

COMPUTER SECTION

AUTOBOX: A Review

*Ronald Bewley**

AUTO BOX can be run either in batch mode or as a menu-driven program and has, as its central feature, *a fully automatic* procedure for identifying and estimating ARIMA models and transfer functions both with and without intervention detection. I was surprised by the ease with which I was able to run through all of the major automatic options on my first attempt, and there is no doubt that its automatic function will lead users to consider these types of analyses more frequently than they otherwise would. In particular it is possible to nominate the set of variables in an equation and let AUTOBOX not only choose an appropriate dynamic structure but also search for any outlying observations, changes in level, and changes in seasonal pattern at the intervention detection stage. Therefore, more traditional econometricians may find it useful to pass a potential relationship through AUTOBOX to investigate data problems and dynamic specification issues before adopting a structural approach to modelling.

AUTOBOX also has a 'manual override' facility that enables standard time-series analysis to be carried out, and there are two methods of synthesizing the two approaches. Either the user can specify an initial model, upon which the automatic procedure attempts to improve, or the insignificant parameters in a user-specified model can be deleted in a step-down procedure akin to step-wise regression.

Much of the review process has been based on Version 1.02 but an (incomplete) advanced copy of Version 2.0 has also been received. The latter offers a number of enhancement but, in essence, the key automatic function is unchanged. The new features are discussed in section 3 of this review.

* Address all communications to Ronald Bewley, Department of Econometrics, University of New South Wales, Kensington, NSW 2033, Australia. First published in *Journal of Applied Econometrics*, Vol. 3 (3) 1988, pp. 240-244.

Version 1.02 runs on an IBM PC-compatible with PC-DOS or MS-DOS at level 2.0 or above. It comprises 10 diskettes and requires 320K RAM. Version 2.0 has one diskette less but requires 400K RAM. The review process was conducted on an Olivetti M24 SP with an 8087 coprocessor installed. AUTOBOX will run without the coprocessor but the manual claims it is up to 10 times faster when one is installed. A hard disk is not essential but floppies do not seem to be a particularly practical alternative, given the storage requirements. Although the programs can be copied onto a hard disk with the COPY command, the program will not operate without a »key« disk in the floppy drive.

1. Input-Output and Documentation

Data can be input either from a file containing only the time-series observations or they can be input via the data manager. If the data file option is used the file must reside in the same directory as the program, and this does tend to lead to long and cluttered directories. The data manager is a little primitive and the only operation permitted is a linear transformation. AUTOBOX requests that such a linear transformation be applied whenever there are negative observations or ones that are large. Since the program failed when I declined the transformation, perhaps the program should assume a positive response and always transform to gain numerical accuracy. Output can be directed to a file or the screen. Unfortunately, it is not possible to create the file and watch the screen scroll with the model development, which I found a little frustrating. In fact, my overall impression is that considerable and worthwhile attention has been placed on developing the model selection procedure, and not enough attention to the input-output features. However, it must be noted that Version 2.0 has been improved somewhat in this direction.

The ultimate test of on-line help in a PC-based program is whether or not a manual is necessary. Without either having read the manual or been familiar with the automatic procedure, I was able to go right through to successful completion in automatic mode. However, when I tried to intervene in the process a few problems arose, but the manual was not then too clear. The major problem appears to be that the author does not wish to reveal how the model selection procedure functions, and this leads to some grey areas in both the manual and the on-line instruction. Much of the manual is devoted to an annotated screen dump, but the author's short notes added little to my understanding. In most cases the questions on the screen are clear and concise, so that the first 70 or so pages of the 120 are largely superfluous. On the other hand I found that the remaining technical notes were inadequate in helping me understand why the program

chose particular models and was apparently satisfied with models that I, as a practitioner, would have rejected. It can be noted from the next section that, from my limited experimentation using AUTOBOX, it is much less successful at fin-tuning user-selected models than if it is permitted to conduct the complete analysis.

2. The Automatic Procedure

The key of this program is certainly its automatic model selection procedure. Indeed, I found the standard 'manual' method a bit too slow and sluggish compared with alternatives such as micro TSP. Since AUTOBOX is a set of FORTRAN programs that are operated via a DOS batch file, programs are frequently being loaded and unloaded, which takes time. Certain special files can be created or modified before a run that will by-pass unwanted features and accommodate user-preferred options, and this will greatly enhance operating time. However, for infrequent users it is preferable to be able to direct the program during execution, and quickly switch between different time series.

The use of an undisclosed selection procedure naturally leads to two questions being raised. Does it choose the same (or a better) model than a practitioner? Does it choose its preferred model in a 'reasonable' number of steps?

