Source: xkcd.com/2586

EXPONENTIAL SMOOTHING MODELS

Dr. Susan Simmons
Institute for Advanced Analytics

INTRODUCTION

Time Dependencies

- Time series data relies on the assumption that the observations at a certain time point depend on previous observations in time.

Time Dependencies

- Time series data relies on the assumption that the observations at a certain time point depend on previous observations in time.

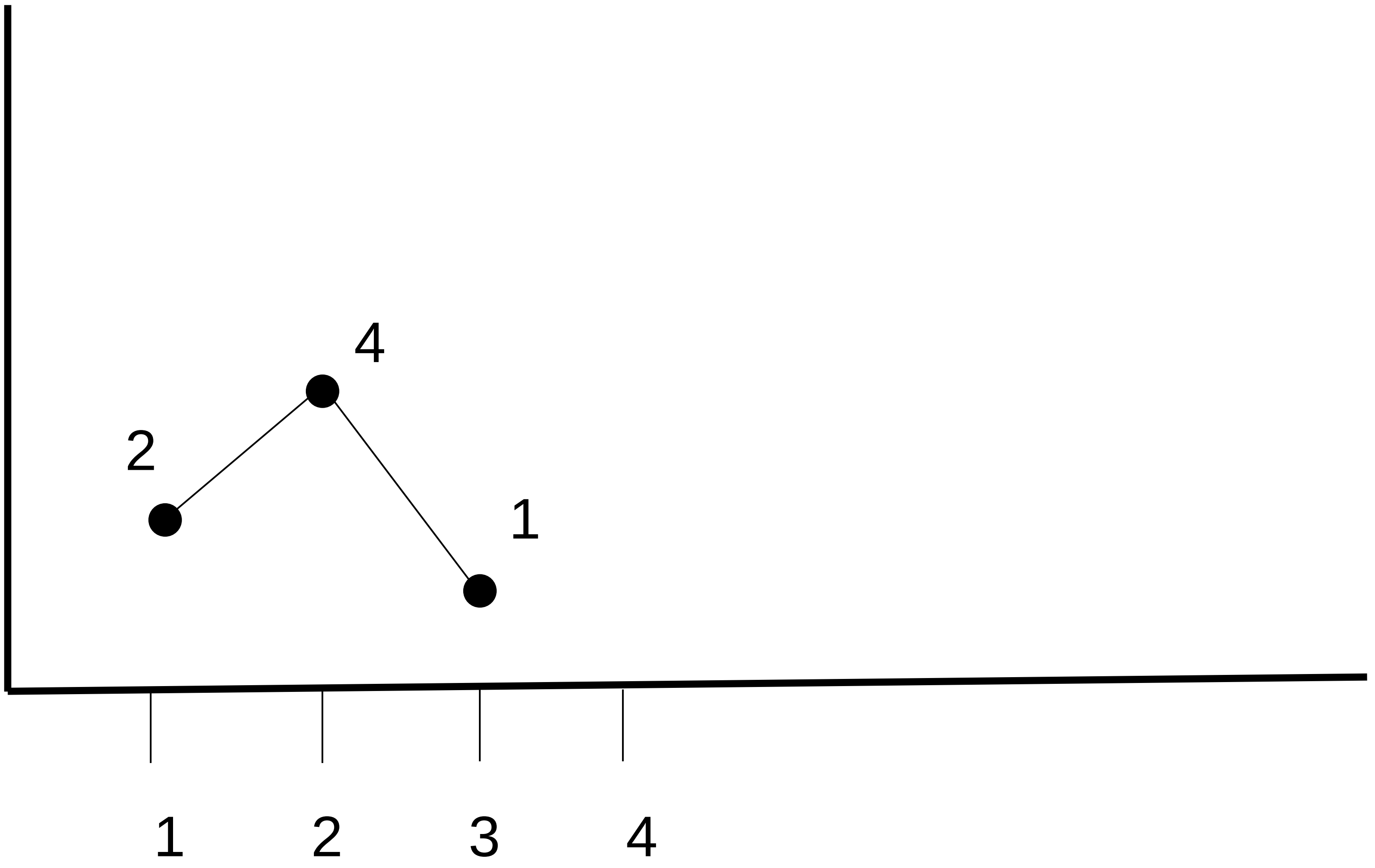
Naïve Model:

$$\hat{Y}_{t+h} = Y_t$$

Average Model:

