Automatic ARIMA modelling, using TSE-AX

H. Njimi, G. Mélard

Abstract— **The paper consists in a brief description of the expert system TSE-AX and the methodology that has been used in the analysis of the set of 111 time series of the NN3 competition.**

I. Introduction

ARIMA processes made popular by Box and Jenkins is one of the statistical methods to forecast data. ARIMA modelling is more difficult to use than other statistical forecasting techniques although, when implemented properly, it can be quite powerful and flexible. Several algorithms of automated ARIMA modelling were developed in order to make the method more applicable and also available to a greater number of users. Most of these algorithms were implemented using expert systems technology. These systems make it possible to program the knowledge of an expert and to reproduce the reasoning carried out by the system. One of these expert systems for building univariate time series models is TSE-AX. Described by Mélard and Pasteels [1], it is included in Time Series Expert 2.3. An improved version [2] is used here for which Njimi et al. [3] gave an early presentation.

II. DESCRIPTION OF TSE-AX

The objective of TSE-AX [2] is to build ARIMA models in an automated way, with and without an intervention analysis, but so that the user receives the intermediate and final results, and is informed of the quality of the final model. The system is adapted to several categories of users from beginners to experts. The later should use such a tool to save time, being qualified to assess the quality of the final model and possibly propose an alternative model. Briefly, TSE-AX covers everything from the specification stage to the forecasting stage, given that the latter is immediate when a final model has been found. The user can specify his or her model building preferences: perform an intervention analysis or not, choose a specification strategy, etc.

The modelling stage of TSE-AX consists of a succession of several phases. At the beginning, the user gives some information to the system like periodicity of the data and the sample to be used. The automated procedure starts with the preliminary stage, where interventions are selected, transformations are performed, and differences, regular and/or seasonal, are chosen to make sure that the series becomes stationary. To select differences, the user can choose between options based on the non-parametric test of Kruskal and Wallis [4] and presence of autocorrelations or comparison of variances [1], [2]. Next follows the specification stage, where an ARIMA model is identified using one of these three strategies: 'expert' [2], 'autoregressive specification' [6] and 'mixed' [2] where a certain number of models are fitted and a choice among them is made. The remaining stages are the estimation stage, where the final model is fitted, the model checking stage, were the adequacy of that model is investigated, and the forecasting stage. In all modelling stages, the parameters are estimated by exact maximum likelihood.

There are more than twenty input commands that enable the user to customize the modelling strategy. They are concerned with the treatment of outliers by intervention analysis (several types are supported like additive outliers (AO) and level shift (LS)), the seasonal component, the Box-Cox transformation and difference operators. These commands are typically entered into a file and can act either on a single series or on a stream of series.

III. Modelling methodology

Our analysis of the NN3-Competition was performed using TSE-AX version 2.4. For the treatment by TSE-AX, default values for all commands were used. We have accepted the treatment of outliers by intervention analysis to avoid extreme values that would badly influence the various steps of the analysis: specification, estimation, test for adequacy, and forecasting. We also used the 'Mixed strategy' in the step of specification because that strategy is the most complete one [3].

Using the mixed strategy, TSE-AX fits at most 32 models. All fits are done in TSE-AX by exact maximum likelihood estimation.

IV. Closing comments

In Section 2, we have provided a brief description of the expert system, TSE-AX, which we used to forecast the set of 111 time series of the NN3 competition. Participating in this competition with an approach different to those used in the neural network domain, we hope to contribute to the evaluation of progress in neural networks forecasting.

Manuscript submitted May 10, 2007.

H. Njimi. CP 210 (bldg NO room 2.O.9.216), Campus Plaine, Université Libre de Bruxelles, Bd du Triomphe, B-1050 Bruxelles, Belgium (email: hnjimi@ulb.ac.be).

G. Mélard. ECARES CP 114 (Solbosch bldg S room S.11.131), Université Libre de Bruxelles, Avenue Franklin Roosevelt, 50, B-1050 Bruxelles,Belgium (email: gmelard@ulb.ac.be. homepages: http://homepages.ulb.ac.be/~gmelard/).

REFERENCES

[1] G.Mélard, J.-M. Pasteels, "Automatic ARIMA modeling including interventions, using time series expert software", *International Journal of Forecasting,* vol. 16, October 2000, pp. 497-508.

[2] H. Njimi, G. Mélard, J.-M. Pasteels, "User's Manual of Time Series Expert (TSE version 2.4) – TSE-AX: An expert system for ARIMA models", Institut de Recherche en Statistique - ECARES, Université Libre de Bruxelles, Bruxelles, March 2007, to be published.

[3] H. Njimi, G. Mélard, J.-M. Pasteels, "Modélisation SARIMA assistée", *Actes des XXVèmes Journées de Statistique, Lyon, France, 13-17 mai 2003*, Société Française de Statistique, Tome 2, pp. 731-734.

[4] W. H. Kruksal, W. A. Wallis, "Use of ranks in the one-criterion variance analysis", *Journal of the American Statistical Association*, vol. 67, 1952, pp. 401–412.

[5] G. Mélard, J-M. Pasteels, "User's Manual of Time Series Expert (TSE version 2.3)", Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles, Bruxelles, 1998.

[6] G. Mélard, "*Méthodes de prévision à court terme*", Editions de l'Université de Bruxelles, Bruxelles, and Editions Ellipses, Paris, 1990, ch. 8.