

# 金融大数据与量化分析 CAPM检验

崔轩宁 1800013083

## 步骤一：数据获取

代码见data\_retrieve.ipynb

通过聚宽平台jqdatasdk包进行数据获取，所有数据均来自于聚宽平台。

浦发银行：600000.XSHG，选取的大盘股

浙江龙盛：600352.XSHG，选取的小盘股

上证综指：000001.XSHG，用此代表基准收益

国债指数：000012.XSHG，用此代表无风险收益

将所有文件数据以csv格式保存在data目录下，包括如下信息：

```
date, open, close, high, low, volume, money
```

以浦发银行为例进行展示：

	Unnamed: 0	open	close	high	low	volume	money
0	2016-05-03	10.43	10.53	10.60	10.36	47791645.0	501830132.0
1	2016-05-04	10.50	10.51	10.60	10.47	39777662.0	418592994.0
2	2016-05-05	10.50	10.50	10.53	10.45	25699440.0	269576075.0
3	2016-05-06	10.50	10.30	10.52	10.29	40272847.0	418454827.0
4	2016-05-09	10.29	10.18	10.35	10.11	37762247.0	386423417.0

## 步骤二：画出收益率图

代码见yield\_draw.ipynb

### 数据处理

所下载数据以日为单位，但最终统计以月为单位，故先计算每月收益率。具体来说，我们将2016年5月1日的开盘价近似为2016年4月的收盘价，此后根据每月最后一天的收盘价计算当月收益率，再进一步计算累计收益率

