# ONE-STEP  PREDICTION  OF  FINANCIAL  TIME  SERIES

**by**

**Srichander Ramaswamy**

**July 1998**

# ONE-STEP PREDICTION OF FINANCIAL TIME SERIES

by

**Srichander Ramaswamy**[*]

**July 1998**

## Abstract

This paper examines the one-step prediction of financial time series from a binary decision theory perspective. Under the assumption that the decision statistic of the binary hypothesis testing problem is a Gaussian random variable, bounds for the forecasting efficiency of the hypothesis testing problem are derived. When the criterion for forecasting performance is the total return over the investment period, an optimisation problem is formulated to compute an optimally weighted decision statistic for the binary hypothesis testing problem. Numerical results are illustrated using weekly time series of excess return between two US dollar bond portfolios having six months duration difference. In particular, it is shown that, on average, a 27 basis point excess return per annum is possible against a given benchmark by carrying out active duration management.

# Contents

# Introduction

Many empirical studies in the literature have examined the predictability of financial time series. A measure that is commonly used in these studies to compare the performance of various forecasting models is the mean square forecasting error. Although it appears reasonable to use this as a performance criterion, the end objective of much financial markets forecasting is to maximise the return from any investment over a given time horizon. If this is indeed the case, then the total return over the investment horizon is the appropriate criterion to use to compare the forecasting performance of different models. This paper addresses the forecasting problem from this perspective. In particular, it is assumed that position taking in the market involves being either long or short against a given benchmark and the investment objective is to outperform the benchmark. The forecasting problem posed here reduces to a binary hypothesis testing problem where the question of interest is to know whether a long or a short position against the given benchmark is to be taken. Posing the forecasting problem in the binary decision theory framework is also appealing given the way financial market participants operate: many investment houses provide traders with a benchmark that they are required to outperform including specific guidelines which indicate to what extent traders can deviate from the given benchmark at any given time.

In this paper, we use a US Treasury bond portfolio as an example and assume that the benchmark portfolio has a modified duration of 3.75 years with bond holdings in the one to ten-year maturity sector. It is further assumed here that one can deviate from the benchmark duration by three months on either side. The forecasting problem is posed as a binary hypothesis testing problem which indicates whether a long or short position of three months duration is to be taken once every week against the benchmark. Performance bounds are then derived to compute the achievable forecasting efficiency and excess return over the benchmark. This problem is of interest to many investment houses since it will provide some evidence on whether active fund management can lead to systematic excess return against a given benchmark.

The numerical results from the study carried out suggest that over a four-year period between January 1994 and December 1997 it was possible to achieve on average a 27 basis point excess return per annum over the benchmark when no transaction costs are involved. Transaction costs were not considered here because they tend to be quite small if the strategy is implemented using the current ten-year Treasury note and bond futures contract. It is of interest to note that, based on the weekly position taking, a perfect foresight model would have produced a 150 basis point excess return per annum over the same period. The results based on this study are quite interesting because the 27 basis point excess return is quite significant compared with a static duration extension by three months over this period. Incidentally, a three months static duration extension would have led to only a 7 basis point excess return per annum over the benchmark portfolio. Further, under the

assumption that the decision statistic of the binary hypothesis testing problem is a Gaussian random variable, it is shown that during the four-year investment period considered here, the maximum theoretically achievable forecasting efficiency is close to 60%.

The rest of the paper is organised in the following manner. Section 1 provides background information and a literature survey for the forecasting problem addressed in this paper. Section 2 briefly describes the nearest neighbour approach of Casdagli (1989) and Mizrach (1993) for forecasting, and a probabilistic interpretation of the nearest neighbour forecasting method is given. Section 3 examines the forecasting problem from the binary decision theory perspective. In this section, the general Gaussian problem and determination of performance bounds for the hypothesis testing problem are also presented. The results from Section 3 are used in Section 4 to formulate a suitable optimisation problem to compute the decision statistic of the binary hypothesis testing problem. The forecasting efficiency and total return over the investment horizon of the proposed technique are compared in Section 5 against the equally weighted nearest neighbour approach and a random walk model. The time series used for this performance comparison is the weekly excess return between two bond portfolios having different duration. The final section concludes.


## 1.    Background

Central to any classical approach to time series analysis is the assumption of stationarity of the time series, which allows one to invoke the Wold decomposition theorem. The Wold decomposition theorem states that stationary processes have moving average representations, and hence any stationary time series can be approximated to an arbitrary degree of accuracy by the rational transfer function of an ARMA model. Moreover, any segment of a stationary time series has the same statistical properties as any other segment of the time series. However, when one considers a time series of asset returns, the statistical properties of different sections of the return series may differ during different time periods for a variety of reasons. In general, when the time series has different probability density functions over different time periods, one is confronted with a heterogeneous stochastic system. Finding predictive models for such a time series is a non-trivial exercise. This is because any parametric model fitted over a given sample period of the time series may not be valid over another sample period. For this reason, when one is considering a time series of asset returns, it is not uncommon to find that the out-of-sample forecasting performance of parametric models is worse than a random walk model. In trying to explain this empirical observation, recent work on forecasting financial time series has focused on non-linearity of the data generating process as a possible key to this puzzle (see Mark (1985), Krugman (1988) and Scheinkman and LeBaron

(1989)). However, the study carried out by Flood et al. (1990) indicates that the use of non-linear models do not result in improved out-of-sample forecasting performance over the random walk model.

In general, the problem of identifying a predictive model based on the sequence of iterates from a data generating process is referred to as the inverse problem in dynamical systems (see Casdagli (1989)). One way to construct such a predictive model is to use approximation functions or interpolating polynomials (see Casdagli (1989) and Powell (1981)). A useful variant to this method is to identify nearest neighbours from a Euclidean-norm point of view and use this for forecasting. This procedure is in some sense similar to the locally weighted regression technique of Cleveland and Devlin (1988). Geometrically, the predictions obtained using the nearest neighbour technique are based on finding local portions of the curve in the past that closely resemble the present. This method has much in common with technical analysis, which is widely used for trading purposes. One potential strength of the nearest neighbour method is that it is a non-parametric method and uses only information local to the points to be predicted. This avoids potential problems such as over-fitting and model order misspecification that arise with other parametric methods.

In a study reported by Mizrach (1993), the nearest neighbour approach has been examined in the context of forecasting exchange rates. Mizrach compares the forecasting efficiency (measured in terms of the mean square forecast error) of the nearest neighbour method with that of a random walk model. On the basis of the study, he concludes that the outperformance of the nearest neighbour method is not statistically significant compared with the forecast from a random walk model. A similar conclusion is reached by Parhizgari and De Boyrie (1997) when comparing the out-of-sample performance of various forecasting models against a random walk model on exchange rate data. The main motivation for the present study is the observation that most studies reported in the literature use mean square forecast error as the criterion to compare the forecasting performance of various models. This, in our opinion, is questionable because the objective of an economic agent in using a forecasting model is to maximise profit over the investment horizon of interest. If this is true, then the return over the investment horizon should be the criterion used for performance comparison.

In the study reported here, the annualised return over the investment horizon will be used to compare the performance of competing forecasting models. The model that will be used to generate the one-step forecast is the nearest neighbour approach (see Dockner et al. (1997)). In order to compute the return over the investment horizon generated by the forecasting model, it will be assumed that a long position (of constant size) will be taken against a given benchmark if the forecast indicates that the market will rally. Similarly, a short position (of constant size) will be assumed to be taken against the benchmark if the forecast indicates the market will fall. The annualised excess return over the benchmark generated by such position taking will be used to measure the forecasting performance.

