

第 1 章 R 语言概述

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

R 语言在金融计量中的角色

- R: 统计计算与图形绘制语言 + 完整开发环境
- 在金融计量经济学中的优势:
 - 强大的回归、时间序列、贝叶斯分析等工具
 - 灵活且高质量的可视化能力
 - 便于处理海量与高维金融数据
- 典型应用场景:
 - 风险管理 (VaR、ES、压力测试)
 - 投资组合优化
 - 衍生品定价与对冲
 - 金融市场预测与资产定价实证

R 的开源生态

- 完全开源：基于 GNU GPL 许可证
- R 核心开发团队（R Core Team）维护语言本身
- CRAN（综合 R 档案网络）：
 - 提供 R 源代码、二进制安装文件
 - 托管上万扩展包（统计、机器学习、金融等）
- R 基金会：非营利组织，为 R 的长期发展提供支持
- 开源带来的优势：
 - 可自由审查与修改代码
 - 促进学术合作与可重复研究

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

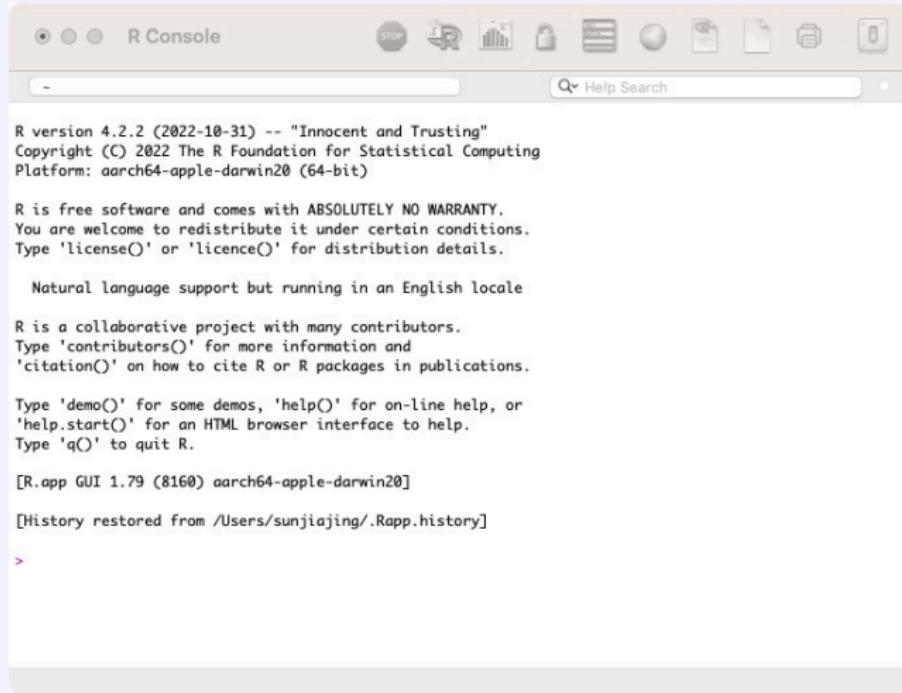
S 语言简史

- 1970s: 钱伯斯等在贝尔实验室开发 S 语言
- 最初: 作为 Fortran 库的一部分, 为内部统计分析服务
- 1980s: S 在实验室外分发, 出版《S》系列专著
- 1988: 核心实现用 C 重写 (S3), 之后发展为 S4 面向对象系统
- S 及其软件生态曾获 ACM 软件系统奖
- 商业化版本: S-PLUS (后来被 TIBCO 收购)

R 语言的起源与发展

- 1990s: Ross Ihaka 与 Robert Gentleman 在奥克兰大学开发 R
- 目标:
 - 免费的 S 语言“方言”
 - 开源、跨平台、易于教学与研究
- 1995: 首个公开版本, 以 GNU GPL 方式发布
- 1997: R Core Team 成立; CRAN 建立
- 2003: R 发布稳定版本; R 基金会成立
- 如今: 主流统计计算平台之一, 广泛应用于统计、数据科学、金融等领域

R 软件界面示意



R 软件界面示意（版本 4.2.2）

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

什么是 IDE ?

IDE (Integrated Development Environment)

为程序员提供集成化的开发环境，一般包含：

- 源代码编辑器
 - 命令/控制台窗口
 - 对象浏览器、调试器
 - 版本控制 (Git) 集成
-
- R 主要通过命令行交互，IDE 能显著提升使用效率
 - Python、Julia 等交互式语言也有各自 IDE (如 PyCharm、Jupyter)

RStudio 概述

- 由 Posit PBC (原 RStudio, PBC) 开发的 R 专用 IDE
- 主要版本：
 - RStudio Desktop (Windows/macOS/Linux)
 - RStudio Server (运行在 Linux 上, 浏览器访问)
- 主要特点：
 - 代码编辑、控制台、对象浏览、绘图整合在一处
 - 包管理、版本控制、项目管理
 - 与 Shiny、R Markdown、Quarto 紧密集成
- 首席科学家：Hadley Wickham (tidyverse 重要作者)

RStudio 界面结构

The screenshot displays the RStudio interface with four main panels:

- 代码编辑器 (Code Editor):** Shows an R script named "Untitled1.R" with the following code:


```
1 x <- rnorm(1000)
2 hist(x)
```
- 环境面板 (Environment Panel):** Shows the global environment with the variable "x" defined as a numeric vector of length 1000.
- 控制台 (Console):** Shows the R startup message and a histogram of the variable "x". The histogram has a maximum frequency of approximately 200 at x=0.
- 选项卡面板 (Plots Panel):** Displays a histogram titled "Histogram of x" with the x-axis labeled "x" and the y-axis labeled "Frequency".

RStudio 四大面板

- 左上: **Source Editor**
 - 编写、编辑、运行脚本/Markdown/Quarto 文档
- 左下: **Console**
 - 交互式执行命令，试验代码片段
- 右上: **Environment / History**
 - 查看当前对象、查看命令历史
- 右下: **Plots / Files / Packages / Help / Viewer**
 - 查看图形、管理包、阅读帮助文档等

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

简单算术与类型

```
# Basic arithmetic
8 + 4 # addition
9 - 2 # subtraction
7 * 6 # multiplication
120 / 40 # division
sqrt(64) # square root
10^2 # power

# Type coercion
3 + TRUE # TRUE is coerced to 1
str(1)
str(as.array(1))
```

- R 是动态类型语言，必要时自动做类型强制（coercion）
- str(): 快速查看对象内部结构的利器

常见数学常数

- 自然常数 e : `exp(1)`
- 圆周率 π : 内置常数 `pi`
- 正负无穷: `Inf`, `-Inf`
- 非数值: `NaN` 表示如 $0/0$ 等未定义结果

自然常数 e 的两种表示

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = \sum_{n=0}^{\infty} \frac{1}{n!}$$

对数收益率（连续复利）

定义

$$r_t = \log \left(\frac{P_t}{P_{t-1}} \right)$$

- 波动大时，对数收益率比简单收益率更稳健
- 具有可加性：

$$\sum_{t=1}^T r_t = \log \left(\frac{P_T}{P_0} \right)$$

- 累计简单收益率：

$$\exp \left(\sum_{t=1}^T r_t \right) - 1$$

- 本书统一用 $\log(\cdot)$ 表示以 e 为底的自然对数

循环结构：for / while / repeat

for 循环

```
for (i in 1:5) {  
  print(i + 1)  
}
```

while 循环

```
i <- 0  
while (i <= 5) {  
  print(i)  
  i <- i + 1  
}
```

repeat + break

```
i <- 0
```

循环的替代：apply 家族

- `apply()`: 对矩阵按行/列操作
- `lapply()`: 对列表/向量元素逐个操作，返回列表
- `sapply()`: 在 `lapply()` 基础上尝试简化结果（向量/矩阵）
- `tapply()`: 按因子分组后，对每组应用函数

```
lapply(0:4, function(a) a + 1)  
sapply(0:4, function(a) a + 1)
```

- 在 R 中，能用向量化或 apply 家族替代循环时，通常更高效、更不容易出错

函数与闭包

```
add_numbers <- function(a, b) {  
  result <- a + b  
  return(result)  
}  
  
add_numbers(5, 3) # 8
```

- 函数有局部作用域：函数内部变量不会污染全局环境
- R 函数是第一类对象：
 - 可以作为参数传入
 - 可以作为返回值
 - 可以存入列表等数据结构
- 闭包 (closure)：函数“记住”其定义时的环境

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

基本数据结构概览

- **向量 (vector)**: 同类型、一维
- **矩阵 (matrix)**: 同类型、二维
- **数组 (array)**: 同类型、多维
- **列表 (list)**: 可容纳不同类型对象
- **数据框 (data.frame)**:
 - 列为向量，行表示观测
 - 每列类型可不同
- **因子 (factor)**: 用于表示分类变量

向量与列表

向量

```
X <- c(2, 4, 6, 8, 10)
print(X)
```

- 一维、同质
- 是 R 的“基本积木”

列表

```
ids <- c(101, 102, 103)
names <- c("Li Hua", "Wang
          Ming", "Zhang Yan")
total <- 3
employeeList <- list(
EmployeeIDs = ids,
Names = names,
TotalEmployees = total
)
```

数据框、矩阵、数组、因子

数据框

```
df <- data.frame(
ID = c(1, 2, 3),
Name = c("Zhang San", "Li Si",
        "Wang Wu"),
Score = c(85, 90, 78)
)
```

矩阵

```
M <- matrix(
1:9,
nrow = 3, byrow = TRUE
)
```

数组

```
B <- array(1:8, dim = c(2,
2, 2))
```

因子

```
fctr <- factor(c("Male",
"Female", "Male"))
```

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

导入数据与基础包

- 常用导入函数：

- `read.csv()`: CSV 文本
- `readxl::read_excel()`: Excel
- `haven::read_sav()`: SPSS 等

- 示例：导入中国 GDP 数据

```
folder <- "Z:/FinancialEconometrics/Rcode"  
setwd(folder)  
China_GDP <- read.csv("China_GDP.csv")  
head(China_GDP)
```

- 基础包 `datasets` 提供内置数据集：

- `mtcars`, `iris`, `airquality` 等
- 使用 `data()` 查看列表

子集、过滤与汇总

- 子集提取：

```
subset_China_GDP <- China_GDP$GDP[2:4]
```

- 条件过滤：

```
filtered <- China_GDP$GDP[China_GDP$GDP > 2000]
```

- dplyr 分组汇总：

```
library(dplyr)
avg_mpg <- mtcars %>%
  group_by(cyl) %>%
  summarise(avg_mpg = mean(mpg),
            .groups = "drop")
```

缺失值插补示例

```
data_set <- data.frame(  
  var1 = c(1, 2, NA, 4, 5),  
  var2 = c(6, NA, 8, 9, 10)  
)  
  
# Mean imputation  
for (col in names(data_set)) {  
  if (is.numeric(data_set[[col]])) {  
    m <- mean(data_set[[col]], na.rm = TRUE)  
    data_set[[col]][is.na(data_set[[col]])] <- m  
  }  
}
```

缺失值插补示例

- 简单方法：均值/中位数插补
- 更复杂方法：
 - kNN 插补：VIM::kNN()
 - 多重插补：mice
 - 回归、卡尔曼滤波等

导出与.RData

- CSV:

```
write.csv(mydata, "mydata.csv", row.names = FALSE)
```

- Excel:

```
library(writexl)  
write_xlsx(mydata, "mydata.xlsx")
```

- SPSS:

```
library(haven)  
write_sav(mydata, "mydata.sav")
```

- 保存 R 对象:

```
save(x, y, z, file = "mydata.RData")  
save.image("workspace.RData")# Save whole workspace
```

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

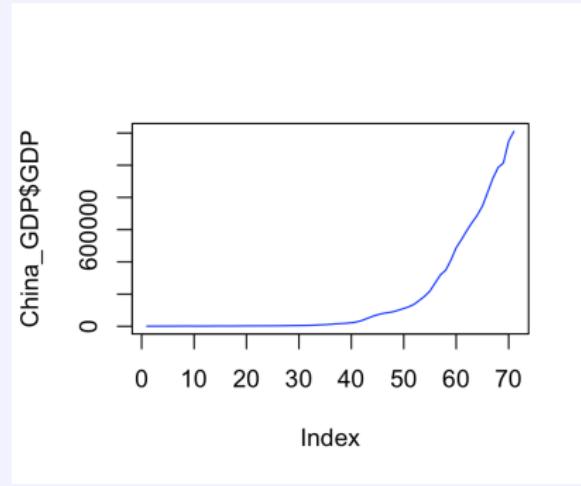
10 API 与开放数据

11 小结

基础折线图：GDP 示例

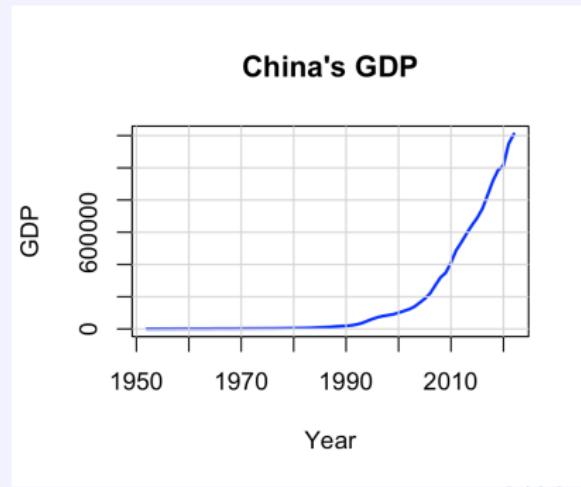
```
plot(China_GDP$GDP, type = "l", col = "blue")
```

- `plot()`: 基础绘图函数
- 将 GDP 向量直接绘制为折线图



时间序列对象与美化

```
GDP.ts <- ts(China_GDP$GDP,  
  start = 1952, frequency = 1)  
plot(GDP.ts, type = "l", col = "blue",  
 xlab = "Year", ylab = "GDP", main = "China's GDP", lwd =  
  2)  
grid(col = "gray85", lty = "solid")
```



其他基础图形

- **散点图**: `plot(x, y)`
- **直方图**: `hist(x)`
- **箱线图**: `boxplot(x)`
- **条形图**: `barplot()`
- **密度图**: `plot(density(x))`
- **QQ 图**: `qqnorm()`, `qqline()`
- **散点图矩阵**: `pairs()`

(示例使用 `mtcars` 数据集)

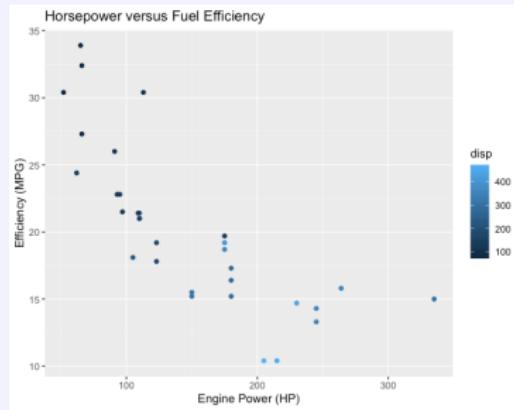
ggplot2 的图形语法

- 基本思想：图形 = 若干“层”的组合
- 主要元素：
 - Data: 数据
 - Aesthetics: 映射 (x, y, color, size, shape, ...)
 - Geoms: 几何对象 (points, lines, bars, boxplots, ...)
 - Facets: 分面
 - Stats: 统计变换 (平滑拟合、汇总等)
 - Coordinates: 坐标系统
 - Themes: 整体外观

简单散点图：ggplot2

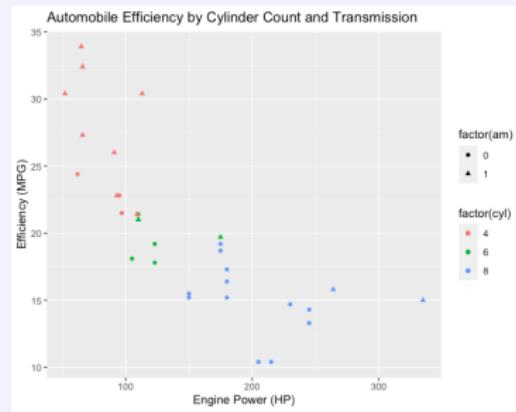
```
library(ggplot2)

ggplot(mtcars, aes(x = hp, y = mpg)) +
  geom_point() +
  labs(title = "Horsepower vs MPG",
       x = "HP", y = "Miles per Gallon")
```



添加颜色、形状、尺寸

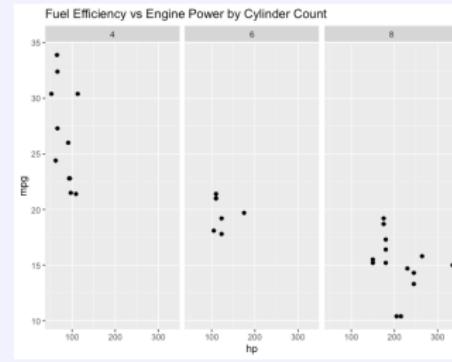
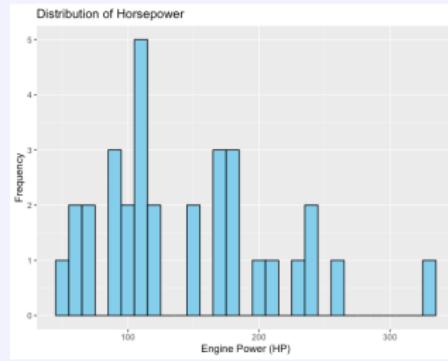
```
ggplot(mtcars,  
       aes(x = hp, y = mpg,  
            col = factor(cyl),  
            shape = factor(am))) +  
  geom_point()
```



直方图与分面

```
ggplot(mtcars, aes(x = hp)) +
  geom_histogram(binwidth = 10,
  fill = "skyblue",
  color = "black")

ggplot(mtcars, aes(x = hp, y = mpg)) +
  geom_point() +
  facet_grid(. ~ cyl)
```



目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

内置随机数生成

- 正态分布: `rnorm(n, mean, sd)`
- 卡方: `rchisq(n, df)`
- t 分布: `rt(n, df)`
- F 分布: `rf(n, df1, df2)`
- 二项分布: `rbinom(n, size, prob)`
- 泊松分布: `rpois(n, lambda)`

```
set.seed(123)
x <- rnorm(10) # N(0, 1)
y <- rnorm(10, 1, 2) # N(1, 2^2)
```

分位数转换（逆 CDF）法

基本思想

- 若 $U \sim \text{Unif}[0, 1]$, 且 F_X 为随机变量 X 的 CDF
- 则 $X = F_X^{-1}(U)$ 服从分布 F_X
- 一般步骤：
 - ① 生成均匀样本 $u \in [0, 1]$
 - ② 计算 $x = F^{-1}(u)$
 - ③ 得到服从目标分布的样本
- 可以用来自定义任意分布的随机数生成器

逆 CDF 法模拟标准正态 (1)

```
set.seed(123456)

# Self-written CDF (demo; usually use pnorm)
compute_phi <- function(z) {
  integrand <- function(t)
    (1 / sqrt(2 * pi)) * exp(-t^2 / 2)
  integrate(integrand,
            lower = -Inf, upper = z)$value
}

inverse_phi <- function(p) {
  f <- function(z) compute_phi(z) - p
  uniroot(f, c(-8, 8))$root
}
```

逆 CDF 法模拟标准正态 (2)

```
u <- runif(1000)
x <- sapply(u, inverse_phi)

hist(x, probability = TRUE,
main = "Standard normal via inverse CDF")
lines(density(x), col = "red", lwd = 2)
```

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

R 帮助系统

- RStudio 中的 **Help** 面板
- 控制台命令：
 - `?rnorm`
 - `help("rnorm")`
 - `example("lm")`
- 帮助页面包括：
 - Description (描述)
 - Usage (用法)
 - Arguments (参数说明)
 - Details (细节)
 - Value、Examples 等
- 学习新函数的第一步：看帮助 + 跑示例

tidyverse 概览

- tidyverse: 一组风格统一的 R 包集合
- 典型成员:
 - ggplot2: 可视化
 - dplyr: 数据操作
 - tidyr: 数据整理
 - readr: 读取文本数据
 - purrr: 函数式编程
 - tibble: 现代数据框

- 安装:

```
install.packages("tidyverse")
```

- 金融计量中的优势:
 - 语法统一, 清晰可读
 - 易于构建可重复的数据处理管线

dplyr 的“动词”与管道

- 行操作: filter(), slice(), arrange()
- 列操作: select(), rename(), mutate(), relocate()
- 分组汇总: group_by() + summarise()
- 管道操作符 %>%:
 - 由 magrittr 引入
 - 让代码“从左到右”阅读

```
library(dplyr)
starwars %>%
  filter(species == "Human") %>%
  select(name, height, mass) %>%
  arrange(desc(height)) %>%
  head(5)
```

tidy: 整洁数据

- 目标: 每列一个变量, 每行一个观测
- 主要函数:

- `gather()` / `pivot_longer()`: 宽 → 长
- `spread()` / `pivot_wider()`: 长 → 宽
- `separate()`: 拆列
- `unite()`: 合列
- `nest()` / `unnest()`: 列表列

```
long <- tidy_dataframe %>%
  tidyverse::gather(Group, Frequency,
  Group.1:Group.3)
```

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

什么是 API ?

API (应用程序编程接口)

- 软件之间通信的“契约”
- 常见：基于 HTTP 的 REST 接口
- 通过 URL + 参数获取 JSON / CSV / XML 数据
- 关键概念：
 - Endpoint (端点)：例如 `https://api.worldbank.org/v2/...`
 - Query parameters: `?format=json&per_page=20000`
 - Rate limit (限流)、分页 (pagination)
 - API Key (密钥) 与认证

在 R 中访问 API 的两条路径

① 使用专门封装包（推荐）

- 世界银行: WDI
- FRED: fredr
- OECD: OECD
- IMF: imfr
- Eurostat: eurostat

② 自己写 HTTP 请求

- httr2 发送请求
- jsonlite 解析 JSON
- 手动处理端点、参数、分页、错误

示例：使用 WDI 获取人均 GDP

```
library(WDI)
library(dplyr)
library(ggplot2)
# Search indicators
WDIsearch("gdp per capita") %>% head()
# China/UK/US GDP per capita 1990-2023
wb <- WDI(
  country = c("CN", "GB", "US"),
  indicator = c(gdppc = "NY.GDP.PCAP.KD"),
  start = 1990, end = 2023
) %>%
  as_tibble() %>%
  arrange(country, year)
ggplot(wb, aes(year, gdppc, color = country)) +
  geom_line()
```

直接请求世界银行 API

```
library(httr2)
library(jsonlite)
library(dplyr)
fetch_wb <- function(iso3, indicator) {
  url <- sprintf(
    "https://api.worldbank.org/v2/country/%s/indicator/%s",
    iso3, indicator
  )
  req <- request(url) %>%
    req_url_query(format = "json", per_page = 20000)
  resp <- req_perform(req)
  j <- resp$body_json(resp, simplifyVector = TRUE)
  as_tibble(j[[2]]) %>%
    transmute(country = country$value, iso3c = country$iso3code
      ,
      year= as.integer(date), value = as.numeric(value))}
```

目录

1 R 语言与金融计量经济学

2 R 与 S 语言的历史

3 RStudio 与 IDE

4 R 编程基础

5 R 的数据结构

6 数据导入、预处理与导出

7 基础可视化与 ggplot2

8 随机数与逆 CDF 法

9 帮助系统与 tidyverse

10 API 与开放数据

11 小结

本章小结

- 了解了 R 与 S 语言的历史、R 的开源生态 (R Core, CRAN, R Foundation)
- 熟悉了 RStudio IDE 的面板与常用功能
- 掌握了：
 - R 基本语法：算术、循环、函数、对数收益率
 - 核心数据结构：向量、列表、数据框、矩阵、数组、因子
 - 数据导入、预处理（子集、过滤、汇总、缺失值插补）与导出
 - 基础可视化与 ggplot2 图形语法
 - tidyverse (dplyr, tidyr) 的数据整理管线
 - 随机数生成与分位数变换法
 - 在 R 中通过 API 获取开放宏观与金融数据
- 为后续金融计量模型与实证分析打下技术基础

第 2 章引言和背景

孙佳婧 † 洪永淼 ‡ 奥利弗·林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

目录

1 金融市场

- 货币市场
- 债券市场
- 股票市场

2 市场类型以及交易方式

- 主要交易所
- 电子交易与订单类型
- 监管框架与全球化

3 收益率

- 收益率的构建
- 股票指数的编制方法
- 收益率的统计性质

4 金融经济学基础

- 效用函数以及风险厌恶
- 均值-方差投资组合模型
- CAPM、APT 与 C-CAPM

5 章节总结

金融市场的基本概念

- 金融市场：涉及货币资金借贷、外汇买卖、有价证券交易、债券与股票发行、黄金等贵金属买卖的复杂网络。
- 被誉为国民经济的“晴雨表”，提供观测与监控经济运行状态的直观指标。
- 通过各类市场的有序运作，为长期项目提供资金，使企业和个人优化资源配置、加强风险管理。

金融市场的主要分类（一）

- 按交易标的：货币市场、债券市场、股票市场、外汇市场、衍生品市场、保险市场、黄金市场等。
- 按交易中介角色：
 - 直接金融市场：经济主体之间直接交易金融资产。
 - 间接金融市场：通过银行、基金等金融中介机构撮合和调节资金供需。
- 按金融工具交易阶段：
 - 发行市场：新金融工具的初次发行与销售。
 - 流通市场：已经发行工具的后续买卖。

金融市场的主要分类（二）

- 按是否有固定交易场所：
 - 场内市场：如证券交易所，具有固定场所和标准化交易程序。
 - 场外市场：多为非标准化合约交易，如部分衍生品。
- 按金融工具本源与从属关系：
 - 传统金融市场。
 - 金融衍生品市场：期货、期权等衍生工具。
- 按地域范围：
 - 国内金融市场。
 - 国际金融市场：跨越国界，涉及多种货币和跨境资本流动。

金融市场与中国经济

- 金融市场是国民经济的“血脉”，保障资金高效流动与资源最优配置。
- 通过支持创新和技术升级，促进实体经济发展与国家战略实施。
- 金融市场稳定关系整体经济稳定：
 - 提供必要金融服务。
 - 保障金融体系安全，促进社会和谐、经济持续健康发展。
- 加快建设金融强国，是中国经济社会发展和全面建设社会主义现代化国家的重要战略任务。

金融市场按工具类型的划分

- 货币市场：主要交易 1 年以内到期的短期工具。
- 债券市场：主要是 1 年以上到期的中长期债券。
- 股票市场：交易上市公司股份。
- 外汇市场：不同货币之间的交易。
- 衍生品市场：期货、期权等衍生工具。
- 商品市场：黄金等贵金属、农产品（如豆粕、玉米）、能源化工产品（如原油）等。

接下来将依次介绍主要市场的运作方式、交易工具及其在全球经济中的作用。

货币市场概览

- 货币市场是固定收益市场的重要组成部分。
- 主要特征：
 - 工具期限较短（通常 1 年以内）。
 - 流动性高、信用风险较低。
 - 多数以折价发行，到期按面值偿还。
- 主要工具包括：
 - 国库券、大额可转让定期存单、欧洲美元存单。
 - 商业票据、银行承兑汇票、回购协议等。

货币市场工具 (1)

定义 (国库券 (Treasury Bills, T-bills))

国库券是一种高流动性的短期政府债务工具，通常以低于面值的价格发行，投资者的收益来自到期时面值与购买价格之间的差额。

定义 (大额可转让定期存单 (Certificates of Deposit, CDs))

大额存单是银行发行的一种定期存款凭证，通常到期前不得提前支取，但可在二级市场转让。

货币市场工具（2）

定义 (欧洲美元存单 (Eurodollar CDs))

以美元计价，由外国银行或美国机构在其境外分支机构发行，不受美联储直接监管。名称源于二战后美元在欧洲银行的广泛使用，与当今“欧洲”地域已无直接对应关系。

定义 (商业票据 (Commercial Paper, CP))

大型企业发行的短期无担保债务工具，常用于在金融市场上直接融资，以替代银行借款。

货币市场工具 (3)

定义 (银行承兑汇票 (Bankers' Acceptances, BA))

由银行承兑并承诺付款的远期汇票，广泛用于国际贸易，有助于在缺乏信用基础的交易中建立信任。通常以贴现形式交易，可在公开市场出售或持有至到期。

定义 (回购协议 (Repurchase Agreements, repos/RPs))

证券卖方承诺在未来以更高价格回购同一证券的交易安排，实质上构成一项有抵押的短期借款。相应的“逆回购”是指证券买方发起、为交易对手提供有抵押贷款的安排。

货币市场的典型利率（美国）

- **联邦基金利率 (Federal Funds Rate)**

- 美国存款类金融机构在联邦基金市场上相互拆借准备金余额的隔夜无担保利率。
- FOMC 设定目标区间，实际成交体现在有效联邦基金利率 (EFFR) 上。

- **伦敦银行同业拆借利率 (LIBOR)**

- 历史上重要的批发货币市场无担保同业资金利率。
- 长期作为全球金融合约的参考利率，美元市场已过渡至 SOFR。

- **担保隔夜融资利率 (SOFR)**

- 美元 LIBOR 的替代基准，由纽约联储发布。
- 基于以美国国债为担保的隔夜回购交易数据，覆盖面广、透明度高。

中国货币市场的主要利率

- **人民银行贷款基准利率**
 - 由中国人民银行设定，用于指导商业银行贷款定价。
 - 利率市场化改革后，银行可在基准利率基础上加减点自主定价。
- **存款基准利率**
 - 用于指导存款定价，可在基准利率基础上适度调整。
- **上海银行间同业拆放利率（SHIBOR）**
 - 反映银行间市场短期资金成本，覆盖隔夜到一年期多个期限。
- **香港银行同业拆息（HIBOR）**
 - 香港银行同业市场的主要基准利率，影响贷款与抵押贷款定价。

债券市场概览

- 债券：在未来特定日期预设支付固定现金流的合约。
- 债券市场规模庞大，例如美国国债总额已达数十万亿美元级别。
- 主要类型：
 - 零息债券、国库券、附息债券。
 - 中长期国债及债券相关衍生品。

债券类型与衍生品

- **零息债券**: 到期一次性偿还本金, 期间无利息支付 (通常折价发行)。
- **国库券**: 原始期限不超过 1 年的政府短期债务。
- **附息债券**:
 - 定期支付票息, 票面利率按面值百分比表示。
 - 到期偿还本金。
- **中长期国债**:
 - 1-10 年为中期, 超过 10 年为长期。
- **衍生品**:
 - 期权、期货、互换、互换期权、利率 cap/floor/collar 等。

联邦机构债务与欧洲债券

- **联邦机构债务：**

- 如吉利美 (Ginnie Mae)、房利美 (Fannie Mae)、房地美 (Freddie Mac)。
- 以住房抵押贷款为基础，提升住房信贷可获得性。
- 存在显性或隐性担保，信用评级较高，融资成本较低。

- **外国债：**

- 在某国本土市场、以该国本币计价、由外国发行人发行。
- 如 Yankee、Samurai 债等。

- **欧洲债券与离岸美元债：**

- 欧洲债券：以某货币计价，但在该货币发行国以外发行。
- 离岸美元债 (Eurodollar Bond)：以美元计价、在美国境外发行。

中国债券市场的主要品种

- **国债：**财政部发行，被视为无风险资产之一。
- **地方政府债券：**
 - 资助地方基础设施等项目。
 - 随规范化管理推进发展迅速。
- **企业债券：**
 - 公司债、中期票据等，用于业务扩张或债务再融资。
- **金融债：**
 - 政策性银行和商业银行发行，包括次级债。
 - 用于提高资本充足率。
- **同业存单：**
 - 银行间市场发行，短期筹资工具，流动性较高。

股票市场基本概念

- 股票代表公司所有权的一部分。
- 普通股股东：
 - AGM 上一股一票，选举董事会。
 - 享有公司剩余收益的剩余索取权。
- 有限责任：
 - 股东最大损失限于原始投资。
 - 债权人不能追索个人资产。
- 上市股票 vs 私募股权：
 - 证券交易所上市为公开交易股票。
 - 未上市为私募股权。

优先股与公司融资结构

优先股：

- 兼具股权与债权特征（可视为夹层资本）。
- 承诺每年支付固定金额，通常无投票权，类似永续债。
- 公司对股息支付无合同义务，但股息累计，在普通股分红前优先支付。
- 可赎回、可转换或股息率可调整（与市场利率挂钩）。
- 与债券利息不同，优先股股息对公司不是税前扣除费用。

存托凭证与跨境投资

存托凭证：

- 代表一定数量外国股份（可为普通股、优先股或债券）的所有权。
- 国内存托银行持有对应股份，并发行可在本国交易的凭证。
- 为投资者提供便捷、成本较低的外国证券投资渠道。
- 通常以美元或欧元计价，面临外汇风险。
- GDR 在美国以外市场交易；ADR 在美国市场交易。

市场指数与 ETF

- 市场指数反映股票市场整体估值水平：
 - 可覆盖几十种或数千种证券。
 - 如道琼斯工业平均指数、标普 500、罗素 3000 等。
- 指数构建方法多样（价格加权、市值加权等），后续章节详细介绍。
- ETF（交易所交易基金）：
 - 允许投资者以类似股票的方式交易一篮子资产。
 - 如 SPDR ETF 跟踪标普 500 指数收益。

股指期货与 E-mini 合约

- E-mini S&P 500:
 - 在 CME Globex 平台交易的股指期货合约。
 - 合约名义价值为标普 500 指数的 50 倍。
 - 1997 年推出，因大型合约价值过大，小型投资者难以参与。
- 特点：
 - 流动性极高，名义成交额长期居前。
 - 对冲基金偏好交易 E-mini 而非传统大合约。
 - 全电子交易，撮合效率高。

中国股指期货市场的发展

- 2010 年：中国金融期货交易所推出沪深 300 股指期货，标志股指期货市场起步。
- 2015 年股市异常波动：
 - 2015/6/15–8/26，上证指数从 5178 点跌至 2850 点，跌幅约 45%。
 - 股指期货被部分分析人士认为加剧了波动。
 - 2015/9/7 起，对非套保交易提高保证金至 40%，10 手以上视为异常交易，成交量一度暴跌 99% 以上。
- 短期抑制了波动率溢出，但显著限制市场功能。

监管与市场功能的动态平衡

- 2017 年起，中金所逐步调整沪深 300、中证 500、上证 50 股指期货交易规则：
 - 放宽异常交易认定。
 - 下调交易保证金和手续费。
- 政策逻辑：
 - “稳定压倒一切”：在极端波动时优先控制风险、防止扩散。
 - 市场平稳后逐步恢复市场功能，提升风险管理能力和价格发现效率。
- 体现新时代金融监管在风险可控前提下促进市场发展的理念。

股指期货与计量经济分析

- 计量模型可用于：
 - 量化股指期货市场的风险与波动特征。
 - 分析股指期货与现货市场的联动与溢出效应。
- 政策应用示例：
 - 利用模型评估不同保证金比例与手数限制的影响。
 - 在防范过度投机和系统性风险的同时，保持市场流动性与价格发现功能。
- 目标：在维护金融安全与稳定的前提下，释放金融市场活力，服务实体经济高质量发展。

目录

1 金融市场

- 货币市场
- 债券市场
- 股票市场

2 市场类型以及交易方式

- 主要交易所
- 电子交易与订单类型
- 监管框架与全球化

3 收益率

- 收益率的构建
- 股票指数的编制方法
- 收益率的统计性质

4 金融经济学基础

- 效用函数以及风险厌恶
- 均值-方差投资组合模型
- CAPM、APT 与 C-CAPM

5 章节总结

三类典型市场结构

定义 (经纪市场)

由拥有专业知识的经纪人提供搜索服务，通过收取佣金撮合买卖双方。例如：房地产市场、证券发行市场（投资银行作为承销商）、大宗股票的“楼上”市场等。

定义 (做市商市场)

做市商以自营账户交易，建立库存，通过报出买入价与卖出价提供流动性。投资者只需查看并比较报价，典型为报价驱动市场，如 OTC 市场。

电子拍卖市场

定义 (电子拍卖市场)

所有交易者在单一场所（物理或电子）集中交易。订单通过匹配引擎按到达和匹配规则自动执行，属于订单驱动市场。投资者无需在不同做市商之间搜索价格。

- 电子限价订单簿是电子拍卖市场的典型形式。
- 大幅降低搜索成本和中介费用。
- 为算法交易和高频交易提供基础设施。

纳斯达克 (NASDAQ) 的演变

- 起源：最初作为场外交易市场，采用做市商报价制度。
- 1971 年之前：
 - 报价手工记录，通过“粉红单”每日发布。
- 1971 年：NASDAQ 自动报价系统成立。
 - 通过计算机网络连接经纪人和做市商，实时显示和更新报价。
 - 初期仅提供报价，交易仍需协商。
- 如今：已发展为具备全电子交易平台的交易所，绝大多数交易无需人工协商。

