Can Dividend Yields Forecast Returns?

Mitual Kotecha
ING Baring
London

and

Sanjay Yadav
Financial Options Research Centre,
Warwick Business School, The University of Warwick

July 1995

We are grateful to Stewart Hodges and Jeremy Smith for many useful suggestions. We would like to thank seminar participants at Mitsubishi Finance, the Bank of England, Essex and Warwick Univerisities for helpful comments. The usual disclaimer applies.

Funding for this work was provided by the Corporate Members of the Financial Options Research Centre: HSBC Markets Ltd., Mitsubishi Finance International Ltd., Morgan Grenfell & Co., Swiss Bank Corporation and UBS Ltd.

Financial Options Research Centre
Warwick Business School
University of Warwick
Coventry
CV4 7AL
Phone: 01203 524118

FORC Preprint: 95/59

Can Dividend Yields Forecast Returns?

Mitul Kotecha

and

Sanjay Yadav

Abstract

There have been many studies on the use of dividend yields to predict returns, most concentrating on US data. There is a paucity of studies based on data from the UK stock market. We investigate the ability of dividend yields to forecast nominal returns on the value weighted FTA All Share Index, for return horizons ranging from one month to four years. We use non-overlapping as well as overlapping observations in our study and the sample period runs from 1965 to 1992.

Following Goetzmann and Jorion (1993), we allow for the lagged price relation between returns and dividend yields in our regression specification and use the bootstrap methodology to model the distribution of test statistics under the null hypothesis that returns and dividend yields are independent. We extend Goetzmann and Jorion's methodology by using WLS rather than OLS to estimate our model and thus correct explicitly for heteroscedasticity in the regression residuals. We also study the yield -return relationship using a methodology that does not impose a particular functional form on the data, a priori.

Our overall conclusion is that when we take adequate account of potential nonlinearities in the yield-return relationship, evidence supportive of return predictability largely disappears. This is consistent with the findings of Goetzmann and Jorion for US data, although their results were obtained within the framework of a linear model. The results obtained from applying the joint tests are in close agreement with the results obtained using individual return horizon statistics. Finally, none of the alternative covariance matrix estimators evaluated in this study are satisfactory; all of them yield standard error estimates which are substantially downward biased in the presence of overlapping observations. Therefore, researchers ought to focus primarily on the regression slope coefficients to conduct inference.

1. Introduction

The notion that the current dividend yield can forecast future stock returns has a long tradition dating back at least to Dow (1920). Early studies of the effects of dividend yields on stock returns, including those of Litzenberger and Ramaswamy (1979), Blume (1990), Gordon and Bradford (1980), and Miller and Scholes (1982) point to a positive and significant relation between dividend yield and stock returns. The intuition underlying this result is that stock prices are low relative to dividends when discount rates and expected returns are high, so that dividend yields should reflect changes in expected stock returns. Formally, in a perfect-certainty model, i.e. where agents know the constant growth rate of dividends (g) and the discount rate (r), the stock price P_t at time t is

$$P_t = \sum_{i=1}^{\infty} \frac{D_t(1+g)^i}{(1+r)^i} = \frac{D_{t+1}}{r-g} \quad (1)$$

The dividend yield is equal to the known discount rate minus the constant growth rate of dividends.

$$\frac{D_{t+1}}{P_t} = r - g \quad (2)$$

In this setup, the discount rate for dividends is the expected return on the stock. In a world of uncertainty, where g and r may vary over time, the link between discount rates and stock returns is not as straightforward as expressed in (2). However, the perfect-certainty model suggests that dividend yields may possibly capture variations in expected returns in the real world.

A nonlinear relationship between long-run dividend yields and returns in January is found by Keim (1985). He finds that the regression coefficients on dividend yields exhibit a significant January seasonal effect, even when controlling for size. When January observations are excluded the predictive power of dividend yields is no longer significant. Keim's results suggest that the observed relationship between long-run dividend yields and stock returns is not solely attributable to differences in marginal tax rates for dividends and capital gains.

Fama and French (1988) employ an ordinary least squares (OLS) regression framework and show that the dividend yield predicts a significant proportion of multiple year returns to the NYSE index. In a prominent study using US data, Goetzmann and Jorion (1993) point out that the results obtained by Fama and French as well as a number of other studies should be interpreted with caution. It is well known that dividend payments follow persistent patterns and so variations in dividend yields are dominated by stock price movements. This being the case, the right-hand-side variable in regression models used by researchers is in fact a lagged dependent variable and not exogenous as assumed when the model is estimated using OLS. This implies that the regression framework typically used to examine the relationship between dividend yields and stock returns is subject to biases which complicates the task of drawing inference.

Hodrick (1992) uses a VAR model in three variables: real stock returns, dividend yield and the T bill return. This formulation is appropriate because it permits the dividend yield to be endogenous rather than predetermined. Hodrick uncovers statistical evidence consistent with the view that stock returns are predictable.

The essential difference between Goetzmann and Jorion's paper and other papers in this area is twofold. First, Goetzmann and Jorion explicitly allow for the lagged price relation between returns and dividend yields in their regression specification. Secondly, they use the bootstrap methodology to model the distribution of test statistics under the null hypothesis that returns and dividend yields are independent. The use of the bootstrap methodology is particularly appropriate when the distribution of the time series under investigation, i.e. stock returns, is unknown and can only be guessed at. One unfortunate consequence of using the bootstrap methodology, however, is that the time varying volatility in stock returns is destroyed and so one is forced to assume that stock returns are homoscedastic under the null hypothesis. One possible procedure for alleviating this problem is to use the Weighted Least Squares (WLS) bootstrap methodology proposed by McQueen (1992). Failure to use this, or some other suitable technique that corrects explicitly for heteroscedasticity in the regression residuals means that the results presented in the Goetzmann and Jorion study are not robust to heteroscedasticity. Also, Goetzmann and Jorion do not entertain the possibility of a nonlinear relation between stock returns and dividend yields in their study.

There have been many studies on the use of dividend yields to predict returns, most concentrating on US data. There is a paucity of studies based on data from the UK stock market. In this paper we investigate the ability of dividend yields to forecast nominal returns on the value weighted FTA All Share Index, for return horizons ranging from one month to four years. We use non-overlapping as well as overlapping observations in our study and the sample period extends from 1965 to 1992. This study is of methodological interest as well, in that it uses a variety of techniques, thereby enabling one to gauge the sensitivity of the results to the specific technique employed. This paper has numerous distinguishing features. In a small Monte Carlo study we show how much statistical power is gained by using overlapping instead of non-overlapping returns. We investigate the performance of alternative covariance matrix estimators for computing standard errors when the regression model has overlapping data. We extend Goetzmann and Jorion's methodology by using WLS rather than OLS to estimate our model and thus correct explicitly for heteroscedasticity in the regression residuals. We explore the extent to which our results are sensitive to the assumption of a linear functional form. The aim is to develop a robust model to predict stock returns over different measurement intervals.