$$\hat{Y}_{t+h} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

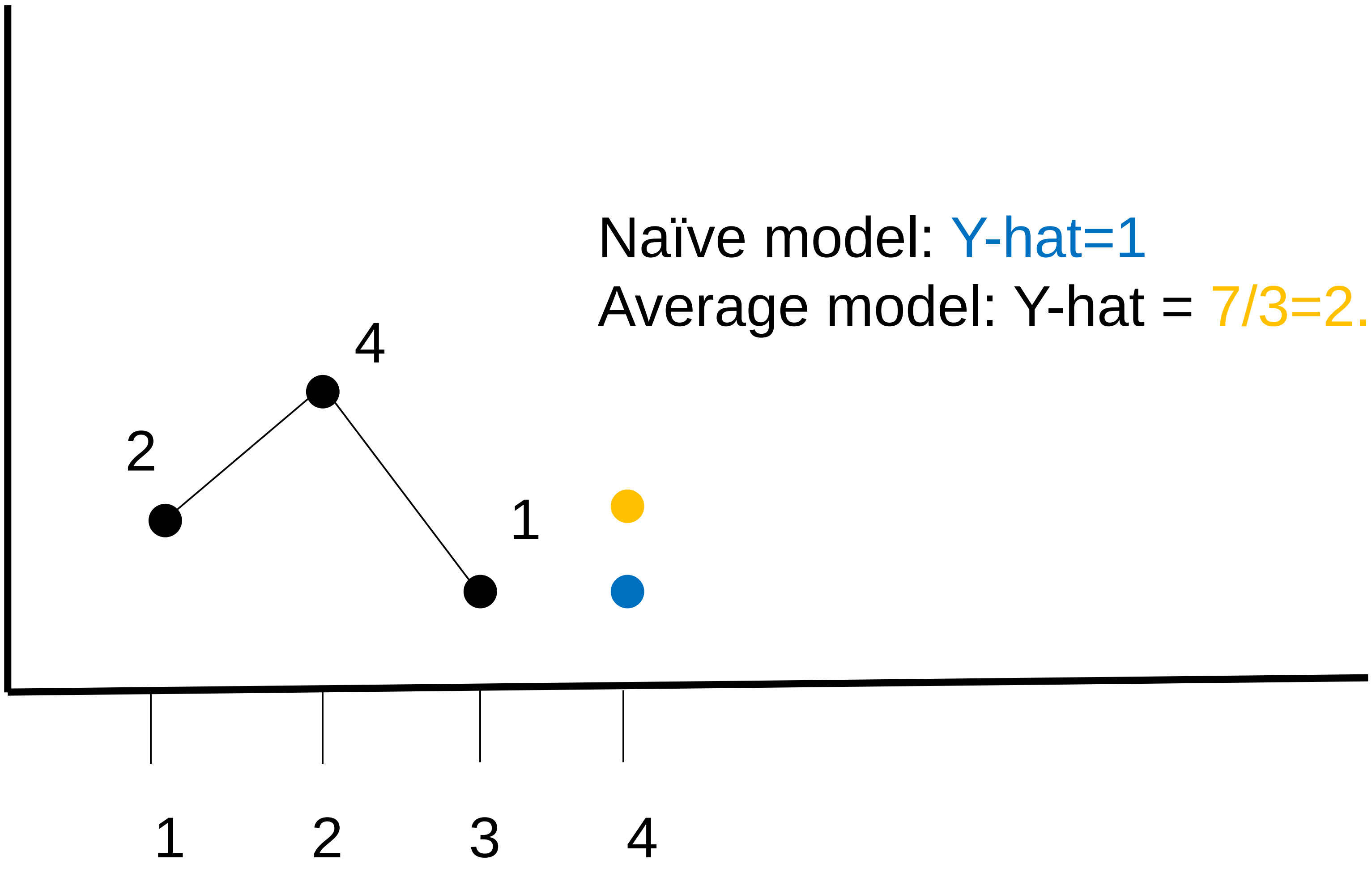
Naïve model versus Average model



Naïve model versus Average model

Naïve model: Y-hat=1

Average model: Y-hat = 7/3=2.333



Exponential Smoothing

- This is what exponential smoothing does (however, it is a WEIGHTED average, not a simple average)
- Models only require a few parameters.
- Equations are simple and easy to implement.

Exponential Smoothing

- There are many different types of exponential smoothing models.
- We will discuss the four common types of Exponential Smoothing:
  - Single
  - Linear / Holt (incorporates trend)
  - Holt-Winters (incorporates trend and seasonality)
- ESM are great for “one-step ahead” forecasting

SINGLE EXPONENTIAL SMOOTHING

Single Exponential Smoothing

- The Single Exponential Smoothing model equates the predictions at time *t* equal to the weighted values of the previous time period along with the previous time period's prediction:

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

Where $\hat{Y}_t$ is the estimate of Y_t (weighted average of previous observations)

Single Exponential Smoothing

- We can apply a weighting scheme that decreases exponentially the further back in time we go.

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

$$\hat{Y}_{t+1} = \theta Y_t + (1-\theta)[\theta Y_{t-1} + (1-\theta)\hat{Y}_{t-1}]$$

$$\hat{Y}_{t+1} = \theta Y_t + \theta(1-\theta)Y_{t-1} + (1-\theta)^2\hat{Y}_{t-1}$$

$$\hat{Y}_{t+1} = \theta Y_t + \theta(1-\theta)Y_{t-1} + \theta(1-\theta)^2 Y_{t-2}$$
$$+(1-\theta)^3\hat{Y}_{t-2}$$
$$\vdots$$

$$\hat{Y}_{t+1} = \theta Y_t + \theta(1-\theta)Y_{t-1} + \theta(1-\theta)^2 Y_{t-2} + \cdots$$