Considering that the model's forecast will be used to take a long or short position against a given benchmark, we will formulate the forecasting problem in the binary hypothesis testing framework.

## 2. Nearest neighbours

Finding predictive models for financial time series that outperform a random walk model out-of-sample is a difficult exercise. This has been attributed to a non-linear data generating process that governs the evolution of such time series and/or the non-stationarity of the time series. In order to model and extract some predictive information from such time series, recent studies have focused on using non-linear models. However, in the context of exchange rate prediction, Diebold and Nason (1990) conclude that non-linearities of exchange rates cannot be exploited to improve out-of-sample forecasts. Considering that, among other reasons, the unsatisfactory out-of-sample performance of forecasting models for financial time series has been attributed to non-linearities and/or non-stationarity, we will consider the nearest neighbour technique (see Casdagli (1989), Mizrach (1993) and Dockner et al. (1997)) for forecasting. The advantage of using the nearest neighbour technique is that it is capable of accounting for both the non-linearities and the non-stationarity of the given time series. The nearest neighbour method consists of embedding the original time series into $m$-vectors where each vector is formed using lagged variables (see Casdagli (1989) and Mizrach (1993)). Each such $m$-vector will represent a point in the $m$-dimensional space and $m$ is referred to as the embedding dimension. The forecasting problem is then reduced to first identifying the nearest neighbours in the $m$-dimensional space to the current $m$-vector whose one step forecast is required, and then computing a weighted average of the last coordinate of the vectors following the nearest neighbours. The details and an example are given below.

Let $\{x(n)\}$ be a discrete time series and let the one-step prediction be required at $n = N$. If $m$ is the embedding dimension, we construct $m$-vectors of the form

$$Z(i) = [x(i+1), x(i+2), \cdots, x(i+m)], \quad i = 0, 1, \cdots, N-m \tag{1}$$

The distance between any two $m$-vectors can be defined in terms of a suitable norm. In the present study, the distance between any two $m$-vectors will be defined by the Euclidean 2-norm $L_2$. If we consider two $m$-vectors $Z(i)$ and $Z(j)$, then the distance $d_{ij}$ between the two vectors is given by

$$d_{ij} = \left[ (x(i+1) - x(j+1))^2 + \cdots + (x(i+m) - x(j+m))^2 \right]^{1/2} \tag{2}$$

The nearest neighbour approach consists of first finding $m$-vectors from the embedding space that are closest to the current $m$-vector $Z(N-m)$ in the $L_2$-norm sense. Let us denote the $M$ nearest

neighbours to the current $m$-vector as $Z(i_k)$, $k = 1,2,\cdots,M$. Then the forecast for $x(N+1)$ is given by the arithmetic average

$$\hat{x}(N+1) = \tfrac{1}{M} \sum_{k=1}^{M} x(i_k + m + 1) \qquad (3)$$

The forecast error is given by

$$e_{N+1} = (\hat{x}(N+1) - x(N+1))^2 \qquad (4)$$

If a total of $K$ forecasts are made, then we define the mean square forecast error as

$$\mathrm{MSE} = \tfrac{1}{\mathrm{VAR}(x_n)} \sum_{i=1}^{K} (\hat{x}(N+i) - x(N+i))^2 \qquad (5)$$

where $\mathrm{VAR}(x_n)$ is the variance of the original time series. Casdagli (1989) suggests that one way to select the embedding dimension is to pick a suitable value of $m$ that minimises the mean square error given by equation (5). Further, it is suggested that a suitable choice for the number of nearest neighbours is $M > 2(m+1)$, which in turn will tend to reduce the forecast error arising from noise.

The following example illustrates the nearest neighbour methodology. Consider the time series given by {2.0, 2.3, 2.1, 2.2, 2.6, 1.9, 2.3, θ}. It is assumed that the embedding dimension is $m = 3$ and the forecast for the last element θ of the time series is required. Table 1 shows the phase space embedding of the time series. The first three columns in this table indicate the three-dimensional embedding vectors and the fourth column indicates the next co-ordinate of the three-dimensional vector in the given time series. The last column gives the $L_2$ norm $d_{ij}$ between the given three-dimensional vector and the current three-dimensional vector (shown in the last row of the table) for which the one-step forecast is required. Examining Table 1, it can be inferred that under the assumption that $M = 2$, the best estimate for θ is 2.25.

Table 1

**Phase space embedding**

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $d_{ij}$ |
|-------|-------|-------|-------|----------|
| 2.0 | 2.3 | 2.1 | 2.2 | 0.7483 |
| 2.3 | 2.1 | 2.2 | 2.6 | 0.3742 |
| 2.1 | 2.2 | 2.6 | 1.9 | 0.5916 |
| 2.2 | 2.6 | 1.9 | 2.3 | 0.9000 |
| 2.6 | 1.9 | 2.3 | θ | |

5

## 2.1 Probabilistic interpretation

The discussion above shows that the nearest neighbour method selects local $m$-segments from the past that are close to the current $m$-segment of the time series in order to generate the one-step ahead forecast. Since such a method avoids fitting global parameters over the entire time series, parameter instability issues that arise when using standard regression models can be avoided, particularly when the time series under consideration is non-stationary. In this section we will show that the estimate $\hat{x}(N+1)$ of the random variable $x(N+1)$ given by equation (3) is the maximum a posteriori (MAP) estimate under the following assumptions.

Let $Z(i_k)$, $k = 1, 2, \cdots, M$ denote the nearest neighbours in the $m$-dimensional embedding space to the current $m$-vector $Z(N-m)$. The nearest neighbour forecast is based on finding the arithmetic average of the past observations $x(i_k + m + 1)$, $k = 1, 2, \cdots, M$ of the time series. Denoting $r_k = x(i_k + m + 1)$, $k = 1, 2, \cdots, M$ we form the observation vector $\vec{r} = [r_1, r_2, \cdots, r_M]$. Let us assume that the unconditional distribution of the random variable $\theta = x(N+1)$ is Gaussian with $N(0, \sigma_\theta)$. Our objective is to estimate on the basis of the data set $\vec{r}$, the value of the random variable $\theta$, and this requires knowledge of the conditional probability density function $p(\theta \mid \vec{r})$. If this function is known, then one can compute the MAP estimate of $\theta$ by solving the following equation:

$$\left. \frac{\partial \ln p(\theta \mid \vec{r})}{\partial \theta} \right|_{\theta = \hat{\theta}(\vec{r})} = 0 \tag{7}$$

In general, it is easier to work with the conditional density function $p(\vec{r} \mid \theta)$, and hence express the conditional density $p(\theta \mid \vec{r})$ as

$$p(\theta \mid \vec{r}) = \frac{p(\vec{r} \mid \theta) p(\theta)}{p(\vec{r})} \tag{8}$$

Let us assume that observations satisfy the equation

$$r_k = \theta + n_k, \quad k = 1, 2, \cdots, M \tag{9}$$

where the additive noise $n_k$ are independent and normally distributed random variables with $N(0, \sigma_n)$. Since the random variable is $N(0, \sigma_\theta)$ and $n_k$ are independent and normally distributed, the conditional density function $p(\vec{r} \mid \theta)$ is given by

$$p(\vec{r} \mid \theta) = \prod_{k=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left( -\frac{(r_k - \theta)^2}{2\sigma_n^2} \right) \tag{10}$$

