

# 基于多因子模型的信用业务 风险度量研究报告

山东大学齐鲁证券金融研究院

“基于多因子模型的信用业务风险度量”项目组

项目负责人：林路 教授

项目组成员：卞晨晔、白凯敏、任鹏飞、刘萃、李邵珉

柴海涛、董平、高强

2015年6月

## 目录

摘要	1
一、项目综述	2
1.1、项目背景	2
1.2、研究目标	2
1.3、前期探索	3
(1) 多因子模型	3
(2) GARCH模型	10
二、泡沫预警模型	15
2.1、模型理论简介	15
2.2、超指数膨胀的泡沫模型简介	18
2.3、模型的参数估计和泡沫的分析	21
2.4、泡沫模型的实证检验	23
三、风险损失度量-VaR	28
3.1、VaR模型简介及计算方法	28
3.2、模型实证分析	35
3.3、结合泡沫模型计算VaR值	42
四、总结与展望	46
附录	47
参考文献	55

## 摘要

本项目针对齐鲁证券的质押贷款等信用业务，建立有效的违约风险计量模型，对未来可能的风险损失情况进行评估和预警，对信用业务的违约损失进行定量分析，保障业务风险可控、可测、可承受。通过研究股票收益率的分布，得到股价未来的涨跌走势和幅度，进而评估质押股票的总体价值分布，得到其违约损失的分布。

在本项目中，我们主要使用了多因子模型、泡沫模型以及 VaR 等方法来对该问题进行研究。

多因子模型用来研究微观指标、宏观指标以及流动性等因素对股票收益率的影响，尤其关注的是出现大幅下跌的股票所具有的共性指标。该模型通过选取合适的指标，根据历史数据对股票收益率进行回归分析和预测。

在传统多因子模型的基础上，考虑到股票收益率之间存在的密切联系，我们又加入了自回归模型。实际数据的分析表明，加入适当的自回归项后，多因子模型的显著性得到了提高。

泡沫模型可以用来检测和识别资产价格泡沫，理性地预测泡沫破灭的临界时点，使人们在其破灭之前采取预防措施，有效地规避资产价格下跌所带来的损失。在本项目中，我们使用泡沫模型检测股票价格中存在的泡沫，并对泡沫的破灭进行预测，即预测股票价格的急剧下跌。实证分析表明，对于有些股票，泡沫模型确实可以取得较好的预警效果。

VaR 是一种得到广泛应用的金融市场风险管理工具，它简单明了、易于操作。在本项目中，我们将泡沫模型和 VaR 模型结合起来。根据泡沫模型的预警结果，分段选择历史数据，分别计算正常时期和股价大跌时期的 VaR 值。实证分析表明，将泡沫模型和 VaR 结合起来的效果更好。

## 一、项目综述

### 1.1、 研究背景

融资融券，又称证券信用交易或保证金交易，是指投资者向具有融资融券业务资格的证券公司提供担保物，借入资金买入证券(融资交易)或借入证券并卖出(融券交易)的行为，包括券商对投资者的融资、融券和金融机构对券商的融资、融券。我们狭义上说的融资融券概念指的是前者，即券商对投资者的融资和融券，在融资融券交易中，投资者不仅可以实现新的盈利途径而且在某种程度上可以规避风险。当预测证券价格上涨时，投资者可以向证券公司融入资金，买入标的证券而后获得盈利；当预测证券价格即将下跌时，投资者可以向证券公司借入证券而后卖出，之后待证券价格下跌时买入证券归还给证券公司获利，或者是通过卖空之前买入证券而对冲风险。融资融券交易的开展对我国证券市场来说是一种信用创新，对券商说是一大利好，它可以成为证券公司的主要业务之一，而且会相应的提高其佣金，取得相应的利息收入以及咨询服务等与融资融券业务相关的收入，在这期间券商的服务也会因融资融券业务的发展而增强，并且由于融资融券业务，很多新的投资者也会加入到股市中来参与其中，对推动我国证券市场的发展具有重要的意义。

随着对融资融券业务研究的逐步深入，我们意识到融资融券业务的开展给投资者提供新的投资方式的同时，也给他们带来了新的更高的风险。融资融券特殊的杠杆机制决定了其风险与一般金融工具相比具有特殊性，主要表现在风险的放大性和风险与机会的共生性。2013 年齐鲁证券的信用业务快速发展，取得了担保证券市值 239 亿，融出资金超百亿的良好业绩。与此同时，公司面临的风险也在不断累积加剧，市场竞争压力使得担保证券质押率不断提高，客户违约造成强制平仓大量发生，仅昌九生化重组失败事件就造成了三名客户违约合计 18 万的坏账损失。

### 1.2、 研究目标

针对齐鲁证券的质押贷款等信用业务潜在的损失风险，建立有效的违约风险计量模型来估计风险损失情况，对未来可能的风险损失情况进行评估和预警，保

障信用业务风险可控、可测、可承受。通过建立合适的统计模型，研究股价或股票收益率的分布得到股价未来的涨跌走势和幅度，即利用均值回复随机平稳临界点模型（简称泡沫模型）在股价大跌之前识别出泡沫破灭的临界点并进行预警，根据产生的警报密切关注股价走势，以此做出相应的信用风险控制决策。然后根据泡沫模型对股价大跌临界点的预警，结合 VaR 模型对违约损失进行定量分析，选择适当的方法对股票收益率进行估计得出股票收益率的近似分布，进而得到股票收益率的 VaR 值，得出的结果有助于在业务开展前提供决策依据。

### 1.3、 前期探索

#### (1) 多因子模型

本项目研究主体是券商对投资者的融资，即客户提供质押物充当保证金，证券公司借款给客户购买证券，客户到期偿还本息。根据《融资融券试点交易实施细则》，客户信用账户维持担保比例不得低于 130%。证券公司在不高于上述交易所规定比例的基础上，可以根据客户资信状况等因素，自行确定维持担保比例的最低标准。当投资者维持担保比例低于 130% 时，证券公司应当通知客户在不超过 2 个交易日的期限内追加担保物，且客户追加担保物后的维持担保比例不得低于 150%。客户未能按期交足担保物或者到期未偿还融资融券债务的，证券公司将采取强制平仓措施，处分客户的担保物，不足部分还可以向客户追索。维持担保比例，即客户担保物价值与其融资融券债务之间的比例，其计算公式如下所示：

$$\text{维持担保比例} = \frac{\text{现金} + \text{股价} \times \text{抵押股票数量} \times \text{折算率} + \text{融资买入金额}}{\text{融资买入金额} + \text{利息及费用总和}}$$

通俗的讲就是：维持担保比例 = (自己的资产 + 借来的资产) / 借来的资产。由维持担保比例的计算公式可以看出，通过相关模型研究客户质押股票的价格波动情况并进行一定的预测很有必要，可以根据产生的预测密切关注股价走势，以此做出相应的信用风险控制决策。在股价或股票收益率的预测模型中，多虑多种因素影响股票收益率的多因子模型最为直观。

#### 1.3.1 模型概述

多因子模型是关于资产定价的模型，但是与资本资产定价模型和单指数模型

不同，多因子模型认为证券价格并不仅仅取决于证券的风险，还取决于其他一些因素，如投资者未来预期收入、未来消费品的相对价格及未来的投资机会等。多因子模型的理论基础是 Ross 基于套利原理创立的套利定价理论(APT)和 Merton 基于均衡原理创立的跨时期资本资产定价模型(ICAPM):

(1) 资本资产定价模型 (CAPM 模型)

股票定价中起作用的是市场风险，股票的单个风险不能得到补偿。对于一个给定的资产  $i$ ，它的期望收益率和市场组合期望收益率的关系可以表示为：

$$E(r_i) = r_f + \beta_{im}[E(r_m) - r_f]$$

其中  $E(r_i)$  是资产  $i$  的期望收益率 (或普通股的资本成本率);  $r_f$  是无风险收益率，通常以短期国债的利率来近似替代;  $\beta_{im}(\beta)$  是资产  $i$  的系统性风险系数，

$$\beta_{im} = \frac{Cov(r_i, r_m)}{\sqrt{Var(r_m)}\sqrt{Var(r_i)}}; E(r_m) \text{ 是市场投资组合 } m \text{ 的期望收益率，通常用股票}$$

价格指数收益率的平均值或所有股票的平均收益率来代替;  $E(r_m) - r_f$  是市场风险溢价 (Market Risk Premium)，即市场投资组合期望收益率与无风险收益率之差。

通过对  $\beta$  进行分析，可以得出结论:在风险资产的定价中，那些只影响股票的方差而不影响该股票与股票市场组合的协方差的因素在定价中不起作用，对定价唯一起作用的是该股票的  $\beta$  系数。由于收益的方差是风险大小的量度，可以说:与市场风险不相关的单个风险，在股票的定价中不起作用，起作用的是有规律的市场风险，这是 CAPM 中最为重要的思想。

(2) 套利定价模型 (APT 模型)

由于 CAPM 理论可能存在设定误差，Ross 提出了 APT 模型。该理论认为风险可由几个因子产生，而不像 CAPM 那样基于一个风险因子，资本资产定价模型可以看做套利定价模型的特例。套利定价理论所假定的收益生成过程的计算公式的普通形式如下：

$$R_i = a_i + b_{i1}I_1 + b_{i2}I_2 + \dots + b_{ij}I_j + e_i$$

式中:  $a_i$  为当所有影响证券的各指数价值皆为 0 时证券  $i$  的平均收益率;  $I_j$  为影响

证券*i*收益水平的第*j*个指数； $b_{ij}$ 为证券*i*对第*j*个指数的敏感度； $e_i$ 为一个随机误差项，其平均值为0，均方差为 $\delta_{e_i}^2$ 。

APT模型的基本原理是根据唯一价格规则，风险性质相同的两种资产不可能产生不同的收益结果，否则市场上将出现套利行为。而套利的结果必然是使两种资产的收益率趋于一致，使市场重新达到平衡。

### 1.3.2模型构建

借鉴多因子模型的思想，我们可以用此模型研究股票收益率的各种影响因素。在选择变量时，可以从微观指标、宏观指标、流动性等方面进行选择。尤其需要关注的是，寻找那些可能出现大幅下跌的股票具有的共性指标，例如发生重组失败的公司股价所具有的共性。模型构建过程如下：

第一步，依据待估公司所属行业类型，选择影响公司股票价格的因素。包括宏观经济、行业发展和公司经营等方面，常见的影响因子如下所述：

#### (1) 公司的微观指标

公司价值决定股票价格，股票价格反映公司价值。一些可以导致公司价值改变的经济指标，必定会影响到股票的价格，从而影响到股票的收益率。这些指标包括：净利润增长、净资产增长、销售收入增长、经营现金流增长。

#### (2) 经济环境的宏观指标

国家的经济政策和公司所处行业的状况对公司的经营发展有着重要的影响。指标的选取通常采用市场指数、工业产值增长率、通货膨胀率、长短期政府债券的收益率之差等。

#### (3) 股票的流动性

从理论上说，股市存在流动性溢价现象：如果股票流动性低，其交易成本越大，投资者对持有该资产的预期收益就越高。

第二步，通过对选出的因素进行显著性检验来确定进入多因子模型的因素种类及其个数，显著性检验过程如下所示：

首先利用 F 检验对股票的对数收益率与各影响因子之间是否存在确定性线性关系进行检验；若选出的因素对股票对数收益率的总体回归效果显著，则利用 t 检验对各影响因子回归系数进行显著性检验，进而可以剔除有效但冗余的因子。

总体回归效果显著并不说明每个影响因子对股票的对数收益率都是重要的,即可能有某个影响因子对股票的对数收益率并不起作用或者能被其它的影响因子的作用所代替,因此对这种影响因子我们希望从回归方程中剔除,这样可以建立更简单有效的回归方程。

第三步,用适当的样本估计并建立多因子股价模型,对股票对数收益率进行回归和预测。

多因子模型的基本形式为

$$\bar{K}_i = a_i + b_{i1}\delta_{1t} + b_{i2}\delta_{2t} + \dots + b_{ik}\delta_{kt} + e_i$$

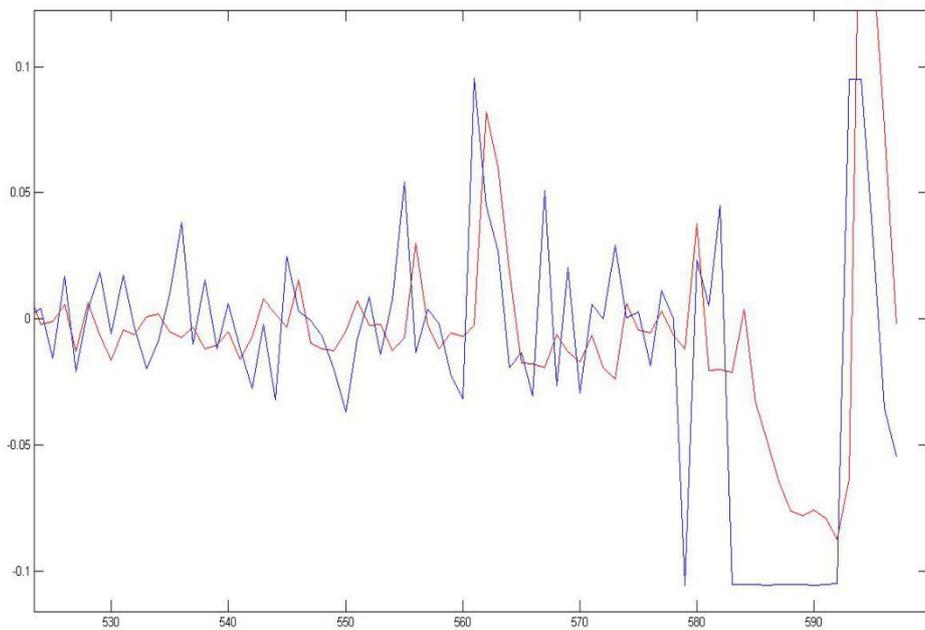
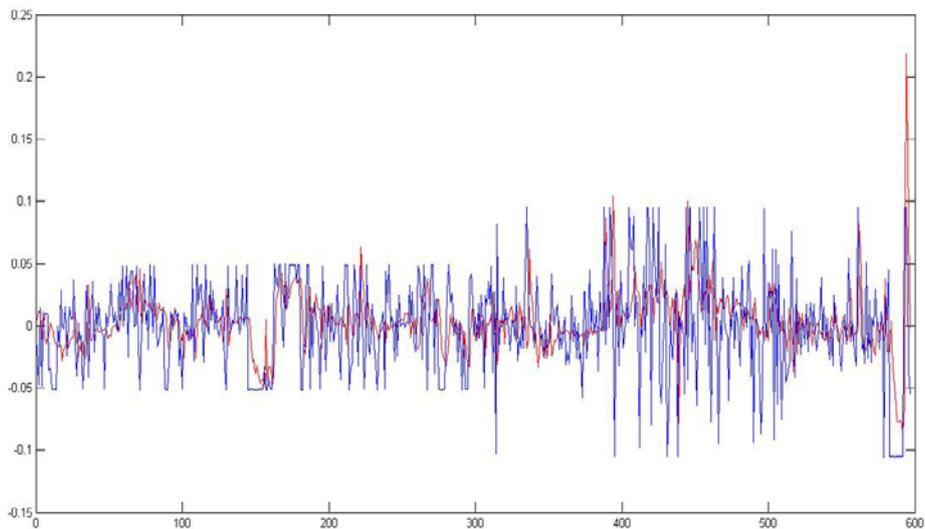
式中:  $\delta_{kt}$  是第  $k$  是个风险因素在时期  $t$  的意外变化;  $b_{ik}$  是资产  $i$  对第  $k$  是个风险因素的敏感系数。

