沪深 300 时间数据近三年收益率分析

一、下载一只股票或市场指数 3 年的日收益率数据，计算样本偏度、峰度，利用 Q-Q 图和 Jarque-Bera 进行正态性检验。

所用数据为沪深 300 时间数据，由于整个数据的起止时间为 2001 年 2 月 19 日至 2018 年 3 月 16 日，考虑选择最近三年的时间数据进行分析，即将 2015 年 3 月 16 日至 2018 年 3 月 16 日的日收益率作为样本。

1. 计算

由 R 包"moment"中的函数计算出样本的偏度和峰度：

1) 样本偏度为：-1.21
2) 样本峰度为：9.36

2. 正态性检验

首先绘制收益率的时序图：

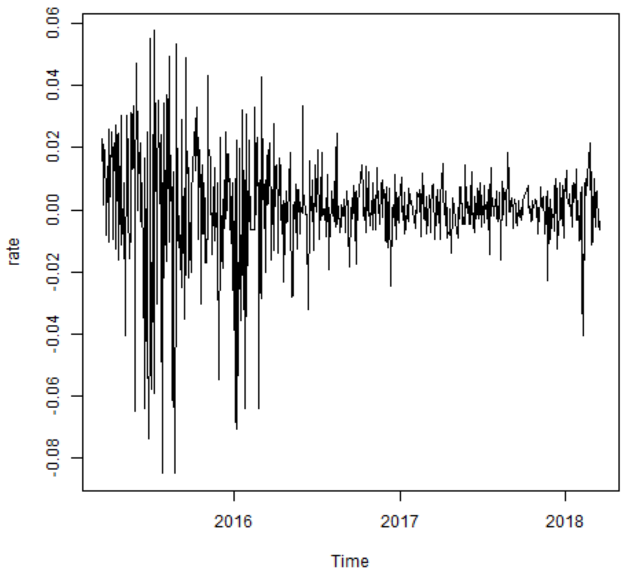

图 1：收益率时序图

从时序图中可以看出收益序列呈现周期不固定的波动变化，因此需要采用一些方法对收益序列进行进一步分析。

对该序列做正态性检验，主要采用四种方法：Q-Q 图、密度函数曲线、Shapiro-Wilk 检验以及 JB 检验：

1) Q-Q 图

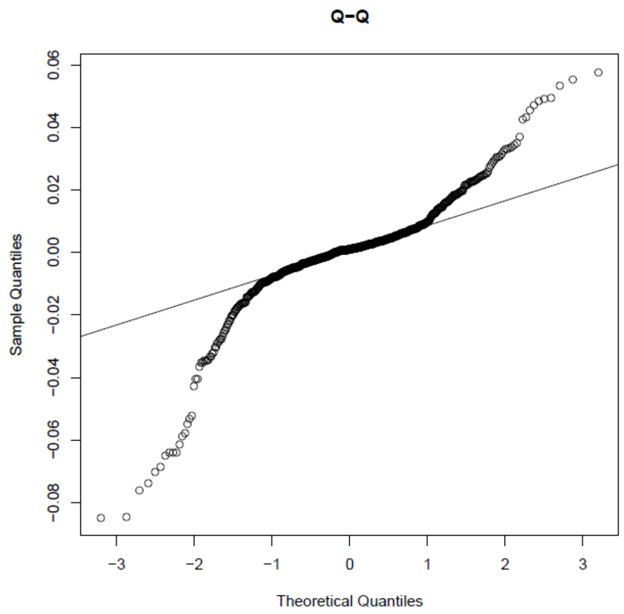

图 2：收益率序列 Q-Q 图

Q-Q 图可以反映出两个概率分布的近似程度。由图 2 可以看出，收益率序列的分布与正态分布有较大差异，基本反映出收益率序列不服从正态分布。

2) 密度函数曲线

为了更加直观，绘制收益率序列的直方图、核密度估计曲线以及正态分布的核密度估计曲线（用收益率百分比作图），如图 3：

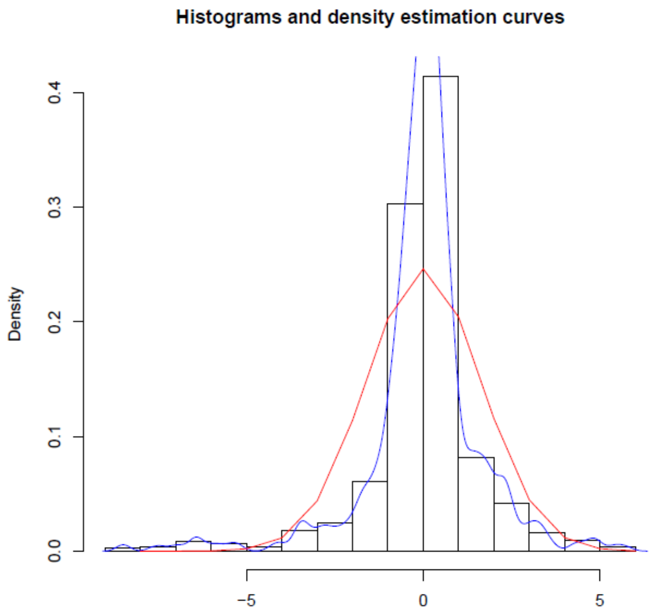

图 3：直方图和核密度估计曲线

