

Economic and Statistical Measures of Forecast Accuracy*

Clive W.J. Granger

M. Hashem Pesaran

University of California, San Diego

Trinity College, Cambridge

May 1999

Abstract

This paper argues in favour of a closer link between the decision and the forecast evaluation problems. Although the idea of using decision theory for forecast evaluation appears early in the dynamic stochastic programming literature, and has continued to be used with meteorological forecasts, it is hardly mentioned in standard academic texts on economic forecasting. Some of the main issues involved are illustrated in the context of a two-state, two-action decision problem as well as in a more general setting. Relationships between statistical and economic methods of forecast evaluation are discussed and useful links between Kuipers score used as a measure of forecast accuracy in the meteorology literature and the market timing tests used in finance are established. An empirical application to the problem of stock market predictability is also provided, and the conditions under which such predictability could be exploited in the presence of transaction costs are discussed.

JEL Classifications: C10, C20, C22

Key Words: Decision theory, forecast evaluation, probabilistic forecasts, economic and statistical measures of forecast accuracy, stock market predictability.

*A preliminary version of this paper was presented at a meeting of the Royal Meteorological Society, London, December 1998. Address for correspondence: Clive W.J. Granger, Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, USA, Tel: (619) 534-3845, Fax: (619) 534-7040; or M. Hashem Pesaran, Faculty of Economics and Politics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD, E-Mail: hashem.pesaran@econ.cam.ac.uk.

1 Introduction

The literature on forecast evaluation in econometrics has been primarily concerned with statistical accuracy measures of point forecasts, and little attention has been paid to the economic importance of these forecasts. Forecasts are not made in isolation but are designed for use in decision making and so there is need for consideration of the interaction between the modeler who produces forecasts and the decision maker who consumes them. It would generally be considered inappropriate for a model to be built specifically for a single user, such as a think-tank building a macro model for a particular political party, then forecasts may be biased in ways that the buyer would like to hear. Clearly this could happen with a forecasting model built “in-house” for a company. Rather, we will consider a model built by an independent provider whose forecasts may potentially be used by many decision makers. A decision theoretical approach is then appropriate for forecast evaluation by each decision maker. This approach will be particularly useful for comparing the relative usefulness of forecasts from different models. Although the idea of using decision theory for forecast evaluation appears early in the dynamic stochastic programming literature, White (1966), and has continued to be used with meteorological forecasts, as discussed by Katz and Murphy (1997) and possibly elsewhere, it is hardly mentioned in standard academic texts on forecasting such as Box and Jenkins (1970), Granger and Newbold (1977, 1986), and Clements and Hendry (1998, 1999). The only clear exception is the book by Theil (2nd edition, 1960) whose Sections 8.4 and 8.5 are similar in spirit to our discussion although quite different in technique. There are, of course, many books on decision theory but they are rarely phrased in terms of forecasting. There are also a number of exceptions in the application of statistical techniques to financial time series. Leith and Tanner (1991), for example, argue that in evaluation of stock market forecasts profits (losses) generated from using the forecasts are more appropriate than the conventional statistical measures of forecast accuracy. Similar arguments are also made in Pesaran and Timmermann (1994, 1995).

This paper, which will largely consist of a simple example considered in detail, illustrates the advantages of taking a decision theoretic approach to forecast evaluation. In particular it will be suggested that point forecasts are often less appropriate than predictive distributions and that evaluation need not be based on cost functions using forecast errors. The paper also discusses some of the relationships that exist between statistical and economic methods of forecast evaluation. It establishes useful links between Kuipers score used as a measure of forecast accuracy in the meteorology literature and the market timing tests used in finance. In particular, it shows that the

market timing test statistic advanced in Pesaran and Timmermann (1992) is in fact a standardized version of the Kuipers score (defined as the difference between the hit and the false alarm rates) used extensively in the meteorology literature. The paper also provides an empirical application to the problem of stock market predictability and discusses the conditions under which such predictability could be exploited in the presence of transaction costs.

The plan of the paper is as follows: Section 2 sets up a slightly generalized version of the two-state, two-action decision problem and derives the optimal decision rule. The problem of evaluation and comparison of forecasts is addressed in Section 3. This section reviews a number of statistical measures used in the meteorology literature for evaluation of probability forecasts, establishes links between economic and statistical measures of forecast accuracy, and derives the algebraic relationship that exists between the Kuipers score and the market timing test statistic. Section 4 considers a simple portfolio decision problem, derives switching rules between stocks and bonds in the presence of transaction costs and provides some empirical evidence on the statistical and economic importance of stock market predictability using data from U.S. stock market. Section 5 reconsiders the problem of forecast evaluation in the context of a more general decision model. Section 6 provides some concluding remarks.

2 Event and Probability Forecasts in the Context of a Simple Model

Consider a situation in which there are two “states” of the world, which for ease will be called “Bad” and “Good”. A forecast will be made on day $t-1$ of the situation on day t , for any t . Let $\hat{\pi}_t$ denote the forecast probability that the Bad event will occur on day t . Thus the forecast probability of the Good event is $1-\hat{\pi}_t$. Note that $\hat{\pi}_t$ and $1-\hat{\pi}_t$ are not point forecasts or even an interval forecast but the whole distribution is given for all possible outcomes. One could think of an applied economist constructing a model from which the forecast probabilities are generated. We presume the existence of an actual generation process for the events and let π_t be the probability of the Bad event occurring on day t according to this process. As an alternative to the probability forecasts, a point forecast \hat{z}_t could be provided with $\hat{z}_t = 1$ if the Bad state is forecast to occur or $\hat{z}_t = 0$ otherwise. Similarly, $z_t = 1$ denotes the situation when the Bad event actually occurs and $z_t = 0$ otherwise. The economists might have some “rule of thumb” which gives $\hat{z}_t = 1$ if $\hat{\pi}_t$ exceeds some specified probability threshold, $p_t \in (0, 1)$. An obvious value is $p_t = 0.5$

but, as will be seen, other values could be used. The economic forecaster thus has two alternative forms of forecast to announce, either $\hat{\pi}_t$, which takes some value in the region $0 \leq \hat{\pi}_t \leq 1$, and represents a probability forecast; or \hat{z}_t which corresponds to the point (or event) forecast. Note that as presented \hat{z}_t can be derived from $\hat{\pi}_t$ but not *vice versa*. The relationship between the event and probability forecasts can also be written as $\hat{z}_t = I(\hat{\pi}_t - p_t)$, where the indicator function $I(\cdot)$, is defined by $I(\mathfrak{A}) = 1$ if $\mathfrak{A} > 0$, and $I(\mathfrak{A}) = 0$, otherwise.

A decision maker will use the forecasts to decide whether or not to take some action. As an example, if the forecast is that prices will rise by more than 2.5% over the next quarter (the Bad event), a central bank may decide to raise interest rates. In another simple example, provided by Stewart (1997), a building contractor has to decide whether or not to pour concrete for a foundation and the decision will be based on a forecast of whether it will rain tomorrow (Bad) or not. A similar example is a local government's road authority which has to either sand/salt local roads to counter icing problems in the evening depending on a forecast of a very low temperature (Bad) or not during the night. To consider the interactions of the states and the decisions, there will be a payoff matrix of "values" or "utilities" of the form

Table 1: Payoff Matrix for a Two-State, Two-Action Decision Problem

		States (s_t)	
		Bad ($s_t = 1$)	Good ($s_t = 0$)
Decisions (d_t)	Yes ($d_t = 1$)	$U_{by}(t)$	$U_{gy}(t)$
	No ($d_t = 0$)	$U_{bn}(t)$	$U_{gn}(t)$

where $U_{by}(t)$ is the economic value to the decision maker at time t if the Bad state occurs after the Yes decision was taken, $U_{bn}(t)$ is the value in the Bad state when no action is taken and so forth. These values could be utilities but in practice it is often easier to think of them as measurable quantities such as benefits or costs. If the convention is used that a larger U is preferred to a smaller one, that a Good State produces higher values than a Bad one, and that a Yes decision produces some benefit in the Bad state one gets the following inequalities:

$$\text{Taking correct action is beneficial: } U_{by}(t) > U_{bn}(t),$$

$$\text{Taking false action is costly: } U_{gn}(t) > U_{gy}(t).$$

We shall refer to $U_{gn}(t) - U_{gy}(t)$ as the cost of taking *false* action, $U_{by}(t) - U_{bn}(t)$ as the benefit of taking *correct* action, and denote the ratio of the cost of taking false action to the benefit of taking correct action by c_t

$$c_t = \frac{U_{gn}(t) - U_{gy}(t)}{U_{by}(t) - U_{bn}(t)} > 0, \text{ for all } t.$$

We refer to c_t as the cost-benefit ratio.¹ To complete the characterization of the decision process, we need to specify the following four conditional probabilities:

$$\begin{aligned} \Pi_{t,gy} &= \text{Prob}(s_t = 0 | d_t = 1, \Omega_{t-1}), \\ \Pi_{t,gn} &= \text{Prob}(s_t = 0 | d_t = 0, \Omega_{t-1}) \\ \Pi_{t,by} &= \text{Prob}(s_t = 1 | d_t = 1, \Omega_{t-1}), \\ \Pi_{t,bn} &= \text{Prob}(s_t = 1 | d_t = 0, \Omega_{t-1}), \end{aligned}$$

where Ω_{t-1} is the information available to the decision maker at time $t - 1$, and d_t and s_t are defined in Table 1. In many situations the probability of the Bad (or Good) state at time t is invariant to whether action is taken by the decision maker at time $t - 1$. For example, we do not expect that our decision to carry an umbrella will alter the probability of rain, or the sale of our individual (small) holdings of stocks is likely to change the probability of a stock market crash. In decision theory these are known as games played against nature and simplify the matrix of conditional probabilities to

$$\pi_t = \Pi_{t,by} = \Pi_{t,bn} = 1 - \Pi_{t,gy} = 1 - \Pi_{t,gn}. \quad (1)$$

However, there are clearly circumstances where the probability of the Bad state occurring *does* depend on whether action is taken. All strategic game problems fall in this category and are subject to the additional uncertainty of the effect of actions on outcomes. In this paper we shall confine our analysis to the former type of decision problems and assume that the decision maker computes expected values of a “yes” and a “no” decision using an estimate of π_t , which we denote by $\hat{\pi}_t$. This estimate could be based on an econometric model, could be purely subjective, or could be estimated unconditionally using sample frequencies known as climatological frequency in the metrological literature.

