

# QAM\_HW4

## Question 1.

To construct the size and BM portfolios several data cleaning processes steps have been done and is listed below:

1. Sample Period: The sample period considered for the replication strategy is for the period of January 1973 - December 2019.
1. Universe of Stocks: The sample has been restricted to only include common shares (SHRCD = 10 and 11) and will also be restricted to securities traded on the New York Stock Exchange, American Stock Exchange, and NASDAQ (EXCHCD = 1, 2, and 3). Financial trading firms have been excluded from the universe of stocks as defined using the KRF SICCODE industry 47 due to their high leverage that is unusual for all the other firms.
2. Missing Returns: The data contains missing returns specifically for RET with the characters (E, D, C, B, A), are treated as they are missing and will be considered to have 0 return. The characters represent no listing information, no valid previous prices, etc. The universe of stocks that have missing returns are quite small and will be assumed to have negligible impact when replicating the value-weighted returns.
3. Delisting return calculation: Similarly, the delisting returns could also be missing and DLRET with the characters (S, T, A, P) will also be considered as missing and will be assumed to have 0 return.
4. Total return calculation: To calculate the total monthly return for each stock I will consider both returns (with dividends) and total delisting returns (with dividends). More specifically, if both the returns and delisting-returns are not missing then I will add both returns in the calculation to calculate the total return, if the returns are not missing and the delisting returns are missing I will only use the returns, and finally if the returns are missing and the delisting returns are not, I will only use the delisting returns.
5. Validation Checks: To ensure that the previous record for each PERMNO is the actual 1 month lagged data, I have created a valid flag variable that checks whether or not the previous record for that PERMNO is the previous lagged month. Furthermore, I have only included records with valid previous lagged month, non-missing lagged market capitalization (since it will be used to calculate the value weighted returns) and non-missing returns.
6. Joining with CCM LINK: Before joining the CRSP data to the COMPUSTAT data to get financial statement data, I have joined the CRSP dataset to the CCM link table. When joining the CRSP dataset to the CCM link table I have joined only when the link type is LU/LC meaning that the link is complete or keeping the reliable links only and that I am only considering the primary links - P/C. Furthermore, the linking data must be valid between the linking start date and the linking end date. After joining the CRSP dataset with the CCM link dataset, I can now have a GVKEY and IID pair for each PERMNO and since each GVKEY and IID Pair is unique to each PERMNO (under the conditions stated above - primary link, etc.), I do not need to clean any duplicated data when joining the CRSP dataset with the CCM link dataset.
7. COMPUSTAT dataset: To get the Book Value Equity for each firm to create the Book-To-Market variable, COMPUSTAT data is needed. The annual fundamentals COMPUSTAT data will be used to get the BE for each firm furthermore, pensions data is obtained from ACO\_PNFNDA dataset. To create the BE variable I have done the following process: First, the shareholder's equity - total is considered (SEQ), if it is not available common/ordinary equity will be used (CEQ) plus the preferred stock capital - total (PSTK). If they're not available the total assets minus total liabilities minus minority interest (AT - LT - MIB) will be used. If they're not available AT - LT will be used. To create the deferred taxes and investment tax credit (DT), I have used deferred taxes and investment tax credit (TXDITC). If it is not available the investment tax credit (ITCB) plus deferred taxes (TXDB) will be used. To create the book value of preferred stock (PS), I have used the preferred stock redemption value (PSTKRV). If it is not available the preferred stock liquidating value will be used (PSTKL). If it is not available the par value - preferred stock capital (PSTK) will be used instead. Finally the BE is defined as  $SHE - PS + DT - PRBA$  where PRBA is the postretirement benefit asset from the COMPUSTAT pension annual data (ACO\_PNFNDA) and is joined to the original COMPUSTAT data using GVKEY.
8. Combining CRSP and COMPUSTAT: after creating the

BE variable in the COMPUSTAT dataset, we can now join the CRSP dataset (that has been joined with the CCM link table) to the COMPUSTAT dataset. Since when creating the BM portfolio we will use the BE used for the last fiscal year end for last year for the portfolio formation to that year (will be expanded further below under stock portfolio formation section), I will create a fiscal year variable based on the datadate and add 1 year to the fiscal year variable to ensure that the last fiscal year end is in t-1 of the portfolio formation at year t. Finally, for each fiscal year there might be some multiple datadates available (could be due to the change in accounting methodology reporting) and I will use the latest datadate available for that fiscal year in that case. After doing so, I will join the CRSP dataset with the COMPUSTAT dataset based on the fiscal year, GVKEY and IID. The resulting final dataset will have all the CRSP dataset (PERMNO level) with the COMPUSTAT fundamentals joined only for July (since portfolios are formed at the end of June of each year).

