STATISTICAL INFERENCE AND MODELLING OF MOMENTUM IN STOCK PRICES

ABSTRACT. The following results are obtained. (i) It is possible to obtain a time series of market data $\{y(t)\}$ in which the fluctuations in fundamental value have been compensated for. An objective test of the efficient market hypothesis (EMH), which would predict random correlations about a constant value, is thereby possible. (ii) A time series procedure can be used to determine the extent to which the differences in the data and the moving averages are significant. This provides a model of the form

$$y(t) - y(t-1) = 0.5\{y(t-1) - y(t-2)\} + \epsilon(t) + 0.8\epsilon(t-1),$$

where $\epsilon(t)$ is the error at time t, and the coefficients 0.5 and 0.8 are determined from the data. One concludes that today's price is not a random perturbation from yesterday's; rather, yesterday's *rate of change* is a significant predictor of today's rate of change. This confirms the concept of momentum that is crucial to market participants. (iii) The model provides out-of-sample predictions that can be tested statistically. (iv) The model and coefficients obtained in this way can be used to make predictions on laboratory experiments to establish an objective and quantitative link between the experiments and the market data. These methods circumvent the central difficulty in testing market data, namely, that changes in fundamentals obscure intrinsic trends and autocorrelations.

This procedure is implemented by considering the ratio of two similar funds (Germany and Future Germany) with the same manager and performing a set of statistical tests that have excluded fluctuations in fundamental factors. For the entire data of the first 1149 days begining with the introduction of the latter fund, a standard runs test indicates that the data is 29 standard deviations away from that which would be expected under a hypothesis of random fluctuations about the fundamental value. This and other tests provide strong evidence against the efficient market hypothesis and in favor of autocorrelations in the data. An ARIMA time series finds strong evidence (9.6 and 21.6 standard deviations in the two coefficients) that the data is described by a model that involves the first difference, indicating that momentum is the significant factor.

The first quarter's data is used to make out-of-sample predictions for the second quarter with results that are significant to 3 standard deviations. Finally, the ARIMA model and coefficients are used to make predictions on laboratory experiments of Porter and Smith in which the intrinsic value is clear. The model's forecasts are decidedly more accurate than that of the null hypothesis of random fluctuations about the fundamental value.

INTRODUCTION. An understanding of deterministic and random forces in markets is complicated by two factors: (i) frequent changes in underlying fundamentals that alter the value of the security, and (ii) noise or randomness that is unrelated to fundamental value. For several decades, the dominant theory has been the efficient market hypothesis (EMH) which exists in several forms and stipulates that changes in fundamental value plus augmented by statistical noise are the only factors in a market that would be confirmed by statistical testing. The statistical testing, however, is complicated by precisely the noise factor, compounded by the fact that quantification of events in terms of fundamental value is not generally precise or unique. For example, the impact of a CEO resignation to a company's stock, or an unexpected political event to a nation's bond market are difficult to quantify. Thus, an attempt to test EMH usually centers on showing that a very large data set is random, e.g., that a random walk model in which the price at time t is given by the price at time $t-1$ plus a noise term is as good a description as any other. Of course, within such a study, it is difficult to extract or compensate for world events that have a great deal of randomness.

Thus the question of distinguishing between random and deterministic behavior is rather sublte. A recent survey of results appears in Griliches and Intriligator [1994]. Various time series methods have been used, e.g., by Brock [1986] and others (see references therein) to measure the extent to which a given set of data is of random origin. This issue is more complex, however, due to the existence of simple nonlinear models which exhibit highly irregular patterns that are difficult to distinguish using standard time series methods from random processes. Consequently, it is important to develop new approaches that focus on a smaller but statistically significant subset of the market data in which the randomness involved in external factors (that affect fundamental value) has been compensated.

As noted above, the classical theories of financial markets that have largely dominated the academic studies of financial markets are generally based on the concept of random price fluctuations about a fundamental value [e.g., Fama (1970), Tirole (1982)]. A theoretical basis for this idea is that any deviation from a realistic value is reason for investors to purchase (due to undervaluation) or sell short (due to overvaluation) in the expectation that a profit can be made as other investors follow suit. The following three prerequisites appear to be part of any microeconomic model of this concept: (i) There are a group of investors with similar assessments of the value of the security. (ii) These investors have capital that exceeds those of the less informed. (Or, alternatively they manage the capital of less informed investors who have learned of their investing skill through a very accurate representation of managers' likely performance.) (iii) These investors not only want to maximize their returns but are also willing to rely on the self-interested maximization of others.

Testing the classical theories by examining the evolution of major US stocks during the past century, Shiller (1981) concluded that volatility over the past century appears to be five to thirteen times too high to be attributed to new information about future real dividends. A number of other studies have drawn similar conclusions on other subsets of market data (e.g. Loderer and Zimmermann [1985]).

An interesting application of these ideas is the issue of the persisent discount from net asset value that afflicts the average closed-end fund (about 12% for domestic equity funds in January 1995). Some recent articles have debated whether the origin of the closed end discount is related to investor sentiment [Anderson and Born (1992), Lee *et al* (1991), Chen *et al* (1993), Chopra *et al* (1993)]. The closed end discount has thus been a puzzle for both practical and academic observers, since investors with cash can attempt to maximize their returns by buying $100 of stock for $88. This immediately invokes the concept of reliance on the self-interested maximization of others, since the profit usually (though not always) cannot be attained in any other way. Experiments of Beard and Beil [1994] indicate the reluctance of most participants to rely on others even while engaging in self-maximizing behavior themselves. This finding may be the fundamental explanation for much of the observed deviations from classical theory of efficient markets.

Another application is the large body of knowledge grouped under the heading of 'technical analysis' that is based on the premise that the markets do exhibit more statistical structure than random fluctuations about fundamental value (see for example Edwards and Magee [1992]). The existence or nonexistence of nonrandom, autocorrelated behavior is crucial for the validity of this largely practical discipline that is generally ignored by academics, although one recent study by Blume, Easley and O'Hara [1994] demonstrates that the recent history of volume could be potentially useful to traders.

Understanding market efficiency or ineffiency is often difficult because 'noise makes it very difficult to test either practical or academic theories about the way that financial or markets work,' according to Black [1986]. Thus, a growing body of research work has addressed the problem of nonclassical market behavior through controlled laboratory experiments in which subjects are allowed to bid through computers on an asset whose value is unambiguous. These include Forsythe *et al* [1982], Plott and Agha [1982], Plott [1986], Porter and Smith [1989a,b, 1994], Smith [1982, 1986], Smith *et al* [1988], Williams [1979, 1987], and Williams and Smith [1984].

