中国大盘+中小板股指聚集性、非对称性与VaR分析及预测、预警

——基于Arma-EGARCH模型的沪深300+中小板2007-2018.12日度指数分析

**课程名称：金融时间序列**

**任课老师：赵 冬 华**

**姓 名：吴 婕**

**学 号： 18210180067**

**摘要：**为防范股票市场上的不确定性和风险，有效地度量股票指数收益率的波动性显得尤为重要。而研究、测量股票市场上的不确定性和风险有助于规避股票市场上的潜在风险，帮助投资者合理投资，增强股市的合理性、抗风险性。因此，本课题借鉴国内外已有的研究文献，运用GARCH族模型拟合了股票指数收益率的波动性方程，实证研究了沪深300指数，和中小板指数。时间节点从2007.12.1持续到2018.12.1，主要是为了囊括进2008和2015年的两次股灾，以及2018年股市寒冬，直觉上，这是波动率非常大、存在集聚效应的时点。本文用R语言，进行了Arma-EGARCH模型的分析，包括计算股票指数的收益率，实现收益率的可视化 ，计算一些基本统计量，绘制股指收益率的ACF和PACF图，检验收益率序列的ARCH效应，估计Arma-EGARCH模型以及标准化残差分析等。此外，本文尝试使用EGARCH模型对股票收益率序列进行滚动预测，并基于VaR曲线提出了股市危机预警信号。结果表明两种指数收益率均有聚集性、持续性，以及不管是A股市场还是中小板市场存在着冲击的非对称性。具体而言，中小板市场对利好利空消息更敏感，体现市场抗风险的薄弱性。

**关键词：GARCH    波动聚集   非对称性   滚动预测     VaR   预警**

# 一、引言

股票至今已有将近400年的历史，它伴随着股份公司的出现而出现。随着企业经营规模扩大与资本需求不足要求一种方式来让公司获得大量的资本金。于是产生了以股份公司形态出现的，股东共同出资经营的企业组织。股份公司的变化和发展产生了股票形态的融资活动；股票融资的发展产生了股票交易的需求；股票的交易需求促成了股票市场的形成和发展；而股票市场的发展最 终又促进了股票融资活动和股份公司的完善和发展。股票最早出现于资本主义国家。

改革开放以来，中国大力发展社会主义市场经济，自1980年，我国第一次股票发行以来，已经有37年的历史，这段历史大致可以分为三个时期，即1980至1991年的初步发展时期，1998年至今的规范调整时期。股市经过这么多年的发展，与我国的社会主义市场经济的发展形成了良性互动，不断发展壮大，不断完善。股票市场是金融市场的重要组成部分，与国民经济发展密切相关。股票市场在促进资本集中，提高企业资本的有机构成，推动经济持续健康发展，优化资源配置等方面起着极其重要的作用。

然而，我国股市的发展中仍然存在着许多的问题，例如波动性较大，相关法律法规仍需进一步完善。在2008年和2015年，我国发生了两次股灾，以及2018年股市不振。根据股市成长阶段论，股市的发展道路虽然不完全一样，但是一般都要经历5个阶段：休眠阶段、操纵阶段投资阶段、崩溃阶段、成熟阶段。沪深三百指数选取的沪深两市A 股中规模大，流动性好，最具有代表性的300支股票构成，可以综合反映沪深两市的整体表现，具有作为投资业绩的评价标准的功能，为指数化投资以及指数衍生品的创新提供了基本条件；中小板块即中小企业板，2004年5月，经国务院批准，中国证监会批复同意深圳证券交易所在主板市场内设立中小企业板块，是指流通盘大约1亿以下的创业板块，是创业板的一种过渡。所以，对沪深300与中小板进行波动性检测与预测，同时对比两种市场的特点异同，对预测A 股市场的整体走势，规避投资风险，具有重要意义。

# 二、文献综述

在《基于EVT的ARMA-EGARCH-M模型VaR研究》中，作者结合极值理论和ARMA-GARCH类模型来捕捉金融资产收益分布的有偏性、厚尾性，自相关性和波动的异方差性等特征，计算了VaR值来度量金融资产收益的风险。

本文在参考以上观点的基础上，研究思路方面，首先，对沪深300以及中小板日度指数、对数化收益率进行了平稳性，在确定他们是平稳的以后，在AIC信息准则下建立了ARIMA模型，并且对模型进行了检验与预测。接着对模型检验arch效应，发现了存在着很明显的arch效应，于是建立了arma-eGARCH模型，对两个模型进行了比较，考察了非对称性对股市的影响，最终利用模型对中国股市进行了有效的预测与预警，预测结果较Arima模型有所改进。

# 三、数据描述

## （一）数据来源

本课题选取了2007.12.1至2018.12.1沪深300和中小板的日度指数，数据来源于数据库——Wind金融终端，从时间维度上看，囊括了2008和2015年的两次股灾。2008年牛市之前，大盘1000点，历经22个月，大盘上涨了5100点，而后暴跌进入熊市，仅花11个月，就下跌了4400点，到达底部1700点；而2015年7月，在近一个月的时间里，沪指连续下降近1500点。因而，所选取的时段内，很有可能存在波动率非常大、存在聚簇的时点。

沪深300指数，是由沪深证券交易所于2005年4月8日联合发布的反映沪深300指数编制目标和运行状况，并能够作为投资业绩的评价标准，为指数化投资和指数衍生产品创新提供基础条件。沪深300的主要特点有：1.严格的样本选择标准，定位于交易性成份指数；2.采用自由流通量为权数；3.采用分级靠档法确定成份股权重；4.样本股稳定性高，调整设置缓冲区；5.指数行业分布状况基本与市场行业分布比例一致。而中小板指数是指流通盘大约1亿以下的创业板块，是相对于主板市场而言的，有些企业的条件达不到主板市场的要求，所以只能在中小板市场上市。中小板市场是创业板的一种过渡，中小板作为我国资本市场的不可缺少的组成部分,为丰富我国股票市场结构做出了不可忽视的贡献,成为了促进了中国金融体系不断走向成熟的重要步骤；同时,中小板市场在资源配置的过程中起到了良好的调节作用,促进产业结构的平衡。

因此,对选取的2007.12.1至2018.12.1沪深300指数和中小板指数，从时间维度和内容维度上，都是有意义的。因此，对其进行分析与预测，对国资本市场的成熟发展具有重要的意义。

## （二）数据预处理与描述性统计

数据录入后，首先将2007.12.1-2018.12.1日的所有数据录入成训练集，用于数据描述性统计和建模；将2018.12.3-2018.12.14作为与预测结果对比的测试集。

将录入的训练集数据画出了沪深300指数以及中小板指数的时间序列折线图形如下：

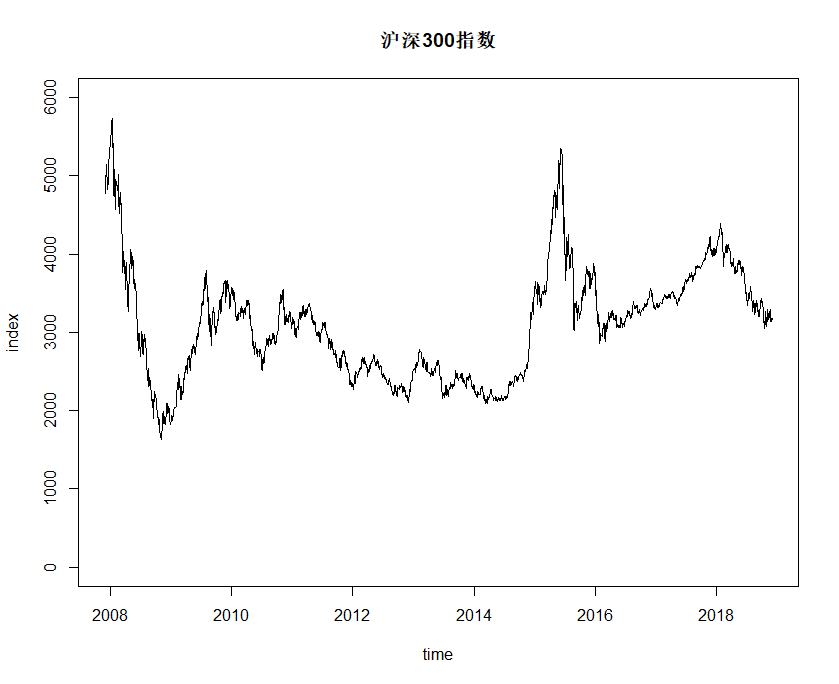


图1 沪深300日度指数

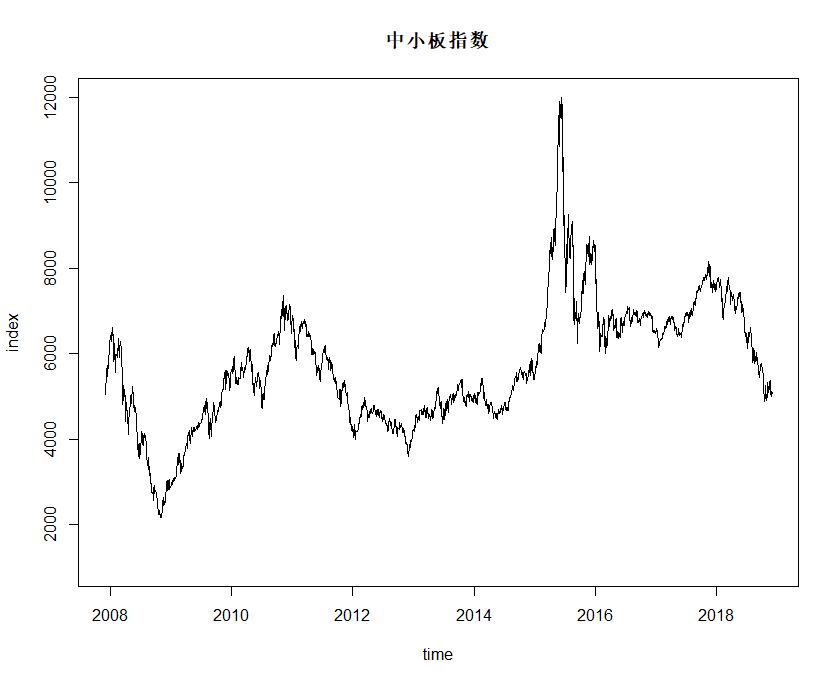


图2 中小板日度指数

通过对折线图的观察，可以直观的看出，在时间的维度上，沪深300指数在2007年至2010年期间波动较大，在2010年至2014年波动较小，而2015年以后有回复较大波动；而中小板指数仅仅在2015年至2016年期间呈现出较大的波动性，在其他时间段的波动性较小，趋势较为平稳。所以从直观的角度上，可以判断，所选取的时间区间上的沪深300指数与中小板指数具有较为明显的聚集效应。

