ISSN 1440-771X
ISBN 0 7326 1076 1

# Bayesian Exponential Smoothing

**Catherine S. Forbes, Ralph D. Snyder and Roland G. Sharmi**

**Working Paper 7/2000**

**August 2000**

**DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS**

MONASH UNIVERSITY

AUSTRALIA

# Bayesian Exponential Smoothing

Catherine S. Forbes [1]
Department of Econometrics and Business Statistics
Monash University
Victoria 3800
Australia
catherine.forbes@buseco.monash.edu.au

Ralph D. Snyder
Department of Econometrics and Business Statistics
Monash University
Victoria 3800
Australia
ralph.snyder@buseco.monash.edu.au

Roland G. Shami
Department of Econometrics and Business Statistics
Monash University
Victoria 3800
Australia
roland.shami@buseco.monash.edu.au

This work was partially supported by an Australian Research Council Small Grant.
[1] Denotes corresponding author.

## ABSTRACT

In this paper, a Bayesian version of the exponential smoothing method of forecasting is proposed. The approach is based on a state space model containing only a single source of error for each time interval. This model allows an improvement to current practices in exponential smoothing by providing both point predictions and measures of the uncertainty surrounding them. The method proposed calculates posterior prediction and parameter distributions via Monte Carlo composition. We evaluate the method with a Monte Carlo simulation study and apply it to forecasting car part demand. The main advantage of the approach is that it produces exact, small sample prediction distributions. It also works very quickly on modern computing machines.



## KEYWORDS

Time series analysis, forecasting, structural model, local level model, prediction interval.

## 1. INTRODUCTION

Exponential smoothing methods are widely used for forecasting demand in supply chain management applications (Gardner, 1985). Their relative simplicity and robustness, together with their reliance on the 'stylised facts' of time series analysis, mean that they are particularly well suited for automated approaches to forecasting. Their central place in time series analysis has been reinforced by repeated successes against more sophisticated approaches in a succession of forecasting competitions (Makridakis, Anderson, Carbone, Fildes, Hibon, Lewandowski, Newton, Paerzen and Winkler, 1982; Makridakis, Chatfield, Hibon, Lawrence, Mills, Ord and Simmons, 1993).

The traditional focus of exponential smoothing implementations has normally been on getting good point predictions. The measurement of the uncertainty surrounding such predictions, however, has been largely ad hoc - see Johnston and Harrison, 1986; Harvey and Snyder, 1990; Snyder, Koehler and Ord, 1999 for the details. Accurate measures of uncertainty can be very important in some applications. They are needed, for example, for the determination of appropriate levels of safety stock in inventory control. The lack of a suitable statistical framework hampered progress in this direction.

An early statistical framework for exponential smoothing was provided by Box and Jenkins (1976). They demonstrated that all linear forms of exponential smoothing could be rationalised in terms of ARIMA models. The links, however, were obscure and so their approach had only limited impact on the exponential smoothing fraternity. Harrison and Stevens (1976) followed with a multi-disturbance state space approach.

They introduced, in the ‘recipes’ section of their papers, special cases that in today’s terminology are called the local level, local trend and local seasonal models. The reliance on the multi-disturbance form of the state space model necessitated the use of the Kalman filter (Kalman, 1960; Kalman and Bucy, 1961) for estimation and prediction purposes. Invariant forms of the Kalman filter converge to a steady state corresponding to exponential smoothing (West and Harrison, 1989, pp51-56). The link between framework and method was therefore more direct than that proposed by Box and Jenkins. Nevertheless, one had to possess an understanding of the Kalman filter before seeing the link with exponential smoothing.

The seeds of another approach may be found in Snyder (1985). A state space framework based on a single source of error, with equations closely resembling those of exponential smoothing, was proposed. For the first time, a statistical framework with clear, direct links with exponential smoothing had been found. It was this work that laid the basis for the nonlinear state space framework in Ord, Koehler and Snyder (1997) that has been used, amongst other things, to rationalise the multiplicative form of the Holt-Winters method of forecasting (Winters, 1960).

The single source of error (SSOE) state space model, called the innovations form by Aoki and Havenner (1991), has advantages over its traditional, multi-source of error (MSOE) counterpart due to its simpler construction. The SSOE state space models have a simple equivalence to the popular ARIMA processes (Snyder, Ord and Koehler, 2001), whereas the MSOE models have more arduous relationships relating parameters to their ARIMA counterparts (Ray, 1989). By exploiting its relationship with exponential smoothing methods, the SSOE model can be rewritten as a simple regression model, conditional on the smoothing constants. This, it will be seen, enables the use of Monte Carlo composition, rather than the Kalman filter combined with Markov chain Monte Carlo (MCMC) methods, to produce samples from the exact posterior predictive distributions.

The MSOE framework postulates random errors in both the measurement of the system, taken to be a function of unobservable state components, and in the evolution of those unobserved components. Bayesian estimation and forecasting using linear forms of these models is detailed in West and Harrison (1989). Their approach was based predominantly on the use of conjugate prior distributions, along with fairly strong restrictions on the relationship between the two error distributions. These restrictions were removed by Carlin, Polsen and Stoffer (1992), who explored the use of MCMC methods for non-linear and non-conjugate forms of the MSOE model. Carter and Kohn (1994) showed that the use of the Kalman filter leads to faster convergence of MCMC algorithms in this context. Harvey (1984) used classical maximum likelihood methods to estimate the unknown MSOE disturbance variances.

The purpose of this paper is to provide a complete discussion of Bayesian statistical exponential smoothing analysis using the SSOE model. To date statistical analysis of the innovations form has only been considered in conjunction with maximum likelihood methods. While maximum likelihood approaches have successfully produced point estimates and approximate confidence intervals for model parameters, no other exact method for producing prediction intervals has been developed for the SSOE model. In addition, classical approaches, such as maximum likelihood methods, do not allow for the incorporation of subjective (non-data) information when available. In contrast the new Baysian approach is easy to implement, and produces exact prediction intervals based on a finite sample of observations.

The plan of this paper is as follows. Motivated by exponential smoothing methods, the SSOE form of the state space model and calculation of its likelihood function are reviewed in Section 2. Our proposal for a Bayesian analysis of the model and computation of the joint posterior distribution via Monte Carlo composition is presented in Section 3. Use of the output from our algorithm for deriving Bayesian marginal posterior parameter estimates, forecasting distributions and related quantities is outlined in Section 4. This method is illustrated in an application to forecasting the demand for a car part presented in Section 5. Results from a Monte Carlo simulation study demonstrating the repeated sampling performance of the proposed forecasting method is presented in Section 6. The paper concludes with some final comments in Section 7.

