

Exponential Smoothing: The State of the Art

EVERETTE S. GARDNER, Jr.
*Navy Fleet Material Support Office, Mechanicsburg,
Pennsylvania, U.S.A.*

ABSTRACT

This paper is a critical review of exponential smoothing since the original work by Brown and Holt in the 1950s. Exponential smoothing is based on a pragmatic approach to forecasting which is shared in this review. The aim is to develop state-of-the-art guidelines for application of the exponential smoothing methodology. The first part of the paper discusses the class of relatively simple models which rely on the Holt-Winters procedure for seasonal adjustment of the data. Next, we review general exponential smoothing (GES), which uses Fourier functions of time to model seasonality. The research is reviewed according to the following questions. What are the useful properties of these models? What parameters should be used? How should the models be initialized? After the review of model-building, we turn to problems in the maintenance of forecasting systems based on exponential smoothing. Topics in the maintenance area include the use of quality control models to detect bias in the forecast errors, adaptive parameters to improve the response to structural changes in the time series, and two-stage forecasting, whereby we use a model of the errors or some other model of the data to improve our initial forecasts. Some of the major conclusions: the parameter ranges and starting values typically used in practice are arbitrary and may detract from accuracy. The empirical evidence favours Holt's model for trends over that of Brown. A linear trend should be damped at long horizons. The empirical evidence favours the Holt-Winters approach to seasonal data over GES. It is difficult to justify GES in standard form—the equivalent ARIMA model is simpler and more efficient. The cumulative sum of the errors appears to be the most practical forecast monitoring device. There is no evidence that adaptive parameters improve forecast accuracy. In fact, the reverse may be true.

KEY WORDS Bibliography—exponential smoothing Comparative methods—ARIMA, exponential smoothing Control charts—CUSUM Evaluation—forecast monitoring systems, exponential smoothing, adaptive Exponential smoothing—adaptive, coefficient choice, higher-order, review, theory Seasonality—estimation, harmonics Tracking signal—methodology Use—inventory control

Exponential smoothing methods are widely used in industry. Their popularity is due to several practical considerations in short-range forecasting. Model formulations are relatively simple.

0277-6693/85/010001-28\$02.80

© 1985 by John Wiley & Sons, Ltd.

Received October 1983

Revised August 1984

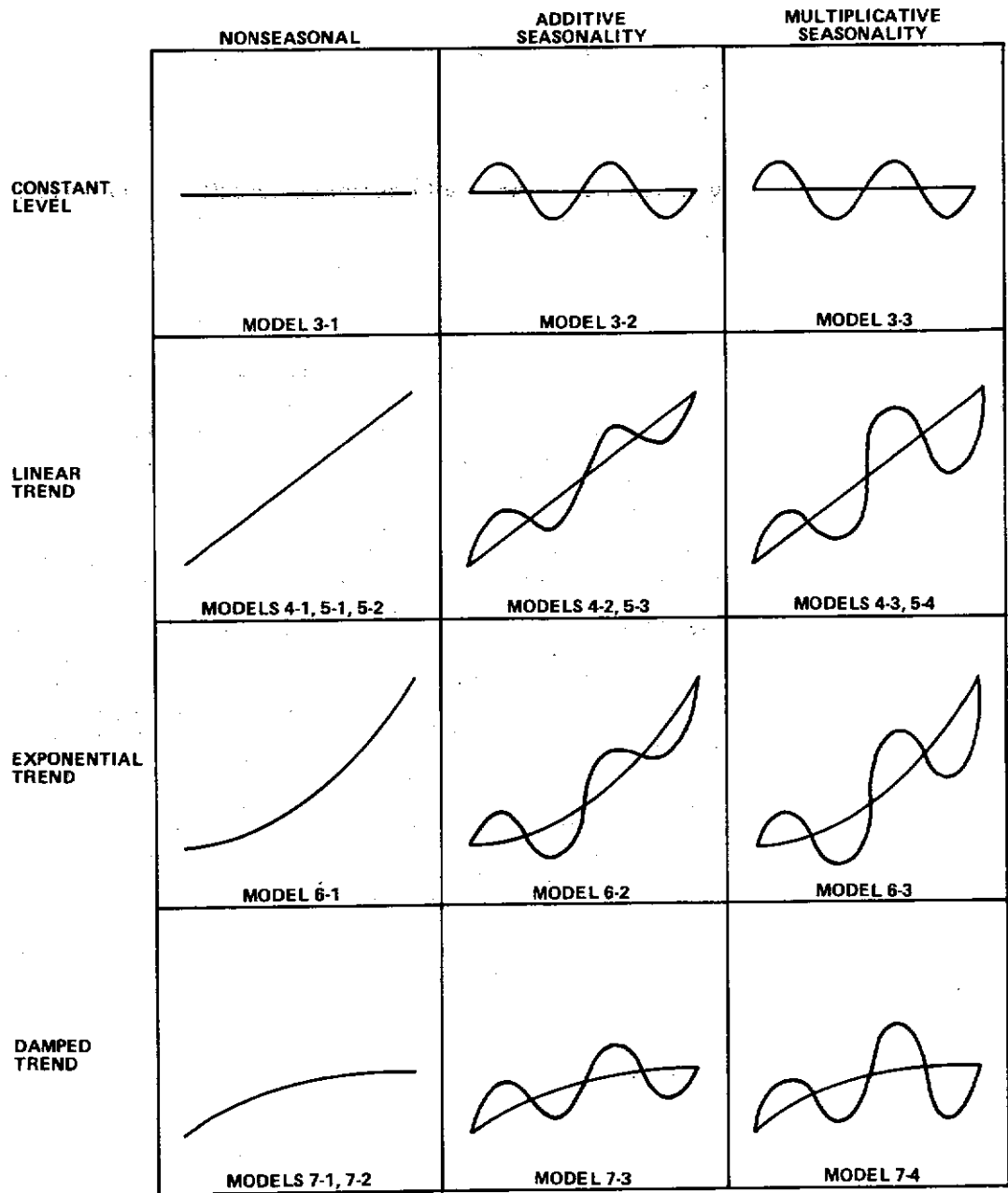


Exhibit 1. Forecast profiles from exponential smoothing

Symbol	Definition
α	Smoothing parameter for the level of the series
γ	Smoothing parameter for trend
δ	Smoothing parameter for seasonal factors
ϕ	Trend modification parameter
β	Discount factor, $0 \leq \beta \leq 1$
S_t	Smoothed level of the series, computed after X_t is observed. Also the expected value of the data at the end of period t in some models
T_t	Smoothed trend at the end of period t . Can be additive (linear) or multiplicative
I_t	Smoothed seasonal index or factor at the end of period t . Can be additive or multiplicative
S_t''	Double-smoothed average (from an application of simple exponential smoothing to S_t)
X_t	Observed value of the time series in period t
m	Number of periods in the forecast lead-time
p	Number of periods in the seasonal cycle
$\hat{X}_t(m)$	Forecast for m periods ahead from origin t
e_t	One-period-ahead forecast error, $e_t = X_t - \hat{X}_{t-1}(1)$. Note that $e_t(m)$ should be used for other forecast origins

Exhibit 2. Standard notation for exponential smoothing

Thus model components and parameters have some intuitive meaning to the user. Only limited data storage and computational effort are required. Tracking signal tests for forecast control are easy to apply.

Perhaps the most important reason for the popularity of exponential smoothing is the surprising accuracy that can be obtained with minimal effort in model identification. Two large-scale empirical studies have found little difference in forecast accuracy between exponential smoothing and ARIMA models identified by the Box-Jenkins (1976) methodology—see Makridakis and Hibon (1979) and Makridakis *et al.* (1982).

Despite the large body of research on exponential smoothing, there has never been a comprehensive review of the subject. This paper reviews the research since the original work by Brown and Holt in the 1950s. Sections 1-3 discuss the class of relatively simple models which rely on the Holt-Winters heuristic decomposition procedure for seasonal data. These models are appropriate in inventory control systems, when the noise component of the time series is relatively large, and when limited historical data rule out more sophisticated models.

The discussion of Holt-Winters is illustrated by Exhibits 1-7. Exhibit 1 shows examples of forecast profiles. There is no agreement in the literature on notation for Holt-Winters so Exhibit 2

Model	Recurrence form	Error-correction form
3-1 Non-seasonal	$S_t = \alpha X_t + (1 - \alpha)S_{t-1}$ $\hat{X}_t(m) = S_t$	$S_t = S_{t-1} + \alpha e_t$ $\hat{X}_t(m) = S_t$
3-2 Additive seasonals	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)S_{t-1}$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = S_{t-1} + \alpha e_t$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t + I_{t-p+m}$
3-3 Multiplicative seasonals	$S_t = \alpha(X_t/I_{t-p}) + (1 - \alpha)S_{t-1}$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t I_{t-p+m}$	$S_t = S_{t-1} + \alpha e_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t/S_t$ $\hat{X}_t(m) = S_t I_{t-p+m}$

Exhibit 3. Constant level models (simple smoothing)

Model	Recurrence form	Error-correction form
4-1 Non-seasonal	$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha \gamma e_t$ $\hat{X}_t(m) = S_t + mT_t$
4-2 Additive seasonals	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha \gamma e_t$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$
4-3 Multiplicative seasonals	$S_t = \alpha(X_t/I_{t-p}) + (1 - \alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t/I_{t-p}$ $T_t = T_{t-1} + \alpha \gamma e_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t/S_t$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$

Exhibit 4. Holt-Winters linear trend models

is proposed as a standard. Exhibits 3-7 contain model formulations corresponding to the forecast profiles in Exhibit 1. With a few exceptions, each model is written in two forms, a recurrence form and an error-correction form. The recurrence forms were used in the original work by Brown and Holt and still have pedagogic value. However, the error-correction forms are equivalent and generally easier to use.

Section 1 discusses simple smoothing (Exhibit 3) for a constant-level process. Section 2 discusses models for linear trends (Exhibits 4 and 5). Section 3 deals with non-linear trends (Exhibits 6 and 7). The research in each section is reviewed according to the following questions. What are the useful properties of each model? What parameters are recommended? How should the model be initialized?

Model	Recurrence form	Error-correction form
5-1* Non-seasonal	$S_t = \alpha X_t + (1 - \alpha)S_{t-1}$ $T_t = \alpha(S_t - S_{t-1}) + (1 - \alpha)T_{t-1}$ $\hat{X}_t(m) = S_t + \left(\frac{1 - \alpha}{\alpha}\right)T_t + mT_t$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha^2 e_t$ $\hat{X}_t(m) = S_t + \left(\frac{1 - \alpha}{\alpha}\right)T_t + mT_t$
5-2* Non-seasonal	$S_t = \alpha X_t + (1 - \alpha)S_{t-1}$ $S''_t = \alpha S_t + (1 - \alpha)S''_{t-1}$ $\hat{X}_t(m) = 2S_t - S''_t + m\left(\frac{\alpha}{1 - \alpha}\right)(S_t - S''_t)$	$S_t = S_{t-1} + T_{t-1} + \alpha(2 - \alpha)e_t$ $T_t = T_{t-1} + \alpha^2 e_t$ $\hat{X}_t(m) = S_t + mT_t$
5-3 Additive seasonals	N/A	$S_t = S_{t-1} + T_{t-1} + \alpha(2 - \alpha)e_t$ $T_t = T_{t-1} + \alpha^2 e_t$ $I_t = I_{t-p} + \delta[1 - \alpha(2 - \alpha)]e_t$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$
5-4 Multiplicative seasonals	N/A	$S_t = S_{t-1} + T_{t-1} + \alpha(2 - \alpha)e_t/I_{t-p}$ $T_t = T_{t-1} + \alpha^2 e_t/I_{t-p}$ $I_t = I_{t-p} + \delta[1 - \alpha(2 - \alpha)]e_t/S_t$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$

N/A, Not applicable.

Exhibit 5. Brown's linear trend models (* denotes equivalent models)

Model	Recurrence form	Error-correction form
6-1 Non-seasonal	$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} T_{t-1})$ $T_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)T_{t-1}$ $\hat{X}_t(m) = S_t T_t^m$	$S_t = S_{t-1} T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha \gamma e_t / S_{t-1}$ $\hat{X}_t(m) = S_t T_t^m$
6-2 Additive seasonals	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)S_{t-1} T_{t-1}$ $T_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)T_{t-1}$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t T_t^m + I_{t-p+m}$	$S_t = S_{t-1} T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha \gamma e_t / S_{t-1}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t T_t^m + I_{t-p+m}$
6-3 Multiplicative seasonals	$S_t = \alpha(X_t / I_{t-p}) + (1 - \alpha)S_{t-1} T_{t-1}$ $T_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)T_{t-1}$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = (S_t T_t^m) I_{t-p+m}$	$S_t = S_{t-1} T_{t-1} + \alpha e_t / I_{t-p}$ $T_t = T_{t-1} + (\alpha \gamma e_t / S_{t-1}) / I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t / S_t$ $\hat{X}_t(m) = (S_t T_t^m) I_{t-p+m}$

Exhibit 6. Exponential trend models

In Section 4, we review the general exponential smoothing methodology according to the same questions. General exponential smoothing differs from Holt-Winters in that Fourier functions of time are used to model seasonality. This introduces a considerable mathematical complexity which has been an obstacle in practical applications. However, recent research has done much to simplify general exponential smoothing.

In Sections 5-7, we turn to problems in the maintenance of forecasting systems based on exponential smoothing. These problems apply to both Holt-Winters and general exponential

Model	Recurrence form	Error-correction form
7-1 Non-seasonal	$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)\phi T_{t-1}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i$	$S_t = S_{t-1} + \phi T_{t-1} + \alpha e_t$ $T_t = \phi T_{t-1} + \alpha \gamma e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i$
7-2 Non-seasonal	N/A	$S_t = S_{t-1} + \phi T_{t-1} + \alpha(2 - \alpha)e_t$ $T_t = \phi T_{t-1} + \alpha(\alpha - \phi + 1)e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i$
7-3 Additive seasonals	N/A	$S_t = S_{t-1} + \phi T_{t-1} + \alpha(2 - \alpha)e_t$ $T_t = \phi T_{t-1} + \alpha(\alpha - \phi + 1)e_t$ $I_t = I_{t-p} + \delta[1 - \alpha(2 - \alpha)]e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i + I_{t-p+m}$
7-4 Multiplicative seasonals	N/A	$S_t = S_{t-1} + \phi T_{t-1} + \alpha(2 - \alpha)e_t / I_{t-p}$ $T_t = \phi T_{t-1} + \alpha(\alpha - \phi + 1)e_t / I_{t-p}$ $I_t = I_{t-p} + \delta[1 - \alpha(2 - \alpha)]e_t / S_t$ $\hat{X}_t(m) = \left(S_t + \sum_{i=1}^m \phi^i T_i \right) I_{t-p+m}$

N/A, Not applicable.