One such experiment (Porter and Smith [1989a,b]) allowed participants the choice of retaining cash or bidding on an asset that pays $0.24 at each of the 15 time periods and is subsequently worthless. The participants usually started the bidding well under the $3.60 fundamental value, then drove the price well above the fundamental value in the middle time periods in a speculative bubble until prices dropped in the ensuing bust. This experiment differs from many of its predecessors e.g. Smith, Suchanek and Williams [1988], Williams and Smith [1987] and Williams [1984] in that there was no uncertainty about the value of the financial instruments. A thorough series of experiments has demonstrated the robustness of these bubbles that are not eliminated by imposing trading costs or limits, using business people instead of students or allowing short selling, etc. (Porter and Smith [1994]). As discussed in these papers and in Caginalp and Balenovich [1994], the data obtained in the experiments would be extremely difficult to explain without the concept of price trend or derivatives influencing trading patterns. The concept that traders are partly influenced by the recent trend is incompatible with EMF since price histories can help predict prices in that case. However, the results of the experiments are not completely incompatible with EMF in principle in two ways: (i) The price does eventually reach the fundamental value though a large bubble forms before it does so. (ii) Players who have attained experience in these games (as a common group) tend to trade closer to the fundamental value, since the participants have a more accurate expectation of the trading preferences of the rest of the group. As noted by Porter and Smith [1989a,b], real markets always have some experienced investors as well as some newcomers and it is not clear *a priori* whether the fundamental value will ever be attained.

These experiments inspire two immediate questions about world markets. First, are the assets of experienced investors, who are ready to exploit inefficiency, adequate to keep trading close to fundamental value, or does the constant influx of new investors dominate markets? Second, is there a time scale in ordinary markets, analogous to the experiments, on which a return to equilibrium occurs?

We address these questions in this paper by focussing on financial instruments that serve to interpolate between the two extremes of a controlled laboratory experiment and a vast array of unsorted data. This situation is possible in the (uncommon) case where two stocks on the same exchange represent essentially the same fundamental value but trade independently of one another. By considering their *ratio*, one can examine statistical properties in a time series that has already extracted changes in fundamental value. Toward this end we study the ratio of two closed end funds, The Germany Fund (GER) and The Future Germany Fund (FGF), both of which invest in common stocks in Germany and have the same managerand nearly identical portfolios. An examination of the changes in net asset value shows that the yearly appreciation or depreciation of the two funds is within 1% of one another, which is not surprising since the portfolios are very similar (Morningstar [1994]).

The EMH would predict a constant ratio for the stock prices FGF/GER plus randomness or noise. Statistically significant nonrandomness would require rejecting this hypothesis for this data. Moreover, autocorrelations and Box-Jenkins ARIMA models (to be defined in Section 2) can be used to compare systematically the random walk hypothesis with the optimal choice of the ARIMA model and coefficients. In particular, we perform 'out of sample' tests by evaluating these coefficients using a portion of the data and then testing the predictive data on future price ratios.

We examine a time period from the inception of the latter fund to be introduced (FGF on March 6, 1990) until the beginning of this study, a period of 1149 days. We begin by performng two nonparametric tests on this data: the runs test (which examines the number of crossings of the mean relative to the expected value) and the turning point test (which examines the number of changes in direction). These are defined precisely in Section 1. Applying these tests to the entire data set, we fall 29 stantard deviations short of the expecterd number of runs, a strong refutation of randomness in the series. Similarly we are 6.8 standard deviations below the expected number of turning points expected in the random series. The Box-Jenkins method implies a strong correlation between the differences of the series, rather than the individual members of the series, and an ARIMA (1,1,1) model (defined in Section 1) is found to have coefficients that differ from the (random) null hypthesis by 21.6 and 9.6 standard deviations. These and other tests argue overwhelmingly for a nonrandom, autocorrelated component and against a trendless random walk that would be required by the EMH.

These tests also indicate the need for a trend based (or derivative) component in any reasonable fit of the data, since the most valuable predictor is the current difference as a multiple of the most recent difference. This provides deductive support for the approach used by Caginalp and Balenovich [1994] in modeling markets using differential equations in which trend based investing plays an important role. Finally, the predictive value of the ARIMA model involving this trend dependency is explored by evaluating the coefficients based on the first quarter (64 trading days) and using these coefficients in the same model to predict the subsequent two weeks of market activity. The predictions are quite accurate, providing another confirmation that the model explains the phenomenon quite well. The ARIMA method is also used to make predictions for the entire second quarter by updating the coefficients at each time in order to predict the next time. The predictions compare favorably with the null hypothesis of the random walk hypothesis.

Our results indicate that the effect of the derivative, or trend based investing, is most pronounced within the first quarter and gradually diminishes so that the last 64 trading days is much closer to randomness. This appears to be compatible with the experimental observations of Porter and Smith [1898a,b] and other experimenters who concluded that experience as a group yields a convergence to fundamental value. The data is also suggestive of damped oscillations that were also implied by the Caginalp and Ermentrout [1990, 1991] numerical studies of the differential equations.

The data analysis in this paper helps to establish a methodology for evaluating the time scale on which convergence to equilibrium can be expected. The rational hypothesis theories would not only expect a monotonic convergence to equilibrium (as a result of first order differential equations) but would also imply a time scale that is at best instantaneous (i.e. since the beginning of FGF trading is known in advance, the adjustment should occur before the first trade), and at worst one or two days if some inefficiency is allowed. Our finding that this time scale is of the order of a calendar quarter is surprising in this context.

The instances in which changes in fundamental value can be extracted in this way can be regarded as a *de facto* world market experiment that addresses the primary concern of the critics of the laboratory experiments, in that everyone with access to the New York Stock Exchange had the opportunity to participate. And the tests of the autocorrelations in the first quarter are remarkably close to those of the laboratory experiments, as discussed in the Conclusion. Our study thereby provides a quantitative response to the most serious criticism of the laboratory experiments: the exclusion of outside arbitrage capital from the experiments. Hence an analysis of this type can be considered as an interpolation between the experiments and world markets which can be applied as other similar situations arise.

1. STATISTICAL TESTS FOR RANDOMNESS. We define the ratio of the two securities as

$$y(t) = \frac{\text{Price of Future Germany Fund at time t}}{\text{Price of Germany Fund at time t}} \tag{1.1}$$

In general, the price need not be equal to the net asset value (NAV) and may trade at a discount (if the price is less than NAV) or premium. The discount of each fund is defined by

$$D(t) = \frac{\text{NAV of security - Price of security}}{\text{NAV of security}} \times 100\%. \tag{1.2}$$

The absolute price of GER and FGF are, of course, arbitrary denominations, so that the absolute ratio is not significant and only the temporal evolution of the ratio is relevant. One may expect that the discount of both FGF and GER would be very close, but our analysis allows for the preference of one over the other. In other words, if for any reason the fund GER merits a lower discount or a higher premium over that of FGF, then one would still expect the ratio to be invariant over time except for noise fluctuations in a perfectly rational market.

To be more precise, one can write y(t) as

$$y(t) = \frac{NAV_{FGF}(t) \cdot [D_{FGF}(t) + 1]}{NAV_{GER}(t) \cdot [D_{GER}(t) + 1]}. \tag{1.3}$$

If the portfolios of any two funds are identical then the ratio of the two NAV's must remain constant. Also, the fraction of that NAV that American investors are willing to pay, i.e., $1 + D(t)$ may well vary with time as American investors' enthusiasm for the German market changes. However, there is no reason, based on fundamentals, for the ratio $(1 + D_{FGF}(t))/(1 + D_{GER}(t))$ to vary in time.

