Quant Asset Management Problem Set 4

Nitish Ramkumar

Question 1 - Construction of Characteristic and factor portfolios

Introduction

The aim of this question is to prepare the necessary data for subsequent analysis for the size, book-to-market, HML and SMB portfolios. This involves cleaning CRSP and compustat data, merging them, using them to rank the various companies based on size (market equity) and Book-to-Market deciles. This will also involve ranking size and book-to-market in such a way that a HML and SMB portfolio can be formed.

Data

This question will involve 4 major data sources

1. **CRSP**, from where we can get

i. Stock and company identification (PERMNO, PERMCO)
ii. Monthly returns (RET)
iii. Exchange code (EXCHCD), share code (SHRCD) and industry code (NAICS or SICCD, explained later)
iv. Price (PRC) and shares outstanding (SHROUT) This can be retrieved out of the CRSP Monthly stock web portal

2. **COMPUSTAT**, from where we can get the fundamental information about various companies. For our analysis, we will be retrieving

i. the unique identifier (GVKEY)
ii. fiscal year (fyear)
iii. the industry format which will help us identify industrial or financial firms(INDFMT)
iv. data format, which helps us identify (DATAFMT)
v. **Shareholder's equity**(SHE) - Total (SEQ), Common/ordinary equity - total (CEQ), preferred / preference stock (capital) - total (PSTK), Assets total (at), Liabilities total (lt), minority balance (MIB)

vi. **Deferred tax and investment tax credit** (DT) - Deferred Taxes and Investment Tax Credit (TXDITC), Investment Tax Credit (Balance Sheet)" (ITCB), Deferred Taxes (Balance Sheet)" (TXDB)
vii. **Book value of preferred stock**(PS) - Preferred Stock Redemption Value (PSTKRV), Preferred Stock Liquidating Value (PSTKL), Preferred Preference Stock (Capital) - Total (PSTK)

3. **COMPUSTAT PENSION DATA** - We can get the unique identifier (GVKEY) and Post Retirement Benefit Asset (PRBA) from this dataset.

4. **LINK TABLE** - This is retrieved out the CRSP data set CRSPA.CCMXPF_LINKTABLE. This can be retrieved directly out of R API to connect to WRDS. Information retrieved out of this table are GVKEY (Compustat identifier), lpermno (the corresponding link to PERMNO in CRSP), lpermco (the link to PERMCO in CRSP), linkdt (the start date for this particular link), linkenddt (the end date for this particular link), linktype and linkprim(types)

The data has not been constrained to a specific data range as part of input.

Methodology

Clean up CRSP data

The basic clean up of the CRSP data involves the following steps:

1. Filter out stocks with share code of 10,11 (ordinary common shares with no further definitions)
2. Filter out stocks with exchange code of 1,2,3 (only NYSE, AMEX and NASDAQ)
3. Set all returns as NA when returns are equal to -99, -88, -77, -66, -55, -44, B, C
4. Set all delisting returns as NA when it is equal to -99, -88, -77, -66, -55, -44, A, P, S, T.
5. **Market Cap** - Calculate Marketcap for each issue (PERMNO), as the product of the absolute value of price and absolute value of shares outstanding. This will further be accumulated within every company in later steps.
6. **Delist returns** need to incorporated into the returns if a company has been delisted. If a delist return is present, it needs to be set as return (if actual return is missing), or the returns should be geometrically combined *((1+Return)(1+Delist return) - 1)*

Non-Financials
As mentioned in the paper, we are expected to ignore financials due to how leverage is treated differently as compared to non-financial firms.

This can be done using 2 different values in the CRSP data set - North American Industry Classification Code System (NAICS) or the Standard Industrial Code (SICCD). To avoid financial firms, we need to make sure that NAICS doesn't begin with 52, or SICCD doesn't start with 6. Both were tested, and the NAICS check gave better results.

Filling Missing Data
We also need to fill in missing months, as it is critical while trying to make sure correct months are chosen while calculating the characteristic and factor portfolios. By calculating the number of months between 2 entries, we can gauge and fill up missing entries for every PERMCO, PERMNO combination.

Once we do this, for example, we will know that a shift of 6 months will surely take us from January to June.