6. Stocks Portfolio Formation: After joining the CRSP and COMPUSTAT dataset I will now summarize the PERMNO level dataset to PERMCO level, hence, I will sum the BE and ME for each PERMNO that the PERMCO has and calculate the value-weighted return for each PERMCO level. Hence, the dataset now is unique for each PERMCO and per month. After doing so, to create the size ranked portfolio, I will use the lagged market cap data on month 7 of year t to rank the stocks for that year and the rank/decile will be kept constant for that portfolio until month 6 of year t + 1 (annual rebalancing). Hence, the portfolio will then be rebalanced on month 7 of year t + 1 and the lagged market cap data on that month (market cap data of June of year t + 1). Similarly for the BM portfolio, the portfolio will be formed at the end of June of year t, and I will be using the BE data obtained from the fiscal year that ended on the previous year (t - 1) whereas the ME data used will be the ME for December of year t-1 for the portfolio formation of month 7 of year t. Hence, when forming the BM portfolio for of month 7 of year t, the market cap data on December of year t-1 must not be missing and the lagged market cap variable for month 7 should not be missing. Furthermore, when creating the BM deciles the firms with negative BMs ( $BM < 0$ ) will not be considered, and the rest of the firms will be considered to create the BM deciles. Finally, when creating the deciles only the NYSE firms are used to create the breakpoints (for both the size and BM portfolio), hence, for each decile, the portfolio has the same number of NYSE firms. Furthermore, when creating the portfolio deciles, I will create breakpoints according to the size/BM rankcalculated and I will keep the right interval of the breakpoints open, meaning that firms located at exactly the boundary will be considered to be included in the next decile portfolio. All of the conditions above must be satisfied to be included in the portfolio.

7. Portfolio returns: Finally for each portfolio formed based on each decile (both size and BM), the value-weighted returns are calculated for each portfolio (decile) based on the lagged market cap to avoid forward bias. Furthermore, when calculating the HML and SMB factor the the size decile used will be the 0-50 percentile and 50-100 percentile. Whereas, the BM deciles used will be split into 3 different groups 0-30 percentile, 30-70 percentile and 70-100 percentile. To calculate the HML factor the following formula is used:  $HML = 0.5(SmallValue + LargeValue) - 0.5(SmallGrowth + LargeGrowth)$ . Whereas to calculate the SMB factor the following formula is used instead  $SMB = 1/3(SmallValue + SmallNeutral + SmallGrowth) - 1/3(LargeValue + LargeNeutral + LargeGrowth)$  \$.

The sample resulting dataset after doing all the data processing is shown in the table below:

Year	Month	decile	size_ret	bm_ret	Rf	smb_ret	hml_ret
2019	1	1	0.10269	0.08658	0.0021	0.02803	-0.01483
2019	1	2	0.09891	0.08814	0.0021	0.02803	-0.01483
2019	1	3	0.10174	0.06947	0.0021	0.02803	-0.01483
2019	1	4	0.10434	0.06870	0.0021	0.02803	-0.01483
2019	1	5	0.13155	0.07797	0.0021	0.02803	-0.01483
2019	1	6	0.11796	0.10501	0.0021	0.02803	-0.01483
2019	1	7	0.12676	0.07111	0.0021	0.02803	-0.01483
2019	1	8	0.10119	0.11473	0.0021	0.02803	-0.01483
2019	1	9	0.10188	0.11079	0.0021	0.02803	-0.01483
2019	1	10	0.07367	0.05398	0.0021	0.02803	-0.01483

## Question 2. Size Decile

For the size decile the annualized average excess returns, volatility, SR and skewness is shown below:

	Mean Excess Return	Standard Deviation	Sharpe Ratio	Skewness(m)
Decile 1	0.08248	0.22271	0.37033	-0.65599
Decile 2	0.08667	0.22908	0.37834	-0.72554
Decile 3	0.09189	0.21489	0.42763	-0.93303
Decile 4	0.08493	0.20799	0.40837	-0.90107
Decile 5	0.08866	0.19986	0.44359	-0.93794
Decile 6	0.08207	0.18511	0.44337	-0.91449
Decile 7	0.08552	0.18347	0.46614	-0.80929
Decile 8	0.08202	0.17640	0.46493	-0.80643
Decile 9	0.07541	0.16160	0.46668	-0.76853
Decile 10	0.06070	0.14946	0.40613	-0.60025
WML	-0.06709	0.16717	-0.40134	-1.53157

The correlation between the replicated portfolio and the one obtain from the KRF's website is shown below for each decile:

	Portfolio
Decile 1	0.99722
Decile 2	0.99695
Decile 3	0.99629
Decile 4	0.99623
Decile 5	0.99580
Decile 6	0.99577
Decile 7	0.99520
Decile 8	0.99541
Decile 9	0.99494
Decile 10	0.99633
WML	0.99299

As seen from the above table, the correlation is very close to 1 and the replicated portfolio is very close to the Size Portfolio from the actual KRF.

### Question 3. B/M Decile

For the B/M decile the annualized average excess returns, volatility, SR and skewness is shown below:

	Mean Excess Return	Standard Deviation	Sharpe Ratio	Skewness(m)
Decile 1	0.05772	0.17858	0.32320	-0.52082
Decile 2	0.07756	0.16316	0.47534	-0.73842
Decile 3	0.08259	0.16396	0.50375	-0.96569
Decile 4	0.07884	0.16606	0.47473	-0.76083
Decile 5	0.07582	0.15786	0.48028	-0.77110
Decile 6	0.07009	0.16264	0.43096	-0.79769
Decile 7	0.08236	0.15334	0.53714	-0.51667
Decile 8	0.08962	0.16005	0.55995	-0.86313
Decile 9	0.09015	0.16321	0.55234	-0.68454
Decile 10	0.11868	0.19197	0.61824	-0.51249
WML	0.01565	0.15522	0.10084	0.34139

The correlation between the replicated portfolio and the one obtain from the KRF's website is shown below for each decile:

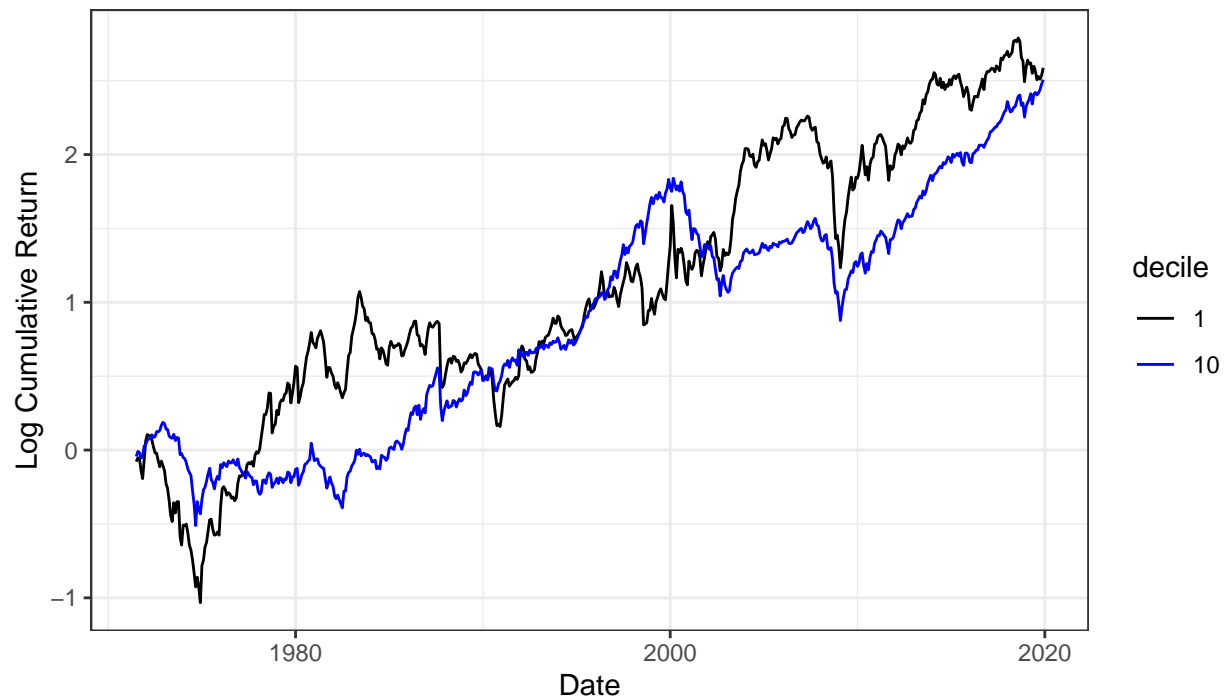
Similarly, the correlation is pretty close to 1, although for some of the deciles the correlation is a lot less.

