Module 5 - Exponential Smoothing Methods

MATH1307 Forecasting

Prepared by: Dr. Haydar Demirhan based on the textbook by Hyndman et al., Forecasting with Exponential Smoothing: The State Space Approach. Springer, 2008.

Introduction

Exponential smoothing methods are used to model time series data by fitting different trend and seasonality patterns. They provide a class of alternative models for time series data. State-space model approach makes the inclusion of additive and multiplicative error terms into the exponential smoothing models possible.

In this module, we will cover

- Assessing forecast accuracy
- Exponential smoothing methods
  - Simple exponential smoothing
  - Holt's linear method
  - Holt-Winters' trend and seasonality method
- State-space models
  - Additive errors
  - Multiplicative errors

Assessing Forecast Accuracy

We can use different error measures to assess the forecast accuracy. Define the one-step-ahead forecast error as $e_t = Y_t - \hat{Y}_t$, then we have the following *scale-dpenedent* errors:

- Mean absolute error (MAE) = $mean(|e_t|)$.

- Mean squared error (MSE) = $mean(e_t^2)$.
- Root mean squared error (RMSE) = $\sqrt{mean(e_t^2)}$.

When comparing forecast methods on a single series, we prefer the MAE, which cannot be used to make comparisons between series as it makes no sense to compare accuracy on different scales.

The percentage error is defined by $p_t = 100e_t / Y_t$, which is scale-independent. We have the following percentage errors:

- Mean absolute percentage error (MAPE) = $mean(|p_t|)$.
- Symmetric mean absolute percentage error (sMAPE) = $mean(200|Y_t - \hat{Y}_t| / (Y_t + \hat{Y}_t))$.

Notice that because the value of sMAPE can be negative, it is not a measure of *absolute percentage errors* at all.

Mean absolute scaled error (MASE) is another generally applicable measure of forecast accuracy. It is obtained by scaling the errors based on the in-sample MAE from the naiıve forecast method. Thus, a scaled error is defined as

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n} |Y_i - Y_{i-1}|},$$

which is independent of the scale of the data. A scaled error is less than one if it arises from a better forecast than the average one-step naiıve forecast computed in-sample. Conversely, it is greater than one if the forecast is worse than the average one-step na ̈ıve forecast computed in-sample. Then,

- Mean absolute scaled error (MASE) = $mean(|q_t|)$.

The MASE can be used to compare forecast methods on a single series, and to compare forecast accuracy between series as it is scale-free. It is the only available method which can be used in all circumstances.

Classification of Exponential Smoothing Methods

In exponential smoothing, we represent the trend component as a combination of a level term (ℓ) and a growth term (b).The level and growth can be combined in a number of ways, giving five future trend types. We will define the following trend types or growth patterns:

$$\begin{aligned}
\text{None:} \quad & T_h = \ell \\
\text{Additive:} \quad & T_h = \ell + bh \\
\text{Additive damped:} \quad & T_h = \ell + (\phi + \phi^2 + \cdots + \phi^h)b \\
\text{Multiplicative:} \quad & T_h = \ell b^h \\
\text{Multiplicative damped:} \quad & T_h = \ell b^{(\phi + \phi^2 + \cdots + \phi^h)}
\end{aligned}$$

In the damped trend types, we dampen the trend as the length of the forecast horizon increases to improve the forecast accuracy particularly at long lead times.

We can also include a seasonal component to the model either additively or multiplicatively. And we include the error in the same way: either additively or multiplicatively. In general, the nature of the error component has often been ignored, because the distinction between additive and multiplicative errors makes no difference to point forecasts.

When the type of error component is ignored, we count fifteen exponential smoothing methods given in the following table. Also, relevant cells include the common cases of the methods.

Table 1. Classification of exponential smoothing methods

Trend Component	Seasonal Component N (None)	 A (Additive)	 M (Multiplicative)
N (None)	N,N Simple exponential smoothing	N,A	N,M
A (Additive)	A,N Holt's linear method	A,A Holt-Winters' additive method	A,M Holt-Winters' multiplicative method
Ad (Additive damped)	Ad,N Additive Damped trend method	Ad,A	Ad,M Holt-Winters' damped method
M (Multiplicative)	M,N Exponential trend method	M,A	M,M
Md (Multiplicative damped)	Md,N Multiplicative damped trend method	Md,A	Md,M

For each method in the above Table 1, there are two possible state-space models. One is corresponding to a model with additive errors and the other is corresponding to a model with multiplicative errors. If the same parameter values are used, these two models give equivalent point forecasts although different prediction intervals. Thus, there are 30 potential models described in this classification. An exponential smoothing method is an algorithm for producing point forecasts only. In addition to the point forecasts, stochastic state-space models give a framework for computing prediction intervals and other properties.

Exponential Smoothing Methods

In this section, we will cover simple exponential smoothing (N,N), Holt's linear method (A,N), the damped trend method (Ad,N) and Holt-Winters' seasonal method (A,A and A,M).

Suppose that we have given the series $Y_t = Y_1, Y_2, \dots, Y_n$, then a forecast of Y_{t+h} given the observation up to time point t is denoted by $\hat{Y}_{t+h|t}$. For one step forecasts, we will use the notation Y_{t+1}.

