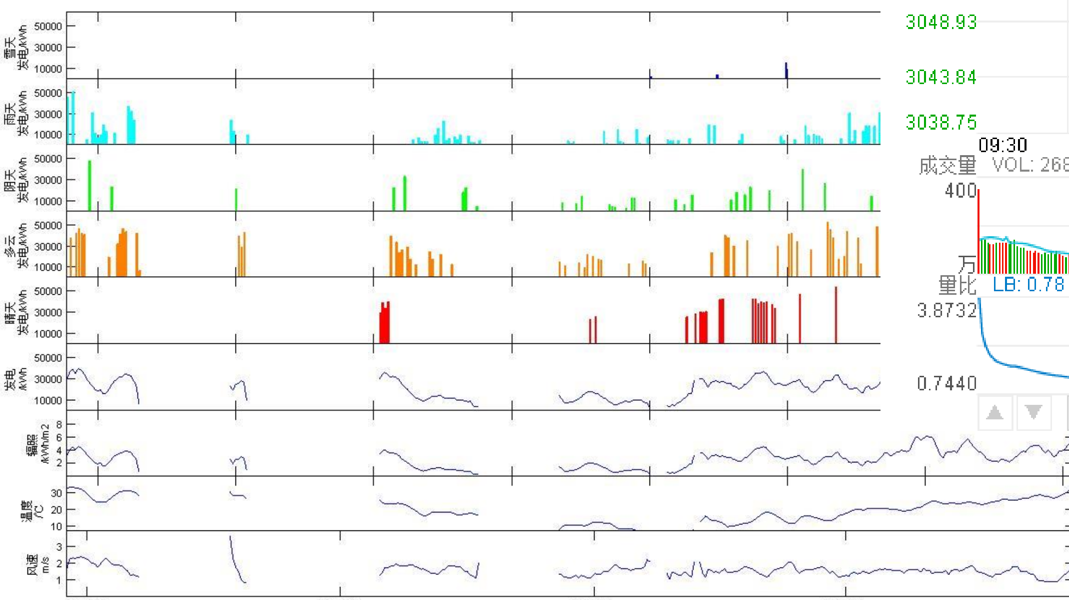
第十一课（第31-33课时）

时间序列分析和金融数据

- 时间序列及其分析的常见任务
- Python中的时间序列分析功能
- 金融数据分析基础
- 金融数据分析常见任务
- 金融数据分析实战（背景介绍）

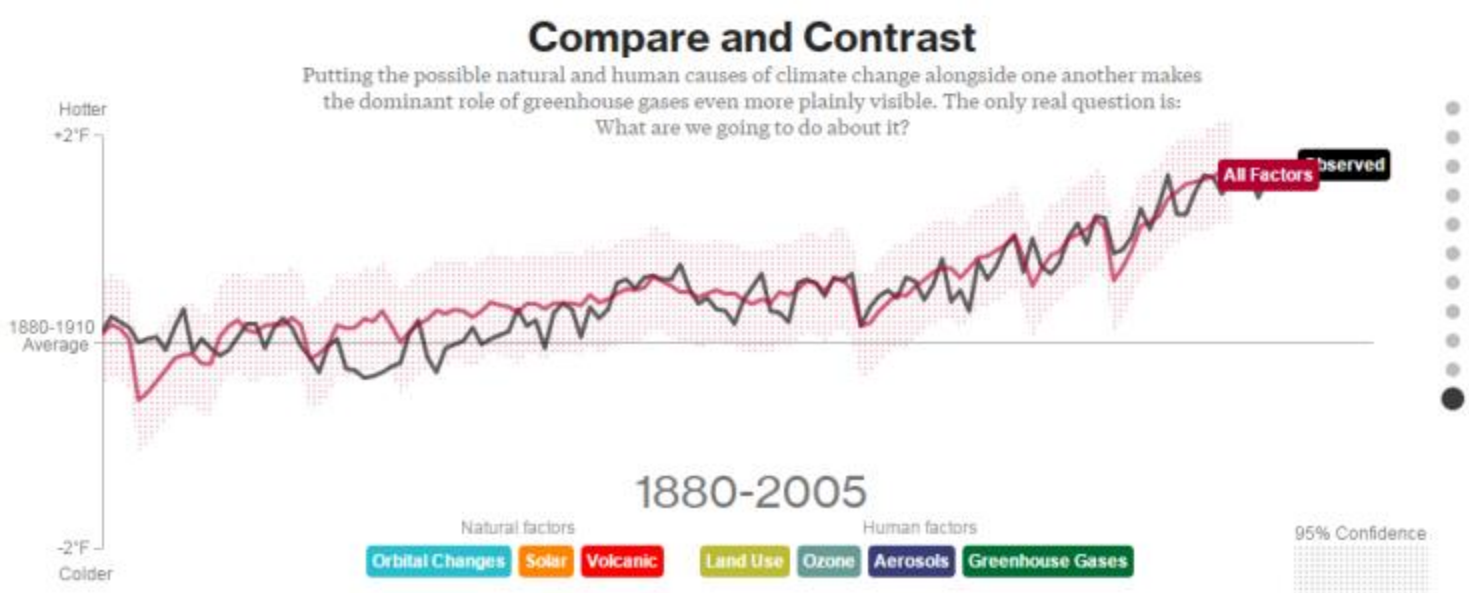
时间序列概述

- **什么是时间序列**
  - 某些量在时间上的变化，自变量为时间
  - 如：
    - 股票数据
    - 客流数据
    - 天气数据
    - 日志数据



时间序列概述

- **时间序列的特性**
  - 趋势
  - 周期性（年、季节、月、周、日）



时间序列分析概述

- **时间序列包含的要素**
  - 时间范围
  - 采样频率（间隔）
  - 随时间变化的变量

- **时间序列分析的组成**
  - 趋势
    - 模型（线性或非线性）
    - 幅度
  - 周期性（季节性）
    - 累加或者累乘
  - 噪声：随机变动，需要估计和减少
  - 其他：异常值，异常波动，丢失数据，或许暗示特殊事件

时间序列分析概述

小象学院 ChinaHadoop.cn

- **时间序列分析的主要任务**
  - 描述：
    - 解释过去：趋势、周期性、不确定性（噪声）
    - 训练模型
  - 预测：
    - 研究未来：预测未来的值
    - 验证模型，使用模型
  - 控制：
    - 锁定现在
    - 使用模型

时间序列分析常见任务

- **平滑：**
  - 去除序列中短期的效应

- **移动平均法**
  - 简单移动平均法 $$s_i = \frac{1}{2k+1}\sum_{j=-k}^{k} x_{i+j}$$
  - 加权移动平均法 $$s_i = \sum_{j=-k}^{k} w_j x_{i+j} \quad 其中 \quad \sum_{j=-k}^{k} w_j = 1$$

时间序列分析常见任务

- 平滑：
- 指数平均法：

（一） 一次指数平滑预测

当时间数列无明显的趋势变化，可用一次指数平滑预测。其预测公式为：

yt+1'=ayt+(1-a)yt' 式中，

- yt+1'--t+1期的预测值，即本期（t期）的平滑值St ；
- yt--t期的实际值；
- yt'--t期的预测值，即上期的平滑值St-1 。

（二） 二次指数平滑预测

二次指数平滑是对一次指数平滑的再平滑。它适用于具线性趋势的时间数列。其预测公式为：

yt+m=(2+am/(1-a))yt'-(1+am/(1-a))yt=(2yt'-yt)+m(yt'-yt) a/(1-a)

