

## 7.5 Innovations state space models for exponential smoothing

---

In the rest of this chapter, we study the statistical models that underlie the exponential smoothing methods we have considered so far. The exponential smoothing methods presented in Table 7.6 are algorithms which generate point forecasts. The statistical models in this section generate the same point forecasts, but can also generate prediction (or forecast) intervals. A statistical model is a stochastic (or random) data generating process that can produce an entire forecast distribution. We will also describe how to use the model selection criteria introduced in Chapter 5 to choose the model in an objective manner.

Each model consists of a measurement equation that describes the observed data, and some state equations that describe how the unobserved components or states (level, trend, seasonal) change over time. Hence, these are referred to as **state space models**.

For each method there exist two models: one with additive errors and one with multiplicative errors. The point forecasts produced by the models are identical if they use the same smoothing parameter values. They will, however, generate different prediction intervals.

To distinguish between a model with additive errors and one with multiplicative errors (and also to distinguish the models from the methods), we add a third letter to the classification of Table 7.5. We label each state space model as ETS(·, ·, ·, ·, ·, ·) for (Error, Trend, Seasonal). This label can also be thought of as ExponentTial Smoothing. Using the same notation as in Table 7.5, the possibilities for each component are: Error = {= {A,M}}, Trend = {= {N,A,A<sub>d</sub>}<sub>d</sub>} and Seasonal = {= {N,A,M}}.

### ETS(A,N,N): simple exponential smoothing with additive errors

Recall the component form of simple exponential smoothing:

Forecast equation	$\hat{y}_{t+1 t} = \ell_t$
Smoothing equation	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$

If we re-arrange the smoothing equation for the level, we get the “error correction” form:

$$\begin{aligned}\ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t\end{aligned}$$

where  $e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$  is the residual at time  $t$ .

The training data errors lead to the adjustment of the estimated level throughout the smoothing process for  $t = 1, \dots, T$ . For example, if the error at time  $t$  is negative, then  $y_t < \hat{y}_{t|t-1}$  and so the level at time  $t-1$  has been over-estimated. The new level  $\ell_t$  is then the previous level  $\ell_{t-1}$  adjusted downwards. The closer  $\alpha$  is to one, the “rougher” the estimate of the level (large adjustments take place). The smaller the  $\alpha$ , the “smoother” the level (small adjustments take place).

We can also write  $y_t = \ell_{t-1} + e_t$ , so that each observation can be represented by the previous level plus an error. To make this into an innovations state space model, all we need to do is specify the probability distribution for  $e_t$ . For a model with additive errors, we assume that residuals (the one-step training errors)  $e_t$  are normally distributed white noise with mean 0 and variance  $\sigma^2$ . A short-hand notation for this is  $e_t = \varepsilon_t \sim \text{NID}(0, \sigma^2)$ ; NID stands for “normally and independently distributed.”

Then the equations of the model can be written as

$$\begin{aligned}y_t &= \ell_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + \alpha \varepsilon_t.\end{aligned}$$

We refer to (7.3) as the *measurement* (or observation) equation and (7.4) as the *state* (or transition) equation. These two equations, together with the statistical distribution of the errors, form a fully specified statistical model. Specifically, these constitute an innovations state space model underlying simple exponential smoothing.

The term “innovations” comes from the fact that all equations use the same random error process,  $\varepsilon_t$ . For the same reason, this formulation is also referred to as a “single source of error” model. There are alternative multiple source of error formulations which we do not present here.

The measurement equation shows the relationship between the observations and the unobserved states. In this case, observation  $y_t$  is a linear function of the level  $\ell_{t-1}$ , the predictable part of  $y_t$ , and the error  $\varepsilon_t$ , the unpredictable part of  $y_t$ . For other innovations state space models, this relationship may be nonlinear.

The state equation shows the evolution of the state through time. The influence of the smoothing parameter  $\alpha$  is the same as for the methods discussed earlier. For example,  $\alpha$  governs the amount of change in successive levels: high values of  $\alpha$  allow rapid changes in the level; low values of  $\alpha$  lead to smooth changes. If  $\alpha = 0$ , the level of the series does not change over time; if  $\alpha = 1$ , the model reduces to a random walk model,  $y_t = y_{t-1} + \varepsilon_t$ . (See Section 8.1 for a discussion of this model.)

## ETS(M,N,N): simple exponential smoothing with multiplicative errors

In a similar fashion, we can specify models with multiplicative errors by writing the one-step-ahead training errors as relative errors:

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$$

where  $\varepsilon_t \sim \text{NID}(0, \sigma^2)$ . Substituting  $\hat{y}_{t|t-1} = \ell_{t-1}$  gives  $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$  and  $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$ .

Then we can write the multiplicative form of the state space model as

$$\begin{aligned} y_t &= \ell_{t-1}(1 + \varepsilon_t) \\ \ell_t &= \ell_{t-1}(1 + \alpha\varepsilon_t). \end{aligned}$$

## ETS(A,A,N): Holt's linear method with additive errors

For this model, we assume that the one-step-ahead training errors are given by  $\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim \text{NID}(0, \sigma^2)$ . Substituting this into the error correction equations for Holt's linear method we obtain

$$\begin{aligned}
 y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\
 \ell_t &= \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t \\
 b_t &= b_{t-1} + \beta \varepsilon_t,
 \end{aligned}$$

where, for simplicity, we have set  $\beta = \alpha$ .

## ETS(M,A,N): Holt's linear method with multiplicative errors

Specifying one-step-ahead training errors as relative errors such that

$$\varepsilon_t = \frac{y_t - (\ell_{t-1} + b_{t-1})}{(\ell_{t-1} + b_{t-1})}$$

and following an approach similar to that used above, the innovations state space model underlying Holt's linear method with multiplicative errors is specified as

$$\begin{aligned}
 y_t &= (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t) \\
 \ell_t &= (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t) \\
 b_t &= b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t
 \end{aligned}$$

where again  $\beta = \alpha$  and  $\varepsilon_t \sim \text{NID}(0, \sigma^2)$ .

## Other ETS models

In a similar fashion, we can write an innovations state space model for each of the exponential smoothing methods of Table 7.6. Table 7.7 presents the equations for all of the models in the ETS framework.

Table 7.7: State space equations for each of the models in the ETS framework.

## ADDITIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
<b>N</b>	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
<b>A</b>	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
<b>A<sub>d</sub></b>	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

## MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
<b>N</b>	$y_t = \ell_{t-1} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha (\ell_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$
<b>A</b>	$y_t = (\ell_{t-1} + b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$
<b>A<sub>d</sub></b>	$y_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$