4-15-2008

Stock Index Pricing with Random Walk and Agent-Based Models

Scott N. Swisher IV '08
Illinois Wesleyan University

Stock Index Pricing with Random Walk and Agent-based Models

Scott N. Swisher IV
Illinois Wesleyan University
April 15, 2008

Abstract: The objective of this work is to empirically test the EMH (efficient market hypothesis) and compare its results to those of a viable agent-based competitor using computational simulation. The models are not directly fit to the data; random walk and agent-based methods of stock price determination are statistically compared using the criteria of stationarity, randomness, and autoregressive behavior. The agent-based approach used, styled the "ant trader" model, is based on the ant model established by Kirman in his 1993 work "Ants, Rationality, and Recruitment". Daily returns of the Hang Seng and Nikkei 225 indices are used over the periods 1987-2007 and 1984-2007, respectively. Preliminary simulations run with the agent-based model indicate high sensitivity to parameter changes; parameter imbalances lead to unrealistic growth in returns. Batch stationarity tests using ADF and PP tests suggest that the two models behave similarly under the chosen parameter conditions. However, the random-walk model is found to be more consistent with the available data when using the Wald-Wolfowitz runs test and the Lo-MacKinlay variance ratio test. We conclude that the EMH can be theoretically challenged by the ant trader model, but not empirically. The agent-based model has more realistic assumptions and is more flexible; however, the random walk model agrees with the stationarity and randomness properties of real-world stock index return.

I. Introduction[1]

"The image one gets from the news is that financial markets are dominated by *people*. In contrast, a reading of a standard finance textbook ... can create the impression that financial markets are nearly devoid of human activity" (Thaler 1993). The field of asset pricing, specifically the valuation of stock market shares, has historically played host to a number of contradictory theories regarding the determination of prices. As the debate currently stands, the efficient market hypothesis (EMH) has assumed a dominant position following the enumeration of rational expectations theory at the University of Chicago (primarily) in the 1960s by Muth, Fama, and Lucas (Sheffrin 1996; Shiller 2000). Fama's specification is that "security prices always fully reflect the available information" in an efficient market (Shleifer 2000). Initial econometric testing regarding the efficient-markets theory confirmed germane hypotheses, but by the mid-1970s academics were increasingly skeptical duc to the restrictive nature of the assumptions and contradictory empirical findings (Sheffrin 1996). As a result, alternative theories involving non-rational actors were developed under the banner of behavioral finance by Shiller, De Bondt, Thaler, Roll, and others; however, the EMH remained the *de facto* central paradigm of finance, a position it has held for over thirty years (Hirshleifer 2001; Sheffrin 1996; Shleifer 2000).

In such a context, questioning the current theory vis-à-vis well-developed alternatives is perfectly reasonable because the consensus is not well-defined (Arthur et al. 1997; Baker and Wurgler 2007; De Bondt and Thaler 1984; Hirshleifer 2001; Hong and Stein 2007; Shleifer 2000; Worthington and Higgs 2003). The objective of this work is to empirically test the EMH and compare its results to those of a viable competitor using computational simulation. Specifically, the individual-agent approach has been gaining momentum recently as the appropriate numerical tools are now widely available (Bonaneau 2002; Cioffi-Revilla 2002; Diks et al. 2007; Gilbert and Bankes 2002; Inchiosa and Parker 2002; Tesfatsion 2002). This fact, coupled with intensifying doubts concerning the validity of efficient-markets theory, has led to intensive use of the agent-based approach with computational agent-based modeling (ABM) of financial markets (Bonabeau 2002). Although multiple theories currently compete with the EMH to varying degrees, we focus explicitly on the use of ABM to generate results consistent with Hang Seng and Nikkei 225 price changes. The agent-based results are compared with output from a random-walk model directly inspired by the tenets of the EMH; model parameters are selected such that each model is run with reduced error. However, the methodology is not entirely based on error-minimizing fitting functions;

[1] I would like to thank the members of my honors research committee at Illinois Wesleyan University: T.X. He, S.H. Lee, N. Jaggi, and M. Seeborg. I. Odinaka, an undergraduate at the university, gave me much-needed advice regarding *Mathematica* programming.

stationarity[2], randomness, and autocorrelograms are used as comparative metrics in order to assess model accuracy and appropriate characterization of historical data. Our results imply that the random-walk model is more consistent with the empirical facts in this particular situation.

Throughout the 1990s, a perception has been developing that efficient-markets theory is inconsistent with the available data; critics cite price volatility in excess of what would be dictated by changes in fundamental value as evidence (Thaler 1993; Shleifer 2000). Additionally, those in behavioral finance argue that no time-constrained individual could ever possess the computing power required to calculate and recalculate the fundamental value of all stocks in a diversified portfolio (Hirshleifer 2001). Such skepticism is countered by those empirical results that do confirm the EMH (Pearce and Roley 1985); efficient financial markets are consistent with *laissez-faire* and the innate wisdom of unconstrained market forces (Ormerod 1998). The net result is a field characterized by theoretical conflict between alternative theories, a situation not uncommon in the economics discipline, but the dispute is as much dogmatic and political as it is empirical and scientific (Schleifer 2000). Stock markets have been traditionally viewed as the apogee of free-market idealism; shares are traded on a daily basis without significant restriction, so each stock price should represent actual (fundamental) value. Deviation from the correct valuation is rapidly purged from the market system by the broad mass of fundamentalist traders; as such, the EMH posits that each stock price reflects the discounted present value of the sum of future earnings.

Confirming the applicability of bounded rationality and the imperfection of market traders leads to a reexamination of previous bubble-corrective incidents that brought financial ruin to millions and persistent negative economic consequences (Shiller 2000). Economists are prompted to question if the market can be manipulated to make mistakes as speculators push asset prices higher for their own self-enrichment at the cost of macroeconomic stability (Raines and Leathers 2000). Regardless of the evidence, the efficient-markets theory has an incumbent advantage that can be nearly impossible to nullify. Resistance to theoretical change is also due to the esoteric nature of the topic, and although data availability is no longer a problem (Pearce 1984), financial data require statistical analysis using complex econometric modeling (Worthington and Higgs 2003). The utility of the ARCH model in performing data analysis in finance, for example, is due to the heteroscedastic and autoregressive properties of stock prices over time; the time-series are characterized by short-lived bursts of volatility (Shumway and Stoffer 2006). Extensive data availability can also be viewed as a mixed blessing: although ample series are available for investigation, data mining (i.e. selecting the data set that maximizes model performance) can become endemic to the study of random walk models (Hirshleifer 2001). Consequently, rejection or

[2] A set of time-series data is said to be stationary if its mean and variance do not change as a function of time.

acceptance of the EMH is a function of the data set used, so no generalized conclusions are drawn by the literature.

Our objective here is to empirically test the EMH and compare its results to those of an agent-based alternative using *Mathematica*-based computational simulation. The two models are empirically compared (without trying to find the SSE-minimizing model outcome) using the criteria of stationarity, autoregressive behavior, and randomncss[3]. The agent-based approach used, termed the "ant trader" model, is based on the ant model established by Kirman in his 1993 work "Ants, Rationality, and Recruitment". Daily returns of the Hang Seng and Nikkei 225 indices are used over the periods 1987-2007 and 1984-2007, respectively. The model that most accurately reflects the conditions present in these real-world markets in terms of statistical testing (stationarity, randomness, autocorrelograms) will be theoretically preferred.

This study is not immune to the issue of limited applicability of results, and the determinations made here regarding the efficient-markets theory are not necessarily extensible to other indices that differ non-trivially in terms of period under consideration, composition of stocks, industrial concentration, or regional factors (Worthington and Higgs 2003; Huber 1995). Disagreement between the EMH and its opponents will continue regardless of new scholarly publications because stock market efficiency is a function of index characteristics and time; some indices are more adept at incorporating information than others[4] (Worthington and Higgs 2003).

II. Contextual Literature

Arguments made for efficient markets were originally theoretical, consisting of the formation and application of rational expectations by Muth, Lucas, and so on; the idea of rational expectations quickly migrated from macroeconomics to finance, resulting in the random-walk model of stock prices (Sheffrin 1996). The EMH is appealing partly because of its implications: stock market prices reflect all available information concerning the discounted expected value of future corporate earnings streams, i.e. the capital asset pricing model, or CAPM (Ibid.). Consequently, no long-term profitable trading rule can be established as stock prices engage in a random walk around fundamental value (Ibid.). The theory provides a fundamentalist trading rule as follows: sell if the price is above the "true" value (overvalued) and buy/hold if price is below the "true" value (undervalued), where the true value reflects the future

---

[3] A set of time-series data is said to be random if no tendency to remain above or below the median or mean exists.

[4] Co-existence of efficient and inefficient markets implies that both sides can find supporting evidence, which leads to contradictory results.

earnings stream of the asset (Ibid.). Efficient-markets theory implies that financial analysts are redundant when all market actors have access to complete information, and technical analysis based on short-run trends is ineffectual (Hirshleifer 2001; Shleifer 2000). Volatile day trading is unproductive, assuming that the underlying true value is not changing in the short-run, so all the rational investor has to do is buy and hold undervalued stocks[5] until they eventually become overvalued in the course of a random walk stochastic process. The discovery of a long-term profitable trading rule would invalidate the efficient-markets theory, however (Thaler 1993).

The EMH can be broken down into three subclasses as defined by Fama in his 1970 work "Efficient Capital Markets: A Review of Theory and Empirical Work": a particular market may exhibit weak, semi-strong, or strong efficiency (Shleifer 2000). A market is said to be weakly efficient if complete awareness of past information does not improve long-run portfolio profitability (Ibid.). Semi-strong efficiency is satisfied if portfolio return cannot be increased using knowledge of publically-available information (Ibid.). The strict criterion of strong-form efficiency is the most difficult to prove; we say that a market is strongly efficient if even insider (non-public) information cannot improve portfolio return. As such, a majority of authors deal with weak-form efficiency in financial markets because it is difficult to properly treat the insider information set econometrically[6] (Sheffrin 1996). Although this paper does not directly test for market efficiency, the definitions are worth noting due to their importance in the EMH framework. The random-walk specification that will be used is weakly efficient; a more stringent specification would require additional evidence regarding the information set, which is beyond the scope of this paper.

A voluminous literature has grown around the efficient market hypothesis; a concise summary is provided by Sheffrin in *Rational Expectations* (1996) as cited previously. The EMH assumes the following: investors are rational actors, imperfectly rational investors trade randomly with zero net effect, and arbitrage undertaken by rational actors nullifies the actions of non-rational traders (Shleifer 2000). Of these three, the most important assumption is arbitrage; if of sufficient number, rational arbitrageurs[7] can effectively purge the market of its irrational elements through fundamentals-based trading. Therefore, the requirement that market agents are perfectly rational can be relaxed and the theory remains consistent under suboptimal (i.e. realistic) conditions. Under these assumptions with a market composed of risk-

[5] Undervalued, i.e. below fundamental value, assuming that fundamental value is well-defined and can be computed.

[6] By definition, insider information is not known to the public; therefore, compiling data on such a topic might prove impossible. Individuals with insider information will not want to divulge the extent of their knowledge due to fear of prosecution or loss of trading advantage.

[7] "Rational arbitrageurs" can be classified as individuals who rationally exploit price differentials (deviation from fundamental value) in order to obtain trading advantage.

neutral rational investors, mathematical economists Samuelson and Mandelbrot proved that returns follow a random walk process in the mid-1970s (Ibid.). At this time, empirical evidence overwhelmingly supported the efficient-markets theory and arbitrage was able to explain away isolated outbreaks of irrational "noise trader" behavior. The 1972 event study of Scholes suggested that arbitrageurs can only operate when near-perfect substitutes are available for an individual stock, but his work generally confirmed the EMH regardless of the prerequisites (Ibid.). Of note here is that the EMH relies upon a multitude of powerful assumptions, mainly the primacy of rational, fundamentalist traders. The entire logical argument is invalidated with a violated assumption, but empirical evidence is also supportive.

Empirical testing has proven effective in validating the efficient-markets hypothesis; see Pearce and Roley "Stock Prices and Economic News" and their subsequent confirmation of the EMH (1985). The authors use S&P500 return data coupled with data sources that address expectations and announcements; expectations reflect the state of the information set, while announcements stochastically shock the information set. Theoretically, they adapt the rational expectations framework to the question of predicting changes in stock index prices as follows:

$$\Delta SP_t = a + bx_t^u + cx_t^e + d_i \sum_{i=1}^{m} x_{t-i}^u + e_t \qquad (2.1)$$

The change in stock price at time t is a function of the unexpected announcements vector $\mathbf{x}^u$, expected announcements vector $\mathbf{x}^e$, all previous unexpected news $\sum \mathbf{x}^u$, and an error term e indexed by t. Coefficient *b* should be significantly non-zero, while coefficients *c* and *d* are predicted to be zero in accordance with the EMH. This is because only newly-presented unexpected news should serve as a stochastic shock; expected news and previous surprises ought to be integrated into the price already.

Pearce and Roley find that unexpected announcements induce nearly-instantaneous changes in S&P 500 price, but expected (anticipated) announcements do not have a statistically significant effect on stock prices (1985). These results concur with the theoretical predictions of the EMH; only surprise changes in the information set lead to non-trivial stock price movements. Therefore, stock prices reflect all available information, which includes expectations about future announcements regarding monetary policy and corporate finance. Pearce and Roley used the efficient-markets theory as their null hypothesis when conducting statistical tests regarding regression coefficients, so more precisely the authors did not disprove the theory. Their paper is representative of an extensive body of work that has failed to reject the EMH, insofar as a failure to reject represents validation and lends credibility to the proponents of rational expectations.

The random walk model has been specified in a number of increasingly sophisticated ways as per Hagerman and Richmond "Random Walks, Martingales and the OTC" in which the authors validated the weakly efficient form of the efficient market hypothesis (1973). After stating that "the evidence

overwhelmingly shows that security returns are independent over time," the authors propose an extension to the EMH in which over-the-counter (OTC) securities would be used in place of stocks. A more direct method is used to investigate the hypothesis in this study: serial correlation coefficients are computed for a set of 253 securities, along with the use of distribution-free runs testing, to test for autocorrelation. Hagerman and Richmond find that 12.3% of the serial correlation coefficients were significantly non-zero at the 5% level under normality, but this result is discounted as flawed for a number of reasons: variable error, covariance with the aggregate market trend, and the normality assumption introduced substantial bias into the estimate of ρ. The runs test fails to find a significantly non-zero proportion of securities with excessive deviation from normality. Therefore, Hagerman and Richmond fail to reject the null hypothesis that the OTC securities market is weakly efficient as posited by the efficient-markets theory; the EMH is not without empirical support.

In summation, we cannot claim that the efficient market hypothesis has explicitly failed. The theory has extensive empirical justification, as shown previously, and the EMH/CAPM duality has been very successful: the models remain essentially intact after more than thirty years of criticism. However, the strict assumptions of efficient-markets theory can appear implausible in the current trading environment and critics are numerous and vocal (Thaler 1993). Economic history is rife with examples of individually-irrational herd behavior, bubble formation along with the inevitable crash: Baker and Wurgler cite the Nifty Fifty and the Black Monday crash of 1987 as examples of rational traders gone awry, violations of the EMH in the short-run (2007). Recent macroeconomic events, particularly the 1990s "dot-com" bubble and the 2005 U.S. housing bubble, have served to discredit the concept of efficient financial markets. The field of behavioral finance has emerged as a center of heterodox thought in this area, proposing alternative theories of stock price formation (Hirshleifer 2001).

