Journal of Forecasting
J. Forecast. **24**, 17–37 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/for.938

Prediction Intervals for Exponential Smoothing Using Two New Classes of State Space Models

ROB J. HYNDMAN,[1]* ANNE B. KOEHLER,[2] J. KEITH ORD[3] AND RALPH D. SNYDER[1]
[1] *Monash University, Australia*
[2] *Miami University, USA*
[3] *Georgetown University, USA*

ABSTRACT
Three general classes of state space models are presented, using the single source of error formulation. The first class is the standard linear model with homoscedastic errors, the second retains the linear structure but incorporates a dynamic form of heteroscedasticity, and the third allows for non-linear structure in the observation equation as well as heteroscedasticity. These three classes provide stochastic models for a wide variety of exponential smoothing methods. We use these classes to provide exact analytic (matrix) expressions for forecast error variances that can be used to construct prediction intervals one or multiple steps ahead. These formulas are reduced to non-matrix expressions for 15 state space models that underlie the most common exponential smoothing methods. We discuss relationships between our expressions and previous suggestions for finding forecast error variances and prediction intervals for exponential smoothing methods. Simpler approximations are developed for the more complex schemes and their validity examined. The paper concludes with a numerical example using a non-linear model. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS forecast distribution; forecast interval; forecast error variance; Holt–Winters method; structural models

INTRODUCTION

Exponential smoothing methods were given a firm statistical foundation by the use of state space models with a single source of error (Ord *et al*., 1997). One of the important contributions following from that work is the ability to provide a sound statistical basis for finding prediction intervals for all the exponential smoothing methods. Traditionally, prediction intervals for the exponential smoothing methods have been found through heuristic approaches or by employing equivalent or approximate ARIMA models.

* Correspondence to: Rob J. Hyndman. Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia. E-mail: rob.hyndman@buseco.monash.edu.au

Copyright © 2005 John Wiley & Sons, Ltd.

In this paper we consider three forms of state space models, using the single source of error (SSOE) formulation introduced by Snyder (1985) and used in later work by Ord *et al*. (1997) and Hyndman *et al*. (2001), among others. The first form is the standard linear scheme for which equivalent ARIMA models exist. The second and third versions both involve a specific non-linear structure whereby the state variables and random errors combine in multiplicative fashion rather than additively. Neither of these classes of model has an ARIMA equivalent, although the second class may be formulated as a kind of GARCH model (Bollerslev, 1986). The third class is not covered by either ARIMA or GARCH structures, but is important as a stochastic description of non-linear forecasting schemes such as the Holt–Winters multiplicative model (cf. Makridakis *et al*., 1998, pp. 161–169). For each of the three schemes, we develop exact expressions for the expected value of multi-step-ahead forecasts and the corresponding variances of the forecast errors (henceforth called 'forecast variances'). In addition, we develop approximations for the more complex models and examine conditions under which these approximations are acceptable.

Given these general results, we are able to provide analytic expressions to be used in computing prediction intervals for many types of exponential smoothing, including all of the widely used methods. In contrast, Ord *et al*. (1997) found prediction intervals by using the model to simulate the entire prediction distributions for each future time period. While simulating prediction intervals may be an excellent method for producing them, many forecasters may prefer analytical formulas for their forecasting software, as they are generally easier to apply and faster to compute. Hyndman *et al*. (2001) describe a framework of 24 models for exponential smoothing, including all of the usual methods as well as some extensions. The procedures in that paper also use simulation to produce prediction intervals for the models. We will provide analytical expressions for the forecast variances for some of those 24 models.

Where an equivalent ARIMA model exists (such as for simple exponential smoothing, and Holt's linear method), our results provide identical forecast variances to those from the ARIMA model.

Previous literature on prediction intervals

State space models with multiple sources of error have also been used to find forecast variances for the simple and Holt exponential smoothing methods (Johnston and Harrison, 1986). In these cases the variances are limiting values in models where the convergence is rapid. The variance formulas in these two cases are the same as in our results.

Prediction intervals for the additive Holt–Winters method and the multiplicative Holt–Winters method have previously been considered by Chatfield and Yar. For the additive Holt–Winters method they found an exact formula for the forecast variance that can be computed directly from the form of the smoothing method (Yar and Chatfield, 1990). For the multiplicative Holt–Winters method, they provided an approximate formula (Chatfield and Yar, 1991). In both papers they assumed that the one-period-ahead forecast errors are independent, but they did not assume any particular underlying model for the smoothing methods.

Using a single source of error state space model, Koehler *et al*. (2001) derived an approximate formula for the forecast variance for the multiplicative Holt–Winters method. Their formula differs from that of Chatfield and Yar (1991) only in how the standard deviation of the one-step-ahead forecast error is estimated. The variance formulas were given only for the first year of forecasts in both of these papers (Chatfield and Yar, 1991; Koehler *et al*., 2001).

The results in this current paper include finding both an exact formula (ignoring the estimation error for the smoothing parameters and initial state) for the forecast variance in all future time periods for the multiplicative Holt–Winters method and a better approximation to this exact formula. Another

Copyright © 2005 John Wiley & Sons, Ltd.

point of difference in our work is that Yar and Chatfield (1990) assumed that the variance of the one-period-ahead forecast error is constant for the additive Holt–Winters method. We include a class of models where this forecast variance is not constant but instead changes with the mean of the time series.

Structure of the paper

In the next section we present the main results of the paper. We use the classification of exponential smoothing methods from Hyndman *et al*. (2001) and show the relationship to three general classes of state space models for exponential smoothing. We present formulas for the h-period-ahead means (i.e., forecasts) and forecast variances for 15 specific exponential smoothing models that correspond to nine exponential smoothing methods, including the most widely used ones.

Then, we examine each of the three general classes of models more closely. We provide general matrix formulas for the means and variances and then specialize these formulas to non-matrix expressions for specific exponential models. Proofs for these results are provided in the appendices. For the Class 3 models, the non-matrix expression is an approximation. Thus, we devote a section to the accuracy of this approximation. We discuss the computation of prediction intervals using our models, and give a numerical example using the multiplicative Holt–Winters method. In this example, we compare our exact forecast variances with approximations and compare prediction intervals obtained by using our exact expression with those obtained by simulating complete prediction distributions.

We conclude with some general comments and a summary.

THE MAIN RESULTS

We describe the exponential smoothing methods using a similar framework to that proposed in Hyndman *et al*. (2001). Each method is denoted by two letters: the first letter denotes the type of trend (none, additive or damped) and the second letter denotes the type of seasonality (none, additive or multiplicative):

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

Cell NN describes the simple exponential smoothing method, cell AN describes Holt's linear method. 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.

Hyndman *et al*. (2001) proposed two state space models for each of these methods: one with additive errors and one with multiplicative errors. To distinguish these models, we will add a third letter (A or M) before the letters denoting the type of trend and seasonality. For example, MAN refers to a model with multiplicative errors, additive trend and no seasonality.

Copyright © 2005 John Wiley & Sons, Ltd.

