

## 实验一 商业数据特征描述与模式识别

### 一、实验目的

分析沪深 300 成分股公司 10 年收益率数据时间序列，学习并练习应用 Matlab 或 Python 进行数据特征描述，并对数据特征中可能存在的模式进行探索识别。

### 二、实验内容

下载试验一数据源，具体数据分析任务如下：

1. 选择 10 家企业，分析整个时间跨度（2006-2015 年）每股收益率的概率分布，判断是否符合正态分布；
2. 继续分析以上企业的每股收益率，判断  $t+1$  时期的收益率与上一时期即  $t$  时期的收益率是否相关。

### 三、实验报告提交

1. 实验报告中应包含可以运行的代码（可拷贝运行检验，不要截图），和数据分析结果（用图形输出方法，尽可能用 `tiff` 格式，不要截图）。
2. 每次下课务必提交，无论有没有完成；下课后，任课老师作第一次打包下载；
3. 课后可以完善，也鼓励完善，完善后的定稿，再次提交文件名注第二次提交；第二次有效提交时间为第一次提交后 24 小时以内。

### 四、简要说明

1. 首先思考、检索、试验如何用程序将分布在不同页面的数据进

行读取、叠加形成一致的时间序列；

2. 绘制直方图, 初步判断是否符合正态分布; 然后试验 `matlab` 或 `python` 正态分布检验方法, 例如 Kolmogorov-Smirnov 检验函数(`kstest` 函数)、lilliefors 检验 (`lillietest` 函数) 等, 可检索、试验以上方法。

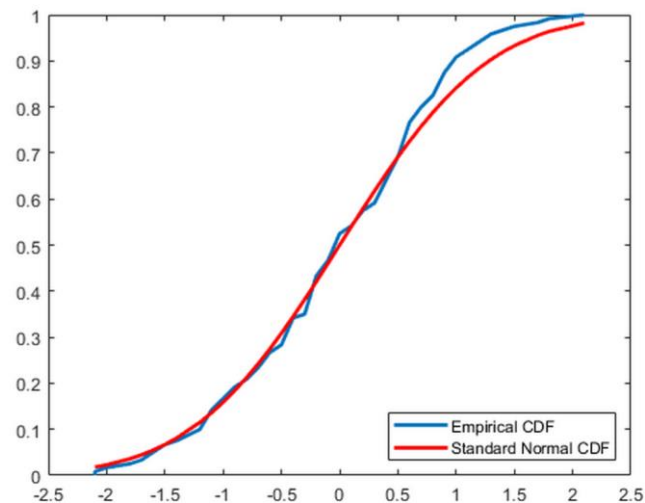


示意图 1. 经验分布与模型正态分布对比图

(注: 数据与本试验不同)

3. 针对任一公司, 将  $t$  时期收益率作  $x$  轴,  $t+1$  时期收益率作  $y$  轴, 绘制如下图所示; 并讨论探索是否具有任何可识别的数据特征。

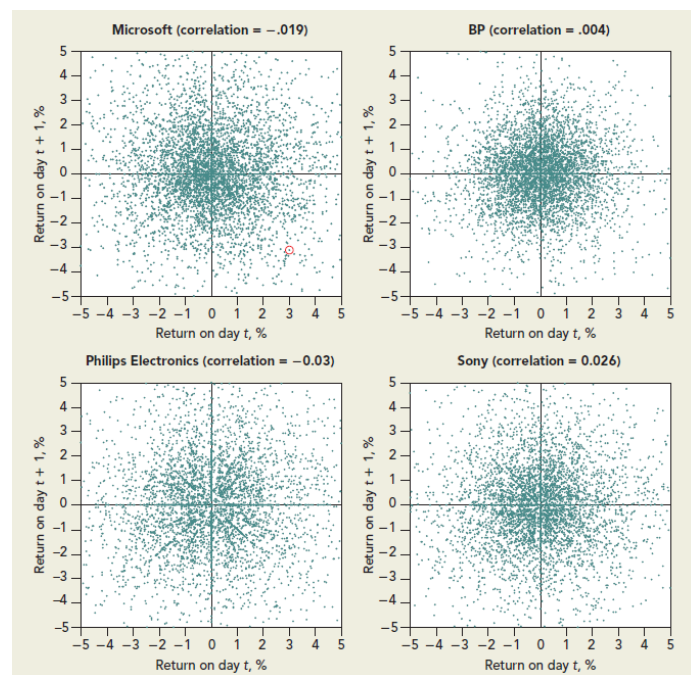


示意图 2. 上市公司个股  $t$  时期与  $t+1$  时期收益率的相关性