



1928-2018



第二章

R与R studio 介绍

上节内容回顾

- 数据的定义、大数据的定义
- 为什么学（Why）？
- 如何学（How）？
- 学什么（What）？

■ 本章内容

1、如何安装R软件和R Studio

2、R的基本操作介绍

3、如何使用R工具包

1、R 软件的安装

- R软件的下載

- 官方网址: <https://www.r-project.org/>
- 通常R软件每年有两个版本更新, 目前最新的版本是R 4.0.2, 发布时间是2020年6月22日。



R 的界面

R Console (64-bit)

文件 编辑 其他 程序包 窗口 帮助

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"  
Copyright (C) 2020 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)
```

R是自由软件，不带任何担保。
在某些条件下你可以将其自由散布。
用'license()'或'licence()'来看散布的详细条件。

R是个合作计划，有许多人为了它做出了贡献。
用'contributors()'来看作者的详细情况
用'citation()'会告诉你如何在出版物中正确地引用R或R程序包。

用'demo()'来看一些示范程序，用'help()'来阅读在线帮助文件，或
用'help.start()'通过HTML浏览器来看帮助文件。
用'q()'退出R。

```
> |
```

R控制台

Information



滚动:

键盘: PgUp, PgDown, Ctrl+方向键, Ctrl+Home, Ctrl+End;
鼠标: 控制滚动条。

编辑:

移动鼠标:

左箭头或Ctrl+B: 左移一字符;

右箭头或Ctrl+F: 右移一字符;

Home或Ctrl+A: 移到行头;

End或Ctrl+E: 移到行尾;

历史命令: 上、下箭头, Ctrl+P, Ctrl+N

删除:

Del或Ctrl+D: 删除在当前位置的字符或所选文本;

Backspace: 删除当前位置之前的字符;

Ctrl+Del或Ctrl+K: 删除当前位置到行尾的所有文字;

Ctrl+U: 删除当前行上所有的文字;

复制和粘贴:

按下鼠标左键选择文字;

用Shift+Del(或Ctrl+C)把选择文字复制到剪贴板;

用Shift+Ins(或Ctrl+V或Ctrl+Y)来粘贴剪贴板的内容(如果有的话);

在控制台中, Ctrl+X表示先复制再粘贴;

其他:

Ctrl+L: 清空控制台;

Ctrl+O或INS: 开关覆盖模式, 初始是关掉的;

Ctrl+T: 将当前所在位置的字符与其左方的字符交换;

注意: 控制台输出可以被缓冲,

用Ctrl+W来开关这一功能。

用ESC来中断解释器的工作。

TAB键可以补全当前单词。

可以用Windows标准热键来转换

绘图设备 (MDI用Ctrl+Tab或Ctrl+F6, SDI用Alt+Tab)

确定

控制台快捷键说明

2、RStudio 的安装

- **Rstudio的优势：**可以让用户在更加友好的界面上操作R软件，获得更好的交互体验。
- Rstudio的下载地址
 - <http://www.rstudio.com/>
 - <https://www.rstudio.com/products/rstudio/download/>
- **注意事项：**安装的时候一定要先安装R软件，然后再安装R studio

<http://www.rstudio.com/>



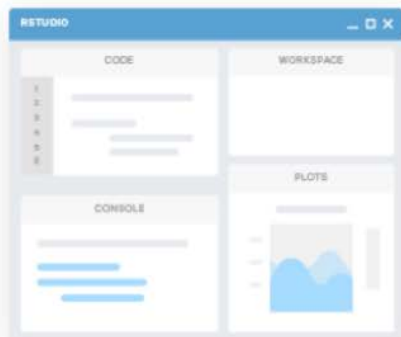
Products

Resources

Pricing

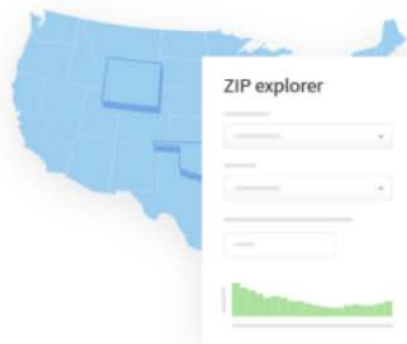
About Us

Blog



RStudio

RStudio makes R easier to use. It includes a code editor, debugging & visualization tools.



Shiny

Shiny helps you make interactive web applications for visualizing data. Bring R data analysis to life.



R Packages

Our developers create popular packages to expand the features of R. Includes ggplot2, dplyr, R Markdown & more.

RStudio 界面

The environment tab shows all the active objects.
The history tab shows a list of commands used so far.

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for loading data and performing statistical tests. The code includes comments and function calls like `library(fBasics)`, `da=read.table("d-ibm3dx7008.txt",header=T)`, and `dim(da)`.
- Console:** Shows the output of the R commands, including the dimensions of the data (9845 rows and 5 columns) and the first row of the data.
- Environment Tab:** Displays the Global Environment with active objects: `aa` (numeric vector), `g` (factor), and `lev` (integer vector).
- Plots Tab:** Shows a scatter plot titled "Edgar Anderson's Iris Data" with four panels: Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width. The plot displays three distinct clusters of data points (red, green, and blue).

The console output shows the following commands and results:

```
d:/Program Files/RStudio/
[33] -0.48698676 1.46134590 -1.09269836 0.59137531 0.46719920 -0.03001046 2.28040199 -0.39784935
[41] 0.26496680 -0.20950195 -0.07238824 0.99856475 -0.30269490 0.59329136 0.73635976 0.13290132
[49] -2.20014636 -0.02841123 -1.45157209 0.77820568 0.18886704 0.49191351 0.53341601 0.01950028
[57] 0.35508844 -1.67646321 2.61936840 -0.01716057 0.92466817 -1.11980213 -1.07918870 -1.92181614
[65] -0.72673988 1.09626176 -0.67038154 1.01787407 -0.02746193 -1.38357762 1.75584299 1.73340015
[73] -1.68227119 0.71354307 -0.01310410 1.16051823 0.43017798 -0.92318433 2.08388874 -0.54806375
[81] -1.09870863 0.44254304 -1.03610991 -2.21728922 -1.29192252 -0.26582230 1.22250745 1.25599664
[89] 0.51512539 -0.56141258 0.74066474 -0.82322423 0.42101535 0.33787949 1.22773648 1.19597528
[97] -0.06310911 -1.43617949 -0.56643957 0.67224100 0.67224100 0.67224100 0.67224100 0.67224100
> head(aa)
[1] 0.07523262 0.02390498 -0.74013386 -0.38178050 0.75385954 -0.22749984
> class(aa)
[1] "numeric"
>
```


R中常见的数据类型

常用数据类型

	类型	说明
1	字符 (character)	它们常常被引号包围
2	数字 (numeric)	实数向量
3	整数 (integer)	整数向量
4	逻辑 (logical)	逻辑向量 (TRUE=T、FALSE=F)
5	复数 (complex)	复数
6	列表 (list)	S 对象的向量
7	因子 (factor)	常用于标记样本

- **data.frame (数据框):** 可理解为松散的数据集，可由不同类型的列（数值、因子、字符等）组成的类矩阵。可以理解为excel中的表格；matlab中也有table类型。

■ 本章内容

1、如何安装R软件和R Studio

2、R的基本操作介绍

3、如何使用R工具包

R 操作简介



- R学习资料包括
 - Additional Note-Basic Operations in R.pdf
 - Additional Note-Introduction to Rstudio.pdf
 - 153分钟学会R.pdf
- R基本操作

R 的基本操作

