I. Code for Computing the TFP with Intangible Capital Part:

1. The SAS Studio Code To get the data from WRDS

```
libname maindata "/wrds/comp/sasdata/naa";

data compustat_data(keep= fyear gvkey conm FIC PPEGT AT SALE OIBDP DP EMP CAPX
DPACT INTAN DFXA AM );

set maindata.funda;

where DATAFMT='STD' and POPSRC='D' and CONSOL='C' and INDFMT='INDL' and
CURCD='USD';

run;

data compustat_names(keep= gvkey sic naics conm year1 year2);

set maindata.names;

run;
```

2. The MATLAB Code To Compute Intangible Capital and Investment in Intangible Capital: The running time: about 40 minutes

```
%The file Intan_Part_Intan_Cap_full.xlsx is getting from  Peter and
%Taylor Total Q. in WRDS with the variables: GVKEY,Firm's Knowledge Capital
%Replacement Cost, Firm's Organization Capital Replacement Cost and Firm's
%Intangible Estimated Replacement Cost. The time period is 1950 to 2017.

% The file Intan_Part_SGA_RD_full.xlsx is getting from Compustat in WRDS
% with the variable: Gvkey;Fiscal Year; Asset Total; Cost of Good Sold; Intangible Asset
% Total; In Process R&D Expense; Research and Development Expense Selling,
% General and Administrative Expense. The Time period is 1950 to 2018


% clean the command window and all variables in the workspace, running time 40 mins
clc; clearvars; close all;

%% Input
data_Cap = readtable('Intan_Part_Intan_Cap_full.xlsx');
data_SGA_RD= readtable('Intan_Part_SGA_RD_full.xlsx');
data_Cap = data_Cap(:,{'GVKEY' 'FiscalYear'
'Firm___sIntangibleCapitalEstimatedReplacementCost'});
data_Cap.Properties.VariableNames{'GVKEY'} = 'GlobalCompanyKey';
data_Cap.Properties.VariableNames{'FiscalYear'} = 'DataYear_Fiscal';
```

```
data_Cap.Properties.VariableNames{'Firm___sIntangibleCapitalEstimatedReplacementCost'} =
'FirmIntangibleCapital';

t_year = 1977;

toDelete = data_SGA_RD.DataYear_Fiscal <t_year;
data_SGA_RD(toDelete,:) = [];
clear toDelete;

toDelete = data_Cap.DataYear_Fiscal <t_year;
data_Cap(toDelete,:) = [];
clear toDelete;


%% Calculate Investment

data_SGA_RD.ResearchAndDevelopmentExpense(isnan(data_SGA_RD.ResearchAndDevelop
mentExpense)) = 0;
data_SGA_RD.InProcessR_DExpense(isnan(data_SGA_RD.InProcessR_DExpense)) = 0;

n= size(data_SGA_RD,1);

for i = 1:n
   if isnan(data_SGA_RD.Selling_GeneralAndAdministrativeExpense(i)) == 1
      data_SGA_RD.SGA(i)=0;
    elseif isnan(data_SGA_RD.Selling_GeneralAndAdministrativeExpense(i)) == 0 &...
          data_SGA_RD.ResearchAndDevelopmentExpense(i)>
data_SGA_RD.Selling_GeneralAndAdministrativeExpense(i) &&...
         data_SGA_RD.ResearchAndDevelopmentExpense(i) <
data_SGA_RD.CostOfGoodsSold(i)

      data_SGA_RD.SGA(i)=data_SGA_RD.Selling_GeneralAndAdministrativeExpense(i);
    else
     data_SGA_RD.SGA(i) = data_SGA_RD.Selling_GeneralAndAdministrativeExpense(i)...
        -data_SGA_RD.ResearchAndDevelopmentExpense(i)...
   - data_SGA_RD.InProcessR_DExpense(i);
    end
end

data_SGA_RD.investment_intan = data_SGA_RD.ResearchAndDevelopmentExpense +
0.3*data_SGA_RD.SGA;
data_SGA_RD = data_SGA_RD(:,{'GlobalCompanyKey' 'DataYear_Fiscal'
'investment_intan'});
data_SGA_RD = outerjoin(data_SGA_RD,data_Cap,'Type','left');

%% Change Name
```

```
data_SGA_RD.Properties.VariableNames{'GlobalCompanyKey_data_SGA_RD'} =
'GlobalCompanyKey';
data_SGA_RD.Properties.VariableNames{'DataYear_Fiscal_data_SGA_RD'} =
'DataYear_Fiscal';
data_SGA_RD = data_SGA_RD(:,{'GlobalCompanyKey' 'DataYear_Fiscal' 'investment_intan'
'FirmIntangibleCapital'});
%% export
writetable(data_SGA_RD,'data_intangible_capital.xlsx');
```

3. The Matlab Code in organizing/selecting/preparing data for computing the TFP with Intangible Capital:

a. Part 1:It would take 1 minutes to run the code

```
%% clean the command window and all variables in the workspace, running time: 1 mins
clc; clearvars; close all;

%% Import data
data = readtable('COMPUSTAT_DATA_MINH3.csv');
data_names = readtable('COMPUSTAT_NAMES_MINH3.csv');

data = outerjoin(data,data_names,'Type','left');
data.gvkey_data_names=[];
data.conm_data_names=[];

%% Remove compustatage<2
data.Compustatage= data.fyear - data.year1;
toDelete = data.Compustatage <2;
data(toDelete,:) = [];
clear toDelete;
%% first year of exit: there are no exits from Compustat before 1966
year2= data_names.year2;
firstexit = min(year2);

%% Remove Financial and Regulated Firm
t1 = 4900;
t2 = 4999;
t3 = 6000;
t4 = 6999;

toDelete = data.sic >= t1 & data.sic <t2;
data(toDelete,:) = [];
clear toDelete;
toDelete = data.sic >= t3 & data.sic <t4;
data(toDelete,:) = [];
```

```
clear toDelete;

%% Remove NaN
toDelete = isnan(data.at);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.emp);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.sale);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.ppegt);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.oibdp);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.dpact);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.dp);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.capx);
data(toDelete,:) = [];
clear toDelete;

%% remove negative and small value
t5 = 0.1;
t6 = 0;
toDelete = data.emp <t5;
data(toDelete,:) = [];
clear toDelete;
toDelete = data.sale <t5;
data(toDelete,:) = [];
clear toDelete;
toDelete = data.at <t5;
data(toDelete,:) = [];
clear toDelete;
toDelete = data.ppegt <t5;
data(toDelete,:) = [];
clear toDelete;
toDelete = data.dp ==t6;
data(toDelete,:) = [];
clear toDelete;
```

```
toDelete = data.dpact ==t6;
data(toDelete,:) = [];
clear toDelete;
toDelete = data.capx <=t6;
data(toDelete,:) = [];
clear toDelete;

%% change variable name

data.Properties.VariableNames{'gvkey_data'} = 'GlobalCompanyKey';
data.Properties.VariableNames{'conm_data'} = 'CompanyName';
data.Properties.VariableNames{'fyear'} = 'DataYear_Fiscal';
data.Properties.VariableNames{'at'} = 'Assets_Total';
data.Properties.VariableNames{'capx'} = 'CapitalExpenditures';
data.Properties.VariableNames{'dp'} = 'DepreciationAndAmortization';
data.Properties.VariableNames{'dpact'} =
'Depreciation_DepletionAndAmortization_Accumulated_';
data.Properties.VariableNames{'emp'} = 'Employees';
data.Properties.VariableNames{'intan'} = 'IntangibleAssets_Total';
data.Properties.VariableNames{'oibdp'} = 'OperatingIncomeBeforeDepreciation';
data.Properties.VariableNames{'ppegt'} = 'Property_PlantAndEquipment_Total_Gross_';
data.Properties.VariableNames{'sale'} = 'Sales_Turnover_Net_';

%% export
writetable(data,'data_part1.xlsx');
```

