Applied Economic Forecasting

6. Moving Averages & Exponential Smoothing

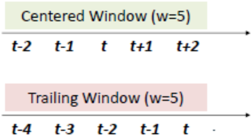
1. Moving Average Method
2. Simple exponential smoothing

Section 1

Moving Average Method

1. Centered Moving Averages

- These are powerful for visualizing the trends because they suppress seasonality and noise (the random component).
- These are not as useful for forecasting.



1. Centered Moving Averages

- In a centered moving average, the value of the moving average at time t (MA_t) is computed by centering the window around time t and averaging across the `k` values within the window:

$$MA_t = \frac{1}{2k+1} \sum_{j=-k}^{k} x_{t+j},$$

where t changes from $k+1$ to $n-k$.

For example, with a window of width `w` = 3,

- the moving average at time point t = 2 means averaging the values of the series at time points 1, 2, 3 ;
- at time point t = 3 the moving average is the average of the values at time points 2, 3, 4 and so on.

1. Centered Moving Averages

In short, the moving average is centered at the middle of the values being averaged.

Time	Values	Centered MA(3)
1	4	N/A
2	5	4
3	3	4
4	4	4
5	5	N/A

2. Simple Moving Averages (Trailing Windows)

- A constant number of data points (fixed window length) are used for forecasting.
- The mean is computed by adding the newest value and dropping the oldest.
- A moving average of order k (the window length), $MA(k)$, is

$$\hat{y}_{t+1} = \frac{y_t + y_{t-1} + \cdots + y_{t-k+1}}{k} = \frac{\sum_{i=1}^{k} y_{t-i+1}}{k}$$

where $\hat{y}_{t+1}$ is the forecast value for the next period. y_t is the actual value at period t, k is the number of terms in the moving average

2. Simple Moving Averages (Trailing Windows)

- Calculates the simple arithmetic mean of the data of the `k` most recent data.
    - `k` is determined by the researcher.
- Equal weights are assigned to each observation.
- Each new data point is included in the average as it becomes available, and the earliest data point is discarded.
- The rate of response to changes in the underlying data pattern depends on the number of periods, `k`, included in the moving average.

Choosing the Forecast Window (`k`)

The choice of the smoothing parameter is a balance between under-smoothing and oversmoothing.

Narrower windows
: the smaller k is better for forecasting in unstable economic environments.
- can more quickly capture the most recent changes in economic conditions. - reveal local trend.

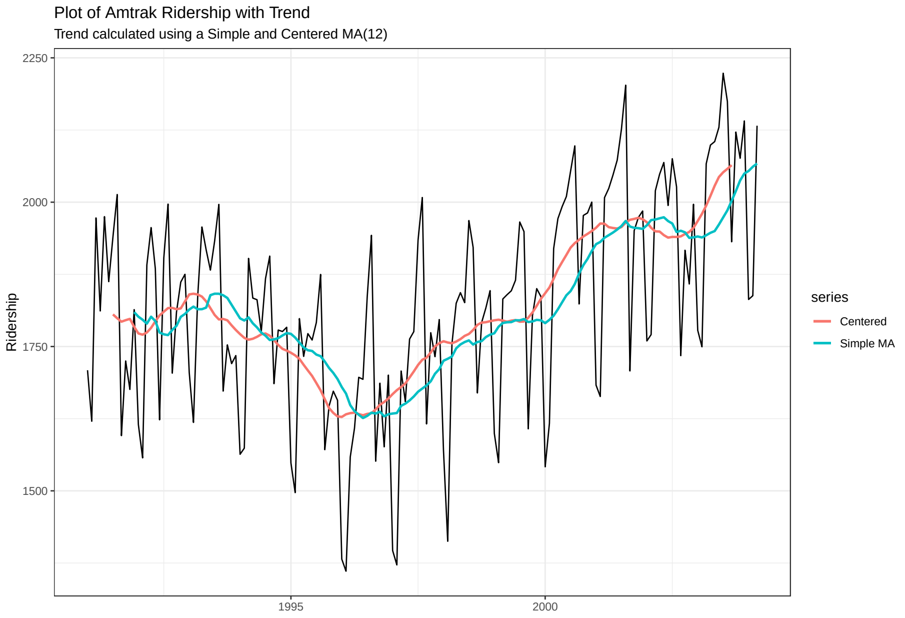
Wider windows
: the larger k is better for forecasting in stable economic environments. - will expose more global trends

When k=1, MA becomes the naive forecasting, as $\hat{y}_{t+1} = y_t$.

