Exponential smoothing:
The state of the art – Part II

Everette S. Gardner, Jr.

Bauer College of Business
334 Melcher Hall
University of Houston
Houston, Texas 77204-6021
Telephone 713-743-4744, Fax 713-743-4940
egardner@uh.edu

June 3, 2005

Exponential smoothing: The state of the art – Part II

Exponential smoothing: The state of the art – Part II

Abstract

In Gardner (1985), I reviewed the research in exponential smoothing since the original work by Brown and Holt. This paper brings the state of the art up to date. The most important theoretical advance is the invention of a complete statistical rationale for exponential smoothing based on a new class of state-space models with a single source of error. The most important practical advance is the development of a robust method for smoothing damped multiplicative trends. We also have a new adaptive method for simple smoothing, the first such method to demonstrate credible improved forecast accuracy over fixed-parameter smoothing. Longstanding confusion in the literature about whether and how to renormalize seasonal indices in the Holt-Winters methods has finally been resolved. There has been significant work in forecasting for inventory control, including the development of new prediction distributions for total lead-time demand and several improved versions of Croston's method for forecasting intermittent time series. Regrettably, there has been little progress in the identification and selection of exponential smoothing methods. The research in this area is best described as inconclusive, and it is still difficult to beat the application of a damped trend to every time series.

Key words

Time series – ARIMA, exponential smoothing, state-space models, identification, stability, invertibility, model selection; Comparative methods – evaluation; Intermittent demand; Inventory control; Prediction intervals; Regression – discount weighted, kernel

1. Introduction

When Gardner (1985) appeared, many believed that exponential smoothing should be disregarded because it was either a special case of ARIMA modeling or an *ad hoc* procedure with no statistical rationale. As McKenzie (1985) observed, this opinion was expressed in numerous references to my paper. Since 1985, the special case argument has been turned on its head, and today we know that exponential smoothing methods are optimal for a very general class of state-space models that is in fact broader than the ARIMA class.

This paper brings the state of the art in exponential smoothing up to date with a critical review of the research since 1985. Prior research findings are included where necessary to provide continuity and context. The plan of the paper is as follows. Section 2 summarizes new information that has come to light on the early history of exponential smoothing. Section 3 gives formulations for the standard Holt-Winters methods and a number of variations and extensions to create equivalences to state-space models, normalize seasonals, and cope with problems such as series with a fixed drift, missing observations, irregular updates, planned discontinuities, multiple seasonal cycles (in the same series), and multivariate series. Equivalent regression, ARIMA, and state-space models are reviewed in Section 4. This section also discusses variances, prediction intervals, and some possible explanations for the robustness of exponential smoothing. Procedures for method and model selection are discussed in Section 5, including the use of time series characteristics, expert systems, information criteria, and operational measures. Section 6 reviews the details of model-fitting, including the selection of parameters, initial values, and loss functions. In Section 6, we also discuss the use of adaptive parameters to avoid model-fitting. Applications of exponential smoothing to inventory control with both continuous and intermittent demand are discussed in Section 7. Section 8 summarizes the many empirical studies in which exponential smoothing has been used. Conclusions and an assessment of the state of the art are offered in Section 9. This plan does not include coverage of tracking signals, a subject that has disappeared from the literature since the earlier paper.

2. Early history of exponential smoothing

Exponential smoothing originated in Robert G. Brown's work as an OR analyst for the US Navy during World War II (Gass and Harris, 2000). In 1944, Brown was assigned to the antisubmarine effort and given the job of developing a tracking model 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 firing depth charges from destroyers. Brown's tracking model was essentially simple exponential smoothing of continuous data, an idea still used in modern fire-control equipment.

During the early 1950s, Brown extended simple exponential smoothing to discrete data and developed methods for trends and seasonality. One of his early applications was in forecasting the demand for spare parts in Navy inventory systems. The savings in data storage over moving averages led to the adoption of exponential smoothing throughout Navy inventory systems during the 1950s. In 1956, Brown presented his work on exponential smoothing of inventory demands at a conference of the Operations Research Society of America. This presentation formed the basis of Brown's first book, *Statistical Forecasting for Inventory Control* (Brown, 1959). His second book, *Smoothing, Forecasting, and Prediction of Discrete Time Series* (Brown, 1963), developed the general exponential smoothing methodology. In numerous later books, Brown integrated exponential smoothing with inventory management and production planning and control.

During the 1950s, Charles C. Holt, with support from the Logistics Branch of the Office of Naval Research (ONR), worked independently of Brown to develop a similar method for exponential smoothing of additive trends and an entirely different method for smoothing seasonal data. Holt's original work was documented in an ONR memorandum (Holt, 1957) and went unpublished until recently (Holt, 2004a, 2004b). However, Holt's ideas gained wide publicity in 1960. In a landmark article, Winters (1960) tested Holt's methods with empirical data, and they became known as the Holt-Winters forecasting system. Another landmark article by Muth (1960) was among the first to examine the optimal properties of exponential smoothing forecasts. Holt's methods of exponential smoothing were also featured in the classic text by Holt, Modigliani, Muth, and Simon, *Planning Production, Inventories, and Work Force* (1960), a book that is still in use today in doctoral programs in operations management.

3. Formulation of exponential smoothing methods

Section 3.1 classifies and gives formulations for the standard methods of exponential smoothing. These methods can be modified to create state-space models as discussed in Section 3.2. Seasonal indices are not automatically renormalized in either the standard or state-space versions of exponential smoothing, and procedures for renormalization are reviewed in Section 3.3. In Section 3.4, we collect a number of variations on the standard methods to cope with special kinds of time series.

3.1 Standard methods

Table 1 contains equations for the standard methods of exponential smoothing, all of which are extensions of the work of Brown (1959, 1963), Holt (1957), and Winters (1960). For each type of trend, and for each type of seasonality, there are two sections of equations. The first section gives recurrence forms and the second gives error-correction forms. Recurrence forms were used in the original work by Brown and Holt and are still widely used in practice, but error-correction forms are simpler and give equivalent forecasts. The notation follows Gardner (1985) and is defined in Table 2. It is worth emphasizing that there is still no agreement on notation for exponential smoothing. An appalling variety of notation exists in the literature, and some authors add to the confusion by changing notation from one paper to the next.

Hyndman et al.'s (2002) taxonomy, as extended by Taylor (2003a), is helpful in describing the methods. Each method is denoted by one or two letters for the trend (row heading) and one letter for seasonality (column heading). Method N-N denotes no trend with no seasonality, or simple exponential smoothing (Brown, 1959). The other nonseasonal methods are Holt's (1957) additive trend (A-N), Gardner and McKenzie's (1985) damped additive trend (DA-N), Pegels' (1969) multiplicative trend (M-N), and Taylor's (2003a) damped multiplicative trend (DM-N). The parameters in the trend methods can be constrained using discounted least squares (DLS) to produce special cases often called Brown's methods, as discussed in Section 4.1. All seasonal methods are formulated by extending the methods in Winters (1960). Note that the forecast equations for the seasonal methods are valid only for a forecast horizon (m) less than or equal to the length of the seasonal cycle (p).

Table 1
Standard exponential smoothing methods

Trend	Seasonality **N** None	 **A** Additive	 **M** Multiplicative
N None	$S_t = \alpha X_t + (1-\alpha)S_{t-1}$ $\hat{X}_t(m) = S_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)S_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = \alpha(X_t / I_{t-p}) + (1-\alpha)S_{t-1}$ $I_t = \delta(X_t / S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t I_{t-p+m}$
	$S_t = S_{t-1} + \alpha e_t$ $\hat{X}_t(m) = S_t$	$S_t = S_{t-1} + \alpha e_t$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = S_{t-1} + \alpha e_t / I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t / S_t$ $\hat{X}_t(m) = S_t I_{t-p+m}$
A Additive	$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = \alpha(X_t / I_{t-p}) + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $I_t = \delta(X_t / S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$
	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha\gamma e_t$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha\gamma e_t$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t / I_{t-p}$ $T_t = T_{t-1} + \alpha\gamma e_t / I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t / S_t$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$
DA Damped Additive	$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\phi T_{t-1}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \phi^i T_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\phi T_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \phi^i T_t + I_{t-p+m}$	$S_t = \alpha(X_t / I_{t-p}) + (1-\alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\phi T_{t-1}$ $I_t = \delta(X_t / S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t + \sum_{i=1}^{m} \phi^i T_t)I_{t-p+m}$
	$S_t = S_{t-1} + \phi T_{t-1} + \alpha e_t$ $T_t = \phi T_{t-1} + \alpha\gamma e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \phi^i T_t$	$S_t = S_{t-1} + \phi T_{t-1} + \alpha e_t$ $T_t = \phi T_{t-1} + \alpha\gamma e_t$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \phi^i T_t + I_{t-p+m}$	$S_t = S_{t-1} + \phi T_{t-1} + \alpha e_t / I_{t-p}$ $T_t = \phi T_{t-1} + \alpha\gamma e_t / I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t / S_t$ $\hat{X}_t(m) = (S_t + \sum_{i=1}^{m} \phi^i T_t)I_{t-p+m}$
M Multiplicative	$S_t = \alpha X_t + (1-\alpha)(S_{t-1}R_{t-1})$ $R_t = \gamma(S_t / S_{t-1}) + (1-\gamma)R_{t-1}$ $\hat{X}_t(m) = S_t R_t^m$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)S_{t-1}R_{t-1}$ $R_t = \gamma(S_t / S_{t-1}) + (1-\gamma)R_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t R_t^m + I_{t-p+m}$	$S_t = \alpha(X_t / I_{t-p}) + (1-\alpha)S_{t-1}R_{t-1}$ $R_t = \gamma(S_t / S_{t-1}) + (1-\gamma)R_{t-1}$ $I_t = \delta(X_t / S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t R_t^m)I_{t-p+m}$
	$S_t = S_{t-1}R_{t-1} + \alpha e_t$ $R_t = R_{t-1} + \alpha\gamma e_t / S_{t-1}$ $\hat{X}_t(m) = S_t R_t^m$	$S_t = S_{t-1}R_{t-1} + \alpha e_t$ $R_t = R_{t+1} + \alpha\gamma e_t / S_{t-1}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t R_t^m + I_{t-p+m}$	$S_t = S_{t-1}R_{t-1} + \alpha e_t / I_{t-p}$ $R_t = R_{t-1} + (\alpha\gamma e_t / S_{t-1}) / I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t / S_t$ $\hat{X}_t(m) = (S_t R_t^m)I_{t-p+m}$
DM Damped Multiplicative	$S_t = \alpha X_t + (1-\alpha)(S_{t-1}R_{t-1}^{\phi})$ $R_t = \gamma(S_t / S_{t-1}) + (1-\gamma)R_{t-1}^{\phi}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m}\phi^i}$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)S_{t-1}R_{t-1}^{\phi}$ $R_t = \gamma(S_t / S_{t-1}) + (1-\gamma)R_{t-1}^{\phi}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m}\phi^i} + I_{t-p+m}$	$S_t = \alpha(X_t / I_{t-p}) + (1-\alpha)(S_{t-1}R_{t-1}^{\phi})$ $R_t = \gamma(S_t / S_{t-1}) + (1-\gamma)R_{t-1}^{\phi}$ $I_t = \delta(X_t / S_t) + (1-\delta)I_{t-1}$ $\hat{X}_t(m) = (S_t R_t^{\sum_{i=1}^{m}\phi^i})I_{t-p+m}$
	$S_t = S_{t-1}R_{t-1}^{\phi} + \alpha e_t$ $R_t = R_{t-1}^{\phi} + \alpha\gamma e_t / S_{t-1}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m}\phi^i}$	$S_t = S_{t-1}R_{t-1}^{\phi} + \alpha e_t$ $R_t = R_{t-1}^{\phi} + \alpha\gamma e_t / S_{t-1}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m}\phi^i} + I_{t-p+m}$	$S_t = S_{t-1}R_{t-1}^{\phi} + \alpha e_t / I_{t-p}$ $R_t = R_{t-1}^{\phi} + (\alpha\gamma e_t / S_{t-1}) / I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t / S_t$ $\hat{X}_t(m) = (S_t R_t^{\sum_{i=1}^{m}\phi^i})I_{t-p+m}$

Table 2. Notation for exponential smoothing

Symbol	Definition
α	Smoothing parameter for the level of the series
γ	Smoothing parameter for the trend
δ	Smoothing parameter for seasonal indices
ϕ	Autoregressive or damping parameter
β	Discount factor, $0 \le \beta \le 1$
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 additive trend at the end of period t
R_t	Smoothed multiplicative trend at the end of period t
I_t	Smoothed seasonal index at the end of period t. Can be additive or multiplicative
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-step-ahead forecast error, $e_t = X_t - \hat{X}_{t-1}$ (1). Note that $e_t(m)$ should be used for other forecast origins
C_t	Cumulative renormalization factor for seasonal indices. Can be additive or multiplicative
V_t	Transition variable in smooth transition exponential smoothing
D_t	Observed value of nonzero demand in the Croston method
Q_t	Observed inter-arrival time of transactions in the Croston method
Z_t	Smoothed nonzero demand in the Croston method
P_t	Smoothed inter-arrival time in the Croston method
Y_t	Estimated demand per unit time in the Croston method (Z_t/P_t)

There are several differences between Table 1 and the tables of equations in Gardner (1985). First, the DA methods are given in recurrence forms that were not included in the earlier paper. Second, the seasonal DA methods were formulated with three parameters in the earlier paper, but the same methods in Table 1 contain four parameters as developed in Gardner and McKenzie (1989). Finally, the DM methods are new.

The DA-N method can be used to forecast multiplicative trends with the autoregressive or damping parameter ϕ restricted to the range $1 < \phi < 2$, a method sometimes called "generalized Holt." As Taylor (2003a) observed, generalized Holt is a clumsy way to model a multiplicative trend because the local slope is estimated by smoothing successive differences of the local level. In contrast, Pegels' multiplicative trends (M-N, M-A, and M-M) estimate the local growth rate by smoothing successive ratios of the local level. In hopes of producing more robust forecasts, Taylor's methods (DM-N, DM-A, and DM-M) add a damping parameter $\phi < 1$ to Pegels' multiplicative trends.

Although many new models underlying exponential smoothing have been proposed since 1985, the damped multiplicative trends are the only new methods in the sense that they create new forecast profiles. Like the damped additive trends, the forecast profiles for Taylor's methods will eventually approach a horizontal nonseasonal or seasonally-adjusted asymptote. However, in the near term, different values of ϕ can be used to produce forecast profiles that are convex, nearly linear, or even concave.

3.2 State-space equivalent methods

There are many equivalent state-space models for each of the methods in Table 1. Here we review the particular modeling framework of Hyndman et al. (2002) that includes all methods in Table 1 except the DM methods. In this framework, each exponential smoothing method has two corresponding state-space models, each with a single source of error (SSOE). One model has an additive error and the other has a multiplicative error. As discussed in Section 4.3, if the parameters are the same, the two models give the same point forecasts but different variances. The methods corresponding to the Hyndman et al. framework are the same as those in Table 1 with two exceptions: we must modify all multiplicative seasonal methods and all damped additive-trend methods.

We proceed as follows to modify the multiplicative seasonal methods. In the N-M standard equations for updating the multiplicative seasonal component I_t, replace the smoothed level S_t with S_{t-1}. This change is made in both recurrence and error-correction forms. In the A-M, DA-M, and M-M standard equations for updating I_t, replace S_t with $S_{t-1} + T_{t-1}$, where T_{t-1} is the previous smoothed trend, again in both recurrence and error-correction forms. In the DA-M method, note that the seasonal state-space modification does not damp T_{t-1} in updating I_t.

Koehler et al. (2001) present several other state-space versions of the A-M method, all with the same multiplicative seasonal modification. One precedent for this modification is found in Williams (1987), who shows that it allows us to update each component independently, while in the standard method the new smoothed level is used in updating the other components. Archibald (1990) made the same point without reference to the work of Williams. Perhaps another reason to use the multiplicative seasonal modification is that, as Ord (2004) observed, this was done in Holt's original work (1957). However, Holt et al. (1960) and Winters (1960) discarded this idea and used the standard equations in Table 1.

What are the practical consequences of adopting the state-space versions of the multiplicative seasonal methods? The answer to this question awaits empirical study. In an analysis of the A-M method, Koehler et al. (2001) show that the difference between the two versions of the equation for updating the seasonal component will be small, provided that all three smoothing parameters are less than about 0.3. However, Koehler et al. warn that negative seasonal components can occur in the state-space version of A-M unless the forecast errors are much less variable than the data.

