TILBURG ◆ UNIVERSITY

Tilburg University

The Economic Value of Predicting Stock Index Returns and Volatility

Marquering, W.; Verbeek, M.J.C.M.

Publication date:
2000

Center
for
Economic Research

No. 2000-78

THE ECONOMIC VALUE OF PREDICTING STOCK INDEX RETURNS AND VOLATILITY

By Wessel Marquering and Marno Verbeek

September 2000

ISSN 0924-7815

The Economic Value of Predicting Stock Index Returns and Volatility*

Wessel Marquering
K.U. Leuven

Marno Verbeek
K.U. Leuven and Tilburg University

September 6, 2000

Abstract

In this paper, we analyze the economic value of predicting index returns as well as volatility. On the basis of fairly simple linear models, estimated recursively, we produce genuine out-of-sample forecasts for the return on the S&P 500 index and its volatility. Using monthly data from 1954 to 1998, we test the statistical significance of return and volatility predictability and examine the economic value of a number of alternative trading strategies. We find strong evidence for market timing in both returns and volatility. Joint tests indicate no dependence between return and volatility timing, while it appears easier to forecast returns when volatility is high. For a mean-variance investor, this predictability is economically profitable, even if short sales are not allowed and transaction costs are quite large.

JEL classification: G11; G14.

Keywords: Predictability of Stock Returns and Volatility; Market Timing; Performance Evaluation.

---

* The authors would like to thank Allan Timmermann for providing the core dataset used in this paper. We also wish to thank Pascal Maenhout for providing some additional data, and Frank de Jong, Christopher Polk and several seminar participants for helpful comments and suggestions. An earlier version of the paper was presented at the 2000 European Finance Association meeting in London. The authors acknowledge the support of the Belgian Program "Interuniversity Poles of Attraction", initiated by the Prime Minister's Office, Science Policy Programming. Address all correspondence to: Wessel Marquering, Center for Economic Studies, K.U. Leuven, Naamsestraat 69, 3000 Leuven, Belgium. Tel.: +32-16-326849. Fax: +32-16-326796. E-mail: Wessel.Marquering@econ.kuleuven.ac.be.

1 Introduction

In the academic literature on stock market predictability the prevalent view until the 1970s was that stock prices are very closely described by a random walk and that no economically exploitable predictable patterns exist. More recent empirical work, however, reports evidence that stock returns are to some extent predictable, either from their own past or from other publicly available information, like dividend yields, price-earnings ratios, short and long interest rates, changes in industrial production, or simply from calendar dummies; see, for example, Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell (1987), Fama and French (1989), Breen, Glosten and Jagannathan (1989), Harvey (1991), Bekaert and Harvey (1995), Pesaran and Timmermann (1995, 2000), Kandel and Stambaugh (1996), and Sullivan, Timmermann and White (1999). While the majority of this literature focuses on the statistical evidence of predictable time variation in expected returns, several authors have explicitly addressed the economic value of this predictability (see, for example, Solnik, 1992, Pesaran and Timmermann, 1995, 2000, and Jacobsen, 1999). In this case, the question is whether a given trading strategy outperforms a passive investment strategy, where trading is based on information that is genuinely available at each time an investment decision is made, taking into account, for example, transaction costs and short-sales constraints.

Another source of predictability that is well-documented exists in the volatility of stock market returns; see, for example, Engle (1982), Bollerslev (1986), French, Schwert and Stambaugh (1987), Pagan and Schwert (1990) and Bollerslev, Engle and Nelson (1994) to name a few. Only a few recent papers (e.g. Fleming, Kirby and Ostdiek, 1999) examine the economic significance of volatility timing, by considering trading rules that exploit the predictability in second moments of returns. In this paper we analyze the economic significance of jointly predicting stock market index returns and their volatility. While this problem has been neglected in the literature, it is of considerable interest. On the one hand, it is possible that any predictability in expected returns is accompanied by a change in the conditional variance, so that the economic value of return predictability is overstated by studies that assume that variances are constant. On the other hand, the exploitation of both sources of predictability is expected to result in investment strategies that outperform strategies that exploit only one source.

While the ability to predict returns from publicly available information is often associated with a violation of market efficiency, this is not necessarily the case. On the one hand, rational asset pricing models may imply time-varying risk premia that lead to predictable patterns in observed asset returns. For example, Lucas (1978) shows that rational expectations can result in predictable variation in expected returns. More recently, Kirby (1998) analyzes the question to what extent several well-known pricing models are consistent with the ability to predict returns on size-based portfolios and concludes that, in general, returns are "too predictable". Further, Lewellen and Shanken (2000) argue that estimation risk, due to investors uncertainty about the parameters of the return generating process, may also lead to return predictability, while Brock and Hommes (1997) show that an evolutionary asset pricing model with heterogenous traders may result in persistent deviations from rational expectations fundamental values. On the other hand, it is likely that most empirical studies evaluating within-sample predictability overstate out-of-sample predictability because of overfitting, finite sample biases and data snooping; see, for example, Bossaerts and Hillion (1999) and Sullivan, Timmermann and White (1999).

In this paper we consider a mean-variance investor, who allocates his wealth every month to the Standard & Poor's 500 index and the short-term Treasury bill (used as a proxy for the risk free rate). Asset allocation is based upon an information set that includes the results of two simple models for predicting index returns and volatility, estimated using data that were genuinely available. This way, we truly focus on *ex ante* predictability. The mean-variance investor will invest relatively little in stocks if the predicted excess return is low and/or the predicted volatility is high. We assume that the investor uses only public information to generate forecasts, and that his transactions do not influence prices. Furthermore, we assume that the investor rebalances his portfolio only once a month. The coefficients of the models are re-estimated each month when new information comes available, so that each month the investor revises his beliefs about expected returns and volatility.

The main goal of this paper is to analyze the economic value of trading strategies based on exploiting the predictable components in both the first and second moment of stock returns, using post-war U.S. data from 1954 to 1998. First, we analyze the statistical evidence for the existence of market timing in returns and volatility, both at the monthly frequency, using the non-parametric tests of market timing skills proposed by Henriksson and Merton (1981), Cumbey and Modest (1987) and some extensions. As volatility is unobserved, even ex post, these test are not directly applicable to volatility timing, and we propose some adjustments using daily data. These enable an easy and relatively reliable way to test for volatility timing. Because in many cases both return and volatility forecasts will be used to determine an investor's portfolio weights, it is important to simultaneously evaluate the existence of return and volatility timing and their interdependence. We do so by extending the existing tests to this multivariate case. This allows us, for example, to investigate whether a good return forecasts is typically associated with a bad volatility forecast or vice versa.

Next, we consider a number of different trading rules: the optimal (conditionally mean-variance efficient) strategy and a sub-optimal switching strategy, using predictions from the return model, the volatility model or both, and three buy-and-hold strategies. Furthermore, we take into account short-sales constraints and transaction costs. In addition to reporting summary statistics for the returns on these strategies, we report Sharpe ratios, Jensen's alphas and Treynor-Mazuy measures. Each of these measures, however, suffers from some drawbacks in the presence of time-varying expected returns and volatility. Therefore, we also express the economic value of each strategy by going back to the investor's utility function and answer the following question: how much would an uninformed investor, with a given risk aversion, be willing to pay to switch from a static to a given dynamic portfolio (compare Fleming, Kirby and Ostdiek, 1999)? That is, what is the maximum fee the investor would be willing to pay to switch from the static strategy to the dynamic one?

The remainder of this paper is organized as follows. Section 2 introduces a mean-variance investor and presents a number of alternative investment strategies. Section 3 describes the data used in our analysis and briefly presents preliminary results based on estimating the models over the entire sample and two subsamples. Section 4 presents the results of several market timing tests and analyzes the statistical significance of predictability of stock index returns and their volatility. In Section 5, we analyze a number of trading strategies to study the economic significance, and, finally, Section 6 concludes.

2 Trading Rules of a Mean-Variance Investor

Consider an investor maximizing a mean-variance utility function and composing his portfolio from a risky asset and a risk free asset. For a given level of (initial) wealth, the investor's optimization problem is given by

$$\max_{w_{t+1}} U\left(E_t\left\{r_{p,t+1}\right\}, Var_t\left\{r_{p,t+1}\right\}\right), \tag{1}$$

where w_{t+1} denotes the proportion of the portfolio allocated to the risky asset, and $r_{p,t+1}$ is the return of the investor's portfolio, which is equal to

$$r_{f,t+1} + w_{t+1}(r_{m,t+1} - r_{f,t+1}), \tag{2}$$

where $r_{m,t+1}$ denotes the return on the risky asset in period $t+1$, and $r_{f,t+1}$ the risk free return.

More specifically we assume the following utility function:

$$E_t\left\{r_{p,t+1}\right\} - \frac{1}{2}\gamma Var_t\left\{r_{p,t+1}\right\}, \tag{3}$$

where γ denotes the coefficient representing the investor's degree of risk aversion. Solving the maximization problem shows that the optimal portfolio weight for the investor is given by

$$w^*_{t+1} = \frac{E_t\left\{r_{m,t+1}\right\} - r_{f,t+1}}{\gamma Var_t\left\{r_{m,t+1}\right\}}. \tag{4}$$

If the risk aversion increases, the fraction invested in the risky asset will decrease. If the expected excess return on the risky asset increases, the investor *ceteris paribus* wants to increase his proportion invested in the risky asset. The conditional variance, which represents a measure of the risk involved, is negatively related to this proportion. If we assume that short selling and borrowing at the risk free rate is not allowed, the portfolio weights must lie between 0 and 1, and the optimal portfolio weight becomes

$$\begin{aligned} w^{ns}_{t+1} &= 0 \quad \text{if} \quad w^*_{t+1} \leq 0, \\ &= w^*_{t+1} \quad \text{if} \quad 0 < w^*_{t+1} \leq 1, \\ &= 1 \quad \text{if} \quad w^*_{t+1} > 1. \end{aligned} \tag{5}$$

We will compare the performance of the optimal rules with a passive (buy-and-hold) strategy and a sub-optimal switching rule (as used in e.g. Pesaran and

Timmermann, 1995), in which 100% is invested in the market return if the expected excess return is positive, and 100% in the risk free return otherwise. In formula:

$$\begin{aligned} w^{**}_{t+1} &= 1 \qquad \text{if} \qquad E_t\left\{r_{m,t+1} - r_{f,t+1}\right\} > 0, \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{6}$$

While w^{*}_{t+1} is an optimal trading rule, w^{**}_{t+1} generally is not.[1] One may suspect that the switching rule involves less transaction costs because it will have periods in which the portfolio remains unchanged.

The important elements in these trading rules are the conditional expectation and the conditional variance of the risky return. In the empirical application we shall approximate these conditional moments by fairly simple functions of historical returns and other observed variables. Let X_t denote a row vector of variables that is observed at time t, including a constant. These variables are used to predict the excess return $r^e_{m,t+1} = r_{m,t+1} - r_{f,t+1}$ on the stock market index (the risky asset), by assuming that

$$r^e_{m,s+1} = X_s\delta + \varepsilon_{s+1}, \qquad s = 1, 2, ..., t, \tag{7}$$

where $E_s\left\{\varepsilon_{s+1}\right\} = 0$ and δ is a vector of unknown coefficients. In the empirical section, the parameters in δ are estimated recursively by OLS, using information from periods 1 to t. We thus use a recursive regression where estimation is based on windows of expanding size. The model in (7) implies that

$$E_t\{r^e_{m,t+1}\} = X_t\delta, \tag{8}$$

so that, with the estimate $\hat{\delta}_t$ for δ, we obtain the conditional forecast for the excess return in period $t+1$, denoted by $\hat{r}^e_{m,t+1}$, as

$$\hat{r}^e_{m,t+1} = X_t\hat{\delta}_t. \tag{9}$$

This forecast is updated every period because new information becomes available $(X_{t+1}, X_{t+2}, ...)$ and because the coefficient estimate is updated as well. Information about future values of $r^e_{m,t}$ or X_t is not used at any point in time, so a comparison of $\hat{r}^e_{m,t+1}$ with realized returns $r^e_{m,t+1}$ is truly a measure of *ex ante* predictability.

[1] Note that the switching rule is optimal when $\gamma = 0$ and short-sale and borrowing restrictions are present.

In a very similar fashion we consider a linear model for the conditional variance of $r^e_{m,t+1}$, which is explained from a set of variables Z_t, observed at time t, and potentially different from X_t. Besides macroeconomic and financial variables, Z_t can contain lagged dependent variables (see, e.g., Kothari and Shanken, 1997). Again, the coefficients are estimated recursively using information from observations 1 to t, applying OLS to

$$e^2_{s+1} = Z_s\phi + \xi_{s+1}, \qquad s = 1, 2, ..., t, \tag{10}$$

where e^2_{s+1} is the squared prediction error of regression (7). That is, $e^2_{s+1} = (r^e_{m,s+1} - X_s\hat{\delta}_s)^2$. A similar model for conditional volatility is used by, e.g., Breen, Glosten and Jagannathan (1989), while Schwert (1989) runs regressions of the absolute prediction error on its own lagged values and lagged values of volatilities in economic fundamentals like output, inflation, the interest rate and the corporate bond yield. The conditional forecast for the variance of the excess return in period $t+1$, is given by

$$V\hat{a}r_t\{r^e_{m,t+1}\} = Z_t\hat{\phi}_t, \tag{11}$$

where the coefficients are estimated using information from period t and before.

