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.

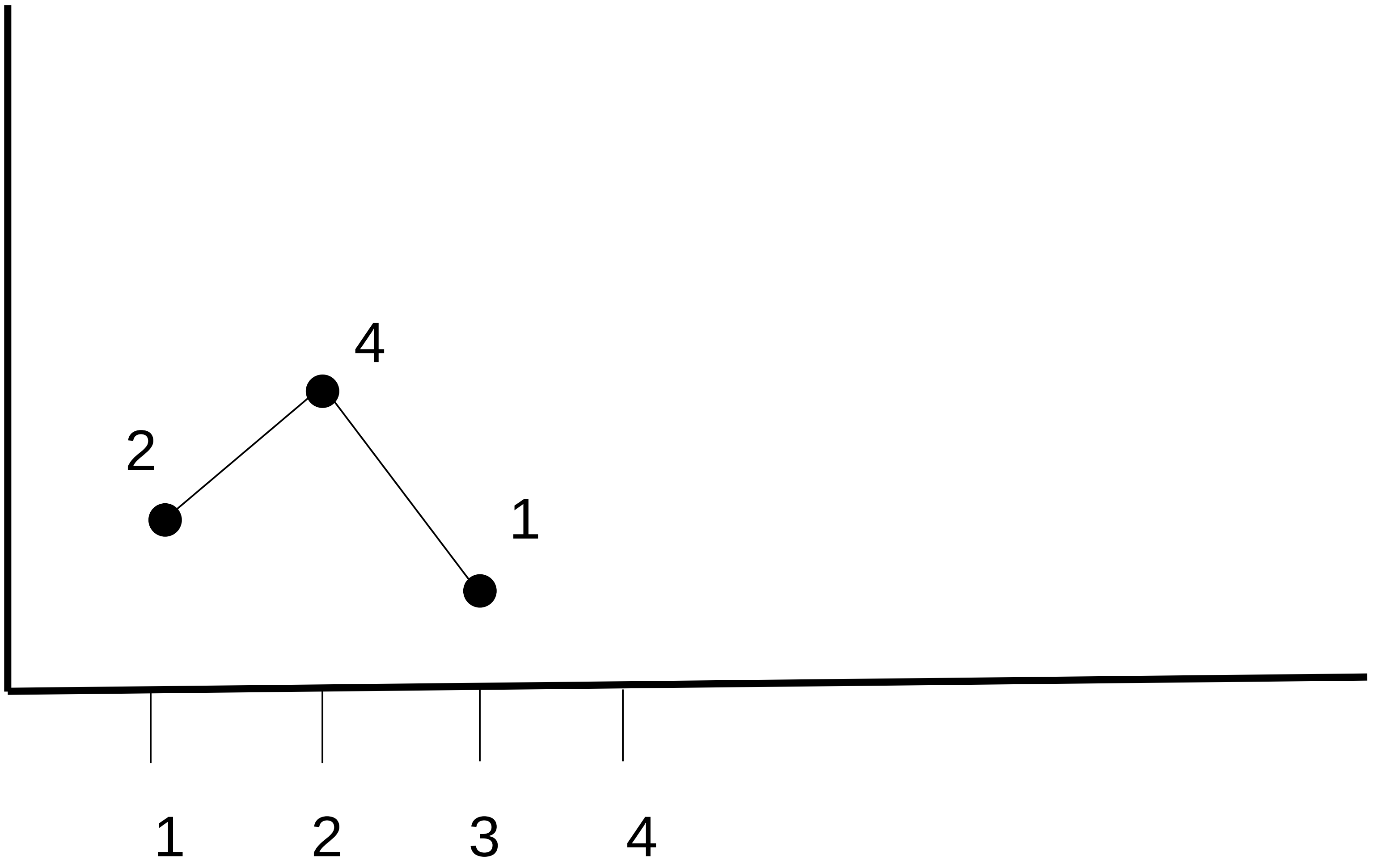
Naïve Model:

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

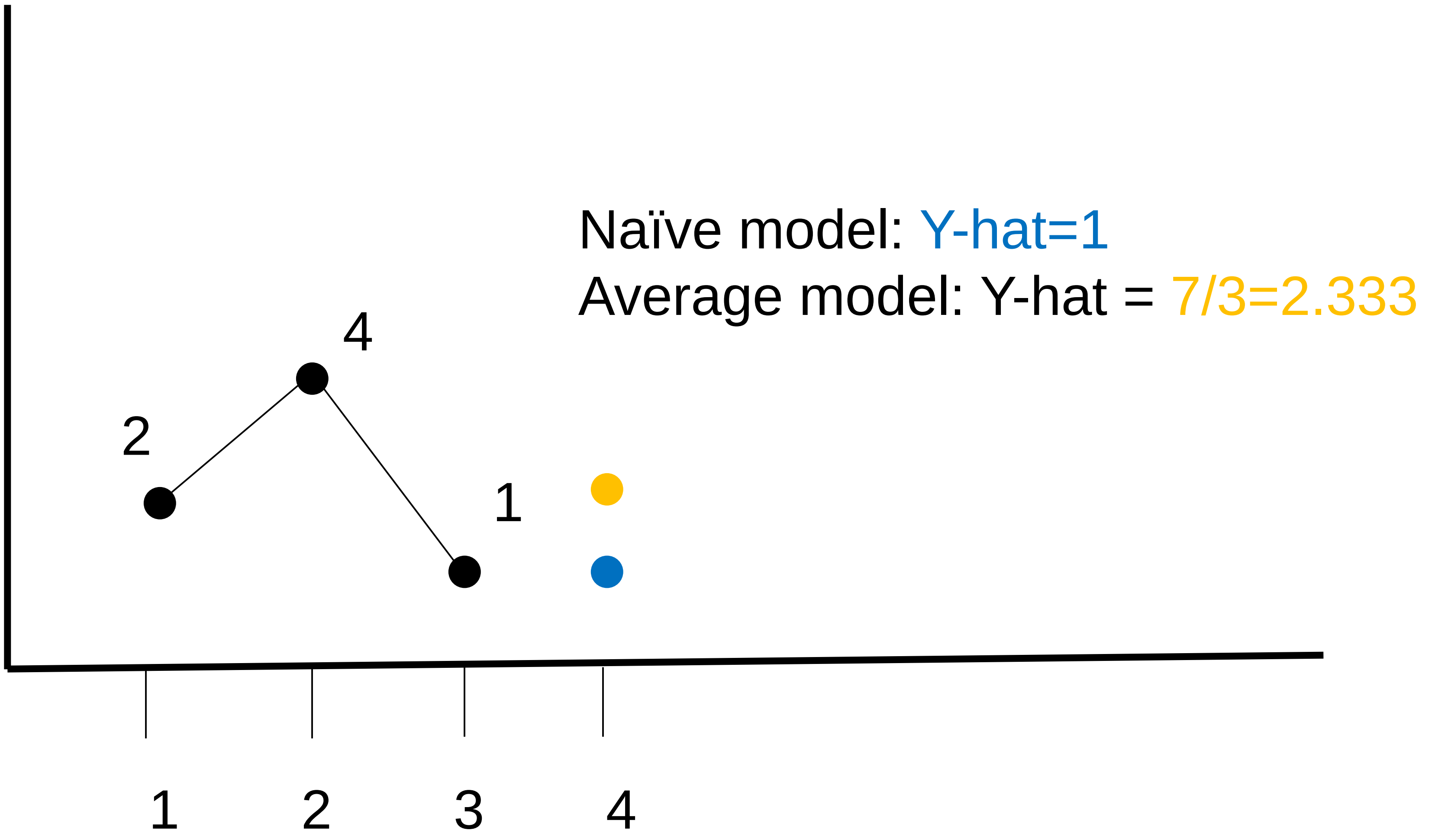
Average Model:

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

Naïve model versus Average model



Naïve model versus Average model



Exponential Smoothing

- This is what exponential smoothing does (however, it is a WEIGHTED average, not a simple average)
- Forecasting should only require a few parameters.
- Forecast equations should be 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)
  - Seasonal
  - Holt-Winters (incorporates trend and seasonality)
- ESM are great for “one-step ahead” forecasting