### 1.3.3 模型实证研究

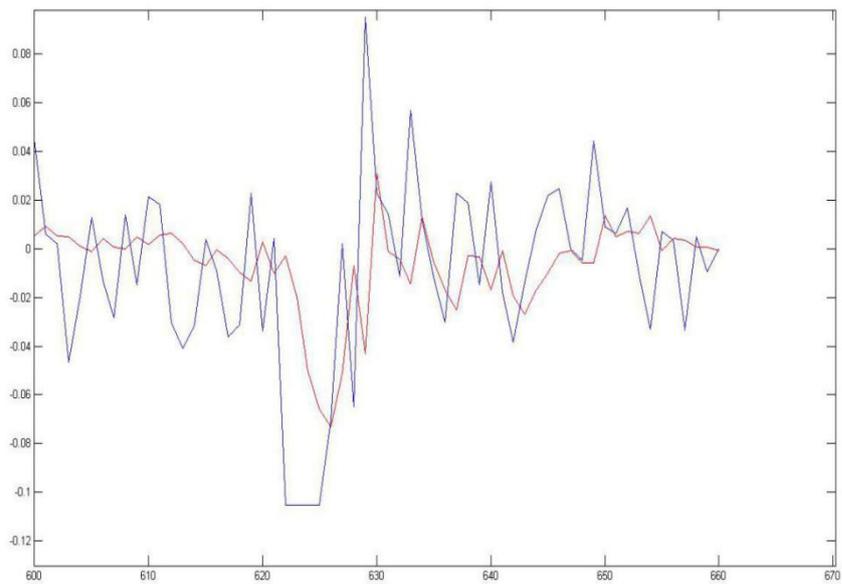
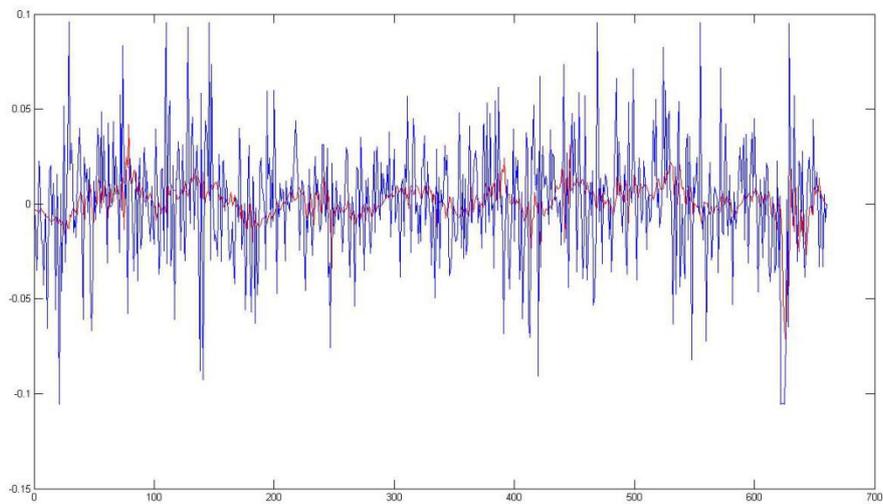
考虑到股票收益率之间存在密切的联系,在多因子模型中加入了自回归模型 (AR),即把前一段时间的股票收益率作为影响因子加入到多因子模型中。经实证检验加入一阶或二阶自回归的模型的显著性有所提高,但加入高阶自回归会导致模型显著性降低,于是只考虑把股票收益率影响最密切的前两天的收益率作为影响因子加入到多因子模型中,以得到股价未来的涨跌走势和幅度

通过一系列显著性检验并加入低阶自回归模型建立新的多因子模型,由于一些影响因子每天更新导致无法及时获取,使用已知的数据代替,这导致了在对股票收益率进行回归预测时出现了延迟现象。

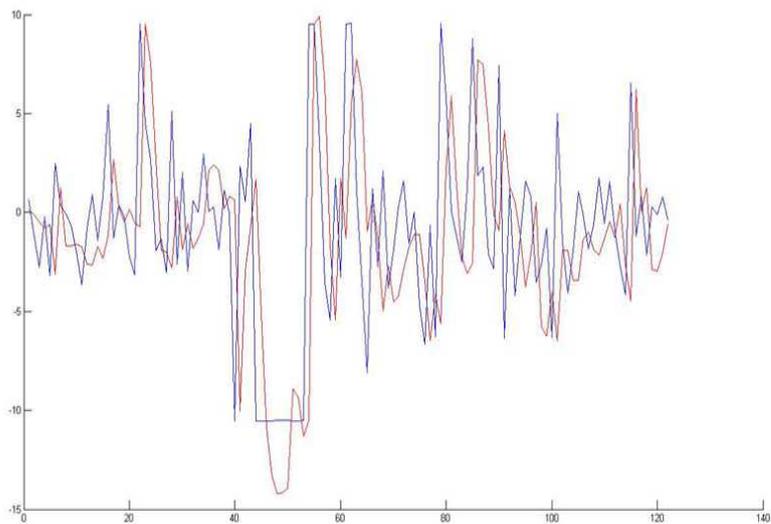
昌九生化



### 重庆啤酒



```
-----回归方程显著性检验 (H0:  $\beta_1 = \beta_2 = \dots = \beta_k = 0$ ) -----  
经过计算: 接受H0, 原假设成立, 无线性关系。 >> check  
第1次计算结果:  
Warning: Matrix is close to singular or badly scaled. Results may be inaccurate.  
RCOND = 2.811152e-35.  
> In check at 12  
 $\beta_0 = 239935402556105.0300$   
 $\beta_1 = -0.0119$   
 $\beta_2 = 72.7952$   
 $\beta_3 = -239935402556104.2800$   
 $\beta_4 = 0.0470$   
 $\beta_5 = -0.0946$   
 $\beta_6 = 0.0097$   
 $\beta_7 = 0.0137$   
 $\beta_8 = -0.0212$   
|  
>>>>>请输入您要求的显著性水平 ( $0 < \alpha < 1$ )  $\alpha = 0.02$   
-----回归方程显著性检验 (H0:  $\beta_1 = \beta_2 = \dots = \beta_k = 0$ ) -----  
经过计算: 接受H0, 原假设成立, 无线性关系。 >> |
```



蓝色代表真实值, 红色代表预测值。可以看出预测值比真实值滞后大约一天, 即对明天收益率的预测大约等于今天的收益率。

## (2) GARCH 模型

传统计量经济学模型在古典假设中要求扰动项是同方差的,但是实践中这一点通常不易满足,扰动项往往是异方差的。特别是在金融时间序列中,常常会发生某一特征的值成群出现的现象。如对股票收益率序列建模,波动的数据类型使得随机扰动项往往在较大幅度波动后面伴随着较大幅度的波动,在较小幅度波动的后面紧接着较小幅度的波动。收益率序列一般不具备独立正态分布,常表现为尖峰厚尾性和负偏性。收益率序列的波动又常常具有非对称性,即股市上坏消息引起的波动要比好消息引起的波动大,即“杠杆效应”。

2003年10月8日,瑞典皇家科学院宣布将当年诺贝尔经济学奖授予计量经济学家、美国纽约大学的 Robert Engle 与加州大学圣迭戈分校的 Clive Granger,以表彰他们在分析金融时间序列数据时所发明的统计方法。Engle 教授于 1982 年提出了自回归条件异方差模型 (ARCH),对描述金融时间序列波动的多变性提供了一种简洁有效的新方法。随着研究的深入,ARCH 模型的一些不足逐渐显现出来,如不能体现收益率厚尾性,为达到最佳拟合效果需要很大的阶数等,为此 Engle 的学生 Bullerslev 于 1986 年将这个模型推广成了广义自回归条件异方差模型 (GARCH)。

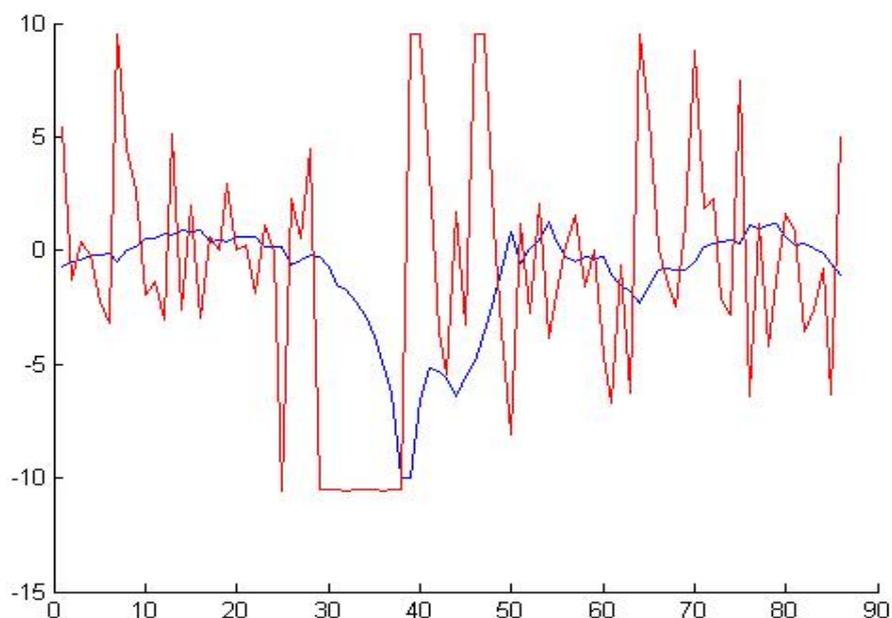
一般的 GARCH (p,q) 模型可以表示为:

$$r_t = c_0 + \sum_{i=1}^N \phi_i r_{t-i} + \sum_{j=1}^M \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

$$\varepsilon_t = \mu_t \sigma_{t|t-1}, \mu_t \sim i.i.d. (0, 1), \sigma_{t|t-1}^2 = k + \sum_{i=1}^q \alpha_i \sigma_{t-i|t-i-1}^2 + \sum_{j=1}^p \beta_j \varepsilon_{t-j}^2 \quad (2)$$

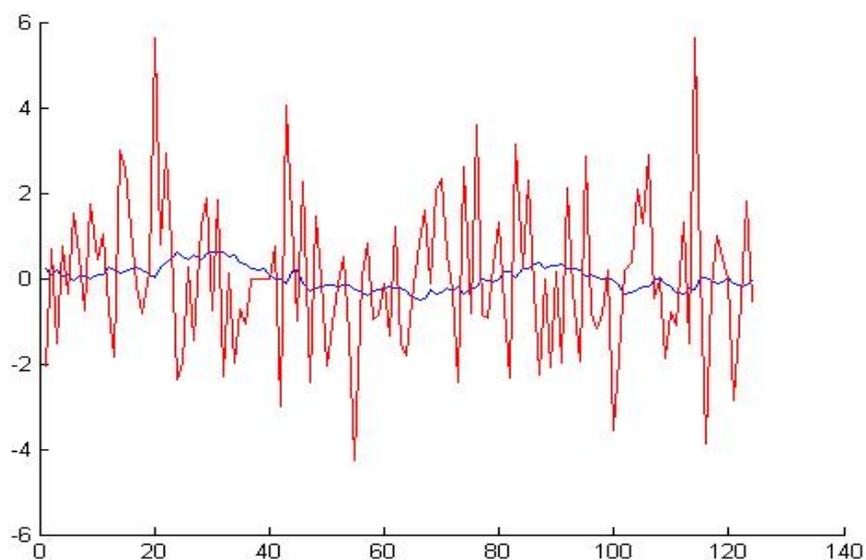
其中  $r_t$  为收益率,  $\sigma_{t|t-1}^2$  为条件方差,  $\mu_t$  为独立同分布的随机变量,(1) 式称为条件均值方程;(2) 式称为条件方差方程,说明时间序列条件方差的变化特征。为了适应收益率序列经验分布的尖峰厚尾特征,也可假设收益率服从其他分布,如 Bollerslev (1987)假设收益率服从广义 t-分布, Nelson(1991)提出的 EGARCH 模型采用了 GED 分布等。对金融时间序列, GARCH 模型的阶数一般选取 p=1 及 q=1。实际验证中,选取了 A 股的一些股票,用 GARCH (1,1) 模型进行拟合以及预测,下面给出一些结果:

下图为昌九生化的结果



红线代表实际对数收益率（百分比的形式），蓝线代表预测的对数收益率，可以看出明显的滞后，在连续跌停的十几天里，蓝线逐渐降到最低，而当预测到跌停时，实际收益率已经恢复到震荡阶段，在其后的十几天预测值才逐渐上升最终恢复震荡。

下图为平安银行的结果



可以看出，在没有长时间大跌或大涨的情况下，GARCH 模型的预测为在 0 附近波动的曲线，这是由于预测是对收益率均值的预测，而正常情况下收益率均值为零。

另外，GARCH 模型还存在如下不足：

①GARCH 模型不能解释股票收益和收益变化波动之间出现的负相关现象。GARCH(p,q)模型假定条件方差是滞后残差平方的函数,因此,残差的符号不影响波动,即条件方差对正的价格变化和负的价格变化的反应是对称的。然而在经验研究中发现,当利空消息出现时,即预期股票收益会下降时,波动趋向于增大;当利好消息出现时,即预期股票收益会上升时,波动趋向于减小。GARCH(p,q)模型不能解释这种非对称现象。

②GARCH(p,q)模型为了保证非负,假定(2)式中所有系数均大于零。这些约束隐含着的任何滞后项增大都会增加因而排除了  $r_t$  的随机波动行为,这使得在估计 GARCH 模型时可能出现震荡现象。

下面对未来收益率小于某一值的概率进行估计。

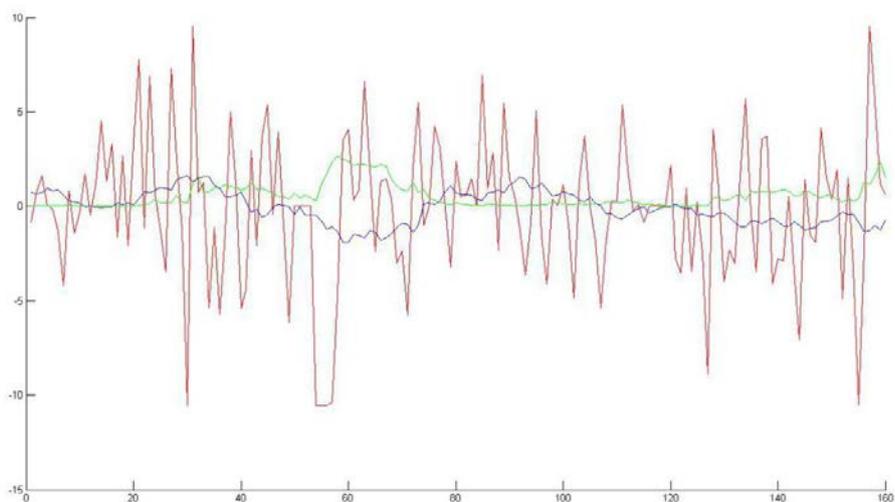
假设由 GARCH(1,1)模型得到的对第 t+1 天收益率的预测近似服从正态分布,期望为第 t+1 天的真实收益率,方差为条件方差方程估计出的第 t+1 天的条件方差:

$$\hat{r}_{t+1} \sim N(r_{t+1}, \hat{\sigma}_{t+1|t})$$

于是,第 t+1 天收益率的小于 a 的概率可用下式近似表示:

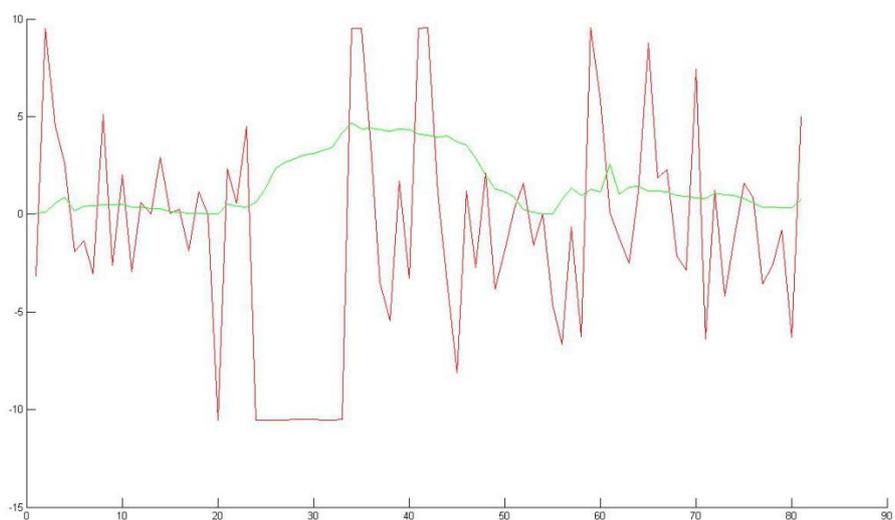
$$P\{r_{t+1} < a\} = \Phi\left\{\frac{a - \hat{r}_{t+1}}{\hat{\sigma}_{t+1}}\right\},$$