Exhibit 7. Damped trend models ($0 < \phi < 1$)

smoothing. Section 5 discusses quality control models for detecting bias in the forecast errors. A related topic, the use of adaptive parameters to improve the response rate of the forecasts to changes in the structure of the time series, is discussed in Section 6. Section 7 reviews various strategies for improvement of the forecasts, such as adjusting for autocorrelation in the errors and combining the forecasts from several different methods.

In Section 8, we evaluate the state of the art from a pragmatic point of view. The aim is to develop guidelines for the application of the exponential smoothing methodology. We also suggest some new directions for future research.

1. MODELS FOR A CONSTANT-LEVEL PROCESS (SIMPLE SMOOTHING)

Model formulations for a constant-level process (Exhibit 3) are explained in Section 1.1. Properties are discussed in Section 1.2. In Sections 1.3 and 1.4, we review recommended parameters and starting values. Two important extensions of constant-level models, to multivariate and intermittent time series, are discussed in Sections 1.5 and 1.6.

1.1. Model formulations

Simple smoothing represents the time series by $X_t = b + \varepsilon_t$ where ε_t is a random component with mean zero and variance σ_ε^2 . The level b is assumed to be constant in any local segment of the series but may change slowly over time. In Model 3-1, the statistic S_t is an unbiased estimator of the level as well as the forecast for any period ahead. Models 3-2 and 3-3 adjust for additive and multiplicative seasonality, respectively. The two seasonal models give much the same forecasts when the level of the series is stable. But if the level changes and the seasonal fluctuations change proportionally, Model 3-3 is more appropriate.

1.2. Properties of simple smoothing

Although the simple smoothing model is a weighted moving average, it is possible to derive a smoothing parameter which gives approximately the same forecasts as an unweighted moving average of any given number of periods. For details see Brown (1959). This relationship has led some researchers to conclude that simple smoothing has no important advantage in accuracy (Adam, 1973, with corrections by McLeavey *et al.*, 1981; Armstrong, 1978; Elton and Gruber, 1972; Kirby, 1966). However, Makridakis *et al.* (1982) found that simple smoothing was significantly more accurate than the unweighted moving average in a sample of 1001 time series.

When the sample size is large, simple smoothing is optimal with respect to a discounted-least-squares (DLS) criterion, with discount factor $\beta = 1 - \alpha$. Muth (1960) was the first of many to prove that simple smoothing is optimal for the ARIMA (0, 1, 1) process:

$$(1 - B)X_t = (1 - \theta B)\varepsilon_t \quad (1)$$

The condition for optimality is $\theta = 1 - \alpha$. A similar process for which simple smoothing is optimal is the random walk:

$$X_t^* = X_{t-1}^* + \varepsilon_{2,t} \quad (2)$$

$$X_t = X_t^* + \varepsilon_{1,t} \quad (3)$$

X_t is the observed value, X_t^* is the 'true' level of the series which is unobserved, and the two error terms are generated by independent white noise processes. Harrison (1967), Nerlove and Wage

(1964), and Theil and Wage (1964) showed that simple smoothing is optimal with α determined by the ratio of the variances of the noise processes. Further interpretation of this result is given by Harvey (1984b), who pointed out that the Kalman filter for (2) and (3) reduces to simple smoothing in the steady state.

These are the only optimal properties of simple smoothing although robustness has been predicted by other research. Cogger (1973b), Cohen (1963), Cox (1961) and Pandit and Wu (1974) argued that more complex models may not yield significantly smaller errors on some series not of the ARIMA (0, 1, 1) type, provided that one-period-ahead forecasting is the only concern. Such series include first-order autoregressive processes (Cohen, Cox) and certain lower-order ARIMA processes (Cogger, Pandit and Wu). Bossons (1966) also argued that exponential smoothing can be relatively insensitive to specification error, especially when the mis-specification arises from an incorrect belief in the stationarity of the generating process.

Robustness was supported by Makridakis *et al.* (1982) (see also Makridakis, 1983). Simple smoothing was the best overall choice for one-period-ahead forecasting, considering 24 time series methods and a variety of accuracy measures, such as mean absolute percentage error (MAPE), the average ranking of each method; and the mean squared error (MSE).

More evidence of robustness is given by the simulation study of Gross and Craig (1974). Simple smoothing and Bayesian forecasting were used to estimate the means of Poisson demand series, which are frequently encountered in inventory systems. Overall there was little difference in accuracy. However, simple smoothing was superior when the series contained step changes in the mean. This is surprising because the major claims for success of Bayesian methods are based on time series of this nature.

1.3. Parameter selection

In certain inventory problems, it is possible to derive a smoothing parameter which is optimal in the sense that it minimizes the costs of stock replenishment policy. Examples are given by Adelson (1966), Howe (1974), Landi and Johnson (1967), and Trigg and Pitts (1962). This line of research depends on highly restrictive assumptions and may be difficult to implement. For example, the only costs considered by Landi and Johnson are due to fluctuations in inventory and order levels.

In practice, the smoothing parameter is usually chosen by a grid search to minimize the *ex post* MSE. This procedure can be justified as an approximation to the exact maximum likelihood estimate of the parameter via the Kalman filter (see Harvey, 1984b).

Most of the research on simple smoothing has assumed a range of 0-1 for α , although a more restricted range of 0.10-0.30 is typical in practice. It is widely held that a more complex model should be entertained if the best α value falls above 0.30 during the model-fitting process. For an example of this argument, see Montgomery and Johnson (1976).

There is no evidence to support such a restricted range of parameters. Both theory and empirical work suggest that a wider range of parameters should be considered. Harrison's (1967) analysis of serial variation functions showed that underestimation of the optimal parameter is always more serious than overestimation. A frequently overlooked property of Model 3-1 is that it is equivalent to a difference equation which is stable in the range $0 < \alpha < 2$ (Brenner *et al.*, 1968). Another way to justify a wider range for α is to recognize that the ARIMA (0, 1, 1) process is invertible in the range $0 < \alpha < 2$.

In the Makridakis *et al.* study (1982), α values above 0.3 were frequently estimated during the model-fitting process. Although Makridakis did not consider Models 3-2 and 3-3, large δ values in the linear trend models (Exhibit 4) were also found. A more limited study by Chatfield (1978) also found relatively large parameter values. These studies show that it is dangerous to guess at values of the smoothing parameters. The parameters should be estimated from the data.

Furthermore, it is desirable to subject the parameters to *ex ante* testing. There is ample evidence (Fildes and Howell, 1979; Fildes, 1979) that *ex post fit* has little correlation with *ex ante* accuracy.

1.4. Starting values

Starting values (as of time 0) for S and I are required by the models in Exhibit 3. Methods for computing S_0 have been developed by a number of researchers. There appears to be no empirical evidence favouring any particular method. Brown's (1959) original suggestion, simply using the mean of the data for S_0 , is popular in practice. Ledolter and Abraham (1984) recommended backcasting to obtain S_0 , since this leads to the same forecasts as the ARIMA (0, 1, 1) model estimated by unconditional least squares. Backcasting is done by reversing the time order of the data and using the most recent data point to start the smoothing process.

$$S_t^* = S_{t+1}^* + ae_t \quad (4)$$

When the beginning of the series is reached, S_0^* is used as S_0 in Model 3-1.

When only a few data points are available, it can be difficult to choose a starting value. Gilchrist (1967, 1976) proposed using an exact DLS formulation for S_t rather than the approximation in Model 3-1. The advantage is that no estimate of starting values is needed. One of many equivalent DLS formulations for S_t is:

$$S_t = \frac{\sum_{r=0}^{t-1} \beta^r X_{t-r}}{\sum_{r=0}^{t-1} \beta^r} \quad (5)$$

S_t can be estimated recursively with:

$$N_t = X_t + \beta N_{t-1} \quad (6)$$

$$D_t = 1 + \beta D_{t-1} \quad (7)$$

$$S_t = N_t / D_t \quad (8)$$

N_0 and D_0 are always zero. Model 3-1 is an approximation to the DLS formulation. In the limit, the two are equivalent. Thus one could switch to the simpler Model 3-1 after enough data have been collected to fit the model. Similar ideas for eliminating the need to estimate starting values are discussed by Cogger (1973a), McClain (1981), Taylor (1981) and Wade (1967).

Another alternative with a limited number of data points is to use Bayesian methods to combine a prior estimate of the level with an average of the available data—see Cohen (1966), Johnson and Montgomery (1974) and Taylor (1981).

A linear regression on dummy variables can be used to obtain starting values for the additive seasonal factors. Heuristic algorithms for estimating starting values for both additive and multiplicative seasonality can be found in Johnson and Montgomery (1974), Montgomery and Johnson (1976) and Winters (1960). Classical time series decomposition methods (see Makridakis and Wheelwright, 1978) are more objective than the heuristics and require about the same computational effort. The seasonal factors from a decomposition correspond directly to the I_t values in Models 3-2 or 3-3. One of the methods for estimating S_0 can then be applied to the deseasonalized data.

1.5. Extensions to multivariate time series

The models in Exhibit 3 treat each time series independently but there are many applications in which a multivariate model would be more appropriate. An example is an inventory control system in which the products are competitive. Simple smoothing has been generalized to multivariate

forecasting by Jones (1966) and Enns *et al.* (1982). The generalization is straightforward. We replace the scalars in Model 3-1 with:

$$S_t = S_{t-1} + \alpha e_t \quad (9)$$

If there are k series, the dimensions of S_t , S_{t-1} , and e_t are $k \times 1$. The smoothing matrix α has dimension $k \times k$.

Enns *et al.* assume that the time series are produced by a multivariate generalization of the random walk, equations (2) and (3) above. The parameters are estimated by maximum likelihood, further assuming that the variances of the disturbance terms are proportional. The algorithm requires the concentrated likelihood function corresponding to (9) to be computed. If there are N observations on each series, this involves construction of the covariance matrix of Nk observations and the diagonalization of a certain $N \times N$ matrix. Recent work by Harvey (1984a) showed that this complexity is unnecessary. Using the same assumptions as Enns *et al.*, Harvey showed that the concentrated likelihood function can be expressed in terms of quantities which make up the likelihood function for the individual series.

The practical result of this insight is that, given the Enns *et al.* assumptions, one can use univariate smoothing models to forecast related time series. The univariate parameters can be chosen by a grid search to minimize the sum of the vector products of the one-period-ahead errors ($\sum e_t e_t'$). Such parameters approximate those obtained via the concentrated likelihood function. Harvey also gives exact maximum likelihood procedures based on univariate application of the Kalman filter. Harvey's results hold for smoothing models containing polynomial trends and seasonal components as well. In either the approximate or exact case, the simplification compared to Enns *et al.* is considerable.

1.6. Extensions to intermittent time series

Another important extension of simple smoothing is to series which are observed intermittently. This is a common situation in production and inventory systems. For example, demand at the retail level may be filtered through a distributor before it affects the production system. The distributor usually replenishes stocks in lots, which makes the demand on production intermittent. If simple smoothing is applied to an intermittent demand series, the forecasts are wildly biased. We underestimate the size of individual demand transactions and overestimate the long term average demand.

Simple smoothing has been modified for intermittent series by Croston (1972) with corrections by Rao (1973). See also Peterson and Silver (1979) for a discussion of applications of the Croston model. The basic idea is to smooth two components of the demand process separately: the size of each demand transaction and the time between consecutive transactions. When a demand transaction occurs, we smooth each component. The forecast for the next period is the ratio of the smoothed size to the smoothed time between transactions. If a time period passes with no demand, we do not change the previous estimates. The forecasts are unbiased if the probability of occurrence of a transaction follows a Bernoulli process.

Problems similar to those described above occur when the data happen to be collected at irregular time intervals or when some data are missing from the series. Modifications to exponential smoothing for these situations are available in Tydeman (1972) and Wright (1983).

2. MODELS FOR LINEAR TRENDS

Several alternative model formulations are commonly used for linear trends, as explained in Section 2.1. The properties of these formulations are reviewed in Section 2.2. Parameter selection and recommended starting values are discussed in Sections 2.3 and 2.4.

2.1. Model formulations

A linear trend process is $X_t = b_0 + b_1t + \varepsilon_t$. The simple smoothing model will lag this process for infinite time. The lag or bias in the forecasts will eventually stabilize at an expected value of $b_1(1 - \alpha)/\alpha$ (Brown, 1963). There are several ways to adjust for the lag in simple smoothing.

Model 4-1, due to Holt *et al.* (1960) and Winters (1960), uses separate parameters to smooth the level and trend of the series. The Brown models in Exhibit 5 use a single parameter to smooth both components. The Holt-Winters models were developed by heuristic reasoning, whereas the Brown models yield a DLS with discount factor $\beta = 1 - \alpha$. Models 5-1 and 5-2 are equivalent and both are special cases of Holt-Winters. This can be seen by comparing the error-correction form of Model 4-1 to that of Model 5-2.

The seasonal Holt-Winters Models 4-2 and 4-3 require three parameters, which can be cumbersome in large forecasting systems. In Exhibit 5, the seasonal models require only two parameters. This is achieved by adding the Holt-Winters seasonal updating procedure to the Brown Model 5-2. The result is a DLS on the seasonally adjusted data. Only the error-correction forms are given for the seasonal models in Exhibit 5—the recurrence forms are too inefficient for practical use.

2.2. Properties

Model 4-1 is optimal for two generating processes. Harrison (1967), Nerlove and Wage (1964), and Theil and Wage (1964) proved optimality for a random walk with a linear growth term:

$$T_t^* = T_{t-1}^* + \varepsilon_{3,t} \quad (10)$$

$$X_t^* = X_{t-1}^* + T_{t-1}^* + \varepsilon_{2,t} \quad (11)$$

$$X_t = X_{t-1}^* + \varepsilon_{1,t} \quad (12)$$

The optimal smoothing parameters are determined by the relative variances of the three white noise processes. Harvey (1984b) showed that the steady-state Kalman filter for (10)–(12) reduces to Model 4-1.

The same authors also proved optimality for the ARIMA (0, 2, 2) process:

$$(1 - B)^2 X_t = (1 - \theta_1 B - \theta_2 B^2) \varepsilon_t \quad (13)$$

The moving average parameters are related to the smoothing parameters as follows: $\theta_1 = 2 - \alpha - \alpha\gamma$ and $\theta_2 = \alpha - 1$.