纽约证券交易所 (NYSE)

- 传统模式：
 - 每种证券由一位专家维护限价订单簿。
 - 在流动性不足时，专家作为做市商介入，维护公平有序，并从佣金和价差中获利。
- 电子化后：
 - 订单可直接提交给专家系统自动执行。
 - 大宗交易仍可能由“楼上”经纪人协商撮合。
- 债券交易：
 - NYSE 运营自动债券系统，但绝大多数债券仍在场外经做市商网络交易。

伦敦证券交易所 (LSE)

- 1997 年之前：
 - 类似早期纳斯达克，为场外市场提供自动报价系统。
 - 证券公司兼任经纪人和做市商。
- 1997 年之后：
 - 主要采用电子限价订单簿进行交易撮合。
 - 对于大宗或流动性较差的证券，仍依赖做市商成交。

中国主要证券交易所

- 上海证券交易所与深圳证券交易所：
 - 兼具经纪市场与做市商市场的特征。
 - 证券公司作为经纪人，为客户提供交易服务。
 - 在部分品种中存在做市商，通过库存与双向报价提供流动性。
- 电子交易为主：
 - 大量交易通过电子系统完成。
 - 正在探索包括暗池在内的创新交易机制，服务大宗交易和降低价格冲击。

电子交易与暗池

- 交易与通信成本下降、信息更易获取、线上经纪低佣金 → 电子交易取代人工撮合。
- 高频交易（HFT）与算法交易快速发展。
- 暗池交易与经纪人交叉网络：
 - 交易前无透明度，参与者看不到订单簿。
 - 价格常设定在公开市场买卖报价中点。
 - 适合大宗交易且不希望显著影响价格的投资者。

电子市场中的订单类型

基本订单类型：

- 市场订单（market order）：
 - 以当前市场价格立即成交的买入或卖出订单。
- 限价订单（limit order）：
 - 买入：在限价或更低价成交。
 - 卖出：在限价或更高价成交。
 - 若价格不具竞争力，可能长期不成交。
- 停止订单（stop order）：
 - 触发价达到后方转为市价或限价订单。
 - 包括停止买入（stop-buy）、止损卖出（stop-loss）等。

高级订单类型

- 冰山订单 (iceberg order):
 - 仅显示部分挂单数量。
 - 成交后剩余数量再逐步显露，减少信息泄露。
- 挂钩订单 (pegged order):
 - 价格随参考价格（如买卖差价中点）自动调整。
 - 有利于保持相对位置而非绝对价格。

时间条件：

- 日订单：当日收盘自动失效。
- 开放订单：有效期可达数月，除非被取消或成交。

GTC 与 FOK

“直到取消” (Good Till Cancelled, GTC) 与 “全部成交或取消” (Fill or Kill, FOK) 是两类常见时间/执行条件：

- GTC：订单一经下达，除非被交易者取消或完全执行，否则持续有效，可跨多个交易日。
- FOK：要求在订单被接收时立即全额成交，否则将被完全取消，不产生部分成交。

全球监管框架：MiFID 与 NMS

- 金融市场全球化背景下，交易所间竞争加剧。
- 欧洲：MiFID（金融工具市场指令）
 - 2007 年实施，2014 年修订，MiFID II 自 2018 年起实施。
 - 着重提升透明度、保护投资者、促进竞争与跨境金融活动规范。
- 美国：Reg NMS（全国市场体系规章）
 - 2005 年推出，提升市场效率和透明度，保护投资者并促进市场整合。
- 智能订单路由（SOR）与多市场并存：
 - 加强市场间互联互通。
 - 流动性提供者竞争加剧，弱化单一做市商垄断地位。

中国资本市场监管与多层次市场体系

- 基础法律：《证券法》
 - 规范证券发行、交易、信息披露与监管。
 - 目标：保护投资者、维护公平公正的市场秩序、促进健康发展。
- 监管机构：中国证监会
 - 通过规章和指引细化《证券法》。
 - 对证券公司和市场参与者合规性进行监督。
- 多层次资本市场体系：
 - 主板、创业板、科创板等，适配不同企业类型与融资需求。
 - 不同板块采用差异化上市标准和监管要求。
 - 原中小板已于 2021 年并入深交所主板。

目录

1 金融市场

- 货币市场
- 债券市场
- 股票市场

2 市场类型以及交易方式

- 主要交易所
- 电子交易与订单类型
- 监管框架与全球化

3 收益率

- 收益率的构建
- 股票指数的编制方法
- 收益率的统计性质

4 金融经济学基础

- 效用函数以及风险厌恶
- 均值-方差投资组合模型
- CAPM、APT 与 C-CAPM

5 章节总结

收益率的作用与本节内容

- 收益率是衡量投资效果的直接指标：
 - 对个人投资者、金融分析师、基金经理与企业决策者都至关重要。
- 收益率的波动率常被视作风险的重要度量。
- 收益率是研究：
 - 资本市场行为、
 - 有效市场假说、
 - 资产定价模型（如 CAPM）的基础数据。
- 本节重点：
 - 收益率的构建；
 - 股指与对应收益率；
 - 收益率的统计性质。

简单收益率的定义与分解

定义 (简单收益率)

时间 s 到 t 之间的简单总收益率为

$$\mathcal{R}_{s,t} = \frac{P_t + D_t}{P_s},$$

其中 D_t 为持有期 $[s, t]$ 内支付的股息总额。相应的简单净收益率为

$$R_{s,t} = \mathcal{R}_{s,t} - 1.$$

- 由于价格 $P_t \geq 0$, 总收益率 $\mathcal{R}_{s,t} \geq 0$ 。
- 净收益率 $R_{s,t}$ 可以为负, 但下界为 -1 (即亏损 100%)。
- 可分解为:

$$\text{股息收益} = \frac{D_t}{P_s}, \quad \text{资本增值率} = \frac{P_t - P_s}{P_s},$$

- 二者之和即 $R_{s,t} = \mathcal{R}_{s,t} - 1$ 。

对数收益率（连续复利）

定义（对数收益率）

在 $[t-1, t]$ 上的连续复利收益率为

$$r_t \equiv \log R_t = p_t - p_{t-1},$$

其中 $p_t \equiv \log P_t$ 。多时段的对数收益率为

$$r_{t,t+k} = \log R_{t,t+k} = \log (R_{t-k+1} \cdots R_t) = r_{t-k+1} + \cdots + r_t.$$

- 对数收益率是**连续复利回报**，在理论和实证中广泛使用。
- 多期可加性：长区间对数收益率等于单期对数收益率之和。

为何常用对数收益率？

- **对称性**: 对数收益率对价格上涨/下跌更对称。
- **可加性**: 多期对数收益率可简单相加，便于长期分析：

$$r_{t,t+k} = r_t + r_{t+1} + \cdots + r_{t+k-1}.$$

- **小变动近似**: 当价格变动不大时，

$$r_t \approx R_t,$$

可用对数收益率近似简单收益率。

- **统计性质好**: 对数收益率常比简单收益率更接近正态分布，便于使用经典统计模型。

对数收益率的年化

- 设 r_t 为日度对数收益率，样本平均为 \bar{r}_{day} 。
- 常见做法：假定一年约有 252 个交易日，则

$$\bar{r}_{\text{year}} \approx 252 \times \bar{r}_{\text{day}}.$$

- 采用 252 的理由：

$$252 \approx 365.2425 \times \frac{5}{7} - 10,$$

即从全年天数中扣除周末与部分节假日。

- 实务中也可直接用样本期内实际交易日数进行年化，更为稳妥。

简单收益率 vs 对数收益率：高波动例子

考虑一个投资组合，价值先从 100 元下降到 50 元，再回升到 100 元。

① 简单收益率

- $100 \rightarrow 50: \frac{50-100}{100} = -50\%;$
- $50 \rightarrow 100: \frac{100-50}{50} = 100\%;$
- 算术平均: $\frac{-50\%+100\%}{2} = 25\%;$
- 但两期累计收益 $(1 - 50\%)(1 + 100\%) - 1 = 0$, 实际应为 0。

② 对数收益率

- $100 \rightarrow 50: \log(0.5);$
- $50 \rightarrow 100: \log(2);$
- 和: $\log(0.5) + \log(2) = 0$, 准确反映最终回到起点。
- 在高波动市场中，对数收益率更能准确衡量长期收益表现。

时间可加性与组合可加性

- 对数收益率时间可加：

$$r_{t,t+k} = r_t + r_{t+1} + \cdots + r_{t+k-1}.$$

- 简单收益率不具时间可加性：

$$R_{t,t+k} \neq R_t + R_{t+1} + \cdots + R_{t+k-1}.$$

- 但简单收益率具有投资组合可加性：

$$R_t(\mathbf{w}) = \sum_{i=1}^n w_{it} R_{it},$$

其中 w_{it} 为资产 i 在 t 时刻的组合权重。

- 对数收益率一般不具有组合可加性，仅在变动极小时可近似为线性组合。

简单收益率的组合可加性推导

设 n_i 为股票 $i = 1, 2$ 的持有数量, $V_t = n_1 P_{1t} + n_2 P_{2t}$ 为组合价值,
 $w_{it} = n_i P_{it} / V_t$ 为权重。

则

$$\begin{aligned}V_{t+1} &= n_1 P_{1,t+1} + n_2 P_{2,t+1} \\&= V_t (1 + w_{1t} R_{1,t+1} + w_{2t} R_{2,t+1}),\end{aligned}$$

于是

$$\frac{V_{t+1} - V_t}{V_t} = w_{1t} R_{1,t+1} + w_{2t} R_{2,t+1}.$$

- 说明：组合简单收益率是单个资产简单收益率的加权和。
- 对数收益率一般无法写成这样的线性组合。

累计收益与几何平均收益率

定义 (累计收益与几何平均收益率)

给定从 t 到 $t+k$ 的 k 个单期简单总收益率 R_t, \dots, R_{t+k-1} , 其中

$R_{t+j} \equiv \frac{P_{t+j} + D_{t+j}}{P_{t+j-1}}$, 则

$$\mathcal{R}_{t,t+k} = \prod_{j=0}^{k-1} R_{t+j}, \quad R_{t,t+k} = \mathcal{R}_{t,t+k} - 1.$$

累计收益与几何平均收益率

定义 (累计收益与几何平均收益率)

若 $r_{t+j} = \log R_{t+j}$, 则

$$\log \mathcal{R}_{t,t+k} = \sum_{j=0}^{k-1} r_{t+j} \iff \mathcal{R}_{t,t+k} = \exp\left(\sum_{j=0}^{k-1} r_{t+j}\right).$$

对应的几何平均收益率为

$$\bar{R}_{t,t+k}^{(G)} = (\mathcal{R}_{t,t+k})^{1/k} - 1 = \exp\left(\frac{1}{k} \sum_{j=0}^{k-1} r_{t+j}\right) - 1.$$

- 几何平均与复利机制一致, 更适合长期业绩评价。

股票指数的两大编制思路

- 算术平均法：

- 对一篮子代表性股票价格求简单算术平均。
- 与基期平均价格比较，得到指数。
- 每只股票权重相同，不能反映市值差异。

- 加权平均法：

- 按市值/成交额/流通股本等赋予不同权重。
- 更能反映不同股票对市场的实际影响力。

价格加权指数：道琼斯工业平均指数 (DJIA)

- 由道和琼斯于 1896 年创建，最初包含 12 家公司。
- 目前包含 30 家大型蓝筹股公司。
- **价格加权：**
 - 指数 = 成分股价格之和 / “道琼斯除数”。
 - 除数会因股票分割、派息、成分股更替等调整，以保持指数连续性。
- **特点：**
 - 高价股权重更大。
 - 结构简单，但对价格高而市值不一定大的股票更敏感。

市值加权指数：标准普尔 500 指数 (S&P 500)

- 覆盖美国 500 家大型、具代表性的上市公司。
- 成分股由标准普尔委员会根据市值、流动性、行业等选择。
- **自由流通市值加权：**
 - 权重按可自由流通股本 \times 价格计算。
 - 剔除政府、大股东等不能自由交易的股份。
- 指数公式：
$$\text{指数} = \frac{\text{总自由流通市值}}{\text{除数}}$$
。
- 除数调整保证因分割、派息等技术性变化不影响指数水平。

中国：上证指数与深证指数

● 上证综合指数：

- 覆盖上交所全部 A 股、B 股。
- 采用自由流通市值加权。
- 基期：1990-12-19，基点 100。
- 公式：

$$\text{指数} = \frac{\sum(\text{个股价格} \times \text{自由流通股本})}{\text{基期自由流通总市值}} \times \text{基点.}$$

● 深证成指/深证综指：

- 覆盖深交所 A 股、B 股。
- 同样采用自由流通市值加权，基期 1991-04-03，基点 100。

收益率与对数收益率的关系

- 收益率 R 与对数收益率 $r = \log(1 + R)$ 一般视为随机变量。
- 詹森不等式：

$$\mathbb{E}[r] = \mathbb{E}[\log(1 + R)] \leq \log(1 + \mathbb{E}[R]).$$

- 通常 $\mathbb{E}[R] \in (-1, 1)$, 在 $|\mathbb{E}[R]| \ll 1$ 且方差较小情形下,

$$\mathbb{E}[r] \simeq \mathbb{E}[R] - \frac{1}{2} \text{Var}(R).$$

- 说明：对数收益率的期望通常**小于**简单收益率的期望。

夏普比率与风险溢价

- 用收益率期望衡量“平均收益水平”，方差衡量“风险”。
- 夏普比率 (Sharpe Ratio)：

$$S = \frac{E[R] - R_f}{\sigma(R)} = \frac{E[R] - R_f}{\sqrt{\text{Var}(R)}},$$

其中 R_f 为无风险收益率。

- $E[R] - R_f$: 风险溢价。
- 解释：
 - $S > 0$: 组合平均回报高于无风险；
 - $S < 0$: 表现不如无风险；
 - S 越高，单位风险获得的超额回报越多。

索提诺比率与信息比率

索提诺比率 (Sortino Ratio) :

$$\text{Sortino} = \frac{\mathbb{E}[R] - R_f}{\sigma_{\text{down}}},$$

其中 σ_{down} 为下行标准差 (只考虑负收益)。

信息比率 (Information Ratio) :

$$\text{IR} = \frac{\mathbb{E}[R - R_b]}{\sqrt{\text{Var}(R - R_b)}},$$

其中 R_b 为基准 (如指数) 收益率。

- 索提诺比率：只惩罚下行风险，适合对回撤敏感的策略。
- 信息比率：评估组合相对基准的超额表现与稳定性。

有限责任与对数正态假设

- 多数金融资产具有**有限责任**:

- 价格下界为 0, 简单收益率下界为 $-100\% (R_t \geq -1)$ 。

- 因此简单收益率本身不可能严格服从正态分布。

- 常见建模假设:

$$1 + R \sim \text{Lognormal}(\mu, \sigma^2) \iff r = \ln(1 + R) \sim \mathcal{N}(\mu, \sigma^2).$$

- 即: **对数收益率近似正态**, 简单收益率近似对数正态。

有限责任的例外与“原油宝”事件

有限责任的一些例外：

- 一般合伙企业：合伙人可能对企业债务承担无限责任；
- 保证金账户交易：价格剧烈波动时可能产生追加保证金通知，损失超过初始投入；
- 裸期权等衍生品：卖出未对冲的期权时，理论上损失可非常巨大。

原油宝事件（2020 年 4 月）：

- “原油宝”是挂钩 WTI 等原油期货的银行理财产品。
- 2020-04-20，WTI 期货价格历史性跌至负值（每桶 -37.63 美元）。
- 由于设计与风控问题，部分投资者出现大额亏损甚至“爆仓”与负余额。

偏度与峰度

- 偏度与峰度刻画分布形态：

$$\kappa_3 \equiv E\left[\frac{(r - \mu)^3}{\sigma^3}\right], \quad \kappa_4 \equiv E\left[\frac{(r - \mu)^4}{\sigma^4}\right].$$

- 对正态分布： $\kappa_3 = 0$ 、 $\kappa_4 = 3$ 。
- 偏度：
 - $\kappa_3 > 0$: 右尾更长，右偏；
 - $\kappa_3 < 0$: 左尾更长，左偏。
- 峰度：
 - $\kappa_4 > 3$: 正峰度 (尖峰厚尾, Leptokurtic);
 - $\kappa_4 < 3$: 负峰度 (平峰薄尾, Platykurtic)。

样本偏度与峰度及 R 示例

- 样本均值与方差：

$$\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t, \quad s^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2.$$

- 样本偏度与峰度：

$$\hat{\kappa}_3 = \frac{\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^3}{\left(\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2 \right)^{3/2}}, \quad \hat{\kappa}_4 = \frac{\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^4}{\left(\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2 \right)^2}.$$

目录

1 金融市场

- 货币市场
- 债券市场
- 股票市场

2 市场类型以及交易方式

- 主要交易所
- 电子交易与订单类型
- 监管框架与全球化

3 收益率

- 收益率的构建
- 股票指数的编制方法
- 收益率的统计性质

4 金融经济学基础

- 效用函数以及风险厌恶
- 均值-方差投资组合模型
- CAPM、APT 与 C-CAPM

5 章节总结

本节目标

- 金融计量经济学中的模型（如 CAPM、C-CAPM）**基于经济理论**：
 - 效用最大化；
 - 风险厌恶；
 - 均衡与套利。
- **本节核心：**
 - 效用函数与风险厌恶；
 - 期望效用理论及其悖论；
 - 风险厌恶的度量（ARA, RRA）与常用效用函数；
 - 均值-方差投资组合理论；
 - CAPM、APT 与 C-CAPM 的理论基础；
 - 均衡定价 vs 套利定价。

圣彼得堡悖论与效用

- 圣彼得堡赌局：掷硬币直到出现正面：

- 第一次正面：赢 1 元；
- 第二次才正面：赢 2 元；
- 第三次才正面：赢 4 元；依此类推。

- 数学期望：

$$E = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot 2^{k-1} = \sum_{k=1}^{\infty} \frac{1}{2} = \infty.$$

- 理论上应愿意支付“无限多”入场费，但现实中没人会这么做。
- 伯努利提出：决策基于效用而非金额本身，且效用具有边际递减。

伯努利的对数效用与期望效用

- 伯努利使用对数效用：

$$u(x) = \ln x.$$

- 对圣彼得堡赌局，期望效用有限：

$$\sum_{t=1}^{\infty} \left(\frac{1}{2}\right)^t \ln(2^{t-1}) = \ln 2 < \infty.$$

- 如果初始财富为 1，等价确定支付约为 2 元。
- 说明：人们在决策中追求的是**期望效用最大化**而非**期望货币额最大化**。

理性偏好与效用表示

定义 (理性偏好)

偏好关系 \succeq 理性, 当且仅当:

- 完备性: 任意 $x, y \in X$, $x \succeq y$ 或 $y \succeq x$;
- 传递性: 若 $x \succeq y$ 且 $y \succeq z$, 则 $x \succeq z$.

定义 (连续性)

若一列偏好关系在极限下仍保持, 则称 \succeq 连续: 若 $x^n \succeq y^n$ 对所有 n 成立, 且 $x = \lim x^n$ 、 $y = \lim y^n$, 则 $x \succeq y$ 。

效用表示定理

Proposition

若偏好关系 \succeq 理性且连续，则存在连续效用函数 $u: X \rightarrow \mathbb{R}$ 表示该偏好：

$$x \succeq y \iff u(x) \geq u(y).$$

- 含义：可以用一个实值函数来“编码”偏好。
- 这使得可以用微积分等工具处理最优决策问题。

简单彩票与复合彩票

定义 (简单彩票)

简单彩票 $L = (p_1, \dots, p_n)$, $p_i \geq 0$ 、 $\sum p_i = 1$, 表示各结果的发生概率。

定义 (复合彩票)

给定若干简单彩票 $L_k = (p_1^k, \dots, p_n^k)$ 与选择概率 α_k ($\sum \alpha_k = 1$), 先以 α_k 选中某张彩票, 再按该彩票决定结果, 称为复合彩票。

- 复合彩票可理解为“彩票的的彩票”, 常用于刻画多阶段风险。

独立性公理与期望效用

Proposition (独立性公理)

对任意彩票 A, B, C 与 $\alpha \in [0, 1]$:

$$A \succeq B \iff \alpha A + (1 - \alpha)C \succeq \alpha B + (1 - \alpha)C.$$

- 含义：与第三个彩票 C 的“混合”不应改变原先对 A, B 的偏好。
- 这是期望效用理论的关键假设之一。

期望效用定理

Proposition (期望效用)

若偏好关系 \succeq 在彩票集合上满足：理性、连续性与独立性，则存在效用函数 $u: X \rightarrow \mathbb{R}$ 使对任何彩票 $L = (p_1, \dots, p_n)$ 与 $L' = (p'_1, \dots, p'_n)$ ：

$$L \succeq L' \iff \sum_{i=1}^n p_i u(x_i) \geq \sum_{i=1}^n p'_i u(x_i).$$

- 决策可以写成“最大化期望效用”问题。
- 风险态度（厌恶/中性/偏好）由效用函数形状决定。

阿莱悖论：对独立性公理的挑战

例 (阿莱悖论)

两组选择：

第一组：

- 1A: 100% 得 100 万;
- 1B: 89% 得 100 万, 10% 得 500 万, 1% 得 0。

第二组：

- 2A: 11% 得 100 万, 89% 得 0;
- 2B: 10% 得 500 万, 90% 得 0。

- 实验结果：多数人选 1A 和 2B。
- 这组偏好与期望效用的**独立性公理**不相容。
- 说明现实决策往往偏好“确定性”，不完全符合期望效用假设。

埃尔斯伯格悖论：模糊厌恶

例 (埃尔斯伯格悖论)

两个盒子：

- 盒 1：50 红 + 50 黑，已知比例；
- 盒 2：共 100 球，红黑比例未知。

抽到红球获奖。

多数人偏好从盒 1 抽，无论中奖颜色设定为红或黑。

- 表明人们倾向回避未知概率（模糊厌恶）。
- 超出了传统期望效用理论的描述范围。

风险厌恶与确定性等值

- 风险厌恶：偏好“确定但略低收益”而非“高期望但有风险”的项目。
- 确定性等值 (Certainty Equivalent, CE)：

$$u(CE) = E[u(X)].$$

- 风险态度：
 - 风险中性： $CE = E[X]$ ，效用线性；
 - 风险厌恶： $CE < E[X]$ ，效用凹；
 - 风险偏好： $CE > E[X]$ ，效用凸。

Arrow-Pratt 绝对风险厌恶系数

定义 (绝对风险厌恶系数 ARA)

$$R_A(y) = -\frac{u''(y)}{u'(y)}.$$

- $R_A(y)$ 描述在财富水平 y 时，对**小额绝对赌注**的局部风险厌恶强度。
- 导出方式：令投资者在赌注

$$\begin{cases} +\delta & \text{概率 } p, \\ -\delta & \text{概率 } 1-p \end{cases}$$

与不参与之间无差异，可求得临界概率 p^* ，

$$p^* = \frac{1}{2} + \frac{\delta}{4} R_A(y).$$

- $R_A(y)$ 越大，需要越高的赢面才愿意参与。

Arrow-Pratt 相对风险厌恶系数

定义 (相对风险厌恶系数 RRA)

$$R_R(y) = -\frac{y u''(y)}{u'(y)}.$$

- 关注“按财富比例”的小赌注：赢/输 cy 。
- 类似推导，可得临界概率：

$$p^* = \frac{1}{2} + \frac{c}{4} R_R(y).$$

- $R_R(y)$ 在理论与实践中都很关键：
 - 用于分析“大额投资”或“占财富较大比例”的决策。

风险厌恶 vs 损失厌恶与前景理论

前景理论 (Prospect Theory):

- 由 Kahneman 和 Tversky 提出。
- 决策基于相对于某个参照点的“收益/损失”。
- 价值函数呈“S”形：
 - 收益区：凹形（边际效用递减）；
 - 损失区：凸形，且**损失厌恶**——对损失比等额收益更敏感。
- 解释许多期望效用理论无法解释的行为（如阿莱悖论、埃尔斯伯格悖论）。

常用效用函数：CARA、CRRA、HARA、线性

CARA (常绝对风险厌恶): $u(y) = -e^{-\alpha y}$, $R_A(y) = \alpha$ 为常数。

CRRA (常相对风险厌恶): $u(y) = \frac{y^{1-\gamma}-1}{1-\gamma}$, $R_R(y) = \gamma$ 。当 $\gamma = 1$ 时,
 $u(y) = \ln y$ 。

HARA 与线性效用

HARA (双曲绝对风险厌恶): 若 $R_A(y) = \frac{1}{ay+b}$, 可得 $u(y) = \frac{(y-c_s)^{1-\gamma}}{1-\gamma}$,
其中 $\gamma = 1/a$, $c_s = -b/a$ 。CARA/CRRA 是其特例。

线性效用 (风险中性): $u(y) = \alpha y$, 绝对与相对风险厌恶系数均为 0。

相关系数与两资产组合风险

- 相关系数：

$$\rho_{A,B} = \frac{\text{Cov}(A, B)}{\sigma_A \sigma_B}.$$

- 两资产组合标准差：

$$\sigma_p = \sqrt{\sigma_A^2 w_A^2 + \sigma_B^2 w_B^2 + 2w_A w_B \rho_{A,B} \sigma_A \sigma_B}.$$

- 若 $\rho_{A,B} < 1$, 组合风险低于简单权重加权平均。
- $\rho_{A,B} = -1$ 且权重合适时, 组合可实现“零风险”。

多资产组合与全局最小方差组合

- $R = (R_1, \dots, R_n)'$, 期望为 μ , 协方差矩阵为 Σ 。
- 权重向量 w , 满足 $w'1 = 1$ 。
- 组合期望:

$$\mu_w = w'\mu.$$

- 组合方差:

$$\sigma_w^2 = w'\Sigma w.$$

- 全局最小方差组合:

$$w_{GMV} = \frac{\Sigma^{-1}1}{1'\Sigma^{-1}1}.$$

有效前沿与两基金分离定理（要点）

- 在给定目标收益 m 下最小化方差，可得到一族最优组合 $w_{\text{opt}}(m)$ 。
- 其 $(m, \sigma_{\text{opt}}(m))$ 在均值-标准差平面上形成一条双曲线，上半部分为**有效前沿**。

定理（两基金分离定理）

任何位于均值-方差有效集上的投资组合，都可以写成两个（任意两个）有效组合的线性组合。

CAPM 的基本思想

- 马科维茨组合理论 + 均衡假设 \Rightarrow 资本资产定价模型 (CAPM)。
- 基本观点：
 - 投资者只为系统性风险得到补偿；
 - 可分散风险不获得额外风险溢价。
- 关键变量： β ——资产相对市场组合的风险度量。

夏普-林特纳 CAPM

- 存在无风险资产时：

$$E[R_i] = R_f + \beta_{im} (E[R_m] - R_f),$$

$$E[Z_i] = \beta_{im} E[Z_m].$$

- 解释：

- 预期超额收益与 β 呈线性关系；
- 只要知道资产的 β 就能在 CAPM 框架下给出“合理报酬”。

布莱克 CAPM（零贝塔 CAPM）

定理

对任意资产或投资组合 i , 以及任一有效前沿上的组合 p , 有

$$E(R_i) = E(R_{Z_p}) + \beta_{ip}(E(R_p) - E(R_{Z_p})), \quad \beta_{ip} = \frac{\text{Cov}(R_i, R_p)}{\text{Var}(R_p)},$$

其中 R_{Z_p} 为相对于 p 的零协方差组合收益率。

- 当不存在真正无风险资产时, 用零贝塔组合替代。

套利定价理论 (APT)

- 由 Ross (1976) 提出。
- 观点：
 - 资产预期收益由多个因子（宏观变量等）线性驱动；
 - 若价格偏离因子结构给出的值，则存在套利机会，价格会被套利推回。
- 抽象形式：

$$\mu \approx \iota \lambda_0 + \mathbf{B} \lambda_K,$$

其中 λ_0 近似无风险或零贝塔收益， λ_K 为因子风险溢价。

- APT 比 CAPM 更一般，但没有指定因子具体是什么，需要实证选择。

C-CAPM 与随机贴现因子

- 基于消费的 CAPM：从代表性消费者的跨期效用最大化出发：

$$\max E_t \left[\sum_{j=0}^{\infty} \delta^j U(C_{t+j}) \right].$$

- 一阶条件得到欧拉方程：

$$U'(C_t) = \delta E_t [(1 + R_{i,t+1}) U'(C_{t+1})].$$

- 定义随机贴现因子（定价核）：

$$M_{t+1} = \delta \frac{U'(C_{t+1})}{U'(C_t)}, \quad 1 = E_t[(1 + R_{i,t+1}) M_{t+1}].$$

- 超额收益与 M_{t+1} 的协方差负相关：协方差越小，要求的风险溢价越高。

均衡定价 vs 套利定价

均衡定价: 基于供需平衡与效用最大化, 典型模型: CAPM、C-CAPM。

套利定价: 基于无套利原则, 价格必须使得不存在零成本无风险套利,

典型模型: APT、多因子模型。

- 实务与研究中, 两类思路往往相辅相成:

- 均衡模型提供“应当如此”的基准;
- 套利模型刻画“若偏离则如何被纠正”。

目录

1 金融市场

- 货币市场
- 债券市场
- 股票市场

2 市场类型以及交易方式

- 主要交易所
- 电子交易与订单类型
- 监管框架与全球化

3 收益率

- 收益率的构建
- 股票指数的编制方法
- 收益率的统计性质

4 金融经济学基础

- 效用函数以及风险厌恶
- 均值-方差投资组合模型
- CAPM、APT 与 C-CAPM

5 章节总结

本章小结

- 系统回顾了：
 - 金融市场结构与交易方式（前面章节）；
 - 收益率的定义、构建与统计性质；
 - 股票指数的编制方法与实际案例；
 - 效用、风险厌恶与期望效用理论；
 - 均值-方差组合理论及 CAPM、APT、C-CAPM。
- 为后续章节的计量建模与检验打下金融经济学基础：
 - 如何从理论出发构建可估计的计量模型；
 - 如何使用收益率数据检验资产定价理论。

第 3 章线性回归模型

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

本章主线

- 线性回归模型是金融计量中最常用的基础工具之一：
 - 资产定价：CAPM、APT 中估计预期收益与风险
 - 风险管理：波动率、相关性估计与 VaR / ES
 - 市场有效性检验：信息如何反映在价格中
 - 宏观 \leftrightarrow 金融：利率、通胀、GDP 与资产价格
- 模型构建的一般流程：
 - ① 明确问题与理论背景
 - ② 选模型类型（线性/非线性，静态/动态）
 - ③ 数据获取、清洗与变换
 - ④ 模型设定：因变量、自变量及预期符号
 - ⑤ 参数估计（OLS 等）
 - ⑥ 模型检验与稳健性分析
 - ⑦ 解释结果，用于预测/政策分析
- 本章：一元 \rightarrow 多元回归 \rightarrow 模型诊断 \rightarrow 金融应用

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

一元线性回归模型：基本形式

模型

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$$

- y : 被解释变量 (因变量)
- x : 解释变量 (自变量、预测变量)
- β_0 : 截距, 当 $x = 0$ 时 y 的期望值
- β_1 : 斜率, x 每变动 1 单位, y 的条件期望变化
- ε : 误差项, 代表除 x 之外的其他影响因素
- 目标: 找出“最佳直线”, 让预测值 \hat{y}_i 尽量接近 y_i

术语: x 与 y 的不同称呼

表: 回归中 x 和 y 的术语

y	x
因变量 (Dependent variable)	自变量 (Independent variable)
被解释变量 (Explained variable)	解释变量 (Explanatory variable)
被预测/响应变量 (Predicted / Response)	预测变量 (Predictor)
被回归变量 (Regressand)	回归变量 (Regressor)
—	控制变量 (Control variable)

- 多元回归中, 这些术语同样适用

一元回归的金融应用举例

- 利率与股市：

- x : 基准利率, y : 股指收益 / 水平
- 分析利率变动对股市整体表现的影响

- 房价与宏观：

- x : GDP 增长率, y : 房地产价格指数

- 汇率与进出口：

- x : 美元兑人民币汇率, y : 美国对中国出口额

- 财务指标与股票收益：

- x : 市盈率、股息率等单一指标
- y : 个股收益率

- 通过斜率 β_1 : 看方向与强度

“回归”与“中庸之道”

[“回归现象”与文化类比]

- 高尔顿 (Galton) 在 19 世纪研究身高遗传:
 - 高个父母 \Rightarrow 子女偏高但不如父母高
 - 矮个父母 \Rightarrow 子女偏矮但高于父母
- 子代身高向总体平均“回归”: **回归现象**
- 与“中庸之道”类似: 避免极端, 向中间靠拢、追求平衡
- 在建模中也要避免极端:
 - 不要过度拟合, 也不要欠拟合
 - 保持模型复杂度与解释力之间的平衡

一元线性回归的基本假设

令 $x = \{x_i\}$, $y = \{y_i\}$, $i = 1, \dots, n$:

- ① **线性**: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
- ② **零均值**: $E(\varepsilon_i) = 0$
等价于 $E(y_i | x_i) = \beta_0 + \beta_1 x_i$

- ③ **同方差性**:

$$\text{Var}(\varepsilon_i) = \sigma^2 \quad \text{或} \quad \text{Var}(\varepsilon_i | x_i) = \sigma^2$$

- ④ **无相关性**:

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$$

- ⑤ **正态性 (可选)**:

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \Leftrightarrow \varepsilon_i \sim N(0, \sigma^2)$$

“零协方差不代表独立性”的例子

例

令随机变量 x 满足 $E(x) = 0$ 且 $E(x^3) = 0$ (如 $x \sim N(0, \sigma^2)$)。令 $y = x^2$ 。显然 x 与 y 不独立, 但

$$\text{Cov}(x, y) = E(xy) - E(x)E(y) = E(x^3) - 0 = 0.$$

\Rightarrow 协方差为 0 $\not\Rightarrow$ 独立。

OLS: 目标函数

最小化残差平方和

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

- 对 β_0, β_1 求偏导并令其为 0:

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0,$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0.$$

- 解出 $\hat{\beta}_0, \hat{\beta}_1$: 正规方程

OLS 显式解与协方差形式

定义样本均值：

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

由一阶条件可解得：

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}},$$

其中

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2.$$

回归直线

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \bar{y} + \frac{S_{xy}}{S_{xx}}(x - \bar{x}).$$

残差与 OLS 的几何性质

残差定义

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i.$$

OLS 拟合线的若干重要性质：

① 拟合线必过样本均值点 (\bar{x}, \bar{y})

② 残差和为零： $\sum_{i=1}^n \hat{\varepsilon}_i = 0$

③ $\sum_{i=1}^n x_i \hat{\varepsilon}_i = 0$, 即 $\text{Cov}(x_i, \hat{\varepsilon}_i) = 0$

④ $\sum_{i=1}^n \hat{y}_i \hat{\varepsilon}_i = 0$, 即 $\text{Cov}(\hat{y}_i, \hat{\varepsilon}_i) = 0$

⇒ 残差在解释变量与拟合值方向上“正负抵消”

误差 vs 残差

- 误差 (Error) :

$$\varepsilon_i = y_i - E(y_i \mid x_i)$$

理论上定义，涉及真实但未知的参数， ε_i 不可观测。

- 残差 (Residual) :

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

是用估计值代替真值后得到的可观测量。

- 模型诊断：用残差检验假设（同方差性、独立性、正态性、异常值等）。

回归系数的 t 检验

在线性回归中，常检验

$$H_0 : \beta_i = 0 \quad \text{vs} \quad H_1 : \beta_i \neq 0.$$

t 统计量

$$t_i = \frac{\hat{\beta}_i - 0}{SE(\hat{\beta}_i)}, \quad i = 0, 1.$$

- 在误差正态且模型假设成立时：

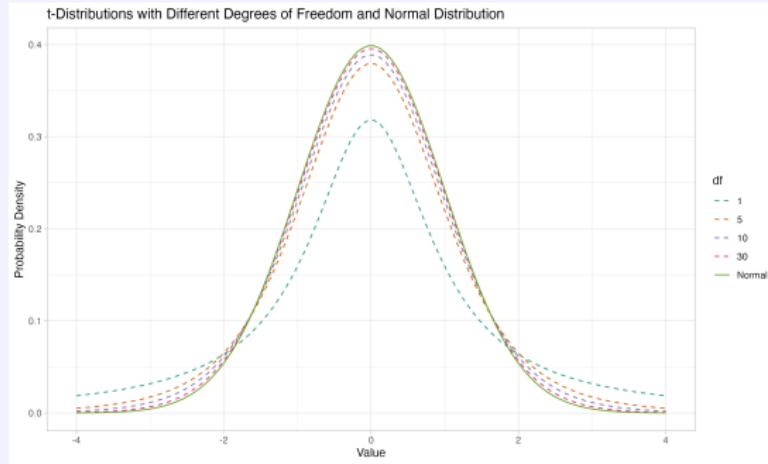
$$t_i \sim t(n - k - 1),$$

其中 k 为自变量个数

- 通过临界值/ p 值判断 β_i 是否显著

t 分布与正态分布

- t 分布自由度 df 小时：尾部更“厚”，更保守
- 随 df 增大（样本量变大），t 分布逐渐逼近标准正态
- 自由度常为：样本量减去估计参数个数



图：不同自由度 t 分布与标准正态分布比较

平方和分解与决定系数 R^2

定义：

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{总平方和}$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{回归平方和}$$

$$SSE = \sum_{i=1}^n \hat{\varepsilon}_i^2 \quad \text{残差平方和}$$

$$SST = SSR + SSE.$$

决定系数

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

- $R^2 \in [0, 1]$: 越大说明样本中 y 的变异越多被回归解释
- 若所有点共线，则 $R^2 = 1$

R^2 与相关系数

在简单线性回归中，

$$R^2 = (\hat{\rho}_{y,\hat{y}})^2,$$

即 样本相关系数的平方。

- 这也是“ R 平方”名称的来源
- 直观理解：回归线拟合的好坏 \Leftrightarrow 拟合值与真实值的相关程度

调整后的 R^2

- 问题：在多元回归中，增加一个解释变量， R^2 几乎不会下降
- 即使新变量没什么用，也会“虚假提高” R^2

调整后的决定系数

$$\bar{R}^2 = 1 - \frac{\sum \hat{\varepsilon}_i^2 / (n - k - 1)}{\sum (y_i - \bar{y})^2 / (n - 1)}.$$

