ISSN 1440-771X
ISBN 0 7326 1065 6

MONASH UNIVERSITY

AUSTRALIA

Forecasting for Inventory Control with Exponential Smoothing

Ralph D. Snyder, Anne Koehler and Keith Ord

Working Paper 10/99
August 1999

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Forecasting for Inventory Control with Exponential Smoothing

Associate Professor Ralph D. Snyder
Department of Econometrics and Business Statistics
Monash University
Clayton, Victoria 3168 Australia

Professor Anne Koehler
Department of Decision Sciences
Miami University
Oxford, OH45056, USA

Professor Keith Ord
Department of Management Science
The Pennsylvania State University
University Park, PA 16802, USA
(After July 1, 1999: McDonough School of Business, Georgetown University,
Washington, DC 20057 USA)

Abstract:
Exponential smoothing, often used for sales forecasting in inventory control, has always been rationalized in terms of statistical models that possess errors with constant variances. It is shown in this paper that exponential smoothing remains the appropriate approach under more general conditions where the variances are allowed to grow and contract with corresponding movements in the underlying level. The implications for estimation and prediction are explored. In particular the problem of finding the prediction distribution of aggregate lead-time demand for use in inventory control calculations is considered. It is found that unless a drift term is added to simple exponential smoothing, the prediction distribution is largely unaffected by the variance assumption. A method for establishing order-up-to levels and reorder levels directly from the simulated prediction distributions is also proposed.

Key Words: Inventory control, demand forecasting, exponential smoothing, bootstrap methods.

1. INTRODUCTION

The conceptualisation of simple exponential smoothing (Brown, 1959) was an important development for demand forecasting in inventory control. Yet implementations of this method have often been surrounded by practices, summarised in Gardner(1985), which possess questionable theoretical roots and which, at best, have enjoyed a mixed success. This paper is written on the assumption that good practice emerges from sound theory and that such a strategy must be built on the statistical models underlying the technique. Our contribution is to suggest that the theory of simple exponential smoothing can be extended to a broader class of models where error variances, instead of remaining constant, can change over time. The implications of this heteroscedasticity for estimation are explored in section 3. Its consequences for prediction, particularly in relation to aggregate lead time demand, a quantity of particular interest in inventory control, are outlined in section 4.

In a recent paper (Ord, Koehler and Snyder, 1997) the methods of exponential smoothing were shown to apply under much more general conditions than those traditionally envisaged in the literature. In this paper we take a more detailed look at some of the important versions of exponential smoothing and explore the consequences with reference to the forecasting requirements in inventory control.

Many expositions of exponential smoothing are related back to associated ARIMA models (Box and Jenkins,). It is implicitly assumed in such expositions that the processes under consideration extend back into the infinite past. Most items carried in a typical inventory system typically possess a finite life cycle so that this semi-infinite time assumption is unrealistic. It is assumed in this paper that items are introduced into an inventory at the start of a period of time that is designated period 1. It is further assumed that n periods have elapsed since the introduction and that the problem is to forecast demand over a lead-time h extending from period $n+1$ to $n+h$. Demand for an item in typical period t is represented by

y_t .

2. SIMPLE EXPONENTIAL SMOOTHING

2.1 LOCAL LEVEL MODELS

Demand over time in basic inventory theory is usually represented by normally and independently distributed random variables with a common mean m and a common standard deviation s . The series is often written as

$$y_t = m + e_t , \tag{2.1}$$

the e_t being $NID\left(0, s^2\right)$ random variables. The errors e_t represent unanticipated demand. In this model the impact of each error is restricted to the period in which it occurs. Each error only has a *transient* effect.

In practice unanticipated demand may spill over into later periods. New customers may cause demand to increase in the long term. New competitors entering a market may permanently reduce market shares. Assuming that a proportion a of unanticipated demand has a permanent effect from causes like these, the model (2.1) may be modified to give

$$y_t = m + a\sum_{j=1}^{t-1} e_{t-j} + e_t . \tag{2.2}$$

Muth (1960) introduced this model, albeit with a semi-infinite past and with $m = 0$, and showed that it underpins simple exponential smoothing. Differencing yields, for $t \geq 2$, the process $\Delta y_t = -\theta e_{t-1} + e_t$ where $\theta = 1 - a$. Working in a semi-infinite time context, Box and Jerkins (1976) also demonstrated that this model underlies simple exponential smoothing.

A 'local level' may be defined as $m_t = m + a\sum_{j=0}^{t-1} e_{t-j}$. The model (2.2) may then be rewritten in terms of a measurement equation $y_t = m_{t-1} + e_t$ and a transition equation $m_t = m_{t-1} + ae_t$. Both the measurement and transition equations define what may be referred to as a local level model (LLM). It is a special case of the linear state space framework in Snyder (1985). The parameter a corresponds to the familiar smoothing constant. An advantage of this representation over its ARIMA counterpart is that the link with the error correction form of exponential smoothing is more transparent.

A generalisation of the local level model that accommodates level dependent variability consists of the measurement equation

$$y_t = m_{t-1} + m_{t-1}^q e_t \tag{2.3}$$

together with the transition equation

$$m_t = m_{t-1} + am_{t-1}^q e_t \tag{2.4}$$

where the parameter *q* determines the degree of heteroscedasticity. It will be designated LLM(*q*). Our primary focus will be on the special cases LLM(0) and LLM(1). LLM(0) corresponds to the original local level model with *additive* errors. LLM(1) represents series with level dependent variability based on *relative* errors.

The behaviour of the local level in LLM(*q*) is governed by

$$m_t = \theta m_{t-1} + ay_t\ . \tag{2.5}$$

This relationship is obtained by eliminating the error term in the LLM(*q*) equations (2.3) and (2.4). Interestingly, it does not depend *q*. The behaviour of the level is independent of the form of heteroscedasticity. The closed form solution for the local level is

$$m_t = \theta^t m + a\sum_{j=0}^{t-1}\theta^j y_{t-j} \ . \qquad (2.6)$$

The local level summarises the past behaviour of the demand process. The weights $\alpha\theta^j$ determine the impact of past time series values. When $|\theta| < 1$ these weights decline with increases in the age index *j*. This discounting of past observations is warranted when markets are subject to structural change.

The inequality $|\theta| < 1$ for LLM(0) corresponds to the invertibility condition for the ARIMA(0,1,1) process (Box and Jenkins, 1976). It is also equivalent to $0 < a < 2$. Thus LLM(0) accomodates structural change for values of the smoothing parameter in excess of 1, a conclusion that is incompatible with the traditional argument above. The importance of this can be gauged by focussing on the first-order autocorrelation coefficient of the first-differences. It can be established that the first-order autocorrelation in the differenced demand series is given by $corr(\Delta y_t \Delta y_{t-1}) = -\theta$. Demand series with positively autocorrelated first differences can only be modelled, within the local level framework, if *a* is allowed to take values above one.

2.2 SMOOTHING OF TIME SERIES

The unknown smoothing parameter a and the seed level m may be assigned trial values. At the start of typical period t the observed values of the series $y_1, y_2, \ldots y_{t-1}$ from earlier periods are also fixed, known quantities. The information set may be designated by $I_{t-1} = \{y_1, y_2, \ldots y_{t-1}, a, m\}$. Let $\overline{m}_{t-1} = (m_{t-1}|I_{t-1})$. According to (2.5) successive local levels can be computed recursively with

$$\overline{m}_t = \overline{m}_{t-1} + a\left(y_t - \overline{m}_{t-1}\right) \tag{2.7}$$

where $\overline{m}_0 = m$. These *conditional* local levels are *fixed* rather than *random* quantities. The equation (2.7) corresponds to the simple exponential smoothing updating relationship (Brown, 1959). The traditional view has always been that simple exponential smoothing can only be rationalised in terms of an ARIMA(0,1,1) process (Box & Jenkins, 1976). A key finding in this paper is that exponential smoothing is also compatible with models where the variation in the series is dependent on the underlying level.

