Problem Set 5

(Due: 1:00 PM, Tuesday, May 2)

1. **Interview questions.**

   (a) In a logistic regression, $F(\mathbf{x}, \theta) = \frac{e^{\theta' \mathbf{x}}}{1 + e^{\theta' \mathbf{x}}}$, what does the coefficient θ_i associated with x_i mean? Answer this in economic terms. For example, say $\theta_i = 1$. What happens when x_i increases/decreases by 1 unit?

   (b) What determines the degrees of freedom of a K-NN model?

2. **Forecasting corporate earnings:** It is the beginning of 2020. You are an equity analyst in charge of building a model to forecast the quarterly earnings of Wal-Mart Stores, Inc. Download the data file "WMT.csv", which contains quarterly earnings (based on net income) for Wal-Mart.

   (a) Denote the log quarterly earnings by x_t. Plot the time series of the first difference $\Delta x_t = x_t - x_{t-1}$, and the seasonal difference $\Delta_4 x_t = x_t - x_{t-4}$. Explain their economic meanings and comment on their statistical properties.

   (b) As a first attempt, build an ARIMA(0,1,1) for x_t,

   $$x_{t+1} - x_t = a_0 + \varepsilon_{t+1} - \theta \varepsilon_t.$$

   Use data from 1971Q1 to 2015Q4 to estimate this model. (In R, you can use the function *arima* to do the estimation.)

   (c) Next, build an "airline model" for x_t,

   $$(x_t - x_{t-1}) - (x_{t-4} - x_{t-5}) = (\varepsilon_t - \theta_1 \varepsilon_{t-1}) - \theta_4 (\varepsilon_{t-4} - \theta_1 \varepsilon_{t-5}).$$

   Again, use data from 1971Q1 to 2015Q4 to estimate this model. Explain the economic meanings of θ_1 and θ_4.

   (d) Use the period from 2016Q1 to 2019Q4 as the testing period. Generate your one-quarter ahead forecast[1] for Wal-Mart's earnings in fiscal quarter Q1 of 2016 through Q4 of 2019 based on the two estimated models. (Hint: You can use the function *predict* in R to produce the forecasts. Also, notice the constant term a_0 in the ARIMA(0,1,1), which makes its forecast different from a simple exponential moving average.)

   ---
   [1] One-quarter-ahead forecast means predicting x_{t+1} while taking all the observations up to x_t as given.

(e) Plot the forecasting errors of the two models from 2016Q1 to 2019Q4. Summarize the forecasting errors of the two models using the mean-squared error (MSE), defined as

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_{t+i} - E_{t+i-1}[x_{t+i}])^2,$$

where x_{t+i} is the realized earnings in quarter $t + i$, and $E_{t+i-1}[x_{t+i}]$ is your prediction for that quarter. Which model performs better?