Rule-based Forecasting vs. Damped-trend Exponential Smoothing

Everette S. Gardner, Jr., Ph.D.
College of Business Administration
University of Houston
Houston, Texas 77204-6282

Published in:
Management Science, Vol. 45, No. 8 (August, 1999), pp. 1169-1176.

Rule-based Forecasting vs. Damped-trend Exponential Smoothing

Abstract

This paper evaluates the *ex ante* performance of rule-based time series forecasting systems proposed in earlier research. We show that comparable performance can be obtained with a simpler alternative, a damped-trend version of exponential smoothing fitted to minimize the MAD criterion. The results suggest that the performance of rule-based systems would be improved through this alternative and that time series forecasters should consider MAD fits in model development.

(*Combining forecasts, Exponential smoothing, Extrapolation, Expert systems, Judgment, Rule-based forecasting*)

Rule-based Forecasting vs. Damped-trend Exponential Smoothing

1. Introduction

In an earlier paper in this journal, Collopy and Armstrong (C&A) (1992) proposed a rule-based approach to time series forecasting. Drawing on protocol analyses of five forecasting experts, the authors developed a comprehensive rule base, including some 99 rules, to combine the forecasts from four extrapolation methods according to 18 features of time series. The four extrapolation methods included a random walk, time series regression, Brown's double exponential smoothing for a linear trend, and Holt's method of exponential smoothing for a linear trend. Rule-based forecasting was tested on annual time series drawn from the Makridakis et al. (1982) forecasting competition and gave promising results. Compared to the popular equal-weights combination of forecasts, rule-based forecasting performed significantly better at both short and long-term forecast horizons.

The C&A system requires considerable human intervention in identifying features of time series. Vokurka et al. (1996) extended C&A's research with a rule-based expert forecasting system (RBEFS) that uses predefined rules to automatically identify time series features and select an extrapolation method from several alternatives, including simple exponential smoothing, Gardner-McKenzie (1985) damped-trend exponential smoothing, classical decomposition, and a combination of all methods. Using the same time series as C&A, Vokurka et al. also conducted experiments in which judgment was incorporated into RBEFS: the user was allowed to select from a fixed set of built-in methods based on a judgmental appraisal of the time series. Vokurka et al. found no significant difference between the performance of RBEFS and the C&A system.

This paper compares the C&A and Vokurka et al. rule-based forecasting systems to a simpler alternative: damped-trend exponential smoothing alone, fitted to minimize the MAD. The C&A system did not use damped-trend smoothing as one of the candidate methods for combining forecasts. Vokurka et al.'s REBFS included damped-trend smoothing but the MSE was used in model-fitting.

2. Model-fitting criteria in time series forecasting

Referees for this paper pointed out that a MAD fit is unconventional in time series forecasting. They are correct. A literature search showed that most empirical research has followed the precedent set in the M-Competition (Makridakis et al., 1982), using the MSE as a model-fitting criterion, with *ex ante* forecast accuracy judged by other criteria. The only exception appears to be the work of Weiss and Andersen (1984), who reexamined the ARIMA results for 80 series drawn from Makridakis' subset of 111 series (31 series were deemed not suitable for ARIMA modeling). In one-step-ahead forecasting, Weiss and Andersen found little difference between the forecasts from OLS estimation and those from models estimated to minimize the MAD or mean APE. However, when trace estimation methods were used, the conclusions changed. A trace is a sequence of forecasts from leadtime one through some maximum value. If the maximum is six periods, and the holdout sample is six, the trace consists of six forecasts at one-step-ahead, five at two-steps-ahead, four at three-steps-ahead, and so on. The mean trace forecast error is the average of all forecast errors over all leadtimes. Models fitted to minimize trace MAD or mean APE were superior to OLS, regardless of whether trace absolute errors or trace squared errors were used for *ex ante* evaluation. Fildes and Makridakis (1988) found the Weiss and Andersen conclusions ambiguous because they apply only to trace forecasting using a trace-estimated model, and I agree. The Weiss and Andersen results are not comparable to Makridakis et al. and it is not clear how the results might be generalized.

In a theoretical argument, Zellner (1986) maintained that one should match the error measure used in model development with the error measure used in *ex ante* model evaluation. Fildes and Makridakis (1988) responded that there was no empirical evidence that matching these error measures made any practical difference in *ex ante* model evaluation. Makridakis and Hibon (1992) agreed.

