White Rose Research Online
eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/82948/

Monograph:
Hong, X. and Billings, S.A. (1997) Time Series Multistep Ahead Predictability Estimation and Ranking. Research Report. ACSE Research Report 687 . Department of Automatic Control and Systems Engineering

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

ETURN
LIBRARY

p5947337 X

Time Series Multistep Ahead Predictability Estimation and Ranking

X. Hong and S. A. Billings
Department of Automatic Control and Systems Engineering,
University of Sheffield, Mappin Street, Sheffield S1 3JD

Abstract — A predictability index was defined as the ratio of the variance of the optimal prediction to the variance of the original time series by Granger and Anderson (1976) and Bhansali (1989). A new simplified algorithm for estimating the predictability index is introduced and the new estimator is shown to be a simple and effective tool in applications of predictability ranking and as an aid in the preliminary analysis of time series. The relationship between the predictability index and the position of the poles and lag p of a time series which can be modelled as an AR(p) model are also investigated. The effectiveness of the algorithm is demonstrated using numerical examples including an application to stock prices.

Keywords — time series, predictability, autocorrelation, moving average model, autoregressive model

Research Report No. 687

Sept 1997

200412726

1 Introduction

Multi-step ahead prediction is concerned with the problem of forecasting a given time series into the future. Given several time series we may wish to evaluate qualitatively and quantitatively the predictability for different prediction steps as a preliminary analysis. The aim is to evaluate predictability as a measure of ranking time series from unknown model structures to aid preprocessing. In so doing we may smooth the data, and compute functions or indicators of the time series to enhance the prediction accuracy. In view of this objective, a proper metric of predictability would be desirable and a way of ranking or assessing the series through a simple reliable estimator of the predictability would be valuable.

Under the assumption that the process is ergodic, the predictability of a time series can be interpreted as the dependency between the future value and the past values. For a k-step ahead prediction the predictability can be measured as the mutual information between the output $y(t+k)$, that is to be predicted, with all the information before the time origin t. The mutual information measures the reduction in uncertainty of $y(t+k)$ due to a knowledge of past information (Zheng and Billings, 1996). The computation of the mutual information involves determining the entropy which is computed based on the estimation of the probability and conditional distribution function of the variables. This method is therefore computationally very expensive. The predictability for a k-step ahead prediction of a time series is also closely related to the significance of the correlation between the two outputs $y(t)$ and $y(t+k)$ which is very simple to compute. The sample autocorrelation coefficients of a time series may therefore provide some indication of the predictability but this is unlikely to be an accurate descriptive measure of predictability.

Based on n_y past values $y(t), y(t-1), \cdots, y(t-n_y+1)$, and starting at the time origin t, an optimal k-step ahead predictor can be obtained by minimising the criterion of mean squared error $E[(y(t+k)-\hat{y}(t+k|t))^2|y(t), y(t-1), \cdots, y(t-n_y+1)]$. This yields

$$\hat{y}(t+k|t) = E[y(t+k)|y(t), y(t-1), \cdots, y(t-n_y+1)] \tag{1}$$

A predictability index which is denoted as $R^2(k)$ has been defined as the ratio of the variance of the optimal prediction to the variance of the original time series (Granger and Anderson, 1976, R. J. Bhansali, 1989). Generally $\hat{y}(t+k|t)$ is a nonlinear function of the n_y past values $y(t), y(t-1), \cdots, y(t-n_y+1)$. Assume that the k-step-ahead prediction error $\xi(t+k|t) = y(t+k)-\hat{y}(t+k|t)$ is unpredictable from all linear and non-linear combinations of $y(t), y(t-1), \cdots, y(t-n_y+1)$. Then it may be assumed that $\xi(t+k|t) = y(t+k)-\hat{y}(t+k|t)$ is zero-mean and uncorrelated to

SHEFFIELD UNIV. LIBRARY ST. GEORGE'S

$\hat{y}(t+k|t)$. Therefore it may be shown that

$$R^2(k) = \frac{E[\hat{y}(t+k|t)^2]}{E[y(t)^2]}$$
$$= 1 - \frac{\sigma_k^2}{E[y(t)^2]} \quad (2)$$

where $\sigma_k^2 = E[\xi(t+k|t)^2]$ is the variance of the k-step ahead prediction error. The k-step ahead predictability index $R^2(k)$ is a non-increasing function of k which is always less than 1. If the series is perfectly predictable $R^2(k)$ will be equal to 1, and if the series is completely unpredictable, it will equal 0.

