

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

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：白凯敏、卞晨晔、李劲岷、刘萃

任鹏飞、柴海涛、董平、高强

2014年11月

目 录

摘要	1
一、信用业务风险度量项目综述	2
1.1、项目背景	2
1.2、研究目标	2
1.3、前期探索—多因子模型	3
1.3.1 模型概述	4
1.3.2 模型构建	4
1.3.3 模型实证研究	4
二、泡沫预警模型	6
2.1、模型理论简介	6
2.2、超指数膨胀的泡沫模型简介	7
2.2.1 SB (super-exponential bubble) 模型	8
2.2.2 JLS 模型	9
2.2.3 均值回复随机平稳临界点模型	10
2.3、模型的参数估计和泡沫的检验方法	11
2.3.1 模型的参数估计方法	12
2.3.2 泡沫的检测和警报的生成	13
2.4、泡沫模型的实证检验	14
2.4.1 数据的选取	15
2.4.2 检验结果及分析	16
三、风险损失度量—VaR	17
3.1、VaR模型简介	17
3.2、计算VaR值的方法	18
3.2.1 参数法	19
3.2.2 非参数法	19
3.3、计算VaR	21
3.4、GUI界面的实现	22

四、展望	49
附录	50
参考文献	53

摘要

一、信用业务风险度量项目综述

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。证券公司要求投资者提供一定的担保品，向投资者拆借资金买入股票或拆解股票予以卖出。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券商获利；或者是通过卖空之前买入证券而对冲风险。融资融券这种多层次交易、新兴的金融工具是我国发展证券场所急需引入的。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。融资融券业务的开展，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。第一阶段的目标是建立合适的统计模型，通过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。信用业务风险不同于一般的股票投资风险，客户提供质押物充当保证金，当维持担保比率低于一定比例将对客户强制平仓，当强制平仓后资金余额低于融资额公司才面临损失，则第二阶段的目标是根据泡沫模型对股价大跌临界点的预警对违约损失进行定量分析，即利用 VaR 模型对可能发生的风险损失进行度量，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，得出的结果有助于在业务开展前提供决策依据。

1.3、多因子模型

1.3.1 模型概述

多因子模型是关于资产定价的模型，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。多因子模型的理论背景是 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，

$\beta_{im} = \frac{Cov(r_i, r_m)}{\sqrt{Var(r_m)}\sqrt{Var(r_i)}}$ ； $E(r_m)$ 是市场投资组合 m 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 β 进行分析, 可以得出结论: 在风险资产的定价中, 那些只影响该证券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 β 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 δ_e^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, 是市场重新达到平衡。

多因子模型的基本形式为

$$\bar{K}_{it} = a_i + b_{i1}\delta_{1t} + b_{i2}\delta_{2t} + \dots + b_{ik}\delta_{kt} + e_{it}$$

式中: δ_{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。

1.3.2 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。在进行变量选择时, 可以从微观指标、流动性、宏观指标等方面进行。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票。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经验分析，根据它们对公司股价影响的重要性，确定进入股价模型的基础因素种类及其个数；

第三步，用适当的样本估计并建立基础多因素股价模型。

依据 APT 模型为理论构架，建立反应经济环境的宏观因素和公司的微观指标以及股票市场中股票流动性的实证检验模型如下：

$$R_{it} = \alpha + \beta_1 \Delta EPS_{it} + \beta_2 \Delta BVPS_{it} + \beta_3 \Delta SPS_{it} + \beta_4 BDM_{it} + \beta_5 SIZE_{it} + \beta_6 TURNOVER_{it} + \beta_7 OUTSHARE_{it} + \beta_8 ROINDEX_t + \beta_9 \Delta GDP_t + \beta_{10} \Delta FI_t + \beta_{11} \Delta P_t + \beta_{12} GB_t + \beta_{13} I_t + \epsilon_{it}$$