## 2. A STATISTICAL FRAMEWORK FOR EXPONENTIAL SMOOTHING

### 2.1 Simple Exponential Smoothing

Exponential smoothing (Brown, 1959), in its simplest form, involves successive applications of the formula

$$l_t = \alpha y_t + (1-\alpha) l_{t-1}, \tag{1}$$

where $y_t$ is the value of a univariate time series in period *t* and $l_t$ is a 'smoothed' value representing the underlying level at the end of period *t*. If this formula was applied to the semi-infinite sample $\{y_{t-j}\}_{j=0}^{\infty}$ then the underlying level may be resolved into an exponentially weighted average

$$l_t = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{t-j} . \quad (2)$$

Provided that $0 \le \alpha \le 1$, the weights decline with increases in the age index *j*. In this ideal situation with an infinite sample, $l_t$ may be thought of as the 'true' underling level.

Equation (1) may be rewritten as $l_t = l_{t-1} + \alpha(y_t - l_{t-1})$. If $l_{t-1}$ is thought of as a prediction, then $e_t = y_t - l_{t-1}$ may be interpreted as a prediction error. At the start of typical period *t* while $y_t$ is unknown, the error $e_t$ is also unknown. If statistical distributions for the prediction errors are provided, such as assuming that the errors have independent normal distributions with zero mean and common variance $\sigma^2$, the result can be viewed as a statistical model for simple exponential smoothing:

$$y_t = l_{t-1} + e_t \quad (3)$$

$$l_t = l_{t-1} + \alpha e_t . \quad (4)$$

This statistical model, appropriately referred to as a local level model, provides a description of the data evolution process. It also provides the opportunity for a comprehensive statistical approach that includes the derivation of associated prediction distributions. Note that if the value of the so-called 'smoothing parameter', $\alpha$, is zero then the level of the overall series remains unchanged through time. Conversely, if $\alpha = 1$, then the level at the end of period *t-1*, is given by the observation $y_{t-1}$. Thus, $\alpha$ can be interpreted as the proportion of the variability in the level that can be attributed to a structural change in the underlying process.

### 2.2 General Exponential Smoothing

A generalisation of simple exponential smoothing allows for additional components to the underlying level. All components, such as level, trend and seasonal effects, may be collected together in a vector designated by $b_t$. This so-called 'state vector' is governed by the first-order linear relationship

$$b_t = Db_{t-1} + \alpha y_t , \tag{5}$$

where $D$ is a $k$x$k$ matrix and $\alpha$ is a $k$-vector of unknown 'smoothing' parameters. Notice that equation (5) only describes a point estimate of the unobserved components and does not automatically afford an approach to quantify the uncertainty associated with those point estimates. Backsolving (5) gives

$$b_t = \sum_{j=0}^{\infty} D^j \alpha y_{t-j} . \tag{6}$$

A statistical model encompassing general exponential smoothing as given in (5) is based on the SSOE state space form (Snyder 1985):

$$y_t = x'b_{t-1} + e_t \qquad \text{measurement equation} \tag{7}$$

$$b_t = Tb_{t-1} + \alpha e_t \qquad \text{transition equation} \tag{8}$$

where $x$ is a fixed $k$-vector, and $T$ is a $k$x$k$ transition matrix typically taken to contain known constants. The $e_t$'s are assumed to be independent normally distributed disturbances with common mean zero and variance $\sigma^2$, and $\alpha$ again is a $k$-vector of 'smoothing' parameters. Here, the value of the observed series at time $t$, $y_t$, is described as arising from a known linear combination of unobservable components, $x'b_{t-1}$, and an independent disturbance term, $e_t$. The two-part model allows the practitioner an opportunity to describe the feedback mechanism involved in the evolution of the data via the transition equation (8). Note that by solving (7) for $e_t$ and substituting the result into (8), we obtain the exponential smoothing updating relationship (5) where $D = T - \alpha x'$. The eigenvalues of $D$ must lie within the unit circle to ensure that the coefficients of the series values in (6) decline with increases in the age index $j$.

It is often thought that models with only one primary disturbance source are more restricted than their multi-disturbance counterparts. For state space models, this is not necessarily the case. It is shown in Snyder, Ord and Koehler (2001) that the SSOE form (7) and (8) is equivalent to the entire ARIMA class of models. Furthermore, Godolphin and Stone (1980) have demonstrated that the conventional MSOE form with independent disturbances is less general than the ARIMA class. Given that the MSOE form is usually applied with independent disturbances to control the number of parameters, there is no loss from using the SSOE form in practice.

### 2.3 Structural Models

The SSOE model is very general, and so for practical purposes is usually narrowed to one containing a local level $l_t$, a local growth rate $g_t$, and local seasonal effects $s_t$:

$$y_t = l_{t-1} + g_{t-1} - \sum_{j=1}^{m-1} s_{t-j} + e_t \tag{9}$$

$$l_t = l_{t-1} + g_{t-1} + \alpha_1 e_t \tag{10}$$

$$g_t = g_{t-1} + \alpha_2 e_t \tag{11}$$

$$s_t = -\sum_{j=1}^{m-1} s_{t-j} + \alpha_3 e_t , \tag{12}$$

where $m$ is the number of seasons. This is the model underlying the additive Holt-Winters method of forecasting (Winters, 1960). It can be put into general SSOE form, as demonstrated below for the quarterly case. Let $b_t' = (l_t, g_t, s_t, s_{t-1}, s_{t-2})$, $x' = (1,1,-1,-1,-1)$, $\alpha' = (\alpha_1, \alpha_2, \alpha_3, 0, 0)$ and

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{13}$$

The reduced form of this model is a seasonal ARIMA model.

### 2.4 The Likelihood Function

The parameters $\alpha$ and $\sigma^2$ are unknown. Furthermore, semi-infinite samples are not available. Assuming that the forecaster possesses the finite sample $\{y_1, y_{2,} \dots y_n\}$, equation (6) can be rewritten as:

$$b_t = \sum_{j=t}^{\infty} D^j \alpha y_{t-j} + \sum_{j=0}^{t-1} D^j \alpha y_{t-j} \, . \tag{14}$$

The first sum involves 'missing' values associated with times before period 1. Note, however, that (14) can be rewritten in terms of a seed vector $b_0$ as

$$b_t = D^t b_0 + \sum_{j=0}^{t-1} D^j \alpha y_{t-j} \, . \tag{15}$$

The term containing $b_0$ in (15) summarises the entire past history of the series up to the start of the observation period. In addition to estimating the unknown parameters $\alpha$ and $\sigma^2$, the value of $b_0$ will need to be estimated as well. Given the assumption of independent, normally distributed errors, the conditional distribution of $y_t$ given the past observed data $Y_{t-1} = \{y_1, y_2, \cdots, y_{t-1}\}'$ is normal, i.e.

