February 1993

Causal Forces: Structuring Knowledge for Time Series Extrapolation

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

Fred Collopy
Case Western Reserve University

Postprint version. Published in *Journal of Forecasting*, Volume 12, Issue 2, February 1993, pages 103-115.
Publisher URL: http://dx.doi.org/10.1002/for.3980120205

Causal Forces: Structuring Knowledge for Time Series Extrapolation

Abstract
This paper examines a strategy for structuring one type of domain knowledge for use in extrapolation. It does so by representing information about causality and using this domain knowledge to select and combine forecasts. We use five categories to express causal impacts upon trends: growth, decay, supporting, opposing, and regressing. An identification of causal forces aided in the determination of weights for combining extrapolation forecasts. These weights improved average *ex ante* forecast accuracy when tested on 104 annual economic and demographic time series. Gains in accuracy were greatest when (1) the causal forces were clearly specified and (2) stronger causal effects were expected, as in longer-range forecasts. One rule suggested by this analysis was: "Do not extrapolate trends if they are contrary to causal forces." We tested this rule by comparing forecasts from a method that implicitly assumes supporting trends (Holt's exponential smoothing) with forecasts from the random walk. Use of the rule improved accuracy for 20 series where the trends were contrary; the MdAPE (Median Absolute Percentage Error) was 18% less for the random walk on 20 one-year ahead forecasts and 40% less for 20 six-year-ahead forecasts. We then applied the rule to four other data sets. Here, the MdAPE for the random walk forecasts was 17% less than Holt's error for 943 short-range forecasts and 43% less for 723 long-range forecasts. Our study suggests that the causal assumptions implicit in traditional extrapolation methods are inappropriate for many applications.

Keywords
causal forces, combining contrary trends, damped trends, exponential smoothing, judgment, rule-based forecasting, selecting methods

Comments
Postprint version. Published in *Journal of Forecasting*, Volume 12, Issue 2, February 1993, pages 103-115.
Publisher URL: http://dx.doi.org/10.1002/for.3980120205

Causal Forces: Structuring Knowledge for Time-series Extrapolation

J. Scott Armstrong
University of Pennsylvania, Philadelphia, PA, U.S.A.

Fred Collopy
Case Western Reserve University, Cleveland, OH, U.S.A.

***Reprinted by permission: Journal of Forecasting*, 12 (1993), 103-115**

ABSTRACT

This paper examines a strategy for structuring one type of domain knowledge for use in extrapolation. It does so by representing information about causality and using this domain knowledge to select and combine forecasts. We use five categories to express causal impacts upon trends: growth, decay, supporting, opposing, and regressing. An identification of causal forces aided in the determination of weights for combining extrapolation forecasts. These weights improved average *ex ante* forecast accuracy when tested on 104 annual economic and demographic time series. Gains in accuracy were greatest when (1) the causal forces were clearly specified and (2) stronger causal effects were expected, as in longer- range forecasts. One rule suggested by this analysis was: "Do not extrapolate trends if they are contrary to causal forces." We tested this rule by comparing forecasts from a method that implicitly assumes supporting trends (Holt's exponential smoothing) with forecasts from the random walk. Use of the rule improved accuracy for 20 series where the trends were contrary; the MdAPE (Median Absolute Percentage Error) was 18% less for the random walk on 20 one-year ahead forecasts and 40% less for 20 six-year-ahead forecasts. We then applied the rule to four other data sets. Here, the MdAPE for the random walk forecasts was 17% less than Holt's error for 943 short-range forecasts and 43% less for 723 long-range forecasts. Our study suggests that the causal assumptions implicit in traditional extrapolation methods are inappropriate for many applications.

KEY WORDS Causal forces Combining Contrary trends Damped trends Exponential smoothing Judgment Rule-based forecasting Selecting methods

Forecasters have often stated the need to use their knowledge about a time series (domain knowledge) when making extrapolations. Researchers suggest that this would be useful. For example, in discussing quantitative extrapolation, Newbold (1983) argued for "approaches where the forecasters *think* (about the subject matter area, the data, and anything relevant)."

In their review of the literature, Bunn and Wright (1991) concluded that judgmental knowledge and quantitative methods should be integrated.

Domain knowledge wis found to be useful in Edmundson (1990). He had forecasters make judgmental forecasts when they were aided by a decision-support system that contained quantitative methods. These subjects made more accurate forecasts than those made by either judgment or quantitative methods alone.

In this paper we consider how to structure certain aspects of domain knowledge for use in extrapolation. Lopes (1983) suggested that 'extrapolative methods would be strengthened by making provisions for causal or explanatory information to be used when such [information] is available.' Following this idea, we describe an approach that uses judgments about causality as inputs to simple quantitative extrapolation methods. This approach offers potential cost savings in comparison with purely judgmental extrapolation procedures. Because it uses more information, it should be more accurate than simple extrapolations. Further, because it uses judgments in a structured way, this approach may help to more effectively integrate judgment into quantitative extrapolation methods.