SINGLE EXPONENTIAL SMOOTHING

Single Exponential Smoothing

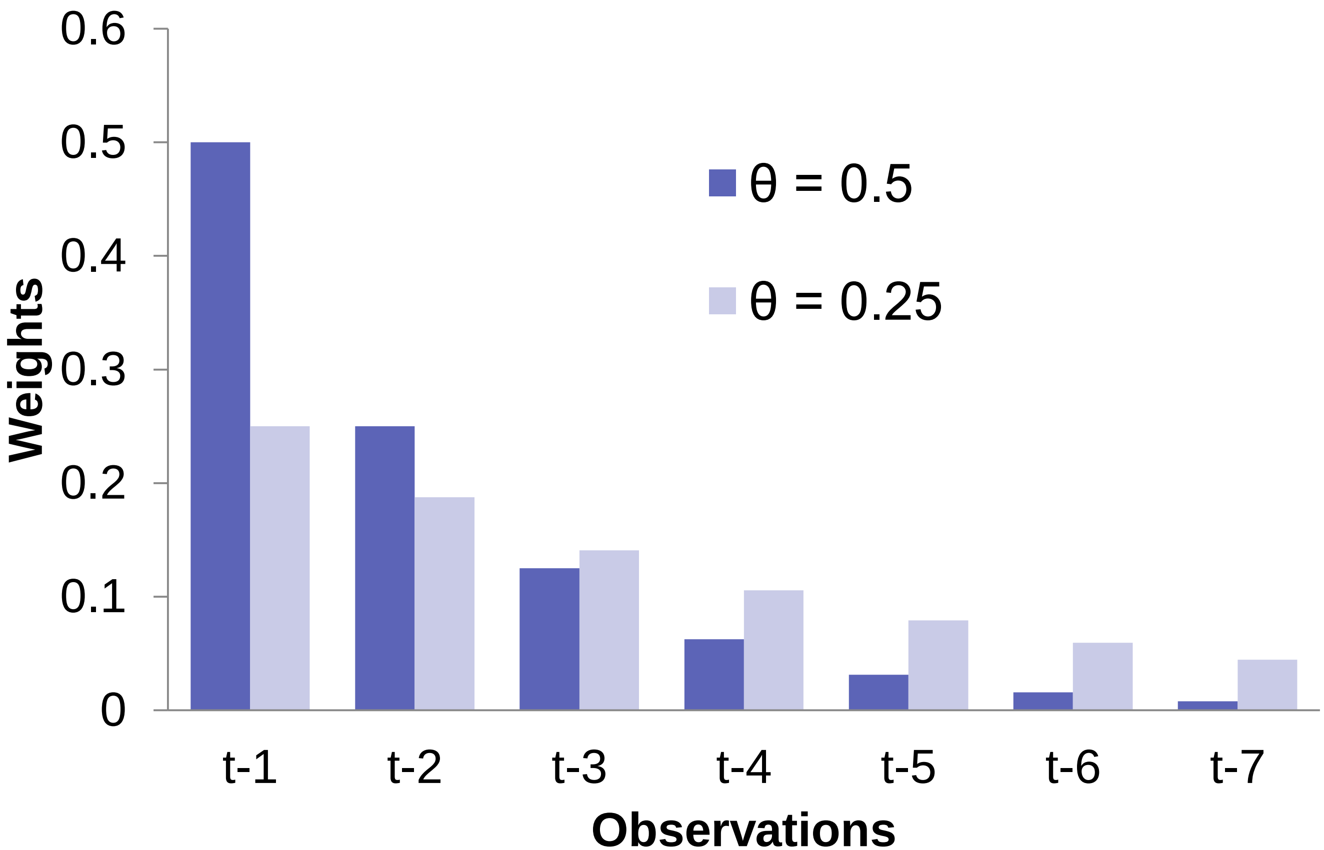
- We can apply a weighting scheme that decreases exponentially the further back in time we go.

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

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

Single Exponential Smoothing

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



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.

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

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

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

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

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

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, you should not look at significance tests UNLESS you validate normality!!).
- Models were originally derived without statistical distribution consideration (estimates are fine even without normality!).

ESM Procedure – SAS

```
proc esm data=Time.Steel print=all plot=all lead=24;
    forecast steelshp / model=simple;
run;
```

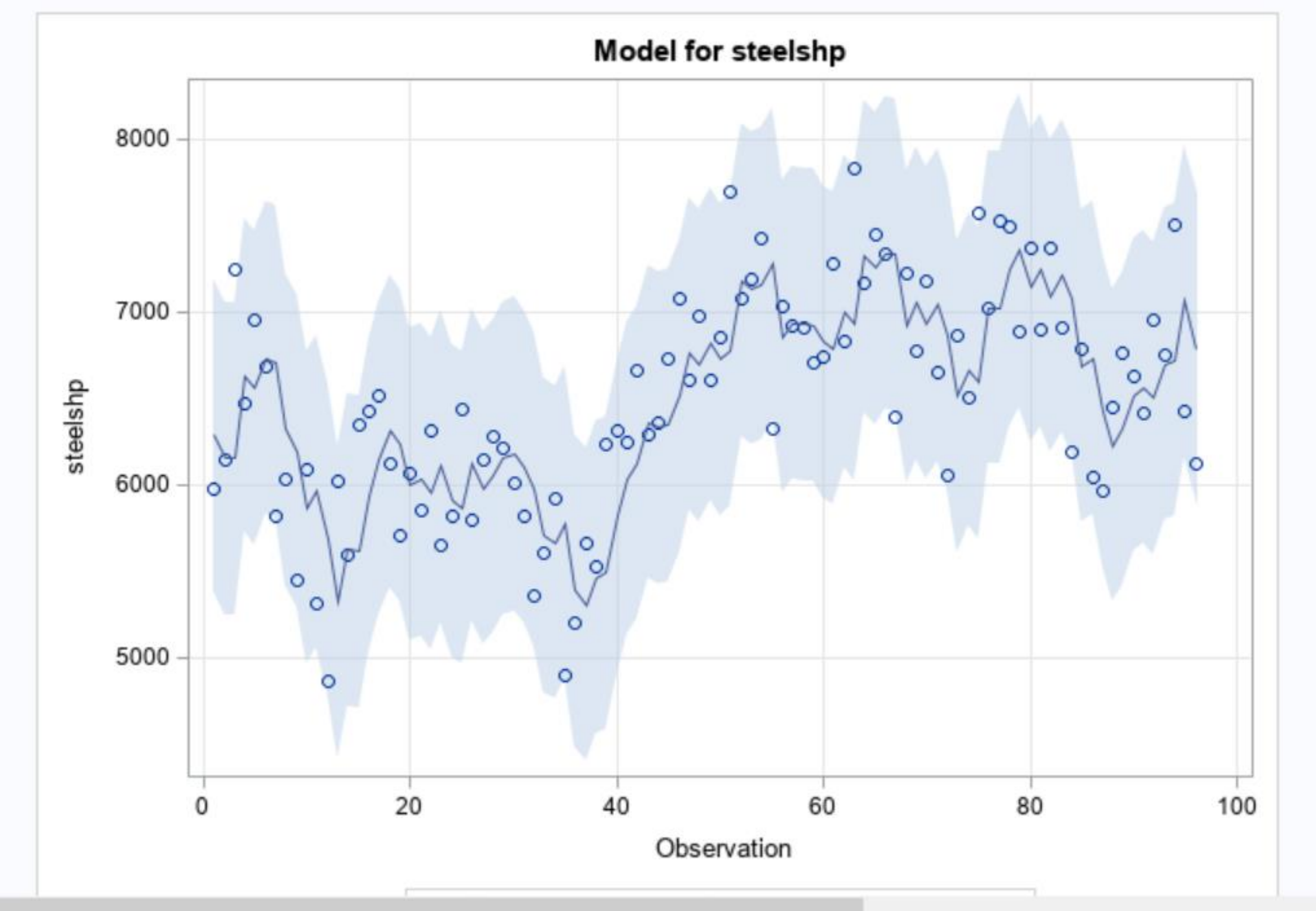
Output from SAS

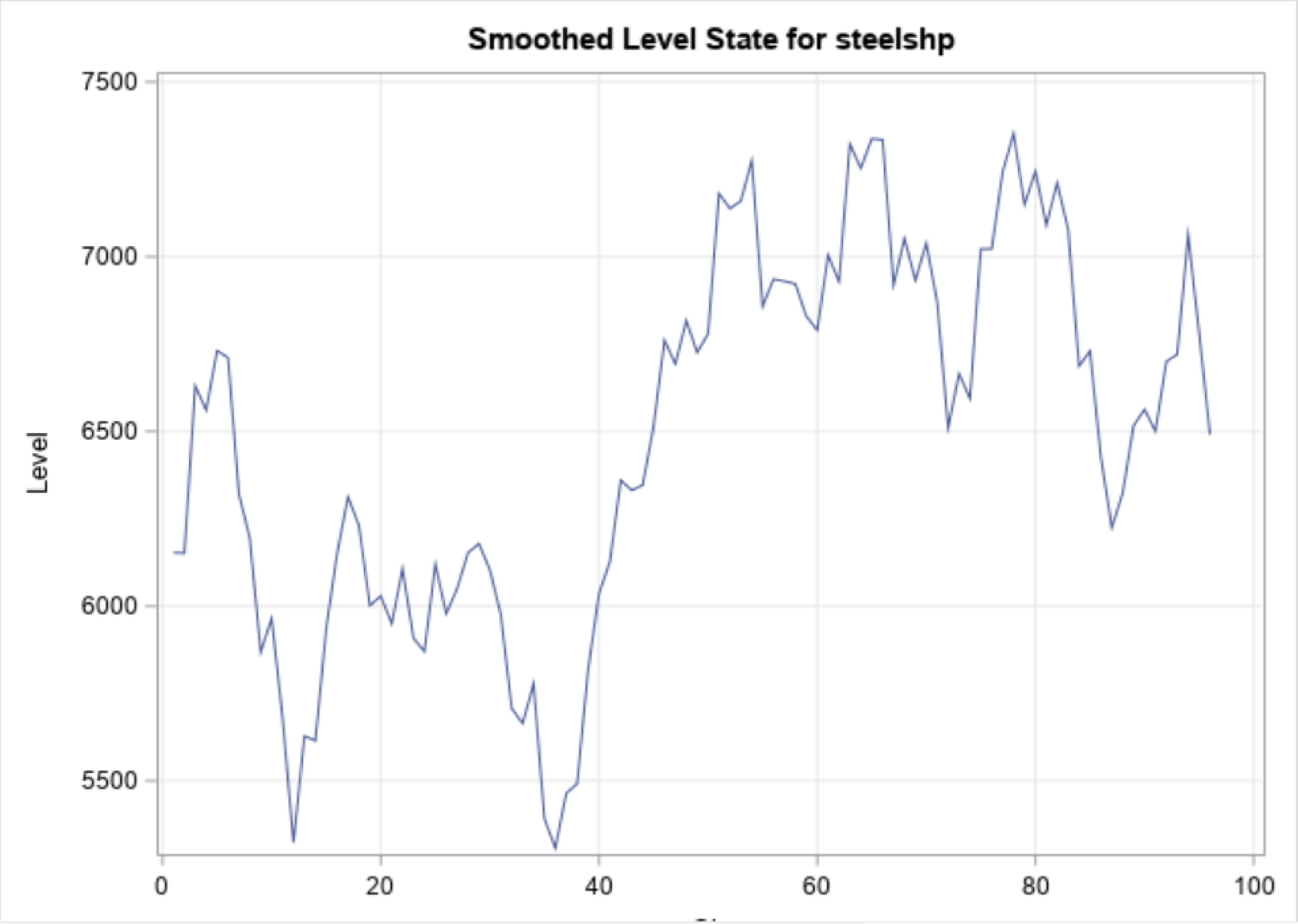
Number of Observations Read	96

Descriptive Statistics	
Variable	steelshp
Number of Observations	96
Number of Missing Observations	0
Minimum	4867
Maximum	7824
Mean	6488.135
Standard Deviation	640.8745

