

This review originally appeared in the Forum, the joint newsletter of the IABF and the IIF. Summer 1995.

The Forum, Summer 1995

AUTOBOX 3.0 and SCA PC-EXPERT: A Review of Two Systems for Automatic ARIMA Forecasting

by Ulrich Kusters
Department of Business Administration
Research Unit for Statistics and Operations Research
Catholic University of Eichstat
EMail Ulrich.Kuesters

1. General Remarks

AUTOBOX 3.0 from Automatic Forecasting Systems (AFS) and PC-EXPERT/XUTS from Scientific Computing Associates - hence referred to as SCA PC-EXPERT- are expert systems which can be used to manually or automatically forecast univariate time series on the basis of the Box-Jenkins methodology. In addition, both programs can model exogenous regressors and outliers via transfer functions.

Both developers have considerable experience in designing forecasting systems. AUTOBOX's David Reilly has worked for more than 20 years on automatic ARIMA modeling and has written several papers about ARIMA modeling strategy. The same is true for Lon-Mu Liu, one of the principal developers of SCA's time series subsystems. [A list of references can be obtained by sending an E-mail to the author or software editor.]

The distinct differences in the design philosophies of AUTOBOX and SCA PC-EXPERT make these packages attractive to different audiences. The base system XUTS of SCA PC-EXPERT, which stands for Extended Univariate Time Series, offers sophisticated analytical tools which can be recommended to teachers, academic researchers and forecasters who usually analyze a hand-full of time series with great care. AUTOBOX, in contrast, offers an integrated data management tool and total automation of the modeling and forecasting tasks. Therefore, it can be recommended to practitioners who must process large amounts of data automatically.

2. Systems Description and Documentation

AUTOBOX embeds a forecasting engine built on the central modeling steps of the Box-Jenkins paradigm. This core is extended by several useful modules,

such as a configuration editor, a data management module, evaluation statistics and a report generator, all of which expedite the forecasting practitioner's tasks.

The complete system for DOS is shipped on two high-density disks for \$1195. This non-copy protected system can process up to 300 time series for univariate forecasting and up to 5 input variables in a multivariate input model. More limited versions are available at lower cost. A Windows version is under development.

SCA PC-EXPERT is one of a family of MS-Windows-based programs offered by SCA. Along with XUTS and SCAGRAF, a high-resolution graphic subsystem, the system is offered on seven high density disks for \$1395. This copyprotected system is restricted to 40000 memory words, but there is no specific limit in terms of the number of series, observations or model parameters. Other members of the SCA family include modules for statistical quality control, multivariate analysis, seasonal decomposition (essentially X11) and econometrics. These have not been included in this review AUTOBOX and SCA PC-EXPERT have been reviewed in the past by several authors - see references at end of review.

AUTOBOX is shipped with a 240 page reference manual/users guide which is cumbersome to read, principally because it has been written with an editor that uses only the ASCII 8-bit character set. This document, moreover, does not cover recent extensions introduced into Version 3. There is, however, extensive on-line documentation that is up to date and yields informative, context-sensitive help messages.

SCA documentation includes a large reference manual covering the time series/forecasting tools and three small reports which describe the expert system, graphics facility and Windows-specific features. I found the detailed documentation to be well-written and to contain many examples.

Lag operator notation is used throughout both manuals as well as in the screen outputs. Most business practitioners will run into difficulty trying to read these outputs. An untrained user can simply ignore all output but the forecasts; however, there is no guide which says which information can be safely ignored.

3. User Environment and Data Management

3.1 SCA PC-EXPERT

The program opens up with three panels - an output window, the SCA modeling menu and the SCA command window, as shown in Figure 1. Commands can be given in two ways. If you simply type a command in the command window, you'll see immediate results in the output window. Alternatively, you may select entries from the SCA menu panel.