Following the arguments in Van Trees (1968) it can be shown that the MAP estimate of the parameter $\theta$ is given by (see Appendix A.1)

$$\hat{\theta}_{MAP} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_n^2/M}\left(\frac{1}{M}\sum_{k=1}^{M} r_k\right) \qquad (11)$$

Under the assumption that $\sigma_\theta^2 >> \sigma_n^2/M$, it follows that the MAP estimate reduces to the arithmetic average of the $M$ nearest neighbours given by

$$\hat{\theta}_{MAP} = \frac{1}{M}\sum_{k=1}^{M} r_k \qquad (12)$$

Comparing equations (3) and (12), one can interpret the conditional forecast made using the nearest neighbour approach as the MAP estimate. This is true under the assumption that the random variable $\theta = x(N+1)$ being estimated is normally distributed and that the a priori knowledge of the mean of $\theta$ is of limited value. Note that if we are using a time series of asset returns, then these assumptions are quite reasonable. It is useful to note here that under the Gaussian assumptions for $\theta$ and the noise, the conditional density $p(\theta|\bar{r})$ is also Gaussian. It follows from this that the estimate given by equation (11) is also the minimum mean square estimate of the random variable $\theta = x(N+1)$.

## 3.    Binary hypothesis testing

The motivation for forecasting financial time series is to explore trading strategies that are superior to a buy-and-hold strategy and hence provide an opportunity for making excess profit. The analysis presented by Mizrach (1993) suggests that the nearest neighbour approach does not outperform a random walk model in forecasting exchange rates. However, the criterion used for making such a comparison is based on the mean square forecast error. This, in our opinion, may not be the most appropriate criterion to use. The reason for this is that the ultimate objective of any forecasting model is to enable a trading position to be taken against a given benchmark (which could be cash, for instance) in order to generate an excess return. If we assume that the size of the trading position is constant (denoted here as long or short against the benchmark), then one way to implement the trading strategy is to use the sign of the forecast to take the position. The performance of the competing forecasting models can then be compared on the basis of the total excess return generated over the investment period. Such a criterion will have economic significance. We will show here by considering a simple example that the mean square forecast error can be a misleading indicator of forecasting performance to a profit-maximising economic agent.

Let us assume that the weekly return in basis points generated by a trading position is given by the difference equation

$$r_{t+1} = 1 - 1.4\,r_t^2 + 0.3\,r_{t-1}, \quad t = 1, 2, \cdots, 104 \tag{13}$$

with initial conditions $r_1 = -0.10$ and $r_0 = 0.12$. Let the following forecasting models be used for position taking:

$$y_{t+1} = 0.3978 - 0.8371\,y_t + 0.5763\,(r_t - y_t), \quad t = 1, 2, \cdots, 104 \tag{14}$$

$$z_{t+1} = 0.52 - 0.8\,z_t + 0.4\,(r_t - z_t), \quad t = 1, 2, \cdots, 104 \tag{15}$$

The initial conditions are taken to be $y_1 = z_1 = 0$ in the above models. We will refer to equation (14) as model A and equation (15) as model B. The trading strategy is to take a long position if the model's forecast is greater than or equal to zero and a short position if the model output is negative. The total return generated over the investment period is given by

$$R_{total} = \prod_{t=1}^{104} (1 + I_t\, r_{t+1}) \;-\; 1 \tag{16}$$

where $I_t$ is an indicator variable taking the value $+1$ or $-1$ depending upon whether the forecast model's output is positive or negative, respectively. The mean square forecast error for model A is defined as

$$\text{MSE} = \sum_{t=1}^{104} (r_{t+1} - y_{t+1})^2 \tag{17}$$

The mean square prediction error for model B is similarly defined. Table 2 shows the total return over the investment period and the mean square prediction error for both models. It can be seen from this table that although model A has a lower mean square prediction error, it has a lower total return than model B over the investment period.

Table 2

**Forecasting performance comparison**

| Model | MSE | $R_{total}$ (bps) |
|:-----:|:-----:|:-----:|
| A | 47.80 | 48.90 |
| B | 50.12 | 57.25 |

The simple example considered above illustrates that the mean square forecast error may not be the appropriate criterion to use when comparing the forecasting performance of different models. Considering that the output from any forecasting model has to be ultimately transformed into a trading position that can be interpreted as being long or short against a given benchmark, the forecasting problem of interest is a binary hypothesis testing problem. In general, the classical binary

hypothesis testing problem involves observing a set of random variables, and on the basis of this one has to decide which of the alternate hypotheses is true. With reference to the forecasting problem of interest here, the competing hypotheses are a short and a long position against a given benchmark. The short position will be denoted as hypothesis $H_0$ and the long position as hypothesis $H_1$. The basic components of a binary decision theory problem are shown in Figure 1. The first is a source that generates an output. The second is a probabilistic transition mechanism that generates a point in the observation space according to some probability law. Finally, on the basis of the outcome in the observation space, a decision rule assigns each point to one of the two hypotheses.

Figure 1

**Components of a decision theory problem**
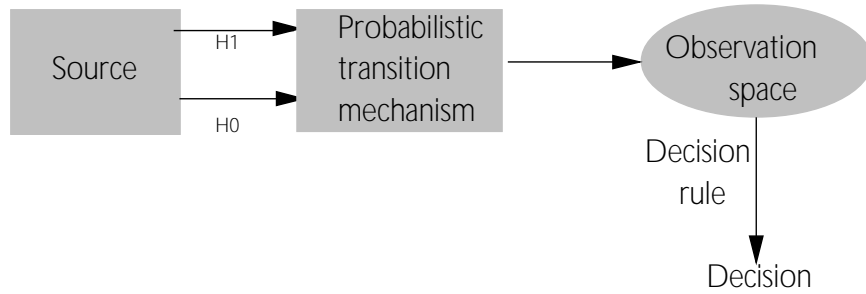


Let us assume that $\vec{r} = [r_1, r_2, \cdots, r_M]$ denotes the set of *M* observations from a source that generates random variables belonging to two known conditional densities $p(\vec{r}|H_0)$ and $p(\vec{r}|H_1)$ depending upon which hypothesis is true. It can be shown that the application of the Bayes or Neyman-Pearson criterion leads to an optimal test that involves processing the *M* observations to find the likelihood ratio and then comparing this value against a threshold to make the decision (see Van Trees (1968)). In general, it is easier to work with the logarithm of the likelihood ratio and this is given by

$$\ell(\vec{r}) = \log\left(\frac{p(\vec{r}|H_1)}{p(\vec{r}|H_0)}\right) > \gamma \text{ then } H_1, \text{otherwise } H_0 \qquad (18)$$

### 3.1 General Gaussian problem

The above discussion shows that the binary hypothesis testing problem leads to a likelihood ratio test involving the conditional density functions of the observations under the two hypotheses. In this section it is shown that if the observation vector is a Gaussian density under both hypotheses, then some useful performance measures can be derived for the test statistic. Such an assumption will allow us to find bounds for achievable forecasting efficiency for the observed time series and to determine the optimal embedding dimension *m*. Let us again assume that the observed

