J. Opl Res. Soc. Vol. 39, No. 9, pp. 863–867, 1988
Printed in Great Britain.
0160-5682/88 $3.00 + 0.00

Model Identification in Exponential Smoothing

EVERETTE S. GARDNER JR[1] and ED. McKENZIE[2]
[1]College of Business, University of Houston, Texas, USA and
[2]Mathematics Department, University of Strathclyde, UK

Model identification has traditionally been ignored in forecasting via exponential smoothing. The usual practice is to apply the same model to every time-series in a collection. This paper develops a procedure for model identification in large forecasting applications based on an examination of variances of differences of the time-series. The order of differencing yielding minimum variance suggests an appropriate model for the series. Empirical results show that this procedure selects models that give reasonable *ex ante* forecast accuracy.

Key words: forecasting, time-series

INTRODUCTION

The aim of this paper is to develop a robust procedure for identifying exponential smoothing models. The procedure should also be simple and efficient so that it can be automated in large forecasting applications such as in inventory control.

In practice, model identification is ignored. It is common to apply the same exponential smoothing model, usually Holt's linear-trend model, to every time-series. The classical references on exponential smoothing[1,2] offer no guidance on model identification except to suggest that visual inspection of a plot of the data may be helpful. However, visual analysis can be misleading. For an example of the pitfalls in this approach, see McKenzie.[3] We agree with Box and Jenkins[4] that it is unsafe to rely on a purely subjective analysis of a time-series.

In Gardner and McKenzie,[5] we developed an indirect procedure for model identification with non-seasonal or deseasonalized data. The procedure is to fit a general, three-parameter model containing a variety of special cases to the time-series. The general model relies on an autoregressive-damping (AD) parameter to control the rate of trend extrapolation. Different values of the AD parameter define a constant-level model, a damped trend and a linear trend. The general model also contains at least six equivalent ARIMA processes. Thus the process of computing minimum-MSE parameters in the general model is a way to identify a more specific model for the data.

Using time-series from the M-competition,[6] we showed that the general model was more accurate than applying a linear trend to every time-series. Accuracy of the general model also compared favourably with more sophisticated models tested in the M-competition, including ARIMA models and the ARARMA models developed by Parzen.[7]

The disadvantage of the general model is that it is inefficient and needlessly complex for many time-series. For example, suppose that a time-series is non-seasonal with a reasonably constant level. A single-parameter model (simple smoothing) is the preferred choice. But three parameters must be computed in the general model to determine that only one is really necessary.

This paper proposes a simpler, more efficient procedure for identifying exponential smoothing models. The procedure is based on a comparison of the variances of relevant differences of the data. The order of differencing yielding minimum variance suggests an appropriate model for the time-series.

The plan of this paper is as follows. The following section summarizes the general model and the special cases it contains. Next, the identification procedure and its rationale are developed. Finally, we test the identification procedure using time-series from the M-competition.

THE GENERAL MODEL

The general model for non-seasonal series is an extension of the Holt model for a linear trend:

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + T_{t-1}), \tag{1}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}, \tag{2}$$

$$\hat{X}_t(m) = S_t + mT_t. \tag{3}$$

S_t and T_t are the level and trend components of the series. The smoothing parameters are α and γ. The m-step-ahead forecast is given by (3).

The trend component is modified with an AD parameter, ϕ, and the general model is

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + \phi T_{t-1}), \tag{4}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)\phi T_{t-1}, \tag{5}$$

$$\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \phi^i T_t. \tag{6}$$

The general model includes three possibilities for trend, depending on the value of ϕ found during model-fitting. If there is no trend in the data, the best fitting ϕ will be near 0, and the model reduces to a constant level (simple smoothing). If there is a strong trend, ϕ will be fitted at a value near 1, and the model gives about the same results as the linear trend in (1)–(3). Finally, if the trend is erratic, model-fitting will yield $0 < \phi < 1$ and the trend is damped. The forecasts approach an asymptote given by the straight line $S_t + T_t\phi/(1 - \phi)$.