Examining the Amtrak Ridership Data

Let us apply the knowledge above the some real world data.

```r
# Import the Amtrak ridership data from github
fileurl <- "https://raw.github.com/Shamar-Stewart/ForecastingS21/master/Lectures/Lecture6/Amtrak_data.csv"
amtrak <- ts(read.csv(fileurl)[,2],
             frequency = 12, start = c(1991,1))

################# Centered MA(12)
order <- 12; ld <- length(amtrak)
ca.ma12 <- ma(amtrak,order = order, centre = TRUE)

################## Creating Rolling MA(12)

order <- 12
ld <- length(amtrak)
sma12 <- matrix(NA,ld)

for (j in 1:(ld - 11)){
sma12[order+(j-1)] <- mean(amtrak[j: (11 +j)])
}
sma12 <- ts(sma12, frequency = 12, start = c(1991,1))

# Using the zoo library
zoo.sma12 <- zoo::rollmean(amtrak,k = 12, fill = NA,
                           align = "right")

# Plot the Amtrak data & MA(12) methods

autoplot(amtrak) + autolayer(ca.ma12 , series = "Centered", lwd = 0.9) +
  autolayer(sma12, series = "Simple MA", lwd = 0.9) +
  labs(title = "Plot of Amtrak Ridership with Trend",
       subtitle = "Trend calculated using a Simple and Centered MA(12)", x = "", y = "Ridership")
```



Plot of Amtrak Ridership with Trend
Trend calculated using a Simple and Centered MA(12)
Ridership
2250
2000
1750
1500
1995
2000
series
Centered
Simple MA

Examining the Amtrak Ridership Data

Discussion:

- The moving average method is inadequate for generating monthly forecasts because it does not capture the seasonality in the data.
  - Seasons with high ridership are under-forecasted, and seasons with low ridership are over-forecasted as the moving average "lags behind" the actual data.

- A similar issue arises when forecasting a series with a trend:
  - the moving average "lags behind", thereby under-forecasting in the presence of an increasing trend and
  - over-forecasting in the presence of a decreasing trend.

Final Words: Moving Averages

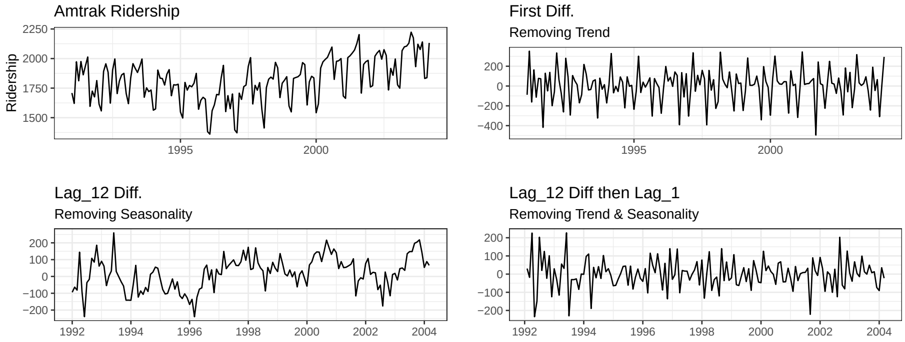
- In general, MAs can be used for forecasting only in series that lack seasonality and trend. There are a few popular methods for removing trends (de-trending) and removing seasonality (deseasonalizing) from a series:
  - Regression models,
  - Advanced exponential smoothing methods, and
  - Differencing.
- The MA method can then be used to forecast such de-trended and deseasonalized series, and then the trend and seasonality can be added back to the forecast (think of the series decompostion from Lecture 5).
- Moving Averages is, in general, better than the simple average method of forecasting.

Differencing

A simple and popular method for removing a trend and/or a seasonal pattern from a series is by the operation of differencing. **Differencing**, as the name suggest, is taking the difference between two values.