变量定义及说明：

信用业务风险度量研究报告

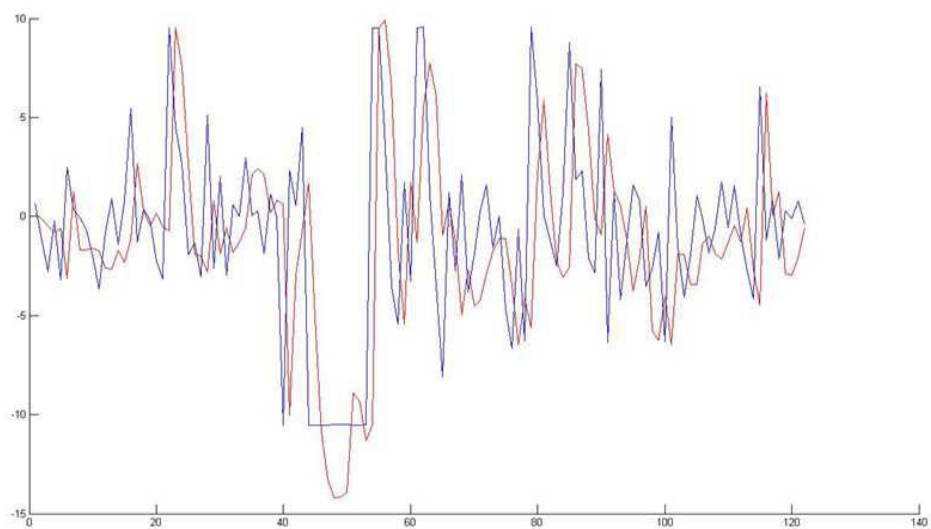
变量名	变量说明	计算方法
R_{it}	个股回报率	1) 年个股累积回报率(年内各月回报率累加) 2) 考虑送股、配股、分发现金股利的可比价格年初年末之差除以年初可比股价 $\frac{P_{it}(1+F_{it}+S_{it}) * C_{it} + D_{it} - P_{i,t-1} - S_{it} * K_{it} * C_{it}}{P_{i,t-1} + S_{it} * K_{it} * C_{it}}$
ΔEPS_{it}	每股盈余增长	$\frac{EPS_{it} - EPS_{i,t-1}}{EPS_{i,t-1}}$
$\Delta BVPS_{it}$	每股净资产增长	$\frac{BVPS_{it} - BVPS_{i,t-1}}{BVPS_{i,t-1}}$
ΔSPS_{it}	每股销售收入增长	$\frac{SPS_{it} - SPS_{i,t-1}}{SPS_{i,t-1}}$
BDM_{it}	账面市值比	$\frac{BV_{i,t-1}}{MV_{i,t-1}}$
$SIZE_{it}$	公司规模	IN(总市值)或者 IN(流通市值)
$TURNOVER_{it}$	换手率	年个股交易股数 / 流通股股数(均值)
$OUTSHARE_{it}$	流通股比例	年初年末流通股比例均值
$ROIINDEX_{it}$	指数回报率	$\frac{INDEX_{it} - INDEX_{i,t-1}}{INDEX_{i,t-1}}$
ΔGDP_{it}	GDP 增长	$\frac{GDP_{it} - GDP_{i,t-1}}{GDP_{i,t-1}}$
ΔFI_{it}	固定资产投资增长	$\frac{FI_{it} - FI_{i,t-1}}{FI_{i,t-1}}$
ΔIP_{it}	工业产值增长	$\frac{IP_{it} - IP_{i,t-1}}{IP_{i,t-1}}$
GB_{it}	利率的期限结构	$LGB_{it} - TB_{i,t-1}$
I_{it}	通货膨胀	$GPI_{it} - CPI_{i,t-1}$

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，则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，但加入高阶的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。

多因子模型研究的是股票收益率与影响股价的因素之间的关系，由于未来几天的某些自变量（如换手率）的取值并不知道，因此不能用来预测，而如果用已知的数据代替未来的数据，则会出现延迟现象，如下图：

信用业务风险度量研究报告



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相差大约一天,即对明天收益率的预测大约等于今天的收益率。