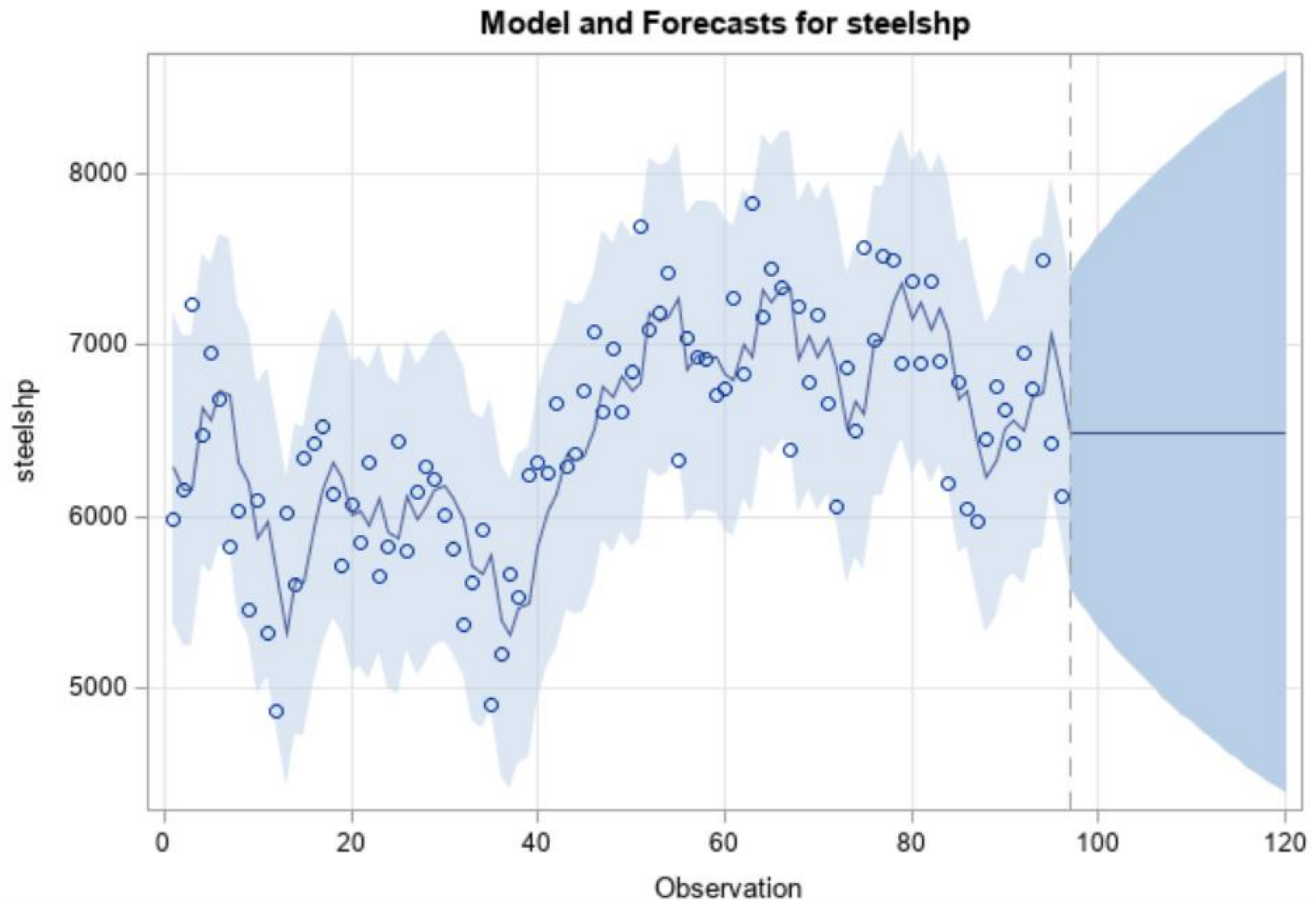
Simple Exponential Smoothing Parameter Estimates

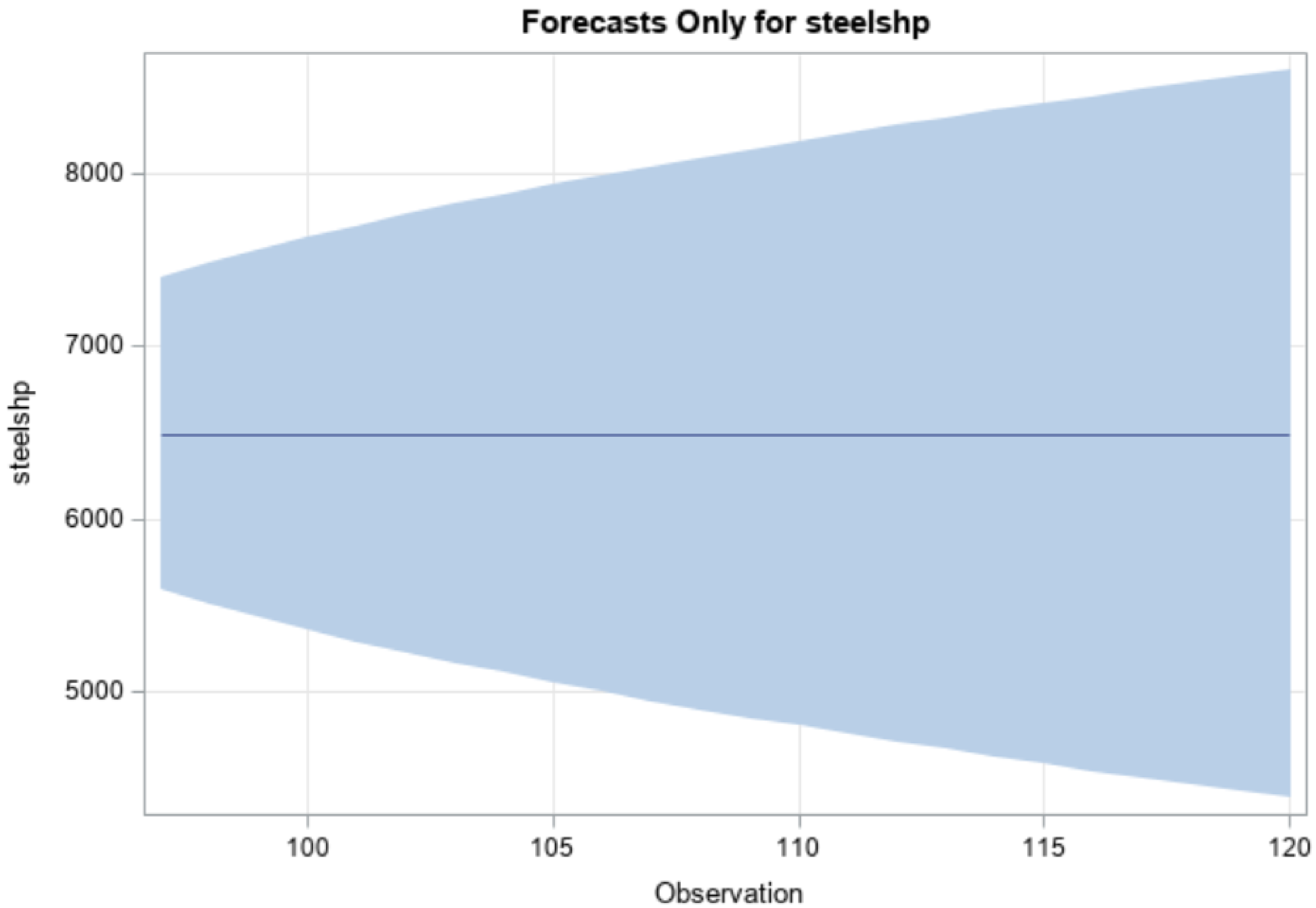
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.43935	0.06037	7.28	<.0001


Model for steelshp
8000
7000
6000
5000
steelshp
0
20
40
60
80
100
Observation



Smoothed Level State for steelshp
Level
7500
7000
6500
6000
5500
0
20
40
60
80
100


Model and Forecasts for steelshp
steelshp
8000
7000
6000
5000
0
20
40
60
80
100
120
Observation