式中，yt= ayt-1'+(1-a)yt-1

显然，二次指数平滑是一直线方程，其截距为：(2yt'-yt)，斜率为：(yt'-yt) a/(1-a)，自变量为预测天数。

时间序列分析常见任务

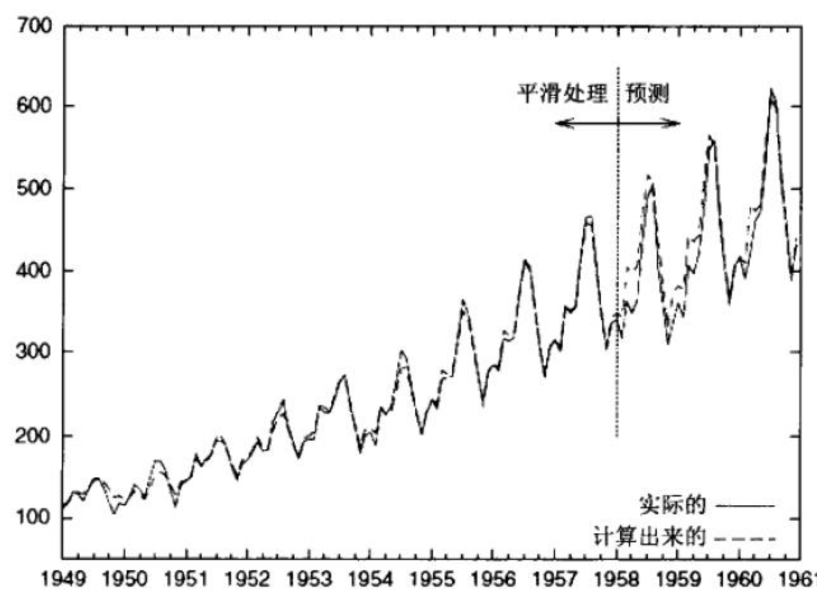
- **平滑：指数平均法：**
  - 预测值是以前观测值的加权和，且对不同的数据给予不同的权，新数据给较大的权，旧数据给较小的权

（三） 三次指数平滑预测

三次指数平滑预测是二次平滑基础上的再平滑。其预测公式是：

yt+m=(3yt'-3yt+yt)+[(6-5a)yt'-(10-8a)yt+(4-3a)yt]*am/2(1-a)2+ (yt'-2yt+yt')*a2m2/2(1-a)2

式中，yt=ayt-1+(1-a)yt-1



时间序列分析常见任务

- **相关函数**
- **自相关**
  - 自相关现象大多出现在时间序列数据中，随机变量之间不再是完全相互独立的，而是存在某种相关性
  - 平滑会增加自相关
  - 自相关数据不适于用线性回归模型

$$c(k)=\frac{\sum_{i=1}^{N-k}(x_i-\mu)(x_{i+k}-\mu)}{\sum_{i=1}^{N}(x_i-\mu)^2}$$

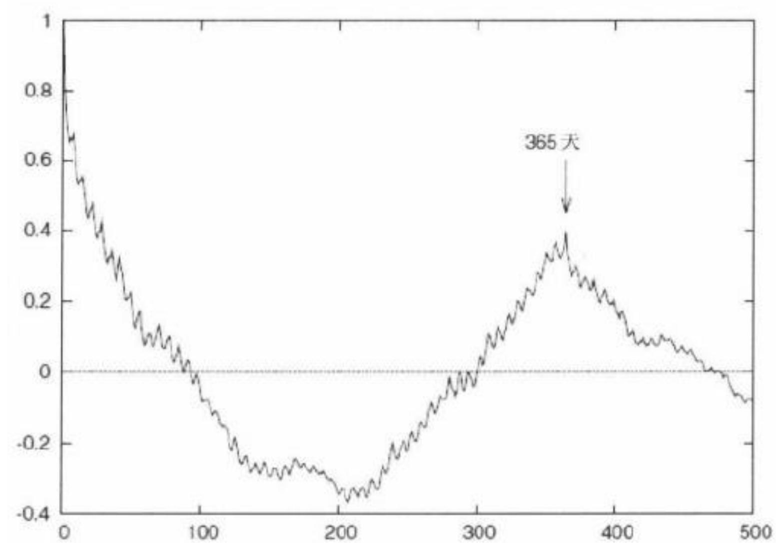


图 4-10　图 4-5 显示的呼叫中心数据的相关函数。在 365 天的地方出现了第二个高峰，数据中还包含一个以星期为周期的结构

Python时间序列模块

- **时间序列的场景**
  - 时间戳 timestamp：特定时刻
  - 固定时期 period：如2016年，2000年9月
  - 时间间隔 interval：起始~结束时间戳（period为其特例）
  - 实验或过程时间：每个时间点相对于特定起始时间的度量
- **python标准模块**
  - from datetime import ...
    - datetime
    - timedelta
    - date
    - time
    - 特定的时间格式：%H，%W等

Python时间序列模块

➢ **pandas，以时间戳作为序列索引**

- from datetime import datetime
- dates=[datetime(2016,7,15),datetime(2016,7,16),datetime(2016,7,17)]
- ts=pd.Series(np.random.randn(3),index=dates)

```
In [331]: ts
Out[331]:
2016-07-15   -0.870052
2016-07-16   -2.174847
2016-07-17    1.112486
dtype: float64
```

```
In [333]: ts.index
Out[333]: DatetimeIndex(['2016-07-15', '2016-07-16', '2016-07-17'],
dtype='datetime64[ns]', freq=None, tz=None)
```

- 自动按时间对齐

```
In [341]: ts+ts[::2]
Out[341]:
2016-07-15    2.218444
2016-07-16         NaN
2016-07-17   -1.066844
dtype: float64
```

Python时间序列模块

- **pd.to_datetime(datestr)**
  - 从包含时间信息的字符串解析时间
- **生成时间序列索引**
  - index=pd.date_range('7/17/2016',period=1000)
  - index=pd.Datetimeindex(['7/15/2016', '7/16/2016', '7/17/2016'])
- **生成日期范围：**
  - pd.date_range('7/17/2016', '7/17/2017', freq='BM')

```
In [346]: pd.date_range('7/17/2016', '7/17/2017', freq='BM')
Out[346]:
DatetimeIndex(['2016-07-29', '2016-08-31', '2016-09-30', '2016-10-31',
               '2016-11-30', '2016-12-30', '2017-01-31', '2017-02-28',
               '2017-03-31', '2017-04-28', '2017-05-31', '2017-06-30'],
              dtype='datetime64[ns]', freq='BM', tz=None)
```

  - 参考《利用Python进行数据分析》

Python时间序列模块

- **timestamp和period的互相转换**
  - pd.to_timestamp
  - pd.to_period
    - ts=pd.Series(np.random.randn(6),index=pd.date_range('7/17/2017',periods=6,freq='M'))
    - ts1=ts.to_period('M')
    - ts1.to_timestamp(how="start")

```
In [354]: ts
Out[354]:
2017-07-31    0.865774
2017-08-31   -0.110886
2017-09-30   -0.290815
2017-10-31   -0.205878
2017-11-30   -0.018218
2017-12-31    0.767338
```

```
In [356]: ts1=ts.to_period('M')

In [357]: ts1
Out[357]:
2017-07    0.865774
2017-08   -0.110886
2017-09   -0.290815
2017-10   -0.205878
2017-11   -0.018218
2017-12    0.767338
Freq: M, dtype: float64
```

```
In [359]: ts1.to_timestamp(how="start")
Out[359]:
2017-07-01    0.865774
2017-08-01   -0.110886
2017-09-01   -0.290815
2017-10-01   -0.205878
2017-11-01   -0.018218
2017-12-01    0.767338
```

Python时间序列模块

小象学院 ChinaHadoop.cn

- **重采样：resampling**
  - ts.resample
    - ts.resample("W-Wed",how='mean')...
  - 高频到低频：降采样
    - 聚合：需要考虑数据点的归属和标记问题
  - 低频到高频：升采样
    - 插值：
  - 其他："W-Wed",→ "W-Fri “
  - OHLC重采样
    - ts.resample('M',how='ohlc')

```
In [364]: ts.resample('M',how='ohlc')
Out[364]:
                open      high       low     close
2017-07-31  0.865774  0.865774  0.865774  0.865774
2017-08-31 -0.110886 -0.110886 -0.110886 -0.110886
2017-09-30 -0.290815 -0.290815 -0.290815 -0.290815
2017-10-31 -0.205878 -0.205878 -0.205878 -0.205878
2017-11-30 -0.018218 -0.018218 -0.018218 -0.018218
2017-12-31  0.767338  0.767338  0.767338  0.767338
```

