THE JOURNAL OF FINANCE • VOL. L, NO. 4 • SEPTEMBER 1995

Predictability of Stock Returns: Robustness and Economic Significance

M. HASHEM PESARAN AND ALLAN TIMMERMANN*

ABSTRACT

This article examines the robustness of the evidence on predictability of U.S. stock returns, and addresses the issue of whether this predictability could have been historically exploited by investors to earn profits in excess of a buy-and-hold strategy in the market index. We find that the predictive power of various economic factors over stock returns changes through time and tends to vary with the volatility of returns. The degree to which stock returns were predictable seemed quite low during the relatively calm markets in the 1960s, but increased to a level where, net of transaction costs, it could have been exploited by investors in the volatile markets of the 1970s.

MANY RECENT STUDIES CONCLUDE that stock returns can be predicted by means of publicly available information, such as time series data on financial and macroeconomic variables with an important business cycle component.[1] This conclusion seems to hold across international stock markets as well as over different time horizons. Variables identified by these studies to have been statistically important for predicting stock returns include interest rates, monetary growth rates, changes in industrial production, inflation rates, earnings-price ratios, and dividend yields. However, the economic interpretation of these results is controversial and far from evident. First, it is possible that the predictable components in stock returns reflect time-varying expected returns, in which case predictability of stock returns is, in principle, consistent with an efficient stock market. A second interpretation takes expected returns as roughly constant and regards predictability of stock returns as evidence of stock market inefficiency. It is, however, clear that predictability of excess returns on its own does not imply stock market inefficiency, and can be

* Pesaran is from Trinity College, Cambridge. Timmermann is from University of California, San Diego. This is a substantially revised and abridged version of the paper "The Use of Recursive Model Selection Strategies in Forecasting Stock Returns," *Department of Applied Economics Working paper No. 9406, March 1994,* University of Cambridge. We would like to thank a referee and the editor, René Stulz, as well as seminar participants at The London School of Economics, The University of Exeter, The Bank of England, University of Southern California, and University of California at San Diego for helpful comments on the earlier version. The first author gratefully acknowledges financial support from the Economic and Social Research Council and the Newton Trust of Trinity College, Cambridge.

[1] See, for instance, the articles by Balvers, Cosimano, and McDonald (1990), Breen, Glosten, and Jagannathan (1990), Campbell (1987), Cochrane (1991), Fama and French (1989), Ferson and Harvey (1993), French, Schwert, and Stambaugh (1987), Glosten, Jagannathan, and Runkle (1993), Pesaran and Timmermann (1994a).

interpreted only in conjunction with, and in relation to, an intertemporal equilibrium model of the economy. Inevitably, all theoretical attempts at interpretation of excess return predictability will be model-dependent, and hence inconclusive (see Fama (1991)).

An alternative approach to evaluating the economic significance of stock market predictability would be to see if the evidence could have been exploited successfully in investment strategies. This can be done in two ways: One method would be to evaluate the track records of portfolio managers in "real time," and see if these portfolios systematically generate excess returns. The main strength of this approach lies in the fact that it ensures that investors' portfolio decisions are based exclusively on historically available information. However, it does not provide much information on which specific factors have been responsible for predicting stock returns, nor does it guarantee that the information used by portfolio managers has been publicly available. An alternative approach, which explicitly addresses these issues, is to simulate investors' decisions in real time using publicly available information on a set of factors thought a priori to have been relevant to forecasting stock returns. Clearly, caution needs to be exercised when following this research strategy. In particular, it is important that, as far as possible, rules for prediction of stock returns are formulated and estimated without the benefit of hindsight. Most articles in the finance literature report excess return regressions estimated on the basis of the entire sample of available observations or on substantial subsamples of the data, which, for the purpose of trading, is clearly inappropriate, as in "real time" no investor could have obtained parameter estimates based on the entire sample. A similar consideration also applies to the choice of the forecasting model. Any analysis of stock market predictability that focuses on a particular forecasting model, taken as known with certainty over the whole sample period, can be criticized for ignoring the problem of "model uncertainty" and the impact this is likely to have on investors' portfolio strategy in "real time." When the same forecasting model is used over the whole sample period, it inevitably raises the possibility that the choice of the model could have been made with the benefit of hindsight.

The purpose of this article is to assess the economic significance of the predictability of U.S. stock returns, explicitly accounting for the forecasting uncertainty faced by investors who only have access to historical information. Rather than assuming that investors somehow historically knew that a specific forecasting model was going to perform well, we make a much weaker assumption about investors' beliefs over the sort of business cycle and financial variables thought as being potentially important in forecasting stock returns. Based on these beliefs, we assume that agents establish a base set of potential forecasting variables and, at each point in time, search for a reasonable model specification, capable of predicting stock returns, across this set. Notably, this procedure assumes that, at each point in time, investors use only historically available information to select a model according to a predefined model selection criterion and then use the chosen model to make one-period ahead predictions of excess returns. The recursive forecasts are then employed in a portfolio switching strategy according to which shares or bonds are held depending on whether excess returns on stocks are predicted to be positive or negative.

A third factor that needs to be taken into account when simulating investors' portfolio decisions in "real time" is transaction costs. We analyze portfolio returns for zero, low, and high transaction cost scenarios to shed light on whether the predictable components in stock returns are economically exploitable net of transaction costs. Finally, we also consider the public availability of statistical methods and computer technology used by investors to compute one-step ahead forecasts of excess returns. In developing our forecasting equations we use simple statistical and computing techniques that clearly were publicly available to any investor throughout the sample period analyzed in this paper. We also provide evidence based on forecasting techniques that were not available until well into the seventies, since this offers important insights into the potential usefulness of our methodology for forecasting of stock returns in the future. The availability of different model selection criteria also raises the issue of uncertainty over the choice of model selection criteria. This problem is addressed in an Appendix where a new profit-based hyper-selection procedure is proposed.

The plan of the article is as follows: Section I discusses how to identify predictability of stock returns. Section II sets up the real time simulation experiments and reports the main prediction results for the monthly observations on U.S. stock returns. The economic significance of the predictions is assessed through their use in a trading strategy in Section III, and Section IV provides a discussion of the main findings.

I. Identifying Predictability of Stock Returns: A Recursive Modeling Approach

Consider an investor who believes that stock returns can be predicted by means of a set of financial and macroeconomic indicators, but does not know the "true" form of the underlying specification, let alone the "true" parameter values. Under these circumstances the best the investor can do is to search for a suitable model specification among the set of models believed a priori to be capable of predicting stock returns. As time progresses and the historical observations available to investors increase, the added information is likely to lead the investor to change the forecasting equation unless, of course, the investor holds very strong prior beliefs in a specific model. Here we consider an open-minded investor with no strong beliefs in any particular model. The evolution of forecasting models over time may reflect the learning process of the investor or the changing nature of the underlying data generating process, or both. In practice it will be difficult to disentangle these two effects.[2]

[2] A similar recursive modelling strategy has also been considered by Phillips (1992) and Phillips and Ploberger (1994) in the context of univariate autoregressive-moving average processes. However, their work is best viewed as recursive estimation of the order of the Autoregressive Moving Average process, and does not involve searching over subset of regressors as we do in this paper. Also see the discussion of the hyper-selection criterion in the Appendix, where the recursive modelling strategy is further extended to allow for the uncertainty over the choice of model selection criterion itself.

Suppose that, at each point in time, t, an investor searches over a base set of κ factors or regressors to make one period ahead forecasts of excess returns using only information that are publicly available at the time.[3] We simulate investor's search for a forecasting model by applying standard statistical criteria for model selection, as well as financial criteria, to the set of regression models spanned by all possible permutations of the κ factors/regressors $\{x_1, x_2, \ldots, x_\kappa\}$ in the base set.[4] This gives a total of 2^κ different models, each of which is uniquely identified by a number, i, between 1 and 2^κ. Consider a $\kappa \times 1$ selection vector, ν_i, composed of ones and zeros where a one in its jth element means that the jth regressor is included in the model, whereas a zero in its jth element means that this regressor is excluded from the model. Then model i (denoted by M_i) can be represented by the κ-digit string of zeros and ones corresponding to the binary code of its number. Denoting the number of regressors included in model M_i by κ_i, then $\kappa_i = e'\nu_i$, where e is a $\kappa \times 1$ vector of ones. Suppose that ρ_τ, the excess return at time τ, is forecast by means of linear regressions

$$M_i : \rho_{\tau+1} = \beta_i' X_{\tau,i} + \varepsilon_{\tau+1,i} \qquad \tau = 1, 2, \ldots, t-1. \tag{1}$$

where $X_{\tau,i}$ is a $(\kappa_i + 1) \times 1$ vector of regressors under model M_i, obtained as a subset of the base set of regressors, X_τ, chosen by the investor at the beginning of the experiments, plus a vector of ones for the intercept term. Conditional on model M_i and given the observations $\rho_{\tau+1}, X_{\tau,i}, \tau = 1, 2, \ldots, t-1$ (with $t \geq \kappa + 2$), parameters of model M_i can be estimated by the ordinary least squares (OLS) method. Denoting these estimates by $\hat{\beta}_{t,i}$ we have

$$\hat{\beta}_{t,i} = \left(\sum_{\tau=0}^{t-1} X_{\tau,i} X'_{\tau,i} \right)^{-1} \sum_{\tau=0}^{t-1} X_{\tau,i}\rho_{\tau+1}, \qquad \begin{array}{ll} \text{for} & t = \kappa+2, \kappa+3, \ldots, T, \\ \text{and} & i = 1, \ldots, 2^\kappa. \end{array} \tag{2}$$

The OLS estimates are fairly simple to compute (even in the early 1960s) and, in view of the Gauss-Markov Theorem, are reasonably robust even in the presence of nonnormal errors in the excess return equation.

The particular choice of $X_{\tau,i}$ to be used in forecasting of $\rho_{\tau+1}$ can be based on a number of statistical model selection criteria suggested in the literature, such as the $\bar{R}^2$, Akaike's Information Criterion (AIC) (Akaike (1973)), or Schwarz's Bayesian Information Criterion (BIC) (Schwarz (1978)).[5] These

[3] In this article we shall make the simplifying assumption that the base regressors remain in effect over the whole sample period. However, in principle, one can also consider the possibility of revising the base set once clear indications of "regime switches" are established.