b. Part 2: It will take 70 minutes to run the code

```
%% clean the command window and all variables in the workspace, running time 65 mins
clc; clearvars; close all;

%% Import data
firstexit=1966;
data = readtable('data_part1.xlsx');
data_intangible_capital = readtable('data_intangible_capital.xlsx');

average_wage = readtable('wage_average.xlsx');
average_wage.Properties.VariableNames = {'DataYear_Fiscal','average_wage'};

data = outerjoin(data,data_intangible_capital,'Type','left');
data.Properties.VariableNames{'GlobalCompanyKey_data'} = 'GlobalCompanyKey';
data.Properties.VariableNames{'DataYear_Fiscal_data'} = 'DataYear_Fiscal';
data.DataYear_Fiscal_data_intangible_capital = [];
data.GlobalCompanyKey_data_intangible_capital = [];
```

```
toDelete = data.FirmIntangibleCapital <0.1;
data(toDelete,:) = [];
clear toDelete;

toDelete = data.investment_intan <=0;
data(toDelete,:) = [];
clear toDelete;

toDelete = isnan(data.FirmIntangibleCapital);
data(toDelete,:) = [];
clear toDelete;

toDelete = isnan(data.investment_intan);
data(toDelete,:) = [];
clear toDelete;


%% labor_expense
data = outerjoin(data,average_wage,'Type','left');
data.Properties.VariableNames{'DataYear_Fiscal_data'} = 'DataYear_Fiscal';
data.DataYear_Fiscal_average_wage=[];
labor_expense = data.average_wage .* data.Employees;
labor_expense = labor_expense ./1000; % we want to make mils in figure
data.labor_expense = labor_expense;
total_sales = data.Sales_Turnover_Net_; % total sales
oibdp=data.OperatingIncomeBeforeDepreciation; % operating income before depr. and amo.

total_expense = total_sales - oibdp;
material = total_expense - data.labor_expense;
value_added = total_sales - material;
data.value_added = value_added;
data.material = material;
data = sortrows(data,'GlobalCompanyKey','ascend');
%% Calculate Capital
%Compute Age of Capital
dpact = data.Depreciation_DepletionAndAmortization_Accumulated_;
dp = data.DepreciationAndAmortization;
cap_age = dpact./dp;
data.capital_age = cap_age;
data.average_age=zeros(size(data, 1), 1);
gvkey = data.GlobalCompanyKey;
gvkey = unique(gvkey,'rows');
n_loop_age = size(gvkey,1);
data_2=table();
for i_k = 1: n_loop_age
gvkey_k = gvkey(i_k,:);
```

```
FinalTable_k = data(data.GlobalCompanyKey == gvkey_k,:);
number_year_k = size(FinalTable_k,1);
for j_k= 1:number_year_k
   if (j_k==2) &&FinalTable_k.DataYear_Fiscal(j_k) - FinalTable_k.DataYear_Fiscal(j_k-
1)==1
       FinalTable_k.average_age(j_k) = (FinalTable_k.capital_age(j_k-
1)+FinalTable_k.capital_age(j_k))/2;
   elseif (j_k>2) && FinalTable_k.DataYear_Fiscal(j_k) - FinalTable_k.DataYear_Fiscal(j_k-
1)==1 && FinalTable_k.DataYear_Fiscal(j_k)-FinalTable_k.DataYear_Fiscal(j_k-2)~=2
        FinalTable_k.average_age(j_k) = (FinalTable_k.capital_age(j_k-
1)+FinalTable_k.capital_age(j_k))/2;
   elseif (j_k>2) && FinalTable_k.DataYear_Fiscal(j_k) - FinalTable_k.DataYear_Fiscal(j_k-
1)==1 && FinalTable_k.DataYear_Fiscal(j_k)-FinalTable_k.DataYear_Fiscal(j_k-2)==2
   FinalTable_k.average_age(j_k) = (FinalTable_k.capital_age(j_k-
2)+FinalTable_k.capital_age(j_k-1)+FinalTable_k.capital_age(j_k))/3;
   else
   FinalTable_k.average_age(j_k) = FinalTable_k.capital_age(j_k);
   end
average_age_each =FinalTable_k(j_k,:) ;
data_2=vertcat(data_2,average_age_each);
end
end

data_2.age = round(data_2.average_age,0);
data_2.cap_year = data_2.DataYear_Fiscal-data_2.age;
for loop_data_cap=1:size(data_2,1)
if data_2.cap_year(loop_data_cap) < 1951
  data_2.cap_year(loop_data_cap) = 1951;
else
end
end

%% Deflator
clear data;
data = data_2;
clear data_2;
past_inv_deflator = readtable('deflator.xlsx');
past_inv_deflator.wage_ind = [];
past_inv_deflator.gdp_ind = [];
past_inv_deflator.Properties.VariableNames = {'cap_year','past_inv_def'};
past_inv_deflator.past_inv_def = past_inv_deflator.past_inv_def;

data = outerjoin(data,past_inv_deflator,'Type','left');
data.cap_year_past_inv_deflator = [];

gdp_deflator = readtable('deflator.xlsx');
```

```
gdp_deflator.wage_ind = [];
gdp_deflator.nonres_ind = [];
gdp_deflator.Properties.VariableNames = {'DataYear_Fiscal','gdp_def'};
gdp_deflator.gdp_def = gdp_deflator.gdp_def;

data = outerjoin(data,gdp_deflator,'Type','left');
data.Properties.VariableNames{'DataYear_Fiscal_data'} = 'DataYear_Fiscal';


inv_deflator = readtable('deflator.xlsx');
inv_deflator.wage_ind = [];
inv_deflator.gdp_ind = [];
inv_deflator.Properties.VariableNames = {'DataYear_Fiscal','inv_def'};
inv_deflator.inv_def = inv_deflator.inv_def;

data = outerjoin(data,inv_deflator,'Type','left');
data.DataYear_Fiscal_inv_deflator = [];
data.DataYear_Fiscal_gdp_deflator = [];
data.Properties.VariableNames{'DataYear_Fiscal_data'} = 'DataYear_Fiscal';

%% compute total investment

data.total_inv=data.CapitalExpenditures+data.investment_intan;
data.fixed_cap = data.CapitalExpenditures;
%% compute the quantities of output, investment, and capital
data.adj_value = data.value_added ./ (data.gdp_def);
data.adj_cap=data.Property_PlantAndEquipment_Total_Gross_ ./ data.past_inv_def;
data.adj_inv=data.total_inv ./ data.inv_def;
data.adj_intan = data.FirmIntangibleCapital ./ data.past_inv_def;
data.adj_fixed_cap = data.fixed_cap ./ data.inv_def;
%% create the lag cap & lag intangible asset

gvkey_lag = data.GlobalCompanyKey;
gvkey_lag = unique(gvkey_lag,'rows');
n_loop_lag = size(gvkey_lag,1);
data_2=table();
for i_lag = 1: n_loop_lag
gvkey_lag_loop = gvkey_lag(i_lag,:);
FinalTable_lag = data(data.GlobalCompanyKey == gvkey_lag_loop,:);
number_year_lag = size(FinalTable_lag,1);
for j_lag= 1:number_year_lag
   if j_lag==1
     FinalTable_lag.lag_cap(j_lag)=NaN;
          FinalTable_lag.lag_intan(j_lag)=NaN;

   elseif FinalTable_lag.DataYear_Fiscal(j_lag) - FinalTable_lag.DataYear_Fiscal(j_lag-1)==1
```

```
        FinalTable_lag.lag_cap(j_lag) = FinalTable_lag.adj_cap(j_lag-1);
            FinalTable_lag.lag_intan(j_lag) = FinalTable_lag.adj_intan(j_lag-1);

    else
        FinalTable_lag.lag_cap(j_lag)= NaN;
                FinalTable_lag.lag_intan(j_lag)=NaN;
    end
    lag_cap_each =FinalTable_lag(j_lag,:) ;
    data_2= vertcat(data_2, lag_cap_each);
end
end
clear data;
data = data_2;
clear data_2;

%% filters for correct calculation of value added

t1 = 0.01;
toDelete = data.material <t1;
data(toDelete,:) = [];
clear toDelete;
toDelete = data.adj_value <t1;
data(toDelete,:) = [];
clear toDelete;

%% create the order of company
gvkey_company = data.GlobalCompanyKey;
gvkey_company = unique(gvkey_company,'rows');
n_order = size(gvkey_company,1);
gvkey_company(:,2) = 1:n_order;
gvkey_company = array2table(gvkey_company,'VariableNames',...
    {'GlobalCompanyKey','company'});
data = outerjoin(data,gvkey_company,'Type','left');
data.Properties.VariableNames{'GlobalCompanyKey_data'} = 'GlobalCompanyKey';
data.GlobalCompanyKey_gvkey_company=[];

data.Properties.VariableNames{'DataYear_Fiscal'} = 'year';
data.Properties.VariableNames{'adj_inv'} = 'i';
data.Properties.VariableNames{'Employees'} = 'l';
data.Properties.VariableNames{'adj_value'} = 'y';
data.Properties.VariableNames{'lag_cap'} = 'k';
data.Properties.VariableNames{'lag_intan'} = 't';
% remove the year before 1986
t_year = 1986;
toDelete = data.year <t_year;
data(toDelete,:) = [];
```

```
clear toDelete;

data.y = log(data.y);
data.i = log(data.i);
data.l = log(data.l);
data.k = log(data.k);
data.t = log(data.t);

toDelete = isnan(data.y);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.i);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.l);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.k);
data(toDelete,:) = [];
clear toDelete;
toDelete = isnan(data.t);
data(toDelete,:) = [];
clear toDelete;


%% export
writetable(data,'data_part2.xlsx');
```

c. **Part 3:**

```
 %% clean the command window and all variables in the workspace, running time 1 mins
clc; clearvars; close all;

%% Input data
firstexit=1966;
data = readtable('data_part2.xlsx');
data = data(:,{'CompanyName' 'company' 'year' 'sic' 'y' 'l' 'k' 'i' 't' 'GlobalCompanyKey' 'year2'
'Compustatage'});

%% generate 3 digit SIC codes
data.sic = floor(data.sic ./10);
last = max(data.year);

%% define exit
data.exit=zeros(size(data, 1), 1);
number_exit = size(data,1);
```

```
for loop_exit = 1:number_exit
if data.year(loop_exit) == data.year2(loop_exit) && data.year(loop_exit) ~= last
   data.exit(loop_exit)=1;
else
end
end

%% Generate required inputs

data.i2=data.i .^2;
data.k2=data.k .^2;
data.t2=data.t .^2;

data.ik=data.i .* data.k;
data.it=data.i .* data.t;
data.kt=data.k .* data.t;
data.ikt=data.i .* data.k .* data.t;

data = data(:,{'CompanyName' 'company' 'year' 'sic' 'y' 'l' 'k' 'i' 't' 'i2' 'k2' 't2' 'ik' 'it' 'kt' 'ikt'
'GlobalCompanyKey' 'year2' 'Compustatage' 'exit'});

data.ind_year = data.year .*1000 + data.sic;
data.capital = zeros(size(data, 1), 1);
data.intancapital = zeros(size(data, 1), 1);
data.labor = zeros(size(data, 1), 1);

%% Remove after year t
t_year = 2017;
toDelete = data.year >t_year;
data(toDelete,:) = [];
clear toDelete;


%% Min Max
min = min(data.year)+1;
max = max(data.year);
size_year = max-min+1;

%% create matrix betas
data.TFP = NaN(size(data, 1), 1);
betas = zeros(2,size_year);
col = 0;

%% Renname
data.Properties.VariableNames{'GlobalCompanyKey'} = 'gvkey';
```

```
%% export

writetable(data,'data_part3_Minh.xlsx');
```

4. Stata Code to Compute TFP with Intangible Capital

```
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Codes Part 1 version 2\data_part3_Minh.xlsx", sheet("Sheet1") firstrow

tsset company year


scalar min= 1987
scalar max=2017
scalar size=max-min+1
matrix betas=J(3,size,0)
scalar firstexit=1966
scalar col=0


* expanding window production function estimation
forvalues q=`=min'/`=max'{

        * generate exit probabilities
        if `q'>`=firstexit' {
                probit exit i k t ik it kt ikt i2 k2 t2 if year <=`q'
                predict exit_prob if year <=`q'
        }
        else {
        gen exit_prob = 0
        }


        * first stage regression, estimate labor coef.
        areg y l k i t t2 i2 k2 ik it kt ikt  if year <=`q', absorb(ind_year)

        predict res, residuals
        ereturn list
        matrix betas_1=e(b)
        scalar col=col+1
        matrix betas[1,col]=betas_1[1,1]
        replace labor=betas_1[1,1] if year==`q'

        * second stage regression to estimate the coef. for capital and intangible capital
```

```
		gen Q= _b[i]*i + _b[k]*k + _b[t]*t + _b[i2]*i2 + _b[k2]*k2 + _b[t2]*t2 + _b[ik]*ik
+_b[it]*it + _b[kt]*kt + _b[ikt]*ikt
		gen y_al= Q + res
		nl (y_al = {b_0=0} + {b_1=0.3}*k + {b_2=0.3}*t + {b_3=0.5}*(L.Q - {b_1=0.3}*L.k
- {b_2=0.3}*L.t )+{b_exit=0}*L.exit_prob) if year <=`q' & !missing(L.k) & !missing(L.t) &
!missing(L.exit_prob)

		ereturn list
		matrix betas_2=e(b)
		matrix betas[2,col]=betas_2[1,2]
		matrix betas[3,col]=betas_2[1,3]

		replace capital=betas_2[1,2] if year==`q'
		replace intancapital=betas_2[1,3] if year==`q'

* compute TFP

		replace TFP=y_al-k*capital -t*intancapital if TFP==.

		drop res Q exit_prob y_al

		}

matrix b=betas'

svmat b

export excel b1 b2 b3 using "C:\Users\nguye\Google Drive\thesis\Draft Code\Codes Part 1
version 2\coeff.xlsx", firstrow(variables)

drop if TFP==.
drop if year==1986

* delete observations if there are fewer than five observations from that industry
* (due to industry specific time dummies)
bysort year sic: gen a=_N
drop if a<5

destring gvkey, replace
keep gvkey year TFP
rename year fyear
order gvkey fyear TFP
sort gvkey fyear

* TFP estimates
```

outsheet using "Minh_TFP_intan_capital.csv", comma replace

II. The Code for the Analysis in Return

1. The SAS code to get the Data

```
libname maindata "/wrds/comp/sasdata/naa";
data compustat_data(keep= fyear gvkey conm FIC CAPX PPEGT AT EMP INVT XRD PPENB
PPENLI LT TXDITC PSTKL PSTKRV PSTK CSHO AJEX IB GP DVP TXDI MKVALT
CSHO PRCC_F DLTT  EXCHG);
set maindata.funda;
where DATAFMT='STD' and POPSRC='D' and CONSOL='C' and INDFMT='INDL' and
CURCD='USD';
run;

data compustat_names(keep= gvkey sic naics conm year1 year2);
set maindata.names;
run;
```