Python时间序列模块

➢ 绘图：读入数据

- sh1=pd.read_csv('sh000001.csv',parse_dates=True,index_col=1)

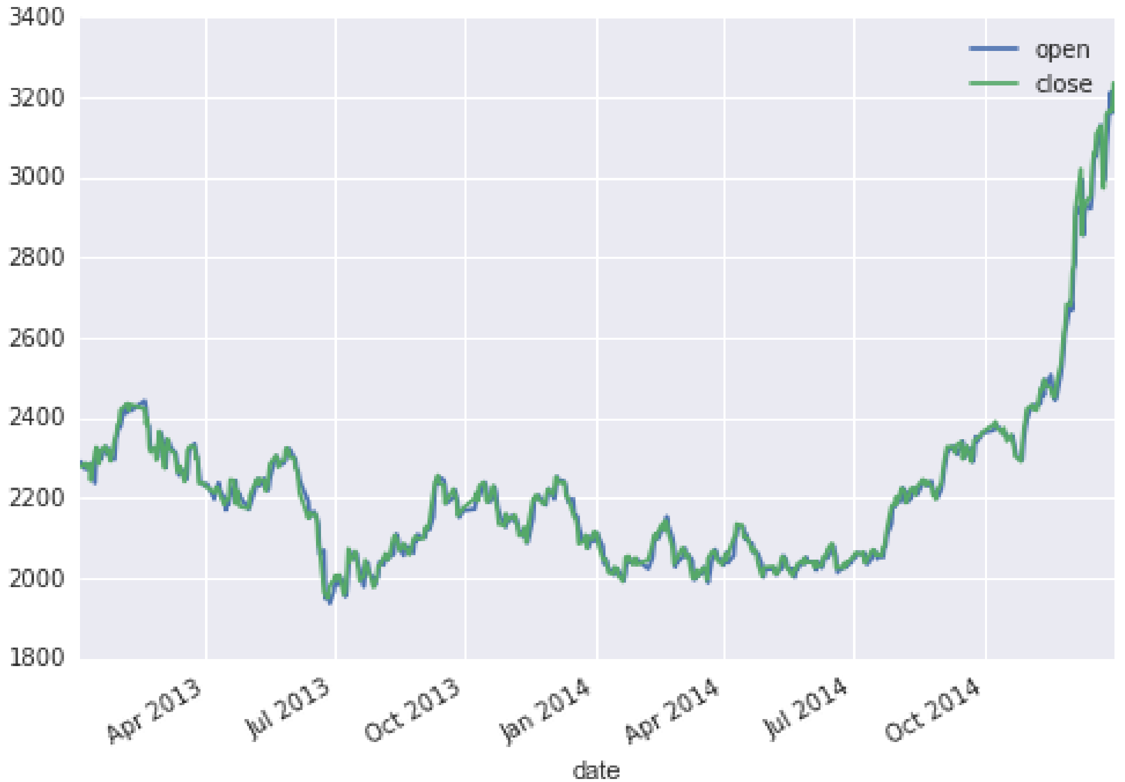
```
In [380]: sh1.head()
Out[380]:
           index_code     open    close      low     high       volume  \
date
2014-12-31   sh000001  3172.60  3234.68  3157.26  3239.36  40599852100
2014-12-30   sh000001  3160.80  3165.82  3130.35  3190.30  39772532000
2014-12-29   sh000001  3212.56  3168.02  3126.94  3223.86  51011143900
2014-12-26   sh000001  3078.01  3157.60  3064.18  3164.16  46070093200
2014-12-25   sh000001  2992.46  3072.54  2969.87  3073.35  37694777600

                   money    change
date
2014-12-31  4.323200e+11  0.021752
2014-12-30  4.372630e+11 -0.000695
2014-12-29  5.559040e+11  0.003298
2014-12-26  4.889150e+11  0.027686
2014-12-25  3.790180e+11  0.033643
```

Python时间序列模块

➢ **绘图**

```
In [385]: sh1[['open','close']].plot()
Out[385]: <matplotlib.axes._subplots.AxesSubplot at 0x12db7730>
```



Python时间序列模块

- **移动窗口函数**
  - rolling_mean
  - rolling_std
  - …
- **指数加权函数**
  - pd.ewma(, span=)
- **二元移动窗口函数**
  - rolling_corr
  - sz1=pd.read_csv('sz399001.csv',parse_dates=True,index_col=1)
  - corr=pd.rolling_corr(sh1.change,sz1.change,125,min_periods=100)

Python时间序列模块

- **二元移动窗口函数**
  - corr.plot()

```
In [392]: corr.plot()
Out[392]: <matplotlib.axes._subplots.AxesSubplot at 0x1344f8f0>
```

- **货币的时间价值**
  - P（Principal）——本金，又称期初额，现值（PV）
  - i（The rate of interest）——利率，通常指每年利息与本金之比
  - I（Interest）——利息
  - S（Summation）——本金与利息之和，本利和，终值（FV）
- **单利**

$$I = P \times i \times t$$

终值计算：$S=P+P\times i\times t$

现值计算：$P=S-I$

金融数据分析基础

- **复利**
  - 每经过一个计息期，要将所生利息加入本金再计利息，逐期滚算，“利滚利”
  - 复利终值：

$$S = P \times (1+i)^n$$

  - 复利初值：

$$P = S \times \frac{1}{(1+i)^n}$$

  - 名义利率和实际利率
    - 当利息在一年内要复利几次，给出的年利率叫做名义利率
    - 例：本金1000元，投资5年，利率8%
    - 多次计息时，实际利率比名义利率更高

- **复利【练习】**
  - 信用卡分期利息计算
    - 分期1万元，12期还清（等额本息，每月1期），每期利息为0.78%，试计算实际利率

金融数据分析基础

计息次数（m）	$F = P\left(1+\frac{i}{m}\right)^{mn}$	$P = F\left(1+\frac{i}{m}\right)^{-mn}$
1次（年）	$F = 100\left(1+\frac{0.08}{1}\right)^{(1)(3)} = 125.97$	$F = 100\left(1+\frac{0.08}{1}\right)^{-(1)(3)} = 79.38$
2次（半年）	$F = 100\left(1+\frac{0.08}{2}\right)^{(2)(3)} = 126.53$	$F = 100\left(1+\frac{0.08}{2}\right)^{-(2)(3)} = 79.03$
4次（季）	$F = 100\left(1+\frac{0.08}{4}\right)^{(4)(3)} = 126.82$	$F = 100\left(1+\frac{0.08}{4}\right)^{-(4)(3)} = 78.85$
特点	一年中计息次数越多，终值越大。	一年中计息次数越多，现值越小。

➢ 连续复利

$$p_n = \lim_{m\to\infty} p_0\left(1+\frac{i}{m}\right)^{mn} = p_0 \lim_{m\to\infty}\left[\left(1+\frac{i}{m}\right)^{\frac{m}{i}}\right]^{ni} = p_0 e^{ni}$$

```
In [396]: (1+0.08)
Out[396]: 1.08

In [397]: np.exp(0.08)
Out[397]: 1.0832870676749586
```

金融数据分析基础

- **现金流**
- **固定现金流：按揭、养老金、保险、分期**
- **变化现金流： 投资**
  - 投资依据1，净现值（NPV）
    - 投资方案所产生的现金净流量以资金成本为贴现率折现之后与原始投资额现值的差额
  - 投资依据2：内部收益率（IRR）
    - 资金流入现值总额与资金流出现值总额相等、净现值等于零时的折现率。