从 A 股中筛选出了一些连续三天的收益率小于负百分之七的股票（如 ST 凤凰、杭钢股份、兴业矿业、昌九生化等）进行实证检验，下面为兴业矿业的结果：



其中红线代表真实对数收益率，（单位为百分数），蓝线为 **GARCH (1,1)** 模型预测的对数收益率，绿线代表预测第二天收益率小于负百分之七的概率乘以十，（乘以十是为了便于比较，绿线的值为五则表示有 0.5 的概率）。在第五十到六十个数据之间有连续四天跌停，在第一天跌停时，预测当天收益率小于负百分之七的概率不到 0.1，连续两天跌停时，预测第三天收益率小于负百分之七的概率仅为 0.25，连续三天跌停时，预测第四天下跌较大的概率仅为 0.32。

下面是昌九生化的结果：



红线为真实对数收益率（乘以一百），绿线为概率（乘以十）。可以看出，在连续跌停的十几天里，**GARCH** 模型预测下一天下跌超过百分之七的概率仅为 0.3 左右，在下跌的最后一天，模型预测下一天跌幅较大的概率为 0.5，而真实情况

却是涨停，在其后十几天收益率处于震荡阶段，而预测跌幅较大的概率却在 0.4 左右，比跌停期间预测的还要大。由此可见，GARCH 模型在实际应用中还有很多问题有待解决。

## 二、泡沫预警模型

金融资产作为实体经济延伸至虚拟经济的价值载体，基于的是对未来回报的信用创造。因此，与实体经济类似，其价格的变化可归结为供需双方的力量对比。然而，金融资产价格的波动更加频繁，并且还会呈现持续的加速上涨并在短时间内极速滑落的大幅度波动特征——表现为泡沫的积累与破灭现象。资产价格泡沫破灭可能导致数以亿计市值的大量蒸发，极大破坏金融市场秩序的同时严重影响宏观经济的配置效率。在泡沫破灭之前及时地检测和识别泡沫，不仅能够帮助政策制定者提前采取预防措施以避免对社会经济的严重冲击，同时也能帮助投资者有效地规避下跌的损失。因此资产泡沫的研究具有重要的理论意义和实用价值。股票作为金融资产的典型代表，对其泡沫的建模、检测和识别历来都是一个热门的研究。

### 2.1、模型理论简介

国际上研究股市泡沫最早的模型可以追溯至 **Blanchard** 等人提出的理性预期泡沫模型。该模型认为，为了对泡沫在未来破裂进行风险补偿，市场中的理性投资者会要求一个对应泡沫成分的超额收益，并且泡沫在理性预期条件下以这个超额收益的速度增长。他们提出用游程（run）、厚尾（fat tail）来检验泡沫的存在性。另一方面，从基本面的数据出发，**Shiller** 和 **LeRoy** 首次提出用方差界检验的思想（variance bounds test）来衡量资产价格与股息之间的方差变异关系，以检测泡沫的存在性。为了改善方差界检验方法的局限性，**West** 还提出了一套两步检测法。随后协整检验被用来判断泡沫的存在性。其基本思想是检测价格与股息序列之间是否具有协整关系。然而 **Evans** 却指出当考查时间段内出现泡沫的反复积累和破灭时，基于单位根检验和协整检验的方法将会失效，因为它们都无法应对这类泡沫过程中的非线性特征。为了讨论这些非线性特征，**Evans** 继而在理性预期泡沫的基础上提出了周期性破灭型投机泡沫模型（periodically collapsing bubbles）。**Van Norden** 和 **Schaller** 以 **Evans** 的模型为基础提出了基于指数超额收益率和相对泡沫度的动态关系区制转换检测方法。另外，**Froot** 和 **Obstfeld** 创立了内生性泡沫模型，通过检验价格与基本面变量存在指数依赖关系来确认泡沫的

存在性。其后，McQueen 和 Thorley 在对理性预期泡沫模型大量分析后创立了持续期依赖检测法（duration independent），其核心思想是泡沫中的正超额收益的游程终止概率将随着游程的长度而递减。目前，国内已有不少学者借助这些经典的股市投机性泡沫模型和检测方法对中国股市的泡沫进行实证。例如，周爱民用动态自回归法检验泡沫。潘国陵采用方差界方法检验泡沫，全登华和贺学会等人采用持续期依赖法来实证检验中国股市泡沫，崔畅和孟庆斌等人使用了门限自回归模型来识别周期破灭型投机泡沫，赵鹏等人则尝试了通过马尔科夫序贯关系的区制转换方法来检测我国股市的周期破灭型投机泡沫。此外，国内的一些学者的还将泡沫检测思路集中于测算合理的市盈率和股票内在价值，有代表性的包括吴世农等人的 CAPM 法，刘焯松的有限期限 F-O 模型，邹辉文的股票内在价值 Kalman 信息滤波，以及徐爱农的企业价值——剩余收益的 Ohlson 模型。

然而，在对金融泡沫模型进行全面考察后，Gurkaynak 却得出了令人吃惊的结论：“传统的泡沫模型并不能给出令人信服的泡沫检测和识别的结果，因为每一篇宣称存在泡沫的文章或模型，总可以找到一个对应的不要求泡沫存在的解释”。究其原因，主要是：1)人们实质上无法正确的估测基本面的大小；2)传统模型假设的泡沫也是指数膨胀的，它始终无法和同样是指数增长的基本面对应的高速价格增长显著的区分开来。可以说，除了持续期依赖检测模型，以上提到的传统模型均无法完全克服 Gurkaynak 指出的问题。然而，持续期依赖检测也有自身的缺陷：它要求用月收益率或者周收益率来检测（消除噪声），因此对三年以及更短时间产生的泡沫无效；另一方面依赖法对于泡沫的破灭可能性没有明确的估计方法。

正如 Gurkaynak 的评论，泡沫不是偏离基本面的更高速的指数膨胀，而是本质上区别于指数增长的更快膨胀——超指数膨胀（super-exponential growth）。与允许无限膨胀下去的指数增长形式不同，这种膨胀的过程内在的蕴含着一个临界时点（critical time），泡沫在临界时点处趋于发散。预示市场对泡沫的追逐需要无限的资金，因此不可能持续下去。临界时点只是理论上的泡沫终结时刻。实际的投机泡沫在接近临界时点积累过程中会越来越不稳定，最终在临界时点之前的某一时刻就破裂掉。目前认为，泡沫超指数膨胀可以归因于金融市场中的正反馈效应。典型的正反馈可来自于理性因素，如利用噪音交易者的做市交易、庞氏骗

局；也可能来自行为因素，如不确定性的理性模仿或者是所谓的“word-of-mouth”社会效仿机制导致的羊群效应（herding）。

Sornette 等人首次提出了一个描述泡沫超指数膨胀的股价模型，称为 S-B 模型（super-exponential bubble）。该模型将 Black-Scholes 股价模型推广，引入泡沫积累时对价格的正反馈，并尝试通过检验表征正反馈的指数  $m$  是否显著大于 1 来识别泡沫，模型中临界时间服从逆高斯分布。另一方面 Sornette 的团队也给出过另一种表述形式的超指数增长投机泡沫模型（JLS 模型）。该模型以 Blachard 的理性预期泡沫模型为基础，价格的超指数增长来自于正反馈效应下泡沫破裂风险概率的超速增加。同时 JLS 模型还内在要求泡沫在逼近临界时点的过程中出现加速的周期振荡。最终，泡沫的检验可以转化为识别价格中 LPPS 模式（log-periodical power signature）的出现。随后的大量实证研究表明，JLS 模型不仅可以识别出超指数数膨胀泡沫的存在性，同时能够对泡沫的破灭时点给出一定的预先估计。JLS 模型已被用于对中国股市近期的两次泡沫（2005 年中期至 2007 年后期，2009 年中期）进行诊断，确认了其中的 LPPS 模式，并发现泡沫的破灭时点与模型预估的时点能较好的匹配。需要指出的是，JLS 模型客观上要求市场在泡沫积累阶段涌现出 LPPS 模式，这对模型限制较强，且 LPPS 的产生机理目前尚不十分清楚。

针对以上泡沫模型的不足，Sornette 等人提出了一类新颖的超指数膨胀泡沫模型——均值回复的平稳随机临界时点模型，并做了实证。同样的，在市场投机泡沫积累的过程中，股价会在正反馈机制的作用下呈现超指数增长，其中蕴含着临界时点。然而由于市场中噪音交易者的存在，以及市场中理性人不断的进出和信念更新，临界时点用随机过程来刻画。一方面，市场的理性投资人能够对随机临界时点产生一致的无偏估计——即理性预期到“潜在的临界时点”。另一方面，理性预期的存在也恰好使得泡沫市场中的随机临界时点成为一个围绕潜在临界时点来回振荡的平稳过程。值得一提的是，理性预期的作用在此模型中被弱化到了预期潜在临界时点而不是价格。该模型可以很好的避开 S-B 模型中的参数估计问题。对泡沫存在性的检验转化为检测“非线性非平稳的股价序列是否蕴含着一个均值回复的平稳的临界时间点序列”。同时，估计得到的潜在临界时点大小还可以用于预警泡沫的破灭。

我国股市交易监管制度虽然不断健全，但是中小投资者仍然普遍的存在投机心理，机构投资者易于价格操纵，再加上缺乏一定的套利机制，市场中正反馈效应依然严重。超指数膨胀模型能够有效的捕捉正反馈导致的泡沫特征。本节尝试引入平稳随机临界时点的超指数膨胀泡沫模型来对我国股市中的泡沫进行建模、检验和识别。

## 2.2、超指数膨胀的泡沫模型简介

按照超指数膨胀模型的基本逻辑，泡沫的本质是金融资产价格在中正反馈机制下产生的暂态的快于指数增长的情形，泡沫如果不破灭则将会在有限的临界时点处发散。然而由于在有限时间内没有无限的资金支持泡沫的膨胀，因而泡沫必将在临界时点之前破灭。临界时点可以视为泡沫的理论终结点。任何泡沫不可能在跨过临界时点后生存。但临界时点决不是泡沫破灭的真正时刻，破灭将发生在此前的某个时刻。另外，一切外在的冲击仅仅是泡沫破灭的导火线，而泡沫在其向临界时点运行的过程中变得越来越不稳定才是其破灭的内在原因。为了理解本节的均值回复平稳随机临界点模型，我们先介绍两个经典的超指数膨胀模型：SB 模型和 JLS 模型。

### 2.2.1 S-B (super-exponential bubble) 模型

该模型假设，在泡沫阶段股票的价格将满足如下的随机微分方程：

$$dp = \mu(p)p^m dt + \sigma(p)p^m dW - \kappa dj, m > 1$$

其中， $W$  是一个维纳过程，描述价格的随机扩散， $j$  是一个跳跃过程代表着泡沫的破灭价格产生暴跌， $\kappa$  衡量暴跌的幅度。另外变量  $\mu(p)$  是一个待定的漂移变量，由无套利条件导出，它使得价格过程满足中性概率测度下的鞅条件： $\tilde{E}_t(dp) = 0$ 。扩散系数  $\sigma(p)$  是对波动率的测度。指数  $m$  表征着正反馈的强度。实际上，在泡沫没有破灭 ( $j = 0$ ) 时 SB 模型可视为是对几何布朗运动的推广，因为当  $\mu, \sigma$  为常数且  $m = 1$  时，模型退化为标准的 Black-Scholes 股价方程。方

程 (1) 的含义可以从没有扩散项的方程看出来。此时，泡沫破灭之前的方程表述为：

$$dp = \mu p^m dt, m > 1$$

易知方程 (2) 的解为：

$$p = K(t_c - t)^{-\beta}$$

其中

$$\beta = \frac{1}{m-1}, K = \left(\frac{\beta}{\mu}\right)^\beta, t_c = \frac{p_0^{-(m-1)}}{\mu(m-1)}.$$

(3) 式表明当存在正反馈时，具有泡沫的价格会在一个有限的时间节点  $t_c$  处发散。 $t_c$  就是临界时点。加入扩散项后，临界时点成为一个随机变量。可以证明，当扩散系数和漂移系数满足一定的条件时，临界时点将服从逆高斯分布。

### 2.2.2 JLS模型

在 JLS 模型中，股票价格在泡沫阶段可以用如下的随机微分方程刻画：

$$dp = \mu(p)pdt + \sigma(p)p dW - \kappa dj, m > 1$$

上面方程中的各项与式 (1) 相同。在套利条件下，价格在中性概率测度之下成为一个鞅。因此，

$$E_t(\tilde{d}p) = E_t[\mu(t)dt - \kappa dj] = 0 \Rightarrow \mu(t)\kappa h(t)$$

其中  $\tilde{E}_t(dj)/dt$  代表的是泡沫破灭的风险率，它是指泡沫在  $t$  时刻没有破灭，但是在未来时刻发生破灭的概率密度。根据理性预期泡沫的思想，在泡沫积累的过程中，价格被推高，破灭的风险率也越来越大，理性投资者要求对风险进行补偿，因此继续推高价格直到泡沫破灭。在 JLS 模型中，泡沫的膨胀是理性投资人对破灭风险的客观要求，而破灭的发生来自于市场中噪声交易者的一致性抛盘行为。噪声交易者发生一致行为的概率实际上就是风险率  $h(t)$ ，JLS 模型假设来自噪声交易者内部相互效仿的正反馈效应将使得它满足：

$$h(t) \propto (t_c - t)^{-\beta}, 0 < \beta < \infty$$

因此由 (5) 式, 对数价格的期望增长将满足:

$$E_{t_0} \ln p_t = \ln p_0 + \int_0^t \mu(t) dt = \ln p_0 + \kappa \int_0^t h(t) dt = A - B(t_c - t)^{1-\beta}$$

其中  $A, B$  是两个待定的参数。由 (7) 式可以看出, 在 JLS 模型中, 临界时点的期望价格不会发散, 而是加速到达有限值  $A$ 。真正会发散的是产生泡沫破裂的概率密度。因此, JLS 模型与 SB 模型有很大的不同。SB 模型的正反馈直接作用在泡沫价格上, 而 JLS 模型的正反馈作用在价格的变化率上——漂移项  $\mu$ , 同时也暗示着收益率之间的一个趋势关系, 表明泡沫的存在使市场存在一定的“惯性”。实际上, 由于巨大的背景噪音, 直接在收益率层面不可能检测这种惯性, 只有到价格层面才能观察到它。为了取得更好的实证结果, JLS 模型实际上还要求风向率具有在临界时点之前具有加速周期振荡的特征。

### 2.2.3 均值回复平稳随机临界点模型

均值回复随机平稳临界点模型是超指数膨胀模型中的一类最新模型, 它的经济学含义很明确: 假定市场中理性的交易者是序贯的进入到市场中, 并且他们都一致的意识到了泡沫的存在, 即泡沫在理论上会在随机的临界时点内终结。然而他们谁都不能准确的估计到泡沫的临界时点 (更不能估计泡沫破灭的时刻), 只能理性的预期泡沫破灭临界时点的一个无偏估计。这个无偏估计被称之为“潜在临界时点”。同时, 理性预期的存在客观上使得随机的临界时点成为一个以潜在临界时点为无条件期望的“均值回复平稳过程”, 模型中采用 Ornstein-Uhlenbeck (O-U) 过程来描述。每个理性投资者根据他进入市场时的“风险收益-成本”最优化策略来决定自己的退出市场的时间  $t^{ex}$ , 由于他们每个人面对的临界时点不同, 导致不同的退出时间, 这种非同步性客观上支持了泡沫的延续, 这与 Abreu 和 Brunnermeier 提出的“同步性问题”(synchronization problem) 的思想是一致的, 但是当时他们处理的是指数型增长的泡沫。

