Penn Libraries
UNIVERSITY of PENNSYLVANIA

University of Pennsylvania
ScholarlyCommons

Marketing Papers

Wharton School

October 2007

Rule-based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

Fred Collopy
Case Western Reserve University

Follow this and additional works at: http://repository.upenn.edu/marketing_papers

Rule-based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations

Abstract

This paper examines the feasibility of rule-based forecasting, a procedure that applies forecasting expertise and domain knowledge to produce forecasts according to features of the data. We developed a rule base to make annual extrapolation forecasts for economic and demographic time series. The development of the rule base drew upon protocol analyses of five experts on forecasting methods. This rule base, consisting of 99 rules, combined forecasts from four extrapolation methods (the random walk, regression, Brown's linear exponential smoothing, and Holt's exponential smoothing) according to rules using 18 features of time series. For one-year ahead *ex ante* forecasts of 90 annual series, the median absolute percentage error (MdAPE) for rule-based forecasting was 13% less than that from equally-weighted combined forecasts. For six-year ahead *ex ante* forecasts, rule-based forecasting had a MdAPE that was 42% less. The improvement in accuracy of the rule-based forecasts over equally-weighted combined forecasts was statistically significant. Rule-based forecasting was more accurate than equal-weights combining in situations involving significant trends, low uncertainty, stability, and good domain expertise.

Rule-Based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations

Fred Collopy
Case Western Reserve University

J. Scott Armstrong
The Wharton School, University of Pennsylvania

This paper examines the feasibility of rule-based forecasting, a procedure that applies forecasting expertise and domain knowledge to produce forecasts according to features of the data. We developed a rule base to make annual extrapolation forecasts for economic and demographic time series. The development of the rule base drew upon protocol analyses of five experts on forecasting methods. This rule base, consisting of 99 rules, combined forecasts from four extrapolation methods (the random walk, regression, Brown's linear exponential smoothing, and Holt's exponential smoothing) according to rules using 18 features of time series. For one-year ahead ex *ante* forecasts of 90 annual series, the median absolute percentage error (MdAPE) for rule-based forecasting was 13% less than that from equally-weighted combined forecasts. For six-year ahead *ex ante* forecasts, rule-based forecasting had a MdAPE that was 42% less. The improvement in accuracy of the rule-based forecasts over equally-weighted combined forecasts was statistically significant. Rule-based forecasting was more accurate than equal-weights combining in situations involving significant trends, low uncertainty, stability, and good domain expertise.

Introduction

In empirical studies of forecasting accuracy, no single extrapolation method has performed well across all types of data and all forecast horizons. In this paper, we describe an approach where the use of extrapolation methods depends upon the features of a time series. This approach incorporates judgment into the extrapolation process. The combination of judgment and quantitative methods was called for in commentary to Makridakis and Hibon (1979) and to the M-competition (see commentary in Armstrong and Lusk 1983). A survey of experts (Armstrong 1988) identified a need to integrate judgment and extrapolation. In reviewing the literature, Bunn and Wright (1991) concluded that judgment and statistical forecasting methods should be integrated. Furthermore, Lawrence, Edmundson, and O'Connor (1986) and Sanders and Ritzman (1990b) have provided empirical support.

We developed and applied a rule base for extrapolations of annual economic and demographic time series. We compared the accuracy of rule-based forecasting with other procedures for making extrapolations, in particular with equally-weighted combined forecasts.

The outline of the paper is as follows. First, we discuss the process that we used for developing a rule base. Then we describe the rule base that we tested. Next, we propose hypotheses about the conditions of time series that favor rule-based forecasting. We then explain the validation procedure. Finally, we present results of the validation tests.

Developing a Rule Base

Our intention for rule-based forecasting was to apply a *validated, fully disclosed,* and *understandable* set of conditional actions to make forecasts. Although rule development begins by specifying rules that seem reasonable to experts, the goal is to go beyond such specifications and to validate rules by prior research and empirical tests. By

fully disclosed, we mean that the forecaster (or decision maker) has access to all of the rules. Understandable is judged from the perspective of forecasters and decision makers. We have developed a rule base that helps to realize our intention.

Our rule base integrates several strategies for extrapolation. Among these are (1) using features of the series to establish weights for combining forecasts, (2) using heuristics to establish parameters for an exponential smoothing model, (3) using separate models for long-range and short-range forecasts, (4) damping the trend under certain conditions, and (5) incorporating domain knowledge in extrapolation. Most of these ideas have been discussed in the forecasting literature and some have been supported by empirical research.

In this study, we focused primarily upon combining. Of the many approaches proposed for combining forecasts, equal weights has been shown to be particularly robust. However, our rule base uses features of the series to produce differential weights.

Obtaining Information for Developing Rules

The forecasting literature provided some research streams that influenced our development of rules. Among these research streams were findings that favor decomposing time series, placing more weight on recent data, using simple methods, and making conservative trend estimates when uncertainty is high. Guidelines from the literature, however, typically lacked a precise statement of the conditions under which particular actions were likely to be useful. Also, such guidelines rarely specified precise actions needed to obtain forecasts.

We conducted a survey to obtain experts' opinions about what factors affect the accuracy of extrapolations (Collopy and Armstrong 1992). We obtained the opinions of 49 practitioners and academics engaged in forecasting and in research on forecasting. These experts had few strong opinions about what factors relate to accuracy. For example, 44% had no opinion with respect to the statement, "Simpler models will more accurately forecast series with a high level of randomness than more statistically sophisticated models." The respondents who did have an opinion often differed from one another. For the same statement, 32% of the experts agreed while 24% of them disagreed. As another example, the respondents were undecided about whether combining forecasts was more appropriate for short- or long-range forecasts. Nevertheless, the survey provided useful information, in particular about what features should be considered in selecting an extrapolation method. For example, the experts thought that the presence of an abrupt change was as important as the presence of trends. They also judged both the short-term and long-term trends to be important in the selection of an extrapolation method.

In retrospective assessments, the first author asked five forecasting experts to describe their rules for selecting and weighting extrapolation methods for forecasting. The experts were J. Scott Armstrong, Christopher Chatfield, Robert Fildes, Spyros Makridakis, and Steven P. Schnaars. During the tape-recorded sessions, the experts described numerous guidelines that were generally consistent with the published literature. For example, the experts cautioned against the use of complex methods and they suggested combining forecasts for many situations. But the experts also offered insights about such procedures as (1) adjusting the data for unusual events, (2) identifying series that are likely to regress to the series mean in the future, (3) using the random walk for unstable series, and (4) relying principally on exponential smoothing for stable series. Collopy and Armstrong (1989) provide details on the procedure used in these direct assessments.

Grabowski (1988) concluded that operational rules were most likely to be discovered when experts performed familiar, realistic tasks. Consistent with this, we gave realistic tasks to the same five experts. We used a concurrent protocol method (Ericsson and Simon 1984) with each expert working individually. They received plots of six time series along with their titles and dates. The experts could view a variety of transformations, statistics, and forecasts. In tape-recorded sessions, they described how they would forecast each series.

We developed most rules through our analysis of protocols. For example, the experts suggested rules for irregularities and changes in patterns. One rule was "If the series has a recent pattern that looks suspicious, then increase the weight given to the random walk forecast. That is, rely more on an extrapolation of the last observation with no trend." (The operational form of this is rule 46 in Appendix B.) We found protocols especially useful in linking actions with conditions. On the other hand, the experts overlooked several conditions that were nut represented in the test series. Collopy and Armstrong (1989) provide further details about some of these protocols.

Specifying the Rules

Our rules combine forecasts using weights that vary according to the features of the series. Given the desire for an understandable system, we chose to combine four simple, widely understood extrapolation methods:

(1)   The random walk emphasizes the short-range perspective; it sets the level to the last observation and is based on the assumption that there is no trend.

(2)   The linear regression, which fits a least squares line to the historical data (or transformed historical data), represents the long range; we refer to its trend estimate as the basic trend.

(3)   Holt's linear exponential smoothing (Holt et al. 1960) captures information about short-range trends, and we call this the recent trend.

(4)   Brown's linear exponential smoothing with trend (Brown 1959) also measures the short-range trend.

Makridakis et al. (1982) describes these methods.

We introduced rules only if they made sense to us. An initial set of 36 time series was used to calibrate the rules. First, we made judgmental (eyeball) extrapolations. We then compared these with the rule-based forecasts to see if the latter looked reasonable to us. When the judgmental forecasts and the rule-based forecasts differed substantially, we searched for explanations. We also examined changes in the forecasts that resulted from adding a rule. These analyses led to many refinements in the rules.

