

This review originally appeared in the February 1996 edition of The American Statistician, Vol. 50, No.1  
©1996 American Statistical Association

## **Automatic Forecasting**

**Keith ORD and Sam LOWE**

---

AUTOBOX, Version 3.0

Available from Automatic Forecasting Systems, Inc.,  
P.O. Box 563,  
Hatboro, PA 19040.  
\$395. Demo diskette free.

AUTOCAST II

Available from Delphus, Inc.,  
103 Washington St., Suite 348  
Morristown NJ 07960.

\$349. Demo diskette free.

FORECAST PRO, Version 2.0

Available from Business Forecast Systems, Inc.,  
68 Leonard St.,  
Belmont, MA 02178.

\$595. Demo diskette free

NCSS

Available from NCSS,  
329 North 1000 East,  
Kaysville, UT 84037.

\$249 (Base system plus three modules, one of which is time series)

4CAST/2

Available from Delphus, Inc.,  
103 Washington St., Suite 348,  
Morristown, NJ 07960.

\$495. Demo diskette free.

---

### **1. AUTOMATIC FORECASTING**

The term *automatic forecasting* describes a forecasting system (FS) that, apart from some initial specifications, requires only the input of an observed time series in order to generate a set of forecasts.

That is, the selection of a forecasting scheme, from some

prespecified set of possibilities, takes place *without user intervention*. For completeness we shall refer to *manual* selection whenever the choice of model requires explicit choices to be made by the analyst.

The motivation for automatic forecasting stems from the large number of time series that a forecaster may face in an operational setting, such as the thousands of components of ten held in inventory by a manufacturing plant. The value of inventories for single components is such that the detailed modeling of individual series would not be cost-effective. A *batch* FS that operates automatically and feeds from and into a company's database is clearly more appropriate. Further, operational experience with automatic selection procedures suggests that they may match up quite well with models identified by an analyst. Thus even when a series is sufficiently important to warrant the analyst's serious attention, the automatically generated forecast will often be a useful place to start.

For further discussion of the relative performance of automatic and manual methods, see Hill and Fildes (1984), Poulos, Kvanli, and Pavur (1987), Texter and Ord (1989), and the earlier software review by Tashman and Leach (1991). Rycroft (1993) provides a detailed appraisal of 103 statistical packages that include a forecasting capability.

By way of background Section 2 of this review deals with the structure of forecasting systems. Section 3 describes the process by which particular programs were selected. Sections 4 and 5 deal, respectively, with the general computational and the statistical forecasting capabilities of the packages. Section 6 reports on a comparative study of the programs using a standard set of series, and Section 7 contains some brief remarks about the individual packages. Future directions for automatic forecasting are outlined in Section 8, and some conclusions are presented in Section 9. Finally, in Section 10 some recent enhancements in the packages are noted.

## **2. FORECASTING SYSTEMS**

In this review we focus upon computer programs that enable us to generate forecasts for a single series, using as inputs only the past values of that series; that is, the forecasts are generated using univariate time series methods. Such approaches involve a forecasting system that incorporates the following components:

- A set of possible *time series models or forecast*

*functions* (FF); an FF may be derived from a time series model, but some forecasters use an FF directly, for which there may or may not be an underlying model.

- A *selection*, or *identification*, mechanism that determines the "best" model or FF according to some preset criterion.
- Procedures for *estimating*, or at least setting, the values of the unknown parameters.

The selection process may be one of three possibilities, depending on whether the software allowed choices from

1. a fixed list
2. a list that could be modified by the user prior to performing the forecast task or
3. a class of time series models, typically ARIMA, for which forecast functions can then be specific.

For (1) and (2) selection was usually based upon fitting all the models on the list and choosing the "best" according to a user specified criterion such as the forecast meansquared error (FMSE); for (3) time series models were selected using the autocorrelation function, partial autocorrelation function, or similar criteria. In all cases the programs also allowed the user to choose a procedure on a "manual" basis. Once the FF has been identified and fitted, generating point forecasts is a simple matter of extrapolation; interval forecasts are considerably more complex (Chatfield 1993).