Simple Exponential Smoothing

The method of simple exponential smoothing basically takes the forecast for the previous period (say for time t, $\hat{Y}_t$) and adjusts it using the forecast error $(Y_t - \hat{Y}_t)$ using the following expression:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t),$$

where α is a constant between 0 and 1. The new forecast is just a function of the old forecast and its error with weight changing according to the value of α. After simple manipulation, we can write Eq. (2) as a **weighted average** of observation and forecast at time point t:

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t.$$

If we replace $\hat{Y}_t$ with its components and expand the Eq. (3), we get

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

If we keep replacing, we obtain the following general form:

$$\begin{aligned}\hat{Y}_{t+1} = \quad & \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \alpha(1-\alpha)^3 Y_{t-3} \\ & + \alpha(1-\alpha)^4 Y_{t-4} + \cdots + \alpha(1-\alpha)^{t-1} Y_1 + (1-\alpha)^t \hat{Y}_1.\end{aligned}$$

In this case, $\hat{Y}_{t+1}$ represents a weighted moving average of all past observations with the weights decreasing exponentially; hence, it's called **exponential smoothing**. An important problem with these models is called *initialization problem*. When the value of α is small, the weight of $\hat{Y}_1$ would be quite large and this would cause problems with relatively short time series.

For longer range forecasts, it is assumed that the forecast function is *flat*: $\hat{Y}_{t+h|t} = \hat{Y}_{t+1}$ for $h = 2, 3, 4, \ldots$ because simple exponential smoothing works best for data that have no trend, seasonality, or other underlying patterns.

Notice that we need to specify $\hat{Y}_1$ and α for the simple exponential smoothing.

We will use the function `ses()` (https://www.rdocumentation.org/packages/forecast/versions/7.3/topics/ses) from the `forecast` package to fit exponential smoothing models. The argument `alpha` is used to specify the value of α and if it is not specified α is estimated. To apply Box-Cox transformation before the model is fitted, we use argument `lambda` which corresponds to the Box-Cox transformation parameter λ. It is also possible to choose the approach to estimate initial values. If the argument `initial` is set to `optimal`, the initial values are optimized along with the smoothing parameters. If it is set to `simple`, the initial values are obtained using simple calculations on the first few observations.

To illustrate the contents, let's apply the simple exponential smoothing to the annual US new freight cars series observed between 1947 and 1993.

```
data("freight")
fit1 <- ses(freight, alpha=0.1, initial="simple", h=5) # Set alpha to a small value
summary(fit1)
```

```
##
## Forecast method: Simple exponential smoothing
##
## Model Information:
## Simple exponential smoothing
##
## Call:
##  ses(y = freight, h = 5, initial = "simple", alpha = 0.1)
##
##   Smoothing parameters:
##     alpha = 0.1
##
##   Initial states:
##     l = 4631.45
##
##   sigma:  1634.026
## Error measures:
##                     ME     RMSE      MAE       MPE     MAPE     MASE
## Training set -615.3614 1634.026 1339.442 -109.3975 125.4587 1.279911
##                   ACF1
## Training set 0.4752577
##
## Forecasts:
##      Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## 1994       1739.251 -354.8369 3833.339 -1463.380 4941.883
## 1995       1739.251 -365.2813 3843.784 -1479.354 4957.856
## 1996       1739.251 -375.6741 3854.177 -1495.248 4973.751
## 1997       1739.251 -386.0161 3864.519 -1511.065 4989.567
## 1998       1739.251 -396.3080 3874.811 -1526.805 5005.307
```

```r
fit2 <- ses(freight, alpha=0.8, initial="simple", h=5) # Set alpha to a large value
summary(fit2)
```

```
##
## Forecast method: Simple exponential smoothing
##
## Model Information:
## Simple exponential smoothing
##
## Call:
##  ses(y = freight, h = 5, initial = "simple", alpha = 0.8)
##
##   Smoothing parameters:
##     alpha = 0.8
##
##   Initial states:
##     l = 4631.45
##
##   sigma:  1445.217
## Error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -62.30317 1445.217 1034.509 -27.74406 59.45001 0.9885304
##                   ACF1
## Training set 0.02924579
##
## Forecasts:
##      Point Forecast       Lo 80    Hi 80      Lo 95    Hi 95
## 1994       2288.851   436.73119 4140.971  -543.7218 5121.424
## 1995       2288.851   -83.01964 4660.722 -1338.6121 5916.314
## 1996       2288.851  -507.78857 5085.491 -1988.2401 6565.942
## 1997       2288.851  -876.05270 5453.755 -2551.4515 7129.153
## 1998       2288.851 -1205.72164 5783.424 -3055.6366 7633.339
```

