International Journal of Forecasting 16 (2000) 497–508

international journal of forecasting

www.elsevier.com/locate/ijforecast

Automatic ARIMA modeling including interventions, using time series expert software

G. Mélard*, J.-M. Pasteels

ISRO CP 210 (bldg NO room 2.O.9.300), Campus Plaine, Université Libre de Bruxelles, Bd du Triomphe, B-1050 Bruxelles, Belgium

Abstract

This article has three objectives: (a) to describe the method of automatic ARIMA modeling (AAM), with and without intervention analysis, that has been used in the analysis; (b) to comment on the results; and (c) to comment on the M3 Competition in general. Starting with a computer program for fitting an ARIMA model and a methodology for building univariate ARIMA models, an expert system has been built, while trying to avoid the pitfalls of most existing software packages. A software package called Time Series Expert TSE-AX is used to build a univariate ARIMA model with or without an intervention analysis. The characteristics of TSE-AX are summarized and, more especially, its automatic ARIMA modeling method. The motivation to take part in the M3-Competition is also outlined. The methodology is described mainly in three technical appendices: (Appendix A) choice of differences and of a transformation, use of intervention analysis; (Appendix B) available specification procedures; (Appendix C) adequacy, model checking and new specification. The problems raised by outliers are discussed, in particular how close they are from the forecast origin. Several series that are difficult to deal with from that point of view are mentioned and one of them is shown. In the last section, we comment on contextual information, the idea of an e − M Competition, prediction intervals and the possible use of other forecasting methods within Time Series Expert.

Keywords: M3-Competition; ARIMA models; Expert system; Intervention analysis; Outliers

1. Introduction

This article has three objectives: (a) to describe the method of automatic ARIMA modeling (AAM), with and without intervention analysis, that has been used in our analysis; (b) to comment on our results; and (c) to comment on the M3 Competition in general.

In 1970, Box and Jenkins (see Box et al., 1994), made ARIMA models popular by proposing a model building methodology comprising several stages: specification, estimation, diagnostic checking and forecasting. The justification for automatic ARIMA modeling is the following: (a) the method for building an ARIMA model is somewhat complex and requires a deep knowledge of the method; (b)

*Corresponding author.

E-mail addresses: gmelard@ulb.ac.be (G. Mélard), pasteels@intracen.org (J.-M. Pasteels).

0169-2070/00/$ – see front matter
PII: S0169-2070(00)00067-4

consequently, building an ARIMA model is often a difficult task for the user, requiring training in statistical analysis, a good knowledge of the field of application, and the availability of an easy to use but versatile specialized computer program; (c) the number of series to be analyzed is often large.

Starting with a computer program for fitting ARIMA models called ANSECH (Mélard, 1982), and a methodology for building univariate ARIMA models (Mélard, 1990), we started to build an expert system (Laforet, Mélard & Pasteels, 1990). The aim was to avoid the pitfalls of most existing software packages which act as black boxes, do not explain their reasoning and give neither intermediate results, nor guarantee for the final model (see Tashman and Leach, 1991, for a detailed criticism). Like others, we have made use of the technology of expert systems to build a univariate ARIMA model with or without an intervention analysis, leading to the development of software called TSE-AX. A first version of the system was integrated in Time Series Expert (TSE) 2.0 in 1993. An improved version included in TSE 2.3 is described here.

The expected performance of the system and the conditions under which it will do well should be the same as those observed for other competitions (see Fildes & Makridakis, 1995). For the M1-Competition, ARIMA modeling performed well (for some horizons) for quarterly series and for series with a weak noise component, see Pasteels and Mélard (2000). However, relatively poor performances were obtained for noisy series (micro-economic) and more generally for horizons 1 and 2.

The main part of the paper is subdivided into 5 sections. In Section 2, we summarize the characteristics of TSE-AX and more especially, the automatic ARIMA modeling method. Our motivation to take part in the M3-Competition is the subject of Section 3, and our methodology is described in Section 4. The problems raised by outliers and the series which are difficult to forecast in that respect are discussed in Section 5. We conclude with some final remarks in Section 6.

