UNIVERSITY OF BRITISH COLUMBIA
Department of Statistics
Stat 443: Time Series and Forecasting
Assignment 3: Analysis in the Time Domain

The assignment is due on **Thursday, March 24** at **9:00pm**.

- Submit your assignment online in the **pdf format** under module "Assignments".
- The assignment must be completed in R Markdown; display all the R code used to perform your data analysis.
- Please make sure your submission is clear and neat. It is the student's responsibility that the submitted file is in good order (e.g., not corrupted and is what you intend to submit).
- **Late submission penalty**: 1% per hour or fraction of an hour. (In the event of technical issues with submission, you can email your assignment to the instructor to get a time stamp but submit on canvas as soon as it becomes possible to make it available for grading.)

1. The file rimouski.csv contains monthly mean max temperatures from Rimouski, Quebec climate station used in Assignment 1. The data starts in January 1954 and ends in August of 2017. The column 'Mean.Max.Temp' is the monthly mean max temperature measured in degrees Celsius. Missing data have been imputed using values from the month in the previous year.

   (a) Import the data into `R` and create a time-series object for the monthly mean temperature. Break the data into training and testing sets:

      i. the training data starts January, 1954 and ends at December, 2010;
      ii. the testing data starts at January, 2011 and ends at December, 2016.

      Plot the training data using `R` along with its acf and pacf. Ensure at least 5 full seasonal cycles can be observed in the acf and pacf. Comment on what you observe.

   (b) Fit a SARIMA $(0,0,0) \times (0,1,0)_s$ model to the monthly mean temperature data using the `arima()` function.

      i. Write down the model equation(s) and express X_t as an ARMA(p,q) process.
      ii. Give estimates for the model parameter(s). Specifically, state the parameters and their estimates. Printing the results of the `arima()` function without further comments is not enough.
      iii. Plot the residuals and the acf and pacf of the residuals. Comment on the plots.

   (c) We propose adding a seasonal MA$(Q = 1)$ component.

      i. Why is this an appropriate choice? Justify using the residual acf and pacf plots from part (b).
      ii. Write down the $(0,0,0)\times(0,1,1)_s$ model equations and express X_t as an ARMA(p,q) process.

iii. Fit the model using the `arima()` function, and specifically identify the model parameters and their estimates (do not just print the function results).

iv. Compare the AIC of this model with the SARIMA $(0,0,0) \times (0,1,0)_s$ model. Which one would you choose based on AIC?

(d) We now add an $AR(p=1)$ component.

i. Plot the acf and pacf of residuals from the SARIMA $(0,0,0) \times (0,1,1)_s$ model and justify the choice of adding an AR(1) component.

ii. Write down the SARIMA $(1,0,0) \times (0,1,1)_s$ model equations and express X_t as an ARMA(p,q) process.

iii. Fit the model to the data using `arima()` and specify each parameter's estimate.

iv. Using the diagnostic plots from `tsdiag()`, say whether this model seems like a good fit.

v. Using AIC, have we improved our model by adding the AR(1) component?

(e) Use the `predict(sarima.object, n.ahead=p)` function to predict the monthly mean temperatures for 2011-2016 using the model fit from part (d). In a single plot, show both the predictions and the test set data, using a legend to distinguish between the predicted and observed values. Comment on whether the predictions are reasonable and how they differ from the test data.

(f) Using the `HoltWinters()` function, fit a Holt-Winters model to the data.

i. Do the results of the Holt-Winters model indicate a trend?

ii. Use `predict()` to predict the monthly mean max temperatures for 2011-2016. Plot the Holt-Winter predictions, the Box-Jenkins predictions from part (e), and the testing data in the same plot. Comment on the comparison between the two forecasting methods.

iii. Calculate the mean squared prediction error (MSPE) for the Box-Jenkins model and the Holt-Winters model. Which method performs better?

2. The file bynd.txt contains the value, in \$, for the daily closing price of a share in Beyond Meat, Inc. (BYND) for 716 consecutive trading days, from May 5th, 2019 to March 3rd, 2022.

(a) Read the data set into `R`, and coerce the data into a time series object. Create a plot of the data, and plot the acf and pacf of the series. Comment on what you observe.

(b) Create a new series by differencing consecutive values in the time series. Plot the acf and pacf of the new series. Does the first-differenced series appear stationary?

(c) Explain briefly why a model of the form

$$X_t = X_{t-1} + Z_t,$$

where $\{Z_t\}$ is white noise with mean zero and constant variance, may be suitable for the share price series.

(d) Assume the model above is appropriate in what follows. Which ARIMA model is it?

(e) Estimate σ^2, the variance of $\{Z_t\}$.

(f) Write down what you would use for the forecast at lead time l, for any $l > 0$.

(g) Write down the 90% prediction interval for the forecast at lead time l. Provide a prediction and 90% prediction interval for the stock price one week in the future (March 10th).

(h) A director of Beyond Meat Inc. is considering selling some shares in the company in the near future. Write a paragraph explaining your recommendations to this director.