CUFE·CAFD

TITLE：Investment Assignment VII

GROUP NAME：GGM Never Die

MEMBERS：高雅萌、郭好格、孟舒晨

DATE：12/5/2019

THE RESULTS:

Match the factors that we computed with the factors from WRDS, and run a correlation procedure between the two files. The results are as follows:

Pearson 相关系数, N = 342 Prob > |r| under H0: Rho=0		
	computed_HML	HML
computed_HML	1.00000	0.94717 <.0001
HML HML	0.94717 <.0001	1.00000

Pearson 相关系数, N = 342 Prob > |r| under H0: Rho=0		
	computed_SMB	SMB
computed_SMB	1.00000	0.98350 <.0001
SMB SMB	0.98350 <.0001	1.00000

Pearson 相关系数, N = 342 Prob > |r| under H0: Rho=0		
	computed_rm_rf	Mkt_RF
computed_rm_rf	1.00000	0.99649 <.0001
Mkt_RF Mkt-RF	0.99649 <.0001	1.00000

From the results, we can see that the correlation coefficient between computed_HML and the HML is 0.94717, the correlation coefficient between computed_SMB and the SMB is 0.98350, the correlation coefficient between computed_RM_RF and the Mkt_RF is 0.99649. The correlation coefficients are all beyond 0.9 which means our construction is good, but in the meantime, not exactly the same with the factors from WRDS which can be explained by that we are ignoring some details in the construction of the factors, for example, the data in French's website is constructed with an expanded dataset (using some book values that are not available on COMPUSTAT).

Our assignment is carried out by three parts:

Part 1: select data from dataset:

First, we start by select all NYSE, AMEX, and NASDAQ securities from the CRSP monthly stock price file 'msf' and in order to get the size data of previous year, we select the period of 1962 to 1992.

```
proc sql;
create table d_ff as
select
    msf.permno,
    msf.ret,
    msf.prc,
    msf.shrout,
    msf.hsiccd,
    msf.hexcd,
    msf.date
from
    worklib.msf
where
    msf.hexcd<=3 and 1962<=year(msf.date)<=1992;
quit;
```

Then we use the data below to select firms with ordinary common equity:

```
data d_june;
    set d_june;
    where shrcd^=. and 10<=shrcd<=12;
```

Then we select firms' total assets data(data6), Earnings per Share (data53) and common equity (data60), and calculate the market size value of each firm according to the previous year.

```
*Collect market size as of december of previous year;
proc sql;
create table d_cst_size as
select
d_cst.*,
       (abs(d_ff.prc)*d_ff.shrout) as size_prev_year,
d_ff.date as date_prev_year
from
d_cst left join d_ff
on
d_cst.permno=d_ff.permno and
month(d_ff.date)=12 and
year(d_ff.date)=year(d_cst.date)-1
order by
date, permno;
run;
libname worklib 'D:\investment\data\lecture8';
data crspcusip;
```

```
    set worklib.crspcusip;
run;
 data permno_shrcd;
    set worklib.permno_shrcd;
run;
*Collect monthly data for stocks, 1963-1991;
proc sql;
create table d_ff as
select
    msf.permno,
    msf.ret,
    msf.prc,
    msf.shrout,
    msf.hsiccd,
    msf.hexcd,
    msf.date
from
    worklib.msf
where
    msf.hexcd<=3 and 1962<=year(msf.date)<=1992;
quit;
%*Collect compustat data and size of previous year;
data d_june;
    set d_ff;
    if prc^=. and shrout^=. then
        size_june=abs(prc)*shrout;
    else
        size_june=.;
run;
%* Collect sharecode;
proc sql;
create table d_june as
select distinct
    d_june.*,
    permno_shrcd.shrcd
from
    d_june left join permno_shrcd
on
    d_june.permno=permno_shrcd.permno and
    permno_shrcd.initial_date<=d_june.date and
    d_june.date<=permno_shrcd.final_date
order by
    permno,date,shrcd;
quit;
```

```
    proc sort data=d_june nodupkeys;
        by permno date;
    data d_june;
        set d_june;
        where shrcd^=. and 10<=shrcd<=12;
    %*Collect COMPUSTAT data;
    proc sql;
    create table d_cst as
    select
        d_june.*,
        crspcusip.gvkey
    from
        d_june left join crspcusip
    on
        d_june.permno=crspcusip.permno and
        year(d_june.date)-1>=crspcusip.begyear and
        year(d_june.date)-1<=crspcusip.endyear;
    quit;
    data compann;
      set worklib.compann;
      where indfmt='INDL' and datafmt='STD' and consol='C' and
  popsrc='D';
      if SEQ>0;                    /* Shareholders' Equity */
      PREF=PSTK;                   /* Preferred stock - Redemption Value
*/
      if missing(pref) then PREF=PSTKR; /* Preferred stock -
Liquidating Value */
      data60 = sum(SEQ, TXDB, -PREF); /* Deferred taxes  and
Investment Tax Credit */
      label data60 = "Book Value of Equity";
      rename at=data6;
      rename EPSPX=data53;
      newgvkey=input(gvkey,8.);
    run;
    data compann;set compann;drop gvkey;rename newgvkey=gvkey;run;
    proc sql;
    create table d_cst as
    select
    d_cst.*,
    compann.data60,
    compann.data6,
    compann.data53,
    compann.datadate as compustat_date
    from
```

```
d_cst left join compann
on
d_cst.gvkey=compann.gvkey and
year(d_cst.date)-1=year(compann.datadate)
order by
date,permno;
run;
*Collect market size as of december of previous year;
proc sql;
create table d_cst_size as
select
d_cst.*,
    (abs(d_ff.prc)*d_ff.shrout) as size_prev_year,
d_ff.date as date_prev_year
from
d_cst left join d_ff
on  d_cst.permno=d_ff.permno and
month(d_ff.date)=12 and
year(d_ff.date)=year(d_cst.date)-1
order by
date, permno;
run;
data d_cst_size;
set d_cst_size;
where size_prev_year^=. and data6^=. and data60^=. and data53^=.;
run;
data worklib.d_cst_size;
    set d_cst_size;
data worklib.d_ff;
    set d_ff;
run;
```