- A lag-1 difference (also called first difference) means taking the difference between every two consecutive values in the series ($\Delta y_t = y_t - y_{t-1}$)
  - Lag-1 differencing measures the changes from one period to the next.
- Differencing at lag-k means subtracting the value from k periods back ($y_t - y_{t-k}$).
  - E.g., for a daily series, lag-7 differencing means subtracting from each value (y_t) the value on the same day in the previous week (y_{t-7}).
- We can difference the original series and obtain a differenced series that lacks trend and seasonality.

```r
g1 <- autoplot(amtrak) + labs(title = "Amtrak Ridership", x = "", y = "Ridership")
g2 <- autoplot(diff(amtrak, 1)) + labs(title = "First Diff.", subtitle = "Removing Trend",
    x = "", y = "")
g3 <- autoplot(diff(amtrak, 12)) + labs(title = "Lag_12 Diff.", subtitle = "Removing Seasonality",
    x = "", y = "")
g4 <- autoplot(diff(diff(amtrak, 12), 1)) + labs(title = "Lag_12 Diff then Lag_1",
    subtitle = "Removing Trend & Seasonality", x = "", y = "")
gridExtra::grid.arrange(g1, g2, g3, g4, ncol = 2)
```



Section 2

Simple exponential smoothing

Simple exponential smoothing

Problem

In the case of both the simple and moving averages, we assigned equal weights to the most recent observations as well as those far into the past.

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

The issue is clear but how do we address this?

Solution: Exponential Smoothing Methods

- Assign less weights to past observations
- Assign higher weights to more recent data

Exponential Smoothing Methods

This method is suitable for forecasting data with no clear trend or seasonal pattern.

- The exponential smoothing equation can take the form

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

where $0 \leq \alpha \leq 1$ is the smoothing constant or weighting factor.

- Eqn. 1 can be rewritten as

$$\begin{aligned} \hat{y}_{t+1} &= \alpha y_t + \hat{y}_t - \alpha \hat{y}_t \\ &= \hat{y}_t + \alpha\,(y_t - \hat{y}_t) \end{aligned}$$

- The new forecast is the old forecast adjusted by α times the forecast error in the old forecast

Exponential Smoothing Methods

Question:

In this equation,

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

1. What happens if $\alpha \to 1$?

Answer: As α is getting bigger, more weight is given to the most recent observation, but less weight to old information.

2. What happens if $\alpha \to 0$?

Answer: In this case, as α is getting smaller, less weight is given to the most recent observation, but more weight to old information.

Exponential Smoothing Methods

- Substitute $\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\,\hat{y}_{t-1}$ into Eqn 1 to yield

$$\begin{aligned}\hat{y}_{t+1} &= \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + (1-\alpha)\,\hat{y}_{t-1}\right] \\ &= \alpha y_t + \alpha(1-\alpha)\,y_{t-1} + (1-\alpha)^2\,\hat{y}_{t-1}\end{aligned}$$

Exponential Smoothing Methods

- Substitute $\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\,\hat{y}_{t-1}$ into Eqn 1 to yield

$$\begin{aligned}\hat{y}_{t+1} &= \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + (1-\alpha)\,\hat{y}_{t-1}\right] \\ &= \alpha y_t + \alpha\,(1-\alpha)\,y_{t-1} + (1-\alpha)^2\,\hat{y}_{t-1}\end{aligned}$$

- Continue this substitution to obtain:

$$\begin{gathered}\hat{y}_{t+1} = \alpha y_t + \alpha\,(1-\alpha)\,y_{t-1} + \alpha\,(1-\alpha)^2\,y_{t-2} + \\ \alpha\,(1-\alpha)^3\,y_{t-3} + \alpha\,(1-\alpha)^4\,y_{t-4} + \cdots \end{gathered} \tag{2}$$

Note α determines the speed of decaying impacts of past observations on the forecast value

Exponential Smoothing Methods

To use exponential smoothing methods, we need to determine:

1. the value of α (can be determined by minimizing MSE)
2. the initial value for the smoothing
   a. the first observation
   b. the average of the first five or six observations

Exponential Smoothing Methods

Example: Use the sales data from Acme Tool Company.

