Rule-Based Forecasting:
Using Judgment in Time-Series Extrapolation

J. Scott Armstrong
The Wharton School, University of Pennsylvania

Monica Adya
Department of Information Systems, U. of Maryland - Baltimore County

Fred Collopy
The Weatherhead School of Management, Case Western Reserve University

ABSTRACT

Rule-Based Forecasting (RBF) is an expert system that uses judgment to develop and apply rules for combining extrapolations. The judgment comes from two sources, forecasting expertise and domain knowledge. Forecasting expertise is based on more than a half century of research. Domain knowledge is obtained in a structured way; one example of domain knowledge is managers= expectations about trends, which we call "causal forces." Time series are described in terms of 28 conditions, which are used to assign weights to extrapolations. Empirical results on multiple sets of time series show that RBF produces more accurate forecasts than those from traditional extrapolation methods or equal-weights combined extrapolations. RBF is most useful when it is based on good domain knowledge, the domain knowledge is important, the series is well behaved (such that patterns can be identified), there is a strong trend in the data, and the forecast horizon is long. Under ideal conditions, the error for RBF's forecasts were one-third less than those for equal-weights combining. When these conditions are absent, RBF neither improves nor harms forecast accuracy. Some of RBF's rules can be used with traditional extrapolation procedures. In a series of studies, rules based on causal forces improved the selection of forecasting methods, the structuring of time series, and the assessment of prediction intervals.

Keywords: accuracy, causal forces, combining forecasts, consistent trends, contrary series, cycles, damped trends, decay forces, decomposition, discontinuities, expert systems, exponential smoothing, extrapolation, growth forces, inconsistent trends, instabilities, judgment, opposing forces, outliers, regressing forces, reinforcing series, start-up series, supporting forces.

Many organizations need sales forecasts for thousands of products. Their forecasters commonly use extrapolation methods. However, extrapolation methods usually ignore managers= knowledge. Managers often have valuable information, including knowledge of events that occurred but are not yet reflected in a time series (e.g., strike, stockout). They might also know what past events are not likely to recur (e.g., natural disasters or new government regulations). Sometimes they know about anticipated events that have not occurred in the past (e.g., major product improvement, emergence of a new competitor).

In commentaries on the M-competition (Armstrong and Lusk 1983), some researchers called for the integration of judgment and statistical extrapolations. In reviewing the literature, Bunn and Wright (1991) also recommended integrating judgment and statistical forecasting methods. Armstrong and Collopy (1998) reviewed the literature on judgmental integration and located 47 empirical studies, mostly published since 1985. They concluded that domain knowledge should be structured and used as an input to statistical models.

Rule Based Forecasting (RBF) is an expert system that uses domain knowledge to combine forecasts from various extrapolation methods. Using production (if-then) rules, RBF determines what weights to give to the forecasts. Features of the situation are identified in the conditional (if) part of the rules, and weights are adjusted to

match features to the underlying assumptions of the methods. In effect, RBF uses structured judgment to tailor extrapolation methods to situations. RBF also draws upon knowledge about forecasting gained through research over the past half century. This includes findings about such practices as combining forecasts and damping trends.

GATHERING KNOWLEDGE FOR RBF

Knowledge for rules can be obtained from expert judgment, theory, and prior empirical research. Fortunately, much useful empirical research has been conducted on time-series forecasting over the past several decades.

- **Initially, use forecasting expertise for knowledge about rules.**

Although forecasting expertise can be found in the literature, these sources often fail to adequately describe conditions under which a method is be expected to be successful. Nevertheless, some rules were apparent from the literature. One example is to combine forecasts when one is uncertain which is the best method.

Interviews are a good source of knowledge. Collopy and Armstrong (1989) asked experts, in structured interviews, to describe rules they considered important for extrapolation. The interviews gave insight into general strategies but produced few usable rules. For example, one expert believed forecasts should rely heavily on the random walk whenever there were "high uncertainties or substantial irregularities in the historical data." The experts tended to state rules in vague terms and seldom identified specific conditions to govern their use.

- **Use protocols to identify knowledge about rules.**

Protocols are often used for acquiring knowledge. In protocols, an analyst observes as an expert performs a task. The expert describes what he is thinking while doing the task.

We found protocols to be the most useful way of gathering knowledge for RBF. The third author asked five experts in forecasting to talk about the processes they used as they examined six time-series (Collopy and Armstrong 1989). All of these experts were active researchers on forecasting methodology, and each had written at least one book on forecasting. They were asked to describe what method would be most appropriate and what adjustments they would make while applying the method. For example, one of the experts said to damp the trend when large changes occurred in the most recent observations. Another expert said that, faced with a change in slope in a time-series, he would truncate the series before fitting a model.

The third author had interviewed some of these experts prior to the protocols. Even experts who thought that they had given good descriptions of their extrapolation processes in that interview provided new insights during the subsequent protocol analysis. The initial protocol sessions did not provide rules for all the conditions forecasters are likely to encounter, so further sessions with one of the experts continued over several months.

FORMULATING RULES

As knowledge is obtained, one can convert it into rules or if-then statements. Formulating rules is a complex task. We discuss general guidelines that apply to extrapolation. We then show how these guidelines are translated into rules by describing some of the roughly one hundred conditional statements encoded in RBF.

- **Give separate consideration to level and trend.**

Decomposition of a time series into level and trend has long proven useful in extrapolation. However, many forecasters ignore this when they assume that they must use a single model to forecast both level and trend. Decomposition is valuable in RBF because domain knowledge can affect the estimates of level and trend differently.

- **Use simple extrapolation methods.**

As discussed in Harvey (2001), forecasters often assume that a complex situation requires complex methods. However, complex methods increase the potential for mistakes and misunderstanding. Also, simple

extrapolation methods are typically as accurate as complex ones (Armstrong 1985, pp. 494-495, summarizes evidence). The issue of complexity versus accuracy is particularly important for new situations. Complex methods may do well under conditions for which they were designed, but tend to suffer when applied to new conditions.

- **Combine forecasts.**

Combining forecasts yields substantial benefits. In empirical studies, the combined forecast=s error is almost always substantially lower than the average error of the component forecasts and it is sometimes lower than the best component's (Armstrong 2001b). The original version of Rule-Based Forecasting was based on combined extrapolations from four methods: the random walk, linear regression against time, Holt=s linear exponential smoothing, and Brown=s linear exponential smoothing. Other extrapolation methods can and probably should be used in RBF. Armstrong (2001b) shows evidence suggesting that there are benefits to combining forecasts from at least five methods. Results from Gardner (1999) suggest that the damped-trend model might make a useful addition. Based on results from Fildes et al. (1998), the "robust trend" model, which uses the median trend value from a time series, is also promising.

Methods that differ substantially from one another are likely to incorporate more information than those that are similar. They can differ with respect to the data used, the way that data are processed, or both. Employing different methods may reduce the effects of biases in the component forecasts and lead to more accurate forecasts (Batchelor and Dua 1995).

- **Use different models for short- and long-term forecasts.**

The analyst can decompose the forecasting problem according to the length of the forecast horizon. One reason for doing so is that causal factors may operate differently over the forecast horizon. For example, while daily changes in the stock market have unknown causal factors, the long-term trends have growth factors. Separate models for short- and long-range extrapolations improved accuracy in a study by Carbone and Makridakis (1986). In RBF, the random walk is used to capture aspects of the short term, exponential smoothing for the medium term, and linear regression for the long term.

Forecasts from the long- and short-term models should be blended. For the short-term, most weight is placed on forecasts from the short-range model. The weight on the long-range model should increase as the forecast horizon increases.