We first define our use of the term 'causal forces.' Next, we describe how causal forces can be used to combine forecasts. Finally, we show how the findings can be used to guide the selection of quantitative extrapolation methods.

DESCRIPTION OF CAUSAL FORCES

Changes in the trend in a time series arise due to a variety of underlying factors. For example, a strong marketing program might involve a reduction in prices, wider distribution, and an increase in advertising. We use the term 'causal forces' to represent the cumulative directional effects of the factors that influence the trends in a time series.

We have classified causal forces by the way that they relate to historical trends. Some forces cause movement in a given direction (up or down) irrespective of current trends; we call these 'growth' and 'decay' forces. Forces that support the current direction of a trend are called 'supporting'. Forces that act against trends are labeled as 'opposing.' Finally, 'regressing' is used to classify forces that cause a series to move toward a mean. We briefly describe each type of force here, along with some actions that might be taken when extrapolating time series.

Growth forces support upward trends but not downward trends in the time series. An example would be product sales where the market is favorable and the firm is actively trying to increase sales. When the trend is rising in a growth series, it would be extrapolated without adjustments. However, if the trend in a growth series were downward, it would be damped. That is, the estimated trend factor would be reduced in magnitude.

Decay forces are the inverse of growth forces. For example, in forecasting real unit production costs for the manufacture of a technical product, costs would be expected to go down because of improved efficiency. An upward trend would not be taken very seriously and would not have as strong an impact on the forecast as would a trend of similar magnitude in the downward direction.

In series with *supporting* forces, the assumption is that the direction of the current trend correctly reflects the impact of the most important causal forces, whether it is upward or downward. Real estate prices might be an example. When the trend is up, people often conclude that the investment will be a good one. When the trend is down, they may assume that the price for these properties will continue to fall.

Series with *opposing* forces resist further movements in the current direction. For instance, when personnel loss rates are increasing, organizations often improve pay and working conditions to attract workers. Conversely, when loss rates are decreasing, pay and conditions may be allowed to become less attractive.

For some series, the forecaster has good reason to assume that there is a tendency to move toward a certain value, such as a grand mean for a type of series (e.g. the batting average for all baseball players) or its own long-term mean (e.g. a given player's lifetime average). For these *regressing* series, forecasts of the long-term trend should reflect this tendency to move toward that mean.

Table I summarizes this categorization of forces and presents examples of series with these forces. The examples represent forces that are typical, but forces may vary for actual series. A sales series, for instance, might be classified as decay instead of growth if the product were competing against a clearly superior product. Also, causal forces acting on a series might change over time; for example, a firm may decide to withdraw marketing support, thus changing forces from growth to decay. Epidemics further illustrate how causal forces can change over time. A rising trend for a communicable disease would, in the early stages, mean an increase in the number of disease carriers, so that the primary forces might be growth. Later, causal forces might oppose upward trends because of preventive measures such as immunization programs. Judgments based on domain knowledge play a central role in identifying causal forces.

Table I. Relationship of causal forces to trends

Type of causal force	Causal force direction when: Trend is up	Causal force direction when: Trend is down	Examples
Growth	Up	Up	Sales (units)
Decay	Down	Down	Epidemics (late)
Supporting	Up	Down	Real estate prices (?)
Opposing	Down	Up	Personnel loss rates
Regressing	(Toward a mean)		Demographic (% male births)

Unknown	?	?	Exchange rates

To illustrate the importance of causal forces, consider the series in Figure l. (This is series 96 from the M-Competition, Makridakis *et al.,* 1982). As an inventory series, it is subject to opposing forces (inventories should not be allowed to get too high because of carrying costs, nor too low because service will suffer). Holt's exponential smoothing, which implicitly assumes supporting forces, extrapolates the recent trend upward. The random walk, an opposing forces model, does not extrapolate a trend. The assumption about causal forces was important in forecasting this series. Over the 6-year period shown, the *ex ante* forecast error from Holt's forecast was about 1.7 times greater