The field of behavioral finance has propagated a number of alternative theories based around a set of common concepts, such as cognitive biases, but no consensus exists. Hong and Stein provide a list of reasons why stock prices would persistently deviate from fundamental value in "Disagreement and the Stock Market" (2007). Momentum, the continuation of upward or downward trend regardless of other factors, is the central tenet of technical analysis (momentum investing), an investment philosophy that argues for the intensive examination and mimicry of trend. The existence of stock price momentum again implies that a long-term profitable trading rule exists, which is inconsistent with the efficient-markets theory. In post-earnings (announcement) drift, returns are abnormally high/low following positive/negative news, respectively; the trading rule in this case is to buy stocks following a positive announcement. Mean reversion is equivalent to the so-called "overreaction hypothesis" of De Bondt, Thaler (1984); good/bad news in the short-run leads to losses/gains in the medium-run (3+ years). Almost all of the alternative hypotheses based on cognitive biases, such as the illusion of control, have not been

explored to the extent required to become serious competition for the EMH. Hirshleifer's survey article, "Investor Psychology and Asset Pricing", is an example of how diverse and disparate the field has become; many scholars are trying to connect psychological concepts with investor behavior as manifested in stock market prices, but a proliferation of applicable theories in cognitive psychology has resulted in a rather wide range of applications to finance (2001). Each cognitive bias has been explored by a limited number of authors, so no one concept has reached the requisite critical mass, so to speak, to genuinely compete with the efficient market hypothesis.

Sources like De Bondt and Thaler "Does the Stock Market Overreact" (1984) find an overreaction effect in stock prices after a significant news announcement. Theoretically, human violation of Bayes' rule implies that traders tend to overestimate the effect of positive unexpected news; therefore, we should empirically see excessive stock gains immediately after favorable announcements. This "overreaction hypothesis" contradicts efficient-markets theory since stock prices temporarily yet persistently overestimate the actual value. Monthly data on NYSE common stock returns are used from Jan. 1926 to Dec. 1982; the authors cite problems with the use of daily data, such as the "bid-ask" effect and infrequent trading. Two groups of stocks are defined: a loser portfolio, stocks that suffered negative news in the recent past, and a winner portfolio, stocks under the influence of positive announcements; the portfolios are tracked for 2-5 years after the news event. Since an unanticipated announcement would tend to overvalue or undervalue stocks with positive or negative reports, respectively, the authors expect that winners will retreat and losers will gain during the subsequent correction. De Bondt and Thaler find empirically that this is the case; the loser portfolio substantially outperformed the winner portfolio in every case considered. Therefore, an investor could formulate a contrarian trading rule as follows: buy stocks on negative news and sell stocks on positive news. Such a strategy could be profitable in the long-run, according to the authors, due to this overreaction effect. The discovery of a profitable trading rule in the long-run implies the invalidity of the EMH as stock prices are not engaging in random walks.

Baker and Wurgler (2007) try to predict stock market return using an index of investor sentiment, and this approach is relatively common in the literature. The theoretical concept is that price changes reflect exogenous changes in investor sentiment, which can be measured by a derived sentiment index. In "Investor Sentiment in the Stock Market", the authors construct an investor sentiment index based on six factors which serve as proxies: trading volume, dividend premium, the closed-end fund discount, the number and preliminary returns of IPOs, and the equity share in new issues. Baker and Wurgler attempt to remove the macroeconomic components of the proxy variables in order to target solely sentiment as opposed to accounting for other exogenous changes. This index is moderately successful in terms of predicting future returns; stocks that are difficult for arbitrageurs tend to be more intensively affected by

changes in sentiment. These results, when combined with the conclusions of others in behavioral finance, imply that sentiment indices can be accurate predictors of stock market return.

Volatility in excess of changes in fundamental value is another cornerstone of the behavioral finance literature, and the topic has been discussed extensively (Thaler 1993; Shleifer 2000; Shiller 2000). As an example, consider "What Moves Stock Prices?" by Cutler, Poterba, and Summers (1989) as reproduced in Thaler (1993). After accounting for changes in publicly-available information, the authors conclude that as much as half of the variance in stock prices remains unexplained; this result rejects the null hypotheses of weak or semi-strong efficiency. Again, this result suggests the fallibility of efficient-markets theory in certain situations.

In *Irrational Exuberance*, Shiller colloquially discusses herd behavior, the idea that individual decision-making is influenced by the choices of others in what is termed an "information cascade" (2000). Such a cascade is characterized by incomplete information: since no agent knows the true fundamental value of an asset at a fixed point in time, the decisions of other agents in the previous period are used as a reference point (Scharfstein and Stein 1990). For example, the Oct. 1987 bull market was partially driven by money managers who wanted to continue their employment at investment firms; no manager wanted to miss out on the record gains that were perpetuated by the traders themselves (Ibid.). Alternatively, discerning the actual value of a stock takes a considerable amount of time and financial resources (Hirshleifer 2001); an investor might find it advantageous to bypass the research process by agreeing with the majority. This principle is an extension of Kirman's agent-based ant model, which is the basis of this paper's approach.

As defined by Kirman (1993), the "ant model" is a well-known agent-based model of ant colony behavior during the search for food. The colony is exposed to two equally favorable non-exhaustible food sources (sites A and B) and pheromone trails from the initial scouts can be modeled as positive feedback (Ormerod 1998). We partition the colony's fixed[8] population into two mutually-exclusive groups: ants currently searching for food at site A (group A) and those foraging at site B (group B). The probability of a new ant selecting site A[9] is directly related to the number of ants in group A, and thus indirectly related to the number of ants in group B. However, the random chance that an ant would spontaneously and independently switch from one group to another is ever-present[10] (Kirman 1993).

---

[8] Under the assumption that the colony experiences zero population growth in the short-run. A more sophisticated long-run model could express population growth as a function of the quantity of food gathered in each period.

[9] Joining group A (visiting site A), leaving group B (ignoring site B).

[10] The analogous stock market situation: group A is the set of optimists (bulls); group B is a collection of pessimists (bears); the ants are traders who engage in a search for return on initial investment given the known risk-reward environment and exogenous macroeconomic variables (theoretically).

The site visited by most initial scouts may become very popular due to the positive feedback mechanism, but sudden switching to the other group can occur if a cluster of ants randomly decides to investigate an alternative site (Ibid.). No long-run equilibrium exists, and rapid changes can still occur regardless of the time horizon due to the model's statistical qualities. Kirman's simple Markov chain is able to explain ant behavior so well because each ant is considered as an individual agent that chooses a food site in each period. Ormerod states that, "the idea that the system as a whole can be understood by the behavior of a single, representative agent is a complete non-starter" (1998); the conventional approach in economics, aggregation with the representative *homo economicus*, cannot apply here due to the ant recruitment method and its reliance on positive feedback.

Computational agent-based modeling is a relatively new simulation technique, at least in economics; see Bonabeau; Cioffi-Revilla; Gilbert and Bankes 2002. The increased availability of simulation tools has led to intensive application of this mathematical framework to a wide range of problems, such as individuals trying to leave a burning building through a single door (Bonabeau 2002). Each person is modeled as an agent with generalized behavioral rules regarding conduct in the group; for example, an individual attempting to escape from a fire might try to avoid or help others on the way to the door (Ibid.). The agent-based approach allows for precisely-defined unique actors: based on parameters, one agent may be more likely to attempt a reckless exit than another. Agent-based models are typically solved via simulation techniques because no closed-form solution can be found analytically. Therefore, we would expect that each trial of an agent-based model generates a unique solution that is not strictly reproducible if probabilistic components are involved in the modeling scheme. Parameter values are important in ABM because parameter inaccuracy can lead to large changes in model outcomes[11]; the parameters of interest are usually exogenously determined, however, making empirical comparison difficult (Kirman 1993).

A recent application of the interacting-agent approach can be seen in Arthur et al. "Asset pricing Under Endogenous Expectations in an Artificial Stock Market" (1997) as the authors construct a self-contained artificial stock market in which each trader is assigned his/her own unique bundle of pricing models. The poorly-performing models are dropped and new models are added, so each agent generates expectations based on the outcomes of his/her respective models. Therefore, expectations are internally generated, not exogenously imposed, and prices interact with expectations in a dynamic fashion.

[11] Final index price is one such outcome, and we usually have a target for that value based on historical information. Therefore, set the expected value of final index price equal to the recorded final price in the data set in order to maximize the likelihood of achieving the actual quantity in a representative simulation run.

III. Theoretical Discussion: Random Walk Model

As shown by Samuelson and Mandelbrot, efficient-markets theory implies that the value of a frequently-traded stock should engage in a random walk about its fundamental value because stock prices fully reflect all available information about the expected value of discounted future earnings. Traders, who primarily concern themselves with the difference between actual and fundamental value, will quickly correct the price of an undervalued or overvalued stock. The availability of complete, accurate information implies that traders are able to integrate changes in the important earnings indicators into the stock price almost instantaneously. Stock market indices are simply bundles of individual stocks, so index value should also deviate from its fundamental value in a random-walk process, where the fundamental value of an index is the summation of the fundamental values of its component stocks. Theoretically, we have that expected returns are positive and constant (Sheffrin 1996); formally,

$$E[\check{Z}_t - \overline{Z}_t | I_{t-1}] = 0 \qquad (3.1)$$

$\check{Z}_t$ = actual return in time t,

$\overline{Z}_t$ = expected return in time t,

I_{t-1} = information set at time $t-1$.

The expected value (E[] operator) of actual return minus (constant) expected return given the previous information set is zero; actual return never deviates from expectations based on available information (Ibid.). If deviation from expected return does occur, this disparity should be extremely short-lived as expectations rapidly adjust.

The particular mechanism through which the stock price random-walk is transmitted can be specified in a number of equivalent ways, three of which are considered here. The simplest case of a random-walk model is the driftless case with finite up or down steps at each time interval; at each decision point, the series either increases by one or decreases by one with equal probability. "Drift" is conceptually defined as the nonrandom per-period change in the dependent variable; drift ought to be representative of long-run change or trend. Var is the variance operator.

$$a_0 = p_0 \qquad (3.2\text{-}3.5)$$

$$a_{n+1} = a_n \pm 1 = a_n - (-1)^{\mathrm{Binomial}\left[1,\frac{1}{2}\right]} \mathrm{Binomial}\left[1, \frac{1}{2}\right]$$

$$E[a_n] = p_0$$

$$\mathrm{Var}[a_n] = n$$

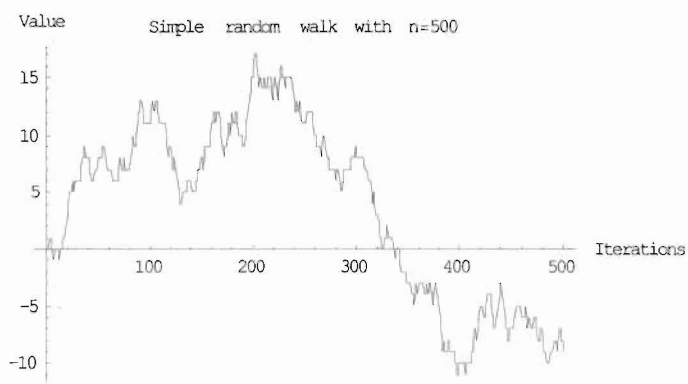
$\{a_n - a_{n-1}\} \sim (-1)^{\text{Binomial}[1,\frac{1}{2}]}\,\text{Binomial}[1, \frac{1}{2}]$ (3.6-3.8)

$E[a_n - a_{n-1}] = p_0$

$Var[a_n - a_{n-1}] = 1$

Figure 3.1 shows a representative simulation of this type of stochastic process. The process is non-stationary[12] because the variance is a function of time index n. The sequence of first differences $\{a_n - a_{n-1}\}$ is stationary, however; in this case, differencing can achieve stationarity. $E[\rho(a_n, a_{n-1})] = 0$ as well.[13]

Fig. 3.1: Simple random walk stochastic process.



One-dimensional random walks without drift are defined as follows in the finance literature:

$a_0 = p_0$ (3.9-3.15)

$a_{n+1} = a_n + N(0,\sigma^2)$

$E[a_n] = p_0$

$Var[a_n] = n\sigma^2$

$\{a_n - a_{n-1}\} \sim N(0,\sigma^2)$

$E[a_n - a_{n-1}] = 0$

$Var[a_n - a_{n-1}] = \sigma^2$

[12] A stochastic process is said to be stationary if its probability distribution function is time-invariant; we would expect that a stationary process has time-independent moments, such as the first and second moments of mean and variance, if they exist. The concept of stationarity is important because non-stationarity implies that the underlying probabilistic process (probability density function) is changing over time.

[13] Where ρ is the correlation coefficient between a_n and a_{n-1}.

The untransformed process is non-stationary, while the differenced sequence is stationary; $E[\rho(a_n, a_{n-1})] = 0$ since $\{a_n - a_{n-1}\} \sim N(0,\sigma^2)$.

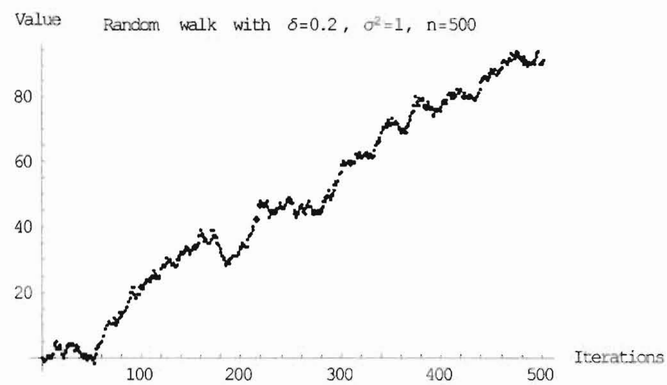
A one-dimensional random-walk process with drift δ is defined recursively as follows:

$a_0 = \rho_0$ (3.16-3.17)

$a_{n+1} = a_n + \delta + N(0,\sigma^2)$

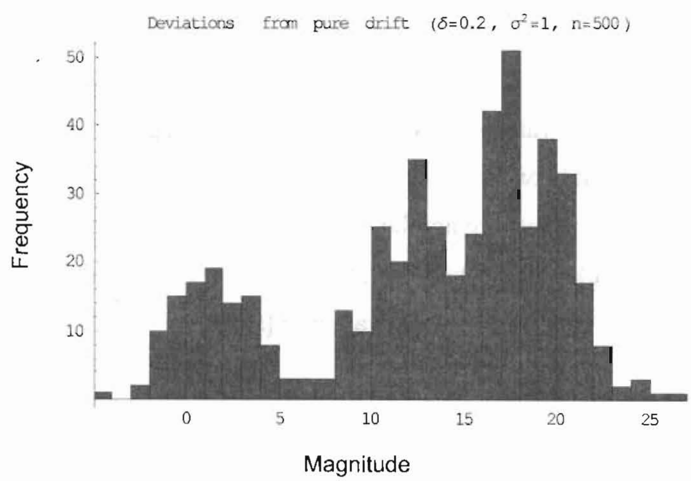
$N(0,\sigma^2)$ is a normally-distributed random variable with mean 0 and constant variance σ^2. A typical result of such a process is plotted below (Figure 3.2).

Fig. 3.2: Random walk process (with drift δ).



Taking the difference $|a_n - n\delta|$, the absolute deviation from pure drift, we find that no particular distribution arises regarding the density of this quantity; each trial distribution looks entirely unique. Clustering occurs when the series oscillates around an arbitrary fixed point for a period of time. Figure 3.3 provides a density plot of deviations from drift for the same simulation that generated Fig. 3.2 above.

Fig. 3.3: Deviation from drift, random walk model.