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文体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，市场中的代表性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)检验、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, Porter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评断泡沫的存在性。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与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-0 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魔心，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) 传统的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告诉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：它要求用月或者周收益率来检测（消除噪声），因此对 3 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。

正对 Gurkaynak 的评论，我们认为，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——炒至熟膨胀 (super-exponential growth)。与允许无线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，典型的正反馈包括可来自于理性因素。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 SB 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 模型已被用于与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 模型，并发现泡沫的破灭时点与模型预估的时点能较好的匹配。另一方面，我们对传统的 JLS、模型做出重大改进，利用基于行为的随机折现因子来表述。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，本文作者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 SB 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”，并且可以通过估计表征正反馈强度的指数来识别泡沫的大小。同时，估计得到的潜在临界时间大小还可以用酒精泡沫的破灭。

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线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，破灭将发生在此之前的某个时刻。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终结模型，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 SB (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， κ 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 ($j = 0$) 是 SB 模型可视为是对几何布朗运动的推广，

因为当 μ, σ 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 μ ，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，同行 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 均值回复随机平稳临界点模型

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 0-U 过程，满足均值回复平稳的性质：

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

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。因此 Black-Scholes 股价方程对应的是超指数模型没有泡沫时的特殊情况。易见， $T_c \gg 0, m \rightarrow 1$ 且 $\alpha \rightarrow 0$ 时

$$r_\tau(t) = \ln p(t+\tau) - \ln p(t) = -\beta \ln\left(1 + \frac{\Delta \tilde{t}_c - \tau}{T_c + \Delta \tilde{t}_c - \tau}\right) \approx \frac{1}{(m-1)(T_c + \Delta \tilde{t}_c - \tau)} (\tau - \Delta \tilde{t}_c) \sim N(\mu, \sigma_{t_c})$$

可见，对数收益率渐近地成为一个独立的正态随机变量，这恰好对应了没有出现泡沫时有效市场的情形。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方法

2.3.1 模型的参数估计方法

平稳随机终结模型中虽然有三个待估参数： K, β, 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

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

对 $\ln K$ 和 β 的估计是通过如下的优化算法得到的：由于就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 β 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为的临界时间越接近，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 β^* 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。入股时间序列 \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 泡沫模型的实证检验

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可的临界时间越接近，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）。

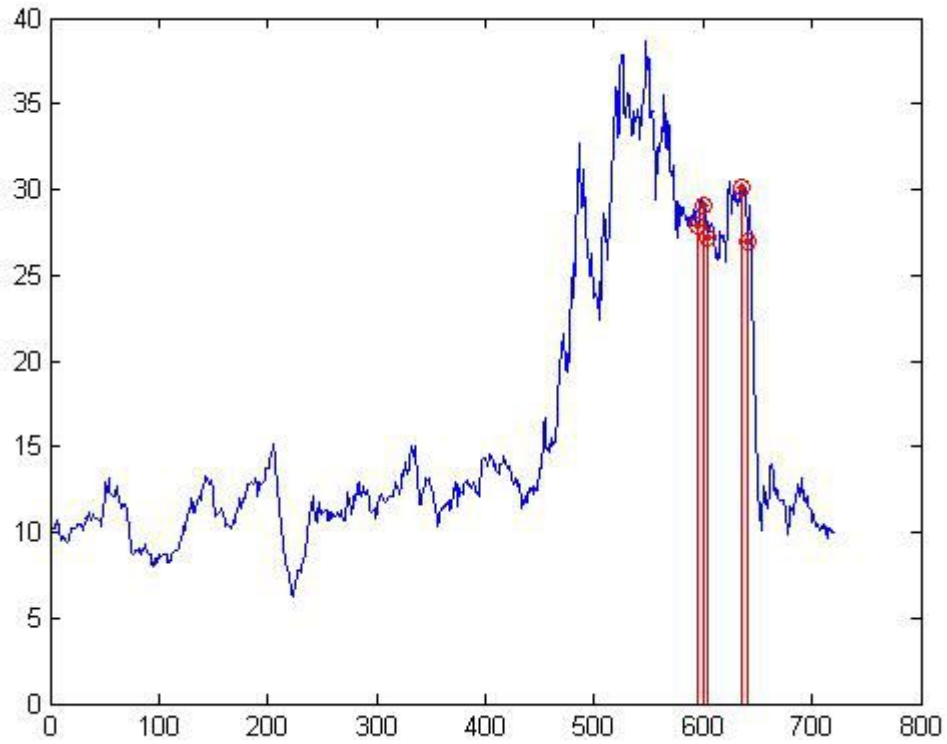
酒鬼酒(2012年11月27日)：3天跌20%负面新闻，塑化剂超标，公司停产整顿。