$M$-vector of nearest neighbours (selected using any appropriate distance measure) is given by $\vec{r} = [r_1, r_2, \cdots, r_M]$. The conditional density of the random vector $\vec{r}$ is assumed to be Gaussian with mean vectors $\vec{\mu}_0$ and $\vec{\mu}_1$ under the hypotheses $H_0$ and $H_1$. The covariance matrix under both hypotheses is assumed to be equal and denoted $\Sigma$. The likelihood ratio test for deciding the true hypothesis involves computing the quantity $\ell(\vec{r})$ given by equation (18). Denoting $\Sigma^{-1} = Q$, it is easy to show that the likelihood test reduces to (see Appendix A.2)

$$\left(\vec{\mu}_1 - \vec{\mu}_0\right)^T Q \, \vec{r} \, \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \gamma \tag{19}$$

where $\gamma$ is a suitably chosen threshold and all vectors under consideration are assumed to be column vectors. For the sake of convenience, we will denote $\Delta\vec{\mu} = \left(\vec{\mu}_1 - \vec{\mu}_0\right)$, so that the likelihood ratio test is given by

$$\ell(\vec{r}) = \Delta\vec{\mu}^T Q \, \vec{r} \, \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \gamma \tag{20}$$

We note that the function $\ell(\vec{r})$ is a linear transformation of Gaussian random variables, and hence is also a Gaussian random variable. This will allow us to characterise the performance of the test using the distance $d$ between the means under the two hypotheses when the variance is normalised to unity. This distance $d$ can be computed from

$$d^2 = \frac{\left[E(\ell \mid H_1) - E(\ell \mid H_0)\right]^2}{\mathrm{Var}(\ell \mid H_0)} \tag{21}$$

It is easy to show by evaluating the conditional expectations that $d^2$ is given by (see Appendix A.3)

$$d^2 = \Delta\vec{\mu}^T Q \, \Delta\vec{\mu} \tag{22}$$

Further simplifications in performance computation are possible if we assume that the Gaussian random variables that constitute the vector $\vec{r}$ are independent so that the covariance matrix is diagonal. In this case it is easy to show that the likelihood ratio test reduces to

$$\ell(\vec{r}) = \sum_{k=1}^{M} \frac{\Delta\mu_k \, r_k}{\sigma_k^2} \, \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \gamma \tag{23}$$

where $\sigma_k^2$ is the variance of the random variable $r_k$. Likewise, it can be shown that under these assumptions $d^2$ is given by

$$d^2 = \sum_{k=1}^{M} \frac{(\Delta\mu_k)^2}{\sigma_k^2} \tag{24}$$

In the following section, we will show how the distance measure given by equation (24) can be used to establish bounds for the test performance. The analysis presented will also enable us to compute the optimal embedding dimension $m$ for selecting the nearest neighbours.

### 3.2    Performance bounds

In order to characterise the performance of the likelihood ratio test, we need to compute the conditional probability of accepting the hypothesis $H_1$ when the true hypothesis is $H_0$ and the conditional probability of accepting the hypothesis $H_1$ when the true hypothesis is $H_1$. In the first case, the conditional probability is referred to as the false alarm probability $P_F$ and in the second case as the detection probability $P_D$. The objective of any hypothesis testing problem will be to maximise the detection probability $P_D$ and minimise the false alarm probability $P_F$. The ability to achieve this trade-off will in general depend on the distance $d$ between means under the two hypotheses with the test performance improving monotonically as $d$ increases. If $\gamma$ is the test threshold to distinguish between the two hypotheses, then it can be shown that $P_F$ and $P_D$ are given by

$$P_F = \int_{\gamma}^{\infty} p\big(\ell(\vec{r})|H_0\big)d\ell \tag{25}$$

$$P_D = \int_{\gamma}^{\infty} p\big(\ell(\vec{r})|H_1\big)d\ell \tag{26}$$

Using the above equations, it is possible to characterise the trade-off between $P_D$ and $P_F$ as a function of the distance $d$ when the threshold $\gamma$ is increased. This is shown in Figure 2.

Considering that the end objective of the forecasting problem of interest here is to maximise the return over the investment period, a reasonable criterion to use for selecting the test threshold for the hypothesis testing problem is to minimise the total probability of decision error. If one is considering a daily or weekly time series of asset returns, then it is reasonable to assume that the a priori probabilities of a negative return (which corresponds to the hypothesis $H_0$) and a positive return (which corresponds to the hypothesis $H_1$) are equally likely over one time step. Using the Bayes criterion (see Appendix A.4) it can be shown that the total probability of error is given by

$$\Re = 0.5P_F + 0.5(1 - P_D) \tag{27}$$

Figure 2

**Test performance as a function of γ and *d***
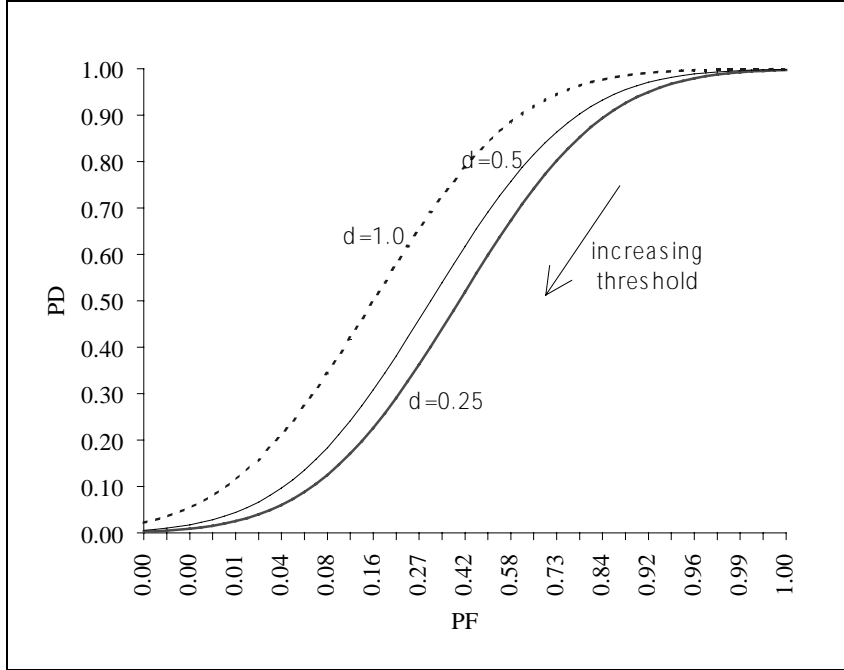


Considering that both $P_F$ and $P_D$ depend explicitly on the test threshold γ and implicitly on the embedding dimension *m* through the test statistic $\ell(\vec{r})$, we will denote the total probability of error in equation (27) as $\Re(\gamma, m)$. For any given embedding dimension *m*, it is possible to compute an optimal test threshold $\gamma^*$ that minimises the total probability given by equation (27), which we denote $\Re_{\min}(\gamma^*, m)$. Further, since the total probability of error is also dependent on the embedding dimension *m*, we will denote the embedding dimension $m = m^*$ that minimises the total probability of error as the optimal embedding dimension and the corresponding total probability of error as $\Re_{\min}(\gamma^*, m^*)$. For the choice of embedding dimension as $m = m^*$, the maximum achievable forecasting efficiency $\eta^*$ for the hypothesis testing problem is given by

$$\eta^* = 100\left[1 - \Re_{\min}(\gamma^*, m^*)\right] \tag{28}$$

Equation (28) provides a statistical upper bound on the achievable forecasting efficiency for the hypothesis testing problem. The choice of the optimal embedding dimension $m^*$ will be examined in Section 5 for the time series of excess return between two bond portfolios with different duration.