Because rules interact with one another, we examined the effects of removing a rule (or group of rules) from the rule base. We then put it back before analyzing other rules. We also examined different formulations for important rules. For example, we varied damping factors to determine the effects on accuracy. Experimentation was constrained because the number of series affected by some rules was small. These searches involved hundreds of runs and we recorded the results in a log.

Description of a Rule Base

One objective of the rule base is to provide more accurate forecasts. A second objective is to provide a systematic summary of knowledge. By expressing the knowledge explicitly and uniformly, we gain benefits generally associated with rule bases. These benefits include automating some tasks associated with maintaining a complex body of knowledge and providing knowledge in an accessible and modifiable form. Besides its usefulness in forecasting, knowledge in this form aids reasoning about forecasting; that is, the knowledge is useful both procedurally and declaratively. Such explicit representation also can help in developing and testing theories (Buchanan and Shortliffe 1985, p. 52).

Our goal in this study was to test the feasibility of rule-based forecasting as a procedure. To do this, we had to establish a set of reasonable rules. We do not presume that they are the best possible set of rules.

Figure 1 shows the elements of the rule base. First, features of the series are identified. Rules are then applied to produce short- and long-range forecasting models. To formulate these models we had to select smoothing factors for the exponential smoothing methods and make estimates of levels and trends for each model. For the long-range model, we formulated rules to damp the trend over the forecast horizon. Finally, there are rules for blending the forecasts - from the short- and long-range models. The rule base in Appendix B follows the structure of Figure 1. Appendix B presents the 99 rules along with brief explanations.

Figure 1
Structure of the Rule Base
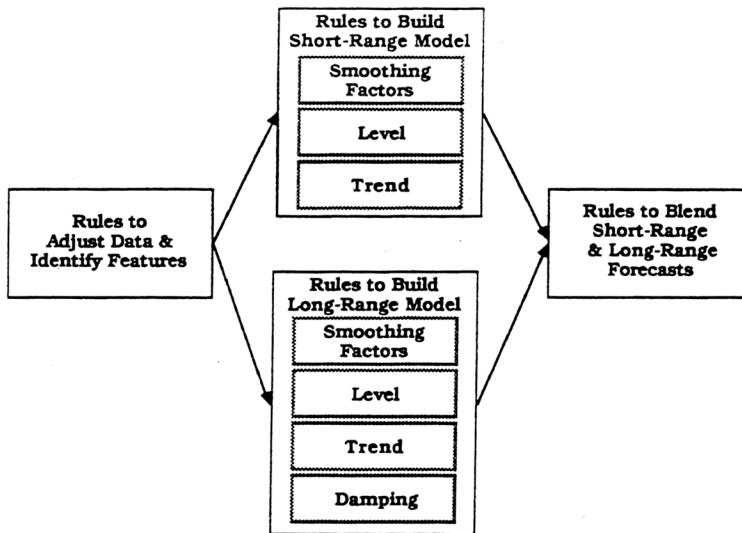
Identifying Series Features

Forecasting actions are conditioned by features of time series. We used 18 features to describe time series. All of the features are binary except the coefficient of variation (which is continuous) and causal forces (described by six categories). Appendix A defines the features.

The analyst identifies certain features (e.g., .irrelevant early data, unstable recent trend) by examining plots of the series. The identification of 'last observation unusual' is of particular concern given that the last observation typically plays an important part in the estimation of both level and trend. We expect that eventually the identification of such features will be performed by rules or analytical procedures.

Domain knowledge (knowledge about the underlying process) is used to determine some features. Based upon knowledge of the underlying process, the analyst specifies the functional form as either additive or multiplicative. For example, many sales series follow a multiplicative process; that is, the trend can be thought of in percentage terms. Domain, knowledge also aids in identifying economic cycles, unusual events, and causal forces. Causal forces represent the net directional effect of the factors that influence a specific series at a given point in time. Causal forces play a role in many of the rules here. For example, sales of a product are often affected by forces that promote growth, such as an increasing number of customers, their greater wealth, their growing awareness of the product, and improvements in product quality. If a sales trend is downward, a firm might try, to reverse it by reducing prices or increasing advertising. Consequently, the rules would extrapolate the estimated upward trends for this series, but they would damp downward trends. (For a discussion and evaluation of causal forces, see Armstrong and Collopy 1992b.)

Rules and analytic procedures determine some features. Currently, these include features such as the direction. of the basic trend, the statistical significance of the basic trend, and the amount of variation about the trend.

Estimating the Short-Range and Long -Range Models

The factors affecting short-range forecasts for a time series often differ from those affecting long-range forecasts. For example, in the short term, sales of automobiles could be affected by price reductions or by a new advertising campaign. Long-range forecasts are influenced by such things as design, service, restrictions on foreign

trade, and product reliability. Short- and long-range models reflect different underlying processes. Separate models for short- and long-range extrapolations improved accuracy in a study by Carbone and Makridakis (1986).

Smoothing Factors. A grid-search procedure provides estimates for smoothing factors that produce the best fit to the historical data for Holt's exponential smoothing. We used the grid search described in Makridakis et al. (1982). Smoothing factors for Brown's linear exponential smoothing were based largely upon rules from Armstrong (1985, pp. 165-168) and Gardner (1985). For example, alpha (for levels) is decreased when the amount of historical variation is large. Beta (for trends) for the short-range model is increased if the recent trend direction is supported by the causal forces.

Level. The rules apply weights to the estimated levels from the four methods-random walk, linear regression, Holt's, and Brown's. Short- and long-range models use different weights. For example, the short-range model puts more weight on the random walk. Weights for levels also vary depending upon the presence of discontinuities in the historical data, the proximity of the last value to a previous extreme in a cyclical series, the occurrence of an unstable recent trend or a changing basic trend, or the presence of a suspicious pattern. The assumption here is that features found in the historical data may occur again in the future. For example, when discontinuities are observed, rules put more weight on the latest observation (the random walk)
.

The rules adjust the level based upon how well the rule base was able to forecast the latest observation (t_0) given data through the preceding period (t_{-1}). Rules adjust the estimate of the Level at t_0 in the direction. implied by the causal forces.

Trend. When the directions of the recent trend (estimated by Holt's exponential smoothing) and the basic trend (estimated by linear regression) different rules make a more conservative trend extrapolation; they do this by increasing the weight on the random walk. Similarly, when the causal forces are not in the same direction as the estimated trend, rules make a more conservative trend estimate.

The short- and long-range models employ different trend weights. The recent trend is more heavily weighted in the short-range model. The basic trend gets more weight in the long-range model than in the short-range model.

Other features also affect the weighting of extrapolation methods. For example, if there has been a long recent run (movements in the same direction for at least six consecutive years), a rule puts more weight on the recent trend and less on the basic trend (see rule 44 in Appendix B for the operational specification).

For series with causal forces that are regressing (that is, tending to return to a mean), a trend line is extended from the latest observation to the overall historical mean of the series. Rules then blend the previously estimated long-range model trend with this re gressing trend.

Damping the Long-Range Model Trend. The damping factor represents the percentage of the long-range model trend that is ignored in the forecast. Rules establish a damping factor based upon the consistency of the recent and basic trends, and on the relationship of the direction of these trends to the causal forces (rules 89 to 91 in Appendix B). Rules then modify the factor by the amount of variation in the series (rule 92) and for suspicious patterns (rule 93) and unstable recent trends (rule 94). The damping factor increases (i.e., the trend is further discounted) as the forecast horizon is extended beyond the first year. For example; if the rules establish a 5% damping factor, the long-range model trend for the two-year horizon is multiplied by 0.95. For the 3-year-ahead forecast, the procedure reduces the trend by $(0.95) \times (0.95) = 0.9025$, and so forth.

Blending the Short-Range and Long-Range Forecasts

Rules blend the forecasts from the short- and long-range models. They place progressively more weight on the long-range model as the forecast horizon increases. Rules also vary the weights according to the direction of the trends in the long-range and short-range models; and according to their relationships with the direction of causal forces. For example, when the directions of the short-range and long-range model trends *differ,* the rules put more weight on the model that is consistent with the direction implied by the causal forces.