Clean up and merge Compustat and Pension Annual data

Based on symbols introduced in data section

The basic clean up of the 2 datasets involves the following steps:

1. For compustat data, we need to only entries which have DATAFMT = STD (standard) and INDFMT = INDL (This will only include non-financial firms). For Pension Annual data, we need to restrict it only for non-financial firms as wel (INDFMT = INDL)
2. Merge the computstat and Pension Annual data using the gvkey field which is common between both tables

Calculate Book Equity (BE) Value
Book equity (BE) can be calculated as: *BE = Shareholder's Equity (SHE) - Book Value of Preferred Stock(PS) + Deferred tax and investment tax credit(DT) - PRBA*

, where SHE = SEQ (when SEQ is valid), else CEQ + PSTK (when the sum is valid), else AT-LT-MIB (when MIB is valid) , else AT - LT
PS = PSTKRV (when PSTKRV is valid), else PSTKL (when PSTKL is valid) , else PSTK
DT = TXDITC (where TXDITC is valid), else ITCB + TXDB (ignoring the term which is NA)

Merge Process

1. Initially the merge is done between the the compustat linked (compustat data and pension annual) data set and the link table. This will be a merge based on the GVKEY in the compustat data and the GVKEY in the link data. We will be performing a link on a company level with compustat.
2. **Link types** - Before we merge, we can filter certain information out in the link data based on link type.

a. Link types (linktype) LC and LU are the most accurate link types.

b. The other link types are either indicate duplicate(LD) or some sort of information doesn't match our criteria (LX-in exchange not in CRSP world, LS - different gvkey for same PERMCO, LN - compustat doesn't contain valid information, LO - no entries in link table, NR/NU - link not available).

c. We also care only about primary links, so we can restrict the linkprim type in the link table to P(primary link) and C (primary link assigned by CRSP to manage issues due to overlap). The USEDFLAG field is not necessary because after all these conditions, all the values of USEDFLAG is already equal to 1.

3. After we perform a link based on gvkey between the 2 data sets, we also need to make sure that the date in compustat is between the link start and link end date. This will make sure that for a given time, there will be only one link between PERMCO and GVKEY

Size Deciles

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 that company in June of year t. This means the rebalancing of the ranking is done annually using end of June information.

2. All marketcap of a company (for each PERMCO) are accumulated together to get one value for each company. It is this marketcap which is used for the ranking process.

3. All companies are allocated a rank between 1 to 10 from july of year t to july of year t+1, based on breakpoints calculated using the ME in June of year t of only NYSE stocks.

4. Once we get the rankings, we can calculate value-weighted returns for all the companies within each decile, for each year month combination. This will use the lagged market capitalization for that stock. (Note that is is not clearly mentioned whether it is value weighted or equal weighted returns in the paper. Value weighted was chosen due to better results).

5. A long short portfolio is created by going long the 1st decile and short the 10th decile for every year month combination.

6. The data has to be restricted between 1973 and 2015.

Checks

1. While calculating breakpoints at time t, a valid CRSP price (in turn a valid marketcap) should be available in June of year t for it to be considered for ranking. (December market cap check has been ignored for size and has been done only for book-to-market calculations)

Book-to-Market Deciles

As per the 1992 paper, the Book-to-Market deciles are based on the Book equity, which is got out of Compustat and Market equity out of the CRSP data. Steps for getting these decile returns are as follows:

1. All fundamental information provided in compustat is on a company level. So as our Book equity calculation will be on a company level, we should aggregate our market cap for every company for each month and year.

2. The BE/ME ratio is calculated using the Book Equity of end of fiscal year t-1 and Market equity (Marketcap) from december year t-1.
3. The BE/ME value of all NYSE stocks should be used for calculating the decile breakpoints.
4. The rankings are rebalanced every year at the end of june, using data from the year before. So ranking of each stock from year t-1 is used for returns from July of year t to June of year t+1 for that stock.
5. Value weighted returns are calculated for each year month and decile combination using the lagged market capital of the stock.
6. A long short portfolio is created by going long the 10th decile and short the 1st decile for each year month combination.
7. When BE value is NA, we check if there is a valid value for that stock and for that year in the historical BE value file, provided in French website.
8. The data is restricted from 1973 to 2015.

Checks

1. While calculating the breakpoints for year t, there should be valid stock price (which is valid market cap) in december of year t-1.

2. Valid Total book assets, book equity in the compustat data at the end of the fiscal year of t-1.

SMB and HML portfolios

For SMB and HML portfolios, we first create 6 portfolios from sorts of stocks on Market Equity (Size) and BE/ME (Book-to-Market) values. We follow these steps to achieve this:

1. For Book-to-Market, Divide the stocks into 3 groups based on the 30%, 70% and 100% breakpoints of BE/ME values of NYSE stocks. As for the characteristic portfolios, while calculating the rank and return for year t, the BE is based on end of fiscal year t-1, the ME is based on market equity in December of year t-1. The rank calculated using data from t-1 year, is the rank for returns from July year t+1 to June year t (annual rebalancing)

2. For Size, we shall divide stocks into two groups, with the breakpoint being the median of the Market Equity(ME) for stocks in year t-1 (while calculating for year t). The rank calculated using data from t-1 year, is the rank for returns from July year t+1 to June year t. (annual rebalancing).