2.3 MAXIMUM LIKELIHOOD ESTIMATION

A wide variety of methods (Gardner, 1985) have been suggested in the context of simple exponential smoothing for estimating the seed level m, the smoothing parameter a and the standard deviation *s*. Holt (1957) recommends the use of the sum of squared one-step ahead prediction errors as a criterion for selecting the smoothing parameter. It also makes sense to apply the same criterion when choosing the seed level. Yet we have seen that simple exponential smoothing is also a legitimate method when demand data is generated by LLM(1). It might be speculated that the same tactic works with the sum of squared *relative* errors. Given that absolute and relative errors are inherently different quantities, one could not compare both types of sum of squared errors criterion to make the choice between LLM(0) and LLM(1). This is a serious drawback with the sum of squared errors criterion.

Likelihood functions for different models, in contrast, are comparable quantities. In forming such likelihood functions we choose to treat the seed level *m* as a parameter. Under the semi-infinite life assumption adopted quite widely in expositions of the state space approach to time series analysis, such a strategy would not be legitimate. Then *m* would be represented as an infinite sum of past errors and therefore would be a random variable with an infinite variance. Because *m* is an unobserved quantity, it would be necessary to integrate it out of the likelihood. In other words it would be necessary to find the marginal likelihood function (Kalbfleish & Sprott). Under our finite life assumption for the inventories this difficulty is avoided because *m* is a fixed quantity. Under our assumption the likelihood can be shown to be

$$\ell(m,a,s|y_1,y_2,\dots y_n)=\left(2\pi s^2\right)^{-n/2}\left(\prod_{t=1}^{n}|\overline{m}_{t-1}|\right)^{-q}\exp\left(-\sum_{i=1}^{n}\overline{e}_t^2\Big/2s^2\right), \quad (2.8)$$

the one-step ahead prediction errors $\overline{e}_t=\left(y_t-\overline{m}_{t-1}\right)\big/\overline{m}_{t-1}^q$ being obtained from the application of simple exponential smoothing relationship (2.7). The maximum likelihood of the variance estimate of the variance is given by the familiar formula $\overline{s}^2=\sum_{i=1}^{n}\overline{e}_t^2\Big/n$.

Substitution of this into (2.8) yields

$\ell(m,a|y_1,y_2,\dots y_n)=\left(2\pi\overline{s}^2\right)^{-n/2}\left(\prod_{t=1}^{n}|\overline{m}_{t-1}|\right)^{-q}\exp(-n/2)$. Thus the maximum likelihood estimates of *a* and *m* may be obtained by minimising the quantity

$$\omega=\overline{s}\left(\sqrt[n]{\prod_{t=1}^{n}|\overline{m}_{t-1}|}\right)^{q}. \quad (2.9)$$

For LLM(0) ω corresponds to the standard error $\overline{s}$. In other words it is appropriate to minimise the standard error or its equivalent, the conventional sum of squared errors $S=\sum_{t=1}^{n}\overline{e}_t^2$. This justifies the extension of Holt's strategy for the selection of the smoothing parameter to the problem of choosing the seed level. The criterion (2.9) is $\omega=\overline{s}\left(\sqrt[n]{\prod_{t=1}^{n}|\overline{m}_{t-1}|}\right)$ for LLM(1). The second term in this expression is the geometric mean of the local levels. Its effective purpose is to convert $\overline{s}$, now measured in relative terms, into a quantity with the same units of measurement as those for the original demand series. By focussing on ω rather than *s* (or the sum of squared errors), comparisons can be made between different models using within sample fit. ω will be referred to as the *generalised* standard error. Those values of *m* and *a* which minimise the generalised standard error will be represented by $\hat{m}$ and $\hat{a}$. The corresponding value of $\bar{s}$ will be designated by $\hat{s}$. The statistics $\hat{m}, \hat{a}$ and $\hat{s}$ are maximum likelihood estimates.

It now has been established that simple exponential smoothing can be rationalised in terms of a broader set of models than has hitherto been appreciated. The form of the updating relationship remains the same for all versions of the local level model. Only the fitting criteria change to reflect different possible assumptions about the behaviour of the variance. Practitioners are therefore faced with the prospect of implementing more elaborate estimation procedures based on criteria other than the traditional sum of squared prediction errors. The question is whether such change is really warranted?

To gain insight into this question it is worthwhile considering the stationary model $y_t = m + m^q e_t$, a special case of LLM(q) obtained when $a = 0$. In this case $\omega = \sqrt{\sum_{t=1}^{n} (y_t - m)^2 / nm^p} \left(\sqrt[n]{\prod_{t=1}^{n} m} \right)^p = \sqrt{\sum_{t=1}^{n} (y_t - m)^2 / n}$. The generalised standard error is independent of the degree of heteroscedasticity *q*. The sample average $\hat{m} = \sum_{t=1}^{n} y_t / n$ minimises the generalised standard error for any value of *q*. In the stationary case LLM(0) and LLM(1) have the same maximum likelihood estimate.

This neat simplification of ω disappears when $a \neq 0$. To gauge the impact of this, a small simulation study was undertaken comparing the estimates obtained from LLM(0) and LLM(1). At each of 1000 replications of the simulation, a time series was generated from LLM(1) with $m = 100$, the sample size *n*, the smoothing parameter *a* and the standard deviation *s* being selected randomly from the values shown in Table 1. Both LLM(0) and

LLM(1) were fitted to the simulated data and their optimal generalised standard errors ω_0^* and ω_1^* compared. Apportioning any ties equally between both approaches, it was found that LLM(1) was correctly selected 63 percent of the time. In 28 percent of cases ω_0^* proved to be more than one percent away from ω_1^*.

Place Table 1 about here

The LLM(1) generalised standard error ω_1^0, when the optimal LLM(0) estimates are used as approximations for the LLM(1) estimates, was also calculated. The ratio $\lambda = \omega_1^0 / \omega_1^*$ can never be less than 1 because the LLM(0) estimates are not optimal for the LLM(1) model. Nevertheless λ averaged 1.0004 and had a standard deviation of only 0.0013. The optimal estimates for both models were usually remarkably close.

LLM(1) is more 'nonlinear' than LLM(0) and therefore potentially more difficult to estimate. The above simulation suggests the following estimation strategy for LLM(1):

1. Find the maximum likelihood estimates $\hat{m}, \hat{a}$ and the one-step predictions $\hat{y}_t$ for the simpler LLM(0).

2. Use $\hat{m}, \hat{a}$ as approximations for the corresponding quantities in LLM(1) and estimate the standard deviation of the relative errors with

$$\hat{s} = \sqrt{n^{-1} \sum_{t=1}^{n} \left(y_t - \hat{y}_t \right)^2 / \hat{y}_t^2} \ . \qquad (2.10)$$

It might be argued, given the above simulation results, that there is little point in using LLM(1) and that this proposed estimation procedure is largely redundant from a practical point of view. In the simulation, however, λ_1 had an average of 1.0166 and a standard deviation of 0.0392. The size of the standard deviation indicates that in some contexts the gains from using LLM(1) may be warranted. From a practical point of view, however, it is usually the predictive capacity of a model that counts. It is this issue that we now explore.

2.4 LEAD TIME DEMAND DISTRIBUTION

The prediction distribution of the typical series value y_{n+j} beyond period n is conditioned on the sample $y_1, y_2, \ldots, y_n$. For convenience it is assumed initially that the seed level m, the smoothing parameter a, and the standard deviation *s* are known exactly. The problem then reduces to finding the distribution of $y_{n+j} | I_n$.

For LLM(0), back-substitution of the recurrence relationship (2.4) yields $\overline{m}_{n+j} = \overline{m}_n + a \sum_{t=n+1}^{n+j} e_t$ where $\overline{m}_{n+j} = \left(\overline{m}_{n+j} | I_n\right)$. The future *conditional* local level $\overline{m}_{n+j}$ is a *random* rather than a *constant* quantity. It follows from this future local level equation that

$$y_{n+j} = \overline{m}_n + a \sum_{t=n+1}^{n+j-1} e_t + e_{n+j} . \tag{2.11}$$

Thus $E\left(y_{n+j} | I_n\right) = \overline{m}_n$ and $Var\left(y_{n+j} | I_n\right) = \left((j-1)a^2 + 1\right)s^2$.