2. CRSP data:

The file CRSP_1970_NEW.dta is the Stata data that we get from CRSP in WRDS with the variable RET, Permco, Permno, Date

3. Stata Code to create the Annual Return for Contemporaneous Case:

```
use "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with Table\CRSP_1970_NEW.dta"

gen int fyear = year(date)


gen log_ret = log(1+ret)

egen annual_ret = total(log_ret), by(permno fyear)

replace annual_ret = exp(annual_ret) - 1


export delimited using "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\CRSP_Contemporaneous_Return_1970.csv", replace
```

4. MatLab Code for Average Contemporaneous Excess Return in TFP sorted Portfolios:

```
%% clean the command window and all variables in the workspace
clc; clearvars; close all;

%% Import data

data = readtable('Minh_TFP_intan_capital.csv');
```

```
data.TFP = exp(data.TFP );
dataFF = readtable('F-F_Research_Data_Factors.csv');
data_CRSP_Annual_Return = readtable('CRSP_Contemporaneous_Return_1970.csv');
data_permno_gvkey = readtable('permno_gvkey.xlsx');
data_CRSP_Annual_Return = data_CRSP_Annual_Return(:,{'permno' 'permco'...
    'fyear' 'annual_ret'});
data_permno_gvkey = data_permno_gvkey(:,{'StandardAndPoor_sIdentifier'...
    'HistoricalCRSPPERMNOLinkToCOMPUSTATRecord'});
data_permno_gvkey.Properties.VariableNames{'HistoricalCRSPPERMNOLinkToCOMPUSTA
TRecord'} = 'permno';
data_permno_gvkey.Properties.VariableNames{'StandardAndPoor_sIdentifier'} = 'gvkey';
data_permno_gvkey = unique(data_permno_gvkey,'rows');

%%
data_CRSP = outerjoin(data_CRSP_Annual_Return,data_permno_gvkey,'Type','left');
data_CRSP.permno_data_CRSP_Annual_Return = [];
data_CRSP.permco = [];
data_CRSP.permno_data_permno_gvkey = [];
data_CRSP = unique(data_CRSP,'rows');


%% remove the year before the min year of data
t_year = min(data.fyear);
toDelete = data_CRSP.fyear <t_year;
data_CRSP(toDelete,:) = [];
clear toDelete;

%% Return with TFP join
data_CRSP_TFP = outerjoin(data,data_CRSP,'Type','left');
data_CRSP_TFP = data_CRSP_TFP(:,{'gvkey_data' 'fyear_data'...
    'TFP' 'annual_ret'});

toDelete = isnan(data_CRSP_TFP.annual_ret);
data_CRSP_TFP(toDelete,:) = [];
clear toDelete;
data_CRSP_TFP.Properties.VariableNames{'gvkey_data'} = 'gvkey';
data_CRSP_TFP.Properties.VariableNames{'fyear_data'} = 'fyear';


%% Merge with RF
data_CRSP_TFP.annual_ret=data_CRSP_TFP.annual_ret .*100;

dataFF = dataFF(:,{'Year' 'RF'});
dataFF.Properties.VariableNames{'Year'} = 'fyear';
data_CRSP_RF = outerjoin(data_CRSP_TFP,dataFF,'Type','left');
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_TFP'} = 'fyear';
```

```
data_CRSP_RF.fyear_dataFF =[];
data_CRSP_RF.excess_ret = data_CRSP_RF.annual_ret - data_CRSP_RF.RF;

clear dataFF data_CRSP_TFP data_permno_gvkey data_CRSP data_CRSP_Annual_Return
data;
%% Export
%writetable(data_CRSP_RF,'data_CRSP_RF.xlsx');

%% Input Compustat Data

Compustat_names = readtable('COMPUSTAT_NAMES_PART2_ver3.csv');
Compustat_data = readtable('COMPUSTAT_DATA_PART2_ver3.csv');

t_year = 2017;
toDelete = data_CRSP_RF.fyear >t_year;
data_CRSP_RF(toDelete,:) = [];
clear toDelete;
%% Joint Compustat data

Compustat_names.conm=[];
Compustat_names.naics=[];

data_WRDS = outerjoin(Compustat_data,Compustat_names,'Type','left');
data_WRDS.gvkey_Compustat_names =[];
data_WRDS.Properties.VariableNames{'gvkey_Compustat_data'} = 'gvkey';
clear Compustat_names Compustat_data

%% Remove Micro firm

data_WRDS.Market_cap = data_WRDS.csho .* data_WRDS.prcc_f;
data_Market_Cap = data_WRDS(:,{'gvkey' 'fyear' 'Market_cap' 'exchg'});
data_CRSP_RF = outerjoin(data_CRSP_RF,data_Market_Cap,'Type','left');
data_CRSP_RF.Properties.VariableNames{'gvkey_data_CRSP_RF'} = 'gvkey';
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_RF'} = 'fyear';
data_CRSP_RF.gvkey_data_Market_Cap = [];
data_CRSP_RF.fyear_data_Market_Cap = [];


clear data_Market_Cap

toDelete = isnan(data_CRSP_RF.Market_cap);
data_CRSP_RF(toDelete,:) = [];
clear toDelete;

data_CRSP_RF = sortrows(data_CRSP_RF,'fyear','ascend');
```

```
data_NYSE=data_CRSP_RF(data_CRSP_RF.exchg==11,:);


year = data_NYSE.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
data_3=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_NYSE(data_NYSE.fyear == year_i,:);
percentile_20 = prctile(FinalTable_k.Market_cap,20);
percentile_20_each = array2table(percentile_20);
percentile_20_each.year=year_i;
data_3=vertcat(data_3,percentile_20_each);
FinalTable_k_nyse_remove_micro = FinalTable_k(FinalTable_k.Market_cap
>=percentile_20,:);
data_2=vertcat(data_2,FinalTable_k_nyse_remove_micro);
end
data_nyse_remove_micro = data_2;
percentile_20_list = data_3;
clear percentile_20_each data_3 year n_NYSE data_2 i_NYSE year_i FinalTable_k
percentile_20 FinalTable_k_nyse_remove_micro

%% Create Break Point for each year

year = data_nyse_remove_micro.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_nyse_remove_micro(data_nyse_remove_micro.fyear == year_i,:);
percentile_all_tfp = prctile(FinalTable_k.TFP,[10 20 30 40 50 60 70 80 90 100]);
percentile_all_tfp_each = array2table(percentile_all_tfp);
percentile_all_tfp_each.year=year_i;
data_2=vertcat(data_2,percentile_all_tfp_each);
end
break_point_list = data_2;

clear year n_NYSE data_2 i_NYSE year_i FinalTable_k percentile_all_tfp
percentile_all_tfp_each

%% Creating Low - High Portfolio
% Portfolio 10
year = break_point_list.year;
```

```
year = unique(year,'rows');
n = size(year,1);
data_2=table();

for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(FinalTable_k.TFP<=break_point_i,:); % 10 portfolio
data_2=vertcat(data_2,FinalTable_k);
end

portfolio_10 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i

% Portfolio 20
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_20 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 30
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
```

```
end
portfolio_30 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 40
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_40 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 50
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_50 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 60
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
```

```
break_point_i = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_60 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 70
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_70 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 80
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_80 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 90
year = break_point_list.year;
year = unique(year,'rows');
```

```
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_90 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 100
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp10(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_100 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j


%% Create Table Excess Return for TFP-Sorted Porfolio

% Average Future excess return
excess_return_sorted_port = table();

year = portfolio_10.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_10(portfolio_10.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
```

```
end
excess_return_sorted_port.Port_Low(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_Low(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_Low(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_Low(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_20.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_20(portfolio_20.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_2(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_2(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_2(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_2(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_30.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_30(portfolio_30.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_3(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_3(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_3(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_3(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_40.fyear;
year = unique(year,'rows');
```

```
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_40(portfolio_40.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_4(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_4(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_4(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_4(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_50.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_50(portfolio_50.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_5(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_5(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_5(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_5(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_60.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_60(portfolio_60.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_6(1) = mean(data_2.excess_ret);
```

```
excess_return_sorted_port.Port_6(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_6(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_6(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_70.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_70(portfolio_70.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_7(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_7(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_7(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_7(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_80.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_80(portfolio_80.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_8(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_8(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_8(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_8(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_90.fyear;
year = unique(year,'rows');
```

```
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_90(portfolio_90.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_9(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_9(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_9(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_9(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_100.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_100(portfolio_100.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_High(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_High(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_High(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_High(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

%% export
writetable(excess_return_sorted_port,'Contemporaneus_return.xlsx');
writetable(portfolio_10,'Contemp_portfolio_10.xlsx');
writetable(portfolio_20,'Contemp_portfolio_20.xlsx');
writetable(portfolio_30,'Contemp_portfolio_30.xlsx');
writetable(portfolio_40,'Contemp_portfolio_40.xlsx');
writetable(portfolio_50,'Contemp_portfolio_50.xlsx');
writetable(portfolio_60,'Contemp_portfolio_60.xlsx');
writetable(portfolio_70,'Contemp_portfolio_70.xlsx');
writetable(portfolio_80,'Contemp_portfolio_80.xlsx');
writetable(portfolio_90,'Contemp_portfolio_90.xlsx');
```

```
writetable(portfolio_100,'Contemp_portfolio_100.xlsx');



function y=NWtest(ret,lag,h0)

T=size(ret,1);
vv=var(ret);

for l=1:1:lag

   cc=cov(ret(1:end-l),ret(l+1:end));
   vv=vv+2*(1-l/lag)*cc(1,2);
end

y=(mean(ret)-h0)/sqrt(vv)*sqrt(T);
end
```

