

Forecasting based on state space models for exponential smoothing

Rob J Hyndman

29 August 2002

In business, there is a frequent need for fully automatic forecasting that takes into account trend, seasonality and other features of the data without need for human intervention. In supply chain management, for example, forecasts of demand are required on a regular basis for very large numbers of time series, so that inventory levels can be planned to provide an acceptable level of service to customers.

One of the most successful automatic forecasting methods in practice is based on exponential smoothing models. There are a variety of such models, each having the property that forecasts are weighted averages of past observations with recent observations given relatively more weight than older observations. The name “exponential smoothing” reflects the fact that the weights decrease exponentially as the observations get older.

Although the exponential smoothing methods have been around since the 1950s, a modelling framework incorporating stochastic models, likelihood calculation, prediction intervals and procedures for model selection, was not developed until the work of Hyndman, Koehler, Snyder and Grose (2002) — hereafter denoted by HKSG.

Pegels (1969) proposed a taxonomy of exponential smoothing methods where each method has a trend component and a seasonal component. This was later extended by Gardner (1985) and modified by HKSG. The methods are given in the following table.

Trend Component	Seasonal Component		
	N (none)	A (additive)	M (multiplicative)
N (none)	NN	NA	NM
A (additive)	AN	AA	AM
M (multiplicative)	MN	MA	MM
D (damped)	DN	DA	DM

Some of these methods are better known under other names. For example, cell NN describes the simple exponential smoothing (or SES) method, cell AN describes Holt’s linear method.

¹This document is based on excerpts from Chapter 3 of Makridakis, Wheelwright and Hyndman (1998) and Hyndman, Koehler, Snyder and Grose (2002).

The additive Holt-Winters' method is given by cell AA and the multiplicative Holt-Winters' method is given by cell AM. The other cells correspond to less commonly used but analogous methods.

For each of the 12 methods in the above table, there are two possible state space models, one corresponding to a model with additive errors and the other to a model with multiplicative errors. These two models give equivalent point forecasts although different prediction intervals. HKSG proposed an automatic forecasting procedure that tries each of the 24 state space models on a given time series and selects the "best" method.

In this document, we denote the observed time series by Y_1, Y_2, \dots, Y_n . A forecast of Y_{t+h} based on all the data up to time t is denoted by $Y_t(h)$.

1 Point forecasts for the best-known models

In this section, we give a simple introduction to the best known exponential smoothing methods — simple exponential smoothing (NN), Holt's linear method (AN) and Holt-Winters seasonal method (AA and AM).

1.1 Simple exponential smoothing (NN model)

Suppose we wish to forecast the next value of our time series Y_t which is yet to be observed. Our forecast is denoted by $Y_{t-1}(1)$. When the observation Y_t becomes available, the forecast error is found to be $Y_t - Y_{t-1}(1)$. The method of simple exponential forecasting, due to Brown (1959), takes the forecast for the previous period and adjusts it using the forecast error. That is, the forecast for the next period is

$$Y_t(1) = Y_{t-1}(1) + \alpha(Y_t - Y_{t-1}(1)) \quad (1.1)$$

where α is a constant between 0 and 1.

It can be seen that the new forecast is simply the old forecast plus an adjustment for the error that occurred in the last forecast. When α has a value close to 1, the new forecast will include a substantial adjustment for the error in the previous forecast. Conversely, when α is close to 0, the new forecast will include very little adjustment.

Another way of writing (1.1) is

$$Y_t(1) = \alpha Y_t + (1 - \alpha)Y_{t-1}(1). \quad (1.2)$$

The forecast $Y_t(1)$ is based on weighting the most recent observation Y_t with a weight value α and weighting the most recent forecast $Y_{t-1}(1)$ with a weight of $1 - \alpha$.

The implications of exponential smoothing can be better seen if equation (1.2) is expanded by replacing $Y_{t-1}(1)$ with its components as follows:

$$\begin{aligned} Y_t(1) &= \alpha Y_t + (1 - \alpha)[\alpha Y_{t-1} + (1 - \alpha)Y_{t-2}(1)] \\ &= \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + (1 - \alpha)^2 Y_{t-2}(1). \end{aligned}$$

If this substitution process is repeated by replacing $Y_{t-2}(1)$ by its components, $Y_{t-3}(1)$ by its components, and so on, the result is

$$\begin{aligned} Y_t(1) &= \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \alpha(1 - \alpha)^3 Y_{t-3} \\ &\quad + \alpha(1 - \alpha)^4 Y_{t-4} + \alpha(1 - \alpha)^5 Y_{t-5} + \cdots + \alpha(1 - \alpha)^{t-1} Y_1 \\ &\quad + (1 - \alpha)^t Y_0(1). \end{aligned} \tag{1.3}$$

So $Y_t(1)$ represents a weighted moving average of all past observations with the weights decreasing exponentially, hence the name “exponential smoothing”.