Single Exponential Smoothing

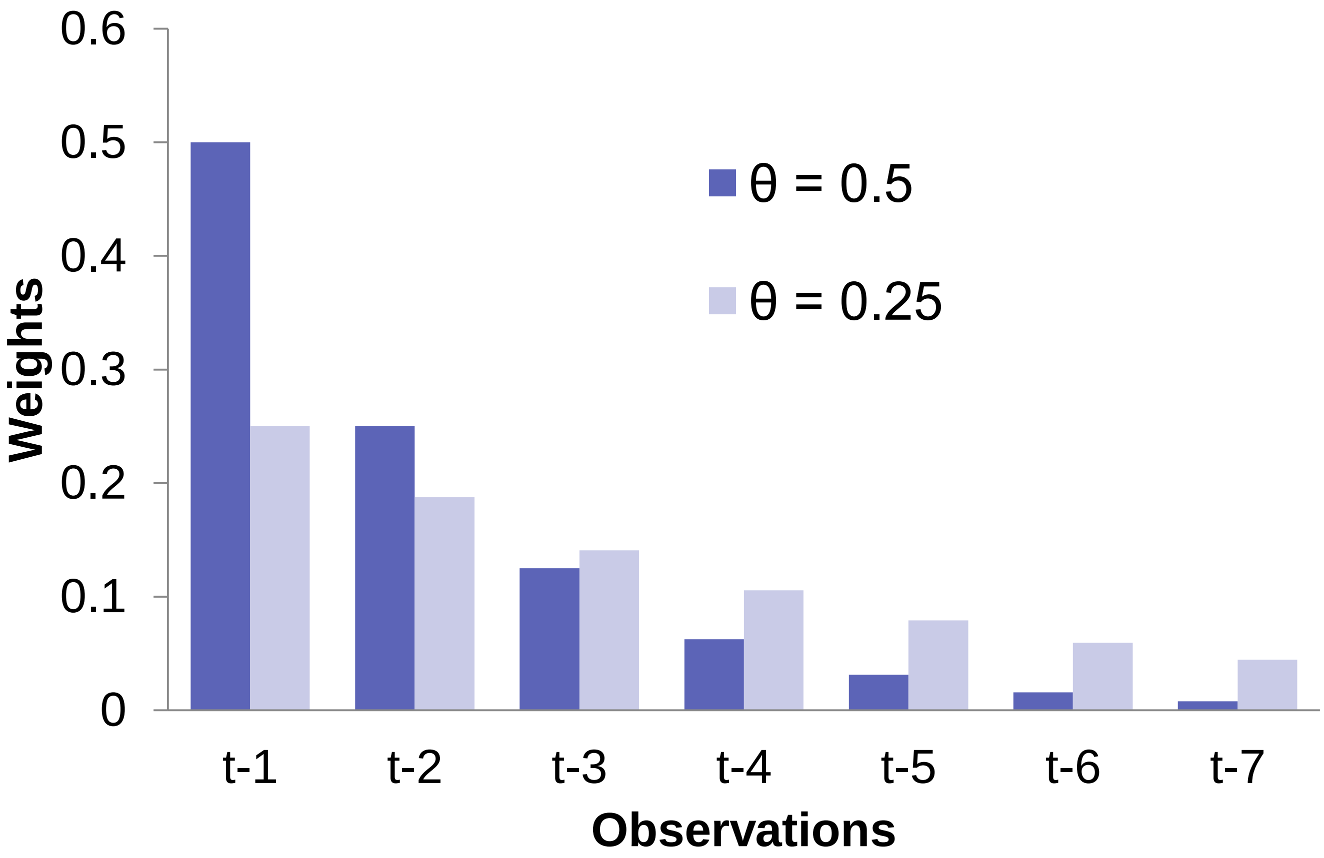
- As you can see, as we go further back in time, the weights decrease exponentially (more weight is put on the most recent observations).

$$\hat{Y}_{t+1} = \theta Y_t + \theta(1-\theta)Y_{t-1} + \theta(1-\theta)^2 Y_{t-2} + \theta(1-\theta)^3 Y_{t-3} + \theta(1-\theta)^4 Y_{t-4} + \cdots$$

$$0 \leq \theta \leq 1$$

Single Exponential Smoothing

- The larger the value of the θ, the more that the most recent observation is emphasized.



Component Form

- The Single ESM can also be written in **component form**:

Forecast Equation: $\hat{Y}_{t+1} = L_t$

Level Equation: $L_t = \theta Y_t + (1 - \theta) L_{t-1}$

Parameter Estimation

$$\hat{Y}_t = \theta Y_{t-1} + (1-\theta)\hat{Y}_{t-1}$$

- The typical method for calculating the optimal value of θ in the Exponential Smoothing model is through one-step ahead forecasts.
- The value of θ that minimizes the one-step ahead forecast errors is considered the optimal value.

