

基于GARCH模型对股票市场进行分析预测

贾 雪, 吴芷婧, 孙佳萍, 欧 圆, 耿 帅, 白晓东*

大连民族大学, 辽宁 大连
Email: *baixd_dlnu@163.com

收稿日期: 2021年3月21日; 录用日期: 2021年4月5日; 发布日期: 2021年4月20日

摘 要

本文研究了上海证券综合指数和深圳成分股指数, 发现两者趋势十分相似, 波动特征几乎相同。为了更好的预测股票发展, 我们对两者对数收益率进行统计分析, 建立GARCH模型。结果表明, 我国股票对数收益率波动具有较高持续性, 投机因素较强, 具有一定的风险。

关键词

时间序列分析, 描述性统计分析, GARCH模型

The Analysis and Forecast of Stock Market Based on GARCH Model

Xue Jia, Zhijing Wu, Jiaping Sun, Yuan Ou, Shuai Geng, Xiaodong Bai*

Dalian Minzu University, Dalian Liaoning
Email: *baixd_dlnu@163.com

Received: Mar. 21st, 2021; accepted: Apr. 5th, 2021; published: Apr. 20th, 2021

Abstract

This paper studies Shanghai Composite Index and Shenzhen Component Index, and finds that they have similar trends and almost identical fluctuation characteristics. In order to better predict the stock development, we make statistical analysis on the logarithmic returns of the two, and establish GARCH model. The results show that the fluctuation of logarithmic return rate of Chinese stock has high persistence, strong speculative factors and certain risks.

*通讯作者。

Keywords

Time Series Analysis, Descriptive Statistical Analysis, GARCH Mode

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 背景

随着我国市场经济的飞速发展,全球经济联系的愈加紧密,股票市场也得到了极大的发展和扩大。我国股市自 1986 年诞生以来,至今已经经历近 30 年的不断磨砺与成长,1990 年 12 月深圳证券交易所和上海证券交易所的建立更是翻开了我国股市的新篇章。然而却因为成立时间较短和发展进程比较落后尚未形成完善的供求机制和市场监管机制,高速发展的股票立即出现了许多问题。并且因为我国自身处于经济转型和世界经济全球化的洪流中,中国股市受到多重不确定因素的影响。从股票市场的发展、完善和规范市场这一基本特征来看,当前我国股票市场仍然还处于比较稚嫩的成长期[1]。

时间序列分析是一种用于处理动态数据的统计方法,它的理论基础和研究方法在金融领域探索中有着不可替代的地位。许多的对于时间序列分析模型如 ARCH 模型、GARCH 模型和 ARMA 模型等都在对于金融时间序列预测中体现出极大的作用,其中由 Bollerslev 在 Engle 的自回归条件异方差 ARCH 的基础上提出的广义条件异方差自回归模型 GARCH 在研究探索金融数据间的特征和规律上有着广泛的应用。当我们选取不同的 GARCH 的尾部时,它们的建模结果并不一样,因此我们需要选取最为合适的 GARCH 尾部来进行建模,探讨出更合适的残差分布来描述我国的股市走向和预测性,并对我国上海和深圳两个股票市场之间进行分析探讨和说明。

2. 预备知识

2.1. 上证综指

上海证券综合指数简称“上证指数”或“上证综指”,其样本股是在上海证券交易所全部上市股票,包括 A 股和 B 股,反映了上海证券交易所上市股票价格的变动情况。上证综合指数是最早发布的指数,是以上证所挂牌上市的全部股票为计算范围,以发行量为权数的加权综合股价指数[2]。

2.2. 深证成指

深证成份指数,简称深证成指(SZSE COMPONENT INDEX)是深圳证券交易所的主要股指。它是按一定标准选出 500 家有代表性的上市公司作为样本股,用样本股的自由流通股数作为权数,采用派氏加权法编制而成的股价指标。以 1994 年 7 月 20 日为基期,基点为 1000 点。