2. Automatic ARIMA modeling

In Section 1, we have mentioned several justifications for an automatic ARIMA model building procedure. They can be summarized by saying that using the Box–Jenkins method on a large scale requires both expertise and time. Aiming to make the method available to people without that expertise, several software vendors have implemented automated time series forecasting methods, e.g. Mandrake (C.E.M.S., Palaiseau, France, see Azencott, 1990), and the list in Tashman (2000). Ireland and Sharda (1988) have tested two of these programs using a collection of 111 time series drawn from the M1-Competition (Makridakis et al., 1984); see also Küsters (1995). These products are based on the technology of expert systems.

Our perception of a system for building ARIMA models is as follows. The system should work in an automated way, but the user should be informed of the steps, receive the intermediate and final results, and be informed of the quality of the final model. The system should be adapted to several categories of users from beginner to expert. The latter should use such a tool to save time, being qualified to assess the quality of the final model and possibly propose an alternative model.

One of the characteristics of our system is to consider only meaningful models for the lag structure concerned. Of course, the user may be able to improve the model given in an automated way by TSE-AX, especially if more information is available about the time series under study than can be used directly. For example, we may possess contextual information, explanations for extreme values, or changes in the definition of the variable. Like other systems, TSE-AX should be used for its convenience (it is faster than a step-by-step analysis) but not as a magic tool. It should shine when a large batch of series needs to be analyzed or when a moderate batch of series needs to be analyzed periodically. In these cases, the user simply has not the time to perform an analysis individually on each series.

Expert systems capture the expertise of a human expert and describe the reasoning used by the system, which gives them some pedagogical characteristics. The architecture of an expert system is mainly composed of:

- a *knowledge base*, which contains the different strategies of problem solving with their application conditions and is often represented as a sequence of production rules;
- a *facts base*, namely the data, contextual information and temporary results;
- an *inference engine*, the executing part of the expert system by which deductions are performed;
- an *interface module*, which describes the pieces of knowledge that have been used to derive the conclusion.

Often, an expert system is built starting from an 'expert system shell', a software tool that includes the inference engine and the interface module, for which it suffices to provide some expertise, under the form of a knowledge base, to obtain an operational expert system. We have not followed that approach. Although the knowledge base does exist (and excerpts are shown in the manual), TSE-AX takes the form of an executable program, provided with the data and additional information, that relies on another program for performing the statistical analysis (computing statistics, fitting models, evaluating forecasts).

Following Anderson (1977), the first step of an analysis is to gain some *familiarity* with the data. It seems difficult to automate that task. The information given to the system are the periodicity and the chosen sample. The user can then correct the data, eliminating part of the series if that step seems useful. The automated procedure starts at the beginning of the second stage, the *preliminary analysis*, where interventions are selected, transformations are performed, and differences are chosen to make sure that the series becomes stationary. The remaining stages are the *specification stage*, where an ARIMA model is identified, the *estimation stage*, where the model is fitted, the *model checking stage*, where the adequacy of the model is explored, the *forecasting stage* and the *interpretation stage*. If the model is rejected at the model checking stage, a new specification should be obtained, and so on, until an acceptable model is found.

Our system covers everything from the specification stage to the forecasting stage, given that the latter is immediate when a final model has been found. It does not explore all possible models and has a limited number of steps. The user can specify his/her model building preferences (performing an intervention analysis or not, choosing a specification strategy, etc.) through a command language which is described in the manual, thereby improving the portability and flexibility of the system. There are more than twenty input commands that enable the user to customize the modeling strategy. They concern the treatment of outliers by intervention analysis (number and type of shocks to accept, action to be taken at the forecasting origin, etc.), the seasonal component (stochastic or deterministic), the Box–Cox transformation (setting the significance level of the test) and difference operators.

The description of the main parts of the system is deferred to Appendices A (differences, transformation and intervention analysis), B (model specification) and C (adequacy, model checking and new specification).

3. Motivations

Our motivation for participating in the M3-Competition is described here. One of our objectives when building an automated system was to avoid some misuse of the Box–Jenkins (BJ) methodology that occurred during the M1-Competition (Makridakis et al., 1984). Indeed, the BJ methodology had been applied to a systematic sample of 111 series among the 1001 series. Therefore, it had been used to forecast short series (12 yearly data points sometimes). The asymptotic properties of the estimators are no longer valid in that context and this may lead to incorrect model selection. For example, Rossana and Seater (1995) show that, in the case of short annual series, random walk and IMA(1,1) processes are identified too systematically.

