Stock Returns Distributions: The Emergent Behaviour of a Multi-Agent Artificial Stock Market

McGregor J. Collie
School of Computer Science and Software Engineering
Monash University, Wellington Road, Vic. 3800, Australia
gjc@mail.csse.monash.edu.au +61-3-9905 9182

Abstract

We present an artificial stock market in which simple trading agents enter an asynchronous double auction market to trade in a single stock. Beginning with a population of random trading agents drawing their bid prices from a normal distribution around the current price, we compare the statistical properties of the emergent stock price return distribution to that observed on a real price series, the daily returns distribution for GE stock listed on the NYSE.

We then introduce simple technical trading agents to this market, and use a genetic algorithm to evolve a population mix of agents, using the Kullback-Leibler (KL) distance to the GE distribution as our fitness function. We show that the agent populations that produce the most realistic returns distribution show predominance of non-random, rule based trading behaviours.

1 Introduction

Artificial agent based stock market simulations have experienced an explosion of interest since the pioneering work of John Holland and Brian Arthur at the Santa Fe Institute in the early 1990s[2, 13, 21, 18]. They have frequently been applied to exploring the assumptions of the Efficient Markets Hypothesis, or EMH. Central to the EMH is the notion that the stock price determined by the collective opinions of market participants should be arbitrage free, and that a sequence of stock prices should have no profitable forward predictive power[9, 10], a viewpoint that is not without criticism [5, 19, 8, 23].

The price sequence generated by an efficient market should follow the path of a random walk, i.e. the increments should be independent, identically distributed (iid) normal random variates. If this were the case then the returns distribution one would expect to observe is the normal, or Gaussian distribution. in reality, stock returns distributions are known to be non-Gaussian[7, 11], and to have distinctive features common across instruments, companies, markets and time scales. Attempting to generate these features from the interaction of traders with different beliefs, expectations and trading strategies has provided fertile ground for artificial agent based simulations (see Yang's 1999 draft paper [26] for a good review and discussion).

In this study we use the real returns distribution of GE stock prices from the NYSE over the period 8 Jul 68 - 9 Oct 01 (8382 data points) obtained from the yahoo finance website as our comparison 'real' returns distribution.

1.1 Random Trading

Efficient markets, if they exist, should consist of stocks which are priced so as to accurately reflect the future earnings potential of the underlying asset based upon all currently available information[9]. The direction of past price changes should hold absolutely no indication as to future price changes, i.e. the expected return on any stock should be zero, a result that predates the formal statement of the EMH, first appearing in the work of Bachelier in 1900[3].

Consider that at each instant in the market, traders are forming different opinions about the future earnings potential of a share. Bénassy-Quéré et al. show that even where traders are equally well informed, they can form quite different expectations[4]. We see that the efficiency of the market, if it exists, must be brought about by the aggregate effect of all the traders.

According to the EMH, each individual active trader experiences a random sequence of wins and losses in their trading decisions. This is because, if all shares are efficiently valued, then at any point in time there is no profitable arbitrage opportunity within the market. Share prices will only change in response to unpredictable future events, or, to put it another way, traders will only trade in response to previously unpredictable information becoming known. Since a random process will be generating their trades then the returns distribution from such trades should be at least as well modeled by a random trading process than by any other process.

1.2 Technical Trading

Technical trading strategies are those that attempt to use the history of past prices to predict future prices. According to the *weak form* (the least restrictive form) of the EMH, past prices have absolutely no value in terms of predicting future prices[9]. The current price is simply the mean value about which future prices may fall, and prices previous to the current price provide no indication as to future prices. A 2002 paper by Kwon and Kish[17] presents a simple set of technical trading strategies, and shows that they return profits above a simple buy-and-hold strategy over a given 34 year period on the NYSE. A good review of technical trading strategies is presented in Lo, Mamaysky and Wang[20].

1.3 Behavioural Finance

Behavioural finance is a branch of economics that attempts to explain market behaviour as the result of human behaviour. It is a direct competitor to the efficient markets hypothesis (EMH) in that it does not propose that the market is efficient, but rather that it is simply the result of human behaviour, and as such is laden with inherent biases, such as those documented by Tversky and Kahneman[25]. Effects described by behavioural finance typically include herd-like behaviour, over and under reaction to news events, and overtrading. For a good review the reader is referred to Thaler[24]. Hong, Kubik and Stein[14] report evidence to indicate that stock market participation is seen at least in part a social activity, and not purely a wealth generating exercise.