The estimation of the variance of the k-step ahead prediction error σ_k^2 is crucial in the estimation of the k-step ahead predictability index $R^2(k)$. In the application of time series prediction, a mathematical model is identified and applied to predict future values. When a model is used in prediction, this will serve as a predictor and the variance of the k-step ahead prediction error can be computed and the predictability index can be readily calculated. However different models for one time series may give rise to different prediction errors, and different time series usually need to be fitted with different models. If one model fails to give good k-step ahead predictions this does not mean that the time series may not be predicted well using other models. The computation of the k-step ahead prediction error by building a non-linear model is not only cumbersome but also model dependent. Therefore, the procedure of evaluating the predictability of a time series by fitting a nonlinear model is not appropriate for the present objective.

Given a time series, one simple estimator of $R^2(k)$ can be constructed based on a locally linearised long moving average model which is constructed from sample autocorrelation coefficients of the series (R. J. Bhansali, 1989). In the present work, a new algorithm for the predictability index estimator is introduced to simplify this procedure. A formula for the variance of the time series is used (Box and Jenkins, 1976) so that the estimation of the variance of the noise, which has been used in Bhansali's procedure, becomes unnecessary. The relationship between the predictability to the position of the poles and the lag p if a given time series is properly fitted into a AR(p) model are also studied. It is shown that the new estimator of the predictability index is a simple and effective tool in applications of time series predictability ranking. The effectiveness is demonstrated with examples of both simulated and economic time series.

2 Multi-step Ahead Predictability Index Estimator

2.1 Algorithm

Given a time series, one simple reliable estimator of $R^2(k)$ can be constructed based on a locally linearised long moving average model which is constructed from sample autocorrelation coefficients of the time series (R. J. Bhansali, 1991). It is assumed that the time series may be represented by a form of moving average (MA) model with a predetermined and sufficiently large lag. The procedure uses the sample autocovariance to fit an autoregressive (AR) model with a predetermined sufficiently large lag, followed by the transfer of the AR model into a long lag MA model.

A new procedure to compute the new predictability estimator which is defined as $R_l^2(k)$, where the subscript l denotes that this is based on a locally linear assumption, is introduced to simplify the above approach. A formula for the variance of the time series is used which has been introduced in Box and Jenkins (1976) so that the estimation of the variance of the noise becomes unnecessary. The procedure is given as follows:

(i). Define a 'window' over a period of the series $y(1), y(2), \cdots, y(N)$, and remove the mean. The time series should be long enough, typically 300 to 1000 points.

(ii). In the AR expression (Box and Jenkins, 1976)

$$y(t) = \sum_{j=1}^{\infty} \pi_j y(t-j) + \xi(t) \tag{3}$$

because $\pi_j \to 0$, as $j \to \infty$, it may be assumed that $\pi_j \approx 0$, when $j > n_y$, where n_y is a predetermined positive integer which is large enough. So a finite AR(n_y) model which is an approximation to an infinite order AR process takes the truncated form of equation Eq.(3) as

$$y(t) = \sum_{j=1}^{n_y} \pi_j y(t-j) + \xi(t) \tag{4}$$

(iii) The π_j's can be computed by solving the n_yth order Yule-Walker equation

$$\begin{cases} \rho_{yy}(1) = \pi_1 + \pi_2\rho_{yy}(1) + \cdots + \pi_{n_y}\rho_{yy}(n_y-1) \\ \rho_{yy}(2) = \pi_1\rho_{yy}(1) + \pi_2 + \cdots + \pi_{n_y}\rho_{yy}(n_y-2) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \rho_{yy}(n_y) = \pi_1\rho_{yy}(n_y-1) + \pi_2\rho_{yy}(n_y-2) + \cdots + \pi_{n_y} \end{cases}$$