Forecasts Only for steelshp
8000
7000
6000
5000
steelshp
100
105
110
115
120
Observation

Forecasts for Variable steelshp				
Obs	Forecasts	Standard Error	95% Confidence Limits	
97	6490.9911	461.3453	5586.7710	7395.2113
98	6490.9911	503.9085	5503.3485	7478.6337
99	6490.9911	543.1465	5426.4435	7555.5388
100	6490.9911	579.7348	5354.7317	7627.2505
101	6490.9911	614.1472	5287.2846	7694.6976
102	6490.9911	646.7311	5223.4214	7758.5609
103	6490.9911	677.7503	5162.6249	7819.3574
104	6490.9911	707.4107	5104.4917	7877.4906
105	6490.9911	735.8765	5048.6997	7933.2825
106	6490.9911	763.2814	4994.9870	7986.9952
107	6490.9911	789.7360	4943.1371	8038.8452
108	6490.9911	815.3326	4892.9686	8089.0136
109	6490.9911	840.1497	4844.3279	8137.6543
110	6490.9911	864.2545	4797.0833	8184.8989
111	6490.9911	887.7051	4751.1212	8230.8611

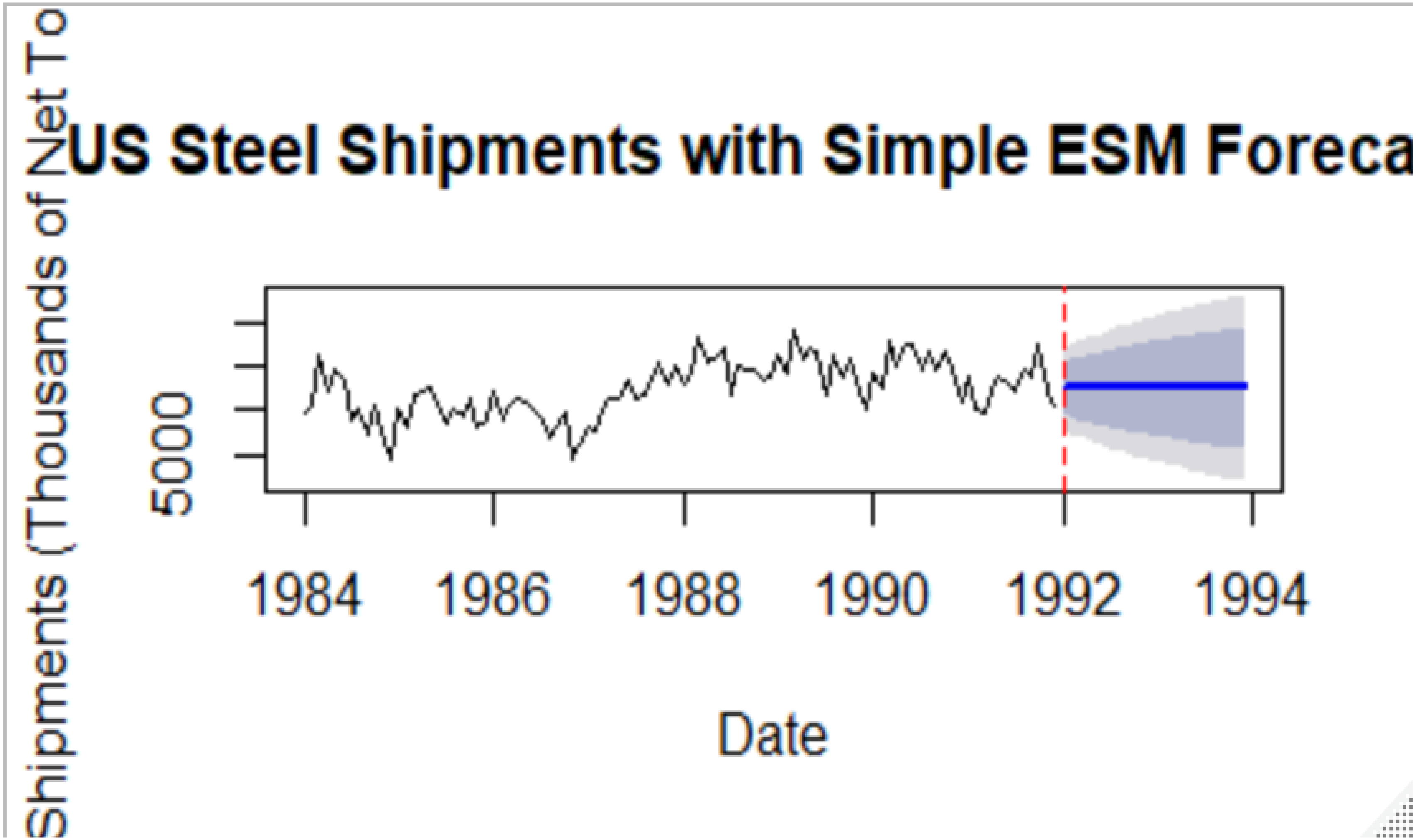
SES Function – R

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

Output from R

```
Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = SteelShp, h = 24, initial = "optimal")

  Smoothing parameters:
    alpha = 0.4393

     AIC     AICc      BIC
1620.928 1621.189 1628.621
```

```
plot(SES.Steel, main = "US Steel Shipments with Simple ESM Forecast", xlab = "Date", ylab = "Shipments (Thousands of Net Tons)")
abline(v = 1992, col = "red", lty = "dashed")
```



LINEAR TREND FOR EXPONENTIAL SMOOTHING

Trending Exponential Smoothing

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

Trending Exponential Smoothing

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

Double / Brown Exponential Smoothing

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

$$\hat{Y}_{t+k} = L_t + kT_t$$

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

Double / Brown Exponential Smoothing

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

$$\hat{Y}_{t+k} = L_t + kT_t$$

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

Double / Brown Exponential Smoothing

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

$$\hat{Y}_{t+k} = L_t + k T_t$$

Number of periods to be forecasted.

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

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

Double / Brown Exponential Smoothing

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

$$\hat{Y}_{t+k} = L_t + kT_t$$

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

Estimate of the slope of the series at time t.

Double / Brown Exponential Smoothing

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

$$\hat{Y}_{t+k} = L_t + kT_t$$

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

Same in both equations!

Double / Brown Exponential Smoothing

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

$$\hat{Y}_{t+k} = L_t + kT_t$$

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

Same in both equations! PROBABLY UNREALISTIC

Trending Exponential Smoothing

- The Single Exponential Smoothing model are better used for short-term forecasts.
- The model can not adequately handle data that is trending up or down.
- There are multiple ways to incorporate a trend in the Exponential Smoothing Model.
  - Double / Brown Exponential Smoothing
  - 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.

$$\hat{Y}_{t+k} = L_t + kT_t$$

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

Linear / Holt Exponential Smoothing

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

$$\hat{Y}_{t+k} = L_t + kT_t$$

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

Different between two equations!

Trending Exponential Smoothing

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

Damped Trend Exponential Smoothing

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

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

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

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

Damped Trend Exponential Smoothing

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

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

Between 0 and 1

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

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

ESM Procedure – SAS

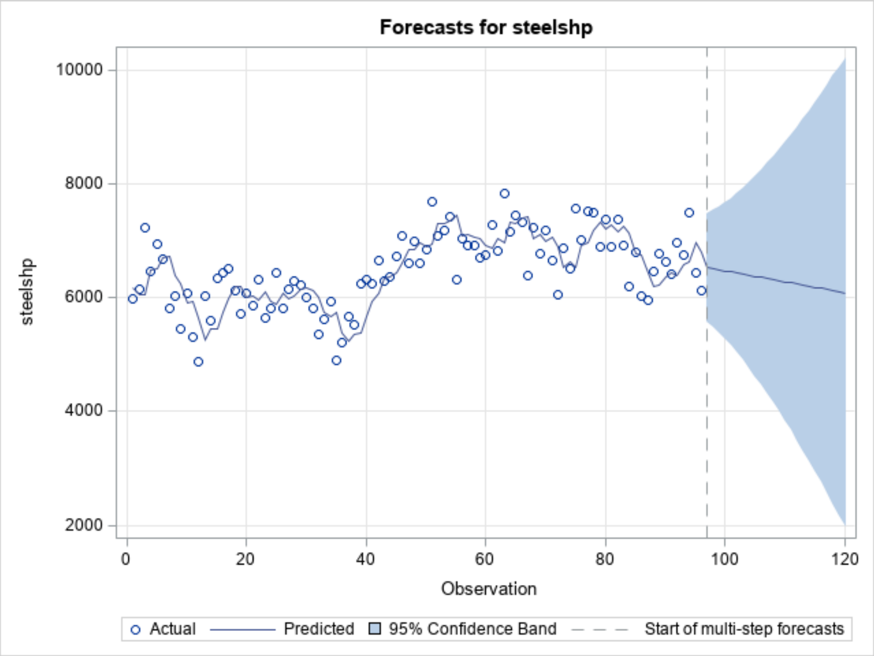
```
proc esm data=Time.Steel print=all plot=all lead=24;
     forecast steelshp / model=double;
run;

proc esm data=Time.Steel print=all plot=all lead=24;
     forecast steelshp / model=linear;
run;

proc esm data=Time.Steel print=all plot=all lead=24;
     forecast steelshp / model=damptrend;
run;
```