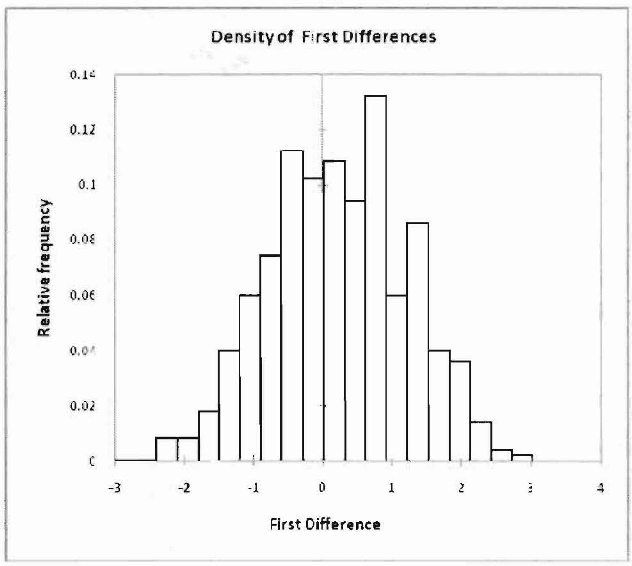
In this particular example, the total deviation from pure drift was 6481.5; this value will be non-zero due to variance present in the normally-distributed random variable. For this simple random-walk model, the first differences are normally distributed with mean δ and variance σ^2. This directly follows from the recursive definition of a_n as specified previously.

$$\sum_{n=0}^{m} |a_n - n\delta| = \text{deviation from drift} = 6481.5 \qquad (3.18\text{-}3.19)$$

$$a_{n+1} - a_n = \delta + N(0,\sigma^2) \sim N(\delta,\sigma^2)$$

Figure 3.4 shows the density of first differences for the $\{a_n - a_{n-1}\}$ terms in our example:

Fig. 3.4: Relative frequency of first differences, random walk model (with drift).



We have $\mu = 0.243 \approx 0.2 = \delta$ and $\sigma^2 = 0.958 \approx 1$ for the set of first differences in our example; $\mu \rightarrow \delta$ and $\sigma^2 \rightarrow 1$ as $n \rightarrow \infty$, so even n=500 is a sufficiently small sample size such that we find significant deviation from the normally-distributed ideal. Because the first differences are normally distributed, we would expect that $\rho(1da_n, 1da_{n-1}) \approx 0$ with sufficiently large n. In other words, a time-series data set generated via an underlying random walk process should have no autoregressive properties[14] whatsoever; the expected correlation coefficient between successive values ought to be zero or near zero. $|\rho(1da_n, 1da_{n-1})| \gg 0$ implies that the data are not the result of a random walk process, regardless of the drift parameter δ. The random walk process with drift, as defined previously, is non-stationary because its variance is a function of the time index *n*. First, we examine $E[a_n]$:

[14] The term "autoregressive" refers to $\rho(a_n, a_{n-p}) \neq 0$, $p > 0$, $(n-p) \geq 0$.

$E[a_n] = E[a_{n-1} + \delta + N(0,\sigma^2)]$ (3.20-3.23)

$E[a_n] = E[a_{n-1}] + E[\delta] + E[N(0,\sigma^2)]$

$E[a_n] = \sum_{n=0}^{m-1} a_n + \delta$

$E[a_n] = \rho_0 + (n-1)\delta + \delta = \rho_0 + n\delta$

Since $E[a_n]$ is a function of the time index n, the process is already proven non-stationary. The variance of the n^{th} term can be expressed as follows since the sum of normally-distributed random variables has the property $N(\mu_1, \sigma^2_1) + N(\mu_2, \sigma^2_2) + \ldots + N(\mu_k, \sigma^2_k) = N(\sum\mu_k, \sum\sigma^2_k)$.

$Var[a_n] = \sum_{n=0}^{m-1} Var[a_n] + \sigma^2 = (n-1)\sigma^2 + \sigma^2 = n\sigma^2$ (3.24-3.25)

$\lim_{n\to\infty} Var[a_n] = \infty$

Therefore, additional evidence is provided for non-stationarity of the random walk with drift process as the variance of the n^{th} term is again a function of time index *n*. As *n* increases, we would expect that the range of possible outcomes becomes larger, but the final position of the series is still distributed approximately normally around $E[a_n]$. Although the random walk with drift process itself is non-stationary, its first differences are stationary as $\{1da_n\} \sim N(\delta,\sigma^2)$.

The three random-walk processes discussed have common characteristics that can be empirically tested for regardless of the particular model specification used, and these traits are shared by all random-walk models:

(i) $E[\rho(1da_n, 1da_{n-1})] = 0$; 1d = first difference;

(ii) $\{a_n\}$ is non-stationary;

(iii) $\{\Delta a_n\}$ is stationary;

(iv) $\lim_{n\to\infty} Var(a_n) = \infty$.

Our purpose in examining multiple specifications was to draw out these useful shared traits. The presence of autoregressive behavior in the first differences of a particular time-series data set indicates that original series was not the product of a random walk process. Additionally, testing for stationarity in the unmodified and first difference financial time-series can validate or invalidate the random walk hypothesis (Diba and Grossman 1988). Tests for non-stationarity include the augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) tests for unit roots; the existence of a unit root implies that the original series is non-stationary, but differencing may be used to obtain a stationary series. Therefore, tests are available that will evaluate the soundness of the EMH as manifested in the random walk hypothesis using data from the Hang Seng and Nikkei 225 stock market indices. Drift parameter δ can be interpreted as the long-run trend regarding the value of the index as determined by corporate finance and macroeconomic fundamentals; the EMH posits that stock prices will

engage in a random walk around this trend as all available information has already been integrated into the price.

IV. Theoretical Discussion: Agent-based Model

As summarized by Ormerod in *Butterfly Economics*, the agent-based approach to time-series modeling defines a finite number of groups that probabilistically interact with each other according to simple behavioral rules (1998). ABM treats each individual separately, and although the behavioral rules may be uniform across individuals and groups, large-scale simplification and aggregation is impossible. Many economic models can be solved by resorting to the representative agent approach and aggregating across a particular group, but agent-based models are defined by the inter-agent or inter-group dynamic; using a single agent to model the behavior of a cluster of agents will remove the micro-level mechanics that enable person-to-person interaction. The statistical nature of ABM implies that each model trial will generate a unique outcome because random variables are embedded into the recursive equations; however, the outcome of the n^{th} period takes the previous *n-1* outcomes as given. Therefore, the Markov chain ABM approach requires the use of simulation and variation of parameters in order to reach any well-supported conclusions as one cannot test directly for agent-based behavior in financial time-series. The ant model framework (two groups, four flows between them) was adapted from Kirman's "Ants, Rationality, and Recruitment" (1993). We will now proceed to the development of the agent-based "ant trader" investor sentiment model.

n : time index
a_n : buyer/bullish/optimistic group of traders
b_n : seller/bearish/pessimistic group of traders
$c_n = a_n - b_n$: difference in group sizes (net optimism, sentiment index)
P(n) : stock (index) price as a function of time
N : total number of traders participating in the market

$a_{n=0} = a_0$ (4.1-4.4)

$b_{n=0} = b_0$

$N = a_0 + b_0 = a_n + b_n$

$\forall n \in [0,\infty)\ \exists a_n, b_n$ s.t. $N = a_n + b_n$

p_1 : random a→b switch probability, p1 ∈ (0,1)

p_2 : a→b persuasion parameter, $[p_2(b_n)] \in (0,1)$

p_3 : random b→a switch probability, p3 ∈ (0,1)

p_4 : b→a persuasion parameter, $[p_4(a_n)] \in (0,1)$

p_5 : translational parameter (c(n) to ΔP(n)), p5 ∈ (-∞,∞)

$p(a \to b)_n = p_1 + p_2[b(n-1)]$ (4.5-4.10)

$p(b \to a)_n = p_3 + p_4[a(n-1)]$

$a_n = \text{Binomial}[a_{n-1}, 1 - p(a \to b)_n] + \text{Binomial}[b_{n-1}, p(b \to a)_n]$

$b_n = \text{Binomial}[a_{n-1}, p(a \to b)_n] + \text{Binomial}[b_{n-1}, 1 - p(b \to a)_n] = N - a_n$

$P(0) = p_0$

$P(n) = P(n-1) + p_5[c(n)]$

An expression can be obtained for the expected value of price, E[P(n)], as a function of the expected value of optimist group a_n, $E[a_n]$. Expected value must be defined recursively as dictated by the agent-based model's structure of repeated binomially-distributed draws.

$$E[a_0] = a_0$$
$$\begin{aligned} E[a_n] &= E[\text{Binomial}[a_{n-1}, 1 - p(a \to b)_n] + \text{Binomial}[b_{n-1}, p(b \to a)_n]] \\ &= (E[a_{n-1}])(1 - E[p(a \to b)_n]) + (E[b_{n-1}])(E[p(b \to a)_n]) \\ &= (E[a_{n-1}])(1 - p_1 - p_2 E[b_{n-1}]) + (E[b_{n-1}])(p_3 + p_4 E[a_{n-1}]) \\ &= (E[a_{n-1}])(1 - p_1 - p_2(N - E[a_{n-1}])) + (N - E[a_{n-1}])(p_3 + p_4 E[a_{n-1}]) \end{aligned}$$
$$E[a_n] = p_3 N + E[a_{n-1}]\,(1 - p_1 - p_2 N - p_3 - p_4 N) - E[a_{n-1}]^2 (p_2 - p_4)$$

Price is a recursive function of the expected value of a_n:

$$E[P(0)] = p_0$$
$$E[P(n)] = E[P(n-1)] + p_5 (2E[a_n] - N)$$

With $p_1 = p_3$, $p_2 = p_4$, so-called "balanced parameters", we have that[15]:

$$E[a_n] = a_0$$
$$E[P(n)] = p_0$$

This implies that the sequence {P(n)} is stationary if and only if the parameters are balanced. {(P(n)} is non-stationary if the parameters are imbalanced. A similar process can be repeated for the variance of the $\{a_n\}$ and {P(n)} series:

[15] From a market perspective, parameters remain balanced only in the absence of stochastic information shocks.

$$
\begin{aligned}
\mathrm{Var}[a_0] &= 0 \\
\mathrm{Var}[a_n] &= \mathrm{Var}[\mathrm{Binomial}[a_{n-1}, 1 - p(a \to b)_n] + \mathrm{Binomial}[b_{n-1}, p(b \to a)_n]] \\
&= (E[a_{n-1}])(1 - E[p(a \to b)_n])(E[p(a \to b)_n]) + (E[b_{n-1}])(E[p(b \to a)_n])(1 - E[p(b \to a)_n]) \\
&= (E[a_{n-1}])(1 - p_1 - p_2 E[b_{n-1}])(p_1 + p_2 E[b_{n-1}]) + (E[b_{n-1}])(p_3 + p_4 E[a_{n-1}])(1 - p_3 - p_4 E[a_{n-1}]) \\
&= (E[a_{n-1}])(1 - p_1 - p_2(N - E[a_{n-1}]))(p_1 + p_2(N - E[a_{n-1}])) + (N - E[a_{n-1}])(p_3 + p_4 E[a_{n-1}])(1 - p_3 - p_4 E[a_{n-1}]) \\
&\dots \\
\mathrm{Var}[a_n] &= Np_3(1 - p_3) + E[a_{n-1}](p_1(1-2Np_2) - p_1^2 + Np_2 - N^2 p_2^2 - p_3 + p_3^2 + Np_4 - 2Np_3 p_4) + E[a_{n-1}]^3 (p_4^2 - p_2^2) \\
&\quad + E[a_{n-1}]^2 [\, p_2^2(2p_1 - 1) + 2Np_2^2 - p_4(1 - 2p_3 + Np_4)]
\end{aligned}
$$

Again, the variance of P(n) is a recursively-defined function of the variance of a_n;

$$\mathrm{Var}[P(0)] = 0$$

$$\mathrm{Var}[P(n)] = \mathrm{Var}[P(n-1)] + 4(p_5)^2\,\mathrm{Var}[a_n]$$

Balanced parameters ($p_1 = p_3$, $p_2 = p_4$) will result in zero variance in the 0^{th} period and constant variance in all subsequent periods. Imbalanced parameters generate a monotonically increasing/decreasing sequence {Var[P(n)]} as $n \to \infty$; therefore, parameter balance is required for variance stationarity as well. The following conclusions regarding the "ant trader" model are consequences of our expressions for the expected value and variance of the P(n) terms.

(i) $\Delta P(n) = P(n) - P(n-1) = p_5 c_n = p_5(2a_n - N)$ from the definition of the agent-based model.

(ii) $E[\Delta P(n)] = 0$ with balanced parameters;

(iii) $E[\Delta P(n)]$ is non-constant with imbalanced parameters.

(iv) $\mathrm{Var}[\Delta P(n)] = 4(p_5)^2 \mathrm{Var}[a_n] = c$ with balanced parameters (c = constant).

(v) $\mathrm{Var}[\Delta P(n)]$ is non-constant with imbalanced parameters.

(vi) Therefore, the {Δ P(n)} series is non-stationary with imbalanced parameters.

Our simulation results indicate that the $\{a_n\}$ series is (approximately) normally distributed with mean $\mu=(N/2)$, as is $\{\Delta a_n\}$ with mean $\mu=0$, given that the parameters are balanced. As a result, $\{\Delta p_n\}$ is also normally distributed; again, this only applies for balanced parameters, which cannot be used with non-zero trend. Please see Figures A.1, A.2, and A.3 in the appendix.

Our "ant trader" model has the following theoretical properties:

Fig. 4.1: Stationarity of agent-based model (based on parameters).

Series	Balanced	Imbalanced
{P(n)}	stationary	non-stationary
{ΔP(n)}	stationary	non-stationary

Since both indices have sample means that are not equal to their initial values, imbalanced parameters are required in order to match this long-run upward trend. Consequently, this implies that the stationarity of first-differences is the decisive difference between the EMH-based random walk model and the ant trader

model; stationarity of the non-differenced series will not be decisive. If $E[\rho(P(n), P(n-1))] \neq 0$, then only the ant trader model can adapt via parameter fitting and the random-walk model is inconsistent with the data. The need to raise the asset (index) price if buyers outweigh sellers in period *t* is an expression of simple supply and demand equilibration; the price adjusts according to investor sentiment in order to clear the market in every period.

V. Algorithm Implementation and Validation

Both the random-walk model and the ant trader model were implemented in *Mathematica* 5.2 and 6.0 environments. Approximately ten independent notebooks were written for various tasks, including model validation and the recording of simulation data for later analysis. The following pseudocode attempts to summarize how this was accomplished, starting with the random-walk model.

RWM Algorithm

Clear all variables (s, k, δ);

Define s, k, δ (starting point, length of simulation, drift term);

Draw k terms from the sequence $\{s_n\}$ such that $s_o = s$, $s_{n+1} = s_n + \delta + NID[0,1]$;

Manually export $\{s_n\}$ to *Stata* 10, compute Mean[$\{s_n\}$] and Variance[$\{s_n\}$]};

Go to start until a set number of entries have been exported.

ABM Algorithm

Clear all variables (N, k, s, $p_1 - p_5$);

Define s, k, N (starting point, simulation length, number of traders);

Define $p_1, p_2, \ldots, p_5$ (as in section IV);

Draw k terms from the $\Delta P(n)$ sequence $\{r_n\}$ such that $r_o = 0$, $r_{n+1} = f(r_n)$, where f operates on r_n;

Define the P(n) sequence $\{s_n\}$ such that $s_o = s$, $s_{n+1} = s_n + r_n$;

Append data file "ABMdata.txt" with new entry {parameters, Mean[$\{s_n\}$], Variance[$\{s_n\}$]};

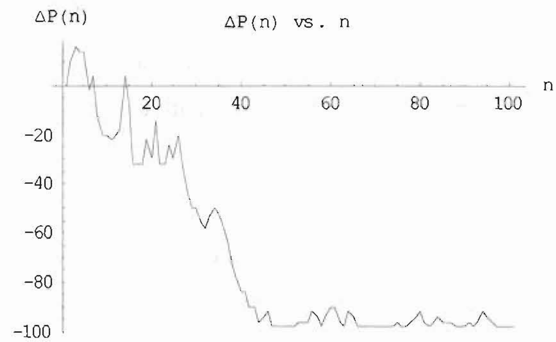
Go to start until a set number of entries have been written to the data file.

