6

Prediction Distributions and Intervals

Point forecasts for each of the state space models were given in Table 2.1 (p. 18). It is also useful to compute the associated prediction distributions and prediction intervals for each model. In this chapter, we discuss how to compute these distributions and intervals.

There are several sources of uncertainty when forecasting a future value of a time series (Chatfield 1993):

1. The uncertainty in model choice—maybe another model is correct, or maybe none of the candidate models is correct.
2. The uncertainty in the future innovations $\varepsilon_{n+1}, \ldots, \varepsilon_{n+h}$.
3. The uncertainty in the estimates of the parameters: α, β, γ, ϕ and $\boldsymbol{x}_0$.

Ideally, the prediction distribution and intervals should take all of these into account. However, this is a difficult problem, and in most time series analysis only the uncertainty in the future innovations is taken into account.

If we assume that the model and its parameters (including $\boldsymbol{x}_0$) are known, then we also know $\boldsymbol{x}_n$, the state vector at the last period of observation, because the error in the transition equation can be calculated from the observations up to time n. Consequently, we define the prediction distribution as the distribution of a future value of the series given the model, its estimated parameters, and $\boldsymbol{x}_n$. A short-hand way of writing this is $y_{n+h|n} \equiv y_{n+h} \mid \boldsymbol{x}_n$.

We briefly discuss how to allow for parameter estimation uncertainty in Sect. 6.1. We do not address how to allow for model uncertainty, although this is an important issue. Hyndman (2001) showed that model uncertainty is likely to be a much bigger source of error than parameter uncertainty.

The mean of the prediction distribution is called the *forecast mean* and is denoted by $\mu_{n+h|n} = \mathrm{E}(y_{n+h} \mid \boldsymbol{x}_n)$. The corresponding *forecast variance* is given by $v_{n+h|n} = \mathrm{V}(y_{n+h} \mid \boldsymbol{x}_n)$. We will find expressions for these quantities for many of the models discussed in this book.

We are also interested in "lead-time demand" forecasting, where we predict the *aggregate* of the next h observations rather than each of the next h observations individually. We discuss this briefly here and in more detail in Chap. 18.

The most direct method of obtaining prediction distributions is to simulate many possible future sample paths from the fitted model, and to estimate the distributions from the simulated data. This approach will work for any time series model, including all of the models discussed in this book. We describe the simulation method in more detail in Sect. 6.1.

While the simulation approach is simple and can be applied to any well-specified time series model, the computations can be time-consuming. Furthermore, the resulting prediction intervals are only available numerically rather than algebraically. Therefore, the approach does not allow for algebraic analysis of the prediction distributions.

An alternative approach is to derive the distributions analytically. Analytical results on prediction distributions can provide additional insight and can be much quicker to compute. These results are relatively easy to derive for some models (particularly the linear models), but very difficult for others. In fact, there are analytical results on prediction distributions for only 15 of the 30 models in our exponential smoothing framework.

When discussing the analytical prediction distributions, it is helpful to divide the thirty state space models given in Tables 2.2 and 2.3 (pp. 21–22) into five classes; Classes 1–4 are shown in Table 6.1.

For each of Classes 1–3, we give expressions for the forecast means and variances. Class 1 consists of the linear models with homoscedastic errors; these are discussed in Sect. 6.2. In Sect. 6.3 we discuss Class 2, which contains the linear models with heteroscedastic errors. Class 3 models are discussed

Table 6.1. The models separated in the exponential smoothing framework split into Classes 1–5.

	A,N,N	A,N,A				
Class 1 $\longrightarrow$	A,A,N	A,A,A				
	A,A_d,N	A,A_d,A				
	M,N,N	M,N,A	M,N,M			
Class 2 $\longrightarrow$	M,A,N	M,A,A	M,A,M			$\longleftarrow$ Class 3
	M,A_d,N	M,A_d,A	M,A_d,M			
Class 4 $\longrightarrow$	M,M,N		M,M,M			
	M,M_d,N		M,M_d,M			
		M,M,A	A,N,M	A,M,N	A,M_d,N	
Class 5 $\longrightarrow$		M,M_d,A	A,A,M	A,M,A	A,M_d,A	
			A,A_d,M	A,M,M	A,M_d,M	

in Sect. 6.4; these are the models with multiplicative errors and multiplicative seasonality but additive trend.

Class 4 consists of the models with multiplicative errors, multiplicative trend, and either no seasonality or multiplicative seasonality. For Class 4, there are no available analytical expressions for forecast means or variances, and so we recommend using simulation to find prediction intervals.

The remaining 11 models are in Class 5. For these models, we also recommend using simulation to obtain prediction intervals. However, Class 5 models are those that can occasionally lead to numerical difficulties with very long forecast horizons. Specifically, the forecast variances are infinite, although this does not usually matter in practice for short- or medium-term forecasts. This issue is explored in Chap. 15.

Section 6.5 discusses the use of the forecast mean and variance formulae to construct prediction intervals even in cases where the prediction distributions are not Gaussian. In Sect. 6.6, we discuss lead-time demand forecasting for Class 1 models.

Most of the results in this chapter are based on Hyndman et al. (2005) and Snyder et al. (2004), although we use a slightly different parameterization in this book, and we extend the results in some new directions.

To simplify some of the expressions, we introduce the following notation:

$$h = mh_m + h_m^+,$$

where[1] h is the forecast horizon, m is the number of periods in each season, $h_m = \lfloor (h-1)/m \rfloor$ and $h_m^+ = \big[(h-1) \bmod m\big] + 1$. In other words, h_m is the number of complete years in the forecast period *prior* to time h, and h_m^+ is the number of remaining times in the forecast period up to and including time h. Thus, h_m^+ can take values $1, 2, \ldots, m$.

6.1 Simulated Prediction Distributions and Intervals

Recall from Chap. 4 that the general model with state vector

$$\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-m+1})'$$

has the form

$$\begin{aligned} y_t &= w(\boldsymbol{x}_{t-1}) + r(\boldsymbol{x}_{t-1})\varepsilon_t, \\ \boldsymbol{x}_t &= \boldsymbol{f}(\boldsymbol{x}_{t-1}) + \boldsymbol{g}(\boldsymbol{x}_{t-1})\varepsilon_t, \end{aligned}$$

where $w(\cdot)$ and $r(\cdot)$ are scalar functions, $\boldsymbol{f}(\cdot)$ and $\boldsymbol{g}(\cdot)$ are vector functions, and $\{\varepsilon_t\}$ is a white noise process with variance σ^2.

One simple approach to obtaining the prediction distribution is to simulate sample paths from the models, conditional on the final state $\boldsymbol{x}_n$. This

[1] The notation $\lfloor u \rfloor$ means the integer part of u.

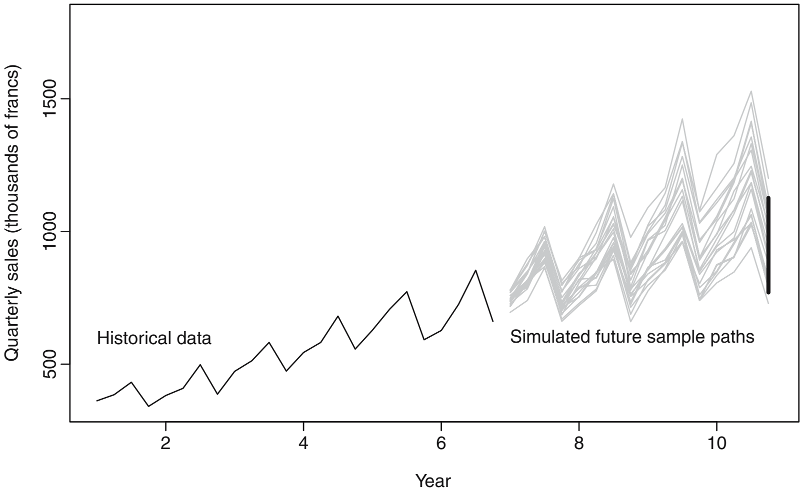

Fig. 6.1. Quarterly French exports data with 20 simulated future sample paths generated using the ETS(M,A,M) model assuming Gaussian innovations. The *solid vertical line* on the right shows a 90% prediction interval for the 16-step forecast horizon, calculated from the 0.05 and 0.95 quantiles of the 5,000 simulated values.

was the approach taken by Ord et al. (1997) and Hyndman et al. (2002). That is, we generate observations $\{y_t^{(i)}\}$, for $t = n+1, \ldots, n+h$, starting with $\boldsymbol{x}_n$ from the fitted model. Each ε_t value is obtained from a random number generator assuming a Gaussian or other appropriate distribution. This procedure is repeated for $i = 1, \ldots, M$, where M is a large integer. (In practice, we often use $M = 5{,}000$.)

Figure 6.1 shows a series of quarterly exports of a French company (in thousands of francs) taken from Makridakis et al. (1998, p. 162). We fit an ETS(M,A,M) model to the data. Then the model is used to simulate 5,000 future sample paths of the data. Twenty of these sample paths are shown in Fig. 6.1.

Characteristics of the prediction distribution of $y_{n+h|n}$ can then be estimated from the simulated values at a specific forecast horizon: $\boldsymbol{y}_{n+h|n} = \{y_{n+h}^{(1)}, \ldots, y_{n+h}^{(M)}\}$. For example, prediction intervals can be obtained using quantiles of the simulated sample paths. An approximate $100(1-\alpha)\%$ prediction interval for forecast horizon h is given by the $\alpha/2$ and $1-\alpha/2$ quantiles of $\boldsymbol{y}_{n+h|n}$. The solid vertical line on the right of Fig. 6.1 is a 90% prediction interval computed in this way from the 0.05 and 0.95 quantiles of the simulated values at the 16-step horizon.

The full prediction density can be estimated using a kernel density estimator (Silverman 1986) applied to $\boldsymbol{y}_{n+h|n}$. Figure 6.2 shows the prediction

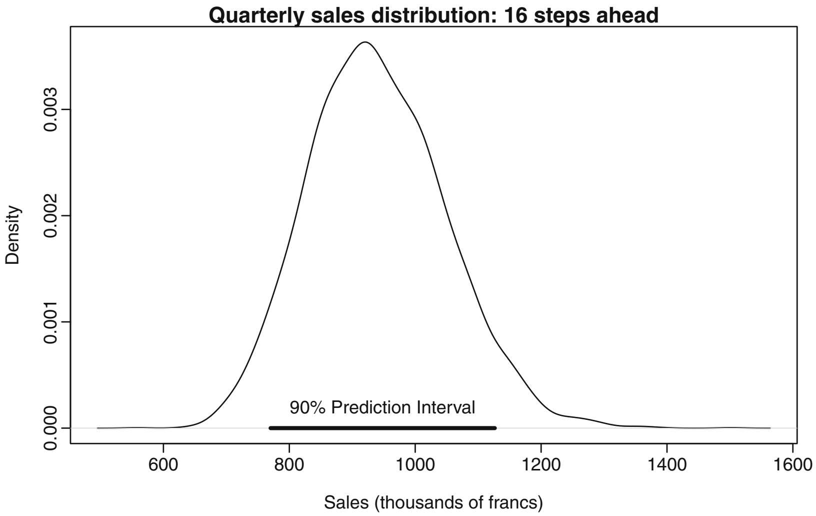

Fig. 6.2. The 16-step forecast density estimated from 5,000 simulated future sample paths. The 90% prediction interval is calculated from the 0.05 and 0.95 quantiles.

density for the data in Fig. 6.1 obtained in this way, along with the 90% prediction interval.