1.4 Related Market Simulations

Similar stock market simulation with random traders include the Genoa artificial stock market due to Raberto et al. [22] and Gode and Sunder [12], and other work on emergent properties of interacting agents is due to Shu-Heng Chen and others, see in particular Chen and Yeh[6].

Gode and Sunder describe a simulation with traders who trade almost randomly, but are subject to a budget constraint on the prices they issue bids at which they show causes traders to allocate their assets highly efficiently.

1.5 Stylized Facts

Real returns distributions contain similar properties for a wide range of markets, trading instruments and time periods. These properties are referred to as *stylized empirical facts*, the most famous of which are arguably the high kurtosis and fat tails (relative to a normal distribution) of the empirical distribution. For more information on the nature of real observed stock return distributions, and efforts to model them, we refer the reader to the review by Farmer [11] and the recent study by Cont [7].

1.6 An Agent Based Trading Simulation

If we want to show that the stylized facts that typify a stock market returns distribution can be explained somewhat by the prevalence of certain trading rules and theories amongst the trading public, then an agent based simulation presents itself as an obvious way to do this. Here we use an artificial agent based simulation to show the unlikeliness of the EMH as a realistic description of the market. This is done in two ways.

The first is to show that agents trading randomly will, under simple realistic constraints, generate returns distributions that appear to be moving away from Gaussian, and the second is to show that by introducing agents with fixed strategies into the trading environment, we begin to generate returns distributions that look more like the real thing.

1.7 Genetic Evolution of Agent Populations

If the random market hypothesis is true, then a model of traders as random trading agents is as at least as good a model as any other. If however there are trading strategies that do influence the market, then a model of random traders will not be the best description.

Here we construct a market filled with populations of different kinds of trading agents, including randomly trading agents. By evolving the population sizes, and keeping the total number of agents, and the total available cash fixed, we can arrive at the population mix that provides the returns distribution closest to the comparison GE distribution, for the set of agents being considered.

We find that by using the genetic algorithm we converge to a population mix of trading agents which provides a relatively good fit to the empirical GE distribution. The existence of such a state, in which rule-based agents form a percentage of the final population, shows that even with a set of simple trading agents we are able to construct markets with realistic features.

The *Kullback-Leibler* (KL) distance[16] is a means of measuring the 'distance' from one probability distribution to another. The KL distance from distribution p to distribution q is defined as

$$KL(p,q) = \sum_i p(x_i) * \log(\frac{p(x_i)}{q(x_i)}). \quad (1)$$

This quantity is always positive, and is minimized for the case $p(x_i) = q(x_i) \forall i$.

Here we use the Kullback-Leibler distance to compare the returns distribution generated by different populations of interacting trading agents to both normal and observed distributions.

2 Experimental Setup

2.1 Market Microstructure

Our virtual trading environment is that of a single stock traded on a continuous double-auction market with no transaction costs, and unfettered and equal access to the same information. The market moves in steps which we hereafter refer to as trading rounds, which do not necessarily correspond in breadth or duration. Multiple times within each trading round a trading agent is selected at random from the trading agent pool, allowed to trade, and then returned to the pool.

Traders submit sealed bids continuously to an asynchronous double auction at fixed volumes, with sellers stating a minimum sell price and buyers a maximum buy price. If bids can be matched with any existing orders then they are matched immediately, such that the lowest priced existing sell orders or the highest priced existing buy orders are cleared first. Note that since bids are matched immediately, then in any one trading round there may be multiple strike prices. Indeed, within the submission of a single bid there may be multiple strike prices as qualifying bids are cleared (although no further bids may be submitted until the current bid has either cleared or is set in the auction list)[1]. For a further discussion of the double auction in the context of artificial stock markets, the reader is referred to the recent work of Hsu and Soo [15].

2.2 Design of the Genetic Algorithm

For the genetic evolution of the agent populations in our market, we have written a wrapper program to run instances of the artificial market with different trading populations. We implement this in the wrapper program with an array of v vectors (I use 16 in practice) of d dimension, containing randomly chosen initial agent populations. i.e. each vector $\tilde{V} = p_1, p_2, .., p_d$, where p_i is the population of agents of the i_{th} type, where $\sum_{i=0}^{d} p_i = constant$. The output of the market simulation is fed into a third program,

[1] note that if two bids exist at the same price, the 'older' bid will be cleared first

which calculates the returns distribution, its corresponding Gaussian approximation, and calculates the KL distance from the generated distribution to both the GE distribution and the Gaussian approximation. The initial wrapper program reads in these results, and uses them as a fitness function to rank the v population vectors. The best performing population vector is then selected, as is another randomly selected vector, and these two are blended together to form a third vector (crossover), which replaces some other vector, excepting that vector which performed the best. The vector that is replaced is selected with probability proportional to the size of its KL distance. The process is completed until improvement is not noted for v generations (i.e. for v^2 separate runs of the market simulation).