$$SSE = \sum_{t=1}^{T} \left(Y_t - \hat{Y}_t\right)^2$$

Parameter Estimation

$$\hat{Y}_t = \theta Y_{t-1} + (1-\theta)\hat{Y}_{t-1}$$

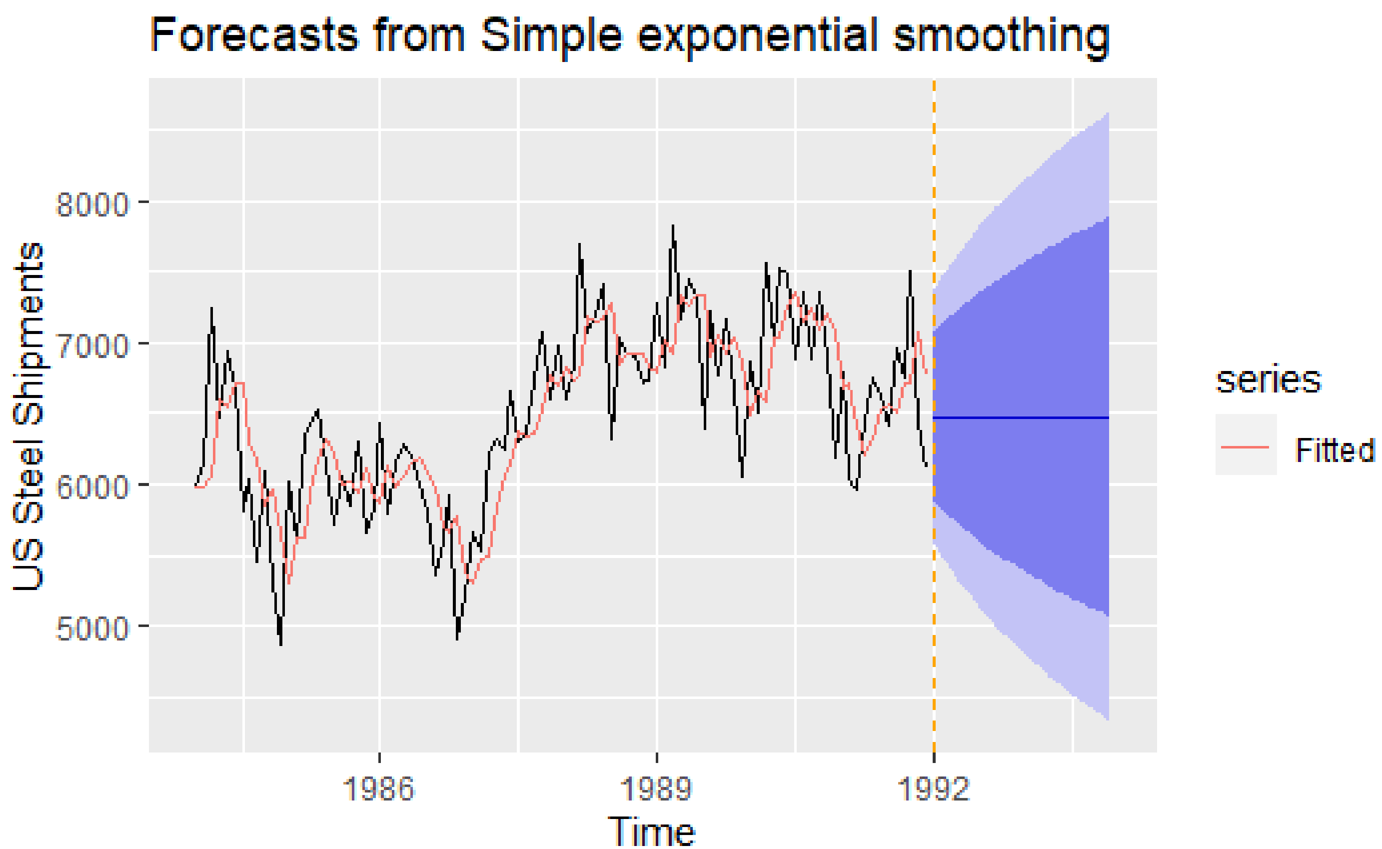
- Estimates that are not statistically significant should not be disqualified (in fact, significance test usually test if θ = 0….which simplifies down to the average model).
- Models were originally derived without statistical distribution consideration (estimates are fine even without normality!).
- HOWEVER, normality is needed if trying to construct a confidence interval.

SES Function

```
ES.Steel <- ses(SteelShp, initial="simple",
                h=24)
summary(ES.Steel)
```

Output from R (edited)

```
Smoothing parameters:
    alpha = 0.4549

  Initial states:
    l = 5980

  sigma:  460.4357
Error measures:
                   ME     RMSE      MAE        MPE     MAPE
Training set 11.43866 460.4357 363.9341 -0.2204828
5.708307
```


Forecasts from Simple exponential smoothing
US Steel Shipments
8000
7000
6000
5000
1986
1989
1992
Time
series
Fitted

LINEAR TREND FOR EXPONENTIAL SMOOTHING

Trending Exponential Smoothing

- The Single Exponential Smoothing model are better used for short-term forecasts.
- The SES model **cannot** adequately handle data that is trending up or down.
- There are multiple ways to incorporate a trend in the Exponential Smoothing Model.
  - Linear / Holt Exponential Smoothing
  - Damped Trend Exponential Smoothing

Linear / Holt Exponential Smoothing

- The Linear Exponential Smoothing model has two components.
- The second component incorporates trending into the model.

$$\hat{Y}_{t+h} = L_t + hT_t$$

$$L_t = \theta Y_t + (1-\theta)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

Linear / Holt Exponential Smoothing

- The Linear Exponential Smoothing model has two components.
- The second component incorporates trending into the model.

$$\hat{Y}_{t+h} = L_t + hT_t$$

$$L_t = \theta Y_t + (1 - \theta)(L_{t-1} + T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1}$$

There are only two parameters to estimate here (both smoothing or “weight” parameters)

Damped Trend Exponential Smoothing

- The Linear Exponential Smoothing model has two components.
- The second component incorporates trending into the model.

$$\hat{Y}_{t+h} = L_t + \sum_{i=1} \phi^i \, T_t$$

$$L_t = \theta Y_t + (1-\theta)(L_{t-1} + \phi T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)\phi T_{t-1}$$

Damped Trend Exponential Smoothing

- The Linear Exponential Smoothing model has two components.
- The second component incorporates trending into the model.