3. We intersect the these 2 rankings and form 6 portfolios. i.e. the small group of the size rankings along with the low BE/ME group of the BTM rankings will be the S/L portfolio. Lets mark this portfolio 1. Similar we can do the same for all the 6 portfolios and mark them with their corresponding ranks (i.e. 1 to 6).

4. Now value weighted returns can be calculated for each year month combination, for each of the 6 portfolios.

HML
HML is the difference between the simple average of the two high BE/ME portfolios and the simple average

of the two low BE/ME portfolios. i.e.
HML = 1/2(Small Value + Big Value) - 1/2(Small Growth + Big Growth)

SMB
SMB is the difference between the simple average of all 3 small portfolios and the simple average of all 3 big portfolios. i.e.
SMB = 1/3(Small Value + Small Neutral + Small Growth) - 1/3(Big Value + Big Neutral + Big Growth)

Checks

1. While calculating ranks for year t, we include firms having valid CRSP stock prices (CRSP market cap) of december year t-1 and june year t.

2. Firms should also have valid book equity for year t-1.
3. Only positive BE entries are used for calclating the breakpoints for BE/ME.

4. Atleast 2 years of historical data for a company should be available before using it for decile ranking.

Sample outputs

Book-to-market deciles

Table 1: Btm Deciles returns

Year	Month	port	BtM_Ret
2015	12	1	-0.0081500
2015	12	2	-0.0270335
2015	12	3	-0.0130918
2015	12	4	-0.0142675
2015	12	5	-0.0161776
2015	12	6	-0.0278255
2015	12	7	-0.0558537
2015	12	8	-0.0278716
2015	12	9	-0.0283692
2015	12	10	-0.0498722

Size deciles

Table 2: Size Deciles returns

Year	Month	port	Size_Ret
2015	12	1	-0.0423789
2015	12	2	-0.0555958
2015	12	3	-0.0531313
2015	12	4	-0.0444725
2015	12	5	-0.0599477
2015	12	6	-0.0466588
2015	12	7	-0.0408498
2015	12	8	-0.0379329
2015	12	9	-0.0252311
2015	12	10	-0.0140657

HML and SMB returns

Table 3: HMB and SML portfolios

Year	Month	HML_Ret	SMB_Ret
2015	1	-0.0301506	-0.0072510
2015	2	-0.0229677	0.0056028
2015	3	-0.0054109	0.0316064
2015	4	0.0208737	-0.0323244
2015	5	-0.0200252	0.0122583
2015	6	-0.0096247	0.0280128
2015	7	-0.0446932	-0.0412047
2015	8	0.0331380	0.0014555
2015	9	0.0081314	-0.0284787
2015	10	-0.0030796	-0.0209682
2015	11	-0.0102432	0.0362418
2015	12	-0.0215062	-0.0302417

Question 2 - Size Portfolio Correlation

Data

Apart from the results from the previous question (from 1973 to 2015), we also need the Fama French returns divided based on size.

Methodology

The output contains the following information for each decile and the long short portfolio :

1. Arithmetically Annualized excess returns (reduced by risk free rate) of each size decile and the long short portfolio. Arithmetic annualization is done by multiplying by 12.

2. Arithmetically annualized standard deviation of each size decile and the long short portfolio. The annualization is done by multiplying by $\sqrt{(}12)$

3. Annualized Sharpe Ratio, which is calculated by dividing the excess annualized returns and the annualized sharpe ratio.

4. Skewness, which is calculated on the actual returns over the time period.

5. Correlation, which is formulated by comparing the returns which were calculated and the returns posted on the fama french website.

Note: The Fama French breakpoints according to the excel sheet from the website, mentions that the breakpoints included the financial firms as well. This might be a reason for the differences in correlation.