In the agent-based algorithm, the function f takes two binomial draws using the previous period's value for a_n, the bullish group. Given $\Delta P(n)$, a_n can be backwards-constructed[16] using parameter values. Both algorithms were designed to run repeatedly in order to generate a sufficient number of end results for statistical testing. The purpose of the data file is to record all simulation results in a text format that is easy for *Stata* 10 to interpret. Please see Section XII (Appendix II) for more information.

[16] With full knowledge of parameter values, a_n can be derived using the definitions established previously.

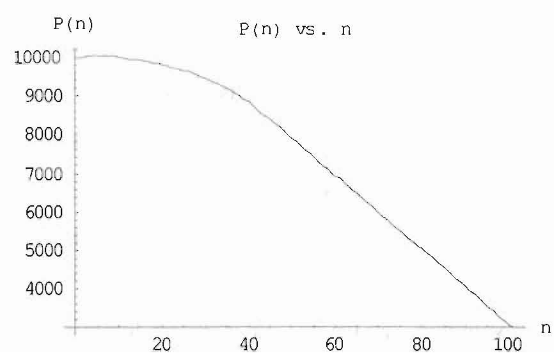
While the random-walk model code was straightforward and could be validated by inspection, the agent-based model was evaluated through the use of two illustrative cases: no random switching and no persuasive switching. In the case of no random switching, we set $p_1 = p_3 = 0$ and expect that one group will come to dominate the other if the absolute magnitudes of p_2 and p_4 are large enough, given that p_2 and p_4 are equal. Let $a_0 = 50$, $b_0 = 50$, $N = 100$, $s_o = 10000$, $k = 100$, $p_2 = p_4 = 0.005$, and $p_5 = 1$; in this situation, p_2 and p_4 can be considered large as they are half their maximum allowable values. The specific simulation that follows is typical of the class of results with this type of parameter scaling. Figures 5.1 and 5.2 plot $\Delta P(n)$ and $P(n)$ versus time index n with no random switching, high persuasion.

Fig. 5.1: $\Delta P(n)$ vs. n; no random, high persuasive component.



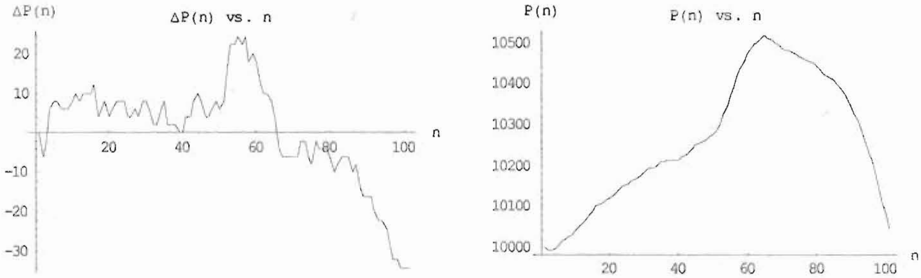
Although the parameter scheme does not suggest a long-run trend for $\Delta P(n)$, the series becomes stuck near $\Delta P(n) = -100$ around $n = 40$. As the majority of traders cluster in one group with $b_n \rightarrow 100$, it becomes increasingly unlikely that the other group, bullish a_n, will regain agents. This persuasive effect, coupled with the lack of random switching to restore the system to equilibrium ($a_n = b_n$), suggests that the system will become fixed at one extreme ($a_n = 100$) or the other ($a_n = 0$). Such an occurrence is most likely when the persuasive parameter is large, because the popular group becomes "locked in" by its own success in gaining the majority share of available members (N).

Fig. 5.2: P(n) vs. n; no random, high persuasive component.



Since $a_n \approx 0$ with $n \geq 40$, as seen above, the price series P(n) achieves its steepest rate of descent at roughly 100 index points lost per time period. Consequently, the index has become completely devalued after a relatively short period of time. When p_2 and p_4 are scaled down by 1-2 orders of magnitude, the result is much less predictable; index price P(n) stays within a few thousand points of its initial value s_o after 100 iterations, and the net price change is near zero. Both of these results correspond to what we would expect theoretically from the agent-based model: the random switching component adds noise and tends to bring the system back to equilibrium, while the persuasive component pulls the system away from equilibrium and represents self-reinforcing momentum investing. In the absence of random switching, the agent-based model spends most of its time in a disequilibrium state; however, this effect is significantly weakened when the persuasive parameters are made less important. Figure 5.3 attempts to show, using a representative simulation result, what happens when p_2 and p_4 decrease by an order of magnitude ($p_2 = p_4 = 0.0005$). Again, this result typifies the entire class of possibilities.

Fig. 5.3: P(n) vs. n; no random, low persuasive component.



Bullish group a_n stays within 15 traders of its initial equilibrium allocation; as a result, the price series P(n) changes much more slowly, which is again consistent with our theoretical expectations.

When persuasive switching is removed ($p_2 = p_4 = 0$), the agent-based model is qualitatively analogous to the random-walk case. The ΔP(n) series reverts to higher-frequency[17] white noise, with the noise increasing in amplitude as p_1 and p_3 increase, assuming that $p_1 = p_3$. Noise in ΔP(n) translates to added noise in the price series P(n) by definition. Therefore, as the random switching parameters increase, possible movement away from initial price s_o increases. The persuasive component of ABM is the central assumption that sets apart the random-walk and agent-based approaches. As such, the proposition that the agent-based model loses its uniqueness when the persuasion parameters are removed is not unreasonable. Without rigorous proof, such a characterization is useful for assessing our untested model. Without going into further detail, we can claim that the agent-based model responded to the hypothetical situations presented by the case studies as expected, which suggests that the model is functioning as defined.

[17] When compared to the other case studies in this section.

VI. Data

Hang Seng and Nikkei 225 stock index data were obtained over the periods 1987-2007 (5080 obs.) and 1984-2007 (5797 obs.); summary statistics, bivariate correlations, and graphical analysis follow here and in the appendix. The two indices in question were chosen for their instability during the period, coupled with the fact that a minority of authors has used data from Asian markets before and after the 1997 financial crisis; i.e. there is sufficient variation for the models to explain. The following tables (Tables 6.1-6.5; Figures 6.1-6.2) characterize the data sets in terms of variable definitions, summary statistics, bivariate correlations, and the density of returns.

Table 6.1: Variable definitions.

Variable	Units	Definition
Returndaily	unitless	Daily percentage change in index price;
Open	HKD/Yen	Index price at open of trading session;
High	HKD/Yen	Highest index price during trading day;
Low	HKD/Yen	Lowest index price during trading day;
Close	HKD/Yen	Index price at close of trading session;
Change	HKD/Yen	Daily change in index price, unweighted;
Volume	shares	Number of shared traded during trading day.

Note that Returndaily is defined as the percentage change in index price; only active trading days were recorded in the data set, so weekends and holidays are excluded. Therefore, bias is introduced since we would expect above-average volatility following weekends as new information needs to be integrated into stock prices.

Table 6.2: Summary statistics for the Hang Seng index variables.

Variable	Obs.	Mean	Std. Dev.	Variance	Skewness	Kurtosis
Returndaily	5079	0.0549875	1.667655	2.781074	-1.929032	48.46569
Open	4516	10559.73	4562.354	$2.08*10^7$	0.0279019	2.569677
High	4516	10636.03	4588.401	$2.11*10^7$	0.0207726	2.566115
Low	4516	10474.56	4527.343	$2.05*10^7$	0.0331988	2.572902
Close	5080	9700.088	4940.557	$2.44*10^7$	0.1162412	2.233254
Change	5080	0.0552217	1.667572	2.780795	-1.929172	48.46553
Volume	1496	451774.7	460654.9	$2.12*10^{11}$	3.358333	17.3229

Hang Seng daily return is skewed to the right (negative skewness), thus the distribution is highly peaked and asymmetric.

Table 6.3: Summary statistics for the Nikkei 225 index variables.

Variable	Obs.	Mean	Std. Dev.	Variance	Skewness	Kurtosis
Returndaily	5796	0.0166975	1.364516	1.861903	0.0557901	10.14701
Open	5146	18592.2	6449.565	$4.16*10^7$	0.7960269	3.457427
High	5146	18723.54	6472.333	$4.19*10^7$	0.7852155	3.436237
Low	5146	18450.42	6422.845	$4.13*10^7$	0.805568	3.473681
Close	5797	17921.21	6408.639	$4.11*10^7$	0.9091101	3.604324
Change	5797	0.0166881	1.364399	1.861584	0.0558421	10.14902
Volume	1275	75584.78	187426.1	$3.51*10^{10}$	2.328616	7.45228

Daily return of the Nikkei has no skew, is relatively symmetric about zero, and is peaked.

Table 6.4: Hang Seng bivariate correlation coefficients.[18]

Hang Seng	Close	Close1d	Close2d	Close3d	Returndaily	Return1d	Return2d	Return3d	Volume	Time
Close	1									
Close1d	0.0234	1								
Close2d	-0.0016	0.6918	1							
Close3d	0.0007	-0.3709	-0.854	1						
Returndaily	0.0072	0.883	0.6148	-0.333	1					
Returndaily1d	-0.001	0.6069	0.8833	-0.7573	0.6943	1				
Returndaily2d	-0.0008	0.3291	0.7542	-0.8841	0.3766	0.8557	1			
Returndaily3d	-0.0012	0.1442	0.5227	-0.7964	0.1649	0.5934	0.9014	1		
Volume	0.7354	-0.0094	-0.025	0.0049	-0.0013	-0.0192	-0.0059	-0.0009	1	
Time	0.8819	0.0074	-0.0004	-0.0003	-0.0073	-0.0001	0.0001	0.0001	0.5814	1

Table 6.5: Nikkei 225 bivariate correlation coefficients.

Nikkei 225	Close	Close1d	Close2d	Close3d	Returndaily	Return1d	Return2d	Return3d	Volume	Time
Close	1									
Close1d	0.0164	1								
Close2d	0.0003	0.7081	1							
Close3d	-0.001	0.3865	0.8566	1						
Returndaily	0.0133	0.953	0.6821	0.3783	1					
Returndaily1d	0.0005	0.6752	0.955	0.8226	0.7131	1				
Returndaily2d	-0.0007	0.3717	0.8177	0.9556	0.3973	0.86	1			
Returndaily3d	-0.0011	0.1753	0.575	0.863	0.1914	0.6088	0.9059	1		
Volume	-0.4401	-0.0506	0.0067	0.003	-0.0597	0.0083	0.0039	0.004	1	
Time	-0.4416	-0.0127	-0.0007	-0.0003	-0.0131	-0.0007	-0.0003	-0.0002	-0.592	1

[18] Note: Close1d = first difference of closing price, Returndaily1d = first difference of daily return, etc.

The correlation coefficients between Returndaily and its differences (Returndaily1d, Returndaily2d, and so on) are negatively related to difference number for both indices, as expected, in Tables 6.4 and 6.5. Upon examination of Figures 6.1 and 6.2, we cannot claim that daily returns are normally distributed; this contradicts the efficient markets hypothesis. Please see the appendix for additional graphical analysis (Figures A.4-A.9).

Figure 6.1: Distribution of Hang Seng daily return.[19]

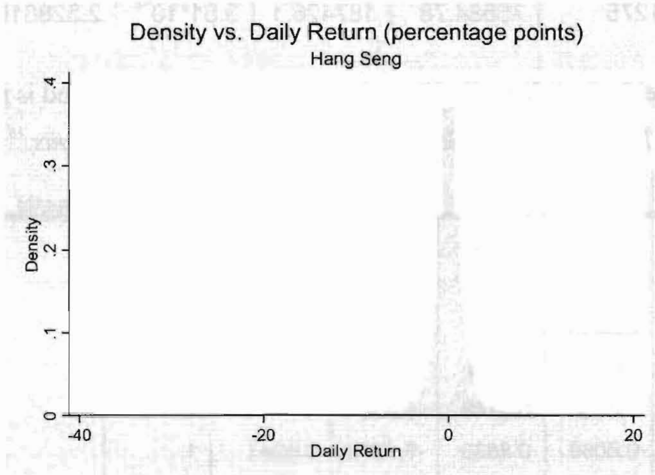
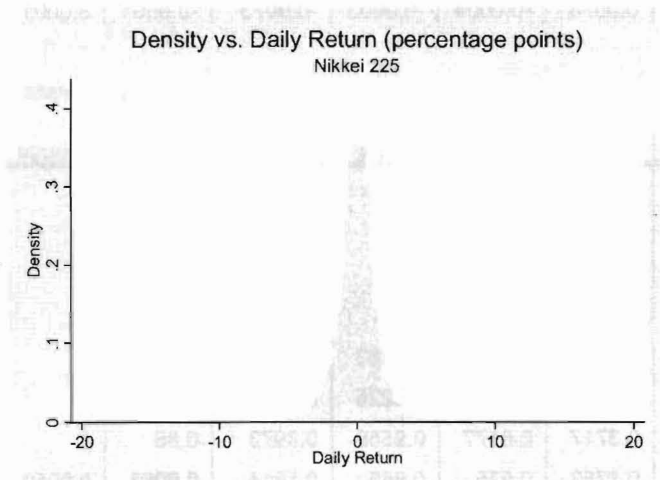


Figure 6.2: Distribution of Nikkei 225 daily return.[20]



[19] Hang Seng returns are not normally distributed; the Shapiro-Wilk normality test indicates that, with the null hypothesis of non-normality, P>Z=0.000. The Shapiro-Francia test agrees with this result.

[20] Nikkei 225 returns are also not normally distributed when using the same normality tests.

VII. Statistical Testing Procedure

The efficient market hypothesis will be examined through three batteries of econometric testing: stationarity/unit roots[21] (ADF, PP tests), autocorrelation, and randomness (Wald-Wolfowitz, Lo-MacKinlay tests). A random walk model has certain statistical properties that can be tested for empirically: non-stationarity of the untransformed series, stationarity of differenced series (or the log return series), and independence of successive values. The drift term δ will be used to approximate the linear trend in fundamental value without actually attempting a regression that tries to derive fundamental value for a number of reasons; previous attempts at predicting "true" fundamental value have not been entirely fruitful due to a number of innate causal reasons, such as stock prices themselves causing changes in fundamental value and the use of stock prices as a macroeconomic indicator. Drift δ will be selected such that final price matches the expectation value of the last term of the simulated series.

The ADF and PP tests assume the null hypothesis of non-stationarity (the existence of a unit root); the alternative hypothesis is stationarity (Dickey and Fuller 1979; Phillips and Perron 1988). In the Wald-Wolfowitz non-parametric test for randomness, serial independence of the data, i.e. randomness, is assumed as the null; the alternative is therefore non-randomness. A "run" is defined as a sequence of successive values that are all above or below the median (by default; the mean can also be used at the decision point). The null hypothesis of the WW test is rejected when the number of runs in the data is small relative to the sample size, which implies that the runs present are longer than would be likely under the serial independence assumption (Wald and Wolfowitz 1943). The Lo-MacKinlay variance ratio test partitions the data set into increments (subintervals) based on aggregation values $q = 2, 4, 8, 16$ and computes the variance of each increment. The Lo-MacKinlay test assumes that the subintervals are homoscedastic under the random walk null hypothesis, which implies that the increments should have similar variances; the alternative is a non-random walk process. After taking a variance ratio that compares the largest incremental variance to the variance of the data set as a whole, the test accepts or rejects the null hypothesis based on the test statistic $z^*(q)$. Therefore, the conclusion will be subject to the value of q, which is an externally-determined parameter (Lo and Mackinlay 1988).