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

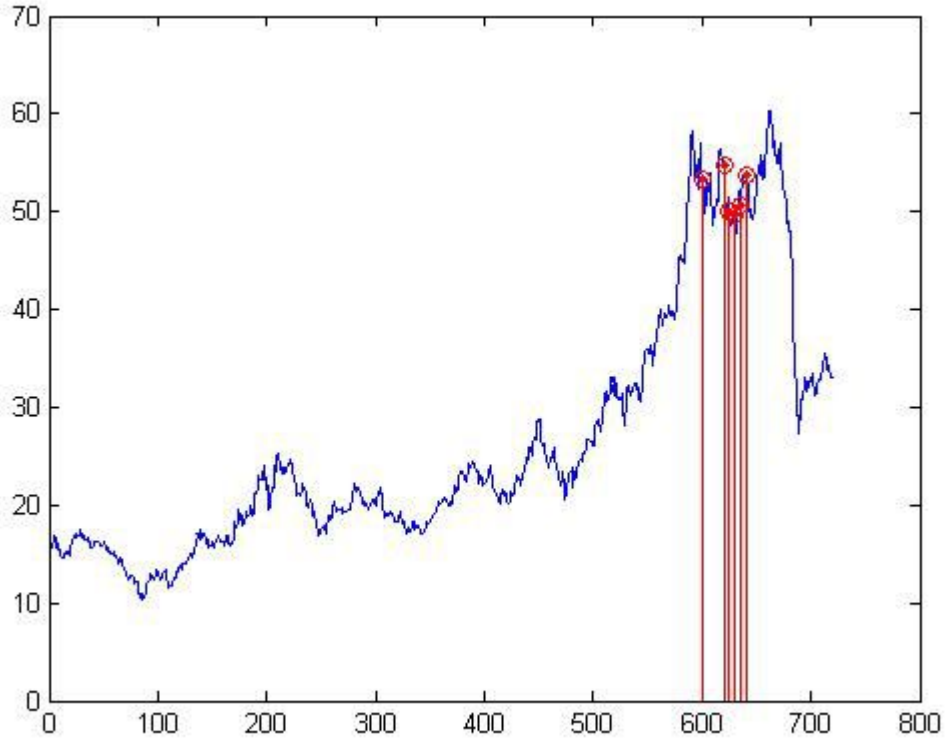
图 1 昌九生化泡沫预警图



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

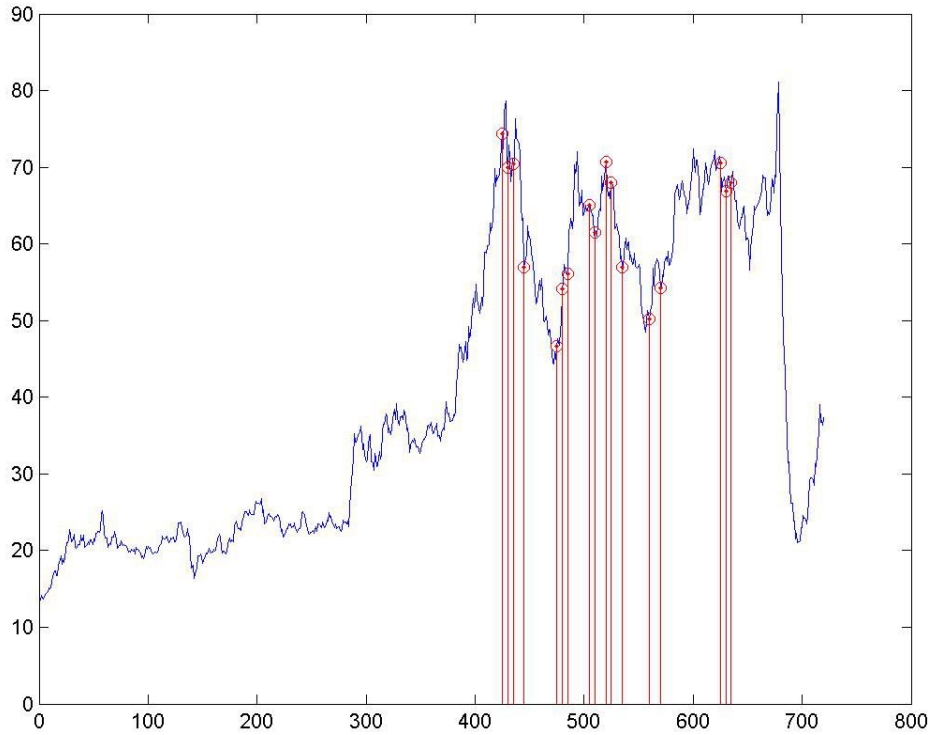
图 2 酒鬼酒泡沫预警图



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

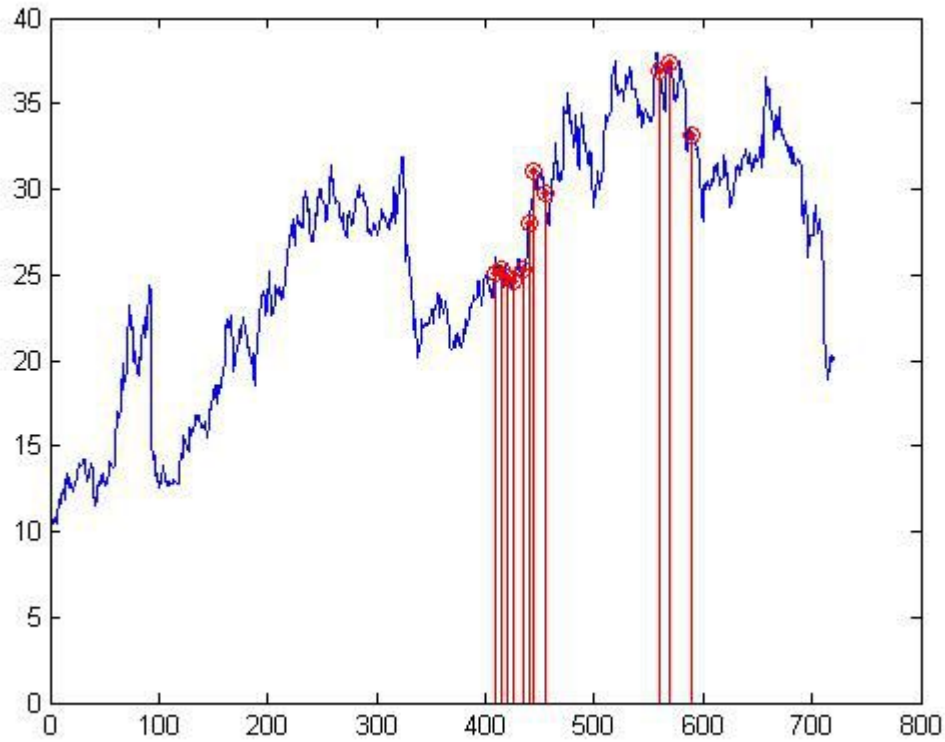
图 3 重庆啤酒泡沫预警图



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，从而可以是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

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，较传统方法而言，具有强大的优势，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的显著性水平下，投资者的资产最大损失值，这一指标更具有合理性好科学性，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 方法的应用逐渐设计到证券公司、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 就是某也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：