There are several advantages in computing prediction distributions and intervals in this way:

- If the distribution of ε_t is not Gaussian, another distribution can be used to generate the ε_t values when simulating the future sample paths.
- The historical ε_t values can be resampled to give bootstrap prediction distributions without making any distributional assumptions.
- The method can be used for nonlinear models where ε_t may be Gaussian but y_t is not Gaussian.
- The method avoids the complex formulae that are necessary to compute analytical prediction intervals for some nonlinear models.
- For some models (those in Classes 4 and 5), simulation is the only method available for computing prediction distributions and intervals.
- It is possible to take into account the error in estimating the model parameters. In this case, the simulated sample paths are generated using the same model but with randomly varying parameters, reflecting the parameter uncertainty in the fitted model. This was done in Ord et al. (1997) for models with multiplicative error, and in Snyder et al. (2001) for models with additive error.
- The increasing speed of computers makes the simulation approach more viable every year.

6.1.1 Lead-Time Forecasting

In inventory control, forecasts of the sum of the next h observations are often required. These are used for determination of ordering requirements such as reorder levels, order-up-to levels and reorder quantities.

Suppose that a replenishment decision is to be made at the beginning of period $n+1$. Any order placed at this time is assumed to arrive a lead-time later, at the start of period $n+h+1$. Thus, we need to forecast the aggregate of unknown future values y_{n+j}, defined by

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

The problem is to make inferences about the distribution of $Y_n(h)$ which (in the inventory context) is known as the "lead-time demand." The results from the simulation of single periods give the prediction distributions and intervals for individual forecast horizons, but for re-ordering purposes it is more useful to have the lead-time prediction distribution and interval. Because $Y_n(h)$ involves a summation, the central limit theorem states that its distribution will tend towards Gaussianity as h increases. However, for small to moderate h, we need to estimate the distribution.

The simulation approach can easily be used here by computing values of $Y_n(h)$ from the simulated future sample paths. For example, to get the distribution of $Y_n(3)$ for the quarterly French exports data, we sum the first three values of the simulated future sample paths shown in Fig. 6.1. This gives us 5,000 values from the distribution of $Y_n(3)$ (assuming the model is correct). Figure 6.3 shows the density computed from these 5,000 values along with a 90% prediction interval.

Here we have assumed that the lead-time h is fixed. Fixed lead-times are relevant when suppliers make regular deliveries, an increasingly common situation in supply chain management. For stochastic lead-times, we could randomly generate h from a Poisson distribution (or some other count distribution) when simulating values of $Y_n(h)$. This would be used when suppliers make irregular deliveries.

6.2 Class 1: Linear Homoscedastic State Space Models

We now derive some analytical results for the prediction distributions of the linear homoscedastic (Class 1) models. These provide additional insight and can be much quicker to compute than the simulation approach. Derivations of the results in this section are given in Appendix "Derivation of Results for Class 1."

The linear ETS models are (A,N,N), (A,A,N), (A,A_d,N), (A,N,A), (A,A,A) and (A,A_d,A). The forecast means are given in Table 6.2. Because of the linear

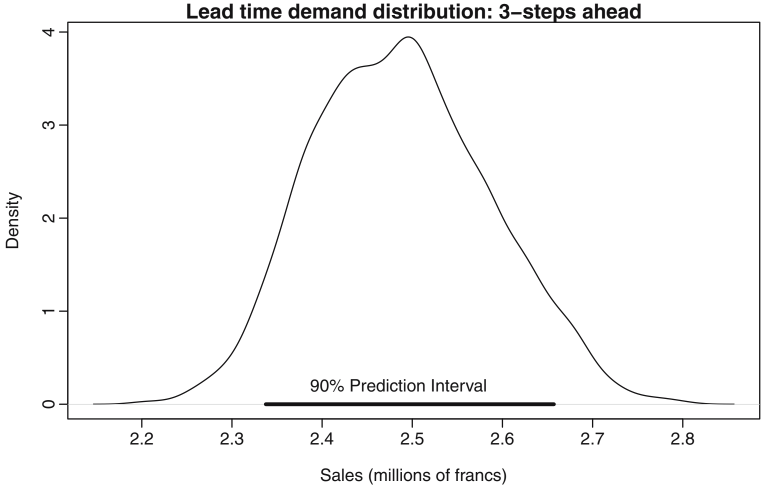

Fig. 6.3. The 3-step lead-time demand density estimated from 5,000 simulated future sample paths assuming Gaussian innovations. The 90% prediction interval is calculated from the 0.05 and 0.95 quantiles.

Table 6.2. Forecast means and c_j values for the linear homoscedastic (Class 1) and linear heteroscedastic (Class 2) state space models.

Model	Forecast mean: $\mu_{n+h\|n}$	c_j
(A,N,N)/(M,N,N)	ℓ_n	α
(A,A,N)/(M,A,N)	$\ell_n + hb_n$	$\alpha + \beta j$
(A,A$_\mathrm{d}$,N)/(M,A$_\mathrm{d}$,N)	$\ell_n + \phi_h b_n$	$\alpha + \beta\phi_j$
(A,N,A)/(M,N,A)	$\ell_n + s_{n-m+h_m^+}$	$\alpha + \gamma d_{j,m}$
(A,A,A)/(M,A,A)	$\ell_n + hb_n + s_{n-m+h_m^+}$	$\alpha + \beta j + \gamma d_{j,m}$
(A,A$_\mathrm{d}$,A)/(M,A$_\mathrm{d}$,A)	$\ell_n + \phi_h b_n + s_{n-m+h_m^+}$	$\alpha + \beta\phi_j + \gamma d_{j,m}$

The values of c_j are used in the forecast variance expressions (6.1) and (6.2). Here, $d_{j,m} = 1$ if $j = 0 \pmod m$ and 0 otherwise, and $\phi_j = \phi + \phi^2 + \cdots + \phi^j$.

structure of the models, the forecast means are identical to the point forecasts given in Table 2.1 (p. 18).

The forecast variances are given by

$$v_{n+h|n} = \mathrm{V}(y_{n+h} \mid \boldsymbol{x}_n) = \begin{cases} \sigma^2 & \text{if } h = 1; \\ \sigma^2 \left[1 + \sum_{j=1}^{h-1} c_j^2\right] & \text{if } h \geq 2; \end{cases} \tag{6.1}$$

where c_j is given in Table 6.2. Note that $v_{n+h|n}$ does not depend on $\boldsymbol{x}_n$ or n, but only on h and the smoothing parameters.

Table 6.3. Forecast variance expressions for each linear homoscedastic state space model, where $v_{n+h|n} = \mathrm{V}(y_{n+h} \mid \boldsymbol{x}_n)$.

Model	Forecast variance: $v_{n+h\|n}$
(A,N,N)	$v_{n+h\|n} = \sigma^2\big[1+\alpha^2(h-1)\big]$
(A,A,N)	$v_{n+h\|n} = \sigma^2\Big[1+(h-1)\big\{\alpha^2+\alpha\beta h+\tfrac{1}{6}\beta^2 h(2h-1)\big\}\Big]$
(A,A$_\mathrm{d}$,N)	$v_{n+h\|n} = \sigma^2\Big[1+\alpha^2(h-1)+\frac{\beta\phi h}{(1-\phi)^2}\{2\alpha(1-\phi)+\beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\Big\{2\alpha(1-\phi^2)+\beta\phi(1+2\phi-\phi^h)\Big\}\Big]$
(A,N,A)	$v_{n+h\|n} = \sigma^2\Big[1+\alpha^2(h-1)+\gamma h_m(2\alpha+\gamma)\Big]$
(A,A,A)	$v_{n+h\|n} = \sigma^2\Big[1+(h-1)\big\{\alpha^2+\alpha\beta h+\tfrac{1}{6}\beta^2 h(2h-1)\big\} + \gamma h_m\big\{2\alpha+\gamma+\beta m(h_m+1)\big\}\Big]$
(A,A$_\mathrm{d}$,A)	$v_{n+h\|n} = \sigma^2\Big[1+\alpha^2(h-1)+\frac{\beta\phi h}{(1-\phi)^2}\{2\alpha(1-\phi)+\beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\Big\{2\alpha(1-\phi^2)+\beta\phi(1+2\phi-\phi^h)\Big\} + \gamma h_m(2\alpha+\gamma) + \frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)}\Big\{h_m(1-\phi^m)-\phi^m(1-\phi^{mh_m})\Big\}\Big]$

Because the models are linear and ε_t is assumed to be Gaussian, $y_{n+h} \mid \boldsymbol{x}_n$ is also Gaussian. Therefore, prediction intervals are easily obtained from the forecast means and variances.

In practice, we would normally substitute the numerical values of c_j from Table 6.2 into (6.1) to obtain numerical values for the variance. However, it is sometimes useful to expand (6.1) algebraically by substituting in the *expressions* for c_j from Table 6.2. The resulting variance expressions are given in Table 6.3.

We note in passing that $v_{n+h|n}$ is linear in h when $\beta = 0$, but cubic in h when $\beta > 0$. Thus, models with non-zero β tend to have prediction intervals that widen rapidly as h increases.

Traditionally, prediction intervals for the linear exponential smoothing methods have been found through heuristic approaches or by employing equivalent or approximate ARIMA models. Where an equivalent ARIMA model exists (see Chap. 11), the results in Table 6.3 provide identical forecast variances to those from the ARIMA model.

State space models with multiple sources of error have also been used to find forecast variances for SES and Holt's method (Harrison 1967; Johnston and Harrison 1986). With these models, the variances are limiting values, although the convergence is rapid. The variance formulae arising from these two cases are the same as in our results.

Prediction intervals for the additive Holt-Winters method have previously been considered by Yar and Chatfield (1990). They assumed that the one-period ahead forecast errors are independent, but they did not assume any particular underlying model for the smoothing methods. The formulae presented here for the ETS(A,A,A) model are equivalent to those given by Yar and Chatfield (1990).

6.3 Class 2: Linear Heteroscedastic State Space Models

Derivations of the results in this section are given in Appendix "Derivation of Results for Class 2."

The ETS models in Class 2 are (M,N,N), (M,A,N), (M,A_d,N), (M,N,A), (M,A,A) and (M,A_d,A). These are similar to those in Class 1 except that multiplicative rather than additive errors are used. Consequently, the forecast means of Class 2 models are identical to the forecast means of the analogous Class 1 model (assuming the same parameters), but the prediction intervals and distributions will be different. The forecast means for Class 2 also coincide with the usual point forecasts. Specific values of the forecast means are given in Table 6.2.

The forecast variance is given by

$$v_{n+h|n} = (1+\sigma^2)\theta_h - \mu^2_{n+h|n}, \tag{6.2}$$

where

$$\theta_1 = \mu^2_{n+1|n} \qquad \text{and} \qquad \theta_h = \mu^2_{n+h|n} + \sigma^2 \sum_{j=1}^{h-1} c_j^2 \theta_{h-j}, \tag{6.3}$$

where each c_j is identical to that for the corresponding additive error model from Class 1 in Table 6.2.