An Example

As an illustration of the impact of the rules, Figure 2 presents six-year-ahead forecasts for General Motors. Figure 2a shows the forecast using an additive model; note the drop in the level. We specified a multiplicative model based on domain knowledge; this corresponds more closely with the historical trend (Figure 2b). Upon examining this series but before looking at the holdout sample, we judged that an unusual event must have occurred in the last year, 1970. (It did. We learned that a strike in 1970 cost General Motors 20% of its annual sales.) Therefore, we adjusted the last period, identified the last point as unusual, and applied the rules to produce a forecast. The rule for handling unusual last observations (rule 4) had a substantial impact. Of lesser importance was our adjustment of the last observation based on domain knowledge about the strike (rule 3). Figure 2c presents the forecast along with the actual sales.
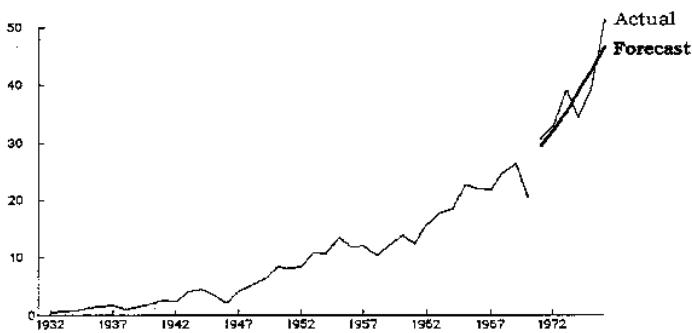
Figure 2
Rule-based Forecasts and Actuals for Net Sales of General Motors in Billions of Dollars
(Forecasts are shown using bold lines.)
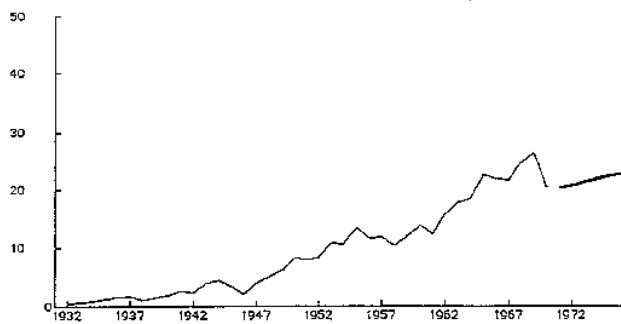
A — Additive Model
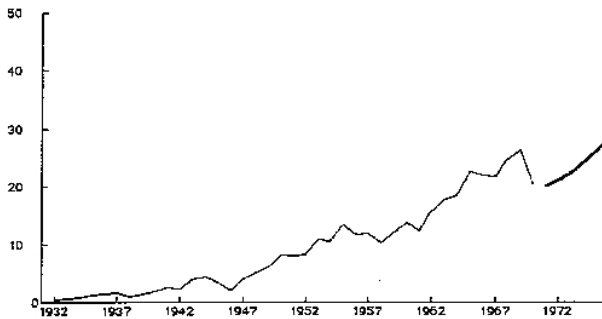
B — Multiplicative Model

C — Multiplicative with Last Observation Identified as Unusual

Hypotheses on Conditions Favoring Rule-Based Forecasting

The performance of rule-based forecasting depends not only on the rule base, but also on the conditions of the series. By conditions, we mean the set of features that describes a series. We expected both equal-weights combining and rule-based forecasting to be more accurate than the random walk when

> (1) historical data. have significant trends, and
> (2) uncertainty is low.

With respect to equal-weights combining, prior research provides much support for the first hypothesis and fair support for the second. We also hypothesized that rule-based forecasting would provide more accurate forecasts than the random walk or equal-weights combining when

> (3) historical data show stable patterns, and
> (4) good domain knowledge is available.

Hypotheses 3 and 4 represent logical extensions of prior research. We also expected rule-based forecasting to be more accurate for longer horizons.

Extrapolating trends is risky because if the trend forecast is in the wrong direction, the resulting forecast will be less accurate than the random walk. Also, if the trend changes substantially, it can produce errors larger than those from the random walk. We expected a trended forecast to be most useful when the basic trend in the historical data is statistically significant. When it is not, the random walk will tend to perform relatively well and the rule-based forecast will tend to move toward the random walk forecast.

The rule base relies upon the patterns in the data to decide which extrapolation methods to use and how to weight them. When uncertainty is low, pasterns are easier to detect, so rule-based forecasting is expected to be more accurate. Conversely, high uncertainty makes it difficult for rule-based forecasting to select the proper method. Here, equal weights combining should perform fairly well according to findings by Sanders and Ritzman (1990a) and Lawrence et al. (1986). When uncertainty is high, rule-based forecasting tends to approximate an equally-weighted combined forecast. It also uses more damping, leading it to more closely approximate a random walk. Therefore, we expected rule-based forecasting to improve accuracy for series with low uncertainty, and to pose little risk for series with high uncertainty.

When series are unstable (with discontinuities, changing trends, etc.), the rule base moves toward a random walk forecast. Hence, rule-based forecasting can help in stable situations, and it is unlikely to harm the forecasts when the historical series show unstable patterns. Some techniques, such as Holt's or Brown's, might encounter difficulties when historical series show unstable patterns.

Knowledge about the subject area that is being forecast should be useful. Domain knowledge has improved accuracy in previous extrapolation studies (e.g., Mathews and Diamantopoulos 1990; Sanders and Ritzman 1990b). Given its ability to incorporate domain knowledge, rule-based forecasting should be more accurate than the other extrapolation procedures when the forecaster has domain knowledge.

Because rule-based forecasting can incorporate causal information, we expected that it would be more useful for long-range than for short-range forecasts, because causal factors have stronger effects in the long term. We also expected rule-based forecasting to be more appropriate for long-range forecasts because it uses information about the patterns in the data.

Validation Procedure

This section describes the time series data used to calibrate and validate the rule base, features of these series, and reliability of the judgmental feature identification process. Next, it proposes criteria for assessing accuracy, followed by the *ex ante* testing procedure. Finally, it describes alternative extrapolation procedures used as a basis of comparison to assess the performance of rule-based forecasting.

Data

       We analyzed 126 of the annual time series from the M-competition data (Makridakis et al. 1982). This 70% sample of the annual series was stratified by demographic (comprising 17% of the series), macroeconomic (32%), organization (32%), and industry (19%) categories. These data were from several countries and covered different time spans. The number of observations varied from 15 to 58, with a mean of 24 years, a median of 21, and a mode of 19 years. About a third of the series were of modal length.

       Some features were common across most of these series. For example, 87% had upward basic trends, 84% had upward recent trends, and 83% had multiplicative functional forms. The most common causal force was growth (68%). However, the series did vary with respect to many features, in particular instability. Table 1 summarizes the features for these series.[1]

Table 1
Presence of Features and Coding Reliability
(Entries are percentages; R means identified by rules)

Feature	Feature Present*	Coder Agreement**	Feature	Feature Present*	Coder Agreement**
Trend			Domain Knowledge		
Direction of Basic Trend Up	87	R	Causal Forces		
Direction of Recent Trend Up	84	R	Growth	68 ⎫	
Significant Basic Trend ($t > 2$)	83	R	Decay	2 ⎪	
Uncertainty			Supporting	2 ⎪	81
Coeff of Variation > 0.2	21	R	Opposing	7 ⎬	
Basic & Recent Trends Differ	12	R	Regressing	3 ⎪	
Instability			Unknown	18 ⎭	
Recent Run Not Long	60	R	Functional Form		
Near a Previous Extreme	49	R	Multiplicative	83 ⎫	75
Irrelevant Early Data	40	94	Additive	17 ⎭	
Changing Basic Trend	32	81	Cycles Present	26	84
Suspicious Pattern	21	94	Adjusted Observations	2	100
Outliers Present	13	R			
Unstable Recent Trend	12	91			
Level Discontinuities	6	91			
Last Observation Unusual	2	100			

* 126 series.
** 36 series.

       Rule-based forecasting relies on the assumption that the features of each series can be reliably identified. Unfortunately, feature identification for these series was done under adverse conditions. First, the titles for the series often were not highly informative (e.g., "Company Data") and the descriptive information was also limited. Further, though we had little domain knowledge, we made the judgmental identifications of features. To assess reliability, we compared features identified independently by the two of us for the last of the validation tests used in our study. This consisted of 36 series. Agreement ranged from 75% to 100%, and averaged 89%. However, better reliability would be desirable for such critical features as causal forces (with 81 % agreement), functional form (75%), and changing basic trend (81%). Table 1 summarizes the reliability for each feature in the column labeled "Coder Agreement."

       Should rule-based forecasting perform well on these data? The presence of trends is favorable; 83% had statistically significant basic trends. Uncertainty did not pose a serious problem; variation was high in only 21 % of

---

[1] A coding of the features for each series can be obtained from the authors.

the series. Domain knowledge, while modest, seemed valuable; for example, we specified causal forces for all but 18% of the series. On the negative side, considerable instability existed; for example, about 32% of the series had changing basic trends, and only 15% of the series were free of any instability features. Overall, despite some instability in the series and limited domain knowledge, we expected rule -based forecasting to provide more accurate forecasts for these data than could be obtained from the random walk or equal-weights combining.

