



清华大学
Tsinghua University

金融大数据-大作业： CAPM模型的实证检验

金融大数据与量化分析

Financial Big Data and Quantitative Analytics



实验要求

- 每位同学都应**完成**作业
- 课程给出了详尽的辅助参考。
 - 参考例子均采用：招商银行（600036.SH）和建设机械（600984.SH）。
 - 供同学们参考关键的python程序，提示：需要根据实际情况酌情修改。



步骤一、数据获取

请统一选取沪市A股：

- (1) 大盘股：600000.SH 浦发银行；
- (2) 小盘股：600352.SH 浙江龙盛；
- (3) 股价指数：000001.SH 上证综指；
- (4) 无风险证券：CBA00621.CS 中债国债1-3年指数，债券指数等

时间范围：2016年5月-2020年5月，月频数据

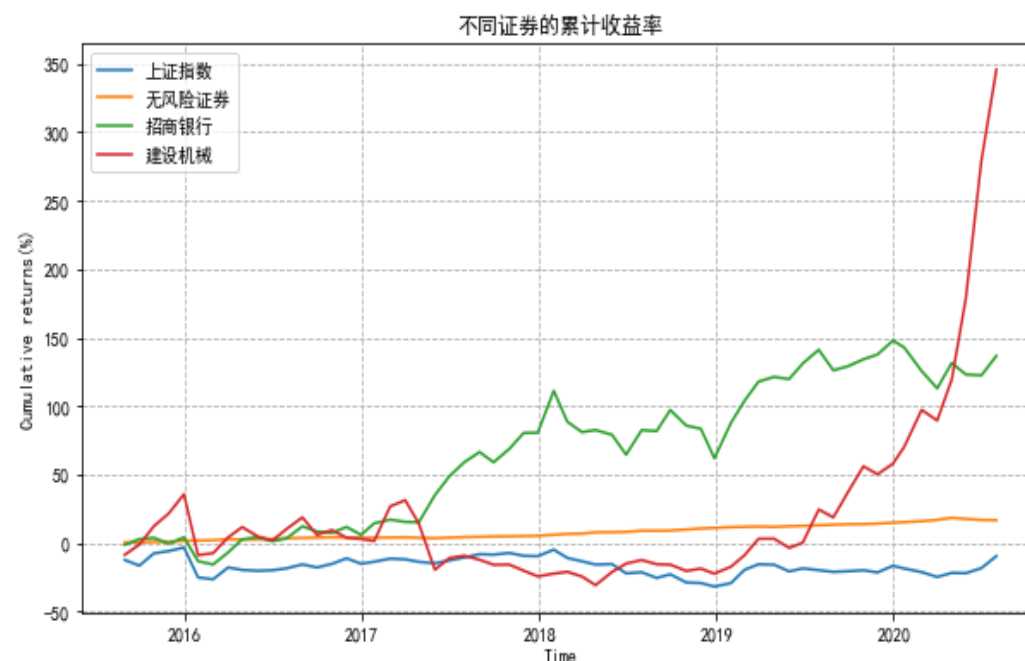
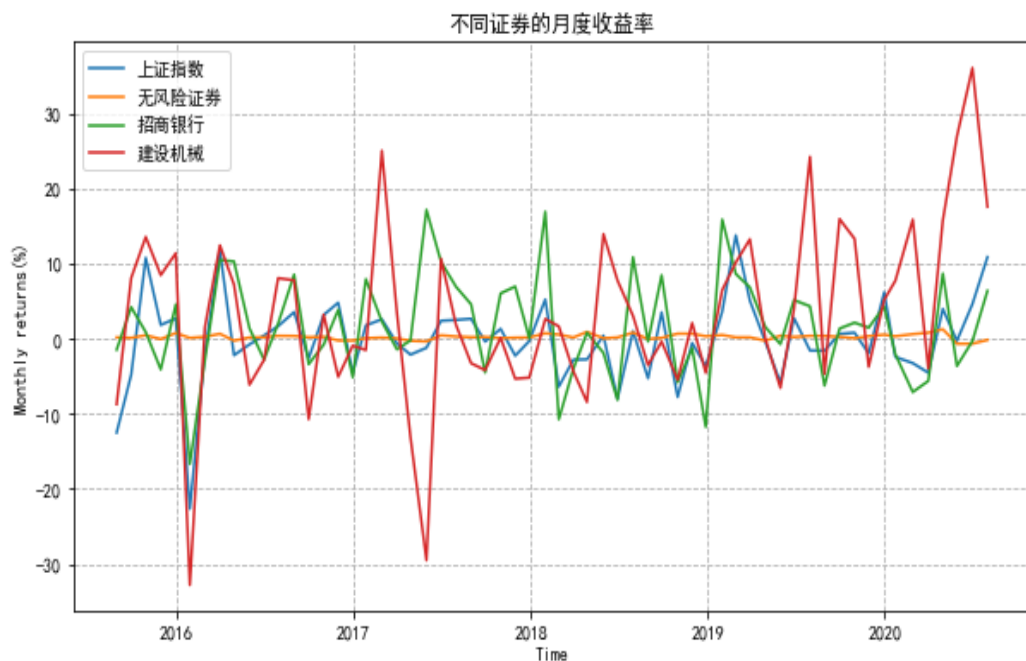
数据来源：Wind、中心GP数据库、聚宽等均可以。