Table 4: Size sort returns for first 6 deciles

	1	2	3	4	5	6
Return	0.0798113	0.0796329	0.0897193	0.0813522	0.0854843	0.0790571

	1	2	3	4	5	6
Standard Deviation	0.2173863	0.2210978	0.2096064	0.2031476	0.1969811	0.1851917
Sharpe Ratio	0.3671403	0.3601707	0.4280369	0.4004583	0.4339723	0.4268936
Skewness	-0.0811834	-0.2428759	-0.4998181	-0.5103414	-0.5229077	-0.5611134
Correlation	0.9968548	0.9987513	0.9988636	0.9987957	0.9982941	0.9986333

Table 5: Size sort returns for last 4 deciles and long short

	7	8	9	10	LongShort
Return	0.0838696	0.0766303	0.0725157	0.0545057	0.0253056
Standard Deviation	0.1831629	0.1786477	0.1645450	0.1527510	0.1658236
Sharpe Ratio	0.4578958	0.4289464	0.4407044	0.3568269	0.1526056
Skewness	-0.4642631	-0.4891281	-0.4449456	-0.3378191	0.9530220
Correlation	0.9984754	0.9991019	0.9990713	0.9997673	0.9945879

Question 3 - BE/ME Portfolio Correlation

Data

Apart from the results from the previous question (from 1973 to 2015), we also need the Fama French returns data divided based on book-to-market ratio.

Methodology

The output contains the following information for each decile and the long short portfolio :

1. Arithmetically Annualized excess returns (reduced by risk free rate) of each book-to-market decile and the long short portfolio. Arithmetic annualization is done by multiplying by 12.

2. Arithmetically annualized standard deviation of each book-to-market decile and the long short portfolio. The annualization is done by multiplying by $\sqrt{(}12)$

3. Annualized Sharpe Ratio, which is calculated by dividing the excess annualized returns and the annualized sharpe ratio.

4. Skewness, which is calculated on the actual returns over the time period.

5. Correlation, which is formulated by comparing the returns which were calculated and the returns posted on the fama french website.

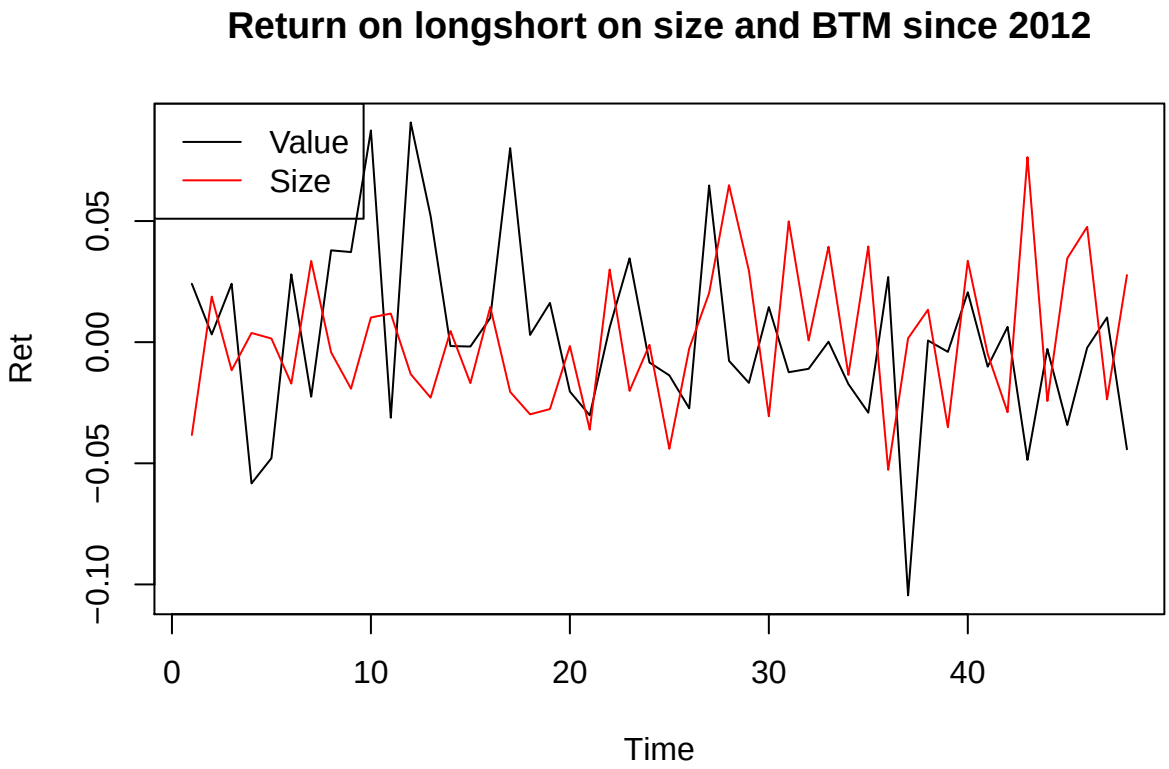
Table 6: Book-to-Market sort returns for first 6 deciles