Later work by Armstrong and Collopy (1992) showed that the MSE is an unreliable statistic for *ex ante* forecast evaluation. Furthermore, the MSE produces poor outlier protection and only fair construct validity. Fildes (1992) made a number of other arguments against the MSE in the context of *ex ante* evaluation in large numbers of time series. Neither Armstrong and Collopy nor Fildes discussed model-fitting and I could find no evidence that the work of these authors has been used to justify the MAD as a model-fitting criterion. This is surprising. In my opinion it is difficult to defend the MSE in model-fitting, given that it is such a poor choice for *ex ante* evaluation. The results below appear to be the first reported test of model-fitting criteria in exponential smoothing, certainly the most widely-used time series methodology.

3. The damped-trend exponential smoothing system

The forecasting system is based on the class of autoregressive-damping systems, also known as damped-trend systems, developed by Gardner and McKenzie (1985). In error-correction form, the damped trend is written:

$$S_t = S_{t-1} + \phi T_{t-1} + h_1 e_t \,, \tag{1}$$

$$T_t = \phi T_{t-1} + h_2 e_t \,, \tag{2}$$

$$\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \phi^i T_t \,. \tag{3}$$

The one-step-ahead forecast error is defined as $e_t = X_t - \hat{X}_{t-1}$ (1). S_t and T_t are the level and trend components of the series. There are two smoothing parameters, h_1 and h_2, for level and trend, and an autoregressive-damping parameter ϕ to control the rate of growth in the forecasts.

It may not be obvious that a number of options are available for fitting the damped-trend system. Each parameter can be constrained to one of at least three regions: the range 0 to 1, the region defined by discounted-least-squares, or to much larger regions of stability. Although a grid search for parameters is common in practice, more sophisticated search algorithms are readily available. Initial values for the level and trend components may be computed through backcasting, time series regression, or simple averages of the first few data observations. Outliers can be identified during model-fitting and adjusted prior to a final fit. The system can be fitted to minimize the MSE, the MAD, or other error measures. Following the fit, the user can also adjust the forecasts to compensate for any first-order autocorrelation found in the residuals. For more details of the options in model-fitting, see Gardner (1985).

4. Fitting the damped trend to the Makridakis data

In this study, all parameters were constrained to the range 0 to 1 using a grid-search algorithm. Initial values of model components were computed using the first five data observations. The average difference amongst these observations was taken as the initial trend, with the initial level set equal to the first data observation. Outliers in the residuals from model-fitting were identified and the original data were adjusted. The procedure was to compute 95% normal probability limits around the residuals from an initial fit. If an error fell outside the probability limits, the corresponding data observation was set equal to the forecast. Next, the model was refitted to the adjusted data. *Ex ante* forecast errors were computed only once, after the refit. The residuals of few series contained significant first-order autocorrelation, so no such adjustments were made.

On *a priori* grounds, the damped-trend model was fitted to minimize the MAD. In the exponential smoothing software used in this study, Peer Planner (Delphus, 1997), it is convenient to use either the MSE or MAD to fit any model. The MAD was selected because both C&A and Vokurka et al. evaluated forecasting performance strictly by absolute error criteria.

Another problem in model-fitting was that early data was irrelevant in some series and distorted the smoothing parameters. We scanned graphs of the fit periods and judgmentally identified 16 of 126 series with irrelevant early data. In 9 of the 16 problem series, the first one or two observations were much smaller than the remainder, which interfered with estimation of the initial trend. In the remaining 7 series, initial trend estimates were distorted by huge discontinuities early in the fit periods. Like C&A, we simply trimmed irrelevant early data rather than attempt any type of data adjustment, although the trimming was not the same in each case. Details of the fit periods by series are available from the author. The effects of trimming are evaluated below.

Although we did not test the effects of all alternative model-fitting options, at the request of the referees we did compute differences in *ex ante* forecast accuracy for all combinations of MAD and MSE fits, data with and without outlier adjustments, and data with and without trimmed fit periods. Table 1 shows median and mean APE results for the complete set of 126 annual time series analyzed by Vokurka et al. In all series, the damped-trend system was fitted through period n-6, where n is the total number of observations. *Ex ante* median and mean APEs were computed through forecast horizons 1-6 and for the cumulative forecasts. The damped-trend results can be replicated using the Peer Planner system, with fit options set as described above.