The models for the conditional expectation and variance are deliberately chosen to be linear and rather simple with fixed selections for the variables included in X_t and Z_t. First, while the recent literature presents more complicated nonlinear models, for example based on neural networks or GARCH-type specifications, these techniques and approaches were certainly not available to investors in the major part of our sample period. Second, we want to limit the specification search for alternative models in order to reduce data snooping biases as much as possible. Pesaran and Timmermann (1995, 2000) and Bossaerts and Hillion (1999) use statistical model selection criteria at each point in time to select the "best" model to predict future returns (that is the best subset of X_t variables). As we *a priori* restrict attention to one specification only, the economic value of our trading strategies is a conservative estimate compared to these studies. Also note that we do not claim in any way that the specification we employ is 'correct'. Within sample it determines the best linear predictor from a given set of explanatory variables, and we simply use it to generate a series of one-month ahead forecasts.

3 Data and Some Preliminary Results

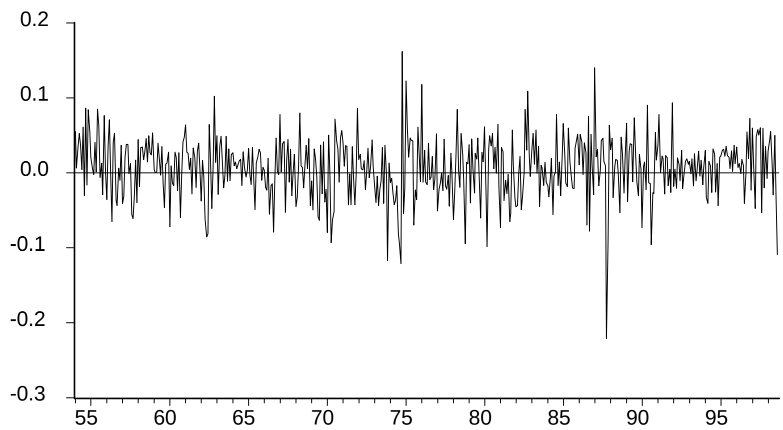
Because we do not perform any statistical tests to determine the "best" forecast model, and no other model selection criteria are employed, our procedure is not subject to data snooping in a strict sense. However, the choice of variables that we include in our analysis is a potential source of indirect data snooping biases, because it is tempting to include variables that appeared to have significant explanatory power in other studies that use partly the same data set. To avoid this, it is important to only include variables that have some economic rationale or whose significance is based on out-of-sample (pre-sample) evidence.

Many early studies, including Angas (1936), Prime (1946), and Dowrie and Fuller (1950), document that the state of the business cycle is an important indicator for the stock market. While financial theory does not provide much guidance on which business cycle indicators to select, fundamental variables suggested in this context include dividend yields, company earnings, changes in short-run and long-run interest rates, industrial production, liquidity measures and the inflation rate. Early studies, most notably Mandelbrot (1963), also document the persistence in stock market volatility, particularly at higher frequencies. The variables in our analysis are chosen to match the above categories.

The data set we use is an updated and expanded version of the one employed by Pesaran and Timmermann (1995), which has a monthly frequency and covers the period 1954:1 to 1992:12. We have updated the data set to capture a more recent time period (until 1998:9). We use the following financial and macroeconomic series: the return on the Standard & Poor's 500 (S&P 500) index, the 3-month Treasury bill, the price-earnings ratio on the S&P 500, dividend yield on the S&P 500, inflation, industrial production, the 12-month Treasury bill, monetary growth, and the commercial paper-Treasury yield spread.

While Pesaran and Timmermann (1995) use the one-month Treasury bill return, we use the return on the three-month Treasury bill to approximate the risk free rate. This measure is also used in related papers, e.g. Harvey (1991) and Pontiff and Schall (1998). Duffee (1996) claims that the one-month Treasury bill return is a poor proxy for the risk free rate, due to its idiosyncratic volatility. Moreover, the 3-month Treasury bill rate has the advantage that its data are more widely available than for 1-month Treasury bill rate, so that it is more likely that

Figure 1: Excess Returns S&P 500 Index (1954:1 - 1998:9).



a portfolio manager would have selected the former. To avoid look-ahead bias, we include the financial variables with one lag, and the macroeconomic variables with two lags, as the macroeconomic variables are typically only available with a two-month lag.

Table 1 presents some summary statistics for the monthly stock index return, the risk free rate, and the excess return. The excess return, which is displayed in Figure 1, does not exhibit any obvious patterns, except perhaps for some business cycle effects. For the entire sample period, January 1954 to September 1998, the average return on the Standard & Poor's 500 index is approximately 1.06% per month, while the average return on the three-month Treasury Bill is approximately 0.45% per month. The standard deviation (which is an indication for the corresponding risk) is, as expected, much higher for the stock index return. Looking at two subperiods (which are chosen to include half of the data points) it appears that for the first subperiod, 1954:1 to 1976:5, the average monthly return on the stock index and on the risk free rate is somewhat lower than in the second subperiod, 1976:6 to 1998:9. Finally, as is well known, the monthly stock index return and excess return exhibit a high level of kurtosis. The appendix provides details about the definitions and sources of the variables employed.

To obtain an indication to what extent the excess returns are predictable using financial ratios and macroeconomic variables, Table 2 presents the estimation results for model (7) using the entire sample period (i.e. within-sample); this is

Table 1: **Summary Statistics**

Mean, standard deviation, skewness, and kurtosis of monthly stock index return, risk free rate, and excess return. The stock index return is measured by the return on Standard & Poor's 500 index and the risk free rate is the 3-month Treasury bill rate published by the Federal Reserve Bank of St. Louis. The first column presents the statistics over the period 1960:1–1998:9; the second and third column correspond to the two subperiods of 1960:1–1979:5 and 1979:6–1998:9 respectively.

	1954:1-1998:9	1954:1-1976:5	1976:6-1998:9
	$T = 537$	$T = 269$	$T = 268$
	Monthly Stock Index Return		
Mean (%)	1.0577	0.8943	1.2218
Std. dev. (%)	4.0565	4.0493	4.0647
Skewness	-0.3930	-0.0151	-0.7683
Kurtosis	5.1819	3.8908	6.5542
	Monthly Risk Free Rate		
Mean (%)	0.4474	0.3323	0.5629
Std. dev. (%)	0.2210	0.1480	0.2221
Skewness	1.0022	0.3974	0.9964
Kurtosis	4.3303	2.5625	3.6959
	Monthly Excess Return on Stocks		
Mean (%)	0.6104	0.5621	0.6589
Std. dev. (%)	4.0813	4.0882	4.0814
Skewness	-0.4114	-0.0537	-0.7701
Kurtosis	5.1190	3.8208	6.4278

a simple linear projection of the monthly excess return on the S&P 500 index on one or two lags of the monthly price-earnings ratio, dividend yield, inflation rate, change in industrial production, short and long interest rates, monetary growth, and the commercial paper-Treasury yield spread. In early studies, these variable are argued to be able to predict stock market returns through their relationship with the business cycle. The price-earnings ratio is employed as a measure of company earnings, while monetary growth and the commercial paper-Treasury yield spread measure liquidity. It should be stressed that the estimates in Table 2 are not used to study the economic significance of predictability. They are merely presented to obtain some idea about the within-sample evidence of predictability of stock market returns. As explained in Section 3, we shall use the recursive modelling approach to generate genuine out-of-sample forecasts. Table 2 presents the estimation results of the entire sample period (1954:1-1998:9) and two subperiods

(1954:1-1976:5 and 1976:6-1998:9), so that we can examine whether the estimation results are robust over the two subsamples. Clearly, there is some statistical evidence of within-sample predictability with R^2s ranging from 11 to 23%, while the F-statistics in Table 2 indicate rejection of the null hypothesis of constant expected returns (with p-values of 0.00). To allow for heteroskedasticity of unknown form, the t-statistics are based upon the heteroskedasticity-consistent covariance matrix estimator proposed by White (1980).

The estimated coefficient for the price-earnings ratio is negative for the entire sample period and both subsamples, and is significant for the entire sample and the second subsample at conventional levels. A negative sign corresponds with the literature and could be interpreted as a mean-reversion effect due to undervaluation or overvaluation. For none of the specified samples the dividend yield has a coefficient which is significantly different from zero (as in Sentana and Wadhwani, 1991), while elsewhere in the literature often a significantly positive value is found. An explanation for this difference in findings is that the dividend yield and the price-earnings ratio may represent the same proxy (namely a variable which indicates whether the market is overvalued or undervalued), see also Kothari and Shanken (1997). Leaving out the price-earnings ratio results in a positive and significant coefficient for the dividend yield, which supports this suspicion.

The inflation variable has a (marginally significant) negative impact on the excess return, which is intuitively plausible because inflation is an indicator of an overheating economy, which is usually followed by a bear market. We include interest rates with one and two lags, to allow excess returns to depend upon differences in interest rates as well as their levels. The coefficients of the short interest rates tend to have a negative sign, while their long term counterparts tend to be positive; this is also found in Solnik (1993). A negative sign of the short term interest rate is consistent with the interpretation that the interest rate is a proxy for business cycle effects in stock returns, as interest rates tend to be high at the peak and just after the peak of a business cycle. The coefficient of the rate of growth in the monetary base is only statistically significant in the first subperiod. The estimated value of the slope coefficient of the commercial paper-Treasury yield spread, which is an indicator for the default risk[2], is negative and statistically significantly different

[2] Adding the default premium on corporate bonds, i.e. the yield spread between Baa-rated and Aaa-rated corporate bonds, in the prediction model produced only marginally different results.

Table 2: **Linear Model for Excess Index Returns**

Regression of monthly stock index excess returns of the S&P 500 index on lagged explanatory variables for the periods 1954:1 to 1998:9, 1954:1 to 1976:5, and 1976:6 to 1998:9 respectively. The conditioning variables are the lagged price-earnings ratio (PE), lagged dividend yield (DIV), twice-lagged inflation rate ($INFL$), twice-lagged change in industrial production (IP), lagged and twice lagged short and long interest rate ($I3_{t-1}, I3_{t-2}, I12_{t-1}, I12_{t-2}$ respectively), twice lagged monetary growth variable (MB), and the lagged commercial paper-Treasury yield spread (CP). Heteroskedasticity corrected t-statistics are reported in square brackets. The values in parentheses, which correspond to the F-statistic, are p-values. Adj. R^2 denotes the R^2 adjusted for the degrees of freedom.

	1954:1-1998:9	1954:1-1976:5	1976:6-1998:9
Explanatory Variables*	$T = 537$	$T = 269$	$T = 268$
$Constant$	0.0918	-0.0203	0.1147
	[2.9881]	[-0.2277]	[3.1033]
PE_{t-1}	-0.2919	-0.1098	-0.3019
	[-3.0026]	[-0.4411]	[-2.5880]
DIV_{t-1}	-0.0460	0.2174	-0.0878
	[-0.9460]	[1.3593]	[-1.2579]
$INFL_{t-2}$	-0.1679	-0.3216	-0.1079
	[-1.8823]	[-1.9622]	[-0.7803]
IP_{t-2}	-0.1260	-0.0772	-0.1506
	[-3.0902]	[-1.1264]	[-2.2431]
$I3_{t-1}$	-0.0528	-0.1558	-0.0171
	[-1.0695]	[-1.1345]	[-0.3177]
$I3_{t-2}$	-0.1209	-0.1732	-0.1857
	[-1.6326]	[-1.0441]	[-1.9039]
$I12_{t-1}$	0.0099	-0.0136	0.1239
	[0.1217]	[-0.0978]	[1.1854]
$I12_{t-2}$	0.1252	0.2834	0.0425
	[2.5223]	[2.8259]	[0.8310]
MB_{t-2}	0.0855	0.4643	-0.0660
	[1.1872]	[3.4041]	[-0.4782]
CP_{t-1}	-0.3265	-0.5120	-0.3180
	[-3.0631]	[-3.0513]	[-2.0316]
R^2	0.128	0.230	0.114
Adj. R^2	0.112	0.200	0.080
DW	2.028	2.026	2.031
F-statistic	7.746 (F(10,526))	7.717 (F(10,258))	3.306 (F(10,257))
	(0.0000)	(0.0000)	(0.0005)

* For a more detailed description of the variables: see the Data Appendix.

Table 3: **Linear Model for Squared Unexpected Returns**

Regression of squared prediction errors on lagged squared prediction errors (SQ_{t-1}) and lagged short interest rate ($I3_{t-1}$) for the periods 1954:1 to 1998:9, 1954:1 to 1976:5, and 1976:6 to 1998:9 respectively. Heteroskedasticity and serial correlation corrected t-statistics with truncation lag 2 are reported in square brackets. The values in parentheses are p-values, which correspond to the F-statistic. Adj. R^2 denotes the R^2 adjusted for the degrees of freedom.