part 2: construct FF factors

First, split all stocks into two size groups, S and B;

In June of each year t from 1963 to 1991, all NYSE stocks on CRSP are ranked on size (price times shares outstanding), the median NYSE size is then used to split NYSE, Amex and NASDAQ stocks into two groups, small and big (denoted as S and B). Most Amex and NASDAQ stocks are smaller than NYSE median. Therefore, the small group contains a large number of stocks, but it contains far less than half of the combined value of the two size groups.

```
*define the NYSE size breakpoint;
proc sort data=d_june;
    by month;
proc means data=d_june noprint;
    var size_june;
    output out=size
    median(size_june)=size_breakpoint;
    by month;
    where hexcd=1;
run;
*group 2 according to median size;
data d_june;
    merge d_june size(keep= month size_breakpoint);
    by month;
    if size_june>size_breakpoint then sizetype="B";
    else sizetype="S";
run;
```

Second, split all stocks into three beme groups, H, M and L;

We also break NYSE, Amex and NASDAQ stocks into three book-to-market equity groups based on the breakpoints for the bottom 30%(L), middle 40%(M) and top 30%(H) of the ranked values of BEME for NYSE stocks. The book equity value is defined as book value of shareholders' equity, plus balanced-sheet deferred taxes and investment tax credit, minus the book value of preferred stock. The market equity value is defined as price times shares outstanding. And the book equity value and market equity value both use the data of December of year t-1.

```
proc univariate data=d_cst_size noprint;
    var beme;
    by month;
    output out=beme
    pctlpre=bemetype
    pctlpts=(30,70);
    where hexcd=1;
run;
proc sort data=d_cst_size;
by month;
data d_cst_size;
    merge d_cst_size beme(keep=month bemetype30 bemetype70);
    by month;
    if beme<=bemetype30 then bemetype="L";
    if beme>=bemetype70 then bemetype="H";
    if bemetype30<=beme<=bemetype70 then bemetype="M";
run;
```

Third, construct market return;

Market returns are computed as the monthly value-weighted return.

```
*compute value weighted market return;
proc sort data=d;
    by permno date;
run;
data d;
    set d;
    value=abs(prc)*shrout;
run;
proc sort data=d;
    by date permno;
run;
proc means data=d noprint;
    weight value;
    output out=mktret
    mean(ret)=vwretd;
    by date;
    where value^=.;
run;
```

Fourth, construct factor SMB;

SMB (small minus big), meant to mimic the risk factor in related to size, is the difference, each month, between the simple average of the returns on the small-stock portfolios(S/L, S/M and S/H) and the simple average of the returns on the three big-stock portfolios (B/L, B/M and B/H). Thus,

Lastly, construct factor HML;

HML (high minus low), meant to the risk factor in returns related to book-to-market equity, is defined similarly. HML is the difference, each month, between the simple average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L).

```
data result;
    set factor;
    computed_SMB=(retS_H+retS_M+retS_L)/3-(retB_H+retB_M+retB_L)/3;
    computed_HML=(retS_H+retB_H)/2-(retS_L+retB_L)/2;
    month=year(date)*100+month(date);
run;
```

Part 3: The Regression Results of Fama_French 3 Factors Model

Fama and French (1993) proposes a multifactor model to explain excess returns on stocks:

$$R_{it} - R_{ft} = \alpha_i + b_i * (R_{mt} - R_{ft}) + s_i * SMB_t + h_i * HML_i + \varepsilon_{it}$$

The factors in the model are:

1. The excess market return (as in CAPM);

2. SMB (“small minus big”): the return on small stocks minus the return on big stocks;

3. HML(“high minus low”): the return on high book to market stocks minus the return on low book to market stocks. The factors in the model are supposed to capture firm characteristics that seem to represent exposures to systematic risk. Fama and French (1992) identified size and book to market as such characteristics.

