

ECONOMIC TIME SERIES FORECASTING: BOX-JENKINS METHODOLOGY AND SIGNAL PROCESSING APPROACH

Milan Marček^{1,2}

¹Institute of Computer Science, Faculty of Philosophy and Science, The Silesian University Opava

²The Faculty of Management Science and Informatics, University of Žilina

Abstract

This paper is devoted to the presentation of methods of economic time series analysis and modelling using the Box-Jenkins methodology and the signal processing approach. Some results of our research on time series modelling with emphasis on potential improving forecast accuracy are presented here. The assessment of the particular models has been made using the root mean square error.

1 Introduction

The goal of this paper is to illustrate two modeling approaches: probabilistic and adaptive signal processing for economic time series. In Section 2, we can see that a random process of time series of stock prices may be generated as the output of linear filter driven by white noise. Section 3 is focused on the behaviour of the GL filter and LSL when they are used to forecast future observations of stationary AR processes. In Section 4 of this paper, we report on the problem of forecast accuracy across proposed models.

2 Application of the Box-Jenkins methodology in the stock prediction problem

We give an example that provides one kind of possible results. We will regard these results as the referential values for the approach of adaptive signal processing procedures. Many of modelling techniques of autoregressive processes are based on recent developments in time series analysis recently consolidated and presented by Box and Jenkins [2].

To illustrate the Box-Jenkins methodology, we consider the stock price time readings of a typical company (say VAHOSTAV company). We would like to develop a time series model for this process so that a predictor for the process output can be developed. The data was collected for the period January 2, 1997 to December 31, 1997, which provided a total of 163 observations (see Fig. 1).

To build a forecast model the sample period for analysis y_1, \dots, y_{128} was defined, i.e. the period over which the forecasting model was developed and the ex post forecast period (validation data set), y_{129}, \dots, y_{163} as the time period from the first observation after the end of the sample period to the most recent observation. By using only the actual and forecast values within the ex post forecasting period only, the accuracy of the model can be calculated.

To determine appropriate Box-Jenkins model, a tentative ARMA model in identification step is identified. In Figure 2, the estimate of autocorrelation (\hat{r}_k) and partial autocorrelation \hat{a}_{kk} function (ACF, PACF) of the data are given. To test whether the autocorrelation and partial autocorrelation coefficients are statistically equal to zero, we use the t-statistic $t_r = \hat{r}_k / S(\hat{r}_k)$ and $t_a = \hat{a}_{kk} / S(\hat{a}_{kk})$ where

$$S(\hat{r}_k) = N^{-1/2} \left[1 + 2 \sum_{j=1}^{k-1} \hat{r}_j \right]$$

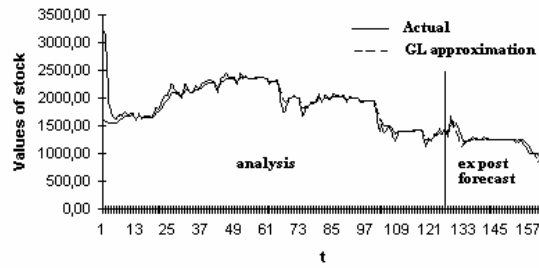


Figure: 1 The data for VAHOSTAV stock prices (January 1997 - December 1997) and the values of the AR(7) model for VAHOSTAV stock prices estimated by GL algorithm

and

$$S(\hat{a}_{kk}) = N^{-1/2}$$

denote standard errors of the k th sample autocorrelation or partial autocorrelation coefficient, respectively, N is the number of data points, k is the lag. Since the ACF decays in an exponential fashion, and the PACF truncates abruptly after lag 2, we may tentatively identify the model for this time series as AR(2).