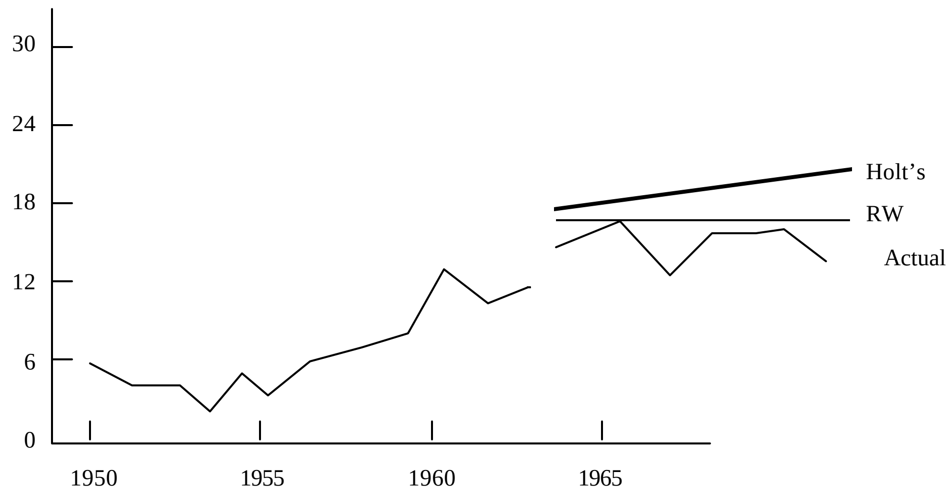


Figure 1. Change in stocks: France. An illustrative series showing Holt's, random walk, and actual hold-out data.

than that for the random walk.

USE OF CAUSAL FORCES IN COMBINING FORECASTS

We believe that causal force information can play a useful role in the combination of forecasts. This section describes how we applied the idea of causal forces to aid in establishing weights for combining forecasts.

We had used causal forces in Collopy and Armstrong (1992), where 99 rules were used to establish weights to combine forecasts. This approach uses rules to estimate the levels and trends for short- and long-range models. The rules select and weight forecasts from a set of four extrapolation methods: random walk, linear trend (from a regression against time), Holt's linear exponential smoothing, and Brown's linear exponential smoothing. We selected these methods because they are based on different assumptions about trends. The random walk is based on the assumption that trends are spurious; the linear regression on the assumption that the long-term historical trend will continue; and Holt's and Brown's methods on the assumption that the recent trend will continue. The methods were also selected because they are well known, inexpensive, and easy to interpret.

The causal force specified by the analyst influences the weights assigned to the trend estimates from each of the four extrapolation methods. These weights depend upon the relationship between causal forces and trends in the historical data. When causal forces and the long and short trends are all in agreement, the system is considered to be consistent and the model extrapolates the trends using a balanced mix of trends. If either the short- or long- term trend is contrary to the causal forces, trend extrapolation is done cautiously by putting more weight on the random walk and by damping the trend as the horizon increases. Caution is also used for trend forecasts when the causal forces are unidentified. This is done by increasing the weight on the random walk and by more damping. Figure 2 illustrates the reasoning behind these differential weighting heuristics.

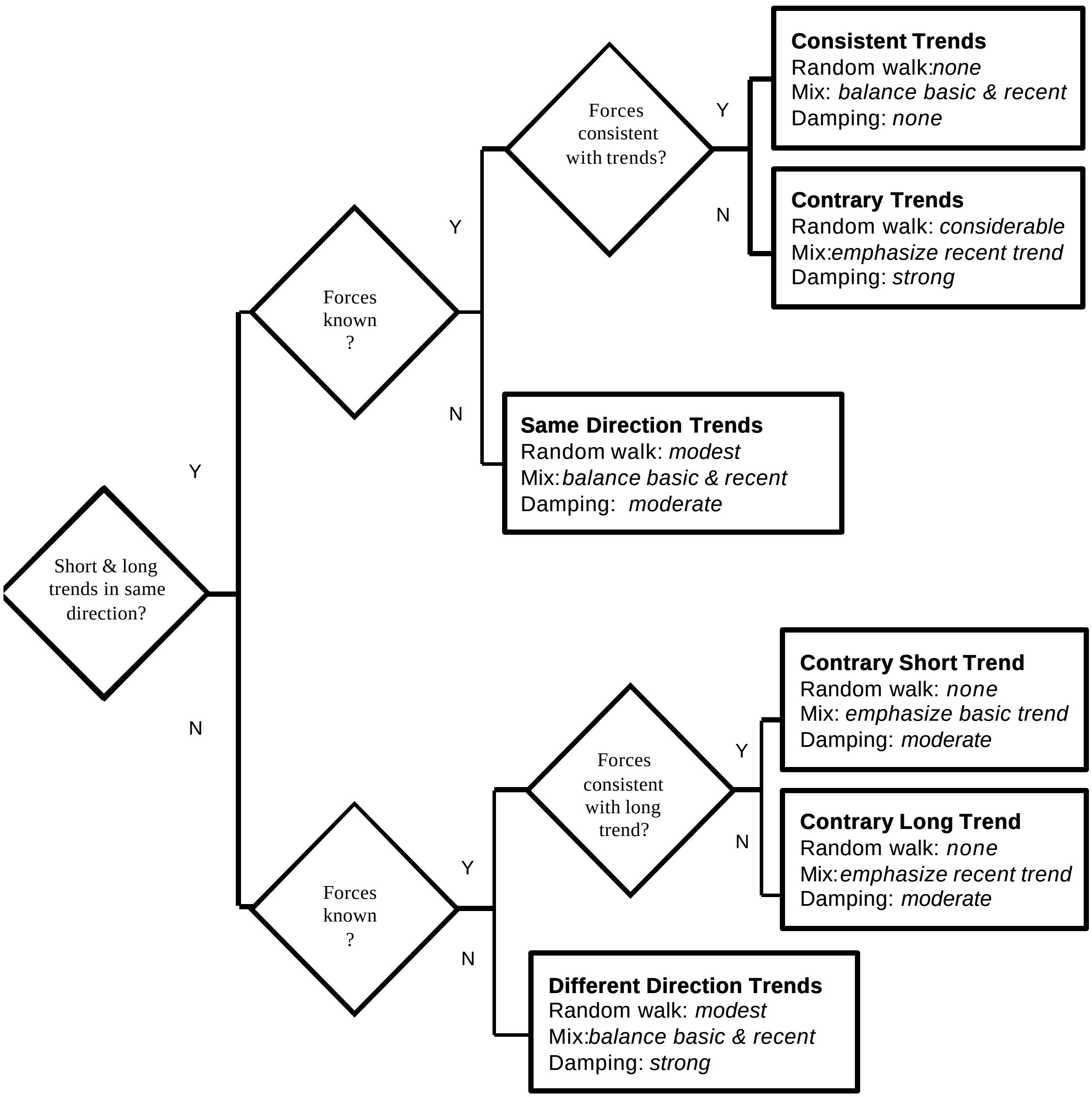
The estimate of level is also influenced by causal forces, and the same four models are used to estimate levels. When the most recent movement of the data is consistent with causal forces, it receives more weight than if it is contrary to the forces. An example of a rule for estimating level is, 'If the causal force direction is the same as the recent trend direction and the coefficient of determination (R^2) is greater than 0.9, then add 0.1 to alpha for the short model [where alpha is the weight on the latest observation] ' (Collopy and Armstrong, 1992, rule 15). For smooth series, this rule makes the forecast more responsive to a recent change when it is consistent with causal forces.

