ASSIGNMENT 2

STAT 485/685 E100/G100: Applied Time Series Analysis
Fall 2020
Simon Fraser University

This week's assignment will be based on trend-fitting in time series analysis. The topics here have been covered in the Week 3 videos, as well as Chapter 3 of the textbook *Time Series Analysis with Applications in R (2nd ed.)* by Cryer & Chan.[1]

Due date: **Friday, Oct. 2nd at 11:59 pm (end of day)** (Pacific Time).

Marks: 10.

Please include your R code for each of the below questions. Some ideas for how you can most easily do this:

- Copy-paste your code/plots into a Word document along with your responses to the questions, and save as a PDF.
- Take images of your code/plots and upload to Crowdmark, along with your responses to the questions.
- Save your code and responses together in an RMarkdown document and save as PDF (if you've worked with RMarkdown before).

Other important policies on assignment submissions:

- Please write each question on a **separate page**!
- Please **show all your code and work**, in order to get full marks.
- Upload your complete answers as PDF files or high-resolution images.
- If you're hand-writing answers, please make sure they are **neat and clearly readable**, and that the photo is high resolution.
- Please **clearly label the question numbers**.

[1] Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R.* Springer Science and Business Media.

For this assignment, we will be using some datasets and functions in the `TSA` package in `R`. For instructions on how to install and load the package, please see the Week 1 module on Canvas.

1. *(8 marks)* The dataset `gold` gives a time series of the daily price of gold (in $ per troy ounce) for the 252 trading days of the year 2005.
   Remember to include all code and plots in your answers below.

   (a) Read in this dataset using the function `data()` in the `TSA` package.

   (b) Plot the dataset. Remember to label the x- and y-axes. Briefly describe what you see.

   (c) Fit a linear time trend to this dataset. What are the estimates of the intercept and the slope?

   (d) Add the fitted line to the plot using the function `abline()`. Show the plot.

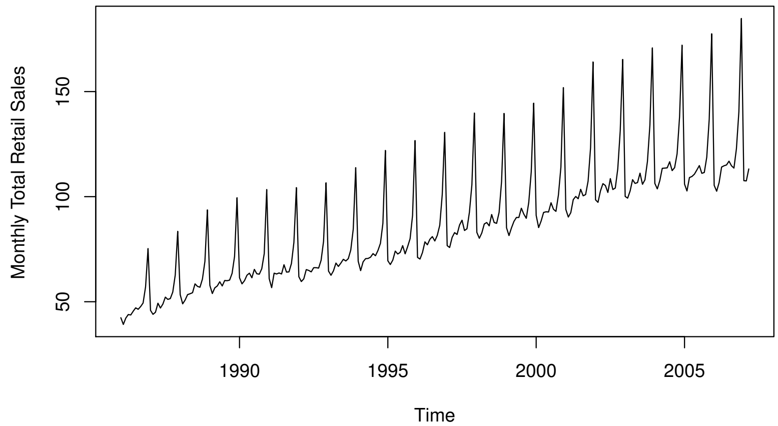
   (e) Using the estimates of the intercept and the slope, calculate the estimated mean price of gold at day 100.

   (f) Plot the (studentized) residuals of the model against time. Describe what you see in the plot.

   (g) The residual plot is telling us that the linear fit may not be appropriate. To fit a quadratic model, use the code
   ```
   my.model <- lm(gold ~ time(gold) + I(time(gold)^2))
   ```
   Note: The `I()` part of the code is necessary, because otherwise the ˆ symbol will not do what we want it to.)
   What are the three parameter estimates?

   (h) Plot the (studentized) residuals of the quadratic model against time. What do you see now?

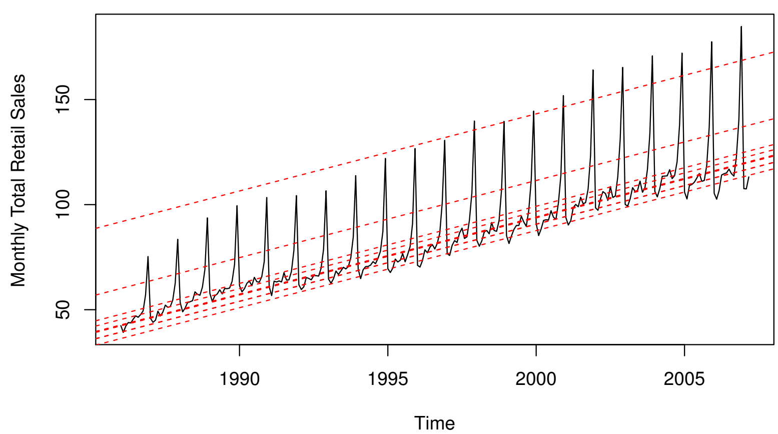
2. *(2 marks)* The `retail` dataset gives the monthly total retail sales (in billions of pounds) in the UK, from 1986 to 2007. The plot is shown below:



In this dataset, we see both seasonal behaviour and a linear time trend. We decide to fit a combination of these two models: a *seasonal means plus linear time trend* model. This can be done using the following code:

```
data(retail)
month. <- season(retail)
model.retail <- lm(retail ~ month.-1 + time(retail))
```

This model fits separate lines of retail sales vs. time for each month, where the lines are assumed to be parallel. In other words, it assumes that the retail sales are increasing at the same rate for each month. Here are the lines we get:



Remember to include all code in your answers below.

