THE PREDICTIVE POWER OF THE LOG-DIVIDEND-PRICE RATIO; EMPIRICAL EVIDENCE FOR THE UNITED STATES, 1926:01-2003:12

GERALDINE RYAN

Department of Economics, University College Cork, Ireland. Email: g.ryan@ucc.ie

Abstract

Using monthly data from 1926:01 to 2003:12 for the United States, this paper examines the predictability of real stock prices based on the dividend-price ratio. In particular, we focus on estimating and forecasting a nonlinear exponential smooth autoregressive model (ESTAR). One motivation for nonlinearity in asset markets is the presence of transaction costs, which result in a nonlinear adjustment process towards equilibrium through arbitrage. Using a novel approach that allows for the joint testing of nonlinearity and nonstationarity, we are able to reject the null hypothesis of linearity and that of a nonlinear unit root. We also find evidence of a nonlinear cointegrating relationship between stock prices and dividends where the error correction term follows a globally stationary ESTAR process. This evidence together with nonlinear impulse response functions, which show that large deviations have faster speeds of mean reversion than small deviations indicates that while stock prices may reflect their fundamentals in the long run, they may deviate substantially from their fundamentals for periods of time. Using an ESTAR-EGARCH model of the dividend-price ratio we find empirical support for in-sample and out-of-sample long-horizon predictability, and we explain why it is often difficult to exploit this predictability using real-time forecasts.

Keywords: Present Value Model of Stock Prices; Nonlinear Unit Root Tests; Nonlinear Cointegration Tests; ESTAR- EGARCH model; Long Horizon Predictability Tests.

JEL: G12, G14, C53

1. INTRODUCTION

Economists have shown considerable interest in the properties of stock prices, with particular attention being paid in the literature to whether stock prices can be characterised as random walk or mean reverting processes. If stock prices follow a mean reverting process, then any shock to the stock price is temporary and there is a tendency for the price level to return to its trend path overtime. Therefore, investors may be able to forecast future returns based on past returns. However, if stock prices follow a random walk process then any shock to the stock price is permanent and there is no tendency for the price level to return to a trend path over time. This suggests that future returns are unpredictable based on historical observations.

Investors and financial economists have expended enormous resources studying the predictability of asset prices. Most early work in finance, such as that by Louis Bachelier (1900) was concerned with finding patterns in asset prices (see also Kendall, 1953; Samuelson, 1965; Mandelbrot, 1966). The findings of these studies, as summarized by Fama (1970), gave little reason to believe that there were any predictable patterns in asset prices that could be consistently exploited by investors to earn abnormal returns.

In recent years, however, empirical research has identified some degree of predictability in asset prices. Fama and French for instance, find that "predictable price variation due to mean reversion … (is) about 40% for 3-5 year return variances (Fama and French, 1988a, 246)." Poterba and Summers (1988) find positive serial correlation over short horizons and strong negative serial correlation over longer horizons. They argue that such patterns may be a result of investors overreacting to news, causing prices to slowly mean revert. McQueen and Thorley (1991) point out that such overreaction stories imply nonlinearities in returns and using a Markov Chain technique that allows for such nonlinearities, they find evidence of non-random behaviour in post-war annual returns.

Campbell and Shiller (1988a,b, 2001) use the present value model to show that earnings and dividends are particularly useful in predicting future returns. According to the present value model, stock prices are fundamentally determined by the discounted present value of expected future dividends (Campbell, Lo and MacKinlay, 1997). Campbell and Shiller (2001) argue that stock prices are not likely to drift too far from their normal levels relative to indicators of fundamental value, such as dividends or earnings. They contend that it seems natural to accentuate the mean-reversion theory that when stock prices are very high relative to these indicators then prices will eventually fall in the future to bring the ratios back to more normal historical levels.

A number of competing theories have tried to explain the deviations of fundamental values from their equilibrium value. These include noise traders (DeLong Shleifer, Summers and Waldmann, 1990; Shleifer, 2000), fads (Shiller, 1989) and stochastic speculative bubbles (Blanchard and Watson, 1982; West, 1988; Evans, 1991), as well as the theory of booms and slumps in economic activity (Phelps, 1994; Phelps and Zoega, 2001). These theories suggest that while stock prices may reflect their fundamentals in the long run, they may deviate substantially from their fundamentals for periods of time (De Long, Shleifer, Summers and Waldmann, 1990).

This paper examines the ability of the log dividend-price ratio to predict excess stock returns at both short and long horizons in the United States, over the period 1926:01-2003:12. Using both equally weighted and value weighted returns data and recent modelling and forecasting techniques[1] we examine the forecast accuracy of Campbell and Shiller' s present value model of stock returns. We test for evidence of nonlinear error correction towards the present value model and then following Kilian (1999) and Kilian and Taylor (2003)[2] we parsimoniously model the nonlinearity using smooth transition autoregressive (STAR) models. Using both equally weighted and value weighted data we find that the exponential smooth transition autoregressive (ESTAR) model appropriately represents the data. This model allows for nonlinear mean-reversion in the dividend-price ratio. Furthermore, we examine whether ESTAR predictors can improve in-sample and out-of-sample forecasts of US excess stock returns using a modified bootstrap procedure based on a nonlinear data generating process.

Evidence of ESTAR suggests that the stock price should be more predictable at longer forecast horizons, at least for large enough sample sizes. Evidence of ESTAR is a very important finding as its existence invalidates the standard errors of long horizon tests based on linear regression analysis. The dividend-price ratio based on the U.S. equally weighted and value weighted data for the period 1926:01-2003:12

[1] See Luukonen, Saikkonen and Teräsvirta, 1988a,b; Teräsvirta and Anderson, 1992; Granger and Teräsvirta, 1993; Teräsvirta, 1994; Mark, 1995; Berkowitz and Giorgianni, 2001; Kilian, 1999; Kilian and Taylor, 2003; and Kapetanois *et al.*, 2003; amongst others.

[2] Kilian (1999) and Kilian and Taylor (2003) concentrate on exchange rate predictability.

is clearly represented by a nonlinear mean reversion process. Therefore, previous empirical research examining long horizon stock price predictability based on linear models should be regarded as invalid. We examine long horizon predictability using a nonlinear data generating process.

This paper is set out as follows: Section 2 briefly reviews the stock price predictability literature. Section 3 introduces the data and discusses some preliminary test statistics while Section 4 carries out unit root tests, cointegration tests, checks for nonlinearities and models any nonlinearities found. In contrast to standard linear methods and empirical studies that test for linearity only, we consider a novel approach that allows for the joint testing of nonlinearity and nonstationarity.[3] We reject the null hypotheses of linearity and nonstationarity indicating nonlinear mean reversion of the dividend-price ratio. Using nonlinear impulse response functions we show that large deviations mean revert at a faster speed than small deviations. We also find a nonlinear cointegrating relationship between stock prices and dividends where the error correction term follows a globally stationary ESTAR process. The evidence of smooth threshold dynamics suggests that stock prices should be predictable at longer horizons, at least for large enough sample sizes.

Section 5 assesses the degree of long-horizon predictability of real stock returns in the presence of smooth-threshold nonlinearities. We use the empirical methodology outlined in Kilian and Taylor (2003) to test the relative forecast accuracy of our long horizon regressions against those of a pure random walk model and a random walk model with a drift. These econometric tests allow us to move beyond the standard analysis of nonlinear models used in existing studies, so that we can better assess the extent of the support for nonlinear models of stock price behaviour. Finally, Section 6 summarises the findings of this paper.

[3] For a growing literature that addresses the joint issues of nonlinearity and nonstationarity see Michael *et al*., 1997; Enders and Granger, 1998; Berben and van Dijk, 1999; van Dijk *et al*., 2001; Kapetanios *et al*., 2003 and Kapetanios *et al*., 2004.

2. STOCK PRICE PREDICTABILITY AND LONG-HORIZON REGRESSION ANALYSIS

A large body of empirical work has accumulated documenting stock return predictability. One of the most popular predictors is the dividend-price ratio.[4] Among those examining the ability of the dividend-price ratio to predict stock price behaviour are Fama and French (1988a, b), Campbell and Shiller (1988a,b, 1998, 2001), Ferson (1989), Goetzmann and Jorion (1995), Hodrick (1992), Stambaugh (1999), Goyal and Welch (2003) and Valkanov (2001).

The foundation for much of the research into stock price predictability is based on the present value model of stock prices. This model states that stock prices are fundamentally determined by the discounted value of their expected future dividends (Campbell, Lo and MacKinlay, 1997). Early research based on the present value model, with constant discount rates found that stock price movements could not be explained solely by dividend variability. Leroy and Porter (1981) and Shiller (1981), for example, found that, under the assumption of a constant discount factor, stock prices were too volatile to be consistent with movements in future dividends. This conclusion, known as the excess volatility hypothesis, argues that stock prices exhibit too much volatility to be justified by fundamental variables. While a number of papers challenged the statistical validity of the variance bounds tests of Leroy and Porter and Shiller, on the grounds that stock prices and dividends were non-stationary processes (Flavin, 1983; Marsh and Merton, 1986), much of the subsequent literature, nonetheless, found that stock price movements could not be explained solely by dividend variability as suggested by the present value model with constant discounting (Campbell and Shiller, 1987; West 1988).

More recent research by Campbell and Shiller (1988a, b); Campbell (1991); Cochrane (1991, 1992) and Timmerman (1995) argue that the present value model with time-varying discount rates can explain fluctuations in stock prices. Results based on this form of the present value model are as mixed as those based on the present value model with constant discount rates. Froot and Obstfeld (1991) are unable to come to a decisive conclusion using data for the United States from 1900 to 1988. They contend that the results are dependant on the specification of the unit root

[4] Other popular predictors include interest rates (see Fama and Schwert, 1977; Glosten, Jagannathan and Runkle, 1993) and the price-earning ratio (see Ang and Bekaert, 2001; Rapach and Wohar, 2005).

tests. Both Lamont (1998) and Balke and Wohar (2002) find evidence of nonstationarity in the log dividend-price ratio for the United States using quarterly data from 1947:01 to 1994:04 and 1953:02-1999:01 respectively. However, while Lamont finds evidence of cointegration between dividends and stock prices using a bivariate Horvath-Watson (1995) test Balke and Wohar (2002) do not concur with this finding.

Gallagher and Taylor (2001), Bohl and Siklos (2004), Coakley and Fuertes (2004) and Kanas (2005) amongst others argue that the failure of the present value model with time varying discount rate is a result of the way in which the dividend-price ratio is modelled. Each of these studies argues that the dividend-price ratio should be modelled as a nonlinear process. They contend that while the present value model with a time varying discount rate may be a valid representation of the long-run behaviour of the stock price, it does not allow for short-term deviations in the ratio.[5]

One set of models which allows the present value model to hold in the long-run but to deviate from its equilibrium for short periods of time is the Smooth Transition Autoregressive (ESTAR) models (see Granger and Teräsvirta, 1993; Teräsvirta, 1994). These models allow the dividend-price ratio to exhibit random walk behaviour when it is close to equilibrium and mean-reverting behaviour as it deviates further away from its equilibrium value. Aslanidis (2002) maintains that this type of model is very appropriate in a stock market where there are a large number of participants, each switching at different times due to a number of reasons including heterogeneous beliefs, varying learning speeds, and different investment horizons. The nonlinear representation of the dividend-price also enables us to allow for 'limits to arbitrage' in our present value model (see Gallagher and Taylor, 2001; Kapetanois et al, 2004).[6] Kilian and Taylor (2003) contend that this type of long-run mean reverting behaviour may improve predictability.

---

[5] This hypothesis is consistent with the view that the stock market is efficient in the long run but deviates from its fundamental value in the short run due to factors such as noise traders, booms and slumps in the economy etc. (see for example Blanchard and Watson, 1982; De Long, Shleifer, Summers and Waldmann, 1990; Evans, 1991; Phelps, 1994; Phelps and Zoega, 2001; Shleifer, 2000; Shiller, 1989; West, 1988).

[6] In reality arbitrage, (defined as the simultaneous purchase and sale of the same, or essentially similar, security in two different markers for advantageously different prices) opportunities are limited by a number of factors like the existence of transaction costs, short-selling constraints, or mispricing of securities deepening in the short run. Given a distribution of degrees of risk aversion across smart traders, arbitrage will increase as the degree of fundamental mispricing increases, so that arbitrage is stabilising and becomes more stabilising in extreme circumstances. Traditional arbitrage models,

The question of stock price predictability is fundamentally a question of stock market efficiency. If the stock market is efficient then the random-walk theory of the stock market states that stock price changes are not predictable. In other words the dividend-price ratio has no ability to forecast movements in stock prices. Campbell and Shiller (2001) argue however, that for the dividend-price ratio to remain within its historical range then the dividend-price ratio must predict future growth in dividends. As there is little empirical evidence to support this claim they continue by questioning whether the dividend-price ratio forecasts future dividend movements as required by the random-walk theory, or whether it forecasts future movements in stock prices. Using graphical analysis they conclude that in some countries such as France, Germany and Italy the dividend-price ratio does not appear to forecast future dividend growth, whereas in other countries such as Australia, Canada, Spain, Japan and the US the dividend-price index appears to forecast stock price behaviour.

Several formal approaches have been adopted in the literature to evaluate stock price predictability. Among these are variance-ratio tests, long-horizon regressions and vector autoregressive techniques. These predictability tests have important implications for asset pricing and market efficiency. In an efficient capital market, equity prices reflect currently available information and one should not be able to predict future returns using historical returns data. Therefore, if returns are predictable, it could imply market inefficiency. To date, the literature indicates that in the absence of market efficiency, deviations of asset prices from their long-run equilibrium value should help predict cumulative future asset returns (Kilian, 1999). This predictability proposition is frequently tested using long-horizon regression tests (see for example: Ang and Bekaert, 2001; Berkowitz and Giorgianni, 1997; Campbell and Shiller, 1998a,b; Fama and French, 1988a,b; Hodrick, 1992; Kilian, 1995; Kilian and Taylor, 2003).

2.1. Long-Horizon Regression Analysis

The long-horizon regression approach entails estimating k individual equations:

$$\Delta s_{t+k}^{k} = \alpha_k + \beta_k z_t + e_{t+k}^{k} \qquad k=1,2,\ldots K \qquad (1)$$

therefore, imply a degree of nonlinearity in asset price dynamics (see Gallagher and Taylor, 2001; Cuthbertson, and Nitzsche, 2004).

Where s_t and z_t are observed data, Δ denotes the first difference, k is the horizon length and α_k and β_k are the parameters to be estimated. In the stock price predictability scenario s_t represents the log of the real stock price and z_t represent the log dividend-price ratio.

We can examine the statistical significance of (1) using either in-sample or out-of-sample tests. In-sample forecasts are based on the full sample of data whereas the out-of-sample forecasts are evaluated using the sample of data available to the trader at each moment in time.

2.2. In-Sample Analysis

In-sample mean reversion in the stock-price dividend ratio may be detected by a t-test of $H_0:\beta_k = 0$ versus $H_1: \beta_k < 0$ for a given horizon for some k in equation (1) for all forecast horizon as $H_0:\beta_k = 0\ \forall$ k versus $H_1:\beta_k <0$ for some k.[7]

Berkowitz and Giorgianni (1997) argue that if the t-statistics and the regression R^2's are found to increase with k, then the researcher can take this as evidence that z_t can predict long-run changes in s_t better than short-run movements. However, Kilian (1995) argues that under the alternative hypothesis, the slope coefficients will increase with the forecast horizon, so that evidence of increasing slopes and R^2 measures do not imply increased long-horizon predictability.

Hodrick (1992) investigates the predictability of stock returns at five horizons, from one month to four years, for the US from 1952 to 1987. Using dividend yields as the regressor, Hodrick finds strong evidence that dividend yields predict stock prices at long horizons. His findings suggest that stock prices are predictable at the 12-month horizon. Campbell and Shiller (1998a, b) find that the price-dividend and the price-earning ratios are useful for forecasting changes in real stock prices at long horizons. Using annual S&P 500 data from 1871 to 1997, they use scatter plots and R^2 measures to indicate a weak ability for the price-dividend ratio to forecast real stock price growth over the next year, but a strong ability for the price-dividend ratio to forecast real stock price growth over the next ten years.

A number of authors including Hodrick (1992), Nelson and Kim (1993) and more recently, Ang and Bekaert (2001) have pointed out that the finite sample distribution of the long-horizon regression coefficient and its associated t-statistic can

[7] See Mark (1995) for further details on this joint hypothesis test.

be quite different from the asymptotic distribution due to persistence in the independent variable and overlap in the returns data.

One key issue with long horizon regression analysis is that when the horizon k is larger than one, the dependent variable in (1) becomes time overlapping. This induces an $MA_{(N-1)}$ component in the error terms, which needs to be corrected for in the estimation. The standard way of doing this correction is to apply the methods of Newey and West (1987). However, statistics based on these corrections will approximate the relevant asymptotic distributions very poorly in finite samples, especially when the degree of time-overlap becomes large. In addition, when k grows Δs_{t+k}^{k} becomes more and more persistent and will, in a finite sample, often be indistinguishable from non-stationary unit root processes.

This dilemma has lead to a large body of literature, which questions the interpretation of long-horizons regression test results. For example, Mankiw, Romer and Shapiro (1991), Hodrick (1992), Nelson and Kim (1993), and Berkowitz and Giorgianni (2001) have all reported that conventional long-horizon regression tests are biased in favour of finding predictability. They argue that severe size distortions may arise from spurious regression fits and from small-sample bias in the estimate of regression coefficients and asymptotic standard errors.

Mark (1995), Chinn and Meese (1995) and Bauer (1995) attempt to correct for these problems by generating bootstrap critical values for the diagnostic tests associated with the long horizon regressions. Mark (1995) generates pseudo-data for long horizon regressions to examine the question of exchange rate predictability. He argues that the change in the exchange rate should be modelled as a random walk process and the fundamental should be modelled as a linear autoregressive process as follows

$$\Delta e_t = a_0 + \varepsilon_{1,t}$$

$$z_t = b_0 + \sum_{j=1}^{p} b_j z_{t-j} + \varepsilon_{2,t} \qquad (2)$$

where e_t is the log of the domestic-currency price of one unit of foreign exchange and $z_t \equiv f_t - e_t$, where f_t is the purchasing power parity relationship calculated as the logarithm of the domestic consumer price index minus the logarithm of the foreign consumer price index. Using data for Canada, Germany, Japan and Switzerland from 1973:01 to 1991:04, Mark finds a pattern of increased long-horizon predictability.

While only some of his long-horizon regression test statistics were significant at conventional levels, Mark conjectured that only the small sample size prevented more of his results from being significant.