The remainder of this paper is organised as follows. In section 2, we describe the data used in this study and discuss the methodology which we use to compute the tests. In section 3, we present and interpret the empirical results. In the final section we summarise the results and offer our conclusions.

2. Methodology

2.1 Data Construction And Model Specification

Goetzmann and Jorion use data on the S&P 500 index over the period 1927 through 1990. Their data series are monthly total, capital, and income returns on the S&P 500. We calculate these series from the FTA All Share index. We use the FTA All Share total return index to calculate the total returns data and the FTA All Share price index to construct the capital returns data. Income returns are calculated from the capital, and total returns data series. The monthly capital, and income returns are used to construct a price series P, exclusive of dividends, from which monthly dividend payments are inferred.

Because of seasonalities in monthly dividends, an annual dividend series D was computed by reinvesting the dividends at the monthly riskless rate (the one-month Treasury bill rate).

Total returns are constructed from the FTA All Share total returns index as follows

$$\frac{TP_{t+1} - TP_t}{TP_t} = TR_{t,t+1}$$

where TP_t is the total returns index at period t, and TR_t is total returns at period t. Capital returns are constructed in a similar manner from the FTA All Share price index.

$$\frac{WDP_{t+1} - WDP_t}{WDP_t} = TRP_{t,t+1}$$

where WDP_t is the without dividend price or the price index at period t and TPR_t is the capital returns series.

Income returns are calculated by subtracting capital returns from total returns.

$$DP_t = TR_t - TPR_t$$

where the variable DP_t is the income returns data series.

To form a price series P_t that excludes the reinvestment of dividends, P_0 is set at 100, then P_t is recursively computed from the series TPR_t (capital returns series).

The monthly dividend D_t is recursively computed from the series DP_t

$$D_t = DP_t \times P_t$$

A monthly annualised dividend series DD_t is computed from continuously compounding twelve monthly dividends at the one-month Treasury bill rate R_t. This annualised dividend is calculated for each observation in the series.

$$DD_t = D_t + (1+R_t)\, D_{t-1} + (1+R_t)\,(1+R_{t-1})\, D_{t-2} + \ldots$$

The actual dividend yield is thus defined as

$$DT_t = \frac{DD_t}{P_t}$$

An OLS regression is then performed with total return TR_t regressed on the dividend yield DT_t:

$$TR(t,t+T) = \alpha(T) + \beta(T)\, DT(t) + \varepsilon(t,t+T) \qquad (3)$$

The null hypothesis being tested is that there is no relation between TR(t, t + T) and DT(t). A rejection of the null hypothesis would indicate that dividend yields can help to predict returns.

2.2 The Use Of Randomisation To Estimate Significance Levels

Goetzmann and Jorion employ the bootstrap methodology to investigate the sampling distribution of the beta statistic under the null, whereas we use the randomisation methodology. Both methods rely on resampling the data; however, the main difference between randomisation and bootstrapping is that in the latter method the data is sampled with replacement. Randomisation is based on the premise that under the null hypothesis one variable is distributed independently of another. The null hypothesis in which we are interested in is that returns are distributed independently of dividend yields.

Randomisation shuffles the data to destroy any time dependence and then recalculates the test statistic for each shuffle to estimate its distribution under the null. This experiment is repeated a thousand times. The empirical probability value (p-value) is the proportion of times the randomised beta coefficient exceeds the historic beta coefficient. This method assumes no knowledge of the distribution of stock returns and therefore avoids the problem of non-normality.

Following Goetzmann and Jorion, the randomisation methodology is adopted to enable us to maintain the temporal relationship between returns, dividends, and prices that is consistent with their historical behaviour. We know that price levels at a particular point in time are dependent upon the capital appreciation return history up until that point and that dividends exhibit a high degree of serial correlation. To capture this relationship in the randomisation model, we implement the following procedure:

1. Randomly shuffle historical total returns data to obtain simulated total returns TR*.
2. Subtract income return series DP* from simulated total returns TR* to obtain capital return series TPR*.
3. Capital return series TPR* is compounded to calculate a pseudo price level series P*, which is used to create pseudo dividend yield DT*, where DT*=DD/P*, DD=actual annual dividend flow and P*=simulated price series.
4. Regress multiple horizon returns on dividend yields and save slope coefficient and R^2.
5. Repeat steps 1 to 4 one thousand times.

As total returns have been shuffled randomly, there is no relationship between returns and dividends, and the dividend series is highly autocorrelated. The price series is similar to its capital appreciation return history because it is formed from the recursive computation of the pseudo capital appreciation return series.

2.3 Alternative Covariance Matrix Estimators

It is well known that the use of overlapping observations yields more powerful tests; however, it is not clear exactly how much more statistical power is gained by using overlapping as opposed to non-overlapping data. The principal drawback associated with the use of overlapping data is that they induce serial correlation in the regression disturbances which leads to biased and inconsistent estimates of the conventional OLS standard errors. Under the null hypothesis, the regression disturbance follows an MA process of order m, where m is equal to the return measurement interval minus one. The point estimates of the slope coefficient, however, are consistent but not efficient.

Various methods have been proposed in the literature to purge the standard errors of this bias. Hansen (1982) proposed a GMM estimator to correct the OLS standard errors for the presence of serial correlation. The covariance matrix for the vector of parameters in equation (3), $\delta' = (\alpha, \beta)$, is calculated as $V(\delta) = (X' X)^{-1}\ \hat{S}_T\ (X' X)^{-1}$,

$$\text{where} \quad \hat{S}_T = \hat{\Omega}_o + \sum_{j=1}^{m} [\hat{\Omega}_j + \hat{\Omega}'_j],\ \hat{\Omega}_j = \sum_{t=j+1}^{T} e_t x'_t x_{t-j} e_{t-j},$$

x_t is the vector of explanatory variables and e_t the residuals in (3). This GMM method is used in Hansen and Hodrick (1980), and they note (p.836) this procedure has greater power than an alternative solution, which involves constructing new series made up of non-overlapping observations of the original series and undertaking the analysis using standard OLS on this subsets of observations.

