Time Series Analysis: Part 2

Fan Liu
liufan@nju.edu.cn
School of Management & Engineering

Getting Started

Now at the end of time period t, we have data $y_1, y_2, \ldots, y_t$ and are going to forecast y_{t+1}.

One naïve method is to let $\hat{y}_{t+1} = y_t$. It means that the forecast of the value at time $t+1$ is determined by the actual value at time t.

Another simple method is to make use of the simple moving average

$$\hat{y}_{t+1} = \frac{1}{N}(y_t + y_{t-1} + \cdots + y_{t-N+1})$$

Can we design one method between the previous two methods?

Simple Exponential Smoothing

First, I think it is necessary to take all the previous values into account.

Second, the earlier the observation is the less effect it has on the forecast.

Thus introducing the simple exponential smoother or exponential smoothing method

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \ldots + \alpha(1-\alpha)^{t-1}y_1 + (1-\alpha)^t l_0$$

At time 0, the forecast of y_1 is l_0 .

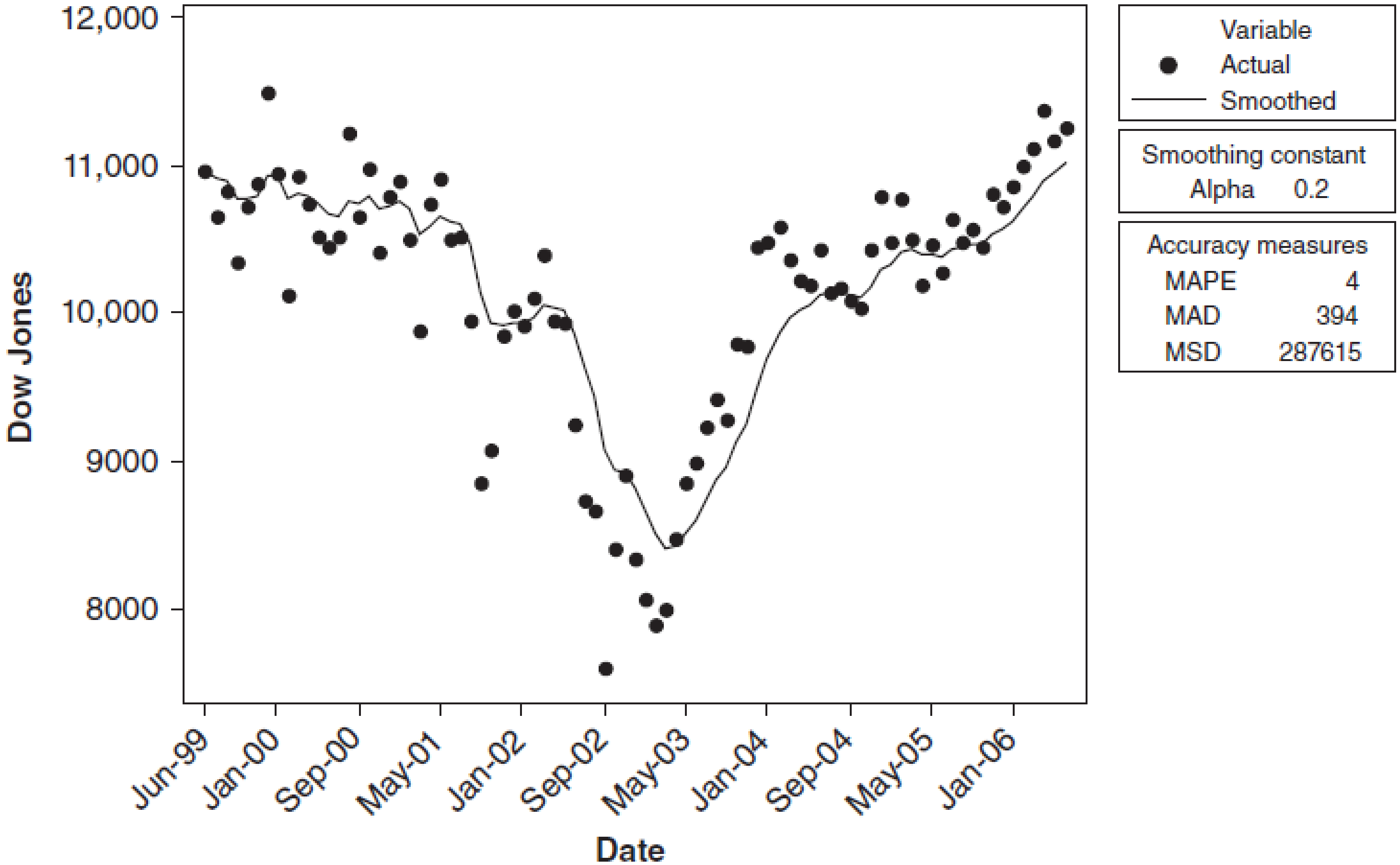
At time 1, the forecast of y_2 is $\alpha y_1 + (1-\alpha)l_0$.