The MAD fit gave better *ex ante* median and mean APEs than the MSE fit at every forecast horizon, regardless of whether outliers were adjusted or fit periods were trimmed. For both MAD and MSE fits, the effect of outlier adjustment was to increase the cumulative median APE, with mixed results at the individual forecast horizons. This was true, regardless of whether the fit periods were trimmed. For example, compare runs 1 and 2 in Table 1. Run 1 is the base case (using original data) and produced a cumulative median APE of 7.8%. In run 2, outlier removal increased the cumulative median APE to 8.1%. Now compare runs 3 and 4. In run 3, outliers were not removed but fit periods were trimmed, producing a cumulative median APE of 7.6%. This value increased to 7.8% in run 4 after outlier removal.

In retrospect, outlier removal was not a good decision for this data. However, trimmed fit periods made a small improvement in the cumulative median APE. The effects of outlier removal and trimmed fit periods were offsetting. Run 4 included both outlier adjustments and trimmed fit periods and produced exactly the same cumulative median APE as the base case in run 1.

TABLE 1.
Effects of model-fitting options on the damped trend, 126 series.

	Outliers	Fit periods	Fit	*Ex ante*	*Ex ante* APE by horizon						
Run	**removed?**	**trimmed?**	**Criterion**	**error measure**	**1**	**2**	**3**	**4**	**5**	**6**	**Cum.**
1	No	No	MAD	Median APE	3.3	**4.6**	**7.2**	11.1	**12.0**	15.9	7.8
2	Yes	No	MAD	Median APE	2.7	4.8	8.2	10.9	12.4	15.3	8.1
3	No	Yes	MAD	Median APE	3.2	**4.6**	**7.2**	10.6	**12.0**	15.4	**7.6**
4	Yes	Yes	MAD	Median APE	**2.6**	**4.6**	7.6	**9.8**	12.3	**14.8**	7.8
5	No	No	MAD	Mean APE	6.9	9.4	14.3	17.7	19.4	23.9	15.3
6	Yes	No	MAD	Mean APE	6.5	8.9	14.0	17.6	19.4	23.6	15.0
7	No	Yes	MAD	Mean APE	6.5	8.8	13.5	**16.9**	**18.2**	22.7	14.4
8	Yes	Yes	MAD	Mean APE	**6.1**	**8.4**	**13.4**	**16.9**	18.4	**22.6**	**14.3**
9	No	No	MSE	Median APE	3.4	5.6	**8.1**	**11.6**	**13.5**	**15.4**	**8.4**
10	Yes	No	MSE	Median APE	**3.3**	**5.2**	9.1	11.9	13.6	15.9	9.0
11	No	Yes	MSE	Median APE	3.4	5.5	8.3	11.9	13.7	**15.4**	8.5
12	Yes	Yes	MSE	Median APE	**3.3**	**5.2**	9.2	11.9	14.1	15.9	9.0
13	No	No	MSE	Mean APE	**6.7**	9.4	14.6	18.5	20.2	24.7	15.7
14	Yes	No	MSE	Mean APE	6.8	9.5	14.9	18.5	20.5	24.3	15.8
15	No	Yes	MSE	Mean APE	6.8	**9.1**	**14.4**	18.2	**20.0**	24.1	**15.4**
16	Yes	Yes	MSE	Mean APE	**6.7**	9.3	14.6	**18.1**	20.2	**23.7**	**15.4**

4. Comparisons to C&A and Vokurka et al.

Tables 2-4 (next page) compare the *ex ante* forecast accuracy of damped-trend smoothing with selected results from C&A and Vokurka et al. The damped-trend results are taken from Run 4 in Table 1, with outlier removal and fit periods trimmed. This option is shown because it was selected *a priori.* In tables 2-4, it was not possible to test for statistical significance of differences from damped-trend smoothing because we did not have details of the individual errors by series in C&A and Vokurka et al.

Table 2 makes median APE comparisons to Vokurka et al. for all 126 series (C&A do not give results for all series, since some were used for model development). In Vokurka et al.'s "user procedure," the user selected one of the built-in methods in RBEFS after a judgmental evaluation of the time series. With the exception of a tie at horizon 2, damped-trend smoothing with a MAD fit was more accurate at all horizons than the user procedure or RBEFS.

