# Decomposition by Causal Forces: A Procedure for Forecasting Complex Time Series

J. Scott Armstrong[*] Fred Collopy[†]
J. Thomas Yokum[‡]

[*]University of Pennsylvania, armstrong@wharton.upenn.edu
[†]Case Western Reserve University
[‡]Angelo State University

Postprint version. Published in *International Journal of Forecasting*, Volume 21, Issue 1, January 2005, pages 25-36.
Publisher URL: http://dx.doi.org/10.1016/j.ijforecast.2004.05.001

# Decomposition by Causal Forces:
# A Procedure for Forecasting Complex Time Series

J. Scott Armstrong
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
Armstrong@wharton.upenn.edu

Fred Collopy
The Weatherhead School of Management
Case Western Reserve University
Cleveland, OH 44106
collopy@cwru.edu

J. Thomas Yokum
Virgil J. Powell Professor of American Economic Principles
Angelo State University
San Angelo, TX 76909
Tyokum@angelo.edu

January 30, 2004

**Decomposition by Causal Forces:**
**A Procedure for Forecasting Complex Time Series**


## Abstract

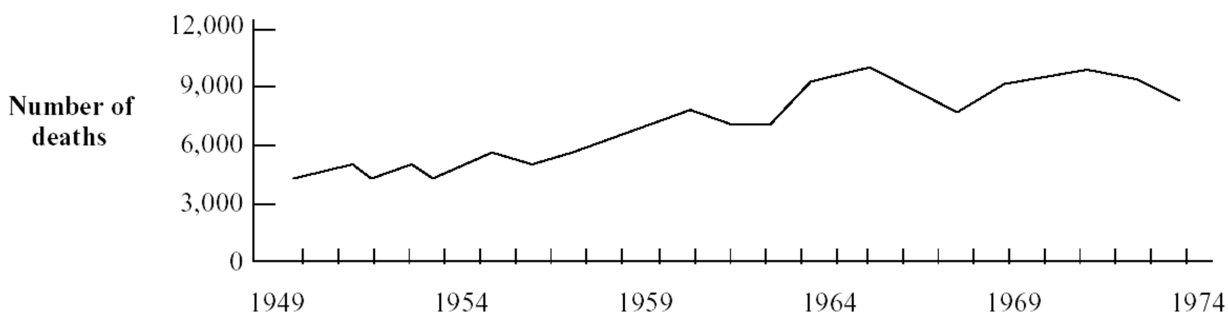
Causal forces are a way of summarizing forecasters' expectations about what will happen to a time series in the future. Contrary to the common assumption for extrapolation, time series are not always subject to consistent forces that point in the same direction. Some are affected by conflicting causal forces; we refer to these as complex times series. It would seem that forecasting these times series would be easier if one could decompose the series to eliminate the effects of the conflicts. Given forecasts subject to high uncertainty, we hypothesized that a time series could be effectively decomposed under two conditions: 1) if domain knowledge can be used to structure the problem so that causal forces are consistent for two or more component series, and 2) when it is possible to obtain relatively accurate forecasts for each component. Forecast accuracy for the components can be assessed by testing how well they can be forecast on early hold-out data. When such data are not available, historical variability may be an adequate substitute. We tested decomposition by causal forces on 12 complex annual time series for automobile accidents, airline accidents, personal computer sales, airline revenues, and cigarette production. The length of these series ranged from 16 years for airline revenues to 56 years for highway safety data. We made forecasts for one to ten horizons, obtaining 800 forecasts through successive updating. For nine series in which the conditions were completely or partially met, the forecast error (MdAPE) was reduced by more than half. For three series in which the conditions were not met, decomposition by causal forces had little effect on accuracy.



*Keywords*: airline accidents, extrapolation, Holt's exponential smoothing, model formulation, personal computers, revenue forecasting, transportation safety.

If you were asked to extrapolate the annual number of deaths on British highways, given the time series presented in Figure 1, how would you proceed?

**Figure 1. Deaths on UK Highways**




We presented this question to a number of forecasting experts, and they suggested several solutions. One suggestion was to make a quantitative extrapolation and then revise it by judgment. This approach has had mixed results in previous studies (e.g., Mathews and Diamantopoulos 1990; Sanders and Ritzman 2001). Others expressed reservations about simply extrapolating the annual number of deaths observed historically. This occurred because, while increases in the safety of highways and automobiles reduce the number of deaths, the greater number of miles driven increases deaths.

We refer to the highway deaths series as a complex time series. For complex time series, experts expect the underlying causal forces to push the series' trend in different directions over the forecast horizon. Such time series can often be represented as the product of two or more observable series. We hypothesized that knowledge of causal forces could be used to better structure forecasting problems with complex series.

## Hypotheses and Prior Research

Decomposition is defined as "the processes of breaking a problem into sub-problems, solving them, and then combining the solutions to get an overall solution" (Armstrong 2001, p.776). In the context of this paper, it would be defined as "dividing a global time series into two or more component series, forecasting each, and then recomposing the components to produce a forecast." We use the term *decomposition* to refer to *multiplicative* breakdowns of a problem (Z = X * Y). We did not examine additive breakdowns (Z = X + Y), often referred to as disaggregation or segmentation.

Decomposition has been widely regarded as a successful strategy for the extrapolation of time series in the traditional approach of using mean, seasonality, trend, and error. The procedure was described by Shiskin (1958). Research has also shown decomposition to be beneficial for judgmental forecasting (MacGregor 2001).