Let $Y_1, \ldots, Y_n$ denote the time series of interest. We consider three classes of SSOE state space models defined as follows:

Class 1 $Y_t = H\boldsymbol{x}_{t-1} + \varepsilon_t$
$x_t = F\boldsymbol{x}_{t-1} + G\varepsilon_t$

Class 2 $Y_t = H\boldsymbol{x}_{t-1}(1 + \varepsilon_t)$
$x_t = (F + G\varepsilon_t)\boldsymbol{x}_{t-1}$

Class 3 $Y_t = H_1\boldsymbol{x}_{t-1}H_2\boldsymbol{z}_{t-1}(1 + \varepsilon_t)$
$x_t = (F_1 + G_1\varepsilon_t)\boldsymbol{x}_{t-1}$
$z_t = (F_2 + G_2\varepsilon_t)\boldsymbol{z}_{t-1}$

where F, G, H, F_1, F_2, G_1, G_2, H_1, H_2 are all matrix coefficients, and $\boldsymbol{x}_t$ and z_t are unobserved state vectors at time t. In each case, $\{\varepsilon_t\}$ is i.i.d. $\mathrm{N}(0, \sigma^2)$ where the lower tail of the distribution is truncated for Classes 2 and 3 so that $1 + \varepsilon_t$ is positive. The truncation is usually negligible as σ is usually relatively small for these models. Let p be the length of vector $\boldsymbol{x}_t$ and q be the length of vector z_t. Then the orders of the above matrices are as follows:

Class 1	$F\ (p \times p)$	$G\ (p \times 1)$	$H\ (1 \times p)$
Class 2	$F\ (p \times p)$	$G\ (p \times p)$	$H\ (1 \times p)$
Class 3	$F_1\ (p \times p)$	$G_1\ (p \times p)$	$H_1\ (1 \times p)$
	$F_2\ (q \times q)$	$G_2\ (q \times q)$	$H_2\ (1 \times q)$

Fifteen of the 18 models described above fall within the three state space model classes above:

Class 1	ANN	AAN	ADN	ANA	AAA	ADA
Class 2	MNN	MAN	MDN	MNA	MAA	MDA
Class 3	MNM	MAM	MDM			

The remaining three models (ANM, AAM and ADM) do not fit within one of these three classes, and will not be considered further in this paper. Hyndman *et al.* (2001) also consider six additional models with multiplicative trend that fall outside the three state space model classes defined above. Note that the above 15 models include two models for simple exponential smoothing, two models for Holt's method, two models for the additive Holt–Winters method and one model for the multiplicative Holt–Winters method.

Equations for the 15 models above are given in Table I using the same notation as in Hyndman *et al.* (2001). As in that paper, we use the SSOE model in our developments. That is, all the observation and state variables are driven by the single error sequence ε_t. For development of this approach, see Snyder (1985) and Ord *et al.* (1997). The variables ℓ_t, b_t and s_t are elements of the state vector and denote the level, slope and seasonal components, respectively; the parameters α, β and γ are the usual smoothing parameters corresponding to the level equation, trend equation and seasonal equation; ϕ is a damping coefficient used for the damped trend models; and m denotes the number of periods in the seasonal cycle.

We derive the forecast means and variances for each of the three model classes, and specifically for each of the 15 models. The point forecast of Y_{n+h} made h steps ahead from the forecast origin n is an estimate of the mean of Y_{n+h} given $\boldsymbol{x}_n$. This mean will be called the 'forecast mean' and denoted

Copyright © 2005 John Wiley & Sons, Ltd.

Table I. Equations defining each of the 15 models

Class 1			
ANN	$Y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	ANA	$Y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$
AAN	$Y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t$	AAA	$Y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \alpha\beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$
ADN	$Y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \alpha\beta\varepsilon_t$	ADA	$Y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \alpha\beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$
Class 2			
MNN	$Y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	MNA	$Y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$
MAN	$Y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \alpha\beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	MAA	$Y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \alpha\beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$
MDN	$Y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \alpha\beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	MDA	$Y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \alpha\beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$
Class 3			
MNM	$Y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$		
MAM	$Y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \alpha\beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$		
MDM	$Y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \alpha\beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$		

by $\mu_h = \mathrm{E}(Y_{n+h} \mid \boldsymbol{x}_n)$. The corresponding forecast variance is given by $v_h = \mathrm{var}(Y_{n+h} \mid \boldsymbol{x}_n)$. Note that we condition on the final state to compute these attributes. The final state is known because we fix the initial state, and because the error in the state equation is observed up to time n. Even if the initial state is assumed unknown, its effect will be negligible for large n and invertible values of the parameters.

The main results are summarized in Tables II and III.

Copyright © 2005 John Wiley & Sons, Ltd.

Table II. h-period-ahead forecast means and variances. Here $\lfloor u \rfloor$ denotes the largest integer less than or equal to u and m denotes the period of seasonality. For Class 3, the expression is exact when $h \le m$ but only approximate for $h > m$

	Mean	Variance
Class 1	μ_h	$v_1 = \sigma^2$ and $v_h = \sigma^2\left(1+\sum_{j=1}^{h-1} c_j^2\right)$
Class 2	μ_h	$v_h = (1+\sigma^2)\theta_h - \mu_h^2$
		where $\theta_1 = \mu_1^2$ and $\theta_h = \mu_h^2 + \sigma^2 \sum_{j=1}^{h-1} c_j^2 \theta_{h-j}$
Class 3	μ_h	$v_h = s_{n-m+h}^2\left[\theta_h(1+\sigma^2)(1+\gamma^2\sigma^2)^k - \tilde{\mu}_h^2\right]$
		where $\theta_1 = \tilde{\mu}_1^2,\quad \theta_h = \tilde{\mu}_h^2 + \sigma^2 \sum_{j=1}^{h-1} c_j^2 \theta_{h-j}$
		and $k = \lfloor (h-1)/m \rfloor$

Table III. Values of μ_h, $\tilde{\mu}_h$ and c_j for the 15 models. Here $\phi_j = 1 + \phi + \cdots + \phi^j = (1 - \phi^{j+1})/(1 - \phi)$, $d_{j,m} = 1$ if $j = m \pmod m$ and 0 otherwise, and $j^* = j \pmod m$

	μ_h	$\tilde{\mu}_h$	c_j
Class 1/Class 2			
ANN/MNN	ℓ_n		α
AAN/MAN	$\ell_n + hb_n$		$\alpha(1 + j\beta)$
ADN/MDN	$\ell_n + \phi_{h-1}b_n$		$\alpha(1 + \phi_{j-1}\beta)$
ANA/MNA	$\ell_n + s_{n-m+1+(h-1)*}$		$\alpha + \gamma d_{j,m}$
AAA/MAA	$\ell_n + hb_n + s_{n-m+1+(h-1)*}$		$\alpha(1 + j\beta) + \gamma d_{j,m}$
ADA/MDA	$\ell_n + \phi_{h-1}s_{n-m+1+(h-1)*}$		$\alpha(1 + \phi_{j-1}\beta) + \gamma d_{j,m}$
Class 3			
MNM	$\ell_n s_{n-m+1+(h-1)*}$	ℓ_n	α
MAM	$(\ell_n + hb_n)s_{n-m+1+(h-1)*}$	$\ell_n + hb_n$	$\alpha(1 + j\beta)$
MDM	$(\ell_n + \phi h_{-1}b_n)s_{n-m+1+(h-1)*}$	$\ell_n + \phi_{h-1}b_n$	$\alpha(1 + \phi_{j-1}\beta)$

Criteria such as maximum likelihood for selection of optimal estimates for the parameters can be found in Hyndman *et al*. (2001) and Ord *et al*. (1997). It is important to notice that estimates for σ^2 are not done in the same manner for all three classes. The estimate for σ^2 would be

$$\hat{\sigma}^2 = \sum_{t=1}^{n} \hat{\varepsilon}_t^2 / n$$

Copyright © 2005 John Wiley & Sons, Ltd.

where

$$\hat{\varepsilon}_t = \begin{cases} Y_t - \hat{Y}_t & \text{for Class 1} \\ (Y_t - \hat{Y}_t)/\hat{Y}_t & \text{for Classes 2 and 3} \end{cases}$$

and

$$\hat{Y}_t = E(Y_t \mid \boldsymbol{x}_{t-1}) = \begin{cases} H\boldsymbol{x}_{t-1} & \text{for Classes 1 and 2} \\ H_1\boldsymbol{x}_{t-1}H_2\boldsymbol{z}_{t-1} & \text{for Class 3} \end{cases}$$

For the special cases in Table II, $\hat{Y}_{n+1} = \mu_1$.