此模型假设泡沫破灭之前, 股票价格可以用如下的随机微分方程描述:

$$dp = \mu(p)p^m(1 + \delta(p, t))dt + \sigma(p)p^m dW$$

其中,

$$\delta(p, t) = \alpha \tilde{t}_c(t) + \frac{m\sigma^2}{2\mu} [p(t)]^{m-1}$$

描述了市场在泡沫阶段存在的一个特殊的控制因子。 $\tilde{t}_c(t)$ 代表随机的临界时点，根据模型的假设，它是一个 O-U 过程，满足均值回复平稳的性质：

$$d\tilde{t}_c = \alpha \tilde{t}_c dt + \sigma_{tc} dW$$

可以证明，结合（9）式，（8）式具有如下的解析解：

$$p(t) = K(\tilde{T}_c - t)^{-\beta}$$

其中，

$$\beta = \frac{1}{m-1}, K = \left(\frac{\beta}{\mu}\right)^\beta, T_c = \frac{p_0^{-(m-1)}}{\mu(m-1)}, \tilde{T}_c = T_c + \tilde{t}_c.$$

（10）式与（3）式的区别在于临界时点不再是一个常数，而是成为了一个均值回复的平稳随机过程。值得注意的是在式（8）中，当控制因子 $\delta(p, t) = 0$ 且正反馈指数 $m = 1$ 时，模型将退化成为几何布朗运动。

依据此模型，Lin 和 Sornette 提出泡沫的检验和识别可以转化为检测：1) 随机临界时间序列是否能够拒绝非平稳的单位根过程假设；2) 潜在临界时点 $T_c$ 与最后观测时点的距离是否足够小；3) 反映正反馈效应的指数  $m$  是否显著的大于 1。

## 2.3、模型的参数估计和泡沫的检验

### 2.3.1 模型的参数估计方法

平稳随机均值模型中虽然有三个待估参数： $K, \beta, T_c$ 。然而只要知道前两个参数，就可以通过（10）式算出随机临界时间序列，从而进一步估计出潜在临界时间 $T_c$ 。假设观测股价序列的长度为 $N$ ，最后一个数据对应的时间为 $t_N$ ，则有序列

$$\tilde{T}_c(t) = \exp\left[\frac{\ln \kappa - \ln p}{\beta}\right] + t, t = t_N - N + 1, \dots, t_N$$

对  $\ln K$  和  $\beta$  的估计是通过如下的优化算法得到的：由于这个泡沫模型的最重要特征是非线性非平稳的股价序列蕴含着无时间趋势的均值回复的平稳临界时间序列，因此首先必须保证无时间趋势，其次再对其进行无截距的单位根检验确保是平稳过程。具体做法是在一定的搜索空间中通过禁忌搜索选择一组最优的备选参数偶  $(\ln K^*, \beta^*)$  即是估计出的模型参数。

### 2.3.2 泡沫的检测和警报的生成

由模型可知，通过股价数据估计出参数  $\beta$  可以间接的求得市场中正反馈效应的强度  $m$ ，数值越大代表泡沫增长速度越快；另一方面潜在临界时间与最后一个数据观测时点的距离  $S_c = T_c - t_N$  能够揭示出泡沫积累的程度。我们称  $S_c$  为“临界时距”。它的值越小说明此时泡沫离市场一致认可的临界时间越接近，泡沫的不稳定程度就越高。但是泡沫的临界时点代表的仅仅是泡沫在没有破灭的情形下产生发散的时点，这是一种假想情形。实际上市场并没有足够的资金支持这种超速增长，所以真正的情况往往是，离临界时间较远时，就会有足够数量的理性交易者退出市场，造成泡沫的提前破灭。Lin 和 Sornette 在对美国的标普 500 指数、纳斯达克指数和香港恒生指数近 30 年的数据分析后建议，可以通过  $250 \leq S_c < 500$  和  $S_c < 250$ （小于一年）来建立泡沫不同程度的预警警报。综上，可以选取如下的准则来检测识别泡沫和预警：

i) 检验由最优参数偶  $(\ln K^*, \beta^*)$  产生的临界时间序列  $\tilde{t}_c$  是否能够拒绝单位根假设。

ii) 由  $\beta^*$  计算正反馈指数  $m$  的大小， $m$  越大可以认为识别到泡沫强度也越大。

iii) 检验由  $(\ln K^*, \beta^*)$  得到的  $S_c$  是否小于 500 天。如果条件满足，可以认为在此时间段内检验到了超指数膨胀的股市泡沫，并产生泡沫破灭的警报。如果时间序列  $\tilde{t}_c$  在一定的显著性水平之下并不能拒绝单位根假设，那么即使  $S_c$  很小，

也不能认为出现了股市泡沫，此时得到的潜在临界时间 $T_c$ 实际上是伪潜在临界时间。

## 2.4、泡沫模型的实证分析

在完成参数估计之后可以得到潜在临界时间序列，潜在临界时间与最后一个数据观测时点的距离能够揭示出泡沫积累的程度，我们称为“临界时距”。它的值越小说明此时泡沫离市场一致认为的临界时间越接近，泡沫的不稳定程度就越高。

### 2.4.1 数据的选取

我们选取历史上曾表现出泡沫积累、加剧，并最终破灭的若干家公司的数据进行检验，例如昌九生化、酒鬼酒、重庆啤酒等公司。

昌九生化(2013年11月7日)：13天跌52.2%，重组失败（开始传言和赣州稀土重组，结果赣州稀土和另外一家公司重组，受到消息冲击大跌）。

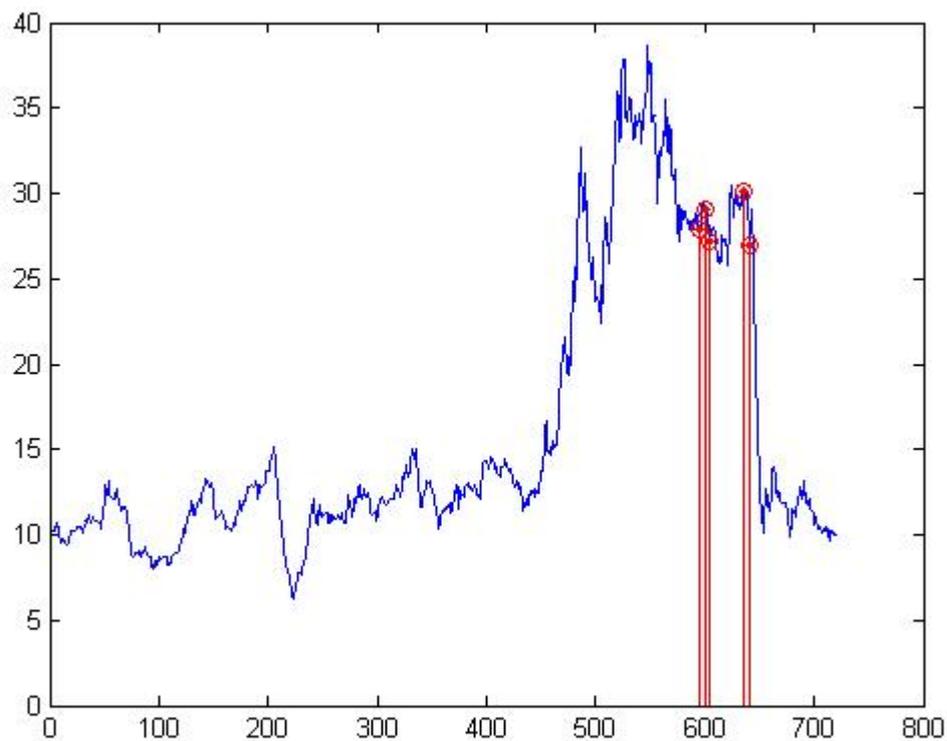
酒鬼酒(2012年11月27日)：3天跌20% 负面新闻，塑化剂超标，公司停产整顿。

重庆啤酒(2011年12月12日)：重啤2011年12月7日公布乙肝疫苗揭盲进展，分析认为数据表明乙肝疫苗无效果，这导致重啤复牌后连续9跌停，重仓的大成基金浮亏巨大。

## 2.4.2 检验结果及分析

### 1) 昌九生化

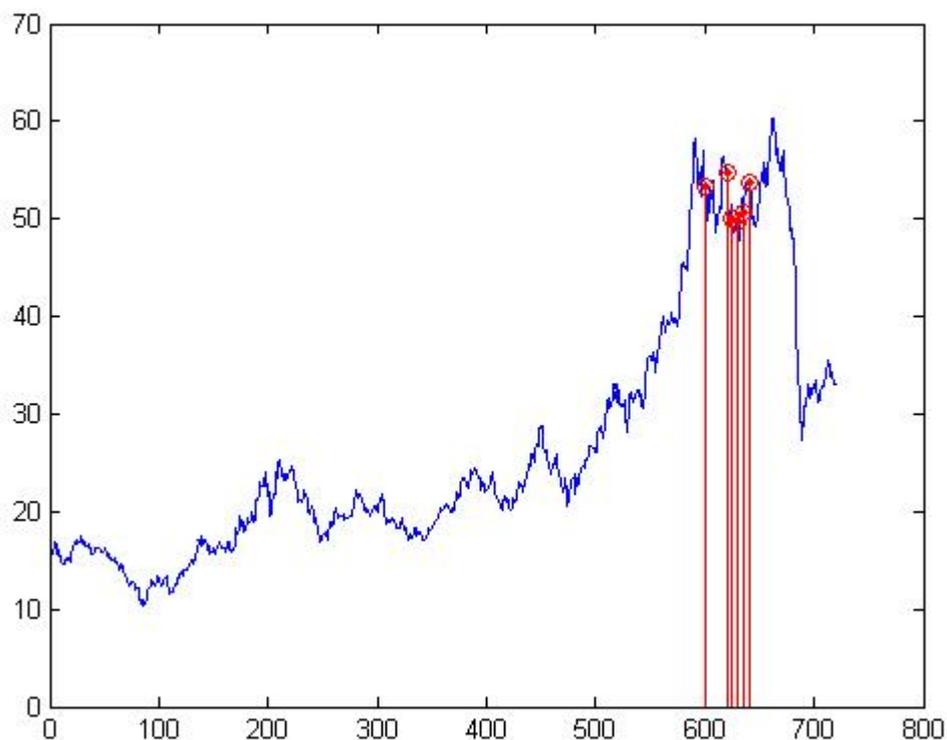
图 1 昌九生化泡沫预警图



上图是昌九生化的泡沫预警图，从图中可以看到，在 600-650 天之间生成了较多泡沫预警警报，根据模型理论，股价在此时存在泡沫，并且即将破灭，随之带来股价的大跌。这与随后股价的走势相吻合，股价在之后的确发生了急剧的下跌，泡沫预警的效果比较好。

## 2)酒鬼酒

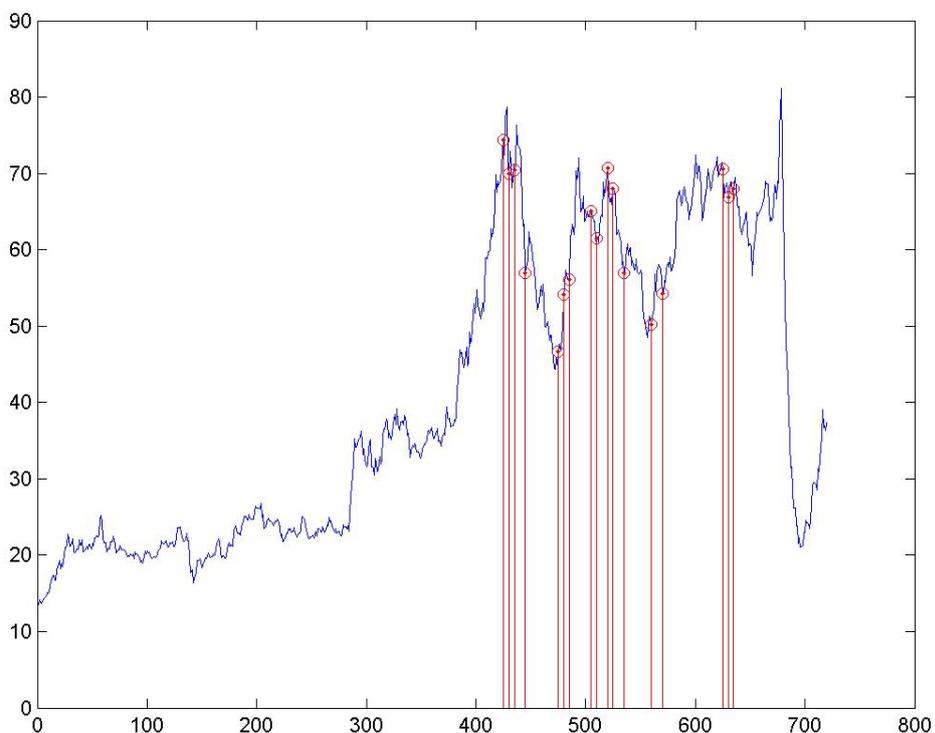
图 2 酒鬼酒泡沫预警图



从上图来看，泡沫模型产生的警报密集分布在在 610-640 之间，说明此时泡沫积累程度较为严重，可能发生股价急跌。但此时距离真实股价大跌还有一定的距离，泡沫的预警大概提前了 20 天。

### 3)重庆啤酒

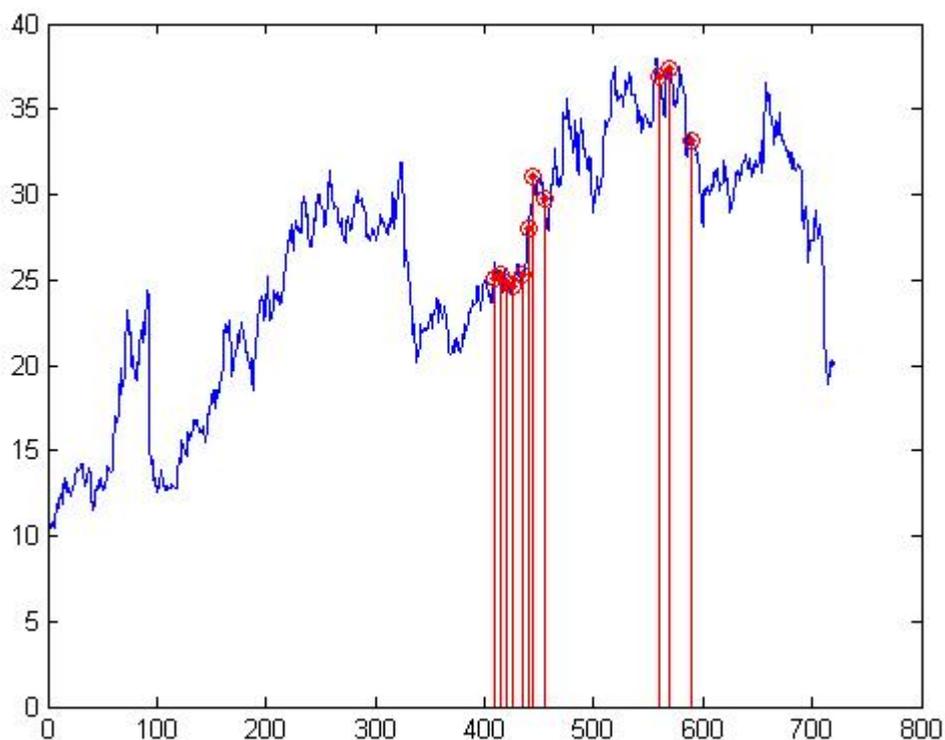
图 3 重庆啤酒泡沫预警图



重庆啤酒的泡沫预警图在多个位置产生了泡沫预警，其中前两个预警密集的区域虽有均有一定程度的股价下跌，表明预警有一定的效果；最后一段密集预警的时期虽然在股价急跌之前，但是提前的时间过早，距离真正的大跌大概 25 天，此时预警效果不够理想。

#### 4) 南岭民爆

图四 南岭民爆泡沫预警图



南岭民爆的泡沫预警图，图中预警点的位置第一段密集预警区域并未在股价急跌之前，其原因在于前期大概 100 天的时候，股价有一个急速的下降。但是这一下降并非由股价泡沫引起，而是由于该公司发生了拆股或者配股等可能导致股价稀释的财务计划。这种情形可能会影响到泡沫模型预警的准确性。