At time 2, the forecast of y_3 is $\alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2 l_0$.

……

Simple Exponential Smoothing

It is a smoother.



Simple Exponential Smoothing

The weight decays exponentially. So it is called as exponential smoothing method.

Look at the formula.

$$\hat{y}_{t+1} = \alpha y_t + \boxed{\alpha(1-\alpha)y_{t-1} + \ldots + \alpha(1-\alpha)^{t-1}y_1 + (1-\alpha)^t l_0}$$

The red part is in fact

$$(1-\alpha)[\alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + \cdots + \alpha(1-\alpha)^{t-2}y_1 + (1-\alpha)^{t-1}l_0]$$

Which equals to $(1-\alpha)\hat{y}_t$.

Simple Exponential Smoothing

So the simple exponential smoothing can be written as

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

The forecast is based on weighting the most recent observation y_t with a weight value α, and weighting the most recent forecast $\hat{y}_t$ with a weight of $(1 - \alpha)$ .Thus, it can be interpreted as a weighted average of the most recent forecast and the most recent observation.

Define a new variable l_t measuring the level of the series. So

$$\hat{y}_{t+1} = l_t \quad l_t = \alpha y_t + (1 - \alpha)l_{t-1}$$

We “update” the information of the level of the series at time t when the actual value is observed, and then use the level to forecast.

Simple Exponential Smoothing

Another format of the SES is

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)$$

It can be seen that the new forecast is simply the old forecast plus an adjustment for the error that occurred in the last forecast.

Let $e_t = y_t - \hat{y}_t$, then the SES can be written using error term

$$\hat{y}_{t+1} = \hat{y}_t + \alpha e_t$$

The unknown parameters involved are α and l_0.

Estimation

The parameters involved is usually estimated by minimizing or maximizing some objective functions.

For example, there are researchers selecting the parameters which minimize the inventory cost.

MSE (mean squared error) is the most commonly applied measure. And the error of MSE in most cases is chosen to be one-step ahead in-sample forecast errors.

Suppose now T observations are included. Then we forecast y_t based on all the previous observations $y_{t-1}, y_{t-2},\ldots,y_2\ y_1$ for each t (1<=t<=T). Then minimize

$$\text{MSE} = \boxed{\frac{1}{T}} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2$$

Sometimes omit it leading to SSE (sum of squared errors)
Result would not be affected

Estimation

Besides minimization, another common method is to select the optimal parameters by a grid search.

For example, in the SES model, α is only parameter involved (now we ignore the initial values of level). The range of α is from 0 to 1. But in practice, the range sometimes can be restricted to (0,0.3) according to expertise experience. There is no theoretical support for such restriction.

The search is still based on MSE or SSE.

In SES, we can assume $0 \leq \alpha \leq 1$ which is in fact the max range. And we increase α from 0 to 1 by 0.05, and calculate the MSE for each α and then select the best one. Totally 201 iterations are needed.

When the number of parameters increased, the combinations make the computation times increase exponentially.

Initialization

The initial values of the level, seasonality and/or trend can be estimated with the parameters. But such method is not suggested.