as

$$\Pi = \mathbf{P}_\rho^{-1}\mathbf{r}_\rho \tag{5}$$

where

$$\Pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_p \end{bmatrix}, \qquad \mathbf{r}_\rho = \begin{bmatrix} \rho_{yy}(1) \\ \rho_{yy}(2) \\ \vdots \\ \rho_{yy}(n_y) \end{bmatrix}$$

$$\mathbf{P}_\rho = \begin{bmatrix} 1 & \rho_{yy}(1) & \rho_{yy}(2) & \cdots & \rho_{yy}(n_y-1) \\ \rho_{yy}(1) & 1 & \rho_{yy}(1) & \cdots & \rho_{yy}(n_y-2) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{yy}(1) & \rho_{yy}(2) & \rho_{yy}(3) & \cdots & 1 \end{bmatrix}$$

in which

$$\rho_{yy}(k) = \frac{cov[y(t), y(t+k)]}{E[y(t)^2]} \tag{6}$$

are the theoretical autocorrelation coefficient. These are replaced by the estimated autocorrelations which are computed from

$$\rho_{yy}(k) = \frac{\sum_{t=1}^{N-k} y(t)y(t+k)}{\sum_{t=1}^{N-k} y(t)^2} \qquad 0 \le k \le n_y \tag{7}$$

(iv). A long moving average (MA) model of order n_y of the time series

$$y(t) = \sum_{j=0}^{n_y} \psi_j \xi(t-j), \qquad \psi_0 = 1 \tag{8}$$

is given by assuming $\psi_j \approx 0$ for $j > n_y$ in the infinite moving average (MA) model (Box and Jenkins, 1976)

$$y(t) = \sum_{j=0}^{+\infty} \psi_j \xi(t-j), \qquad \psi_0 = 1$$

The coefficients $\psi_j's$ are computed (Box and Jenkins, 1976) as

$$\psi_j = \sum_{i=1}^{j} \psi_{j-i}\pi_i, \qquad \psi_0 = 1 \tag{9}$$

(v). An estimator for the variance of the k-step-ahead prediction error is given as (Box and Jenkins, 1976)

$$\sigma_k^2 = (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{k-1}^2)\sigma^2 \tag{10}$$

where σ^2 is the variance of the noise $\xi(t)$, and the variance of the time series is given (Box and Jenkins, 1976) as

$$E[y(t)^2] = \frac{\sigma^2}{1 - \sum_{i=1}^{n_y} \rho_{yy}(i)\pi_i} \tag{11}$$

(vi). Substituting Eq.(10) and Eq.(11) into Eq.(2), the estimate of the k-step ahead predictability index $R^2(k)$ takes the form

$$R_l^2(k) = 1 - (1 + \sum_{i=1}^{k-1} \psi_i^2)(1 - \sum_{i=1}^{n_y} \rho_{yy}(i)\pi_i) \tag{12}$$

Notice that Eq.(12) does not contain the term σ^2, the variance of the noise, and the estimation of σ^2 becomes unnecessary.

Remarks

(i). The asymptotic distribution of the parameters of the long moving average model as well as the prediction error variance was vigorously analysed by Bhansali (1989) and Saikkonen (1986). The result is that under the condition that the order n_y approaches ∞ simultaneously but sufficiently slowly with N, the distributions are asymptotically normal. In practice, it is reasonable to assume that if N is the data length, n_y can be selected as approximately $\frac{1}{10} of N$.

(ii). Eq.(10) shows that the variance of the k-step ahead prediction error σ_k^2 is proportional to the variance σ^2, the one-step-ahead prediction error under the condition that the coefficients of the moving average model are correctly identified. For a linear model the distinction between the criterion of minimising the one-step-ahead prediction error and the k-step-ahead prediction error is trival if the order of the model is not underestimated.