Information about causal forces affects the estimates of the levels and of the trends for both short- and long-range models. In all, 23 of the 99 rules in the rule base are dependent on information about causal forces.

To examine the extent to which the use of causal forces improved *ex ante* forecast accuracy in rule-based forecasting, we compared rule-based forecasts with and without causal force information. In operational terms, we specified causal forces as 'unidentified' in the latter run of the rule-based forecasting model. We used the sample of 126 annual economic and demographic time scries that was described in Collopy and Armstrong (1992). This was a stratified random sample selected from the M-Competition data. 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, particularly instability.

Each author independently coded the forces for each of the 126 series. In addition, one author specified forces at two different times, roughly three months apart. The identification was considered unclear if (1) the authors did not agree, (2) the coding was not reliable across time, or (3) either of the authors reported that he was uncertain. We were unable to identify causal forces for 22 of the 126 series, so we excluded them from further analysis. Of the remaining 104 series, causal forces were clear for 79 and unclear for the other 25.[1] Table II summarizes the features of the clear and unclear series. While we do not know the extent to which these series might be representative of those faced by decision makers, we suspect that growth forces are common and that supporting forces are rare. Considering that the traditional methods such as Holt's make the assumption that causal forces are supporting, it is striking that *none* of the 79 clearly identified series fell into this category.



---

[1]The coding for each series is available from the first author.

We expected that when causal forces could be clearly identified, they would aid in the selection and weighting of extrapolation methods. We were unsure about whether specifying forces would improve accuracy if the forces were unclear. Furthermore, because the causal forces have a greater impact over longer time periods, we expected use of this information to have a larger effect on long-range forecasts than on short-range ones.

The primary evaluation criterion was the Median Absolute Percentage Error (MdAPE) because it has been used in earlier comparative studies on these data and it is not affected by outliers. We also examined the Relative Absolute Error (RAE) because of its superior sensitivity and reliability (Armstrong and Collopy, 1992). The RAE, which is similar to Theil's *U2*, is the ratio of the absolute forecast error from a proposed model to the absolute forecast error from the random walk. Because the RAE is a ratio, we used the geometric mean of the RAEs (GMRAE) to summarize across series. We Winsorized the RAE at 0.01 and at 10 because very large or very small errors have a strong impact on geometric means. (Winsorizing occurred about 2Vo of the time.) Finally, we calculated the cumulative RAE (CumRAE), which sums the model's errors across each year in the forecast horizon, and divides by the corresponding sum for the random walk forecast errors.

Six years were held out to allow for a single h-year-ahead forecast for each series for h = 1 – 6. Only one starting point was used for each series and the dates of the starting points varied across series. We examined *ex ante* forecasts from one-year-ahead to six-years-ahead. This procedure is the same as that used in Makridakis *et al.* (1982).