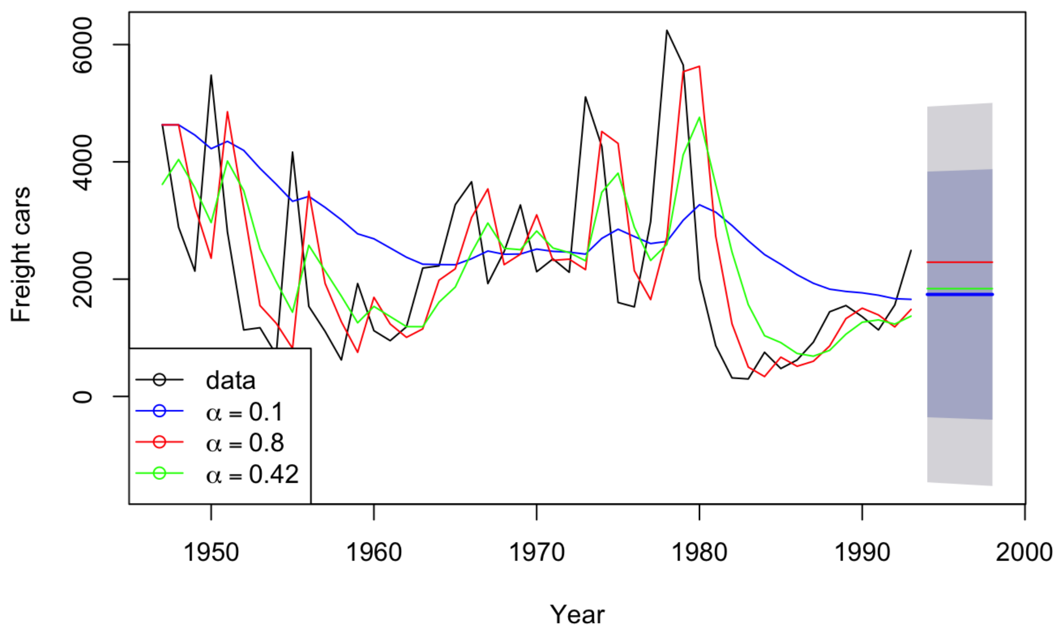
```r
fit3 <- ses(freight, h=5) # Let the software estimate alpha
summary(fit3)
```

```
##
## Forecast method: Simple exponential smoothing
##
## Model Information:
## Simple exponential smoothing
##
## Call:
##  ses(y = freight, h = 5)
##
##   Smoothing parameters:
##     alpha = 0.4179
##
##   Initial states:
##     l = 3616.2556
##
##   sigma:  1451.738
##
##      AIC     AICc      BIC
## 869.2817 869.8398 874.8321
##
## Error measures:
##                     ME     RMSE      MAE       MPE     MAPE     MASE
## Training set -90.54977 1420.514 1123.457 -51.68427 81.36899 1.073525
##                   ACF1
## Training set 0.2185174
##
## Forecasts:
##      Point Forecast      Lo 80    Hi 80     Lo 95    Hi 95
## 1994       1837.766  -22.71028 3698.243 -1007.587 4683.120
## 1995       1837.766 -178.62927 3854.162 -1246.045 4921.578
## 1996       1837.766 -323.32813 3998.861 -1467.342 5142.875
## 1997       1837.766 -458.92853 4134.461 -1674.725 5350.258
## 1998       1837.766 -586.95743 4262.490 -1870.529 5546.062
```

The output of `ses()` function mainly shows the measures of model fit and the forecasts. For some of the error measures, the model with $\alpha = 0.8$ gives better results than the one with estimated α.

The following code chunk displays the fitted models, observed series and forecasts for 5-time points ahead.

```
#main = "Exponential smoothing fits and forecasts for annual US new freight cars ser
  ies."
plot(fit1, plot.conf=FALSE, ylab=" Freight cars", xlab="Year", main="", type="l")
lines(fitted(fit1), col="blue", type="l")
lines(fitted(fit2), col="red", type="l")
lines(fitted(fit3), col="green", type="l")
lines(fit1$mean, col="blue", type="l")
lines(fit2$mean, col="red", type="l")
lines(fit3$mean, col="green", type="l")
legend("bottomleft",lty=1, col=c(1,"blue","red","green"),  c("data", expression(alph
  a == 0.1), expression(alpha == 0.8),
                                                             expression(alpha == 0.42
  )),pch=1)
```



Holt's Linear Method

Holt's linear method extends simple exponential smoothing to linear exponential smoothing to be able **include trends to the forecasting model**. In this model, we have two smoothing constants α and β^*. Both constants are in between 0 and 1. The forecasting model for Holt's linear method is composed of two components.

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

Here ℓ_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the slope (or growth) of the series at time t. The constant 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.

When $\alpha = \beta^*$, Holt's method is equivalent to *Brown's double exponential smoothing*. Another special case occurs when $\beta^* = 0$, which makes $b_t = b_{t-1} = b$. For this case, we have the following forecasting model:

$$\begin{aligned} \text{Level:} \quad & \ell_t &= \alpha Y_t + (1-\alpha)(\ell_{t-1} + b), \\ \text{Forecast:} \quad & \hat{Y}_{t+h|t} &= \ell_t + bh. \end{aligned}$$

This method is known as **simple exponential smoothing with drift**.

Damped Trend Method

This is a modification of Holt's linear method. The forecasting equations include another parameter ϕ to create a damped trend. The forecast equations of this model are as follows:

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

The dampening factor ϕ is in between 0 and 1. When $\phi = 1$, this method gives the same forecasts as Holt's linear method. For $0 < \phi < 1$, as $h \to \infty$ the forecasts approach an asymptotic value given by $\ell_t + \phi b_t / (1-\phi)$.

Exponential Trend Method

A variation from Holt's linear trend method is the one with the level and the slope to be multiplied rather than added Hyndman and Athanasopoulos, 2014 (https://www.otexts.org/fpp/7/3):

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

where b_t now represents an estimated growth rate which is multiplied rather than added to the estimated level. The trend in the forecast function is now exponential rather than linear so that the forecasts project a constant growth rate rather than a constant slope.

To fit Holt's linear, exponential, and dampened trend models, we will use the function `holt()` (https://www.rdocumentation.org/packages/forecast/versions/7.3/topics/ses) from the package `forecast` . We can fit an exponential trend instead of a linear trend by setting the argument `exponential` to `TRUE` . Also, we can fit a damped trend by setting `damped = TRUE` . If the arguments `alpha` and `beta` are not specified, they are estimated from the data. We can apply the Box-Cox transformation before fitting the model by setting the argument `lambda` to the value of λ of the Box-Cox transformation. The `initial` argument specifies the approach to estimates the initial values as before.