Brown's Model 5-1 is optimal for the ARIMA (0, 2, 2) process but for a smaller subclass than Holt-Winters. The Brown model is equivalent to an equal-root ARIMA (0, 2, 2) model (Cogger, 1974).

Thus the Holt-Winters model is somewhat more general than Brown's. Does the added generality significantly improve forecast accuracy? The answer depends on the generating process. Harrison (1967) derived the variance of the forecast errors for the ARIMA (0, 2, 2) process and concluded that the maximum penalty for the use of Brown's model would be an increase in the standard deviation of the errors (one-period-ahead) of less than 2 per cent.

In empirical studies involving other generating processes, Holt-Winters has generally had a wider margin in accuracy over the Brown model—see Gardner and Dannenbring (1980), Makridakis and Hibon (1979) and Makridakis *et al.* (1982). The reason is the additional flexibility of the two-parameter model when applied automatically to a large number of series. For example, if a series contains no trend, the corresponding Holt-Winters parameter will be fitted at a level near zero and the model will behave much like the simple smoothing model. But the Brown model will always extrapolate some trend from any series.

Turning to the seasonal models, the additive Model 4-2 is optimal for a certain ARIMA process derived by McKenzie (1976a). However, this generating process is so complex that there is no point in reviewing it here—see equation (16) of McKenzie (1976a). As Chatfield (1977) observed, the additive version of Holt–Winters would never be identified through Box–Jenkins procedures. The multiplicative Model 4-3 does not appear to have an ARIMA equivalent. Thus the seasonal Holt–Winters models are not special cases of the class of ARIMA models recommended by Box and Jenkins.

Both the Holt–Winters and Brown models were robust at short horizons in the Makridakis *et al.* study (1982). At longer horizons (more than three or four periods ahead) both models had a tendency to overshoot the data. As discussed in Section 3, it may be advisable to damp a linear trend at long horizons.

2.3. Parameter selection

Moderate parameters, generally less than 0.3, have been recommended for the Holt–Winters models in Exhibit 4. For the Brown models, α values of 0.2 or less are generally accepted. Examples of these recommendations may be found in Brown (1963), Coutie *et al.* (1964), Harrison (1967) and Montgomery and Johnson (1976).

Although these parameter ranges have been criticized as arbitrary (Chatfield, 1978), they are appropriate in inventory control applications where forecasts are generated automatically. It is important to detect biased errors as quickly as possible in inventory control. Quick detection increases the lead time available to adjust the flow of material into stock. Gardner (1984) showed that using moderate parameters makes it easier to detect biased errors, regardless of the tracking signal used. The reason is that a trend-adjusted model has a tendency to overshoot sudden changes in the series, such as a step increase in level. Before the model catches up to the new level, the forecast errors are negative in sign; after the overshoot, the errors are positive which confounds the tracking signal. It should be noted that the simple smoothing parameter has no effect on bias detection with most tracking signals (Gardner, 1983a). Tracking signals are discussed further in Section 5.

In other applications, a wider range of parameters should be considered. Makridakis *et al.* (1982) and Chatfield (1978) found that the most accurate parameters were frequently in the range 0.3–1. Should parameters above 1.0 be considered? Theory suggests that they should. McClain and Thomas (1973) showed that the non-seasonal Model 4-1 is stable (invertible) over the range $0 < \alpha < 2$ and $0 < \gamma < (4 - 2\alpha)/\alpha$.

The search for parameters within the region of stability can be narrowed by avoiding areas of oscillation. The largest value of α in Model 4-1 which does not lead to oscillation is given by $\alpha < 4\gamma/(1 + \gamma)^2$ (McClain and Thomas, 1973; McClain, 1974). There is no evidence that restricting parameters in this manner improves accuracy. Makridakis *et al.* (1982) for example had great success with oscillatory parameters.

The Brown Model 5-1 is stable for $0 < \alpha < 2$. The model is also critically damped. This means that one obtains the fastest possible response to a structural change in the time series without oscillation. Again, it is not clear that this property is relevant to forecast accuracy.

Much like simple smoothing, Harrison's (1967) analysis of serial variation functions showed that underestimation of the optimal parameters in Model 4-1 is always more serious than overestimation. He also found that the variance of the forecast errors is more sensitive to departures from the optimum in α than γ . McClain and Thomas (1973) reached the same conclusions.

Work on the regions of stability of seasonal model parameters has been done by Brenner *et al.* (1968) and McClain (1974). Unfortunately, the regions are so complex that it is difficult to

generalize from this work. Using computational methods, Sweet (1983a, b) reached the following conclusions on the parameters for seasonal models (both additive and multiplicative): if the length of the seasonal cycle is four periods, the model is always stable for parameters between 0 and 1. If the cycle is 12 periods, the model is not necessarily stable for parameters in this range. The conditions for stability are complex. Sweet gives procedures for checking the stability of any set of parameters.

A grid search can be used to find the parameter set which minimizes the fitted MSE although sophisticated search procedures such as the Hooke-Jeeves pattern search algorithm are more efficient. Examples favouring Hooke-Jeeves are available in Berry and Bliemel (1974) or Flowers (1980). A FORTRAN IV program for the Hooke-Jeeves algorithm is also available in Buffa and Taubert (1972).

Does the forecasting horizon used during the model-fitting process make any difference in *ex ante* accuracy? For example, should the one-period-ahead MSE be used to select parameters when the *ex ante* horizon will be longer? A simulation study by Dalrymple and King (1981) found that the horizon used during model-fitting made no difference in *ex ante* accuracy. Makridakis *et al.* (1982) reached the same conclusion.

2.4. Starting values

Starting values for the linear trend models can be obtained with backcasting although this procedure should be used with caution. If the trend is erratic, it is easy to generate a negative starting value for S_0 . Problems with backcasting can be avoided by using ordinary least squares estimates of the level and trend. Decomposition methods can be used to start the seasonal factors. When the historical data are limited, exact DLS models can be used as with simple smoothing—see Cogger (1973a), Gilchrist (1976) and McClain (1981).

3. MODELS FOR NON-LINEAR TRENDS

The linear models in Exhibits 4 and 5 can be extended in several different ways to accommodate non-linear trends. In Section 3.1, we deal with exponential trends, which are dangerous at long horizons but may be useful for short-range forecasting early in the product life cycle. In Section 3.2, we discuss damped exponentials. This form of trend was the most accurate at long horizons in the Makridakis *et al.* study (1982). Methods for smoothing polynomial trends of any order are reviewed in Section 3.3. Since the polynomial models are of little practical interest, formulations are not included in the exhibits.

3.1. Exponential trends

An exponential trend is generated by $X_t = b_0 b_1^t$ with $b_1 > 0$. Several smoothing models have been suggested for this process—see Pegels (1969), Brenner *et al.* (1968) and Roberts (1982). The simplest approach is that of Pegels in Exhibit 6. The local slope of the process is estimated by smoothing successive ratios of the level of the series (S_t/S_{t-1}). No analysis or empirical research has been reported for this model, although least squares estimates of the level and slope together with a decomposition provide obvious starting values.

3.2. Damped exponential trends

The first model in Exhibit 7 adds an autoregressive parameter ϕ to the Holt-Winters linear Model 4-1. If $\phi = 1$, Models 7-1 and 4-1 are identical. If $\phi > 1$, the growth in the forecasts is exponential (but this model requires one more parameter than Pegels). If $\phi < 1$, the growth has a damped

exponential form, declining in both relative and absolute terms each period. The difference between a linear and a damped trend can be substantial at long horizons, even with a relatively large ϕ of say 0.9 or 0.95.

In Model 7-2 there is one less parameter than in Model 7-1. This is achieved by modifying Brown's linear model. If $\phi = 1$, Models 7-2 and 5-2 are identical. Model 7-2 is extended to seasonal data in Models 7-3 and 7-4. Both seasonal models require three parameters. Model 7-1 is not extended to seasonal data since four parameters would be needed.

Although the parameters in Model 7-2 are simple and intuitively reasonable, they do not satisfy a DLS criterion unless $\phi = 1$. The general DLS solution for a damped exponential trend is: $h_1 = [1 - (\beta/\phi)^2]$ and $h_2 = [1 - (\beta/\phi)][1 - (\beta/\phi)^2]$, where h_1 and h_2 are the parameters for the level and trend, respectively, and the discount factor is $\beta = 1 - \alpha$. This solution follows directly from the work of McKenzie (1976b).

Roberts (1982) showed that Model 7-1 is optimal for the ARIMA (1, 1, 2) process:

$$(1 - \phi B)(1 - B)X_t = (1 - \theta_1 B - \theta_2 B^2)\epsilon_t \quad (14)$$

The conditions on the parameters are $\theta_1 = 1 + \phi - \alpha - \phi\alpha\gamma$ and $\theta_2 = -\phi(1 - \alpha)$.

If ϕ is unknown, which is almost always the case, ordinary least squares cannot be used to estimate starting values for the damped exponential models. Thus backcasting is the most practical way to obtain starting values for level and trend. Again, decomposition can be used to start the seasonal factors.

No direct empirical evidence is available on the damped exponential models, although some indirect evidence is available from the performance of Lewandowski (1982) time series methods in the Makridakis *et al.* (1982) study. Lewandowski used a damped trend like the models in Exhibit 7 on most time series. The rate of decay in the trend increased with the noise in the series. This strategy was far more accurate than competing time series methods at long horizons.

3.3. Polynomial trends

An n th-degree polynomial trend process is written:

$$X_t = b_0 + b_1 t + \frac{1}{2} b_2 t^2 + \dots + (1/n!) b_n t^n \quad (15)$$

The $n + 1$ coefficients can be estimated by taking linear combinations of the first $n + 1$ orders of exponential smoothing. This result is known as the fundamental theorem of exponential smoothing (Brown and Meyer, 1961). D'Esopo (1961) proved that coefficients estimated in this manner are optimal with respect to a DLS criterion with $\beta = 1 - \alpha$.

To illustrate, the orders of smoothing are designated by $S_t, S_t', \dots, S_t^{n+1}$. In Model 3-1, S_t is the first order of smoothing and estimates the level of a constant series (a zero-degree polynomial). In Model 5-2, S_t is smoothed to yield S_t' and they are combined to estimate the coefficients of a linear trend process (a first-degree polynomial). Continuing in the same fashion, linear combinations of the first three orders of smoothing would estimate the coefficients in a quadratic trend process.

Polynomial trend models above the first degree are of little practical interest in business and economic forecasting. These models have some undesirable feedback properties, particularly a tendency to amplify noise in the series (Morris and Glassey, 1963). In the Makridakis *et al.* (1982) study, quadratic smoothing was unstable and perhaps the worst time series method tested. Numerous authors have also questioned the need for higher-order polynomial models because of the excessive differencing implied in the equivalent ARIMA process. In general, exponential smoothing of order n is optimal for the ARIMA (0, n , n) process (Cogger, 1974; Godolphin and Harrison, 1975; Goodman, 1974; Ledolter and Box, 1978; McKenzie, 1974). Thus quadratic smoothing implies the ARIMA (0, 3, 3) process which is rarely, if ever, observed in practice.

It should be mentioned that one empirical study (Markland, 1970) found that quadratic smoothing was more accurate than Brown's linear model, using a sample of extremely volatile series (sudden changes in level and trend) from an inventory system. It is difficult to generalize from this study because the results appear to be distorted by intermittent observations of the data. There is also some question as to whether reasonable alternatives to the quadratic model were considered. For example, the series were non-seasonal although a seasonal version of Holt-Winters was compared to the quadratic model.

4. GENERAL EXPONENTIAL SMOOTHING

General exponential smoothing (GES), direct smoothing and adaptive smoothing are terms used interchangeably to describe the use of DLS to fit certain functions of time to the data. The functions considered are polynomials, exponentials, sinusoids, and their sums and products. Such functions are sufficient to model any ARIMA process. The main difference from the Box-Jenkins approach is that there is little emphasis on identification. The main difference from Holt-Winters is the use of sinusoids to model seasonality.

In Section 4.1 we summarize how GES models are formulated and estimated. In Sections 4.2-4.4, properties, recommended starting values and parameters are reviewed. Several extensions of the GES methodology, designed to improve accuracy and streamline the calculations, are discussed in Sections 4.5 and 4.6. The empirical evidence comparing GES to Holt-Winters is reviewed in Section 4.7. The GES notation is from McKenzie (1976a).

4.1. Model formulation and estimation

A GES model is formulated as a multiple linear regression. The forecast equation is:

$$\hat{X}_t(m) = \sum_{i=1}^n a_i(t) f_i(m) = \mathbf{a}' \mathbf{f}(m) \quad (16)$$

The vector of coefficients is \mathbf{a} . The vector \mathbf{f} is composed of known functions of time. The only allowable functions are those mentioned above, usually called 'fitting functions'.

Details of estimation procedures for \mathbf{a} are thoroughly discussed by Brown (1963). Here we state only the criterion for \mathbf{a} and the solution. \mathbf{a} is chosen to minimize

$$\sum_{j=0}^t \beta^j [X_{t-j} - \mathbf{a}' \mathbf{f}(-j)]^2 \quad (17)$$

The solution requires that we smooth the model parameters rather than the components of the time series as in the Holt-Winters approach. The error-correction form of the solution is:

$$\mathbf{a}_t = \mathbf{L}' \mathbf{a}_{t-1} + \mathbf{h} \mathbf{e}_t \quad (18)$$

\mathbf{L} is a constant square matrix dependent only on the fitting functions and defined such that $\mathbf{f}_t = \mathbf{L} \mathbf{f}_{t-1}$. The smoothing vector is \mathbf{h} defined by $\mathbf{h} = \mathbf{F}^{-1} \mathbf{f}(0)$. Vector \mathbf{h} depends on both the fitting functions and the discount factor β . The vector $\mathbf{f}(0)$ is composed of the fitting functions at the time origin. For large t , \mathbf{F} is given by:

$$\mathbf{F}_t = \mathbf{F} = \sum_{j=0}^{\infty} \beta^j \mathbf{f}(-j) \mathbf{f}'(-j) \quad (19)$$

Explicit expressions for \mathbf{h} have been derived in some cases (Brown, 1967, 1982; Dobbie, 1963; McKenzie, 1976a; Sweet, 1981). Usually, \mathbf{h} is obtained computationally. FORTRAN programs

for computing \mathbf{h} for additive seasonal patterns are available in Abraham and Ledolter (1983), and Groff and Muth (1972). Tables of \mathbf{L} , \mathbf{f} , \mathbf{F}^{-1} and \mathbf{h} , mostly for additive seasonal patterns, are available in Brown (1963, 1967, 1977, 1982) and Montgomery and Johnson (1976).