The automatic identification and estimation directives can be invoked by typing in the IARIMA (identification) command together with relevant options or by selecting Automatic ARIMA Modeling in the SCA menu. The latter opens a submenu, in which the user can fill in items like the seasonality period, the order of differencing and so on. Unfilled items are either left empty and/or preset with presumably suitable defaults. Pressing the available help button yields a help screen with the corresponding paragraph syntax. Suitable examples and context sensitive help screens are not available.

High resolution graphs cannot be produced directly. However, there is an export facility for numeric data which can be read by SCAGRAF, an accompanying graphic program. This system produces neat graphs which can be cut and pasted using the MS-Windows clipboard or saved to a BMP or TIF graphic file.

The SCA indexing scheme for time series uses only observation indices, which makes it cumbersome to manipulate the data and the forecasts. Only SCAGRAF has a command to provide calendar information (week, month, quarter, year) in the plots. Data series can be read in various formats, including the SCA specific-file format.

SCA PC-EXPERT offers a macro facility which contains a rather large subset of matrix and statistical functions as well as some flow control statements like IF-THEN-ELSE. This macro facility can also be used to program forecast comparisons.

The SCA system does not include an internal data base facility, possibly limiting its appeal to practitioners who must process large amounts of data. However, the SCA macro facility may be used to transfer data and command switches between PC-EXPERT and other applications such as EXCEL and ORACLE via ASCII files. According to SCA, several of their customers have done that successfully. The new version, which is under development, will simplify this communication task considerably by adding DDE/OLE and ODBC support.

Data can be saved and recalled from session to session by the WORKSPACE paragraph. Unfortunately, this feature cannot be used for command switches.

3.2 AUTOBOX

AUTOBOX opens up with a functional menu system that asks you to choose between three user levels - novice, intermediate and advanced. All levels allow for completely automatic modeling even with exogenous regressors; however, moving from the novice to the intermediate and advanced opens up more option switches. Determining appropriate settings for the options, however, may require advanced statistical knowledge. On the other hand, the novice menu does not permit the user to choose the forecast horizon, which is something any novice should be able to do. All menu items are organized in a hierarchical way.

A newcomer needs time to become familiar with this kind of menu system; however, with experience, it turns out to be a fast way to give instructions.

AUTOBOX is a configurable system. The menu system writes files which contain parameters and input data for the computations. Configuration states can be easily recalled for subsequent sessions by rereading the command files. Because utilities to simplify the administration of different profiles are not available, different versions of the profiles can be administered only with DOS commands.

Several configuration screens in the intermediate menu, in the advanced menu and in the engine setup allow the specification of databases, time series, time and forecast ranges, analysis options and reporting options which can be used to define defaults for novices and statistically untrained users. The observations can be indexed not only by number but also with major/minor indices (usually year/month). This simplifies greatly the time reference of observations and forecasts.

The data base is a rather unique feature. This subsystem allows a direct association of the time series and associated forecasts, fitted values, confidence bounds for the forecasts, associated input series and outlier series. This feature is very useful for analyzing and forecasting a great number of time series. ARIMA models and transfer function components can also be associated with each individual time series.

Second, the AUTOBOX data base can attach user-defined attributes (fields) like brand names, sales regions and packing types to each series. This feature helps the user to analyze and forecast several related series (like aggregate components) in an organized way. Moreover, the data base allows statistical functions like means, sums and minima to be defined on actual and forecasted values, permitting the user to easily define new forecast evaluation formulas, forecast aggregates and similar quantities, all without macro programming.

The graphical features of are AUTOBOX are out-of-date. First, the user is restricted to VGA mode. Second, time series plots can be printed out only on HP-PCL compatible printers using the screen dump utility HPPS. Presentation graphs must be produced elsewhere; however, data import/export facilities expedite this task. Hopefully, the windows version, AUTOBOX 5.0, which is under development, will remedy the graphical shortcoming.