Kilian (1995) and Berkowitz and Giorgianni (2001) disagree with Mark's (1995) findings. They assert that the bootstrap procedure used by Mark is not entirely correct, and may result in spurious inference. Berkowitz and Giorgianni (2001) explain that the linear data generating framework postulated by Mark (1995) implies that real stock prices should be predictable at all horizons or at no horizon. This results from the fact that in a linear framework, long-horizon forecasts are simple extrapolations of short-horizon forecasts. As Mark's result depends on the stationarity of $z_{t,}$ the bootstrap is likely to be unreliable unless one corrects the bias in the initial slope coefficients of the lags of z_t. Re-estimating Mark's dataset correcting for the implied bias, Berkowitz and Giorgianni (2001) only report one significant slope coefficient and this is at the 90 percent confidence interval.

One possible explanation for the pattern of stock price predictability in the data focuses around statistical power. According to Berkowitz and Giorgianni (2001) while the argument that a linear framework implies predictability at all horizons or no horizon is logically correct, it may be the case in certain circumstances that the power to detect predictability in a linear framework is greater at long horizons. Berben and van Dijk (1998), Mark and Sul (2002), Campbell (2001), Kilian (1999) and Kilian and Taylor (2003) investigate whether there are power advantages at long horizons in predictive regression tests using various asymptotic frameworks and Monte Carlo simulations for finite samples. While the results are somewhat mixed the preponderance of studies find potential power gains at long horizons.

Recent developments in the literature indicate that the underlying data-generating process (DGP) of many variables including exchange rates (See Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001; Kilian and Taylor, 2003) should be represented as nonlinear processes. For example, Kilian and Taylor (2003) argue that the data generating progress for a long horizon regression model examining the predictability of exchange rates should be characterised as follows

$$\Delta e_t = a_0 + \varepsilon_{1,t}$$

$$z_t = \mu + \lambda\,(z_{t-1}-\mu) - [1-\exp\{-\gamma\,(z_{t-d} - \mu)^2\}][\lambda(z_{t-1}-\mu)]+\varepsilon_{2,t} \qquad (3)$$

Where similar to Mark (1995) the change in the exchange rate is modelled as a random walk process, however here the fundamental is modelled as a nonlinear smooth transition autoregressive (STAR) process.

2.3. Out-of-Sample Tests

Even a sophisticated trader can only use prevailing information to estimate the long horizon regression model, (1) and therefore it is important that we evaluate our model using only real time data. The random walk model is a natural benchmark in judging forecast performance. Many statistics have been identified in the literature to compare the performance of the augmented model with the performance of the respective benchmark model. In this section we focus on four of these, two tests of equal forecast accuracy and two tests for forecast encompassing. In particular, we consider the t-statistic for equal MSE developed by Diebold and Mariano (1995) and West (1996) and the F-statistic proposed by McCracken (2000). We also consider the t-statistic for forecast encompassing developed in Harvey, Leybourne and Newbold (1998) and West (2001) and the variant proposed by Clark and McCracken (2001).

Comparison of the forecasts from a benchmark model with those from an augmented model enables us to determine the added value of the features of the augmented model, if any. To carry out these forecasts, the total sample, of T observations are divided into an in-sample and an out-of-sample portion, where the in-sample portion spans the first R observations and the out-of-sample portion the last (P-k+1) observations. We then estimate the benchmark and augmented models using the in-sample portion of the total sample, and we use the estimated models to generate two series of (P-k+1) one-step-ahead out-of-sample forecasts, one corresponding to the fitted benchmark model and the other to the fitted augmented model. We denote the one-period out-of-sample forecast errors for the benchmark model as $\{\hat{u}_{A,t+1}\}_{t=R}^{T-k}$ and the augmented model as $\{\hat{u}_{B,t+1}\}_{t=R}^{T-k}$. Forecasts are recursively updated to generate a time series of one-period ahead forecasting errors $\{\hat{u}_{i,t+k}\}_{t=R}^{T-k}$, where i = A, B and t = R+1, …, T+1, giving a total of P = T+1-R observations.

The first test is the Diebold and Mariano (1995) predictive accuracy test. This statistic provides a statistical comparison of the accuracy of two competing forecasts, A and B, using the loss differential. This is computed as the difference between the forecast errors:

$$\hat{d}_{t+\tau} = (\hat{u}_{A,t+k})^2 - (\hat{u}_{B,t+k})^2 \qquad t = 1, 2, \ldots T \tag{4}$$

Where $(\hat{u}_{i,t+k})^2 = (p^i_{t+k} - \hat{p}^i_{t+k})^2$ and i = A, B. The large-sample N(0, 1) statistic for testing the null of equal forecast accuracy is given by

$$MSE - T = (P - k + 1)^{1/2} * \frac{\bar{d}}{\sqrt{\phi_{dd}}} \tag{5}$$

where $\bar{d} = (P - k + 1)^{-1} \sum_{t=R}^{T-k} \hat{d}_{t+k}$ = MSE_A-MSE_B, is the average loss differential, ϕ is its asymptotic variance that, as suggested by Diebold and Mariano (1995) can be estimated by an unweighted sum of d_t's autocovariance: $\hat{\phi}_{dd} = \sum_{i=-(k-1)}^{k-1} \hat{\gamma}_t(d)$, where k is the forecast horizon.[8] The null hypothesis is that there is no significant difference in the accuracy of the competing models conditioning on being in a particular regime, hence the difference in the MSE's will be less than or equal to 0. Under the alternative, MSE_B should be smaller than MSE_A. Hence the MSE-T test and the other equal accuracy test are one-sided to the right.[9]

The second test is the ENC-T test. Drawing on the methodology of Diebold and Mariano (1995), Harvey, Leybourne, and Newbold (1998) propose this encompassing test which uses a t-statistic for the covariance between $u_{A,t+k}$ and $u_{A,t+k}$-$u_{B,t+k}$. To estimate this statistic, we estimate

$$\hat{c}_{t+\tau} = \hat{u}_{A,t+k}(\hat{u}_{A,t+k} - \hat{u}_{B,t+k}) \tag{6}$$

The large-sample N(0, 1) statistic for testing the null of equal forecast accuracy is given by

$$ENC - T = (P - k + 1)^{1/2} * \frac{\bar{c}}{\sqrt{\phi_{cc}}} \tag{7}$$

where $\bar{c} = (P - k + 1)^{-1} \sum_{t=R}^{T-\tau} \hat{c}_{t+\tau}$ and $\hat{\phi}_{cc} = \sum_{i=-(k-1)}^{k-1} \hat{\gamma}_t(\hat{c})$. Under the null that model A's forecast encompasses model B, the covariance between $u_{A,t+\iota}$ and $u_{A,t+\iota}$- $u_{B,t+\iota}$ will be less than or equal to 0. Under the alternative that model B contains added

---

[8] Note, if the DM statistic is computed based on a one-step-ahead forecast then $\hat{\phi}$ reduces to $\hat{\gamma}_0$, the variance of d_t.

[9] See Clark and McCracken (2004) for further details.

information, the covariance should be positive. Hence the ENC-T test as well as the other encompassing test described below is one-sided to the right.

The third test, the ENC-NEW test, is also a forecast encompassing test. Clark and McCracken (2001) derive the asymptotic distribution of the ENC-NEW statistic under the null hypothesis that the augmented model encompasses the information of the benchmark. The statistic of the ENC-New test is

$$\mathrm{ENC-NEW} = (\mathrm{P}-\mathrm{k}+1) * \frac{(\mathrm{P}-\mathrm{k}+1)^{-1} \sum_{t=R}^{T} \left[\hat{u}_{A,t+1}^{2} - \hat{u}_{A,t+1} \hat{u}_{B,t+1}\right]}{(\mathrm{P}-\mathrm{k}+1)^{-1} \sum_{t=R}^{T} \hat{u}_{B,t+1}^{2}} \tag{8}$$

The fourth test, the MSE-F test, is

$$\mathrm{MSE-F} = (\mathrm{P}-\mathrm{k}+1) * \frac{\sum_{t=R}^{T} \left[\hat{u}_{A,t+1}^{2} - \hat{u}_{B,t=1}^{2}\right]}{\sum_{t=R}^{T} \hat{u}_{B,t+1}^{2}} \tag{9}$$

The null hypothesis here states that the benchmark model has a mean-squared forecasting error less than or equal to the error of the augmented model; the alternative is that the augmented model has a smaller mean-squared error. Clark and McCracken show that these two tests - the ENC-NEW and the MSE-F tests - have the best overall power and size properties.[10]

It is well known that asymptotic critical values for these test statistics are severely biased in small samples. In order to mitigate these size distortions critical values may be calculated based on the bootstrap approximation of the finite sample distribution of the test statistic under the null hypothesis of no predictability in the cointegrated model or some equivalent representation of the data-generating process (McCracken, 2000). Unlike asymptotic critical values, bootstrap critical values based on the percentiles of the bootstrap distribution automatically adjust for the increase in the dispersion of the finite-sample distribution of the test statistic that occurs in near-spurious regressions as the sample size grows. (Kilian, 1999) As a result, bootstrap inference is immune from the near-spurious regression problem discussed in Berkowitz and Giorgianni (2001).

[10] We examine the size and power of each of the out-of-sample tests as applied to our data generating process in Section 5 below.

Before examining the in-sample and out-of-sample results of our long horizon stock price predictability regressions we need to examine the data and the nature of the nonlinearity in the log dividend-price index in the US for the period 1927:01-2003:12, if any.

3. DATA DESCRIPTION

Prior to setting up the empirical model we briefly outline our data set. The variables of interest are: the value weighted stock market price and its associated dividend yield; the equally weighted stock market price and its associated dividend yield; the one-month Treasury bill return and the inflation rate. All of the data is obtained from the Centre for Research in Security Prices of the University of Chicago's School of Business (CRSP) Database. The data is available from 1925:12-2003:12.[11]

The monthly value-weighted return series without dividends (RWD_t) is used to calculate the nominal value-weighted stock price. The value-weighted nominal stock price, NP_t, is calculated as $(1+RWD_t)*NP_{t-1}$. The value-weighted nominal dividend series is calculated using both the monthly value-weighted return series with dividends (RD_t) and the monthly value-weighted return series without dividends (RWD_t). The value-weighted nominal dividend series, ND_t, is calculated as $(RD_t - RWD_t)*NP_{t-1}$. The equally-weighted stock price and dividend series are calculated in the same way.

Monthly inflation rates, π_t, are used to calculate the monthly nominal goods price level. The normalised nominal goods price level series, CPI_t is produced by setting the price in December 1925 equal to 1 and recursively setting $CPI_t = (1+\pi_t)*CPI_{t-1}$. The nominal stock price and dividend series are deflated by the CPI to give the real stock price (P_t) and the real dividend (D_t) series. The log of real stock prices is denoted by p_t and log dividends by d_t.

The one-period real return series is calculated as $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$, where P_t is the end-of-month real stock price and D_t is the real dividends paid during month t. Excess returns are calculated as the one-period real return series minus the one-period return on a one-month Treasury bill.

[11] The first twelve observations are used to calculate the annualised dividend index.

To compute the dividend-price ratio we follow the approach of Hodrick (1992). Since dividend payments are highly seasonal, a monthly annualised dividend series, MD_t, is computed from compounding twelve monthly dividends at the 1-month Treasury bill rate r_t:

$$MD_t = D_t + (1 + r_t)D_{t-1} + (1 + r_t)(1 + r_{t-1})D_{t-2} + ... + (1 + r_t)(1 + r_{t-1})D_{t-11} \quad (10)$$

The annual dividend-price series is defined as: $Z_t = D_t / P_t$.

Table 1 reports some summary statistics on the log of the stock price, the log of the dividend series, the change in the log of the stock price, the change in the log of the dividend series, the returns series, the excess returns series, the log dividend price series and change in the dividend price series.[12] The results are largely as expected. Normality is rejected in each of the series and there is strong evidence of both skewness and kurtosis. The latter may imply that there are outlying observations; that the error process is heteroskedastic; or that the data would be better described by using a nonlinear time-series model.

The sample autocorrelations of the price series and the dividend series, both in levels and logs, reveal some degree of persistence. The first-order autocorrelation values are close to one, which suggests that these series are non-stationary. On the other hand, the sample autocorrelations of the first difference series are low and insignificant indicating that the series are first difference stationary.

4. MODELLING THE DIVIDEND-PRICE RATIO

The empirical analysis begins by testing for nonlinearity in the value-weighted and equally-weighted dividend-price ratio series. If evidence of nonlinearity is found we fit appropriate nonlinear models to the series. Next, model diagnostics are used to examine the appropriateness of our models and graphs of the transition functions are used to characterise our models. Any heteroskedasticity in the data is modelled using (G)ARCH models. Finally, the most appropriate parsimonious models are used to generate impulse response functions for both the value-weighted and equally-weighted dividend price ratio series. These functions investigate whether the speed of

[12] In each of the tables which follow, part (a) refers to the value-weighted series whereas part (b) refers to the equally-weighted series.

adjustment towards equilibrium increases with the size of a shock to dividend-price ratio.

4.1. Nonlinearity Test

The first step in testing linearity is to select the order of the AR(p).[13] We follow Tsay (1989) in using partial autocorrelation functions (PACF) to select the appropriate lag order of the AR. Figure 1, part (a) examines the PACF of the equilibrium error of the value-weighted log dividend-price ratio. It reveals correlations up to the order one. The PACF of the equilibrium of the equally-weighted log dividend-price ratio is shown in Figure 1, part (b). This figure reveals correlation up to order two. Therefore we conclude that the value-weighted log dividend-price series is best modelled as an AR(1), whereas the equally-weighted series is best modelled as an AR(2).

Next we examine whether the log dividend-price ratio contains any nonlinearity. To do this we estimate the following artificial regression

$$y_t = \beta_{00} + \sum_{i=1}^{p} \left[\beta_{1i} y_{t-1} + \beta_{2i} y_{t-i} y_{t-d} + \beta_{3i} y_{t-i} y_{t-d}^2 + \beta_{4i} y_{t-i} y_{t-d}^3 \right] + \varepsilon_t \qquad (11)$$

where y_t represents the demeaned dividend-price ratio, i represents the order of the autoregressive component and d represent the order of the delay function. Based on the partial autocorrelation functions discussed above we set the autoregressive component, i, equal to one in each country, while we estimate the delay parameter, d, using a grid search procedure. To examine whether our series is linear we test the null hypothesis whether $\beta_2 = \beta_3 = \beta_4 = 0$. If we reject this null hypothesis then our series is nonlinear. In our grid search if we reject the null hypothesis for more than one delay parameter, d, then we select the delay parameter with the smallest probability.

Once we have identified the delay parameter we can examine whether the nonlinearity is best characterised by an exponentially smooth transition autoregressive (ESTAR) process or by a logarithmic smooth transition autoregressive (LSTAR) process. We can examine which process is viable, if any, using a sequence of nested tests based on our artificial regression (11).[14] These tests are as follows:

$$H_{03}: \beta_{4i} = 0 \qquad i = 1, \ldots, p \qquad (12a)$$

---

[13] In modelling and testing for nonlinearity we use the demeaned log dividend-price ratio.

[14] These tests were proposed by Teräsvirta (1994).

$$H_{02}: \beta_{3i} = 0 \mid \beta_{4i} = 0 \qquad i = 1, \ldots, p \tag{12b}$$

$$H_{01}: \beta_{2i} = 0 \mid \beta_{3i} = \beta_{4i} = 0 \qquad i = 1, \ldots, p \tag{12c}$$

If we reject H_{02} then the log dividend-price ratio is modelled as an ESTAR process, otherwise it is modelled as an LSTAR process.

Table 2 (a) reports the results for the value-weighted series, while Table 2 (b) reports the results for the equally-weighted series. In the first column of the tables we report the results of the nonlinearity test H_L. We clearly reject the null hypothesis of linearity and select a delay of 2 for the value-weighted series and a delay of 3 for the equally-weighted series. Armed with this information we can now establish whether the log dividend-price ratios should be modelled as ESTAR or LSTAR processes.

Columns two, three and four of Table 2 report the results for the null hypotheses H_{03}, H_{02} and H_{01} respectively as outlined in (12a) to (12c) above. Examining the results for the value-weighted series in part (a) of Table 2, we see that the tests conclude that the log dividend-price ratio may be modelled as either an ESTAR or an LSTAR process at the 10% level of significance, however using the 5% level of significant, the results select an ESTAR process. The results in part (b) of the table relating to the log of the equally-weighted dividend-price ratio clearly select an ESTAR process.

As equation (11) is an artificial regression and hypotheses (12a) to (12c) are based on this regression we need to confirm that our results are correct by estimating the smooth transition autoregressive (STAR) model:

$$y_t - \mu = \sum_{j=1}^{p} \lambda_j \left[y_{t-j} - \mu \right] + \sum_{j=1}^{p} \lambda_j^* \left[y_{t-j} - \mu \right] \Phi\left[\langle y_{t-d} \rangle_{d=1}^{\bar{d}} ; \mu ; \gamma \right] + \varepsilon_t \tag{13}$$

Using a grid search procedure and both the Logistic Smooth Autoregressive Transition (LSTAR):

$$\Phi\left[\langle y_{t-d} \rangle_{d=1}^{\bar{d}} ; \mu ; \gamma \right] = \left(1 + \exp\left[-\gamma \sum_{d=1}^{\bar{d}} \left[y_{t-d} - \mu \right] \right] \right)^{-1} \tag{14a}$$

and the Exponential Smooth Autoregressive Transition (ESTAR)

$$\Phi\left[\langle y_{t-d} \rangle_{d=1}^{\bar{d}} ; \mu ; \gamma \right] = \left(1 - \exp\left[-\gamma \sum_{d=1}^{\bar{d}} \left[y_{t-d} - \mu \right]^2 \right] \right) \tag{14b}$$

characterisations of the transition function so that we can verify that we have selected the most appropriate delay parameter and form of the transition function. Here we select the delay parameter based on the probability level associated with the transition parameter, γ. We select the transition parameter with the smallest probability parameter. The results of these tests[15] confirm that the log of the value-weighted dividend-price ratio is an ESTAR process with a delay of 2 and the log of the equally-weighted dividend-price ratio is an ESTAR process with a delay of 3.

4.2. Stationarity and Cointegration Tests

Now that linearity has been rejected and an ESTAR model has been chosen for both the value-weighted and the equally-weighted series we can test for nonlinear stationarity and nonlinear cointegration using the procedure developed by Kapetanios *et al.* (2003, 2004). For the purpose of comparison, we also report the conventional Augmented Dickey-Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test statistics, denoted by t_{ADF}, t_{PP}, $KPSS_{\mu}$ and $KPSS_{\tau}$ respectively (see Dickey and Fuller, 1979, 1981; Kwiatkowski *et al.*, 1992; Perron, 1988).