- 对自变量个数 k 做了惩罚
- 一般有 $\bar{R}^2 \leq R^2$
- 做变量选择、模型比较时更有参考价值

残差标准差 (RSE)

定义

$$\text{RSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{或除以自由度}) .$$

- 衡量“平均预测误差”的绝对水平
- 单位与 y 相同
- 通常与 y 的均值相比，判断误差是否“大”

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

多元线性回归：标量形式

模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon.$$

- y : 被解释变量
- x_1, \dots, x_k : 多个解释变量
- β_j : 在其他变量控制不变时, x_j 对 y 的边际影响
- ε : 其他未入模因素

$$\min_{\beta_0, \dots, \beta_k} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2.$$

多元线性回归：矩阵形式

矩阵表示

$$Y = \mathbf{X}\beta + \mathbf{u},$$

- Y : $n \times 1$ 向量, 所有 y_i
- \mathbf{X} : $n \times (k+1)$ 设计矩阵
- β : $(k+1) \times 1$ 参数向量
- \mathbf{u} : 误差向量

OLS 闭式解

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y.$$

- 要求: \mathbf{X} 满秩 (无完全多重共线性)

Gauss-Markov 定理 (直观版)

在线性模型

$$Y = \mathbf{X}\beta + \varepsilon$$

下, 如果

- 线性且外生: $E(\varepsilon | \mathbf{X}) = \mathbf{0}$
- 同方差: $\text{Var}(\varepsilon | \mathbf{X}) = \sigma^2 \mathbf{I}$
- 无自相关: $\text{Cov}(\varepsilon_i, \varepsilon_j | \mathbf{X}) = 0$
- \mathbf{X} 满秩

则 OLS 估计量是最佳线性无偏估计量 (BLUE):

- 线性: $\hat{\beta}$ 是 Y 的线性函数
- 无偏: $E(\hat{\beta}) = \beta$
- 最小方差: 在所有线性无偏估计量中方差最小

高斯-马尔可夫：矩阵形式的假设

以 \mathbf{X} 为条件，常写为：

- ① $Y = \mathbf{X}\beta + \varepsilon$
- ② \mathbf{X} 为 $n \times (k+1)$ 且秩为 $k+1$
- ③ $E(\varepsilon | \mathbf{X}) = \mathbf{0}$
- ④ $E(\varepsilon\varepsilon' | \mathbf{X}) = \sigma^2 \mathbf{I}$
- ⑤ 若进一步假设 $\varepsilon | \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ ，则 OLS 也是 MVUE，可做 t/F 检验

F 统计量：整体检验

多元线性回归：

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon.$$

检验所有斜率是否同时为零

$$H_0 : \beta_1 = \cdots = \beta_k = 0 \quad \text{vs} \quad H_1 : \text{至少一个 } \beta_j \neq 0.$$

F 统计量与 R^2 的关系

$$F = \frac{(n - k - 1)}{k} \cdot \frac{R^2}{1 - R^2} = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}.$$

- 分子自由度 k , 分母自由度 $n - k - 1$
- 用来检验“模型整体是否有解释力”

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

异方差性：概念与后果

- 理想假设: $\text{Var}(\varepsilon_i | x_i) = \sigma^2$, 对所有 i 相同
- 异方差性: $\text{Var}(\varepsilon_i | x_i) = \sigma_i^2$ 随 x_i 变化
- 后果:
 - OLS 仍然无偏, 但**不再有效** (不是 BLUE)
 - 标准误估计有偏, t/F 检验不可靠
- 可能来源:
 - 截面数据规模差异大 (国家、公司大小)
 - 重要变量遗漏
 - 模型设定错误 (本质非线性)
 - 结构性变化、制度变迁

异方差性的识别与处理

- 识别：

- 图形：残差平方 vs 解释变量
- White 检验
- Breusch-Pagan (BP) 检验

- 处理：

- 加权最小二乘 (WLS)
- 使用稳健标准误 (robust / HAC)
- 重新设定模型 (变换变量等)

多重共线性：含义与影响

- 多重共线性：一些解释变量之间高度相关
- 后果：
 - OLS 估计仍无偏，但方差放大
 - t 统计量变小，部分系数“不显著”
 - 样本稍变，系数估计大幅波动，甚至符号反常

“和而不同”：变量之间既要“和”（协同解释 y ），又要“不同”（提供独立信息）；过度相似 \Rightarrow 信息冗余，模型不稳。

诊断工具：VIF 与条件数

定义 (方差膨胀因子 (VIF))

对某解释变量 x_i ,

$$\text{VIF}_i = \frac{1}{1 - R_i^2},$$

其中 R_i^2 是用其他解释变量回归 x_i 得到的决定系数。

- $\text{VIF}_i = 1$: 与其他变量不相关
- $\text{VIF}_i > 5$ 或 > 10 : 共线性问题严重

条件数: 对相关矩阵 $\mathbf{R} = \mathbf{X}'\mathbf{X}$ 的特征值 $\lambda_{\min}, \lambda_{\max}$,

$$\kappa = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}.$$

一般 $\kappa > 10$ 或更高表示显著共线性。

缓解多重共线性的策略

- 变量中心化：尤其在多项式回归中减少相关性
- 删除高度相关的变量：
 - 使用逐步回归、最优子集回归等方法
- 使用正则化：
 - 岭回归、Lasso、主成分回归
 - 以引入偏差换取方差大幅降低
- 参考文献：
 - Greene 《Econometric Analysis》 (Greene 2000)
 - James 等 《Introduction to Statistical Learning》 (James et al. 2013)

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

案例：mtcars 中重量与 MPG

- 研究问题：车重 wt 是否降低燃油效率 mpg ?
- 数据：R 内置 mtcars 数据集

```
# 加载 mtcars 数据集并回归  
data(mtcars)
```

```
model <- lm(mpg ~ wt, data = mtcars)  
summary(model)
```

```
# 可视化回归线  
library(ggplot2)  
ggplot(mtcars, aes(x = wt, y = mpg)) +  
  geom_point() +  
  geom_smooth(method = "lm", se = FALSE)
```

mtcars 回归结果解读

Call:

```
lm(formula = mpg ~ wt, data = mtcars)
```

Coefficients:

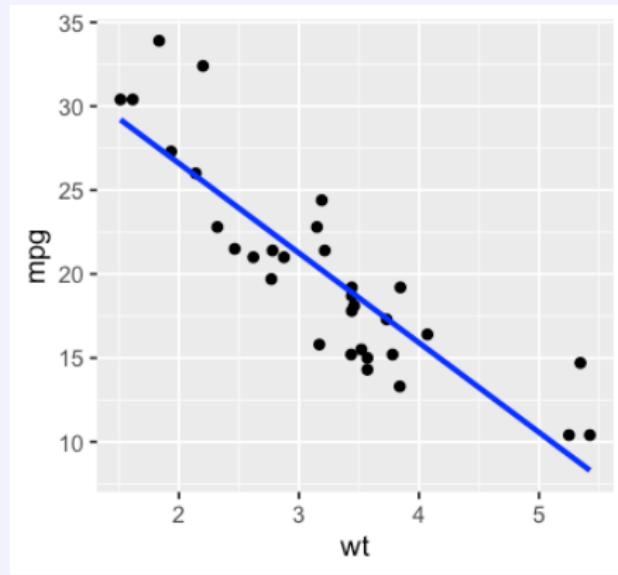
(Intercept)	37.2851	(t = 19.858, p < 2e-16)
wt	-5.3445	(t = -9.559, p = 1.29e-10)

Residual standard error: 3.046 on 30 df

Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446
 F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10

- 斜率为 -5.34: 车重每增加 1 (1000 磅), mpg 平均下降约 5.3
- $R^2 \approx 0.75$: 模型解释了约 75% 的变异
- F 统计量显著: wt 对 mpg 有显著整体影响

散点图与回归线



汽车重量与 MPG 的线性关系

- 图形有助于：
 - 直观检视线性关系、异常值
 - 检查残差方差是否随 x 变化

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

CAPM：理论形式

CAPM 模型

$$\text{E}(R_i) = R_f + \beta_{im} (\text{E}(R_m) - R_f).$$

- R_i : 资产 i 的收益率
- R_f : 无风险利率
- R_m : 市场组合收益率
- β_{im} : 相对市场的系统性风险

超额收益形式（用于实证）

$$R_{it} - R_{ft} = \alpha_i + \beta_{im} (R_{mt} - R_{ft}) + \varepsilon_t.$$

- α_i : Jensen's alpha, 捕捉“异常收益”

比亚迪 (002594) 超额收益回归：R 代码

```
library(ggplot2)
library(zoo)

BYD <- read.csv("BYD.csv")
BYD$Date <- as.Date(BYD$Date)

# 缺失值插值
BYD$BYD <- na.approx(BYD$BYD)
BYD$SZ <- na.approx(BYD$SZ)
BYD$Rf <- na.approx(BYD$Rf)
```

比亚迪 (002594) 超额收益回归：R 代码

```
# 对数收益率  
N <- nrow(BYD)  
BYD$BYD_RR <- c(NA, diff(log(BYD$BYD)))  
BYD$SZ_RR <- c(NA, diff(log(BYD$SZ)))  
  
# 超额收益  
BYD$ExcessReturn_SZ <- BYD$SZ_RR - BYD$Rf/365  
BYD <- BYD[-1,]  
  
CAPM_model <- lm(BYD_RR ~ ExcessReturn_SZ, data = BYD)  
summary(CAPM_model)
```

比亚迪 CAPM 回归结果概览

Coefficients:

(Intercept)	0.008144	(t = 7.846, p < 1e-13)
ExcessReturn_SZ	1.261996	(t = 17.293, p < 2e-16)

Residual standard error: 0.0177

Multiple R-squared: 0.4476, Adjusted R-squared: 0.4461
 F-statistic: 299 on 1 and 369 DF, p-value: < 2.2e-16

- $\hat{\beta} \approx 1.26 > 1$: 比亚迪系统性风险高于市场
- 截距 $\hat{\alpha} \approx 0.8\%$ 日超额收益（样本期内）
- $R^2 \approx 0.45$: 市场因子解释了约一半波动
- 实务中需进一步检验残差（异方差、自相关等）

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

Fama–French 多因子模型

四因子模型（含动量）

$$R_{i,t} - R_{f,t} = \alpha + \beta_m(Mkt - RF)_t + \beta_sSMB_t + \beta_hHML_t + \beta_uUMD_t + \varepsilon_t.$$

- $Mkt - RF$: 市场超额收益
- SMB: Small minus Big (规模因子)
- HML: High minus Low (价值因子)
- UMD: Up minus Down (动量因子)

五因子 + 动量（六因子）

进一步加入 RMW (盈利)、CMA (投资) 两个因子

AAPL 月度超额收益：多因子回归结论小结

- 市场因子： $\hat{\beta}_m \approx 1.2\text{--}1.3$ ，显著大于 1：
 - Apple 对市场超额收益高度敏感
- 风格暴露：
 - SMB：不显著（大盘股特征）
 - HML：显著为负 \Rightarrow 成长股特征
 - RMW：显著为正 \Rightarrow 盈利能力因子正向暴露
 - CMA、UMD：不显著
- α ：在 CAPM 与 FF(4/6) 下均为显著正，但随着因子增多略有下降
 - 表明部分异常收益被风格因子“吸收”
 - 剩余 α 可能与遗漏因子或结构变化相关

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

章节总结

- 从一元到多元线性回归：
 - 模型设定、OLS 估计、假设与性质
 - t 检验、F 检验、 R^2 、 \bar{R}^2 、RSE
- 模型诊断：
 - 异方差性：识别与稳健处理
 - 多重共线性：VIF、条件数与应对策略
- 金融应用案例：
 - mtcars：车重与燃油效率
 - BYD CAPM： β 、 α 的含义
 - AAPL 多因子模型：风格因子与剩余超额收益
- 连接后续章节：
 - 时间序列回归、协整、长期方差（LRV）与 HAC

目录

1 一元线性回归模型

- 一元回归的基本假设
- 普通最小二乘法 (OLS)
- t 检验与决定系数 R^2
- 决定系数 R^2 以及调整后的 R^2

2 多元线性回归模型

- Gauss–Markov 定理
- F 统计量与整体显著性

3 异方差性与多重共线性

4 案例：汽车重量与燃油效率

5 案例：CAPM 下的比亚迪股票

6 案例：Fama–French 因子模型（简要）

7 章节总结

8 参考文献

参考文献 |

-  Greene, William H (2000). *Econometric Analysis (4th Edition)*. New Jersey: Prentice Hall.
-  James, Gareth et al. (2013). *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.

第 4 章自回归移动平均模型

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

- 时间序列：按时间顺序、等间隔观测的一列随机变量

$$\{Y_t\}_{t \in \mathbb{Z}}, \quad t = \dots, -1, 0, 1, \dots$$

- 一维：单变量；多维：向量时间序列（如多资产收益、宏观变量向量）
- 典型应用
 - 宏观：GDP、通胀、利率、汇率
 - 金融：收益率、波动率、成交量
 - 其他：气象、医学信号、高频工业数据

- **ARMA 模型**: 一元时间序列的均值动态

AR 部分 + MA 部分

- **VAR 模型**: 多变量时序，所有变量对所有变量的滞后回归
- **SVAR 模型**: 在 VAR 基础上加入结构性约束，识别“结构冲击”
- **应用**:
 - 货币政策、财政政策冲击分析
 - 金融市场对宏观冲击的动态响应

波动率模型一瞥

- 一维: ARCH、GARCH
- 多维: BEKK, DCC-GARCH 等
- 主要刻画: 条件方差/协方差的动态
- 本章重点: 均值方程 ARMA; 波动率模型见后续章节

严格平稳与弱平稳

严格平稳 (strong stationarity)

若对任意 k 和任意 t ,

(Y_t, \dots, Y_{t-k+1}) 的联合分布不依赖于 t ,

则 $\{Y_t\}$ 为严格平稳。

弱平稳 (covariance stationarity)

若对所有 t :

$$E(Y_t) = \mu, \quad \text{Var}(Y_t) = \gamma_0 < \infty,$$

$$\text{Cov}(Y_t, Y_{t-k}) = \gamma_k \text{ 仅依赖于 } k,$$

则称为 (协方差) 弱平稳。

严格平稳 vs 弱平稳：反例

- 柯西分布序列：每期 Y_t 都服从 Cauchy
- 性质：
 - 分布不随时间变 严格平稳
 - 但均值和方差不存在 不满足弱平稳
- 提示：强平稳并不必然蕴含弱平稳
- 类似地，一些 GARCH 过程可以是强平稳但非弱平稳

遍历性：为什么光“平稳”还不够？

- 例： $Y_t = Z$, 所有 t 共用同一个 $N(0, 1)$ 随机变量 Z
- 性质：
 - 严格平稳（每期分布相同）
 - 样本均值 $\bar{Y}_T = Z$, 不会收敛到 0
- 结论：
 - 平稳 \neq “样本平均一定是好估计”
 - 需要更强条件：**均值遍历性**

均值遍历性与绝对可和条件

均值遍历性

若

$$\bar{Y}_T = \frac{1}{T} \sum_{t=1}^T Y_t \xrightarrow{P} E(Y_t) = \mu,$$

则称 $\{Y_t\}$ 在均值上遍历。

充分条件（绝对可和）

若

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty \quad \text{或} \quad \sum_{k=0}^{\infty} |\rho_k| < \infty,$$

则均值遍历性成立，且 $\rho_k \rightarrow 0$ 。

自相关系数

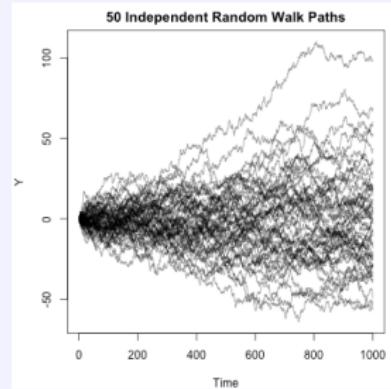
$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad \gamma_k = \text{Cov}(Y_t, Y_{t-k})$$

- 反映 Y_t 与 Y_{t-k} 的线性相关程度
- 取值 $\in [-1, 1]$
- 平稳过程中： $\rho_{-k} = \rho_k$

随机游走过程：单次实现与多路径比较



(a) 随机游走过程的单次实现



(b) 50 条独立随机游走路径

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T Y_t = -3.192129$$

$$\hat{\mu}_{100} = \frac{1}{50} \sum_{s=1}^{50} Y_{100}^s = -1.015433$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{\mu})^2 = 48.2165$$

$$\hat{\sigma}_{100}^2 = \frac{1}{50} \sum_{s=1}^{50} (Y_{100}^s - \hat{\mu}_{100})^2 = 92.785795$$

混合条件：相关性如何“衰减”

- **混合**: 描述“过去-未来”依赖随时间增大如何衰减
- 常见混合系数:
 - α -mixing (强混合)
 - β -mixing (绝对正则)
 - ϕ -mixing
 - ρ -mixing (最大相关系数)
- **共性**: 定义某种“过去事件-未来事件”的依赖度, 要求该依赖度随滞后 $n \rightarrow \infty$ 时趋于 0
- **作用**: 为 LLN、CLT 等渐近理论提供条件

定义

白噪声 $\{\varepsilon_t\}$ 满足：

- $E(\varepsilon_t) = 0$
 - $\text{Var}(\varepsilon_t) = \sigma^2$
 - $\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0, k \neq 0$
-
- 各期“扰动”不相关、同方差
 - 频谱平坦，类似“白光”含所有频率

自回归 AR(p) 模型

- AR(1):

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$$

- AR(p):

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- 应用:

- 金融: 收益率的短期自相关
- 宏观: 短期利率、通胀率等

移动平均 MA(q) 模型

- MA(1):

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

- MA(q):

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- 描述: 当前值由**当前和过去误差线性组合决定**

滞后算子 L 与多项式

- 定义: $LY_t = Y_{t-1}$
- $L^p Y_t = Y_{t-p}$, $L^{-1} Y_t = Y_{t+1}$
- AR(1) 可写为

$$(1 - \phi_1 L) Y_t = c + \varepsilon_t$$

- 一般滞后多项式:

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$$

- 在根位于单位圆外时可逆: $(1 - \phi L)^{-1} = \sum_{j=0}^{\infty} \phi^j L^j$

- 在可逆性条件下：

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad \Rightarrow \quad Y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

- 即 AR(1) \iff MA(∞)
- 类似地，MA(q) 也可表示为 AR(∞)
- 这也是 ARMA 模型中 ACF / PACF 都拖尾的原因

定义

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}.$$

- 同时利用“均值的线性自回归结构”和“误差的短期相关结构”
- 一般比纯 AR 或纯 MA 更灵活、参数更节省

ACF：自协方差与自相关

自协方差

$$\gamma_k = \text{Cov}(Y_t, Y_{t-k}) = E[(Y_t - \mu)(Y_{t-k} - \mu)]$$

自相关

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad \gamma_0 = \text{Var}(Y_t)$$

- 将 ρ_k 随 k 作图：ACF 图
- 观察衰减形态 判断模型类型

AR 与 MA 的 ACF 结构

- AR(1): $Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$

$$\rho_k = \phi_1^k, \quad k \geq 1$$

- 一般 AR(p): ACF 拖尾, 满足 Yule–Walker 递推
- MA(1): $y_t = \theta \varepsilon_{t-1} + \varepsilon_t$

$$\rho_1 = \frac{\theta}{1 + \theta^2}, \quad \rho_k = 0 \quad (k \geq 2)$$

- MA(q): ACF 在 q 阶截尾

PACF：偏自相关

- 第 k 阶偏自相关：在控制 $Y_{t-1}, \dots, Y_{t-k+1}$ 后， Y_t 与 Y_{t-k} 的“净”相关
- 形式上：AR(k) 中 Y_{t-k} 的估计系数

$$Y_t = \delta + \phi_1 Y_{t-1} + \cdots + \phi_k Y_{t-k} + \varepsilon_t \Rightarrow \text{PACF}_k = \phi_k$$

- 将 PACF 随 k 作图 PACF 图

ACF / PACF 判别规则

模型	ACF	PACF
AR(p)	拖尾	p 阶截尾
MA(q)	q 阶截尾	拖尾
ARMA(p, q)	拖尾	拖尾
其他线性过程	截尾	截尾

- AR: 看 PACF 截尾位置 $\Rightarrow p$
- MA: 看 ACF 截尾位置 $\Rightarrow q$
- ARMA: **两边都拖尾** 用信息准则 (AIC/BIC) 定阶

Yule–Walker 方程: AR(p) 估计

- AR(p):

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- Yule–Walker 方程:

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}, \quad k = 1, 2, \dots$$

- 矩阵形式 (略) 用样本自协方差代替真值, 解线性方程组得到
 $\hat{\phi}_1, \dots, \hat{\phi}_p$
- R 中: `ar.yw()` |

极大似然估计 (MLE) 回顾

似然与对数似然

$$L(\boldsymbol{\theta} \mid x_1, \dots, x_n) = f(x_1, \dots, x_n \mid \boldsymbol{\theta}), \quad \ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$$

- MLE: 选取使 $\ell(\boldsymbol{\theta})$ 最大的参数 $\hat{\boldsymbol{\theta}}$
- 通常对 ℓ 取导 $= 0$, 解方程
- 对正态样本 $N(\mu, \sigma^2)$:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$$

ARMA(p,q) 的 MLE 思路

- 模型：

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$$

- 给定参数与初始值，可递推得到 ε_t
- 条件对数似然：

$$\ell(\boldsymbol{\theta}) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2$$

- 数值优化（如 BFGS）最大化 ℓ
- R 中：arima(..., method = "ML") |

定义

$$\text{AIC} = -2\ell(\hat{\theta}) + 2k$$

$$\text{BIC} = -2\ell(\hat{\theta}) + k \log T$$

- ℓ : 极大化后对数似然
- k : 参数个数
- T : 样本长度
- 准则值越小越好 (拟合好 + 模型简约)
- R 中: `auto.arima(ic="aic") |, ic="bic" |`

一致性与无偏性

无偏性

$$E(\hat{\theta}) = \theta$$

一致性

$$\hat{\theta} \xrightarrow{P} \theta \quad (T \rightarrow \infty)$$

- 无偏并不保证一致；一致也不必然无偏
- 例：正态方差的 MLE $\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$ 有偏但一致

得分、Fisher 信息与 Hessian

- 得分函数 (Score) :

$$S(\theta) = \frac{\partial}{\partial \theta} \log f(X; \theta), \quad E[S(\theta)] = 0$$

- Fisher 信息 :

$$\mathcal{I}(\theta) = E [S(\theta)^2] = -E \left[\frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right]$$

- Hessian (对数似然二阶导) :

$$H(\theta) = \frac{\partial^2}{\partial \theta^2} \log f(X; \theta)$$

Cramér–Rao 下界与渐近正态性

- Cramér–Rao 不等式:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)}$$

- 在常规正则条件下, MLE 渐近有效:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}(\theta)^{-1})$$

- 大样本下, MLE 近似正态, 可用于构造置信区间与检验

- AR(1):

$$y_t = \theta y_{t-1} + \varepsilon_t$$

- 若 $\theta = 1$:

$$y_t = y_{t-1} + \varepsilon_t \Rightarrow \text{随机游走}$$

- 此时:

- 无条件方差随 t 增大而发散
- 过程非平稳，具有随机趋势
- 冲击影响是永久性的

ARMA 的平稳条件与单整过程

- 一般 ARMA(p,q):

$$\phi(L)y_t = \alpha(L)\varepsilon_t$$

- 平稳 AR 多项式 $\phi(z) = 0$ 的根都在单位圆外
- 若恰有一个单位根:

$$(1 - L)\theta^*(L)y_t = \alpha(L)\varepsilon_t$$

则 Δy_t 为平稳 ARMA y_t 为 $I(1)$

- 若需两次差分才平稳 $I(2)$

I(0) vs I(1) 的对比

- I(0):
 - 平稳，方差有限
 - 冲击的影响是暂时的
 - ACF 快速衰减
- I(1):
 - 非平稳，存在随机趋势
 - 冲击影响持久（“记忆长”）
 - ACF 缓慢衰减
- 建模策略不同：I(0) 直接 ARMA；I(1) 常需差分或做协整分析

ADF 检验（简述）

- 检验 y_t 是否存在单位根
- 典型形式：

$$\Delta y_t = \alpha y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \text{截距/趋势} + \varepsilon_t$$

- 原假设： $H_0: \alpha = 0$ (单位根)
- 备择： $H_1: \alpha < 0$ (平稳)
- 使用特殊的临界值 (非标准 t 分布)

目录

- ① 结构模型与时间序列模型之间的关系
- ② 长期方差 (Long-run Variance)
- ③ 采用 ARMA 模型进行预测
- ④ 案例：上证指数收益率的 ARMA 建模与预测
- ⑤ 向量自回归模型 VAR
- ⑥ 结构向量自回归模型 SVAR
- ⑦ SVAR 的 R 案例：A 模型与 B 模型
- ⑧ 章节总结

结构模型 vs 时间序列模型：是否矛盾？

- 结构模型（回归模型）：基于经济理论，关注因果关系和机制。

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

- 时间序列模型（如 ARMA）：基于时间依赖性，关注预测和动态特征。
- 问题：二者是否互相矛盾？
- 结论：不矛盾，而是互补——理论驱动的因果关系可以通过时间序列模型在数据中获得支持。

示例：结构模型叠加时间序列过程

假设

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t,$$

其中 ε_t 为白噪声；若

$$X_t = u_t + \alpha u_{t-1},$$

为 MA(1) 过程， u_t 与 ε_t 独立，且为白噪声。则

$$Y_t = \beta_0 + \beta_1 u_t + \alpha \beta_1 u_{t-1} + \varepsilon_t.$$

由此可得：

- ① $E(Y_t) = \beta_0;$
- ② $\text{Var}(Y_t) = \beta_1^2(1 + \alpha^2)\sigma_u^2 + \sigma_\varepsilon^2;$
- ③ $\text{Cov}(Y_t, Y_{t-1}) = \beta_1^2 \alpha \sigma_u^2;$
- ④ 对 $k \geq 2$, $\text{Cov}(Y_t, Y_{t-k}) = 0.$

即 Y_t 亦为 MA(1) 过程：因果结构与时间序列特征完全可以兼容。

VAR 与 SVAR：结构与时间序列的结合

VAR 模型：

- 多变量时间序列模型；
- 每个变量作为自身及其他变量滞后项的线性函数；
- 可用于预测和政策分析，刻画变量之间的动态相互影响。

SVAR 模型：

- 在 VAR 基础上加入结构假设；
- 引入经济理论，识别「谁是冲击、谁是反应」；
- 有助于识别结构性冲击（货币政策冲击、需求冲击等）。

结论：结构模型强调因果，时间序列模型强调动态与预测——结合使用可获得对经济现象更全面的理解。

目录

- 1 结构模型与时间序列模型之间的关系
- 2 长期方差 (Long-run Variance)
- 3 采用 ARMA 模型进行预测
- 4 案例：上证指数收益率的 ARMA 建模与预测
- 5 向量自回归模型 VAR
- 6 结构向量自回归模型 SVAR
- 7 SVAR 的 R 案例：A 模型与 B 模型
- 8 章节总结

长期方差的动机

- 金融时间序列常存在**相关性与异方差**，普通方差低估不确定性。
- 长期方差 (Long-run Variance, LRV) 刻画

$$\text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T y_t \right)$$

的极限——用于构造渐近标准误与稳健推断。

- HAC (Heteroskedasticity and Autocorrelation Consistent) 估计量：
 - 在存在自相关和异方差时仍能给出一致的方差估计；
 - 是计量经济学实证中非常核心的一类工具。

从部分和出发的长期方差

假设 $E(y_t) = 0$ 且 $E(y_t^2) \leq C < \infty$, 记

$$S_T = \sum_{t=1}^T y_t.$$

则

$$E(S_T^2) = \sum_{t=1}^T E(y_t^2) + \sum_{t=1}^T \sum_{\substack{s=1 \\ s \neq t}}^T E(y_t y_s).$$

若 $\{y_t\}$ 弱平稳, 记自协方差 $\gamma(k) = E(y_0 y_k)$, 有

$$E(S_T^2) = T\gamma(0) + 2 \sum_{k=1}^{T-1} (T-k)\gamma(k) = T\gamma(0) + 2T \sum_{k=1}^{T-1} \left(1 - \frac{k}{T}\right) \gamma(k).$$

再假设

$$\sum_{k=1}^{\infty} |\gamma(k)| < \infty,$$

则

$$1 - \frac{T}{T-1} = \left[\left(1 - \frac{T}{T-1} \right)^2 \right]$$

长期方差的定义

如果均值不为零，只需考虑 $y_t - E(y_t)$ ，则

$$\text{Var}\left(\frac{1}{\sqrt{T}}S_T\right) \rightarrow \text{Var}(y_0) + 2 \sum_{k=1}^{\infty} \text{Cov}(y_0, y_k) = \text{lrvar}.$$

- **长期方差**: 与短期方差（单期方差）相对，反映长期累积波动。
- 在谱分析中，这是频率零处谱密度的 2π 倍。

Parzen-Newey-West 估计量:

$$\widehat{\text{lrvar}} = \sum_{|k| \leq M_T} \left(1 - \frac{|k|}{M_T}\right) \widehat{\gamma}(k),$$

其中 $M_T \rightarrow \infty$ 且 $M_T/T \rightarrow 0$ ， $\widehat{\gamma}(k)$ 为样本自协方差。

在回归残差场景下即常说的 **Newey-West 估计量**。

回归模型中的长期方差

回归模型：

$$Y_t = X_t' \beta + u_t, \quad t = 1, \dots, T,$$

其中 X_t, β 为 $q \times 1$ 向量, u_t 外生,

$E(u_t | X_t) = 0$, u_t 允许自相关和条件异方差。

OLS 估计量:

$$\hat{\beta} = \left(\sum_{t=1}^T X_t X_t' \right)^{-1} \sum_{t=1}^T X_t Y_t.$$

记 $v_t = X_t u_t$, 在大数定律和中心极限定理条件下:

$$\frac{1}{T} \sum_{t=1}^T X_t X_t' \xrightarrow{p} Q, \quad \frac{1}{\sqrt{T}} \sum_{t=1}^T v_t \xrightarrow{} N(0, \Omega_v),$$

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{} N\left(0, Q^{-1} \Omega_v Q^{-1}\right).$$

其中 Ω_v 正是 $\{v_t\}$ 的长期方差。

鞅差序列与混合条件

鞅差序列：若

$$E(y_t | \mathcal{F}_{t-1}) = 0,$$

则 $\{y_t\}$ 为关于滤子 $\{\mathcal{F}_t\}$ 的鞅差序列。

- 在计量经济模型中常假设误差项是鞅差；
- 保证误差与解释变量不相关，从而保证估计的一致性。

定理 (Herrndorf (1984))

设 $E(y_t) = 0$, $E(|y_t|^{2+\delta}) \leq C < \infty$, 且存在 $\sigma^2 = \text{lrvvar} > 0$,

$$\frac{1}{T} E(S_T^2) \rightarrow \sigma^2, \quad \sum_{k=1}^{\infty} \alpha(k)^{\frac{\delta}{2+\delta}} < \infty,$$

则

$$\frac{1}{\sigma \sqrt{T}} S_T \implies N(0, 1),$$

其中 $\alpha(k)$ 为 α -混合系数。

矩条件与混合系数衰减

Herrndorf 定理体现出：

- 矩条件（高阶矩有限）；
- 混合条件（依赖性随滞后衰减）。
- 若 $\delta = \infty$, 只需

$$\sum_{k=1}^{\infty} \alpha(k) < \infty,$$

允许较慢的衰减，例如 $\alpha(k) = (k \log k)^{-1}$ 。

- 若 δ 较小，则需要更快的衰减速度。

这些结果为长期方差存在性与中心极限定理提供了充分的一般性理论基础。

目录

- ① 结构模型与时间序列模型之间的关系
- ② 长期方差 (Long-run Variance)
- ③ 采用 ARMA 模型进行预测
- ④ 案例：上证指数收益率的 ARMA 建模与预测
- ⑤ 向量自回归模型 VAR
- ⑥ 结构向量自回归模型 SVAR
- ⑦ SVAR 的 R 案例：A 模型与 B 模型
- ⑧ 章节总结

AR(1) 模型的一步预测

考虑 AR(1):

$$y_t = \theta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } (0, \sigma^2).$$

基于样本 $\{y_1, \dots, y_T\}$, 预测 y_{T+1} :

$$y_{T+1} = \theta y_T + \varepsilon_{T+1},$$

条件期望预测:

$$\hat{y}_{T+1|T} = E(y_{T+1} | y_1, \dots, y_T) = \theta y_T.$$

预测误差:

$$e_{T+1|T} = y_{T+1} - \hat{y}_{T+1|T} = \varepsilon_{T+1},$$

$$E(e_{T+1|T}) = 0, \quad \text{Var}(e_{T+1|T}) = \sigma^2.$$

条件期望在均方误差意义下是最优预测。

多步预测与预测误差方差

对 r 步预测：

$$y_{T+r} = \theta^r y_T + \theta^{r-1} \varepsilon_{T+1} + \cdots + \varepsilon_{T+r}.$$

预测值：

$$\hat{y}_{T+r|T} = \theta^r y_T.$$

预测误差：

$$y_{T+r} - \hat{y}_{T+r|T} = \theta^{r-1} \varepsilon_{T+1} + \cdots + \varepsilon_{T+r},$$

$$\text{Var}(y_{T+r} - \hat{y}_{T+r|T}) = \sigma^2 (1 + \theta^2 + \cdots + \theta^{2(r-1)}).$$

随着 r 增大，预测误差方差增加；当 $|\theta| < 1$ 时，该方差收敛到

$$\sigma^2 \frac{1}{1 - \theta^2}.$$

参数估计误差与预测区间

实务中使用估计量 $\hat{\theta}$:

$$\hat{y}_{T+r|T} = \hat{\theta}^r y_T.$$

预测误差可拆分为

$$\begin{aligned} y_{T+r} - \hat{y}_{T+r|T}(\hat{\theta}) &= [y_{T+r} - \hat{y}_{T+r|T}(\theta)] + [\hat{y}_{T+r|T}(\theta) - \hat{y}_{T+r|T}(\hat{\theta})] \\ &= \theta^{r-1} \varepsilon_{T+1} + \cdots + \varepsilon_{T+r} + (\theta^r - \hat{\theta}^r) y_T. \end{aligned}$$

若 $\hat{\theta} \xrightarrow{P} \theta$, 则 $(\hat{\theta}^r - \theta^r) \xrightarrow{P} 0$, 参数估计误差在大样本中可以忽略。

在常规正态假设下, 可以构造 r 步预测的置信区间:

$$\hat{y}_{T+r|T} \pm z_{\alpha/2} \sqrt{\hat{\sigma}^2 (1 + \hat{\theta}^2 + \cdots + \hat{\theta}^{2(r-1)})}.$$

一个重要的注意点

注意：预测区间的长度不会因样本量 T 增加而收缩到 0。

- OLS 系数的估计误差会随 T 增大而减小；
- 但未来冲击 $\varepsilon_{T+1}, \dots, \varepsilon_{T+r}$ 的不确定性不会消失。