5. **Stata Code to make Future annual rate from July year t+1 to June yeare t+2**

```
use "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with Table\CRSP_1970_NEW.dta"

gen int fyear = year(date)-1

replace fyear = fyear -1 if month(date) < 7


gen log_ret = log(1+ret)

egen annual_ret = total(log_ret), by(permno fyear)

replace annual_ret = exp(annual_ret) - 1

export delimited using "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\CRSP_Future_Annual_Return_1970.csv", replace
```

6. **Stata Code to make future risk free rate from July year t+1 to June year t+2, The file F-F_Research_Data_Factors_Monthly.csv was downloaded from Fama French Website**

```
import delimited "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with Table\F-
F_Research_Data_Factors_Monthly.csv"

tostring v1, replace

gen date = date(v1, "YM")
```

```
format date %td


gen int fyear = year(date)-1

replace fyear = fyear -1 if month(date) < 7


gen log_rf = log(1+rf/100)

egen annual_rf = total(log_rf), by(fyear)

replace annual_rf = exp(annual_rf)-1

replace annual_rf = annual_rf *100


export delimited using "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\FF_Future_Risk_Free_Rate.csv", replace
```

7. MatLab Code for Average Future Excess Return in TFP sorted Portfolios:

```
%% clean the command window and all variables in the workspace
clc; clearvars; close all;

%% Import data

data = readtable('Minh_TFP_intan_capital.csv');
data.TFP = exp(data.TFP );
dataFF = readtable('FF_Future_Risk_Free_Rate.csv');
dataFF.Properties.VariableNames{'fyear'} = 'Year';
dataFF.Properties.VariableNames{'annual_rf'} = 'RF';
dataFF = dataFF(:,{'Year' 'RF'});
dataFF = unique(dataFF,'rows');

data_CRSP_Annual_Return = readtable('CRSP_Future_Annual_Return_1970.csv');
data_permno_gvkey = readtable('permno_gvkey.xlsx');
data_CRSP_Annual_Return = data_CRSP_Annual_Return(:,{'permno' 'permco'...
    'fyear' 'annual_ret'});
data_permno_gvkey = data_permno_gvkey(:,{'StandardAndPoor_sIdentifier'...
    'HistoricalCRSPPERMNOLinkToCOMPUSTATRecord'});
data_permno_gvkey.Properties.VariableNames{'HistoricalCRSPPERMNOLinkToCOMPUSTA
TRecord'} = 'permno';
data_permno_gvkey.Properties.VariableNames{'StandardAndPoor_sIdentifier'} = 'gvkey';
```

```
data_permno_gvkey = unique(data_permno_gvkey,'rows');

%%
data_CRSP = outerjoin(data_CRSP_Annual_Return,data_permno_gvkey,'Type','left');
data_CRSP.permno_data_CRSP_Annual_Return = [];
data_CRSP.permco = [];
data_CRSP.permno_data_permno_gvkey = [];
data_CRSP = unique(data_CRSP,'rows');


%% remove the year before the min year of data
t_year = min(data.fyear);
toDelete = data_CRSP.fyear <t_year;
data_CRSP(toDelete,:) = [];
clear toDelete;

%% Return with TFP join
data_CRSP_TFP = outerjoin(data,data_CRSP,'Type','left');
data_CRSP_TFP = data_CRSP_TFP(:,{'gvkey_data' 'fyear_data'...
    'TFP' 'annual_ret'});

toDelete = isnan(data_CRSP_TFP.annual_ret);
data_CRSP_TFP(toDelete,:) = [];
clear toDelete;
data_CRSP_TFP.Properties.VariableNames{'gvkey_data'} = 'gvkey';
data_CRSP_TFP.Properties.VariableNames{'fyear_data'} = 'fyear';


%% Merge with RF
data_CRSP_TFP.annual_ret=data_CRSP_TFP.annual_ret .*100;

dataFF = dataFF(:,{'Year' 'RF'});
dataFF.Properties.VariableNames{'Year'} = 'fyear';
data_CRSP_RF = outerjoin(data_CRSP_TFP,dataFF,'Type','left');
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_TFP'} = 'fyear';
data_CRSP_RF.fyear_dataFF =[];
data_CRSP_RF.excess_ret = data_CRSP_RF.annual_ret - data_CRSP_RF.RF;

clear dataFF data_CRSP_TFP data_permno_gvkey data_CRSP data_CRSP_Annual_Return
data;
%% Export
%writetable(data_CRSP_RF,'data_CRSP_RF.xlsx');

%% Input Compustat Data

Compustat_names = readtable('COMPUSTAT_NAMES_PART2_ver3.csv');
```

```
Compustat_data = readtable('COMPUSTAT_DATA_PART2_ver3.csv');

t_year = 2017;
toDelete = data_CRSP_RF.fyear >t_year;
data_CRSP_RF(toDelete,:) = [];
clear toDelete;
%% Joint Compustat data

Compustat_names.conm=[];
Compustat_names.naics=[];

data_WRDS = outerjoin(Compustat_data,Compustat_names,'Type','left');
data_WRDS.gvkey_Compustat_names =[];
data_WRDS.Properties.VariableNames{'gvkey_Compustat_data'} = 'gvkey';
clear Compustat_names Compustat_data

%% Remove Micro firm

data_WRDS.Market_cap = data_WRDS.csho .* data_WRDS.prcc_f;
data_Market_Cap = data_WRDS(:,{'gvkey' 'fyear' 'Market_cap' 'exchg'});
data_CRSP_RF = outerjoin(data_CRSP_RF,data_Market_Cap,'Type','left');
data_CRSP_RF.Properties.VariableNames{'gvkey_data_CRSP_RF'} = 'gvkey';
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_RF'} = 'fyear';
data_CRSP_RF.gvkey_data_Market_Cap = [];
data_CRSP_RF.fyear_data_Market_Cap = [];


clear data_Market_Cap

toDelete = isnan(data_CRSP_RF.Market_cap);
data_CRSP_RF(toDelete,:) = [];
clear toDelete;

data_CRSP_RF = sortrows(data_CRSP_RF,'fyear','ascend');

data_NYSE=data_CRSP_RF(data_CRSP_RF.exchg==11,:);


year = data_NYSE.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
data_3=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_NYSE(data_NYSE.fyear == year_i,:);
```

```
percentile_20 = prctile(FinalTable_k.Market_cap,20);
percentile_20_each = array2table(percentile_20);
percentile_20_each.year=year_i;
data_3=vertcat(data_3,percentile_20_each);
FinalTable_k_nyse_remove_micro = FinalTable_k(FinalTable_k.Market_cap
>=percentile_20,:);
data_2=vertcat(data_2,FinalTable_k_nyse_remove_micro);
end
data_nyse_remove_micro = data_2;
percentile_20_list = data_3;
clear percentile_20_each data_3 year n_NYSE data_2 i_NYSE year_i FinalTable_k
percentile_20 FinalTable_k_nyse_remove_micro

%% Create Break Point for each year

year = data_nyse_remove_micro.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_nyse_remove_micro(data_nyse_remove_micro.fyear == year_i,:);
percentile_all_tfp = prctile(FinalTable_k.TFP,[10 20 30 40 50 60 70 80 90 100]);
percentile_all_tfp_each = array2table(percentile_all_tfp);
percentile_all_tfp_each.year=year_i;
data_2=vertcat(data_2,percentile_all_tfp_each);
end
break_point_list = data_2;

clear year n_NYSE data_2 i_NYSE year_i FinalTable_k percentile_all_tfp
percentile_all_tfp_each

%% Creating Low - High Portfolio
% Portfolio 10
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();

for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(FinalTable_k.TFP<=break_point_i,:); % 10 portfolio
data_2=vertcat(data_2,FinalTable_k);
end
```

```
portfolio_10 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i

% Portfolio 20
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_20 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 30
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_30 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 40
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
```

```
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_40 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 50
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_50 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 60
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_60 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 70
year = break_point_list.year;
```

```
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_70 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 80
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_80 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 90
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
```

```
portfolio_90 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 100
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp10(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_100 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j


%% Create Table Excess Return for TFP-Sorted Porfolio

% Average Future excess return
excess_return_sorted_port = table();

year = portfolio_10.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_10(portfolio_10.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_Low(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_Low(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_Low(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_Low(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_20.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
```

```
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_20(portfolio_20.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_2(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_2(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_2(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_2(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_30.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_30(portfolio_30.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_3(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_3(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_3(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_3(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_40.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_40(portfolio_40.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_4(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_4(2) = NWtest(data_2.excess_ret,1,0);
```

```
excess_return_sorted_port.Port_4(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_4(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_50.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_50(portfolio_50.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_5(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_5(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_5(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_5(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_60.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_60(portfolio_60.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_6(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_6(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_6(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_6(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_70.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
```

```
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_70(portfolio_70.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_7(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_7(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_7(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_7(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_80.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_80(portfolio_80.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_8(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_8(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_8(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_8(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_90.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_90(portfolio_90.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_9(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_9(2) = NWtest(data_2.excess_ret,1,0);
```

```
excess_return_sorted_port.Port_9(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_9(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_100.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_100(portfolio_100.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_High(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_High(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_High(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_High(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

%% export
%writetable(excess_return_sorted_port,'future_return_all_state.xlsx');

writetable(portfolio_10,'future_portfolio_10.xlsx');
writetable(portfolio_20,'future_portfolio_20.xlsx');
writetable(portfolio_30,'future_portfolio_30.xlsx');
writetable(portfolio_40,'future_portfolio_40.xlsx');
writetable(portfolio_50,'future_portfolio_50.xlsx');
writetable(portfolio_60,'future_portfolio_60.xlsx');
writetable(portfolio_70,'future_portfolio_70.xlsx');
writetable(portfolio_80,'future_portfolio_80.xlsx');
writetable(portfolio_90,'future_portfolio_90.xlsx');
writetable(portfolio_100,'future_portfolio_100.xlsx');
writetable(data_CRSP_RF,'future_data_CRSP_RF.xlsx');




function y=NWtest(ret,lag,h0)

T=size(ret,1);
vv=var(ret);
```

```
for l=1:1:lag

   cc=cov(ret(1:end-l),ret(l+1:end));
   vv=vv+2*(1-l/lag)*cc(1,2);
end

y=(mean(ret)-h0)/sqrt(vv)*sqrt(T);
end
```