Criteria

Our primary criterion for the comparison of extrapolation procedures was accuracy. Academicians and practitioners surveyed by Carbone and Armstrong (1982) also identified accuracy as their primary criterion, as did practitioners in a survey by Mentzer and Cox (1984).

To assess accuracy, we used the median absolute percentage error (MdAPE) for comparisons across series because it is easy to interpret, not dependent on scale, unaffected by outliers, reliable, and valid. We also calculated the Mean Absolute Percentage Error (MAPE), the Median Relative Absolute Error (MdRAE), and the Geometric Mean of the Relative Absolute Error (GMRAE). (The RAE is similar to Theil's U2 as it controls for scale and for the amount of change over the forecast horizon.. It is calculated by dividing the absolute forecast error from the proposed model by the corresponding error from the random walk. To summarize across a set of time series, we used the geometric mean of the RAE.)[2] Armstrong and Collopy (1992a) analyzed these error measures and summarized the advantages and disadvantages of each.

The holdout sample and *ex ante* validation procedure were the same as those used in the M -competition to provide a common benchmark for comparis on. Six years were held out for *ex ante* forecast validation for each series, providing a single *h*-year-ahead forecast for each series for *h* = 1 to 6. A single starting year was used for each series, and the starting years differed among series.

We selected stratified random samples for the calibration and validation subsamples. The tests used three validation subsamples, labeled as $V1$, $V2$, and $V3$.[3] Our rationale for selecting several validation subsamples was to evaluate the rule -based forecasting process, instead of focusing on a particular set of rules.

We calibrated the rule base on 36 annual series and produced ex *ante* forecasts for the 18 series in validation set $V1$. We updated the rule base by recalibrating it with the original 36 series plus the 18 from $Y1$. This set. of rules was used to produce *ex ante* forecasts for the 36 series of $V2$. Finally, we included the 36 series from $V2$ and recalibrated the rules again to produce *ex ante* forecasts for the 36 series of $V3$. In other words, the calibration of the rules yielded three models that differed somewhat with respect to the number of rules and to details of their formulation.

Alternative Extrapolation Procedure*s*

We used equal-weights combining as the primary basis for comparison because empirical research showed that this strategy is robust. Equal weights typically provides extrapolations that are at least as accurate as other strategies across a variety of conditions. In addition, we compared the accuracy of rule -based forecasts against three other procedures: random walk, typical method-nine, and typical method-five. We obtained forecasts for these procedures from tapes used in the M -competition study. All four procedures are described here.

Random Walk. The random walk has been a basis for comparison in some prior studies. Sometimes it has been a strong competitor. For example, in Schnaars' (1984) study of annual unit sales forecasts, the random walk

---

[2] We trimmed the RAE at 0.01 on the low side and at 10 on the high side. Trimming occurred for about 2% of the error measures for the one- and six-year-ahead forecasts, and for about 0.3% of the cumulative errors.
[3] We selected the series randomly from ten stratified samples. The series ending in the numbers 3 and 6 were used for calibration. The first validation used series ending in 5, the second validation used those ending in 2 and 7, and the third used those ending in 4 and 8.

was as accurate as other methods, one of which was equal-weights combining of forecasts from seven extrapolation methods.

Typical Method-Nine. Typical method-nine assumed that the forecaster chose with equal likelihood from among nine quantitative methods employed for annual data in the M-competition: moving averages, single exponential smoothing, adaptive response rate exponential smoothing, Brown's quadratic exponential smoothing, linear trend analysis (regression), automatic AEP filtering, Bayesian forecasting, Holt's exponential smoothing, and Brown's linear exponential smoothing. Makridakis, et al. (1982) provide descriptions of these methods. Each was proposed for these data by experts in forecasting. A survey by Fildes and Lusk (1984)found that forecasting experts disagreed about which of these methods to use for such series. So we calculated errors for each of the nine methods across the series, and then selected the method that had the median error. Consequently, this approach estimates the level of accuracy one would achieve given the a priori knowledge that existed at the time the M-competition study was conducted.

Typical Method-Five. Typical method-five is like typical method-nine except that it included only the five smoothing methods from the M-competition's combining strategies (Makridakis et al. 1982): single exponential smoothing, adaptive response rate exponential smoothing, automatic AEP filtering, Holt's exponential smoothing, and Brown's linear exponential smoothing. We used typical method-five to determine' whether the typical method estimate is sensitive to the selection of the methods. More importantly, however, it tests the value of equal-weights combining for these series.

Equal-Weights Combining. Combining has performed well in published empirical comparisons (Clemen 1989). Equal-weights combining has generally been more accurate than the average accuracy of the individual forecasts used in the combination. Sometimes it has been more accurate than the best of the individual forecasts. Equal-weights combining is recommended when one does not know what is the best method, a situation that seems common. We use the results from Makridakis et al. (1982) for Combining A, which was an equally-weighted combination of forecasts from the methods used in typical method-five.[4] We sometimes refer to this method as 'equal-weights.'

Results

Table 2 summarizes the median absolute percentage errors for the three ex ante forecast validation tests. For these data, the "typical method" procedures reflect the benefits of quantitative methods developed since the 1950s. The accuracy of the two typical method approaches was comparable. Each produced more accurate forecasts than did the random walk. On average, the typical method-five error was 22% less than that for the random walk for the one-year-ahead forecasts and 25% less for six-year-ahead forecasts.

---

[4] For details see Makridakis et al. (1982). In that study equal-weights combining was called Combining A. This is described as being composed of six methods. In addition to the five methods included in typical method-five, Combining A included Holt-Winters'. Because Holt-Winters' is equivalent to Holt's for annual data, we excluded it from the typical methods. An alternative procedure, designed to estimate the best fit, was called Combining B. That procedure was not as accurate as Combining A in the M-competition study. We examined Combining B for the series used in our study and found it to be less accurate than equal-weights.

Table 2
Ex-Ante Forecast Errors for Extrapolation Procedures (Median Absolute Percentage Errors = MdAPEs)

Extrapolation Procedure	One-Year-Ahead Forecasts				Six-Year-Ahead Forecasts			
	$V1$ (18)	$V2$ (36)	$V3$ (36)	Weighted Average	$V1$ (18)	$V2$ (36)	$V3$ (36)	Weighted Average
Random Walk	6.4	5.7	5.6	5.8	30.1	24.7	25.2	26.0
Typical Method—nine	5.5	4.3	4.9	4.8	23.3	18.0	18.0	19.0
Typical Method—five	5.0	3.9	4.8	4.5	23.9	18.1	18.8	19.5
Equal-Weights	2.8	3.1	4.3	3.5	22.8	21.9	18.4	20.7
Rule-Based Forecasting	2.5	3.1	3.2	3.0	13.0	9.1	14.2	11.9

Note. The number of series is indicated in parentheses.

Equal-weights combining yielded substantial gains over typical method-five for one year ahead forecasts, as its MdAPE was 22% lower. On six-year-ahead forecasts, it did not improve on the performance of typical method-five; on the contrary, its errors were larger. This finding suggests that equal-weights combining is more appropriate for short-range forecasts, perhaps because short-range forecast errors from the individual methods are more likely to be compensating.

As we hypothesized, rule-based forecasting produced more accurate forecasts than did equal-weights. Its errors were lower in each of the three validations for both one-year ahead and six-year-ahead forecasts, except for one tie. For one-year-ahead forecasts, rule-based forecast errors were 13% lower on average across the three validations than those from equal-weights. For the six-year-ahead forecasts the rule-based forecast errors were 42% less. For $V3$, the absolute percentage errors from the rule-based forecasts were lower than those from equal-weights for 61 % of the one-year-ahead forecasts; the same result was obtained for the six-year-ahead and cumulative-horizon forecasts.

We tested the superiority of the rule-based forecasts over those from equal-weights using the Wilcoxon matched-pairs signed-ranks test for the Absolute Percentage Error in each series. We calculated significance tests separately for each set of validation data ($V1$; $V2$, and $V3$) for one-year, six-year, and cumulative horizons. Each of the nine tests was in the predicted direction, and five were significant at the $p < 0.05$ level. Following Rosenthal's (1978) procedure for adding Z's, we found that rule-based forecasting was significantly more accurate than equal-weights for one-year-ahead ($p = 0.009$), six-year ahead ($p = 0.001$), and cumulative horizon ($p = 0.006$) forecasts. [5]

We examined other measures of accuracy for Validation 3. These results supported the above findings. Table 3 shows errors for the one-year-ahead and six-year-ahead forecasts and for the cumulative horizon for the primary selection procedures. The cumulative error measure includes more information (forecast errors for horizons 2, 3, 4, and 5), and it is less susceptible to outliers (Armstrong and Collopy 1992a). Rule-based forecasting was more accurate than equal-weights and than the random walk over all horizons for all criteria except the MAPE, where it was less accurate than equal-weights. Its relative superiority was greater for the longer horizon and for the more reliable measures (relative errors and medians). We also performed the same analysis by applying the same rule base (Appendix B) to all 126 annual series; the rule-based forecasts were more accurate on all criteria for all horizons.