We also found it worthwhile to include some treatment of outliers in the ARIMA modeling procedures. Interventions were not considered in the ARIMA modeling approach of the M1-Competition. By running our own experiment on the M1-Competition set (see Pasteels and Mélard, 1999), we found that, for certain kinds of series (monthly macro-economic and industrial series), gains in forecasting accuracy can be obtained by ARIMA modeling with automatic intervention analysis.

Another problem in the M1-Competiton was that the forecast origins of the 1001 series were too concentrated around the mid-seventies (see Fig. 1), and also in the subset of 111 series that was used to test the accuracy of the BJ methodology. It is common knowledge that the seventies were characterized by unexpected shocks that affected the world economy. The M3-Competition presents a much more balanced distribution of forecast origins.

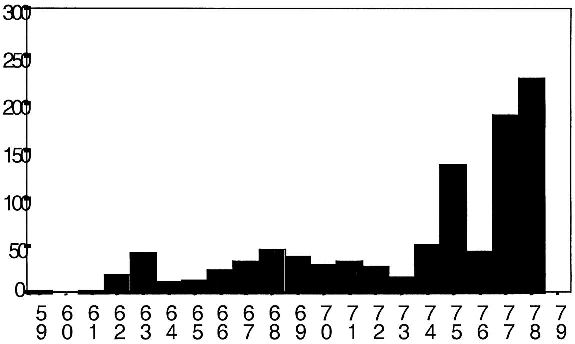

Fig. 1. Histogram of the forecast origins in the M1-Competition.

4. Methodology

Our analysis of the M3-Competition series is limited to quarterly and monthly series (2184 series out of 3003). Yearly series are discarded because most of them are too short to be modeled by ARIMA. It is better to put aside all yearly series in order not to corrupt the sample collected by the organizers. Series with unknown time intervals between successive observations are also excluded because our method requires that information.

We used the automatic ARIMA modeling system TSE-AX, described briefly in Section 2. Two modeling strategies have been applied. They are called respectively AAM1 and AAM2 and are described as follows:

- AAM1: automatic ARIMA modeling (no treatment of outliers);
- AAM2: same as AAM1 but with selective intervention analysis (for macro, industrial and demographic series, quarterly or monthly).

Our implementation of intervention analysis for AAM2 is outlined in Appendix A.3. For the two above-mentioned strategies, we used a high significance level for the transformation test; see Appendix A.2. By doing so, at least a squared root is applied to the series (and possibly a logarithmic transformation if needed). It is motivated by the fact that all data of the M3-

Competition are positive. Taking a squared root transformation guarantees that all the forecast values will be positive.

5. Checking for series difficult to forecast because of outliers

Let T be the number of data points and h, the forecast lead time. We have used the first $T-h$ observations in order to check if our forecasts for the last h values are consistent and to detect the series presenting potential difficulty to be forecast.

We also ran an outlier detection procedure on the complete series (with T observations) in order to assess if there are extreme values close to the forecasting origin, on observation number T, $T-1$, $T-2$, or $T-s$, where s is the seasonal periodicity. We have used the outlier detection procedure on the series after transformations and differences but before model specification. This procedure gives an indication of potential problems in computing extrapolative forecasts for some series. Some contextual information (what we called domain knowledge) is expected to help significantly to forecast those series. There are several examples in Pasteels and Mélard (1999) with series of the M1-Competition contaminated by shocks at the forecast origin. It has been found that for a few of them (the QNC23 series for example) the domain knowledge helped us to identify the nature of the shock and could reduce the forecasting errors drastically (for the QNC23 series the MAPE has been reduced from 125% to 20%).

We mention in Fig. 2 the percentages of series (for this Competition) which present shocks at the forecast origin or close to it. Of course, these percentages are somewhat overestimated because the outlier detection is done on a potentially autocorrelated series (see Chen & Liu, 1993, for a complete discussion). There might be some spurious outliers (redundant with some ARMA parameters). Nevertheless, this procedure gives an indication of the importance of the phenomenon.

Timing of the more recent extreme values	at T	at T or $T-1$	at T, $T-1$ or $T-2$	between T and $T-s$
Quarterly	1.4%	2.6%	4.5%	8.2%
Monthly	2.0%	3.4%	3.6%	13.6%
All	1.7%	2.9%	4.2%	10.6%

Fig. 2. Percentages of series of the M3-Competition presenting extreme values at the forecast origin or close to it.