- 见Additional Note-Basic Operations in R.pdf
- 课堂演示


Rstudio操作介绍

- Additional Note-Introduction to Rstudio.pdf
- 有用的网址:
<https://dss.princeton.edu/training/>




Introduction to RStudio
(v 1.3)

Shield
Oscar Torres-Reyna
otorres@princeton.edu



August 2013

<http://dss.princeton.edu/training/>



■ 本章内容

- 1、如何安装R软件和R Studio
- 2、R的基本操作介绍
- 3、如何使用R工具包

R 工具包(packages)

- R有很多的工具包可供使用，截止目前，R大约有16,285个工具包。通过这些软件包，R的使用者可以很方便地做各种分析。
- R工具包安装：
 - 1、菜单操作：tools菜单->install packages
 - 2、控制台操作：console->packages->install
 - 3、命令操作：install.packages(“car”)，安装car工具包
- R工具包调用：
 - 命令操作：library(car)

[A3](#)
[aaSEA](#)
[AATtools](#)
[ABACUS](#)
[abbyyR](#)
[abc](#)
[abc.data](#)
[ABC.RAP](#)
[abcADM](#)
[ABCanalysis](#)
[abcdeFBA](#)
[ABCOptim](#)
[ABCP2](#)
[abcrf](#)
[abcrlda](#)
[abctools](#)
[abd](#)
[abdiv](#)
[abe](#)
[abf2](#)
[ABHgenotypeR](#)

Available CRAN Packages By Name

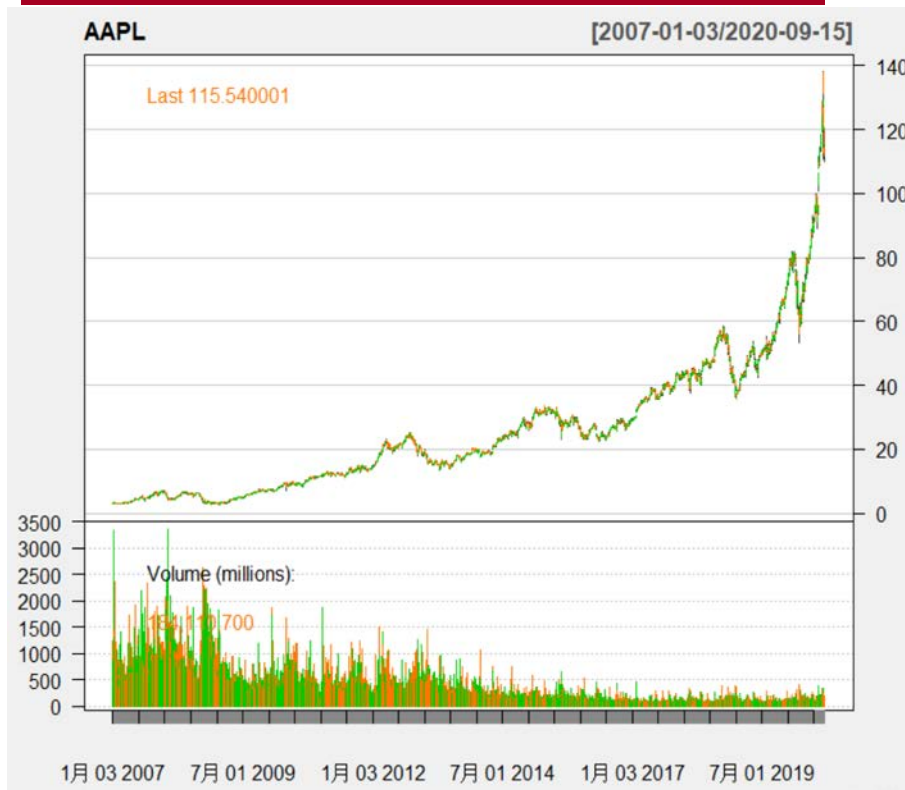
[ABCDEF](#)[GHIJ](#)[KLMN](#)[OPQR](#)[STUV](#)[WXYZ](#)

Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
Amino Acid Substitution Effect Analyser
Reliability and Scoring Routines for the Approach-Avoidance Task
Apps Based Activities for Communicating and Understanding Statistics
Access to Abbyy Optical Character Recognition (OCR) API
Tools for Approximate Bayesian Computation (ABC)
Data Only: Tools for Approximate Bayesian Computation (ABC)
Array Based CpG Region Analysis Pipeline
Fit Accumulated Damage Models and Estimate Reliability using ABC
Computed ABC Analysis
ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
Implementation of Artificial Bee Colony (ABC) Optimization
Approximate Bayesian Computational Model for Estimating P2
Approximate Bayesian Computation via Random Forests
Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis
Tools for ABC Analyses
The Analysis of Biological Data
Alpha and Beta Diversity Measures
Augmented Backward Elimination
Load Gap-Free Axon ABF2 Files
Easy Visualization of ABH Genotypes

例1: quantmod 工具包

- Quantmod工具包：可以从一些开放数据源直接下载金融数据，例如雅虎财经、谷歌财经以及联邦储备经济数据库(FRED)等。
- Quantmod工具包的安装： `install.packages("Quantmod")`
- Quantmod工具包的应用：
 - `library(quantmod)` # Load the package
 - `getSymbols("AAPL")` # Download daily prices of Apple stock from Yahoo, default source is Yahoo.
 - `dim(AAPL)` # (dimension): See the size of the downloaded data.
 - `head(AAPL)` # See the first 6 rows of the data
 - `tail(AAPL)` # See the last 6 rows of the data
 - `chartSeries(AAPL,theme="white")` # Plot the daily price and volume
 - `chartSeries(AAPL)#` Giving the same plot with black background.
 - `getSymbols("AAPL",from="2005-01-02", to="2010-12-31")` # set the range of the date
 - `head(AAPL)`

例1: quantmod 工具包



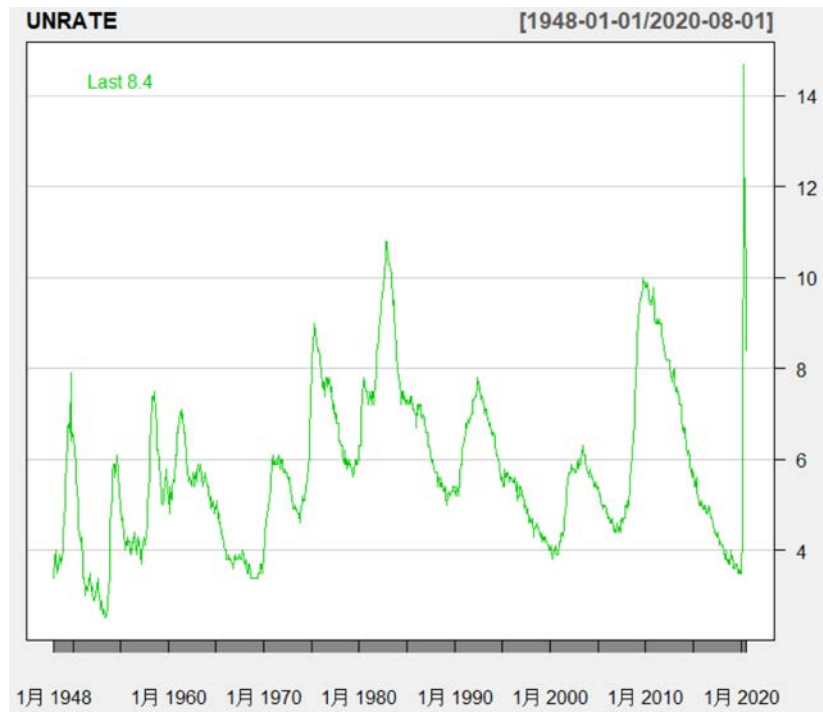
例1: quantmod 工具包

- Quantmod工具包的应用:

- `getSymbols("UNRATE",src="FRED")` #Download unemployment rates from FRED.
- `head(UNRATE)` # See the first 6 rows of the data
- `chartSeries(UNRATE,theme="white")` # Plot monthly unemployment rates
- `tail(UNRATE,8)` # See the last 8 rows of the data