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

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。

3.2、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。

3.2.1、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就是移动平均法和 GARCH 法。

(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：

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

这里 R_t 表示金融资产第 t 期的回报， μ 表示 T 期内的回报均值， σ_t^2 表示回报序列的标准差， $\hat{\sigma}_t^2$ 是 σ_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。

(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 λ 来表示这个权重, 参数 λ 也称为衰减因子(deeayfaetor)。

估计式如下:

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

整理可得

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

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 的形 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 λ , 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 的形

skmetrios 经过对 λ 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 λ 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入值不同,所以在 Risktuetries 中,所有资产都采用同一值入是不正确的。

(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。

3.2.2、非参数法

(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​​来损益分布,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 Δp_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 Δp_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) 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 Δt 内, $\Delta S_t = S_{t-1}(\mu \Delta t + \sigma \epsilon \Delta t)$, $\epsilon \sim N(0, \sigma_t^2)$

σ_t 用窗口数据的历史波动率表示, μ_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 ϵ 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。

(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。

(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。

(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 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 $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ 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:

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

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

(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

$$p = F(x) = \int_{-\infty}^{x_p} \hat{f}(x) dx$$

$$1-p = 1 - F(x) = 1 - \int_{-\infty}^{x_p} \hat{f}(x) dx = \int_{x_p}^{+\infty} \hat{f}(x) dx$$

$$= \int_{x_p}^{+\infty} \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}} dx = \frac{1}{n\sqrt{2\pi}} \sum_{i=1}^n \int_{x_p}^{+\infty} e^{-\frac{(x-x_i)^2}{2h^2}} d\frac{x-x_i}{h}$$