	Portfolio
Decile 1	0.99483
Decile 2	0.98294
Decile 3	0.97021
Decile 4	0.97088
Decile 5	0.96409
Decile 6	0.92600
Decile 7	0.95463
Decile 8	0.96360
Decile 9	0.94788
Decile 10	0.93960
WML	0.89191

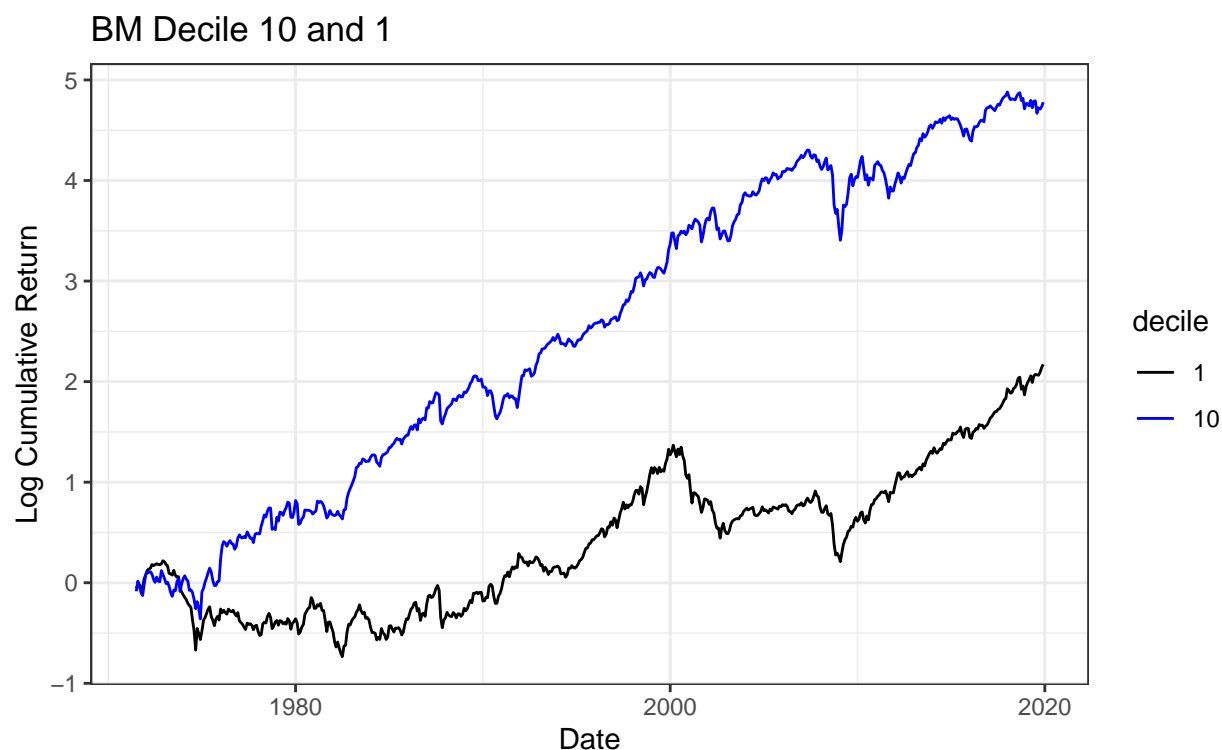
This could be caused by the treatment of the boundary ranking discussed above as well as that I have not used Davis, Fama and French (2000)'s data to fill in missing historical book equity value.