The process by which crossover occurs is to take a simple average of the two vector elements. I.e. if $\tilde{V}_3$ and $\tilde{V}_5$ are selected to produce a new vector $\tilde{V}_{new}$, then

$$\tilde{V}_{new} = ((p_1^3 + p_1^5)/2, .., (p_i^3 + p_i^5)/2, .., (p_d^3 + p_d^5)/2), \tag{2}$$

where p_i^j refers to the i_{th} agent type's population in the population mix vector $\tilde{V}_j$, the j_{th} population mix vector.

The averaging in equation 2 is done by rounding the average value down. After the crossover has taken place, the total number of agents in the vector is calculated. Where this number is less than the total number of agents needed to keep the population constant overall, the difference is added back in a random way, introducing a random mutation process into our genetic algorithm.

2.3 Randomly Trading Agents

In our simulation we start initially with trading agents whom we attempt to have behave as randomly as possible. Subject only to cash constraints and an expected log normal return distribution for stock prices changes, these agents do not follow any trading rules. When a random agent is selected in a trading round, the trader chooses to buy, sell or hold with equal probability. If an agent chooses not to hold then they sample for a random price. This same price is to be used whether the trader decides to submit either a buy or sell order. The sampling for a price is done by the following method: For the last strike price $S(t)$, the random agent wishes to determine a new price,

$$\hat{S} = S(t) * \exp(N), \tag{3}$$

where N is some normally distributed value with $\mu = 0$, and for ease of modelling $\sigma = 1/100$. We expect intuitively that such agents may well generate a log normal stock return distribution, since

$$\log \frac{\hat{S}}{S(t)} = N, \tag{4}$$

so any deviation from this should be a consequence of exogenous constraints.

If the agent wishes to buy, we denote the percentage of their cash holdings that they wish to spend as m. The random buyer then calculates a volume, $v = int(\frac{m}{\hat{S}})$, and submits their bid. When the seller decides to sell shares, they determine a maximum realisable price for their shares, i.e. $\hat{S} * volume$. They then randomly determine a percentage of this total sum that they wish to realise in the sale, m. This amount mirrors the random percentage of cash that selling agents wish to spend. The volume is then determined in the same way as with the buyers, i.e. $vol = int(\frac{m}{\hat{S}})$, creating a balanced market.

2.4 Rule-Based Trading Agents

Into our random trading environment we introduce traders who follow simple technical trading rules. We introduce **momentum traders** which record the current price when they are active until they have a history of five previous prices. When they have collected these five prices they then calculate the gradient of the prices as though they were all collected at fixed intervals apart, and use the figure arrived at to make a trading decision. If the value is higher than some threshold value then they will buy shares, if is below some lower bound then they will sell.

Stop-loss traders behave initially like random traders but differ once they have purchased shares in that they will attempt to cut their losses if prices fall beyond a certain level, and will take profits if the share price rises above their target values. These traders also submit bounded random prices, that is, they will submit a minimum strike price less than or equal to the current price if they are trying to cut their losses, but will submit a random (normally distributed about the current price) priced bid when buying and when taking profits.

2.5 Social Trading Agents

Following on from ideas present in the work of Hong, Kubik and Stein, [14] we present also the **social trader** agents, who trade with far greater probability if they know that others are trading. They are allowed limited communication with each other, and will tend to trade in the same direction as each other. These agents go some way to modelling *herding behaviour* (see Anderson and Holt [1] for a review), and, we hope, will induce *volatility clustering*, another stylized fact of real returns distributions.

These types of trading agents do not necessarily follow the trend of the markets, as the technical traders in the previous section try to, but rather the trading decisions of each other. All social trading agents belong to a *social club* to which they report news of their last trade decision. When a social trader is selected in a trading round, they consult with their social club on the stock they are considering, and are returned the percentage of trades (n.b. not the percentage of traders) which were submitted as buy orders. If the social club reports that P percent of all shares offered for trading by social club members in that round were buy (sell) orders, then the trader will decide to buy (sell) with probability $p(buy(or\ sell)) = 0.8 * P + 0.1$.