3.3 Hardware/Operating System

The tests of these programs were carried out on a Toshiba laptop T3600CT, with a 80486DX50 processor, 16 MB RAM and a 250MB hard disk, operating under DOS 6.2 and Microsoft Windows for Workgroups.

Both systems occasionally failed, giving unreproducible run-time errors, usually when asked to do something unusual. When operating in a standard fashion, however, both systems were stable.

Common to both systems are their capabilities to produce huge amounts of output. This output can be browsed with different tools but, neither program makes it easy to restrict the output to sensible subsets, that can be retrieved in an organized way. This is a problem which is common to nearly all statistical systems. In AUTOBOX, several output files became so large that the DOS-editor gave the message 'insufficient memory'. Note that both systems offer a way to generate a screen 'snapshot'.

4. Statistical Features

Both systems consider the same model class, namely the seasonal autoregressive integrated moving average model with dynamic regressors (transfer components), abbreviated as SARIMAX(p,d,q)(P,D,Q)s[X]

Nevertheless the identification, estimation and outlier diagnosis issues are attacked in rather different ways. Unfortunately neither manual describes its statistical modeling strategy (rules) in sufficient detail. Hence, the description must remain a bit sketchy.

4.1 SCA PC-EXPERT

Univariate modeling (no exogenous regressors) is implemented in SCA PC-EXPERT as a black-box command. The rule-based expert system utilizes a version of the linear filtering method developed by Lon-Mu Liu (1989). This approach employs tentative models to represent the regular and seasonal component, whose autocorrelation structures are identified separately. The combination of these models yields a new, tentative model. According to the manual, the ordinary, partial and extended sample autocorrelation functions are used to identify the proper ARMA structure. But we are not told precisely how these tools are applied in an automatic way.

An important gap is the lack of suitable tests for the determination of the differencing order. Neither unit root tests nor seasonal unit root tests have been implemented yet. The system relies on a direct estimation method which evaluates the significance of the estimated AR(1) parameter in a filtering model. When the upper bound based on a 95-percent confidence test of the estimated Phi parameter exceeds unity, differencing is taken.

In contrast to univariate modeling, the influence of exogenous regressors cannot be automatically modeled. Therefore, the user must identify the transfer

components manually, a major limitation for the novice. (SCA does make automatic transfer function modeling available in its version for UNIX machines.)

One can use what is called the HOLD sequence to save the disturbance part of the model, whose ARIMA-structure can then be identified. After that the user can combine both models into a transfer function model with the identified error model. Unfortunately, this procedure may induce biases into the analyses due to its sequential nature.

SCA PC-EXPERT does not record intermediate results either in a file or audit trail. Therefore the reliability of its identification strategy can be checked only indirectly by an evaluation of the structure of the estimated residuals and the accuracy of the implied forecasts.

In principle, PC-EXPERT's macro facility can be used to extend the automatic ARIMA identification procedure into an automatic forecasting system. To do this, a user requires not only a thorough knowledge of the Box- Jenkins methodology and some SCA macro programming experience, but also some experience in designing data analysis strategies. The user's knowledge should also cover suitable techniques for trapping numerical run-time errors, avoiding infinite identification cycles and so on. Beyond the automatic identification modules, there are several highly sophisticated procedures to detect and handle outliers, be they additive outliers, innovational outliers, level shifts or temporary changes. For temporary changes the declination parameter (the parameter of the denominator polynomial) may be set by the user (default decay parameter is 0.7). This may be difficult for nonexperts.

Noteworthy is the fact that estimation and outlier detection can be handled simultaneously. This is exceptional because most other programs, including AUTOBOX, handle these issues sequentially. Exceptional as well is the program's ability to estimate calendar effects, including trading day and Easter effects, which are very important for sales and production forecasts. But neither the outlier diagnostics nor the calendar effects are integrated within the automatic identification module. Therefore they must be implemented by the forecaster.