#### Question 4. Value and Size Anomaly

##### Size Decile 10 and 1



Looking at the plot above, the size anomaly has not been working very well for the past few years as the spread between the decile 10 portfolio and decile 1 portfolio is almost non-existent. Furthermore, the WML portfolio also exhibits negative skewness as seen from the table above on the size return for the sample period considered.



Looking at the plot above, the value anomaly is still working pretty well for the past few years. The BM decile 10 has been consistently outperforming the BM decile 1 and thus doing a WML portfolio will result in positive average returns. Furthermore, as seen from the table above of the summary statistics the WML portfolio of using BM has positive skewness.

#### Question 5. HML SMB Portfolios

For the SMB portfolios the annualized average excess returns, volatility, SR and skewness is shown below:

##	Mean Excess Return	Standard Deviation	Sharpe Ratio	Skewness(m)
##	0.01749418	0.10889279	0.16065509	0.28661025

and for the HML portfolio:

##	Mean Excess Return	Standard Deviation	Sharpe Ratio	Skewness(m)
##	0.02383274	0.07558353	0.31531663	0.04616459

The correlation between the replicated SMB factor with the one from KRF website is :

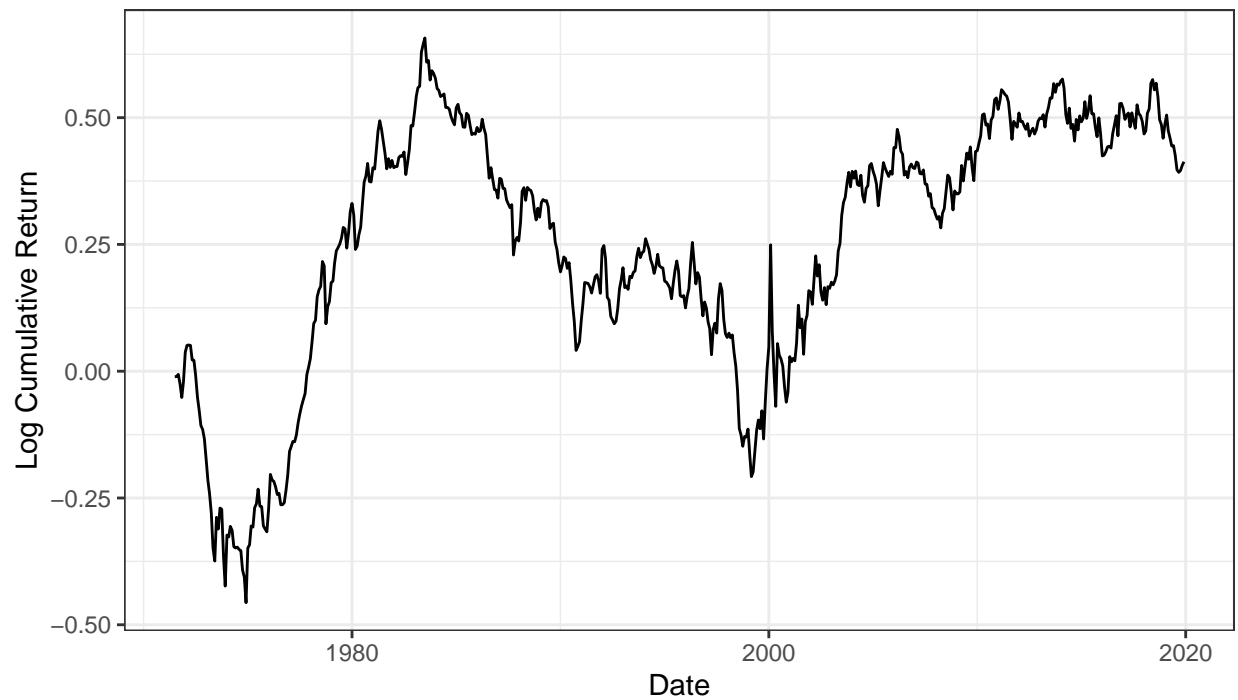
```
## [1] 0.9890475
```

whereas the correlation between the replicated HML factor with the one from KRF website is :

```
## [1] 0.8635701
```