为了进一步研究，将沪深300以及中小板日度指数转化为对数收益率进行计算

画出转化后的图像：

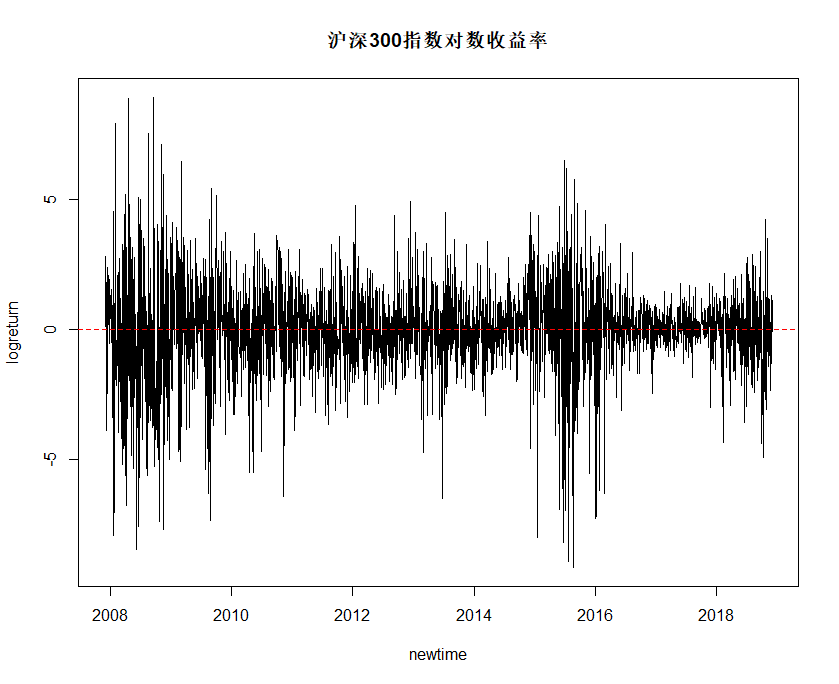


图3 沪深300对数收益率折线图

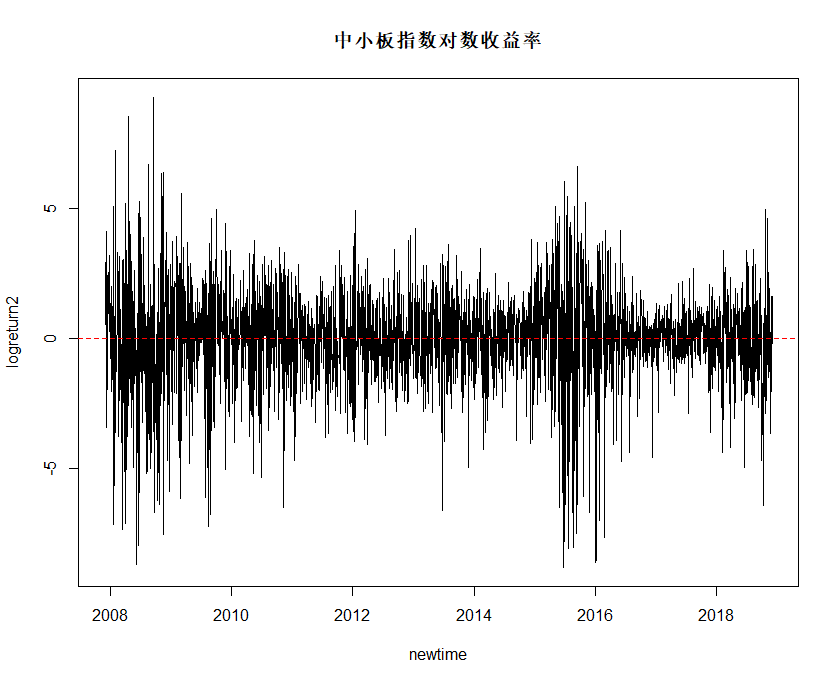


图4 中小板对数收益率折线图

通过对对数收益率的图像的观察，可以比较直观的看出，在2007年至2010年以及2015年至2016年期间，沪深300的方差与其他时间段的方差显然不同，要比其他时间段来得大，可以猜测是由于股灾引起的，18年也出现明显的波动；对于中小板指数也是类似的情况，所以不妨猜测股灾会引起股票市场波动性增大。另外一方面，仔细观察图像波动特点，不难发现，高的波动紧跟着高的波动，小的波动紧跟着小的波动，即沪深300与中小板的对数收益率具有波动集聚性。波动集聚性体现了经济周期：在持续的高波峰位置，如果该波峰是正的，则可能形成牛市，如果该波峰是负的，则可能是熊市；若处于波谷时刻，则市场比较稳定，市场风险较低，适合比较稳健的投资者进行投资。

在对数图像的基础上画出大致的波动趋势如下：

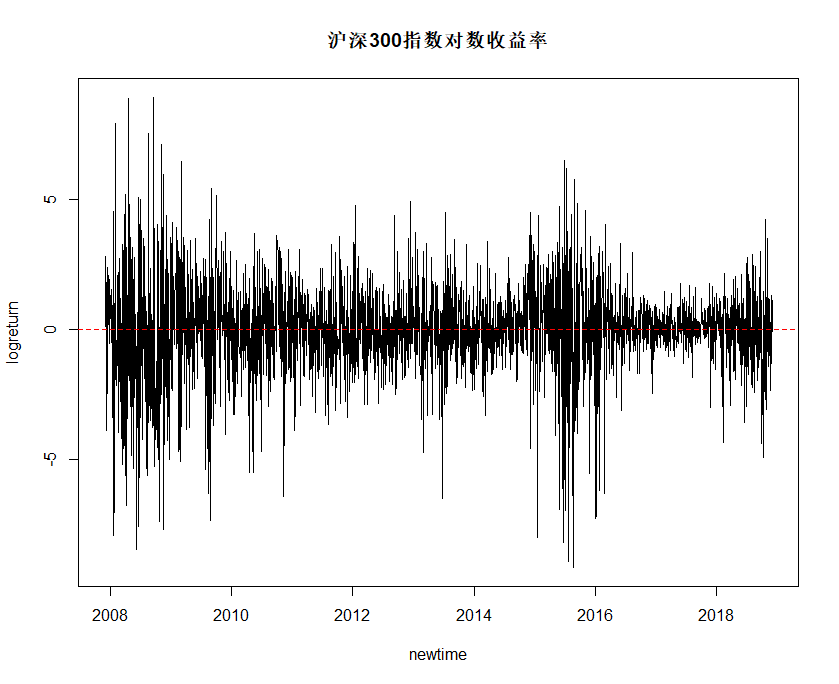


图5 沪深300对数收益率波动图

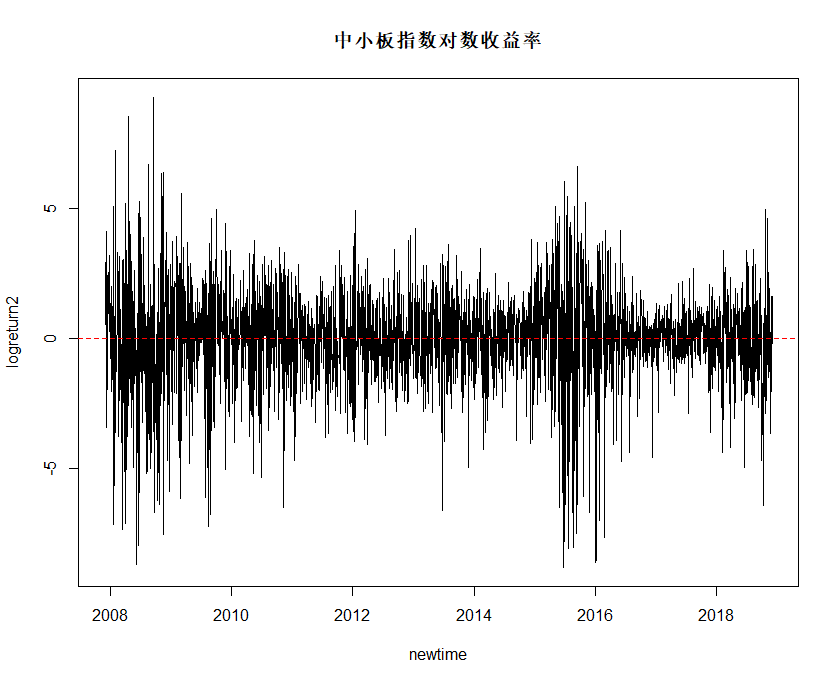


图6 中小板对数收益率波动图

观察图像，可以发现，上浪和下浪是不对称的，根据杠杆效应理论：好消息总是没有坏消息对市场的影响大。在股票价格发生下跌时，公司的净股东权益会下降但是负债不会发生改变。根据：

**杠杆率=负载/所有者权益**

所以公司的杠杆率会提高，杠杆率的提高将会引起公司信用恶化的问题，进而进一步导致公司的股票价格下降。在沪深300与中小板对数收益率图像中，可以看到，负的波浪比正的波浪要来得大，即坏消息对市场的影响比好消息要来得大。杠杆效应的存在，使得ARCH模型不足以描述市场的运行，而是需要加入方差方程调整成为新的GARCH模型。

画出沪深300与中小板收益率的分布图：

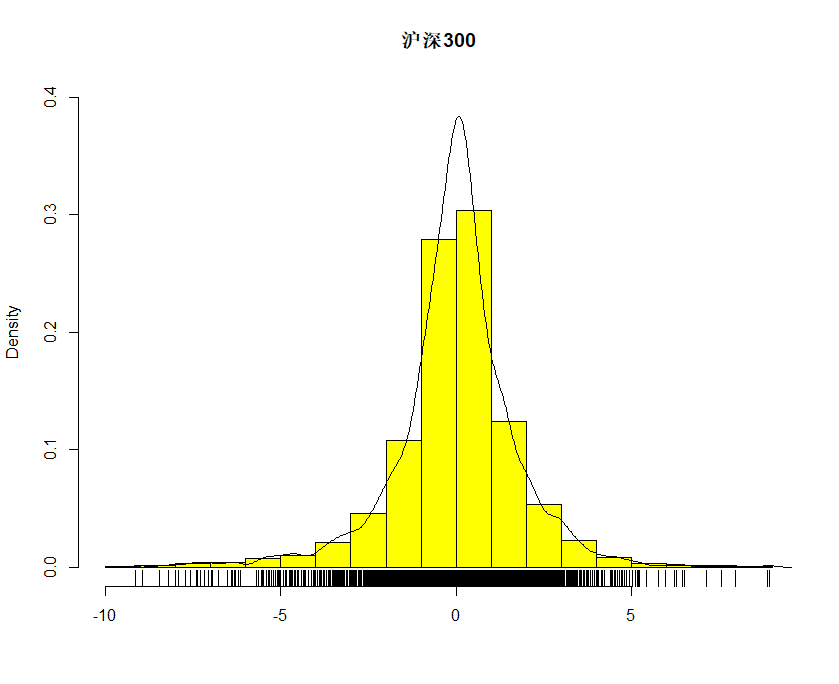


图7 沪深300对数收益率分布图

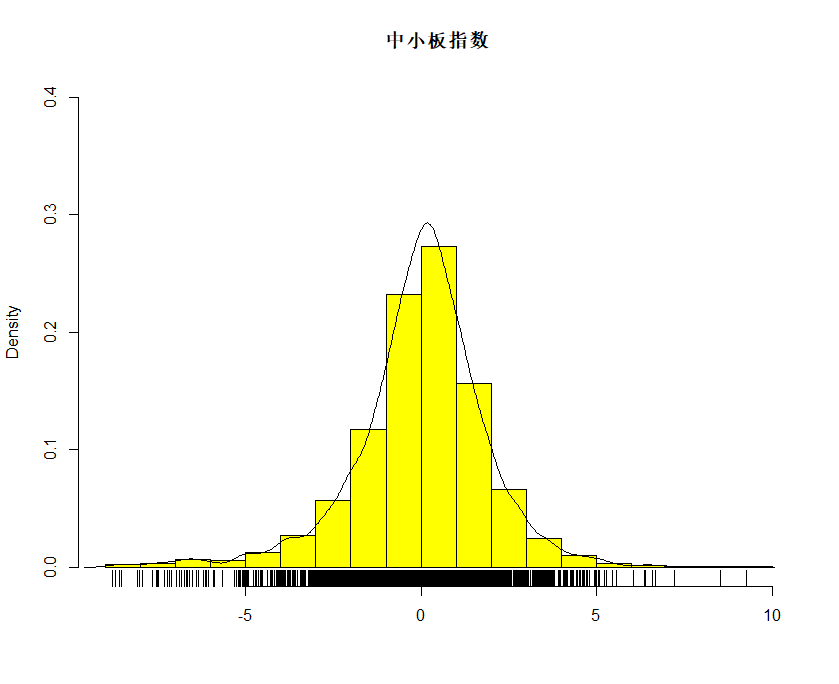


图8 中小板对数收益率分布图

观察图像可以发现，图像并非是正态分布，而是“尖峰厚尾”的t分布。可以猜测原因是：与自然界的分布不同，股市收益率时间序列的数值差异过大，波动性强，极端值较自然界的正态分布更易出现，所以在金融界更容易出现“尖峰厚尾”的t分布。收益率分布的这种性质也就导致了资产回报的两头大，中间空，市场流动性差在卖出资产时，有时甚至会遇到需要降价很多才能卖出的情况。另外，可以观察到，中小板的偏度峰度较沪深300来得更大。

利用R计算出沪深300与中小板的四阶距见下表：

**表1：收益率各阶距**

|  |  |  |
| --- | --- | --- |
|  | 沪深300对数收益率 | 中小板对数收益率 |
| 均值 | -0.01525882 | 0.0002513144 |
| 方差 | 1.755342 | 1.894895 |
| 偏度 | -0.490518 | -0.5805791 |
| 峰度 | 6.865434 | 5.734946 |

可以看出：1.两者的均值都接近于零，如果长期持有，持股者可能可以获得一个正的收益，但是收益不大；2.中小板的方差略大于沪深300的方差，可以推断中小板的风险要大于沪深300的风险；3.与中小板相比，沪深300的偏度较小，中小板比沪深300更加左偏，负的收益率更多4. 沪深300的峰度比中小板的要更大，所以沪深300的收益率更为集中。

**表2：沪深300与中小板JB正态性检验结果**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | x-squared | | df | | P-value |
| 对数收益率 | 沪深300 | 1775.6 | | 2 | | <2.2e-16 | |
| 中小板 | | 985.71 | | 2 | | <2.2e-16 |

首先对样本进行JB正态性检验，结果得到：沪深300样本的p值小于2.2e-16，中小板的p也小于2.2e-16。因此认为样本不服从正态分布。综合上述，可以猜测沪深300与中小板的对数收益率都符合t分布。