---

[5] This procedure required the calculation of Z scores for the three validation tests, adding them, and dividing by the square root of the number of experiments (three) to get an overall level of statistical significance.

Table 3
Alternate Measures of Ex Ante Forecast Errors for Validation 3
(For 1-year-ahead, 6-year-ahead, and cumulative of 1- to 6-year-ahead)

Extrapolation Procedure	Median Relative Absolute Error (MdRAE)			Geometric Mean Relative Absolute Error (GMRAE)			Median Absolute Percentage Error (MdAPE)			Mean Absolute Percentage Error (MAPE)		
	1-yr	6-yr	Cum	1-yr	6-yr	Cum	1-yr	6-yr	Cum	1-yr	6-yr	Cum
Random Walk	1.00	1.00	1.00	1.00	1.00	1.00	5.6	25.2	16.3	7.6	26.1	17.8
Equal-Weights	0.70	0.65	0.74	0.69	0.74	0.71	4.3	18.4	10.4	5.6	20.5	13.1
Rule-Based Forecasting	0.63	0.48	0.55	0.67	0.48	0.67	3.2	14.2	8.7	6.3	23.6	15.0

Conditions Analysis

The conditions analysis examines the performance of alternative procedures for series having different sets of features. Our hypotheses had been specified in term of conditions. We expected rule-based forecasting to produce more accurate forecasts in conditions involving significant trends, low uncertainty, stability, and high domain knowledge. In other conditions, we expected rule-based forecasting to perform as well as the better of the random walk or equal-weights. Because we developed the hypotheses prior to analyzing the data, the test was conducted on the full set of 126 series: To examine these hypotheses, we compared errors from rule-based forecasting (using the rules in Appendix B) with those from equal-weights and the random walk using the Median Cumulative Relative Absolute Error (MdCumRAE). The MdCumRAE controls for scale, amount of change, and outliers. Table 4 summarizes the results when the series are classified by conditions. The MdCumRAE for the random walk is 1.0 by definition.

Table 4
Testing the Relative Accuracy of Rule-Based Forecasting under Various Conditions

Conditions	Number of Series	Median CumRAE	
		Equal-Weights Combining	Rule-Based Forecasting
Trend (Basic)			
Not Statistically Significant	22	1.06	0.96
Statistically Significant ($t > 2$)	104	0.67	0.54*
Uncertainty			
Coef of Variation > 0.2	26	0.87	0.89
Coef of Variation < 0.2	100	0.68	0.55*
Instability Features			
More than Two	20	0.91	0.92
Two	48	0.73	0.73
One	39	0.71	0.55*
None	19	0.45	0.23*
Domain Knowledge			
Causal Forces Unknown	22	0.83	0.87
Causal Forces Known	104	0.71	0.55*

* Rule-Based Forecasting more accurate than Equal-Weights at $p \leq 0.01$.

For series with no trends, rule-based forecasting performed as well as the random walk. As expected, rule-based forecasting was more accurate than the random walk for series with statistically significant trends in the historical data; its error was 46% less than that for the random walk. Here it was also more accurate than equal weights although we had not hypothesized any difference ($p = 0.02$ using a two-tailed Wilcoxon test of the CumRAE).

As expected, rule-based forecasting showed little superiority over the random walk for series with high uncertainty. But for series with low uncertainty, rule-based forecasting was more accurate than the random walk, as the MdCumRAE was 0.55. We had not hypothesized rule-based forecasts to be more accurate than equal-weights forecasts for series with low uncertainty. Nevertheless, the rule-based forecasts were more accurate ($p = 0.002$ using a two-tailed Wilcoxon test of the CumRAE).

For series with much instability (two or more instability features), we did not expect rule-based forecasting to be superior to equal weights or the random walk. In fact, it was a bit more accurate than the random walk, but slightly less accurate than equal-weights. When only one instability feature was present, rule-based forecasting was significantly more accurate than equal-weights ($p = 0.01$ using a one-tail Wilcoxon test of the CumRAE). A similar superiority was found when the series had no instabilities.

As the stability of the series increased, rule-based forecasting improved in comparison to the random walk. This result is also consistent with our hypotheses; for the four categories, the Spearman rank-order correlation coefficient was 1.0 ($p < 0.05$). Rule based forecasting also improved relative to equal-weights as stability increased. This improvement supports our hypothesis and is statistically significant ($p < 0.05$).

This analysis supported the hypothesis that rule-based forecasting performs as well as equal weights or the random walk under adverse conditions. This is important as some methods did not perform well for all conditions. For example, for the 22 series that did not have statistically significant historical trends, Holt's exponential smoothing yielded an error that was 50% greater than the error from the random walk.

We examined the effect of domain knowledge by comparing the errors for series where the causal forces were known against those for series where they were unknown. As expected, rule-based forecasting was more accurate than the other procedures for series with known causal forces. For such series it was 45% more accurate than the random walk (MdRAE of 0.55) and 23% more accurate than equal-weights ($p = 0.01$ using a one-tailed Wilcoxon test on the CumRAE). Where the causal forces were unknown, rule-based forecasting and equal-weights were comparable.

Overall, this analysis of the CumRAE was consistent with the hypotheses on nine of 10 hypotheses; it showed statistically significant differences in the direction hypothesized and no significant differences where none were hypothesized. We repeated the analysis using the RAE for one-ahead and six-ahead forecasts and obtained further support with agreement on 15 of 20 Wilcoxon comparisons. Finally, we repeated the analysis using the Absolute Percentage Error for one- and six-ahead forecasts and found consistency on 18 of the 20 tests.

Limitations

The extent of calibration was limited for many rules. For example, rule 4 was employed for only two series. With so few uses, many rules are likely to be poorly specified. Consequently, we expect that significant improvements can be made in the rule base.

Domain expertise was general and limited. The descriptions of the data were not highly informative and we had only general knowledge. The use of specific and accurate domain expertise should improve accuracy.

We did not establish that the judgmental aspects of the feature identification process can be replicated. Although this process was reliable, the two authors share a common experience.

Although the M-competition data. represent a variety of economic and demographic series, we know little about the extent to which the rule base generalizes to other types of data. We expect that new situations will create a need to add rules and to refine the existing rules.

Because the M-competition data have been subjected to extensive prior study, we may have learned about rules that are relevant only for these types of data. Furthermore, the calibration may have yielded rules that apply well only for the mix of series encountered in these data. We plan to extend the validation to other types of data.

The results apply to annual data. We expected rule-based forecasting to be appropriate for annual data because they typically have significant trends, stable patterns, and strong causal forces. We do not know how rule-based forecasting will perform for quarterly, monthly, weekly, and daily data, especially where seasonal effects are strong. This study also used long calibration series (median of 15 years); rule-based forecasting benefits from long series because it uses information about patterns in the data.. We do not know how the procedure will perform for short series.

Discussion

Many of the rules can be applied to existing systems. This can be done using groups of rules (such as those relating to causal forces) or individual rules.

Rule-based forecasting applies the cumulative knowledge of forecasting to a given situation. Knowledge about forecasting methods can be passed along in the rule base. As new research findings are established, they can be incorporated into the rule base. Because it provides for the systematic integration of research on forecasting, rule-based forecasting should aid further research. It could help to identify areas for further research. Rule-based forecasting offers a vehicle for testing generalizations about forecasting.

Additional Validation

Our data did not allow for calibration of all rules. Some rules were used on all 126 series, while some were never used. Several potentially important rules, such as adjustments in trends when the last observation is unusual (rules 48 and 84), were employed on few series. Further validation should use large data bases with heterogeneous series. In addition, simulated data might be used to test the effects of variations in trends, uncertainty, stability, and domain knowledge.

Rules will change as researchers examine different types of data. Before rule-based forecasting can be generalized to the many types of data encountered in practice, additional validation and model refinement will be required. Researchers should examine data from a variety of environments.

Our results only drew upon *general* domain knowledge. We anticipate significant gains from using specific domain knowledge.

Feature Identification

Currently, it is necessary to make judgments about many of the features. Some of these judgments could be replaced by statistical procedures. The automatic identification of features would improve reliability, which, in turn, should improve accuracy. It also would reduce costs and thus aid in validations using a large number of series.