For most models, there is no non-recursive expression for the variance, and we simply substitute the relevant c_j values into (6.2) and (6.3) to obtain numerical expressions for the variance. However, for the ETS(M,N,N) model, we can go a little further (Exercise 6.1).

6.4 Class 3: Some Nonlinear Seasonal State Space Models

Derivations of the results in this section are given in Appendix "Derivation of results for Class 3."

The Class 3 models are (M,N,M), (M,A,M) and (M,A_d,M). These are similar to the seasonal models in Class 2 except that the seasonal component is multiplicative rather than additive.

Table 6.4. Values of $\mu_{n+h|n}$, $\tilde{\mu}_{n+h|n}$ and c_j for the Class 3 models.

	Approx $\mu_{n+h|n}$	$\tilde{\mu}_{n+h|n}$	c_j
ETS(M,N,M)	$\ell_n s_{n-m+h_m^+}$	ℓ_n	α
ETS(M,A,M)	$(\ell_n + hb_n)s_{n-m+h_m^+}$	$\ell_n + hb_n$	$\alpha + \beta j$
ETS(M,A$_d$,M)	$(\ell_n + \phi_h b_n)s_{n-m+h_m^+}$	$\ell_n + \phi_h b_n$	$\alpha + \beta\phi_j$

Here, $\phi_j = \phi + \phi^2 + \cdots + \phi^j$. Values of c_j are used in the forecast variance expressions (6.5).

6.4.1 Approximate Forecast Means and Variances

For these models, the exact forecast means and variances are complicated to compute when $h \geq m$. However, by noting that σ^2 is usually small (much less than 1), we can obtain approximate expressions for the mean and variance which are often useful. Let $\hat{y}_{n+h|n}$ be the usual point forecast as given in Table 2.1. Then,

$$\mu_{n+h|n} \approx \hat{y}_{n+h|n} \tag{6.4}$$

and

$$v_{n+h|n} \approx s^2_{n-m+h_m^+}\Big[\theta_h(1+\sigma^2)(1+\gamma^2\sigma^2)^{h_m} - \tilde{\mu}^2_{n+h|n}\Big], \tag{6.5}$$

where

$$\tilde{\mu}_{n+h|n} = \hat{y}_{n+h|n}/s_{n-m+h_m^+}$$

is the seasonally adjusted point forecast, $\theta_1 = \tilde{\mu}^2_{n+1|n}$, and

$$\theta_h = \tilde{\mu}^2_{n+h|n} + \sigma^2 \sum_{j=1}^{h-1} c_j^2 \theta_{h-j}, \qquad h \geq 2. \tag{6.6}$$

These expressions are exact for $h \leq m$, but are only approximate for $h > m$. The variance formula (6.5) agrees with those in Koehler et al. (2001) and Chatfield and Yar (1991) (who only considered the first year of forecasts).

Specific values for $\mu_{n+h|n}$, $\tilde{\mu}_{n+h|n}$ and c_j for the particular models in Class 3 are given in Table 6.4.

Example 6.1: ETS(M,N,M) model

For the ETS(M,N,M) model, $\theta_1 = \ell_n^2$, and for $h \geq 2$,

$$\begin{aligned}\theta_h &= \ell_n^2 + \alpha^2\sigma^2 \sum_{j=1}^{h-1} \theta_{h-j} \\ &= \ell_n^2 + \alpha^2\sigma^2(\theta_1 + \theta_2 + \cdots + \theta_{h-1}).\end{aligned}$$

Then, by induction, we can show that $\theta_h = \ell_n^2(1+\alpha^2\sigma^2)^{h-1}$. Plugging this into (6.5) gives the following simpler expression for $v_{n+h|n}$:

$$v_{n+h|n} \approx s^2_{n-m+h_m^+}\ell_n^2\Big[(1+\sigma^2)(1+\alpha^2\sigma^2)^{h-1}(1+\gamma^2\sigma^2)^{h_m}-1\Big].$$

The expression is exact for $h \le m$.

6.4.2 Exact Forecast Means and Variances

To obtain the exact formulae for $h > m$, we first write the models in Class 3 using the following nonlinear state space model:

$$\begin{aligned} y_t &= \boldsymbol{w}_1'\boldsymbol{x}_{t-1}\boldsymbol{w}_2'\boldsymbol{z}_{t-1}(1+\varepsilon_t),\\ \boldsymbol{x}_t &= (\boldsymbol{F}_1+\boldsymbol{G}_1\varepsilon_t)\boldsymbol{x}_{t-1},\\ \boldsymbol{z}_t &= (\boldsymbol{F}_2+\boldsymbol{G}_2\varepsilon_t)\boldsymbol{z}_{t-1}, \end{aligned}$$

where $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, $\boldsymbol{w}_1'$ and $\boldsymbol{w}_2'$ are all matrix or vector coefficients, and $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ are unobserved state vectors at time t. As for Class 2, $\{\varepsilon_t\}$ is $\text{NID}(0,\sigma^2)$, where the lower tail of the distribution is truncated so that $1+\varepsilon_t$ is positive.

Let k be the length of vector $\boldsymbol{x}_t$ and q be the length of vector $\boldsymbol{z}_t$. Then the orders of the above matrices are as follows:

$$\begin{array}{lll} \boldsymbol{F}_1\ (k\times k) & \boldsymbol{G}_1\ (k\times k) & \boldsymbol{w}_1'\ (1\times k)\\ \boldsymbol{F}_2\ (q\times q) & \boldsymbol{G}_2\ (q\times q) & \boldsymbol{w}_2'\ (1\times q) \end{array}$$

- For the ETS(M,N,M) model, $\boldsymbol{x}_t = \ell_t$, $\boldsymbol{z}_t = (s_t,\dots,s_{t-m+1})'$, and the matrix coefficients are $\boldsymbol{w}_1 = 1$, $\boldsymbol{w}_2' = [0,\dots,0,1]$,

$$\boldsymbol{F}_1 = 1,\quad \boldsymbol{F}_2 = \begin{bmatrix} \mathbf{0}'_{m-1} & 1\\ I_{m-1} & \mathbf{0}_{m-1}\end{bmatrix},\quad \boldsymbol{G}_1=\alpha,\quad \text{and}\quad \boldsymbol{G}_2 = \begin{bmatrix} \mathbf{0}'_{m-1} & \gamma\\ \boldsymbol{O}_{m-1} & \mathbf{0}_{m-1}\end{bmatrix}.$$

- For the ETS(M,A$_\text{d}$,M) model, $\boldsymbol{x}_t = (\ell_t, b_t)'$, $\boldsymbol{w}_1' = [1,1]$,

$$\boldsymbol{F}_1 = \begin{bmatrix} 1 & \phi\\ 0 & \phi\end{bmatrix},\quad \boldsymbol{G}_1 = \begin{bmatrix} \alpha & \alpha\\ \beta & \beta\end{bmatrix},$$

  and $\boldsymbol{z}_2$, $\boldsymbol{w}_2$, $\boldsymbol{F}_2$ and $\boldsymbol{G}_2$ are the same as for the ETS(M,N,M) model.
- The ETS(M,A,M) model is equivalent to the ETS(M,A$_\text{d}$,M) model with $\phi = 1$.

For models in this class,

$$\mu_{n+h|n} = \boldsymbol{w}_1' \boldsymbol{M}_{h-1} \boldsymbol{w}_2 \tag{6.7}$$

and

$$v_{n+h|n} = (1+\sigma^2)(\boldsymbol{w}_2' \otimes \boldsymbol{w}_1') \boldsymbol{V}_{n+h-1|n} (\boldsymbol{w}_2' \otimes \boldsymbol{w}_1')' + \sigma^2 \mu^2_{n+h|n}, \tag{6.8}$$

where $\otimes$ denotes a Kronecker product (Schott 2005, Sect. 8.2), $\boldsymbol{M}_0 = \boldsymbol{x}_n \boldsymbol{z}_n'$, $\boldsymbol{V}_0 = \boldsymbol{O}_{2m}$, and for $h \geq 1$,

$$\boldsymbol{M}_h = \boldsymbol{F}_1 \boldsymbol{M}_{h-1} \boldsymbol{F}_2' + \boldsymbol{G}_1 \boldsymbol{M}_{h-1} \boldsymbol{G}_2' \sigma^2 \tag{6.9}$$

and

$$\begin{aligned}
\boldsymbol{V}_{n+h|n} &= (\boldsymbol{F}_2 \otimes \boldsymbol{F}_1) \boldsymbol{V}_{n+h-1|n} (\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)' \\
&+ \sigma^2 \Big[(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1) \boldsymbol{V}_{n+h-1|n} (\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)' + (\boldsymbol{G}_2 \otimes \boldsymbol{G}_1) \boldsymbol{V}_{n+h-1|n} (\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)' \Big] \\
&+ \sigma^2 (\boldsymbol{G}_2 \otimes \boldsymbol{F}_1 + \boldsymbol{F}_2 \otimes \boldsymbol{G}_1) \Big[\boldsymbol{V}_{n+h-1|n} + \overrightarrow{\boldsymbol{M}}_{h-1} \overrightarrow{\boldsymbol{M}}'_{h-1} \Big] (\boldsymbol{G}_2 \otimes \boldsymbol{F}_1 + \boldsymbol{F}_2 \otimes \boldsymbol{G}_1)' \\
&+ \sigma^4 (\boldsymbol{G}_2 \otimes \boldsymbol{G}_1) \Big[3\boldsymbol{V}_{n+h-1|n} + 2\overrightarrow{\boldsymbol{M}}_{h-1} \overrightarrow{\boldsymbol{M}}'_{h-1} \Big] (\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)',
\end{aligned} \tag{6.10}$$

where $\overrightarrow{\boldsymbol{M}}_{h-1} = \text{vec}(\boldsymbol{M}_{h-1})$. (That is, the columns of $\boldsymbol{M}_{h-1}$ are stacked to form a vector.) Note, in particular, that $\mu_{n+1|n} = (\boldsymbol{w}_1' \boldsymbol{x}_n)(\boldsymbol{w}_2' \boldsymbol{z}_n)$ and $v_{n+1|n} = \sigma^2 \mu^2_{n+1|n}$. While these expressions look complicated and provide little insight, it is relatively easy to compute them using computer matrix languages such as **R** and Matlab.

In Appendix "Derivation of results for Class 3," we show that the approximations (6.4) and (6.5) follow from the exact expressions (6.7) and (6.8). Note that the usual point forecasts for these models are given by (6.4) rather than (6.7).

6.4.3 The Accuracy of the Approximations

In order to investigate the accuracy of the approximations (6.4) and (6.5) for the exact mean and variance given by (6.7) and (6.8), we provide some comparisons for the ETS(M,A,M) model in Class 3.

These comparisons are done for quarterly data, where the values for the components are assumed to be the following: $\ell_n = 100$, $b_n = 2$, $s_n = 0.80$, $s_{n-1} = 1.20$, $s_{n-2} = 0.90$ and $s_{n-3} = 1.10$. We use the following base level values for the parameters: $\alpha = 0.2$, $\beta = 0.06$, $\gamma = 0.1$, and $\sigma = 0.05$. We vary these parameters one at a time as shown in Table 6.5.