因此，即使模型估计得再精确，只要存在随机冲击，预测区间总会保持一个正的宽度，这与参数估计的置信区间随样本量缩小的直觉不同。

一般 ARMA 过程的预测

设 y_t 有无穷阶 AR 表示：

$$y_t = \mu + \sum_{j=1}^{\infty} \theta_j(y_{t-j} - \mu) + \varepsilon_t.$$

对 ARMA 可逆过程，总能写成上述形式。

目标：基于 $\{y_1, \dots, y_T\}$ 预测 y_{T+1} ：

$$\begin{aligned}\hat{y}_{T+1|T} &= \mu + \sum_{j=1}^{\infty} \theta_j(y_{T+1-j} - \mu) \\ &= \mu + \sum_{j=1}^T \theta_j(y_{T+1-j} - \mu) + \sum_{j=T+1}^{\infty} \theta_j(y_{T+1-j} - \mu).\end{aligned}$$

实际上既看不到 y_0, y_{-1}, \dots ，也不知道 μ, θ_j 真值。

常见做法：

$$\hat{y}_{T+1|T} = \hat{\mu} + \sum_{j=1}^T \hat{\theta}_j(y_{T+1-j} - \hat{\mu}),$$

即用有限阶截断 + 参数估计替代。

样本外 R^2 与预测性能评价

将总样本 $\{y_1, \dots, y_n\}$ 分为：

估计样本 $\{y_1, \dots, y_T\}$, 评估样本 $\{y_{T+1}, \dots, y_{T+K}\}$,

$$n = T + K.$$

Campbell 等提出的样本外 R^2 :

$$R^2_{OOS} = 1 - \frac{\sum_{j=1}^K (y_{T+j} - \hat{y}_{T+j|T})^2}{\sum_{j=1}^K (y_{T+j} - \bar{y})^2},$$

其中 \bar{y} 为全样本均值。

- 分母对应简单「常数均值」预测；
- $R^2_{OOS} > 0$ 表示预测优于简单平均；
- 是比较不同预测模型在未来样本表现的自然指标。

滚动预测与逐步更新的 R^2

若采用一步一步的滚动预测 (one-step ahead forecast)，在预测 $T+j$ 时使用样本

$$\{y_1, \dots, y_{T+j-1}\},$$

并更新均值为 $\bar{y}_{1:T+j-1}$ ，则定义

$$R_{OOS,j}^2 = 1 - \frac{\sum_{j=1}^K (y_{T+j} - \hat{y}_{T+j|T+j-1})^2}{\sum_{j=1}^K (y_{T+j} - \bar{y}_{1:T+j-1})^2}.$$

- 更贴近「实时预测」场景；
- 同样用于衡量相对简单基准（滚动均值）是否被改进。

等预测准确性检验：Diebold–Mariano

两个预测方法，误差分别为

$$\hat{\varepsilon}_j, \quad \hat{\varepsilon}_j^*, \quad j = 1, \dots, K.$$

定义「损失差」：

$$d_j = \hat{\varepsilon}_j - \hat{\varepsilon}_j^*.$$

在原假设下， $E(d_j) = 0$ （两种预测的平均损失相同）。

Diebold–Mariano 检验统计量：

$$S = \frac{\frac{1}{\sqrt{K}} \sum_{j=1}^K d_j}{\widehat{\text{lrvar}}(d_j)},$$

其中 $\widehat{\text{lrvar}}(d_j)$ 为 $\{d_j\}$ 的长期方差估计。

若 K 足够大且 $\{d_j\}$ 近似平稳，则 $S \approx N(0, 1)$ ，可据此检验两种预测是否存在显著性能差异。

过拟合与样本外表现

- 一个在样本内拟合极好 (R^2 接近 1 甚至等于 1) 的模型，未必在样本外预测表现良好。
- 过拟合模型往往过度追踪历史数据的「噪声」和偶然特征。
- 即便某个方法在某段时间表现突出，也无法保证在其他时期仍然有效：
 - 经济环境变化；
 - 市场结构改变；
 - 政策调整等。

因此：要持续评估和调整预测模型，不能迷信单一模型。

目录

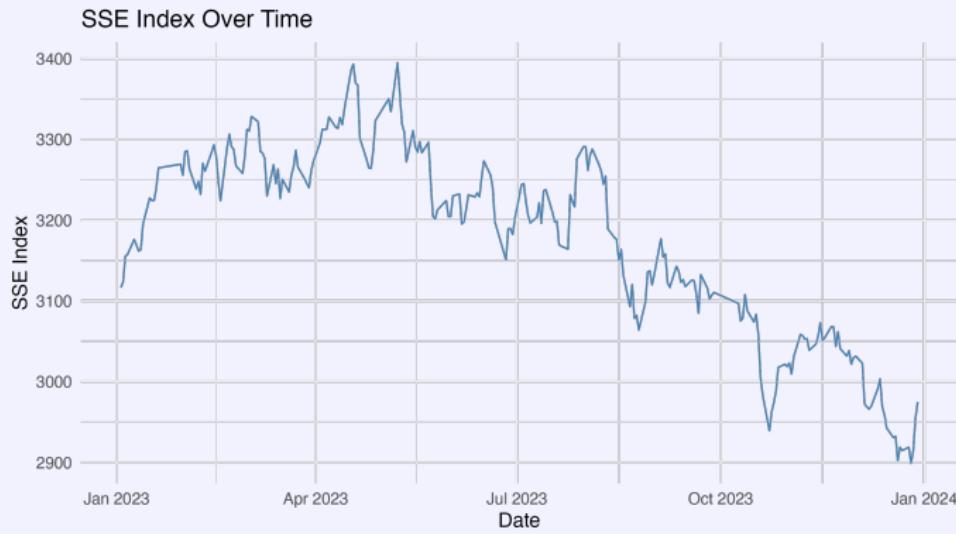
- 1 结构模型与时间序列模型之间的关系
- 2 长期方差 (Long-run Variance)
- 3 采用 ARMA 模型进行预测
- 4 案例：上证指数收益率的 ARMA 建模与预测
- 5 向量自回归模型 VAR
- 6 结构向量自回归模型 SVAR
- 7 SVAR 的 R 案例：A 模型与 B 模型
- 8 章节总结

数据介绍：上证综合指数

- 上证综合指数最初源于「静安指数」，后被上证指数取代。
- 采用市值加权，反映上海证券交易所挂牌股票的整体走势。
- 本节数据：
 - 2023 年 1 月 3 日至 2023 年 12 月 29 日；
 - 使用**日度收盘价**，构造成对数收益率：

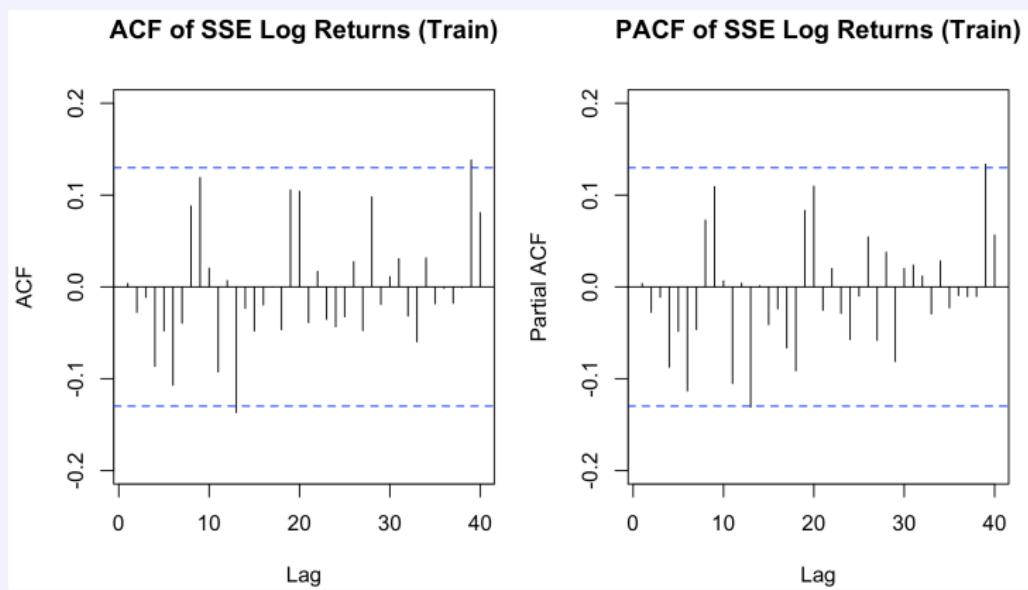
$$r_t = \log \left(\frac{p_t}{p_{t-1}} \right).$$

指数走势



2023 年上证综合指数的日度走势。之后我们基于该指数构造对数收益率并进行 ARMA 建模与预测。

ACF/PACF：收益率的相关性特征



- ACF 图：无显著滞后超出置信区间，说明自相关性不明显。
- PACF 图：同样无显著滞后。
- 这是资产收益的一个常见特征：在处理平稳后，收益率本身通常接近白噪声。

auto.arima 模型选择

使用 R 语言 forecast 包的 auto.arima:

- 通过极大似然估计不同 ARIMA 规格；
- 比较 AIC/AICc/BIC 选出最佳模型；
- 过程自动完成，适合实务中的初步建模。

模型输出示意：

$\text{sigma}^2 = 5.308e-05 : \loglikelihood = 798.65 AIC = -1595.31 AICc = -1595.29 BIC = -1591.88$

即拟合结果为

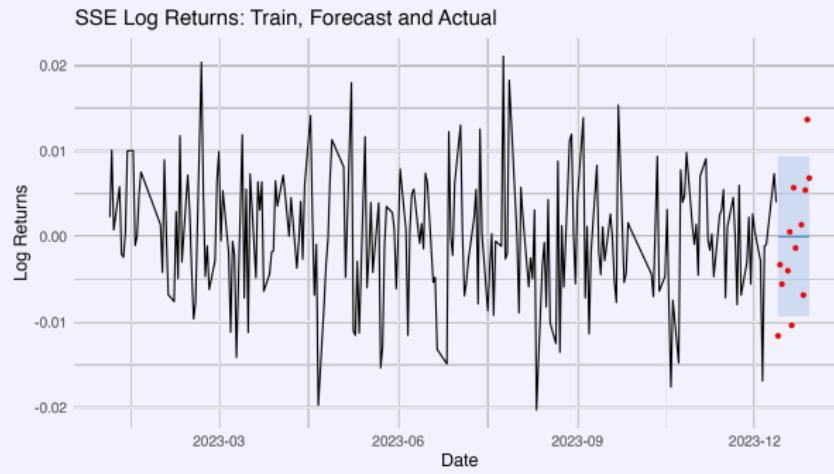
$$r_t = \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2),$$

说明在该样本中，线性自相关结构极弱。

训练误差指标的含义

- ME: Mean Error, 平均误差, 反映整体偏差方向与大小。
- RMSE: Root Mean Squared Error, 均方根误差, 放大较大误差的影响。
- MAE: Mean Absolute Error, 平均绝对误差, 比 RMSE 稳健。
- MPE: Mean Percentage Error, 带符号的平均百分比误差。
- MAPE: Mean Absolute Percentage Error, 平均绝对百分比误差, 便于不同量纲比较。
- MASE: Mean Absolute Scaled Error, 与基准模型（如随机游走或均值）比较, $MASE < 1$ 表示优于基准。
- ACF1: 一阶残差自相关, 越接近 0 表明残差相关性越弱。

预测结果与置信区间



- 黑线：训练样本对数收益率；
- 蓝线：预测期内的点预测；
- 蓝色带：预测区间；
- 红点：真实收益率。

可见，大多数真实值落在预测区间内，与模型「近似白噪声」的结论一致：
最好的线性预测基本就是零。

样本外 R^2 计算

样本外 R^2 的 R 代码（略作修正）：

```
mean_train <- mean(train_data)
ss_total <- sum( (test_data - mean_train)^2 )
ss_residual <- sum( (test_data - forecasts$mean)^2 )

R_OOS_2 <- 1 - (ss_residual / ss_total)
print(paste("Out-of-sample R^2:", R_OOS_2))
```

- 若 $R_{OOS}^2 > 0$, 说明 ARIMA 预测优于「常数均值」基准；
- 在许多股票收益数据上, R_{OOS}^2 往往非常接近 0, 表明线性预测能力有限；
- 这也凸显了引入波动建模（如 GARCH）及非线性方法的重要性。

上证指数：数据读取与预处理

```
# ----- 上证指数：数据读取、可视化、对数收益率计算与 ARIMA 预测 -----

# 1) 将工作目录设为当前脚本所在目录（在 RStudio 下有效）
if (requireNamespace("rstudioapi", quietly = TRUE) &&
    rstudioapi::isAvailable() &&
    !is.null(rstudioapi::getActiveDocumentContext()$path)) {
  setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
}

# 2) 加载必要的库
if (!requireNamespace("ggplot2", quietly = TRUE)) install.packages("ggplot2")
if (!requireNamespace("forecast", quietly = TRUE)) install.packages("forecast")
if (!requireNamespace("tseries", quietly = TRUE)) install.packages("tseries")
if (!requireNamespace("dplyr", quietly = TRUE)) install.packages("dplyr")
library(ggplot2)
library(forecast)
library(tseries)
library(dplyr)

# 3) 读取数据并预处理
sse <- read.csv("sse.csv", stringsAsFactors = FALSE)
sse$date <- as.Date(sse$date)
sse <- sse[order(sse$date), ]
```

R 代码（续）：图形与样本划分

```

# 4) 指数随时间的折线图
sse_plot <- ggplot(sse, aes(x = Date, y = SSE)) +
  geom_line(color = "steelblue", linewidth = 0.5) +
  labs(title = "SSE Index Over Time", x = "Date", y = "SSE Index") +
  theme_minimal(base_size = 12)

ggsave("sse_index_plot.png", plot = sse_plot,
       width = 8, height = 4.5, dpi = 300)

# 5) 计算对数收益率并划分训练/测试（训练 95%，测试 5%）
sse$Log_Returns <- c(NA, diff(log(sse$SSE)))
sse <- na.omit(sse)
split_index <- floor(0.95 * nrow(sse))
train_data <- sse$Log_Returns[1:split_index]
test_data <- sse$Log_Returns[(split_index + 1):nrow(sse)]
train_date <- sse$Date[1:split_index]
test_date <- sse$Date[(split_index + 1):nrow(sse)]

# 6) ACF / PACF 图
png("sse_acf_pacf.png", width = 8, height = 4.5,
    units = "in", res = 300)
par(mfrow = c(1, 2), mar = c(4, 4, 3, 1) + 0.1)
Acf(train_data, main = "ACF of SSE Log Returns (Train)", lag.max = 40)
Pacf(train_data, main = "PACF of SSE Log Returns (Train)", lag.max = 40)
dev.off()

```

R 代码（续）：ARIMA 拟合与预测图

```
# 7) 拟合 ARIMA 并预测
set.seed(123)
model <- auto.arima(train_data)
print(summary(model))

forecasts <- forecast(model, h = length(test_data))

# 8) 组织绘图数据
train_df <- data.frame(Date = train_date,
                        Value = as.numeric(train_data))
fc_df <- data.frame(
  Date      = test_date,
  Forecast = as.numeric(forecasts$mean),
  Lower    = as.numeric(forecasts$lower[, "80%"]),
  Upper    = as.numeric(forecasts$upper[, "80%"]))
)
actual_df <- data.frame(Date = test_date,
                         Actual = as.numeric(test_data))
```

R 代码（续）：训练 + 预测 + 真实值合图

```
# 9) 训练+预测+真实值合图
final_plot <- ggplot() +
  geom_line(data = train_df,
            aes(x = Date, y = Value),
            color = "black", linewidth = 0.5) +
  geom_line(data = fc_df,
            aes(x = Date, y = Forecast),
            color = "steelblue", linewidth = 0.6) +
  geom_ribbon(data = fc_df,
              aes(x = Date, ymin = Lower, ymax = Upper),
              fill = "steelblue", alpha = 0.2) +
  geom_point(data = actual_df,
             aes(x = Date, y = Actual),
             color = "red", size = 1.2) +
  scale_x_date(date_breaks = "3 month", date_labels = "%Y-%m") +
  labs(title = "SSE Log Returns: Train, Forecast and Actual",
       x = "Date", y = "Log Returns") +
  theme_minimal(base_size = 12)

ggsave("sse_forecast_plot.png", plot = final_plot,
       width = 8, height = 4.5, dpi = 300)
```

R 代码（续）：样本外 R^2

```
# 10) 样本外 R^2
mean_train <- mean(train_data)
ss_total    <- sum( (test_data - mean_train)^2 )
ss_residual <- sum( (test_data - forecasts$mean)^2 )

R_OOS_2 <- 1 - (ss_residual / ss_total)
print(paste("Out-of-sample R^2:", R_OOS_2))
```

目录

- 1 结构模型与时间序列模型之间的关系
- 2 长期方差 (Long-run Variance)
- 3 采用 ARMA 模型进行预测
- 4 案例：上证指数收益率的 ARMA 建模与预测
- 5 向量自回归模型 VAR
- 6 结构向量自回归模型 SVAR
- 7 SVAR 的 R 案例：A 模型与 B 模型
- 8 章节总结

VAR(p) 模型的定义

向量自回归 (VAR) 是 AR 模型的多元扩展：

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t,$$

- \mathbf{y}_t : $n \times 1$ 向量 (如多种宏观/金融变量)。
- \mathbf{c} : $n \times 1$ 常数向量。
- Φ_j : $n \times n$ 系数矩阵, $j = 1, \dots, p$ 。
- ε_t : $n \times 1$ 白噪声向量,

$$E(\varepsilon_t) = 0, \quad E(\varepsilon_t \varepsilon_\tau') = \begin{cases} \Omega, & t = \tau, \\ 0, & t \neq \tau, \end{cases}$$

其中 Ω 为对称正定矩阵。

VAR 可以刻画多个时间序列之间的动态相互作用。

滞后算子形式

使用滞后算子 L ($L\mathbf{y}_t = \mathbf{y}_{t-1}$) 可写为

$$(\mathbf{I}_n - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p) \mathbf{y}_t = \mathbf{c} + \varepsilon_t,$$

或者记为

$$\Phi(L) \mathbf{y}_t = \mathbf{c} + \varepsilon_t,$$

其中 $\Phi(L)$ 是 $n \times n$ 滞后多项式矩阵，其 (i,j) 元素为

$$\Phi_{ij}(L) = \delta_{ij} - \phi_{ij}^{(1)} L - \cdots - \phi_{ij}^{(p)} L^p,$$

δ_{ij} 为 Kronecker 符号。

这种表示形式有利于分析：

- 平稳性与特征根；
- VAR 与 VMA 表示之间的转换；
- 进一步构造 IRF、预测等。

VAR(p) 的 companion 形式

Hamilton 将 VAR(p) 写成 VAR(1):

$$\xi_t = \mathbf{F} \xi_{t-1} + \mathbf{v}_t,$$

其中

$$\xi_t = \begin{bmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p+1} - \boldsymbol{\mu} \end{bmatrix}, \quad \mathbf{v}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}.$$

VAR(p) 的 companion 形式

误差协方差矩阵：

$$\mathbf{Q} = \begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

平稳性与特征值条件

由

$$\xi_{t+s} = \mathbf{v}_{t+s} + \mathbf{F}\mathbf{v}_{t+s-1} + \cdots + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^s\xi_t$$

可见，若 $\mathbf{F}^s \rightarrow 0$ ($s \rightarrow \infty$)，则冲击会衰减。

- VAR(p) 协方差平稳的充要条件：矩阵 \mathbf{F} 的所有特征值均在单位圆内（模小于 1）。
- 这等价于滞后多项式的根在单位圆之外。

平稳性与特征值条件

Proposition (Hamilton 命题 10.1)

矩阵 Φ 的特征值 λ 满足

$$\det(\lambda^p \mathbf{I}_n - \Phi_1 \lambda^{p-1} - \cdots - \Phi_p) = 0.$$

当所有满足上述方程的 λ 的模小于 1 时, VAR(p) 协方差平稳。等价地,

$$\det(\mathbf{I}_n - \Phi_1 z - \cdots - \Phi_p z^p) = 0$$

的所有根 z 位于单位圆外。

高斯 VAR(p) 与条件似然

高斯 VAR(p):

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{i.i.d. } N(\mathbf{0}, \Omega).$$

- 变量维度: n ;
- 有效样本量: T (前 p 期视为给定)。

令

$$\mathbf{x}_t = \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix}, \quad \Pi' = [\mathbf{c} \quad \Phi_1 \quad \cdots \quad \Phi_p],$$

则

$$\mathbf{y}_t \mid \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p} \sim N(\Pi' \mathbf{x}_t, \Omega).$$

对数似然与 MLE

条件对数似然：

$$\begin{aligned} L(\theta) &= \sum_{t=1}^T \log f(\mathbf{y}_t \mid \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \theta) \\ &= -\frac{Tn}{2} \log(2\pi) + \frac{T}{2} \log |\Omega^{-1}| - \frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \Pi' \mathbf{x}_t)' \Omega^{-1} (\mathbf{y}_t - \Pi' \mathbf{x}_t). \end{aligned}$$

最大化可得闭式解：

$$\hat{\Pi}' = \left(\sum_{t=1}^T \mathbf{y}_t \mathbf{x}_t' \right) \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1},$$

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t', \quad \hat{\varepsilon}_t = \mathbf{y}_t - \hat{\Pi}' \mathbf{x}_t.$$

在高斯假设下，这些也是 OLS 的估计量，具有良好的渐近性质（渐近正态）。

为什么阶数选择很重要？

VAR(p) 中的 p 决定了保留多少历史信息：

- p 过大：模型过于复杂，过拟合，参数多、自由度低；
- p 过小：无法刻画所有动态相关结构，欠拟合。

常用方法：

- 信息准则：AIC, BIC, HQIC；
- 残差相关性检验（LM 检验等）；
- 经济直觉与实务经验。

信息准则：AIC、BIC 与 HQIC

假设模型参数个数为 k , 最大似然值为 L , 样本量为 T :

$$\text{AIC} = 2k - 2 \log L,$$

$$\text{BIC} = \log(T) k - 2 \log L.$$

HQIC 的惩罚强度介于 AIC 与 BIC 之间。

直觉：

- AIC 惩罚相对较弱, 倾向选择较高阶 (更复杂的模型);
- BIC 惩罚随 $\log T$ 增长, 大样本下更容易选择较低阶;
- HQIC 折中, 兼顾拟合与简约性。

实务中常比较多个准则, 并结合残差诊断综合判断。

格兰杰因果：基本思想

核心问题：变量 Y 的过去是否有助于预测 X 的未来？

定义

若对所有 $s > 0$,

$$\text{MSE}[\text{E}(x_{t+s} \mid x_t, x_{t-1}, \dots)] = \text{MSE}[\text{E}(x_{t+s} \mid x_t, x_{t-1}, \dots, y_t, y_{t-1}, \dots)],$$

则 y 不是 x 的格兰杰原因。若加入 y 的历史显著降低 MSE，则称 y 是 x 的格兰杰原因。

关键词：预测性因果，是关于预测能力而非哲学上的真因果。

VAR 表示与约束检验

对二元 VAR(p):

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \sum_{j=1}^p \begin{bmatrix} \phi_{11}^{(j)} & \phi_{12}^{(j)} \\ \phi_{21}^{(j)} & \phi_{22}^{(j)} \end{bmatrix} \begin{bmatrix} x_{t-j} \\ y_{t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

若 y 对 x 没有格兰杰因果，则对所有 j 有

$$\phi_{12}^{(j)} = 0.$$

- **约束模型**: 强制 $\phi_{12}^{(j)} = 0$;
- **无约束模型**: 自由估计 $\phi_{12}^{(j)}$;
- 通过 F 统计量或似然比统计量检验这些约束是否成立。

若拒绝原假设，则认为 y 是 x 的格兰杰原因。

格兰杰因果 真正因果

- Hamilton: 股价是股息的格兰杰原因, 但从经济逻辑看, 股息是股价的真正原因。
- 原因: 市场前瞻性——股价已经反映了对未来股息的预期。

Lutkepohl 给出反例:

- 原模型参数矩阵为下三角, x_t 不是 z_t 的格兰杰原因;
- 但经过线性可逆变换后, 表面上似乎出现 x_t 对 z_t 的格兰杰因果;
- 说明在变量变换下格兰杰因果关系可能发生表面上的变化。

结论: 格兰杰因果是有用的预测工具, 但不能机械地等同于「真实因果」, 仍需结合经济理论与结构分析。

从 VAR 到 VMA

将平稳 VAR 模型写成无穷阶向量 MA (VMA)：

$$\mathbf{y}_t = \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots$$

对某个期 t 的冲击 ε_t , 其对未来的影响:

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \varepsilon'_t} = \Psi_s, \quad s = 0, 1, 2, \dots$$

- Ψ_s 的第 (i, j) 元素 $\frac{\partial y_{i,t+s}}{\partial \varepsilon_{j,t}}$ 就是「 j 变量对 i 变量的 s 期脉冲响应函数」;
- 描述了一个单位冲击在系统中随时间的传播路径。

多重冲击与矩阵形式

设在 t 时刻的冲击为

$$\delta = (\delta_1, \dots, \delta_n)'$$

则对 \mathbf{y}_{t+s} 的影响为

$$\Delta \mathbf{y}_{t+s} = \Psi_s \delta.$$

- 不同分量的冲击可以叠加；
- 各个分量的 IRF 共同构成系统的动态响应；
- 在宏观和金融中，IRF 是理解「政策/冲击 \rightarrow 变量」动态关系的核心工具。

正交脉冲响应函数 (OIRF)

难点： ε_t 各元素一般相关。

- 直接对 $\varepsilon_{j,t}$ 加一单位，会同时改变其它误差分量；
- 不利于解释「纯粹的 j 变量冲击」。

解决方案：对 Ω 做 LDL（或 Cholesky）分解：

$$\Omega = ADA'$$

其中 A 为对角线为 1 的下三角矩阵， D 为对角正定矩阵。定义

$$\mathbf{u}_t = A^{-1}\varepsilon_t, \quad E(\mathbf{u}_t\mathbf{u}_t') = D,$$

则 \mathbf{u}_t 的各分量不相关。

OIRF 的表达式

对 $u_{j,t}$ 的单位冲击，其对未来 y_{t+s} 的影响为

$$\Psi_s \mathbf{a}_j,$$

其中 \mathbf{a}_j 是矩阵 A 的第 j 列。

在样本中：

- 用 OLS/MLE 得到 $\hat{\Phi}_1, \dots, \hat{\Phi}_p$;
- 由 $\hat{\Phi}_j$ 得到 $\hat{\Psi}_s$;
- 用

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

估计协方差，再做分解

$$\hat{\Omega} = \hat{A} \hat{D} \hat{A}'$$

- 得到样本 OIRF： $\hat{\Psi}_s \hat{\mathbf{a}}_j$ 。

将其作为 s 的函数绘图，即得到常见的脉冲响应图。

目录

- 1 结构模型与时间序列模型之间的关系
- 2 长期方差 (Long-run Variance)
- 3 采用 ARMA 模型进行预测
- 4 案例：上证指数收益率的 ARMA 建模与预测
- 5 向量自回归模型 VAR
- 6 结构向量自回归模型 SVAR
- 7 SVAR 的 R 案例：A 模型与 B 模型
- 8 章节总结

VAR 的局限与 SVAR 的动机

Reduced-form VAR:

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim (0, \Omega).$$

- 只能刻画滞后关系；
- 误差项之间可能相关，反映同期结构，但未显式建模；
- 无法直接用来检验经济理论、解释结构性冲击。

SVAR 的目标：

- 明确刻画变量之间的同期（即时）关系；
- 利用经济理论对冲击进行结构性识别；
- 服务于政策分析、冲击分解等任务。

结构形式与识别问题

一般结构形式（简化常数项）：

$$\mathbf{y}_t = A_1^* \mathbf{y}_{t-1} + \cdots + A_p^* \mathbf{y}_{t-p} + \mathbf{e}_t,$$

其中

$$A_j^* = A\Phi_j, \quad \mathbf{e}_t = A\varepsilon_t, \quad E(\mathbf{e}_t \mathbf{e}_t') = \Sigma_e = A\Omega A'.$$

未识别问题：

- 结构形式参数个数： $(p+1)n^2 + \frac{n(n+1)}{2}$ ；
- Reduced form 只有 $pn^2 + \frac{n(n+1)}{2}$ 个可辨识参数；
- 差额： n^2 个结构参数需要额外约束；
- 若不施加限制，结构参数不可唯一确定。

A 模型、B 模型与 AB 模型

A 模型：

- 假定 Ω 为对角矩阵；
- 用下三角矩阵 A （对角为 1）刻画同期关系：

$$A\mathbf{y}_t = A_1^*\mathbf{y}_{t-1} + \cdots + A_p^*\mathbf{y}_{t-p} + \mathbf{e}_t;$$

- 结构性约束体现在 A 的零元素上。

B 模型：

- 写成

$$\mathbf{y}_t = \Phi_1\mathbf{y}_{t-1} + \cdots + \Phi_p\mathbf{y}_{t-p} + B\mathbf{e}_t, \quad \mathbf{e}_t \sim (0, I_n);$$

- 即 $\varepsilon_t = B\mathbf{e}_t$ ；
- 对 B 的结构施加约束（如下三角）。

AB 模型与长期约束 (Blanchard–Quah)

AB 模型：

$$A\epsilon_t = B\mathbf{e}_t,$$

其中 A, B 同时受结构约束，一般需要更多限制条件才能识别。

Blanchard–Quah 长期约束：

- 不直接约束 A 或 B ；
- 考虑不同冲击的长期效应（如永久 vs 暂时）；
- 例如技术冲击对产出有永久效应，而需求冲击的长期效应为零；
- 通过对 VMA 表示中长期响应施加约束来识别结构冲击。

这种方法在宏观分析中非常常见，如区分供给/需求冲击。

目录

- 1 结构模型与时间序列模型之间的关系
- 2 长期方差 (Long-run Variance)
- 3 采用 ARMA 模型进行预测
- 4 案例: 上证指数收益率的 ARMA 建模与预测
- 5 向量自回归模型 VAR
- 6 结构向量自回归模型 SVAR
- 7 SVAR 的 R 案例: A 模型与 B 模型
- 8 章节总结

R 案例: 模拟三元 VAR 过程

```

rm(list = ls())
set.seed(123)

timeLength <- 500

# VAR(1) 系数矩阵
CoeffMatrix <- matrix(c(0.25, 0.05, 0.20,
                         0.10, 0.28, 0.22,
                         0.65, 0.45, 0.28),
                        3, byrow = TRUE)

# 冲击矩阵 (下三角)
ImpactMatrix <- diag(1, 3)
ImpactMatrix[lower.tri(ImpactMatrix)] <- c(-0.1, -0.06, 0.25)

# 生成时间序列
timeSeriesData <- matrix(rnorm(3 * (timeLength + 1), 0, 1),
                           3, timeLength + 1)
for (i in 2:(timeLength + 1)) {
  timeSeriesData[, i] <- CoeffMatrix %*% timeSeriesData[, i - 1] +
    ImpactMatrix %*% rnorm(3, 0, 1)
}
timeSeriesData <- ts(t(timeSeriesData))
dimnames(timeSeriesData)[[2]] <- c("Series1", "Series2", "Series3")
plot.ts(timeSeriesData, main = "模拟的时间序列数据")

```

估计 Reduced-form VAR 与 A 模型

```
library(vars)

# 简化形式 VAR(1)
varModelEstimate <- VAR(timeSeriesData, p = 1, type = "none")

# A 模型: A 为下三角, 对角为 1
A_matrix <- diag(1, 3)
A_matrix[lower.tri(A_matrix)] <- NA

SVAR_A_Model <- SVAR(varModelEstimate, Amat = A_matrix,
                       max.iter = 1000)

SVAR_A_Model          # A 矩阵估计
SVAR_A_Model$Ase       # A 矩阵系数标准误
```

A 模型估计结果示意

```
> SVAR_A_Model
SVAR Estimation Results:
=====
Estimated A matrix:
  Series1 Series2 Series3
Series1  1.0000  0.0000    0
Series2  0.0506  1.0000    0
Series3  0.0407 -0.3191    1

> SVAR_A_Model$Ase
  Series1 Series2 Series3
Series1 0.000000 0.000000    0
Series2 0.044721 0.000000    0
Series3 0.044778 0.044721    0
```

解释：

- 对角元素为 1：标准化自身同期效应；
- 非对角元素：变量间的同期影响（如 Series3 对 Series2 为负）；
- 标准误反映估计精度，对角线为 0 因为这些元素被固定。

B 模型与 A 的逆

```
# A 的逆矩阵
solve(SVAR_A_Model$A)

# B 模型：B 为下三角，对角为 1
B_matrix <- diag(1, 3)
B_matrix[lower.tri(B_matrix)] <- NA

SVAR_B_Model <- SVAR(varModelEstimate, Bmat = B_matrix)

SVAR_B_Model      # B 矩阵估计
SVAR_B_Model$Bse  # B 系数标准误
```

B 模型估计结果示意

```

> solve(SVAR_A_Model$A)
      Series1   Series2   Series3
Series1  1.0000  0.00000     0
Series2 -0.0505  1.00000     0
Series3 -0.0568  0.31909     1

> SVAR_B_Model
Estimated B matrix:
      Series1   Series2   Series3
Series1  1.0000  0.0000     0
Series2 -0.0506  1.0000     0
Series3 -0.0568  0.3191     1

> SVAR_B_Model$Bse
      Series1   Series2   Series3
Series1 0.000000 0.000000     0
Series2 0.044721 0.000000     0
Series3 0.046943 0.044721     0

```

- B 矩阵近似为 A^{-1} ;
- 对角线为 1: 结构冲击对自身的即时效应归一为 1;
- 非对角元素刻画结构冲击在各变量间的同期传导。

脉冲响应函数的计算

```
# A 模型的脉冲响应函数  
irf_A_model <- irf(SVAR_A_Model, n.ahead = 10,  
                      boot = TRUE, ci = 0.95)  
plot(irf_A_model)  
  
# B 模型的脉冲响应函数  
irf_B_model <- irf(SVAR_B_Model, n.ahead = 10,  
                      boot = TRUE, ci = 0.95)  
plot(irf_B_model)
```

- `irf()` 自动计算（正交）脉冲响应并给出置信区间；
- 可用于分析结构冲击在多变量系统中的动态传播。

目录

- ① 结构模型与时间序列模型之间的关系
- ② 长期方差 (Long-run Variance)
- ③ 采用 ARMA 模型进行预测
- ④ 案例：上证指数收益率的 ARMA 建模与预测
- ⑤ 向量自回归模型 VAR
- ⑥ 结构向量自回归模型 SVAR
- ⑦ SVAR 的 R 案例：A 模型与 B 模型
- ⑧ 章节总结

本章主要内容回顾 (1/2)

- ARMA 模型：

- 平稳性、可逆性、ACF/PACF 与阶数识别；
- 尤尔-沃克方程、极大似然估计与渐近性质；
- AIC/BIC 阶数选择，ADF 单位根检验；
- 长期方差 (LRV) 与 HAC/Newey-West 估计。

- 预测：

- ARMA 模型的多步预测、预测误差方差；
- 样本内/样本外 R^2 ，模型比较；
- 以上证指数收益率为例的 ARMA 建模与预测。

本章主要内容回顾 (2/2)

- VAR 模型：

- VAR(p) 的定义与滞后算子表示；
- 平稳性条件 (companion 矩阵特征值)；
- 高斯 VAR 的极大似然估计；
- 阶数选择、格兰杰因果检验；
- 脉冲响应函数与正交脉冲响应。

- SVAR 模型：

- Reduced-form VAR 的局限与同期结构问题；
- 结构形式、识别问题与参数个数；
- A 模型、B 模型、AB 模型及长期约束方法；
- R 案例：三变量 VAR 的模拟与 SVAR 估计。

第 5 章波动率模型

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

为什么要研究波动率？

- 在金融时间序列中常见 **波动聚集现象**：
 - 大冲击后往往伴随一段时间的大冲击；
 - 小冲击后往往伴随一段时间的小冲击。
- 在较高频数据中尤为明显（如日度、周度收益率）。
- 传统模型假定误差方差为常数，难以刻画这种特征。
- **波动率模型**（尤其是 ARCH/GARCH）允许误差的方差随时间变化。

波动率在金融中的作用

- 波动率（特别是 **条件标准差**）是：
 - 期权定价的关键输入；
 - 风险管理（VaR、ES 等）的核心要素。
- 波动率越大，期权价值通常越高。
- 波动率模型帮助：
 - 更精确地估计和预测市场波动；
 - 区分高波动时期和低波动“平稳”时期；
 - 制定更有效的交易与风险管理策略。