In some situations, humans May be better able to identify features than are statistical procedures. For example, we obtained more accurate forecasts when we used our judgment to identify the functional form (original or log transformed) than when we used the functional form that provided the best fit to the historical data. Improvements are needed in the reliability of this judgmental process.

Selection Rules

The conditions analysis suggested that some extrapolation methods may dominate in certain situations. If these situations can be reliably identified, a selection procedure might be used instead of combining. For example, for series where causal forces are unknown, instability is high, and no significant long-term trend exists, one might select the random walk. The random walk is simple and inexpensive. Further research is needed on the value of rule-based selection procedures.

Conclusions

With rule-based forecasting, one can apply expertise about forecasting methods and domain knowledge that is appropriate to the conditions of time series. Prior literature suggests that such an approach should improve the accuracy of extrapolation forecasts under certain conditions.

Rule-based forecasting produced substantially more accurate forecasts than the random walk for annual series in three validation tests. Rule-based forecasting was also more accurate than equal-weights combining, a procedure that has performed well in prior studies. These improvements in accuracy with respect to equal-weights were statistically significant. As expected, the gains in accuracy were largest for series with significant trends, low uncertainty, stability, and good domain knowledge. The gains in accuracy were larger for long-term forecasts; in three validations involving six-year-ahead ex *ante* forecasts for a total of 90 series, errors for rule-based forecasting were 42% less than those from equal-weights combining. In general, the improvement will depend on the conditions of the data.

The rule-based forecasting procedure offers promise. We provide our rules as a starting point. Hopefully, they will be replaced by simpler and fewer rules.[6]

Appendix A: Definitions of the Features

Identified by Rules

Coefficient of Variation. The standard deviation divided by the mean for the trend adjusted data in original units. The trend-adjusted series removes the effect of the basic trend as estimated from the regression.

Direction of Basic Trend. The direction of the trend (up or down) that results from fitting a linear regression to the historical data.

Direction of Recent Trend. The direction of the trend that results from fitting Holt's model to the historical data.

Near a Previous Extreme. A last observation that is greater than 90% of the highest, or less than 110% of the lowest historical observations previously recorded for the trend-adjusted series, provided that the previous extreme observation is not the one immediately preceding the last one.

Outlier. An isolated observation, other than the last observation, that is outside a two-standard-deviation band of the linear regression.

Recent Run not Long. The period-to-period movements for the last six observations have not all been in the same direction.

Significant Basic Trend. The *t*-statistic for the linear regression is greater than 2.0.

Variation about Trend. The coefficient of determination (R^2) for the linear regression on the chosen functional form for the series.

Input by the Analyst Based on Inspection of the Data

Changing Basic Trend. An underlying trend that is obviously changing over the historical period. This determination is based on an understanding of the process and an examination of the data (or transformed data where relevant).

Irrelevant Early Data. An early portion of a series that is believed to have resulted from a substantially different underlying process. For example, the start-up period for sales of a product would be eliminated once the pattern has stabilized. In truncating a series to eliminate such a period, avoid starting the truncated series with an extreme observation.

Last Observation Unusual. A last observation that deviates substantially from the previous data. Adjustments of historical observations should be made before identifying an unusual last observation.

Level Discontinuities. Changes in the level of the series (steps). One can identify such steps based on a knowledge of the series (e.g., knowledge that a change was made in the measurement process for the data) and an examination of a plot of the series (or transformed data where relevant).

Suspicious Pattern. Series that show a substantial change in the recent pattern.

Unstable Recent Trend. Series that show marked changes in the recent trend pattern. To identify unstable trends, one can examine a graph of the series (or transformed data where relevant). One must be careful with cyclical series to determine whether there is instability in addition to the cycle.

Input by the Analyst Based on Domain Knowledge

Adjusted Observations. Revisions based on knowledge of unusual events, such as strikes, stockouts, wars, or natural disasters, to remove the short-term effects.

Causal Forces. The net directional effects of the principal factors acting-on the series. There are five types of causal forces. "Growth" exerts an upward force at all times. "Decay" exerts a downward force at all times. "Supporting" forces push in the direction of the historical trend. "Opposing" forces work against the trend. Finally, "Regressing" forces exert pressure towards a mean. Forces should only be specified when they are obvious; otherwise, they should be identified as "Unknown."

Cycles Present. Regular movements of the series about the basic trend.

Functional Form. Expresses the expected pattern of the trend of the series. This rule base is limited to multiplicative and additive forms.

Appendix B: The Rule Base

The rules used in Validation 3 are presented here. Explanations are provided in braces. Additions and subtractions of weights are constrained to produce positive weights that sum to one.

Adjusting the Historical Data and Identifying Features

1. *Truncation.* IF there are irrelevant early data, THEN delete these data. {Prior data are relevant for extrapolation only if the underlying process is stable.}

2. *Functional Form.* IF the functional form is multiplicative, THEN use a log transformation of the original series. {Fitting a log transformation of the original series assumes a multiplicative growth.}

3. *Observations Adjusted.* IF observations are judged to be irregular based on domain knowledge, THEN adjust the observations prior to analysis to remove their short-term effects. {The transitory effects of unusual events should not have an impact on the forecast, thus the observations should be adjusted.}

4. *Last Observation Unusual.* IF the last observation is unusual, THEN replace it by -the average of the last observation and the rule-based forecast for this period that was made at t_{-1}. {The last observation is especially important. This rule reduces the impact of the last observation when it is subject to much uncertainty.}

5. *Outliers.* IF an observation is an outlier, THEN set the observation equal to two standard deviations from the regression line. {This guards against mistakes or events that may have only transitory effects.}

6. *Recent Trend Down.* IF the trend component from Holt's is down, THEN set the direction of the recent trend to down.

7. *Basic Trend Down.* IF the trend component from the regression is down, THEN set the direction of the basic trend to down.

8. *Significant Basic Trend.* IF the t-statistic for the regression's trend is greater than 2, THEN set the significant basic trend to true.

9. *Recent Run not Long.* IF there has been a movement in the same direction for at least the last six observations, THEN set the 'recent run not long' to false.

10. *Near a Previous Extreme.* IF the last observation is greater than 90% of the highest, or less than 110% of the lowest historical observation for the trend adjusted series, AND it is not immediately preceded by the previous extreme observation, THEN set near a previous extreme to true.

Short-Range Model: Selecting Smoothing Factors for Level (Brown's Alpha)

11. *Base Value.* Set alpha for the short-range model to 0.7. (A starting value is needed. Given a lack of other information, this rule places much weight on the last observations.)

12. *Variation.* Multiply alpha for the short-range model by the variation about the trend. {When there is much variation about the trend, uncertainty is high: It is then desirable to base the estimate of the current level on a larger number of observations.}

13. *Last Observation Unusual.* IF the last observation is unusual, THEN subtract 0.2 from alpha for the short-range model. {Given uncertainty about this observation, its impact is reduced.}

14. *Discontinuities in Level.* IF there are level discontinuities, AND the variation about the trend is greater than 0.9, THEN add 0.1 to alpha for the short-range model. {.Smooth series with previous discontinuities are candidates for future shifts. A higher alpha helps the series to incorporate information about a shift in the level.}

15. *Causal Force.* IF the direction implied by the causal force is the same as the direction of the recent trend AND the variation about the trend is greater than 0.9, THEN add 0.1 to alpha for the short-range model. {For smooth series, the estimate of the level should be more responsive to recent changes when they are in the expected direction.}

16. *Unstable Recent Trend.* IF there is an unstable recent trend, THEN add 0.1 to alpha for the short-range model. {The estimate of the level should be updated more rapidly when trend changes are taking place.}

17. *Maximum Value.* IF alpha for the short-range model is calculated to be greater than 0.7, THEN use 0.7. { This avoids putting a very high weight on the last observation.

18. *Minimum Value.* IF alpha for the short-range model is less than 0.2, THEN use 0.2. {This avoids putting a very low weight on the last observation.

