

## Calculate ROA remotely

```
/* PC-SAS/Connect Communication Block */
%let wrds = wrds.wharton.upenn.edu 4016;
options comamid=TCP remote=WRDS;
signon username=_prompt_;
/* Submit SAS code to WRDS Unix Server */
rsubmit;

/* SAS CODE submitted to WRDS Sever */
proc sql;
create table work.demo as

select gvkey, tic, fyear, ni/at as roa

from comp.funda

where not missing(at) and consol="C" and indfmt="INDL" and
datafmt="STD" and popsrc="D" and fyear=2019

order by gvkey, fyear;

quit;

/* Remote SAS CODE submission Ends */
end submit;
/* Sign off from PC-SAS/Connect */
Signoff;
```

## Some non-SAS coding

### ► PostgreSQL Data Extraction

```
comp = conn.raw_sql("""
select gvkey, tic, fyear, ni, at,
       case when at=0 then null else ni/at end as roa
from comp.funda
where at is not null and indfmt='INDL' and datafmt='STD'
and popsrc='D' and consol='C' and fyear=2019
order by gvkey, fyear
""")
```

### ► Python (Pandas) Data Extraction

```
msi=pd.read_sas("/wrds/crsp/sasdata/a_stock/msi.sas7bdat")
```

## Compustat Data Handling

- “We define book common equity, BE as the COMPUSTAT book value of stockholders equity, plus balance-sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock.”
  - “Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the value of preferred stock.”

```
proc sql;
create table comp as
select distinct gvkey, fyear, datadate,
coalesce(PSTKRVP,PSTKL,PSTK,0) as ps, /* Formula used by FF(1993) */
seq + sum(TXDIRTC,0) ~ calculated ps as be 'Book Value', /* Formula used by FF(1993) */
year(datadate) as year /*Calendar Year*/
from comp.funda
where indfmt='INDL' and datafmt='STD' and popsrc='D' and consol='C' /*Standard Compustat Filters*/;

create table comp as
select distinct *, fyear-min(fyear) as age /*Checking existing years*/
from comp
group by gvkey;

quit;
```

♦ Note Inclusion of Investment Tax Credit (ITCB) would reduce the correlation coefficient of SML and HTML outputs of this code and the ones from Professor Kenneth R. French's Data Library.

## CRSP Data Handling

- The Fama and French (1993) and many follow up papers do not adjust return of equity with delisted return, which would be discussed more in the Event Study Notes.
- The delisted return treatment, however, is more appropriate for construction of mimicking portfolio as stock delisting is rarely a foreseeable event.

```
proc sql;
create table msf as
select distinct a.permco, a.permno, a.date, pro, ret, shrout, retx, shred, exchcd
from crsp.msif(where=(date>='01Dec1961'd)) as a inner join
crsp.msenames(where=(exchcd in (1,2,3))) as b
/*1:NYSE(-2), AMEX(-2), and NASDAQ (-3)*/
on a.permno=b.permno and NAMEEDT<=date<=NAMEENDT;

create table crspm as
select a.*, b.dlret, sum(1,ret)*sum(1,dlret)-1 as retadj
/*Return adjusted for delisting*/(2)Adding Delisted Returns*/
abs(a.pro)*a.shrout as meq 'Market Value of Equity at Issue Level'
from msf as a left join crsp.msdelist(where=(missing(dlret)=0)) as b
on a.permno=b.permno and
intnx('month',a.date,0, 'END')=intnx('month',b.DLSTDT,0, 'END')
order by a.date, a.permco, MEQ;

quit;
```

## CRSP Data - Market Value of Equity

```
proc sql;
    create table crspm as
    select distinct *, sum(meq) as me
    from crspm
    group by permco, date;
quit;
```

TICKER	PERMCO	PERMNO	DATE	MEQ	ME
GOOGL	45483	90319	20180531	328,521,600	707,131,027
GOOG	45483	14542	20180531	378,609,427	707,131,027
GOOGL	45483	90319	20180629	337,239,351	726,547,658
GOOG	45483	14542	20180629	389,308,307	726,547,658

❖ Note one way to determine primary shares is to use the shares with larger Market Cap, though this approach many not be always right. Stock with ticker GOOG is, for example, Class-C shares (shares without voting rights)