Hansen's (1982) method does not ensure that the covariance matrix is positive definite. Newey and West (1987) proposed an alternative method for obtaining serial correlation consistent and heteroscedasticity consistent standard errors, with a positive definite covariance matrix. This new procedure modified the Hansen (1982) GMM estimator by weighting the sample autocovariance function, such that the weights decline as j increased. For the Bartlett weighting method,

$$\hat{S}_T = \hat{\Omega}_o + \sum_{j=1}^{m} \omega(j,m)\,[\hat{\Omega}_j + \hat{\Omega}'_j], \text{ where } \omega(j,m) = 1 - [j/(m+1)].$$

Alternatively, for the Quadratic Spectral (QS) weights, $\hat{S}_T = \hat{\Omega}_o + \sum_{j=1}^{n-1} \omega\ (j)\ [\hat{\Omega}_j + \hat{\Omega}'_j]$, where

$$\omega(j) = \frac{25}{12\pi^2 u^2}\left(\frac{\sin(6\pi u/5)}{6\pi u/5} - \cos(6\pi u/5)\right), \textit{ and } u = j/(m+1)\cdot$$

The Newey-West method with Bartlett weights has been used by Frankel and Froot (1987), however, there are few examples of an application of Newey-West with QS weights.

Finally, Andrews and Monahan (1991) extended the class of GMM estimators, by first prewhitening the series and then using the Newey-West covariance matrix with QS weights on the prewhitened series. The prewhitening entails fitting a Vector AutoRegressive model of order b (VAR(b)) to the series $z_t = x_t e_t$. In their study Andrews and Monahan (1991) use a VAR(1) model to approximate both autoregressive (AR) and moving average (MA) processes.

For these alternative covariance matrices the choice of the bandwidth parameter, m, is important for obtaining good standard error estimates. Hansen and Hodrick (1980) use a bandwidth parameter, m = order of the moving average process. Frankel and Froot (1987) consider two values of m, m = order of the moving average process and m = twice the order of the moving average process. Recently, Andrews (1991) developed a method for choosing the optimal bandwidth parameter, $\hat{Z}_T$, according to the type of weighting scheme used. For Bartlett weights, $\omega(j) = 1 - (j/\hat{Z}_T)$, $\hat{Z}_t = 1.1447(\hat{\alpha}(1)T)^{1/3}$, where

$$\hat{\alpha}(1) = \frac{\sum_{a=1}^{p} w_a \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^6(1+\hat{\rho}_a)^2}}{\sum_{a=1}^{p} w_a \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4}} \quad \text{and} \quad w_a = \begin{bmatrix} 0, & a=1 \\ 1, & a\neq 1 \end{bmatrix}, \quad \{(\hat{\rho}_a, \hat{\sigma}_a^2), a = 1,\ldots,p\}$$

are the parameter estimate and innovation variance of an AR(1) model fitted to each of the series formed as the product of the p explanatory variables, x_{at}, and the residuals, e_t. For the QS weights the parameter u is calculated as $u = j/\hat{Z}_T$ where $\hat{Z}_T = 1.3221(\hat{\alpha}(2)T)^{1/5}$, and

$$\hat{\alpha}(2) = \frac{\sum_{p=1}^{p} w_a \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8}}{\sum_{a=1}^{p} w_a \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4}} .$$

It would be interesting to verify, for our particular empirical model, exactly how much statistical power is gained by using overlapping data. Also, it would be worthwhile to investigate the relative performance of the alternative covariance matrix estimators. To shed light on these issues we undertake a Monte Carlo study with a sample size that corresponds to the sample size that we use in our empirical study. For the Monte Carlo study, the total returns series is created by sampling randomly from a normal distribution with mean and variance equal to those of the historical data.

2.4 Joint tests

Richardson (1989) is critical of studies which focus on individual return horizon statistics, typically the return horizon which provides strongest evidence consistent with deviations from a random walk model, on the grounds that individual test statistics are not independent and so reliance on them may yield misleading results. In this paper we conduct joint tests over multiple return horizons. The joint test we use in this study is simply the average of the individual regression betas.

2.5 Stratified randomisation and WLS randomisation

UK stock return variances were higher in the stock market boom in 1975, the crash around September 1981, and the crash in October 1987. It is also apparent that volatility reverts to pre crash/boom levels after these periods. Thus, one would expect that the return series is heteroscedastic and that the regression errors exhibit variation through time. Randomisation, however, destroys the temporal pattern of heteroscedasticity present in the data. Consequently, under randomisation we are forced to assume that returns are homoscedastic under the null hypothesis. It would be interesting to investigate how the presence of heteroscedasticity under the null alters inference. To preserve the pattern of heteroscedasticity present in the actual return series we carry out a stratified randomisation of the total return data. We partition the sample into five regimes or strata, where each regime displays a different level of return variance. To generate an artificial return series, the return data are sampled without replacement from each regime.

One drawback is that the partitioning of data into different regimes based on return variance is very arbitrary. Researchers may disagree about the number of different regimes based on return variance required to ensure that the pattern of heteroscedasticity in the data is maintained. In the limit, one would require as many variance regimes as there are data points, thereby rendering this technique impractical. An alternative and preferable approach is to use WLS which deflates the dependent as well as the independent variables by estimates of the monthly standard deviation of returns. This method corrects explicitly for heteroscedasticity in the sense that it renders the regression residual homoscedastic and thus yields more efficient parameter estimates.

Following French, Schwert and Stambaugh (1987) and McQueen (1992) we utilise daily returns on the FTA All Share index to construct unconditional estimates of the monthly variance of stock returns. The monthly return variance is computed as:

$$\hat{\sigma}_t^2 = \sum_{i=1}^{n} TR_{i,t}^2 + 2 \sum_{i=1}^{n-1} TR_{i,t} \times TR_{i+1,t}$$

where n is the number of trading days in month t and $TR_{i,t}$ is the return on the FTA All Share index on day i in month t.

Under the null, stock returns and dividend yields are independent and so the standard deviation of the error term in (3) is

$$\hat{\sigma}_{t,t+T} = (\hat{\sigma}^2{}_t + \hat{\sigma}^2{}_{t+1} + \ldots + \hat{\sigma}^2{}_{t+T})^{0.5} . \qquad (4)$$

WLS technique is implemented by using $\frac{1}{\hat{\sigma}_{t,t+T}}$ as weights.

The WLS Randomisation technique is very similar to OLS Randomisation. The only difference is that under WLS Randomisation, monthly returns and their corresponding variances (computed using daily data) are shuffled together. Shuffled returns and variances are subsequently aggregated to obtain overlapping observations on returns and variances. The reciprocal of the square root of the aggregated variance is then used as weights and WLS estimation is implemented.