Although not shown in Table 2, any of the damped trend with MSE fit results (runs 9-12 in Table 1) are better than Vokurka et al.'s expert system. It is interesting that the damped trend with MSE fit, without outlier removal or trimming of fit periods (run 9 in Table 1) gives about the same accuracy as the Vokurka et al. user procedure.

In Table 3, some rather involved comparisons are made to both C&A and Vokurka et al. for a subset of 90 series taken from the 126 used in Table 2. Following C&A, these series are further subdivided into validation sets 1, 2, and 3, consisting of 18, 36, and 36 series respectively. C&A reported results only for horizons 1 and 6, so other horizons are not shown.

TABLE 2.
Damped trend vs. Vokurka et al., 126 series.

	Ex ante median APE by horizon						
Method	1	2	3	4	5	6	Cum.
Damped trend (MAD fit)	**2.6**	**4.6**	**7.6**	**9.8**	**12.3**	**14.8**	**7.8**
User procedure	3.6	**4.6**	8.4	12.1	13.1	15.1	8.4
Expert system (RBEFS)	3.4	5.3	8.0	13.7	13.9	17.7	10.5
Equal-weights combination	6.9	10.1	13.7	15.6	17.5	21.8	12.4
Random walk	5.6	10.6	15.0	19.1	21.9	26.2	15.7

TABLE 3.
***Ex ante* median APE comparisons for 90 series (taken from the 126 series in Table 2 above).**

		1-step-ahead median APE				6-step-ahead median APE			
Source	Method	V1(18)	V2(36)	V3(36)	Wtd.Avg.	V1(18)	V2(36)	V3(36)	Wtd.Avg.
Gardner	Damped trend (MAD fit)	**1.6**	**2.7**	3.3	**2.8**	16.9	16.1	12.8	14.8
C&A	Rule-based forecasting	2.5	3.1	**3.2**	3.0	**13.0**	**9.1**	14.2	**11.9**
Vokurka et al.	User procedure	3.2	3.1	4.0	3.5	22.5	15.8	**11.6**	15.5
	Expert system (RBEFS)	4.2	3.0	4.0	3.6	22.4	16.8	14.5	17.0
	Equal-weights comb.	7.0	6.7	7.1	6.9	25.1	21.3	19.4	21.3
	Random walk	6.4	5.7	5.6	5.8	30.1	24.7	25.2	26.0

TABLE 4.
***Ex ante* median APE comparisons for 36 series (validation set 3 in Table 3 above).**

		Median APE			Mean APE		
Source	Method	1-yr.	6-yr.	Cum.	1-yr.	6-yr.	Cum.
Gardner	Damped trend (MAD fit)	3.3	12.8	8.9	**6.2**	**19.1**	13.3
C&A	Rule-based forecasting	**3.2**	14.2	8.7	6.3	23.6	15.0
Vokurka et al.	User procedure	4.0	**11.6**	**7.7**	6.7	20.0	**11.4**
	Expert system (RBEFS)	4.0	14.5	9.8	6.7	23.6	14.0
	Equal-weights comb.	7.1	19.4	10.4	8.5	24.7	15.7
	Random walk	5.6	25.2	16.3	7.6	26.1	17.8

In interpreting Table 3, it is important to understand that the weighted average is not the median APE of all forecast errors but rather a weighted average of the medians for each validation set. It would be useful to know the true cumulative median APE over all forecast horizons for all series, but these were not reported by C&A, except for validation set 3 as discussed below. For the damped trend, cumulative median APEs differ substantially from weighted averages.

In Table 3, compared to C&A, the damped trend is significantly more accurate in validation sets 1 and 2 at 1-step-ahead and gives about the same results in validation set 3. At 6-steps-ahead, the C&A system is significantly more accurate in validation sets 1 and 2, while the damped trend does better in validation set 3.

How do we explain the differences in performance among validation sets? This is a difficult question to answer because the C&A rule base is complex. We can say that most series in validation set 3 are relatively well-behaved and easy to forecast while validation sets 1 and 2 contain some ill-behaved series. The C&A rule base appears to be less sensitive to such series, especially at long forecast horizons.