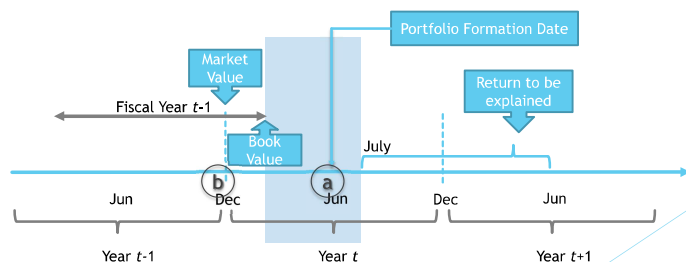
## CRSP Data -Rebalance

- The code below is conceptually the most straightforward solution to construct a value-weighted portfolio
  - All dividends are assumed to be reinvested proportionally based on the market value of stocks at the previous month end.
  - The code appendix provides another approach suggested by Professor Robert F. Stambaugh at Wharton School, which may further improve the correlation coefficients of calculated SMB and HML with the ones from Professor Kenneth R. French's Data Library.

```
proc sql;
    create table crspm_w as
    select distinct a.permno, a.date, a.retadj, a.me, a.exchcd, a.shrod, b.meq as weight_port
    from crspm as a inner join crspm as b
    on a.permno=b.permno and intnx('month',a.date,-1,'E')=intnx('month',b.date,0,'E');
quit;
```

## CRSP Data - Size and B/M Preparation

```
proc sql;
    create table crspjune as
    select a.permno, a.date, a.exchcd, a.shrod, a.me as JUN_ME "Market Cap in current June"/"a"/,
    b.me as DEC_ME "Market Cap in previous December"/"b"/
    from crspm(where=(month(date)=6)) as a inner join
    (select distinct * from crspm(where=(month(date)=12))) as b
    on a.permno=b.permno and intnx('month',b.date,a.date)=6;
quit;
```



## CRSP & COMPUSTAT Matching

```
proc sql;
    create table ccm as
    select a.*, b.permno as permno, b.linkprin
    from comp as a, crsp.ccmxpf_inkhist as b
    where a.gvkey=b.gvkey
    and linktype in ('LC', 'LU') /*1)Confirmed Linkage Only*/
    and linkprin in ('P', 'C') /*2)Take Primary Shres Only*/
    and linkin<=datadate<=coalesce(linkenddt, today())
    /*3)Matched based on the first valid June after Fiscal Year END*/
    group by permno, fyear
    /*4)Keep the Last Datadate for a Calendar Year*/
    having datadate=max(datadate)
    /*5)Remove Duplicated Records through disagreed primary shares*/
    order by permno, year, datadate;
quit;
```

Potential duplications:

- ❖ A few cases contain different GVKEYS for a same PERMNO-Date combination
- ❖ Some companies change length of fiscal year end

- For Matching CRSP and Compustat with CUSIP, please see Case Study at the end of this class.

## CCM - Size and B/M Calculation

- ▶ “[T]o avoid the survival bias inherent in the way COMPUSTAT adds firms to its tapes, we do not include firms until they have appeared on COMPUSTAT for two years.”
  - ▶ Adding this condition would slightly reduce the correlation coefficient, which may imply that the official “Fama French” factors may no longer take consideration of the survivorship bias documented in early papers.
  - ▶ In fact, Compustat now updates (and corrects) its data on daily basis, so that this previously documented bias may no longer exist in recent Compustat data.

```
proc sql;
    create table ccm_june as
    select a.*, b.BE, (1000*b.BE)/a.DEC_ME as BEME, b.age, b.datadate
    from crspjune as a, ccm as b
    where a.permno=b.permno and age > 2 and be > . and
    intnx('month',a.date,-12,'E')<datadate<intnx('month',a.date,0,'E')
    order by a.date;
quit;
```

## NYSE Break Point Calculation

```
proc univariate data=ccm_june noprint;
    where exched=1 and beme>0 and shrd in (10,11) /*Common US Stocks*/;
    var JUN_ME BEME;
    by date;
    output out=nyse_breaks median = SIZEMEDN
    pctlpre=JUN_ME BEME pctlpts=30 70;

run;

proc sql;
    create table ccm.nyse_port as
    select a.permno, a.date, a.exched, a.shrd,
    case when JUN_ME <= sizemedn then 'S' else 'B' end as sizeport,
    case when beme <= beme30 then 'L' when beme > beme70 then 'H'
    else 'M' end as btmport
    from ccm_june as a inner join nyse_breaks as b
    on a.date=b.date and shrd in (10,11);

quit;
```

	Small	Big
High B/M (Value)	S&H	B&H
Middle B/M (Neutral)	S&M	B&M
Low B/M (Grow)	S&L	B&L

❖ In the original Fama & French (1993), size and B/M break breakpoints are created by NYSE stocks only to ensure consistency introduced by the small stocks listed in the later joined NASDAQ (1973).

❖ This treatment becomes a standard way to handle breaking points among Asset Pricing literature, even in the ones using data after CRSP NASDAQ data integration.