### 三、风险损失度量-VaR

如何有效的防范和控制风险已经日渐成为各个金融类公司所关注的主要问题,无论是股票市场上股票价格的突然下跌还是债券市场上债券的违约行为都会对其持有者造成资产的损失。因此,如何对可能发生的损失进行预警以及度量可能发生的损失已经显得十分的必要。在前面部分,我们已经提出了对损失进行预警的泡沫模型,在股票价格开始下跌之前产生警报,从而可以使股票持有者能够提前做出反应。而接下来的部分,我们将对可能发生的损失进行度量,以便使股票持有者能量化可能发生的损失。风险度量的方法经过了一系列的演变,到 20 世纪 90 年代,国外发展流行起来一种金融市场管理风险的工具---VaR。VaR 提出以后已经在金融市场中得到了广泛的应用。在我们的项目中,我们也选择了 VaR 模型来对风险损失进行度量。在应用 VaR 时我们选择非参的方法对股票收益率进行核密度估计,从而得出股票收益率服从的一个近似分布,进而通过求分位数的方法来得到股票收益率的 VaR 值。

#### 3.1、VaR 模型简介及计算方法

VaR 是在 20 世纪 90 年代由 JP 摩根公司提出的一种金融风险度量方法。VaR 是一种简单易懂且易于操作而又严密的数学方法,它测量的是风险价值的绝对值,更为简洁和直观,而且它关注的是一定显著性水平下,投资者资产的最大损失值,这一指标更具有合理性和科学性,因此能为各种金融机构起到更好的防范风险的指导作用。近十年来,VaR 方法的应用逐渐涉及到证券公司、投资银行、商业银行、养老基金以及非金融企业等。而且,VaR 的应用已不仅局限在金融市场风险的度量,在信用风险、流动性风险、现金流风险和操作风险方面也在逐步得到应用。

VaR 的概念简单明了,某种资产头寸的 VaR 就是在这样的一个数值:在一定时间内,该资产头寸市场价值的损失大小超过 VaR 的概率等于预先确定的置信度  $P$ ,从统计学的角度来看,VaR 就是某置信水平下的分位数。假设一只股票今天价格为  $P_t$ ,如果根据过去的历史数据能预测到明天这只股票的价格  $P_{t+1}$  小于

$P_t - 1.5$  元的概率为 5%，也即是预测这只股票明天只有 5% 的可能性下跌 1.5 元，也即是：

$$\Pr(P_{t+1} < P_t - 1.5) = 0.05$$

那么此时就可以说有 95% 把握明天这只股票不会下跌低于  $P_t - 1.5$  元，此时  $P_t - 1.5$  就是要求的 VaR 值。

### 3.1.1 计算 VaR 值的方法

VaR 方法是目前国际金融领域衡量市场风险的主流技术。迄今为止，已经有很多种方法可以来计算 VaR 的数值。这些方法主要可以分为两大类：参数法和非参数方法。

#### 一)、参数法

在采用参数法计算单个资产 VaR 时，通常需要假设分布是零均值正态分布，主要是想得到金融资产未来特定时间的波动性，即收益率的方差，进而求得 VaR 值。模拟这种波动性通常采用的方法是移动平均法和 GARCH 法。

#### 1.1) 简单移动平均

波动性和相关性的估计和预测最早是采用简单移动平均方法。简单移动平均法也称“等加权移动平均”，假定产生回报序列的随机过程是独立同分布的，并且在计算中采用等权重的移动平均。用数学公式表示如下：

$$\mu = \frac{1}{T} \sum_{t=1}^T R_t, \hat{\sigma}_T^2 = \frac{1}{T-1} \sum_{t=1}^T (R_t - \mu)^2$$

这里  $R_t$  表示金融资产第  $t$  期的回报， $\mu$  表示  $T$  期内的回报均值， $\sigma_T^2$  表示回报序列的标准差， $\hat{\sigma}_T^2$  是  $\sigma_T^2$  的一个无偏估计。大量的金融实证研究表明，如果  $\mu = 0$  模型将能得到更好的预测效果。即：

$$\hat{\sigma}_T^2 = \frac{1}{T-1} \sum_{t=1}^T R_t^2$$

简单移动平均法一个最大的优点就是概念容易理解，计算简单。但是，等权重也带来了相应的缺陷。因为，过去的信息对于未来市场的影响应该是随着时间延长而逐渐递减的，不能对所有过去的信息赋予相同的权重。而且，相同的权重可能使市场中发生的极端事件对未来市场波动性造成巨大的影响，同时还会将这个

影响延续至未来很长一段时间，这种现象也被称为“幽灵效应”。

### 1.2) 指数加权移动平均

为了避免简单移动平均上述的缺陷，人们提出了另外一种移动平均方法，那就是指数移动平均方法。指数移动平均法也就是对不同时期的资料赋予不同的权重。对于那些离当前较近的历史资料赋予较大的权重，而那些离当前时刻较远的历史资料赋予较小的权重。为了表达和计算的简单，采用指数的形式来表示权重。这里引入一个参数  $\lambda$  来表示这个权重，参数  $\lambda$  也称为衰减因子(decay factor)。

估计式如下：

$$\hat{\sigma}_T^2 = (1 - \lambda) \sum_{t=1}^T \lambda^{t-1} (R_t - \mu)^2$$

整理可得

$$\hat{\sigma}_T^2 = \lambda \sigma_{T-1}^2 + (1 - \lambda) R_{T-1}^2$$

对协方差的估计也可以采用同样的加权方式来获得，用式子表示如下

事实上，上面的指数移动平均模型式就是 IGARCH 模型，因此，指数移动模型具有 GARCH 模型的部分优点，可以更准确地模拟回报序列的自相关性和聚集性。JP.Morgan 的 Riskmetrics 就是采用这种方法来估计和预测波动性和相关性。

采用指数移动模型最关键的就是确定衰减因子  $\lambda$ ，它可以采用均方根误差法来确定。当  $\lambda = 1$  时，则指数移动平均模型退化为简单移动平均。JP.Morgan 的 skmetrios 经过对  $\lambda$  的估计，发现使用日资料时， $\lambda = 0.94$  比较合理；使用月资料时， $\lambda = 0.97$  比较合理。指数加权移动平均除了需要选择最佳的衰减因子  $\lambda$  外，还要选择最佳的历史窗口长度。历史窗口也就是历史数据包含的过去的时间长度。最佳历史窗口长度也会随着入的不同而不同。

事实上，如果时间序列是自相关的，则指数移动平均的预测效果更好，因为这种方法给予最近的历史数据最大的权重；相反，如果实际时间序列不是自相关的，显然指数移动平均的预测效果没有简单移动平均方法的效果好。尽管指数移动平均方法在一定程度上改进了简单移动平均方法，但它也表现出许多局限性：

- (1) 指数移动平均只在一步向前预测时才有效。
- (2) 目前还没有最佳的理论方法来估计衰减因子。
- (3) 使用指数移动平均方法在预测时间间隔较长的波动性时(如月波动性)，往

往会遇到数据量不足问题，如果处理不当，会产生很大误差。

(4)对于不同类型的资产，由均方根误差法原则得到的 $\lambda$ 值不同，所以在 Risktueries 中，所有资产都采用同一值 $\lambda$ 是不正确的。

(5)衰减因子是随时间显著变化的，所以使用常数衰减因子是不适当的。

## 二)、非参数法

众所周知，股票市场上经常有突发事件的产生，因此股票收益率的分布也会呈现出尖峰厚尾的特征，从而不满足正态分布，这种情况下再用参数方法进行计算 VaR 就显得不太妥当。而非参数方法的提出则很好的解决了这个问题，因为非参数方法计算 VaR 不需要对股票收益率分布做出假设，而且在计算过程上也不复杂。

### 2.1) 历史模拟法

历史模拟法的核心在于根据金融资产的价格或者是影响其价值的风险因子的历史样本变化来模拟金融资产的未损益分布，利用分位数(quantile)给出一定置信度下的 VaR 估计。历史模拟法是直接由金融资产或者组合价格的历史资料来构建实际的价格或者回报分布，可较为真实地反应金融资产或组合过去的风险状况。此外，历史模拟法假设过去价格变化的轨迹会一再的重复，于是可以将价格或回报的历史分布形态当成预期的概率分布。因此，历史法不需要对金融资产或组合的价格或者回报分布做任何假设，也不需要估计任何参数。

历史模拟法的计算相当简单。例如要计算一个组合的 VaR，通过前 T 期的历史价或者  $p_t(t = 1, 2, \dots, T)$  就可以获得一个价格波动序列  $\Delta p_t(t = 1, 2, \dots, T - 1)$ 。如果已知当前时刻的价格  $p_0$ ，则未来可能的价格序列  $p_{0+t}(t = 1, 2, \dots, T - 1)$  就可以用当前的价格  $p_0$  加上前面获得的价格波动序列  $\Delta p_t$ ，即： $p_0 + \Delta p_t$  将所得到的未来价格波动序列按照从小到大的顺序排列，即可获得投资组合的历史分布，再根据风险管理者设定的信赖水平，找出相对应的分位数即可求得 VaR。例如，如果我们得到了 1000 个未来的可能价格序列，在 95%的信赖水平下，预期将会有 50 个观测值超过 VaR，于是第 51 个值就是 VaR 的估计值。如果分位数不是整数则可以采用插值法求解。通过历史法即可以计算绝对 VaR，也可以计算相对 VaR。上面的例子获得的就是绝对 VaR。只要将上面

例子中的价格换成收益，就可以得到相对 VaR。

## 2.2) 蒙特卡罗模拟法

历史模拟法要正确估计收益率的分布需要比较多的历史数据，而且无法预测到历史上未发生的极端情况和较低概率事件。而蒙特卡罗的方法通过模拟未来特定期间内可能发生的不同情境下的相对应之资产价格变量，利用一个随机过程重复仿真，得到投资组合于未来特定期间之损益(或报酬)及其分布函数，进而估算出投资组合的风险值，因此其不需要大量的数据且能够预测到历史之外的事件的发生。蒙特卡罗模拟法主要有三个步骤:首先，选择标的资产价格变量的随机过程及其参数，这些参数可由历史数据求出;然后，利用随机数产生器及所设定之价格路径(Process)的随机过程模式，仿真出标的资产的价格路径。最后，基于仿真出的价格，计算收益率，并由此得到收益率的经验分布，从而找到一定分位上的 VaR 估计量。

该方法的具体步骤为:

(1)假定资产服从  $dS_t = \mu_t S_t dt + \sigma_t S_t dW$  的随机过程，其中  $dW \sim N(0, \sigma_t^2)$

(2)在一个很短的时间  $\Delta t$  内， $\Delta S_t = S_{t-1}(\mu \Delta t + \sigma \varepsilon \Delta t)$ ， $\varepsilon \sim N(0, \sigma_t^2)$ ;

$\sigma_t$  用窗口数据的历史波动率表示， $\mu_t$  用均值表示;

(3)对于向前一天的 VaR 的预测，设定  $\Delta t = 1 / 5000$  (天)，结合产生的 5000 个服从正态分布的随机变量  $\varepsilon$  的伪随机数，可以产生 5000 个相对应的向前一天的资产价格的预测值  $S_{t+1}$ ，从而计算出收益率的预测值  $r_{t-1}$ ;

(4)重复以上步骤 5000 遍，可以得到 5000 个向前一天的收益率预测，将之排序得到经验概率分布从而得到 VaR 的值。

蒙特卡罗方法不需要假定风险因子属于特定分布，可以照顾到收益率的厚尾性和波动率的时变性，可以较好的处理非线性、非正态问题。然而该方法的最大的缺点是计算量巨大，而且随着需要考虑的风险因子的增加，模型变得越来越复杂。Jamshidan and zhu (1997)提出了情景模拟的蒙特卡罗模拟法，运用主成分分析的方法减少需要分析的风险因子数量，并通过实现事前分析可能发生的情况来减少需要模拟的情景数量。Glasserman, Heidelberger and Shahabuddin (2000)将蒙特卡罗模拟方法和均值一方差分析结合起来减少了蒙特卡罗模拟的计算量，但

是并没有影响模型在预测 VaR 方面的表现。

### 2.3) 核密度估计方法

正如上文中所提到的，历史模拟法虽然简单易行，但对需要比较多的历史数据，对时间变化不敏感，而且无法预测到历史上未发生的极端情况和较低概率事件。蒙特卡罗方法虽然克服了这方面的缺陷，但它的计算量巨大，而且随着需要考虑的风险因子的增加，模型会变得越来越复杂。因此，一些学者引入了统计学中的非参数密度估计方法来进行 VaR 的计算。非参数密度估计始于直方图法，后来发展为最近邻法、核估计法等，其中理论发展最完善的是核密度估计法。

#### (2.3.1)核密度估计基本概念

核密度估计是在直方图的基础上由和加以改进得到的一种估计方法—非参数核密度估计方法。设 $(x_1, x_2, \dots, x_n)$ 为离散的随机样本，则单变量的核密度估计定义为

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right)$$

其中 $\hat{f}(x)$ 为总体未知密度函数 $f(x)$ 一个核密度估计， $k(\cdot)$ 为核函数， $h$ 为窗宽， $n$ 为样本容量。可以看出，核函数是一种权函数；该函数利用数据点 $x_i$ 到 $x$ 的距离 $(x - x_i)$ 来决定 $x_i$ 在估计点 $x$ 的密度时所起的作用。如果核函数选择标准正态密度函数 $\Phi(\cdot)$ ，则离 $x$ 点越近的样本点加的权就越大影响也越大

核密度估计结果既与样本有关，又与核函数及窗宽的选取有关 在给定样本以后，一个核估计的好坏，取决于核及窗宽的选取是否得当核函数和窗宽的选择直接影响密度函数的估计精度。

#### (2.3.2)核函数和窗宽的选择

##### ① 核函数的选择

一般核函数要求为对称的密度函数族  $\mathbf{P}$ ，即核函数  $k(\cdot)$ 满足如下条件：  
 $k(-x) = k(x)$ ； $k(x) > 0$ ； $\int k(x)dx = 1$ 。根据使用不同核函数进行估计时得到的 IMSE 可以发现核密度函数对密度估计的结果影响较小，文中将使用一个常见

核函数正态核函数  $k(u) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{u^2}{2}\right)}$  进行分析。

② 窗宽确定方法—交叉印证法

对核密度估计结果有影响的另外一个因素是窗宽，并且理论和经验均证明，窗宽对估计结果有重要的影响。窗宽选取中一个常用的方法是交叉印证法 (Cross-Validation Method 简记: CV)。该方法由 Rudemo(1982) and Bowman(1984) 提出后，成为较为流行的窗宽估计方法，CV 方法目标是使 ISE 最小，其中 ISE 定义为：

$$ISE(h) = \int (\hat{f}(x) - f(x))^2 dx$$

该方法的主要思路是:首先在全部数据中挑出一个观察值，然后用其余部分建立模型，并用刚才挑出的观察值检验整个估计的效果来选择使 ISE 最小的 h 即：

$$ISE(h) = \int \hat{f}^2(x)dx - 2 \int f(x)\hat{f}(x)dx + \int f^2(x)dx$$

在样本独立的条件下有以下结论：

$$CV(h) = \int \hat{f}(x)dx - \frac{2}{n} \sum_{s=1}^n f_n^{(-s)}(X_s)$$

其中： $f_n^{(-s)}(X_s) = \frac{1}{n-1} \sum_{t \neq s} K(X_t - x)$ ， $X_s$  挑出印证模型效果变量。

(2.3.3)核密度估计计算 VaR

正如上文中提到的，通过核密度估计可以得到单个变量的密度函数估计，有了密度函数再求 VaR 就只是求一个对应水平下的分位数。

下面就正态核函数下的核密度估计求取 VaR 值进行理论上的推导。

假设我们要求置信水平 p 下的 VaR 值，那么则有

$$\begin{aligned} p &= F(x) = \int_{-\infty}^{x_p} \hat{f}(x)dx \\ 1 - p &= 1 - F(x) = 1 - \int_{-\infty}^{x_p} \hat{f}(x)dx = \int_{x_p}^{+\infty} \hat{f}(x)dx \\ &= \int_{x_p}^{+\infty} \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}} dx = \frac{1}{n\sqrt{2\pi}} \sum_{i=1}^n \int_{x_p}^{+\infty} e^{-\frac{(x-x_i)^2}{2h^2}} d \frac{x - X_i}{h} \end{aligned}$$

令  $y = \frac{x - x_i}{h}$  则

$$1 - p = \frac{1}{n\sqrt{2\pi}} \sum_{i=1}^n \int_{x_p}^{+\infty} e^{-\frac{y^2}{2}} dy$$