Simulation data derived from the ant trader model will be subject to stationarity tests, the examination of autoregressive traits (visual inspection of autocorrelograms), and randomness testing. We expect to find that the untransformed series and first differences in price are non-stationary, and significant autoregressive behavior can only be explained by the agent-based model. Empirical modeling in this case is the application of previously discussed theoretical models to the specific case of Nikkei 225

[21] ADF: Augmented Dickey-Fuller test, PP: Phillips-Perron test; both are stationarity tests.

(1984-2007) and Hang Seng (1987-2007) index prices using parameter estimation by fitting expected mean and variance to the data. Statistical testing, with the exception of the Lo-MacKinlay procedure, will be executed with built-in *Stata* 10 routines. Specifically, a *Stata* module implementation of the Lo-MacKinlay variance ratio test written by Baum in 2006 will be used[22].

VIII. Results

We first need to obtain a theoretical estimate for the random walk model's mean and variance. Using this estimate, E[μ] from the model can be matched to the sample mean in real-world cases (Hang Seng and Nikkei 225). Regarding the agent-based model, only a large-sample simulation could provide the necessary relationship between parameter values and μ or σ^2, so such a simulation is attempted. Given the poor results of this effort, *ceteris paribus* studies are done in which a selected parameter is allowed to vary within ±5% of its initial value. Next, E[μ] and E[σ^2] are derived for the agent-based model using the definitions established in Section IV; again, the purpose of this is to match E[μ] to the sample mean by varying the model's parameters: p_1, p_2, …, p_5. Finally, stationarity and randomness testing is done using *Stata* 10 in order to determine which model best reflects market conditions for the two indices.

The random walk with drift and "ant trader" models are fit to the data using the first and second central moments of mean and variance, respectively. Therefore, we need to obtain E[μ] and E[σ^2] in each case in order to plausibly match real-world behavior. For the random walk model, these expected values can be found analytically; *k* stands for the number of model iterations.

$$
\begin{aligned}
E[\mu] &= E[\tfrac{1}{k} \textstyle\sum_{n=0}^{k} a_n] \\
&= \tfrac{1}{k} \textstyle\sum_{n=0}^{k} E[a_n] \\
&= \tfrac{1}{k} \textstyle\sum_{n=0}^{k} (p_0 + n\delta) \\
E[\mu] &= p_0 + \frac{(k+1)\delta}{2}
\end{aligned}
$$

$$
\begin{aligned}
E[\sigma^2] &= E[\tfrac{1}{k-1} \textstyle\sum_{n=0}^{k} (a_n - \mu)^2] \\
&= \tfrac{1}{k-1} \textstyle\sum_{n=0}^{k} (E[a_n] - E[\mu])^2 \\
&= \tfrac{\delta^2}{k-1} \textstyle\sum_{n=0}^{k} \left(n - \frac{n-1}{2}\right)^2 \\
&= \ldots \\
E[\sigma^2] &= \frac{\delta^2 (2k^2 - 5k - 6)(k+1)}{24(k-1)}
\end{aligned}
$$

Since *k* and p_0 are taken as given based on the data, we can only change the drift term δ in order to retain consistency with the sample mean and variance. The assumption made here was that the random (non-drift) component is normally distributed with mean 0 and variance 1; additional flexibility requires

[22] Baum's module, revised as of 2007, is available at: http://ideas.repec.org/c/boc/bocode/s456740.html.

changing these parameters, which will not be considered here. δ will be selected to match the sample mean, so this current scheme allows for no control over variance.

Simulations were run using the agent-based model with randomly drawn parameter values uniformly distributed on finite intervals. p_0 and *k* were fixed based on the index used (Hang Seng or Nikkei); N was fixed at 100; p_1, p_2, p_3, p_4, and p_5 were allowed to vary as Figure 8.1 shows.

Fig. 8.1: Allowed parameter ranges, m=2450 simulation.

$$p_1, p_3 \in [0.01, 0.1]$$
$$p_2, p_4 \in [0.0005, 0.005]$$
$$p_5 \in [1, 10]$$

The simulation program, executed in *Mathematica* 6.0, calculated the first four central moments (mean, variance, skewness, and kurtosis) for each trial. Each trial was represented by a line in the data file, which recorded the central moments coupled with relevant parameter values, as shown in Figure 8.2.[23]

Fig. 8.2: Mathematica-generated data file.

ABMdata.txt :

$$\{p_{11}, p_{21}, p_{31}, p_{41}, p_{51}, \mu_1, \sigma^2_1, \gamma_{11}, \gamma_{21}\}$$
$$\{p_{12}, p_{22}, p_{32}, p_{42}, p_{52}, \mu_2, \sigma^2_2, \gamma_{12}, \gamma_{22}\}$$
$$\ldots$$
$$\{p_{1m}, p_{2m}, p_{3m}, p_{4m}, p_{5m}, \mu_m, \sigma^2_m, \gamma_{1m}, \gamma_{2m}\}$$

We then attempted to regress these moments on the recorded parameter values for each draw using *Stata* 10. The following results are for the Nikkei 225 simulation with $p_0 = 9927$, $k = 5796$, $N = 100$, $p_5 = 5$ fixed for all trials; $m = 2450$ total trials were run, so the data file had 2450 lines. Sample mean was regressed on parameter values p_1, p_2, p_3, and p_4 initially; the result was surprising as no variables are significant at the 0.10 level, although p_4 is close (Table A.5). However, even this weakly-significant variable is contradictory because all parameters should matter theoretically, according to our definitions. Regressing variance on the set of parameter values $\{p_1, p_2, p_3, p_4\}$ generated a similar result; the constant term was the most significant, indicating high variance regardless of parameters (Table A.6). We did achieve significance of p_2 at the 0.10 level; again, the other parameters are not significant. This implies that parameters reduction in the agent-based model may be possible, as only a subset of the available parameters is important in explaining simulation mean and variance. Please see the appendix for a full listing of attempted regressions (Tables A.1-A.6).

---

[23] γ_1 = skewness, γ_2 = kurtosis.

This null result[24] implied that our model specification was incorrect. The probabilistic definition of the "ant trader" model suggested that only the relative parameter values were important, i.e. the ratios Randomratio:= p_3 / p_1 and Persuasratio:= p_4 / p_2. Independent variables Randomdiff:= $p_3 - p_1$ and Persuasdiff:= $p_4 - p_2$ could have explanatory power if the difference in parameters was influential. Although most regressions did not assign significance to these new variables, the persuasion ratio was significant at α=0.05 when predicting sample mean. Our simulation results yielded one definite conclusion: model outcomes "explode" when the parameters become imbalanced; Highly unequal parameters ($p_1 >>> p_3$, $p_2 >>> p_4$, or vice versa) lead to large sample means of $\pm 10^6$ or more. A plot of frequency vs. mean for the simulation data (m = 2450) clearly shows a twin-peaked distribution that is skewed away from zero and towards extreme values (please see Appendix, Figure A.10). These mixed results suggest that another analytical tool is necessary, specifically variation of parameters around an arbitrary starting point, in order to maintain relative stability while simultaneously exploring individual parameter effects on the central moments.

Given a starting point of $\{p_1, p_2, p_3, p_4, p_5, N\} = \{0.05, 0.001, 0.05, 0.001, 5, 100\}$, parameters p_1, p_2, p_3, and p_4 were varied individually with tolerance ± 5% *ceteris paribus*. Initial price p_o and duration k (total number of iterations) were specified according to real-world Hang Seng and Nikkei 225 data. Parameter values used are specified in Table 8.3, as shown below:

Table 8.3: Allowed ranges, variation of parameters.

Parameter	Hang Seng	Nikkei 225
p_1	0.05 ± 5%	0.05 ± 5%
p_2	0.001 ± 5%	0.001 ± 5%
p_3	0.05 ± 5%	0.05 ± 5%
p_4	0.001 ± 5%	0.001 ± 5%
p_5	5	5
N	100	100
p_0	2583	9954
k	5080	5797

Samples of size n=50 were used for each set of parameters, i.e. fifty runs of the model each time, where the parameters $p_1, \ldots, p_4$ were discretely varied from p_x-5% to p_x+5% in intervals of 1%. The averages of the first four central moments were recorded for each parameter set across the fifty model runs. Results were graphed with SSE-minimizing linear (sample mean case) and polynomial of order 2 (sample variance case) interpolating functions. We will consider the $E[\mu]$ vs. p_2 case as representative, and the remainder of the graphs can again be found in the appendix (Figures A.11-A.20). In this case, simulation

[24] These unsuccessful regression results have no bearing on the remainder of our simulations or statistical tests.

results indicated that E[μ] is indirectly related to p_2 and therefore directly related to p_4; the relationships in Figure 8.3 and Figure 8.4 are subsequently linear.

Fig 8.3: E[μ] vs. p_2, Hang Seng model.

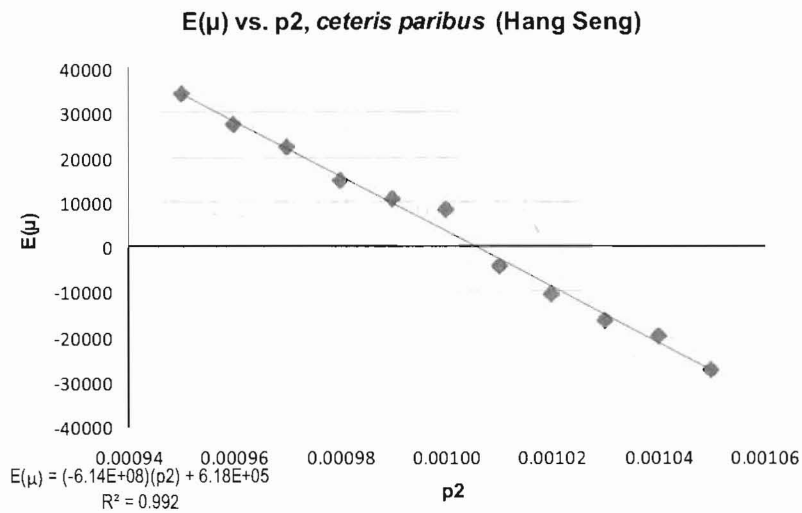
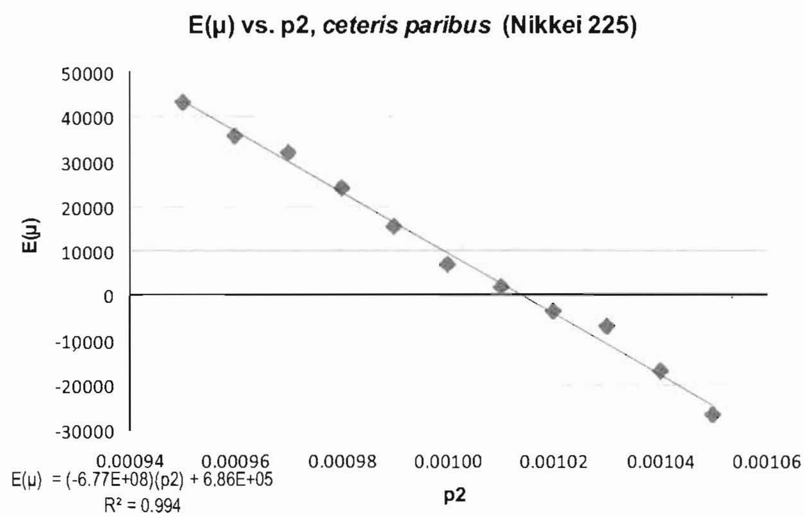


Fig 8.4: E[μ] vs. p_2, Nikkei 225 model.



Variance was found to be directly related to the difference $|p_4 - p_2|$; increased deviation of p_2 from the fixed value of p_4 results in an exponential increase in the sample variance. As expected, p_1 was also found to be indirectly related to E[μ], which implies that p_3 is directly related in an analogous fashion. The difference $|p_3-p_1|$ affects sample variance directly, behaving in the same way as $|p_4-p_2|$. Parameter p_5 was not a significant predictor of sample mean or variance within the 5% tolerance.

We can rely on our explicit forms of $E[a_n]$ and $Var[a_n]$ in order to compute the needed quantities $E[\mu]$ and $E[\sigma^2]$; the rest follows directly in Figure 8.5.

Fig. 8.5: Expectation values for first and second (central) moments, "ant trader" model.

$$E[\mu] = \frac{1}{k}\sum_{n=0}^{k} E[a_n]$$

$$E[\sigma^2] = \frac{1}{k-1}\sum_{n=0}^{k} (E[a_n] - E[\mu])^2$$

Recall that $E[a_n]$ has been recursively defined previously. Therefore, given a parameter set, we can use *Mathematica* to solve for these expectation values, which will be set equal to sample mean and variance. Now that $E[\mu(p_1,p_2,...)]$ and $E[\sigma^2(p_1,p_2,...)]$ are well-defined for both the random walk and agent-based models, we can proceed to parameter selection, point-by-point simulation, stationary testing, and comparison to the actual data. Table 8.4 lists the parameters selected for each model, based on the index, in order to match the first two central moments; sample mean was prioritized over variance. Thirty trials will be computed for each set of parameter values, and stationarity/randomness tests will be individually applied to each run. Augmented Dickey-Fuller and Phillips-Peron unit root tests are used in combination with the Wald-Wolfowitz runs test and the Lo-MacKinlay variance ratio test. All routines were performed in *Stata* 10, with the results reported in Tables 8.5 – 8.6. The ADF test has no lags or drift/trend term, and the PP test uses the default number of lags (nearest integer value of $4(k/100)^{2/9}$, where the series is k periods long (k iterations in the model).

Table 8.4: Parameters for statistical testing.

Parameter	Hang Seng (RWM)	Nikkei 225 (RWM)	Hang Seng (ABM)	Nikkei 225 (ABM)
δ	2.8014	2.7568		
p_1			0.04944	0.04945
p_2			0.001	0.001
p_3			0.05	0.05
p_4			0.001	0.001
p_5			5	5
N			100	100
p_0	2583	9954	2583	9954
k	5080	5797	5080	5797

In the results presented over the next three pages, the simulation program was run without checking for non-negativity of the price series at each iteration. Since negative price is nonsensical and cannot be transformed via logarithms, these results will be referred to as "possibly negative" whereas the trials that were later hand-picked for non-negativity will be labeled "non-negative".

Table 8.5: Stationarity tests, Hang Seng Index, $\{p_n\}$ series, possibly negative.[25]

Run	ABM: ADF	ABM: PP	RWM: ADF	RWM: PP
1	0.900	0.767	0.932	0.932
2	0.997	0.890	0.708	0.709
3	0.000	0.134	0.950	0.950
4	0.800	0.724	0.510	0.517
5	0.980	0.804	0.966	0.967
6	0.995	0.892	0.924	0.925
7	0.847	0.787	0.292	0.298
8	0.992	0.858	0.974	0.974
9	0.627	0.753	0.982	0.982
10	0.522	0.653	0.994	0.995
11	1.000	0.993	0.552	0.573
12	0.163	0.571	0.881	0.881
13	1.000	0.994	0.993	0.992
14	0.906	0.789	0.990	0.991
15	0.997	0.953	0.999	0.999
16	0.965	0.861	0.944	0.945
17	0.995	0.662	0.581	0.605
18	0.726	0.750	0.962	0.962
19	1.000	0.998	0.956	0.956
20	0.061	0.246	0.977	0.977
21	1.000	0.986	0.991	0.992
22	0.413	0.510	0.969	0.969
23	0.829	0.773	0.267	0.253
24	1.000	1.000	0.798	0.787
25	0.980	0.654	0.850	0.846
26	0.986	0.863	0.997	0.997
27	0.971	0.930	0.995	0.995
28	0.869	0.611	0.986	0.986
29	0.858	0.450	0.951	0.951
30	0.711	0.572	0.986	0.986

[25] Null hypothesis: existence of a unit root (non-stationary), p-values are reported, 5% critical value: -2.86.