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。

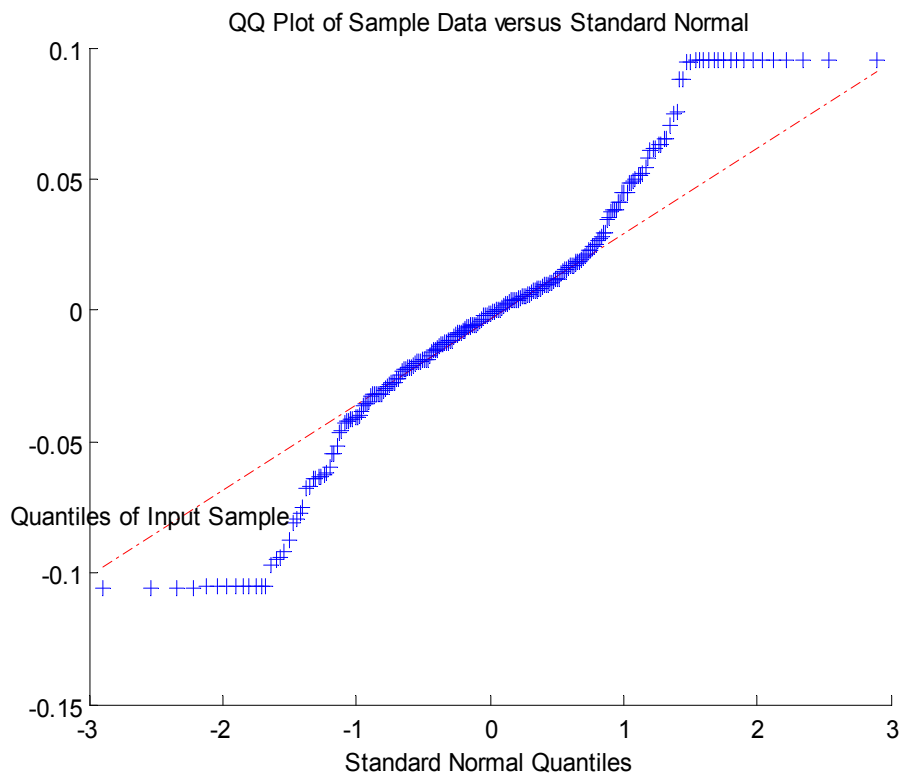
3.3、计算 VaR

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：

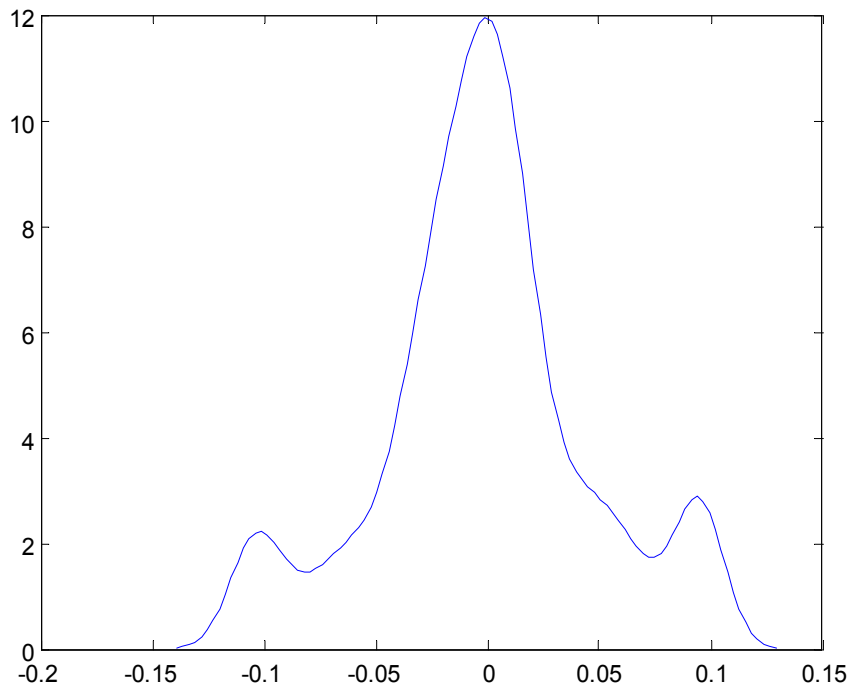
	mean	std	skewness	kurtosis
旭光股份	-0.00065	0.035158	-0.33078	4.574035
蓝帆股份	-0.00015	0.036734	-0.18366	4.247648
南岭风暴	-0.00128	0.030059	-0.16797	5.526263
南方食品	-0.00128	0.030059	-0.16797	5.526263
ST贤成	6.09E-05	0.028419	-0.45391	5.095591
如意集团	0.000414	0.036893	-0.02346	3.950466
酒鬼酒	-0.00111	0.029771	-0.40726	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的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，因此不适合用参数的方法去计算 VaR 值，因为基本假设不满足。因此，我们就采用非参方法进行计算，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 `ksdenstiy` 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：