The results in Table 6.5 show that the mean and approximate mean are always very close, and that the percentage difference in the standard

Table 6.5. Comparison of exact and approximate means and standard deviations for the ETS(M,A,M) model in Class 3.

Period ahead h	Exact mean (6.7) $\mu_{n+h\|n}$	Approximate mean (6.4)	Exact SD (6.8) $\sqrt{v_{n+h\|n}}$	Approximate SD (6.5)	SD percent difference
$\sigma = 0.05$, $\alpha = 0.2$, $\beta = 0.06$, $\gamma = 0.1$					
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
$\sigma = \mathbf{0.1}$, $\alpha = 0.2$, $\beta = 0.06$, $\gamma = 0.1$					
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
$\sigma = 0.05$, $\alpha = \mathbf{0.6}$, $\beta = 0.06$, $\gamma = 0.1$					
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
$\sigma = 0.05$, $\alpha = 0.2$, $\beta = \mathbf{0.18}$, $\gamma = 0.1$					
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
$\sigma = 0.05$, $\alpha = 0.2$, $\beta = 0.06$, $\gamma = \mathbf{0.3}$					
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

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.

6.5 Prediction Intervals

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

In applying these formulae, the maximum likelihood estimator for σ^2 (see p. 68) is simply

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

where $\hat{\varepsilon}_t = y_t - \mu_{t|t-1}$.

The prediction distributions for Classes 2 and 3 are non-Gaussian because of the nonlinearity of the state space equations. However, prediction intervals based on the above (Gaussian) formula will usually give reasonably accurate results, as the following example shows. In cases where the Gaussian approximation may be unreasonable, it is necessary to use the simulation approach of Sect. 6.1.

6.5.1 Application: Quarterly French Exports

As a numerical example, we consider the quarterly French exports data given in Fig. 6.1, and use the ETS(M,A,M) model. We estimate the parameters to be $\alpha = 0.8185$, $\beta = 0.01$, $\gamma = 0.01$ and $\sigma = 0.0352$, with the final states $\ell_n = 757.3$, $b_n = 15.7$, and $\boldsymbol{z}_n = (0.873, 1.141, 1.022, 0.964)'$.

Figure 6.4 shows the forecast standard deviations calculated exactly using (6.8) and approximately using (6.5). The approximate values are so close to the exact values in this case (because σ^2 and γ are both very small) that it is almost impossible to distinguish the two lines.

The data with three years of forecasts are shown in Fig. 6.5. In this case, the conditional mean forecasts obtained from model ETS(M,A,M) 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 $\mu_{n+h|n} \pm 1.96\sqrt{v_{n+h|n}}$, 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.

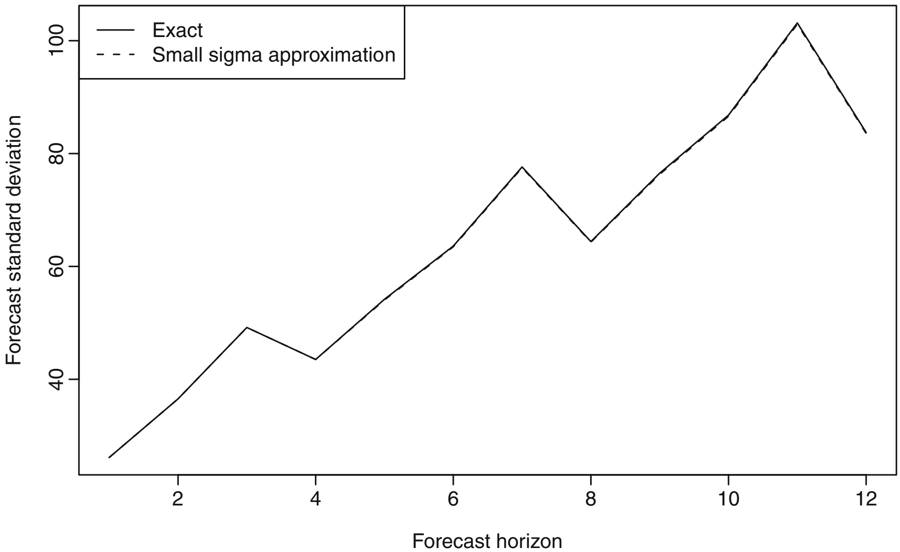

Fig. 6.4. Forecast standard deviations calculated (**a**) exactly using (6.8); and (**b**) approximately using (6.5).

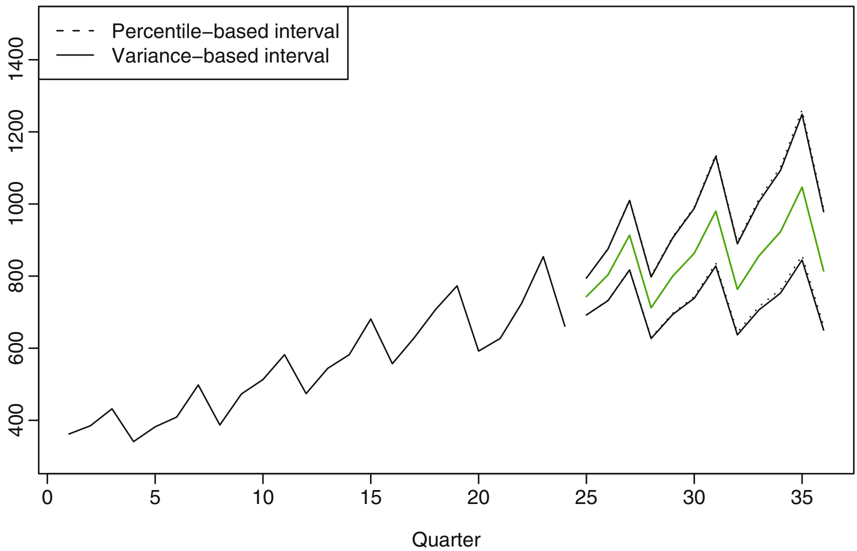

Fig. 6.5. Quarterly French exports data with 3 years of forecasts. The *solid lines* show prediction intervals calculated as $\mu_{n+h|n} \pm 1.96\sqrt{v_{n+h|n}}$, and the *dotted lines* show prediction intervals computed by generating 20,000 future sample paths from the fitted model and finding the 2.5 and 97.5% quantiles at each forecast horizon.

Clearly, the variance-based intervals are a good approximation despite the non-Gaussianity of the prediction distributions.

6.6 Lead-Time Demand Forecasts for Linear Homoscedastic Models

For Class 1 models, it is also possible to obtain some analytical results on the distribution of lead-time demand, defined by

$$Y_n(h) = \sum_{j=1}^{h} y_{n+j}. \tag{6.11}$$

In particular, the variance of lead-time demand can be used when implementing an inventory strategy, although the basic exponential smoothing procedures originally provided only point forecasts, and rather ad hoc formulae were the vogue in inventory control software.

Harrison (1967) and Johnston and Harrison (1986) derived a variance formula for lead-time demand based on simple exponential smoothing using a state space model with independent error terms. They utilized the fact that simple exponential smoothing emerges as the steady state form of the model predictions in large samples. Adopting a different model, Snyder et al. (1999) were able to obtain the same formula without recourse to a restrictive large sample assumption. Around the same time, Graves (1999) obtained the formula using an ARIMA(0,1,1) model.

Harrison (1967) and Johnston and Harrison (1986) also obtained a variance formula for lead-time demand when trend-corrected exponential smoothing is employed. Yar and Chatfield (1990), however, suggested a slightly different formula. They also provide a formula that incorporates seasonal effects for use with the additive Holt-Winters method.

The approach we adopt here is based on Snyder et al. (2004), although the parameterization in this book is slightly different from that used in Snyder et al. (2004). The results obtained subsume those in Harrison (1967), Johnston and Harrison (1986), Yar and Chatfield (1990), Graves (1999) and Snyder et al. (1999). In addition, for ETS(A,A,A), the recursive variance formula in Yar and Chatfield (1990) has been replaced with a closed-form counterpart.

6.6.1 Means and Variances of Lead-Time Demand

In Appendix "Derivation of C_j values" we show that

$$y_{n+j} = \mu_{n+j|n} + \sum_{i=1}^{j-1} c_{j-i}\varepsilon_{n+i} + \varepsilon q_{n+j},$$

where $\mu_{n+j|n}$ and c_k are given in Table 6.2. Substitute this into (6.11) to give

$$Y_n(h) = \sum_{j=1}^{h} \Big(\mu_{n+j|n} + \sum_{i=1}^{j-1} c_{j-i}\varepsilon_{n+i} + \varepsilon_{n+j}\Big) = \sum_{j=1}^{h} \mu_{n+j|n} + \sum_{j=1}^{h} C_{j-1}\varepsilon_{n+h-j+1}, \tag{6.12}$$

where

$$C_0 = 1 \quad \text{and} \quad C_j = 1 + \sum_{i=1}^{j} c_i \quad \text{for} \quad j = 1, \dots, h-1. \tag{6.13}$$

Thus, lead-time demand can be resolved into a linear function of the uncorrelated level and error components.

From (6.12), it is easy to see that the point forecast (conditional mean) is simply

$$\hat{Y}_n(h) = \mathrm{E}(Y_n(h) \mid \boldsymbol{x}_n) = \sum_{j=1}^{h} \mu_{n+j|n} \tag{6.14}$$

and the conditional variance is given by

$$\mathrm{V}\left(Y_n(h) \mid \boldsymbol{x}_n\right) = \sigma^2 \sum_{j=0}^{h-1} C_j^2. \tag{6.15}$$

The value of C_j for each of the models is given in Table 6.6. These expressions are derived in Appendix "Derivation of C_j values."

As with the equations for forecast variance at a specific forecast horizon, we can substitute these expressions into (6.15) to derive a specific formula for each model. This leads to a lot of tedious algebra that is of limited value. Therefore we only give the result for model ETS(A,N,N):

Table 6.6. Values of C_j to be used in computing the lead-time variance in (6.15).

Model	C_j
(A,N,N)	$1 + j\alpha$
(A,A,N)	$1 + j\left[\alpha + \frac{1}{2}\beta(j+1)\right]$
(A,A$_\mathrm{d}$,N)	$1 + j\alpha + \frac{\beta\phi}{(1-\phi)^2}\left[(j+1)(1-\phi) - (1-\phi^{j+1})\right]$
(A,N,A)	$1 + j\alpha + \gamma j_m$
(A,A,A)	$1 + j\left[\alpha + \frac{1}{2}\beta(j+1)\right] + \gamma j_m$
(A,A$_\mathrm{d}$,A)	$1 + j\alpha + \frac{\beta\phi}{(1-\phi)^2}\left[(j+1)(1-\phi) - (1-\phi^{j+1})\right] + \gamma j_m$

Here m is the number of periods in each season and $j_m = \lfloor j/m \rfloor$ is the number of complete seasonal cycles that occur within j time periods.

$$\begin{aligned}
\mathrm{V}(Y_n(h) \mid \boldsymbol{x}_n) &= \sum_{j=0}^{h-1}(1+j\alpha)^2 \\
&= \sigma^2 h\left[1+\alpha(h-1)+\tfrac{1}{6}\alpha^2(h-1)(2h-1)\right]. \qquad (6.16)
\end{aligned}$$

6.6.2 Matrix Calculation of Means and Variances

The mean and variance of the lead-time demand, and the forecast mean and variance for a single period, can also be computed recursively using matrix equations. From Chap. 3, we know that the form of the Class 1 models is

$$\begin{aligned}
y_t &= \boldsymbol{w}'\boldsymbol{x}_{t-1} + \varepsilon_t, \\
\boldsymbol{x}_t &= \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t,
\end{aligned}$$

where $\boldsymbol{w}'$ is a row vector, $\boldsymbol{g}$ is a column vector, $\boldsymbol{F}$ is a matrix, $\boldsymbol{x}_t$ is the unobserved state vector at time t, and $\{\varepsilon_t\}$ is $\mathrm{NID}(0,\sigma^2)$.

