CPEE Final Project- Advanced target

YIN Youxin, WANG Hao, WANG Yujing

本题旨在探究中国股票市场是否存在动量或者反转现象

(a) First, read the datasets return_monthly.xlsx and me_lag.xlsx into MATLAB and reshape them into the long-table formats as you have seen in class. Then merge these two datesets and drop observations with missing lagged market capitalization. The output of your code should be similar to return_m.mat . You may or may not build on the code provided in momentum_class.m.

导入 return_monthly.xlsx 和 me_lag.xlsx 数据，连接之后，使用 stack 函数将宽列表转化为长列表，并且去除带有缺失值的行。

```
% (a)
%%
%读取数据
return_m_hor = readtable('return_monthly.xlsx','ReadVariableNames',true,'PreserveVariableNames',true,'Format','auto');
me_lag = readtable('me_lag.xlsx','ReadVariableNames',true,'PreserveVariableNames',true,'Format','auto');
%%
%转化为long数据
long_return_m = stack(return_m_hor,3:127,"NewDataVariableName",'return','IndexVariableName','date');
long_me_lag = stack(me_lag,3:127,"NewDataVariableName",'me_lag','IndexVariableName','date');
%%
%两个表连接一下成为mydata
mydata = innerjoin(long_return_m,long_me_lag,"Keys",{'code','date','name'});
%%
%去除mydata中的缺失值成为mydata1
mydata1 = mydata(~isnan(mydata.return)&~isnan(mydata.me_lag)&(mydata.me_lag>0),:);
%上面就已经完成了（1），融合后的表是mydata1
```

(b) Every K months, sort stocks into five groups based on previous K months' return and hold this position for K months. What is the average equal-weighted return spread between high and low previous stock returns portfolios for K = 1; 3; 6; 12; 24. Do you find that momentum exists in Chinese stock markets?

为了获得在 K 的不同取值下，拥有最高历史收益和最低历史收益的组的未来收益的 spread，我们首先需要获得每个月每个公司的前 K 个月历史平均收益，和之后的 K 个月的未来平均收益。

首先，我们做一些简单的数据处理，获得我们需要的数据：

```
%下面对mydata1按照时间分组，获得yymm列
[G,date] = findgroups(mydata1.date);% 按照时间进行分组
mydata1_eachday_equal_weighted_return = splitapply(@(x)mean(x),mydata1.return,G);

return_eachday = table(date,mydata1_eachday_equal_weighted_return);

mydata1.date = char(mydata1.date);
mydata1.year = year(mydata1.date);
mydata1.month = month(mydata1.date);
mydata1.yymm = mydata1.year*12+mydata1.month;
%%
%截取mydata1中有用的列，成为mydata4
mydata4 = table(mydata1.code,mydata1.name,mydata1.date,mydata1.return,mydata1.yymm,mydata1.me_lag,'VariableNames',{'code','name','date','return','yymm','
```

之后，我们遍历 K 的不同取值，获得不同 K 取值下的历史平均收益和未来平均收益。

```
%%
%将return统一平移（考虑到了公司之间的接合处），并求一下前K月的收益，与包括这个月和之后K-1个月的收益
for K = [1, 3, 6, 12, 24]
    mydata1_try = mydata1;
    mydata2 = mydata1;
    mydata2.yymm = mydata1.yymm+K;
    newReturnName = strcat('return_lag', num2str(K));
    mydata2 = renamevars(mydata2, 'return', newReturnName);
    mydata3 = outerjoin(mydata1_try, mydata2, 'Keys', {'code','yymm'},'MergeKeys',true,'Type','left');
    mydata4.(newReturnName) = mydata3.(newReturnName);
end

for K = [1, 3, 6, 12, 24]
    newReturnName = strcat('return_lag', num2str(K));
    newReturnName_mean = strcat('return_lag_mean', num2str(K));
    % 计算滑动平均并将结果赋值给新列
    mydata4.(newReturnName_mean) = movmean(mydata4.(newReturnName), [0 K-1]);
    mydata4 = removevars(mydata4, newReturnName);
    newReturnName_lead = strcat('return_lead_mean', num2str(K));
    mydata4.(newReturnName_lead) = movmean(mydata4.return, [0 K-1]);
end
%这里就获得了新的mydata4，包含基础信息，以及K不同取值下，lag和lead的收益平均值
```

