

Reprinted from *International Journal of Forecasting*
7

Copyright 1991, pp. 209-230, with permission from
Elsevier Science

Computer reviews

Automatic forecasting software: A survey and evaluation^{1[1]}

Leonard J. Tashman

School of Business Administration, University of Vermont, Burlington, VT 05405, USA

Michael L. Leach

Burlington Electric Department, Burlington, VT 05401, USA

Abstract: Forecasting software, through the incorporation of automatic features for model selection and estimation, has made heretofore "complex" methods more accessible to the practitioner, giving rise concomitantly to claims that software now relieves the practitioner of the burden of technical knowledge. Academics, however, have questioned the wisdom of forecasting without foretraining.

This paper presents a survey and evaluation of automatic forecasting, based on the features of thirteen forecasting packages which perform single-equation methods on time series data. Our goals are (a) to clarify for the practitioner the virtues and limitations of automatic forecasting, and (b) to assess whether the software encourages if not nurtures good forecasting practice in the identification, evaluation, and defense of a forecasting model.

Our principal conclusions: forecasting software can provide substantial and reliable assistance to the practitioner in the selection of appropriate specifications for extrapolative models. With regard to important tasks involving the evaluation and presentation of forecasts, as well as for the determination of whether the introduction of causal variables is worthwhile, the practitioner is left largely to his own devices, expertise, and judgment. The serious danger to the untrained practitioner is the "closed-world" problem of knowing what you don't know.

Keywords: Automatic parameter optimization, Automatic method selection, Automatic data decomposition, Fit criteria, Post-sample simulation, M-competition, Forecasting formula.

I. Introduction

In recent years, forecasting software has become more comprehensive, as well as far simpler to use. As a result, what once were considered to be complex methods have now become accessible to the practitioner.

^{1[1]} Note from the editor: this article has been refereed in the usual fashion.

These improvements come partly as a result of the incorporation of automatic features for model selection, estimation, and evaluation. Indeed, some software developers have suggested that their products relieve the forecasting practitioner of the burden of technical knowledge. "We have a forecasting tool easy enough for those who slept through their statistics course", once proclaimed the package *Wisard* (now *Trendsetter*).

It is less certain, however, whether these developments have upgraded the practice of forecasting. To some in the academic world, it is not the claim of simplicity that is raising eyebrows; it is the wisdom of forecasting without foretraining. In a recent software review, Beaumont (1987, p. 72) expressed his "opinion that forecasting expert systems will, if used by the untrained, be a backward step for the profession." Jenkins (1982, p. 14) was more pointed: "The notion that data can be fed into a computer like a 'sausage machine' and forecasts produced automatically is manifestly absurd."

Exhibit 1
The programs.

Software program	Version	Company
<i>Autobox Plw</i>	2.0	Automatic Forecasting Systems, P.O. Box 563, Hatboro, PA 19040, USA Tel: (215)-675-0652
<i>Autocast II</i>	1.16	Levenbach Associates, 103 Washington St. Suite 348, Morristown, NJ 07960, USA Tel: (201)-285-9248
<i>Forecast!</i>	2.1	Intex Solutions, 161 Highland Ave., Needham, MA 02194, USA Tel: (617)-449-6222
<i>Forecast Plus</i>	2.1	Walonick Associates, 6500 Nicollet Ave. So., Minneapolis MN 55423, USA Tel: (612)-866-9022
<i>Forecast Pro</i>	1.01	Business Forecasting Systems, 68 Leonard St., Belmont, MA 02178, USA Tel: (617)-484-5050
<i>Forman</i>	2.12	Professor Keith A. Yeomans University of Witwatersrand P.O. Box 98, Wits, Johannesburg 2050, South Africa Tel: (011)-643-6641
<i>Number Cruncher</i>	5.0	NCSS, 865 East 400 North, Kaysville, UT 84037, USA

<i>PC/Sibyl</i>	5.06	Tel: (801)-54600445 Applied Decision Systems, 33 Hayden Ave., Lexington, MA 02173, USA
<i>Smartforecasts II</i>	2.15	Tel: (617)-861-7580 Smart Software, 392 Concord Ave., Belmont, MA 02178, USA
<i>Smooth</i>	1.2	Tel: (617)-489-2743 True Basic, 39 South Main St., Hanover, NH 03755, USA
<i>Time Machine</i>	2.1	Tel: (603)-643-3882 Research Services, 797 East 5050 South, Ogden, UT 84403, USA
<i>Trendsetter Expert</i>	1A	Tel: (801)-479-3553 Concentric Data Systems, 18 Lyman St., Westboro, MA 01581, USA
<i>Ystat</i>	1.28	Tel: (508)-366-1122 Ming Telecomputing, P.O. Box 101, Lincoln Center, MA 01773, USA
		Tel: (617)-259-0391

In this paper, we will describe and evaluate the automatic forecasting features found in current forecasting programs. Section 2 describes the different forms of automatic forecasting; i.e., the types of methodological decisions which can be made without requirement of significant input or direction from the user. Section 3 offers an evaluation of how effectively automatic forecasting features serve the needs of the forecasting practitioner. Section 4 provides a summary of our findings.

Our primary purpose is not to weigh the pros and cons of the various software packages; for example, we give little attention to file management capabilities, an important dimension of software selection. Rather our goal is to clarify for the practitioner of quantitative forecasting the software's virtues and limitations in the selection, evaluation and defense of a forecasting method.

Thirteen software programs (Exhibit 1) were selected for review. The common denominator of these programs is the emphasis upon single-equation modeling of time series data. (See Appendix A for a description of the selection process.) Not included are those packages whose primary orientation is econometric modeling or multivariate time series, methodologies designed principally for statistically sophisticated users. For a good review of econometric packages, see van Ness and ten Cate (1988).

The software on our list constitutes a market segment analogous to that of the autofocus compact in the 35 mm camera market. Purchasers of autofocus cameras tend to be photographic amateurs, for whom convenience and ease of use are major considerations. Such buyers rely on the camera's internal judgment to make key decisions regarding lighting, shutter speed and focus. In this fashion, users of automatic forecasting software may expect the program to determine many of the key settings required to bring an appropriate forecasting method into focus.

Within our market segment of forecasting packages, we found considerable variety in the menus of forecasting methods offered. To provide some standards for evaluation, we chose to focus our attention upon three common types of methods: exponential smoothing, ARIMA, including models with transfer functions for input variables, and single-equation regression, including dynamic regression for lagged variable and error terms.

2. The faces of automatic forecasting

What is a software program offering the practitioner under the rubric of automatic forecasting?

Our survey reveals a corpus of many faces. The most fundamental distinction to be made is that between automatic parameter optimization (APO) and automatic method selection (AMS).

2.1. Automatic parameter optimization

Automatic parameter optimization is pertinent to exponential smoothing, to ARIMA (including transfer function) models, and to dynamic regression. The basis of APO is an algorithm which iteratively searches for the optimal values of one or more constants (weights), in accordance with the particular specification.

In ARIMA and certain dynamic regression specifications, APO releases the user from the requirement to supply technical information such as initial parameter values. In the case of exponential smoothing, APO pardons the user from the substantial burden of searching for an optimal fit by trial and error.

If this is the program's only conception of automatic forecasting, the practitioner would still retain the responsibility to:

- (1) select a general method (smoothing, ARIMA, regression) and identify a suitable specification for that method;
- (2) designate certain fit settings, such as the portion of the data to be set aside for forecast evaluation and the statistical criterion (loss function) underlying parameter estimation.

The benefits of automatic parameter estimation nonetheless can be substantial, in terms of savings in time or reduction in technical demands. The practitioner of exponential smoothing, moreover, need not be concerned with rules for the selection of appropriate smoothing constants, rules which typically have lead to the assignment of low values (0.1 to 0.3) for the smoothing constants.

Indeed, Chatfield (1978) warns how dangerous it can be to designate plausible values for the smoothing constants. "Typical values quoted in the literature are often way off from the estimated values. [Hence], smoothing parameters should not be guessed at but estimated from the data." (p. 272) The danger can be insidious; for, as Newbold and Bos (1989) point out, assertions in the literature about the choice of smoothing constants are often predicated upon erroneous assumptions about the probability process underlying the time series.

Despite the advantages, the practitioner must recognize that there are variations on the theme of automatic parameter optimization, as a consequence of which the programs may produce very disparate estimates of the optimal smoothing or ARIMA parameters. Chatfield and Yar (1988), for example, show how a large divergence in smoothing weights among certain exponential smoothing programs can result from differences in the way starting values for the optimization algorithm are set. They warn the practitioner (p. 506) to "be aware of how dependent the results can be upon the choice of starting values."

On the other hand, when Makridakis and Hibon (1989) compared a variety of methods for choosing starting values (for non-seasonal exponential smoothing procedures), they found that the choice did not significantly affect post-sample accuracy, when accuracy is assessed over the 1001 series in the M-competition. To some degree, apparently, we may expect the optimization of smoothing weights to absorb different starting values in a way which preserves the accuracy of fit.