2.6 Nonparametric Model

It is not entirely clear from (3) that the link between dividend yields and stock returns is linear and stable over time. Indeed, Keim finds a nonlinear relationship between dividend yields and returns for the month of January. In a recent paper based on data from the US stock market, Christie and Huang (1994) study the yield-return relationship using a methodology that does not compel them to impose a particular functional form on the data, a priori.

Failure to take account of possible nonlinear dependence between stock returns and dividend yields will lead to functional form misspecification. It is well known that OLS will yield biased and inconsistent estimates if the model is misspecified. To rectify this situation, we make use of a nonparametric estimation technique, Nearest Neighbour (NN), to estimate the regression function. Other related nonparametric techniques, namely kernels and series could be used. However, our choice is motivated by the superior performance of the NN estimator relative to the other estimators in finite samples.

Additional justification for a nonlinear model is provided by another strand of the literature. The notion that stock returns exhibit strong ARCH effects is amply documented in the literature. It is not inconceiveable that ARCH effects may be a symptom of omitted nonlinearities in the regression model. Diebold and Nason (1990) point out that strong ARCH effects may show up in the data as a consequence of our failure to take adequate account of possible nonlinearities in the model. It would seem to be worthwhile to explore the robustness of our results to alternative assumptions regarding the functional form of the model.

Therefore, in what follows, we do not assume that the functional form of the model is linear, when even casual inspection of the scatterplot (Figure 1) suggests a nonlinear specification would be more suitable. Hence we avoid reliance on conventional econometric techniques which impose a particular functional form, a priori.

The model is $TR_{t,t+T} = F(DT_t) + \varepsilon_{t,t+T}$ (5)

where F(.) is a smooth function and the residual term has 0 mean and constant variance. NN technique entails estimating F at a point DT*. Let $H(DT^*,DT_t) = |DT^*-DT_t|$ be the Euclidean distance between DT* and DT_t. The neighbourhood weight function is given by:

$$W_t(DT^*, DT_t, DT_k) = V[H(DT^*, DT_t)/H(DT^*, DT_k)],$$

where $H(DT^*,DT_k)$ is the Euclidean distance between DT^* and the kth nearest DT_t. In line with the recommendations of Cleveland and Devlin (1988) and Diebold and Nason (1990), we utilise the tricube weight function for V(.):

$$V(u) = (1-u^3)^3 \quad \text{if } u < 1,$$
$$= 0 \quad \text{if } u \geq 1.$$

The fitted value of the dependent variable at the point DT*, i.e. F(DT*) is given by:

$$TR^*_{t,t+T} = F(DT^*) = \beta DT^*$$

To obtain an estimate of the slope coefficient β, WLS regression is performed with weight $W_t(DT^*,DT_t,DT_k)$. In other words, to obtain an estimate of β we minimise:

$$\Sigma W_t (DT^*, DT_t, DT_k) \times (TR_{t,t+k} - \beta DT_t),$$

where the summation is over the number of observations in the estimation period. Following standard practise, the slope coefficient is the first derivative of $F(DT_t)$, evaluated at the average value of the dividend yield variable in the estimation period.

The use of NN estimator requires that the researcher make some choices which may be considered judgemental, especially regarding the number of nearest neighbours (k) to be used. However, in our empirical study we investigate the sensitivity of our results to different choices of k. We allow k to vary between 0.4 and 0.8 times the sample size, in 5 steps, and find that the results are remarkably robust. To conserve space, beta coefficients for k equal to 0.5 times the sample size are reported in the paper.

3. Results

We begin with a discussion of the results obtained from the Monte Carlo simulations. Table 1 presents the standard errors of the slope coefficients across the 10000 regressions. From Table 1 it is evident that the standard errors are higher for non-overlapping observations than overlapping observations. It is clear that more accurate estimates are obtained using overlapping data, especially as the return horizon lengthens. For one to six month return horizons, however, there is little gain in the accuracy with which slope coefficients are estimated using overlapping data.

Next, we consider the performance of alternative covariance matrix estimators in the presence of overlapping observations. A variety of alternative covariance matrix estimation methods are considered: standard OLS (OLS), Newey-West with Bartlett weights and m = order of MA process (NWB), Newey-West with Quadratic Spectral weights and m = order of MA process (NWQS), Newey-West with Bartlett weights and m chosen using Andrews (1991) procedure (NWBA), Newey-West with Quadratic Spectral weights and m chosen using Andrews (1991) procedure (NWQSA), the Andrews and Monahan (1992) method with QS weights and m chosen using Andrews (1991) procedure (AM), and the Hansen (1982) method (HAN). The results are based on 10000 replications and are presented in Table 2.

We find, in general, that the corrected standard errors obtained using the various covariance matrix estimators are unsatisfactory; they underestimate the true standard errors and the magnitude of the downward bias increases with the return measurement interval. Consequently, inference based on the slope t statistics is likely to be seriously misleading. Therefore, we eschew reliance on slope t statistics to conduct inference and focus attention on the p-value of the slope coefficient instead.

Table 3 summarises the results obtained from regressing overlapping and non-overlapping returns on dividend yields for the period 1965-1992. All the slope coefficients are positive and increase over the return measurement interval, from 0.369 for one month returns to 8.695 for the 48 month period. The empirical p-values associated with all the regression slopes are well below 0.05, which implies that the slope coefficients are statistically significant. Notice also that R^2 increases in the time horizon, from almost 0 to around 62% for the 48 month return. All the R^2 values are highly significant. The results from applying the joint test confirms the conclusion obtained using individual return horizon regression statistics, that dividend yields can predict stock returns.

Table 4 reports the results obtained from performing a stratified randomisation of the returns data under the null. In general, stratified randomisation produces p-values that are larger than those under randomisation. However, the basic conclusion remains unaltered as none of the empirical p-values exceed 0.05. In other words, all the beta coefficients and R^2 remain statistically significant. As pointed out earlier, stratified randomisation is not a technique which is satisfactory in ensuring that the temporal pattern of return volatility is maintained under the null. Hence, we should treat the results with caution. Instead, more attention will be focused on the WLS estimates of slope coefficients in order to draw inference.