For the effecient market hypothesis to remain valid for these two stocks, one would need to show that the time series for all y(t), with t between 1 and 1149, is best described as a constant (reflecting initial denominations) plus a noise term. A very useful set of statistical tools are nonparametric tests which do not make any assumptions on the distributions of the underlying data. The first of these, called the 'sign test' (Krishnaiah and Singh [1984]), is a test for trend or autocorrelation and is based on the premise that any observation from a time series with independent error terms is equally likely to be above or below the median of the series. Each block of observations that is entirely above or below the median is called a run. For an odd number, n, of observations, the expected number of runs, μ_R, and the standard deviation of the number of runs, σ_R, are given by

$$\mu_R = \frac{n+1}{2} \qquad\qquad \sigma_R = \frac{1}{2}\sqrt{\frac{(n-1)(n-3)}{n-2}}. \tag{1.4}$$

Hence either a straight line or sine function would be at the opposite extreme from a series of random terms. We apply this test to the series $\{y(t)\}$ for (i) the first quarter, or 64 data points, and (ii) the entire data of 1149 days. For the first 64, we find 14 runs instead of the expected 33. Since the standard deviation is $\sigma = 3.94$, the result is 4.77 standard deviations away from the mean. In case (ii) we obtain 71 runs which is 29 standard deviations from the expected 561.8 runs! Note that 4.77 is a large number of standard deviations in the first case as well since the formula implies a maximum of about 10 standard deviations if there were the minimum of 2 runs.

A more subtle test for autocorrelation that is more sensitive to data away from the median is the turning point test. This test uses the number of changes in direction to test for randomness. The number of changes in direction, v, has a mean and standard deviation given by

$$\mu_v = \frac{2}{3}(n-2) \qquad\qquad \sigma_v = \sqrt{\frac{16n-29}{90}}. \tag{1.5}$$

For the first 64 days we find 30 turning points versus the expected 41.33 with a standard deviation of 3.32 so that our observation is 3.41 standard deviations below the mean. For the entire data we have 667 turning points versus the expected 764.6 with a standard deviation of 14.3 so that we are 6.8 standard deviations below the mean.

This analysis then argues strongly against the efficient market hypothesis with noise (the null hypothesis) and suggests that a more complex approach is needed.

An examination of the data for the first 64 days (Figure 1) reveals that the oscillations about a constant value are damped in time. In applying these tests to the last 64 days, for example, the runs test and the turning point test both are markedly different from the behavior near the first 64. The last 64 days exhibits 28 runs which is rather close to the expected 33 runs. Similarly, the turning point test counts 40 turning points, compared with the expected 41.3, showing good agreement with the hypothesis of random behavior for the quotient for the last part of the data. Specifically, 40 and 41.3 are separated by less than half of one standard deviation.

2. STATISTICAL MODELING OF DATA. Once the null hypothesis (described above) has been rejected by the nonparametric tests of the data, the use of Box-Jenkins ARIMA methods for understanding the essential nature of the time series can lead to a simple statistical model for obtaining optimal predictions (Farnum and Stanton [1989]).

The basic ARIMA models involve autoregressive (AR) components that link the present observation y(t) with those up to p times earlier, $\{y(t-1), \ldots, y(t-p)\}$ and the moving averages (MA) of the error terms experienced in the previous q members of the time series, $\{\epsilon(t-1), \ldots, \epsilon(t-q)\}$. The observations, y(t), can be differenced (denoted I) if the original series is nonstationary, so that the methods are applied to the sequence $w(t) := \Delta^d y(t)$ which is the sequence $\{y(t)\}$ differenced d times.

The Box-Jenkins madeling procedure involves the following steps:

(i) The series is differenced, if necessary, to make it stationary. This determines the parameter d. For example, either a runs or turning point test may be used to determine the level d at which the differenced data is within one standard deviation of the expected number of runs or turning points.

(ii) The autocorrelation function (acf) and the partial autocorrelation function (pacf) are computed to determine the autoregressive parameter, p, and the moving average parameter, q. This can be established by a formal 95% confidence rule.

(iii) the coefficients of the model, defined below, are then determined by a least squares estimate.

(iv) The residuals from the estimated model should be random; otherwise further analysis is needed.

The general model can then be written as:

$$\begin{aligned} w(t) &= \phi_1 w(t-1) + \phi_2 w(t-2) + \cdots + \phi_p w(t-p) \\ &\quad + \epsilon(t) + \theta_0 - \theta_1 \epsilon(t-1) - \cdots - \theta_q \epsilon(t-q) \end{aligned} \tag{2.1}$$

in terms of the coefficients or 'process parameters' ϕ_i and θ_i. In particular, ARIMA (0,1,0), i.e. $p = 0$, $d = 1$, $q = 0$, is just ordinary random walk, while ARIMA (0,1,1) is simply an exponential smoothing scheme.

Applying part (i) of this procedure to our entire data, we find that $d = 1$ is necessary and sufficient. Indeed, whereas the data $\{y(t)\}$ displays a slow upward drift, the first difference $\{y(t) - y(t-1)\}$ appears to be stationary. Evidence is provided by the nonparametric tests for centrality, such as the sign test and the Wilcoxon test. The former shows 512 negative signs, 91 zeros, and 545 positive signs; this yields a p-value of 32.5% with regard to the hypothesis that the median of the series is zero (versus that it is not zero). Wicoxon's test, on the other hand, yields the much higher p-value of 56.9% for the same hypothesis. The two high p-values indicate that the first difference is consistent with being centered at zero (as should be the case if it is stationary). Though we prefer nonparametric tests, a regression of $\{y(t) - y(t-1)\}$ on time produces a slope that is not significantly different fron zero. This provides good evidence that the series $\{y(t) - y(t-1))$ is stationary.

Examination of the acf and pacf's (Figure 2) results in $p = 1$ and $q = 1$, as the correlations drop dramatically in the next order. This points to the ARIMA(1,1,1) model, rather than the (0,1,0) associated with the random walk. The ARIMA(1,1,1) model selected by the data using this procedure is therefore

$$y(t) - y(t-1) = 0.5\{y(t-1) - y(t-2)\} + \epsilon(t) + 0.8\epsilon(t-1). \tag{2.2}$$

The coefficients 0.5 and 0.8 above are respectively 9.6 and 21.6 standard deviations away from the null hypothesis values of 0.