- **Damp the trend as the forecast horizon lengthens.**

As uncertainty increases, the trend component of the forecasts should become more conservative. In RBF, to reflect this increased uncertainty, the magnitude of the trend is reduced as the forecast horizon lengthens.

THE ELEMENTS OF RULE-BASED FORECASTING

Exhibit 1 shows the basic structure of the Rule-Based Forecasting system. First, adjustments are made to the data and the features of the series are identified. Rules are then applied to produce short and long-range model forecasts. To formulate these models, RBF makes estimates of levels (at t_0) and trends for each model. For the long-range model, the rules may damp the trend over the forecast horizon. Finally, rules are applied to blend the forecasts from the short- and long-range models.

Exhibit 1
Structure for Rule-Based Forecasting
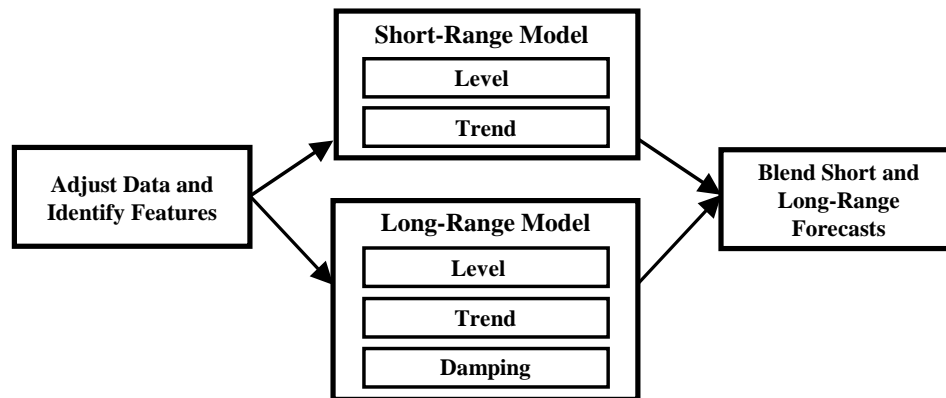
The "If" Part of the Rules

Critical to using RBF successfully is to describe the features of the time series accurately. These come from two sources: domain knowledge and historical data.

- **Use domain knowledge to define conditions that have affected or will affect the time series.**

We identified several conditions that relied on domain knowledge. These included information about the expected functional form, cycles, whether the series represents a start-up, the forecast horizon, historical adjustments of observations due to unusual events, and factors affecting trends.

Expectations about trends are especially important. This knowledge can be structured by asking domain experts to identify the type of "causal forces" affecting a series. The domain expert assesses the net directional effect of factors expected to affect the trend over the forecast horizon. For example, in forecasting unit sales of computers, one might consider rising incomes, increasing population, improvements in product capabilities, availability of software, and reduction in prices.

If causal forces tend to drive the series up, they are called *growth*. For example, when products are actively marketed in growing markets, the forces would be classified as growth. As noted, unit sales of computers have growth forces.

If forces tend to drive a series down, they are called *decay*. An example would be the costs of producing technical products, such as computers. Historical trends might fluctuate, but as long as the underlying forces are downward, the series is classified as decay.

If forces are expected to move against the historical trend, they are *opposing*. An example would be inventory levels relative to sales. When inventories get high, holding costs lead managers to reduce their levels. When they are too low, service suffers, prompting decisions to hold larger inventories.

If forces tend to move the series toward a mean, they are *regressing*. An example would be the performance of a professional athlete, such as a batting average; his average for the first three games of the current season would tend to regress toward his historical average. If he were a new player, his average might regress to the average for new players.

If forces reinforce the historical trend, they are called *supporting*. Here we have difficulty finding examples because information about the trend in a series is assumed to be the dominant factor affecting behavior, and other factors would be unimportant. This might occur over specific periods for sales of fashion crazes or fad items, or for market prices. For example, if real estate prices were going down, the perceived value of the neighborhood might go

down. If prices were going up, people might perceive this neighborhood as the place to live. In our experience with hundreds of time series, we have yet to encounter a series with supporting forces.

Exhibit 2 lists the types of causal forces. It shows the direction of the forces depending upon whether the historical trend is up or down. When one has little information about the factors that would affect a series over a forecast horizon, it is best to code the causal forces as "unknown." Also, when factors are known, but they operate in different ways so that their net effect is unknown, it is best to code the series as unknown.

Exhibit 2
Direction of Causal Forces Given Historical Trends

| Type of Causal Force | Causal force direction when . . . | | Examples |
	trend has been up	trend has been down	
Growth	Up	Up	Sales (units) Macroeconomic data
Decay	Down	Down	Production costs (units)
Opposing	Down	Up	Inventory as percent of sales
Regressing	Toward a known mean value	Toward a known mean value	Demographic (percent male births) Athletic performance
Supporting	Up	Down	Real estate prices
Unknown	?	?	Exchange rates

Coding causal forces (as to whether they are growth, decay, etc.) is generally simple, requiring domain experts to spend less than a minute per series. In some cases, the same causal forces apply to a set of time series. To obtain reliable codings of causal forces, one should rely on knowledgeable domain experts. If possible, obtain codings from more than one expert, especially when there is uncertainty about the codings. The forces tend to endure over time, so that an expert needs only to revise them when changes are expected.

The specification of causal forces can have a strong influence on forecasts. For example, we made rule-based forecasts for U.S. congressional mailings. Based on our knowledge about congress, it seemed obvious that the causal forces were growth. But if the rules had changed in 1976 so that members of congress would have had to pay for their mailings, the resulting decay forces would generate a much different forecast (Exhibit 3).

Exhibit 3
Rule-based Forecasts Can Vary Substantially Given Different Causal Force Specifications

(U.S. Congressional mailings)
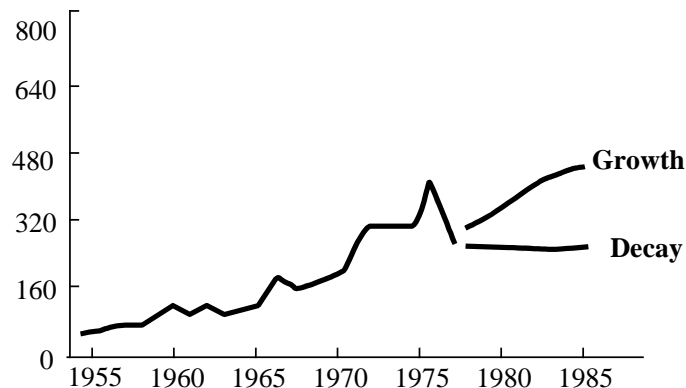
We tested the benefits of identifying causal forces by examining 126 annual series from the M-Competition data (Makridakis et al. 1982). The codings were those used in Collopy and Armstrong (1992a). Despite our limited domain knowledge, we achieved an acceptable interrater reliability for the codings of causal forces with 81 percent agreement. The result is better than chance as there were five possible categories (or four, if "supporting" were excluded). Excluding the 22 series that we classed as "unknown," tests on the remaining 104 series showed that the use of causal forces improved accuracy (Armstrong and Collopy 1993). For one-year-ahead forecasts, the use of causal forces reduced the MdAPE (median absolute percentage error) by more than four percent. The improvements were, as expected, larger for the six-year-ahead forecasts, where they were 12 percent.

- **Use domain knowledge to adjust observations for events.**

RBF calls for adjusting data based on domain knowledge. For example, a one-time event such as a strike may have affected sales in the past. Or other periodic events such as sales promotions may have affected sales. The forecaster should make adjustments to remove the temporal effects of such events. One approach would be to replace the value for the period with the average of the observations preceding and following it. Similarly, one could replace an unusual last observation with the estimated value from a linear regression.

