

A Trading Approach to Testing for Predictability

Stanislav Anatolyev* Alexander Gerko
New Economic School Deutsche Bank

Abstract

We propose a market timing test for conditional mean independence of financial returns. The new excess predictability (EP) test statistic has an interpretation of a properly normalized return of a certain trading strategy. We discuss similarities of the EP test to the popular directional accuracy (DA) test of Pesaran and Timmermann (1992). Power properties of the EP test are advantageous, and size properties are comparable to those of the DA test. We illustrate application of the test using weekly data on the S&P500 index.

Key words: Trading strategy; Mean predictability; Directional accuracy test; S&P500 index.

*Corresponding author. Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru. We are grateful to the editors Eric Ghysels and Torben Andersen, an associate editor, and two referees whose comments helped improve the presentation. We also thank for useful suggestions all other members of the NES research project “Dynamics and Predictability in Russian Financial Markets”, and the participants of the Time-Series Workshop in Linz in October 2003.

1 Introduction

It is a common perception nowadays that predictability of financial returns should be judged by how successful out-of-sample exercises are rather than by in-sample fit and coefficient significance. Out-of-sample exercises should in turn be tied to profitability of virtual investors' decisions rather than to simple statistical measures like the mean squared prediction error (MSPE). Such measures aim at minimizing an unrelated to profitability loss function rather than at getting a significant outcome from the viewpoint of profit maximization, the ultimate goal of making predictions of returns in financial markets (e.g., see Leitch and Tanner (1991), Brock, Lakonishok and LeBaron (1992), Pesaran and Timmermann (1995), Satchell and Timmermann (1995)). The proportion of correctly predicted signs of returns at the heart of the popular directional accuracy (DA) test of Pesaran and Timmermann (1992) also belongs to the class of such statistical measures.

In this paper we show that the trading approach can be used to construct a formal test for mean predictability. We develop a market timing test that is explicitly tied to a virtual investor's simple trading strategy. This trading strategy issues a buy signal if a forecast of next period return is positive and a sell signal otherwise. Our *excess profitability (EP)* test is based on a suitably normalized profitability of the position implied by the trading strategy above a certain benchmark. The EP statistic is asymptotically distributed as a standard normal.

A close inspection of how the EP and DA statistics are constructed reveals certain similarities. In particular, both are Hausman type tests, and both are derived under stronger assumptions placed on the relationship between predictors and predictands than the tested features presume. However, in addition to economic interpretability of the EP statistic, the EP test formally tests for mean predictability rather than sign predictability (see Christoffersen and Diebold (2003) for subtle differences between the two concepts). We show analytically that the EP test exhibits strictly higher power against some local alternatives with either linear or nonlinear predictability. The power dominance is also confirmed via simulations in realistic setups. We also discuss the issue of possible size distortions, and conduct a simulation experiment to assess them numerically. Even though the size distortions may come from a number of sources, they are likely to be small in practice.

We illustrate the application of the test using half a century of weekly data on the

S&P500 index. By computing the EP statistic from data in a moving time window, we observe a remarkable pattern of dynamics of the level of mean predictability throughout the last half of the 20th century. In particular, we provide some evidence that it has substantially increased during the last decade.

The paper is organized as follows. In Section 2 we introduce the trading strategy, build the EP test and compare its form to that of the DA test. In Section 3 we study power and size properties of the two tests. Section 4 describes the empirical illustration, and Section 5 concludes. Three appendices contain technical derivations.

2 Trading strategy and test for mean predictability

Let the variable y_t represent the return on some financial asset or index, and \hat{y}_t be a continuously distributed forecast of y_t . The forecast \hat{y}_t is allowed to depend only on the data from $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$, or, more generally, from the extended information set $\mathcal{I}_{t-1} \supset \{y_{t-1}, y_{t-2}, \dots\}$ which may include other historical variables. Consider the following trading rule based on \hat{y}_t :

$$\begin{cases} \text{buy shares worth current wealth,} & \text{if } \hat{y}_t \geq 0, \\ \text{sell shares worth current wealth,} & \text{otherwise.} \end{cases} \quad (1)$$

That is, an investor goes long if the prediction of the next period return is positive, and goes short otherwise. For brevity, we will call this rule *the trading strategy*. Equipped with the trading strategy, the investor modifies her position each trading period closing it at the end of the period. Then the one period return of the trading strategy is

$$r_t = \text{sign}(\hat{y}_t)y_t, \quad (2)$$

where $\text{sign}(\cdot)$ takes value -1 when its argument is negative, and value $+1$ when its argument is non-negative. We implicitly assume that the distribution of $\text{sign}(\hat{y}_t)$ is non-degenerate.

The trading strategy (1) describes the behavior of a risk neutral “artificial technical analyst”, in the terminology of Skouras (2001). The profitability of the trading strategy (1) was evaluated by Gençay (1998) to measure whether forecasts have economic value in practice. Using two and a half decades of DJIA data Gençay (1998) finds that this trading strategy is able to provide perceptible profits relative to the “buy-and-hold” strategy. We

instead use this trading approach to construct a formal test of mean predictability of returns which is based on the out-of-sample profitability of the trading strategy. The reader should keep in mind that the trading process is only a thought experiment, and it makes no difference whether or not market limitations (like transactions costs and short selling constraints) allow that the trading strategy is exercised. However, market limitations weaken the links to the notion of market efficiency, as the presence of predictability is equivalent to inefficiency under the strong assumption of constant risk premia and no operational market imperfections.