之后，我们对处理好的数据进行分组，计算每组的平均收益：

```
%下面按照lag的收益做分组，计算lead的收益
group_return_diffK = zeros(5,5);
K_list = [1, 3, 6, 12, 24];
for K_index = 1:5
    K = K_list(K_index);
    selectedColumns = [1:6, 6 + 2*K_index-1, 6 + 2*K_index];
    mydata_K_try = mydata4(:, selectedColumns);
    newReturnName_mean = strcat('return_lag_mean', num2str(K));
    mydata_K_try = mydata_K_try(~isnan(mydata_K_try.(newReturnName_mean)), :);

    [G,yymm] = findgroups(mydata_K_try.yymm);
    %这里，表示根据jdate分组，G表示每一行所属组的序号，jdate则会是每一组的代表性元素
    mydata_K_try_breaks = table(yymm);

    prctile_20 = @(input)prctile(input,20);
    prctile_40 = @(input)prctile(input,40);
    prctile_60 = @(input)prctile(input,60);
    prctile_80 = @(input)prctile(input,80);

    mydata_K_try_breaks.rt20 = splitapply(prctile_20,mydata_K_try.(newReturnName_mean),G);
    mydata_K_try_breaks.rt40 = splitapply(prctile_40,mydata_K_try.(newReturnName_mean),G);
    mydata_K_try_breaks.rt60 = splitapply(prctile_60,mydata_K_try.(newReturnName_mean),G);
    mydata_K_try_breaks.rt80 = splitapply(prctile_80,mydata_K_try.(newReturnName_mean),G);

    mydata_K_try = outerjoin(mydata_K_try,mydata_K_try_breaks,'Keys',{'yymm'},'MergeKeys',true,'Type','left');

    rtport = rowfun(@rt_bucket,mydata_K_try(:,{newReturnName_mean,'rt20','rt40','rt60','rt80'}),'OutputVariableNames','ce
    mydata_K_try.rtport = table2array(rtport);

    if K==3
        save return_K3.mat mydata_K_try;
    end

    [G,~] = findgroups(mydata_K_try.rtport);
    newReturnName_lead = strcat('return_lead_mean', num2str(K));
```

```
    rt_group_mean = splitapply(@mean,mydata_K_try.(newReturnName_lead),G);
    group_return_diffK(:,K_index) = rt_group_mean;
end
%这里就获得了最后的group_return_diffK，是不同K取值，不同组的平均return
```

得到如下结果：不同分组在不同 K 取值下的平均收益是：

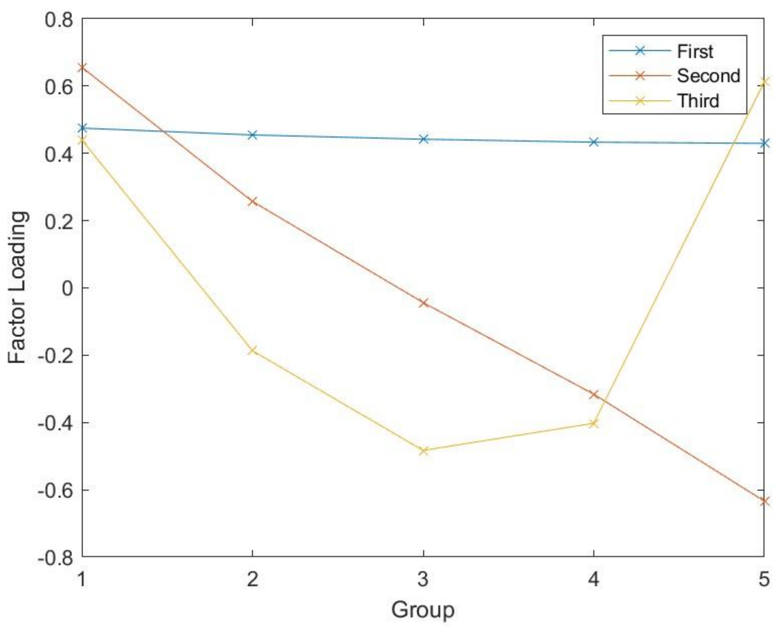
K:	1	3	6	12	24
组1	1.40	1.25	0.84	1.10	1.52
组2	1.34	1.10	1.01	1.15	1.52
组3	1.05	1.03	1.04	1.05	1.36
组4	0.58	0.80	0.95	0.88	1.17
组5	-0.14	0.37	0.63	0.68	0.78