$$\hat{Y}_{t+h} = L_t + \sum_{i=1}^{h} \phi^i T_t$$

Between 0 and 1

$$L_t = \theta Y_t + (1-\theta)(L_{t-1} + \phi T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)\phi T_{t-1}$$

HOLT Function – R

```
LES.Steel <- holt(SteelShp, initial="optimal",
                  h=24)
summary(LES.Steel)

LDES.Steel <- holt(SteelShp, initial="optimal",
                   h=24, damped=TRUE)
summary(LDES.Steel)
```

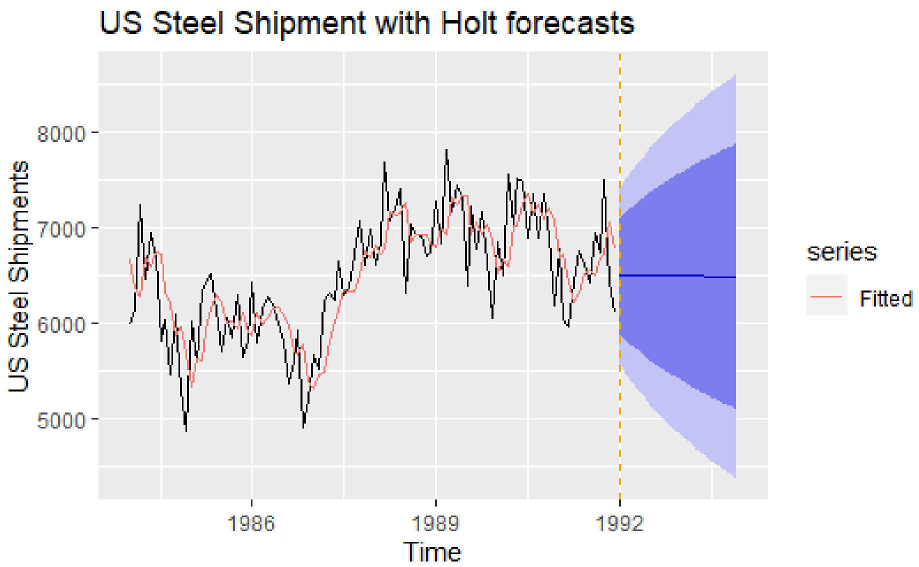
Holt output from R (edited)

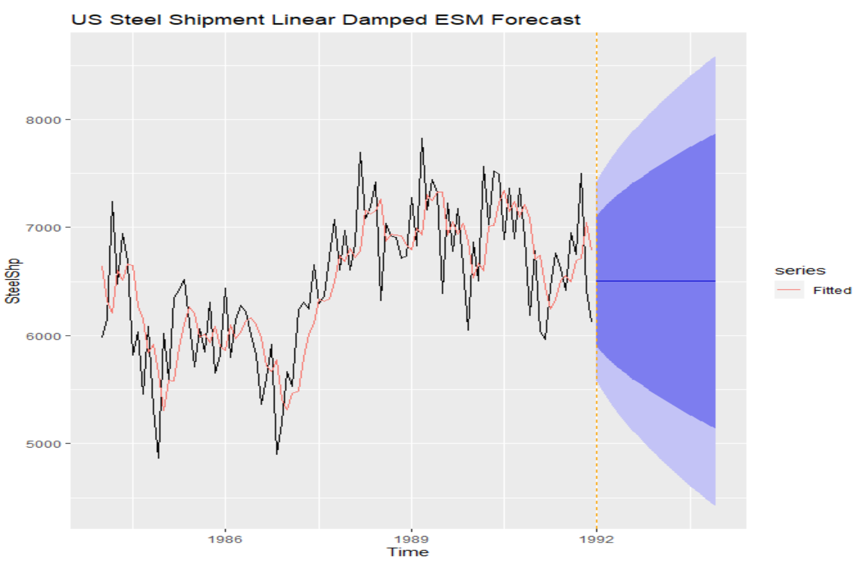
```
Smoothing parameters:
    alpha = 0.4329
    beta  = 1e-04



    AIC     AICc      BIC
1626.001 1626.667 1638.822

Error measures:
                    ME     RMSE      MAE        MPE     MAPE
Training set -4.167318 461.5062 369.9177 -0.4760441
5.818476
```

Damped Holt output from R (edited)

```
Smoothing parameters:
    alpha = 0.5721
    beta  = 1e-04
    phi   = 0.8057


     AIC     AICc      BIC
4906.527 4906.924 4926.862

Error measures:
                   ME     RMSE      MAE        MPE     MAPE
Training set 196.2283 4818.336 3529.079 -0.2896149
7.191284
```

Linear ESM


US Steel Shipment with Holt forecasts
US Steel Shipments
8000
7000
6000
5000
1986
1989
1992
Time
series
Fitted


US Steel Shipment Linear Damped ESM Forecast
SteelShp
8000
7000
6000
5000
1986
1989
1992
Time
series
Fitted