## （三）平稳性分析

平稳性是时间序列具有的一种统计特征，也是建立时间序列模型的重要前提。对于平稳的序列可以建立 AR 模型、MA 模型、ARMA 模型对其进行相应的分析和预测。因此对数据进行平稳性检验是时间序列分析的关键一步。根据限制条件的严格程度，可将时间序列的平稳性定义分为严平稳时间序列和宽平稳时间序列。由于实际生活中，很多的时间序列不仅具有随机性，而且具有非平稳性。特别是大多数经济时间序列均表现出一定的趋势性，即时间序列值随时间的变化呈少的趋势，以及季节性和方差的不稳定性。时间序列的非平稳性带有普遍性，它在客观反映经济活动现实的同时，也带来了运用数学模型描绘经济变量变化过程的困难。因此，在对时间序列建立合适地模型之前必须对时间序列做平稳性检验。

时间序列的平稳性检验主要有下列两种方法：一种是根据 ACF 图和 PACF 图的特征做出判断的图检验方法；另一种是构造检验统计量进行假设检验的方法。通常采用时间序列图和 ACF 图、PACF 图相结合来判定序列的平稳性。

进行ADF单位根检验，检验结果如下表：

**表3：沪深300与中小板平稳性检验结果**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Dickey-Fuller | Lag order | p-value | 检验结果 |
| 沪深300 | 指数 | -1.165 | 1 | 0.245 | 不平稳 |
| 对数收益率 -37.2599 | | 1 | 0.01 | 平稳 |
| 中小板 | 指数 | -0.5421 | 1 | 0.4436 | 不平稳 |
| 对数收益率 -37.2192 | | 1 | 0.01 | 平稳 |

对数收益率的数据都是平稳的，而指数都是不平稳的。

再用R画出ACF图如下：

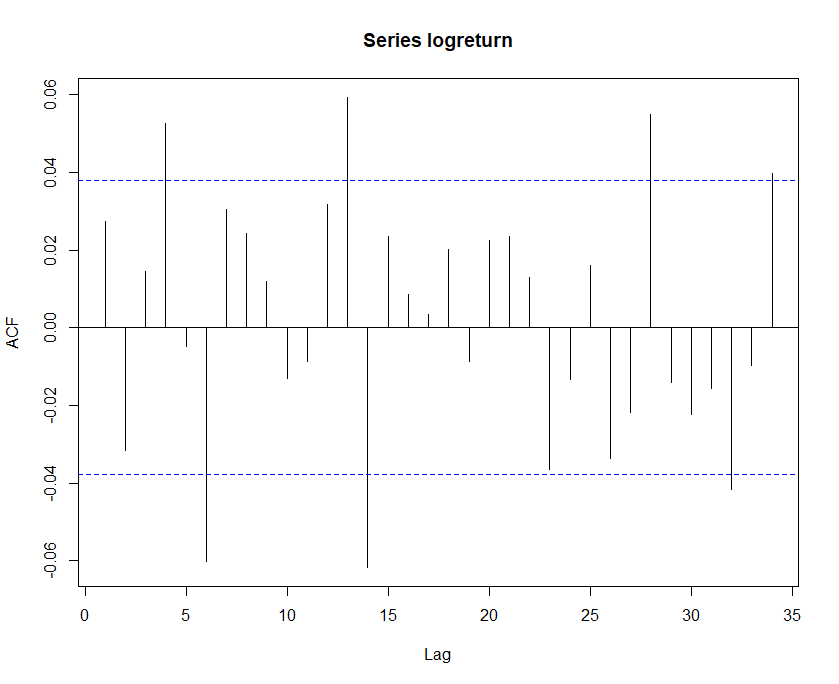


图9 沪深300 ACF图

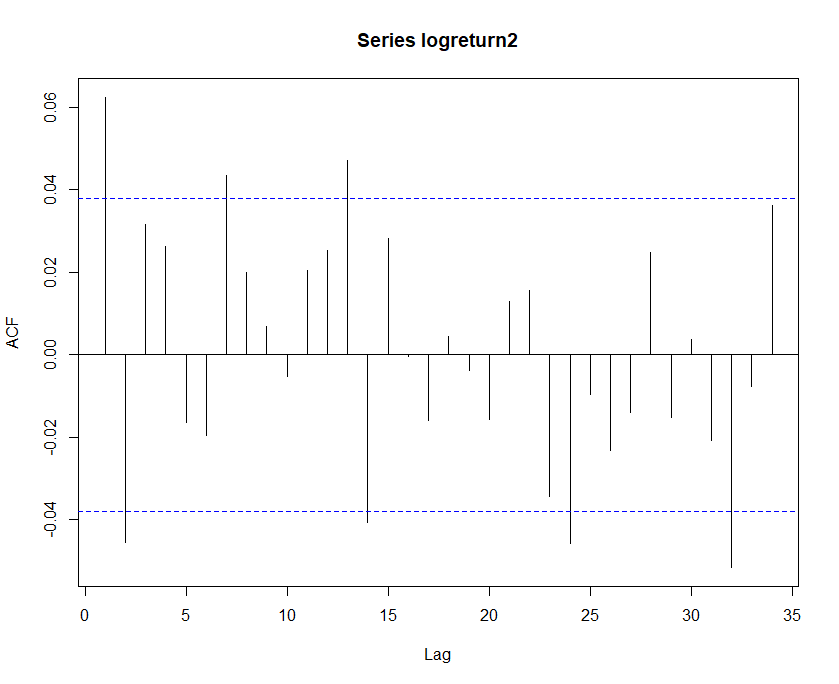


图10 中小板 ACF图

由ACF图可知沪深300与中小板均自相关，满足基本要求。

# 四、实证分析

## （一）ARIMA模型建立、检验与预测

**1.建模思路**

根据之前对数据的分析，沪深300与中小板收益率均通过了平稳性分析以及ADF单位根检验，所以考虑用ARIMA模型。ARIMA模型全称为自回归积分滑动平均模型，是由博克思(Box)和詹金斯(Jenkins)于70年代初提出一著名时间序列预测方法。在ARIMA（p，d，q）模型中，AR是自回归， p为自回归项； MA为移动平均，q为移动平均项数，d为时间序列成为平稳时所做的差分次数。

**2.实证结果**

直接利用R对模型参数进行估计，估计结果见下表：

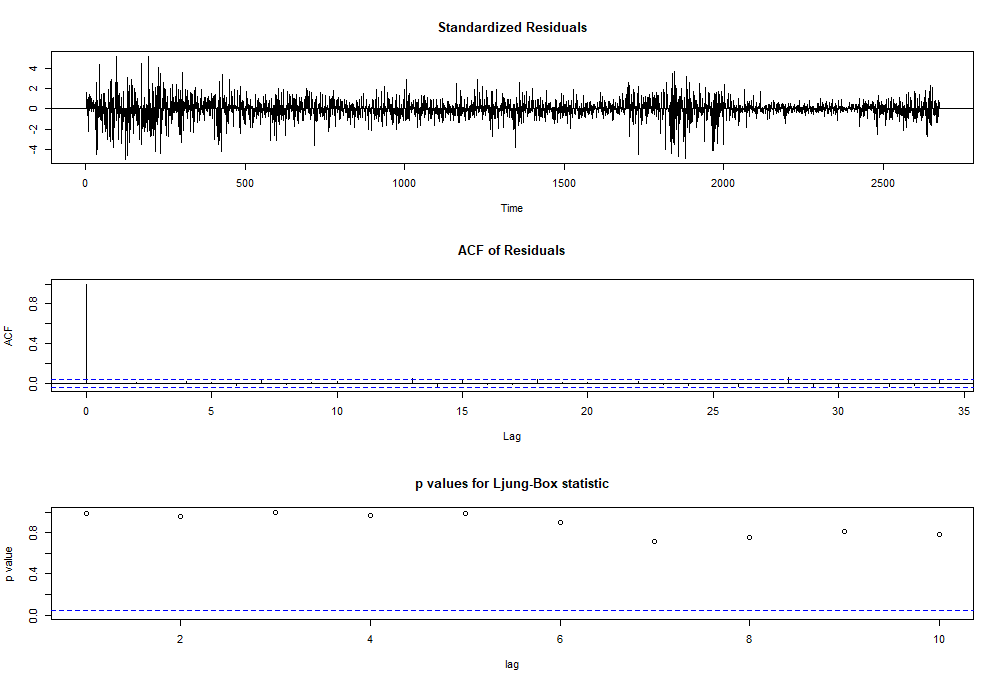
**表4：沪深300与中小板ARIMA模型参数**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 模型 | | Ar1 | Ar2 | Ar3 | Ma1 | | Ma2 | | Mean |
| 沪深300 | ARIMA(3,0,2) | Coe | 0.2109 | -0.9443 | 0.0500 | -0.1829 | 0.9106 | | -0.0161 | |
| s.e | 0.0383 | 0.0319 | 0.0204 | 0.0334 | | 0.0392 | | 0.0346 |
|  |  |  | Ar1 | Ar2 | Ma1 | Ma2 | | Ma3 | | Mean |
| 中小板 | ARIMA(2,0,3) | Coe | 0.1820 | -0.8426 | -0.1136 | 0.8040 | | 0.0937 | | 0.0001 |
| s.e | 0.0672 | 0.1787 | 0.0690 | 0.1908 | | 0.0204 | | 0.0391 |
|  | 模型 |  | Sigma^2 | Log likelihood | AIC | AICC | | BIC | | |
| 沪深300 | ARIMA(3,0,2) |  | 3.052 | -5287.17 | 10588.33 | 10588.37 | | 10629.58 | | |
| 中小板 | ARIMA(2,0,3) |  | 3.56 | -5493.28 | 11000.04 | 11000.08 | | 11041.28 | | |