The default settings for fit criteria are another source of variation in the results of an optimization algorithm. These include the restrictions placed upon the range of permissible weights for the smoothing constants, the time period over which the fit of the designated method is evaluated, the number of iterations the program will attempt in the search for an optimum, and the type of optimization algorithm used (grid search or simplex for smoothing, least-squares or maximum likelihood for ARIMA). So, one cannot assume that APO results in a uniquely optimal set of parameter values. It would be unwise, in consequence, for the practitioner to accept APO results without being aware of the settings used to fit the model.

Among programs which offer exponential smoothing, APO is now virtually a standard feature. Indeed, most of the programs give the user the choice between an optimization algorithm and manual entry of values for the smoothing weights. *Time Machine* and *Number Cruncher* represent partial exceptions. While *Time Machine* contains a grid-search algorithm for optimizing the smoothing constant in Brown's single-parameter specifications -commonly called single and double smoothing -it requires user specification of the smoothing weights in order to implement a Holt- Winters specification. There would not seem to be a logical basis for this inconsistency. *Number Cruncher* consistently requests user input of smoothing weights, explicitly recommending values between 0.1 and 0.3, an example of the afore-mentioned concern of Newbold and Bos.

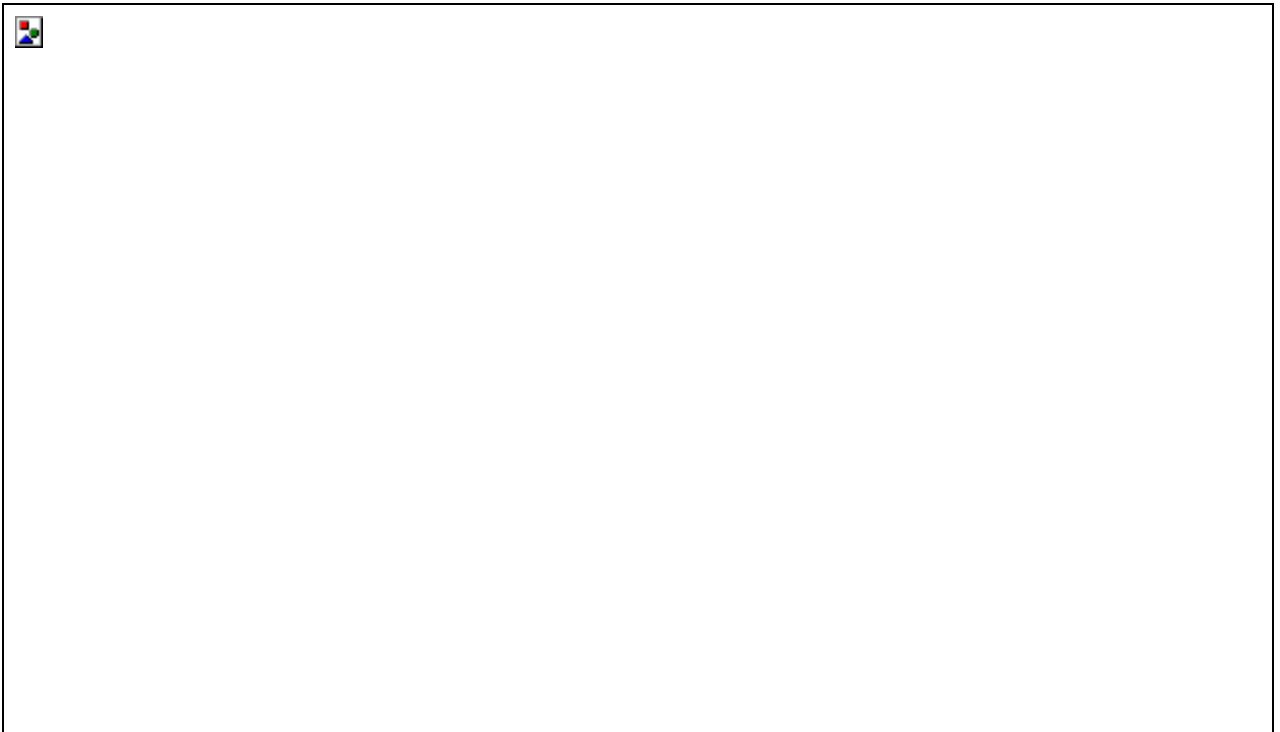
2.2. Automatic method selection

The selection of a forecasting method is the heart of the forecasting process. The experienced analyst will generally employ a winnowing procedure in which a number of preliminary methods are identified, estimated and evaluated until the choices are reduced to a single best or to a few acceptable specifications.

Graphs usually play an important role in the model identification process. The analyst views time plots and correlograms, among other graphics, to determine the presence and form of trend in the data, the length and type of seasonality, the strength of the autocorrelations, the timing and magnitude of outliers, and the need for transforming and differencing a time series. The process is not precise and experts can differ on interpretations of the graphics. Nevertheless, as Sharda and Rock (1985, p. 205) point out, graphics "promote user involvement in the [modeling] process [which] may lead to more insightful forecasting."

In about half the programs reviewed here, the practitioner is spared the need to read and interpret graphs for method selection. Indeed, the only demands on the user are to supply the time series, and (optionally) to designate fit settings. Then, literally at the touch of a key, the user will be shown a "best" forecasting method.

The approach to AMS varies both across and within software programs. We may classify the variants into five categories: rule-based logic (expert system), automatic specification tests, a unified framework, a forecasting contest, and 'all possible specifications'. Exhibit 2 provides a summary.



a - Key: 1: Exponential smoothing, 2: Standard least-squares regression, 2': Dynamic regression, 3: Univariate ARIMA, 3': ARIMA with transfer function models. *AUTO:*

Program performs procedure without user request, *USER*: Program performs procedure upon user provided instructions, *N.A.*: Procedure is not available in or not applicable to the program.

b	Within exponential smoothing, only single smoothing is offered.
<i>Autobox Plus:</i>	
<i>Autocast II:</i>	User can select the MSE or MAD as the fit criterion and choose either the standard Grid Search or the more computationally efficient Simplex search algorithm for optimizing the smoothing constants. The program also offers a "variance analysis" -a comparison of the variability of the level, first and second differences of the time series -to help the user identify an appropriate option from the trend menu.
<i>Forecast Plus:</i>	Provides three choices of "period of optimization": earliest 25% of the data, most recent 25%, and the entire time series.
<i>Forecast Pro:</i>	Uses the Bayesian Information criterion (BIC) to designate a best method from among those already fit; however, minimization of the MSE is the criterion employed in the fitting algorithms.
<i>Forecast!:</i>	Brown's single and double smoothing methods are offered (no Holt-Winters) but these can be automatically applied to the seasonally adjusted data.
<i>Forman:</i>	Within ARIMA, only non-seasonal specifications are offered. For seasonal data, the user may apply ARIMA specifications to the seasonally adjusted data.
<i>Number Cruncher:</i>	Offers robust weighting procedures within the regression option.
<i>PC/Sibyl:</i>	Offers linear and damped trend smoothing specifications for the original and seasonally-adjusted data, as well as Winters' three-parameter method.
<i>Smartforecasts II:</i>	Methods in the forecasting tournament are fit to minimize the "average forecast error", which is an average MAD in a (rolling) post-sample simulation.
<i>Smooth:</i>	Has independent period-of-fit and period-of-optimization settings.
<i>Time Machine:</i>	APO is offered for single and double smoothing but program requires the user to enter the three smoothing weights for Winters' method.
<i>Trendsetter Expert:</i>	The user may choose the starting point but not the ending point of the period of fit. Hence, it does not permit the most recent data to be reserved for post-sample evaluation.
<i>Ystat:</i>	Within ARIMA, only non-seasonal specification are offered. For seasonal data, the user must apply ARIMA specifications to the seasonally adjusted data.
<i>Methodological note:</i>	Dynamic regression refers to a regression specification that includes, in addition to the explanatory variables of standard regression, lagged values of the dependent variable and/or lagged error terms, Transfer function models are ARIMA

models that contain one or more input (explanatory) variables. While a transfer function model can be reduced mathematically to a type of dynamic regression model called an Armax model (Fildes, 1985), dynamic regression and transfer functions represent different approaches to the design and estimation of a forecasting model, and their application to a given data set can produce dissimilar forecasts.

2.2.1. Rule-based logic

Rule-based logic is used to emulate the analytical process of the human expert; in effect, to reproduce human reasoning in the selection of a method.

The *Forecast Pro* 'expert system' uses a few basic rules to select a method from among exponential smoothing, univariate ARIMA, and dynamic regression. With but one time series to analyze, the rules are predicated upon the " autocorrelation structure" as well as stability of variance between the first and second half of the series. (After differencing the series as necessary to achieve stationarity, the program fits a states-space model of autoregressive terms. If only low-order terms (AR1 or less) are significant, the autocorrelation structure is deemed to be 'low or weak'; if higher-order terms are also significant, the auto-correlation structure is termed "strong".)

The expert will recommend ARIMA if it finds the order of the autocorrelation structure to be " strong or moderately strong". And when the autocorrelation structure is low, the expert will still recommend ARIMA if a variance stability (Chow) test reveals that the series is "stable". The combination of weak autocorrelation structure and unstable variance leads the expert to recommend exponential smoothing. (The expert will also recommend smoothing when the sample size is very small.)