The best idea is to estimate them independently.

About l_0, we can initialize it as y_1, or the average value of $y_1, y_2, \ldots, y_t$ where t is less than T.

About b_0, we can initialize it as $(y_t - y_1)/(t-1)$, where t is less than T.

About the initial values of seasonality, we have to initialize $s_0, s_1, \ldots, s_{m-1}$, where m is the length of the season. For s_i, $i \leq m-1$, the initial value could be

$$s_i = \frac{1}{K}\sum_{k=0}^{K} y_{i+km}$$

The methods in this class are all commonly used. There are alternatives which suit specific problems.

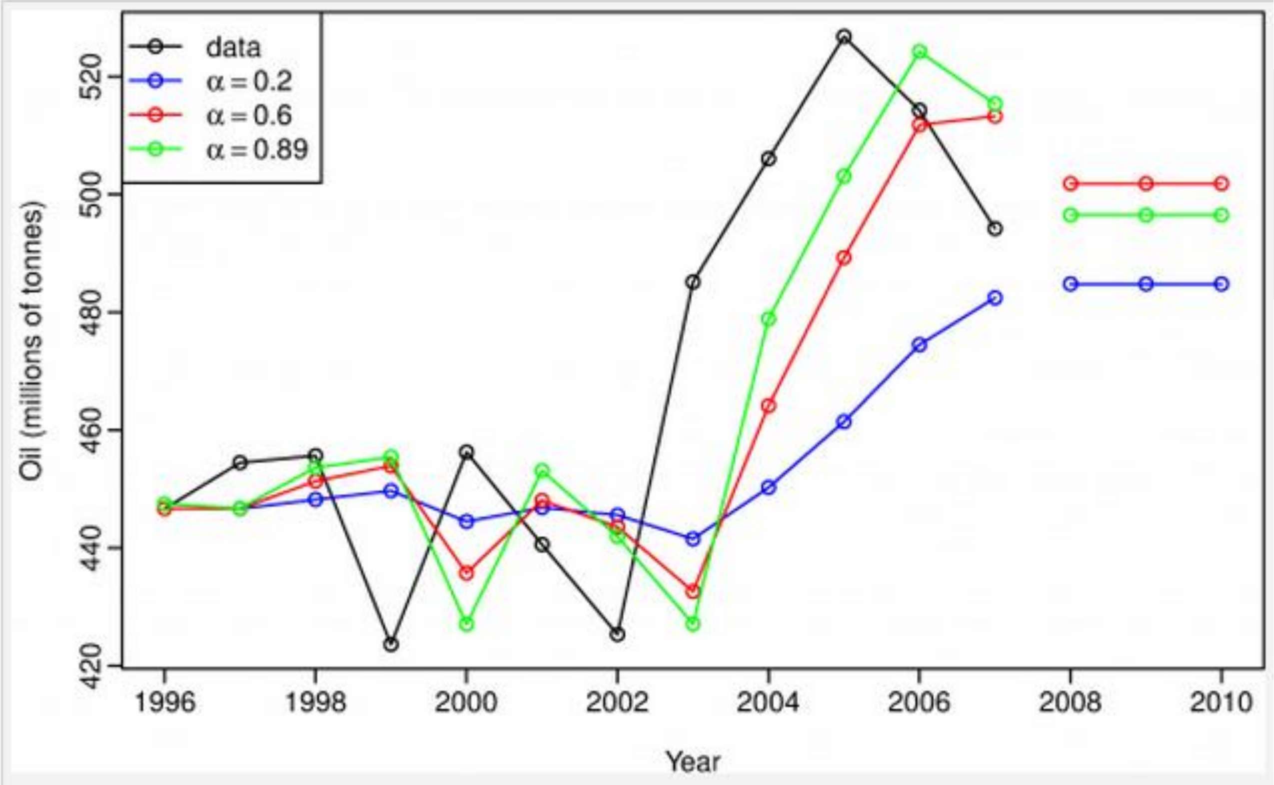
Simple Exponential Smoothing

In this example, simple exponential smoothing is applied to forecast oil production in Saudi Arabia.