As suspected the linear unit root tests suggest that the null hypothesis of a unit root cannot be rejected for the log stock price, excess stock price, dividend or dividend-price ratio series (see Table 3). In each case the test statistics suggest that the variables are first difference stationary.

The nonlinear unit root tests strongly reject the null hypothesis of a unit root in the ESTAR modelled dividend-price series (see Table 4). The major implication of the finding of nonlinear stationarity in the dividend-price series is that although real stock prices consistently deviate from their long run equilibrium, the deviation is nonlinearly mean-reverting. The strong results of the ADF, PP and KPSS tests and the result in the nonlinear unit root test suggests that stocks and dividends have roots of the same order, i.e. they may be cointegrated.

The loglinear present value model shows that when the log of real stock prices and the log of real dividends are first-difference stationary they are cointegrated with a cointegrating vector $(1,-1)'$. Therefore, the long-run equilibrium relationship described by the present value model is given by $p_t = d_t$. The Ordinary Least Squares

---
[15] These results are not presented here but they are available on request.

(OLS) regression of log real stock prices on log real dividends and a constant are presented in Table 5.

Examining the linear Engle Granger and the Error Correction Model statistics for the value-weighted series, we cannot reject the null hypothesis of no linear cointegration at the 5% level. Examining the results of our nonlinear cointegration tests we can easily reject the null hypothesis of no cointegration in favour of a globally stationary nonlinear ESTAR cointegration.

The equally-weighted series, on the other hand, rejects the null hypothesis of no cointegration in both linear and nonlinear versions of our tests. Overall, the test results clearly demonstrate adjustment towards equilibrium over the long run. This suggests that the log dividend-price ratio is in fact mean-reverting.

4.3. ESTAR Estimation Results

ESTAR models are fitted to the dividend-price series. We report the parsimonious form of the estimated ESTAR model. The results are based on the demeaned log real dividend-price model.

We begin by estimating an exponential smooth transition autoregressive model expressed as:

$$z_t - \mu = \sum_{j=1}^{p} \lambda_j \left[z_{t-j} - \mu \right] + \sum_{j=1}^{p} \lambda_j^* \left[z_{t-j} - \mu \right] \Phi\left[\langle z_{t-d} \rangle_{d=1}^{\bar{d}} ; \mu ; \gamma \right] + \varepsilon_t \qquad (15)$$

where z_t is the demeaned log dividend-price ratio and d is the delay parameter. To find a proper initial value for the transition parameter, γ, we standardise the model by dividing the exponential part of the transition function by σ_z^2, the sample variance of z_t (Teräsvirta, 1994). Granger and Terasvirta (1993) argue that scaling the exponential term by the sample variance speeds the convergence and improves the stability of the nonlinear least squares estimation algorithm. It also makes it possible to compare estimates of transition parameters across equations. Since we are unable to reject the null hypothesis that $\lambda_j = -\lambda_j^*$ we re-parameterise the ESTAR model as follows:

$$z_t - \mu = \sum_{j=1}^{p} \lambda_j \left[z_{t-j} - \mu \right] - \sum_{j=1}^{p} \lambda_j \left[z_{t-j} - \mu \right] \Phi\left[\langle z_{t-d} \rangle_{d=1}^{\bar{d}} ; \mu ; \gamma \right] + \varepsilon_t \qquad (16)$$

Table 6 reports the results. The estimated models perform well in terms of providing goodness of fit and statistically significant coefficients. The log-likelihood tests show a clear preference for the ESTAR models over an autoregressive alternative. The speed of adjustment of the transition variables can be clearly identified in Figure 2. The Variance Ratio (VR) indicates a reduction of 2.6% in the unexplained component of the equilibrium error for the value-weighted log dividend-price ratio and a reduction of 2.7% for the log of the equally-weighted dividend-price ratio.

The ARCH test results indicate that there is substantial conditional heteroskedasticity in both the value-weighted and equally-weighted series. Therefore, it is imperative that we re-estimate the ESTAR models accounting for this heteroskedasticity. In examining various possible forms of heteroskedasticity it was found that an ESTAR-EGARCH model is the most appropriate in both cases. The ESTAR-EGARCH model used in this study is based on equation (16). The error process is $\varepsilon_t = e_t\sqrt{\sigma_t^2}$, where $e_t \sim$ N.I.D(0,1) and the conditional variance is $\ln\left(\sigma_t^2\right) = \beta_1 + \beta_2\left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] + \beta_3\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta_4\ln\left(\sigma_{t-1}^2\right)$ and is not dependent on e_t.

The estimated ESTAR-EGARCH models perform well in terms of providing goodness of fit, statistically significant coefficients and satisfactory residual diagnostics (see Table 7). The estimated standardised transition parameters, γ, appears to be significantly different from zero on the basis of the individual 't-ratios'. However, previous empirical studies (see for example Kilian, 1999; Kilian and Taylor, 2003) have noted that these 't-ratios' must be carefully interpreted since if under the null hypothesis the transition parameters are equal to zero, then the dividend-price series is generated by a unit root process.

We therefore calculate the empirical marginal significance level of the transition parameters using Monte Carlo simulations assuming that the true data generating process for the dividend-price series is a first-order unit root process, with slope and innovation variance parameters calibrated using the actual value-weighted and equally-weighted dividend-price series respectively (Details relating to the program are contained in Appendix 1). The empirical significance level is based on 5000 simulations of length 1424, from which the first 500 were discarded (leaving

924 data points, corresponding to the size of our data set). At each replication, an ESTAR-EGARCH equation was estimated for each artificial data set, identical in form to those reported in Table 7. The percentage of ‘t-ratios’ for the estimated transition parameters, greater in absolute value then that reported in Table 7 were obtained and taken as the respective empirical significance level. Our estimated transition parameters are significantly different from zero at the 5% level. This result indicates strong evidence of nonlinear mean reversion in the dividend-price ratio.

The next step is model evaluation. Obvious assumptions to be tested following Eitrheim and Teräsvirta (1996) are the null hypotheses of no remaining nonlinearity, no residual autocorrelation, and parameter consistency as we assume that the parameters are constant when we estimate the models.[16] Table 7 also outlines these results. The models capture all the nonlinear features of the data: the NRL statistic reports the p-value for the Lagrange Multiplier test of the null of no remaining nonlinearity. If the null were rejected the models should be re-estimated, that is not the case here. Testing for parameter consistency is also important in this nonlinear framework since the model has been estimated assuming constant parameters. In contrast to the linear case where the alternative to the null of parameter constancy is a single structural break, the statistics ET1, ET2 and ET3[17] in Table 7 test for parameter consistency in ESTAR model under a parametric alternative, which explicitly allows the parameters to change smoothly. According to these test statistics the model is stable.

The speed of adjustment can be visualized from the plots of the estimated transition function, $F[y_{t-d}]$, against the corresponding lagged values of the series, y_{t-d}. Figure 3 shows that the transition functions are mildly explosive near the equilibrium and mean reverting away from the equilibrium level. The inverted-bell shaped plots in Figure 3 show that the adjustment processes of the negative and positive deviations are acceptably symmetrical in nature. This finding is in accordance with most literature (Gallagher and Taylor, 2001).

To gain more insights into how the adjustment of the dividend-price ratio transfers from one regime to the other, we plot the estimated transition functions against time. While the transition appears to be slightly more volatile in the equally

[16] See Appendix 10.1 for details relating to these three tests.

[17] ET1 refers to Smooth Monotonic parameter consistency; ET2 refers to Symmetric Non-Monotonic parameter consistency and ET3 refers to Monotonic and Non-Monotonic parameter consistency.

weighted case, Figure 4(b) shows that both functions reach their uppermost value at times or booms. For example, Figure 4 captures the Great Depression (1929-1932), World War II (1941-1945), the period of phenomenal economic growth after World War II (mid 1940's – mid 1950's), Oil Crisis (1973), serious recession during Ford's Presidency (1974) where unemployment rose to over 12%, rise in oil prises (1980), recession in Regan's Presidency (1982), Stock Market Crash (1987), technology boom (1990's) etc. This provides strong visual justification for the use of nonlinear models, in particular ESTAR models, in estimating and forecasting the stock dividend-price ratio.

To obtain further insights into the dynamic structure of dividend-price ratio, we perform impulse response function analysis to evaluate the propagation mechanism of shocks to the dividend-price ratio. Figure 5 presents the impulse response analysis for our value-weighted and equally-weighted nonlinear models respectively. As pointed out by Taylor and Peel (2000) the impulse response functions in nonlinear models are not independent from the initial conditions, the size of the shock and the future path of the exogenous innovations. As a result, the impulse response functions must be computed by Monte Carlo integration.

In this paper the impulse response functions are calculated as follows: starting with the first eleven observations set to their historical values, we estimate 5000 simulations of length 200 of our model with and without a shock of size 's' at time eleven. Thus, for every simulation, we obtain two realisations of the deviations of the dividend-price ratio from its fundamental. The difference between the two simulated paths, one allowing for a shock 's' and the other without it, are stored and averaged over all 5000 simulations, so that this average is taken as the estimated impulse response function for a size shock 's'. We consider six shocks: 1, 5, 10, 15, 20 and 30 percent. Figure 5 clearly shows that the speed of the adjustment towards the equilibrium increases with the size of the shock, that is, when the dividend-price ratio is driven further away from its fundamental.[18] Thus, shocks of small size appear to be persistent and the dividend-price ratio apparently does not follow its fundamental.

[18] This is discernible from the half-life estimates of the shocks. The half-life of the value-weighted estimates are 1% = 114 months, 10% = 107 months, 20% = 98 months, 30% = 88 months, 40% = 79 months and 50% = 70 months. The half-life of the equally-weighted estimates are 1% = 43 months, 10% = 43 months, 20% = 43 months, 30% = 42 months, 40% = 40 months and 50% = 39 months.

Although the combined evidence from the nonlinear impulse response functions may be difficult to interpret and generalise it is clearly indicative of the presence of nonlinearities in the dynamic structure of the dividend-price ratio. These nonlinearities call into question the results of many studies, which have generated forecasts conditional on the adequacy of a linear dynamic structure for the dividend-price ratio.

Now that we have established that the log of the value-weighted and equally-weighted dividend-price ratios are best represented as ESTAR(1)-EGARCH(1) and ESTAR(2)-EGARCH(1) processes respectively, we can proceed with estimating the long horizon regressions and determining the degree with which we can predict stock returns, if any.

5. STOCK PRICE PREDICTABILITY: EMPIRICAL RESULTS

The in-sample and out-of-sample tests are based on the long-horizon regression approach. This approach entails estimating k individual equations:

$$r_{t+k}^{k} = \alpha_k + \beta_k z_t + e_{t+k}^{k} \qquad k=1, 4, 8, 12, 18, 24, 36, 48 \tag{17}$$

where r_{t+k}^{k} represents either the continuously compounded k-period rate of return or the continuously compounded k-period rate of return minus the risk-free rate of return, z_t is the dividend-price ratio, k is the horizon length and α_k and β_k are the parameters to be estimated. The error term e_{t+k}^{k} is an element of the time t+k information set.

We examine the in-sample and out-of-sample tests using, (i) the value-weighted excess returns series and (ii) the equally-weighted excess returns series. Our in-sample tests are based on the H_0: $\beta_k = 0$ versus H_1: $\beta_k < 0$ for a given horizon for some k, or jointly for all forecast horizons as $H_0{:}\beta_k = 0\ \forall$ k versus $H_1{:}\beta_k <0$ for some k (Mark, 1995), while our out-of-sample tests are based on comparing the forecast accuracy of (17) with that of (a) a random walk model and (b) a random walk model with a drift,[19] using the MSE-T, MSE-F, ENC-T and ENC-F test statistics.

[19] This is essentially a test of predictability as it involves comparing $r_{t+k}^{k} = \alpha_k + \beta_k z_t + e_{t+k}^{k}$ with $r_{t+k}^{k} = \alpha_k + e_{t+k}^{k}$.

Since much of the literature[20] criticises long-horizon regressions for size distortions and low power, we begin by examining the power of our long horizon stock price predictability regressions. In contrast to the stock price predictability literature we generate the critical values for the log-dividend price ratio using an ESTAR-EGARCH data generating process.

5.1. Power and Size Properties of our Long Horizon Regressions

To estimate the size and power properties of our long horizon model we use a Monte Carlo experiment; which generates pseudo stock price and dividend-price data series, estimates a long horizon regressions for each data set and assesses the significance of each of these regressions. We calculate both the power and size results using the parameters estimated from the equally-weighted returns series.[21]

Similar to Kilian and Taylor (2003) our bootstrap approach follows a two-step process. In step 1, we write down the unrestricted reduced form representation of the data. This reduced form, which is compatible with the data, encompasses the restricted model under the null hypothesis and encompasses the unrestricted model under the alternative hypothesis. In step 2, we generate critical values by estimating the process subject to the restrictions under the null hypothesis and simulating the distribution of the test statistics in repeated replications.

We postulate that the unrestricted data generating process may be represented as a bivariate nonlinear model for $(r_t, z_t)'$ such that z_t follows an ESTAR process and r_t is possibly predictable based on historical data.

$$r_t \sim (p_t, d_t, p_{t-1}, d_{t-1}, \ldots) + u_{1t} \tag{18}$$

$$z_t - \mu = \sum_{j=1}^{p} \lambda_j [z_{t-j} - \mu] - \sum_{j=1}^{p} \lambda_j [z_{t-j} - \mu] \Phi\left[\langle z_{t-d} \rangle_{d=1}^{\bar{d}}; \mu; \gamma\right] + u_{2t} \tag{19}$$

Here we can clearly see that the dividend-price variable is always represented as an ESTAR model, z_t. The returns series, r_t, on the other hand, is not as narrowly specified. This model is broad enough to encompass both the random walk behaviour

[20] See for example Mankiw, Romer and Shapiro (1991), Hodrick (1992), Nelson and Kim (1993), and Berkowitz and Giorgianni (2001).

[21] Similar power and size statistics are generated using the value weighted series and therefore are not reported.

of stock returns and more complicated linear or nonlinear serially correlated processes.

Although, the nature of the unrestricted model affects the power of the test, we never have to estimate this fully unrestricted model in practice. Since we are interested in testing the null hypothesis that stock returns are unpredictable based on past information, we can represent the null:

$$H_0 : r_t - \mu = u_{1t} \tag{20}$$

Here, we have restricted the returns series to follow a random walk process, whilst the log dividend-price process, z_t, remains as a nonlinear process.

We examine the marginal significance of each forecast, in step 2, based on a bootstrap procedure with the following data generating process for r_t and z_t:

$$r_t - \mu = u_{1t} \tag{21}$$

$$z_t - \mu_z = \sum_{j=1}^{p} \lambda_j [z_{t-j} - \mu] - \sum_{j=1}^{p} \lambda_j [z_{t-j} - \mu] \Phi\left[\langle z_{t-d} \rangle_{d=1}^{\bar{d}} ; \mu ; \gamma\right] + u_{2t} \tag{22}$$

This ESTAR process is consistent with our dividend price series, whilst the returns series is generated under the null hypothesis that returns follow a random walk process. The innovations $u_t = (u_{1t}, u_{2t})'$ in practice will be treated as independent and identically distributed. Bootstrap p-values for the long horizon regression test statistics under the null hypothesis may be obtained by generating repeated trials from this bootstrap data generating process, and by then re-estimating the long-horizon regression test statistics for each set of bootstrap data and evaluating the empirical distribution of the resulting long-horizon regression test statistic. We repeat this process 500 times to obtain our marginal significance values.[22]

5.1.1. Size Test Results

To compute the size of our tests we conduct Monte Carlo experiments with the data generated under the null hypothesis of no predictability. In other words, the data generating process for the size tests are based on equations (21) and (22). The size of the test is the proportion of the t-statistics at each horizon that have a p-value less than

[22] The ESTAR-EGARCH procedure is shown in Appendix 1.

10%. The results from the in-sample size tests are reported in Figure 6, while the out-of-sample test results are presented in Figure 7. The in-sample results indicate that our bootstrap is remarkably accurate. The effective size of each test is reasonably close to the nominal significance level of 10% and remains fairly constant across forecast horizons. Therefore, we can reasonably conclude that size distortions do not lead to increased long-horizon predictability. Some of the out-of-sample results are slightly less accurate, particularly at long horizons. The MSE-T (i.e. the Diebold Mariano Test) is accurate at short horizons, up to the one-year horizon, after which it tends to under reject the null hypothesis. The ENC-T test is slightly more accurate at short horizons, but under rejects the null hypothesis of no predictability after the two-year horizon. The MSE-F and ENC-F results are more accurate.[23] Therefore we can conclude that while the MSE-T and ENC-T test statistics have good size properties at short horizons, the MSE-F and ENC-F test statistics have good size properties at all horizons.

5.1.2. Power Test Results

To investigate the power of our long horizon regressions, we conduct Monte Carlo simulations based on the best fitting model under the alternative hypothesis of predictability. As pointed out by Kilian and Taylor (2003), the power of the test will in general depend on the specific form of the alternative model. As it is difficult to identify the actual underlying nonlinear process at work in the stock price from the actual data, we focus instead on the easier task of finding a reasonable approximation to the time series process of the fundamental, Δd_t. Given the DGP for z_t, selecting a DGP for Δd_t will pin down the implied DGP for r_t, where r_t is calculated as[24]

$$r_{t+1} = \ln(\exp(dp_t - dp_{t+1}) + \exp(dp_t)) + \Delta d_{t+1} \tag{23}$$

d_t represents the log dividend series and dp_t represents the log dividend-price ratio. Based on preliminary analysis we can conclude that Δd_t follows a linear process.[25] The lag structure for the fundamental is selected using the AIC. This procedure

[23] Harvey, Leybourne and Newbold (1998), West (2001) and Clark and McCracken (2001) come to a similar conclusion.
[24] See Goyal and Welch (2003).
[25] Tests of the form shown in equation (11) indicated no significant evidence of nonlinear behaviour in the dividend growth series.

selected 6 lags for the change in dividends and 2 lags for the price variable. All insignificant lags were then dropped resulting in:

$$\Delta d_t = \gamma_1 \Delta d_{t-1} + \gamma_2 \Delta d_{t-3} + \gamma_3 \Delta d_{t-6} + \gamma_4 \Delta p_{t-2} + u_{1t} \tag{24}$$

The data generating process under predictability is thus given by equations (21) and (24). By randomly re-sampling the residuals and using the estimated equations, we can build up a pseudo-sample of data for r_t (using equation (23)) and z_t which matches the original sample size.[26]

Next we estimate the long-horizon regressions and the t-statistic corresponding to $\hat{\beta}_k$. We calculate the p-value corresponding to each t-statistic using the bootstrap procedure. The power results reported are the proportion of the p-values that are less than 10% for each k. As the actual size is close to the nominal size, there is no need for size corrections.