Some exponential smoothing procedures are also provided; however, the smoothing parameters must be preset because they cannot be optimized by these procedures. As it is well known, some very simple smoothing models are equivalent to unrestricted ARIMA-models while others are restricted submodels of SARIMA-models. But estimating smoothing constants via ARMA models is unnecessary and far too complicated.

4.2 AUTOBOX

For the end-user, the simple selection of the dependent variable, the candidate set of regressors and the time ranges suffice to produce automatic forecasts. The

way this is done is controlled by 56 switches within the submenu, Change analysis defaults, shown in Figure 4.

Switches control the extent of automatic identification steps, such as the use of Box-Cox-transformation for variance stabilization, the setting of I-value thresholds, the way how to handle differences to achieve stationarity and so on. Despite the on-line-help, it is not always easy to understand the meaning of these switches.

The results of the different alternatives can be evaluated by setting all 45 switches governing 'output report options' to yes. This yields huge amounts of output, but reveals at least some part of the inherent modeling wisdom of the program. In doing so, I found that AUTOBOX uses several advanced and clever techniques to identify suitable forecast models, including necessity and sufficiency tests for the parameters, a stability check of the ARMA polynomials (stationarity and invertibility), and a recently developed test by Franses to discriminate between seasonal unit roots and seasonal dummy model components.

Surprisingly, no simple unit root test along the lines of Dickey and Fuller, which could be used to discriminate between deterministic and stochastic trends, has been implemented as a strategic part of the model identification step. Only a simple Dickey-Fuller-test without a deterministic trend is provided as a purely descriptive tool.

Outlier handling is quite different from that in the SCA program. Only the handling of additive outliers and level shifts coincide in both systems. AUTOBOX does not include the choice of innovational outliers; however, it does recognize several other types of outliers which cannot be handled by the other program. One of these is the seasonal pulse, which can be very useful to model seasonal patterns which occur only in a few months (like christmas effects in December).

AUTOBOX can detect dynamic patterns such as transient changes in the data. (Level shifts in the data can be detected in both systems as additive outliers in the first differences.) It is also able to detect deterministic trend changes, which are increasingly used by econometricians, but whose relevance for forecasting has not yet been proven.

Different types of exponential smoothing schemes also can be estimated. Moreover, the parameters of the smoothing procedures are automatically optimized, an improvement over SCA PC-EXPERT.

More significantly for ARIMA users, AUTOBOX extends its automatic modeling to the exogenous regressor cases. According to the manual, it offers two model identification options. The first technique is the well known prewhitening method developed by Box and Jenkins. As an alternative the user can select the

common filter/least squares technique. But neither the documentation nor the help screens tells the user how the system selects the numerator/denominator polynomial orders for the rational transfer function weights. It might be by the corner method but this is only a conjecture. The same is true for the selection of the applied common filter (The factors which are nearest to 1.0?).

AUTOBOX permits the user to easily evaluate the forecast accuracy of the identified model. First, the user can withhold some observations for forecast evaluation using an out-of-sample-analysis. Second, different forecast evaluation measures including the mean error (bias) and mean absolute percentage error (MAPE) can be computed for different forecast horizons without any additional programming. Moreover, the user can select either the minimization of the forecast error within the sample or within the withheld observations as a model selection device (but not for model estimation). The user of SCA would have to write a macro for out-of-sample evaluations.

5. Conclusions

SCA PC-EXPERT/XUTS contains very useful and sophisticated analysis tools. Some of these tools, however, havenot been integrated yet into a coherent data analysis and forecasting strategy. Data management to handle different but related series and their forecasts is not available. As a result, many steps in the forecasting process must be managed by the knowledgeable user, who knows a lot about Box-Jenkins time series analysis and who has time to administer the results. For the practicing forecaster in industry who must process large amounts of data under heavy time and budget constraints, this package is probably too inconvenient. Improvement may come, however, with the next release of SCA PC-EXPERT, which will contain DDE/OLE and ODBC support. Such features should allow the integration of the program's forecasting tools into spreadsheets and data bases.