Year	Time Period t	Observed values y_t	Level ℓ_t $\alpha = 0.2$	Level ℓ_t $\alpha = 0.6$	Level ℓ_t $\alpha = 0.89^*$
--	0	--	446.7	446.7	447.5*
1996	1	446.7	446.7	446.7	446.7
1997	2	454.5	448.2	451.3	453.6
1998	3	455.7	449.7	453.9	455.4
1999	4	423.6	444.5	435.8	427.1
2000	5	456.3	446.8	448.1	453.1
2001	6	440.6	445.6	443.6	441.9
2002	7	425.3	441.5	432.6	427.1
2003	8	485.1	450.3	464.1	478.9
2004	9	506.0	461.4	489.3	503.1
2005	10	526.8	474.5	511.8	524.2
2006	11	514.3	482.5	513.3	515.3
2007	12	494.2	484.8	501.8	496.5

Simple Exponential Smoothing

The longer range forecasts, SES assumes that the forecast function is "flat". That is $\hat{y}_{t+h} = \hat{y}_{t+1}$.



Forecasts $\hat{y}_{T+h\|T}$		
484.8	501.8	496.5
484.8	501.8	496.5
484.8	501.8	496.5

Holt's Linear Method

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

$$\begin{aligned} \text{Level:} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}), \\ \text{Growth:} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}, \\ \text{Forecast:} \quad & \hat{y}_{t+h|t} = \ell_t + b_t h. \end{aligned}$$

Here ℓ_t is an estimate of the level of the series at time t and b_t an estimate of the slope (or growth) of the series at time *t*. Note that b_t is a weighted average of the previous growth b_{t-1} and an estimate of growth based on the difference between successive levels

Holt's Linear Method

The error correction form of the level and the trend equations show the adjustments in terms of the within-sample one-step forecast errors:

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$$
$$b_t = b_{t-1} + \alpha\beta^* e_t$$

where $e_t = y_t - (\ell_{t-1} + b_{t-1}) = y_t - \hat{y}_{t|t-1}$

We can further define that $\beta = \alpha\beta^*$.

To estimate the parameters, we shall initialize the level l_0 and slop b_0 first.

Holt's linear model is sometimes called *addictive trend model*.

Addictive Damped Trend Method

Gardner and McKenzie (1985) proposed a modification of Holt's linear method to allow the "damping" of trends. The equations for this method are

$$\begin{aligned}
\text{Level:} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}), \\
\text{Growth:} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}, \\
\text{Forecast:} \quad & \hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t.
\end{aligned}$$

Thus, the growth for the one-step forecast of y_{t+1} is ϕb_t , and the growth is dampened by a factor of ϕ for each additional future time period.

Multiplicative (Exponential) Trend Method

A variation from Holt's linear trend method is achieved by allowing the level and the slope to be multiplied rather than added:

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t b_t^h \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}) \\
b_t &= \beta^* \frac{\ell_t}{\ell_{t-1}} + (1-\beta^*) b_{t-1}
\end{aligned}$$

Where b_t now represents an estimated growth rate (in relative terms rather than absolute) which is multiplied rather than added to the estimated level.

The error correction form is

$$\begin{aligned}
\ell_t &= \ell_{t-1} b_{t-1} + \alpha e_t \\
b_t &= b_{t-1} + \alpha\beta^* \frac{e_t}{\ell_{t-1}}
\end{aligned}$$

Where $e_t = y_t - (\ell_{t-1} b_{t-1}) = y_t - \hat{y}_{t|t-1}$.