条件均值 vs 条件方差

- 之前的 ARMA 模型是 **条件均值模型**:

$$Y_t = \mathbb{E}(Y_t | \mathcal{F}_{t-1}) + \varepsilon_t$$

- 刻画收益的可预测部分。

- 本章关注条件方差模型:**

$$\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$$

- 刻画收益的 风险 (波动率)。

- 金融计量实践:**

- 条件均值模型刻画收益;
- 条件方差模型刻画风险;
- 二者并不冲突, 可以叠加使用。

波动聚集现象示例

- 从美元兑人民币对数日收益率可以观察到：
 - 存在高波动期与低波动期，即波动聚集；
 - 波动率随时间连续演变，剧烈“跳变”较少；
 - 波动率有“上界”，在一定范围内波动 \Rightarrow 通常是平稳的。
- 在股票市场中，**价格大涨与大跌对未来波动率影响不对称——杠杆效应**：负向冲击往往提高波动率更多。

波动聚集现象示例

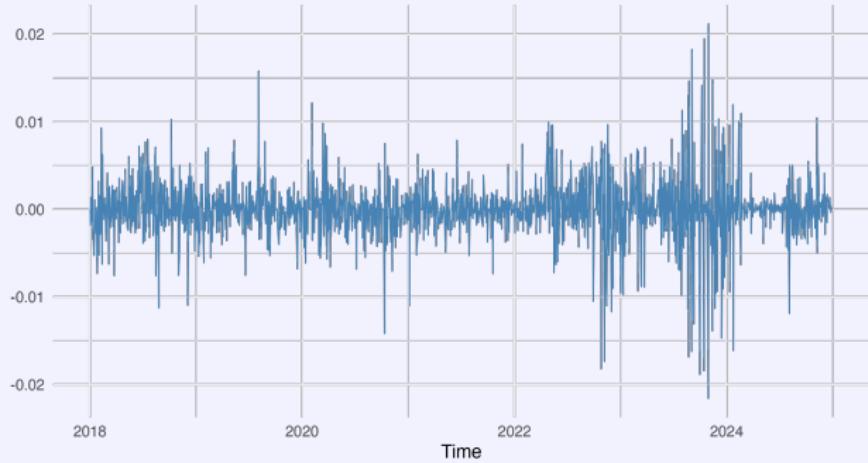


图: 美元兑人民币汇率对数日收益率 (2018-01-01 至 2024-12-31)

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

ARMA 模型中的误差项假设

- 在 ARMA 模型中常假定：

$$\varepsilon_t \sim \text{i.i.d. } (0, \sigma^2).$$

- 这是一个 **很强的假设**，比弱平稳所需条件要强。
- 例：MA(1) 模型

$$Y_t = \mu + \varepsilon_t + \alpha \varepsilon_{t-1}.$$

- 在 ε_t 零均值、常方差且不相关时，已可推出：

$$E(Y_t) = \mu, \quad \text{Var}(Y_t) = (1 + \alpha^2)\sigma^2, \quad \text{Cov}(Y_t, Y_{t-1}) = \alpha\sigma^2.$$

白噪声的定义

- 上述结果仅需误差满足：

$$\mathbb{E}(\varepsilon_t) = 0,$$

$$\mathbb{E}(\varepsilon_t^2) = \sigma^2,$$

$$\mathbb{E}(\varepsilon_t \varepsilon_{t-k}) = 0, \quad k \neq 0.$$

- 满足 (i)–(iii) 的过程称为 **白噪声**：

$$\varepsilon_t \sim \text{WN}(0, \sigma^2).$$

- 注意：白噪声只要求 **线性不可预测**，不排除

$$\text{Cov}(\varepsilon_t^2, \varepsilon_{t-k}^2) \neq 0$$

等非线性依赖结构。

i.i.d.、MDS 与白噪声

- i.i.d.: 独立同分布，限制最强。
- 鞅差分序列 (MDS):

$$E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0.$$

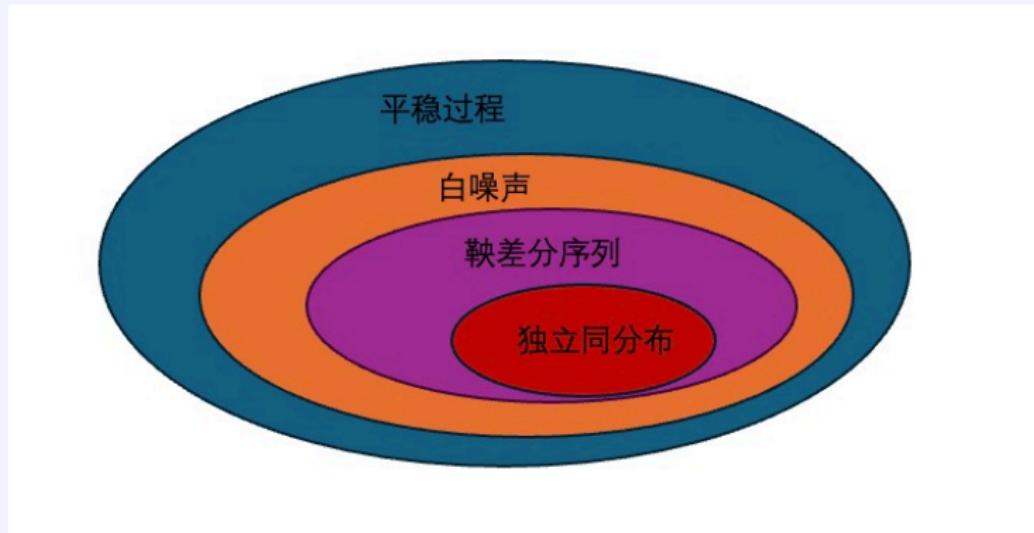
表示在给定过去信息时，当前的最佳预测为 0。

- 白噪声：仅要求不同时间点之间不存在线性可预测模式：

$$\rho(\varepsilon_t, \varepsilon_s) = 0, \quad t \neq s.$$

- 关系 (集合包含): i.i.d. \subset MDS \subset 白噪声 \subset 弱平稳。

i.i.d.、MDS 与白噪声



图：不同时间序列过程的关系示意

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

ARCH 模型的基本思想

- **ARCH**: Autoregressive Conditional Heteroskedasticity (自回归条件异方差)。
- 思路: 当前期的 条件方差依赖于过去误差项的 平方。
- 适合描述:
 - 波动率随时间变化;
 - 波动聚集现象。
- 均值方程用来刻画条件均值, 方差方程刻画条件方差:

$$Y_t = \mu_t + \varepsilon_t, \quad \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2.$$

Engle 与 ARCH 模型

罗伯特·F. 恩格尔 (Robert F. Engle III) 提出 ARCH 模型 Engle (1982)，并因此获得 2003 年诺贝尔经济学奖。

- 1969 年获康奈尔大学经济学博士；
- 曾任教于 MIT、UCSD，现任 NYU Stern 教授；
- 计量经济学会会士、美国艺术与科学院院士等。

ARCH 模型在经济、金融和统计等领域有广泛应用。

ARCH(1) 模型

- 最简单形式：ARCH(1)

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + u_t,$$

其中 u_t 是新的白噪声过程。

- 等价地，条件方差形式：

$$\sigma_t^2 = E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.$$

- 参数约束（保证 $\sigma_t^2 \geq 0$ ）：

$$\alpha_0 > 0, \quad \alpha_1 \geq 0.$$

- 注意：这并不意味着 ε_t 非平稳，只是说明 ε_{t-1}^2 与 σ_t^2 相关。

ARCH(1) 的无条件方差

- 设 $\sigma^2 = E(\varepsilon_t^2)$, 有

$$\sigma^2 = E(\sigma_t^2) = \alpha_0 + \alpha_1 E(\varepsilon_{t-1}^2) = \alpha_0 + \alpha_1 \sigma^2.$$

- 若 $0 \leq \alpha_1 < 1$, 则平稳解为

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1},$$

与时间 t 无关。

- 说明: 在此条件下, ARCH(1) 过程是方差有限且无条件方差不随时间变化的。

ARCH(p) 模型

- 条件方差形式:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2.$$

- 等价的 AR(p) 形式:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + u_t.$$

- 参数约束:

$$\alpha_0 > 0, \quad \alpha_j \geq 0, \quad j = 1, \dots, p,$$

$$\alpha_1 + \cdots + \alpha_p < 1$$

(保证平稳且方差有限)。

- 无条件方差:

$$\sigma^2 = \frac{\alpha_0}{1 - (\alpha_1 + \cdots + \alpha_p)}.$$

ARCH 模型与 Ljung–Box 检验

- 回顾：Ljung–Box 检验主要用于检测 **自相关性**。
- 在 ARCH 上的应用：
 - ARCH(p) 模型是 ε_t^2 的 AR(p)；
 - 若 ε_t^2 存在自相关，则意味着条件异方差存在。
- 检验步骤（思路）：
 - 拟合 ARCH(p)，得到残差 \hat{u}_t ；
 - 构造标准化残差：

$$z_t = \frac{\hat{u}_t}{\hat{\sigma}_t};$$

- 对 $\{z_t\}$ 进行 Ljung–Box 检验。

- Q 统计量：

$$Q = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{T-k},$$

其中 $\hat{\rho}_k$ 是残差的样本自相关。

ARCH(p) 模型：R 模拟与估计

```

library(rugarch)
library(ggplot2)
library(FinTS)

# 1. 模拟 ARCH( $p$ ) 模型
set.seed(123)
n      <- 1000
omega  <- 0.1
alphas <- c(0.5, 0.2, 0.1) #  $p = 3$ 
eps    <- rnorm(n)
sigma2 <- rep(0, n)
sigma2[1:length(alphas)] <- omega / (1 - sum(alphas))

for (i in (length(alphas) + 1):n) {
  sigma2[i] <- omega + sum(alphas * eps[(i-1):(i-length(alphas))])^2
  eps[i]    <- rnorm(1, sd = sqrt(sigma2[i]))
}

# 2. 估计不同阶数的 ARCH( $p$ ) 并用 AIC/BIC 选模
p_max     <- 5
aic_values <- bic_values <- rep(NA, p_max)

for (p in 1:p_max) {
  spec <- ugarchspec(
    variance.model = list(model = "sGARCH", garchOrder = c(p, 0)),
    mean.model     = list(armaOrder = c(0, 0), include.mean = FALSE)
  )
  fit <- try(ugarchfit(spec, data = eps), silent = TRUE)
  aic_values[p] <- infocriteria(fit)[1]
  bic_values[p] <- infocriteria(fit)[2]
}

```

ARCH(p) 模型：R 代码（续）

```
best_p_aic <- which.min(aic_values)
best_p_bic <- which.min(bic_values)

# 使用最优 p 拟合，并做 Ljung-Box 检验
best_spec <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(best_p_aic, 0)),
  mean.model     = list(armaOrder = c(0, 0), include.mean = FALSE)
)
best_fit <- ugarchfit(best_spec, data = eps)
summary(best_fit)

# Ljung-Box 检验（标准化残差）
ljung_box_test <- Box.test(
  best_fit@fit$residuals / best_fit@fit$sigma,
  lag = 10,
  type = "Ljung-Box"
)
ljung_box_test
```

ARCH 模型与 ARMA 模型的关系

- ARCH 模型：针对 **条件方差建模**，刻画 **波动聚集**。
- ARMA 模型：针对 **条件均值建模**，刻画 **线性依赖结构**。
- 实务中：
 - 先用 ARMA 模型拟合收益；
 - 将 ARMA 残差作为 ARCH/GARCH 模型中的 ε_t ；
 - 同时刻画“收益可预测性”和“风险随时间的变化”。

ARCH(1) 的平稳性分析

- 考虑 ARCH(1):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.$$

- 假设 $\{\varepsilon_t\}$ 弱平稳, 令

$$\text{Var}(\varepsilon_t) = \sigma^2 < \infty,$$

且不随 t 变化, 则必须满足

$$\sigma^2 = \text{E}(\sigma_t^2) = \alpha_0 + \alpha_1 \text{E}(\varepsilon_{t-1}^2) = \alpha_0 + \alpha_1 \sigma^2.$$

- 若 $|\alpha_1| < 1$, 有唯一解

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1}.$$

ARCH(1) 的渐近弱平稳性

- 从某个初始值 σ_1^2 出发,

$$\begin{aligned} E(\sigma_t^2) &= \alpha_0 + \alpha_1 E(\sigma_{t-1}^2) \\ &= \alpha_0 + \alpha_1 \alpha_0 + \alpha_1^2 E(\sigma_{t-2}^2) \\ &= \alpha_0 \sum_{j=1}^t \alpha_1^{j-1} + \alpha_1^t E(\sigma_1^2). \end{aligned}$$

- 若 $|\alpha_1| < 1$ 且 $E(\sigma_1^2) < \infty$, 则当 $t \rightarrow \infty$:

$$E(\sigma_t^2) \rightarrow \frac{\alpha_0}{1 - \alpha_1}.$$

- 说明: ARCH(1) 是 渐近弱平稳的; 若一开始就取 $E(\sigma_1^2) = \frac{\alpha_0}{1 - \alpha_1}$, 则对所有 t 都弱平稳。

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

从 ARCH 到 GARCH

- ARCH 能刻画波动聚集，但实际中常需要 较高阶。
- 高阶 ARCH：
 - 参数多，估计不稳定；
 - 浪费自由度，违反“简约原则”。
- Bollerslev (1986) 提出 **GARCH** (generalized ARCH)：
 - 将过去的条件方差也纳入递推；
 - 更灵活、更节省参数。

GARCH(p, q) 一般形式

- GARCH(p, q) 条件方差:

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^p \alpha_k \varepsilon_{t-k}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2.$$

- 滞后算子形式:

$$\sigma_t^2 = \alpha_0 + \alpha(L) \varepsilon_{t-1}^2 + \beta(L) \sigma_{t-1}^2,$$

其中

$$\alpha(L) = \sum_{k=1}^p \alpha_k L^k, \quad \beta(L) = \sum_{j=1}^q \beta_j L^j.$$

- GARCH 模型在实证中非常常用, 尤其是 GARCH(1,1)。

GARCH(1,1) 模型

- 形式:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

- 参数约束 (保证 $\sigma_t^2 > 0$):

$$\alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \beta_1 \geq 0.$$

- 若弱平稳成立, 则无条件方差满足

$$\sigma^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma^2 \quad \Rightarrow \quad \sigma^2 = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)},$$

$$\alpha_1 + \beta_1 < 1.$$

GARCH(1,1) 与无限阶 ARCH

- 递推展开：

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{j=1}^{\infty} \beta_1^{j-1} \varepsilon_{t-j}^2.\end{aligned}$$

- 说明：

- GARCH(1,1) 等价于 **系数按几何级数衰减的无限阶 ARCH**；
- 过去冲击对当前波动率的影响逐步减弱。

- 优点：

- 用少量参数刻画长期记忆；
- 比高阶 ARCH 模型更简洁。

GARCH 模型的 R 示例

```
library(rugarch)

set.seed(123)
n      <- 1000
eps    <- rnorm(n)
sigma2 <- rep(1, n)
y      <- rep(0, n)

alpha0 <- 0.01
alpha1 <- 0.05
beta1  <- 0.9
sigma2[1] <- alpha0 / (1 - alpha1 - beta1)

for (i in 2:n) {
  sigma2[i] <- alpha0 + alpha1 * y[i-1]^2 + beta1 * sigma2[i-1]
  y[i]       <- rnorm(1, sd = sqrt(sigma2[i]))
}
```

GARCH 模型的 R 示例

```
# 搜索不同 (p, q) 的 GARCH 模型
p_max <- 3; q_max <- 3
aic_values <- bic_values <- matrix(NA, p_max, q_max,
                                     dimnames = list(p = 1:p_max, q = 1:q_max))

for (p in 1:p_max) {
  for (q in 1:q_max) {
    spec <- ugarchspec(
      variance.model = list(model = "sGARCH", garchOrder = c(p, q)),
      mean.model     = list(armaOrder = c(0, 0), include.mean = FALSE)
    )
    fit <- try(ugarchfit(spec = spec, data = y), silent = TRUE)
    aic_values[p, q] <- infocriteria(fit)[1]
    bic_values[p, q] <- infocriteria(fit)[2]
  }
}
```

GARCH(p, q) 的弱平稳性

- 一般的 GARCH(p, q):

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2.$$

- 若

$$\sum_{j=1}^p \alpha_j + \sum_{k=1}^q \beta_k < 1,$$

则 $\{\varepsilon_t\}$ 是弱平稳的，并且

$$\text{Var}(\varepsilon_t) = E(\sigma_t^2) = \frac{\alpha_0}{1 - \sum_{j=1}^p \alpha_j - \sum_{k=1}^q \beta_k}.$$

- 条件类似 ARMA 模型中 AR 多项式根在单位圆之外的条件。

GARCH(1,1) 的强平稳性 (Nelson 定理)

定理 (Nelson 定理)

考慮

$$\varepsilon_t = \sigma_t \nu_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

其中 $\{\nu_t\}$ 独立同分布, $E(\nu_t) = 0$, $\text{Var}(\nu_t) = 1$, 且 $\alpha_0, \alpha_1 > 0$, $\beta_1 \geq 0$ 。
若

$$E[\log(\beta_1 + \alpha_1 \nu_t^2)] < 0,$$

则存在一个严格平稳解 $\{\sigma_{t,\text{st}}^2\}$, 且

$$\sigma_t^2 - \sigma_{t,\text{st}}^2 \rightarrow 0 \quad a.s.$$

并且 σ_t^2 的分布收敛到一个非退化的极限分布。

强平稳 vs 弱平稳

- 当 $E(\nu_t^2) = 1$ 时, 由 Jensen 不等式:

$$E[\log(\beta_1 + \alpha_1 \nu_t^2)] < \log(\alpha_1 + \beta_1).$$

- 即使 $\alpha_1 + \beta_1 \geq 1$, 也有可能

$$E[\log(\beta_1 + \alpha_1 \nu_t^2)] < 0,$$

从而 GARCH 过程仍可能是 **强平稳的**。

- 但此时可能不存在二阶矩, 即 **弱平稳失败**。
- 结论: GARCH 过程可以是强平稳而非弱平稳, 类似于前文提到的柯西分布例子。

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

为什么需要非对称波动率模型？

- 标准 ARCH / GARCH 的一个重要局限：**对称性**。
 - 只看新息的绝对值，不看符号；
 - 同幅度正负冲击对未来波动率的影响被设为相同。
- 现实中：**负向冲击通常带来更大波动**。
 - 风险厌恶：坏消息 \Rightarrow 恐慌抛售 \Rightarrow 不确定性上升；
 - 卖空限制：悲观预期难完全通过做空表达，只能离场，加剧波动；
 - 好消息虽有买入，但担心“利好难以持续”，买入力度相对克制。
- 因此需要能体现**杠杆效应**的非对称波动率模型。

EGARCH 模型：基本形式

- Nelson (1991) 提出的 **指数 GARCH (EGARCH)**:

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \alpha \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}}.$$

- 特点:

- 对数形式保证 $\sigma_t^2 > 0$, 参数无需非负约束;
- 含有 $\frac{\varepsilon_{t-1}}{\sigma_{t-1}}, \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}}$ 两项:
 - α 控制冲击绝对值对波动率的影响;
 - γ 控制冲击符号的非对称效应。
- 只要 $\gamma \neq 0$, EGARCH 就体现杠杆效应; 特别地,

$\gamma < 0 \Rightarrow$ 同幅度负向冲击引致的波动率上升大于正向冲击。

GJR-GARCH 模型

- Glosten, Jagannathan, and Runkle (1993) 提出的 GJR-GARCH(1,1):

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \gamma y_{t-1}^2 + \delta y_{t-1}^2 \mathbf{1}\{y_{t-1} < 0\},$$

其中 $\mathbf{1}\{\cdot\}$ 为指示函数。

- 解释:

- γ : 新息平方对波动率的“基准”影响;
- δ : **额外的负向冲击效应**;
- 当 $y_{t-1} < 0$ 时, 冲击对波动率的影响系数为 $\gamma + \delta$, 为正向冲击的影响 (γ) 加上额外部分。
- $\delta \neq 0$ 时, 新息冲击曲线呈现不对称形状。

新息冲击曲线 (news impact curve)

定义

新息冲击曲线 (news impact curve) 用于刻画不同规模与符号的“新息”(冲击)对未来条件方差(波动率)的影响。它揭示波动率的非对称(杠杆)效应：在许多市场中，同幅度的负向新息对未来波动率的提升通常大于正向新息。

新息冲击曲线 (news impact curve)

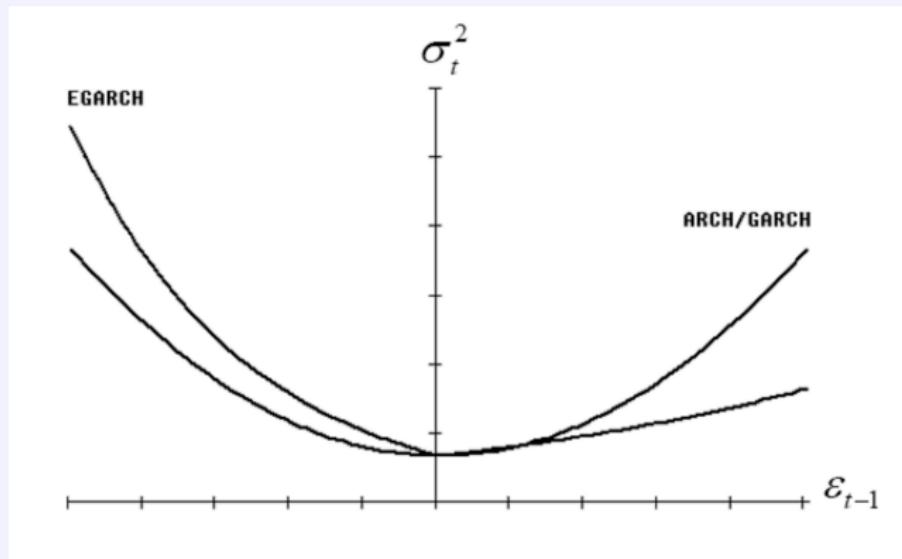


图: 新息冲击曲线示意图

更多非对称波动率模型

- 常见的非对称模型包括：
 - 门限 GARCH (Threshold GARCH) (Glosten, Jagannathan, and Runkle 1993);
 - 门限 ARCH (Zakoian 1994);
 - 马尔可夫区制转换 (Markov Regime-Switching) GARCH (Hamilton and Susmel 1994) 等。
- 核心思想：对不同符号、不同状态的冲击赋予不同的方差响应。

门限 GARCH 模型 (TGARCH)

- 门限 GARCH(1,1) 模型：

$$\begin{cases} \varepsilon_t = \nu_t \sigma_t, \\ \sigma_t^2 = \psi + \beta \sigma_{t-1}^2 + \alpha_1 \varepsilon_{t-1}^2 \mathbf{1}\{\nu_{t-1} < 0\} + \alpha_2 \varepsilon_{t-1}^2 \mathbf{1}\{\nu_{t-1} \geq 0\}, \\ \{\nu_t\} \sim \text{i.i.d. } (0, 1). \end{cases}$$

- 当 $\alpha_1 \neq \alpha_2$ 时，负向和正向冲击的方差响应不同：
 - 多数高频金融数据中常见 $\alpha_1 > \alpha_2$ ，表明负面冲击对条件方差影响更强。

门限 ARCH 模型

- 门限 ARCH 模型：

$$\sigma_t = \psi + \beta \sigma_{t-1} + \alpha_1 |\varepsilon_{t-1}| \mathbf{1}\{\nu_{t-1} > 0\} + \alpha_2 |\varepsilon_{t-1}| \mathbf{1}\{\nu_{t-1} \leq 0\}.$$

- 若 $\alpha_1 \neq \alpha_2$, 则波动率具有非对称性。
- 一般情形下 $\alpha_1 > \alpha_2$:
 - 负向冲击（坏消息）对条件方差 h_t 的影响更大；
 - 与前面讨论的杠杆效应一致。

马尔可夫区制转换 GARCH 模型 (MS-GARCH)

- MS-GARCH 模型形式：

$$\begin{cases} \varepsilon_t = \nu_t \sigma_t, \\ \sigma_t^2 = \psi(S_t) + \beta(S_t) \sigma_{t-1}^2 + \alpha(S_t) \varepsilon_{t-1}^2, \\ \{\nu_t\} \sim \text{i.i.d. } (0, 1), \end{cases}$$

其中状态变量 S_t 为马尔可夫链。

- 状态转移例：

$$P(S_t = 1 \mid S_{t-1} = 0) = p_{01}, \quad P(S_t = 0 \mid S_{t-1} = 1) = p_{10}.$$

- 系数依赖于状态 S_t ：
 - 可表示“扩张期 vs 衰退期”不同的波动水平；
 - 实证发现：股票收益率在衰退期波动率通常更高。
- 若允许扰动分布的形状参数也随状态变化，则 $\{\nu_t\}$ 成为 MDS 而非 i.i.d.

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

IGARCH 的动机：从 GARCH(1,1) 出发

- 考虑 GARCH(1,1):

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2.$$

- 若弱平稳成立且 $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, 则

$$E(\varepsilon_t^2) = E(\sigma_t^2) = \sigma^2,$$

且

$$\sigma^2 = \omega + (\alpha + \beta)\sigma^2 \Rightarrow \sigma^2 = \frac{\omega}{1 - \alpha - \beta}.$$

- 因此需满足

$$\omega > 0, \quad \alpha \geq 0, \quad \beta \geq 0, \quad \alpha + \beta < 1$$

才有有限无条件方差。

IGARCH: 带“单位根”的 GARCH

- 当

$$\alpha + \beta = 1$$

时，称为单整 GARCH(1,1)，即 **IGARCH(1,1)**。

- 设

$$u_t = \varepsilon_t^2 - \sigma_t^2 \quad (\text{MDS}),$$

由 $\varepsilon_{t-1}^2 = u_{t-1} + \sigma_{t-1}^2$ 得

$$\sigma_t^2 = \omega + (\alpha + \beta)\sigma_{t-1}^2 + \alpha u_{t-1}.$$

- 在 IGARCH 情形 ($\alpha + \beta = 1$):

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 + \alpha u_{t-1}.$$

IGARCH: 带“单位根”的 GARCH

- 递推展开:

$$\sigma_t^2 = \omega t + \sigma_0^2 + \alpha \sum_{l=1}^{t-1} u_{t-l},$$

从而

$$E(\sigma_t^2 | \mathcal{F}_0) = \omega t + \sigma_0^2.$$

IGARCH 与平方过程的 ARMA 表示

- 一般 GARCH(1,1) 的平方过程表示：

$$\varepsilon_t^2 = \omega + (\alpha + \beta) \varepsilon_{t-1}^2 - \beta u_{t-1} + u_t,$$

即 ARMA(1,1)。

- IGARCH(1,1) 情形：

$$\varepsilon_t^2 = \omega + \varepsilon_{t-1}^2 - \beta u_{t-1} + u_t.$$

- 特点：

- 波动率过程或平方过程具有“单位根”；
- 对任意滞后 j , 冲击 u_{t-j} 对 σ_t^2 或 ε_t^2 的影响不会消失；
- $\text{Var}(\varepsilon_t)$ 非良好定义（发散）。

- 对资产超额收益而言，无条件方差无限大往往不合理。

RiskMetrics：IGARCH 的一个特例

定义 (RiskMetrics)

J.P. Morgan (1997) 提出的 RiskMetrics 模型是 IGARCH(1,1) 的一个特例：

$$\begin{cases} \varepsilon_t = \nu_t \sigma_t, \\ \sigma_t^2 = (1 - \beta) \sum_{j=1}^{\infty} \beta^{j-1} \varepsilon_{t-j}^2, \quad \beta \in (0, 1), \\ \{\nu_t\} \sim \text{i.i.d. } N(0, 1). \end{cases}$$

等价递推形式：

$$\sigma_t^2 = \beta \sigma_{t-1}^2 + (1 - \beta) \varepsilon_{t-1}^2.$$

这对应于 IGARCH(1,1) 过程中的 $\omega = 0$, $\alpha = 1 - \beta$ 。

- β 为衰减因子，控制历史冲击的权重衰减速度；
- 日度金融数据中常用 $\beta \approx 0.94$ ；
- RiskMetrics 广泛用于风险管理（如 VaR 计算）。

IGARCH 的平稳性讨论与估计

- 若 $\omega > 0$, 则

$$E(\varepsilon_t^2) = \infty,$$

IGARCH(1,1) 不具弱平稳性。

- 严平稳性可通过 Jensen 不等式分析:

$$E[\log(\beta + \alpha u_t^2)] \leq \log(\beta + \alpha E[u_t^2]) = \log(\beta + \alpha) = 0.$$

- 即使 $\beta + \alpha$ 略大于 1, 也有可能 $E[\log(\beta + \alpha u_t^2)] < 0$, 但无条件方差仍发散。
- R 中可用 rugarch 估计 IGARCH(1,1):

IGARCH(1,1) 的 R 估计示例

```
# 设置 IGARCH(1,1) 模型
spec <- ugarchspec(
  variance.model = list(model = "iGARCH", garchOrder = c(1, 1)),
  mean.model     = list(armaOrder = c(0, 0)),
  distribution.model = "norm"
)

# 估计
fit <- ugarchfit(spec, data)
```

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

GARCH(1,1) 的模型设定

- 以 GARCH(1,1) 为例：

$$\begin{aligned}\varepsilon_t &= \nu_t \sigma_t, \\ \sigma_t^2(\theta) &= \omega + \beta \sigma_{t-1}^2(\theta) + \alpha \varepsilon_{t-1}^2,\end{aligned}$$

其中 $t = 2, \dots, T$, $\theta = (\omega, \beta, \alpha)'$ 。

- ε_t 可为条件均值模型（如 ARMA）的残差。
- 误差分布假设：

- 最常用： $\nu_t \sim N(0, 1)$;
- 对金融数据：也常用 Student- t 等重尾分布。

对数似然与 QMLE

- 条件高斯对数似然：

$$\ell(\theta) = \sum_{t=2}^T \ell_t(\theta), \quad \ell_t(\theta) = -\frac{1}{2} \log \sigma_t^2(\theta) - \frac{1}{2} \left(\frac{\varepsilon_t}{\sigma_t(\theta)} \right)^2,$$

其中 $\sigma_t^2(\theta)$ 由初始值递推得到。

- 记 $\hat{\theta} = \arg \max_{\theta \in \Theta} \ell(\theta)$ 。
- 若条件均值与条件方差设定正确，则
 - 在较强条件（如条件正态）下，高斯 MLE 一致、渐近正态且有效；
 - 在较弱条件下（误差非正态），高斯 MLE 仍是一致、渐近正态——即 QMLE。

初始条件的几种设定

- 对 GARCH(1,1) 的 σ_1^2 , 常见设定:

$$\textcircled{1} \quad \sigma_1^2(\theta) = \frac{\omega}{1 - \beta - \alpha} \text{ (假定弱平稳, 取无条件方差);}$$

$$\textcircled{2} \quad \sigma_1^2 = \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 \text{ (样本平均方差);}$$

$$\textcircled{3} \quad \sigma_1^2 = \varepsilon_1^2 \text{ (任意设定);}$$

$\textcircled{4}$ 将 σ_1^2 视作额外未知参数, 与 θ 一同估计。

- 不同选择对应不同的“先验”平稳假设, 有限样本表现略有区别, 大样本下影响较小。

对数似然的导数递推

- 对数似然一阶导数：

$$\frac{\partial \ell_t(\theta)}{\partial \theta} = -\frac{1}{2} (\varepsilon_t^2(\theta) - 1) \frac{\partial \log \sigma_t^2(\theta)}{\partial \theta},$$

$$\varepsilon_t^2(\theta) = \frac{\varepsilon_t^2}{\sigma_t^2(\theta)},$$

$$\frac{\partial \log \sigma_t^2(\theta)}{\partial \theta} = \frac{1}{\sigma_t^2(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta}.$$

- 递推形式：

$$\frac{\partial \sigma_t^2}{\partial \omega} = 1 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \omega}, \quad \frac{\partial \sigma_t^2}{\partial \beta} = \sigma_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \beta}, \quad \frac{\partial \sigma_t^2}{\partial \alpha} = \varepsilon_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \alpha}.$$

渐近分布与“信息矩阵等价”

定理

在 *Bollerslev and Wooldridge (1992)* 的条件下,

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{J}^{-1} \mathcal{I} \mathcal{J}^{-1}),$$

其中

$$\mathcal{J} = E\left[-\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'}\right], \quad \mathcal{I} = E\left[\frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'}\right].$$

渐近分布与“信息矩阵等价”

- 若 ν_t i.i.d., 则 \mathcal{I} 与 \mathcal{J} 成比例。
- 特别地：

$$\mathcal{I} = \frac{1}{4} E[(\nu_t^2 - 1)^2] E\left(\frac{\partial \log \sigma_t^2}{\partial \theta} \frac{\partial \log \sigma_t^2}{\partial \theta'}\right),$$

$$\mathcal{J} = \frac{1}{2} E\left(\frac{\partial \log \sigma_t^2}{\partial \theta} \frac{\partial \log \sigma_t^2}{\partial \theta'}\right).$$

高斯与非高斯误差下的方差估计

- 若 $\nu_t \sim N(0, 1)$, 则

$$E[(\nu_t^2 - 1)^2] = E(\nu_t^4) - 1 = 2,$$

因此 $Avar(\hat{\theta}) = 2 \mathcal{J}^{-1}$ 。

- 若 ν_t 为 i.i.d. 非高斯:

$$Avar(\hat{\theta}) = E[(\nu_t^2 - 1)^2] \mathcal{J}^{-1},$$

可用一致估计

$$\left(\frac{1}{T} \sum_{t=1}^T (\hat{\nu}_t^2 - 1)^2 \right) \hat{\mathcal{J}}^{-1}.$$

- 若仅假设鞅差分 (MDS) 结构, 则使用 **三明治估计量**:

$$\hat{\mathcal{J}}^{-1} \hat{\mathcal{I}} \hat{\mathcal{J}}^{-1},$$

用于稳健标准误、置信区间和假设检验。

Wald 检验与显著性检验

- 一般线性约束:

$$\mathbb{H}_0: R\theta = r,$$

其中 R 为 $q \times p$ 矩阵、 r 为 $q \times 1$ 向量, $q < p$ 。

- Wald 统计量:

$$W = T(R\hat{\theta} - r)' \left(R \hat{\mathcal{J}}^{-1} \hat{\mathcal{I}} \hat{\mathcal{J}}^{-1} R' \right)^{-1} (R\hat{\theta} - r),$$

在大样本下 $W \sim \chi_q^2$ 。

- 示例: 检验“无异方差效应”, 即 $\beta = \alpha = 0$ 。

上证指数收益率的 GARCH 建模：R 举例

```
set.seed(123)

# 如在 RStudio 中，可设置工作目录为当前脚本目录
if (requireNamespace("rstudioapi", quietly = TRUE) &&
rstudioapi::isAvailable()) {
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
}

library(rugarch)
library(forecast)
library(zoo)

# 载入数据
sse_data <- read.csv("sse.csv", stringsAsFactors = FALSE)
sse_data$date <- as.Date(sse_data$date)

# 计算对数收益率
sse_data$logReturns <- c(NA, diff(log(sse_data$SSE)), lag = 1))
sse_data <- sse_data[-1,]

# 自动选择 ARMA 模型
arma_fit <- auto.arima(
sse_data$logReturns,
seasonal      = FALSE,
stepwise       = FALSE,
approximation = FALSE
)
summary(arma_fit)
```

上证指数：多种 GARCH 家族模型拟合

```

# 标准 GARCH(1,1)
spec_garch <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
  mean.model     = list(armaOrder = c(0, 0), include.mean = TRUE)
)
garch_fit <- ugarchfit(spec = spec_garch, data = sse_data$LogReturns)

# EGARCH(1,1)
spec_egarch <- ugarchspec(
  variance.model = list(model = "eGARCH", garchOrder = c(1, 1)),
  mean.model     = list(armaOrder = c(0, 0), include.mean = TRUE)
)
egarch_fit <- ugarchfit(spec = spec_egarch, data = sse_data$LogReturns)

# GJR-GARCH(1,1)
spec_gjrgarch <- ugarchspec(
  variance.model = list(model = "gjrGARCH", garchOrder = c(1, 1)),
  mean.model     = list(armaOrder = c(0, 0), include.mean = TRUE)
)
gjrgarch_fit <- ugarchfit(spec = spec_gjrgarch, data = sse_data$LogReturns)

```

门限 GARCH 与马尔可夫区制转换 GARCH: R 示例

```

# 门限 GARCH (TARCH)
tgarch_spec <- ugarchspec(
  variance.model = list(model = "fGARCH",
    submodel = "TGARCH",
    garchOrder = c(1, 1)),
  mean.model     = list(armaOrder = c(0, 0), include.mean = TRUE)
)

# 马尔可夫区制转换模型 (均值方程示意)
library(MSwM)
data_df <- data.frame(log_returns = sse_data$LogReturns)
formula <- log_returns ~ 1

# k = 2 个状态，sw 长度与系数个数匹配
model_ms_garch <- msmFit(formula, data = data_df, k = 2,
  sw = c(TRUE, TRUE))

# 汇总输出
garch_fit
egarch_fit
gjrgarch_fit
# tarch_fit (若拟合 TARCH)
model_ms_garch

