# 数据科学工具期末作业

**2017312271 王枕戈 信息学院 信息管理与信息系统专业**

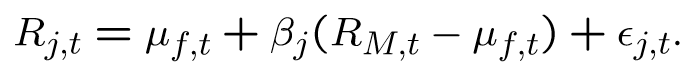
## 实验概述

验证股票数据是否满足CAPM模型的理论。运用python处理数据

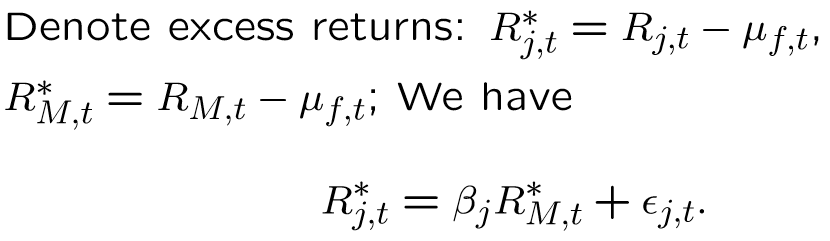
## 实验步骤

### CAMP

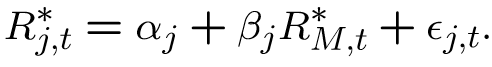
投资学、金融学中存在一个基础的理论--CAPM，主要用于寻找：单个股票的收益率与市场收益率以及无风险利率的一些关系。基本公式为：



为了用已有的中国A股数据对该模型进行验证，我们先将模型的形式进行改变。



将该模型抽象成线性规划模型，即：



如果能用数据进行建模，并对回归参数进行统计检验，如果不显著不为0（因为对参数的检验中，原假设为参数不为0），即可验证CAPM模型的正确性。

### 准备工作

课程发的数据只有每个股票的收盘价，没有市场的收益率，以及无风险利率。为此我们在以下网站进行下载：

**中国一年期国债收益率历史数据：<https://cn.investing.com/rates-bonds/china-1-year-bond-yield-historical-data>**

**上证指数：<http://quotes.money.163.com/trade/lsjysj_zhishu_000001.html>**

用一年期国债代替无风险利率，需要注意，国债的收益率都已经转化为年化收益率。应转化为日收益率，即除以（100\*253）。除以100以便转化为百分数，除以253因为一年有253个交易日。

股票每日价格，以及上证指数都取收盘价。

### 数据处理

首先利用python中的pandas包中的read\_csv方法导入数据。详细过程见源代码，process.py文件。

模型计算的日期为2006.6.8-2016.6.8共十年的时间。

由于国债数据已经是收益率，所以除了注意转化为日收益率，其他不需要改动。

股票数据为价格，应转化成收益率，在这我们用log return方法计算，rt = log(Pt/P(t-1))。当天收益率为今昨两天取ln再做差分。上证指数同理。

经过process\_stock(), process\_debt(),process\_market()三个函数的处理，最后得到data,debt, market三个DataFrame类型的数据。

进行差分运算后第一个数据类型为NaN应予以剔除。

为了之后的回归，将数据转化为数组类型：np.array(data).flatten().T

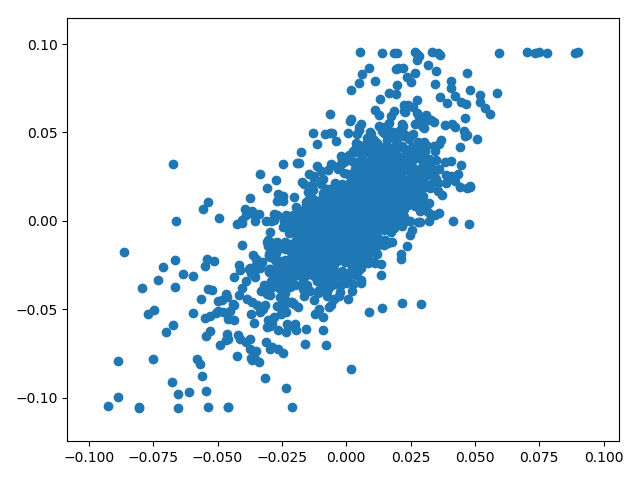
计算Ri, Rm:

Ri = data - debt  
Rm = market - debt

### 运行结果及分析

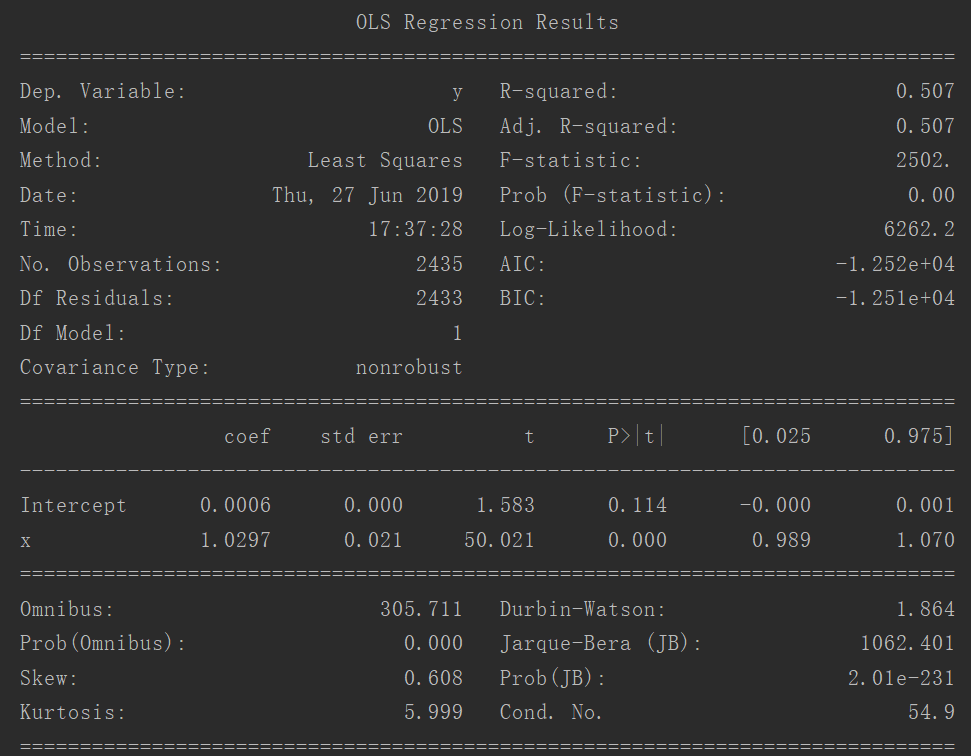
绘图：

plt.scatter(Rm, Ri)



可以看出数据大致符合线性方程。

参数的假设检验：



可以看出回归方程中，截距项为0.0006，p值为0.114，在0.05置信度的假设下，不拒绝不为0。模型拟合度R^2 = 0.507，拟合效果良好。

## 小结

实验过程中发现，国债市场的交易日期与股票市场的交易日期并不一致，即存在某些天股票开市，债市不开始；债市开市，股市不开始。

本次作业由于时间原因，只分析了第一只股票“浦发银行”，且只分析了10年的股票情况，这些是以后可以改进的地方。

通过一学期的课程学习，以及最后花费10余个小时的debug经历。我从这门课学到了不少新知识，收获颇丰。