2.2 General Properties

$R_l^2(k)$ is an estimate of $R^2(k)$ based on the assumption that the process is locally linear. Consider a time series which is properly fitted by an AR(p) model which is given by

$$(1 - \sum_{j=1}^{p} \pi_j z^{-j}) y(t) = 0$$

Note that p is not equal to but often smaller than the order of the Yule-Walker equation n_y that is preset as a very large number in the procedure of computing the predictability index. It is shown in the following how the predictability estimator $R_l^2(k)$ is dependent on the characteristic polynomial of the system which is written as

$$1 - \sum_{j=1}^{p} \pi_j z^{-j} = 0 \tag{13}$$

This can also be expressed as

$$\prod_{j=1}^{p} (1 - z_j z^{-1}) = 0 \tag{14}$$

where $z_j, j = 1, \ldots, p$ are all the poles which are assumed to be inside the unit circle. The estimated predictability will be influenced by the number and the position of the poles. Consider a first order AR process as an example

$$y(t) = z_1 y(t-1) + \xi(t) \tag{15}$$

After some simple calculations,

$$R_l^2(k) = z_1^{2k} \tag{16}$$

It is seen that if the pole z_1 is close to 1, then $R_l^2(k)$ will be high. Now consider a first order AR process with a lag p:

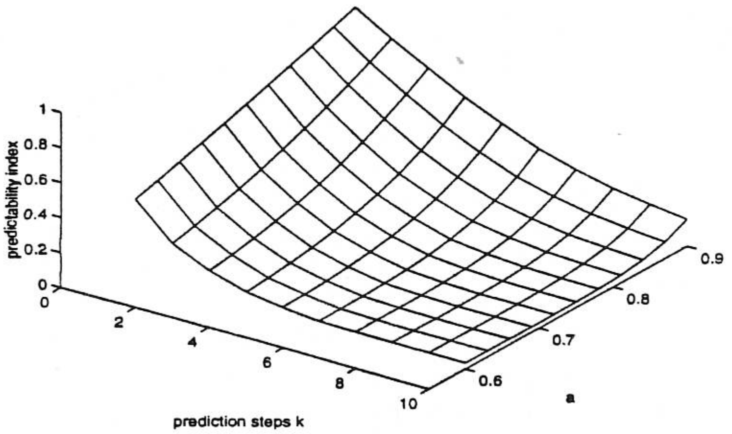
$$y(t) = ay(t-p) + \xi(t) \tag{17}$$

After some simple calculation,

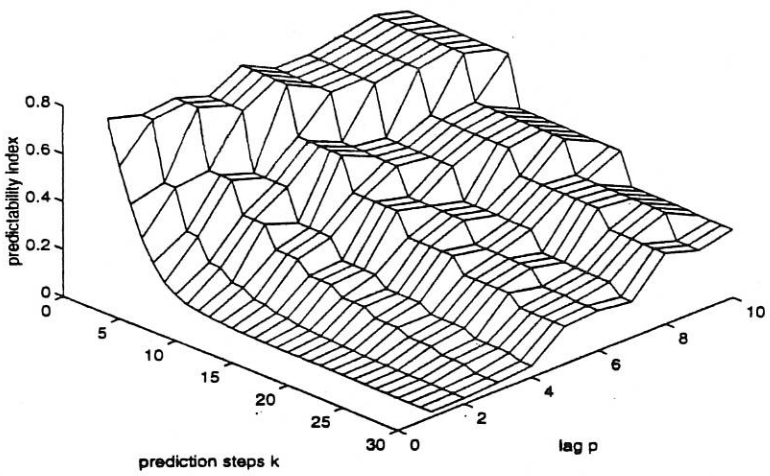
$$R_l^2(k) = a^{2(\mathrm{Int}(k/p)+1)} \tag{18}$$

where $\mathrm{Int}(\bullet)$ denotes the integer part of $\bullet$. This implies that the closer the poles are to the unit circle, the higher $R_l^2(k)$ will be. The larger the lag p (the number of the poles), the higher $R_l^2(k)$ as k grows.

This conclusion is illustrated in Fig.1. At first set $p = 1$ and let a gradually vary from 0.6 to 0.9. For each pair of fixed p and a, a time series of 500 points was generated using Eq.(17) and the order of the Yule-Walker equation was set as 50. Then $R_l^2(k)$ was computed and plotted against a. This is shown in Fig.1(a). Then, set $a = 0.8$ and let p vary from 1 to 10. For each pair of fixed p and a, a time series of 500 points was generated using Eq.(17) and the order of the Yule-Walker equation was set as 50. The $R_l^2(k)$ was computed and plotted against p. This is shown in Fig.1(b).