```

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

为什么要先检验 ARCH 效应？

- 在建模波动率之前，应先检验是否存在波动聚类（ARCH 效应）。
- 高频金融数据：
 - 波动聚类通常很明显；
- 宏观序列：
 - 更平滑、样本小，证据可能较弱。
- 一般框架：

$$Y_t = \mu(\mathcal{F}_{t-1}, \theta_0) + \varepsilon_t,$$

先建模条件均值，再在残差上检验条件方差是否时变。

检验框架：条件同方差 vs 条件异方差

- 原假设（条件同方差）：

$$\mathbb{H}_0 : \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_0^2, \quad \exists \sigma_0^2 > 0.$$

- 备择假设（条件异方差）：

$$\mathbb{H}_A : \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) \text{ 非常数 a.s.}$$

等价地：

$$\forall \sigma^2 > 0, \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) \neq \sigma^2 \text{ a.s.}$$

- 关键前提：条件均值模型 $\mu(\mathcal{F}_{t-1}, \theta)$ 设定正确。

条件均值误设对 ARCH 检验的影响

- 若 $\mu(\mathcal{F}_{t-1}, \theta)$ 设定正确，则存在 θ_0 使

$$\mu_t = \mu(\mathcal{F}_{t-1}, \theta_0), \quad E(\varepsilon_t | \mathcal{F}_{t-1}) = 0.$$

- 若条件均值误设，则估计残差：

$$\hat{\varepsilon}_t = \varepsilon_t + [\mu_t - \mu(\mathcal{F}_{t-1}, \theta_0)],$$

其中误设项依赖于 \mathcal{F}_{t-1} 。

- 即便真实误差 ε_t 条件同方差， $\hat{\varepsilon}_t$ 也可能呈现 ARCH 效应。
- 因此：先搞定条件均值，再讨论波动率。

Engle 的 LM 检验：步骤

- Engle (1982) 的 LM 检验步骤：

① 拟合条件均值模型，得到残差：

$$\hat{\varepsilon}_t = Y_t - \mu(\mathcal{F}_{t-1}, \hat{\theta}).$$

② 进行辅助回归：

$$\hat{\varepsilon}_t^2 = \phi_0 + \sum_{j=1}^p \phi_j \hat{\varepsilon}_{t-j}^2 + u_t, \quad t = p+1, \dots, T.$$

③ 计算 TR^2 或 $(T-p)R^2$ ，其中 R^2 为辅助回归的决定系数。

LM 检验的渐近分布

- 在原假设 H_0 (无 ARCH 效应) 下, 当 $T \rightarrow \infty$:

$$TR^2 \xrightarrow{d} \chi_p^2.$$

- 参数估计不确定性对极限分布无影响:
 - 可以用 $\hat{\theta}$ 替代未知的 θ_0 ;
 - 无需调整 χ^2 自由度。
- 当 p 较大时,
 - 辅助回归对各滞后阶的权重相同;
 - 检验功效可能下降。

Box–Pierce 型混成检验

- McLeod and Li (1983) 提出的统计量：

$$\text{ML}(p) = T \sum_{j=1}^p \hat{\rho}_2^2(j),$$

其中 $\hat{\rho}_2(j)$ 为残差平方序列 $\{\hat{\varepsilon}_t^2\}$ 的样本自相关系数：

$$\hat{\rho}_2(j) = \frac{\hat{\gamma}_2(j)}{\hat{\gamma}_2(0)}.$$

- 在 H_0 下，当 $T \rightarrow \infty$ ：

$$\text{ML}(p) \xrightarrow{d} \chi_p^2.$$

- 该检验在渐近上与 Engle 的 LM 检验等价。
- 同样地，条件均值参数估计的不确定性不影响其极限分布。

LM 与 Box-Pierce 型检验：R 示例

```
library(lmtest)
library(fUnitRoots)

# 模拟一个 AR(1) 过程
set.seed(123)
n <- 100
Y <- arima.sim(model = list(ar = 0.5), n = n)

# 拟合 AR(1) 模型
model <- lm(Y[2:n] ~ Y[1:(n-1)])

# 残差
residuals <- model$residuals
residuals_squared <- residuals^2

# —— LM 检验 ——
p <- 1
auxiliary_model <- lm(
  residuals_squared[(p + 1):n] ~ residuals_squared[1:(n - p)])
)

lm_stat <- summary(auxiliary_model)$r.squared * (n - p)
lm_stat
```

Box-Pierce 型混成检验：R 示例（续）

```

# —— Box-Pierce 型混成检验 ——
sigma_hat_squared <- mean(residuals_squared)
demeaned_resid_sq <- residuals_squared - sigma_hat_squared

gamma_hat_0 <- var(demeaned_resid_sq)
p <- 1 # 阶数，可调整
gamma_hat_j <- sapply(
  1:p,
  function(j)
  mean(demeaned_resid_sq[(j + 1):n] * demeaned_resid_sq[1:(n - j)],
  na.rm = TRUE)
)

rho_hat <- gamma_hat_j / gamma_hat_0

#  $ML(p)$  统计量
ml_stat <- n * sum(rho_hat^2)
ml_stat

```

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

为什么需要多元波动率模型？

- 在金融市场中，多种资产的价格与波动率往往具有显著相关性：
 - 尤其在危机或极端事件期间，多资产可能同时出现大幅波动；
 - 资产间的联动性是风险管理与资产配置的核心问题。
- 多元波动率模型提供刻画多资产联合波动的工具：
 - 可研究波动溢出 (volatility spillover)；
 - 可度量动态相关性与传染效应 (contagion)。
- 常见的多元波动率模型：
 - BEKK 模型 (Baba–Engle–Kraft–Kroner)；
 - DCC–GARCH 模型 (Dynamic Conditional Correlation GARCH)；
 - 以及 VEC、因子–GARCH 等更高维扩展。

文献与本节内容

- Tsay (2013) 第 10 章:
 - 指数加权估计、多元 GARCH、VEC、BEKK 等;
 - 常数相关、时变相关、动态相关模型。
- Linton (2019) 第 11.14 节:
 - 更高维波动率模型与因子-波动率模型。
- 本节主要通过两个案例介绍:
 - ① 基于 BEKK-GARCH 的股指期现货波动率溢出;
 - ② 基于 DCC-GARCH 的股指联动性分析。

股指期货与“杠杆—双刃剑”效应

- 1982 年：美国堪萨斯期货交易所推出全球首张股指期货。
- 股指期货的两面性：
 - 正面：套期保值、风险对冲与价格发现；
 - 负面：高杠杆、低成本易吸引投机者，可能放大波动。
- 2015 年二季度 A 股波动案例：
 - 上证综指：5178.19 点 → 2850.71 点（8 月 26 日）；
 - 沪深 300：5380.43 点 → 2952.01 点；
 - 舆论一度将异常波动部分归因于股指期货。
- 研究问题：
 - 股指期货与现货之间是否存在显著波动率溢出效应？
 - 溢出方向与强度如何？

VAR-BEKK-GARCH 模型设定

- 样本区间：2010/04/16–2018/08/23。
- 使用 VAR-BEKK-GARCH 模型 (Engle and Kroner 1995)：

$$R_t = \mu + \sum_{i=1}^{k-1} \Gamma_i R_{t-i} + \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim N(0, H_t),$$

$$H_t = C' C + A' \varepsilon_{t-1} \varepsilon'_{t-1} A + B' H_{t-1} B.$$

- 记号说明：

$$R_t = \begin{pmatrix} r_{f,t} \\ r_{s,t} \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad H_t = \begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{pmatrix}.$$

- $r_{f,t}$: 股指期货当月合约对数收益率； $r_{s,t}$: 股指（现货）对数收益率。

VAR 阶数选择与 BEKK(1,1) 设定

- VAR 部分滞后阶数 k :
 - 依据 HQ、AIC、BIC 等信息准则选择；
 - 注意：这些准则通常基于同方差假定，仅作初步参考。
- 需要对残差向量 ε_t 进行：
 - 自相关性检验；
 - 异方差性（ARCH 效应）检验。
- 对于本案例（2 维资产）：
 - 采用二元 BEKK-GARCH(1,1) 进行估计；
 - 在保持可识别性与可估计性的前提下，参数相对精简。

二元 BEKK-GARCH(1,1) 模型

$$\begin{aligned}
 H_t = & \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix}' \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix} \\
 & + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}' \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}\varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (1) \\
 & + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}' \begin{pmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.
 \end{aligned}$$

- C 、 A 、 B 为 2×2 系数矩阵；
- BEKK 结构保证 H_t 半正定。

波动溢出的三条传递渠道

- 方差方程 (79) 展示信息和波动率在期现货市场间的传递，主要通过三条渠道：

① 滞后均方误差：

$$(\varepsilon_{1,t-1}^2, \varepsilon_{2,t-1}^2)$$

代表来自收益率冲击角度的波动溢出 (return innovation effect)。

② 滞后条件方差与协方差：

$$(h_{11,t-1}, h_{22,t-1}, h_{12,t-1})$$

自身条件方差刻画波动持续性，协方差项反映历史信息的跨市场溢出。

③ 误差交叉乘积：

$$\varepsilon_{1,t-1}\varepsilon_{2,t-1},$$

可理解为潜在的双向冲击 (bi-directional shock)。

溢出方向的参数刻画

- 当期冲击通道：
 - 从股票 → 期货的信息： a_{21}^2 ；
 - 从期货 → 股票的信息： a_{12}^2 。
- 历史信息通道：
 - 股票历史 → 期货历史的溢出： b_{21}^2 ；
 - 期货历史 → 股票历史的溢出： b_{12}^2 。
- 若 $|a_{12}| > |a_{21}|$ ，说明期货市场当期冲击对股票波动影响更强；若 $|b_{21}| > |b_{12}|$ ，说明期货市场在历史信息渠道对股票波动溢出更大。

描述性统计：沪深 300 期现货收益

表：沪深 300 股指期货（当月合约）与沪深 300 股指的描述性统计量

	均值	标准差	偏度	峰度
沪深 300 股指期货（当月合约）				
日度收益率	-1.66081×10^{-5}	0.01667454	-0.6091457	9.400874
沪深 300 股指 日度收益率	-5.34889×10^{-6}	0.01489299	-0.7382722	5.066927

数据来源：中国金融期货交易所（WIND 数据库）

BEKK-GARCH 实证：数据与 VAR 阶数选择

```
# 清空工作空间
rm(list = ls())

# 设置工作目录（如在 RStudio 中）
if (requireNamespace("rstudioapi", quietly = TRUE) &&
rstudioapi::isAvailable()) {
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
}

# 加载所需的库
library(readxl)    # 读取 Excel
library(forecast)   # 时间序列分析
library(vars)       # VAR 模型
library(mgarchBEKK) # BEKK-GARCH 模型

# 读取数据
dat <- read_excel("沪深300股指期现货收盘价.xlsx")
dat <- na.omit(dat)

# 提取并取对数
date <- dat$日期
LnS <- log(dat$"沪深300股指收盘价")
LnF <- log(dat$"沪深300股指期货当月收盘价")

n <- length(LnS)

# 计算收益率
date.ret <- date[-1]
RS <- LnS[-1] - LnS[-n]
RF <- LnF[-1] - LnF[-n]
```

BEKK-GARCH 实证：模型拟合与诊断

```
# 合并收益率  
dat.var <- cbind(RF, RS)  
  
# VAR 阶数选择  
VARselect(dat.var, lag.max = 12, type = "const")  
  
# 拟合 VAR 模型 (例如 p = 3)  
var.2c <- VAR(dat.var, p = 3, type = "const")  
  
# 提取残差  
ut <- na.omit(resid(var.2c))  
  
# 拟合 BEKK-GARCH 模型  
estimated <- BEKK(ut)  
  
# 模型诊断  
diagnoseBEKK(estimated)
```

BEKK-GARCH 估计结果示例（节选）

```
Number of estimated series : 4060
Length of estimated series : 2030
BEKK order : 1 1
```

C estimates:

```
[,1]      [,2]
[1,] -0.002390247 -0.002540273
[2,]  0.000000000 -0.002144994
```

ARCH estimates:

```
[,1]      [,2]
[1,] -0.09025784  0.6614547
[2,] -0.18887369 -0.9614542
```

GARCH estimates:

```
[,1]      [,2]
[1,]  0.5162522 -0.08867566
[2,] -1.5260502 -0.82329481
```

- C estimates $\Rightarrow \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix}$;
- ARCH estimates $\Rightarrow \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$;
- GARCH estimates $\Rightarrow \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$.

BEKK 结果的经济含义

- 若估计结果满足

$$|a_{12}| > |a_{21}|,$$

则期货市场当期冲击对股票市场波动的影响更强。

- 若

$$|b_{21}| > |b_{12}|,$$

则在历史信息渠道上，期货对股票的波动溢出效应更大。

- 样本期（2010/04/16–2018/08/23）内：

- 中国金融期货交易所多次调整交易规则（见表 ??）；
- 建议按制度变动对样本进行子样本划分，检验结果稳健性。

DCC-GARCH 模型：基本思想

- 单变量 GARCH：只刻画单个资产的条件方差。
- DCC-GARCH：在此基础上引入动态条件相关性：
 - 每个资产各自一个 GARCH(1,1) 方程；
 - 标准化残差用于构造时变相关矩阵；
 - 条件相关矩阵随时间动态更新。
- 优点：
 - 维度较高时仍相对简洁；
 - 能刻画市场联动与风险传染的动态演化。

二元 DCC-GARCH 模型形式 (1)

- 设 ARMA 调整后的收益为 \tilde{r}_t :

$$\tilde{r}_t = H_t^{1/2} \epsilon_t, \quad H_t = D_t S_t D_t,$$

$$h_{11,t} = c_{11} + a_{11,1} \tilde{r}_{1,t-1}^2 + g_{11,1} h_{11,t-1},$$

$$h_{22,t} = c_{22} + a_{22,1} \tilde{r}_{2,t-1}^2 + g_{22,1} h_{22,t-1}.$$

- 其中：

- H_t : 条件协方差矩阵；
- $D_t = \text{diag}(\sqrt{h_{11,t}}, \sqrt{h_{22,t}})$ ；
- S_t : 时变条件相关矩阵。

二元 DCC-GARCH 模型形式 (2)

- 条件相关矩阵：

$$S_t = \begin{pmatrix} 1 & s_{12,t} \\ s_{12,t} & 1 \end{pmatrix}, \quad s_{12,t} = \frac{q_{12,t}}{\sqrt{q_{11,t} q_{22,t}}}.$$

- Q_t 的 DCC 更新：

$$q_{12,t} = (1 - \alpha - \beta)\bar{q}_{12} + \alpha \frac{\tilde{r}_{1,t-1}\tilde{r}_{2,t-1}}{\sqrt{h_{11,t-1} h_{22,t-1}}} + \beta q_{12,t-1},$$

$$q_{11,t} = (1 - \alpha - \beta)\bar{q}_{11} + \alpha \left(\frac{\tilde{r}_{1,t-1}}{\sqrt{h_{11,t-1}}} \right)^2 + \beta q_{11,t-1},$$

$$q_{22,t} = (1 - \alpha - \beta)\bar{q}_{22} + \alpha \left(\frac{\tilde{r}_{2,t-1}}{\sqrt{h_{22,t-1}}} \right)^2 + \beta q_{22,t-1}.$$

DCC 参数的经济含义

- 模型参数向量：

$$\theta = (\alpha, \beta, c_{11}, c_{22}, a_{11,1}, a_{22,1}, g_{11,1}, g_{22,1}).$$

- DCC 部分：

- α : 条件相关性对短期冲击的即时响应；
- β : 相关性的持久性（长期衰减速度）。

- 单变量 GARCH 部分：

- c_{ii} : 基线方差（截距），即无新信息时的平均波动；
- $a_{ii,1}$: ARCH 项，滞后平方创新对当前方差的影响；
- $g_{ii,1}$: GARCH 项，滞后条件方差对当前方差的影响。

- 参数越大，波动聚集效应越明显，历史波动对当前市场状态的影响越强。

DCC-GARCH 实证：数据与研究设计

- 资产：道琼斯工业平均指数（DJIA）与标普/TSX 综合指数（S&P/TSX Composite）。
- 样本区间：2012/01/01–2020/12/31。
- Hong et al. (2024) 使用该数据对 DCC-GARCH 参数进行结构性变化检验。
- 实证步骤：
 - ① 从 CSV 读入价格数据，剔除周末；
 - ② 使用 price2return 函数计算对数收益率并插补缺失值；
 - ③ 对各收益率拟合 ARIMA 模型，提取残差；
 - ④ 基于残差拟合 DCC-GARCH 模型；
 - ⑤ 提取动态条件相关系数并作图。

DCC-GARCH 实证：数据准备与收益率

```
# 清除所有变量
rm(list = ls())

# 设置工作目录
if (requireNamespace("rstudioapi", quietly = TRUE) &&
rstudioapi::isAvailable()) {
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
}

# 载入必要库
library(rmgarch) # 多变量 GARCH
library(ggplot2) # 绘图
library(zoo)      # na_interpolation 所需

# 收益率函数
price2return <- function(p){
N <- length(p)
return.series <- (log(p[2:N]) - log(p[1:(N-1)])) * 100
return.series <- na_interpolation(return.series)
return(return.series)
}

# 读取数据并剔除周末
temp.dat <- read.csv("stock_indices.csv")
dat <- temp.dat[setdiff(1:nrow(temp.dat),
which(temp.dat$weekdays %in% c("Saturday", "Sunday"))), ]

Dow_Jones <- dat$Dow_Jones_Index_Industrial_Average
SP       <- dat$S.P.TSX_Composite_Index

# 计算收益率
```

DCC-GARCH 实证：拟合模型与提取相关性

```
# 合并收益率
return_series <- cbind(Dow_Jones.return, SP.return)

# ARIMA 拟合与残差提取
arima_dj <- auto.arima(Dow_Jones.return)
arima_sp <- auto.arima(SP.return)

residuals_dj <- residuals(arima_dj)
residuals_sp <- residuals(arima_sp)

# DCC-GARCH 数据
ut <- cbind(residuals_dj, residuals_sp)

# GARCH(1,1) 规格
garch11.spec <- ugarchspec(
mean.model = list(armaOrder = c(0, 0), include.mean = FALSE),
variance.model = list(garchOrder = c(1, 1), model = "sGARCH"),
distribution.model = "norm"
)

# DCC-GARCH 规格
dcc.garch11.spec <- dccspec(
uspec = multispec(replicate(2, garch11.spec)),
dccOrder = c(1, 1),
distribution = "mvnrm"
)

# 拟合 DCC-GARCH 模型
dcc.fit <- dccfit(dcc.garch11.spec, data = ut)
dcc.fit
```

DCC-GARCH 实证：动态相关性的提取与绘图

```
# 条件相关矩阵
dynamic_correlation <- rcor(dcc.fit)

# 取非对角元素，DJIA 与 S&P/TSX 的相关系数
off_diagonal_correlation <- dynamic_correlation[2, 1, ]

# 时间序列索引（若有日期信息可替换）
time_index <- as.Date(dimnames(dynamic_correlation)[[3]])

dynamic_corr_df <- data.frame(
  time      = time_index,
  correlation = off_diagonal_correlation
)

# 绘图
ggplot(dynamic_corr_df, aes(x = time, y = correlation)) +
  geom_line() +
  labs(title = "Dynamic Conditional Correlation (Off-Diagonal)",
       x = "Time", y = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

DCC-GARCH 拟合结果示例（节选）