不同 K 取值下的 spread(Group5-Group1)取值是：

K:	1	3	6	12	24
	-1.54	-0.87	-0.21	-0.41	-0.74

尽管我们未做假设检验，但是从计算的结果中，我们可以明显的看出，随着历史收益的增加，未来的收益明显有逐渐变小的趋势，这表明至少从分析用到的数据集整体来看，中国股票市场不存在动量现象，与之相反，反转现象倒是比较明显。

(c) Use the principal component analysis for the equal-weighted return of five portfolios created by sorting on previous K = 3 months' return. How do you interpret the first and second factor from the PCA in economic terms? You may explore whether these factors are useful to explain portfolios sorted by previous stock returns. Also compare PCA factors to the factor MOM defined as the difference in equal-weighted return between the high previous stock returns portfolio and low previous stock returns portfolio.

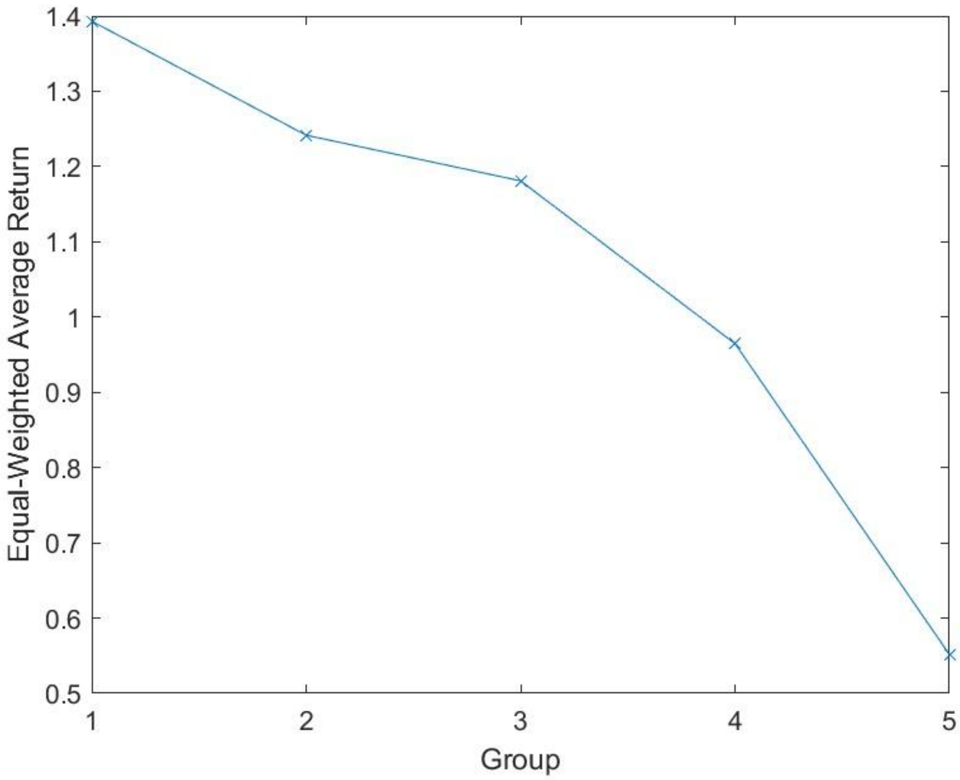


第一主成分代表 level，在各组基本水平不变，对于解释收益差异没有帮助；

第二主成分代表 slope，它的 factor loading 在历史低收益组为正，历史高收益组为负，呈单调下降趋势，这意味着第二主成分的增加会使得历史低收益组在未来的收益增加，历史高收益组在未来的收益减少，一定程度上解释了收益的差异。

以如下方式构建 MOM 因子：MOM=Return(Group5)-Return(Group1)

MOM 因子和 PCA 五个主成分的相关性依次为-0.227, -0.967, 0.061, 0.103, 0.000，可见 MOM 因子与第二主成分高度负相关，说明 MOM 因子具有斜率结构，以及 MOM 因子的减小会使得历史高收益组在未来的收益减少。



从 K=3 的平均收益图像可以看出，第二主成分因子决定整体斜率趋势，历史高收益组在未来收益较低，至于 Group2~3 未来收益下降放缓可能来源于第三主成分因子的影响。总体来看，中国股票市场存在反转效应，动量效应不明显。