	1954:1-1998:9	1954:1-1976:5	1976:6-1998:9
Explanatory Variables*	$T = 537$	$T = 269$	$T = 268$
$Constant$	0.0009	0.0007	0.0013
	[5.1810]	[2.0066]	[2.9746]
SQ_{t-1}	0.1391	0.1591	0.1046
	[2.5385]	[2.0066]	[2.1128]
$I3_{t-1}$	0.0007	0.0012	0.0000
	[2.0025]	[1.2758]	[0.0521]
R^2	0.023	0.038	0.011
Adj. R^2	0.020	0.030	0.003
F-statistic	6.342 (F(2,534))	5.182 (F(2,266))	1.461 (F(2,265))
	(0.0020)	(0.0062)	(0.2339)

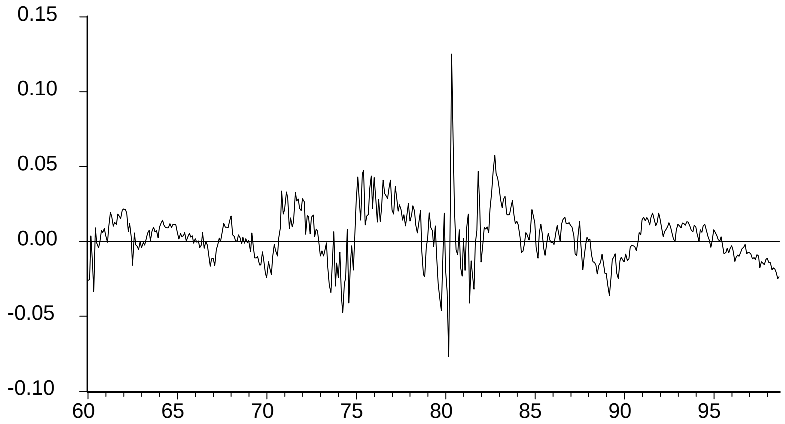
* For a more detailed description of the variables: see the Data Appendix.

from zero. We have not included any calendar effects in the model, like a January dummy, because these effects typically lack economic rationale and are potentially the mere result of data snooping. However, when we include a January dummy in the model the results are basically unchanged. Note that because the January effect is especially documented for small-cap stocks, its effect on the value-weighted stock index is expected to be small.

Table 3 reports the results from estimating the volatility model over the entire sample and the two subsamples. These are simply OLS results where the squared prediction error in (7) is explained from its lag and the short interest rate. To allow for autocorrelation and heteroskedasticity of unknown form, the t-statistics are based upon the heteroskedasticity and autocorrelation adjusted standard errors of Newey and West (1987). The lagged squared prediction error has a positive and significant impact on current volatility, which indicates that large monthly shocks in returns tend to be followed by large shocks. The estimated coefficient of the

Apparently, the impact of this variable is captured rather well by the commercial paper-Treasury yield spread.

Figure 2: Out-of-Sample Predicted Excess Returns S&P 500 Index (1960:1 - 1998:9).



lagged short term interest rate has a positive sign, but is only marginally significant when using the entire sample. Similar results were found in Breen, Glosten and Jagannathan (1989).

4 Nonparametric Tests for Market Timing

The empirical results in the previous section do not necessarily imply that there is any statistically significant out-of-sample predictability in index returns or volatility. To analyze the out-of-sample forecasting power of the linear model for the S&P 500 index, we use the predicted excess returns, $\hat{r}^e_{m,t}$, based on the recursive OLS estimates and thus using information up to time $t-1$ only. We assume that the investor starts investing in January 1960, such that he uses the information prior to 1960, i.e. from 1954:1–1959:12, to estimate his initial model. Consequently, we use the first 72 months as the base estimation period. Figure 2 displays the one-month ahead forecasts for the S&P 500 returns in excess of the risk free return. The figure shows a substantial degree of persistence in the predicted returns, with an estimated first order autocorrelation coefficient of 0.72. The number of zero crossings is limited, so that the switching strategy, as discussed above, involves reasonable long periods without switching. Over the entire period, the squared correlation coefficient between the predicted and actual excess returns is 6.3%, which is only half of the within-sample R^2 of 12.8%.

A more formal comparison of the predicted ($\hat{r}^e_{m,t}$) and actual excess returns ($r^e_{m,t}$) can be obtained in a variety of ways. First, we present in Table 4 the contingency table corresponding to the Henriksson and Merton (1981) (HM henceforth) test for market timing. This table is simply a cross-tabulation of the signs of $\hat{r}^e_{m,t}$ and $r^e_{m,t}$. Table 4 shows that, over the entire sample period, the sign of $r^e_{m,t}$ is predicted correctly in 266 out of 465 months (57.2%). The idea behind the HM test is that there is an indication of market timing if the sum of the (estimated) conditional probabilities of a correct forecast exceeds one (which makes it a one-tailed tests). Given the figures in Table 4 for the entire sample period, the estimated probability of a correct forecast conditional on a down market, p_1, equals $93/(93+104) = 0.4721$, and the probability of a correct forecast conditional on an up market, p_2, equals $173/(465-93-104) = 0.6455$. Consequently, the sum of the conditional probabilities of a correct forecast equals 1.1176, which exceeds unity, thus providing an indication of market timing ability. A formal way to test this is to use the nonparametric HM test statistic, which is asymptotically standard normally distributed under the null hypothesis. It is given by

$$HM = \frac{n_{11} - \frac{n_{10}n_{01}}{n}}{\sqrt{\frac{n_{10}n_{01}n_{20}n_{02}}{n^2(n-1)}}}, \tag{12}$$

where n_{11} is the number of correct bear market forecasts, n_{01}, n_{10} are the numbers of bear markets and bear market forecasts, respectively, while n_{02}, n_{20} denote the number of bull markets and bull market forecasts, respectively. The total number of evaluation periods is n. As the test statistic exceeds the one-sided 5% critical value of 1.64 for the entire period and both subperiods, the absence of market timing has to be rejected statistically. Note that the evidence for the last two decades is slightly weaker.

The HM test, testing whether $p_1 + p_2 = 1$ against the alternative that the sum exceeds unity, is asymptotically equivalent to a one-tailed test on the significance of the slope coefficient α_1 in[3]

$$I_{\{\hat{r}^e_{m,t}>0\}} = \alpha_0 + \alpha_1 I_{\{r^e_{m,t}>0\}} + \nu_t, \tag{13}$$

[3] To see this note that $\alpha_0 = P\{\hat{r}^e_{m,t} > 0 | r^e_{m,t} \leq 0\} \equiv 1 - p_1$, and $\alpha_1 = P\{\hat{r}^e_{m,t} > 0 | r^e_{m,t} > 0\} - P\{\hat{r}^e_{m,t} > 0 | r^e_{m,t} \leq 0\} \equiv p_1 + p_2 - 1$.

Table 4: **Nonparametric Market Timing Test**

This 2×2 contingency table is a cross-tabulation of the signs of $r^e_{m,t}$ and $\hat{r}^e_{m,t}$, which are obtained from recursive out-of-sample estimations. The Henriksson-Merton test for market timing has a standardized normal distribution in large samples. The critique value of the one-tailed test with a 5% significance level is 1.64. Figures correspond to the period 1960:1–1998:9 ($T = 465$); the figures between brackets correspond to two subperiods of 1960:1–1979:5 ($T = 233$) and 1979:6–1998:9 ($T = 232$) respectively.

	$r^e_{m,t} \leq 0$	$r^e_{m,t} > 0$	Total
$\hat{r}^e_{m,t} \leq 0$	93 (45; 48)	95 (33; 62)	188 (78; 110)
$\hat{r}^e_{m,t} > 0$	104 (66; 38)	173 (89; 84)	277 (155; 122)
Total	197 (111; 86)	268 (122; 146)	465 (233; 232)

Proportion of correctly predicted signs: 57.2% (57.5%; 56.9%)

Henriksson-Merton test: 2.5507 (2.1748; 1.9624)

where $I_{\{.\}}$ denotes the indicator function, equal to one if its argument is true and zero otherwise. Breen, Glosten and Jagannathan (1989), BGJ henceforth, conduct a very similar test, interchanging the role of the two variables in (13). That is, $I_{\{r^e_{m,t}>0\}}$ is the dependent variable, and $I_{\{\hat{r}^e_{m,t}>0\}}$ the independent variable. The corresponding t-statistic is exactly the same as the one obtained from (13). Cumbey and Modest (1987), CM henceforth, extend this test such that not just the sign of the realized excess return matters, but also its magnitude. This involves a regression of $r^e_{m,t}$ upon a constant and $I_{\{\hat{r}^e_{m,t}>0\}}$. Finally, Bossaerts and Hillion (1999), BH henceforth, use a regression of $r^e_{m,t}$ upon $\hat{r}^e_{m,t}$. In both cases the null hypothesis is that the slope coefficient is zero, which is tested against the one-sided alternative that it is positive.[4] Table 5 presents the results of these regression-based tests for market timing. As expected, the HM and BGJ tests result in statistics which are virtually the same as those reported in Table 4. The CM and BH tests, however, yield substantially larger tests statistics of 3.66 and 5.62 over the entire sample period, respectively. This means that the model is not only capable to predict the sign of the excess return, but also capable the order of magnitude.

Without exception, each of the tests indicates significant out-of-sample forecasting power of the recursive regression model for the S&P 500 index. This is

[4] Bossaerts and Hillion (1999) employ a two-sided alternative, which seems less appropriate.

consistent with a large body of recent literature, for example, Campbell (1987), Breen, Glosten and Jagannathan (1989), Pesaran and Timmermann (1995, 2000), but deviates from the conclusion of Bossaerts and Hillion (1999), who do not find any significant out-of-sample predictability for excess returns on the S&P 500 for the period 1990:6-1995:5. If we focus upon the most recent half of our sample period, the evidence for out-of-sample predictability is still clear, but somewhat less pronounced than for the first half. This is in line with the within-sample R^2s reported in Table 2.

The tests reported in Tables 4 and 5 do not take into account volatility timing. In fact, the different tests are not directly applicable to test for volatility timing because the conditional volatility is unobserved, even *ex post*. While the majority of volatility forecast evaluations in the literature rely upon the use of ex post squared (unexpected) returns, it is recently stressed by Andersen and Bollerslev (1998) that such a comparison is inappropriate and substantially underestimates the predictive performance of volatility models. The reason for this is that squared returns provide very noise measurements of actual volatility due to the idiosyncratic noise in the return process. To overcome this problem we essentially follow, among others, Schwert (1989) and Andersen and Bollerslev (1998) and construct a measure of ex post (monthly) volatility that is based on cumulative squared *daily* returns. Using high-frequency data to calculate the low-frequency ex post volatilities, substantially reduces the noise in the series, thus improving the ex post volatility measurements.

If daily returns do not exhibit any autocorrelation, the variance of the return in month t can be estimated as

$$\tilde{\sigma}_t^2 = \sum_{i=1}^{N_t} (r_{i,t} - \overline{r_t})^2, \tag{14}$$

where N_t is the number of trading days in month t, $r_{i,t}$ the return on day i in month t, and $\overline{r_t}$ denotes the average daily return in month t. Andersen and Bollerslev (1998) use a similar method to compute daily volatility using intradaily returns. It is well known, however, that daily stock index returns exhibit positive first-order autocorrelation. The most common explanation of this phenomenon is infrequent or non-synchronous trading of securities (see, e.g., Lo and MacKinlay, 1990). If daily returns are positively correlated, the estimator in (14) will underestimate the true volatility of monthly returns. Therefore we follow French, Schwert and

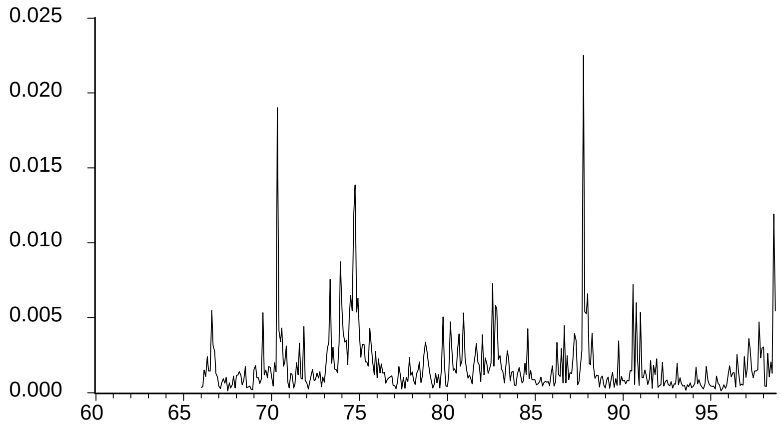
Table 5: **Regression-Based Tests for Market Timing**

Market timing tests based upon simple regressions, using the sign and level of actual excess returns and predicted excess returns, obtained from recursive out-of-sample estimations. The test statistics of the different versions of the test correspond to the t-statistics of the slope coefficient. A slope coefficient of zero corresponds to the case of no market timing ability. The t-statistics are given in parentheses. The first column presents the estimation results over the period 1960:1 to 1998:9; the second and third column correspond to the two subperiods of 1960:1–1979:5 and 1979:6–1998:9 respectively.