(a)



(b)

Figure 1: Estimated predictability index of the time series: $y(t) = ay(t-p)+\xi(t)$, $\xi(t) : N(0,1)$; (a) p=1, a varies from 0.6 to 0.9 and (b) a=0.8, p varies from 1 to 10

3 Applications to Time Series Predictability Ranking

In this section, two examples are given to show how the predictability index estimator is used in time series predictability ranking. The first is a simulated example which is used to show the effectiveness of the predictability index estimator. The second example is an application to real data using bank stock prices.

Example 1: Consider a simulated time series described by

$$x(t) = \sin(0.31415926 * t) \tag{19}$$

A set of 5 time series were generated as $y_i(t) = x(t) + \xi_i(t)$, $i = 1, 2, \cdots, 5$ where the noise $\xi_i(t) : N(0, \sigma_i^2)$. The variance of the noise was $\sigma_i^2 = 0.1^2, 0.2^2, 0.3^2, 0.4^2$ and 0.5^2 for $i = 1, 2, \cdots, 5$ respectively. So the S/N ratios of the 5 series are 40, 28, 21, 16 and 12db respectively. The original series $x(t)$ was a pure deterministic time series. It is clear that as the variance of the additive noise grows larger, the predictability of the corrupted time series should gradually decrease. 500 points was generated for each of the 5 time series which are plotted in Fig.2(a), (b), (c), (d) and (e). The predictability index estimator was computed for each of them in which the order of the Yule-Walker equation was chosen to be 50. The result of ranking these time series according to the predictability index estimator Eq.(12) is shown in Fig.2(f). This shows that the larger the variance of the additive noise, the lower the predictability index of the corrupted time series.

Example 2: A set of 6 main British bank stock prices, Abbey-nt, Barclay, Bank-Scot, Natwest, Royal-Bk and Hsbc which are denoted by $x_i(t)$, $i = 1, 2, \cdots, 6$ respectively were studied. The raw stock price data are plotted in Fig.3. Each of these contains 300 points. Because the time series were nonstationary, the relative movements of the modified stock prices over one step which are given by

$$y_i(t) = \frac{x_i(t+1) - x_i(t)}{x_i(t)} \tag{20}$$

for $i = 1, 2, \cdots, 6$ was used. The new transformed series are stationary and these were ranked according to predictability. The six new series are shown in Fig.4. The predictability index estimator was used to rank these six new series and the result is shown in Fig.5. The result shows that all of these are of low predictablity but the Royal-Bk is more predictable than the others.