We run the regression again, using the factors we constructed, the results are shown below, which is closed to Fama-French’s results:

Stats on Intercept

观测	portfolio_size	int1	int2	int3	int4	int5	t_int1	t_int2	t_int3	t_int4	t_int5
1	1	-0.51324	-0.09911	-0.05104	0.047915	0.11100	-3.39033	-0.90837	-0.54997	0.58827	1.24922
2	2	-0.22467	-0.11320	0.04595	0.045013	-0.06045	-2.04153	-1.30230	0.60247	0.67411	-0.77896
3	3	-0.04750	0.00529	-0.08107	0.055787	0.08745	-0.49512	0.05655	-1.04981	0.81324	0.93566
4	4	0.00027	-0.10773	-0.08374	0.008873	0.09884	0.00322	-1.22090	-0.98328	0.11028	0.95545
5	5	0.20136	-0.06346	-0.12320	-0.003730	-0.03039	2.48826	-0.85715	-1.32678	-0.04506	-0.30419

Stats on rm-rf

观测	portfolio_size	rm_rf1	rm_rf2	rm_rf3	rm_rf4	rm_rf5	t_rm_rf1	t_rm_rf2	t_rm_rf3	t_rm_rf4	t_rm_rf5
1	1	1.03210	0.96354	0.91107	0.87183	0.88960	27.8070	36.0182	40.0373	43.6565	40.8330
2	2	1.13434	1.03205	0.97630	0.94925	1.06245	42.0397	48.4255	52.2081	57.9807	55.8435
3	3	1.08622	1.01173	0.95884	0.95681	1.08201	46.1800	44.0830	50.6445	56.8881	47.2145
4	4	1.05261	1.06387	1.01884	1.01538	1.13728	50.7507	49.1766	48.7957	51.4690	44.8370
5	5	0.96771	1.01171	0.95549	1.00217	1.09555	48.7728	55.7346	41.9681	49.3755	45.0687

Stats on SMB

观测	portfolio_size	smb1	smb2	smb3	smb4	smb5	t_smb1	t_smb2	t_smb3	t_smb4	t_smb5
1	1	1.44530	1.39224	1.29751	1.25427	1.28500	26.1350	34.9304	38.2702	42.1542	39.5872
2	2	1.13329	0.95407	0.93664	0.81082	0.92743	28.1899	30.0460	33.6172	33.2404	32.7173
3	3	0.82031	0.67804	0.56958	0.52176	0.74958	23.4072	19.8289	20.1919	20.8212	21.9533
4	4	0.35455	0.29634	0.28150	0.25338	0.44651	11.4732	9.1938	9.0486	8.6202	11.8152
5	5	-0.19199	-0.16598	-0.25513	-0.16572	0.01781	-6.4944	-6.1372	-7.5213	-5.4801	0.4933

Stats on HML

观测	portfolio_size	hml1	hml2	hml3	hml4	hml5	t_hml1	t_hml2	t_hml3	t_hml4	t_hml5
1	1	-0.13476	0.11764	0.31624	0.39678	0.56613	-2.1663	2.62388	8.29198	11.8547	15.5048
2	2	-0.39954	0.03892	0.25835	0.44805	0.73945	-8.8349	1.08950	8.24310	16.3290	23.1902
3	3	-0.47534	-0.00662	0.27837	0.47753	0.70080	-12.0578	-0.17209	8.77268	16.9405	18.2462
4	4	-0.45877	0.00697	0.23974	0.49150	0.64326	-13.1979	0.19224	6.85098	14.8652	15.1318
5	5	-0.43143	-0.02731	0.14416	0.55715	0.79869	-12.9741	-0.89755	3.77799	16.3784	19.5861

These tables show that when the excess market return is also in the regressions, each of the three stock-market factors captures variation in returns. Not surprisingly. the three stock-market factors capture strong common variation in stock returns.

The t-statistics on the SMB slopes for stocks are greater than 4; most are greater than 10. SMB, the mimicking return for the size factor, clearly captures shared variation in stock returns that is missed by the market and by HML. Moreover, the slopes on SMB for stocks are related to size. In every book-to-market quintile, the slopes on SMB decrease monotonically from smaller- to bigger-

size quintiles.

Similarly, the slopes on HML, the mimicking return for the book-to-market factor, are systematically related to BE/ME. In every size quintile of stocks, the HML slopes increase monotonically from strong negative values for the lowest- quintile to strong positive values for the highest-BE/ME quintile.

The βs for these two portfolios are 1.04 and 1.06. In general adding SMB and HML to the regressions collapses the βs for stocks toward 1.0: low βs move up toward 1.0 and high Bs move down. This behavior is due, of course, to correlation between the market and SMB or HML. Although SMB and HML are almost uncorrelated, the correlations between RM-RF and the SMB and HML returns are 0.32 and - 0.38.

The most of the t-statistics on the intercept are not significant, which means these factors can explain the average premium of stock returns.