It is commonly assumed that domain knowledge can improve the accuracy of extrapolations. While domain knowledge is seldom used in a formal way in time-series forecasting, the topic is gaining attention. In a review of research, Armstrong and Collopy (1998) found 47 papers on the integration of judgment and statistical methods, most from the previous ten years; they concluded that integration generally improves accuracy when experts have domain knowledge and when significant trends are involved.

Decomposition is likely to improve accuracy when, based on domain knowledge, trends in the components are expected to differ from one another. For example, the highway deaths series includes the effects of changes in the number of miles driven in the UK as well as effects of safety improvements. We expected that since the forces differ, the forecast errors would be less likely to be correlated with one another. Armstrong and Collopy (2001) found that errors from extrapolation methods tended to be in the direction of the causal forces (e.g., for growth forces, the actual values were much more likely to exceed the forecast values.) Thus, the forecast errors for the components are likely to compensate for one another, which should reduce errors in the overall recomposed forecast.

In addition, domain knowledge can be used to select the functional form (e.g., additive or multiplicative). In many downward sloping economic series, negative numbers are not sensible and a multiplicative trend can be chosen to reflect this.

Decomposition can be risky because errors in the components multiply when the forecasts are recombined. For example, a 20% increase in forecast error for one component would increase the overall error by 20%, all other things being equal. Furthermore, when the errors in the forecasts of the components are in the same direction, the errors can be explosive; an increase of 20% in the forecast errors for two components translates into a 44% increase in the forecast error for the global variable (1.2 * 1.2 = 1.44). By comparison, for a time series that was disaggregated (additively decomposed), a 20% increase in the forecast errors for two components of equal size would produce only a 20% forecast error for the global series.

Decomposition should be done so that the errors in each of the components are not excessive. MacGregor (2001) also found this to be important for judgmental decomposition. We decided that the ideal way to determine if the errors from the decomposition would be greater than from the global series would be to simulate the forecasting situation. We proposed two operational rules: Our preferred rule was that each of the components could be forecast over a simulation period with less error than could the aggregate. The second rule was that the coefficient of variation of each of the components would be less than that for the global series. This latter rule was expected to be useful for short series.

Decomposition is only expected to be useful when there is substantial uncertainty in forecasting the global series. As noted by MacGregor (2001), decomposition is expected to be more valuable in situations involving high uncertainty. This might be reflected by the coefficient of variation of the global series. It also implies that decomposition would be more useful as the forecast horizon increases.

**Causal Forces to Represent Domain Knowledge**

To use domain knowledge, forecasters must have reliable information beyond what is available in the historical series. We structure this knowledge though a scheme that we refer to as “causal forces.” The purpose is to capture an expert’s expectations about the direction of a trend and the functional form to best represent that trend.

The use of causal forces first occurred to us in response to a request for forecasts of epidemics by the Chinese Academy of Medicine in Beijing. Some researchers had used standard time-series extrapolation procedures to forecast epidemics (e.g., Broughton 1991). We believed that using those procedures was inappropriate because they are based on the assumption that trends will continue, whereas time series for epidemics change when cures take effect. When forecasters have domain knowledge (say that most people had been vaccinated), they should expect a change in the trend.

Forecasters can use causal forces to structure much of the domain knowledge about time series trends. After examining hundreds of times series, we classified causal forces into six categories that relate historical trends in the data to expectations based upon domain knowledge.

In the first category, *growth*, we expect forces to push the trends upward, irrespective of historical trends. Managers might make this assumption for sales of a product marketed aggressively in a healthy economy.

In the second category, *decay*, we expect the forces to push the trend downward, irrespective of historical trends. For example, decay would be used to represent a product from which marketers are withdrawing support.

The third category, *supporting*, involves forces that are expected to reinforce the historical trend's direction. This assumption is implicit in traditional extrapolation methods. We have had difficulty finding examples, although, real estate prices might be one.

*Opposing* forces, the fourth category, occur when the forces act in a direction opposite to the historical trend. In this case, the time interval must be long enough for decision makers to take actions to affect the data in the following time period. For example, consider a quarterly time series for inventory as a percent of sales; low inventories damage service so managers increase stocks, but high inventories increase holding costs and prompt managers to reduce inventory.

In the fifth category, *regressing*, the forces cause the series to move toward a mean value. Time series for athletic performance are often subject to regressing forces.

Finally, there are series for which the forces are *unknown*. In such cases, domain experts either lack knowledge or cannot separate the directional effects of conflicting forces.

Armstrong and Collopy (2001) noted that the benefit of using causal forces increases as the forecast horizon lengthens because the causal effects increase accordingly. This reinforces our expectation that decomposition by causal forces is more advantageous as the horizon increases.

In research on rule-based forecasting, causal forces have been used to improve the weights for combining extrapolation forecasts (Collopy and Armstrong 1992). They have also been used to produce simple heuristics for selecting among extrapolation methods. For example, the rule "Do not use trend extrapolation if the historical trend is contrary to causal forces" produced substantial improvements in the forecast accuracy of extrapolation methods (Armstrong and Collopy 1993). Finally, causal forces help to explain why forecast errors for economic data are often asymmetric, even when expressed as logs (Armstrong and Collopy 2001); this allowed for improvements in calibrating prediction intervals. The current research builds upon these previous studies by using causal forces in decomposing time series.