1. Consider $\alpha = 0.1,\ 0.6$
2. Consider the initial period forecast: $\hat{y}_1 = 500$

Time Year	Time Quarters	Actual Value Y_t	Smoothed Value $\hat{Y}_t(\alpha = .1)$	Forecast Error e_t	Smoothed Value $\hat{Y}_t(\alpha = .6)$	Forecast Error e_t
2000	1	500	500.0	0.0	500.0	0.0
	2	350	500.0	−150.0	500.0	−150.0
	3	250	485.0 (1)	−235.0 (2)	410.0	−160.0
	4	400	461.5 (3)	−61.5	314.0	86.0
2001	5	450	455.4	−5.4	365.6	84.4
	6	350	454.8	−104.8	416.2	−66.2
	7	200	444.3	−244.3	376.5	−176.5
	8	300	419.9	−119.9	270.6	29.4
2002	9	350	407.9	−57.9	288.2	61.8
	10	200	402.1	−202.1	325.3	−125.3
	11	150	381.9	−231.9	250.1	−100.1
	12	400	358.7	41.3	190.0	210.0
2003	13	550	362.8	187.2	316.0	234.0
	14	350	381.6	−31.5	456.4	−106.4
	15	250	378.4	−128.4	392.6	−142.6
	16	550	365.6	184.4	307.0	243.0
2004	17	550	384.0	166.0	452.8	97.2
	18	400	400.6	−0.6	511.1	−111.1
	19	350	400.5	−50.5	444.5	−94.5
	20	600	395.5	204.5	387.8	212.2
2005	21	750	415.9	334.1	515.1	234.9
	22	500	449.3	−50.7	656.0	−156.0
	23	400	454.4	−54.4	562.4	−162.4
	24	650	449.0	201.0	465.0	185.0
2006	25	850	469.0	381.0	576.0	274.0

Figure 1: Exponentially Smoothed Values for Acme Tools Company

Exponential Smoothing Methods

3. Use 1 at 2000Q2 to forecast for 2000Q3 with $\alpha = 0.1$

$$\hat{y}_{2000Q3} = \alpha y_{2000Q2} + (1-\alpha)\hat{y}_{2000Q2} = 0.1 * 350 + 0.9 * 500 = 485$$

4. The forecast error is

$$e_{2000Q3} = y_{2000Q3} - \hat{y}_{2000Q3} = 250 - 485 = -235$$

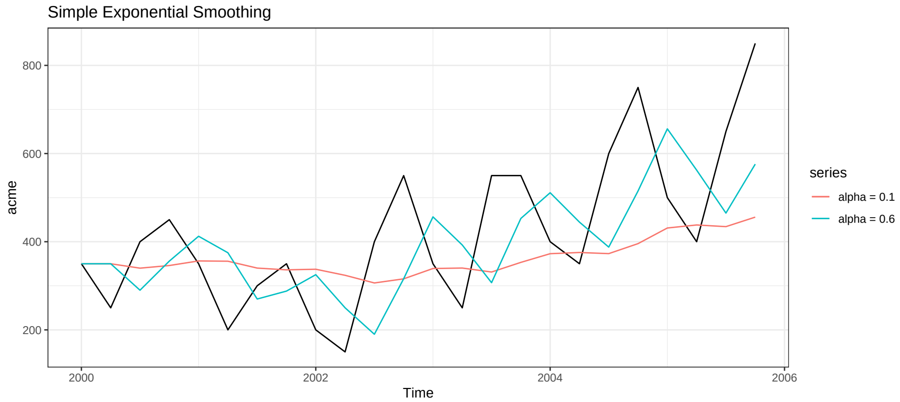
5. Use 1 at 2000Q3 to forecast for 2000Q4 with $\alpha = 0.1$

$$\hat{y}_{2000Q4} = \alpha y_{2000Q3} + (1-\alpha)\hat{y}_{2000Q3} = 0.1 * 250 + 0.9 * 485 = 461.5$$