$$y_t | b_0, \alpha, \sigma^2, Y_{t-1} \sim N\left(x' b_{t-1}, \sigma^2\right), \tag{16}$$

with $b_{t-1}$ given by (15) for $t = 2, \cdots, n$. Although $b_0$ is unknown, and therefore must be estimated, for a given value of $\alpha$ it can be seen as the constant coefficient in a linear regression model. Thus, conditional on a particular value of $\alpha$, the SSOE state space model can be transformed into a simple linear regression model

$$\tilde{y}_t = \tilde{x}_t' b_0 + e_t \ , \tag{17}$$

where

$$\tilde{y}_t = y_t - x' \sum_{j=1}^{t-1} D^{j-1} \alpha y_{t-j} \quad \text{and} \quad \tilde{x}_t' = x' D^{t-1} \tag{18}$$

for $t = 1, \ldots, n$. The $\tilde{y}_t$ are computed easily and quickly using the following modification of exponential smoothing. Using a trial seed value of $\bar{b}_0 = 0$ and a given value of $\alpha$, the one step ahead trial predictions, $\bar{b}_{t-1}$, can be produced using the exponential smoothing recursion in (5). The transformed values of the series are then given by

$$\tilde{y}_t = y_t - x' \bar{b}_{t-1}, \tag{19}$$

along with $\tilde{x}_t = \tilde{x}_{t-1} D$, where $\tilde{x}_0 = x$. Once the transformed data values $\tilde{y}_t$ and regression variables $\tilde{x}_t$ have been found, the likelihood function can be computed for any value of $\left(b_0, \alpha, \sigma^2\right)$ using

$$L\left(b_0, \alpha, \sigma^2\right) \propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} \left(\tilde{y}_t - \tilde{x}_t' b_0\right)^2\right\}, \tag{20}$$

or in vector notation,

$$L\left(b_0,\alpha,\sigma^2\right) \propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\tilde{Y}-\tilde{X}b_0\right)'\left(\tilde{Y}-\tilde{X}b_0\right)\right\}, \tag{21}$$

where $\tilde{X}'=\left(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n\right)$ and $\tilde{Y}'=\left(\tilde{y}_1,\tilde{y}_2,\cdots,\tilde{y}_n\right)$.

There are several points to note regarding the above likelihood function. As noted earlier, for any given value of $\alpha$, the likelihood function is of a standard linear regression form. In addition, $D$, $\tilde{Y}$ and $\tilde{X}$ are all functions of $\alpha$ alone, and not functions of $b_0$ or $\sigma^2$. As a consequence, however, the choice of $\alpha$ is critical and complicated by the fact that the log-likelihood function is a highly non-linear function of $\alpha$.

## 3. A BAYESIAN ANALYSIS

The exponential smoothing model provides a description of the evolution of time series components based on the past values of these components and the relative impact of noise. This description is simpler than the MSOE state space model and is characterised by the unknown parameters $b_0$, $\alpha$ and $\sigma^2$. Given these parameters, the likelihood function is easy to compute as the model can be transformed into a linear regression. Bayesian analysis for linear regression under non-informative priors has been available for some time; see, for example, Zellner (1971). However, in this case the linear regression model is conditional on the smoothing parameter vector, $\alpha$. Regression problems such as this one have been encountered before in many other models; see Geweke and Terui (1993) for a complete discussion in the context of threshold autoregressive (TAR) processes.

To complete a Bayesian analysis, a joint prior distribution for the unknown parameters, $b_0$, $\alpha$ and $\sigma^2$ must be specified. To enable computation of the posterior distribution using Monte Carlo composition, a standard non-informative regression prior is used for $b_0$ and $\sigma^2$. In our examples, where $\alpha$ is a scalar, we take a uniform marginal prior distribution for $0<\alpha<1$. As the choice of this prior will not alter the algorithm, however, alternatives are possible within this framework. In general, we consider the joint prior density

$$p\left(b_0, \alpha, \sigma^2\right) \propto \sigma^{-d}\, p(\alpha), \tag{22}$$

where typically $d$ = 1 or 2. We take $d$ = 2 in our examples. The density of the posterior distribution is proportional to the product of the likelihood function in (21) and the prior density in (22). That is,

$$p\left(b_0, \alpha, \sigma^2 | Y_n\right) \propto \sigma^{-(n+d)} p(\alpha) \exp\left\{-\frac{1}{2\sigma^2}\left(\tilde{Y} - \tilde{X} b_0\right)' \left(\tilde{Y} - \tilde{X} b_0\right)\right\}. \tag{23}$$

The posterior density can be decomposed into the product

$$p\left(b_0, \sigma^2, \alpha | Y_n\right) = p\left(b_0 | \sigma^2, \alpha, Y_n\right) \cdot p\left(\sigma^2 | \alpha, Y_n\right) \cdot p\left(\alpha | Y_n\right). \tag{24}$$

Using standard Bayesian algebraic manipulations on (23), it can be shown that $p\left(b_0 | \alpha, \sigma^2, Y_n\right)$ is the density of a $k$ dimensional normal distribution with mean vector and variance-covariance matrix,

$$\hat{b}_0 = \left(\tilde{X}'\,\tilde{X}\right)^{-1} \tilde{X}'\tilde{Y} \qquad \text{and} \qquad \sigma^2\left[\tilde{X}'\,\tilde{X}\right]^{-1}, \tag{25}$$

respectively. The distribution of $\sigma^2$, given $(\alpha, Y_n)$, can also easily be shown to have an inverted gamma distribution with shape and scale parameters

$$a = (n + d - k - 2)/2 \qquad \text{and} \qquad c = 2/(n-k)s^2, \tag{26}$$

respectively, where $(n-k)\tilde{s}^2 = \left(\tilde{Y} - \tilde{X}\hat{b}_0\right)'\left(\tilde{Y} - \tilde{X}\hat{b}_0\right)$. The resulting probability density function is

$$p\left(\sigma^2|\alpha, Y_n\right) = \Gamma(a)^{-1} c^{-a} \sigma^{2(a+1)} \exp\left\{-1/c\sigma^2\right\}. \tag{27}$$

To derive the marginal posterior distribution of $\alpha$, note that the product of the likelihood and prior density functions can be analytically integrated with respect to both $b_0$ and $\sigma^2$, and that the resulting function is proportional to the marginal posterior density of $\alpha$. That is,

$$p(\alpha|Y_n) \propto \int_{R^k} \int_0^\infty L\left(b_0, \sigma^2, \alpha\right) p\left(b_0, \sigma^2, \alpha\right) d\sigma^2 db_0, \tag{28}$$

and hence

$$p(\alpha|Y_n) \propto \left|\tilde{X}'\tilde{X}\right|^{-1/2} \tilde{s}^{-(n-k+d-2)} p(\alpha). \tag{29}$$

Note that as both $\tilde{X}$ and $\tilde{s}$ are functions of $\alpha$, (29) cannot be further reduced without specification of $p(\alpha)$. Notice also that this parameter $\alpha$ can be interpreted as relating to the portion of variability in the state vector attributed to unanticipated structural change, and hence the incorporation of prior information may be available, and quite beneficial to the forecasting performance of the model.