- **Decompose time series to avoid conflicting causal forces.**

Sometimes the causal forces in a time series do not act in a single direction. In such cases, one may be able to improve forecasts by decomposing the series. Consider the prediction of automobile deaths on UK highways. The original time series (top of Exhibit 4) incorporates growth (number of miles driven) and decay (safety improvements) forces. These forces are isolated in the components (traffic volume and death per traffic volume) as shown in the lower panel.

Exhibit 4
Time Series with Mixed Causal Forces Can be Decomposed

**UK Highway Deaths: Causal Forces are Mixed**
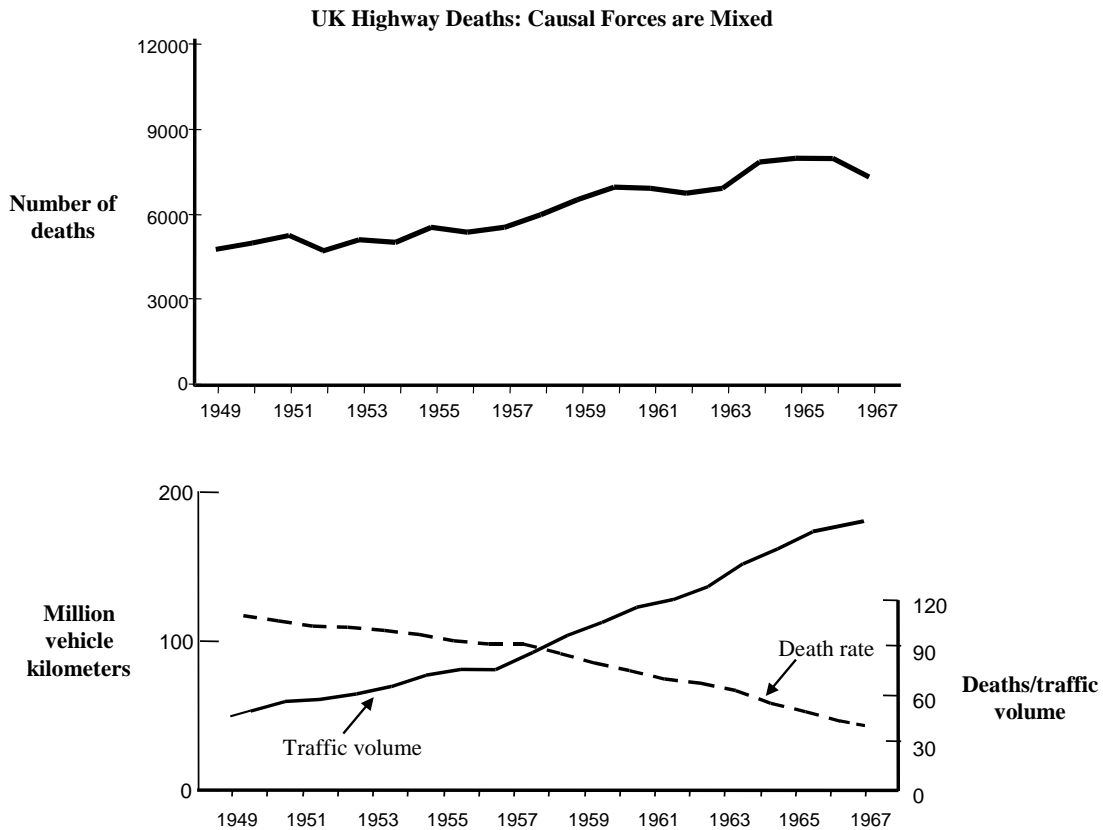
- **Decompose only when each component can be forecasted as well as the target series.**

A decomposition that is multiplicative can be risky because recombining the forecast will cause errors in the components to be multiplied. For example, a 20 percent increase in error for one component would increase the overall error by 20 percent, all other things remaining equal. Furthermore, when errors in components are in the same direction, combining them can be explosive; an increase of 20 percent in the error for each of two component forecasts translates into a 44 percent increase in the error of the target variable=s forecast (1.2 times 1.2). Collopy and Armstrong (1996), therefore, tested the "forecastability" of the components and the target series before determining whether to forecast by decomposition.

By selecting an appropriate method for extrapolating each component and then synthesizing forecasts, one would expect improved accuracy. Exhibit 5 shows the improved accuracy of a recomposed for a ten-year-ahead forecast of highway deaths based on forecasts made in 1967.

Exhibit 5
Recomposed Time Series Forecasts Were More Accurate
than Global Forecasts When Compared with Actual Deaths on UK Highways
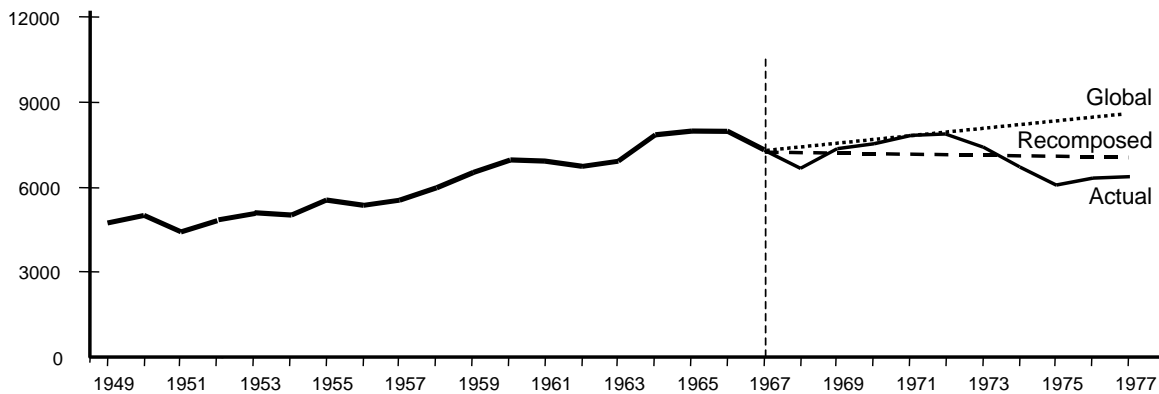
To test the value of decomposition, we used successive updating from this series on deaths, and also series on auto injuries and accidents in the U.K. For 150 forecasts for which tests showed decomposition to be recommended, the procedure reduced forecast error by 16 percent in comparison with the global forecast. For the 75 forecasts for which they did not, decomposition increased the error by 42 percent. MacGregor (2001), in decomposing judgments, also found it to be useful but risky.

- **Use features of the historical time series to define conditions.**

Forecasters have traditionally characterized time series by type of data, trend, length of series, seasonality, and uncertainty. We have developed features to describe each of these conditions. We expanded the conditions and developed nine features to represent instabilities and one to represent the level. Including domain knowledge and historical data, we suggest using 28 features for characterizing time series (Exhibit 6), yielding well over a billion possible types of series.

Exhibit 6
Time Series Features

Domain Knowledge	Historical Data	
$ Causal forces	**Types of Data**	**Uncertainty**
Growth	$ Only positive values possible	$ Coefficient of variation about
Decay	$ Bounded (e.g., percentages,	trend > 0.2
Supporting	asymptotes)	$ Basic and recent trends differ
Opposing	$ Missing observations	**Instability**
Regressing	**Level**	$ Irrelevant early data
Unknown	$ Biased	$ Suspicious pattern
$ Functional form	**Trend**	$ Unstable recent trend
Multiplicative	$ Direction of basic trend	$ Outliers present
Addictive	$ Direction of recent trend	$ Recent run not long
$ Cycles expected	$ Significant basic trend (t > 2)	$ Near a previous extreme
$ Forecast horizon	**Length of Series**	$ Changing basic trend
$ Subject to events	$ Number of observations	$ Level discontinuities
$ Start-up series	$ Time interval (e.g., annual)	$ Last observation unusual
$ Related to other series	**Seasonality**	
	$ Seasonality present	

Some of these features can be determined analytically. For instance, one can obtain the direction of the basic trend by fitting a linear regression to the historical data, and the direction of the recent trend by fitting Holt=s exponential smoothing model to the same data. Other such features include significance of the trend, variation about the trend, whether the current value of the series is near a historical limit, or whether the series level or trend has changed suddenly.