[4] An intercept term is included in all the excess return regressions considered by the investor.

[5] There are, of course, other criteria that could be used for choosing the subset of regressors used to compute the forecasts. Prominent examples are Mallows' (1973) C_p criterion, Amemiya's (1980) prediction criterion, and the Posterior Information Criterion recently proposed by Phillips (1992) and Phillips and Ploberger (1994). However, in this article we focus on the more often used and familiar model selection criteria.

criteria are likelihood-based and assign different weights to the "parsimony" and "fit" of the models. The "fit" is measured by the maximized value of the log-likelihood function ($\widehat{\mathrm{LL}}$), and the "parsimony" by the number of freely estimated coefficients. At time t, and under model M_i, we have

$$\widehat{\mathrm{LL}}_{t,i} = \frac{-t}{2}\{1 + \log(2\pi\hat{\sigma}^2_{t,i})\}, \tag{3}$$

where

$$\hat{\sigma}^2_{t,i} = \sum_{\tau=0}^{t-1} (\rho_{\tau+1} - X'_{\tau,i}\hat{\beta}_{t,i})^2/t. \tag{4}$$

The Akaike and Schwarz model selection criteria can be written as

$$\mathrm{AIC}_{t,i} = \widehat{\mathrm{LL}}_{t,i} - (\kappa_i + 1), \tag{5}$$

$$\mathrm{BIC}_{t,i} = \widehat{\mathrm{LL}}_{t,i} - \tfrac{1}{2}(\kappa_i + 1)\log(t). \tag{6}$$

The $\bar{\mathrm{R}}^2$ criterion, originally suggested by Theil (1958) as a criterion for selecting regressors in a linear regression model, is given by

$$\bar{R}^2_{t,i} = 1 - \frac{\tilde{\sigma}^2_{t,i}}{S^2_{\rho,t}} \tag{7}$$

where $\tilde{\sigma}^2_{t,i}$ is the unbiased estimator of σ^2 given by $\tilde{\sigma}^2_{t,i} = \sum_{\tau=0}^{t-1} (\rho_{\tau+1} - X'_{\tau,i}\hat{\beta}_{t,i})^2/(t - \kappa_i - 1)$, and $S^2_{\rho,t} = \sum_{\tau=1}^{t} (\rho_\tau - \bar{\rho}_t)^2/(t - 1)$ is the sample variance for the first t observations on ρ, and $\bar{\rho}_t = t^{-1}\sum_{\tau=1}^{t}\rho_\tau$. The $\bar{\mathrm{R}}^2$ criterion can also be written explicitly as a trade-off between fit and parsimony:

$$\mathrm{TC}_{t,i} = \widehat{\mathrm{LL}}_{t,i} - \tfrac{1}{2}\log\left(\frac{t}{t - k_i - 1}\right). \tag{8}$$

It is easy to show that, in the context of linear regression models, the $\bar{\mathrm{R}}^2_{t,i}$ and the $\mathrm{TC}_{t,i}$ criteria are equivalent, in the sense that they select the same model.

We also considered a model selection criterion based on a measure of directional accuracy, on the grounds that investors in practice often are interested in predicting the switches in the sign of the excess return function and not necessarily the magnitude of changes in the excess returns. The derivation of the 'sign' criterion (SC) based on the directional accuracy of the forecasts involves two steps. In the first step, one finds the set of regressors that maximize the proportion of correctly predicted signs of the excess returns given by

$$\mathrm{SC}_{t,i} = \frac{1}{t}\sum_{\tau=1}^{t}\{\mathrm{I}(\rho_\tau)\mathrm{I}(\hat{\rho}_{\tau,i}) + (1 - \mathrm{I}(\rho_\tau))(1 - \mathrm{I}(\hat{\rho}_{\tau,i}))\}, \tag{9}$$

where $I(\rho_\tau)$ is an indicator function that takes the value of unity if $\rho_\tau > 0$, and zero otherwise, and $\hat{\rho}_{\tau,i}$ is the forecast of ρ_τ based on model M_i. In the case of a draw, i.e., when two or more models correctly predict the same (maximum) proportion of signs of excess returns, a second step selects a model recursively according to the $\bar{R}^2$ criterion.

From the point of view of assessing the market value of predictability of stock returns, the above statistical criteria can, however, be criticized on the grounds that they do not take account of transaction costs, and are not necessarily in accordance with the investor's loss function. To deal with these shortcomings, we use a forecasting strategy that directly maximizes financial criteria. In particular we consider a "recursive wealth" criterion and a recursive Sharpe ratio.[6] The recursive wealth criterion maximizes the cumulated wealth obtained using forecasts from model M_i in a switching portfolio, which we refer to as portfolio i, at time t. The cumulative wealth from such a portfolio at time t is given by

$$W_{t,i} = W_0 \prod_{\tau=1}^{t} (1 + r_{\tau,i}), \tag{10}$$

where $r_{\tau,i}$ is the period τ return (net of transaction costs) on portfolio i, constructed on the basis of the excess return forecasts from model M_i. The returns $r_{\tau,i}$ depend on the nature of the trading rule, the transaction costs and the whole sequence of excess return forecasts, risk free interest rates, stock prices and dividends prior to period τ. For an example of the sort of trading rule that we have in mind, see Section III.

The Recursive Sharpe Criterion maximizes the ratio of the mean excess return on portfolio i to its standard deviation:

$$\text{Sharpe}_{t,i} = \frac{\frac{1}{t}\sum_{\tau=1}^{t}(r_{\tau,i} - I1_{\tau-1})}{\sqrt{\frac{1}{t-1}\sum_{\tau=1}^{t}(r_{\tau,i} - \bar{r}_{t,i})^2}}, \tag{11}$$

where I1 is the return on a 1-month T-bill held from the end of one month to the next, and $\bar{r}_{t,i} = t^{-1}\sum_{\tau=1}^{t} r_{\tau,i}$.

We applied all the above model selection criteria to the linear regressions. For each of the criteria, and based on data up to period t, the model with the highest value for the criterion function was chosen to forecast excess returns for period $t + 1$. For example, in the case of the recursive Sharpe criterion, the selected model was that which maximized $\text{Sharpe}_{t,i}$ given by (11) over all of the 2^k portfolios, $i = 1, 2, 3, \ldots, 2^k$. Notice that our approach makes only very

[6] We are grateful to a referee for drawing our attention to this work.

weak assumptions about the underlying data generating process (DGP). We do not assume that the DGP is necessarily fixed throughout the sample period, and at each point in time we use the different model selection criteria simply as a tool for obtaining an *approximate* forecasting equation. This procedure treats all models under consideration as equally likely. Choosing a particular model at time t does not necessarily restrict the model choice at subsequent periods.

When interpreting the results reported in Sections II and III below, however, it is important to recognize that neither the AIC nor the BIC was publicly available until well into the 1970s. The main property of the BIC criterion is that, under certain regularity conditions, it will asymptotically select the "true" model, provided of course that the 'true' model is contained in the set of models over which the search is conducted.[7] Neither Akaike's criterion nor the $\bar{R}^2$ criterion is consistent, in the sense of selecting the "true" model, as the sample size increases without bounds. In the context of forecasting stock returns where the "true" model or the correct list of regressors is clearly unknown and may be changing over time, the consistency property of a model selection criterion is not as important as it may appear at first. Both the $\bar{R}^2$ and the AIC, although statistically inconsistent, have the important property of yielding an approximate model. The primary aim is to select a forecasting equation that could be viewed *at the time* as being a reasonable approximation to the data generating process. Akaike's criterion has been shown by Shibata (1976) to strike a good balance between giving biased estimates when the order of the model is too low, and the risk of increasing the variance when too many regressors are included. Finally, for the purpose of our exercise, the $\bar{R}^2$ has the advantage of being extensively used by economists to evaluate model performance.

II. Recursive Predictions of U.S. Stock Returns

When simulating the historical process through which an investor may attempt to forecast stock returns, it is important to establish the sort of variables the investor is likely to consider using in modeling stock returns, the criteria he/she adopts to select a particular forecasting model, and the estimation procedure applied.

In this section we explain the choice of the base variables that we assume will be considered by investors in forecasting stock returns, and in the next section we explain the estimation and forecasting procedure in more detail. The starting point of our analysis is the long tradition in finance that links movements in stock returns to business cycle indicators. For instance, in his book *Investment for Appreciation. Forecasting Movements in Security Prices. Techniques of Trading in Shares for Profit* published in 1936, Angas writes that "The major determinant of price movements on the stock exchange is the business cycle." (p. 15). Other examples of early studies that emphasize the

[7] See, for example, Pötscher (1991) and the references cited therein.

systematic variation of stock returns over the business cycle include Prime (1946). Dowrie and Fuller (1950), Rose (1960), and Morgan and Thomas (1962). Variables suggested by these studies to be systematically linked with stock returns include short and long interest rates, dividend yields, industrial production, company earnings, liquidity measures, and the inflation rate.

Based on a review of the early literature (see Pesaran and Timmermann (1994c)) we established a benchmark set of regressors over which the search for a "satisfactory" prediction model could be conducted by a potential investor. The set consists of a constant, which is always included in the model, as well as nine regressors, namely $X_t = \{YSP_{t-1}, EP_{t-1}, I1_{t-1}, I1_{t-2}, I12_{t-1}, I12_{t-2}, \Pi_{t-2}, \Delta IP_{t-2}, \Delta M_{t-2}\}$, where YSP is the dividend yield, EP is the earnings-price ratio, I1 is the 1-month T-bill rate, I12 is the 12-month T-bond rate, Π is the year-on-year rate of inflation, ΔIP is the year-on-year rate of change in industrial output, and ΔM is the year-on-year growth rate in the narrow money stock. All variables computed using macroeconomic indicators, such as, ΔIP and ΔM, were measured using 12-month moving averages to decrease the impact of historical data revisions on the results.