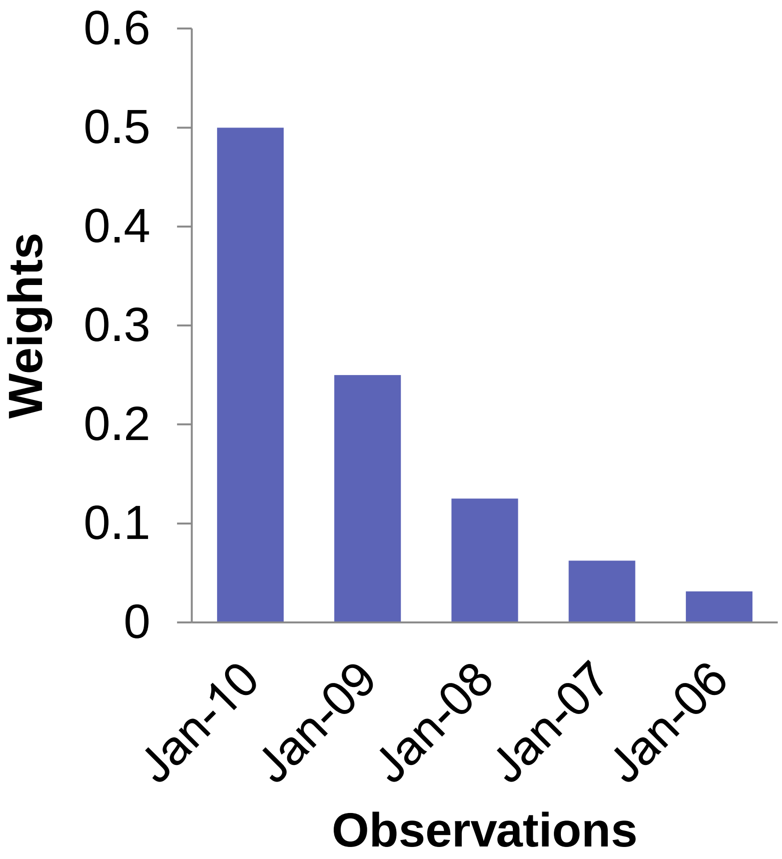
SEASONAL EXPONENTIAL SMOOTHING

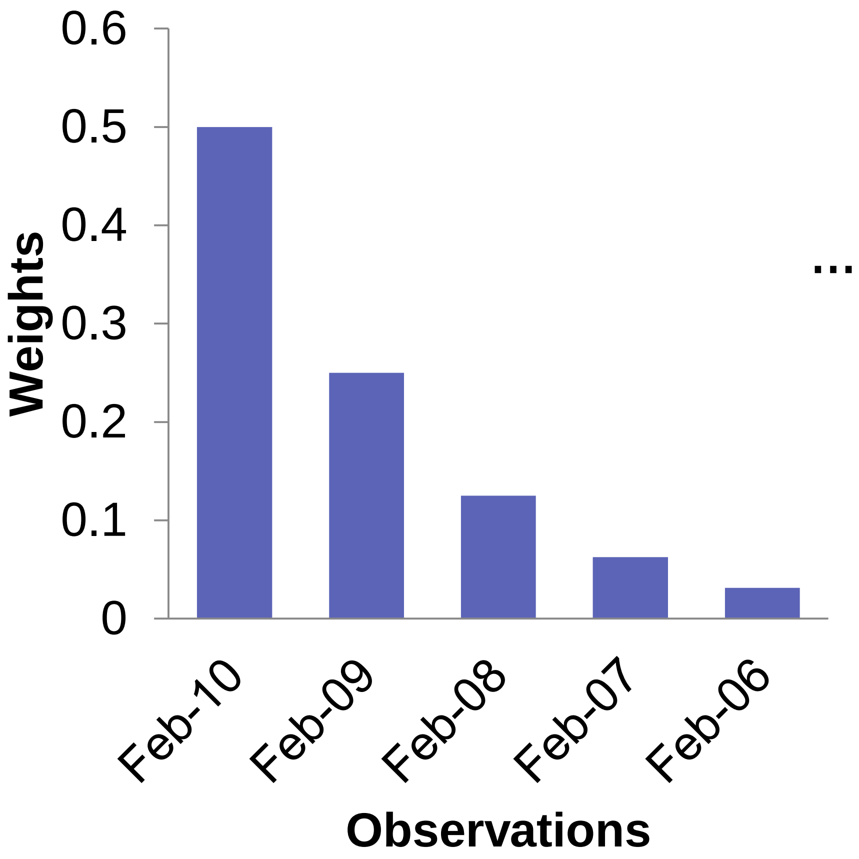
Seasonal Exponential Smoothing

- Exponential Smoothing models can also be adapted to account for seasonal factors.
- Seasonal models can be additive or multiplicative in the seasonal effect in the Exponential Smoothing Model.
  - Holt Winters Additive Exponential Smoothing (includes trend)
  - Holt Winters Multiplicative Exponential Smoothing (includes trend)

Seasonal Exponential Smoothing

- In seasonal exponential smoothing, weights decay with respect to the seasonal factor.





Winters / Triple Exponential Smoothing (Additive)

- The Linear Exponential Smoothing model has three components.
  - Level, Trend and Seasonal

$$\hat{Y}_{t+h} = L_t + hT_t + S_{t-p+h}$$

$$L_t = \theta\left(Y_t - S_{t-p}\right) + (1-\theta)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$
$$S_t = \delta(Y_t - L_{t-1} - T_{t-1}) + (1-\delta)S_{t-p}$$

Winters / Triple Exponential Smoothing (Multiplicative)

- The Linear Exponential Smoothing model has three components.

$$\hat{Y}_{t+h} = (L_t + hT_t)S_{t-p+h}$$

$$L_t = \theta\left(Y_t/S_{t-p}\right) + (1-\theta)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$
$$S_t = \delta(Y_t/(L_{t-1} + T_{t-1})) + (1-\delta)S_{t-p}$$