The 'Then' Part of the Rules

RBF uses the features described in Exhibit 6 to determine how to create a combined forecast by weighting forecasts from the various methods. We illustrate this for the relationship between causal forces and historical trends. Series in which causal forces agree with statistical trends are called reinforcing series. Series in which causal forces and statistical trends differ are called contrary series. Series in which causal forces are unknown are classified by whether their long- and short-term trends are consistent (Exhibit 7).

Exhibit 7
Forecasting of Trends Depends on the Directions of Causal Forces and Historical Trends
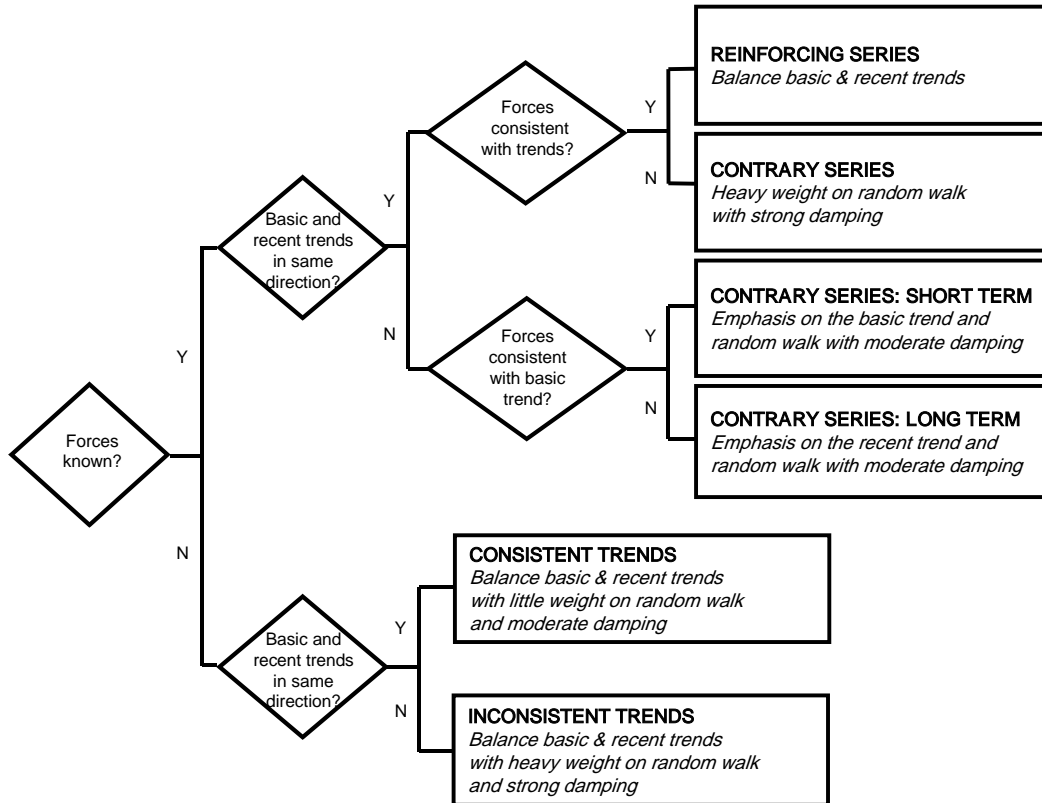
- **Use full trend extrapolation for reinforcing series.**

When the basic and recent trends and the causal forces are all in the same direction, RBF places a heavy weight on the trend estimates. In these cases – common in situations such as sales forecasting – traditional extrapolation models do well because the direction of the trend happens to agree with the direction implied by the causal forces.

- **Place little weight on the trends in contrary series.**

Contrary series arise when expectations conflict with long-term or short-term trends. When the causal forces conflict with both the basic and recent trends, RBF places little weight on the statistical trend estimates.

If only the recent trend is opposite to expectations, RBF relies primarily on the basic trend. Conversely, if only the basic trend conflicts with expectations, more weight is given to the recent trend. Causal forces can be used with any extrapolation method by applying the principle:

- **If expected trends (from causal forces) are contrary to historical trends, do not use that trend.**

Here is a simple version of the contrary series principle: ignore trends if they conflict with causal forces. The story of the late Julian Simon's 1980 challenge illustrates the power of the contrary-series principle. Simon says, "Pick any natural resource and any future date. I'll bet the [real] price will not rise." He based this on long-term trends in the prices of natural resources, and argued that major changes in the long-term causal factors are rare. Paul Ehrlich, an ecologist from Stanford University, accepted the challenge; he selected ten years and five metals (copper, chromium, nickel, tin, and tungsten) whose prices had been rising in recent years. The causal forces for the prices of resources were decay because of improvements in procedures for prospecting and extraction, reductions in energy and transportation costs, development of substitutes, and improvements in free trade. The exhaustion of resources might lead to increased prices; however, this seldom has a strong effect because of human ingenuity in developing

better ways to find resources and to recycle. For example, Ascher (1978, pp. 139-141) reported that estimates of the ultimate available petroleum reserves *increased* from the late 1940s to the mid-1970s. This has continued; The *Wall Street Journal* (April 16, 1999, p. 1) reported that between 1976 and 1996, estimated global oil reserves grew by 72 percent.

Ehrlich assumed that recent price trends would continue, an assumption of supporting trends. To implement this assumption, we used Holt=s exponential smoothing to extrapolate trends for the five metals. The resulting forecasts showed sharply rising prices. Exhibit 8 shows the forecasts for chromium. RBF is especially useful when domain knowledge indicates that recent trends may not persist. Although RBF initially forecasted an increase in prices (because it allows for the possibility that short-term trends might continue), over the 10-year horizon the forecast became dominated by the long-term trend, which is downward and consistent with the causal forces. We found the same pattern for each of the five metals. (We prepared the forecasts using a version of RBF described by Adya et al. 2000 and data from *Metals Week.*). Simon won the bet; his directional predictions were correct for all five metals (Tierney 1990).