To make the damped additive (DA) methods fit the Hyndman et al. (2002) framework, we begin with the level equations. The equivalent state-space model does not damp the previous trend in the level equations, so we delete ϕ (replace ϕT_{t-1} with T_{t-1}). Next, the forecast equations must be changed to begin damping at two steps ahead, rather than immediately as in Table 1. The forecast equation in the nonseasonal state-space equivalent method (DA-N) is:

$$\hat{X}_t(m) = (S_t + \sum_{i=0}^{m-1} \phi^i T_t) \qquad (1)$$

At first glance, it looks as if the state-space DA method will always extrapolate more trend at any horizon than the standard method, but this may not be true if fitted parameter values differ substantially between the two versions. Given the success of the standard DA method, it is difficult to understand why Hyndman et al. (2002) chose to start trend damping at two steps ahead. There appears to be no statistical reason for this choice. In contrast to Hyndman et al., the text by Bowerman et al. (2005) includes a comprehensive treatment of state-space models for exponential smoothing in which trend damping starts immediately.

3.3 Renormalization of seasonal indices

The standard seasonal methods are initialized so that the average seasonal index is 0 (additive) or 1 (multiplicative); thereafter, normalization goes astray because only one seasonal index is updated each period. The problem of renormalization was overlooked in Gardner (1985) and there has been much confusion in the literature about whether it is necessary to renormalize the seasonal indices, and if so when and how this should be done.

Lawton (1998) analyzed an equivalent state-space model for the A-A method and reached several conclusions. First, if seasonal indices in the A-A method are not renormalized, estimates of trend are correct although estimates of level and seasonals are biased. Fortunately, the errors in estimating level and seasonals are counter-balancing and do not impact the forecasts. If renormalization of seasonal indices alone is carried out, a very common procedure in practice, this must be done at every time period or the forecasts will be biased until the A-A equations have sufficient time to adjust the level. Lawton gives an example in which the bias is serious compared to not renormalizing. If we choose to renormalize at an interval other than every time period, the procedure is as follows: (1) subtract a constant value from each seasonal index to force the sum to zero, and (2) add the same constant to the level.

Other A-A renormalization equations are found in Roberts (1982), McKenzie (1986), and Newbold (1988). These authors go about renormalization in different ways, but their point forecasts turn out to be the same. Furthermore, the point forecasts from all of the alternative sets of renormalization equations are the same as the point forecasts from the standard equations. McKenzie showed that the link between the standard and renormalized versions of the A-A method is very simple. The two methods give equivalent forecasts if we replace the level parameter α in the standard error-correction form with $(\alpha - \delta/p)$, where δ is the smoothing parameter for the seasonal component and p is the number of periods in one season. This parameter adjustment should occur automatically during model-fitting.

Prior to Archibald and Koehler (2003), the only renormalization equations for the A-M method were those of Roberts (1982) and McKenzie (1986). Archibald and Koehler found that the Roberts and McKenzie equations result in point forecasts that differ from each other and also from the standard-equation forecasts. Therefore, Archibald and Koehler set out to make sense of the renormalization problem. First, they developed new renormalization equations for the A-M method that give the same point forecasts as the standard equations. Second, they developed analogous A-A renormalization equations. Finally, they derived equations that compute cumulative renormalization correction factors for the A-A and A-M methods. These correction factors should prove to be popular in practice because they allow the user to keep the standard equations and renormalize the seasonal indices at any point in time.

The cumulative renormalization correction factor C_t for the A-A method is computed iteratively using a simple equation:

$$C_t = C_{t-1} + \delta e_t / p \,, \tag{2}$$

To renormalize at any time, add C_t to the level and subtract it from each seasonal index.

Archibald and Koehler derived the cumulative renormalization correction factor for the A-M method using the state-space version. Here we give the correction factor for the standard A-M version in Table 1:

$$C_t = C_{t-1}(1 + \delta e_t / pS_t) \tag{3}$$

To renormalize at any time, multiply level and trend by C_t and divide each seasonal index by C_t. If the state-space A-M version is used, replace S_t with $S_{t-1} + T_{t-1}$ in equation (3).

To sum up the research in this area, there is no reason to renormalize in the additive seasonal methods if forecast accuracy is the only concern. It is not known whether renormalization can safely be ignored in the multiplicative methods. If we choose to renormalize to provide reliable estimates of either additive or multiplicative model components, the simplest approach is to apply the correction factors of Archibald and Koehler. Their A-A and A-M correction factors are easily extended to the other seasonal methods in either standard or state-space form.

There is little empirical evidence on the problem of renormalization. The only reference is Archibald and Koehler (2003), who tested the 401 monthly series from the M1-Competition (Makridakis et al., 1982). In 12 series, they found that the standard A-A method produced values of level, trend, and seasonal indices that were off more than 5% (compared to renormalized A-A).

3.4 Other variations on the standard methods

This section collects special versions of the standard Holt-Winters methods to cope with missing or irregular observations, irregular update intervals, planned discontinuities, series containing a fixed drift, and series containing two or more seasonal cycles. We can also simplify the A-A method by merging the level and seasonal components, and adapt several methods to multivariate series. Discussion of one additional variation, Croston's (1972) method for inventory series with intermittent demands, is deferred until Section 7 on inventory control.

For missing observations, Wright's (1986a, 1986b) solution is straightforward. Missing observations receive zero weight, while the others are exponentially weighted according to the age of the observation. Wright gives modified formulas for the N-N and A-N methods that automatically adjust the weighting pattern for all observations following a gap. These formulas also work for the equivalent problem of observations that naturally occur at irregular time intervals. Wright's procedure was extended by Cipra et al. (1995) to seasonal methods. Although Wright's procedure looks sensible, Aldrin and Damsleth (1989) made a complaint that has been repeated many times in the literature of exponential smoothing, that the procedure is *ad hoc* with no statistical rationale. Aldrin and Damsleth developed an elaborate alternative procedure that computes optimal weights in the equivalent ARIMA models. It is not clear that the ARIMA procedure is worth the trouble because the authors analyzed two time series and got about the same results as the Wright procedure.

If the time between updates of the N-N method is irregular, the data for several periods may be reported as a combined observation. Obviously, the smoothing parameter should be increased to give more weight to combined observations. Johnston (1993) derived a formula for optimal adjustment of the smoothing parameter, while Anderson (1994) and Walton (1994) derived simpler alternative formulas. Anderson's idea is the simpler of the two, and gives values very close to Johnston's optimal formula. When a combined observation occurs, in the N-N equation we replace X_t with X_t / k, where k is the number of periods combined, and we replace α with the expression $1-(1-\alpha)^k$. This adjustment assumes that the data are spread evenly over the combined periods.

There may be planned discontinuities in a time series. For example, we may expect a disruption in demand following a price change or a new product introduction. There are three ways of dealing with planned discontinuities in exponential smoothing. If discontinuities are recurring, Carreno and Madinaveitia (1990) add an index similar to a seasonal index to the A-N method to model the effects. When the effects of discontinuities cannot be estimated from history, judgmental adjustments to the forecasts are usually necessary. Williams and Miller

(1999) recommend making such adjustments within the exponential smoothing method rather than as a second-stage correction outside the method. The basic idea is to add an adjustment factor to the forecast equation and otherwise allow the updating equations to operate normally. If the discontinuity occurs approximately as planned, this will be more accurate than making second-stage corrections outside the method. It may be possible to express planned discontinuities as a set of linear restrictions on the forecasts from a linear exponential smoothing method. If so, Rosas and Guerrero (1994) show that one can compute weights that meet the restrictions in the moving-average representation of the equivalent ARIMA model. Once this is done, the weights can be converted into smoothing parameters.

The N-N method can be enhanced by adding a drift (fixed trend) term, making the method equivalent to the "Theta method of forecasting" (Assimakopoulos and Nikolopoulos, 2000) that performed well in the M3 competition (Makridakis and Hibon, 2000). In a mathematical tour de force, Hyndman and Billah (2003) showed that the Theta method is the same thing as simple smoothing with drift equal to half the slope of a linear trend fitted to the data. Another way to match the Theta method is to use the same drift choice in the A-N method with the trend parameter set to zero. We do not know why the particular drift choice in the Theta method or its equivalents is better than any other, nor is it clear when one should prefer a fixed drift over a smoothed trend.

For time series containing two seasonal cycles, Taylor (2003b) adds one more seasonal component to the A-M method. The new method, called double seasonal exponential smoothing, was applied to electricity demand recorded at half-hour intervals, with one seasonal equation for a within-day seasonal cycle and another for a within-week cycle. As often happens in complex time series forecasted with exponential smoothing, Taylor found significant first-order autocorrelation in the residuals. Thus he fitted an AR(1) model to remove it, estimating the

AR(1) parameter at the same time as the smoothing parameters. The resulting forecasts outperformed those from the standard A-M method as well as a double seasonal ARIMA model.

Rather than add a seasonal component, Snyder and Shami (2001) eliminate it from the A-A method. The seasonal component is incorporated into the level, which depends on the level a year ago and is augmented by the total growth in all seasons during the past year. Thus their parsimonious method requires only two parameters. Snyder and Shami found that the two-parameter version of A-A was less accurate than the standard three-parameter version, although the differences were not statistically significant. Snyder and Shami overlooked a simpler way to reduce the number of parameters in any of the trend and seasonal methods, which is to apply DLS, as discussed in Section 4.1.

Some of the univariate methods in Table 1 have been generalized to the multivariate case by Jones (1966), Enns et al. (1982), Harvey (1986), and Pfefferman and Allon (1989). For the N-N method, Jones and Enns et al. simply replaced the scalars with matrices. In error-correction form, the multivariate version of N-N is then:

$$\mathbf{S}_t = \mathbf{S}_{t-1} + \boldsymbol{\alpha}\mathbf{e}_t \tag{4}$$

With k series, the dimensions of $\mathbf{S}_t$, $\mathbf{S}_{t-1}$, and $\mathbf{e}_t$ are $k \times 1$, and the dimension of $\boldsymbol{\alpha}$ is $k \times k$. Enns et al. assume that the series are produced by a multivariate random walk and estimate the parameters by a complex maximum likelihood procedure. Harvey achieved a profound simplification by proving that one can forecast the individual series using univariate methods. The univariate parameters are chosen by a grid search to minimize the sum of vector products of the one-step-ahead errors, a procedure that approximates maximum-likelihood estimation. Harvey also developed multivariate models with trend and seasonal components. Again, we replace the scalars with matrices in the error-correction forms of the univariate trend and seasonal methods, with parameters chosen in the same way as for multivariate N-N. Pfefferman and Allon analyzed the multivariate A-A method and derived several structural models that produce optimal forecasts. Unlike the earlier multivariate research, Pfefferman and Allon give detailed instructions for initialization and model-fitting. Pfefferman and Allon also present what appears to be the only empirical evidence on multivariate exponential smoothing. In forecasting two bivariate time series of Israeli tourism data, multivariate A-A was significantly more accurate than univariate A-A.

4. Properties

Each exponential smoothing method in Table 1 is equivalent to one or more stochastic models. The possibilities include regression, ARIMA, and state-space models, as discussed in Sections 4.1 – 4.3. The associated research on variances and prediction intervals is discussed in Section 4.4. The most important property of exponential smoothing is robustness, reviewed in Section 4.5. Discussion of the property of invertibility is deferred until Section 6.1 on parameter selection. The theoretical relationships between judgmental forecasting and exponential smoothing are beyond our scope, although we note in passing that several exponential smoothing methods are treated as models of judgmental extrapolation (see for example Andreassen and Kraus, 1990).

4.1 Equivalent regression models

In large samples, exponential smoothing is equivalent to an exponentially-weighted or DLS regression model. Simple smoothing corresponds to DLS with discount factor $\beta = 1 - \alpha$. In the other methods, DLS reduces the number of parameters, an advantage in large inventory control systems (Gardner, 1990). In the A-N method, DLS produces a special case known as double exponential smoothing (Brown, 1963) that has only one parameter. To construct double exponential smoothing from the error-correction form of the A-N method, proceed as follows. In the level equation, replace the multiplier for the error α with $\alpha(2-\alpha)$; in the trend equation, replace the multiplier for the error $\alpha\gamma$ with α^2. In the error-correction form of the DA-N method, Gardner and McKenzie (1985) show that DLS reduces the number of parameters from three to two as follows. First, replace the multiplier for the error α with $1-(\beta/\phi)^2$. Next, in the trend equation, replace the multiplier for the error $\alpha\gamma$ with $(1-\beta/\phi)(1-(\beta/\phi^2))$.

General exponential smoothing (GES) (Brown, 1963) also relies on DLS regression with either one or two discount factors to fit a variety of functions of time to the data, including polynomials, exponentials, sinusoids, and their sums and products. A detailed review of GES is available in Gardner (1985), and since that time only a few papers on the subject have appeared. Gijbels et al. (1999) and Taylor (2004c) showed that GES can be viewed in a kernel regression framework. Gijbels et al. found that simple smoothing (N-N) is actually a zero-degree local polynomial kernel model. A normal kernel is commonly used in kernel regression, but Gijbels et al. used an exponential kernel to show the equivalence to simple smoothing. The main conclusion in this paper is that choosing the minimum-MSE parameter in simple smoothing is equivalent to choosing the regression bandwidth by cross-validation, a procedure that divides the data into two disjoint sets, with the model fitted in one set and validated in another. We note that previous research has established theoretical support for choosing minimum-MSE parameters in exponential smoothing (see for example, Harvey, 1984). Gijbels et al. went on to suggest extensions to trends and seasonality although the details are unpleasant.

Taylor (2004c) proposed another type of kernel regression, an exponentially-weighted quantile regression (EWQR). The rationale for EWQR is that it is robust to distributional assumptions. EWQR turns out to be equivalent to simple exponential smoothing of the cumulative density function, which is the inverse of the quantile function. We can also think of EWQR as an extension of GES to quantiles. Just as DLS delivers exponential smoothing for the mean, EWQR delivers the analogy for quantiles. Unlike Gijbels et al., Taylor gives empirical results. Using a collection of 256 deseasonalized time series of daily supermarket sales, Taylor experimented with a number of likely methods for generating point forecasts from the quantiles. He discovered that the best methods were a weighted average of the forecasts of the median and the 0.33 and 0.67 quantiles, and simple exponential smoothing of a time series created by trimming all observations below the 0.25 quantile and above the 0.75 quantile. These methods produced forecasts that were more accurate than a variety of standard exponential smoothing methods as well as a modification of simple smoothing developed by the supermarket company. Taylor also presented versions of EWQR with linear and damped trends and with seasonal terms, although these enhancements were not necessary for the data analyzed. A special case of EWQR was developed by Cipra (1992), who extended GES to the median by replacing the DLS criterion with discounted least absolute deviations.

The only other GES research since 1985 is by Bartolomei and Sweet (1989), who compared GES to the A-A and A-M methods using 47 time series from the M1 competition. The authors found little difference in forecast accuracy, although they speculated that one of the damped-trend methods might have done better.

4.2 Equivalent ARIMA models

All linear exponential smoothing methods have equivalent ARIMA models. The easiest way to see the nonseasonal models is through the DA-N method, which contains at least six ARIMA models as special cases (Gardner and McKenzie, 1988). If $0 < \phi < 1$, the DA-N method is equivalent to the ARIMA (1, 1, 2) model, which can be written as:

$$(1-B)(1-\phi B)X_t = [1-(1+\phi-\alpha-\phi\alpha\gamma)B-\phi(\alpha-1)B^2]e_t \qquad (5)$$

We obtain an ARIMA (1, 1, 1) model by setting $\alpha = 1$. With $\alpha = \gamma = 1$, the model is ARIMA (1, 1, 0). When $\phi = 1$, we have a linear trend (A-N) and the model is ARIMA (0, 2, 2):

$$(1-B)^2 X_t = [1-(2-\alpha-\alpha\gamma)B-(\alpha-1)B^2]e_t \qquad (6)$$

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

$$(1-B)X_t = [1-(1-\alpha)]e_t \qquad (7)$$

The ARIMA (0, 1, 0) random walk model can be obtained from (7) by choosing $\alpha = 1$. ARIMA-equivalent seasonal models for the linear exponential smoothing methods exist, although most are so complex that it is unlikely they would ever be identified through the Box-Jenkins methodology.