As hypothesized, using causal forces reduced errors in comparison with a rule-based forecast that did not identify causal forces. For one-year-ahead forecasts on the 104 series, the use of causal forces reduced the MdAPE by more than 4% and the GMRAE by more than two 2%. For the six-year-ahead annual forecasts, specifying causal forces reduced the MdAPE by 12% and the GMRAE by 15%. Averaging the MdAPE across the one-ahead and six-ahead horizons, the reduction was about 8%. The use of causal forces also reduced the cumulative MdAPE over the six-year horizon by about 8%. The bottom two rows of Table III, labeled 'all identified series,' summarize the results.

The above improvements in accuracy were achieved using a model that had performed well in comparison with equally weighted combined forecasts for these 104 series. The rule-based forecasting model without causal forces was more accurate than equal-weights combining for both error measures and for all horizons. For one-ahead MdAPEs for these series, Combining A (an equally weighted forecast from the M-competition) averaged 3.47, whereas rule-based forecasting (Collopy and Armstrong, 1992) without causal forces averaged 2.98. For six-ahead forecasts, the MdAPEs were 19.6 for Combining A and 14.26 for the rule-based forecasting model with unspecified causal forces.

Table II. Clarity and features of the series (percentage of series in each category)

	Forces	
Feature	Clear[a]	Unclear[b]
Trend		
Direction of basic trend up	96	72
Direction of recent trend up	96	26
Significant basic trend (t >2)	96	64
Uncertainty		
Coeff. of variation > 0.2	18	24
Basic and recent trends differ	6	20
Instability		
Recent run not long	47	28
Near a previous extreme	47	52
Irrelevant early data	33	24
Changing basic trend	30	44
Suspicious pattern	8	20
Unstable recent trend	14	28
Level discontinuities	15	8
Last observation unusual	5	0
Domain knowledge		
Causal forces		
Growth	92	60
Decay	1	4
Supporting	0	8
Opposing	4	20
Regressing	3	8
Functional form		
Multiplicative	91	80
Additive	9	20
Cycles present	30	24
Adjusted observations	1	4

[a] 79 series [b] 25 series

As we expected, the specification of causal forces was valuable only when forces were clear. For the 79 series with clear forces, the specification of forces reduced the MdAPE of the six- ahead forecast by 14% and the GMRAE by 22% (top of Table III, labeled 'Clear' forces). When causal forces were not clear, forecasts were slightly less accurate if forces were speci6ed than if they were not. This difference, which was not statistically significant, is shown in the middle of Table III, labeled 'Unclear' forces.

Forces are likely to be clear in most series of interest to decision makers. When they are not clear, we suggest that they not be used in forecasting.

Table III. *Ex ante* errors by causal force specification

Causal forces	Number of series	One-ahead		Six-ahead		Cumulative	
		GMRAE	MdAPE	GMRAE	MdAPE	GMRAE	MdAPE
Clear	79						
Unspecified		0.466	2.88	0.419	14.34	0.453	9.12
Specified		0.443	2.85	0.328	12.28	0.409	8.19
Unclear	25						
Unspecified		0.692	3.52	0.708	13.03	0.934	10.15
Specified		0.740	3.36	0.814	13.93	0.976	10.04
All identified series	104						
Unspecified		0.514	2.98	0.479	14.26	0.543	9.34
Specified		0.501	2.85	0.109	12.53	0.504	8.62

USE OF CAUSAL FORCES FOR SELECTING AMONG METHODS

In addition to their use in combining, we were interested in the application of causal forces to the selection of extrapolation methods, so we examined whether causal forces can be used to decide when it is appropriate to use traditional extrapolation methods. A single selection rule was used. This stated that forecast accuracy would be improved by ignoring trend estimates when they conflict with causal forces. We ignored trend by using a random walk for such series.

As discussed above, traditional exponential smoothing methods assume supporting forces. That is, they assume that the causal forces that have caused a trend will continue to have the same effect in the future. Such methods may perform poorly when the trend is contrary to the causal forces. Contrary trends occur for a declining trend with growth forces, an increasing trend with decay forces, any series with opposing forces, or a trend away from the target mean in a regressing series.

None of the 79 clearly identi5ed series in the M-competition was classified as supporting. Furthermore, we fnd it difficult to suggest many types of series that might have supporting forces. What might happen if supporting trends are assumed inappropriately? In many cases there may be no damage because the force is not be contrary to the trend. To examine the impact of using traditional methods where they are expected to be inappropriate, we classified the 126 series into three groupings. The 'consistent' group included those series where the recent trend (estimated by Holt's additive exponential smoothing) was in the same direction as the causal forces; the 'unidentified' group contained those series where the direction of causal forces was not known; and the 'contrary' group consisted of those series where the recent trend conflicted with the causal forces. For each type of series we compared forecasts from Holts' exponential smoothing with a random walk forecast. We expected that Holt's would be more accurate than the random walk for consistent series because its assumptions are more appropriate. For contrary series, we expected that Holt's would not do as well as the random walk. For series with unidentified causal forces, we expected that trends would contain some information, but that the forecast accuracy of Holt's and the random walk would be comparable.