Our power study shows that the proposed long horizon regressions are highly accurate under the null hypothesis of no return predictability. The result from the in-sample power test is reported in Figure 8, while the results from the out-of-sample tests are reported in Figure 9. The in-sample test results show that the power of the test is very accurate at all horizons. The power of the test only falls slightly at the four-year horizon.

As expected the power of the ENC-F test is greater than the power of the standard MSE-T test.[27, 28] Similar to Clark and McCracken (2001), we find that the power of MSE-T< ENC-T< MSE-F< ENC-F. We note that the power for the out-of-sample test, Figures 9, is lower than that in the in-sample tests, Figure 8.

5.2. Stock Price Predictability in the United States, 1926:01-2003:12

Now that we have established that our ESTAR-EGARCH models have appropriate size and power levels we continue by examining whether the dividend-price ratio has predictive power in the United States from 1926:01 to 2003:12. In particular, using value-weighted and equally-weighted data respectively, we examine whether the log dividend-price ratio can predict one-step ahead excess stock return forecasts.

[26] As in previous estimations our simulations here are estimated using 1424 observations from which the first 500 are discarded, leaving 924 observations corresponding to the sample size.
[27] Previous empirical research has traditionally examined the MSE-T statistic only.
[28] See, Harvey, Leybourne and Newbold (1998), West (2001) and Clark and McCracken (2001).

We begin by examining the results from the in-sample t-tests. If our model of stock price determination is correct then we should expect to see a clear pattern of increased long-horizon predictability in the form of an increased adjusted r-squared measure and p-values that fall as the horizon grows. This is indeed what we find. Table 8 shows the adjusted r-squared measure for each series. Similar to Hodrick (1992) we find that the $\overline{R}^2$ adjusted never exceeds 20%. As expected the $\overline{R}^2$ becomes stronger as one increases the horizon, k.

Figure 10 shows the bootstrap p-values from our long-horizon regression tests. We examine the ability of the log-dividend price ratio to predict four different returns series. These returns series are: (1) the excess value-weighted returns series and (2) the excess equally-weighted returns series In each case separate results are shown for horizons of k=1, 4, 8, 12, 18, 24, 36 and 48 months. The marginal significant levels for the value-weighted series are generated using an ESTAR(1)-EGARCH(1) data generating process, while the marginal significance levels for the equally-weighted series are generated using an ESTAR(2)-EGARCH(1) data generating process. As expected, the p-values fall as stock prices become more predictable at longer horizons. Figure 10 also shows the results for a joint test[29] for all forecast horizons as H_0: $\beta_k = 0 \ \forall$ k versus H_1: $\beta_k < 0$ for some k. The joint test is significant at the 5% level for each of the four returns series.

Next we examine the out-of-sample forecasts. Again, we look at the ability of the log dividend-price series to predict our four returns series.[30] The out-of-sample tests are based on a sequence of rolling forecasts and involve comparing the null hypothesis of equal forecast accuracy against the one-sided alternative that forecasts from the long horizon regressions are more accurate than random walk forecasts. Comparison of the forecasts from the long horizon model with those from benchmark random walk models[31] enables us to determine the ability of the log dividend-price ratio to forecast movements in the stock price. To carry out these forecasts, we divide the total sample of T observations into an in-sample and an out-of-sample portion, where the in-sample portion spans the first 697 observations and the out-of-sample portion the last 227 observations for stock returns. In other words we consider one-step-ahead forecasts over the period 1985:01 – 2003:12.

[29] See Mark (1995) for details regarding this joint test statistic.

[30] These return series are (i) the value-weighted excess returns series and (ii) the equally-weighted excess returns series.

[31] We examine both a pure random walk model and a random walk with a drift model.

We will begin by examining whether stock returns based on the log-dividend price ratio out-predict those based on a pure random walk model.[32] As the asymptotic critical values from our out-of sample test statistics are severely biased in small samples, we generate the marginal significance level using a bootstrap approximation of the finite sample distribution of the test statistic under the null hypothesis of no predictability (see McCracken, 2000). Unlike asymptotic critical values, bootstrap critical values based on the percentiles of the bootstrap distribution automatically adjust for the increase in the dispersion of the finite-sample distribution of the test statistic that occurs in near-spurious regressions as the sample size grows (Kilian, 1999).

5.2.1. Dependent Variable: Excess Returns. Alternative Model: Pure Random Walk

Figures 11 and 12 show the bootstrap p-value for the MSE-T, ENC-T, MSE-F and ENC-T test statistics for the equally-weighted returns series and value-weighted excess returns series respectively. The benchmark model in this case is the pure random walk model. We consider the present value model to have superior predictive ability if the marginal significant level is less than 10%. If, on the other hand, the marginal significance level is greater than 10% then the benchmark model has superior power.

The MSE-T statistic tells us that the dividend-price ratio is superior at the 48-month horizon; however this test fails the joint test. The MSE-F test, on the other hand, concludes that the log dividend-price series has greater predictive ability at the 12, 18, 36 and 48 month horizon. The ENC-T test finds the log dividend-price ratio is only superior at the 18-month horizon, and the ENC-F test concludes that the log dividend-price series is superior at the 18, 24 and 36 month horizons. Bearing in mind that the size and the power properties of the ENC-F are more accurate than those for the remaining tests, we conclude that the log dividend-price ratio has the ability to beat the pure random walk model at medium horizons (i.e. from one and a half years to three years).

Figure 12 presents the bootstrap p-values relating to the ability of the dividend-price ratio to forecast one-step-ahead changes in the value-weighted excess returns series. Here we are unable to identify any level of predictability. It is likely

[32] The pure random walk model states that $P_t = P_{t-1}$.

that the finding of a unit root in log dividend-price ratio (see Table 3) is driving this result. .

5.2.2. Dependent Variable: Excess Returns. Alternative Model: Random Walk with a Drift

Next, we consider an alternative benchmark model. In this scenario we compare the forecasts from our long horizon regressions, estimated as $r^k_{t+k} = \alpha_k + \beta_k z_t + e^k_{t+k}$ with those from a random walk model with a drift, estimated as $r^k_{t+k} = \alpha_k + e^k_{t+k}$. Figure 13 presents the results generated using the equally-weighted returns series and Figure 14 presents the results generated using the value-weighted returns series.

As before, the one-step-ahead forecasts based on the equally weighted returns series indicate some degree of predictability, while those based on the value weighted series have no out-of-sample predictive ability. Therefore, we will only examine the results from the equally-weighted series.

The out-of sample forecasts generated using equally-weighted excess returns are reported in Figure 17. As the ENC-F test statistic has the most accurate size and power characteristics we will only examine this statistic here. This statistic finds that the log-dividend price series produces superior forecasts at the 4, 8, 12, 18, 24, 36 and 48-month horizon. The joint test is also significant. These results are similar to those calculated using the returns series, therefore in this case adjusting for the risk-free interest rate does not reduce the predictive ability of the log dividend-price ratio.

5.2.3. Summary

We conclude that the in-sample and out-of-sample results for the equally-weighted series are in agreement. This finding is at odds with recent studies (see Goyal and Welch, 2003;[33] Lewellen and Shanken, 2001, and Bossaerts and Hillion, 2001[34]) and

[33] Goyal and Welch (2003) suggest that the lack of out-of-sample predictability may be a consequence of learning in the marketplace. That is, the best in-sample investment strategies may not persist into subsequent periods because the market adjusts to the new information.

[34] On the other hand, Lewellen and Shanken (2001) and Bossaerts and Hillion (2001) argue that the Bayesian learning of economic agents can generate *ex-post* predictable patterns that are *ex-ante* rational and therefore not real-time tradable opportunities. In this case, predictability is just an *ex-post* illusion. For example, suppose you know that the time-series of stock returns is mean-reverting. In real time, you still do not know if stock prices will be higher or lower next period because you do not know the true mean of the distribution. Nonetheless, a pattern of mean reversion is easily detected ex post relative to the sample mean.

with the findings from the value-weighted return series. It is likely that the results using the equally-weighted series are driven by small firms while those of the value-weighted series are driven by larger firms (see DeFusco, Geppert, Zorn, 2005). This indicates that the success of the equally-weighed series may be due to size of the firm.

6. CONCLUSION

The results from tests of the traditional present value model produce mixed results. Moreover, recent studies on the effects of transaction costs and limits to arbitrage suggest that the cointegrating relationship between real stock prices and dividends should be approximated by a nonlinear adjustment. Since previous empirical studies modelled the cointegrating relationship between stock prices and dividends as a linear process, they were biased in favour of finding predictability and were subject to severe size distortions. Thus their results are invalid.

The purpose of this paper is to contribute to the debate on the relevance of nonlinear models in financial markets. To that end, we employed nonlinear unit root tests, cointegration tests and ESTAR models to show that the reason for poor predictive performance of stock price models is due to nonlinearity in the adjustment of the dividend-price ratio to its long run equilibrium path. We examine both value-weighted and equally-weighted stock price data.

The evidence presented reveals that the estimated cointegrating residual of the present value model is approximated well by an ESTAR(1)-EGARCH(1) model using value-weighted data and by an ESTAR(2)-EGARCH(1) model using equally-weighted data. In other words, the error correction towards the cointegrating equilibrium implied by the present value model is nonlinear. The parameters of the nonlinear models and the generalised impulse response functions imply random walk behaviour for small deviations and fast mean-reverting adjustment for large deviations from equilibrium. This finding is consistent with features of the stock market, such as transaction costs and limits to arbitrage and noise trader activity.

Taken together, the evidence presented in this paper confirms the results of recent studies that emphasise the importance of allowing for nonlinearity in the adjustment of the dividend-price ratio towards its long-run equilibrium path. Hence, we offer a potential reason why stock price models may have failed to out-perform the forecasting performance of the random walk model in the past.

Allowing for the exponentially smooth autoregressive behaviour of the dividend-price ratio allows us to account for nonlinearities which result from limits to arbitrage, such as the existence of transaction costs, short-selling constraints, or mispricing of securities deepening in the short run. Our long horizon regression tests designed to detect nonlinear long-horizon predictability provided strong in-sample evidence against the random walk model. For example, our in-sample long-horizon tests were superior to the random walk model at the 12, 18, 24, 32 and 48-month horizons at the 5% level of significance.

Our out-of-sample results are mixed. While we find that the present value model has an ability to predict the equally-weighted returns series, particularly at medium horizons; we find that it has no ability to predict the value-weighed returns series at any horizon. Previous studies concur with our finding for the value-weighted returns series. For example, Bossaerts and Hillion (1999) and Goyal and Welch(2003) document substantial in-sample predictability but find no evidence of out-of-sample predictability. Similarly, while Lo and MacKinlay (1997) find some evidence of market timing profits from 1967 to 1993, Pesaran and Timmermann (1995) conclude that a real-time investor could have profited only during the 1970's from timing the stock market, not in the 1960's or the 1980's.

Our out-of-sample tests demonstrate why it is difficult to predict stock prices based on the dividend-price ratio in real time. In practice, only very large deviations from the dividend-price ratio will reveal a stock returns inherent tendency to mean revert. The plots of the transition functions over time highlight that such large deviations are rare, particularly in the value-weighted return series. The ability of the present value model to predict the equally-weighted returns series may be related to the fact that deviations from equilibrium for this series has a tendency to mean revert quite quickly, as shown in the impulse response function.

REFERENCES

Ang, A., and G. Bekaert. (2001) "Stock Return Predictability: Is It there?" *Working Paper No 8207, National Bureau of Economic Research*.

Aslanidis, N. (2002) "Smooth Transition Regression Models in UK Stock Market Returns." *Working paper,* Department of Economics, University of Crete.

Bachelier, L. (1900) "Théorie de la Spéculation." *Annales de l'Ecole normale superiure*.

Balke, N.S., and M.E. Wohar. (2002) "Low-Frequency Movements in Stock Prices: A State-Space Decomposition." *The Review of Economics and Statistics*, 84(4), 649-667.

Bauer, G. (1995) "The Foreign Exchange Risk Premium Over the Long Run." Manuscript, Department of Finance, University of Pennsylvania.

Berben, R.P., and D. Van Dijk. (1999) "Unit Roots and Asymmetric Adjustment – a Reassessment." Econometric Institute Report 101, Erasmus University Rotterdam, Econometric Institute.

Berkowitz, J., and L. Giorgianni. (2001) "Long-Horizon Exchange Rate Predictability?" *Review of Economics and Statistics*, 83, 1, 81-91.

Blanchard, O.J., and M.W. Watson. (1982) "Bubbles, Rational Expectations and Financial Markets." In *Crises in the Economic and Financial Structure: Bubbles, Bursts, and Shocks* (ed ., P. Wachtel). Lexington, MA: Lexington Books, 295-315.

Bohl, M.T., and P.L. Siklos. (2004) "The Present Value Model of US Stock Prices Redux: a New Testing Strategy and Some Evidence." *The Quarterly Review of Economics and Fina*nce, 44, 208-223.

Bossaerts, P. and P. Hillion. (2001) "IPO Post-Issue Markets: Questionable Predictions But Diligent Learners?" *The Review of Economics and Statistics*, 83(2), 333-347.

Campbell, J.Y. (1991) "A Variance Decomposition for Stock Returns." *Economic Journal,* 101(127), 157-79.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. (1997) *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Campbell, J.Y., and R.J. Shiller. (1987) "Cointegration and Tests of Present Value Models." *Journal of Political Economy*, 95(5), 1062-1088.

Campbell, J.Y., and R.J. Shiller. (1988a) "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance*, 43(3), 661-76.

Campbell, J.Y., and R.J. Shiller. (1988b) "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, 1(3), 195-228.

Campbell, J.Y., and R.J. Shiller. (1998) "Valuation Ratios and the Long-Run Stock Market Outlook." *Journal of Portfolio Management*, Winter, 11-26.

Campbell, J.Y., and R.J. Shiller. (2001) "Valuation Ratios and the Long-Run Stock Market Outlook: An Update." Cowles Foundation Discussion Paper No 1295, Yale University.

Chinn, M.D., and R.A. Meese. (1995) "Banking on Currency Forecasts: How Predictable is Change in Money?" *Journal of International Economics*, no 38, 53-69.

Clark, T.E and M.W. McCracken. (2001) "Tests of Equal Forecast Accuracy and Forecast Encompassing Models." *Journal of Econometrics*, 105, 85-110.

Clark, T.E and M.W. McCracken. (2004) "Evaluating Long-Horizon Forecasts." *Manuscript Federal Reserve Bank of Kansas.*

Coakley, J., and A-M. Fuertes. (2004) "Momentum and Reversal Dynamics in US Valuation Ratios." *Financial Econometrics Research Centre, Working Paper, forthcoming.*

Cochrane, J. H. (1991) "Volatility Tests and Efficient Markets: A Review Essay." *Journal of Monetary Economics,* 27, 463-485.

Cochrane, J.H. (1992) "Explaining the Variance of Price-Dividend Ratios." *Review of Financial Studies*, 5(2), 243-80.

Cuthbertson, K., and D. Nitzsche. (2004) Quantitative Financial Economics - *Stocks, Bonds and Foreign Exchange.* Second Edition. John Wiley and Sons Ltd.

Daniel, K. (2001) "The Power and Size of Mean Reversion Tests." *Journal of Empirical Finance*, 8(5), 493-535.

DeFusco, R.A., J.M. Geppert and T.S. Zorn. (2005) "Is There a Presidential Puzzle?" FMA 2005 Annual Meeting.

De Long, J.B., A. Shleifer, L.H. Summers, R.J. Waldmann. (1990) "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98, 703-38.

Dickey, D.A., and W.A. Fuller. (1979) "Distribution of the Estimates for Autoregressive Time Series with a Unit Root." *Journal of American Statistical Association,* 74, 427-31.

Dickey, D.A., and W.A. Fuller. (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica*, 49, 1057-72.

Diebold, F.X., and R. Mariano. (1995) "Comparing Predictive Accuracy." *Journal of Business and Economics Statistics*, 13, 253-62.

Eitrheim, Ø., and T. Teräsvirta. (1996) "Testing the Adequacy of Smooth Transition Autoregressive Models." *Journal of Econometrics*, 74(1), 59-75.

Enders, W. and C.W.J. Granger. (1998) "Unit Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates." *Journal of Business and Economic Statistics*, 16, 304-311.

Evans, G.W. (1991) "Pitfalls in Testing for Explosive Bubbles in Asset Prices." *American Economic Review*, September, 81(4), 922-30.

Fama, E.F. (1970) "Multi-period Consumption-Investment Decisions." *American Economic Review*, 60, 163-174.

Fama, E.F., and K.R. French. (1988a) "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy*, 96, 246- 273.

Fama, E.F., and K.R. French. (1988b) "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22, 3-25.

Fama, E.F., and G.W. Schwert. (1977) "Asset Returns and Inflation." *Journal of Financial Economics,* November, 5, 115-46.

Ferson, W.E. (1989) "Changes in Expected Security Returns." *Journal of Finance*, 44, 1191-1217.

Flavin, M. (1983) "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence." *Journal of Political Economy,* 91, 929-956.

Froot, K.A., and M. Obstfeld. (1991) "Intrinsic Bubbles: The Case of Stock Prices." *American Economic Review*, December, 81(5), 1189-214.

Gallagher, L.A., and M.P. Taylor. (2001) "Risky Arbitrage, Limits of Arbitrage, and Nonlinear Adjustment in the Dividend-Price Ratio. *Economic Inquiry*, 39(4), 524-36.

Glosten, L.R., R. Jagannathan, and D. Runkle. (1993) "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance,* 48, 1779-1801.

Goetzmann, W.N., and P. Jorion. (1995) "A Longer Look at Dividend Yields." *Journal of Business*, 68(4), 483-508.

Goyal, A. and I. Welch. (2003) "Predicting the Equity Premium with Dividend Ratios." *Management Science*, 49(5), 639-654.

Granger, C.J., and T. Teräsvirta. (1993) *Modelling Nonlinear Economic Relationships.* Oxford, U.K.: Oxford University Press.

Harvey, D.I., S.J. Leybourne and P. Newbold. (1998) "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics*, 16(2), 254-59.

Hodrick, R.J. (1992) "Dividend Yields and Expected Stock returns: Alternative Procedures for Inference and Measurement." *The Review of Financial Studies*, 5(3), 357-86.

Horvath, M.T.K and M.W. Watson. (1995) "Testing for Cointegration When Some of the Cointegrating Vectors are Prespecified." *Econometric Theory*, 11, 984-1014.

Kanas, A. (2005) "Nonlinearity in the Stock Price-Dividend Relation." *Journal of International Money and Finance*, forthcoming, 2005.

Kapetanois, G., Y. Shin and A. Snell. (2003) "Testing for a Unit Root in the Nonlinear STAR Framework." *Journal of Econometrics*, 112, 359-379.