通过求得上式中  $x_p$  即可求得密度函数对应概率  $p$  的下侧分位数，也即是对应水平  $p$  下的 VaR 值。

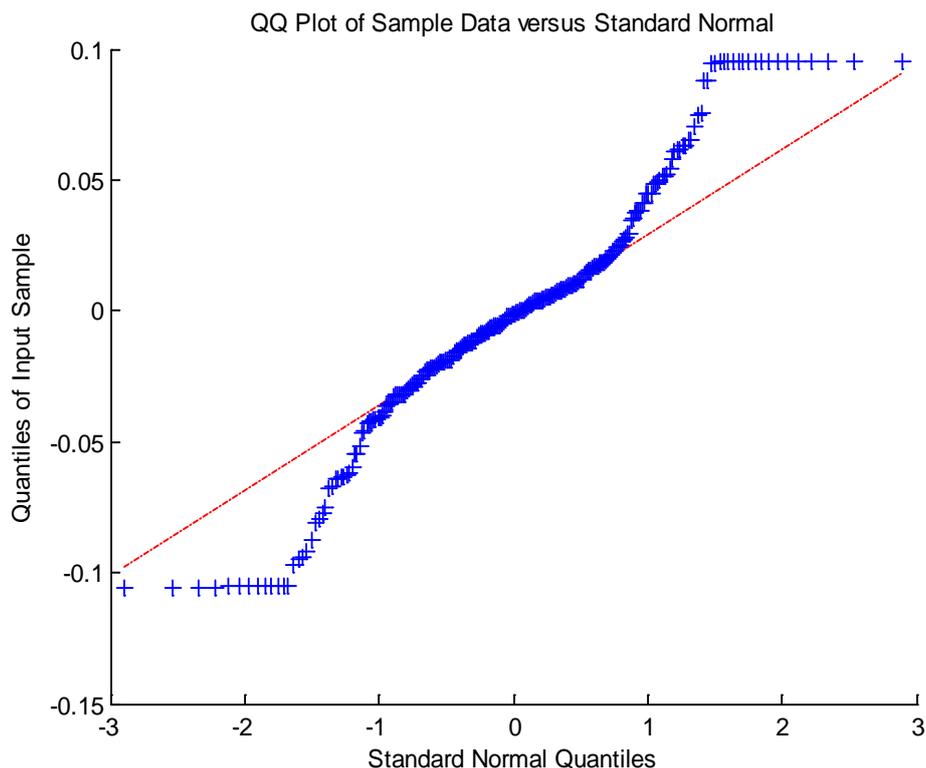
### 3.2、模型实证分析

项目中我们根据齐鲁证券公司提供的历史上曾经发生过大跌的公司，选取这些公司的股票价格数据，来通过上文中所提到的方法来计算 VaR 值。我们选取了昌九生化，酒鬼酒等十家公司股票收盘价数据，通过股票收盘价来得到每天的收益率，进而对收益率求取 VaR 值。

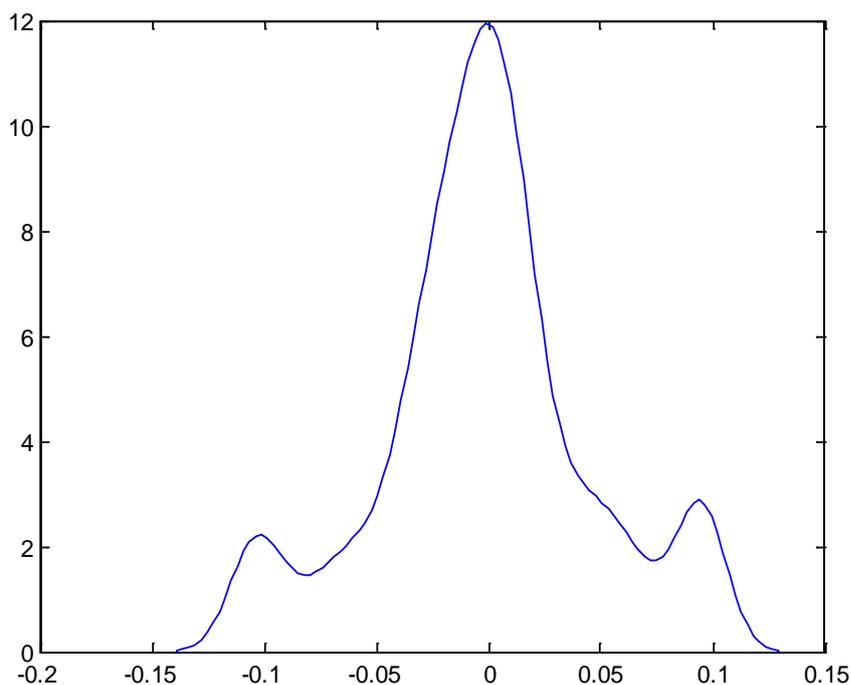
首先，对得到的收益率数据进行描述性统计，得到每个公司收益率数据的均值，标准差，偏度以及峰度。结果如下表所示：

	mean	std	skewness	kurtosis
旭光股份	-0.00065	0.035158	-0.33078	4.574035
蓝帆股份	-0.00015	0.036734	-0.18366	4.247648
南岭风暴	-0.00128	0.030059	-0.16797	5.526263
南方食品	-0.00128	0.030059	-0.16797	5.526263
ST 贤成	6.09E-05	0.028419	-0.45391	5.095591
如意集团	0.000414	0.036893	-0.02346	3.950466
酒鬼酒	-0.00111	0.029771	-0.40726	5.071775
昌九生化	-0.00209	0.031786	-0.10384	4.791881
百圆裤业	-0.00458	0.044906	-0.02164	3.473621
洪都航空	0.001041	0.026808	-0.79382	6.088397

通过上表可以看出这些公司股票收益率数据均值有些和 0 较为接近，有些显著不为 0，而偏度值都为负值，且峰度值都明显的比 3 大（正态分布偏度为 0，峰度为 3），说明这些收益率数据明显异于正态分布。我们选取昌九生化的数据进行更进一步的分析，用昌九生化收益率数据，做出 Q-Q 图，如下所示：



从上图也可以看出，昌九生化收益率数据呈现出很明显的厚尾特性。偏度值和峰度值，以及上面的 Q-Q 图，都证明收益率数据均值并不为 0，而且收益率的分布也不是正态分布，此时如果用上文中提到的参数方法，那这些基本假设不满足，因此不适合用参数的方法去计算 VaR 值。此时我们可以采用上文中提到的非参方法进行计算，鉴于核密度估计的优良特性，我们选取核密度的估计方法进行计算 VaR。用 MATLAB 中自带的 ksdensity 函数，可以轻松的实现核密度估计，而且可以选取不同的核密度函数和带宽进行估计。我们选取常用的正态核函数，窗宽的选取则采用正态核函数相应的窗宽。我们首先选取昌九生化近两年的收益率数据进行核密度估计，得到估计的密度分布函数如下图：

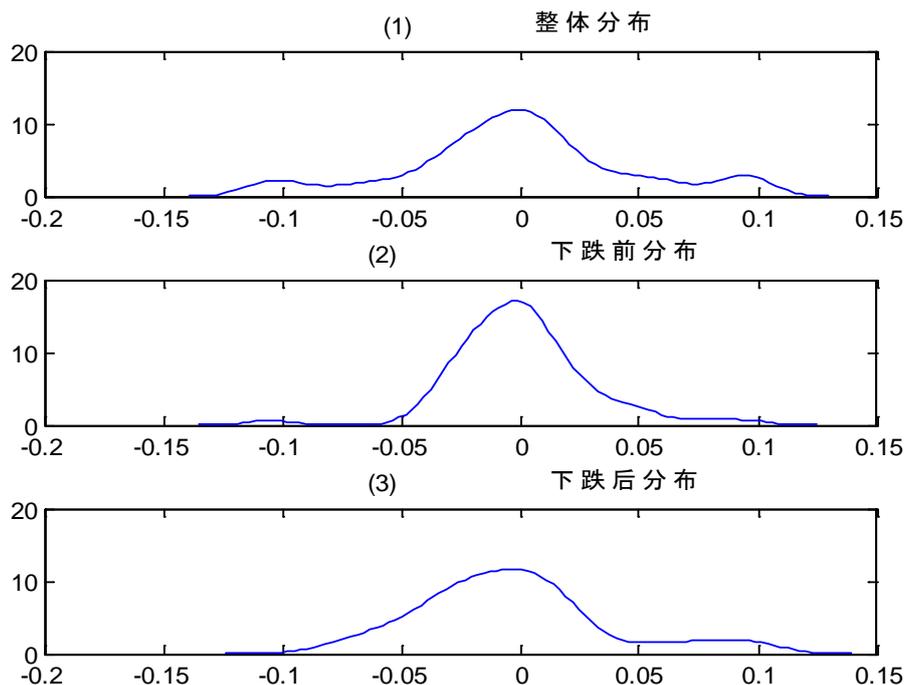


因为股票价格有涨跌停的限制，所以收益率的数据很少有超过 0.1 的，而核密度估计对估计进行了平滑处理，因此得到密度函数估计图像在 0.1 处有凸起部分，当然另外一个原因也是厚尾现象的影响。

为了验证大跌前和大跌后的分布是否一样，进一步将数据分成大跌之前的数据和大跌之后的数据分别进行分析，描述性统计如下：

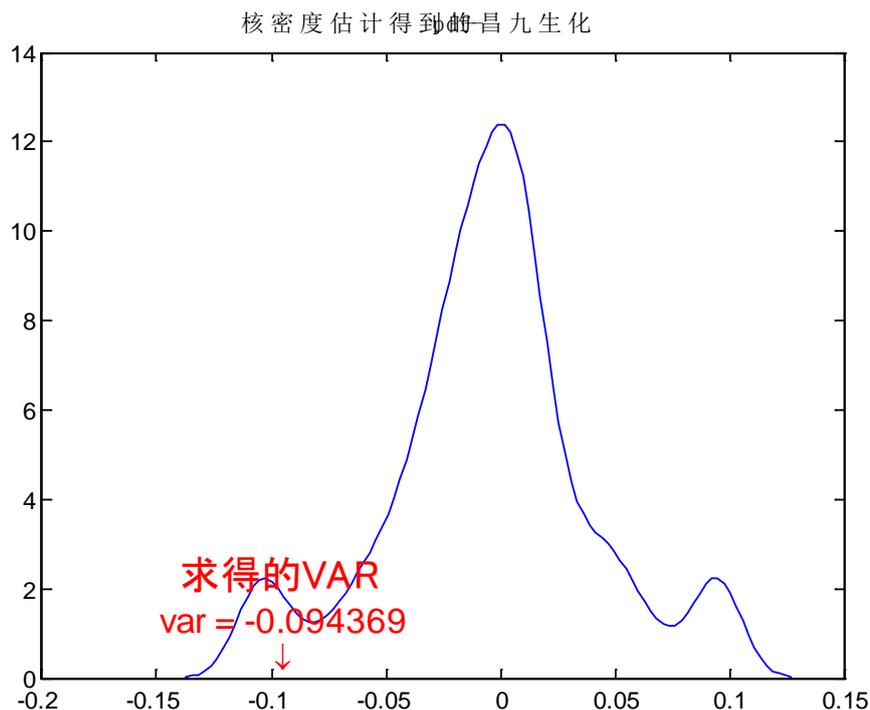
	下跌前	下跌后
均值	0.000325	-0.003473
方差	0.028877	0.039194
偏度	0.202445	0.760729
峰度	6.200169	3.671842

可以看出，下跌之前和下跌之后均值发生微小变化，但是偏度和峰度却发生了十分显著地变化。更进一步，对下跌前和下跌后的收益率数据用核密度估计方法拟合的结果如下图：



从上面的三幅图也可以看出，选取不同时段的数据，得到的核密度估计会出现明显的不同，而且核密度估计对数据量的要求也不是很大，既能保持估计对数据的敏感性，又能得到理想的估计结果。

接着，我们继续用核密度估计的方法对昌九生化收益率数据计算 VaR 值。有了估计的密度分布函数，只需要用密度函数求出对应概率的分位数即求得了相应概率的 VaR 值。由于核密度估计对数据选择的敏感性，选择数据就显得十分重要，我们选择昌九生化大跌前后的一小段时间进行 VaR 的计算，选取的数据截止到 2014/8/4。因为这些数据包含了比较多的下跌数据，所以估计得到的密度分布函数厚尾性就会更加明显，求得的 VaR 值相应的也会比较大。下面我们就按照上文中提到的方法对昌九生化进行 VaR 计算，核密度函数仍然选取正态核函数，带宽也用 MATLAB 中对应正态核函数的最优带宽，估计密度分布函数和 VaR 如下图所示：

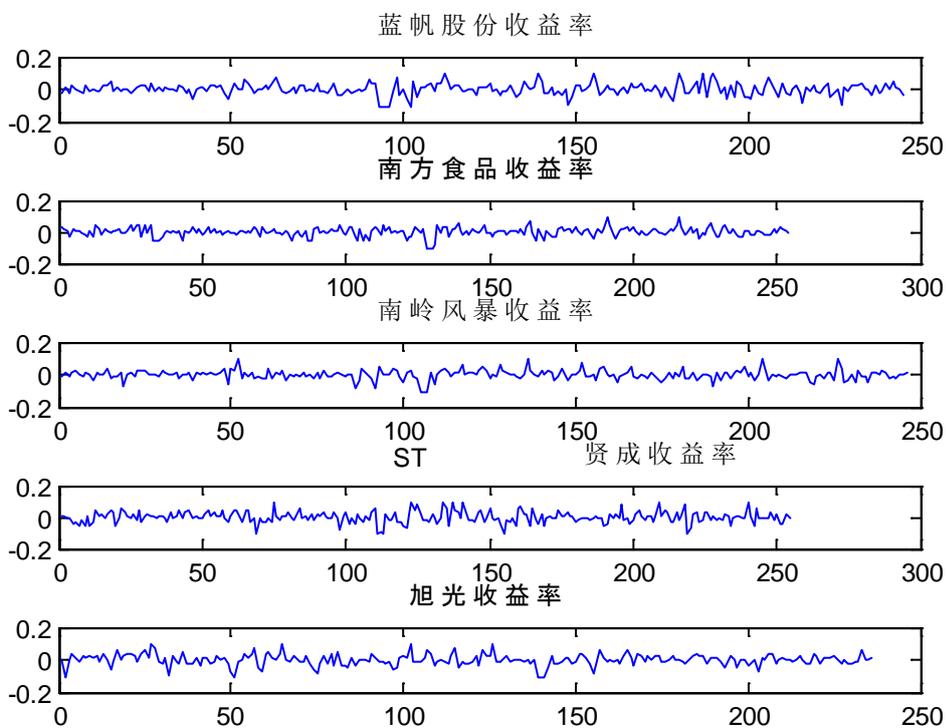


从图中可以看到，由核密度估计得到的正如我们所预料的，绝对值的确比较大。这样算出来的 VaR 值比较保守，因为我们选取的数据是昌九生化历史上下跌最厉害的一段的数据，这段数据中包含了较多的下跌比较严重的数据，也就是收益率为负数较多的数据，这样的数据的分布就有负的偏度值，而且厚尾性也更加严重，那计算得出的 VaR 值的绝对值相应的也就越大。因此我们就算出的 VaR 值为-0.094369，很接近-0.01，这时我们就可以说明天收盘时，也就是 2014-8-5 日，收盘价下跌的幅度较今天的收盘价有 95%的可能不会下跌超过 9.4369%，或者说明天收盘价相对于今天的收盘价有 5%的可能会下降超过 9.4369%。