```
> head(UNRATE)
              UNRATE
1948-01-01    3.4
1948-02-01    3.8
1948-03-01    4.0
1948-04-01    3.9
1948-05-01    3.5
1948-06-01    3.6

> tail(UNRATE,8)
              UNRATE
2020-01-01    3.6
2020-02-01    3.5
2020-03-01    4.4
2020-04-01   14.7
2020-05-01   13.3
2020-06-01   11.1
2020-07-01   10.2
2020-08-01    8.4
```



例1: quantmod 工具包

- Quantmod工具包的应用:
 - getSymbols("^TNX") # Download CBOE 10-year Treasures Notes
 - head(TNX)
 - tail(TNX)
 - chartSeries(TNX,theme="white",TA=NULL) # Obtain plot without volume.

> head(TNX)

	TNX.Open	TNX.High	TNX.Low	TNX.Close	TNX.Volume	TNX.Adjusted
2007-01-03	4.658	4.692	4.636	4.664	0	4.664
2007-01-04	4.656	4.662	4.602	4.618	0	4.618
2007-01-05	4.587	4.700	4.583	4.646	0	4.646
2007-01-07	NA	NA	NA	NA	NA	NA
2007-01-08	4.668	4.678	4.654	4.660	0	4.660
2007-01-09	4.660	4.670	4.644	4.656	0	4.656

> tail(TNX)

	TNX.Open	TNX.High	TNX.Low	TNX.Close	TNX.Volume	TNX.Adjusted
2020-09-09	0.690	0.710	0.674	0.703	0	0.703
2020-09-10	0.705	0.723	0.679	0.685	0	0.685
2020-09-11	0.692	0.697	0.664	0.669	0	0.669
2020-09-13	NA	NA	NA	NA	NA	NA