Kapetanois, G., Y. Shin and A. Snell. (2004) "Testing for Cointegration in Nonlinear STAR Error Correction Models." *Working Paper No. 497*, Department of Economics, Queen Mary University of London.

Kendall, M.G. (1953) "The Analysis of Economic Time-Series – Part 1: Prices." *Journal of Royal Statistical Society*, 96, 11-25.

Kilian, L. (1995) "Small-Sample Confidence Intervals for Impulse Response Function." *The Review of Economics and Statistics*, 218-30.

Kilian, L. (1999) "Exchange Rates and Monetary Fundamentals: What do We Learn from Long-Horizon Regressions?" *Journal of Applied Econometrics*, 14, 491-510.

Kilian, L., and M.P. Taylor. (2003) "Why is it so Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, 60, 85-107.

Kwiatkowski, D., P.C. Phillips, P. Schmidt and Y. Shin. (1992) "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root." *Journal of Econometrics,* 54(1-3), 159-178.

Lamont, O. (1998) "Earnings and Expected Returns." *Journal of Finance*, 53, 1563-87.

Lewellen, J. and J. Shanken. (2001) "Learning, Asset-Pricing Tests, and Market Efficiency." *Journal of Finance*, 57, 1113-1145.

LeRoy, S., and R. Porter. (1981) "The Present Value Relation: Tests Based on Variance Bounds." *Econometrica* 49, 555-574.

Luukonen, R., P. Saikkonen and T. Teräsvirta. (1988a) "Testing Linearity in Univariate Time Series Models." *Scandinavian Journal of Statistics*, 15, 161-75.

Luukonen, R., P. Saikkonen and T. Teräsvirta. (1988b) "Testing Linearity Against Smooth Transition Autoregressive Models." *Biometrika*, 75, no 3, 491-9.

Mandelbrot, B.B. (1966) "Forecasts of Future Prices, Unbiased Markets and Martingale Models," *The Journal of Business*, January, (Special Supplement), 39, 242 - 255. Reprinted as Chapter E19 “Nonlinear Forecasts, Rational Bubbles, and Martingales,” with annotations and corrections, in *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer Verlag, New York, 1997, 471 - 491.

Mankiw, N., D. Romer and M.D. Shapiro. (1991) “Stock Market Forecastability and Volatility: A Statistical Appraisal.” *Review of Economic Studies*, 5, 455-77.

Mark, N.C. (1995) “Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability.” *American Economic Review*, 85, 201-18.

Mark, N.C., and D. Sul. (2002) “Nominal Exchange Rate and Monetary Fundamentals: Evidence From a Small Bretton Woods Panel.” *Journal of International Economics*, 53, 29-52.

Marsh, T.A., and R.C. Merton. (1986) “Dividend variability and variance bounds tests for the rationality of stock market prices.” *American Economic Review,* 76, 483-498.

McCracken, M.W. (2000) “Robust Out of Sample Inference.” *Journal of Econometrics,* 99, 195-223.

McQueen, G. and S. Thorley. (1991) “Are Stock Returns Predictable? A Test Using Markov Chains.” *Journal of Finance*, 46(1), 239-263.

Michael, P., A.R. Nobay and D.A. Peel. (1997) “Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation.” *Journal of Political Economy*, 104, 862-879.

Nelson, C.R., and M.J. Kim. (1993) “Predictable Stock Returns: Evidence from Nonparametric and Threshold Models.” *International Review of Economics and Finance,* 10, 353-68.

Newey, W., and K. West. (1987) “A Simple Positive-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55, 703-708.

Pesaran, M.H., and A. Timmermann. (1995) “Predictability of Stock Returns: Robustness and Economic Significance.” *The Journal of Finance*, 1(4), 1201-1227.

Perron, P. (1988) “Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach.” *Journal of Economic Dynamics and Control*, 12, 297-332.

Phelps, E.S. (1994) *Structural Slumps: The Modern Equilibrium Theory of Unemployment, Interest and Assets.* Cambridge Mass: Harvard University Press.

Phelps, E.S., and G. Zoega. (2001) “Structural Booms.” *Economic Policy*, 32, 83-126.

Potterba, J.M., and l.H. Summers. (1988) “Mean Reversion in Stock Prices: Evidence and Implications.” *Journal of Financial Economics*, 22, 27-59.

Rapach, D.E., and M.E. Wohar. (2005) “Valuation Ratios and Long-Horizon Stock Price Predictability.” *Journal of Applied Econometrics*, 20(3), 327-344.

Samuelson, P.A. (1965) “Proof that Properly Anticipated Prices Fluctuate Randomly.” *Industrial Management Review*, Spring.

Shleifer, A. (2000) *Inefficient Markets: An Introduction to Behavioural Finance*. Oxford University Press, Oxford.

Shiller, R.J. (1981) “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review*, 71, 421-435.

Shiller, R.J. (1989) “Comovements in Stock Prices and Comovements in Dividends.” *Journal of Finance*, 44: 719–729.

Stambaugh, R.F. (1999) "Predictive Regressions." *Journal of Financial Economics*, 54, 375-421.

Taylor, M.P., and D.A. Peel. (2000) "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals." *Journal of International Money and Finance*, 19, 33-53.

Taylor, M.P., D.A. Peel, and L. Sarno. (2001) "Nonlinear Adjustment in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles." *International Economic Review*, 42, 1015-42.

Teräsvirta, T., and H.M. Anderson. (1992) "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models." *Journal of Applied Econometrics*, 7, 119-36.

Teräsvirta, T. (1994) "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models." *Journal of the American Statistical Association*, March, 89(425), 208-18.

Timmermann, A. (1995) "Cointegration Tests of Present Value Models with a Time-Varying Discount Factor." *Journal of Applied Econometrics*, 10, 17-31.

Tsay, R.S. (1989) "Testing and Modelling Threshold Autoregressive Processes." *Journal of the American Statistical Association*, March, 84(405), 231-40.

Valkanov, R. (2001) *"Long Horizon Regressions: Theoretical Results and Applications*." Working paper, University of California, Los Angeles.

Van Dijk, D., T. Teräsvirta and P.H. Franses. (2001) "Smooth Transition Autoregressive Models – A Survey of Recent Developments." Unpublished Manuscript, Erasmus University Rotterdam.

West, K.D. (1988) "Dividend Innovations and Stock Price Volatility." *Econometrica*, January, 56(1), 37-61.

West, K.D. (1996) "Asymptotic Inference about Predictive Ability." *Econometrica*, 64, 1067-84.

West, K.D. (2001) "Tests for Forecast Encompassing When Tests Depend on Estimated Regression Parameters." *Journal of Business and Economic Statistics*, 19, 29-33.

Table 1: Summary Statistics

United States – Value Weighted									
	Mean	*Std Dev*	*Skew*	*Kurt*	*JB*	*ρ(1)*	*ρ(2)*	*ρ(3)*	*ρ(4)*
$\mathbf{p_t}$	0.718	0.647	0.177 (0.02)	-0.575 (0.00)	17.58 (0.00)	0.996*	0.992*	0.987*	0.984*
$\mathbf{d_t}$	-2.530	0.357	-0.587 (0.00)	-0.657 (0.00)	69.84 (0.00)	0.998*	0.997*	0.995*	0.993*
$\mathbf{\Delta p_t}$	0.002	0.054	-0.435 (0.00)	6.834 (0.00)	1825.86 (0.00)	0.097	-0.016	-0.102*	0.008
$\mathbf{\Delta d_t}$	0.000	0.015	-0.085 (0.29)	16.279 (0.00)	10193.5 (0.00)	0.204*	0.191*	0.439*	0.190*
$\mathbf{R_t}$	0.005	0.054	-0.415 (0.00)	6.802 (0.00)	1808.02 (0.00)	0.097	-0.017	-0.107*	0.008
$\mathbf{ER_t}$	0.004	0.054	-0.412 (0.00)	6.776 (0.00)	1794.00 (0.00)	0.098*	-0.015	-0.105*	0.010
$\mathbf{d_{t-1}-p_t}$	-3.249	0.376	-0.351 (0.00)	0.224 (0.16)	20.96 (0.00)	0.988*	0.974*	0.960*	0.948*
$\mathbf{\Delta(d_{t-1}-p_t)}$	-0.001	0.056	-0.005 (0.94)	6.750 (0.00)	1752.44 (0.00)	0.111*	-0.004	-0.082	0.027

United States – Equally Weighted									
	Mean	*Std Dev*	*Skew*	*Kurt*	*JB*	*ρ(1)*	*ρ(2)*	*ρ(3)*	*ρ(4)*
$\mathbf{p_t}$	2.055	1.165	-0.382 (0.00)	-0.721 (0.00)	42.55 (0.00)	0.998*	0.995*	0.993*	0.991*
$\mathbf{d_t}$	-1.240	1.100	-0.741 (0.00)	-0.227 (0.15)	86.74 (0.00)	0.999*	0.999*	0.998*	0.997*
$\mathbf{\Delta p_t}$	0.004	0.070	0.453 (0.00)	9.855 (0.00)	3766.83 (0.00)	0.159*	0.002	-0.098	-0.057
$\mathbf{\Delta d_t}$	0.003	0.020	0.865 (0.00)	18.208 (0.00)	12866.1 (0.00)	0.609*	0.524*	0.635*	0.455*
$\mathbf{R_t}$	0.007	0.069	0.450 (0.00)	9.849 (0.00)	3766.00 (0.00)	0.160*	0.002	-0.103*	-0.057
$\mathbf{ER_t}$	0.006	0.069	0.453 (0.00)	9.826 (0.00)	3749.34 (0.00)	0.162*	0.004	-0.101*	-0.056
$\mathbf{d_{t-1}-p_t}$	-3.299	0.333	-0.118 (0.14)	0.478 (0.00)	10.95 (0.00)	0.975*	0.940*	0.904*	0.871*
$\mathbf{\Delta(d_{t-1}-p_t)}$	0.000	0.074	-0.972 (0.00)	11.656 (0.00)	5371.02 (0.00)	0.199*	0.022	-0.058	-0.038

Note: p_t is the log of real stock prices series, d_t is the log of the annualised real dividend series, R_t is the returns series, ER_t is the excess returns series and $\Delta = (1\text{-}L)$ denotes the first difference. Skew, Kurt and JB denotes the standard skewness, kurtosis and Jarque-Bera statistics as reported in Kendall and Stuart (1958). $\rho(k)$ is the autocorrelation between x_t and x_{t-k}. The sample period is 1927:01 – 2003:12.

Table 2: P-Values for the linearity Tests

(a) Value -Weighted

	H_L	H_{01}	H_{02}	H_{03}
D=1	0.5416	0.7902	0.1831	0.7078
D=2	0.0687	0.0504	0.0297	0.2774
D=3	0.2294	0.1206	0.0599	0.3972
D=4	0.4546	0.9514	0.3661	0.4052
D=5	0.2869	0.4337	0.3094	0.0862
D=6	0.2967	0.7957	0.0730	0.3066
D=7	0.1689	0.2967	0.0321	0.5251
D=8	0.2143	0.9103	0.0931	0.2257

(b) Equally-Weighted

	H_L	H_{01}	H_{02}	H_{03}
D=1	0.0105	0.7408	0.0187	0.9539
D=2	0.0001	0.0238	0.1394	0.0065
D=3	0.0000	0.1361	0.0578	0.2079
D=4	0.3417	0.7926	0.1108	0.7604
D=5	0.2768	0.2782	0.0991	0.5193
D=6	0.0304	0.0426	0.0157	0.3069
D=7	0.0001	0.0004	0.0152	0.0356
D=8	0.0499	0.0150	0.0250	0.0541

Note: Values are estimated including the robusterrors option in RATS.

Table 3: Linear Unit Root Test Results

	DF	***PP***	***KPSS$_\mu$***	***KPSS$_\tau$***	***DF***	***PP***	***KPSS$_\mu$***	***KPSS$_\tau$***
	VW	***VW***	***VW***	***VW***	***EW***	***EW***	***EW***	***EW***
p_t	0.785	-0.844	13.44	0.87	-0.95	-1.03	17.10	1.13
Δp_t	-12.10	-27.44	0.07	0.04	-8.80	-25.64	0.02	0.01
d_t	-1.15	-1.19	15.48	0.83	-0.99	-1.16	16.67	2.16
Δd_t	-7.41	-29.65	0.07	0.06	-7.98	-18.35	0.12	0.06
Returns	-12.15	-27.45	0.04	0.04	-8.76	-25.61	0.02	0.02
Excess Returns	-12.08	-27.40	0.05	0.04	-8.71	-25.58	0.03	0.02
$(d_{t-1} - p_t)$	-2.34	-2.17	7.04	0.94	-4.83	-4.04	1.60	0.62
$\Delta(d_{t-1} - p_t)$	-11.41	-27.07	0.04	0.02	-9.29	-24.77	0.02	0.01

Note: VW refers to the value weighted series and EW refers to the equally weighted series. p_t is the log of the real stock prices series, d_t is the log of the annualised real dividend series as calculated in Hodrick (1992), $\Delta = (1\text{-}L)$ denotes the first difference. The unit root tests are the Augmented Dickey Fuller (ADF), the Phillips-Perron Z_t (PP) and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test statistics (see Dickey and Fuller, 1979, 1981, Kwiatkowski, Phillips, Schmidt and Shin, 1992, Perron, 1988). The null hypothesis for the ADF and the PP test statistics are that the series are I(1). These statistics are estimated with a constant and without a trend and the number of lags is selected using the Bayesian Information Criterion. The critical values for the DF and the PP test statistics are –2.587 at the 10% level, -2.86 at the 5% level and -3.43 at the 1% level. The null hypothesis for the Kwiatkowski, Phillips, Schmidt and Shin test is that the series is stationary. The number of lags in these tests is set at four. This statistic is estimated with a constant, KPSSμ and with a constant and a trend, KPSSι. The 1% critical level for the KPSSμ unit root test is 0.739; the 5% critical level is 0.463 and the 10% level is 0.347. The 1% critical level for the KPSSι unit root test is 0.216; the 5% critical level is 0.146 and the 10% level is 0.119.

Table 4: Nonlinear Unit Root Test Results

	t_{KSS1} VW	t_{KSS2} VW	t_{KSS1} EW	t_{KSS2} EW
AR	-2.58	-3.02	-4.07	-5.19
AR-ARCH	-5.33	-6.44	-3.98	-4.69

Note: VW refers to the value weighted series and EW refers to the equally weighted series. The number of lags is selected using the Akaike Information Criterion. t_{KSS1} refers to a t-test of the H_0: $\delta=0$ in $\Delta y_t = \delta y_{t-1}^3 + \text{error}$, whereas t_{KSS2} refers to a t-test H_0: $\delta=0$ in $\beta\,\Delta y_t = \sum_{j=1}^{p} \rho_j \Delta y_{t-j} + \delta y_{t-1}^3 + \text{error}$. In the AR-ARCH model heteroskedasticity was explicitly modelled using an ARCH(1) model. The 10%, 5% and 1% critical values for KSS test statistics are, respectively -2.55, -2.88 and -3.48.

Table 5: Cointegration Tests

(a) Value Weighted Series

Value Weighted

$$d_{t-1} = \underset{\substack{(0.008)\\ [339.1]}}{-2.87} + \underset{\substack{(0.008)\\ [55.16]}}{0.483p_t} \qquad R^2 = 0.77 \qquad DW = 0.04$$

Linear Engle Granger	Nonlinear Engle Granger	Linear ECM	Nonlinear ECM
-2.70	-3.21	-5.23	-4.62

(b) Equally Weighted Series

Equally Weighted

$$d_{t-1} = \underset{\substack{(0.021)\\ [147.1]}}{-3.10} + \underset{\substack{(0.008)\\ [101.41]}}{0.905p_t} \qquad R^2 = 0.91 \qquad DW = 0.04$$

Linear Engle Granger	Nonlinear Engle Granger	Linear ECM	Nonlinear ECM
-3.25	-3.56	-10.57	-11.55

Note: p_t and d_t are the logs of real stock price and dividend series respectively. All tests are carried out under the null hypothesis of no cointegration. All the tests are based on demeaned series. The critical values for the linear Engle Granger tests are –3.02 at the 10% level, -3.37 at the 5% level and 4.00 at the 1% level of significance. The critical values for the linear Error Correction Model (ECM) tests are –1.28 at the 10% level, -1.64 at the 5% level and -2.32 at the 1% level of significance The critical values for the nonlinear Engle Granger tests are -2.98 at the 10% level, -3.28 at the 5% level and -3.84 at the 1% level. The critical values for the nonlinear Error Correction Model (ECM) tests are -2.92 at the 10% level, -3.22 at the 5% level and -3.78 at the 1% level. The number of lags selected for each test was selected using the Akaike Information Criteria (AIC).

Table 6: ESTAR Model

(a) Value Weighted Series

$$y_t = 0.19 + [1.02*(y_{t-1} + 0.19)] *[\exp\{-0.064(y_{t-2} + 0.19)^2\}] + \hat{u}_t$$

(-3.23) (134.8) (-3.23) (-3.85) (-3.23)

[0.001]

R^2	= 0.97	SEE	= 0.056
DW	= 1.81	VR	= 0.973
ARCH (1)	= 182.4 (0.000)	LR(1)	= 24.99 (0.000)

(b) Equally Weighted Series

$$y_t = (1.19y_{t-1} + (1- 1.19)y_{t-2}) * [\exp\{-0.075(y_{t-3})^2\}] + \hat{u}_t$$

(36.77) (36.77) (-4.76)

[0.000]

R^2	= 0.95	SEE	= 0.072
DW	= 1.99	VR	= 0.974
ARCH (1)	= 278.4 (0.000)	LR(1)	= 23.80 (0.000)

Note: The figures in parenthesis are t-ratios and the figures in the square brackets are marginal probability values. The marginal probability value is calculated from 5000 Monte Carlo experiments – see text for details. R^2 is the proportion of the variation in y_t explained by the model, s is the standard error of the estimate. DW is the Durbin-Watson statistic, ARCH(n) is a tests for autoregressive conditional heteroscedastity with n lags of the squared residual, LR(1) is a log-likelihood test against an AR(1) model, VR is the variation ratio.