Information theoretic criteria provide another tool for model selection. Such criteria attempt to strike a balance between the size of residuals and the number of parameters present in the model. Their primary use is to guard against overfitting a model. We provide a summary of three information criteria: the Bayes, Akaike, and the final prediction error (FPR), as they apply to several competing models. The class of competing models are linear recurrences involving up to second order differences. Considering competing nonlinear recurrences would be of limited value, since the main candidate, ARIMA(1,1,1), which is linear, yields (as we shall see) good out-of-sample estimates, has excellent interpretability in terms of the "market momentum", yet it only uses two parameters. So the use of nonlinear models with the aim of reducing the number of parameters is not appealing. Table 1 summarizes the results. The ARIMA models are listed in the first column of the table. [The first model listed, ARIMA(0,0,0), corresponds to the EMH; it takes value 1 for all values of the time variable t. The second and third columns display the average of the sum of squares of residuals, denoted by $\hat{\sigma}_k^2$, for for the whole data ($n = 1149$). The index k denotes the number of parameter in the model and is found in column three. Columns four, five, and six show the values of the Bayes' information criterion $ln(\hat{\sigma}_k^2) + kn^{-1}ln(n)$, the Akaike criterion $ln(\hat{\sigma}_k^2) + 2kln(ln(n))$, and the FPE $\frac{n+k+1}{n-k-1}\hat{\sigma}_k^2$, for the whole data set of $n = 1149$ *points*. A discussion of these three criteria can be found in Shumway ([1988], p. 154). A model is "best" with respect to a criterion if the criterion takes a minimal value for that model.

For the whole data set ARIMA(1,1,1) minimizes the Bayes criterion with a value of -7.53015, and comes in a close second with the other two critera. The ARIMA that is slightly better has 4 parameters and involves second order differences. We believe its complexity does not warrant the slight gain in the performance of the other two criteria. This presents additional strong evidence that ARIMA(1,1,1) is the model of choice, especially in view of the fact that it clearly outperforms the EMH, which has a Bayes information criterion value of -3.660. Since these criteria are on a logarithmic scale the differences between ARIMA(1,1,1) and other models, particularly the EMH is very considerable indeed.

Consequently, we find that the ARIMA(1,1,1) model emerges quite naturally from our data as it is confirmed by the stationarity and ACF, PACF tests as well as the information criteria. The model expressed in equation (2.2) is a relation between today's rate of change, $y(t) - y(t-1)$, compared with yesterday's, $y(t-1) - y(t-2)$. Thus the ARIMA procedure leads to the conclusion that the best predictor of prices is not a random perturbation from yesterday's price. Rather, the rate of change of yesterday is a significant predictor of today's rate of change. This is statistical confirmation of the concept of momentum that is so widely used by practitioners in a variety of forms.

3. OUT-OF-SAMPLE FORECASTS. An important test for a statistical model is its capacity to make out-of-sample predictions that preclude the possibility of "overfitting the data". The ARIMA(1,1,1) model renders a prediction once the coefficients have been evaluated, a process that generally requires at least 50 or 60 data points (Farnum and Stanton [1989]). We perform two types of forecasts. The first uses the coefficients that have been evaluated using the first quarter to predict the next ten days. The second involves using the first N days to evaluate the coefficients in order to predict the $(N+1)th$ day.

To perform the first set of forecasts, we use the first quarter's data of 64 observations to evaluate the ARIMA coefficients as in the previeous section with the analogous result

$$y(t) - y(t-1) = 0.8\{y(t-1) - y(t-2)\} + \epsilon(t) + 0.9\epsilon(t-1) \tag{3.1}$$

relating today's rate of change in the quotient of the two stocks with that of yesterday.

Thus the ARIMA model based on the first quarter's data can be used to predict the values of the quotient under study for an additional ten trading days. Without updating its coefficients, the ARIMA model yields a set of forecasts that can be compared with the actual outcomes. Specifically, one has

(Day, Forecast, Actual) $\longrightarrow$ {(65, 0.9889, 1.0169), (66, 0.9932, 1.0256), (67, 0.9968, 1.0344), (68, 0.9997, 1.0446), (69, 1.0020, 1.0263), (70, 1.0038, 1.0026), (71, 1.0054, 1.0454), (72, 1.0066, 0.9827), (73, 1.0076, 0.9739), (74, 1.0084, 0.9824)}.

The actual values are well within the 95% confidence regions centered at the predicted values in all cases. There is also a "catch-up" effect that is evident during the first few days of prediction. During days 65 through 69 the actual values are always higher that the forecast values, and the latter show a steady increasing trend attempting to catch-up with the actual response. The last five days of forecasting show that the forecasts actually exceed the true values, since the latter peak around the middle of the predictive period. Even ten days is a rather long forecasting period without updating the ARIMA coefficients, and it is unwise to go further.

The predictions we obtain appear quite good and indicate that the model used provides a suitable explanation of the data and of the phenomenon under study. Furthermore, a study of the residuals (i.e. the data minus the statistical model) indicates that the model fits very well indeed. A plot of the residuals versus the fitted values provides no indication of any additional trends in the data; see Figure 3. Formal nonparametric tests yield supporting evidence that what is left unexplained after fitting the ARIMA model is simply random error. The runs test, for example, observes 32 runs out of 32.7 expected under the hypothesis of randomness, a nearly perfect fit. The turning point test notes 44 turning points out of 41.3 expected. The standard deviation is in this case 3.32, and we therefore exceed the expected number of turning points by less than one standard deviation; this is rather consistent with the hypothesis of random residuals, as was stated earlier.

To perform the ARIMA (1,1,1) forecasts with updated coefficients, we use the first N days, beginning with the 64th to predict the following day, until the end of the second quarter, or 128th day. The forecasts and actual values are displayed in Figure 4. The forecasts are quite close to the actual values, as all of the latter values are well within the corresponing 95% confidence intervals. The 64 updated forecasts were obtained using the Minitab software.

A comparison that is most relevant to understanding the fundamental forces of market dynamics is that between the ARIMA (1,1,1) and the usual random walk that best expresses the predictions of the efficient market hypothesis. The random walk (with no drift term) predicts that the $(N+1)$ data point is equal to the Nth day plus random error. The ARIMA model gives forecasts that are closer to the actual values than the random walk in 44 of the 64 days of the second quarter, which is exactly 3 standard deviations away from the expected 32, assuming a binomial distribution. A more extensive test of the forecasts can be made by the Wilcoxon matched pairs signed ranks test, which is a nonparametric test that can distinguish between two possible explanations without the distortion that would be involved in assumptions on the distribution of the underlying data. One defines $c_1(t)$ as the magnitude of the difference between the forecast and the actual data. Similarly, $c_2(t)$ is the magnitude of the difference between the random walk prediction and the actual data. These values are ranked with respect to the magnitude of $|c_1(t)-c_2(t)|$ and signed with respect to $c_1(t)-c_2(t)$. Then the sum of the positive ranks is $T_+=580$, which is just over three standard deviations from the expected value. This provides strong statistical evidence that the statistically determined model which effectively chooses momentum as its key component is a better predictor than the random walk hypothesis.

Note that the predictions of the EMH could also be interpreted to predict a single constant value, i.e. the value of the 64th day, which would have even less predictive value than the random walk.

The previous two sections justify the ARIMA(1,1,1) model on the following counts:

1. It is the model that minimizes the Bayes information criterion.

2. It uses the market momentum as its key predictive feature, having thereby a wide intuitive appeal. We showed that is yields better out-of-sample prediction than the random walk model, which also enjoys good interpretability but has a poor overall performance for the entire data set according to the three information criteria (see Table 1).