Table 5 reports the WLS estimates of slope coefficients. None of the slope coefficients for regressions utilising non-overlapping observations differ significantly from those obtained under the null hypothesis that returns and dividend yields are independent. However, the reverse is true for slope coefficients obtained using overlapping observations; the empirical p-values are well below 0.05 for 12, 24 and 36 month return horizons and marginally below 0.05 for 48 month returns. In contrast to the findings reported using OLS randomisation, WLS randomisation only yields regression statistics suggestive of a significant positive relation between returns and dividend yields over longer return horizons.

Table 6 reports the beta coefficients obtained from using the nonparametric NN technique due to Cleveland and Devlin. The p-values associated with the beta coefficients suggests that dividend yields cannot forecast returns, with the sole exception of returns for the 24 month period. Specifically, for some values of k, the beta coefficient on 24 month returns is marginally significant as evidenced by a p-value that drops slightly below 0.05. Subperiod analysis using this technique demonstrates that the results are not sensitive to the particular sample period under scrutiny. Compared to OLS and WLS Randomisation, the p-values produced under NN randomisation are substantially higher, implying that use of either OLS or WLS is likely to lead to an over rejection of the null hypothesis of no predictability.

Figure 2 graphs the empirical distribution of the beta coefficient for 36 month returns that is obtained from the yield-return model under the null. The OLS beta coefficient appears approximately normal; it fails to capture the skewness in the beta coefficient as compared to the NN beta coefficient. This failure to capture the skewness is evident for overlapping as well as non-overlapping observations (see Figure 3).

Table 7 presents the results of OLS randomisation for subperiods. The results for the two subperiods are strikingly different. For the 1965-1978 subperiod, the regression slope coefficients are significant but for the latter subperiod, the evidence appears to be in favour of the null hypothesis that dividend yields cannot forecast returns. The p-value for the 36 month return is 0.018 in the first subperiod and 0.299 in the second subperiod. Highest statistically significant value of R^2, 68%, is obtained for the 36 month return horizon.

Further subperiod empirical analysis reveals that return predictability is largely confined to the 1971-1975 period. This period witnessed massive swings in the level of the market. Annualised return standard deviation for the 1971-1975 period was higher than the 1965-1970 and 1976-1992 period by 25% and 30%, respectively. Evidence consistent with the notion that US stock returns contain predictable components has been documented by Kim, Nelson and Startz (1991) and McQueen (1992) for a highly volatile period, i.e. the period following the 1929 Crash and prior to WWII.

Table 8 reports results using the WLS technique to reestimate the model over the two subperiods. Once again, significant slope coefficients are only found for the early subperiod.

It would thus appear that evidence supportive of the ability of dividend yield to forecast returns is concentrated entirely in the first subperiod, and then only if we assume a linear relationship between returns and dividend yield. However, if we allow the return-yield relationship to be nonlinear, there is little evidence to reject the null that dividend yields cannot predict stock returns.

4. Conclusions

The aim of this study was to test the null hypothesis that the dividend yield has no predictive content. The results reported in this paper are robust to the non-normality and heteroscedasticity of stock returns. We find that failure to correct explicitly for heteroscedasticity in the regression residuals is likely to overstate the evidence that dividend yields can forecast returns. The OLS randomisation yields probability values (p-values) that are lower than that obtained under WLS randomisation, implying that

OLS randomisation rejects the null hypothesis more frequently. Researchers should be wary of reporting results based exclusively on OLS and WLS randomisation.

The results obtained from applying the joint tests are in close agreement with the results obtained using individual return horizon statistics. None of the alternative covariance matrix estimators evaluated in this study are satisfactory; all of them yield standard error estimates which are substantially downward biased. Therefore, researchers ought to focus primarily on the p-values of the regression slope coefficients to conduct inference.

Our overall conclusion is that when we take adequate account of the potential nonlinearities in the return-yield relationship, evidence supportive of return predictability largely disappears. This result is consistent with the findings of Goetzmann and Jorion for US data, although their results were obtained within the framework of a linear model. The difference in the results using the nonparametric NN technique can be explained by the fact that local fitting is less sensitive to the presence of extreme observations compared with OLS.

One extension of the empirical analysis conducted here would be to reestimate the regression model with real and excess returns as the dependent variable in order to discover whether the conclusions reported in this paper are materially affected. Another important extension would be to explore the robustness of the WLS slope coefficient estimates to alternative weights.

Table 1

Monte Carlo Results
Standard Deviation Across 10000 regressions

Return Horizon (Months)	Overlapping Observations True Standard Deviation	Non-overlapping Observations True Standard Deviation
3	0.3653	0.3695
6	0.7140	0.7330
9	1.0422	1.1041
12	1.3501	1.4902
24	2.4220	3.0765
36	3.3696	4.4212
48	4.2406	8.6604

Table 2

Monte Carlo Results
Mean of Corrected Standard Errors
Overlapping Sample

	Return Horizon (Months)			
Alternative Covariance Matrices	12	24	36	48
TRUE	1.3501	2.4220	3.3696	4.2405
OLS	0.4058	0.5564	0.6620	0.7361
NWB[1]	0.9913	1.6844	2.1744	2.5317
NWB[2]	1.0384	1.6825	2.0412	2.2156
HAN[1]	1.1450	1.8230	2.2354	2.4468
HAN[2]	1.0328	1.5820	1.7711	2.0085
NWBA[3]	1.0194	1.6403	2.0168	2.3265
NWQS[1]	1.0737	1.7957	2.2936	2.6231
NWQSA[3]	1.2728	1.5918	1.8695	2.1292
AM[3]	1.1561	1.8360	2.4392	2.7790

Notes:

1. The bandwidth parameter, m is equal to the order of the moving average process

2. The bandwidth parameter, m is equal to twice the order of the moving average process.

3. The bandwidth parameter, m is chosen according to the method of Andrews (1991).

Table 3

OLS Regression of FTA All Share Total Returns on Dividend Yields
Sample Period: 1965-1992
Randomisation probability (p) values are based on 1000 shuffles

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta	R^2	Empirical p-value of R^2
Non-overlapping				
1	0.369	0.005	0.025	0.010
3	1.111	0.004	0.072	0.008
6	2.412	0.002	0.130	0.002
Joint	1.298	0.007		
Overlapping				
12	4.023	0.014	0.244	0.001
24	5.897	0.016	0.354	0.009
36	7.825	0.015	0.521	0.004
48	8.695	0.024	0.618	0.000
Joint	6.610	0.011		

Table 4

OLS Regression of FTA All Share Total Returns on Dividend Yields
Sample Period: 1965-1992
Stratified Randomisation probability (p) values are based on 1000 shuffles