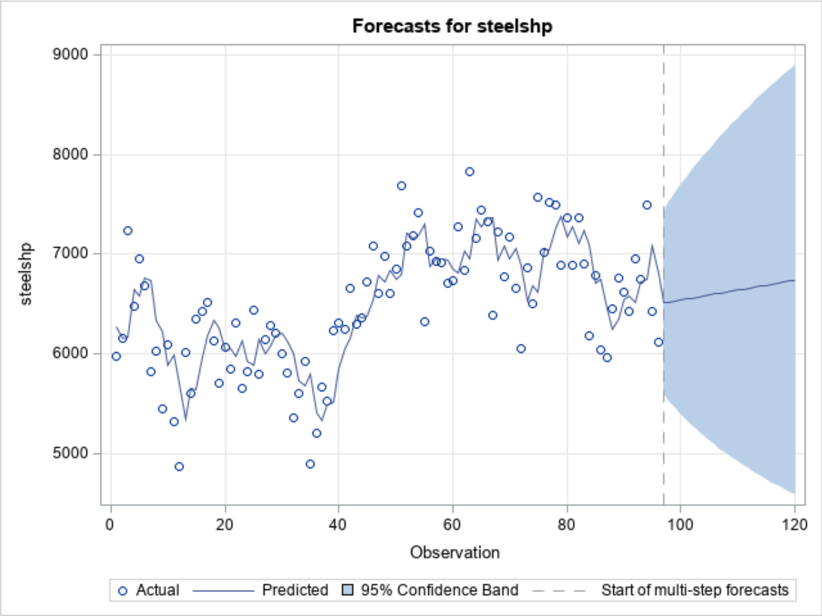
Parameter Estimates

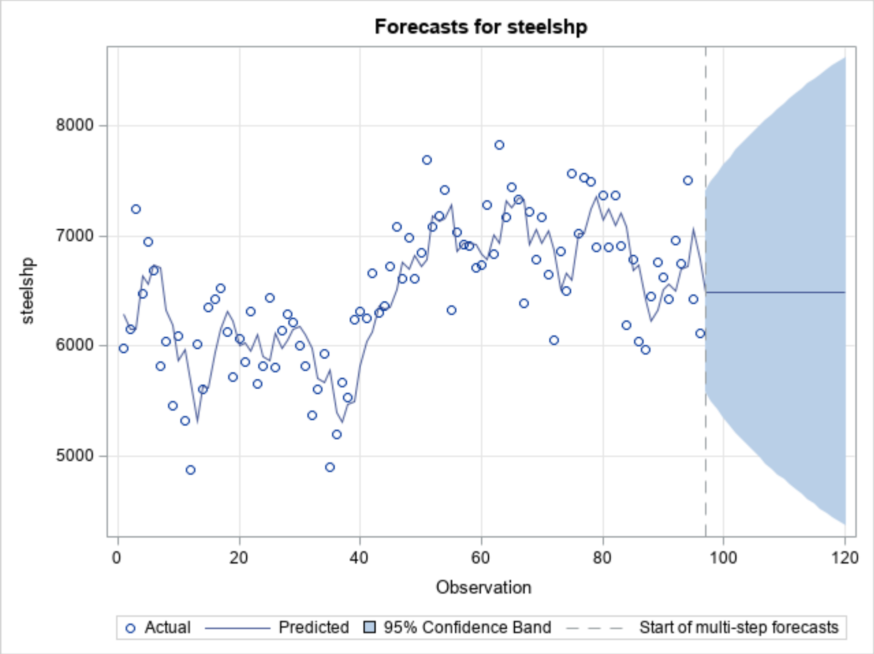
Double Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level/Trend Weight	0.19993	0.02173	9.20	<.0001

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.44202	0.06110	7.23	<.0001
Trend Weight	0.0010000	0.01091	0.09	0.9272

Damped-Trend Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.43874	36.18159	0.01	0.9904
Trend Weight	0.99900	0.0003538	2823.43	<.0001
Damping Weight	0.0010000	82.54435	0.00	1.0000


Forecasts for steelshp
steelshp
10000
8000
6000
4000
2000
0
20
40
60
80
100
120
Observation
Actual
Predicted
95% Confidence Band
Start of multi-step forecasts

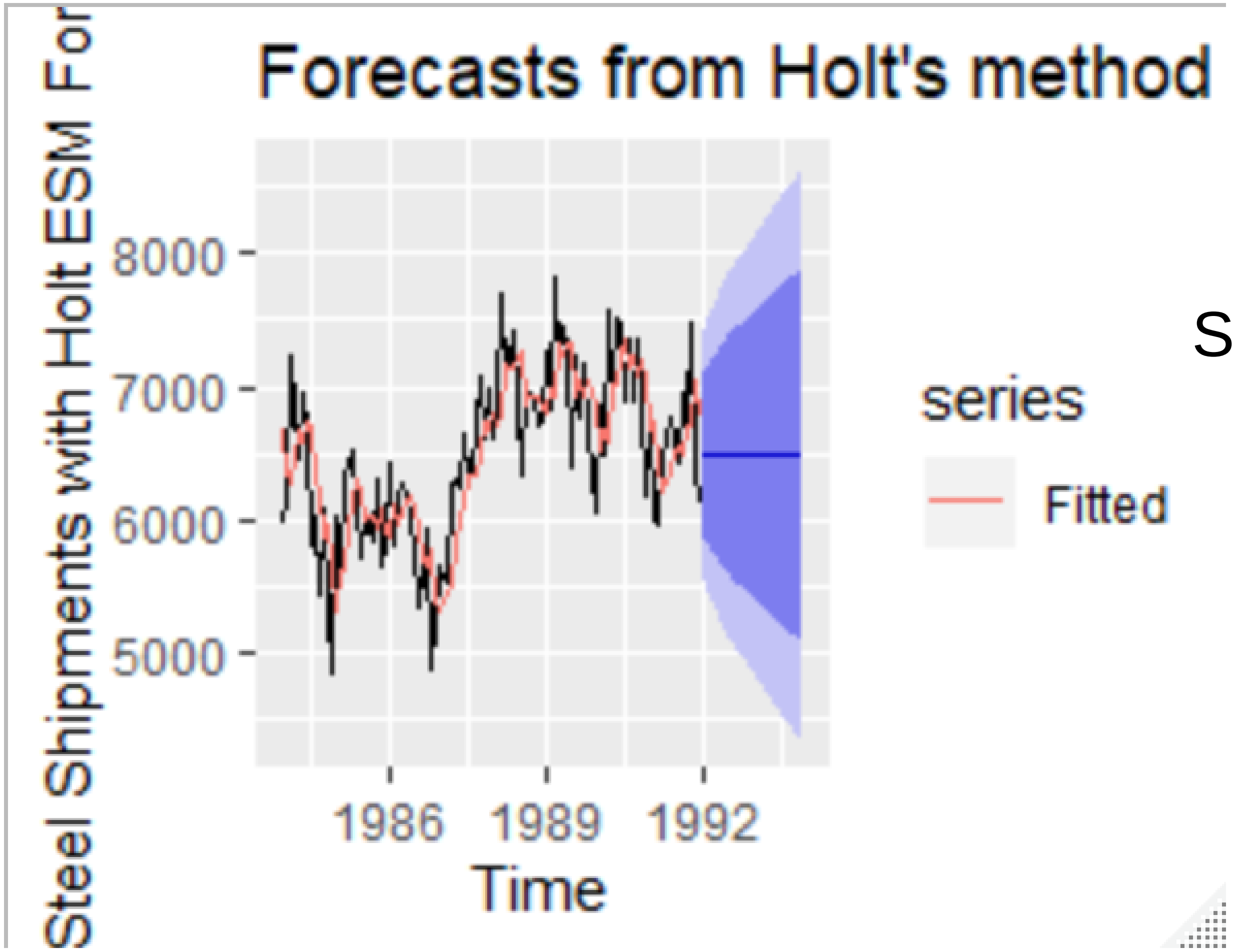

Forecasts for steelshp
steelshp
9000
8000
7000
6000
5000
0
20
40
60
80
100
120
Observation
Actual
Predicted
95% Confidence Band
Start of multi-step forecasts


Forecasts for steelshp
steelshp
8000
7000
6000
5000
0
20
40
60
80
100
120
Observation
Actual
Predicted
95% Confidence Band
Start of multi-step forecasts

