EXPONENTIAL SMOOTHING MODELS: MEANS AND VARIANCES FOR LEAD-TIME DEMAND

Ralph D. Snyder

Department of Econometrics and Business Statistics
P.O. Box 11E, Monash University, VIC 3800, Australia
Email: ralph.snyder@buseco.monash.edu.au

Anne B. Koehler

Department of Decision Sciences and Management Information Systems
Miami University, Oxford, OH 45056, USA.
Email: koehlerab@muohio.edu

Rob J. Hyndman

Department of Econometrics and Business Statistics
Monash University, VIC 3800, Australia
Email: rob.hyndman@buseco.monash.edu.au

J. Keith Ord
[Corresponding Author]

McDonough School of Business
320 Old North, Georgetown University, Washington, DC 20057, USA.
Email: ordk@georgetown.edu

ABSTRACT

Exponential smoothing is often used to forecast lead-time demand for inventory control. In this paper, formulae are provided for calculating means and variances of lead-time demand for a wide variety of exponential smoothing methods. A feature of many of the formulae is that variances, as well as the means, depend on trends and seasonal effects. Thus, these formulae provide the opportunity to implement methods that ensure that safety stocks adjust to changes in trend or changes in season. An example using weekly sales shows how safety stocks can be seriously underestimated during peak sales periods.

KEYWORDS

Forecasting; inventory; lead-time demand; exponential smoothing; forecast variance.

1. INTRODUCTION

Inventory control software typically contains a forecasting module that predicts the mean and variance of lead-time demand. These values are incorporated into an inventory control module for the determination of ordering parameters such as reorder levels, order-up-to levels and reorder quantities. These forecasting modules often rely upon exponential smoothing methods (initially introduced by R.G. Brown, 1959), as they are intuitively appealing, easy to update and have minimal computer storage requirements. Brown's initial methods, combined with Holt's (1957) local linear trend method and the Holt-Winters (Winters, 1960) schemes for series displaying both trend and seasonal patterns provide reasonably good coverage of likely behaviors to be met in practice, particularly when the damped trend method of Gardner and McKenzie (1985) is included. Overall, exponential smoothing methods have a proven record for generating sensible point forecasts (Gardner, 1985; Makridakis and Hibon, 2000). For a review of recent developments of statistical models for exponential smoothing, see Chatfield et al. (2001).

The basic problem in inventory control may be formulated as follows. Suppose that a replenishment decision is to be made at the beginning of period n+1. Any order placed at this time is assumed to arrive a lead-time later at the start of period $n+\lambda$. Inventory theory dictates that the primary focus should be on lead-time demand, an aggregate of unknown future values y_{n+j} defined by

$$Y_n(\lambda) = \sum_{j=1}^{\lambda} y_{n+j} . \qquad (1)$$

The problem is to make inferences about the distribution of lead-time demand. Typically an appropriate form of exponential smoothing is applied to past demand data $y_1, \ldots, y_n$, the results being used to predict the mean of the lead-time demand distribution. For most of the paper, we assume that λ is fixed, but in section 5 we briefly consider stochastic lead-times. Fixed lead-times are relevant when suppliers make regular deliveries, an increasingly common situation in supply chain management.

Many inventory management systems require the variance of lead-time demand in order to implement an inventory strategy, but the basic exponential smoothing procedures originally provided only point forecasts and rather ad-hoc formulae were the vogue in inventory control software. Then Johnston and Harrison (1986) derived a variance formula for use with simple exponential smoothing. Using a simple state space model, Johnston and Harrison utilized the fact that simple exponential smoothing emerges as the steady state form of the associated Kalman filter in large samples. Adopting a different model, Snyder, Koehler and Ord (1999) were able to obtain the same formula without recourse to the Kalman filter strategy. The advantage of their approach is that no restrictive large sample assumption is needed. Johnston and Harrison (1986) also obtained a variance formula for lead-time demand when trend-corrected exponential smoothing is employed. Yar and Chatfield (1990), however, have suggested a slightly different formula. They also provide a formula that incorporates seasonal effects for use with the additive Winters (1960) method. Harvey and Snyder (1990) obtain similar variance formulae for level, trend and seasonal cases using a structural time series framework. They rely on continuous time models so that the links with exponential smoothing are more obscure.

Most of the work discussed so far makes the (sometimes implicit) assumption that the variance of the DPUT (demand per unit time) process is constant. Yet, as Brown (1959, p. 94) observed "you will be very likely to find that the standard deviation of demand is nearly proportional to the total annual usage, or to the average monthly usage". Indeed, some authors in the inventory literature have built upon this idea, notably Miller (1986) and Lovejoy (1990). However, these authors assume zero lead-times. Heath and Jackson (1994) generate forecasts for individual future time periods, but do not examine lead-time demand. Thus, a systematic framework for the development of forecast variances for lead-time demand has not been available.

The purpose of this paper is to take a fresh look at the problem. We use the linear version of the single source of error model from Ord, Koehler and Snyder (1997) to unify the derivations. We also provide useful extensions to accommodate errors that depend on trend and seasonal effects. This aspect of the results is particularly important since the variance typically increases during peak sales periods so that safety stocks could be seriously underestimated at precisely those times that are potentially most profitable.

1.1 Structure of the paper

The model and its special cases are introduced in Section 2. Associated formulae for means and variances of lead-time demand are presented in Section 3. The general principles used in their derivation are presented in the Appendix. Some numerical examples, and the results from applying these formulae to real demand data, are explored in Section 4. Issues associated with stochastic lead-times are examined in Section 5 and conclusions and directions for further research are discussed in section 6.

Throughout the paper, we adopt a convention concerning the sum operator Σ. In those cases where the upper limit is less than the lower limit, the sum should be equated to zero.