2.3. 收益率的定义以及公式

股票收益率指投资于股票所获得的收益总额与原始投资额的比率。在股票研究时,我们往往研究收益率,因为收益率受其他干扰因素较小,相比价格更容易分析[3]。

设定 P_t 为某资产在 t 时刻的价格,那么收益 R_t 则为:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

常用的收益率还有一种就是对数收益率，其定义如下：

$$r_t = \ln \frac{P_t}{P_{t-1}}$$

对上式两种收益率进行比较，发现对数收益率更易于进行建模运算，且对数处理使数据更加平滑，更易于描绘数据的特征。因此在文章接下来的研究中，主要采用对数收益率的计算公式。

3. 股票趋势研究以及 Garch 模型

3.1. 数据的波动特征

首先做出两股收益率时序图，展开分析讨论。

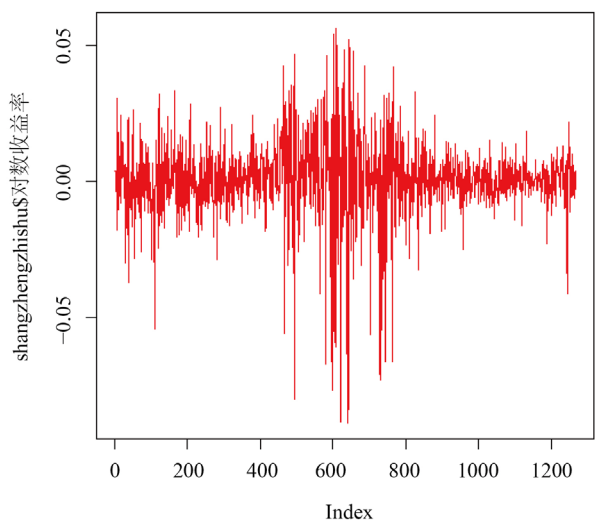


Figure 1. The log-return distribution of the Shanghai Composite Index distribution

图 1. 上证综合指数对数收益率分布图

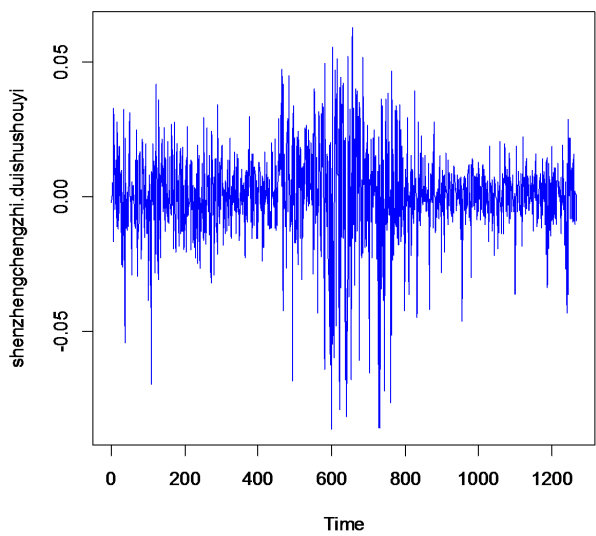


Figure 2. The log-return of Shenzhen Stock Index

图 2. 深证成指对数收益率分布图

从图 1 和图 2 中可以明显的看出，上证综合指数、深证成指对数收益率的数值在以 0.00 的水平线附近上下波动，起伏范围在-0.1 至 0.06 之间，着属于正常范围，未超出我国股市 10%的涨停板的限制。图中显示，收益率分布序列出现多次峰值，波动变化较为剧烈，说明了股市收益率的突变性。同时，上证综指、深证成指也呈现出来波动群聚现象，一波股市的波动会带动另一波股市的波动，具有明显的金融时间序列的特征。