Observe that the lead-time demand can be determined recursively by

$$Y_n(j) = Y_n(j-1) + y_{n+j}, \qquad (6.17)$$

where $Y_n(0) = 0$ and $Y_n(j) = \sum_{i=1}^{j} y_{n+i}$. Consequently, (6.17) can be written as

$$Y_n(j) = Y_n(j-1) + \boldsymbol{w}'\boldsymbol{x}_{n+j-1} + \varepsilon_{n+j}. \qquad (6.18)$$

So, if the state vector $\boldsymbol{x}_{n+j}$ is augmented with $Y_n(j)$, the first-order recurrence relationship

$$\begin{bmatrix} \boldsymbol{x}_{n+j} \\ Y_n(j) \end{bmatrix} = \begin{bmatrix} \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{w}' & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{n+j-1} \\ Y_n(j-1) \end{bmatrix} + \begin{bmatrix} \boldsymbol{g} \\ 1 \end{bmatrix} \varepsilon_{n+j}$$

is obtained. This has the general form $\boldsymbol{z}_{n+j} = \boldsymbol{A}\boldsymbol{z}_{n+j-1} + \boldsymbol{b}\varepsilon_{n+j}$. If the mean and variance of the $\boldsymbol{z}_{n+j}$ are denoted by $\boldsymbol{m}^z_{n+j|n} = \mathrm{E}(\boldsymbol{z}_{n+j} \mid \boldsymbol{x}_n)$ and $\boldsymbol{V}^z_{n+j|n} = \mathrm{V}(\boldsymbol{z}_{n+j} \mid \boldsymbol{x}_n)$, then they can be computed recursively using the equations

$$\begin{aligned}
\boldsymbol{m}^z_{n+j|n} &= \boldsymbol{A}\boldsymbol{m}^z_{n+j-1|n}, \\
\boldsymbol{V}^z_{n+j|n} &= \boldsymbol{A}\boldsymbol{V}^z_{n+j-1|n}\boldsymbol{A}' + \sigma^2\boldsymbol{b}\boldsymbol{b}'.
\end{aligned}$$

The mean of the lead-time demand $Y_n(h)$ is the last element in $\boldsymbol{m}^z_{n+h|n}$, and the variance of $Y_n(h)$ is the bottom right element of $\boldsymbol{V}^z_{n+h|n}$.

This same procedure of using an augmented matrix can also be applied to find the forecast mean and variance of y_{n+h} for any single future time period $t = n+h$. In this case, the state vector $\boldsymbol{x}_{n+j}$ is augmented with y_{n+j} in place of $Y_n(j)$, and

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{w}' & 0 \end{bmatrix}.$$

Then, the mean and variance of y_{n+h} are the last elements in $\boldsymbol{m}^z_{n+h|n}$ and $\boldsymbol{V}^z_{n+h|n}$ respectively. Of course, one can use $\boldsymbol{A} = [\boldsymbol{F}, \boldsymbol{w}']'$ and the general form $\boldsymbol{z}_{n+j} = \boldsymbol{A}\boldsymbol{x}_{n+j-1} + \boldsymbol{b}\varepsilon_{n+j}$ to remove the unnecessary multiplications by 0 in an actual implementation.

6.6.3 Stochastic Lead-Times

In practice, lead-times are often stochastic, depending on various factors including demand in the previous time periods. We explore the effect of stochastic lead-times on forecast variances in the case of the ETS(A,N,N) model for simple exponential smoothing.

Let the lead-time, T, be stochastic with mean $\mathrm{E}(T) = h$. The mean lead-time demand, given the level at time n, is

$$\mathrm{E}(Y_n(T) \mid \ell_n) = \mathrm{E}_T[\mathrm{E}(Y_n(T) \mid T, \ell_n)] = h\ell_n,$$

as in the case of a fixed lead-time. The variance of the lead-time demand reduces to

$$\begin{aligned}
\mathrm{V}(Y_n(T) \mid \ell_n) &= \mathrm{V}_T[\mathrm{E}(Y_n(T) \mid T, \ell_n)] + \mathrm{E}_T[\mathrm{V}(Y_n(T) \mid T, \ell_n)] \\
&= \mathrm{V}_T(\ell_n T) + \mathrm{E}_T\left[\sigma^2 \sum_{j=1}^{T} C_{j,T}^2\right] \\
&= \ell_n^2 \mathrm{V}(T) + \sigma^2 \mathrm{E}_T\left[\sum_{j=1}^{T} \left\{1 + 2\alpha(T-j) + \alpha^2(T-j)^2\right\}\right] \\
&= \ell_n^2 \mathrm{V}(T) + \sigma^2 h + \sigma^2\alpha\left[(1 + \tfrac{1}{2}\alpha)h_{[2]} + \tfrac{1}{3}\alpha h_{[3]}\right],
\end{aligned}$$

where $h_{[j]} = \mathrm{E}[T(T-1)\dots(T-j+1)]$, $j = 1, 2, \dots$, is known as the jth factorial moment of the distribution of T.

For example, when the lead-time is fixed, $h_{[j]} = h(h-1)\dots(h-j+1)$. When the lead-time is Poisson with mean h, then $h_{[j]} = h^j$. Therefore, the lead-time demand variance becomes

$$\mathrm{V}(Y_n(T) \mid \ell_n) = (\ell_n^2 + \sigma^2)h + \sigma^2\alpha\left[(1 + \tfrac{1}{2}\alpha)h^2 + \tfrac{1}{3}\alpha h^3\right].$$

Compare this with the variance for a fixed lead-time as given in (6.16). The two variances are plotted in Fig. 6.6 for $\alpha = 0.1$, $\sigma = 1$ and $\ell_n = 2$, showing that a stochastic lead-time can substantially increase the lead-time demand variance.

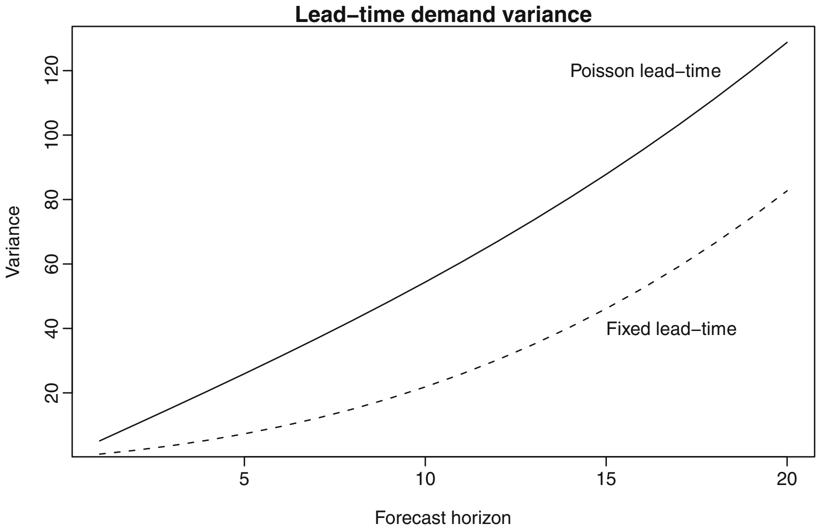

Fig. 6.6. Lead-time demand variance for an ETS(A,N,N) model with fixed and stochastic lead-times. Here, $\alpha = 0.1$, $\sigma = 1$ and $\ell_n = 2$.

6.7 Exercises

Exercise 6.1. For the ETS(M,N,N) model, show that

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

and

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

Exercise 6.2. For the ETS(A,A,A) model, use (6.23) replacing ϕ_j by j to show that

$$v_{n+h|n} = \sigma^2 \Big[1 + (h-1) \{ \alpha^2 + \alpha\beta h + \tfrac{1}{6}\beta^2 h(2h-1) \} \\ + \gamma h_m \{ 2\alpha + \gamma + \beta m(h_m + 1) \} \Big].$$

Exercise 6.3. Monthly US 10-year bonds data were forecast with an ETS(A,A_d,N) model in Sect. 2.8.1 (p. 28). Find the 95% prediction intervals for this model algebraically and compare the results obtained by simulating 5,000 future sample paths using **R**.

Exercise 6.4. Quarterly UK passenger vehicle production data were forecast with an ETS(A,N,A) model in Sect. 2.8.1 (p. 28). Find the 95% prediction intervals for this model algebraically and compare the results obtained by simulating 5,000 future sample paths using **R**.

Appendix: Derivations

Derivation of Results for Class 1

The results for Class 1 models are obtained by first noting that all of the linear, homoscedastic ETS models can be written using the following linear state space model, introduced in Chap. 3:

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1} + \varepsilon_t \tag{6.19}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t, \tag{6.20}$$

where $\boldsymbol{w}'$ is a row vector, $\boldsymbol{g}$ is a column vector, $\boldsymbol{F}$ is a matrix, and $\boldsymbol{x}_t$ is the unobserved state vector at time t. In each case, $\{\varepsilon_t\}$ is $\text{NID}(0, \sigma^2)$.

Let $\boldsymbol{I}_k$ denote the $k \times k$ identity matrix, and $\mathbf{0}_k$ denote a zero vector of length k. Then

- The ETS(A,N,N) model has $\boldsymbol{x}_t = \ell_t$, $\boldsymbol{w} = \boldsymbol{F} = 1$ and $\boldsymbol{g} = \alpha$;
- The ETS(A,A$_\text{d}$,N) model has $\boldsymbol{x}_t = (\ell_t, b_t)'$, $\boldsymbol{w}' = [1 \;\; \phi]$,

$$\boldsymbol{F} = \begin{bmatrix} 1 & \phi \\ 0 & \phi \end{bmatrix} \quad \text{and} \quad \boldsymbol{g} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix};$$

- The ETS(A,N,A) model has $\boldsymbol{x}_t = (\ell_t, s_t, s_{t-1}, \ldots, s_{t-(m-1)})'$, $\boldsymbol{w}' = [1 \; \mathbf{0}'_{m-1} \; 1]$,

$$\boldsymbol{F} = \begin{bmatrix} 1 & \mathbf{0}'_{m-1} & 0 \\ 0 & \mathbf{0}'_{m-1} & 1 \\ \mathbf{0}_{m-1} & \boldsymbol{I}_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{g} = \begin{bmatrix} \alpha \\ \gamma \\ \mathbf{0}_{m-1} \end{bmatrix};$$

- The ETS(A,A$_\text{d}$,A) model has $\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-(m-1)})'$, $\boldsymbol{w}' = [1 \; \phi \; \mathbf{0}'_{m-1} \; 1]$,

$$\boldsymbol{F} = \begin{bmatrix} 1 & \phi & \mathbf{0}'_{m-1} & 0 \\ 0 & \phi & \mathbf{0}'_{m-1} & 0 \\ 0 & 0 & \mathbf{0}'_{m-1} & 1 \\ \mathbf{0}_{m-1} & \mathbf{0}_{m-1} & \boldsymbol{I}_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{g} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \mathbf{0}_{m-1} \end{bmatrix}.$$

The matrices for (A,A,N) and (A,A,A) are the same as for (A,A$_\text{d}$,N) and (A,A$_\text{d}$,A) respectively, but with $\phi = 1$.