表4的模型是根据最小AIC原则确定的。

**3.模型检验**

分别画出沪深300与中小板的ARIMA模型残差图



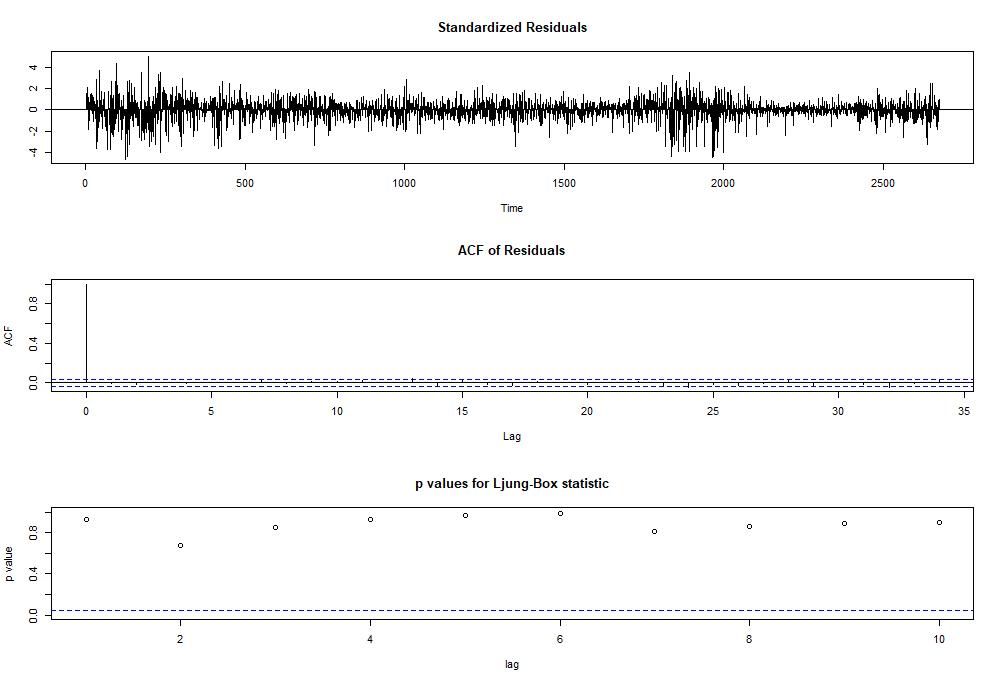


图11 残差检验图

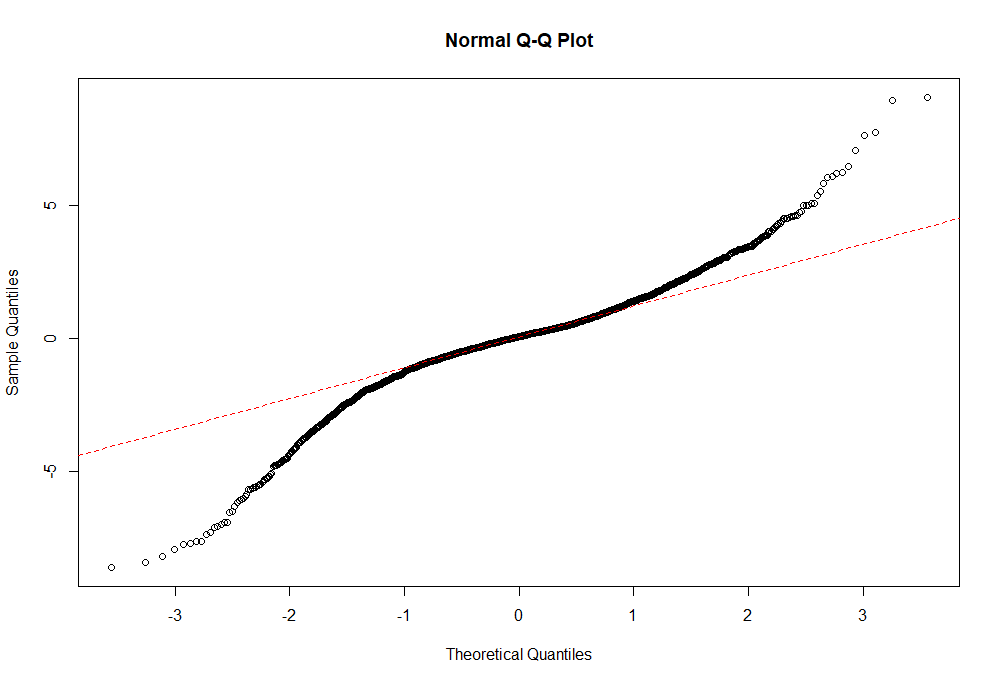


图12 沪深300残差Q-Q图

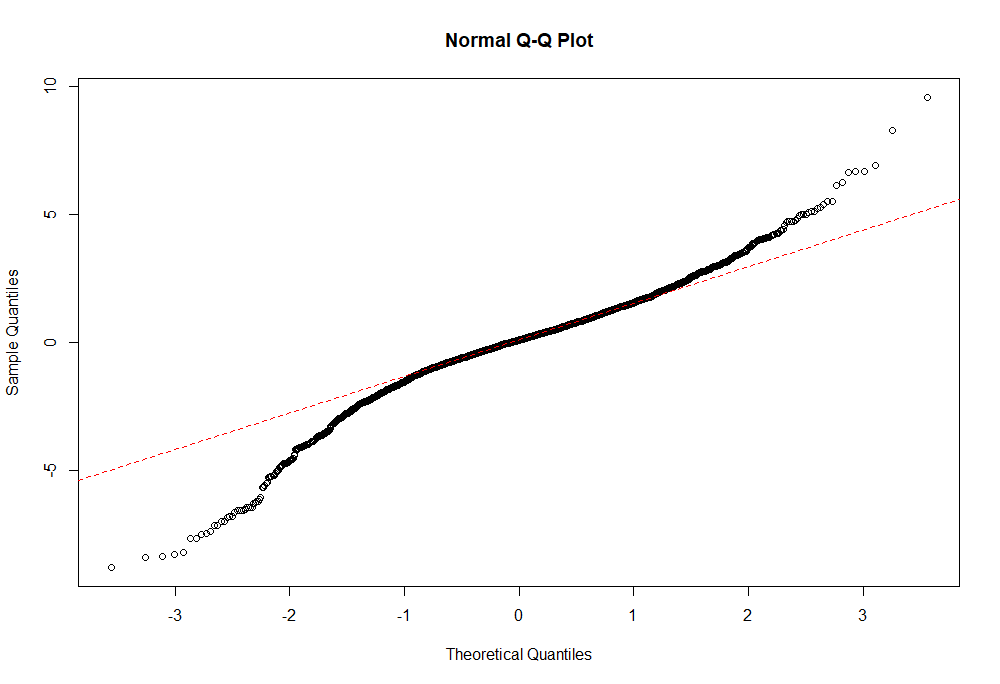


图13 中小板残差Q-Q图

观察残差图与Q-Q图可以发现，在08年左右有很多CI之外的异常点，可以猜测是由于08年的股灾引起了集聚效应严重。在集聚效应十分严重的情况下，ARCH模型分布的尾部相比于正态分布尾部更厚，主要用来解决时间序列的波动性问题，即其‘扰动’比高斯白噪声序列更容易产生异常值，这与实证研究中资产收益率表现出的异常值多于正态分布产生的异常值的结论一致，而且ARCH模型能够很好地刻画波动率聚集效应。因而，进一步检验Arch效应。

**4.模型预测**

利用arma模型进行10期预测，并且与测试集的10期真实数据进行比较，结果如下：

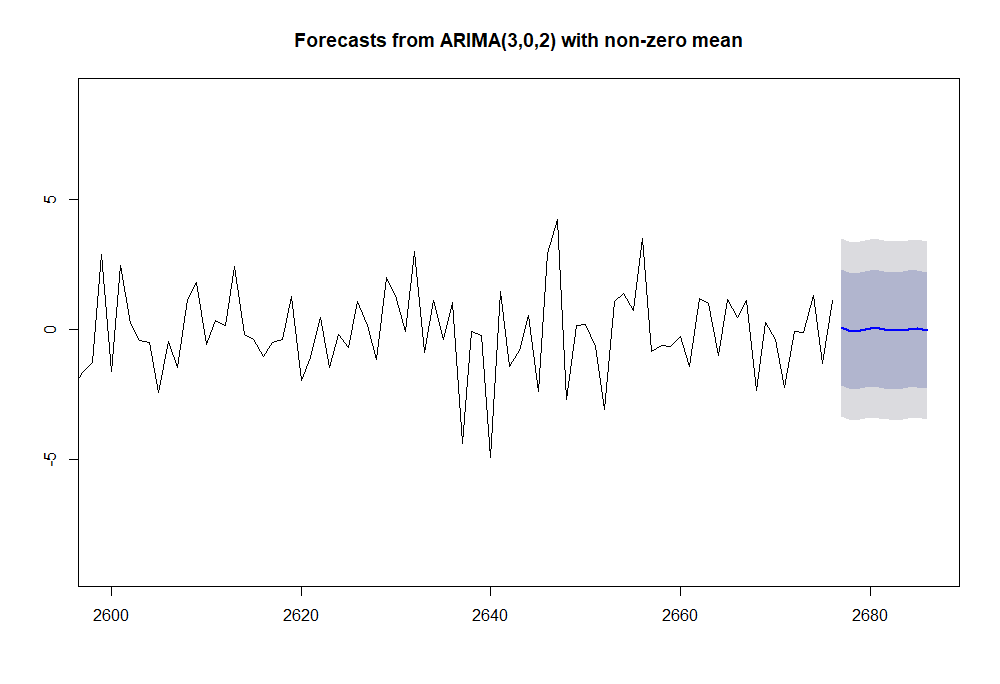


图14 arma对沪深300未来10期预测图

10期真实值数据如下：



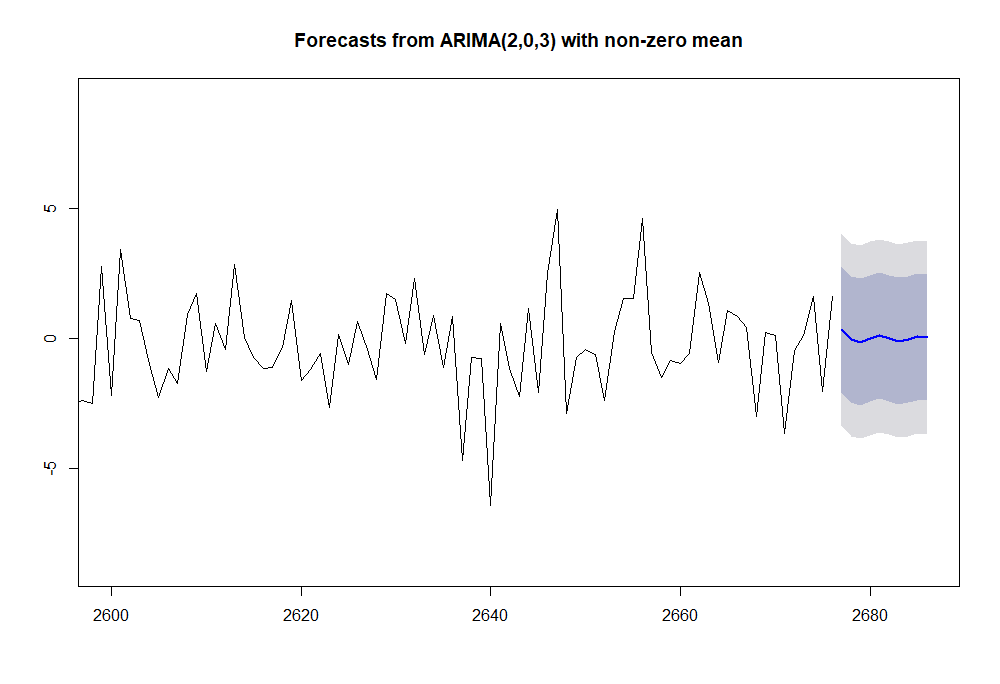


图15 arma对中小板未来10期预测图

10期真实数据



可以看到虽然实际值还是位于预测置信区间内，但预测的点估计和实际还是差别非常大的。

## （二）ARCH效应检验

为了检验arch效应，对残差做Ljung-box及Mc.Li.test分析。结果见下表：

**表5：沪深300与中小板arch效应Ljung-box检验结果**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | x-squared | | df | | P-value |
| 对数收益率 | 沪深300 | 882.53 | 12 | | <2.2e-16 | |
| 中小板 | 838.26 | | 12 | | <2.2e-16 |