Exhibit 8
Use of Appropriate Causal Forces can Improve Accuracy
(example using chromium prices)
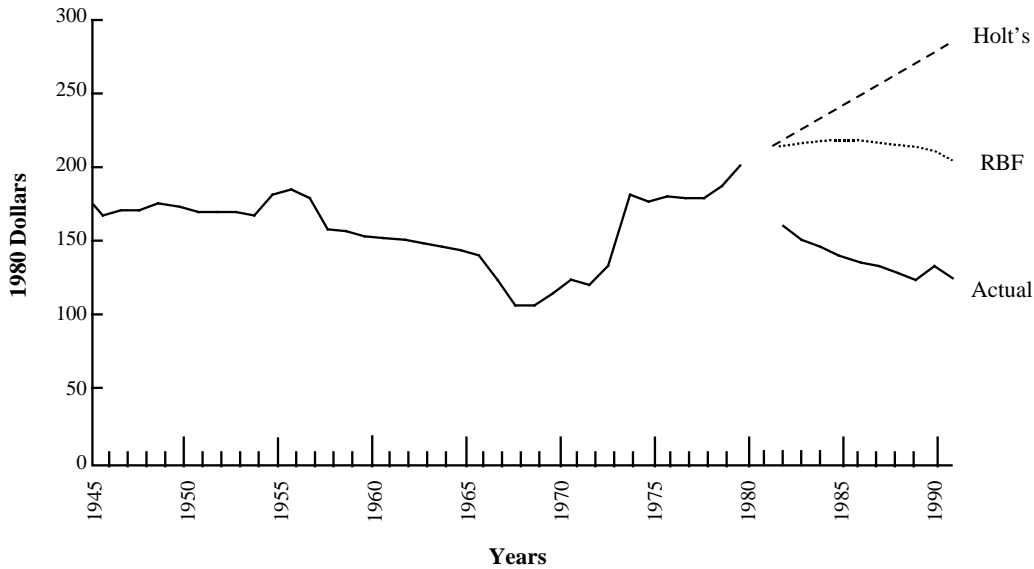
Traditional extrapolation procedures make a false assumption that the causal forces will always support the trend. We hypothesized that for contrary series, traditional extrapolation would not perform well. To test this, we examined the accuracy of Holt=s exponential smoothing for 126 annual series from the M-competition. As expected, Holt=s was accurate for the 84 reinforcing series where the assumption of "supporting forces" caused no problem. Its Geometric Mean of the Cumulative Relative Absolute Error (GMCumRAE) over a six-year horizon was 0.52; that is, the error was about half that for the random walk. It even did well for the 22 series in which causal forces were unidentified, where its GMCumRAE was 0.67. But for the 20 contrary series, the corresponding error was 1.10, which is less accurate than the random walk.

To further evaluate the value of contrary series in selecting an extrapolation method, Armstrong and Collopy (1993) examined forecasts from four data sets. These included ten annual series on epidemics in China, quarterly personnel data for nine US Navy pay grades, annual unit product sales for 51 consumer products (based on a 50 percent probability sample from Schnaars 1984), and 26 economic and demographic series from a variety of published sources that were collected at the Weatherhead School at Case Western Reserve University. In each data set, the trend extrapolation of contrary series produced forecast errors that were substantially larger than those from the random walk, all differences being significant at $p < .05$ (Exhibit 9).

Exhibit 9
Random Walk is More Accurate than Holt=s for Contrary Series
(ex ante Median Absolute Percentage Errors)

| | Short term (one-ahead) | | | | Long term (6-ahead for annual, 18 for quarterly) | | | |
| | | MdAPE | | | | MdAPE | | |
Data Set	Number of Forecasts	Holt's (H)	Random Walk (R)	Ratio (H/R)	Number of Forecasts	Holt's (H)	Random Walk (R)	Ratio (H/R)
Annual								
Chinese epidemics	121	27.7	25.0	1.11	95	133.0	71.8	1.81
Unit product sales	60	9.3	7.6	1.22	32	28.1	18.0	1.56
Weatherhead	74	5.5	4.4	1.25	61	35.6	16.2	2.20
Quarterly								
Navy personnel	688	4.0	3.2	1.25	535	20.1	12.8	1.57

A selection rule stating that the trend should be ignored for contrary series would have improved accuracy for all four data sets, and for both short and long horizons. Also, as expected, the rule produced greater accuracy improvements for the long-term forecasts. In this study, we used a no-trend model, but other options might be superior. For example, based on evidence summarized by Webby, O'Connor and Lawrence (2001), judgmental extrapolations might be useful.

- **Use a conservative trend estimate if the basic and recent trends are inconsistent.**

When causal forces are unknown, one can compare the directions of the recent and basic trends. If they are the same, trend extrapolations should be aggressive. When they differ, a conservative trend extrapolation should be used by increasing the weight on the random walk. This is in line with the general principle to be conservative in the face of uncertainty.

To test this, we assumed the causal forces were unknown for the 126 annual time series that were used in Armstrong and Collopy (1992). Forecasts were made for a six-year horizon from a single origin. For the 109 series where the long-term trend forecast (from a regression against time) was in the same direction as the short-term trend forecast (from exponential smoothing), Holt's exponential smoothing reduced the Cumulative RAE by 40% in comparison with the random walk. But for the 17 series where the forecasts had inconsistent trends, Holt's reduced the Cumulative RAE by only 8%.

- **Tailor extrapolation weights to the time interval of the series.**

Short-term models, particularly Holt=s, are accurate for short-period (monthly) series having low variability and low uncertainty. For short-period (e.g., monthly) data, RBF shifts weight from the long-trend model (linear regression) to short-term models.

- **To estimate levels for the short-term model, heavily weight the latest observation, particularly in the presence of discontinuities.**

The levels from the component methods are combined using weights determined by the rules. Different weights are used for short-term and long-term models. The short-term model relies more on the random walk (latest observation) to provide initial estimates of levels. When discontinuities have occurred, the rules put more weight on the latest observation. When instabilities have occurred and where uncertainty is high, weights should be distributed more equally across time.

For the long-model level estimation, RBF places more emphasis on regression. This incorporates more of the historical data than does the random walk or exponential smoothing, thus providing a more reliable estimate of the level.

- **Adjust the estimate of the level in the direction implied by the causal forces.**

Once an initial estimate of the current level is made, it can be adjusted in the direction implied by the causal forces. A mechanical adjustment can also be made to the level, based on how well the rule-base forecasted the observation at the origin (t_0), given data through the preceding period (t_{-1}).

EVIDENCE ON THE VALUE OF RBF

We tested RBF using annual series from the M-competition (Collopy and Armstrong 1992a). Annual data are ideal because causal forces play a stronger role when the data covers a longer period. In addition, the series were well behaved in that they had strong trends, modest uncertainty, and few instabilities. RBF proved more accurate than alternative methods, including the random walk, the typical method (of nine methods) used in the M-competition, and most important, equal-weights combining (Exhibit 10). The improvement that RBF showed over the other methods depended upon the forecast horizon. For example, for six-year ahead forecasts, the MdAPE for RBF was on average 57 percent that of equal-weights combining, and both were much more accurate than the random walk.

Exhibit 10
RBF has Lower Ex-Ante Forecast Errors than Other Extrapolation Procedures

| Extrapolation Procedure	Median Absolute Percentage Errors (number of series)							
	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	5.5	4.3	4.9	4.8	23.3	18.0	18.0	19.0
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: V1 was the first validation sample, V2 was the second, and V3 the third.

Because RBF is designed to tailor the extrapolation to the features, we expected it to be more accurate than traditional extrapolation methods in many situations. Our major findings were that RBF improves accuracy when:

$ long-interval (e.g., annual) data are used,
$ good domain knowledge is available,
$ causal forces can be clearly identified,
$ domain knowledge conflicts with the historical trend,
$ long-range forecasts are needed,
$ significant trends exist,
$ uncertainty is modest to low, and
$ instability is modest to low.

An example that meets most of the above conditions is General Motors sales after a strike in 1970. (These data come from series 5 from the M-competition data, Makridakis et al. 1983.) Using domain knowledge to specify a multiplicative trend and to adjust the last observation produced a more accurate forecast than was provided by other methods, such as Holt's, in the M-competition.