2. Models for Exponential Smoothing

Future values of a time series are unknown and must be treated as random variables. Their behavior must be linked to a statistical model in order to derive prediction distributions. A model should have the potential to include unobserved components such as levels, growth rates and seasonal effects, because various forms of exponential smoothing are based on these concepts. Common cases of exponential smoothing and their models are shown in Table 1. The column marked 'Code' uses nomenclature from Hyndman et al (2001). Here N designates 'None', 'A' designates 'Additive' and D designates 'Damped'. All codes involve two letters. The first letter is used to describe the trend. The second letter describes the seasonal component. The various components are ℓ_t for local level, b_t for local growth rate, s_t for local seasonal effect and e_t for a random variable designating the unpredictable component. The α, β, γ are so-called smoothing parameters. The ϕ, another parameter, is a damping factor. The purpose of the caret symbol is outlined later.

Case	Code	Model	Smoothing Method	Description
1	NN	$y_t = \ell_{t-1} + e_t$ $l_t = l_{t-1} + \alpha e_t$	$\hat{y}_t = \hat{\ell}_{t-1}$ $\hat{\ell}_t = \hat{\ell}_{t-1} + \alpha(y_t - \hat{y}_t)$	Simple exponential smoothing (Brown, 1959)
2	AN	$y_t = \ell_{t-1} + b_{t-1} + e_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$ $b_t = b_{t-1} + \alpha\beta e_t$	$\hat{y}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1}$ $\hat{\ell}_{t-1} = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \alpha(y_t - \hat{y}_t)$ $\hat{b}_t = \hat{b}_{t-1} + \alpha\beta(y_t - \hat{y}_t)$	Trend-corrected exponential smoothing (Holt, 1957)
3	AD	$y_t = \ell_{t-1} + b_{t-1} + e_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$ $b_t = \phi b_{t-1} + \alpha\beta e_t$	$\hat{y}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1}$ $\hat{\ell}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \alpha(y_t - \hat{y}_t)$ $\hat{b}_t = \phi\hat{b}_{t-1} + \alpha\beta(y_t - \hat{y}_t)$	Damped trend (Gardner and McKenzie, 1985)
4		$y_t = s_{t-m} + e_t$ $s_t = s_{t-m} + \gamma e_t$	$\hat{y}_t = \hat{s}_{t-m}$ $\hat{s}_t = \hat{s}_{t-m} + \gamma(y_t - \hat{y}_t)$	Elementary seasonal case
5	AA	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + e_t$ $l\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$ $b_t = b_{t-1} + \alpha\beta e_t$ $s_t = s_{t-1} + \gamma e_t$	$\hat{y}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \hat{s}_{t-m}$ $\hat{\ell}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \alpha(y_t - \hat{y}_t)$ $\hat{b}_t = \hat{b}_{t-1} + \alpha\beta(y_t - \hat{y}_t)$ $\hat{s}_t = \hat{s}_{t-m} + \gamma(y_t - \hat{y}_t)$	Winters additive method (Winters, 1960)
6	DA	$y_t = \ell_{t-1} + b_{t-1} + c_{t-m} + e_t$ $l\ell_t = l\ell_{t-1} + b_{t-1} + \alpha e_t$ $b_t = \phi b_{t-1} + \alpha\beta e_t$ $s_t = s_{t-1} + \gamma e_t$	$\hat{y}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \hat{s}_{t-m}$ $\hat{\ell}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \alpha(y_t - \hat{y}_t)$ $\hat{b}_t = \phi\hat{b}_{t-1} + \alpha\beta(y_t - \hat{y}_t)$ $\hat{s}_t = \hat{s}_{t-m} + \gamma(y_t - \hat{y}_t)$	Damped trend with seasonal effects

Table 1. Models for Common Linear Forms of Exponential Smoothing.

Each model in Table 1 contains a measurement equation that specifies how a series value is built from unobserved components. It contains transition equations that describe how the unobserved components change over time in response to the effects of structural change. It involves a random variable representing the unpredictable component.

All the models in Table 1 are special cases of what is best called a single source of error state space model, introduced by Snyder (1985). The unobserved components are stacked to give a vector x_t. It is assumed that all components combine linearly to give the series value, so the measurement equation is specified as

$$y_t = h'x_{t-1} + e_t \tag{1}$$

where h is a fixed vector of coefficients. The lag on x_t is used to reflect the assumption that the conditions at time t-1 determine what happens during the period t. The evolution of the unobserved components is governed by the first-order transition relationship

$$x_t = Fx_{t-1} + ge_t \tag{2}$$

where F is a fixed matrix and g is a fixed vector that reflects the impact of structural change.

Example 1: For the AN model in Table 1, $h' = (1,1), x_t' = (\ell_t, b_t), F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, and\ g' = (\alpha, \alpha\beta).$

The first component of (1) is the underlying mean level, or one-step-ahead forecast, and we may designate it by $m_t = h'x_{t-1}$. The second component represents the unpredictable error or disturbance term. It is possible that the disturbance is completely independent of the mean level, but it is also possible that its variance increases with this level. For example, whenever sales variation is naturally thought of in terms of percentage changes, rather than absolute changes, the standard deviation is likely to depend on the mean. Both possibilities are captured by the assumption that the disturbance is governed by the relationship

$$e_t = m_t^{\,r}\varepsilon_t \quad \text{for} \quad r = 0,1 \tag{3}$$

where the $\{\varepsilon_t\}$ are independent and identically distributed with zero mean and variance σ^2, written as IID(0, σ^2). . The measurement equation may now be written as $y_t = m_t + \varepsilon_t$ when $r = 0$ or $y_t = m_t(1+\varepsilon_t)$ when $r = 1$. In the latter case, the ε_t is a unit-less quantity, conveniently thought of as a relative error. It means that the unpredictable component potentially depends on the other components of a time series, something that can be very important in practice. The elements h, F, g potentially depend on a vector of parameters designated by ω.