¹Notice that c_t is invariant to affine transformations of the utilities $U_{ij}(t)$, $i = y, n$, $j = g, b$; suggesting that utilities are at best measurable from observations on the decision maker’s actions only up to an affine transformation.

Under the above set up the expected value (utility) of **taking** action is given by²

$$U_{by}(t)\hat{\pi}_t + U_{gy}(t)(1 - \hat{\pi}_t),$$

and the expected value of **not taking** action by

$$U_{bn}(t)\hat{\pi}_t + U_{gn}(t)(1 - \hat{\pi}_t),$$

and so action is taken if

$$\hat{\pi}_t > \frac{U_{gn}(t) - U_{gy}(t)}{U_{by}(t) - U_{bn}(t) + U_{gn}(t) - U_{gy}(t)} = \frac{c_t}{1 + c_t} = q_t. \quad (2)$$

Alternatively, this condition can be written as

$$\frac{\hat{\pi}_t}{1 - \hat{\pi}_t} > c_t > 0. \quad (3)$$

Namely, action will be taken if the predicted odds of the bad event occurring exceeds the cost-benefit ratio, c_t , associated with taking action. In the meteorology literature a simple version of the above decision problem is used where q_t (taken to be fixed) is referred to as the “cost-loss” ratio. In the more general set up of this paper we shall refer to q_t as the “payoff” ratio. It is clear that $0 \leq q_t \leq 1$. The extreme values of $q_t = 0$ and $q_t = 1$ correspond to the cases where cost-benefit ratio of the decision is either zero or infinitely large. Action will always be taken if $q_t = 0$. On the other extreme action will never be taken if $q_t = 1$. Neither of these extremes are of much practical interest. So in what follows we focus on situations where $0 < q_t < 1$.

The above framework can be readily extended to decision problems with 3 or more states and/or actions. An m -state, two-action generalization is considered in Granger and Pesaran (1996). An example with two states and three actions is discussed in Section 4.

2.1 Economic Value of Forecasts

The time t realized value of the economic benefit of the decision based on the probability forecast $\hat{\pi}_t$, is given by

$$\begin{aligned} v_t(\hat{\pi}_t) = & U_{by}(t)z_t I(\hat{\pi}_t - q_t) \\ & + U_{gy}(t)(1 - z_t)I(\hat{\pi}_t - q_t) \\ & + U_{bn}(t)z_t \{1 - I(\hat{\pi}_t - q_t)\} \\ & + U_{gn}(t)(1 - z_t) \{1 - I(\hat{\pi}_t - q_t)\}, \end{aligned} \quad (4)$$

²Here for simplicity of exposition we are assuming that $U_{ij}(t)$ are known to the decision maker with certainty. But it is relatively easy to relax this assumption.

which can also be written equivalently as

$$v_t(\hat{\pi}_t) = a_t + b_t(z_t - q_t)I(\hat{\pi}_t - q_t), \quad (5)$$

where

$$a_t = z_t U_{bn}(t) + (1 - z_t)U_{gn}(t), \quad (6)$$

and

$$b_t = U_{by}(t) - U_{bn}(t) + U_{gn}(t) - U_{gy}(t) > 0. \quad (7)$$

The *ex ante* economic value of using the probability forecast $\hat{\pi}_t$ is given by³

$$E[v_t(\hat{\pi}_t) | \Omega_{t-1}] = \pi_t U_{bn}(t) + (1 - \pi_t)U_{gn}(t) + b_t(\pi_t - q_t)I(\hat{\pi}_t - q_t). \quad (8)$$

where only the last term, $b_t(\pi_t - q_t)I(\hat{\pi}_t - q_t)$, varies across different forecasts, $\hat{\pi}_t$. It is clear that this term can be negative for some values of π_t , $\hat{\pi}_t$ and q_t , unless $\hat{\pi}_t = \pi_t$. Due to the discreteness of the event space the optimum solution for a particular realization is not unique. There exists an optimum solution set defined by all $\hat{\pi}_t$, with values $\hat{\pi}_t \geq q_t$ if $\pi_t \geq q_t$, or $\hat{\pi}_t < q_t$ if $\pi_t < q_t$. In this solution set the solution $\hat{\pi}_t = \pi_t$, if it can be achieved, has the particular advantage that it does not depend on the payoff ratio, q_t , and therefore is superior to all the other solutions. For this reason we refer to the solution $\hat{\pi}_t = \pi_t$ as the “supreme” solution.⁴

It should be noted that the supreme solution provides the complete forecast distribution, conditional on a particular information set. If there are several users, with different values of q_t , they can all use the supreme forecasts, but this is not true for other solutions in the optimal set. To obtain the supreme solution one will need to know the “true” conditional probability distribution function of the event; and the above analysis suggests that the forecaster should attempt to obtain a good estimate of this probability. One can view the forecasters as producers of goods and the users of forecasts as the consumers of these goods. Without a precise knowledge of the payoff functions of the users of the forecasts, the most appropriate course open to the forecasters is to do their best to obtain the supreme solution.

³Recall that the payoffs $U_{bn}(t)$, $U_{by}(t)$, etc. are known in period $t-1$, and $E(z_t | \Omega_{t-1}) = \pi_t$.

⁴For further details see Granger and Pesaran (1996).

3 Evaluation and Comparison of Forecasts

In this section we establish the link between forecast evaluation and the simple decision problem set out above. We can imagine having a span of dates $t = 1, 2, \dots, T$ for which z_t is observed and also probability forecasts from two competing models giving $\hat{\pi}_t^{(1)}$ and $\hat{\pi}_t^{(2)}$. The aim is to evaluate and compare these probability forecasts.

It is widely recognized that evaluation and comparison of forecasts critically depends on the choice of the loss function, $L(z_t, \hat{\pi}_t)$.⁵ A number of different loss functions have been used in the literature, but by far the most popular is the quadratic loss function, $L(z_t, \hat{\pi}_t) = (z_t - \hat{\pi}_t)^2$. This loss function is routinely used for evaluation of point forecasts, but it has also been applied extensively in the meteorology literature for evaluation of probability and event forecasts. The resultant accuracy measure is known as the Brier score (Brier, 1950) which is defined by⁶

$$B = \frac{1}{T} \sum_{t=1}^T (z_t - \hat{\pi}_t)^2. \quad (9)$$

As with the conventional mean squared forecast errors (MSFE) for point forecasts, the Brier score is identically zero only in the case of perfect deterministic forecasts and has a maximum value of unity in the case of forecasts that are consistently wrong. An interesting decomposition of the Brier score into measures of “reliability”, “resolution” and “uncertainty” has been derived by Murphy (1973). This decomposition is based on the classification of the forecast probabilities into m (say) groups with each group containing T_i observations. Denoting these groups by Υ_i and the corresponding group-specific forecast probabilities by y_i (where $y_1 = 0$, and $y_m = 1$) Murphy’s decomposition can then be written as

$$B = \frac{1}{T} \sum_{i=1}^m T_i (y_i - \bar{z}_i)^2 - \frac{1}{T} \sum_{i=1}^m T_i (\bar{z}_i - \bar{z})^2 + \bar{z}(1 - \bar{z}), \quad (10)$$

where

$$\bar{z}_i = \frac{\sum_{t \in \Upsilon_i} z_t}{T_i}, \quad \text{and} \quad \bar{z} = \frac{\sum_{t=1}^T z_t}{T}.$$

⁵See, for example, West et al. (1993), Diebold and Mariano (1995), Weiss (1996), and Granger and Pesaran (1996).

⁶See, for example, Wilks (1995, Ch. 7) and Katz and Murphy (1997, Ch. 2). As Wilks (1995, p. 260) points out the ordinal measure proposed by Brier (1950) was in fact twice the measure commonly used.

In this decomposition only the first two terms (known as reliability and resolution) depend on forecast probabilities. The third term, $\bar{z}(1 - \bar{z})$, depends only on the sample observations and in the present setup is not affected by the forecasts. Based on Murphy’s decomposition the following forecasting skill score (SS), which measures the skill of the forecaster relative to the climatological (unconditional) forecasts, is routinely used in practice:

$$SS_b = \frac{\text{Resolution-Reliability}}{\text{Uncertainty}} = 1 - \frac{B}{\bar{z}(1 - \bar{z})}.$$

SS_b , lies between $1 - 1/[\bar{z}(1 - \bar{z})]$ and unity, and clearly can be negative if the Brier score is larger than the uncertainty component, $\bar{z}(1 - \bar{z})$. Only forecasts with a positive skill score ($SS_b > 0$) are said to be skillful.