Forecast Mean

Let $\boldsymbol{m}_{n+h|n} = \text{E}(\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n)$. Then $\boldsymbol{m}_{n|n} = \boldsymbol{x}_n$ and

$$\boldsymbol{m}_{n+h|n} = \boldsymbol{F}\boldsymbol{m}_{n+h-1|n} = \boldsymbol{F}^2\boldsymbol{m}_{n+h-2|n} = \cdots = \boldsymbol{F}^h\boldsymbol{m}_{n|n} = \boldsymbol{F}^h\boldsymbol{x}_n.$$

Therefore

$$\mu_{n+h|n} = \mathrm{E}(y_{n+h}|\boldsymbol{x}_n) = \boldsymbol{w}'\boldsymbol{m}_{n+h-1|n} = \boldsymbol{w}'\boldsymbol{F}^{h-1}\boldsymbol{x}_n.$$

Example 6.2: Forecast mean of the ETS(A,A$_d$,A) model

For the ETS(A,A$_\mathrm{d}$,A) model, $\boldsymbol{w}' = [1\ \phi\ \boldsymbol{0}'_{m-1}\ 1]$ and

$$\boldsymbol{F}^j = \begin{bmatrix} 1 & \phi_j & 0 & 0 & \dots & 0 \\ 0 & \phi^j & 0 & 0 & \dots & 0 \\ 0 & 0 & d_{j+m,m} & d_{j+m+1,m} & \dots & d_{j+2m-1,m} \\ 0 & 0 & d_{j+m-1,m} & d_{j+m,m} & \dots & d_{j+2m-2,m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & d_{j+1,m} & d_{j+2,m} & \dots & d_{j+m,m} \end{bmatrix},$$

where $\phi_j = \phi + \phi^2 + \cdots + \phi^j$, and $d_{k,m} = 1$ if $k = 0 \pmod{m}$ and $d_{k,m} = 0$ otherwise. Therefore,

$$\boldsymbol{w}'\boldsymbol{F}^j = [1, \phi_{j+1}, d_{j+1,m}, d_{j+2,m}, \dots, d_{j+m,m}] \tag{6.21}$$

and

$$\mu_{n+h|n} = \ell_n + \phi_h b_n + s_{n-m+h_m^+}.$$

The forecast means for the other models can be derived similarly, and are listed in Table 6.2

Forecast Variance

Define the state forecast variance as $\boldsymbol{V}_{n+h|n} = \mathrm{V}(\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n)$. Note that $\boldsymbol{V}_{n|n} = \boldsymbol{O}$, where $\boldsymbol{O}$ denotes a matrix of zeros. Then, from (6.20),

$$\boldsymbol{V}_{n+h|n} = \boldsymbol{F}\boldsymbol{V}_{n+h-1|n}\boldsymbol{F}' + \boldsymbol{g}\boldsymbol{g}'\sigma^2,$$

and therefore

$$\boldsymbol{V}_{n+h|n} = \sigma^2 \sum_{j=0}^{h-1} \boldsymbol{F}^j\boldsymbol{g}\boldsymbol{g}'(\boldsymbol{F}^j)'.$$

Hence, using (6.19), the forecast variance for h periods ahead is

$$\begin{aligned} v_{n+h|n} &= \mathrm{V}(y_{n+h} \mid \boldsymbol{x}_n) \\ &= \boldsymbol{w}'\boldsymbol{V}_{n+h-1|n}\boldsymbol{w} + \sigma^2 = \begin{cases} \sigma^2 & \text{if } h = 1; \\ \sigma^2 \left[1 + \sum_{j=1}^{h-1} c_j^2\right] & \text{if } h \geq 2; \end{cases} \end{aligned} \tag{6.22}$$

where $c_j = \boldsymbol{w}'\boldsymbol{F}^{j-1}\boldsymbol{g}$.

Example 6.3: Forecast variance for the ETS(A,A$_d$,A) model

Using (6.21), we find that $c_j = \boldsymbol{w}'\boldsymbol{F}^{j-1}\boldsymbol{g} = \alpha + \beta\phi_j + \gamma d_{j,m}$. Consequently, from (6.22) we obtain

$$v_{n+h|n} = \sigma^2\left[1 + \sum_{j=1}^{h-1}(\alpha + \beta\phi_j + \gamma d_{j,m})^2\right]$$
$$= \sigma^2\left[1 + \sum_{j=1}^{h-1}\Big(\alpha^2 + 2\alpha\beta\phi_j + \beta^2\phi_j^2 + \{\gamma^2 + 2\alpha\gamma + 2\beta\gamma\phi_j\}d_{j,m}\Big)\right]. \quad (6.23)$$

In order to expand this expression, first recall the following well known results for arithmetic and geometric series (Morgan 2005):

$$\sum_{j=1}^{p} j = \tfrac{1}{2}p(p+1), \quad \sum_{j=1}^{p} j^2 = \tfrac{1}{6}p(p+1)(2p+1) \quad \text{and} \quad \sum_{j=1}^{p} a^j = \frac{a(1-a^p)}{1-a},$$

where $a \neq 1$, from which it is easy to show that

$$\sum_{j=1}^{p} ja^j = \frac{a[1-(p+1)a^p + pa^{p+1}]}{(1-a)^2}, \qquad \sum_{j=1}^{p} j(p-j+1) = \tfrac{1}{6}p(p+1)(p+2)$$

and $\phi_j = \phi(1-\phi^j)/(1-\phi)$ when $\phi < 1$. Then the following expressions also follow for $\phi < 1$:

$$\sum_{j=1}^{h-1}\phi_j = \frac{\phi}{(1-\phi)^2}\left[h(1-\phi) - (1-\phi^h)\right]$$

and

$$\sum_{j=1}^{h-1}\phi_j^2 = \frac{\phi^2}{(1-\phi)^2}\sum_{j=1}^{h-1}(1 - 2\phi^j + \phi^{2j})$$
$$= \frac{\phi^2}{(1-\phi)^2(1-\phi^2)}\left[h(1-\phi^2) - (1+2\phi-\phi^h)(1-\phi^h)\right].$$

Furthermore, $\sum_{j=1}^{h-1} d_{j,m} = h_m$. If $h-1 < m$ (i.e., $h_m = 0$), then $\sum_{j=1}^{h-1}\phi_j d_{j,m} = 0$, and if $h-1 \geq m$ (i.e., $h_m \geq 1$), then

$$\sum_{j=1}^{h-1}\phi_j d_{j,m} = \sum_{\ell=1}^{h_m}\phi_{\ell m} = \frac{\phi}{1-\phi}\sum_{\ell=1}^{h_m}(1-\phi^{\ell m})$$
$$= \frac{\phi}{(1-\phi)(1-\phi^m)}\left[h_m(1-\phi^m) - \phi^m(1-\phi^{mh_m})\right].$$

(continued)

Using the above results, we can rewrite (6.23) as

$$v_{n+h|n} = \sigma^2\Bigg[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2}\{2\alpha(1-\phi)+\beta\phi\} \qquad (6.24)$$
$$- \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\left\{2\alpha(1-\phi^2)+\beta\phi(1+2\phi-\phi^h)\right\}$$
$$+ \gamma h_m(2\alpha+\gamma) + \frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)}\left\{h_m(1-\phi^m)-\phi^m(1-\phi^{mh_m})\right\}\Bigg].$$

This is the forecast variance for the ETS(A,A_d,A) model when $h \geq 2$.

Example 6.4: Forecast variance for the ETS(A,A,A) model

To obtain the forecast variance for the ETS(A,A,A) model, we could take the limit as $\phi \to 1$ in (6.24) and apply L'Hospital's rule. However, in many ways it is simpler to go back to (6.23) and replace ϕ_j with j. This yields (Exercise 6.2)

$$v_{n+h|n} = \sigma^2\Big[1 + (h-1)\{\alpha^2 + \alpha\beta h + \tfrac{1}{6}\beta^2 h(2h-1)\} \qquad (6.25)$$
$$+ \gamma h_m\{2\alpha + \gamma + \beta m(h_m+1)\}\Big].$$

The forecast variance expressions for all other models can be obtained as special cases of either (6.24) or (6.25):

- For (A,A_d,N), we use the results of (A,A_d,A) with $\gamma = 0$ and $s_t = 0$ for all t.
- For (A,A,N), we use the results of (A,A,A) with $\gamma = 0$ and $s_t = 0$ for all t.
- The results for (A,N,N) are obtained from (A,A,N) by further setting $\beta = 0$ and $b_t = 0$ for all t.
- The results for (A,N,A) are obtained as a special case of (A,A,A) with $\beta = 0$ and $b_t = 0$ for all t.

Derivation of Results for Class 2

The models in Class 2 can all be written using the following state space model:

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1}(1+\varepsilon_t), \qquad (6.26)$$
$$\boldsymbol{x}_t = (\boldsymbol{F} + \boldsymbol{g}\boldsymbol{w}'\varepsilon_t)\boldsymbol{x}_{t-1}, \qquad (6.27)$$

where $\boldsymbol{w}$, $\boldsymbol{g}$, $\boldsymbol{F}$, $\boldsymbol{x}_t$ and ε_t are the same as for the corresponding Class 1 model. The lower tail of the error distribution is truncated so that $1+\varepsilon_t$ is positive.

The truncation is usually negligible as σ is usually relatively small for these models.