4.3 Equivalent state-space models

The equivalent ARIMA models do not extend to the nonlinear exponential smoothing methods. The only statistical rationale for exponential smoothing that includes nonlinear methods is due to Ord et al. (1997). Prior to this work, state-space models for exponential smoothing were formulated using multiple sources of error (MSOE). For example, simple exponential smoothing (N-N) is optimal for a model with two sources of error (Muth 1960). The observation and state equations are written:

$$X_t = \ell_t + \nu_t \tag{8}$$

$$\ell_t = \ell_{t-1} + \eta_t \tag{9}$$

The unobserved state variable ℓ_t denotes the local level at time t, and the error terms ν_t and η_t are generated by independent white noise processes. Using different methods, various authors (Nerlove and Wage, 1964; Theil and Wage, 1964; Harrison, 1967; Chatfield, 1996) showed that simple smoothing is optimal with α determined by the ratio of the variances of the noise processes. Harvey (1984) also showed that the Kalman filter for (8) and (9) reduces to simple smoothing in the steady state.

For the trend and seasonal versions of exponential smoothing, the MSOE models are complex, as demonstrated in Proietti (1998, 2000), who gives examples of models that are equivalent to linear versions of exponential smoothing. Another limitation of the MSOE approach is that researchers have been unable to find such models that correspond to multiplicative-seasonal versions of exponential smoothing. In response to these problems, Ord et al. (1997) built on the work of Snyder (1985) to create a general, yet remarkably simple class of state-space models with a single source of error (SSOE). For example, the SSOE model with additive errors for the N-N method is written as follows:

$$X_t = \ell_{t-1} + \varepsilon_t \tag{10}$$

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t \tag{11}$$

Note that the observation equation (10) includes ℓ_{t-1} rather than ℓ_t as in equation (8) of the MSOE model. The error term ε_t in the observation equation is then the one-step-ahead forecast error assuming knowledge of the level at time $t-1$. The correspondence to simple smoothing is seen in the state equation (11), which is the error-correction form of simple smoothing in Table 1, except that the level ℓ is substituted for the smoothed level S.

For the multiplicative-error N-N model, we alter the additive-error SSOE model as follows:

$$X_t = \ell_{t-1} + \ell_t \varepsilon_t \tag{12}$$

$$\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t) = \ell_{t-1} + \alpha\ell_{t-1}\varepsilon_t \tag{13}$$

In this case, the one-step-ahead forecast error is still $X_t - \ell_{t-1}$, but it is no longer the same as ε_t. The state equation (13) becomes:

$$\ell_t = \ell_{t-1} + \alpha\ell_{t-1}\left(\frac{X_t - \ell_{t-1}}{\ell_{t-1}}\right) = \ell_{t-1} + \alpha(X_t - \ell_{t-1}) \tag{14}$$

Thus we have shown that the multiplicative-error state equation can be written in the error-correction form of simple smoothing. It follows that the state equations are the same in the additive- and multiplicative-error cases, and this is true for all SSOE models.

Following similar logic, Hyndman et al. (2002) extended Ord et al.'s class of SSOE models to include all the methods of exponential smoothing in Table 1 except the DM methods. Because the state equations for all models are the same as the error-correction forms of exponential smoothing (with modifications as discussed in Section 3.2), the observation equations are obvious. In the Hyndman et al. framework, there are 12 basic models, each with additive or multiplicative errors, in effect giving 24 models in total. Hyndman et al. (2002) remark that the additive- and multiplicative-error models give the same point forecasts, but this is true only if the same parameters are found during model-fitting, an improbable occurrence.

The additive-error models are usually fitted to minimize squared errors, but the multiplicative-error models are fitted to minimize squared relative errors, where the errors are relative to the one-step-ahead forecasts rather than the data.

The theoretical advantage of the SSOE approach to exponential smoothing is that the errors can depend on the other components of the time series. As an illustration, consider the N-N method/model. For the additive-error version, the variance of the one-step-ahead forecast errors is $Var(\varepsilon_t) = \sigma^2$, while the variance for the multiplicative-error model changes with the level component, that is $Var(\ell_{t-1}\varepsilon_t) = \ell_{t-1}^2\sigma^2$. In the more complex models, multiplicative-error effects can be profound because the variance changes with every component of the time series (level, trend, and seasonality).

To put the theoretical advantage of the SSOE approach another way, each of the linear exponential smoothing models with additive errors has an ARIMA equivalent. However, the linear models with multiplicative errors and the nonlinear models are beyond the scope of the ARIMA class. As Koehler et al. (2001) and Hyndman et al. (2002) observed, their state-space models are not unique and many other such models could be formulated. Some additional possibilities are discussed in Chatfield et al. (2001), the most readable reference on the state-space foundation for exponential smoothing.

The only theoretical criticism of the SSOE approach appears to be an *OR Viewpoint* by Johnston (2000) on a paper by Snyder et al. (1999) discussed in Section 7 on inventory control. Johnston's primary argument is that the SSOE model for simple smoothing is not really a model at all, and should be viewed instead as nothing more than an estimation procedure. However, Snyder et al. (2000) pointed out that the SSOE and MSOE models for simple smoothing are both special cases of a more general state-space model. For additional discussion of the theoretical relationships amongst these models, see Harvey and Koopman (2000) and Ord et al. (2005).

4.4 Variances and prediction intervals

Variances and prediction intervals for point forecasts from exponential smoothing can be computed using either empirical or analytical procedures. Empirical procedures are available in Gardner (1988) and Taylor and Bunn (1999). Because post-sample forecast errors are usually much larger than fitted errors, I used the Chebyshev distribution to compute probability limits from DA-N fitted errors at different forecast horizons. For data from the M1 competition, coverage percentages were very close to targets. Nevertheless, Chatfield and Yar (1991) complained that this procedure often results in constant variance as the lead time increases, while Chatfield (1993) observed that the intervals are sometimes too wide to be of practical use. These criticisms do not apply to the work of Taylor and Bunn, who proposed another way to avoid a normality assumption. They used quantile regression on the fitted errors to obtain prediction intervals that are functions of forecast lead time as suggested by theoretical variance expressions. For the N-N, A-N, and DA-N methods, Taylor and Bunn obtained excellent results in both simulated and M1 data.

Analytical prediction intervals can be computed in several different ways. The wrong way to do so is to use $s\sqrt{m}$ as the standard deviation of m-step-ahead forecast errors, where s is the standard deviation of the one-step-ahead errors. This expression has been used in the literature for various exponential smoothing methods but is correct only when the optimal model is a random walk. For other models, the expression can be seriously misleading, as discussed in Koehler (1990), Yar and Chatfield (1990), and Chatfield and Koehler (1991). The expression

$s\sqrt{m}$ has also been used for the standard deviation of cumulative lead time demand *m* steps ahead, but this is also wrong as discussed in Section 7.1.

The simplest analytical approach to variance estimation is based on the assumption that the series is generated by deterministic functions of time (plus white noise) that are assumed to hold in a local segment of the time series. Brown (1963) followed this approach in deriving variances for the N-N method, the double smoothing version of A-N, and the GES methods. Sweet (1985) and McKenzie (1986) also followed this approach in deriving variances for the A-A and A-M methods. However, Newbold and Bos (1989) called the use of deterministic functions of time “grossly inaccurate” in criticizing the work of Brown, Sweet, McKenzie, Gardner (1983, 1985), and many other authors. Newbold and Bos state that “any amount” of empirical evidence supports their criticism, although it is curious that they give no references. There is no such empirical evidence in the references listed below, or in the references to Gardner (1985).

For the A-A method, an analytical variance expression was derived by Yar and Chatfield (1990), who assumed only that one-step-ahead errors are uncorrelated. But for this to be true, the equivalent ARIMA model must be optimal. Thus Yar and Chatfield’s variance expression turns out to be the same as that of the equivalent ARIMA model. In a follow-on study, Chatfield and Yar (1991) found an approximate formula for the A-M method, again by assuming that the one-step-ahead errors are uncorrelated. In contrast to the additive case, they showed that the width of the multiplicative prediction intervals depends on the time origin and can change with seasonal peaks and troughs.

For the SSOE state-space models, there are numerous recent papers containing variance results that can be sorted out as follows. Empirical procedures for variance estimation, including bootstrapping and simulation from an assumed model, in both cases with either additive or multiplicative errors, are found in Ord et al. (1997), Snyder (2002), Snyder et al. (1999, 2002, 2004), and Hyndman et al. (2002). Analytical variance expressions for various models, with prediction intervals computed from the normal distribution, are found in Ord et al. (1997), Koehler et al. (2001), Snyder et al. (1999, 2001, 2002, 2004), and Hyndman et al. (2005b). We can also classify the papers according to whether they deal with the variance around cumulative or point forecasts. Variances for cumulative forecasts are found in Snyder (2002) and Snyder et al. (2001, 2002, 2004), and are most used in inventory control, as discussed in Section 7.1, while the other papers deal with point forecasts. Hyndman et al. (2005b) is an extremely valuable reference because it contains all known results for variances and prediction intervals around point forecasts. The models are divided into three classes. The first class includes linear models with additive errors and ARIMA equivalents, corresponding to the N-N, A-N, DA-N, N-A, A-A, and DA-A methods. The second class includes the same models, but now the errors are assumed to be multiplicative to enable the variance to change with the level and trend of the time series. In the third class, including the N-M, A-M, and DA-M methods, the variance changes with level, trend, and the multiplicative seasonal pattern. Equations for some of the exact prediction intervals are tedious, so Hyndman et al. give handy approximations. Note that a few state-space models are not included in the Hyndman et al. (2005b) classification and may prove to be intractable.

Thus, for most state-space models, we have four options for prediction intervals. They can be empirical or analytical, and each type can have additive or multiplicative errors. There is no guidance on how one should choose from these options. Hyndman et al. (2005b) do not test their analytical prediction intervals with real data, so there is no way to compare performance to the empirical results in earlier papers. Because of the normality assumption, the analytical prediction intervals will almost certainly prove to be too narrow. This was also the case with their empirical prediction intervals in the M1 and M3 data (Hyndman et al., 2002).

4.5 Robustness

The equivalent models help explain the general robustness of exponential smoothing, although there are other possible explanations for the performance of several methods. For the DA-N method, the process of computing minimum-MSE parameters is an indirect way to identify a more specific model from the special cases it contains. For the A-A method, a simulation study by Chen (1997) showed that forecast accuracy was not sensitive to the assumed data generating process. It seems reasonable to assume that Chen's conclusion applies to the other additive seasonal methods.

Simple smoothing (N-N) is certainly the most robust forecasting method and has performed well in many types of series not generated by the equivalent ARIMA (0, 1, 1) process. Such series include the very common first-order autoregressive processes and a number of lower-order ARIMA processes (Cogger, 1973; Cohen, 1963; Cox, 1961; Pandit and Wu, 1974; Tiao and Xu, 1993). Bossons (1966) showed that simple smoothing is generally insensitive to specification error, especially when the mis-specification arises from an incorrect belief in the stationarity of the generating process. Related work by Hyndman (2001) shows that ARIMA model selection errors can inflate MSEs compared to simple smoothing. Hyndman simulated time series from an ARIMA (0, 1, 1) process and fitted a restricted set of ARIMA models of order (0, 1, 1), (1, 1, 0), and (1, 1, 1), each with and without a constant term. The best model was selected using Akaike's Information Criterion (AIC) (Akaike, 1970). The ARIMA forecast MSEs were significantly larger than those of simple smoothing due to incorrect model selections, a problem that became worse when the errors were non-normal.

Simple smoothing has done especially well in forecasting aggregated economic series with relatively low sampling frequencies. Rosanna and Seater (1995) show that such series often can be approximated by an ARIMA (0, 1, 1) process. This finding has been misinterpreted by some researchers. The series examined by Rosanna and Seater were not generated by an ARIMA (0, 1, 1) process. The series were sums of averages over time of data generated more frequently than the reporting interval. The effects of averaging and temporal aggregation were to destroy information about the generating process, producing series for which the ARIMA (0, 1, 1) process was merely an artifact. Much the same problem can occur in company-level data. For example, simple exponential smoothing was a very competitive method in Schnaars' (1986) study of annual unit sales series for a variety of products.

Satchell and Timmerman (1995) give a different explanation for the performance of simple smoothing in economic time series. In Muth (1960), simple smoothing was shown to be equivalent to a random walk with noise model, assuming that the process began an infinite number of periods ago. Satchell and Timmerman re-examined this model and derived an explicit formula for weights when the time series has a finite history. They found that exponentially declining weights are surprisingly robust as long as the ratio of the variance of the random walk process to the variance of the noise component is not exceptionally small.

5. Method selection

The definitions of aggregate and individual method selection in the work of Fildes (1992) are useful in exponential smoothing. Aggregate selection is the choice of a single method for all time series in a population, while individual selection is the choice of a method for each series. In commentary on the M-3 competition, Fildes (2001) summed up the state of the art in time series method selection: In aggregate selection, it is difficult to beat the damped-trend version of exponential smoothing. In individual selection, it may be possible to beat the damped trend, but it is not clear how one should proceed. The evidence reviewed below supports this judgment, and the research on individual selection of exponential smoothing methods is best described as inconclusive.

Individual method selection can be done in a variety of ways. In Section 5.1, we review method selection using time series characteristics. Most such procedures were not specifically designed for exponential smoothing, and we consider here only those that include exponential smoothing in some form as a candidate method. Section 5.2 reviews expert systems that include some form of exponential smoothing. The most sophisticated approach to method selection is through information criteria in Section 5.3, while the most tedious approach is through projected operational or economic benefits in Section 5.4. Finally, in Section 5.5, we briefly consider the problems in model identification as opposed to method selection, an area where almost nothing has been done. Most studies on method selection do not include benchmark results, that is comparisons were not made to likely alternatives. This problem is noted in the discussion, and benchmark DA-N and DA-M results are given when available. The question of whether out-of-sample criteria should be used for method selection is beyond our scope – see Tashman (2000) for a review.

5.1 Time series characteristics

Method-selection procedures using time series characteristics have been proposed by Gardner and McKenzie (1988), Shah (1997), and Meade (2000). The aim of the Gardner-McKenzie procedure is not to improve accuracy but to avoid fitting a damped trend when simpler methods serve just as well. Method selection rules are given in Table 3:

Table 3
Method selection rules

Case	Series yielding minimum variance	Method
A	X_t	N-N
B	$(1-B)X_t$	DA-N
C	$(1-B)^2 X_t$	A-N
D	$(1-B^p)X_t$	N-M
E	$(1-B)(1-B^p)X_t$	DA-M
F	$(1-B^2)(1-B^p)X_t$	A-M

In the first Case, the N-N method is recommended because we should not allow a trend or seasonal pattern if differencing serves only to increase variance. In Case B, the DA-N method is recommended because it is equivalent to an ARIMA process with a difference of order 1 (see Section 4.2). Although the N-N method is also equivalent to an ARIMA process with a difference of order 1, the DA-N method is suggested for reasons of robustness. In Case C, the A-N method is justified by its equivalence to an ARIMA process with a difference of order 2. A multiplicative trend is another possibility in Case C, although Gardner and McKenzie argue that such trends are dangerous in automatic forecasting systems (later evidence, such as Taylor, 2003a, suggests otherwise). In Cases D, E, and F, a seasonal method is called for because a seasonal difference reduces variance. For reasons of simplicity, only multiplicative seasonality is considered. Using M1-competition data, the Gardner-McKenzie procedure identified simpler methods than the damped trend about 40% of the time. Forecast accuracy was slightly better than the DA-N method applied to all nonseasonal series, with the DA-M method applied to all seasonal series.

The Gardner-McKenzie procedure was tested by Tashman and Kruk (1996), who made comparisons to two alternatives, a condensed version of rule-based forecasting (discussed in Section 5.2) and selection using the Bayesian Information Criterion (BIC) (Schwarz, 1978). Tashman and Kruk tested these procedures using 103 short, annual time series (Schnaars, 1986), and 29 series (6 quarterly and 23 monthly) from the M2 competition (Makridakis et al, 1993). Tashman and Kruk's results are complex but the main conclusions can be summarized as follows. There was little agreement among the selection procedures about the best method for many time series. Gardner-McKenzie and rule-based forecasting gave similar accuracy that was better than the BIC. All three procedures had trouble differentiating between appropriate and inappropriate applications of both the damped trend and simple smoothing. For example, the damped trend performed better in series containing strong trends than in series for which the damped trend was deemed appropriate.