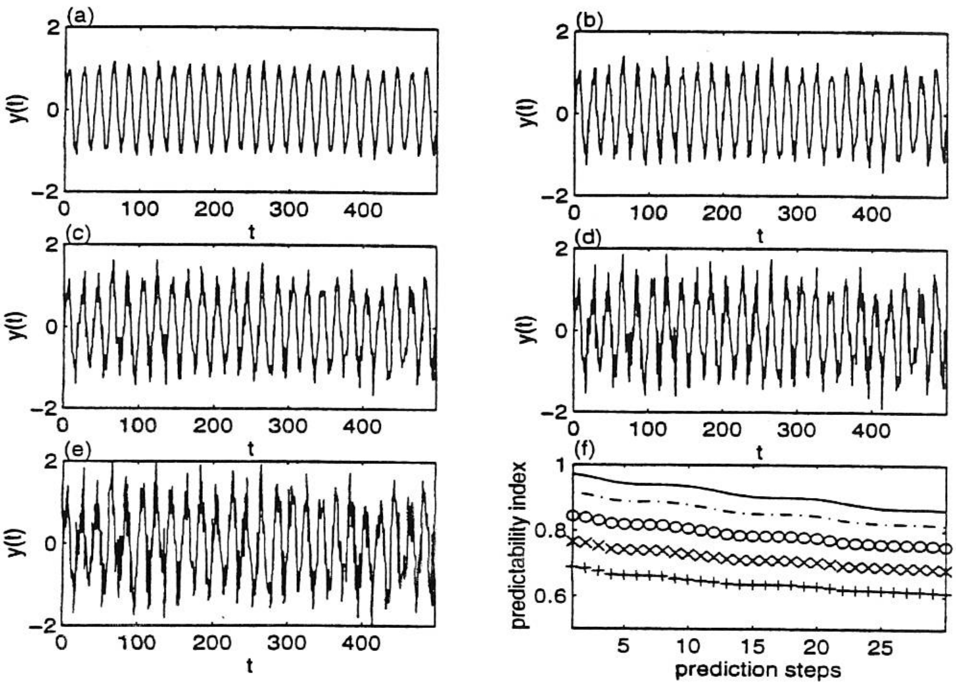


Figure 2: Estimated predictability index of the time series: $x(t) = \sin(0.31415926 * t)$ corrupted by noise $\xi(t) : N(0, \sigma_i^2), y_i(t) = x(t) + \xi_i(t)$; (a) $y_1(t), \sigma_1^2 = 0.1^2$; (b) $y_2(t), \sigma_2^2 = 0.2^2$; (c) $y_3(t), \sigma_3^2 = 0.3^2$; (d) $y_4(t), \sigma_4^2 = 0.4^2$; (e) $y_5(t), \sigma_5^2 = 0.5^2$ and (f) Predictability index $R_l^2(k)$ (Solid: $y_1(t)$; Dashdot: $y_2(t)$; Circle: $y_3(t)$; X mark: $y_4(t)$ and Plus: $y_5(t)$)

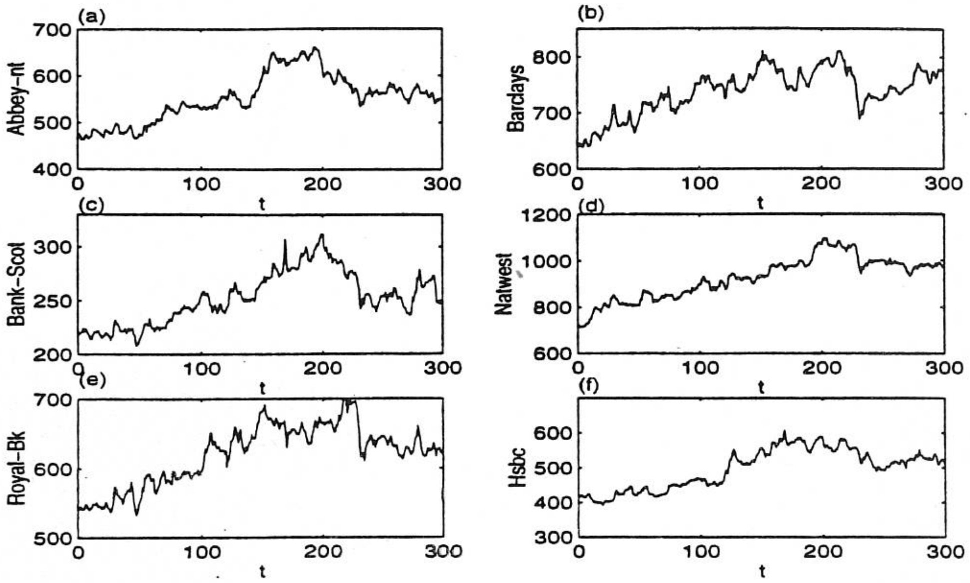


Figure 3: Raw stock prices data of 6 main British banks; (a) Abbey-nt; (b) Barclays; (c) Bank-Scot; (d) Natwest; (e) Royal-Bk and (f) Hsbc