3.2. 正态性检验

如图所示：图 3 与图 4 是上证综指和深证成指的对数收益率的直方图，我们可以很直观看出其正态性。并且通过表 1 得到两者的偏度都为负，可以看出其分布都呈左偏，说明上证综合指数和深证成指走

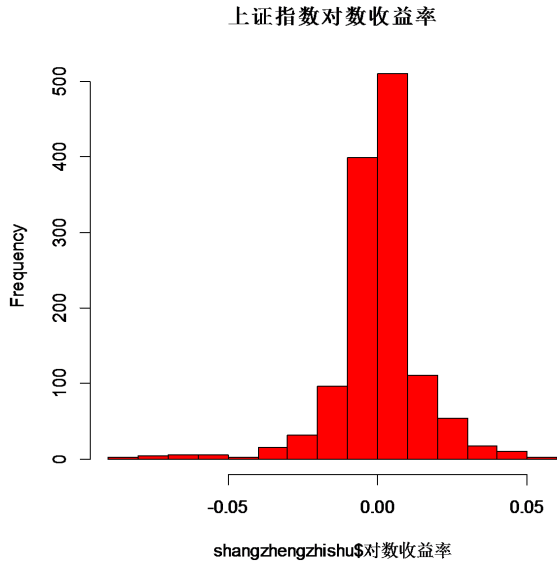


Figure 3. Shanghai Composite Index log-return histogram
图 3. 上证综合指数对数收益直方图

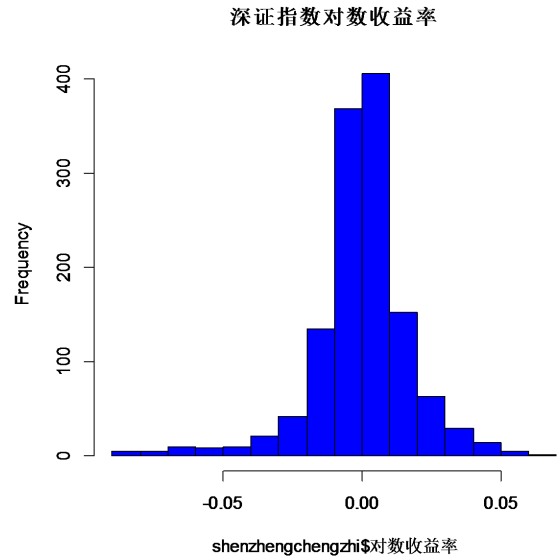


Figure 4. Shenzhen composite index log-return histogram
图 4. 深证综合指数的对数收益直方图

Table 1. Basic statistical characteristics of logarithmic returns of Shanghai and Shenzhen Stock Exchange
表 1. 沪深两股对数收益率基本统计特征

	上证综合指数	深证综合指数
均值	0.0002910	0.0001468
标准差	0.01476152	0.01720294
偏度	-1.257995	-0.9824075
峰度	7.391537	4.58369
JB 检验	3232.1	1319.3
P 值	0.000000	0.000000

势状态具有不平衡性，且收益率低于自身均值状况的交易日占多数，同时表中 J-B 统计量的 P 值也能看出，P 值为 0，说明两个序列都不是正态分布。上证综合指数和深证成指的峰度系数大于 3，在此说明了上证综合指数和深证成指的对数收益率存在尖峰厚尾特征。同时，深证成指的平均值大于上证综合指数，但上证综合指数和深证成指的均值都大于 0，这与我国近年来良好的经济增长形式是紧密相关的。标准差表示出波动性，且表中数据反映深证成指的标准差大于上证综合指数，相差 0.02，比较小，说明深证成指比上证综合指数波动性的幅度稍大一些，且标准差过低也说明了我国股市尚未发展完善。