It is assumed that the same model governs both past and future values of a time series. Past values are known, in which case it is possible to make a pass through the data, applying a compatible form of exponential smoothing in each period. Suppose, at the beginning of typical period *t*, past applications of exponential smoothing have yielded the estimated value $\hat{x}_{t-1}$ for the state vector x_{t-1}. After observing y_t at the end of period *t*, it is possible to calculate the error $\hat{e}_t = y_t - h'\hat{x}_{t-1}$. The error can be substituted into the transition equation to give $\hat{x}_t = F\hat{x}_{t-1} + g\left(y_t - h'\hat{x}_{t-1}\right)$ for the estimated value of the state vector x_t. Given the progressive nature of this algorithm, it is clear that this estimate depends on the parameters, the starting values of the state variables and the observations through time *t*, which we write as $\hat{x}_t = x_t \mid y_1, ..., y_t, x_0, \omega$. Induction may be used to confirm that $\hat{x}_t$ is a fixed value.

A special case of the above model, best termed a composite model, is now considered. The state vector x_t is partitioned into random sub-vectors designated by $x_{1,t}$ and $x_{2,t}$. The measurement equation has the form

$$y_t = h_1' x_{1,t-1} + h_2' x_{2,t-1} + e_t \tag{4}$$

where h_1 and h_2 are sub-vectors of h. The sub-vectors of the state vector are governed by transition equations

$$x_{k,t} = F_k x_{k,t-1} + g_k e_t \quad \left(k = 1, 2\right) \tag{5}$$

where F_1, F_2 are transition matrices and g_1, g_2 are sub-vectors of *g*. The special feature of this composite model is that the transition equation for $x_{1,t}$ does not contain $x_{2,t}$ and vice versa. It is shown in the Appendix that the results for a composite model can be built directly from those of its constituent models.

All the models in Table 1 are special cases of the single source of error model or the composite model. The links with these general models are provided in Table 2. Here 0_k refers to a k-vector of zeros and I_k refers to a $k \times k$ identity matrix. Note that although the seasonal cases are governed by mth-order recurrence relationships, they are converted to equivalent first-order relationships. Also note that ω is a vector formed from some or all of the parameters $\alpha, \beta, \gamma, \phi$.

Case	x_t	h	F	g
1	$x_t = l_t$	$h = 1$	$F = 1$	$g = \alpha$
2	$x_t = [l_t, b_t]'$	$h' = [1 \quad 1]$	$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$g = [\alpha \quad \alpha\beta]'$
3	$x_t = [l_t, b_t]'$	$h' = [1 \quad 1]$	$F = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix}$	$g = [\alpha \quad \alpha\beta]'$
4	$x_t = [s_t, \ldots, s_{t-m+1}]'$	$h' = [0'_{m-1} \quad 1]$	$F = \begin{bmatrix} 0'_{m-1} & 1 \\ I_{m-1} & 0_{m-1} \end{bmatrix}$	$g = [\gamma \quad 0'_{m-1}]'$
5	$x_{1t} = [l_t, b_t]'$ $x_{2t} = [s_t, \ldots, s_{t-m+1}]'$	$h_1' = [1 \quad 1]$ $h_2' = [0'_{m-1} \quad 1]$	$F_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ $F_2 = \begin{bmatrix} 0'_{m-1} & 1 \\ I_{m-1} & 0_{m-1} \end{bmatrix}$	$g_1 = [\alpha \quad \alpha\beta]'$ $g_2 = [\gamma \quad 0'_{m-1}]'$
6	$x_{1t} = [l_t, b_t]'$ $x_{2t} = [s_t, \ldots, s_{t-m+1}]'$	$h_1' = [1 \quad 1]$ $h_2' = [0'_{m-1} \quad 1]$	$F_1 = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix}$ $F_2 = \begin{bmatrix} 0'_{m-1} & 1 \\ I_{m-1} & 0_{m-1} \end{bmatrix}$	$g_1 = [\alpha \quad \alpha\beta]'$ $g_2 = [\gamma \quad 0'_{m-1}]'$

Table 2. Conformity of Special Cases to the General Model or Composite Model.

In the homoscedastic cases, only the mean potentially depends on trend and seasonal effects. However, in the heteroscedastic cases, both the mean and the variance of the random error component depend on trend and seasonal effects. Thus, prediction variances reflect trend and seasonal effects in the heteroscedastic case, a feature that is potentially quite useful in practice.

An intriguing insight from Table 2 is that each smoothing method applies for both a homoscedastic and a heteroscedastic model. Now, each homoscedastic case is equivalent to an ARIMA process (Box, Jenkins and Reinsel, 1994). However, no heteroscedastic case is equivalent to an ARIMA process. Thus, exponential smoothing applies for a wider class of models than the ARIMA class (Ord, Koehler and Snyder, 1997).

Many other cases are conceivable when addition operators are replaced in the measurement equation by multiplications. Examples of such cases are presented in Hyndman, Koehler, Snyder and Grose (2002). A variety of models underlying the multiplicative version of Winters multiplicative method have been introduced in Koehler, Snyder and Ord (2001). The complexity of these nonlinear possibilities precludes the derivation of results using the methodology of this paper.

3. MEANS AND VARIANCES OF LEAD TIME DEMAND

For the purposes of the present discussion, we assume that methods similar to those described in Ord, Koehler and Snyder (1997) have been applied to past demand data to estimate the parameters of an appropriate model. The problem is now to find the mean and variance of the lead-time demand distribution. Our analysis is built, in part, on prediction variance results from Hyndman, Koehler, Ord and Snyder (2001) for conventional prediction distributions. As noted earlier, we assume the lead-time λ to be fixed; this assumption is relaxed for a special case in section 5.

