Stock Market Simulation and Inference Technique

McGregor J. Collie, David L. Dowe and Leigh J. Fitzgibbon
Clayton School of Information Technology
Monash University, Wellington Road, Vic. 3800, Australia
gjc@csse.monash.edu.au

Abstract

We present an agent-based stock market simulation in which traders utilise a hybrid mixture of common information criteria based inference procedures, including minimum message length (MML) inference. Traders in our model compete with each other using a range of different inference techniques to infer the parameters and appropriate order of simple autoregressive (AR) models of stock price evolution. We show that such traders are initially profitable while a significant population of random traders exist, and that MML inference traders outperform other inference traders in the presence of a noisy AR signal.

1. Introduction

The efficient markets hypothesis (EMH) as popularised by Fama [13] and others (e.g., Jensen [22] and Malkiel [26]) presents us with the claim that the market is 'efficient with respect to [an] information set ... if it is impossible to make economic profits by trading on the basis of [that] information' [22]. This model ignores the behaviour of individual trading agents in the system, relegating them to an arbitrage role in which more efficient traders exploit less efficient traders to keep the market correctly priced, with respect to the information set they are basing their decisions upon. It has also been disputed on various other conceptual grounds [14], including provability in [12], where it is shown that the undecidability of Kolmogorov complexity [36] means that we can rarely prove our inference technique to be superior, and could rarely (if ever) prove a market to be efficient. Agent-based models of stock market phenomena work on the basis of a ground-up approach (see Tesfatsion [33] for a review of agent-based approaches in finance, and LeBaron [23] for a survey of agent-based stock market simulations in particular), in that the macroscopic properties of the market are an emergent property of the individual interaction of trading agents through the microstructure of the auction mechanism [8]. Within the context of heterogeneous agent-based simulation, one can introduce agents that correspond with rational expectations (RE) assumptions, (see, e.g., [6]), as well as agents that correspond to behavioural models of individual economic behaviour[1].

In this simulation we introduce a set of trading agents into an artificial single stock trading environment who attempt to model the evolution of the stock price using an autoregressive (AR) model. Furthermore, the AR agents are divided into subsets of agents who use different *information criteria* (IC) to select an autoregressive model order.

The use of an information criterion (IC) as a means to selecting a parsimonious model to explain observed data is fairly controversial in terms of implementation, if not in principle. Based upon the work of Fitzgibbon, Dowe and Vahid [15] in which various inference techniques are compared on various generated AR signals, we implement within this simulation agents that embody a range of different information criterion based inference techniques, and allow them to compete directly with each other.

The stock market simulation used here is an extension of the artificial stock market presented in Collie [8], in which agents are selected randomly from a trader pool of fixed total size to appraise the market (consider the sequence of past prices) and potentially submit bids to a double-auction process. Other stock market simulations using an agent-based trading methodology include the pioneering work of the Santa Fe simulation [19, 2], the recent Genoese simulation [28] and others, e.g. [25, 21, 7]. See LeBaron [24] and also his website[2] for further references.

The next section outlines the set-up of the stock market simulation in more detail, and presents a closer examination of the information criteria used by the trading agents.

2. Simulation Design

Agents participate in multiple rounds of a continuous double-sided auction of a single tradable asset, submitting

[1]For a review of behavioural finance the reader is referred to Shiller [31] and Barberis & Thaler [3]

[2]http://people.brandeis.edu/∼blebaron/acf/index.htm

buy and sell orders at fixed prices ('at limit' bidding). Unmatched or partially matched orders are submitted to an 'order book', as commonly employed in modern exchanges. New trades are matched against existing orders in the book.

Simulations are run for an exogenously determined number of trading rounds, or until some other termination condition is reached, such as the cessation of trade by the trading agents.

Traders are initially allocated equal numbers of shares and an equal value of virtual currency with which to trade. The total amount of shares and currency within the simulation is held constant throughout, but the total amount of wealth available at any one time fluctuates with the current trading value of the asset. There are 100 traders in total in each simulation, of which 40 are random or AR signal generating, with the remaining 60 divided evenly amongst the 6 inference techniques examined[3].

We present two simulations here, one in which inference traders act in a market with each other and randomly trading 'noise' agents, and one in which the random traders are replaced by a set of traders who calculate future price changes as following an exogenously specified noisy AR process.

2.1. Agent Design

Trading agents participating in these simulations are either of the randomly trading variety, or are one of a number of different types of autoregressive inference trader.