The early studies of stock returns are not always clear on what they consider to be the appropriate time lags between the changes in the business cycle variables and stock returns. Here, following standard practice in finance, we decided to include the most recently available values of the macroeconomic variables in the base set of regressors. The lag associated with the publication of macroeconomic indicators means that these variables must be included in the base set with a 2-month time lag. Since the dividend and earning yields are based on 12-month moving averages, only a one period lag of these variables was included in the base set. To allow for the possibility, often mentioned in financial studies, that changes in interest rates rather than their absolute levels affect stock returns, we also included a two month as well as a one month lagged value of the interest variables.

A. Data Sources

All variables were measured at monthly frequencies over the period 1954(1) to 1992(12), and the data sources were as follows: Stock prices were measured by the Standard & Poor's 500 index at close on the last trading day of each month. These stock indices, as well as a monthly average of annualized dividends and earnings, were taken from Standard & Poor's Statistical Service. The 1-month T-bill rate was measured on the last trading day of the month and computed as the average of the bid and ask yields. The source was the Fama-Bliss risk free rates file on the Center for Research in Security Prices (CRSP) tapes. Similarly, the 12-month discount bond rate was measured on the last trading day of the month, using the Fama-Bliss discount bonds file on the CRSP tapes as the data source. The inflation rate was computed using the producer price index for finished goods (source: Citibase), and the rate of change in industrial production was based on a seasonally adjusted index for industrial production (source: Citibase). The monetary series were based on the narrow monetary aggregates published by the Federal Reserve Bank of St. Louis and provided by Citibase. Finally, the dependent variable, excess returns on stocks, ρ_t, was computed as $\rho_t = (P_t + D_t - P_{t-1})/P_{t-1} - I1_{t-1}$, where P_t is the stock price, D_t is dividends and $I1_{t-1}$ is the return from holding a 1-month T-bill from the end of month $t - 1$ to the end of month t.

The recursive model selection and estimation strategy was based on monthly observations over the period 1954(1) to 1992(12). The year 1954 was chosen as the start of the sample for estimation, since reliable monthly measures for most macroeconomic time series start to become available only after the Second World War. Also, it was only after the "accord" between the Fed and the Treasury in March 1951, and after the presidential election in 1952, that the Fed stopped pegging interest rates and began to pursue an independent monetary policy (see Mishkin (1992), p. 453). As far as trading in stocks and bonds are concerned, we took a rather conservative stand and commenced with the trading at the start of 1960, thus using 6 years of monthly observations as a preliminary "training" period for estimation. As noted above, by the early 1960s, a number of studies had already suggested the possibility that stock returns may be varying systematically over the course of the business cycle.

In each case the model selection criteria set out in Section I were applied to linear regression models using the excess returns on the S & P 500 portfolio as the dependent variable and subsets of the base set of regressors as the independent variables. For our set of nine regressors, this means comparing $2^9 = 512$ models at each point in time, and over the period 1959(12) to 1992(11) this gives a total of 202,752 regressions to be computed. Of course, we do not literally assume that investors proceeded with these computations, but we emphasize the simplicity of the individual steps involved in such a forecasting procedure: OLS estimation of the models, followed by model selection using simple choice criteria and then computation of a one-step-ahead forecast. We also computed forecasts of monthly excess returns based on a model that included the entire set of regressors. There is no specification uncertainty associated with this last procedure, according to which only the parameter estimates are updated recursively in light of new monthly observations.

To summarize, the recursive model specification proceeds as follows: in 1959(12) the values of the selection criteria are computed for each of the 512 possible combinations of regressors from the base set using monthly data over the period 1954(1) to 1959(12). An intercept term is included in all the regressions. The model that maximizes the discriminant function of a given model selection criterion is chosen, and the parameter values estimated with observations over the 1954(1) to 1959(12) period are used to forecast excess returns for 1960(1). To forecast monthly excess returns for 1960(2), the procedure is repeated for all the 512 models using monthly data over the period 1954(1) to 1960(1), and so on. Thus, although computationally demanding, our selection procedure clearly simulates the search procedure which an investor could have carried out in real time. It also captures the possibility that an investor may switch from one model to another in light of new empirical evidence obtained as the sample size expands.

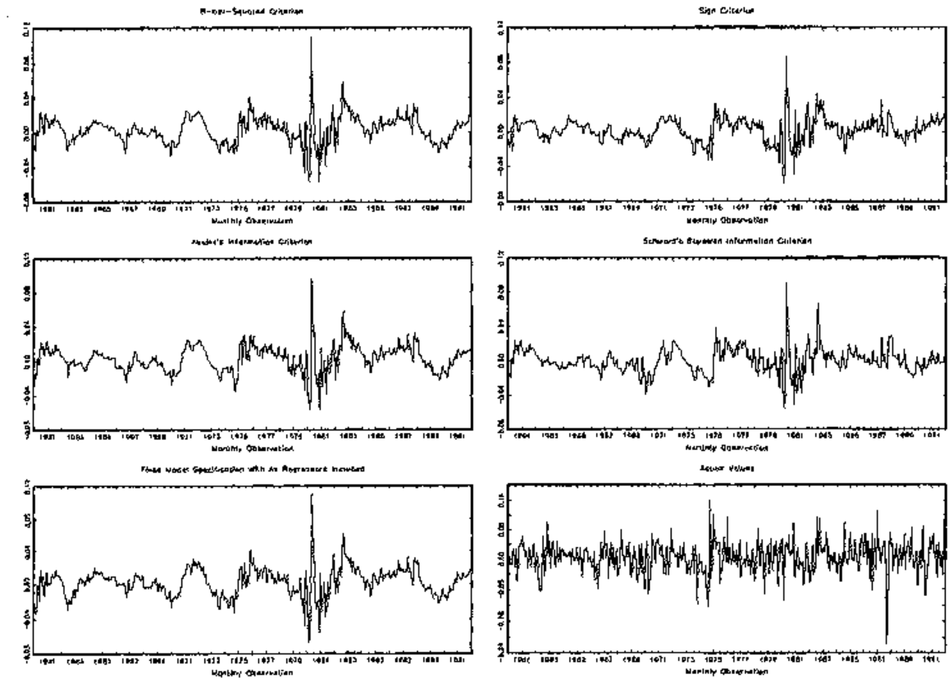

Figure 1. Recursive excess return forecasts under alternative model selection strategies, 1960(1) to 1992(12).

B. Empirical Results: How Robust is the Predictability of Stock Returns?

As the previous section makes clear, we estimated a total of 202,752 models over the 1959(12) to 1992(1) period. Clearly, we cannot supply the reader with all the details of the estimation results.[8] Here we provide some graphic displays of the main results.

Figure 1 shows the predicted excess returns based on the linear regression models estimated by OLS and selected recursively according to the four model selection criteria, namely $\bar{R}^2$, AIC, BIC, and SC. The bottom panel of this figure also shows the actual values of the excess returns and predictions from a recursively estimated equation with all the nine variables in the base set included as regressors.[9] The recursive predictions based on the various model selection criteria have very similar patterns showing quite a high degree of volatility, especially during the early 1980s. This coincides with the period of high volatility of nominal interest rates resulting from changes in the operat-

[8] The data files and the details of the estimations and forecasting results are available from the authors on request.

[9] Notice, however, that the figures giving the recursive forecasts have a different vertical scale than the actual values of the excess returns. Due to the relatively high variance of actual excess returns plotting the excess return forecasts and their realizations on the same scale would have obscured the differences that exist between the different recursive forecasts.

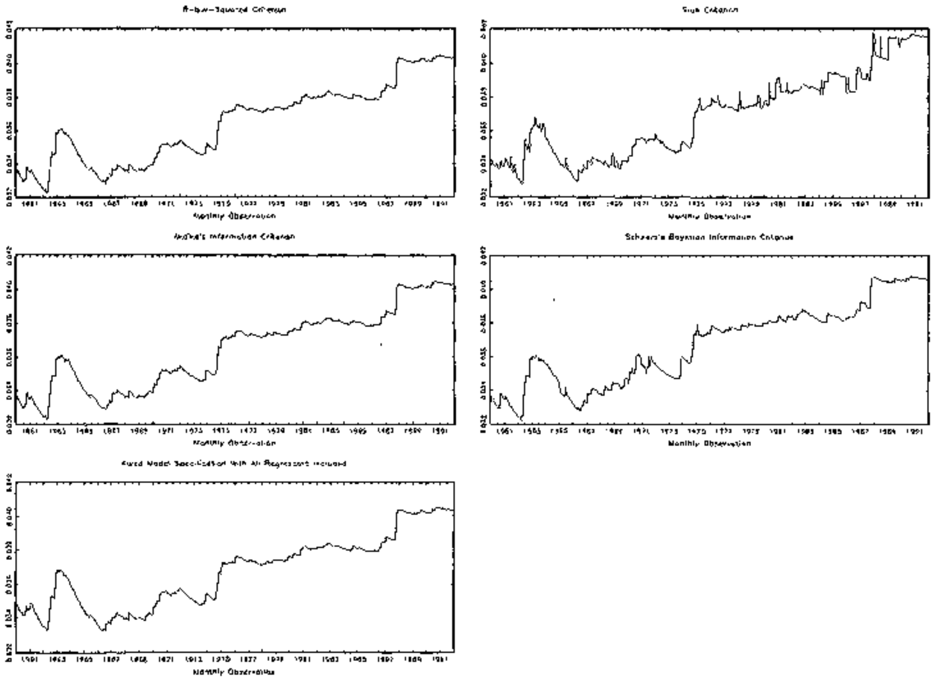

Figure 2. Standard errors of recursive excess return equations under alternative model selection strategies, 1960(1) to 1992(12).

ing procedures followed by the Federal Reserve between September 1979 and October 1982. Not surprisingly, the volatility of the predictions is much smaller than the volatility of the actual excess returns.