12

## 4.    Optimal test statistic

In Section 3, the binary hypothesis testing problem was examined and it was shown that, under the assumption that the test statistic is a Gaussian random variable, performance bounds can be derived in terms of the distance between the conditional means under the two hypotheses. The assumption that the test statistic is a Gaussian random variable is not as restrictive as it may first appear. In particular, if we consider the time series of asset returns, the assumption that the returns are normally distributed is quite reasonable to a first approximation. Moreover, since the observation vector $\vec{r}$ in this case is comprised of asset returns from different periods of the data sample, it is again reasonable to assume that the observations are uncorrelated. These two assumptions together will ensure that the random vector $\vec{r}$ is jointly Gaussian with diagonal covariance matrix, and hence the analysis given in Section 3 can be applied. Under the Gaussian assumption, the test statistic given by equation (23) can be rewritten as

$$\ell(\vec{r}) = \sum_{k=1}^{M} w_k \, r_k \underset{H_0}{\overset{H_1}{\underset{<}{\overset{>}{\gtrless}}}} \gamma \tag{29}$$

where

$$w_k = \frac{\Delta\mu_k}{\sigma_k^2}, \quad k = 1,2,\cdots,M \tag{30}$$

Using historical data it is easy to compute the mean and variance of the random variables that constitute the vector $\vec{r}$ of nearest neighbours, and hence the weights $w_k$ in equation (30). We will denote by $b_{max}$ the maximum absolute value of $w_k$ in equation (30), which can be computed using the historical data.

The choice of the weights $w_k$ in equation (29) plays an important role in the forecasting performance of the hypothesis testing problem. Although the choice of these weights using equation (30) is one possible alternative, from a forecasting performance point of view this may not be necessarily optimal. In the context of forecasting financial time series, the performance criterion of interest is usually the total return over the given investment horizon. Let us consider a weekly trading strategy where a long or a short position against the benchmark is taken based on the output from the binary hypothesis testing problem, and let $\{x(n)\}$ denote the time series of weekly returns for the asset under consideration. If we consider a $P$-period investment horizon starting from time $n = N$, then the total return $R$ arising from the trading strategy is given by

$$R = \prod_{i=1}^{P} \left(1 + I_i \, x(N+i)\right) - 1 \tag{31}$$

where $x(N+i)$ denotes the one period realised return at time $n = N + i$. In equation (31), $I_i$ is an indicator variable that takes the value +1 or -1 depending on whether the test statistic given by equation (29) indicates that the true hypothesis is $H_1$ or $H_0$, respectively. Note that the indicator variable is a function of the weights $w_k$. In this study, we will use the total return $R$ or, equivalently, the average annualised return over the investment horizon as the criterion to compare the performance of various forecasting models. In the numerical study, the investment horizon of interest was taken to be four years so that with weekly data we have $P = 208$.

In order to maximise the return over the investment period, the optimal weights for computing the decision statistic $\ell(\vec{r})$ can be determined by carrying out an in-sample optimisation of an objective function as in equation (31). In such an in-sample optimisation, we use the observed time series up to $n = N$ to compute the weights $w_k$, and use these weights to evaluate the out-of-sample forecasting performance from time step $n = N + 1$ onwards. However, a better investment performance criterion to use for computing the weights would be the one that maximises the worst case investment return during any sub-period of the in-sample investment horizon. Such a criterion will also lead to more robust out-of-sample performance and will be the one used in the present study. If we denote the length of the sub-period as $L$, then the weights $w_k$ for computing the test statistic in equation (29) in order to ensure a robust out-of-sample performance can be determined by solving the following optimisation problem:

$$\underset{w}{\text{Max}} \left[ \underset{j \in [0,1,\cdots,K-L]}{\text{Min}} \prod_{i=1}^{L} \left(1 + I_i x(N-i-j)\right) \right] \tag{32}$$

subject to constraints

$$-b_{\max} \le w_i \le b_{\max}, \quad i = 1, 2, \cdots, M \tag{33}$$

In the objective function given above, $K$ denotes the historical $K$-period investment horizon over which the in-sample optimisation is carried out. In the numerical example presented in the next section, it was assumed that the historical period over which the in-sample optimisation is carried out is two years so that $K = 104$ for a time series of weekly returns. The optimisation problem formulated above is non-convex in the design variables $w_k$. Hence, the search for a global minimum can be carried out using the modified simulated annealing algorithm (see Srichander (1995)). The performance of different forecasting models with respect to the annualised return over the investment horizon will be examined in the next section.

14

## 5.    Numerical results

This section examines the out-of-sample performance of a program trading strategy that uses an optimally weighted test statistic to decide whether a long or a short position is to be taken in the market over the period of the forecast. The performance of the optimal program trading strategy is compared against both a random walk model and the equally weighted nearest neighbour approach of Mizrach (1993) where the $L_2$ norm is used to select the nearest neighbours. The number of nearest neighbours used in this study is $M = 10$. The performance comparison is based on the average annualised excess return generated over a given investment horizon by taking either a long or a short position against a given benchmark. Such a position is taken on the basis of the directional return forecast made by the particular trading model under consideration. The time series data used for performance comparison of the forecasting models is the weekly excess return between two US dollar portfolios, one with 4.0 year modified duration and another with 3.5 year modified duration. The weekly excess return time series was computed using the J.P. Morgan total bond return index in the one to ten-year sector. The motivation for considering this time series was to explore whether it is possible to achieve excess returns against a given benchmark by pursuing an active portfolio duration management strategy. Table 3 shows the average annualised return over the period January 1994 to December 1997 for holding a US dollar portfolio with fixed duration of 3.5, 3.75 and 4.0 years.

Table 3

**Average annual return (January 1994 - December 1997)**

| **Duration** | 3.50 years | 3.75 years | 4.0 years |
|---|---|---|---|
| **Annual return** | 6.26 % | 6.33 % | 6.40 % |

It can be inferred from Table 3 that during the period January 1994 to December 1997 there was on average a 7 basis point excess return for a static duration extension of every 0.25 years in the US Treasury market. If we consider the benchmark duration to be 3.75 years, then any active duration management strategy with a permissible duration in the range 3.5 to 4.0 years that produces more than a 7 basis point excess return over the benchmark during the four-year investment period will be deemed to be efficient. The performance of the various forecasting models which generate trading rules once every week to take a duration position of either 3.5 or 4.0 years will be compared in this section. In order to provide enough historical data while searching for nearest neighbours, weekly returns for each day of the week starting from January 1986 has been used in the present study. Before proceeding with the out-of-sample performance analysis, we will illustrate the choice of the optimal embedding dimension $m^*$ for selecting the nearest neighbours and establish performance bounds for the achievable forecasting efficiency.