Among others, special attention has to be given to series numbers 800, 802, 803, 819, 1122, 1195, 1465, 2340, 2359, 2479, 2725 and 2748, which all present a highly significant outlier (with a significance level less than 0.0001) at the forecast origin. Series 802 is a quarterly series of a micro-economic nature starting at 1984/II and ending at 1992/IV. A significant drop is registered for the last quarter of 1992 (see Fig. 3). The interpretation of this extreme value as either a regular observation, or an additive outlier, or an outlier on the innovations, or an effect due to a law or announcement would yield totally different forecasts. For example, Fig. 3 represents three different forecasts:

- (F1) those obtained by our automatic ARIMA modeling system, with intervention analysis, by assuming that the extreme value is due to an additive shock;
- (FEXP) those obtained by applying Holt's exponential smoothing, without intervention analysis, by assuming that no seasonal adjustment is needed and that the shock is of a permanent nature (i.e. the drop reflects the poor economic health of the firm concerned);
- (FJUDGE) those obtained by our 'poor' judgment in absence of more information, by

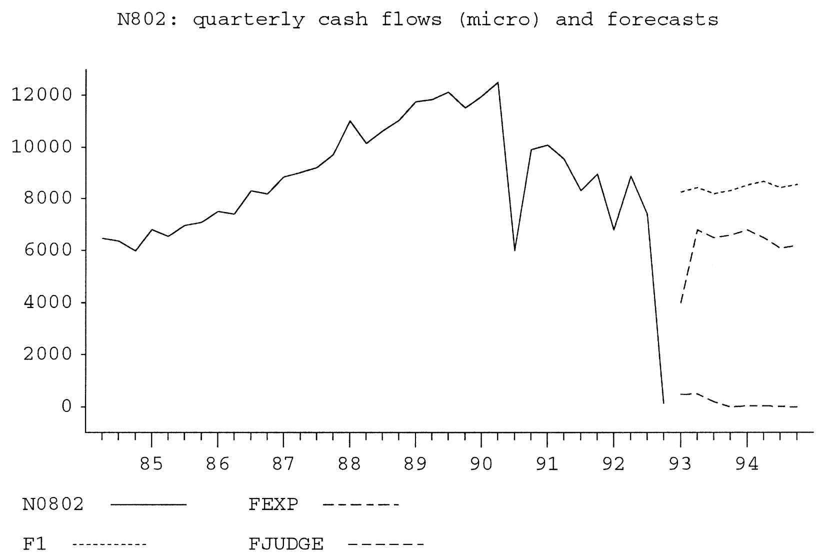

Fig. 3. The 802 series (quarterly cash flows of a firm) and forecasting scenarios.

proceeding by analogy and assuming that the drop registered for the last quarter 1992 is of the same nature as the one observed for the third quarter 1990; consequently, an increase of around 4000 units is expected for the next period.

In the framework of this competition, it is not fair to look for domain knowledge (so we did not) because it is above all a competition of extrapolative methods. As a result, our forecasts for series 802 are of course full of uncertainty. It would also be interesting for any participant in the competition to assess the performances of his/her method on this subset of '*problematic*' series.

6. Final remarks

6.1. Contextual information

When forecasting any time series (using an extrapolation method or not), all available contextual information (including information concerning possible shocks) should be used. By discarding the domain knowledge, extrapolative methods are compared on a sub-optimal information set. This point was discussed with several specialists (F. Collopy, B. Edmonds, W. Gor, F. Diebold, S. Moss) in the Internet forum concerning this competition.

6.2. e-M competition

With the opportunities offered by the Internet, a real-time competition dealing with contextual information could be organized at low cost. We leave that suggestion to the organizers and participants.

6.3. Prediction intervals

This competition, like the previous ones, is based on point forecasts accuracy. Prediction intervals are too often neglected in competitions. Even if the diversity of procedures to compute prediction intervals make it difficult to compare them (see Chatfield, 1993); it would be interesting to include their analysis in forthcoming forecasting competitions.