The next issue concerns the evaluation of $p(\alpha|Y_n)$. Consider the case with $k = 1$ where $\alpha$ is a scalar. To conform to the ideas of exponential smoothing, $\alpha$ is taken to lie within the interval between zero and one. Thus it is easy to evaluate

$$m(\alpha) = \left|\tilde{X}(\alpha)'\tilde{X}(\alpha)\right|^{-1/2} \tilde{s}(\alpha)^{-(n-k+d-2)} p(\alpha) \tag{30}$$

numerically on a grid of J+1 points $\alpha_j$ over the segment (0,1) and normalise

$$\hat{p}\left(\alpha_j|Y_n\right)=\frac{m\left(\alpha_j\right)\left(\alpha_{j+1}-\alpha_j\right)}{\sum_{j=0}^{J} m\left(\alpha_j\right)\left(\alpha_{j+1}-\alpha_j\right)}, \tag{31}$$

where $\alpha_0 = 0$ and $\alpha_J = 1$. In general, $\alpha$ is typically not of very high dimension and also is usually restricted to lie within a bounded region. Hence, numerical methods such as the simple one presented above, are available and appear to work reasonably well in the examples we have considered. To minimise redundant difficult calculations, we compute $\hat{p}\left(\alpha_j|Y_n\right)$ from (30) once, and save $\tilde{s}^2$, $\hat{b}_0$ and $\left(\tilde{X}'\tilde{X}\right)^{-1}$ associated with each $\alpha_j$.

## 4. SMOOTHING AND PREDICTION

### 4.1 Marginal Posterior Distributions

To obtain useful summaries of marginal posterior distributions, Monte Carlo composition is used in conjunction with Rao-Blackwellised estimation (Gelfand and Smith, 1990). An independent sample of *r* values $\left(b_0^{(i)},\sigma^{2(i)},\alpha^{(i)}\right)$ are simulated by first taking a sample of $\alpha^{(i)}$ values drawn from $\hat{p}\left(\alpha_j|Y_n\right)$ using the inverse cumulative distribution method. Next, $\sigma^{2(i)}$ is drawn from an inverted gamma distribution with parameters given in (26), where $\tilde{X}$, $\tilde{Y}$, and $\tilde{s}^{2(i)}$ are calculated using $\alpha = \alpha^{(i)}$. Finally $b_0^{(i)}$ is drawn from a normal distribution with mean $\hat{b}_0^{(i)}$ and variance $\sigma^{2(i)}\left(\tilde{X}^{(i)}{}'\tilde{X}^{(i)}\right)^{-1}$ as given in (25). Rao-Blackwellised density estimates for both of the remaining marginal posterior densities are computed using a simple average of the $r$ Inverted Gamma densities for the marginal density of $\sigma^2$,

$$\hat{p}\left(\sigma^2|Y_n\right) = \frac{1}{r}\sum_{i=1}^{r}\Gamma\left(a^{(i)}\right)^{-1} c^{(i)^{-a^{(i)}}} \sigma^{2\left(a^{(i)}+1\right)} \exp\left\{-1/c^{(i)}\sigma^2\right\} \tag{32}$$

and a simple average of the $r$ normal densities for the marginal density of $b_0$,

$$\begin{aligned}\hat{p}(b_0|Y_n) = \frac{1}{r}\sum_{j=1}^{r}\left(2\pi\sigma^{2(i)}\right)^{-k/2}\left|\tilde{X}\left(\alpha^{(i)}\right)'\,\tilde{X}\left(\alpha^{(i)}\right)\right|^{1/2} \\ \cdot\exp\left\{-\frac{1}{2\sigma^{2(i)}}\left(b_0-\hat{b}_0^{(i)}\right)'\,\tilde{X}\left(\alpha^{(i)}\right)'\,\tilde{X}\left(\alpha^{(i)}\right)\left(b_0-\hat{b}_0^{(i)}\right)\right\}.\end{aligned} \tag{33}$$

### 4.2 Prediction Distributions

In a similar way Rao-Blackwellised estimators can be used to calculate predictive densities. Given values for $b_0, \alpha, \sigma^2$ and past data $Y_n$, the distribution of $y_t$ at some future point $t = n + f$ is normal with mean $m_{n+f}$ and variance $V_{n+f}$, where

$$m_{n+f} = x'b_{n+f-1}, \qquad V_{n+f} = \sigma^2\left(1+\sum_{j=0}^{f-2} x'T^j\alpha\right), \tag{34}$$

$$b_{n+f-1} = T^{f-1}b_n, \qquad \text{and} \qquad b_n = D^n b_0 + \sum_{j=0}^{n-1} D^j y_{t-j}\,. \tag{35}$$

Note that the values of $b_n$ can be obtained recursively. Thus, the Rao-Blackwellised estimator of the predictive density of a future value $y_{n+f}$ is given by

$$\hat{p}\left(y_{n+f}|Y_n\right) = \frac{1}{r}\sum_{i=1}^{r}\left(2\pi V_{n+f}^{(i)}\right)^{-1/2} \exp\left\{-\frac{1}{2V_{n+f}^{(i)}}\left(y_{n+f}-m_{n+f}^{(i)}\right)^2\right\}. \tag{36}$$

Similarly, Rao-Blackwellised estimators of the mean and variance of the predictive distributions are as follows

$$\hat{E}\left[y_{n+f}|Y_n\right] = \frac{1}{r}\sum_{i=1}^{r} m_{n+f}^{(i)} \tag{37}$$

and

$$V\hat{a}r\left[y_{n+f}|Y_n\right] = \frac{1}{r}\sum_{i=1}^{r}\left\{V_{n+f}^{(i)} + m_{n+f}^{(i)2}\right\} - \hat{E}\left[y_{n+f}|Y_n\right]^2 . \quad (38)$$

Any quantile of these predictive distributions can be estimated by numerically integrating the appropriate (already normalised) density estimate, or by using empirical quantiles from simulated values.

**4.3 Smoothing the Data using a State Vector Estimator**

Estimates of the unobserved state vector, or system components $b_{t-1}$, may be desired as additional output of the statistical analysis. At each time $t > 0$, given a sample from the joint posterior distribution of $\alpha$ and $b_0$, a Rao-Blackwellised estimator of the marginal posterior mean of the state vector at time *t-1* is given by

$$\hat{E}\left[b_{t-1}|Y_n\right] = \frac{1}{r}\sum_{i=1}^{r} b_{t-1}^{(i)} , \quad (39)$$

where the $b_{t-1}^{(i)}$ are obtained by using the *r* sample values of $\alpha^{(i)}$ and $b_0^{(i)}$, and (15). Similarly, a smoothed version of the observed data can be obtained by using $x'\hat{E}\left[b_{t-1}|Y_n\right]$. Clearly, interval estimates are also available as are Rao-Blackwellised variance estimators. As these estimates depend on the entire sample $Y_n$, they are appropriately called ‘smoothed’ estimates. This is in contrast to simple exponential smoothing point estimates, which are generally considered ‘filtered’ estimates. As simple exponential smoothing assumes ‘known’ values of $\alpha$ and $b_0$, its forecast at time *t* depends only on past data $Y_{t-1}$.