More detail concerning the results for each class are given in the following sections. Derivations of these results are given in the appendices.

EXPECTED VALUES AND FORECAST VARIANCES

Class 1

Derivations are given in Appendix A.

In this case, the general results for the mean and variance are

$$\mu_h = HF^{h-1}\boldsymbol{x}_n, \tag{1}$$

$$v_1 = \sigma^2 \tag{2}$$

and

$$v_h = \sigma^2\left[1 + \sum_{j=1}^{h-1} c_j^2\right], \quad h \geq 2 \tag{3}$$

where $c_j = HF^{j-1}G$. Specific values for μ_h and c_j for the particular models in Class 1 are given in Tables II and III.

Note that forecast means from ANN are equivalent to the usual point forecasts obtained by simple exponential smoothing (SES) and AAN gives forecasts equivalent to Holt's method. SES with drift is obtained from AAN by setting $\beta = 0$ so that $b_n = b$ for all n. The additive Holt–Winters method is equivalent to the forecast means from AAA. Furthermore, ANN is equivalent to an ARIMA(0, 1, 1) model where $\theta = 1 - \alpha$ and AAN is equivalent to an ARIMA(0, 2, 2) model.

The expressions for v_h can be simplified as shown below:

ANN $\quad v_h = \sigma^2[1 + \alpha^2(h-1)]$

AAN $\quad v_h = \sigma^2\left[1 + \alpha^2(h-1)\left\{1 + \beta h + \frac{1}{6}\beta^2 h(2h-1)\right\}\right]$

ADN $\quad v_h = \sigma^2\left[1 + \alpha^2(h-1) + \alpha^2\beta\sum_{j=1}^{h-1}\phi_{j-1}(2 + \phi_{j-1}\beta)\right]$

ANA $\quad v_h = \sigma^2[1+\alpha^2(h-1)+\gamma(2\alpha+\gamma)\lfloor(h-1)/m\rfloor]$

AAA $\quad v_h = \sigma^2\left[1+\alpha^2(h-1)\left\{1+\beta h+\frac{1}{6}\beta^2 h(2h-1)\right\}+\gamma k\{\gamma+\alpha[2+\beta m(k+1)]\}\right]$

where $k = \lfloor(h-1)/m\rfloor$

ADA $\quad v_h = \sigma^2\left[1+\sum_{j=1}^{h-1}\left\{\alpha^2(1+\phi_{j-1}\beta)^2+\gamma d_{j,m}[\gamma+2\alpha(1+\phi_{j-1}\beta)]\right\}\right]$

Class 2

Derivations are given in Appendix B.

In this case, the general result for the forecast mean is the same as for Model 1, namely

$$\mu_h = HF^{h-1}\boldsymbol{x}_n \tag{4}$$

The forecast variance is given by

$$v_h = HV_{h-1}H'(1+\sigma^2)+\sigma^2\mu_h^2 \tag{5}$$

where

$$V_h = FV_{h-1}F'+\sigma^2 GV_{h-1}G'+\sigma^2 P_{h-1}, \quad h=1,2,\ldots \tag{6}$$

$V_0 = O$ and $P_j = GF^j\boldsymbol{x}_n\boldsymbol{x}_n'(F^j)'G'$.

For the six models we consider in this class, we obtain the following simpler expression:

$$v_h = (1+\sigma^2)\theta_h - \mu_h^2$$

where $\theta_1 = \mu_1^2$ and

$$\theta_h = \mu_h^2 + \sigma^2\sum_{j=1}^{h-1} c_j^2\theta_{h-j} \tag{7}$$

and c_j depends on the particular model. Note that c_j is identical to that for the corresponding additive error model from Class 1. Specific values for μ_h and c_j for the particular models in Class 2 are given in Tables II and III.

Note that forecast means from MNN are equivalent to simple exponential smoothing (SES) but that the variances are different from ANN. Similarly, MAN gives forecast means equivalent to Holt's method but with different variances from AAN and MAA gives forecast means equivalent to the additive Holt–Winters method but with different variances from AAA.

In the case of MNN, a non-recursive expression for v_h can be obtained:

$$v_h = \ell_n^2\left[(1+\alpha^2\sigma^2)^{h-1}(1+\sigma^2)-1\right]$$

Class 3

Derivations are given in Appendix C.

Copyright © 2005 John Wiley & Sons, Ltd.

For models in this class:

$$\mu_h = H_1 M_{h-1} H_2' \tag{8}$$

and

$$v_h = (1+\sigma^2)(H_2 \otimes H_1)V_{h-1}(H_2 \otimes H_1)' + \sigma^2 \mu_h^2 \tag{9}$$

where $\otimes$ denotes a Kronecker product, $M_0 = \boldsymbol{x}_n \boldsymbol{z}_n'$, $V_0 = O_{2m}$ and for $h \geq 1$:

$$M_h = F_1 M_{h-1} F_2' + G_1 M_{h-1} G_2' \sigma^2 \tag{10}$$

and

$$\begin{aligned} V_h = {} & (F_2 \otimes F_1)V_{h-1}(F_2 \otimes F_1)' + \sigma^2\left[(F_2 \otimes F_1)V_{h-1}(G_2 \otimes G_1)' + (G_2 \otimes G_1)V_{h-1}(F_2 \otimes F_1)'\right] \\ & + \sigma^2(G_2 \otimes F_1 + F_2 \otimes G_1)\left[V_{h-1} + \mathrm{vec}M_{h-1}(\mathrm{vec}M_{h-1})'\right](G_2 \otimes F_1 + F_2 \otimes G_1)' \\ & + \sigma^4(G_2 \otimes G_1)\left[3V_{h-1} + 2\mathrm{vec}M_{h-1}(\mathrm{vec}M_{h-1})'\right](G_2 \otimes G_1)' \end{aligned} \tag{11}$$

Note, in particular, that $\mu_1 = (H_1 \boldsymbol{x}_n)(H_2 \boldsymbol{z}_n)'$ and $v_1 = \sigma^2 \mu_1^2$.

Because σ^2 is usually small (much less than 1), approximate expressions for the mean and variance can be obtained:

$$\begin{aligned} \mu_h &= \mu_{1,h-1}\mu_{2,h-1} + O(\sigma^2) \\ v_h &\approx (1+\sigma^2)[v_{1,h-1} + \mu_{1,h-1}^2][v_{2,h-1} + \mu_{2,h-1}^2] - \mu_{1,h-1}^2 \mu_{2,h-1}^2 \end{aligned}$$

where $\mu_{1,h} = H_1 F_1^h \boldsymbol{x}_n$, $\mu_{2,h} = H_2 F_2^h \boldsymbol{z}_n$, $v_{1,h} = \mathrm{var}(H_1 \boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n)$ and $v_{2,h} = \mathrm{var}(H_2 \boldsymbol{z}_{n+h} \mid \boldsymbol{z}_n)$.

In the three special cases we consider, these expressions can be written as

$$\mu_h = \tilde{\mu}_h s_{n-m+1+(h-1)^*} + O(\sigma^2) \tag{12}$$

and

$$v_h \approx s_{n-m+1+(h-1)^*}^2 \left[\theta_h (1+\sigma^2)(1+\gamma^2\sigma^2)^k - \tilde{\mu}_h^2\right] \tag{13}$$

where $k = \lfloor (h-1)/m \rfloor$, $\theta_1 = \tilde{\mu}_1^2$ and

$$\theta_h = \tilde{\mu}_h^2 + \sigma^2 \sum_{j=1}^{h-1} c_j^2 \theta_{h-j}, \quad h \geq 2$$

These expressions are exact for $h \leq m$. Specific values for μ_h, $\tilde{\mu}_h$ and c_j for the particular models in Class 3 are given in Tables II and III.