2.1.1 Random Traders

Random trading agents (or 'noise traders') are generally introduced into agent-based stock market simulations as a means of providing market liquidity. Probably the most well known artificial randomly agent trading model is the 'zero intelligence' model of Gode and Sunder [16], where randomly trading agents subject to a 'budget constraint' achieve high allocative efficiency in a double auction.

The random trading agents used in this model differ slightly from those of the previous random traders in [8]. When a randomly trading agent is selected from the trader pool, they clear any existing, previously unfilled orders remaining in the order book, and choose a uniformly distributed random number[4] from 0 to 1. This number is then compared to the trader's current ratio of stocks to cash, and (if necessary) an order is submitted to the market to adjust their current position. The strike price of this new order is drawn from a Gaussian distribution around the last price change. These generated prices are not bounded below or above by the current price, so that (for example) a randomly generated sell order may be submitted at a price greater than the current price.

2.1.2 Noisy and Inferential AR Traders

Autoregressive time series processes for a time series $y(t)$ are of the general form

$$y_t = \sum_{i=1}^{p} [\alpha_i y_{t-i}] + \epsilon_t \tag{1}$$

where ϵ_t is a $N(\mu_t, \sigma_t^2)$ Gaussian i.i.d. error term with average μ_t and variance σ_t^2.

In our second simulation we introduce both a noisy AR signal-generating trader and an AR signal-detecting inference trader. The signal is introduced through replacing the random noise traders of the previous simulation with noisy AR signal generating traders, who generate trade prices using an AR model like that given above in eq. (1).

The order p of the autoregressive function chosen is varied randomly from one to eight when the simulation is initialised, with the parameters (α_i) of eq. (1) chosen not necessarily to guarantee stationarity. When the order p of the model is greater than one, the individual parameters are initially chosen as $A_i = 1.0+\epsilon_i$, where ϵ_i is assumed Gaussian i.i.d. with standard deviation of 0.01, and the final parameters are then normalised; $\alpha_i = A_i/(\sum_{j=1}^{p} A_j)$. For AR models of order one the stationarity condition is imposed, the parameter α is chosen as $N(0.99, 0.01^2)$, subject to α being no greater than one.

In [15] clear differences in inferential power amongst different IC based inference techniques were shown. We introduce groups of different inference technique based trading agents into the second simulation to potentially exploit the noisy AR signal in the price series, and to compare advantages to using different types of inference techniques.

The use of IC in model selection has generally been used as a means of augmenting maximum likelihood (ML) techniques so as to identify not simply the model that best fits the data, but rather the model that best *explains* the data; that is, the most parsimonious model. Such models are generally chosen on the basis of minimisation of model complexity [5], or minimum message (hypothesis and data given hypothesis) length [35, 37, 36]. As noted by Hanlon and Forbes [17], these IC in general take the form

$$-\log(\sigma^2) + Penalty(p, T), \tag{2}$$

where the penalty term is a function of the number of parameters, p, and the sample size, T. The IC above is minimised by the appropriate selection of model order, p.

The four inference trader types in this simulation that use an IC of the form of eq. (2) use: Akaike's information criteria [1] (hereafter AIC), corrected AIC [20](CAIC),

[3] The proportion of the number of random to inference traders is exogenously determined, and is arrived at by attempting to gather enough random traders to provide necessary liquidity, whilst not so many that the emergent properties of the market take too long to appear.

[4] Generated using the *Mersenne Twister* [27]

Schwartz's Bayesian IC [30] (which is here equivalent to the 1978 MDL [29] technique) (BIC), and Hannan and Quinn's information criteria [18] (HQIC).

The IC formalism of eq. (2) does not explicitly take into account factors relating to prior probability of potential model choices (see, e.g., [37, p251], [17, 4], [36, pp279-280]). The Minimum Message Length (MML) formalism of Wallace et al. [35, 37, 36, 34] (see also [32, sec.5] and [10]) differs from the usual informational criteria in that it uses an explicitly Bayesian approach, in that additional terms are included in the information criterion specifying the Bayesian prior distribution over the model parameters. The inclusion of a term involving the determinant of the expected Fisher information matrix captures further information about the appropriate weighting of observed data from different regions of the parameter space ([17, section2]). As outlined in Fitzgibbon, Dowe and Vahid [15], the MML87 IC approximation [37] used in this simulation is described by

$$-\log\left(f(y_1,...,y_N|\theta)\right)-\log\left(\frac{h(\theta)\epsilon^N}{\sqrt{|I(\theta)|}}\right) + \frac{p}{2}\left(1+\log\left(\kappa_p\right)\right)-\log\left(h(p)\right), \qquad (3)$$

where $f(y_1,...,y_N|\theta))$ is the likelihood function of the N observed data points y for model θ. $|I(\theta)|$ is the determinant of the Fisher information matrix, κ_p is a space-quantising lattice constant[5], $h(p)$ is a prior over the number of parameters, p, and ϵ is an estimate of data measurement error.

