Problem Set 4

Name: Minqi Tan

Question 1

Before calculating the portfolio time series, I conduct a series of data cleaning as part of my PS4_Q1 function. Next, I describe my data cleaning process and their respective assumptions:

1. **Procedure and Decisions:**
   I follow Ken French's procedure in selecting data sets, type of firms, period, and in calculating book equity and market equity. I also made some decisions on my own in details such as filling missing prices and filling the first lag in market equity.
2. **Data Sets:**
   CRSP monthly equity data (PERMNO, PERMCO, RET, SHRCD, EXCHCD, SHROUT, PRC)
   Compustat Fundamental Annual
   (gvkey,fyear,indfmt,datafmt,SEQ,CEQ,PSTK,AT,LT,MIB,DT,TXDITC,ITCB,TXDB,PSTKRV,PSTKL)
   Compustat Pension Annual (gvkey, PRBA)
   CRSP Compustat Linking Table(gvkey,lpermco/no,linkprim,linktype,linkdt,linkenddt
   Historical Book Equity data (Ken French's Data library)

3. **CRSP Clean Up**:
   1. Select EXCHCD(1,2,3) and SHRCD(10,11)
   2. RET/DLRET: Convert all values of "A", "B", "C", "-99", "-88", "-77", etc. to NA.
   3. Holding period return = (1+RET) * (1+DLRET) – 1. Missing Returns are omitted.
   4. If there is a gap in PRC, fill the gap with previous PRC
   5. Market cap=absolute value of price (PRC)*shares outstanding/1000 (Then sum up per PERMCO)
   6. Lagged Market cap is the one-month lag of the Market Cap. For first period with no lag, assume lagged market cap = market cap/(1+firm arithmetic average return)
   7. We also check the date to make sure we only include the true lags.
4. **Compustat Clean Up:**
   1. Select industry format only "INDL" to take out financial firms. Select only USD currency and STD Data format. Same goes for Compustat Pension data.
   2. Merge Compustat with Compustat Pension by gvkey and date
   3. Calculate BE: follow Ken French exactly as described in the problem set.

5. **Merge CRSP with Compustat**

1. First, merge CRSP with Linktable by PERMCO
2. select only the PERMCO whose date is within the link start and end date range. For PERMCO that are active, the link end date show 'E' , replace it with NA.
3. Filter on variable LINKTYPE, if there are a few types on PERMCO and the type is not LC, only keep LC.
4. Filter on variable LINKPRIM, only keep 'P' if there are different values for LINKPRIM on the same PERMCO.
5. Filter on variable LIID , only keep '1' if there are different values for LIID on the same PERMCO.
6. Filter on the variable LINKENDDT, use the link that's current
7. Find the difference between the start and end date of the link, use the link that's been around the longest.
8. Choose the gvkey that has been around the longest.
9. Use the smaller gvkey if there are multiple with same PERMCO on the same date.
10. Now we can finally merge CRSP with Compustat, by gvkey and year.

6. **Construct Portfolio formed on ME(size):**
   1. As per the 1992 paper, the size deciles are calculated based on the rankings measured annually for returns from July of year t to June of year t+1, which is based on the market cap for each company in June of year t.
   2. All companies are allocated a rank between 1 to 10, using breakpoints based on the ME in June of year t of only NYSE stocks.
   3. Compute value weighted returns in each decile each month, weight is lagged ME.
   4. Restrict the data from 1973.01 to 2019.12
   5. Create a long short portfolio by going long on the 1st decile and shorting the 10th decile

7. **Construct Portfolio formed on BE-ME (HML):**
   1. Use historical BE data from Ken French Data Library, and fill in any NAs using PERMNO link.
   2. BE/ME portfolio is calculated by matching Market cap and book equity at December of year t-1 with return from July of year t to June of year t + 1. Use only returns on December for decile sorting.
   3. All companies are allocated a rank between 1 to 10, using breakpoints based on the BE/ME in December of year t of only NYSE stocks.
   4. Compute value weighted returns in each decile each month, weight is lagged ME.
   5. Restrict the data from 1973.01 to 2019.12
   6. Create a long short portfolio by going long on the 1st decile and shorting the 10th decile

8. **Construct SMB and HML factors:**

1. Cut the portfolio into 2 sub group: small (size decile <=5) and big (size decile>=6)
2. Cut the portfolio into 3 sub group: value(hml decile8-10), growth (1-3) and neutral(4-7)
3. Now we can construct the factors based on below formulas:
   SMB = (Small Value + Small Neutral + Small Growth - Big Value - Big Neutral – Big Growth)* 1/3
   HML = (Small Value + Big Value - Small Growth - Big Growth)* ½
4. For each of the 6 sub group, calculate value weighted returns, weight is lagged market cap.
5. Restrict data from 1973.01 to 2019.12

Question 2

Calculations:
1.Annualized excess returns: arithmetic mean(excess of risk free) * 12
2.Annualized volatility: monthly volatility * sqrt(12)
3.Sharp ratio: step1/step2
4.Skewness/Correlation: use R functions on monthly returns
Output:

	1	2	3	4	5	6	7	8	9	10	LongShort
Excess Return	0.08281438	0.08588751	0.09504779	0.08536991	0.08824591	0.08508852	0.08580319	0.0830432	0.0767202	0.06176723	0.02111438
Standard Deviation	0.21641385	0.22129662	0.20673444	0.20198079	0.19659411	0.18273947	0.18121112	0.1748086	0.1618037	0.14959037	0.16296269
Sharpe Ratio	0.38266674	0.38811037	0.45975786	0.42266349	0.44887359	0.46562749	0.47349849	0.4750522	0.4741561	0.41290911	0.12956575
Skewness	-0.18021656	-0.25187874	-0.50923636	-0.53922098	-0.49533305	-0.58518434	-0.52228464	-0.4517200	-0.4887848	-0.34532002	0.83605770
Correlation	0.99819477	0.99817103	0.99831360	0.99804231	0.99709231	0.99720134	0.99723234	0.9971007	0.9971840	0.99846764	0.99654521

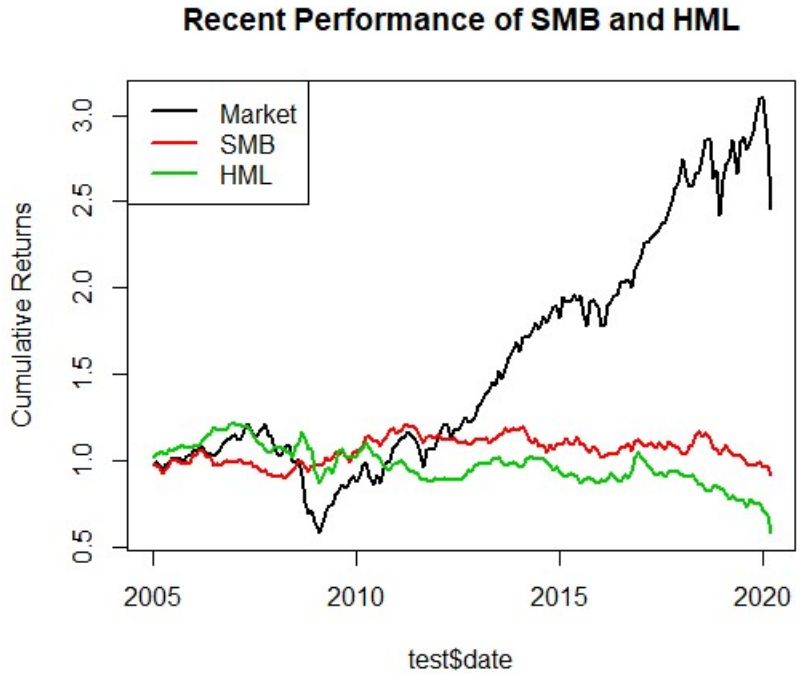
Question 3

Calculations: same as in question 2.
Output:

	1	2	3	4	5	6	7	8	9	10	LongShort
Excess Return	0.06826883	0.08045981	0.08873855	0.0908385	0.0758920	0.08156006	0.0907583	0.08185724	0.09880738	0.11678928	0.05698044
Standard Deviation	0.18162711	0.16368150	0.16061487	0.1676003	0.1600510	0.15743885	0.1544615	0.16142861	0.16848206	0.20599704	0.15742973
Sharpe Ratio	0.37217620	0.48836079	0.54977404	0.5393588	0.4706992	0.51490674	0.5850343	0.50395697	0.58424152	0.56526992	0.36245800
Skewness	-0.22549671	-0.52968184	-0.46109134	-0.4369801	-0.4719726	-0.46891908	-0.1682238	-0.49401516	-0.47233094	0.09869599	0.78198000
Correlation	0.99711024	0.97827867	0.97647941	0.9604875	0.9716558	0.95499010	0.9607355	0.96271952	0.96765115	0.96393452	0.96012440

Question 4

Let's examine how size and value have performed in the past 15 years.
We can plot SMB and HML's cumulative returns over the period, using Ken French's factor data. As we can see in below plot, both size and value have underperformed against the market by a big margin.



Question 5

Calculation: same as in question 2 and 3
Output:

	SMB
Excess Return	-0.03468405
Standard Deviation	0.11110581
Sharpe Ratio	-0.31217136
Skewness	0.45895441
Correlation	0.98024683

	HML
Excess Return	-0.02657963
Standard Deviation	0.10738569
Sharpe Ratio	-0.24751553
Skewness	-0.05518426
Correlation	0.95047953

Question 6

Characteristic portfolios

The characteristic portfolios (1992) aimed at showing that the two easily measured variables, size (ME) and book-to-market equity (BE/ME) can provide a powerful explanation of the cross-sectional stock returns. The paper proves that these two variables together capture the cross-sectional variation in stock returns associated with size, book-to-market and E/P and leverage. The characteristic portfolio is based on the Fama Macbeth regression**.**

Factor Portfolios

The factor portfolios(1993) on the other hand shows that the mimicking portfolios on size and book-to-market equity can be used in a time-series regression to find factor loadings. The paper proves that these mimicking factors, separately or together, capture strong common variation in returns, therefore, are essential risk factors. The time series regression also indicates these factors alone can't explain the average return on stocks as the market risk factor also plays a big role.