红色的线代表以收益率序列的均值为均值，标准差为标准差的正态分布密度曲线；蓝色的线为收益率序列实际的密度函数曲线，可以看到二者之间存在差异。

3) Shapiro-Wilk 检验

对收益率序列做 SW 检验，经计算，收益率序列的 P 值为 2.2e-16，远远小于 0.05，因此拒绝原假设，即收益率序列不服从正态分布。

4) JB 检验

对收益率序列做 JB 检验，结果如图 4：

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 1418.3317
  P VALUE:
    Asymptotic p Value: < 2.2e-16
```

图 4：JB 检验结果

可以看到 JB 检验统计量的值非常大，因此有理由认为收益率序列不服从正态分布。

二、绘制收益率和收益率绝对值的 ACF 图像，进行分析和 Ljung-Box 检验。

1. ACF 图

利用“ggplot2”包绘制收益率和收益率绝对值的 ACF 图像。

1) 收益率序列的 ACF 图像如图 5：

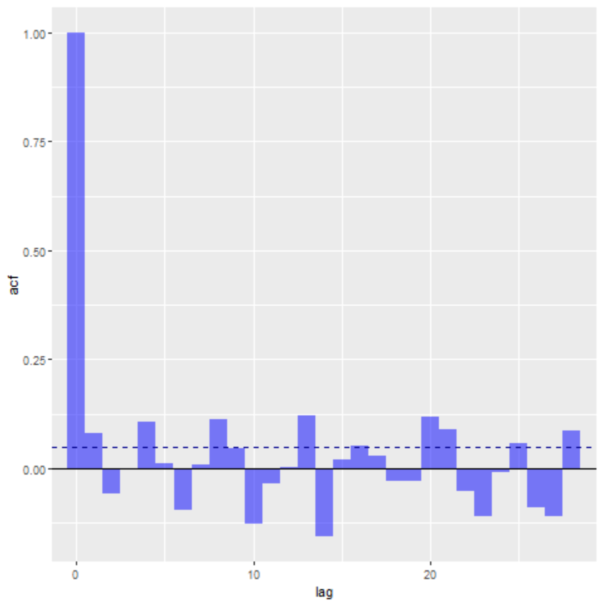

图 5：收益率的 ACF 图像

由收益率序列的 ACF 图像可以基本确定序列存在自相关，不是白噪声序列。

1) 收益率绝对值的 ACF 图像如图 6：

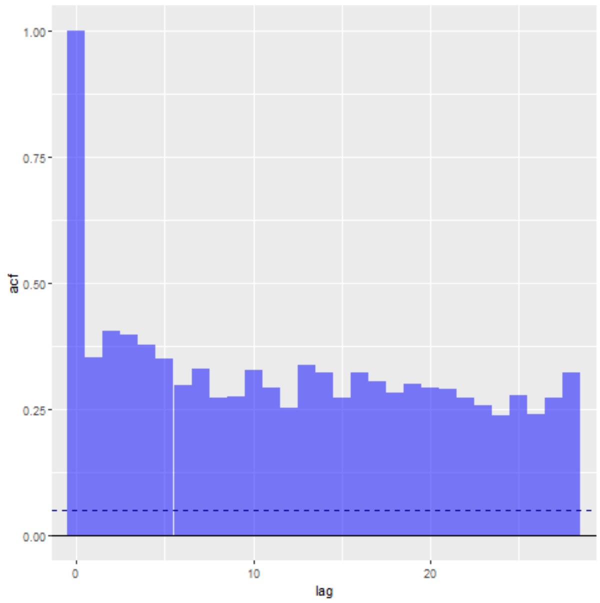

图 6：收益率绝对值的 ACF 图像

由图 6 可以看出，收益率绝对值序列有很强的自相关性。

2. Ljung-Box 检验

Ljung-Box 检验是对时间序列是否存在滞后相关的一种统计检验。首先对收益率序列进行 Ljung-Box 检验，检验结果分别如图 7 和图 8：