Note that the usual point forecasts for these models are given by (12) rather than (8). Also, the forecast means from MAM are equivalent to the multiplicative Holt–Winters method.

Copyright © 2005 John Wiley & Sons, Ltd.

For the MNM model, a simpler expression for v_h is available:

$$v_h \approx s^2_{n-m+1+(h-1)^*}\left[(1+\alpha^2\sigma^2)^{h-1}(1+\sigma^2)(1+\gamma^2\sigma^2)^k - \ell_n^2\right]$$

(The expression is exact for $h \leq m$.)

THE ACCURACY OF THE APPROXIMATIONS

In order to investigate the accuracy of the approximations for the mean (12) and standard deviation (13) to the exact expressions in (8) and (9), we provide some comparisons for the MAM model in Class 3.

These comparisons are done for quarterly data where the values for the components are assumed to be the following: $\ell_n = 100$, $b_n = 2$, $s_n = 0.80$, $s_{n-1} = 1.20$, $s_{n-1} = 0.90$, $s_{n-1} = 1.10$. We use the following base level values for the parameters: $\alpha = 0.2$, $\beta = 0.3$ (i.e., $\alpha\beta = 0.06$), $\gamma = 0.1$, $\sigma = 0.05$. We vary these parameters one at a time as shown in Table IV.

The results in Table IV show that the mean and approximate mean are always very close and that the percentage difference in the standard deviations only becomes substantial when we increase γ. This result for the standard deviation is not surprising because the approximation is exact if $\gamma = 0$. In fact, we recommend that the approximation not be used if the smoothing parameter for γ exceeds 0.10.

PREDICTION INTERVALS

The forecast distributions for Class 1 are clearly normal as the models are linear and the errors are normal. Consequently, $100(1 - \alpha)\%$ prediction intervals can be calculated from the forecast means and variances in the usual way, namely $\mu_h \pm z_{\alpha/2}\sqrt{v_h}$ where z_q denotes the qth quantile of a standard normal distribution.

The forecast distributions for Classes 2 and 3 are non-normal because of the non-linearity of the state space equations. However, prediction intervals based on the above (normal) formula will usually give reasonably accurate results, as the following example shows.

Example

As a numerical example, we consider the quarterly sales data given in Makridakis *et al.* (1998, p. 162) and use the multiplicative Holt–Winters method (model MAM). Following the approach outlined in Hyndman *et al.* (2001), we estimate the parameters to be $\alpha = 0.8$, $\beta = 0.1$, $\gamma = 0.1$, $\sigma = 0.0384$ with the final states $\ell_n = 757.2$, $b_n = 17.6$, $z_n = (0.873, 1.146, 1.031, 0.958)'$.

Figure 1 shows the forecast standard deviations calculated exactly using (9) and approximately using (13). We also show the approximation suggested by Koehler *et al.* (2001) for $1 \leq h \leq m$. Clearly, both approximations are very close to the exact values in this case (because σ^2 is so small here).

The data with three years of forecasts are shown in Figure 2. In this case, the conditional mean forecasts obtained from model MAM are virtually indistinguishable from the usual forecasts because σ is so small (they are identical up to $h = m$). The solid lines show prediction intervals calculated as $\mu_h \pm 1.96\sqrt{v_h}$ and the dotted lines show prediction intervals computed by generating 20,000 future sample paths from the fitted model and finding the 2.5% and 97.5% quantiles at each forecast

Copyright © 2005 John Wiley & Sons, Ltd.

Table IV. Comparison of exact and approximate means and standard deviations for MAM model in Class 3 (i.e., (8) and (9) versus (12) and (13))

Period ahead h	Mean (8) μ_h	Approximate mean (12)	SD (9) $\sqrt{v_h}$	Approximate SD (13)	SD percent difference
$\sigma = 0.05$, $\alpha = 0.2$, $\alpha\beta = 0.06$, $\gamma = 0.1$					
5	121.01	121.00	7.53	7.33	2.69
6	100.81	100.80	6.68	6.52	2.37
7	136.81	136.80	9.70	9.50	2.07
8	92.81	92.80	7.06	6.93	1.80
9	129.83	129.80	10.85	10.45	3.68
10	108.03	108.00	9.65	9.34	3.21
11	146.44	146.40	13.99	13.60	2.81
12	99.22	99.20	10.13	9.88	2.47
$\sigma = \mathbf{0.1}$, $\alpha = 0.2$, $\alpha\beta = 0.06$, $\gamma = 0.1$					
5	121.05	121.00	15.09	14.68	2.73
6	100.84	100.80	13.39	13.07	2.40
7	136.86	136.80	19.45	19.04	2.11
8	92.84	92.80	14.15	13.89	1.84
9	129.93	129.80	21.77	20.96	3.75
10	108.11	108.00	19.39	18.75	3.29
11	146.55	146.40	28.11	27.30	2.89
12	99.30	99.20	20.35	19.83	2.55
$\sigma = 0.05$, $\alpha = \mathbf{0.6}$, $\alpha\beta = 0.06$, $\gamma = 0.1$					
5	121.02	121.00	10.87	10.60	2.47
6	100.82	100.80	9.96	9.76	2.04
7	136.83	136.80	14.76	14.51	1.72
8	92.82	92.80	10.86	10.70	1.47
9	129.86	129.80	16.64	16.19	2.71
10	108.05	108.00	14.83	14.48	2.37
11	146.46	146.40	21.45	21.00	2.09
12	99.24	99.20	15.45	15.16	1.86
$\sigma = 0.05$, $\alpha = 0.2$, $\alpha\beta = \mathbf{0.18}$, $\gamma = 0.1$					
5	121.03	121.00	10.19	9.87	3.08
6	100.82	100.80	9.88	9.66	2.27
7	136.83	136.80	15.55	15.29	1.69
8	92.82	92.80	12.14	11.98	1.28
9	129.87	129.80	19.67	19.16	2.56
10	108.06	108.00	18.41	18.04	2.03
11	146.48	146.40	27.86	27.41	1.64
12	99.26	99.20	20.93	20.65	1.35
$\sigma = 0.05$, $\alpha = 0.2$, $\alpha\beta = 0.06$, $\beta = \mathbf{0.3}$					
5	121.04	121.00	8.10	7.53	7.12
6	100.83	100.80	7.13	6.68	6.36
7	136.84	136.80	10.28	9.70	5.64
8	92.83	92.80	7.42	7.05	4.97
9	129.90	129.80	11.89	10.77	9.46
10	108.08	108.00	10.47	9.59	8.42
11	146.51	146.40	15.04	13.91	7.49
12	99.27	99.20	10.79	10.07	6.67

Copyright © 2005 John Wiley & Sons, Ltd.