**Research Design**

We compared the accuracy of direct extrapolation of the global series with extrapolation using decomposition by causal forces. Direct extrapolation represents current practice and is recommended in major forecasting texts. Our hypothesis was that decomposition by causal forces would improve forecast accuracy when

(1) uncertainty is high, (2) forecasters can use domain knowledge to decompose the problem such that different forces can be identified for two or more component series, (3) the causal forces imply trends that differ in direction, and (4) it is possible to obtain forecasts for each component that are more accurate than the forecast for the global series.

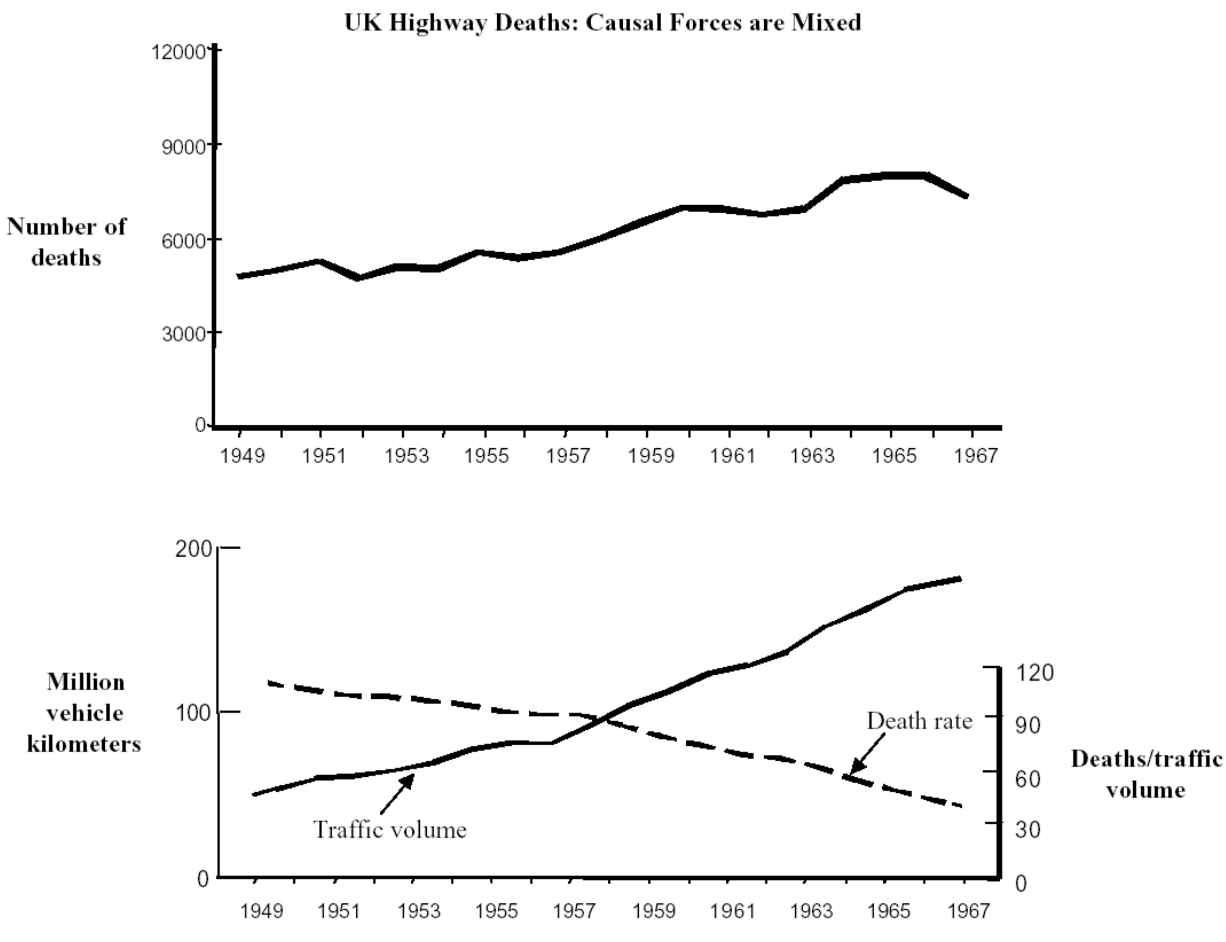
We first describe our initial analysis, which was of British data on motor vehicle deaths, injuries, and accidents, noting that some forces drive the number of deaths up, while others drive them down. Using the same procedures, we analyzed nine series in five other areas that involved U.S. motor vehicle safety, airline safety, airline yield, personal computer sales, and cigarette production.

**U.K. Motor Vehicle Deaths, Injuries, and Accidents,**

We obtained data from a study of highway safety in Great Britain (Broughton 1991, and personal correspondence with Broughton). These data included the annual numbers of deaths, serious injuries, and accidents on highways in Great Britain from 1949 to 2000. We selected these series because, in our judgment, they were complex, domain knowledge was available, and we expected that the components would be fairly easy to forecast. Our model-calibration data consisted of 32 observations from 1949 through 1980; we withheld the data from 1981 to 2000 for *ex ante* forecast validation.

All three global series were affected by growth forces (an increase in the amount of traffic) and decay forces (safety improvements). We isolated the forces by using data on traffic volume (a growth series) to calculate the rates for each of the three global values. The resulting accident rate, injury rate , and death rate series were as decay series. Figure 2 shows highway deaths, along with the two components for 1949 through 1967.

**Figure 2**




We used domain knowledge as follows:

- Growth forces affected traffic over the forecast horizon as population, affluence, and the number and quality of highways all increased. To reduce the risk of large errors, we used an additive trend.
- The rate component *(e.g.*, deaths per million vehicle kilometers) was affected by decay forces, as roads, cars, and safety practices among drivers *(e.g.*, using seat belts) all improved. We used a

multiplicative (log) form to reflect that these cannot have negative values and that the rate of decrease in the units slows as the series approaches zero.