Other measures of forecast accuracy, arguably more suited to evaluating probability (or event) forecasts, are also considered in the metrology literature. Most of these measures are derived from the contingency table of realizations and forecasts as in Table 2. A prominent example of such measures is the Kuipers score (KS) defined by⁷

$$KS = H(\mathbf{q}) - F(\mathbf{q}), \quad (11)$$

where $\mathbf{q} = (q_1, q_2, \dots, q_T)'$, $H(\mathbf{q})$ is the proportion of Bad events that were correctly forecast to occur, and $F(\mathbf{q})$ is the proportion of Good events that were incorrectly forecast (or were acted upon). These two proportions are known as the “hit rate” and the “false alarm rate”, respectively.

Table 2: Contingency Matrix of Realizations and Forecasts/Actions

		<i>Realizations (z_t)</i>	
		Bad ($z_t = 1$)	Good ($z_t = 0$)
<i>Forecasts/Actions</i>	Yes ($\hat{\pi}_t > q_t$)	<i>Hits (T_{by})</i>	<i>False Alarms (T_{gy})</i>
	No ($\hat{\pi}_t \leq q_t$)	<i>Misses (T_{bn})</i>	<i>Correct Rejections (T_{gn})</i>

In terms of the information summarized in the above contingency matrix we have

$$H(\mathbf{q}) = \frac{T_{by}}{T_{by} + T_{bn}}, \quad F(\mathbf{q}) = \frac{T_{gy}}{T_{gy} + T_{gn}}. \quad (12)$$

⁷Kuiper’s score, also known as the Hanssen-Kuipers discriminant or Kuiper’s performance index, was originally proposed by Peirce (1884). For an account of this and other measures of forecasting skill see Murphy and Dann (1985) and Wilks (1995, Ch. 7).

Clearly, $T = T_{by} + T_{bn} + T_{gy} + T_{gn}$.⁸ In Table 2, the probability forecasts, $\hat{\pi}_t$, $t = 1, \dots, T$, are in effect converted into deterministic or event forecasts, $\hat{z}_t = I(\hat{\pi}_t - p_t)$, using the payoff ratio, q_t , as the appropriate threshold probability. Without reference to the decision problem it is difficult to decide what value should be chosen for p_t . This also sheds doubt on the usual practice of using a “rule of thumb” such as setting $\hat{z}_t = 1$, if $\hat{\pi}_t > 0.5$. Also, in general, there is no reason for p_t to be the same for forecasts obtained at different time periods and/or for different individuals.

The Kuipers score has the desirable feature that random forecasts or forecasts that consistently predict Bad or Good events will score zero; while the same need not be true of Brier score. This is known as the “equitability” property, see Gandin and Murphy (1992). For example, a doomsday forecaster who systematically predicts Bad events will have a hit rate of 100 per cent, but will also have a false alarm rate of 100 per cent; scoring zero on the Kuipers measure. The Brier score of a doomsday forecaster could be quite respectable as compared to forecasters with non-zero skills on the Kuipers scale. Nevertheless, the KS measure is also somewhat arbitrary and does not fully relate to the decision problem. Given the realizations and forecasts, z_t and $\hat{\pi}_t$, $t = 1, 2, \dots, T$, it is possible to compute the average economic value of the probability forecasts to the decision maker using (5), namely

$$V(\hat{\boldsymbol{\pi}}, \mathbf{q}) = \bar{a} + T^{-1} \sum_{t=1}^T b_t(z_t - q_t)I(\hat{\pi}_t - q_t), \quad (13)$$

where $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_T)'$, and $\bar{a} = T^{-1} \sum_{t=1}^T a_t$, is the same across all probability forecasts and therefore can be normalized to zero without loss of generality. Also recall that $b_t > 0$.

Based on (13), a value-based forecasting skill score, SS_v , can also be constructed:

$$SS_v = \frac{\sum_{t=1}^T b_t(z_t - q_t) [I(\hat{\pi}_t - q_t) - I(\pi_t^0 - q_t)]}{\sum_{t=1}^T b_t(z_t - q_t) [I(z_t - q_t) - I(\pi_t^0 - q_t)]},$$

where π_t^0 is the reference probability forecast, often taken to be the climatological probability of the Bad event. As with other skill scores we have $SS_v \leq 1$, and SS_v attains its maximum in the unlikely event where $(\hat{\pi}_t - q_t)(z_t - q_t) > 0$ for *all* t .

The average economic value, $V(\hat{\boldsymbol{\pi}}, \mathbf{q})$, can also be used to directly compare two or more probability forecasts. For example, the probability forecasts $\hat{\pi}_t^{(1)}$,

⁸Strictly speaking all the four components of T also depend on \mathbf{q} . But this is suppressed for notational convenience.

$t = 1, 2, \dots, T$, are preferable to the alternative forecasts $\hat{\pi}_t^{(2)}$ if

$$T^{-1} \sum_{t=1}^T b_t(z_t - q_t) I(\hat{\pi}_t^{(1)} - q_t) > T^{-1} \sum_{t=1}^T b_t(z_t - q_t) I(\hat{\pi}_t^{(2)} - q_t), \quad (14)$$

and *vice versa*. A comparison of particular interest is the one between probability forecasts, $\hat{\pi}_t$, and the associated event forecasts $\hat{z}_t = I(\hat{\pi}_t - p_t)$. Applying the above criterion, $\hat{\pi}_t$ will be preferred to \hat{z}_t if

$$T^{-1} \sum_{t=1}^T b_t(z_t - q_t) I(\hat{\pi}_t - q_t) > T^{-1} \sum_{t=1}^T b_t(z_t - q_t) I(\hat{z}_t - q_t).$$

But since $0 < q_t < 1$, $I(I(\hat{\pi}_t - p_t) - q_t) = I(\hat{\pi}_t - p_t)$ and the condition becomes:

$$T^{-1} \sum_{t=1}^T b_t(z_t - q_t) I(\hat{\pi}_t - q_t) > T^{-1} \sum_{t=1}^T b_t(z_t - q_t) I(\hat{\pi}_t - p_t). \quad (15)$$

Clearly the two methods of reporting forecasts will be equivalent if the threshold probability, p_t , is chosen to be the same as the payoff ratio q_t for all t . But since q_t differs across different decision makers and event forecasts are typically generated by a forecaster using the same probability threshold value, in general probability and event forecasting can lead to different outcomes. To obtain conditions under which probability forecasts are superior to event forecasts consider the following decomposition of (15):

$$\begin{aligned} & T^{-1} \sum_{t=1}^T b_t(\hat{\pi}_t - q_t) [I(\hat{\pi}_t - q_t) - I(\hat{\pi}_t - p_t)] + \\ & T^{-1} \sum_{t=1}^T b_t(z_t - \hat{\pi}_t) [I(\hat{\pi}_t - q_t) - I(\hat{\pi}_t - p_t)] > 0 \end{aligned}$$

Since $b_t > 0$, and $0 < p_t, q_t < 1$ it is then easily established that all the elements in the first sum are non-negative. The elements in the second sum, have ambiguous signs and for a finite T the sign of the second term will also be ambiguous. It is therefore not possible to rank the probability and event forecasts in finite samples. But for sufficiently large T probability forecasting is superior to event forecasting under the condition

$$\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T b_t(z_t - \hat{\pi}_t) [I(\hat{\pi}_t - q_t) - I(\hat{\pi}_t - p_t)] = 0.$$

This condition can be simplified further by noting that

$$\begin{aligned} I(\widehat{\pi}_t - q_t) - I(\widehat{\pi}_t - p_t) &= 1 \quad \text{if } p_t > \widehat{\pi}_t > q_t, \\ &= -1 \quad \text{if } q_t > \widehat{\pi}_t > p_t, \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

Therefore,

$$\begin{aligned} & T^{-1} \sum_{t=1}^T b_t(z_t - \widehat{\pi}_t) [I(\widehat{\pi}_t - q_t) - I(\widehat{\pi}_t - p_t)] \\ &= T^{-1} \sum_{p_t > \widehat{\pi}_t > q_t} b_t(z_t - \widehat{\pi}_t) - T^{-1} \sum_{q_t > \widehat{\pi}_t > p_t} b_t(z_t - \widehat{\pi}_t), \end{aligned}$$

and for the superiority of $\widehat{\pi}_t$ over \widehat{z}_t , it is sufficient that

$$\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{p_t > \widehat{\pi}_t > q_t} b_t(z_t - \widehat{\pi}_t) = 0,$$

and

$$\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{q_t > \widehat{\pi}_t > p_t} b_t(z_t - \widehat{\pi}_t) = 0.$$

In the standard case discussed in the literature where b_t is a fixed constant these conditions reduce to asymptotic unbiasedness conditions of the probability forecasts, $\widehat{\pi}_t$, grouped according to whether $p_t > \widehat{\pi}_t > q_t$, or $q_t > \widehat{\pi}_t > p_t$.

The economic valuation approach to forecast comparisons can also be applied to the problem of combining probability forecasts. The two probability forecasts could be combined in any relevant fashion, such as linearly

$$\widehat{\pi}_t^{(c)} = \theta \widehat{\pi}_t^{(1)} + (1 - \theta) \widehat{\pi}_t^{(2)} \quad 0 \leq \theta \leq 1,$$

giving an average economic value

$$V(\widehat{\boldsymbol{\pi}}^{(c)}, \mathbf{q}) = T^{-1} \sum_{t=1}^T b_t(z_t - q_t) I\left(\theta \widehat{\pi}_t^{(1)} + (1 - \theta) \widehat{\pi}_t^{(2)} - q_t\right),$$

and one could search over $\theta \in [0, 1]$ to obtain the maximum average value available from such combinations. Clearly, from its method of construction, the optimum average value will be no less than $\text{Max}\left(V(\widehat{\boldsymbol{\pi}}^{(1)}, \mathbf{q}), V(\widehat{\boldsymbol{\pi}}^{(2)}, \mathbf{q})\right)$, as one could select $\theta = 0$ or 1.