- **风险与收益**
  - 风险越大，预期回报也就越高
    - 假定投资国债的回报为5%，投资股票的回报和相应的概率如下

概率	回报
0.05	+50%
0.25	+30%
0.40	+10%
0.25	−10%
0.05	−30%

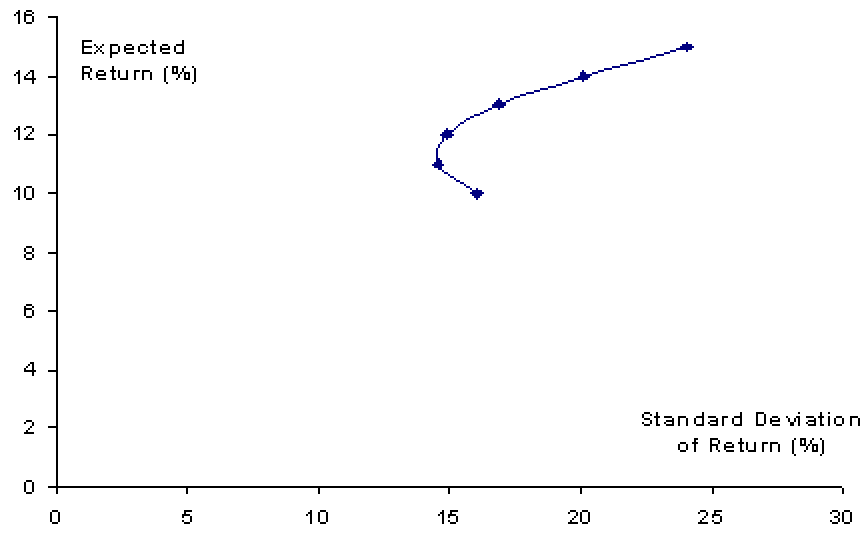
预期回报 =10%

回报的标准差 =18.97%

➢ 风险投资组合

$$\mu_P = w_1\mu_1 + w_2\mu_2 \qquad \sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho w_1 w_2 \sigma_1 \sigma_2}$$

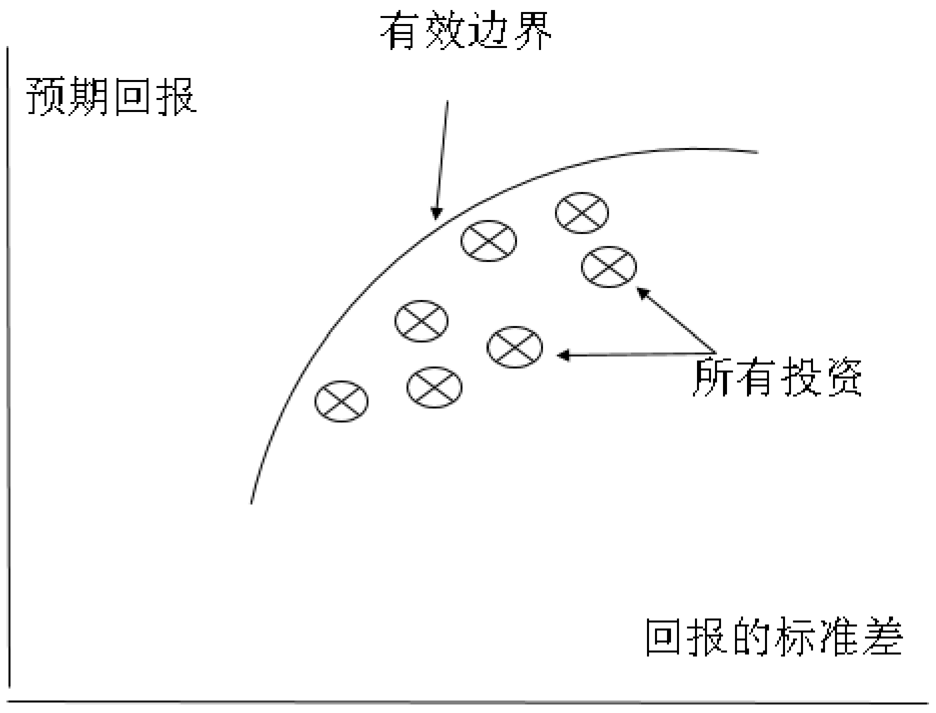
$$\mu_1 = 10\%$$
$$\mu_2 = 15\%$$
$$\sigma_1 = 16\%$$
$$\sigma_2 = 24\%$$
$$\rho = 0.2$$



金融数据分析基础

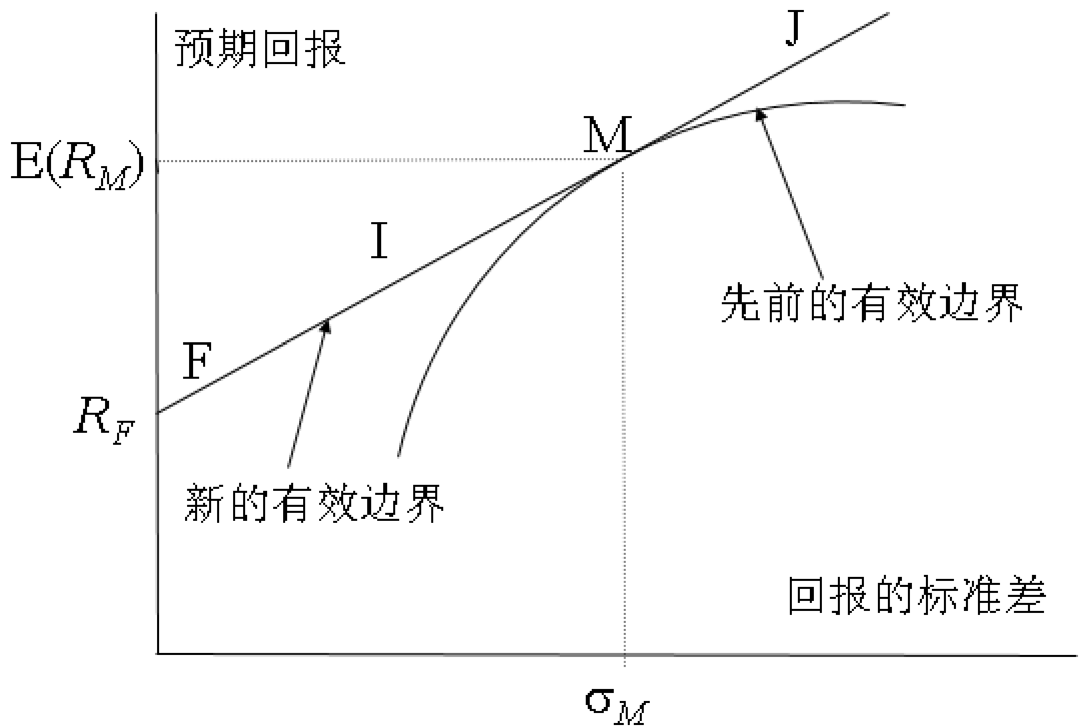
小象学院 ChinaHadoop.cn

- **风险资产的有效边界**



金融数据分析基础

- 所有投资资产的有效边界



- **系统与非系统风险**
  - 投资收益与市场收益之间的最佳线性拟合关系

$$R = \alpha + \beta R_M + \varepsilon$$

系统风险(不可分散)

非系统风险(可分散)