Figure 2 shows the standard errors of recursive excess return regressions under the different model selection strategies. In all cases, the recursively estimated standard errors have a tendency to increase over time. In particular, substantial increases in the estimated standard errors can be clearly seen in 1962, 1974, and after the October 1987 stock market crash.[10] Important information on the ex ante forecasting performance of the different model selection strategies is also provided in Figure 3, which displays the recursively computed squared values of the simple correlation coefficient (r^2) between the recursive forecasts obtained under the different model selection criteria and the actual excess returns. The fit of the recursive forecasts is relatively high in the early 1960s (with values of the r^2 being around 0.20), and increases substantially during 1962, but then starts to decline until the early 1970s. With increased volatility of the markets in 1974, the fit of the recursive

[10] The recursive standard error estimates in Figure 2 show the trend in volatility of excess returns conditional on the information in the base set of regressors. Similar trends can also be seen in unconditional measures of volatility, such as the recursively estimated standard deviations of the actual excess return series.

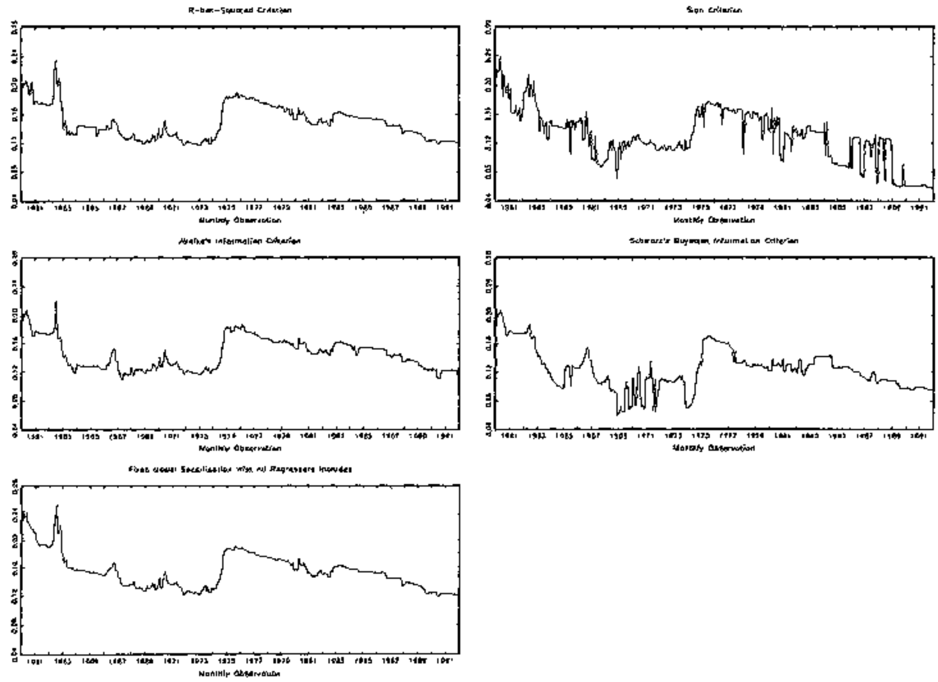

Figure 3. Fit of recursive excess return equations, 1960(1) to 1992(12). Note: The fit is measured by the recursively computed squared correlation coefficient between the recursive forecasts (from Figure 1) and the actual values of the excess returns.

forecasts, as measured by r^2, jumps from around 0.12 to around 0.18, and then once again starts to decline steadily up to the end of the sample, where it takes a value of around 0.12 for models selected according to the $\bar{R}^2$ and the Akaike Criteria, and 0.10 for models selected using the Schwarz criterion.

Taken together, Figures 2 and 3 show that although the volatility in the U.S. stock market increased significantly around 1974, the predictability of stock returns, as measured by the fit of the recursive forecasts relative to the actual values, also increased in this period. This suggests that predictability of stock returns may indeed be particularly pronounced in periods of economic "regime switches" where the markets are relatively unsettled and investors are particularly uncertain of which forecasting model to use for trading. The episode early in 1962 where the volatility in the stock market also went up leads to a similar conclusion. In contrast, the significant increase in the standard errors of the recursive forecasting equations after the October 1987 stock market crash coincided with a decline in the value of r^2 for all the model selection criteria, indicating that this episode was somewhat different from the ones in 1962 and 1974.

For each of the variables in the base set of regressors Table I presents the percentage of periods where a regressor is included in the recursively selected models. As to be expected, on balance the $\bar{R}^2$ and the Akaike criteria tend to

Table I

Percentage of Periods Where a Regressor is Included in Forecasting Equations

The results are based on monthly excess return equations selected and estimated recursively over the period 1960(1) to 1992(12). Each month the set of regressors that maximizes a given model selection criterion was determined and used to forecast stock returns one month ahead. For a definition of the statistical model selection criteria, see Section II of the article. The regressors are YSP(−1) = dividend yield lagged one period, EP(−1) = earnings-price ratio lagged one period I1(−1) = one-month T-bill rate lagged one period, I1(−2) = one-month T-bill rate lagged two periods, I12(−1) = twelve-month T-bill rate lagged one period, I12(−2) = twelve-month T-bill rate lagged two periods, Π(−2) = inflation rate lagged two periods, ΔIP(−2) = change in industrial production lagged two periods, and ΔM(−2) = monetary growth rate lagged two periods.

Selection	Percentages								
Criteria	YSP(−1)	EP(−1)	I1(−1)	I1(−2)	I12(−1)	I12(−2)	Π(−2)	ΔIP(−2)	ΔM(−2)
Akaike	70.5	9.6	98.5	25.0	27.0	31.3	51.8	74.7	84.6
Schwarz	62.9	0.0	93.9	22.0	2.8	34.8	0.0	8.8	28.3
$\bar{R}^2$	69.7	20.5	99.2	23.0	44.7	42.9	55.8	87.9	89.6
Sign	62.4	44.2	72.5	54.8	49.7	50.0	59.6	58.6	76.5

select more regressors than does the Schwarz criterion. The latter criterion imposes a much heavier penalty for inclusion of an additional regressor than do the former criteria, and this difference becomes particularly marked as the sample size is increased from 72 to 468 observations over the 1960 to 1992 period.

An alternative, and in many respects more comprehensive, method of examining the robustness of the factors contributing to the predictability of stock returns would be to consider the time-profile of the inclusion frequencies of the different factors in the forecasting model. The graphs in Figure 4 provide such time-profiles for the $\bar{R}^2$ criterion, and show the periods in which a regressor is included (the graph takes a value of 1) or excluded (the graph takes a value of 0) from the model used for the one-period ahead prediction of excess returns. Similar graphs for the other model selection criteria are also available from the authors, and are not presented here to save space.[11] Since we are searching for a model specification over a large number of time periods and across different combinations of regressors, it is possible that a regressor occasionally gets included by chance and not because it is statistically significant for prediction of excess returns. In such cases, however, it is unlikely that the regressor under consideration will be included in the forecasting equation for long in subsequent periods. In contrast, when a regressor is selected in a large proportion of the time periods and on a continuous basis, then it is reasonable to

[11] The results for the Akaike criterion were very close to those for the $\bar{R}^2$ criterion. The Sign criterion did not provide much information about periods in which specific regressors are included. Because of the discontinuous nature of this criterion, the number of switches between periods where a variable is included in the model and periods where it is excluded from the model is much higher for the Sign criterion than for any of the other model selection criteria.

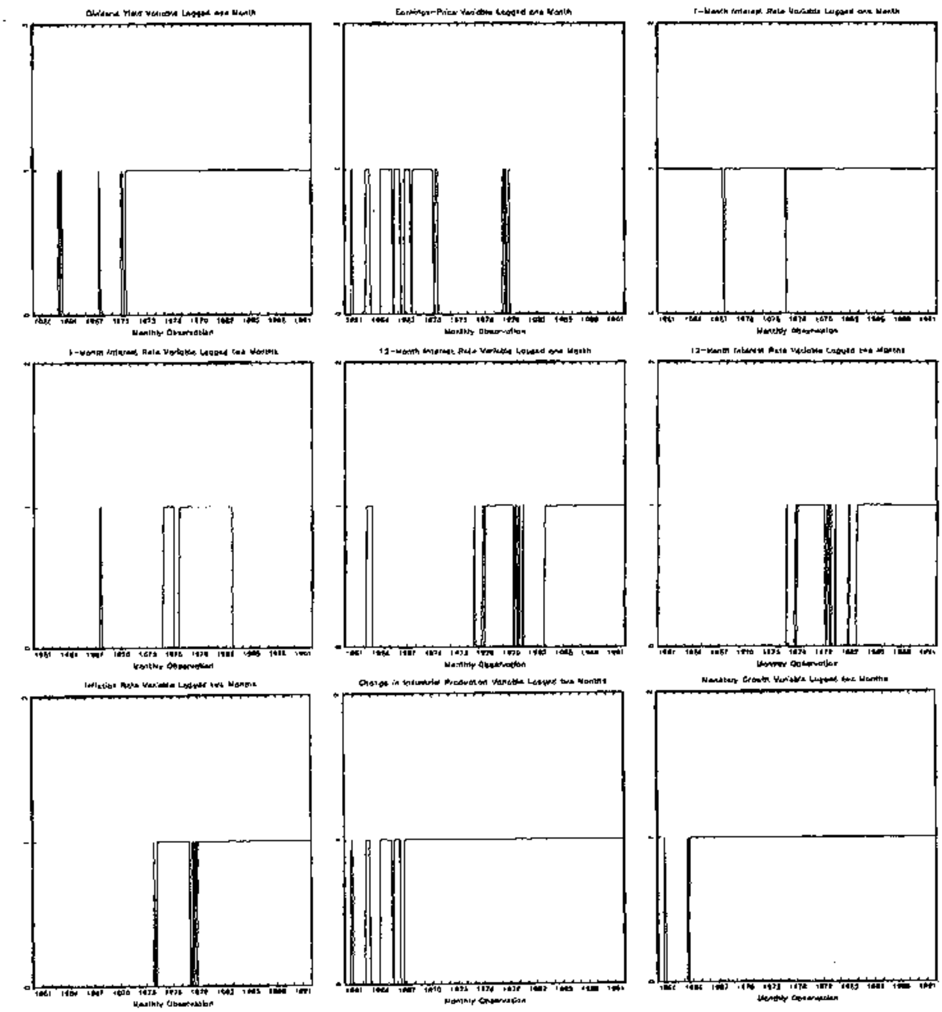

Figure 4. Inclusion frequency of the variables in the base set under the $\bar{R}^2$ model selection criterion, 1960(1) to 1992(12). Note: The inclusion of the variable in the regression model is depicted by unity, and zero otherwise.

expect the regressor in question to be an important factor in generating the observed predictability of stock returns.