8. MatLab Code for Average Future Excess Return in TFP sorted Portfolios in Expansion Period:

```
%% clean the command window and all variables in the workspace
clc; clearvars; close all;

%% Import data

data = readtable('Minh_TFP_intan_capital.csv');
data.TFP = exp(data.TFP );
dataFF = readtable('FF_Future_Risk_Free_Rate.csv');
dataFF.Properties.VariableNames{'fyear'} = 'Year';
dataFF.Properties.VariableNames{'annual_rf'} = 'RF';
dataFF = dataFF(:,{'Year' 'RF'});
dataFF = unique(dataFF,'rows');

data_CRSP_Annual_Return = readtable('CRSP_Future_Annual_Return_1970.csv');
data_permno_gvkey = readtable('permno_gvkey.xlsx');
data_CRSP_Annual_Return = data_CRSP_Annual_Return(:,{'permno' 'permco'...
   'fyear' 'annual_ret'});
data_permno_gvkey = data_permno_gvkey(:,{'StandardAndPoor_sIdentifier'...
   'HistoricalCRSPPERMNOLinkToCOMPUSTATRecord'});
data_permno_gvkey.Properties.VariableNames{'HistoricalCRSPPERMNOLinkToCOMPUSTA
TRecord'} = 'permno';
data_permno_gvkey.Properties.VariableNames{'StandardAndPoor_sIdentifier'} = 'gvkey';
data_permno_gvkey = unique(data_permno_gvkey,'rows');

%%
data_CRSP = outerjoin(data_CRSP_Annual_Return,data_permno_gvkey,'Type','left');
data_CRSP.permno_data_CRSP_Annual_Return = [];
data_CRSP.permco = [];
data_CRSP.permno_data_permno_gvkey = [];
```

```
data_CRSP = unique(data_CRSP,'rows');


%% remove the year before the min year of data
t_year = min(data.fyear);
toDelete = data_CRSP.fyear <t_year;
data_CRSP(toDelete,:) = [];
clear toDelete;

%% Return with TFP join
data_CRSP_TFP = outerjoin(data,data_CRSP,'Type','left');
data_CRSP_TFP = data_CRSP_TFP(:,{'gvkey_data' 'fyear_data'...
    'TFP' 'annual_ret'});

toDelete = isnan(data_CRSP_TFP.annual_ret);
data_CRSP_TFP(toDelete,:) = [];
clear toDelete;
data_CRSP_TFP.Properties.VariableNames{'gvkey_data'} = 'gvkey';
data_CRSP_TFP.Properties.VariableNames{'fyear_data'} = 'fyear';


%% Merge with RF
data_CRSP_TFP.annual_ret=data_CRSP_TFP.annual_ret .*100;

dataFF = dataFF(:,{'Year' 'RF'});
dataFF.Properties.VariableNames{'Year'} = 'fyear';
data_CRSP_RF = outerjoin(data_CRSP_TFP,dataFF,'Type','left');
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_TFP'} = 'fyear';
data_CRSP_RF.fyear_dataFF =[];
data_CRSP_RF.excess_ret = data_CRSP_RF.annual_ret - data_CRSP_RF.RF;

clear dataFF data_CRSP_TFP data_permno_gvkey data_CRSP data_CRSP_Annual_Return
data;
%% Export
%writetable(data_CRSP_RF,'data_CRSP_RF.xlsx');

%% Input Compustat Data

Compustat_names = readtable('COMPUSTAT_NAMES_PART2_ver3.csv');
Compustat_data = readtable('COMPUSTAT_DATA_PART2_ver3.csv');

t_year = 2017;
toDelete = data_CRSP_RF.fyear >t_year;
data_CRSP_RF(toDelete,:) = [];
clear toDelete;
%% Joint Compustat data
```

```
Compustat_names.conm=[];
Compustat_names.naics=[];

data_WRDS = outerjoin(Compustat_data,Compustat_names,'Type','left');
data_WRDS.gvkey_Compustat_names =[];
data_WRDS.Properties.VariableNames{'gvkey_Compustat_data'} = 'gvkey';
clear Compustat_names Compustat_data

%% Expansion

data_CRSP_RF = data_CRSP_RF(data_CRSP_RF.fyear ~= 1990&data_CRSP_RF.fyear
~=2001&data_CRSP_RF.fyear ~=2007&data_CRSP_RF.fyear ~=2008,:);


%% Remove Micro firm

data_WRDS.Market_cap = data_WRDS.csho .* data_WRDS.prcc_f;
data_Market_Cap = data_WRDS(:,{'gvkey' 'fyear' 'Market_cap' 'exchg'});
data_CRSP_RF = outerjoin(data_CRSP_RF,data_Market_Cap,'Type','left');
data_CRSP_RF.Properties.VariableNames{'gvkey_data_CRSP_RF'} = 'gvkey';
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_RF'} = 'fyear';
data_CRSP_RF.gvkey_data_Market_Cap = [];
data_CRSP_RF.fyear_data_Market_Cap = [];


clear data_Market_Cap

toDelete = isnan(data_CRSP_RF.Market_cap);
data_CRSP_RF(toDelete,:) = [];
clear toDelete;

data_CRSP_RF = sortrows(data_CRSP_RF,'fyear','ascend');

data_NYSE=data_CRSP_RF(data_CRSP_RF.exchg==11,:);


year = data_NYSE.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
data_3=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_NYSE(data_NYSE.fyear == year_i,:);
percentile_20 = prctile(FinalTable_k.Market_cap,20);
```

```
percentile_20_each = array2table(percentile_20);
percentile_20_each.year=year_i;
data_3=vertcat(data_3,percentile_20_each);
FinalTable_k_nyse_remove_micro = FinalTable_k(FinalTable_k.Market_cap
>=percentile_20,:);
data_2=vertcat(data_2,FinalTable_k_nyse_remove_micro);
end
data_nyse_remove_micro = data_2;
percentile_20_list = data_3;
clear percentile_20_each data_3 year n_NYSE data_2 i_NYSE year_i FinalTable_k
percentile_20 FinalTable_k_nyse_remove_micro

%% Create Break Point for each year

year = data_nyse_remove_micro.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_nyse_remove_micro(data_nyse_remove_micro.fyear == year_i,:);
percentile_all_tfp = prctile(FinalTable_k.TFP,[10 20 30 40 50 60 70 80 90 100]);
percentile_all_tfp_each = array2table(percentile_all_tfp);
percentile_all_tfp_each.year=year_i;
data_2=vertcat(data_2,percentile_all_tfp_each);
end
break_point_list = data_2;

clear year n_NYSE data_2 i_NYSE year_i FinalTable_k percentile_all_tfp
percentile_all_tfp_each

%% Creating Low - High Portfolio
% Portfolio 10
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();

for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(FinalTable_k.TFP<=break_point_i,:); % 10 portfolio
data_2=vertcat(data_2,FinalTable_k);
end
```

```
portfolio_10 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i

% Portfolio 20
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_20 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 30
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_30 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 40
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
```

```
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_40 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 50
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_50 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 60
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_60 = data_2;
```

```
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 70
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_70 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 80
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_80 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 90
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
```

```
break_point_i = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_90 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j

% Portfolio 100
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp10(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_100 = data_2;

clear year n data_2 i year_i FinalTable_k break_point_i break_point_j


%% Create Table Excess Return for TFP-Sorted Porfolio

% Average Future excess return
excess_return_sorted_port = table();

year = portfolio_10.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_10(portfolio_10.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_Low(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_Low(2) = NWtest(data_2.excess_ret,1,0);
```

```
excess_return_sorted_port.Port_Low(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_Low(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_20.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_20(portfolio_20.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_2(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_2(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_2(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_2(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_30.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_30(portfolio_30.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_3(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_3(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_3(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_3(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_40.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
```

```
year_i = year(i,:);
FinalTable_k = portfolio_40(portfolio_40.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_4(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_4(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_4(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_4(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_50.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_50(portfolio_50.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_5(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_5(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_5(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_5(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_60.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_60(portfolio_60.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_6(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_6(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_6(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_6(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);
```

```
clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_70.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_70(portfolio_70.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_7(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_7(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_7(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_7(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_80.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_80(portfolio_80.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_8(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_8(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_8(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_8(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_90.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
```

```
year_i = year(i,:);
FinalTable_k = portfolio_90(portfolio_90.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_9(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_9(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_9(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_9(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_100.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_100(portfolio_100.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_High(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_High(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_High(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_High(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

%%

function y=NWtest(ret,lag,h0)

T=size(ret,1);
vv=var(ret);

for l=1:1:lag

   cc=cov(ret(1:end-l),ret(l+1:end));
   vv=vv+2*(1-l/lag)*cc(1,2);
end

y=(mean(ret)-h0)/sqrt(vv)*sqrt(T);
```

```
end
```