3. ARIMA(1,1,1) offers a degree of "universality", in the sense that it fits well the first quarter of the data, and it fits best the entire data in accordance to the Bayes criterion. It is the simplest model that hints to some degree to a certain general "principle" independent of the data sets. It is this feature that motivates us to use it in the next section for the purpose of explaining the Porter and Smith experiments.

4. PREDICTIONS OF LABORATORY EXPERIMENTS. As discussed in the introduction, the statistical methods have the potential to establish a quantitative link between the laboratory experiments and the world markets. Toward this end we implement the ARIMA model established in (3.1), using the first quarter of the FGF/GER data, in order to forecast the experiments done by Porter and Smith [1989b]. This means that we will not only use a model with the first difference, but with precisely the same coefficients that had been optimized with the first quarter of $y(t) = FGF(t)/GER(t)$ data. In the set of experiments we consider, the participants make trades by the usual nominal bidding process on a financial instrument that pays 24 cents during each of 15 periods. Hence, the fundamental values of the instrument are given by

$$P_a(t) = 3.60 - 0.24t \qquad \text{for } t = 1,2, \dots\ 15. \tag{4.1}$$

We let $P(t)$ denote the experimental values which typically start well under the initial fundamental value of 3.60. By defining

$$x(t) = P(t)/P_a(t) \tag{4.2}$$

we can cast the issue is the same framework of the problem of time series for $y(t) = FGF/GER$. The two time series $y(t)$ and $x(t)$ both possess the key property that the temporal changes in their fundamental value have been eliminated. Consequently, the efficient market hypothesis predicts a constant value (in time) for both [in particular, EMH predicts $x_{EMH}(t) = 1$ for the time series $x(t)$]. If one of these time series can be used to predict the other, then it would provide evidence of fundamental deterministic forces that underlie both experimental and world markets. This is what we establish below by using the ARIMA model and coefficients determined by $y(t)$ in order to make successful predictions for $x(t)$. The implementation of the ARIMA forecast for $x(t)$ can begin with the third trading period, since the first two are needed for the initial conditions. The remaining 13 prices can be forecast with the ARIMA model (3.1) for each of the experiments. In any experiment the forecast at time t, $x_f(t)$, is given by (3.1) as

$$x_f(t) = x(t-1) + 0.8\{x(t-1) - x(t-2)\}. \tag{4.3}$$

The data points have been forecast in this way can then be tested for statistical validity. We apply this analysis to each of the Porter and Smith [1989b] experiments for which the initial trade is at least one-half of the fundamental value (see Table 2). We note that two of the experiments begin with values that are 27% and 42% of the fundamental value, and exhibit an extremely nonlinear price evolution that cannot be described by a first difference ARIMA model. Using the EMH prediction of $x_{EMH}(t) = 1$ as the null hypothesis we test the forecasts $x_f(t)$ against the experimental values, x_{EXP}(t), by implementing statistical tests discussed in Section 3. In particular, the deviations of the model forecasts from the experimental data, namely the vector M with entries $|x_f(t) - x(t)|$, are compared with the deviations of the EMH predictions from the data, i.e., the vector E with entries $|x_{EMH}(t) - x(t)|$. In all there are $13(5) - 1 = 64$ data points in the experiments as no trading occured at one time period.

The Wilcoxon paired difference test defined in Section 3 then results in T_+ with $n = 64$ which has a p-value of 0.007. This means that the probability of observing a median for E that is less than or equal to M is less than 1%. In summary, this test confirms with a very high confidence level that the forecasts are superior to those of the null hypothesis. We find this result remarkable in view of the fact it is not only the particular model but the coefficients as well that have been determined entirely from New York Stock Exchange data.

Furthermore, the Box-Jenkins procedure that underlies the ARIMA methods is deductive and objective, so that the comparison between the NYSE data and the Porter and Smith experiments is done in an unbiased manner.

5. CONCLUSION. A ratio of two closed end mutual funds with very similar portfolios and an identical manager, renders a financial entity whose fundamental value is constant in time. This procedure removes the effects of noise in the valuations of the underlying securities in the funds' portfolios and allows an examination of basic forces in the marketplace. The introduction of the latter fund establishes a time frame so that the time scales involving relaxation from disequilibria can be established. The strong autocorrelations we observe can be compared with the laboratory data of Porter and Smith [1989b], in which observations over 15 time periods generally led to 7 standard deviations from the expected value in the runs test, and 3.7 standard deviations from expected value for the turning point test (Caginalp and Balenovich [1991, 1994]). Thus, the tests on the ratio for the early portion (e.g. up to day 64) indicate an autocorrelated behavior that is similar to the laboratory observations. The ARIMA tests on the entire data also lead to the conclusion that the derivative of price is very significant. We note that Caginalp and Balenovich [1994] used the Wilcoxon signed pairs test and the Porter and Smith [1989b] data to compare the predictions of their differential equations model involving price derivative and valuation to the EMH and to the model in which derivative based strategy was eliminated. Their statistical finding that the model with the derivative behavior was overwhelmingly more likely than the efficient market hypothesis with noise is corroborated by our results.

An important aspect of our statistical modelling involves the establishment of quantiative comparison tests between the laboratory experiments and world financial markets. By utilizing market data in which the fundamental value can be factored out, we are able to consider a situation that is comparable to the controlled laboratory experiments in which the fundamental value is clear. The fact that the ARIMA model established from NYSE data is an excellent predictor of the experiments provides an indication that the underlying forces of price dynamics in the experiments must be very close to those of the world markets, and that momentum is an important part of both.

The methodology we use in this paper may provide an explanation for the paradox or controversy that has existed for some time between the experimenters who find bubbles under very robust conditions and time series analysts who generally seem to find randomness in testing a large amount of market data. By focusing on large amounts of unsorted data - for example, the New York Stock Exchange now contains a significant number of stocks which are in fact closed end *bond* funds - there are an uncontrolled number of random events that effect the fundamental value of the stocks. Consequently, the observation that there is little or no autocorrelation in such a sample does not fully address the question of whether there are significant autocorrelations in the marketplace. Realistically, while these studies provide a great deal of insight into market dynamics, they do not conflict with the observations of the bubbles experiments, and other studies that suggest deviations from the rational market behavior. Similar comments apply to the studies that attempt to predict the behavior of an individual stock, such as the White [1993] analysis of IBM stock, which showed that the previous three days' market closes could not be used to predict the fourth day's close. The influences on such a stock are so diverse (e.g., the company's expected earnings, the outlook for the computer industry, interest rate changes, overnight changes in the foreign markets, analysts' changes in ratings of IBM stock, changes in company's executives, etc., and the ever changing emphasis that investors place upon these factors) that the randomness associated with these fundamental changes may obliterate a significant amount of autocorrelation that one finds in the laboratory experiments. Furthermore, a convincing and objective methodology seems to require *not* utilizing an understanding of the major factors that affect a stock's price. Any study that does isolate the major factors and models the price accordingly is generally dismissed as suspect and contrived, and consequently insignificant on a statistical basis.