	1960:1-1998:9	1960:1-1979:5	1979:6-1998:9
Explanatory Variables	$T = 465$	$T = 233$	$T = 232$
	HM test*): $I_{\{\hat{r}^e_{m,t}>0\}} = \alpha_0 + \alpha_1 I_{\{r^e_{m,t}>0\}} + v_{1,t}$		
Constant	0.5279	0.5946	0.4419
	(15.1728)	(13.3545)	(8.2398)
$I_{\{r^e_{m,t}>0\}}$	0.1176	0.1349	0.1335
	(2.5660)	(2.1926)	(1.9746)
	BGJ test: $I_{\{r^e_{m,t}>0\}} = \delta_0 + \delta_1 I_{\{\hat{r}^e_{m,t}>0\}} + v_{2,t}$		
Constant	0.5053	0.4231	0.5636
	(14.0905)	(7.5263)	(12.2893)
$I_{\{\hat{r}^e_{m,t}\}>0}$	0.1192	0.1511	0.1249
	(2.5660)	(2.1926)	(1.9746)
	CM test: $r^e_{m,t} = \phi_0 + \phi_1 I_{\{\hat{r}^e_{m,t}>0\}} + v_{3,t}$		
Constant	-0.0036	-0.0091	0.0003
	(-1.2057)	(-1.9847)	(0.0825)
$I_{\{\hat{r}^e_{m,t}\}>0}$	0.0141	0.0166	0.0141
	(3.6631)	(2.9429)	(2.6304)
	BH test: $r^e_{m,t} = \theta_0 + \theta_1 \hat{r}^e_{m,t} + v_{4,t}$		
Constant	0.0029	-0.0021	0.0071
	(1.5121)	(-0.7516)	(2.6962)
$\hat{r}^e_{m,t}$	0.5827	0.7094	0.5377
	(5.6248)	(4.5321)	(3.8606)

*)It is easily checked that $\hat{\alpha}_0 = 1 - \hat{p}_1$, and $\hat{\alpha}_1 = \hat{p}_1 + \hat{p}_2 - 1$ (cf. footnote 4).

Figure 3: Actual Volatility (Estimated from Daily Data) (1966:1 - 1998:9).



Stambaugh (1987) and Akgiray (1989), and use an adjusted estimator for the variance of the return in month t, based upon the assuming that daily returns in month t are appropriately described by a first-order autoregressive process with coefficient ϕ_t. In particular, we can use the following estimator for the ex post monthly variance

$$\tilde{\sigma}_t^2 = \sum_{i=1}^{N_t}(r_{i,t} - \overline{r_t})^2 \left[1 + 2N_t^{-1} \sum_{j=1}^{N_t-1}(N_t - j)\hat{\phi}_t^j\right], \qquad (15)$$

where $\hat{\phi}_t$ is the first-order autocorrelation coefficient estimated using daily returns within month t.

To determine the ex post monthly variance we have data available of daily returns on the S&P 500 index from the beginning of 1966 to the end of 1998, which makes a total of 8453 observations (trading days). As a result, we can estimate the ex post volatility for a total of 393 months, starting in January 1966.[5] While this measure for monthly volatility is an estimate, we shall occasionally refer to it as the actual volatility, to contrast it with the predicted volatility according to the linear model in (10).

[5] The average estimated daily autocorrelation coefficient equals 0.12, and varies between -0.50 and 0.61. We also considered an alternative where the first-order autocorrelation coefficient was estimated over the entire sample period. This produced an estimate of 0.043. While this resulted in different values for the test statistics, these were not uniformly larger or smaller. Qualitatively, all conclusions remained the same.

Figure 4: Out-of-Sample Predicted Volatility (1960:1 - 1998:9).

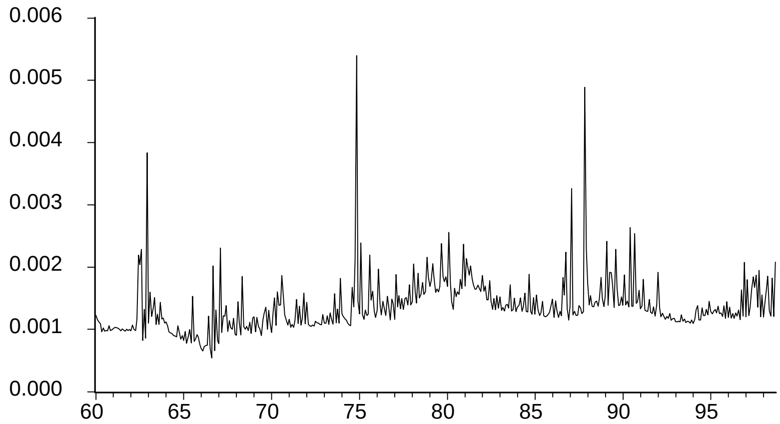


Figure 3 shows how actual ex post volatility varies over the period 1966:1–1998:9. In this figure, the volatility for each month is calculated according to formula (15). It is clear from the figure that the variance is not constant over time, and it tends to cluster: periods with high volatility are followed by some less volatility periods. Furthermore the crash of October 1987 is easily recognized, as well as the volatile period in 1998.

The out-of-sample predictions for the monthly variance of the S&P 500 returns are displayed in Figure 4. The figure shows some small swings in the predicted volatility combined with many high frequency changes. The squared correlation coefficient between the predicted and "actual" volatility figures is as low as 2.5%. This is partly due to the high variation in actual volatilities and its skewed distribution. On average, predicted volatilities are somewhat below the actual volatilities.[6]

To compare predicted and actual volatilities, we use the same range of tests as in Table 5. However, because volatilities are nonnegative, the benchmark in the contingency table has to be chosen at some positive number. In this case we choose a benchmark of $\xi = 0.0012$, corresponding to a monthly standard deviation of 3.46%. This number closely corresponds to the sample median of the actual volatility. The resulting dummy variables thus distinguish periods with "high" and periods with "low" (predicted) volatility. Note that the choice of benchmark

[6]This can be explained by the fact that predicted volatilities ignore estimation error in the predicted means (see below).

Table 6: **Regression-Based Tests for Volatility Timing**

Volatility timing tests based upon simple regressions, using the sign and level of actual volatility in excess of its benchmark ξ (0.0012, which is approximately the sample median) and predicted volatility in excess of its benchmark, obtained from recursive out-of-sample estimations. The test statistics of the different versions of the test correspond to the t-statistics of the slope coefficient. A slope coefficient of zero corresponds to the case of no volatility timing ability. The t-statistics are given in parentheses. The first column presents the estimation results over the period 1966:1 to 1998:9; the second and third column correspond to the two subperiods of 1966:1 to 1979:5 and 1979:6 to 1998:9 respectively.

Explanatory Variables	1966:1-1998:9 $T = 393$	1966:1-1979:5 $T = 161$	1979:6-1998:9 $T = 232$
HM volatility test: $I_{\{V\hat{a}r_{t-1}\{r^e_{m,t}\}>\xi\}} = \alpha_0 + \alpha_1 I_{\{\tilde{\sigma}^2_{m,t}>\xi\}} + v_{1,t}$			
Constant	0.6684	0.4231	0.8033
	(22.0760)	(7.6526)	(27.4372)
$I_{\{\tilde{\sigma}^2_{m,t}>\xi\}}$	0.1692	0.2275	0.1513
	(3.9567)	(2.9549)	(3.5577)
BGJ volatility test: $I_{\{\tilde{\sigma}^2_{m,t}>\xi\}} = \delta_0 + \delta_1 I_{\{V\hat{a}r_{t-1}\{r^e_{m,t}\}>\xi\}} + v_{2,t}$			
Constant	0.3299	0.3919	0.1724
	(6.6102)	(6.8852)	(1.9017)
$I_{\{V\hat{a}r_{t-1}\{r^e_{m,t}\}>\xi\}}$	0.2275	0.2288	0.3448
	(3.9567)	(2.9549)	(3.5577)
CM volatility test: $\tilde{\sigma}^2_{m,t} = \phi_0 + \phi_1 I_{\{V\hat{a}r_{t-1}\{r^e_{m,t}\}>\xi\}} + v_{3,t}$			
Constant	0.0012	0.0015	0.0008
	(5.6312)	(5.3912)	(2.0322)
$I_{\{V\hat{a}r_{t-1}\{r^e_{m,t}\}>\xi\}}$	0.0007	0.0010	0.0010
	(2.9032)	(2.6674)	(2.5480)
BH volatility test: $\tilde{\sigma}^2_{m,t} = \theta_0 + \theta_1 V\hat{a}r_{t-1}\{r^e_{m,t}\} + v_{4,t}$			
Constant	0.0007	0.0005	0.0007
	(1.7811)	(0.8407)	(1.3403)
$V\hat{a}r_{t-1}\{r^e_{m,t}\}$	0.8010	1.1669	0.6482
	(3.1465)	(2.9686)	(1.8815)

is irrelevant for the implementation of the BH test. The results of the different tests are reported in Table 6. For all of the reported samples, all tests soundly reject the null hypothesis of no predictability at the 5% level. The results indicate the presence of statistically significant volatility timing at a monthly frequency. This is in line with the findings of Fleming, Kirby and Ostdiek (1999), who find (economically) significant volatility timing using daily data. While volatility is much more persistent at daily than at monthly frequencies, some statistically significant predictability remains at the monthly frequency. The fact that the HM and BGJ tests provide higher values for the test statistics than the CM and BH tests, indicates that it is relatively more easy to predict whether volatility is high or low, relative to the median, than to predict its magnitude.

The two sets of tests above considered market timing in returns and volatility separately. While this provides useful information, it is possible that a good return forecast more than average corresponds to a bad volatility forecast or vice versa. It is therefore useful to consider the joint forecasting power of the two models. A first way to jointly test for the presence of timing in both moments, implies choosing a function of returns and volatility one is interested in. Instead of making some ad hoc choice, we decided to remain within the nonparametric framework and extend the above contingency tables to incorporate both dimensions simultaneously. To this end, we constructed a 4×4 table, in which we cross-tabulate the signs of the pairs $(\hat{r}^e_{m,t}, V\hat{a}r_{t-1}\{r^e_{m,t}\} - \xi)$ and $(r^e_{m,t}, \tilde{\sigma}^2_{m,t} - \xi)$, i.e. we compare the signs of the excess return and conditional variance in excess of its median, with their predicted counterparts. This novel approach provides important insights into the interdependencies of the predictability of the two components. Moreover, evidence of the joint predictive power can be obtained from this table that classifies observations according to the 16 different outcomes.

Table 7 presents the 4×4 contingency table and the test statistic corresponding to the nonparametric test for market timing in both moments. In this table the diagonal cells represent the correctly predicted pairs; from this we see that the signs of the pairs are predicted correctly in 132 out of 393 cases, i.e. 33.6% are classified correctly. Under the null of no predictability, we would expect a percentage of 25. A suitable test statistic in case the contingency table is larger than 2×2 is the generalized Henriksson and Merton test statistic for a $m \times m$ contingency table,

Table 7: **Nonparametric Market Timing Test for both Moments**

This 4 × 4 contingency table is a cross-tabulation of the pairs ($r^e_{m,t}$, $\tilde{\sigma}^2_{m,t} - \xi$) and ($\hat{r}^e_{m,t}$, $V\hat{a}r_{t-1}\{r^e_{m,t}\} - \xi$), which are obtained from recursive out-of-sample forecasts, and where $\xi = 0.0012$. The nonparametric test for market timing in first and second moment, which is a generalization of the Henriksson and Merton test statistic, is a χ^2 test with 9 degrees of freedom. The critique value of the one-tailed test with a 5% significance level is 16.92. Figures correspond to the period 1966:1–1998:9 ($T = 393$).

	$r^e_{m,t} \leq 0$, $\tilde{\sigma}^2_{m,t} > \xi$	$r^e_{m,t} \leq 0$, $\tilde{\sigma}^2_{m,t} \leq \xi$	$r^e_{m,t} > 0$, $\tilde{\sigma}^2_{m,t} > \xi$	$r^e_{m,t} > 0$, $\tilde{\sigma}^2_{m,t} \leq \xi$	Total
$\hat{r}^e_{m,t} \leq 0$, $V\hat{a}r_{t-1}\{r^e_{m,t}\} > \xi$	47	18	30	36	131
$\hat{r}^e_{m,t} \leq 0$, $V\hat{a}r_{t-1}\{r^e_{m,t}\} \leq \xi$	15	7	6	12	40
$\hat{r}^e_{m,t} > 0$, $V\hat{a}r_{t-1}\{r^e_{m,t}\} > \xi$	33	25	49	52	159
$\hat{r}^e_{m,t} > 0$, $V\hat{a}r_{t-1}\{r^e_{m,t}\} \leq \xi$	5	21	8	29	63
Total	100	71	93	129	393

Proportion of correctly predicted pairs: 33.6% (under H_0: 25%)

χ^2-test: 42.5544 ($\chi^2_{9;0.05} = 16.92$)

given by

$$HM = \sum_{i,j=1}^{m} \frac{\left(n_{ij} - \frac{n_{i0}n_{0j}}{n}\right)^2}{\frac{n_{i0}n_{0j}}{n}}, \tag{16}$$

where m is the number of categories, n_{ij} the number of observations in the category (i, j), and n_{i0} and n_{0j} are the ith row and the jth column totals. Under the null hypothesis, the test statistic is asymptotically Chi-squared distributed with $(m - 1)^2 = 9$ degrees of freedom. The realization of this statistic is 42.55 (with a 5% critical value of 16.92), which clearly shows that there is a significant relationship between the predicted and realized pairs. The null hypothesis that is actually tested here is independence between forecasts and realizations. Strictly speaking, the fact that the test rejects does not imply *positive* timing ability, just that there is dependence between forecasts and realizations.[7]

To evaluate the question whether the rejection of the test is due to positive timing ability, we explore several alternative tests based upon auxiliary regres-

[7] Recall that in the 2 × 2 case, the alternative hypothesis was one-sided, corresponding to positive timing ability.