- We multiplied the component forecasts to produce recomposed forecasts (e.g., traffic volume times death rate).

As a benchmark, we used Holt's exponential smoothing (Holt, et al. 1960) to forecast the global series. This widely used method weights both levels and trends. We used SAS ETS with the parameters estimated from an ARIMA (0,2,2) model to estimate and produce the Holt forecasts. Like most extrapolation methods, Holt's is based on the assumption that the forces acting on a series over the forecast horizon will tend to be in the same direction as the recent trend in the series; in other words, as it is typically used, it incorrectly assumes supporting causal forces.

**Component Forecasts and Variability**

To compare the forecastability of the components with that of the global series, we conducted a test. Using the data available for calibrating the model, we divided each series into an estimation (or fit) portion and a validation portion.

Using the fit data through 1960 we produced *ex ante* forecasts for 1961 through 1970 and calculated the resulting forecast errors. Then we added the next observation, 1961, to the estimation data, re-estimated the models, made new forecasts, and calculated errors. We repeated this procedure until all but the last observation was included in the estimation data, producing 55 forecasts.

Following the recommendations in Armstrong and Collopy (1992) we used median absolute percentage errors (MdAPEs). For the analysis of the estimation sample, we defined the accuracy of the components for a given horizon to be superior to the forecasts for the whole series (the global forecasts) if the MdAPE of each of the components forecasts was less than the MdAPE of the global forecast for the vast majority of the horizons.

Table 1 shows the errors from forecasts for 1961-1970. MdAPEs for the direct forecasts were compared with those for each of the components. Boldface indicates instances in which the global forecasts were more accurate than either the miles or rate forecasts. This occurs only for accidents; as a result, our hypothesis was that decomposition would not be expected to help—and might be risky—in forecasting the accidents series.

**Table 1. British Motor Vehicle Safety:**
**MdAPEs for Global and Component Forecast -1961-1970 by Horizon**

**(**Boldface numbers show where the global forecasts were more accurate)

| | | Deaths | | Injuries | | Accidents | |
|---|---|---|---|---|---|---|---|
| **Horizon** | **Traffic** | **Global** | **Rate** | **Global** | **Rate** | **Global** | **Rate** |
| 1 | 1.5 | 5.8 | 4.9 | 3.4 | 3.4 | **3.2** | 4.2 |
| 2 | 2.1 | 7.9 | 7.3 | 6.2 | 4.8 | 5.7 | 5.6 |
| 3 | 3.6 | 9.5 | 7.3 | 8.9 | 6.8 | **6.2** | 7.7 |
| 4 | 5.1 | 10.7 | 5.9 | 10 | 9.6 | **7.9** | 9.8 |
| 5 | 5.0 | 14.9 | 3.6 | 15.6 | 13.1 | **8.9** | 11.6 |
| 6 | 5.3 | 19.9 | 3.7 | 16.2 | 14.8 | **14.1** | 15.4 |
| 7 | 7.3 | 23.3 | 5.3 | 22.7 | 15.2 | 23.1 | 16.1 |
| 8 | 8.8 | 26.2 | 7.3 | 28.8 | 21.3 | 31.8 | 17.0 |
| 9 | 8.4 | 30.5 | 10.7 | 33.6 | 27.7 | 34.7 | 19.1 |
| 10 | 10.8 | 41.0 | 10.6 | 34.7 | 32.3 | 40.1 | 21.2 |

**An Alternative Test of Forecastability**

It may not always be possible to conduct a simulation like that described above. Often there is not enough historical data to calibrate and then test the models. For this reason, we wanted to examine whether an alternative test might be workable. For that alternative, we compared the coefficients of variation (CV) about the trend line.

For the deaths series, the coefficient of variation about the trend line for the global series was 6.4, while for traffic and death rate the CVs were 5.6 and 1.2, respectively. For injuries, the global series CV was 6.1, and for the components, the CVs were 5.6 and 1.9 for the traffic volume and injury rate, respectively. For accidents, however, the CV for the global series was 3.9, for the rate, 2.0, and for the traffic component, 5.6; the CV for traffic volume being higher than the global series CV. In other words, this decision rule produced the same result for these series, suggesting that if the hypotheses were confirmed, it might be a viable alternative.

Both decision rules indicated that decomposition by causal forces should be used for the series on deaths and serious injuries, but not for the accidents series.

**Results from U. K. Highway Safety Data**

Figure 3 shows global and recomposed forecasts for deaths that we obtained using data through 1967. The direct forecast is that deaths will continue to increase, while the recomposed forecast is the opposite. In the actual series, deaths decreased. There were 6,614 deaths for the tenth year, yielding an absolute percentage error (APE) of 4.3 for the recomposed forecast, compared with 29.5 for the direct forecast. The mean absolute percentage error (MAPE) for the direct forecasts across all of the ten horizons was 14% versus 5% for the recomposed forecasts.

**Figure 3. Recomposed Forecasts Were More Accurate for Deaths on UK Highways as of 1967**


We then extended the estimation data for the three British highway series to 1980. By using the process described and conducting successive updates over the 20-year hold-out period from 1981 to 2000, we obtained comparisons of the global and recomposed forecasts. This produced the twenty one-year-ahead forecast errors averaged in Figure 4 under horizon 1, nineteen two-year-ahead forecast errors averaged under horizon 2, down to 10 ten-year ahead forecasts.