Comparison

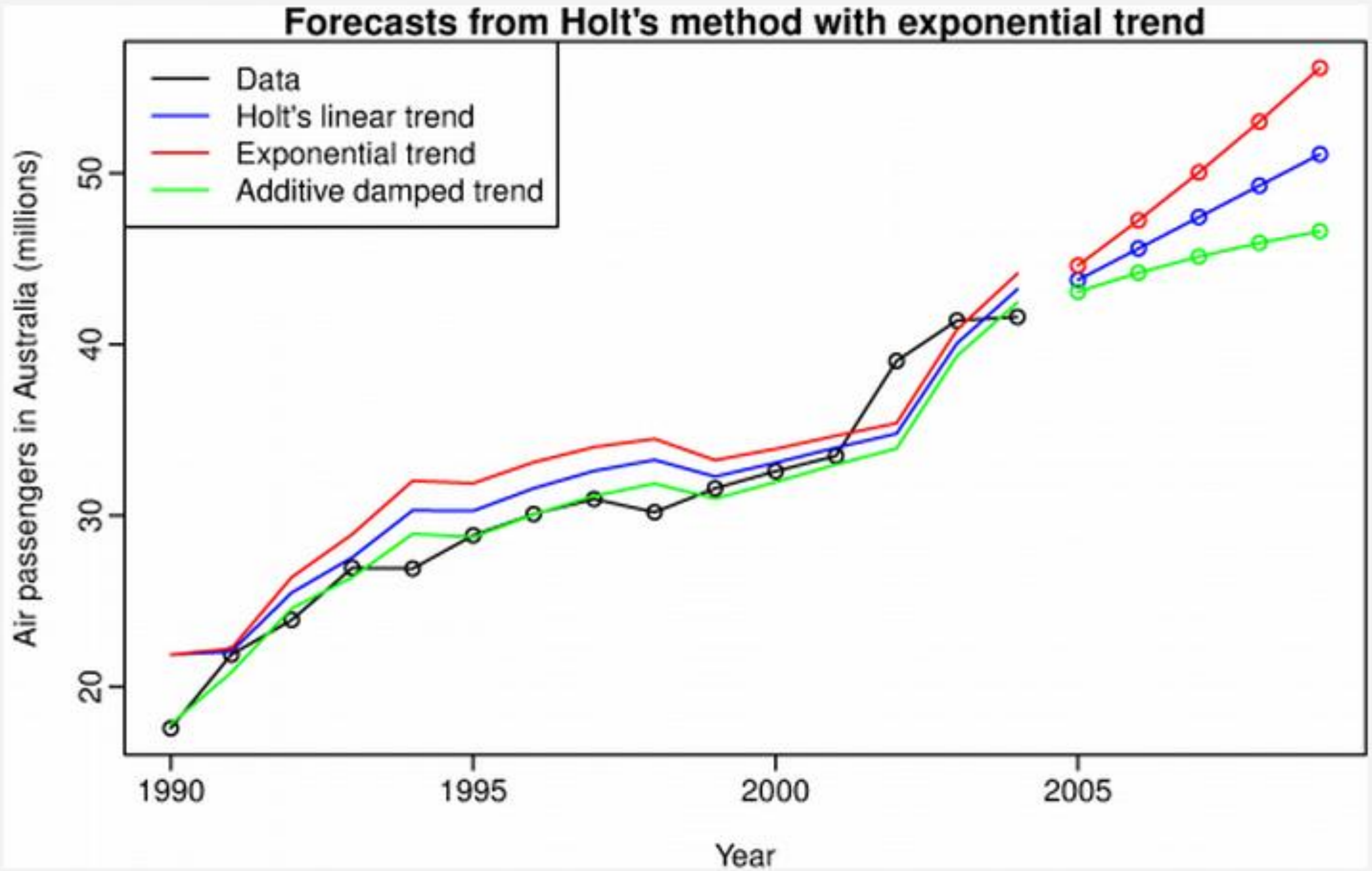


Figure 7.3: Forecasting Air Passengers in Australia (thousands of passengers). For all methods α=0.8 and β=0.2, and for the additive damped trend method with φ=0.85.*

Multiplicative Damped Trend Method

Motivated by the improvements in forecasting performance seen in the additive damped trend case, Taylor (2003) introduced a damping parameter to the multiplicative trend method resulting to a multiplicative damped trend method:

$$\begin{aligned}\hat{y}_{t+h|t} &= \ell_t b_t^{(\phi+\phi^2+\cdots+\phi^h)} \\ \ell_t &= \alpha y_t + (1-\alpha)\ell_{t-1} b_{t-1}^{\phi} \\ b_t &= \beta^* \frac{\ell_t}{\ell_{t-1}} + (1-\beta^*) b_{t-1}^{\phi}\end{aligned}$$

