

5-1-2006

Choosing Smoothing Parameters For Exponential Smoothing: Minimizing Sums Of Squared Versus Sums Of Absolute Errors

Terry E. Dielman

Texas Christian University, t.dielman@tcu.edu

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Dielman, Terry E. (2006) "Choosing Smoothing Parameters For Exponential Smoothing: Minimizing Sums Of Squared Versus Sums Of Absolute Errors," *Journal of Modern Applied Statistical Methods*: Vol. 5 : Iss. 1 , Article 11.

Choosing Smoothing Parameters For Exponential Smoothing: Minimizing Sums Of Squared Versus Sums Of Absolute Errors

Terry Dielman
M.J. Neeley School of Business
Texas Christian University

When choosing smoothing parameters in exponential smoothing, the choice can be made by either minimizing the sum of squared one-step-ahead forecast errors or minimizing the sum of the absolute one-step-ahead forecast errors. In this article, the resulting forecast accuracy is used to compare these two options.

Key words: Exponential smoothing, forecasting accuracy, M-competition, outliers, parameter selection, Simulation

Introduction

In a number of comparisons of forecasting methods, exponential smoothing methods have been shown to be simple but relatively accurate techniques for generating forecasts (See Makridakis et al., 1982; Makridakis et al., 1993; Makridakis & Hibon, 2000). When using exponential smoothing methods to forecast a time series, a smoothing parameter (or parameters) must be chosen. One way this choice can be made is to choose the parameter or parameters that minimize some error criterion over the history of the data available. Typically, the choice made is to minimize the sum of squared one-step-ahead forecast errors (SSE). Another option would be to minimize the sum of the absolute one-step-ahead forecast errors (SAE). Minimizing SSE is the most often used criterion for choosing the smoothing parameter, but minimizing SAE could provide protection against outliers in the time series. This article examines the question of which of these choices might be best in practice.

Terry Dielman is Professor of Decision Sciences in the Information Systems and Supply Chain Management department, M.J. Neeley School of Business, Texas Christian University. Email: t.dielman@tcu.edu

In the context of regression models, forecasts generated from least squares (equivalent to SSE) coefficient estimates and least absolute value (equivalent to SAE) coefficient estimates were studied by Dielman (1986). When the disturbance distribution was long-tailed, presenting the opportunity for outliers, the least absolute value based forecasts were, on the whole, superior to the least squares based forecasts. These results were obtained from a simulation study assuming that an exogenous independent variable was available for use in the regressions. Whether the superiority of a least absolute value type criterion could exist for smoothing parameter choice and subsequent generation of forecasts in exponential smoothing methods is the issue considered in this article.

The analyses presented in this article support three main conclusions: First, while instances where outliers will degrade forecast performance may not be common, such instances do occur in practice. Second, minimizing SAE to determine exponential smoothing parameters can provide protection against such outliers. Finally, on average, minimizing SAE does not result in much, if any, deterioration in forecast accuracy over minimizing SSE when conditions are optimal for SSE.

Methodology

M1-Competition Data

Three exponential smoothing techniques are examined in this part of the study: single exponential smoothing, Brown's double exponential smoothing, and Holt's two-parameter exponential smoothing.

The one-period-ahead forecast for single exponential smoothing can be written as

$$\hat{y}_{T+1} = \alpha y_T + (1 - \alpha) \hat{y}_T \quad (1)$$

All subsequent forecasts have the same value. The smoothing parameter, α , must be chosen to implement this forecasting technique. The choice is made by performing a grid search over the range 0.01, 0.02, ..., 0.99 and choosing the value of α from this range that minimizes either the SSE or SAE.

Brown's double exponential smoothing is often suggested when data are trended. The m -period-ahead forecasts are generated from the following equations:

$$S'_t = \alpha y_t + (1 - \alpha) S'_{t-1} \quad (2)$$

$$S''_t = \alpha S'_t + (1 - \alpha) S''_{t-1} \quad (3)$$

$$\hat{y}_{T+m} = a_T + m b_T \quad (4)$$

where

$$a_t = 2S'_t - S''_t \quad (5)$$

and

$$b_t = \frac{\alpha}{1 - \alpha} (S'_t - S''_t) \quad (6)$$

As with single exponential smoothing, the smoothing parameter, α , is chosen by performing a grid search over the range 0.01, 0.02, ..., 0.99 and choosing the value of α from this range that minimizes either the SSE or SAE.

