WAGES FORECASTING USING BOX – JENKINS METHODOLOGY

Dušan Marček[1,2]

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

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

Abstract

The purpose of the paper is to demonstrate the overall developing steps of the Box-Jenkins time series modeling, i. e. selecting an appropriate model, tentative estimating the parameters of the selected model, checking the validity of the estimated model and generating forecasts of future observations. In this paper we concentrate on model identification by Hannan-Rissanen method using the Matlab program. Based on work [3], the paper presents the use of Box-Jenkins approach applied to wages forecasts in the Slovak economy.

Keywords: *ARMA Processes, Mean Square Error, Steps in Box-Jenkins Methodology*

Introduction

The wages time series are in fact stochastic in which successive observations are dependent and can be represented by a linear combination of independent random variables $\varepsilon_t, \varepsilon_{t-1}, \ldots$. If the successive observations are highly dependent, we should use in model past values of the time series variable and (or) current and past values of the error terms $\{\varepsilon_t\}$. There are available techniques which are designed to exploit this dependency and which will generally produce superior forecasts. Many of these techniques are based on developments in time series analysis recently presented by Box and Jenkins [1].

Most models for the time series of wages have centered about autoregressive (AR) processes. In Section 1 attention is confined to the application of Box-Jenkins steps (identification, estimation, diagnostic checking and forecasting) for time series modeling of wages. Central to the interest of the ARMA model will be the basic concept of last squares estimation when applied to the linear model and testing of its adequacy. The major goal of next section is to develop a classical ARMA model to predict the wages.

1 Application of ARMA modeling in wages prediction problem

To illustrate of the Box-Jenkins methodology, consider the wages time readings $\{y_t\}$ of the Slovak economy. We would like to develop a time series model for this process so that a predictor for the process output can be developed. The quarterly data were collected for the period January 1, 1991 to December 31, 2006 which provides total of 64 observations (displayed in Fig. 1). To build a forecast model we define the sample period for analysis $x_1, \ldots, x_{60}$, i.e. the period over which we estimate the forecasting model and the ex post forecast period (validation data set), $x_{61}, \ldots, x_{64}$ 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 ex post forecasting period only, the accuracy of the model can be calculated.

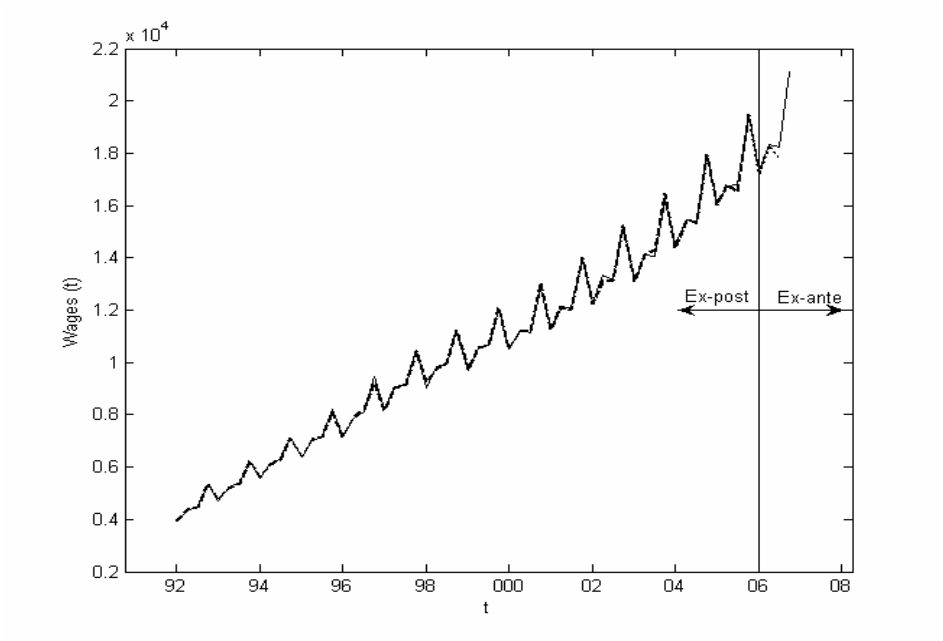

Fig. 1: Nominal wages (January 1991 - December 2006)

To determine appropriate Box-Jenkins model, a tentative ARMA model in identification step is identified. In order to fit a time series to data, first the data were transformed to a stationary ARMA type process, i.e. the data must be modeled by a zero-mean and constant variability. After eliminating trend and seasonal component, the natural logarithms of the once differenced data $y_t = x_t - x_{t-4}$ are shown in Fig. 2.

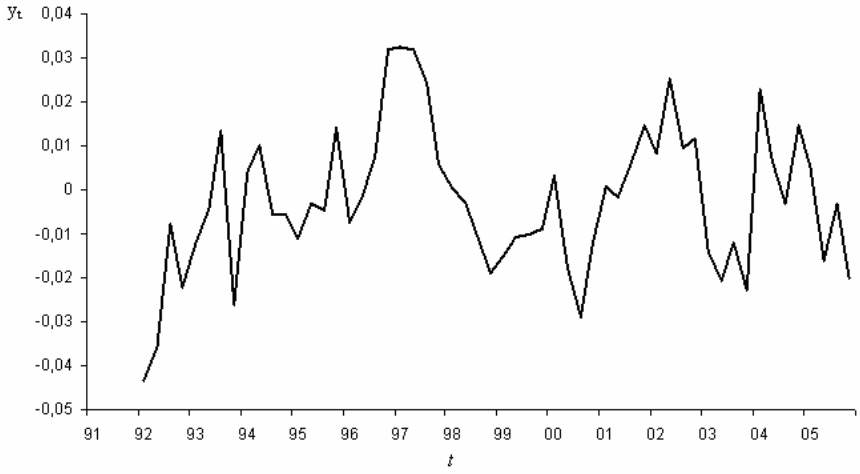

Fig. 2: The wages data after transformation to a stationary ARMA type process

There are various methods and criteria for selecting of an ARMA model. In this section we concentrate on model identification by Hannan-Rissanen procedure [2]. The Matlab program developed in [3] selects as well as estimates the model. Using this program the model for $\{y_t\}$ time series was tentatively identified as ARMA(1,3) with preliminary estimates of the model parameters as follows