Our methodology overcomes these obstacles by examining a ratio that has extracted the randomness of fundamental changes in a simple, objective and irrefutable manner. The validity of our out-of-sample predictions is further support for our approach. As markets expand, and similar stocks and funds are traded around the globe, this methodology can be utilized to examine an increasing set of data around the world. Currently, these methods can be used to test a number of pairs of similar closed end funds run by distinct managers. The very rapid increase in American Depository Receipts (ADR's) also can lead to a vast data set for a comprehensive test. In particular, one would compare the ADR with the stock in the home country, adjusted for the currency changes.

Finally, there is a basic issue that is at the core of any dynamical theory, namely establishing a time scale for the return to equilibrium after a perturbation. Within our study, the introduction of the second fund marks a starting point for the time series, so that one can study the time interval required for the ratio of the two stocks to exhibit random behavior. Of course, since the introduction of second fund in known in advance, this time scale would vanish under complete efficiency. Even with some inefficiency, one might expect that equilibrium would be established within a few days. Our finding that randomness is not achieved for at least several quarters is quite surprising in this context.

ACKNOWLEDGEMENT: The authors thank an anonymous referee for suggesting the use of information theoretic criteria pertaining to model selection, and for additional changes that led to an improved exposition.

REFERENCES:

Anderson, S. and J. Born 1992: *Closed-End Investment Companies: Issues and Answers* Kluwer, Boston.

Beard, T.R. and R.A. Beil 1994: *Do people rely on the self-interested maximization of others? An experimental test*, Management Sciences **40**, 252-262.

Blume, L., D. Easley and M. O'Hara 1994: *Market statistics and technical analysis: the role of volume,* J. Finance 69, 153-181.

Brock, W.A. 1986: *Distinguishing random and deterministic systems*, J. Economic Theory **40**, 168-195.

Caginalp, G. and D. Balenovich 1991: *Mathematical models for the psychology of fluctuations in financial markets*, paper delivered at the Economic Science Association Meeting in Tucson, Arizona.

Caginalp, G. and D. Balenovich 1994: *Market oscillations induced by the competition between value-based and trend-based investment strategies* Applied Mathematical Finance.

Caginalp, G. and G.B. Ermentrout 1990: *A kinetic thermodynamic approach to the psychology of fluctuations in financial markets*, Applied Mathematics Letters **4**, 17-19.

Caginalp, G. and G.B. Ermentrout 1991: *Numerical studies of differential equations related to theoretical financial markets*, Applied Mathematics Letters **4**, 35-38.

Chen, N., R. Kan and M. Miller 1993: *Are the discounts on closed-end funds a sentiment index?* J. Finance 48, 795-800.

Chopra, N., C. Lee, A. Shleifer and R. Thaler 1993: *Yes, discounts on closed-end funds are a sentiment index,* J. Finance 48, 801-808.

DeLong, J., A. Shleifer, L. Summers and R. Waldmam 1990: *Positive feedback investment strategies and destabilizing specualtion,* J. Finance **45**, 379-395.

Edwards, R. and J. Magee 1992: *Technical Analysis of Stock Trends, Sixth Ed.*, New York Inst. of Finance, New York.

Fama, E.F. 1970: *Efficient capital markets: a review of theory and empirical work*, J. Finance **25**, 383-417.

Farnum, N.R., and L.W. Stanton 1989: *Quantitative Forecasting Methods,* PWS-Kent, Boston.

Forsythe, R., T. Palfrey and C. Plott 1982: *Asset Valuation in an experimental market*, Econometrica **50**, 537-567.

Griliches, Z. and M. Intriligator 1994: *Handbook of Econometrics IV* No. Holland, Amsterdam.

Krishnaiah, P.K. and P.K. Sen, eds. 1984: *Handbook of Statistics 4*, North Holland, New York.

Lee, C., A. Shleifer, and R. Thaler 1991: *Investor sentiment and the closed-end fund puzzle,* J. Finance 48, 75-109.

Loderer, C., and H. Zimmermann 1985: *Rights issues in Switzerland some findings to consider in the debate over financing decisions,* Unpublished manuscript.

Miller, R.M., C.R. Plott and V.L. Smith 1977: *Intertemporal competitive equilibrium: An empirical study of speculation*, Quarterly J. Economics **91**, 599-624.

Morningstar 1994: *Closed-End Funds*, Chicago.

Plott, C.R. and G. Agha 1982: *Intertemporal Speculation with a random demand in an experimental market,* Aspiration levels in bargaining and economic decision making, R. Toertz, ed., Springer-Verlag, Berlin.

Plott, C.R., 1986: *Rational choice in experimental markets*, J. Business **59**, 301-328.

Porter, D. and V.L. Smith 1989a: *Futures markets, dividend certainty and common expectations in asset markets*, working paper, University of Arizona.

Porter, D. and V.L. Smith 1989b: *The scope of bubbles and crashes in experimental asset markets*, working paper, University of Arizona.

Porter, D. and V.L. Smith 1994: *Stock market bubbles in the Laboratory*, Applied Mathematical Finance.

Rosenthal, R., 1981: *Games of perfect information, predatory pricing and the chain store paradox*, J. Economic Theory **24**, 92-100.

Shiller, Robert 1981: *Do stock prices move too much to be justified by subsequent changes in dividends,* American Economic Review 75, 1071-1082.

Shumway, R. H. 1988: *Applied Statistical Time Series Analysis*, Prentice-Hall, Englewood Cliffs.

Smith, V.L. 1982: *Microeconomic systems as an experimental science*, American Economic Review **72**, 923-955.

Smith, V.L., 1986: *Experimental methods in the political economy of exchange*, Science **234**, 167-173.

Smith, V.L., G.L. Suchanek and A.W. Williams 1988: *An experimental analysis of stock market bubbles: Prices, expectations and market efficiency*, Finanzmarkt und Portfolio Management **3**, 19-32.

Smith, V.L., G.L. Suchanek and A.W. Williams 1988: *Bubbles, crashes and endogenous expectations in experimental spot asset markets*, Econometrica **56**, 1119-1151.

Tirole, J., 1982: *On the possibility of speculation under rational expectations*, Econometrica **50**, 1163-1181.

White, H. 1993: *Economic prediction using neural networks: the case of IBM daily returns,* Neural Networks in Finance and Investing (Eds. R. Trippi and E. Turban) pp. 315-328 Probus, Chicago.

Williams, A.W. 1979: *Intertemporal competitive equilibrium: On further experimental results*, Research in Experimental Economics **I**, JAI Press.

Williams, A.W. 1987: *The formulation of price forecast in experimental markets*, J. Money, Credit and Banking **19**, 1-18.

Williams, A.W. and V.L. Smith 1984: *Cyclical double-auction on markets with and without speculators*, J. Business **57**, 1-33.