The model in (4)–(6) can be simplified considerably. Using the one-step-ahead forecast error, $e_t = X_t - S_{t-1} - \phi T_{t-1}$, we can rewrite (4) and (5) as

$$S_t = S_{t-1} + \phi T_{t-1} + h_1 e_t, \tag{7}$$

$$T_t = \phi T_{t-1} + h_2 e_t, \tag{8}$$

where

$$h_1 = \alpha; \quad h_2 = \alpha\gamma.$$

The non-seasonal forecasting system given by (6)–(8) has at least six equivalent ARIMA processes. By equivalent, we mean that the forecasts are minimum mean-squared-error (MSE) forecasts for the corresponding ARIMA processes. These processes provide a statistical rationale for the model. If $0 < \phi < 1$, the trend is damped and the equivalent process is ARIMA (1, 1, 2), which can be written as

$$(1 - B)(1 - \phi B)X_t = [1 - (1 + \phi - h_1 - \phi h_2)B - \phi(h_1 - 1)B^2]e_t. \tag{9}$$

We can also obtain an ARIMA (1, 1, 1) process by setting $h_1 = 1$. With $h_1 = h_2 = 1$, the process is ARIMA (1, 1, 0).

When $\phi = 1$, the trend is linear and the process is ARIMA (0, 2, 2):

$$(1 - B)^2 X_t = [1 - (2 - h_1 - h_2)B - (h_1 - 1)B^2]e_t. \tag{10}$$

When $\phi = 0$, we have simple smoothing and the equivalent ARIMA (0, 1, 1) process:

$$(1 - B)X_t = [1 - (1 - h_1)]e_t. \tag{11}$$

The ARIMA (0, 1, 0) process, or random-walk model, can be obtained from (11) by choosing $h_1 = 1$.

The general model can be extended to seasonal time-series. For example, the general model with multiplicative seasonality is

$$S_t = S_{t-1} + \phi T_{t-1} + h_1 e_t / I_{t-p}, \tag{12}$$

$$T_t = \phi T_{t-1} + h_2 e_t / I_{t-p}, \tag{13}$$

$$I_t = I_{t-p} + h_3 e_t / S_t, \tag{14}$$

$$\hat{X}_t(m) = (S_t + \sum_{i=1}^{m} \phi^i T_t) I_{t-p+m}. \tag{15}$$

There are p periods in one season, and the seasonal factors are denoted by $I_k, k = 1, 2, \ldots, p$. The seasonal factors are smoothed separately from the level and trend components with h_3. For the derivation of this model and a discussion of other forms of seasonal exponential smoothing, see Gardner and McKenzie.[8]

MODEL IDENTIFICATION PROCEDURE

Fitting the general model is a cumbersome way to identify a specific exponential smoothing model for a time-series. Another difficulty in automating the general model is that the time-series must be tested for seasonality in order to decide whether to use the seasonal or non-seasonal version.

At first glance, one alternative to fitting the general model is to examine the autocorrelations of the data in an attempt to identify an equivalent ARIMA process. However, this type of analysis is likely to be beyond the practitioners who use exponential smoothing. Furthermore, not all exponential smoothing models have equivalent ARIMA processes. For example, exponential smoothing models with multiplicative seasonality such as (12)–(15) are non-linear and thus have no equivalent ARIMA processes.[3]

For practical applications, a simple model identification procedure is needed. Our proposal is to examine the variances of relevant differences of the data. For a given time-series, six variances are computed, as shown in Table 1. For reasons of simplicity, we propose to use multiplicative seasonality whenever a seasonal model is indicated. Previous research[8] has found little difference in forecast accuracy between additive and multiplicative seasonal versions of exponential smoothing.

The rationale for the decision rules in Table 1 is straightforward. In case A, the variance of the original data X_t is the minimum. Certainly we do not wish to allow a trend or seasonal component in the model if differencing serves only to increase variance. The only reasonable model when this occurs is a constant level.