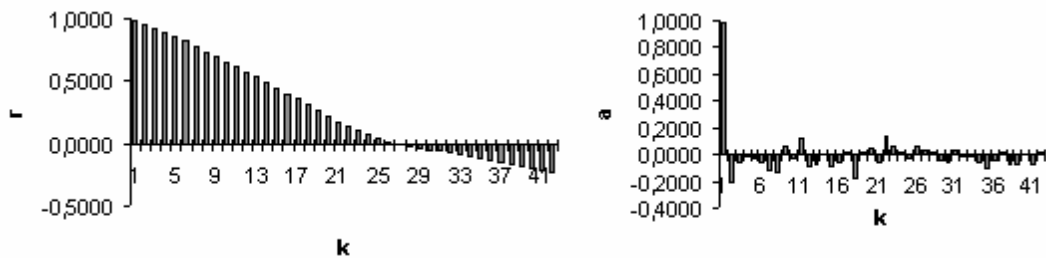


Figure 2: Autocorrelation function and partial autocorrelation function of the data for VAHOSTAV stock prices (period for analysis)

In the estimation step, we compute estimates for the parameters of the AR(2) model

$$y_t = \xi + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \quad (1)$$

$$t = 1, 2, \dots, N - 2$$

or with of obvious matrix notation

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \varepsilon$$

by OLS (Ordinary Least Squared)

$$\hat{\mathbf{a}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \begin{bmatrix} \hat{\xi} \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} 26.693 \\ 1.113 \\ -0.127 \end{bmatrix} \quad (2)$$

In the diagnostic checking step, we test adequacy and closeness of fit of the model to the data by sample autocorrelation function of the residuals say

$$e_t = y_t - \hat{y}_t, \quad t = 1, 2, \dots, N - 2 \quad (3)$$

The sample autocorrelation function of the residuals is shown in Fig. 3.

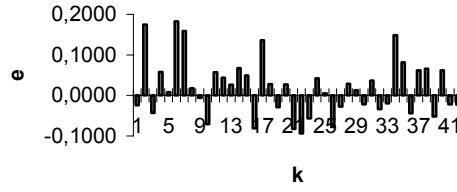


Figure 3: Sample autocorrelation function of the residuals for model (2)

The modified Box-Pierce statistic Q is used for collectively testing the magnitudes of the residual autocorrelations for insignificance. The statistic is [3]

$$Q = (N - d) \sum_{k=1}^K r_{ek}^2 \quad (4)$$

where r_{ek}^2 is the square of the residual autocorrelation coefficients, for lags $k = 1, 2, \dots, K$, d is d th differences of the data. For our stock price time series the Box-Pierce statistic for lag $k = 42$ was computed to be 27.78. This value is less than the critical chi square value of 55.7585 (degrees of freedom is $42 - 2 = 40$, $\alpha = 0.005$). Hence, we can conclude that the error terms are random and the model (1) is an adequate model.

3 Stock price prediction using adaptive signal processing procedures

In practice, the modelling of a set of data as we shown in Section 2, is a much more complex process than the one of fitting and testing. Most of deterministic methods to signal processing have the goal of representing a given sequence $\{y_t\}$ as the impulse response of a rational linear system [1], [4]. The AR model involves a linear filter with the transfer function $H(z)$, where

$$H(z) = [A(z)]^{-1} = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (5)$$

and is also known as the all-pole model which has only a nontrivial denominator polynomial, generating the random process $\{y_t\}$ from the white noise $\{\varepsilon_t\}$. The linear filter is represented by the inverse of the polynomial $[A(z)]^{-1}$. The difference equation for the input-output relationship for this filter is

$$y_t + \sum_{k=1}^p a_k y_{t-k} = \varepsilon_t \quad \text{or} \quad (6)$$

$$y_t = -\sum_{k=1}^p a_k y_{t-k} + \varepsilon_t \quad (7)$$

where a_k are the filter parameters that determine the location of the poles of linear filter $H(z)$. To forecast the observation y_t we can take expectation at origin $t-l$ of the model (7) written at time $t+l$ namely

$$E[y_t] = \hat{y}_t = -\sum_{k=1}^p a_k y_{t-k} \quad (8)$$

For selecting the model order p , we will now monitor prediction Mean Square Error (MSE). If the data is truly described by a finite-order AR model, then the theoretical MSE becomes constant once the model order is reached. A realization of this criterion is shown in Fig. 4.