sions. This allows us to focus the tests upon combinations of predicted signs and actual signs which are particularly interesting. To do so, we define two sets of dummy variables. The first set includes four dummy variables corresponding to four different outcomes in the columns of Table 7. That is,

$$\begin{aligned} I_{A,t} &= 1 \quad \text{if} \quad r^e_{m,t} \leq 0 \text{ and } \tilde{\sigma}^2_{m,t} > \xi, \quad 0 \text{ otherwise}, \\ I_{B,t} &= 1 \quad \text{if} \quad r^e_{m,t} \leq 0 \text{ and } \tilde{\sigma}^2_{m,t} \leq \xi, \quad 0 \text{ otherwise}, \\ I_{C,t} &= 1 \quad \text{if} \quad r^e_{m,t} > 0 \text{ and } \tilde{\sigma}^2_{m,t} > \xi, \quad 0 \text{ otherwise}, \\ I_{D,t} &= 1 \quad \text{if} \quad r^e_{m,t} > 0 \text{ and } \tilde{\sigma}^2_{m,t} \leq \xi, \quad 0 \text{ otherwise}. \end{aligned} \tag{17}$$

The second set includes four dummies corresponding to the different row outcomes, and are denoted as $I_{\hat{A},t}$, $I_{\hat{B},t}$, $I_{\hat{C},t}$ and $I_{\hat{D},t}$, respectively. They are defined in a similar way replacing actual outcomes by predicted outcomes. Now consider the following set of equations

$$\begin{aligned} I_{A,t} &= \alpha_{10} + \alpha_{11} I_{\hat{A},t} + \alpha_{12} I_{\hat{B},t} + \alpha_{13} I_{\hat{C},t} + v_{A,t}, \\ I_{B,t} &= \alpha_{20} + \alpha_{21} I_{\hat{A},t} + \alpha_{22} I_{\hat{B},t} + \alpha_{24} I_{\hat{D},t} + v_{B,t}, \\ I_{C,t} &= \alpha_{30} + \alpha_{31} I_{\hat{A},t} + \alpha_{33} I_{\hat{C},t} + \alpha_{34} I_{\hat{D},t} + v_{C,t}, \\ I_{D,t} &= \alpha_{40} + \alpha_{42} I_{\hat{B},t} + \alpha_{43} I_{\hat{C},t} + \alpha_{44} I_{\hat{D},t} + v_{D,t}. \end{aligned} \tag{18}$$

To prevent perfect multicollinearity, in each equation the dummy variable corresponding to the forecast with two incorrect signs is omitted. The coefficients α_{jj}, $j = 1, 2, 3, 4$, correspond to the increase in the probability of observing a given pair of outcomes if the recursive models predict these two outcomes, relative to the case where the two opposite outcomes are predicted. Similarly, the coefficients α_{ij}, $i \neq j$, $j = 1, 2, 3, 4$, measure the increase in probability due to forecasting only one given outcome correctly.

The system of equations in (18) is singular, because $v_{A,t}+v_{B,t}+v_{C,t}+v_{D,t} = 1$ for each t. This means that the coefficients in one equation can be expressed as linear functions of those in the other three equations.[8] When the system is estimated by the seemingly unrelated regressions (SUR) estimator, this means that either equation can be dropped with equivalent results (Barten, 1969). However, for ease

[8]It can be shown that the following restrictions hold: $\alpha_{40} = 1-\alpha_{10}-\alpha_{20}-\alpha_{30}-\alpha_{11}-\alpha_{21}-\alpha_{31}$; $\alpha_{42} = \alpha_{11} + \alpha_{21} + \alpha_{31} - \alpha_{12} - \alpha_{22}$; $\alpha_{43} = \alpha_{11} + \alpha_{21} + \alpha_{31} - \alpha_{13} - \alpha_{33}$; and $\alpha_{44} = \alpha_{11} + \alpha_{21} + \alpha_{31} - \alpha_{24} - \alpha_{34}$.

Table 8: **Regression-Based Tests for Market Timing in both Moments**

The tests for market timing, which provides a joint test of predictability in both moments, are based upon SUR (OLS) of $I_{i,t} = \alpha_{j0} + \alpha_{j1} I_{\hat{A},t} + \alpha_{j2} I_{\hat{B},t} + \alpha_{j3} I_{\hat{C},t} + \alpha_{j4} I_{\hat{D},t} + v_{i,t}$, $i = A, B, C, D$, $j = 1, 2, 3, 4$. In each equation the dummy that corresponds to the forecasts with two incorrect signs is omitted. The t-statistics are given in parentheses. The predictive failure test is a χ^2 test with 9 degrees of freedom. Estimation results over the period 1966:1–1998:9.

Explanatory Variable	$I_{A,t}$	$I_{B,t}$	$I_{C,t}$	$I_{D,t}$
Constant	0.0794	0.1572	0.1500	0.2748
	(1.4898)	(5.2349)	(2.2622)	(6.7227)
$I_{\hat{A},t}$	0.2794	-0.0198	0.0790	-
	(4.3100)	(-0.4437)	(1.0429)	
$I_{\hat{B},t}$	0.2956	0.0178	-	0.0252
	(3.4583)	(0.2652)		(0.2981)
$I_{\hat{C},t}$	0.1282	-	0.1582	0.0522
	(2.0363)		(2.1324)	(0.9462)
$I_{\hat{D},t}$	-	0.1761	-0.0230	0.1855
		(3.1234)	(-0.2715)	(2.5861)
Wald test: 42.0026 ($\chi^2_{9;0.05} = 16.92$)				

of interpretation we present results for the full system. Note that because of the nature of the regressors OLS applied to each equation is numerically equivalent to SUR.[9] The SUR (OLS) estimates for the above system, for the period January 1966 to September 1998, are presented in Table 8.

The null hypothesis of independence implies that all partial slope coefficients in (18) are equal to zero, which corresponds to 9 linearly independent restrictions. The standard Wald test for these restrictions produces a value of 42.00, so that the null hypothesis that both returns and volatility are not predictable is soundly rejected. Note that the Wald statistic based on (18) is very close to the value of the non-parametric test statistic of 42.55, given in Table 7, which is no surprise given that these tests are asymptotically equivalent.

Ideally, the diagonal coefficients α_{11} to α_{44} are positive (and significant), indicating positive timing abilities. Note that three out of four diagonal elements are significantly positive. If we jointly test the set of restrictions $\alpha_{11} = \alpha_{22} = \alpha_{33} = \alpha_{44} = 0$, we obtain a test statistic of 24.16, which is again highly significant for a

[9]It is well-known result that OLS and SUR provide identical estimates in case each equation in the system has the same regressors. This result also holds if regressors are different across equations but span the same space, as is the case in (18); see Greene (2000, Sect. 15.4).

Chi-squared distribution with 4 degrees of freedom. When considering the coefficients in Table 8 a desirable feature is that, in each of the equations, the dummy for the best pair of forecasts has a coefficient that is larger than the other two dummies. This means that the probability of a given pair of outcomes is larger if both signs are correctly predicted than if only one sign is correctly predicted. The most notable violation of this implication is that the estimate for α_{24} exceeds that of α_{22}, indicating that the model is not very well capable to time whenever simultaneously excess returns are negative and volatility is low ($r^e_{m,t} \leq 0$ and $\tilde{\sigma}^2_{m,t} \leq \xi$), which – in this sample period – occurs in 18% of the months. For the other combinations the simultaneous timing is reasonably well. Further, we tested whether the probability of a correct return forecast is independent of the probability of a correct volatility forecast. This resulted in a test statistic of 0.302, which is highly insignificant for a standard normal distribution. Apparently, the fact that the sign of the excess return is predicted correctly does not increase or decrease the probability that the sign of volatility (relative to 0.0012) is predicted correctly (or vice versa). Finally, Pesaran and Timmermann (1995) suggest that the predictability of excess returns is larger at times when volatility is high. From the above results, we can easily test the null hypothesis of independence between a correctly signed excess return forecast and volatility, as measured by the high-low dummy. This results in a value of 2.238, which implies sound rejection based on a standard normal distribution. Apparently, periods with larger shocks more than average correspond to periods with a correctly predicted up- or down-market. As noted by Pesaran and Timmermann (1995) this finding could be consistent with incomplete learning after a large shock to the economy, as well as with the existence of time-varying risk premia.

Overall, the results in this section indicate the presence of statistically significant out-of-sample predictability for both S&P 500 returns and its volatility. For predicting returns, the evidence is typically somewhat stronger for the period up to May 1975 than for the period after this date. Pesaran and Timmermann (1995) also find that forecasting performance is relatively low in the 1980s. For predicting volatility the strength of the statistical evidence depends upon the question whether the evaluation takes into account the actual levels or just the signs (and forecasts refer to periods of “high” and “low” volatility). Overall, it appears somewhat harder to predict the actual volatility than to predict its sign.

None of the tests in this section requires the implementation of any particular trading strategy. This way, these tests do not allow for the incorporation of transaction costs or short-sales restrictions or the comparison of alternative strategies. In a sense, one could interpret the rejection of some of the above tests as a necessary condition for any economically significant predictability. To elaborate upon the economic value of predicting returns and volatility we shall, in the next section, analyze the results of a number of trading strategies based upon return and volatility predictions from our two recursive models.

5 Evaluating Empirical Trading Rules

In this section we consider the mean-variance investor from Section 2, who uses recursive OLS estimates of model (7) and (10) to determine his optimal portfolio weights. This results in a portfolio that is estimated to be *ex ante* conditional mean-variance efficient. We assume that the investor has a risk aversion coefficient γ of 6. This choice corresponds with a moderately risk-averse investor that, in a world without conditioning information, would invest about 50% in the risky asset. Most of the results below (e.g. Sharpe ratios) are insensitive to the choice of γ, while occasionally we consider alternative levels of risk aversion as well. Similar to Section 4, the investor starts investing in January 1960, such that he initially uses the information prior to 1960, i.e. from 1954:1-1959:12. This means that the base estimation period consists of 72 months. Besides the optimal trading rule and the switching rule, as described in Section 2, we will also consider three simple benchmark strategies: holding only the market portfolio, holding only the risk free asset, and holding 50% of the wealth in the market portfolio and 50% in the risk free asset.[10] This latter strategy is especially interesting for the utility-based measure, as the average weight of most of the dynamic strategies corresponds more or less with this strategy. The optimal portfolio is evaluated with and without tak-

[10] Alternatively, we could compare the results of the active strategies with a so-called uninformed strategy. This strategy, which is also an active strategy, uses only past returns to construct the expected return and volatility in a recursive way. The results of this strategy are not reported here, as this strategy yielded an underperformance relative to the strategies with 50 and 100 percent of the wealth in the market portfolio. Consequently, the economic value of the active (informed) strategies is a conservative estimate when compared to the strategies with 50 and 100 percent in the market portfolio.

Table 9: **Overview of Passive and Active Strategies**

Strategy	Description
Passive:	
I: 100% market	100% is invested in the market portfolio
II: 50% market	50% is invested in the market portfolio, and 50% in the risk free asset
III: 0% market	100% is invested in the risk free asset
Active:	
IV: Switching (0-1)	100% is invested in the market portfolio if $E_t\{r^e_{m,t+1}\} \geq 0$, and 0% otherwise
V: Optimal μ	optimal portfolio as in (4) with conditional volatility assumed constant
VI: Optimal μ, σ	optimal portfolio as in (4) (with volatility timing)
VII: Optimal μ (0-1)	optimal portfolio with short-sale and borrowing constraints as in (5) with conditional volatility assumed constant
VIII: Optimal μ, σ (0-1)	optimal portfolio with short-sale and borrowing constraints as in (5) (with volatility timing)

ing into account future volatility, and with and without short-sale and borrowing restrictions.[11] For convenience, Table 9 presents an overview of the strategies that are used in this section.

To obtain some idea about the different trading rules, first note that the average weight of the S&P 500 index in the switching portfolio is 0.596. This also follows directly from Table 4. For the unrestricted mean-variance portfolios using predicted returns and predicted returns and volatility, the average weights are 0.188 and 0.494, respectively. The variation in the weights however, is much higher than for the switching strategy, particularly when volatility timing is used. In some periods the weights of the unrestricted mean-variance portfolios imply substantial short positions in the market index or in the risk free asset. If we restrict the weights to be between 0 and 1, the average weights for the last two strategies become 0.363 and 0.501, respectively. For each of the strategies the portfolio returns are determined. Table 10 presents the average returns and their estimated standard deviations over the sample period and the two subsamples, and a number of additional statistics for investment performance and market timing.

[11]Note that introducing short-sale and borrowing restrictions does not affect the switching portfolio in (6).

Table 10: **Evaluation of Various Trading Rules**

The results are based on recursive least squares using various trading rules. Mean and Std. dev. denote the mean return and the standard deviation of the return on the corresponding strategy in %, respectively. The optimal strategies are based on a risk aversion coefficient of 6 (see Table 9 for more details). The Sharpe ratio equals the average excess return of the strategy divided by the sample standard deviation. Alpha, is the OLS estimate of the intercept in the regression of excess return of the strategy on the excess return of the S&P500 index. The t-statistics for this coefficient are given in parentheses. The Treynor-Mazuy (TM) test static is the t-value of the squared return coefficient in a regression of the excess returns upon a constant, the excess return and the squared excess return of the S&P500 index. Panel A refers to the period 1960:1-1998:9, while panel B and C refer to the two subperiods of 1960:1 to 1979:5 and 1979:6 to 1998:9 respectively.