In inventory control applications the primary interest is in *total* demand over a lead-time h. Aggregation of (2.11) gives $\sum_{t=n+1}^{n+h} y_t = h\overline{m}_n + \sum_{j=0}^{h-1} (1 + ja) e_{n+h-j}$. Thus the mean lead-time demand is given by the usual formula $E\left(\sum_{t=n+1}^{n+h} y_t | I_t\right) = hm_n$. The variance, however, has the more complex formula $Var\left(\sum_{t=n+1}^{n+h} y_t | I_n\right) = f(h,a)^2 s^2$ where $f(h,a) = \sqrt{h\left(1 + a(h-1)\left(1 + (2h-1)a/6\right)\right)}$. In other words, conditional total lead-time demand is normally distributed with mean hm_n and standard deviation $f(h,a)s$. In practice the $\sqrt{h}$ is often used instead of $f(h,a)$. This leads to a serious under-estimation of the prediction standard deviation and this has serious consequences for safety stock determination and customer service, a matter that is more fully explored in Snyder, Kohler and Ord (1997).

For the LLM(1)

$$y_{n+j} = \overline{m}_n \prod_{t=n+1}^{n+j-1} (1+ae_t)(1+e_{n+j}) . \qquad (2.12)$$

Again the conditional mean, which may be used as a forecast, is $E(y_{n+j}|I_n) = \overline{m}_n$. Being expressed in relative terms the errors are fairly small. Products of the errors are negligible so that the random variable y_{n+j} in (2.12) can be approximated by the quantity $\tilde{y}_{n+j}$ defined by $\tilde{y}_{n+j} = \overline{m}_n \left(1 + a \sum_{t=n+1}^{n+j-1} e_t + e_{n+j}\right)$. The conditional variance may be approximated by $Var(\tilde{y}_{n+j}|I_n) = (1+(j-1)a^2)\overline{m}_n^2 s^2$. For lead-time demand it can be established that the mean is again given by $E\left(\sum_{t=n+1}^{n+h} y_t | I_t\right) = h\overline{m}_n$ and that the conditional variance can be approximated by the slightly different formula $Var\left(\sum_{t=n+1}^{n+h} \tilde{y}_t | I_n\right) = f(h,a)^2 \overline{m}_n^2 s^2$. It is tempting to approximate the lead time demand distribution by a normal distribution with mean $h\overline{m}_n$ and standard deviation $f(a,h)\overline{m}_n s$. Simulation studies indicated that provided it is assumed that $\overline{m}_n$ is known with certainty then this approximation works well.

From a practical point of view it is probably simpler to bypass normal approximations based on the above moments formulae and simulate lead-time demand distributions directly from the relationships (2.11) and (2.12). The quantities m, a and s are usually unknown. A parametric bootstrap based on the approximations $m = \hat{m}$, $a = \hat{a}$ and $s = \hat{s}$ can be used. Exponential smoothing, seeded with the maximum likelihood estimates, is used to calculate

$\hat{m}_n$, the corresponding value of $\overline{m}_n$. These quantities are used together with the formulae the LLM(q) formulae to generate bootstrap samples of lead-time demand. A simulation for comparing the results of the bootstrap method for LLM(0) and LLM(1) was undertaken. First, a sample was simulated from the LLM(1) with $m = 100$, $s = 0.05$ and $n = 30$. Maximum likelihood estimates were obtained for both models. These were then used obtain both bootstrap samples of 1000 lead-time demands, the lead-time being $h = 10$. Figure 1 shows the quantile-quantile plot of the samples. The plot is quite close to the 45^0 line reflecting the similarity of the LLM(0) and LLM(1) lead-time demand distributions. This result is typical of those obtained when the simulation conditions were varied.

Place Figure 1 about here

The parametric bootstrap approach ignores the effect of estimation error. An appropriate adaptation of the simulation method described in Ord, Koehler and Snyder (1997) would account for this source of error. It is anticipated that greater differences would then emerge between the two models.

3. SIMPLE EXPONENTIAL SMOOTHING WITH DRIFT

3.1 *LOCAL LEVEL MODELS WITH DRIFT*

Intuitively, it makes sense that heteroscedasticity related to the magnitude of the fluctuations in the mean, may be modest when the mean is locally constant. It is likely to have more of an impact if there is a tendency for the series to increase or decrease over time. We therefore introduce a growth rate *b* into (2.2) to give

$$y_t = m + bt + \alpha \sum_{j=1}^{t-1} e_j + e_t . \qquad (3.1)$$

This can be written as $y_t = m + bt + u_t$ where $u_t = u_{t-1} - \theta e_{t-1} + e_t$. Differencing (3.1) yields $\Delta y_t = b - \theta e_{t-1} + e_t$ so that the associated time series is 'difference stationary'. Contrast this with the more common case where the trend is accompanied by first-order autocorrelated disturbances governed by $u_t = \phi u_{t-1} + e_t$, the associated series now being 'trend stationary' provided that parameter ϕ satisfies the condition $|\phi| < 1$.

The local level may be defined as $m_t = m + bt + \alpha \sum_{j=1}^{t} e_t$. Then (3.1) may be written as

$y_t = m_{t-1} + b + e_t$ where $m_t = m_{t-1} + b + \alpha e_t$. This local level model with drift is the special case of the model underlying Holts trend corrected exponential smoothing where the growth rate is restricted to a constant value. The generalisation to include heteroscedastic variation is

$$y_t = (m_{t-1} + b) + (m_{t-1} + b)^q e_t \qquad (3.2)$$

where

$$m_t = (m_{t-1} + b) + \alpha (m_{t-1} + b)^q e_t . \qquad (3.3)$$

It will be designated LLDM(q). Again $q = 1$ corresponds to the relative error case.

The local level, for any LLDM(q) can be written as $m_t = \theta^t m + \sum_{j=1}^{t} \theta^j b + \alpha \sum_{j=0}^{t-1} \theta^j y_{t-j}$.

The local level is still a discounted linear function of series values when the invertibility condition $|\theta| < 1$ holds. In what follows, however, we continue to use the common restriction $0 \le \alpha \le 1$.

3.2 SMOOTHING AND ESTIMATION

Smoothing and estimation is quite similar to the case where there is no drift. For given values of m, b and α the augmented smoothing relationship $\overline{m}_t = \overline{m}_{t-1} + b + \alpha\left(y_t - \overline{m}_{t-1} - b\right)$ may be applied where $\overline{m}_0 = m$. The errors may be calculated with

$\overline{e}_t = \left(\overline{y}_t - \overline{m}_{t-1} - b\right) / \left(\overline{m}_{t-1} + b\right)^q$. Using the principle of maximum likelihood, it can be argued that m, b and α should be chosen to minimise the generalised standard error

$$\omega = \overline{s}\left(\sqrt[n]{\prod_{t=1}^{n} \left|\overline{m}_{t-1} + b\right|}\right)^q .$$

A simulation study similar to the first study described in section 2.3 was again conducted, but now with a drift of $b = 0.5$. Interestingly the simulated ratio λ now had a moderately larger average of 1.0015 and standard deviation of 0.0040. It seems that even with a drift term, the LLDM(0) estimates are close enough for most practical purposes to use as approximations for their LLDM(1) counterparts. Interestingly, the optimal generalised standard error ω_1^* was now greater than ω_0^* about 80 percent of the 1000 replications. Furthermore, the gap between ω_0^* and ω_1^* now exceeded one percent in 67 percent of the replications. The differences between LLDM(0) and LLDM(1) when there is drift can be quite marked.

3.3 *Lead Time Demand Distribution*

It is possible to obtain formulae for the mean and standard deviation of the lead-time demand distributions. It is simplest, however, to undertake a parametric bootstrap of the lead-time distribution. Figure 2 shows the quantile-quantile plot obtained when the simulation was conducted under essentially the same conditions as those depicted in section 2.4. The only difference is that now a drift of $b = 0.5$ is assumed. The quantile-quantile plot of simulated lead-time demands is now steeper than the 45^0 line. This indicates that the lead-time demand distribution for LLDM(1) is more spread than that for LLDM(0). It suggests that larger safety stocks may be required if level dependent errors are present and that a pronounced drift is observed in demand.