3.1 Relationship Between Economic and Statistical Measures of Forecast Accuracy

In general, a simple one-to-one relationship between the purely statistical measures of forecast accuracy such as the Brier or the Kuipers scores, and the value-based measures such as $V(\hat{\boldsymbol{\pi}}, \mathbf{q})$ does not exist. However, in the case where $b_t = b$ and $q_t = q$ for all t (the standard case considered in the literature) a simple relation between the average economic value of the Kuipers score can be derived.⁹ From the definitions of the hit rate given by (12) it is easily seen that

$$H(q) = \frac{\sum_{t=1}^T z_t I(\hat{\pi}_t - q)}{T \bar{z}}.$$

Similarly,

$$F(q) = \frac{\sum_{t=1}^T (1 - z_t) I(\hat{\pi}_t - q)}{T(1 - \bar{z})}.$$

Therefore,

$$T^{-1} \sum_{t=1}^T z_t I(\hat{\pi}_t - q) = \bar{z} H(q), \text{ and } T^{-1} \sum_{t=1}^T I(\hat{\pi}_t - q) = (1 - \bar{z}) F(q) + \bar{z} H(q).$$

Using these results in (13), we have

$$V(\hat{\boldsymbol{\pi}}, q) = b [(1 - q) \bar{z} H(q) - q(1 - \bar{z}) F(q)]. \quad (16)$$

There is a one-to-one relationship between $V(\hat{\boldsymbol{\pi}}, q)$ and the Kuipers score only in the case where the payoff ratio is constant over time and equal to the unconditional forecast probability, namely if $q = \bar{z}$. In this case

$$V(\hat{\boldsymbol{\pi}}, q) = b \bar{z} (1 - \bar{z}) K S, \quad (17)$$

and the use of Kuipers' score as an indicator of economic value is justified.

In practice where the same probability forecasts are used by many different decision makers with different payoff ratios the economic value of forecasts varies across individuals and comparison of alternative forecasts needs to take this into account. Suppose there are N individual decision makers indexed by i , and denote the individual-specific economic values by $V_i(\hat{\boldsymbol{\pi}}, q_i)$, where

⁹See also Palmer et al. (1998) and Richardson (1998) for a discussion of such relationships in the meteorology literature.

q_i is individual i 's payoff ratio. The average economic value to the society of using the probability forecasts $\hat{\pi}$, issued by a forecasting agency is given by

$$\bar{V}(\hat{\pi}) = N^{-1} \sum_{i=1}^N b_i [\bar{z}(1 - q_i)H_T(q_i) - (1 - \bar{z})q_i F_T(q_i)].$$

Characterizing the distribution of the payoff ratios across individuals by the density function $f(q)$ on $q \in [0, 1]$, and assuming that b_i and q_i are independently distributed across i , for sufficiently large N , we have

$$\bar{V}(\hat{\pi}) \approx b \left\{ \bar{z} \int_0^1 (1 - q)H_T(q)f(q)dq - (1 - \bar{z}) \int_0^1 qF_T(q)f(q)dq \right\},$$

where $b = E(b_i)$. A good choice for $f(q)$ is the Beta distribution, which could be calibrated to match the in-sample mean and variance of the payoff ratios across individuals.

3.2 Relationship Between Statistical Measures of Forecast Accuracy and Tests of Market Timing

Despite their attractive features statistical measures such as the Brier and Kuipers scores do not seem to have been used in the econometrics or finance literature. This is partly explained by economists' pre-occupation with point forecasts. But even when the focus of the analysis is on forecasting turning points or direction of market changes the tendency has been to employ model-based approaches such as probit or logit models or market timing tests originally developed by Henriksson and Merton (1981). In this section we investigate the relationship between the Kuipers score and a particular version of the market timing test statistic advanced in Pesaran and Timmermann (PT, 1992). The PT test statistic applied to the probability forecasts, $\hat{\pi}_t$, and the associated realizations, z_t , for $t = 1, 2, \dots, T$ is given by

$$PT = \frac{\hat{P} - \hat{P}_*}{\left\{ \hat{V}(\hat{P}) - \hat{V}(\hat{P}_*) \right\}^{\frac{1}{2}}} \stackrel{a}{\sim} N(0, 1), \quad (18)$$

where \hat{P} is the proportion of events (Bad or Good) that are correctly predicted, \hat{P}_* is the estimate of the probability of correctly predicting the events assuming predictions and realizations are independently distributed, $\hat{V}(\hat{P})$ and $\hat{V}(\hat{P}_*)$ are consistent estimates of the variances of \hat{P} and \hat{P}_* , respec-

tively.¹⁰ In terms of z_t and $I(\hat{\pi}_t - q_t)$ we have

$$\hat{P} = T^{-1} \sum_{t=1}^T z_t I(\hat{\pi}_t - q_t) + T^{-1} \sum_{t=1}^T (1 - z_t) [1 - I(\hat{\pi}_t - q_t)],$$

$$\hat{P}_* = \hat{P}_z \hat{P}_\pi + (1 - \hat{P}_z)(1 - \hat{P}_\pi),$$

$$\hat{V}(\hat{P}) = T^{-1} \hat{P}_*(1 - \hat{P}_*),$$

$$\begin{aligned} \hat{V}(\hat{P}_*) &= T^{-1} (2\hat{P}_z - 1)^2 \hat{P}_\pi (1 - \hat{P}_\pi) + T^{-1} (2\hat{P}_\pi - 1)^2 \hat{P}_z (1 - \hat{P}_z) \\ &\quad + 4T^{-2} \hat{P}_z \hat{P}_\pi (1 - \hat{P}_z)(1 - \hat{P}_\pi), \end{aligned}$$

where $\hat{P}_z = \bar{z}$, and $\hat{P}_\pi = T^{-1} \sum_{t=1}^T I(\hat{\pi}_t - q_t)$. It is now easily seen that¹¹

$$T^{-1} \sum_{t=1}^T z_t I(\hat{\pi}_t - q_t) = \bar{z} H(\mathbf{q}),$$

$$T^{-1} \sum_{t=1}^T (1 - z_t) I(\hat{\pi}_t - q_t) = (1 - \bar{z}) F(\mathbf{q}),$$

where the hit and the false alarm rates, $H(\mathbf{q})$ and $F(\mathbf{q})$, are defined by (12). Using the above results we have

$$\begin{aligned} \hat{P} &= \bar{z} H(\mathbf{q}) + (1 - \bar{z}) [1 - F(\mathbf{q})], \\ \hat{P}_* &= \bar{z} \hat{P}_\pi + (1 - \bar{z}) (1 - \hat{P}_\pi), \\ \hat{P}_\pi &= \bar{z} H(\mathbf{q}) + (1 - \bar{z}) F(\mathbf{q}), \end{aligned}$$

which establishes the following exact relationship between the numerator of the PT test statistic and the Kuipers score (KS) defined by (11):

$$\hat{P} - \hat{P}_* = 2\bar{z}(1 - \bar{z}) [H(\mathbf{q}) - F(\mathbf{q})] = 2\bar{z}(1 - \bar{z}) KS.$$

The PT test, however, goes one step further and asks whether the Kuipers score is statistically significant. Keeping only terms of order T^{-1} in $\hat{V}(\hat{P}_*)$ and after some algebra we have

$$PT = \frac{\sqrt{T} KS}{\left\{ \frac{\hat{P}_\pi (1 - \hat{P}_\pi)}{\bar{z}(1 - \bar{z})} \right\}^{1/2}}. \quad (19)$$

¹⁰It is interesting to note that $\hat{P} - \hat{P}_*$ is in fact identical to the Heidke score (Heidke, 1926). For a discussion of the use of the Heidke score see Wilks (1995, pp. 248-249).

¹¹Notice that in this sub-section we are allowing the payoff ratio, q_t , to vary over time.

Since \hat{P}_π varies with the probability forecasts, $\hat{\pi}_t$, it is possible for two sets of probability forecasts to have identical Kuipers scores but different PT test statistics.

4 An Application to the Stock Market

There now exists a substantial empirical literature showing that a statistically significant part of the variations in monthly or quarterly stock market returns are predictable.¹² This evidence naturally raises the question of the economic importance of such predictability. In this section we consider this issue in the case of an investor who wishes to decide between holding stocks or bonds. To simplify the analysis and to ensure that the simple two-state decision model of Section 2 is directly applicable to the present problem we shall assume that the investor's utility function is logarithmic. It is well known that in the case of log-utility a multi-period decision problem can be reduced to a sequence of independent single period decision problems.¹³ We further assume that sales and purchases of stocks and bonds are subject to transaction costs. The marginal cost of transactions for stocks and bonds will be denoted by ξ_{at} and ξ_{bt} , respectively. The transaction cost of trading in stocks is composed of a commission fee and the implied costs associated with the bid-ask spread. It varies with the size of the transaction and could even differ depending on whether stocks are sold or bought. But here we assume the same rate applies to all transactions but allow the transaction costs to vary over time.

4.1 Derivation of the Switching Rules in the Presence of Transaction Costs

In the context of the above setup consider an individual investor who at the beginning of period t owns N_t units of a portfolio of stocks with a unit price of P_t that closely tracks a market index and wishes to decide whether to stay in stocks or to switch his/her portfolio into government bonds paying per period holding return of r_t . To simplify the analysis we confine our attention to an all-or-nothing strategy where the investor decides to stay

¹²See, for example, Campbell (1987), Fama and French (1989), Breen et al. (1990), and Pesaran and Timmermann (1994, 1995) on predictability of stock returns in the US, and Clare, Thomas and Wickens (1994), Black and Fraser (1995) and Pesaran and Timmermann (1999) for evidence of stock market predictability in the UK.