Holt's two-parameter exponential smoothing is also suggested when data are trended, but is somewhat more flexible than Brown's method because separate parameters

are allowed for the two smoothing equations. The m -period-ahead forecasts are generated from the following equations:

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (7)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (8)$$

$$\hat{y}_{T+m} = L_T + mT_T \quad (9)$$

Values for two parameters, α and β , must be chosen in this case. Again, a grid search is used with values of 0.01, 0.02, ..., 0.99 for each parameter. All possible parameter value combinations are examined and the pair of values that minimizes either the SSE or SAE is chosen.

The 1001 time series used in the M1 forecasting competition (See Makridakis et al., 1982) are used to evaluate the choice of criteria for choosing the smoothing parameter. The optimal values of the smoothing parameter(s) are chosen for each of the time series. The smoothing parameters for each method that minimize either the SSE or the SAE for each individual time series are chosen. One to six-period-ahead out-of-sample forecasts are then generated using the optimal values under the two criteria. The out-of-sample forecasts are compared to the actual values and accuracy measures are computed for the forecasts. The three accuracy measures reported in this article are the mean absolute percentage error (MAPE), the root mean square error (RMSE), and the mean absolute deviation (MAD). These accuracy measures will be presented to compare the forecasting accuracy for the parameter choices of each criterion.

A Brief Simulation

A small simulation was run to further compare forecast performance for the SAE and SSE criteria. Only single exponential smoothing was examined in this simulation. Single exponential smoothing provides optimal forecasts when the data generation process is ARIMA (0,1,1). This was the process used to generate the data for the simulation experiment. The procedures outlined in Dunne (1992) were used to generate data from an ARIMA (0, 1, 1)

process. The following were factors considered in the experiment:

1. Sample sizes of $T = 20, 30$ and 50 were used.
2. The error distributions considered were:
 - a) Normal with mean zero and standard deviation one (Normal). The following distributions will be referred to as outlier-producing distributions:
 - b) Contaminated Normal with 0.75 probability of observations coming from a $N(0,1)$ distribution and 0.25 probability from a $N(0,5)$ distribution. The contamination was introduced in three different ways to assess potential situations where the minimum SAE criterion might outperform the minimum SSE criterion.

CNR5: The contamination was allowed to occur randomly throughout the time series.

CNB5: The first 25% of the observations were from the $N(0, 5)$ distribution.

CNE5: The last 25% of the observations were from the $N(0, 5)$ distribution.
 - c) Same as b but the contaminating distribution was $N(0, 10)$ (CNR10, CNB10 and CNE10).
 - d) Cauchy with median zero and scale parameter one (Cauchy). These errors represent a pathological situation where extreme outliers are possible and should be the best-case scenario for minimizing SAE.
3. The true value of the exponential smoothing parameter was set at $0.2, 0.3, 0.5, 0.7$ and 0.8 .

For each experimental setting of the simulation, $10,000$ time series were generated, the optimal value of the smoothing parameter was estimated using a grid search over the values $0.01, 0.02, \dots, 0.99$, and one period ahead forecasts were computed using this parameter

value. Out-of-sample forecasts were computed and were compared to the actual values (which were generated from the process used in the simulation) and the MAPE, RMSE, and MAD were computed for these $10,000$ forecasts. All programs were written in FORTRAN and IMSL subroutines were used for random number generation.

Results

M1-Competition Results

Each of the three exponential smoothing methods was applied to each of the 1001 time series from the M1-competition. Optimal smoothing parameters to minimize both SSE and SAE were chosen and forecasts were generated. Table 1 shows the values of the accuracy measures for the one through six period ahead forecasts (combined). Table 2 shows the values for the one period ahead forecast. Cases where minimizing SAE results in greater accuracy are highlighted in bold. The choice of criterion is dependent to some extent on the accuracy measure. For example, in Table 1 the MAPE is smaller for the SAE criterion for single exponential smoothing, although the RMSE and MAD are both smaller for the SSE criterion. This experiment was conducted using seasonally adjusted data as well (where appropriate) with little difference in the results of the comparison. The forecast accuracy was improved regardless of criterion (because of the presence of seasonal series in the data set), but the difference in forecast accuracy between SAE and SSE did not change appreciably. The tables for the seasonally adjusted results have not been included in the article.

The results suggest that there are instances where the SAE forecasts provide improvement over the SSE forecasts according to some accuracy criterion. In other words, there are cases with outliers present that can affect forecast accuracy. The results from the simulation are intended to shed additional light on situations when the SAE forecasts might be most beneficial.