Place Figure 2 about here

---

4. Trend Corrected Exponential Smoothing

Another possible model is $y_t = m + bt + \alpha_1 \sum_{j=1}^{t-1} e_j + \alpha_2 \sum_{i=1}^{t-1} \sum_{j=1}^{i} e_j + e_t$. It may be written as

$y_t = m + bt + u_t$ where $\Delta^2 u_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$ for $t > 2$, the parameters being related by the equations $\theta_1 = 2 - \alpha_1 - \alpha_2$ and $\theta_2 = \alpha_1 - 1$. It is a trend line with a particular form of autocorrelated disturbances. A local level may be defined as

$m_t = m + bt + \alpha_1 \sum_{j=1}^{t} e_j + \alpha_2 \sum_{i=1}^{t} \sum_{j=1}^{i} e_j$. Then $y_t = m_{t-1} + b + \alpha_2 \sum_{j=1}^{t-1} e_j + e_t$. If, in addition, a

local growth rate is defined as $b_t = b + \alpha_2 \sum_{j=1}^{t} e_j$, then the model may be written is state space

form as $y_t = m_{t-1} + b_{t-1} + e_t$ where $m_t = m_{t-1} + b_{t-1} + \alpha_1 e_t$ and $b_t = b_{t-1} + \alpha_2 e_t$. Unlike LLDM(0), the growth rate is now allowed to change over time. This is the so-called local trend model.

A generalisation to accommodate heteroscedastic variation is

$$y_t = m_{t-1} + b_{t-1} + \left(m_{t-1} + b_{t-1}\right)^q e_t \qquad (4.1)$$

$$m_t = m_{t-1} + b_{t-1} + \alpha_1 \left(m_{t-1} + b_{t-1}\right)^q e_t \qquad (4.2)$$

$$b_t = b_{t-1} + \alpha_2 \left(m_{t-1} + b_{t-1}\right)^q e_t \tag{4.3}$$

It will be designated LTM(q). Again $q = 1$ corresponds to the relative error case.

Assuming that *m*, *b*, α_1 and α_2 have been assigned trial values, the information available at the end of typical period *t* is $I_t = \left\{y_1, \ldots, y_t, m, b, \alpha_1, \alpha_2\right\}$. Let $\overline{m}_0 = m$ and $\overline{b}_0 = b$. Furthermore let $\overline{m}_t = \left(m_t | I_t\right)$ and $\overline{b}_t = \left(b_t | I_t\right)$ for $t \geq 1$. These conditional quantities must be consistent with the equations for LTM(q). They may be computed recursively with the relationships

$$\overline{m}_t = \overline{m}_{t-1} + \overline{b}_{t-1} + \alpha_1 \left(y_t - \overline{m}_{t-1} - \overline{b}_{t-1}\right) \tag{4.4}$$

$$\overline{b}_t = \overline{b}_{t-1} + \alpha_2 \left(y_t - \overline{m}_{t-1} - \overline{b}_{t-1}\right), \tag{4.5}$$

These recurrence relationships are obtained by eliminating the e_t from the LTM(*q*) equations. They correspond to the error correction form of Holts trend corrected exponential smoothing (Gardner, 198*). Thus this traditional method is applicable under much broader conditions than those traditionally stated in the literature. It applies when the variation depends on the underlying level.

Maximum likelihood estimates of *m*, *b*, α_1 and α_2 can be obtained by minimising the generalised standard error $\omega = \overline{s} \left(\sqrt[n]{\prod_{t=1}^{n} |\overline{m}_{t-1} + \overline{b}_{t-1}|} \right)^q$. The standard deviation $\overline{s}$ is still calculated from the formula $\overline{s}^2 = \sum_{i=1}^{n} \overline{e}_t^2 \Big/ n$ but now the errors are obtained from trend corrected exponential smoothing using the formula $\overline{e}_t = \left(\overline{y}_t - \overline{m}_{t-1} - \overline{b}_{t-1}\right) \Big/ \left(\overline{m}_{t-1} + \overline{b}_{t-1}\right)^q$. LTM(*q*) reduces to LLDM(*q*) when $\alpha_2 = 0$. Both models have similar properties so the conclusions reached with the simulations for LLDM(*q*) also apply to LTM(*q*).

5. Seasonal Effects

LTM(0) can be augmented by a seasonal cycle of length *p* if required. Let c_t denote the seasonal effect associated with typical period *t*. The resulting model, when the seasonal effects are additive, is $y_t = m_{t-1} + b_{t-1} + c_{t-p} + e_t$ where $m_t = m_{t-1} + b_{t-1} + \alpha_1 e_t$, $b_t = b_{t-1} + \alpha_2 e_t$ and $c_t = c_{t-p} + \alpha_3 e_t$. It is easily seen that this model underpins the additive version of Holts seasonal exponential smoothing. Its multiplicative counterpart is $y_t = (m_{t-1} + b_{t-1}) c_{t-p} (1 + e_t)$ where $m_t = (m_{t-1} + b_{t-1})(1 + \alpha_1 e_t)$, $b_t = b_{t-1} + \alpha_2 (m_{t-1} + b_{t-1}) e_t$ and $c_t = c_{t-p}(1 + \alpha_3 e_t)$. If the errors are substituted out of these equations and the appropriate conditioning on past information is undertaken, the equations for Winters method of exponential smoothing is obtained (Winters, 1960). The details are covered in Ord, Koehler and Snyder (1997).

To reduce the number of parameters it is often better to use a Fourier representation of seasonal cycles (Brown, 196X). Winters method would not be practical if applied to say weekly demand data. One possibility is a linear local level model with drift and seasonal cycle LLDSM(0)

$$
\begin{aligned}
y_t &= m_{t-1} + b + c_{t-p} + e_t \\
m_t &= m_{t-1} + b + a e_t \\
c_t &= \sum_{j=1}^{r} \left(\alpha_j \sin(\omega_j t) + \gamma_j \cos(\omega_j t) \right)
\end{aligned}
$$

where the ω_j are the frequencies and the α_j and β_j are coefficients. Note that $r \le [(p+1)/2]$. Usually *r* is much smaller than $(p+1)/2$. Otherwise there would be no advantage in using Fourier representations. In this model the seasonal cycle is deterministic. Stochastic error representations of a seasonal cycle are possible. These imply, however, that the seasonal cycle follows a random walk, something that is difficult to believe. Another possible generalisation involves local growth rates as found in the local trend model.

A nonlinear heteroscedastic generalisation of this model is:

$$\begin{aligned} y_t &= (m_{t-1} + b)(1 + c_{t-p})(1 + e_t) \\ m_t &= (m_{t-1} + b)(1 + ae_t) \\ c_t &= \sum_{j=1}^{r} \left(\alpha_j \sin(\omega_j t) + \gamma_j \cos(\omega_j t) \right) \end{aligned} .$$

The seasonal and irregular components increase with the trend. This model will be designated LLDSM(1).

Again the generalised standard error may be used as the estimation criterion. It is

$$\omega = \sqrt{\sum_{t=1}^{n} \left(y_t - m_{t-1} - b - c_{t-p} \right)^2 \Big/ n}$$

and

$$\omega = \sqrt{\sum_{t=1}^{n} \left(\frac{y_t - (m_{t-1} + b)(1 + c_{t-p})}{(m_{t-1} + b)(1 + c_{t-p})} \right)^2 \Big/ n} \sqrt[n]{\prod_{t=1}^{n} \left| (m_{t-1} + b)(1 + c_{t-p}) \right|}$$

for LLDSM(0) and LLDSM(1) respectively.

