W271 Assignment 3

Due 11:59pm Pacific Time Sunday November 24 2019

Instructions (Please Read Carefully):

- No page limit, but be reasonable
- Do not modify fontsize, margin or line_spacing settings
- This assignment needs to be completed individually; this is not a group project. Each student needs to submit their homework to the course github repo by the deadline; submission and revisions made after the deadline will not be graded
- Answers should clearly explain your reasoning; do not simply 'output dump' the results of code without explanation
- Submit two files:

  1. A pdf file that details your answers. Include all R code used to produce the answers. Do not suppress the codes in your pdf file
  2. The R markdown (Rmd) file used to produce the pdf file

  The assignment will not be graded unless **both** files are submitted
- Use the following file-naming convensation:
  - StudentFirstNameLastName_HWNumber.fileExtension
  - For example, if the student's name is Kyle Cartman for assignment 1, name your files follows:
    - KyleCartman_assignment1.Rmd
    - KyleCartman_assignment1.pdf
- Although it sounds obvious, please write your name on page 1 of your pdf and Rmd files
- For statistical methods that we cover in this course, use the R libraries and functions that are covered in this course. If you use libraries and functions for statistical modeling that we have not covered, you must provide an explanation of why such libraries and functions are used and reference the library documentation. For data wrangling and data visualization, you are free to use other libraries, such as dplyr, ggplot2, etc.
- For mathematical formulae, type them in your R markdown file. Do not e.g. write them on a piece of paper, snap a photo, and use the image file.
- Incorrectly following submission instructions results in deduction of grades
- Students are expected to act with regard to UC Berkeley Academic Integrity

Question 1 (1 point):

Backshift Operator Expression

a. Write down the $ARIMA(2,0,2)(1,0,1)_4$ in terms of (1) backshift operators and (2) the fully-expressed form as y_t as a function of lags of y_t and the shock ω_t.

   For example, for the $ARIMA(1,0,1)(0,0,0)_4$ model, you will write down:

   1. $(1-\phi_1 B)y_t = (1+\theta_1 B)\omega_t$
   2. $y_t = \phi_1 y_{t-1} + \omega_t + \theta_1 \omega_{t-1}$

b. Write down the $ARIMA(2,1,2)(1,0,1)_4$ in terms of (1) backshift operators and (2) the fully-expressed form.

c. Write down the $ARIMA(2,1,2)(1,1,1)_4$ in terms of (1) backshift operators and (2) the fully-expressed form.

Parameter Redundancy, Stationarity, and Invertibility

In each of the following cases, (1) check for parameter redundancy and ensure that the $ARMA(p,q)$ notation is expressed in the simplest form, and (2) determine whether they are stationary and/or invertible.

a. $y_t = y_{t-1} - \frac{1}{4}y_{t-2} + \omega_t + \frac{1}{2}\omega_{t-1}$

b. $y_t = 2y_{t-1} - y_{t-2} + \omega_t$

c. $y_t = \frac{7}{10}y_{t-1} - \frac{1}{10}y_{t-2} + \omega_t + \frac{3}{2}\omega_{t-1}$

d. $y_t = \frac{7}{10}y_{t-1} - \frac{1}{10}y_{t-2} + \omega_t + \frac{1}{3}\omega_{t-1}$

Question 2: Time series merging and interpolation (2 points)

The file `AMZN.csv` contains Amazon share price data obtained from Yahoo! Finance. The file `UMCSENT.csv` contains the University of Michigan Consumer Sentiment index obtained from the Federal Reserve Economic Database (FRED).

Read `AMZN.csv` and `UMCSENT.csv` into R as dataframes and convert them to time series objects. You may find it advantageous to work with `xts` rather than `ts` objects for the following questions (refer to `xts.Rmd` in the github repo's `live_session_7` folder).

a. Merge the two set of series together, preserving all of the observations in both set of series

b. Fill all of the missing values of the UMCSENT series with -9999

c. Create the following new series from the original UMCSENT series:

    - UMCSENT02, replacing all of the -9999 with NAs
    - UMCSENT03, replacing the NAs with the last observation
    - UMCSENT04, replacing the NAs using linear interpolation

    Print a few observations to ensure that your merge as well as the missing value imputation are done correctly. Choose a reasonable number of observations (do not print out the entire dataset).

d. Calculate the daily return of the Amazon closing price, where daily return is defined as $(x_t - x_{t-1})/x_{t-1}$. Plot the daily return series.

e. Create a 20-day and a 50-day rolling mean series from the AMZN close series.

Question 3: Atmospheric CO2 Concentration (4 points)

The file `mauna_loa.csv` contains weekly observations of atmospheric carbon dioxide concentration measured at the Mauna Loa observatory in Hawaii, obtained from the National Oceanic and Atmospheric Administration (NOAA), dating from 1974 to 2019.

a. Conduct a thorough EDA of the time series and develop a model that captures both trend and seasonality in the series, following all appropriate steps and conducting suitable diagnostics. Use the model to generate a 2-year-ahead forecast and plot this. In what year does your model predict that atmospheric CO2 will first reach 420 parts per million?

b. Load the `co2` dataset from the R's `datasets` package. This is a time series of monthly observations from 1959 to 1997. Use averages from the NOAA data to extend the monthly series to 2019. Repeat the previous analysis and forecast for this monthly 1959-2019 series. Assess your model's pseudo-out-of-sample forecasting performance using a rolling test set window, and plot the distribution of RMSEs over this range of test sets.

Question 4: Vector Autoregression (3 points)

Use the series contained in `Q4.txt` to conduct a multivariate time series analysis and build a model to forecast the series. In model estimation, do not use the observations in 1993. All relevant time-series model building steps are applicable. Measure and discuss your model's performance, using both in-sample and out-of-sample model performance. When training your model, exclude all the observations in 1993. For the out-of-sample forecast, measure your model's performance in forecasting 1993. Discuss the model performance and forecast a 12-month forecast beyond the last observed month in the given series.