Formally, the null hypothesis is that of conditional mean independence, $H_0 : E[y_t | \mathcal{I}_{t-1}] = \text{const}$. Technically, we require that under the null a stronger property holds, that \hat{y}_t be independent from y_t for all lags and leads (see the discussion of this technical issue in the next section). The expected one period return of the trading strategy (1) is $E[r_t]$, which is consistently estimable under the null by the following two estimators:

$$A_T = \frac{1}{T} \sum_t r_t \quad (3)$$

and

$$B_T = \left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) \right) \left(\frac{1}{T} \sum_t y_t \right). \quad (4)$$

Indeed, under the null, $A_T \xrightarrow{p} E[r_t]$ and

$$B_T \xrightarrow{p} E[\text{sign}(\hat{y}_t)] E[y_t] = E[\text{sign}(\hat{y}_t) E[y_t]] \stackrel{H_0}{=} E[\text{sign}(\hat{y}_t) E[y_t | \mathcal{I}_{t-1}]] = E[\text{sign}(\hat{y}_t) y_t] = E[r_t].$$

While A_T is the average return resulting from use of the trading strategy, B_T is (an estimate of) the average return of a benchmark strategy that issues buy/sell signals at random with probabilities corresponding to the proportion of “buys” and “sells” implied *ex post* by the trading strategy. When y_t is predictable, real-time forecasting and investing according to the trading strategy will generate a higher return than the benchmark, the difference between A_T and B_T will be sizable, and the test will have power.

To complete the construction of the test, it remains to compute the variance of $A_T - B_T$ under the null. Let $p_{\hat{y}} = \Pr\{\text{sign}(\hat{y}_t) = 1\}$, then (see appendix A)

$$\text{var}(A_T - B_T) = 4 \frac{T-1}{T^2} p_{\hat{y}} (1 - p_{\hat{y}}) \text{var}(y_t). \quad (5)$$

The most straightforward way to estimate this variance is

$$\hat{V}_{EP} = \frac{4}{T^2} \hat{p}_{\hat{y}} (1 - \hat{p}_{\hat{y}}) \sum_t (y_t - \bar{y})^2,$$

where we corrected for degrees of freedom when estimating the variance of y_t , and where

$$\hat{p}_{\hat{y}} = \frac{1}{2} \left(1 + \frac{1}{T} \sum_t \text{sign}(\hat{y}_t) \right)$$

is a consistent estimator of $p_{\hat{y}}$. The estimate \hat{V}_{EP} is positive by construction unless by chance all forecasts have the same sign. The Hausman-type *excess profitability (EP)* test statistic and its asymptotic distribution are

$$EP \equiv \frac{A_T - B_T}{\sqrt{\hat{V}_{EP}}} \xrightarrow{d} N(0, 1). \quad (6)$$

The EP statistic (6) is reminiscent of the directional accuracy (DA) statistic of Pesaran and Timmermann (1992) that is routinely used as a predictive-failure test in constructing forecasting models (e.g., see Pesaran and Timmermann (1995) and Qi (1999)). Let us have a look at the construction of the DA statistic that results after a change of variables. When the forecasts do not have predictive power, the (recentered and renormalized) success ratio (see appendix B)

$$\tilde{A}_T = \frac{1}{T} \sum_t \text{sign}(\hat{y}_t) \text{sign}(y_t),$$

does not differ much from the expected success ratio that would obtain in case y_t and \hat{y}_t were independent. A natural estimate of the latter is

$$\tilde{B}_T = \left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) \right) \left(\frac{1}{T} \sum_t \text{sign}(y_t) \right).$$

Let

$$\hat{p}_y = \frac{1}{2} \left(1 + \frac{1}{T} \sum_t \text{sign}(y_t) \right)$$

be a consistent estimate of $p_y = \Pr\{\text{sign}(y_t) = 1\}$. Under independence of \hat{y}_t and y_t at all lags and leads, the Hausman-type DA test statistic, an appropriately scaled difference between \tilde{A}_T and \tilde{B}_T , is asymptotically standard normal:

$$DA \equiv \frac{\tilde{A}_T - \tilde{B}_T}{\sqrt{\hat{V}_{DA}}} \xrightarrow{d} N(0, 1), \quad (7)$$

where (see appendix B)

$$\hat{V}_{DA} = 16 \frac{T-1}{T^2} \hat{p}_{\hat{y}} (1 - \hat{p}_{\hat{y}}) \hat{p}_y (1 - \hat{p}_y).$$

The DA test is not unrelated to the possibility of obtaining excess profits, as the ability to predict market's direction is certainly useful for investors. However, as Skouras (2000)

argues, the ability of an investor to predict the market's direction may not necessarily lead to extracting excess profits if at times when mistakes on direction are made returns are greater in absolute value than at times when no mistakes on direction are made. While the DA statistic will likely be significant under these circumstances, the EP statistic based on the profitability itself will not be significant.