It is shown in the Appendix that lead-time demand can be resolved into a linear function of the uncorrelated level and error components:

$$Y_n(\lambda) = \sum_{j=1}^{\lambda} \mu_{n+j} + \sum_{j=1}^{\lambda} C_j e_{n+j} \ . \qquad (6)$$

where

$$\mu_{n+j} = h' F^{j-1} x_n \qquad (7)$$

is the mean of the *j*-step prediction distribution. It is further established that the coefficients of the errors in (6) are given by

$$C_j = 1 + \sum_{i=1}^{\lambda-j} c_i \quad \text{for } j = 1, \ldots, \lambda, \tag{8}$$

where

$$c_i = h' F^{i-1} g \,. \tag{9}$$

Particular cases of the formulae for the means μ_{n+j} and the coefficients C_j are shown in Table 3. Note that $\phi_j = \sum_{i=0}^{j-1} \phi^i$; $\phi_j^{(2)} = \sum_{i=1}^{j-1} i\phi^i$; $p = \left\lceil \frac{j+m-1}{m} \right\rceil$; $d_{j,m} = 1$ if j is a multiple of *m* and $d_{j,m} = 0$ otherwise. The results for Case 5 and Case 6 are constructed by adding the corresponding results for constituent basic models, an approach that is also rationalized in the Appendix.

Case	μ_{n+j}	c_j	C_j
1	$\hat{\ell}_n$	α	$1+(\lambda-j)\alpha$
2	$\hat{\ell}_n + j\hat{b}_n$	$\alpha(1+j\beta)$	$1+(\lambda-j)\alpha+\frac{(\lambda-j)(\lambda-j+1)}{2}\alpha\beta$
3	$\hat{l}_n + \phi_j \hat{b}_n$	$\alpha(1+\beta\phi_j)$	$1+(\lambda-j)\alpha+(\lambda-j)\alpha\beta\phi_{\lambda-j}-\alpha\beta\phi_{\lambda-j}^{(2)}$
4	$\hat{s}_{n+j-pm}$	$d_{j,m}\gamma$	$1+\gamma\sum_{i=1}^{\lambda-j} d_{i,m}$
5	$\hat{\ell}_n + j\hat{b}_n + \hat{s}_{n+j-pm}$	$\alpha(1+j\beta)+d_{j,m}\gamma$	$1+(\lambda-j)\alpha+\frac{(\lambda-j)(\lambda-j+1)}{2}\alpha\beta+\gamma\sum_{i=1}^{\lambda-j} d_{i,m}$
6	$\hat{\ell}_n + \phi_j\hat{b}_n + \hat{s}_{n+j-pm}$	$\alpha(1+\beta\phi_j)+d_{j,m}\gamma$	$1+(\lambda-j)\alpha+(\lambda-j)\alpha\beta\phi_{\lambda-j}-\alpha\beta\phi_{\lambda-j}^{(2)}+\gamma\sum_{i=1}^{\lambda-j} d_{i,m}$

Table 3. Key Results for Basic models.

From (6), the conditional variance is given by

$$\operatorname{var}\left(Y_n(\lambda) \mid x_n, \omega\right) = \sigma^2 \sum_{j=1}^{\lambda} C_j^2 \,. \tag{10}$$

in the homoscedastic case. All the information needed to evaluate the grand mean and the grand variance is available in Table 3. In the heteroscedastic case the grand variance is

$$\operatorname{var}\left(Y_n(\lambda) \mid x_n, \omega\right) = \sigma^2 \sum_{j=1}^{\lambda} C_j^2 \theta_{n+j} \tag{11}$$

where $\theta_{n+j} = E\left(m_{n+j}^2 \mid x_n, \omega\right)$. It is established, in the Appendix, that the heteroscedastic formulae may be computed using the recurrence relationship

$$\theta_{n+j} = \mu_{n+j}^2 + \sum_{i=1}^{j-1} c_{j-i}^2 \theta_{n+i} \sigma^2 \tag{12}$$

where the c_j are also given in Table 3.

In common with most of the literature on inventory systems, we have derived only the mean and variance for lead-time demand (LTD). Safety stocks are then determined assuming the LTD to be normally distributed. In the homoscedastic case, LTD will be normal if the errors are normal, but the LTD is only approximately normal in the heteroscedastic case even when a normal error process is assumed. However, a numerical study in Hyndman, Koehler, Ord and Snyder (2001) indicates that there is little error involved in approximating these distributions by the normal. The same conclusion must apply to lead-time distributions where aggregation must help to further reduce the approximation error.

4. Examples

To gauge whether a move to multiplicative models from the simpler additive models could be worthwhile in practice, we examine the differences between them with two sets of examples. In the first, we explore the effect of changes in lead-times. In the second, we provide evidence based upon weekly sales data for a particular product with a seasonal sales pattern.

4.1 Numerical comparisons

We consider the process corresponding to simple exponential smoothing, given in Table 1 as case NN. Consider the following parameter settings:

Parameter	Description	Values
λ	Length of lead-time	5, 20
m_t	Mean level at start of period t+1	25, 100
α	Smoothing constant	0.1, 0.5
σ	Standard deviation of ε_t in (3), additive scheme	5, 15
κ	Standard deviation of ε_t in (3), multiplicative scheme	See text

For the multiplicative scheme, the standard deviation of the error, e_t in equation (3) is κm_t where $\mathrm{var}(\varepsilon_t) = \kappa^2$. In our calculations, the value of κ is set equal to σ/m_t to maintain the same variance for DPUT between the two schemes.