Table 7: ESTAR-EGARCH Model

(a) Value Weighted Series

$$y_t = 0.18 + [1.01*(y_{t-1} + 0.18)] * [\exp\{-0.04(y_{t-2} + 0.18)^2\}] + \hat{u}_t$$

(2.55) (139.7) (2.55) (2.46) (2.55)
[0.030]

$$\ln\left(\sigma_t^2\right) = -0.143 + 0.186\left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] - 0.337\frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + 0.975\ln\left(\sigma_{t-1}^2\right)$$

(-3.96) (7.88) (-3.46) (158.9)

R^2 =	0.97	VR =	0.972
LR(1) =	25.30 (0.000)	H_0: No Remaining Nonlinearity	0.002 (0.999)
H_0: No Error Autocorrelation:		H_0: Parameter Consistency:	
No of residual lags = 1	2.969 (0.085)	Smooth Monotonic	0.002 (0.999)
No of residual lags = 2	1.541 (0.214)	Symmetric Non-Monotonic	0.002 (0.999)
No of residual lags = 3	2.395 (0.066)	Monotonic and Non-Monotonic	0.002 (0.999)

(b) Equally Weighted Series

$$y_t = (1.21y_{t-1} + (1- 1.21)y_{t-2}) * [\exp\{-0.056(y_{t-3})^2\}] + \hat{u}_t$$

(35.35) (35.35) (-2.83)
[0.000]

$$\ln\left(\sigma_t^2\right) = -0.081 + 0.194\left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] - 0.400\frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + 0.984\ln\left(\sigma_{t-1}^2\right)$$

(-3.17) (6.49) (-3.44) (218.7)

R^2 =	0.95	VR =	0.974
LR(1) =	24.07 (0.000)	H_0: No Remaining Nonlinearity	0.004 0.999
H_0: No Error Autocorrelation:		H_0: Parameter Consistency:	
No of residual lags = 1	0.861 (0.353)	Smooth Monotonic	0.004 (0.999)
No of residual lags = 2	0.518 (0.595)	Symmetric Non-Monotonic	0.003 (0.999)
No of residual lags = 3	1.476 (0.219)	Monotonic and Non-Monotonic	0.002 (0.997)

Note: The figures in parenthesis are t-ratios and the figures in the square brackets are marginal probability values. The marginal probability value is calculated from 5000 Monte Carlo experiments – see text for details. R^2 is the proportion of the variation in y_t explained by the model, LR(1) is a log-likelihood test against an AR(1)-EGARCH model, VR is the variation ratio. The autocorrelation test examines first to third order serial correlation. The test of no remaining nonlinearity uses an additive ESTAR model. An ESTAR-EGARCH with a delay of 2 is used for the Value Weighted process and a delay of 3 is used in the Equally-Weighted process.

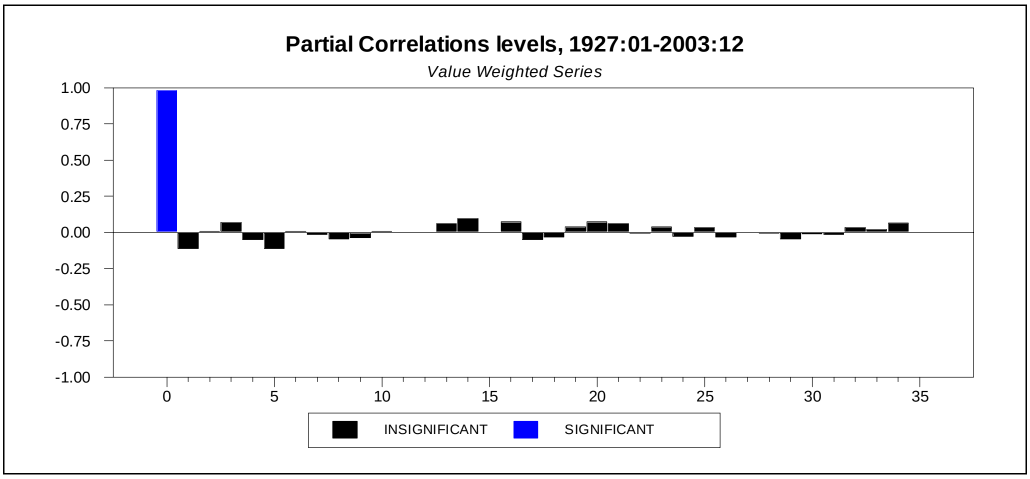
Table 8: Long Horizon Regression: Adjusted R- Squared

	$\overline{R}^2$	
k	***Equally-Weighted Excess Returns Series***	***Value-Weighted Excess Returns Series***
1	0.00030	0.00065
4	0.00733	0.00823
8	0.01560	0.02055
12	0.03814	0.04069
18	0.07785	0.06899
24	0.11284	0.08775
36	0.15336	0.11887
48	0.19631	0.14375

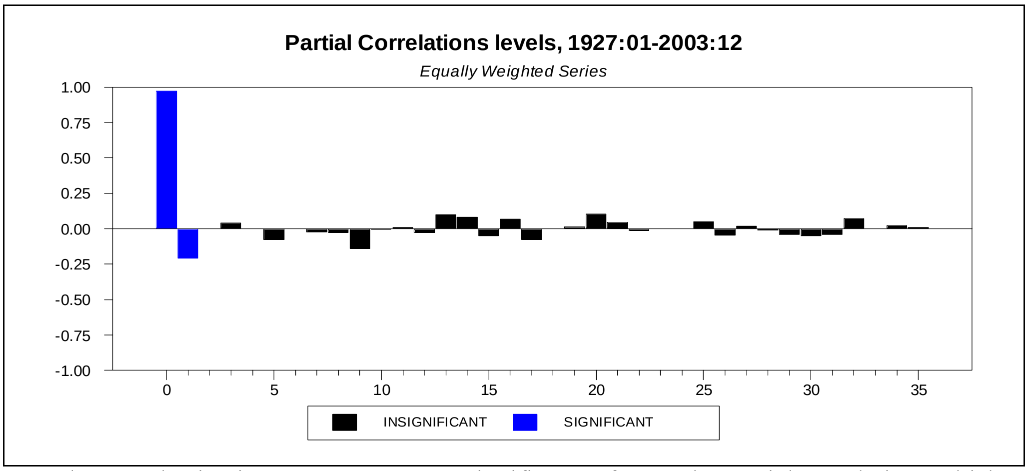
Note: The sample period is 1927:01-2003:12. $\overline{R}^2$ is the r-squared adjusted measure of multiple determination.

Figure 1: Partial Autocorrelation Function, Dividend-Price Ratio

(a) Value Weighted Series



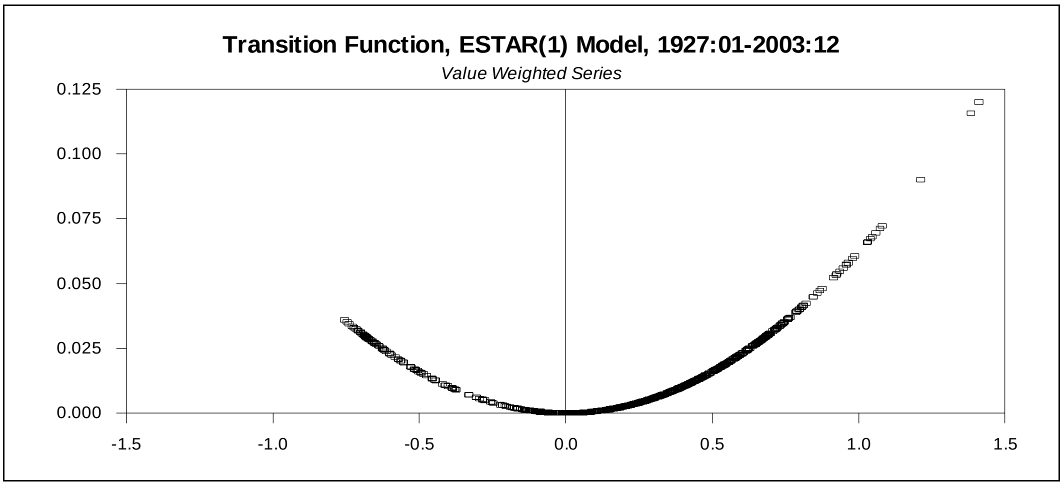
(b) Equally Weighted Series



Note: The sample size is 1927:01-2003:12. Significant refers to the partial correlations which are at least three standard deviations from the mean.

Figure 2: Transition Function, ESTAR Model

(a) Value Weighted Series



(b) Equally Weighted Series

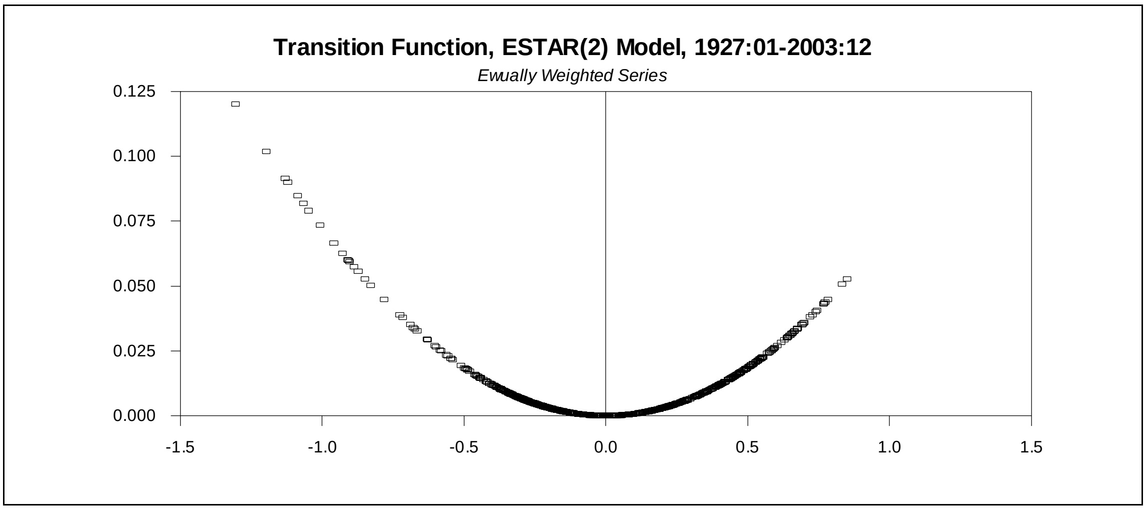
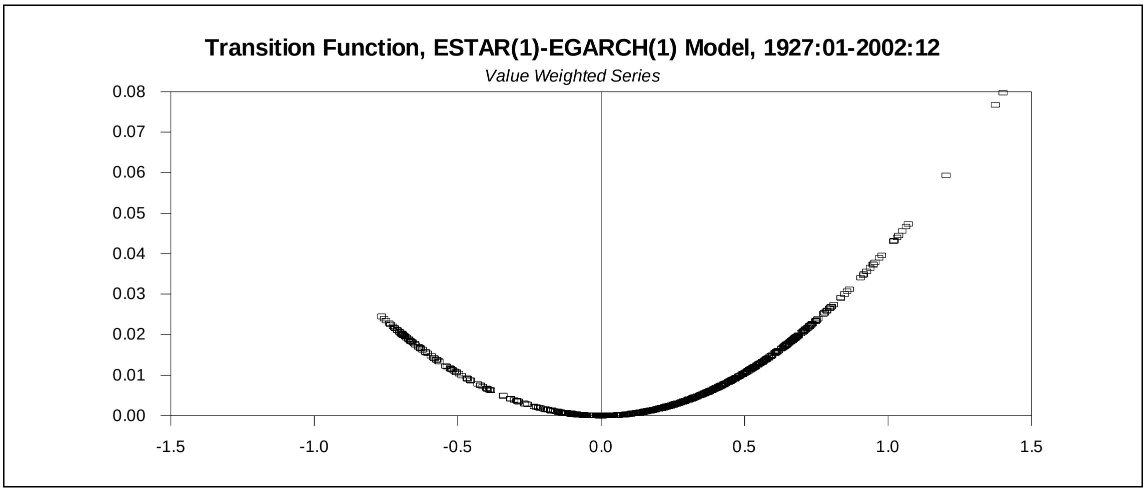


Figure 3: Transition Function, ESTAR-EGARCH Model

(a) Value Weighted Series



(b) Equally Weighted Series

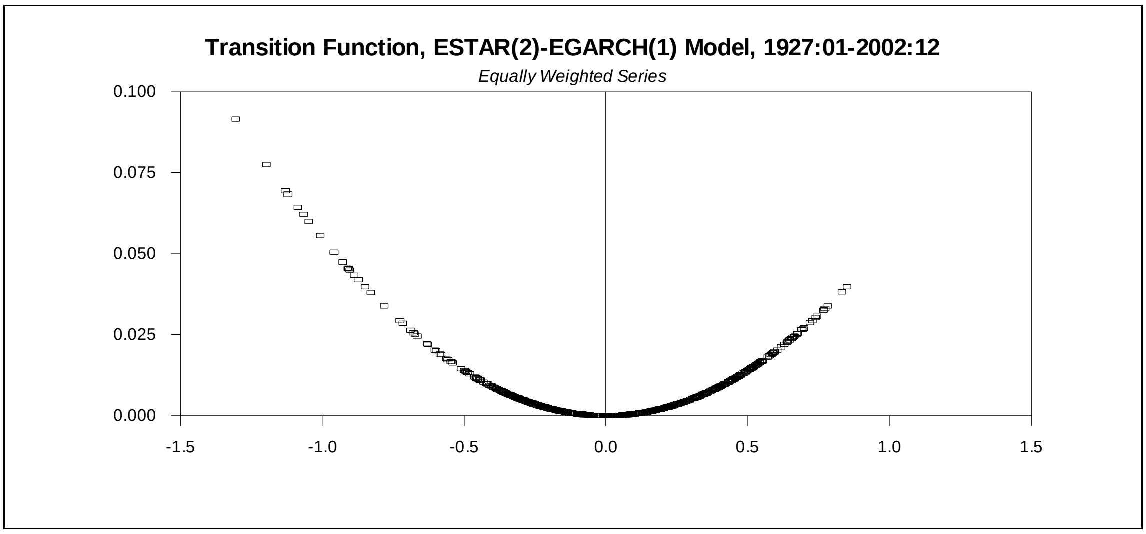
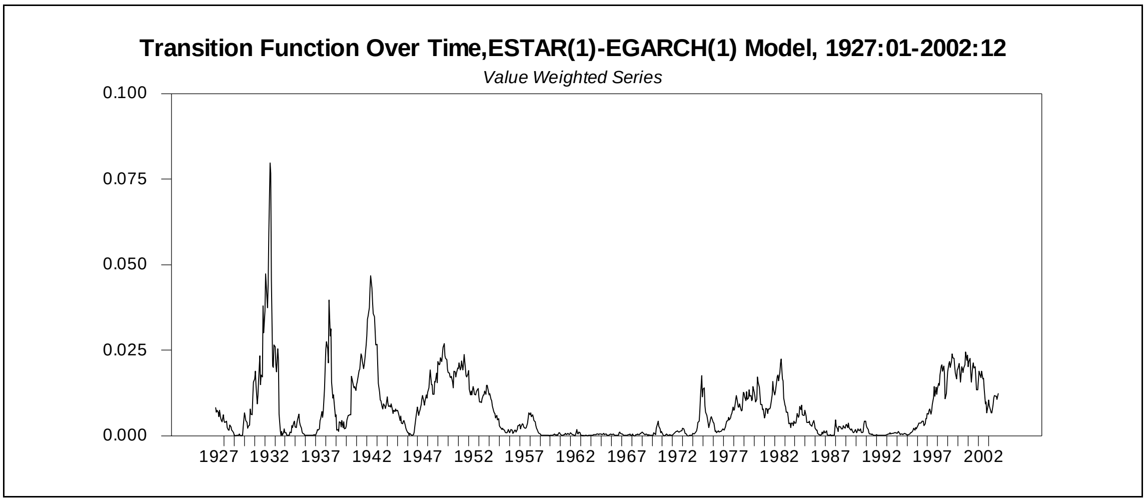


Figure 4: Transition Function over Time, ESTAR-EGARCH Model

(a) Value Weighted Series



(b) Equally Weighted Series

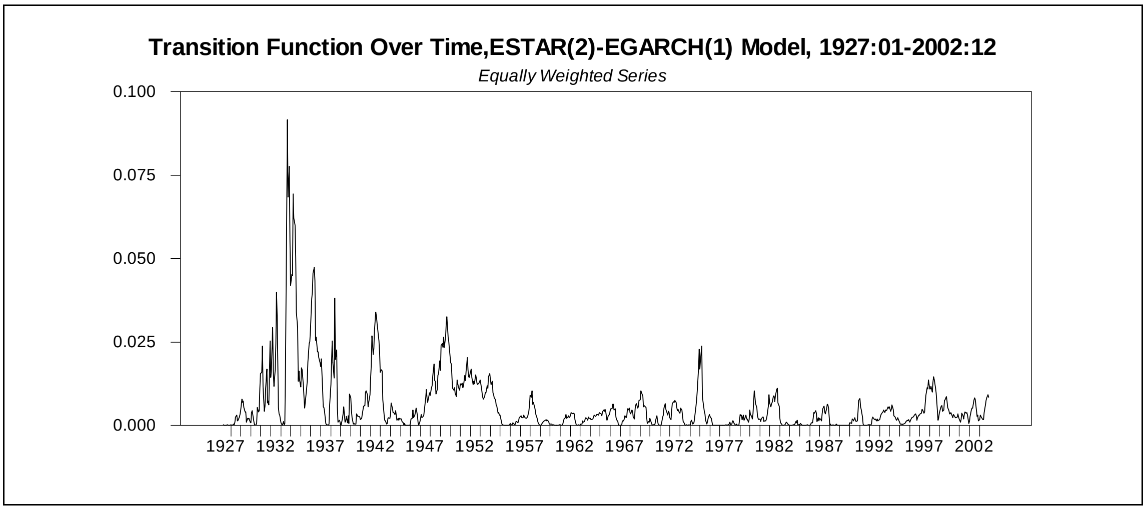
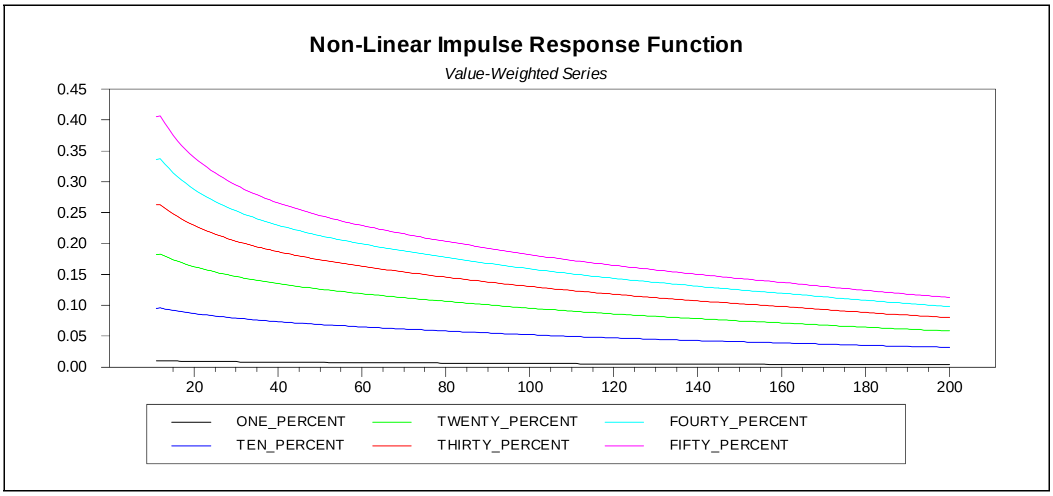
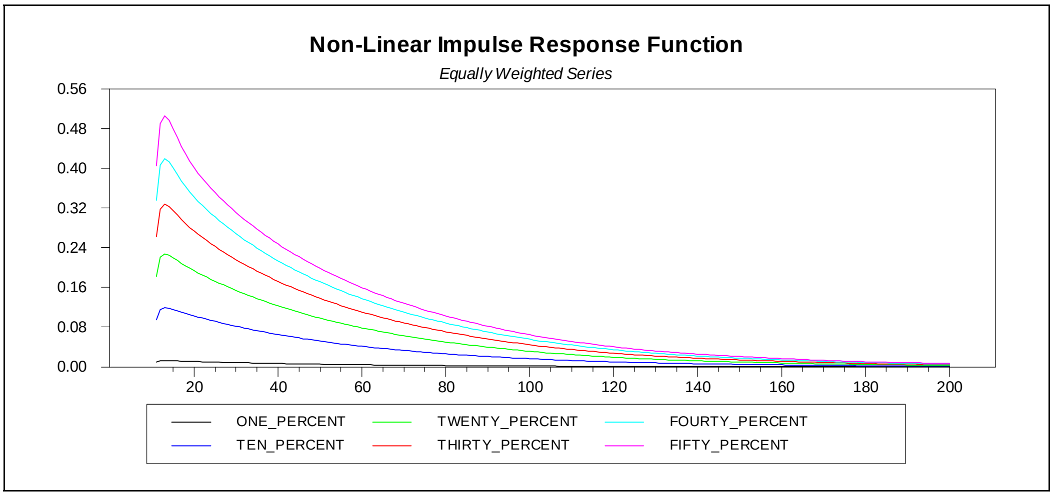


