computational abilities due to lack of time, lack of intelligence, lack of information and so forth [Simon, 1972]. Muth’s idea of rationality was interpreted differently by Lucas [Lucas and Prescott, 1971], who popularized the idea, turning it into an important tool for economic modeling. This reinterpreted idea of rational expectations is that, in a stochastic dynamic general equilibrium model framework, agents’ forecasts of the distribution of economic quantities are equal to the true distributions of the quantities they are predicting, i.e. their predictions are not biased. Note that this is a statement about individual, rather than collective, behavior.

Indeed, expectation formation is important in economic dynamic models. Rational expectations (Lucas’ version) are useful not only since it uniquely characterizes agent’s inner workings, but also because agents’ predictions are, by construction, automatically consistent with the model. In reality, however, agents do not always conform to this type of behavior. In some experiments, subjects learn to forecast in accordance with rational expectations, while in other situations this does not happen [Wagener, 2014]. Clearly, different economic situations warrant different behavior.

In recent decades the rational expectations hypothesis has been challenged by roughly two streams of criticisms, behavioral economics and complexity economics. The two differ in their distance from the mainstream paradigm (i.e. rational expectations); the latter is more extreme than the first. Complexity economics focuses as much on the interaction between agents as on the agents themselves. A possible reason for this shifted focus is the following.

Comparing economics to the natural sciences, one observes that there are similar simplifications. These have to do with the physical reality economic agents operate in, for example the assumption of no transaction costs, only one risky/one risk-free asset and homogeneous markets.
These assumptions, which might be called physical assumptions, should be robust to small perturbations, e.g. if transaction costs are non-zero but very small, then the predictions should also change only slightly.

Since economics is a social science, dealing with human behavior, there are also assumptions related to the way economic agents think about the economic system they operate in, which determines how they act. When contemplating an economic system it is tempting to model agents as if they were operating in the environment generated by the physical assumptions. However, doing so means that this second set of assumptions is correlated to the physical assumptions and might, therefore, be biased. For example, it is much more reasonable to assume that an agent has rational expectations in an environment with one stock than in an environment with a thousand stocks.

The complexity economics approach motivates the use of simple behavioral rules, also called heuristics, to model economic agents. Simple rules may not be unique, which introduces an element of heterogeneity into models: different agents might use different rules or the same agent might use different rules in different circumstances. The first of these models date back to the seventies [Zeeman, 1974] and studies the dynamics of fundamentalists (whose prediction is based on fundamentals) and chartists (whose prediction is based on previous prices) interacting in a financial market. In the late eighties and early nineties similar models were applied to the foreign exchange market, including Frankel and Froot [Frankel and Froot, 1986, Frankel and Froot, 1990a, Frankel and Froot, 1990b] and Kirman [Kirman, 1991] (where the relative proportion of fundamentalists and chartists is determined by a Markov chain).

Inspired by these models and the aforementioned observed stylized facts in financial markets, several heterogeneous agent models have been proposed. Examples of these are [Arthur et al., 1997b] and [LeBaron et al., 1999] (in a computational setting), in addition to [Lux, 1995] and [Brock and Hommes, 1998] (simple, analytically tractable models). Specifically, the Brock-Hommes (BH) model assumes that there are multiple types of agents that have different future beliefs on prices. The evolution of the fraction of agents in each category is then determined by how well their forecasts (strategies) perform, compared to others, based on the payoff. More recently, [Diks and Van der Weide, 2003, Diks and Van der Weide, 2005] introduced so-called continuous beliefs systems, which combine some of the advantages of the computational and analytical approaches.

Secondly, outside of economics, econophysics is an interdisciplinary field that tries to explain stylized facts, using agent based models among other methods. One interesting example is Cont & Bouchaud [Cont and Bouchaud, 2000]. In this study, the benchmark is the case where all traders buy and sell stocks independently of each other. A consequence of the central limit theorem is then that returns on the stock will converge to a Gaussian. However, the assumption that choices of agents to buy or sell stocks are independent is too strong (in reality agents frequently interact, directly or indirectly). Violations of the independence assumption may lead to herding behavior,
where agents influence each others' decisions. This type of behavior is also investigated in [Cont and Bouchaud, 2000] using network theory, resulting in fat tails in the return distribution.

Finally, closer to the field of probability theory, there is the work of [Föllmer et al., 2005], which also studies markets with different types of agents, but in the framework of stochastic processes. In earlier work Föllmer [Föllmer and Schweizer, 1993, Föllmer, 1994] discussed the convergence of a discrete time stochastic process in the presence of a random environment (possibly generated by the interaction of different agents) to a continuous time process. Interestingly, this procedure was not applied to the later model [Föllmer et al., 2005].

Summarizing, in these models the decisions of agents to buy and sell stocks become correlated, rendering the Muthian rational expectations assumption inapplicable. Such behavior can become correlated in different ways: by social interaction (i.e. recruitment behavior), evolutionary pressure on forecasting rules or implicitly by background stochastic processes. Since the econophysics literature is generally not concerned with expectation formation, while the heterogeneous agents literature does not focus on social interactions (a notable exception is [Brock and Durlauf, 1997]) it might therefore be interesting to look at a model which combines both ideas. Finally, endogenous uncertainty is an important feature of asset prices, so it may be more realistic to model these by a stochastic process. Kirman [Kirman, 1991] does this, but in his study the population dynamics are independent of the price process, which seems unrealistic; also, he does not consider a framework with a continuous state space.

3 Two modeling frameworks

In order to accommodate social interaction as in Kirman [Kirman, 1991] using Markov chains, in addition to evolutionary updating as in Brock and Hommes [Brock and Hommes, 1998], the basic framework for the model will be Stochastic Evolutionary Game Theory (SEGT), where payoffs change over time due to market dynamics. More specifically, a Markov chain will be constructed based on the interactions between traders and their expectations, after which a Stochastic Differential Equation (SDE) is obtained in the limit where the number of traders goes to infinity and the time steps go to zero. The market price (and therefore the payoff) is determined by supply and demand at each time point. Together, these combine to form a 2D system.

In the following, these two building blocks will be explained. First, the Ant Process is discussed as well as the mathematical machinery behind it. Then, the BH model is presented, in addition to the underlying economic theory.

3.1 Generalized Ant Process

Alan Kirman [Kirman, 1993] used the Ant Process, related to Polya’s urn model, originally to explain the foraging behavior of ants. In this subsection a more general version is considered. The
model considers a population of \( N \) agents, each of which can be of type 1 or type 2. The number of individuals of type 1, \( k \), changes over time according to a time-homogeneous Markov chain with state variable \( K_t \in S = \{0, 1, \ldots, N\} \) (the number of individuals of type 2 is then \( N - K_t \)).

It is assumed that at each time step the following occurs. First, one individual is selected randomly from the population. Then, the individual changes his type independently with probability \( \epsilon e_+ + (X_t, X_{t-1}) \) if it is of type 2 and \( \epsilon e_- + (X_t, X_{t-1}) \) if it is of type 1, where \( 0 < \epsilon \ll 1 \), \( e_+ \) and \( e_- \) are functions that take values in \( I := [0, 1] \) and \( X_t \) is a dynamic stochastic variable, which is determined by another equation, depending on \( K_t \). In this thesis, it will be of the form

\[
X_t = H(X_{t-1}, W_t),
\]

where \( W_t = K_t/N \) is the fraction of the population of type 1 at time \( t \) and \( H \) is differentiable in both arguments. (In the original Ant Process the self-conversion probability was simply \( \epsilon \) for both types and there was no additional variable \( X_t \).) Since the self-conversion probabilities depend on \( X_t \), these probabilities are formally conditional probabilities. However, since \( X_t \) is itself determined by \( K_t \), it is not of practical interest to treat them as such.

If no self-conversion takes place, the individual meets another ant, which is randomly picked from the remaining population. The first individual is converted to the type of the second one with probability \( 1 - \delta \). The Ant Process, therefore, models social interaction and it is clear that when a large portion of the population is of one type, it is likely that next time step this portion will be even larger, leading to herding behavior. Note that it is a one-step process, i.e. at every time step the population of type 1 (2) can either increase/decrease by one or stay the same. From this description, it is clear that the conversion and self-conversion probabilities only depend on the current state of the system. In addition, note that since (1) can be inserted in \( e_+ \) and \( e_- \), the vector \((W_{t+1}, X_t)\) only depends on \((W_t, X_{t-1})\), hence the 2D system is Markov as well. Therefore, we will write, slightly abusing notation, \( e_+(W_t, X_{t-1}) \) when we mean \( e_+(H(X_{t-1}, W_t), X_{t-1}) \) (and similarly for \( e_- \)).

First, note that at each time step the probability of a change only directly relates to \( W_t \); the stochasticity only affects \( X_t \) indirectly through (1). Since the variable \( X_t \) is trivial from a stochastic point of view, the fact that the system under investigation is 2D will be mostly ignored in this subsection. Now, let \((K_t)_{t \geq 0} \) be the \( S \)-valued Markov chain and let \((\Omega, \mathcal{F}, P)\) be the underlying probability space. A natural candidate for this space is \( \Omega = S^{\{0,1,2,\ldots\}} \), which consists of all paths \( \omega \), i.e. \( \omega_t \in S \ \forall t \). The random variables that make up the chain are then given by the projections \( K_t : \Omega \to S \) defined by \( K_t(\omega) = \omega_t \). The \( \sigma \)-algebra is defined by the smallest collection of sets in \( \Omega \) such that each projection is measurable, i.e. \( \mathcal{F} = \sigma(K_t, t \geq 0) = 2^{\Omega} \). Finally, the probability measure associated to the Markov chain is induced by the tridiagonal time-dependent transition matrix (which easily follows from the dynamic rules introduced above)
\[ T(k+1|k) := P(K_{t+1} = k+1|K_t = k) = \left( 1 - \frac{k}{N} \right) \left( \epsilon e_+(k/N, x_{t-1}) + (1 - \delta) \frac{k}{N-1} \right) \quad (2) \]

\[ T(k-1|k) := P(K_{t+1} = k-1|K_t = k) = \frac{k}{N} \left( \epsilon e_-(k/N, x_{t-1}) + (1 - \delta) \frac{N-k}{N-1} \right) \quad (3) \]

\[ T(k|k) := P(K_{t+1} = k|K_t = k) = 1 - T(k+1|k) - T(k-1|k) \quad (4) \]

Note that it is required that

\[ T(k+1|k) + T(k-1|k) \leq 1 \iff \epsilon + 2(1 - \delta) \frac{k}{N} \frac{N-k}{N-1} \leq 1 \iff \epsilon \leq \delta, \]

since otherwise these expressions do not define probabilities. (Clearly, \( N \geq 2 \); otherwise, the process is ill-defined.)

### 3.1.1 FP equation

The primary quantity of interest is \( P(k,t) \), the probability to be in state \( k \) at time \( t \). An expression for this quantity is \( (T^t P(0))_k \), where \( T \) is the transition matrix and \( P(0) = (P(0,0), P(1,0), ..., P(N,0)) \) is the vector of probabilities at time \( t = 0 \). However, this solution is difficult to deal with analytically. In addition, for later purposes the state space needs to be continuous. Therefore, the most straightforward way to proceed is to approximate the Markov chain by a process that has a continuous state space and as a side effect will have to be continuous in time as well. Therefore, in the following the time steps of the original Markov chain go to zero and the state variable becomes continuous, i.e. \( N \) goes to infinity. In this way the Fokker-Planck (FP) equation, also known as the Kolmogorov forward equation, is obtained from the stochastic process defined by (2), (3) and (4). Initially, a heuristic derivation is given to provide intuition and obtain a tentative result. Afterwards, this result is proven to be true by rigorous arguments.