We used the MdAPE as the primary criterion in analyzing the holdout forecasts. We also examined the mean absolute percentage error (MAPE) and the median relative absolute error (MdRAE). (The RAE – the error in the proposed model divided by the error for the naive forecast – is described in Armstrong and Collopy, 1992, and in the dictionary found at forecastingprinciples.com).

The results agreed with our expectations (Table 2). For deaths, the recomposed forecasts were more accurate for nine of the ten horizons, and their average error was 61.4% less than the error of the global forecasts (MdAPE of 11.7 versus 30.4). For serious injuries, the recomposed forecasts were more accurate for all ten horizons, and their average error was 76.8% less than that of the global forecasts (12.9 versus 55.7). Averaging across both series, decomposition reduced forecast errors by two-thirds. The superiority of the recomposed forecasts over the global forecasts was statistically significant at p=.0001 for deaths and p=.007 for injuries (using the one-tailed Wilcoxon signed ranks test for paired differences).

In the case of accidents, for which both decision rules judged decomposition by causal forces to be inappropriate, decomposition would have increased the error on all ten forecast horizons. Overall, its use would have increased the error by 128% in comparison with the global forecasts.

We obtained similar results when we examined the other error measures. Using MAPE, the recomposed forecasts for the deaths series had an average error that was 62.1% less than that of the global forecasts (MAPE of 12 versus 31.7), and for serious injuries the recomposed forecasts had 77.1% less error. For accidents, recomposed forecasts had 52.7% more error.

**Table 2. British Motor Vehicle Safety (1981-2000) –Gain from Recomposed Forecast Errors**

| | **Deaths** | | | **Injuries** | | | **Accidents** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Horizon** | **Global** | **Recomposed** | **Gain** | **Global** | **Recomposed** | **Gain** | **Global** | **Recomposed** | **Gain** |
| 1 | 4.3 | 5.0 | -0.7 | 8.3 | 1.9 | 6.4 | 2.8 | 5.7 | -2.9 |
| 2 | 8.7 | 5.2 | 3.5 | 16.8 | 4.7 | 12.1 | 3.0 | 9.1 | -6.1 |
| 3 | 12.5 | 8.0 | 4.5 | 25.2 | 7.1 | 18.1 | 5.4 | 11.4 | -6.0 |
| 4 | 15.9 | 9.5 | 6.4 | 33.4 | 8.2 | 25.2 | 5.5 | 13.2 | -7.7 |
| 5 | 23.5 | 11.7 | 11.8 | 44.5 | 13.5 | 31.0 | 8.1 | 15.0 | -6.9 |
| 6 | 26.5 | 12.0 | 14.5 | 57.9 | 16.6 | 41.3 | 4.7 | 16.7 | -12.0 |
| 7 | 37.0 | 16.5 | 20.5 | 73.7 | 18.4 | 55.3 | 8.3 | 17.3 | -9.0 |
| 8 | 51.4 | 16.5 | 34.9 | 86.0 | 19.2 | 66.8 | 8.9 | 18.2 | -9.3 |
| 9 | 59.9 | 15.3 | 44.6 | 99.9 | 18.9 | 81.0 | 10.2 | 21.4 | -11.2 |
| 10 | 63.7 | 17.3 | 46.4 | 111.9 | 20.7 | 91.2 | 10.2 | 24.7 | -14.5 |
| MdAPE | 30.4 | 11.7 | 61.4% | 55.7 | 12.9 | 76.8% | 6.7 | 15.3 | -128.4% |
| MAPE | 31.7 | 12.0 | 62.1% | 55.4 | 12.7 | 77.1% | 7.9 | 16.7 | -52.7% |
| MdRAE | 1.5 | 0.4 | 73.3% | 2.1 | 0.5 | 76.2% | 1.4 | 2.1 | -50.0% |

Using the median of the relative absolute errors (MdRAEs), we found that the recomposed forecasts of deaths had 73.3% less error than the direct forecasts (MdRAE of .4 versus 1.5) and for serious injuries they had 76.2% less error. For accidents, where decomposition was not recommended, the MdRAEs for the recomposed forecasts had 50% more error than the global forecasts.

Across all three series, 305 of the 465 comparisons (66%) were in the predicted direction ($p<.0001$). Of course, all of the above tests overstate statistical significance because the errors at each horizon are not independent of one another. More important than the tests of statistical significance, however, is (1) the consistency of our results with prior research on causal forces, and (2) the large effect sizes for the two series for which decomposition by causal forces was recommended.

The accuracy gains increased as the forecast horizon increased. The correlation between the gain in accuracy and the forecast horizon was .96 for deaths and .99 for injuries.

Encouraged by these results, we conducted extensions to determine if the findings would hold for other series and to see whether the decision rules involving the tests of the CVs would be satisfactory when there was insufficient data to use the pretest approach for accuracy.

**U.S. Motor Vehicle Safety**

We forecast U.S. motor vehicle deaths, injuries, and accidents, using 56 observations from 1945 through 2000 from *Accident Facts* and the National Highway Administration. The three series mirrored the British highway safety data in that they were affected by the same conflicting causal forces: growth, due to an increase in traffic, and decay, due to improvements in the safety of automobiles, highways and drivers.

We tested the forecast accuracy of individual components using the data from 1945 to 1980 with forecasting horizons from 1981 through 1990. The components were more accurate than the global series for all data except for one-period-ahead horizons on the accidents and deaths data. The coefficients of variation (about the trend) for the components, 2.9 and 6.5, were lower than the CV for the deaths series, 13.2. The same was true for injuries for which the component CVs were 7.1 or 6.5 and the global CV was 37.1; and for accidents, where the component CVs were 1.3 and 5.6, and the global CV was 6.3. Both tests showed that conditions were suitable for applying decomposition by causal forces to all three series.