If the data are non-seasonal, there is no reason to use GES. An equivalent and simpler model from the Holt-Winters class can always be formulated. For example, it can be shown (Brown, 1963) that GES for a polynomial model is equivalent to multiple smoothing (Section 3.3) with $\alpha = 1 - \beta$.

If the data are seasonal, GES differs considerably from Holt-Winters. The seasonal terms in GES are coefficients of cosine and sine functions, whereas in Holt-Winters the seasonal terms are indexes of the typical level of demand each period. Although it is difficult to know how many cosine/sine terms are required, we can state a rule (Brown, 1982) for the maximum number of terms (based on the highest frequency that could be observed in the data). Assume that the seasonal pattern goes through a complete cycle in 1 year. If the data are monthly, at most 11 seasonal terms are ever needed. If the data are quarterly, at most three terms are needed. To illustrate, assume that the seasonal pattern is additive and the data are monthly. The seasonal terms are:

$$\begin{aligned} & a_3 \cos \omega t + a_4 \sin \omega t + a_5 \cos 2\omega t + \\ & a_6 \sin 2\omega t + a_7 \cos 3\omega t + a_8 \sin 3\omega t + \\ & a_9 \cos 4\omega t + a_{10} \sin 4\omega t + a_{11} \cos 5\omega t + \\ & a_{12} \sin 5\omega t + a_{13} \cos 6\omega t \end{aligned}$$

The terms are numbered starting with a_3 because we usually reserve a_1 and a_2 for the level and trend. ω is defined as $2\pi/p$, where p is the number of periods in the seasonal cycle. It may seem that we should end the series with a sine term but this always turns out to be zero when p is even and can be neglected.

If the seasonal pattern is multiplicative, note that it is necessary to multiply each cosine or sine term by the trend. This makes the model far more complex, which may be the reason that only additive seasonal patterns are typically used in practice. It is interesting that multiplicative seasonal patterns are typically used in the Holt-Winters models.

4.2. Properties

GES has two advantages compared to Holt-Winters. First, the use of a single parameter β in the range 0-1 means that the forecast errors always have finite variance. Secondly, the structure of GES is such that all seasonal terms are revised with each observation. This should make the forecasts more responsive to changing seasonal patterns.

McKenzie (1976b) showed that GES is optimal for the ARIMA process:

$$\phi(B)X_t = \theta(\beta B)\epsilon_t \quad (20)$$

in which $\theta(B) = 1 - \theta_1 B - \dots - \theta_n B^n$ and $\phi(B) = 1 - \phi_1 B - \dots - \phi_n B^n$. If there are no exponentials in the model, ϕ can be substituted for θ on the right hand side of (20) and the optimal process becomes:

$$\phi(B)X_t = \phi(\beta B)\epsilon_t \quad (21)$$

It is not clear that the equivalent ARIMA model could always be identified from the data. Abraham and Ledolter (1983) argued that many GES models would not be identified because the use of DLS to fit the models forces unreasonable restrictions on the ARIMA parameters. Roberts (1982) argued to the contrary.

Whatever the outcome of this dispute, the problem of identification is not important here. The practical implication of McKenzie's work is this: once a GES model has been selected, the ARIMA

equivalent is often more efficient. The reason is that \mathbf{a}_t can be revised with explicit linear equations, avoiding the need to store the transition matrix \mathbf{L} . Many of the components of the fitting functions turn out to be zero, making an explicit expression for the forecasts more efficient. See McKenzie (1976b, 1984) for examples comparing the computational efficiency of GES models and ARIMA equivalents.

4.3. Parameter selection

Brown (1977) recommended β in the range 0.75 ('fast smoothing' for shorter time series) to 0.90 (normal smoothing). Once β has been selected, we still require the components of \mathbf{h} , which may not be available explicitly. However, McKenzie showed that the equivalent ARIMA model usually yields an explicit expression for \mathbf{h} .

4.4. Starting values

Starting values in GES can be obtained with a least-squares fit of a multiple linear regression model to the available data. This can be done in two ways. One is to set the time origin at period 1 and use the regression coefficients as \mathbf{a}_0 . Then we smooth the data until the current value of \mathbf{a}_t is reached. An alternative is to set the time origin at the most recent data point in the regression model. The coefficients then correspond directly to \mathbf{a}_t .

In any event, \mathbf{a} depends on β and it is cumbersome to experiment with different values of β . Note that \mathbf{F}^{-1} must be computed for each value of β . It is usually simpler to initialize \mathbf{a} using the equivalent ARIMA model.

4.5. Extensions of GES to the Holt-Winters seasonal models

The Holt-Winters seasonal models require three parameters. Sweet (1981) developed GES versions of Holt-Winters and related seasonal models which require only one parameter. The seasonal fitting functions in Sweet's models are vectors of zeros and ones, corresponding to dummy variables. This property of the fitting functions allowed Sweet to derive explicit expressions for \mathbf{F}^{-1} and \mathbf{h} , considerably simplifying the forecasting and revision process. McKenzie's result above for the equivalent ARIMA process also holds for the Sweet models (McKenzie, 1984). For example, the Holt-Winters model with linear trend and multiplicative seasonality is optimal for:

$$(1 - B^n)^2 X_t = (1 - \beta^n B^n)^2 \varepsilon_t \quad (22)$$

4.6. GES with two discount factors

Harrison (1965) and Ameen and Harrison (1984) have criticized the use of a single discount factor for both trend and seasonality in GES. It seems reasonable to expect better performance if the two components are modelled separately, with individual discount factors.

Several alternatives to GES have been developed that appear to be virtually unknown in the U.S. The simplest is known as 'SEATREND' (Harrison, 1965). This method is similar to Model 5-2 above (where we combined Brown's double smoothing model with Holt-Winters seasonal factors) except that Harrison's seasonal factors are Fourier coefficients estimated by:

$$a_i = (1/2p) \sum_{m=1}^p I_{t-p+m} \cos(i\mu_m), \quad (23)$$

$$b_i = (1/2p) \sum_{m=1}^p I_{t-p+m} \sin(i\mu_m), \quad i = 1, 2, \dots, p/2 \quad (24)$$

μ_m is defined as $[2(m-1)\pi/p] - \pi$. Each coefficient is tested for significance and the smoothed seasonal factors are estimated by:

$$I_{t-p+m} = 1 + \sum_m [a_i \cos(i\mu_m) + b_i \sin(i\mu_m)], \quad i = 1, 2, \dots, p/2 \quad (25)$$

The summation includes only significant harmonics.

Another alternative, called 'DOUBTS' (Harrison, 1965), is essentially a double application of GES, first to the trend and then to the seasonal component. Again the structure is similar to Model 5-2. The first step is to use a moving average to remove seasonality from the series. The double smoothing model is fitted and deviations from the trend are computed. Finally, GES is used to fit cosine and sine terms to the deviations. Different values of β are used for the trend and seasonal components. DOUBTS is widely used in the U.K. in preference to SEATREND although Harrison found little difference in forecast accuracy between the two.

The third alternative, 'Discount Weighted Estimation' (Ameen and Harrison, 1984), is more general in that two discount factors are used directly in the matrix F of the weighted fitting functions. The result is a model in which the trend and seasonal components each give a DLS in isolation. Ameen and Harrison derived modified recurrence relations for this approach and showed the resulting models have ARIMA equivalents. Roberts (1982) presented a number of examples of this type of model.

4.7. Empirical evidence on GES

Which version of GES should be used? Is GES in any form more accurate than Holt-Winters? The evidence is sketchy but suggests that one of the GES models with two discount factors should be preferred to standard GES (with a single discount factor). Whether the GES models with two discount factors are more accurate than Holt-Winters awaits further evidence.

Groff (1973) compared standard GES to Holt-Winters on 63 monthly sales series and found little difference in accuracy. Reid (1975) compared standard GES, SEATREND and Holt-Winters on various subsets of 113 macroeconomic series. Standard GES was applied to every series, SEATREND to 47 series and Holt-Winters to 69. In one-period-ahead forecasting, Holt-Winters did better than standard GES on 72 per cent of the series. SEATREND did better than standard GES on 79 per cent of the series. SEATREND was also better than Holt-Winters 57 per cent of the time.

A study by Wagle *et al.* (1968) used the same models as Reid in one-period-ahead forecasting with 20 monthly sales series. Standard GES was best on four series, SEATREND on two series, and Holt-Winters on 14 series.

Another study by Davies and Huitson (1967) compared DOUBTS and Holt-Winters at various forecasting horizons using 22 monthly sales series. For horizons 1-6 there was little difference in accuracy. At longer horizons, DOUBTS was somewhat more accurate.

5. FORECAST MONITORING

In most forecasting systems, it is important to automatically monitor the forecast errors to ensure that the system remains in control. This is especially true when the system is based on simple exponential smoothing, which will lag any trend in the data. In inventory control, forecast monitoring is essential because of the need to take off-line action when there is a significant change in demand. If demand goes up, new orders should be placed on a priority basis, whereas orders

currently outstanding should be expedited into stock. If demand goes down, any unneeded orders should be cancelled to prevent excess inventory investment.

In Section 5.1, we review the operation of monitoring devices (tracking signals) used to keep watch for signs of bias in the forecast errors. The establishment of control limits for the signals is discussed in Section 5.2. The empirical evidence on the choice of tracking signal is discussed in Section 5.3.

5.1. Tracking signals

The first tracking signal used in forecasting was the simple cumulative sum (cusum) of the errors, developed by Brown (1959). The simple cusum is defined as the ratio of the sum of the errors at the end of each period to the smoothed mean absolute deviation (MAD) of the errors. The ratio should fluctuate around zero if the errors are unbiased. If the ratio exceeds a control limit, an exception report is issued to the user.

One possible difficulty with the simple cusum is that it may give an unreasonable number of false alarms. That is the cusum may wander away from zero over time due to nothing more than randomness. To mitigate this problem, Brown (1971, 1982) developed the parabolic mask tracking system which operates as follows: we keep a record of the last 8–12 values of the simple cusum. To test for bias, a parabola is constructed and centred over the most recent cusum. The arms of the parabola point backward in time. If any previous cusum falls outside the area of the parabola, an exception is reported. This should be a more powerful statistical test because the distance from zero of the last cusum is irrelevant—we test only the recent changes in the cusums themselves.

The most thorough tracking signal was developed by Harrison and Davies (1964). This signal is also based on cusums, but they are computed backward in time. The first backward cusum is just the current error; the second is the sum of the last two errors; the third is the sum of the last three errors, and so on. Each cusum is tested individually for bias. The number of cusums needed to operate this control system quickly gets out of hand in any time series. However, Harrison and Davies devised a system whereby all possible cusums can be implicitly tested for bias by storing only four quantities. The best proof of this result is by Coutie *et al.* (1964).

Trigg (1964) developed a signal for the mean error which is widely used in practice. The signal is simply the ratio of the smoothed error to the MAD. The value of the signal will fluctuate between 0 (perfect forecasts) and 1 (extreme bias in the forecasts). Gardner (1983a) proposed a signal for autocorrelation defined as the ratio of the smoothed covariance in errors (at lag 1) to the smoothed MSE. This signal will also fluctuate between 0 and 1.

5.2. Control limits

Control limits for tracking signals can be set according to the probability of getting a false report, defined as a Type I error, or a case where the control limit is exceeded due to chance. Cumulative probability distributions for all signals described above except the backward cusum are available in Gardner (1983a), which extends earlier work by Batty (1969), Brown (1963), and Montgomery and Johnson (1976).

McKenzie (1978) also discussed the distribution of Trigg's signal in the case where the smoothing parameter in the tracking signal is allowed to differ from that in the forecasting model. McKenzie argued that the smoothing parameter in the tracking signal should generally be smaller than that in the forecasting model. Given a good starting value, this idea should reduce the variance of the signal and make it easier to detect bias in the errors.

Another approach is to set control limits on the basis of average run lengths, defined as the number of time periods required to detect bias in the forecast errors. Control limits of this nature are also available in Gardner (1983a).

5.3. Empirical comparisons

There have been four empirical studies comparing the ability of tracking signals to detect bias in the forecast errors. The first was by Golder and Settle (1976), who found that Harrison and Davies' backward cusum was superior to Trigg's signal. Gardner (1983a, 1984, 1985) compared the ability of all five signals to detect bias in forecast errors. Performance comparisons were made on the basis of average run lengths to detect bias in simulated time series. Surprisingly, the simple cusum was recommended as the best signal. The simple cusum gave about the same performance as the sophisticated cusum models and performed significantly better than the Trigg and autocorrelation signals.

Gardner (1985) found that using a relatively small parameter in the Trigg signal improved performance as McKenzie predicted. However, the signal was unstable and unresponsive to bias at any smoothing parameter (in the forecasting model) above 0.1, regardless of the parameter used in the tracking signal.

6. ADAPTIVE CONTROL OF THE SMOOTHING PARAMETERS

In the previous section, the goal of forecast monitoring was to alert the user to the need for intervention in the forecasting system. Here the goal is more ambitious, to make the forecasting system completely automatic.

Thus far we have treated the smoothing parameters as constants. Adaptive parameters are designed to improve performance by automatically changing in a controlled manner as the structure of the time series changes. Some measure of forecast accuracy is monitored to detect changes in the series. If recent accuracy has been good, adaptive-control systems assume that the structure of the time series is stable and should apply moderate parameters. If accuracy deteriorates, the assumption is that the structure of the series has somehow changed and the parameters are increased in order to shorten the response lag in the forecasts.

Section 6.1 compares a number of methods for controlling adaptive parameters. Section 6.2 reviews the empirical evidence on the effectiveness of adaptive parameters.

6.1. Control methods

The most popular control method is based on Trigg's tracking signal described above (see also Trigg and Leach, 1967). As each error is observed, this method sets the value of α , equal to the value of the smoothed forecast error divided by the MAD. A smaller idea was developed by Van Dobben De Bruyn (1964), except that α increases along an S-shaped path as accuracy deteriorates. This makes α nearly constant for small errors but accelerates the response to large errors.

Another popular control method, called evolutionary operation, was developed by Chow (1965). This approach requires that three forecasts be computed each period. One forecast is computed using a base value of α . The others are computed using $\alpha_H = \alpha + 0.05$ and $\alpha_L = \alpha - 0.05$. If the forecast error using α is less than that using α_H and α_L , no change in α is made. If the error from α_H or α_L is lower, α is reset to α_H or α_L . New values of α_H and α_L are computed and the process begins anew. Chow's work was extended by Roberts and Reed (1969), and later by Montgomery (1970), to automatically control two or more parameters in the same model. See also Raine (1971) for criticisms of the theory of evolutionary operation in forecasting.