HW Function

```
HWES.USAir <- hw(Passenger, seasonal="additive")
summary(HWES.USAir)

HWES.USAir <- hw(Passenger, seasonal="multiplicative")
summary(HWES.USAir)
```

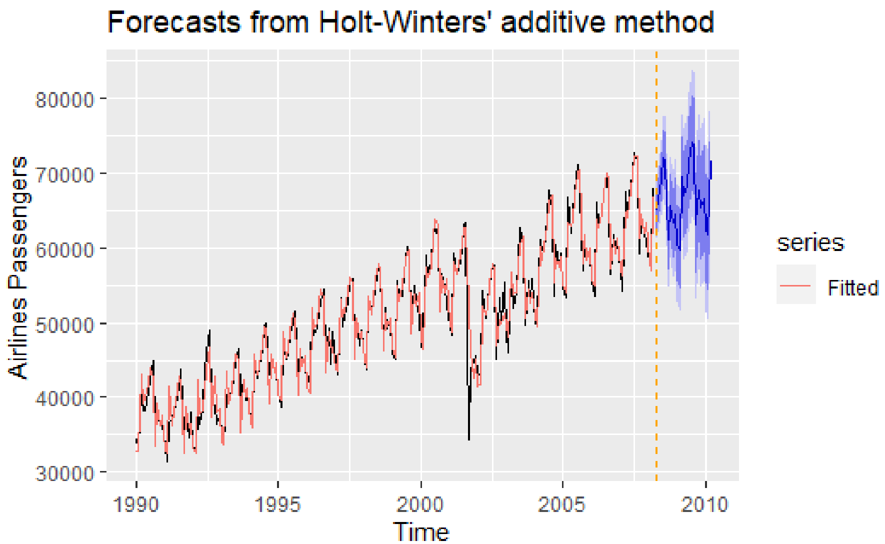
```
Call:
 hw(y = Passenger, seasonal = "additive")

  Smoothing parameters:
    alpha = 0.5967
    beta  = 1e-04
    gamma = 1e-04

sigma:  1949.79

     AIC     AICc      BIC
4515.651 4518.696 4573.265

Error measures:
                    ME     RMSE      MAE        MPE     MAPE      MASE
Training set -84.80235 1877.214 1168.093 -0.2917412 2.495749 0.4389788
```

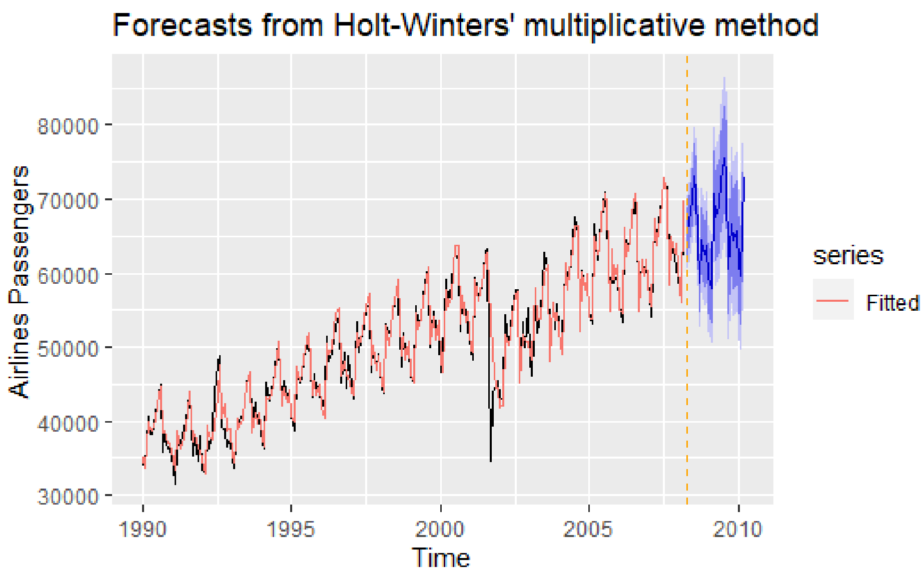
ADDITIVE

```
Call:
 hw(y = Passenger, seasonal = "multiplicative")

  Smoothing parameters:
    alpha = 0.4372
    beta  = 1e-04
    gamma = 0.2075


 sigma:  0.0381

     AIC     AICc      BIC
4504.228 4507.272 4561.842

Error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE
Training set -113.1889 1848.797 1090.105 -0.383246 2.303162 0.4096702
```

MULTIPLICATIVE

Additive Model versus Multiplicative Model





EVALUATING FORECASTS

Forecasting Strategy

- Accuracy of forecasts depends on your definition of accuracy.
  - Different across different fields of industry.

- Good forecasts should have the following characteristics:
  - Be highly correlated with actual series values
  - Exhibit small forecast errors
  - Capture the important features of the original time series.

Judgment Forecasting

- When using data, forecasts are found using quantitative (or modeling) approaches. However, there are instances where models are not available (or potentially past data is not available) and a qualitative or judgement forecast is used.
- Occasionally a qualitative and quantitative approach are merged together.

Accuracy vs. Goodness-of-Fit

- A diagnostic statistic calculated using the same sample that was used to build the model is a ***goodness-of-fit*** statistic.
- A diagnostic statistic calculated using a hold out sample that was not used in the building of the model is an ***accuracy*** statistic.