The results appear in Table 4. By construction, the means for the two schemes are identical. The last column of the table shows the ratios of the standard deviations for the multiplicative to additive schemes. As might be expected, these ratios are typically close to 1.0, and increase substantially only when the lead-time is long, the coefficient of variation [ratio of standard deviation to mean] for DPUT is (unreasonably) high, and the correlation between demands for successive periods is high [high alpha]. These results are consistent with those given by Hyndman, Koehler, Ord and Snyder (2001).

Now suppose that demand is seasonal. For simplicity, we consider a simplified version of case AA in Table 1, for which there is no slope and the seasonal effects are fixed, so that $\beta = \gamma = b_0 = 0$. Further, to make the interpretation more direct, we assume that the seasonal effect is an upward shift in mean DPUT, so that the expected level at the beginning of period (t+1) becomes $(1+c)\ell_t$ and is sustained throughout the lead-time. Comparison of the variance formulae given in (10) and (11) reveals that the ratio of standard deviations (multiplicative to additive) reduces to (1+c). The clear implication is that if the additive scheme is used to compute safety stock, the service level will be well below that target figure, with consequent likely increases in lost sales. Conversely, in a period of low DPUT, inventories would be excessive. This question is examined further in section 5.

Table 4: Comparison of additive and multiplicative models, with Lead Times fixed

LAMBDA	LEVEL	ALPHA	SIGMA	KAPPA	MEAN	SD(A)	SD(M)	SD ratio
5	25	0.1	5	0.2	125	13.5	13.5	1.00
5	25	0.1	15	0.6	125	40.5	40.6	1.00
5	25	0.5	5	0.2	125	23.7	23.8	1.01
5	25	0.5	15	0.6	125	71.2	74.8	1.05
5	100	0.1	5	0.05	500	13.5	13.5	1.00
5	100	0.1	15	0.15	500	40.5	40.5	1.00
5	100	0.5	5	0.05	500	23.7	23.7	1.00
5	100	0.5	15	0.15	500	71.2	71.4	1.00
20	25	0.1	5	0.2	500	45.5	45.5	1.00
20	25	0.1	15	0.6	500	136.4	138.0	1.01
20	25	0.5	5	0.2	500	143.8	147.4	1.03
20	25	0.5	15	0.6	500	431.5	551.5	1.28
20	100	0.1	5	0.05	2000	45.5	45.5	1.00
20	100	0.1	15	0.15	2000	136.4	136.5	1.00
20	100	0.5	5	0.05	2000	143.8	144.1	1.00
20	100	0.5	15	0.15	2000	431.5	437.5	1.01

MEAN: mean lead-time demand
SD(A), SD(M): standard deviations for additive and multiplicative schemes respectively
SD ratio = SD(M)/SD(A)

4.2 Demand for costume jewelry

To examine the potential in practice for such effects as those described in section 4.1, we chose the real time series shown in Figure 1. This series shows the weekly demand for a particular costume jewelry product[1] in the United States, covering the time period 1998, week 5 to 2000, week 24. This product is one of several hundred produced by the company and many of them show seasonal patterns similar to this one. The pronounced increase in sales in the period between Thanksgiving (end of November) and Christmas is obvious [corresponding to observations 43-47 and 95-99 in the figure] and is widely anticipated in the

[1] We are grateful to Bill Sichel for providing this data set.

retail trade. Given that the series possesses such pronounced seasonal peaks, case 5 of the models from Table 1 was fitted using the conditional maximum likelihood approach described in Ord, Koehler and Snyder (1997). The maximum likelihood estimates of the smoothing parameters turned out to be $\alpha = 0.35$ and. $\beta = \gamma = 0$. These results indicate the presence of a constant growth rate and an invariant seasonal cycle; in other words, precisely the seasonal case examined in section 4.1. The point predictions for the demands in individual future weeks are plotted in Panel A of Figure 2, using 2000 week 24 as origin.

Panel B of Figure 2 is to be interpreted as follows. The variance of the lead-time demand is plotted on the vertical axis, and the lead-time is plotted on the horizontal axis, corresponding to successive lead-times of 1, 2, …, 52 weeks. The additive model would produce a variance plot that is almost quadratic; the standard deviations would produce almost a straight line. By contrast, the variances of aggregate demand show a marked rate of increase in response to peaking seasonal effects. From Figure 1, the increase in variability during the seasonal peaks is very evident so that if the company were to use a constant variance model, it would seriously underestimate the safety stocks required during these peak periods.