AUTOBOX performs three checks after each model has been estimated in automatic mode. *Necessity* requires that all coefficients are significant; *sufficiency* requires that the acf terms are insignificant; and *invertibility* checks that the roots of *both* AR and MA polynomials are outside the unit circle. Apart from this, no information is given on the selection procedure. The user not only has no control over significance levels, but also the pre-assigned levels are not revealed. My limited experimentation suggested that the significance level is not constant, and I was informed by the author that the significance level depends on sample size. I noted that, in my trials, three standard errors was not significant (with 95 observations) but, in another problem, 1-5 standard errors was significant.

There is little point in judging a program with a few straight forward examples. I concentrated on three series which had previously given me some trouble, to see whether I would have been better off if I had used AUTOBOX. In each case I used the full automatic procedure and then restarted the automatic procedure with models of my choice. While I was more than happy with the fully automatic mode, I have some reservations about interfering with AUTOBOX's initial model selection. The first contained 95 non-seasonal observations and were taken from Bowerman and O'Connell (1979. p. 359). The first few terms of the acf and pacf are given in the table.

	1	2	3	4	5
acf	.438	-.112	-.343	-.241	.003
pacf	.438	-.376	-.159	-.024	.042

There is a little doubt that an AR(2) model would be close to a final model and this was chosen by both Bowerman and O'Connell, and AUTO BOX at the first attempt. There does not appear to be an 'overfitting' procedure in AUTO BOX but, in manual mode, an AR(3) model produced three significant t-ratios at the 5 per cent level, the smallest being 2.33. Clearly there is a problem with the first observation in comparing AR(2) and AR(3) process.

Since the automatic procedure can be started with a user-supplied model, I used a variety of starting points to test how it might be used for fine-tuning. As the author points out in the manual, a 'particularly poor model' as an initial choice may prevent the algorithm from recovering.

To remove the effects of the first few observations, observations 4-95 were used and an AR(3) was selected by AUTOBOX, but when an AR(2) was used to start the process, no alternative was offered by the program. Similarly, ARMA(1,1) and a variety of high-order MA models did not converge on the AR(3), or to any model that was remotely acceptable in my opinion. It frequently left a term at lag 3 in the residual acf of the final model in excess of 0.3. From my limited experience it appears that the program first deletes one insignificant parameter at a time and then (possibly) builds on that base. While it is not too surprising to find that acceptable. Even with moderate deviations from an AR(3) as my initial choice, I was not satisfied with AUTOBOX's model development or its final selection. Perhaps AUTOBOX should flag that it is experiencing difficulty in improving on a model more readily than it does.

My second example comprises 220 highly seasonal monthly observations on a high-growth telecommunications series. AUTOBOX claims to use a Box-Cox option to transform a series to variance stationarity but, in fact, restricts its attention to values of the Box-Cox parameter equal to 1, 1/2, 0, -1/2, and -1. The optional parameter value is chosen, in this case 0, on an error sum of squares criterion.

The manual is not very precise in many of its descriptions and, for example, does not actually say that the Box-Cox error sum of squares is adjusted to account for the transformation of the dependent variable (in private correspondence, I have been assured that the adjustment is made). Moreover, the manual does not make it clear that only a five-point grid search over these values is made, and I spent some time trying to work out if and how estimated value of the Box-Cox parameter (of any value) could be found.

AUTOBOX took five steps to arrive at its preferred model in the second example and, during the course of its selection procedure, AUTOBOX

considered first-differencing and seasonal differencing but not both simultaneously. AUTOBOX's preferred $(2,0,0) * (0,1,1)_{12}$ model included a trend constant and the autocorrelations at lags 12 and 24 are 0.130 and 0.152 with standard errors of 0.072 and 0.075, respectively. Given the importance of lags 12 and 24 with monthly series, I considered that AUTOBOX should have done more in an attempt to improve the model, particularly at seasonal lags.

Using the same program in non-automatic mode, I re-estimated my preferred model, $(1, 1, 1) * (0, 1, 1)_{12}$, without a constant. All coefficients in both my model and AUTOBOX's are greater than two standard errors. The autocorrelations at lags 12 and 24 in my model are 0.076 and 0.108 with standard errors of 0.071 and 0.072. The difference between the two models is not great (AUTOBOX's m.s.e. is 3.9 per cent higher than that for my preferred model) but it does highlight that one must accept AUTOBOX's preference for, say, including constant terms over higher-order differencing. Naturally, the difference in the forecast functions had a major impact on the forecast of the two models, of the order of 10 per cent in the first year, but the 95 per cent confidence intervals were of a similar size (approximately ± 17 per cent). Interestingly, the automatic procedure did not depart from my model when it was used as a starting point.