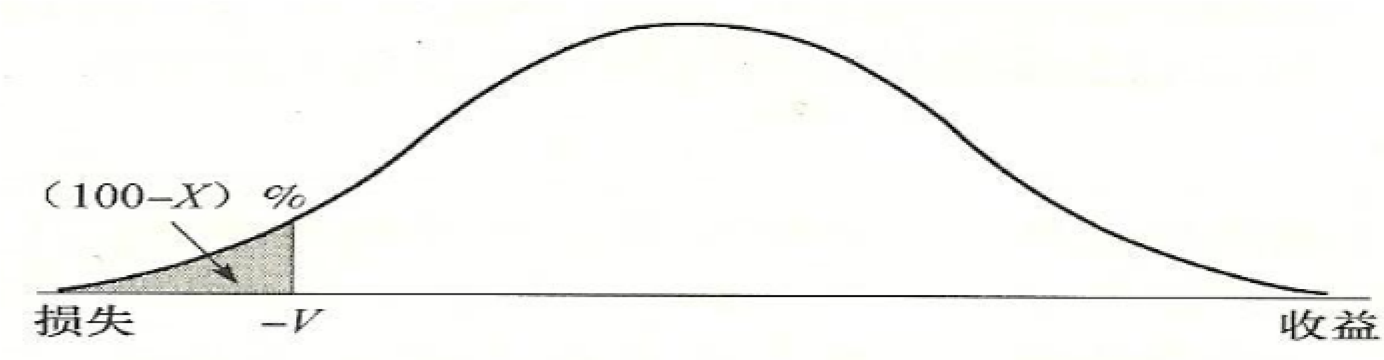
金融数据分析的常见任务

- **量化分析**
  - 应用：量化投资
  - 市场分析：趋势、波动率

- **风险计量与管理：**
  - 市场风险、利率风险、信用风险
  - 应用：风险评估、征信

金融数据分析中的常见任务和方法

- **风险计量：VaR在险价值**
  - 某交易组合T天时损失不超过V元的概率为X%



  - 测算方法
    - 历史模拟方法
    - Monte Carlo模拟方法

金融数据分析中的常见任务和方法

小象学院 ChinaHadoop.cn

- **波动率计量**
  - 某个变量的波动率定义为该变量在单位时间内连续复利收益率的标准差
    - 假设S_i是某个变量在第i天的取值，那么日波动率就可以表示为$\ln(S_i/S_{i\text{-}1})$的标准差；时间周期为T的波动率为$\ln(S_T/S_0)$的标准差

$$S_T = S_{T-1}e^{\delta_T} = S_{T-2}e^{\delta_{T-1}}e^{\delta_T}\ldots = S_0e^{\delta_1}\cdots e^{\delta_T} = S_0e^{\delta_1+\cdots\delta_T}$$

$$T\text{天连续复利收益率}\Delta_T = \delta_1 + \cdots + \delta_T = \ln\left(\frac{S_T}{S_0}\right);$$

当T很小时，连续复利收益率Δ_T也较小，那么

$$S_T = S_0e^{\Delta_T} \approx S_0(1+\Delta_T) \Rightarrow \Delta_T \approx \frac{S_T - S_0}{S_0}$$

$\Rightarrow$连续复利收益率的方差$\approx$变量百分比变化的方差

金融数据分析中的常见任务和方法

- **波动率计量**

【定律】不定性随时间的平方根成正比！

在$\{\delta_i, i=1,2,\ldots\}$独立同分布的假设下，有：

$$\sqrt{\mathrm{Var}(\delta_1+\cdots+\delta_T)}=\sqrt{\mathrm{T}\cdot Var(\delta_i)}=\sqrt{\mathrm{T}}\cdot\sqrt{Var(\delta_i)}$$

$$\Rightarrow \sigma_{\text{year}}=\sqrt{252}\sigma_{\text{day}};\quad \sigma_{\text{year}}=\sqrt{52}\sigma_{\text{week}}$$

- **方法1：使用历史数据**

金融数据分析常见任务和方法

小象学院 ChinaHadoop.cn

➤ 波动率计量

天数	股票闭市价	价格比 S_i/S_{i-1}	每天回报 $u_i=\ln(S_i/S_{i-1})$	天数	股票闭市价	价格比 S_i/S_{i-1}	每天回报 $u_i=\ln(S_i/S_{i-1})$
0	20.00			11	21.00	1.012 05	0.011 98
1	20.10	1.005 00	0.004 99	12	21.10	1.004 76	0.004 75
2	19.90	0.990 05	−0.010 00	13	20.90	0.990 52	−0.009 52
3	20.00	1.005 03	0.005 01	14	20.90	1.000 00	0.000 00
4	20.50	1.025 00	0.024 69	15	21.25	1.016 75	0.016 61
5	20.25	0.987 80	−0.012 27	16	21.40	1.007 06	0.007 03
6	20.90	1.032 10	0.031 59	17	21.40	1.000 00	0.000 00
7	20.90	1.000 00	0.000 00	18	21.25	0.992 99	−0.007 03
8	20.90	1.000 00	0.000 00	19	21.75	1.023 53	0.023 26
9	20.75	0.992 82	−0.007 20	20	22.00	1.011 49	0.011 43
10	20.75	1.000 00	0.000 00				

小象学院 ChinaHadoop.cn

- **波动率计量**

市场变量在第i天末的价格为S_i,第i天连续复利收益率$u_i=\ln(\frac{S_i}{S_{i-1}})$

波动率Var(u_i)的无偏估计：$\sigma_n^2=\frac{1}{m-1}\sum_{i=1}^{m}(u_{n-i}-\bar{u})^2 \ldots\ldots\ldots\ldots(9-2)$

其中：$\bar{u}=\frac{1}{m}\sum_{i=1}^{m}u_{n-i}$

金融数据分析中的常见任务和方法

小象学院 ChinaHadoop.cn

- **波动率计量**

1.
计算简单，在前一天波动率预测值基础上更新一下即可得到第二天波动率预测值
2.
λ衡量波动率对最新市场价格百分比变化的敏感度
3.
Morgan(1994)研究表明，λ=0.94常见

指数加权移动平均模型(*EWMA*)

$\alpha_{i+1}=\lambda\alpha_i$，其中$\lambda\in(0,1)$，那么：

$$\alpha_1+\alpha_2+...+\alpha_n=\alpha_1+\lambda\alpha_1+...+\lambda^{m-1}\alpha_1=1$$

$$\Rightarrow\frac{1-\lambda^m}{1-\lambda}\alpha_1=1\Rightarrow\alpha_1=\frac{1-\lambda}{1-\lambda^m}\approx1-\lambda$$

$$\sigma_n^2=\alpha_1u_{n-1}^2+\alpha_2u_{n-2}^2+...+\alpha_mu_{n-m}^2$$
$$=\alpha_1u_{n-1}^2+\lambda\alpha_1u_{n-2}^2+...+\lambda^{m-1}\alpha_1u_{n-m}^2$$

$$\sigma_{n-1}^2=\alpha_1u_{n-2}^2+\alpha_2u_{n-3}^2+...+\alpha_mu_{n-m-1}^2$$
$$=\alpha_1u_{n-2}^2+\lambda\alpha_1u_{n-3}^2+...+\lambda^{m-1}\alpha_1u_{n-m-1}^2$$

$$\sigma_n^2-\lambda\sigma_{n-1}^2=\alpha_1u_{n-1}^2-\lambda^m\alpha_1u_{n-m-1}^2\approx(1-\lambda)u_{n-1}^2$$

$$\Rightarrow\sigma_n^2-\lambda\sigma_{n-1}^2=(1-\lambda)u_{n-1}^2..............(9-8)$$

$$\Rightarrow\sigma_n^2=\sum_{i=1}^{m}\alpha_iu_{n-i}^2=\frac{1-\lambda}{1-\lambda^m}\sum_{i=1}^{m}\lambda^{i-1}u_{n-i}^2\approx(1-\lambda)\sum_{i=1}^{m}\lambda^{i-1}u_{n-i}^2$$

- **波动率计量**

GARCH(1,1)模型

模型：

$$\sigma_n^2=\gamma V_L+\alpha\cdot u_{n-1}^2+\beta\sigma_{n-1}^2 \cdots\cdots\cdots\cdots\cdots\cdots (10-9)$$

满足：γ，α，β皆为正数，且$\gamma+\alpha+\beta=1$

(1)当$\gamma=0$，$\alpha=1\text{-}\lambda$，$\beta=\lambda$时，模型退化为EWMA模型

(2)该模型可推广为GARCH(p,q)