A simulation study similar to the one described in section 3.3 was undertaken to determine the differences in the estimates for the linear and nonlinear cases. The series for the simulation were generated from the nonlinear model with $c_t = 0.5 \sin(2\pi t/52)$. This corresponds to a pronounced seasonal cycle in the time series but the size of the amplitude of the cycle is quite plausible in practice. On average the generalised standard error turned out to be about 20 percent higher for the estimates based on the wrong LLDSM(0). Using the generalised standard error as the selection criterion, the correct model was chosen 99 percent of the time. Thus, for the first time, we have detected a major difference between the homoscedastic and heteroscedastic models. The implications of this will be explored in greater depth using criteria from inventory control.

INVENTORY CONTROL

In this section we attempt to gauge the impact of differences arising from the linear and nonlinear seasonal models in the context of an inventory problem. The focus will be on an order level system with periodic reviews and the backlogging of excess demand.

The state of an order level system at any point of time is represented by the stock position, a quantity governed by the formula *StockPosition* = *Stock* – *Backlog* + *OnOrder*. The order level represents the appropriate level for the stock position following the placement of a new replenishment order. It is assumed that such orders are placed at the start of each review period.

The size of the order level determines the service given to customers. It is assumed that service is summarised by the fill-rate (customer service level), a statistic that measures the proportion of demand satisfied without delays caused by shortages. It is further assumed that managers specify a target value for the fill-rate, the problem then being to choose the order level to meet this target.

A theory for the determination of order levels using the fill-rate (customer service level) statistic appears to have been first proposed by Brown (19??). The theory involves the use of exponential smoothing in combination with what Brown refers to as a ‘partial expectation’. The approach was a major breakthrough in its day and its influence may still be found in modern inventory control software. It has, however, two weaknesses that can nowadays be circumvented with the common availability of powerful computers.

a) It employed heuristics, based on mean absolute deviations, to measure the variability of lead-time demand.

b) It relied on an approximation for the fill-rate that was a necessary convenience when calculations were done manually, but which is known to be inaccurate when review periods are short in length – see below for details.

The theory of this paper provides an opportunity to circumvent the heuristics for measuring variability while using the exact formula for the customer service level. We now outline a parametric bootstrap approach that would have been impractical until recent times. It is based on the assumption that an order placed at time *n* is delivered at time $n+h$ and that the problem is therefore to use such an order to influence the performance of the inventory system in the period $(n+h, n+h+1)$. The fill-rate in this period is defined as

$\beta = 1 - E\left(x_{n+h+1|n}\right) / E\left(y_{n+h+1|n}\right)$ where $x_{n+h+1|n}$ is the excess demand in period *n*+*h*+1 and $y_{n+h+1|n}$ is the demand in period *n*+*h*+1 given the information set I_n. Since $x_{n+h+1|n} \le y_{n+h+1|n}$ the fill-rate always lies in the interval [0, 1]. This measure of service should not be confused with the tail of the lead time demand distribution commonly used in some approaches to inventory control (Buffa, 19??).

Demand in period *n*+*h*+1 may be easily simulated from the model underlying the forecast method using parameter estimates in place of the unknown parameters. Given a particular order level *S*, the corresponding excess demand in the same period can also be calculated with

$x_{n+h+1|n} = \left(y_{(n+1:n+h+1)|n} - S\right)^{+} - \left(y_{(n+1:n+h)|n} - S\right)^{+}$. Each RHS term, being the excess of lead-time demand over total supply, is a backlog. They correspond to the closing and opening backlog in period *n*+*h*+1 given the information I_n. Being the increase in the backlog, the RHS corresponds to the excess demand in period *n*+*h*+1. It is possible to follow Brown (1959) and assume that the opening backlog is small enough to be ignored. The second term on the RHS of the formula for excess demand would then disappear. In practice, particularly when review periods are relatively short, a delivery may be insufficient to completely eliminate an existing backlog. It is better not to make this approximation.

A bootstrap involving *R* replications may be used to estimate the fill rate. Denoting the *r*th replication of excess demand and demand by $x_{n+h+1|n}^{(r)}$ and $y_{n+h+1|n}^{(r)}$ respectively, a bootstrap estimate of the fill-rate is $\hat{\beta} = 1 - \sum_{r=1}^{R} x_{n+h+1|n}^{(r)} \Big/ \sum_{r=1}^{R} y_{n+h+1|n}^{(r)}$.

The fill-rate depends on the order level *S*, a relationship that may be represented by the function $\beta(S)$. The problem is to find that value of *S* which satisfies the condition $\beta(S) = \overline{\beta}$ where $\overline{\beta}$ is the target fill rate. The 'true' implicit function $\beta(S)$ is unknown. However $\hat{\beta}$ also depends on *S*, a relationship which that may be designated by $\hat{\beta}(S)$. Using $\hat{\beta}(S)$ as an approximation for $\beta(S)$, the problem can be revamped to one of finding the solution $\hat{S}$ of the equation $\hat{\beta}(S) = \overline{\beta}$.

The parametric bootstrap procedure consists of the following steps:

a) Simulate from the appropriate exponential smoothing demand model the $y_{n+j|n}^{(r)}$ for $j = 1$ to $h+1$, $r = 1$ to *R*.

b) Use a binary search procedure to solve the implicit function equation $\hat{\beta}(S) = \overline{\beta}$ for $\hat{S}$.

Note that $\hat{\beta}(S)$ is evaluated at step (b) for each trial value of *S* using the demands from step (a). There is no need to regenerate the demands for each function evaluation.

This bootstrap procedure is easily implemented on modern computers. But it is likely to yield values for the order level slightly below those actually required because the parametric bootstrap method ignores the effects of estimation error. It is possible to adapt this procedure to largely overcome this problem using a more complex prediction methodology from Ord, Koehler and Snyder (1997). This option is not pursued here.

The fill rate is an appropriate criterion for evaluating whether there are significant gains from using a relative error rather than an additive error approach to forecasting when it is known that demands are generated by a relative error model. Any differences that might occur can be gauged from a simulation study. Let LLDSM(q,θ) denote the local level with drift and seasonal cycle model with parameter vector θ. Furthermore, let $\hat{\theta}_q$ denote the maximum likelihood estimate of the parameter vector θ from LLDSM(q). The steps in each replication of the simulation are:

a) Generate a time series of length *n* from the 'true' model LLDSM(1, θ).

b) Estimate the time series on the assumption that the LLDM(0, θ) is the appropriate model to yield estimate $\hat{\theta}_0$.

c) Use LLDM(0, $\hat{\theta}_0$) with the bootstrap method to find the order level denoted by $\hat{S}_0$.

d) Estimate the time series on the assumption that the LLDM(1, θ) is the appropriate model to yield estimate $\hat{\theta}_1$.

e) Use LLDM(1, $\hat{\theta}_1$) with the bootstrap method to find the order level denoted by $\hat{S}_1$.

f) Generate an ensemble of future demands from the 'true' model LLDM(1, θ) and evaluate the fill rates achieved with $\hat{S}_0$ and $\hat{S}_1$ respectively. These fill rates are designated by $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively.

The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ from each replication of the above steps potentially change. These values can be collected into a sample. The two samples may be compared to determine whether there are significant differences between the additive and relative error demand models. They can also be compared with the nominal fill rate $\overline{\beta}$ to gauge the effect of ignoring the estimation error in the parametric bootstrap or any bias in the forecast procedure.

In the simulation study it was assumed that:

a) The stock position is reviewed at the beginning of each week.

b) Orders are delivered after a delay of 9 weeks. Thus the aim is to control inventories in the week following the delivery, namely week 114 (2*52+9+1).

c) Deliveries occur at the start of a week, immediately following the review.

d) Weekly demand is governed by the LLDSM(1) with $m = 100$, $b = 0.1$, $c_t = 0.5\sin(2\pi t/52)$, $a = 0.5$ and $s = 0.05$. The growth rate, in annual terms, is 5.2 (ie 0.1 * 52 weeks). This, relative to the initial level, is a little over 5 percent per annum.

e) Weekly demand data for two years is available for forecasting purposes so that the current review occurs at the beginning of period 105 (ie 2*52+1).

f) The target fill-rate is $\overline{\beta} = 95\%$.