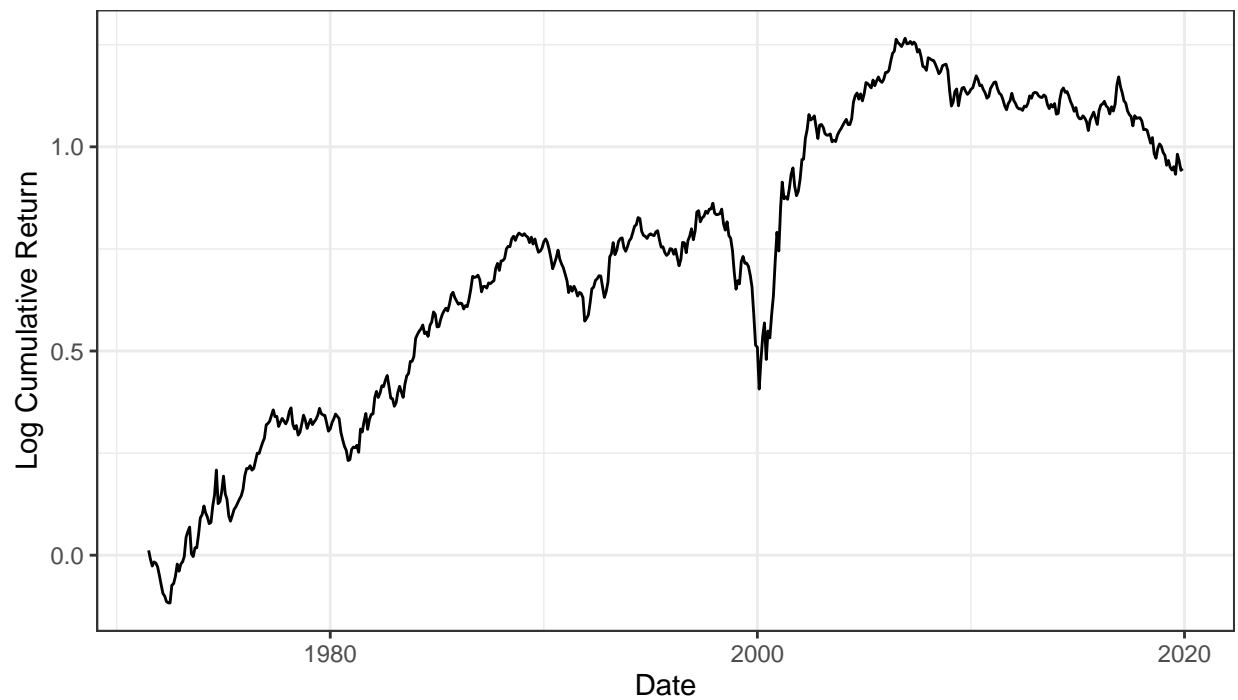
The correlation for the replaction HML factor is a quite bit lower, and this could be caused by the treatment of the boundary ranking discussed above again and that I have not used Davis, Fama and French (2000)'s data to fill in missing historical book equity value.

### SMB Factor Cumulative Returns



The SMB factor has not been consistent across time as of recently as seen from the cumulative plot of the SMB factor returns above.

### HML Factor Cumulative Returns



Whereas, the HML factor been quite consistent across time as can be seen from the cumulative plot of the HML factor returns above. However, the HML factor seems to also be declining by quite a bit as of recently.

**Question 6.**

As seen above, using the factor portfolios instead of the characteristic portfolios result in a lower standard deviation of returns, higher sharpe ratio and lower skewness. The factor portfolios HML and SMB result both have higher sharpe rations than the WML portfolios of either the size/BM portfolio. Similarly, the average excess returns for the factor portfolios are both higher and they also have lower standard deviation, and thus the factor portfolios performed better during the sample period than the WML portfolios of the characteristic portfolios.