Hold-out Sample

- A hold out sample in time series analysis is different than cross-sectional analysis.
- The hold-out sample is always at the end of the time series, and doesn't typically go beyond 25% of the data.
- IF YOU HAVE A SEASONAL TIME SERIES (Fall 2) Ideally, an entire season should be captured in a hold-out sample.

Hold-out Sample

1. Divide the time series into two or three segments – training and validation (hold-out) and/or test.
2. Derive a set of candidate models.
3. Calculate the chosen *accuracy* statistic by forecasting the validation data set.
4. Pick the model with the best accuracy statistic.
5. Provide the accuracy of the model on the *test* data set.

Model Diagnostic Statistics

1. Mean Absolute Percent Error:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|$$

2. Mean Absolute Error:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$$

Model Diagnostic Statistics

1. Mean Absolute Percent Error:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \longrightarrow$$

Problems:
- Overweight of Over-predictions
- Actual of 0

2. Mean Absolute Error:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| \longrightarrow$$

Problems:
- Not scale invariant

Model Diagnostic Statistics

3. Square Root of Mean Square Error:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$$

4. Symmetric Mean Absolute Percent Error:

$$\text{sMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{\left(|Y_t| + |\hat{Y}_t|\right)}$$

Model Diagnostic Statistics

3. Square Root of Mean Square Error:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$$

Problems:
- Overweight of larger errors
- Not scale invariant

4. Symmetric Mean Absolute Percent Error:

$$\text{sMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{\left(|Y_t| + |\hat{Y}_t|\right)}$$

Problems:
- Divide by 0
- Still asymmetric

Comparison Across Diagnostics

	$Y_t = 1$, $\hat{Y}_t = 3$	$Y_t = 2$, $\hat{Y}_t = 3$	$Y_t = 3$, $\hat{Y}_t = 3$	$Y_t = 4$, $\hat{Y}_t = 3$	$Y_t = 15$, $\hat{Y}_t = 3$	MEAN
APE	200%	50%	0%	25%	80%	71%
AE	2	1	0	1	12	3.2
SE	4	1	0	1	144	30
Sym. APE	50%	20%	0%	14.3%	66.7%	30.2%

Comparison Across Diagnostics

	$Y_t = 0$, $\widehat{Y}_t = 3$	$Y_t = 2$, $\widehat{Y}_t = 3$	$Y_t = 3$, $\widehat{Y}_t = 3$	$Y_t = 4$, $\widehat{Y}_t = 3$	$Y_t = 15$, $\widehat{Y}_t = 3$	MEAN
APE	∞	50%	0%	25%	80%	…?
AE	3	1	0	1	12	3.4
SE	9	1	0	1	144	31
Sym. APE	100%	20%	0%	14.3%	66.7%	40.2%

Model Diagnostic Statistics

5. Akaike's Information Criterion

$$\text{AIC} = -2\log(L) + 2k$$

$$\text{AIC} = n\log\left(\frac{SSE}{n}\right) + 2k$$

Likelihood Based

6. Schwarz's Bayesian Information Criterion:

$$\text{SBC} = -2\log(L) + k\log(n)$$

$$\text{SBC} = n\log\left(\frac{SSE}{n}\right) + k\log(n)$$

Model Diagnostic Statistics

5. Akaike's Information Criterion

$$\mathrm{AIC} = -2\log(L) + 2k$$

$$\mathrm{AIC} = n\log\left(\frac{SSE}{n}\right) + 2k$$

Error Based

6. Schwarz's Bayesian Information Criterion:

$$\mathrm{SBC} = -2\log(L) + k\log(n)$$

$$\mathrm{SBC} = n\log\left(\frac{SSE}{n}\right) + k\log(n)$$

R output

```
training=subset(Passenger,end=length(Passenger)-12)
test=subset(Passenger,start=length(Passenger)-11)
HWES.USAir.train <- hw(training, seasonal =
"multiplicative",initial='optimal')
test.results=forecast(HWES.USAir.train,h=12)

error=test-test.results$mean
MAE=mean(abs(error))
MAPE=mean(abs(error)/abs(test))
MAE
[1] 1134.58
MAPE
[1] 0.01763593
```

ETS

ETS (Error, Trend, Season)

- ETS is an automated search procedure that will try to identify the “best” model based on treating the data as a state space problem (think back to the “components” of ESM’s)
  - For “Error”, the choices are Additive (A) or Multiplicative (M)
  - For “Trend”, the choices are None (N), Additive (A), Multiplicative (M)….if you wanted “Damped”, you can specify that damped=TRUE, which indicates the trend
  - For “Seasonal”, the choices are None (N), Additive (A), Multiplicative (M)
- You can choose which one you want, OR you can let the computer choose (either specify “Z” for each EST feature or leave them blank)

ETS function

```
ets.passenger<-ets(training)
summary(ets.passenger)
ets.forecast.passenger<-forecast(ets.passenger,h=12)
error=mean(abs(test-test.results$mean))
error
```

R output (edited)

```
ETS(M,Ad,M)

Call:
 ets(y = training)

  Smoothing parameters:
    alpha = 0.6485
    beta  = 1e-04
    gamma = 1e-04
    phi   = 0.9755

 AIC     AICc      BIC
4225.929 4229.567 4285.918

Training set error measures:
                    ME      RMSE      MAE       MPE     MAPE
Training set 157.1888 1747.817 1043.596 0.1980283 2.19083
```