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta	R^2	Empirical p-value of R^2
Non-overlapping				
1	0.369	0.023	0.025	0.044
3	1.111	0.015	0.072	0.032
6	2.412	0.016	0.130	0.014
Joint	1.298	0.022		
Overlapping				
12	4.023	0.042	0.244	0.014
24	5.897	0.033	0.354	0.016
36	7.825	0.022	0.521	0.033
48	8.695	0.048	0.618	0.011
Joint	6.610	0.039		

Table 5

WLS Regression of FTA All Share Total Returns on Dividend Yields
Sample Period: 1965-1992
WLS Randomisation probability (p) values are based on 1000 shuffles

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta	R^2	Empirical p-value of R^2
Non-overlapping				
1	0.294	0.060	0.017	0.316
3	0.481	0.141	0.027	0.194
6	1.212	0.086	0.072	0.099
Joint	0.662	0.092		
Overlapping				
12	2.793	0.034	0.157	0.041
24	5.244	0.029	0.324	0.020
36	6.805	0.039	0.491	0.004
48	8.106	0.046	0.638	0.000
Joint	5.737	0.039		

Table 6

NN Estimates of the Beta Coefficient
Sample Period: 1965-1992

NN Randomisation probability (p) values are based on 1000 shuffles

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta
	Non-overlapping	
1	0.175	0.403
3	0.553	0.352
6	1.063	0.419
Joint	2.597	0.459
	Overlapping	
12	2.061	0.404
24	5.253	0.052
36	6.946	0.181
48	9.647	0.128
Joint	5.977	0.378

Table 7

OLS Regression of FTA All Share Total Returns on Dividend Yields
Randomisation probability (p) values are based on 1000 shuffles

First Period 1965 - 1978

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta	R^2	Empirical p-value of R^2
1	0.229	0.015	0.039	0.016
3	0.712	0.012	0.118	0.023
6	1.577	0.001	0.180	0.036
Joint	0.839	0.031		
12	6.236	0.006	0.391	0.019
24	8.911	0.029	0.530	0.028
36	10.496	0.018	0.682	0.006
48	9.923	0.014	0.606	0.005
Joint	8.892	0.024		

Second Period 1979 - 1992

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta	R^2	Empirical p-value of R^2
1	0.185	0.090	0.075	0.088
3	0.554	0.140	0.180	0.079
6	1.062	0.189	0.331	0.075
Joint	0.600	0.121		
12	0.765	0.395	0.023	0.601
24	1.247	0.417	0.046	0.612
36	3.500	0.299	0.202	0.418
48	5.424	0.227	0.427	0.284
Joint	2.734	0.344		

Table 8
WLS Regression of FTA All Share Returns on Dividend Yields
WLS Randomisation probability (p) values are based on 1000 shuffles.

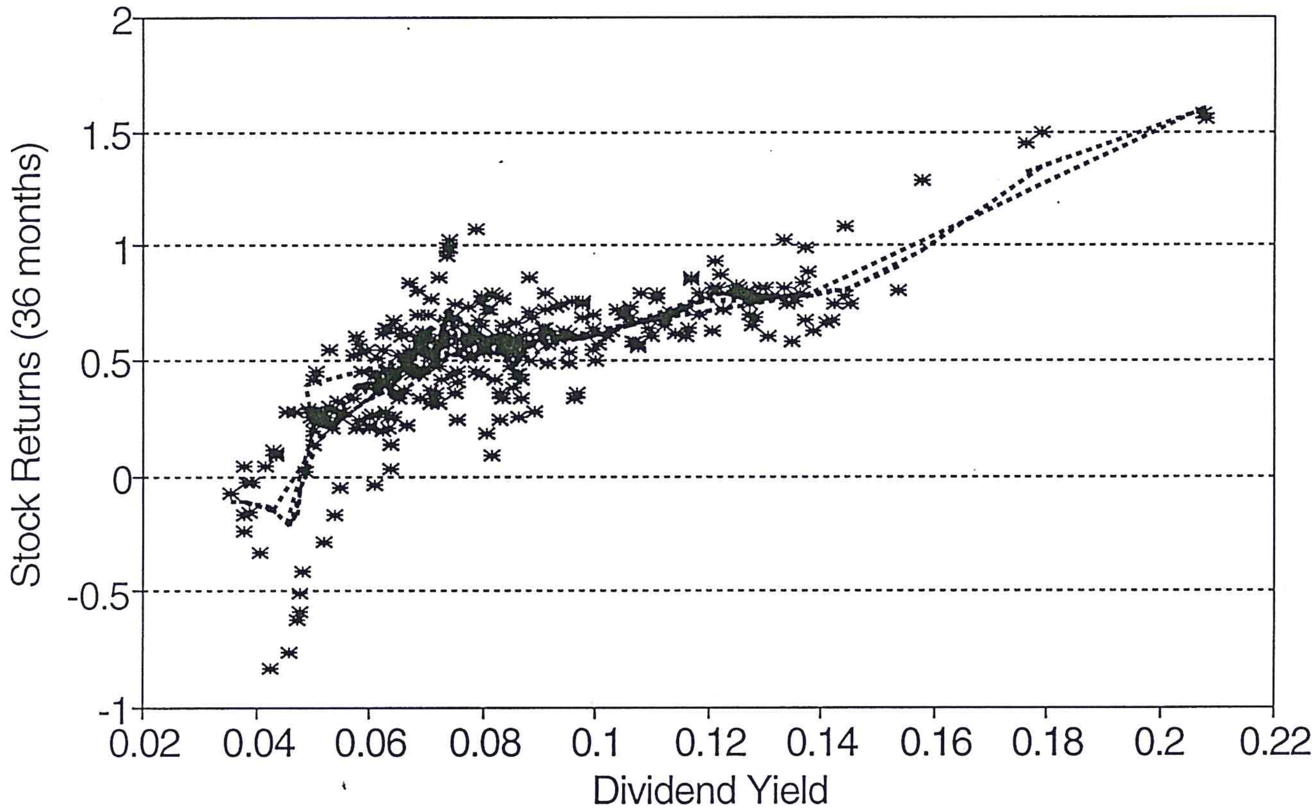
First Period 1965 - 1978

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta	R^2	Empirical p-value of R^2
1	0.331	0.152	0.014	0.528
3	0.382	0.349	0.003	0.796
6	1.400	0.192	0.031	0.565
Joint	0.704	0.241		
12	6.042	0.022	0.241	0.111
24	11.317	0.012	0.496	0.044
36	11.347	0.018	0.533	0.045
48	9.305	0.078	0.416	0.060
Joint	9.503	0.044		