¹³The solution to the problem of optimal portfolio weights in a multi-period context in the case of the power utility function is considered in Campbell and Vicaria (1998) and Brandt (1998). These authors consider alternative approximate solution approaches and do not allow for transaction costs.

either fully in stocks or switch completely into bonds. The payoff matrix of this decision problem is laid out in Table 3, where the rate of return on stocks over the period t to $t + 1$, in the event of market falling is denoted by $R_{t+1}^f = (P_{t+1}^f + D_{t+1}^f - P_t)/P_t$, and in the event of market rising is denoted by $R_{t+1}^r = (P_{t+1}^r + D_{t+1}^r - P_t)/P_t$. D_{t+1}^f and D_{t+1}^r are dividends paid per share (net of transaction costs) in the two market states.

Table 3: Return Payoffs for the Decision Problem of Converting Stocks to Bonds

		Market at the Beginning of Period $t + 1$	
		Falls	Rises
Convert Stocks to Bonds	Yes	$N_t P_t (1 - \xi_{at})(1 - \xi_{bt})(1 + r_t)$	$N_t P_t (1 - \xi_{at})(1 - \xi_{bt})(1 + r_t)$
	No	$N_t P_t (1 + R_{t+1}^f)$	$N_t P_t (1 + R_{t+1}^r)$

Denoting the forecast probability of a market fall (namely $R_{t+1} < 0$) by $\hat{\pi}_{t+1}$ (formed at the beginning of period t) and using the logarithm of the payoffs in Table 3, the investor will decide to convert stocks into bonds if

$$\hat{\pi}_{t+1} > \frac{\ln \left[(1 + R_{t+1}^r)/(1 + r_t) \right] - \ln \left[(1 - \xi_{at})(1 - \xi_{bt}) \right]}{\ln \left[(1 + R_{t+1}^r)/((1 + R_{t+1}^f)) \right]} = q_{ts}. \quad (20)$$

To utilize this decision rule the investor also needs to predict the extent to which market is likely to rise or fall. We assume that historical averages of positive and negative stock returns are used to predict the values of R_{t+1}^r and R_{t+1}^f , respectively. In the case of most market portfolios these historical averages have been remarkably stable. The sample means of R_t^r and R_t^f computed recursively using monthly returns, R_t , on Standard and Poor 500 over the period 1954(1) to 1992(12) are displayed in Figure 1.¹⁴ In the light of these estimates and for purpose of illustration we set $R_{t+1}^r = 0.035$ and $R_{t+1}^f = -0.030$. For given values of R_{t+1}^r and R_{t+1}^f , the payoff ratio, q_{ts} , is decreasing in r_t and increasing in ξ_{at} and ξ_{bt} . In general, $q_{ts} > 0$, but for sufficiently high transaction costs it could be larger than unity. As in Pesaran and Timmermann (1995) we consider two transaction cost scenarios; a low-cost scenario where $\xi_{at} = .005$ and $\xi_{bt} = 0.001$, and a high-cost scenario with $\xi_{at} = .01$ and $\xi_{bt} = 0.001$. Using monthly observations on the Treasury Bill rate for r_t , the values of the payoff ratio, q_{ts} , show considerable variations over time ranging from 0.51 to 0.68 for the high-cost scenario and 0.43 to 0.59 for the low cost scenario. This clearly shows that in periods of high interest

¹⁴The source of the data used in this Section is Pesaran and Timmermann (1995).

rates even relatively low probability of a stock market fall could initiate a switch from stocks to bonds. But the same need not be true in periods where the interest rate is low.

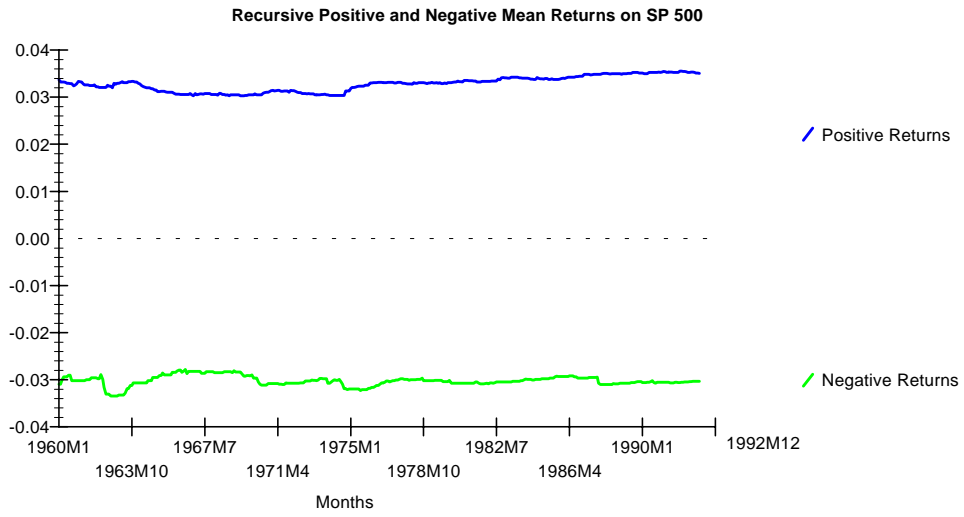


Figure 1

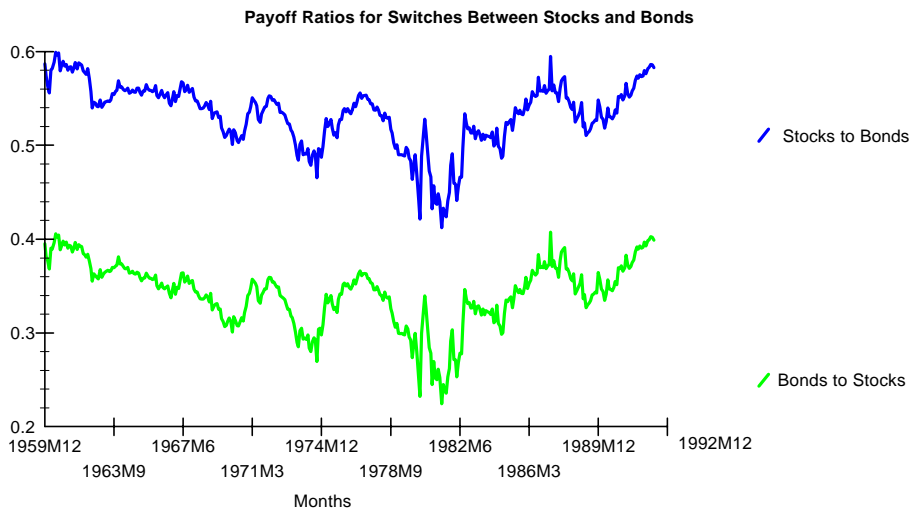


Figure 2

Consider now the problem of an investor who holds government bonds, B_t , and wishes to decide whether to stay in bonds or to switch into stocks. The payoffs for this decision problem are set out in Table 4.

Table 4: Return Payoffs for the Decision Problem of Converting Bonds to Stocks

		Market at the Beginning of Period $t + 1$	
		Falls	Rises
Convert Bonds to Stocks	Yes	$B_t(1 + R_{t+1}^f)(1 - \xi_{at})(1 - \xi_{bt})$	$B_t(1 + R_{t+1}^r)(1 - \xi_{at})(1 - \xi_{bt})$
	No	$B_t(1 + r_t)$	$B_t(1 + r_t)$

Therefore, the investor will decide to switch from bonds into stocks if

$$\hat{\pi}_{t+1} < \frac{\ln [(1 + R_{t+1}^r)/(1 + r_t)] + \ln [(1 - \xi_{at})(1 - \xi_{bt})]}{\ln [(1 + R_{t+1}^r)/((1 + R_{t+1}^f))]} = q_{tb}. \quad (21)$$

The two decision problems are equivalent only under zero transaction costs. But in the presence of transaction costs we have $q_{ts} > q_{tb}$, and the investor is faced with a third possibility, namely to do nothing if

$$q_{tb} \leq \hat{\pi}_{t+1} \leq q_{ts}. \quad (22)$$

The width of this no-transaction band is given by

$$q_{ts} - q_{tb} = \frac{-2 \ln [(1 - \xi_{at})(1 - \xi_{bt})]}{\ln [(1 + R_{t+1}^r)/((1 + R_{t+1}^f))]} > 0.$$

Figure 2 shows the values of q_{ts} and q_{tb} under the low-cost scenario, $\xi_{at} = 0.005$ and $\xi_{bt} = 0.001$, and using the recursive estimates of the historical means of R_{t+1}^r and R_{t+1}^f displayed in Figure 1. According to these estimates on average forecast probabilities, $\hat{\pi}_{t+1}$, should lie outside the range (0.34, 0.53) before it is worthwhile for the investor to make any transactions. This is an example where probability forecasts, even if accurate, may have little economic value to a decision maker.¹⁵

¹⁵The issue of the economic value of predictability of returns in the presence of transaction costs is discussed recently by Luttmer (1997) who also shows that when transaction costs are sufficiently large investors may have little incentives to exploit predictable patterns in stock returns in setting of their portfolio weights.