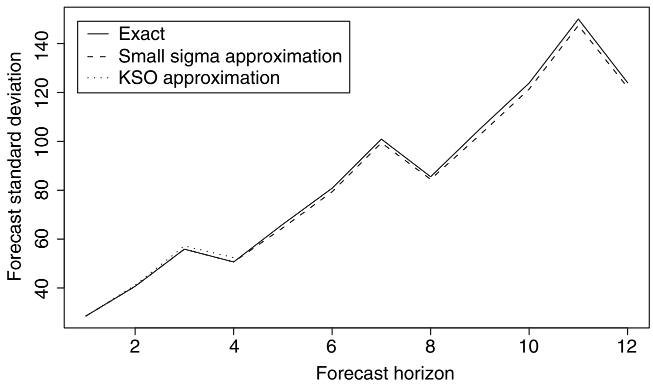

Figure 1. Forecast standard deviations calculated (a) exactly using (9); (b) approximately using (13); and (c) using the approximation suggested by Koehler *et al.* (2001) for $1 \leq h \leq m$

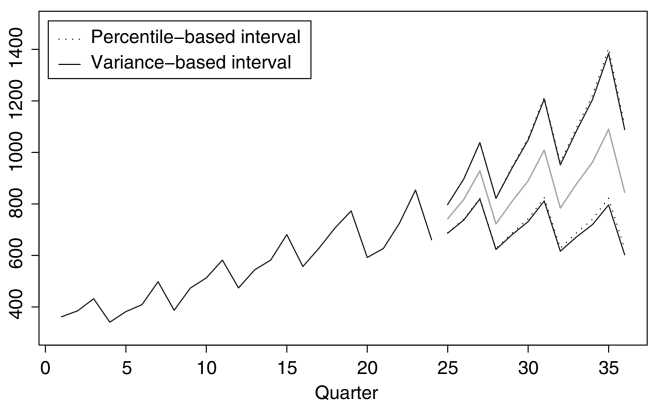

Figure 2. Quarterly sales data with three years of forecasts. The solid lines show prediction intervals calculated as $\mu_h \pm 1.96\sqrt{v_h}$ and the dotted lines show prediction intervals computed by generating 20,000 future sample paths from the fitted model and finding the 2.5% and 97.5% quantiles at each forecast horizon

horizon. Clearly, the variance-based intervals are a good approximation despite the non-normality of the forecast distributions.

SUMMARY

For three general classes of state space models, we have provided derivations of exact matrix expressions for the means and variances of prediction distributions. We relate these three classes of state space models to the commonly used exponential smoothing methods (simple, Holt, additive and multiplicative Holt–Winters) and to other known exponential smoothing methods (Hyndman *et al.*, 2001). We provide a summary of these models and the corresponding non-matrix expressions of the means and variances in Tables I, II and III. These means and variances may be used to

Copyright © 2005 John Wiley & Sons, Ltd.

construct analytical prediction intervals when using the exponential smoothing methods for forecasting.

The non-matrix formulas for the Class 3 models are not exact for $h > m$. In Table IV we compare our exact matrix formulas with our approximate formulas for the model that corresponds to the multiplicative Holt–Winters method (MAM). We find that the approximation is very good as long as the smoothing parameter for the seasonal component remains small (i.e., less than 0.1). We also consider an example in which we compare our forecast standard deviations and prediction intervals with the values from some of the previously used approaches.

In summary, we have provided, for the first time, exact analytical formulas for the variances of prediction distributions for all the exponential smoothing methods. More generally, we have exact formulas for variances of the general state space models, of which the exponential smoothing models are special cases. Where possible, we have presented both matrix and non-matrix expressions.

Simulation methods have been the only comprehensive approach to handling the prediction distribution problem for all exponential smoothing methods to date. Our formulas provide an effective alternative, the advantage being that they involve much lower computational loads.

APPENDIX A: PROOFS OF RESULTS FOR CLASS 1

Let

$$\begin{aligned} \boldsymbol{m}_h &= E(\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n) \\ V_h &= \operatorname{var}(\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n) \end{aligned}$$

Note that $\boldsymbol{m}_0 = \boldsymbol{x}_n$ and $V_0 = O$.

For Class 1

$$\boldsymbol{m}_h = F\boldsymbol{m}_{h-1} = F^2\boldsymbol{m}_{h-2} = \ldots = F^h\boldsymbol{m}_0 = F^h\boldsymbol{x}_n$$

and therefore

$$\mu_h = H\boldsymbol{m}_{h-1} = HF^{h-1}\boldsymbol{x}_n$$

The state forecast variance is given by

$$V_h = FV_{h-1}F' + GG'\sigma^2$$

and therefore

$$V_h = \sigma^2 \sum_{j=0}^{h-1} F^j GG'(F^j)'$$

Hence, the prediction variance for h periods ahead is

$$v_h = HV_{h-1}H' + \sigma^2 = \begin{cases} \sigma^2 & \text{if } h = 1 \\ \sigma^2\left[1 + \sum_{j=1}^{h-1} c_j^2\right] & \text{if } h \geq 2 \end{cases}$$

Copyright © 2005 John Wiley & Sons, Ltd.

where $c_j = HF^{j-1}G$.

We now consider the particular cases.

ADA

We first derive the results for the ADA case. Here the state is $\boldsymbol{x}_n = (\ell_n, b_n, s_n, s_{n-1}, \ldots, s_{n-m+1})$:

$$H = [1 \quad 1 \quad \mathbf{0}'_{m-1} \quad 1], \qquad F = \begin{bmatrix} 1 & 1 & \mathbf{0}'_{m-1} & 0 \\ 0 & \phi & \mathbf{0}'_{m-1} & 0 \\ 0 & 0 & \mathbf{0}'_{m-1} & 1 \\ \mathbf{0}_{m-1} & \mathbf{0}_{m-1} & I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}, \qquad G = \begin{bmatrix} \alpha \\ \alpha\beta \\ \gamma \\ \mathbf{0}_{m-1} \end{bmatrix}$$

where I_n denotes the $n \times n$ identity matrix and $\mathbf{0}_n$ denotes a zero vector of length n. Therefore, $HF^i = [1, \phi_i, d_{i+1,m}, d_{i+2,m}, \ldots, d_{i+m,m}]'$ where $\phi_i = 1 + \phi + \cdots + \phi^i$ and $d_{j,m} = 1$ if $j = m$ (mod m) and $d_{j,m} = 0$ otherwise. Hence we find $c_j = HF^{j-1}G = \alpha(1 + \phi_{j-1}\beta) + \gamma d_{j,m}$,

$$\mu_h = \ell_n + \phi_{h-1}b_n + s_{n-m+1+(h-1)^*}$$

and for $h \geq 2$:

$$\begin{aligned} v_h &= \sigma^2\left\{1 + \sum_{j=1}^{h-1}[\alpha(1+\phi_{j-1}\beta) + \gamma d_{j,m}]^2\right\} \\ &= \sigma^2\left\{1 + \sum_{j=1}^{h-1}\left[\alpha^2(1+\phi_{j-1}\beta)^2 + \gamma d_{j,m}[\gamma + 2\alpha(1+\phi_{j-1}\beta)]\right]\right\} \end{aligned}$$

These formulas agree with those of Yar and Chatfield (1990), except that we apply the dampening parameter ϕ beginning in the second forecast time period, $n + 2$, instead of in the first forecast time period, $n + 1$.

Other cases

All other cases of Class 1 can be derived as special cases of ADA:

- For ADN, we use the results of ADA with $\gamma = 0$ and $s_t = 0$ for all t.
- For AAN, we use the results of ADN with $\phi = 1$.
- The results for ANN are obtained from AAN by further setting $\beta = 0$ and $b_t = 0$ for all t.
- For AAA, the results of ADA hold with $\phi = 1$.
- The results for ANA are obtained as a special case of AAA with $\beta = 0$ and $b_t = 0$ for all t.

APPENDIX B: PROOFS OF RESULTS FOR CLASS 2

Let $\boldsymbol{m}_h$ and V_h be defined as in Appendix A. The forecast means for Class 2 have the same form as for Class 1, namely

$$\mu_h = H\boldsymbol{m}_{h-1} = HF^{h-1}\boldsymbol{x}_n$$

Copyright © 2005 John Wiley & Sons, Ltd.

To obtain V_h, first note that $V_h = FV_{h-1}F' + G\ \mathrm{var}(\boldsymbol{x}_{n+h-1}\varepsilon_{n+h})G'$ and that

$$\mathrm{var}(\boldsymbol{x}_{n+h-1}\varepsilon_{n+h}) = \mathrm{E}[\boldsymbol{x}_{n+h-1}\boldsymbol{x}'_{n+h-1}]\mathrm{E}(\varepsilon^2_{n+h}) - 0 = \sigma^2[V_{h-1} + \boldsymbol{m}_{h-1}\boldsymbol{m}'_{h-1}]$$

Therefore

$$V_h = FV_{h-1}F' + \sigma^2 GV_{h-1}G' + \sigma^2 P_{h-1}$$

where $P_j = GF^j\boldsymbol{x}_n\boldsymbol{x}'_n(F^j)'G'$.