**Table 1. Summary of Operating Characteristics for Each Program**

| System requirements  | AUTOBOX | AUTOCAST II | FORECAST PRO | NCSS | 4CAST/2 |
|----------------------|---------|-------------|--------------|------|---------|
| Math coprocessor (a) | REC     | REC         | REC          | REC  | OP      |
| Hard disk            | 3 MB    | <1 MB       | 1 MB         | no   | 1 MB    |
| Minimum RAM (b)      | 640K    | 640K        | 2 MB         | 512K | 640K    |
| Input/Output         |         |             |              |      |         |
| Format of input (c)  | A       | C, T        | A            | A    | C       |
| Graphics-character   | yes     | no          | yes          | yes  | yes     |
| -exportable?         | yes     | no          | yes          | yes  | no      |

|                        |      |     |          |        |     |
|------------------------|------|-----|----------|--------|-----|
| Operations             |      |     |          |        |     |
| Windows available (d)  | no   | no  | yes      | no     | no  |
| Max # observations     | 2000 | 500 | no limit | 13,200 | 999 |
| Ease of use (e)        |      |     |          |        |     |
| Installation           | 3    | 4   | 4        | 3      | 4   |
| Tutorials/help screens | 2    | 3   | 4        | 2      | 3   |
| Output                 | 4    | 4   | 3        | 4      | 4   |
| Documentation          | 3    | 3   | 4        | 3      | 2   |
| Overall                | 3    | 4   | 4        | 3      | 3   |

- (a) Math coprocessor: REC = recommended, OP = optional.
- (b) Minimum RAM: Network connections may need to be switched off to provide sufficient RAM, depending upon the machine and the configuration.
- (c) Format of input: R = rows, C = columns, T = tables, A = all.
- (d) WINDOWS availability: See Section 10 of paper for updates.
- (e) Ease of use four-point scale: 4 = good/easy, 3 = only minor problems, 2 = could be improved, 1 = not acceptable

The time series models underlying the forecast process are straightforward, and we do not elaborate upon them here; for a full discussion see, for example, Abraham and Ledolter (1983), Chatfield (1989), or Kendall and Ord (1990).

### 3. SELECTION OF SOFTWARE

A recent directory of forecasting packages was compiled by Aghazadeh and Romal (1992). From this listing we identified all those packages that featured "automatic model selection," and requested copies for testing. The five packages considered in the review represent the totality of positive responses for which we had access to the current version of a commercially available program. All programs were run on IBM or compatible platforms. Among the generalpurpose statistical software companies only NCSS and SAS have automatic forecasting programs either available or under test.

The NCSS software is currently available, and is evaluated in this review. The SAS System (a modifiable list system that runs in WINDOWS) is still under development so we have not reported upon it here.

During the time that we were testing the software a detailed summary of the capabilities and requirements of a number of forecasting packages appeared in *OR/MS Today*, produced by Yurkiewicz (1993). Our summary, Table 1, relies heavily upon this source for those packages common to both studies, and we have cross-checked the appropriate entries for consistency.

The capabilities and requirements listed in Tables 1 and 2 show considerable variations. Our evaluation is designed to point to the performance characteristics of each program, and to leave the final judgment to the reader in light of his or her own requirements.

#### **4. CRITERIA: REQUIREMENTS AND PERFORMANCE**

These refer to program requirements, their capabilities, and their performance. Certain common features may be noted. In all cases

1. The minimal configuration is a 286 system.
2. It is possible both to read and create ASCII files and to interface with major spreadsheets.
3. The programs may run in batch mode to handle a large number of series.
4. A data editor is available, although the degree of sophistication varies.
5. Systems have basic (pixel) graphics capabilities, but some go much beyond this minimum.
6. Systems are menu-driven.
7. The user may select the start and end points within a series, although the method is not always transparent;
8. The option of *manual*, rather than *automatic*, selection is available.

The additional criteria, which varied across packages, are defined below and summarized by package in Table 1.

*System requirements:* These items are generally selfexplanatory although some judgment was involved; for example, some programs did not require a math coprocessor, but ran very slowly without it. Inputs could be *row* only, *column* only, *tabular*, or all of these options.

