STAT 1221/2220
Homework 22 Solutions

1. Redditors across the country flocked to Robinhood at the beginning of 2021 to send the price of Gamestop's stock to the moon. Prices were extremely volatile for weeks. This regression will analyze Gamestop's stock prices from January 4, 2021 through March 31, 2022. (Keep in mind that the stock market is closed on the weekend so although the time variable increments by 1, there will be a gap in dates between Friday and Monday in the data.)

   (a) Create a time series plot of Gamestop stock price against time in R.

   (b) Describe the relationship between price and time according to the plot in part (a).

   > **Solution:** In general, there does not appear to be a linear trend. Prices were extremely volatile during the first 50 days. They were still volatile between days 50 and 150, but less so. After around day 150, the random fluctuations in stock price became smaller in magnitude. Prices declined fairly consistently between days 225 and 300, but have been rising steadily since then.

   (c) Fit the simple linear regression that uses time to predict Gamestop stock price. Print the model summary.

   (d) Calculate and save the residuals to the data frame. Then create and print a time series plot of the residuals against time.

   (e) Run the Durbin-Watson test in R on the model created in part (c). Print the results of the test.

   (f) Does the independence condition appear to hold for the data? Justify your answer using the plot in part (d) and the Durbin-Watson test in part (e).

   > **Solution:** The independence condition does not hold. There appears to be positive autocorrelation in the time series plot as positive residuals tend to be followed by several more positive residuals. The same tends to happen once the residuals flip to being negative. The Durbin-Watson tests supports the notion of positive autocorrelation with a test statisic of 0.17 and a p-value close to 0.

   (g) Use the **ar.ols** function to generate the autoregressive model, letting R choose the optimal order. Set **order.max** equal to 10. Print the model at the end.

   (h) Report the regression equation from the autoregressive model.

   > **Solution:** $\hat{Y} = 16.15 + 0.7738L_1 + 0.2921L_2 - 0.0002L_3 - 0.3258L_4 + 0.1627L_5 + E$

   (i) Calculate the predicted Gamestop stock price for April 1 according to the autoregressive model. (Hint: Open the data frame in R to obtain the five most recent prices to use as the lagged values.)

   > **Solution:**
   > $\hat{Y} = 16.15+0.7738(166.58)+0.2921(166.85)-0.0002(179.90)-0.3258(189.59)+0.1627(151.95) = 156.70$

   (j) Refer back to part (g) where R chose the order of the model. Create the lagged terms in R up to and including the order of the model, and save the values to the data frame. Then generate the correlation matrix between the price and the lagged terms. Use the code **cor(df[c(6:314), c(3:8)])** where df is the name of the data frame. Row 6 should be the first row with a complete set of lagged observations and the prices and lagged terms should be in columns 3 through 8.

   (k) Look at the diagonal entries of the correlation matrix. Why do all of the diagonal entries have the same correlation?

Solution: Because the predictors are created by lagging, the same values occur in the predictors but in different rows. For example, consecutive columns are all offset by one day so they will always share the same correlation (0.906). Columns that have one column between them are offset by two days and also share the same correlation (0.853).

(l) Generate the autoregressive model using the 'lm' function and the lagged terms created in part (j). Save the residuals from the autoregressive model to the data frame. Then create and print the residual plot of residuals against time. (Recall that the 'ar.ols' function does not allow you to make predictions so the model has to be created a second time.)

(m) Run the Durbin-Watson test on the autoregressive model and print the results.

(n) Does the autoregressive model appear to satisfy the independence condition? Justify your answers using the results in parts (l) and (m).

Solution: The autoregressive model does appear to satisfy the independence condition. There is no indication of any kind of pattern in the residuals as they tend to bounce back and forth between positive and negative randomly. The Durbin-Watson test supports the notion of independence with a test statisic of 2.03 and a p-value that tends to be between 0.80 and 0.85. (Remember that the p-value is calculated through a simulation so an exact value is not reported. However, it is certainly going to be large enough to fail to reject H_0.)