6. The forecast error is

$$e_{2000Q4} = y_{2000Q4} - \hat{y}_{200Q4} = 400 - 461.5 = -61.5$$

```r
acme <- ts(read.csv("Acme.csv"),frequency = 4,start=c(2000,1))
ses.acme1 <- ses(acme,initial = "simple", alpha = 0.1)
ses.acme2 <- ses(acme,initial = "simple", alpha = 0.6)

autoplot(acme) + autolayer(ses.acme1$fitted, series = "alpha = 0.1") +
  autolayer(ses.acme2$fitted, series = "alpha = 0.6") +
  labs(title = "Simple Exponential Smoothing")
```



```r
acme1_MSE <- mean(ses.acme1$residuals^2)
acme2_MSE <- mean(ses.acme2$residuals^2)
acme1_MAPE <- mean(abs(ses.acme1$residuals/acme)*100)
acme2_MAPE <- mean(abs(ses.acme2$residuals/acme)*100)
c1 <- c("MSE" = acme1_MSE, "MAPE" = acme1_MAPE)
c2 <- c("MSE" = acme2_MSE, "MAPE" = acme2_MAPE)
knitr:: kable(cbind('$\\alpha = 0.1$'= c1, '$\\alpha = 0.6$' = c2),
  format.args = list(big.mark = ","),
  caption = "Comparison of forecasting Performance")
```

Table 2: Comparison of forecasting Performance

	$\alpha = 0.1$	$\alpha = 0.6$
MSE	26,866.22	24,015.47
MAPE	30.94	35.38

$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2; \quad MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|e_t|}{|y_t|} \times 100$$

Notice: Both MSE and MAPE are large. This means that the methods are not doing a great job of capturing the dynamics of the data.

Solution: In addition to α, we try different initial values.

Simple Exponential Smoothing

Alternatively, we can allow `R` to do the optimization and provide us with the optimal values of α, and the initial forecast value:

```
ses(acme, initial = "optimal", alpha = NULL)
```

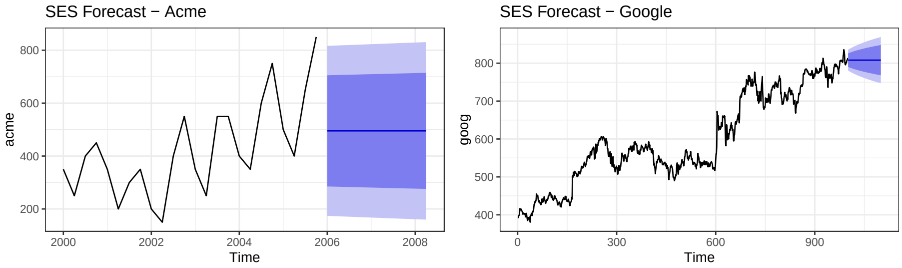
We can then view the optimal model using

```
ses(acme, initial = "optimal", alpha = NULL)$model
```

```
## Simple exponential smoothing
##
## Call:
##  ses(y = acme, initial = "optimal", alpha = NULL)
##
##   Smoothing parameters:
##     alpha = 0.3225
##
##   Initial states:
##     l = 334.754
##
##   sigma:  156
##
```

Forecast from Simple Exponential Smoothing

```r
g1 <- autoplot(ses.acme1) + labs(title = "SES Forecast - Acme")
ses.goog <- ses(goog, alpha = .2, h = 100)
g2 <- autoplot(ses.goog) + labs(title = "SES Forecast - Google")
gridExtra::grid.arrange(g1,g2,ncol = 2)
```



CAUTION!!!

We see that our forecast projects a flatlined estimate into the future. This does not capture the possible positive trend in the data.
This is why a simple exponential smoothing should not be used on data with a trend or seasonal component.

Exponential Smoothing Adjusted for Trend

Holt's Method

Holt's Method

- allows for evolving local linear trends in a time series
- can be used to generate forecasts
- Advantage: flexible to track changing in level and trend

Equations for Holt's smoothing

1. The exponentially smoothed series, or current level estimate

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \tag{3}$$

where L_t is the new smoothed value (estimate of current level).

Eq. 3 is very similar to the equation for simple exponential smoothing, except that a term (T_{t-1}) has been incorporated to properly update the level when a trend exists.