The forecast variance is given by

$$\begin{aligned} v_h &= HV_{h-1}H'(1+\sigma^2) + \sigma^2 H\boldsymbol{m}_{h-1}\boldsymbol{m}'_{h-1}H' \\ &= HV_{h-1}H'(1+\sigma^2) + \sigma^2\mu_h^2 \end{aligned}$$

In the special case where $G = QH$ we obtain a simpler result. In this case, $\boldsymbol{x}_t = F\boldsymbol{x}_{t-1} + Qe_t$ where $e_t = y_t - H\boldsymbol{x}_{t-1} = H\boldsymbol{x}_{t-1}\varepsilon_t$. Thus, we obtain the linear exponential smoothing updating rule $\boldsymbol{x}_t = F\boldsymbol{x}_{t-1} + Q(y_t - H\boldsymbol{x}_{t-1})$. Define θ_h such that $\mathrm{var}(e_{n+h} \mid \boldsymbol{x}_n) = \theta_h\sigma^2$. Then it is readily seen that $V_h = FV_{h-1}F' + QQ'\ \mathrm{var}(e_{n+h} \mid \boldsymbol{x}_n)$ and so, by repeated substitution:

$$V_h = \sigma^2 \sum_{j=0}^{h-1} F^j QQ'(F^j)'\theta_{h-j}$$

and

$$HV_{h-1}H' = \sigma^2 \sum_{j=1}^{h-1} c_j^2\theta_{h-j} \tag{B1}$$

where $c_j = HF^{j-1}Q$. Now

$$e_{n+h} = [H(\boldsymbol{x}_{n+h-1} - \boldsymbol{m}_{h-1}) + H\boldsymbol{m}_{h-1}]\varepsilon_{n+h}$$

which we square and take expectations to give $\theta_h = HV_{h-1}H' + \mu_h^2$. Substituting (B1) into this expression for θ_h gives

$$\theta_h = \sigma^2 \sum_{j=1}^{h-1} c_j^2\theta_{h-j} + \mu_h^2 \tag{B2}$$

where $\theta_1 = \mu_1^2$. The forecast variance is then given by

$$v_h = (1+\sigma^2)\theta_h - \mu_h^2 \tag{B3}$$

We now consider the particular cases.

MDA

We first derive the results for the MDA case. Here the state is $\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-m+1})'$, $H = [1, 1, 0, \ldots, 0, 1]$,

Copyright © 2005 John Wiley & Sons, Ltd.

$$F = \begin{bmatrix} 1 & 1 & \mathbf{0}'_{m-1} & 0 \\ 0 & \phi & \mathbf{0}'_{m-1} & 0 \\ 0 & 0 & \mathbf{0}'_{m-1} & 1 \\ \mathbf{0}_{m-1} & \mathbf{0}_{m-1} & I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} \alpha & \alpha & \mathbf{0}'_{m-1} & \alpha \\ \alpha\beta & \alpha\beta & \mathbf{0}'_{m-1} & \alpha\beta \\ \gamma & \gamma & \mathbf{0}'_{m-1} & \gamma \\ \mathbf{0}_{m-1} & \mathbf{0}_{m-1} & O_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}$$

Then from (4) we obtain $\mu_h = \ell_n + \phi_{h-1}b_n + s_{n-m+1+(h-1)^*}$ where $\phi_i = 1 + \phi + \cdots + \phi^i$ and $j^* = j \bmod m$.

To obtain the expression for v_h, note that this model satisfies the special case $G = QH$ where $Q = [\alpha, \alpha\beta, \gamma, \mathbf{0}'_{m-1}]'$. Thus we can use the expression (B3) where $c_j = HF^{j-1}Q = \alpha(1 + \phi_{j-1}\beta) + \gamma d_{j,m}$ (the same as c_j for the corresponding model from Class 1).

Other cases

All other cases of Class 2 can be derived as special cases of MDA:

- For MDN, we use the results of MDA with $\gamma = 0$ and $s_t = 0$ for all t.
- For MAN, we use the results of MDN with $\phi = 1$.
- For MAA, the results of MDA hold with $\phi = 1$.
- The results for MNA are obtained as a special case of MAA with $\beta = 0$ and $b_t = 0$ for all t.
- The results for MNN are obtained from MAN by further setting $\beta = 0$ and $b_t = 0$ for all t. In this case, a simpler expression for v_h can be obtained. Note that $c_j = \alpha$, $\theta_1 = \ell_n^2$ and for $j \geq 2$:

$$\theta_j = \ell_n^2 + \sigma^2\alpha^2 \sum_{i=1}^{j-1} \theta_{j-i} = \ell_n^2 + \alpha^2\sigma^2(\theta_1 + \theta_2 + \ldots + \theta_{j-1})$$

Hence

$$\theta_j = \ell_n^2 (1 + \alpha^2\sigma^2)^{j-1}$$

and

$$v_h = \ell_n^2\left[(1 + \alpha^2\sigma^2)^{h-1}\right](1 + \sigma^2) - \ell_n^2 = \ell_n^2\left[(1 + \alpha^2\sigma^2)^{h-1}(1 + \sigma^2) - 1\right]$$

APPENDIX C: PROOFS OF RESULTS FOR CLASS 3

Note that we can write Y_t as

$$Y_t = H_1 \boldsymbol{x}_{t-1} \boldsymbol{z}'_{t-1} H'_2 (1 + \varepsilon_t)$$

So let $W_h = \boldsymbol{x}_{n+h}\boldsymbol{z}'_{n+h}$, $M_h = \mathrm{E}(W_h \mid \boldsymbol{x}_n, \boldsymbol{z}_n)$ and $V_h = \mathrm{var}(W_h \mid \boldsymbol{x}_n, \boldsymbol{z}_n)$ where (by standard definitions)

$$V_h = \mathrm{var}(\mathrm{vec}\, W_h \mid \boldsymbol{x}_n, \boldsymbol{z}_n) \qquad \text{and} \qquad \mathrm{vec}\, A = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_r \end{bmatrix}$$

where matrix A = $[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r]$

Note that

$$\begin{aligned} W_h &= (F_1\boldsymbol{x}_{n+h-1} + G_1\boldsymbol{x}_{n+h-1}\varepsilon_{n+h})(\boldsymbol{z}'_{n+h-1}F'_2 + \boldsymbol{z}'_{n+h-1}G'_2\varepsilon_{n+h}) \\ &= F_1W_{h-1}F'_2 + (F_1W_{h-1}G'_2 + G_1W_{h-1}F'_2)\varepsilon_{n+h} + G_1W_{h-1}G'_2\varepsilon^2_{n+h} \end{aligned}$$

It follows that $M_0 = \boldsymbol{x}_n\boldsymbol{z}'_n$ and

$$M_h = F_1M_{h-1}F'_2 + G_1M_{h-1}G'_2\sigma^2 \tag{C1}$$

For the variance of W_h, we find $V_0 = 0$ and

$$\begin{aligned} V_h &= \mathrm{var}\{\mathrm{vec}(F_1W_{h-1}F'_2) + \mathrm{vec}(F_1W_{h-1}G'_2 + G_1W_{h-1}F'_2)\varepsilon_{n+h} + \mathrm{vec}(G_1W_{h-1}G'_2)\varepsilon^2_{n+h}\} \\ &= (F_2\otimes F_1)V_{h-1}(F_2\otimes F_1)' + (G_2\otimes F_1 + F_2\otimes G_1)\mathrm{var}(\mathrm{vec}\ W_{h-1}\varepsilon_{n+h})(G_2\otimes F_1 + F_2\otimes G_1)' \\ &\quad + (G_2\otimes G_1)\mathrm{var}(\mathrm{vec}\ W_{h-1}\varepsilon^2_{n+h})(G_2\otimes G_1)' \\ &\quad + (F_2\otimes F_1)\mathrm{cov}(\mathrm{vec}\ W_{h-1}, \mathrm{vec}\ W_{h-1}\varepsilon^2_{n+h})(G_2\otimes G_1)' \\ &\quad + (G_2\otimes G_1)\mathrm{cov}(\mathrm{vec}\ W_{h-1}\varepsilon^2_{n+h}, \mathrm{vec}\ W_{h-1})(F_2\otimes F_1)' \end{aligned}$$