Figure 4 shows that the dividend yield variable is selected in most periods from 1970 onwards. This finding supports the many recent studies that find the yield variable to be statistically significant in predicting stock returns (e.g., Campbell (1987), Fama and French (1989)). Our results, however, suggest that the correlation between yields and excess returns has become particularly strong only after 1970.[12] Compared to the dividend yield variable, the earnings-price ratio lagged one month is not selected as often in the forecasting models; in fact, this regressor is never chosen by the Schwarz criterion.

The 1-month interest rate variable lagged one month is excluded in the forecasting models in most periods (see Figure 4), whereas the two-month lagged value of the 1-month interest rate, $I1_{t-2}$, is included as a regressor in the forecasting models mainly during the 1975 to 1982 period. In the case of the longer interest rate lagged one month, $I12_{t-1}$, it is clear from Figure 4 that this variable is selected by the $\bar{R}^2$ criterion (but not by the Schwarz criterion), during the 1976 to 1979 periods, and again after 1982. This is an interesting finding since the periods when the $I12_{t-1}$ variable is not included in the forecasting model coincide closely with the period from October 1979 to September 1982, when the Federal Reserve ceased to target interest rates. It suggests that the 12-month interest rate has predictive power over excess returns in regimes where the Federal Reserve targets interest rates, but not in regimes where monetary aggregates are targeted. A similar conclusion also emerges with respect to the 12-month interest rate lagged two months, $I12_{t-2}$.

Using the $\bar{R}^2$ model selection criterion, the inflation variable lagged two periods (Π_{t-2}) is included in the forecasting equations primarily after the first oil shock. In contrast, the Schwarz criterion does not select this variable at all. Figure 4 also shows that the rate of change in industrial production lagged two periods (ΔIP_{t-2}) tends to be included in the forecasting models from 1964 onwards, when the $\bar{R}^2$ model selection criterion is adopted. Once again, by comparison the Schwarz criterion rarely selects the ΔIP_{t-2} variable. Finally, using the $\bar{R}^2$ criterion, the money growth variable, ΔM_{t-2}, gets selected continuously from 1964 onwards, but the same is not true when the Schwarz criterion is considered, which selects the money growth variable 28 percent of the times.

The different outcomes obtained under the $\bar{R}^2$ and the Akaike Criteria on the one hand, and under the Schwarz criterion on the other, lies in the fact that the Schwarz criterion penalizes inclusion of variables more heavily than either the Akaike or the $\bar{R}^2$ criteria. This may be a drawback for the Schwarz criterion since, in the event of a structural break in the underlying data generating process, this criterion may detect the change at a slower rate than the other criteria.

In a recent study, Bossaerts and Hillion (1994)[13] also find evidence supporting the conclusion that the best prediction model for monthly stock returns changes over time. Bossaerts and Hillion compare the in-sample and out-of-sample forecasting performance of models selected according to a variety of standard model selection criteria and develop a new model selection criterion

[12] On this point see also the recent papers by Goetzmann and Jorion (1993) and Nelson and Kim (1993), who find that, after correcting for lagged endogenous variable bias, the correlation between lagged values of the dividend yield and excess returns on stocks is particularly strong after the Second World War.

[13] We are grateful to an anonymous referee for bringing this research to our attention.

that can be used to detect nonstationarities in the underlying data generating process. Based on an examination of the out-of-sample forecasts of monthly returns in 15 countries over the period 1989 to 1993, Bossaerts and Hillion conclude that there is strong evidence of nonstationarities in stock returns in the U.S. and several other markets. Notice that an important difference between our study and theirs is that we explicitly allow the forecasting model to change through time.

C. The Market Timing Value of The Recursive Forecasts

In a recent study, Leitch and Tanner (1991) found that traditional measures of forecasting performance, such as the $\bar{R}^2$, were not as strongly correlated with profits from a trading strategy based on a set of predictions as were a measure of the directional accuracy (e.g., the proportion of times the sign of excess returns is correctly predicted) of the forecasts. In Pesaran and Timmermann (1992, 1994b) we have developed a new market timing test statistic that is based on the directional accuracy of the forecasts and hence may provide important information on the economic value of the recursive forecasts. In the present case with only two directions, our test statistic is asymptotically equivalent to the Henriksson and Merton (1981) test. (Pesaran and Timmermann (1994b)).[14] Values of this market timing test statistic, which are asymptotically normally distributed, are reported in Table II. For a one-sided test of the null of no market timing against the alternative of market timing skills, the 5 percent critical value is 1.64. Using the whole sample, 1960(1) to 1992(12), the null hypothesis is strongly rejected for the predictions based on models recursively selected according to all criteria except for the Schwarz criterion. Predictability seems particularly high during the 1970s and is not statistically significant during the other two decades. This suggests that the forecasting performance is not primarily a function of the length of the "learning period" of the models, as one might have suspected if the underlying data-generating process had remained constant over time.

The percentage of correctly predicted signs of the excess returns also conveys important information to an investor. Once again, with the exception of the Schwarz criterion, the recursive predictions get the sign of the excess returns right in at least 58 percent of all months over the 1960 to 1992 period. The results over the three subperiods, 1960 to 1969, 1970 to 1979 and 1980 to 1989, also show the substantially higher proportion of correct signs achieved by all the recursive forecasts over the 1970s as compared to the other two subperiods.

III. Assessing the Economic Significance of Predictability of Excess Returns in a Simple Trading Strategy

In the finance literature, the efficient market hypothesis is often interpreted as the impossibility of constructing a trading rule, based on publicly available

[14] The largest deviation between the values of our test statistic and the values of the nonparametric Henriksson-Merton test statistic reported in Table II was 0.03.

Table II

Predictive Accuracy of Excess Return Forecasts over 1960(1) to 1992(12) and Three Sub-Periods

The *PT* statistic is the nonparametric test statistic for market timing proposed in Pesaran and Timmermann (1992). This test, which is asymptotically equivalent to the Henriksson-Merton (1981) test of market-timing, has a standardized normal distribution in large samples. All selection criteria were applied recursively (see Section II). The hyper-selection criterion is described in the Appendix. The recursive wealth procedure uses as a forecasting model the one that, at each point in time, generated the largest wealth (net of transaction costs) when its forecasts were used in a simple trading strategy. The recursive Sharpe criterion is based on a similar procedure with the difference that it is now the Sharpe ratio corresponding to trading results based on the set of predictions generated by a particular regression model that is being maximized recursively.

	1960 to 1992		1960 to 1969		1970 to 1979		1980 to 1989	
Selection Criteria	PT-Statistics	Proportion of Correct Signs	PT-Statistics	Proportion of Correct Signs	PT-Statistics	Proportion of Correct Signs	PT-Statistics	Proportion of Correct Signs
Panel A: Statistical Selection Criteria								
$\bar{R}^2$	3.39*	59.6	1.64†	58.3	2.64*	61.7	0.76	56.7
Akaike	2.88*	58.3	1.33	56.7	2.67*	61.7	0.43	55.0
Schwarz	1.34	53.3	−1.20	43.3	2.39*	60.8	0.73	54.2
Sign	2.93*	58.3	1.92†	59.2	1.88†	58.3	0.79	55.8
All Regressors	3.41*	59.6	1.14	56.7	2.81*	62.5	1.11	57.5
Panel B: Hyper-Selection Procedure								
Transaction Costs								
Zero	2.88*	58.1	1.92†	59.2	1.88†	58.3	0.85	55.8
Medium	3.06*	58.6	1.92†	59.2	1.71†	57.5	0.95	56.7
High	3.06*	58.6	1.92†	59.2	1.51	56.7	1.11	57.5
Panel C: Financial Criteria								
Recursive Wealth Procedure								
Transaction Costs								
Zero	3.73*	60.4	0.98	55.8	2.64*	61.7	1.67†	60.0
Medium	3.51*	59.6	1.39	56.7	2.25*	60.0	1.51	59.2
High	3.27*	58.8	1.39	56.7	1.88†	58.3	1.58	58.3
Recursive Sharpe Procedure								
Transaction Costs								
Zero	3.71*	60.4	1.29	57.5	2.62*	61.7	1.90†	60.8
Medium	3.14*	58.8	1.14	56.7	2.62*	61.7	1.07	56.7
High	3.31*	59.1	0.98	55.8	2.62*	61.7	1.20	56.7

* Indicates statistically significant evidence of market timing at the 10 percent level.

† Indicates statistically significant evidence of market timing at the 1 percent level.

information, that is capable of yielding positive excess profits (discounted at an appropriate risk-adjusted rate). Jensen (1978) puts it this way: "A market is efficient with respect to information set Ω_t if it is impossible to make economic profits by trading on the basis of information set Ω_t."

Predictability of stock returns in itself does not, however, guarantee that an investor can earn profits from a trading strategy based on such forecasts. First, monthly excess returns on stocks do not follow a standard distribution but are considerably more leptokurtic than, say, the normal distribution. Standard measures of predictive performance, such as the $\bar{R}^2$, may not be reliable in terms of indicating opportunities for profit-making. Secondly, transaction costs may erode the profits from trading in the financial markets based on recursive forecasts of excess returns. Compared to the natural benchmark of a buy-and-hold strategy in the market portfolio, which is a relatively passive investment strategy and hence incurs low transaction costs, an investment strategy based on recursive forecasts is likely to incur considerably higher transaction costs and may not be as profitable as the buy-and-hold strategy when transaction costs are appropriately taken into account.