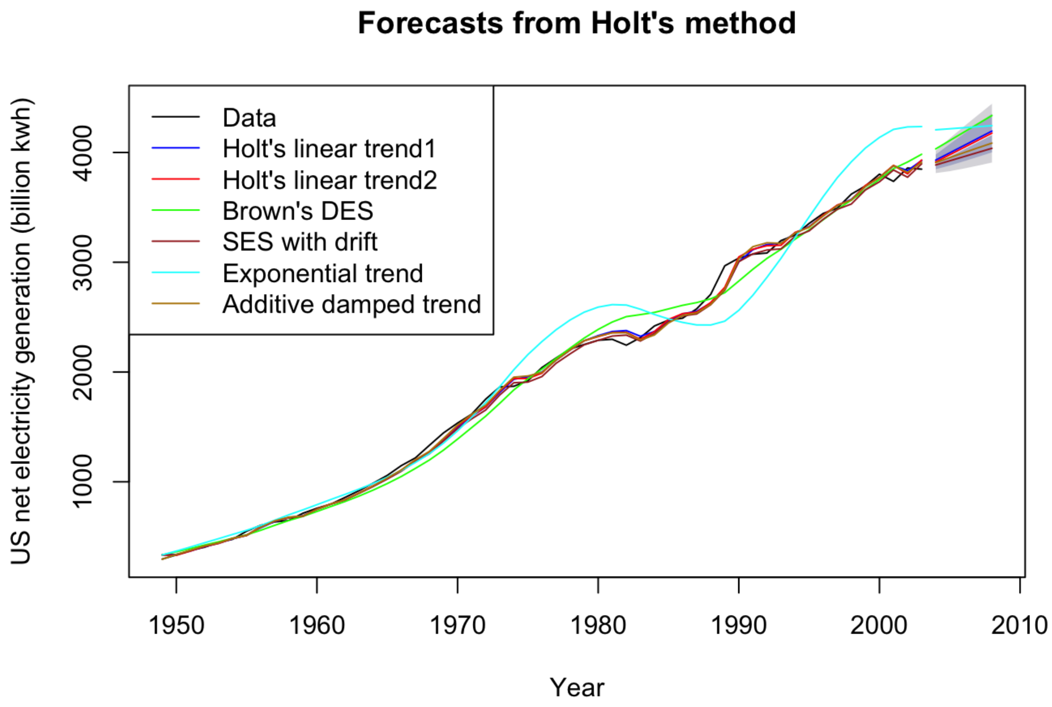
To illustrate the implementation of the methods, let's consider the annual US net electricity generation series.

```
fit1 <- holt(usnetelec, alpha=0.8, beta=0.1, initial="simple", h=5) # Set alpha and beta
fit2 <- holt(usnetelec, initial="simple", h=5) # Let the software estimate both alpha and beta
fit3 <- holt(usnetelec, alpha=0.2, beta=0.2, initial="simple", h=5) # Brown's double exponential smoothing
fit4 <- holt(usnetelec,  beta=0, initial="simple", h=5) # Simple exponential smoothing with drift
fit5 <- holt(usnetelec, alpha=0.1, beta=0.8, initial="simple", exponential=TRUE, h=5) # Fit with exponential trend
fit6 <- holt(usnetelec, alpha=0.8, beta=0.2, damped=TRUE, initial="simple", h=5) # Fit with additive damped trend
summary(fit2) # Which has the smallest RMSE
```

```
##
## Forecast method: Holt's method
##
## Model Information:
## Holt's method
##
## Call:
##  holt(y = usnetelec, h = 5, initial = "simple")
##
##   Smoothing parameters:
##     alpha = 1
##     beta  = 0.0875
##
##   Initial states:
##     l = 296.1
##     b = 38
##
##   sigma:  51.6245
## Error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 5.733357 51.62446 37.64055 0.4248986 2.265907 0.5333341
##                    ACF1
## Training set 0.08139712
##
## Forecasts:
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2004       3913.583 3847.423 3979.742 3812.401 4014.765
## 2005       3979.165 3881.424 4076.907 3829.683 4128.648
## 2006       4044.748 3919.865 4169.631 3853.755 4235.741
## 2007       4110.331 3960.093 4260.569 3880.561 4340.100
## 2008       4175.913 4001.135 4350.692 3908.613 4443.214
```

We obtained the smallest RMSE with Holt's linear trend model when the constants are estimated from the data. The following code chunk displays the fitted models, observed series and forecasts for 5-time points ahead.

```
plot(fit2, type="l", ylab="US net electricity generation (billion kwh)", xlab="Year"
  ,
     fcol="white", plot.conf=FALSE)
lines(fitted(fit1), col="blue")
lines(fitted(fit2), col="red")
lines(fitted(fit3), col="green")
lines(fitted(fit4), col="brown")
lines(fitted(fit5), col="cyan")
lines(fitted(fit6), col="darkgoldenrod")
lines(fit1$mean, col="blue", type="l")
lines(fit2$mean, col="red", type="l")
lines(fit3$mean, col="green", type="l")
lines(fit4$mean, col="brown", type="l")
lines(fit5$mean, col="cyan", type="l")
lines(fit6$mean, col="darkgoldenrod", type="l")
legend("topleft", lty=1, col=c("black","blue","red","green","brown","cyan","darkgold
  enrod"),
       c("Data","Holt's linear trend1","Holt's linear trend2", "Brown's DES", "SES w
  ith drift" , "Exponential trend","Additive damped trend"))
```



Holt-Winters' Trend and Seasonality Method

If there is a seasonality exists in the series, Holt's-Winter method provides a solution to the model fitting problem. Holt-Winters' method has the smoothing equations for the level, for trend, and seasonality. There are two different Holt-Winters' methods, depending on whether seasonality is modelled in an additive or multiplicative way.