Next we find that

$$\begin{aligned} \mathrm{var}(\mathrm{vec}\ W_{h-1}\varepsilon_{n+h}) &= \mathrm{E}[\mathrm{vec}\ W_{h-1}(\mathrm{vec}\ W_{h-1})'\varepsilon^2_{n+h}] = \sigma^2(V_{h-1} + \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})') \\ \mathrm{var}(\mathrm{vec}\ W_{h-1}\varepsilon^2_{n+h}) &= \mathrm{E}(\mathrm{vec}\ W_{h-1}(\mathrm{vec}\ W_{h-1})'\varepsilon^4_{n+h}) - \mathrm{E}(\mathrm{vec}\ W_{h-1})\mathrm{E}(\mathrm{vec}\ W_{h-1})'\sigma^4 \\ &= 3\sigma^4(V_{h-1} + \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})') - \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})'\sigma^4 \\ &= \sigma^4(3V_{h-1} + 2\ \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})') \end{aligned}$$

and

$$\begin{aligned} \mathrm{cov}(\mathrm{vec}\ W_{h-1}, \mathrm{vec}\ W_{h-1}\varepsilon^2_{n+h}) &= \mathrm{E}(\mathrm{vec}\ W_{h-1}(\mathrm{vec}\ W_{h-1})'\varepsilon^2_{n+h}) - \mathrm{E}(\mathrm{vec}\ W_{h-1})\mathrm{E}(\mathrm{vec}\ W_{h-1})'\sigma^2 \\ &= \sigma^2(V_{h-1} + \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})') - \sigma^2\ \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})' \\ &= \sigma^2 V_{h-1} \end{aligned}$$

It follows that

$$\begin{aligned} V_h &= (F_2\otimes F_1)V_{h-1}(F_2\otimes F_1)' + \sigma^2[(F_2\otimes F_1)V_{h-1}(G_2\otimes G_1)' + (G_2\otimes G_1)V_{h-1}(F_2\otimes F_1)'] \\ &\quad + \sigma^2(G_2\otimes F_1 + F_2\otimes G_1)[V_{h-1} + \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})'](G_2\otimes F_1 + F_2\otimes G_1)' \\ &\quad + \sigma^4(G_2\otimes G_1)[3V_{h-1} + 2\ \mathrm{vec}\ M_{h-1}(\mathrm{vec}\ M_{h-1})'](G_2\otimes G_1)' \end{aligned}$$

The forecast mean and variance are given by

$$\mu_h = \mathrm{E}(Y_{n+h}\,|\,\boldsymbol{x}_n, \boldsymbol{z}_n) = H_1M_{h-1}H'_2$$

Copyright © 2005 John Wiley & Sons, Ltd.

and

$$\begin{aligned}v_h = \operatorname{var}(Y_{n+h} \mid \boldsymbol{x}_n, z_n) &= \operatorname{var}[\operatorname{vec}(H_1 W_{h-1} H_2' + H_1 W_{h-1} H_2 \varepsilon_{n+h})] \\ &= \operatorname{var}[(H_2 \otimes H_1)\operatorname{vec} W_{h-1} + (H_2 \otimes H_1)\operatorname{vec} W_{h-1}\varepsilon_{n+h}] \\ &= (H_2 \otimes H_1)[V_{h-1}(1+\sigma^2) + \sigma^2 \operatorname{vec} M_{h-1}(\operatorname{vec} M_{h-1})'](H_2' \otimes H_1') \\ &= (1+\sigma^2)(H_2 \otimes H_1)V_{h-1}(H_2 \otimes H_1)' + \sigma^2 \mu_h^2\end{aligned}$$

When σ is small (much less than 1), it is possible to obtain some simpler but approximate expressions. The second term in (C1) can be dropped to give $M_h = F_1^{h-1} M_0 (F_2^{h-1})'$ and so

$$\mu_h \approx H_1 F_1^{h-1} \boldsymbol{x}_n (H_2 F_2^{h-1} z_n)'$$

The order of this approximation can be obtained by noting that the observation equation may be written as $Y_t = U_{1,t} U_{2,t} U_{3,t}$ where $U_{1,t} = H_1 \boldsymbol{x}_{t-1}$, $U_{2,t} = H_2 z_{t-1}$ and $U_{3,t} = 1 + \varepsilon_t$. Then

$$\mathrm{E}(Y_t) = \mathrm{E}(U_{1,t}U_{2,t}U_{3,t}) = \mathrm{E}(U_{1,t}U_{2,t})\mathrm{E}(U_{3,t})$$

since $U_{3,t}$ is independent of $U_{1,t}$ and $U_{2,t}$. Since $\mathrm{E}(U_{1,t}U_{2,t}) = \mathrm{E}(U_{1,t})\mathrm{E}(U_{2,t}) + \operatorname{cov}(U_{1,t}, U_{2,t})$, we have the approximation:

$$\mu_h = \mathrm{E}(Y_{n+h} \mid \boldsymbol{x}_n, z_n) = \mathrm{E}(U_{1,n+h} \mid \boldsymbol{x}_n)\mathrm{E}(U_{2,n+h} \mid z_n)\mathrm{E}(U_{3,n+h}) + O(\sigma^2)$$

When $U_{2,n+h}$ is constant the result is exact. Now let

$$\begin{aligned}\mu_{1,h} &= \mathrm{E}(U_{1,n+h+1} \mid \boldsymbol{x}_n) = \mathrm{E}(H_1 \boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n) = H_1 F_1^h \boldsymbol{x}_n \\ \mu_{2,h} &= \mathrm{E}(U_{2,n+h+1} \mid z_n) = \mathrm{E}(H_2 z_{n+h} \mid z_n) = H_2 F_2^h z_n \\ v_{1,h} &= \operatorname{var}(U_{1,n+h+1} \mid \boldsymbol{x}_n) = \operatorname{var}(H_1 \boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n) \\ v_{2,h} &= \operatorname{var}(U_{2,n+h+1} \mid z_n) = \operatorname{var}(H_2 z_{n+h} \mid z_n) \\ v_{12,h} &= \operatorname{cov}(U_{1,n+h+1}^2, U_{2,n+h+1}^2 \mid \boldsymbol{x}_n, z_n) = \operatorname{cov}\left([H_1 \boldsymbol{x}_{n+h}]^2, [H_2 z_{n+h}]^2 \mid \boldsymbol{x}_n, z_n\right)\end{aligned}$$

Then

$$\mu_h = \mu_{1,h-1}\mu_{2,h-1} + O(\sigma^2) = H_1 F_1^{h-1} \boldsymbol{x}_n H_2 F_2^{h-1} z_n + O(\sigma^2)$$

By the same arguments, we have

$$\mathrm{E}(Y_t^2) = \mathrm{E}(U_{1,t}^2 U_{2,t}^2 U_{3,t}^2) = \mathrm{E}(U_{1,t}^2 U_{2,t}^2)\mathrm{E}(U_{3,t}^2)$$

and

$$\begin{aligned}\mathrm{E}(Y_{n+h}^2 \mid z_n, \boldsymbol{x}_n) &= \mathrm{E}(Y_{1,n+h}^2 U_{2,n+h}^2 \mid \boldsymbol{x}_n z_n) = \mathrm{E}(U_{3,n+t}^2) \\ &= [\operatorname{cov}(U_{1,n+h}^2, U_{2,n+h}^2 \mid \boldsymbol{x}_n z_n) + \mathrm{E}(U_{1,n+h}^2 \mid \boldsymbol{x}_n)\mathrm{E}(U_{2,n+h}^2 \mid z_n)]\mathrm{E}(U_{3,n+h}^2) \\ &= (1+\sigma^2)[v_{12,h-1} + (v_{1,h-1} + \mu_{1,h-1}^2)(v_{2,h-1} + \mu_{2,h-1}^2)]\end{aligned}$$

Copyright © 2005 John Wiley & Sons, Ltd.