The results, shown in Table IV, generally supported our expectations. For consistent series, Holt's error was about half that from the random walk. Holt's forecasts were more accurate for the unidentified series, but the gain in accuracy was not as large as that for the consistent series. For the 20 contrary series, Holt's was less accurate than the random walk; its CumRAE was 1.10. As we were particularly interested in the 20 contrary series, we extended the analysis to include other error measures. These also showed the random walk to be more accurate for these series. For one-ahead forecasts, the MdAPE was 8.3 for Holt's and 6.8 for the random walk. For six-ahead forecasts, the MdAPE was 19.9 for Holt's and 11.9 for the random walk. In other words, using the contrary trends selection rule reduced the short-range forecast errors by about 18% and the long-range errors by about 40% for the contrary series.

Table IV. Errors for Holt's method by consistency of forces and trends

Relationship of forces and trends	Number of series	Geometric mean of the RAE		
		One-ahead	Six-ahead	Cumulative
Consistent	84	0.60	0.45	0.52
Unidentified	22	0.82	0.64	0.67
Contrary	20	1.30	0.90	1.10

To assess whether the rule about contrary series can be generalized, we examined four additional data sets:[2]

(1) We received quarterly data on the number of US Navy personnel in each of nine pay grades from the Navy Personnel Research and Development Center.

[2]These four data sets are available from the second author.

(2) We obtained annual data on ten disease epidemics in China from the Chinese Academy of Preventive Medicine.

(3) Steven Schnaars provided us with data on 103 annual unit product sales (described in Schnaars, 1984). We took a 50% systematically selected subsample and were able to identify causal forces for about half of the series.

(4) We collected annual economic and demographic data for 26 series from a variety of public sources. These data are called Weatherhead I. Four of these were eliminated because the causal forces were unidentified.

For each data set, we compared the accuracy of Holt's exponential smoothing with the random walk for all forecasts where there were contrary trends. For this analysis, contrary trends were defined as those where the direction expected due to causal forces was contrary to the trend estimated by Holt's exponential smoothing.

Given that the Navy personnel data represent inventory levels, we assumed that the causal forces would often be opposing. Although we expected that other causal forces, such as growth, would occur for these inventories, we had no information about these forces; thus, we defined all these forecasts as having contrary trends. For the Chinese epidemics data, we used inflection points to determine when the series became contrary. Operationally, this was defined as times when the change in the trend (second difference) was negative, which occurred for 74% of the starting points. Coding of causal forces for the unit product sales was done by the authors using the same procedure as for the M-competition and focusing only on clearly identified series; contrary trends occurred for 241% of the 255 starting points used for the unit product sales data. The same procedure was used for the Weatherhead data and contrary trends were encountered for 18% of the 417 starting points.

We encountered no series with supporting causal forces. Holt's was inappropriate, then, and it should have difficulty whenever contrary trends are encountered. In fact, contrary trends occurred for 62% of the forecasts in these four data sets.

Table V. *Ex ante* forecast errors of Holt's versus random walk for contrary series (median absolute percentage errors – MdAPEs)

Data set	One-ahead				*h*-ahead			
	No. of forecasts	Holt's	Random walk	Ratio	No. of forecasts	Holt's	Random walk	Ratio
Chinese epidemics	121	27.7	25.0[a]	1.11	95	133.0	71.8[a]	1.81
Unit product sales	60	9.3	7.6[a]	1.22	32	28.1	18.0[a]	1.56
Navy personnel	688	4.0	3.2[a]	1.25	535	20.1	12.8[a]	1.57
Weatherhead I	74	5.5	4.4[a]	1.25	61	35.6	16.2[a]	2.20

Other than the epidemics data, we made no changes in the causal forces over the forecast horizon. Of course, one would expect added benefits if the specification of the causal forces were updated each time period.

For the initial calibration of Holt's forecasts, we used 24 periods for the quarterly data and 12 for the annual data, except for the unit product sales data, where six periods were used because the series were substantially shorter. We then made 18-ahead forecasts for the quarterly data and six-ahead for the annual data. Successive updating was performed after each period. This procedure used the first *t* periods of the series to fit the models and made forecasts for *h* periods into the future. Then period t+ 1 was added into the Qt data, and forecasts were made up to *h* periods. The procedure continued until the calibration data included all but the last observation.

The contrary series selection rule improved accuracy for each forecast horizon in each of the four data sets and Table V shows these results. In each of the eight comparisons, the random walk was more accurate than Holt's at $p < 0.001$, based on the one-tailed Wilcoxon signed ranks test. The differences were substantial.