The stability of these control methods has been widely criticized (see for example Fildes, 1979). Several modifications have been suggested in an attempt to avoid unstable forecasts. Shone (1967) suggested that the changes in α using the Trigg and Leach method be delayed one period—that is, we set α , equal to the value of the ratio at $t - 1$. Flowers (1980) suggested that the range of

permissible values for α be constrained. Whybark (1973) (see also Roberts and Whybark, 1974) developed a conservative model which changes α only when certain control limits on the size of each error are broken. Dennis (1978) developed a similar model which changes α only when a control limit on the number of consecutive errors with the same sign has been broken.

Another conservative model was proposed by Rao and Shapiro (1970), who used the spectrum of the time series to detect changes in structure. Successive spectra of overlapping portions of the series are computed, with α determined as a function of the maximum change in the various frequency components. This procedure should make the forecasts highly resistant to outlying data points.

Finally, the Kalman filter has been used to control the simple smoothing model (Bunn, 1981; Enns *et al.*, 1982). In the context of simple smoothing, the Kalman filter is based on the structure of the random walk given in equations (2) and (3) in Section 1.2. α_t is taken as the ratio of the variance of X_t^* to the total variance of X_t . Bunn gives a simple recursive approximation to the variances whereas Enns *et al.* give a maximum likelihood algorithm.

Kalman filtering, at least in the forms used by Bunn and Enns *et al.*, has two limitations as an adaptive control method. One is the implicit assumption that the underlying covariance structure of the generating white noise processes is not changing with time. If this assumption is true, the recursions will quickly yield the steady-state solution for α . If the assumption is not true, the recursions are incorrect. Another limitation is that α is overly restricted to the range 0–1. There appears to be no way to modify the filter to expand the range for α . See Harvey (1984a, b) for more discussion of the relationships between Kalman filtering and exponential smoothing.

6.2. Empirical evidence

Despite the apparent popularity of adaptive parameters in practice, the empirical research is not encouraging. There have been 10 studies comparing the accuracy of adaptive and constant parameters. A summary of the results: six studies favoured adaptive parameters, but later researchers raised serious questions regarding the validity of five of these studies. The sixth favourable study is difficult to evaluate for reasons explained below. The seventh study was indifferent between adaptive and constant parameters although we show below that some of the comparisons may be biased. The remaining three studies, all based on large sample sizes, favoured constant parameters.

The six studies favouring adaptive parameters were by Bunn (1980), Dennis (1978), Hollier *et al.* (1981), Whybark (1973), Chow (1965) and Smith (1974). The first four studies were re-examined by Ekern (1981, 1982), who found no convincing evidence favouring adaptive parameters. Gardner (1983b) reached the same conclusion regarding the study by Chow. Smith gives simulation results favouring a certain adaptive method which is described only in broad terms. Unfortunately there is not enough information to apply Smith's method to any other time series.

The one indifferent study was by Dancer and Gray (1977), who used adaptive parameters (Trigg and Leach, Whybark) on 359 time series. However, for 97 series Model 5-2 was used in error— S_t^* was taken as the forecast. If the correct model had been used, the constant-parameter version may have done better relative to the adaptive version.

Studies favouring constant parameters were by Gardner and Dannenbring (1980), Makridakis and Hibon (1979) and Makridakis *et al.* (1982). Gardner and Dannenbring used 9000 simulated time series to evaluate all the methods described above with the exceptions of Rao and Shapiro Kalman filtering. Adaptive parameters gave unstable forecasts even when the structure of the series was stable. The instability offset any response advantage when changes in structure occurred. The Gardner and Dannenbring results are mostly based on *ex post* fit, a viewpoint which is often misleading, as discussed in Section 2. However, other large-sample studies agree with the

conclusions. Makridakis and Hibon used the Trigg and Leach method with simple smoothing in 111 series. A constant parameter gave better results using a variety of *ex ante* error measures, a conclusion which was further confirmed using 1001 series in the study by Makridakis *et al.* (1982).

Thus no convincing advantage for adaptive parameters has been demonstrated as yet. There may even be some penalty in accuracy for the use of some of these methods. An alternative to adaptive parameters is to simply refit the forecasting model at regular intervals, using only recent data (Eilon and Elmaleh, 1970). One could also refit immediately after an exception report issued by one of the tracking signals discussed in Section 5 (Buffa, 1975). Using either strategy, the refitting can be done automatically if one is willing to specify a set of permissible α values, say 0.1, 0.2, 0.3, ..., 1.9. Eilon and Elmaleh as well as Buffa recorded worthwhile improvements in accuracy compared to models fitted only once to the early part of each series.

7. TWO-STAGE FORECASTING

Diagnostic checking to determine whether an exponential smoothing model is statistically adequate is usually ignored. There appears to be no solution except to test the adequacy of the corresponding ARIMA model (if such a model exists). If we cannot easily validate a smoothing model, we can at least attempt to improve the forecasts through what Gilchrist (1976) calls two-stage forecasting. The forecasts from any of the models discussed above are treated as the first stage. They are modified in a second stage by one of several procedures discussed below.

7.1. Adjustments for autocorrelation in the errors

If significant autocorrelation is found in the errors, an autoregressive model can be used to modify the forecasts. The most common pattern is first-order autocorrelation (in the one-period-ahead errors). This pattern can be estimated recursively with a simple DLS model. If the estimated autoregressive parameter is R , the one-period-ahead forecast is modified by Re_t . The m -period-ahead forecasts are modified by $R^m e_t$. The modification decreases rapidly with m and is of little help for more than a few periods ahead. Improvements in accuracy through this procedure were reported by Chatfield (1978) and Reid (1975).

7.2. Combining the forecasts from several models of the data

A number of researchers have found advantages in combining exponential smoothing with causal models or with other time series methodologies (Chen and Winters, 1966; Corcoran, 1978; Crane and Crotty, 1967; Gardner, 1979; Kao and Pokladnik, 1978; Newbold and Granger, 1974).

One might also combine several different smoothing models. A great deal of empirical evidence favouring combinations of smoothing models was reported by Makridakis *et al.* (1982) and Makridakis and Winkler (1983). The first study tested two methods of combining forecasts. One was a simple average of five exponential smoothing models (two constant-level, three linear trend) and adaptive filtering. The second was a weighted average based on the sample covariance matrix of percentage errors of the six models. Surprisingly, the first method was better using most error criteria (on 1001 time series). Compared to the other time series methods tested, the first combining method was clearly the best performer in MAPE and average ranking did about the same as several other methods in median APE and average MSE. Further results are given by Makridakis and Winkler (1983), who showed that the MAPE decreased continuously as more models were combined. Of course the rate of decrease eventually approached zero.

Thus a simple average of several exponential smoothing models seems worth trying in practice. But the details of implementation are not clear. How should the particular models be selected?

When should this approach be preferred to a single model? The most vexing problem is that of monitoring the forecasts. It is easy to imagine a situation in which tracking signals would report one or more individual models to be out of control whereas the combined forecast appears to be in control (or vice versa).

7.3. Regressing actual data on the forecasts

The objective in this approach is to eliminate systematic bias in the forecasts. We obtain the second-stage forecasts by a linear regression of actual data on the first-stage forecasts. Favourable empirical evidence is reported by Brandon *et al.* (1983) who applied the procedure to several exponential smoothing models in 33 time series (composed of quarterly corporate earnings data).

8. THE STATE OF THE ART

A judicial summing up: choosing parameters and models on the basis of *ex post* fit cannot be justified. The restricted ranges of parameters and starting values typically used in practice are arbitrary and may detract from accuracy. There is no evidence that choosing parameters in regions of oscillation has any effect on accuracy. The horizon over which the parameters are fitted is not important. Exact DLS is recommended when there are few data to estimate starting values. With interrelated or intermittent series, modified versions of exponential smoothing should improve accuracy. It is widely held that Brown's linear trend model is as accurate as Holt-Winters. Brown's model also has the theoretical advantage of being critically damped. However, Holt-Winters has been more accurate in empirical studies.

A linear trend is typically used for any forecasting horizon although there is evidence the trend should be damped as the horizon increases. GES in standard form is difficult to justify. The equivalent ARIMA model is simpler, more efficient and easier to initialize. The GES models with two discount factors and the Sweet models are attractive alternatives to GES in standard form. Trigg's smoothed-error tracking signal is ubiquitous although the evidence favours any of the cusum signals. No advantage for adaptive parameters has been demonstrated as yet. Adaptive parameters may even detract from accuracy.

As a methodology, exponential smoothing suffers from the lack of an objective procedure for model identification. There is also no procedure for diagnostic checking of the chosen model (although the use of tracking signals mitigates this problem to some extent). Numerous Box-Jenkins forecasters have been critical of exponential smoothing because of these deficiencies. This criticism is valid but the Box-Jenkins approach is infeasible in many practical applications. Furthermore, the multiplicative Holt-Winters seasonal models do not have ARIMA equivalents. The additive Holt-Winters models have ARIMA equivalents but they are so complex as to make identification through their autocorrelation structures impossible.

More research is clearly needed on the problems of model identification and validation in exponential smoothing. The only guidance in the literature appears to be that of McKenzie (1984), who suggested that the variances of the possibly relevant differences of the data be used to assist in model choice. The order of differencing yielding minimum variance is a simple indicator of the appropriate ARIMA model (and its GES equivalent). McKenzie's procedure could be used automatically with any version of GES.

Several other questions deserve more research. Are the GES models with two discount factors more accurate than Holt-Winters? It would be valuable to have results using the Makridakis *et al.* data on this question. It would also be valuable to have results for the Sweet models so that the penalty for the use of a single parameter in the Holt-Winters approach could be assessed. Should

an autoregressive model be fitted to the forecast errors as a matter of course? Again the Makridakis data could help answer this question.

When should a combination of exponential smoothing forecasts be used? How should the models to be combined be selected? Makridakis *et al.* (1982) offer little guidance on these questions.

Exponential smoothing is frequently the only reasonable time series methodology in large forecasting systems. In smaller applications, a wide range of alternative methodologies is available. There is presently no theoretical basis for choosing among these alternatives. In the Makridakis *et al.* study, the best methodology varied considerably depending on the error criterion, the forecasting horizon, whether the data were aggregated, and even whether the data were collected on an annual, quarterly, or monthly basis.

Because there is so little agreement on the most accurate time series methodology, robustness may be the most practical basis for selection of a methodology, a view shared in Fildes' (1979) state-of-the-art survey on time series. There is substantial evidence that exponential smoothing models are robust, not only to different types of data but to specification error. Thus exponential smoothing should not be dismissed out of hand.

The fact that much of the literature has dismissed exponential smoothing as a special case of the Box-Jenkins methodology has been an obstacle to the advancement of time series forecasting. This attitude is no longer supportable considering the difficulties with ARIMA equivalences for the Holt-Winters seasonal models and the fact that exponential smoothing was at least as accurate as Box-Jenkins in the studies by Makridakis and Hibon (1979) and Makridakis *et al.* (1982). A more realistic attitude is that both exponential smoothing and the Box-Jenkins methodology have merit. The challenge for future research is to establish some basis for choosing among these and other approaches to time series forecasting.

ACKNOWLEDGEMENTS

The author wishes to express appreciation to Chris Chatfield, Kenneth Cogger, Johannes Ledolter, Ed. McKenzie, and Arnold Sweet for comments and suggestions on an earlier version of this paper. None of these people necessarily agrees with the opinions expressed in the paper. The bibliography was compiled with the help of Robert Fildes, who made suggestions on the scope of the research and provided a number of references not readily available in the U.S.

REFERENCES

- Abraham, B. and Ledolter, J., *Statistical Methods for Forecasting*, New York: Wiley, 1983.
- Adam, E. E., 'Individual item forecasting model evaluation', *Decision Sciences*, 4 (1973), 458-470.
- Adelson, R. M., 'The dynamic behavior of linear forecasting and scheduling rules', *Operational Research Quarterly*, 17 (1966), 447-462.
- Ameen, J. R. M. and Harrison, P. J., 'Discount weighted estimation', *Journal of Forecasting*, 3 (1984), 285-296.
- Armstrong, J. S., *Long-Range Forecasting*, New York: Wiley, 1978, Chapter 7.
- Batty, M., 'Monitoring an exponential smoothing forecasting system', *Operational Research Quarterly*, 20 (1969), 319-325.
- Berry, W. L. and Bliemel, F. W., 'Selecting exponential smoothing constants: an application of pattern search', *International Journal of Production Research*, 12 (1974), 483-499.
- Bossons, J., 'The effects of parameter mis-specification and non-stationary on the applicability of adaptive forecasts', *Management Science*, 12 (1966), 659-669.

