THE UNIVERSITY OF TEXAS AT AUSTIN
McCombs School of Business

STA 372.5 **Spring 2018**

HOMEWORK #5 – Due Wednesday, February 28

1. Problem #3 on the 2015 midterm exam.

2. Suppose the following simple exponential smoothing model is fit to the data:

$$Y_t = M_{t-1} + \varepsilon_t \qquad \varepsilon_t \text{ iid } N(0, \sigma^2 = 4)$$

$$M_{t-1} = 0.7Y_{t-1} + 0.3M_{t-2}.$$

Note that $\alpha = 0.7$ and $\sigma^2 = 4$ are known. Suppose the $t = 25^{th}$ observation is 3.0 and $M_{24} = 3.1$.

(a) What is the distribution of Y_{26} given information through time period $t = 25$? What is the forecast of Y_{26} given information through $t = 25$? What is the probability the actual value of Y_{26} that occurs is within one unit of the forecasted value?

(b) What is the distribution of Y_{27} given information through time period $t = 25$? What is the forecast of Y_{27} given information through $t = 25$? What is the probability the actual value of Y_{27} that occurs is within one unit of the forecasted value?

3. Read the case "Wachovia Bank and Trust Company, N.A. (B)." The data for this case is in STA372_Homework5_Question3.dat (for use in R) and STA372_Homework5_Question3.xlsx (for use in Excel) on the *Data sets* page of the Canvas class website. Let Y_t represent desseasonalized volume in week *t*. The .dat file contains *Week* in the first column, *Y* in the second column, and 730 in each row of the third column.

Parts (a)-(f) should be done using R.

(a) Read the data into a data table called *data_table* and name the columns *Week*, *Y* and *FixedForecast*.

Plot *Y* vs. *Week* using the *ggplot2* package. Is there a pattern in *Y* through time?

(b) Suppose the predecessor's long-run forecast of weekly volume of 730,000 is used each week. What is the RMSE for the in-sample forecasts?

To compute this RMSE, use the R commands:

```
data_table$Error <- data_table$Y - data_table$FixedForecast
RMSE_FixedForecast <- sqrt(sum(data_table$Error^2)/nrow(data_table))
```

(c) Another forecasting method the bank is considering is to use the current week's deseasonalized volume as the forecast of next week's deseasonalized volume. This is the forecast given by a random walk model. What is the RMSE for the in-sample forecasts?

Hint: A random walk model is a special case of the simple exponential smoothing model used in part (d) with $\alpha = 1$. Note that you can use the *ses* command in R with *alpha=1* to give you the results for a random walk model.

(d) Use the simple exponential smoothing model for deseasonalized volume

$$Y_t = M_{t-1} + \varepsilon_t \qquad \varepsilon_t \text{ iid } N(0, \sigma^2)$$

$$M_{t-1} = \alpha Y_{t-1} + (1 - \alpha) M_{t-2}$$

with α and M_0 estimated. Plot Y and the in-sample forecasts on the same graph. What is the RMSE for the in-sample forecasts?

(e) Plot the residuals from the model in part (d) and compute their autocorrelation function. Are the residuals independent? What does this imply about whether the information in the original time series about future observations has been properly extracted?

(f) What model should you use to forecast deseasonalized volume for the week of April 10-14 (i.e. week 67)? Using the appropriate forecast of deseasonalized volume and the seasonal index given in the case for the week of April 10-14, what is your forecast of volume (not deseasonalized volume)? What is the 95% confidence interval for your forecast of volume?

(g) **Part (g) does not need to be turned in.** However, you should do it to make sure you understand the mechanics of estimation for the simple exponential smoothing model.

Set up a spreadsheet in Excel to compute the in-sample forecasts using the model in part (d). Use Solver to estimate α and M_0. The estimates of these values should be the same as though obtained using the *ses* command in R in part (d) (subject to small differences due to different optimization routines being used in R and Solver).