As a preliminary, assume that the time-dependent transition probabilities have the following form

\[ P(k+1|k; \Delta t) = T(k+1|k)\Delta t + o(\Delta t) \]

\[ P(k-1|k; \Delta t) = T(k-1|k)\Delta t + o(\Delta t) \]

\[ P(k|k; \Delta t) = 1 - P(k-1|k; \Delta t) - P(k+1|k; \Delta t) \]

where \( \Delta t \) is the size of each time step, \( o(\Delta t) \) means that \( o(\Delta t)/\Delta t \) tends to zero as \( \Delta t \to 0 \) and the other probabilities are zero. Now,
\[ P(k, t + \Delta t) = P(k|k - 1; \Delta t)P(k - 1, t) + P(k|k + 1; \Delta t)P(k + 1, t) + P(k|k; \Delta t)P(k, t) \]

\[ = \sum_{k'} P(k'|k; \Delta t)P(k', t), \]

where \( k' \) in this case runs over \( \{k - 1, k, k + 1\} \), but in general could run over the entire state space.

Now, consider the general case that \( K_t \) is a continuous random variable (i.e. in the limit \( N \to \infty \) and \( w = k/N \)) and assume that all states can be reached in time step \( \Delta t \). In this case all probabilities are functions of \( w \) and the sum can be replaced by an integral to get

\[ P(w, t + \Delta t) = \int P(w|w'; \Delta t)P(w', t)dw'. \]

Assuming a Taylor series around \( w' = w \) exists, we can write the integrand as

\[ P(w + \Delta w, t) = P(w, t + \Delta t|w', t)P(w', t) \]

\[ = \sum_{l=0}^{\infty} \frac{(\Delta w)^l}{l!} \left[ \frac{\partial^l}{\partial w^l} (P(w + \Delta w|w'; \Delta t)P(w', t)) \right] |_{w'=w} = \]

\[ \sum_{l=0}^{\infty} \frac{(\Delta w)^l}{l!} \frac{\partial^l}{\partial w^l} (P(w + \Delta w|w; \Delta t)P(w, t)) \]

where \( \Delta w = w - w' \). Integrating this expression with respect to \( w' \) (or, equivalently, \( \Delta w \)) then gives

\[ P(w, t + \Delta t) = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \int (\Delta w)^l \frac{\partial^l}{\partial w^l} (P(w + \Delta w|w; \Delta t)P(w, t))d\Delta w \]

\[ = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \frac{\partial^l}{\partial w^l} (M_l(w, \Delta t)P(w, t)) \]

where we have used that under the integral sign \( \Delta w \) is independent of \( w \) (i.e. the partial derivatives can be moved out of the integral) and \( M_l(w, \Delta t) \) is the \( l \)-th jump moment, defined by

\[ \int (w' - w)^l P(w'|w; \Delta t)dw' = E[(W_{t+\Delta t} - w)^l|W_t = w]. \]

This is the Kramers-Moyal (KM) expansion. Now, the only quantities in this equation that are characteristic of a particular Markov process, are the jump moments. Furthermore, the jump moments can also be calculated for discrete processes after which \( N \to \infty \) so that the jump moments are obtained for the case that \( k \) is continuous. The resulting KM expansion is then associated to the limit stochastic process of the Markov chain.

As the jump moments are derived from the transition probabilities \( P(k'|k; \Delta t) \) they can be written as \( D^{(l)}(k)\Delta t + o(\Delta t) \), where \( D^{(l)}(k) \) is some function of \( k \), which has to be determined.
In the case of the Ant Process there are only three terms in the jump moments since \(k' \in \{k − 1, k, k + 1\}\), one of which vanishes since \(k − k = 0\). The jump moments are then given by

\[
M_i(k, \Delta t) = (T(k + 1|k) + (-1)^iT(k - 1|k))\Delta t + o(\Delta t)
\]

If the fraction \(w = k/N\) is the state variable instead, the jump moments are given by \((l > 0)\)

\[
M_i(w, \Delta t) = (T(wN + 1|wN) + (-1)^iT(wN - 1|wN))\frac{1}{N^l\Delta t} + o(\Delta t)
\]

Now, to truncate the KM expansion, the FP approximation comes into play, which demands that time is scaled as \(\tau = t/N^2\), so the jump moments become

\[
M_i(w, \Delta \tau) = (T(wN + 1|wN) + (-1)^iT(wN - 1|wN))\frac{1}{N^{l-2}\Delta \tau} + o(\Delta \tau).
\]

Then put \(\epsilon = \alpha/N\), \(\delta = 2\alpha/N\), where \(\alpha > 0\), (this setup ensures that \(\delta > \epsilon\)) and let \(N \to \infty\); all terms with order higher than two will go to zero and the first and second jump moments become

\[
M_1(w, \Delta \tau) = (\lim_{N \to \infty} N(T(k + 1|k) - T(k - 1|k)))\Delta \tau + o(\Delta \tau)
\]

\[
M_2(w, \Delta \tau) = (\lim_{N \to \infty} (T(k + 1|k) + T(k - 1|k)))\Delta \tau + o(\Delta \tau)
\]

while higher order jump moments are zero. In the limit \(D^{(1)}(k)\) and \(D^{(2)}(k)\) are equal to, respectively:

\[
\lim_{N \to \infty} N(T(k + 1|k) - T(k - 1|k)) = \alpha(1 - w)e_+(w, x_{t-\Delta t}) - \alpha w e_-(w, x_{t-\Delta t})
\]

\[
= \alpha(e_+(w, x_{t-\Delta t}) - (e_+(w, x_{t-\Delta t}) + e_-(w, x_{t-\Delta t}))w)
\]

(5)

\[
\lim_{N \to \infty} (T(k + 1|k) + T(k - 1|k)) = (1 - w)w + w(1 - w) = 2w(1 - w)
\]

(6)

Note that \(M_0(w, \Delta \tau) = \int P(w'|w; \Delta \tau)dw' = 1\) (which continues to hold when \(N \to \infty\)). Therefore, the truncated KM expansion for the Ant Process becomes

\[
P(w, \tau + \Delta \tau) = P(w, \tau) - \frac{\partial}{\partial w}((\alpha(e_+(w, x_{t-\Delta t}) - (e_+(w, x_{t-\Delta t}) + e_-(w, x_{t-\Delta t}))w)\Delta \tau
\]

\[
+ o(\Delta \tau)P(w, \tau)) + \frac{\partial^2}{\partial w^2}((w(1 - w)\Delta \tau + o(\Delta \tau))P(w, \tau))
\]

Now, subtracting \(P(w, t)\) from both sides, dividing by \(\Delta \tau\), letting \(\Delta \tau\) go to zero and replacing \(\tau\) by \(t\) again, the FP equation is obtained

\[
\frac{\partial}{\partial t}P(w, t) = - \frac{\partial}{\partial w}(\alpha(e_+(w, x_t) - (e_+(w, x_t) + e_-(w, x_t))w)P(w, t)) + \frac{\partial^2}{\partial w^2} (w(1 - w)P(w, t))
\]

(7)
3.1.2 Formal derivation

In this subsection, it will be shown that the stochastic process associated to the FP equation in the previous part is indeed the limit process of the Markov chain. First, some notation is introduced. Let $\mathcal{W}_N = K^N / N$, where $K^N := K$ is the Ant Process (the superscript is added to explicitly show dependence on $N$). Then, put $E_N = \{w | w = k/N, k \in S\}$, so $W_N$ takes values in $E_N$. For $w, w' \in E_N$, the transition probability $q_{w,w'}$ then equals $p_{Nw,Nw'}^N$, where $p_{Kt,N} = P(K_{t+1} = k | K_t = k')$ is the original transition probability matrix of the Ant Process. Denote by $P_N$ the associated transition probability function: for $w \in E_N$ and $A \subseteq E_N$, $P_N(w,A) = \sum_{w' \in A} q_{w,w'}$. Then, for $f \in C(I)$ the transition semi-group is defined by $T_N f(w) = \int f(w') P_N(w, dw')$, where the integral is a Lebesgue integral with respect to the measure $P_N(w, dw')$; note the integral reduces to an ordinary sum in this case.

In a similar fashion, the transition semi-group of the limit process, $W$, is defined by $T_t f(w) = \int f(w') P(t, w, dw')$, where $P(t, w, A)$ is the transition probability. The discrete time process becomes a continuous time process as follows. Let $W_N(t) := W_N(\lfloor Nt \rfloor)$ for $t \geq 0$, where $\lfloor \cdot \rfloor$ is the floor function. Note that $W_N$ is a continuous time cadlag (right continuous paths with left limits) stochastic process.

For the theorem below, we wish to introduce the generator of a Markov process. A generator of a Markov process, $A$, is related to the semi-group $T_t$ by $A f = \lim_{t \to 0} (T_t f - f) / t$. In addition, it is related to the FP equation as follows. The FP equation can generally (writing $b(w)$ for the drift term and $a(w)$ for the volatility term) be written as $\dot{P} = -L^* P$, where

$$L^* P(w) = -\frac{\partial}{\partial w} (b(w) P(w)) + \frac{1}{2} \frac{\partial^2}{\partial w^2} (a(w) P(w))$$

is the operator adjoint to the generator. The generator is then given by

$$LP(w) = b(w) \frac{\partial}{\partial w} P(w) + \frac{1}{2} a(w) \frac{\partial^2}{\partial w^2} P(w).$$

The equation $\dot{P} = -LP$ is called the Kolmogorov backward equation.

The following quantities (which are reminiscent of the first two Markov jump processes) will also turn out to be useful for the theorem:

$$a^N(w) = N \int (v - u(w))^2 P^N(u(w), dv)$$

$$b^N(w) = N \int (v - u(w)) P^N(u(w), dv)$$

We now present the proof regarding the limit process of the Ant Process, which is loosely based on a similar proof in [Abundo et al., 1998].

**Theorem 1.** The FP (or Kolmogorov forward) equation associated to the limit process $W$ is given by (7).
Proof. It has to be established that

\[
\lim_{N \to \infty} \sup_{0 \leq t \leq t_0} \sup_{w \in E_N} |T_N^{[Nt]} f(w) - T_t f(w)| = 0
\]

for every \( t_0 \geq 0 \) and \( f \in C(I) \). If this holds and \( W^N(0) \) converges weakly to \( W(0) \), then \( W^N \) converges weakly to \( W \) in the Skorohod space \( \mathcal{D}(\mathbb{R}^+, I) \), the space of cadlag functions. This is, therefore, a necessary condition for \( W \) to be the limit process. According to Theorem 6.5, p.31 from [Ethier and Kurtz, 1986], (8) holds if

\[
\lim_{N \to \infty} \sup_{w \in E_N} |(T^N - 1)f(w) - \mathcal{L}f(w)| = 0
\]

for every \( f \in C^2(I) \), where \( \mathcal{L} \), as before, is the generator of the limit process \( W \). First, we get that

\[
\lim_{N \to \infty} \sup_{w \in E_N} |(T^N - 1)f(w) - \mathcal{L}f(w)| \leq \lim_{N \to \infty} \sup_{w \in E_N} |a^N(w) - a(w) + b^N(w) - b(w)|C
\]

where \( C \) is some positive constant due to the fact that all \( f \), \( f' \) and \( f'' \) are bounded since they are continuous on a compact interval. Using the triangle inequality, the quantity on the right hand side can be bounded by

\[
C \lim_{N \to \infty} \sup_{w \in E_N} [|a^N(w) - a(w)| + |b^N(w) - b(w)|]
\]

Clearly, by (5) and (6), which hold for any \( w \), both terms are zero. This shows that (8) holds. As the FP equation can be derived from the generator, this concludes the proof.

3.1.3 SDE

One may try to solve (7) directly using, for instance, separation of variables or a Fourier transformation since the PDE is linear. Alternatively, the Feynman-Kac formula can be used, which states that this PDE is associated to the following SDE (that is, the solution of the FP equation is a time dependent PDF and the stochastic process associated to this PDF satisfies the SDE)

\[
W_t = W_0 + \int_0^t \alpha(e_+(W_s, X_s) - (e_+(W_s, X_s) + e_-(W_s, X_s))W_s ds + \int_0^t \sqrt{2W_s(1 - W_s)} dB_s,
\]

(9)

where the second integral is a stochastic integral, \( B \) is a Wiener process and \( W_0 \) is a random variable (which can often be taken constant). In the following, whenever an SDE is written down, it is always understood that an initial condition is specified.