编程方式：python，涉及numpy, pandas, matplotlib, statmodels等库。



步骤二、画出收益率图

- 在图上分别画出两只股票，股指，及无风险证券月收益和累计收益的时间序列。
- (例图如下)





收益率画图的参考程序（部分）

■ 月度收益率 $r_{t+1} = \frac{p_{t+1}}{p_t} - 1$ 。

■ 累计收益率 $R_{t+1} = (1 + r_1) \times (1 + r_2) \cdots (1 + r_{t+1}) - 1$

#导入相关的库:

```
import os
import time,datetime
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm
```

#计算累计收益率:

```
def log_ret(x):
    test=x.map(lambda x: np.log(1+x/100))
    return (np.exp(test.sum())-1)
```

#计算累计收益率

```
def get_cum_ret(ret):
    cum_ret = ret.copy()
    for i in range(len(ret)):
        for j in range(ret.shape[1]-1):
            cum_ret.iloc[i,j+1] = log_ret(ret.iloc[:i+1,j+1])*100
    return cum_ret
```



步骤三、计算两只票的相关性

- 计算两只股票的月收益率的均值(mean)、方差(variance)、标准差(standard variance), 以及两只股票月收益率的协方差(Covariance)和相关系数(Correlation coefficient)。
- (参考示例列表如下)

股票	均值	方差	标准差	协方差	相关系数
招商银行	0.0167	0.0046	0.0681	0.0014	0.1765
建设机械	0.0321	0.0139	0.1181		

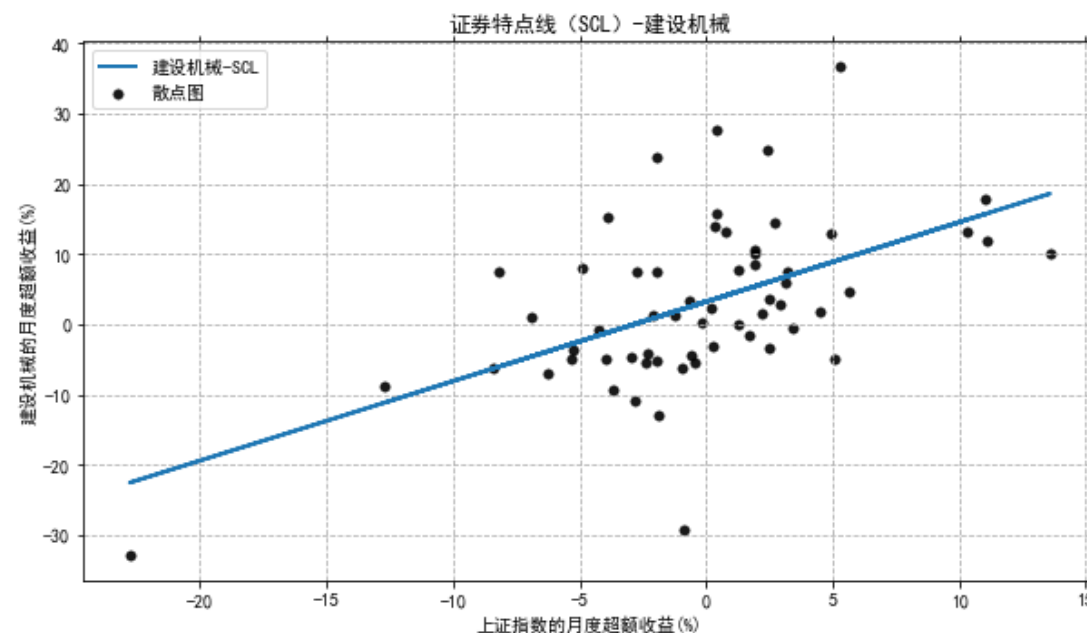
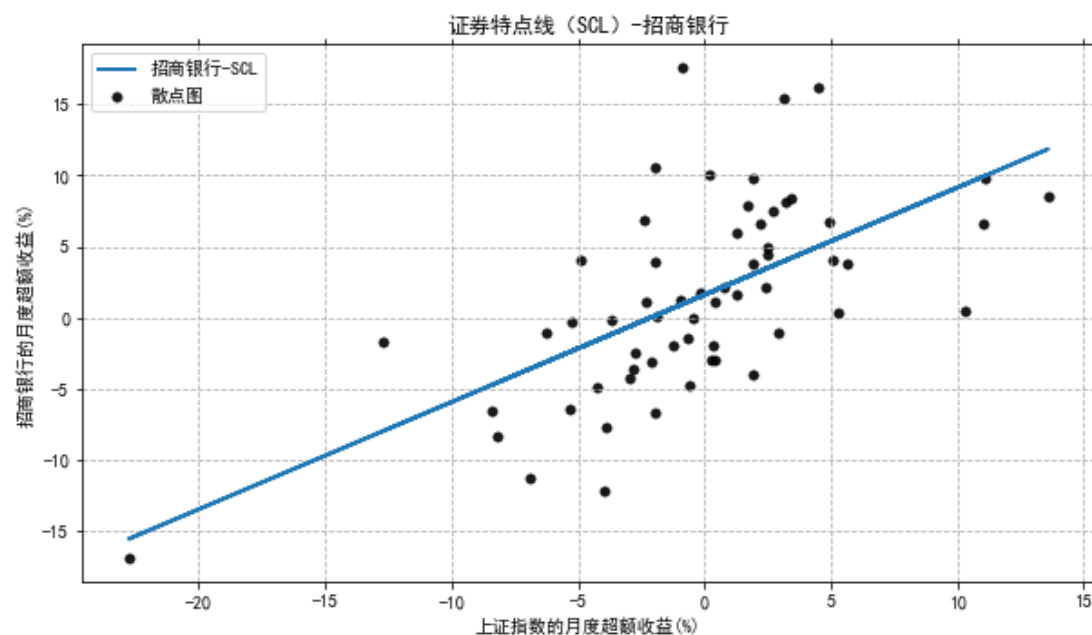