Exhibit 11
Rule-Based Forecasts for General Motors Sales
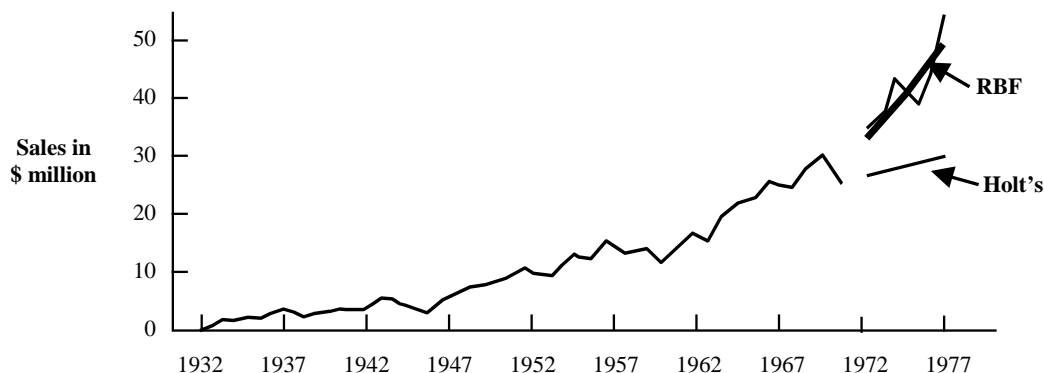
RBF is more accurate than other methods for long-range forecasts because causal factors and trends are more important in the long run. The MdRAE for one-year-ahead forecasts for the 36 series in V3 was 0.63, while it was 0.48 for six-year-ahead forecasts. (The RAE is less than 1.0 when the method is more accurate than the random walk). Collopy and Armstrong (1992a) obtained similar results when the procedure was extended to other sets of time series.

To examine the accuracy of RBF under a variety of conditions, we again used relative absolute errors (RAEs). We cumulated RAEs over the forecast horizon and reported their medians (MdCumRAE). Although the number of series was small, the results indicate that RBF is especially accurate when one has domain knowledge for trended, stable time series (Exhibit 12).

Exhibit 12
Rule-Based Forecasting is More Accurate than Equal Weights under a Variety of Conditions
(Annual Series from the M-Competition)

| | | Median CumRAE | |
| | | | |
Conditions	Number of Series	Equal-Weights Combining	Rule-Based Forecasting
Domain Knowledge			
Causal forces unknown	22	0.83	0.87
Causal forces known	104	0.71	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*
Trend (Basic)			
Not statistically significant	22	1.06	0.96
Statistically significant ($t > 2$)	104	0.67	0.54*
Uncertainty			
Coefficient of variation > 0.2	26	0.87	0.89
Coefficient of variation < 0.2	100	0.68	0.55*

* Rule-based forecasting is more accurate than equal-weights at $p < 0.01$.

In the absence of domain knowledge, RBF showed no advantage over equal weights (first row of Exhibit 12). This was surprising to us as we are convinced that rules based on prior research help in weighting extrapolations. However, the sample contained only 22 series, so there is uncertainty about this finding.

The M3-competition allowed for a better test of the value of rules based only on forecasting expertise Makridakis and Hibon (2000). This competition called for forecasts for 3,003 series that were a mix of annual, quarterly, monthly, and other short-period data. Due to the absence of much in the way of domain knowledge, forecasts were prepared under the assumption that no domain knowledge was available. This removed what we believe to be one of RBF's primary advantages. Automatic identification procedures were used to identify six features that were previously identified using judgment. We simplified the rule-base by removing one method from the four that were used in the original implementation. Although this resulted in some loss in accuracy, it reduced the number of rules in the rule-base from 99 to 64. Results from the M3-competition series were consistent with those observed with other independent samples. RBF was substantially more accurate than the random walk and equal-weights combining for annual series. For shorter-period data, RBF did not improve on combining in a situation that lacked domain knowledge for such data. This may be due to our failure to properly calibrate the rules (Adya et al. 2000a).

Given the limited domain knowledge employed in the studies to date, we believe that the forecast validity of RBF had been underestimated. We did find that the benefits RBF provided were greater when the raters agreed on causal forces. Of the 104 series of the M-competition for which we were able to identify causal forces, the two coders agreed that causal forces were 'clear' for 79 series, but not for 25. When we analyzed the 79 clear series as if the causal forces were unknown, there was reduction in accuracy, especially for the long-range forecasts. For example, without causal forces, the MdAPE for six-year-ahead forecasts would have been increased by about 17 percent (Armstrong & Collopy 1993, Table 3). In contrast, for the 25 series for which we did not agree, the error was slightly *higher* when using causal forces. Based on this small sample, specifying causal forces seems valuable only when one has reasonably good information about them.

Extensions of RBF have been conducted by Vokurka, Flores and Pearce (1996) and Adya et al. (2000). Despite using different base methods, and in the case of Vokurka et al., a somewhat different procedure, both produced results similar to Collopy and Armstrong's (1992a). Adya et al. (2000) introduced rules for identifying historical features automatically. This version of RBF was validated on an independent sample of 458 series and on 122 of the 126 series reported in Collopy and Armstrong (1992a). Results from both samples were consistent with those reported in Collopy and Armstrong (1992a).

When the conditions are not favorable to RBF, using it is not risky. Typically, RBF reverts to one of the component models that is expected to do well. For example, the random walk tends to perform well when trends are not expected. When the series contains much uncertainty or many instabilities, equal-weights combining performs well.

LIMITATIONS

Academicians and practitioners regard accuracy as the most important criterion for selecting a forecasting method. However, they also put emphasis on understandability, computational requirements, data requirements, cost, and ease of use (Armstrong 2001c). Researchers developing RBF have addressed primarily accuracy.

Cost poses a potential problem. While RBF is less expensive to develop and use than econometric methods, it is more expensive than standard extrapolation methods. This issue is being addressed by studies using automatic feature identification. They have reduced costs to code features, but this area needs more work. For example, attempts to automatically identify the functional form have produced ambiguous results.

RBF has been developed, refined, and validated primarily with annual data. Yet, quarterly, monthly, weekly, and daily data are used for many applications of extrapolation. Rules need to be calibrated for such data. Domain knowledge might be less important for these data.

IMPLICATIONS FOR PRACTITIONERS

RBF benefits from managers= knowledge. If this knowledge is good, one can expect improved forecast accuracy.

RBF represents the accumulation of expertise and empirical conclusions from prior research. To the extent that the rules are updated with the latest findings, those using RBF benefit. They would have to take action to avoid them.

The identification of contrary series should be of particular interest to practitioners. Forecasts of these series are likely to be outside the prediction intervals in the direction of the causal forces (Armstrong and Collopy 2000). RBF can highlight these cases for additional attention when making forecasts and estimating prediction intervals.

No commercial RBF program is available. However, one can use some of the rules with existing programs. For example, the contrary series rule for trends can be applied to any trend extrapolation method. Software vendors can easily implement this and other rules, such as damping the trend when a series has discontinuities or inconsistent trends.

IMPLICATIONS FOR RESEARCHERS

Researchers have tried to specify extrapolation procedures purely on the basis of statistical considerations for many decades, and the results have been disappointing. Little knowledge in RBF came from such research as that stimulated by Box and Jenkins (1970). Instead, it has come from empirical studies that compared reasonable competing methods.

While the method of competing hypotheses is useful, progress has been limited because researchers have generally failed to adequately define the conditions associated with the forecasts. For RBF to benefit, researchers must describe the conditions in their studies. We suggest that information be provided on the 28 features of time-series shown in Exhibit 6. This information should help to identify which methods are most effective under which conditions. For example, some methods might be particularly effective when discontinuities are encountered, while other techniques might be useful when facing high uncertainty.

To aid further research, Collopy and Armstrong (1992a) provided full disclosure of the original rules. Some errors in these rules have been corrected, leading to small improvements in the accuracy of RBF when tested on the original validation series (Adya 2000). The corrected rule-base is provided at the Forecasting Principles website (hops.wharton.upenn.edu/forecast). A PC version of the RBF code in C++ is available to researchers for replications and extensions. An object-oriented extension of rule-based forecasting was published by Assimakopoulos and Konida (1992).We hope further research will focus on ways to improve RBF, including refining rules, developing new rules, testing in new situations, identifying features, and understanding the conditions affecting the accuracy of RBF.