3.3. 自相关图和偏自相关图

如图所示：图 5 是上证指数的对数收益率的自相关图和偏自相关图，可以看出两个图大部分函数值在置信区间内(图中蓝色虚线区域)上下跳跃，并且通过表 2 自相关表发现，自相关系数很小，所以收益率序列自相关性很低，或者说具有很弱的自相关性，因此在条件期望模型中不需要引入自相关性部分，满足 GARCH 模型中的均值方程，收益率由一个常数项加上一个随机扰动项组成[4]。

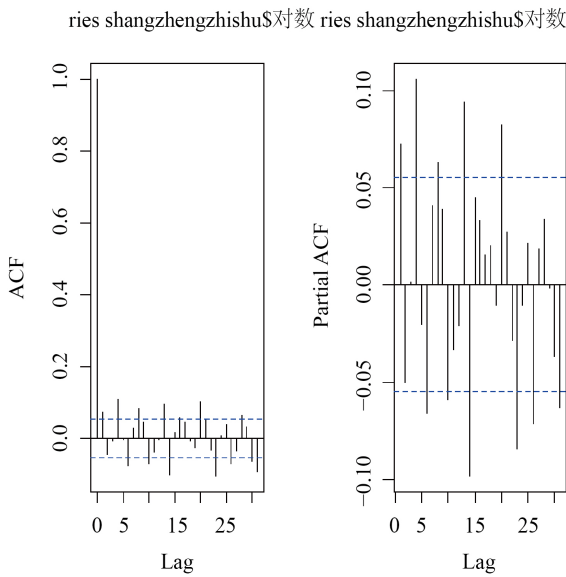


Figure 5. Autocorrelation and partial autocorrelation of logarithmic return of Shanghai Composite Index

图 5. 上证指数的对数收益率的自相关图和偏自相关图

Table 2. Autocorrelation
表 2. 自相关系数表

0	1	2	3	4	5	6	7	8	9	10
1.000	0.072	-0.045	-0.005	0.107	-0.004	-0.076	0.029	0.084	0.046	-0.071
11	12	13	14	15	16	17	18	19	20	21
-0.038	-0.003	0.094	-0.101	0.016	0.056	0.045	-0.007	-0.026	0.101	0.050
22	23	24	25	26	27	28	29	30	31	
-0.031	-0.104	0.007	0.038	-0.070	-0.036	0.063	0.033	-0.064	-0.093	

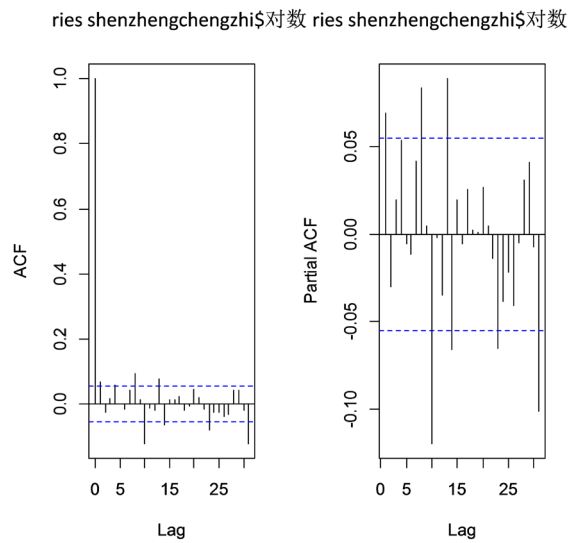


Figure 6. Autocorrelation and partial autocorrelation of log-return
图 6. 深证成指的对数收益率的自相关图和偏自相关图

Table 3. Autocorrelation
表 3. 自相关系数表

0	1	2	3	4	5	6	7	8	9	10
1.000	0.069	-0.025	0.015	0.057	0.001	-0.015	0.042	0.092	0.014	-0.121
11	12	13	14	15	16	17	18	19	20	21
-0.012	-0.020	0.078	-0.063	0.012	0.012	0.022	-0.021	-0.007	0.044	0.021
22	23	24	25	26	27	28	29	30	31	
-0.014	-0.083	-0.025	-0.027	-0.039	-0.032	0.041	0.040	-0.019	-0.121	