Table 8.6: Stationarity tests, Nikkei 225 Index, $\{p_n\}$ series, possibly negative.

Run	ABM: ADF	ABM: PP	RWM: ADF	RWM: PP
1	0.995	0.940	0.979	0.979
2	0.901	0.639	0.751	0.759
3	0.000	0.027	0.971	0.971
4	0.090	0.402	0.632	0.623
5	0.964	0.788	0.996	0.996
6	0.789	0.694	1.000	1.000
7	0.998	0.972	0.993	0.993
8	0.991	0.903	0.988	0.988
9	0.997	0.893	0.995	0.995
10	0.187	0.394	0.706	0.716
11	0.996	0.928	0.994	0.993
12	0.997	0.890	0.973	0.972
13	0.994	0.880	0.912	0.913
14	0.076	0.576	0.776	0.784
15	0.988	0.952	0.877	0.876
16	0.050	0.519	0.960	0.960
17	0.100	0.367	0.992	0.992
18	1.000	0.998	0.988	0.987
19	0.000	0.025	0.891	0.889
20	0.926	0.812	0.977	0.977
21	0.963	0.883	0.984	0.984
22	0.688	0.578	0.976	0.977
23	0.001	0.173	0.923	0.924
24	0.788	0.750	0.948	0.948
25	1.000	0.997	0.944	0.944
26	0.375	0.481	0.521	0.526
27	1.000	0.994	0.951	0.951
28	0.998	0.955	0.584	0.627
29	0.000	0.257	0.863	0.862
30	0.005	0.154	0.893	0.892

The ADF and PP test results for the differenced series all had p-values of 0.000 or less; therefore, the first differences are stationary without exception in all cases. One possible explanation for this result regarding the agent-based model, as this was expected for the random walk model, is that our parameters were too close to the balanced case to make much of an impact. Test results imply that the ant trader model was biased towards stationarity even in the price series, which is a surprising result that goes against our expectations and the empirical realities of the data. Table 8.7 attempts to summarize the results of the stationarity tests; the major implication is that the random walk model was consistent with the actual data. Since the actual indices were non-stationary in the $\{p_n\}$ series and stationary in the $\{\Delta p_n\}$ series, the agent-based model presents a problem when its $\{p_n\}$ terms are stationary. Although the agent-based model generated a more realistic estimate for test p-values in the N225 case, the random walk model is perfectly accurate in predicting stationarity.

Table 8.7: Stationarity test results, summary, model comparison, possibly negative.

n=30 (per model)		ADF: {p} series	PP: {p} series	ADF: {Δp} series	PP: {Δp} series
Hang Seng		0.863	0.851	0.000	0.000
Nikkei 225		0.425	0.451	0.000	0.000
HSI: ABM average		0.803	0.748	0.000	0.000
HSI: RWM average		0.862	0.863	0.000	0.000
N225: ABM average		0.629	0.661	0.000	0.000
N225: RWM average		0.898	0.900	0.000	0.000
% stationary, ABM	$\alpha = 0.10$	6.67%	0%	100%	100%
HSI	$\alpha = 0.05$	3.33%	0%	100%	100%
	$\alpha = 0.01$	3.33%	0%	100%	100%
% stationary, RWM	$\alpha = 0.10$	0%	0%	100%	100%
HSI	$\alpha = 0.05$	0%	0%	100%	100%
	$\alpha = 0.01$	0%	0%	100%	100%
% stationary, ABM	$\alpha = 0.10$	30.00%	6.67%	100%	100%
N225	$\alpha = 0.05$	20.00%	6.67%	100%	100%
	$\alpha = 0.01$	16.67%	6.67%	100%	100%
% stationary, RWM	$\alpha = 0.10$	0%	0%	100%	100%
N225	$\alpha = 0.05$	0%	0%	100%	100%
	$\alpha = 0.01$	0%	0%	100%	100%

Autocorrelation plots of returns can provide qualitative information regarding the adherence of a model to the empirical ideal. Again, the random walk model outperformed the ant trader model in this dimension of comparability; please see the appendix, Figures A.21-A.24. The RWM generates alternating, seemingly random correlation coefficients between the n^{th} period and previous periods, which is in accordance with empirical reality. However, the ant trader model exhibits strong autocorrelation in returns which is not present in daily data. We conclude that the random walk model has outperformed the agent-based model in terms of autoregressive behavior.

From this point onward, all results deal with non-negative simulation data. The *Mathematica* program was repeatedly executed until the price series did not fall below 20% of its original value at any point during the run; only those runs which satisfied this requirement were added to the data set. This stipulation was made because we felt that the possibility of negative price should be ruled out entirely, even if the runs selected do have systematic bias away from zero. After recording a total of 120 non-negative trial runs, two different transformations, log price[26] and first differences, were applied to the data in order to generate a stock index return series for testing. Weekly data was taken by reducing the sampling rate to 1/5 in order to compensate for daily trading biases, such as the bid-ask effect and

[26] The log price series is defined by $(\text{Log price})_t = \text{Ln}[(P_t)/(P_{t-1})]$. This transformation works properly only if the series is non-negative and far enough away from zero.

infrequent trading. Tables 8.8 – 8.11 report the general results obtained; "100% stationary" indicates that the 30 samples for that case were all stationary, i.e. the null hypothesis of a unit root (non-stationarity) was rejected at the $\alpha = 0.05$ level for every individual sample. Tables 8.8-8.9 below outline our results when the log price transformation was used.

Table 8.8: Random walk test results, log price series, non-negative.

Hang Seng, LN, α=0.05	ABM	RWM	Actual
Aug. Dickey-Fuller	100% stationary	100% stationary	stationary
Phillips-Perron	100% stationary	100% stationary	stationary
Wald-Wolfowitz	100% non-random	100% non-random	random (p-value=0.27)
Lo-MacKinlay (q = 2)	63.3% random	100% non-random	non-random
Lo-MacKinlay (q = 4)	90% non-random	100% non-random	non-random

Table 8.9: Random walk test results, log price series, non-negative.

Nikkei 225, LN, , α=0.05	ABM	RWM	Actual
Aug. Dickey-Fuller	100% stationary	100% stationary	stationary
Phillips-Perron	100% stationary	100% stationary	stationary
Wald-Wolfowitz	100% non-random	100% non-random	random (p-value=0.62)
Lo-MacKinlay (q = 2)	90% random	100% non-random	non-random
Lo-MacKinlay (q = 4)	100% non-random	100% non-random	non-random

Both the agent-based and random walk models generated stationary series after transformation, which is consistent with our previous results and the actual stock price data. As the real-world price series is random according to the Wald-Wolfowitz test, neither model managed to match this behavior. The Lo-MacKinlay variance ratio test provided conflicting results depending on the aggregation parameter q, but the RWM was in agreement with the actual data regardless of q while the agent-based approach was random in the majority with $q = 2$. Therefore, the log transformation again suggests that the random walk model is more harmonized with the actual data. The Hang Seng and Nikkei indices appear to behave quite similarly insofar as both are efficient according to the WW test and inefficient via Lo-MacKinlay.

The log price series is viewed as the best return specification available, but the results of the Wald-Wolfowitz and Lo-MacKinlay tests, that the random walk series is non-random, are contrary to our expectations. Taking the ratio of $(P_t)/(P_{t-1})$ is problematic in this case because the innovations (random component) in our random walk model were drawn from a normal distribution that did not depend on the magnitude of the series itself. Consequently, as the upward trend term is larger than the innovations, the relative price change will decrease as P_t increases over time. Therefore, $(\text{log price})_t$ heteroscedastically goes to zero as t increases since variance as a function of time declines monotonically as well. In response

to this problem, we decided to compare the log price results to the case of first differences, as shown below in Tables 8.10-8.11.

Table 8.10: Random walk test results, first differences, non-negative.

Hang Seng, 1d, α=0.05	ABM	RWM	Actual
Aug. Dickey-Fuller	100% stationary	100% stationary	stationary
Phillips-Perron	100% stationary	100% stationary	stationary
Wald-Wolfowitz	100% non-random	96.6% random	random (p-value=0.47)
Lo-MacKinlay (q = 2)	100% random	100% non-random	non-random
Lo-MacKinlay (q = 4)	100% non-random	100% non-random	non-random

Table 8.11: Random walk test results, first differences, non-negative.

Nikkei 225, 1d, α=0.05	ABM	RWM	Actual
Aug. Dickey-Fuller	100% stationary	100% stationary	stationary
Phillips-Perron	100% stationary	100% stationary	stationary
Wald-Wolfowitz	100% non-random	90% random	random (p-value=0.62)
Lo-MacKinlay (q = 2)	93.3% random	100% non-random	non-random
Lo-MacKinlay (q = 4)	100% non-random	100% non-random	non-random

Using the first differenced price series, the runs test now concludes that return is randomly distributed around its median for the random walk model, which makes intuitive sense. As a result, the RWM is now even more compatible with the Hang Seng and Nikkei stock index price data. The agent-based results have not changed significantly; the Lo-MacKinlay test result still depends on q when comparing the $q = 2$ and $q = 4$ outcomes. Although the Lo-MacKinlay conclusion differs from that of the runs test in every case, the agent-based model and its ambiguous variance ratio test remains inconsistent with the actual result for both indices. Therefore, despite the contrarian Lo-MacKinlay results, we can again conclude that the random walk model is closer to agreement with the real-world price data using both the log price and first differenced series. We found that the RWM outperforms the agent-based approach in both the negative and non-negative simulation samples. The final result is an unambiguous classification of the Nikkei and Hang Seng markets as weakly efficient, thus the random walk model can better explain stock price behavior in this type of situation.

IX. Concluding Remarks

When compared to the ant trader model, the random-walk model is more consistent with the data available on the Hang Seng and Nikkei 225 stock market indices over the 1987-2007 and 1984-2007 periods, respectively. This result, which is supported using both negative and non-negative simulation samples, cannot be easily generalized to other markets. Various empirical papers have found inconsistent results regarding the efficient-markets theory; some financial markets appear to be efficient in the short/long run, while others are not (Worthington, Higgs 2003). Therefore, our results only apply to the particular situation examined and are possibly a byproduct of the data at hand, which includes the 1997 Asian financial crisis. We did not want to selectively isolate any financial crises in order to make the statistical tests as robust as possible, and an obvious extension of this work is the consideration of different periods of time. It is plausible that the market operates efficiently over certain time scales and not others; an adjustment period immediately following a financial disaster may temporarily inhibit market efficiency, for example. The results in this case imply that efficient-markets theory cannot be challenged on empirical grounds using stationarity when the comparison group is simulation data generated using the ant trader model. Although the EMH assumptions are unpalatable, inflexible, and unrealistic, the resultant simulation data are consistent with actual data when using return autocorrelograms, stationarity (ADF, PP tests), and randomness (Wald-Wolfowitz runs test, Lo-MacKinlay variance ratio test) as cross-model comparative tools. A number of useful properties of the agent-based model have been established regarding mean and variance, and variation of parameters yielded insight into the underlying agent-agent dynamic. We hope to continue to improve on the agent-based approach[27] as manifested in the ant trader model as this concept is still in its infancy when compared to the thirty years of refinement that the random walk model has undergone.

[27] The next step is to allow agent-based model parameters to vary over time in response to generalized exogenous changes in consumer confidence, macroeconomic health, or consumer confidence.

X. References

Arthur, W. Brian; John H. Holland; Blake LeBaron; Richard Palmer and Paul Tayler. 1997. "Asset Pricing Under Endogenous Expectations in an Artificial Stock Market," in *The Economy as an Evolving Complex System, II.* W. Brian Arthur, Steven Durlauf and David Lane, eds. Boston, MA: Addison-Wesley, pp. 15-44.

Baker, Malcolm and Jeffery Wurgler. 2007."Investor Sentiment in the Stock Market." *Journal of Economic Perspectives* 21:2, pp. 129-51.

Bonabeau, Eric. 2002."Agent-Based Modeling: Methods and Techniques for Simulating Human Systems." *Proceedings of the National Academy of Sciences of the United States of America* 99:10, pp. 7280-7.

Cioffi-Revilla, Claudio. 2002."Invariance and Universality in Social Agent-Based Simulations." *Proceedings of the National Academy of Sciences of the United States of America* 99:10, pp. 7314-6.

Diba, Behzad T. and Herschel I. Grossman. 1988."Explosive Rational Bubbles in Stock Prices?" *Am. Econ. Rev.* 78:3, pp. 520-30.

De Bondt, Werner F. M. and Richard Thaler. 1984."Does the Stock Market Overreact?" *Journal of Finance* 40:3, pp. 793-805.

Dickey, David A. and Wayne A. Fuller. 1979."Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74:366, pp. 427-31.

Diks, Cees; Cars Hommes; Valentyn Panchenko and van der Weide, Roy. 2007."E&F Chaos: A User Friendly Software Package for Nonlinear Economic Dynamics." *Center for Nonlinear Dynamics in Economics and Finance:* Working paper 06-15:pp. 9/10/2007.

Gilbert, Nigel and Steven Bankes. 2002."Platforms and Methods for Agent-Based Modeling." *Proceedings of the National Academy of Sciences of the United States of America* 99:10, pp. 7197-8.

Hagerman, Robert L. and Richard D. Richmond. 1973."Random Walks, Martingales and the OTC." *The Journal of Finance* 28:4, pp. 897-909.

Hirshleifer, David. 2001."Investor Psychology and Asset Pricing." *Journal of Finance* 56:4, pp. 1533-97.

Hong, Harrison and Jeremy C. Stein. 2007."Disagreement and the Stock Market." *Journal of Economic Perspectives* 21:2, pp. 109-28.

Huber, Peter. 1995."Random Walks in Stock Exchange Prices and the Vienna Stock Exchange." *Economics Series 2, Institute for Advanced Studies.*

Inchiosa, Mario E. and Miles T. Parker. 2002."Overcoming Design and Development Challenges in Agent-Based Modeling using ASCAPE." *Proceedings of the National Academy of Sciences of the United States of America* 99:10, pp. 7304-8.

Kirman, Alan. 1993."Ants, Rationality, and Recruitment." *The Quarterly Journal of Economics* 108:1, pp. 137-56.

Lo, Andrew W. and A. Craig MacKinlay. 1988."Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." *The Review of Financial Studies* 1:1, pp. 41-66.

Ormerod, Paul. 1998. *Butterfly Economics.* New York, NY: Basic Books.

Pearce, Douglas K. 1984."An Empirical Analysis of Expected Stock Price Movements." *Journal of Money, Credit and Banking* 16:3, pp. 317-27.

Pearce, Douglas K. and V. Vance Roley. 1985."Stock Prices and Economic News." *The Journal of Business* 58:1, pp. 49-67.

Phillips, Peter and Pierre Perron. 1988."Testing for a Unit Root in Time Series Regression." *Biometrika* 7:2, pp. 335-46.

Raines, Patrick J. and Charles G. Leathers. 2000. *Economists and the Stock Market: Speculative Theories of Stock Market Fluctuations.* Cheltenham, UK: Edward Elgar.

Scharfstein, David S. and Jeremy C. Stein. 1990."Herd Behavior and Investment." *Am. Econ. Rev.* 80:3, pp. 465-79.

Schleifer, Andrei. 2000. *Inefficient Markets: An Introduction to Behavioral Finance.* Oxford UK: Oxford University Press.

Sheffrin, Steven M. 1996. *Rational Expectations.* Cambridge, UK: Cambridge University Press.

Shiller, Robert J. 2000. *Irrational Exuberance.* New York, NY: Broadway Books.

Shumway, Robert H. and David S. Stoffer. 2006. *Time Series Analysis and its Applications.* New York, NY: Springer.