15

### 5.1 Optimal embedding dimension

In this section, the method used to select the optimal embedding dimension will be presented. Under the assumption that the decision statistic of the hypothesis testing problem is a Gaussian random variable, the analysis in Section 3.1 showed that the distance between the conditional means under the two hypotheses can be used to characterise the performance of the test. This distance measure was used in Section 3.2 to find performance bounds for the hypothesis testing problem. Since the conditional distribution of the test statistic $\ell(\vec{r})$ under the hypotheses $H_0$ and $H_1$ is a function of the nearest neighbours, the choice of the nearest neighbours will influence the error probability given by equation (27). In order to minimise this error probability, we will select the embedding dimension $m$ such that for a given number of nearest neighbours $M = 10$, the distance $d$ between the means under the two hypotheses as computed from equation (24) is maximised. The embedding dimension $m^*$ that maximises the distance $d$ will be referred to as the optimal embedding dimension. For a given $m^*$ it is possible to compute the minimum error probability $\Re_{\min}(\gamma^*, m^*)$, and hence the maximum achievable forecasting efficiency $\eta^*$ from equation (28). Table 4 shows the distance $d$ and the forecasting efficiency $\eta^*$ for various choices of the embedding dimension $m$ computed using the historical data.

Table 4

**Performance bounds using historical data**

| Time period | Embedding dimension $m$ | Distance $d$ | Optimal threshold $\gamma^*$ | Forecasting efficiency $\eta^*$ |
|---|---|---|---|---|
| January 1992 - December 1995 | 2 | 0.413 | 0.22 | 58.18 % |
| | 3 | 0.466 | 0.25 | 59.21 % |
| | 4 | 0.309 | 0.15 | 56.14 % |
| | 5 | 0.378 | 0.20 | 57.49 % |
| | 6 | 0.316 | 0.18 | 56.28 % |
| January 1994 - December 1997 | 2 | 0.403 | 0.22 | 57.98 % |
| | 3 | 0.464 | 0.25 | 59.17 % |
| | 4 | 0.468 | 0.25 | 59.25 % |
| | 5 | 0.340 | 0.20 | 56.75 % |
| | 6 | 0.314 | 0.18 | 56.24 % |

Examining Table 4 it can be seen that over the two different time periods considered in the study, the maximum forecasting efficiency achievable is close to 60%. This maximum forecasting

efficiency occurs for the embedding dimension $m^* = 3$, and hence we will choose this as the optimal embedding dimension.

### 5.2    Out-of-sample performance

In this section the out-of-sample performance of the various forecasting models is compared. In each case, weekly forecasts were made every Wednesday and, depending on the output of the forecasting model, a duration position of either 3.5 or 4.0 years was assumed to be taken for a one-week period. The annualised excess return generated from such a trading strategy over the benchmark portfolio duration of 3.75 years during the four-year investment period January 1994 to December 1997 is used to compare the performance of competing models. It has been assumed here that a long position in the market is to be taken when the one-period forecast is positive for the random walk model. In the case of the trading model based on nearest neighbours, the decision statistic $\ell(\vec{r})$ was compared against the optimal threshold $\gamma^* = 0.25$ for deciding whether to take the long or short position. The parameters for the random walk model were selected on the basis of the mean and variance of the excess return time series over the period January 1986 to December 1993. The weights for the optimal decision statistic $\ell(\vec{r})$ of the binary hypothesis testing problem were selected on the basis of the optimisation formulation given in Section 4. In particular, it was assumed that the length of the sub-period is six months, and the historical time period used for the in-sample optimisation is two years. This implies that $L = 26$ and $K = 104$ in the objective function given by (32). The optimal weights of the decision statistic were recomputed once every six months while carrying out the out-of-sample performance analysis. The out-of-sample performance for the various forecasting models during the investment period January 1994 to December 1997 is shown in Table 5. The performance figures for the random walk model are the average values after running the random walk forecast with 50 different starting seed values to generate the random numbers.

Table 5

**Forecasting performance of different models**

| Model | Embedding dimension | Forecasting efficiency | Annual excess return (bps) |
|---|---|---|---|
| Optimised | 3 | 58.0% | 27.0 |
| Equal weights | 3 | 58.0% | 17.0 |
| Equal weights | 4 | 57.5% | 16.5 |
| Equal weights | 5 | 58.0% | 16.0 |
| Random walk | - | 49.8% | 1.5 |

It can be seen from Table 5 that the optimised nearest neighbour approach suggested in this paper outperforms the other models in terms of the annualised excess return over the investment horizon. Although in terms of forecasting efficiency the model with optimal weights is only marginally better than the model with equal weights, this model performs much better than the other models considered here in terms of the return over the investment period. In fact, the use of the optimal decision statistic to decide whether to take a long or short position against the benchmark leads to an outperformance of 27 basis points on an annualised basis. In other words, the dynamic portfolio duration management using the model suggested in this paper leads to a 20 basis point excess return over a static duration extension of 0.25 years. Clearly, such an excess return is quite significant over the investment horizon considered in the study and could be particularly interesting for active portfolio duration management against a given benchmark.

The model with equal weights also performs reasonably well over the investment period since it has an excess return of 10 basis points over a static duration extension by 0.25 years. It is of interest to note here that the model with equal weights with $m = 3$ is the one that performs best. This again suggests that the optimal embedding dimension for this time series is $m^* = 3$. The performance of the random walk model, on the other hand, is the worst in terms of the excess return over the investment horizon. In fact, the performance of the random walk model is very similar to that of a portfolio with a static duration of 3.75 years. The random walk model has a forecasting efficiency of 50%, indicating that it has little forecasting ability.


## Conclusions

In this paper we examined the predictability of financial time series using the nearest neighbour approach. Posing the problem of one-step prediction as a binary hypothesis testing problem, it was shown how the embedding order $m$ for the given time series can be determined. It was suggested that a suitable performance criterion that has economic significance is the average annualised return over the investment horizon. Such a criterion was used to compare the performance of competing forecasting models. Although there has been evidence in the literature for the support of a random walk model for forecasting financial time series, the criterion used for performance comparison can play an important role. Since the end objective for forecasting any financial time series is to maximise total investment return over a given time horizon, the use of such a performance criterion is easy to motivate. Using the excess return time series between two US dollar bond portfolios, it was shown that a trading model that uses the directional forecast from a nearest neighbour forecasting model outperforms the random walk model.

Some of the findings in this paper are significant. In particular, the demonstration that bond returns tend to follow past return patterns is significant as it provides theoretical justification for the use of technical analysis in trading over the short term. Moreover, under the assumption that the decision statistic of the hypothesis testing problem is Gaussian, it was shown that the use of past return patterns for forecasting leads to a maximum predictive efficiency of 60% for the US dollar bond returns over the investment period studied. In general, the out-of-sample forecasting performance, which is measured in terms of average annualised return over the investment horizon, depends on the ability to predict big market moves. The present study shows that models with similar forecasting efficiency (measured in terms of the percentage of correct predictions) can have significantly different annualised return over the investment horizon. Using the total return over the investment horizon as the objective function, the study showed how an optimal decision statistic of the hypothesis testing problem can be computed.