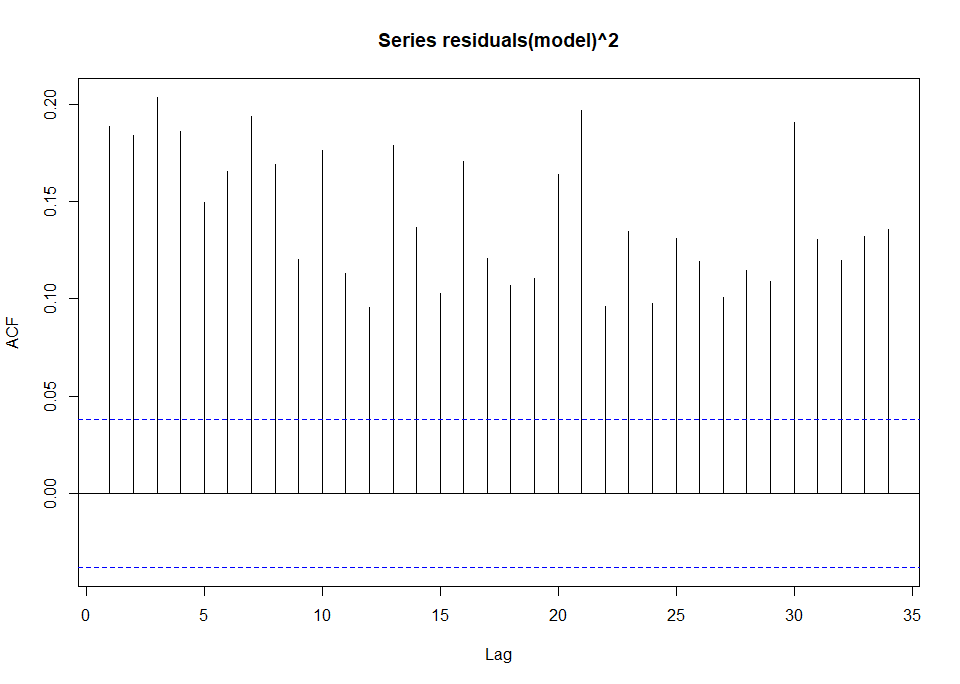


图16 沪深300 lag为12的Ljung-box检验图

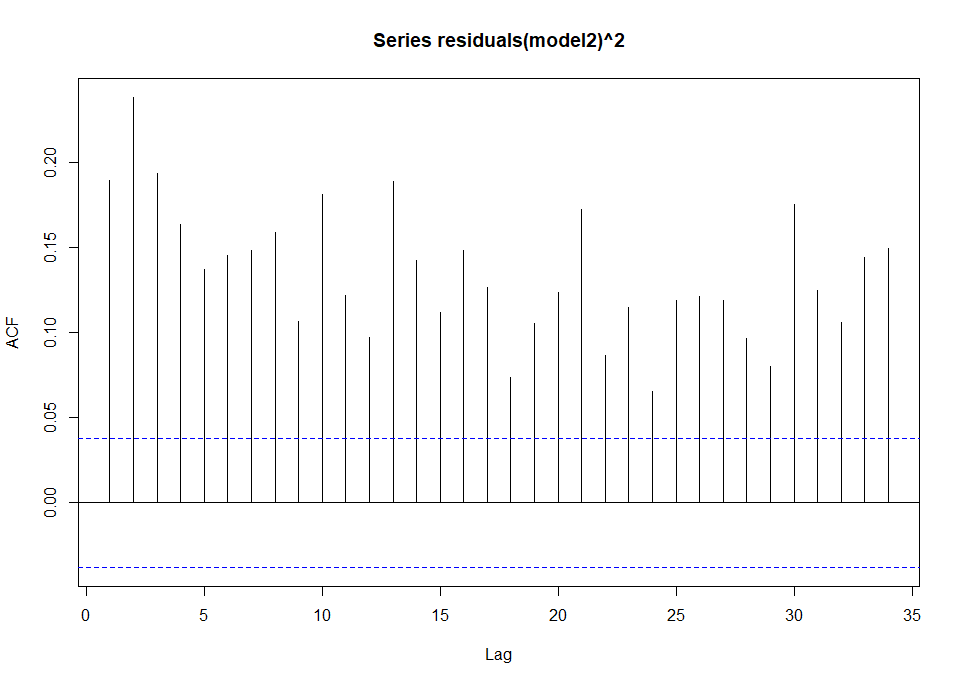


图17 中小板 lag为12的Ljung-box检验图

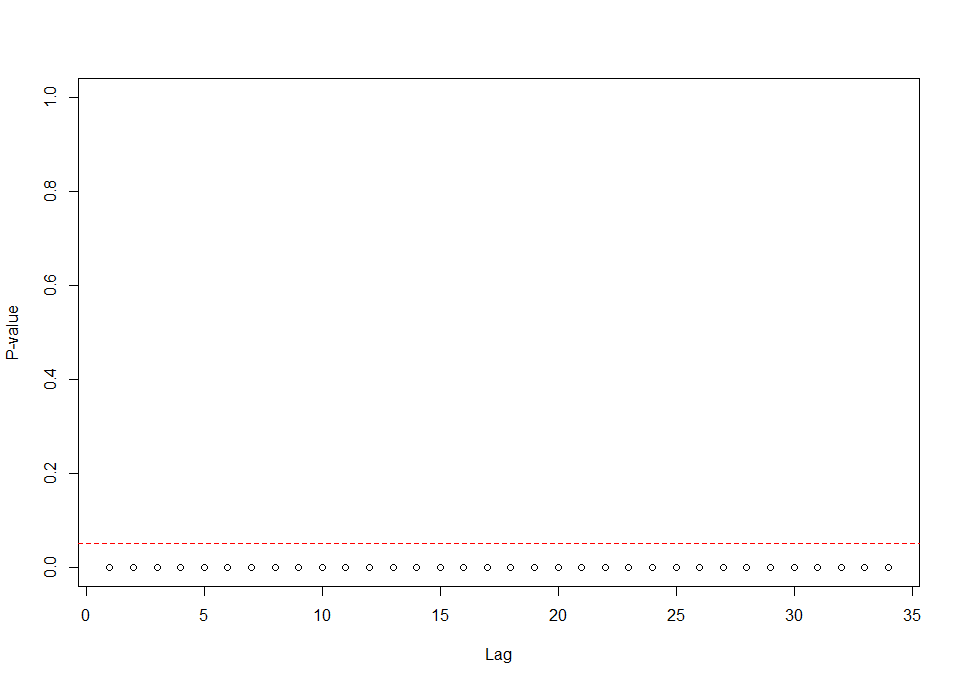


图18 沪深300 Mc.Li.test图

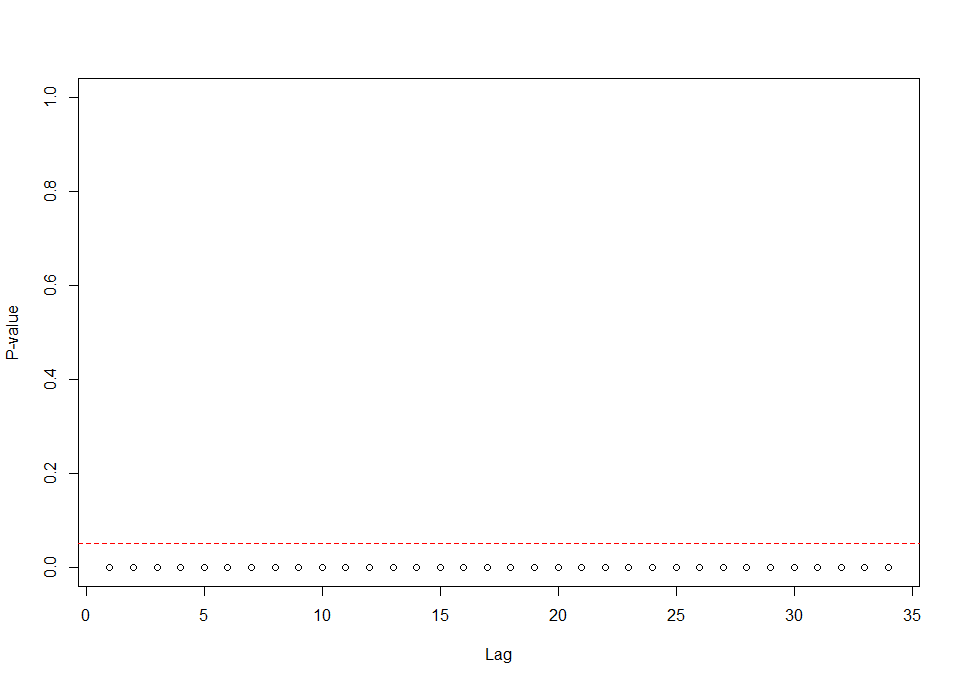


图19 中小板 Mc.Li.test图

由于Ljung-box的p值小于2.2e-16，以及MC.LI.test的点都处于红线以下，因此发现两个市场皆具有十分显著的arch效应，即波动聚集性对模型的影响十分的大，进一步说明了中国的大盘和中小板都具有明显的聚集效应。所以，转向Arma-GARCH的建立。

## （三）Arma-EGARCH模型建立、检验与预测

**1.建模思路**

由于之前得到了的结果具有显著的ARCH效应，所以考虑选用GARCH(p,q)族模型, ARCH模型全称“自回归条件异方差模型”，解决了传统的计量经济学对时间序列变量的第二个假设（方差恒定）所引起的问题。GARCH模型称为广义ARCH模型，是ARCH模型的拓展，由Bollerslev(1986)发展起来的。在GARCH之后又衍生GARCH族。GARCH族模型如下：

(a)

(b)

其中，当＝0,用b式时为Arma-egarch模型。

**2.实证结果**

因为本文不仅要研究聚集性，还要研究两个股市的非对称性，考虑到garch不能衡量正负随机扰动对下一期方差造成的非对称影响，本文最后考虑用到的是egarch模型。同时，这里假设是随机扰动项满并非足正态分布，而是满足skew T分布。对沪深300与中小板，分别利用R进行参数估计，在AIC准则下对比多次模型，最终选择egarch(1,1)（b式）与arma(1,1),得到arma-egarch模型。如以下结果：

**沪深300：**

Arma(1,1):

EGARCH(1,1):)

Skew T分布的参数为：

skew=0.93777 shape=5.251389

**中小板：**

Arma(1,1):

EGARCH(1,1):)