*Outputs:* These items include the production of ASCII- type output files, the availability of graphics, as well as the types of plot, etc., available from each program.

*Operations:* The ability to run in batch mode without user intervention between successive series is important when a large number of series must be forecast; conversely, the flexibility to develop forecasts manually is desirable for important series where the user may wish to explore beyond the confines of the automatic system.

| <b>Table 2. Forecasting Capability: The Main Statistical Features Available in Each Program</b> |               |             |              |      |         |
|---|---------------|-------------|--------------|------|---------|
|   | AUTOBOX       | AUTOCAST II | FORECAST PRO | NCSS | 4CAST/2 |
| Exponential Smoothing   |               |             |              |      |         |
| Single  | yes(a)        | yes         | yes          | yes  | yes     |
| Double  | no            | no          | no           | no   | yes     |
| Holt  | yes(a)        | yes         | yes          | yes  | yes     |
| Adaptive  | no            | no          | no           | no   | yes     |
| Damped trend  | yes(a)        | yes         | no           | no   | yes     |
| Winter's seasonal   | no            | yes         | yes          | yes  | yes     |
| Harrison's seasonal   | no            | no          | no           | yes  | no      |
| Parameter estimation  | yes(a)        | yes         | yes          | no   | yes     |
| Analysis of outliers  | yes(a)        | yes         | no           | no   | yes     |
| ARIMA modeling  | generally yes | no          | yes          | yes  | yes     |
| Polynomial trends   | no            | no          | no           | yes  | no      |
| Seasonal models   | yes           | no          | yes          | yes  | yes     |
| Seasonal dummies  | yes           | no          | no           | no   | no      |
| Noncontiguous lags  | yes           | no          | no           | no   | no      |
| General   |               |             |              |      |         |

|                           |        |     |             |         |                      |
|---------------------------|--------|-----|-------------|---------|----------------------|
| capabilities              |        |     |             |         |                      |
| Transforms available(b)   | yes    | yes | yes         | LN      | yes                  |
| Multiple criteria         | no     | yes | yes         | no      | no                   |
| Model selections          | C      | F   | C           | C       | C                    |
| Produces ACF, PACF        | yes    | yes | yes         | no      | yes                  |
| Detailed diagnostics      | yes    | yes | yes         | no      | yes                  |
| Interval forecasts        | yes    | yes | yes         | no      | yes                  |
| -level at choice          | yes    | yes | yes         | no      | yes                  |
| -time-dependent variances | yes    | no  | no          | no      | no                   |
| Multiple forecast origins | yes    | yes | yes         | no      | yes                  |
| Rolling simulation        | yes    | yes | yes         | no      | yes                  |
| -with reestimation        | yes    | yes | yes         | no      | no                   |
| -model reselection        | yes    | no  | no          | no      | no                   |
| Other techniques(d)       | IA, TF | TSD | TAR, MA, DR | TR, TSD | C, TSD, PT, X11, SWR |

- (a) Exponential smoothing in AUTOBOX available with ARIMA framework.
- (b) Transformations available: LN = logarithmic, SR = square root.
- (c) Model selection: F = fixed list, C = class of models.
- (d) Other techniques: TSD = time series decomposition, IA = intervention analysis, TAR = trend & AR errors, TF = transfer function, X11 = Census X11, PT = polynomial trend, MA = moving averages, TR = trigonometric regression, C = combinations of forecasts, SWR = stepwise regression, DR = dynamic regression.

*Ease of use:* Comparisons were made by at least two users operating independently, and the rating represents a composite of their assessments on a four-point scale: 4 = good/easy. 3 = only minor

problems, 2 = could be improved, 1 = not acceptable. The reported scores denote averages across users. Not every user scored every attribute. Users were assigned to particular packages in such a way as to ensure that no user had previous experience in the operation of that program, and every user compared two or three programs. In making the "ease of use" comparisons it should be noted that we went with the "plain vanilla" options in each package. Thus a package with an extensive set of options, such as AUTOBOX, requires a greater initial investment of time, but can provide a wider range of analyses. However, we do not feel that our "ease of use" scores were influenced by the complexity factor. Packages with more options allow an analyst greater flexibility in followup investigations, but a detailed appraisal of such benefits is beyond the scope of our study.