When there is at least one other time series with which the variable to be forecast is " significantly" correlated, the *Forecast Pro* recommendation will always be dynamic regression. *Forecast Pro* uses "dynamic regression" in a very general sense, to refer to a regression specification which includes dynamic terms, by which it means lagged values of the dependent variables and/or lagged error terms. In the spirit of ARIMA, it drops the assumption of standard OLS regression that the error term represents a random (white noise) process, and instead permits explicit modeling of autocorrelation in the errors.

Forecast Pro does not cite specific empirical support for its rules, nor does it enunciate these rules with precision. The user is not sure of the boundary between strong and weak autocorrelation, or between homogeneous (stable) and unstable behavior. One wonders, moreover, how the expert handles situations which may be " too close to call".

2.2.2. Automatic specification tests

Autobox Plus employs a sequence of classical significance tests to find an appropriate ARIMA specification. The program will first (optionally) transform and difference the data in pursuit of a stationary series and tentative ARMA identification, called a starter model. Alternatively, the practitioner can designate a personal starter model.

The heart of the procedure is a battery of automatic "diagnostic checks" on the starter model, tests which determine the need to (a) retain existing ARMA terms, (b) incorporate additional

ARMA terms, (c) perform further differencing of the data, and (d) (optionally) adjust for (several types of) interventions in the time series.

If the user wishes to include one or more independent variables in the analysis, further tests are conducted to identify a suitable transfer function. The ultimate goal of the testing process is the discovery of a specification whose errors are informationless (white noise) over the period of fit. In this regard, the program monitors the residuals after each test and reports on how successful each modeling adjustment has been.

Automatic specification tests also underlie *Forecast Pro's* search for a particular ARIMA specification, as well as for an acceptable dynamic regression specification. In the former, the program starts with a low-order ARMA model (separate specifications for the regular and seasonal components) and gradually builds up terms until the BIC no longer significantly improves. The program's claim that it has "made Box-Jenkins available with two keystrokes" is essentially correct. But not entirely correct.

Recommended transformations will not automatically be implemented. Rather, the user must switch into data editing mode to make the transformation, and then ask the expert system to repeat its task, this time on the transformed data.

In dynamic regression, *Forecast Pro* automatically performs tests on behalf of each explanatory variable. These tests determine the statistical significance of (a) the first-order terms (X), (b) the second-order terms (X^2), (c) all lags of the dependent variable, and (d) key lagged-error terms (equivalent to a generalized version of the well-known Cochrane-Orcutt (1949) procedure). Ambitious as this process is, it omits investigation of lags on the explanatory variables, as well as interactions among the explanatory variables. But, as one can see, automatic method selection in a regression context is very involved.

2.2.3. Unified framework

In *Autocast II*, the practitioner of exponential smoothing may request a particular smoothing specification (of the Holt-Winters variety) by making a selection from the trend menu (Constant-level, Damped, Linear, Exponential) and from the seasonality menu (Multiplicative, Additive, Non-seasonal). For anyone of these twelve specifications, the program will respond by performing automatic parameter optimization. But the practitioner can also bypass the trend and seasonality menus in favor of an automatic method selection.

In automatic mode, *Autocast II* employs neither rule-based logic nor specification tests. Rather, Gardner, its developer, showed (1985) that the 12 smoothing specifications (4 choices for trend x 3 choices for seasonality) are special cases of a unified framework based upon 4 parameters: level, trend, seasonal, and trend-modification parameters. For example, depending upon the value of the trend-modification parameter (ϕ), the unified method yields an exponential trend ($\phi > 1$), linear trend ($\phi = 1$), a damped trend $0 < \phi < 1$), or

no trend ($\phi=0$). In effect, AMS in *Autocast II* is the application of automatic parameter optimization to this unified framework.

For monthly and quarterly data, however, AMS in *Autocast II* always assumes seasonality of a multiplicative form, effectively excluding 3 of 12 smoothing options from the unified framework. This restriction is built in to reduce computation time. The practitioner should realize, however, that specifications based on additive seasonality are sometimes the more appropriate for the data at hand. Accordingly, if a time plot of the time series does not clearly indicate that the seasonal spread increases as the level of the series increases (the multiplicative pattern), the practitioner would be wise to compare the AMS forecast with one based on the additive seasonality option from the seasonal menu. For an example, see Chatfield (1978).

2.2.4. Forecasting contest

Two of the programs reviewed here permit automatic selection of a forecasting method by conducting a contest among a prescribed group of forecasting procedures. A winner is chosen, and its forecasts are reported.

In *Smartforecasts II*, five methods do battle in a "tournament of automatic forecasts": two specifications for horizontal data (simple moving average and simple exponential smoothing); two for trend (linear moving average and double exponential smoothing); and one for seasonal data (Winters' exponential smoothing). The last is a restricted version of Winters' three-parameter method, the restriction requiring the level, trend, and seasonal smoothing weights to be equal. The method generating the lowest average (absolute) forecast error is declared the winner.

This forecasting contest is quite narrowly drawn. The moving average and Brown's methods are virtually identical in the way they smooth the data. More importantly, there is only one procedure for seasonal data, Winters', and the tournament's single-parameter version is certainly a significant restriction on the method's performance. It is not uncommon to find a time series with a stable seasonal pattern and erratic trend. For such a series, Winters' method, if unrestricted, would likely result in the assignment of a seasonal weight close to zero and a trend and/or level weight closer to unity.

The competitors in *Trendsetter Expert* include four techniques for non-seasonal data: (adaptive rate) single and double exponential smoothing, and least-squares trend lines for the most recent half as well as the entire data series. For a seasonal time series, the four methods will be implemented on the deseasonalized data.

Unlike *Smartforecasts II*, the *Trendsetter Expert* contest does not necessarily result in a winning specification. Rather the program also "combines the four sets of forecasts to arrive at a single set of forecast values." The procedure for the combining of forecasts is strictly withheld from the user, as the creators consider it proprietary. (One may surmise that both simple and weighted averages of the forecasts are derived and compared against an accuracy criterion. It is possible too that rule-based logic is used to select an appropriate combination of forecasts.) In effect, this program represents the ultimate

black box: The user is not told (1) which technique or combination of techniques has been selected, or (2) what criteria have been employed for method selection.

2.2.5. "All possible specifications"

The *Time Machine's* "Suggest-a-Model" and *Number Cruncher's* "ARMA Search" assist the practitioner to find a specification for a univariate ARIMA model. After preliminary diagnoses are made of the need to transform and difference the series to achieve stationarity, Suggest-a-Model will automatically fit all first and second order regular ARMA specifications [except an ARMA (2,2)] to the level and to various differences of the data. The practitioner is shown up to three "likely" or "possible" specifications, together with the numeric starting values for the estimation algorithm. This is a comprehensive list for non-seasonal data, but excludes consideration of seasonal ARMA terms. The approach of multiple recommendations is a good idea, however, since a variety of ARIMA specifications can fit the data equally well and still generate very divergent forecasts.

Number Cruncher tries all ARMA (p, p- 1) specifications, after asking the user to request a maximum order, p, and selects the lowest-order specification whose residual sum of squares is within a user-designated percentage of the lowest RSS. This approach leads to anything but parsimonious specifications for seasonal data; monthly data, for example, will require the user to set the maximum order at 12 or above. Hence, it is practical only if performed on deseasonalized data, a significant limitation.

2.3. Automatic data decomposition

Three programs do not contain any of the versions of AMS, but do offer, virtually automatically, a preliminary data decomposition. Based upon the classical decomposition of a time series,

Smooth, *PC/Sibyl* and *Forecast!* analyze the trend/cycle and seasonality of a series. Their "front-end analysis" enables the practitioner to view the seasonal indexes, and to choose either the original or seasonally adjusted data for subsequent modeling. The later would be a significant enhancement in *Time Machine* and *Number Cruncher* whose automatic ARIMA features are unwieldy for seasonal data. Several other packages facilitate analysis of the seasonally adjusted data by offering a classical or Census decomposition as one of the forecasting methodologies.

Further, *Smooth* and *PC/Sibyl* analyze the volatility of a time series by calculating the average absolute change per time period (AAC), and identifying the portion of the AAC attributed to trend, cycle, and seasonality. Such a feature may remind the practitioner not to overlook the need to account for seasonality and business cycle sensitivity in choosing a methodology.

2.4. Settings for fit criteria

A fit criterion defines what is meant by "best" in a process for choosing the best forecasting procedure. Most practitioners are familiar with the least-squares criterion for

fitting a regression line. On this basis of fit, "best" means lowest mean squared error (MSE) while giving equal weight to all values in the time series, recent and distant.

A fit criterion has two principal dimensions: a statistical standard (also called a loss function) and a time-period setting. Minimization of a squared error loss function is a virtually universal statistical standard for fitting regression and many time series methods; and, with two exceptions, the programs reviewed here perform parameter estimation exclusively on this basis. The smoothing algorithm in *Smartforecasts II* uses an absolute error loss function, while *Autocast II*, an exponential smoothing program, gives the user an option, MSE or MAD, as the statistical standard for the optimization algorithm.