Refining Rules

Refine rules based on data analysis and expert judgment: Researchers should state their expectations about the effects of a new or revised rule before testing it. Starting with certain expectations (e.g., what weights to use), one should change the rules based upon evidence from data analyses. For example, when an expert indicated that the trend component should be reduced if the recent trend showed irregular movements, he proposed that this reduction should be by a "fair amount." We decided that a 25 percent reduction per year would serve as a good initial representation of a fair amount. We then conducted searches starting with data on 18 time series, examining weights of 15 percent, 20 percent, 30 percent, and 35 percent. When the accuracy of the resulting forecasts improved, we moved in the direction that produced improvements. However, we were averse to setting weights that deviated greatly from the initial estimates. When weights moved in unexpected directions, we removed the changes and reexamined our reasoning. It is important that the rules be consistent with the beliefs of experts and prior evidence.

Because rules interact with one another, it was necessary to return to earlier calibration runs after making changes in other rules. We tested whether the revised weights remained near optimal. For example, did modifying the rules that dealt with uncertainty affect the previously calibrated rules for significant trends?

We experimented with computer search routines in order to identify optimum weights. To make the search manageable, we simplified the system by removing Brown=s exponential smoothing. Searches using Hooke and Jeeves (1961) were computationally demanding. A more serious problem, however, was that while these searches modified the basic rules in a manner that improved in-sample accuracy, they often produced results that did not seem sensible or that harmed accuracy when tested on a validation sample. Our conclusion is that one must rely heavily on forecasting expertise and domain knowledge.

Use large samples of forecasts to refine rules: Our initial work involved variations on 21 features. This implies an immense number of possible types of situations, or cells. We initially used only 126 annual series and evaluated forecasts for six forecast horizons, which provided 756 forecasts. The sample size was based on the time available, deadlines, and our computer capabilities at the time. We recommend the use of much larger samples of time series in future work. We have done some initial work with quarterly and monthly series, and here, especially, one needs many forecasts (Adya et al. 2000a). We believe that studies with fewer than 100 monthly time series will produce unreliable results. Calibration of rules should ideally be done using many thousands of series.

The number of forecasts can be expanded by forecasting for a number of horizons. For the annual M-competition data, forecasts were made for six years into the future. Obviously, these are not independent observations, but they can provide evidence about the relative accuracy of forecasting methods and the consistency of the findings.

The sample sizes of ex ante forecasts can also be increased by using successive updating (rolling horizons or moving origin). Again, although the observations are not independent of one another, they provide some evidence on the relative accuracy of forecasting methods.

Use out-of-sample error measures that control for scale and difficulty: Following the principles described in Armstrong (2001a), error measures should be unit-free; it should not matter if series are measured as percentages, temperatures, dollars, or billions of dollars. Otherwise, comparisons might be dominated by a few series having large values, and conclusions could vary if arbitrary changes were made in the scaling of one or more series. Errors should be stated as percentages, except for series with values near zero.

Some series are more difficult to forecast than others because they fluctuate wildly. Relative error measures help to control for this. We suggest the RAE (Relative Absolute Error) because it compares errors from a given forecasting method with errors produced by a naive (no change) forecast and it is easy to understand (Armstrong and Collopy 1992).

Use sensitive error measures for calibration: Error measures for rule calibration should be sensitive, to insure that the effects of changes are evident. Median error measures and ordinal measures, such as percent-better-than-random-walk, are not sensitive, because once a method is more accurate than the benchmark method on a series, improved accuracy will have no effect. The MAPE and the Geometric Mean of the RAE (GMRAE) are sensitive measures. We recommend using the GMRAE because the MAPE is biased in favor of low forecasts (Armstrong 2001a). We also recommend trimming high and low values for GMRAEs so that the conclusions are not dominated by a few extreme errors.

Conduct tests of rules on "wind tunnel" data: Early in the development of RBF, one of the authors returned from a visit to the Wright Brother=s museum impressed that the brothers succeeded in flying largely because of their use of a wind tunnel. The wind tunnel enabled them to observe the effects of design changes under known conditions. We found this strategy to be useful when validating refinements of RBF. Changes in the effects of new rules are easier to identify and mistakes are more readily apparent when tested on the same data set. Fortunately, researchers in forecasting have shared data freely. In particular, data from the M-competitions can be obtained from the Forecasting Principles website.

Developing New Rules

RBF provides a summary of knowledge on extrapolation. As such, it can help researchers identify gaps in knowledge, and thus lead to ideas for research on new rules. What do these rules or procedures contribute to accuracy compared with those used previously? The importance of new rules can be tested on benchmark data,. By controlling both the model and the data, it is possible to examine the contribution of a new rule.

RBF might draw upon domain knowledge to develop rules for pooling time-series data (Duncan, Gorr and Szczypula, 2001, discuss pooling time series). For example, while various products for a firm might have different sales levels, they might be subject to similar causal forces and thus have similar trends. A trend factor based on the average of a pool of products might be more accurate than one limited to the product of interest.

Testing in New Situations

It would be useful to test RBF on other types of data, particularly data that differ substantially from the M-competition data. One might also test different ways of using RBF. For example, Tashman and Kruk (1996) used rules from RBF to help in the selection of extrapolation methods.

One of the major advantages in using RBF is the ability to incorporate domain knowledge. However, the level of domain expertise incorporated in studies to date has been low. It would be useful to test RBF in situations in which there is much domain knowledge.

Little research has been done on short series. The shorter the series, the greater the uncertainty. In an (unpublished) analysis using annual sales data provided by Schnaars (1984), we found it difficult to improve accuracy over that provided by a simple rule: "if there are fewer than eight observations, use the random walk."

Although some of the data sets included bounded data, little consideration was given to procedures for handling such data. Likewise, little work has been done using data that can take on negative values or dealing with missing observations.

Automatic Feature Identification

Expert judgments sometimes lack reliability and this can harm forecast accuracy (Stewart 2001). Experts might disagree in identifying time-series features. One obvious improvement is to use domain experts who know much about the area. Another way is to use more coders (we typically used two coders), but this adds expense. Automatic identification of time-series features can improve reliability and avoid biases. Assuming that the automatic procedures are valid, improved reliability would increase accuracy. From a practical standpoint, the primary advantage of automatic identification is cost reduction because it automates time-consuming judgments. From a research point of view, automatic identification can aid replication and extension efforts.

In our original work, we had rules for automatic identification of 11 features of the time series. Another 17 features were determined judgmentally. Some of these 17, such as start up series and seasonality, were not relevant, and other features, such as "subject to events" were ignored due to our lack of domain knowledge. Judgmental coding constrained our ability to handle large volumes of time series because it took two coders five to eight minutes each to code each series. This included time for the coders to discuss discrepancies. Seven features required domain knowledge. Adya et al. (2000b) added rules for automatically identifying five of the remaining features: outliers, unusual last observations, level discontinuities, basic trend changes, and unstable recent trends. To automate the identification of the five features, we developed heuristics that use simple statistical measures, such as first differences and regression estimates. For instance, the identification of a change in historical trend is done by comparing slopes in various parts of the historical data. If there are large differences in the slopes, a change in the basic trend is assumed to have occurred. Outliers, level discontinuities, and unusual observations had to first be adjusted so that the regression fit could approximate the basic trends as closely as possible.

Features that require domain knowledge cannot be automatically identified. But there is little need to automate domain judgments because they take little time to identify, and once set, they tend to remain constant over time. Also, causal forces often apply to all the series in a group.

Although the selection of a functional form depends upon domain knowledge, we examined automatic rules for identifying the best form. This included such things as using an additive form if the series can take on negative values. The automatic codings often conflicted with the subjective coding. Nevertheless, the resulting rules proved to be about as effective as the original subjective-coding procedure.

