MSA8200 Mini Project

List of Topics:

1. Predicting sea surface temperature

Data: https://github.com/asbates/bayes-time-series/blob/master/data/gilbralter_time_series_r_2.csv

Implementation: https://www.youtube.com/watch?v=pPO5av4HD90

Main methods covered: BSTS (Bayesian Structured Time Series) Model

Additional variables: temperature at different sea level

2. Predicting Walmart sales

Data and description: https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting

Implementation: https://rpubs.com/spillai/walmart_store_sales_forecast

3. Predicting stock price using dividend

Data: Stock_Returns_1931_2002.xlsx @ https://www.princeton.edu/~mwatson/Stock-Watson_3u/Students/EE_Datasets/

Data description: https://www.princeton.edu/~mwatson/Stock-Watson_3u/Students/EE_Datasets/

Implementation 1: https://www.econometrics-with-r.org/14-9-can-you-beat-the-market-part-ii.html (may also refer to the part I)

Implementation 2: https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/ (though using a different dataset)

Additional variables: dividend

4. Predicting Google Trend using ads spending and weather

Data and implementation: https://github.com/loganguerry/Google-Search-time-series-prediction

Additional variables: weather, ads spending

List of Methods:

1. Facebook Prophet:
   a. tutorial with R/python codes:
      https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/
   b. The official document: https://facebook.github.io/prophet/
2. BSTS (Bayesian structural time series): tutorial with R codes:
   http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html

Things to do:

A group of 3-4 students will form a team. Each team can pick a data set to perform a comparative study of three methods: 1. SARIMA without including additional variables 2. SARIMA with additional variables. 3. A method of your choice: prophet, BSTS.

Make sure the cover the following steps:

1. Data exploration and initial insights, this includes data visualization, pattern discovery, relationship discovery (between y and x, ACF, PACF, …), etc.

2. Split the data into training (the first 80% obs) and testing (the remaining 20% obs)

- Train your model using the training set following the procedure described. Make sure to explain your model component, model fitting results and model diagnostics.
- Then perform prediction for the remaining 20% observations and calculate MSPE. For time series model, using the 1-step-ahead prediction.

3. A discussion of what works vs. what doesn't work.

Each team will prepare a 10 minutes' presentation and a report. The presentation is on March 11th. The report and the slides should be submitted by 11:59 pm, March 15th.

Please signup using the google sheet:

https://docs.google.com/spreadsheets/d/12B4Cdxp3kwoXehA-UNflFS7c0Vzfe7bEao3hkudctZk/edit?usp=sharing