4.2 Statistical Measures of Performance

To get some idea of the empirical importance of the transaction costs in switching trading strategies we computed recursive forecast probabilities of a fall in the U.S. stock market using monthly observations on the SP 500 market index, PSP , and a number of standard factors used in the literature to forecast stock returns. In particular, we estimated the following return regression *recursively* over the period 1954(1)-1992(12).

$$R_{t+1} = a_{0t} + a_{1t}YSP_t + a_{2t}\Delta I1_t + a_{3t}PI12_{t-1} + a_{4t}\Delta IP12_{t-1} + \varepsilon_{t+1} = \boldsymbol{\beta}'_t \mathbf{x}_t + \varepsilon_{t+1}, \quad (23)$$

where YSP is the dividend yield based on 12-month moving average of dividends paid on SP 500, $\Delta I1$ is the first-difference of the 1-month T-Bill rate, $PI12$ is the inflation rate (computed using the 12-months moving average of producer price index), and $\Delta IP12$ is the rate of change of industrial production (again based on 12 months moving average of the index of industrial production).¹⁶ Denoting the recursive forecasts of R_{t+1} (formed at time t) and their associated standard errors by \hat{R}_{t+1} and $\hat{\sigma}_{t+1}$, respectively, we computed the probability forecasts, $\hat{\pi}_{t+1}$, assuming the conditional distribution of the errors, ε_{t+1} , is approximately normal. Normality is clearly rejected in the case of high frequency (daily or intra-daily) data, but the evidence against normality is less clear cut in the case of quarterly and monthly observations. As can be seen from Figure 3, overall the normal curve provides a reasonably good fit to the recursive residuals; the major exception being the October 1987 crash.

¹⁶For further details and data sources see Pesaran and Timmermann (1994, pp. 363-356).

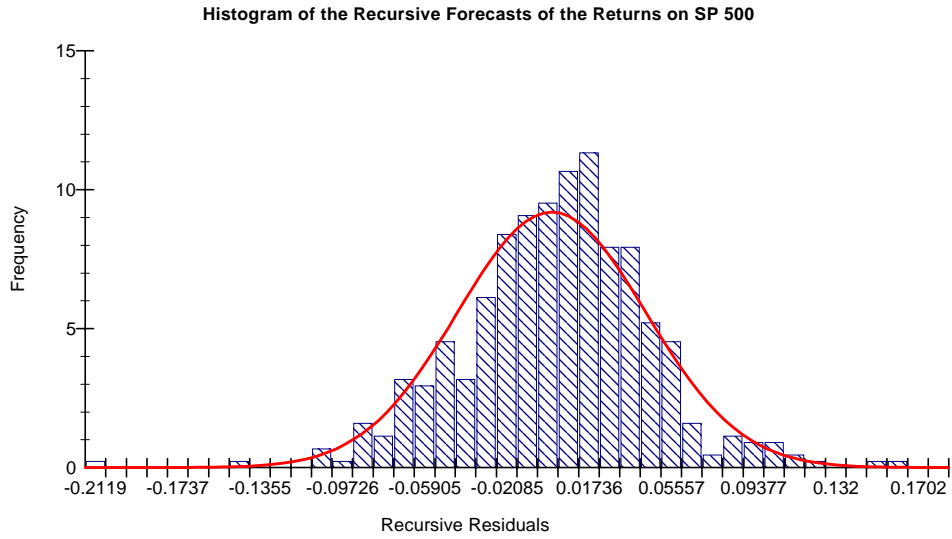


Figure 3

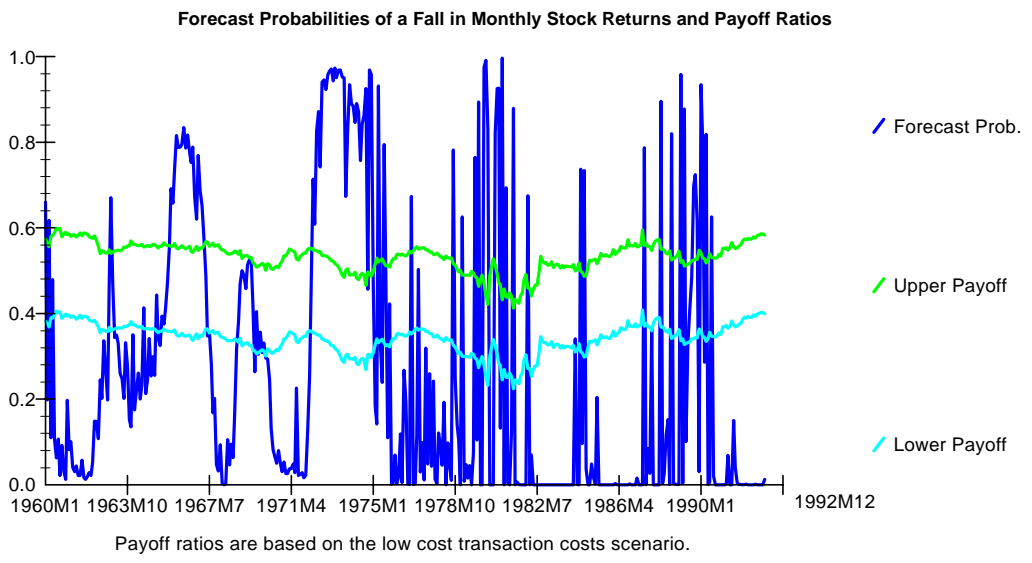


Figure 4

Under the normality assumption the probability forecast of a fall in the stock market in month $t + 1$, based on information available in month t , is given by

$$\widehat{\Pr}(R_{t+1} < 0 \mid \Omega_t) = \hat{\pi}_{t+1} = \Phi(-\hat{R}_{t+1}/\hat{\sigma}_{t+1}). \quad (24)$$

The estimates of $\hat{\pi}_{t+1}$ together with the payoff ratios, q_{ts} and q_{tb} , computed under the low cost scenario are displayed in Figure 4.¹⁷ According to these estimates in 34 months out of the total number of 396 monthly observations (or just over 8.6% of the periods) $\hat{\pi}_{t+1}$ will fall within the no transaction band (q_{ts}, q_{tb}), and on these occasions the investor lacks the necessary economic incentive to trade. The number of such periods rises to 15.7% of the total observations when the payoff ratios are computed for the high transaction cost scenario. The number of cases for which the probability forecasts fall inside and outside the range (q_{ts}, q_{tb}), cross-tabulated according to whether $R_{t+1} < 0$ or not, are summarized in Table 5. Over the sample period the market fell in 40% of the months, while based on probability forecasts and the payoff ratios the investor would have acted decisively to leave the market in only 26.5% of the times in the case of zero transaction costs and even less under the low and the high transaction cost scenarios, namely 21.5% and 20.5%, respectively

Table 5: Contingency Matrix of Realizations and Actions Under Different Transaction Costs Scenarios

Actions	Falls ($R_{t+1} < 0$)			Rises ($R_{t+1} \geq 0$)			Totals		
	Transaction Costs			Transaction Costs			Transaction Costs		
	Zero	Low	High	Zero	Low	High	Zero	Low	High
$\hat{\pi}_{t+1} > q_{ts}$	56	47	46	49	38	35	105	85	81
$q_{tb} \leq \hat{\pi}_{t+1} \leq q_{ts}$	0	15	24	0	19	38	0	34	62
$\hat{\pi}_{t+1} < q_{tb}$	102	96	88	189	181	165	291	277	253
Totals	158			238			396		

Various statistical measures of fit discussed in Section 4 can also be computed using the results in Table 5. It is, however, important to note that in the presence of transaction costs we need to distinguish between cases where

¹⁷In order to avoid the excessive uncertainty associated with the forecasts at the start of the sample period the forecast probabilities for the initial period 1954(1)-1959(12) are dropped only probability forecasts over the subsequent period shown in Figure 4.

$\hat{\pi}_{t+1} > q_{ts}$ and $\hat{\pi}_{t+1} < q_{tb}$. Only under zero transaction costs the 3×2 contingency matrix will be reduced to the standard 2×2 matrix of the previous sections. In this case the hit and false alarm rates are equal to 35.4% and 20.6%, respectively, yielding a Kuipers score of KS=14.8%. In comparison to the Kuipers scores typically obtained in short-term weather forecasting this is small, but statistically significant.¹⁸ The associated value of the market timing test statistic (which is equivalent to a test based on the Kuipers score) is 3.28 which is substantially above the critical value of a one-sided normal test even at the 0.01% level. When transaction costs are taken into account the KS measures and the PT statistics associated with switches out of the market (namely when $\hat{\pi}_{t+1} > q_{ts}$ at time t) are 13.8% and 3.27, respectively. The corresponding statistics for switches from bonds into stocks ($\hat{\pi}_{t+1} < q_{tb}$) are 15.3% and 3.25. Therefore, there is statistically significant evidence of market timing in both directions. Increasing the transaction costs from the low to the high scenario does not qualitatively change this conclusion. See Table 6.

¹⁸For example, the KS measures reported in Richardson (1998, Table 3) for prediction of temperature anomalies for January and February 1998 over Europe lie in the range 0.367-0.468.

Table 6 : Statistical Performance Measures for the Probability Forecasts of Stock Returns			
Transaction Costs	Statistics	Falls($R_{t+1} < 0$)	Rises ($R_{t+1} \geq 0$)
<i>Zero</i>			
	Hits(%)	35.4	79.4
	False(%)	20.6	64.5
	KS(%)	14.9	14.9
	PT	3.28	3.28
<i>Low</i>			
	Hits(%)	29.7	76.1
	False(%)	16.0	60.8
	KS(%)	13.7	15.3
	PT	3.27	3.25
<i>High</i>			
	Hits(%)	29.1	69.3
	False(%)	14.7	55.7
	KS(%)	14.4	13.6
	PT	3.48	2.77

H and F refer to the hit and the false alarm rates, KS is the Kuipers score defined as $H - F$, and PT is the Pesaran-Timmermann market timing test defined by (19).