Let $\boldsymbol{m}_{n+h|n} = \mathrm{E}(\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n)$ and $\boldsymbol{V}_{n+h|n} = \mathrm{V}(\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n)$ as in Sect. 6.2. The forecast means for Class 2 have the same form as for Class 1, namely

$$\mu_{n+h|n} = \boldsymbol{w}'\boldsymbol{m}_{n+h-1|n} = \boldsymbol{w}'\boldsymbol{F}^{h-1}\boldsymbol{x}_n.$$

From (6.26), it can be seen that the forecast variance is given by

$$\begin{aligned} v_{n+h|n} &= \boldsymbol{w}'\boldsymbol{V}_{n+h-1|n}\boldsymbol{w}(1+\sigma^2) + \sigma^2\boldsymbol{w}'\boldsymbol{m}_{n+h-1|n}\boldsymbol{m}'_{n+h-1|n}\boldsymbol{w} \\ &= \boldsymbol{w}'\boldsymbol{V}_{n+h-1|n}\boldsymbol{w}(1+\sigma^2) + \sigma^2\mu^2_{n+h|n}. \end{aligned}$$

To obtain $\boldsymbol{V}_{n+h-1|n}$, first note that $\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}e_t$, where $e_t = y_t - \boldsymbol{w}'\boldsymbol{x}_{t-1} = \boldsymbol{w}'\boldsymbol{x}_{t-1}\varepsilon_t$. Then it is readily seen that $\boldsymbol{V}_{n+h|n} = \boldsymbol{F}\boldsymbol{V}_{n+h-1|n}\boldsymbol{F}' + \boldsymbol{g}\boldsymbol{g}'\mathrm{V}(e_{n+h} \mid \boldsymbol{x}_n)$. Now let θ_h be defined such that $\mathrm{V}(e_{n+h} \mid \boldsymbol{x}_n) = \theta_h\sigma^2$. Then, by repeated substitution,

$$\boldsymbol{V}_{n+h|n} = \sigma^2 \sum_{j=0}^{h-1} \boldsymbol{F}^j\boldsymbol{g}\boldsymbol{g}'(\boldsymbol{F}^j)'\theta_{h-j}.$$

Therefore,

$$\boldsymbol{w}'\boldsymbol{V}_{n+h-1|n}\boldsymbol{w} = \sigma^2 \sum_{j=1}^{h-1} c_j^2\theta_{h-j}, \tag{6.28}$$

where $c_j = \boldsymbol{w}'\boldsymbol{F}^{j-1}\boldsymbol{g}$. Now

$$e_{n+h} = \left[\boldsymbol{w}'(\boldsymbol{x}_{n+h-1} - \boldsymbol{m}_{n+h-1|n}) + \boldsymbol{w}'\boldsymbol{m}_{n+h-1|n}\right]\varepsilon_{n+h},$$

which we square and take expectations to give $\theta_h = \boldsymbol{w}'\boldsymbol{V}_{n+h-1|n}\boldsymbol{w} + \mu^2_{n+h|n}$. Substituting (6.28) into this expression for θ_h gives

$$\theta_h = \sigma^2 \sum_{j=1}^{h-1} c_j^2\theta_{h-j} + \mu^2_{n+h|n}, \tag{6.29}$$

where $\theta_1 = \mu^2_{n+1|n}$. The forecast variance is then given by

$$v_{n+h|n} = (1+\sigma^2)\theta_h - \mu^2_{n+h|n}. \tag{6.30}$$

Derivation of Results for Class 3

Note that we can write (see p. 85)

$$y_t = \boldsymbol{w}_1'\boldsymbol{x}_{t-1}\boldsymbol{z}'_{t-1}\boldsymbol{w}_2(1+\varepsilon_t).$$

So let $\boldsymbol{Q}_h = \boldsymbol{x}_{n+h}\boldsymbol{z}'_{n+h}$, $\boldsymbol{M}_h = \mathrm{E}(\boldsymbol{Q}_h \mid \boldsymbol{x}_n, \boldsymbol{z}_n)$ and $\boldsymbol{V}_{n+h|n} = \mathrm{V}(\overrightarrow{\boldsymbol{Q}}_h \mid \boldsymbol{x}_n, \boldsymbol{z}_n)$ where $\overrightarrow{\boldsymbol{Q}}_h = \mathrm{vec}(\boldsymbol{Q}_h)$. Note that

$$\begin{aligned}
\boldsymbol{Q}_h &= (\boldsymbol{F}_1\boldsymbol{x}_{n+h-1} + \boldsymbol{G}_1\boldsymbol{x}_{n+h-1}\varepsilon_{n+h})(\boldsymbol{z}'_{n+h-1}\boldsymbol{F}'_2 + \boldsymbol{z}'_{n+h-1}\boldsymbol{G}'_2\varepsilon_{n+h}) \\
&= \boldsymbol{F}_1\boldsymbol{Q}_{h-1}\boldsymbol{F}'_2 + (\boldsymbol{F}_1\boldsymbol{Q}_{h-1}\boldsymbol{G}'_2 + \boldsymbol{G}_1\boldsymbol{Q}_{h-1}\boldsymbol{F}'_2)\varepsilon_{n+h} + \boldsymbol{G}_1\boldsymbol{Q}_{h-1}\boldsymbol{G}'_2\varepsilon^2_{n+h}.
\end{aligned}$$

It follows that $\boldsymbol{M}_0 = \boldsymbol{x}_n\boldsymbol{z}'_n$ and

$$\boldsymbol{M}_h = \boldsymbol{F}_1\boldsymbol{M}_{h-1}\boldsymbol{F}'_2 + \boldsymbol{G}_1\boldsymbol{M}_{h-1}\boldsymbol{G}'_2\sigma^2. \tag{6.31}$$

For the variance of $\boldsymbol{Q}_h$, we find $\boldsymbol{V}_0 = 0$, and

$$\begin{aligned}
\boldsymbol{V}_{n+h|n} = \mathrm{V}\big[&\mathrm{vec}(\boldsymbol{F}_1\boldsymbol{Q}_{h-1}\boldsymbol{F}'_2) + \mathrm{vec}(\boldsymbol{F}_1\boldsymbol{Q}_{h-1}\boldsymbol{G}'_2 + \boldsymbol{G}_1\boldsymbol{Q}_{h-1}\boldsymbol{F}'_2)\varepsilon_{n+h} \\
&+ \mathrm{vec}(\boldsymbol{G}_1\boldsymbol{Q}_{h-1}\boldsymbol{G}'_2)\varepsilon^2_{n+h}\big] \\
= {}& (\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)\boldsymbol{V}_{n+h-1|n}(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)' \\
&+ (\boldsymbol{G}_2 \otimes \boldsymbol{F}_1 + \boldsymbol{F}_2 \otimes \boldsymbol{G}_1)\mathrm{V}(\overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon_{n+h})(\boldsymbol{G}_2 \otimes \boldsymbol{F}_1 + \boldsymbol{F}_2 \otimes \boldsymbol{G}_1)' \\
&+ (\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)\mathrm{V}(\overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon^2_{n+h})(\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)' \\
&+ (\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)\mathrm{Cov}(\overrightarrow{\boldsymbol{Q}}_{h-1}, \overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon^2_{n+h})(\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)' \\
&+ (\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)\mathrm{Cov}(\overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon^2_{n+h}, \overrightarrow{\boldsymbol{Q}}_{h-1})(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)'.
\end{aligned}$$

Next we find that

$$\begin{aligned}
\mathrm{V}(\overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon_{n+h}) &= \mathrm{E}[\overrightarrow{\boldsymbol{Q}}_{h-1}(\overrightarrow{\boldsymbol{Q}}_{h-1})'\varepsilon^2_{n+h}] \\
&= \sigma^2\big[\boldsymbol{V}_{n+h-1|n} + \overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})'\big], \\
\mathrm{V}(\overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon^2_{n+h}) &= \mathrm{E}\big[\overrightarrow{\boldsymbol{Q}}_{h-1}(\overrightarrow{\boldsymbol{Q}}_{h-1})'\varepsilon^4_{n+h}\big] - \mathrm{E}(\overrightarrow{\boldsymbol{Q}}_{h-1})\mathrm{E}(\overrightarrow{\boldsymbol{Q}}_{h-1})'\sigma^4 \\
&= 3\sigma^4\big[\boldsymbol{V}_{n+h-1|n} + \overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})'\big] - \overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})'\sigma^4 \\
&= \sigma^4\big[3\boldsymbol{V}_{n+h-1|n} + 2\overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})'\big],
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}(\overrightarrow{\boldsymbol{Q}}_{h-1}, \overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon^2_{n+h}) &= \mathrm{E}\big[\overrightarrow{\boldsymbol{Q}}_{h-1}(\overrightarrow{\boldsymbol{Q}}_{h-1})'\varepsilon^2_{n+h}\big] - \mathrm{E}(\overrightarrow{\boldsymbol{Q}}_{h-1})\mathrm{E}(\overrightarrow{\boldsymbol{Q}}_{h-1})'\sigma^2 \\
&= \sigma^2(\boldsymbol{V}_{n+h-1|n} + \overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})') - \sigma^2\overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})' \\
&= \sigma^2\boldsymbol{V}_{n+h-1|n}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\boldsymbol{V}_{n+h|n} &= (\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)\boldsymbol{V}_{n+h-1|n}(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)' \\
&\quad + \sigma^2\Big[(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)\boldsymbol{V}_{n+h-1|n}(\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)' + (\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)\boldsymbol{V}_{n+h-1|n}(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)'\Big]
\end{aligned}$$

$$+ \sigma^2(\boldsymbol{G}_2 \otimes \boldsymbol{F}_1 + \boldsymbol{F}_2 \otimes \boldsymbol{G}_1)\Big[\boldsymbol{V}_{n+h-1|n} + \overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})'\Big]$$
$$\times (\boldsymbol{G}_2 \otimes \boldsymbol{F}_1 + \boldsymbol{F}_2 \otimes \boldsymbol{G}_1)'$$
$$+ \sigma^4(\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)\Big[3\boldsymbol{V}_{n+h-1|n} + 2\overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})'\Big](\boldsymbol{G}_2 \otimes \boldsymbol{G}_1)'.$$

The forecast mean and variance are given by

$$\mu_{n+h|n} = \mathrm{E}(y_{n+h} \mid \boldsymbol{x}_n, \boldsymbol{z}_n) = \boldsymbol{w}_1' \boldsymbol{M}_{h-1} \boldsymbol{w}_2$$

and

$$\begin{aligned} v_{n+h|n} &= \mathrm{V}(y_{n+h} \mid \boldsymbol{x}_n, \boldsymbol{z}_n) = \mathrm{V}[\mathrm{vec}(\boldsymbol{w}_1' \boldsymbol{Q}_{h-1} \boldsymbol{w}_2 + \boldsymbol{w}_1' \boldsymbol{Q}_{h-1} \boldsymbol{w}_2' \varepsilon_{n+h})] \\ &= \mathrm{V}[(\boldsymbol{w}_2' \otimes \boldsymbol{w}_1')\overrightarrow{\boldsymbol{Q}}_{h-1} + (\boldsymbol{w}_2' \otimes \boldsymbol{w}_1')\overrightarrow{\boldsymbol{Q}}_{h-1}\varepsilon_{n+h}] \\ &= (\boldsymbol{w}_2' \otimes \boldsymbol{w}_1')[\boldsymbol{V}_{n+h-1|n}(1+\sigma^2) + \sigma^2 \overrightarrow{\boldsymbol{M}}_{h-1}(\overrightarrow{\boldsymbol{M}}_{h-1})'](\boldsymbol{w}_2 \otimes \boldsymbol{w}_1) \\ &= (1+\sigma^2)(\boldsymbol{w}_2' \otimes \boldsymbol{w}_1')\boldsymbol{V}_{n+h-1|n}(\boldsymbol{w}_2' \otimes \boldsymbol{w}_1')' + \sigma^2 \mu^2_{n+h|n}. \end{aligned}$$

When σ is sufficiently small (much less than 1), it is possible to obtain some simpler but approximate expressions. The second term in (6.31) can be dropped to give $\boldsymbol{M}_h = \boldsymbol{F}_1^{h-1}\boldsymbol{M}_0(\boldsymbol{F}_2^{h-1})'$, and so

$$\mu_{n+h|n} \approx \boldsymbol{w}_1' \boldsymbol{F}_1^{h-1} \boldsymbol{x}_n (\boldsymbol{w}_2' \boldsymbol{F}_2^{h-1} \boldsymbol{z}_n)'.$$

