Problem Set 1

Nikhil Guruji

April 11, 2018

Names of students I discussed with: Aman Jain, Nakul Thakare

Question 1

Construct the value-weighted market return using CRSP data, replicating the market return time series available in Kenneth French website. Also calculate the equal-weighted market return, and the lagged total market capitalization. Your output should be from January 1926 to December 2017, at a monthly frequency.

Before calculating the portfolio time series, I conduct a series of data cleaning as part of my PS Q1 function. Next, I describe my data cleaning process and their respective assumptions:

1. **Universe of Stocks**: Following Ken French procedure, I restrict the sample to common shares (share codes 10 and 11) and to securities traded in the New York Stock Echange, American Stock Exchange, or the Nasdaq Stock Exchange (exchange codes 1, 2, and 3).
2. **Missing Returns**: I use the rule given on slide 29/110 in the Lecture_1_2018_MFE.pdf. I take $r_{i,t}$ as:

   $r^h_{i,t}$ if $r^d_{i,t}$ is missing,

   $r^d_{i,t}$ if $r^h_{i,t}$ is missing, and

   $(1 + r^h_{i,t})(1 + r^d_{i,t}) - 1$ if both are not missing.

   If both are missing, I ignore all the data of that day for that stock.
3. **Delisting Return Calculation**: The delisted returns are taken from the CRSP website
4. **Market Capitalization Calculation**: I calculated the market capitalization on every day for every stock by multiplying the price (PRC) by the number of shares outstanding (SHROUT). To deal with negative values in PRC or SHROUT, I took the absolute value of the respective PRC or SHROUT values.
5. **Portfolio Weights**: I calculated the portfolio weights using data table library in R. I followed the following steps:

i) Calculated the lagged market capitalization (because value weights change the next month based on a given month's market capitalization).

ii) To calculate weights, for each day, I calculated the ratio of market capitalization (now lagged) for each share and the total market capitalization (also lagged) for that month

iii) I multiplied the weight of each stock on a given day with it's total return $r_{i,t}$ calculated above and summed this product for all stocks to get the market return on a particular date.

6. **Sample Period**: I chose the sample period from January 1926 to December 2017 as given in the Question.
7. **Data used**: I used the following data from CRSP>Annual Update>Stock>Security Files>CRSP Monthly Stock:

i) PERMNO

ii) date (format: YYYYMMDD) (downloaded as integer format)

iii) SHRCD

iv) EXCHCD

v) RET

vi) DLRET

vii) PRC

viii) SHROUT

8. **Libraries**: I use the R libraries data.table, xts, zoo, moments, and lubridate

I store the output in the variable "Monthly_CRSP_Stocks" which has 1104 rows (corresponding to each month from January 1926 to December 2017) and 7 columns (date, Stock_Vw_Ret, Stock_Ew_Ret, Stock_lag_MV, Year, Month, and YM)

Note: While running the code, please ignore the warnings that NA's are generated. The NA's have already been dealt with.

Question 2

Using the risk-free rate of return from French's website3, report the following moments of the market excess returns for both time series (4 decimal digits): annualized return, annualized volatility, annualized Sharpe ratio, skewness, and excess kurtosis. Annualized values should be calculated geometrically. You should be comparing between July 1926 to December 2017, at a monthly frequency.

The summary statistics are in Table 1 below. I report the following five statistics: annualized mean, annualized standard deviation, annualized sharpe ratio, skewness, and excess kurtosis. In Column 1, I report the statistics for the value-weighted market portfolio of stocks from Ken French's website. In Column 2, I report the statistics for the replicated value-weighted market portfolio of stocks calculated in the previous question.

Table 1: Summary Statistics

	Fama French's(1)	Replication (2)
Annualized Mean Excess Returns	0.08255045	0.08279037
Annualized Standard Deviation	0.200563368	0.20040477
Annualized Sharpe Ratio	0.41159222	0.41311576
Skewness	0.18788515	0.19379316
Excess Kurtosis	7.88145193	7.88016368

From question, we have a times series of value-weighted market returns, namely $\{r_t\}_{t=1}^{T}$ Let the market return from French's website be given by $\{r_t^F\}_{t=1}^{T}$. I compute these statistics as follows:

1. **Sample Period**: Monthly from July 1926 to December 2017.
2. **Skewness**: I calculate the skewness of r_t from the monthly time series directly (no annualization, and no logs) using the full sample through the "skewness()" function in the "moments" library.
3. **Excess Kurtosis**: I calculate the excess kurtosis (kurtosis -3) from the monthly time series directly (no annualization, and no logs) using the full sample through the "kurtosis()" function in the "moments" library.

4. **Annualized Mean**: Using the monthly values of $r_t - r_t^f$ (where r_t^f is the risk-free rate at time t) in decimal form, I do the following calculation:

$$\mu_{monthly}^{excess} = \frac{1}{T}\sum_{t=1}^{T}(r_t - r_t^f)$$
$$\mu_{annualized}^{excess} = (1 + \mu_{monthly})^{12} - 1$$

5. **Annualized Standard Deviation**: Using the monthly values of $r_t - r_t^f$ in the decimal form, I calculate the standard deviation $\sigma_{monthly}$ using the inbuilt "sd()" function in R, and then do the following:

$$\sigma_{annualized} = \sqrt{(\sigma_{monthly}^2 + (1 + \mu_{monthly}^{excess})^2)^{12} - (1 + \mu_{monthly}^{excess})^{24}}$$

However, for the sake of simplicity, one could also use:

$$\sigma_{annualized} = \sigma_{monthly}\sqrt{12}$$

But that would be incorrect (underestimating the standard deviation)

6. **Annualized Sharpe Ratio**: I calculate the Annualized Sharpe ratio by dividing the Annualized mean by Annualized Standard deviation

$$SR_{annualized} = \frac{\mu_{annualized}}{\sigma_{annualized}}$$

The moments are very similar. The difference is **not** because of the sample dates because an inner join is done which selects data in the common dates.

Question 3

Report (up to 8 decimal digits) the correlation between your time series and French's time series, and the maximum absolute difference between the two time series. It is zero? If not, justify whether the difference is economically negligible or not. What are the reasons a nonzero difference? You should be comparing between July 1926 to December 2017, at a monthly frequency.

In Table 2 below , I report the time-series correlation between the replicated value-weighted market portfolio of stocks and the value-weighted market portfolio of stocks from Ken French's website. I also report the maximum absolute difference between the two series. I limit the sample to be between July 1926 and December 2017 as the data from Ken French's website starts from July 1926.

It can be observed that The difference between the replicated portfolio and the one from French's website is not zero. We are talking about monthly returns. So, 0.2% per month difference is, economically speaking, certainly not negligible, but also not too much. (Compared to the monthly returns, which are, on an average around 8%)

This is because there were some problems with the data from CRSP such as negative prices, missing returns, missing market capitalization, etc which were removed. Hence, the replication is not perfect. (correlation is not 1 and the moments are not exactly the same)

Table 2: Correlation and Maximum absolute difference

Correlation	Maximum Absolute Difference
0.99999409	0.00206886