Multiplicative seasonality case

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 like the number of months or quarters in a year, 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 $h + m = [(h-1) \mod m] + 1$. The parameters $\alpha, \beta^*,$ and γ are all restricted in between 0 and 1.

Additive seasonality case

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:

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

The only differences in the other equations are that the seasonal indices are now added and subtracted instead of taking products and ratios.

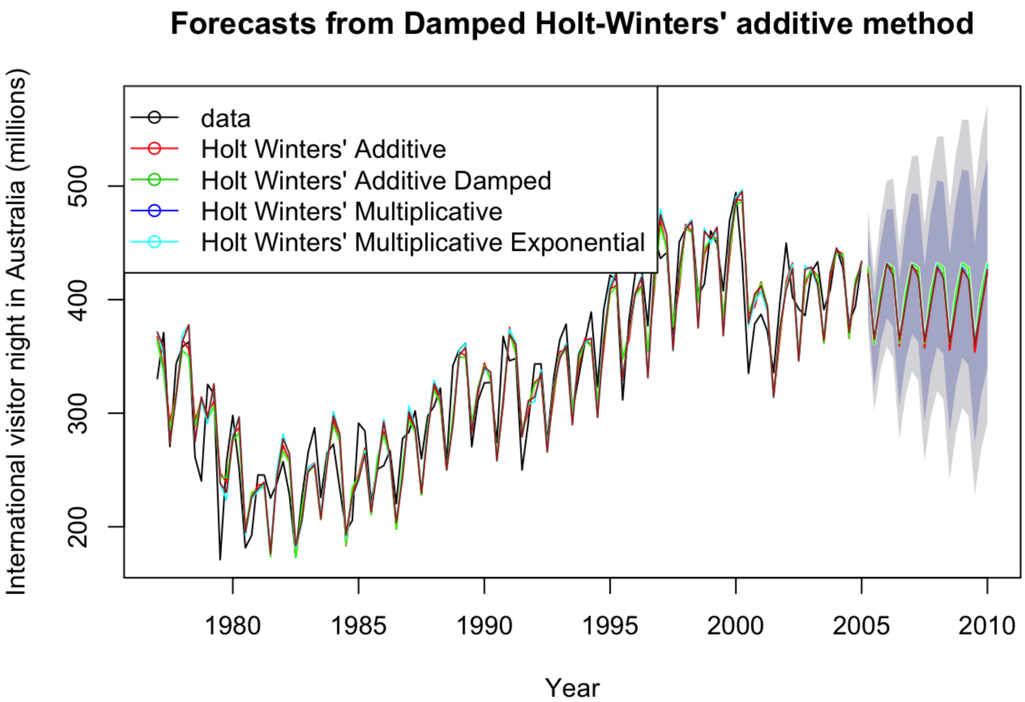
To implement Holt-Winters method, we will use `hw()` (https://www.rdocumentation.org/packages/forecast/versions/7.3/topics/ses) function from the package `forecast`. In addition to the parameter explained for `ses()` and `holt()` functions, we use the argument `seasonal=c("additive","multiplicative")` to set the type of model.

Let's revisit the quarterly motor vehicle production in the UK (thousands of cars) for the first quarter of 1977 through the first quarter of 2005 and fit various Holt-Winter models using the following code chunk.

```
data("ukcars")
fit1 <- hw(ukcars,seasonal="additive", h=5*frequency(ukcars))
fit2 <- hw(ukcars,seasonal="additive",damped = TRUE, h=5*frequency(ukcars))
fit3 <- hw(ukcars,seasonal="multiplicative", h=5*frequency(ukcars))
fit4 <- hw(ukcars,seasonal="multiplicative",exponential = TRUE, h=5*frequency(ukcar
  s))
```

An additive model with damped trend gives the smallest RMSE value within the set of considered models. The following code chunk displays the fitted models, observed series and forecasts for 5-time points ahead.

```
plot(fit2,ylab="International visitor night in Australia (millions)",
     plot.conf=FALSE, type="l", fcol="white", xlab="Year")
lines(fitted(fit1), col="red", lty=1)
lines(fitted(fit2), col="green", lty=1)
lines(fitted(fit3), col="cyan", lty=1)
lines(fitted(fit4), col="brown", lty=1)
lines(fit1$mean, type="l", col="red")
lines(fit2$mean, type="l", col="green")
lines(fit3$mean, type="l", col="cyan")
lines(fit4$mean, type="l", col="brown")
legend("topleft",lty=1, pch=1, col=1:5,
       c("data","Holt Winters' Additive", "Holt Winters' Additive Damped", "Holt Win
  ters' Multiplicative", "Holt Winters' Multiplicative Exponential"))
```



For all the methods discussed above, we can get some special cases for extreme values of parameters. For example,

- if $\alpha = 0$, the level is constant over time,
- if $\beta = 0$, the slope is constant over time, and

- if $\gamma = 0$, the seasonal pattern is constant over time,
- if $\alpha = 1$, naive forecasts are obtained using the simple exponential smoothing method, and
- the additive and multiplicative trend methods are special cases of their damped counterparts obtained by letting $\phi = 1$.