The error correction form of the smoothing equations is

$$\begin{aligned}\ell_t &= \ell_{t-1} b_{t-1}^{\phi} + \alpha e_t \\ b_t &= b_{t-1}^{\phi} + \alpha\beta^* \frac{e_t}{\ell_{t-1}}\end{aligned}$$

Holt-Winters seasonal method

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

Holt (1957) and Winters (1960) extended Holt's method to capture seasonality. Holt-Winters' method is based on three smoothing equations—one for the level, one for trend, and one for seasonality. It is similar to Holt's linear method, with one additional equation for dealing with seasonality. In fact, there are two different Holt-Winters' methods, depending on whether seasonality is modeled in an additive or multiplicative way.

Holt-Winters multiplicative method

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

$$
\begin{aligned}
\text{Level:} \quad & \ell_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Growth:} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
\text{Seasonal:} \quad & s_t = \gamma y_t/(\ell_{t-1} + b_{t-1}) + (1-\gamma)s_{t-m} \\
\text{Forecast:} \quad & \hat{y}_{t+h|t} = (\ell_t + b_t h)s_{t-m+h_m^+},
\end{aligned}
$$

Where m is the length of seasonality, ℓ_t represents the level of the series, b_t denotes the growth, s_t is the seasonal component, $\hat{y}_{t+h|t}$ is the forecast for h periods ahead, and $h_m^+ = \lfloor (h-1) \bmod m \rfloor + 1$.

As with other smoothing methods, we need initial values of the components and estimates of the parameter values.

Holt-Winters multiplicative method

The error correction representation is of HW multiplicative method is:

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \frac{e_t}{s_{t-m}}$$
$$b_t = b_{t-1} + \alpha\beta^* \frac{e_t}{s_{t-m}}$$
$$s_t = s_t + \gamma \frac{e_t}{(\ell_{t-1} + b_{t-1})}$$

Holt-Winters addictive method

The seasonal component in Holt-Winters' method may also be treated additively, although in practice this seems to be less commonly used. The basic equations for Holt-Winters' additive method are as follows:

$$
\begin{aligned}
\text{Level:} \quad & \ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Growth:} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
\text{Seasonal:} \quad & s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\
\text{Forecast:} \quad & \hat{y}_{t+h|t} = \ell_t + b_t h + s_{t-m+h_m^+}.
\end{aligned}
$$

About the third equation, the equation for the seasonal component is often expressed in other books as

$$s_t = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{t-m}$$

$$s_t = \gamma^*(1-\alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1-\gamma^*(1-\alpha)]s_{t-m}$$