We then extended the estimation data to 1990 and prepared forecasts for 1991 through 2000. For the injuries and accidents series, the recomposed forecasts were more accurate than the global forecasts for all ten horizons. For deaths, the average error was 63.2% less for the recomposed forecasts (MdAPE of 4.6 versus 12.5) and the recomposed forecasts were more accurate for eight of the ten horizons. For injuries, the average error was 95.7% less than that of the global series (MdAPE of 3.8 versus 87.5), and for accidents, the average error for the recomposed forecasts was 79.6% less (MdAPE of 11.5 versus 56.3). Averaging across all three U.S. highway safety series, the overall forecast error was 80% lower for the recomposed series than for the global forecasts.

Similar results were obtained using the other error measures. The MAPE of the recomposed forecasts was lower for all three series. Over the ten horizons, the decrease in the MAPE averaged 77.4%. The decrease in MdRAE for the three series averaged 61.1%.

The recomposed forecasts were again statistically superior to the global forecasts. Using a one-tail Wilcoxon signed rank test, the recomposed forecast errors were significantly lower (at p<.0001) than the global forecast errors (Table 3). This was true for all three individual series, even though deaths had the smallest mean difference and the p-value for injuries was .032. The percentage of positive differences in the error (i.e., showing a smaller recomposed error) was significantly larger than the null hypothesis of no difference (at p<.0001). The correlation between the forecast horizon and the gain in accuracy was .75 for death, .99 for injuries, and .14 for accidents.

**Table 3. US Motor Vehicle Safety (1991-2000) –**
**Summary of Forecast Errors for Horizons 1 to 10**

| | Deaths | | | Injuries | | | Accidents | | |
|---|---|---|---|---|---|---|---|---|---|
| **Horizon** | **Global** | **Recomposed** | **%Gain** | **Global** | **Recomposed** | **%Gain** | **Global** | **Recomposed** | **%Gain** |
| MdAPE | 12.5 | 4.6 | 63.2 | 87.5 | 3.8 | 95.7 | 56.3 | 11.5 | 79.6 |
| MAPE | 12.8 | 4.7 | 63.3 | 86.5 | 3.9 | 95.5 | 56.3 | 14.8 | 73.7 |
| MdRAE | 2.8 | 0.6 | 78.6 | 2.1 | 0.8 | 61.9 | 1.4 | 0.8 | 42.9 |

**Airline Safety**

The airline accident data, which consisted of 56 observations from 1945 to 2000, was taken from the *Statistical Abstract of the United States*. In the preliminary test, the forecasts of the individual components were more accurate than the global series forecasts for the accident series for all horizons from 1981 to 1990.

The CV for the mileage component exceeded the CV for the global series (44.3 vs. 25.8), but the CV was only 4.0 for the accident rate series. Still, because we had enough data and the predictability pretest was favorable to decomposition by causal forces, decomposition was advised.

The results for seven of ten forecasting horizons (1991 to 2000) supported decomposition by causal forces (Table 4). For accidents, forecasts from recomposed series were more accurate (35.5) than global forecasts (40.1) , an 11.5% decrease.

**Table 4. Airline Accidents (1991-2000) –**
**Summary of Forecast Errors for Horizons 1 to 10**

| | Global | Recomposed | %Gain |
|---|---|---|---|
| MdAPE | 40.1 | 35.5 | 11.5 |
| MAPE | 36.7 | 36.9 | -.5 |
| MdRAE | 2.6 | 0.9 | 65.4 |

Based on the MAPE, the gains in accuracy provided by the recomposed forecasts were unclear. In the airline accident series, the MAPE *increased* by .5% (MAPE of 36.9 versus 36.7). Based on the MdRAE, the recomposed forecasts decreased forecast error by 65.4% for airline accidents (MdRAE of .9 versus 2.6). The correlation between the forecast horizon and the gain in accuracy was .90.

To assess the sensitivity of the decomposition to choice of denominator in a rate series, we analyzed 27 observations on U.S. airline accidents from 1973 through 1999 using five horizons (1995 through 1999). We divided by two different bases (either by mileage or by departures) and kept the same numerator.

For these short series, one can compare predictability of the individual components based only on the coefficients of variation. For the global series, the CV was 28.7 and for the component series for accidents/miles, the CVs were 5.1 and 7.2, while for components series for accidents/departures, the CVs were 4.6 and 4.9. Since all components were less variable, decomposition by causal forces was advised.

The recomposed forecasts were almost always more accurate than the global forecasts (Table 5). The average MdAPE for accidents, when miles traveled was used as the base for the rate, was 20 for the recomposed series versus 41.6 for the global series, a 51.9% decrease. The forecast error decreased by 35% when the base variable was departures.