HOLT Function – R

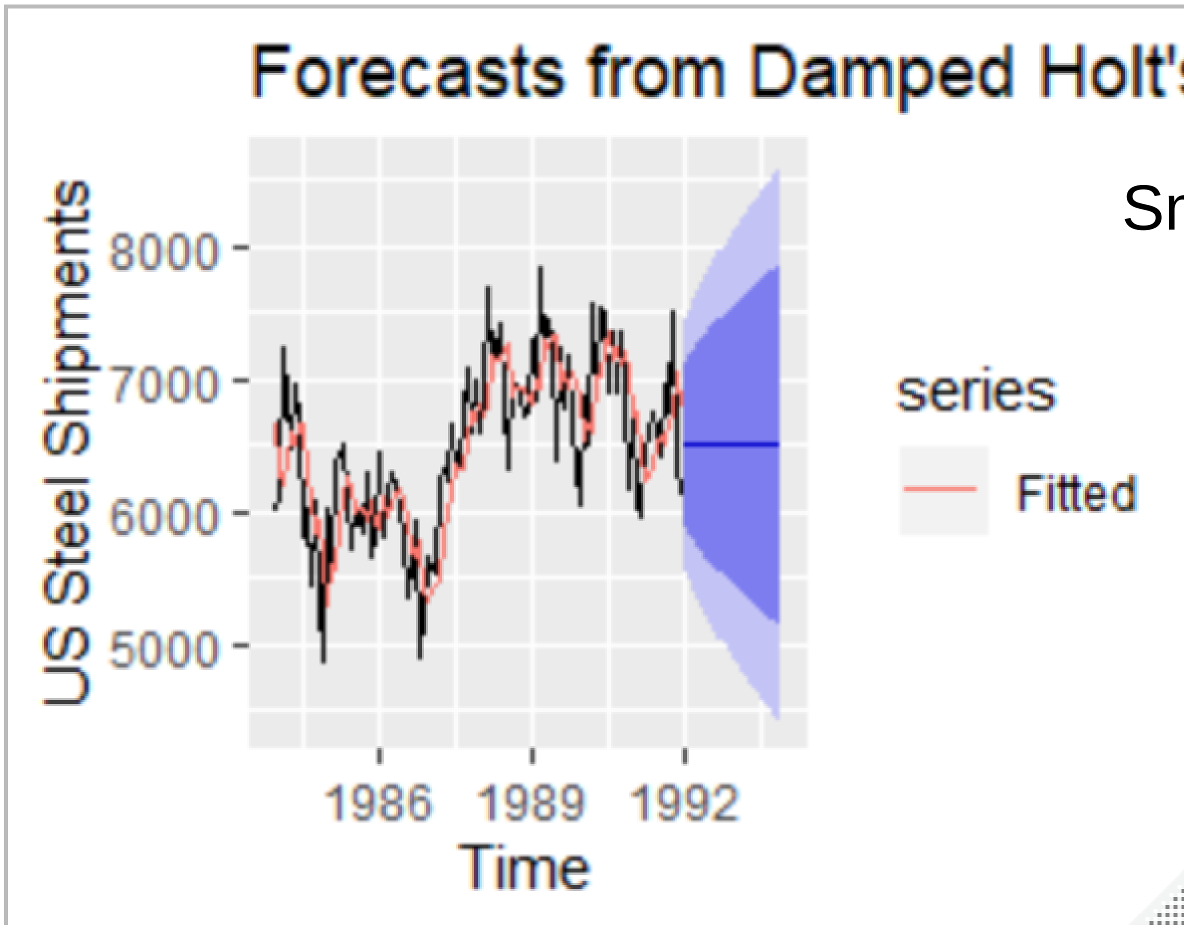
```
LES.Steel <- holt(SteelShp, initial="optimal",
                  h=24)
summary(LES.Steel)

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

Linear ESM



Smoothing parameters
alpha = 0.4329
beta = 1e-04

Damped Trend



Smoothing parameters:

alpha = 0.4202

beta = 1e-04

phi = 0.9083

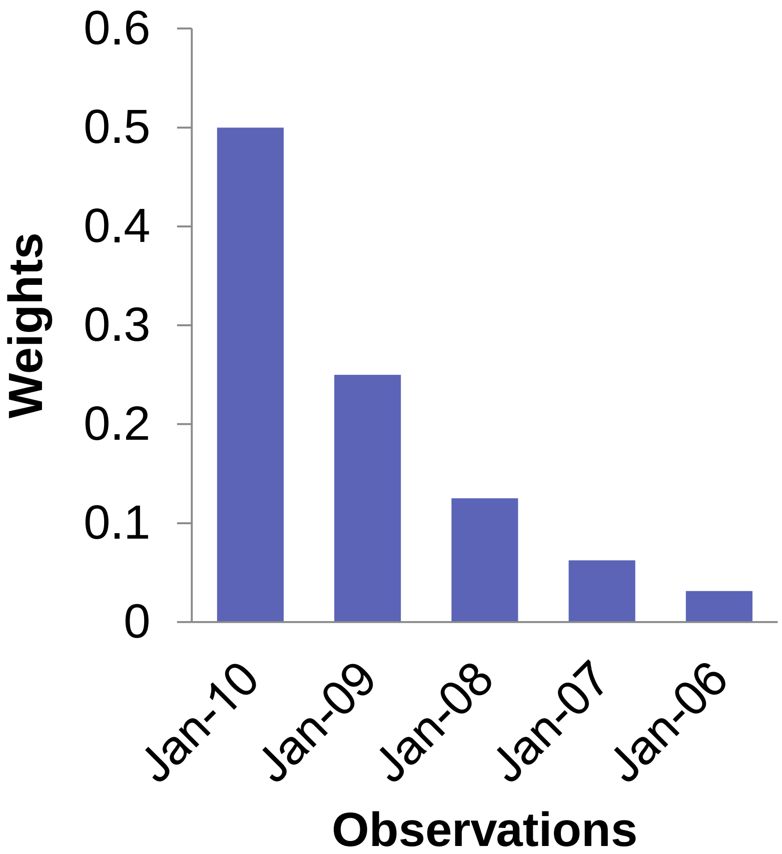
SEASONAL EXPONENTIAL SMOOTHING

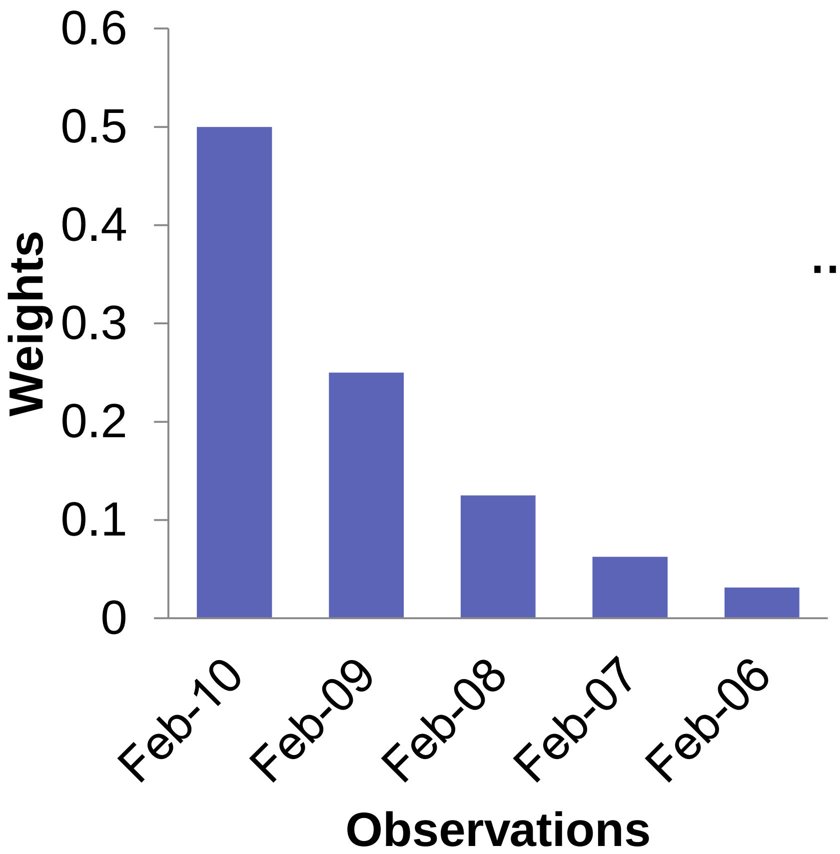
Seasonal Exponential Smoothing

- Exponential Smoothing models can also be adapted to account for seasonal factors.
- There are multiple ways to incorporate a seasonal effect in the Exponential Smoothing Model.
  - Additive Seasonal Exponential Smoothing
  - Multiplicative Seasonal Exponential Smoothing
  - Winters Additive Exponential Smoothing (includes trend)
  - Winters Multiplicative Exponential Smoothing (includes trend)

Seasonal Exponential Smoothing

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





...

Seasonal Exponential Smoothing (Additive)

$$Y_t = \mu_t + s_p(t) + \varepsilon_t$$

Seasonal Exponential Smoothing (Additive)

- The Seasonal Exponential Smoothing model has two components as well.
- The second component incorporates seasonality into the model.

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

$$L_t = \theta\left(Y_t - S_{t-p}\right) + (1-\theta)L_{t-1}$$

$$S_t = \delta(Y_t - L_{t-1}) + (1-\delta)S_{t-p}$$

Seasonal Exponential Smoothing (Multiplicative)

$$Y_t = (\mu_t)s_p(t) + \varepsilon_t$$

Seasonal Exponential Smoothing (Multiplicative)

- The Seasonal Exponential Smoothing model has two components as well.
- The second component incorporates seasonality into the model.

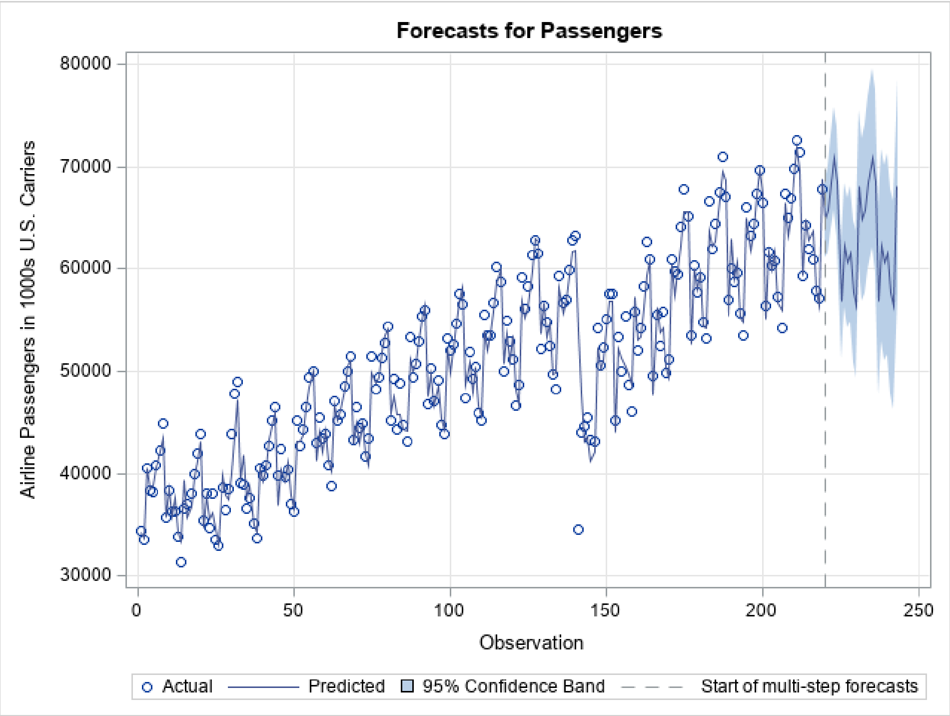
$$\hat{Y}_{t+k} = (L_t)(S_{t-p+k})$$

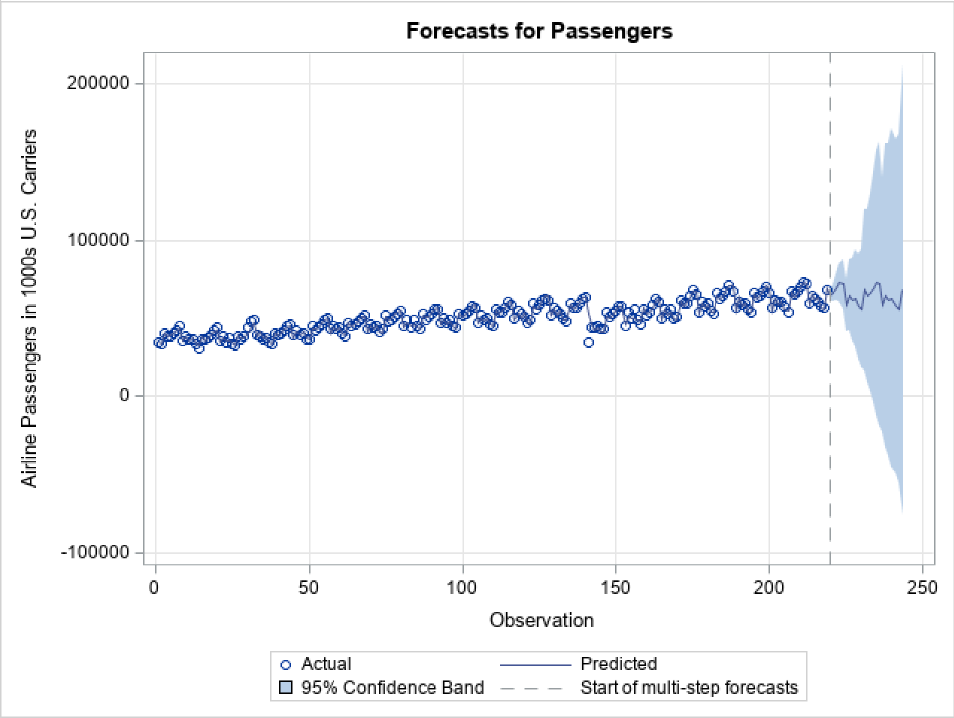
$$L_t = \theta\left(Y_t / S_{t-p}\right) + (1-\theta)L_{t-1}$$
$$S_t = \delta(Y_t / L_{t-1}) + (1-\delta)S_{t-p}$$

ESM Procedure – SAS