Figure 5: Nonlinear Impulse Response Functions

(a) Value Weighted Series

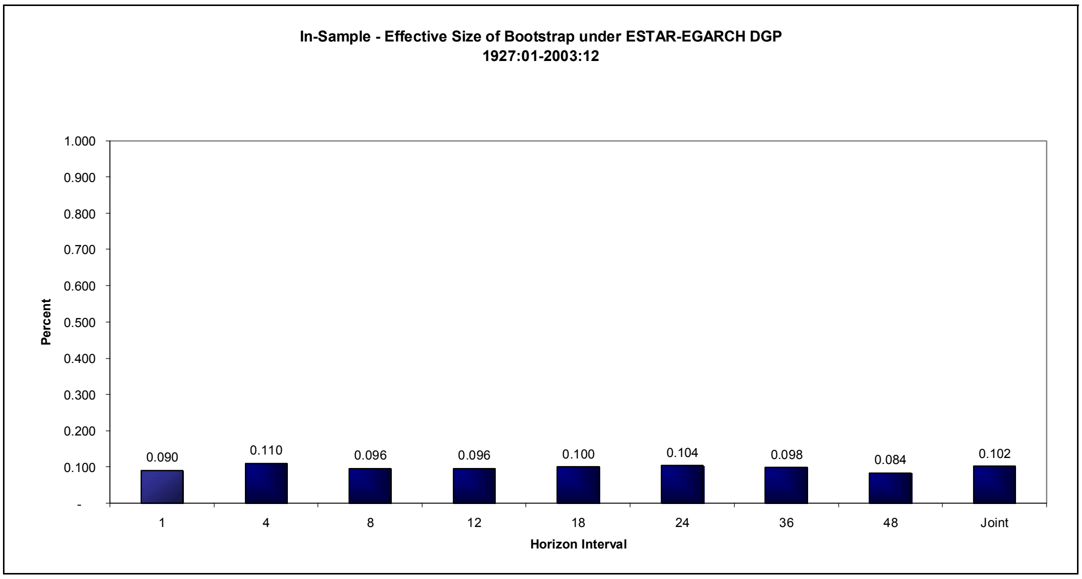


(b) Equally Weighted Series



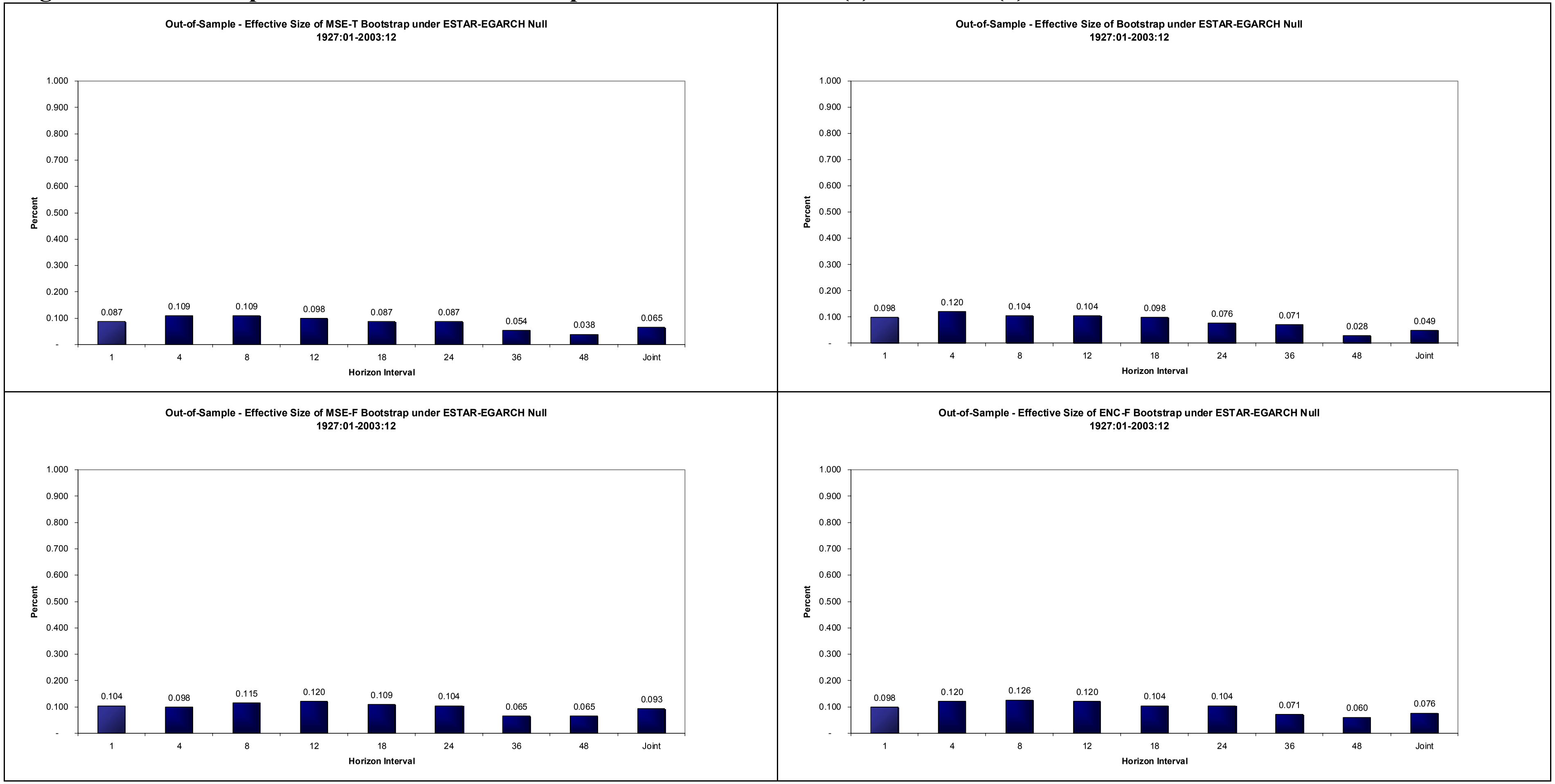
Note: These impulse responses are based on 1000 simulations.

Figure 6: In-Sample Effective Size of Bootstrap Test under ESTAR-EGARCH Null



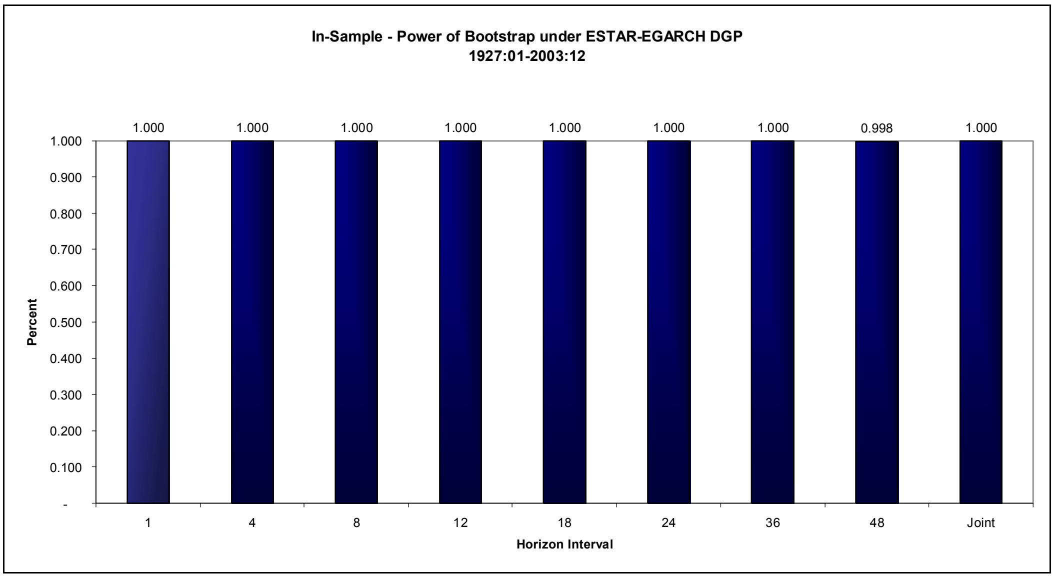
Note: This contains 500 Monte Carlos each with 500 bootstraps.

Figure 7 Out-of-Sample Effective Size of Bootstrap Test under an ESTAR(2)-EGARCH(1) Null



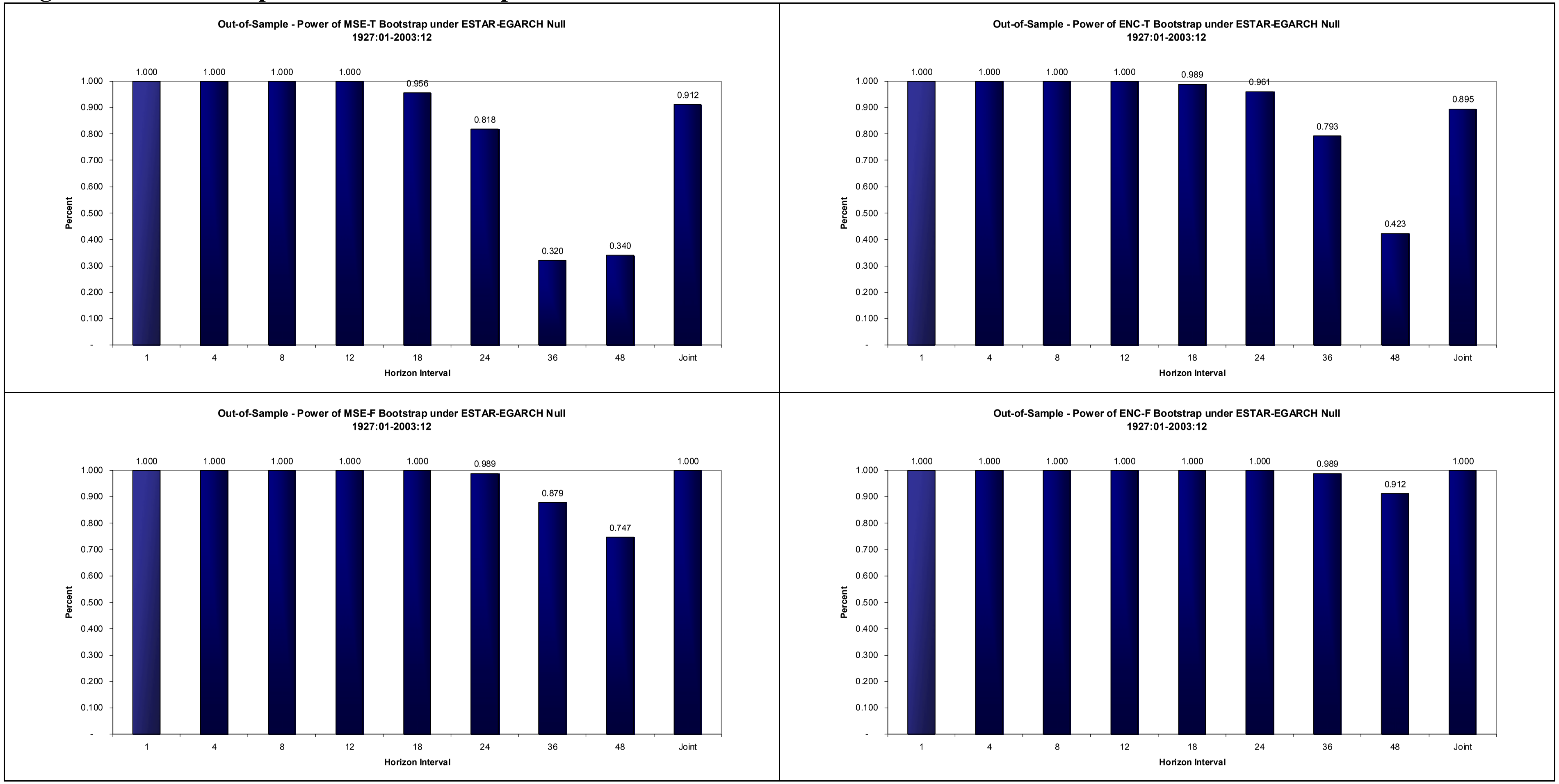
Note: Based on 500 Monte Carlos each with 500 bootstraps.

Figure 8: In-Sample Power of Bootstrap under ESTAR-EGARCH Null.



Note: Based on 500 Monte Carlos each with 500 bootstraps.

Figure 9: Out-of-Sample Power of Bootstrap under ESTAR-EGARCH Null.



Note: Based on 500 Monte Carlos each with 500 bootstraps.

Figure 10: In-Sample Forecasts – Equally Weighted and Value Weighted Series

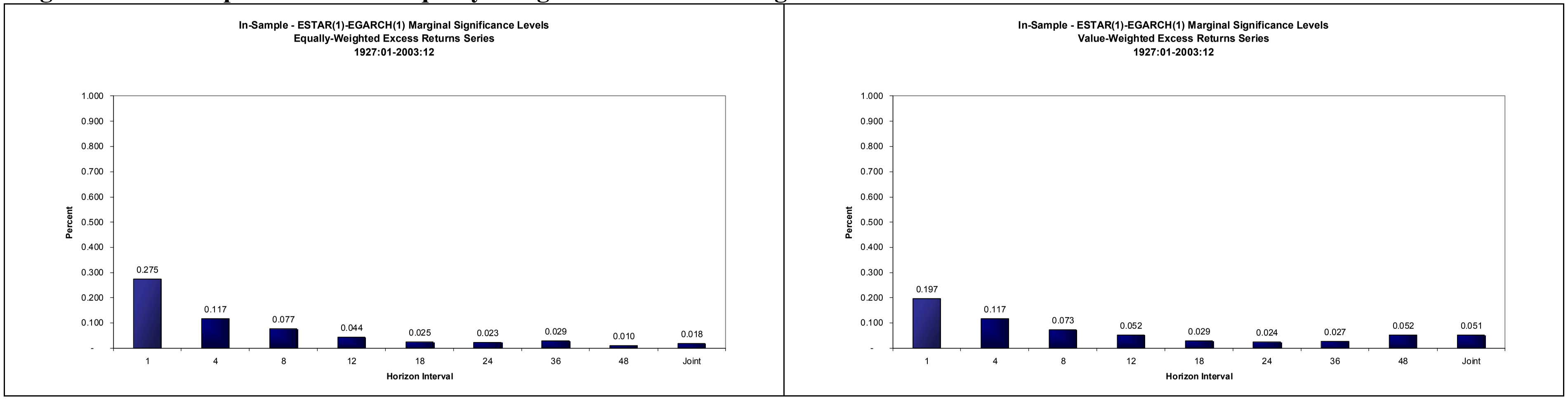
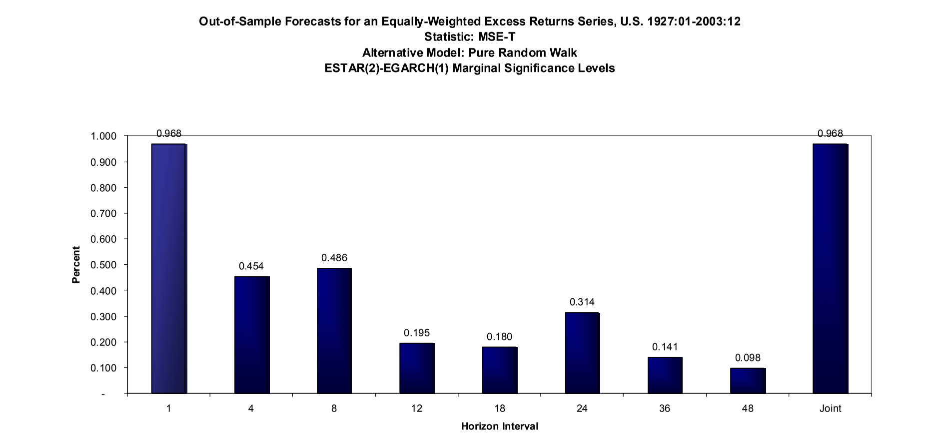
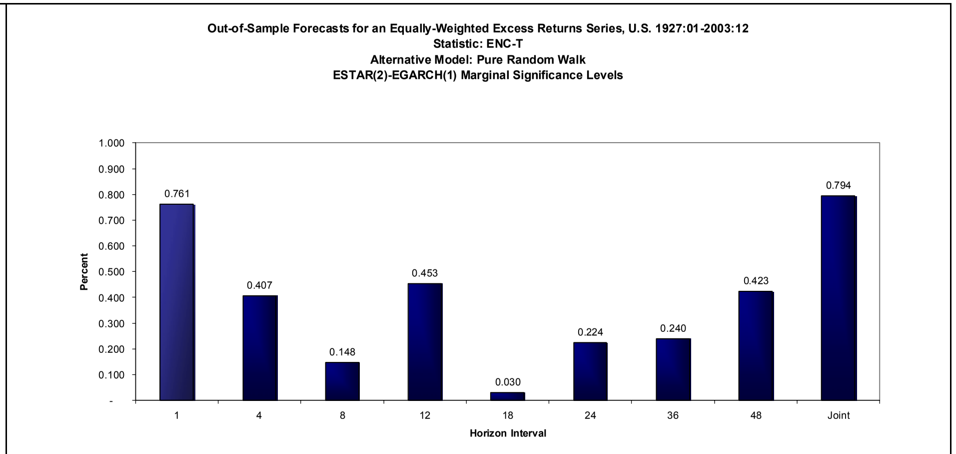
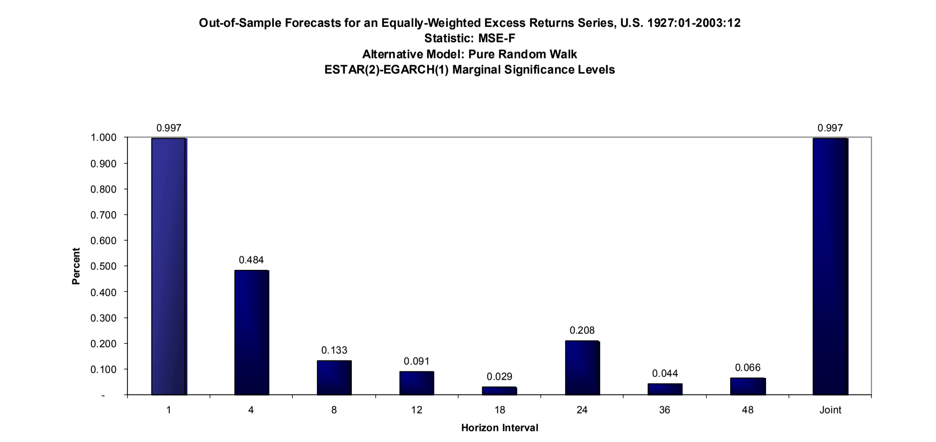
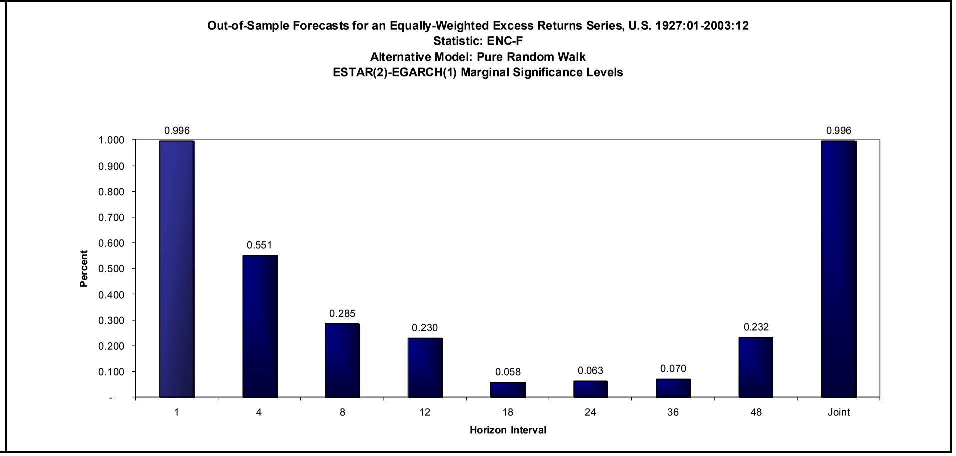


