# MSDS 604 Time Series Analysis

Shan Wang

Fall 2023 Module 2

Homework 4 Due Monday Nov 20 11:59pm

**Introduction**: submit one .pdf file containing all you answers for the homework, including screenshot of python code, output or plot if asked. The .pdf can be converted from Latex file, pictures of your handwriting solutions, word files, markdown files and .etc (anything that can be converted into .pdf). If there are coding problems, only include the answer of the question in the .pdf, and upload a separate notebook for Python code. This homework requires the submission of both pdf file and python notebook.

1. Statement questions:

   (a) True or False: If $\{X_t\}$ is stationary, then $(1-B)X_t$ (difference once) is still stationary.

   (b) Do a little research on the internet, and briefly state the problem with over-differencing. Cite the source of your information in the answer.

2. **In this question you will become acquainted with the relationship between $ARIMA$ models and $ARMA$ models**: in general, an $ARIMA(p,d,q)$ can be represented as a non-stationary $ARMA(p+d,q)$ model.

   - Consider the $ARIMA(1,1,1)$ model given by

     $$(1-\phi B)(1-B)Y_t = (1+\theta B)\epsilon_t$$

     by expanding the backward shift operator notation show that this $ARIMA(1,1,1)$ model can be represented as a non-stationary $ARMA(2,1)$ model. No need to show non-stationary, just show the order matches.

3. **In this question you will acquainted with fitting a data with trend but no seasonality.** The *sb_history.csv* dataset contains the the monthly adjusted closing price data on Starbucks from March 1993 to December 2003. The *sb_test.csv* dataset contains the monthly adjusted closing price data on Starbucks from Jan 2004 to December 2004. The goal of this question is to work on the history data to forecast the next 12 months. We put the test data in the box; do not use it until mentioned.

   (a) In this part, we will practice differencing data manually.

      i. Difference the sb_history.csv data until you decide the parameter value d: each time, run the time series plot, ACF plot and ADF test to check for stationary.

ii. Design a 3-fold cross-validation to select a ARMA(p,q) model for the differenced data.

iii. Fit the selected ARMA(p,q) model on the differenced data.

iv. Forecast the differenced data for the next 12 steps.

v. Convert differencing to get the forecast for the original variable for the next 12 steps.

vi. Use sb_test.csv data and the forecast from the above step to calculate the test RMSE.

vii. Draw the plot with the history, test, and forecast.

(b) In this part, we will practice choosing all the ARIMA parameters automatically, including the *trend* parameter.

i. Design a 3-fold cross-validation to select a ARIMA(trend_order=(p,d,q), trend) model for sb_history.csv data. Hint: the trend parameter is defined as below, and the minimum order of the trend polynomial needs to be greater or equal to d.

> **trend** : `str` {'n','c','t','ct'} `or` iterable, `optional`
>
> Parameter controlling the deterministic trend. Can be specified as a string where 'c' indicates a constant term, 't' indicates a linear trend in time, and 'ct' includes both. Can also be specified as an iterable defining a polynomial, as in *numpy.poly1d*, where *[1,1,0,1]* would denote $a + bt + ct^3$. Default is 'c' for models without integration, and no trend for models with integration. Note that all trend terms are included in the model as exogenous regressors, which differs from how trends are included in `SARIMAX` models. See the Notes section for a precise definition of the treatment of trend terms.

ii. Fit the selected ARIMA model with sb_history.csv data.

iii. Use the fitted model to forecast the next 12 steps.

iv. Use sb_test.csv data and the forecast from the above step to calculate the test RMSE.

v. Draw the plot with the history, test, and forecast.

(c) Did you get the same forecast results from (a) and (b)? Briefly discuss why.