For academic users the situation is very different. Academic researchers who usually analyze a handful of time series with great care and in great detail will be able to accomplish their tasks readily. For this group, the expert system component should play at most the part of a data analysis support tool. For researchers and especially for students the entire SCA system may be a serious alternative to other statistical packages.

Due to the high integration of the different time series analysis tools into an integrated framework as well as to its strong data management tools AUTOBOX 3.0 can be strongly recommended to practitioners who must handle many data sets. Practitioners with only little knowledge about Box-Jenkins models can use the program as a pure black-box tool, assuming some knowledgeable user with experience in forecasting methods has configured the system to match the specific features of the data and the specific needs of the end user.

Without the support of an experienced forecaster, a novice may unwittingly run into pitfalls. But this is probably true for any sophisticated forecasting program, not just this one. I believe that the fruitful use of AUFOBOX 3.0 requires a basic forecasting course which covers at least the methodological issues which underlie the BoxJenkins methodology, including transfer functions and outliers. SCA PC-EXPERT requires even more knowledge.

In teaching, with the exception of advanced classes, AIJTOBOX cannot be recommended, mainly due to its inadequate documentation and the black box character of several routines. The user interface and graphical facilities we hope will improve with the arrival of the MS-Windows version.

Addresses

SCA PC-EXPERT
Scientific Computing Associates,
913 Van Buren St., Ste. 3H,
Chicago, IL, 60607-3528,
USA
Tel: 312-455-0222
, Fax: 312-455-1652

AUTOBOX 3.0
Automatic Forecasting Systems
PO Box 563
Hatboro, PA, 19040
USA
Tel: 215-675-0652
Fax: 215-672-2534

Acknowledgments

I am obliged to Gerhard Armingier and Carlo Bianchi for their critical remarks. Especially, I would like to thank the software review editor of The Forum, Leonard Tashman, for his tireless efforts in editing this publication. Obviously any faults and omission are my responsibility.

References: Other Reviews and Applications

Tim C. Ireland and Ramesh Sharda, A Test of Box-Jenkins Forecasting Expert Systems for Microcomputers, Department of Management, Oklahoma State University, College of Business Administration, Stillwater, OK 74078, Phone: (405) 744-8642, (405) 744-8638.

John C. Picket (1991), AUTOBOX 3.0, International Journal of Forecasting, 7:395-398.

John C. Picket (1991), AUTOBOX 3.0, OR/MS Today, pp. 32-35, April 1991.

Lon-Mu Liu (1989), Identificatzon of SeasonalARIMA Models using a Filtering Method, Communications in Statistics, 18:2279-2288.

J. Keith Ord (1986), AUTOBOX - Software Review, International Journal of Forecasting, 2:511-513.

R.H. Shumway (1986), AUTOBOX (Version 1.02), The American Statistician, 40(4):299-300.

Leonard J. Tashman and Michael L. Leach (1991), Automatic forecast software: A survey and evaluation, International Journal of Forecasting, 7:209-230.

Pamela A. Texter and J. Keith Ord, (1989), Forecasting using automatic identification procedures: A comparative analysis, International Journal of Forecasting, 5:209-215.

Pamela Texter-Geriner and J. Keith Ord, (1991), Automatic forecasting using explanatory variables: A comparative study, International Journal of Forecasting, 7:127-140.

Barry R. Weller (1994), SCA SCA SCA PC-EXPERT for Windows, International Journal of Forecasting, 10(3):481 -487.

Matthew Witten (1994), Numbers to Pictures - Comparison: Visualization Software, Advanced Systems, pp. 34-40, November 1994

[Click here to the reviews page](#)

P.O. Box 563
Hatboro, PA 19040
Tel: (215) 675-0652
Fax: (215) 672-2534
sales@autobox.com