Taylor (2003a) also obtained somewhat disconcerting results with the Gardner-McKenzie procedure. In tests using the 1,428 monthly series from the M3 competition, the damped trend generally did well in series containing strong trends. However, some series that were clearly trending were classified as stationary due to high levels of variance.

Shah (1997) proposed method selection based on discriminant analysis of descriptive statistics for individual series. Using a sample of series from the population of interest, the first step in Shah's procedure is to fit all candidate methods and compute *ex post* forecast accuracy results. Next, we estimate discriminant scores from standard statistics such as autocorrelations and coefficients of skewness and kurtosis. The sample accuracy results are combined with discriminant scores to determine the best method for each series in the population. Shah found that his procedure identified methods significantly more accurate than use of the same method for all time series, a conclusion that is difficult to generalize because of the limited range of methods and data considered. Shah used only three candidate methods (N-N, A-M, and Harvey's basic structural model) and applied them only to the quarterly time series in the M1 collection. It would be helpful to have discriminant analysis results when selection is made from a larger group of candidate methods such as that in Table 1.

The most exhaustive study of method selection, and a paradigm of research design in comparative methods, is found in Meade (2000), whose candidates included two naïve methods, a deterministic trend, the robust trend of Fildes (1992), methods selected automatically from the ARIMA and ARARMA classes, and three exponential smoothing methods (N-N, A-N, and DA-N) applied to seasonally-adjusted data when appropriate. Meade simulated time series from a wide range of ARIMA and ARARMA processes, fitted all alternative methods, and computed descriptive statistics for data used in model-fitting. The statistics are comprehensive and borrow heavily from the rule-based forecasting procedure of Collopy and Armstrong (1992): the number of observations, percentage of outliers, R^2 and variances from regressions (X_t on time, X_t on lagged values, $(1-B)X_t$ on lagged values), binary variables for the direction and consistency of trend, binary variables for the differencing analysis discussed above, and autocorrelations. These statistics were used as explanatory variables in a regression-based performance index for each method.

Meade tested his procedure with additional simulated series as well as the 1,001 series from the M1 competition and Fildes' collection of 261 telecommunications series. In the simulated series, Meade's procedure consistently selected the best method from all candidates.

This was expected because the series were generated from one of the candidate methods. In the M1 series, the results were less encouraging, with selected methods ranking fifth in median performance and second in mean performance. In the Fildes series, the selected methods ranked fourth for both median and mean performance, although it is by now well established that Fildes' robust trend is the only method that gives reasonable results for these series. Meade went on to experiment with selecting combinations of forecasts, a problem beyond our scope. Meade's procedure, like that of Shah, may have merit in selection from the exponential smoothing class, but it is difficult to tell.

5.2 Expert systems

Expert systems for individual selection have been proposed by Collopy and Armstrong (C&A) (1992), Vokurka et al. (1996), Adya et al. (2001), Arinze (1994), and Flores and Pearce (2000). C&A's rule-based forecasting system includes 99 rules constructed from time series characteristics and domain knowledge. These rules combine the forecasts from four methods: a random walk, time series regression, Brown's double exponential smoothing, and the A-N method. This is an odd set of candidate methods because Brown's method is a special case of the A-N method, as discussed in Section 4.1. Because the C&A approach requires considerable human intervention in identifying features of time series, Vokurka et al. (1996) developed a completely automatic expert system that selects from a different set of candidate methods: the N-N and DA-N methods, classical decomposition, and a combination of all candidates. C&A and Vokurka et al. tested their systems using 126 annual time series from the M1 competition and concluded that they were more accurate than various alternatives. However, they did not compare their results to aggregate selection of the DA-N method. Gardner (1999) made this comparison and found that aggregate selection of the DA-N method was more accurate at all forecast horizons than either version of rule-based forecasting.

Another version of rule-based forecasting by Adya et al. (2001) reduced Collopy and Armstrong's rule base from 99 to 64 rules for data with no domain knowledge. They also deleted Brown's double exponential smoothing from the list of candidate methods. Adya et al. tested their system in the M3 competition and obtained better results. Rule-based forecasting was slightly more accurate than aggregate selection of DA-N in annual data and performed about the same as DA-N in seasonally-adjusted monthly and quarterly data. In future research, Adya et al. plan to reintroduce Brown's method and add the DA-N method to the list of candidate methods. Reintroduction of Brown's method will likely detract from performance, but it may be that the DA-N method can offset the loss.

Arinze (1994) developed a rule-induction type of expert system to select from the N-N, A-N, and A-M methods, adaptive filtering, moving averages, and time series decomposition. For unknown reasons, seasonal exponential smoothing methods and the damped-trend methods were not considered. Arinze tested his system using 85 aggregate economic series and found that it picked the best method about half the time. Another rule-induction expert system was developed by Flores and Pearce (2000) and tested with M3 competition data. Flores and Pearce were pessimistic about their results, which at best were mixed.

5.3 Information criteria

Numerous information criteria are available for selection of an exponential smoothing method. Information criteria have an advantage over the procedures discussed in Sections 5.1–5.2 in that they can distinguish between additive and multiplicative seasonality. The disadvantage of information criteria is that the computational burden can be significant. For example, Hyndman et al. (2002) recommend fitting all models (from their set of 24 alternatives) that might conceivably be appropriate for a time series, then selecting the one that minimizes the AIC. In the M1 and M3 data, this procedure gave accuracy results that compared favorably to commercial software and rule-based forecasting, although, like most of the selection procedures discussed above, they did not compare their results to aggregate selection of the DA-N method. Table 4 makes such comparisons. The first part of the table gives MAPEs for aggregate selection of DA-N (as reported in Gardner and McKenzie, 1985) vs. individual selection of state-space methods using the AIC (as reported in Hyndman et al., 2002). In the 1,001 series, for the average of all forecast horizons, the DA-N method was better than individual selection using the AIC. DA-N was also better at every individual horizon save horizon 2, in which there was a tie. At longer horizons, the advantage of the DA-N method was substantial. For example, at horizon 18, the difference was more than 7 percentage points. In the subset of 111 series, overall comparisons are about the same as in the 1,001 series although the AIC-selected method was better at horizons 2 and 15.

The second part of Table 4 gives symmetric APEs for the M3 competition data as reported in Makridakis and Hibon (2000), Hyndman et al. (2002), and Taylor (2003a). For the annual and quarterly series, we compare DA-N to the state-space models. In the annual series, overall and at every horizon, DA-N was more accurate. In the quarterly series, the state-space models have a small advantage at horizons 1 and 2, but overall DA-N was more accurate. In the monthly series, we have results for DA-N, DM-N, and the state-space models. Again the state-space models have a small advantage in the short term, but overall there is little to choose among the three alternatives.

Table 4
APE Comparisons

M1 Competition: MAPE

	1,001 series		111 series	
Horizon	Damped add. trend	State-space framework	Damped add. trend	State-space framework
1	**8.3**	9.0	**7.6**	8.7
2	**10.8**	**10.8**	9.7	**9.2**
3	**12.1**	12.8	**11.5**	11.9
4	**13.0**	13.4	**13.1**	13.3
5	**15.7**	17.4	**14.5**	16.0
6	**17.9**	19.3	**15.9**	16.9
8	**17.7**	19.5	**16.6**	19.2
12	**16.7**	17.2	**13.6**	15.2
15	**21.0**	23.4	29.0	**28.0**
18	**21.7**	29.0	**29.5**	31.0
Overall	**16.2**	17.6	**16.1**	17.3

M3 Competition: Symmetric APE

	645 annual series		756 quarterly series		1,428 monthly series		
Horizon	Damped add. trend	State-space framework	Damped add. trend	State-space framework	Damped add. trend	Damped mul. trend	State-space framework
1	**8.0**	9.3	5.1	**5.0**	11.9	11.8	**11.5**
2	**12.4**	13.6	6.8	**6.6**	11.4	11.0	**10.6**
3	**17.0**	18.3	**7.7**	7.9	13.0	12.7	**12.3**
4	**19.3**	20.8	**9.1**	9.7	14.2	**13.3**	13.4
5	**22.3**	23.4	**9.7**	10.9	12.9	12.4	**12.3**
6	**24.0**	25.8	**11.3**	12.1	12.6	12.9	**12.3**
8			**12.8**	14.2	**13.0**	13.3	13.2
12					13.9	**13.6**	14.1
15					17.5	**17.4**	17.6
18					18.9	**18.2**	18.9
Overall	**17.2**	18.5	**9.3**	9.9	14.6	**14.4**	14.5

Later work by Billah et al. (2005) compared eight information criteria used to select from four exponential smoothing methods. The criteria included the AIC, BIC, and other standards, as well as two new Empirical Information Criteria (EIC) that penalize the likelihood of the data by a function of the number of parameters in the model. One of the EIC penalty functions is linear, while the other is nonlinear, and neither depends on the length of the time series (they are intended for use in groups of series with similar lengths). Billah et al.'s candidate exponential smoothing methods included N-N, N-N with drift (see Section 3.4), A-N, and the state-space version of DA-N. Billah et al. tested the criteria with simulated time series and seasonally-adjusted M3 data. Although the EIC criteria performed better than the others, this study is not benchmarked, and we do not know whether the EIC criteria picked methods better than aggregate selection of the DA-N method. Another problem is that, although the authors used the M3 data, they reported MAPE rather than symmetric APE results. This is frustrating because we cannot make comparisons to other M3 results or to the AIC results in Hyndman et al. (2002).

One other idea for selecting from alternative state-space models for the A-M method, called the correlation method, was suggested by Koehler et al. (2001). The correlation method chooses the model that gives the highest correlation between the absolute value of the residuals and (1) estimates for the level and trend components, (2) the same components multiplied by the seasonal index, and (3) the seasonal indices themselves. In simulated time series, this idea worked very well compared to a maximum likelihood method, but we have no results for real data. It should be possible to use the correlation method to choose from a wider range of models, although nothing has been reported.

5.4 Projected operational or economic benefits

In production and inventory control, forecasting is a major determinant of inventory costs, service levels, scheduling and staffing efficiency, and many other measures of operational performance (Adshead and Price, 1987; Fildes and Beard, 1992; Lee et al. 1993). In the broader context of supply chains, forecasting determines the value of information sharing, a function that reduces costs and improves delivery performance (Zhao et al., 2002). Forecast errors also contribute to the bullwhip effect, the tendency of orders to increase in variability as one moves up a supply chain (Chandra and Grabis, 2005; Dejonckheere et al., 2003, 2004; Zhang, 2004). It follows that forecasting methods in operating systems should be selected on the basis of benefits, although this was done in only a few studies. The lack of research is understandable because of the expense – usually, one must build a model of the operating system in order to project benefits.

The only study of method selection for a manufacturing process is by Adshead and Price (1987), who developed a cost function to select a forecasting method for a producer of industrial fasteners with annual sales of £4 million. Total costs affected by forecasting included inventory carrying costs, stock-out costs, and overtime. Using real data, the authors developed a detailed simulation model of the plant, including six manufacturing operations carried out on 33 machines. They computed costs for a range of parameters in the N-N method, the double smoothing version of the A-N method, and Brown’s (1963) quadratic exponential smoothing, a method that performed very poorly in empirical studies and thus disappeared from the literature. Stock-out costs proved difficult to measure, and the authors were forced to test several assumptions in the cost function. Regardless of the assumption, the N-N method was the clear winner and was implemented by the company.

In a US Navy distribution system with more than 50,000 inventory items, Gardner (1990) compared the effects of a random walk and the N-N, A-N, and DA-N methods on the average delay time to fill backorders. Delay time was estimated in a simulation model using nine years of real daily demand and lead time history. The DA-N method proved superior for any level of inventory investment. For example, at a typical investment level of $420 million, the damped trend reduced delay time by 19% (6 days) compared to the N-N method, the standard Navy forecasting method prior to this study. Rather than reduce delay time, management opted to hold it constant and reduce inventory investment by 7% ($30 million). The cost of this study was $150,000, mostly in programming time.

For a distributor of electronics components, Flores et al. (1993) compared methods on the basis of costs due to forecast errors, defined as the sum of excess inventory costs (above targets) and the margin on lost sales. The authors used a sample of 967 demand series to compute costs for the N-N method with fixed and adaptive parameters, the double smoothing version of the A-N method, and the median value of historical demand. For items with margins greater than 10%, the N-N method with a fixed parameter was best, while the median was best for items with lower margins. The relative performance of the median was surprising, and Flores et al. remarked that a broader study could well change the conclusions. Essentially the same cost function as that of Flores et al. was used to evaluate methods for steel and aluminum sales by Mahmoud and Pegels (1990), although this paper is impossible to evaluate because several smoothing methods were not defined and references were not given.

The only other study of method selection using operational or economic benefits is by Eaves and Kingsman (2004), who selected a modified version of the Croston method for intermittent demand on the basis of inventory investment. This paper is discussed further in Section 7.2.

5.5 Identification vs. selection

Although state-space models for exponential smoothing dominate the recent literature, very little has been done on the identification of such models as opposed to selection using information criteria. The only possibly relevant papers here are by Koehler and Murphree (1988) and Andrews (1994). Koehler and Murphree identified and fitted MSOE state-space models to 60 time series (all those with a minimum length of 40 observations) from the 111 series in the M1 competition. Their identification and fitting routine is best described as semi-automatic, with some human intervention required. Koehler and Murphree did not attempt to match their model selections to equivalent exponential smoothing methods. They compared forecast accuracy (mean and median APEs) to simple exponential smoothing and ARIMA models identified by an expert. In general, the identification process was disappointing. Although there were some differences in subsets of the data, simple exponential smoothing ranked first in overall accuracy by a significant margin, with state-space second, and ARIMA third. For the complete set of 111 series, Andrews identified and fitted MSOE models (all with exponential smoothing equivalents), again using a semi-automatic procedure. His results appear to be better than the Box-Jenkins results, although he did not give enough details to be sure, and he did not make comparisons to the exponential smoothing results reported for the M1 competition.

Rather than attempt to identify a model, we could attempt to identify the best exponential smoothing method directly. Chatfield and Yar (1988) call this a “thoughtful” use of exponential smoothing methods that are usually regarded as automatic. For the Holt-Winters class, Chatfield and Yar give a common-sense strategy for identifying the most appropriate method. This strategy is expanded in Chatfield (1988, 1995, 1997, 2002), and here we give the strategy in a nutshell. First, we plot the series and look for trend, seasonal variation, outliers, and changes in structure that may be slow or sudden and may indicate that exponential smoothing is not appropriate in the first place. We should examine any outliers, consider making adjustments, and then decide on the form of the trend and seasonal variation. At this point, we should also consider the possibility of transforming the data, either to stabilize the variance or to make the seasonal effect additive. Next, we fit an appropriate method, produce forecasts, and check the adequacy of the method by examining the one-step-ahead forecast errors, particularly their autocorrelation function. The findings may lead to a different method or a modification of the selected method. For a sample of reasonable size, it would be useful to have results for this strategy as a validation of the automatic method selection procedures discussed above. It does not appear that any of the automatic procedures have been validated in such a manner.

6. Model-fitting

In order to implement an exponential smoothing method, the user must choose parameters, either fixed or adaptive, as well as initial values and loss functions. The user must also decide whether to normalize the seasonals, a problem considered earlier in Section 3.3. The research in choosing fixed parameters, discussed in Section 6.1, is not particularly helpful, and there are several open research questions. To avoid model-fitting for the N-N method, we can use adaptive parameters, reviewed in Section 6.2. Parameter selection is not independent of initial values and loss functions, as discussed in Section 6.3.

6.1 Fixed parameters

There is no longer any excuse for using arbitrary parameters in exponential smoothing given the availability of good search algorithms, such as the Excel Solver. For examples of using the Solver in parameter searches, see Bowerman et al. (2005) and Rasmussen (2004). One cautionary note is that, in the trend and seasonal models, the response surface is not necessarily convex. Thus it may be advisable to start any search routine from several different points to evaluate local minima.

A theoretical paper by Farnum (1992) suggests that search routines might be improved by taking account of the autocorrelation structure of the time series. Farnum gives a search routine of this type for the N-N method, although it is obvious that a simple grid search will produce the same result for this method. It is not clear how autocorrelation might be used in searches for the more complex methods.

