Homework #4

Le Wang

Instruction: Do all the following empirical exercises using R. Turn in your R markdown file with answers and supporting tables and graphs, if any. Refer to the R output whenever appropriate when discussing your results. It does not matter which method you use to plot a time series variable.

Question 1: [Google Searches]

Stock investment often depends not only on fundamentals about a firm, but also on consumer interest in the firm. One way to gauge the latter is probably see how many people have searched for this firm online. Let's use Google Trends to study google stock.

1. Lets download the google searches of **Google Stock** for the past 30 days. You have to modify my code to achieve this. See manual for instructions here.
2. Plot your time series of "interest".
3. Obtain and plot stock prices on Google using the `quantmod` package for the same time period. Do you see any obvious relationship between goolge searches and stock prices?
4. Lets focus on the searches within U.S. Which state has most interest in Google Stock? Plot the frequency by states.

Question 2: [Twitter and Sentiment Analysis]

Follow the instructions to register an app to use use Twitter API. Click here for instructions. Save your credentials and then

1. dev.twitter.com
2. Roll down to "Manage Your Apps"
3. Log into your account
4. Obtain Keys and Access Tokens

Then use my example code on Github and fill in your own credentials.

1. Find the coordinates for your hometown. Read this article to find out how in Google Maps.
2. Specify a topic or person that you care about (for example, a candidate for an upcoming election) and would like to find out how your neighbors (within 30 miles) feel about this topic/person.
3. Follow my example code and use `nrc` lexicons. Conduct a naive sentiment analysis with the Twitter data that you download.
4. Is the result consistent with what you think before this analysis?

Question 3.[Learn How to Calculate Value at Risk (VaR)]

1. Go to Yahoo Finance and download the historical stock prices data for Google. In what follows, we will focus on the adjusted prices (the 6th column of the data)
2. Construct a new variable, daily returns (based on the formulae $\frac{(P_t - P_{t-1})}{P_{t-1}}$).
3. Based on the distribution for the daily returns, calculate the VaR (value at risk) for daily returns at 95 percent confidence level.
4. Based on your results in (3), calculate the VaR in the next day for $10,000 at 95 confidence level. Interpret the value that you obtain.

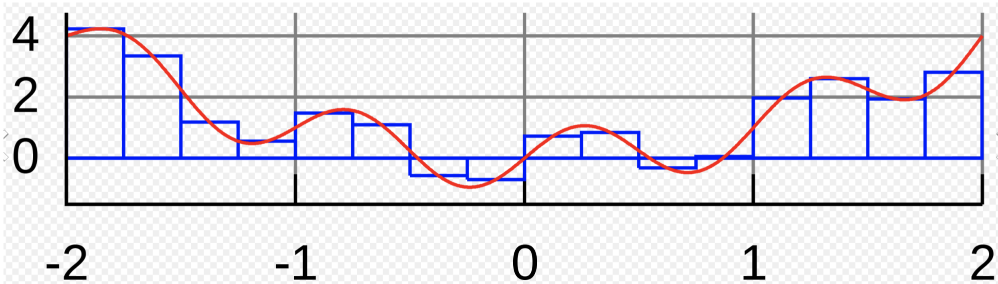

Figure 1: Rectangle Rule for Simple Numerical Integration

5. Cacculate the VaR in the next day for $10,000 at 99 percent confidence level. Interpret the VaR that you obtain.
6. Suppose that your client does not tell you what risk he or she is willing to tolerate. What you can do is to calculate the VaR at every confidence level, say, from 1 to 99 percent confidence level and plot it. This is called **cumulative risk profile**.
7. Suppose that there was a fundamental change in Google's operations in 2012 (change of the CEO perhaps), which in turn led to a systematic change in the distribution of daily returns. So, we will focus on the post-2012 data. Again, repeat the analysis above for this subset of the data and report the VaR in the next day for $10,000 at 95 confidence level. Interpret the value that you obtain.
8. Calculate the VaR in the next **week** for $10,000 at 95 confidence level. Interpret the value that you obtain. Use two different approaches
    a. Use the built-in command in `R`
    b. Use the T rule discussed in class.

Question 4. [Simple Numerical Integration]

Let's perform **numerical integration** here, the opposite operation of numerical differentiation. The idea is to approximate the area under a function by the sum of a series of rectangles given by

$$\int_a^b f(x_i)dx \approx \sum_{i=1}^{K} f(x_i)dx$$

where $x_1, x_2, x_3, \ldots, x_K$ are the right end points for K equally partitioned subintervals of $[a, b]$. The length of each subinterval is equal to dx. Lets illustrate this idea using a graph first (Figure 1).

Lets do it for the following example

$$\int_0^1 x^3 dx = .25$$

Below I will be walking you through this process. First, lets define $dx = 0.001$

```
dx <-
```

Second, lets use the command `seq()` to generate a set of x values from 0 to 1 with an increament of 0.001.

```
x <- seq()
```

Third, lets calculate $f(x) = x^3$ for every value

```
f.x <-
```

Finally, lets calculate the sum of $f(x) \cdot dx$ (if your answer is correct, it should be very close to .25)

```
sum <-
sum
```