Skew T分布的参数为：

skew=0.7425 shape=6.960587

**3.模型检验**

下面对残差和残差平方进行acf检验，得到以下结果：

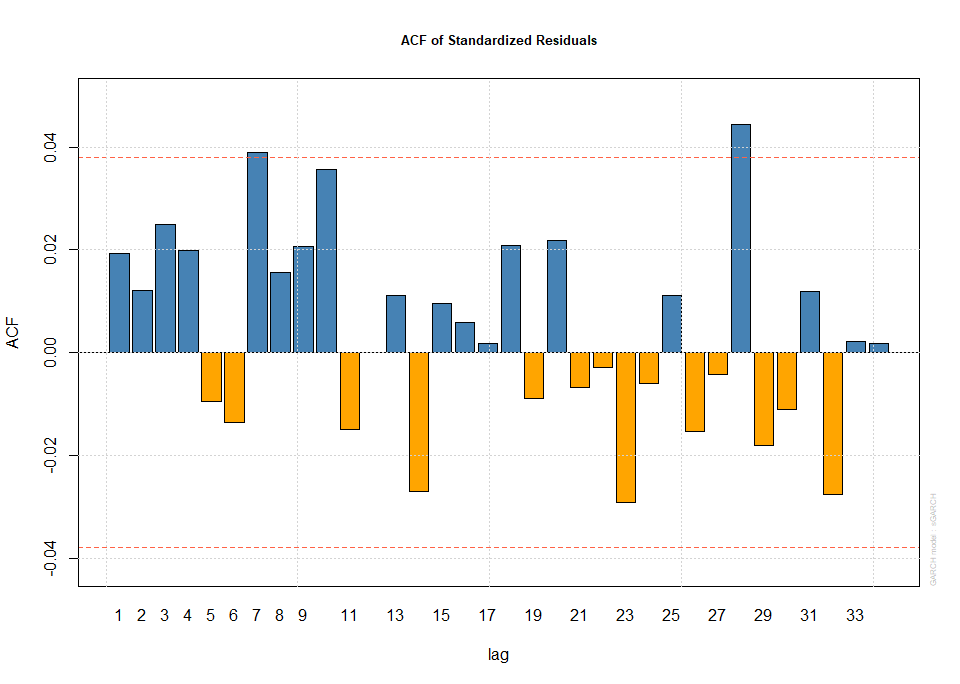


图20 沪深300 残差检验图

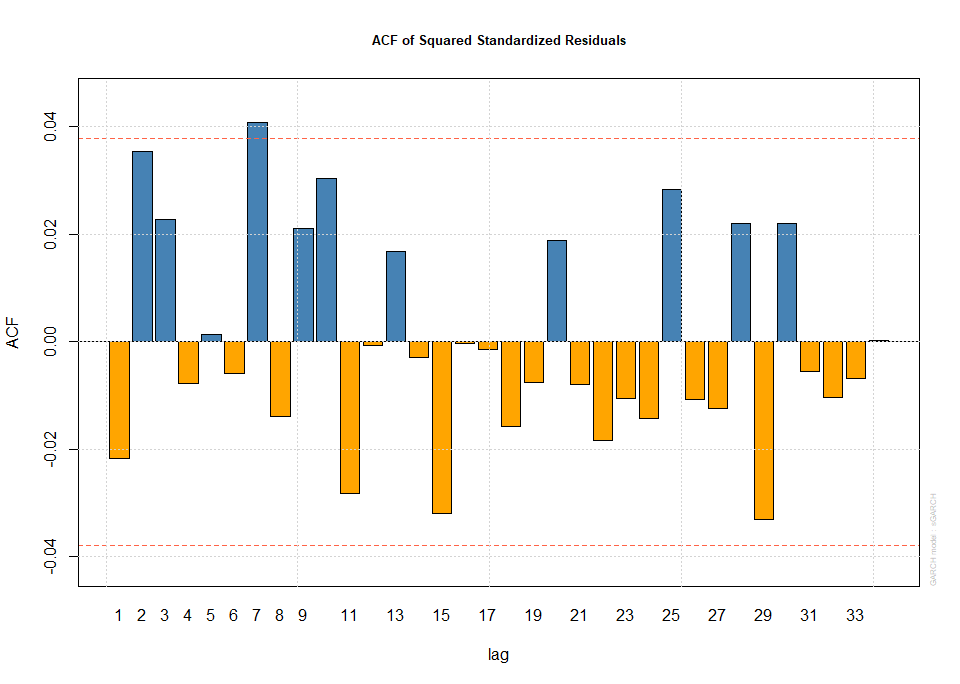


图21 沪深300 残差平方检验图

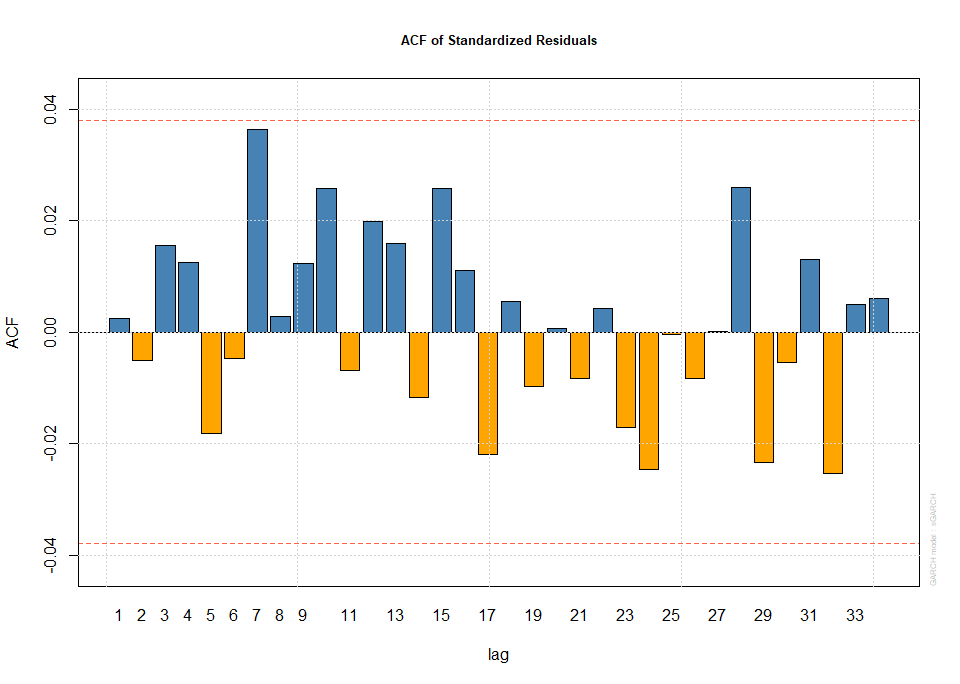


图22 中小板 残差检验图

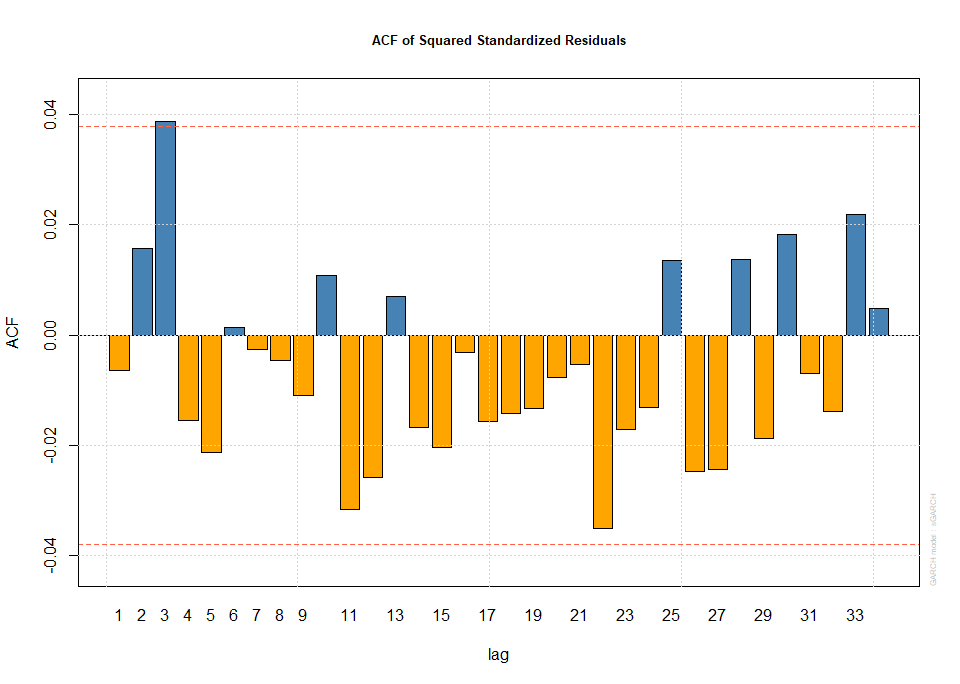


图23 中小板 残差检验图

**表6：沪深300与中小板arch效应检验结果**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | x-squared | | df | | P-value |
| 残差 | 沪深300 | 26.707 | 12 | | 0.008513 | |
| 中小板 | 15.733 | | 12 | | 0.2038 |
| 残差平方 | 沪深300 | 932.18 | | 12 | | < 2.2e-16 |
| 中小板 | 837.54 | | 12 | | < 2.2e-16 |