We hope that our search routine comes to rest at a set of invertible parameters, but this may not happen, as discussed below. Invertible parameters create a model in which each forecast can be written as a linear combination of all past observations, with the absolute value of the weight on each observation less than one, and with recent observations weighted more heavily than older ones. This definition is generally accepted, but the words stability and invertibility are often used interchangeably in the literature, which can be confusing. One definition of stability comes from control theory. If we view an exponential smoothing method as a system of linear difference equations, a stable system has an impulse response that decays to zero over time. The stability region for parameters in control theory is the same as the invertibility region in time series analysis (McClain and Thomas, 1973). But from the time series perspective, stability has another definition related to stationarity and is not relevant here. For a detailed comparison of the properties of stability, stationarity, and invertibility, see Pandit and Wu (1983). Examples of authors that use stability in the control theory sense are McClain and Thomas (1973), McClain (1974), Sweet (1985), Gardner and McKenzie (1985, 1988, 1989), Chatfield and Yar (1991), and Lawton (1998).

In the linear non-seasonal methods, the parameters are always invertible if they are chosen from the usual [0, 1] interval. The same conclusion holds for quarterly seasonal methods, but not for monthly seasonal methods (Sweet, 1985), whose invertibility regions are complex. For the monthly A-A and A-M methods, Sweet (1985) and Archibald (1990) give examples of some apparently reasonable combinations of [0, 1] parameters that are not invertible. Both authors test A-M parameters using the A-A invertibility region. Non-invertibility usually occurs when one or more parameters fall near boundaries, or when trend and/or seasonal parameters are greater than the level parameter.

For all seasonal exponential smoothing methods, we can test parameters for invertibility using an algorithm by Gardner and McKenzie (1989), assuming that additive and multiplicative invertible regions are identical. However, this test may fail to eliminate some troublesome parameters. An extraordinary finding in Archibald's study is that some combinations of [0, 1] parameters near boundaries fall within the ARIMA invertible region, but the weights on past data diverge. The result is that some older data are weighted more heavily than recent data. Archibald found that diverging weights occur in both standard and state-space versions of the A-M method. Through trial and error, Archibald found a more restrictive parameter region for state-space A-M that seemed to prevent diverging weights. The lesson from Archibald's study is that one should be skeptical of parameters near boundaries in all seasonal models.

Archibald's work was extended by Hyndman et al. (2005a), who give equations that define an "admissible" parameter space for all additive seasonal methods except the DM methods. Combinations of parameters that fall within the admissible space produce truly invertible models. Although the admissible space is complex for all methods considered, it is a simple matter to program the equations as a final check on fitted parameters. Hyndman et al.

also make a case similar to that of Archibald and Koehler (2003) (see Section 3.3) for renormalization of seasonals in state-space models.

In the N-N method, Johnston and Boylan (1994) argue that α should not exceed 0.50. This argument is based on the use of the N-N method to estimate the level in the corresponding MSOE model – see Section 4.3, equations (8) and (9). Johnston and Boylan decompose the error into three parts, namely a term for residual random noise, a term that accounts for sampling error in estimating α, and a term that reflects the approximate nature of the model. At short forecast horizons, the residual random noise term dominates. As the horizon increases, the importance of this term is superseded by sampling error. Eventually, the approximate nature of the model becomes dominant and Johnston and Boylan argue that it is unwise to extrapolate beyond this point. When $\alpha = 0.50$, the boundary for safe extrapolation works out to be only one time period. The authors make no attempt to reconcile this argument with the many examples of time series in which the N-N method with $\alpha > 0.50$ is optimal.

Once the parameters have been selected, another problem is deciding how frequently they should be updated. When forecasting from multiple time origins, Fildes et al. (1998) compared three options for choosing parameters in the N-N, A-N and DA-N methods: (1) arbitrarily, (2) optimize once at the first time origin, and (3) optimize each time forecasts are made. These options were tested in the Fildes collection of 261 telecommunications series, and the best option was to optimize each time forecasts were made. It remains to be seen whether this conclusion applies to series that are not so well behaved.

6.2 Adaptive smoothing

The term adaptive smoothing is used to mean many different things in the literature. Here we mean only that the parameters are allowed to change automatically in a controlled manner as the characteristics of the time series change. In Gardner (1985), I concluded that there was no credible evidence in favor of any of the numerous forms of adaptive smoothing. See also Armstrong (1984) for a similar conclusion. Since then, a number of new ideas for adaptive smoothing have appeared.

The Kalman filter can be used to compute the parameter in the N-N method. Snyder (1988) developed such an algorithm, assuming a random walk with a single source of error. The algorithm is similar to Gilchrist's (1976) exact DLS version of the N-N method in that no initial values or model-fitting are necessary, although the user must pre-specify a "long-run" smoothing parameter. Snyder's method is adaptive in the sense that the smoothing parameter is time-varying and will eventually converge to the long-run parameter. Using the 111 series from the M1 competition, Snyder's MAPE results were about the same as the fixed-parameter N-N method in monthly data, but slightly better in annual and quarterly data. In Snyder (1993), his filter was implemented in a system for forecasting auto parts sales, although problems within the company made it difficult to assess forecasting performance.

A more elaborate Kalman filtering idea, by Kirkendall (1992), uses adaptive parameters in four MSOE state-space models designated as steady, outlier, level shift, and a mixed model with mean and variance based on a weighted average of the first three models. The steady model is the N-N method and the others are variations. Separate model estimates and separate posterior probabilities are maintained for each of the models, and the state transitions from one model to another according to the probabilities. Kirkendall gives empirical results for two time series, but there is no benchmark to judge the performance of the proposed system. Similar proposals for adapting to changes in structural models corresponding to exponential smoothing are available in Jun and Oliver (1985) and Jun (1989), but again the empirical results are limited and not benchmarked.

An unpromising scheme for adapting the N-N method was suggested by Pantazopoulos and Pappis (1996), who set the parameter equal to the absolute value of the two-step-ahead forecast error divided by the one-step-ahead error. The practical consequence is that the smoothing parameter frequently exceeds 1.0. When this happens, the authors reset the parameter to 1.0, thus producing a random-walk forecast.

The only adaptive method that has demonstrated significant improvement in forecast accuracy compared to the fixed-parameter N-N method is Taylor's (2004a, 2004b) smooth transition exponential smoothing (STES). Smooth transition models are differentiated by at least one parameter that is a continuous function of a transition variable, V_t. The formula for the adaptive parameter α_t is actually a logistic function:

$$\alpha_t = 1/(1 + \exp(a + bV_t)) \tag{15}$$

There are several possibilities for V_t, including e_t, $|e_t|$, and e_t^2. Whatever the transition variable, the logistic function restricts α_t to [0, 1]. The drawback to STES is that model-fitting is required to estimate a and b; thereafter, the method adapts to the data through V_t.

In Taylor (2004a), STES was arguably the best method overall in volatility forecasting of stock index data compared to the fixed-parameter version of N-N and a range of GARCH and autoregressive models. The application of exponential smoothing to volatility forecasting is very different to the usual exponential smoothing applications. With financial returns, the mean is often assumed to be zero or a small constant value, and attention turns to predicting the variance.

In Taylor's study, he recommended using $V_t = e_t$ or $|e_t|$ in order to replicate the variance dynamics of smooth transition GARCH models. In Taylor (2004b), $V_t = e_t^2$ was the best choice for simulated time series with level shifts and outliers as well as the 1,428 M3 monthly series. As benchmarks for STES, Taylor computed results for numerous other exponential smoothing methods, with both fixed and adaptive parameters. STES performed well in the simulated series, as expected. In the many re-examinations of the M3 series, Taylor is the only researcher who has followed the advice of Fildes (1992) and Fildes et al. (1998) and evaluated forecast performance across time. Using the last 18 observations of each series, Taylor computed successive one-step-ahead monthly forecasts, for a total of 25,704 forecasts. Judged by MAPE and median APE, STES was the most accurate method tested, significantly so for the MAPE. The results were not as good for the median symmetric APE and root-mean-squared APE, but STES was still among the best methods.

Only a few authors have proposed adapting the parameters in the trend methods. In the A-A method, Williams (1987) contends that only the level parameter should be adapted. Mentzer (1988) and Mentzer and Gomes (1994) agree with Williams and recommend setting the level parameter in the A-A method equal to the absolute percentage error in the current period (if the error exceeds 100%, the level parameter is set equal to 1.0). Mentzer and Gomes present results for the M1 data that are the best of all methods reported to date. But in the M3 data, Taylor (2004b) found that the Mentzer and Gomes version of the A-A method was certainly the worst exponential smoothing method tested, regardless of the error measure or whether the parameters were fixed or adaptive. There seems to be no explanation for this contradiction in performance.

6.3 Initial values and loss functions

Standard exponential smoothing methods are usually fitted in two steps, by choosing fixed initial values (see Gardner, 1985, for a review of the alternatives), followed by an independent search for parameters. In contrast, the new state-space methods are usually fitted using maximum likelihood, a procedure that makes the choice of initial values less of a concern because they are refined simultaneously with the smoothing parameters during the optimization process. Unfortunately, maximum likelihood may require significant computation times, as discussed in Hyndman et al. (2002). For example, in monthly seasonal models with a damped trend, there are 13 initial values and 4 parameters, so the optimization is done in 17-dimensional space. Hyndman et al. used heuristics to reduce the number of dimensions and speed up computation, although whether the heuristics made any difference in forecast accuracy is not known. Given the findings of Makridakis and Hibon (1991) discussed below, this seems unlikely.

Another maximum likelihood procedure differing in many details from Hyndman et al. is found in Broze and Melard (1990), who give meticulous instructions for fitting all of the linear exponential smoothing methods in Table 1. The Broze and Melard procedure is difficult to evaluate because they give no empirical results or computation times. An alternative to maximum likelihood is Segura and Vercher's (2001) nonlinear programming model that optimizes initial values and parameters simultaneously, but again the authors are silent about empirical results and computation times.

In an exhaustive re-examination of the M1 series, Makridakis and Hibon (1991) measured the effect of different initial values and loss functions in fitting the N-N, A-N, and DA-N methods, using seasonally-adjusted data where appropriate. Initial values were computed by least squares, backcasting, and several simple methods such as setting all initial values to zero or setting the initial level equal to the first observation, with initial trend equal to the difference between the first two observations. Loss functions included the MAD, MAPE, Median APE, MSE, the sum of the cubed errors, and a variety of non-symmetric functions computed by weighting the errors in different ways. There was little difference in average post-sample accuracy regardless of initial values or loss function. Furthermore, sample size or type of data (annual, quarterly, or monthly) did not make any consistent difference in the best choice of initial values or loss function. The authors repeated the study in the Fildes telecommunications data with much the same findings.

The major conclusion from the Makridakis and Hibon study is that the common practice of initializing by least squares, choosing parameters from the [0, 1] interval, and fitting models to minimize the MSE provides satisfactory results. The authors caution that this conclusion applies to automatic forecasting of large numbers of time series and may not hold for individual series, especially those containing significant outliers. To cope with outliers, I argue for a MAD loss function (Gardner, 1999). However, I point out that there are exceptions, making it advisable to evaluate both MSE and MAD loss functions in many series.

7. Forecasting for inventory control

In inventory control with continuous (non-intermittent) demand, exponential smoothing methods are the same as in other applications, but variance estimates are considerably different. Variances of cumulative demand over the complete reorder lead time are required, as discussed in Section 7.1. If demand is intermittent, we need both specialized smoothing methods and variance estimates, as discussed in Section 7.2. Our discussion is concerned only with these topics, and the vast literature on inventory decision rules constructed from forecasting systems is beyond our scope.

7.1 Continuous demand

For the N-N method, what might be called the traditional estimate of the standard deviation of total lead time demand is $s\sqrt{m}$, where s is the one-step-ahead standard deviation and m is the lead time (Brown, 1959, 1967). This estimate has been persistent in the literature, including a recent paper by Willemain et al. (2004) discussed below, but it is biased. The correct multiplier for the standard deviation was derived using an MSOE state-space model by Johnston and Harrison (1986) and an SSOE model by Snyder et al. (1999):

$$f(\alpha, m) = \sqrt{(m + \alpha(m-1)m(1 + \alpha(2m-1)/6))} \qquad (16)$$

The effect of this multiplier is significant for any value of α at any lead time greater than one period. For example, with $\alpha = 0.2$, and a lead time of two periods, the correct standard deviation is more than twice the size of the traditional estimate. For the same α at a lead time of six periods, the correct standard deviation is almost four times the traditional estimate.

For the linear methods A-N, DA-N, A-A, and DA-A, Snyder et al. (2004) used SSOE models to develop variance expressions for cumulative lead-time demand, assuming both additive and multiplicative errors. Snyder et al. (2004) can be viewed as a companion paper to Hyndman et al. (2005b), which contains prediction intervals around point forecasts for the same methods and several others. For those who prefer MSOE models, some limited and far more complex variance results are available in Harvey and Snyder (1990). It is important to understand how the error assumptions in the SSOE models affect the distribution of cumulative lead-time demand. If the errors are additive and normal, cumulative lead-time demand will of course be normal. If the errors are normal and multiplicative, cumulative lead-time demand will not be normal, although Hyndman et al. (2005b) suggest that the normal distribution is a safe approximation. For the smoothing methods without analytical variance expressions, there are many bootstrapping procedures in the literature that can be used to develop empirical variance estimates. The parametric bootstrapping procedure of Snyder (2002) and Snyder et al. (2002) should appeal to the practical forecaster because it is tailored to lead-time demand, and can be used when the distribution of demand is non-normal, when the lead time is stochastic, and when demand is intermittent.

Snyder et al. (2002) used the parametric bootstrap to study an important practical question about state-space modeling: Do the assumptions of additive and multiplicative errors make any difference in estimating variances? The authors used multiplicative errors in generating data with no trend or seasonal pattern, and then fitted both additive- and multiplicative-error versions of the N-N method. This research design should have produced results substantially in favor of the multiplicative-error version, but it did not. Differences in simulated fill rates and order-up-to-levels between the two N-N versions were very small except when a major step change in the series occurred. When data with trends and seasonality were simulated, the multiplicative-error N-N method did better, but this finding is misleading because the method was not appropriate for the data.

Stockouts are not treated in the research discussed above, although the parametric bootstrap could easily be adapted to do so. Stockouts truncate the distribution of demand, causing systematic bias in estimates of the mean and variance. To correct for such bias in the N-N method, Bell (1978) replaced X_t with the conditional mean of the demand distribution for periods that include stockouts. The conditional mean is defined as the expected value of demand, given that observed demand is greater than or equal to the quantity actually available for sale. Demand is assumed normal, with variance estimated by the smoothed MAD. The normality assumption may seem doubtful, but Bell (1978, 2000) and Artto and Plykkanen (1999) argue that product stocking methods based on the normal distribution work well in practice. Through simulation, Bell (1981, 2000) found that his procedure works well so long as the number of stockouts does not exceed 50%. For larger numbers of stockouts, Bell (2000) gives adjustments to his procedure.

7.2 Intermittent demand

If time series of inventory demands are observed intermittently, we cannot recommend the N-N method because the forecasts are biased low just before a demand occurs and biased high just afterward, resulting in excessive stock levels. The standard method of forecasting intermittent series was developed by Croston (1972), and works as follows. Using the N-N method, we smooth two components of the time series separately, the observed value of nonzero demand (D_t) and the inter-arrival time of transactions (Q_t). The smoothed estimates are denoted Z_t and P_t, respectively, and their recurrence equations are:

$$Z_t = \alpha D_t + (1-\alpha) Z_{t-1} \tag{17}$$

$$P_t = \alpha Q_t + (1-\alpha) P_{t-1} \tag{18}$$

The value of α is the same in both equations. The expected value of demand per unit time (Y_t) is then:

$$E(Y_t) = Z_t / P_t \tag{19}$$

If there is no demand in a period, Z_t and P_t are unchanged. When demand occurs every period, the Croston method gives the same forecasts as the conventional N-N method.