2020-09-14	0.676	0.676	0.653	0.671	0	0.671
2020-09-15	0.685	0.689	0.674	0.679	0	0.679

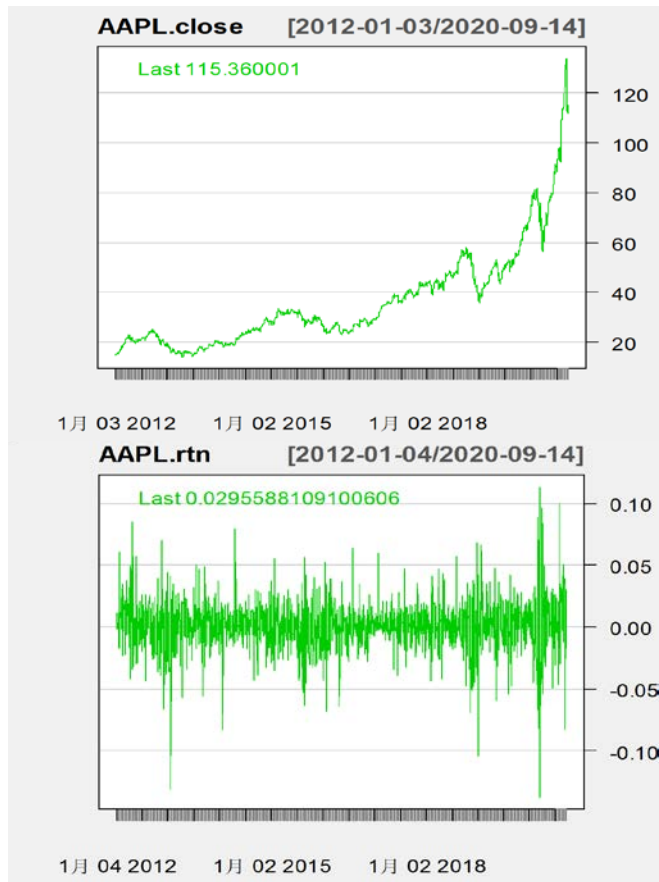


例2: rugarch工具包

- **rugarch工具包**: 广义自回归条件异方差（GARCH）类模型的识别、估计、预测和模拟的工具包，功能强大，是金融数据建模的重要工具包之一。
 - The rugarch package aims to provide a flexible and rich univariate GARCH modelling and testing environment. Modelling is a simple process of defining a specification and fitting the data. Inference can be made from summary, various tests and plot methods, while the forecasting, filtering and simulation methods complete the modelling environment.
- rugarch工具包的安装: `install.packages("rugarch")`
- rugarch工具包的调用: `library("rugarch")`
- rugarch工具包使用手册: <https://www.r-project.org/>

例2: rugarch工具包

- **rugarch工具包的应用**
 - `install.packages("rugarch")`
 - `library(rugarch)`
 - `library(quantmod)`
 - `getSymbols("AAPL",from="2012-01-03",to="2020-9-15")`
#Specify period
 - `AAPL.rtn=diff(log(AAPL$AAPL.Adjusted))` # Compute log returns
 - `AAPL.close = AAPL$AAPL.Close`
 - `chartSeries(AAPL.close,theme="white")` # Get the plot of close price
 - `chartSeries(AAPL.rtn,theme="white")` # Get the plot of log returns