Parameter estimation for the AR models is done using a standard ordinary least squares (OLS) regression. A more explicit discussion of the OLS technique, the form of the likelihood function and the MML model used here are given in [15] and [9]. We include agents that assume both a stationary and non-stationary process, and model data accordingly.

3. Results

Figures 1 and 2 show average wealth levels for the different classes of traders across 10 simulations, where the results for each individual simulation are averages across each agent class over 150 trading rounds. These figures can be seen more clearly, along with some additional results, in [9]. Within each trading round there are between 3,600 and 8,200 individual potential trading opportunities for each of the 100 agents in each simulation.

In our first simulation inference agents attempt to model an AR time series from the prices generated by the noise traders and their interaction with them. With little or no signal to detect, the wealth levels of the inference traders don't show much variation; they manage to take most of the noise trader's wealth, but their own wealth, affected by the declining price of the stock they are trading, does not increase significantly after an initial, highly volatile trading period. In fig. (1) we can see the average wealth levels for the random (AVG_randtrade) and inference agents (AVG_all_IC), for approximately 150 trading rounds. Smaller jumps in average inference trader wealth later in the trading rounds reflect the transfer of wealth into fewer and fewer inferential agents' control. The type of inference trader that captures most of the wealth appears to fluctuate randomly, and did not show any clear outperformer across many simulations.

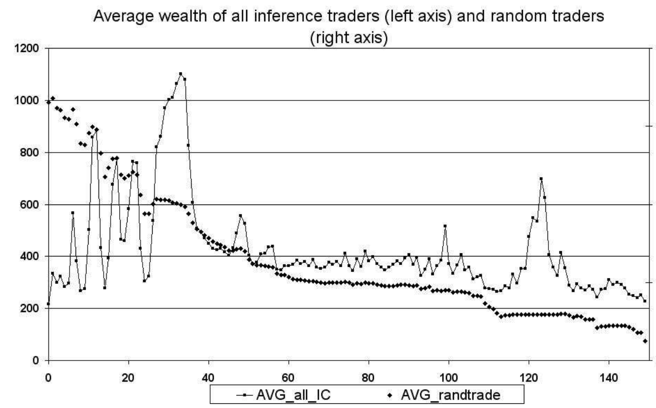

Figure 1. Random noise traders and inference agents average wealth.

In the second simulation with a noisy AR signal being generated, inference technique traders do significantly better than in the presence of only random traders. Each of the 10 individual simulations has a different randomly selected autoregressive model built in to the noisy AR signal traders within it.

In figure 2 we can see that wealth levels for AR series modelling inference traders are higher, and more stable, for longer, than in the previous simulation. It is difficult to detect much difference between the performance of inference agents embodying non-MML techniques, which show slight variation based upon the IC they use, but in general the more sophisticated ICs outperform the simpler ones. Using the labels from section 2.1.2, we have outperformance of BIC (AVG_BIC) over HQ (AVG_HQ) over CAIC (AVG_CAIC) over AIC (AVG_AIC). As the trading rounds increase beyond 80, the MML inference trader wealth levels split into two, with those MML traders using a model assuming a stationary AR time series (AVG_MMLs) drifting down towards the IC traders, whilst the MML non-stationarity in-

[5]See ref. 5 in [15], [37, p248], or the appendix to [17] for more on the space-quantising lattice.