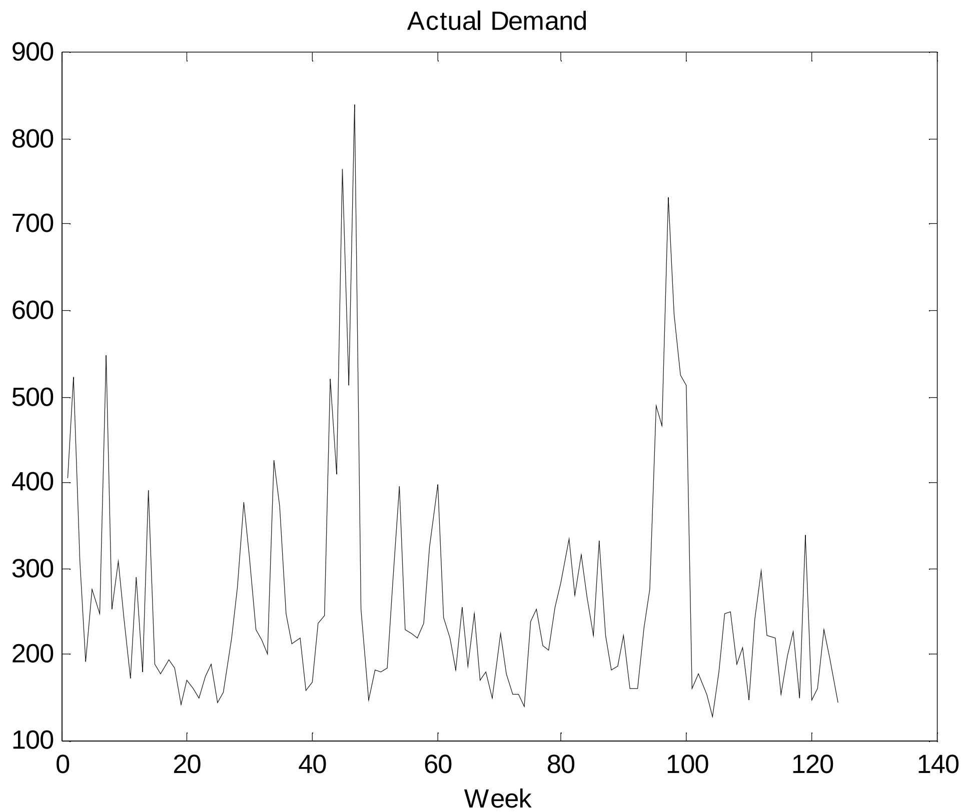

Figure 1. Weekly Demand

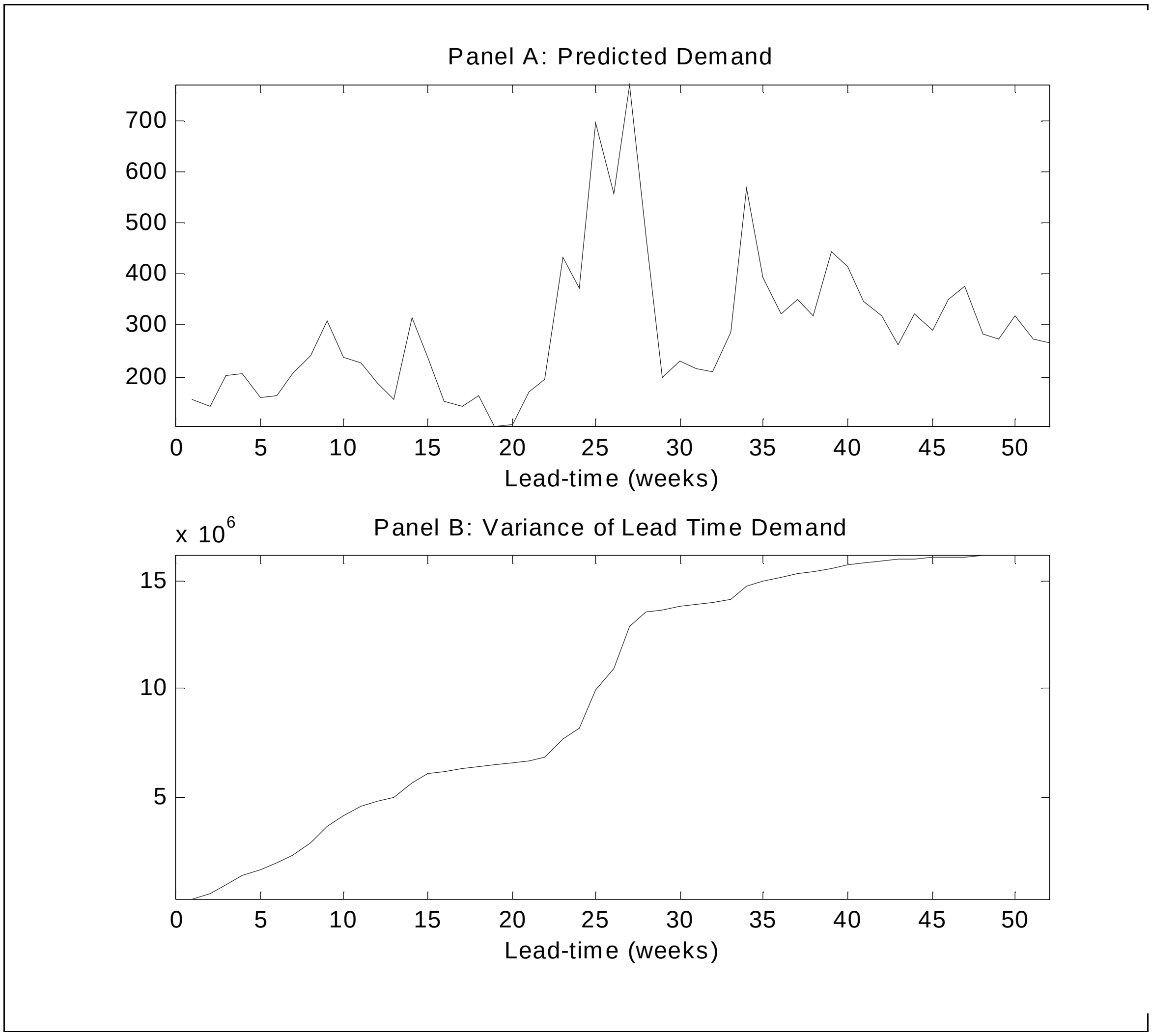

Figure 2: Panel A shows the predicted demand for individual weeks. Panel B shows the variance of lead-time demand when the lead-time is as on the horizontal axis.