Figure 11: Out-of-Sample Forecasts: Equally-Weighted Excess Returns Series. Alternative Model: Pure Random Walk



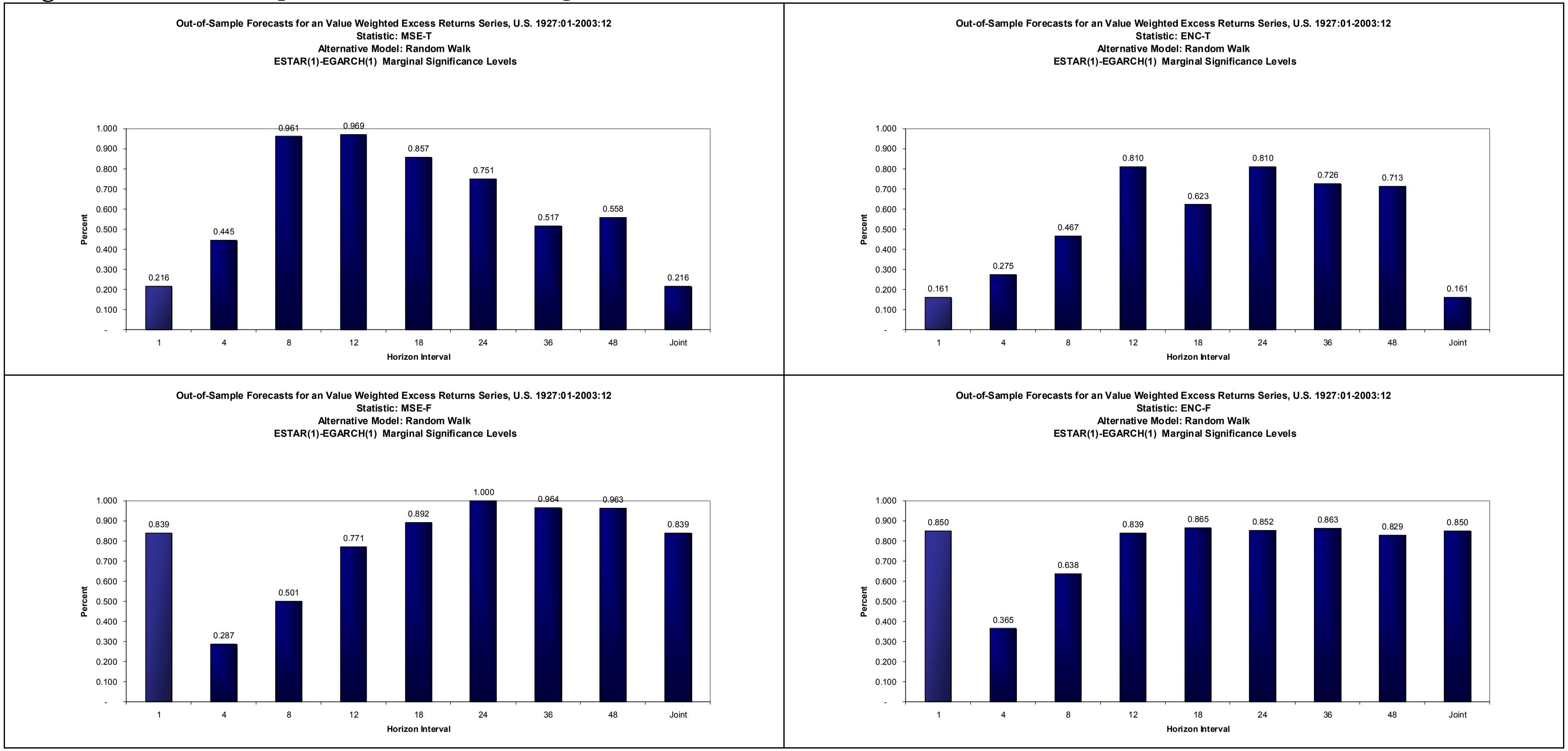






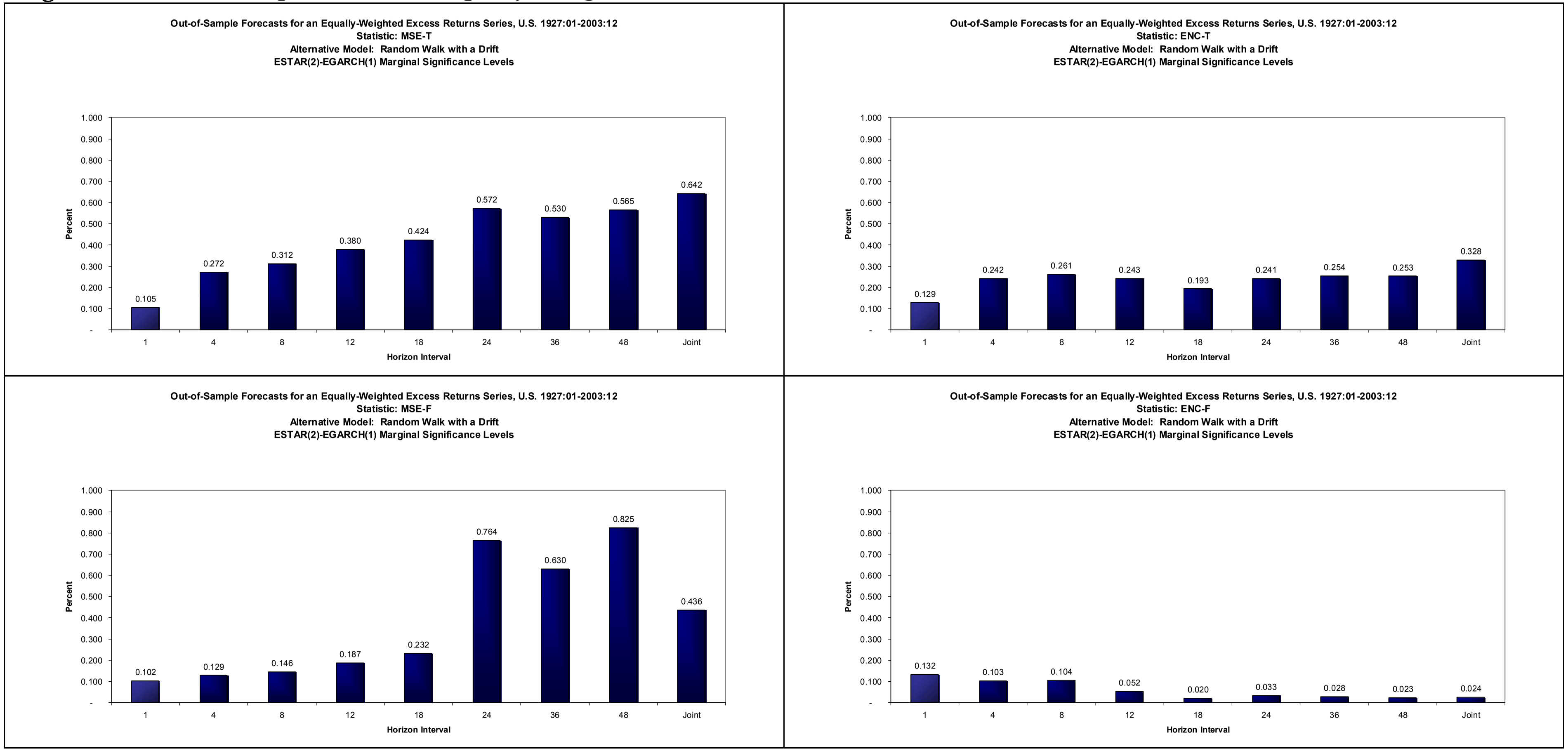
Note: Based on 2000 bootstraps.

Figure 12: Out-of-Sample Forecasts: Value-Weighted Excess Returns Series. Alternative Model: Pure Random Walk



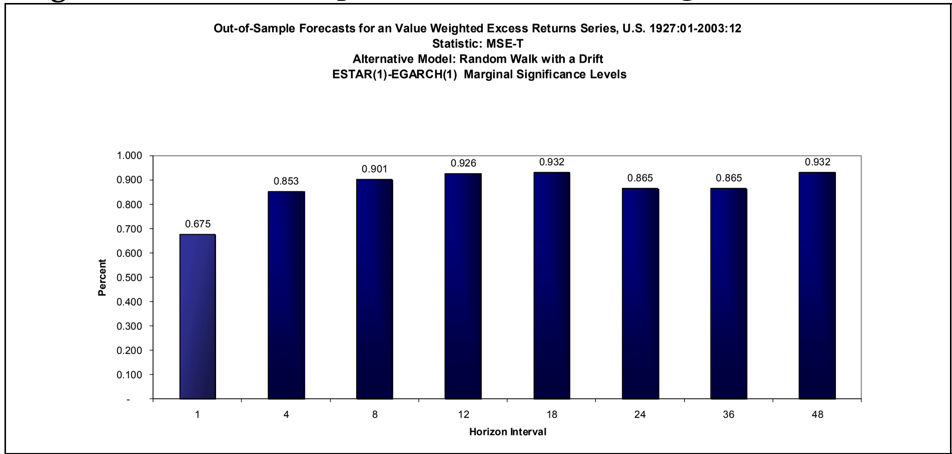
Note: Based on 2000 bootstraps.

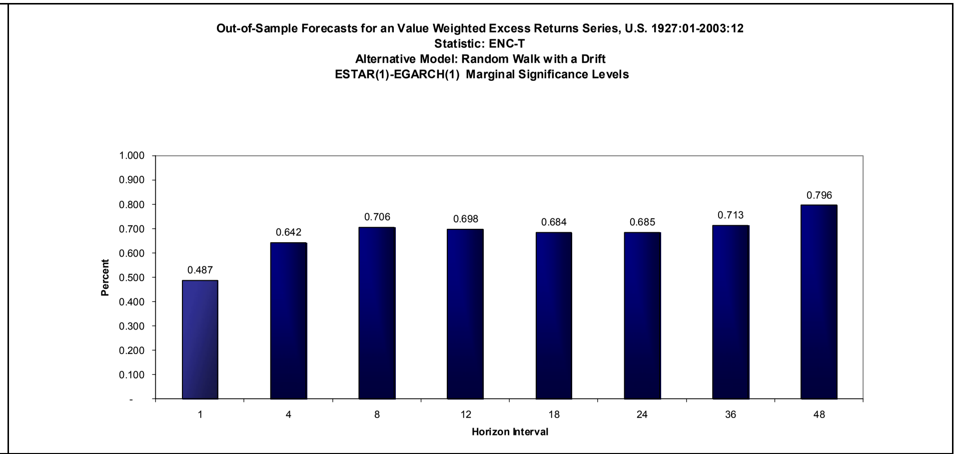
Figure 13: Out-of-Sample Forecasts: Equally-Weighted Excess Returns Series. Alternative Model: Random Walk with Drift

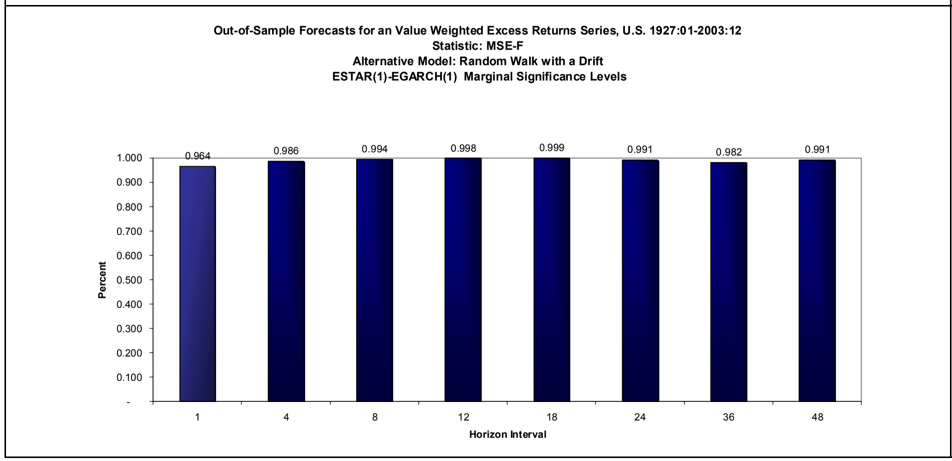


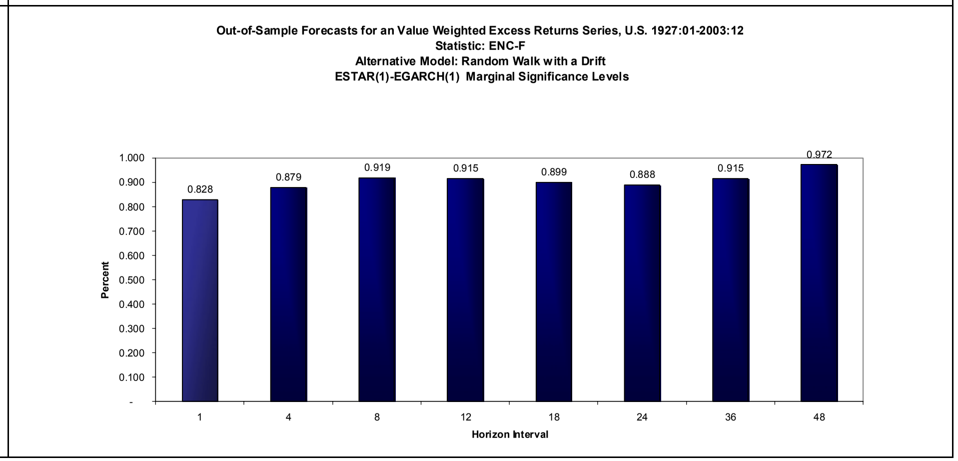
Note: Based on 2000 bootstraps.

Figure 14: Out-of-Sample Forecasts: Value-Weighted Excess Returns Series. Alternative Model: Random Walk with Drift









Note: Based on 2000 bootstraps.

Appendix 1: Bootstrap Technique for an ESTAR GARCH Process

1. Estimate the ESTAR(p)-EGARCH model:

$$\varepsilon_t = -[z_t - \mu] + \sum_{j=1}^{p} \lambda_j [z_{t-j} - \mu_y] + \sum_{j=1}^{p} \lambda_j [z_{t-j} - \mu_y] \Phi \left[\langle z_{t-d} \rangle_{d=1}^{\bar{d}} ; \mu ; \gamma \right] \sim N(0, \sigma_\varepsilon) \quad \text{(A1)}$$

where, $\ln\left(\sigma_t^2\right) = \beta_1 + \beta_2 \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] + \beta_3 \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta_4 \ln\left(\sigma_{t-1}^2\right)$

The autoregressive order p is determined by the partial autocorrelation function and the parameters are estimated through maximisation of the log-likelihood function.

2. Due to the assumed normality of the disturbances ε_t in (A1), the bootstrap residuals $\{\varepsilon_t^*\}$ are constructed accordingly; let u_t^* be an independent draw from a N(0,1) distribution, then the bootstrap residuals are computed as

$$\ln\left(\hat{\sigma}_t^2\right) = \hat{\beta}_1 + \hat{\beta}_2 \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] + \hat{\beta}_3 \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \hat{\beta}_4 \ln(\hat{\sigma}_{t-1}^2)$$

$$\varepsilon_t^* = u_t^* \sqrt{\hat{\sigma}_t^2}$$

3. The bootstrap samples are created recursively.