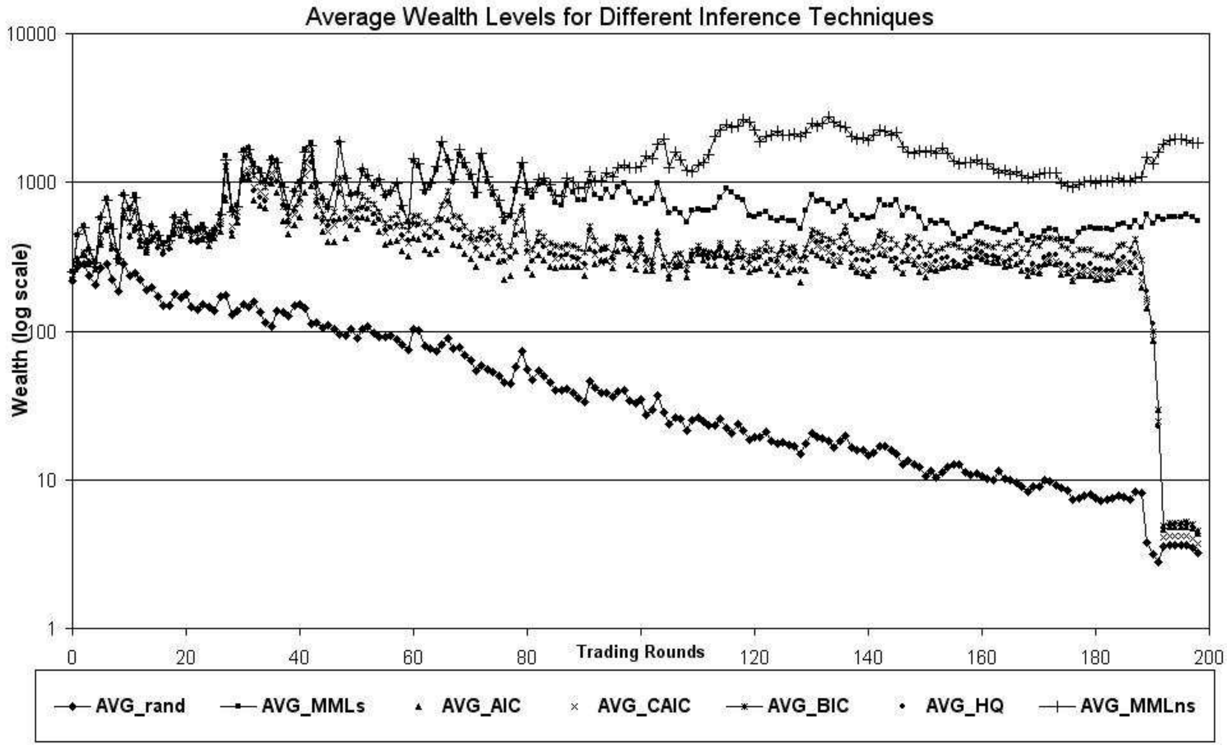

Figure 2. AR noise traders and inference agent wealth levels by type.

ference (AVG_MMLns) outperforms. The collapse of non-MML IC trader wealth levels towards 200 trading rounds represents the almost complete transfer of wealth to the MML agents.

4. Conclusions

We have demonstrated a simulated stock market in which trading agents embody a class of different AR time series inference techniques, and shown that in the presence of a noisy signal, MML-inference technique based agents significantly outperform other traders using different inference techniques. We have shown such performance advantages to be persistent as long as there exists a noisy signal to exploit, and that a level of outperformance exists amongst inferential agents even in the absence of an explicit noisy signal. Such types of agent-based hybrid inference technique models appear to be a reasonable technique to apply to real markets, and in future work we intend to combine inferential agents with genetic algorithm based search techniques in building superior simulations of empirical market characteristics, and to demonstrate the ability of such models to perform in real trading environments.

The performance of the MML-based inference techniques here is not surprising given the success of MML in earlier applications by Dowe et al. (e.g., [15, 10, 32] and references therein) and others, and, we hope, provides further impetus to greater recognition of this methodology and its relevance to practical statistical inference. In terms of market efficiency, we see here that an agent using a superior inference technique will consistently outperform a lesser one. Furthermore (recalling section 1, and as stated in [12]), since in most markets we can never prove that our inference technique is superior, we can neither in general establish that there does not exist some (as yet unused) trading technique that will outperform, nor in general that a market is efficient[6].

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions On Automatic Control*, 19(6):716–723, December 1974.

[2] W. Arthur, J. Holland, B. LeBaron, R. Palmer, and P. Tayler. Asset pricing under endogenous expectations in an artificial stock market. In W. Arthur, S. Durlauf, and D. Lane, editors, *The economy as an evolving complex system II*, pages 15–44. Addison-Wesley, Reading, MA, 1997.

[6]This research was partially supported by Australian Research Council (ARC) Discovery Grant DP0343650

[3] N. Barberis and R. Thaler. A survey of behavioral finance. In G. Constantinides, M. Harris, and R. Stulz, editors, *Handbook of the Economics of Finance*, chapter 18. Elsevier Science, 2003. also NBER Working Paper 9222 (2002).