DISCUSSION

In the four new data sets, we encountered 943 contrary trends among the 1524 starting points. For the one-ahead forecasts, the unweighed geometric mean of the ratios of Holt's to the random walk error was 1.21. This represents an error reduction of about 17%. This result is of the same order as the 18% reduction for the M-competition data. By including the M- Competition results with the other four data sets, the error reduction would be about 17% for one-ahead forecasts.

For the longer-range forecasts, we found contrary trends for 723 of forecasts in the four data sets. The geometric mean ratio of Holt's error to the random walk error was 1.77. This represents an error reduction in excess of 43%. Combining this with the error reduction of over 40% in the M-competition, the average error reduction across the five data sets was about 43% for h-ahead forecasts.

The unweighed geometric mean of the short- and long-range forecast error was about 31%. Because the contrary rule was invoked for 62% of these forecasts, the overall error reduction was about (0.31)(0.62) = 0.192. That is, by using this rule instead of always forecasting with Holt's for all series in these four data sets, the average error reduction would be 19%.

Considering the three sets with economic data, unit product sales, Weatherhead, and the M-competition, the error reductions were 18%, 20%, and 18%, respectively. The unweighed geometric mean was 19%. Contrary trends were encountered for 24% of the unit product sales data, 18% of the Weatherhead data, and 19% of the M-competition data, or an average of about 20%. Thus, the overall gain for short-range economic forecasts would be about (0.18)(0.2) = 0.036, or less than 4%.

The error reductions for the six-year ahead economic forecasts were substantially larger. Given the error reductions of 36% for the unit product sales, 55% for Weatherhead, and 41% for the M-competition, the average error reduction was 44%. Using the estimate of 20'Vo of the forecasts as contrary, the average error reduction for the economic data was (0.44)(0.2) = 0.088, or almost 9%.

To put the above results into perspective, consider that two decades of research on combining forecasts has produced more than 200 studies (Clemen, 1989). A rough estimate is that a combination of forecasts from two extrapolation methods can reduce errors by about 7% (Armstrong, 1986). The results reported above are of the same order.

Why not use causal variables directly? First, the forecaster may not know what variables are involved. For example, using the causal forces approach, one can say that the net impact of the principal factors acting on a product category is 'growth' while lacking information about the factors contributing to this growth such as improvements in product performance, wider availability in retail stores, more advertising, lower prices, shortage of a substitute, or an increase in the number of potential buyers.

Cost represents another problem associated with the direct use of causal variables. Even when causal variables can be identified, it is necessary to collect relevant data on them (sometimes not available), to estimate their effects, and to forecast changes in them (which sometimes can be more difficult than forecasting the variable of interest). For applications involving forecasts of many thousands of items (e.g. inventory control), cost might be an important factor. Extrapolation with causal forces offers a low cost alternative to causal models. The simple selection rule that we presented could be easily implemented in existing forecasting systems.

Excepting the data on Chinese epidemics, we assumed that causal forces would be constant over time for each series. In many situations, though, forces on a series might change. In such cases, experts' domain knowledge might be useful to specify forces for each starting point.

LIMITATIONS

Our study has several limitations. First, except for the Navy personnel inventories, we used only annual data. We suspect that causal forces will be less useful when trends are subject to the increased uncertainty or to the transient factors often associated with shorter time periods. Similarly, causal forces might be less valuable for unstable annual series or if trends are weak relative to random variation.

We coded the causal forces using only general knowledge. With better domain knowledge, the forces might be more accurately identified. This would be expected to strengthen the value of causal forces.

The study specified a single causal force for each series. This procedure is problematic for series subject to different types of forces. Referring to our example of French inventories (Figure 1), one might specify this series as subject to opposing *and* growth forces (the latter because of inflation and economic development).

FURTHER RESEARCH

Additional studies are needed to determine the extent to which our results can be generalized. An increase in the number of series with different causal forces along with a rise in the overall number of series would allow for additional calibration for the combination of forecasts. It is also important to learn whether the results can be generalized to short-interval data such as monthly time series, where the causal forces are likely to have less influence.

It is not clear how to deal with series subject to multiple causal forces. One approach might be to combine forecasts. Another is to dampen the trends. We believe that the most promising approach, however, would be to decompose the series so that each element is influenced by a single type of force. For French inventories (Figure 1) growth could be estimated by using GNP in current prices, while inventory would be the current series divided by GNP in current prices. However, even with this procedure, causal forces will not always be clearly identifiable.

We tested only one of many possible selection rules. Further research might aid in the development of additional guidelines for the selection of extrapolation methods.

This study examined only one aspect of domain knowledge. Other aspects also appear to be useful, judging from the results in Collopy and Armstrong (1992). We expect that the use of actual domain experts will show additional ways to structure domain knowledge. Such studies would be especially useful if conducted in actual forecasting situations because this will enable the use of richer domain knowledge.