 - `spec = ugarchspec(mean.model=list(armaOrder=c(0,0),`
include.mean=T),distribution.model="norm") # model specification
 - `aapl_ret = zoo(AAPL.rtn)` # To get the zoo type data
 - `gfit_norm = ugarchfit(spec=spec,data=aapl_ret[2:2189])`
 - `plot(gfit_norm)` # Get the plots associated with the GARCH fitting



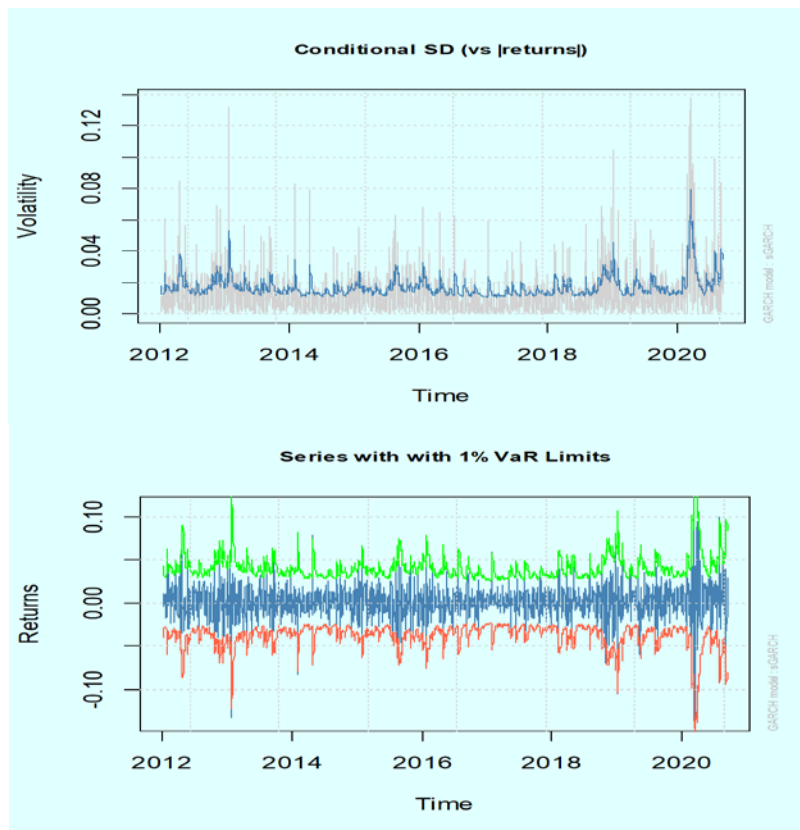
例2: rugarch工具包

- rugarch工具包的应用

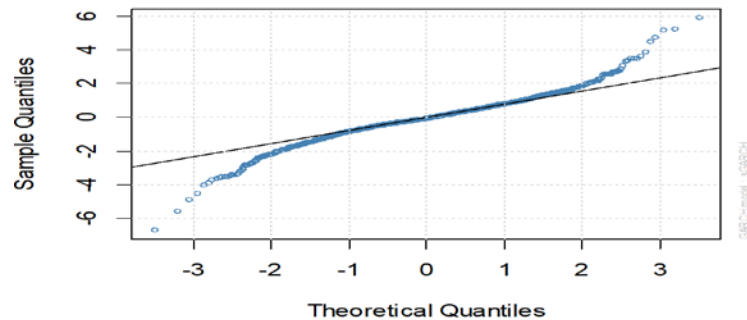
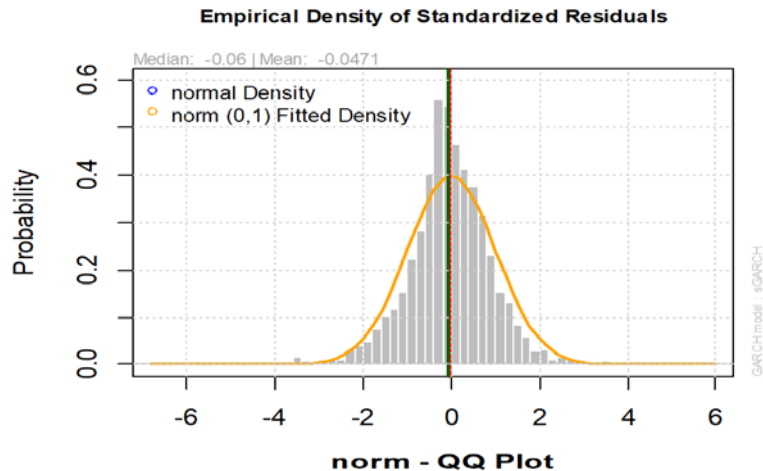
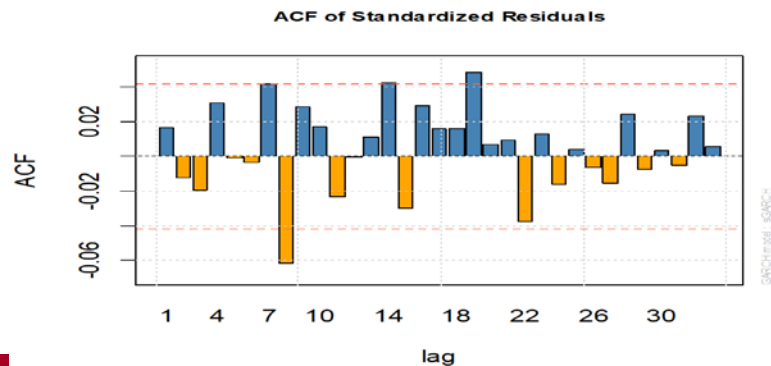
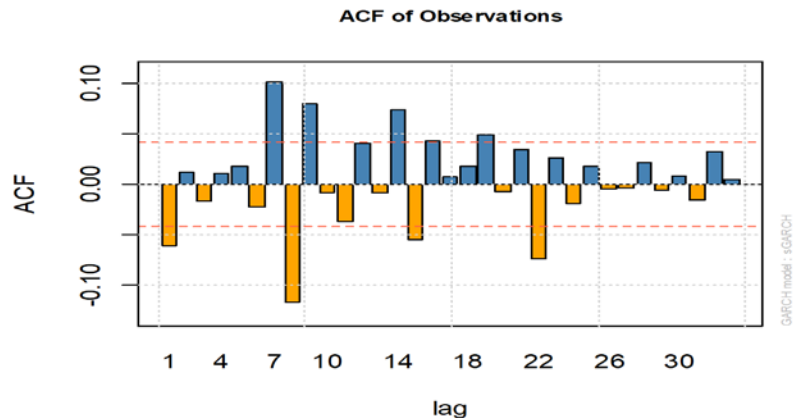
```
> plot(gfit_norm)
```

Make a plot selection (or 0 to exit):

- 1: Series with 2 Conditional SD Superimposed
- 2: Series with 1% VaR Limits
- 3: Conditional SD (vs |returns|)
- 4: ACF of Observations
- 5: ACF of Squared Observations
- 6: ACF of Absolute Observations
- 7: Cross Correlation
- 8: Empirical Density of Standardized Residuals
- 9: QQ-Plot of Standardized Residuals
- 10: ACF of Standardized Residuals
- 11: ACF of Squared Standardized Residuals
- 12: News-Impact Curve



例2: rugarch工具包





感谢大家的聆听！