## NYSE Break Portfolio Construction

```
proc sql;
    create table port_ccm as
    select a.*, b.sizeport, b.btmport, b.date as portdate format date9.
    from crspm_w as a, ccm.nyse_port as b
    where a.permno=b.permno and 1 <= intck('month',b.date, a.date) <= 12
    /*1) Match Previous June to current July to June*/
    order by date, sizeport, btmport;

    create table port_ccm_vwret as
    select distinct date, sizeport, btmport, count(distinct permno) as n_firms,
    sum(weight_port*retadj)/sum(weight_port) as vwret
    from port_ccm
    group by date, sizeport, btmport;

quit;

proc transpose data=port_ccm_vwret (keep=date sizeport btmport vwret)
    out=ff_vwret (drop=_name_);
    by date ;
    ID sizeport btmport;
    Var vwret;

run;
```

$$\diamond \text{Value Weighted Return}_t = \sum_i^n (w_{i,t-1} \times R_t) = \sum_i^n \left( \frac{\text{Market Cap of Stock}_{i,t-1}}{\sum_i^n (\text{Market Cap of Stock}_{i,t-1})} \times R_t \right) = \frac{\sum_i^n (\text{Market Cap of Stock}_{i,t-1} \times R_t)}{\sum_i^n (\text{Market Cap of Stock}_{i,t-1})}$$

A Constant

## Fama & French Factors

```
data ff_factors; set ff_vwret;
    WH = (bh + sh)/2;
    WL = (sl + bl)/2;
    WHML = WH - WL;
    WB = (bl + lm + bh)/3;
    WS = (sl + sm + sh)/3;
    WSMB = WS - WB;
    label WH = 'WRDS High'
    WL = 'WRDS Low'
    WHML = 'WRDS HML'
    WS = 'WRDS Small'
    WB = 'WRDS Big'
    WSMB = 'WRDS SMB';

run;
```

	Small	Big
High B/M (Value)	S&H	B&H
Middle B/M (Neutral)	S&M	B&M
Low B/M (Grow)	S&L	B&L

- $HML_t = \text{AVG}(R_{t, \text{Port}_{S\&H}}, R_{t, \text{Port}_{B\&H}}) - \text{AVG}(R_{t, \text{Port}_{S\&L}}, R_{t, \text{Port}_{B\&L}})$
- $SMB_t = \text{AVG}(R_{t, \text{Port}_{S\&H}}, R_{t, \text{Port}_{S\&M}}, R_{t, \text{Port}_{S\&L}}) - \text{AVG}(R_{t, \text{Port}_{B\&H}}, R_{t, \text{Port}_{B\&M}}, R_{t, \text{Port}_{B\&L}})$

# Result Comparison

```
proc sql;
  create table ff_test as
  select distinct a.*, smb, hml
  from ff_factors as a inner join ff_factors_monthly as b
  on intnx('month', a.date, 0, 'e')=intnx('month', b.date, 0, 'e');
quit;

proc corr data=ff_test;
  var smb hml;
  with wsmb whml;
run;

proc corr data=ff_test(where=(year(date)>1980));
  var smb hml;
  with wsmb whml;
run;
```

Pearson Correlation Coefficients, N = 690 Prob >  r  under H0: Rho=0		
	smb	hml
WSMB	0.98461 <.0001	-0.13715 <.0001
WHML	-0.18290 <.0001	0.96858 <.0001

1962 - Present

Pearson Correlation Coefficients, N = 468 Prob >  r  under H0: Rho=0		
	smb	hml
WSMB	0.99280 <.0001	-0.18808 <.0001
WHML	-0.25655 <.0001	0.97669 <.0001

1980 - Present

# CRSP Data - Rebalance (Alternative)

```
data crspm_w; set crspm; by permno date; retain me_base cumretx;
  LME=lag(meq); /*lagged month Market Values*/

  if first.permno then do;
    LME=me/(1+retx); me_base=LME;
    weight_port=.; cumretx=sum(1,retx);
  end; /*1) Initiate lagged Values*/

  if month(date)=7 then do; /*2) Portfolio Formation Date at June*/
    weight_port=LME; /*3) RESET Weights at the beginning of July*/
    me_base=LME;
    cumretx=sum(1,retx);
  end;

  else if first.permno=0 then do;
    if LME>0 then weight_port=cumretx*me_base;
    else weight_port=.;
    cumretx=cumretx+sum(1,retx); /*4) Dividend is not considered*/
  end;

  keep permno date retadj weight_port MB exched shrcd;
run;
```

- When a stock start records before July, its previous month weights are set to NULL
- When it is July, the previous month weights are set to Jun's one and cumulative RETX is to July one
- If it is not July and not the first obs of a permno, then the market weight is calculated based on cumulative RETX and market cap on previous June.