Statistics 3540
Assignment #4: Solutions

1. Use the unemployment data from Assignment #3 to answer the following:

   (a) Analyze the data by using the multiplicative decomposition method. Make sure to comment on any features you see in the plots of (i) the deseasonalized and (ii) detrended values. In particular, does it seem reasonable to assume a linear trend in the detrended values? Finally, tell how many observations are used to calculate each moving (or centered moving) average.

   The Minitab output is below:

```
Time Series Decomposition for Unemploy

Multiplicative Model


Data      Unemploy
Length    192
NMissing  0


Fitted Trend Equation

Yt = 11.0251 - 0.028308*t


Seasonal Indices

Period    Index
     1  1.08020
     2  1.07254
     3  1.07129
     4  1.04334
     5  1.01160
     6  0.93961
     7  1.02553
     8  1.01637
     9  0.91820
    10  0.91833
    11  0.95073
    12  0.95227
```

```
Accuracy Measures

MAPE  6.23247
MAD   0.50015
MSD   0.38541
```

(i) The plot of the deseaonlized values has a negative (downward) trend that may be linear.

(ii) The plot of the detrended values has the seasonal component left in the series; there is the peak in the summer of each year and the drop in the winter. There is also a dip in the plot around the midpoint of the series. This may be suggesting a pattern in the data that appears on a longer time scale, such as every few years.

(iii) In this time series we have an even number of observations, so we use 12-point centered moving averages.

Finally, the question asked to discuss if there was a linear trend in the detrended values. This doesn't make any sense, because detrended values better not show any trend! The detrended values should show the seasonal effect. My apologies for this.

(b) Repeat (a) using the additive decomposition method.

```
Time Series Decomposition for Unemploy

Additive Model


Data      Unemploy
Length    192
NMissing  0


Fitted Trend Equation

Yt = 11.0290 - 0.028306*t

Seasonal Indices

Period      Index
     1   0.634375
     2   0.571875
     3   0.526042
     4   0.330208
     5   0.088542
```

```
 6  -0.428125
 7   0.176042
 8   0.126042
 9  -0.628125
10  -0.611458
11  -0.390625
12  -0.394792

Accuracy Measures

MAPE  6.28883
MAD   0.50169
MSD   0.38957
```

(i) The deseasonalized values have a decreasing trend that may be linear.

(ii) As in (a), the detrended values retain the seasonal component of the series, plus a drop in the plot around the midpoint of the series.

(iii) As in (a), 12 observations are used in finding each centered moving average.

(c) Do the two methods appear to give similar results for this data?

The two methods give very similar results on the general features. The also give almost identical regression equations for modelling the trend of the deseasonalized values.

We also note that multiplicative decomposition gives a smaller value of MSD, so this approach may be more suitable for this data. However, the different is MSD values is pretty small.

You also see that the 2 methods give quite different estimates of the SN_t values. That is to be expected, since one uses an additive seasonal factor, and the other uses a multiplicative one.

2. Refer to the data on yearly U.S. lumber production in Table 6.6, p. 315. The data is on the course webpage in the file **lumber.mtw**.

For the purpose of this question, assume the series runs from 1946–1975.

(a) Plot the data vs. time. Is there any trend apparent?

There does not appear to be any trend in the time series. Therefore, simple (single) exponential smoothing will be appropriate.

(b) Analyze the data using simple exponential smoothing, setting (i) $\alpha = 0.5$, (ii) $\alpha = 0.2$, (iii) $\alpha = 0.1$, (iv) $\alpha = 0.05$.

The output is below.

```
Single Exponential Smoothing for Lumber

Data    Lumber
Length  30

Smoothing Constant

Alpha  0.5

Accuracy Measures

MAPE        5
MAD      1779
MSD   4802687

Smoothing Constant

Alpha  0.2

MAPE        5
MAD      1727
MSD   4350848

Smoothing Constant

Alpha  0.1

MAPE        5
MAD      1691
MSD   4305743

Smoothing Constant

Alpha  0.05

MAPE        5
MAD      1645
MSD   4346532
```

(c) *Using the most appropriate* α *from (b)* , calculate **by hand** a 90% prediction interval for lumber production in 1978. Make sure to explain your choice of α.

From (b), the most appropriate α value is $\alpha = 0.1$, since it gives the smallest value of $SSE = T(MSD) = 129172290$.

Find a 90% PI for y in 1978.

Since the time series ran until 1975, we want the PI when $\tau = 3$.