In case B, a non-seasonal difference of order 1 minimizes variance, which indicates the presence of some trend. The damped trend is recommended because it is equivalent to an ARIMA process with a difference of order 1, the (1, 1, 2), as discussed above. Another reason for using the damped trend is that case B is likely to occur more often in practice than case C, as discussed in Gardner and McKenzie.[5] The damped trend should be preferred because it is more robust. Although the constant-level model is also equivalent to an ARIMA process with a difference of order 1, the (0, 1, 1), we propose using a constant level only in cases where no differencing is required.

In case C, a non-seasonal difference of order 2 is necessary to minimize variance. In this case the trend is so strong that damping seems inadvisable. Furthermore, the linear trend is justified by its equivalence to an ARIMA process with a difference of order 2, the (0, 2, 2). One complication

TABLE 1. *Model identification rules*

Case	Series yielding minimum variance	Model selected
A.	X_t	Constant level
B.	$(1-B)X_t$	Damped trend
C.	$(1-B)^2X_t$	Linear trend
D.	$(1-B^p)X_t$	Constant level, seasonal
E.	$(1-B)(1-B^p)X_t$	Damped trend, seasonal
F.	$(1-B)^2(1-B^p)X_t$	Linear trend, seasonal

in case C is that an exponential trend is a possibility when the minimum-variance difference is of order 2. However, we consider an exponential trend to be a dangerous option in an automatic model-identification procedure.

For similar reasons, non-seasonal differencing beyond order 2 was not considered. Higher orders of differencing correspond to polynomial trends that appear to be of little use in practical time-series work. For example, a difference of order 3 suggests the ARIMA (0, 3, 3) process, which is equivalent to exponential smoothing with a quadratic trend.[4] The quadratic trend, also known as triple exponential smoothing, gave ridiculous results in the M-competition.

It may not be obvious that choosing the damped trend in case B does not always rule out the constant level and linear possibilities in cases A and C. The formulation of the damped trend is the same as the general model in equations (6)–(8). In some cases, model-fitting may result in $\phi = 0$, which is the same as a constant level, or $\phi = 1$, the same as a linear trend.

In cases D, E and F, a seasonal model is indicated because a seasonal difference reduces variance. The justification for the trend components of the seasonal models is similar to that above.

These models are all special cases of either the non-seasonal or seasonal versions of the general model discussed above. Thus we should expect that the identification procedure will give about the same forecasting results as fitting a version of the general model to every time series. The advantage of the identification procedure is that it should frequently yield more parsimonious models. For example, when a constant level is identified, only one parameter must be estimated, rather than three in the general model. When a linear trend is identified, only two parameters must be estimated.

We acknowledge that limiting the identification procedure to a set of only six models may overlook other models giving better fits to the data. For example, as the referee for this paper pointed out, if the time-series is ARMA (1, 1), then the variances of the differenced data are such that a constant-level model will be identified incorrectly. We should add to this comment that our identification procedure rules out all stationary models; instead, every time-series is smoothed. The justification is that the identification procedure must be simple and robust in order to handle large numbers of time-series in automatic forecasting applications. We do not recommend sole reliance on the procedure when it is feasible to perform a detailed analysis of the modelling possibilities for each time-series. When such analysis is feasible, models outside the exponential smoothing class should be considered.

Another problem raised by the referee is that the variance associated with double seasonal differencing, that is, the variance of $(1 - B^p)^2 X_t$, is not considered in identifying the linear-trend, multiplicative-seasonal model. We agree that this variance is important when enough data are available. Unfortunately, we do not have access to sufficient numbers of appropriate time-series on which to test the effects of double seasonal differencing. The problem is that two entire seasons of data are lost in the variance computation. In the Makridakis time-series discussed below, double seasonal differencing in many cases would have left only one or two seasons of data on which to make the trend decision.

EMPIRICAL RESULTS

As a test of the identification procedure, we used the sample of 111 series from the M-competition. This sample includes 20 annual, 23 quarterly and 68 monthly series. *Ex ante* forecasts from one time-origin were made, using hold-out samples of six periods for annual data, eight for quarterly and 18 for monthly. Table 2 shows mean absolute percentage error (MAPE) results, with comparisons to other modelling strategies for this data. For the general model and the linear trend, autocorrelations were tested to determine whether to use a seasonal or non-seasonal model. As expected, we found little difference in MAPE between the variance identification procedure and the general model.