- Box, G. E. P. and Jenkins, G. M., *Time Series Analysis, Forecasting and Control* (Revised edn), San Francisco: Holden-Day, 1976.
- Brandon, C. H. *et al.*, 'Revising forecasts of accounting earnings: a comparison with the Box-Jenkins method', *Management Science*, **29** (1983), 256-264.
- Brenner, J. L. *et al.*, 'Difference equations in forecasting formulas', *Management Science*, **14** (1968), 141-159.
- Brown, R. G., *Statistical Forecasting for Inventory Control*, New York: McGraw-Hill, 1959.
- Brown, R. G. and Meyer, R. F., 'The fundamental theorem of exponential smoothing', *Operations Research*, **9** (1961), 673-685.
- Brown, R. G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, NJ: Prentice-Hall, 1963.
- Brown, R. G., *Decision Rules for Inventory Management*, New York: Holt, Rinehart and Winston, 1967, Part 2.
- Brown, R. G., 'Detection of turning points in a time series', *Decision Sciences*, **2** (1971), 383-403.
- Brown, R. G., *Materials Management Systems: A Modular Library*, New York: Wiley, 1977, Chapters 5-8.
- Brown, R. G., *Advanced Service Parts Inventory Control*, Norwich, VT: Materials Management Systems, Inc., 1982, Chapters 3 and 9.
- Buffa, E. S. and Taubert, W. H., *Production-Inventory Systems: Planning and Control* (revised edn), Homewood, IL: Irwin, 1972, Chapter 2.
- Buffa, F. P., 'The application of a dynamic forecasting model with inventory control properties', *Decision Sciences*, **6** (1975), 298-306.
- Bunn, D. W., 'A comparison of several adaptive forecasting procedures', *Omega*, **8** (1980), 485-491.
- Bunn, D. W., 'Adaptive forecasting using the Kalman filter', *Omega*, **9** (1981), 323-324.
- Chatfield, C., 'Some recent developments in time series analysis', *Journal of the Royal Statistical Society (A)*, **140** (1977), 492-510.
- Chatfield, C., 'The Holt-Winters forecasting procedure', *Applied Statistics*, **27** (1978), 264-279.
- Chen, G. K. C. and Winters, P. R., 'Forecasting peak demand for an electric utility with a hybrid exponential model', *Management Science*, **12** (1966), B531-B537.
- Chow, W. M., 'Adaptive control of the exponential smoothing constant', *Journal of Industrial Engineering*, **16** (1965), 314-317.
- Cogger, K. O., 'Extensions of the fundamental theorem of exponential smoothing', *Management Science*, **19** (1973a), 547-554.
- Cogger, K. O., 'Specification analysis', *Journal of the American Statistical Association*, **68** (1973b), 899-905.
- Cogger, K. O., 'The optimality of general-order exponential smoothing', *Operations Research*, **22** (1974), 858-867.
- Cohen, G. D., 'A note on exponential smoothing and autocorrelated inputs', *Operations Research*, **11** (1963), 361-367.
- Cohen, G. D., 'Bayesian adjustment of sales forecasts in multi-item inventory control systems', *Journal of Industrial Engineering*, **17** (1966), 474-479.
- Corcoran, A., 'The use of exponentially smoothed transition matrices to improve forecasting of cash flows from accounts receivable', *Management Science*, **24** (1978), 732-739.
- Coutie, G. *et al.* *Short-Term Forecasting*, ICT Monograph No. 2, Edinburgh: Oliver and Boyd, 1964.
- Cox, D. R., 'Prediction by exponentially weighted moving averages and related methods', *Journal of the Royal Statistical Society (B)*, **23** (1961), 414-422.
- Crane, D. B. and Crotty, J. R., 'A two-stage forecasting model: exponential smoothing and multiple regression', *Management Science*, **13** (1967), B501-B507.
- Croston, J. D., 'Forecasting for intermittent demand', *Operational Research Quarterly*, **23** (1972), 289-303.
- Dalrymple, D. J. and King, B. E., 'Selecting Parameters for short-term forecasting techniques', *Decision Sciences*, **12** (1981), 661-669.
- Dancer, R. E. and Gray, C. F., 'An empirical evaluation of constant and adaptive computer forecasting models for inventory control', *Decision Sciences*, **8** (1977), 228-238.
- Davies, R. and Huitson, A., 'A sales forecasting comparison', *The Statistician*, **17** (1967), 269-278.
- Dennis, J. D., 'A performance test of a run-based adaptive exponential forecasting technique', *Production and Inventory Management*, **19** (1978), 43-46.
- D'Esopo, D. A., 'A note on forecasting by the exponential smoothing operator', *Operations Research*, **9** (1961), 686-687.
- Dobbie, J. M., 'Forecasting periodic trends by exponential smoothing', *Operations Research*, **11** (1963), 908-918.

- Eilon, S. and Elmaleh, J., 'Adaptive limits in inventory control', *Management Science*, **16**(1970), B533-B548.
- Ekern, S., 'Adaptive exponential smoothing revisited', *Journal of the Operational Research Society*, **32** (1981), 775-782.
- Ekern, S., 'On simulation studies of adaptive forecasts', *Omega*, **10** (1982), 91-93.
- Elton, E. J. and Gruber, H. J., 'Earnings estimates and the accuracy of expectational data', *Management Science*, **18** (1972), B409-B424.
- Enns, P. G. *et al.*, 'Forecasting applications of an adaptive multiple exponential smoothing model', *Management Science*, **28** (1982), 1035-1044.
- Fildes, R., 'Quantitative forecasting—the state of the art: extrapolative methods', *Journal of the Operational Research Society*, **30** (1979), 691-710.
- Fildes, R. and Howell, S., 'On selecting a forecasting model', in Makridakis, S. and Wheelwright, S. C. (eds), *TIMS Studies in the Management Sciences, Forecasting*, Amsterdam: North-Holland, Vol. 12, 1979.
- Flowers, A. D., 'A simulation study of smoothing constant limits for an adaptive forecasting system', *Journal of Operations Management*, **2** (1980), 85-94.
- Gardner, E. S., 'Box-Jenkins vs. multiple regression: some adventures in forecasting the demand for blood tests', *Interfaces*, **9** (1979), 49-54.
- Gardner, E. S. and Dannenbring, D. G., 'Forecasting with exponential smoothing: some guidelines for model selection', *Decision Sciences*, **11** (1980), 370-383.
- Gardner, E. S., 'Automatic monitoring of forecast errors', *Journal of Forecasting*, **2** (1983a), 1-21.
- Gardner, E. S., 'Evolutionary operation of the exponential smoothing parameter: revisited', *Omega*, **11** (1983b), 621-623.
- Gardner, E. S., 'Forecast control with parabolic masks', working paper, 1984. (Copy available from the author at Navy Fleet Material Support Office, Mechanicsburg, PA 17055, U.S.A.)
- Gardner, E. S., 'Cusum vs. smoothed-error forecast monitoring schemes: some simulation comparisons', *Journal of the Operational Research Society*, forthcoming, 1985.
- Gilchrist, W. G., 'Methods of estimation using discounting', *Journal of the Royal Statistical Society (B)*, **29** (1967), 355-369.
- Gilchrist, W. G., *Statistical Forecasting*, London: Wiley, 1976.
- Godolphin, E. J. and Harrison, P. J., 'Equivalence theorems for polynomial-projecting predictors', *Journal of the Royal Statistical Society (B)*, **37** (1975), 205-215.
- Golder, E. R. and Settle, J. G., 'Monitoring schemes in short-term forecasting', *Operational Research Quarterly*, **27** (1976), 489-501.
- Goodman, M. L., 'A new look at higher-order exponential smoothing for forecasting', *Operations Research*, **22** (1974), 880-888.
- Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, New York: Academic Press, 1977.
- Groff, G. K. and Muth, J. F., *Operations Management: Analysis for Decisions*, Homewood, IL: Irwin, 1972, Chapter 11.
- Groff, G. K., 'Empirical comparison of models for short-range forecasting', *Management Science*, **20** (1973), 22-31.
- Gross, D. and Craig, R. J., 'A comparison of maximum likelihood, exponential smoothing, and Bayes forecasting procedures in inventory modeling', *International Journal of Production Research*, **12** (1974), 617-622.
- Harrison, P. J. and Davies, O. L., 'The use of cumulative sum (CUSUM) techniques for the control of routine forecasts of product demand', *Operations Research*, **12** (1964), 325-333.
- Harrison, P. J., 'Short-term sales forecasting', *Applied Statistics*, **14** (1965), 102-139.
- Harrison, P. J., 'Exponential smoothing and short-term sales forecasting', *Management Science*, **13** (1967), 821-842.
- Harvey, A. C., 'Analysis and generalization of a multivariate exponential smoothing model', *LSE Econometrics Programme Discussion Paper No. A.41*, 1984a. (Copy available from the author at London School of Economics, Houghton St., London WC2A 2AE, U.K.)
- Harvey, A. C., 'A unified view of statistical forecasting procedures', *Journal of Forecasting*, **3** (1984b), 245-275.
- Hollier, R. H. *et al.*, 'A comparison of short-term adaptive forecasting methods', *Omega*, **9** (1981), 96-98.
- Holt, C. C. *et al.*, *Planning Production, Inventories, and Work Force*, Englewood Cliffs, NJ: Prentice-Hall, 1960, Chapter 14.
- Howe, W. G., 'A new look at Brown's dynamic inventory system', *Operations Research*, **22** (1974), 848-857.

- Johnson, L. A. and Montgomery, D. C., *Operations Research In Production Planning, Scheduling, and Inventory Control*, New York: Wiley, 1974.
- Jones, R. H., 'Exponential smoothing for multivariate time series', *Journal of the Royal Statistical Society (B)*, **28** (1966), 241-251.
- Kao, E. and Pokladnik, F., 'Incorporating exogenous factors in adaptive forecasting of hospital census', *Management Science*, **24** (1978), 1677-1686.
- Kirby, R. M., 'A comparison of short and medium range forecasting methods', *Management Science*, **13** (1966), B202-B210.
- Landi, D. and Johnson, B., 'Optimal linear inventory control and first order exponential smoothing', *Operations Research*, **15** (1967), 342-347.
- Ledolter, J. and Box, G. E. P., 'Conditions for the optimality of exponential smoothing forecast procedures', *Metrika*, **25** (1978), 77-93.
- Ledolter, J. and Abraham, B., 'Some comments on the initialization of exponential smoothing', *Journal of Forecasting*, **3** (1984), 79-84.
- Lewandowski, R., 'Sales forecasting by FORSYS', *Journal of Forecasting*, **1** (1982), 205-214.
- Makridakis, S. and Wheelwright, S. C., *Interactive Forecasting: Univariate and Multivariate Methods* (2nd edn), San Francisco: Holden-Day, 1978.
- Makridakis, S. and Hibon, M., 'Accuracy of forecasting: an empirical investigation (with discussion)', *Journal of the Royal Statistical Society (A)*, **142** (1979), 97-145.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R. and Winkler, R., 'The accuracy of extrapolation (time series) methods: results of a forecasting competition', *Journal of Forecasting*, **1** (1982), 111-153.
- Makridakis, S., 'Empirical evidence versus personal experience', *Journal of Forecasting*, **2** (1983), 295-306.
- Makridakis, S. et al., *Forecasting: Methods and Applications* (2nd edn), New York: Wiley, 1983.
- Makridakis, S. and Winkler, R., 'Averages of forecasts: some empirical results', *Management Science*, **29** (1983), 987-996.
- Markland, R. E., 'A comparative study of demand forecasting techniques for military helicopter spare parts', *Naval Research Logistics Quarterly*, **17** (1970), 103-119.
- McClain, J. O. and Thomas, L. J., 'Response-variance tradeoffs in adaptive forecasting', *Operations Research*, **21** (1973), 554-568.
- McClain, J. O., 'Dynamics of exponential smoothing with trend and seasonal terms', *Management Science*, **20** (1974), 1300-1304.
- McClain, J. O., 'Restarting a forecasting system when demand suddenly changes', *Journal of Operations Management*, **2** (1981), 53-61.
- McKenzie, E., 'A comparison of standard forecasting systems with the Box-Jenkins approach', *The Statistician*, **23** (1974), 107-116.
- McKenzie, E., 'A comparison of some standard seasonal forecasting systems', *The Statistician*, **25** (1976a), 3-14.
- McKenzie, E., 'An analysis of general exponential smoothing', *Operations Research*, **24** (1976b), 131-140.
- McKenzie, E., 'The monitoring of exponentially weighted forecasts', *Journal of the Operational Research Society*, **29** (1978), 449-458.
- McKenzie, E., 'General exponential smoothing and the equivalent ARMA process', *Journal of Forecasting*, **3** (1984), 333-344.
- McLeavey, D. W. et al., 'An empirical evaluation of individual item forecasting models', *Decision Sciences*, **12** (1981), 708-714.
- Montgomery, D. C., 'Adaptive control of exponential smoothing parameters by evolutionary operation', *AIIE Transactions*, **2** (1970), 268-269.
- Montgomery, D. C. and Johnson, L. A., *Forecasting and Time Series Analysis*, New York: McGraw-Hill, 1976.
- Morris, R. H. and Glassey, C. R., 'The dynamics and statistics of exponential smoothing operators', *Operations Research*, **11** (1963), 561-569.
- Muth, J. F., 'Optimal properties of exponentially weighted forecasts', *Journal of the American Statistical Association*, **55** (1960), 299-306.
- Nerlove, M. and Wage, S., 'Some observations on adaptive forecasting', *Management Science*, **10** (1964), 207-224.
- Newbold, P. and Granger, C. W. J., 'Experience with forecasting univariate time series and the combination of forecasts (with discussion)', *Journal of the Royal Statistical Society (A)*, **137** (1974), 131-165.

- Packer, A. H., 'Simulation and adaptive forecasting as applied to inventory control', *Operations Research*, **15** (1967), 660-679.
- Pandit, S. M. and Wu, S. M., 'Exponential smoothing as a special case of a linear stochastic system', *Operations Research*, **22** (1974), 868-879.
- Pegels, C., 'Exponential forecasting: some new variations', *Management Science*, **15** (1969), 311-315.
- Peterson, R. and Silver, E., *Decision Systems for Inventory Management and Production Planning*, New York: Wiley, 1979, Chapter 4.
- Raine, J. E., 'Self-adaptive forecasting reconsidered', *Decision Sciences*, **2** (1971), 181-191.
- Rao, A. G. and Shapiro, A., 'Adaptive smoothing using evolutionary spectra', *Management Science*, **17** (1970), 208-218.
- Rao, A. G., 'A comment on forecasting and stock control for intermittent demand', *Operational Research Quarterly*, **24** (1973), 639-640.
- Reid, D. J., 'A review of short-term projection techniques', in Gordon, H. A. (ed.), *Practical Aspects of Forecasting*, London: Operational Research Society, 1975.
- Roberts, S. A., 'A general class of Holt-Winters type forecasting models', *Management Science*, **28** (1982), 808-820.
- Roberts, S. D. and Reed, R., 'The development of a self-adaptive forecasting technique', *AIIE Transactions*, **1** (1969), 314-322.
- Roberts, S. D. and Whybark, D. C., 'Adaptive forecasting techniques', *International Journal of Production Research*, **12** (1974), 634-645.
- Shone, M. L., 'Viewpoints', *Operational Research Quarterly*, **18** (1967), 318-319.
- Smith, D. E., 'Adaptive response for exponential smoothing: comparative system analysis', *Operational Research Quarterly*, **25** (1974), 421-435.
- Sweet, A. L., 'Adaptive smoothing for forecasting seasonal series', *AIIE Transactions*, **13** (1981), 243-248.
- Sweet, A. L., 'Computing the variance of the forecast error for the additive seasonal model', *Research Memo No. 83-5* (1983a), School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, U.S.A.
- Sweet, A. L., 'Computing the variance of the forecast error for the multiplicative (Holt-Winters) seasonal model', *Research Memo No. 83-7* (1983b), School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, U.S.A.
- Taylor, S. G., 'Initialization of exponential smoothing forecasts', *AIIE Transactions*, **13** (1981), 199-205.
- Theil, H. and Wage, S., 'Some observations on adaptive forecasting', *Management Science*, **10** (1964), 198-206.
- Trigg, D. W. and Pitts, E., 'The optimal choice of the smoothing constant in an inventory control system', *Operational Research Quarterly*, **13** (1962), 287-298.
- Trigg, D. W., 'Monitoring a forecasting system', *Operational Research Quarterly*, **15** (1964), 271-274.
- Trigg, D. W. and Leach, A. G., 'Exponential smoothing with an adaptive response rate', *Operational Research Quarterly*, **18** (1967), 53-59.
- Tydeman, J., 'A note on short-term forecasting using an irregular time interval', *Operational Research Quarterly*, **23** (1972), 381-383.
- Van Dobben De Bruyn, C. S., 'Prediction by progressive correction', *Journal of the Royal Statistical Society (B)*, **26** (1964), 113-122.
- Wade, R. C., 'A technique for initializing exponential smoothing forecasts', *Management Science*, **13** (1967), 601-602.
- Wagle, B. et al., 'A program for short-term sales forecasting', *The Statistician*, **18** (1968), 141-147.
- Ward, D. H., 'Comparison of different systems of exponentially weighted prediction', *The Statistician*, **13** (1963), 173-185.
- Whybark, D. C., 'A comparison of adaptive forecasting techniques', *Logistics Transportation Review*, **8** (1973), 13-36.
- Winters, P. R., 'Forecasting sales by exponentially weighted moving averages', *Management Science*, **6** (1960), 324-342.
- Wright, D., 'Extension of forecasting methods for irregularly spaced data', *Proceedings of the International Time Series Meeting*, Toronto, Canada, 1983. (Copy available from the author at Faculty of Administration, University of Ottawa, Ottawa V1N C85, Canada.)