```
*
* DCC GARCH Fit
*
```

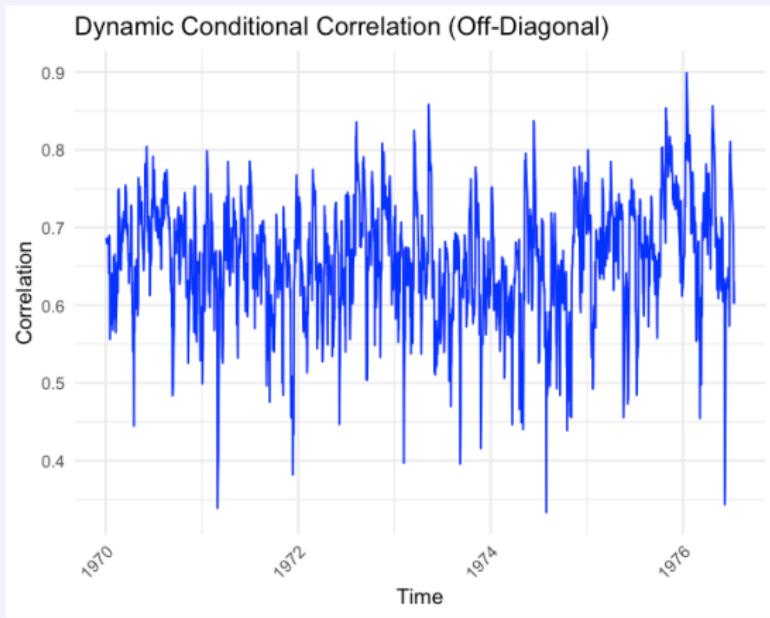
Distribution : mvnorm
 Model : DCC(1,1)
 No. Parameters : 9
 [VAR GARCH DCC UncQ] : [0+6+2+1]
 No. Series : 2
 No. Obs. : 2385

Optimal Parameters

	Estimate	Std. Error	t value	Pr(> t)
[residuals_dj].omega	0.041274	0.008534	4.8364	0.000001
[residuals_dj].alpha1	0.222284	0.032933	6.7496	0.000000
[residuals_dj].beta1	0.734947	0.030983	23.7210	0.000000
[residuals_sp].omega	0.020112	0.005474	3.6738	0.000239
[residuals_sp].alpha1	0.188567	0.035891	5.2539	0.000000
[residuals_sp].beta1	0.785539	0.033671	23.3301	0.000000
[Joint]dcca1	0.072732	0.014655	4.9630	0.000001
[Joint]dccb1	0.793087	0.044881	17.6708	0.000000

- $[Joint]dcca1 = \alpha$ 、 $[Joint]dccb1 = \beta$: 动态相关的短期冲击效应与持久性；
- 各资产的 GARCH(1,1) 参数均显著，说明存在明显的波动聚集。

DCC-GARCH 的动态相关图



图：DCC-GARCH 模型估计的动态条件相关性

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

本章小结

- 本章聚焦于金融时间序列的波动率建模：
 - 介绍了 ARCH 模型及其平稳性条件；
 - 扩展到 GARCH 模型及其弱平稳与强平稳性质；
 - 讨论了非对称波动率模型（EGARCH、GJR、TGARCH、MS-GARCH）；
 - 分析了具有“单位根”特性的 IGARCH 模型及 RiskMetrics。
- 模型估计与检验：
 - 给出了 ARCH/GARCH 的（拟）极大似然估计与渐近性质；
 - 介绍了 LM 检验与 Box-Pierce 型混成检验等 ARCH 效应诊断工具。
- 多元扩展与应用：
 - 引入多元波动率模型（BEKK、DCC-GARCH）；
 - 通过股指期现货与国际股指联动两个案例，展示了多元 GARCH 在波动溢出与动态相关分析中的具体应用。
- 通过本章，读者应能将单变量与多变量波动率模型从理论理解扩展到实务应用。

目录

- 1 白噪声、i.i.d. 与鞅差分
- 2 ARCH 模型
- 3 GARCH 模型
- 4 非对称波动率模型
- 5 IGARCH 模型
- 6 ARCH、GARCH 模型的估计
- 7 ARCH、GARCH 模型的诊断检验
- 8 多元波动率模型
 - 案例：基于 BEKK–GARCH 的股指期现货波动率研究
 - 案例：基于 DCC–GARCH 的股指联动性分析
- 9 章节总结
- 10 参考文献

参考文献 I

-  Bollerslev, Tim (1986). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3, pp. 307–327.
-  Bollerslev, Tim and Jeffrey M Wooldridge (1992). "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances". In: *Econometric reviews* 11.2, pp. 143–172.
-  Engle, Robert F (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica*, pp. 987–1007.
-  Engle, Robert F and Kenneth F Kroner (1995). "Multivariate simultaneous generalized ARCH". In: *Econometric Theory* 11.1, pp. 122–150.
-  Glosten, Lawrence R, Ravi Jagannathan, and David E Runkle (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks". In: *The journal of finance* 48.5, pp. 1779–1801.

参考文献 II

-  Hamilton, James D and Raul Susmel (1994). "Autoregressive conditional heteroskedasticity and changes in regime". In: *Journal of econometrics* 64.1-2, pp. 307–333.
-  Hong, Yongmiao et al. (2024). "Kolmogorov–Smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach". In: *Journal of Econometrics* 238.2, p. 105603.
-  Linton, Oliver (2019). *Financial econometrics*. Cambridge University Press.
-  McLeod, Allan I and William K Li (1983). "Diagnostic checking ARMA time series models using squared-residual autocorrelations". In: *Journal of Time Series Analysis* 4.4, pp. 269–273.
-  Morgan, JP et al. (1997). "Creditmetrics-technical document". In: *JP Morgan*, New York 1, pp. 102–127.
-  Nelson, Daniel B (1991). "Conditional heteroskedasticity in asset returns: A new approach". In: *Econometrica* 59.2, pp. 347–370.

参考文献 III

-  Tsay, Ruey S. (2013). *Analysis of Financial Time Series*. 3rd. Hoboken, NJ: Wiley. ISBN: 978-1-118-47512-3.
-  Zakoian, Jean-Michel (1994). “Threshold heteroskedastic models”. In: *Journal of Economic Dynamics and Control* 18.5, pp. 931–955.

第 6 章收益可预测性与有效市场假说

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

目录

1 收益可预测性与有效市场假说

2 有效市场假说

3 弱式有效市场的检验

4 半强式与强式有效性检验

5 异常收益与信息集

6 随机游走价格模型

7 对 EMH 的假设检验

- 序列相关性检验
- 方差比检验
- 依照 AR 模型对 EMH 进行检验
- 对 rw_2 与 rw_3 进行假设检验

8 案例：贵州茅台的弱式有效性检验

9 章节总结

10 参考文献

收益可预测性与有效市场假说：动机

- 本章核心：讨论资产收益可预测性与有效市场假说（EMH）的关系。
- EMH 的核心主张：所有相关信息已反映在价格中，无法持续获得超额回报。
- 研究收益的可预测性：
 - 是对 EMH 的直接检验；
 - 有助于理解市场动态与投资者行为；
 - 揭示市场何时、因何而低效（异常与模式）。
- 实务意义：
 - 指导投资组合管理与投资策略设计；
 - 提升市场透明度与资源配置效率；
 - 为监管政策制定提供依据。

目录

1 收益可预测性与有效市场假说

2 有效市场假说

3 弱式有效市场的检验

4 半强式与强式有效性检验

5 异常收益与信息集

6 随机游走价格模型

7 对 EMH 的假设检验

- 序列相关性检验
- 方差比检验
- 依照 AR 模型对 EMH 进行检验
- 对 rw_2 与 rw_3 进行假设检验

8 案例：贵州茅台的弱式有效性检验

9 章节总结

10 参考文献

EMH 的提出与发展：Samuelson 与 Fama

- 有效市场假说（EMH）是金融经济学中的核心概念。
- 历史源头：
 - 1960s: 保罗·A·萨缪尔森 (Paul A. Samuelson)
 - 尤金·F·法玛 (Eugene F. Fama)
- Samuelson (1965):
 - 文章: *Proof that Properly Anticipated Prices Fluctuate Randomly.*
 - 观点: 若价格变动恰当反映所有信息与预期, 则价格变动应不可预测 (随机)。
 - 研究背景: 可储存商品的跨期定价 (考虑储存成本与损耗)。

EMH 的提出与发展：Fama 的路径

- Fama 对以下问题有浓厚兴趣：
 - 股票价格的统计特性；
 - 技术分析 vs 基本面分析。
- 早期实证研究：
 - 较早使用计算机进行金融实证；
 - 首次系统使用“有效市场”一词；
 - 提出“价格完全反映所有可用信息”。
- 贡献：
 - 将 EMH 与可操作的信息集合关联起来；
 - 推动事件研究法等一系列方法论；
 - 发现大量价格行为的实证规律与异常。

EMH 的直觉与后续发展

- 信息效率的“反直觉”特征：
 - 市场越高效，价格变动越接近随机；
 - “最有效”的市场表现为不可预测的随机波动。
- 原因：众多投资者争夺信息优势
 - 任意可利用信息迅速被反映到价格；
 - 初始套利空间迅速被消除。
- 1970s 之后的“新古典 EMH”：
 - 引入风险厌恶投资者与资产定价模型；
 - 价格变动不可预测，因为价格已充分反映所有可得信息。

EMH 的扩展与争论

- EMH 的扩展方向：
 - 非交易资产（如人力资本）；
 - 状态依赖偏好与投资者异质性；
 - 信息不对称与交易成本；
 - 多种资产类别：股票、债券、货币、商品等。
- 总体结论：在任一时点，证券价格已充分反映所考虑的信息集。
- 更系统的综述：Lo (2008), Lim and Brooks (2011), and Sewell (2011)。

EMH 的三种形式

- **弱式有效：**

- 价格已吸收全部历史价格信息；
- 基于历史价格的技术分析无法持续获得超额收益。

- **半强式有效：**

- 价格迅速反映所有公开信息（财报、公告、宏观数据等）；
- 公募信息难以用于获取系统性超额回报。

- **强式有效：**

- 价格反映所有信息（包括内部信息）；
- 即便内部人士也难以持续获得超额收益。

有效市场假说的检验层次

市场假说 类型	特点	检验方法	典型结论
弱式有效	历史价格信息已反映在当前股价中	游程检验、方差比检验，自相关检验等	历史信息不能预测未来收益，难以获得超额收益
半强式有效	所有公开信息已反映在股价中	事件研究法（财报、分红、并购等）	公开信息被迅速反映于价格，难以获得超额收益
强式有效	所有信息（含非公开信息）已反映在股价中	内幕信息持有者绩效分析	即便掌握内部信息亦难以持续获得超额收益

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

弱式 EMH 的实证证据（一）

- 多数实证研究支持弱式 EMH:
 - 标普 500 指数 (1980–1984) vs 抛硬币模型；
 - 基于历史趋势的技术交易策略**不能持续获利** (Brealey, Myers, and Allen 2014)。
- Solnik (1973):
 - 9 个国家股票市场的日收益率；
 - 序列相关性接近零，支持弱式 EMH。
- 本书中上证指数案例（第 ?? 节）:
 - ARMA 建模下收益率无显著自相关；
 - 历史信息已反映在当前价格中。

弱式 EMH 的实证证据（二）

- Barber and Odean (2000):
 - 比较个人投资者与标普 500 指数基金的长期表现；
 - 指数基金净回报稳定，整体高于平均个人投资者；
 - 频繁交易降低净收益，长期分散投资更稳健。
- 投资含义：
 - 在弱式有效市场下，基于历史价格的短期交易策略难以持续超额获利；
 - 更应重视长期投资、分散配置与成本控制。

弱式有效性检验方法

- 检验收益序列的“随机性”：
 - 自相关函数 (ACF)、偏自相关函数 (PACF);
 - 游程检验 (runs test)、方差比检验 (variance ratio test)。
- 检验技术交易规则的获利性：
 - 均线策略、突破策略等;
 - 若不能持续获得风险调整后的超额收益，则支持弱式 EMH。
- 常用统计检验：
 - 非参数：游程检验、方差比检验;
 - 参数：ADF 单位根检验、Phillips–Perron 检验等。

技术分析与弱式 EMH 的矛盾

不难看出，有效市场假说（EMH）与部分著名投资策略（尤其是技术分析策略）存在一定的矛盾：弱式 EMH 认为技术分析难以持续获得超额收益。下面列出技术分析中 10 种常见图表形态：

- 双顶形态 (Double Tops)
- 双底形态 (Double Bottoms)
- 趋势线与突破 (Trendline & Breakout)
- 头肩形态 (Head and Shoulders)
- 三角形 (Triangle)
- 楔形 (Wedge)
- 三角旗形 (Pennant)
- 旗形 (Flag)
- 圆弧底 (Rounding Bottom)
- 杯柄形 (Cup with Handle)

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

半强式有效性：公开信息的价格反应

- 半强式 EMH：价格迅速、无偏地反映所有公开信息。
- 典型实证研究：
 - Ball and Brown (1968)：财报发布后股价迅速调整；
 - Foster (1973)：股价对盈利公告反应迅速；
 - Keown and Pinkerton (1981)：股利公告和并购事件对股价影响即时显现；
 - Asquith and Mullins Jr (1986)：市场对股票分拆等事件反应有效；
 - Hayn (1995)：盈余反应系数 (ERC) 衡量财报信息的价格影响。
- 总体结论：在许多场景下，公开信息会被较快地反映在价格中。

强式有效性与资产管理表现

- 强式 EMH：价格反映所有信息（包括内部信息）。
- 资产管理表现研究：
 - Jensen (1969)：资本管理者难以长期战胜市场；
 - 支持“在给定信息集下，获取超额收益非常困难”的观点。
- 内幕交易研究：
 - Meulbroek (1992)：非法内幕交易可获得正的累积异常收益 (CAR)；
 - 与 Manne (1966)、Carlton and Fischel (1982) 的观点呼应：
 - 内幕交易可能加快价格发现；
 - 提高资本配置效率，但具有严重的公平性与监管问题。

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

异常收益 (Abnormal Returns) 的定义

- 核心概念：异常收益 = 实际回报 – 正常回报。
- 正常回报 R_t^* ：
 - 由某一资产定价模型给出；
 - $R_t^* = R_{ft} + \pi_t$, R_{ft} 为无风险利率, π_t 为风险溢价。
- 在 EMH 框架下：

$$\mathbb{E}(R_T | \mathcal{F}_t) = R_t^*,$$

- 若观察到长期显著偏离 R_t^* 的收益，则被视为“异常收益”，挑战 EMH。

信息集的层次结构

- 信息集的三层结构：
 - \mathcal{H}_t : 价格历史;
 - \mathcal{F}_t : 所有公开信息 (含 \mathcal{H}_t);
 - \mathcal{G}_t : 所有信息 (公开 + 非公开), 含 \mathcal{F}_t 。
- 对任意 $\mathcal{F}'_t \subseteq \mathcal{F}_t$, 有迭代期望:

$$\mathbb{E}(R_T - R_t^* \mid \mathcal{F}'_t) = 0.$$

- 信息集扩展 \Rightarrow 条件方差减小:

$$\text{Var}(R_{t+j} \mid \mathcal{F}'_t) \geq \text{Var}(R_{t+j} \mid \mathcal{F}_t), \quad (\mathcal{F}'_t \subseteq \mathcal{F}_t).$$

- 意味着: 更多信息 \Rightarrow 预测更精确、风险评估更充分。

EMH 检验的困难与联合假设问题

- 实证检验 EMH 时，通常同时假定：
 - 某一特定资产定价模型（CAPM、FF 三因子等）；
 - 市场在给定信息集下有效。
- 若拒绝“无法获得超额收益”的原假设：
 - 原因可能是市场无效；
 - 也可能是资产定价模型设定错误；
 - 这就是“联合假设问题（joint hypothesis problem）”。
- 结论：
 - 我们很难“彻底证伪”弱式 EMH；
 - 对半强式、强式 EMH 的检验更受信息集定义与数据可得性限制。

信息效率的形式化定义

设 \mathcal{H}_t 为时间 t 的市场价格历史, \mathcal{F}_t 为包含 \mathcal{H}_t 的公共信息集, \mathcal{G}_t 为包含 \mathcal{F}_t 的全部信息集。若对所有 $t \geq 0$ 与 $x \in \mathbb{R}$,

$$\Pr(P_T \leq x | \mathcal{F}_t) = \Pr(P_T \leq x | \mathcal{H}_t),$$

则说明市场已完全吸收公共信息集 $\{\mathcal{F}_t\}$ 。

- 含义：在给定价格历史的情况下，更多公共信息不再改变未来价格分布。
- 在半强式/强式有效市场中，价格历史已足以刻画未来价格分布。

更弱的可检验条件

- 直接检验分布等式非常困难，通常转而检验条件期望：

$$\mathbb{E}(P_T | \mathcal{F}_t) = \mathbb{E}(P_T | \mathcal{H}_t),$$

$$\mathbb{E}(R_T | \mathcal{F}_t) = \mathbb{E}(R_T | \mathcal{H}_t).$$

- 由迭代期望：

$$\mathbb{E}(R_T - \mathbb{E}(R_T | \mathcal{H}_t) | \mathcal{F}_t) = 0.$$

- 为可操作起见，通常进一步假设：

- 期望收益率、无风险利率与风险溢价在检验窗口内近似恒定；
- 或由特定资产定价模型刻画。

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型**
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

随机游走与弱式 EMH

- 由于强式信息难以观测，实证研究多聚焦：
 - 弱式与半强式有效市场。
- 半强式检验往往需要财报、公告等复杂信息处理，本节主要：
 - 通过随机游走模型检验弱式有效性。
- 弱式 EMH 的经典刻画：**价格服从随机游走**。

无漂移与有漂移的随机游走

- 基本随机游走模型：

$$P_t = P_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots$$

- 有漂移的随机游走：

$$P_t = \mu + P_{t-1} + \varepsilon_t,$$

$$P_t = P_0 + t\mu + \sum_{s=1}^t \varepsilon_s.$$

- 金融意义：

- P_t : 价格或对数价格；
- 漂移 μ : 长期“正常回报”的来源；
- 高频数据中常假设 $\mu \approx 0$, 中低频数据则需要为 μ_t 建模。

风险与预期收益的关系

- 一般认为预期收益是风险的函数：

$$\mu_t = h(\text{Var}(r_t | \mathcal{F}_{t-1})),$$

其中 $h(\cdot)$ 单调递增。

- 若 $\text{Var}(r_t | \mathcal{F}_{t-1})$ 随时间变化，则 $E(r_t | \mathcal{F}_{t-1})$ 也应是时变的。
- 日度/周度数据：
 - 常近似视 $\mu_t \approx \mu$ 为常数。
- 季度/年度数据：
 - 通常需要显式建模 μ_t 的动态。

创新项 ε_t 的假设: rw1–rw3

为刻画随机游走的精确含义, 需要对创新项 ε_t 进行约束 (Campbell, Lo, and MacKinlay 1997):

- **rw1:**

- ε_t 独立同分布 (i.i.d.), 且 $E(\varepsilon_t) = 0$;
- 最强假设: 排除一切时间依赖与结构变化。

- **rw2:**

- ε_t 在时间上独立, $E(\varepsilon_t) = 0$;
- 允许分布随时间变化, 较 rw1 更灵活。

- **rw3:**

- 对所有 k, t , $E(\varepsilon_t) = 0$, 且

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0;$$

- 只要求不相关性, 依赖结构可以更一般。

rw1–rw3 的层次结构与应用

- 强弱关系：

$$\text{rw1} \Rightarrow \text{rw2} \Rightarrow \text{rw3},$$

反之不成立。

- 实务选择：

- 高频数据（稳定市场结构）：可考虑 rw1 或 rw2；
- 低频数据（结构变化更明显）：常采用 rw3。

- 研究者需在模型简洁性与现实刻画之间做权衡。

回报是按交易时间还是日历时间生成？

在金融分析中，理解回报是按交易时间还是按日历时间生成非常重要，因为这会直接影响如何分析和解释数据。

- **交易时间假设：**

- 回报在市场开放期间产生；
- 使用日回报进行分析较为直接；
- 假定每天存在交易、间隔固定。

- **日历时间假设：**

- 即使非交易日也会产生经济回报；
- 观察到的“交易日回报”可能是多个日历日的累积；
- 需对周末、节假日等非交易日进行调整。

交易时间 vs 日历时间：一个简单刻画

- 在 rw1 假设下，若 $E(r_t) = \mu$, $\text{Var}(r_t) = \sigma^2$, 则观测到的收益率 r_t^O 满足：

$$E(r_t^O) = \mu D_t, \quad \text{Var}(r_t^O) = \sigma^2 D_t,$$

其中：

- D_t 为上一次交易日至当前观察日前的日历天数；
- 如：周二-周五 $D_t = 1$, 周一 $D_t = 3$ 。
- 本书中，如无特别说明，**默认数据按交易时间生成**。

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

序列相关性检验的思路

- 若市场有效，价格变动应类似随机游走：
 - 价格变动不可预测；
 - 价格变动（收益）之间不应存在显著线性关系。
- 等价表述：价格增量（或收益率）序列的自相关系数应接近 0。
- 若某个滞后期的自相关显著不为零：
 - 说明历史信息对未来价格变动具有预测力；
 - 与 EMH 中的信息效率观点相冲突。

自协方差与自相关函数

- 考虑价格增量 $Y_t = P_t - P_{t-1}$ 。
- 自协方差函数：

$$\gamma(j) = \text{Cov}(Y_t, Y_{t-j}) = E[(Y_t - \bar{Y})(Y_{t-j} - \bar{Y}_{t-j})].$$

- 自相关函数：

$$\rho(j) = \frac{\gamma(j)}{\gamma(0)}, \quad j = 0, 1, 2, \dots$$

- 若序列非平稳，可将自协方差定义为

$$\gamma(j) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=j+1}^T \text{Cov}(Y_t, Y_{t-j}),$$

若该极限存在。

- $\rho(j) \in [-1, 1]$ ，在尺度变换（如日度 \rightarrow 年度）下保持不变。

最佳线性预测与样本自相关

- 在滞后 j 期信息下的最佳线性预测：

$$\mathbb{P}(Y_t | Y_{t-j}) = \mu + \beta_j(Y_{t-j} - \mu),$$

其中 $\mu = E(Y_t)$ 。

- 实证中使用样本自协方差与样本自相关：

$$\hat{\gamma}(j) = \frac{1}{T} \sum_{t=j+1}^T (Y_t - \bar{Y})(Y_{t-j} - \bar{Y}), \quad \bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t,$$

$$\hat{\rho}(j) = \frac{\sum_{t=j+1}^T (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})}{\sqrt{\sum_{t=j+1}^T (Y_t - \bar{Y})^2 \sum_{t=j+1}^T (Y_{t-j} - \bar{Y})^2}}.$$

- 在 rw1 假设下 (i.i.d. 且方差有限)， $\hat{\rho}(j)$ 的渐近分布可求。

样本自相关的渐近分布

定理

设 Y_t 为具有有限方差的独立同分布随机变量。对任意固定的 p , 当 $T \rightarrow \infty$ 时,

$$\sqrt{T}(\hat{\rho}(1), \dots, \hat{\rho}(p))' \implies N(0, I_p),$$

即

$$\Pr(\sqrt{T}\hat{\rho}(1) \leq x_1, \dots, \sqrt{T}\hat{\rho}(p) \leq x_p) \rightarrow \prod_{j=1}^p \Phi(x_j),$$

其中 $Z = (Z_1, \dots, Z_p) \sim N(0, I_p)$, Φ 为标准正态分布的 CDF。

- 推论：对任意固定 k , 有

$$\sqrt{T}\hat{\rho}(k) \implies N(0, 1).$$

Bartlett 区间与 Box–Pierce 统计量

- 在 rw1 下检验 $H_0 : \rho(k) = 0$ 。
- Bartlett 区间（置信水平 $1 - \alpha$ ）：

$$\left[-\frac{z_{\alpha/2}}{\sqrt{T}}, \frac{z_{\alpha/2}}{\sqrt{T}} \right],$$

若 $\hat{\rho}(k)$ 落在区间外，则拒绝 H_0 。

- 联合检验 (Box–Pierce)：

$$Q = T \sum_{j=1}^p \hat{\rho}(j)^2$$

在 rw1 下， $Q \stackrel{d}{\approx} \chi_p^2$ 。

- 检验：

$$H_0 : \rho(1) = \dots = \rho(p) = 0, \quad H_A : \exists j, \rho(j) \neq 0.$$

小样本修正与 Ljung–Box 统计量

- 小样本下, $\hat{\rho}(j)$ 可能偏差较大:
 - 倾向于错误拒绝或错误接受 H_0 。
- 修正一: 用 $T-j$ 替代 T 计算自协方差:

$$\tilde{\gamma}(j) = \frac{1}{T-j} \sum_{t=j+1}^T (Y_t - \bar{Y}_j)(Y_{t-j} - \bar{Y}_j),$$

\bar{Y}_j 为 $t = j+1, \dots, T$ 的均值。

- 修正二: 偏差校正样本自相关

$$\hat{\rho}^{bc}(j) = \hat{\rho}(j) + \frac{T-j}{(T-1)^2} (1 - \hat{\rho}(j)^2).$$

- 修正三: Ljung–Box 统计量

$$Q = T(T+2) \sum_{j=1}^P \frac{\hat{\rho}(j)^2}{T-j},$$

小样本表现通常优于原始 Box–Pierce 统计量。

方差比检验的基本思想

- 方差比 (Variance Ratio) 用于检验收益是否符合随机游走。
- 最早由 Poterba and Summers (1988) 与 Lo and MacKinlay (1988) 引入。
- 设 r_t 为对数收益率, 平稳且方差有界, 定义

$$r_t(p) = r_{t+1} + \cdots + r_{t+p}, \quad p = 2, 3, \dots$$

- 方差比:

$$VR(p) = \frac{\text{Var}(r_t(p))}{p \cdot \text{Var}(r_t)}.$$

- 若各期回报独立 (随机游走), 应有 $VR(p) = 1$ 。

方差比与自相关结构

- 若序列平稳但存在自相关,

$$VR(p) = 1 + 2 \sum_{j=1}^{p-1} \left(1 - \frac{j}{p}\right) \rho(j).$$

- 含义:

- $VR(p)$ 是前 $p - 1$ 阶自相关的加权平均;
- 保留了相关性的符号信息:
 - 若 $\rho(j) > 0$, 则 $VR(p) > 1$;
 - 若 $\rho(j) < 0$, 则 $VR(p) < 1$ 。

- 对 EMH 的含义:

- 若 $VR(p)$ 显著偏离 1, 表明存在可预测结构;
- 有违弱式 EMH 的随机游走设定。

实证计算：高频与低频收益

- 给定 $np + 1$ 个对数价格观测 p_0, \dots, p_{np} :

- 高频收益: r_1, \dots, r_T , 其中 $T = np$;
- 低频收益: $r_0(p), r_p(p), \dots, r_{(n-1)p}(p)$ 。

- 高频收益方差估计:

$$\hat{\sigma}_H^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \hat{\mu})^2, \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t.$$

- 非重叠低频方差估计:

$$\hat{\sigma}_L^2(p) = \frac{1}{n} \sum_{t=0}^{n-1} (r_t(p) - p\hat{\mu})^2.$$

- 方差比估计量:

$$\widehat{VR}(p) = \frac{\hat{\sigma}_L^2(p)}{p \cdot \hat{\sigma}_H^2}.$$

方差比的渐近分布 (rw1)

- 在 rw1 假设下，且 p 固定，有

$$\sqrt{T}(\widehat{VR}(p) - 1) \implies N(0, 2(p-1)).$$

- 定义检验统计量：

$$S = \frac{\sqrt{T}(\widehat{VR}(p) - 1)}{\sqrt{2(p-1)}}.$$

- 拒绝域：

$|S| > z_{\alpha/2} \Rightarrow$ 拒绝 EMH (随机游走) 原假设。

- 实务细节：

- 上式基于**非重叠低频回报**；
- 起始日可不同 (周一到周一、周二到周二等)。

重叠回报与更高效率的方差比

- 定义重叠低频收益序列 $\{r_0(p), \dots, r_{T-p}(p)\}$ 。
- 重叠低频方差估计：

$$\hat{\sigma}_{\text{LO}}^2(p) = \frac{1}{T-p+1} \sum_{t=0}^{T-p} (r_t(p) - p \hat{\mu})^2.$$

- 对应的渐近方差为

$$\omega_1(p) = \frac{4}{6p} (2p-1)(p-1), \quad \omega_1(p) < 2(p-1).$$

- 含义：
 - 使用重叠回报 \Rightarrow 渐近方差更小；
 - 相当于“有效样本量”更大。

基于自相关的方差比估计

- 方差比的另一种估计：

$$\widetilde{VR}(p) = 1 + 2 \sum_{j=1}^{p-1} \left(1 - \frac{j}{p}\right) \hat{\rho}(j),$$

其中 $\hat{\rho}(j)$ 用高频数据估计。

基于自相关的方差比定理

定理

在 $rw1$ 条件下,

$$\sqrt{T}(\widetilde{VR}(p) - 1) \xrightarrow{} N(0, \omega_1(p)),$$

其中 $\omega_1(p) = \frac{4}{6p}(2p - 1)(p - 1)$ 。

- 当存在异方差 ($rw3$) 时, 可对 $\omega_1(p)$ 进行修正;
- 实务中常使用 wild bootstrap 来估计长期方差 (Choi 1999)。

Bootstrap 与 Wild Bootstrap 简介

Bootstrap 是一种非参数统计推断方法，通过重抽样构造统计量的近似抽样分布，用于估计均值、标准误、置信区间等，尤其适用于：

- 样本量较小；
- 模型复杂或传统理论方法不适用。

Wild Bootstrap 是特别适用于存在异方差情形的 Bootstrap 方法，由 Wu (1986) 提出：

- 在回归残差上引入随机权重；
- 不直接对原始观测值重抽样；
- 更适合时间序列与异方差回归框架。

系统介绍可参见 Davison and Hinkley (1997)。

基于 AR(p) 模型的 EMH 检验

- 考虑 AR(p) 模型：

$$Y_t = \mu + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_t,$$

其中 $E(\varepsilon_t | Y_{t-1}, \dots, Y_{t-p}) = 0$ 。

- 在弱式 EMH 下：

- 历史收益（滞后项）不应对当前收益具有预测力；
- 原假设：

$$H_0 : \beta_1 = \cdots = \beta_p = 0.$$

- 备择假设：至少存在某个 j 使 $\beta_j \neq 0$ 。

Wald 统计量与异方差稳健估计

- 令 X 为 $(T - p - 1) \times (p + 1)$ 矩阵：

- 第一列全为 1；
- 其余列为 Y_{t-1}, \dots, Y_{t-p} 。

- Wald 统计量：

$$W = T(\hat{\beta} - \beta)' \hat{V}^{-1} (\hat{\beta} - \beta),$$

$\hat{\beta}$ 为 OLS 估计， \hat{V} 为渐近方差估计。

- 同方差情形 (i.i.d.)：

$$\hat{V} = \hat{\sigma}_\varepsilon^2 (X' X / T)^{-1}.$$

- 存在异方差时 (rw2、rw3)：

$$\hat{V}_W = (X' X / T)^{-1} (X' D X / T) (X' X / T)^{-1},$$

$$D = (\hat{\varepsilon}_{p+1}^2, \dots, \hat{\varepsilon}_T^2).$$

- 在 H_0 下， $W \sim \chi_p^2$ 。

异方差性与 EMH

- 金融市场中异方差性（波动率时变）十分常见：
 - 宏观新闻、风险事件、政策变化等都会影响波动。
- 重要的是：
 - 异方差性本身并不与 EMH 冲突；
 - EMH 关注的是信息被吸收的速度与方式，而非波动是否稳定。
- 因此：
 - 必须在 rw2、rw3 等更一般设定下检验 EMH；
 - 需要异方差稳健的统计方法（如 White 标准误、wild bootstrap）。

rw2 情形下的极限分布

- 在 rw2 下: $\bar{Y}_t = Y_t - E(Y_t)$ 独立但非同分布。
- 定义:

$$\lambda_{jj} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=j+1}^T E(\bar{Y}_{t-j}^2) E(\bar{Y}_t^2),$$

$$\gamma_0 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\bar{Y}_t^2),$$

$$V_{2;jj} = \frac{\lambda_{jj}}{\gamma_0^2}, \quad V_2(p) = (V_{2;11}, \dots, V_{2;pp}),$$

$$c_{j,p} = 2\left(1 - \frac{j}{p}\right), \quad \omega_2(p) = \sum_{j=1}^{p-1} c_{j,p}^2 V_{2;jj}.$$

rw2 情形下的 CLT

定理

在 rw2 下，若 $E(|Y_t|^{2+\delta}) \leq M$ 且 γ_0 与 λ_j 存在 ($j = 1, \dots, p$)，则当 $T \rightarrow \infty$ ，

$$\sqrt{T}(\hat{\rho}(1), \dots, \hat{\rho}(p))' \xrightarrow{d} N(0, V_2(p)),$$

$$\sqrt{T}(\widetilde{VR}(p) - 1) \xrightarrow{d} N(0, \omega_2(p)).$$

- 说明：在均值与二阶矩“平均意义上稳定”的条件下，依然可以进行标准化与正态近似。

rw3 情形下的假设 (1/2)

Assumption

设 $\tilde{Y}_t = Y_t - E(Y_t)$ 。对所有 t 与任意 $j \geq 0$:

$$E(\tilde{Y}_t) = 0, \quad E(\tilde{Y}_t \tilde{Y}_{t-j}) = 0,$$

且对任意 $s \neq t$ 与 $j, k = 1, \dots, p$,

$$E(\tilde{Y}_t \tilde{Y}_{t-j} \tilde{Y}_s \tilde{Y}_{s-k}) = 0.$$

Assumption

\tilde{Y}_t 为强混合过程, 混合系数 $\alpha(m)$ 的衰减率为 $O(m^{-r})$, 其中 $r > 1$ 。存在 $\delta > 0$ 使得对所有 t 与任意 $j \geq 0$,

$$E\left(\left|\tilde{Y}_t \tilde{Y}_{t-j}\right|^{2(r+\delta)}\right) \leq C < \infty.$$

rw3 情形下的假设 (2/2)

Assumption

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\tilde{Y}_t^2) = \sigma^2 < \infty.$$

Assumption

对所有 t 与任意 $j, k \geq 0$ (且 $j \neq k$), 有

$$E(\tilde{Y}_t^2 \cdot \tilde{Y}_{t-j} \cdot \tilde{Y}_{t-k}) = 0.$$

Assumption

对 $j = 1, \dots, p$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\tilde{Y}_t^2 \tilde{Y}_{t-j}^2) = \lambda_j < \infty.$$

rw3 情形下的极限结果与实务操作

- linton2019financial 引入 $E(\tilde{Y}_t \tilde{Y}_{t-j} \tilde{Y}_s \tilde{Y}_{s-k}) = 0$ 以避免渐近方差为无穷和。
- 在上述假设下，定理 3 的渐近正态性结果仍然成立。
- 实务操作：
 - 以样本估计量替代 γ_0, λ_j 等参数；
 - 具体公式参见 linton2019financial (2019, 第 108 页, 式 (3.40)–(3.43))。
- 总结：
 - 在 rw2、rw3 这类更一般设定下，仍可基于自相关与方差比检验 EMH；
 - 只是渐近方差需要相应调整与估计。

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

案例背景：贵州茅台（600519）

- 研究对象：贵州茅台酒股份有限公司（股票代码 600519）。
- 时间区间：2023 年 1 月 3 日至 2023 年 12 月 29 日。
- 公司概况：
 - 中国著名的大曲酱香型白酒生产商；
 - 2023 年品牌价值 875.24 亿美元；
 - 被视为中国文化与高端消费的代表品牌。
- 研究目的：
 - 采用序列相关性检验与方差比检验；
 - 检验 2023 年贵州茅台价格是否符合弱式 EMH 下的 $rw1$ 随机游走模型。

R 代码：数据处理与检验（1/2）

```
# 清除环境中的所有对象
rm(list = ls())
# 设置工作目录（在 RStudio 中）
if (requireNamespace("rstudioapi", quietly = TRUE) &&
    rstudioapi::isAvailable()) {
  setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
}

library(tseries)
library(ggplot2)
library(vrtest)

Moutaidata <- read.csv("Moutai.csv")

# 提取数据
Moutaidata <- data.frame( Date = as.Date(Moutaidata$date),
                           MoutaiPrice = Moutaidata$GuizhouMoutai_600519_ClosePrice_Adj)
```

R 代码：数据处理与检验（2/2）

```
# 计算差分与对数收益率
Moutaidata$DiffPrice <- c(NA, diff(Moutaidata$MoutaiPrice))
Moutaidata$LogReturns <- c(NA, diff(log(Moutaidata$MoutaiPrice)))
Moutaidata <- Moutaidata[-1, ]  
  
# Box-Pierce & Ljung-Box 检验（价格差分）
Box_Pierce_result <- Box.test(Moutaidata$DiffPrice,
                                 lag = 10, type = "Box-Pierce")
Box_Ljung_result <- Box.test(Moutaidata$DiffPrice,
                               lag = 10, type = "Ljung-Box")  
  
# 自动方差比检验（对数收益率）
AutoBoot_result <- AutoBoot.test(Moutaidata$LogReturns,
                                   nboot = 500, wild = "Normal")
```

Box-Pierce 与 Ljung-Box 检验结果

```
> Box_Pierce_result
```

Box-Pierce test

```
data: Moutaidata$DiffPrice  
X-squared = 11.957, df = 10, p-value = 0.2879
```

```
> Box_Ljung_result
```

Box-Ljung test

```
data: Moutaidata$DiffPrice  
X-squared = 12.384, df = 10, p-value = 0.2602
```

- 滞后 10 期下，两种检验的 p 值均大于 0.05：
 - 未发现价格差分（近似收益）存在显著自相关；
 - 支持 rw1 下“无自相关”的假设。

自动方差比检验（AutoBoot.test）结果（1/2）

```
> AutoBoot_result
```

```
$test.stat
```

```
[1] -0.49597
```

```
$VRsum
```

```
[1] 0.9459821
```

```
$pval
```

```
[1] 0.324
```

```
$CI.stat
```

```
2.5% 97.5%
```

```
-1.120601 1.254277
```

```
$CI.VRsum
```

```
2.5% 97.5%
```

```
0.8533894 1.1777045
```

自动方差比检验 (AutoBoot.test) 结果 (2/2)

- 检验结果解读：

- `test.stat = -0.49597`: 统计量接近 0，未显示显著偏离；
- `VRsum = 0.946`: 方差比接近 1；
- `pval = 0.324`: 大于 0.05，无法拒绝随机游走假设；
- `CI.stat` 区间包含 0；
- `CI.VRsum` 区间包含 1。

案例小结：贵州茅台与弱式 EMH

- 序列相关性检验 (Box–Pierce、Ljung–Box):
 - 未发现显著自相关；
 - 差分价格序列与随机游走相容。
- 方差比检验 (AutoBoot.test + wild bootstrap):
 - 方差比接近 1，置信区间包含 1；
 - 统计量置信区间包含 0， p 值较大。
- 结论（就 2023 年样本期而言）：
 - 贵州茅台价格走势与弱式有效市场相符；
 - 历史价格与收益信息未表现出可用于预测未来价格的统计证据。

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

本章小结

- 本章围绕收益可预测性与有效市场假说（EMH）展开：
 - 介绍 EMH 的三种形式及其含义；
 - 引入随机游走价格模型作为检验弱式 EMH 的理论框架。
- 对 EMH 的假设检验：
 - 序列相关性检验：基于样本自相关、Box–Pierce 与 Ljung–Box 统计量；
 - 方差比检验：比较不同时间尺度收益方差，刻画自相关结构；
 - AR 模型检验：利用自回归系数的显著性检验可预测性；
 - 在 rw2、rw3 等更一般假设下，对上述统计量的渐近分布进行修正。
- 实证案例：
 - 以 2023 年贵州茅台数据为例；
 - 综合序列相关性与方差比检验；
 - 结果支持该股票在样本期内的大致弱式有效性。

目录

- 1 收益可预测性与有效市场假说
- 2 有效市场假说
- 3 弱式有效市场的检验
- 4 半强式与强式有效性检验
- 5 异常收益与信息集
- 6 随机游走价格模型
- 7 对 EMH 的假设检验
 - 序列相关性检验
 - 方差比检验
 - 依照 AR 模型对 EMH 进行检验
 - 对 rw_2 与 rw_3 进行假设检验
- 8 案例：贵州茅台的弱式有效性检验
- 9 章节总结
- 10 参考文献

参考文献 |

-  Asquith, Paul and David W Mullins Jr (1986). "Equity issues and offering dilution". In: *Journal of Financial Economics* 15.1-2, pp. 61–89.
-  Ball, Ray and Philip Brown (1968). "An Empirical Evaluation of Accounting Income Numbers". In: *Journal of Accounting Research* 6.2, pp. 159–178. ISSN: 00218456, 1475679X. URL:
<http://www.jstor.org/stable/2490232> (visited on 09/29/2025).
-  Barber, Brad M and Terrance Odean (2000). "Trading is hazardous to your wealth: The common stock investment performance of individual investors". In: *The Journal of Finance* 55.2, pp. 773–806.
-  Brealey, Richard A, Stewart C Myers, and Franklin Allen (2014). *Principles of corporate finance*. McGraw-hill.
-  Campbell, John Y, Andrew W Lo, and A Craig MacKinlay (1997). *The econometrics of financial markets*. Princeton University press.
-  Carlton, Dennis W and Daniel R Fischel (1982). "The regulation of insider trading". In: *Stan. L. Rev.* 35, p. 857.

参考文献 II

-  Choi, In (1999). "Testing the random walk hypothesis for real exchange rates". In: *Journal of Applied Econometrics* 14.3, pp. 293–308.
-  Davison, Anthony Christopher and David Victor Hinkley (1997). *Bootstrap methods and their application*. 1. Cambridge university press.
-  Foster, George (1973). "Stock market reaction to estimates of earnings per share by company officials". In: *Journal of Accounting Research*, pp. 25–37.
-  Hayn, Carla (1995). "The information content of losses". In: *Journal of accounting and economics* 20.2, pp. 125–153.
-  Jensen, Michael C (1969). "Risk, the pricing of capital assets, and the evaluation of investment portfolios". In: *The Journal of Business* 42.2, pp. 167–247.
-  Keown, Arthur J and John M Pinkerton (1981). "Merger announcements and insider trading activity: An empirical investigation". In: *The Journal of Finance* 36.4, pp. 855–869.

参考文献 III

-  Lim, Kian-Ping and Robert Brooks (2011). "The evolution of stock market efficiency over time: A survey of the empirical literature". In: *Journal of Economic Surveys* 25.1, pp. 69–108.
-  Lo, Andrew W. (2008). "Efficient markets hypothesis". In: *The New Palgrave Dictionary of Economics Online*. Ed. by Steven N. Durlauf and Lawrence E. Blume. 2nd ed. New York: Palgrave Macmillan. DOI: 10.1057/9780230226203.0454.
-  Lo, Andrew W and A Craig MacKinlay (1988). "Stock market prices do not follow random walks: Evidence from a simple specification test". In: *The Review of Financial Studies* 1.1, pp. 41–66.
-  Manne, Henry (1966). *Insider Trading and the Stock Market*. New York: The Free Press.
-  Meulbroek, Lisa K (1992). "An empirical analysis of illegal insider trading". In: *The Journal of Finance* 47.5, pp. 1661–1699.

参考文献 IV

-  Poterba, James M and Lawrence H Summers (1988). "Mean reversion in stock prices: Evidence and implications". In: *Journal of Financial Economics* 22.1, pp. 27–59.
-  Sewell, Martin (2011). "History of the efficient market hypothesis". In: *Rn* 11.04, p. 04.
-  Solnik, Bruno H (1973). "Note on the validity of the random walk for European stock prices". In: *The Journal of Finance* 28.5, pp. 1151–1159.
-  Wu, Chien-Fu Jeff (1986). "Jackknife, bootstrap and other resampling methods in regression analysis". In: *The Annals of Statistics* 14.4, pp. 1261–1295.

第 7 章非参数方法

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

非参数方法在金融计量中的作用

- 非参数方法不依赖具体分布假设，具有更强的灵活性；
- 可用于估计资产价格的波动率：
 - 传统 GARCH 等参数模型可能难以刻画复杂非线性特征；
 - 非参数方法（核平滑、局部线性回归等）能更好适应尾部与异常值。
- 在高频金融数据分析中尤为重要：
 - Zhou (1996) 首次系统讨论微观结构噪声；
 - 利用核方法估计高频数据的波动率；
 - 能处理不规则时间间隔与微观结构噪声。
- 在风险管理中：
 - 用于估计损失分布尾部，进而计算 VaR 与 CVaR；
 - 对极端风险与“黑天鹅”事件的刻画尤为关键。

- **一维核密度估计：** i.i.d. 情形下的概率密度估计；
- **多维核密度估计：** 渐近性质与“维数诅咒”；
- **非参数回归：** Nadaraya-Watson 估计与局部多项式（本节只预热）；
- **带宽选择与 MSE：** 偏差-方差权衡与最优带宽；
- **金融应用示例：**
 - 收益率分布与尾部风险；
 - 多市场联合密度（如 FTSE & DAX）；
 - 利率期限结构与仿射模型检验的简述。

目录

1 非参数概率密度估计量

2 多维核密度估计

3 非参数回归模型

- Nadaraya-Watson 估计量
- 局部多项式估计量

4 参考文献

核密度估计：基本思想

- 核密度估计通过在每个样本点 X_t 上放置一个“核函数”，并做加总平滑：

$$\hat{g}(x) = \frac{1}{Th} \sum_{t=1}^T K\left(\frac{x - X_t}{h}\right).$$

- 核函数 $K(\cdot)$** : 对邻近样本赋权的平滑函数;
- 带宽 h** : 控制平滑程度的关键参数:
 - h 小 \Rightarrow 估计更“粗糙”、波动大（欠平滑）；
 - h 大 \Rightarrow 估计过于平滑，细节丢失（过平滑）。
- 优点**:
 - 自动适应偏态、多峰、厚尾等复杂分布；
 - 不要求正态性或具体参数形式；
 - 特别适合刻画金融收益的重尾与偏度。

核函数与带宽的作用

- 核函数 $K(\cdot)$ 的典型性质：

- 非负， $\int K(u) du = 1$ ；
- 一般为对称“钟型”，权重随 $|u|$ 增大而递减；
- 常用核：高斯核、Epanechnikov 核、四次核等。

- 带宽 h 决定核在 x 周围“看到”的数据范围：

小 h ：强调局部、噪声大 大 h ：强调整体、更平滑但偏差大

- 实务中带宽选择：

- MSE/AMISE 准则（理论最优）；
- 经验法则（如 Silverman rule-of-thumb）；
- 交叉验证（cross-validation）。

二阶核函数与 q 阶核

定义 (二阶核函数 $K(\cdot)$)

二阶 (或正) 核函数是满足:

- $\int_{-\infty}^{\infty} K(u) du = 1;$
- $\int_{-\infty}^{\infty} uK(u) du = 0;$
- $\int_{-\infty}^{\infty} u^2 K(u) du = C_K < \infty;$
- $\int_{-\infty}^{\infty} K^2(u) du = D_K < \infty.$

二阶核函数与 q 阶核

定义 (q 阶核)

若 $K(\cdot)$ 满足：

- $\int K(u) du = 1;$
- 对 $1 \leq j \leq q - 1$, $\int u^j K(u) du = 0;$
- $\int u^q K(u) du < \infty;$
- $\int K^2(u) du < \infty,$

则称 K 为 q 阶核。

q 阶核的意义

- 高阶核可以减少估计偏差：
 - 更好地拟合目标函数的高阶平滑特征；
 - 对局部曲率变化大的情形更有效。
- 但高阶核也可能增加方差，对小样本不友好，需要权衡。

常用二阶核

常用二阶核：

- 均匀核 (Uniform) :

$$K(u) = \frac{1}{2} \mathbf{1}(|u| \leq 1).$$

- 高斯核 (Gaussian) :

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \quad -\infty < u < \infty.$$

- Epanechnikov 核:

$$K(u) = \frac{3}{4}(1 - u^2) \mathbf{1}(|u| \leq 1).$$

- 四次方核 (Quartic) :

$$K(u) = \frac{15}{16}(1 - u^2)^2 \mathbf{1}(|u| \leq 1).$$

R 代码：常见核函数定义

```
# 均匀核
func_uniform <- function(x) {
  y <- numeric(length(x))
  idx <- abs(x) <= 1
  y[idx] <- 1/2
  return(y)
}

# 高斯核
func_gaussian <- function(x) {
  y <- (1 / sqrt(2 * pi)) * exp(-0.5 * x^2)
  return(y)
}

# Epanechnikov 核
func_e <- function(x) {
  y <- numeric(length(x))
  idx <- abs(x) <= 1
  y[idx] <- (3 / 4) * (1 - x[idx]^2)
  return(y)
}
```

R 代码：常见核函数定义

```
# 四次方核
func_quartic <- function(x) {
  y <- numeric(length(x))
  idx <- abs(x) <= 1
  y[idx] <- (15 / 16) * (1 - x[idx]^2)^2
  return(y)
}
```

R 代码：绘制核函数图像

```
library(ggplot2)
library(dplyr)

# 自变量网格
x_vals <- seq(-3, 3, by = 0.01)

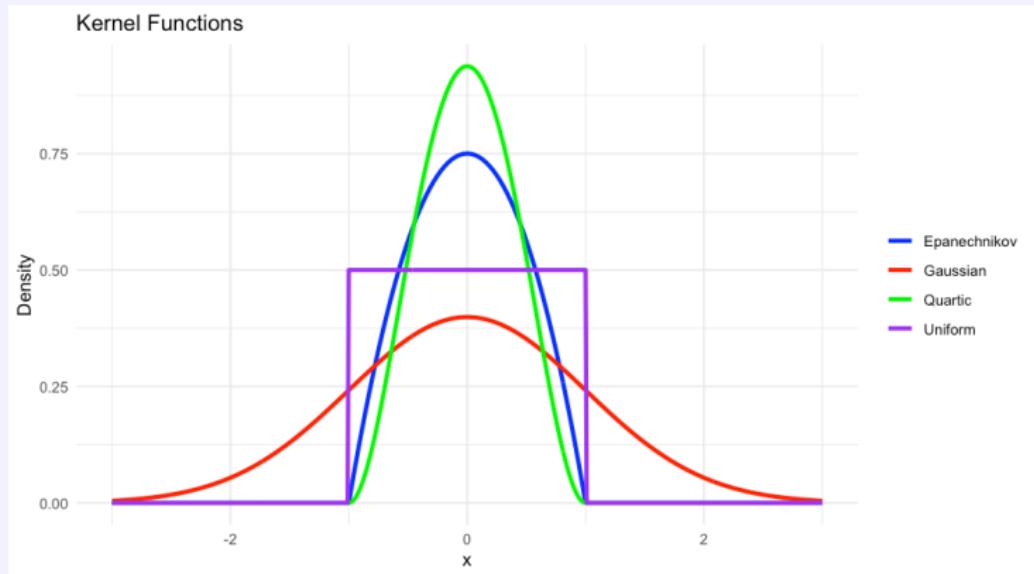
# 构造四类核的数据框
df_uniform <- data.frame(x = x_vals,
                           y = func_uniform(x_vals),
                           kernel = "Uniform")
df_gaussian <- data.frame(x = x_vals,
                           y = func_gaussian(x_vals),
                           kernel = "Gaussian")
df_e <- data.frame(x = x_vals,
                     y = func_e(x_vals),
                     kernel = "Epanechnikov")
df_quartic <- data.frame(x = x_vals,
                           y = func_quartic(x_vals),
                           kernel = "Quartic")
```

R 代码：绘制核函数图像

```
df_kernels <- bind_rows(df_uniform, df_gaussian,
                         df_e, df_quartic)

ggplot(df_kernels, aes(x = x, y = y, color = kernel)) +
  geom_line(size = 1.2) +
  labs(title = "Kernel Functions", x = "x", y = "Density") +
  theme_minimal() +
  theme(legend.title = element_blank())
```

核函数图像与选择



图：均匀核、高斯核、Epanechnikov 核与四次方核

核函数图像与选择

- 高斯核支撑集为 \mathbb{R} , 其余核支撑在 $[-1, 1]$;
- 均匀核在支撑内等权重; 其他核采用“钟形”权重;
- 对给定带宽, Epanechnikov 核可最小化 MSE;
- 实务上, 估计结果对核的选择不敏感, 对带宽非常敏感。

核估计与直方图的关系

例 (均匀核与直方图)

若 $K(u) = \frac{1}{2}\mathbf{1}(|u| \leq 1)$, 则

$$\hat{g}(x) = \frac{1}{2hT} \sum_{t=1}^T \mathbf{1}(|x - X_t| \leq h).$$

- 这相当于以 x 为中心、长度 $2h$ 区间内的样本相对频率；
- 当 h 很小、 T 很大时：

$$E\mathbf{1}(|x - X_t| \leq h) = \int_{x-h}^{x+h} g(y) dy \approx 2h g(x).$$

- 因此直方图是 $g(x)$ 的合理估计；在 $h \rightarrow 0$ 且 $Th \rightarrow \infty$ 时， $\hat{g}(x)$ 一致。

示例：DAX 指数日收益率的密度估计

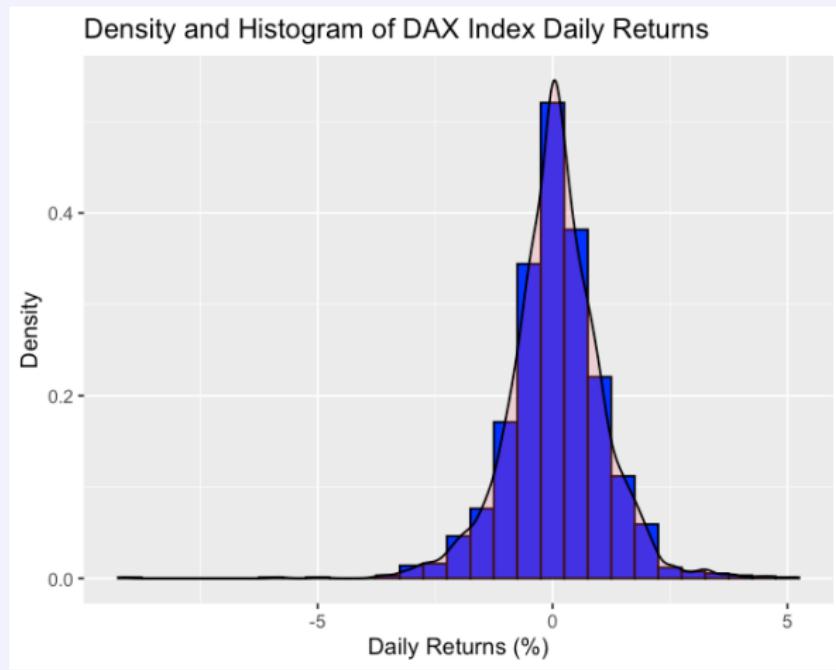
```
library(datasets)
data("EuStockMarkets")