From the output below, we see that the forecast for 1978, given data up to (and including 1975) is

$$\hat{y}_{T+\tau}(T) = \hat{y}_{30+3}(30) = 35644.8$$

```
Single Exponential Smoothing for Lumber

Alpha  0.1

Forecasts

Period  Forecast    Lower    Upper
1976     35644.8  31502.2  39787.4
1977     35644.8  31502.2  39787.4
1978     35644.8  31502.2  39787.4
```

90% PI means $\alpha^* = 0.1$, so we need $z_{.1/2} = z_{0.05} = 1.64$ or 1.65.

Finally, we know from above that $s^2 = SSE/(T-1) = 129172290/29 = 4454217$.

Then the 90% PI is

$$\begin{aligned} \hat{y}_{30+3}(30) \pm z_{0.05} s\sqrt{1+(\tau-1)\alpha^2} &= 35644.8 \pm 1.64\sqrt{4454217}\sqrt{1+2(.1)^2} \\ &= 35644.8 \pm 3495.663 = (32149.14, 39140.46) \end{aligned}$$

Note that this PI is quite different from the one the Minitab found. Also note that the PI that Minitab found is the same for any value of τ, which is not the case with the PI formula we are using.

3. For the Dow Jones data used in previous assignments:

   (a) Use trend-corrected double exponential smoothing to model the data, using the following combinations of (α, γ):
   (0.03, 0.1), (0.05, 0.1) (0.1, 0.2), (0.3, 0.2), (0.9,0.1), (0.1, 0.9).
   The output is below.

```
Double Exponential Smoothing for DJIA

Data    DJIA
Length  25

Smoothing Constants

Alpha (level)  0.03
Gamma (trend)  0.10
```

```
MAPE      33
MAD      423
MSD   233994

Smoothing Constants

Alpha (level)  0.05
Gamma (trend)  0.10

MAPE      34
MAD      432
MSD   246505

Smoothing Constants

Alpha (level)  0.1
Gamma (trend)  0.2

MAPE      37
MAD      495
MSD   313149

Smoothing Constants

Alpha (level)  0.3
Gamma (trend)  0.2

MAPE      26
MAD      319
MSD   132575

Smoothing Constants

Alpha (level)  0.9
Gamma (trend)  0.1

MAPE     15.2
MAD     188.4
MSD   57630.7

Smoothing Constants

Alpha (level)  0.1
Gamma (trend)  0.9
```

```
MAPE       40
MAD       500
MSD    321148
```

(b) Which choice of (α, γ) from (a) is best for this dataset? Explain.

From (a), the best choice of (α, γ) is (0.9, 0.1), since it gives the smallest value $MSD = 57630.7$, so the smallest SSE.

(c) Examine the plot of the smoothed estimates and actual DJIA values. Does it appear the exponential smoothing has done well? Explain.

It seems to do reasonably well, but there is some pattern in the results. We see that it seems to overestimate the data for the first few years of the series, then overestimate in the remaining years.

(d) Using the choice of (α, γ) from (b), find a 95% prediction interval for the DJIA in 1996. How does this interval compare with the one found in Assignment #2?

In this case, $\tau = 2$, $T = 25$, $\alpha = 0.9$, $\gamma = 0.1$.

From the output below, $\hat{y}_{T+2}(T) = 4233.28$.

```
Forecasts

Period  Forecast    Lower    Upper
1995     4038.79  3577.11  4500.47
1996     4233.28  3573.04  4893.52
```

95% PI means $z_{.05/2} = z_{.025} = 1.96$.

From the MSD value,

$$s^2 = SSE/(T-2) = T/(T-2)57630.7 = (25/23)57630.7 = 62642.07$$

We'll also need $c_\tau = c_2 = 1 + \alpha^2(1+\gamma)^2 = 1 + (0.9)^2(1+.1)^2 = 1.9801$.

Then the 95% PI is

$$\begin{aligned} \hat{y}_{T+2}(T) \pm z_{.025} s\sqrt{c_2} &= 4233.28 \pm 1.96\sqrt{62642.07}\sqrt{1.9801} \\ &= 4233.28 \pm 690.2918 = (3542.99, 4923.57) \end{aligned}$$

On Assignment #3, we found the 95% PI for y (based on the linear regression model) was (2306.645, 4411.355).

The double exponential smoothing PI is narrower, and its lower and upper bounds are quite a bit larger than those from the regression model.