Author's biography:

Everette S. Gardner, Jr. is a Commander in the U.S. Navy and Director of the Operations Analysis Department at the Navy Fleet Material Support Office in Mechanicsburg, Pennsylvania. His department is

responsible for operations research on business management in the Navy. This includes the development and maintenance of models used to manage more than one million inventory items. He holds the Ph.D. in Business Administration from the University of North Carolina at Chapel Hill. Dr. Gardner has published on production, inventory management, and forecasting in the *Journal of Forecasting*, *Journal of the Operational Research Society*, *Naval Research Logistics Quarterly*, *Management Science*, *Decision Sciences*, *Interfaces* and *Omega*. Dr. Gardner is Associate Professor of Operations Management, Pennsylvania State University and Associate Editor of the *Journal of Forecasting*.

Author's address:

Everette S. Gardner, Jr., Operations Analysis Department (Code 93), Navy Fleet Material Support Office, P.O. Box 2010, Mechanicsburg, PA 17055, U.S.A.

COMMENTARIES

**'Exponential Smoothing: The State of the Art'
by E. S. Gardner, Jr.**

Introduction to the Commentaries

KENNETH O. COGGER
University of Kansas, U.S.A.

The comprehensive survey of the state of the art in exponential smoothing by Gardner was impressive on my first reading. When asked by the editors to organize commentaries on the paper, I was happy to take on this task, given my perception of its importance.

Although much of the paper is not controversial, it was nevertheless desired that the widest range of people be contacted and asked to comment. Nine experts in the topic area with a variety of perspectives were invited to formally comment on the paper. Practitioners as well as academics were represented. Some were closely identified with early developments in the area of exponential smoothing, and others were associated with continuing developments in the ARIMA modeling area. Several of those invited to comment did not submit formal commentaries but did review the work very favourably and provided reactions directly to Gardner. In some cases, these comments were incorporated in the manuscript.

The commentaries which follow present Gardner's paper, I believe, in the proper perspective. As McKenzie states, the paper contributes greatly to our understanding of the development of smoothing procedures as well as our understanding of their properties. Knowledge of these properties can, as Chatfield points out, help practitioners sort out which procedures might be applied in a given situation. This selection problem is sometimes difficult and Hillmer's agreement with Gardner on the need for an organized approach to the selection of a smoothing variant seems reasonable and suggests potential future research avenues.

I would like to add an additional comment. Early work, including some by myself, into the identification of ARIMA equivalents to exponential smoothing has sometimes been incorrectly characterized as revealing the 'inferiority' of smoothing procedures, since in many cases an exponential smoothing model is demonstrably a special case of a broad class of ARIMA models. This characterization is not shared by me. Analytically, it was shown in an early paper that smoothing procedures are rather robust. Thus, their 'inferiority' in a practical sense has not been supported. The importance of knowing what these ARIMA equivalences are, however, cannot be overemphasized. Any progress in developing objective selection procedures for exponential smoothing will probably benefit from this work, which is fully described in Gardner's paper. A further important point on model equivalencies is that I have come to the point of view, since the development of damped trend smoothing, and other variants, that every ARIMA model can probably be described in exponential smoothing terminology. There are exponential smoothing models which have no ARIMA equivalent (e.g. Winters multiplicative seasonal smoothing) as well as models which have very non-parsimonious ARIMA equivalents (e.g. Winters additive seasonal approach). Thus, the debate is not over between provincial advocates of

0277-6693/85/010029-10\$01.00

© 1985 by John Wiley & Sons, Ltd.

Received October 1984

these two modelling procedures. Many of us, I suspect, fall somewhere in between, with the specific application and setting dictating a practical choice.

Author's biography:

Kenneth O. Cogger is a Joyce C. Hall Faculty Fellow and Professor of Business Administration at the University of Kansas. He has a Ph.D. in Statistics and Management Science from the University of Michigan. His research interests include time series analysis, forecasting and applied stochastic processes and he has published articles in *Management Science*, *Operations Research*, *Journal of the American Statistical Association* and other journals.

Author's address:

Kenneth O. Cogger, School of Business, University of Kansas, Lawrence, KA 66045, U.S.A.

Comments on 'Exponential Smoothing: The State of the Art' by E. S. Gardner, Jr.

C. CHATFIELD

University of Bath, U.K.

It is my impression that a high proportion of projection forecasts are produced by some variant of exponential smoothing. When I became involved in forecasting, I was initially highly confused by the complex interrelationships between Holt-Winters, Brown, GES, Box-Jenkins and the like, and it took me many years to sort it all out in my own mind. A review paper clarifying the whole area is long overdue and this important paper is therefore particularly welcome. It should make life easier not only for the new researcher but also for more experienced workers by providing a handy reference source.

One of the problems in comparing different forecasting methods is the bewildering variety of notation. I hope, therefore, that the author's plea for a standard notation as in his Exhibit 2 will be respected by everyone even though most of us will no doubt be able to think of (different!) ways in which it might be 'improved'. I also endorse the author's remarks that starting values and smoothing parameters should generally not be chosen arbitrarily but by looking at the data.

The main message to come out of forecasting competitions such as the M-competition, is that there are many different types of forecasting problem requiring different treatment, and that no single class of models is superior in every case. The choice depends on the data, the objectives, the skill of the analyst, the programs available and so on. My own 'favourite' is the Holt-Winters method which is easy to implement and understand, and seems robust. I am pleased to see that this method comes out well in the paper. I also think that the full Box-Jenkins procedure is sometimes worth trying, though not when the series is dominated by trend and seasonal variation. I have never been attracted to Brown's method because of the use of only one smoothing parameter and its effect on the error-correction form of the model (e.g. αe_t and $\alpha^2 e_t$ in Exhibit 5). Nor have I ever been attracted to GES because of its formulation in multiple regression terms. Multiple regression is a most misused technique. Thus the paper has served to consolidate my view that the Holt-Winters method is generally a sensible technique to use when a simple projection forecast is required.

Author's biography:

C. Chatfield is a senior lecturer in Statistics at the University of Bath, U.K. He has a Ph.D. in Statistics from Imperial College, London and has held teaching appointments at London University and McGill University, Montreal. He has been a Council member of the Royal Statistical Society and an editor of *Applied Statistics*. His research interests include time-series analysis, marketing models and statistical inference. He has published widely and his three books include *The Analysis of Time Series*, 3rd edn, published by Chapman & Hall, 1984.

Author's address:

C. Chatfield, University of Bath, School of Mathematics, Claverton Down, Bath BA2 7AY, U.K.

**Comments on 'Exponential Smoothing:
The State of the Art' by E. S. Gardner, Jr.**

S. C. HILLMER
University of Kansas, U.S.A.

I found Everette Gardner's review of exponential smoothing to be very interesting. It is successful in presenting a broad view of the current state of exponential smoothing. The paper is non-controversial in that its aim is to summarize previous research on exponential smoothing rather than to come to any definitive conclusions about the 'best' variant of exponential smoothing or the 'best' forecasting method. I would like to thank Gardner for doing an excellent job of organizing a vast amount of material in a readable manner. My discussion of the paper is concerned with a number of observations about exponential smoothing based upon the paper.

One thing that struck me is that there are many different forecasting methods which fall in the general category of exponential smoothing. For example, Exhibit 1 suggests that there are as many as 17 different versions. Why are there so many variations? One explanation is that there are many different types of time series behaviour occurring in the real world and each different method was designed to best forecast each different time series structure. Another explanation is that some of the variations are proposed solutions to time series behaviour which is perceived to exist by researchers who rarely or never deal with real data but in reality there are few time series for which these variations are relevant. Some of the variations may be able to be dismissed for this reason; however, based upon the popularity of exponential smoothing as a forecasting technique it is clear that many of the versions have some merit. Thus, an important question for the practitioner in need of a forecast method for a particular time series is: 'which of the variants of exponential smoothing is best for my particular data set?' In a sense much of Gardner's paper is attempting to provide advice about this question.

My interpretation of the advice is that to pick the correct method, it takes experience, skill and some knowledge of the characteristics of the data you desire to forecast. A hypothetical practitioner may have to answer questions such as the following. Is your data seasonal? Is the seasonality additive or multiplicative? Does the data fluctuate locally around a constant level, or a linear trend, or a non-linear trend? How stable are the parameter estimates? Once these kinds of questions are answered a decision can be made about the particular variant of exponential smoothing that is appropriate. Then there are other more technical questions to answer. What values should be chosen for the smoothing parameters? What are the starting values? The review by Gardner suggests that there are a variety of opinions about these technical issues. Furthermore, the complexity of the issues seems to increase the more removed the methods get from the basic exponential smoothing models. My point is that unless a practitioner is willing to limit the possibilities to only the basic forms of exponential smoothing, the process which leads to getting the forecasts can apparently get very complicated. I suspect that experts in exponential smoothing have had a great deal of experience with real data and would admit that there is a great deal of skill involved in choosing the right alternative. If my suspicions are correct, I would argue that one of the virtues frequently claimed for exponential smoothing, its simplicity, may be illusory.

Suppose a practitioner is faced with the problem of choosing which approach within the realm of exponential smoothing methods will provide the best forecasts. One way to approach this problem is to arbitrarily select a simple form of exponential smoothing such as a non-seasonal constant level method or a non-seasonal linear trend method. Suppose this method is going to be applied to a number of time series. If some of the series' behaviour is inconsistent with what is implicitly

assumed by the exponential smoothing method selected (e.g. if they are seasoned) then the forecast performance for these series may be poor. Thus, there is a potential cost to be paid for the simple-minded approach. Another approach is to attempt to match the particular variant of exponential smoothing to homogenous groups of series. This may be difficult for people who are inexperienced in applying exponential smoothing. What such people need is an organized approach to choosing a particular exponential smoothing variant and methods to validate that the variant chosen is appropriate. Gardner recognizes this need when he argues that 'more research is clearly needed on the problems of model identification and validation in exponential smoothing'. The results of this research are most critical to beginning exponential smoothers since these results can partially alleviate their lack of experience. It seems to me that one of the greatest contributions of Box and Jenkins to the area of time series analysis is their organized approach to the modelling of time series. They made it easier for people with limited experience to begin modelling real time series. A similar contribution would be helpful to novices in the area of exponential smoothing.

I was somewhat relieved to discover that exponential smoothing has evolved to the state where there are many complex alternatives. In my own experience in the modelling of time series, I know that real data can behave in many different ways. I view the many alternatives in exponential smoothing as an attempt to deal with the variety of behaviour which exists in real time series. I have believed for some time that there is a certain amount of art involved with the analysis of data and the developments in exponential smoothing seem to confirm that belief. If it is true that some of the more advanced methods discussed in the paper are relevant, then it is worth while to spend time in deciding which method is most appropriate for the data being forecast. This view is consistent with the common-sense attitude that if you spend time thinking about a problem (in this case forecasting a time series) and the peculiarities involved with that problem then you should be in a better position to pick the best available solution than you would be if you hadn't bothered to think.

Author's biography:

Steven C. Hillmer is an associate professor of Business Administration at the University of Kansas. He has a Ph.D. in Business Statistics from the University of Wisconsin at Madison. His research interests include time series analysis, seasonal adjustments and forecasting and he has published articles in *Journal of the American Statistical Association*, *Biometrika*, *Journal of Business and Economic Statistics* and *Journal of Forecasting*.

Author's address:

S. C. Hillmer, The University of Kansas, School of Business, Lawrence, KA 66045, U.S.A.

**Comments on 'Exponential Smoothing:
The State of the Art' by E. S. Gardner, Jr.**

ED. MCKENZIE

Naval Postgraduate School, U.S.A. and University of Strathclyde, U.K.

The author is to be congratulated on a good practical review of the models and techniques known collectively as exponential smoothing (ES). As his references show, we have made great progress in the use and understanding of ES procedures in the last 30 years. However, this has been accompanied by considerable argument and controversy about models and their performance compared with other approaches. My own view is that some of the associated algebra and numerical legerdemain may be obscuring more basic aspects of the problem. Some of these will be discussed in this note.

The underlying model which is assumed to be generating the data consists of a linear combination of known functions of time plus an error term. Usually, these errors are assumed to be independent and identically distributed (i.i.d.) random variables with zero-mean. The modeller takes the view that the coefficients in the linear combination may change in time. He seeks to cope with this kind of instability by means of ES. There are two distinct approaches available to him, and we consider these now.

The first encompasses the procedures associated with the work of R. G. Brown which are optimal with respect to the discounted least squares (DLS) criterion. Although others have advanced the state of knowledge in this area, Brown's contribution is singularly worthy of note. More than anyone he is responsible for the popularity of this approach. His books on the subject are landmarks in forecasting theory and practice. They are well-written, lucid and, more importantly for textbooks, easy and often enjoyable to read.

The ideas behind DLS are simple and attractive. The forecasts of future values are weighted averages of observations with the more recent having greater weight than those in the more distant past. The motivation for such weighting is that, in this unstable situation, more recent data are inherently more relevant to the future than earlier values. The idea has great intuitive appeal, but can be too simple. Consider, for example, the prediction of seasonal data. The data most relevant to the forecast of sales next April may well include the sales figures for the most recent Aprils as well as the most recent months.