- $\alpha \in (0, 1)$ is the smoothing constant for the level
- y_t is the actual value at time t
- T_{t-1} is the trend estimate at time $t - 1$

Holt's Linear Trend Method

2. Trend Estimate

$$T_t = \beta \left(L_t - L_{t-1} \right) + (1 - \beta) \, T_{t-1} \tag{4}$$

where $\beta \in (0, 1)$ is the smoothing constant for the trend estimate.

3. The forecast for p periods into the future

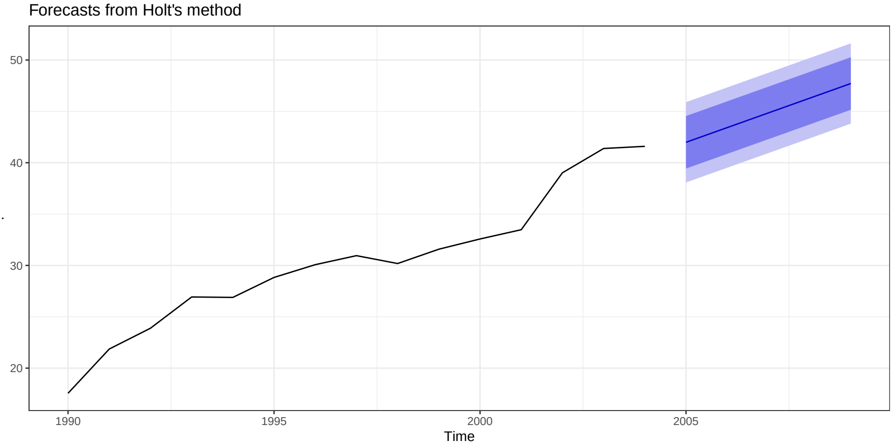
$$\hat{y}_{t+p} = L_t + pT_t \tag{5}$$

where p is the periods to be forecast into the future. Note that the forecasts for future periods lie along a straight line with the slope T_t and intercept L_t.

Two smoothing parameters α and β $(0 \leq \alpha, \beta \leq 1)$.

Holt's method in R

```r
air <- window(ausair, start=1990, end=2004)
air %>% holt(h=5) %>% autoplot()
```

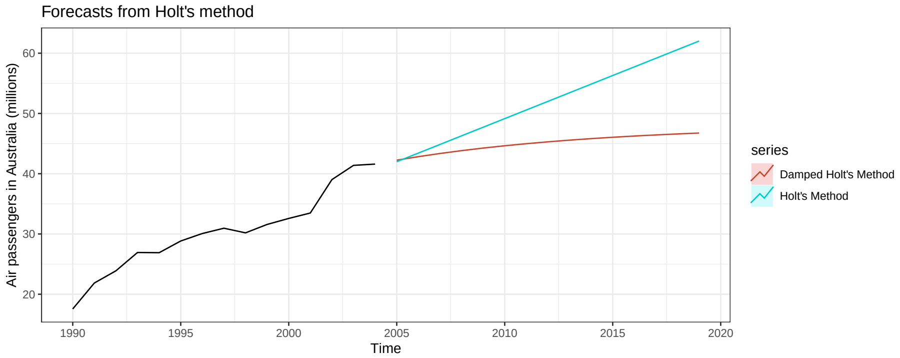


Holt's Linear Trend Method with a Damped Trend

- The forecasts generated by Holt's linear method display a constant trend (increasing or decreasing) indefinitely into the future.
- Empirical evidence indicates that these methods tend to over-forecast, especially for longer forecast horizons.
- Gardner & McKenzie (1985) introduced a parameter that "dampens" the trend to a flat line some time in the future.
- Damping parameter $0 < \phi < 1$.
- If $\phi = 1$, identical to Holt's linear trend.

Example: Air passengers

```r
air %>% holt(damped=TRUE, h=15, phi = 0.9) -> w.damped
air %>%  holt(damped=FALSE, h=15) -> n.damped
autoplot(air) +
  autolayer(w.damped, series = "Damped Holt's Method", PI = FALSE) +
  autolayer(n.damped, series = "Holt's Method", PI = FALSE) +
  labs(title = "Forecasts from Holt's method",
       y = "Air passengers in Australia (millions)")
```



Exponential Smoothing Adjusted for Trend and Seasonal Variation: Winters' Method

- Three parameter linear and seasonal exponential smoothing method
- an extension of Holt's method
- one additional equation for seasonality
- Has both an additive and multiplicative method.