4.3 Economic Measures of Performance

Statistical measures of performance such as the Kuipers score or its associated market timing test statistic (PT) at best provide a general indicator of the economic value of the probability forecasts, and a more satisfactory approach, whenever possible, would be to directly compute measures of economic performance associated with the use of probability forecasts in decision making. In the context of the present application, an appropriate measure is the logarithm of the investor's terminal wealth obtained using the probability forecasts as compared to being fully invested in the market at all times, (namely a buy-and-hold strategy). Under the active strategy the investor decides to switch from stocks to bonds if $\hat{\pi}_{t+1} > q_{ts}$, from bonds to stocks if $\hat{\pi}_{t+1} < q_{tb}$, and stay put if $q_{tb} \leq \hat{\pi}_{t+1} \leq q_{ts}$. Let h_t take the value of unity if the investor is in stocks at time t and zero otherwise. It is then easily seen

that

$$h_{t+1} = (1 - i_{t+1} - j_{t+1})h_t + j_{t+1}, \quad t = 0, 1, 2, \dots \quad (25)$$

where

$$i_{t+1} = I(\hat{\pi}_{t+1} - q_{ts}) \text{ and } j_{t+1} = I(q_{tb} - \hat{\pi}_{t+1}).$$

Notice that i_{t+1} and j_{t+1} both are known at time t . Also since $q_{tb} < q_{ts}$ they cannot both be *zero* or *unity* at the same time. Suppose that initially the funds available to the investor at time $t = 0$ is given by W_0 which is held in cash, so that $h_0 = 0$. Then the realized return (net of transaction costs) of the switching investment strategy is given by

$$1 + R_{t+1}^s = W_{t+1}/W_t = h_t X_{t+1} + (1 - h_t)Y_{t+1}, \quad t = 0, 1, \dots \quad (26)$$

where

$$X_{t+1} = i_{t+1}(1 - \xi_{at})(1 - \xi_{bt})(1 + r_t) + (1 - i_{t+1})(1 + R_{t+1}), \quad (27)$$

and

$$Y_{t+1} = j_{t+1}(1 - \xi_{at})(1 + R_{t+1}) + (1 - j_{t+1})(1 - \xi_{bt})(1 + r_t). \quad (28)$$

It is assumed that costs are paid at the time of the transaction. To allow for costs incurred when dividends are re-invested in stocks the nominal return on stock holdings is computed as $R_{t+1} = [P_{t+1} - P_t + (1 - \xi_{at})D_{t+1}]/P_t$.¹⁹

Using the recursive forecasts $\hat{\pi}_{t+1}$, and the payoff ratios q_{ts} and q_{tb} we computed R_t^s , $t = 1960(1), \dots, 1992(12)$, under the three transaction costs scenarios: *Zero* ($\xi_{at} = 0, \xi_{bt} = 0$), *Low* ($\xi_{at} = 0.005, \xi_{bt} = 0.001$), and *High* ($\xi_{at} = 0.01, \xi_{bt} = 0.001$). For purposes of comparison we also computed two other sets of recursive probability forecasts and switching portfolio returns; one without the inflation variable, $PI12_{t-1}$, and another without the output growth variable, $\Delta IP12_{t-1}$. We shall refer to these as models M_1 and M_2 , respectively. The base model containing all the four variables will be called M_0 . The logarithm of the terminal wealth, $\ln(W_T)$, generated by these portfolios, their mean returns as well as the mean return of the buy and hold strategy together with a number of other related statistics are summarized in Table 7. These results generally favour model M_0 , and at the same time show the importance of allowing for transaction costs when evaluating the

¹⁹These recursive relations generalize the results in Pesaran and Timmermann (1994, pp. 350-351) derived for the simple switching strategy where $j_{t+1} = 1 - i_{t+1}$ for all t .

economic value of stock market forecasts. The results also show that dropping the inflation and the output growth variables from the basic model is not justifiable from an economic viewpoint, with inflation being relatively more important than the output growth variable over the period under analysis. Similar conclusions are reached if one considers the statistical measures of performance such as the Kuipers score or the market timing test statistic.

Table 7 : Performance Measures for Alternative Forecasting Models of U.S. Stock Returns (1960(1) – 1992(12))					
		Forecasting Models			Market
Transaction		M_0	M_1	M_2	Portfolio
Costs					
<i>Zero</i>	Mean Return (%)	13.43	12.07	12.65	10.91
	S.D. of Return (%)	43.67	47.34	42.71	51.75
	Sharpe Ratio	0.176	0.133	0.162	0.100
	$\ln(W_T)$	8.7541	8.2658	8.5124	7.8254
	Out of Stocks (%)	26.5	12.9	19.4	0.0
<i>Low</i>	Mean Return (%)	12.11	10.28	11.24	10.88
	S.D. of Return (%)	44.52	49.43	42.73	51.77
	Sharpe Ratio	0.143	0.092	0.128	0.099
	$\ln(W_T)$	8.3111	7.6505	8.0516	7.8142
	Out of Stocks (%)	22.7	12.1	18.2	0.0
<i>High</i>	Mean Return (%)	10.91	9.98	10.91	10.85
	S.D. of Return (%)	43.92	48.96	43.09	51.80
	Sharpe Ratio	0.117	0.087	0.120	0.098
	$\ln(W_T)$	7.9258	7.7559	7.9393	7.8030
	Out of Stocks (%)	24.7	12.1	18.2	0.0

*Model M_0 is defined by (23). Models M_1 and M_2 are obtained from M_0 by dropping the inflation and the output growth variables, respectively. “Mean Return” is the arithmetic mean of asset returns over the switching portfolio based on the recursive forecasts. “S.D. of Return” is the standard deviation of the returns. “Sharpe Ratio” is computed as the ratio of the mean of the excess return (return on the portfolio minus return on T-Bill) divided by its standard error. W_T is the value of the terminal wealth, measured at the end of December 1992. “Out of Stocks” is the proportion of periods that the portfolio is not in stocks.

5 The General Case

The approach to forecast evaluation advocated in this paper can be readily extended to multi-period decision problems with a general utility (loss) function. To simplify the exposition we focus on a single period decision problem where at time t the decision variable d_t is chosen so that $E[U(d_t, x_{t+1}) | \Omega_t]$ is maximized, where $U(\cdot)$ is a globally concave utility function, and x_{t+1} is a state variable with the conditional probability distribution function $F_t(x) = \text{Prob}(x_{t+1} < x | \Omega_t)$. As before, Ω_t is the information set available at time t , and includes at least past observations on x , namely x_{n-j} , $j \geq 0$. It is assumed that the choice of F influences d_t , but not *vice versa*. This is equivalent to the assumption (1) made in the case of the simple decision problem discussed in Section 2. At time t a forecaster provides a predictive distribution $\hat{F}_t(x)$ as an estimate of $F_t(x)$. A decision, $\hat{d}_t = \Psi_u(\hat{F}_t)$, is then made at time t , based on $\hat{F}_t(x)$, and at time $t + 1$ when the realization x_{t+1} becomes known, a value is generated for the decision maker that depends on x_{t+1} and on the decision made at time t , namely $U[\Psi_u(\hat{F}_t), x_{t+1}]$. A decision rule, $\Psi_u(F_t)$, is said to be optimal if for all feasible decision values, d_t , it satisfies the following condition

$$E[U(\Psi_u(F_t), x_{t+1}) | \Omega_t] \geq E[U(d_t, x_{t+1}) | \Omega_t].$$

In general, the “optimal” decision rule, $d_t^* = \Psi_u(F_t)$, depends on the whole of the predictive distribution function and knowledge of point forecasts would not be sufficient for a complete solution to the decision problem; unless, of course, the utility function is quadratic in d_t and x_{t+1} .

When a run of decisions and realizations $\{d_t, x_{t+1}; t = 1, 2, \dots, T\}$ are available the “closeness” of actual decisions, d_t , to the optimal decisions, d_t^* , can be measured in terms of the utility-based criterion:

$$S_u(d^* : d) = T^{-1} \sum_{t=1}^T \{U(d_t^*, x_{t+1}) - U(d_t, x_{t+1})\}.$$

On average, we would expect $S_u(d^* : d) \geq 0$. This result follows immediately from the optimality of d_t^* . Notice that

$$E[S_u(d^* : d)] = T^{-1} \sum_{t=1}^T E\{E[U(d_t^*, x_{t+1}) | \Omega_t] - E[U(d_t, x_{t+1}) | \Omega_t]\},$$

and by virtue of the optimality of d_t^* ,

$$E[U(d_t^*, x_{t+1}) | \Omega_t] \geq E[U(d_t, x_{t+1}) | \Omega_t].$$

Hence, it must be that $E[S_u(d^* : d)] \geq 0$. This is an interesting result, although by itself is of little practical use since the derivation of d_t^* requires perfect knowledge of the predictive distribution function, $F_t(x)$, and this is not usually known, either to the decision maker or to the econometrician. However, S_u can be used to discriminate between two sets of decisions based on two different estimates of $F_t(x)$. Denote these estimates by $\hat{F}_t^{(1)}$ and $\hat{F}_t^{(2)}$ and the corresponding optimal decisions by $\hat{d}_t^{(1)} = \Psi_u(\hat{F}_t^{(1)})$ and $\hat{d}_t^{(2)} = \Psi_u(\hat{F}_t^{(2)})$, respectively. The mean loss (in utility terms) associated with the use of these probability forecasts are given by

$$S_u(d^* : d^{(1)}) = T^{-1} \sum_{t=1}^T \left\{ U(d_t^*, x_{t+1}) - U(\Psi_u(\hat{F}_t^{(1)}), x_{t+1}) \right\},$$

and

$$S_u(d^* : d^{(2)}) = T^{-1} \sum_{t=1}^T \left\{ U(d_t^*, x_{t+1}) - U(\Psi_u(\hat{F}_t^{(2)}), x_{t+1}) \right\}.$$

Although $S_u(d^* : d^{(1)})$ and $S_u(d^* : d^{(2)})$ are not separately observable, from the above results it is reasonable to select the estimate which yields the lowest mean utility loss relative to the optimal decision. Namely, select $\hat{F}_t^{(1)}$ over $\hat{F}_t^{(2)}$ if (for sufficiently large T)

$$S_u(d^* : d^{(1)}) < S_u(d^* : d^{(2)}),$$

or equivalently if

$$T^{-1} \sum_{t=1}^T U \left[\Psi_u(\hat{F}_t^{(1)}), x_{t+1} \right] > T^{-1} \sum_{t=1}^T U \left[\Psi_u(\hat{F}_t^{(2)}), x_{t+1} \right].$$

This is a generalization of (14) which no longer depends on d_t^* , and contrasts the normal practice where $\hat{d}_t^{(1)}$ and $\hat{d}_t^{(2)}$ are compared directly using purely statistical measures.²⁰ It is also worth noting that since different decision makers will have different value functions, they need not rank the models (or forecasts) the same. However, if one model produces the true data generating process, it will be preferred by all decision makers over any alternatives, regardless of their value functions.