The process has several properties. First, note that the process is not a martingale since the drift term is non-zero. Secondly, the process is Markov like the original process, since it is an Itô diffusion. Furthermore, it achieves its highest volatility when \( w = 1/2 \), while having zero volatility at the boundary points \( w = 0 \) and \( w = 1 \). Similarly, the drift term is zero when
\[ W_s = \frac{e_+(W_s, X_s)}{e_+(W_s, X_s) + e_-(W_s, X_s)} \]

and it has a positive (negative) value when \( W_s = 0 \) (\( W_s = 1 \)), by definition of the \( e_+ \) and \( e_- \) functions. Also, note that it is required that \( 0 \leq W_t \leq 1 \); this is because \( w \) is a fraction, but it is also necessary to have a real-valued process, since the root in the volatility term becomes complex otherwise. Because the volatility is zero at the boundary, in continuous time the boundary will never be reached. If, moreover, the process starts at the boundary, it will move to the center, because of the drift term. Finally, in general SDE’s do not have an explicit solution, i.e. a solution in terms of \( t \) and \( B_t \).

For simulation and estimation purposes the SDE has to be discretized. Therefore, it is shown here how this is done in general. Note first that for any SDE that defines a stochastic process \( X \) the drift function \( b(t, X_t) \) and the volatility function \( \sigma(t, X_t) \) have the following infinitesimal representation, given that the volatility term is a square integrable martingale. Let \( (\mathcal{F}_t)_{t \geq 0} \) be the natural filtration of the process. Then we have that \( E[X_{t+h} - X_t | \mathcal{F}_t] = E[\int_t^{t+h} b(s, X_s) ds | \mathcal{F}_t] \). For small \( h \), this can be approximated by \( b(t, X_t)h \). Similarly, the conditional variance of the difference \( X_{t+h} - X_t \) is

\[ E[\left( \int_t^{t+h} \sigma(s, X_s) dB_s \right)^2 | \mathcal{F}_t] = E[\int_t^{t+h} \sigma(s, X_s)^2 dB_s | \mathcal{F}_t]. \]

This, in turn, is roughly equal to \( \sigma(t, X_t)^2 h \) for small \( h \). Based on these observations, it follows that the discretized version of such an SDE is

\[ W_{n+1} = W_n + b(t_n, W_n)h + \sigma(t_n, W_n)\Delta B_n \]

where \( W_n = W(t_n), \ t_{n+1} = t_n + h \) and \( t_0 = 0 \) (so \( t_n = nh \)). Also, \( \Delta B_n = B(t_{n+1}) - B(t_n) \) has a normal distribution with mean zero and variance \( h \). When \( h \to 0 \), the original SDE is obtained.

In Section 4 explicit formulas for \( e_+ \) and \( e_- \) will be introduced, so that more can be said about the SDE.

### 3.2 BH model

Subsection 3.1 dealt with a stochastic process that might be used for modeling the evolution of types. This section will describe how a price actually is determined in a financial market depending on these types. In addition, an alternative evolution mechanism is discussed. Brock & Hommes [Brock and Hommes, 1998] constructed an asset pricing model where in each period the price is obtained from the equilibrium between supply and demand for that asset and a risk free asset. The evolution of types is deterministically determined by how well the strategy associated to each type performs in predicting the price, which is a form of reinforcement learning. First, the market equation (11) is introduced, after which the learning mechanism is taken into account.
3.2.1 Market equation

Following Brock & Hommes [Brock and Hommes, 1998] it is assumed that an economic agent can choose between a risk free and a dividend paying risky asset; \( p_t \) is the risky asset’s price and \( y_t \) is its dividend at time \( t \). Let \( H \) be the number of trader types. Wealth dynamics is then given by

\[
Q_{t+1} = (1 + r)Q_t + \left( p_{t+1} + y_{t+1} - (1 + r)p_t \right)z_t
\]

where \( Q \) is wealth (i.e. the amount of money the trader has), \( r \) is the risk-free return and \( z_t \) is the number of shares bought at time \( t \). Assuming that agents of all types are mean-variance maximizers (which means that they only take into account a linear combination of mean and variance, where the relative weight is determined by \( a \), the risk aversion parameter), each agent type \( h \) wishes to maximize the criterion \( E_{ht}[Q_{t+1}] - \frac{a}{2}\text{Var}_{ht}[Q_{t+1}] \), where \( E_{ht} \) and \( \text{Var}_{ht} \) are the conditional expectation and variance of trader type \( h \) with respect to the filtration generated by the dividend process and the price process respectively. That is, these expectations are induced by the probability measures corresponding to these agents beliefs. These probability measures will not be modeled explicitly, however.

As in Brock & Hommes [Brock and Hommes, 1998] it is assumed that the expected variances are constant \((\sigma^2)\) and equal among trader types. As each trader maximizes the aforementioned criterion function, the demand \( z_{ht} \) for risky assets by trader type \( h \) is

\[
z_{ht} = \frac{E_{ht}[p_{t+1} + y_{t+1} - (1 + r)p_t]}{as^2}.
\]

Assuming that there is no outside supply of shares, these agents comprise the entire market, so aggregate demand must be zero and the market equilibrium equation becomes

\[
\sum_{h=1}^{H} w_{ht} E_{ht}[p_{t+1} + y_{t+1} - (1 + r)p_t] = 0 \iff (1 + r)p_t = \sum_{h=1}^{H} w_{ht} E_{ht}[p_{t+1} + y_{t+1}], \quad (10)
\]

where \( w_{ht} \) is the fraction of agents of type \( h \).

The following assumptions simplify the market (equilibrium) equation. First, since the Ant Process is tailored to two types (stochastic processes get much more complicated when they have dimension higher than one) and two types already gives the modeler much flexibility, the model is studied using two types only. Secondly, it is assumed that \( E_{ht}[y_{t+1}] = E_t[y_{t+1}] = E[y_t] = \bar{y} \) for all \( h \). This assumption makes sure that there exists a constant fundamental price \( p^* = \bar{y}/r \) and traders agree on the average dividend stream. Finally, defining the price in deviations from the fundamental price, i.e. \( x_t = p_t - p^* \), it is assumed that \( E_{ht}[p_{t+1}] = p^* + f_h(x_{t-1}) \) for all \( h \). The last assumption means that all traders make their forecast (1) at the beginning of each period so they can use information from the previous period and (2) based on information from this previous period only. Under this assumption it holds that that the price tomorrow is a weighted sum of the price forecasts of the day after tomorrow of the traders today. Although these assumptions may
not be fully realistic, they are necessary to obtain a tractable model [Brock and Hommes, 1998].

Taking all the assumptions together (10) becomes

\[(1 + r)x_t = w_t f_1(x_{t-1}) + (1 - w_t) f_2(x_{t-1}) = f_2(x_{t-1}) + (f_1(x_{t-1}) - f_2(x_{t-1}))w_t.\]  (11)

From this expression it is immediately clear that the price is an affine function of the fraction \(w_t\). Note, however, that \(w_t\) need not be exogenous and can depend on the price process, which results in a more complicated dependency. Indeed, Brock and Hommes [Brock and Hommes, 1998] use a deterministic evolution rule that depends on the price (which will be discussed next), but a candidate for the evolution of these fractions is also the Ant Process.

Heterogeneous agent models contain multiple types of agents, at most one of which can be rational. Therefore, the resulting model is boundedly rational and the modeler encounters the problem of how to deviate from rationality, i.e. what forecasting rules traders use. General insights from behavioral finance, neuroeconomics etc. can be used to discipline the large amount of possible rules. The main insight is that people tend to use heuristics, simple rules that performed well in the past.

### 3.2.2 Evolutionary component

In the BH model the types are driven by an evolutionary process, where fitness is determined by the price process. At each time \(t\) the fractions are given by the Gibbs probabilities (multinomial logit probabilities)

\[n_{ht} = P_h = \frac{e^{\beta \pi_{h,t-1}}}{\sum_{h=1}^{H} e^{\beta \pi_{h,t-1}}},\]

where \(\pi_{h,t-1}\) is the payoff (fitness) in the previous period and \(\beta\) is the intensity of choice, i.e. the degree to which agents select the optimal strategy. Note that if \(\beta = 0\), all strategies are equally likely, while if \(\beta \to \infty\) the strategy with the highest payoff is chosen with certainty. The payoff can be taken as minus the squared prediction error, i.e. \(- (x_t - f_h(x_{t-1}))^2\), which is a measure of how close the prediction was to the realized price. It is assumed that the squared prediction error is a proxy for profit (i.e. the increased wealth between periods due to the risky asset): when the error decreases, the profit increases.

One interesting interpretation of the probability distribution \(\{P_h\}\) is the following [Weisbuch et al., 2000]. Consider the criterion function \(F = \beta G + S\), where \(S = - \sum_h P_h \ln P_h\) (the entropy of the probability distribution) and \(G = \sum_h P_h \pi_h\) (the expected profits). Maximizing \(F\) with respect to \(P_h\) given the constraint that \(\sum_h P_h = 1\), the following first order conditions hold

\[\beta \pi_h - \log P_h - 1 - \lambda = 0 \quad \forall h, \quad \sum_h P_h = 1.\]

The first \(h\) equations solve to \(P_h = e^{\beta \pi_h} e^{-(1+\lambda)}\). The last equation gives \(e^{1+\lambda} = \sum_h P_h\). Therefore,
the Gibbs probabilities maximize $F$. This means that, depending on $\beta$, the resulting probabilities take into account the trade off between immediate profit and trial and error.

Another interpretation of the Gibbs probabilities is in terms of a random utility model [Hommes and Wagener, 2009]. Assume that agent $i$’s observed pay-off is $\hat{\pi}_{ht} = \pi_{ht} + \epsilon_{ih}$, where $\pi_{ht}$ is the actual pay-off and $\epsilon_{ih}$ is i.i.d. with a double exponential distribution. When $N \to \infty$, the probability that an agent chooses strategy $h$ is $P_h$. The parameter $\beta$ is inversely proportional to the variance of $\epsilon_{ih}$, so $\beta$ indirectly determines to what extent agents favor immediate payoff over (long-term) search for higher profits.

The BH model has been analyzed for simple examples with evolutionary learning [Brock and Hommes, 1998]. Typically, the model shows chaotic behavior in various regions of the parameter space. In addition, complicated bifurcations (as the BH model is a deterministic dynamical system) occur when $\beta$ increases from 0 to $\infty$. When noise is added to the market equation, the model becomes stochastic; in this case some stylized facts, such as volatility clustering, are observed [Brock and Hommes, 1998].

4 The model

In this section the general model is introduced. First, some intuition is given regarding the original Ant Process in Subsection 4.1. Then, the model is introduced, both the continuous and discrete version. Finally, the equilibrium distribution of the model is analyzed in Subsection 4.4. The price in deviation of the fundamental and the fraction of agents in type 1 are generally denoted by capital letters, except when it is clear that they are interpreted pointwise, i.e. $x_t = X_t(\omega)$ and $w_t = W_t(\omega)$ for $\omega \in \Omega$; this will usually be the case when we analyze the market equation.

4.1 A special case

In the simple case of $r = 0$, $f_1 = 1$, $f_2 = 0$ and $e_+ = e_- = 1$, the market equation becomes $x_t = (1-0)w_t$ or $X_t = W_t$ (i.e. the price is equal to the fraction of type 1) and the generalized Ant Process reduces to the ordinary Ant Process. It is therefore of some interest to examine whether the original Ant Process exhibits the stylized facts outlined above. In addition, the stochasticity of the general model is in a sense driven by the original Ant Process. (Clearly, it is not realistic that $0 \leq X_t \leq 1$, so this is an artificial example.)