Table 1: Accuracy Measures One Through Six Period Ahead Forecasts

	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Single	17.5	17.7	578348	572521	32884	32668
Brown	20.7	19.9	290890	272913	19475	18056
Holt	22.5	22.3	290928	389576	19657	25428

Table 2: Accuracy Measures One Period Ahead Forecasts

	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Single	11.1	11.1	297704	291140	14001	13695
Brown	13.3	13.5	120026	119844	9713	10088
Holt	14.0	14.0	123836	170598	10714	13391

Simulation Results

Tables 3 through 17 summarize the simulation results. In all experimental settings when the disturbances were normal, there was little difference between accuracy measures for minimizing SAE versus SSE. In cases where there was a difference, the accuracy measures for minimizing SSE were smaller. In most of the outlier-producing distributions, the accuracy measures for minimizing SAE were smaller than those for minimizing SSE. The differences in the

accuracy measures in favor of SAE are more pronounced in cases where the true smoothing constant is larger and where outliers are more likely. When the contaminated normal disturbances were used, the differences in the accuracy measures in favor of SAE occurred when the standard deviation was larger (10 rather than 5) and when the occurrence of the outliers was at the end or throughout the series rather than at the beginning.

Table 3: Accuracy Measures for Simulation using $T = 50$ and $\alpha = 0.2$

<u>Errors</u>	MAPE		RMSE		MAD	
	SAE	SSE	SAE	SSE	SAE	SSE
Normal	0.82	0.81	1.03	1.02	0.82	0.81
CNR5	1.64	1.67	2.69	2.71	1.63	1.66
CNR10	2.71	2.84	5.14	5.21	2.63	2.77
CNB5	0.83	0.83	1.03	1.04	0.82	0.83
CNB10	0.84	0.85	1.03	1.05	0.83	0.84
CNE5	4.16	4.18	5.19	5.23	4.14	4.16
CNE10	8.54	8.58	10.47	10.55	8.34	8.38
Cauchy	7.24	7.90	85.14	86.00	6.07	6.72

Table 4: Accuracy Measures for Simulation using $T = 30$ and $\alpha = 0.2$

<u>Errors</u>	MAPE		RMSE		MAD	
	SAE	SSE	SAE	SSE	SAE	SSE
Normal	0.82	0.81	1.03	1.02	0.82	0.81
CNR5	1.65	1.70	2.69	2.73	1.64	1.69
CNR10	2.73	2.91	5.13	5.25	2.67	2.85
CNB5	0.82	0.83	1.03	1.05	0.82	0.83
CNB10	0.83	0.86	1.03	1.07	0.82	0.85
CNE5	4.15	4.22	5.21	5.31	4.13	4.20
CNE10	8.54	8.67	10.56	10.76	8.37	8.50
Cauchy	7.59	9.26	63.97	98.24	5.10	7.21

Table 5: Accuracy Measures for Simulation using $T = 20$ and $\alpha = 0.2$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.83	0.83	1.04	1.04	0.83	0.83
CNR5	1.68	1.73	2.73	2.77	1.68	1.72
CNR10	2.79	2.96	5.20	5.34	2.74	2.91
CNB5	0.84	0.85	1.05	1.07	0.83	0.85
CNB10	0.85	0.89	1.05	1.11	0.84	0.88
CNE5	4.12	4.22	5.15	5.28	4.11	4.21
CNE10	8.36	8.75	10.30	10.80	8.21	8.59
Cauchy	10.15	11.86	1574.45	1570.05	23.62	23.60

Table 6: Accuracy Measures for Simulation using $T = 50$ and $\alpha = 0.3$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.82	0.82	1.03	1.02	0.82	0.82
CNR5	1.64	1.67	2.69	2.71	1.62	1.66
CNR10	2.72	2.85	5.13	5.21	2.62	2.76
CNB5	0.83	0.84	1.03	1.04	0.83	0.83
CNB10	0.85	0.86	1.04	1.06	0.83	0.85
CNE5	4.15	4.19	5.18	5.23	4.13	4.16
CNE10	8.58	8.63	10.46	10.54	8.33	8.39
Cauchy	6.62	7.17	85.13	86.17	6.04	6.68

Table 7: Accuracy Measures for Simulation using $T = 30$ and $\alpha = 0.3$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.82	0.81	1.03	1.02	0.82	0.81
CNR5	1.64	1.70	2.68	2.73	1.64	1.69
CNR10	2.72	2.91	5.11	5.23	2.65	2.84
CNB5	0.83	0.84	1.03	1.06	0.82	0.84
CNB10	0.84	0.87	1.04	1.09	0.83	0.86
CNE5	4.15	4.20	5.20	5.27	4.13	4.18
CNE10	8.52	8.60	10.51	10.62	8.34	8.42
Cauchy	9.38	10.39	64.54	103.00	5.16	7.36