The four equations used in **Holt-Winter**'s (multiplicative) smoothing are:

1. The exponentially smoothed series, or level estimate.

$$L_t = \alpha \frac{y_t}{S_{t-s}} + (1-\alpha)\left(L_{t-1} + T_{t-1}\right) \tag{6}$$

Notice:

In Eq. 6, y_t is divided by S_{t-s} which adjusts y_t for seasonality. Think back to when you did the classical decomposition in Lecture 5.

2. The trend estimate:

$$T_t = \beta\left(L_t - L_{t-1}\right) + (1-\beta)\,T_{t-1} \tag{7}$$

Holt-Winter's Exponential Smoothing Methods

3. The seasonality estimate:

$$S_t = \gamma \frac{y_t}{L_t} + (1-\gamma)\, S_{t-s} \tag{8}$$

We can consider y_t/L_t as a seasonal index ratio that can be used in a multiplicative fashion to adjust a forecast to account for seasonal peaks and valleys as in Eq. 9.

4. The forecast for p periods into the future:

$$\hat{y}_{t+p} = (L_t + pT_t)\, S_{t-s+p} \tag{9}$$

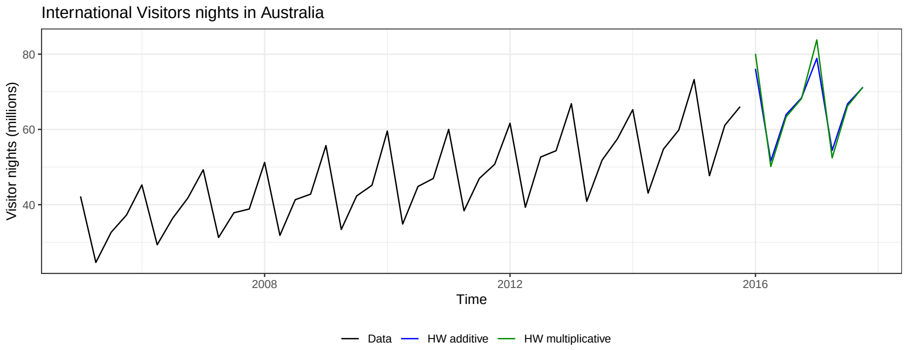
where S_t is the seasonal estimate, γ is the smoothing constant for the seasonality estimate, s is the length of seasonality.

The difference compared to (5) is it is multiplied by S_{t-s+p} to adjust the forecast for seasonality.

Holt-Winter's Exponential Smoothing Methods

- Winter's method provides an easy way to account for seasonality when data have a seasonal pattern.
- In general, exponential smoothing methods have the major advantages of low cost and simplicity.
- The methods assign weights that decline exponentially as the observations get older.

```r
aust <- window(austourists, start=2005)
fit1 <- hw(aust,seasonal="additive")
fit2 <- hw(aust,seasonal="multiplicative")
tmp <- cbind("Data"=aust, "HW additive" = fit1[["mean"]],
             "HW multiplicative" = fit2[["mean"]])
autoplot(tmp) +
  scale_color_manual(name="", values=c('black','blue','green4')) +
  labs(title = "International Visitors nights in Australia",
    y = "Visitor nights (millions)") + theme(legend.position = "bottom")
```



Your turn

1. Return to the Acme Dataset amd see if you can improve the fit by using the Holt and Holt Winter models.

2. Apply Holt-Winters' multiplicative method to the `gas` data.

a. Why is multiplicative seasonality necessary here?
b. Experiment with a dampened trend.
c. Check that the residuals from the best method look like white noise.

R functions

- Simple exponential smoothing: no trend: `ses(y)`
- Holt's method: linear trend: `holt(y)`
- Damped trend method: `holt(y, damped=TRUE)`
- Holt-Winters methods: `hw(y, damped=TRUE, seasonal = "additive")`

  ```
  hw(y, damped=FALSE, seasonal = "additive")
  hw(y, damped=TRUE, seasonal =
  "multiplicative")
  hw(y, damped=FALSE, seasonal =
  "multiplicative")
  ```
- Combination of no trend with seasonality not possible using these functions.