THE UNIVERSITY OF TEXAS AT AUSTIN
McCombs School of Business

STA 372.5 **Spring 2019**

HOMEWORK #4 – Due Wednesday, February 27

1. Problem #3 on the 2015 midterm exam.

2. The time series Y_t shown below with n = 40 observations appears to "meander" through time. This means a reasonable model to consider for modeling its pattern is the simple exponential smoothing model:

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

and

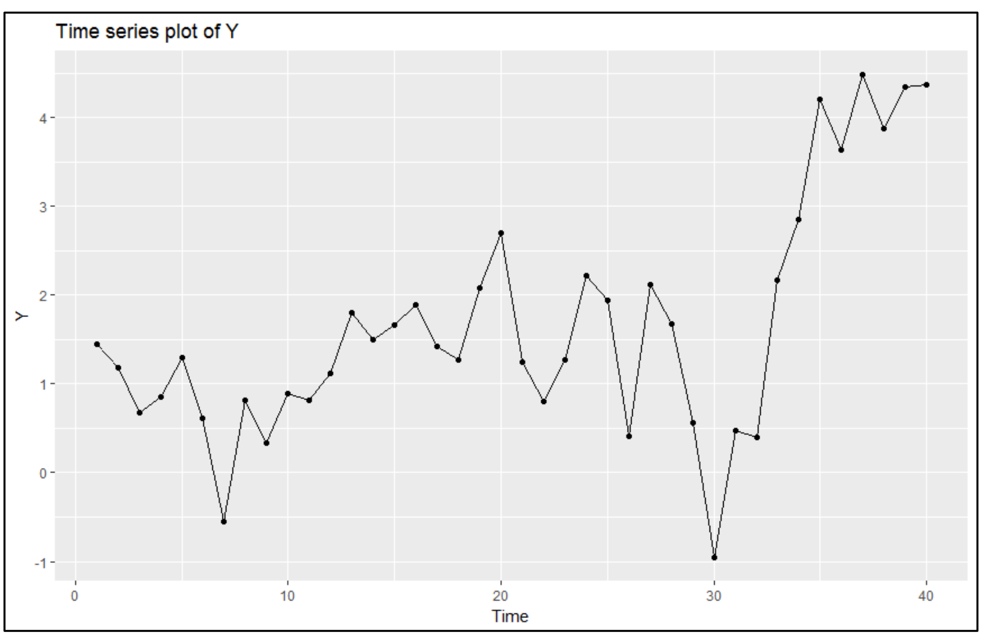
$$L_t = \alpha Y_t + (1 - \alpha)L_{t-1}.$$

Using all available information in the R output on the following pages, answers parts (a)-(c).

(a) What are the three forecasts for Y_{41}, Y_{42}, and Y_{43} that have been replaced with asterisks under *result* in the R output?

(b) What is the 95% confidence interval for Y_{42} that has been replaced with asterisks under *result* in the R output?

(c) What evidence is there that this model has done a good job modeling the pattern in the data?