The motivation for the Ant Process was the observation that ants collectively spent most of the time either at food source 1 (type 1) or at food source 2 (type 2), i.e. most of the time in the neighborhood of $k = 0$ and $k = N$. In order to show that the Ant Process exhibits this behavior [Kirman, 1993] calculated the equilibrium distribution in the limit where $N \to \infty$. We will now proceed to compute this equilibrium distribution using a different route.
4.1.1 Equilibrium distribution

It follows that in the present case, \( e_+(w_t, x_{t-1}) = e_-(w_t, x_{t-1}) = 1 \), \( D^{(1)}(w) = \alpha(1 - 2w) \) and \( D^{(2)}(w) = 2(1 - w)w \). In this case, (7) reduces to

\[
\frac{\partial}{\partial t} P(w, t) = -\frac{\partial}{\partial w}(\alpha(1 - 2w)P(w, t)) + \frac{\partial^2}{\partial w^2}(w(1 - w)P(w, t)).
\]

Setting the time derivative to zero and integrating once with respect to \( w \), the following Ordinary Differential Equation (ODE) for the equilibrium distribution is obtained (using \( f \) instead of \( P \))

\[
-\alpha(1 - 2w)f(w) + \frac{\partial}{\partial w}(w(1 - w)f(w)) = E
\]

where \( E \) is the integration constant. Since \( f(w) \) is a probability density function, boundary conditions dictate that \( E = 0 \). The resulting ODE is then

\[
-\alpha(1 - 2w)f(w) + (1 - 2w)f(w) + w(1 - w)\frac{df(w)}{dw} = 0 \Leftrightarrow \frac{df(w)}{dw} = \frac{(\alpha - 1)(1 - 2w)}{w(1 - w)}.
\]

Integrating this equation gives \( \log f(w) = (\alpha - 1)(\log w + \log 1 - w) + D \), where \( D \) is chosen such that \( f(w) \) is normalized. It follows that \( f(w) = C[w(1 - w)]^{\alpha-1} (C = e^D) \), which is the Probability Density Function (PDF) of the symmetric beta distribution. The equilibrium behavior of the system depends on \( \alpha \): if \( \alpha < 1 \) the system will spend most of its time in one of the extremes \( (w = 0 \text{ or } w = 1) \) while if \( \alpha > 1 \) the system spends most of its time in the middle \( (w = 1/2) \).

4.1.2 SDE

In this simple case (9) reduces to

\[
W_t = W_0 + \int_0^t \alpha(1 - 2W_s)ds + \int_0^t \sqrt{2W_s(1 - W_s)}dB_s.
\]

This is a special form of the same SDE that defines a Pearson diffusion with \( \theta = 2\alpha, \mu = 1/2, a = -1/\theta, b = 1/\theta \) and \( c = 0 \) (although, strictly speaking, a Pearson diffusion is defined as stationary process that solves the SDE) [Forman and Sørensen, 2008]. It may also be viewed as a combination of the Ornstein-Uhlenbeck process and the Moran process of genetic drift. Furthermore, note that the process is mean reverting to \( w = 1/2 \) with mean reversion parameter \( 2\alpha \). Therefore, the drift and volatility terms seem to counteract each other: as soon as the process is close to \( w = 1/2 \), it is likely to move away again. Which of these is dominant depends on \( \alpha \); for small \( \alpha \) one expects the volatility to dominate, while for large \( \alpha \) the drift term dominates. The above SDE has a strong solution for any constant initial condition [Abundo, 2009].

4.1.3 Moments and stylized facts

Even though no explicit solution exists, it is in this case possible to calculate moments of the process, which is useful for understanding its dynamics (and thus stylized facts). First, note that
the volatility term is a martingale, since \( E \int_0^t X_s (1 - X_s) ds \) is finite, because the integrand is bounded. This last property is used extensively below, since it implies that the expectation of the volatility term is zero. Taking expectations on both sides of the SDE gives

\[
E[W_t] = E[W_0] + \int_0^t \alpha (1 - 2E[W_s]) ds \Leftrightarrow \frac{d\mu(t)}{dt} = \alpha (1 - 2\mu(t))
\]

where \( \mu(t) = E[W_t] \). This simple differential equation has solution \( \mu(t) = \frac{1}{2} + e^{-2\alpha t}(\mu(0) - \frac{1}{2}) \).

Clearly, if \( \mu(0) = \frac{1}{2} \) the process has a constant mean. Otherwise, the mean will return to this level with velocity depending on \( \alpha \). Note that \( w = \frac{1}{2} \) is also the mean of the symmetric beta distribution, the equilibrium distribution of the process.

The variance of the process \( W_t \) can be computed easily as well by defining a new process \( V_t = W_t - \mu_t \). This process then has mean zero by definition and Ito’s lemma gives, in shorthand notation,

\[
dV_t = d(W - \mu)_t = dW_t - d\mu_t = \alpha (1 - 2W_t) dt + \sqrt{2W_t (1 - W_t)} dB_t - \alpha (1 - 2\mu_t) dt
\]

which is an SDE solely for \( V \) since \( \mu \) is known explicitly. To obtain the variance Ito’s lemma is applied again with \( f(w) = w^2 \) to obtain

\[
dV_t^2 = 2V_t dV_t + \frac{1}{2} 2d\langle V \rangle_t = 2[V_t [-\alpha 2V_t dt + \sqrt{2(V_t + \mu_t) (1 - (V_t + \mu_t))} dB_t] + 2(V_t + \mu_t)(1 - (V_t + \mu_t)) dt \Rightarrow d\sigma_t^2 = -\alpha 4\sigma_t^2 dt - 2\sigma_t^2 dt + 2(\mu_t - \mu_t^2) dt,
\]

where \( \sigma_t^2 \) is the variance at time \( t \) and it was used that \( E[V_t] = 0 \). Plugging in \( \mu_t \), this leads to the following differential equation

\[
\frac{d\sigma_t^2}{dt} = -2(2\alpha + 1)\sigma_t^2 + 2 \left( \frac{1}{4} - e^{-4\alpha t} \left( \mu_0 - \frac{1}{2} \right)^2 \right)
\]

with solution

\[
\sigma_t^2 = \frac{1}{4(1 + 2\alpha)} \left( 1 - e^{-2(1 + 2\alpha)t} \right) - \left( 1 - e^{-2t} \right) e^{-4\alpha t} \left( \mu_0 - \frac{1}{2} \right)^2 + e^{-2(1 + 2\alpha)t} \sigma_0^2,
\]

where \( \sigma_0^2 \) is the variance of the initial distribution, or initial variance. The variance always decreases exponentially over time with speed depending on \( \alpha \), which makes sense since the process is defined on a compact interval. In the special case that the initial distribution is a constant and equal to \( 1/2 \), the variance is not constant, in contrast to \( \mu \). Furthermore, the variance is bounded from below by \( 1/(4 + 8\alpha) \) to which it converges as \( t \to \infty \). Note that this can still be large relative to \( I \) if \( \alpha \) is small and that it equals the variance of the symmetric beta distribution which is the...
equilibrium distribution of the process. In addition, the variance converges to this value as $t \to \infty$, which again is in agreement with its equilibrium distribution.

Using Ito’s formula with $f(w) = w^3$ we get

$$dV_t^3 = 3V_t^2 dV_t + \frac{1}{2} 6V_t d\langle V \rangle_t = 3V_t^2 \left[-\alpha 2V_t dt + \sqrt{2(V_t + \mu_t)}(1 - (V_t + \mu_t)) dB_t\right] +$$

$$6V_t(V_t + \mu_t)(1 - (V_t + \mu_t)) dt.$$

Taking expectations and plugging in $\mu_t$ and $\sigma^2_t$, the following differential equation is obtained for the third central moment, $\mu_{3,t}$,

$$\frac{d\mu_{3,t}}{dt} = 6[-\alpha \mu_{3,t} + \sigma^2_t - \mu_{3,t} - 2\mu_t \sigma^2_t] = -6[(\alpha + 1)\mu_{3,t} + 2e^{-2\alpha t}\left(\mu_0 - \frac{1}{2}\right)\sigma^2_t].$$

However, as this equation has a solution whose expression would take three lines, the focus is on the solution for the case that the initial random variable has mean equal to $1/2$ to get

$$\mu_{3,t} = \mu_{3,0} e^{-6(\alpha + 1)t}.$$

If $t \to \infty$, the third central moment (and therefore the skewness) converges to zero. In addition, if initial skewness is zero, skewness is zero for all time $t$. It can be shown that the Ant Process is symmetric around $1/2$, so these results are expected.

Finally, for the fourth central moment, $\mu_{4,t}$, using the same trick with $f(w) = w^4$ one obtains

$$\frac{d\mu_{4,t}}{dt} = 12 \left[- \left(\frac{2}{3}\alpha + 1\right)\mu_{4,t} + (1 - 2\mu_t)\mu_{3,t} + (1 - \mu_t)\mu_t \sigma^2_t\right] =$$

$$-12[(\alpha + 1)\mu_{4,t} + 2e^{-2\alpha t}\left(\mu_0 - \frac{1}{2}\right)\mu_{3,t} - \left(\frac{1}{4} - e^{-4\alpha t}\left(\mu_0 - \frac{1}{2}\right)^2\right)\sigma^2_t].$$

As for the third moment, the focus is on the solution for the case that $\mu_0 = 1/2$. It is given by

$$\mu_{4,t} = \frac{3}{16(1 + 2\alpha)(5 + 2\alpha)} + \frac{3((4 + 8\alpha)\sigma^2_0 - 1)e^{-2(1 + 2\alpha)t}}{8(1 + 2\alpha)(5 + 2\alpha)} +$$

$$\frac{3}{16(3 + 2\alpha)(5 + 2\alpha)} e^{-4(3 + 2\alpha)t} - \frac{3\sigma^2_0 e^{-4(3 + 2\alpha)t}}{2(5 + 2\alpha)} + \mu_{4,0} e^{-4(3 + 2\alpha)t}.$$

In the limit of $t \to \infty$, the fourth central moment is given by

$$\frac{3}{16(1 + 2\alpha)(5 + 2\alpha)}.$$

The variance in this limit is equal to $1/(4 + 8\alpha)$. The excess kurtosis (i.e. the kurtosis minus three, which is the kurtosis of the normal distribution) is therefore equal to $-6/(2\alpha + 3)$, which coincides with the excess kurtosis of the symmetric beta distribution. The results for $t \to \infty$ are expected since the invariant (or equilibrium) distribution of the process is the symmetric beta distribution.
For $t = 0$, however, the fourth moment, and therefore the kurtosis, can be large because it is determined by the initial distribution. Since the fourth central moment decays exponentially to the invariant distribution fat tails become thin rapidly; for small $\alpha$ this takes longer than for large $\alpha$.

First, note that it was found above that the model has no excess kurtosis, so it does not exhibit fat tails (which makes sense since it is defined on $I$). Secondly, since the model is fully symmetric around $w = 1/2$, there can be no bubbles and crashes (this implies a certain asymmetry between increasing and decreasing $w$ and it was shown that the skewness is close to zero).

However, the Ant Process shows the other two properties (at least in the discretized SDE, as shown in Subsection 3.1). Since differences (i.e. returns) are driven by an i.i.d. normal distribution, there are no correlations in returns. This can also be seen by simulations of the model (see Figures 1 and 2). In addition, the process shows some volatility clustering because of the following. Suppose $w$ is close to 0 or 1 (which will be the case most of the time in the regime of $\alpha < 1$). If there is a large deviation up or down respectively, the next return will have higher volatility (volatility increases when moving towards $w = 1/2$), so that large deviations are likely to be followed by large deviations (of opposite sign). However, a large deviation in any of the other two directions is likely to be followed by a smaller deviation, but the magnitude of this effect is smaller. In sum, volatility clustering is present, but not strongly and it is conditional on $\alpha$ being small (as can be seen in Figures 1 and 2).

In the continuous framework the concept of returns has no meaning. However, the autocorrelation function of the process and its square can be computed, which indicates whether the continuous process satisfies the corresponding stylized facts to some extent. First, the autocovariance function is computed.