	Mean	**Std. dev.**	**Sharpe**	**Alpha**		**TM test**
		Panel A: 1960:1-1998:9 ($T = 465$)				
Passive:						
I: 100% market	0.9697	4.1191	0.1172		–	–
II: 50% market	0.7284	2.0565	0.1174		–	–
III: 0% market	0.4870	0.2088	0		–	–
Active:					(t-stat.)	(t-stat.)
IV: Switching (0-1)	1.1148	2.7656	0.2270	0.0041	(4.2947)	(7.4211)
V: Optimal μ	1.6185	4.9949	0.2265	0.0114	(4.9010)	(9.2325)
VI: Optimal μ, σ	2.9199	10.8453	0.2243	0.0242	(4.7788)	(9.5158)
VII: Optimal μ (0-1)	0.9083	2.0740	0.2031	0.0028	(3.5662)	(7.9551)
VIII: Optimal μ, σ (0-1)	1.0112	2.5109	0.2088	0.0034	(3.7296)	(7.5733)
		Panel B: 1960:1-1979:5 ($T = 233$)				
Passive:						
I: 100% market	0.5967	4.0918	0.0466		–	–
II: 50% market	0.5014	2.0354	0.0463		–	–
III: 0% market	0.4061	0.1423	0		–	–
Active:					(t-stat.)	(t-stat.)
IV: Switching (0-1)	0.9015	2.8854	0.1717	0.0040	(3.0082)	(2.1467)
V: Optimal μ	1.5446	4.4654	0.2550	0.0112	(3.8539)	(7.1187)
VI: Optimal μ, σ	3.0319	10.5041	0.2500	0.0258	(3.7655)	(6.9675)
VII: Optimal μ (0-1)	0.7692	2.2045	0.1647	0.0030	(2.6643)	(3.7123)
VIII: Optimal μ, σ (0-1)	0.8342	2.6429	0.1620	0.0034	(2.6977)	(2.3150)
		Panel C: 1979:6-1998:9 ($T = 232$)				
Passive:						
I: 100% market	1.3444	4.1213	0.1883		–	–
II: 50% market	0.9563	2.0567	0.1887		–	–
III: 0% market	0.5683	0.2324	0		–	–
Active:					(t-stat.)	(t-stat.)
IV: Switching (0-1)	1.3291	2.6287	0.2894	0.0045	(3.2711)	(7.7768)
V: Optimal μ	1.6927	5.4842	0.2050	0.0122	(3.3587)	(5.8924)
VI: Optimal μ, σ	2.8075	11.1991	0.1999	0.0236	(3.1481)	(6.4289)
VII: Optimal μ (0-1)	1.0481	1.9288	0.2487	0.0027	(2.5860)	(7.2088)
VIII: Optimal μ, σ (0-1)	1.1890	2.3633	0.2626	0.0035	(2.7911)	(7.9477)

Let us first consider the results for the full evaluation period of January 1960 to September 1998. The average monthly return on the S&P 500 index over this period is 0.97% with an estimated (unconditional) standard deviation of 4.12%, while the average risk free return is 0.49%. The standard deviation of the risk free return is 0.21%. To see whether a trading rule outperforms another one we will first consider the Sharpe ratio (SR; see Sharpe, 1966), defined as the ratio of the mean excess return on the (managed) portfolio and the standard deviation of the portfolio return. For example, if a strategy's SR exceeds the market SR the active portfolio dominates (in an unconditional mean-variance sense) the market portfolio. In general, the SR answers the question whether it is attractive for a mean-variance investor to invest all his wealth in this active portfolio, apart from a fraction in the risk free asset. For empirical applications, the (*ex post*) Sharpe ratio is usually estimated as the ratio of the sample mean of the excess return on the portfolio and the sample standard deviation of the portfolio return. The SR of the market index is 0.117, while the SR of the risk free asset is zero by construction.

Next, consider the returns of the switching strategy (Strategy IV). In this strategy wealth is completely allocated to the risky asset if the predicted excess return is positive and to the risk free asset otherwise. The average return of this strategy is 1.11% per month with a standard deviation of 2.77%. Apparently, this strategy is able to increase average returns while reducing risk, which is reflected in the Sharpe ratio of 0.227.

A common way to measure the outperformance or "abnormal" return of an investment strategy is provided by Jensen's (1968) alpha, which is the estimated intercept in a regression of excess returns upon market excess returns, i.e.

$$r_{p,t}^{e} = \alpha_p + \beta_p r_{m,t}^{e} + \varepsilon_{p,t}, \tag{19}$$

where $r_{p,t}^{e}$ denotes the excess return on the portfolio. A positive value for Jensen's alpha indicates that it is reasonable to invest some (positive) part of your wealth in this portfolio. This measure is equivalent to the more recent unconditional stochastic discount factor performance measures in which no particular model is assumed (see, e.g., Chen and Knez, 1996, or Söderlind, 1999). For the switching strategy, Jensen's alpha corresponds to an abnormal return of 0.41% per month, which is statistically highly significant. Note, however, that Jensen's alpha is known to suffer from biases in the presence of market timing (see Dybvig and Ross, 1985 and Grinblatt and Titman, 1989). These biases arise because a successful market timing strategy implies that beta varies over time, as a function of market returns, so that inference based on OLS in (19) is invalid.

A simple way to avoid this bias, is proposed by Treynor and Mazuy (1966). Assume that beta depends linearly on the conditional expected market excess return, i.e. $\beta_p = \varphi_0 + \varphi_1 E_t\{r^e_{m,t}\}$, such that β_p varies over time. We can obtain the Treynor and Mazuy's measure by substituting this into (19), which yields the following regression equation:[12]

$$r^e_{p,t} = \alpha_p + \varphi_0 r^e_{m,t} + \varphi_1 [r^e_{m,t}]^2 + v_{p,t}. \tag{20}$$

A positive φ_1 coefficient in (20) is an indication of successful market timing. Henriksson and Merton (1981) suggest to treat (20) as a standard regression model. The resulting Treynor and Mazuy measures (the t-statistics of the estimate for φ_1) are reported in Table 10. According to this measure the switching portfolio again significantly outperforms any passive investment strategy.

Next, we consider strategies V to VIII based upon the (estimated) optimal mean-variance weights. These strategies differ in the amount of information that is exploited and the restrictions that are imposed. Strategy V ignores any information contained in the volatility model and only exploits the return predictions from the recursive model, while Strategy VI uses information from both models. The average return on the optimal portfolio using predictability in first moments only is 1.62%, which increases to 2.92% if predictability in second moments is exploited as well. The unconditional standard deviations of these returns are 4.99% and 10.84%, respectively, so that the Sharpe ratios of these two strategies are only 0.227 and 0.224. It should be noted that the portfolio weights for these mean-variance strategies without short-sales constraints are occasionally extreme.[13] This is often profitable, for example in October 1987 these two strategies imply negative portfolio weights of −0.581 and −1.358, respectively, while the actual excess return on

[12] Glosten and Jagannathan (1994) explore the inclusion of other nonlinear functions of the excess return on the market index based on option-like pay offs. A simple case, suggested by Henriksson and Merton (1981) involves the inclusion of $\max\{r^e_{m,t}, 0\}$. Using this term rather than the quadratic term in (20) results in values for the t-test statistics that are very similar to the ones reported.

[13] Although the computed weights occasionally imply substantial leverage or short positions, they are not unusual for a typical hedge fund (see, e.g., Fung and Hsieh, 1997).

the S&P 500 index was -22% and "actual" volatility was 15%. However, investors often face short-sales constraints or would otherwise prefer dynamic strategies that involve less trading (or less trading costs). Therefore, we also consider the "optimal" strategies in which we impose *a priori* that the portfolio weights are to be within the [0,1] interval. For the resulting strategies, VII and VIII, the Sharpe ratios reduce further to 0.203 and 0.209, respectively. While this implies that each of the dynamic strategies outperforms the static strategies, it also suggests that the simple switching strategy outperforms any of the "optimal" strategies. However, the risk of the dynamic strategies is typically overestimated by the sample standard deviation, particularly in the presence of volatility timing, because the ex post (unconditional) standard deviation is an inappropriate measure for the (conditional) risk an investor was facing at each point in time. Note, for example, that the unconditional standard deviation of the risk free asset is 0.22%, while the actual risk in any given month is zero. This indicates a potentially severe disadvantage of the use of Sharpe ratios to evaluate dynamic strategies. Below we shall therefore compute an alternative measure for the economic gains involved in a dynamic trading strategy that takes this problem into account.

Considering Jensen's alphas for the "optimal" strategies V to VIII, as reported in Table 10, we see that their estimates are significantly positive for each of the dynamic trading rules. While the t-values are of the same order of magnitude across the different strategies, the estimates are quite different. For the "optimal" portfolios without short-sales constraints, Jensen's alpha is much larger than for the switching strategy. When short-sales restrictions are imposed, the Jensen measures of the optimal portfolios are, however, smaller.

Looking at the t-values of the Treynor and Mazuy measures it appears that these are all exceeding the 5% critical value, indicating the presence of statistically significant market timing for each of the dynamic trading strategies for all subperiods. For the full sample period and for the subperiod 1960–1979, the order of magnitude of the Treynor and Mazuy's measures is higher for the optimal trading rule (Strategy VI) than for the switching rule. For the more recent subperiod starting in 1979, the opposite appears to be the case. Except for the latter subperiod, introducing short-sale and borrowing constraints lowers the values of these measures.

For the last three years, as can be seen from Figure 2, the predicted excess returns on the S&P 500 index were continuously negative. This is mainly caused by the very high price-earnings ratios and low dividend yields over these years. For the active strategies this implies zero or negative investments in the stock market. Of course, with hindsight and knowing the huge increase in US stock prices over these years, one can conclude that these strategies did not do very well over the last part of our sample period. However, this does not necessarily mean that these strategies were ex ante unattractive. In the future it remains to be seen whether stock prices return to levels that can be rationalized with reasonable dividend or earnings forecasts (as predicted by Campbell, 2000, and Shiller, 2000), more or less consistent with our historical forecasting model, or whether structural changes have occurred that invalidate the use of forecasting models estimated on the basis of fairly long historical data sets. In the latter case, the specification of a forecasting model should be more dynamic allowing changing sets of explanatory variables over time combined with a moving window to estimate the parameters (see, for example, Pesaran and Timmermann, 2000), or allowing for structural breaks.

Until now, the issue of transaction costs was ignored. Clearly, in determining the economic value of any dynamic trading strategy transaction costs play an important role. To examine the economic value of the trading rules when transaction costs are present, we assume that the transaction costs are equal to τ percentage points of the value traded, such that the transaction costs equal (in money terms): $\tau W_t|\Delta w_{t+1}|$,where W_t denotes the wealth at time t, and $\Delta w_{t+1} = w_{t+1} - w_t$. Consequently the return after transaction costs is equal to $r_{p,t+1} - \tau|\Delta w_{t+1}|$. To see this note that the wealth at time $t+1$ is equal to $W_{t+1} = W_t r_{p,t+1} - \tau W_t|\Delta w_{t+1}| = W_t(r_{p,t+1} - \tau|\Delta w_{t+1}|)$. Table 11 presents the performance measures and tests for the each of the strategies in the presence of transaction costs. We apply three scenarios with low, medium and high transaction costs of 0.1%, 0.5% and 1% of the value of the trade, respectively. For a passive strategy the inclusion of transaction costs does not matter.[14]

[14] This assumption is a simple approximation and ignores the costs associated with rebalancing the portfolio. This may result in a slight overestimation of the returns of "passive" strategies I and II, consisting of the market portfolio, and the portfolio holding 50% risky asset and 50% riskless asset respectively. In practice, to hold the market portfolio one needs to regularly buy and sell shares to obtain the right proportions of the portfolio. More specifically, holding the market portfolio is contrarian. That is, winning assets must be sold and losing assets must be purchased to keep the portfolio weights constant from period to period. We however neglect this

Table 11: **Evaluation of Various Trading Rules in Presence of Transaction Costs***

Transaction costs are introduced into the trading environment. The transaction costs are either low: 0.1% (Panel A); medium: 0.5% (Panel B); or high: 1% of the value traded (Panel C). Figures correspond to the period 1960:1-1998:9 ($T = 465$).