On a related matter, we note that exponential smoothing is so named because of the exponentially declining impact that $y_{t-j}$ has on the forecast of $y_t$. However, the Bayesian analysis presented here produces an entire posterior distribution over each forecast that is a mixture of distributions conditional on given initial values and smoothing constants. In particular, past observations will not in general maintain an exponentially declining effect on future forecasting distributions.

It is worth noting that the above smoothed state estimators, Rao -Blackwellised density and moment estimators need not always be the most efficient means of estimating these quantities in situations where $\alpha$ is of low dimension. For example, averaging the densities of Student-t distributions in determining the posterior forecast distributions, would result in a more efficient estimator. However, it is more appealing from a practical point of view simply to consider averages of multivariate normal densities.

As a final note, frequently in applied time series analysis the data must first be transformed before a model is deemed appropriate. These transformations typically affect the interpretation of parameters. Also, care must be taken to produce forecast distributions on the original data scale. For a detailed discussion of these and related issues, see West and Harrison (1989, pp 359-364).

## 5. ANALYSIS OF CAR PART DEMAND

To demonstrate the methodology, we analyse thirty-one months of demand data for a particular car part using the local level with constant growth model given by

$$y_t = l_{t-1} + g + e_t \tag{40}$$

$$l_t = l_{t-1} + g + \eta e_t \,. \tag{41}$$

The data, from an Australian subsidiary of a Japanese car manufacturer, are given in Table 5.1. This model can be put into the general SSOE state space form by taking $b_t' = (l_t, g_t), b_0' = (l_0, g), x' = (1,1)$, $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, and $\alpha' = (\eta, 0)$. Figure 5.1 shows the original data, four months of forecasts and the 90% posterior prediction intervals along with the actual demand for this part for October 1996-January 1997. The

| | | | | | |
|---|---|---|---|---|---|
| *Mar-94* | 8 | *Mar-95* | 31 | *Mar-96* | 10 |
| *Apr-94* | 4 | *Apr-95* | 16 | *Apr-96* | 19 |
| *May-94* | 8 | *May-95* | 25 | *May-96* | 10 |
| *Jun-94* | 5 | *Jun-95* | 12 | *Jun-96* | 21 |
| *Jul-94* | 6 | *Jul-95* | 12 | *Jul-96* | 16 |
| *Aug-94* | 9 | *Aug-95* | 12 | *Aug-96* | 18 |
| *Sep-94* | 11 | *Sep-95* | 18 | *Sep-96* | 27 |
| *Oct-94* | 7 | *Oct-95* | 18 | *Oct-96* | 23 |
| *Nov-94* | 12 | *Nov-95* | 14 | *Nov-96* | 24 |
| *Dec-94* | 7 | *Dec-95* | 16 | *Dec-96* | 16 |
| *Jan-95* | 16 | *Jan-96* | 14 | *Jan-97* | 24 |
| *Feb-95* | 11 | *Feb-96* | 24 | | |

**Table 5.1** Monthly car part demand data

forecast distributions are calculated using only data from March 1994 through September 1996, and the point forecasts given are posterior means. The 90% intervals are calculated using the lower 5% and 95% quantiles of the relevant forecast distribution. In addition, the estimated smoothed series, calculated using (39), is also shown in Figure 5.1. In this case, the value of $x'b_{t-1}$ at time *t* is given by $l_{t-1} + g$. As can be seen, the smoothed series appears to do a good job of tracking the 'signal' underlying the noisy original data.

The uncertainty in the forecasts, as reflected by the 90% prediction intervals, combines the uncertainty in estimating each of the parameters $l_0, g, \eta, \sigma^2$, along with the uncertainty in future observed $e_{n+f}$ values, for $f$ = 1, 2, 3, 4. These individual forecast distributions are given in Figure 5.2. In addition, marginal density estimates for the initial level, constant growth term, smoothing and variance parameters are given in Figure 5.3. The marginal mode of $\alpha$ = 0.24 suggests that approximately a quarter of the variation in demand is due to unanticipated structural change, the remaining variation having only a transient effect. An advantage of using the Bayesian approach is that posterior interval estimates for these quantities are available, resulting in $-0.01 < g < 1.1$ with approximately 90% probability, and $0 < \alpha < 0.48$ again with approximately 90% probability.

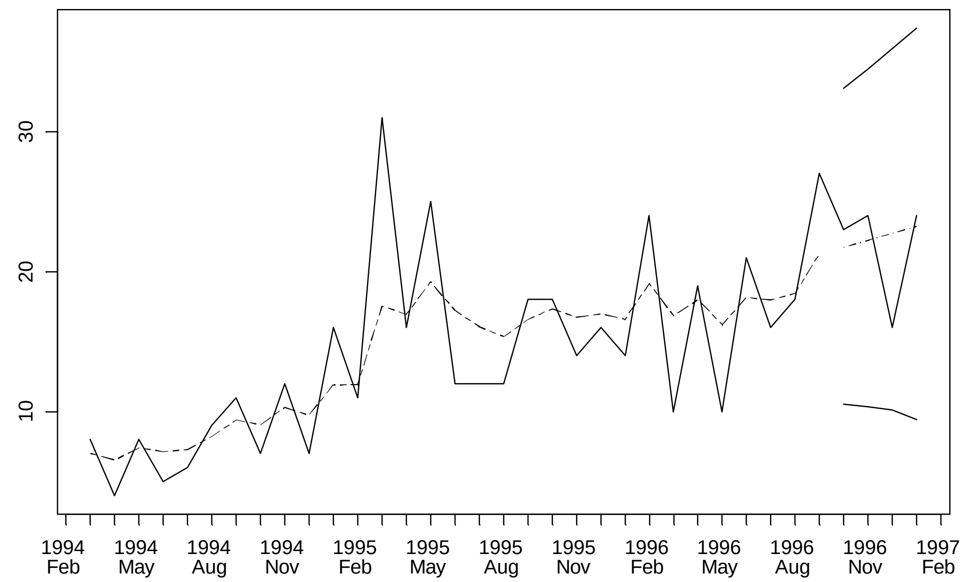


**Figure 5.1** Car part demand data, smoothed series and forecasts along with 90% prediction intervals.

The modes of the forecast distributions in Figure 5.2 demonstrate an expected positive growth $E[g|Y_n] = 0.49$ as their locations gradually increase with the forecast horizon, *f*. The relatively small estimated $\sigma^2$ ($22.5 < \sigma^2 < 57$ with 90% probability) gradually reduces the precision in the forecasts.