Tesfatsion, Leigh. 2002."Economic Agents and Markets as Emergent Phenomena." *Proceedings of the National Academy of Sciences of the United States of America* 99:10, pp. 7191-2.

Thaler, Richard H., ed. 1993. *Advances in Behavioral Finance.* New York, NY: Russell Sage Foundation.

Wald, A. and J. Wolfowitz. 1943."An Exact Test for Randomness in the Non-Parametric Case Based on Serial Correlation." *The Annals of Mathematical Statistics* 14:4, pp. 378-88.

Worthington, Andrew C. and Helen Higgs. 2003."Tests of Random Walks and Markct Efficiency in Latin American Stock Markets: An Empirical Note." *School of Economics and Finance Discussion Papers and Working Papers, Queensland University of Technology* 157:pp. 9/15/2007.

XI. Appendix I

Figure A.1: Density vs. a_n for simulated $\{a_n\}$
(run for 10,000 periods, p_1=p_3=0.05, p_2=p_4=0.001, p_5=5, N=100).

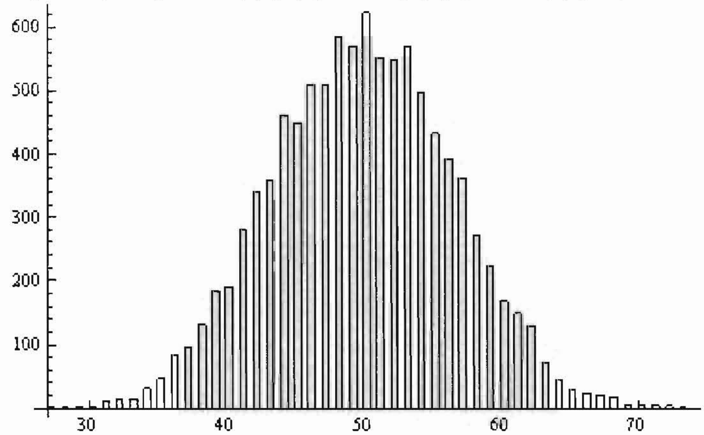


Figure A.2: Density vs. Δa_n for simulated $\{\Delta a_n\}$ (run for 10,000 periods).

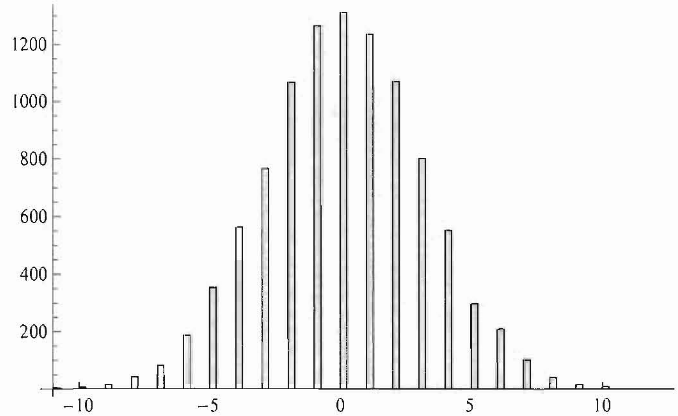


Figure A.3: Density vs. Δp_n for simulated $\{\Delta p_n\}$ (run for 10,000 periods).

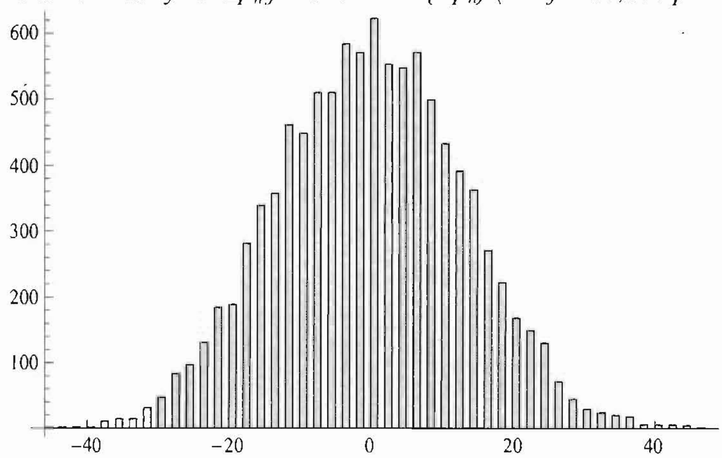


Figure A.4: Weekly return of the Hang Seng index vs. time in trading days: 1987-2007.[28]

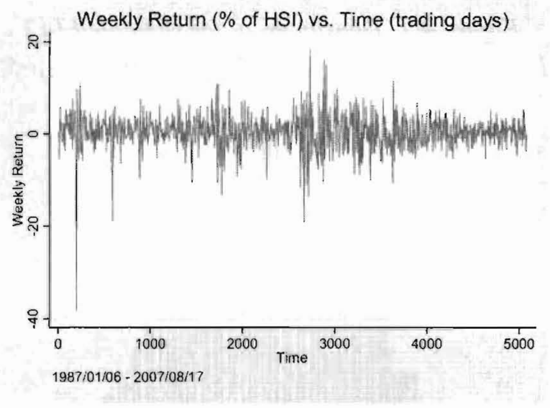


Figure A.5: Weekly return of the Nikkei index vs. time in trading days: 1984-2007.

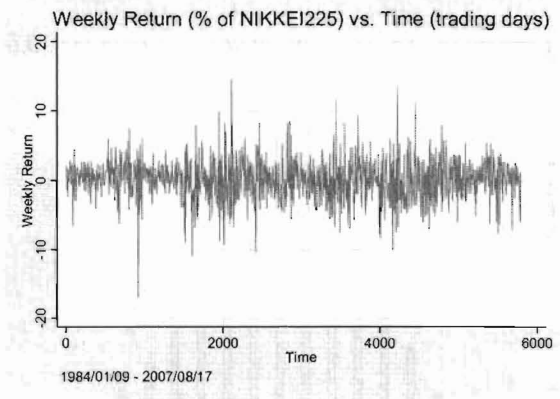


Figure A.6: Daily closing price of the Hang Seng index vs. time in trading days: 1987-2007.

[28] Note: return is calculated as the percentage change of the index over the appropriate time period.

Figure A.7: Daily closing price of the Nikkei 225 index vs. time in trading days: 1984-2007.

Figure A.8: Autocorrelations of daily return for the Hang Seng index, 0-20 period lags.[29]

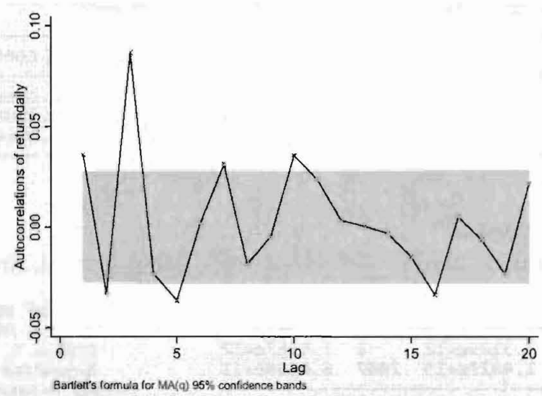
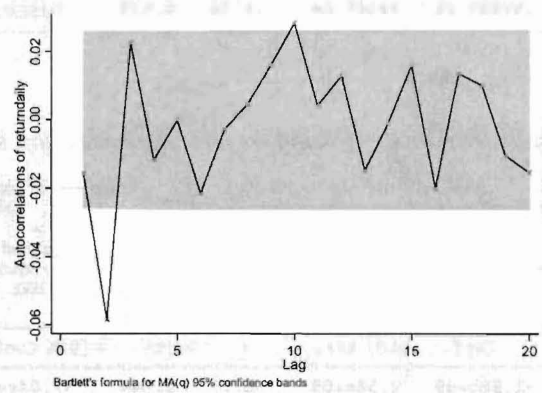


Figure A.9: Autocorrelations of daily return for the Nikkei 225 index, 0-20 period lags.



[29] Note: shaded box indicates 95% confidence band.

Table A.1: Regression results, mean = f(Randomdiff, Persuasdiff), simulated data (m=2450).

Source	SS	df	MS		
Model	1.4067e+12	2	7.0333e+11	Number of obs =	2450
Residual	1.4846e+15	2447	6.0672e+11	F(2, 2447) =	1.16
				Prob > F =	0.3139
				R-squared =	0.0009
				Adj R-squared =	0.0001
Total	1.4860e+15	2449	6.0679e+11	Root MSE =	7.8e+05

mean	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
randomdiff	386892.3	430497	0.90	0.369	-457283.9	1231069
persuasdiff	1.07e+07	8560967	1.25	0.210	-6050560	2.75e+07
_cons	6459.175	15750.16	0.41	0.682	-24425.85	37344.2

Table A.2: Regression results, variance = f(Randomdiff, Persuasdiff), simulated data (m=2450).

Source	SS	df	MS		
Model	6.9231e+22	2	3.4616e+22	Number of obs =	2450
Residual	6.9335e+25	2447	2.8335e+22	F(2, 2447) =	1.22
				Prob > F =	0.2949
				R-squared =	0.0010
				Adj R-squared =	0.0002
Total	6.9404e+25	2449	2.8340e+22	Root MSE =	1.7e+11

variance	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
randomdiff	-4.58e+10	9.30e+10	-0.49	0.623	-2.28e+11	1.37e+11
persuasdiff	2.72e+12	1.85e+12	1.47	0.142	-9.10e+11	6.35e+12
_cons	2.03e+11	3.40e+09	59.61	0.000	1.96e+11	2.10e+11

Table A.3: Regression results, mean = f(Randomratio, Persuasratio), simulated data (m=2450).

Source	SS	df	MS		
Model	3.3135e+12	2	1.6567e+12	Number of obs =	2450
Residual	1.4827e+15	2447	6.0594e+11	F(2, 2447) =	2.73
				Prob > F =	0.0651
				R-squared =	0.0022
				Adj R-squared =	0.0014
Total	1.4860e+15	2449	6.0679e+11	Root MSE =	7.8e+05

mean	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
randomratio	4794.755	11938.55	0.40	0.688	-18615.95	28205.46
persuasratio	26921.57	11643	2.31	0.021	4090.413	49752.74
_cons	-38893.41	28545.08	-1.36	0.173	-94868.42	17081.6

Table A.4: Regression results, variance = f(Randomratio, Persuasratio), simulated data (m=2450).

Source	SS	df	MS		
Model	7.4929e+22	2	3.7465e+22	Number of obs =	2450
Residual	6.9329e+25	2447	2.8332e+22	F(2, 2447) =	1.32
				Prob > F =	0.2667
				R-squared =	0.0011
				Adj R-squared =	0.0003
Total	6.9404e+25	2449	2.8340e+22	Root MSE =	1.7e+11

variance	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
randomratio	-1.98e+09	2.58e+09	-0.77	0.444	-7.04e+09	3.08e+09
persuasratio	3.57e+09	2.52e+09	1.42	0.157	-1.37e+09	8.50e+09
_cons	2.01e+11	6.17e+09	32.51	0.000	1.89e+11	2.13e+11

Table A.5: Regression results, mean = f(p1,p2,p3,p4).

Source	SS	df	MS		
Model	2.384*10^12	4	5.961*10^11	P > F	0.415
Residual	1.483*10^15	2445	6.068*10^11	Adj. R^2	-0.000
Total	1.486*10^15	2449	6.067*10^11	RMSE	7.8*10^5
				N	2450

Mean	Coef.	Std. Err.	t	P>|t|	95% CI LB	95% CI UB
p1	-52871.02	605941.1	-0.09	0.93	-1241082	1135340
p2	-2035137	1.21*10^7	-0.17	0.867	-2.58*10^7	2.17*10^7
p3	717052.9	611050.5	1.17	0.241	-48177.3	1915283
p4	1.98*10^7	1.22*10^7	1.62	0.106	-4187150	4.38*10^7
constant	-78461.9	69319.92	-1.13	0.258	-214393.7	57469.94

Table A.6: Regression results, variance = f(p1,p2,p3,p4).

Source	SS	df	MS		
Model	9.122*10^22	4	2.280*10^22	P > F	0.5222
Residual	6.931*10^25	2445	2.834*10^22	Adj. R^2	-0.0003
Total	6.940*10^25	2449	2.834*10^22	RMSE	1.7*10^11
				N	2450

Variance	Coef.	Std. Err.	t	P>|t|	95% CI LB	95% CI UB
p1	3.37*10^10	1.31*10^11	0.26	0.797	-2.23*10^11	2.90*10^11
p2	-4.34*10^12	2.62*10^12	-1.66	0.098	-9.47*10^12	7.99*10^11
p3	-5.63*10^10	1.32*10^11	-0.43	0.670	-3.15*10^11	2.03*10^11
p4	1.06*10^12	2.64*10^12	0.40	0.687	-4.12*10^12	6.25*10^12
constant	2.13*10^11	1.50*10^10	14.22	0.000	1.84*10^11	2.42*10^11

Figure A.10: Frequency vs. sample mean for simulated Nikkei 225 data, uniformly distributed parameters $\{p_1, p_2, p_3, p_4\}$ in ranges specified in the text, m=2500. Note the local minimum near $\mu=0$.

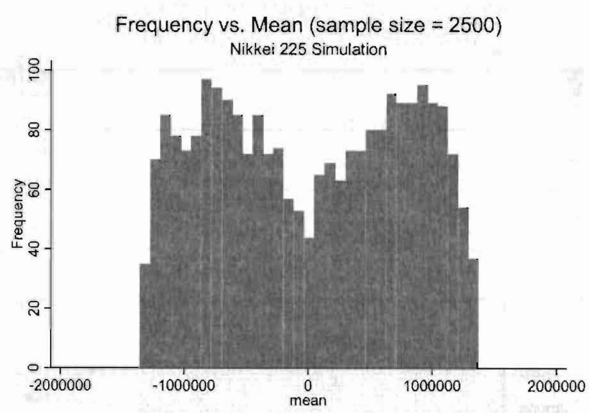


Figure A.11: Variance estimates vs. p_2, Hang Seng simulation.

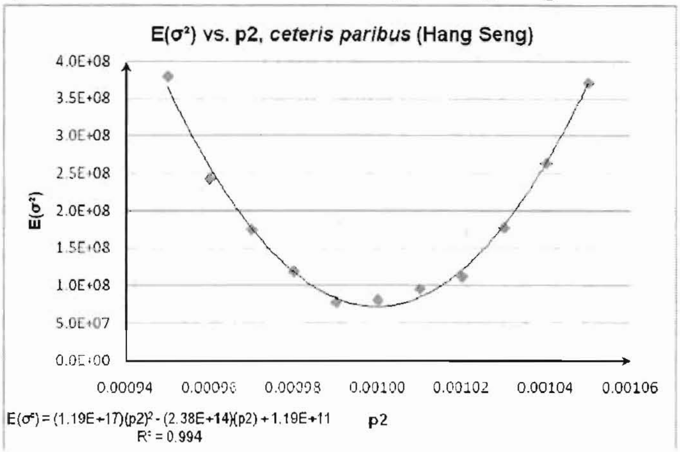


Figure A.12: Variance estimates vs. p_2, Nikkei 225 simulation.

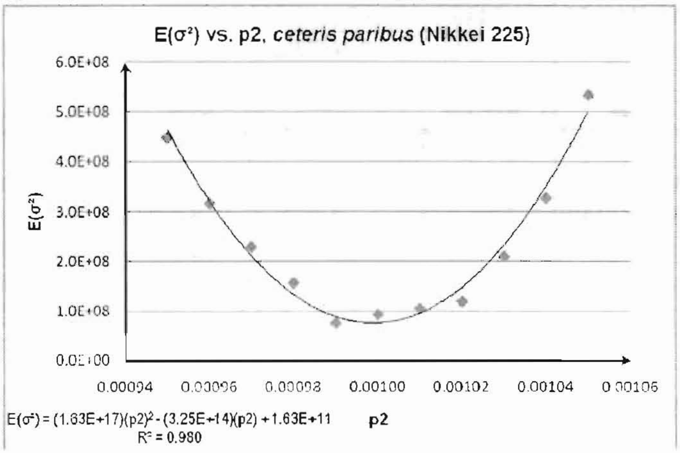


Figure A.13: Mean estimates vs. p_1, Hang Seng simulation.