**Table 5. Airline Accidents (1995-1999)**
**Summary of Forecast Errors for Horizons 1 to 5**

| | Accidents/Mile | | | Accidents/Departure | | |
|---|---|---|---|---|---|---|
| **Horizon** | **Global** | **Recomposed** | **%Gain** | **Global** | **Recomposed** | **%Gain** |
| MdAPE | 41.6 | 20.0 | 51.9 | 39.7 | 25.8 | 35.0 |
| MAPE | 38.8 | 18.1 | 53.4 | 37.6 | 28.2 | 25.0 |
| MdRAE | 0.9 | 0.8 | 11.1 | 2.3 | 0.6 | 73.9 |

We obtained similar results with the MAPE for accidents/mile. The recomposed forecast error of 18.1 was 53.4% lower than the average error for the global forecasts. For accidents/departure, the error of 28.2 was 25% less than that produced by the global series. Likewise, the MdRAE for the two recomposed series were 11.1% and 73.9% less. The gains in accuracy were correlated with the forecast horizon .88 for accidents/mile but it was -.10 for accidents/departure. Bear in mind that in this test the forecast horizon was only five years.

**Airline Revenue**

Airline revenues are affected by growth and decay forces. Airline passenger miles were a growth series, while revenue per passenger mile was classified as decay, due to improvements in technology and to the deregulation of airlines in Europe and the U.S.

The data on airline revenues covered only 1984 through 1999, so it was difficult to test for relative ease of forecasting the components. We used an additive trend for the mileage to reduce the risk of large errors, and a multiplicative trend for the yield. We withheld data from 1995 through 1999 for *ex ante* forecast validation.

The CV test led us to reject decomposition, because the CV of one of the components exceeded the CV for the global series (4.9 vs. 3.9). Despite this, forecasts for the recomposed series were more accurate than the global forecasts for four of the five horizons. The MdAPE was reduced by 47.3% with decomposition. The MAPE decreased by 46.6%, and the MdRAE by 25% (Table 6). The correlation between the accuracy gain and the forecast horizon was .99.

**Table 6. Airline Revenues (1995-1999)**
**Summary of Forecast Errors for Horizons 1 to 5**

| | Global | Recomposed | %Gain |
|---|---|---|---|
| MdAPE | 16.9 | 8.9 | 47.3 |
| MAPE | 16.1 | 8.6 | 46.6 |
| MdRAE | 0.8 | 0.6 | 25.0 |

**Personal Computers**

We analyzed prices and U.S. shipments of personal computers based on 23 annual observations from 1976 through 2000. The data were taken from the *U.S. Industrial Outlook (1978-1988), Computer Industry Almanac (1989-1995)*, and *Gartner Dataquest (1998-2000)*. We constructed a price index from data provided in Berndt and

Rappaport (2001) and by Berndt (personal communication). The projected decline in the prices of personal computers (related to Moore's Law) is caused by improved technology and competition. The growth in unit sales is caused by growth in PC applications and decreases in prices. We used multiplicative log transformations for unit sales and prices. For the component series, the CVs were 4.0 and 5.3, which were less than the CV for the global series (16.0), so we deemed decomposition appropriate.

We withheld data from 1996-2000 for forecast validations and again used successive updating. The recomposed forecasts were more accurate than the global forecasts. The average MdAPE was 25.2% lower, (MdAPE of 10.7 versus 14.3). The reduction in MAPE was about 10% and the reduction in MdRAE was 58.7% (Table 7). The gains in accuracy were larger for the longer horizons with a correlation of .80.

**Table 7. Personal Computers (1996-2000)**
**Summary of Forecast Errors for Horizons 1 to 5**

| | Global | Recomposed | % Gain |
|---|---|---|---|
| MdAPE | 14.3 | 10.7 | 25.2 |
| MAPE | 13.6 | 12.3 | 9.6 |
| MdRAE | 1.2 | 0.5 | 58.7 |

**Cigarette Production**

We analyzed the world production of cigarettes using 1950 through 1995 data from the United States Department of Agriculture. The components were per capita production (a proxy for per capita consumption) and population. There were 46 observations, including ten forecast horizons (1986 to 1995). In the U.S., there were decay forces because of efforts to make smoking less convenient, less socially acceptable, and more expensive, but in the rest of the world, these decay forces were weak or nonexistent.

Using data from 1950 through 1975, we made forecasts for the ten forecasting horizons from 1976 to1985. In our tests, the components proved more difficult to forecast than the global series. Also, the coefficient of variation test showed that both of the components were more variable than the global series. Because it was a contrary series (i.e., the historical trend was in a different direction than expected) we used a random walk forecast for per-capita production, a rule developed by Collopy and Armstrong (1992).

Although we judged decomposition to be inappropriate, decomposition by causal forces was more accurate than the global forecasts. The average global error was 5.9 and the recomposed forecast error was 2.8, a decrease of

52.5%, while the MAPE decreased by 36.5 (Table 8). The correlation between the accuracy gain and the forecast horizon was .90.

**Table 8. Cigarette Production (1986-1995) –**
**Summary of Forecast Errors for Horizons 1 to 10**

| | Global | Recomposed | %Gain |
|---|---|---|---|
| MdAPE | 5.9 | 2.8 | 52.5 |
| MAPE | 5.2 | 3.3 | 36.5 |
| MdRAE | 5.4 | 1.3 | 75.9 |

**Further Analyses of Conditions**

We assumed that decomposition by causal forces would be most useful for series with high uncertainty. In the extreme (series that do not vary), decomposition would have no value. Therefore, we examined only series that had high uncertainty. To assess the importance of this assumption, we compared the gain in accuracy with the global CV for each of the nine series where decomposition was appropriate. The correlation between accuracy gain and CV for the global series was .43, so even among series with much uncertainty, the gains were greatest for those series with the highest uncertainty.

As a further test related to uncertainty, we tested each of the series to examine whether the gains were larger for the longer horizons. The median correlation between the gain in accuracy and the forecast horizon was 0.88. Using Rosenthal's (1978) method for summing statistical significance across studies, the null hypothesis that accuracy gain is not related to the forecast horizon was rejected at $p < .0001$ ($t = 11.5$).