如图所示：图 6 是深证成指的对数收益率的自相关图和偏自相关图，可以看出两个图大部分函数值在置信区间内(图中蓝色虚线区域)上下跳跃，并且通过表 3 自相关表发现，自相关系数很小，所以收益率序列自相关性很低，或者说具有很弱的自相关性，因此在条件期望模型中不需要引入自相关性部分，满

足 GARCH 模型中的均值方程, 收益率由一个常数项加上一个随机扰动项组成。

3.4. ARCH 效应检验

利用 ArchTest 函数检验上证指数序列是否有异方差效应

ARCH LM-test; Null hypothesis: no ARCH effects

data: shangzhengzhishu\$对数收益率

Chi-squared = 214.1, df = 12, p-value < 2.2e-16

分析: 检验结果为卡方统计量的值为 214.1, 对应的 P 值几乎为 0, 也就是说在 1% 的显著性水平上拒绝原假设, 从而拒绝不存在 ARCH 效应的假设, 收益率序列存在 ARCH 效应, 可以进行 GARCH 模型的拟合[5]。

利用 R 中 garchFit 函数进行 GARCH 模型的拟合得到下列结果:

Error Analysis:

	Estimate	Std. Error	t value	Pr(> t)
mu	3.476e-04	2.499e-04	1.391	0.1643
omega	6.289e-07	2.564e-07	2.453	0.0142*
alpha1	6.588e-02	1.044e-02	6.312	2.76e-10***
beta1	9.330e-01	9.188e-03	101.553	<2e-16***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

均值方程: $y_t = 3.476 \times 10^{-4} + \varepsilon_t, \varepsilon_t \sim N(0,1)$

Garch 方程: $H_t^2 = 6.289 \times 10^{-7} + 6.588 \times 10^{-2} \times \varepsilon_{t-1}^2 + 9.330 \times 10^{-1} \times H_{t-1}^2$

分析: 通过运行结果可知所有系数都是显著的, 说明上证指数收益率过去时刻的波动对现在的影响还是明显的, 具有波动聚集现象, 此外 α_1 和 β_1 的和约为 1, 说明, 条件异方差序列具有长记忆性, 表明了收益率的波动具有较高的持续性, 投机因素较强, 具有一定的风险。

在拟合完 GARCH 模型之后, 我们对模型结果的残差进行分析。

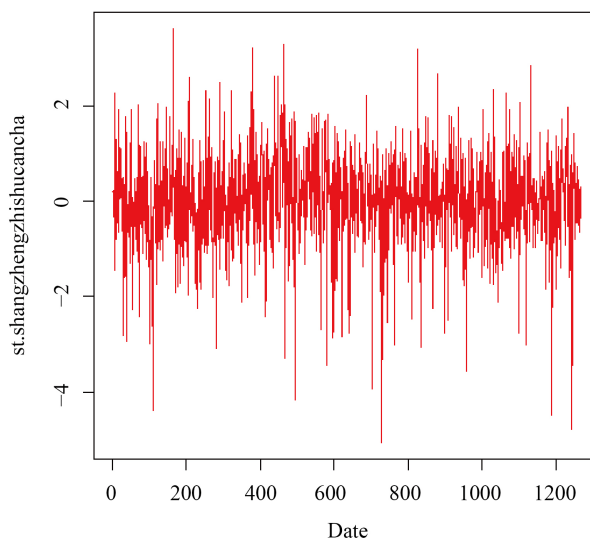


Figure 7. Shanghai composite index residuals

图 7. 上证指数残差

如图所示：图 7 是残差序列时序图，可以看到残差序列没有明显的波动聚集效应。

如图所示：图 8 是残差的自相关图和偏自相关图。可以看到残差序列的 ACF 和 PACF 图大部分函数值在置信区间内(图中蓝色的虚线区域)上下跳跃，所以标准化残差序列并不具有自相关性，或者具有一定的弱相关性(PACF 图结果)。结果表明此 GARCH 模型可以有效地解释收益率序列[6]。