To find out if our recursive predictions could have been used to generate a higher profit than that earned from following a buy-and-hold strategy in the market portfolio, we used our predictions in a simple switching strategy that has been employed extensively in the finance literature. According to this strategy, investors should hold equity in periods where the business cycle indicators suggest that equity returns are going to outperform returns from holding bonds (i.e., the predicted excess return on stocks is positive), and otherwise hold bonds. We do not allow for short-selling of assets, nor do we assume that investors can use leverage when selecting their portfolios. Also, in the absence of a published time series on transaction costs during the period 1960 to 1992, we made the simplifying assumption that these are constant through time and symmetric with respect to whether the investor is buying or selling assets. We further assume that transaction costs are proportional to the value of the trade, letting c_1 and c_2 be the percentage transaction costs on shares and bonds, respectively.[15]

To analyze how the value of the investor's portfolio evolves through time, we first introduce some notations. Let W_t be the funds available to the investor at the end of period t, N_t the numbers of shares held at the end of period t (after trading), P_t the price of shares at the end of period t, D_t the dividends per share paid during period t, r_t the rate of return on bonds in period t, B_t the investor's position in bonds at the end of period t (after trading), and $I_{t+1}(\hat{\rho}_{t+1})$ the indicator variable, as defined earlier, which takes the value 1 or 0 according to the predicted sign of excess returns in period $t + 1$, namely: $I_{t+1}(\hat{\rho}_{t+1}) = 1$, if $\hat{\rho}_{t+1} > 0$, and $I_{t+1}(\hat{\rho}_{t+1}) = 0$, otherwise. For simplicity we also refer to this function as I_{t+1}.

At a particular point in time, t, we assume that the funds are held entirely either in bonds or in stocks, according to the value predicted for I_{t+1}. Net of transaction costs the investor can allocate the funds either to shares or to bonds:

$$N_t = W_t I_{t+1}(1 - c_1)/P_t, \tag{12}$$

$$B_t = W_t(1 - I_{t+1})(1 - c_2). \tag{13}$$

[15] For a more detailed discussion of transaction costs and their relation to commission fees and bid-ask spreads, see Pesaran and Timmermann (1994a).

For period $t + 1$, the budget constraint of the investor becomes

$$W_{t+1} = N_t(P_{t+1} + D_{t+1}) + B_t(1 + r_t). \tag{14}$$

Based on the forecast of excess returns for period $t + 2$, the portfolio allocation procedure is repeated at the end of period $t + 1$. The size of the transaction costs incurred through the reallocation of funds depends on the composition of the investor's existing portfolio in bonds (B_t) or in stocks (N_t), and on the selected portfolio composition for period $t + 1$. This gives four different cases to be considered:

Case I (Reinvest Cash Dividends in Shares)

$$\begin{gathered} I_{t+1} = 1 \quad \text{and} \quad I_{t+2} = 1, \\ N_{t+1} = N_t + N_t D_{t+1}(1 - c_1)/P_{t+1}, \\ B_{t+1} = 0. \end{gathered}$$

Case II (Sell Stocks and Buy Bonds)

$$\begin{gathered} I_{t+1} = 1 \quad \text{and} \quad I_{t+2} = 0, \\ N_{t+1} = 0, \\ B_{t+1} = (1 - c_2)[(1 - c_1)N_t P_{t+1} + N_t D_{t+1}]. \end{gathered}$$

Case III (Bonds Mature, Buy Shares)

$$\begin{gathered} I_{t+1} = 0 \quad \text{and} \quad I_{t+2} = 1, \\ N_{t+1} = (1 - c_1)B_t(1 + r_t)/P_{t+1}, \\ B_{t+1} = 0. \end{gathered}$$

Case IV (Bonds Mature, Buy Bonds)

$$\begin{gathered} I_{t+1} = 0 \quad \text{and} \quad I_{t+2} = 0, \\ N_{t+1} = 0, \\ B_{t+1} = (1 - c_2)(1 + r_t)B_t. \end{gathered}$$

Using these formulae the value of the investor's funds at the end of period $t + 2$ becomes

$$W_{t+2} = N_{t+1}(P_{t+2} + D_{t+2}) + B_{t+1}(1 + r_{t+1}). \tag{15}$$

Extension of these rules to subsequent periods is straightforward.

A. Empirical Results from Trading Based on the Recursive Forecasts

Table III presents the trading results for a number of switching portfolios constructed using forecasts from models selected recursively according to the different selection strategies. Table III also reports the Sharpe index for the various portfolios. These computations assume that investors start off with \$100 at the beginning of 1960 and reinvest the portfolio income every month. In the case of the market and bond portfolios, only the dividends or interests

Table III

Performance Measures for the S&P 500 Switching Portfolio Relative to the Market Portfolio and T-Bills (Monthly Results: 1960(1) to 1992(12))

The switching portfolios are based on recursive least squares regressions of excess returns on an intercept term and a subset of regressors selected from a base set of 9 variables according to different statistical model selection criteria and financial performance criteria. See Section II of the article for a definition of the statistical and financial selection criteria. The hyper-selection criterion is described in the Appendix. The columns headed Zero, Low, and High refer to the portfolio returns under the three transaction costs scenarios described in Section III. A of the article. The final wealth figures assume that investors start off with $100 at the beginning of 1960 and reinvest portfolio income every month. The recursive wealth procedure uses as a forecasting model the one that, at each point in time, generated the largest wealth (net of transaction costs) when its forecasts were used in a simple trading strategy. The recursive Sharpe criterion is based on a similar procedure with the only difference that it is now the Sharpe ratio corresponding to trading results based on the set of predictions generated by a particular regression model that is being maximized recursively. S.D. is standard deviation.

Panel A: Benchmarks, Statistical Selection Criteria

Transaction costs	Market Portfolio			Bonds			Akaike			Schwarz		
	Zero	Low	High	Zero	Low	High	Zero	Low	High	Zero	Low	High
Mean return	11.39	11.34	11.29	5.92	4.66	4.66	14.06	12.51	11.48	11.68	9.46	7.91
S.D. of return	15.71	15.72	15.73	2.61	2.59	2.59	11.16	11.47	11.69	10.73	10.85	10.95
Sharpe's index	0.35	0.43	0.42	—	—	—	0.73	0.69	0.58	0.54	0.44	0.30
Final wealth ($)	**2503**	**2463**	**2424**	**660**	**445**	**445**	**6601**	4144	**3024**	**3305**	**1687**	**1049**

Panel B: Statistical Selection Criteria

Transaction costs	$\bar{R}^2$			Sign			All Regressors			Hyper-Selection Criterion		
	Zero	Low	High	Zero	Low	High	Zero	Low	High	Zero	Low	High
Mean return	14.83	13.22	12.13	13.69	11.70	10.25	14.46	12.85	11.78	13.58	12.32	11.28
S.D. of return	11.37	11.51	11.58	9.06	9.24	9.40	11.61	11.79	11.91	9.78	10.32	10.77
Sharpe's index	0.78	0.74	0.65	0.86	0.76	0.60	0.74	0.70	0.60	0.74	0.74	0.62
Final wealth ($)	**8218**	**5113**	**3694**	**6236**	**3452**	**2233**	**7329**	**4556**	**3288**	**5943**	**4044**	**2932**

Panel C: Financial Criteria

Transaction costs	Recursive Wealth Procedure			Recursive Sharpe Procedure		
	Zero	Low	High	Zero	Low	High
Mean return	14.00	11.76	10.83	15.14	11.81	10.73
S.D. of return	12.24	11.98	11.22	11.60	10.66	10.97
Sharpe's index	0.66	0.59	0.55	0.79	0.67	0.55
Final wealth ($)	**6242**	**3247**	**2508**	**8859**	**3429**	**2458**

are reinvested on a monthly basis. In contrast, the switching portfolios may reallocate funds between bonds and shares, depending on whether a change in the sign of the excess return is predicted.

First consider the results based on zero transaction costs. The mean annual return on the market index over the period 1960 to 1992 is 11.39 percent, and is smaller than the mean return on all the switching portfolios under consideration. Comparing the performance of the switching portfolios based on forecasts using different model selection criteria, the portfolio based on Schwarz's criterion only pays a marginally higher mean return than the buy-and-hold strategy in the market index. At the other end of the performance spectrum, the portfolios based on the forecasts using the Akaike, the $\bar{R}^2$, and the recursive Sharpe criteria pay mean returns of 14.06, 14.83, and 15.14 percent, respectively, which are substantially above the mean return on the market index. Interestingly, the annual mean return on the switching portfolio based on predictions maximizing recursive wealth is smaller than the annual mean returns on most of the other switching portfolios. Clearly, there are important differences in the performance of the model selection criteria in identifying predictability of stock returns. A possible method of resolving the uncertainty surrounding the choice of model selection criterion is discussed in the Appendix.

These differences in mean returns are reflected in the end-of-period wealth accrued to the investment strategies based on reinvesting the funds in either bonds or shares at the end of every month. Under the zero transaction cost scenario, the end-of-period funds of the switching portfolios were approximately twice as large as the end-of-period funds of the market portfolio (in the case of the Akaike, Sign, and recursive wealth criteria) or three times as large (in the case of the $\bar{R}^2$, the recursive Sharpe criterion and the fixed model specification that includes all regressors). The fact that the recursive predictions based on models selected according to the $\bar{R}^2$ or the recursive Sharpe criterion perform better than the recursive predictions from the model that included all regressors suggests that it may pay off for investors to engage in an active search process to find an adequate forecasting equation rather than just basing their forecasts on a fixed model specification that includes the entire set of regressors.[16]

Turning next to the standard deviation of the returns on the switching portfolios, these lie in the range from 9.1 to 12.2 percent per annum, which is substantially lower than the standard deviation of returns on the market portfolio (15.7 percent). The standard deviation of the returns on the switching portfolio selected on the basis of predictions using the sign criterion is particularly low. Taken together, the higher mean and lower standard deviation of the returns on the switching portfolios result in high values of the Sharpe ratio of these portfolios. Notice that the switching portfolio based on recursively maximizing the Sharpe ratio does not produce the highest Sharpe statistic among all the switching portfolios.