Considerations such as these lead naturally to the idea of smoothing distinct components of the forecast, e.g. level, trend and seasonal patterns, separately, and with different rates. This is the basis of the linear trend system of Holt (Model 4-1) and the seasonal systems of Winters (Models 4-2, 4-3). The relationships between the systems obtained by the two different approaches is well displayed by their error-correction forms in Exhibits 3-7. In passing, we note that Models 5-3, 5-4 are not DLS optimal. Such models are discussed in Section 4.5. The structure of the different ES systems is the same for the same models. Only the error-correction coefficients change. The highlighting of this structure and the discussion of the choice of these coefficients is an important aspect of the paper.

Another useful feature of the paper is the presentation of the equivalent ARIMA process for each system. Although the author does not say so, some care is required in the use of the term equivalent here. Initially, it is derived as the process whose minimum mean square error (MMSE) one-step ahead forecasts are given by the ES system. The process is referred to as equivalent only if the MMSE forecasts for all lead times are given by the ES system. This is a real distinction. There are linear forecasting systems whose one-step ahead forecasts are optimal for a particular ARIMA process but whose forecasts for other lead times are not. For more discussion, see Godolphin and Harrison (1975).¹ All the equivalent processes in Gardner's paper are indeed equivalent.

The practical value of the equivalent ARIMA lies in the fact that since the two sets of forecasts are identical the procedures for obtaining them may be considered interchangeable. A large number of practical and theoretical benefits can be derived from this relationship. Some of these are noted in the paper. They and others are discussed in detail in McKenzie's papers referenced therein. There are two aspects, however, worthy of some discussion here.

First, we consider the equivalent ARIMA for the GES system given by equation (20). It is $\phi(B)X_t = \theta(\beta B)e_t$. In order to make any use of this result we must obtain ϕ and θ , and this is not discussed in the paper. We note that the model assumed for the data is in the usual form of a deterministic component plus an error term. The former consists of a linear combination of the fitting functions, as given by equation (16), and the error terms, $\{z_t\}$, are usually assumed to be i.i.d.

¹ References refer to the list in Gardner's paper.

with zero mean. Now, ϕ corresponds to the difference equation whose solutions are the fitting functions $\{f_i(t)\}$. Thus, $\phi(B)f_i(t) = 0, i = 1, 2, \dots, n$. It is easier to derive ϕ than it may at first appear since the factors of ϕ correspond to different fitting functions. This is discussed in more detail and examples are given by McKenzie (1976b, 1984). We can obtain θ directly from ϕ using $\theta_k = \phi_{n-k}/\phi_m, k = 1, 2, \dots, n$.

We now consider the one-step ahead error process $\{e_t\}$. Note that e_t is the error for both the ES system and the equivalent ARIMA. Note also that it is not the same as z_t . If we filter the GES model using $\phi(B)$ then, by the definition of ϕ , the deterministic component in equation (16) must vanish. Thus, $\phi(B)X_t = (B)z_t$. Hence, from equation (20), $\theta(\beta B)e_t = \phi(B)z_t$; i.e. the forecast error process from the GES forecasts follows an ARIMA process when the underlying model really is as supposed. Note that, contrary to ARMA methodology, DLS forecast errors are expected to be correlated. The DLS forecasts are not optimal in an MMSE sense for the assumed model (as defined above and by equation (16)). This is an important point and relates to the underlying philosophy of ES systems. We shall return to it later.

Apart from any intrinsic interest, the error ARIMA process is important because it allows us to derive all the statistical properties of the error sequence directly. These are important for monitoring purposes. The mean and variance are easily obtained. For the GES system they are available from DLS theory, but for other ES systems, e.g. Winters' additive seasonal system, the variance would be very difficult to obtain otherwise. The correlation structure of $\{e_t\}$ is useful if we wish to use a tracking signal such as the CUSUM or smoothed error, or any statistic which uses more than one error at a time.

There is one aspect of monitoring which is not discussed in the literature but may be worth noting here. The philosophy of ES is that the underlying model may change but the forecast will respond and 'home-in' on the new model. Thus, we monitor the forecast errors to ensure that the response is adequate. Sometimes, however, we may wish to know that a model change has occurred, even though the forecast response has been adequate and the errors are now acceptable again. This can be achieved by monitoring $\{z_t\}$. This process is unobservable but can be estimated using $\phi(B)z_t = \phi(B)X_t$. It has two useful properties. First, by the model assumptions, the process is uncorrelated and so can be monitored in a standard way, e.g. using a standard CUSUM. Secondly, if there is a model change, then that change is conveyed to z_t and held there indefinitely. As an illustration of this, consider forecasting a constant level using simple smoothing, i.e. Model 3-1. If the level changes by an amount Δ then, N periods after the change occurred, the mean value of e_t is $\beta^N \Delta$, whereas the mean of z_t is still Δ .

Another area in which the equivalent ARIMA can be helpful is model selection. As the author notes, this is a sadly neglected topic of considerable importance. The lack of useful procedures is also highlighted in the empirical comparisons referenced in the paper. In all these studies comparisons are made with particular ES models, and never with any general ES methodology (comparable to the ARMA methodology). Naturally, this tends to favour the ARMA approach, and renders the results of the more recent studies all the more remarkable.

In order to create a truly automatic ES forecasting system we would need to develop a robust model selection procedure. The idea of using an automatic system in which models are selected after only the most cursory and subjective of examinations is clearly foolish. This is not the place to attempt to develop such a procedure, but there are some pertinent observations we may make. In the usual situations in which such a routine may be expected to be used the model would be chosen from quite a small set of models. The corresponding decisions are easily listed. We need to decide whether the data are seasonal (S) or non-seasonal (NS); if seasonal, then whether additive (AS) or multiplicative (MS); and finally whether the model needs a linear trend (L), a damped trend (D) or is trend-free (F). Table 1 outlines these decisions and the models involved.

Table 1

Seasonality	Trend	Model type
NS	F	3-1
	D	7-1, 7-2
	L	4-1; 5-1, 5-2
AS	F	3-2
	D	7-3
	L	4-2; 5-3
MS	F	3-3
	D	7-4
	L	4-3; 5-4

The indicated models denote structure only. The parameters used in any particular case may be chosen to have different forms from those in the paper. Some of the decisions can be made by fitting the data to various models, or passing them through various filters, e.g. differencing. Others, however, are more complex. It is clearly going to be difficult, for example, to separate the problems of seasonal type and trend type for seasonal data. Nevertheless, such an automatic procedure is urgently required for many applications. Further, there is no doubt that were such a system available it would herald considerable success in the routine application of ES techniques. It would also make future empirical comparisons somewhat fairer.

Our final topic concerns perhaps the most intriguing consequence of the ES-ARIMA equivalence. It involves the frequently quoted view that most ES procedures can be ignored because they are special cases of ARIMA models. A statement conveying this sentiment appears in almost every reference in Gardner's paper where the equivalent ARIMA process for an ES procedure is discussed. I shall refer to it as the 'special case' argument. The theoretical justification seems clear enough. To make it more concrete, we will examine a particular situation. Consider the constant level predictor given in the paper by Model 3-1. The equivalent ARIMA process is given by $(1 - B)X_t = (1 - \beta B)e_t$. It is an ARIMA (0, 1, 1) process. Note, however that β is usually chosen from the interval (0, 1) whereas if we were modelling an ARIMA (0, 1, 1) process we would be free to select β from (-1, 1). Thus, ES Model 3-1 is a special case of the ARIMA (0, 1, 1) model, and enjoys only a restricted range of parameter values. This, in essence, is the 'special case' argument.

In the empirical comparisons carried out in the early seventies, the truth of this was apparently borne out. However, the more recent studies do not reflect the same confidence in its validity or practical significance. Indeed, in his reply to the Commentary on the M-Competition, Makridakis notes that the superior performance of the ARIMA procedures as predicted by this 'special case' argument did not materialize. He urges further research to determine the reason for this.

We may begin a response to this by noting that in any application we are not comparing simply different models or procedures. Any sensible comparison should really be between the two different approaches with their different criteria and philosophies. It is certainly true, as the 'special case' argument suggests, that if we wish to forecast a known ARIMA process with MMSE predictions we would never consider using ES. On the other hand, if we wish to model a deterministic function using DLS (perhaps to allow for possible variation in the model parameters) we would never use the ARIMA methodology. It would be inferior by our chosen criterion, and counter to our forecasting philosophy.

To illustrate this idea, we return to our example of the constant level predictor. Our choice of simple smoothing, i.e. Model 3-1, is made on the assumption that, even if the data so far have a

fairly constant level, we may expect the future to bring changes. The equivalent ARIMA, as noted before, is an ARIMA (0, 1, 1). Note, however, that the ARIMA modeller is very unlikely to fit such a model to the data as described. He would probably fit the model: $X_t = \theta_0 + a_t$; and θ_0 would be estimated by the mean of the data available. This example illustrates the main difference between the two approaches and their conflicting philosophies. The ARIMA modeller assumes that the stationarity observed in the data will be preserved into the future. If this should prove to be the case he can safely anchor his forecasts in the fitted period by using an estimate of the constant level obtained there. If he is correct, his predictions are MMSE optimal with all the corresponding 'nice' statistical properties. The ES modeller, on the other hand, may note the stability of the level in the fitting period, but he fears the worst for the future. As a consequence, his ES model will probably not yield such a good fit on the available data. Nor will it be as good in the future as the ARIMA model if the data really is stationary there. However, should there be variation in the level of the process then these two positions are completely reversed. The ES modeller's primary aim in model building has to be robustness. This follows directly from his attitude towards the data and the underlying process. It is important to note that a direct consequence of this philosophy is that the ES model cannot attain the 'nice' statistical properties, such as uncorrelated residuals, which are the goals of the ARMA modeller. Indeed, the ES modeller does not seek them. They are the price of his goal which is robustness.

It is important to emphasize that relationships between individual models may be entirely irrelevant to performance in practice. We cannot with any sense of reality compare the behaviour of particular models in a vacuum. What we must consider are the two approaches and how the different criteria and conflicting philosophies react to yield a model. To declare ES procedures to be redundant because of the 'special case' argument is at best naïve and simplistic. At worst it is misleading.

In his 1962 book, Brown opens the section on error analysis with a quotation from Sherlock Holmes. The great detective urges the redoubtable Dr. Watson to warn him if he should ever become complacent about his powers or careless in his work. Watson is to issue the warning by whispering in his ear the single word 'Norbury'. To those who continue to use the 'special case' argument I would like to say that I think the time for whispering is past.

NORBURY!

Author's biography:

Ed. McKenzie is a lecturer in Statistics in the Mathematics Department of the University of Strathclyde. His research interests include the relationships between forecasting systems, with special reference to those employing the exponential smoothing concept, and the modelling of non-Gaussian time series.

Author's address:

Ed. McKenzie, University of Strathclyde, Department of Mathematics, Glasgow G1 1XW, Scotland, U.K.

Author's Response to Comments on 'Exponential Smoothing: The State of the Art'

E. S. GARDNER, Jr.

Navy Fleet Material Support Office, Mechanicsburg, PA, U.S.A.

McKenzie's mention of Brown's contributions to exponential smoothing prompts this question: who invented the methodology? Like many other great ideas in management science, it appears that exponential smoothing was conceived by at least two researchers working independently, in this case Brown and Holt.

Brown developed exponential smoothing during World War II when he worked for the U.S. Navy's Operations Evaluation Group. One of his assignments involved the design of a tracking system for fire-control information on the location of submarines. This information was used in a mechanical computing device (a ball-disk integrator) to estimate target velocity and the lead angle for dropping depth charges. This tracking model was essentially exponential smoothing of continuous data, an idea still used in modern fire-control equipment.

After the war, Brown developed the principles of exponential smoothing for discrete data. One of his early applications was in forecasting the demand for spare parts in Navy inventories. The savings in data storage over moving averages led to the adoption of exponential smoothing throughout the Navy's inventory systems in the late 1950s.

Brown's first book (1959) presented a set of simple models for inventory applications. His second book (1963) developed the GES methodology.

Holt worked independently of Brown in the 1950s to develop the same constant-level model, a somewhat different model for smoothing linear trends, and an entirely different approach to seasonal data. Holt's early work was sponsored by the Office of Naval Research and was circulated in an unpublished memorandum in 1957 (often erroneously cited as a published book in the literature). Winters assisted in the development of Holt's method, which were completed in Winters (1960) and Holt *et al.* (1960).

McKenzie notes that there are occasions when tracking signal reports are useful even though the forecasts have caught up to the data. This may seem a trivial idea to some but it is a critical consideration in forecasting for inventory control. As discussed in Section 5, one must take off-line action to adjust the pipeline of material flowing into stock after a significant change in demand. Such action is necessary regardless of whether the forecast response has been adequate.

The need for taking off-line action is the reason for the popularity of simple smoothing in inventory control applications. The errors from simple smoothing quickly reflect changes in demand. Use of a trend component in the forecasting model may obscure these changes. That is, the tracking signal may not have time to report the change because of the faster response induced by the trend component.

In the interests of full disclosure I should mention that I share Chatfield's mistrust of GES. My experience has been that the Fourier functions used to model seasonality in GES are incomprehensible to most practitioners. In contrast the seasonal terms in the Holt-Winters approach have intuitive meaning to the model user. Indeed the equivalent ARIMA to any GES model has more intuitive appeal and ought to be easier to implement.

Hillmer speculates that the simplicity of exponential smoothing (the Holt-Winters class of models) is illusory. I agree, and the point is worth emphasizing. If one attempts to model individual series, the methodology is by no means simple. This may explain why model identification has never been done in published empirical work. I have often wondered how the results of empirical

studies such as those of Makridakis *et al.* (1982) or Newbold and Granger (1974) would change if some attempt at model identification were made.

The problem of model identification is of course not unique to exponential smoothing. Advocates of state space models, in particular the Bayesians, have traditionally ignored the question of identification. Whatever the defects in the Box-Jenkins approach, at least it provides an organized approach to model-building as Hillmer points out.

Hillmer's notion of modelling homogeneous groups of series is intriguing. There are many inventory applications in which thousands of forecasts are made each time period with a single exponential smoothing model. Could accuracy be improved at a reasonable cost in data processing efficiency by automatically selecting from several different models depending on the characteristics of demand? This problem has been overlooked in the literature although it is a promising opportunity for both theoretical and empirical research.