For longer range forecasts, it is assumed that the forecast function is “flat.” That is,

$$Y_t(h) = Y_t(1), \quad h = 2, 3, \dots$$

A flat forecast function is used because simple exponential smoothing works best for data which have no trend, no seasonality, or other underlying pattern.

Another way of writing this is to let $\ell_t = Y_t(1)$. Then $Y_t(h) = \ell_t$ and $\ell_t = \alpha Y_t + (1 - \alpha)\ell_{t-1}$. The value of ℓ_t is a measure of the “level” of the series at time t . This notation is used in the remaining paper to define generalizations of SES.

In order to calculate the forecasts using SES, we need to specify the initial value $\ell_0 = Y_0(1)$ and the parameter value α . We describe how this can be done in Section 4.

1.2 Holt’s linear method (AN model)

Holt (1957) extended simple exponential smoothing to linear exponential smoothing to allow forecasting of data with trends. The forecast for Holt’s linear exponential smoothing is found using two smoothing constants, α and β (with values between 0 and 1), and three equations:

$$\text{Level} \quad \ell_t = \alpha Y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \tag{1.4}$$

$$\text{Trend} \quad b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}, \tag{1.5}$$

$$\text{Forecast: } Y_t(h) = \ell_t + b_t h. \quad (1.6)$$

Here ℓ_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the slope of the series at time t .

Holt's method is sometimes called "double exponential smoothing." In the special case where $\alpha = \beta$, the method is equivalent to "Brown's double exponential smoothing."

In Section 4 we describe how the procedure is initialized and how the parameters α and β are estimated.

1.3 Holt-Winters' trend and seasonality method

If the data have no trend or seasonal patterns, then simple exponential smoothing is appropriate. If the data exhibit a linear trend, Holt's linear method is appropriate. But if the data are seasonal, these methods, on their own, cannot handle the problem well.

Holt's method was extended by Winters (1960) to capture seasonality directly. The Holt-Winters' method is based on three smoothing equations—one for the level, one for trend, and one for seasonality. It is similar to Holt's method, with one additional equation to deal with seasonality. In fact there are two different Holt-Winters' methods, depending on whether seasonality is modelled in an additive or multiplicative way.

Multiplicative seasonality (AM method)

The basic equations for Holt-Winters' multiplicative method are as follows:¹

$$\text{Level: } \ell_t = \alpha \frac{Y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (1.7)$$

$$\text{Trend: } b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad (1.8)$$

$$\text{Seasonal: } s_t = \gamma Y_t / (\ell_{t-1} + b_{t-1}) + (1 - \gamma)s_{t-m} \quad (1.9)$$

$$\text{Forecast: } Y_t(h) = (\ell_t + b_t h) s_{t-m+h} \quad (1.10)$$

where m is the length of seasonality (e.g., number of months or quarters in a year), ℓ_t represents the level of the series, b_t denotes the trend, s_t is the seasonal component, and $Y_t(h)$ is the forecast for h periods ahead. As with all exponential smoothing methods, we need initial values of the components and parameter values. This is discussed in Section 4.

¹These are slightly different from the usual Holt-Winters equations such as those in Makridakis, Wheelwright and Hyndman (1998). The modification was proposed by HKSG to make the state space formulation simpler.

Additive seasonality (AA method)

The seasonal component in Holt-Winters' method may also be treated additively, although this is less common. The basic equations for Holt-Winters' additive method are as follows:

$$\text{Level: } \ell_t = \alpha(Y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (1.11)$$

$$\text{Trend: } b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad (1.12)$$

$$\text{Seasonal: } s_t = \gamma(Y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (1.13)$$

$$\text{Forecast: } Y_t(h) = \ell_t + b_t h + s_{t-m+h}. \quad (1.14)$$

The second of these equations is identical to (1.8). The only differences in the other equations are that the seasonal indices are now added and subtracted instead of taking products and ratios.