Second Period 1979 - 1992

Return Horizon (Months)	Beta Coefficient	Empirical p-value of Beta	R^2	Empirical p-value of R^2
1	0.277	0.097	0.015	0.198
3	0.336	0.342	0.035	0.243
6	0.712	0.331	0.097	0.152
Joint	0.442	0.245		
12	0.329	0.511	0.154	0.175
24	0.858	0.500	0.306	0.143
36	3.083	0.345	0.425	0.165
48	5.013	0.240	0.621	0.093
Joint	2.321	0.388		

Figure 1: Stock Returns and Dividend Yields, 1968-1992



---- NN Fit

Figure 2: Empirical Distribution of
Thirty Six Month Beta Coefficient

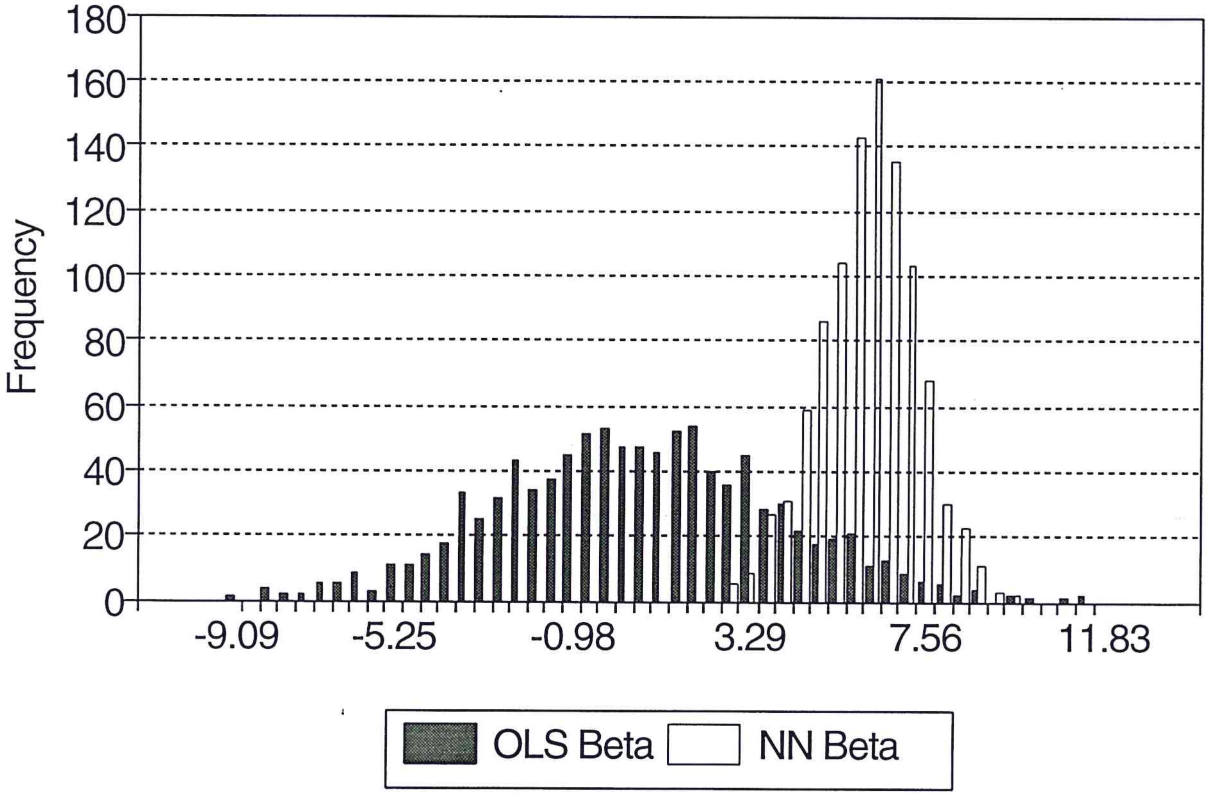
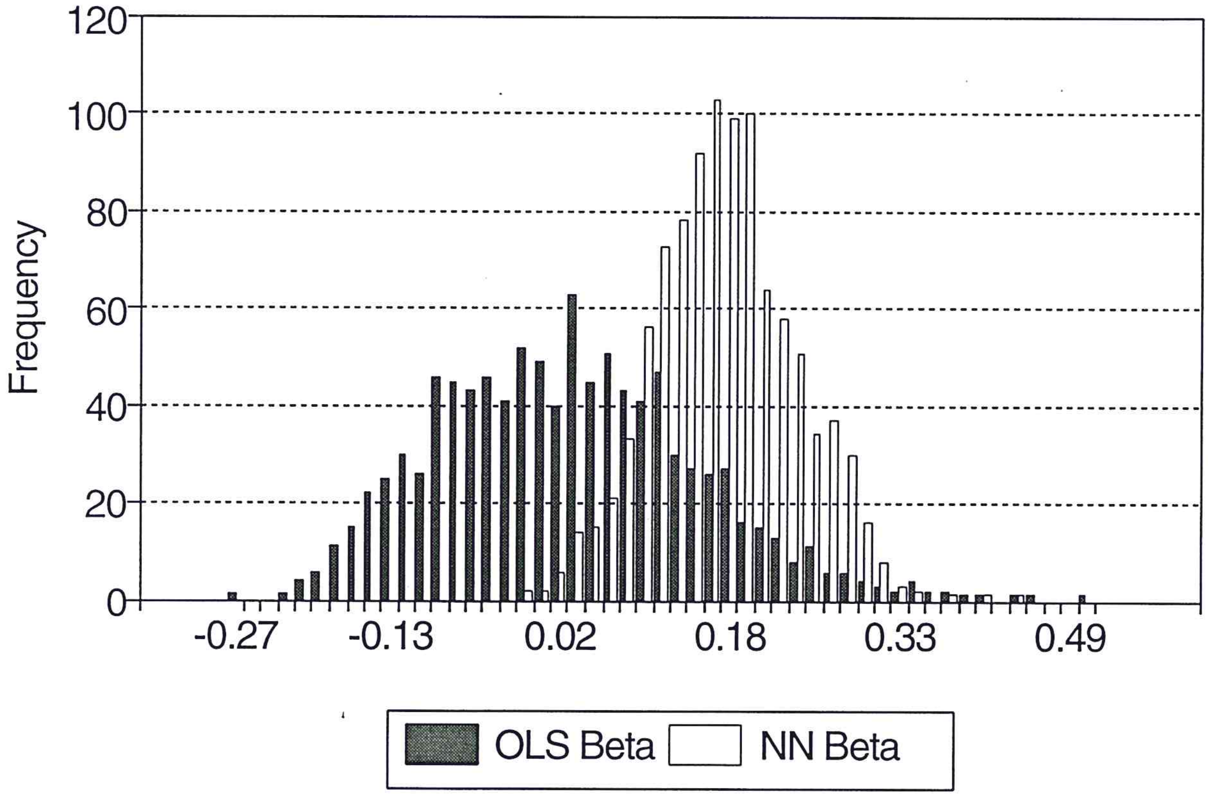


Figure 3: Empirical Distribution of
One Month Beta Coefficient



References

Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-858.