Further note that a typical summand in \tilde{A}_T equals

$$\text{sign}(\hat{y}_t) \text{sign}(y_t),$$

while that in A_T equals

$$\text{sign}(\hat{y}_t) \text{sign}(y_t) |y_t|.$$

This implies that if the series of returns y_t is predictable, the EP test takes fuller advantage of this predictability, while the DA test ignores the fact that higher returns are associated with better forecasts. This is likely to lead to a higher power of the EP test, which will be confirmed in the next section. Also note that the DA test will successfully detect deviations from the null of no predictability when y_t exhibits sign predictability, while the profitability statistic (6) will when y_t exhibits conditional mean dependence. Christoffersen and Diebold (2003) show that the sign predictability may be merely a result of volatility dependence in the absence of mean predictability, and that sign dependence does not imply violation of market efficiency. Thus, our EP statistic is better suited for mean predictability considerations.

3 Power and size

Consider simple departures from the null, a linear autoregression of first order

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid \text{N}(0, 1), \quad (8)$$

with the optimal forecast $\hat{y}_t = \alpha y_{t-1}$, and a threshold model with a mean shift

$$y_t = \begin{cases} -\alpha, & y_{t-1} \leq 0 \\ \alpha, & y_{t-1} > 0 \end{cases} + \varepsilon_t, \quad \varepsilon_t \sim iid \text{N}(0, 1) \quad (9)$$

with the optimal forecast $\hat{y}_t = \alpha \text{sign}(y_{t-1})$. For simplicity, unconditional means are set to zero and conditional variances to unity, and we abstract from parameter uncertainty (see the discussion of test sizes below). Under the null of no predictability, $\alpha = 0$.

We investigate the power of the EP and DA tests for these models under the sequence of local alternatives $H_A^\delta : \alpha = \delta/\sqrt{T}$, $\delta > 0$. In appendix C it is shown that for the linear model (8), $EP \xrightarrow{d} N(\delta E[|y|], 1)$ and $DA \xrightarrow{d} N(2\phi(0)\delta E[|y|], 1)$, and for the nonlinear model (9), $EP \xrightarrow{d} N(\delta, 1)$ and $DA \xrightarrow{d} N(2\phi(0)\delta, 1)$. One can see that under both local linear and local nonlinear predictability, the power of the EP test is strictly larger than that of the DA test, as the noncentrality parameters differ by the factor $2\phi(0) = \sqrt{2/\pi} < 1$ in both cases.

Next, we investigate the power of both tests in a Monte-Carlo experiment with a more realistic setup. We use the following two data generating processes (DGPs) that exhibit different types of predictability and whose coefficients are calibrated using the S&P500 index during the period from 1954 to 1973 (when the index seems to be most predictable):

$$\text{“AR”} \quad y_t = 0.1256 \cdot y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid N(0, 0.000249); \quad (10)$$

$$\text{“SETAR”} \quad y_t = \begin{cases} 0.000844 + 0.2453 \cdot y_{t-1}, & |y_{t-1}| \leq 0.01848 \\ 0.002679 + 0.0664 \cdot y_{t-1}, & |y_{t-1}| > 0.01848 \end{cases} + \varepsilon_t, \quad (11)$$

$$\varepsilon_t \sim iid N(0, 0.000245).$$

The sample length is 1,000. In each of 10,000 simulations, estimation is performed over a rolling window of 100 observations. The EP and DA statistics are computed over 900 predictions. The top panel of Table 1 shows actual rejection frequencies corresponding to nominal ones of 10%, 5%, and 1% for two-sided alternatives. We use the least squares linear predictor (labeled OLS) and nearest neighbors local regression (labeled NN), which are detailed in section 4. The power of the EP test appreciably exceeds that of the DA test, sometimes by a factor greater than 1.5. One can also see how much the power of either test may vary with the choice of a forecasting model. When the DGP is truly linear, fitting a linear model has a much greater predictive power than a nonparametric method, which in turn (slightly) increases power when the series is nonlinearly predictable.

Another concern is actual sizes of the tests. It should be noted that the asymptotics for the EP and DA tests is derived under the strong presumption that under H_0 there is independence of y_t and \hat{y}_t at all lags and leads. (For the DA, see the manipulations below equation (4) in Pesaran and Timmermann (1992, p. 462).) This presumption is evidently stronger than the tested features. Indeed, even under no mean or sign predictability it is highly unlikely that the forecast is completely unrelated to the variable being forecast. For example, when a lagged value y_{t-1} is used to make a prediction of y_t , this prediction is by

construction not independent from y_t at lag one. The only case when the serial independence of predictors and predictands holds under the null is when predictor variables are strictly exogenous, and no estimated model is used to generate predictions. In reality, however, rarely either condition holds. Thus, in practical situations both tests may have wrong asymptotic sizes.