相关性计算的参考程序（部分）

```
# 计算相关性
def get_cov(ret):
    data = ret.iloc[:,[3,4]]/100
    output = pd.DataFrame()
    output['均值'] = data.mean()
    output['方差'] = data.var()
    output['标准差'] = data.std()
    covxy = np.cov(data.iloc[:,0].values, data.iloc[:,1].values)
    coefxy = np.corrcoef(data.iloc[:,0].values,data.iloc[:,1].values)
    output['协方差'] = [np.NaN,covxy[0,1]]
    output['相关系数'] = [np.NaN,coefxy[0,1]]
    return output
```

步骤四、CAPM模型估计和画图

- 用回归的方法分别估计你所选择的股票的单因子CAPM模型，在散点图上画出回归线。
- (例图如下)





CAPM模型估计的参考程序（部分）

```
x = ret.iloc[:,[1,2]]
x.columns = ['常数项','市场超额收益']
x['常数项'] = 1
x['市场超额收益'] = ret.iloc[:,1]-ret.iloc[:,2] #股票收益率减去无风险利率
y1 = ret.iloc[:,3]-ret.iloc[:,2]
plt.rcParams['font.sans-serif'] = ['SimHei'] #下面这两行为了中文显示
plt.rcParams['axes.unicode_minus'] = False
plt.figure(figsize=(10,12)) #图的大小设置
plt.scatter(x.iloc[:,1],y1,s = 30, c = 'black', marker = 'o', alpha = 0.9, linewidths = 0.3, label = '散点图')
reg1 = LinearRegression().fit(x, y1) #线性拟合
pred1 = reg1.predict(x)
plt.plot(x.iloc[:,1], pred1, linewidth = 2, label = '招商银行-SCL')
plt.grid(ls='--')
plt.xlabel('上证指数的月度超额收益(%))')
plt.ylabel('招商银行的月度超额收益(%))')
```

步骤五、分解 α 和 β

- 分别写出两只股票的单因素回归方程；
 - 指出两只股票的 α 和 β 分别是什么，并且讨论 α 和 β 的含义和其统计显著性（t-value）；
 - 对比两只股票的 β ，你得到了什么结论？
- 本部分可参考以下第七次课内容。（示例例子如下）

招商银行： $R_t - Rf_t = 1.61\% + 0.7522 \times (Rm_t - Rf_t)$

建设机械： $R_t - Rf_t = 3.26\% + 1.1315 \times (Rm_t - Rf_t)$

股票	α	t-value	含义	β	t-value	含义
招商银行	1.61%	2.324	超额收益	0.7522	6.032	与市场相关性
建设机械	3.26%	2.506	超额收益	1.1315	4.849	与市场相关性