$$\hat{y}_t = -0.0017 - 0.46y_{t-1} + 0.905\varepsilon_{t-1} + 0.588\varepsilon_{t-2} + 0.365\varepsilon_{t-3}$$

2 Diagnostic checking

After fitting a model to a given data set, the goodness of fit of the model is usually examined to see if it is indeed an appropriate model. There are various ways of checking if a model is satisfactory. A good way to check the adequacy of an overall Box-Jenkins model is to analyze the residuals y_t - $\hat{y}_t$. If the residuals are truly random, the autocorrelations and partial autocorrelations calculated using the residuals should be statistically equal to zero. Since the residuals are also ordered in time, we can treat them as a time series and calculate the sample correlation function of the residuals and see if it behaves to be a stationary random sequence.

Instead of looking at the correlation function we used the portmanteau test based on the Ljung-Box statistic. The test statistic is [5], [4]

$$Q_h = N(N+2)\sum_{k=1}^{K}\frac{r_e(k)^2}{N-k}$$

which has an asymptotic chi square (χ^2) distribution with *K-p-q* degrees of freedom if the model is appropriate. If $Q_K > \chi^2_{1-\alpha,(K-p-q)}$ the adequacy of the model is reject at the level α. The chi-square statistic applied to these autocorrelation is 13.79, and so we have no evidence to reject the model.

3 Forecasting

Because the model is written in terms of a stationary time series to obtain a point forecast, the final model must be rewritten in terms of the original data and then solved algebraically for x_t. The forecasts obtained from this model for time t = 61,62,63,64 are shown in Fig. 3.

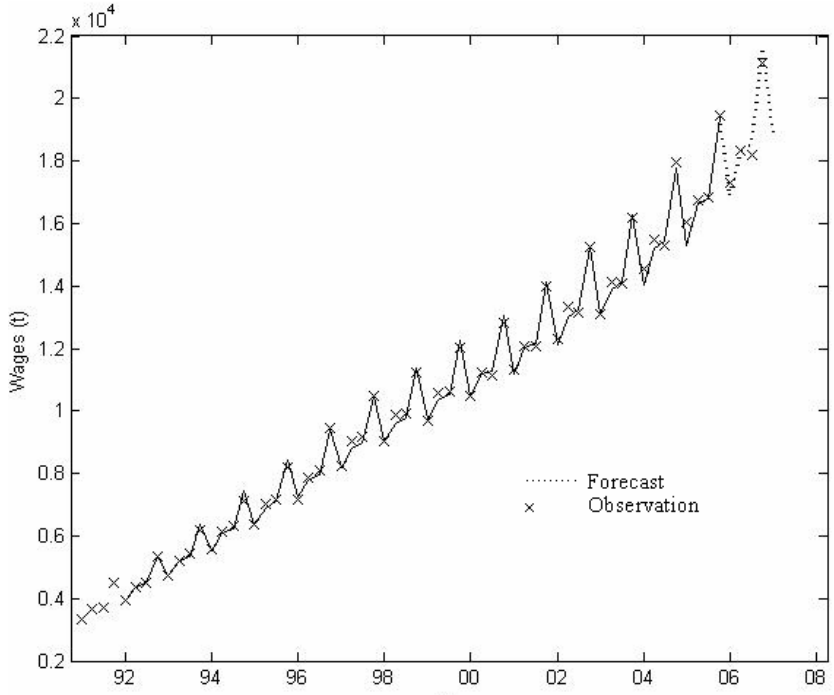

Fig. 3: Forecasts of wages data ARMA(1,3) model

4 Conclusion

This paper has focused on the problems associated with forecasting economic variables, where the assumption of independent errors, and hence independent observations, is not unwarranted. That is, the successive observations of wages data are highly dependent. There are various methods criteria for selecting the orders of an ARMA model. To select an appropriate ARMA model and to estimate the ARMA parameters the Hannan-Rissanen procedure was used. The ARMA(1,3) model for the wages time readings $\{y_t\}$ of the Slovak economy has the statistical summary measure of forecast accuracy MSE better then an alternative model based on exponential smoothing [3].

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. I thank B. Kočvara for computational support.

References:

[1] BOX, G.,E., JENKINS, G., M.: *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco, CA 1976.

[2] GRANGER, C. W. J, and NEWBOLD, P.: *Forecasting Economic Time Series*. Academic Press, NY, 1986

[3] KOČVARA, B.: *Time Series Modeling Using Statistical (Econometric) Methods and Machine Learning*. Diploma work, Faculty of Management Science and Informatics, University of Žilina, April 2007.

[4] MARČEK, D., MARČEK, M.: Time Series Analysis, Modelling and Forecasting with Applications in Economics. EDIS University of Žilina, 2001

[5] MONGOMERY, D.C., JOHNSTON, L.A., GARDINER, J.S.: *Forecasting and Time Series Analysis.* McGraw-Hill, Inc., 1990.

Dušan Marček
e-mail: marcek@fria.fri. uniza.sk