## 6. A SIMULATION STUDY OF FORECAST PERFORMANCE

In order to assess the forecasting performance of the proposed Bayesian approach, we designed and implemented a simulation study to determine the actual coverage probability of the 90% prediction intervals. We simulated data series of two lengths,

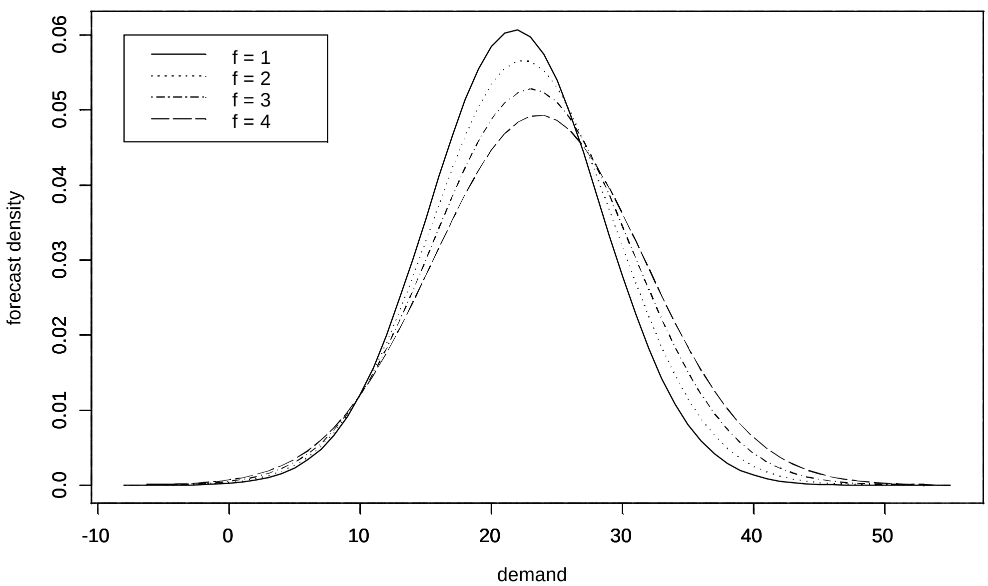


**Figure 5.2** Car part demand forecast distributions

$n$ = 32 and $n$ = 92, for each of the local level model and the local level with constant growth model, under several different parameter settings. Next we computed Bayesian 90% forecast intervals based on 2000 draws from the posterior distribution for forecast horizon $f$ = 1, 2, 3, 4 and $f$ = 1, 2, … 8, respectively, corresponding to the two different sample sizes. We then generated 2000 realisations of the observed data corresponding to the forecast horizon and computed the coverage of the forecast intervals. This process was replicated 100 times to find the distribution of the coverage. The parameter settings chosen for the local level model include all combinations of $l_0 = 100$, $\sigma = (8,\ 16)$ and $\alpha = (0.05,\ 0.2,\ 0.5,\ 0.95)$. For the local level with constant growth model the parameter settings chosen include the same combinations as in the local level model, with $g = 5$. The results of this simulation study are given in Tables 6.1 and 6.2.

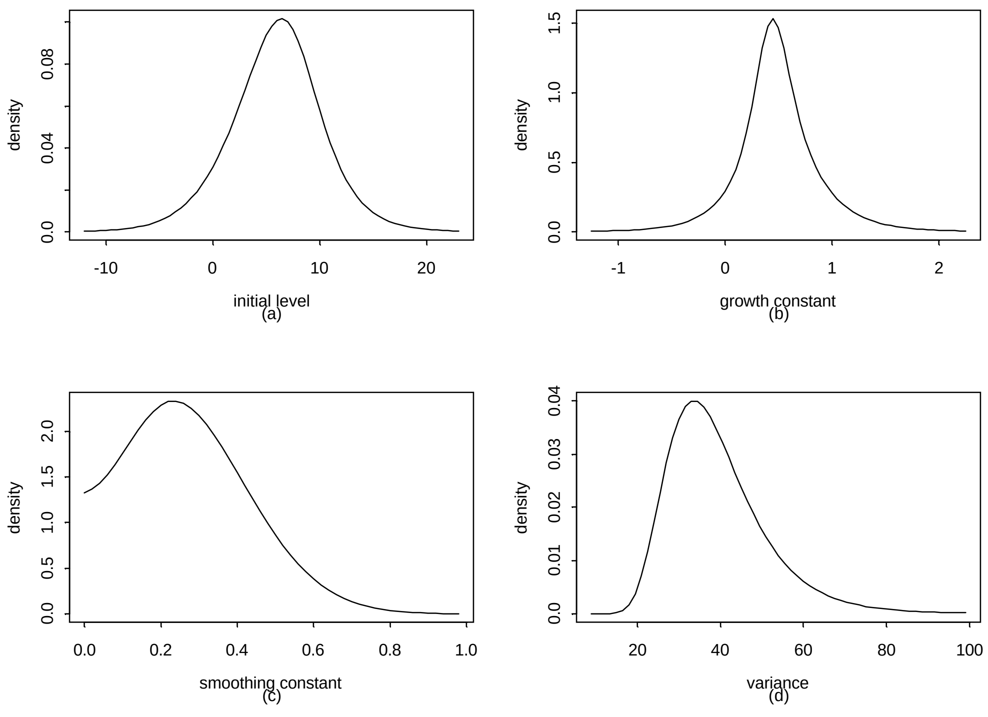


**Figure 5.3** Marginal posterior density estimates of (a) initial level $l_0$, (b) average growth $g$, (c) smoothing parameter $\alpha$, and (d) variance $\sigma^2$ for car part demand data.

The distributions of coverage, while relatively unbiased, are not symmetric, as the discrepancies between the mean and median show. To get a better understanding of these distributions, we have included boxplots of two of the above experiments in Figures 6.1 and 6.2. These results are fairly representative of the others not displayed here, though complete results are available from the authors. In general, the coverage distribution tends to be very compact for small values of $\alpha$, and more dispersed for values of $\alpha$ near unity. As expected, a larger sample size significantly tightens the coverage distributions; and, at least for data generated from the models studied here, the coverage distributions appear to do very well for short horizons, while performance declines for larger horizons.

| | | | Local Level Model | | | | Local Level with Drift Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $f = 1$ | $f = 2$ | $f = 3$ | $f = 4$ | $f = 1$ | $f = 2$ | $f = 3$ | $f = 4$ |
| $\sigma = 8$ | $\alpha = 0.05$ | Mean | .90 | .91 | .92 | .92 | .91 | .92 | .93 | .94 |
| | | Median | .91 | .92 | .92 | .93 | .92 | .94 | .95 | .96 |
| | $\alpha = 0.2$ | Mean | .90 | .91 | .91 | .92 | .90 | .91 | .92 | .93 |
| | | Median | .90 | .92 | .92 | .93 | .91 | .93 | .93 | .94 |
| | $\alpha = 0.5$ | Mean | .90 | .90 | .91 | .90 | .90 | .91 | .90 | .90 |
| | | Median | .91 | .92 | .92 | .92 | .91 | .92 | .92 | .92 |
| | $\alpha = 0.95$ | Mean | .91 | .88 | .87 | .87 | .89 | .86 | .85 | .84 |
| | | Median | .92 | .89 | .88 | .88 | .90 | .88 | .87 | .86 |
| $\sigma = 16$ | $\alpha = 0.05$ | Mean | .90 | .91 | .92 | .92 | .90 | .92 | .92 | .94 |
| | | Median | .91 | .92 | .93 | .93 | .91 | .92 | .94 | .94 |
| | $\alpha = 0.2$ | Mean | .91 | .92 | .93 | .93 | .92 | .92 | .93 | .93 |
| | | Median | .92 | .93 | .94 | .94 | .92 | .93 | .94 | .94 |
| | $\alpha = 0.5$ | Mean | .90 | .91 | .91 | .91 | .89 | .89 | .88 | .88 |
| | | Median | .91 | .91 | .92 | .93 | .90 | .90 | .88 | .89 |
| | $\alpha = 0.95$ | Mean | .90 | .88 | .87 | .87 | .89 | .87 | .86 | .85 |
| | | Median | .92 | .89 | .88 | .88 | .90 | .89 | .88 | .87 |

**Table 6.1** Simulation mean and medians of coverage for Bayesian 90% prediction intervals. Based on 100 replications of coverage from 2000 forecasts drawn from the local level model ($l_0$=100) and local level with drift model ($l_0$=100, $g$ = 5) and sample size $n$ = 32.