5. EFFECT OF STOCHASTIC LEAD-TIMES

We now restrict attention to model NN in Table 1, for simple exponential smoothing, but allow the lead-time, *T*, to be stochastic with mean E(*T*) = λ. In the interests of space, we omit details of the derivations, but simply report the results. The mean lead-time demand [LTD] for both additive and multiplicative models, given the level at time *n*, is:

$$E[Y_n] = E_T[E[Y_n \mid T] = \lambda \ell_n .$$

The variance of LTD for the additive scheme reduces to:

$$V[Y_n] = \ell_n^2 V(T) + \sigma^2[\lambda + (\alpha + 0.5\alpha^2)\lambda_{[2]} + \frac{\alpha^2}{3}\lambda_{[3]}]$$

where $\lambda_{[j]} = E[T(t-1)...(T-j+1)],\ j = 2,3,$ known as the factorial moments of the distribution.

Example 2: When the lead-time is fixed, $\lambda_{[j]} = \lambda(\lambda-1)...(\lambda-j+1).$ When the lead-time is Poisson with mean λ, $\lambda_{[j]} = \lambda^j.$

For the multiplicative scheme, the variance of LTD reduces to:

$$V[Y_n] = \{\ell_n^2/(\alpha\sigma)^2\}[\{1+\sigma^2 + 2(1+\alpha\sigma^2)/(\alpha\sigma)^2\}\{E(B^T)-1\} - (\alpha\sigma)^2\lambda^2 - 2\lambda(1+\alpha\sigma^2)]$$

where $B = 1 + (\sigma\alpha)^2$, and $E(B^T) = B^\lambda$ for T fixed, $E(B^T) = \exp[\lambda(\alpha\sigma)^2]$ for the Poisson.

Using the same parameter configurations as before, we compute the ratios of the standard deviations for the two models, as shown in Table 5.

Table 5: Comparison of additive and multiplicative models, with Poisson Lead Times

LAMBDA	LEVEL	ALPHA	SIGMA	KAPPA	MEAN	SD(A)	SD(M)	SD ratio
5	25	0.1	5	0.2	125	57.7	57.7	1.00
5	25	0.1	15	0.6	125	70.2	70.3	1.00
5	25	0.5	5	0.2	125	62.5	62.6	1.00
5	25	0.5	15	0.6	125	100.5	106.2	1.06
5	100	0.1	5	0.05	500	224.1	224.1	1.00
5	100	0.1	15	0.15	500	227.6	227.6	1.00
5	100	0.5	5	0.05	500	225.3	225.3	1.00
5	100	0.5	15	0.15	500	238.7	238.9	1.00
20	25	0.1	5	0.2	500	121.3	121.3	1.00
20	25	0.1	15	0.6	500	180.1	181.5	1.01
20	25	0.5	5	0.2	500	189.5	193.1	1.02
20	25	0.5	15	0.6	500	472.5	624.1	1.32
20	100	0.1	5	0.05	2000	449.7	449.7	1.00
20	100	0.1	15	0.15	2000	469.0	469.0	1.00
20	100	0.5	5	0.05	2000	472.7	472.8	1.00
20	100	0.5	15	0.15	2000	640.9	646.1	1.01

MEAN: mean lead-time demand
SD(A), SD(M): standard deviations for additive and multiplicative schemes respectively
SD ratio = SD(M)/SD(A)

The conclusions are unchanged: the ratio increases substantially only when the lead-time is long, the coefficient of variation for DPUT is high, and the correlation between demands for successive periods is high [high alpha]. However, comparison of Tables 4 and 5 shows that the variance of lead-time demand generally increases substantially in the presence of uncertain lead-times, as we would expect.

We now assume the onset of a seasonal increase, as in section 4.1. The impact of the seasonal increases, formulated as in section 4.1, is shown in Table 6 for fixed lead-times and for two variants of Poisson lead-times. For a fixed lead-time, the standard deviation is always increased by the factor (1+c). For the Poisson schemes, the increase lies in the range [1, 1+c]. The service levels corresponding to each case, for different values of c, are given in the table, showing the expected drop in performance.

Table 6: Comparison of service levels [SL] for given shifts in demand per unit time

when the multiplicative scheme is correct [target level =0.99].

	Lead-time					
Seasonal	Fixed		Poisson (1)		Poisson (2)	
Factor, c	SD ratio	SL	SD ratio	SL	SD ratio	SL
0	1.0	0.99	1.00	0.99	1.00	0.99
0.5	1.5	0.94	1.18	0.98	1.46	0.94
1	2.0	0.88	1.39	0.95	1.99	0.88
2	3.0	0.78	1.86	0.89	2.88	0.79

Poisson (1): lead-time = 5, level = 25, alpha = 0.5, SD = 15

Poisson (2): lead-time =20, level = 100, alpha = 0.1, SD = 15

6. CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

We have derived formulae for the mean and variance of lead-time demand for many common forms of exponential smoothing. For the general cases, we have assumed the lead-time to be fixed, as is increasingly common in managed supply chain systems. However, in the last part of the paper we have examined the impact of stochastic lead-times for the special case corresponding to simple exponential smoothing. By using the single source of error state space model, we have unified the derivation of the formulae. In the homoscedastic cases, many of the formulae obtained in this paper agree with those found in earlier work (Johnston and Harrison, 1986; Yar and Chatfield, 1990; Snyder, Koehler and Ord, 1999). In addition, for the Winters' additive seasonal method, the recursive variance formula in Yar and Chatfield (1990) has been replaced by a closed- form counterpart. Furthermore, we have obtained, for the first time, formulae for the variance of lead-time demand for the damped trend cases. The results for the heteroscedastic cases are also new.

It has been argued in the paper that the random error component of a demand series can depend on trend and seasonal effects. Thus, a major part of our contribution has been the provision of lead-time demand variance formulae for heteroscedastic extensions to exponential smoothing. Such formulae admit the possibility of smarter approaches to safety stock determination. It is now possible to implement schemes that tailor levels of safety stock to changes in trend or changes in season.