It is important to note that the eight properties listed above are common to the five packages reviewed; they are by no means available on all of the packages currently available.

## 5. FORECASTING CAPABILITY

Table 2 provides a summary of the overall forecasting capabilities possessed by each program; this table is designed to describe potential rather than performance. The following features were common to all programs:

1. Manual selection was allowed as an alternative to automatic.
2. Details of the model selection process could be printed out as an option.
3. Forecasts could be made for multiple horizons; that is, the programs were not restricted to one-step-ahead forecasting.

*Exponential smoothing:* The basic methods are single and double smoothing (Brown's approach), Holt's two-parameter linear smoothing, and multiplicative Winters (or Holt-Winters) three-parameter scheme for seasonal series. In addition, smoothing with an adaptive rate has its band of devotees. The use of a damped trend, corresponding in the homoscedastic additive error case to an ARIMA (1, 1, 2) scheme, has become increasingly popular. Harrison's harmonic smoothing procedure is favored for seasonal series in some programs. The mode of implementation of exponential smoothing methods is important. Some programs use default values for the parameters, whereas others search for the best fitting values. Outliers detection and adjustment are also available in

some cases.

ARIMA modeling: At the most basic level a program may select one of a fixed list of ARIMA models based on some criterion such as minimum mean-square error, with no other guidance on identification and no diagnostics. Beyond this basic structure we would hope to find the availability of plots for the autocorrelation (ACF) and partial autocorrelation (PACF) functions, as well as detailed diagnostics. Although regular and seasonal differences are the most popular way of dealing with trends and changing seasonal patterns, the use of polynomial trends and seasonal dummies is attracting renewed interest; these options are now available in some programs. Nonlinear transformations have long been popular as mechanisms to induce stationarity; generally the programs had only limited automatic options available, if any; the choice was usually restricted to the logarithm (LN) and the square root (SR). The ability to identify noncontiguous lags is useful both in the interests of parsimony and as a way of detecting perhaps unsuspected seasonal patterns, such as a three-monthly effect due to quarterly reporting requirements. Finally, although our interest focused upon univariate forecasting procedures, we have noted where a program included intervention analysis and transfer function capabilities, either within the standard configuration or as an add-on from other systems produced by the vendor.

General capabilities: Under this heading we have included a number of other features that are important to users. The use of a model selection criterion such as minimum mean-square error may lead to overfitting, so it is desirable to have the option of using other measures such as information criteria; the list varies considerably by program.

The forecasting process should not be limited to point forecasts, but should include interval forecasts, preferably with the width of the prediction interval as a choice. Most such intervals assume a normal distribution with constant variance for the error process, but time-dependent variances are slowly being incorporated into the programs (cf. Chatfield 1993). The flexibility to vary both forecast origins and forecast horizons enables the user to assess the stability of the forecasting procedures identified, and thereby increase the comfort level with the selection process.

## **6. FORECAST PERFORMANCE**

In order to test each program, we used a set of six series, given in Table 3.

| <b>Table 3. Series Used in Study</b> |           |                        |  |
|--------------------------------------|-----------|------------------------|--|
| Series                               | Periods   | Number of observations | Source                                 |
| (1) Air conditioner sales            | Monthly   | 60                     | Makridakis and Wheelwright (1978)      |
| (2) PA employment                    | Monthly   | 156                    | Pennsylvania Economic Analysis Project |
| (3) U.S. GNP                         | Quarterly | 100                    | Business Conditions Digest             |
| (4) PA income                        | Quarterly | 54                     | Pennsylvania Economic Analysis Project |
| (5) Sheep population                 | Yearly    | 73                     | Kendall and Ord (1990)                 |
| (6) Utah employment                  | Yearly    | 23                     | Makridakis and Wheelwright (1978)      |

Performance was evaluated by holding out the last 12/8/6 observations for monthly/quarterly/yearly series; a policy followed in Makridakis et al. (1982) and a number of other studies. The principal characteristics of each series are as follows:

1. strongly seasonal, little or no trend
2. seasonal, increases then levels out
3. seasonal with a strong upward trend
4. strong upward trend
5. declines somewhat erratically, then increases at the end
6. strong upward trend.