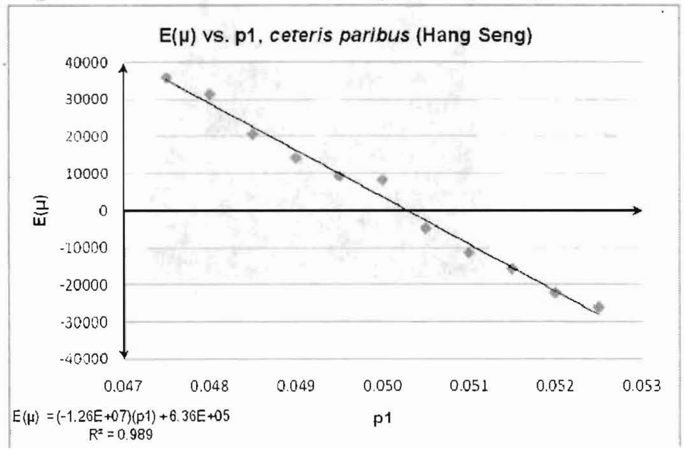


Figure A.14: Mean estimates vs. p_1, Nikkei 225 simulation.

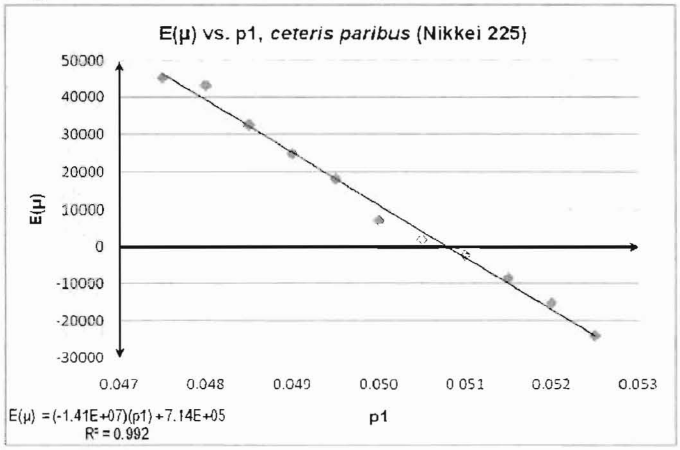


Figure A.15: Variance estimates vs. p_1, Hang Seng simulation.

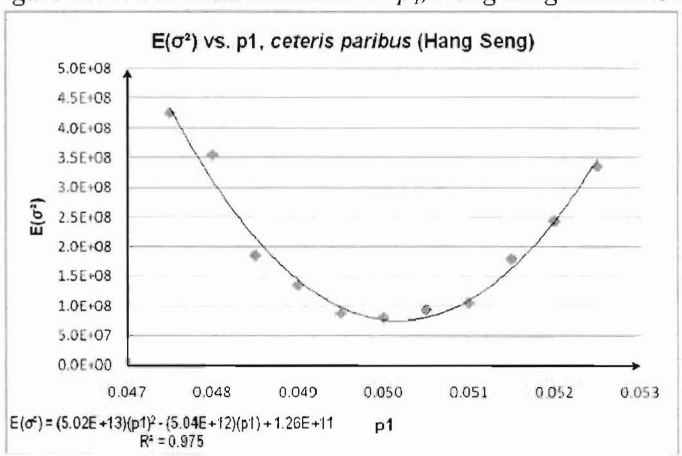


Figure A.16: Variance estimates vs. p_1, Nikkei 225 simulation.

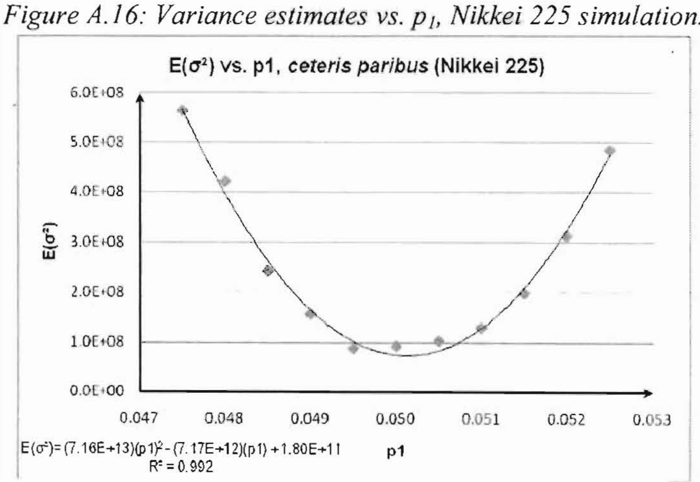


Figure A.17: Mean estimates vs. p_5, Hang Seng simulation.

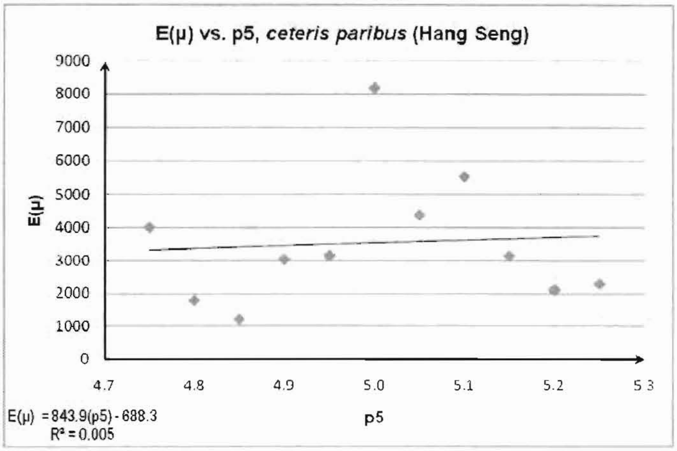


Figure A.18: Mean estimates vs. p_5, Nikkei 225 simulation.

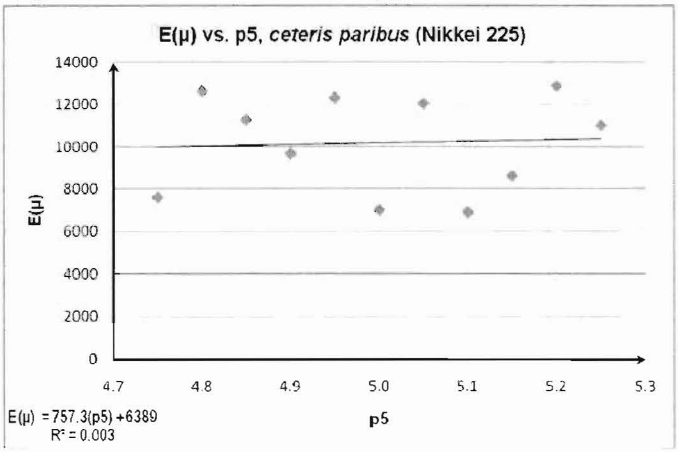


Figure A.19: Variance estimates vs. p_5, Hang Seng simulation.

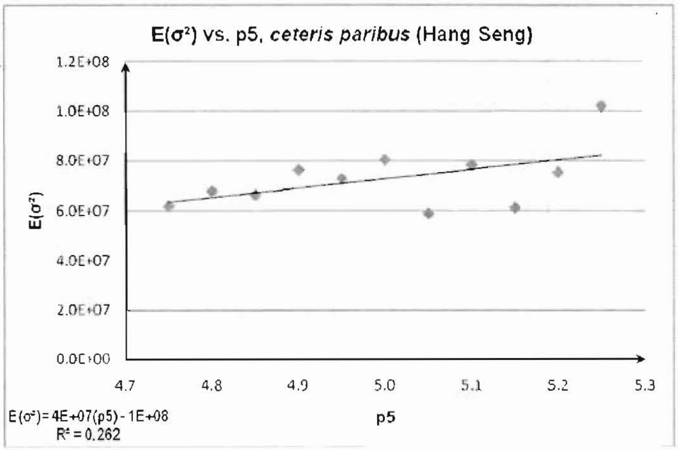


Figure A.20: Variances estimates vs. p_5, Nikkei 225 simulation.

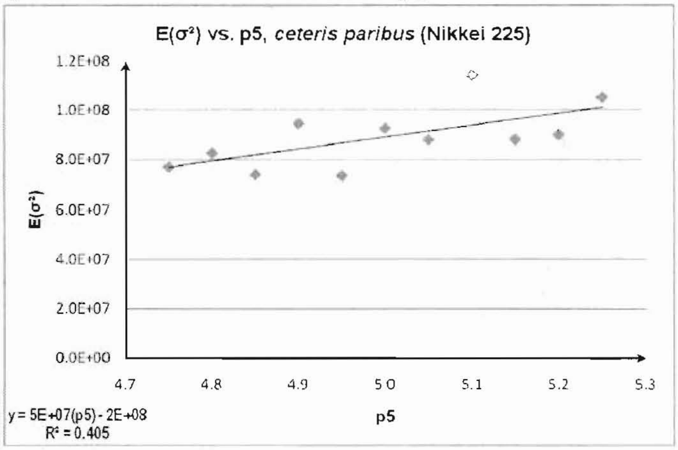


Figure A.21: Autocorrelations, 100 lags, daily return of HSI.

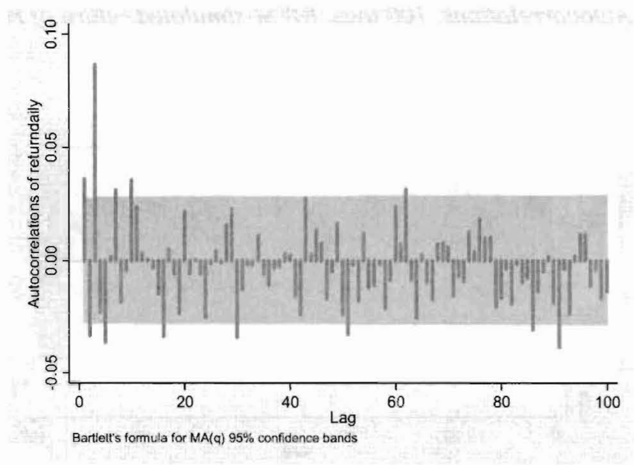


Figure A.22: Autocorrelations, 100 lags, daily return of N225.

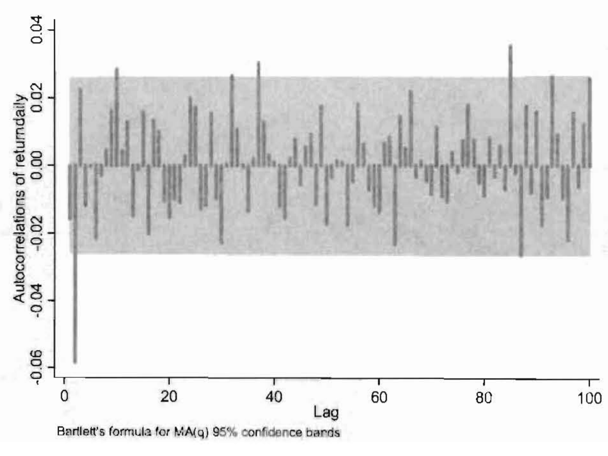


Figure A.23: Autocorrelations, 100 lags, ABM-simulated return of HSI (trial #10).

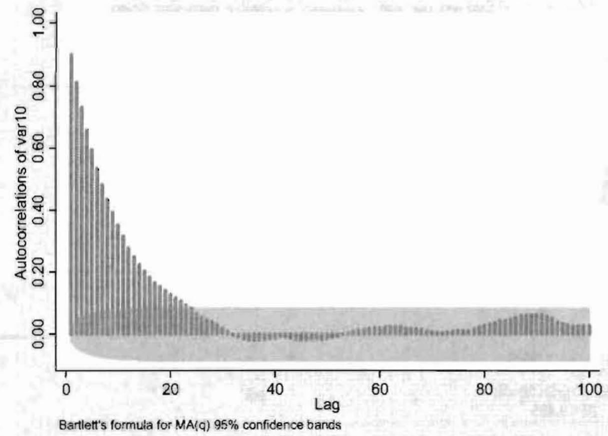
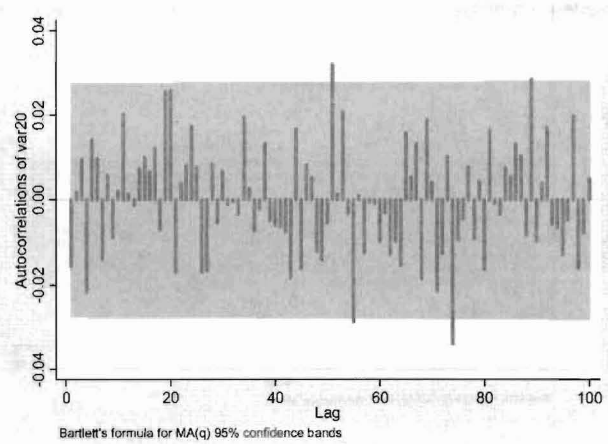


Figure A.24: Autocorrelations, 100 lags, RWM-simulated return of HSI (trial #20).



XII. Appendix II

Figure A2.1: Mathematica code, random-walk model.

```
Clear[α, β, n, k, s, t, p1, p2, p3, p4, p5, A, B, G, H, J,
 f, g, h, j]

s := 9954
k := 5797
δ := 2.7568

draw[k_Integer] :=
 NestList[# + δ + RandomArray[NormalDistribution[0, 1], 1] &, s, k]

A = Flatten[draw[k]];
```

Figure A2.2: Mathematica code, agent-based model.

```
q = 0; Label[start]; Clear[α, β, n, k, s, t, p1, p2, p3, p4, p5, A, B, G, H, J, f, g, h, j];
δ := 0.01; k := 5796; s := 9927; n := 100;
p1 := RandomReal[{0.001 - δ * 0.001, 0.001 + δ * 0.001}]; p2 := 0.05; p3 := 0.001;
p4 := 0.05; p5 := 5; j[z_] := s + z;
f[n_, s_, p1_, p2_, p3_, p4_, p5_, k_] :=
 NestList[
  (p5) *
    (2 *
      Clip[
       (First[RandomInteger[BinomialDistribution[Round[(((# / p5) + n) / 2)],
            1 - p1 * n + p1 * (((# / p5) + n) / 2) - p2], 1]] +
          First[RandomInteger[BinomialDistribution[n - Round[(((# / p5) + n) / 2)],
            (p3) (((# / p5) + n) / 2) + (p4)], 1]]), {1, n - 1}] - n) &, 0, k];
A = f[n, s, p1, p2, p3, p4, p5, k];
B = Thread[j[Table[Apply[Plus, Take[A, i]], {i, 1, k + 1}]]];
G = {k, p1, p2, p3, p4, p5, Round[N[Mean[B]]], Round[N[Variance[B]]],
  N[Skewness[B]], N[Kurtosis[B]]}; G >>> ABMdata; q += 1;
If[q < 149, Goto[start]]
```

ABMdata is a data file which would be exported later by the user when the simulation is complete (q = 150). *Mathematica* 6.0 was used in both cases. This code will not work properly in a *Mathematica* 5.2 environment as certain functions are not available. The binomial distribution is used because each agent makes an independent switching decision in every period; the agent either leaves or remains in its current group with fixed probabilities, thus the final size of a group is binomially-distributed. In other words, a group of traders changes size based on its agents engaging in repeated Bernoulli trials, where "success" is staying in the group and "failure" is switching to the other group.