(a) Fit the model using the code above. Show the table of parameter estimates you get as a result.

(b) If you look closely at the output from `time(retail)`, you will see that the times are not coded as $t = 1, 2, 3, 4, \ldots$, but as $t = 1986.000, 1986.083, 1986.167, 1986.250, \ldots$. In other words, each time unit denotes 1 full year. That doesn't change things too much – we just have to keep it in mind when talking about t.

Recall that the seasonal means model is defined as

$$\mu_t = \begin{cases} \beta_1 & \text{for all Januarys} \quad (t = 1986.000, 1987.000, \ldots) \\ \beta_2 & \text{for all Februarys} \quad (t = 1986.083, 1987.083, \ldots) \\ \beta_3 & \text{for all Marches} \quad (t = 1986.167, 1987.167, \ldots) \\ \vdots & \vdots \\ \beta_{12} & \text{for all Decembers} \quad (t = 1986.917, 1987.917, \ldots) \end{cases}$$

(As we've discussed above, the times are no longer $1, 2, 3, \ldots$ but fractions of the years.)

When we added a linear trend term to this model in part (a), the model became:

$$\mu_t = \begin{cases} \beta_1 + \alpha\, t & \text{for all Januarys} \quad (t = 1986.000, 1987.000, \ldots) \\ \beta_2 + \alpha\, t & \text{for all Februarys} \quad (t = 1986.083, 1987.083, \ldots) \\ \beta_3 + \alpha\, t & \text{for all Marches} \quad (t = 1986.167, 1987.167, \ldots) \\ \vdots & \vdots \\ \beta_{12} + \alpha\, t & \text{for all Decembers} \quad (t = 1986.917, 1987.917, \ldots) \end{cases}$$

where α is the slope for time, which is also given to us in the parameter estimates table.

What is the estimate of the mean trend at $t = 1987.167$? *(Hint: First determine which month this corresponds to.)*

If you wish to better understand the "seasonal means + linear trend" model, more information can be found in the appendix on the next page. We will be going over this model again in Week 4.

Appendix: Seasonal Means + Linear Trend Model

Before we can better understand the model in Question 2, we need to first be familiar with *indicator variables*. You may have seen these in an earlier course on linear regression.

Recall the definition of the *seasonal means model*, used for seasonal data:

$$\mu_t = \begin{cases} \beta_1 & \text{for all Januarys} \quad (t = 1986.000, 1987.000, \dots) \\ \beta_2 & \text{for all Februarys} \quad (t = 1986.083, 1987.083, \dots) \\ \beta_3 & \text{for all Marches} \quad (t = 1986.167, 1987.167, \dots) \\ \vdots & \vdots \\ \beta_{12} & \text{for all Decembers} \quad (t = 1986.917, 1987.917, \dots) \end{cases}$$

In this model β_j represents the mean of the model at month j. An alternative way of representing this model is:

$$\mu_t = \beta_1 I_{January} + \beta_2 I_{February} + \cdots + \beta_{12} I_{December}$$

where the I's are just indicator variables for each month: They take on the value of 1 for their corresponding month, and 0 otherwise.

For example: For any month that happens to be a January, $I_{January} = 1$ and all the other I's are zero, so μ_t reduces down to just β_1. You can check for yourself that these two model formulations are equivalent.

Then, if we wish to add a linear time trend to the seasonal means model, we write:

$$\mu_t = \beta_1 I_{January} + \beta_2 I_{February} + \cdots + \beta_{12} I_{December} + \alpha\, t,$$

where α is just another parameter for us to estimate.

What does this model reduce down to, for different months? If the month t is a January, we get that $\mu_t = \beta_1 + \alpha\, t$. If the month t is a February, we get $\mu_t = \beta_2 + \alpha\, t$, and so on. In other words, we have separate lines for each month, where the lines have the same slope but different intercepts. So they are parallel, but some are higher in the plot than others. This is the same as writing:

$$
\mu_t = \begin{cases}
\beta_1 + \alpha\, t & \text{for all Januarys} \quad (t = 1986.000, 1987.000, \dots) \\
\beta_2 + \alpha\, t & \text{for all Februarys} \quad (t = 1986.083, 1987.083, \dots) \\
\beta_3 + \alpha\, t & \text{for all Marches} \quad (t = 1986.167, 1987.167, \dots) \\
\vdots & \vdots \\
\beta_{12} + \alpha\, t & \text{for all Decembers} \quad (t = 1986.917, 1987.917, \dots)
\end{cases}
$$