2 General point forecasting equations

Following Makridakis, Wheelwright and Hyndman (1998), we can write each of the 12 exponential smoothing methods as follows.

$$\ell_t = \alpha P_t + (1 - \alpha)Q_t \quad (2.1)$$

$$b_t = \beta R_t + (\phi - \beta)b_{t-1} \quad (2.2)$$

$$s_t = \gamma T_t + (1 - \gamma)s_{t-m} \quad (2.3)$$

where ℓ_t denotes the series level at time t , b_t denotes the slope at time t , s_t denotes the seasonal component of the series at time t and m denotes the number of seasons in a year; the values of P_t , Q_t , R_t , and T_t vary according to which of the cells the method belongs, and α , β , γ and ϕ are constants.

Table 1 shows the values of P , Q , R , and T and the formulae for computing point forecasts h periods ahead.

The formulas for damped trend are appropriate when there is trend in the time series, but one believes that continuing to use the final estimate for the growth rate at the end of the historical data would lead to unrealistic forecasts. Thus, the equations for damped trend do what the name indicates: dampen the trend as the length of the forecast horizon increases. In Table 1, one can see that the forecast for h -periods-ahead is $Y_t(h) = \ell_t + (1 + \phi + \dots + \phi^{h-1})b_t$. The trend is dampened by a factor of ϕ for each additional future time period.

Trend component	Seasonal component		
	N (none)	A (additive)	M (multiplicative)
N (none)	$P_t = Y_t$ $Q_t = \ell_{t-1}$ $\phi = 1$ $Y_t(h) = \ell_t$	$P_t = Y_t - s_{t-m}$ $Q_t = \ell_{t-1}$ $T_t = Y_t - Q_t$ $\phi = 1$ $Y_t(h) = \ell_t + s_{t+h-m}$	$P_t = Y_t/s_{t-m}$ $Q_t = \ell_{t-1}$ $T_t = Y_t/Q_t$ $\phi = 1$ $Y_t(h) = \ell_t s_{t+h-m}$
A (additive)	$P_t = Y_t$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $\phi = 1$ $Y_t(h) = \ell_t + hb_t$	$P_t = Y_t - s_{t-m}$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t - Q_t$ $\phi = 1$ $Y_t(h) = \ell_t + hb_t + s_{t+h-m}$	$P_t = Y_t/s_{t-m}$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t/Q_t$ $\phi = 1$ $Y_t(h) = (\ell_t + hb_t)s_{t+h-m}$
M (multiplicative)	$P_t = Y_t$ $Q_t = \ell_{t-1}b_{t-1}$ $R_t = \ell_t/\ell_{t-1}$ $\phi = 1$ $Y_t(h) = \ell_t b_t^h$	$P_t = Y_t - s_{t-m}$ $Q_t = \ell_{t-1}b_{t-1}$ $R_t = \ell_t/\ell_{t-1}$ $T_t = Y_t - Q_t$ $\phi = 1$ $Y_t(h) = \ell_t b_t^h + s_{t+h-m}$	$P_t = Y_t/s_{t-m}$ $Q_t = \ell_{t-1}b_{t-1}$ $R_t = \ell_t/\ell_{t-1}$ $T_t = Y_t/Q_t$ $\phi = 1$ $Y_t(h) = \ell_t b_t^h s_{t+h-m}$
D (damped)	$P_t = Y_t$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $\beta < \phi < 1$ $Y_t(h) = \ell_t + (1 + \phi + \dots + \phi^{h-1})b_t$	$P_t = Y_t - s_{t-m}$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t - Q_t$ $\beta < \phi < 1$ $Y_t(h) = \ell_t + (1 + \phi + \dots + \phi^{h-1})b_t + s_{t+h-m}$	$P_t = Y_t/s_{t-m}$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t/Q_t$ $\beta < \phi < 1$ $Y_t(h) = [\ell_t + (1 + \phi + \dots + \phi^{h-1})b_t]s_{t+h-m}$

Table 1: Formulae for recursive calculations and point forecasts.

Writing (2.1)–(2.3) in their error-correction form we obtain

$$\ell_t = Q_t + \alpha(P_t - Q_t) \quad (2.4)$$

$$b_t = \phi b_{t-1} + \beta(R_t - b_{t-1}) \quad (2.5)$$

$$s_t = s_{t-m} + \gamma(T_t - s_{t-m}). \quad (2.6)$$

The method with fixed level (constant over time) is obtained by setting $\alpha = 0$, the method with fixed trend (drift) is obtained by setting $\beta = 0$, and the method with fixed seasonal pattern is obtained by setting $\gamma = 0$. Note also that the additive trend methods are obtained by letting $\phi = 1$ in the damped trend methods.