	Mean	**Std. dev.**	**Sharpe**	**Alpha**		**TM test**
Panel A: Low Transaction Costs (0.1%)						
Passive:						
I: 100% market	0.9697	4.1191	0.1172		–	–
II: 50% market	0.7284	2.0565	0.1174		–	–
III: 0% market	0.4870	0.2088	0		–	–
Active:					(t-stat.)	(t-stat.)
IV: Switching (0-1)	1.0991	2.7659	0.2213	0.0039	(4.1285)	(7.4357)
V: Optimal μ	1.5733	4.9733	0.2184	0.0109	(4.7244)	(9.2423)
VI: Optimal μ, σ	2.8152	10.7986	0.2136	0.0231	(4.5892)	(9.5249)
VII: Optimal μ (0-1)	0.8936	2.0742	0.1960	0.0026	(3.3750)	(7.9660)
VIII: Optimal μ, σ (0-1)	0.9941	2.5107	0.2020	0.0032	(3.5394)	(7.5887)
Panel B: Medium Transaction Costs (0.5%)						
Passive:						
I: 100% market	0.9697	4.1191	0.1172		–	–
II: 50% market	0.7284	2.0565	0.1174		–	–
III: 0% market	0.4870	0.2088	0		–	–
Active:					(t-stat.)	(t-stat.)
IV: Switching (0-1)	1.0364	2.7717	0.1982	0.0033	(3.4529)	(7.4674)
V: Optimal μ	1.3927	4.8933	0.1851	0.0091	(3.9979)	(9.2675)
VI: Optimal μ, σ	2.3965	10.6256	0.1797	0.0189	(3.8095)	(9.5475)
VII: Optimal μ (0-1)	0.8349	2.0773	0.1675	0.0020	(2.6064)	(7.9950)
VIII: Optimal μ, σ (0-1)	0.9257	2.5136	0.1745	0.0025	(2.7719)	(7.6299)
Panel C: High Transaction Costs (1%)						
Passive:						
I: 100% market	0.9697	4.1191	0.1172		–	–
II: 50% market	0.7284	2.0565	0.1174		–	–
III: 0% market	0.4870	0.2088	0		–	–
Active:					(t-stat.)	(t-stat.)
IV: Switching (0-1)	0.9579	2.7896	0.1688	0.0025	(2.5969)	(7.4481)
V: Optimal μ	1.1669	4.8086	0.1414	0.0068	(3.0472)	(9.2650)
VI: Optimal μ, σ	1.8731	10.4413	0.1327	0.0136	(2.7901)	(9.5415)
VII: Optimal μ (0-1)	0.7615	2.0861	0.1316	0.0013	(1.6648)	(7.9979)
VIII: Optimal μ, σ (0-1)	0.8401	2.5252	0.1398	0.0016	(1.8084)	(7.6349)

* See the notes to Table 10.

From Table 11, it appears that the average returns for each of the dynamic strategies decrease substantially when transaction costs increase. Not surprisingly, this is especially the case for the optimal strategy that also takes into account future volatility.[15] This strategy yields high transaction costs since it entails some extreme weights. The standard deviations are not much influenced by the transaction costs. Consequently, the Sharpe ratios also decrease rapidly when transaction costs increase. However, even for high transaction costs, each of the dynamic strategies outperform the (passive) market portfolio in terms of the Sharpe ratio. Furthermore, the Jensen's alphas and their corresponding t-statistics decrease monotonically. Finally, the Treynor-Mazuy measure is not very sensitive to the transaction costs. This is because it merely tests the ability to time to marketrather than to measure the economic value of it. In general we can conclude that our findings are robust to reasonable transaction costs. All active strategies outperform the passive ones, and the ranking of the alternative trading strategies do not change.

None of the three measures above provides an accurate estimate for the economic value of return and volatility timing. While the Sharpe ratio overestimates the risk in case of time-varying volatility, Jensen's alpha suffers from biases in case of time-varying portfolio weights (betas). The Treynor-Mazuy measure is at best a test for the existence of market timing (in beta), but does not provide an estimate of its economic value. To compare the economic value of the different trading strategies we go back to our mean-variance investor and ask the question: how much would this investor be willing to pay to switch from a passive strategy to a given dynamic strategy. In a recent paper, Fleming, Kirby and Ostdiek (1999) use a utility-based measure to determine the economic value of a dynamic strategy based on volatility timing (of daily returns) relative to a passive strategy. We follow a similar approach and calculate the maximum fee, in terms of percentage per month, an investor with a given level of risk aversion, would be willing to pay

fact, and consequently all predictability results using dynamic trading rules (where transaction costs are present) compared to static ones are conservative.

[15] Note that the mean-variance portfolios are no longer optimal in the presence of transaction costs. While trading rules could be adjusted by taking into account future transaction costs in the portfolio allocation, we decided not to pursue this. In this sense, our estimates of the economic value are again conservative.

Table 12: **Average Realized Utilities**

The average realized utilities, obtained using formula (23), for different values of the risk aversion coefficient are presented in percentages.

1966:1-1998:9	$\gamma = 2$	$\gamma = 6$	$\gamma = 12$	$\gamma = 20$	Ranking	Ranking SR
Passive:						
I: 100% market	0.8874	0.7067	0.4357	0.0745		
II: 50% market	0.7310	0.6858	0.6181	0.5277		
III: 0% market	0.5294	0.5294	0.5294	0.5294		
Active:						
IV: Switching (0-1)	1.1409	1.0618	0.9432	0.7850	3	1
V: Optimal μ	3.3363	1.4739	1.0016	0.8127	2	2
VI: Optimal μ, σ	3.8859	1.6483	1.0888	0.8651	1	3
VII: Optimal μ (0-1)	1.0686	0.9232	0.8134	0.7197	5	5
VIII: Optimal μ, σ (0-1)	1.0977	0.9789	0.8555	0.7755	4	4

to switch from a static to one of the dynamic strategies.

Recall that the (ex ante) utility function of a mean-variance investor is given by

$$E_t\{r_{p,t+1}\} - \frac{1}{2}\gamma Var_t\{r_{p,t+1}\}, \tag{21}$$

which – for a given weight $w_{p,t+1}$ for the risky asset – becomes

$$r_{f,t+1} + w_{p,t+1}E_t\{r^e_{m,t+1}\} - \frac{1}{2}\gamma w^2_{p,t+1}Var_t\{r^e_{m,t+1}\}. \tag{22}$$

Ex post, we can estimate the average utility level as

$$\hat{U}_p = \frac{1}{T}\sum_{t=0}^{T-1}\left[r_{p,t+1} - \frac{1}{2}\gamma w^2_{p,t+1}\tilde{\sigma}^2_{m,t+1}\right], \tag{23}$$

where $\tilde{\sigma}^2_{m,t+1}$ denotes the *ex post* actual variance of the risky return, as presented in Section 4. Note that for the computation of the ex post mean-variance utility given in (23) it is essential to use an accurate measure of actual volatility. The use of squared monthly returns rather than an estimate based on daily data, would produce an unreliable estimation of the economic value of a dynamic trading rule (see Andersen and Bollerslev, 1998). Also note that actual volatility is irrelevant whenever the weight of the risky asset is zero.

The above approach enables us to compare alternative investment strategies by calculating the associated average utility levels. A given utility level can be interpreted as the certain return, in percentage per month, that provides the same utility to the investor as the risky investment strategy. This way, we can determine the economic value of market timing by calculating the maximum fee, in % per month, an investor should be willing to pay for holding the dynamic portfolio rather than a static one. It also allows comparison of alternative dynamic strategies. This maximum fee for holding portfolio a rather than portfolio b, Δ_{ab} say, can be found by solving

$$\frac{1}{T}\sum_{t=0}^{T-1}\left[(r_{a,t+1}-\Delta_{ab})-\frac{1}{2}\gamma w_{a,t+1}^{2}\tilde{\sigma}_{m,t+1}^{2}\right]=\frac{1}{T}\sum_{t=0}^{T-1}\left[r_{b,t+1}-\frac{1}{2}\gamma w_{b,t+1}^{2}\tilde{\sigma}_{m,t+1}^{2}\right], \quad (24)$$

where the indices a and b refer to the two different strategies. From this, it follows that the maximum fee (on a monthly basis), Δ_{ab}, can be straightforwardly computed by taking the difference between two alternative average utility levels.

For different values of γ, Table 12 presents the average realized utility values in percentages per month.[16] For example, an investor with $\gamma = 6$ who currently holds 50% of his wealth in the market portfolio (Strategy II), has an ex post average utility of 0.69%. The value of market timing for the different dynamic strategies, relatively to this passive portfolio is 0.38% per month for the switching strategy (Strategy IV), 0.79% for the optimal portfolio without future volatility (Strategy V), 0.96% for the optimal portfolio with future volatility (Strategy VI), and 0.24% and 0.29% respectively for the strategies with constraints without and with future volatility (Strategy VII and VIII). These numbers indicates sizeable gains due to market and volatility timing. While these gains are lower for more risk-averse investors, any of the dynamic strategies clearly outperforms the static ones, even if the risk-aversion coefficient is as large as 20. Table 12 also contains the ranking of the alternative strategies according to this measure and according to the Sharpe ratio. Obviously, the disadvantage of the Sharpe ratio, which overestimates the risk involved in the dynamic strategies, is overcome with our utility-based measure.

The results for the utility-based performance measure with transaction costs are presented in Table 13. Again it is assumed that static strategies do not entail transaction costs. Especially the dynamic strategies with extreme weights (strategies IV and V) suffer substantially from transaction costs. While for low and medium

[16] In each column, the "optimal" strategies are estimated to be optimal for a mean-variance investor with the corresponding coefficient of risk aversion γ.

Table 13: **Average Realized Utilities in Presence of Transaction Costs**

The average realized utilities, obtained using formula (23), are presented in percentages. The transaction costs are either low: 0.1%; medium: 0.5%; or high: 1% of the value traded. The risk aversion coefficient is equal to 6. Break-even transaction costs equalize average utility levels with those of Strategy II.

1966:1-1998:9				Break-even
Transaction costs:	**0.1%**	**0.5%**	**1%**	transaction costs
Passive:				
II: 50% market	0.6858	0.6858	0.6858	
Active:				
IV: Switching (0-1)	1.0465	0.9855	0.9091	2.4%
V: Optimal μ	1.4265	1.2374	1.0006	1.7%
VI: Optimal μ, σ	1.5413	1.1134	0.5785	0.9%
VII: Optimal μ (0-1)	0.9092	0.8532	0.7831	1.6%
VIII: Optimal μ, σ (0-1)	0.9634	0.9016	0.8243	1.8%

transaction costs, the optimal strategies (without constraints) still outperform the simple switching strategy, this is no longer uniformly the case if transaction costs are as high as 1%. When we consider the three dynamic strategies that do not allow short-sales and leverage, the impact of transaction costs is relatively limited and does not affect the relative ranking of these strategies. While the switching strategy beats its two competitors, each of the three restricted dynamic strategies beats a passive strategy of holding a fixed proportion in the market portfolio. This is also reflected in the break-even transaction costs, defined as the transaction costs, in percentages, that yield an average utility equal to Strategy II (holding 50% in market portfolio). In general, transaction costs have to be rather high in order to make the dynamic strategies uninteresting for a mean-variance investor. For the switching strategy the break-even transaction costs are as large as 2.4%. This means that even with transactions costs somewhat below 2.4%, a moderately risk-averse mean-variance investor[17] would be better off, in terms of average utility, to hold the switching portfolio rather than the static 50/50 allocation.

A final issue that calls attention is that of estimation error. The optimal portfolio weights employed above are based on estimates for the expected return and its conditional variance and are therefore not genuinely optimal. While this does not invalidate the analyses so far, it may indicate that the employed weights are

[17] With a risk aversion coefficient of 6.

too volatile compared to the true but unknown optimal weights, which may explain the relative poor performance, according to some measures, of the strategies referred to as optimal, compared to the switching strategy. It is well known that in general sample efficient portfolios are very sensitive to estimation errors in the expected returns and (co)variances (see Alexander and Resnick, 1985, Best and Grauer, 1991, and Britten-Jones, 1999). Several approaches are proposed to mitigate the effect of estimation error. For example, ter Horst, de Roon and Werker (2000) and Maenhout (2000) propose the use of a "pseudo risk aversion coefficient", that exceeds the investors genuine risk aversion coefficient γ, to adjust portfolio weights for estimation risk. Because in our case the presence of only one risky asset implies that, in each period, the weights of each strategy are – by construction – conditionally mean-variance efficient, such solutions do not translate directly to our problem. Nevertheless, it is an interesting question whether the estimated optimal portfolios can be adjusted to account for estimation error so as to improve performance of the resulting strategy.

We consider two alternative approaches. First, similar to ter Horst, de Roon and Werker (2000), we increased the risk aversion coefficient when computing the optimal weights for the mean-variance investor. In particular, we considered the economic value of a strategy based on a risk aversion coefficient of 12, evaluated for an investor that has an actual risk aversion coefficient of 6. In the absence of transaction costs, this reduces the economic average realized utility for this investor for each of the four portfolios involved. The last four numbers in the second column of Table 12 reduce to 1.094, 1.509, 0.849 and 0.927%, respectively. While this does not affect the relative ranking of the strategies, the use of a higher risk aversion coefficient to compute the optimal portfolio weights does not seem to lead to strategies that are economically more valuable.

A second approach takes into account the estimation error in the expected return by increasing the predicted volatility with a component that reflects the estimation uncertainty in the return forecast. This additional component is equal to the variance of the (expected) return forecast and depends upon the covariance matrix of the recursive OLS estimators for δ in (7). We compute this variance using the White (1980) covariance matrix estimator. On average, this increases the predicted volatilities by 7% and thus reduces the gap between average predicted volatilities and average actual volatilities in our sample. The effect of this correction is that the weights of the risky asset become somewhat smaller. The correction is relatively high when the forecast variance is high, which is in periods with extreme (combinations of) values for the explanatory variables in the forecasting model. The effects of this correction are very minor. The average utilities of strategies V to VIII, adjusted in the above way, for a mean-variance investor with $\gamma = 6$, are equal to 1.429, 1.665, 0.924 and 0.975%, respectively, and thus hardly different from those reported in Table 12. Similar results hold for the respective Sharpe ratios. Apparently, these simple corrections for estimation risk do not result in economically more attractive investment strategies.