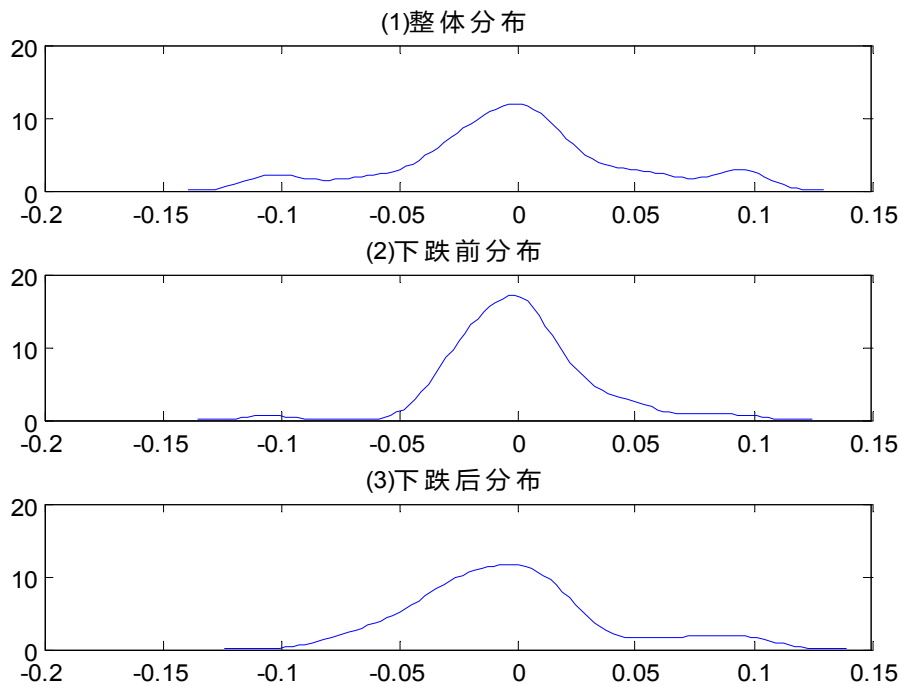


因为股票价格有涨跌停的限制，所以收益率的数据不会超过 0.1，而核密度估计对估计进行了平滑处理，因此得到密度函数估计图像在 0.1 处有凸起部分，当然另外一个原因也是厚尾现象的影响。

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			
mean	std	skewness	kurtosis
0.000325	0.028877	0.202445	6.200169
下跌后			
mean	std	skewness	kurtosis
-0.003473	0.039194	0.760729	3.67184

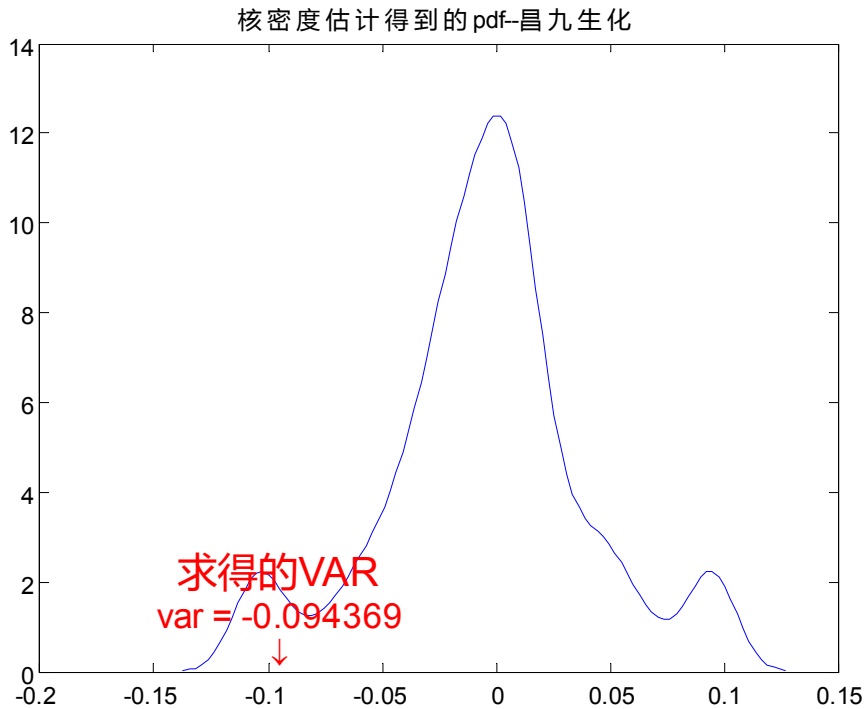
可以看出，下跌之前和下跌之后均值发生微小变化，但是偏度和峰度却发生了十分显著地变化。更进一步，对下跌前和下跌后的收益率数据用核密度估计方法，拟合的结果如下图：



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 的估算，因为这些数据包含了比较多的下跌比较多的数据，所以估计得到的密度分布函数厚尾行 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。实际中，可以根据当前行情进行主观的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，得到的 VaR 值如下表所示：

旭光股份	-0.0647	如意集团	-0.053
蓝帆股份	-0.0619	酒鬼酒	-0.0539
南岭风暴	-0.0507	昌九生化	-0.0944
南方食品	-0.0499	百圆裤业	-0.0417
ST贤成	-0.0598	洪都航空	-0.049

3.4、GUI 界面的实现