6.4. Other forecasting methods

Note that TSE 2.3 is not restricted to ARIMA modeling. A seasonal deterministic component is available, and also a polynomial trend. Furthermore, most basic forecasting methods including simple moving averages and exponential smoothing are also implemented. A large class of 27 methods of exponential smoothing methods is included, combining 7 types of level and trend evolution, including damping, with 4 types of seasonal components, including normalization of seasonal. The technology described by Broze and Mélard (1990) is used. Hence, the parameters are estimated by exact quasi-maximum likelihood, 1-step ahead forecast errors are analyzed and forecast intervals are produced.

Acknowledgements

We thank the Ministry of Education of the 'Communauté Française Wallonie–Bruxelles' which has helped us through two research grants. At the beginning of this project, we received a grant 'Fonds de la Recherche Fondamentale et Collective' (FRFC-IM). More recently we have benefitted from a grant 'Action de Recherche Concertée' 96/01/205. We thank Michèle Hibon and Keith Ord for their judicious remarks and Spyros Makridakis for his comments on the thesis of the second author.

Appendix A. Preliminary analysis

During this phase, the choice of a difference or a combination of differences is made, a possible transformation of the data (power, logarithms) is considered, and an (optional) intervention analysis is performed. The order of *selecting* these steps is of fundamental importance. If everybody would agree to choose the difference(s) before the Box–Cox transformations, it is not sure that intervention analysis at this stage of the analysis would be of common use. Of course, interventions are first *applied* to the data, then the transformation, then the differences. In the description of our strategy, however, the *selection* order of these steps is followed rather than the *application*.

A.1. Choice of a difference and a seasonal difference

In order to make the series stationary as far as the level is concerned, the traditional approach is to analyze plots, compare variances, and to look at the autocorrelation function $\{r_k\}$ and the partial autocorrelation function $\{p_k\}$ of the series $\{y_t\}$, $\{\nabla y_t\}$, $\{\nabla_s y_t\}$, $\{\nabla\nabla_s y_t\}$, $\{\nabla^2 y_t\}$). The choice of a difference is not always obvious but is however important for the sequel. A graphical analysis is not feasible in an automated system. Inverse autocorrelations are recommended by several authors, including Wei (1990). In our automatic procedure, we have chosen to use a criterion of over-differencing based on autocorrelations as well as a test for seasonality: we use the non-parametric test of Kruskal and Wallis, 1952, where the statistic KW is computed using the ranks of the observations.

The criterion of over-differencing is justified as follows. Let us compare some autocorrelation coefficients before and after differencing. If ∇y_t has a lag 1 autocorrelation coefficient larger than that of y_t, we may guess that there is over-differencing. For the seasonal difference, we use the lag s autocorrelation coefficient. We have implemented two numerical criteria to replace the delicate analysis of correlograms performed by the human eye. Those criteria, CACF for the autocorrelation coefficients and CPACF for the partial autocorrelation coefficients, are functions of the most significant coefficients among the L first lags that should be minimized.

We use four heuristics, see Mélard and Pasteels (1998) for details and formulas. Heuristic 1 which makes use of the Kruskal and Wallis test and CACF deals with seasonal differences and over-differencing. Two conditions are necessary for the choice of a seasonal difference: a positive seasonality test as well as no indication of over-differencing. The test for seasonality assumes however that the seasonal profile is stable across time. There exist several cases where the two conditions may not be fulfilled simultaneously:

1. the series shows a regular seasonal profile but with a small amplitude;
2. the series shows an irregular seasonal profile but with a large amplitude;
3. the series shows irregular seasonal variations and with a small amplitude.

In the first case, the condition of over-differencing is often not fulfilled. We recommend not taking seasonal differences but would inform the user of an apparently conflicting situation. In the second case, the Kruskal–Wallis test may not be significant at the 5% level; if that test indicates presence of a seasonal component with a less strict significance level (say, 10%) and there is no over-differencing, we recommend choosing a seasonal difference operator. In the third case, the two tests give the same type of information (absence of seasonal difference).

Heuristic 2 deals with a regular (i.e. non seasonal) difference. Unit roots tests could have been considered here but the risk of a bad specification is high. We preferred to select from among the series $\{y_t\}$, $\{\nabla y_t\}$, and $\{\nabla^2 y_t\}$, taken in that order, the last one for which there is no over-differencing in the sense mentioned above.

Heuristic 3 only plays a role for smooth (probably artificial) series, not very frequent in the real world. There is no value in finding the data generating process of those series but it is good to inform the user. Heuristic 4 allows to check if both CACF and CPACF criteria agree for the choice of differences.