To illustrate the problems in forecasting validation sets 1 and 2, eight time series are plotted in Figure 1. These series were selected due to their disproportionate influence on summary *ex ante* error measures. In Figure 1, symbols mark the forecast periods (which are always the last six periods). Five series (numbers 27, 32, 62, 92, and 165) contain a cycle or abrupt trend reversal during the forecast periods. For example, in series 32, the data grew by 47% during the first five forecast periods but dropped by 75% in the last forecast period. No time series model can be expected to cope with such an anomaly in the last forecast period. The remaining series (numbers 5, 47, and 87) are characterized by drastic changes in rate of growth from fit to forecast periods. The worst of these is series 5, in which the data declined by 10% during the last six fit periods. The result is that the damped-trend model projected a negative trend into the forecast periods but the data actually grew by 250% during the forecast periods.

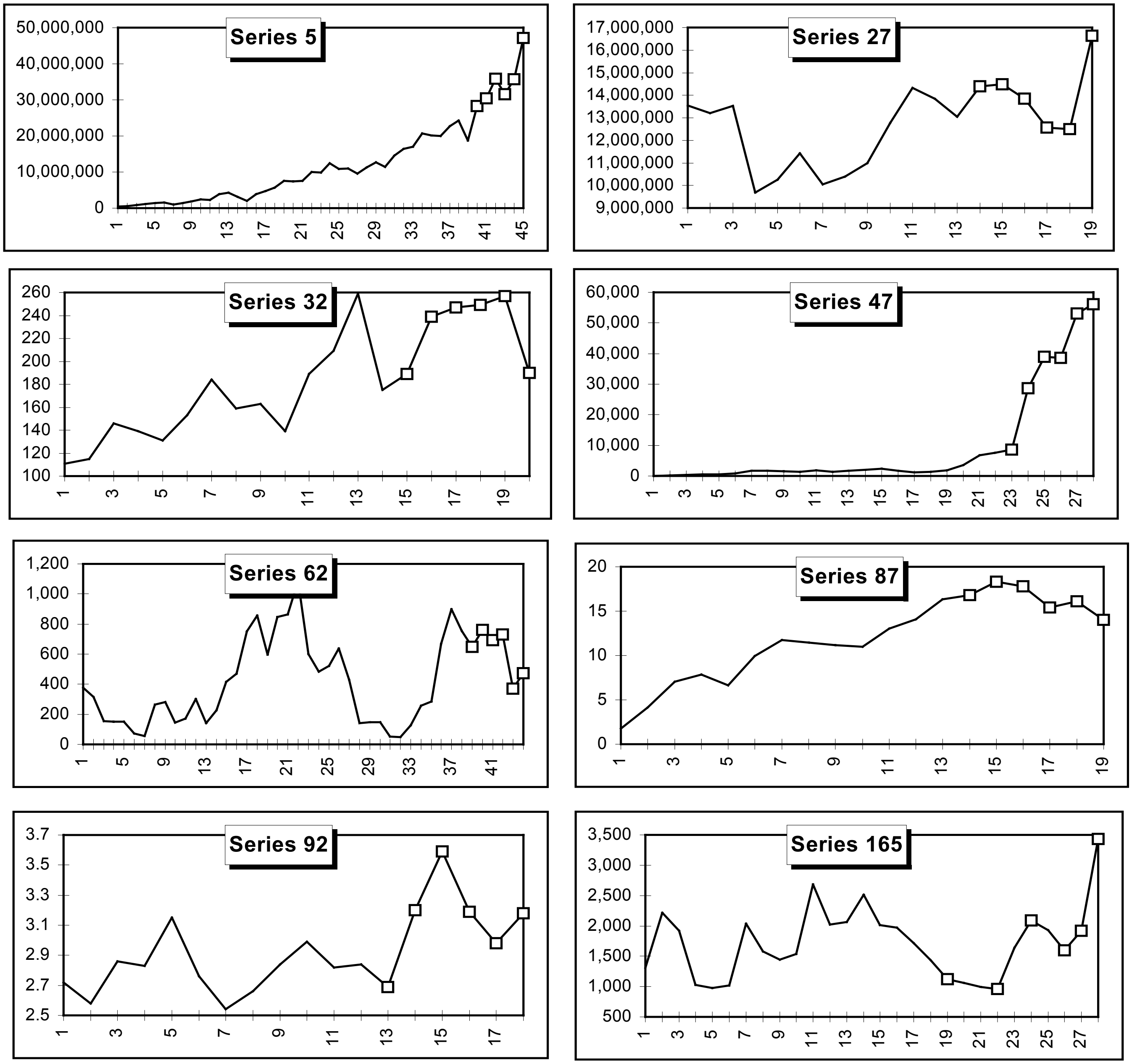
In general, the damped trend was confounded by the series in Figure 1 and produced huge errors during most forecast periods. If the two series from validation set 1 (numbers 5 and 165) are dropped, the 1-step-ahead median APE for the damped trend falls from 1.6% to 1.2%, while the 6-step-ahead value falls from 16.9% to 12.3%. All other series in Figure 1 are from validation set 2. Dropping them reduces the 1-step-ahead median APE from 2.7% to 1.8%, with a reduction from 16.1% to 15.9% at 6-steps-ahead.