α 和 β 分解的参考程序 (部分)

#招商银行的拟合模型

```
def get_OLS1(ret):  
    x = (ret.iloc[:,1]-ret.iloc[:,2])/100  
    y1 = (ret.iloc[:,3]-ret.iloc[:,2])/100  
    X=sm.add_constant(x)  
    model1 = sm.OLS(y1,X)  
    model1 = model1.fit()  
    return model1
```

#建设机械的拟合模型

```
def get_OLS2(ret):  
    x = (ret.iloc[:,1]-ret.iloc[:,2])/100  
    y2 = (ret.iloc[:,4]-ret.iloc[:,2])/100  
    X=sm.add_constant(x)  
    model2 = sm.OLS(y2,X)  
    model2 = model2.fit()  
    return model2
```

```
model1 = get_OLS1(ret)  
model1.summary()
```

步骤六、风险分析

■ 分析由回归产生的误差变量：利用回归的预测值及误差变量数据，分别计算两只股票的系统性风险、公司风险和总风险。

■ 本部分可参考第七次课内容。（例子如下）

总风险 = 股票月度收益率的标准差

系统性风险 = 回归预测值的标准差

公司风险 = 残差部分的标准差

关系：总风险² = 系统性风险² + 公司风险²

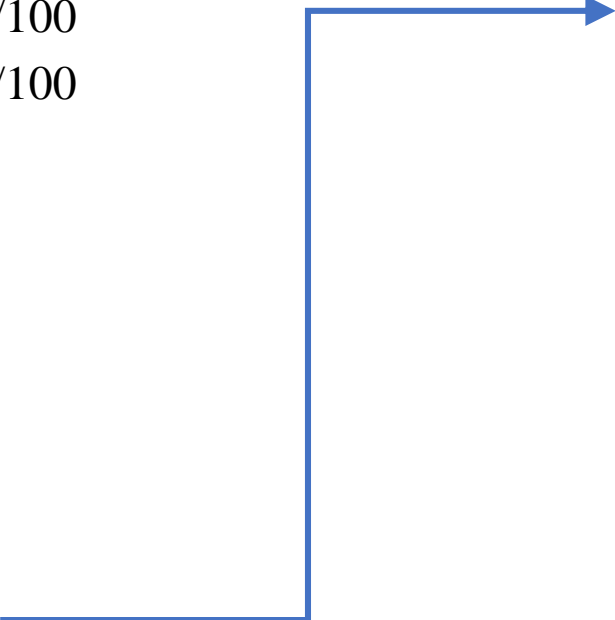
股票	总风险	系统性风险	公司风险
招商银行	6.80%	4.22%	5.33%
建设机械	11.82%	6.35%	9.97%



风险分析的参考程序（部分）

```
def get_risk(ret):
```

```
    x = (ret.iloc[:,1]-ret.iloc[:,2])/100  
    y1 = (ret.iloc[:,3]-ret.iloc[:,2])/100  
    y2 = (ret.iloc[:,4]-ret.iloc[:,2])/100  
    X=sm.add_constant(x)  
    model1 = sm.OLS(y1,X)  
    model1 = model1.fit()  
    y_hat1=model1.predict(X)  
    y_error1 = y1 - y_hat1  
    model2 = sm.OLS(y2,X)  
    model2 = model2.fit()  
    y_hat2=model2.predict(X)  
    y_error2 = y2 - y_hat2
```



```
    risk = np.zeros((2,3))  
    risk[0,0] = y1.std()  
    risk[0,1] = y_hat1.std()  
    risk[0,2] = y_error1.std()  
    risk[1,0] = y2.std()  
    risk[1,1] = y_hat2.std()  
    risk[1,2] = y_error2.std()  
    return risk
```



提示

- (1) 数据获取部分，可以用聚宽平台或中心gp数据库的内容。
- (2) 画图部分，推荐使用matplotlib或者seaborn。
- (3) 均值、标准差与协方差和相关系数，可以用pandas或numpy库来完成
- (4) 单因子回归用后面的OLS模型，可以用statmodels包。

作业提交与验收

- 作业截止日期：校历第12周，12月4日 中午12:00
- 作业提交方式：请将获取数据、notebook计算程序、实验结果MD文件等打包后提交至网络学堂，用XXX表示学号：
 - ✓ XXX_capm-{1,2,3,4}.csv （4个文件，或 *.xlsx格式）
 - ✓ XXX_capm.ipynb
 - ✓ XXX_capm.md
- 作业验收方式：在线会议或当堂分组验收
- 作业验收时间：校历第12周，12月4日下午4：05-5：00



清华大学
Tsinghua University

THANKS