A number of simulation experiments were conducted under a variety of conditions. Each simulation experiment involved 200 replications. At each replication the bootstrap method for finding the order level itself involved 1000 replications. The results are summarised in Table 2. The benchmark case represents a situation that we think may be fairly typical in the inventory control context. The other cases were obtained by varying one factor at a time from its benchmark value.

Insert Table 2 and Figure 3 about here

The following observations can be made.

a) The means are generally below the medians. The distributions of the simulated fill-rate must possess a left skew. This is exemplified by the distribution in Figure 3 for the benchmark case.

b) The medians obtained with the bootstrap method from LLDM(1) are usually about one percent below the target fill-rate of 95 percent. This gap is probably due to the fact that the bootstrap method ignores the effect of estimation error. Given the size of this gap, refinements geared to eliminating this problem appear to be unwarranted.

c) The median fill-rates associated with LLDM(0) are little lower again. There appear to be some gains from using the relative error approach when the data generating process involves relative errors.

d) The gains from using the bootstrap method with LLDM(1) instead of LLDM(0) increase with higher growth rates. The changes in the underlying level are larger and the fluctuations of the irregular component increase as a consequence. Nevertheless, the growth rate has to reach unrealistic levels before the differences become pronounced.

e) Variations in most factors have little impact on the median fill-rate for the LLDM(1) bootstrap method.

4. CONCLUSIONS

In this paper we have proposed a generalisation of the additive local level model or its equivalent, the ARIMA(0,1,1) model, to incorporate a general form of conditional heteroscedasticity. It was demonstrated that simple exponential smoothing, in its traditional form, remains the valid updating relationship under this more general class of models. The only change required is in the form of the criterion function used for selecting the estimates. A simulation indicated that the maximum likelihood estimates obtained with simple exponential smoothing under a level dependent form of heteroscedasticity are almost identical to those for the homoscedastic case. Since the homoscedastic case is inherently easier to estimate than its multiplicative counterpart we recommend the use of the former for estimation purposes.

The issue of heteroscedasticity becomes more critical in the prediction context. Analytical formulae become unreliable for the multiplicative case. We therefore recommend a two-stage procedure:

a) estimate m, a and s using the additive model

b) use the estimates from the previous step in conjunction with the multiplicative model to simulate the prediction intervals.

Appendix

This appendix contains the derivation of the formulae for the mean and variance of the linear and multiplicative local trend/seasonal models. To simplify notation the origin for forecasting, designated period *n* in the body of the paper, will be relabelled period 0. The prediction horizon is designated by *h*. Thus the random *h*-vector $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_h \end{bmatrix}'$ designates *h* unknown future values of the time series. Furthermore, ℓ_0 and b_0 denote the local level and local rate at the start of the prediction origin. They no longer represent the seed values for these quantities in the period prior to the sample. These quantities are known exactly. The vector $\gamma = \begin{bmatrix} c_{-r+1} & c_{-r+2} & \cdots & c_0 \end{bmatrix}'$ of seasonal factors required for forecasting is also known exactly. The formulae to be derived in this appendix are therefore based on the assumption that there is no estimation error.

The derivations rely extensively on a matrix *B* called the 'backward shift' matrix. It is the matrix counterpart of the backward shift operator used so extensively in Box and Jenkins (19XX). The notation employed, together with explanations, is shown in the following Table. To simplify matters it is assumed that *h* is an exact multiple of the seasonal lag *r*. Let $m = h/r$. It is the length of the forecast horizon measured in years.

Notation

ξ	the unit *h*-vector $\begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix}'$
1	the ones *h*-vector $\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}'$
τ	the arithmetic series *h*-vector $\begin{bmatrix}1 & 2 & \cdots & h\end{bmatrix}'$
$\tau^{(2)}$	the series *h*-vector $\begin{bmatrix}1 & 3 & 6 & 10 & \cdots & h(h+1)/2\end{bmatrix}$
B	the backward shift matrix where $b_{ii-1} = 1$ for $i = 1, \ldots, n$ and $b_{ij} = 0$ otherwise. eg $\begin{bmatrix}0 & 0 & 0\\1 & 0 & 0\\0 & 1 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}0\\x_1\\x_2\end{bmatrix}$
B^r	the backward shift matrix of lag *r* where $b_{ii-r} = 1$ for $i = r+1, \ldots, n$ and $b_{ij} = 0$ otherwise. eg. if $B = \begin{bmatrix}0 & 0 & 0 & 0\\1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\end{bmatrix}$ then $B^2 = \begin{bmatrix}0 & 0 & 0 & 0\\0 & 0 & 0 & 0\\1 & 0 & 0 & 0\\0 & 1 & 0 & 0\end{bmatrix}$ and $\begin{bmatrix}0 & 0 & 0 & 0\\0 & 0 & 0 & 0\\1 & 0 & 0 & 0\\0 & 1 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}0\\0\\x_1\\x_2\end{bmatrix}$. Note that $B^j = O$ for all $j \geq n$.

S	partial sum matrix being a unit lower triangular matrix with all elements below the diagonal equal to 1. eg. $S = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ so that $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 + x_2 \\ x_1 + x_2 + x_3 \end{bmatrix}$. Note that $S = (I - B)^{-1} = I + B + B^2 + \cdots + B^{n-1}$
$S^{(r)}$	defined by $S^{(r)} = (I - B^r)^{-1} = I + B^r + B^{2r} + \cdots + B^{mr}$
Ξ	$\Xi = [I_r \quad O_r \quad \cdots \quad O_r]'$ is the matrix counterpart of ξ
Z	$Z = [I_r \quad I_r \quad \cdots \quad I_r]'$

Useful Relationships

$S = (I - B)^{-1}$	$I + BS = I + SB = S$	$S\xi = \mathbf{1}$	$S\mathbf{1} = \tau$
$\xi + B\mathbf{1} = \mathbf{1}$	$\mathbf{1} + B\tau = \tau$	$S^{(r)}\Xi = Z$	$B^r S^{(r)} + I = S^{(r)}$

Linear Seasonal Model

Proposition

A random vector y governed by the linear seasonal local trend model has a mean $\mu_1 + \mu_2$ and variance matrix AA' where $\mu_1 = \mathbf{1}\ell_0 + \tau b_0$, $\mu_2 = Z\gamma$ and

$$A = I + BS(\alpha_1 I + \alpha_2 S) + \alpha_3 B^r S^{(r)}.$$

1 *Proof*

Stack the equations of LTSM(0) to give

$$y = \xi(\ell_0 + b_0) + \Xi\gamma + B(\ell + b) + B^r c + e \quad \text{(F.1)}$$

$$\ell = \xi(\ell_0 + b_0) + B(\ell + b) + \alpha_1 e \quad \text{(F.2)}$$

$$b = \xi b_0 + Bb + \alpha_2 e \quad \text{(F.3)}$$

$$c = \Xi\gamma + B^r c + \alpha_3 e. \quad \text{(F.4)}$$

Solve (2.2) for *c* to give

$$c = \mu_2 + \alpha_3 S^{(r)} e.$$

Also solve (2.5) for *b* to yield

$$b = \mathbf{1} b_0 + \alpha_2 Se.$$

Similarly, the equation (3.1) may be solved for ℓ to give

$$\ell = \mu_1 + (\alpha_2 SB + \alpha_1) Se.$$

Substituting these results into (2.6) we obtain

$$y = \mu + Ae$$

where $\mu = \mu_1 + \mu_2$ and $A = I + BS(\alpha_1 I + \alpha_2 S) + \alpha_3 B^r S^{(r)}$. The result follows.