In exponential smoothing, the lack of choice in loss function is not necessarily important. The results of a recent Makridakis and Hibon study (1989) imply that little or nothing of practical significance would be gained in terms of post-sample accuracy by offering alternatives to MSE, such as minimization of a MAD, MAPE, or Median Error, at least for non-seasonal exponential smoothing. But ARIMA and regression are another matter. Weiss and Anderson (1984) had found cases in which ARIMA models estimated to minimize the MAD and MAPE showed improved forecast accuracy. Dielman (1986) showed, that relative to least squares, least absolute value regression often improved post-sample accuracy, especially for the type of "fat-tail" distributions which are prone to generate outliers.

Unfortunately, automatic forecasting software to date has virtually ignored robust methods of estimation in regression. *Number Cruncher* is the sole exception among the programs on our list, offering several robust weighting procedures in its multiple regression option.

While model estimation is rooted in a squared-error loss function, we see a bit more variety in the statistical standard employed to select a "best" specification from among alternatives estimated. In *Smartforecasts II's* automatic forecasting tournament, the winner is designated by virtue of having the lowest MAD in post-sample testing (which is the same criterion used to estimate each method in the tournament). Forecast Pro's choice of best specification is based on the Bayesian information criterion (BIC), calculated over the period of fit. The BIC more severely penalizes method complexity than does its counterpart, the Akaike information criterion (AIC), according to Schwarz (1978).

The other dimension of a fit criterion is the time-period setting. Many programs allow the user to define the beginning and end of the period of fit: that portion of the historical data used to estimate the coefficients of the forecasting equation. The remaining portion of the historical time series is reserved for a post-sample evaluation of forecasting accuracy.

The opportunity to select a period of fit is of paramount importance in a forecasting program. For one, practitioners may wish to exclude some data from analysis, either out of concern for the accuracy or relevance of a portion of the data, or in order to compare the performance of a forecasting method over different time periods.

Secondly, it is wise to evaluate the out-of-sample accuracy of a forecasting method, particularly when the method is fit to minimize the error (MSE) over the period of fit. The empirical evidence, according to Belsley (1986, p. 45) "shows unequivocally that models selected on the basis of best sample-period fit are a lamentably poor guide to those models that best predict post-sample evidence." Finally, a period of fit setting enables the user to conveniently update the preferred model (i.e., to reincorporate the portion of the data held-out) when ready to perform ex ante forecasting.

In the absence of a programmatic facility to automatically select a period of fit, the practitioner must undertake potentially tedious file manipulations to configure the desired period of fit. In this circumstance, it is more likely that the practitioner will simply not bother with out-of-sample evaluation.

The programmatic choices for period of fit are shown in Exhibit 2. *Forecast Plus* and *Smooth* offer interesting variations on a period of fit setting. *Forecast Plus* allows the user three choices for the "period of optimization", which is that portion of the historical data used in the optimization of the smoothing weights or ARIMA model coefficients: the entire sample, the initial 25% of the data, and the most recent 25% of the data. The latter option permits the practitioner who senses a recent change in the historical data pattern to optimize parameter values based upon the more current observations. It is no substitute, however, for a period of fit setting; without the latter, the user lacks a convenient way to change the time origin from which forecasts are generated.

In this sense, *Smooth* offers the most flexibility. It permits the user not only to choose a period of fit but also to designate a sub-period of optimization, which it calls the "test set". If we have a time series of 72 months, we can first ask *Smooth* to hold out the most recent 12 months by designating 1-60 as the period of fit. We then can decide to optimize parameter values over the months 13-60 by designating 13 as the start of the test set.

Allowing for such a learning or adjustment period can be important in exponential smoothing.

3. Does automatic forecasting facilitate good forecasting practice?

Having illustrated the various forms that automatic forecasting has taken, we will now address some claims and concerns about the safety and effectiveness of automatic forecasting programs.

A prominent marketing theme underlying automatic forecasting programs is that even the unsophisticated analyst can expect to obtain good forecasts. In one brochure, we read that the program "does in minutes the same thing a forecasting expert can do in a day or week, for \$100 an hour." One thinks again of the autofocus camera: If we simply point and shoot, we can expect the subject to develop with clarity and depth.

Of course, some exaggeration in advertising is expected; hence, claims will be discounted. No self-respecting forecasting practitioner will believe that a solution to the

company's forecasting problem can be effected in a few key strokes. It is nevertheless important to ask what the practitioner can and should reasonably expect from the product. (It would be easier to define expectations in this realm had we already in hand a commonly accepted checklist for judging the performance of professional forecasters.)

We propose the following criterion for evaluation: does the software permit if not encourage good forecasting practice in the selection, evaluation and presentation of a suitable forecasting method? And as a logical addendum: what technical background must the user possess to be able to utilize the software in the appropriate manner?

3.1. Method selection issues

The automatic forecasting programs reviewed here embrace those branches of the forecasting methodology tree which deal with single equation modeling of time series data. Granted that these branches are fruitful for forecasting, we may ask whether the software provides a wholesome variety of pickings!

3.1.1. Does the software permit eclectic forecasting?

There is little disagreement among educators that forecasting should be eclectic; that is to say, the practitioner should examine and compare a variety of methods for generating forecasts.

Armstrong (1985, p. 63) expressed the belief that "for a given research budget, it may be better to use a number of different approaches, even though crudely done, than a single approach done well."

If forecasting should be eclectic, the present state of automatic forecasting software must be considered confining. Virtually all the programs which automate method selection limit the user to a single forecasting method, as well as to a restricted number of specifications within this methodology.

In *Autocast II*, automatic method selection is based entirely upon exponential smoothing; and as noted earlier, to a reduced set of those smoothing specifications which the practitioner can request from the program's trend and seasonality menus. Similarly, only smoothing specifications are allowed to compete in *Smartforecasts II's* tournament, and to be combined by the expert in *Trendsetter Expert*. *Time Machine* and *Number Cruncher* offer regression, ARIMA, and some smoothing methods in manual mode but their AMS facility is limited to univariate ARIMA. In automatic mode, *Autobox Plus* applies only the Box-Jenkins methodology; however this methodology is not restricted to univariate ARIMA models. Rather, via the transfer function capability, the program will automatically determine if designated input (explanatory) variables can significantly enhance a univariate model.

The limited range of automatic method selection restricts the practitioner's options, and gainsays potentially valuable insights. The user of automatic smoothing programs cannot tap a class of modeling refinements that is rooted in analysis of autocorrelation and cross

correlation, precepts of the Box-Jenkins approach. Automatic ARIMA programs typically offer limited flexibility in the way one can represent trend and seasonality in the data.

In addition, the limited domain of the automatic forecasting programs restricts the practice of combining forecasts; a practice which, according to Clemen (1989, p. 567) is "practical, economical, and useful ... and many empirical tests have demonstrated the value of composite forecasting." Only *Trendsetter Expert* and *PC/Sibyl* among the programs reviewed here automatically combine forecasts. *PC/Sibyl* will combine forecasts by calculating the average of up to five user-designated methods. *Trendsetter Expert*, as noted, presents a combined forecast, but the user does not know the components.

We echo Clemen's call for the more widespread inclusion in the software of combined forecasting: this with the caveat to the practitioner that a mechanical combination of forecasts should never substitute, as Oiebold argues (1989, p. 591), for serious efforts to find a satisfying model.

The distinction between a method type and specification must be stressed. As noted, the present-day versions of AMS effectively select a specification from a single method type. Accordingly, the practitioner really needs to decide in advance of the commitment to a software product that the methods within the program's automatic mode are adequate for the organization's needs. But this decision may require a sophisticated understanding of the breadth of forecasting methodologies. Bell's words about the "closed-world problem of the expert system not knowing what it does not know" (1985, p. 16) are provocative. The ingenuous practitioner will not know what he is missing.

3.1.2. Can explanatory variables be utilized?

In a widely-cited commentary on the M-competition, Lopes (1983, p. 271) expressed the wish that developers build into forecasting programs "some of the beliefs that knowledgeable forecasters bring to their art ... [including] substantive beliefs about the system that generates the data. It seems that extrapolative methods would be strengthened by making provision for causal or explanatory information to be used, when such is available."

There has been little heed of Lopes' recommendation. Only two of the programs reviewed here – *Autobox Plus* and *Forecast Pro* – make provision for the automatic incorporation of explanatory variables. Eight of the other programs perform regression, but the regression facility is not part of the system of automatic method selection. So the practitioner cannot readily answer the question: can the forecasts be improved by using an explanatory variable?

The preceding paragraph should perhaps be considered less a criticism of the software than a lament on the limitations of the present state of the methodological art. For example, to the practitioner of exponential smoothing, one cannot offer a commonly accepted, systematic means for introducing potential explanatory variables.

As noted, practitioners of ARIMA modeling can utilize the transfer function methodology to build in information on selected "input" variables. However, this methodology can result in "complicated lag structures and, hence, models which are hard to understand." Fildes (1985, p. 506). Being data driven, moreover, it does not permit the practitioner to design a model around a set of core variables -variables whose inclusion is warranted on theoretical grounds, even if their contribution does not meet the tests of statistical significance.

3.1.3. Do the software's modeling decisions promote understanding?

The prevailing ethic of automatic forecasting software seems to be to hand the practitioner a set of forecasts as expeditiously as possible, hoping he will raise as little fuss as possible about the means by which the forecasts were obtained.