To understand the minor reduction at 6-steps-ahead, note that the last observation in the forecast periods was usually out of character with the rest of the forecast periods and reverted to the long-term trend in the series. As an example, see series 32 in which the last observation falls back to a level near that at the beginning of the forecast periods.

In Table 3, compared to Vokurka et al.'s user procedure and RBEFS, the damped trend is more accurate at 1-step-ahead in all validation sets. The damped trend is also more accurate at 6-steps-ahead in validation set 1 and less accurate in validation sets 2 and 3.

Table 4 gives more detailed results for the 36 series in validation set 3. The damped trend and C&A produce about the same cumulative median APE over all forecast horizons but the Vokurka et al. user procedure is superior to both. On the mean APE criterion, the damped trend gives a better cumulative value than C&A, but again the Vokurka et al. user procedure wins.

Figure 1. Examples of series with changing patterns during the forecast periods.



5. Conclusions

Damped-trend smoothing is more accurate than either Vokurka et al.'s user procedure or RBEFS at all forecast horizons, using the complete sample of 126 series. When the 126 series are divided into validation sets or subsamples, relative performance depends on the subsample. The relative performance of damped-trend smoothing and C&A also depends on the subsample, although it may be that damped-trend smoothing is a better short-term forecaster, with C&A superior in the long term.

Therefore it is difficult to argue that the C&A or Vokurka et al. systems are consistently more accurate than damped-trend smoothing alone. This conclusion is important because damped-trend smoothing is much simpler, in terms of operation as well as understanding on the part of the user.

While the C&A and Vokurka et al. systems in their present form contain worthwhile features, it should be possible to improve the performance of both. In the case of C&A, damped-trend smoothing should be incorporated as one of the candidate models for combining forecasts, perhaps as a replacement for one of the two linear-trend smoothing methods. In the Vokurka et al. systems, damped-trend smoothing is already a candidate, but median and mean APE performance should improve if the damped trend is fitted as described above.

When should forecasters prefer a MAD rather an MSE fit? Given the odd time series displayed in Figure 1, my opinion is that it is unreasonable to generalize that a MAD fit should produce better *ex ante* median and mean APEs. The time series in this study are also relatively short and there is no risk of confounding seasonality with trend. In longer, more complex time series, MAD fits may well produce different results.

More research is planned to evaluate the effects of MAD vs. MSE fits. Pending further research, forecasters should compare *ex ante* APE results for both model-fitting options. When forecasting with exponential smoothing, the choice of model-fitting criterion can easily be automated.

References

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

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

Fildes, R. and Makridakis, S., "Forecasting and Loss Functions," *Intl. J. of Forecasting*, 4 (1988), 545-550.

Fildes, R., "The evaluation of extrapolative forecasting methods," *Intl. J. of Forecasting*, 8 (1992), 81-98.

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

Gardner, E. S., Jr., and McKenzie, E., "Forecasting trends in time series," *Management Sci.*, 31 (1985), 1237-1246.

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

Makridakis, S. and Hibon, M., "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," *Intl. J. of. Forecasting*, 7 (1991), 317-330.

Delphus, Inc., *Peer Planner*, Morristown, NJ: 1997.

Vokurka, R. J. et al., "Automatic feature identification and graphical support in rule-based forecasting: a comparison," *Intl. J. of Forecasting*, 12 (1996), 495-512.

Weiss, A. and Andersen, A., "Estimating time series models using the relevant forecast evaluation criterion," *J. R. Statist. Soc. A* 147, Part 3 (1984), 484-487.

Zellner, A., "A tale of forecasting 1001 series: The Bayesian Knight strikes again, *Intl. J. of Forecasting*, 2 (1986), 491-494.