3 Results

The KL distance[2] between the empirical distribution obtained from the daily GE price history (*gedist*) with an assumed Bayesian prior and a Gaussian renormalised over the same range with the same standard deviation and mean gives a figure of

$$KL(gedist, N(\mu_{ge}, \sigma_{ge})) = 1.357. \tag{5}$$

ge subscripts here refer to the statistics of the GE stock price return distribution. Note that for the distributions generated below, points are included in the underlying series only where a trade has been executed in a round.

3.1 Random Agents

3.1.1 Agents with Infinite Resources; *randncc* and *gerandncc*

If we introduce randomly trading agents into our double auction bid matching process and examine the resultant distribution (which we label *randncc*) over 50,000 trading cycles, we find it follows a highly Gaussian form, as can be seen in Figure 2. The KL distance between *randncc* and its descriptive Gaussian is

$$KL(randncc, N(\mu_{randncc}, \sigma_{randncc})) = 0.00717285. \tag{6}$$

If this agent type is modified to draw new bid prices from the Gaussian approximation of the GE distribution *gegauss*, then we expect the resultant distribution to be much closer to *gedist*. We call this new distribution *gerandncc*. The KL distance between *gerandncc* and the empirical GE distribution, *gedist* is

$$KL(gerandncc, gedist) = 0.183387. \tag{7}$$

.

3.1.2 The Cash Constrained Return Distributions; *ccdist* and *geccdist*

Next we introduce cash constraints to the random trading agents. Using the same distribution as *randncc* to draw their bid prices from, we allow 100 agents with equal initial quantities and values of cash and shares (100) to trade for up to 55,555 trading 'days' of 480 time steps each. We denote the resultant distribution *ccdist*. In Figure 1 we can see the probability distribution of the log of price changes with a monetary constraint (*ccdist*), compared with a Gaussian distribution of identical mean and standard deviation. The simulation data appears more peaked, and appears to have fatter tails than the Gaussian.

Over a sample of 80,000+ trading cycles we generate a returns distribution with a KL distance to a Gaussian with the same standard deviation and mean (subscripted *ccdist*) of

[2] Note that all KL distances are given in natural log units of information, *nats*.

$$KL(ccdist, N((\mu_{ccdist}), \sigma_{ccdist})) = 0.0121983. \quad (8)$$

Our cash constrained random agent return distribution is much closer to being Gaussian than is our empirical GE distribution, as we can see by comparison with the KL distance in (5). We can see by comparison with the KL distance for the unconstrained agent returns distribution (6) that the cash constraint has induced some non-Gaussian features in the *ccdist* distribution, increasing the KL distance to its Gaussian approximation.

We then used the standard deviation and mean of our GE sample as parameters for a Gaussian for our cash constrained random traders to draw prices from. This results in a new return distribution (denoted *geccdist*). With a sample of over 50,000 trading rounds we obtain a KL distance to the empirical GE return distribution *gedist* of

$$KL(geccdist, gedist) = 0.177651. \quad (9)$$

We note here that (9) is significantly smaller than (7), i.e. the cash constraint has resulted in a more realistic returns distribution.

3.2 Population Mixtures

I use three different fitness functions in developing agent population mixes. The first is the KL distance from the generated distribution to the corresponding Gaussian, the second fitness function is the KL distance to the GE distribution, and the third fitness function is the ratio of the KL distance to the GE distribution, to the KL distance to the Gaussian. In all cases we attempt to minimise the value.

Firstly, minimising KL to the Gaussian. A set of sixteen population vectors evolved to a final average population of approximately 39% random traders, 20% momentum traders, 10.5% stop-loss traders and 30.5% social traders. As expected, random trading agents dominate in a market that produces a Gaussian returns distribution, even when the agents are acting under a cash constraint. The KL distance for this configuration was found to be slightly higher than that obtained for the case where the population of trading agents consisted solely of random traders (*geccdist* above), clearly identifying a problem in that the GA had identified a local minima. This led to a reconfiguration of the GA to include within the initialization a set of boundary cases, which consisted of including a pure single agent population for each agent type in the search set, as well as random mixtures. This change did not affect any results other than for this special case.