State Space Models

State-space models allow considerable flexibility in the specification of the parametric structure. The textbook uses the *innovations* formulation of the model. Let Y_t denote the observation at time t, and let $\boldsymbol{X}_t$ be a **state vector** containing unobserved components that describe the level, trend and seasonality of the series. Then a linear innovations state-space model is written as

$$\begin{aligned} Y_t &= \boldsymbol{\omega}'\boldsymbol{X}_t + \epsilon_t, \\ \boldsymbol{X}_t &= \boldsymbol{F}\boldsymbol{X}_{t-1} + \boldsymbol{g}\epsilon_t, \end{aligned}$$

where $\{\epsilon_t\}$ is a white noise series and $\boldsymbol{F}, \boldsymbol{g},$ and $\boldsymbol{\omega}$ are coefficients. The first line of Eq. (12) is known as the *measurement* (or observation) equation; it describes the relationship between the unobserved states $\boldsymbol{X}_{t-1}$ and the observation Y_t. The second line is known as the *transition* (or state) equation; it describes the evolution of the states over time. The use of identical errors (or innovations) in these two equations makes it an **innovations** state-space model.

Nonlinear state-space models are also possible. One form that we will use is

$$\begin{aligned} Y_t &= w(\boldsymbol{X}_{t-1}) + r(\boldsymbol{X}_{t-1})\epsilon_t, \\ \boldsymbol{X}_t &= f(\boldsymbol{X}_{t-1}) + g(\boldsymbol{X}_{t-1})\epsilon_t. \end{aligned}$$

where $\{\epsilon_t\}$ is a white noise process with $N(o, \sigma^2)$.

For each exponential smoothing method, there are two corresponding state-space models according to their error terms (one additive and one multiplicative). When the same parameters values are used the point forecasts for both types of model are identical. We will use the triplet to denote the state-space model correspondents of exponential smoothing models. Each triplet will include an additional letter either A or M to denote the type of error terms and the remaining two words will have the same interpretation as those in Table 1.

The triplet (E,T,S) refers to the three components: error, trend and seasonality. For example, ETS(A,A,N) means that the corresponding model has additive errors, additive trend and no seasonality. From Table 1, this is Holt's linear method with additive errors. ETS(M,Md,M) refers to a model with multiplicative errors, a damped multiplicative trend and multiplicative seasonality. ETS can also be considered an abbreviation of **ExponenTial Smoothing**.

We denote the conditional mean of of a future observation given knowledge of the past by $E(Y_{t+h}|X_t)$, where X_t contains the unobserved components such as $|ell_t, b_t,$ and s_t. For $h = 1$, we will use $\mu_{t+1} = \mu_{t+1|t}$. For most of the models, $\mu_{t+h|t} = \hat{Y}_{t+h|t}$. However, for the models with multiplicative trend or multiplicative seasonality, the conditional mean and the point forecast will differ slightly for $h \geq 2$.

Now, we will proceed onto the specific state-space models corresponding to the exponential smoothing models.

State-space models for Holt's linear method

We will illustrate the additive and multiplicative error models.

Additive error model: ETS(A,A,N)

Let $\hat{\mu}_t = \hat{Y}_t = \ell_{t-1} + b_{t-1}$ denote he one-step forecast of Y_t assuming we know the values of all parameters. And let $\epsilon_t = Y_t - \mu_t$ denote one-step ahead forecast error at time t. From the forecast equation in Eq. (6), we find

$$Y_t = \ell_{t-1} + b_{t-1} + \epsilon_t.$$

Using the level and growth equations of Eq. (6), we get

$$
\begin{aligned}
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\epsilon_t,\\
b_t &= b_{t-1} + \beta^*(\ell_t - \ell_{t-1} - b_{t-1}) + \alpha\beta^*\epsilon_t.
\end{aligned}
$$

If we simplify the last equation in (15) by setting $\beta = \alpha\beta^*$, equations in (14) and (15) constitute a state space model underlying Holt's method. We can write it in standard state space formulation state vector as $\boldsymbol{X}_t = (\ell_t, b_t)'$ such as

$$
\begin{aligned}
Y_t &= \begin{bmatrix} 1 & 1 \end{bmatrix}\boldsymbol{X}_{t-1} + \epsilon_t\\
\boldsymbol{X}_t &= \begin{bmatrix} 1 & 1\\ 0 & 1 \end{bmatrix}\boldsymbol{X}_{t-1} + \begin{bmatrix} \alpha\\ \beta \end{bmatrix}\epsilon_t.
\end{aligned}
$$

The model is fully specified once we state the distribution of the error term ϵ_t. Usually we assume that $\epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

Multiplicative error model: ETS(M,A,N)

We will a similar way to derive the model with multiplicative error terms. First we set $\epsilon_t = (Y_t - \mu_t)/\mu_t$ to define ϵ_t as a relative error. Then, we find

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

or in a state-space model form

$$
\begin{aligned}
Y_t &= \begin{bmatrix} 1 & 1 \end{bmatrix}\boldsymbol{X}_{t-1}(1 + \epsilon_t),\\
\boldsymbol{X}_t &= \begin{bmatrix} 1 & 1\\ 0 & 1 \end{bmatrix}\boldsymbol{X}_{t-1} + \begin{bmatrix} 1 & 1 \end{bmatrix}\boldsymbol{X}_{t-1}\begin{bmatrix} \alpha\\ \beta \end{bmatrix}\epsilon_t.
\end{aligned}
$$

Again, we have the assumption $\epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

This kind of nonlinear state-space models are usually considered difficult to handle in estimating and forecasting. However, we can still compute forecasts, the likelihood and prediction intervals for this nonlinear model with no more effort than is required for the additive error model.