(3)一般要求$\alpha+\beta<1$

(4)有时记$\omega=\gamma V_L$

GARCH(1,1)模型

$$\sigma_n^2=\gamma V_L+\alpha\cdot u_{n-1}^2+\beta\sigma_{n-1}^2 \cdots\cdots\cdots\cdots\cdots\cdots (10-9)$$

若记$\omega=\gamma V_L$，则：

$$\sigma_n^2=\omega+\alpha\cdot u_{n-1}^2+\beta\sigma_{n-1}^2=\omega+\alpha\cdot u_{n-1}^2+\beta(\omega+\alpha\cdot u_{n-2}^2+\beta\sigma_{n-2}^2)$$

$$=\omega+\beta\omega+\beta\omega^2+\alpha\cdot u_{n-1}^2+\alpha\cdot\beta\cdot u_{n-2}^2+\alpha\cdot\beta^2\cdot u_{n-3}^2+\beta^3\cdot\sigma_{n-3}^2$$

$$=\cdots\cdots$$

可见：

(1)u_{n-i}^i的权重为$\alpha\beta^{i-1}$，以β的指数速度下降。

(2)数据越新，权重越大

- **波动率计量**

GARCH(1,1)模型

$$\sigma_n^2 = \gamma V_L + \alpha \cdot u_{n-1}^2 + \beta \sigma_{n-1}^2 \quad\cdots\cdots\cdots\cdots\cdots\cdots(10-9)$$

若记$\omega = \gamma V_L$，则：

$$\sigma_n^2 = \omega + \alpha \cdot u_{n-1}^2 + \beta \sigma_{n-1}^2 = \omega + \alpha \cdot u_{n-1}^2 + \beta(\omega + \alpha \cdot u_{n-2}^2 + \beta \sigma_{n-2}^2)$$

$$= \omega + \beta\omega + \beta\omega^2 + \alpha \cdot u_{n-1}^2 + \alpha \cdot \beta \cdot u_{n-2}^2 + \alpha \cdot \beta^2 \cdot u_{n-3}^2 + \beta^3 \cdot \sigma_{n-3}^2$$

$$= \cdots\cdots$$

可见：

(1)u_{n-i}^i的权重为$\alpha\beta^{i-1}$，以β的指数速度下降。

(2)数据越新，权重越大

- **估计波动率计量模型中的参数**

估计$GARCH(1,1)$或者$EWMA$模型中参数

假如随机变量U_i服从均值$N(0,v_i)$。其观察值为$u_1,u_2,...,u_m$.

m个观察值刚好为$u_1,u_2,...,u_m$的概率为

$$\prod_{i=1}^{m}\left[\frac{1}{\sqrt{2\pi v_i}}\exp\left\{\frac{-u_i^2}{2v_i}\right\}\right]$$

要使上式最大化，等同于使下式最大化，即

$$\sum_{i=1}^{m}\left[-\ln(v_i)-\frac{u_i^2}{v}\right]$$

用迭代法可使得上式的值达到最大

➢ 使用模型预测波动率

采用$GARCH(1,1)$模型来预测波动率

$GARCH(1,1)$模型下，第n-1天结束时估算第n天方差为

$$\sigma_n^2 = (1-\alpha-\beta)V_L + \alpha \cdot u_{n-1}^2 + \beta \cdot \sigma_{n-1}^2$$

$$\Rightarrow \sigma_n^2 - V_L = \alpha \cdot (u_{n-1}^2 - V_L) + \beta \cdot (\sigma_{n-1}^2 - V_L)$$

$$\Rightarrow \sigma_{n+t}^2 - V_L = \alpha \cdot (u_{n+t-1}^2 - V_L) + \beta \cdot (\sigma_{n+t-1}^2 - V_L)$$

因为$Eu_{n+t-1}^2 = \sigma_{n+t-1}^2$,所以

$$\begin{aligned}\mathrm{E}[\sigma_{n+t}^2 - V_L] &= \alpha \cdot (\sigma_{n+t-1}^2 - V_L) + \beta \cdot (\sigma_{n+t-1}^2 - V_L) \\ &= (\alpha+\beta) \cdot (\sigma_{n+t-1}^2 - V_L)\end{aligned}$$

$$\mathrm{E}[\sigma_{n+t}^2] = V_L + (\alpha+\beta)^t \cdot (\sigma_n^2 - V_L) \ldots\ldots\ldots\ldots\ldots\ldots (9-14)$$

【分析】

(1)$\alpha+\beta=1$时，退化为EWMA模型，$\mathrm{E}[\sigma_{n+t}^2] = V_L$

(2)$\alpha+\beta<1$时,方差具有均值回归特性,$GARCH$模型

(3)$\alpha+\beta>1$时,方差逃离

金融数据分析中的常见任务和方法

- **市场风险度量**
- **历史模拟法**
  - 假设使用至今总共n 天的历史数据
  - vi 为第 i天的某市场变量的值
  - 共 n-1 个场景（simulation trials）
  - 根据第 i个场景，则明天的市场变量值 (第 n+1天) 可被预测为

$$v_n \frac{v_i}{v_{i-1}}$$

金融数据分析中的常见任务和方法

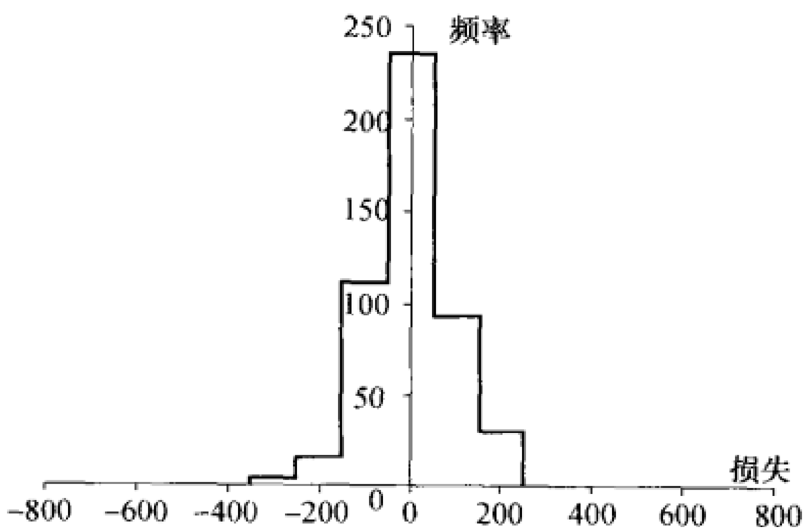
➢ 市场风险度量

表 12-1 用于演示 VaR 计算过程的投资组合

指数	组合价值（以 1 000 美元计）
DJIA	4 000
FTSE 100	3 000
CAC 40	1 000
Nikkei 225	2 000
总计	10 000

表 12-2 采用历史模拟法计算 VaR 所需要的数据

天数	日期	DJLA	FTSE 100	CAC 40	Nikkei 225
0	Aug. 7，2006	11 219. 38	5 828. 8	4 956. 34	15 154. 06
1	Aug. 8，2006	11 173. 59	5 818. 1	4 967. 95	15 464. 66
2	Aug. 9，2006	11 076. 18	5 860. 5	5 025. 15	15 656. 59
3	Aug. 10，2006	11 124. 37	5 823. 4	4 976. 64	15 630. 91
⋮	⋮	⋮	⋮	⋮	⋮
499	Sept. 24，2008	10 825. 17	5 095. 6	4 114. 54	12 115. 03
500	Sept. 25，2008	11 022. 06	5 197. 0	4 226. 81	12 006. 53



如 11. 6 节所示，10 天展望期及 99% 置信区间的 VaR 等于 $\sqrt{10}$ 乘以 1 天展望期及 99% 置信区间的 VaR，10 天的 VaR 等于

$$\sqrt{10} \times 247\,571 = 782\,889$$

即 782 889 美元。

金融数据分析中的常见任务和方法

- **市场风险度量**
- **VaR值的置信区间**
  - 自助法（ bootstrap ）
  - 设有 500 个场景（500个变化率值）
  - 从500个场景中有放回的抽样 500,000 次，得到1000 个包含500个场景的样本
  - 计算每个样本的95% VaR ，可得VaR的置信区间