This disposition is unfortunate. Software has great potential for teaching the practitioner the ropes of good model-building. To do so, however, the automatic method selector must preach what it practices. Its attitude should be to assist the practitioner to find an acceptable forecasting method, rather than to simply show the practitioner what it has wrought on its own.

3.2. Method evaluation issues

When an analyst identifies a plausible forecasting method from an examination of the data, it is standard forecasting practice to evaluate the pattern of the forecasting errors, both within and out-of sample. The practitioner must realize that it is especially critical to evaluate those methods selected automatically by the software. While the program's recommendation is likely to be the best among the alternatives considered, these alternatives are limited both in the scope of methods considered as well as in the variety of fit criteria employed. So the method deemed to be the best among this limited number of options may not be good enough. Evaluation may detect the need for further refinements, or may reveal that the accuracy of the forecasts is not much better than that of a simpler model.

A selection of error evaluation capabilities is provided in Exhibit 3. One can see that the range of offerings is wide. Certain of the programs lack a mechanism for serious error evaluation. Others are comprehensive, at least in certain aspects of the task.

At the one extreme, *Trendsetter Expert* lacks any analysis of forecast errors. The rationale may be that little purpose would be served by such analysis, since the program does not necessarily generate its forecasts from a single specification. Rather, *Trendsetter Expert*, as with its predecessor *Wisard*, claims that its "forecasts are superlative" when assessed on the 111 time series of the earliest M-competition (Makridakis and Hibon, 1979). The specific claim is that *Trendsetter Expert* was more accurate than the best of the 24 techniques (in the M-competition) for 88% of the time series, and better than average for the rest. The practitioner should not view this sort of claim as a warranty. Indeed, the professional commentary on the results of the M-competition, especially those in Armstrong and Lusk (1983), reveal how difficult it is to declare winners and losers. On the other hand, the evidence from the M-competition and elsewhere is quite

strong that combinations of forecasts - the essence of *Trendsetter Expert's* approach to AMS - often turn out to be superior to every one of the individual forecasts.



a *Key*: Times series methods applies to all methods but standard regression. *AUTO*: Program performs procedure without user request, *USER*: Program performs procedure upon user-provided instructions, *N.A.*: Procedure is not available in or not applicable to the program, *FIXED*: Fixed simulation. Forecasts are made from a single origin. Program automatically compares forecasts against known actual values, and displays the same error statistics as are shown for the period of fit, *ROLLING*: Rolling simulation. The program generates forecasts using each post-sample period in succession as the forecast origin. In addition to the results of *FIXED* simulation, the program automatically displays error statistics sorted by lead time, *NAIVE*: A method in which the value for the current time period serves as the forecast for the next time period; i.e., a model which forecasts 'no change' between the current and following time period. *NFR*, for Naive Forecast Regular, is the traditional basis of comparison for non-seasonal data. *NFS*, for Naive Forecast Seasonal, first seasonally adjusts the data to remove the seasonal component of the carryover from one time period to the next.

b Bias: *ME*: Mean error, *MPE*: Mean percent error, *BETA*: Slope coefficient in regression of actual vs. forecast values. Precision: *MAD*: Mean absolute deviation, *MAPE*: Mean absolute percent error, *RMSE*: root mean square error, *SE*: Standard error (= *RMSE* adjusted for degrees of freedom), *R2*: *R*-square statistic (in regression of actual vs. forecast values), *AIC*: Akaike

Information Criterion, *BIC*: Bayesian Information Criterion.

c *Autocast II*: The naive comparisons measure the precision (MSE) of a smoothing specification in relation to that of a naive method. For quarterly and monthly data the naive forecasts have been adjusted for seasonality by the method of classical decomposition. The user is given several options to adjust for outliers - options which replace an actual data

	point with a value closer to that being forecast.
<i>Autobox Plus:</i>	Automatic intervention detector in ARIMA provides means of evaluating and adjusting for the effects of outliers.
<i>Forecast Plus:</i>	Although outliers are not identified, the user can request box-plots of the residuals for effective outlier identification.
<i>PC/Sibyl:</i>	Outlier detection and adjustment is performed in the Harmonic Smoothing procedure, although not in the mainstream smoothing methods of Brown, Holt, and Winters.
<i>Smartforecasts II:</i>	The APE or average forecast error is a measure based on the absolute values of the errors in a rolling simulation; however, Smartforecasts II does not separately tabulate the forecast errors by lead time. The ACF lacks the numerical values of the coefficients.
<i>Smooth:</i>	The ACF lacks the numerical values of the coefficients.

3.2.1. Do the programs permit residual analysis?

With the exception of *Trendsetter Expert* and *Forecast!*, the programs all provide time plots and autocorrelation functions of residuals. Most do so automatically. In *Time Machine* and *Smartforecasts II*, however, the practitioner must first request that the residuals be saved, and then revert to a graphics menu to plot the newly saved data. This extra activity is a minor, perhaps negligible, inconvenience. On the other hand, the practitioner without background or statistical knowledge may not know the importance of analyzing these graphs. Hence, if the graphs are not provided automatically, the initiative to obtain them may be lacking.

The autocorrelation functions in *Smartforecasts II* and *Smooth* lack numerical indications of the size and statistical significance of the autocorrelation coefficients, omissions which seriously restrict the usefulness of the ACF. This is a simple problem to correct.

The omission of an outlier detection facility can be particularly insidious in AMS programs, since the practitioner may never know how sensitive the chosen specification is to data anomalies. Yet, only three of the programs, *Autobox Plus*, *Autocast II*, and *Ystat* incorporate automatic outlier detection. In one other program, *Forecast Plus*, the user can request box plots of the residuals, which are particularly valuable for outlier detection, as demonstrated by Courcelle and Tashman (1989).

What you can do to account for the effects of outliers varies considerably among the programs. *Autocast II* permits the replacement of an outlier with a value estimated from a smoothing specification. The practitioner must decide however whether the freak value is to be considered an outlier or is a feature of the process at work. *Ystat* allows known or suspected causes of outliers to be modeled as dummy variables. *Autobox Plus* can automatically estimate potential (known or unknown) outlier effects jointly with the parameters of an underlying ARMA model. Tsay (1988) describes and illustrates the approach.

3.2.2. Do the programs provide broad perspective on accuracy?

The practitioner should be given a broad picture of the direction (bias) and magnitude (precision) of the forecast errors. For convenience, error measures should be reported both in absolute (volume or currency) and relative terms (percent errors). Moreover, it is good practice to compare the performance of a chosen method with that of a naive forecasting method.

For non-seasonal data, a naive forecast is usually understood to be a forecast of no change from the present (a random walk). For seasonal data, there are various naive specifications: *Autocast II*, for example, has three different naive methods depending upon the form of seasonality.

The list of error measures in Exhibit 3 shows a tendency for the software to skimp on measures of bias -6 of the 13 report none at all -as well as upon relative measures of precision -missing from 4 of the 13. And only 2 of the programs, both specialists in exponential smoothing, provide comparisons between a chosen method and a naive method.

Practitioners might also benefit from the adoption of a standard naive model for seasonal data. We propose the following: compute the naive forecast for any month (quarter) by adding the data value of the same month last year to the change between the immediate past month and its own predecessor last year. So a naive forecast for June 1991 would be the sum of the data value for June 1990 and the change from May 1990 to May 1991. This procedure is equivalent to differencing the series once both regularly and seasonally.

3.2.3. Do the programs offer post-sample simulations?

Simulation is an effective means for evaluating the accuracy of a forecasting method beyond the period of fit. In a fixed simulation -also called a static or constant-origin simulation -the final time period of the period of fit serves as the origin for all forecasts. From this point, forecasts are made for each of 1 to m steps (time periods) ahead, where m stands for the length of the simulation.