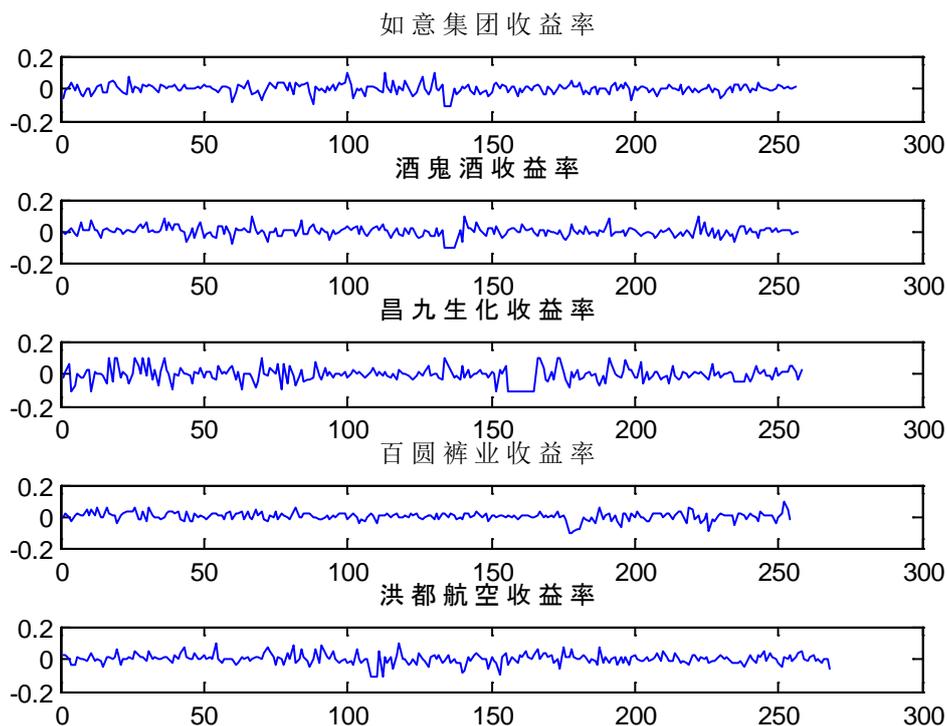
实际中，对于数据的选择可以根据当前行情进行主观的判断，根据对股票的预期来选择不同时段的历史数据进行计算，这样就会得到比较符合当前行情 VaR 值。对其他几家公司选取类似的数据，按照与计算昌九生化 VaR 一样的步骤得到的这些公司 VaR 值如下表所示：

蓝帆股份	-0.061926	如意集团	-0.052973
南方食品	-0.049875	酒鬼酒	-0.053894
南岭风暴	-0.050763	昌九生化	-0.094369
ST 贤成	-0.059824	百圆裤业	-0.041717
旭光股份	-0.064676	洪都航空	-0.049019

从上表中可以看出，洪都航空和南方食品的 VAR 较小，我们画出昌九生化、洪都航空和南方食品的股票价格以及对数收益率图形如下所示：



,



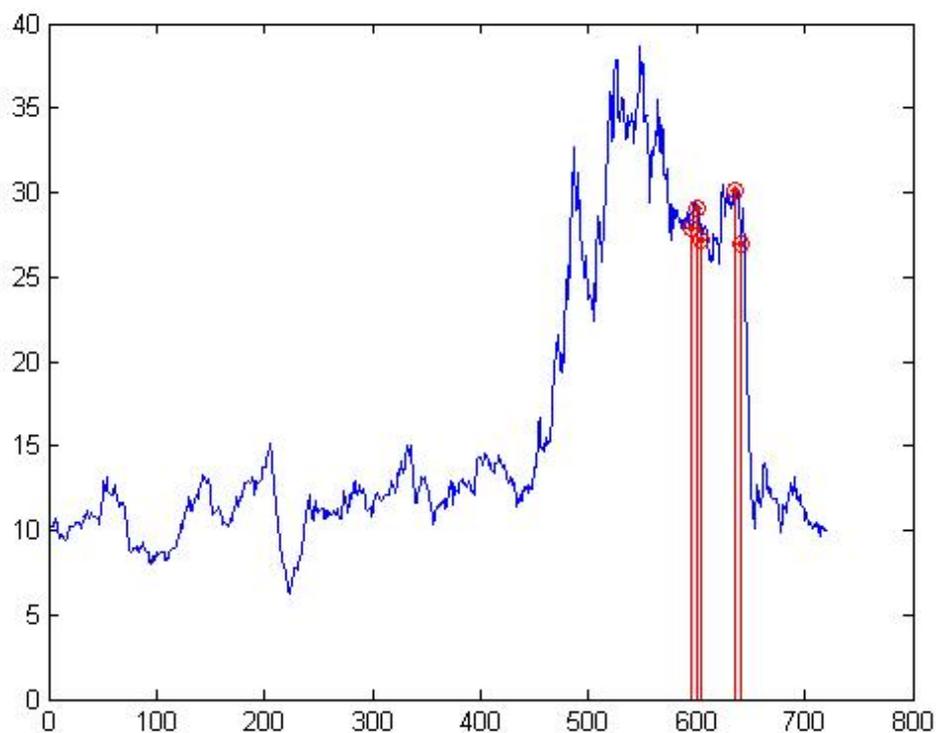
从上面第一幅图中我们可以看出，南方食品大部分数据都较为平稳，只有在第 100 到 150 个数据间出现了较少的收益率的下跌，而且在第 200 个数据附近出现了比较多的收益率上涨的数据，它的 VaR 值为-0.049875，也是第一幅图中五家公司最小的。而旭光股份、蓝帆风暴以及 ST 贤成的收益率波动都比较大，而且收益率下跌的数据也较多，它们的 VaR 值计算出来相对也就较大。南岭风暴的收益率数据看上去也比较大，看上去收益率下降的数据很多，但是它收益上涨的数据也很多，因此最后计算出来的 VaR 值也相对较小。而第二幅图中，昌九生化的收益率波动明显比其他几家公司的波动大，而且收益率下降为负的数据也较多，因此最后计算出来昌九的 VaR 值也是五家公司中最大的为-0.094369，这时昌九数据在我们数据截止日的下一个交易日就很有可能会接近跌停。百圆裤业是五家公司中收益率相对最平稳的一家公司，只在第 175 个数据附近出现了一次较大的负的下落，其他大多数时间，收益率波动都不大，因此最后计算出来的 VaR 值也是五家公司中最小的一个。其他三家的收益率数据也都有一定幅度的波动，但较昌九都要小，较百圆裤业都要大一些，因此计算出来的 VaR 绝对值也是比昌九生化的小，比百圆裤业的大。从以上十家的收益率数据和最后计算出来的 VaR 值可以看出，我们的 VaR 计算方法很好的反应出了不同公司股票收益率

的差异，从而很好的反应出了不同股票价格的表现。股票价格表现出较大波动幅度的公司，其收益率也就较大，计算出来的 VaR 值就相对较大，而股票价格表现出较小波动幅度的公司，其收益率就较小，其相应的 VaR 值就相对较小。

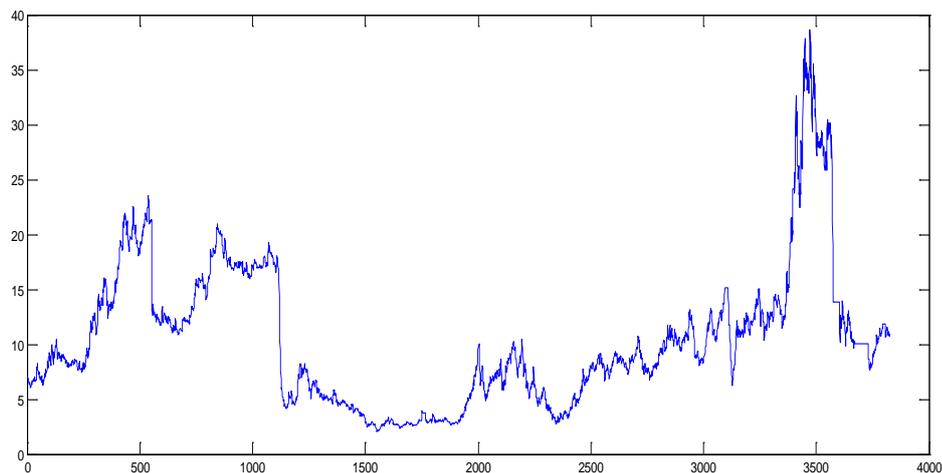
### 3.3、结合泡沫模型计算 VaR 值

正如上文中所提到的，在计算上面的十家公司 VaR 时，我们选取的是该公司股票在出现大跌前后的数据，由于这些数据包含了下跌比较严重的一段数据，因此最终得出的 VaR 绝对值可能会偏大，也即是我们对损失的预期较为保守。由于非参数方法计算 VaR 时对数据的敏感性，实际中在预测 VaR 时，数据的选取很重要。上文中关于数据的选取似乎显得太为主观，没有依据，但其实泡沫模型的预警为我们提供了一个重要的数据选取依据。基于泡沫模型，我们提出了一下数据选取标准：当泡沫模型产生预警时，我们就选择股票最近一段时间（T1）的收益率，同时选取历史上出现过的较为严重下跌的几段时间的收益率（T2, T3, T4....），用所有这些收益率来就进行计算，从而得出预期的 VaR 值，而如果泡沫模型没有产生预警，那我们就仅选取最近一段时间的（T1）的收益率数据，从而计算出 VaR 值。这个数据选取的标准依赖于一个假设，那就是历史可能会重现，当泡沫出现时，我们就假设股票会出现历史上出现过的大跌，而没有泡沫出现，那就认为股票价格是正常增长，不存在泡沫，从而不会出现历史上出现过的大跌。

接下来我们以昌九生化为例来说明如何结合泡沫模型来计算 VaR 值，首先，我们对昌九生化近期的收益率数据建立泡沫模型，得出如下图结果：

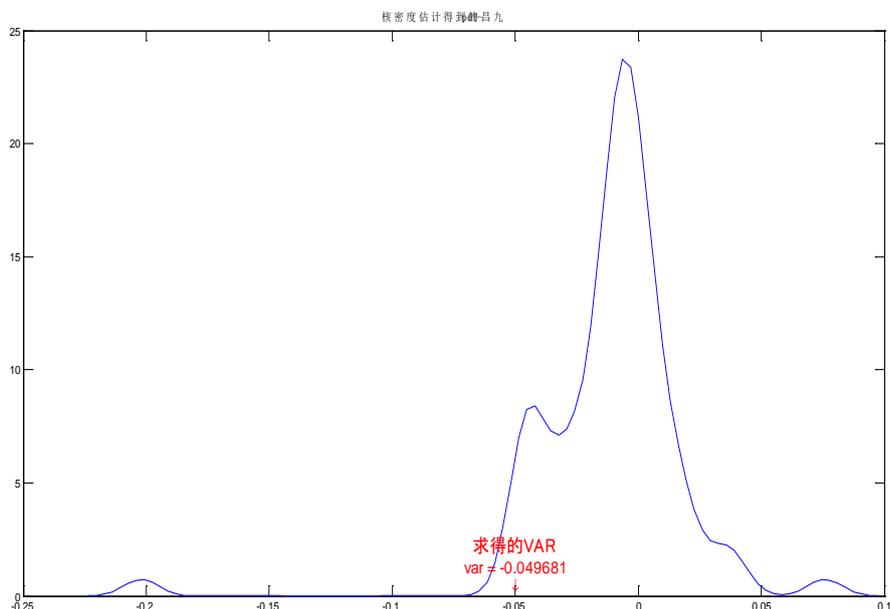


从图中可以看出，在大概第 600 个位置产生了泡沫预警，那我们就选取在第 600 天前 30 天（T1）的股票收益率数据，之后我们再看历史上曾经出现过的大跌情况，昌九生化从上市至今的股票价格数据如下所示：

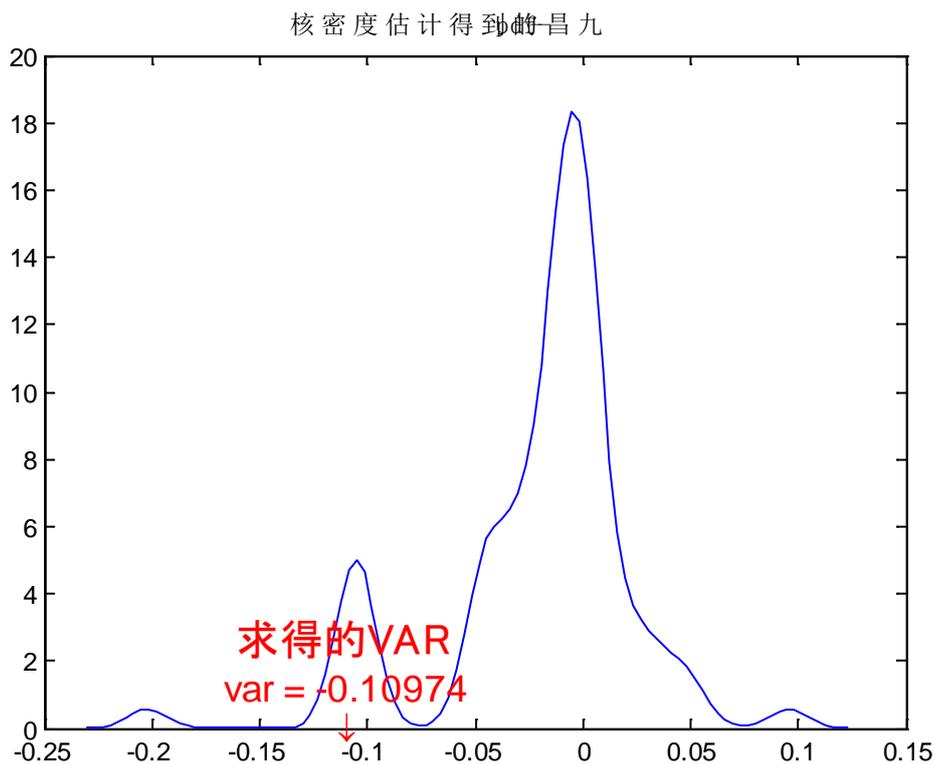


从图中我们可以看出在大概第 550 个数据以及 1100 个数据曾经出现过较为严重的大跌，对应到估计日期数据就是 2001-05-18 和 2003-09-08，对于较近 2003-09-08 的大跌我们选取和 T1 时间段一样个数的数据即选取 30 个数据（T2），对于稍远的 2001-05-18 的下跌我们就选取个数为 T1 一半的数据也即是选取 15 个数据

(T3)，之后我们首先对股价的数据进行处理，得出收益率数据，之后按照(三)中计算 VaR 的方法进行计算，得出预测的概率密度函数估计图和 VaR 值如下图所示：



得到预测的 VaR 值为-0.049，而实际中近期的第 601 个数据为-0.012，我们预测的 VaR 数据较实际有些偏大，这也正如上面我们所说明的，我们的 VaR 值由于数据的选取是近期的数据加上历史上出现过大跌时间段的数据，所以计算出的 VaR 会偏保守，也即是绝对值会比较大。但是从股价的图上我们可以看出，更大幅度的大跌实际发生在第 650 个数据位置附近，因此我们选取第 650 个数据前 30 天的数据做为 T1 时间段的数据，T2 和 T3 时间段的数据不变，再 VaR 的计算，得出下图：



这次计算得出的 VaR 值为-0.1097，而实际的第 651 个数据的值为-0.1054，这时我们预测的 VaR 值较实际的相差就比较小了，可见我们偏保守的 VaR 在实际发生较为严重的大跌时预测的还比较准确，但是在没有发生很严重下跌时，我们的 VaR 计算就较为保守，较实际值会比较大。

## 四、总结与展望

在项目现有研究成果的基础上，我们希望接着研究不同股票的维持担保比率。当强制平仓大额数量的股票时，可能导致股价急剧下降，造成额外的冲击成本。在这样的情况下，就需要考虑股票的抗压能力及其流动性的好坏。对于抗压能力强、流动性较好的股票，极端情况对其造成的冲击成本较小，这样的股票可以设置较低的维持担保比率；而对于抗压能力较弱、流动性较差的股票，极端情况会对其造成较大的冲击成本，对于这样的股票就需要设置相对高的维持担保比率，并对其加以监控，以免造成损失。

从 2014 年 4 月份开题以来，在项目负责人林路教授的严格要求下，项目组几乎每周都会组织正式的讨论班专门讨论项目进展和下一步的研究计划；此外，我们还积极与齐鲁证券的项目对接人员联系，了解相关业务的具体开展情况，了解齐鲁证券公司的具体目标，交流双方对项目所要解决的问题的理解。正是在林教授的严格要求和项目组全体成员的共同努力下，项目一步一个脚印，基本上达到了预期的目标。项目组全体成员感谢齐鲁证券对本项目的大力支持，感谢金融研究院为项目组提供的良好的研究条件！