A typical parametric or nonparametric forecast \hat{y}_t that possibly uses estimated parameters can be viewed as a complicated function of observations dated by $t - 1$ and earlier provided that no future observations are employed in estimating unknown parameters. The complicatedness of such forecasts precludes analytical evaluation of possible size distortions. In cases when an estimated model is used to construct forecasts, one may separate size distortion arising from the use of predictors that are serially dependent on the predictand, and size distortions arising from the presence of parameter uncertainty (apart from distortions due to model uncertainty and, of course, due to the approximating nature of the asymptotic approach). The former distortion may be zero as in the examples used for local power computations. In a more complicated setup, however, its absence is not guaranteed. The type of distortion that is due to estimated parameters has been recently discussed in the econometric literature. West and McCracken (2002) attest that the parameter uncertainty is largely ignored in the applied literature. West (1996) provides formulae for corrections that can be used to attain the asymptotic standard normal distribution for statistics that are differentiable in parameters; McCracken (2000) repeats this for statistics that are not differentiable but whose expected values under the null are differentiable and allow a mean value expansion. Unfortunately, both the EP and DA statistics involve functions that are not differentiable in either sense (see McCracken 2000, discussion following assumption 4). Hence, in the absence of a tractable theory an applied researcher is advised to evaluate possible size distortions via simulations.

To assess size distortions that are likely to arise in practice, we perform a Monte–Carlo experiment using the following DGPs (again, calibrated using the S&P500 index):

$$\text{“Const”} \quad y_t = 0.001526 + \varepsilon_t, \quad \varepsilon_t \sim iid \text{ N}(0, 0.0000250); \quad (12)$$

$$\begin{aligned} \text{“GARCH”} \quad y_t &= 0.002483 + \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim iid \text{ N}(0, \sigma_t^2), & (13) \\ \sigma_t^2 &= 0.0000223 + 0.1773 \cdot \varepsilon_{t-1}^2 + 0.7397 \cdot \sigma_{t-1}^2. \end{aligned}$$

Note that both parametric and nonparametric methods estimate a correct model for the

conditional mean. Under the DGP “Const”, the impact of serial dependence is brought to a minimum; under the DGP “GARCH”, we deliberately invoke strong GARCH effects to exacerbate the dependence between returns and their forecasts. The bottom panel of Table 1 shows actual sizes corresponding to nominal ones of 10%, 5%, and 1%. For the DGP “Const”, both tests exhibit negligible size distortions. This indicates that the parameter uncertainty and asymptotic approximation do not have significant effects with the estimation design used. For the DGP “GARCH”, size distortions are larger but still are small. Note that the distortions are smaller when a nonparametric method is used rather than a parametric one, even though the former provides less precise estimates during the estimation stage. This can be explained by a sort of the “whitening by windowing” effect (Hart 1996) of local nonparametric estimation in time series: kernel estimates evaluated at different histories employ different observations and hence are less dependent than parametric estimates that employ the *entire* estimation sample. The power results, however, indicate that nonparametric methods capture linear predictability less successfully.

4 Empirical illustration

We provide an application of the test using the weekly S&P500 stock market index for the period from January 03, 1950 to May 05, 2003, totaling 2783 observations, which we obtained from `finance.yahoo.com`. Before the analysis we take first differences of logarithms of the index to obtain a series of returns. We construct one step ahead forecasts using rolling regressions with a window of 104 observations corresponding to two years. This scheme results in 2677 ($= 2783 - 1 - 1 - 104$) predictions.

We use the “naive” random walk (RW) forecasts as a benchmark (note that the trading strategy (1) and the RW forecasts (i.e. \hat{y}_t are zero for all t) generate the “buy-and-hold” strategy). Parametric forecasts are provided by a linear OLS model, and nonparametric forecasts are provided by the nearest neighbor (NN) local linear regression (Härdle 1990, Section 3.1.1). Both parametric and nonparametric regressions use the first lag of returns as a regressor. Forecasts on the basis of simple OLS regressions are quite intensively used in applied work (e.g., in Pesaran and Timmermann 1995). Among nonparametric methods, nearest neighbor methods seem to be most popular in empirical work (e.g., Diebold and Nason 1990, Meese and Rose 1991, Mizrach 1992), and the number of neighbors is usually

selected manually. We follow the tradition and set it to $n = 10$ (the number of neighbors selected by cross-validation tends to be too large). The normal kernel is employed.

Test results are reported in Table 2. As follows from its column 2, no model succeeded in improving upon RW forecasts in terms of MSPE, which is in line with the empirical literature. However, the results of EP and DA tests in columns 3 and 4 show strong predictability even in the case of the linear model. It is clear that the conditional mean independence hypothesis is strongly rejected for the weekly S&P500 index. Illuminating patterns unfold on Figure 1 which shows graphs of log cumulative returns on positions based on our trading strategy and different forecasting models. That the NN local regression dominates the linear OLS model makes one think that nonlinear features prevail in the predictable component of the S&P500 index.