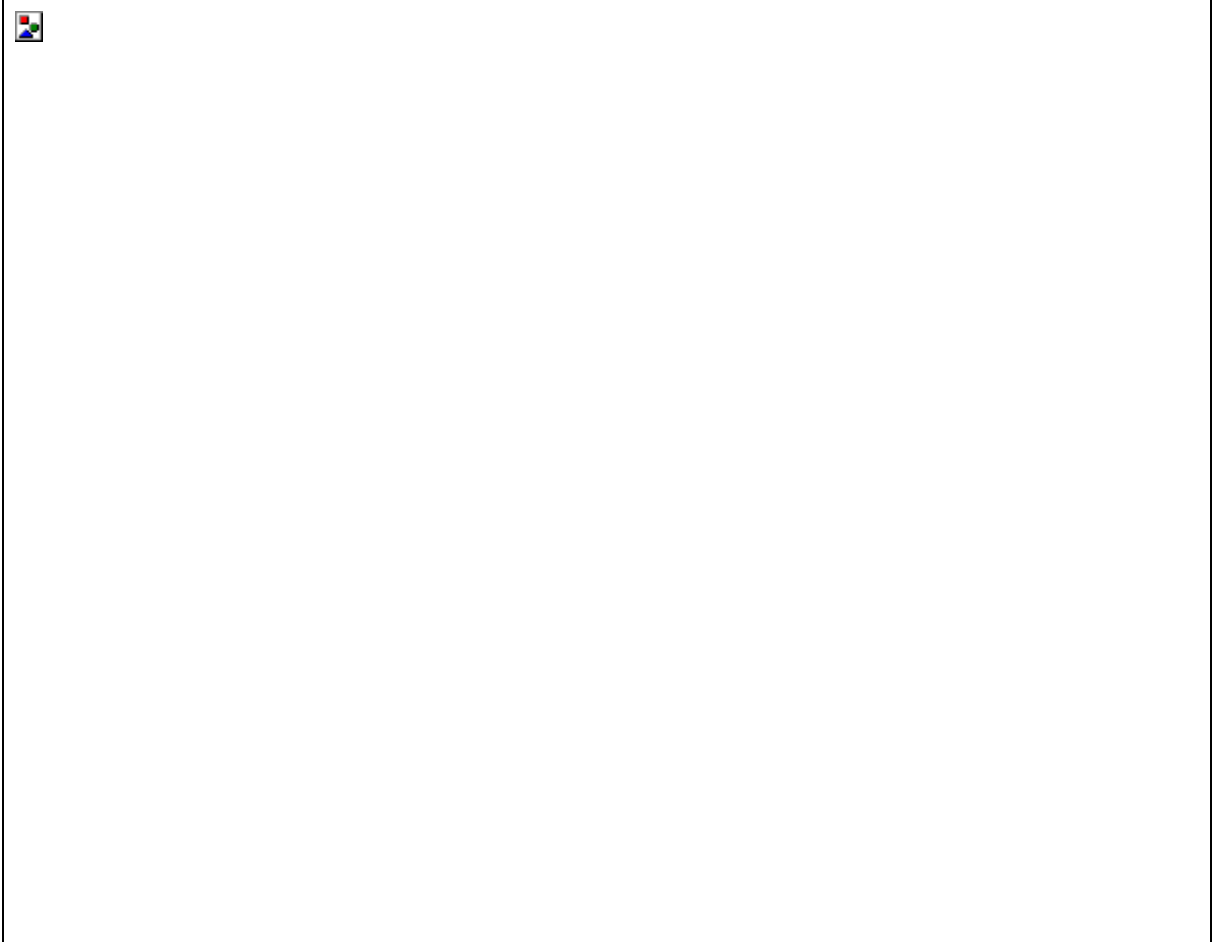
While a fixed simulation gives you one forecast and hence one forecast error for each step ahead, a rolling (or sliding or dynamic) simulation generates a distribution of forecast errors at each lead time. It accomplishes this as follows: after forecasting 1 to m steps ahead with the initial origin, the program shifts the origin forward one step in time, and generates a new stream of forecasts, 1 to $m-1$ steps ahead. This process is repeated for all $m-1$ forecast origins. The result is $m-1$ one-step-ahead forecasts, $m-2$ two-step-ahead forecasts, and so on down to a single m -step-ahead forecast. In this way the rolling simulation can not only supply all the information of the fixed simulation but also can reveal patterns of deterioration in the forecasts as the lead time increases.

Only six of the programs perform a post-sample simulation, and only two of these offer a rolling simulation. Lacking the simulation capability, the practitioner is liable to be overly confident about the future performance of a forecasting method.

3.2.4. Is there consistency between standard regression and the time series methods?

Notable inconsistencies exist in the way the software treats standard regression in comparison to the time series methods. These are manifest in the opportunity to perform post-sample simulations, in the composition of error measures, and in the generation of forecasts. Exhibit 4 summarizes the pertinent features of the regression-modeling facility. Only *Forman* and *RC/Sibyl* extend their (fixed) simulation capability to regression. In *Autobox Plus* one can simulate the post-sample performance of ARIMA models, both with and without input variables; however, the program's standard least-squares regression facility does not allow this operation. In *Ystat*, simulations can be automatically implemented for smoothing and ARIMA methods, but again this feature does not extend to regression.

Among the ten programs offering both regression and time series methods, only *Forecast Pro* and *Forman* display the very same set of error measures for both, which facilitates comparison of forecasting performance. Such comparisons can be quite cumbersome in some of the remaining programs. In *Forecast Plus*, *Smartforecasts II*, and *Time Machine*, the two types of methods share not one fit statistic in common.



These inconsistencies may be a matter of tradition - classical regression has its roots in the analysis of cross-section rather than time series data - but they are unwarranted in a forecasting package.

The practitioner of regression needs to appreciate the likelihood that the use of times series data will lead to the violation of one or more of the standard assumptions rationalizing the method of ordinary least squares. Hence, residual diagnostic tests are especially important. In this regard, we find that the software offers virtually the same opportunities for residual analysis (time plots, ACF, outlier detection) in regression that is provided in time series options. (So the columns on residual analysis in Exhibit 3 carry over to Exhibit 4 as well.) The main exception is *Number Cruncher*, which automatically displays a residual ACF for time series methods but not for standard regression.

When a forecasting program offers methods which include explanatory variables such as regression (standard or dynamic) and transfer functions, it should be expected to provide an efficient procedure for (a) projecting values of these inputs, and (b) converting the projections into forecasts for the dependent variable.

Exhibit 4 reveals (see the column entitled "Entry of projections for explanatory variables") that of the ten packages containing regression, only two provide such a feature. *Smartforecasts II* will allow the user to apply any of its univariate methods to project an explanatory variable, with the program automatically calculating point and interval forecasts for the dependent variable. *Ystat* will do likewise, but allows only two projection methods – linear trend and linear moving average – to be automatically applied to an explanatory variable. The transfer function option in *Autobox Plus* allows the user to automatically generate univariate ARIMA forecasts for any of the input variables, a procedure which supplies an added attraction for the calculation of prediction intervals, as discussed in the next section.

Most of the programs permit the user to make keyboard entries of assumed future values for the explanatory variables. In doing so, a program provides merely an elementary spreadsheet function, but fails to link its extrapolative methods to its regression capability. The standard regression options in *Time Machine* and *Autohox Plus* give the practitioner no option but to calculate regression forecasts manually, i.e., outside the program's regression routine. In the former, the omission is mitigated somewhat by the presence of an internal spreadsheet for implementing such calculations. In the latter, the philosophical emphasis is upon use of the transfer function, not standard regression, to incorporate input variables.

3.2.5. Do the programs calculate confidence and prediction intervals in regression?

Since both confidence and prediction intervals for regression forecasts are standard fare in forecasting textbooks, the practitioner might expect these calculations to be done by the forecasting software.

The final two columns of Exhibit 4 reveal a lack of consistency in program offerings. Of the ten regression programs, three provide point forecasts only and five produce interval forecasts. When an interval forecast is shown, the user frequently is not told (*Forecast Pro*, *Ystat*, and *Number Cruncher*) whether it is a confidence interval for the mean of a probability distribution or a prediction interval for a new observation. If one of the two

has to be chosen, the prediction interval would be preferable to the practitioner, and all but *Ystat* appear to have made this choice.

In *Forecast!*, a menu gives the user the choice to request either confidence intervals, prediction intervals or both. This is an admirably simple option, and a standard for the other packages to consider.

The confidence and prediction intervals we see reported for regression forecasts may be called conditional intervals; that is, they assume that the data input for the explanatory variables are either known, actual values or are projected without error. Such an interval is certainly too narrow, probably much too narrow (Ashley, 1983), when applied as an expression of the total uncertainty underlying a forecast. In contrast, the user of *Autohox Plus'* transfer function can obtain an unconditional prediction interval, one which reflects uncertainty in the (univariate ARIMA) projections of the input variables in addition to the inherent randomness in the time series to be forecast.

3.3. Forecast presentation issues

"Gaining acceptance for a forecast", writes Adams (1986, p. 138) "is often an educational process showing management that a realistic and consistent picture lies behind the forecast." The wise practitioner, therefore, will strive to demystify the forecasting process, and to demonstrate that the results are a logical progression from the past and current data. "Explanation is extremely important", Zellner (1988) maintains: "No one will accept the predictions from a black box."

3.3.1. The forecasting formula

A good test of whether a forecasting method is simple enough to gain managerial acceptance, says Armstrong (1987, p. 541), is that practitioners "should be able to calculate the forecasts by hand [as well as] describe the method to someone else." To do so, one needs to be shown the forecasting formula, the term used by Gilchrist (1976) for the equation which generates the forecasts for a particular method.

The practitioner can use the forecasting formula not only to demonstrate to management how the model works in issuing its forecasts, but also to test the sensitivity of the forecasts to new assumptions and conditions. When combining forecasts, moreover, the forecasting formula may offer insight into how the whole relates to the sum of its parts.

With standard regression and curve-fitting methods, the forecasting formula is merely the fitted equation. Exponential smoothing and ARIMA procedures, however, require assembly of the forecasting formula from the estimated coefficients - the smoothing weights or ARMA parameters.

Most desirable for the practitioner is to be shown the forecasting formula outright, and to be given an illustration (in the reference manual) of its use. For exponential smoothing, only one of the programs, *Number Cruncher*, fulfills this request. As shown in Exhibit 5, the others supply either the components -the Level, Trend, and Seasonal Index values - of the forecasting formula or less helpfully, the weights which reflect the relative emphasis

given the recent and distant past. It seems that the practitioner who wishes to demonstrate the mechanics of a smoothing model is given little assistance by the software.