Assuming the covariance $v_{12,h-1}$ is small compared to the other terms, we obtain

$$v_h \approx (1+\sigma^2)(v_{1,h-1}+\mu_{1,h-1}^2)(v_{2,h-1}+\mu_{2,h-1}^2)-\mu_{1,h-1}^2\mu_{2,h-1}^2$$

We now consider the particular cases.

MDM

We first derive the results for the MDM case where $\boldsymbol{x}_t = (\ell_t, b_t)'$ and $z_t = (s_t, \ldots, s_{t-m+1})'$, and the matrix coefficients are $H_1 = [1, 1]$, $H_2 = [0, \ldots, 0, 1]$,

$$F_1 = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix}, \qquad F_2 = \begin{bmatrix} \mathbf{0}'_{m-1} & 1 \\ I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}, \qquad G_1 = \begin{bmatrix} \alpha & \alpha \\ \alpha\beta & \alpha\beta \end{bmatrix}, \qquad \text{and} \qquad G_2 = \begin{bmatrix} \mathbf{0}'_{m-1} & \gamma \\ O_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}$$

Many terms will be zero in the formulas for the expected value and the variance because of the following relationships: $G_2^2 = O_m$, $H_2G_2 = \mathbf{0}'_m$ and $(H_2 \otimes H_1)(G_2 \otimes X) = \mathbf{0}'_{2m}$ where X is any 2×2 matrix. For the terms that remain, $H_2 \otimes H_1$ and its transpose will only use the terms from the last two rows of the last two columns of the large matrices because $H_2 \otimes H_1 = [\mathbf{0}'_{2m-2}, 1, 1]$.

Using the small σ approximations and exploiting the structure of the MDM model, we can obtain simpler expressions that approximate μ_h and v_h.

Note that $H_2F_2^jG_2 = \gamma d_{j+1,m}H_2$. So for $h < m$, we have

$$H_2 z_{n+h} \mid z_n = H_2 \prod_{j=1}^{h} (F_2 + G_2\varepsilon_{n+h-j+1}) z_n = H_2 F_2^h z_n = s_{n-m+h+1}$$

Furthermore,

$$\mu_{2,h} = s_{n-m+1+h^*}$$

and

$$v_{2,h} = s^2_{n-m+1+h^*}\left[(1+\gamma^2\sigma^2)^k - 1\right]$$

where $k = \lfloor (h-1)/m \rfloor$.

Also note that $\boldsymbol{x}_n$ has the same properties as for MDN in Class 2. Thus

$$\mu_{1,h} = \ell_n + \phi_{h-1} b_n$$

and

$$v_{1,h} = (1+\sigma^2)\theta_h - \mu_{1,h}^2$$

Combining all the terms, we arrive at the approximations

$$\mu_h = \tilde{\mu}_h s_{n-m+1+(h-1)^*} + O(\sigma^2)$$
$$v_h \approx s^2_{n-m+1+(h-1)^*}\left[\theta_h(1+\sigma^2)(1+\gamma^2\sigma^2)^k - \tilde{\mu}_h^2\right]$$

Copyright © 2005 John Wiley & Sons, Ltd.

where $\tilde{\mu}_h = \ell_n + \phi_{h-1}b_n$, $\theta_1 = \tilde{\mu}_1^2$ and

$$\theta_h = \tilde{\mu}_h^2 + \sigma^2\alpha^2\sum_{j=1}^{h-1}(1+\phi_{j-1}\beta)^2\theta_{h-j}, \quad h \ge 2$$

These expressions are exact for $h \le m$. Also, for $h \le m$, the formulas agree with those in Koehler *et al.* (2001) and Chatfield and Yar (1991) if the $O(\sigma^4)$ terms are dropped from the expression.

Other cases

The other cases of Class 3 can be derived as special cases of MDM:

- For MAM, we use the results of MDM with $\phi = 1$.
- The results for MNM are obtained as a special case of MAM with $\beta = 0$ and $b_t = 0$ for all t. The simpler expression for v_h is obtained as for MNN in Class 2.

REFERENCES

Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**: 307–327.

Chatfield C, Yar M. 1991. Prediction intervals for multiplicative Holt–Winters. *International Journal of Forecasting* **7**: 31–37.

Hyndman RJ, Koehler AB, Snyder RD, Grose S. 2001. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* **18**(3): 439–454.

Johnston FR, Harrison PJ. 1986. The variance of leadtime demand. *Journal of the Operational Research Society* **37**: 303–308.

Koehler AB, Snyder RD, Ord JK. 2001. Forecasting models and prediction intervals for the multiplicative Holt–Winters method. *International Journal of Forecasting* **17**: 269–286.

Makridakis S, Wheelwright SC, Hyndman RJ. 1998. *Forecasting: Methods and Applications*, 3rd edn. John Wiley & Sons: New York.

Ord JK, Koehler AB, Snyder RD. 1997. Estimation and prediction for a class of dynamic nonlinear statistical models, *Journal of the American Statistical Association* **92**: 1621–1629.

Snyder RD. 1985. Recursive estimation of dynamic linear models. *Journal of the Royal Statistical Society: Series B* **47**: 272–276.

Yar M, Chatfield C. 1990. Prediction intervals for the Holt–Winters forecasting procedure. *International Journal of Forecasting* **6**: 127–137.

Authors' biographies:

Rob J. Hyndman is professor of statistics in the Department of Econometrics and Business Statistics at Monash University, Australia. He has published extensively in leading statistical and forecasting journals, and is co-author of the well-known text on business forecasting, *Forecasting: Methods and Applications* (Makridakis, Wheelwright & Hyndman, 3rd edn., 1998, Wiley). He is also Editor-in-Chief of the *International Journal of Forecasting*.

Anne B. Koehler is a professor of decision sciences and the Panuska Professor of Business Administration at Miami University, Ohio. Professor Koehler has numerous publications, many of which are on forecasting models for seasonal time series and exponential smoothing methods. She is a co-author of the fourth edition of *Forecasting, Time Series, and Regression: An Applied Approach*, published by Duxbury in 2004.

Keith Ord is a professor at the McDonough School of Business at Georgetown University, Washington, DC. He is co-author of the latest editions of *Kendall's Advanced Theory of Statistics*. His research interests include

Copyright © 2005 John Wiley & Sons, Ltd.

time series and forecasting, inventory modelling and spatial processes. He is a fellow of the American Statistical Association.

Ralph D. Snyder is an associate professor in the Department of Econometrics and Business Statistics at Monash University, Australia. His main areas of research are business forecasting and inventory management. The primary focus of his work has been on the development of a sound statistical framework for the widely used exponential smoothing forecasting methods.

Authors' addresses:
Rob J. Hyndman and Ralph D. Snyder, Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia.

Anne B. Koehler, Department of Decision Sciences and Management Information Systems, Miami University, Oxford, OH 45056, USA.

J. Keith Ord, 320 Old North, Georgetown University, Washington, DC 20057, USA.

Copyright © 2005 John Wiley & Sons, Ltd.