Trend component	Seasonal component		
	N (none)	A (additive)	M (multiplicative)
N (none)	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A (additive)	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
M (multiplicative)	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t/\ell_{t-1}$	$\mu_t = \ell_{t-1}b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
D (damped)	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \alpha\beta\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \alpha\beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \alpha\beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$

Table 2: State space equations for each additive error model in the classification. Multiplicative error models are obtained by replacing ε_t by $\mu_t\varepsilon_t$ in the above equations.

3 State space models

HKSG describe the state space models that underlie the exponential smoothing methods. For each method, there are two models—a model with additive errors and a model with multiplicative errors. The pointwise forecasts for the two models are identical, but prediction intervals will differ.

The general model involves a state vector $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-(m-1)})$ and state space equations of the form

$$Y_t = \mu_t + k(\mathbf{x}_{t-1})\varepsilon_t \quad (3.1)$$

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t \quad (3.2)$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance σ^2 and $\mu_t = Y_{t-1}(1)$. The model with additive errors has $k(\mathbf{x}_{t-1}) = 1$, so that $Y_t = \mu_t + \varepsilon_t$. The model with multiplicative errors has $k(\mathbf{x}_{t-1}) = \mu_t$, so that $Y_t = \mu_t(1 + \varepsilon_t)$. Thus, $\varepsilon_t = (Y_t - \mu_t)/\mu_t$ is a relative error for the multiplicative model.

All the methods in Table 1 can be written in the form (3.1) and (3.2). The underlying equations are given in Table 2. The models are not unique. Clearly, any value of $k(\mathbf{x}_{t-1})$ will

lead to identical point forecasts for Y_t . The only difference between the additive error and multiplicative error models is in the observation equation (3.1).

Note that not all of the 24 state space models are appropriate for all data. The multiplicative error models are not well defined if there are zeros or negative values in the data. Similarly, we don't consider the additive error models with multiplicative trend or multiplicative seasonality if any observations are zero. Further, if the data are not quarterly or monthly (and do not have some other obvious seasonal period), then we do not consider any of the seasonal methods.

4 Estimation and model selection

Let

$$L^*(\boldsymbol{\theta}, \mathbf{X}_0) = n \log \left(\sum_{t=1}^n e_t^2 / k^2(x_{t-1}) \right) + 2 \sum_{t=1}^n \log |k(x_{t-1})|. \quad (4.1)$$

Then L^* is equal to twice the negative logarithm of the conditional likelihood function of the state space model (with constant terms eliminated).

The parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \phi)$ and initial states $\mathbf{X}_0 = (\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$ can be estimated by minimizing L^* . Alternatively, estimates can be obtained by minimizing the one-step mean-squared-error (MSE), minimizing the one-step mean-absolute-percent-error (MAPE), minimizing the residual variance σ^2 or via some other criterion for measuring forecast error.

HKSG found that the 3-step AMSE (an average of the one-step, two-step and three-step MSEs) provided the best parameter estimates for forecasting.

We constrain the estimation by restricting the parameters to lie within the following intervals

$$0.1 \leq \alpha \leq 0.9, \quad 0.1 \leq \beta \leq 0.9, \quad 0.1 \leq \gamma \leq 0.9, \quad \beta \leq \phi \leq 1.$$

Usually, α , β and γ are restricted to values in $(0, 1)$. However we use a smaller range to avoid instabilities occurring. We also constrain the initial states \mathbf{X}_0 so that the seasonal indices add to zero for additive seasonality, and add to m for multiplicative seasonality.

Models are selected using Akaike's Information Criterion:

$$\text{AIC} = L^*(\hat{\boldsymbol{\theta}}, \hat{\mathbf{X}}_0) + 2p$$

where p is the number of parameters in $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{X}}_0$ denote the estimates of $\boldsymbol{\theta}$ and \mathbf{X}_0 . We select the model that minimizes the AIC amongst all of the 24 models that are appropri-

ate for the data. Using the AIC for model selection is preferable to other measurements of forecast error such as the MSE or MAPE as it penalizes against models containing too many parameters.

The AIC also provides a method for selecting between the additive and multiplicative error models. Point forecasts from the two models are identical so that standard forecast accuracy measures such as MSE or MAPE are unable to select between the error types. The AIC is able to select between the error types because it is based on likelihood rather than one-step forecasts.