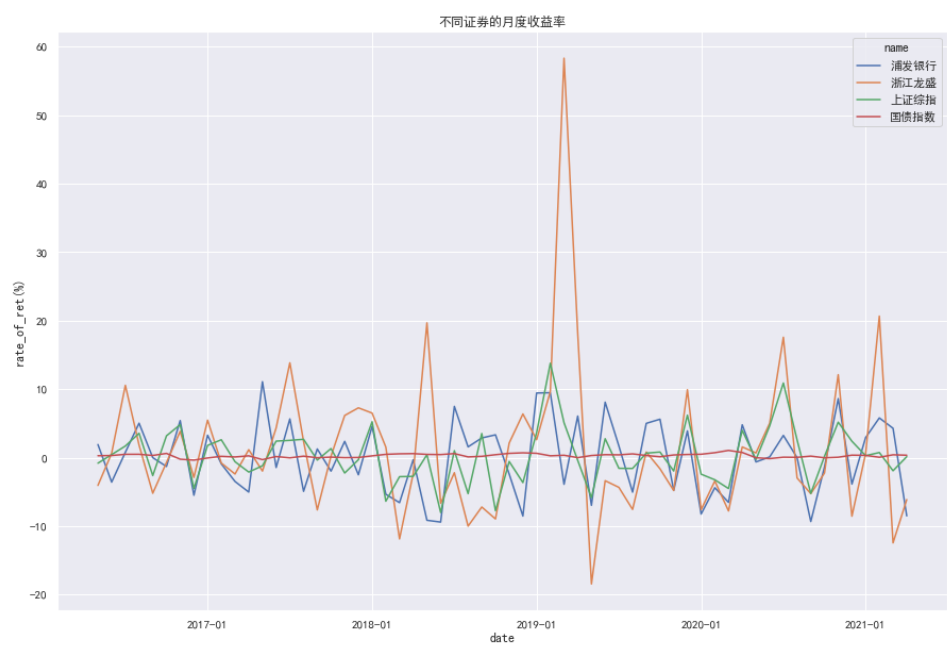
### 画图

用seaborn进行作图，需要先设置如下指令以使得其可以显示中文及负号，当然前提是对应库中有SimHei字体

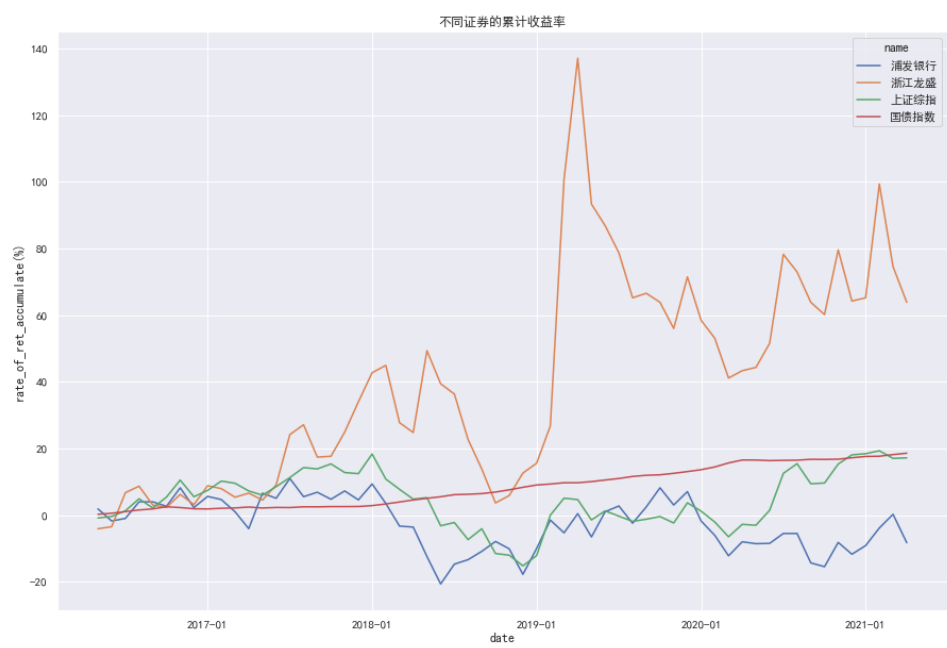
```
sns.set(font="simhei")
plt.rcParams["axes.unicode_minus"] = False
```

利用所求数据进行作图，结果如下：

不同证券的月度收益率



不同证券的累计收益率



步骤三：评价指标计算

代码见compute\_metrics.ipynb

数据处理

以第一日收盘价为基准计算新的净值序列

指标计算

持有收益

只需根据最后一日的收盘价和起初的收盘价即可得

年化收益率

假定一年天数为days\_per\_year，其默认值为245，根据简单公式容易计算出

最大回撤

两层循环找到最大回撤，当然也可以预处理后单层循环搞定

夏普比率

策略年化收益已经得出，无风险利率用国债指数来代替，策略年化收益波动率用日化收益波动率乘上每年天数的开方得到

根据公式即可计算

步骤四：相关性计算

代码见compute\_parameters.ipynb

所需所有指标均可以通过库函数计算得出，结果列举如下

股票	均值	方差	标准差	协方差	相关系数
浦发银行	-9.4034	0.0029	0.0539	0.0011	0.1782
浙江龙盛	0.01340	0.0121	0.1098		