9. **MatLab Code for Average Future Excess Return in TFP sorted Portfolios in Recession Period:**

```
%% clean the command window and all variables in the workspace
clc; clearvars; close all;

%% Import data

data = readtable('Minh_TFP_intan_capital.csv');
data.TFP = exp(data.TFP );
dataFF = readtable('FF_Future_Risk_Free_Rate.csv');
dataFF.Properties.VariableNames{'fyear'} = 'Year';
dataFF.Properties.VariableNames{'annual_rf'} = 'RF';
dataFF = dataFF(:,{'Year' 'RF'});
dataFF = unique(dataFF,'rows');

data_CRSP_Annual_Return = readtable('CRSP_Future_Annual_Return_1970.csv');
data_permno_gvkey = readtable('permno_gvkey.xlsx');
data_CRSP_Annual_Return = data_CRSP_Annual_Return(:,{'permno' 'permco'...
    'fyear' 'annual_ret'});
data_permno_gvkey = data_permno_gvkey(:,{'StandardAndPoor_sIdentifier'...
    'HistoricalCRSPPERMNOLinkToCOMPUSTATRecord'});
data_permno_gvkey.Properties.VariableNames{'HistoricalCRSPPERMNOLinkToCOMPUSTA
TRecord'} = 'permno';
data_permno_gvkey.Properties.VariableNames{'StandardAndPoor_sIdentifier'} = 'gvkey';
data_permno_gvkey = unique(data_permno_gvkey,'rows');

%%
data_CRSP = outerjoin(data_CRSP_Annual_Return,data_permno_gvkey,'Type','left');
data_CRSP.permno_data_CRSP_Annual_Return = [];
data_CRSP.permco = [];
data_CRSP.permno_data_permno_gvkey = [];
data_CRSP = unique(data_CRSP,'rows');


%% remove the year before the min year of data
t_year = min(data.fyear);
toDelete = data_CRSP.fyear <t_year;
data_CRSP(toDelete,:) = [];
clear toDelete;
```

```
%% Return with TFP join
data_CRSP_TFP = outerjoin(data,data_CRSP,'Type','left');
data_CRSP_TFP = data_CRSP_TFP(:,{'gvkey_data' 'fyear_data'...
    'TFP' 'annual_ret'});

toDelete = isnan(data_CRSP_TFP.annual_ret);
data_CRSP_TFP(toDelete,:) = [];
clear toDelete;
data_CRSP_TFP.Properties.VariableNames{'gvkey_data'} = 'gvkey';
data_CRSP_TFP.Properties.VariableNames{'fyear_data'} = 'fyear';


%% Merge with RF
data_CRSP_TFP.annual_ret=data_CRSP_TFP.annual_ret .*100;

dataFF = dataFF(:,{'Year' 'RF'});
dataFF.Properties.VariableNames{'Year'} = 'fyear';
data_CRSP_RF = outerjoin(data_CRSP_TFP,dataFF,'Type','left');
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_TFP'} = 'fyear';
data_CRSP_RF.fyear_dataFF =[];
data_CRSP_RF.excess_ret = data_CRSP_RF.annual_ret - data_CRSP_RF.RF;

clear dataFF data_CRSP_TFP data_permno_gvkey data_CRSP data_CRSP_Annual_Return
data;
%% Export
%writetable(data_CRSP_RF,'data_CRSP_RF.xlsx');

%% Input Compustat Data

Compustat_names = readtable('COMPUSTAT_NAMES_PART2_ver3.csv');
Compustat_data = readtable('COMPUSTAT_DATA_PART2_ver3.csv');

t_year = 2017;
toDelete = data_CRSP_RF.fyear >t_year;
data_CRSP_RF(toDelete,:) = [];
clear toDelete;
%% Joint Compustat data

Compustat_names.conm=[];
Compustat_names.naics=[];

data_WRDS = outerjoin(Compustat_data,Compustat_names,'Type','left');
data_WRDS.gvkey_Compustat_names =[];
data_WRDS.Properties.VariableNames{'gvkey_Compustat_data'} = 'gvkey';
clear Compustat_names Compustat_data
```

```
%% Remove Micro firm

data_WRDS.Market_cap = data_WRDS.csho .* data_WRDS.prcc_f;
data_Market_Cap = data_WRDS(:,{'gvkey' 'fyear' 'Market_cap' 'exchg'});
data_CRSP_RF = outerjoin(data_CRSP_RF,data_Market_Cap,'Type','left');
data_CRSP_RF.Properties.VariableNames{'gvkey_data_CRSP_RF'} = 'gvkey';
data_CRSP_RF.Properties.VariableNames{'fyear_data_CRSP_RF'} = 'fyear';
data_CRSP_RF.gvkey_data_Market_Cap = [];
data_CRSP_RF.fyear_data_Market_Cap = [];


clear data_Market_Cap

toDelete = isnan(data_CRSP_RF.Market_cap);
data_CRSP_RF(toDelete,:) = [];
clear toDelete;

data_CRSP_RF = sortrows(data_CRSP_RF,'fyear','ascend');

data_NYSE=data_CRSP_RF(data_CRSP_RF.exchg==11,:);


year = data_NYSE.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
data_3=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_NYSE(data_NYSE.fyear == year_i,:);
percentile_20 = prctile(FinalTable_k.Market_cap,20);
percentile_20_each = array2table(percentile_20);
percentile_20_each.year=year_i;
data_3=vertcat(data_3,percentile_20_each);
FinalTable_k_nyse_remove_micro = FinalTable_k(FinalTable_k.Market_cap
>=percentile_20,:);
data_2=vertcat(data_2,FinalTable_k_nyse_remove_micro);
end
data_nyse_remove_micro = data_2;
percentile_20_list = data_3;
clear percentile_20_each data_3 year n_NYSE data_2 i_NYSE year_i FinalTable_k
percentile_20 FinalTable_k_nyse_remove_micro

%% Create Break Point for each year
```

```
year = data_nyse_remove_micro.fyear;
year = unique(year,'rows');
n_NYSE = size(year,1);
data_2=table();
for i_NYSE = 1: n_NYSE
year_i = year(i_NYSE,:);
FinalTable_k = data_nyse_remove_micro(data_nyse_remove_micro.fyear == year_i,:);
percentile_all_tfp = prctile(FinalTable_k.TFP,[10 20 30 40 50 60 70 80 90 100]);
percentile_all_tfp_each = array2table(percentile_all_tfp);
percentile_all_tfp_each.year=year_i;
data_2=vertcat(data_2,percentile_all_tfp_each);
end
break_point_list = data_2;

clear year n_NYSE data_2 i_NYSE year_i FinalTable_k percentile_all_tfp
percentile_all_tfp_each

%% Creating Low - High Portfolio
% Portfolio 10
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();

for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(FinalTable_k.TFP<=break_point_i,:); % 10 portfolio
data_2=vertcat(data_2,FinalTable_k);
end

portfolio_10 = data_2;
portfolio_10 = portfolio_10(portfolio_10.fyear == 1990|portfolio_10.fyear
==2001|portfolio_10.fyear ==2007|portfolio_10.fyear ==2008,:);

clear year n data_2 i year_i FinalTable_k break_point_i

% Portfolio 20
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
```

```
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp1(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_20 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_20 = portfolio_20(portfolio_20.fyear == 1990|portfolio_20.fyear
==2001|portfolio_20.fyear ==2007|portfolio_20.fyear ==2008,:);

% Portfolio 30
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp2(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_30 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_30 = portfolio_30(portfolio_30.fyear == 1990|portfolio_30.fyear
==2001|portfolio_30.fyear ==2007|portfolio_30.fyear ==2008,:);

% Portfolio 40
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp3(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_40 = data_2;
```

```
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_40 = portfolio_40(portfolio_40.fyear == 1990|portfolio_40.fyear
==2001|portfolio_40.fyear ==2007|portfolio_40.fyear ==2008,:);

% Portfolio 50
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp4(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_50 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_50 = portfolio_50(portfolio_50.fyear == 1990|portfolio_50.fyear
==2001|portfolio_50.fyear ==2007|portfolio_50.fyear ==2008,:);

% Portfolio 60
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp5(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_60 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_60 = portfolio_60(portfolio_60.fyear == 1990|portfolio_60.fyear
==2001|portfolio_60.fyear ==2007|portfolio_60.fyear ==2008,:);

% Portfolio 70
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
```

```
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp6(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_70 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_70 = portfolio_70(portfolio_70.fyear == 1990|portfolio_70.fyear
==2001|portfolio_70.fyear ==2007|portfolio_70.fyear ==2008,:);

% Portfolio 80
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp7(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_80 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_80 = portfolio_80(portfolio_80.fyear == 1990|portfolio_80.fyear
==2001|portfolio_80.fyear ==2007|portfolio_80.fyear ==2008,:);

% Portfolio 90
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp8(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP &
FinalTable_k.TFP<=break_point_i,:);
```

```
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_90 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_90 = portfolio_90(portfolio_90.fyear == 1990|portfolio_90.fyear
==2001|portfolio_90.fyear ==2007|portfolio_90.fyear ==2008,:);

% Portfolio 100
year = break_point_list.year;
year = unique(year,'rows');
n = size(year,1);
data_2=table();
for i = 1: n
year_i = year(i,:);
FinalTable_k = data_CRSP_RF(data_CRSP_RF.fyear == year_i,:);
break_point_i = break_point_list.percentile_all_tfp10(break_point_list.year==year_i);
break_point_j = break_point_list.percentile_all_tfp9(break_point_list.year==year_i);
FinalTable_k = FinalTable_k(break_point_j<FinalTable_k.TFP,:);
data_2=vertcat(data_2,FinalTable_k);
end
portfolio_100 = data_2;
clear year n data_2 i year_i FinalTable_k break_point_i break_point_j
portfolio_100 = portfolio_100(portfolio_100.fyear == 1990|portfolio_100.fyear
==2001|portfolio_100.fyear ==2007|portfolio_100.fyear ==2008,:);


%% Create Table Excess Return for TFP-Sorted Porfolio

% Average Future excess return
excess_return_sorted_port = table();

year = portfolio_10.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_10(portfolio_10.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_Low(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_Low(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_Low(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_Low(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);
```

```
clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_20.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_20(portfolio_20.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_2(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_2(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_2(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_2(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_30.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_30(portfolio_30.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_3(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_3(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_3(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_3(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_40.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_40(portfolio_40.fyear == year_i,:);
```

```
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_4(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_4(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_4(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_4(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_50.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_50(portfolio_50.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_5(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_5(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_5(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_5(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_60.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_60(portfolio_60.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_6(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_6(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_6(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_6(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i
```

```
year = portfolio_70.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_70(portfolio_70.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_7(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_7(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_7(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_7(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

year = portfolio_80.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_80(portfolio_80.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_8(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_8(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_8(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_8(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_90.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_90(portfolio_90.fyear == year_i,:);
```

```
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_9(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_9(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_9(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_9(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i


year = portfolio_100.fyear;
year = unique(year,'rows');
n_loop = size(year,1);
data_2=table();
for i = 1: n_loop
year_i = year(i,:);
FinalTable_k = portfolio_100(portfolio_100.fyear == year_i,:);
average_each_year=table();
average_each_year.excess_ret =mean(FinalTable_k.excess_ret);
data_2=vertcat(data_2,average_each_year);
end
excess_return_sorted_port.Port_High(1) = mean(data_2.excess_ret);
excess_return_sorted_port.Port_High(2) = NWtest(data_2.excess_ret,1,0);
excess_return_sorted_port.Port_High(3) = std(data_2.excess_ret);
excess_return_sorted_port.Port_High(4) = mean(data_2.excess_ret)/std(data_2.excess_ret);

clear data_2 FinalTable_k average_each_year year_i year n_loop i

%%

function y=NWtest(ret,lag,h0)

T=size(ret,1);
vv=var(ret);

for l=1:1:lag

   cc=cov(ret(1:end-l),ret(l+1:end));
   vv=vv+2*(1-l/lag)*cc(1,2);
end

y=(mean(ret)-h0)/sqrt(vv)*sqrt(T);
end
```