$$\gamma = \gamma^*(1-\alpha) \quad 0 \le \gamma \le 1-\alpha$$

Holt-Winters addictive method

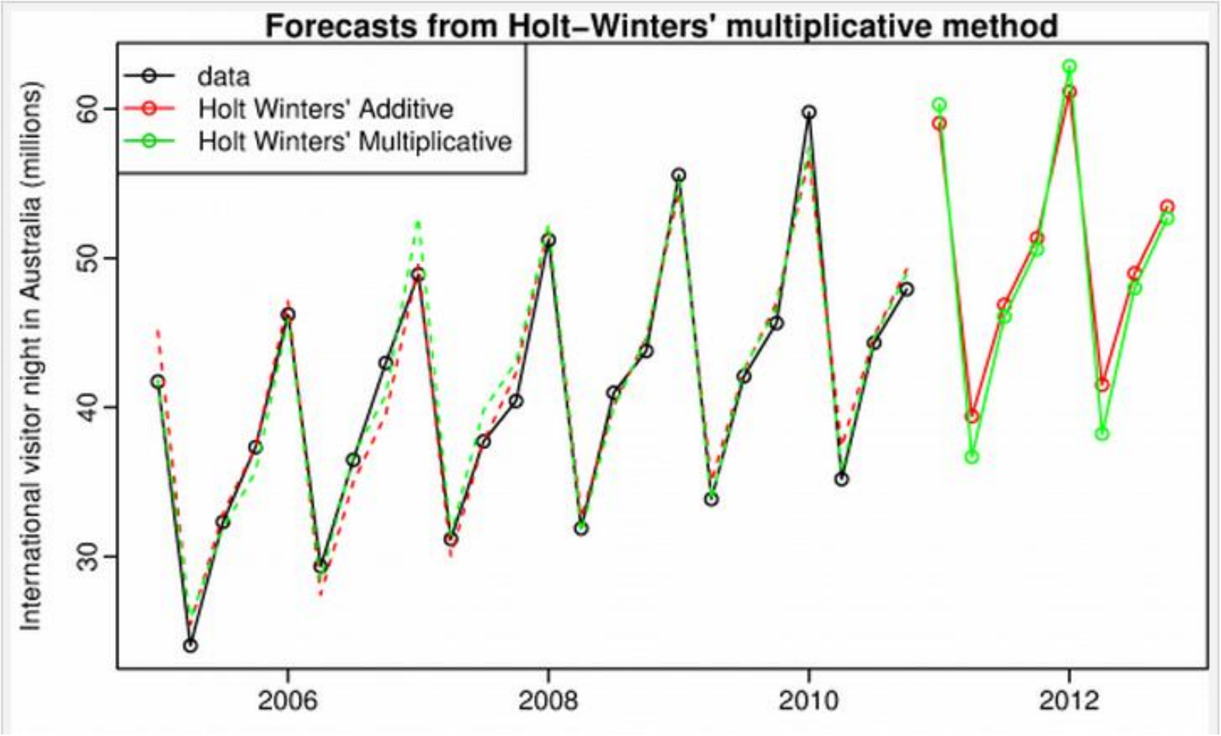
The error correction form of the smoothing equations is:

$$\begin{aligned}
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha e_t \\
b_t &= b_{t-1} + \alpha\beta^* e_t \\
s_t &= s_{t-m} + \gamma e_t.
\end{aligned}$$

where $e_t = y_t - (\ell_{t-1} + b_{t-1} + s_{t-m}) = y_t - \hat{y}_{t|t-1}$ are the one-step training forecast errors.

Comparison

In this example we employ the Holt-Winters method with both additive and multiplicative seasonality to forecast tourists visitor nights in Australia by international arrivals.



A taxonomy of exponential smoothing methods

Exponential smoothing methods are not restricted to those we have presented so far. By considering variations in the combination of the trend and seasonal components, fifteen exponential smoothing methods are possible.

Each method is labeled by a pair of letters (T,S) defining the type of 'Trend' and 'Seasonal' components. For example, (A,M) is the method with an additive trend and multiplicative seasonality; (M,N) is the method with multiplicative trend and no seasonality; and so on.

		Seasonal Component	
Trend	N	A	M
Component	(None)	(Additive)	(Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)
M (Multiplicative)	(M,N)	(M,A)	(M,M)
M_d (Multiplicative damped)	(M_d,N)	(M_d,A)	(M_d,M)

A taxonomy of exponential smoothing methods

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t / \ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
$\mathbf{A_d}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$
M	$\hat{y}_{t+h\|t} = \ell_t b_t^h$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$
$\mathbf{M_d}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t / (\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$

Exponential Smoothing and ARIMA Models

Recall that the first-order exponential smoother is given as

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

and the forecast error is defined as

$$e_t = y_t - \hat{y}_t$$

Then we have

$$\begin{aligned}
e_T - (1-\lambda)e_{T-1} &= (y_T - \hat{y}_{T-1}) - (1-\lambda)(y_{T-1} - \hat{y}_{T-2}) \\
&= y_T - y_{T-1} - \hat{y}_{T-1} + \underbrace{\lambda y_{T-1} + (1-\lambda)\hat{y}_{T-2}}_{=\hat{y}_{T-1}} \\
&= y_T - y_{T-1} - \hat{y}_{T-1} + \hat{y}_{T-1} \\
&= y_T - y_{T-1}.
\end{aligned}$$

Exponential Smoothing and ARIMA Models

We can rewrite the equation as

$$y_T - y_{T-1} = e_T - \theta e_{T-1}$$

Such model is called integrated moving average model denoted as IMA(1,1).