Other pieces of information are also displayed: for each series, the number of $\{r_k\}$ or $\{p_k\}$ which are significantly different from 0 and the number of these for which no explanation can be found. We consider lags for which an explanation can be found to be 1, 2, 3, 4 and also $s, 2s, 3s, \ldots$ (where s represents the seasonal period).

Remark. *Several authors* (*Hylleberg*, 1992 *and Franses*, 1996) *propose a seasonal unit root test which helps to discriminate between a deterministic and a stochastic representation of seasonality. Makridakis and Hibon* (1997) *observe that for the collection of series of the M*1*-Competition, a global gain of accuracy of the forecasts is obtained by using a deterministic representation of the seasonality. We have therefore implemented the latter within the methodology of Box–Jenkins in TSE-AX, but it was not used in the M*3*-Competition.*

A.2. Choice of a transformation

A stationarity test of the series for the variance (possibly after differentiation) is performed. The test is based on the rank correlation coefficient τ of Kendall (Kendall & Ord, 1990) computed between a scale parameter of the residual time series (after differences have been applied) and time. The scale parameter computed for each year is the interquartile range I50 = Q75 − Q25, where Q75 and Q25 represent, respectively, the first and third quartiles. If the test of $\tau = 0$ is rejected, a square root transformation is performed. The test is done again on the new residual time series (after the square root transformation and differences have been applied). In the case of rejection, a logarithmic transformation is performed. The method has the advantage of being resistant against outliers. The level of significance of the test can be selected by the user.

A.3. Intervention analysis

At this stage of building the time series model, we have determined if the series needs to be differenced or/and transformed. An intervention analysis, performed at the user's request, may be employed to avoid extreme values that would influence the next steps of the analysis: specification, estimation, test for adequacy, and forecasting. The justification for performing an intervention analysis at this stage is given by Pasteels and Mélard (2000).

It is customary to distinguish several types of intervention: 'pulse' interventions or additive outliers (AO), 'level shift' (or LS), 'compensation', 'ramp', 'temporary change' (TC), and 'innovation outlier' (IO). The ANSECH program used in the study includes all except innovation outliers. In Pasteels and Mélard (2000), we describe how pulses and level shifts are detected and how their dates are chosen.

Appendix B. Specification

This is usually the step which is the most difficult. At the beginning, our objective was to formalize the specification method described by Mélard (1990) which is based on the analysis of autocorrelations. The '*expert*' strategy which has been implemented differs, however, from the approach proposed initially by Box and Jenkins in the sense that no attempt is made to identify the functional form of the autocorrelation and partial autocorrelation functions except for truncation and that only plausible lags are considered. Nevertheless, the first results have shown that models are often underparametrized, leading us to develop two additional specification methods which are described here. They are called: '*autoregressive specification procedure*' and '*mixed*' respectively. For more details on the '*expert*' strategy, see Mélard and Pasteels (1998).

B.1. Autoregressive specification procedure

This procedure was initially used in a multivariate framework by Tiao and Box (1981) and is related to the concept of extended autocorrelation (Tsay & Tiao, 1984). It has been applied to the univariate case by Mélard (1990) under the name of the *autoregressive specification method*. It consists of fitting successively the parameters of autoregressive models of order 1, 2, and so on, and then examining the autocorrelations of the residual series. If for some p the fit of an AR(p) model leads to a residual series whose autocorrelations are roughly truncated above lag q, that residual series can be represented by an MA(q) model, and an ARMA(p, q) model can be fitted to the original series. That approach appears to be more fruitful than the corner method (Gouriéroux & Monfort, 1990) and similar methods. Those methods do not allow easy specification of seasonal models, and lead to models with terms of higher lag, which cannot be interpreted easily.

We have extended the autoregressive specification procedure to seasonal models using a two-stage method. In the next paragraph, 'AR(p) and SAR(1)' denotes a multiplicative AR scheme involving a regular autoregressive polynomial of degree p and a seasonal autoregressive polynomial of degree 1 in B^s. Also 'SMA(1)' means a seasonal moving average polynomial of degree 1 in B^s.