Obviously, other model selection criteria (such as the BIC) could also be used in a similar manner.

4.1 Initialization

The non-linear optimization requires some initial values. We use $\alpha = \beta = \gamma = 0.5$ and $\phi = 0.9$. The initial values of ℓ_0 , b_0 and s_k ($k = -m + 1, \dots, 0$) are obtained using the following heuristic scheme.

- For seasonal data, compute a $2 \times m$ moving average through the first few years of data (we use up to four years if the data are available). Denote this by $\{f_t\}$, $t = m/2 + 1, m/2 + 2, \dots$
- For additive seasonality, we detrend the data to obtain $Y_t - f_t$. For multiplicative seasonality, we detrend the data to obtain Y_t / f_t . Then compute initial seasonal indices, s_{-m+1}, \dots, s_0 , by averaging the detrended data for each season over the first 3 years available (from $t = m/2 + 1$ to $t = 7m/2$). We normalize these seasonal indices so they add to zero for additive seasonality, and add to m for multiplicative seasonality.
- For seasonal data, compute a linear trend using OLS regression on the first 10 seasonally adjusted values (using the seasonal indices obtained above) against a time variable $t = 1, \dots, 10$.
- For non-seasonal data, compute a linear trend on the first 10 observations against a time variable $t = 1, \dots, 10$.
- Then set ℓ_0 to be the intercept of the trend.
- For additive trend, set b_0 to be the slope of the trend.
- For multiplicative trend, set $b_0 = 1 + b/a$ where a denotes the intercept and b denotes the slope of the fitted trend.

These heuristic values of the initial state \mathbf{X}_0 are then refined by estimating them as parameters along with the elements of θ .

4.2 Automatic forecasting

We combine the preceding ideas to obtain a robust and widely applicable automatic forecasting algorithm. The steps involved are summarized below.

- For each series, we apply the models that are appropriate, optimizing the parameters of the model in each case.
- We select the best of the models according to the AIC.
- We produce forecasts using the best model (with optimized parameters) for as many steps ahead as required.
- To obtain prediction intervals, we use a bootstrap method by simulating 5000 future sample paths for $\{Y_{n+1}, \dots, Y_{n+h}\}$ and finding the $\alpha/2$ and $1 - \alpha/2$ percentiles of the simulated data at each forecasting horizon. The sample paths are generated using the normal distribution for errors (parametric bootstrap) or using the resampled errors (ordinary bootstrap).

With the linear state space models, it is possible to compute analytical prediction intervals (see Hyndman, Koehler, Ord and Snyder, 2001). However, this is difficult with the non-linear models, and so we prefer to use a simulation approach which is applicable to all the models and is very easy to implement.

Application of the automatic forecasting strategy to the M-competition data and IJF-M3 competition data has demonstrated that the methodology is particularly good at short term forecasts (up to about 6 periods ahead) and especially for seasonal short-term series (beating all other methods in the competition for these series). This was achieved without any preprocessing of the data, identification of outliers or level shifts, or other strategy designed to improve the forecasts. It is expected that these results could be improved further if some sophisticated data preprocessing techniques was used (as was done by some of the competitors in the M3 competition).

5 References

- BROWN, R.G. (1959) *Statistical forecasting for inventory control*, New York: McGraw-Hill.
- GARDNER, E.S. (1985) Exponential smoothing: the state of the art, *Journal of Forecasting*, **4**, 1–28.
- HOLT, C.C. (1957) Forecasting seasonal and trends by exponentially weighted moving averages, Office of Naval Research, Research Memorandum No. 52.
- HYNDMAN, R.J., KOEHLER, A.B., ORD, J.K., and SNYDER, R.D. (2001) Prediction intervals for exponential smoothing state space models. Working paper 11/2001, Department of Econometrics & Business Statistics, Monash University.
- HYNDMAN, R.J., KOEHLER, A.B., SNYDER, R.D., GROSE, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International J. Forecasting*, **18**(3), 439–454.
- MAKRIDAKIS, S., WHEELWRIGHT, S.C., and HYNDMAN, R.J. (1998) *Forecasting: methods and applications*, 3rd ed., John Wiley & Sons: New York.
- PEGELS, C.C. (1969) Exponential forecasting: some new variations, *Management Science*, **12**, No. 5, 311–315.
- WINTERS, P.R. (1960) Forecasting sales by exponentially weighted moving averages, *Management Science*, **6**, 324–342.