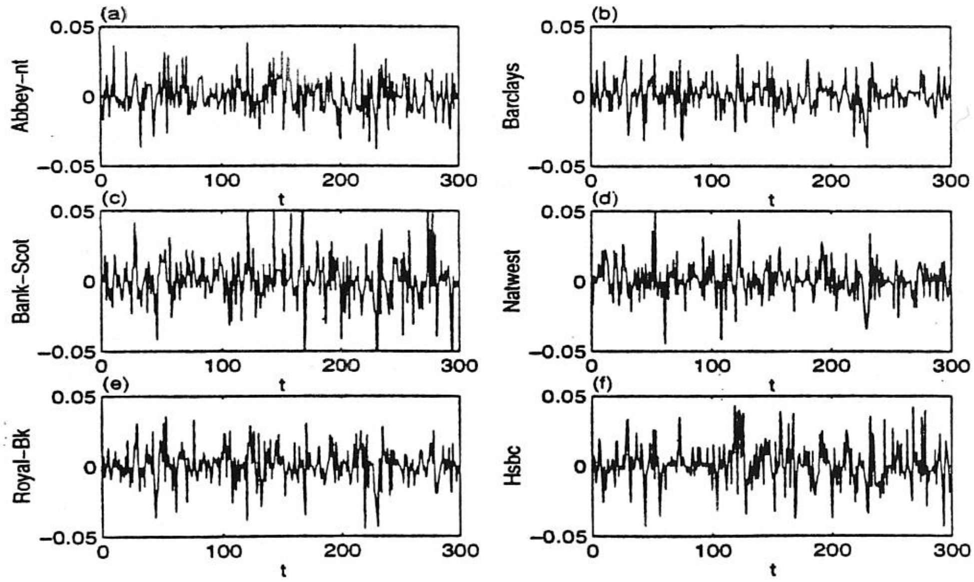


Figure 4: Relative movements of stock prices over the modified data; (a) Abbey-nt; (b) Barclays; (c) Bank-Scot; (d) Natwest; (e) Royal-Bk and (f) Hsbc

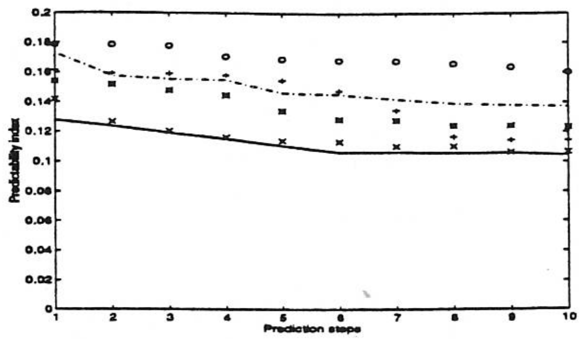


Figure 5: Estimated predictability index $R_l^2(k)$; (Solid: Abbey-nt; Dashdot: Barclays; Circle: Bank-Scot, Star mark: Natwest, Plus: Royal-Bk and X mark: Hsbc)

4 Conclusions

A new algorithm for a predictability index estimator has been introduced and shown to be a simple and effective tool in applications of time series preliminary analysis, such as time series predictability ranking. The relationship between the predictability to the pole position and the lag p of a time series which can be properly fitted into an AR(p) model was also studied. Numerical examples including a practical stock price time series are included to show the effectiveness of the approach.

5 Acknowledgements

SAB gratefully acknowledges that part of this work was supported by EPSRC. XH expresses her thanks for the award of an ORS scholarship which made this study possible.

References

[1] Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control.* Holden-Day Inc., San Francisco.

[2] Bhansali, R. J. (1989). Estimation of the moving-average representation of a stationary process by autoregressive model fitting *J. of Time Series Analysis* . Vol. 10, No. 3, pp215-231.

[3] Bhansali, R. J. (1991). Autoregressive estimation of the prediction mean squared error and an R^2 Measure: an application. in D. Brillinger et al. Eds:*New Directions in Time Series Analysis.* Part I, Springer-Verlag

[4] Granger, C. W. J. and Andersen, A. P. (1976). Non-linear Time Series Modelling. in: David F. Findley (Eds). *Applied Time Series Analysis.* pp.25-38.

[5] Saikkonen, P. (1986). Asymptotic properties of some preliminary estimators for autoregressive moving average time series models. *Journal of Time Series*, Vol. 7, No. 2, pp133-155.

[6] Zheng, G. L. and Billings, S. A. (1996). Radial basis function network configuration using mutual information and the orthogonal least squares algorithm. *Neural Networks.* Vol. 9, No. 9, pp1619-1637.

SHEFFIELD UNIV. LIBRARY ST. GEORGE'S