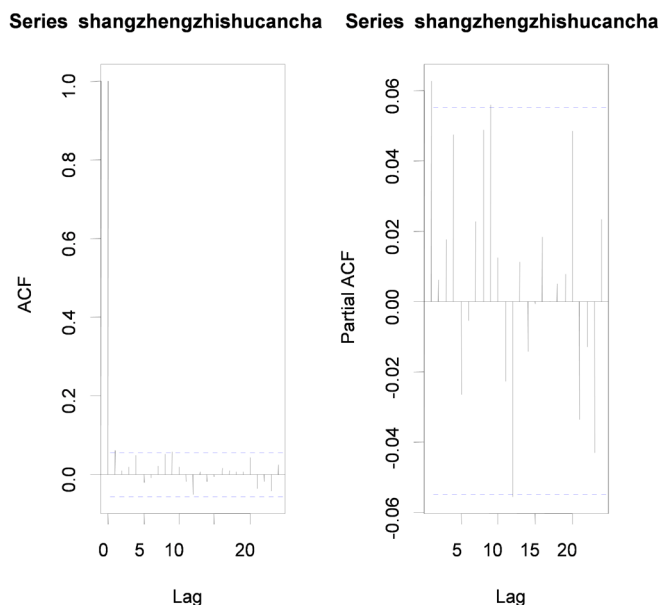


Figure 8. Autocorrelation and partial autocorrelation of residual error of Shanghai Composite Index

图 8. 上证指数残差残差的自相关图和偏自相关图

同理对深证成指进行上述一系列的操作，检验发现：

1) 是否有异方差效应：检验结果为卡方统计量的值为 225.02，对应的 P 值几乎为 0，也就是说在 1% 的显著性水平上拒绝原假设，从而拒绝不存在 ARCH 效应的假设，收益率序列存在 ARCH 效应，可以进行 GARCH 模型的拟合。

2) 进行 GARCH 模型的拟合：均值方程： $y_t = 2.170 \times 10^{-4} + \varepsilon_t, \varepsilon_t \sim N(0,1)$

Garma 方程： $H_t^2 = 1.880 \times 10^{-6} + 5.307 \times 10^{-2} \times \varepsilon_{t-1}^2 + 9.396 \times 10^{-1} \times H_{t-1}^2$ 。

通过运行结果可知所有系数都是显著的，说明深证指数收益率过去时刻的波动对现在的影响还是明显的，具有波动聚集现象，此外 α_1 和 β_1 的和约为 1，说明，条件异方差序列具有长记忆性，表明了收益率的波动具有较高的持续性，投机因素较强，具有一定的风险。

3) 残差序列的时序图：可以看到残差序列没有明显的波动聚集效应。

4) 残差相关性：从深证指数残差的自相关图和偏自相关图可以看到残差序列的 ACF 和 PACF 图大部分函数值在置信区间内上下跳跃，所以标准化残差序列并不具有自相关性，或者具有一定的弱相关性。结果表明此 GARCH 模型可以有效地解释收益率序列。

3.5. 预测

利用求得的模型进行对数收益率预测，所得预测值为表 4 上证指数预测和表 5 深证成指预测中的数据。

1) 上证指数预测

Table 4. Shanghai Composite Index forecasts
表 4. 上证指数预测

1	0.0003476246	0.009683680	0.009683680
2	0.0003476246	0.009710877	0.009710877
3	0.0003476246	0.009737969	0.009737969
4	0.0003476246	0.009764956	0.009764956
5	0.0003476246	0.009791840	0.009791840
6	0.0003476246	0.009818622	0.009818622
7	0.0003476246	0.009845301	0.009845301
8	0.0003476246	0.009871880	0.009871880