CONCLUSIONS

This study tested an approach to structuring knowledge for use in quantitative extrapolation. The results showed that causal forces might play an important role in weighting extrapolation forecasts to produce a combined forecast. For one-year-ahead forecasts of 104 economic and demographic series from the M-Competition, specifying causal forces in a rule-based forecasting model reduced the MdAPE by about 4%. The corresponding error reduction for six-year-ahead forecasts was about 12%.

Traditional extrapolation methods make an implicit assumption that causal forces support the trends. This assumption is often violated. We tested a selection rule stating that trends should not be extrapolated when they are contrary to causal force directions. Thus, we compared Holt’s exponential smoothing (often applied to such series) against the random walk, which does not rely on the assumption of supporting forces. Application of this rule improved accuracy in a test on M-competition data; the *ex ante* MdAPE for contrary series forecasts was reduced by 18% for one-year-ahead forecasts and by 40% for six-year-ahead forecasts. Further tests were conducted on contrary series from four data sets. Here, the rule reduced the one-ahead forecast error by about 17%, and the *h*-ahead forecast by about 43%. It is interesting to note how well the results from the M-competition held up when extended to four new, and substantially different, data sets.

Further research needs to be done on the use of causal forces for short-interval data, on procedures to aid the identification of causal forces (such as the use of domain experts), and on the treatment of series that fall into more than one of the causal force categories. We expect that this research on causal forces will improve the accuracy of forecasts by leading to improvements in rules both for the selection of forecasting methods and for weighting combined forecasts. In this study, initial efforts in the use of causal forces for combining and for selection produced significant error reductions, especially for longer-range forecasts. Substantial gains were achieved with a simple selection rule: If the extrapolation method produces a trend that is contrary to expectations, ignore it. This rule can be easily implemented with existing extrapolation methods with the result that traditional methods will be appropriate in more situations.

ACKNOWLEDGMENTS

The US Navy Personnel R&D Center in San Diego supported part of this research. Dennis A. Ahlburg, Paul Andreassen, William R. Ferrell, Robert Fildes, Keith Ord, Janet Sniezek, and J. Thomas Yokum and anonymous referees made valuable comments on earlier drafts. Jennifer L. Armstrong and Phan Lam provided editorial assistance.

REFERENCES

Armstrong, J. S, "Research on forecasting: A quarter-century review, 1960 – 1984," *Interfaces,* 16, issue 1 (1986), 89-109.

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

Brown, R. G., *Statistical Forecasting and Inventory Control,* New York: McGraw-Hill, 1959.

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

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

Collopy, F. and Armstrong, J. S., "Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations," *Management Science,* (forthcoming, 1992).

Edmundson, R. H., "Decomposition: A strategy for judgmental forecasting," *Journal of Forecasting,* 9 (1990), 305-314.

Holt, C. C., Modigliani, F., Muth, J. F. and Simon, H. A., *Planning Production Inventories and Workforce,* Englewood Cliffs, NJ: Prentice Hall, 1960.

Lopes, L. L., "Pattern, Pattern – who's got the pattern?," *Journal of Forecasting,* 2 (1983), 269 – 272.

Makridakis, S. *et al.*, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting,* 1 (1982), 111 – 153.

Newbold, P., “The competition to end all competitions,” *Journal of Forecasting*, 2 (1983), 276 – 279.

Schnaars, S. P., “Situational factors affecting forecast accuracy,” *Journal of Marketing Research*, 21 (1984), 290 – 297.

Authors' biographies:

J. Scott Armstrong was an industrial engineer for Eastman Kodak, systems analyst for Xerox, and marketing researcher for Polaroid. He then received a PhD from MIT and has been a Professor at the Wharton School since 1968. He taught at the Stockholm School of Economics, University of Hawaii, IMEDE (Switzerland), University of Canterbury (New Zealand), University of Capetown, Chulalongkorn University (Bangkok), University of Auckland, and IDEA (Buenos Aires). He has published papers on survey research, educational methods, social responsibility, scientific methodology, marketing, applied statistics, and strategic planning.

Fred Cullopy is a Professor of Management Information and Decision Systems at the Weatherhead School of Management, Case Western Reserve University. He received his PhD in Decision Sciences from the Wharton School of the University of Pennsylvania. He has worked extensively in the area of information systems for over twenty years, and has conducted research related to the organizational impacts of information systems, decision support systems, and user-interface design. He was the founder and president of Conceptual Instruments, a software development firm, and has lectured and consulted extensively in the areas of forecasting and information systems. In addition to rule-based forecasting, his research interests include decision support, objective setting in organizations, and organizational impacts of information systems.

Authors' addresses:
J. Scott Armstrong, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA.
Fred Collopy, The Weatherhead School of Management, Case Western Reserve University, Cleveland, OH 44106, USA.