## 7. CONCLUSION

Bayesian methods applied to MSOE state space models have provided a useful approach to analysing and responding to unanticipated structural change in business. However, these methods rely on the use of the Kalman filter, which is frequently not well understood by business analysts. In this paper we have discussed the SSOE state space model and its links with exponential smoothing methods. We have developed and tested a Bayesian estimation and prediction methodology for the general SSOE state space model and demonstrated our approach using the local level with constant growth model on car part demand data. Estimates of the unobserved state vector and a smoothed value of the data series are readily available from the approach. Results from a simulation study indicate that the prediction intervals produced from the method have good coverage properties, at least for short horizons and for data generated under the assumed model.

| | | | Local Level Model | | | | | | | | Local Level with Drift Model | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $f = 1$ | $f = 2$ | $f = 3$ | $f = 4$ | $f = 5$ | $f = 6$ | $f = 7$ | $f = 8$ | $f = 1$ | $f = 2$ | $f = 3$ | $f = 4$ | $f = 5$ | $f = 6$ | $f = 7$ | $f = 8$ |
| $\sigma = 8$ | $\alpha = 0.05$ | Mean | .90 | .90 | .91 | .91 | .91 | .91 | .91 | .91 | .90 | .91 | .91 | .91 | .92 | .92 | .92 | .92 |
| | | Median | .91 | .90 | .91 | .91 | .91 | .92 | .91 | .92 | .91 | .91 | .91 | .92 | .92 | .92 | .92 | .92 |
| | $\alpha = 0.2$ | Mean | .90 | .90 | .90 | .90 | .90 | .91 | .90 | .91 | .90 | .91 | .91 | .91 | .91 | .91 | .91 | .91 |
| | | Median | .90 | .91 | .90 | .91 | .91 | .92 | .92 | .92 | .91 | .91 | .91 | .91 | .92 | .91 | .92 | .92 |
| | $\alpha = 0.5$ | Mean | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 |
| | | Median | .90 | .90 | .91 | .90 | .90 | .90 | .90 | .90 | .91 | .90 | .90 | .91 | .91 | .91 | .91 | .91 |
| | $\alpha = 0.95$ | Mean | .90 | .89 | .89 | .88 | .88 | .88 | .88 | .88 | .90 | .89 | .88 | .88 | .88 | .88 | .87 | .87 |
| | | Median | .90 | .89 | .89 | .89 | .88 | .89 | .88 | .88 | .91 | .89 | .89 | .89 | .88 | .88 | .89 | .89 |
| $\sigma = 16$ | $\alpha = 0.05$ | Mean | .90 | .90 | .90 | .90 | .91 | .90 | .91 | .91 | .90 | .90 | .90 | .91 | .91 | .91 | .91 | .91 |
| | | Median | .90 | .90 | .90 | .91 | .91 | .90 | .91 | .91 | .90 | .90 | .91 | .91 | .91 | .91 | .92 | .92 |
| | $\alpha = 0.2$ | Mean | .89 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .91 | .90 | .91 | .91 | .91 | .91 | .91 |
| | | Median | .89 | .90 | .90 | .91 | .90 | .91 | .91 | .91 | .90 | .91 | .91 | .91 | .91 | .91 | .91 | .91 |
| | $\alpha = 0.5$ | Mean | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .90 | .91 | .90 | .90 | .90 |
| | | Median | .90 | .90 | .91 | .90 | .91 | .90 | .91 | .91 | .90 | .91 | .91 | .91 | .92 | .92 | .92 | .92 |
| | $\alpha = 0.95$ | Mean | .90 | .89 | .88 | .88 | .88 | .88 | .88 | .88 | .90 | .89 | .88 | .88 | .88 | .88 | .87 | .87 |
| | | Median | .91 | .90 | .89 | .89 | .89 | .89 | .88 | .88 | .90 | .89 | .88 | .88 | .88 | .88 | .88 | .88 |

**Table 6.2** Simulation mean and medians of coverage for Bayesian 90% prediction intervals. Based on 100 replications of coverage from 2000 forecasts drawn from local level model ($l_0$=100) and local level with drift model ($l_0$=100, $g$ = 5) and sample size $n$ = 92.

The SSOE model is preferable to the MSOE because it is both simpler to analyse and simpler to interpret. Non-Bayesian forecasting methods for the SSOE model have been available from earlier research, but our proposed approach provides the only available exact small sample solution. In addition, the Bayesian approach provides the opportunity to incorporate subjective information and expert knowledge into the process of estimation and prediction. Because of the link with exponential smoothing, prior information on the proportion of variability due to unanticipated structural change can easily be incorporated via the so-called 'smoothing parameter'.