In the case of the KL distance to the empirical GE distribution, the final population mix shows clear differences to the population found attempting to minimise the KL to the Gaussian. Approximately 18.5% random, 11% momentum, 22.5% stop-loss and a dramatic 47.5% social traders emerged as the final mixture. With this configuration we achieved a minimal KL distance of approximately 0.11, which can be compared with (9). Here we see the population of social traders clearly dominating the trading population.

Lastly, in minimising the ratio of KL distances (KL to GE / KL to Gaussian), the smallest ratios I observed were about 1:3 with the agents described in this report. This shows it is possible to develop simple agent populations that generate returns distributions much closer to empirical distributions than to a Gaussian model. The final average population mixtures were approximately 9% random, 36.5% momentum, 45.5% stop-loss, and 9% social traders. Such a mixture is quite different than than the other populations described here, and is most remarkable for the very low population levels for randomly trading and social trading agents.

4 Conclusion

In this paper we have presented a description of a new artificial market simulation, with some preliminary results. We have managed with only simple assumptions and relatively uncomplicated structure to reproduce some of the effects of the real market using our model. By imposing simple monetary constrains on randomly trading agents, we have generated returns exhibiting high kurtosis and fat tails, as is seen in real stock price distributions. In addition, by including a mixture of rule-based trading agents we have been able to generate more realistic returns distributions than is possible with randomly trading agents alone.

We have introduced the KL distance as a fitness function for a GA to evolve a population of trading agents, and have applied this to growing agent population mixes which generate returns distributions which are a better fit to real returns distributions than they are to their Gaussian approximations. We have shown using this technique that populations of trading agents can be evolved which appear to show that real returns distributions are better modelled by populations of trading agents which are not predominantly random in their behaviour. Such a result is in conflict with classical versions of the efficient markets hypothesis, but lends weight to theories of behavioral finance.

We have also shown that without proper boundary conditions, local minima can hide the best solutions, and we have presented a solution to this problem.

We intend to apply our model to evolving populations of more sophisticated agents that better reproduce the stylized facts of real observed returns distributions.

In the randomly trading agent based markets, our results appear to indicate the gradual dominance of trading by a limited number of traders of the market after sufficient time. Such a result is consistent with some Pareto-type wealth distribution theories. In future work we intend to examine the emergence of dominant traders in the simulated marketplace for a better comparison with Pareto-type wealth distribution theories.

Acknowledgements

The author would like to thank his supervisors, Dr. Kevin Korb and Dr. David Dowe, and also Dr. Lloyd Allison for their assistance, suggestions and helpful feedback. In particular he would like to note that the suggestion to use the KL distance as a fitness function, and to genetically evolve groups of agents rather than individual agents were both initially Dr. Korb's suggestions.

References

[1] L. Anderson and C. Holt. *Encyclopedia of Cognitive Science*, chapter Informational Cascades and Rational Conformity. Macmilland Reference Ltd., 2000.

[2] B. Arthur. On learning and adaption in the economy. SFI Working Paper, 92-07-038, 1992.

[3] L. Bachelier. Théorie de la spéculation. *Annales de l'Ecole Normale Supérieure*, 17:21–86, 1900.

[4] A. Bénassy-Quéré, R. MacDonald, and S. Larribeau. Models of exchange rate expectations: how much heterogeneity? www.cepii.fr/anglaisgraph/pagepers/Webabq/Papers/Heterogeneity 0302.pdf, 2002.

[5] Werner F. M. De Bondt and Richard H. Thaler. Further evidence of investor overreaction and stock market seasonality. *Journal of Finance*, 42:557–581, 1987.

[6] S. Chen and C. Yeh. On the emergent properties of artificial stock markets: the efficient market hypothesis and the rational expectations hypothesis. *Journal of Economic Behavior and Organization*, 49:217–239, 2002.

[7] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[8] David L. Dowe and Kevin B. Korb. Conceptual difficulties with the efficient market hypothesis: Towards a naturalized economics. In D. Dowe, K. Korb, and J. Oliver, editors, *Proceedings: ISIS (Information, Statistics and Induction in Science)*, pages 212–223. World Scientific, 1996.

[9] Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.

[10] Eugene F. Fama. Efficient capital markets: Ii. *The Journal of Finance*, 46(5):1575–1617, December 1991.

[11] J. Doyne Farmer. Physicists attempt to scale the ivory towers of finance. *Computing in Science and Engineering (IEEE)*, pages 26–39, 1999. November-December.