步骤五：CAPM模型

代码见CAPM.ipynb

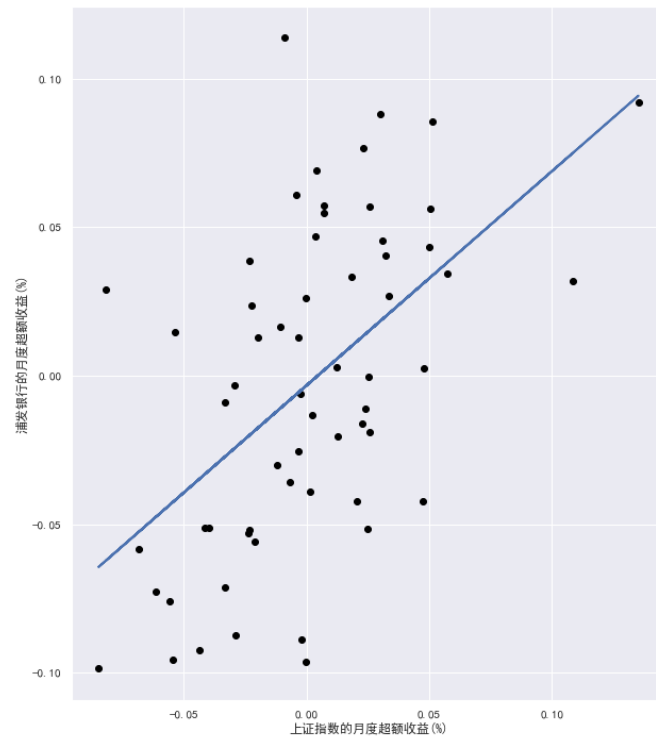
数据处理

以国债指数作为无风险收益，计算其他三只股的超额收益

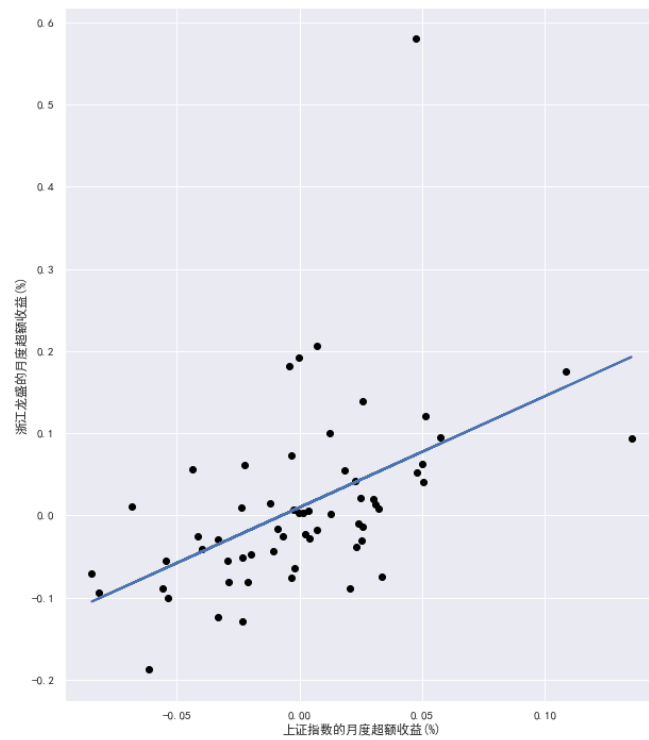
画图

拟合采用sklearn的LinearRegreesion方法，画图采用matplotlib的pyplot方法，结果如下

浦发银行的CAPM估计



## 浙江龙盛的CAPM估计



## 步骤六：分解alpha和beta

代码见model.ipynb

采用statsmodels的OLS方法进行模型拟合，打印模型信息如下

浦发银行的拟合模型OLS

model: 浦发银行									
OLS Regression Results									
=====									
Dep. Variable:		y	R-squared:		0.291				
Model:		OLS	Adj. R-squared:		0.278				
Method:		Least Squares	F-statistic:		23.76				
Date:		Sat, 13 Nov 2021	Prob (F-statistic):		8.85e-06				
Time:		12:44:43	Log-Likelihood:		376.28				
No. Observations:		60	AIC:		-748.6				
Df Residuals:		58	BIC:		-744.4				
Df Model:		1							
Covariance Type:		nonrobust							
=====									
		coef	std err	t	P> t	[0.025	0.975]		
-----									
const		-3.277e-05	6e-05	-0.546	0.587	-0.000	8.74e-05		
x1		0.7215	0.148	4.874	0.000	0.425	1.018		
=====									
Omnibus:		1.253	Durbin-Watson:		2.368				
Prob(Omnibus):		0.534	Jarque-Bera (JB):		1.299				
Skew:		0.300	Prob(JB):		0.522				
Kurtosis:		2.600	Cond. No.		2.47e+03				
=====									

浙江龙盛的拟合模型OLS

model: 浙江龙盛

OLS Regression Results

=====

Dep. Variable:

y

R-squared:

0.252

Model:

OLS

Adj. R-squared:

0.239

Method:

Least Squares

F-statistic:

19.54

Date:

Sat, 13 Nov 2021

Prob (F-statistic):

4.38e-05

Time:

12:44:43

Log-Likelihood:

332.66

No. Observations:

60

AIC:

-661.3

Df Residuals:

58

BIC:

-657.1

Df Model:

1

Covariance Type:

nonrobust

=====

coef

std err

t

P>|t|

[0.025

0.975]

-----

const

9.751e-05

0.000

0.785

0.436

-0.000

0.000

x1

1.3539

0.306

4.421

0.000

0.741

1.967

=====

Omnibus:

60.621

Durbin-Watson:

1.911

Prob(Omnibus):

0.000

Jarque-Bera (JB):

415.188

Skew:

2.760

Prob(JB):

6.97e-91

Kurtosis:

14.645

Cond. No.

2.47e+03

=====

表格总结如下

股票	alpha	t-value	含义	beta	t-value	含义
浦发银行	-3.277e-05	-0.546	超额收益	0.7215	4.874	与市场相关性
浙江龙盛	9.761e-05	0.785	超额收益	1.3539	4.421	与市场相关性

## 结论

- 浦发银行的超额收益为负，说明该股票相对于上证指数的超额收益能力相对较弱；浙江龙盛的超额收益为正，说明该股票相对于上证指数的超额收益能力相对较强；但两者均接近于0，说明两股收益中与大盘波动无关的比例很小，几乎全部受大盘影响
- 浙江龙盛的beta值相较于浦发银行的beta值更大，意味着前者与市场具有更强的相关性，更容易受大盘波动影响
- beta的t-value值普遍大于alpha的t-value值，说明前者相较于后者更为显著

## 步骤七：风险分析

代码见model.ipynb

各数据都已经获取得到，统计表格如下

股票	总风险	系统性风险	公司风险
浦发银行	0.0543	0.0392	0.1512
浙江龙盛	0.1094	0.1481	0.0387