（四) 实际金融数据回报率的统计分析和检验

陆一啸 171470235

数据文件ibm2013daily.txt股票IBM在2013年1月1日至2013年12月31日期间每日价格数据; 以收盘价数据来考虑下面的计算和统计分析:

计算日对数回报率并且给出日对数回报率的时间序列图;

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

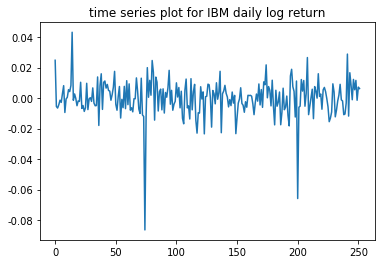
data = pd.read\_csv("./ibm2013daily.txt",index\_col=0, parse\_dates=True)

rtn = np.diff(np.log(data['Close']))

plt.plot(rtn)

plt.title('time series plot for IBM daily log return')

plt.show()



给出日对数回报率基本描述统计结果;

rtn = pd.DataFrame(data=rtn)

rtn.describe()

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| mean(%) | std(%) | min(%) | Q1(%) | median(%) | Q3(%) | max(%) |
|  | -0.011893 | -0.086419 |  |  |  | 0.04312 |

从日对数回报率数据计算月对数回报率;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 0.058402 | -0.01109 | 0.060241 | -0.05176 | 0.026697 | -0.08479 |
| 7 | 8 | 9 | 10 | 11 | 12 |
| 0.020355 | -0.06772 | 0.015839 | -0.03277 | 0.002619 | 0.042975 |

rtn = pd.DataFrame(data=rtn, columns=['rtn'])

rtn.index = data.index[1:]

rtn\_m = rtn.groupby(rtn.index.month).sum()

print(rtn\_m)

检验IBM在2013年的日对数回报率均值是否等于0;

T统计量是-0.11121628, p值是0.91153373. 以为置信水平, 不能拒绝原假设. 所以IBM在2013年的日对数回报率均值等于0.

from scipy import stats

stats.ttest\_1samp(rtn, 0)

探索\*: 测试IBM是否存在 “一月效应” (即股票在一月份回报率不同于其他月份): 首先从雅虎财经网站下载历史价格数据: 股票IBM从1962年1月1日至今每日价格数据 (IBM1962end.csv文件), 从日回报率计算月回报率数据; 之后, 把月回报率数据分为两组: 一月份的回报率和其他月份的回报率来测试这两组的方差是否相等.

Bartlett统计量40.43593506765188, p值2.031697472624353e-10. 以为置信水平, 拒绝原假设. 所以IBM存在 “一月效应”.

data\_ = pd.read\_csv('./IBM1962end.csv', index\_col=0, parse\_dates=True)

rtn\_ = np.diff(np.log(data\_['Close']))

rtn\_ = pd.DataFrame(data=rtn\_, columns=['rtn'])

rtn\_.fillna(0)

rtn\_.index = data\_.index[1:]

rtn\_jan = rtn\_[rtn\_.index.month == 1]

rtn\_others = rtn\_[rtn\_.index.month != 1]

stats.bartlett(rtn\_jan.values[:,-1], rtn\_others.values[:,-1])