UNIVERSITY OF BRITISH COLUMBIA
Department of Statistics
STAT 443: Time Series and Forecasting
Assignment 1: Analysis in the Time Domain

1. (10 marks) This question concerns a test to determine whether it seems likely that a time series is a realization of a white noise process. As usual, we assume we have observed a series $x(1), \ldots, x(N)$ from a stochastic process $X(t)$, $t = 1, 2, \ldots$ Use the `rnorm` command in R to generate an i.i.d. series of size 200 from the $N(0, 2^2)$ distribution. Find the acf of your data and find out how to extract the values from R up to a given lag (you will find `help(acf)` useful).

   (a) One test for white noise is the *portmanteau lack–of–fit test.* This test has test statistic

   $$Q := N \sum_{k=1}^{M} r_k^2,$$

   where r_k is the sample autocorrelation at lag k, N is the number of terms in the series and M is an integer rather less than N, usually between 15 and 30. If the observations are from a white noise process, then approximately $Q \sim \chi^2_M$. Otherwise the value of Q is inflated. The choice of M is not straightforward, however. Perform this test for two different values of M for the data you generated, selecting the values of M at random between 15 and 30 inclusive. Quote the P–value of the test on each occasion, and comment on your results.

   (b) The data file `SP500.txt` contains 523 consecutive closing values for Standard and Poor's 500 Index. Read the data set into R, and coerce the data into a time series object. Create a plot of the data, and of the acf of the series. Comment on what you observe.

   (c) Use a portmanteau lack–of–fit test with $M = 25$ to decide whether the series appears to be a realization from a white noise process.

   (d) Apply an operator to the SP500 series that might reasonably be expected to remove the non-stationary component. Plot the new series and its acf, and comment.

(e) Use a portmanteau lack–of–fit test with $M = 25$ to decide whether the series you created in (d) appears to be a realization from a white noise process.

2. (10 marks) The variable `Close` in the data file `FeedOneCloseJan15Feb16` gives closing prices (in Yen) of Feed One Co. Ltd., as listed on the Tokyo Stock Exchange, 5th January 2015 to 29th January 2016.

   (a) Plot the time series and the acf of the series, and comment on what you observe. Does the closing price time series appear to be stationary?

   (b) Apply the first difference operator to the time series. Plot the new series, the acf, and pacf of the new series, and comment on their patterns. Does the new series appear to be stationary?

   (c) Fit various models from the ARMA(p, q) family to the new series. You should only consider models for which $p + q \leq 6$, and you should state which three models you would consider to be the best options and clarify why you chose those three models. Write out your three possible models in full, providing all the parameter estimates.
   Note 1: Fit your models directly to the series created in (b), as `arima(x, order = c(p, 1, q))` gives a different model to `arima(diff(x), order = c(p, 0, q))`. Clarify this for yourself and identify how the two differ.
   Note 2: For models with an AR component, R's use of the term "intercept" is not intuitive.

   (d) An alternative, and modern, way of model selection is to split the data into a *training set* and a *test set.* The idea is that we use the training set to determine fit and explore competing models, and then assess how well the models perform when fitting values in the test set. A popular criterion to adopt in assessing a model's performance on the test set is mean squared error.
   The above can be applied to time series in R. Suppose the time series is `x` and we wish to use the final `m` values as the training set. Then
   ```
   train <- 1:(length(x)-m)
   trainx <- x[train]
   ```

```
testx <- x[-train]
# sets up training set and test set
model <- arima(trainx, order = c(p, 0, q))
foremodel = predict(model, m)
# fits model to training set, uses model to predict test
set
error <- sum((testx - foremodel$pred)^2)
# computes squared errors over test set
```

Using the differences in January 2016 closing prices as the test set for your models and the remaining data as the training set, compare the three models you selected in (c) for performance over the test set as above.

(e) Use your winning model from (d), when fitted to the entire series, to forecast the next two values in the differenced series. Hence forecast the next two values in the closing price series.

BD