During the *first stage*, 8 pure autoregressive models are fitted (4 models in the non-seasonal case): (1) white noise; (2) AR(1); (3) AR(2); AR(3); SAR(1); AR(1) and SAR(1); AR(2) and SAR(1), AR(3) and SAR(1).

During the *second stage*, the residual autocorrelations are analyzed for the four or the eight models fitted during the first stage. The model showing the smallest number of significant r_k is selected. The procedure stops if there is no significant r_k for the selected model. Otherwise, let k be the largest lag for which r_k is significant; moving average terms of order 1 to the minimum of k and K, where K is the maximum lag specified by the user as being explainable. A term SMA(1) is also added if $r_k$are significant for k multiples of s. Finally if, for a selected model, there remains some significant r_k but only for k not explainable, use of the mixed strategy is recommended.

B.2. Mixed strategy

This strategy is the most complete. It has been implemented to circumvent the pitfalls of the 'expert' strategy and the autoregressive specification procedure; for the latter, it is not always obvious which model to select at the second stage. This strategy combines the first two approaches, hence its name. It requires, however, fitting a larger number of models. Its use assumes that enough computation power is available. It is composed of three steps:

1. fitting of eight autoregressive models (four in the non-seasonal case);
2. for *each* autoregressive model, possible addition of moving average term(s) and fitting of eight ARMA models;
3. selection of the final model among these mixed models.

The first step is the same as in the autoregressive specification procedure. For the second step, the choice of the MA terms is done for each model, not just for the model showing the smallest number of significant autocorrelations. Valid models are retained on the basis of the Ljung–Box Q statistic of order m (where m is a function of n, the length of the series), denoted by Q_m. For the third step, among all the valid models, the final model is the one which minimizes the SBIC criterion, since that criterion gives a heavy penalty to over-parametrized models. If no model is considered as being valid, the choice falls on the model minimizing Q_m.

Appendix C. Adequacy, model checking and new specification

C.1. Adequacy and model checking

At this stage, first check whether or not the optimization process for the likelihood function has converged. Then, the roots of the AR and MA polynomial should be analyzed with respect to the stationarity and invertibility conditions. This is always true in ANSECH when the optimization has converged because these conditions are imposed as constraints during model fitting. Then, the residuals are analyzed:

- tests for adequacy (based on the Ljung and Box statistic, Q_m) are performed for several lags ($m=6, 12, 18, 24, \ldots$ for monthly series);
- the r_k and p_l of the residuals are tested individually with respect to the Bartlett limits;
- the extreme values of the residuals are detected.

We have defined two criteria of model adequacy:

First criterion (*valid model*)

- The Ljung–Box tests at the different lags m ($m=6, 12, \ldots,$ up to a maximum $\leq n/5$) should not reject the null hypothesis (at a significance level of 5%).
- There should be no more than 3 significant coefficients r_k, p_k.

Second criterion (probably valid model)

- The Ljung–Box tests at the different lags m ($m=6$, 12, . . . , up to a maximum $\leq n/5$) should not reject the null hypothesis (at a significance level of 5%).

Introduction of a mean is also tested.

C.2. New specification

Contrary to strategies proposed by other authors, we stop at this stage, which means that we are happy with a specification procedure of one, two or three steps (according to the chosen strategy). To keep things simple, we have avoided consideration of multiple models. The management of a list or better a tree structure of possible models can become extremely complex and there is a risk that the procedure loops indefinitely while looking for the best specification. Furthermore, it is not wise to build a model using a large number of statistical tests.

That option can be justified by the fact that we do not simply want to find a model at any price. We also seek to avoid models based on high-order polynomials. The strategy which has been developed aims at being a first, fast, automated approach, allowing the user to be productive. The final interpretations and the refinements are always supposed to be performed by the user. We do not attempt to suppress entirely a human intervention. On the contrary, the objective is to perform a careful automated analysis which is also aware of its limitations. The user needs to be informed on possible improvements of his/her model.

References

Anderson, O. D. (1977). Time series analysis and forecasting: a further look at the Box–Jenkins approach. *Cahiers du Centre d'Etudes de Recherche Opérationnelle 19*, 223–256.

Azencott, R. (1990). Modélisation ARIMA automatique: le système MANDRAKE. *Cahiers du Centre d'Etudes de Recherche Opérationnelle 32*, 229.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis forecasting and control*, 3rd ed, Prentice Hall Press, Englewood Cliffs, N.J.