We use the function `ets()` (https://www.rdocumentation.org/packages/forecast/versions/7.3/topics/ets) from the `forecast` package to fit state-space models corresponding to the exponential smoothing methods. We need to specify the considered model using the argument `model`, which get a three-character string. The characters and their meanings are `"N"=none`, `"A"=additive`, `"M"=multiplicative`, and `"Z"=automatically`. The first letter denotes the error type ("A", "M" or "Z"); the second letter denotes the trend type ("N","A","M" or "Z"); and the third letter denotes the season type ("N","A","M" or "Z"). So For example, "ANN" is the simple exponential smoothing with additive errors, "MAM" is multiplicative Holt-Winters' method with multiplicative errors, and so on. The arguments `alpha`, `beta`, and `gamma` are for the corresponding constants and it is possible to apply a Box-Cox transformation using the argument `lambda` before fitting the model. There are also some other useful arguments with this function.

Let's illustrate the application of the models over the annual US net electricity generation series, which we have fitted Holt's linear method.

```
fit.AAN = ets(usnetelec, model="AAN")
summary(fit.AAN)
```

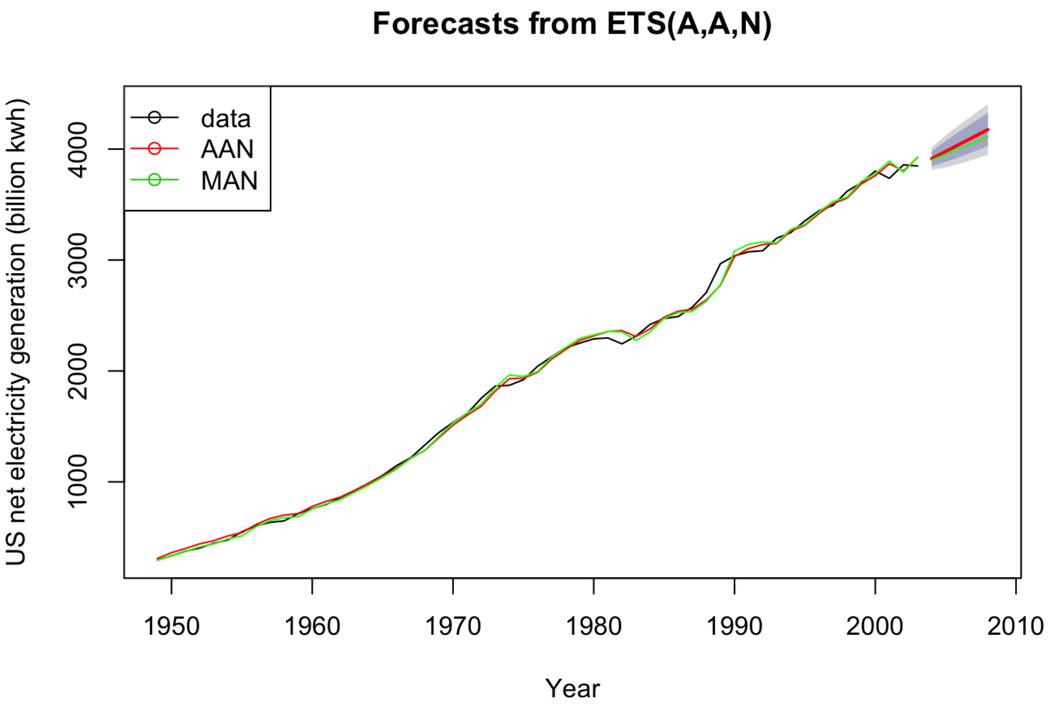
```
## ETS(A,A,N)
##
## Call:
##  ets(y = usnetelec, model = "AAN")
##
##   Smoothing parameters:
##     alpha = 0.9985
##     beta  = 1e-04
##
##   Initial states:
##     l = 244.2092
##     b = 65.5768
##
##   sigma:  51.8636
##
##      AIC     AICc      BIC
## 660.5982 661.8227 670.6349
##
## Training set error measures:
##                       ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -0.04060982 49.94203 36.44997 -0.8111995 2.440206 0.5164647
##                    ACF1
## Training set 0.1263413
```

```
fit.MAN = ets(usnetelec, model="MAN")
summary(fit.MAN)
```

```
## ETS(M,A,N)
##
## Call:
##  ets(y = usnetelec, model = "MAN")
##
##   Smoothing parameters:
##     alpha = 0.9999
##     beta  = 0.2191
##
##   Initial states:
##     l = 254.9338
##     b = 38.3125
##
##   sigma:  0.0259
##
##      AIC     AICc      BIC
## 634.0437 635.2682 644.0803
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 1.162583 52.00363 36.77721 0.2629582 1.942062 0.5211014
##                     ACF1
## Training set 0.006113498
```

The model with additive error terms gives better fits and forecasts in terms of AIC and RMSE. The following display shows the original series, fits and forecasts from both models.

```
frc.AAN = forecast(fit.AAN, h =5) # Produce forecasts for AAN model
frc.MAN = forecast(fit.MAN, h =5) # Produce forecasts for MAN model
plot(frc.AAN, ylab="US net electricity generation (billion kwh)",plot.conf=FALSE, ty
  pe="l", fcol="red", xlab="Year")
lines(fitted(fit.AAN), col="red", lty=1)
lines(fitted(fit.MAN), col="green", lty=1)
lines(frc.MAN$mean,col="green", type="l")
legend("topleft",lty=1, pch=1, col=1:3, c("data","AAN", "MAN"))
```



As another illustration, let's focus on the quarterly motor vehicle production in the UK (thousands of cars) series, which we have fitted a bunch of different seasonal models with different trend structures. Let's fit Holt-Winters' additive method with damped trend and additive and multiplicative errors using state-space approach. Recall that the additive model with damped trend was giving the smallest RMSE value for this series.