Recent literature points out that the degree of predictability may be different in different periods (e.g., Timmermann and Granger 2004). It is interesting to track its evolution through time using the EP statistic computed from data in a moving time window of fixed length. We set the length of the moving time window to 520 which corresponds to approximately 10 years. As the NN local regression showed most promise, we report the results with this predictor. The track of the EP statistic along with the upper 5% critical value line is depicted on Figure 2 where values of the EP statistic over periods of 10 years should be interpreted as measures of mean predictability over whole periods. One can clearly observe two points when the structural breaks in the market efficiency may have occurred. The first corresponds to the global maximum of the EP statistic over the whole sample, which occurs for the period from 1963 to 1973. Starting from the end of this period and up to the year 1992 one can see an apparent trend towards conditional mean independence, and during the period from 1982 to 1992 the EP statistic is close to zero. However, after 1992 this trend is reversed, and for the period from 1993 to 2003 the hypothesis of conditional mean independence can be only marginally accepted. If this 11-year trend continues, we may see the evidence of strong mean predictability of the S&P500 index in the near future.

5 Conclusions

We have proposed a new test for mean predictability of returns based on a properly normalized excess return of a simple trading strategy over the return of a certain benchmark.

The new statistic is similar to that of the popular directional accuracy test of Pesaran and Timmermann (1992), but exhibits higher power against alternatives with linear or nonlinear predictability. We have illustrated application of the test using the weekly S&P500 index for which the hypothesis of conditional mean independence is strongly rejected, and discovered interesting patterns in the degree of predictability throughout the last half of the 20th century.

A Appendix: Computation of variance of $A_T - B_T$

Observe that $E[\text{sign}(\hat{y}_t)] = 2p_{\hat{y}} - 1$ and $\text{var}[\text{sign}(\hat{y}_t)] = 4p_{\hat{y}}(1 - p_{\hat{y}})$. Recall that it is assumed that under H_0 the series \hat{y}_t and y_t are independent for all lags and leads. We have

$$\begin{aligned}
\text{var}(A_T) &= \frac{1}{T} E[\text{sign}(\hat{y}_t)^2 y_t^2] - \frac{1}{T} E[\text{sign}(\hat{y}_t) y_t]^2 \stackrel{H_0}{=} \frac{1}{T} E[y_t^2] - \frac{1}{T} (2p_{\hat{y}} - 1)^2 E[y_t]^2, \\
\text{var}(B_T) &= E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right)^2 \left(\frac{1}{T} \sum_t y_t\right)^2\right] - E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right) \left(\frac{1}{T} \sum_t y_t\right)\right]^2 \\
&\stackrel{H_0}{=} E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right)^2\right] E\left[\left(\frac{1}{T} \sum_t y_t\right)^2\right] - E\left[\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right]^2 E\left[\frac{1}{T} \sum_t y_t\right]^2 \\
&= \frac{1}{T} (2p_{\hat{y}} - 1)^2 \text{var}(y_t) + \frac{4}{T} p_{\hat{y}} (1 - p_{\hat{y}}) \left(\frac{1}{T} E[y_t^2] + \left(1 - \frac{1}{T}\right) E[y_t]^2 \right). \\
\text{cov}(A_T, B_T) &= E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) y_t\right) \left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right) \left(\frac{1}{T} \sum_t y_t\right)\right] \\
&\quad - E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) y_t\right)\right] E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right) \left(\frac{1}{T} \sum_t y_t\right)\right] \\
&\stackrel{H_0}{=} \frac{1}{T^2} E[\text{sign}(\hat{y}_t)^2 y_t^2] + \frac{T-1}{T^2} E[\text{sign}(\hat{y}_t)^2 y_t] E[y_t] \\
&\quad + \frac{T-1}{T^2} E[\text{sign}(\hat{y}_t) y_t^2] E[\text{sign}(\hat{y}_t)] + \frac{(T-1)^2}{T^2} E[\text{sign}(\hat{y}_t)]^2 E[y_t]^2 \\
&\quad - E[\text{sign}(\hat{y}_t)]^2 E[y_t]^2 \\
&= \frac{1}{T} (2p_{\hat{y}} - 1)^2 \text{var}(y_t) + \frac{4}{T} p_{\hat{y}} (1 - p_{\hat{y}}) \left(\frac{1}{T} E[y_t^2] + \left(1 - \frac{1}{T}\right) E[y_t]^2 \right) \\
&= \text{var}(B_T).
\end{aligned}$$

Taking things together, we obtain (5). That the test is of the Hausman type can be seen from the following argument: the estimator B_T is semiparametrically efficient for $E[\text{sign}(\hat{y}_t)] E[y_t]$ as it is a product of independent semiparametrically efficient estimators of $E[\text{sign}(\hat{y}_t)]$ and $E[y_t]$. However, this argument is based on *asymptotic* considerations, while we showed above explicitly that $\text{cov}(A_T, B_T)$ equals $\text{var}(B_T)$ in *finite samples*.