Table 8: Accuracy Measures for Simulation using $T = 20$ and $\alpha = 0.3$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.83	0.83	1.05	1.04	0.83	0.83
CNR5	1.68	1.73	2.72	2.78	1.67	1.73
CNR10	2.77	2.98	5.19	5.33	2.72	2.92
CNB5	0.83	0.85	1.04	1.07	0.83	0.85
CNB10	0.84	0.89	1.05	1.10	0.83	0.88
CNE5	4.13	4.22	5.16	5.27	4.11	4.20
CNE10	8.38	8.69	10.32	10.71	8.22	8.52
Cauchy	32.31	32.89	1572.09	1569.45	23.20	23.50

Table 9: Accuracy Measures for Simulation using $T = 50$ and $\alpha = 0.5$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.82	0.82	1.03	1.02	0.82	0.81
CNR5	1.65	1.68	2.69	2.71	1.62	1.65
CNR10	2.79	2.91	5.14	5.19	2.62	2.73
CNB5	0.84	0.84	1.04	1.05	0.83	0.83
CNB10	0.87	0.89	1.04	1.07	0.83	0.85
CNE5	4.17	4.21	5.18	5.23	4.12	4.16
CNE10	8.73	8.79	10.45	10.55	8.32	8.39
Cauchy	5.84	6.25	85.13	87.23	6.03	6.68

Table 10: Accuracy Measures for Simulation using $T = 30$ and $\alpha = 0.5$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.82	0.81	1.03	1.02	0.82	0.81
CNR5	1.64	1.69	2.67	2.72	1.63	1.68
CNR10	2.76	2.92	5.10	5.21	2.64	2.80
CNB5	0.83	0.85	1.04	1.06	0.82	0.84
CNB10	0.85	0.89	1.04	1.10	0.83	0.86
CNE5	4.14	4.17	5.17	5.22	4.11	4.14
CNE10	8.50	8.58	10.40	10.50	8.26	8.34
Cauchy	9.07	9.84	65.21	124.25	5.22	7.77

Table 11: Accuracy Measures for Simulation using $T = 20$ and $\alpha = 0.5$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.84	0.83	1.05	1.04	0.84	0.83
CNR5	1.67	1.73	2.71	2.77	1.66	1.72
CNR10	2.78	2.97	5.17	5.31	2.70	2.89
CNB5	0.85	0.86	1.06	1.08	0.84	0.86
CNB10	0.86	0.90	1.06	1.12	0.85	0.88
CNE5	4.15	4.21	5.19	5.26	4.13	4.19
CNE10	8.45	8.69	10.38	10.69	8.26	8.50
Cauchy	8.39	10.76	1568.82	1569.82	22.45	23.53

Table 12: Accuracy Measures for Simulation using $T = 50$ and $\alpha = 0.7$

<u>Errors</u>	MAPE		RMSE		MAD	
	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>	<u>SAE</u>	<u>SSE</u>
Normal	0.82	0.82	1.03	1.02	0.82	0.81
CNR5	1.66	1.68	2.69	2.70	1.62	1.64
CNR10	2.97	3.07	5.13	5.18	2.62	2.71
CNB5	0.84	0.85	1.03	1.04	0.82	0.83
CNB10	0.91	0.93	1.04	1.07	0.83	0.85
CNE5	4.19	4.22	5.16	5.20	4.11	4.14
CNE10	8.97	9.05	10.40	10.52	8.28	8.37
Cauchy	9.10	9.42	85.15	89.06	6.04	6.74

Table 13: Accuracy Measures for Simulation using $T = 30$ and $\alpha = 0.7$

<u>Errors</u>	MAPE		RMSE		MAD	
	SAE	SSE	SAE	SSE	SAE	SSE
Normal	0.82	0.81	1.02	1.02	0.82	0.81
CNR5	1.65	1.69	2.67	2.70	1.63	1.67
CNR10	2.87	3.00	5.10	5.19	2.64	2.77
CNB5	0.83	0.84	1.03	1.05	0.82	0.83
CNB10	0.87	0.90	1.04	1.08	0.83	0.85
CNE5	4.11	4.15	5.13	5.17	4.07	4.11
CNE10	8.52	8.63	10.30	10.43	8.18	8.30
Cauchy	7.98	8.52	65.47	161.36	5.21	8.51