2) 深证成指预测

Table 5. Forecast of Shenzhen Composite Index
表 5. 深证成指预测

1	0.0002170391	0.01264380	0.01264380
2	0.0002170391	0.01267148	0.01267148
3	0.0002170391	0.01269890	0.01269890
4	0.0002170391	0.01272607	0.01272607
5	0.0002170391	0.01275297	0.01275297
6	0.0002170391	0.01277962	0.01277962
7	0.0002170391	0.01280602	0.01280602
8	0.0002170391	0.01283217	0.01283217

4. 结论

- 1) 经上证综指和深证成指的线状分布图来看，两者的收益率的分布都具有多次峰值，波动变化剧烈，具有较强的突变性，且都有明显的波动聚群现象
- 2) 通过上证综指和深证成指的直方图，以及偏度得到两者的分布都成左偏，且峰度系数大于 3，有尖峰厚尾特征。且两者的平均值大于 0，这说明我国最近的经济形式是增长状态，但标准差过低，也说明我国的股市尚不成熟[7]。
- 3) 通过求得的公式分析得上证综指和深证成指过去时刻的波动大小对当前波动大小有明显的影响，具有波动聚集效应。上证综指和深证成指的收益率波动具有较高持续性，投机因素较强，总体风险较大。且我们对两股票未来发展做了预测。

5. 模型的优缺点

- 优点：1) 本文采用对数收益率，使数据更加平滑，更易于描绘数据的特征。
- 2) 本文采用 GARCH 模型，相较于 ARCH 模型，更能反映实际数据中的长期记忆性质。同时特别适用于波动性的分析和预测，这样的分析对投资者的决策能起到非常重要的指导性作用，其意义很多时候超过了对数值本身的分析和预测。

3) 本文研究数据来自专门的股票平台, 数据真实, 来源可靠。

缺点: 1) 本文只研究了 GARCH 模型, 没有与其他模型进行对比, 无法判断模型是否是最优、最准确模型。

2) 本文只根据收盘价的起伏变化说明股市之间具有联动性, 没有多因素考虑。

3) 本文采用 GRACH 模型, 存在缺点。

6. 展望

由于对模型的了解还不够深入, 分析还不够透彻, 我们希望以后在以下几个方面进行优化:

1) 未来我们可以利用更加复杂的参数 GRACH 模型比如 Beta 分布, Laplace 分布进行拟合, 或许可以得到更优模型[8]。

2) 未来可以考虑股票的多方面来研究股市之间联动性, 比如一天中股票的最高价, 最低价, 寻找其中是否有某种联系。

基金项目

国家级大创项目(202012026039)。

参考文献

- [1] 刘梦佳. 不同分布假设 GARCH 模型择优及其在股市波动溢出效应中的研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2017.
- [2] 刘湖, 王莹. 股票市场波动性研究——基于 ARMA-TGARCH-M 模型的实分析[J]. 北京航空航天大学学报(社会科学版), 2017, 30(4): 56-66.
- [3] 杨帷. 基于 ARIMA_GARCH 模型的上证综指实证分析[D]: [硕士学位论文]. 湘潭: 湘潭大学, 2017.
- [4] 杨格. 统计金融模型及金融市场波动持续时间序列的研究[D]: [博士学位论文]. 北京: 北京交通大学, 2017.
- [5] 侍成. 几种不同类型金融时间序列模型预测的比较研究[D]: [硕士学位论文]. 北京: 中国矿业大学, 2017.
- [6] 郝博乾. 基于时间序列分析的股票预测模型研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2011.
- [7] 姜乐. 基于时间序列的股票价格分析研究与应用[D]: [硕士学位论文]. 大连: 大连理工大学, 2015.
- [8] 张贵生. 数据驱动的金融时间序列预测模型研究[D]: [博士学位论文]. 太原: 山西大学, 2016.