To test automatic procedures for detecting features, we relied on the wind-tunnel data. We compared judgmental and automated feature identification to determine whether automating the process reduced forecast accuracy. The sample consisted of 122 of the 126 M-competition series used in Collopy and Armstrong (1992a). Results reported in Adya et al. (2000b) indicated that there was only a minor loss in accuracy as a consequence of automatic feature identification.

Vokurka, Flores and Pearce (1996) also used rules to automate identifying features of RBF. They identified and adjusted irrelevant early data and outliers, identified functional forms, and allowed for user interventions at several points in the forecasting process. They used different base methods than those used by Collopy and Armstrong (1992a): Simple exponential smoothing, Gardner=s damped trend exponential smoothing, and classical decomposition. Their results were similar, however, showing improved accuracy in comparison with the random walk and equal weights benchmarks. User intervention did not produce any substantial improvements.

Identifying Situations in which RBF is More Useful than Other Methods

Given a sufficient budget, data on the causal variables, and situations involving large changes, we would expect econometric models to provide more accurate forecasts than RBF. Under what conditions does RBF provide forecasts that are as accurate as those from econometric models? No researchers have conducted studies to find out.

CONCLUSIONS

For almost three decades, the dominant paradigm for forecasting research has been statistical modeling. These efforts have done little to incorporate domain knowledge into extrapolations.

By drawing upon the cumulative findings from research on forecasting and incorporating structured domain knowledge, RBF improves forecasting accuracy. Under some conditions, it has reduced errors by more than a third in comparison to equal weights. Much remains to be done in further elaborating and testing rules especially for data other than annual. Meanwhile, results have shown enough consistency that we can recommend some rules as practical guides for extrapolation.

Fairly elementary domain knowledge can be used to improve extrapolations. Patterns in the data can be helpful as well, particularly signs of significant departures from assumptions made by extrapolation methods, as occurs for contrary series. Integrating knowledge from prior research has produced a system that is more accurate than widely used extrapolation methods, including simple exponential smoothing and combining forecasts. Because these findings resulted from the application of theory and empirical testing, we are optimistic that continued refinement of this research program will produce further improvements.

REFERENCES

Adya, M. (2000), "Corrections to rule-based forecasting: Results of a replication," *International Journal of Forecasting*, 16, 125-127.

Adya, M., J. S. Armstrong, F. Collopy & M. Kennedy (2000a), "An application of rule-based forecasting to a situation lacking domain knowledge," *International Journal of Forecasting* (forthcoming).

Adya, M., J. S. Armstrong, F. Collopy & M. Kennedy (2000b), "Automatic identification of time series features for rule-based forecasting," Working paper available from Monica Adya (adya@umbc.edu).

Armstrong, J. S. (2001a), "Evaluating forecasting methods," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S. (2001b), "Combining forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S. (2001c), "Selecting forecasting methods," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S. (1985), *Long Range Forecasting: From Crystal Ball to Computer.* New York, NY: John Wiley. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J.S. & F. Collopy (2000), "Identification of asymmetric prediction intervals through causal forces," *Journal of Forecasting* (forthcoming).

Armstrong, J. S. & F. Collopy (1998), "Integration of statistical methods and judgment for time series forecasting: Principles from empirical research," in G. Wright & P. Goodwin (eds.), *Forecasting with Judgment*, John Wiley, pp. 269-293. Full text at hops/wharton.upenn.edu/forecast.

Armstrong, J. S. & F. Collopy (1993), "Causal forces: Structuring knowledge for time series extrapolation," *Journal of Forecasting*, 12, 103-115. Full text at hops/wharton.upenn.edu/forecast.

Armstrong, J. S. & F. Collopy (1992), "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, 8, 69-80. Full text at hops/wharton.upenn.edu/forecast.

Armstrong, J. S. & E. Lusk (1983), "The accuracy of alternative extrapolation models: Analysis of a forecasting competition through open peer review," *Journal of Forecasting*, 2, 259-311. Full text at hops/wharton.upenn.edu/forecast.

Ascher, W. (1978), *Forecasting: An Appraisal for Policy-makers and Planners.* Baltimore: Johns Hopkins University Press.

Assimakopoulos, V. & A. Konida (1992), "An object-oriented approach to forecasting," *International Journal of Forecasting,* 8, 175-185.

Batchelor R. & P. Dua (1995), "Forecaster diversity and the benefits of combining forecasts," *Management Science*, 41, 68-75.

Box, G. E. & G. M. Jenkins (1970), *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

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

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

Chatfield, C. (1988), "Apples, oranges and mean square error," *International Journal of Forecasting*, 4, 515-518.

Collopy, F. & J. S. Armstrong (1996), "Decomposition by causal forces: An application to highway deaths," Full text at hops.wharton.upenn.edu/forecast.

Collopy, F. & J. S. Armstrong (1992a), "Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations," *Management Science,* 38, 1394-1414. Full text at hops/wharton.upenn.edu/forecast.

Collopy, F. & J. S. Armstrong (1992b), "Expert opinions about extrapolation and the mystery of the overlooked discontinuities," *International Journal of Forecasting,* 8, 575-582. Full text at hops/wharton.upenn.edu/forecast.

Collopy, F. & J. S. Armstrong (1989), "Toward computer aided forecasting systems: Gathering, coding and validating the knowledge," in G. R. Widmeyer (ed.), *DSS 90 Transactions. TIMS College on Information Systems.* Providence, R.I, pp. 103-119. Full text at hops/wharton.upenn.edu/forecast.

Duncan, G., W. Gorr & J. Szczypula (2001), "Forecasting analogous time series," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Fildes, R., M. Hibon, S. Makridakis & N. Meade (1998), "Generalizing about univariate forecasting methods: Further empirical evidence," *International Journal of Forecasting,* 14, 339-358. (Commentaries follow on pp. 359-366

Gardner, E. S., Jr. (1999), "Rule-based forecasting vs. damped-trend exponential smoothing," *Management Science*, 45, 1169-1176.

Gardner, E. S., Jr. (1985), "Exponential smoothing: The state of the art," (with commentary), *Journal of Forecasting*, 4, 1-38.

Harvey, N. (2001), "Improving judgmental forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Hooke, R. & T. A. Jeeves (1961), "Direct search solution of numerical and statistical problems," *Journal of the ACM*, 8 (April), 212-229.

MacGregor, D. G. (2001), "Decomposition for judgmental forecasting and estimation," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Makridakis, S., A. Anderson, R. Carbone, R. Fildes, M. Hibon, J. Newton, E. Parzen & R. Winkler (1982), "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1, 111-153.

Makridakis, S. & M. Hibon (2000), "The M3-competition: Results, conclusions and implications," *International Journal of Forecasting* (forthcoming).

Makridakis, S. & M. Hibon (1979), "Accuracy of forecasting: An empirical investigation," *Journal of the Royal Statistical Society*, Series A, 142, Part 2, 97-145.

Schnaars, S. (1984), "Situational factors affecting forecast accuracy," *Journal of Marketing Research*, 21, 290-297.

Stewart, T. (2001), "Improving reliability of judgmental forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Tashman, L. J. & J. M. Kruk (1996), "The use of protocols to select exponential smoothing procedures: A reconsideration of forecasting competitions," *International Journal of Forecasting*, 12 , 235-268.

Tierney, J. (1990), "Betting the planet," *New York Times Magazine*, December 2, pp. 52.

Vokurka, R. J., B. E. Flores & S. L. Pearce (1996), "Automatic feature identification and graphical support in rule-based forecasting: A comparison," *International Journal of Forecasting,* 12, 495-512.

Webby, R., M. O'Connor & M. Lawrence (2001), "Judgmental time series forecasting using domain knowledge," in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.