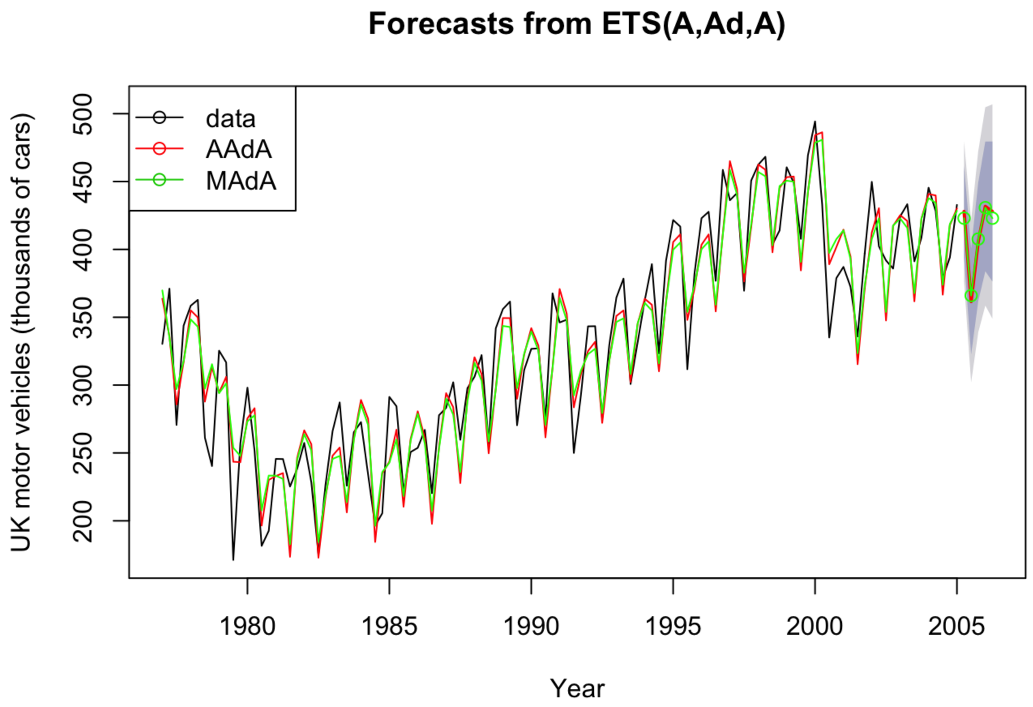
```
fit.AAdA = ets(ukcars, model = "AAA", damped = TRUE)
summary(fit.AAdA)
```

```
## ETS(A,Ad,A)
##
## Call:
##  ets(y = ukcars, model = "AAA", damped = TRUE)
##
##   Smoothing parameters:
##     alpha = 0.5814
##     beta  = 1e-04
##     gamma = 1e-04
##     phi   = 0.9284
##
##   Initial states:
##     l = 343.6012
##     b = -5.3444
##     s = -1.1652 -45.1153 21.2507 25.0298
##
##   sigma:  26.2512
##
##      AIC     AICc      BIC
## 1283.319 1285.476 1310.593
##
## Training set error measures:
##                    ME     RMSE      MAE     MPE     MAPE      MASE
## Training set 2.009896 25.18409 20.44382 0.10939 6.683841 0.6662543
##                    ACF1
## Training set 0.03323651
```

```
fit.MAdA = ets(ukcars, model = "MAA", damped = TRUE)
summary(fit.MAdA)
```

```
## ETS(M,Ad,A)
##
## Call:
##  ets(y = ukcars, model = "MAA", damped = TRUE)
##
##   Smoothing parameters:
##     alpha = 0.5271
##     beta  = 1e-04
##     gamma = 0.0142
##     phi   = 0.8859
##
##   Initial states:
##     l = 353.5972
##     b = -7.1641
##     s = 1.1374 -37.8311 14.1752 22.5184
##
##   sigma:  0.0895
##
##      AIC     AICc      BIC
## 1305.171 1307.328 1332.445
##
## Training set error measures:
##                    ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 1.812806 26.02203 20.95131 -0.1646796 6.833628 0.6827932
##                    ACF1
## Training set 0.06436705
```

```
frc.AAdA = forecast(fit.AAdA, h =5)
frc.MAdA = forecast(fit.MAdA, h =5)
plot(frc.AAdA, ylab="UK motor vehicles (thousands of cars)",plot.conf=FALSE, type=
  "l", fcol="red", xlab="Year")
lines(fitted(fit.AAdA), col="red", lty=1)
lines(fitted(fit.MAdA), col="green", lty=1)
lines(frc.MAdA$mean,col="green", type="o")
legend("topleft",lty=1, pch=1, col=1:3, c("data","AAdA ", "MAdA "))
```



We get a better fit and forecasts from additive damped trend Holt-Winters' model with the additive errors state-space model.

Summary

In this module, we focused on exponential smoothing methods that provide a bunch of models to handle trend and seasonality in the series using additive and multiplicative seasonal and damped and exponential trend patterns. We considered 15 different exponential smoothing approaches.

Then, we moved on to state-space model formulation of the exponential smoothing models which enriches the class of exponential smoothing models by introducing additive and multiplicative error terms. In total, we can fit any of 30 different exponential smoothing models by going over the state-space model formulation. Also, we considered auto-selection of the state space models for exponential smoothing.

References

Hyndman, R.J., Koehler, A.B., Ord, J.K., and Snyder, R.D. (2008). Forecasting with exponential smoothing: the state space approach (http://www.exponentialsmoothing.net), Springer-Verlag.

Hyndman, R.J., Athanasopoulos, G. (2014). *Forecasting: Principles and Practice*. OTEXTS.