附 录

```
shangzhengzhishu<-read.csv("E:上证指数.csv",header=T)
shenzhengchengzhi<-read.csv("E:深证成指.csv",header=T)
shangzhengzhishu.shoupan<-ts(shangzhengzhishu$收盘价,start=c(2013,3),frequency=238)
shenzhengchengzhi.shoupan<-ts(shenzhengchengzhi$收盘价,start=c(2013,3),frequency=238)
par(mfrow=c(1,2))
plot(shangzhengzhishu.shoupan,col="red")
plot(shenzhengchengzhi.shoupan,col="blue")
par(mfrow=c(1,1))
shangzhengzhishu.duishushouyi<-ts(shangzhengzhishu$对数收益率,start=c(2013,3),frequency=238)
plot(shangzhengzhishu$对数收益率,col="red",type="l")
par(mfrow=c(1,2))
acf(shangzhengzhishu$对数收益率)
pacf(shangzhengzhishu$对数收益率)
a<-acf(shangzhengzhishu$对数收益率)
par(mfrow=c(1,1))
hist(shangzhengzhishu$对数收益率,col="red",main="上证指数对数收益率")
summary(shangzhengzhishu$对数收益率)
sd(shangzhengzhishu$对数收益率)
skewness(shangzhengzhishu$对数收益率)
kurtosis(shangzhengzhishu$对数收益率)
jarque.bera.test(shangzhengzhishu$对数收益率)
ArchTest(shangzhengzhishu$对数收益率)
shangzhengzhishu<-garchFit(~garch(1,1),data=shangzhengzhishu$对数收益率,trace=F)
summary(shangzhengzhishu)
shangzhengzhishu.pre<-predict(shangzhengzhishu,10)
shangzhengzhishu.pre
shangzhengzhishucancha<-residuals(shangzhengzhishu,standardize=T)
shangzhengzhishucancha<-ts(shangzhengzhishucancha)
par(mfrow=c(1,1))
plot(shangzhengzhishucancha,xlab='Date',ylab='st.shangzhengzhishucancha',type='l,col="red")
par(mfrow=c(1,2))
acf(shangzhengzhishucancha,lag=24)
pacf(shangzhengzhishucancha,lag=24)
shenzhengchengzhi.duishushouyi<-ts(shenzhengchengzhi$对数收益率)
plot(shenzhengchengzhi$duishushouyi)
par(mfrow=c(1,2))
acf(shenzhengchengzhi$对数收益率)
pacf(shenzhengchengzhi$对数收益率)
```

```
b<-acf(shenzhengchengzhi$对数收益率)
par(mfrow=c(1,1))
plot(shenzhengchengzhi.duishushouyi,col="blue")
hist(shenzhengchengzhi$对数收益率,col="blue",main="深证指数对数收益率")
summary(shenzhengchengzhi$对数收益率)
skewness(shenzhengchengzhi$对数收益率)
kurtosis(shenzhengchengzhi$对数收益率)
jarque.bera.test(shenzhengchengzhi$对数收益率)
sd(shenzhengchengzhi$对数收益率)
ArchTest(shenzhengchengzhi$对数收益率)
shenzhengchengzhi=garchFit(~garch(1,1),data =shenzhengchengzhi$对数收益率,trace=F)
summary(shenzhengchengzhi)
shenzhengchengzhi.pre<-predict(shenzhengchengzhi,10)
shenzhengchengzhi.pre
shenzhengchengzhicanha<-residuals(shenzhengchengzhi,standardize=T)
shenzhengchengzhicanha<-ts(shenzhengchengzhicanha)
par(mfrow=c(1,1))
plot(shenzhengchengzhicanha,xlab='Date',ylab='st.shenzhengchengzhicanha',type='l',col="blue")
par(mfrow=c(1,2))
acf(shenzhengchengzhicanha,lag=24)
pacf(shenzhengchengzhicanha,lag=24)
```