III. Regression CAPM, FF-three factor

1. Matlab code to prepare data for regression:

```
%% clean the command window and all variables in the workspace
clc; clearvars; close all;

%% Input data
FF_factor=readtable('F-F_Research_Data_Factors.csv');
portfolio_10=readtable('Contemp_portfolio_10.xlsx');
portfolio_20=readtable('Contemp_portfolio_20.xlsx');
portfolio_30=readtable('Contemp_portfolio_30.xlsx');
portfolio_40=readtable('Contemp_portfolio_40.xlsx');
portfolio_50=readtable('Contemp_portfolio_50.xlsx');
portfolio_60=readtable('Contemp_portfolio_60.xlsx');
portfolio_70=readtable('Contemp_portfolio_70.xlsx');
portfolio_80=readtable('Contemp_portfolio_80.xlsx');
portfolio_90=readtable('Contemp_portfolio_90.xlsx');
portfolio_100=readtable('Contemp_portfolio_100.xlsx');

%% Data

% remove the year before 1987 FF
t_year = 1987;
toDelete = FF_factor.Year <t_year;
FF_factor(toDelete,:) = [];
clear toDelete t_year
FF_factor.Properties.VariableNames{'Year'} = 'fyear';
FF_factor.RF=[];

portfolio_10 = portfolio_10(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio10 = outerjoin(FF_factor,portfolio_10,'Type','left');
data_portfolio10.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio10.fyear_portfolio_10=[];
toDelete = isnan(data_portfolio10.excess_ret);
data_portfolio10(toDelete,:) = [];
clear toDelete;


portfolio_20 = portfolio_20(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
```

```
data_portfolio20 = outerjoin(FF_factor,portfolio_20,'Type','left');
data_portfolio20.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio20.fyear_portfolio_20=[];
toDelete = isnan(data_portfolio20.excess_ret);
data_portfolio20(toDelete,:) = [];
clear toDelete;

portfolio_30 = portfolio_30(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio30 = outerjoin(FF_factor,portfolio_30,'Type','left');
data_portfolio30.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio30.fyear_portfolio_30=[];
toDelete = isnan(data_portfolio30.excess_ret);
data_portfolio30(toDelete,:) = [];
clear toDelete;

portfolio_40 = portfolio_40(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio40 = outerjoin(FF_factor,portfolio_40,'Type','left');
data_portfolio40.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio40.fyear_portfolio_40=[];
toDelete = isnan(data_portfolio40.excess_ret);
data_portfolio40(toDelete,:) = [];
clear toDelete;

portfolio_50 = portfolio_50(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio50 = outerjoin(FF_factor,portfolio_50,'Type','left');
data_portfolio50.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio50.fyear_portfolio_50=[];
toDelete = isnan(data_portfolio50.excess_ret);
data_portfolio50(toDelete,:) = [];
clear toDelete;

portfolio_60 = portfolio_60(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio60 = outerjoin(FF_factor,portfolio_60,'Type','left');
data_portfolio60.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio60.fyear_portfolio_60=[];
toDelete = isnan(data_portfolio60.excess_ret);
data_portfolio60(toDelete,:) = [];
clear toDelete;

portfolio_70 = portfolio_70(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio70 = outerjoin(FF_factor,portfolio_70,'Type','left');
```

```
data_portfolio70.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio70.fyear_portfolio_70=[];
toDelete = isnan(data_portfolio70.excess_ret);
data_portfolio70(toDelete,:) = [];
clear toDelete;

portfolio_80 = portfolio_80(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio80 = outerjoin(FF_factor,portfolio_80,'Type','left');
data_portfolio80.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio80.fyear_portfolio_80=[];
toDelete = isnan(data_portfolio80.excess_ret);
data_portfolio80(toDelete,:) = [];
clear toDelete;

portfolio_90 = portfolio_90(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio90 = outerjoin(FF_factor,portfolio_90,'Type','left');
data_portfolio90.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio90.fyear_portfolio_90=[];
toDelete = isnan(data_portfolio90.excess_ret);
data_portfolio90(toDelete,:) = [];
clear toDelete;

portfolio_100 = portfolio_100(:,{'gvkey' 'fyear' 'excess_ret'});
% Join data
data_portfolio100 = outerjoin(FF_factor,portfolio_100,'Type','left');
data_portfolio100.Properties.VariableNames{'fyear_FF_factor'} = 'fyear';
data_portfolio100.fyear_portfolio_100=[];
toDelete = isnan(data_portfolio100.excess_ret);
data_portfolio100(toDelete,:) = [];
clear toDelete;

%% export
writetable(data_portfolio10,'data_portfolio10.xlsx');
writetable(data_portfolio20,'data_portfolio20.xlsx');
writetable(data_portfolio30,'data_portfolio30.xlsx');
writetable(data_portfolio40,'data_portfolio40.xlsx');
writetable(data_portfolio50,'data_portfolio50.xlsx');
writetable(data_portfolio60,'data_portfolio60.xlsx');
writetable(data_portfolio70,'data_portfolio70.xlsx');
writetable(data_portfolio80,'data_portfolio80.xlsx');
writetable(data_portfolio90,'data_portfolio90.xlsx');
writetable(data_portfolio100,'data_portfolio100.xlsx');
```

2. Stata Code for running Regression in 10 TFP sorted Portfolios:

```
* Portfolio 10
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\data_portfolio10.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)

* Portfolio 20
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\data_portfolio20.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)

* Portfolio 30
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\data_portfolio30.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)


* Portfolio 40
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\data_portfolio40.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
```

```
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)


* Portfolio 50
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\data_portfolio50.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)


* Portfolio 60
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\data_portfolio60.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)


* Portfolio 70
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with
Table\data_portfolio70.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)


* Portfolio 80
```

```
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with Table\data_portfolio80.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)


* Portfolio 90
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with Table\data_portfolio90.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)


* Portfolio 100
import excel "C:\Users\nguye\Google Drive\thesis\Draft Code\Code Part 2 with Table\data_portfolio100.xlsx", sheet("Sheet1") firstrow
duplicates report  gvkey fyear
duplicates list gvkey fyear
duplicates tag  gvkey fyear, gen(isdup)
drop if isdup
xtset gvkey fyear
reg excess_ret Mkt_RF SMB HML gvkey, vce(cluster gvkey)
reg excess_ret Mkt_RF gvkey, vce(cluster gvkey)
```