For each of the 12 series, we tested whether decomposition by causal forces was advisable by examining the coefficients of variation and, where possible, the comparative forecasting accuracy of the components versus that of the global series. These two approaches yielded the same recommendations on seven of the eight cases where both could be used to evaluate predictability.

## Limitations

Our tests were limited to annual data, for which we expected the influence of causal forces to be strong. We do not know whether decomposition by causal forces might be useful for shorter interval data, such as monthly or daily. Also, we used mostly long series for our tests in order to determine whether the components could be forecasted with greater accuracy than the global values. For the short series we were able to use only the coefficient of variation test. Despite these limitations, we achieved substantial gains at low cost.

With the exception of the cigarette data, the tests were limited to data involving growth and decay series. In such cases, the expectations for the direction of the trends differ. While we expect that the results would apply to other series where the causal forces differ on direction, one should be cautious about generalizing beyond growth and decay series.

Our rules recommended against the use of decomposition for three series. Although decomposition reduced accuracy substantially for one of these series, it would have improved accuracy in the other two. Because we have no theoretical bases for expecting these improvements in other situations, we must still advise against the use of decomposition by causal forces for such cases. But further work should be done in this area.

## Conclusions

The forecasting errors were 56% lower for the nine series for which decomposition by causal forces was appropriate (Table 9). The costs of the approach are low because causal forces are easy to specify and forecasters need intervene only when the causal forces change for a series (which, based on research to date, seems infrequent).

**Table 9**
**Percentage Error Reduction using Decomposition by Causal Forces**
(US data unless noted otherwise)

| | **Decomposition recommended?** | | **Percent Reduction** |
|---|---|---|---|
| **Conditions Met?** (series length/ # forecasts) | **Accuracy** | **Coefficient of variation** | **MdRAE** |
| **Yes** | | | |
| Highway Deaths -UK (52/155) | Yes | Yes | 73.3 |
| Highway Injuries -UK (52/155 | Yes | Yes | 76.2 |
| Highway Deaths (56/55) | Yes | Yes | 78.6 |
| Highway Injuries (56/55) | Yes | Yes | 61.9 |
| Highway Accidents (54/55) | Yes | Yes | 42.9 |
| **Average gain** | | | **66.6** |
| **Partly** | | | |
| Airline Accidents (56/55) | Yes | No | 65.4 |
| Airline Accidents/Miles (27/15) | N/A | Yes | 11.1 |
| Airline Accidents/Depart (27/15) | N/A | Yes | 73.9 |
| Personal Computers (23/15) | N/A | Yes | 58.7 |
| **Average gain** | | | **52.3** |
| **No** | | | |
| Airline Revenues (16/15) | N/A | No | 25.0 |
| Highway Accidents-UK (52/155) | No | No | -50.0 |
| Cigarette Prod. (world) (46/55) | No | No | 75.9 |
| **Average gain** | | | **16.7** |

Forecasters should identify complex time series. Although such series are not expected to be common, they are likely to be subject to large forecast errors. For series with uncertainty, when the trend directions implied by causal forces conflict, and when the components can be forecasted more accurately than the global series for complex time series, decomposition by causal forces can reduce forecast errors by about half. Forecasters can easily adapt decomposition by causal forces to current extrapolation methods. To aid in this, software programs might be extended to guide users through the decomposition process.

**Acknowledgement**s: Partial support for this research was provided by the U.S. Navy Personnel R & D Center and by the Office of Naval Research (under grant number N00014-92-J-1544). Monica Adya, Dennis A. Ahlburg, Ulrich Kuesters, Steve DeLurgio, Donald G. MacGregor, Richard Webby, Nigel Meade, and Eric Stellwagen provided helpful comments. Michael Halperin provided much of the data. Two anonymous reviewers made suggestions that led to major improvements. Mary Haight and Marian Lee provided excellent editorial advice.

## References


*Accident Facts*, National Safety Council, 1945-1949, 1951-1958, 1963, 1966.

Armstrong, J. S. (2001), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

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

Armstrong, J. S. and F. Collopy (1993), “Causal forces: Structuring knowledge for time-series extrapolation,” *Journal of Forecastin*g, 12, 103-115.

Armstrong, J. S. and F. Collopy (1998), “Integration of statistical methods and judgment for time series forecasting: Principles from empirical research,” in G. Wright and P. Goodwin (eds.), *Forecasting with Judgment*. Chichester: John Wiley, pp. 269-293.

Armstrong, J. S. and F. Collopy (2001), “Identification of asymmetric prediction intervals through causal forces,” *Journal of Forecasting*, 20, 273-283.

Berndt, E. and N. Rappaport (2001), “Price and quality of desktop and mobile personal computers: A quarter-century historical overview,” *American Economic Association Papers and Proceedings*, 91, 2, 268-273.

Broughton, J. (1991), “Forecasting road accident casualties in Great Britain,” *Accident Analysis and Prevention,* 5, 353-362.

Collopy, F. and J. S. Armstrong (1992), “Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations,” *Management Science,* 38, 1394-1414.

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

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

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

Rosenthal, R. (1978), “Combining results of independent studies,” *Psychological Bulletin*, 85, 185-193.

Sanders, N. and L. Ritzman (2001), “Judgmental adjustment of statistical forecasts,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers, pp. 405-416.

Shiskin, J. (1958), “Decomposition of economic time series,” *Science*, 128, 1539-1546.