B Appendix: Another representation of DA statistic

Let

$$\bar{A}_T = \frac{1}{T} \sum_t \mathbb{I}[y_t \hat{y}_t > 0],$$

$$\bar{B}_T = \left(\frac{1}{T} \sum_t \mathbb{I}[y_t > 0] \right) \left(\frac{1}{T} \sum_t \mathbb{I}[\hat{y}_t > 0] \right) + \left(1 - \frac{1}{T} \sum_t \mathbb{I}[y_t > 0] \right) \left(1 - \frac{1}{T} \sum_t \mathbb{I}[\hat{y}_t > 0] \right).$$

Using the change of variables $\mathbb{I}[x > 0] = \frac{1}{2}(1 + \text{sign}(x))$ when $\Pr[x = 0] = 0$, and the property $\text{sign}(x_1 x_2) = \text{sign}(x_1) \text{sign}(x_2)$, one can see that the numerator of the DA test is

$$\bar{A}_T - \bar{B}_T = \frac{1}{2} \left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) \text{sign}(y_t) - \left(\frac{1}{T} \sum_t \text{sign}(y_t) \right) \left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) \right) \right) = \frac{\tilde{A}_T - \tilde{B}_T}{2}.$$

The denominator of the DA test is a square root of $\widehat{\text{var}}(\bar{A}_T) - \widehat{\text{var}}(\bar{B}_T)$. But (Pesaran and Timmermann (1992), equations below (6))

$$T\widehat{\text{var}}(\bar{A}_T) = (\hat{p}_y \hat{p}_y + (1 - \hat{p}_y)(1 - \hat{p}_y)) (1 - \hat{p}_y \hat{p}_y - (1 - \hat{p}_y)(1 - \hat{p}_y)),$$

$$T\widehat{\text{var}}(\bar{B}_T) = (2\hat{p}_y - 1)^2 \hat{p}_y (1 - \hat{p}_y) + (2\hat{p}_y - 1)^2 \hat{p}_y (1 - \hat{p}_y) + 4T^{-1} \hat{p}_y \hat{p}_y (1 - \hat{p}_y)(1 - \hat{p}_y).$$

Expanding and subtracting, we get

$$\widehat{\text{var}}(\bar{A}_T) - \widehat{\text{var}}(\bar{B}_T) = 4 \frac{T-1}{T^2} \hat{p}_y (1 - \hat{p}_y) \hat{p}_y (1 - \hat{p}_y).$$

C Appendix: Local power computations

It is easy to see that under H_A^δ , for both models (8) and (9), $T\hat{V}_{EP} \xrightarrow{p} 1$, $T\hat{V}_{DA} \xrightarrow{p} 1$, $\sqrt{T}B_T \xrightarrow{p} 0$ and $\sqrt{T}\tilde{B}_T \xrightarrow{p} 0$. For the linear model (8),

$$A_T = \alpha \frac{1}{T} \sum_t \text{sign}(y_{t-1}) y_{t-1} + \frac{1}{T} \sum_t \text{sign}(y_{t-1}) \varepsilon_t,$$

$$\tilde{A}_T = \frac{1}{T} \sum_t \text{sign}(y_{t-1}) \text{sign}(\alpha y_{t-1} + \varepsilon_t).$$

Under H_A^δ , $\sqrt{T}A_T$ and hence EP converge in distribution to $N(\delta E[|y|], 1)$. For \tilde{A}_T , we have

$$\begin{aligned} E[\tilde{A}_T] &= \Pr[y_{t-1}(\alpha y_{t-1} + \varepsilon_t) \geq 0] - \Pr[y_{t-1}(\alpha y_{t-1} + \varepsilon_t) < 0] = \\ &= \int_{-\infty}^{+\infty} (\Pr[y_{t-1}(\alpha y_{t-1} + \varepsilon_t) \geq 0 | y_{t-1}] - \Pr[y_{t-1}(\alpha y_{t-1} + \varepsilon_t) < 0 | y_{t-1}]) f(y_{t-1}) dy_{t-1} \\ &= \int_0^{+\infty} (\Phi(\alpha y) - \Phi(-\alpha y)) \phi(y) dy + \int_{-\infty}^0 (\Phi(-\alpha y) - \Phi(\alpha y)) \phi(y) dy \\ &= \int_0^{+\infty} 2\phi(0)\alpha y \phi(y) dy - \int_{-\infty}^0 2\phi(0)\alpha y \phi(y) dy + o(\alpha) = 2\phi(0)\alpha E[|y|] + o(\alpha) \end{aligned}$$

and $\text{var}(\tilde{A}_T) = T^{-1} + o(T^{-1})$, so that under H_A^δ , $\sqrt{T}\tilde{A}_T$ and hence DA converge in distribution to $N(2\phi(0)\delta E[|y|], 1)$. For the nonlinear model (9),

$$\begin{aligned} A_T &= \alpha + \frac{1}{T} \sum_t \text{sign}(y_{t-1})\varepsilon_t, \\ \tilde{A}_T &= \frac{1}{T} \sum_t \text{sign}(\alpha + \text{sign}(y_{t-1})\varepsilon_t). \end{aligned}$$

Under H_A^δ , $\sqrt{T}A_T$ and hence EP converge in distribution to $N(\delta, 1)$. For \tilde{A}_T , we have

$$\begin{aligned} E[\tilde{A}_T] &= \Pr[\alpha + \text{sign}(y_{t-1})\varepsilon_t \geq 0] - \Pr[\alpha + \text{sign}(y_{t-1})\varepsilon_t < 0] = \\ &= \Pr[\varepsilon_t \geq -\alpha] - \Pr[\varepsilon_t < -\alpha] = \Phi(\alpha) - \Phi(-\alpha) = 2\phi(0)\alpha + o(\alpha) \end{aligned}$$

(where the second equality follows from independence of y_{t-1} and ε_t and symmetry of ε_t), and $\text{var}(\tilde{A}_T) = T^{-1} + o(T^{-1})$, so that under H_A^δ , $\sqrt{T}\tilde{A}_T$ and hence DA converge in distribution to $N(2\phi(0)\delta, 1)$.