Consider again the process $V$, as defined above. We are then interested in the autocovariance of $W$, $\text{E}[V_t V_s]$, where it is assumed, without loss of generality, that $s \geq t$. We have that $\text{E}[V_t V_s] = \text{E}[\text{E}[V_t V_s | \mathcal{F}_t]] = \text{E}[V_t \text{E}[V_s | \mathcal{F}_t]]$, where $\mathcal{F}_t$ is the natural filtration. Therefore, the quantity of interest is $\hat{V}_s = \text{E}[V_s | \mathcal{F}_t]$. Note that $\hat{V}_t = V_t$. To compute this quantity, first write

$$V_s = V_t + \int_t^s \alpha 2V_u du + \int_t^s \sqrt{2(V_u + \mu_u)}(1 - (V_u + \mu_u)) dB_u.$$ 

Taking conditional expectations on both sides and noting that by the independent increments of BM the conditional expectation of the stochastic integral reduces to an ordinary expectation, which equals zero by similar arguments as before, we get

$$\hat{V}_s = V_t - 2\alpha \int_t^s \hat{V}_u du.$$
Figure 1: Simulation of Ant Process with $h = 0.01$ and $\alpha = 0.1$

Figure 2: Simulation of Ant Process with $h = 0.01$ and $\alpha = 30$
This form implies that $\hat{V}_s$ is pathwise differentiable and the resulting, pathwise, differential equation has solution

$$\hat{V}_s = e^{-2\alpha(s-t)}V_t.$$ 

Inserting this result in the original expectation gives

$$E[V_tV_s] = e^{-2\alpha(s-t)}\sigma_t^2,$$

so that the autocorrelation is $e^{-2\alpha(s-t)}\sigma_t/\sigma_s$. As the variance converges to its equilibrium value with exponential speed, most of the time $\sigma_t/\sigma_s$ is roughly equal to one, so the autocorrelation is roughly equal to $e^{-2\alpha(s-t)}$. This means that the autocorrelation goes to zero with exponential speed as $s \to \infty$; again, if $\alpha$ is small, this takes more time. In general this shows that the Ant Process has little autocorrelation.

Next, the autocovariance of $W^2$, $E[V_t^2V_s^2]$, is computed. For the same reasons as before, the quantity of interest is now $\hat{V}_s^2 = E[V_s^2|\mathcal{F}_t]$. The SDE for $V^2$ was previously derived. By using similar arguments as before, the resulting, pathwise, differential equation for $\hat{V}_s^2$ is

$$\frac{d\hat{V}_s^2(\omega)}{ds} = -2(2\alpha + 1)\hat{V}_s^2(\omega) + 2(1 - \mu_s)\hat{V}_s(\omega) + 2(1 - \mu_s)\mu_s.$$

If we now assume, as above, that $\mu_0 = 1/2$, the solution to the inhomogeneous differential equation is

$$\hat{V}_s^2 = e^{-2(2\alpha+1)(s-t)} \left( V_t^2 - \frac{1}{4(2\alpha+1)} \right) + \frac{1}{4(2\alpha+1)}.$$

Inserting this into the original expectation then gives the autocovariance of the squares

$$E[V_t^2V_s^2] = e^{-2(2\alpha+1)(s-t)} \left( \mu_{4,t} - \frac{\sigma_t^2}{4(2\alpha+1)} \right) + \frac{\sigma_t^2}{4(2\alpha+1)}.$$

As the corresponding autocorrelation turns out to have a complicated expression and the concern is whether there is non-zero autocorrelation, we may only analyze the autocovariance. Now, even if $s \to \infty$, the autocovariance remains positive (i.e. equal to the second and last term on the right), so there is long-range dependence in the squares. This term becomes larger (smaller) as $\alpha$ becomes smaller (larger). In addition, for finite (small) $s$ the value of the autocovariance depends also on the difference $\mu_{4,t} - \sigma_t^2/(8\alpha + 4)$. However, it is clear that even for moderate $s$ (relative to $\alpha$) the exponential factor dominates the first term, so this will be close to zero. Even for small $s$, the aforementioned difference is smaller than the last term on the right in terms of absolute value, since $\mu_{4,t}$ is positive, so the autocovariance is likely to be positive even then. Therefore, there seems to be volatility clustering, especially for small $\alpha$. (Note that we have computing the autocovariance of the squares, not the absolute value; although there may be some differences we expect the results to be qualitatively similar.)
4.2 General model

In the general model the evolution of types in the market equation is driven by the Ant Process which is modified to introduce coupling by multiplying the $\epsilon$ term in the transition probabilities with the corresponding Gibbs’ probabilities, introduced in Subsection 3.2. In this way the probability that an agent converts to another type independently depends on how large the prediction error of that type is. Therefore, the population dynamics combines both social interaction and reinforcement learning, but in a different way than [Brock and Durlauf, 1997], since social interaction is not modeled globally. This ensures that the probability to independently convert to another type increases with the fitness of that type. The evolution of types is thus a combination of the Ant Process and the evolution rule used in the BH model.

The above model can be viewed as the SEGT analogue of the BH model: a stochastic evolutionary game (a Markov chain associated to pay-offs), coupled to a market equation, since the pay-offs change due to market dynamics. In the rest of this section, the forecasting rules are assumed to be affine functions of $x_{t-1}$, i.e. $f_1(x_{t-1}) = a_1 + b_1 x_{t-1}$ and $f_2(x_{t-1}) = a_2 + b_2 x_{t-1}$, where $a_1, a_2, b_1, b_2 \in \mathbb{R}$. When $a_i = 0$ the type is a pure trend follower; when $b_i = 0$ it is a biased type (optimist or pessimist); and when both are zero the trader is a fundamentalist. Other combinations are also possible.

Alternatively, one might let the probability of interaction be proportional to the Gibbs probabilities (i.e. multiply $1 - \delta$ with the Gibbs probabilities), since it could be argued that people are more likely to be persuaded to a different strategy if it is more successful, which was suggested in the original paper by Kirman [Kirman, 1993] and the subject of a simulation study [Westerhoff, 2009]. This will not be pursued here since it turns out that there are problems with the continuous time limit in this case, which has to do with the drift term:

$$N \left[ (1 - k) \left( \epsilon + (1 - \delta)a \frac{k}{N - 1} \right) - k \left( \epsilon + (1 - \delta)b \frac{N - k}{N - 1} \right) \right]$$

$$= \alpha (1 - 2w) + \frac{N}{N - 1} (1 - \delta) [(1 - w)awN - wb(1 - w)N]$$

where $k = wN$. Clearly, the right term diverges when $N \to \infty$ if $a \neq b$ and at least one of them is independent of $N$. Otherwise it is zero and the above expression reduces to the standard drift term. Replacing $a$ and $b$ by $a/N$ and $b/N$ respectively solves the problem for $a \neq b$, but the Gibbs probabilities should not depend on $N$, so the above suggestion is impossible to implement in this context.

Returning to the current framework, (11) becomes

$$(1 + r)x_t = f_{2t} + (f_{1t} - f_{2t})w_t = a_2 + b_2 x_{t-1} + (a_1 - a_2 + (b_1 - b_2)x_{t-1})w_t$$

(12)

and the transition functions $T(k + 1|k)$ and $T(k - 1|k)$ are now given by

22
\[
T(k+1|k) = \left(1 - \frac{k}{N}\right) \left(\epsilon e_+(k/N, X_{t-1}) + (1 - \delta) \frac{k}{N - 1}\right)
\]

\[
T(k-1|k) = \frac{k}{N} \left(\epsilon e_-(k/N, X_{t-1}) + (1 - \delta) \frac{N-k}{N - 1}\right)
\]

where

\[
e_+(w, x_{t-1}) = \frac{e^{\beta \pi_{1,t}(w, x_{t-1})}}{Z_t}, \quad e_-(w, x_{t-1}) = \frac{e^{\beta \pi_{2,t}(w, x_{t-1})}}{Z_t},
\]

are the Gibbs probabilities, where

\[
\pi_{1,t}(w, x_{t-1}) = -(H(x_{t-1}, w_t) - a_i - b_i x_{t-1})^2
\]

\[
= -\frac{1}{(1 + r)^2} (a_2 + b_2 x_{t-1} + (a_1 - a_2 + (b_1 - b_2) x_{t-1}) w_t - a_i (1 + r) - b_i (1 + r) x_{t-1})^2
\]

is the prediction error of type \(i\) at time \(t\).

### 4.2.1 Continuous dynamics

To get a continuous 2D system, the market equation has to become a differential, instead of a difference equation and the Markov process has to be replaced by an SDE.

To deal with the market equation, assume that at each time step, \(\Delta t\), only a fraction of the agents, \(\gamma\), is active, so that the market equation becomes

\[
x_t = (1 - \gamma)x_{t-\Delta t} + \gamma \left[\frac{1}{1 + r} (f_2(x_{t-\Delta t}) + (f_1(x_{t-1}) - f_2(x_{t-\Delta t})) w_t)\right].
\]

In order to obtain a continuous time framework the time steps should approach zero. Therefore, let \(\gamma\) be proportional to \(\Delta t\) (i.e. \(\gamma = c \Delta t\) for some constant \(c\)) and replace \((1 + r)\) by \((1 + r \Delta t)^{1/\Delta t}\).

Note that in this way as the time steps go to zero the fraction of active agents goes to zero as well, which makes sense in a continuous framework. The equation can now be written as

\[
\frac{x(t) - x(t - \Delta t)}{c \Delta t} = \frac{1}{(1 + r \Delta t)^{1/\Delta t}} (f_2(x(t - \Delta t)) + (f_1(x(t)) - f_2(x(t))) w(t)) - x(t - \Delta t).
\]

Letting \(\Delta t \to 0\) gives

\[
\frac{1}{c} \frac{dx}{dt} = e^{-\tau}[f_2(x(t)) + (f_1(x(t)) - f_2(x(t))) w(t)] - x(t).
\]

In the case of our model this becomes (setting \(c = 1\) for convenience)

\[
\frac{dx}{dt} = e^{-\tau}[a_2 + b_2 x(t) + (a_1 - a_2 + (b_1 - b_2) x(t)) w_t] - x(t).
\]
Next, the prediction error has to be discussed, since it also depends indirectly on the market equation. When assuming that part of the traders are inactive at smaller time steps, the squared prediction errors are affected as follows. We have that
\[
x(t) - f_h(x(t - \Delta t)) =
(1 - \gamma) x_{t-\Delta t} + \gamma \left[ \frac{1}{1 + r} (f_2(x_{t-\Delta t}) + (f_1(x_{t-\Delta t}) - f_2(x_{t-\Delta t}))w_t) \right] - f_h(x(t - \Delta t))
\]
Clearly, as $\Delta t$ goes to zero, the prediction errors reduce to $x(t) - f_h(x(t))$. Indeed, much of the economic interpretation of the forecasting rules and prediction errors is lost if $x(t - \Delta t) \to x(t)$.

This is the biggest disadvantage of the continuous time approach. Note also that the traders are assumed to be inactive in terms of trading behavior only; there are, therefore, no direct implications for the Ant Process, which is discussed next.

When obtaining the FP equation of the above Markov chain (as was shown in 3.1), the prediction errors converge as well, as described above. The SDE (9) now reduces to
\[
dW_t = \alpha(e_+ (W_t, X_t) - W_t)dt + \sqrt{2W_t(1-W_t)} dB_t.
\]
Together with the market equation (12) this comprises a 2D system, which will be referred to as the Market Process. Note also that $W_t$ always remains in $I$ because of the following. If $W_t = 0$ or $W_t = 1$, then the volatility term is zero. Therefore we must have that the drift term at $W_t = 0$ is positive, while it is negative at $W_t = 1$; this ensures that the process moves back up and down respectively. Since $e_+$ is always between one and zero, this condition is always satisfied.

The volatility term in (14) remains unchanged, while the drift term has become quite complicated. Since the BH model is a deterministic model it makes sense that the corresponding dynamics appear in the drift term (which consists of the diagonal elements of the transition matrix). Note that for $\beta = 0$ the drift term reduces to the one in the ordinary Ant Process with $2\alpha$, instead of $\alpha$.

In addition, the model may be looked at as an Ant Process with parameter $\alpha/2$ and a changing mean; the mean normally is equal to $1/2$, but now the term $e_+(W_t, X_t)$ fluctuates endogenously and exogenously (depending on both the process itself, $W_t$, and an additional process, $X_t$) between 0 and 1.