Multiplicative Seasonal Model

The situation is more complicated for the multiplicative model. It is necessary to form diagonal matrices from certain vectors. Thus diag(*a*) represents the matrix with diagonal elements $a_1, a_2, \ldots, a_n$. We utilise the following properties of diagonal matrices:

$$\operatorname{diag}(a+b) = \operatorname{diag}(a) + \operatorname{diag}(b) \qquad \operatorname{diag}(a)b = \operatorname{diag}(b)a$$

LTSM(1) is approximated by the model

$$y_t = (\ell_{t-1} + b_{t-1})c_{t-r} + \left(E(\ell_{t-1}|I_n) + E(b_{t-1}|I_n)\right)E(c_{t-r}|I_n)e_t \qquad \text{(F.5)}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1\left(E(\ell_{t-1}|I_n) + E(b_{t-1}|I_n)\right)e_t \qquad \text{(F.6)}$$

$$b_t = b_{t-1} + \alpha_2\left(E(\ell_{t-1}|I_n) + E(b_{t-1}|I_n)\right)e_t \qquad \text{(F.7)}$$

$$c_t = c_{t-r} + \alpha_3 E(c_{t-r}|I_n)e_t \qquad \text{(F.8)}$$

The coefficients of the errors are converted by this approximation from stochastic to fixed quantities that are easier to manipulate.

Let the mean trend and seasonal vectors be defined by $\mu_1 = \mathbf{1}\ell_0 + \tau b_0$ and $\mu_2 = Z\gamma$. The equations (2.6)-(2.7) can be stacked to give

$$y = \operatorname{diag}\left((\ell_0 + b_0)\xi + B(\ell + b)\right)\left(\Xi\gamma + B^r c\right) + \operatorname{diag}(\mu_1)\operatorname{diag}(\mu_2)e \qquad \text{(F.9)}$$

$$\ell = \xi(\ell_0 + b_0) + B(\ell + b) + \alpha_1 \operatorname{diag}(\mu_1)e \qquad \text{(F.10)}$$

$$b = \xi b_0 + Bb + \alpha_2 \operatorname{diag}(\mu_1)e \qquad \text{(F.11)}$$

$$c = \Xi\gamma + B^r c + \alpha_3 \operatorname{diag}(\mu_2)e \qquad \text{(F.12)}$$

Equation (F.12) may be solved for *c* to give

$$c = \mu_2 + \alpha_3 S^{(r)} \text{diag}(\mu_2) e \,. \qquad \text{(F.13)}$$

Solving for *b* in (F.11) gives

$$b = \mathbf{1} b_0 + \alpha_2 S \text{diag}(\mu_1) e \,. \qquad \text{(F.14)}$$

Substituting (F.14) into (F.10) and then solving for ℓ gives

$$\ell = \mu_1 + (\alpha_1 I + \alpha_2 BS) S \text{diag}(\mu_1) e \,. \qquad \text{(F.15)}$$

The expressions (3.1)-(2.7) can be substituted into (F.9) to give a nonlinear relationship for y in terms of the errors e. Given that the errors are in relative terms, products of the form $e_i e_j$ are relatively small. The linear component of the relationship, with the general form $y_L = \mu + Ae$ must be a good approximation for *y*. Noting that $\ell + b$ simplifies to $\ell + b = \mu_1 + \mathbf{1} b_0 + (\alpha_1 I + \alpha_2 S) S diag(\mu_1) e$ it can be shown that $\ell = \mu_1 + (\alpha_1 I + \alpha_2 BS) S \text{diag}(\mu_1) e$ and

$$A = \text{diag}(\mu_2) BS(\alpha_1 I + \alpha_2 S) \text{diag}(\mu_1) + \alpha_3 \text{diag}(\mu_1) B^r S^{(r)} \text{diag}(\mu_2) + \text{diag}(\mu_1) \text{diag}(\mu_2) \,.$$

An approximation for the variance matrix of the prediction distribution is therefore given by $s^2 AA'$. Although it is relatively complex, the formula for *A* is readily calculated in a matrix oriented computer language such as Gauss or Matlab. Provided that the approximations made during its derivation combined with the normal approximation for a distribution that is not normal, do not lead to serious error, this option is a convenient way to derive the prediction distribution.

REFERENCES

Box, G.E.P and G.M. Jenkins, “Time Series Analysis: Forecasting and Control”, Holden-Day, San Francisco, 1976.

Brown, R.G., “Statistical Forecasting for Inventory Control”, Mc.Graw-Hill, New York, 1959.

Brown, R. G., “Decision Rules for Inventory Management”, Dryden Press, Hinsdale, 1967.

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

Harrison, P.J., “Exponential Smoothing and Short-term Sales Forecasting”, Management Science, 13 (1967), 821-842.

Harrison, P.J. and C.F.Stevens, “Bayesian Forecasting”, Journal of the Royal Statistical Society, Series B, 38, 205-247.

Harvey, A.C., “Forecasting, Structural Time Series Models and the Kalman Filter”, Cambridge University Press, Cambridge, 1991.

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

Holt, C.E. (1957), “Forecasting Trends and Seasonals by Exponentially Weighted Averages”, ONR Memorandum, 52, Carnegie Institute of Technology, Pittsburgh, 1957.

Johnston, F.R. and P.J. Harrison, “The Variance of Lead Time Demand”, Journal of the Operational Research Society, 37 (1986), 303-308.

Makridakis, S. et. al., “The Accuracy of Extrapolative (Time Series) Methods: Results of a Forecasting Competition”, Journal of Forecasting, 1, (1982), 111-153.

Muth, J.F., “Optimal Properties of Exponentially Weighted Averages”, Journal of the

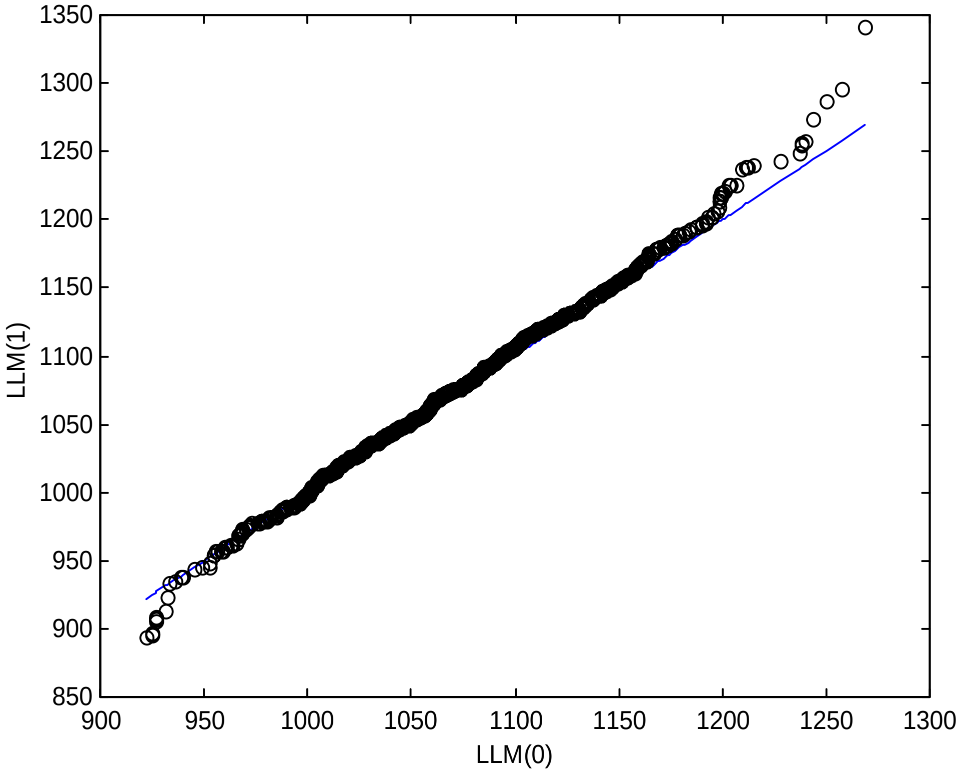
American Statistical Association, 55 (1960), 299-306.

Ord, J.K. , A. B. Koehler, and R. D. Snyder, “Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models”, Journal of the American Statistical Association, forthcoming.

Schweppe, F. “Evaluation of Likelihood Functions for Gaussian Signals”, IEEE Transactions on Information Theory, 11 (1965), 61-70.