References

- Brock, W.A., J. Lakonishok, and B. LeBaron (1992) Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance* 47, 1731–1764.
- Christoffersen, P.F. and F.X. Diebold (2003) Financial asset returns, direction-of-change forecasting, and volatility dynamics. Manuscript, University of Pennsylvania.
- Diebold, F.X. and J.A. Nason (1990) Nonparametric exchange rate prediction. *Journal of International Economics* 28, 315–332.
- Gençay, R. (1998) Optimization of technical trading strategies and the profitability in security markets. *Economics Letters* 59, 249–254.
- Härdle W. (1990) *Applied Nonparametric Regression*. Cambridge University Press, New York.
- Hart, J.D. (1996) Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics* 6, 115–142.
- Leitch, G. and J.E. Tanner (1991) Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review* 81, 580–590.
- McCracken, M.W. (2000) Robust out-of-sample inference. *Journal of Econometrics* 99, 195–223.

Meese, R.A. and A.K. Rose (1991) An empirical assessment of non-linearity in models of exchange rate determination. *Review of Economic Studies* 58, 603–619.

Mizrach B. (1992) Multivariate nearest-neighbor forecasts of EMS exchange rates. *Journal Of Applied Econometrics* 7, 151–163.

Pesaran, M.H. and A. Timmermann (1992) A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* 10, 561–565.

Pesaran, M.H. and A. Timmermann (1995) Predictability of stock returns: robustness and economic significance. *Journal of Finance* 50, 1201–1228.

Qi, M. (1999) Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics* 17, 419–429.

Satchell, S. and A. Timmermann (1995) An assessment of the economic value of non-linear foreign exchange rate forecasts. *Journal of Forecasting* 14, 477–497.

Skouras, S. (2000) Risk neutral forecasting. Manuscript, Santa Fe Institute.

Skouras, S. (2001) Financial returns and efficiency as seen by an artificial technical analyst. *Journal of Economic Dynamics and Control* 25, 213–244.

Timmermann, A. and C. Granger (2004) Efficient market hypothesis and forecasting. *International Journal of Forecasting* 20, 15–27.

West, K.D. (1996) Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

West, K.D. and M.W. McCracken (2002) Inference about predictive ability. In: M. Clements and D. Hendry (eds.) *Companion to Economic Forecasting*, Oxford: Blackwell.

Table 1												
Model	DA statistic						EP statistic					
	10%		5%		1%		10%		5%		1%	
	OLS	NN	OLS	NN	OLS	NN	OLS	NN	OLS	NN	OLS	NN
Power												
AR	40.7	25.0	30.1	16.3	13.5	5.8	51.9	32.5	40.9	22.6	21.5	9.15
SETAR	39.3	39.8	29.3	29.6	14.3	13.9	50.2	52.1	40.1	41.1	22.6	22.1
Size												
Const	9.8	10.1	4.6	4.8	1.1	1.0	9.8	10.4	5.1	5.2	1.7	1.1
GARCH	9.3	9.9	4.5	4.9	0.7	1.0	14.1	11.3	7.9	6.1	2.5	1.2

Notes: Entries are actual rejection frequencies from 10,000 simulations of series with sample size 1,000, corresponding to nominal rejection frequencies of 10%, 5%, and 1%. In each simulation, estimation is performed over a rolling window of 100 observations. The DA and profitability statistics are computed over 900 predictions. DGPs (10) and (11) are used for power, DGPs (12) and (13) – for size comparisons. OLS refers to the OLS linear regression, NN – to the nearest neighbors local regression. The EP and DA statistics defined in (6) and (7), respectively, are asymptotically distributed as $N(0,1)$ under the null of no predictability.

Table 2			
Predictor	Sum of squared prediction errors	DA statistic	EP statistic
RW		1.07	
OLS		1.10	2.52
NN		1.14	4.03

Notes: RW refers to the “buy-and-hold” strategy, OLS – to the OLS linear regression, NN – to the nearest neighbors local regression. Estimation is performed over a rolling window of 104 observations. The EP and DA statistics are computed over 2677 predictions. The EP and DA statistics defined in (6) and (7), respectively, are asymptotically distributed as $N(0,1)$ under the null of no predictability.