## 附录

### 项目核心程序:

```
function test_BubbleAlarm_8(FileName,WinWidth,StepLen,range)
format long g
price0 = xlsread(strcat(FileName, '.xls'),1,'B:B');
price = price0(range(1):range(2));
DataLen = length(price);
WinNum = floor((DataLen - WinWidth) / StepLen);
[~,time_0] = xlsread(strcat(FileName, '.xls'),1,'A:A');
time = time_0(range(1):range(2));
price_mean = mean(price);
if price_mean < 15
    L = [8,1.2];
    U = [11,1.4];
elseif price_mean > 15 && price_mean < 30
    L = [13,1.3];
    U = [16,1.6];
else
    L = [15,1.4];
    U = [18,1.7];
end
best_so_far = zeros(1,3);
best_history = ones(1050,3);
tabu_list = zeros(1020,4);
Tt = zeros(DataLen,1);
tc = zeros(DataLen,1);
tcc = zeros(DataLen,1);
Sc = ones(WinNum,1);
R_best = zeros(1010,2);
variance = ones(10,1);
p_value1 = ones(WinNum,1);
lnK = ones(1,1);
b = ones(1,1);
Ta = 0;
initialpoint = mod(DataLen - WinWidth, StepLen);
for t=1:WinNum
    R0=unifrnd(L,U);
    windowfirst = ((t-1) * StepLen + initialpoint + 1);
    windowlast = ((t-1) * StepLen + WinWidth + initialpoint);
    for j = windowfirst : windowlast
        Tt(j) = exp((R0(1)-log(price(j)))/R0(2))+j;
        Ta = Ta + Tt(j);
    end
end
```

```

end
T = Ta/WinWidth;
for j = windowfirst : windowlast
    tc(j) = Tt(j) - T;
end
tcc=tc(windowfirst : windowlast);
[h,p_value,~]=adftest(tcc);
p_R0 = p_value;
best_so_far = [p_R0(1,1),R0(1)-0.02,R0(2)-0.001,R0(1)+0.02,R0(2)+0.001];
best_history(1,:) = [p_R0(1,1),R0(1),R0(2)];
tabu_list(1,:) = [R0(1)-0.02,R0(2)-0.001,R0(1)+0.02,R0(2)+0.001];
LU = [R0(1)-0.02,R0(2)-0.001,R0(1)+0.02,R0(2)+0.001;
      R0(1)-0.04,R0(2)-0.002,R0(1)+0.04,R0(2)+0.002;
      R0(1)-0.08,R0(2)-0.004,R0(1)+0.08,R0(2)+0.004;
      R0(1)-0.16,R0(2)-0.008,R0(1)+0.16,R0(2)+0.008];
R = zeros(3,2);
for i = 1:3
    R(i,:) = unifrnd(LU(i+1,1:2),LU(i+1,3:4));
    while ((LU(i,1)<=R(i,1))&&(R(i,1)<=LU(i,3)))&&...
          ((LU(i,2)<=R(i,2))&&(R(i,2)<=LU(i,4)))
        R(i,:) = unifrnd(LU(i+1,1:2),LU(i+1,3:4));
    end
    Ta = 0;
    for j = windowfirst : windowlast
        Tt(j) = exp((R(i,1)-log(price(j)))/R(i,2))+j;
        Ta = Ta+Tt(j);
    end
    T = Ta/WinWidth;
    for j = windowfirst : windowlast
        tc(j) = Tt(j)-T;
    end
    tcc = tc(windowfirst : windowlast);
    [h(floor(i)),p_value(floor(i)),~] = adftest(tcc);
    p = p_value';
end
[MIN1] = find(p==min(p));
loca_pmin = MIN1(1,1);
neighbor_pmin = p(loca_pmin);
best_history(2,:) = [neighbor_pmin,R(loca_pmin,1),R(loca_pmin,2)];
tabu_list(2,:) = [R(loca_pmin,1)-0.02,R(loca_pmin,2)-0.001,...
                R(loca_pmin,1)+0.02,R(loca_pmin,2)+0.001];
for k = 1:1000
    R_best(k,:) = best_history(k+1,2:3);
    LU1 = [R_best(k,1)-0.02,R_best(k,2)-0.001,R_best(k,1)+0.02,R_best(k,2)+0.001];

```

```

R_best(k,1)-0.04,R_best(k,2)-0.002,R_best(k,1)+0.04,R_best(k,2)+0.002;
R_best(k,1)-0.08,R_best(k,2)-0.004,R_best(k,1)+0.08,R_best(k,2)+0.004;
R_best(k,1)-0.16,R_best(k,2)-0.008,R_best(k,1)+0.16,R_best(k,2)+0.008];
    R_1 = zeros(3,2);
    for i = 1:3
        R_1(i,:) = unifrnd(LU1(i+1,1:2),LU1(i+1,3:4));
        for n = k:(k+9)
            while ((tabu_list(n,1)<=R_1(i,1))&&(R_1(i,1)<=tabu_list(n,3))&&...
(tabu_list(n,2)<=R_1(i,2))&&(R_1(i,2)<=tabu_list(n,4)))||...
((LU1(i,1)<=R_1(i,1))&&(R_1(i,1)<=LU1(i,3))&&...
(LU1(i,2)<=R_1(i,2))&&(R_1(i,2)<=LU1(i,4)))
                R_1(i,:) = unifrnd(LU1(i+1,1:2),LU1(i+1,3:4));
            end
        end
        Ta = 0;
        for j = windowfirst : windowlast
            Tt(j) = exp((R_1(i,1)-log(price(j)))/R_1(i,2))+j;
            Ta = Ta+Tt(j);
        end
        T = Ta/WinWidth;
        for j = windowfirst : windowlast
            tc(j) = Tt(j)-T;
        end
        tcc = tc(windowfirst : windowlast);
        [h(floor(i)),p_value(floor(i)),~] = adftest(tcc);
        p = p_value';
    end
    [MIN1] = find(p==min(p));
    loca_pmin = MIN1(1,1);
    neighbor_pmin = p(loca_pmin);
    best_history(k+2,:) = [neighbor_pmin,R_1(loca_pmin,1),R_1(loca_pmin,2)];
    tabu_list(k+10,:) = [R_1(loca_pmin,1)-0.02,R_1(loca_pmin,2)-0.001,...
        R_1(loca_pmin,1)+0.02,R_1(loca_pmin,2)+0.001];
    if neighbor_pmin<0.005&& k>10
        break;
    end
end
rank_best = sortrows(best_history,1);
for ii = 1:10
    lnK = rank_best(ii,2);
    b = rank_best(ii,3);
    Ta = 0;
    for j = windowfirst : windowlast
        Tt(j) = exp((lnK-log(price(j)))/b)+j;
    end
end

```

```

        Ta = Ta+Tt(j);
    end
    T = Ta/WinWidth;
    for j = windowfirst : windowlast
        tc(j) = Tt(j)-T;
    end
    tcc = tc(windowfirst : windowlast);
    variance(ii) = var(tcc);
end
location = find(variance==min(variance));
lnK_last = rank_best(location,2);
b_last = rank_best(location,3);
Ta = 0;
for j = windowfirst : windowlast
    Tt(j) = exp((lnK_last-log(price(j)))/b_last)+j;
    Ta = Ta+Tt(j);
end
T = Ta/WinWidth;
Sc(t) = T-windowlast;
for j = windowfirst : windowlast
    tc(j) = Tt(j)-T;
end
tcc = tc(windowfirst : windowlast);
[~,p_value1(t),tstat] = adftest(tcc);

end
Sc_p = [Sc,p_value1];
loc = find(Sc_p(:,1) < WinWidth & Sc_p(:,2) < 0.1);
loc_alarm = (loc-1)*StepLen + WinWidth + initialpoint ;
time_label = 1:DataLen;
time_label = time_label';
date = xlsread(strcat(FileName,'.xls'),1,'C:C');
date_1 = date(range(1):range(2));
range_low = WinWidth/2;
range_up = DataLen;
plot(time_label(range_low:range_up), price(range_low:range_up),'b');
set(gcf, 'Units', 'normal', 'Position',[0,0,1,1]);
set(gca, 'yLim', [min(price)-5, max(price)+5], 'Units', 'normal', ...
    'Position',[0.1,0.2,0.8,0.8]);
hold on
plot(time_label(loc_alarm),price(loc_alarm),'r');
hold on
stem(time_label(loc_alarm),price(loc_alarm),'r');
index_0 = sort([1:100:DataLen,loc_alarm]);

```

```
index = unique(index_0);
set(gca, 'XTick',time_label(index), 'XTickLabel', time(index));
rotateticklabel(gca,'x',-90);
saveas(gcf, strcat(FileName, '.jpg'));
col_name = {'警报时间','股价','临界时距','检验p值'};
xlswrite(strcat(FileName, '_alarm', '.xls'), col_name, 'Sheet1', 'A1');
xlswrite(strcat(FileName, '_alarm', '.xls'), time(loc_alarm), 'Sheet1', 'A2');
xlswrite(strcat(FileName, '_alarm', '.xls'), price(loc_alarm), 'Sheet1', 'B2');
xlswrite(strcat(FileName, '_alarm', '.xls'), Sc_p(loc,:), 'Sheet1', 'C2');
end
```

```

clear;clc;
data=xlsread('rawdata.xls','sheet4');
xi=data(:,1:end-1);
[n,k]=size(data);
k=k-1;
index_of_xi_array=ones(1,k);
X=[ones(n,1) xi];
Y=data(:,end);
fprintf('第1次计算结果:\r')
beta_mao=((X'*X)\X'*Y);
fmt_str0="";
for i0=1:k+1
    fmt_str0=[fmt_str0 'Â' num2str(i0-1) '= %0.4f\r'];
end
fprintf(fmt_str0,beta_mao)
fprintf('\r')
x_ba=mean(xi);
y_ba=mean(Y);
St_square=sum(Y.^2)-n*y_ba^2;
lxy=sum((xi-ones(n,1)*x_ba).*((Y-y_ba)*ones(1,k)));
Sr_square=sum(beta_mao(2:end).*lxy);
Se_square=St_square-Sr_square;
c_flag=Sr_square/Se_square;
F_alpha=input('>>>>>>ÇëÊäËëÄúÒªÇóμÄÏÖøÐÔË®Æ½(0<Á<1)!:Á= ');
while ~(isscalar(F_alpha) && F_alpha<1 && F_alpha>0)    F_alpha=input('>>>>>>请输入您
    要求的显著性水平(0<α<1) α=');
end
F_fenweidian=finv(1-F_alpha,k,n-k-1);
c=k/(n-k-1)*F_fenweidian;
if c_flag>c
fprintf(['\r-----回归方程显著性检验 (H0: β1=β2=...=βk=0) ' ...
    '\r经过计算: 拒绝H0, 原假设不成立, 有线性关系。'])
else    fprintf(['\r-----回归方程显著性检验 (H0: β1=β2=...=βk=0) -----\r经过计算:
接受H0, 原假设成立, 无线性关系。'])
    break;
end
fprintf(['\r\r-----回归系数显著性检验 (分别对β1、β2、...、βk进行) -----'])
flag_go_on=1;
num_of_loop=0;
while flag_go_on
cij=inv(X'*X);
cii=diag(cij);
F_fenweidian_1=finv(1-F_alpha,1,n-k-1);
ci=sqrt(cii(2:end)*Se_square*F_fenweidian_1/(n-k-1));

```

```

format_str='%15.4f';
for ii=1:k-1
    format_str=[format_str '%13.4f'];
end
fprintf(['\r第%d次检验: \rcii: ' format_str '%13.4f\r ci: '
        format_str '\rβi: ' format_str '%13.4f'],num_of_loop+1,cii,ci,beta_mao)
if ~all(abs(beta_mao(2:end))>ci)
    flag_go_on=1;
    beta_1tok=beta_mao;
    beta_1tok(1)=[];
    fi_xin=beta_1tok.^2./cii(1:end-1);
    min_fi=min(fi_xin);
    beta_index=find(fi_xin==min_fi)+1;
    fprintf('\rx%d对y的线性影响最不显著 ( |β%d|= %0.4f )。删除x%d, 进行第%d次计算: ',
            beta_index-1+num_of_loop,beta_index-1+num_of_loop,...
            abs(beta_mao(beta_index)),beta_index-1+num_of_loop,...
            num_of_loop+2)
else
    fmt_str2='x%d';
    index_of_xi=find(index_of_xi_array);
    for i2=1:length(find(index_of_xi))-1
        fmt_str2=[fmt_str2 '¡x%d'];
    end
    fprintf(['\r\r经过检验, 剩余所有变量:' ...
            fmt_str2 '对y的线性影响均显著。检验结束。 \r'],index_of_xi)
    flag_go_on=0;
end
if flag_go_on
    num_of_loop=num_of_loop+1;
    k=k-1;
    if ~k
        fprintf('\r\r警告: 通过一一对所有变量做显著性检验, 已剔除所有变量! ');
        break;
    end
beta_mao=beta_mao-beta_mao(beta_index)/cii(beta_index)*cij(beta_index,:);
beta_mao(beta_index)=[];
fmt_str1="";
for i1=1:beta_index-2
    fmt_str1=[fmt_str1 'Â' num2str(i1) ' = %0.4f\r'];
end
for i1=beta_index:k+1
    fmt_str1=[fmt_str1 'Â' num2str(i1-1+num_of_loop) ' = %0.4f\r'];
end
fprintf(['\r'Â0 = %0.4f\r' fmt_str1],beta_mao)

```

```
X(:,beta_index)=[];
index_of_xi_array(beta_index-1+num_of_loop-1)=0;
xi=X(:,2:end);
x_ba=mean(xi);
lxy=sum((xi-ones(n,1)*x_ba).*((Y-y_ba)*ones(1,k)));
Sr_square=sum(beta_mao(2:end).*lxy);
Se_square=St_square-Sr_square;
end
end
```

## 参考文献

- [1]. Bikker, J. A., Spierdijk, L., & Van Der Sluis, P. J. (2007). Market impact costs of institutional equity trades. *Journal of International Money and Finance*, 26(6), 974-1000.
- [2]. Taylor, J. W. (1999). A quantile regression approach to estimating the distribution of multiperiod returns. *The Journal of Derivatives*, 7(1), 64-78.
- [3]. vanden Goorbergh, R. W., & Vlaar, P. J. (1999). *Value-at-Risk analysis of stock returns historical simulation, variance techniques or tail index estimation?*. De Nederlandsche Bank NV.
- [4]. Sorge, M. (2004). Stress-testing financial systems: an overview of current methodologies.
- [5]. 陈金盛. (2008). 大额交易冲击成本测算与最优变现策略选择研究.
- [6]. 姚颐, 刘志远, & 相二卫. (2012). 基金投资对股市价格压力效应的量化分析[J]. *上海金融*, 6, 013.
- [7]. 陈启欢, & 杨朝军. (2005). 证券市场流动性综合指标及计量理论. *数量经济技术经济研究*, 22(8), 114-128.
- [8]. 房振明, 王春峰, & 曹媛媛. (2005). 上海证券市场流动性模式的研究. *管理工程学报*, 19(2), 33-39.
- [9]. 孙云辉. (2007). 中国股市流动性度量指标设计与实证. *统计与决策*, (9), 17-18.
- [10]. 李娜, & 李敏. (2006). 压力测试在风险管理中的应用. *技术经济与管理研究*, (4), 66-67.
- [11]. 董天新, & 杜亚斌. (2005). 压力测试及其在金融机构风险管理中的运用. *海南金融*, (5), 36-40.
- [12]. 孙云辉. (2007). 中国股市流动性度量指标设计与实证. *统计与决策*, (9), 17-18.
- [13]. 黄中文, 刘亚娟, & 朱芳芳. (2009). 股票质押贷款业务及质押率探析. *金融理论与实践*, (9), 92-95.
- [14]. 董珊珊. (2012). 股票质押贷款质押率测算——基于EGARCH 模型的VaR方法. *中国证券期货*, 5, 038.
- [15]. 王志诚. (2003). 股票质押贷款质押率评定的VaR 方法. *金融研究*, 12, 64-71.
- [16]. 陆静, & 唐小我. (2006). 基于流动性风险的多因素定价模型及其实证研究. *中国管理科学*, 14(5), 45-51.
- [17]. 崔兰伟. (2009). 考虑冲击成本的积极投资组合决策模型研究(Master's thesis, 天津大学).
- [18]. 苏冬蔚, & 麦元勋. (2004). 流动性与资产定价: 基于我国股市资产换手率与预期收益的实证研究. *经济研究*, 2, 95-105.