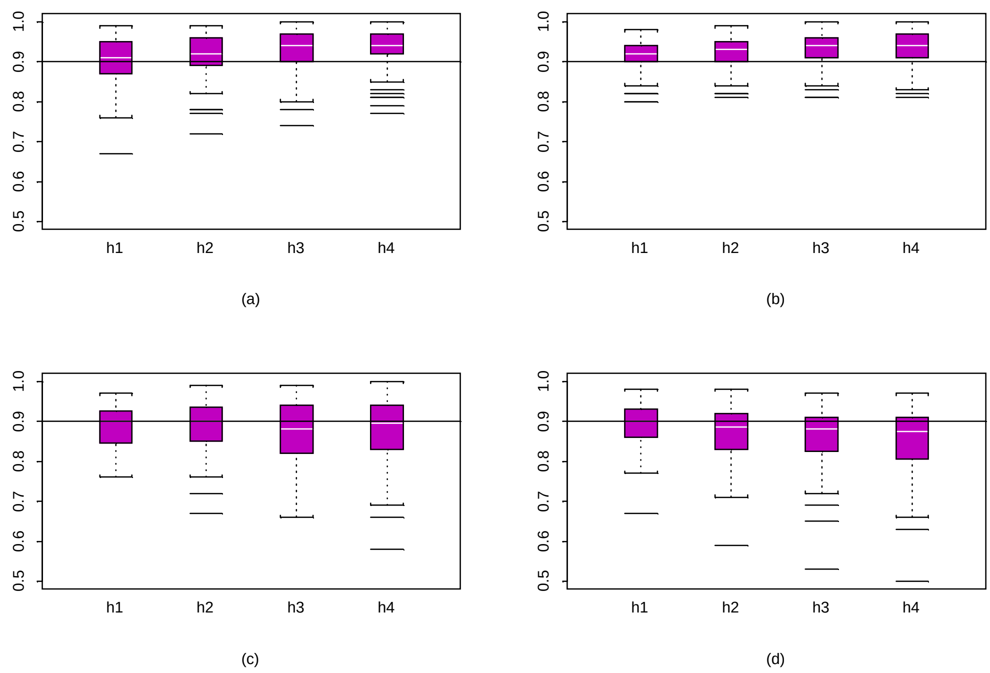


**Figure 6.1** Boxplots for simulated coverage probabilities for Bayesian 90% prediction intervals. Based on 100 replications of 2000 forecasts drawn from local level with constant growth model with sample size $n$ = 32, initial level $l_0$=100, constant growth coefficient $g$ = 5, error variance $\sigma^2$=64 and smoothing constant $\alpha$ equal to (a) 0.05, (b) 0.2, (c) 0.5 and (d) 0.95.

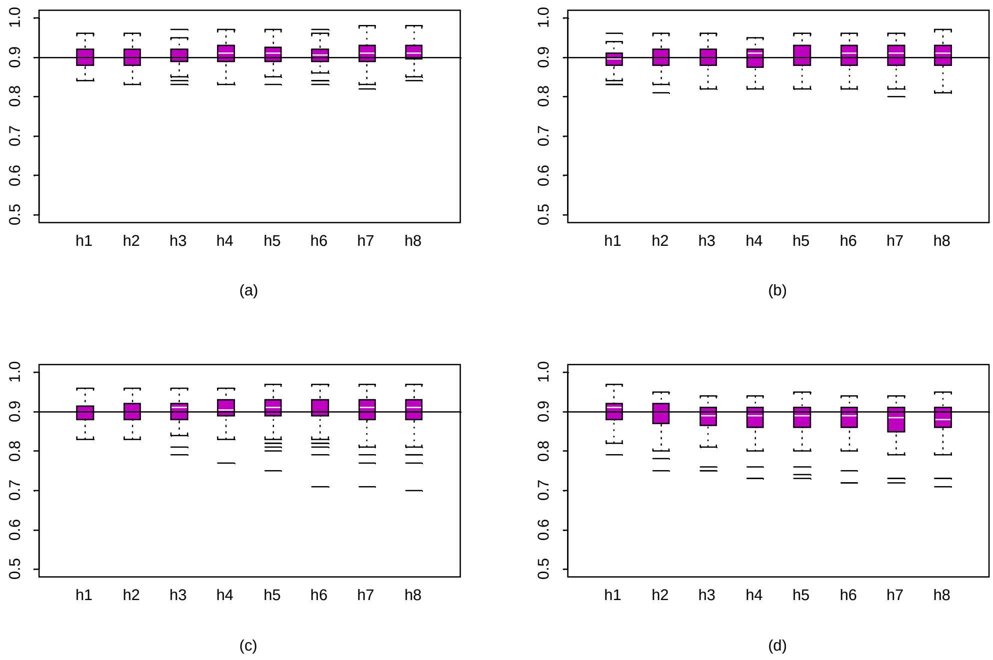


**Figure 6.2** Boxplots for simulated coverage probabilities for Bayesian 90% prediction intervals. Based on 100 replications of 2000 forecasts drawn from the local level model with sample size $n$ = 92, initial level $l_0$=100, error variance $\sigma^2$=64 and smoothing constant $\alpha$ equal to (a) 0.05, (b) 0.2, (c) 0.5 and (d) 0.95.

## REFERENCES

Aoki, M. and Havenner, A. (1991), State Space Modelling of Multiple Time Series. *Econometric Reviews*, 10, 1-59.

Box G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis Forecasting and Control* (2nd ed.), San Francisco: Holden-Day.

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

Carlin, B. P., Polsen, N. G. and Stoffer, D. S. (1992), A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, Vol. 87, No. 418, 493-500.

Carter, C. K. and Kohn, R. (1994) On Gibbs Sampling for State Space Models. *Biometrika*, Vol. 81, No. 3, 541-553.

Chatfield, C. (1977), Some Recent Developments in Time Series Analysis. *Journal of the Royal Statistical Society* (A), 140, 492-510.

Gardner, E.S. Jr. (1985), Exponential Smoothing: The State of the Art. *Journal of Forecasting*, V4, 1-28.

Gelfand, A. E., and Smith, A. F. M. (1990), Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85, 398-409.

Geweke, J., and Terui, N. (1993), Bayesian Threshold Autoregressive Models for Nonlinear Time Series. *Journal of Time Series Analysis*, 14, 441-454.

Godolphin, E. and Stone, M. (1980), On the structural representation for polynomial predictor models. *Journal of the Royal Statistical Society*, *Series B*, 42, 35-45.

Harrison, P. J. and Stevens, C. F. (1976), Bayesian Forecasting (with discussion). *Journal of the Royal Statistical Society*, *Ser. B*, 38, 205-247.

Harvey A. C. (1984), A Unified View of Statistical Forecasting Procedures. *Journal of Forecasting*, 3, 245-75.

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

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

Kalman, R. E (1960), A New Approach to Linear Filtering and Prediction Problems". *Journal of Basic Engineering*, 82, 34-45.

Kalman, R. E and Bucy, R. S. (1961), New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, 83, 95-108.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Paerzen, E. and Winkler, R. (1982), The Accuracy of Extrapolative (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, 1, 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. and Simmons, L. F. (1993), The M2-Competition: A Real-Time Judgmentally Based Forecasting Study. *International Journal of Forecasting,* 9 (1), 5-22.

Ord J. K., Koehler, A. B. and Snyder,R. D. (1997), Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association*, 92, 1621-29.

Ray, W. D. (1989), Rates of Convergence to steady state for the linear growth version of a dynamic linear model (DLM). *International Journal of Forecasting*, 5, 537-545.

Snyder R. D. (1985), Recursive Estimation of Dynamic Linear Models. *Journal of the Royal Statistical Society, Ser B*, 47, 272-76.

Snyder, R. D., Koehler, A. B. and Ord, J. K. (1999), Lead Time Demand for Simple Exponential Smoothing: An Adjustment Factor for the Standard Deviation. *Journal of the Operational Research Society*, 50, 1079-1082.

Snyder, R. D., Ord, J. K. and Koehler, A. B. (2001), Prediction Intervals for ARIMA Models. *Journal of Business and Economic Statistics*, (forthcoming).

West, M. and Harrison, J. (1989), *Bayesian Forecasting and Dynamic Models*, 1989, New York: Springer.

Winters P. R. (1960), Forecasting Sales by Exponentially Weighed Moving Averages. *Management Science*, 6, 324-42.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Economics*, New York: Wiley.