The third example was based on Australia's highly volatile monthly current account data with 125 observations. The series were analysed in isolation and in a transfer function with an SDR/A \$ exchange rate series. Transfer modelling is a key feature of AUTOBOX and, because of the additional degree of complexity in estimating such model over univariate models, automation is even more welcome here. I have worked extensively with these series in a different modelling framework so I was pleased to note that AUTOBOX chose a model that was compatible with my prior.

Two types of transfer function modelling are possible within AUTOBOX. The first involves separate prewhitening of each input series followed by a transformation of the output series with the convolution of the input filters. With the second option, all series are transformed with the same filter, a technique which is more appropriate in multiple input modelling. Naturally, the automatic transfer function modelling option is time-consuming, but so is user-specified development. Note that this and all other analyses can be operated in batch mode with AUTOBOX.

There have been a number of economic events that have altered the generating process of Australia's current account. I used the automatic intervention analysis option to see whether it was able to detect these structural changes. Three types of interventions are possible: single-impulse dummy variables; step functions; and pulses of the seasonal interval. The composition of the current account has changed substantially over the period and, with it, there has been a change in the seasonal pat-

tern. This was picked up by AUTO BOX, as was the change in the level of the process that followed Australia's floating (and subsequent devaluation) of its currency. Furthermore, AUTOBOX detected three single observations as outliers.

I was definitely impressed by the information on a relationship that can be obtained for so little effort and the models selected by AUTOBOX, either as part of an exploratory analysis in structural modelling or as an end-product in time-series work, are worthy of serious consideration. However, my reservation about using AUTOBOX for fine-tuning univariate models make me wary of using AUOTBOX in a similar role with multiple time-series.

3. VERSION 2.0

Unfortunately my copy of Version 2.0 arrived without a complete manual, and with one or two bugs left in the program. However, it was extremely useful and helpful to be given this preview. The input-output feature have been improved, although there is scope for further improvement. For example, the user is now warned if the output file will overwrite an existing file, but data and output files still must reside in the same directory as the programs, a feature which I find frustrating, and it would be simple to generalize or make it as an option to be set when configuring the program.

A utilities module has been added that allows the user to plot series and compute simple descriptive statistics in a convenient form. However, the main additions to AUTOBOX allow users to perform 'traditional regression analysis' either as a simple regression or as a step-wise to specify a model and randomly generate simulated series.

Standard regressions can be computed with or without a step-down feature at a pre-assigned level of significance. Clearly the algorithm draws upon the time-series 'automatic version'. Since each data series has to reside in a different file, and the program must be reloaded for each regression, it is not a serious contender or basic regression. Of course, relatively few econometricians use step-wise procedures in economic model-building these days but this option may occasionally be useful in exploratory analysis.

4. CONCLUSIONS

My overall impression is that this program has an excellent role to play in model-building, providing that the fully automatic procedure is not tampered with. Thus, it can replace the human capital needed in a business environment or provide an 'independent' check for more experienced model-builders. It is a program I would certainly welcome for my own research and in the preparation of lecture course material. I doubt if I would want my students to use the fully automatic version of AUTOBOX until they had mastered the basic principles of time-series analysis, or had no intention of ever becoming such an analyst. AUTOBOX is very good at producing final model estimates but the path it follows can be quite different to the way this practitioner would approach a problem (or as detailed in standard texts) and may, as a result, confuse a student.

AUTOBOX was too slow and cumbersome to operate as a non-automatic ARIMA program for my taste, even with a coprocessor. Although the program is not fast in automatic mode, the 5.10 minutes required in my experiments is probably much faster than most practitioners could estimate a model when thinking time and mistakes are included.

In my opinion the program would be substantially improved if the user were allowed to impose personal preferences (e.g. on the level of differencing in the model) and if some indication of overfitting was more apparent.

There are many versions of AUTOBOX, depending on which options, are requested and whether or not a site licence is required. The complete package (AUTOBOX PLUS) is \$1195 and this is reduced to \$995 (AUTOBOX) if the step-wise regression and data simulation models are excluded. A univariate (automatic) program (AUTOBJ) cost \$595 and naturally does not permit intervention analysis. A non-automatic version (BOXX) of AUTOBOX PLUS is available for \$695 and AFSEZF is a simple univariate program at \$195. Finally, the data simulation program (SIMBOXJ) is available for \$195. Student versions are available for \$10 or \$25 depending on the options but will only run on a supplied data set, although for \$250, AFS will tailor the data set to an instructor's requirement. Site licences begin at \$250 plus \$100 per year.

David Riley is the author and AUTOBOX can be obtained from: Automatic Forecasting Systems, Inc., PO Box 563, Hatboro, Pennsylvania 19040, USA

Reference: Bowerman, B.L., and R. T. O'Connell (1979) Time Series and Forecasting, Wadsworth, Belmont, CA.