6 Conclusions

While it is by now well-known that stock market returns and volatility exhibit predictable patterns, the explanatory power of the associated models is typically limited. Moreover, the within-sample goodness-of-fit measures are likely to overestimate the predictive performance of such models, because of overfitting, data snooping and small sample biases. This makes it an interesting question to evaluate the economic value of this predictability. In this paper we analyzed the joint economic significance of exploiting predictability in both returns and volatility. First, we estimated two simple linear models for monthly returns on the S&P 500 index for the squared unexpected returns using recursive least squares. The predictions from these models were used to conduct a number of nonparametric tests for out-of-sample market timing and volatility timing. The results of these tests unambiguously indicate the presence of statistically significant out-of-sample predictability over the period 1960-1998 and its two subperiods. In order to examine the interaction of the two sources of predictability, a joint test for the presence of market timing in both moments is proposed. This test soundly rejects the null hypothesis that both returns and volatility are not predictable. Moreover, there is no systematic relationship between the quality of the return and volatility forecasts. This indicates, for example, that it is not the case that a good return forecast typically corresponds to a bad volatility forecast or vice versa. Finally, it appears that the predictability of returns is larger in times when volatility is high.

To evaluate the economic value of this out-of-sample predictability, a number of alternative investment strategies were constructed. Besides a simple switching strategy, we also considered the estimated (ex ante) optimal mean-variance efficient portfolio based upon predicted returns and/or volatility, with and without imposing short-sales restrictions. Summarizing our results, we find that irrespective of the measure employed all dynamic trading rules outperform static ones. The Sharpe ratio, alpha, the Treynor-Mazuy test and the ex post average utility level all indicated sizeable gains from market timing over the period 1960-1998 and its two subperiods. The ranking of the alternative active strategies is not uniform across the different measures. When the Sharpe ratio is used to evaluate the strategies' performances, a switching strategy that invests 100% in the market portfolio if the predicted excess return is positive and 0% otherwise, unambiguously outperforms the four alternative dynamic strategies, even if transaction costs are zero or low. On the contrary, Jensen's alpha uniformly favors the strategy that exploits both return and volatility forecasts and has no restrictions on its weights. Note however that the alpha estimates are relatively inaccurate, given that this strategy is highly risky. Also note that successful market timing may imply that alpha estimates are biased downwards (Grinblatt and Titman, 1989).

To overcome the problems with the Sharpe ratio and Jensen's alpha, we propose a utility based measure of portfolio performance. This allows us to compute average utility, in terms of an equivalent certain monthly return, of each of the strategies. This measure also indicates that the economic value of the strategy that exploits predictability in both moments is largest, unless transaction costs are moderate or high. For a moderately risk-averse investor with a mean-variance utility function, the economic value of market timing in returns and volatility corresponds to more than 11.5% per year. This value is higher than for any other of the strategies. However, if short-sales and leverage are not allowed, none of the strategies based on the estimated optimal mean-variance weights manages to outperform the simple switching strategy. Unless the transaction costs are quite large (2.4%), the latter strategy dominates each passive portfolios. Finally, note that our findings are conservative, minimally subject to data snooping, and most results stand up to reasonable transaction costs and estimation error.

7 Data Appendix

In this appendix the description and source of the variables are given. The data we use are an updated and expanded version of those in Pesaran and Timmermann (1995). The frequency of the data is monthly and covers the period 1954:1 to 1992:12. Using different sources (as indicated below) we expanded the data set to capture a more recent time period: 1954:1 to 1998:9. The price and dividend data are based on the end of the month's Standard & Poor's 500 index at close. For more details see Pesaran and Timmermann (1995).

Table 14: **Description and Sources of Data**

Variable	Description and Source
P_t	Value of the price index of the Standard & Poor's 500 at the end of the month. Source: *Standard & Poor's Statistical Service**
D_t	12-month moving average of the dividends per share for the Standard & Poor's 500 index. Source: *Standard & Poor's Statistical Service**
RET_t	Stock index return, calculated as $(P_t + D_t - P_{t-1})/P_{t-1}$.
DIV_t	Dividend yield on the Standard & Poor's 500 index, calculated as D_t/P_t.
PE_t	Price-earnings ratio. Source: *Standard & Poor's Statistical Service**
$INFL_t$	Year-on-year rate of inflation computed using producer price index for finished goods. Source: *Citebase**
IP_t	Year-on-year rate of change in industrial production. Source: *Citebase, partly provided by Federal Reserve Statistical Release*
$I3_t$	Return on the 3-month Treasury Bill, converted to a monthly rate. Source: *Federal Reserve Bank of St. Louis*
$I12_t$	Return on the 12-month Treasury Bond, converted to a monthly rate. Source: *CRSP tapes, the Fama-Bliss discount bonds file*
MB_t	Year-on-year growth rate in the narrow money stock. Source: *Federal Reserve Bank of St. Louis*
COM_t	Return on 3-month Commercial Paper, converted to a monthly rate. Source: *Federal Reserve Bank of St. Louis*
CP_t	Commercial Paper-Treasury yield spread, calculated as $COM_t - I3_t$.
$EXRET_t$	Stock index return minus the three month T-bill rate converted to a monthly rate, calculated as $RET_t - I3_t$.

* Partly provided by DataStream.

References

[1] Alexander, G.J. and B.G. Resnick (1985), "More on Estimation Risk and Simple Rules for Optimal Portfolio Selection", *Journal of Finance*, 40, 125-133.

[2] Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts", *International Economic Review*, 39, 885-905.

[3] Angas, L.L.B (1936), "*Investment for Appreciation. Forecasting Movements in Security Prices. Technique of Trading in Shares for Profit*", McMillan, London.

[4] Akgiray, V. (1989), "Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts", *Journal of Business*, 62, 55-80.

[5] Barten, A. (1969), "Maximum Likelihood Estimation of a Complete System of Demand Equations", *European Economic Review*, 1, 7-73.

[6] Best, M.J. and R.R. Grauer (1991), "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results", *Review of Financial Studies*, 4, 315-342.

[7] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

[8] Bollerslev, T., R.F. Engle and D. Nelson (1994), "ARCH models" in *Handbook of Econometrics*, vol. IV, edited by R. Engle and D. McFadden, Elsevier, Amsterdam.

[9] Bollerslev, T. and R.J. Hodrick (1995), "Financial Market Efficiency Tests", *Handbook of Applied Econometrics*, vol I, edited by M.H. Pesaran and M.R. Wickens, Oxford and Malden, Mass., Blackwell.

[10] Bossaerts, P. and P. Hillion (1999), "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?", *Review of Financial Studies*, 12, 405-428.

[11] Breen, W., L.R. Glosten and R. Jagannathan (1989), "Predictable Variations on Stock Index Returns", *Journal of Finance*, 44, 1177-1189.

[12] Britten-Jones, M. (1999), "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights", *Journal of Finance*, 54, 655-671.

[13] Brock, W.A. and C.H. Hommes (1997), "A Rational Route to Randomness", *Econometrica*, 65, 1059-1095.

[14] Campbell, J.Y. (1987), "Stock Returns and the Term Structure", *Journal of Financial Economics*, 18, 373-399.

[15] Campbell, J.Y. (2000), "Asset Pricing at the Millennium", *Journal of Finance*, forthcoming.

[16] Chen, Z. and P.J. Knez (1996), "Portfolio Measurement: Theory and Applications", *Review of Financial Studies*, 9, 511-555.

[17] Cumbey R.E. and D.M. Modest (1987), "Testing for Market Timing Ability, A Framework for Forecast Evaluation", *Journal of Financial Economics,* 19, 169-189.

[18] Dowrie, G.W. and D.R. Fuller (1950), "*Investments*", second edition, John Wiley, New York.

[19] Duffee, G.R. (1996), "Idiosyncratic Variation of Treasury Bill Yields", *Journal of Finance*, 51, 527-551.

[20] Dybvig, P.H. and S.A. Ross (1985), "Differential Information and Performance Measurement Using a Security Market Line", *Journal of Finance*, 40, 383-399.

[21] Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the U.K. inflation", *Econometrica*, 50, 987-1008.

[22] Fama, E. and K. French (1989), "Business Conditions and Expected Returns on Stocks and Bonds", *Journal of Financial Economics*, 25, 23-49.

[23] Fama, E. and G.W. Schwert (1977), "Asset Returns and Inflation", *Journal of Financial Economics*, 5, 115-146.

[24] Ferson W.E. and R.W. Schadt (1996), "Measuring Fund Strategy and Performance in Changing Economic Conditions", *Journal of Finance*, 51, 425-462.

[25] Fleming, J., C. Kirby and B. Ostdiek (1999), "The Economic Value of Volatility Timing", working paper, Rice University, Houston (forthcoming *Journal of Finance).*

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

[27] Fung, W. and D.A. Hsieh (1997), "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds", *Review of Financial Studies*, 10, 275-302.

[28] Glosten, L.R. and R. Jagannathan (1994), "A Contingent Claim Approach to Performance Evaluation", *Journal of Empirical Finance*, 1, 133-160.

[29] Graham, J.R. and C.R. Harvey (1996), "Market Timing Ability and Volatility Implied by Investment Newsletters' Asset Allocation Recommendations", *Journal of Financial Economics*, 42, 397-421.

[30] Greene, W.H. (2000), *Econometric Analysis*, 4th edition, Prentice Hall.

[31] Grinblatt, M. and S. Titman (1989), "Portfolio Performance Evaluation: Old Issues and New Insights", *Review of Financial Studies*, 2, 3, 393-421.

[32] Harvey, C.R. (1991), "The World Price of Covariance Risk", *Journal of Finance*, 46, 1, 111-157.

[33] 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.

[34] Hodrick, R. (1992), "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement", *Review of Financial Studies*, 5, 357-386.

[35] ter Horst, J.R., F.A. de Roon and B.J.M. Werker, (2000), "Incorporating Estimation Risk in Optimal Portfolios", Discussion Paper, CentER, Tilburg University.

[36] Jacobsen, B. (1999), "The Economic Significance of Simple Time Series Models of Stock Return Predictability", Working Paper, University of Amsterdam.

[37] Jensen, M.C. (1968), "Problems in Selection of Security Portfolios: the Performance of Mutual Funds in the Period 1945-64", *Journal of Finance*, 23, 389-416.

[38] Kandel S. and R.F. Stambaugh (1996), "On the Predictability of Stock Returns: An Asset-Allocation Perspective", *Journal of Finance*, 51, 385-424.

[39] Keim, D. and R. Stambaugh (1986), "Predicting Returns in the Stock and Bond Markets", *Journal of Financial Economics*, 17, 357-390.

[40] Kirby (1998), "The Restrictions on Predictability Implied by Rational Asset Pricing Models", *Review of Financial Studies*, 11, 343-382.

[41] Kothari S.P. and J. Shanken (1995), "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis", *Journal of Financial Economics*, 44, 169-203.

[42] Lewellen, J. and J. Shanken (2000), "Estimation Risk, Market Efficiency, and the Predictability of Returns,", NBER working paper 7699, May 2000.

[43] Lo, A.W. and A.C. MacKinlay (1990), "Data-Snooping Biases in Tests of Financial Asset Pricing Models", *Review of Financial Studies,* 3, 431-467.

[44] Lucas, R.E. (1978), "Asset Prices in an Exchange Economy", *Econometrica*, 46, 1429-1445.

[45] Maenhout, P. (2000), "Robust Portfolio Rules and Asset Pricing", Working Paper, Harvard University.

[46] Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices", *Journal of Business*, 36, 394-419.

[47] Newey, W.K. and K. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica", *Econometrica*, 55, 703-708.

[48] Pagan, A.R. and G.W. Schwert (1990), "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics*, 45, 267-290.

[49] Pesaran, M.H. and A.G. Timmermann (1995), "Predictability of Stock Returns: Robustness and Economic Significance", *Journal of Finance*, 50, 1201-1228.

[50] Pesaran, M.H. and A.G. Timmermann (2000), "A Recursive Modelling Approach to Predicting UK Stock Returns", *Economic Journal*, 110, 159-191.

[51] Pontiff, J. and L.D. Schall (1998), "Book-to-Market Ratios as Predictors of Market Returns", *Journal of Financial Economics*; 49, 141-160.

[52] Prime, J.H. (1946), "*Investment Analysis*", Prentice Hall, New York.

[53] Schwert (1989), G.W., "Why Does Stock Market Volatility Change over Time?", *Journal of Finance*, 44, 1115-1153.

[54] Sentana, E. and S. Wadhwani (1991), "Semi-parametric Estimation and the Predictability of Stock Market Returns: some Lessons from Japan", *Review of Economic Studies*, 58, 547-563.

[55] Sharpe W.F. (1966), "Mutual Fund Performance", *Journal of Business*, 39, 119-138.

[56] Shiller, R.J. (2000), "*Irrational Exuberance*", Princeton University Press, Princeton, New Jersey.

[57] Söderlind, P. (1999), "An Interpretation of SDF Based Performance Measures"*, European Finance Review,* 3, 233-237.

[58] Solnik, B. (1993), "The Performance of International Asset Allocation Strategies using Conditioning Information", *Journal of Empirical Finance*, 1, 33-55.

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

[60] Sullivan, R., A. Timmermann and H. White (1999), "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap", *Journal of Finance*, 54, 1647-1691.

[61] Treynor J.L. and K. Mazuy (1966), "Can Mutual Funds Outguess the Market?", *Harvard Business Review,* 44, 131-136.

[62] White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity" *Econometrica*, 48, 817-835.