Syntetos and Boylan (2001) showed that $E(Y_t)$ is biased high and derived a corrected version of equation (19), although this version is not mentioned in later research by Syntetos and Boylan (2005) and Syntetos et al. (2005). Instead, the later papers give a different corrected version of equation (19):

$$E(Y_t) = (1 - \alpha / 2)(Z_t / P_t) \qquad (20)$$

The modified Croston forecasting system defined by equations (17), (18), and (20) is also found in Eaves and Kingman (2004), who tested the system using a sample of 11,203 repair parts from Royal Air Force inventories. The results varied somewhat depending on the degree of aggregation of the data (weekly, monthly, quarterly) and the type of demand pattern (ranging from smooth to highly intermittent). However, in general the modified Croston method was more accurate than the original, and both methods performed significantly better than the N-N method. To compute safety stocks, Eaves and Kingman relied on a variance expression developed by Sani and Kingsman (1997) (discussed below). The authors extrapolated the sample savings to the entire inventory, with convincing results. The conventional N-N method produced an additional 13.6% in inventory investment (£285 million) over the modified Croston method.

Another idea to correct for bias in the Croston method is given in Leven and Segerstedt (2004). Rather than smooth size and inter-arrival time separately as in equations (17) and (18), the authors proposed a method that can be shown to be equivalent to smoothing both components in the same equation. That is, in the N-N method the authors suggest replacing X_t with D_t / Q_t. The authors give no explanation of how this idea corrects for bias.

Snyder (2002) took a state-space approach to the study of Croston's method. The underlying model assumes that nonzero demands are generated by an ARIMA (0, 1, 1) process, while the inter-arrival times follow the Geometric distribution. The latter assumption means that the probability of demand occurrence is constant (one of Croston's stated assumptions), or equivalently that the mean inter-arrival time of the demand series is constant. Therefore, equation (18) is not used and P in equation (19) has no time subscript. This model can generate negative values, so an alternative model using the logarithms of nonzero demands was specified. Variance estimates for the models were developed using a parametric bootstrap from the normal distribution. Snyder gives encouraging results for his models for a few time series, but more evidence is needed to support a constant mean inter-arrival time. This idea is contrary to the philosophy of exponential smoothing, a problem acknowledged by Snyder.

Further analysis of Snyder's models is given in Shenstone and Hyndman (2005), who developed analytical prediction intervals for them. Shenstone and Hyndman also found that there is no underlying stochastic model for Croston's method or the two variants proposed by Syntetos and Boylan (2001, 2005). Any models that might be considered as candidates simply do not match the properties of intermittent data. Thus, if we wish to have analytical prediction intervals for intermittent data, the only option is to adopt one of Snyder's models.

Shenstone and Hyndman's work creates doubts about the assumptions behind the variance expressions for Croston's method found in the literature, including Croston (1972) as corrected by Rao (1973), Schultz (1987), Johnston and Boylan (1996a), and Sani and Kingsman (1997). All of these variance expressions must be regarded as approximations, although they have generally worked well in empirical studies. The best approximation for the variance of mean demand may be that of Sani and Kingsman:

$$Var(Y_t) = \max[Var(Z_t)/P_t,\ \ 1.1Z_t/P_t] \tag{21}$$

The second term on the right-hand side looks peculiar, but the purpose is to make certain that the variance is larger than the mean, a relationship required by the assumption that demands are generated by the negative binomial distribution. The variance of Z_t is estimated by the smoothed MAD, assuming normality and using the same α as in equations (17) and (18). In an empirical study of forecasting the demand for repair parts, Sani and Kingsman showed that equation (21) gave much better service level performance than Croston's original variance expression.

When should we use Croston's method or one of its variants? Johnston and Boylan (1996a, 1996b) found that Croston's method is superior to the N-N method when the average inter-arrival time is greater than 1.25 times the interval between updates of the N-N method. This finding was thoroughly substantiated by simulating different inter-demand intervals and patterns, different distributions of order size, different forecast horizons, and different parameters in the smoothing methods.

Syntetos et al. (2005) extended Johnston and Boylan's work by developing rules for choosing from three methods: N-N, original Croston, and modified Croston in equations (17), (18), and (20). The rules are based on approximate theoretical MSE values for each method, assuming that the data have a constant mean, and that the errors in forecasting the data are autocorrelated. For inter-arrival times greater than 1.32 and a squared coefficient of variation of demand sizes greater than 0.49, the Croston method with bias correction should be superior to the other methods. For smaller inter-arrival times and coefficients of variation, the original Croston method is theoretically expected to perform best. That is, the original Croston method should always beat the N-N method, a very surprising conclusion that the authors attribute to either the relatively high variability of the N-N errors or the quality of their approximation for the original Croston MSE. Based on a sample of 3,000 real time series, Syntetos et al. argued that their rules worked well.

Unfortunately, problems arise in any attempt to generalize from Syntetos et al. The results contradict Johnston and Boylan (1996a, 1996b), who found that the variability of order size had almost no effect on the relative performance of methods. Syntetos et al. (2005) also appears to contradict Syntetos and Boylan (2005). These two papers do not acknowledge each other even though they use the same data to rank the same methods. The difference is that Syntetos et al. based their rankings on the MSE, while Syntetos and Boylan employed other error measures. In Syntetos et al., the original Croston method was consistently better than the N-N method, but this was not true in Syntetos and Boylan.

8. Empirical studies

Table 4 is a guide to all papers published since 1985 that present empirical results for exponential smoothing, excluding the M-competitions, the many re-examinations of the M-competitions, papers based entirely on simulated time series, and several papers that are impossible to evaluate. This last category includes Shoesmith and Pinder (2001), who found that vector autoregression was more accurate than exponential smoothing in forecasting several inventory demand series. Unfortunately, the authors did not disclose the particular smoothing methods used. Mahmoud and Pegels (1990) and Snyder (1993) were also omitted for reasons explained in Sections 5.4 and 6.2, respectively.

Several generalizations can be made about the 66 papers listed in Table 4. Seasonal methods were rarely used, even though most studies were based on seasonal data. It may be

Table 4. Empirical studies

Data	Methods	Reference
Airline passengers	DA-A	Grubb and Mason (2001)
Ambulance demand calls	A-M	Baker and Fitzpatrick (1986)
Australian football margins of victory	N-N	Clarke (1993)
Auto parts	N-N	Gardner and Diaz (2000)
Auto parts	N-N, Croston	Snyder (2002)
Auto parts	N-N, Croston	Syntetos and Boylan (2005)
Auto parts	N-N, Croston	Syntetos et al. (2005)
Call volumes to telemarketing centers	A-A, A-M	Bianchi et al. (1998)
Chemical products	N-N, Croston	Garcia-Flores et al. (2003)
Computer network services	N-N	Masuda and Whang (1999)
Computer parts	DA-N	Gardner (1993)
Confectionery equipment repair parts	N-N, Croston	Strijbosch et al. (2000)
Consumer product sales (annual)	N-N, A-N, DA-N	Schnaars (1986)
Consumer food products	N-N	Koehler (1985)
Cookware sales	DA-N	Gardner and Anderson (1997)
Cookware sales	DA-N	Gardner et al. (2001)
Crime rates	N-N, A-N	Gorr et al. (2003)
Currency exchange rates	N-N, A-N, A-M	Dhieeriya and Raj (2000)
Department store sales	N-N, A-N	Guerts and Kelly (1986)
Economic data (various)	N-N, A-N	Geringer and Ord (1991)
Economic, environmental data (various)	A-N	Wright (1986b)
Electric utility loads	A-N	Huss (1985a)
Electric utility sales	A-N	Huss (1985b)
Electricity demand	N-N, A-N	Price and Sharp (1986)
Electricity demand	A-M	Taylor (2003b)
Electricity demand forecast errors	N-N	Ramanathan (1997)
Electricity supply	A-N	Sharp and Price (1990)
Electrical service requests	A-M	Weintraub et al. (1999)
Electronics components	N-N, A-N	Flores et al. (1993)
Exports	N-N	Mahmoud et al. (1990)
Financial futures prices	N-N	Sharda and Musser (1986)
Financial returns	N-N	Taylor (2004a)
Food product demand	N-N	Fairfield and Kingsman (1993)
Food product demand	N-N	Mercer and Tao (1996)
Hospital patient movements	A-M	Lin (1989)
Hotel revenue data	N-N, A-N	Weatherford and Kimes (2003)
IBM product sales	A-M	Wu et al. (1991)
Industrial data (various)	N-N, Croston	Willemain et al. (1994, 2004)
Industrial fasteners	N-N, A-N	Adshead and Price (1987)
Industrial production differences	N-N	Öller (1986)
Industrial production index	A-A	Bodo and Signorini (1987)
Leading indicators	A-N	Holmes (1986)
Macroeconomic variables	A-M	Thury (1985)
Mail order sales	N-N	Chambers and Eglese (1988)

Table 4. Empirical studies (continued)

Data	Methods	Reference
Mail volumes	A-M	Thomas (1993)
Manpower retention rates	A-N	Chu and Lin (1994)
Medicaid expenses	A-N	Williams and Miller (1999)
Medical supplies	A-N	Matthews & Diamantopoulos 1994)
Natural gas demand	N-N, A-N, A-M	Lee et al. (1993)
Stock index direction	N-N	Leung et al. (2000)
Supermarket product sales	Many	Taylor (2004c)
Point-of-sale scanner data	N-N	Curry et al. (1995)
Printed banking forms	N-N, A-N	Chan et al. (1999)
Process industry sales	DA-M	Miller and Liberatore (1993)
Royal Air Force spare parts	N-N, Croston	Eaves and Kingsman (2004)
Telephone service times	N-N	Samuelson (1999)
Telecommunications demand	N-N, A-N, DA-N	Fildes et al. (1998)
Tourism	A-A	Pfefferman and Allon (1989)
Tourism	A-N	Martin and Witt (1989)
Travel speeds in road networks	N-N	Hill and Benton (1992)
Truck sales	A-M	Heuts and Bronckers (1988)
US Navy inventory demands	N-N, A-N, DA-N	Gardner (1990)
Utility demand (water and gas)	A-M, N-A, N-M	Fildes et al. (1997)
Vehicle/agricultural machinery parts	N-N, Croston	Sani and Kingsman (1997)
Water quality, Divorce rates	A-N	Wright (1986b)

surprising that there have been no reported applications of the N-A or N-M methods, and only three applications of the A-A method, the subject of a large body of theoretical research. It may also be surprising that there have been few applications of the damped-trend methods. In most cases, little attention was given to method selection, a generalization that is substantiated by the large number of studies with only one method listed.

How often was exponential smoothing successful in these studies? Forecast performance was sometimes difficult to evaluate because many of the studies were not designed to be comparative in nature. However, my interpretation is that there are only seven studies in which exponential smoothing did not produce reasonable forecast accuracy, and all of these can be explained. In Holmes' (1986) analysis of leading indicator series characterized by dramatic turning points, it is unsurprising that transfer function models performed better than the A-N method. In forecasting IBM product sales (Wu et al., 1991), several Box-Jenkins models defeated the A-M method. The data suggest that the damped trend methods would have performed better, but the authors did not consider them, perhaps because they were relatively new at that time.

In forecasting point-of-sale scanner data (Curry et al., 1995), the univariate N-N method was applied to a multivariate problem with predictably poor results. Fildes et al. (1997) developed models for short-term forecasting of water and gas demand and found that complex multivariate methods (beyond the capability of the exponential smoothing methodology) were necessary to capture all the influences on the data. In Fildes et al. (1998), the telecommunications data contained little noise and no structure except very consistent negative trends, making the robust trend method the best choice. In Bianchi et al.'s (1998) study of incoming calls to telemarketing centers, the A-A and A-M methods did not perform as well as ARIMA modeling with interventions that were essential in the data.

The last study in which exponential smoothing did not perform well is by Willemain et al. (2004), who claimed that their patented bootstrap method made significant improvements in forecast accuracy over the N-N and Croston methods. However, as discussed in Gardner and Koehler (2005), Willemain et al. was published with mistakes and omissions that bias the results in favor of the patented method. First, the authors estimated the variance of lead-time demand for both simple smoothing and the Croston method as the length of the lead time multiplied by the one-step-ahead error variance. As shown in Section 7.1, this is a serious under-estimate of variance. Second, the authors did not consider any of the published modifications to the Croston method. Finally, although Willemain et al. mentioned other bootstrap methods in their paper, they did not test any of them, nor did they consider likely alternatives.

How often was Croston's method or one of its variants successful in empirical research? There are nine relevant studies in Table 4. For reasons explained above, the studies by Syntetos and Boylan (2005) and Syntetos et al. (2005) are difficult to interpret, and the results in Willemain et al. (2004) are biased. In the six studies that remain, Croston's method or one of its variants appeared to give reasonable performance, although it is somewhat difficult to generalize because the degree of success depended on the type of data and the way in which forecast errors were measured.

9. The state of the art

Exponential smoothing methods can be justified in part through equivalent kernel regression and ARIMA models, and in their entirety through the new class of SSOE state-space models, which have many theoretical advantages, most notably the ability to make the errors dependent on the other components of the time series. This kind of multiplicative error structure is not possible with the ARIMA class, which makes exponential smoothing a much broader class of models, neatly reversing the "special case" argument discussed in Gardner (1985).

The problem now is to determine whether the SSOE modeling framework has practical as well as theoretical advantages. This has yet to be demonstrated. In the M1 data, aggregate selection of the damped additive trend was a better choice than individual selection of SSOE models through information criteria. The same conclusion holds for annual and quarterly M3 data. For monthly M3 data, individual selection of SSOE models was superior only at short horizons. At longer horizons and overall, there was little to choose between damped additive and multiplicative trends and the SSOE models. I cannot explain these results, and they cannot be ignored if there is to be any hope of practical implementation.

A number of possibilities might be explored to improve the performance of the SSOE models. First, both additive and multiplicative trends could be damped. The effects of immediate trend-damping, rather than at two steps ahead, should also be investigated; this would have some effect on forecast accuracy, and would likely change the results of model selection according to various criteria. Not all of the 24 models in the SSOE framework can be expected to be robust, so the range of candidates might be reduced. The use of information criteria for model selection should be re-examined. In Hyndman (2001), the AIC often failed to select an ARIMA (0, 1, 1) model even when the data were generated by an ARIMA (0, 1, 1) process. It seems unreasonable to believe that the AIC should do any better in selection from the SSOE framework, and other selection procedures should be considered. We note that other information criteria were used in Billah et al. (2005), but the results are not benchmarked.

In general, researchers have avoided the problem of method selection in exponential smoothing, and there is as yet no evidence that individual selection can improve forecast accuracy over aggregate selection of one of the damped trend methods. However, Shah's (1997) discriminant analysis procedure and Meade's (2000) regression-based performance index are promising alternatives for individual method selection and deserve empirical research using an exponential smoothing framework.

From the practitioner's viewpoint, the aim in method selection must be robustness, especially in large forecasting systems. Several new methods have demonstrated robustness: Taylor's (2003a) damped multiplicative trends, Taylor's (2004a, 2004b) adaptive version of simple smoothing, and the Theta method of Assimakopoulos and Nikolopoulos (2000), shown to be equivalent to the N-N method with drift by Hyndman and Billah (2003). All of these methods deserve more research to determine when they should be preferred over competing methods.

There are a number of other opportunities for empirical research in exponential smoothing. The SSOE models yield analytical variance expressions for point forecasts that have eluded researchers for many years. Surely these expressions are better than the variance expressions used in the past, but they have not been evaluated with real data. Perhaps this could be done in something like an M-competition for prediction intervals. Such a competition should also test the SSOE variance expressions for cumulative forecasts at different lead times. In forecasting intermittent demand, we have several new versions of Croston's method that require further testing with real data. Another idea that merits empirical research is Fildes et al.'s (1998) recommendation that parameters be re-optimized each time forecasts are made.

Since Winters (1960) appeared, there has been confusion in the literature about whether and how seasonals should be renormalized in the Holt-Winters methods. Today, it seems foolish not to renormalize using the efficient Archibald-Koehler (2003) system, a major practical advance that resolves conflicting results and puts the renormalization equations in a common form. In the additive seasonal methods, it is not necessary to renormalize the seasonal indices if forecast accuracy is the only concern, but this is rarely the case in practice when repetitive forecasts are made over time. Forecasting methods require regular maintenance, a job that is easier to accomplish when the method components can be interpreted without bias. With multiplicative seasonality, we do not know if renormalization can safely be ignored, so certainly we should use the Archibald-Koehler system.

In fitting additive seasonal models, it is alarming that some combinations of [0,1] parameters fall within the ARIMA invertible region, yet the weights on past data diverge. This problem can be avoided by checking the weights using Hyndman et al.'s (2005b) boundary equations. In fitting multiplicative seasonal models, there is little guidance on parameter choice. Research is also needed on parameter choice for the new damped multiplicative trend methods.