[12] K. Gode and S. Sunder. Allocative efficiency of markets with zero intelligence traders: Markets as a partial substitute for individual rationality. *The Journal of Political Economy*, 101(1):119–137, 1993.

[13] J. Holland and J. Miller. Artificial adaptive agents in economic theory. *American Economic Review: Papers and Proceedings*, 81(2):365–370, 1991.

[14] Harrison Hong, Jeffrey D. Kubik, and Jeremy C. Stein. Social interaction and stock-market participation, July 2001. NBER Working Paper No. 8358.

[15] W. Hsu and V. Soo. Market performance of adaptiver trading agents in synchronous double auctions. In S. Yuan and M. Yokoo, editors, *Intelligent Agents: Specification, Modelling and Applications, PRIMA 2001*, pages 108–121, Berlin; London, 2001. Springer.

[16] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[17] K. Kwon and R. Kish. Technical trading strategies and return predictability: Nyse. *Applied Financial Economics*, 12:639–653, 2002.

[18] B. LeBaron. Building markets with artificial agents: Desired goals, and present techniques. In G. Karakoulas, editor, *Computational Markets*. MIT Press, 1999.

[19] Andrew W. Lo and A. Craig MacKinlay. Stock prices do not follow random walks: Evidence from a simple specifications test. *Review of Financial Studies*, 1(1):41–66, 1988.

[20] A.W. Lo, H. Mamaysky, and J. Wang. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, 55(4):1705–1770, August 2000.

[21] G. Palmer, B. Arthur, J. Holland, B. LeBaron, and P. Tayler. Artificial economic life: A simple model of the stock market. *Physica D*, 75:264–274, 1994.

[22] M. Raberto, S. Cincotti, S. Focardi, and M. Marchesi. Agent-based simulation of a financial market. *Physica A*, 299:319–327, 2001.

[23] R. Shiller. *Irrational Exuberance*. Princeton University Press, New York, 2000.

[24] Richard H. Thaler. The end of behavioural finance. *Financial Analysts Journal*, 55(6), November 1999.

[25] A. Tversky and D. Kahneman. Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[26] J. Yang. The efficiency of an artificial stock market with heterogenous intelligent agents. Draft Paper, Economics Department, Concordia University, February 1999.

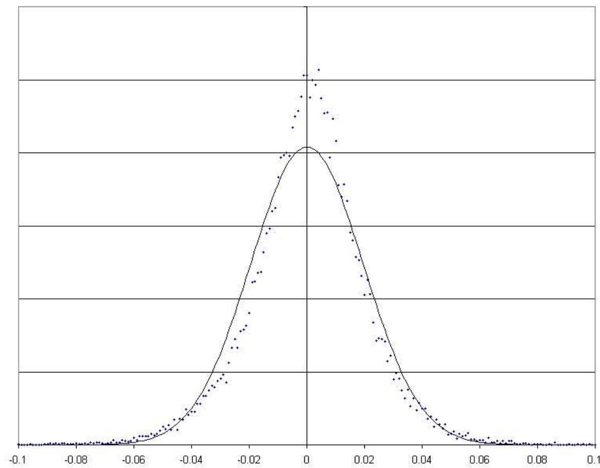

Figure 1: Probability distribution of daily changes generated by monetarily constrained random traders. The smooth curve is a Gaussian with the same standard deviation and mean, for comparison. Note the sharply peaked simulation data, with fatter tails than the Gaussian.

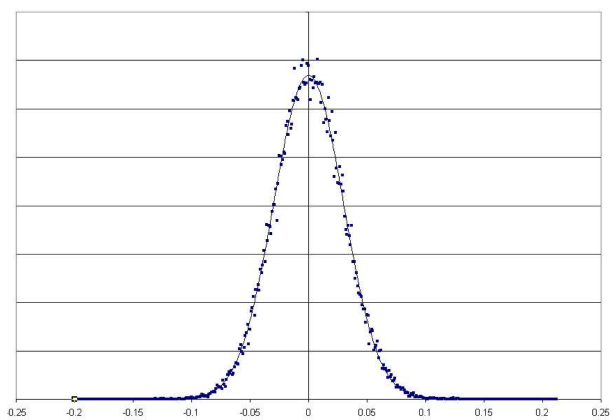

Figure 2: Probability distribution of daily changes generated by random traders with no cash constraints. The smooth curve is a Gaussian with the same standard deviation and mean, for comparison.