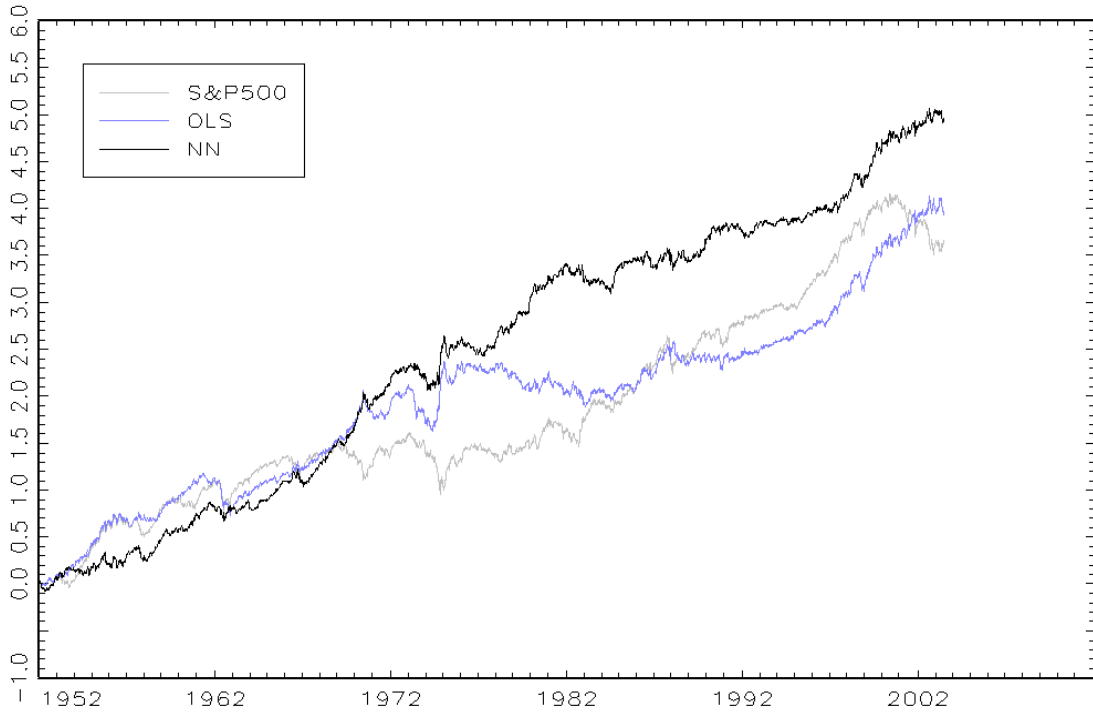


Figure 1: Log cumulative returns, weekly S&P500 index.

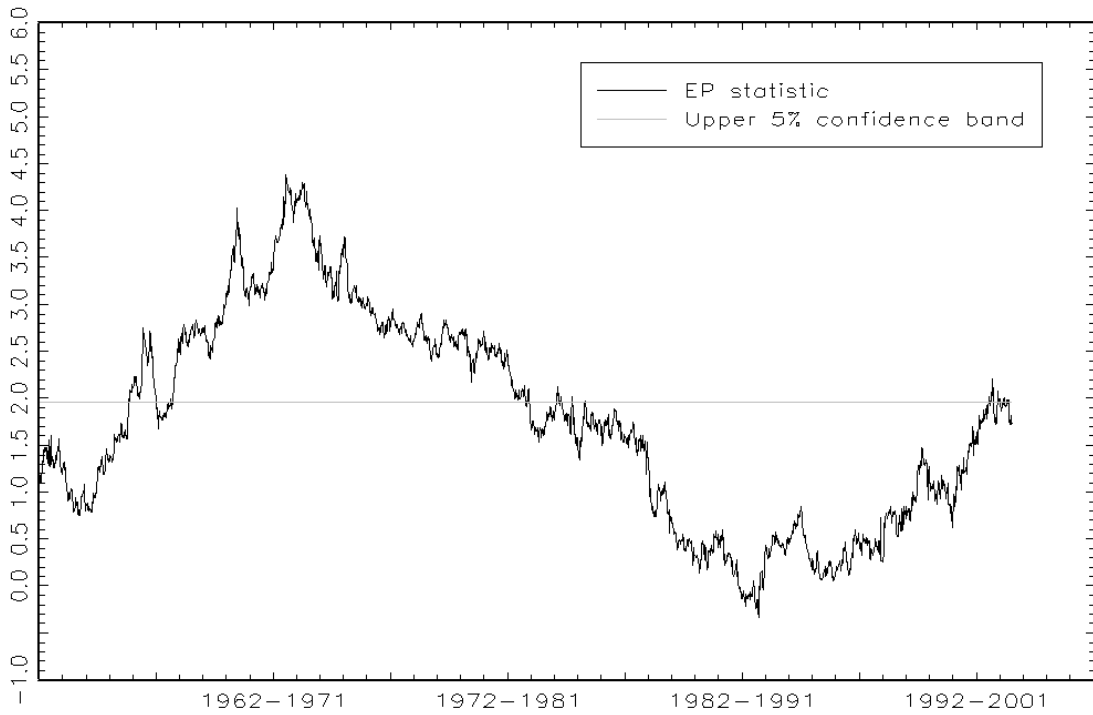


Figure 2: Evolution of mean predictability, weekly S&P500 index.