Allowing for "low" transaction costs of 0.5 of a percent on trading in shares ($c_1 = 0.005$) and 0.1 of a percent on trading in bonds ($c_2 = 0.001$), the mean payoffs on the switching portfolios decline by between 1.6 and 3.3 percent per

[16] This should be compared to the findings in Phillips (1992) which report that a model specification that includes a large number of autoregressive lags, as well as deterministic trends, tends to produce worse predictions than models selected recursively according to his Bayesian Posterior Information Criterion.

annum. Thus, under the low transaction cost scenario, mean returns on the switching portfolios based on predictions from models selected according to Schwarz's criterion are now lower than the mean returns on the market portfolio. In comparison, transaction costs hardly affect the mean return on the buy-and-hold strategy, since the only turnover associated with this portfolio arises from reinvestment of the dividends. Even so, under this low transaction cost scenario, the switching portfolios based on the forecasts produced by models that either included all regressors or were recursively selected according to any of the criteria (apart from the Schwarz's criterion) continue to pay higher mean returns with a lower standard deviation than the market index. This is the case despite the fact that the number of portfolio switches is between two and three per year, depending on which set of forecasts is used.

With "high" transaction costs of 1 percent on shares and 0.1 of a percent on bonds, the mean return on the switching portfolios based on forecasts using the $\bar{R}^2$ and the Akaike criteria still exceed the mean return on the market portfolio. Furthermore, as witnessed by the values of the Sharpe ratios, even with high transaction costs the switching portfolios based on predictions using these model selection criteria still offer a better risk-return trade-off than the market portfolio. Notice the sharp decline in the performance of the switching portfolios based on financial criteria as transaction costs are introduced. This suggests that a two-step procedure using statistical model selection criteria to compute recursive forecasts of stock returns and then trading on the basis of these forecasts may be better at identifying predictability in the stock market than a more direct procedure based on a forecasting model selected according to a financial criterion.

Using the test statistic suggested by Gibbons, Ross, and Shanken (GRS) (1989), we computed the joint significance of the intercept terms in regressions of monthly excess returns of the eight switching portfolios on a constant and the excess return on the market portfolio. In the case of zero transaction costs the value of the GRS test statistic was 3.26 (0.001). But as to be expected, the GRS test statistic declined to 2.36 (0.017) and 1.96 (0.05) for the low and high transaction cost scenarios, respectively. Rejection probability values are provided in brackets after the value of the test statistics.[17] Clearly, the mean-variance efficiency of the buy-and-hold strategy is rejected.

We also analyzed the performance of the switching portfolios over the sub-periods 1960 to 1969, 1970 to 1979 and 1980 to 1989 (see Table IV). For all three subperiods, the portfolios based on forecasts using the Akaike, $\bar{R}^2$, or the sign criteria paid a higher mean return than the buy-and-hold strategy under the zero transaction cost scenario. The switching portfolios based on the remaining model selection criteria only paid a higher mean return than the market during the 1970s and 1980s, while the switching portfolio using the Schwarz criterion paid a higher mean return than the market index only during the 1970s. When transaction costs are introduced, it becomes even

[17] Under the null hypothesis that the market portfolio is mean-variance efficient the Gibbons, Ross, and Shanken (1989) test statistic has a central F distribution.

Table IV

Risk and Returns of Different Portfolios for Subperiods: 1960s, 1970s, and 1980s

	Zero Transaction Costs			Low Transaction Costs			High Transaction Costs		
Risk and Returns	1960 to 1969	1970 to 1979	1980 to 1989	1960 to 1969	1970 to 1979	1980 to 1989	1960 to 1969	1970 to 1979	1980 to 1989
Panel A: Mean (Annualized)									
Portfolios									
Market	8.63	7.42	18.06	8.56	7.40	18.03	8.49	7.38	18.01
Bond	3.71	6.02	8.23	2.47	4.76	6.94	2.47	4.76	6.94
Akaike	8.78	13.35	19.14	7.37	12.06	17.03	6.55	11.31	15.40
Schwarz	5.20	11.94	16.70	3.28	10.13	14.09	2.09	8.96	12.15
$\bar{R}^2$	9.17	14.22	20.28	7.45	12.87	18.25	6.31	12.09	16.69
Sign	11.04	11.19	18.37	9.03	9.79	15.55	7.67	8.96	13.31
All regressors	8.13	15.03	19.26	6.29	13.64	17.42	4.99	12.84	16.08
Hyper selection	11.04	11.19	18.53	9.03	9.38	17.17	7.67	7.89	16.71
Recursive wealth proc.	7.81	14.26	18.07	6.90	11.44	15.16	6.28	9.84	14.57
Recursive sharpe proc.	8.49	15.06	21.25	6.61	13.55	13.96	4.28	12.64	13.43
Panel B: Standard Deviation (Annual)									
Portfolios									
Market	14.36	19.24	12.65	14.39	19.24	12.65	14.42	19.24	12.65
Bond	1.32	1.80	2.66	1.31	1.78	2.63	1.31	1.78	2.63
Akaike	8.30	10.64	12.90	9.16	10.62	13.41	9.72	10.56	13.88
Schwarz	9.92	9.29	10.07	10.53	9.24	10.53	10.97	9.10	11.02
$\bar{R}^2$	8.99	12.52	10.99	9.70	12.10	11.34	10.20	11.64	11.68
Sign	8.62	7.69	8.94	9.39	7.94	8.78	9.99	8.02	8.75
All regressors	7.14	13.17	12.29	8.37	12.60	12.46	9.36	12.03	12.54
Hyper selection	8.62	7.69	10.81	9.39	9.32	10.79	9.99	8.53	11.67
Recursive wealth proc.	10.51	13.44	10.81	11.49	13.02	10.21	11.77	13.10	7.27
Recursive sharpe proc.	10.60	11.91	10.07	11.02	11.13	8.88	11.92	10.47	8.47
Panel C: Sharpe Ratio									
Portfolios									
Market	0.342	0.073	0.776	0.424	0.137	0.877	0.418	0.136	0.875
Bond	-	-	-	-	-	-	-	-	-
Akaike	0.610	0.689	0.846	0.535	0.688	0.752	0.420	0.621	0.609
Schwarz	0.150	0.637	0.842	0.078	0.581	0.679	−0.034	0.461	0.473
$\bar{R}^2$	0.607	0.655	1.097	0.513	0.670	0.998	0.376	0.630	0.835
Sign	0.849	0.672	1.130	0.699	0.633	0.980	0.521	0.524	0.727
All regressors	0.619	0.683	0.898	0.456	0.705	0.841	0.269	0.672	0.729
Hyper selection	0.849	0.672	0.952	0.699	0.496	0.948	0.521	0.367	0.837
Recursive wealth proc.	0.390	0.613	0.910	0.386	0.513	0.805	0.324	0.388	1.050
Recursive sharpe proc.	0.450	0.759	1.290	0.376	0.790	0.790	0.152	0.753	0.766

See the notes to Table III for details of the various procedures.

clearer that the higher mean returns on the switching portfolios are concentrated in the 1970s. Since the standard deviation of returns in the stock market was particularly high during the 1970s, these results seem to indicate that, if ever there was a possibility that investors could improve their market timing based on a simple forecasting procedure similar to ours, this was during the volatile periods in the 1970s where macroeconomic risks and volatility in nominal magnitudes, such as the rate of inflation and nominal interest rates, mattered the most.

Because the portfolios based on forecasts using models recursively selected according to the $\bar{R}^2$, Akaike, and the Sign criteria were quite successful, while portfolios based on the Schwarz criterion were not as successful in terms of the values generated for the financial performance measures, it raises the issue of whether our results can be explained by uncertainty over the choice of model selection criterion. This problem is addressed in the Appendix, where it is shown that the switching portfolios continue to strongly outperform the buy-and-hold strategy, even if the choice of the model selection criterion is made endogenous to the forecasting and the trading processes. The hyper-selection procedure advanced in the Appendix for the resolution of the uncertainties over the choice of model selection criteria and the forecasting model can also be viewed as an artificial intelligence system capable of detecting unexploited profit opportunities in the market. Clearly, it would be possible to devise more comprehensive and sophisticated artificial intelligence systems for the analysis of stock market predictability. However, a comparative analysis of such systems fall outside the scope of the present paper.

IV. Concluding Remarks and a Discussion of the Main Results

We have proposed in this article a new approach for simulating the behavior of an investor in real time, using as little hindsight as possible and specifically accounting for the effect of model specification uncertainty which is crucially important to any investor trying to forecast asset returns in real time.

As far as the issue of robustness of the predictability of U.S. stock returns is concerned, our plots of the inclusion frequency of the different factors in the forecasting models clearly show the importance of allowing for changes in the underlying process of excess returns in the U.S. stock market that seem to have taken place during the 1960 to 1992 period. The only variable to be included in the forecasting models throughout the entire sample period 1960 to 1992 is the one-month lagged value of the one-month T-bill rate. Monetary growth and industrial production are included in the forecasting models more or less continuously from the mid and late 1960s, respectively, and the dividend yield variable starts to get included as a regressor in the forecasting models around 1970. The frequency with which the inflation rate variable and the 12-month interest rate are included in the forecasting equations is, however, closely related to economic "regime switches": the inflation rate gets included after the first oil shock while, according to the models selected by the

$\bar{R}^2$ criterion, the 12-month interest rate is selected from the mid-seventies onwards during periods where the Federal Reserve targeted interest rates.

Thus, in general our findings confirm the results of recent research which has emphasized the importance of predictable components in stock returns related to the business cycle, (see the references cited in Footnote 1). But we find that predictability of stock returns of a magnitude that is economically exploitable seems to depend not just on the evolution of the business cycle, but also on the magnitude of the shocks. Also there does not seem to be a robust forecasting model in the sense that the determinants of the predictability of stock returns in the U.S. seem to have undergone important changes throughout the period under consideration. The timing of the episodes where many of the regressors get included in the forecasting model seems to be linked to macroeconomic events such as the oil price shock in 1974 and the Fed's change in its operating procedures during the 1979 to 1982 period. If we are right in our conclusion that important episodes of predictability of stock returns are closely linked to incidence of sudden shocks to the economy, then in analyzing stock return predictability it is advisable to use forecasting procedures that allow for possible regime changes.