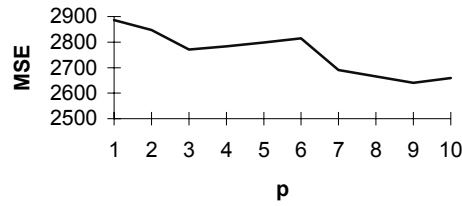


Figure: 4 Graph of MSE (analysis period) versus model order p .

As can be seen from Fig. 4, the MSE's seem to become constant at $p \geq 7$. Then the linear filter (7) has the form

$$y_t = -\sum_{k=1}^7 a_k y_{t-k} + \varepsilon_t \quad (9)$$

The final estimates of predictor parameters (9) are obtained using two of adaptive filtering algorithms in signal processing. The Gradient Lattice (GL) adaptive algorithm and Least Squares Lattice (LSL) form for the parameter estimates of the predictor (9) were used. These algorithms are coded in MATLAB and described in [1]. In this case the process of calculating predictor parameters are updated as each new data point becomes available to track the changing statistics. In Tab.1 parameters of AR process and the corresponding RMSE's (Root Mean Square Errors) for models (1) and (9) are given.

Table 1: OLS, GL AND LSL ESTIMATES OF AR MODELS

Model	Order	Est.proc.	$\hat{\xi}$	\hat{a}_1	\hat{a}_2	\hat{a}_3	\hat{a}_4	\hat{a}_5	\hat{a}_6	\hat{a}_7	RMSE*
(1)	2	OLS	26.639	1.113	0.127						67.787
(1)	7	OLS	45.930	1.085	0.0861	-0.2531	0.0836	-0.0057	0.2081	-0.2281	76.548
(9)	7	GL		-0.7513	-0.1701	-0.0230	-0.0128	-0.0028	-0.0472	0.0084	68.540
(9)	7	LSL		-0.8941	-0.6672	0.7346	-0.2383	0.1805	-0.5692	0.4470	94.570

*ex post forecast period

The RMSE's are called standard deviations of the single-period-ahead forecast errors. For this measure, the AR models estimated by OLS and GL procedures not exceed 5% limit of the variation coefficient, while the variation coefficient (V) computed as $V = \text{RMSE}/\bar{y}$, where \bar{y} is the mean of the stock price time series, for AR model estimated by LSL procedure is 5.59%. Fig. 1 shows the GL prediction results and current values for stock price time series in both analysis and ex post forecast period. The GL approximations in both intervals visually match the current stock series quite well. However the corresponding RMSE value for analysis period is 214.48. This is greater than that for ex post forecast period. The RMSE statistic is here not adequate, because most errors are fairly small, i.e. the model is a good fit to the historical data but there are only first three large errors, these are magnified by using RMSE (since all the errors are squared). This phenomena is produced by GL algorithm itself and this does not influence the ex post forecast errors.

Acknowledgement: This work was supported by Slovak grant foundation under the grant No. 1/2628/05 and from the Grant Agency of the Czech Republic under the grant No. 402/05/2768.

References

- [1] BAYAHAN, G., M.: *Sales Forecasting Using Adaptive Signal Processing Algorithms*. Neural Network World 4-5/1997, Vol. 7, pp. 579-589
- [2] BOX, G. E. P., AND G. M. JENKINS: *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA, 1970
- [3] BOX, G.,E., PIERCE, D.,A.:“*Distribution of Residual Autocorrelation in Autoregressive-Integrated Moving Average Time Series Models*“. Journal of the American Statistical Association, vol. 64, 1977
- [4] THERIEN, C., W.: *Discrete Random Signals and Statistical Signal Processing*. Prentice-Hall, USA, 1992

Author1

Milan Marček: e-mail: marcek@fria.utc.sk