The logic behind ARIMA models is difficult to describe in layman's terms, making it perhaps all the more important that the ARIMA practitioner be provided with forecasting formulas. In this context, the forecasting formula is a difference equation revealing the important dates in the past and the emphasis or weight given each of them. None of the eight programs that offer ARIMA modeling, however, displays the forecasting formula as part of the output for the final model. In several of the programs, moreover, even an expert ARIMA analyst would be hard pressed to derive the forecasting formula from the information tabulated. These software programs thus perpetuate the erroneous belief that the forecasting process underlying ARIMA is inherently incapable of description.

3.3.2. Graphing the forecasts

A forecasting program's graphing capability should enable the user to (a) append forecasts to the plot of the historical series, (b) compare forecasts from several different methods, and (c) obtain sub-series plots. So equipped, the practitioner will be able to demonstrate how the forecasts reflect and extend the patterns in the historical data. In addition, the viewer may discover that a certain specification, although it admirably fits the historical data, issues forecasts which appear entirely unreasonable.

Columns 3-5 of Exhibit 5 describe each program's amenities in the graphing of forecasts. Almost all the programs will automatically append the forecasts to a graph of the historical data. Hence a keystroke will be all that is needed to view the forecasts in historical context. By exception, Forecast Pro requires the user to designate which variables are to be plotted. This is a simple matter, but does require up to ten extra keystrokes per graph. PC/Sibyl does not itself offer a plot of the forecasts. Rather, it assumes that the practitioner will export the forecasts to a spreadsheet for further treatment, including graphing.

Multiple plots are valuable for comparing the forecasts of two or more different specifications. The feature is not available in 8 of the 13 programs, as a result of which one can graph forecasts only for the method in progress. Most of the remaining programs will permit the user to save the forecasts from any number of methods, and then to utilize a general graphing facility to plot some of these.



Two programs have noteworthy capabilities. *Autocast II* can automatically plot the forecasts from the previous as well as current smoothing specification, a feature that would be enhanced if it could encompass a third specification as well. For example, the practitioner may wish to visually compare the forecasts using a global linear trend, a local linear trend and a damped trend. *Time Machine* automatically saves the fitted/forecast values of all methods tried, permitting simultaneous graphing for up to nine sets of

forecasts plus the original series. However, a comparison of three methods is probably the visual saturation point.

Sub-series plots can be especially valuable if the historical series is lengthy or has undergone a change in trend in the recent past. In this circumstance, it can be informative to graphically compare extrapolations of global vs. local trends. Sub-series plots, moreover, act as a magnifying glass for improved outlier detection.

Four of the thirteen programs will do sub-series plots automatically; that is, the user will be allowed, within the graphics mode, to designate the portion of the data to be plotted. In six other programs the user must create a new file of forecasted values, a notable inconvenience.

3.3.3. Reversing transformations

Forecasters often wish to perform a transformation of the original time series. Logarithmic and, more generally, power transformations of the Box-Cox variety (1964) have been applied in smoothing, ARIMA, and regression to stabilize a time series whose variance has been widening with the increasing level of the series. As well, the practitioner may wish to deseasonalize a time series and find an appropriate specification for the de-seasonalized series. In these cases, the method chosen to fit the data will generate forecasts of the transformed or deseasonalized series. The practitioner of course will wish to present the forecasts in terms of the original series. So a reverse transformation or reseasonalization of the forecasts is required.

Column 6 of Exhibit 5 denotes the programs' facility for reversing transformed forecasts. Most desirable is the automatic procedure found in *Autobox Plus* and *Forecast Plus* for ARIMA, and in *Autocast II* for exponential smoothing. These programs not only show the forecasts after they have been transformed back into the original units but present the fit statistics in terms of the original data as well. This is a valuable option for deciding whether the transformation has been beneficial.

Lacking an automatic transformation reversal, the user will need to call upon a program's edit facility to undo the transformation; or worse, to create a new file. In either case, the calculated fit statistics remain in the terms of the transformed data. The danger arises that the practitioner will erroneously employ fit statistics measured in the units of the transformed data (e.g., log units) for the purpose of evaluating the most suitable transformation. Several of the programs inadvertently encourage such a practice.

An automatic procedure for reseasonalization of the forecasts of a deseasonalized series is offered in four programs: *Autocast II*, *Forecast!*, *PC/Sibyl*, and *Trendsetter Expert*. The seasonal adjustment itself is based upon the classical decomposition procedure. While certain other programs make internal calculations of seasonal index values, none enables the user to save the results for purposes of modeling the seasonally adjusted series.

3.3.4. Comparative summary tables

We well know that the best qualification of a prophet is a good memory. Forecasting programs, however, seem reticent about sharing their memory with the user. Certain programs maintain a log (or audit trail), which is a sequential record of the major screen displays. But only two programs, *Smooth* and *PC/Sibyl* offer a summary table of comparative results. *Smooth*'s presentation is excellent: a side-by-side listing, for up to seven specifications, of the parameter estimates, error measures, and forecasts. For the remaining programs, the practitioner must use an external means of recording inputs and outputs of a forecasting session.

This omission is the most surprising finding of our survey, and reminds us of Sherlock Holmes' puzzlement in the account of Silver Blaze.

*...the strange incident, Watson, of the dog barking in the night.
But Holmes, there was no dog barking in the night.
That is what was strange.*

4. Conclusions

We have reviewed the capabilities of thirteen single-equation method forecasting programs. Every program automates either the optimization of parameters, the process of method selection, or both. Our principal goal was to evaluate the ability of the forecasting practitioner to utilize such software in a manner consistent with good forecasting practice in the selection, evaluation, and defense of a forecasting method. We will now summarize our findings.

4.1. Method selection

(1) For half of these programs, the conception of automatic forecasting is limited to automatic parameter optimization in smoothing, regression or ARIMA specifications. There is nothing inherently inimical to good forecasting practice in automatic parameter optimization; indeed, APO saves the practitioner of exponential smoothing from the kind of trial and error search that can lead one to overfit the past.

(2) However, it is the practitioner's responsibility to realize that no set of parameter estimates is uniquely optimal; but rather that the parameter estimates will vary with the system for choosing starting values, with the settings for period of fit and with the choice of statistical criterion of best fit.

(3) The principal problem in this regard is that the software by and large fails to facilitate practitioner selection of fit criteria. Very few of the programs offer the opportunity to implement estimation procedures more robust than MSE minimization. Too often as well, the programs fail to permit the user to designate a period of fit or period of optimization without going through the steps to create a new data file.

(4) In those programs which offer automatic method selection, we do not doubt that the program's first choice is certainly a valuable starter method, against which the practitioner can test his own attempts at modeling. However, the software does not encourage such a procedure. The practitioner is shown a best specification or tournament winner, but is given no basis for understanding why the procedure has won or whether

further improvements should be sought. The absence of meaningful feedback may safeguard the program's secret recipes (in this way, preserving its individuality as well) but it also saps the program of its pedagogical utility: practitioners cannot emulate the AMS procedures in order to improve their modeling skills.

(5) In general, these programs do not automate such useful identification aids as (a) plots of seasonally adjusted data, (b) comparisons of various transformations for stabilizing and normalizing the data, and (c) screening for outliers and discontinuities. Moreover, the graphing/plotting options often are inadequate for the preliminary data analysis so important to model building. Without good graphics, the practitioner is entirely dependent on the program's choice of specification.

(6) To rely unquestioningly on the results of AMS would be poor practice. This is so because the range of options from which the expert chooses is limited, both in the type of method, in the variety of specifications of that method, and in the allowable fit settings. In addition, the absence of screening procedures for outliers and discontinuities can, as Tsay (1987) put it, "easily distort specification of the underlying model." The practitioner should know that automatic method selection does not gainsay preliminary data analysis and appropriate data cleansing.

4.2. Method evaluation

Careful evaluation of a method's residuals and (post-sample) forecasting errors is all the more important when the method itself has emerged from the shaded box of automatic method selection. One would hope in this case that AMS would be accompanied by automatic error evaluation. Indeed, if not directed to undertake further evaluation, the practitioner might have the false sense of security that the prescribed method has succeeded in jumping through all important hoops.

(7) Our section on method evaluation documents the extremely wide range of programmatic offerings, from the virtual absence of error analysis at one extreme to sophisticated simulation capabilities at the other. On the positive side, most all the software automatically shows the user a time plot and autocorrelation function of model residuals. Most provide a bevy of statistics measuring the precision of the historical fit. Thereafter the record is uneven.

(8) In general there is under-attention to measurement of bias in the forecasts (only seven of the thirteen programs report at least one bias measure), to comparisons of a given specification with a naive or starter method (three programs), and to identification of potential outliers (five programs). Post-sample simulation opportunities are available in only half the programs for time series methods and in only two of the ten standard regression programs.

(9) The *M*-competitions have emphasized the use of rolling simulation as a basis for evaluation of forecast errors at alternative lead times. But this feature has not been generally assimilated into forecasting software. Only one program gives the practitioner feedback on the magnitude of the forecasting errors sorted by lead time. Rolling

simulations are unavailable throughout in regression and ARIMA. The risk of poor method selection from neglect of post-sample simulation is magnified since automatic method selection tends to be based on historical rather than post-sample error measures.