²⁰For an exception see West, Edison and Cho (1993) who consider a utility-based comparison of a number of volatility models. The use of cost/utility based comparisons of statistical models has also been emphasized by Christofferson and Diebold (1996) and Weiss (1996).

6 Conclusions

In this paper we have argued in favour of a closer link between the decision and the forecast evaluation problems. We have also emphasized the need for producing and reporting predictive distribution functions rather than point, or even interval forecasts. Point forecasts are sufficient only in relatively simple cases where the utility/loss function is quadratic and the constraints (if any) are linear. We have illustrated some of the main issues involved both in the context of a two-state, two-action decision problem as well as in more general settings, investigated the relationships that may exist between statistical and economic methods of forecast evaluation, and provided an empirical application to the problem of stock market predictability and the conditions under which such predictability can be exploited in the presence of transaction costs.

However, we are aware that in general economies, it is unusual to have estimates of the values arising from correct decisions, based on forecasts. An exception is in finance where in some cases estimates of values are possible. An area that has made progress in this field is meteorology. For example, the California Applications Project, organized by the Scripps Institution of Oceanography, San Diego, is determining the value to water managers, in dollars, of a specific improvement in weather forecasts. It seems that eventually these techniques will be usable more generally, until then the idea of presenting a forecast as a predictive distribution, instead of mere summary statistics and measuring the value in terms of the effect on a decision maker are worth keeping in mind when undertaking a forecasting project.

References

- Box, G.E.P. and G.M. Jenkins (1970)** Time Series Analysis, Forecasting and Control. Holder Day, San Francisco.
- Balvers, R.J., Cosimano, T.F. and B. MacDonald (1990)** "Predicting Stock Returns in an Efficient Market". *Journal of Finance*, 45, 1109-28.
- Black, A. and Fraser, P. (1995)** "UK Stock Returns: Predictability and Business Conditions". *The Manchester School Supplement*, 85-102.
- Brandt, M.W. (1998)** "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach". Working paper, The Wharton School, University of Pennsylvania.
- Breen, W., L.R. Glosten, and R. Jagannathan (1990)** "Predictable Variations in Stock Index Returns". *Journal of Finance*, 44, 1177-1189.
- Brier, G.W. (1950)** "Verification of Forecasts Expressed in Terms of Probabilities". *Monthly Weather Review*, 78, 1-3.
- Campbell, J.Y. (1987)** "Stock Returns and the Term Structure". *Journal of Financial Economics*, 373- 99.
- Campbell, J.Y., and Viceira, L.M. (1998)** "Consumption and Portfolio Decisions when Expected Returns are Time Varying". Working paper, Department of Economics, Harvard University.
- Christoffersen, P.F. and F.X. Diebold (1996)** "Further Results on Forecasting and Model Selection under Asymmetric Loss". *Journal of Applied Econometrics*, 11, 561-571.
- Clare, A.D., Thomas, S.H., and M.R. Wickens (1994)** "Is the Gilt-Equity Yield Ratio Useful for Predicting UK Stock Return?". *Economic Journal*, 104, 303-15.
- Clements, M.P. and D.F. Hendry (1998)** *Forecasting Economic Time Series*. Cambridge University Press.
- Clements, M.P. and D.F. Hendry (1999)** *Forecasting Non-Stationary Economic Time Series*. MIT Press, Cambridge, MA.
- Diebold, F.X. and R. Mariano (1995)** "Comparing forecast accuracy". *Journal of Business Economics and Statistics*, 13, 253-65.
- Ehrendorfer, M. and A.H. Murphy (1988)** "Comparative Evaluation of Weather Forecasting Systems Sufficiency, Quality and Accuracy". *Monthly Weather Review*, 116, 1757-1770.
- Fama, E.F., and K.R. French, (1989)** "Business Conditions and Expected Returns on Stocks and Bonds". *Journal of Financial Economics*, 25, 23-49.
- Heidke, P. (1926)** Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geogr. Ann.*, 8, 310-349.
- 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.

- Gandin, L.S., and A.H. Murphy (1992)** "Equitable Skill Scores for Categorical Forecasts". *Mon. Weather Rev.*, 120, 361-370.
- Granger, C.W.J. (1992)** "Forecasting Stock Market Prices: Lessons for Forecasters". *International Journal of Forecasting*, 8, 3-13.
- Granger, C.W.J. and P. Newbold (1977, 2nd ed. 1986)** *Forecasting Economic Time Series*. Academic Press, San Diego.
- Granger, C.W.J. and M.H. Pesaran (1996)** "A Decision Theoretic Approach to Forecast Evaluation". (University of California, San Diego Discussion Paper No.96-23; University of Cambridge DAE Working Paper No.9618).
- Katz, R.W., and A.H. Murphy (1990)** "Quality/Value Relationships for Imperfect Weather Forecasts in a Prototype Multistage Decision-Making Model". *Journal of Forecasting*, 9, 75-86.
- Katz, R.W. and A.H. Murphy (editors) (1997)** *Economic Value of Weather and Climate Forecasts*. Cambridge University Press, Cambridge.
- Leitch, G. and J.E. Tanner (1991)** "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures". *American Economic Review*, 81, 580-90.
- Luttmer, E.G.J. (1997)** "What Level of Fixed Costs Can Reconcile Asset Returns and Consumption Choices". Working paper, London School of Economics.
- Murphy, A.H. (1973)** "A New Vector Partition of the Probability Score". *Journal of Applied Meteorology*, 12, 595-600.
- Murphy, A.H. (1973)** "Hedging and Skill Scores for Probability Forecasts". *Journal of Applied Meteorology*, 12, 215-223.
- Murphy, A.H. (1977)** "The Value of Climatological, Categorical and Probabilistic Forecasts in the Cost-Loss Ratio Situation. *Monthly. Weather Review*". 105, 805-816.
- Murphy, A.H. (1988)** "Skill Scores Based on the Mean Square Error and their Relationships to the Correlation Coefficient". *Monthly. Weather Review*, 116, 2417-2424.
- Murphy, A.H., and H. Dann (1985)** "Forecast Evaluation". In eds., A.H. Murphy and R.W. Katz, *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, Westview, Boulder. 379-437.
- Palmer, T.N., C. Brankovic, D.S., Richardson (1998)** "A Probability and Decision-Model Analysis of PROVOST Seasonal Multi-model Ensemble Integrations", unpublished manuscript, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK.
- Murphy, A.H. and R.L. Winkler (1987)** "A General Framework for Forecast Verification". *Monthly Weather Review*, 115, 1330-1338.
- Peirce, C.S. (1884)** "The Numerical Measure of the Success of Predictions". *Science*, 4, 453-454.
- Pesaran, M.H. and (1992)** "A Simple Nonparametric Test of Predictive Performance". *Journal of Business and Economic Statistics*, 10, 461-65.
- Pesaran, M.H. and A. Timmermann (1994)** "Forecasting Stock Returns. An Ex-

amination of Stock Market Trading in the Presence of Transaction Costs". *Journal of Forecasting*, 13, 330-365.

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

Pesaran, M.H. and A. Timmermann (1999) "Recursive Modelling Approach to Predicting U.K. Stock Returns". *Economic Journal* (forthcoming 2000).

Richardson, D.S. (1998) "Skill and Relative Economic Value of the ECMWF Ensemble Prediction System", Research Department, Technical Memorandum No. 262, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK.

Stewart, T.R. (1997) "Forecast Value: Descriptive Decision Studies" in *Economic Value of Weather and Climate Forecasts*. Katz and Murphy (ed.), Cambridge University Press.

Theil, H. (1960) *Economic Forecasts and Policy*, 2nd edition. North-Holland, Amsterdam.

Weiss, A.A. (1996) "Estimating Time Series Models Using the Relevant Cost Function". *Journal of Applied Econometrics*, 11, 539-560.

White, D.J. (1966) "Forecasts and Decision Making". *Journal of Mathematical Analysis and Applications* 14, 163-73.

West, K.D., Edison, H.J., and D. Cho (1993) "A Utility-Based Comparison of Some Models of Exchange Rate Volatility". *Journal of International Economics*, 35, 23-45.

Wilks, D.S. (1995) *Statistical Methods in the Atmospheric Sciences*. Academic Press.