```
proc esm data=Time.USAirlines print=all plot=all
                    seasonality=12 lead=24 outfor=test1;
          forecast Passengers / model=addseasonal;
run;

proc esm data=Time.USAirlines print=all plot=all
                    seasonality=12 lead=24;
          forecast Passengers / model=multseasonal;
run;
```

Additive versus Multiplicative Seasonal





Parameter Estimates

Multiplicative Seasonal Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.54523	0.03551	15.36	<.0001
Seasonal Weight	0.25876	0.04102	6.31	<.0001

Forecast

Forecasts for Variable Passengers				
Obs	Forecasts	Standard Error	95% Confidence Limits	
220	64724.5688	1832.8834	61132.1834	68316.9542
221	65970.8088	2740.3577	60599.8064	71341.8111
222	69380.7232	4255.0973	61040.8857	77720.5607
223	72579.3078	6221.1804	60386.0182	84772.5974
224	71557.1022	8131.4381	55619.7764	87494.4280
225	58497.1721	8498.5514	41840.3174	75154.0269
226	64610.5628	11537.999	41996.5005	87224.6250
227	61843.8594	13306.143	35764.2983	87923.4204
228	62608.7949	15911.452	31422.9228	93794.6671
229	57726.4983	17077.420	24255.3693	91197.6273
230	56098.2433	19066.249	18729.0824	93467.4042
231	68201.0895	26316.044	16622.5919	119779.5872
232	64724.5688	28214.654	9424.8623	120024.2753
233	65970.8088	32066.836	3120.9656	128820.6520
234	69380.7232	37325.960	-3776.8148	142538.2613
235	72579.3078	42941.058	-11583.6200	156742.2357
236	71557.1022	46299.297	-19187.8519	162302.0563
237	58497.1721	41208.974	-22270.9333	139265.2776
238	64610.5628	49269.933	-31956.7312	161177.8567
239	61843.8594	50872.814	-37865.0239	161552.7426
240	62608.7949	55345.784	-45866.9482	171084.5380
241	57726.4983	54673.422	-49431.4391	164884.4357
242	56098.2433	56750.358	-55130.4145	167326.9010
243	68201.0895	73465.066	-75787.7930	212189.9721

Winters Exponential Smoothing (Additive)

$$Y_t = \mu_t + \beta_t t + s_p(t) + \varepsilon_t$$

Winters / Triple Exponential Smoothing (Additive)

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

$$\hat{Y}_{t+k} = L_t + kT_t + S_{t-p+k}$$

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

Winters Exponential Smoothing (Multiplicative)

$$Y_t = (\mu_t + \beta_t t) s_p(t) + \varepsilon_t$$

Winters / Triple Exponential Smoothing (Additive)

- The Linear Exponential Smoothing model has three components.

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

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

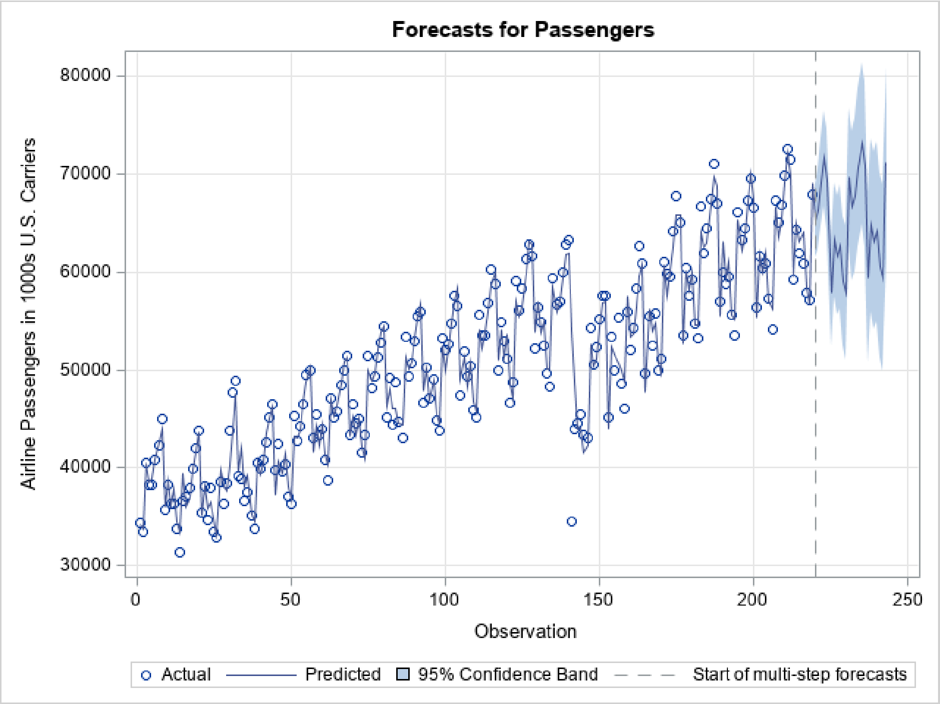
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

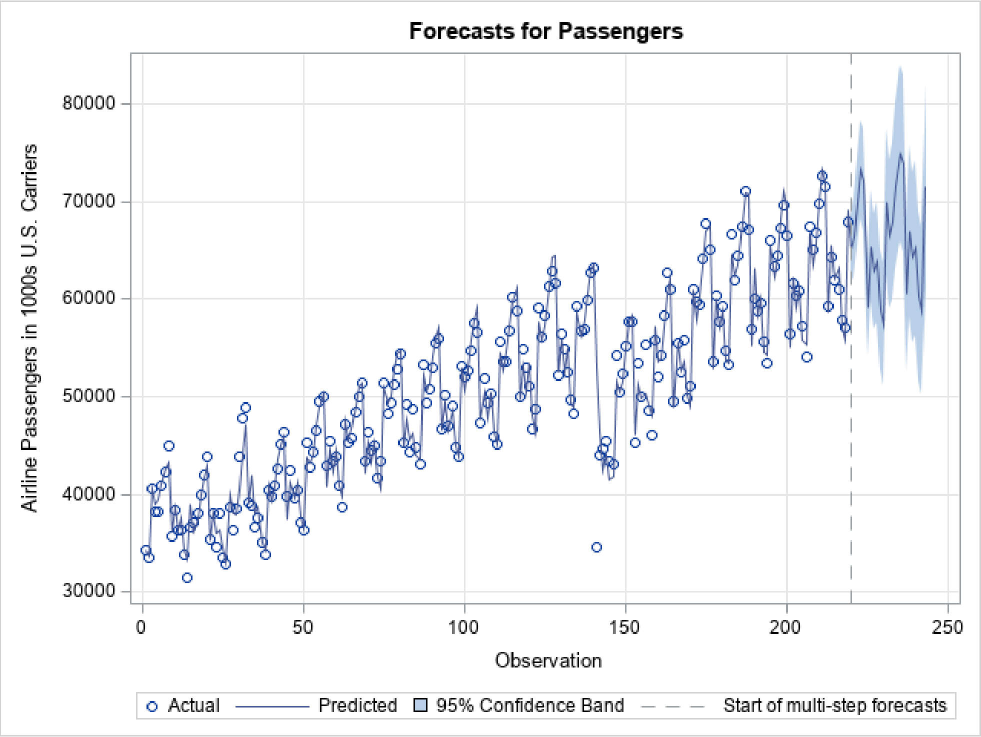
$$S_t = \delta(Y_t/L_{t-1}) + (1-\delta)S_{t-p}$$

ESM Procedure – SAS