金融数据分析实战（背景介绍）

- ➢ **数据**
- ➢ **综合股指数据（1年期）**

sh000001.csv
sh000016.csv
sh000300.csv
sz399001.csv
sz399005.csv
sz399006.csv
sz399905.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	index_cod	date	open	close	low	high	volume	money	change				
2	sh000001	2014/12/31	3172.6	3234.68	3157.26	3239.36	4.06E+10	4.32E+11	0.021752				
3	sh000001	2014/12/30	3160.8	3165.82	3130.35	3190.3	3.98E+10	4.37E+11	-0.00069				
4	sh000001	2014/12/29	3212.56	3168.02	3126.94	3223.86	5.1E+10	5.56E+11	0.003298				
5	sh000001	2014/12/26	3078.01	3157.6	3064.18	3164.16	4.61E+10	4.89E+11	0.027686				
6	sh000001	2014/12/25	2992.46	3072.54	2969.87	3073.35	3.77E+10	3.79E+11	0.033643				
7	sh000001	2014/12/24	3039.21	2972.53	2934.91	3050.51	3.77E+10	3.79E+11	-0.01981				
8	sh000001	2014/12/23	3085.08	3032.61	3025.67	3136.84	4.38E+10	4.19E+11	-0.03032				
9	sh000001	2014/12/22	3129.27	3127.45	3090.51	3189.87	6.79E+10	6.24E+11	0.006064				
10	sh000001	2014/12/19	3053.08	3108.6	3018.42	3117.53	5.21E+10	5.16E+11	0.016705				
11	sh000001	2014/12/18	3062.8	3057.52	3030.32	3089.79	4.36E+10	4.67E+11	-0.00114				
12	sh000001	2014/12/17	3031.95	3061.02	2993.33	3076.6	5.43E+10	5.80E+11	0.013074				
13	sh000001	2014/12/16	2953.81	3021.52	2943.91	3021.9	4.54E+10	4.93E+11	0.023057				
14	sh000001	2014/12/15	2921.45	2953.42	2890.9	2960.23	4E+10	4.11E+11	0.00519				
15	sh000001	2014/12/12	2929.36	2938.17	2914.96	2962.51	4.09E+10	4.20E+11	0.004248				
16	sh000001	2014/12/11	2912.35	2925.74	2892.61	2965.68	4.83E+10	4.80E+11	-0.00485				
17	sh000001	2014/12/10	2855.94	2940.01	2807.68	2946.71	5.13E+10	5.35E+11	0.029317				
18	sh000001	2014/12/9	2992.49	2856.27	2834.59	3091.32	7.72E+10	7.93E+11	-0.0543				
19	sh000001	2014/12/8	2907.82	3020.26	2879.85	3041.66	5.88E+10	5.93E+11	0.028121				
20	sh000001	2014/12/5	2926.57	2937.65	2813.05	2978.03	6.41E+10	6.39E+11	0.013172				
21	sh000001	2014/12/4	2783.47	2899.46	2772.43	2900.51	5.33E+10	5.09E+11	0.043148				
22	sh000001	2014/12/3	2768.68	2779.53	2733.87	2824.18	5.62E+10	5.30E+11	0.005782				
23	sh000001	2014/12/2	2667.82	2763.55	2665.69	2777.37	4.38E+10	3.97E+11	0.031114				
24	sh000001	2014/12/1	2691.73	2680.16	2668.84	2720.74	4.47E+10	4.01E+11	-0.001				
25	sh000001	2014/11/28	2629.63	2682.84	2622.06	2683.18	4.66E+10	4.02E+11	0.019901				
26	sh000001	2014/11/27	2615.37	2630.49	2599.11	2631.4	3.64E+10	3.39E+11	0.010037				
27	sh000001	2014/11/26	2572.65	2604.35	2570.4	2605.07	3.37E+10	3.17E+11	0.014312				
28	sh000001	2014/11/25	2532	2567.6	2527.08	2568.38	3.14E+10	2.82E+11	0.013707				
29	sh000001	2014/11/24	2505.53	2532.88	2495.52	2546.75	3.63E+10	3.30E+11	0.018533				
30	sh000001	2014/11/21	2452.64	2486.79	2446.65	2488.2	2.12E+10	1.98E+11	0.013916				