The numerical results in the paper indicate the following conclusions, some of them familiar:

- Lead-time uncertainty can lead to considerable increases in safety stocks, making careful management of supplier delivery schedules a valuable strategy.
- The failure to recognize that the variability in demand may be proportional to the mean level (rather than constant) can lead to service levels much lower than desired during peak periods (and excess inventory during periods of low demand).
- Incorporating known seasonal and trend patterns into safety stock planning leads to improved inventory management.

The principal direction where further research would be useful lies in the impact of estimation error upon safety stock planning decisions. In common with nearly all of the literature, we have not allowed for the uncertainty in the estimation of model parameters from short series. The combined perils of estimation error and model misspecification have been clearly detailed in Chatfield (1993) for prediction intervals, and they apply equally to the current problem.

References

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis: Forecasting and Control (third edition), Prentice-Hall, Englewood Cliffs.

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

Gardner, E.S. Jr. 1985. Exponential Smoothing: The State of the Art. Journal of Forecasting 4, 1-28.

Gardner, E.S. and McKenzie, E., 1985. Forecasting Trends in Time Series. Management Science 31, 1237-1246.

Harvey, A.C., Snyder, R.D., 1990. Structural Time Series Models in Inventory Control. International Journal of Forecasting 6, 187-198.

Heath, D.C. and Jackson, P.L., 1994. Modeling the Evolution of Demand Forecasts with Application to Safety Stock Analysis in Production / Distribution Systems. Institute of Industrial Engineers, Transactions 26, 17-30.

Holt, C.E. 1957. Forecasting Trends and Seasonal by Exponentially Weighted Averages. ONR Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh, USA.

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., Koehler, A.B., Ord, J.K., Snyder, R.D., 2001. Prediction Intervals for Exponential Smoothing State Space Models. Working Paper 11/2001, Department of Econometrics and Business Statistics, Monash University.

Johnston, F.R., Harrison, P.J. 1986. The variance of lead-time demand. Journal of the Operational Research Society 37, 303-308.

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.

Lovejoy, W.S., 1990. Myopic Policies for some Inventory Models with Uncertain Demand Distributions. Management Science 36, 724-738.

Miller, B. 1986. Scarf's State Reduction Method, Flexibility, and a Dependent Demand Inventory Model. Operations Research 34, 83-90.

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.

Snyder, R.D. 1985. Recursive Estimation of Dynamic Linear Statistical Models. Journal of the Royal Statistical Society, B. 47, 272-276.

Snyder, R.D., Koehler, A.B., Ord, J.K., 1999. Lead-time Demand for Simple Exponential Smoothing. Journal of the Operational Research Society 50, 1079-1082.

Winters, P.R. 1960. Forecasting Sales by Exponentially Weighted Moving Averages. Management Science 6, 324-342.

Yar, M., Chatfield, C., 1990. Prediction intervals for the Holt-Winters Forecasting Procedure. International Journal of Forecasting 6, 127-137.

APPENDIX

General results governing the formulae in Table 3 are derived in this Appendix. To get the formulae governing Cases 1-4, back solve the transition equation (2) from period $n+j$ to period n, to give

$$x_{n+j} = F^j x_n + \sum_{i=1}^{j} F^{j-i} g e_{n+i} \qquad \text{(A1)}$$

Lag (A1) by one period, pre-multiply the result by h', and use the definitions (7) and (9) to get

$$m_{n+j} = \mu_{n+j} + \sum_{i=1}^{j-1} c_{j-i} e_{n+i} \, . \qquad \text{(A2)}$$

Recall that e_t is given by (3) so that $E(e_{n+i}^2 |) = \sigma^2 E(m_{n+i}^2)$. Then we may square (A2) and take expectations to give the recurrence relationship (12) for the heteroscedastic factors. Substitute (A2) into (1) to give $y_{n+j} = \mu_{n+j} + \sum_{i=1}^{j-1} c_{j-i} e_{n+i} + e_{n+j}$. Substitute this into (1) to give $Y_n(j) = \sum_{j=1}^{\lambda} \left(\mu_{n+j} + \sum_{i=1}^{j-1} c_{j-i} e_{n+i} + e_{n+j} \right)$. Rearrange terms to yield the required result (6) where the C_j are defined by (8). Note that the derivation of the C_j is expedited using the following equations: $C_\lambda = 1$ and $C_j = C_{j+1} + c_{\lambda-j}$ for $j = \lambda - 1, \ldots, 1$.

Cases 5 and 6 are composite models. Each transition equation (5), for a composite model, has the same structure as (2). Thus,

$$x_{k,n+j} = F_k^j x_{k,n} + \sum_{i=1}^{j} F_k^{j-i} g_k e_{n+i} \, . \qquad \text{(A3)}$$

Lag (11) by one period and pre-multiply the result by h_k' to give

$$m_{k,n+j} = \mu_{k,n+j} + \sum_{i=1}^{i-1} c_{k,j-i} e_{n+i} \tag{A4}$$

where

$$\mu_{k,n+j} = h_k' F_k^{j-1} x_{k,n} \tag{A5}$$

and

$$c_{k,i} = h_k' F_k^{i-1} g_k \,. \tag{A6}$$

Substitute (A4) into $m_{n+j} = m_{1,n+j} + m_{2,n+j}$ to yield the earlier equation (A2) where

$$\mu_{n+j} = \mu_{1,n+j} + \mu_{2,n+j} \tag{A7}$$

and

$$c_i = c_{1,i} + c_{2,i} \,. \tag{A8}$$

Thus, the formula $C_i = C_{1,i} + C_{2,i} - 1$ may be used to derive the results for Case 5 and Case 6 from their constituent basic cases. In the heteroscedastic cases, the appropriate factors are still derived with the relationship (12).