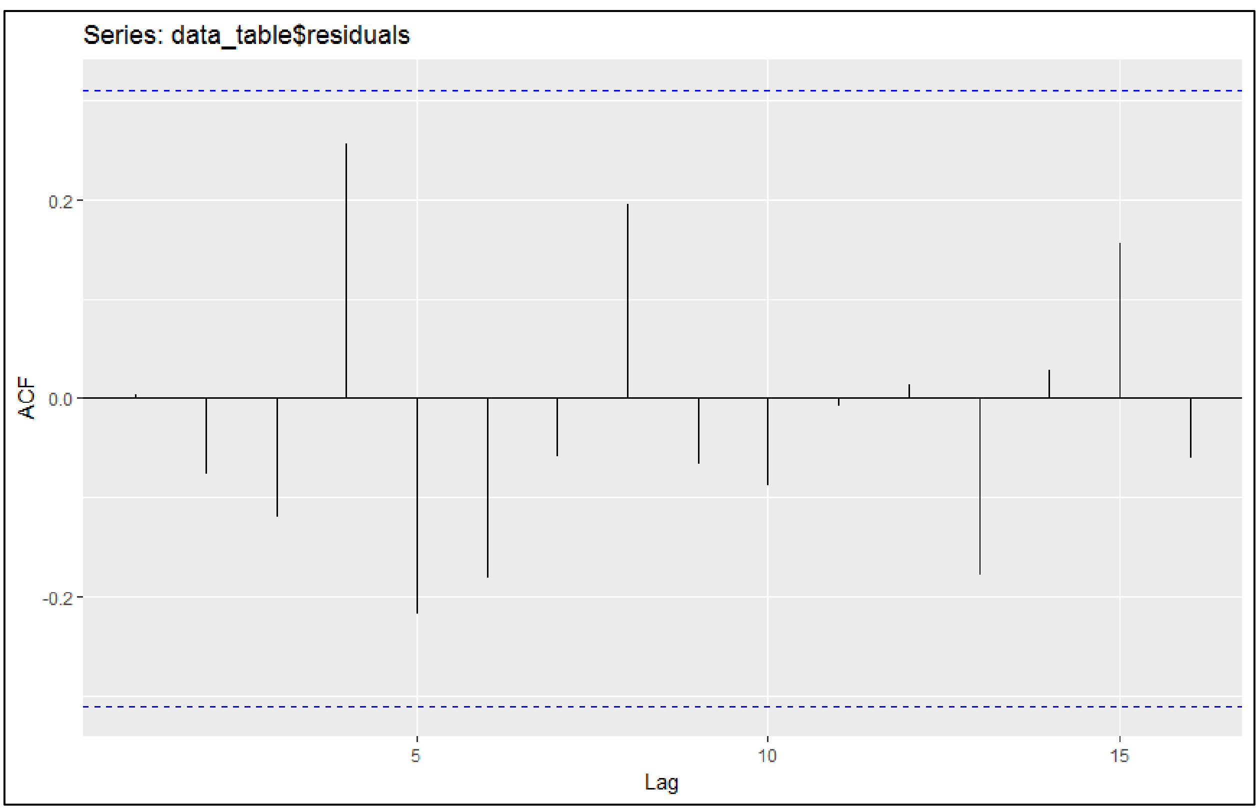


```
----------------------------------------------------------------------------------------
head(data_table)
tail(data_table)
----------------------------------------------------------------------------------------

  Time         Y
1    1 1.4528090
2    2 1.1866877
3    3 0.6786963
4    4 0.8505772
5    5 1.2934060
6    6 0.6071821

   Time        Y
35   35 4.205890
36   36 3.636372
37   37 4.482471
38   38 3.871737
39   39 4.339101
40   40 4.367359


----------------------------------------------------------------------------------------
#
#   Save data as a time series object
#
y_time_series <- ts(data_table[,2])
#
#   Use ses to obtained smoothed series
#
result <- ses(y_time_series, initial="optimal", h=3)
result$model
----------------------------------------------------------------------------------------

Simple exponential smoothing

Call:
 ses(y = y_time_series, h = 3, initial = "optimal")

  Smoothing parameters:
    alpha = 0.786

  Initial states:
    l = 1.375

  sigma:  0.848

     AIC     AICc      BIC
138.3174 138.9841 143.3841


----------------------------------------------------------------------------------------
result
----------------------------------------------------------------------------------------
   Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
41       ******** 3.257413 5.431032 2.682091 6.006354
42       ******** 2.961883 5.726562 ******** ********
43       ******** 2.719240 5.969206 1.859026 6.829420
```

```
-----------------------------------------------------------------------------
data_table$L <- result$model$state[2:41,1]
data_table$forecast <- result$fitted
data_table$residuals <- result$residuals
head(data_table)
tail(data_table)
-----------------------------------------------------------------------------
```

```
  Time         Y         L  forecast   residuals
1    1 1.4528090 1.4361596 1.3750092  0.07779975
2    2 1.1866877 1.2400754 1.4361596 -0.24947188
3    3 0.6786963 0.7988329 1.2400754 -0.56137911
4    4 0.8505772 0.8395038 0.7988329  0.05174428
5    5 1.2934060 1.1962697 0.8395038  0.45390216
6    6 0.6071821 0.7332485 1.1962697 -0.58908756

   Time        Y        L forecast  residuals
35   35 4.205890 3.867759 2.625858  1.5800323
36   36 3.636372 3.685890 3.867759 -0.2313870
37   37 4.482471 4.312000 3.685890  0.7965814
38   38 3.871737 3.965955 4.312000 -0.4402634
39   39 4.339101 4.259247 3.965955  0.3731462
40   40 4.367359 4.344223 4.259247  0.1081128
```

```
-----------------------------------------------------------------------------
#
#   Compute the autocorrelation function of the residuals
#
ggAcf (data_table$residuals)
-----------------------------------------------------------------------------
```



Suggested problem from previous exams

The following problem from the 2018 exam does not need to be turned in, although I strongly recommend that you do it. The answer will be distributed with the answers to this homework.

2018 exam: 5

3. Read the case “Wachovia Bank and Trust Company, N.A. (B).” The data for this case is in STA372_Homework4_Question3.dat (for use in R) and STA372_Homework4_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 frame called *data_table* and name the columns *Week*, *Y* and *FixedForecast*.

       Plot *Y* vs. *Week* using *ggplot2*. 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 following *ses* command in R with *alpha=1* to give you the results for a random walk model.

```
y_time_series <- ts(data_table[,2])
result_random_walk <- ses(y_time_series, alpha=1, initial="simple", h=1)
```

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

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

and

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

with α and L_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 L_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).