	1	2	3	4	5	6
Return	0.0488105	0.0587514	0.0709362	0.0770862	0.0634466	0.0756634
Standard Deviation	0.1832134	0.1687164	0.1643203	0.1667463	0.1617747	0.1599277
Sharpe Ratio	0.2664136	0.3482254	0.4316949	0.4622961	0.3921909	0.4731097
Skewness	-0.2206038	-0.4400086	-0.5058909	-0.4589052	-0.4401935	-0.4913277
Correlation	0.9938295	0.9862461	0.9806990	0.9790734	0.9796473	0.9718574

Table 7: Book-to-Market sort returns for last 4 deciles and long short

	7	8	9	10	LongShort
Return	0.0771589	0.0815928	0.0784705	0.1157575	0.0669470
Standard Deviation	0.1573046	0.1617718	0.1671650	0.2050114	0.1627214
Sharpe Ratio	0.4905060	0.5043697	0.4694191	0.5646393	0.4114206
Skewness	-0.1283219	-0.5752685	-0.3851551	-0.0566273	0.5896358
Correlation	0.9722715	0.9712876	0.9600777	0.9727203	0.9403025

Question 4



As can be seen in the graph for anomaly returns from 2012 to 2017, even though both have had losses during the period, the long-short value portfolio has more consistently provided positive returns as compared to the long-short size portfolio.

The previous results also suggests that the BE/ME characteristic (anomaly) has a much higher sharpe ratio compared to the Size characteristic. Based on the tests conducted by Bonferroni, Holm and BHY (Benjamin, Hochberg and Yekutieli), it has been proven that the t statistics for the size portfolio is very low, which doesn't warrant any significant return. Whereas the value anomaly has highly significant returns.

Question 5 - HML and SMB Portfolio

Data

Apart from the results from the previous question (from 1973 to 2015), we also need the Fama French returns data for the SMB and HML factors.

Methodology

The output contains the following information for the HML and SMB factor portfolio:

1. Arithmetically Annualized excess returns (*not reduced by risk free rate*) for the SMB and HML portfolio. Arithmetic annualization is done by multiplying by 12.

2. Arithmetically annualized standard deviation for SMB and HML portfolio. The annualization is done by multiplying by $\sqrt{(}12)$

3. Annualized Sharpe Ratio, which is calculated by dividing the excess annualized returns and the annualized sharpe ratio.

4. Skewness, which is calculated on the actual returns over the time period.

5. Correlation, which is formulated by comparing the returns which were calculated and the returns posted on the fama french website.

P.S. The arithmetic excess return in the first row just indicates the actual return of the portfolio minus how much was spent in funding the portfolio. If return was to compare how much more return than risk free rate is produced, then another row has been added with this information.

	HML	SMB
Excess Return	0.0383746	0.0212569
Standard Deviation	0.1012521	0.1081456
Sharpe Ratio	0.3790005	0.1965580
Skewness	-0.2147971	0.6383005
Correlation	0.9799284	0.9944340
Excess Return - Rf	-0.0100138	-0.0271315

Question 6

Characteristic portfolios

The characteristic portfolios (1992) are aimed at showing that two easily measurable variables, size (ME) and book-to-market equity (BE/ME) provide a simple and powerful characterization of the cross-section of average stock returns. For example, it has been shown that there is a high positive correlation between average return and book-to-market equity, which is unlikely a beta effect (as beta changes very less across portfolios ranked on BE/ME values). On the whole, it was proved that size and book-to-market captures the cross-sectional variation in average stock returns associated with size, book-to-market and E/P and leverage (while used in combination).

The characteristic portfolio is based on the **Fama Macbeth regression** of the average returns against the variables hypothesized to explain the returns.

Factor Portfolios

The factor portfolios(1993) on the other hand identified the mimicking portfolios for size and book-to-market equity which can then used in a **time-series regression** approach to identify factor loadings on these mimicking portfolios. This goes along with the theme that if assets are priced rationally, then variables related to average returns must proxy for sensitivity to common risk factors. The paper proves that these mimicking factors, seperately and together, capture strong common variation in returns.

The time series regression also indicates these factors alone can't explain the average return on stocks. Major difference between the average return on stocks and T-bill is explained by the market factor.