Table 14: Accuracy Measures for Simulation using $T = 20$ and $\alpha = 0.7$

<u>Errors</u>	MAPE		RMSE		MAD	
	SAE	SSE	SAE	SSE	SAE	SSE
Normal	0.83	0.83	1.05	1.04	0.83	0.83
CNR5	1.67	1.72	2.71	2.75	1.66	1.70
CNR10	2.82	2.97	5.18	5.28	2.69	2.84
CNB5	0.85	0.86	1.06	1.08	0.84	0.86
CNB10	0.88	0.91	1.07	1.11	0.85	0.88
CNE5	4.16	4.21	5.20	5.26	4.14	4.18
CNE10	8.50	8.78	10.40	10.75	8.28	8.55
Cauchy	7.98	8.92	1567.46	1571.71	21.81	23.71

Table 15: Accuracy Measures for Simulation using $T = 50$ and $\alpha = 0.8$

<u>Errors</u>	MAPE		RMSE		MAD	
	SAE	SSE	SAE	SSE	SAE	SSE
Normal	0.82	0.81	1.02	1.02	0.82	0.81
CNR5	1.67	1.69	2.69	2.70	1.62	1.64
CNR10	3.37	3.45	5.13	5.17	2.62	2.70
CNB5	0.84	0.85	1.03	1.04	0.82	0.83
CNB10	1.00	1.01	1.03	1.06	0.82	0.84
CNE5	4.19	4.22	5.14	5.18	4.10	4.12
CNE10	9.26	9.34	10.35	10.48	8.24	8.33
Cauchy	11.10	11.77	85.17	89.92	6.05	6.73

Table 16: Accuracy Measures for Simulation using $T = 30$ and $\alpha = 0.8$

<u>Errors</u>	MAPE		RMSE		MAD	
	SAE	SSE	SAE	SSE	SAE	SSE
Normal	0.81	0.81	1.02	1.02	0.81	0.81
CNR5	1.66	1.68	2.67	2.69	1.63	1.66
CNR10	4.22	5.07	5.10	5.17	2.64	2.75
CNB5	0.83	0.84	1.03	1.05	0.82	0.83
CNB10	0.88	0.91	1.04	1.08	0.82	0.85
CNE5	4.10	4.14	5.11	5.15	4.06	4.10
CNE10	8.55	8.69	10.25	10.43	8.14	8.29
Cauchy	8.33	8.70	65.71	181.03	5.26	8.91

Table 17: Accuracy Measures for Simulation using $T = 20$ and $\alpha = 0.8$

<u>Errors</u>	MAPE		RMSE		MAD	
	SAE	SSE	SAE	SSE	SAE	SSE
Normal	0.83	0.82	1.04	1.03	0.83	0.82
CNR5	1.67	1.71	2.71	2.75	1.65	1.69
CNR10	2.86	2.98	5.18	5.26	2.69	2.81
CNB5	0.85	0.86	1.06	1.08	0.84	0.85
CNB10	0.88	0.91	1.07	1.11	0.85	0.88
CNE5	4.16	4.21	5.18	5.27	4.13	4.18
CNE10	8.50	8.85	10.38	10.83	8.26	8.59
Cauchy	9.66	10.72	1567.28	1572.77	21.71	23.72

Conclusion

The analyses presented suggest that minimizing SAE to determine exponential smoothing parameters can provide protection against outliers. Analysis of the M1-competition data suggests that cases where parameters selected by minimizing SAE result in superior forecasts do occur in practice. However, on average, minimizing SSE appears to provide forecasts that are reasonably robust to most outliers encountered. The simulation recommends that use of the SAE criterion would be most beneficial with the presence of outliers in conjunction with one or more of the following: larger values of the true smoothing parameter, outliers occurring near the end or throughout the series where forecasts are to be generated rather than at the beginning, and, obviously, cases where larger outliers are more likely. Further, even if outliers are not present, using the SAE criterion will not result in much deterioration in forecast accuracy.

References

- Dielman, T. (1986). A comparison of forecasts from least absolute value and least squares regression. *Journal of Forecasting*, 5, 189-195.
- Dunne, A. (1992). Time series simulation. *The Statistician*, 41, 3-8.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods. *Journal of Forecasting*, 1, 111-153.
- Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simpson, L. F. (1993). The M-2 competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, 9, 5-23.
- Makridakis, S. & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, 15, 451-476.