```
proc esm data=Time.USAirlines print=all plot=all
                    seasonality=12 lead=24;
          forecast Passengers / model=addwinters;
run;

proc esm data=Time.USAirlines print=all plot=all
                    seasonality=12 lead=24;
          forecast Passengers / model=multwinters;
run;
```

Plot of data and forecasts





Parameter Estimates

Winters Method (Additive) Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.48692	0.03466	14.05	<.0001
Trend Weight	0.0010000	0.02010	0.05	0.9604
Seasonal Weight	0.45494	0.05805	7.84	<.0001

Winters Method (Multiplicative) Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.52088	0.03558	14.64	<.0001
Trend Weight	0.0010000	0.01860	0.05	0.9572
Seasonal Weight	0.25596	0.03962	6.46	<.0001

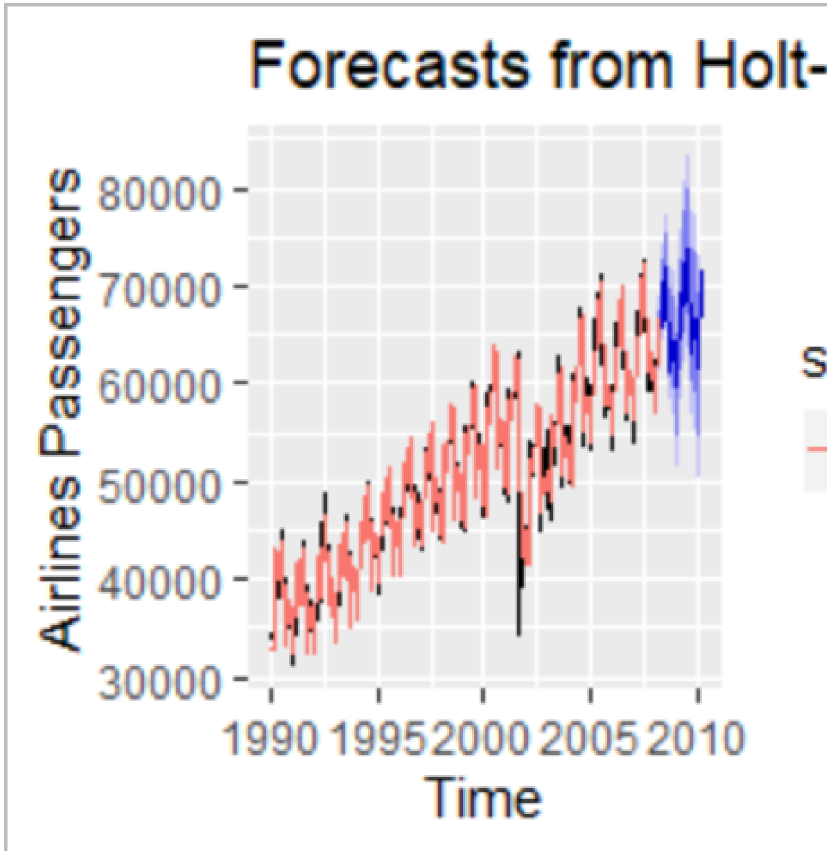
HW Function – R

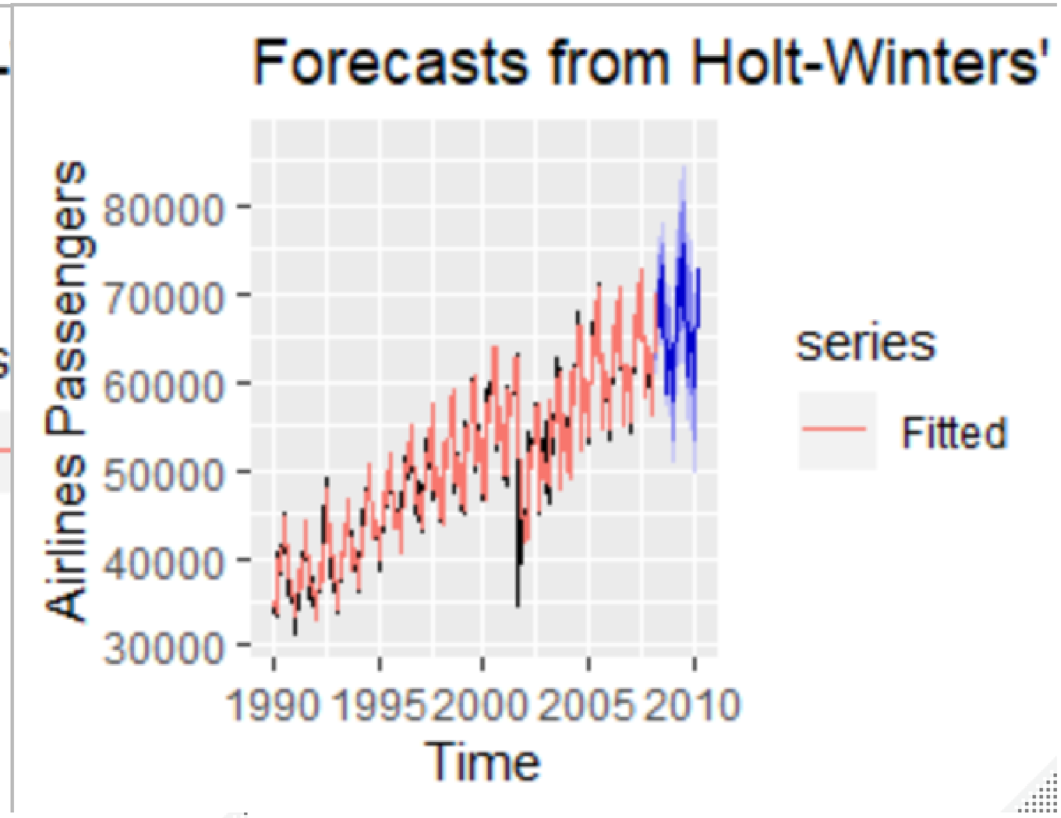
```
HWES.USAir <- hw(Passenger, seasonal="additive")
summary(HWES.USAir)

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
                  ACF1
Training set 0.06636172
```

```
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
                  ACF1
Training set 0.1713934
```

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.
- Occassionally 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.
- Ideally, an entire season should be captured in a hold-out sample.

Hold-out Sample

1. Divide the time series into two segments – training and validation (hold-out).
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.

Model Diagnostic Statistics

1. Mean Absolute Percent Error:

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

2. Mean Absolute Error:

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

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:

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

4. Symmetric Mean Absolute Percent Error:

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

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

	$Y_t = 1$, $\widehat{Y}_t = 3$	$Y_t = 2$, $\widehat{Y}_t = 3$	$Y_t = 3$, $\widehat{Y}_t = 3$	$Y_t = 4$, $\widehat{Y}_t = 3$	$Y_t = 15$, $\widehat{Y}_t = 3$	MEAN
APE	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

	$Y_t = 0$, $\hat{Y}_t = 3$	$Y_t = 2$, $\hat{Y}_t = 3$	$Y_t = 3$, $\hat{Y}_t = 3$	$Y_t = 4$, $\hat{Y}_t = 3$	$Y_t = 15$, $\hat{Y}_t = 3$	MEAN
APE	∞	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

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

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

Likelihood Based

6. Schwarz's Bayesian Information Criterion:

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

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

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:

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

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

Evaluating forecasts

```
proc esm data=Time.USAirlines print=all plot=all
                    seasonality=12 lead=12 back=12 outfor=test;
         forecast Passengers / model=multwinters;
run;

data test2;
set test;
if _TIMEID_>207;
abs_error=abs(error);
abs_err_obs=abs_error/abs(actual);
run;

proc means data=test2 mean;
var abs_error abs_err_obs;
run;
```

SAS Output

Variable	Mean
abs_error	1096.27
abs_err_obs	0.0169736

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
```