My experience is that practitioners happily ignore most of the problems discussed in this paper. In the future, we must validate the substantial body of theory in exponential smoothing and communicate it to practitioners. Writing of exponential smoothing vs. the Box-Jenkins methodology, I concluded Gardner (1985) with the following opinion: "The challenge for future research is to establish some basis for choosing among these and other approaches to time series forecasting." This conclusion still holds, although we have many more alternatives today.

Acknowledgments

I am grateful to Chris Chatfield, Robert Fildes, Anne Koehler, and James Taylor for many helpful comments and suggestions. None of these people necessarily agrees with the opinions expressed in the paper, and any errors that remain are my own. I also wish to acknowledge my assistant, Thomas Bayless, of the Honors College, University of Houston, who prepared the bibliography and was a great help with the manuscript.

References

Adshead, N.S., & Price, D.H.R. (1987). Demand forecasting and cost performance in a model of a real manufacturing unit, *International Journal of Production Research*, 25, 1251-1265.

Adya, M., Collopy, F., Armstrong, J.S., & Kennedy, M. (2001). Automatic identification of time series features for rule-based forecasting, *International Journal of Forecasting*, 17, 143-157.

Akaike, H. (1970). Statistical predictor identification, *Annals of the Institute of Statistical Mathematics*, 22, 203-217.

Aldrin, M., & Damsleth, E. (1989). Forecasting non-seasonal time series with missing observations, *Journal of Forecasting*, 8, 97-116.

Anderson, J.R. (1994). Simpler exponentially weighted moving averages with irregular updating periods, *Journal of the Operational Research Society*, 45, 486.

Andreassen, P.B., & Kraus, S.J. (1990). Judgmental extrapolation and the salience of change, *Journal of Forecasting*, 9, 347-372.

Andrews, R.L. (1994). Forecasting performance of structural time series models, *Journal of Business and Economic Statistics*, 12, 129-133.

Archibald, B.C. (1990). Parameter space of the Holt-Winters' model, *International Journal of Forecasting*, 6, 199-209.

Archibald, B.C., & Koehler, A.B. (2003). Normalization of seasonal factors in Winters' methods, *International Journal of Forecasting*, 19, 143-148.

Arinze, B. (1994). Selecting appropriate forecasting models using rule induction, *Omega*, 22, 647-658.

Armstrong, J.S. (1984). Forecasting by extrapolation: Conclusions from 25 years of research, *Interfaces*, 14, 52-66.

Armstrong, J.S., & Collopy, F. (1993). Causal forces: Structuring knowledge for time-series extrapolation, *Journal of Forecasting*, 12, 103-115.

Artto, K.A., & Pylkkänen, E. (1999). An effective procedure for the distribution of magazines, *International Transactions in Operational Research*, 6, 289-310.

Assimakopoulos, V. & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting, *International Journal of Forecasting*, 16, 521-530.

Baker, J.R., & Fitzpatrick, K.E. (1986). Determination of an optimal forecast model for ambulance demand using goal programming, *Journal of the Operational Research Society*, 37, 1047-1059.

Bartolomei, S.M., & Sweet, A.L. (1989). A note on a comparison of exponential smoothing methods for forecasting seasonal series, *International Journal of Forecasting*, 5, 111-116.

Bell, P.C. (1978). A new procedure for the distribution of periodicals, *Journal of the Operational Research Society*, 29, 427-434.

Bell, P.C. (1981). Adaptive sales forecasting with many stockouts, *Journal of the Operational Research Society*, 32, 865-873.

Bell, P.C. (2000). Forecasting demand variation when there are stockouts, *Journal of the Operational Research Society*, 51, 358-363.

Bianchi, L., Jarrett, J., & Hanumara, R.C. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention, *International Journal of Forecasting*, 14, 497-504.

Billah, B., Hyndman, R.J., & Koehler, A.B. (2005). Empirical information criteria for time series forecasting model selection, *Journal of Statistical Computation and Simulation*, forthcoming.

Bodo, G., & Signorini, L.F. (1987). Short-term forecasting of the industrial production index, *International Journal of Forecasting*, 3, 245-259.

Bossons, J. (1966). The effects of parameter misspecification and non-stationary on the applicability of adaptive forecasts, *Management Science*, 12, 659-669.

Bowerman, B.L., O'Connell, R., & Koehler, A.B. (2005). *Forecasting, Time Series, and Regression (4th edition)*, Pacific Grove, CA: Duxbury Press.

Brown, R.G. (1959). *Statistical Forecasting for Inventory Control*, New York: McGraw-Hill.

Brown, R.G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, NJ: Prentice-Hall.

Brown, R.G. (1967). *Decision Rules for Inventory Management*, New York: Holt, Rinehart, and Winston.

Broze, L., & Mélard, G. (1990). Exponential smoothing: Estimation by maximum likelihood, *Journal of Forecasting*, 9, 445-455.

Carreno, J.J., & Madinaveitia, J. (1990). A modification of time series forecasting methods for handling announced price increases, *International Journal of Forecasting*, 6, 479-484.

Chambers, M.L., & Eglese, R.W. (1988). Forecasting demand for mail order catalogue lines during the season, *European Journal of Operational Research*, 34, 131-138.

Chan, C.K., Kingsman, B.G., & Wong, H. (1999). The value of combining forecasts in inventory management – A case study in banking, *European Journal of Operational Research*, 117, 199 - 210.

Chandra, C., & Grabis, J. (2005). Application of multiple-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand, *European Journal of Operational Research*, 166, 337-350.

Chatfield, C. (1988). What is the 'best method' of forecasting?, *Journal of Applied Statistics*, 15, 19-38.

Chatfield, C. (1993). Calculating interval forecasts, *Journal of Business and Economic Statistics*, 11, 121-135.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference, *Journal of the Royal Statistical Society, Series A*, 158, 419-466.

Chatfield, C. (1996). Model uncertainty and forecast accuracy, *Journal of Forecasting*, 15, 495-508.

Chatfield, C. (1997). Forecasting in the 1990s, *Journal of the Royal Statistical Society, Series D*, 46, 461-473.

Chatfield, C. (2002). Confessions of a pragmatic statistician, *Journal of the Royal Statistical Society, Series D*, 51, 1-20.

Chatfield, C. (2004). *The analysis of time series: An introduction*, 6th edition, Boca Raton: Chapman & Hall/CRC Press.

Chatfield, C., & Koehler, A.B. (1991). On confusing lead time demand with h-period-ahead forecasts, *International Journal of Forecasting*, 7, 239-240.

Chatfield, C., Koehler, A.B., Ord, J.K., & Snyder, R.D. (2001). A new look at models for exponential smoothing, *Journal of the Royal Statistical Society, Series D*, 50, 147-159.

Chatfield, C., & Yar, M. (1988). Holt-Winters forecasting: Some practical issues, *Journal of the Royal Statistical Society, Series D*, 37, 129-140.

Chatfield, C., & Yar, M. (1991). Prediction intervals for multiplicative Holt-Winters, *International Journal of Forecasting*, 7, 31-37.

Chen, C. (1997). Robustness properties of some forecasting methods for seasonal time series: A Monte Carlo study, *International Journal of Forecasting*, 13, 269-280.

Chu, S.C.K., & Lin, C.K.Y. (1994). Cohort analysis technique for long-term manpower planning: The case of a Hong Kong tertiary institution, *Journal of the Operational Research Society*, 45, 696-709.

Cipra, T. (1992). Robust exponential smoothing, *Journal of Forecasting*, 11, 57-69.

Cipra, T., Trujillo, J., & Rubio, A. (1995). Holt-Winters method with missing observations, *Management Science*, 41, 174-178.

Clarke, S.R. (1993). Computer forecasting of Australian rules football for a daily newspaper, *Journal of the Operational Research Society*, 44, 753-759.

Cogger, K.O. (1973). Specification analysis, *Journal of the American Statistical Association*, 68, 899-905.

Cohen, G.D. (1963). A note on exponential smoothing and autocorrelated inputs, *Operations Research*, 11, 361-366.

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

Cox, D.R. (1961). Prediction by exponentially weighted moving averages and related methods, *Journal of the Royal Statistical Society, Series B*, 23, 414-422.

Croston, J.D. (1972). Forecasting and stock control for intermittent demands, *Operational Research Quarterly*, 23, 289-303.

Curry, D.J., Divakar, S., Mathur, S.K., & Whiteman, C.H. (1995). BVAR as a category management tool: An illustration and comparison with alternative techniques, *Journal of Forecasting*, 14, 181-199.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., & Towill, D.R. (2003). Measuring and avoiding the bullwhip effect: A control theoretic approach, *European Journal of Operational Research*, 147, 567-590.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., & Towill, D.R. (2004). The impact of information enrichment on the Bullwhip effect in supply chains: A control engineering perspective, *European Journal of Operational Research*, 153, 727-750.

Eaves, A.H.C., & Kingsman, B.G. (2004). Forecasting for the ordering and stock-holding of spare parts, *Journal of the Operational Research Society*, 55, 431-437.

Enns, P.G. , Machak, J.A., Spivey, W.A., & Wrobleski, W.J. (1982). Forecasting applications of an adaptive multiple exponential smoothing model, *Management Science*, 28, 1035-1044.

Fairfield, R.P., & Kingsman, B.G. (1993). Control theory in production/inventory systems: A case study in a food processing organization, *Journal of the Operational Research Society*, 44, 1173-1182.

Farnum, N.R. (1992). Exponential smoothing: Behavior of the *ex-post* sum of squares near 0 and 1, *Journal of Forecasting*, 11, 47-56.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods, *International Journal of Forecasting*, 8, 81-98.

Fildes, R. (2001). Beyond forecasting competitions, *International Journal of Forecasting*, 17, 556-560.

Fildes, R., & Beard, C. (1992). Forecasting systems for production and inventory control, *International Journal of Operations and Production Management*, 12, 4-27.

Fildes, R., Hibon, M., Makridakis, S., & Meade, N. (1998). Generalising about univariate forecasting methods: Further empirical evidence, *International Journal of Forecasting*, 14, 339-358.

Fildes, R., Randall, A., & Stubbs, P. (1997). One day ahead demand forecasting in the utility industries: Two case studies, *Journal of the Operational Research Society*, 48, 15-24.

Flores, B.E., Olson, D.L., & Pearce, S.L. (1993). Use of cost and accuracy measures in forecasting method selection: A physical distribution example, *International Journal of Production Research*, 31, 139-160.

Flores, B.E. & Pearce, S.L. (2000). The use of an expert system in the M3 competition, *International Journal of Forecasting*, 16, 485-496.

García-Flores, R., Wang X.Z., & Burgess, T.F. (2003). Tuning inventory policy parameters in a small chemical company, *Journal of the Operational Research Society*, 54, 350-361

Gardner, E.S., Jr. (1983). Automatic monitoring of forecast errors, *Journal of Forecasting*, 2, 1-21.

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

Gardner, E.S., Jr. (1988). A simple method of computing prediction intervals for time series forecasts, *Management Science*, 34, 541-546.

Gardner, E.S., Jr. (1990). Evaluating forecast performance in an inventory control system, *Management Science*, 36, 490-499.

Gardner, E.S., Jr. (1993). Forecasting the failure of component parts in computer systems: A case study, *International Journal of Forecasting*, 9, 245-253.

Gardner, E.S., Jr. (1999). Note: Rule-based forecasting vs. damped-trend exponential smoothing, *Management Science*, 45, 1169-1176.

Gardner, E.S., Jr., & Anderson, E.A. (1997). Focus forecasting reconsidered, *International Journal of Forecasting*, 13, 501-508.

Gardner, E.S., Jr., Anderson-Fletcher, E.A., & Wicks, A.M. (2001). Further results on focus forecasting vs. exponential smoothing, *International Journal of Forecasting*, 17, 287-293.

Gardner, E.S., Jr., & Diaz-Saiz, J. (2002). Seasonal adjustment of inventory demand series: A case study, *International Journal of Forecasting*, 18, 117-123.

Gardner, E.S., Jr., & Koehler, A.B. (2005). Comments on a patented forecasting method for intermittent demand, *International Journal of Forecasting*, forthcoming.

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

Gardner, E.S., Jr., & McKenzie, E. (1988). Model identification in exponential smoothing, *Journal of the Operational Research Society*, 39, 863-867.

Gardner, E.S., Jr., & McKenzie, E. (1989). Seasonal exponential smoothing with damped trends, *Management Science*, 35, 372-376.

Gass, S.I., & Harris, C.M. (Eds.) (2000). *Encyclopedia of Operations Research and Management Science (Centennial edition)*, Dordrecht, The Netherlands: Kluwer.

Geriner, P.T., & Ord, J.K. (1991). Automatic forecasting using explanatory variables: A comparative study, *International Journal of Forecasting*, 7, 127-140.

Geurts, M.D., & Kelly, J.P. (1986). Forecasting retail sales using alternative models, *International Journal of Forecasting*, 2, 261-272.

Gijbels, I., Pope, A., & Wand, M.P. (1999). Understanding exponential smoothing via kernel regression, *Journal of the Royal Statistical Society, Series B*, 61, 39-50.

Gilchrist, W.W. (1976). *Statistical Forecasting*, London: Wiley.

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime, *International Journal of Forecasting*, 19, 579-594.

Grubb, H., & Mason, A. (2001). Long lead-time forecasting of UK air passengers by Holt-Winters methods with damped trend, *International Journal of Forecasting*, 17, 71-82.

Harrison, P.J. (1967). Exponential smoothing and short-term sales forecasting, *Management Science*, 13, 821-842.

Harvey, A.C. (1984). A unified view of statistical forecasting procedures, *Journal of Forecasting*, 3, 245-275.

Harvey, A.C. (1986). Analysis and generalisation of a multivariate exponential smoothing model, *Management Science*, 32, 374-380.

Harvey, A.C., & Koopman, S.J. (2000). Signal extraction and the formulation of unobserved components models, *Econometrics Journal*, 3, 84-107.

Harvey, A.C., & Snyder, R.D. (1990). Structural time series models in inventory control, *International Journal of Forecasting*, 6, 187-198.

Heuts, R.M.J., & Bronckers, J.H.J.M. (1988). Forecasting the Dutch heavy truck market: A multivariate approach, *International Journal of Forecasting*, 4, 57-79.

Hill, A.V., & Benton, W.C. (1992). Modelling intra-city time-dependent travel speeds for vehicle scheduling problems, *Journal of the Operational Research Society*, 43, 343-351.

Holmes, R.A. (1986). Leading indicators of industrial employment in British Columbia, *International Journal of Forecasting*, 2, 87-100.

Holt, C.C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages, *ONR Memorandum (Vol. 52)*, Pittsburgh, PA: Carnegie Institute of Technology. Available from the Engineering Library, University of Texas at Austin.

Holt, C.C. (2004a). Forecasting seasonals and trends by exponentially weighted moving averages, *International Journal of Forecasting*, 20, 5-10.

Holt, C.C. (2004b). Author's retrospective on 'Forecasting seasonals and trends by exponentially weighted moving averages,' *International Journal of Forecasting*, 20, 11-13.

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

Huss, W.R. (1985a). Comparative analysis of company forecasts and advanced time series techniques using annual electric utility energy sales data, *International Journal of Forecasting*, 1, 217-239.

Huss, W.R. (1985b). Comparative analysis of load forecasting techniques at a southern utility, *Journal of Forecasting*, 4, 99-107.

Hyndman, R.J. (2001). It's time to move from 'what' to 'why,' *International Journal of Forecasting*, 17, 567-570.

Hyndman, R.J., & Billah, B. (2003). Unmasking the Theta method, *International Journal of Forecasting*, 19, 287-290.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting*, 18, 439-454.

Hyndman, R.J., Akram, M., & Archibald, B. (2005a). The admissible parameter space for exponential smoothing models, working paper, Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia.

Hyndman, R.J., Koehler, A.B., Ord, J.K., & Snyder, R.D. (2005b). Prediction intervals for exponential smoothing using two new classes of state space models, *Journal of Forecasting*, 24, 17-37.

Johnston, F.R. (1993). Exponentially weighted moving average (EWMA) with irregular updating periods, *Journal of the Operational Research Society*, 44, 711-716.

Johnston, F.R. (2000). Viewpoint: Lead time demand adjustments or when a model is not a model, *Journal of the Operational Research Society*, 51, 1107-1110.

Johnston, F.R., & Boylan, J.E. (1994). How far ahead can an EWMA model be extrapolated?, *Journal of the Operational Research Society*, 45, 710-713.