Another conclusion emerging from our study is that there appears to be a relation between periods with high volatility in the markets and periods with higher-than-normal predictability of excess returns on shares. Our results suggest that during the relatively calm markets of the 1960s, there were no excess returns to be gained from following a switching strategy based on forecasts using the recursively selected regression models. In contrast, during the more volatile 1970s there seems to have been important gains to be made from following such a forecasting and trading strategy, while in the 1980s the gains were much smaller. This finding could be consistent both with incomplete learning in the aftermath of a large shock to the economy (see Timmermann (1993)) as well as with a story where the predictability of excess returns is reflecting time-varying risk premia. In the context of the latter it is, however, difficult to explain why return on the switching portfolio exceeds the return on the market when the markets are volatile. It is well known that there is no theoretical reason why required returns on stocks cannot be lower during periods with relatively high volatility. For instance, risk averse investors may want to increase their savings, thereby bidding down the equilibrium return on stocks when the markets are particularly volatile. Furthermore, it is quite possible that the price of risk is time-varying so that there is no constant, proportional relationship between the first and second conditional moments of stock returns. Given the existence of a risk-free T-bill rate, which establishes a lower bound for the nominal return, it seems difficult, however, in the context of an equilibrium model to explain the predictions of negative risk premia on the market portfolio apparent in the 1970s.[18] On the other hand,

[18] One possibility is that, during the periods of high volatility and high inflation in the early 1970s, stocks acted as an inflation hedge making investors more willing to accept low or even negative expected excess returns of stocks. For a more detailed discussion of the equilibrium conditions under which the predicted excess return on stocks can be negative in certain states, see Pesaran and Potter (1993).

it is quite possible that in the event of a major regime switch in the economy, such as the one induced by the first oil shock in 1973, learning may take longer than usual to complete as investors would need extra time to model, say, the new relationship between inflation, nominal interest rates, and stock returns. Learning about the market is a continuous process that involves both the routine updating of the parameter estimates of a given model as well as searching for new models when there are clear signs that the old established relations are no longer valid. It is this latter form of learning that is likely to be time consuming, and seems to account for the increased predictability of stock returns in periods with large and sudden shocks and important regime switches.

Finally, our results do not appear to be sensitive to the particular choice of trading rule. In fact, we investigated the returns from following a trading rule that explicitly accounts for the prediction uncertainty. According to this rule, investors stay in the stock market unless their predictions indicate that there is at least 90 percent probability that bonds will pay a higher return than stocks. This is a relatively conservative trading rule in the sense that the investor stays in the stock market unless he is fairly confident that it will be better to stay in bonds. Consequently this rule generates fewer switches between the two types of assets, and transaction costs are much lower. We found that this trading rule would generate returns with a higher mean and lower standard deviation than the market index, provided transaction costs are zero or low. Under high transaction costs the mean return on the switching portfolio was similar to the market, whereas the volatility of returns was much lower.

Appendix

Uncertainty Over the Choice of Model Selection Criterion—A Profit-Based Hyper-Selection Criterion

It is clear from the empirical results reported in Section III that although most of the model selection criteria considered generate a profit relative to the market index, this is not the case for the Schwarz criterion under the low or high transaction cost scenarios. Without some rule for choosing a model selection criterion an investor could not, without the benefit of hindsight, have been guaranteed to choose one of the more successful selection criteria. In this Appendix we address this issue and consider the problem of how an investor could resolve the uncertainty surrounding the choice of model selection criterion.

In view of the objective of the exercise we consider a profit-based hyper-selection criterion that we employ recursively to choose the model selection criterion to be used subsequently for selecting the forecasting model. The idea is similar to choosing a subset of regressors from the base set of regressors explained in detail in Section II. Here, however, there are two levels at which the search for a suitable forecasting equation needs to be conducted. First, a model selection criterion needs to be chosen from the base set consisting of the five statistical model selection criteria, namely the $\bar{R}^2$, Akaike, Schwarz, and the Sign criteria, as well as the general model specification that includes all the regressors. Having chosen a model selection criterion we then proceed to select the subset of the regressors from the base set for computing one-period ahead forecasts. More specifically, the profit-based hyper-selection criterion works as follows: at each point in time we calculate the funds that would have been generated by using the forecasts from the models selected according to the various model selection criteria. Then the model selection criterion corresponding to the fund with the highest cumulative wealth is chosen to select the model to be used for forecasting the excess return for the next period.

In application of the hyper-selection rule we used information up to 1959(12) to choose a forecasting model and parameter estimates according to the five different ways of selecting a model. Generating in-sample forecasts for the period 1954(1) to 1959(12) and using these in a trading strategy beginning in 1954(1), we computed the value of the funds for the five different portfolios (associated with the five different model selection criteria) at the end of 1959(12). The model selection criterion for which the simulated wealth at this point was highest was then used to select a model to forecast excess returns for 1960(1). Given the dependence of the portfolio values on transaction costs, it is clear that the choice of the model selection criterion will in general also depend on the assumed level of transaction costs. In the low transaction costs scenario, which is likely to be the most relevant one in practice, the final wealth of the switching portfolio selected using the profit-based hyper-selection criterion was 4044 dollars. This compares with final funds of 2463 dollars for the market index under the low transaction cost scenario. (See the results for the Hyper-Selection Criterion in Table III).

In the case of zero transaction costs, the Sign criterion was chosen by the profit-based hyper-selection criterion from 1960 to 1986. Thereafter, the $\bar{R}^2$ criterion was chosen up to 1992, apart from a brief spell in 1990, when the Sign criterion was again chosen. Under the low transaction costs scenario, the Sign criterion was again chosen by the hyper-selection criterion over the periods 1960 to 1975 and 1980 to 1985, while the $\bar{R}^2$ criterion was chosen in the remaining years. These findings are in accordance with the subsample results reported in Table IV, which show that the switching portfolio based on the forecasts using the Sign criterion paid the highest return during the 1960s, but did less well relative to other selection criteria during the 1970s and 1980s.

REFERENCES

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, in B.N. Petrov and F. Csaki, Eds., *Second International Symposium on Information Theory*, Akademiai Kiado, (Budapest, Hungary), 267–281.

Amemiya, T., 1980, Selection of regressors, *International Economic Review* 21, 331–345.

Angas, L.L.B., 1936, Investment for appreciation. Forecasting movements in security prices. *Technique of Trading in Shares for Profit* (Macmillan, London).

Balvers, R.J., T.F. Cosimano, and B. McDonald, 1990, Predicting stock returns in an efficient market, *Journal of Finance* 45, 1109–1128.

Bossaerts, P., and P. Hillion, 1994, Implementing statistical criteria to select return forecasting models, Working paper, Caltech.

Breen, W., L.R. Glosten, and R. Jagannathan, 1990, Predictable variations in stock index returns, *Journal of Finance* 44, 1177–1189.

Campbell, J.Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

Cochrane, J.H., 1991, Production-based asset pricing and the link between stock returns and economic fluctuations, *Journal of Finance* 46, 209–238.

Dowrie, G.W. and D.R. Fuller, 1950, *Investments* (second ed.). (John Wiley, New York).

Fama, E.F., 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

Fama, E.F., and K.R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Ferson, W.E., and C.R. Harvey, 1993, The risk and predictability of international equity returns, *Review of Financial Studies* 6, 527–566.

French, K.R., G.S. Schwert, and R.F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–30.

Gibbons, M.R., S.A. Ross, and J. Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Glosten, C.R., R. Jagannathan, and D.E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess returns on stocks, *Journal of Finance* 48, 1779–1802.

Goetzmann, W.N., and P. Jorion, 1993, Testing the predictive power of dividend yields, *Journal of Finance* 48, 663–680.

Henriksson, R.D., and R.C. Merton, 1981, On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513–533.

Jensen, M.C., 1978, Some anomalous evidence regarding market efficiency, *Journal of Financial Economics* 6, 95–101.

Leitch, G., and J.E. Tanner, 1991, Economic forecast evaluation: Profits versus the conventional error measures, *American Economic Review* 81, 580–590.

Mallows, C.P., 1973, Some comments on C_p, *Technometrics* 15, 661–675.

Mishkin, F.S., 1992, The economics of money, banking, and financial markets, Third ed. (Harper Collins Publishers, New York).

Morgan, E.V., and W.A. Thomas, 1962, The stock exchange. Its history and functions, (Elek Books, London).

Nelson, C.R., and M.J. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–662.

Pesaran, M.H., and S. Potter, 1993, Equilibrium asset pricing models and predictability of excess returns, Department of Applied Economics, Working paper, Cambridge University.

Pesaran, M.H., and A. Timmermann, 1992, A simple nonparametric test of predictive performance, *Journal of Business and Economic Statistics* 10, 461–465.

Pesaran, M.H., and A. Timmermann, 1994a, Forecasting stock returns. An examination of stock market trading in the presence of transaction costs, *Journal of Forecasting* 13, 335–367.

Pesaran, M.H., and A. Timmermann, 1994b, A generalisation of the nonparametric Henriksson-Merton test of market timing, *Economics Letters* 44, 1–7.

Pesaran, M.H., and A. Timmermann, 1994c, The use of recursive model selection strategies in forecasting stock returns, Department of Applied Economics, University of Cambridge, Working paper No. 9406.

Phillips, P.C.B., 1992, Bayesian model selection and prediction with empirical applications. Cowles Foundation Discussion Paper No. 1023, Yale University.

Phillips, P.C.B., and Ploberger, 1994, Posterior odds testing for a unit root with data-based model selection, *Econometric Theory* 10, 744–808.

Pötscher, B.M., 1991, Effects of model selection on inference, *Econometric Theory* 7, 163–185.

Prime, J.H., 1946, Investment analysis (Prentice Hall, New York).

Rose, H.B., 1960, The economic background to investment (Cambridge University Press).

Schwarz, G., 1978, Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

Shibata, R., 1976, Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika* 63, 117–126.

Theil, H., 1958, Economic forecasts and policy, (Amsterdam, North Holland).

Timmermann, A., 1993, How learning in financial markets generates excess volatility and predictability in stock prices, *Quarterly Journal of Economics*, 108, 1135–1145.