```
        Box-Ljung test

data:  newX[, i]
X-squared = 46.175, df = 10, p-value = 1.333e-06
```

图 7：收益率序列 Ljung-Box 检验结果

```
        Box-Ljung test

data:  newX[, i]
X-squared = 869.09, df = 10, p-value < 2.2e-16
```

图 8：收益率绝对值序列 Ljung-Box 检验结果

由以上结果，检验均应拒绝原假设，即认为两序列均存在自相关性。

三、分别利用高斯新息和 t 分布新息的 ARMA(1,1)-GARCH(1,1)模型进行拟合，对标准化残差进行分析（ACF 和 Q-Q 图），利用 AIC 和 BIC 比较高斯新息和 t 分布新息模型。

1. 绘图

首先分别利用高斯新息和 t 分布新息的 ARMA(1,1)-GARCH(1,1)模型进行拟合，绘制标准化残差的 Q-Q 图和 ACF 图，如图 9 和图 10：

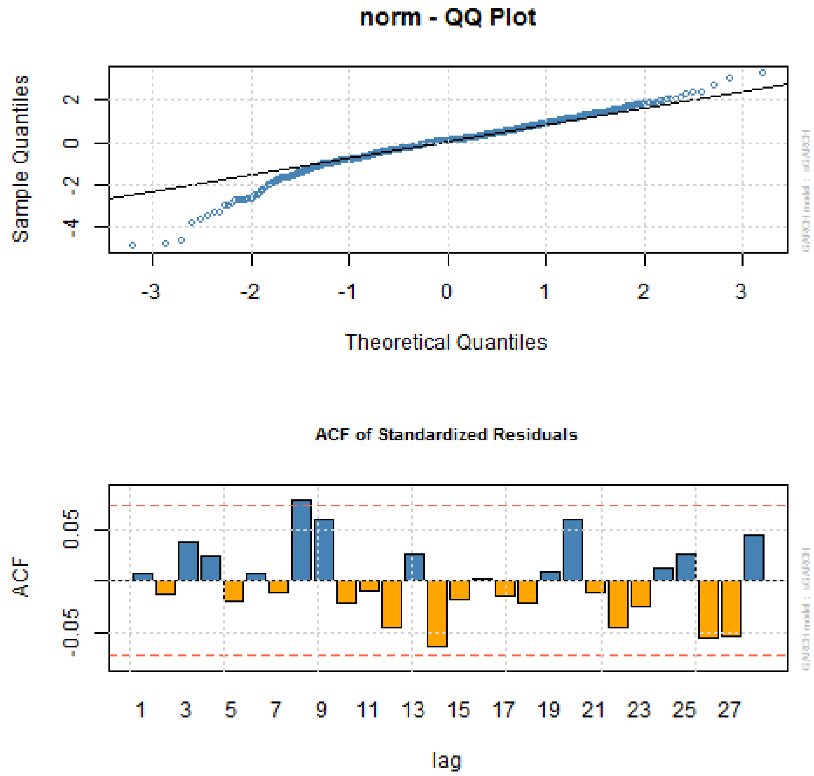

图 9：利用高斯新息拟合的标准化残差的 Q-Q 图和 ACF 图

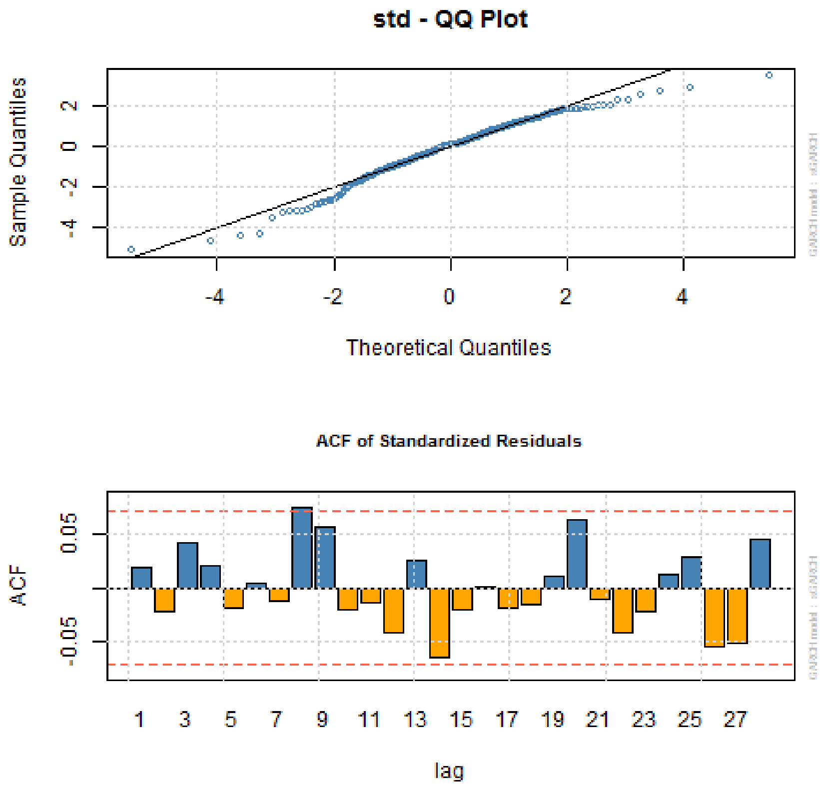

图 10：利用 t 分布新息拟合的标准化残差的 Q-Q 图和 ACF 图

通过 Q-Q 图检验模型标准化残差的分布：

由图 9 的 Q-Q 图可以看出，利用高斯新息拟合的模型的残差的分布与正态分布比较类似，即标准化残差具有正态性；由图 10 的 Q-Q 图可以看出，利用 t 分布新息拟合的模型的残差的分布与 t 分布比较类似，即标准化残差近似服从 t 分布。

通过 ACF 图像检验模型标准化残差的相关性：

由图9图10的ACF图像可以看出，利用两种新息的ARMA(1,1)-GARCH(1,1)模型进行拟合的标准化残差均不具有自相关性，即标准化残差序列是白噪声序列。

2. 利用 AIC 和 BIC 比较高斯新息和 t 分布新息模型

分别由 AIC 和 BIC 的公式进行计算，结果如表 1：

表 1：两种模型的 AIC 和 BIC 值

	AIC	BIC