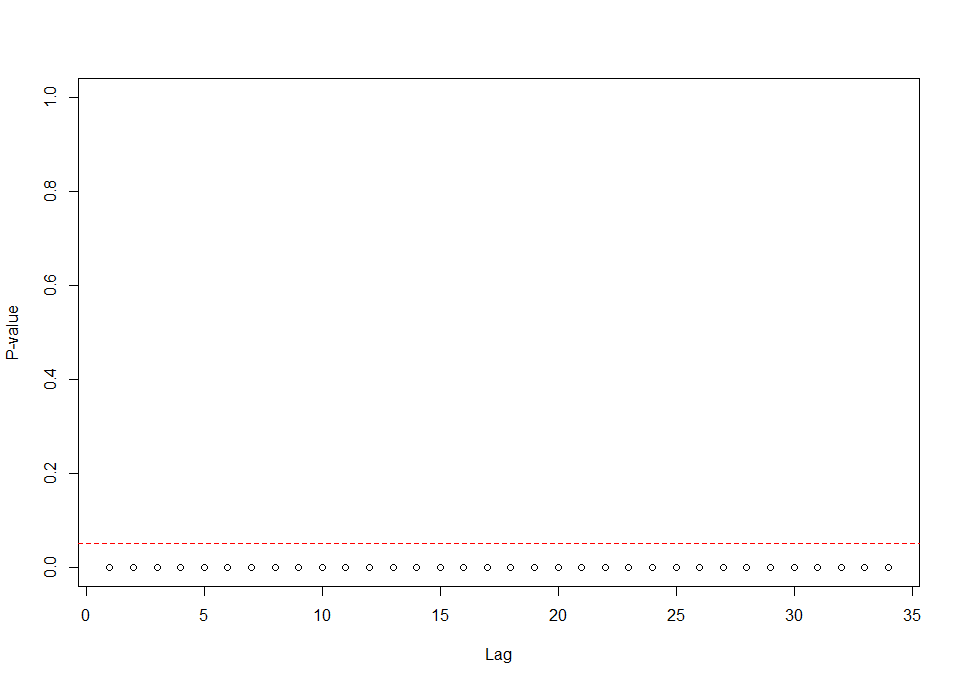


图24 沪深300 Mc.Li.test图

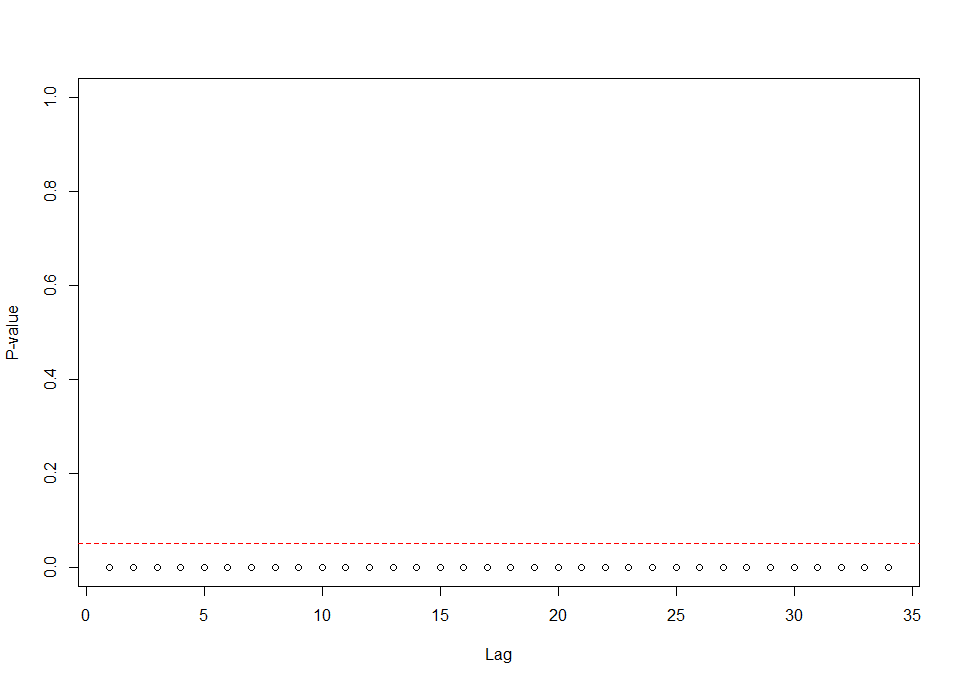


图25 中小板 Mc.Li.test图

由图20-图25以及表6的结果可以看出，残差和残差平方都已经没有自相关、无arch效应，已经是白噪声了。

继续对残差分布做分析，结果如下：

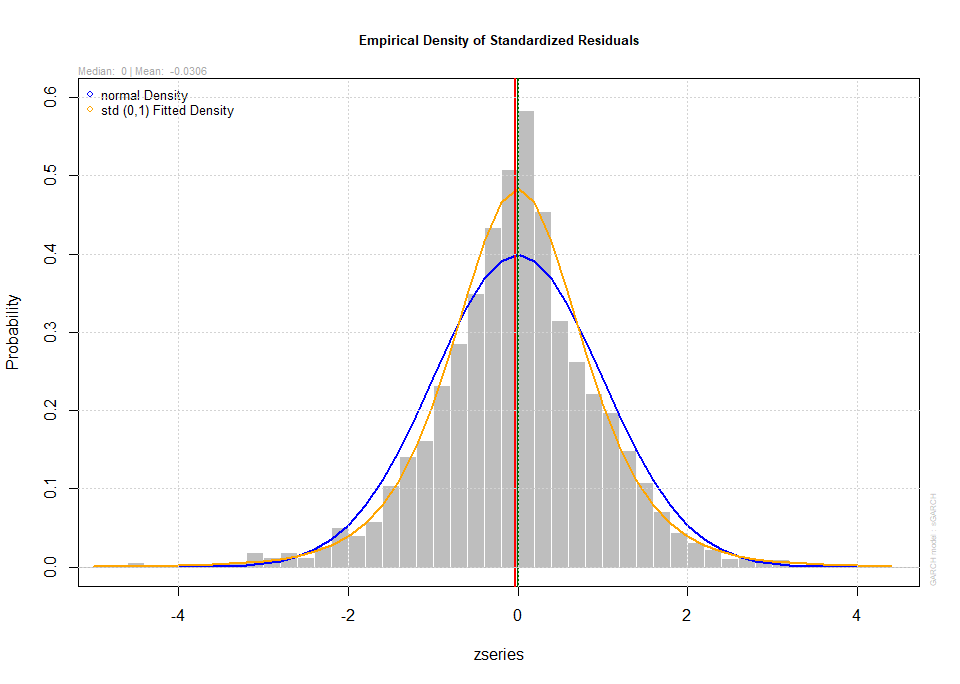


图26 沪深300残差分布图

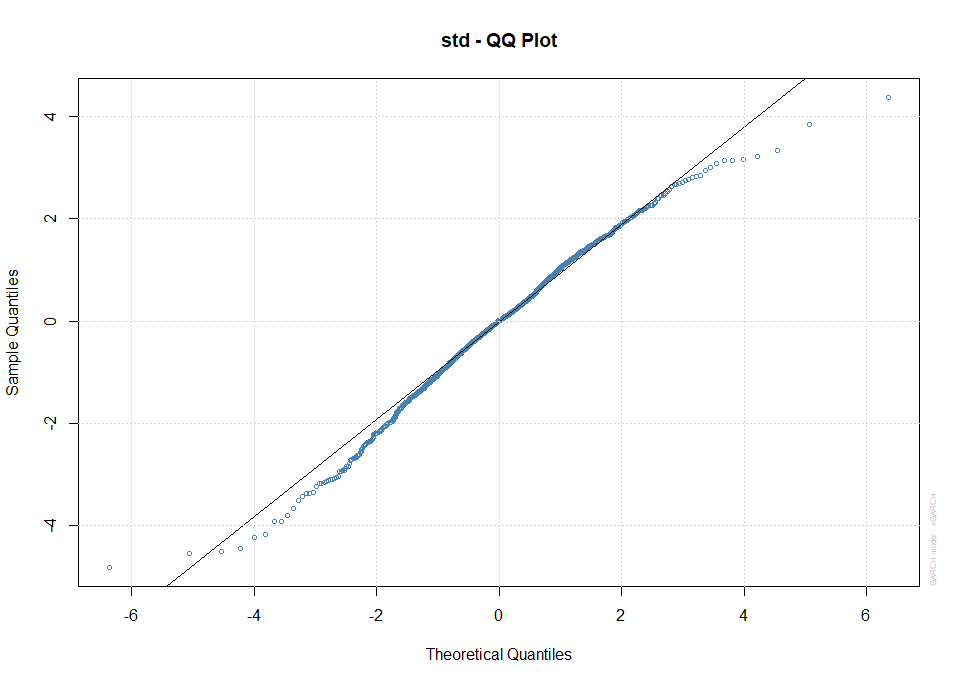


图27 沪深300残差Q-Q图

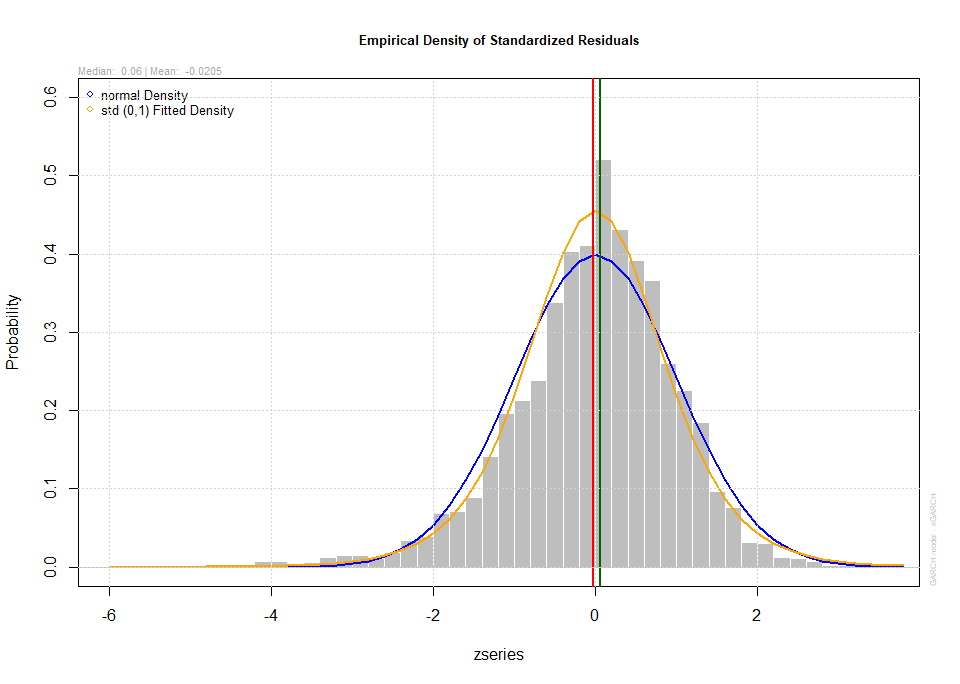


图28 中小板残差分布图

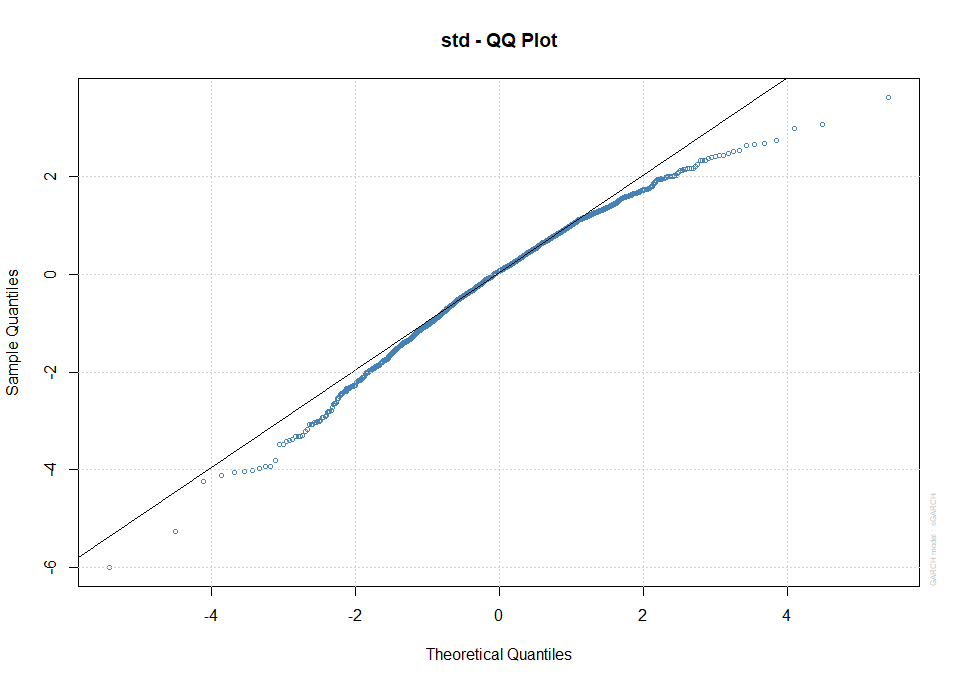


图29 中小板残差Q-Q图

通过对分布图像的观察可以发现，残差分布均为有偏的t分布，其中中小板的偏度更大，与最初描述性统计得到的结果相同。

**4.模型拟合与预测**

下面进行模型拟合和预测，对历史数据的条件方差做拟合，结果如下：

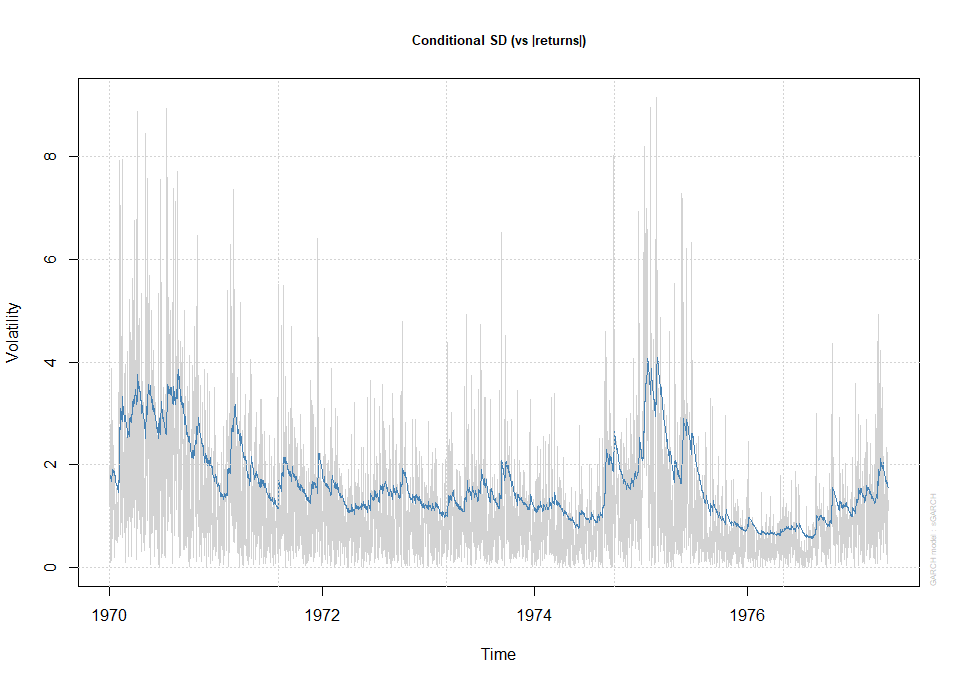


图32 沪深300条件方差拟合

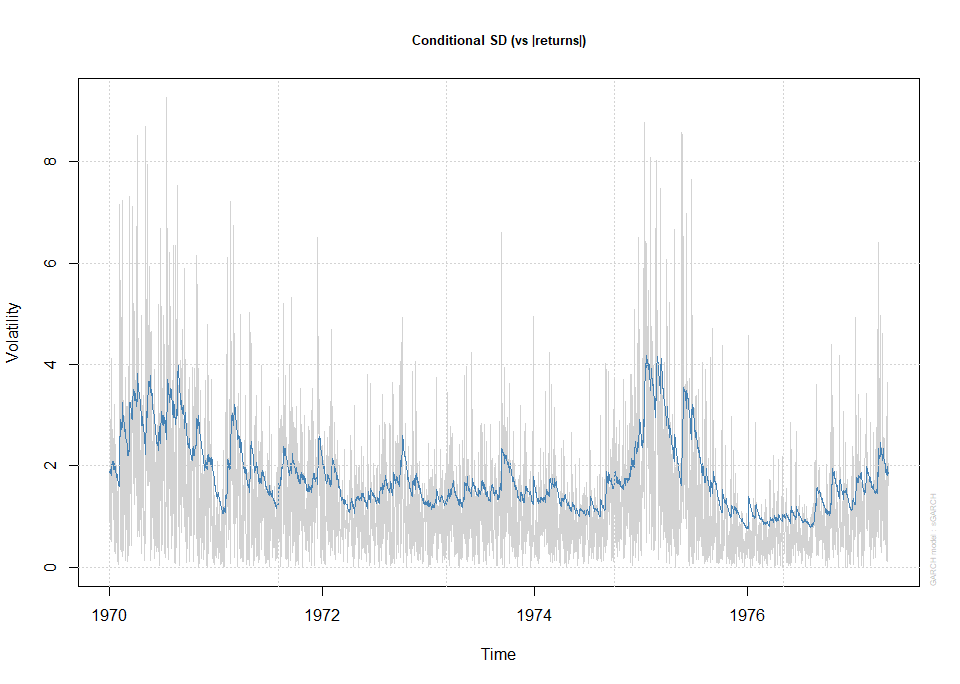


图33 中小板条件方差拟合

由图28-31可以看出，聚簇的地方、方向、大致的波峰波谷等等都还拟合的不错，该模型优于之前。

方差预测见下表：

表7：方差预测表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 沪深300 | | 中小板 | |
| 时间 | 预测值 | 方差 | 预测值 | 方差 |
| T+1 | 0.100918 | 1.531 | 0.2601419 | 1.871 |
| T+2 | -0.002945 | 1.533 | -0.1449981 | 1.873 |
| T+3 | 0.086860 | 1.535 | 0.1482920 | 1.875 |
| T+4 | 0.009210 | 1.537 | -0.0640275 | 1.877 |
| T+5 | 0.076350 | 1.539 | 0.0896755 | 1.879 |
| T+6 | 0.018297 | 1.541 | -0.0215936 | 1.881 |
| T+7 | 0.068492 | 1.542 | 0.0589567 | 1.884 |
| T+8 | 0.025091 | 1.544 | 0.0006445 | 1.886 |
| T+9 | 0.062618 | 1.546 | 0.0428580 | 1.888 |
| T+10 | 0.030170 | 1.548 | 0.0122986 | 1.890 |

## （三）考察非对称性

考察-10~+10的t-1时刻的随机扰动项对t时刻的方差的影响，得到下图：

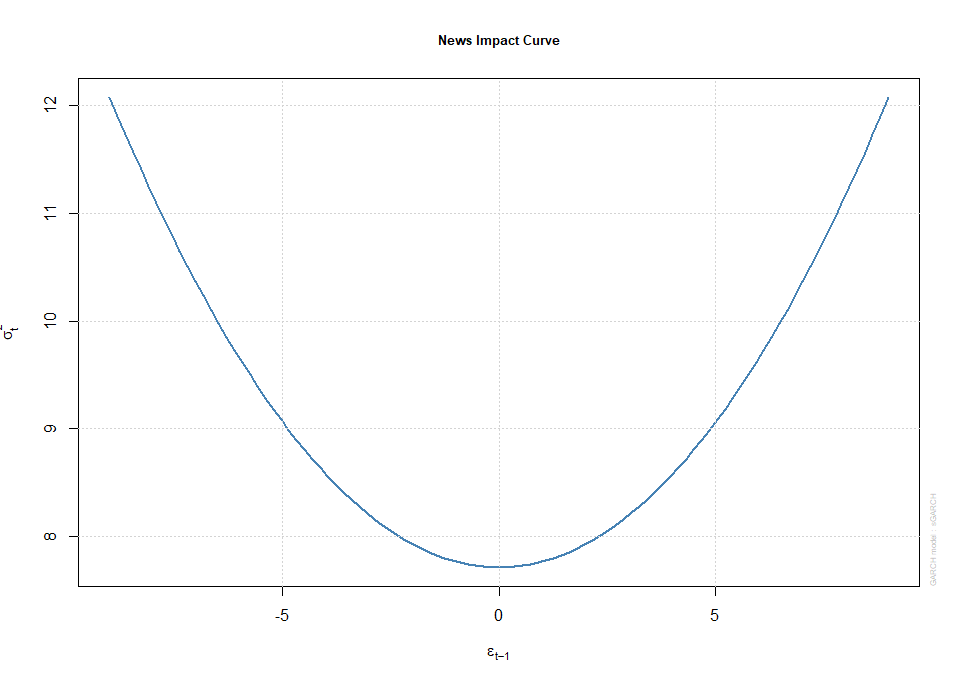


图34 沪深300信息冲击影响图

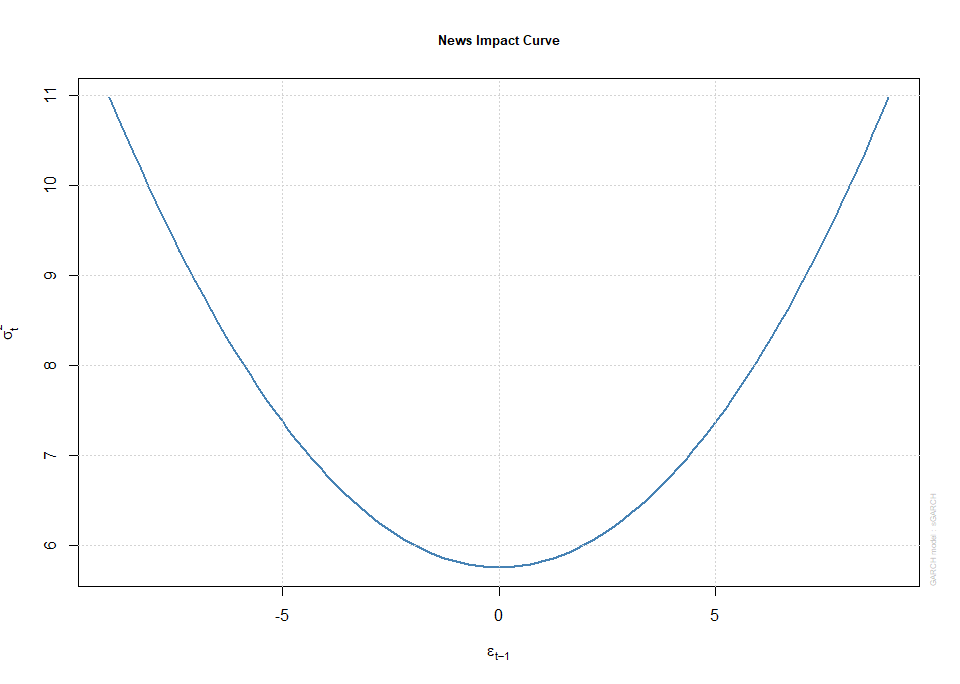


图35 中小板信息冲击影响图

收益率在信息冲击小于零（即利空消息或负冲击）时比较陡峭，而在信息冲击大于零（即利好消息或正冲击时）则比较平缓。说明中国A股还是对负面消息太敏感，抗风险能力较弱。这或许和中国股市散户较多，一有风吹草动就大撤离有关，中小板市场非对称性同样保持，而且相同冲击下，对波动的影响比沪深更大。