The variance procedure resulted in a constant-level model for 29% of the series, a linear trend for 12% and a damped trend for 59%. Note that the damped trend has the same number of parameters as the general model. Thus simpler models than the damped trend were identified 41% of the time, a significant computational savings.

TABLE 2. *Post-sample MAPE comparisons, average of forecast horizons 1–18, 111 time-series*

Model identification strategy	MAPE
Parzen	15.4%
Variances of differences	15.6
General smoothing model	15.8
Box-Jenkins	18.0
Lewandowski	18.6
Holt's linear-trend model	19.5

The variance procedure identified 53 seasonal time-series amongst the 111. Autocorrelation analysis identified 60 seasonal series. Examination of the seven series 'missed' by the variance procedure showed that seasonality was either weak or erratic. In several cases, the series switched between seasonal and non-seasonal behaviour. It made little difference (less than 0.1% in *ex ante* MAPE) whether these series were counted as seasonal or non-seasonal.

Results for the Parzen, Box–Jenkins and Lewandowski identification strategies are from the M-competition. The variance procedure compares favourably in accuracy with the more sophisticated strategies. Compared to Parzen, the variance procedure gives about the same accuracy. Compared to Box–Jenkins and Lewandowski, the variance procedure gives much better accuracy.

Accuracy results by type of data or by forecast horizon are not presented here since they are very similar to previously reported results using the general model.[5] The variance identification procedure performed about the same as the general model in all situations. This was also true using the 1001 series from the M-competition.

CONCLUSIONS

As a forecasting methodology, exponential smoothing suffers from the lack of any objective procedure for model identification (for a review of the literature, see Gardner[9]). The consequence is that exponential smoothing is often misused, both in practice and in research. For example, studies of forecast accuracy have applied Holt's linear-trend model to every time-series in a collection, with no regard to whether a linear trend was in any sense appropriate for the data. Consider the study by Newbold and Granger,[10] who found that the accuracy of models identified through the Box–Jenkins approach was superior to Holt's linear-trend model. This study may well have reached different conclusions if smoothing models appropriate for each series had been selected.

Although much remains to be done, the model-identification procedure discussed in this paper is objective, and the trend components of the models selected are justified by equivalent ARIMA processes. The procedure includes only a limited range of models, but it is simple and efficient, making it suitable for large forecasting applications. In future work, we intend to refine the procedure by investigating its performance with other databases. We also plan to expand the range of models considered.

REFERENCES

1. R. G. BROWN (1983) *Smoothing, Forecasting, and Prediction of Discrete Time Series.* Prentice–Hall, Englewood Cliffs, NJ.
2. D. C. MONTGOMERY and L. A. JOHNSON (1976) *Forecasting and Time Series Analysis.* McGraw–Hill, New York.
3. E. MCKENZIE (1984) General exponential smoothing and the equivalent ARMA process. *J. Forecast.* **3,** 333–344.
4. G. E. P. BOX and G. JENKINS (1976) *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco, Calif.
5. E. S. GARDNER and E. MCKENZIE (1985) Forecasting trends in time series. *Mgmt Sci.* **31,** 1237–1246.
6. S. MAKRIDAKIS, A. ANDERSEN, R. CARBONE, R. FILDES, M. HIBON, R. LEWANDOWSKI, J. NEWTON, E. PARZEN and R. WINKLER (1982) The accuracy of extrapolation (time series) methods: results of a forecasting competition. *J. Forecast.* **1,** 111–153.
7. E. PARZEN (1982) ARARMA models for time series analysis and forecasting. *J. Forecast.* **1,** 67–82.
8. E. S. GARDNER and E. MCKENZIE (1988) Seasonal exponential smoothing with damped trends. *Mgmt Sci.* In press.
9. E. S. GARDNER (1985) Exponential smoothing: the state of the art. *J. Forecast.* **4,** 1–28.
10. P. NEWBOLD and C. W. J. GRANGER (1974) Experience with forecasting univariate time series and the combination of forecasts. *J. R. Statist. Soc. (A)* **137,** 131–165.