The market equation could be solved explicitly (pointwise) in terms of $w(t)$. Since the above equation is a inhomogeneous first order linear differential equation its solution can easily be seen to be
\[
x(t) = e^{f_0} e^{-r(b_2 + (b_1 - b_2)w(t')) - 1 dt'} e^{-r(a_2 + (a_1 - a_2)w(t))} dt',
\]
where $C = x(0)$. This expression can then be plugged into the SDE, so that it is in terms of $W$ only. However, this would lead to a highly complicated SDE which cannot be dealt with analytically. We can, therefore, only hope to study this system through its equilibrium distribution.
4.3 Discrete dynamics

Since a discrete framework is useful for simulation and estimation purposes, the first step is to discretize the Market Process. To obtain the SDE, consistency required letting \( x(t - \Delta) \rightarrow x(t) \). However, when discretizing the SDE it seems most natural to return to the situation where the squared prediction error depended on \( x_{t-1} \) instead of \( x_t \) (i.e. this is the way the squared prediction error was originally constructed in a discrete time framework). In summary, the following stochastic difference equation is obtained (where \( W_n = W(t_n) \) and \( t_n = nh \) as before)

\[
W_{n+1} = W_n + \alpha(e_+ (W_n, X_{n-1}) - W_n)h + \sqrt{2W_n(1-W_n)}\Delta B_n,
\]  

(15)

where again \( \Delta B_n \) has a normal distribution with mean zero and variance \( h \). The parameter \( h \) should be chosen according to the relevant time scale. Note that the time scale does not directly correspond to the time scale of the original Markov chain. (15) is a discrete time approximation of (14). Now, it is not clear anymore whether the process remains in \( I \), although for small \( h \) this is unlikely. In addition, in the case of simulation the process can be artificially bound to \( I \), by letting it be equal to one if it becomes larger than one and similarly for zero. When estimating the model, it will turn out that there is a restriction on the fractions anyway, because the fractions are not observed. Since simulations show that if \( h \) is much larger than 0.01, the volatility becomes too large for the process to remain in the \( I \), let \( h = 0.01 \). Together with the market equation

\[
(1 + r)X_n = a_2 + b_2 X_{n-1} + (a_1 - a_2 + (b_1 - b_2)X_{n-1})W_n,
\]

this constitutes the discrete Market Process. Note that in this setup it is assumed that the market clears at the same rate as the ant Ant Process. The market equation is simply the original equation (i.e. before making it continuous).

In Figures 3 and 4, simulation results of the Market Process are shown (only \( X \)) for the case where one of the traders is a fundamentalist and the other is a chartist. Both figures show the feedback between the price and the fractions. In the case of Figure 3, if the price increases, the Gibbs probability increases, so that on average the fraction of chartists increases, which in turn causes the price to increase. The stochasticity of the model ensures that bubbles cannot last forever: at some point, a downward jump in fractions causes the payoffs of the chartists to drop enough so that the price decreases, which causes the price to decrease even further and so forth. Note, that the process \( X \) exhibits fat tails, which means large deviations occur like in financial markets.

In Figure 4, the same behavior is observed, although it is a mirror-image (through the \( x \)-axis). This is because the chartists now believe that the price is below the fundamental. If the chartists are not biased, \( x = 0 \) is an absorbing state and the dynamics are less interesting. If one wants to study unbiased chartists non-trivially, it is possible to add some noise to the market equation; this ensures steady perturbation away from the fundamental price. Clearly, the fact that the second graph is a mirror image of the first calls for some reservations, since in real markets 'upward crashes' do not occur.
fractions, prices and Gibbs probabilities

\[ X_n, W_n, e^{+(X_n, W_{n+1})} \]

Figure 3: Simulation of the Market Process with \( h = 0.01, \alpha = 0.1, \beta = 5, r = 0.05, a_1 = 0.02, a_2 = 0, b_1 = 1.02 \) and \( b_2 = 0 \)

fractions, prices and Gibbs probabilities

\[ X_n, W_n, e^{+(X_n, W_{n+1})} \]

Figure 4: Simulation of the Market Process with the same parameters as 3, except that \( a_1 = -0.02 \)
4.4 Equilibrium distribution

In equilibrium, the time derivative in (13) has to be set to zero to obtain (writing capital $X$ and $W$ again)

$$e^r X = a_2 + b_2 X + (a_1 - a_2 + (b_1 - b_2) X) W \Leftrightarrow X = \frac{a_2 + (a_1 - a_2) W}{e^r - b_2 - (b_1 - b_2) W},$$

which gives an equilibrium relation between the price $X$ and the fraction $W$, which we denote by $X = h(W)$. Note that $h$ is only well defined if $e^r - b_2 \neq (b_1 - b_2) W$. The equilibrium distribution for $X$ can then be obtained from the equilibrium distribution of $W$. To find the latter, consider the modified FP equation

$$\frac{\partial}{\partial t} P(w,t) = -\frac{\partial}{\partial w}(\alpha(x(t),w) - w) P(w,t)) + \frac{\partial^2}{\partial w^2} (w(1-w) P(w,t)),$$

The equilibrium behavior is obtained by setting the time derivative equal to zero, integrating once with respect to $w$ (setting the integration constant equal to zero as before) and let the density functions and $x(t)$ not depend on time anymore to get (using $f$ instead of $P$

$$-(\alpha(x(w),w) - w)f(w)) + \frac{\partial}{\partial w} (w(1-w)f(w)) = 0 \Leftrightarrow \frac{f'(w)}{f(w)} = 2w - 1 + \alpha(x(w),w) w(1-w).$$

Now, the expression for $x = h(w)$ can be inserted to obtain a first order ODE in terms of $w$, i.e. $f'(w) = g(w, f(w))$. Note that $g$ is clearly Lipschitz continuous in $f$ and continuous in $w$, except at the boundary points (which does not matter, as in the case of the ordinary Ant Process). By the Picard-Lindelof theorem, this ODE must have a solution. Dividing both sides by $f(w)$ and integrating with respect to $w$ gives

$$\ln f(w) = \int_0^w \frac{2w' - 1 + \alpha(x(h(w'),w') - w')}{w'(1-w')} dw' + C$$

However, since $e_+(h(w),w)$ depends in a complicated way on $w$ and six other parameters, the integral is rather complicated and cannot be analytically evaluated.

To make the integral tractable $e_+(h(w),w)$ can be expanded in a Taylor series around $w = 1/2$. Note that even though the process may spend most of the time away from $w = 1/2$, the most interesting behavior occurs in the middle. Furthermore, $w = 1/2$ lies exactly in the middle of $I$, making it a natural choice. In the case of a zeroth order approximation, the above integral becomes

$$\int_0^w \frac{2w' - 1 + \alpha(A - w')}{w'(1-w')} dw'$$

where $A = e_+(h(1/2),1/2)$ depends only on the parameters. The solution to this ODE is

$$f(w) = C w^{r-1} (1-w)^{t-1},$$
where $s = \alpha A$ and $t = \alpha (1 - A)$. Note that this only defines a probability density if $s, t > 0$. This is always the case, since $0 \leq A \leq 1$ because it is a probability and $\alpha > 0$ by construction. In addition, this is just an asymmetric beta distribution, which reduces to a symmetric beta distribution only if $A = 1/2$. If the approximation is first order one also obtains an asymmetric beta distribution with coefficients $\alpha (A - B/2)$ and $\alpha (1 - A - B/2)$ respectively, where

$$B = \frac{de_+ (h(w), w)|_{w=1/2}}{dw}.$$ 

However, the space of feasible parameters such that the coefficients are positive is very complicated. Higher order approximations lead to distributions more complicated than the beta distribution.

As the coefficients of the beta distribution depend on $A$, it is of interest to determine how $A$, in turn, depends on the underlying parameters of the model. For simplicity assume, as was done in the simulations, that $a_2 = b_2 = 0$. Now,

$$A = \frac{e^{-\beta \pi_1}}{e^{-\beta \pi_2} + e^{-\beta \pi_2}},$$

where equilibrium prediction errors at $w = 1/2$ are given by

$$\pi_1 = a_1^2 \left(\frac{1 - 2e^r}{2e^r - b_1}\right)^2 \quad \pi_2 = a_1^2 \left(\frac{1}{2e^r - b_1}\right)^2.$$ 

Now, if $\beta a_1^2$ increases (decreases), $A$ will move closer to (away from) 0 or 1, depending on the relative difference between $\pi_1$ and $\pi_2$. Increases in either $a_1$ or $\beta$ then clearly have the same effect. As the second factors in $\pi_1$ and $\pi_2$ differ only in the numerator of the fraction in the square, the numerators are essential. If $r = 0$ the quantities are the same after squaring. However, as $r$ increases, $\pi_1$ becomes larger, so $A$ decreases. The size of this increase depends on the denominator: if $b_1$ is close to $2e^r$ the effect is largest.

We are interested in the equilibrium distribution of $X$. To build on the previous example assume again that $a_2 = b_2 = 0$. The equilibrium relation is then

$$X = h(W) = \frac{a_1 W}{e^r - b_1 W}.$$ 

We proceed to analyze properties of $h$ in this case. As soon as $e^r = b_1 W$, there is a vertical asymptote within $I$. Therefore, a distinction has to be made between the case $e^r < b_1$, in which case there is a vertical asymptote in the interior of $I$, and $e^r \geq b_1$, in which case there is no vertical asymptote there. Note that in the second case, the range of $h$ is $[0, a_1/(e^r - b_1)] \cup [0, \infty)$. In the first case, however, $x$ can have the values in $(-\infty, a_1/(e^r - b_1)] \cup [0, \infty)$. Apparently, in the second case the value of $b_1$ is not big enough (i.e. the trader does not attach enough weight to the previous price) relative to $e^r$ for equilibrium prices to have sign different from $a_1$. In the first case, $b_1$ is big enough but if equilibrium prices are to be of different sign than $a_1$, they must be large (i.e. there is a gap between zero and $a_1/(e^r - b_1)$ where prices are apparently to close to zero to be equilibrium prices, given that $a_1$ is of opposite sign).
In addition to the previous distinction, note that

\[ h'(w) = \frac{a_1 e^r}{(e^r - b_1 w)^2}, \]

so \( h \) is strictly monotone on \( I \) if \( e^r < b_1 \) and strictly monotone on \( I_+ := [0, e^r/b_1) \) and \( I_- = (e^r/b_1, 1] \) separately (only) if \( e^r \geq b_1 \). The sign of \( a_1 \) determines whether \( h \) is strictly increasing or strictly decreasing on these subsets of \( I \). (Clearly, if \( a_1 = 0, x = 0 \), so \( X \) has a trivial distribution in this case.) As a result, \( h \) is invertible if \( e^r < b_1 \) with inverse function

\[ w = h^{-1}(x) = t(x) = \frac{e^r x}{a_1 + b_1 x}, \]

while this is not the case when \( e^r \geq b_1 \). However, restricted to either \( I_- \) or \( I_+ \), it is invertible and \( t \) has the same form. Finally, note that \( h \) is injective in all cases (provided its domain does not include the singularity).

Combining the two sets of cases, there are four different cases in total. Now, the CDF transformation method can be used to obtain the CDF for \( X, F_X \), given that we know the CDF for \( W, F_W \), (which we do since we know the distribution of \( W \)). This transformation method is applied for each of the four cases. If \( e^r \geq b_1 \) and \( a_1 > 0 \), then we get that

\[ F_X(x) = P(X \leq x) = P(h(W) \leq x) = P(W \leq t(x)) = F_W(t(x)). \]

If \( e^r \geq b_1 \) and \( a_1 > 0 \), there is a vertical asymptote present, so the situation is more complicated. Suppose first that \( x \geq 0 \). Then

\[ P(X \leq x) = P(h(W) \leq x) = P(W \leq t(x)) \quad \text{or} \quad W \geq e^r/b_1. \]

As the events \( \{W \leq t(x)\} \) and \( \{W \geq e^r/b_1\} \) are disjoint when \( x \geq 0 \), we have

\[ P(X \leq x) = P(W \leq t(x)) + P(W \geq e^r/b_1) = F_W(t(x)) + 1 - F_W(e^r/b_1). \]

Secondly, suppose that \( x \leq a_1/(e^r - b_1) \). Then we get that

\[ P(X \leq x) = P(W \leq t(x)) = F_W(t(x)). \]

In the other two cases, \( a_1 < 0 \), so \( h \) is strictly decreasing on the associated subsets. This means that if \( e^r \geq b_1 \) we have

\[ F_X(x) = 1 - F_W(t(x)) \]

and if \( e^r \geq b_1 \) we have
\[ F_X(x) = \begin{cases} 
1 - F_W(t(x)) & \text{if } x \leq 0 \\
1 - F_W(t(x)) + F_W(e^r/b_1) & \text{if } x \geq a_1/(e^r - b_1) \\
1 - F_W(t(0)) & \text{otherwise} 
\end{cases} \]

The PDF can be derived from the CDF by differentiation. It follows that if \( a_1 > 0 \) \( f_X(x) = f_W(t(x))t'(x) \), where the domain is \((-\infty, a_1/(e^r - b_1)] \cup [0, \infty)\) if \( e^r < b_1 \) and \([0, a_1/(e^r - b_1))\) if \( e^r \geq b_1 \). Furthermore, if \( a_1 < 0 \) \( f_X(x) = -f_W(t(x))t'(x) \), where the domain is \((-\infty), 0] \cup [a_1/(e^r - b_1), \infty)\) if \( e^r < b_1 \) and \([0, a_1/(e^r - b_1))\) if \( e^r \geq b_1 \). Note that the PDF can be extended outside of this natural domain (much like the distribution function) if it is taken to be zero there.