[4] R. A. Baxter and J. J. Oliver. MDL and MML: Similarities and differences. Tech. Report 207, Department of Computer Science, Monash University, 1994. Amended March 1995.

[5] H. Bozdogan. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, September 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] C. Chiarella, M. Gallegati, R. Leombruni, and A. Palestrini. Asset price dynamics among heterogeneous interacting agents. *Computational Economics*, 22:213–223, 2003.

[8] M. J. Collie. Stock returns distributions: The emergent behaviour of a multi-agent artificial stock market. In M. Mohammadian, editor, *International Conference on Intelligent Agents, Web Technologies, and Internet Commerce (IAWTIC'03), Vienna, Austria*. University of Canberra, February 2003. Available online at http://www.csse.monash.edu.au/∼gjc/Collie03stockreturndistribs.pdf.

[9] M. J. Collie, D. L. Dowe, and L. J. Fitzgibbon. Trading rule search with autoregressive inference agents. Technical Report 2005/174, Clayton School of Information Technology, Monash Univ, Melbourne, Australia, September 2005.

[10] J. W. Comley and D. L. Dowe. *Minimum Message Length, MDL and Generalised Bayesian Networks with Asymmetric Languages*, chapter 11, pages 265–294. M.I.T. Press, 2005. [Camera ready copy submitted October 2003].

[11] D. L. Dowe. The contribution of variance to utility functions. In *Econometric Society Australasian Meeting (ESAM) Conference Proceedings*, volume 3, pages 1–6, July 1997.

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

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

[14] J. D. Farmer and A. W. Lo. Frontiers of finance: Evolution and efficient markets. *Proceedings of the National Academy of Sciences, USA*, 96:9991–9992, August 1999.

[15] L. J. Fitzgibbon, D. L. Dowe, and F. Vahid. Minimum message length autoregressive model order selection. In M. Palanswami, C. C. Sekhar, G. Venayagamoorthy, S. Mohan, and M. K. Ghantasala, editors, *International Conference on Intelligent Sensing and Information Processing (ICISIP)*, pages 439–444. Chennai, India, January 2004.

[16] 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.

[17] B. Hanlon and C. Forbes. Model selection for segmented time series from a Bayesian approach to information compression. Department of Econometrics and Business Statisics Working Paper 8, Monash University, 2002.

[18] E. Hannan and B. Quinn. The determination of the order of an autoregression. *Journal of the Royal Society. Series B (Methodological)*, 41(2):190–195, 1979.

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

[20] C. Hurvich and C. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, June 1989.

[21] G. Iori. A microsimulation of traders activity in the stock market: the role of heterogeneity, agents' interactions and trade frictions. *Journal of Economic Behavior and Organization*, 49:269–285, 2002.

[22] M. Jensen. Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6:95–102, 1978.

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

[24] B. LeBaron. A builder's guide to agent-based financial markets. *Quantitative Finance*, 1:254–261, 2001.

[25] T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397:498–500, February 1999.

[26] B. G. Malkiel. *A Random Walk Down Wall Street*. W.W. Norton and Co., New York, 1 edition, 1973.

[27] M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modelling and Computer Simulation*, 8(1):3–30, January 1998.

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

[29] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[30] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, March 1978.

[31] R. Shiller. From efficient market theory to behavioral finance. Discussion Paper 1385, Cowles Foundation, October 2002.

[32] P. J. Tan and D. L. Dowe. MML inference of oblique decision trees. In *Proc. 17th Australian Joint Conference on Artificial Intelligence (AI04), Cairns, Qld., Australia*, pages 1082–1088. Springer-Verlag, December 2004.

[33] L. Tesfatsion. Agent-based computational economics. ISU Economics Working Paper 1, Iowa State University, 2003. Revised August 24, 2003.

[34] C. S. Wallace. *Statistical and Inductive Inference by Minimum Message Length*. Springer, 2005.

[35] C. S. Wallace and D. M. Boulton. An information measure for classification. *The Computer Journal*, 11(2):185–194, August 1968. Available online at http://www3.oup.co.uk/computer_journal/hdb/Volume_11/Issue_02/.

[36] C. S. Wallace and D. L. Dowe. Minimum message length and Kolmogorov complexity. *Computer Journal*, 42(4):270–283, 1999. [See also replies and rejoinders on pp330-337 and pp345-347].

[37] C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(3):240–265, 1987.