## （四）基于VAR曲线的收益风险预警

既然股票对数收益率计算公式为,其中为股票指数收益率，第t期的股票指数收盘价，有波动聚集性，意味着大波动不是突然发生，而是逐步发生的，大波动发生前应该有相应的信号出现，比如出现写异常低的收益率。

基于以上的假设，定义VaR（value at risk）曲线：

VaR=

其中EGARCH模型的预测方差，q为对应分布1%水平的下分位点（这里是skew T分布的下分位数），FitValue为EGARCH模型的拟合值，VaR曲线下的点即为实际波动率在预测波动率99%置信区间以外的点。

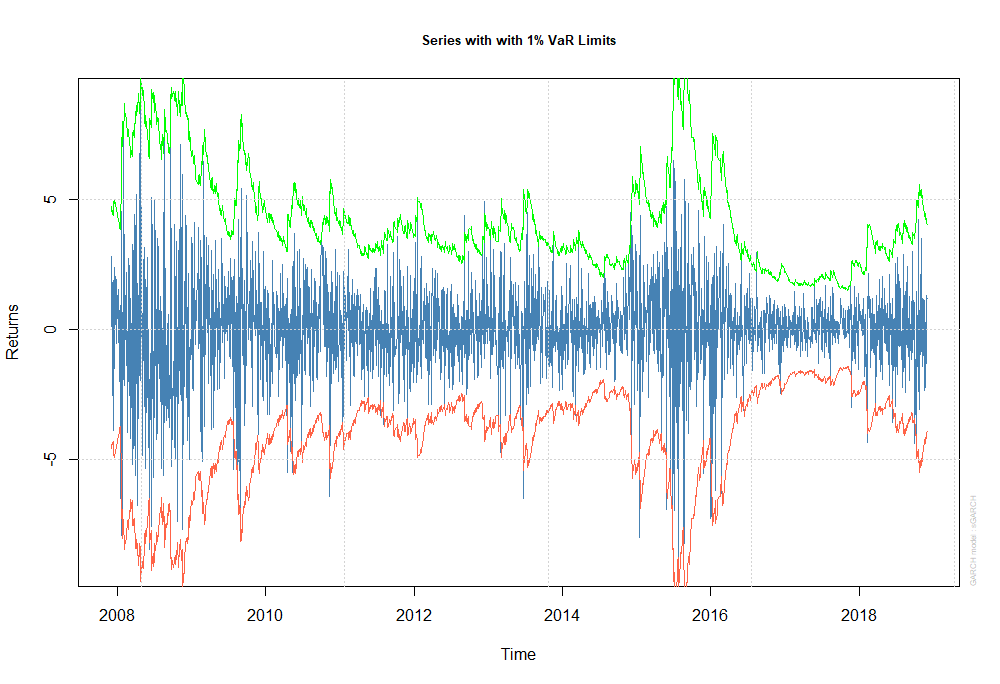


图36 沪深300在1%VaR限制下的序列

由上图可知在较大波动率之前确实有一些点落在VaR曲线以下。但在波动平稳期，模型预警正确率不高（如2013左右）所以该模型在波动率非平稳期有一定参考价值。

但也可见，现在上证指数的大跌在之前是有留下一定信号的。

再看看中小板的情况：

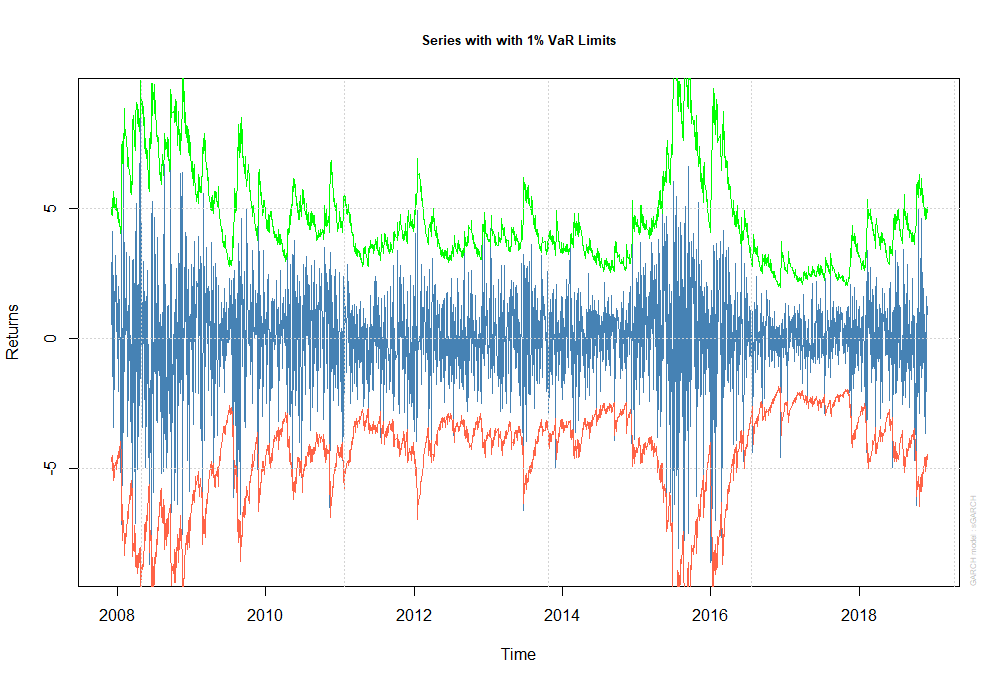


图38 中小板在1%VaR限制下的序列

观察可以发现该图非对称性很强，下边的区间要比上面区间大很多，这个VaR图中明显可以看到负VaR的绝对值要大很多，和非对称分布一致，所以低于VaR的值也更少了，异常值也几乎没有。虽然还起不到预警的效果，但说明至少模型本身还是具有一定参考价值，拟合的效果不错。

# 参考文献

罗攀攀，2013： 基于EVT的ARMA-EGARCH-M模型VaR研究 南昌大学 O211.4;O212.1

YAN DONG ，2012，ARMA and GARCH-type Modeling Electricity Prices ,CHALMERS UNIVERSITY OF TECHNOLOGY,SE-412 96 Göteborg, Sweden

# 附录

**代码**：（以沪深300为例。中小板的分析代码同理）

library(zoo)

library(forecast)

library(TSA)

library(tseries)

library(fUnitRoots)

library(forecast)

library(rugarch)

#提取数据

setwd("D:/WJ")

hushendata<-read.csv("hushen.csv",header=F,fill=TRUE)

value<-read.table("value.txt")

value<-unlist(value)

zxdata<-read.csv("zhongxiao.csv",header=F,fill=TRUE)

value2<-read.table("value2.txt")

value2<-unlist(value2)

zxdata<-zxdata[-1,]

zxdata<-zxdata[1:(length(zxdata$V1)-2),]

hushendata<-hushendata[-1,]

hushendata<-hushendata[1:(length(hushendata$V1)-2),]

time<-as.Date(hushendata$V3)

#描述性统计

plot(time,value,type="l",main="沪深300指数",xlab="time",ylab="index",ylim=c(0,6000))

plot(time,value2,type="l",main="中小板指数",xlab="time",ylab="index",ylim=c(1000,12000))

logreturn <- diff(log(value))\*100

logreturn2 <-diff(log(value2))\*100

newtime <- time[1:(length(time)-1)]

plot(newtime,logreturn,type="l",main="沪深300指数对数收益率")

abline(h=0,lty=2,col="red")

plot(newtime,logreturn2,type="l",main="中小板指数对数收益率")

abline(h=0,lty=2,col="red")

hist(logreturn,main="沪深300",col="yellow",xlab="",ylim=c(0,0.4),probability=T)

lines(density(logreturn),lwd=1);rug(logreturn)

hist(logreturn2,main="中小板指数",col="yellow",xlab="",ylim=c(0,0.4),probability=T)

lines(density(logreturn2),lwd=1);rug(logreturn2)

mean(logreturn)

summary(logreturn)

n<-length(logreturn)

u<-sum(logreturn)/n # 求均值

e<-sqrt(sum((logreturn-u)^2)/(n-1)) # 求标准差

s<-sum((logreturn-u)^3)/((n-1)\*e^3) # 求偏度

k<-sum((logreturn-u)^4)/((n-1)\*e^4) # 求峰度

jarque.bera.test(logreturn) # JB正态性检验，拒绝，符合上述大致t分布的猜想

adfTest(value)

adfTest(logreturn)

adfTest(value2)

adfTest(logreturn2)

acf(logreturn)

acf(logreturn2)

model<-Arima(logreturn,c(3,0,2))

model2<-Arima(logreturn2,c(2,0,3))

tsdiag(model)

tsdiag(model2)

qqnorm(residuals(model))

qqline(residuals(model),lty=2,col="red")

qqnorm(residuals(model2))

qqline(residuals(model2),lty=2,col="red")

newzx<-c(5275.91,5282.01,5278.66,5132.72,5128.49,5050.66,5077.64,5080.2,5140.13,5014.62)

newhushen<-c(3260.95,3267.71,3252,3181.67,3181.56,3144.76,3159.82,3170.61,3219.69,3165.91)

pred.model<-predict(model,n.ahead=11,trace=F,plot=T)

plot(forecast(model),xlim=c(2600,2686))

forecast(model)

future<-diff(log(newhushen))\*100

future

#points(x=c(2678:2686),y=future,pch=3)

pred.model2<-predict(model2,n.ahead=11,trace=F,plot=T)

plot(forecast(model2),xlim=c(2600,2686))

forecast(model2)

future2<-diff(log(newzx))\*100

future2

Box.test(residuals(model)^2,type = "Ljung-Box",lag=12)

Box.test(residuals(model2)^2,type = "Ljung-Box",lag=12)

acf(residuals(model)^2)

acf(residuals(model2)^2)

McLeod.Li.test(y=residuals(model))

McLeod.Li.test(y=residuals(model2))

logreturn<-zoo(logreturn,time)

logreturn2<-zoo(logreturn2,time)

#模型拟合2

myspec=ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 1), submodel = NULL,external.regressors = NULL, variance.targeting = FALSE),

mean.model = list(armaOrder = c(1, 1), include.mean = TRUE, archm = FALSE, archpow = 1, arfima = F,external.regressors = NULL, archex = FALSE), distribution.model = "std")

myfit=ugarchfit(myspec,data=logreturn,solver="gosolnp")

myfit

myfit2=ugarchfit(myspec,data=logreturn2,solver="gosolnp")

myfit2

#残差检验

Box.test(residuals(myfit),type = "Ljung-Box",lag=12)

Box.test(residuals(myfit)^2,type = "Ljung-Box",lag=12)

McLeod.Li.test(y=residuals(myfit))

plot(myfit,which=10)

plot(myfit,which=11)

plot(myfit,which=3)

Box.test(residuals(myfit2),type = "Ljung-Box",lag=12)

Box.test(residuals(myfit2)^2,type = "Ljung-Box",lag=12)

McLeod.Li.test(y=residuals(myfit2))

plot(myfit2,which=10)

plot(myfit2,which=11)

plot(myfit2,which=3)

#模型分布检验（std，normal，sstd）

plot(myfit,which=8)

plot(myfit,which=9)

plot(myfit2,which=8)

plot(myfit2,which=9)

shapiro.test(coredata(residuals(myfit)))

shapiro.test(coredata(residuals(myfit2)))

plot(myfit,which=12)

plot(myfit2,which=12)

plot(myfit,which=2)

plot(myfit2,which=2)

#模型预测2

plot(ugarchforecast(myfit,n.ahead=15))

ugarchforecast(myfit,n.ahead=10)

ugarchforecast(myfit2,n.ahead=10)