(10) Regression forecasting is often a de-emphasized feature in these forecasting programs. Of the ten programs offering least-squares regression, only two materially assist the user to integrate projection of the explanatory variables with a regression forecast of the dependent variable. Only four facilitate comparison of causal and extrapolative forecasts. Only two offer post-sample (fixed) simulation, and only one provides an alternative loss function to least-squares estimation.

4.3. Presentation and defense of forecasts

To help management reduce uncertainty, wrote Adams (1986, p. 138), "a black-box forecast, no matter how accurate it is, is not sufficient. It must be possible to support the forecast persuasively to the corporate executives who will be using it as a basis for making potentially costly decisions. ..."

Our section on the presentation of forecasts examined ways by which the forecasting software could assist the practitioner to defend the forecasting method, and gain acceptance for the forecasts.

Most importantly, we considered whether the software enables the practitioner to demonstrate how the method generates forecasts, as well as whether it facilitates visual and numerical comparisons of the forecasts from different methods.

(11) Illumination of the black box is particularly important for smoothing and ARIMA, where the forecasting formula is not a transparent outcome of the estimation process. Yet guidance to the practitioner in this respect is provided in only half the packages for smoothing and none at all for ARIMA.

(12) Virtually all the programs offer graphs which automatically append a model's forecasts to the plot of the historical data. However, the multiple plotting and sub-series plotting capabilities that facilitate visual comparison of two or more methods are available in only five of the thirteen programs. These limitations are ironic in a field like forecasting, whose psychology is so strongly visual.

(13) Perhaps the most puzzling finding is the virtual absence of summary tables to compare the specifications which have been investigated, as well as the forecasts they have issued. Aggravating the problem of comparing specifications is the lack in more than half the programs of an automatic facility for reversing transformations.

4.4. The bottom line

Automatic forecasting software is providing significant and appropriate assistance to the practitioner in the selection of a specification for a time series method. Moreover, by dramatically simplifying the user's tasks at the keyboard, it has improved the

practitioner's productivity, permitting time to be spent analyzing data that might have had to be devoted to "learning the system".

Nevertheless, for the important tasks of method evaluation and forecast presentation, the practitioner is left largely to his own devices.

Perhaps our responsibility as forecasters should be, as Jenkins argued (1982, p. 16), "not only to present the forecasts to management but also the assumptions, some 'feel' for what the model is saying in common sense terms, and some appreciation of the uncertainty in the forecasts." Relative to this goal, these software programs make very limited inroads into removing the burden of pursuing good forecasting practice from the practitioner's shoulders. Technical expertise is hardly redundant, as some developers claim. "In general", said Belsley (1986, p. 45), "one must know what one is doing if one is to do it well."

Appendix A. Selection of forecast packages

Our goal in program selection was to review a representative majority of those forecasting software programs which were designed to appeal to the non-expert practitioner of basic forecasting methods. To this end we began by holding discussions with and receiving demonstrations by software developers at several annual conferences of the International Institute of Forecasters as well as the International Association of Business Forecasters. Packages whose primary orientation was felt to be (a) econometric, (b) multivariate time series, or (c) general purpose statistical were eliminated. Most vendors market more than a single program, and frequently offer several variations on the same general structure. Accordingly, we chose to include only one package per vendor.

Certain of the programs fit our sense of the automatic forecasting market but were based on "unique" methodologies, while not performing either smoothing, ARIMA, or regression. For example, we did not include Stamp because it implements only the "structural models" developed primarily by Harvey (1984).

As a result of conference contacts, ten programs were retained. Then, as a follow-up, a screening was done of the "Rycroft List". Rycroft (1989) enumerated selected attributes of 104 programs (from 65 vendors) in the categories of forecasting, econometric, and statistical software. Interestingly, three of our original ten programs were omitted from the Rycroft List: *Forman*, *Time Machine*, and *Smooth*. However, the list led us to 20 additional vendors of potential interest. We wrote to these vendors, showed them the introduction to our paper, and requested examination of one (the most appropriate in their judgment) of their programs. Twelve vendors responded – some to take themselves out of consideration – and eight additional programs were received and analyzed. Of these, three (*PC/Sibyl*, *Forecast!*, and *Number Cruncher*) were considered appropriate for this review, netting the thirteen packages listed in Exhibit 1.

The authors feel that the set of thirteen programs examined reveals a reasonably comprehensive portrait of the faces of automatic forecasting as the decade of the 1990s begins.

References

- Adams, F.G., 1986, *The Business Forecasting Revolution* (Oxford University Press, New York).
- Armstrong, J.S., 1985, *Long-Range Forecasting* (Wiley-Interscience, New York).
- Armstrong, J.S., 1987, "The forecasting audit", in: S. Makridakis and S. Wheelwright, eds., *The Handbook of Forecasting; A Manager's Guide*, 2nd ed. (Wiley-Interscience, New York) ~h. 32.
- Armstrong, J.S. and E.J. Lusk, 1983, "Commentary on the Makridakis time series competition", *Journal of Forecasting*, 2, 259-311.
- Ashley, R., "On the usefulness of macroeconomic forecasts as inputs to forecasting models", *Journal of Forecasting*, 2, 211-223.
- Beaumont, C., 1987, "Autobj" (Software Review), *Journal of Forecasting*, 6, 71- 74.
- Bell, M.Z., 1985, "Why expert systems fail", *Journal of the Operational Research Society*, 36, 613-619.
- Box, G.E.D. and D.R. Cox, 1964, "An analysis of transformations", *Journal of the Royal Statistical Society Series B*, 26, 211-252.
- Chatfield, G., 1978, "The Holt-Winters forecasting procedure", *Applied Statistics*, 27, 264-269.
- Chatfield, C. and M. Yar, 1988, "Autocast" (Software Review), *International Journal of Forecasting*, 4, 503-508.
- Clemen, R.T., 1989, "Combined forecasts: A review and annotated bibliography", *International Journal of Forecasting*, 5, 559-583.
- Cochrane, D. and G. Orcutt, 1949, "Application of least squares regression to relationships containing autocorrelated error terms", *Journal of the American Statistical Association*, 44, 32-61.
- Courcelle, R.J. and L.J. Tashman, 1989, "Box plots: Another graphical aid in forecasting", *Journal of Business Forecasting*, 7, 12-17.
- Diebold, F.X., 1989, "Forecast combination and encompassing: Reconciling two divergent literatures", *International Journal of Forecasting*, 5, 589-592.
- Dielman, T.E., 1986, "A comparison of forecasts from least absolute value and least squares regression", *Journal of Forecasting*, 5, 189-195.
- Fildes, R., 1985, "Quantitative forecasting -the state of the art: Econometric models", *Journal of the Operational Research Society*, 36, 549-580.
- Gardner, Jr., E.S., 1985, "Exponential smoothing: The state of the art", *Journal of Forecasting*, 4, 1-28.
- Harvey, A.C., 1984, "A unified view of statistical forecasting procedures", *Journal of Forecasting*, 3, 245-275.
- Jenkins, G.M., 1982, "Some practical aspects of forecasting in organizations", *Journal of Forecasting*, 1, 3-21.
- Lopes, L.L., 1983, "Pattern, pattern -Who's got the pattern", *Journal of Forecasting*, 2, 269-272.
- Makridakis, S. et al., 1982, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition", *Journal of Forecasting*, 1, 111-153.
- Makridakis, S. and M. Hibon, 1979, "Accuracy of forecasting: An empirical investigation", *Journal of the Royal Statistical Society Series A*, 142, 97-145.

- Makridakis, S. and M. Hibon, 1990, "Exponential smoothing: The effect of initial values and loss functions on post-sample forecast accuracy", INSEAD Working Paper.
- Newbold, P. and T. Bos, 1989, "On exponential smoothing and the assumption of deterministic trend plus white noise data-generating models", *International Journal of Forecasting*, 5, 523-527.
- Rycroft, R.S., 1989, "Microcomputer software of interest to forecasters in comparative review", *International Journal of Forecasting*, 5, 437-462.
- Schwarz, G., 1978, "Estimating the dimension of a model", *Annals of Statistics*, 6, 461-464.
- Sharda, R. and J.F. Rock, 1986, "Forecasting software for microcomputers", *Computers and Operations Research*, 13, 197-209.
- Tsay, R.S., 1988, "Outliers, level shifts, and variance changes in time series", *Journal of Forecasting*, 7, 1-20.
- Van Ness, F. and A. ten Cate, 1988, "Software for econometric research with a personal computer", *International Journal of Forecasting*, 5, 263-V8.
- Weiss, A.A. and A.P. Anderson, 1984, "Estimating time series models using relevant forecast evaluation criteria", *Journal of the Royal Statistical Society Series A*, 147, 484-487.
- Zellner, A., 1988, Keynote address before the 1988 International Symposium on Forecasting, Amsterdam.

Biographies: Leonard J. TASHMAN is an Associate Professor of Business Administration at the University of Vermont. He holds a Ph.D. in Economics from Brown University. He is co-author of *The Ways and Means of Statistics* (HBJ), and has published articles in the *National Tax Journal*, *Journal of Education Finance*, *Southern Economic Journal*, and *Journal of Business Forecasting*.

Michael L. LEACH is Load Research Analyst at the Burlington Electric Department (Vermont). He received an M.S. in Statistics from the University of Vermont (1988), where he engaged in research in sequential analysis.