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

Figure 1. Simulated lead-time demand: exponential smoothing



n	30, 50, 100, 200
a	0, 0.2, 0.5, 1.0
s	0.02, 0.05, 0.1

Table 1. Simulation options

Conditions				LLDM(0) fill rate		LLDM(1) fill rate	
b	*s*	*a*	*n*	mean	median	mean	median
Benchmark case							
0.1	0.05	0.5	104	91	93	91	93
Growth rate effect							
1	0.05	0.5	104	89	90	93	94
Variability effect							
0.1	0.1	0.5	104	91	92	90	94
Sample size effect							
0.1	0.05	0.5	260	93	93	94	94
Smoothing parameter effect							
0.1	0.05	0.1	104	92	93	92	94

Table 2. Summary of simulated percentage fill-rates with $m = 100$ and $\bar{\beta} = 95\%$.

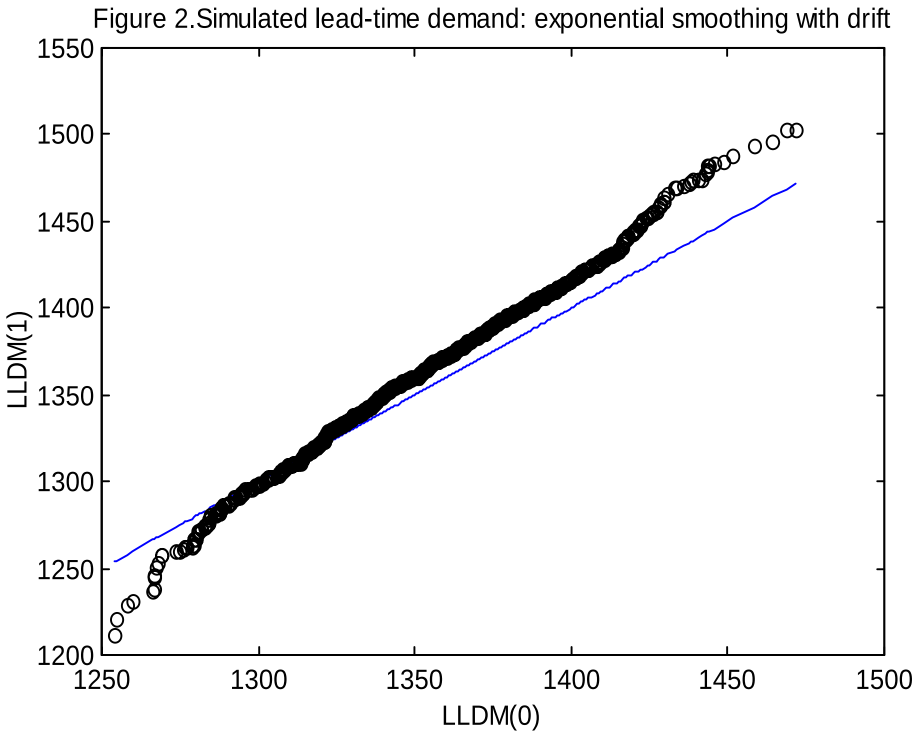


Figure 2.Simulated lead-time demand: exponential smoothing with drift
1550
1500
1450
1400
1350
1300
1250
1200
LLDM(1)
1250
1300
1350
1400
1450
1500
LLDM(0)

Figure 3. Comparison of simulated fill rates. Target fill rate = 95%

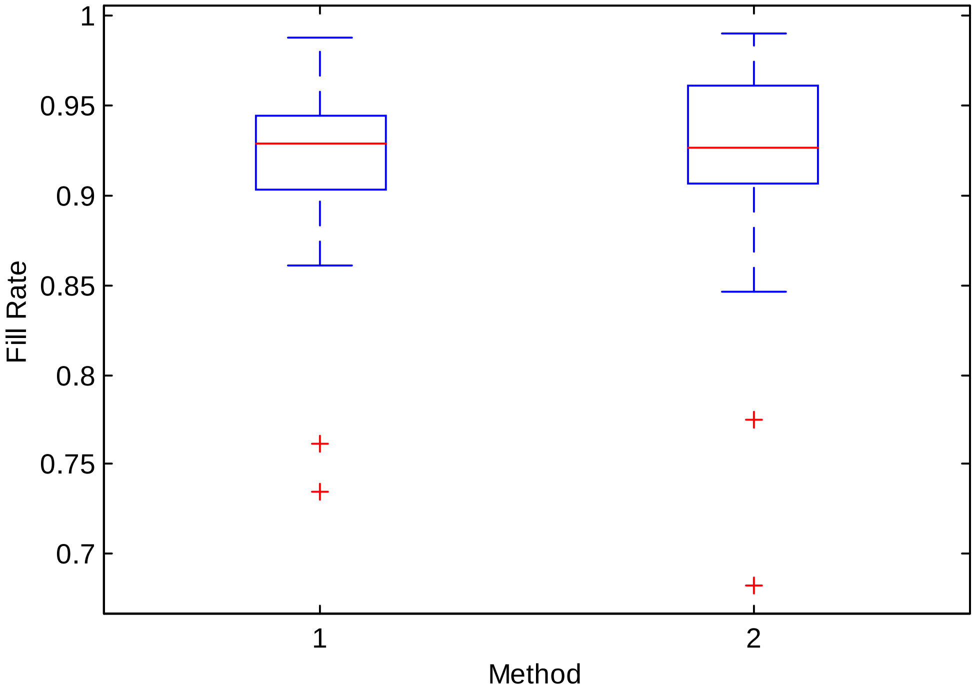
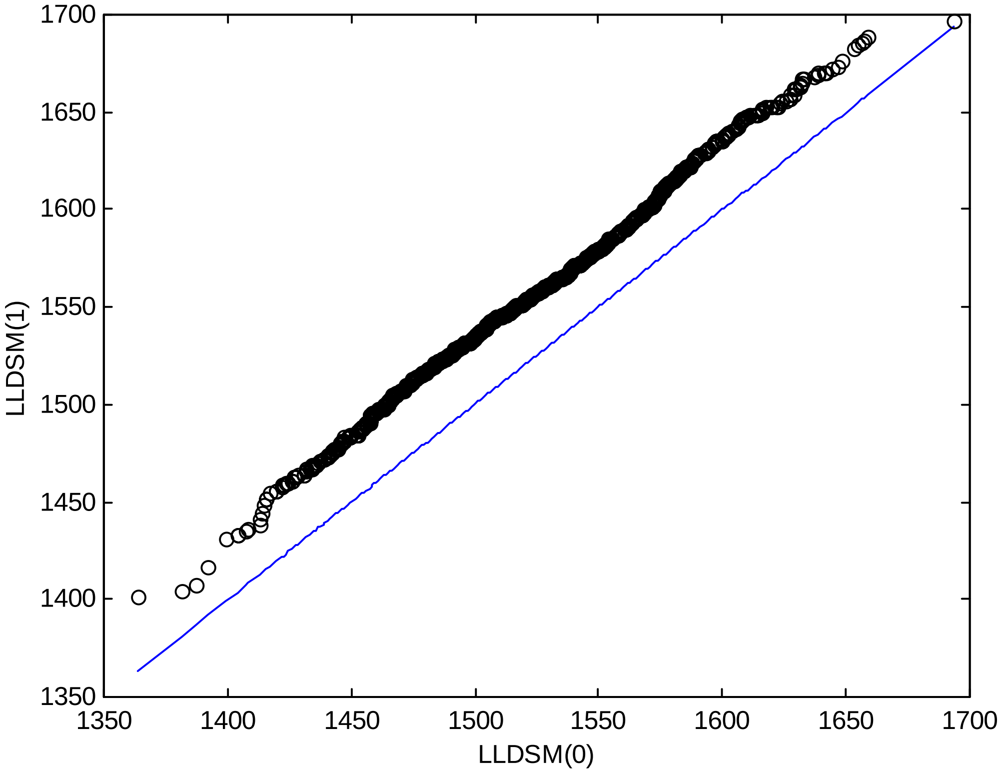


Figure 3.Simulated lead-time demand: seasonal exponential smoothing with drift



Seasonal case

	LLDSM(0)	LLDSM(1)
Mean	0.82	0.94
Median	0.84	0.95

Figure 3. Comparison of simulated fill rates. Target fill rate = 95%