Short-Range Model: Selecting Smoothing Factors for Recent Trend (Brown's Beta)

19. *Base Value.* Set beta for the short-range model to 0.7. {A starting value is needed. Given a lack of other information, beta is set equal to alpha for the short-range model.}

20. *Variation.* Multiply beta for the short-range model by the variation about the trend. {More variation about the trend implies more uncertainty. More uncertainty calls for a larger number of observations.}

21. *Last Observation Unusual.* IF the last observation is unusual, THEN subtract 0.4 from beta for the short-range model. {The last change has a strong impact upon the trend estimate. Less emphasis is given to the most recent change when there is reason to be cautious. }

22. *Discontinuities in Level*. IF there are level discontinuities, AND the variation about the trend is greater than 0.9, THEN subtract 0.1 from beta for the short-range model. {For smooth series, a shift in the level should not unduly affect the trend estimate. A lower beta incorporates more data, thus putting relatively less emphasis on the shift in the trend.}

23. *Causal Force.* IF the direction implied by the causal force is the same as the direction of the recent trend AND the coefficient of variation is greater than 0.9, THEN add 0.1 to beta for the short-range model. {For smooth series, this puts more emphasis on the recent trend when its direction is consistent with the causal forces.}

24. *Unstable Recent Trend.* IF there is an unstable recent trend, THEN subtract 0.2 from beta for the short-range model. {When the recent trend is likely to shift, less weight should be placed on the most recent estimates of trend.}

25. *Changing Basic Trend.* IF the basic trend has been changing, THEN add 0.3 to beta for the short-range model. {when the basic trend has been changing, this puts more weight on the recent trend.}

26. *Maximum Value.* IF beta for the short-range model is greater than 0.7, THEN use 0.7. {This avoids a very high weight on recent observations.}

27. *Minimum Value.* IF beta for the short-range model is less than 0.2, THEN use 0.2. {This avoids a very low weight on recent observations.}

Short-Range Model: Estimating Levels

28. *Benchmark Weights.* Use 20%, 0%, 40%, and 40% to weight the estimates of the level from the random walk, linear regression, Holt's, and Brown's methods, respectively. {Whereas some credence should be given to the use of

equally-weighted combined estimates, the regression estimate normally is not appropriate as its mean lags the data. The random walk is subject to measurement error, so it is weighted less.}

29. *Level Discontinuities.* If there are level discontinuities, THEN add 10% to the weight on the random walk and subtract it from that on Brown's and Holt's. {In a series subject to discontinuities in the level, the last observation may contain information about a new shift.}

30. *Near Extreme.* IF the last observation is near a previous extreme AND the series has cycles present, THEN subtract 10% from the weight on the random walk and add it to that on the regression and on Brown's. {When the last observation is near a previous extreme, it may be there due to chance or to transient factors. There may also be a bound on the series. Accordingly, less emphasis is given to the latest observation.}

31. *Suspicious Pattern.* IF there is a suspicious pattern, THEN add 10% to the weight on the random walk and subtract it from that on the other three methods. {If a pattern in the series looks suspicious, fundamental changes may be taking place. Therefore, more reliance is placed on the last observation in estimating the level.}

32. *Unstable Recent Trends.* IF there is an unstable recent trend, THEN add 30% to the weight on the random walk and subtract it from that on the other three methods. {This makes the estimate of the level more responsive to recent information when the recent trend is changing in unusual ways.}

33. *Changing Basic Trend.* IF there is a changing basic trend, THEN add 15% to the weight on the random walk and subtract it from that on the other three methods. {This makes the estimate of the level more responsive to recent information when the basic trend is changing.}

34. *Last Observation Consistent with Causal Forces.* IF the difference between the last observation and the current level is in the same direction as the causal forces, THEN replace the level by $L' = L + 0.3$ {abs $(x - L)$/S] }, where x is the last observation, L is the level as determined from application of the preceding rules, and S is the standard deviation of the trend-adjusted series. {When the last observation is higher than the estimated current level and movement toward this observation agrees with the causal forces, more weight is given to this latest observation, especially if the series is smooth.}

35. *Last Observation Inconsistent with Causal Forces.* IF the difference between the last observation and the current level is not in the same direction as the causal forces, THEN replace the level by $L' = L - 0.3$ {abs [$(x - L)$/S]}, where x is the last observation, L is the level as determined from application of the preceding rules, and S is the standard deviation of the trend adjusted series. {This pays less attention to the last observation when movement toward it would be against the causal forces.}

36. *Mechanical Adjustment—Causal Forces Unknown.* IF NOT 'last observation unusual' AND the direction of the causal forces is unknown, THEN add 12.5% of the difference between the last observation and the one-ahead rule-based forecast made at t_{-1}. {This rule adjusts the level for potential systematic errors.}

37. *Mechanical Adjustment—Causal Forces Agree with Direction of Forecast Error for Last Observation.* IF NOT last observation unusual AND the direction of the causal forces is the same as the last observation minus the one-ahead rule-based forte made at t_{-1}, THEN add 15% of the difference between the last observation and that corresponding rule-based forecast. {When the actual minus the forecast for period t is in the direction expected by the causal forces, more credence should be placed on the last observation.}

38. Mechanical Adjustment—Causal Forces Oppose Direction of Forecast Error for Last Observation. IF NOT last point unusual AND the direction of the causal forces is opposite to the last observation minus the one-ahead forecast from the rule-based forecast at t_{-1}, THEN add 10% of the difference between the last observation and that corresponding rule-based forecast. {This is the converse of rule 37.}

Short-Range Model: Estimating Trends

39. *Benchmark Methods.* Use 0%, 20%, 40%, and 40% to weight estimates of the trends from the random walk, linear regression, Holt's, and Brown's, respectively. {The exponential smoothing methods, with their heavier weight on recent data, are most appropriate for short-range forecasts.

40. *Causal Forces Unknown.* IF the causal forces are unknown, THEN add 5% to the weight on the random walk, and subtract it from that on the regression trend estimate. {When the causal forces are unknown, trend extrapolation should be more cautious.}

41. *Dissonance.* IF the direction of the basic trend and the direction of the recent trend are not the same OR if the trends agree with one another but differ from the causal forces, THEN add 15% to the weight on the random walk and subtract it from that on the other trend estimates. {Dissonance calls for conservatism in the trend estimate.}

42. *Inconsistent Trends.* IF the direction of the basic trend and the direction of the recent trend are not the same AND the basic trend is not changing, THEN add 20% to the weight on the regression trend and subtract it from that on Brown's and Holt's. {When the trends are inconsistent, the recent and basic trends should be more evenly balanced.}

43. *Incoherent Long Series.* IF the causal forces are not in the same direction as the basic trend, THEN subtract 30% from the weight on the regression and add it *to that on* Brown's and Holt's. {When the forces oppose the basic trend, less weight is placed on that trend estimate, and more is placed on the recent trend.}

44. *Recent Run.* IF the recent run is long, THEN subtract l0% from the weight on the regression and add it to that on Brown's and Holt's. {A long run suggests that the recent trend is stable; so that more weight should be placed on the recent trend.}

45. *Unstable Recent Trend.* IF there is an unstable recent trend, THEN add 20% to the weight on the random walk and subtract it from that on Brown's and Holt's. {When the recent trend is unstable, it should have less effect on the trend estimate.}

46. *Suspicious Pattern.* IF there is a suspicious pattern, THEN add 10% to the weight on the random walk and subtract it from that on the other three methods. {Suspicion calls for conservatism with respect to trend extrapolation.}

47. *Insignificant Basic Trend.* IF NOT a significant basic trend, THEN add 5% to the weight on the random walk and subtract it from that on the regression. {When the long-term trend is not statistically significant, put less weight on it.}

48. *Last Observation Unusual.* IF the last observation is unusual, THEN add 10% to the weight on the regression and subtract it from that on Brown's and Holt's. {This makes the trend estimate more reliant on earlier observations.}

Long-Range Model: Selecting Smoothing Factors for Level (Brown's Alpha)

49. *Base Value.* Set alpha for the long-range model to 0.6. {To incorporate more data in the long-range model, lower values are selected for alpha than for the short-range model.}

50 *through* 54. Same as rules 12 through 16.

55. *Maximum Value.* IF alpha for the long-range model is greater than 0.6, THEN use 0.6. {This avoids a very heavy weight on the last observation.}

56. *Minimum Value.* IF alpha for the long-range model is less than 0.1, THEN use 0.1. {This avoids a very low weight on the recent observation.}

Long-Range Model: Selecting Smoothing Factors for Trend (Brown's Beta)

57. *Base Value. Set* beta for the long-range model to 0.6. {For the long-range model, lower values are selected for beta than in the short-range model in order to incorporate more data.}

58 *through* 63. Same as 20 through 25.

64. *Maximum Value.* IF beta for the long-range model is greater than 0.6, THEN use 0.6. {This avoids a very heavy weight on the recent observations.}

65. *Minimum Value.* IF beta for the long-range model is less than 0.1, THEN use 0.1. {This avoids a very low weight on the recent observations.}

Long-Range Model: Estimating Level

66-68. (Same as rules 28-30.)