高斯新息模型	-4532.20	-4504.59
t 分布新息模型	-4609.30	-4577.08

高斯新息模型的 AIC 值为-4532.20，BIC 值为-4504.59；t 分布新息模型的 AIC 值为-4609.30，BIC 值为-4577.08。经过比较发现 t 分布新息模型拟合得更好。

四、基于 3 中的模型，预测一天后的收益率以及方差，计算 VaR 和 ES 指标（α=0.95）。

1. 高斯新息模型

绘制利用高斯新息拟合的 ARMA(1,1)-GARCH(1,1)模型的波动率时序图、α=0.95 的 VaR 时序图以及残差时序图如图 11、图 12 以及图 13。

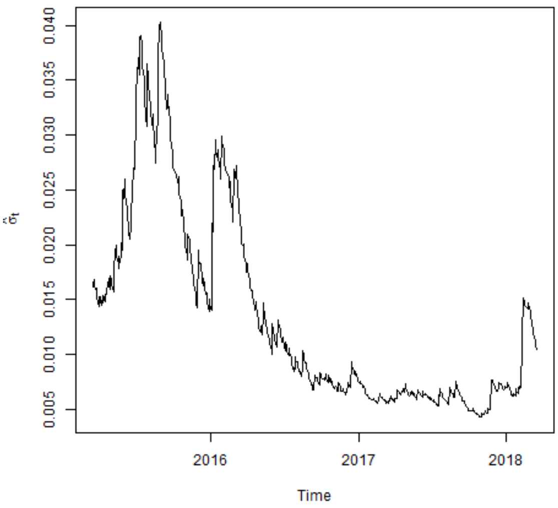

图 11：高斯新息波动率时序图

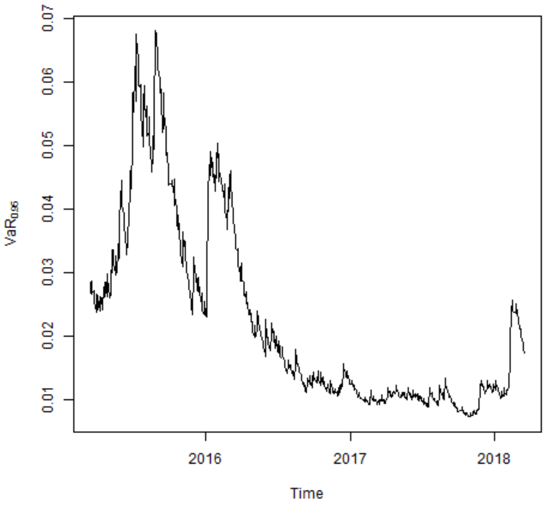

图 12：高斯新息 VaR 时序图

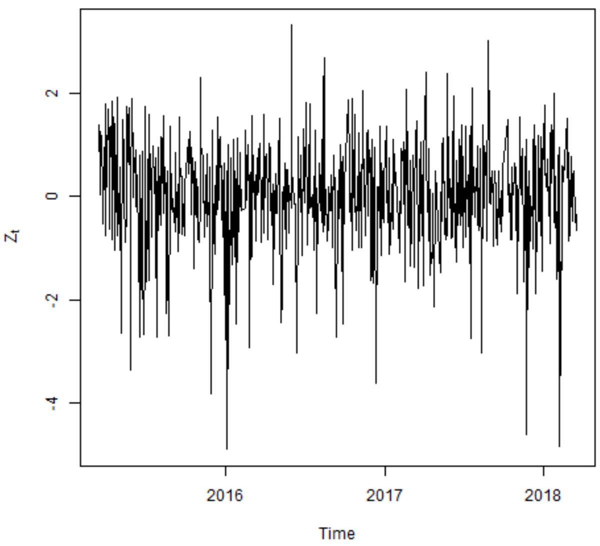

图 13：高斯新息残差时序图

利用高斯新息模型进行后一天的预测结果如图 14：