Andrews, D. W. K. and Monahan, J. C. (1992), An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica*, 60, 953-966.

Bachelier, L., 1900, Theorie de la Speculation, *Paris : Gauthier-Villars.*

Blume, M., 1980, Stock returns and dividend yields : Some more evidence, *Review of Economics and Statistics*, 62, 567-577.

Christie, W. G., and Huang, R. D., 1994, The Changing Functional Relation Between Stock Returns and Dividend Yields, *Journal of Empirical Finance*, 1, 161-191.

Cleveland, W. S. and Devlin, S. J., 1988, Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association*, 83, 596-610.

Cochran, S. J., DeFina, R. H., and Mills, L. O., 1993, International evidence on the predictability of stock returns, *Financial Review*, 28, 159-180.

Copeland, T. E. and Weston, J. F., 1988, Financial Theory and Corporate Policy, *Addison-Wesley Publishing Company.*

DeBondt, W. F. M. and Thaler, R. H., 1985, Does the stock market overreact, *Journal of Finance*, 793-805.

Diebold F. X. and Nason, J. A., 1990, Nonparametric Exchange Rate Prediction, *Journal of International Economics*, 28, 315-332.

Dow, C. H., 1920, Scientific stock speculation, *The Magazine of Wall Street (New York).*

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

Fama E. F., 1991, Efficient capital markets : II, *Journal of Finance*, 5, 1575-1617.

Fama, E. F. and French, K. R., 1988, Dividend yields and expected stock returns, *Journal of Financial Economics*, 22, 3-25.

Ferson, W. E. and Harvey, C. R., 1991, The variation of economic risk premiums, *Journal of Political Economy*, 99, 385-415.

Fisher, L., 1966, Some new stock market indexes, *Journal of Business*, 39, 191-225.

Fisher, R. A., 1935, The design of experiments, *New York : Hafner Publishing Co.*

Frankel, J. A. and Froot, K. A. (1987), Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations, *American Economic Review*, 77, 133-153.

French, K. R., Schwert, G. W. and Stambaugh, R. F., 1987, Expected Stock Returns and Volatility, *Journal of Financial Economics*, 19, 3-29.

Fuller, R. J and Kling, J. L., 1990, Is the stock market predictable?, *Journal of Portfolio Management*, 16, 28-36.

Goetzmann, W. N. and Jorion, P., 1993, Testing the predictive ability of dividends yields, *Journal of Finance*, 2, 663-678.

Gordon, R. H. and Bradford, 1980, Taxation and the stock market valuation of capital gains and dividends : Theory and empirical results, *Journal of Public Economics*, 14, 109-136.

Hansen, L. P. and Hodrick, R. J., 1980, Forward exchange rates as optimal predictors of future spot rates : An econometric analysis, *Journal of Political Economy*, 88, 829-853.

Harvey, C. R., 1991, The world price of covariance risk, *Journal of Finance*, 46, 111-157.

Hodrick, R. J., 1992, Dividend yields and expected stock returns : Alternative procedures for inference and measurement. *Review of Financial Studies*, 5, 357-386.

Jegadeesh, N., 1991, Seasonalities in stock price mean reversion : Evidence from the U.S. and the U.K., *Journal of Finance*, 4, 1427-1444.

Keim, D. B., 1985, Dividend Yields and Stock Returns : Implications of abnormal January returns, *Journal of Financial Economics*, 14, 473-89.

Keim, D. B. and Stambaugh, R. F., 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics*, 17, 357-390.

Kendall, M. G., 1953, The analysis of economic time series, part I : Prices, *Journal of the Royal Statistical Society,* 96, 11-25.

Kim, M. J., Nelson C. R., and Startz R., 1991, Mean Reversion in Stock prices? A Reappraisal of the Empirical Evidence, *Review of Economic Studies*, 58, 515-528.

Kleidon, A. W., 1986, Bubbles, fads and stock price volatility tests : A partial evaluation: Discussion, *Journal of Finance*, 43, 656-659.

LeRoy, S. F., 1973, Risk aversion and the martingale property of stock prices, *International Economic Review*, 14, 437-446.

Lindley, D. V. and Scott, W. F., New Cambridge elementary statistical tables, *Cambridge University Press.*

Lintner, J., 1956, Distribution of incomes of corporations among dividends, retained earnings, and taxes, *American Economic Review*, 46, 97-113.

Litzenberger, R. H. and Ramaswamy, K., 1979, Dividends, short selling, restrictions, tax induced investor clienteles and market equilibrium, *Journal of Finance*, 35, 469-482.

Lo, A. W. and MacKinlay, A. C., 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies*, 1, 41-66.

Maddala, G. S., 1992, Introduction to Econometrics, *Macmillan Publishing Company.*

Marsh, T. A. and Merton, R. C., 1984, Dividend variability and variance bounds tests for the rationality of stock market prices, *Working Paper No,1584-84, Cambridge, Sloan School of Management, M.I.T.*

McQueen, G., 1992, Long-horizon mean-reverting stock prices revisited, *Journal of Financial and Quantitative Analysis*, 1-18.

Miller, M. H. and Scholes, M. S., 1982, Dividends and taxes : Some empirical evidence, *Journal of Political Economy*, 90, 1118-1141.

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

Poterba, J. M., and Summers, 1988, L. H., Mean reversion in stock prices, *Journal of Financial Economics*, 22, 27-59.

Richardson, M., 1989, Temporary components of stock prices: a skeptic's view, Unpublished working paper, Stanford University.

Rozeff, M., 1984, Dividend yields are equity risk premiums, *Journal of Portfolio Management,* 68-75.

Shiller, R. J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity,* 2, 457-498.

Shiller, R. J., 1981, Do stock market prices move to much to be justified by subsequent changes in dividends? *American Economic Review*, 71, 421-436.

Summers, L. H., 1986, Does the stock market rationally reflect fundamental values?, *Journal of Finance*, 41, 591-601.