Assume that \( a_1 > 0 \) for concreteness. The explicit functional form of the PDF is then given by

\[ f_X(x) = C \left( \frac{a_1 + (b_1 - e^r)x}{a_1 + b_1 x} \right)^{\alpha(A-1)-1} \left( \frac{e^r x}{a_1 + b_1 x} \right)^{\alpha A - 1} \frac{a_1 e^r}{a_1 + b_1 x}. \]

Note that this function probably not have an anti-derivative that is an elementary function. In addition, it is difficult to obtain moments analytically. However, a plot of the density already gives some important insights. The most interesting parameters are \( \alpha \) and \( b_1 \) (relative to \( e^r \)). The parameter \( a_1 \) mostly amounts to a rescaling of the system and \( \beta \) has a minor impact on the model since it is only present in \( A \). As was shown earlier, when \( a_1 \) increases \( \beta \) should decrease to keep \( A \) fixed (and vice versa), so it makes sense to keep both \( \beta \) and \( a_1 \) fixed and for convenience we take them equal to 1. In addition, \( r \) is fixed at 0.05. Qualitatively, there are then four conditions corresponding to combinations of low/high \( \alpha \) and \( b_1 \) smaller/larger than \( e^r \). The plots are shown below in Figures 5 and 6.

![Figure 5: Approximate equilibrium distribution with \( a_2 = b_2 = 0, \beta = a_1 = 1, \alpha = 0.1 \) (left: \( b_1 = 0.5 \); right: \( b_1 = 1.5 \))](image-url)
When $\alpha = 0.1$, the distribution clearly is leptokurtic: it peaks around $x = h(0)$ (i.e. zero) and $x = h(1)$, but there is a high probability mass in the tails. The high relative probability around these points means that most of the time a large fraction of traders is either of one type or the other (as we saw in the case of the ordinary Ant Process). Also, when $b_1 = 1.5$, negative $x$ has positive probability, but there is an interval between $h(1)$ and $h(0)$, which has no support. Summarizing, for small $b_1$ the equilibrium price remains within an interval, but shows a tendency to stick to one of the endpoints, while for large $b_1$ the equilibrium price can be negative and has fat tails on both the negative and positive side.

On the other hand, when $\alpha = 10$, the distribution is more like a normal distribution, skewed towards the left. Surprisingly, for $b_1 = 1.5$ the probability that $x$ is negative is so small, that it can barely be seen on the graph; with such large $\alpha$, it is unlikely that a significant fraction of the population is an ‘optimistic chartist’. It appears that the tail on the right side of the right graph is fatter than its counterpart for $\alpha = 0.1$, but the tail drops at an exponential rate; however, there is much more probability mass to the right than in the case of $b_1 = 0.5$. In addition, there is little probability mass for the negative region, making this case less realistic than when $\alpha = 0.1$.

In conclusion, for low $\alpha$ and high $b_1$, there is behavior which resembles at least one of the stylized facts observed in the financial markets, namely fat tails (properties having to do with autocorrelation cannot be studied in equilibrium). This also makes sense since large $b_1$ is indicative of explosive behavior (the optimistic chartist chasing ever higher prices), while in the low $\alpha$ regime the system spends more time in a state where one of the two types is dominant, increasing the likelihood of explosive behavior. There is also a hint of bubbles and crashes since there is some asymmetry between the tails (i.e. the price has a larger probability of being positive, but the tail seems slightly fatter in the negative region).
5 Estimation

Since it is the aim of this thesis to explain the stylized features that are observed in the real world, it is informative to test the model using stock market data. For estimation purposes the framework outlined in [Abundo, 2009] is used. After modifying the 2D discrete system to make it suitable for estimation purposes, the estimation method is discussed. Then we show what data is used and the results are shown and discussed. Finally, we simulate the model using the estimation results as input.

Because only data with respect to prices is available, the discrete system, which was obtained earlier, has to be rewritten. First, the market equation (12) can be solved for \( W_t \) to give

\[
W_t = \frac{(1 + r)X_t - a_2 - b_2X_{t-1}}{a_1 - a_2 + (b_1 - b_2)X_{t-1}}
\]

Note that since \((1 + r)X_t\) is a weighted sum of \(a_1 + b_1X_{t-1}\) and \(a_2 + b_2X_{t-1}\), it must hold that if \((1 + r)X_t > a_2 + b_2X_{t-1}\), then \(a_1 + b_1X_{t-1} > a_2 + b_2X_{t-1}\) and vice versa. In the original equation this is automatically taken care of because \(W_t\), as a driving process, remains between 0 and 1. In this case, the constraint has to be satisfied by hand, so in this model it is required that

\[
0 \leq \frac{(1 + r)X_t - a_2 - b_2X_{t-1}}{a_1 - a_2 + (b_1 - b_2)X_{t-1}} \leq 1
\]

for every \(t\). The aforementioned expression for \(W_t\) can now be inserted in the discretized SDE to give one equation for the dynamics of \(X\) (where \(\hat{\alpha} = \alpha h\))

\[
\frac{(1 + r)X_{t+1} - a_2 - b_2X_t}{a_1 - a_2 + (b_1 - b_2)X_t} = (1 - \hat{\alpha}) \frac{(1 + r)X_t - a_2 - b_2X_{t-1}}{a_1 - a_2 + (b_1 - b_2)X_{t-1}}
\]

\[
+\hat{\alpha} \epsilon_t \left( \frac{(1 + r)X_t - a_2 - b_2X_{t-1}}{a_1 - a_2 + (b_1 - b_2)X_{t-1}}, X_{t-1} \right) + \sqrt{2 \left( \frac{(1 + r)X_t - a_2 - b_2X_{t-1}}{a_1 - a_2 + (b_1 - b_2)X_{t-1}} \right) \left( 1 - \frac{(1 + r)X_t - a_2 - b_2X_{t-1}}{a_1 - a_2 + (b_1 - b_2)X_{t-1}} \right) \Delta B_t}
\]

\[
\Leftrightarrow X_{t+1} = \mu(X_t, X_{t-1}) + \sigma(X_t, X_{t-1}) \Delta B_t
\]

where

\[
\mu(X_t, X_{t-1}) = \frac{1}{(1 + r)}(a_2 + b_2X_t) + (1 - \hat{\alpha})A(X_t, X_{t-1})((1 + r)X_t - a_2 - b_2X_{t-1}) + \hat{\alpha} \frac{(a_1 - a_2 + (b_1 - b_2)X_t)e^{-\beta(X_t-a_1-b_1X_{t-1})^2}}{(1 + r)(e^{-\beta(X_t-a_1-b_1X_{t-1})^2} + e^{-\beta(X_t-a_2-b_2X_{t-1})^2})}
\]

(16)

\[
\sigma(X_t, X_{t-1}) = A(X_t, X_{t-1}) \sqrt{2(a_1 + b_1X_{t-1} - (1 + r)X_t)((1 + r)X_t - a_2 - b_2X_{t-1})}
\]

32
where

\[ A(X_t, X_{t-1}) = \frac{1}{1 + r a_1 - a_2 + (b_1 - b_2) X_{t-1}} \]

is a factor that is close to 1 if the interest rate is close to zero and \( X_t \) is close to \( X_{t-1} \). Note, furthermore, that \((X_t, X_{t-1})\) is a 2D Markov discrete time Markov process. In addition, \( X_{t+1} \) conditionally on \( X_t \) and \( X_{t-1} \), is normally distributed with mean \( \mu(X_t, X_{t-1}) \) and variance \( h(\sigma(X_t, X_{t-1}))^2 \).

Now, since the conditional probability density is known, Maximum Likelihood Estimation (MLE) can be used to estimate the model. In addition we have to deal with the dynamic constraint

\[ 0 \leq \frac{(1 + r)X_t - a_2 - b_2 X_{t-1}}{a_1 - a_2 + (b_1 - b_2) X_{t-1}} = g(X_t, X_{t-1}) \leq 1 \]

so that the fractions are well defined.

### 5.1 MLE

In MLE one wishes to find the parameters that maximize the likelihood of obtaining the sample. Denote the observations by \( X_1, ..., X_N \). The joint density can then be written as (ignoring the dependence on parameters for now)

\[
    f_{X_1, ..., X_N}(x_1, ..., x_N) = \prod_{t=2}^{N} f_{X_{t+1}|X_t, X_{t-1}}(x_{t+1}, x_t, x_{t-1}) f_{X_2|X_1}(x_2, x_1) f_{X_1}(x_1)
\]

where it was used that \((X_t, X_{t-1})\) is Markov and

\[
    f_{X_{t+1}|X_t, X_{t-1}}(x_t, x_{t-1}) = \frac{1}{\sqrt{2\pi h(\sigma(x_t, x_{t-1}))}} e^{-\frac{(x_{t+1} - \mu(x_t, x_{t-1}))^2}{2h(\sigma(x_t, x_{t-1}))^2}}
\]

The last two factors in the likelihood (i.e. the first two observations) should be discarded since they give no information regarding the parameters (i.e. there is no model for them). Inserting the data, this expression can be maximized with respect to the parameters subject to the inequality constraint. Note that the number of inequality constraints is \( 2N - 2 \).

The log-likelihood function based on this density is

\[
    \log L(\theta) = \sum_{n=2}^{N-1} - \log(\sqrt{2\pi h(\sigma(x_n, x_{n-1}))}) - \frac{(x_{n+1} - \mu(x_n, x_{n-1}))^2}{2h(\sigma(x_n, x_{n-1}))^2}
\]

where \( \theta \) denotes the parameters. It is also required that \( 0 \leq g_n(\theta) := g(X_n, X_{n-1}) \leq 1 \). The optimization problem is therefore \( \max_{\theta} \log L(\theta) \) subject to \( g_n(\theta) - 1 \leq 0 \) and \( -g_n(\theta) \leq 0 \) \( \forall n \).

The number of constraints in addition to the fact that they are nonlinear, makes the estimation problem difficult.