```
*------------------------------------*
*       GARCH Model Forecast         *
*------------------------------------*
Model: sGARCH
Horizon: 1
Roll Steps: 0
Out of Sample: 0

0-roll forecast [T0=2018-03-16]:
        Series   Sigma
T+1 -3.134e-05 0.01025
```

图 14：高斯新息模型预测结果

当 α=0.95 时，由高斯新息模型预测后一天的收益率 $\mu_{t+1}=-3.13\times10^{-5}$，标准差 $\sigma_{t+1}=0.01025$；

由公式：$VaR_{\alpha}^{t+1}=\mu_{t+1}+\sigma_{t+1}F_{Z}^{\leftarrow}(\alpha)$ 以及公式 $ES_{\alpha}^{t+1}=\mu_{t+1}+\sigma_{t+1}ES_{\alpha}(Z)$，计算出的 VaR 值为 0.015，ES 值为 0.019。

2. t 分布新息模型

绘制利用 t 分布新息拟合的 ARMA(1,1)-GARCH(1,1)模型的波动率时序图、α=0.95 的 VaR 时序图以及残差时序图如图 15、图 16 以及图 17。

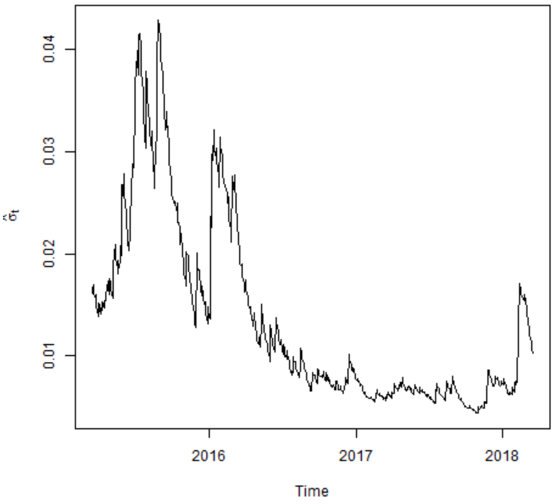

图 15：t 分布新息波动率时序图

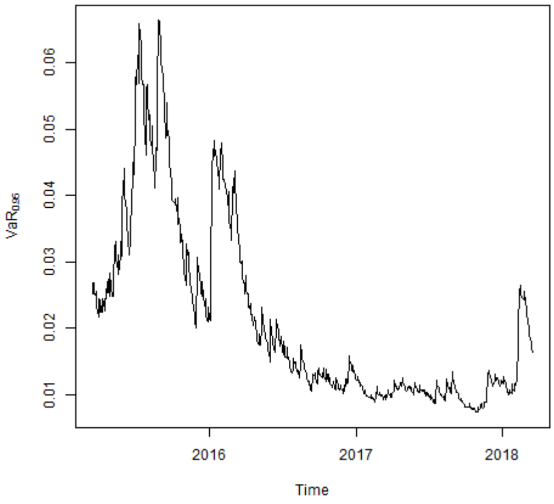

图 16：t 分布新息 VaR 时序图

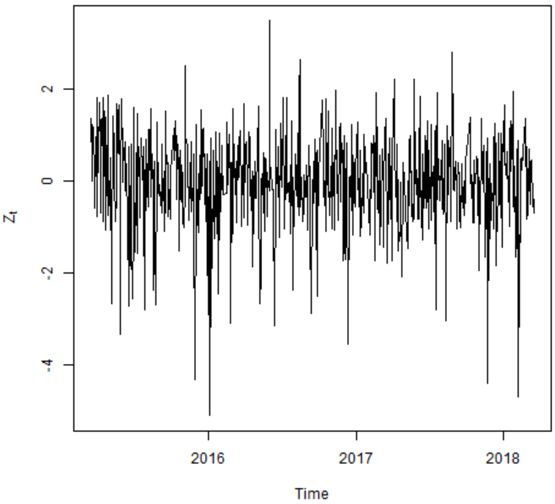

图 17：t 分布新息残差时序图

利用高斯新息模型进行后一天的预测结果如图 18：

```
*-------------------------------------*
*       GARCH Model Forecast          *
*-------------------------------------*
Model: sGARCH
Horizon: 1
Roll Steps: 0
Out of Sample: 0

0-roll forecast [T0=2018-03-16]:
       Series    Sigma
T+1 0.0004877 0.01009
```

图 18：t 分布新息新息模型预测结果

当 α=0.95 时，由 t 分布新息模型预测后一天的收益率 $\mu_{t+1}=0.0004877$，标准差 $\sigma_{t+1}=0.01009$；

由公式：$VaR_{\alpha}^{t+1}=\mu_{t+1}+\sigma_{t+1}F_Z^{\leftarrow}(\alpha)$ 以及公式 $ES_{\alpha}^{t+1}=\mu_{t+1}+\sigma_{t+1}ES_{\alpha}(Z)$，计算出的 VaR 值为 0.014，ES 值为 0.019。

两种模型的结果如表 2：

表 2：两种模型的结果

	μ_{t+1}	σ_{t+1}	$VaR_{0.95}$	$ES_{0.95}$
高斯新息模型	-3.134e-05	0.01025	0.015	0.019
t 分布新息模型	0.0004877	0.01009	0.014	0.019

当 α=0.95 时，利用 t 分布新息拟合模型进行预测和利用高斯新息拟合模型进行预测的结果有所不同：在 t 分布新息模型下预测结果为正，代表收益增加、在高斯新息模型下预测结果为负，代表收益减小；在 t 分布新息模型下计算出的 VaR 值小于在高斯新息模型下计算出的 VaR 值，表明在 95%的置信水平下，市场波动正常时，利用 t 分布新息模型预测的损失较小，即有 95%的把握判断在下一个交易日内的损失小于由高斯新息模型预测的损失。

感谢中国人民大学统计学院杨昊宇同学提供的数据、代码和报告