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 $u_{1,t} = \boldsymbol{w}_1'\boldsymbol{x}_{t-1}$, $u_{2,t} = \boldsymbol{w}_2'\boldsymbol{z}_{t-1}$ and $u_{3,t} = 1 + \varepsilon_t$. Then

$$\mathrm{E}(y_t) = \mathrm{E}(u_{1,t}u_{2,t}u_{3,t}) = \mathrm{E}(u_{1,t}u_{2,t})\mathrm{E}(u_{3,t}),$$

because $u_{3,t}$ is independent of $u_{1,t}$ and $u_{2,t}$. Therefore, because $\mathrm{E}(u_{1,t}u_{2,t}) = \mathrm{E}(u_{1,t})\mathrm{E}(u_{2,t}) + \mathrm{Cov}(u_{1,t}, u_{2,t})$, we have the approximation:

$$\mu_{n+h|n} = \mathrm{E}(y_{n+h} \mid \boldsymbol{x}_n, \boldsymbol{z}_n) = \mathrm{E}(u_{1,n+h} \mid \boldsymbol{x}_n)\mathrm{E}(u_{2,n+h} \mid \boldsymbol{z}_n)\mathrm{E}(u_{3,n+h}) + O(\sigma^2).$$

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

$$\begin{aligned} \mu_{1,h} &= \mathrm{E}(u_{1,n+h+1} \mid \boldsymbol{x}_n) = \mathrm{E}(\boldsymbol{w}_1'\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n) = \boldsymbol{w}_1'\boldsymbol{F}_1^h\boldsymbol{x}_n, \\ \mu_{2,h} &= \mathrm{E}(u_{2,n+h+1} \mid \boldsymbol{z}_n) = \mathrm{E}(\boldsymbol{w}_2'\boldsymbol{z}_{n+h} \mid \boldsymbol{z}_n) = \boldsymbol{w}_2'\boldsymbol{F}_2^h\boldsymbol{z}_n, \\ v_{1,h} &= \mathrm{V}(u_{1,n+h+1} \mid \boldsymbol{x}_n) = \mathrm{V}(\boldsymbol{w}_1'\boldsymbol{x}_{n+h} \mid \boldsymbol{x}_n), \\ v_{2,h} &= \mathrm{V}(u_{2,n+h+1} \mid \boldsymbol{z}_n) = \mathrm{V}(\boldsymbol{w}_2'\boldsymbol{z}_{n+h} \mid \boldsymbol{z}_n), \\ \text{and} \qquad v_{12,h} &= \mathrm{Cov}(u^2_{1,n+h+1}, u^2_{2,n+h+1} \mid \boldsymbol{x}_n, \boldsymbol{z}_n) \\ &= \mathrm{Cov}([\boldsymbol{w}_1'\boldsymbol{x}_{n+h}]^2, [\boldsymbol{w}_2'\boldsymbol{z}_{n+h}]^2 \mid \boldsymbol{x}_n, \boldsymbol{z}_n). \end{aligned}$$

Then

$$\mu_{n+h|n} = \mu_{1,h-1}\mu_{2,h-1} + O(\sigma^2) = \boldsymbol{w}_1' \boldsymbol{F}_1^{h-1} \boldsymbol{x}_n \boldsymbol{w}_2' \boldsymbol{F}_2^{h-1} \boldsymbol{z}_n + O(\sigma^2).$$

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

$$\begin{aligned}
&\mathrm{E}(y_{n+h}^2 \mid \boldsymbol{z}_n, \boldsymbol{x}_n) = \mathrm{E}(u_{1,n+h}^2 u_{2,n+h}^2 \mid \boldsymbol{x}_n, \boldsymbol{z}_n)\mathrm{E}(u_{3,n+h}^2) \\
&= \Big[\mathrm{Cov}(u_{1,n+h}^2, u_{2,n+h}^2 \mid \boldsymbol{x}_n, \boldsymbol{z}_n) + \mathrm{E}(u_{1,n+h}^2 \mid \boldsymbol{x}_n)\mathrm{E}(u_{2,n+h}^2 \mid \boldsymbol{z}_n)\Big]\,\mathrm{E}(u_{3,n+h}^2) \\
&= (1+\sigma^2)[v_{12,h-1} + (v_{1,h-1} + \mu_{1,h-1}^2)(v_{2,h-1} + \mu_{2,h-1}^2)].
\end{aligned}$$

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

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

We now simplify these results for the ETS(M,A$_\mathrm{d}$,M) case where $\boldsymbol{x}_t = (\ell_t, b_t)'$ and $\boldsymbol{z}_t = (s_t, \dots, s_{t-m+1})'$, and the matrix coefficients are $\boldsymbol{w}_1' = [1, \phi]$, $\boldsymbol{w}_2' = [0, \dots, 0, 1]$,

$$\boldsymbol{F}_1 = \begin{bmatrix} 1 & \phi \\ 0 & \phi \end{bmatrix}, \quad \boldsymbol{F}_2 = \begin{bmatrix} \mathbf{0}_{m-1}' & 1 \\ I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix},$$

$$\boldsymbol{G}_1 = \begin{bmatrix} \alpha & \alpha \\ \beta & \beta \end{bmatrix}, \quad \text{and} \quad \boldsymbol{G}_2 = \begin{bmatrix} \mathbf{0}_{m-1}' & \gamma \\ \boldsymbol{O}_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}.$$

Many terms will be zero in the formulae for the expected value and the variance because of the following relationships: $\boldsymbol{G}_2^2 = \boldsymbol{O}_m$, $\boldsymbol{w}_2'\boldsymbol{G}_2 = \mathbf{0}_m'$, and $(\boldsymbol{w}_2' \otimes \boldsymbol{w}_1')(\boldsymbol{G}_2 \otimes \boldsymbol{X}) = \mathbf{0}_{2m}'$ where $\boldsymbol{X}$ is any 2×2 matrix. For the terms that remain, $\boldsymbol{w}_2' \otimes \boldsymbol{w}_1'$ and its transpose will only use the terms from the last two rows of the last two columns of the large matrices because $\boldsymbol{w}_2' \otimes \boldsymbol{w}_1' = [\mathbf{0}_{2m-2}', 1, 1]$.

Using the small σ approximations and exploiting the structure of the ETS(M,A$_\mathrm{d}$,M) model, we can obtain simpler expressions that approximate $\mu_{n+h|n}$ and $v_{n+h|n}$.

Note that $\boldsymbol{w}_2' \boldsymbol{F}_2^j \boldsymbol{G}_2 = \gamma d_{j+1,m} \boldsymbol{w}_2'$. So, for $h < m$, we have

$$\boldsymbol{w}_2' \boldsymbol{z}_{n+h} \mid \boldsymbol{z}_n = \boldsymbol{w}_2' \prod_{j=1}^{h} (\boldsymbol{F}_2 + \boldsymbol{G}_2 \varepsilon_{n+h-j+1}) \boldsymbol{z}_n = \boldsymbol{w}_2' \boldsymbol{F}_2^h \boldsymbol{z}_n = s_{n-m+h+1}$$

Furthermore,

$$\begin{aligned}
&\mu_{2,h} = s_{n-m+h_m^+} \\
\text{and} \qquad &v_{2,h} = [(1+\gamma^2\sigma^2)^{h_m} - 1] s_{n-m+h_m^+}^2.
\end{aligned}$$

Also note that $\boldsymbol{x}_n$ has the same properties as for ETS(M,A$_d$,N) in Class 2. Thus

$$\begin{aligned} \mu_{1,h} &= \ell_n + \phi_h b_n \\ \text{and} \qquad v_{1,h} &= (1+\sigma^2)\theta_h - \mu_{1,h}^2. \end{aligned}$$

Combining all of the terms, we arrive at the approximations

$$\begin{aligned} \mu_{n+h|n} &= \tilde{\mu}_{n+h|n} s_{n-m+h_m^+} + O(\sigma^2) \\ \text{and} \qquad v_{n+h|n} &\approx s_{n-m+h_m^+}^2 \Big[\theta_h (1+\sigma^2)(1+\gamma^2\sigma^2)^{h_m} - \tilde{\mu}_{n+h|n}^2\Big], \end{aligned}$$

where $\tilde{\mu}_{n+h|n} = \ell_n + \phi_h b_n$, $\theta_1 = \tilde{\mu}_{n+1|n}^2$, and

$$\theta_h = \tilde{\mu}_{n+h|n}^2 + \sigma^2 \sum_{j=1}^{h-1} (\alpha + \beta\phi_j)^2 \theta_{h-j}, \qquad h \geq 2.$$

These expressions are exact for $h \leq m$. The other cases of Class 3 can be derived as special cases of ETS(M,A$_d$,M).

Derivation of C_j Values

We first demonstrate that for Class 1 models, lead-time demand can be resolved into a linear function of the uncorrelated level and error components. Back-solve the transition equation (6.20) from period $n+j$ to period n, to give

$$\boldsymbol{x}_{n+j} = \boldsymbol{F}^j \boldsymbol{x}_n + \sum_{i=1}^{j} \boldsymbol{F}^{j-i} \boldsymbol{g} \varepsilon_{n+i}.$$

Now from (6.19) and (6.20) we have

$$\begin{aligned} y_{n+j} &= \boldsymbol{w}' \boldsymbol{x}_{n+j-1} + \varepsilon_{n+j} \\ &= \boldsymbol{w}' \boldsymbol{F} \boldsymbol{x}_{n+j-2} + \boldsymbol{w}' \boldsymbol{g} \varepsilon_{n+j-1} + \varepsilon_{n+j} \\ &\vdots \\ &= \boldsymbol{w}' \boldsymbol{F}^{j-1} \boldsymbol{x}_n + \sum_{i=1}^{j-1} \boldsymbol{w}' \boldsymbol{F}^{j-i-1} \boldsymbol{g} \varepsilon_{n+i} + \varepsilon_{n+j} \\ &= \mu_{n+j|n} + \sum_{i=1}^{j-1} c_{j-i} \varepsilon_{n+i} + \varepsilon_{n+j}, \end{aligned}$$

where $c_k = \boldsymbol{w}' \boldsymbol{F}^{k-1} \boldsymbol{g}$. Substituting this into (6.11) gives (6.15).

To derive the value of C_j for the ETS(A,A$_\text{d}$,A) model, we plug the value of c_i from Table 6.2 into (6.13) to obtain

$$\begin{aligned}
C_j &= 1 + \sum_{i=1}^{j} (\alpha + \beta\phi_i + \gamma d_{i,m}) \\
&= 1 + \alpha j + \beta \sum_{i=1}^{j} \phi_i + \gamma \sum_{i=1}^{j} d_{i,m} \\
&= 1 + \alpha j + \frac{\beta\phi}{(1-\phi)^2} \Big[(j+1)(1-\phi) - (1-\phi^{j+1}) \Big] + \gamma j_m,
\end{aligned}$$

where $j_m = \lfloor j/m \rfloor$ is the number of complete seasonal cycles that occur within j time periods.

A similar derivation for the ETS(A,A,A) model leads to

$$C_j = 1 + \sum_{i=1}^{j} (\alpha + i\beta + \gamma d_{i,m}) = 1 + j\Big[\alpha + \tfrac{1}{2}\beta(j+1)\Big] + \gamma j_m.$$

The expressions for C_j for the other linear models are obtained as special cases of either ETS(A,A$_\text{d}$,A) or ETS(A,A,A) and are given in Table 6.6.