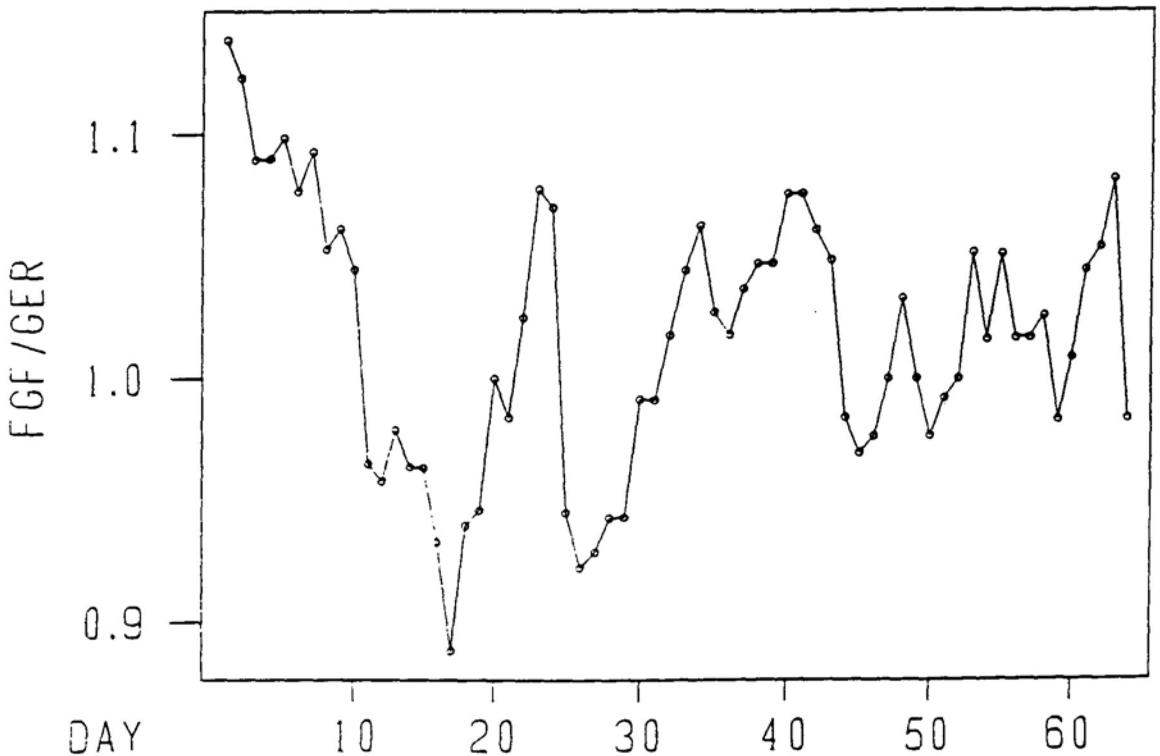

Fig. 1. First quarter of FGF/GER. A daily plot of the prices of the Future Germany Fund divided by the Germany Fund during the first quarter of the former's existence displays the damped oscillations that are confirmed by statistical analysis. For the entire 1149 days, a runs test shows 29 standard deviations from the expected value for a random distribution.

```
ACF

            -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
              +----+----+----+----+----+----+----+----+----+----+
 1    0.745                            XXXXXXXXXXXXXXXXXXXX
 2    0.515                            XXXXXXXXXXXXXX
 3    0.328                            XXXXXXXXX
 4    0.199                            XXXXXX
 5    0.111                            XXXX
 6    0.018                            X
 7   -0.053                           XX
 8   -0.045                           XX
 9   -0.104                         XXXX
10   -0.134                         XXXX
11   -0.212                       XXXXXX
12   -0.228                      XXXXXXX
13   -0.186                       XXXXXX
14   -0.123                         XXXX
15   -0.136                         XXXX
16   -0.168                        XXXXX
17   -0.104                         XXXX
18   -0.062                          XXX
```

```
PACF

            -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
              +----+----+----+----+----+----+----+----+----+----+
 1    0.745                            XXXXXXXXXXXXXXXXXXXX
 2   -0.090                          XXX
 3   -0.050                           XX
 4   -0.007                            X
 5   -0.012                            X
 6   -0.088                          XXX
 7   -0.037                           XX
 8    0.099                            XXX
 9   -0.176                        XXXXX
10   -0.003                            X
11   -0.162                        XXXXX
12    0.040                            XX
13    0.043                            XX
14    0.043                            XX
15   -0.155                        XXXXX
16   -0.092                          XXX
17    0.202                            XXXXXX
18   -0.112                         XXXX
```

Fig. 2. Autocorrelations in the first quarter. A plot of the autocorrelations (*acf*) and the partial correlations (*pacf*) indicates that the ARIMA model that is first order is adequate for describing the statistical properties of the data.

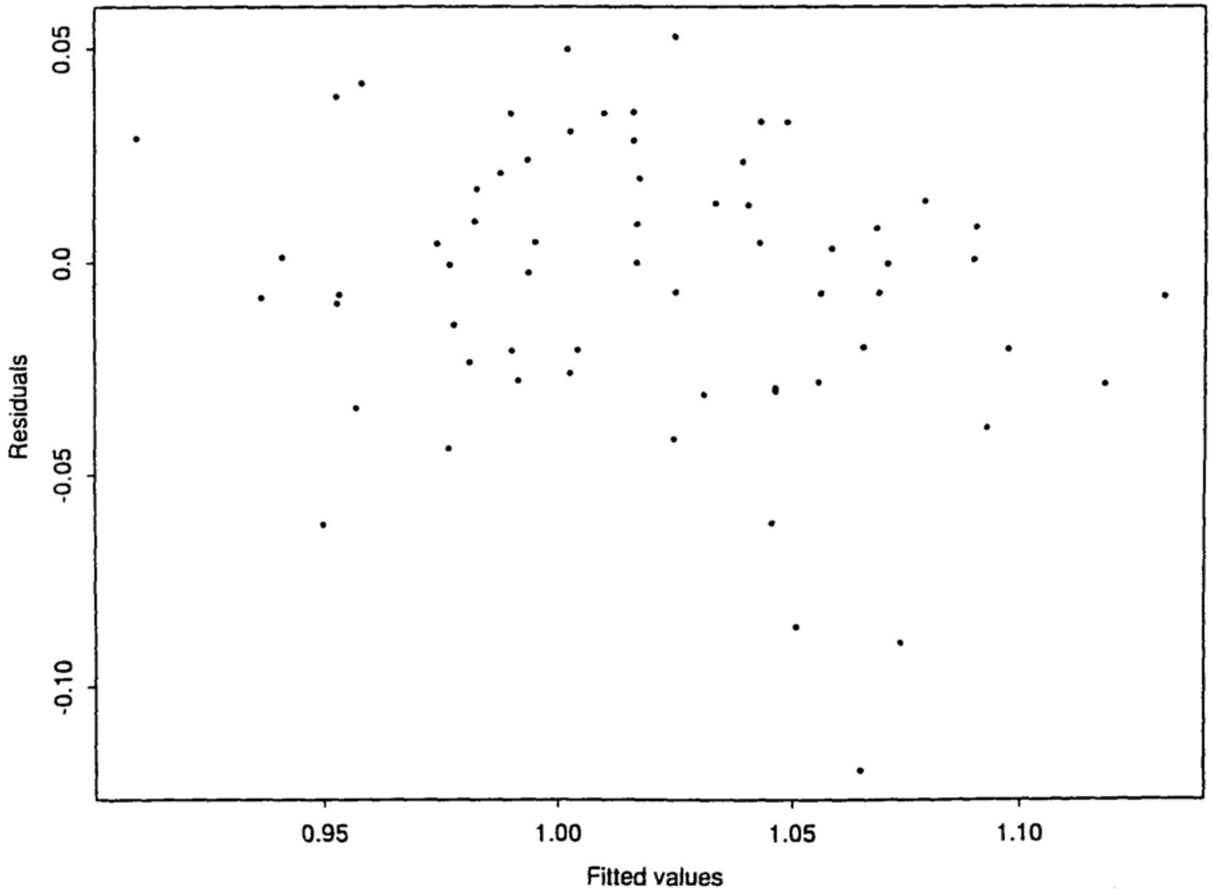

Fig. 3. Display of residuals in first quarter's data. A plot of the residuals (i.e. data minus statistical model) provides no indication of any trends in the residuals, providing a confirmation that the ARIMA (1,1,1) model suitably describes the data.