For more discussion of the equivalence between exponential smoothing techniques and the ARIMA models, see Abraham and Ledolter (1983), Cogger (1974), Goodman (1974), Pandit and Wu (1974), and McKenzie (1984).

ES and ARIMA are not one-to-one mapping!

State Space Model

Now we study the statistical models that underlie the exponential smoothing methods we have considered so far.

The exponential smoothing methods presented are algorithms that 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.

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

State Space Model

To distinguish between a model with additive errors and one with multiplicative errors (and to also distinguish the models from the methods), we label each state space model as ETS(·,·,··,·,·) for (Error, Trend, Seasonal).

The possibilities for each component are: Error ={A,M}, Trend ={N,A,A_d,M,M_d} and Seasonal ={N,A,M}. Therefore, in total there exist 30 such state space models: 15 with additive errors and 15 with multiplicative errors.

For example, ETS(A,N,N): simple exponential smoothing with additive errors.

State Space Model

As discussed before, the error correction form of simple exponential smoothing is given by $\ell_t = \ell_{t-1} + \alpha e_t$, $e_t = y_t - \hat{y}_{t|t-1}$ represents a one-step forecast 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 one-step forecast errors e_t are normally distributed white noise with mean 0 and variance σ^2. Here we denote e_t as $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$.

Usually we use e_t to denote the observed error while ε_t the error which has not been observed. The SES can be written as

$$y_t = \ell_{t-1} + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$$

We refer to the first equation as the *measurement* (or observation) equation and the second as the *state* (or transition) equation.

State Space Model

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.

Another example, 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 random errors as relative errors:

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

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

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

State Space Model

For the two methods with addictive error and multiplicative error underlying the simple exponential smoothing. The point forecasts produced by the models are identical if they use the same smoothing parameter values. They will, however, generate different prediction intervals.

ADDITIVE ERROR MODELS

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

State Space Model

MULTIPLICATIVE ERROR MODELS

Trend	**Seasonal**		
	N	**A**	**M**
N	$y_t = \ell_{t-1}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$	$y_t = (\ell_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+\alpha(\ell_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1}+b_{t-1})(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1}+b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+b_{t-1}+\alpha(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1}+b_{t-1})s_{t-m}(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
$\mathbf{A_d}$	$y_t = (\ell_{t-1}+\phi b_{t-1})(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1}+\phi b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+\phi b_{t-1}+\alpha(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1}+\phi b_{t-1})s_{t-m}(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}(1+\beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}+\alpha(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}+\beta(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m}+\gamma(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}(1+\beta\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
$\mathbf{M_d}$	$y_t = \ell_{t-1}b_{t-1}^{\phi}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1+\beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}+\alpha(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^{\phi}+\beta(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m}+\gamma(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1+\beta\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$

Example

MANAGEMENT SCIENCE
Vol. 54, No. 2, February 2008, pp. 253–265
ISSN 0025-1909 | EISSN 1526-5501 | 08 | 5402 | 0253
DOI 10.1287/mnsc.1070.0786

A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center

James W. Taylor
Saïd Business School, University of Oxford, Oxford OX1 1HP, United Kingdom,
james.taylor@sbs.ox.ac.uk

MANAGEMENT SCIENCE
Vol. 58, No. 3, March 2012, pp. 534–549
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)
http://dx.doi.org/10.1287/mnsc.1110.1434

Density Forecasting of Intraday Call Center Arrivals Using Models Based on Exponential Smoothing

James W. Taylor
Saïd Business School, University of Oxford, Oxford OX1 1HP, United Kingdom,
james.taylor@sbs.ox.ac.uk

Q and A!