However, the constraints do not depend on \( \alpha \) and \( \beta \), so it might be possible to obtain analytical expressions for these estimators. It turns out that this is only possible for \( \alpha \). We proceed to compute this estimator.
The equation for the mean (16) can be written as

\[ \mu_n := \mu(X_n, X_{n-1}) = D_n + E_n + \alpha h (F_n - E_n) \]

where \( D_n = 1/(1+r)(a_2 + b_2 x_t) \), \( E_n = A(X_t, X_{t-1})((1+r)X_t - a_2 - b_2 x_{t-1}) \) and

\[ F_n = \frac{(a_1 - a_2 + (b_1 - b_2) x_t) e^{-\beta (x_t - a_1 - b_1 x_{t-1})^2}}{(1+r)((e^{-\beta (x_t - a_1 - b_1 x_{t-1})^2} + e^{-\beta (x_t - a_2 - b_2 x_{t-1})^2})} \]

are quantities that do not depend on \( \alpha \). The log-likelihood for \( \alpha \) alone can then be written as

\[ \log L(\alpha) = \sum_{n=2}^{N-1} \log(\sqrt{2\pi h} \sigma(x_n, x_{n-1})) - \frac{(x_{n+1} - D_n - E_n - \alpha h (F_n - E_n))^2}{2h(\sigma(x_n, x_{n-1}))^2}, \]

where \( \sigma_n^2 = (\sigma(x_n, x_{n-1}))^2 \). We then get that

\[ \frac{d \log L(\alpha)}{d\alpha} = \sum_{n=2}^{N-1} \frac{(x_{n+1} - D_n - E_n - \alpha h (F_n - E_n))}{\sigma_n^2} (F_n - E_n), \]

which can be solved for \( \alpha \) to get

\[ \hat{\alpha} = \frac{\sum_{n=2}^{N-1} (x_{n+1} - D_n - E_n)(F_n - E_n)1/\sigma_n^2}{h \sum_{n=2}^{N-1} (F_n - E_n)^21/\sigma_n^2}. \] \hspace{1cm} (17)

Note that \( \hat{\alpha} \) is inversely related to \( h \) in a simple way. This makes sense since \( \alpha \) and \( h \) both determine the velocity of the drift. However, \( h \) also affects the volatility term, unlike \( \alpha \), so there is no redundancy in the parameters.

Now, it still has to be shown that \( \hat{\alpha} \) is a maximum for every value of the parameters and data. The second derivative of the log-likelihood is

\[ \frac{d^2 \log L(\alpha)}{d\alpha^2} = -h \sum_{n=2}^{N-1} \frac{(F_n - E_n)^2}{\sigma_n^2} < 0, \]

since \( h > 0 \). Therefore \( \log L(\alpha) \) is strictly concave, so the estimator defined by (17) is indeed a maximum. Note that there is one restriction on \( \alpha \), namely \( \alpha > 0 \). However, as the log-likelihood is concave in \( \alpha \), if it is the case that the estimator is negative, then \( \alpha \) should be chosen positive but as small as possible (i.e. effectively equal to zero).

For \( \beta \), there is no analytical solution. Regular numerical methods, such as Newton-Raphson, do not apply since the log-likelihood also depends on other parameters which are unknown. A simple approach that does not use gradient information is a method based on a grid. First, from the 4D space (i.e. \( a_1, a_2, b_1 \) and \( b_2 \)) a number of points are extracted such that the resulting collection of points resembles a grid. Then it is checked which of these points satisfy the \( N - 1 \) constraints. Then, to each of these feasible points, multiple values of \( \beta \) are associated. Finally, the maximum is computed from the remaining this set of points. Many different ranges for the grid have been tried to come up with the smallest grid possible (so that the precision is higher).
5.2 Data

To estimate the model data is needed. Since only one risky asset is analyzed, it seems best to take a financial asset that proxies for as many of these assets as possible, i.e. a stock index. The S&P is one of the most important indices in the world, so it is suitable for the investigation.

In terms of time scale, daily closing data has many advantages. Firstly, they are easy to get. Secondly, in studying stock prices over a couple of years, daily data already gives sufficiently many observations. Thirdly, when getting intra day data the underlying market micro-structure (the way in which orders are executed etc.) has a significant impact. Finally, daily data is frequent enough with respect to the Ant Process. At every time step an agent changes its mind independently or meets with another agent; clearly, with many traders, this happens every day, so monthly data are not as realistic in this respect.

There is also the problem of determining the fundamental price, which has to be subtracted from the data to obtain $X_t$. It is assumed that the fundamental value is constant over the period of eight years, as in the BH model. The fundamental value $p^*$ is equal to $\hat{y}/r$. Therefore, the average dividend payment needs to be extracted from yearly dividend data in addition to the interest rate, which is assumed to be constant. Once the fundamental value is obtained, the prices in deviation from the fundamental can be computed. The average of interest rates over the period will be used as a constant interest rate. In [Boswijk et al., 2007] the BH model is studied empirically with over a century of data. They use a non-stationary dividend process, which makes sense for such a long time interval; this is not as important in our case.

More specifically, daily adjusted closing prices of the S & P 500 will be used from 02-01-04 to 31-12-10, obtained from Yahoo Finance. The average dividend and constant interest rate are computed using an average of monthly observations over this time period, using monthly data from Shiller's website. The data is shown in figure 7. In addition, for out-of-sample prediction the same data, but then from 03-01-2011 to 10-03-2014 is used.
5.3 Results

As mentioned earlier, $\alpha$ scales inversely with $h$, while $h$ represents the time-scale of the discretized SDE. It is not clear a priori what the value of $h$ should be such that the discretized SDE and the market equation are synchronized. The data can give an informed guess on what the value of $h$ should be for estimation. We found that $h = 1$ gave the most intuitive results for $\alpha$. In addition, a large value of $h$ (compared to $h = 0.01$, which was originally used for simulation purposes) means there are more fluctuations in the price. Looking at the plot of the data, there are more fluctuations in the price than in the simulated paths, shown in Subsection 4.1. Therefore, increasing $h$ might be advantageous on this front as well. Another solution for that problem might be to add a noise term to the market equation, but this has huge consequences for the estimation procedure, since there would then be a second source of stochasticity in the model.

The results are shown in the table below. The values of the parameters in the forecasting rules are intuitive: the ‘trend-following’ coefficients (i.e. $b_1$ and $b_2$) consist of one smaller than 1 and one larger than 1, indicating two different regimes; the ‘bias’ coefficients (i.e. $a_1$ and $a_2$) are both negative and positive. As expected, $\alpha$ takes a value smaller than 1. It is unexpected that $\beta$ takes the value 0. We have tried many orders of magnitude for the range. Since the prediction
errors generally take values of the order $-100^2$ to $-1000^2$, a small value for $\beta$ was expected, but even when sampling in the range $[0, 10^{-6}]$, the algorithm picked $\beta = 0$. Perhaps this would change if an error term was added to the market equation. The p-values were obtained using the inverse information matrix. The parameters are all significant (at a 5-% significance level), except for $\beta$. As is argued in [Hommes and Wagener, 2009] this is because large variations in this parameter cause only small changes in the Gibbs probabilities and is not a bad result if there is enough heterogeneity in the estimated model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$a_1$</th>
<th>$b_1$</th>
<th>$a_2$</th>
<th>$b_2$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-200</td>
<td>1.20</td>
<td>133</td>
<td>0.933</td>
<td>0</td>
<td>0.654</td>
</tr>
<tr>
<td>P-value</td>
<td>0.043</td>
<td>0.022</td>
<td>0.048</td>
<td>0.024</td>
<td>0.38</td>
<td>0.03</td>
</tr>
</tbody>
</table>

It is also interesting to see what the fractions look like over time, given the estimated parameters. As they are a function of the data and the parameters, they are in a sense also estimated. The fractions are shown in 8. The fractions do not show the same behavior as in simulations of the ordinary Ant Process (or the general model, which is not depicted in the thesis but shows even more polarized behavior). This may be caused by the same problem as the one that causes $\beta$ to be equal to zero. Generally, flexibility of the parameters $a_1$, $a_2$, $b_1$ and $b_2$ is lost due to the constraint that the market equation is strictly satisfied for all data points.

![Figure 8: Estimated fractions (02-01-04 to 31-12-10)](image_url)
5.4 Estimation-based simulation

From a prediction point of view, it might be interesting to simulate the model using parameters obtained from the estimation. As was discussed above, the estimates were not completely realistic due to ignored noise in the market equation. We proceed by adding noise to this equation and changing the parameters somewhat by making an educated guess of how the estimation result would change if noise would be added to the market equation. After simulating $M$ times, it will be checked whether the actual price fell within two standard deviations of the mean. In addition, a Monte Carlo (MC) simulation will be performed on the same $M$ simulations to compute an estimate of the kurtosis, skewness, and autocorrelation (AC; both for the normal and absolute returns).

For the simulation, we used the parameter-values obtained from the estimation procedure, but slightly modified. First, the bias coefficients are probably too large since they have to capture the huge decrease in $7$ (if noise would be included such a fall could, partly, be seen as a large error). Therefore, these coefficients are both divided by 2. Secondly, since setting $\beta = 0$ in the simulation yielded explosive results, we decided to set it unequal to zero, namely $\beta = 0.1$. The remaining parameters ($\alpha$, $b_1$, and $b_2$) were unmodified. In addition, we used a noise term with standard deviation equal to 30 and $h = 0.01$. It might seem strange to use $h = 0.01$, while the estimation results were obtained using $h = 1$, but this has to do with the difference between estimation and simulation: in simulation, the fractions sometimes jump outside of $I$, so a small $h$ is needed, while no such thing occurs in estimation. Furthermore, the standard deviation of the noise was chosen on the basis of eye-balling. Finally, as initial conditions we used the price on the first day, $X_1$, and $W_2$ (which was computed from $X_1$ and $X_2$, using the market equation). A typical simulation run with these parameters is shown in 9.
Figure 9: Typical simulation run

Clearly, this realization is different from the actual price path, but it does show some of the qualitative features: mostly, the system is in a low-volatility regime, but it occasionally undergoes a transition to periods of extreme volatility. Two clear differences stand out. Firstly, the actual data seem non-stationary, even in this short period. Secondly, it seems there is not as much asymmetry between large decreases and large increases (i.e. the bubble and crash stylized fact).

With these parameters, the model is run $M = 10000$ times. We then obtain several MC estimates and, for each time point, a mean and standard deviation. The results of the MC are shown in the table below. Note that the standard deviation is the standard deviation of the mean, which is the standard deviation of the sample divided by $\sqrt{M}$. Note that the AC in the absolute returns is clearly larger than the AC in the returns, hinting at absence of autocorrelation, but presence of volatility clustering, although the difference is not very large. In addition, kurtosis is far above 3, so clearly this distribution has fat tails. Finally, skewness is negative indicating an asymmetry between positive and negative returns. Note that the estimates have small standard deviations, so that they are relatively stable. Clearly, the model shows some of the stylized facts observed in financial markets.

| Quantity | $AC(\Delta X)$ | $AC(|\Delta X|)$ | Kurt | Skew |
|----------|----------------|------------------|------|------|
| Estimate | 0.078          | 0.13             | 11   | -1.4 |
| StDev    | 0.0010         | 0.0016           | 0.077| 0.016|
The mean and standard deviations for each time point are shown in Figure 10. In the beginning the actual data fall within the bounds, but as time progresses it moves out of them. Again, this indicates that the actual process is non-stationary, while the model is not. Clearly, this is a severe limitation when trying to explain actual data. Also note that the bounds are quite wide (±300), which is reminiscent of fat tails.

Figure 10: Out-of-sample prediction (03-01-2011 to 10-03-2014)

6 Conclusion

In this thesis a stochastic heterogeneous agent model for a financial market was constructed that integrated two modeling frameworks. It was investigated whether this model explains certain stylized facts observed in these markets.

It was found that, for certain parameter values, the model shows most of the stylized facts, both in its dynamics and its equilibrium distribution. With regard to estimation, there were some problems due to the rigidness of the market equation (i.e. without an error term). A clear cut solution to this problem is to add an error term, but this makes the estimation problem more difficult. Such an investigation is left to future research.

In addition, in the real world interest rates are not constant and dividends are not stationary. It was clear that the stationary model could not explain the non-stationary data very well, even over a short time period. The model could be improved to incorporate these features.
Furthermore, the bubble and crash phenomenon might be captured better with different forecasting rules, such as non-affine forecasting rules. For instance, one can imagine using a forecasting rule that is piecewise linear with a kink at zero (the negative linear piece is then steeper than the positive linear piece, so that heavier crashes are expected).

Finally, it appeared to be difficult to merge a discrete time model and a continuous time model. Perhaps, the discrete time model is best (i.e. the one used for estimation and simulation), since this captured the dynamics of the prediction errors, while the continuous time model failed in this respect.

It is often difficult to construct models of adaptive complex systems that stay valid for long periods, since such models should be 'fundamental' enough to capture the underlying mechanism by which such a system evolves. This is especially true for financial markets, since participants in these markets might actually use the models designed to describe them. As a result, there is a never ending cycle where markets adjust to new models and models are improved when new data is available. In that light, even a model that explains stylized facts well is only a temporary achievement.
References