IV. TFP-Sorted Descriptive data and Industrial Sorted Descriptive data

```
%% clean the command window and all variables in the workspace
```

```
clc; clearvars; close all;

%% Import data

data_US = readtable('final_Minh_TFP_intan_capital.csv');
data_US = table2array(data_US);
data_US(:,3) = exp(data_US(:,3));
DATA = data_US;

a = unique(data_US(:,1));
writematrix(a,'gvkey_same.txt');

tic
dataWRDS_raw = readtable('final_280619_WRDS.xlsx');
toc
dataWRDS_raw(:,2) = [];
dataWRDSraw = table2array(dataWRDS_raw);
for i = 1 : length(dataWRDSraw(1,:))
    K = find(isnan(dataWRDSraw(:,i)));
    dataWRDSraw(K,i) = 0;
end
    K = find(dataWRDSraw(:,1) == 1985);
    dataWRDSraw(K,:) = [];
    K = find(dataWRDSraw(:,1) == 2018);
    dataWRDSraw(K,:) = [];

data_WRDS = zeros(length(DATA(:,1)),22);
data_WRDS(:,1:2) = DATA(:,1:2);

tic
for i = 1 : length(DATA(:,1))
    K_1 = find(dataWRDSraw(:,1) == DATA(i,1));
    K_2 = find(dataWRDSraw(:,2) == DATA(i,2));
    K = intersect(K_1,K_2);
    data_WRDS(i,3:end) = dataWRDSraw(K,3:end);
end
toc

tic
data_sup_raw = readtable('data_Sup.xlsx'); % the deflator-adjusted numbers are here
toc
data_sup_needed = data_sup_raw(:,[1 2 18 19 27 40 41]);
data_sup = table2array(data_sup_needed);

Charac(:,1:2) = data_WRDS(:,1:2);
```

```
K = find(data_WRDS(:,21) == 0); % remove no stock price
Charac(K,:) = [];
DATA(K,:) = [];
data_WRDS(K,:) = [];
K = find(data_WRDS(:,5) == 0);
Charac(K,:) = [];
DATA(K,:) = [];
data_WRDS(K,:) = [];

Charac(:,3) = data_WRDS(:,5) .* data_WRDS(:,21); % market_Cap

Charac(:,4) = log(data_WRDS(:,5) .* data_WRDS(:,21)); % size
Charac(:,5) = (data_WRDS(:,4) - data_WRDS(:,10) + data_WRDS(:,18) -
sum(data_WRDS(:,15:16),2)) ./  (data_WRDS(:,5) .* data_WRDS(:,21)); %BM
K = find(Charac(:,5) <= 0);
Charac(K,:) = [];
DATA(K,:) = [];
data_WRDS(K,:) = [];
K = find(isinf(Charac(:,5)));
Charac(K,:) = [];
DATA(K,:) = [];
data_WRDS(K,:) = [];

Charac(:,5) = log(Charac(:,5)); % log(BM)

Charac(:,6) = data_WRDS(:,19) ./ data_WRDS(:,12); % RD/PPEGT
K = find(isnan(Charac(:,6)));
Charac(K,6) = 0;
K = find(isinf(Charac(:,6)));
Charac(K,6) = 0;

%Charac(:,7) = data_WRDS(:,13) ./ data_WRDS(:,12); % RER
%K = find(isinf(Charac(:,7)));
%Charac(K,7) = 0;

Charac(:,7) = data_WRDS(:,5) .* data_WRDS(:,3); % split adjusted shares


Charac(:,8) = data_WRDS(:,6) ./ ( data_WRDS(:,6) + data_WRDS(:,5) .* data_WRDS(:,21) ); %
leverage
K = find(isnan(Charac(:,8)));
Charac(K,:) = [];

Charac(:,9) =  data_WRDS(:,9) ./ ( data_WRDS(:,4) - data_WRDS(:,10) + data_WRDS(:,18) -
sum(data_WRDS(:,14:16),2) ); % ROE
```

```
Charac(:,10) = ( data_WRDS(:,9) - data_WRDS(:,7) + data_WRDS(:,17) ) ./ data_WRDS(:,4);
% ROA
K = find(isinf(Charac(:,10)));
Charac(K,:) = [];
DATA(K,:) = [];
data_WRDS(K,:) = [];
Charac(:,11) = data_WRDS(:,8) ./ data_WRDS(:,4); % GPR
tic
for i = 1 : length(data_WRDS(:,1))
   K_1 = find(data_sup(:,1) == data_WRDS(i,1));
   K_2 = find(data_sup(:,2) == data_WRDS(i,2));
   K = intersect(K_1,K_2);
   data_WRDS(i,23) = data_sup(K,6); % adjusted Fixed-Cap
   data_WRDS(i,24) = data_sup(K,7); % k
   data_WRDS(i,25) = data_sup(K,5); % adjusted age of capital
   data_WRDS(i,26) = data_sup(K,4) - data_sup(K,3);
end
toc
Charac(:,12) = data_WRDS(:,23) ./ data_WRDS(:,24); % I/K
Charac(:,13) = data_WRDS(:,25); % age of capital
Charac(:,14) = data_WRDS(:,26); % age of year??

data_CRSP_raw = readtable('data_CRSP_RF.xlsx');
data_CRSP_all = table2array(data_CRSP_raw);
data_CRSP(:,1:3) = DATA(:,1:3);
data_CRSP(:,3) = zeros();
for i = 1 : length(DATA(:,1))
   K_1 = find(DATA(i,1) == data_CRSP_all(:,1));
   K_2 = find(DATA(i,2) == data_CRSP_all(:,2));
   K = intersect(K_1,K_2);
   if isempty(K)
      data_CRSP(i,3) = 0;
   end
   if nnz(K) > 1
      data_CRSP(i,3) = data_CRSP_all(K(end),6);
   end
   if nnz(K) == 1
      data_CRSP(i,3) = data_CRSP_all(K,6);
   end
end
K = find(data_CRSP(:,3)==0);
Charac(K,:) = [];
DATA(K,:) = [];
data_WRDS(K,:) = [];
data_CRSP(K,:) = [];
```

```
data_run = DATA% Us

%% Correlation
CorrMatrix = zeros(length(Charac(:,1)),length(Charac(1,:))-2);
CorrMatrix(:,1) = data_run(:,3);
CorrMatrix(:,2:end) = Charac(:,4:end);
Corr = corrcoef(CorrMatrix);
corrcoef(Charac(:,8), data_run(:,3))


%% reform the data (sort TFP by year)
% decide which data you wanna to run with

min_year = min(data_run);
min_year = min_year(2);
max_year = max(data_run);
max_year = max_year(2);
no_Years = max_year - min_year + 1;

n3 = 1;
n4 = 1;
cnt(no_Years,1) = zeros;

data_2(:,1:3) = data_run(:,1:3);
% data_2(size(data_1)) = zeros;

n2 = 1;
a = 1;
for n1 = min_year : max_year
    K = find(data_run(:,2) == n1);
    cnt(n2,1) = nnz(K);
    z = cnt(n2,1) + a - 1;
    data_2(a:z,1) = n1;
    data_2(a:z,2) = data_run(K,1);
    data_2(a:z,3) = data_run(K,3);
    a = a + cnt(n2,1);
    n2 = n2 + 1;
end

check_num_data_2 = sum(cnt); % if checknum = length(data), means sorted all data

for n = 1 : no_Years % column as time (1963 to 2013)
    D_set(n+1,1) = min_year + n - 1;
end
firm_sort = sort(data_run(:,1));
firm_sort = unique(firm_sort,'rows');
```

```
no_Firms = length(firm_sort);
for n = 1 : no_Firms % row as firms
   D_set(1,n+1) = firm_sort(n);
end

D_1 = D_set;
FF_Y = D_set;
Factor_Size = D_set;
Factor_BM = D_set;

% create portfolios for later sorted
for i = 1:10
   all.ptf{1,i} = D_set; % seperate into portfolios with all firms
   nyse20.ptf{1,i} = D_set; % seperate into portfolios with all firms

   all.MarketCap{1,i} = D_set;
   all.size{1,i} = D_set;
   all.BM{1,i} = D_set;
   all.ln_BM{1,i} = D_set;
   all.RD_PPE{1,i} = D_set;
   all.RER{1,i} = D_set;
   all.SplitAdjustedS{1,i} = D_set;
   all.Leverage{1,i} = D_set;
   all.ROE{1,i} = D_set;
   all.ROA{1,i} = D_set;
   all.GPR{1,i} = D_set;
   all.I_K{1,i} = D_set;
   all.Age{1,i} = D_set;
   all.CapitalAge{1,i} = D_set;

   nyse20.MarketCap{1,i} = D_set;
   nyse20.size{1,i} = D_set;
   nyse20.BM{1,i} = D_set;
   nyse20.ln_BM{1,i} = D_set;
   nyse20.RD_PPE{1,i} = D_set;
   nyse20.RER{1,i} = D_set;
   nyse20.SplitAdjustedS{1,i} = D_set;
   nyse20.Leverage{1,i} = D_set;
   nyse20.ROE{1,i} = D_set;
   nyse20.ROA{1,i} = D_set;
   nyse20.GPR{1,i} = D_set;
   nyse20.I_K{1,i} = D_set;
   nyse20.Age{1,i} = D_set;
   nyse20.CapitalAge{1,i} = D_set;
end
```

```
% create D_1 as Panel Data 1
tic
for i = 1 : length(data_run(:,1))
   K_1 = find(D_set(:,1) == data_run(i,2)); % time as index of row
   K_2 = find(D_set(1,:) == data_run(i,1));
   D_1(K_1,K_2) = data_run(i,3);
   FF_Y(K_1,K_2) = data_CRSP(i,3);
   for j = 3 : length(Charac(1,:))
      Factor{1,j-2}(K_1,K_2) = Charac(i,j);
   end
end
toc

check_num_D1 = nnz(D_1(2:end,2:end));

%% get all firms included percentiles
for i = 1 : no_Years
   K = find(Charac(:,2) == D_1(i+1,1));
   percentile_all(i,:) = prctile(data_run(K,3),[0 10 20 30 40 50 60 70 80 90 100]);
end
percentile_all(:,end) = percentile_all(:,end) + 0.0000001;

%% get NYSE list, Market Capitalization and breaking point_20% and percentile

K = find(data_WRDS(:,20) == 11); % NYSE Exchange Codes
nyse_org = Charac(K,1:3); % MarketCap
nyse_org(:,4) = data_run(K,3); % tfp

for i = 1 : no_Years
   K_1 = find(nyse_org(:,2) == D_1(i+1,1));
   bp20(i) = prctile(nyse_org(K_1,3),20);
   K_2 = find(nyse_org(K_1,3) > bp20(i)); % exclude the nyse firms that smaller than bp20
   numcheck(i) = nnz(K_2);
   percentile_20bp(i,:) = prctile(nyse_org(K_1(K_2),4),[0 10 20 30 40 50 60 70 80 90 100]);
end
percentile_20bp(:,end) = percentile_20bp(:,end) + 0.0000001;
%percentile_20bp(:,end) = percentile_all(:,end);
sum(numcheck);


%% seperate all firms into portfolios with NYSE_20% percentiles

for n = 1 : no_Years
   tic
   K = find(Charac(:,2) ==  D_1(n+1,1));
```

```
    for j = 1 : length(K)
        L = find(D_1(1,:) == Charac(K(j),1));
        for i = 1 : 10 % represent each portfolio
            if  ((percentile_20bp(n,i) <= data_run(K(j),3)) && (data_run(K(j),3) <
percentile_20bp(n,i+1)))
                nyse20.ptf{1,i}(n+1,L) = data_run(K(j),3);
                %nyse20.MarketCap{1,i}(n+1,L) = Charac(K(j),3);
                nyse20.size{1,i}(n+1,L) = Charac(K(j),4);
                nyse20.ln_BM{1,i}(n+1,L) = Charac(K(j),5);
                nyse20.RD_PPE{1,i}(n+1,L) = Charac(K(j),6);
                nyse20.SplitAdjustedS{1,i}(n+1,L) = Charac(K(j),8);
                nyse20.Leverage{1,i}(n+1,L) = Charac(K(j),9);
                nyse20.ROE{1,i}(n+1,L) = Charac(K(j),10);
                nyse20.ROA{1,i}(n+1,L) = Charac(K(j),11);
                nyse20.GPR{1,i}(n+1,L) = Charac(K(j),12);
                nyse20.I_K{1,i}(n+1,L) = Charac(K(j),13);
                nyse20.Age{1,i}(n+1,L) = Charac(K(j),14);
                nyse20.CapitalAge{1,i}(n+1,L) = Charac(K(j),13);
                nyse20.ExcessReturn{1,i}(n+1,L) = data_CRSP(K(j),3);
            end
        end
    end
    toc
end

% how
for i = 1 : 10
    num_nyse20(i) = nnz(nyse20.ptf{1,i}(2:end,2:end));
    num_Pft_bp20(i) = round( num_nyse20(i) / 31); % count the amount of the data in each
portfolio
end
%% create table1

for i = 1 : 10
    tbl_nyse20(1,i) = sum(sum(nyse20.ptf{1,i}(2:end,2:end),2));
    %tbl_nyse20(2,i) = sum(sum(nyse20.MarketCap{1,i}(2:end,2:end),2));
    tbl_nyse20(2,i) = sum(sum(nyse20.size{1,i}(2:end,2:end),2));
    tbl_nyse20(3,i) = sum(sum(nyse20.ln_BM{1,i}(2:end,2:end),2));
    tbl_nyse20(4,i) = sum(sum(nyse20.RD_PPE{1,i}(2:end,2:end),2));

    tbl_nyse20(5,i) = sum(sum(nyse20.SplitAdjustedS{1,i}(2:end,2:end),2));
    tbl_nyse20(6,i) = sum(sum(nyse20.Leverage{1,i}(2:end,2:end),2));
    tbl_nyse20(7,i) = sum(sum(nyse20.ROE{1,i}(2:end,2:end),2));
    tbl_nyse20(8,i) = sum(sum(nyse20.ROA{1,i}(2:end,2:end),2));
    tbl_nyse20(9,i) = sum(sum(nyse20.GPR{1,i}(2:end,2:end),2));
    tbl_nyse20(10,i) = sum(sum(nyse20.I_K{1,i}(2:end,2:end),2));
```

```
    tbl_nyse20(11,i) = sum(sum(nyse20.Age{1,i}(2:end,2:end),2));
    tbl_nyse20(12,i) = sum(sum(nyse20.CapitalAge{1,i}(2:end,2:end),2));
    tbl_nyse20(:,i) = tbl_nyse20(:,i) ./ num_nyse20(i);
    K = find(isnan(tbl_nyse20(:,i)));
    tbl_nyse20(K,i) = 0;
end



%% Seperated by Industry

% 600 transportation    935 Energy  940 Techonoly
% 700 Utilities         970 Basic Materials
% 800 Financial         974 Communication
% 905 Health Care       976 Consumer Cyclics
% 925 Capital Goods     978 Consumer Staples

tfp_firm(1,:) = D_1(1,2:end);
for i = 1 : no_Firms
    tfp_firm_num(1,i) = nnz(D_1(2:end,i+1));
end
tfp_firm(2,:) = sum(D_1(2:end,2:end))./tfp_firm_num;
tfp_firm(3,:) = sum(Factor{1,2}(2:end,2:end))./tfp_firm_num;
tfp_firm(4,:) = sum(Factor{1,3}(2:end,2:end))./tfp_firm_num;
tfp_firm(5,:) = sum(Factor{1,8}(2:end,2:end))./tfp_firm_num;
for i = 1 : no_Firms
    K = find(tfp_firm(1,i) == data_WRDS(:,1));
    if nnz(K) > 1
        K = K(1);
    else
        K = K;
    end
    tfp_firm(6,i) = data_WRDS(K(1),22);
end

list_Eco = unique(data_WRDS(:,22));
toDelete = find(isnan(list_Eco));
list_Eco(toDelete) = [];
toDelete = find(list_Eco == 0);
list_Eco(toDelete) = [];
list_Eco(:,2:6) = zeros();

for i = 1 : length(tfp_firm(1,:))
    K = find(tfp_firm(6,i) == list_Eco(:,1));
    list_Eco(K,2) = list_Eco(K,2) + 1;
    L = find(tfp_firm(1,i) == tfp_firm(1,:));
```

```
    list_Eco(K,3) = list_Eco(K,3) + tfp_firm(2,L);
    list_Eco(K,4) = list_Eco(K,4) + tfp_firm(3,L);
    list_Eco(K,5) = list_Eco(K,5) + tfp_firm(4,L);
    list_Eco(K,6) = list_Eco(K,6) + tfp_firm(5,L);

end
for i =  1 : 11
    list_Eco(i,3) =  list_Eco(i,3)/list_Eco(i,2);
    list_Eco(i,4) =  list_Eco(i,4)/list_Eco(i,2);
    list_Eco(i,5) =  list_Eco(i,5)/list_Eco(i,2);
    list_Eco(i,6) =  list_Eco(i,6)/list_Eco(i,2);
end
```