Broze, L., & Mélard, G. (1990). Exponential smoothing: estimation by maximum likelihood. *Journal of Forecasting 9*, 445–455.

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics 11*, 121–135.

Chen, C., & Liu, L. -M. (1993). Joint estimation of model parameters and outliers in time series models. *Journal of the American Statistical Association 28*, 284–297.

Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review 63*, 289–308.

Franses, P. H. (1996). *Periodicity and stochastic trends in economic time series*, Oxford University Press, Oxford.

Gouriéroux, C., & Monfort, A. (1990). *Séries temporelles et modèles dynamiques*, Economica, Paris.

Hylleberg, S. (1992). *Modelling seasonality*, Oxford University Press, Oxford.

Ireland, T. C., & Sharda, R. (1988). *A test of Box–Jenkins forecasting expert systems for microcomputers*. Working paper, Oklahoma State University, Department of Management (unpublished).

Kendall, M. G., & Ord, J. K. (1990). *Time-series*, 3rd ed, Arnold, London.

Küsters, U. (1995). AUTOBOX 3.0 and SCA PC-EXPERT, A review of two systems for automatic ARIMA forecasting. *The Forum, Joint Newletter of the International Association of Business Forecasting and the International Institute of Forecasters 8*(2), 1–6.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in the one-criterion variance analysis. *Journal of the American Statistical Association 67*, 401–412.

Laforet, A., Mélard, G., & Pasteels, J. -M. (1990). Vers un système expert de prévision et de statistique économique. *Cahiers du Centre d'Etudes de Recherche Opérationnelle 32*, 285–302.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1984). *The forecasting accuracy of major time series methods*, Wiley, Chichester.

Makridakis, S., & Hibon, M. (1997). ARMA models and the Box Jenkins methodology. *Journal of Forecasting 16*, 147–163.

Mélard, G. (1982). Software for time series analysis. In: Caussinus, H., Ettinger, P., & Tomassone, R. (Eds.), *Compstat 1982, Part I, Proceedings in Computational Statistics*, Physica-Verlag, Vienna, pp. 336–341.

Mélard, G. (1990). *Méthodes de prévision à court terme,* Bruxelles: Editions de l'Université de Bruxelles, and Paris: Editions Ellipses.

Mélard, G., & Pasteels, J. -M. (1998). *User's manual of Time Series Expert (TSE version 2.3),* Institut de Statistique, Université Libre de Bruxelles.

Pasteels, J.-M., & Mélard, G. (2000). *Automatic ARIMA modeling and intervention analysis, empirical findings,* submitted.

Rossana, R. J., & Seater, J. J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics 13*, 441–451.

Tashman, L. J. (2000). Out-of sample tests of forecasting accuracy: a tutorial and review. *International Journal of Forecasting 16*, 437–450.

Tashman, L. J., & Leach, M. L. (1991). Automatic forecast software: a survey and evaluation. *International Journal of Forecasting 7*, 209–230.

Tiao, G. C., & Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association 76*, 802–816.

Tsay, R. S., & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association 79*, 84–96.

Wei, W. W. S. (1990). *Time series analysis: univariate and multivariate methods,* Addison-Wesley, Redwood City.

Biographies: Guy MÉLARD is Professor of Statistics and Computer Science, Faculté des Sciences sociales, politiques et économiques, and Institute of Statistics and Operations Research, Université Libre de Bruxelles, Belgium. He has published articles in the *Journal of Time Series Analysis*, *Journal of the American Statistical Association*, *Journal of Forecasting*, *IEEE Transactions on Signal Processing*, and *Annals of Statistics*. He has written a book on short-term forecasting methods and has developed many algorithms for computational statistics. His research interests include theory and applications of time series analysis and forecasting.

Jean-Michel PASTEELS was Teacher Assistant at the Department of Applied Economics (DULBEA), Université Libre de Bruxelles, where he obtained a PhD in Economics. He has published articles in the *Cahiers du Centre d'Etudesde Recherche Opérationnelle* and *Cahiers Economiques de Bruxelles*. He has developed, in collaboration with the first author, an expert system for time series analysis. His research interests include theory and applications of time series analysis, economic forecasting, expert systems and international trade theory. He is presently 'Trade Analyst' with the United Nations in Geneva, Switzerland.