69. *Stable Basic Trend.* IF NOT a changing basic trend, THEN add 5%d to the weight on the regression and subtract it from that on the random walk. {Because the regression estimate incorporates more data than does the random walk, the long-range model should place a heavier weight on the regression.}

70-74. (Same as rules 31-35.)

Long-Range Model: Estimating Trend

75-84. (Same as rules 39-48.)

85. *Stable Basic Trend.* IF NOT a changing basic trend, THEN add 15% a to the weight on the regression and subtract it from that on Holt's and Brown's. {Basic trend is more appropriate for the long-range model.)

86. *Inconsistent Trends.* IF the directions of the recent and basic trends are not the same, THEN subtract 10% from the weight on the regression and add it to that on the other methods. {When the causal factors are unknown and the trends are inconsistent, less emphasis is given to the basic trend, even for the long-range model. This makes the allocation among the various trend estimates more equal.}

87. *Changing Basic Trend.* IF there is a changing basic trend, THEN add 20% to the random walk and 5% to Brown's and subtract it from that on the regression. {Given that the process seems to be changing, the basic trend estimate is less relevant and the random walk is used for conservatism.}

88. *Regression to Mean.* IF the causal force is regressing, THEN set the trend for the long-range model equal to $T' = 0.2 ? T + 0.8 (M - L)/(P - R)$, subject to $R > 0.5 P$, where T is the trend for the long-range model as estimated from the preceding rules, M is the long-range mean to which the series tends, L is the level as determined by the rules, F is the number of periods it takes to go from an extreme to the long-range level, and R is the number of periods during which the series has been moving toward the long-range level. {When the forces are regressing, a trend is calculated by going from the last observation to the "mean" to which the series is regressing. The constraint is designed to move the forecast cautiously in the direction of the long-range mean.}

Long-Range Model: Damping Trend

89. *Unknown Causality.* IF causal forces are unknown, THEN damp the long-range model trend estimate by 5%. {With less reason to expect a continuation of the trend, the trend estimate should be cautious.}

90. *Inconsistent Model Trends.* IF the directions of the basic and recent trends are different, THEN damn the trend by 5%. {Given conflicting information about trend, its' estimate should be conservative.}

91. *Incoherent Forces:* IF the directions of the basic and recent trends are contrary to the causal force, THEN damp the trend by 5% for each contrary trend. (Thus, if the forces are contrary to both the recent and basic trends, damp the trend by 10%.) {When the trends go against causal expectations, the trend estimate should be more conservative.}

92. *Uncertainty.* IF the direction of the causal forces is the same as the direction of the trend from long-range model, THEN add $(1 - R^2)/B$ to the damping factor; ELSE add $2(1 - R^2)/B$ to the damping factor. (B is the number of periods before the long-range model reaches 100% of the forecast, and R^2 from the linear, regression is expressed in decimal form.) {Increase the damping as the uncertainty about the regression line increases. Damping is further increased if the direction of the trend for the long-range model is inconsistent with the causal forces.}

93. *Suspicious Pattern.* IF there is a suspicious pattern, THEN damp the trend by 5%. {For a suspicious pattern, the trend estimate should be more conservative.}

94. *Unstable Recent.* IF there is an unstable recent trend, THEN damp the trend by 10%. {With added uncertainty about the trend, the trend estimate should be more conservative.}

95. *Horizon.* Multiply the trend for the long-range model by $(1 - D)^{h-1}$, where D is the damping factor and h is the forecast horizon. {As the forecast horizon lengthens, uncertainty increases. Accordingly, the damping factor is cumulative.}

Blending the Forecasts from the Short-Range and Long-Range Models

96. *Blend Period.* IF data are annual, THEN the blend period is 6. {It, is assumed that a six-year period is long enough for the causal forces to have had a significant impact for most series.}

97. *Standard Blend.* IF the trends from the short-range and long-range models are in the same direction. OR if the causal forces are unknown, THEN $L_h = 1 - \{(100/B) \times [(1 + B - h)/100]\}$, where L_h *is the* percentage of the long-range model used in forecasting horizon h, and B is the blend period, the number of periods over the forecast horizon until the long-range model equals 100%. {This rule blends the two models gradually from 100% short-range model in horizon 1 to 100% long-range model in horizon B. B can exceed the forecast horizon.}

98. *Quick Blend.* IF the short-range model direction conflicts with the long-range model direction AND the causal force direction is the same as the long-range model, THEN set the share of the long-range model to $\sum_1^h(x)/\sum_1^B(y)$ where h is the horizon and B is the blend period. {When the causal forces agree with the long-range model direction and disagree with the short-range model, this moves the forecast more quickly to the long-range model.}

99. *Slow Blend.* IF the short-range model direction conflicts with the long-range model direction AND the *causal force* direction is the same as the short-range model, THEN set the share of the long-range model to $\sum_1^{1+B-h}(x)/\sum_1^B(y)$ where h is the horizon and B is the blend period. {This is the converse of rule 98.}

References

Armstrong, J. S. (1985), *Long-Range Forecasting*. (2nd Ed.), John Wiley, New York.

Armstrong, J. S. (1988), "Research needs in forecasting," *International Journal of Forecasting*, 4, 449-465.

Armstrong, J. S. and F. Collopy (1992a), "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, 8, 69-80.

Armstrong, J. S. and F. Collopy (1992b), "Causal forces: Structuring knowledge for time series extrapolation," *Journal of Forecasting*, (forthcoming).

Armstrong, J. S. and E. J. Lusk (1983), "The accuracy of alternative extrapolation models: Analysis of a forecasting competition through open peer review," (with commentary), *Journal of Forecasting*, 2, 259-311.

Brown, R. G. (1959). *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York.

Buchanan, B. G. and E. H. Shortliffe (1985). *Rule-Based Expert Systems*. Addison-Wesley, Reading, MA.

Bunn, D. and G. Wright (1991), "Interaction of judgmental and statistical forecasting methods: Issues & analysis," *Management Science*, 37, 501-518.

Carbone, R. and J. S. Armstrong (1982), "Evaluation of extrapolative forecasting methods: Results of a survey of academicians and practitioners," *Journal of Forecasting*, 1, 215-217.

Carbone, R. and S. Makridakis (1986), "Forecasting when pattern changes occur beyond the historical data," *Management Science*, 32, 257-271.

Clemen, R. T. (1989), "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, 5, 559-583

Collopy, F. and J. S. Armstrong (1989), "Toward computer-aided forecasting systems," in G. R. Widemeyer (Ed.), *DSS 89 Transactions, TIMS* College on Information Systems. Providence, RI, 103-119. (Copies of this paper are also available from the authors.)

Collopy, F. and J. S. Armstrong (1992), "Expert opinions about extrapolation and the mystery of the overlooked discontinuities," *International Journal of Forecasting* (forthcoming).

Ericsson, K. A. and H. A. Simon (1984). *Protocol Analysis: Verbal Reports as Data.* MIT Press, Cambridge, MA.

Fildes, R. and E. J. Lusk (1984), "The choice of a forecasting model," *Omega*, 12, 427-435.

Gardner, E. S., Jr., "Exponential smoothing: The state of the art," *Journal of Forecasting,* 4 (1985), 1-28.

Grabowski, M. (1988), "Knowledge acquisition methodologies: Survey and empirical assessment," In J. I. deGross and M. H. Olson (Eds.), *Proc. 9th International Conference Information Systems,* 45-54.

Holt, C. C., F. Modigliani, J. F. Muth, and H. A. Simon (1960). *Planning Production Inventories and Work Force.* Prentice-Hall, Englewood Cliffs, NJ.

Lawrence, M. J., R. H. Edmundson and M. J. O'Connor (1986), "The accuracy of combining judgmental and statistical forecasts," *Management Science*, 32, 1521-1532.

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 Royal Statististics Society Series A*, 142, 97-145.

Mathews, B. P. and A. Diamantopoulos (1990), "Judgmental revision of sales forecasts: Effectiveness of forecast selection," *Journal of Forecasting,* 9, 407-415.

Mentzer, J. T. and J. E. Cox (1984), "Familiarity, Application, and Performance of Sales Forecasting Techniques," *Journal of Forecasting,* 3, 27-36.

Rosenthal, R. (1978), "Combining Results of Independent Studies," *Psychological Bulletin,* 85, 185-193.

Sanders, N. R. and L. P. Ritzman (1990a), "Some empirical findings on short-term forecasting: Technique complexity and combinations," *Decision Science,* 20, 635-640.

Sanders, N. R. and L. P. Ritzman (1990b), "Improving short-term forecasts," *Omega*, 18, 365-373.

Schnaars, S. P. (1984), "Situational Factors Affecting Forecast Accuracy," *Journal of Marketing Research*, 21, 290-297.