# Appendix

## A.1 Nearest neighbour estimate

It will be shown here that the equally weighted nearest neighbour estimate can be interpreted as the MAP estimate under the following conditions. Let the set of nearest neighbours $r_k = x(i_k + m + 1)$, $k = 1, 2, \cdots, M$ satisfy the following observation equation:

$$r_k = \theta + n_k, \quad k = 1, 2, \cdots, M \tag{A1}$$

In the above equation, the random variable $\theta = x(N + 1)$ is assumed to be Gaussian $N(0, \sigma_\theta)$ and $n_k$ are each independent Gaussian random variables given by $N(0, \sigma_n)$. We then have

$$p(\vec{r} \mid \theta) = \prod_{k=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(r_k - \theta)^2}{2\sigma_n^2}\right) \tag{A2}$$

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\theta} \exp\left(-\frac{\theta}{2\sigma_\theta}\right) \tag{A3}$$

To compute the MAP estimate, we need the density function $p(\vec{r})$. However, noticing that $p(\theta \mid \vec{r})$ is a probability density function with respect to $\theta$ for any observation vector $\vec{r}$, the density function $p(\vec{r})$ only contributes to the constant needed to make

$$\int_{-\infty}^{\infty} p(\theta \mid \vec{r}) d\theta = 1 \tag{A4}$$

From equation (8) it follows that

$$p(\theta \mid \vec{r}) = \left[\frac{\frac{1}{\sqrt{2\pi}\sigma_\theta} \prod_{k=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_n}}{p(\vec{r})}\right] \exp\left(-\frac{1}{2}\left[\frac{\sum_{k=1}^{M}(r_k - \theta)^2}{\sigma_n^2} + \frac{\theta^2}{\sigma_\theta^2}\right]\right) \tag{A5}$$

The term in the exponent can be rewritten as

$$-\frac{1}{2\sigma_p^2}\left[\theta^2 - \frac{2\theta\sigma_\theta^2}{M\sigma_\theta^2 + \sigma_n^2}\sum_{k=1}^{M} r_k + \frac{\sigma_\theta^2}{M\sigma_\theta^2 + \sigma_n^2}\sum_{k=1}^{M} r_k^2\right] \tag{A6}$$

where

$$\sigma_p^2 = \frac{\sigma_\theta^2 \sigma_n^2}{M\sigma_\theta^2 + \sigma_n^2} \tag{A7}$$

Adding and subtracting the term $\dfrac{\sigma_\theta^4}{(M\sigma_\theta^2 + \sigma_n^2)^2}\left(\displaystyle\sum_{k=1}^{M} r_k\right)^2$ from (A6) and completing the square, we have

$$-\frac{1}{2\sigma_p^2}\left[\theta - \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_n^2/M}\left(\frac{1}{M}\sum_{k=1}^{M} r_k\right)\right]^2 - \frac{1}{2\sigma_p^2}\left[\frac{\sigma_\theta^2}{M\sigma_\theta^2 + \sigma_n^2}\sum_{k=1}^{M} r_k^2 - \frac{\sigma_\theta^4}{(M\sigma_\theta^2 + \sigma_n^2)^2}\left(\sum_{k=1}^{M} r_k\right)^2\right] \qquad (A8)$$

Absorbing the terms containing the observation vector $\vec{r}$ in the constant $\Psi(\vec{r})$, we have

$$p(\theta\,|\,\vec{r}) = \Psi(\vec{r})\exp\left\{-\frac{1}{2\sigma_p^2}\left[\theta - \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_n^2/M}\left(\frac{1}{M}\sum_{k=1}^{M} r_k\right)\right]^2\right\} \qquad (A9)$$

Examining (A9) we see that $p(\theta\,|\,\vec{r})$ is just a Gaussian density. In this case the maximum value of $p(\theta\,|\,\vec{r})$ occurs at the conditional mean so that

$$\hat{\theta}_{\text{MAP}} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_n^2/M}\left(\frac{1}{M}\sum_{k=1}^{M} r_k\right) \qquad (A10)$$

Further, since the minimum mean square estimate is just the conditional mean of $p(\theta\,|\,\vec{r})$, it follows that the minimum mean square estimate is equal to the MAP estimate, i.e. $\hat{\theta}_{\text{MSE}} = \hat{\theta}_{\text{MAP}}$.

### A.2   Likelihood ratio test

We will derive the likelihood ratio test statistic under the assumption that the conditional densities $p(\vec{r}\,|\,H_0)$ and $p(\vec{r}\,|\,H_1)$ are Gaussian. Let the conditional mean under the two hypothesis be given by

$$E[\vec{r}\,|\,H_0] = \begin{bmatrix} E(r_1\,|\,H_0) \\ \vdots \\ E(r_M\,|\,H_0) \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \vdots \\ \mu_{0M} \end{bmatrix} \equiv \vec{\mu}_0 \qquad (A11)$$

$$E[\vec{r}\,|\,H_1] = \begin{bmatrix} E(r_1\,|\,H_1) \\ \vdots \\ E(r_M\,|\,H_1) \end{bmatrix} = \begin{bmatrix} \mu_{11} \\ \vdots \\ \mu_{1M} \end{bmatrix} \equiv \vec{\mu}_1 \qquad (A12)$$

We will assume that the covariance matrix under the two hypotheses is equal and given by

$$\Sigma = E\left[(\vec{r} - \vec{\mu}_0)(\vec{r}^T - \vec{\mu}_0^T)\big|H_0\right] = E\left[(\vec{r} - \vec{\mu}_1)(\vec{r}^T - \vec{\mu}_1^T)\big|H_1\right] \qquad (A13)$$

The conditional densities under the two hypotheses are then given by

$$p(\vec{r}\,|\,H_0) = \left[(2\pi)^{1/M}\,|\Sigma|^{1/2}\right]^{-1}\exp\!\left[-\tfrac{1}{2}(\vec{r}^{\,T}-\vec{\mu}_o^{\,T})\Sigma^{-1}(\vec{r}-\vec{\mu}_0)\right] \qquad\text{(A14)}$$

$$p(\vec{r}\,|\,H_1) = \left[(2\pi)^{1/M}\,|\Sigma|^{1/2}\right]^{-1}\exp\!\left[-\tfrac{1}{2}(\vec{r}^{\,T}-\vec{\mu}_1^{\,T})\Sigma^{-1}(\vec{r}-\vec{\mu}_1)\right] \qquad\text{(A15)}$$

It follows from equation (18) that the likelihood ratio test is given by

$$\ell(\vec{r}) = \tfrac{1}{2}(\vec{r}^{\,T}-\vec{\mu}_0^{\,T})\Sigma^{-1}(\vec{r}-\vec{\mu}_0) - \tfrac{1}{2}(\vec{r}^{\,T}-\vec{\mu}_1^{\,T})\Sigma^{-1}(\vec{r}-\vec{\mu}_1) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \qquad\text{(A16)}$$

Setting $\Sigma^{-1} = Q$ we can simplify (A16) to

$$\ell(\vec{r}) = (\vec{\mu}_1^{\,T}-\vec{\mu}_0^{\,T})\,Q\,\vec{r} \underset{H_0}{\overset{H_1}{\gtrless}} \eta + \tfrac{1}{2}\!\left(\vec{\mu}_1^{\,T}Q\,\vec{\mu}_1 - \vec{\mu}_0^{\,T}Q\,\vec{\mu}_0\right) \equiv \gamma \qquad\text{(A17)}$$

### A.3  Hypothesis test performance

The distance $d$ between the means under the two hypotheses when the variance is normalised to one can be used to characterise the test performance. The square of this distance $d^2$ is given by

$$d^2 = \frac{\left[E(\ell\,|\,H_1) - E(\ell\,|\,H_0)\right]^2}{Var(\ell\,|\,H_0)} \qquad\text{(A18)}$$

where $\ell = \ell(\vec{r})$ denotes the likelihood ratio given by (A18). Taking conditional expectations, we have

$$E(\ell\,|\,H_0) = (\vec{\mu}_1^{\,T}-\vec{\mu}_0^{\,T})\,Q\,\vec{\mu}_0 = \Delta\vec{\mu}^{\,T}\,Q\,\vec{\mu}_0 \qquad\text{(A19)}$$

$$E(\ell\,|\,H_1) = (\vec{\mu}_1^{\,T}-\vec{\mu}_0^{\,T})\,Q\,\vec{\mu}_1 = \Delta\vec{\mu}^{\,T}\,Q\,\vec{\mu}_1 \qquad\text{(A20)}$$