Initially, individual investigators used these series to form their judgments about the performance and ease of use of the programs. Their assessments served as inputs to Tables 1 and 2.

Whenever a small number of series is used to evaluate performance, the choice of such series is open to criticism. Our series are dominantly economic and relatively long. We chose the series to reflect a variety of structures of potential interest, and do not regard them as in any way "representative" of some "population of series"; see the discussion following Makridakis et al. (1982) on this issue. For this reason we have not provided any aggregate statistics (across series) in Table 4 because the rankings that might be inferred from such summaries are not meaningful. In particular, a retrospective analysis revealed that, at the start of their holdout periods, the PA employment series had a major change of direction and the air

conditioner series had a change of level.

Also, we note that out-of-sample forecasts for successive time periods are very highly related, so that the performance measures for individual series have a high degree of variability, as is evident from Table 4.

Finally, we note that none of the series is recorded more frequently than once a month, although a major virtue of automatic forecasting software is that it can handle a large number of very short-term forecasting tasks (e.g., weekly data) very economically, where simple methods will often suffice.

All the programs were then run on all series to determine forecast performance over the hold-out samples. The forecast functions were selected and used to 12/8/6 periods ahead for monthly/quarterly/yearly data. The next observation was added to the series, and a new set of forecasts computed; the process was repeated until the end of the series was reached. This process is known as rolling simulation, and is available as a standard feature in several of the packages; see Table 2. Rolling simulation is a valuable way of checking forecast performance since "in-sample" measures of fit often prove unreliable (cf. Makridakis et al. 1982). Ideally, rolling simulation should include model reestimation at each stage, and even the reselection of the model. The availability of such features is noted in Table 2.

**Table 4. Aggregate Forecast Error Measures: Mean Absolute (FMAE), Mean Absolute Percentage (FMAPE), and Root Mean-Square (FRMSE) by Program and Series**

|        |           | AUTOBOX(a) |     |             |              |      |         |                            |
|--------|-----------|------------|-----|-------------|--------------|------|---------|----------------------------|
| Series | Criterion | (a)        | (b) | AUTOCAST II | FORECAST PRO | NCSS | 4CAST/2 | Manual ARIMA(b) model      |
| (1)    | FMAE      | 100        | 127 | 160         | 131          | 210  | 178     | 110<br>(0,0,0)(0,1,1)12    |
|        | FMAPE     | 100        | 125 | 133         | 128          | 208  | 166     | 106                        |
|        | FRMSE     | 100        | 131 | 132         | 116          | 179  | 158     | 101                        |
| (2)    | FMAE      | 112        | 125 | 100+        | 100          | 147  | 117     | 115 (0,1,1)                |
|        | FMAPE     | 112        | 124 | 100         | 102          | 143  | 116     | 114                        |
|        | FRMSE     | 111        | 126 | 100+        | 100          | 152  | 117     | 119                        |
| (3)    | FMAE      | 293        | 267 | 113         | 100          | 301  | 304     | 113<br>(0,2,2)(0,0,1)8(+C) |
|        | FMAPE     | 309        | 280 | 120         | 100          | 315  | 306     | 120                        |
|        | FRMSE     | 260        | 219 | 100         | 100+         | 251  | 274     | 98                         |

|     |       |     |     |     |     |     |      |                 |
|-----|-------|-----|-----|-----|-----|-----|------|-----------------|
| (4) | FMAE  | 267 | 312 | 362 | 166 | 100 | 220  | 228 (0,1,0)(+C) |
|     | FMAPE | 267 | 312 | 362 | 171 | 100 | 214  | 228             |
|     | FRMSE | 262 | 304 | 353 | 163 | 100 | 229  | 228             |
| (5) | FMAE  | 100 | 105 | 102 | 102 | 165 | 173  | 117 (3,1,0)     |
|     | FMAPE | 100 | 105 | 102 | 102 | 164 | 142  | 135             |
|     | FRMSE | 100 | 106 | 104 | 105 | 172 | 142  | 123             |
| (6) | FMAE  | 127 | 127 | 105 | 128 | 210 | 100  | 127 (0,1,0)(+C) |
|     | FMAPE | 121 | 121 | 100 | 125 | 196 | 103  | 122             |
|     | FRMSE | 123 | 123 | 100 | 127 | 197 | 100+ | 123             |

NOTE: The smallest entry in each row is scaled to 100: 100+ means that the entry was not the smallest, but was very close.

