Problem Set 1

Name: Xiyuan Dong
Names of whom I discussed this problem set with: Liyao Wei

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 2018, 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 Exchange, American Stock Exchange, or the Nasdaq Stock Exchange (exchange codes 1, 2, and 3).

2. **Missing returns:**
   If both the delisting return (DLRET) and the holding period return (RET) is NA, then delete such record.

3. **Delisting return calculation:**
   For most records, it only has RET, if there is only RET, we let return = RET; if there is only DLRET, we let return = DLRET; if there are both RET and DLRE, we let return = (1+RET)*(1+DLRET)-1

4. **Market Capitalization calculation:**
   First, we let price (PRC) = |PRC| as there is negative value in the raw data and only the magnitude matters. Then, we let market cap (CAP) = PRC * shares outstanding (SHROUT). After then, by stocks, we also acquire the one-month lag CAP (CAP of the stock of the previous month), if the lagged CAP is NA, we drop the record.

5. **Portfolio weights:**
   For Equally-valued return, we assign each stock with the same weight, that is 1/ the

number of stocks in that month. For Value-valued return, we weight the stock by their lagged CAP, that is lag_CAP_i/ sum(lag_CAP)

6. **Sample period:**
   1926.07– 2018.12 as required.

7. **Definition of portfolio weights:**
   As defined in 5.

The first 23 records are displayed as follows:

Table 1: Record Sample

Year	Month	Stock_lag_MV	Stock_Ew_Ret	Stock_Vw_Ret
1926	2	27.01723	-0.05431	-0.034
1926	3	26.12723	-0.09646	-0.06496
1926	4	24.32352	0.032343	0.036971
1926	5	25.26952	0.001966	0.012262
1926	6	25.49971	0.051031	0.054579
1926	7	26.44901	0.013033	0.031813
1926	8	27.31789	0.031021	0.028912
1926	9	28.34824	-0.00645	0.00591
1926	10	28.15444	-0.03484	-0.02924
1926	11	27.25417	0.024461	0.028441
1926	12	28.27267	0.024367	0.028981
1927	1	29.60879	0.012013	0.001899
1927	2	29.74728	0.051842	0.044297
1927	3	31.08431	-0.01934	0.004259
1927	4	31.12537	0.007589	0.007143
1927	5	31.34812	0.062324	0.057453
1927	6	33.43319	-0.01965	-0.0208
1927	7	32.61392	0.05858	0.075563
1927	8	34.75348	0.001751	0.022411
1927	9	35.71698	0.03155	0.049726
1927	10	38.80914	-0.03818	-0.04057
1927	11	37.22267	0.094329	0.067341
1927	12	39.33743	0.030062	0.02311

Question 2

Using the risk-free rate of return from French's website, 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 2018, at a monthly frequency.

The summary statistics are in Table 2 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 replicated value-weighted market portfolio of stocks calculated in the previous question. In Column 2, I report the statistics for the value-weighted market portfolio of stocks from Ken French's website.

Table 2: Summary statistics

	Replication (1)	French's (2)
Annualized Mean	0.0783	0.0781
Annualized Standard Deviation	0.1846	0.1848
Annualized Sharpe Ratio	0.4242	0.4226
Excess Skewness	0.1899	0.184
Kurtosis	7.8297	7.831

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

1. **Sample period:** Monthly from July 1926 to December 2018.
2. **Excess Skewness**: I calculate excess skewness of r_t from the monthly time series directly (no annualization, and no logs) using the full sample.
3. **Kurtosis** I calculate kurtosis from the monthly time series directly (no annualization, and no logs) using the full sample.

4. **Annualized Mean:** I take the mean of monthly return and multiply it by 12. (APR)
5. **Annualized Standard Deviation:** I take the standard deviation of monthly return and multiply it by sqrt(12).
6. **Sharpe Ratio:** Annualized Mean/ Annualized Standard Deviation

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 2018, at a monthly frequency.

In Table 3 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 difference between the two series.

I limit the sample to be between July 1926 and December 2018.

The difference between the replicated portfolio and the one from French's website is almost zero as we can see that the correlation is very close to 0 and the maximum absolute difference is only 0.002, which mean my replication methodology is correct.

Table 3: Correlation and maximum difference

1. Correlation	0.99999435
2. Maximum absolute difference	0.00193910