The conditional variance is given by

$$Var(\ell\,|\,H_0) = E\!\left\{[\Delta\vec{\mu}^{\,T}\,Q\,(\vec{r}-\vec{\mu}_0)][(\vec{r}^{\,T}-\vec{\mu}_0^{\,T})\,Q\,\Delta\vec{\mu}]\right\} \qquad\text{(A21)}$$

This can be simplified by taking expectations of the variables to

$$Var(\ell\,|\,H_0) = \Delta\vec{\mu}^{\,T}\,Q\,E\!\left\{(\vec{r}-\vec{\mu}_0)(\vec{r}^{\,T}-\vec{\mu}_0^{\,T})\right\}Q\,\Delta\vec{\mu} = \Delta\vec{\mu}^{\,T}Q\,\Sigma\,Q\,\Delta\vec{\mu} \qquad\text{(A22)}$$

Noting that $\Sigma^{-1} = Q$, we have

$$d^2 = \Delta\vec{\mu}^{\,T}Q\,\Delta\vec{\mu} \qquad\text{(A23)}$$

### A.4   Bayes criterion

In any binary hypothesis testing problem, there are four possible outcomes and to each of these we assign a cost. These are:

(a)  $H_0$ is true and the test decides $H_0$ (cost $C_{00}$)

(b)  $H_0$ is true and the test decides $H_1$ (cost $C_{01}$)

(c)  $H_1$ is true and the test decides $H_0$ (cost $C_{10}$)

(d)  $H_1$ is true and the test decides $H_1$ (cost $C_{11}$)

The application of the Bayes criterion assumes that the a priori probabilities of occurrence of the events $H_0$ and $H_1$ are known and are given by $p_0$ and $p_1$, respectively. The choice of the Bayes decision rule is based on minimising the average cost of the various outcomes of the hypothesis test. If we assume that $\gamma$ is the test threshold, then denoting the expected value of cost as risk $\Re$, we have

$$\Re = C_{00} p_0 \; \Pr(\textit{say } H_0 \mid H_0 \textit{ is true}) + C_{10} p_0 \; \Pr(\textit{say } H_1 \mid H_0 \textit{ is true})$$

$$+ C_{01} p_1 \; \Pr(\textit{say } H_0 \mid H_1 \textit{ is true}) + C_{11} p_1 \; \Pr(\textit{say } H_1 \mid H_1 \textit{ is true}) \tag{A24}$$

If we assume that making a right decision does not involve any cost, then we can set $C_{00} = C_{11} = 0$. Using equations (25) and (26) and assuming that the cost of incorrect decisions is equal and given by $C_{10} = C_{01} = 1$, we have

$$\Re = p_0 \, P_F + p_1 \, (1 - P_D) \tag{A25}$$

Under the assumption that the a priori probabilities of occurrence of the events $H_0$ and $H_1$ are equally likely, we have

$$\Re = 0.5 \, P_F + 0.5 \, (1 - P_D) \tag{A26}$$

# References

Casdagli, Martin (1989): "Non-linear Prediction of Chaotic Time Series". *Physica D*, Vol. 35, pp. 335-56.

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

Diebold, F. X. and J. A. Nason (1990): "Nonparametric Exchange Rate Prediction?". *Journal of International Economics*, Vol. 28, pp. 315-32.

Dockner, E. J., A. Prskawetz and G. Feichtinger (1997): "Non-linear Dynamics and Predictability in the Austrian Stock Market", in *System Dynamics in Economic and Financial Models*, edited by C. Heij, H. Schumacher, B. Hanzon and K. Praagman.

Flood, R. P., A. K. Rose and D. J. Mathieson (1990): "An Empirical Exploration of Exchange Rate Target Zones". *NBER Working Paper* 3543.

Krugman, P. (1988): "Target Zones and Exchange Rate Dynamics". *NBER Working Paper* 2481.

Mark, N. (1985): "On Time Varying Risk Premia in Foreign Exchange Market: An Econometric Analysis". *Journal of Monetary Economics*, Vol. 16, pp. 3-24.

Mizrach, B. (1993): "Multivariate Nearest-Neighbour Forecasts of EMS Exchange Rate", in *Non-linear Dynamics, Chaos and Econometrics*, edited by Hashem Pesaran and Simon Potter. John Wiley.

Parhizgari, A. M. and Maria Eugenia De Boyrie (1997): "Predicting Spot Exchange Rates in a Non-linear Estimation Framework Using Futures Prices". *Journal of Futures Markets*, Vol. 17, pp. 935-56.

Powell, M. J. D. (1981): "Approximation Theory and Methods". Cambridge University Press.

Scheinkman, J. A. and B. LeBaron (1989): "Non-linear Dynamics and Stock Returns". *The Journal of Business*, Vol. 62, pp. 311-37.

Srichander, R. (1995): "Efficient Schedules for Simulated Annealing". *Engineering Optimization*, Vol. 24, pp. 161-76.

Van Trees, H. L. (1968): "Detection, Estimation and Modulation Theory". John Wiley.

# Recent BIS Working Papers

| No. | Title | Author |
| --- | --- | --- |
| 41<br>May 1997 | The euro and European financial markets | Robert N. McCauley<br>and William R. White |
| 42<br>June 1997 | Measuring monetary policy shocks in France, Germany and Italy: the role of the exchange rate | Frank Smets |
| 43<br>July 1997 | Exchange rate regimes and the expectations hypothesis of the term structure | Stefan Gerlach<br>and Frank Smets |
| 44<br>July 1997 | Is there excess comovement of bond yields between countries? | Gregory D. Sutton |
| 45<br>August 1997 | A multi-country comparison of the linkages between inflation and exchange rate competitiveness | Steven B. Kamin |
| 46<br>September 1997 | Global asset allocation in fixed income markets | Srichander Ramaswamy |
| 47<br>September 1997 | Financial asset prices and monetary policy: theory and evidence | Frank Smets |
| 48<br>September 1997 | Some multi-country evidence on the effects of real exchange rates on output | Steven B. Kamin<br>and Marc Klau |
| 49<br>September 1997 | Why does the yield curve predict economic activity? Dissecting the evidence for Germany and the United States | Frank Smets and<br>Kostas Tsatsaronis |
| 50<br>November 1997 | The euro and the dollar | Robert N. McCauley |
| 51<br>November 1997 | Forecast errors and financial developments | Palle S. Andersen |
| 52<br>January 1998 | Inflation and disinflation in Iceland | Palle S. Andersen and<br>Már Guðmundsson |
| 53<br>March 1998 | Exchange rate regimes and inflation and output in Sub-Saharan countries | Marc Klau |
| 54<br>June 1998 | The coming transformation of continental European banking? | William R. White |
| 55<br>June 1998 | Spread overreaction in international bond markets | Gregory D. Sutton |
| 56<br>June 1998 | Commercial banks in the securities business: a review | João A. C. Santos |