- (a) version (a) is the standard; version (b) includes automatic intervention detection and reestimation.
- (b) The ARIMA model selection was done manually using SAS; the standard notation is used: (p, d, q)(P, D, O)S, except that (+C) denotes that a constant term should be included.

The results of the forecasting exercise are summarized in Table 4. The three measures reported are forecast mean absolute error (FMAE), forecast mean absolute percentage error (FMAPE), and forecast root mean-square error (FRMSE). All are in aggregate form, that is, we averaged across all replicates and for all different time horizons. A more disaggregated analysis could be presented (cf. Makridakis et al. 1982), but the overall summary presented here is consistent with the more detailed results.

In order to achieve some comparability of performance across selection procedures we did not use transformations in the final analysis (one of many decisions debated at some length). In addition to the standard analyses for the five packages—we also included an AUTOBOX analysis with automatic intervention detection to assess the effects of outliers, and an analysis based upon manual selection of ARIMA models. The SAS procedure ARIMA was used for this exercise to avoid any hint of bias; selection was based upon the complete series, but the forecasting and model estimation used the same framework as the automatic schemes. We felt that this compromise avoided undue bias in favor of either manual or automatic selection processes.

The results are presented in Table 4. The most striking features are as follows.

1. Automatic methods perform about as well as manual

approaches. This conclusion has been reached previously in several empirical studies, as noted in the Introduction.

2. Performance differed across series for different packages. In general, NCSS performed rather less well than the others, but no other clear preferences emerge.
3. Outlier detection may or may not be beneficial; the reasons for this variability in performance are not evident.

Clearly, conclusions (2) and (3) are very tentative and would require substantial further testing. Conclusion (1) indicates that the potential of automatic methods, noted in the studies cited, appears to be realized in currently available packages.

## 7. INDIVIDUAL PROGRAMS

In this section we detail comments on individual programs that are not easily reduced to tabular form.

*AUTOBOX*: For the schemes the program selects an initial model using checks for stationarity and the ACF and PACF. AUTOBOX then uses a succession of necessity and sufficiency checks to delete or add elements. AUTOBOX includes an outlier detection scheme that may be used to add intervention variables. The support materials were comprehensible, but could be improved. AUTOBOX has the most complete rolling simulation facility, as noted in Table 2. AUTOBOX allows greater flexibility in postforecast analysis, in that each vector of forecasts may be stored and analyzed. Also, of the five packages considered, AUTOBOX has the most extensive data management facilities. The system is available in a number of variants that extend to include transfer functions and multivariate time series.

*AUTOCAST II*: This program concentrates on exponential smoothing. AUTOCAST checks first for seasonality, and then for a trend component (constant, linear, or damped); if both trend and seasonal are included the seasonal element is multiplicative. Good diagnostics are provided. AUTOCAST is easy to use and well-documented. The package has a rolling simulation facility. For inventory planning AUTOCAST provides an analysis of the Economic Order Quantity (EOQ) model.

*FORECAST PRO*: FP provides a rule-based expert system that starts out with basic statistics and a classical decomposition of the series. Summary measures based on out-of-sample forecasts are used to

choose the preferred method, including the choice between exponential smoothing and ARIMA. Good diagnostics are provided. FP is easy to use, and had the best support (documentation, etc.) of all the programs considered. The package also has a rolling simulation facility.