sh000001

金融数据分析实战（背景介绍）

小象学院
ChinaHadoop.cn

- 数据

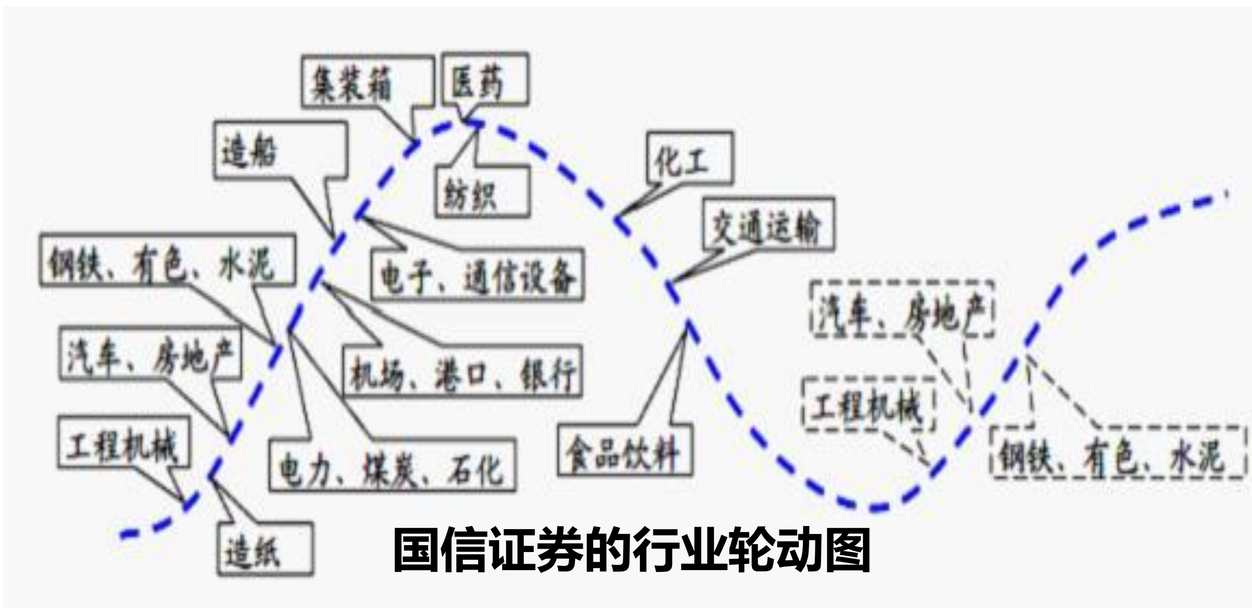
- 个股日线（1年期）

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	code	date	open	high	low	close	change	volume	money	traded_ma	market_val	turnover	adjust_pric	report_typ	report_dat	PE_TTM	PS_TTM	PC_TTM	PB
2	sh600000	2014/12/31	15.45	15.79	15.11	15.69	0.021484	4.69E+08	7.27E+09	2.34E+11	2.93E+11	0.031417	130.2546	#######	#######	6.375742	2.515326	3.401388	1.265
3	sh600000	2014/12/30	14.95	15.5	14.83	15.36	0.027425	4.44E+08	6.77E+09	2.29E+11	2.87E+11	0.02973	127.5151	#######	#######	6.241646	2.462423	3.32985	1.238
4	sh600000	2014/12/29	15.4	15.88	14.71	14.95	0.012187	6.25E+08	9.56E+09	2.23E+11	2.79E+11	0.041897	124.1114	#######	#######	6.075039	2.396694	3.240966	1.205
5	sh600000	2014/12/26	14.3	14.84	14.19	14.77	0.036491	4.67E+08	6.79E+09	2.2E+11	2.76E+11	0.031293	122.617	#######	#######	6.001893	2.367837	3.201944	1.19
6	sh600000	2014/12/25	13.75	14.27	13.53	14.25	0.058692	4.56E+08	6.36E+09	2.13E+11	2.66E+11	0.030539	118.3001	#######	#######	5.790589	2.284474	3.089216	1.149
7	sh600000	2014/12/24	14.11	14.2	13.31	13.46	-0.04607	4.18E+08	5.72E+09	2.01E+11	2.51E+11	0.027983	111.7418	#######	#######	5.469569	2.157827	2.917955	1.085
8	sh600000	2014/12/23	14.4	14.98	14.08	14.11	-0.03883	4.42E+08	6.4E+09	2.11E+11	2.63E+11	0.029591	117.138	#######	#######	5.733704	2.262032	3.058868	1.137
9	sh600000	2014/12/22	14.18	15.2	14.14	14.68	0.041874	6.83E+08	1E+10	2.19E+11	2.74E+11	0.045799	121.8699	#######	#######	5.965325	2.35341	3.182436	1.183
10	sh600000	2014/12/19	13.96	14.2	13.63	14.09	0.014399	4.36E+08	6.1E+09	2.1E+11	2.63E+11	0.029233	116.9719	#######	#######	5.725573	2.258824	3.05453	1.136
11	sh600000	2014/12/18	14.22	14.34	13.71	13.89	-0.01559	5E+08	7.01E+09	2.07E+11	2.59E+11	0.033481	115.3115	#######	#######	5.6443	2.226761	3.011172	1.119
12	sh600000	2014/12/17	13.49	14.39	13.33	14.11	0.061701	8.7E+08	1.2E+10	2.11E+11	2.63E+11	0.05828	117.1379	#######	#######	5.7337	2.262031	3.058866	1.137
13	sh600000	2014/12/16	12.7	13.3	12.65	13.29	0.041536	4.39E+08	5.71E+09	1.98E+11	2.48E+11	0.029397	110.3304	#######	#######	5.400485	2.130572	2.881099	1.071
14	sh600000	2014/12/15	12.8	12.82	12.46	12.76	-0.01695	3.35E+08	4.23E+09	1.9E+11	2.38E+11	0.022451	105.9305	#######	#######	5.185116	2.045606	2.766202	1.02
15	sh600000	2014/12/12	13.05	13.38	12.78	12.98	-0.00536	3.11E+08	4.07E+09	1.94E+11	2.42E+11	0.020811	107.7569	#######	#######	5.274514	2.080875	2.813895	1.046
16	sh600000	2014/12/11	12.96	13.55	12.85	13.05	-0.00836	4.43E+08	5.85E+09	1.95E+11	2.43E+11	0.029657	108.338	#######	#######	5.302959	2.092097	2.82907	1.052
17	sh600000	2014/12/10	12.7	13.25	12.21	13.16	0.040316	6.2E+08	7.9E+09	1.96E+11	2.45E+11	0.041565	109.2512	#######	#######	5.34766	2.109732	2.852918	1.061
18	sh600000	2014/12/9	13.56	14.16	12.46	12.65	-0.084	8.69E+08	1.18E+10	1.89E+11	2.36E+11	0.058246	105.0173	#######	#######	5.140419	2.027972	2.742357	1.019
19	sh600000	2014/12/8	13.39	14.04	13.2	13.81	0.020695	7.04E+08	9.62E+09	2.06E+11	2.58E+11	0.047171	114.6474	#######	#######	5.611792	2.213936	2.99383	1.113
20	sh600000	2014/12/5	13.42	14	12.9	13.53	0.019593	8.6E+08	1.16E+10	2.02E+11	2.52E+11	0.057606	112.3228	#######	#######	5.498011	2.169048	2.933129	1.090
21	sh600000	2014/12/4	12.58	13.3	12.37	13.27	0.054849	7.2E+08	9.31E+09	1.98E+11	2.48E+11	0.048251	110.1644	#######	#######	5.392359	2.127366	2.876764	1.069
22	sh600000	2014/12/3	12.84	13.29	12.32	12.58	-0.02253	7.31E+08	9.38E+09	1.88E+11	2.35E+11	0.048956	104.4362	#######	#######	5.111972	2.01675	2.727181	1.014
23	sh600000	2014/12/2	12.03	13.08	12.03	12.87	0.058388	5.89E+08	7.4E+09	1.92E+11	2.4E+11	0.039495	106.8437	#######	#######	5.229815	2.063241	2.790049	1.03
24	sh600000	2014/12/1	12.45	12.98	12.12	12.16	-0.01936	6.07E+08	7.59E+09	1.81E+11	2.27E+11	0.040694	100.9494	#######	#######	4.941303	1.949418	2.636131	0.980
25	sh600000	2014/11/28	11.53	12.48	11.48	12.4	0.080139	8.32E+08	9.95E+09	1.85E+11	2.31E+11	0.055766	102.9419	#######	#######	5.038829	1.987894	2.68816	0.999
26	sh600000	2014/11/27	11.48	11.72	11.28	11.48	0.014134	4.4E+08	5.06E+09	1.71E+11	2.14E+11	0.029468	95.30429	#######	#######	4.664982	1.840405	2.488717	0.925
27	sh600000	2014/11/26	11.25	11.43	11.1	11.32	0.021661	4.52E+08	5.09E+09	1.69E+11	2.11E+11	0.030307	93.97604	#######	#######	4.599966	1.814756	2.454032	0.912
28	sh600000	2014/11/25	10.81	11.09	10.75	11.08	0.020258	2.89E+08	3.16E+09	1.65E+11	2.07E+11	0.019399	91.98358	#######	#######	4.502439	1.77628	2.402002	0.893
29	sh600000	2014/11/24	10.57	10.99	10.45	10.86	0.006487	4.74E+08	5.09E+09	1.62E+11	2.03E+11	0.031767	90.15718	#######	#######	4.41304	1.74101	2.354308	0.87
30	sh600000	2014/11/21	10.58	10.82	10.46	10.79	0.020814	2.1E+08	2.23E+09	1.61E+11	2.01E+11	0.014072	89.5761	#######	#######	4.384597	1.729789	2.339134	0.870
31	sh600000	2014/11/20	10.45	10.66	10.37	10.57	0.009551	1.62E+08	1.71E+09	1.58E+11	1.97E+11	0.010883	87.74968	#######	#######	4.295197	1.694519	2.29144	0.852
32	sh600000	2014/11/19	10.43	10.54	10.39	10.47	0.001914	1.45E+08	1.52E+09	1.56E+11	1.95E+11	0.009739	86.91951	#######	#######	4.254561	1.678488	2.269762	0.844

sh600000

金融数据分析实战（背景介绍）

- **板块轮动效应**
  - 证券板块与板块之间出现轮动



国信证券的行业轮动图

金融数据分析实战（背景介绍）

小象学院 ChinaHadoop.cn

- **任务：**
- **1. 根据股指进行预测,大盘（股指）数据的EWMA模型，求λ**
- **2. 找出权重股，与真实权重股进行对比,**
- **3. 根据个股数据对个股进行聚类，形成“板块”**
- **4. 发现各股与大盘的关系,尝试挖掘板块之间的关系**
- **5. 尝试对各股进行板块聚类：**
  - 如何定义邻近性？
- **6. 尝试验证板块轮动效应**

$$R = \alpha + \beta R_M + \varepsilon$$

小象学院 ChinaHadoop.cn

联系我们：

- 新浪微博：ChinaHadoop
- 微信公号：ChinaHadoop
- 网站：http://chinahadoop.cn
- 问答社区：http://wenda.ChinaHadoop.cn