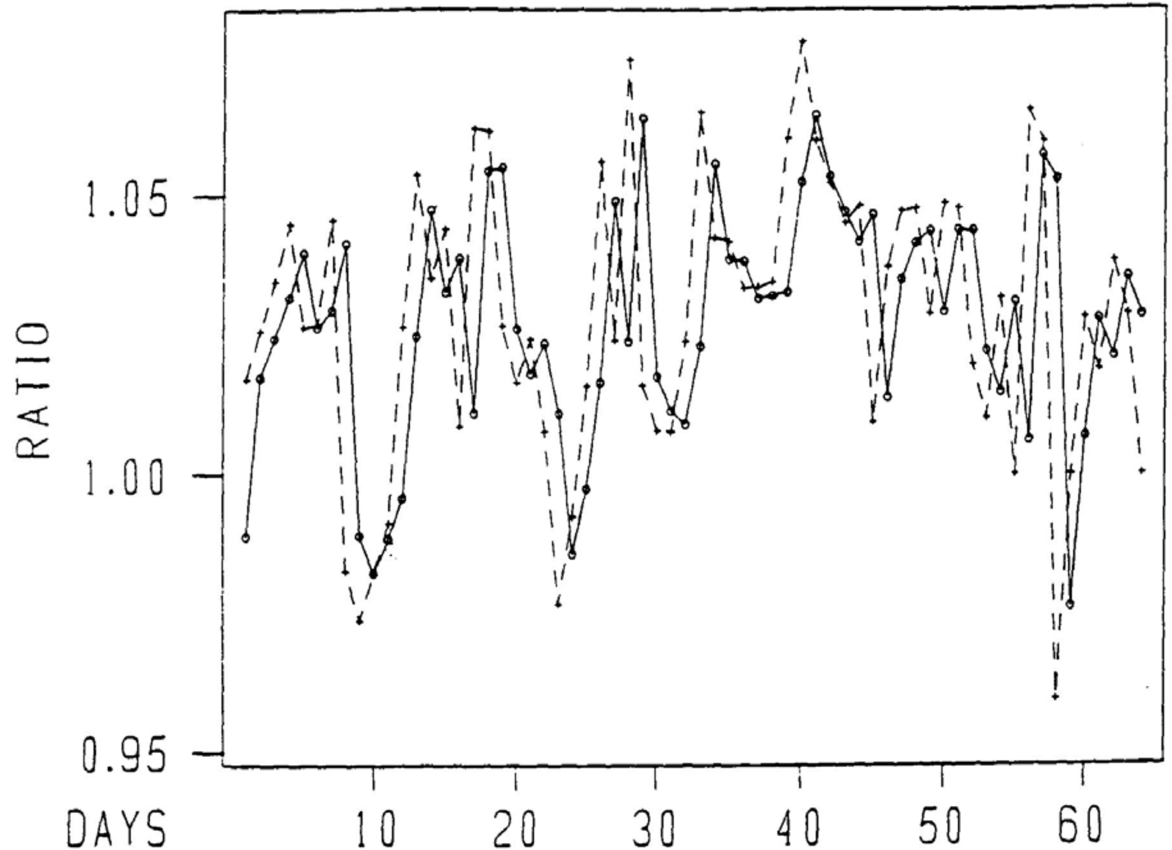

Fig. 4. Out-of-sample ARIMA forecasts and data. The updated ARIMA forecasts for the second quarter of data is graphed along with the data. The predictions are better than the null hypothesis of random walk by three standard deviations (——) forecast; (– – –) data.

Table 1

ARIMA models	Average residual SS	Number of parameters	Bayes	Akaike	FPE
000	29.5584	0	−3.66028	−3.66028	29.6099
100	0.6694	2	−7.43580	−7.44127	0.6729
200	0.6359	3	−7.48103	−7.48923	0.6403
101	0.6270	3	−7.49508	−7.50328	0.6314
201	0.6088	4	−7.51839	−7.52933	0.6141
202	0.6079	5	−7.51373	−7.52740	0.6143
110	0.6407	1	−7.48567	−7.48840	0.6430
210	0.6343	2	−7.48967	−7.49514	0.6376
111	0.6091	2	−7.53015	−7.53562	0.6123
112	0.6080	3	−7.52581	−7.53401	0.6123
212	0.6025	4	−7.52873	−7.53967	0.6078

Table 2. Predictions of laboratory data from market data. Using the ARIMA model developed from the FGF/GER data we make predictions (PRED 1–5) on each of five experiments (EXP 1–5) of Porter and Smith's experiments. The first two data points are required for the initial conditions of the ARIMA model, so that an out-of-sample prediction can be made for the remaining 13 points of each experiment. In comparison with the null hypothesis of an efficient market, the result is significant with a p-value of less than 1 %.

EXP 1	PRED 1	EXP 2	PRED 2	EXP 3	PRED 3	EXP 4	PRED 4	EXP 5	PRED 5
0.66	0.78	0.92	0.99	1.02	1.12	0.62	0.50	0.51	0.57
0.70	0.89	0.96	1.08	1.14	1.27	0.89	0.54	*	*
0.92	0.98	1.02	1.14	1.18	1.40	1.32	0.57	0.97	0.76
1.01	1.05	1.05	1.20	1.38	1.49	1.43	0.60	1.04	0.83
1.14	1.11	1.06	1.24	1.54	1.57	1.62	0.62	1.22	0.89
1.28	1.16	1.09	1.28	1.72	1.64	1.18	0.63	1.22	0.93
1.40	1.20	1.13	1.30	1.97	1.69	1.36	0.65	1.32	0.97
1.56	1.23	1.11	1.33	2.32	1.73	1.56	0.66	1.37	1.00
2.09	1.26	1.25	1.35	2.83	1.77	1.85	0.67	0.86	1.03
2.42	1.28	1.32	1.36	3.59	1.79	2.10	0.67	0.87	1.05
2.66	1.30	1.48	1.37	4.76	1.82	2.44	0.68	1.15	1.06
3.04	1.31	0.95	1.38	4.72	1.83	0.89	0.68	1.29	1.07
1.37	1.32	0.87	1.39	1.12	1.85	1.75	0.69	1.87	1.08