*NCSS*: NCSS is a general statistical package, of which the automatic forecasting system forms only a small part. Any evaluation of its other features is beyond the scope of this study. NCSS handles seasonal ARMA series using an ARMA ( $S + 2, S + 1$ ) scheme for seasonality of  $S$  periods (cf. Pandit and Wu 1983, chap. 4, 9). On occasion the appropriate order model could not be fitted, and a lower order scheme had to be used. A useful feature of NCSS is the provision of classical time series decomposition plots. Models are rated by their residual mean-square errors, and this criterion produces a tendency to overfit. Overall, NCSS seemed somewhat less user-friendly than the other programs, but this may not be a problem if the complete system is used on a regular basis.

*4CAST/2*: Covers both exponential smoothing (ES) and ARIMA schemes, although the set of possible ARIMA models is restricted. The choice between ES and ARIMA is made at the start of the analysis. In our study we restricted attention to the smoothing methods, which may account for the results in Table 4.

## 8. FUTURE DIRECTIONS

As noted in our opening review, automatic forecasting procedures for ARIMA models have now reached the stage where their results are comparable to those achieved by competent analysts. Given the huge potential for cost savings in operating a forecasting system automatically, with only exception reporting, the advantages of such software become evident. At the same time we must recognize that such systems will be used by nonexperts so that the decision rules need to be reliable.

By and large the software we reviewed had good datahandling facilities, with the ability to select part of a series for estimation purposes, and then to withhold the later observations for out-of-sample model evaluation. The out-of-sample evaluation should be performed by updating the forecasts at successive origins (rolling simulation) and, ideally, by reestimating the model each time, as is already done in AUTOBOX. Clearly, the data management facilities must also allow running in batch mode, whereby a large number of series can be processed sequentially without intervention, as allowed by all the software reviewed. However, given the likely use by

nonexperts, it is desirable that systems flag possible exceptions; that is, series that have behaved erratically in recent periods. Given that rolling simulations are already in place, this step should not be too difficult, although adequate criteria need to be devised.

On the methodological side a number of developments come to mind, such as models with time-varying parameters and nonlinear schemes; likewise, we would like to see exponential smoothing developed more systematically through structural models. However, many of these features do not yet exist in standard forecasting software so it may be unreasonable to expect them in automatic schemes any time soon.

Another area for development is that of multivariate series. Although this topic was beyond the scope of our study we note that AUTOBOX has a multivariate time series system (MTS) that allows automatic development for vector models.

Finally, there is room for improvement in the provision of prediction intervals. However, given the problems noted earlier (Chatfield 1993), this topic remains one where further theoretical developments are urgently needed.

## **9. SUMMARY**

All the programs reviewed have been available for some time and, as such, are among the most successful products in a rather crowded field, as noted by Aghazadeh and Romal (1992) and Rycroft (1993). One effect of this competition between packages is an element of convergent evolution to systems that include batch processing, spreadsheet interfaces, and multiple platforms. The configurations described in this review were correct at the time of proofreading, but a number of developments are in the pipeline (see Section 10), and the potential user should check with vendors.

In conclusion, users seeking an automatic forecasting package should be aware that differences do exist in certain key areas, and they should weigh their requirements and select accordingly on their needs for:

- accessibility
- ARIMA models versus exponential smoothing
- advanced features such as intervention analysis and transfer functions
- rolling simulations and

- data transformations.

## 10. UPDATES

AUTOBOX: Version 5.0 for WINDOWS has just been released; enhancements include improved reporting and help facilities.

AUTOCASTII: Has been integrated into a general operational system known as PEER Planner for WINDOWS.

FORECAST PRO: No major new developments reported.

NCSS: Version 6.0 for WINDOWS has just been released; it will include an update of the time series component by the end of fall 1995.

4CAST 2: A WINDOWS version is due for release in late 1995 when the system will be extensively revised.

*[Received August 1995. Revised .]*

Keith Ord is with the Department of Management Science and Information Systems, Pennsylvania State University, University Park, PA 16802.

Sam Lowe is with the Business Operations Analysis Group, ATT Bell Laboratories, Somerset, NJ 08873.

The authors are indebted to a number of colleagues, who helped them carry out the first round of testing of these programs: Sherry Bowman, Paul Fields, Duncan Fong, Ram Ganeshan, Gina Gempesaw, Jack Hayya, Ralph Snyder, and Susan Xu.