Johnston, F.R., & Boylan, J.E. (1996a). Forecasting for items with intermittent demand, *Journal of the Operational Research Society*, 47, 113-121.

Johnston, F.R., & Boylan, J.E. (1996b). Forecasting intermittent demand: A comparative evaluation of Croston's method. Comment, *International Journal of Forecasting, 12*, 297-298.

Johnston, F.R., & Harrison, P.J. (1986). The variance of lead-time demand, *Journal of the Operational Research Society*, 37, 303-309.

Jones, R.H. (1966). Exponential smoothing for multivariate time series, *Journal of the Royal Statistical Society, Series B*, 28, 241-251.

Jun, D.B. (1989). On detecting and estimating a major level or slope change in general exponential smoothing, *Journal of Forecasting*, 8, 55-64.

Jun, D.B., & Oliver, R.M. (1985). Bayesian forecasts following a major level change in exponential smoothing, *Journal of Forecasting*, 4, 293-302.

Kirkendall, N.J. (1992). Monitoring for outliers and level shifts in Kalman Filter implementations of exponential smoothing, *Journal of Forecasting*, 11, 543-560.

Koehler, A.B. (1985). Simple vs. complex extrapolation models, *International Journal of Forecasting*, 1, 63-68.

Koehler, A.B. (1990). An inappropriate prediction interval, *International Journal of Forecasting*, 6, 557-558.

Koehler, A.B., & Murphree, E.S. (1988). A comparison of results from state space forecasting with forecasts from the Makridakis competition, *International Journal of Forecasting*, 4, 45-55.

Koehler, A.B., Snyder, R.D., & Ord, J.K. (2001). Forecasting models and prediction intervals for the multiplicative Holt-Winters method, *International Journal of Forecasting*, 17, 269-286.

Lawton, R. (1998). How should additive Holt-Winters estimates be corrected?, *International Journal of Forecasting*, 14, 393-403.

Lee, T.S., Cooper, F.W., & Adam, E.E., Jr. (1993). The effects of forecasting errors on the total cost of operations, *Omega*, 21, 541-550.

Lee, T.S., Feller, S.J., & Adam, E.E., Jr. (1992). Applying contemporary forecasting and computer technology for competitive advantage in service operations, *International Journal of Operations and Production Management*, 12, 28-42.

Leung, M.T., Daouk, H., & Chen, A.-S. (2000). Forecasting stock indices: A comparison of classification and level estimation models, *International Journal of Forecasting*, 16, 173-190.

Levén, E., & Segerstedt, A. (2004). Inventory control with a modified Croston procedure and Erlang distribution, *International Journal of Production Economics*, 90, 361-367.

Lin, W.T. (1989). Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches, *International Journal of Forecasting*, 5, 195-208.

Mahmoud, E., Motwani, J., & Rice, G. (1990). Forecasting US exports: An illustration using time series and econometric models, *Omega*, 18, 375-382.

Mahmoud, E., & Pegels, C.C. (1990). An approach for selecting times series forecasting models, *International Journal of Operations and Production Management*, 10, 50-60.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *Journal of Forecasting*, 1, 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, J.K., & Simmons, L.F. (1993). The M2-Competition: A real-time judgmentally based forecasting study, *International Journal of Forecasting*, 9, 5-22.

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

Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476.

Martin, C.A., & Witt, S.F. (1989). Forecasting tourism demand: A comparison of the accuracy of several quantitative methods, *International Journal of Forecasting*, 5, 7-19.

Masuda, Y., & Whang, S. (1999). Dynamic pricing for network service: Equilibrium and stability, *Management Science*, 45, 857-869.

Mathews, B.P., & Diamantopoulos, A. (1994). Towards a taxonomy of forecast error measures: A factor-comparative investigation of forecast error dimensions, *Journal of Forecasting*, 13, 409-416.

McClain, J.O. (1974). Dynamics of exponential smoothing with trend and seasonal terms, *Management Science*, 20, 1300-1304.

McClain, J.O. & Thomas, L.J. (1973). Response-variance tradeoffs in adaptive forecasting, *Operations Research*, 21, 554-568.

McKenzie, E. (1985). Comments on 'Exponential Smoothing: The State of the Art' by E.S. Gardner, Jr., *Journal of Forecasting*, 4, 32-36.

McKenzie, E. (1986). Error analysis for Winters' additive seasonal forecasting system, *International Journal of Forecasting*, 2, 373-382.

Meade, N. (2000). Evidence for the selection of forecasting methods, *Journal of Forecasting*, 19, 515-535.

Mentzer, J.T. (1988). Forecasting with adaptive extended exponential smoothing, *Journal of the Academy of Marketing Science*, 16 issue 3/4, 62-70.

Mentzer, J.T., & Gomes, R. (1994). Further extensions of adaptive extended exponential smoothing and comparison with the M-Competition, *Journal of the Academy of Marketing Science*, 22, 372-382.

Mercer, A., & Tao, X. (1996). Alternative inventory and distribution policies of a food manufacturer, *Journal of the Operational Research Society*, 47, 755-765.

Miller, T., & Liberatore, M. (1993). Seasonal exponential smoothing with damped trends: An application for production planning, *International Journal of Forecasting*, 9, 509-515.

Muth, J.F. (1960). Optimal properties of exponentially weighted forecasts, *Journal of the American Statistical Association*, 55, 299-306.

Nerlove, M., & Wage, S. (1964). On the optimality of adaptive forecasting, *Management Science*, 10, 207-224.

Newbold, P. (1988). Predictors projecting linear trend plus seasonal dummies, *Journal of the Royal Statistical Society, Series D*, 37, 111-127.

Newbold, P., & Bos, T. (1989). On exponential smoothing and the assumption of deterministic trend plus white noise data-generating models, *International Journal of Forecasting*, 5, 523-527.

Öller, L.-E. (1986). A note on exponentially smoothed seasonal differences, *Journal of Business and Economic Statistics*, 4, 485-489.

Ord, J.K. (2004). Charles Holt's report on exponentially weighted moving averages: An introduction and appreciation, *International Journal of Forecasting*, 20, 1-3.

Ord, J.K., Koehler, A.B., & Snyder, R.D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models, *Journal of the American Statistical Association*, 92, 1621-1629.

Ord, J.K., Snyder, R.D., Koehler, A.B., Hyndman, R.J., & Leeds, M. (2005). Time series forecasting: The case for the single source of error state space approach, working paper, Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia.

Pandit, S.M., & Wu, S.M. (1974). Exponential smoothing as a special case of a linear stochastic system, *Operations Research*, 22, 868-879.

Pandit, S.M., & Wu, S.M. (1983). *Time series and systems analysis with applications*, New York: Wiley.

Pantazopoulos, S.N., & Pappis, C.P. (1996). A new adaptive method for extrapolative forecasting algorithms, *European Journal of Operational Research*, 94, 106-111.

Pegels, C. (1969). Exponential forecasting: Some new variations, *Management Science*, 15, 311-315.

Pfeffermann, D., & Allon, J. (1989). Multivariate exponential smoothing: Method and practice, *International Journal of Forecasting*, 5, 83-98.

Price, D.H.R., & Sharp, J.A. (1986). A comparison of the performance of different univariate forecasting methods in a model of capacity acquisition in UK electricity supply, *International Journal of Forecasting*, 2, 333-348.

Proietti, T. (1998). Seasonal heteroscedasticity and trends, *Journal of Forecasting*, 17, 1-17.

Proietti, T. (2000). Comparing seasonal components for structural time series models, *International Journal of Forecasting*, 16, 247-260.

Ramanathan, R., Engle, R., Granger, C.W.J., Vahid-Araghi, F., & Brace, C. (1997). Short-run forecasts of electricity loads and peaks, *International Journal of Forecasting*, 13, 161-174.

Rao, A. (1973). A comment on: Forecasting and stock control for intermittent demands, *Operational Research Quarterly*, 24, 639-640.

Rasmussen, R. (2004). On time series data and optimal parameters, *Omega*, 32, 111-120.

Roberts, S.A. (1982). A general class of Holt-Winters type forecasting models, *Management Science*, 28, 808-820.

Rosanna, R.J., & Seater, J.J. (1995). Temporal aggregation and economic time series, *Journal of Business and Economic Statistics*, 13, 441-451.

Rosas, A.L., & Guerrero, V.M. (1994). Restricted forecasts using exponential smoothing techniques, *International Journal of Forecasting*, 10, 515-527.

Samuelson, D.A. (1999). Predictive dialing for outbound telephone call centers, *Interfaces*, 29 issue 5, 66-81.

Sani, B., & Kingsman, B.G. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items, *Journal of the Operational Research Society*, 48, 700-713.

Satchell, S., & Timmermann, A. (1995). On the optimality of adaptive expectations: Muth revisited, *International Journal of Forecasting*, 11, 407-416.

Schnaars, S.P. (1986). A comparison of extrapolation models on yearly sales forecasts, *International Journal of Forecasting*, 2, 71-85.

Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics,* 6, 461-464.

Segura, J.V., & Vercher, E. (2001). A spreadsheet modeling approach to the Holt-Winters optimal forecasting, *European Journal of Operational Research*, 131, 375-388.

Shah, C. (1997). Model selection in univariate time series forecasting using discriminant analysis, *International Journal of Forecasting*, 13, 489-500.

Sharda, R., & Musser, K.D. (1986). Financial futures hedging via goal programming, *Management Science*, 32, 933-947.

Sharp, J.A., & Price, D.H.R. (1990). Experience Curve models in the electricity supply industry, *International Journal of Forecasting*, 6, 531-540.

Shenstone, L., & Hyndman, R.J. (2005). Stochastic models underlying Croston's method for intermittent demand forecasting, *Journal of Forecasting,* forthcoming.

Shoesmith, G.L., & Pinder, J.P. (2001). Potential inventory cost reductions using advanced time series forecasting techniques, *Journal of the Operational Research Society,* 52, 1267-1275.

Snyder, R.D. (1985). Recursive estimation of dynamic linear models, *Journal of the Royal Statistical Society, Series B*, 47, 272-276.

Snyder, R.D. (1988). Progressive tuning of simple exponential smoothing forecasts, *Journal of the Operational Research Society*, 39, 393-399.

Snyder, R.D. (1993). A computerised system for forecasting spare parts sales: A case study, *International Journal of Operations and Production Management*,13, 83-92.

Snyder, R.D. (2002). Forecasting sales of slow and fast moving inventories, *European Journal of Operational Research*, 140, 684-699.

Snyder, R.D., Koehler, A.B., & Ord, J.K. (1999). Lead time demand for simple exponential smoothing: An adjustment factor for the standard deviation, *Journal of the Operational Research Society*, 50, 1079-1082.

Snyder, R.D., Koehler, A.B., & Ord, J.K. (2000). Viewpoint: A reply to Johnston, *Journal of the Operational Research Society*, 51, 1108-1110.

Snyder, R.D., Koehler, A.B., & Ord, J.K. (2002). Forecasting for inventory control with exponential smoothing, *International Journal of Forecasting*, 18, 5-18.

Snyder, R.D., Koehler, A.B., Hyndman, R.J., & Ord, J.K. (2004). Exponential smoothing models: Means and variances for lead-time demand, *European Journal of Operational Research*, 158, 444-455.

Snyder, R.D., Ord, J.K., & Koehler, A.B. (2001). Prediction intervals for ARIMA models, *Journal of Business and Economic Statistics*, 19, 217-225.

Snyder, R.D., & Shami, R.G. (2001). Exponential smoothing of seasonal data: A comparison, *Journal of Forecasting*, 20, 197-202.

Strijbosch, L.W.G., Heuts, R.M.J., & van der Schoot, E.H.M. (2000). A combined forecast-inventory control procedure for spare parts, *Journal of the Operational Research Society*, 51, 1184-1192.

Sweet, A.L. (1985). Computing the variance of the forecast error for the Holt-Winters seasonal models, *Journal of Forecasting*, 4, 235-243.

Syntetos, A.A., & Boylan, J.E. (2001). On the bias of intermittent demand estimates, *International Journal of Production Economics*, 71, 457-466.

Syntetos, A.A. & Boylan, J.E. (2005). The accuracy of intermittent demand estimates, *International Journal of Forecasting*, 21, 303-314.

Syntetos, A.A., Boylan, J.E., & Croston, J.D. (2005). On the categorization of demand patterns, *Journal of the Operational Research Society*, 56, 495-503.

Tashman, L.J., & Kruk, J.M. (1996). The use of protocols to select exponential smoothing procedures: A reconsideration of forecasting competitions, *International Journal of Forecasting*, 12, 235-253.

Tashman, L.J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review, *International Journal of Forecasting*, 16, 437-450.

Taylor, J.W. (2003a). Exponential smoothing with a damped multiplicative trend, *International Journal of Forecasting*, 19, 715-725.

Taylor, J.W. (2003b). Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, 54, 799-805.

Taylor, J.W. (2004a). Volatility forecasting with smooth transition exponential smoothing, *International Journal of Forecasting*, 20, 273-286.

Taylor, J.W. (2004b). Smooth transition exponential smoothing, *Journal of Forecasting*, 23, 385-404.

Taylor, J.W. (2004c). Forecasting with exponentially weighted quantile regression, working paper, Said Business School, University of Oxford, Park End St., Oxford OX1 1HP, UK.

Taylor, J.W., & Bunn, D.W. (1999). A quantile regression approach to generating prediction intervals, *Management Science*, 45, 225-237.

Theil, H., & Wage, S. (1964). Some observations on adaptive forecasting, *Management Science*, 10, 198-206.

Thomas, R.J. (1993). Method and situational factors in sales forecast accuracy, *Journal of Forecasting*, 12, 69-77.

Thury, G. (1985). Macroeconomic forecasting in Austria: An analysis of accuracy, *International Journal of Forecasting*, 1, 111-121.

Tiao, G.C., & Xu, D. (1993). Robustness of maximum likelihood estimates for multi-step predictions: The exponential smoothing case, *Biometrika*, 80, 623-641.

Vokurka, R.J., Flores, B.E., & Pearce, S.L. (1996). Automatic feature identification and graphical support in rule-based forecasting: a comparison, *International Journal of Forecasting* 12, 495-512.

Walton, J.H.D. (1994). Inventory control—A soft approach?, *Journal of the Operational Research Society*, 45, 485-486.

Weatherford, L.R., & Kimes, S.E. (2003). A comparison of forecasting methods for hotel revenue management, *International Journal of Forecasting*, 19, 401-415.

Weintraub, A., Aboud, J., Fernandez, C., Laporte, G., & Ramirez, E. (1999). An emergency vehicle dispatching system for an electric utility in Chile, *Journal of the Operational Research Society*, 50, 690-696.

Willemain, T.R., Smart, C.N., Shockor, J.H., & DeSautels, P.A. (1994). Forecasting intermittent demand in manufacturing: A comparative evaluation of Croston's method, *International Journal of Forecasting*, 10, 529-538.

Willemain, T.R., Smart, C.N., & Schwarz, H.F. (2004). A new approach to forecasting intermittent demand for service parts inventories, *International Journal of Forecasting*, 20, 375-387.

Williams, D.W., & Miller, D. (1999). Level-adjusted exponential smoothing for modeling planned discontinuities, *International Journal of Forecasting*, 15, 273-289.

Williams, T.M. (1987). Adaptive Holt-Winters forecasting, *Journal of the Operational Research Society*, 38, 553-560.

Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages, *Management Science*, 6, 324-342.

Wright, D.J. (1986a). Forecasting data published at irregular time intervals using an extension of Holt's method, *Management Science*, 32, 499-510.

Wright, D.J. (1986b). Forecasting irregularly spaced data: An extension of double exponential smoothing, *Computers and Industrial Engineering*, 10, 135-147.

Wu, L.S.-Y., Ravishanker, N., & Hosking, J.R.M. (1991). Forecasting for business planning: A case study of IBM product sales, *Journal of Forecasting*, 10, 579-595.

Yar, M., & Chatfield, C. (1990). Prediction intervals for the Holt-Winters forecasting procedure, *International Journal of Forecasting*, 6, 127-137.

Zhang, X. (2004). The impact of forecasting methods on the bullwhip effect, *International Journal of Production Economics*, 88, 15-27.

Zhao, X., Xie, J., & Leung, J. (2002). The impact of forecasting model selection on the value of information sharing in a supply chain, *European Journal of Operational Research*, 142, 321-344.