# 提取 DAX 收盘价并计算日收益率（百分比）
dax_prices <- EuStockMarkets[, "DAX"]
dax_returns <- diff(dax_prices) /
    lag(dax_prices, -1) * 100
dax_returns <- na.omit(dax_returns)

dax_returns_df <- data.frame(DAX_Returns = dax_returns)

library(ggplot2)
ggplot(dax_returns_df, aes(x = DAX_Returns)) +
  geom_histogram(aes(y = ..density..),
                 binwidth = 0.5,
                 color = "black") +
  geom_density(alpha = .2, fill = "#FF6666") +
  labs(title = "Density and Histogram of DAX Index Daily Returns",
       x = "Daily Returns (%)", y = "Density")
```

DAX 日收益率分布的特征



图：DAX 指数日收益率的核密度与直方图估计

DAX 日收益率分布的特征

- 分布常呈**非对称**（偏度显著）；
- 尾部比正态更“厚”，极端波动发生概率更大；
- 这类重尾、偏态特征是金融收益的典型现象；
- 非参数密度估计比简单正态假设更能反映真实风险结构。

一维核密度估计的理论假设

Assumption (概率密度函数的平滑性)

- (i) $\{X_t\}$ 为严平稳过程, 边际密度为 $g(x)$;
- (ii) $g(x)$ 在 $[a, b]$ 上连续两次可微;
- $g''(\cdot)$ 在 $[a, b]$ 上 Lipschitz 连续:

$$|g''(x_1) - g''(x_2)| \leq C|x_1 - x_2|, \quad \forall x_1, x_2 \in [a, b].$$

Assumption (二阶正值核, 紧支撑)

$K(u)$ 为二阶正值核, 且支撑集为 $[-1, 1]$ 。

- 有界支撑假设可简化推导, 但不是必要条件。

偏差-方差分解与 MSE

对于内部点 $x \in [a+h, b-h]$, 核密度估计:

$$\hat{g}(x) = \frac{1}{Th} \sum_{t=1}^T K\left(\frac{x-X_t}{h}\right).$$

偏差-方差分解:

$$\text{MSE}[\hat{g}(x)] = \{\text{E}\hat{g}(x) - g(x)\}^2 + \text{Var}[\hat{g}(x)].$$

在适当正则条件下, 可得:

$$\text{E}\hat{g}(x) - g(x) = \frac{1}{2}h^2 C_K g''(x) + o(h^2),$$

$$\text{Var}[\hat{g}(x)] = \frac{1}{Th} g(x) D_K + o((Th)^{-1}).$$

因此:

$$\text{MSE}[\hat{g}(x)] = \frac{1}{Th} g(x) D_K + \frac{1}{4} h^4 [g''(x)]^2 C_K^2 + o(Th^{-1} + h^4).$$

勒贝格控制收敛定理的作用（偏差项）

定理（勒贝格控制收敛定理（简述））

若可测函数列 $f_n(x) \rightarrow f(x)$ 几乎处处，且存在可积函数 $g(x)$ 使

$$|f_n(x)| \leq g(x)$$

对所有 n 与几乎处处的 x 成立，则

$$\lim_{n \rightarrow \infty} \int f_n(x) dx = \int \lim_{n \rightarrow \infty} f_n(x) dx = \int f(x) dx.$$

在偏差推导中：

- 被控函数： $[g''(x + \lambda h u) - g''(x)] u^2 K(u)$ ；
- 控制函数： $|g''(\cdot)| \cdot |u^2 K(u)|$ ，其积分有限；
- 利用控制收敛定理可交换 $h \rightarrow 0$ 的极限与积分，从而得到偏差的 $O(h^2)$ 结论。

核密度估计的边界问题

- 当 x 接近支撑端点（如 $[a, a+h]$ 或 $(b-h, b]$ ）时：
 - 核函数无法在左右两侧对称地“看到”数据；
 - 导致“质量流失”（kernel mass lost），偏差为 $O(1)$ 而非 $O(h^2)$ ；
 - 即所谓的 边界效应。
- 边界区域的估计往往系统性偏低（或偏高），对尾部密度影响显著；
- 在金融应用中，尾部密度对应极端风险（如市场崩盘），不能简单丢弃。

边界问题的若干修正方法

- **简单但不推荐：**只在内部区间 $[a + h, b - h]$ 上使用估计；
 - 丢失尾部信息；
 - 对金融尾部风险分析不利。
- **边界核方法** (Hong and Li 2005): $\hat{g}(x) = \frac{1}{T} \sum_{t=1}^T K_h(x, X_t)$, 其中

$$K_h(x, y) = \begin{cases} h^{-1} K\left(\frac{x-y}{h}\right), & x \in [0, h], \\ \frac{\int_{-x/h}^1 K(u) du}{\int_{-x/h}^1 K(u) du}, & x \in [h, 1-h], \\ h^{-1} K\left(\frac{x-y}{h}\right), & x \in (1-h, 1]. \end{cases}$$

- **优点：**修正核使其在边界处仍积为 1; 保持非负；
- **缺点：**边界偏差阶为 $O(h)$, 比内部 $O(h^2)$ 慢。

数据反射法 (Schuster, 1985)

- 思想：构造增广样本

$$\{X_t\}_{t=1}^T \cup \{-X_t\}_{t=1}^T,$$

在 $[0, 1]$ 支撑上进行估计；

- 对边界点 $x \in [0, h)$ ，估计量为

$$\hat{g}(x) = \frac{1}{2T} \sum_{t=1}^T \left\{ K_h(x - X_t) + K_h(x + X_t) \right\}.$$

- 若 x 离边界远，则 $K_h(x + X_t)$ 项为 0，退化为普通核估计；
- 该方法仅在边界区域修正估计；
- 后续被 Chen and Hong (2012) 拓展，用于时变函数估计。

核估计的方差与独立同分布假设

Assumption (独立同分布观测)

样本 $\{X_t\}_{t=1}^T$ 独立同分布。

- 定义

$$Z_t(x) = K_h(x - X_t) - \mathbb{E} K_h(x - X_t).$$

则 $\{Z_t\}$ i.i.d., 均值 0;

- 由此可得

$$\text{Var}[\hat{g}(x)] = \frac{1}{Th} g(x) D_K + o((Th)^{-1}).$$

- Th 可以看作落在区间 $[x - h, x + h]$ 的有效样本量；
- 在适当 α -混合条件下，弱依赖序列也有同样的渐近方差 (Hart 1996)。

均方误差与最优带宽

$$\text{MSE}[\hat{g}(x)] = \frac{1}{Th} g(x) D_K + \frac{1}{4} h^4 [g''(x)]^2 C_K^2 + o(Th^{-1} + h^4).$$

- 小 h :
 - 偏差小 ($O(h^2)$), 方差大 ($O((Th)^{-1})$), 欠平滑;
- 大 h :
 - 方差小, 偏差大, 过平滑;
- 最优带宽 h_0 通过最小化 MSE 得到:

$$h_0 = \left[\frac{D_K}{C_K^2} \frac{1/g(x)}{[g''(x)/g(x)]^2} \right]^{1/8} T^{-1/5}.$$

- 对应的最优收敛率:

$$\hat{g}(x) - g(x) = O_p(T^{-2/5}),$$

明显慢于参数估计的 $T^{-1/2}$ 。

目录

1 非参数概率密度估计量

2 多维核密度估计

3 非参数回归模型

- Nadaraya-Watson 估计量
- 局部多项式估计量

4 参考文献

多维核密度估计与记号

- 设 $X_t = (X_{1t}, \dots, X_{dt})'$, 联合密度 $f(x)$;
- 使用乘积核:

$$\hat{f}(x) = \frac{1}{T} \sum_{t=1}^T \prod_{i=1}^d K_h(x_i - X_{it}) = \frac{1}{T} \sum_{t=1}^T \mathcal{K}_h(x - X_t).$$

- x 为内部点: 每个维度上 $x_i \in [a_i + h, b_i - h]$;
- 目标: 分析 $\hat{f}(x)$ 的偏差、方差与 MSE。

多维核密度估计的偏差

对内部点 x , 在类似一维假设下, 可得:

$$\mathbb{E}\hat{f}(x) - f(x) = \frac{1}{2}h^2 C_K \sum_{i=1}^d f_{ii}(x) + o(h^2),$$

其中

$$f_i(x) = \frac{\partial}{\partial x_i} f(x), \quad f_{ij}(x) = \frac{\partial^2}{\partial x_i \partial x_j} f(x),$$

$\sum_{i=1}^d f_{ii}(x)$ 是 f 的拉普拉斯算子 (Laplacian)。

- 偏差仍为 $O(h^2)$;
- 但涉及所有维度的二阶偏导之和;
- 凸显多维密度在各方向曲率对偏差的重要性。

多维核密度估计的方差与 MSE

定义

$$Z_t(x) = \mathcal{K}_h(x - X_t) - E\mathcal{K}_h(x - X_t).$$

若 $\{X_t\}$ i.i.d., 则 $\{Z_t\}$ i.i.d., 均值为 0。进一步可得：

$$\text{Var}[\hat{f}(x)] = \frac{1}{Th^d} f(x) D_K^d + o((Th^d)^{-1}).$$

因此：

$$\text{MSE}[\hat{f}(x)] = \frac{1}{Th^d} f(x) D_K^d + \frac{1}{4} C_K^2 h^4 \left[\sum_{i=1}^d f_{ii}(x) \right]^2 + o(Th^{-d} + h^4).$$

多维核密度估计的方差与 MSE

最优带宽：

$$h_0 = \left[\frac{dD_K^2}{C_K^2} \frac{1/f(x)}{\left(\sum_{i=1}^d f_{ii}(x)/f(x) \right)^2} \right]^{1/(d+4)} T^{-1/(d+4)}.$$

对应最优 MSE 收敛率：

$$\text{MSE}[\hat{f}(x)] \propto T^{-4/(d+4)}.$$

“维数诅咒”

- 当维度 d 增大时：

$$\text{MSE}[\hat{f}(x)] \propto T^{-4/(d+4)},$$

收敛速度明显变慢；

- 例：

- $d = 1: T^{-4/5};$

- $d = 2: T^{-2/3};$

- $d = 3: T^{-4/7};$

- d 越大，所需样本量呈指数级增长。

- 这就是维数诅咒：

- 高维非参数估计需要巨大样本；

- 实证金融与经济中，非参数维度通常不超过 3–5。

缓解维数诅咒的若干思路

- 结构性假设：

- 各分量独立： $f(x) = \prod_{i=1}^d g_i(x_i)$ ；
- 降维为多个一维密度估计。

- 马尔可夫结构：

$$f(X_t | \mathcal{F}_{t-1}) = f(X_t | X_{t-1}) = \frac{f(X_t, X_{t-1})}{g(X_{t-1})},$$

将无限维历史压缩为低维状态。

- 投影寻踪 (projection pursuit)：

- 假设多变量函数可表示为若干线性组合的未知函数之和；
- 线性组合系数与非参数函数同时估计；
- 在保持部分灵活性的同时降低有效维度。

应用：仿射期限结构模型与密度估计

定义（仿射期限结构模型）

即期利率 $r(t)$ 被表示为状态变量 $X(t)$ 的仿射函数：

$$r(t) = a + b'X(t),$$

其中 a 为常数、 b 为参数向量， $X(t)$ 通常满足某种随机微分方程。

仿射期限结构模型：

- 优点：
 - 给出债券价格与利率衍生品的解析表达式；
 - 便于定价、风险管理与敏感度分析。
- 非参数密度估计的作用：
 - 检验给定仿射模型是否与数据兼容；
 - 通过 $\hat{g}(x)$ 、 $\hat{f}(x, y)$ 检验扩散项设定；
 - 如 Aït-Sahalia (1996)、Hong and Li (2005) 对利率模型的设定检验。

R 示例：FTSE 与 DAX 收益的联合密度

```
library(datasets)
library(ks)

data("EuStockMarkets")

# 提取 FTSE 与 DAX 收盘价
ftse_prices <- EuStockMarkets[, "FTSE"]
dax_prices   <- EuStockMarkets[, "DAX"]

# 计算日收益率（百分比）
ftse_returns <- diff(ftse_prices) /
                  lag(ftse_prices, -1) * 100
dax_returns   <- diff(dax_prices) /
                  lag(dax_prices, -1) * 100

ftse_returns <- na.omit(ftse_returns)
dax_returns   <- na.omit(dax_returns)

market_returns <- data.frame(
  FTSE_Returns = ftse_returns,
  DAX_Returns  = dax_returns
)
```

R 示例：FTSE 与 DAX 收益的联合密度

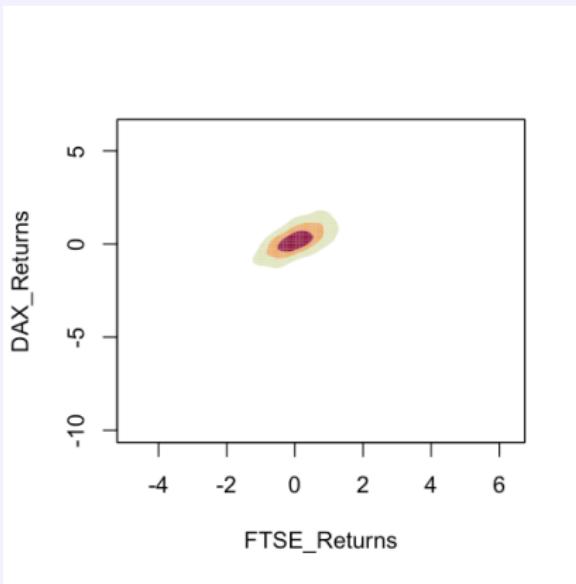
```
data_matrix <- as.matrix(market_returns)

# 多变量核密度估计
kde_result <- kde(x = data_matrix)

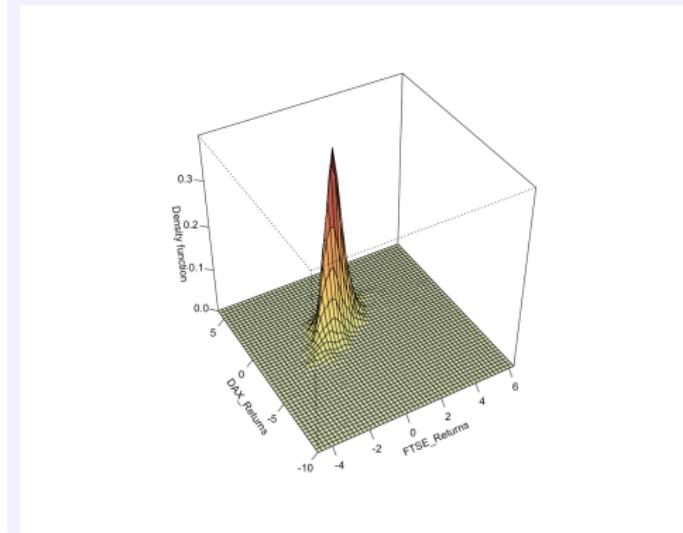
# 2D 等高线图
plot(kde_result, display = "filled.contour")

# 3D 透视图
plot(kde_result, display = "persp")
```

多维核密度估计的可视化



(a) 二维等高线图



(b) 三维透视图

图: FTSE 与 DAX 日收益率的联合核密度估计

目录

1 非参数概率密度估计量

2 多维核密度估计

3 非参数回归模型

- Nadaraya-Watson 估计量
- 局部多项式估计量

4 参考文献

为什么需要非参数回归？

- 无需预设函数形式：
 - 不要求 Y 与 X 存在“线性”或“多项式”关系；
 - 更适合刻画复杂、未知的非线性结构。
- 对模型设定更稳健：
 - 不依赖严格分布与函数假设；
 - 面对非线性、异方差、复杂依赖结构时，减少设定偏差 (misspecification bias)。
- 完全数据驱动 (data-driven)：
 - 特别适合探索性数据分析；
 - 可帮助发现数据中的新模式与规律。
- 本节重点：
 - Nadaraya-Watson (局部常数) 估计量；
 - 局部多项式估计量 (含局部线性)。

示例：线性回归 vs LOESS 非参数回归

```
library(ggplot2)
library(ggpubr)
set.seed(123)

# 模拟非线性数据:  $y = \sin(x) + \text{噪声}$ 
x_vals <- seq(-3, 3, length.out = 100)
y_vals <- sin(x_vals) + rnorm(length(x_vals), sd = 0.3)
data <- data.frame(x = x_vals, y = y_vals)

# 线性回归
linear_model <- lm(y ~ x, data = data)

# 非参数回归: LOESS
loess_model <- loess(y ~ x, data = data)

# 预测与残差
data$y_pred_linear <- predict(linear_model, newdata = data)
data$y_pred_loess <- predict(loess_model, newdata = data)
data$residuals_linear <- data$y - data$y_pred_linear
data$residuals_loess <- data$y - data$y_pred_loess
```

示例图：拟合曲线与残差

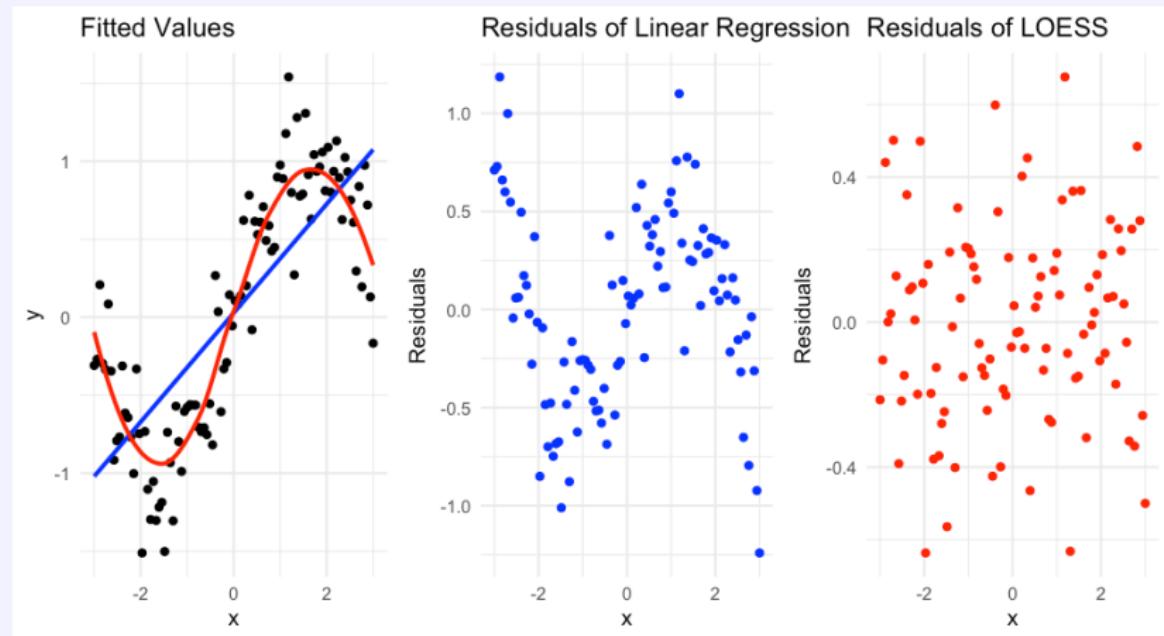
```
plot1 <- ggplot(data, aes(x = x)) +
  geom_point(aes(y = y), color = "black") +
  geom_line(aes(y = y_pred_linear), color = "blue", size = 1) +
  geom_line(aes(y = y_pred_loess), color = "red", size = 1) +
  labs(title = "Fitted Values", x = "x", y = "y") +
  theme_minimal()

plot2 <- ggplot(data, aes(x = x)) +
  geom_point(aes(y = residuals_linear), color = "blue") +
  labs(title = "Residuals of Linear Regression",
       x = "x", y = "Residuals") +
  theme_minimal()

plot3 <- ggplot(data, aes(x = x)) +
  geom_point(aes(y = residuals_loess), color = "red") +
  labs(title = "Residuals of LOESS",
       x = "x", y = "Residuals") +
  theme_minimal()

ggarrange(plot1, plot2, plot3, ncol = 3, nrow = 1)
```

示例结论：非参数回归的优势



图：线性回归 vs 非参数回归（LOESS）

示例结论：非参数回归的优势

- 黑点： $y = \sin x + \varepsilon$ 模拟数据；
- 蓝线：线性回归 \Rightarrow 明显无法捕捉非线性结构；
- 红线：LOESS 非参数回归 \Rightarrow 更好跟随真实曲线；
- 残差图：
 - 线性回归残差存在系统性模式；
 - LOESS 残差更接近“随机云”。

Nadaraya-Watson 核回归估计量

- 回归模型: $Y_t = r(X_t) + \varepsilon_t$, $E(\varepsilon_t | X_t) = 0$;
- 条件均值: $r(x) = E(Y_t | X_t = x)$ 为目标函数;
- Nadaraya-Watson 估计量:

$$\hat{r}(x) = \frac{\hat{m}(x)}{\hat{g}(x)},$$

其中

$$\hat{m}(x) = \frac{1}{T} \sum_{t=1}^T Y_t K_h(x - X_t), \quad \hat{g}(x) = \frac{1}{T} \sum_{t=1}^T K_h(x - X_t).$$

- 定义权重:

$$\hat{W}_t(x) = \frac{K_h(x - X_t)}{\sum_{s=1}^T K_h(x - X_s)}, \quad \sum_{t=1}^T \hat{W}_t(x) = 1,$$

则

$$\hat{r}(x) = \sum_{t=1}^T \hat{W}_t(x) Y_t.$$

直观理解：局部加权平均

- $K_h(x - X_t) = \frac{1}{h}K\left(\frac{x-X_t}{h}\right)$ 为 “距离权重”：
 - X_t 距离 x 越近，权重越大；
 - 若 K 支撑在 $[-1, 1]$ ，则 $|X_t - x| > h$ 的点权重为 0。
- 若核为均匀核 $K(u) = \frac{1}{2}\mathbf{1}(|u| \leq 1)$ ，则

$$\hat{r}(x) = \frac{\sum_{t=1}^T Y_t \mathbf{1}(|X_t - x| \leq h)}{\sum_{t=1}^T \mathbf{1}(|X_t - x| \leq h)},$$

即局部样本均值。

- Tukey (1961) 将其称为回归直方图 (regressogram)。

R 示例：Nadaraya-Watson 回归 (mtcars)

```
if (!requireNamespace("KernSmooth", quietly = TRUE)) {  
  install.packages("KernSmooth")  
}  
library(KernSmooth)  
  
data(mtcars)  
  
x <- mtcars$hp    # 解释变量：马力  
y <- mtcars$mpg   # 被解释变量：油耗 (mpg)  
  
# 使用 dpill 自动选择带宽  
bw <- dpill(x, y)  
  
# Nadaraya - Watson 核回归 (degree = 0)  
fit <- locpoly(x, y,  
                 bandwidth = bw,  
                 degree     = 0,  
                 kernel     = "normal",  
                 gridsize   = 100)
```

R 示例: Nadaraya-Watson 回归 (mtcars)

```
plot(x, y,
      main = "Nadaraya-Watson Estimation",
      xlab = "Horsepower",
      ylab = "Miles per Gallon",
      pch = 19)
lines(fit, col = "blue")
```

示例图：马力与油耗的非参数关系

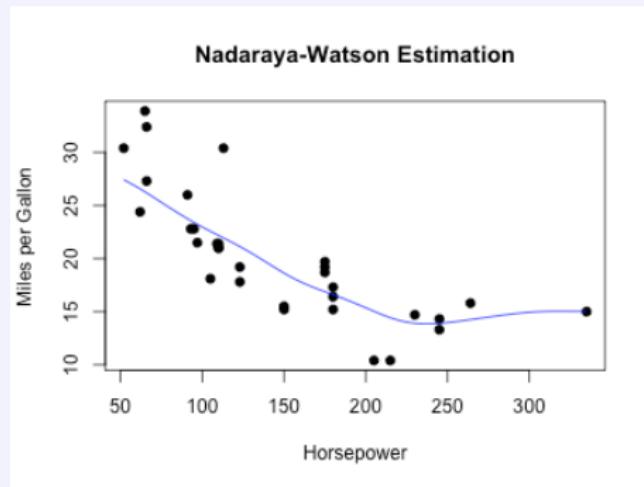


图: Nadaraya-Watson 估计: HP 对 MPG 的影响

- 蓝线为 Nadaraya-Watson 回归曲线；
- 不强制线性关系，可捕捉非线性“边际效应”；
- 在汽车数据中，往往马力越大，MPG 越低，但关系不一定是直线。

Nadaraya-Watson 的误差分解

- 设 $Y_t = r(X_t) + \varepsilon_t$, $E(\varepsilon_t | X_t) = 0$;

- 回顾

$$\hat{r}(x) = \frac{\hat{m}(x)}{\hat{g}(x)}, \quad \hat{m}(x) = \frac{1}{T} \sum Y_t K_h(x - X_t).$$

- 为分析渐近行为, 对任意固定 x , 分解:

$$\hat{r}(x) - r(x) = \frac{\hat{m}(x) - r(x)\hat{g}(x)}{\hat{g}(x)}.$$

- 进一步写作

$$\hat{m}(x) - r(x)\hat{g}(x) = \hat{V}(x) + \hat{B}(x),$$

其中

$$\hat{V}(x) = \frac{1}{T} \sum_{t=1}^T \varepsilon_t K_h(x - X_t) \quad (\text{方差部分}),$$

$$\hat{B}(x) = \frac{1}{T} \sum_{t=1}^T [r(X_t) - r(x)] K_h(x - X_t) \quad (\text{偏差部分}).$$

方差项：局部有效样本量的作用

假设 $\{(X_t, Y_t)\}$ i.i.d., 定义 $\sigma^2(x) = E(\varepsilon_t^2 | X_t = x)$, 则

$$E[\widehat{V}^2(x)] = \frac{1}{T} E[\varepsilon_t^2 K_h^2(x - X_t)] = \frac{1}{Th} \sigma^2(x) g(x) \int K^2(u) du [1 + o(1)].$$

由于

$$E\widehat{g}(x) \rightarrow g(x) \quad (h \rightarrow 0),$$

可得（标准化后）

$$E\left[\frac{\widehat{V}(x)}{E\widehat{g}(x)}\right]^2 = \frac{1}{Th} \frac{\sigma^2(x)}{g(x)} D_K [1 + o(1)], \quad D_K = \int K^2(u) du.$$

- 方差 $\propto (Th)^{-1}$;
- Th 约等于落在 $[x - h, x + h]$ 内的有效样本量；
- $\sigma^2(x)$ 越大、 $g(x)$ 越小，局部方差越大。

偏差项: r 与 g 的平滑性

- 记 $m(x) = r(x)g(x)$;
- 在内部点 $x \in [a+h, b-h]$, 若 r, g 足够平滑, 可得

$$\mathbb{E}\widehat{B}(x) = \frac{1}{2}h^2 \left[m''(x) - r(x)g''(x) \right] C_K + o(h^2),$$

其中 $C_K = \int u^2 K(u) du$;

- 由链式法则:

$$m''(x) = r''(x)g(x) + 2r'(x)g'(x) + r(x)g''(x),$$

故

$$\mathbb{E}\widehat{B}(x) = \frac{1}{2}h^2 \left[r''(x)g(x) + 2r'(x)g'(x) \right] C_K + o(h^2).$$

- 标准化后

$$\mathbb{E} \left[\frac{\widehat{B}(x)}{\mathbb{E}\widehat{g}(x)} \right] = \frac{1}{2}h^2 \left[r''(x) + 2r'(x) \frac{g'(x)}{g(x)} \right] C_K + o(h^2).$$

Nadaraya-Watson 的 MSE 与最优带宽

$$\mathbb{E}[\hat{r}(x) - r(x)]^2 = \frac{1}{Th} \frac{\sigma^2(x)}{g(x)} D_K + \frac{1}{4} h^4 \left[r''(x) + 2r'(x) \frac{g'(x)}{g(x)} \right]^2 C_K^2 + o(Th^{-1} + h)$$

- 方差项 $\propto (Th)^{-1}$;
- 均方偏差项 $\propto h^4$;
- 小 h : 偏差小、方差大 (欠平滑);
- 大 h : 方差小、偏差大 (过平滑)。

Nadaraya-Watson 的 MSE 与最优带宽

最小化 MSE 得最优带宽

$$h^* = c^* T^{-1/5},$$

其中

$$c^* = \left[\frac{D_K}{C_K^2} \frac{\sigma^2(x)g(x)}{(r''(x) + 2r'(x)g'(x)/g(x))^2} \right]^{1/5}.$$

- $\sigma^2(x)$ 大 \Rightarrow 最优 h^* 较大;
- $r(x)$ 越“不平滑”(导数大) \Rightarrow 最优 h^* 较小;
- 收敛率为 $T^{-2/5}$ (比参数 $T^{-1/2}$ 慢)。

AMISE：渐近积分均方误差

定义 (渐近积分均方误差)

渐近积分均方误差 (Asymptotic Mean Integrated Squared Error, 简称 AMISE) 是

$$= E \left[\int (\hat{\theta}(x) - \theta(x))^2 dx \right]$$

的大样本近似主项，其中 θ 为目标函数（密度或回归函数）， $\hat{\theta}$ 为其估计量。

AMISE：渐近积分均方误差

以 Nadaraya-Watson 回归为例，AMISE 可写为

$$(h) \approx \frac{R(K)}{Th} \int \frac{\sigma^2(x)}{g(x)} dx + \frac{1}{4} \mu_2(K)^2 h^4 \int [r''(x)]^2 g(x) dx,$$

其中 $R(K) = \int K^2(u) du$, $\mu_2(K) = \int u^2 K(u) du$.

- 最小化 AMISE $\Rightarrow h_{\text{opt}} \propto T^{-1/5}$;
- 各种核只影响常数项 $(R(K), \mu_2(K))$;
- 因此：带宽比核函数更重要。

Nadaraya-Watson 的另一种视角：局部常数

- 全局最小二乘：

$$\min_r \sum_{t=1}^T (Y_t - r)^2 \Rightarrow \hat{r} = \bar{Y}.$$

- 局部加权最小二乘：

$$\min_r \sum_{t=1}^T (Y_t - r)^2 K_h(x - X_t).$$

- 一阶条件：

$$\sum_{t=1}^T (Y_t - \hat{r}(x)) K_h(x - X_t) = 0,$$

由此得到

$$\hat{r}(x) = \frac{\sum Y_t K_h(x - X_t)}{\sum K_h(x - X_t)} = \frac{\hat{m}(x)}{\hat{g}(x)}.$$

- 即：Nadaraya-Watson 是局部常数回归 (local constant)，在每个小邻域内假设 $r(z) \approx \text{常数}$ 。

局部多项式：基本思想

- 在估计点 x 的邻域中，用 p 阶多项式逼近 $r(z)$ ：

$$r(z) \approx \sum_{j=0}^p \alpha_j(x)(z-x)^j,$$

其中 $\alpha_j(x) = \frac{1}{j!} r^{(j)}(x)$ 。

- 更形式化：在 x 的小邻域内进行泰勒展开：

$$r(z) = \sum_{j=0}^p \frac{1}{j!} r^{(j)}(x)(z-x)^j + \text{高阶余项}.$$

- 只要局部样本数足够，可以通过局部加权最小二乘估计 $\alpha_j(x)$ ；
- Nadaraya-Watson = 局部多项式 ($p = 0$)；**
 - $p = 0$: 局部常数；
 - $p = 1$: 局部线性；
 - $p \geq 2$: 局部二次、三次等。

局部多项式的加权最小二乘形式

给定估计点 x , 定义

$$Z_t(x) = [1, (X_t - x), \dots, (X_t - x)^p]', \quad W_t(x) = K_h(x - X_t).$$

局部加权残差平方和:

$$\sum_{t=1}^T (Y_t - \alpha' Z_t(x))^2 W_t(x) = (Y - Z\alpha)' W (Y - Z\alpha),$$

其中

- $Y = (Y_1, \dots, Y_T)'$;
- Z 为 $T \times (p + 1)$ 矩阵, 行向量为 $Z_t'(x)$;
- $W = \text{diag}(W_1(x), \dots, W_T(x))$.

一阶条件:

$$\sum_{t=1}^T Z_t W_t (Y_t - Z_t' \hat{\alpha}) = 0,$$

即

$$\hat{\alpha}(x) = (Z' W Z)^{-1} Z' W Y.$$

局部多项式估计：函数值与导数

- 局部多项式拟合：

$$\hat{r}(z) = \sum_{j=0}^p \hat{\alpha}_j(z)(z-x)^j,$$

对所有 z 在 x 的邻域内；

- 点 x 处的函数估计：

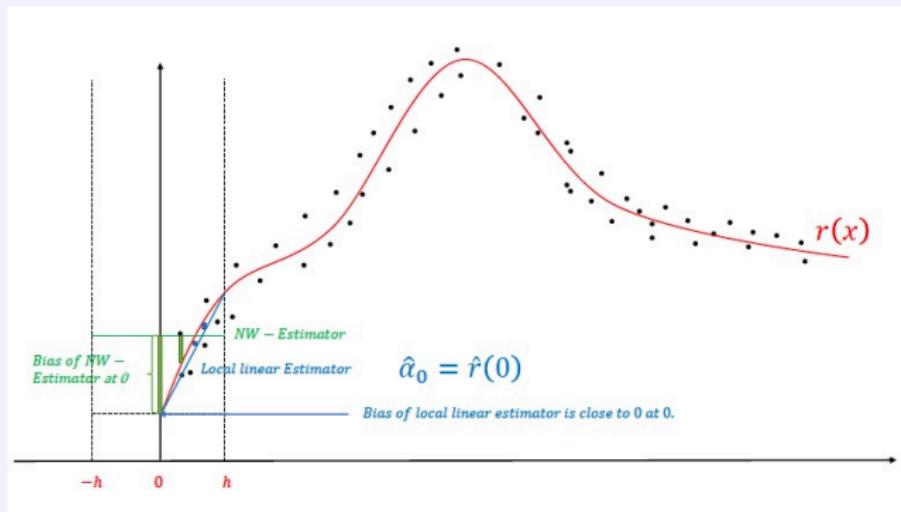
$$\hat{r}(x) = \sum_{j=0}^p \hat{\alpha}_j(x)(x-x)^j = \hat{\alpha}_0(x).$$

- 第 ν 阶导数在 x 的估计：

$$\hat{r}^{(\nu)}(x) = \nu! \hat{\alpha}_\nu(x), \quad \nu = 1, \dots, p.$$

- 因此，局部多项式平滑可以同时估计 $r(x)$ 及其若干阶导数。

局部线性 vs Nadaraya-Watson：偏差比较



图：Nadaraya-Watson 与局部线性估计的比较

- 局部线性估计 ($p = 1$) 在内部与边界的偏差普遍更小；
- Nadaraya-Watson 在边界处存在明显“向内拉”的偏差；
- 局部多项式（特别是局部线性）能显著缓解边界问题。

局部多项式的实践设置

- 需要选择的要素：
 - 多项式阶数 p (常用 $p = 1$ 或 2)；
 - 带宽 h (可用交叉验证、plug-in 等数据驱动方法)；
 - 核函数 $K(\cdot)$ (常用 Epanechnikov、高斯等)。
- 当 $p = 1$ 时：
 - 得到局部线性平滑 (local linear smoothing)；
 - 在理论上等价于某种加权形式的 Nadaraya-Watson，且假设 $g(x)$ 已知时具有更好的性质。
- 在 R 中：
 - 将前面 `locpoly(..., degree = 0)` 改为 `degree = 1` 或 `2` 即可使用局部线性/二次回归。

金融中的应用与小结

- 局部多项式在金融中的典型应用：
 - 估计时变风险溢价或利率曲线；
 - 非线性因子模型中因子暴露的估计；
 - 对资产收益-协变量关系的非参数刻画。
- 相比 Nadaraya-Watson：
 - 拟合更灵活、偏差更小；
 - 边界表现更好；
 - 可自然估计导数（如斜率、曲率）。
- 总结：
 - Nadaraya-Watson：局部常数，是最简单的核回归；
 - 局部多项式：在其基础上进行局部线性 / 二次扩展；
 - 带宽选择依然是关键，核函数的影响相对次要。

目录

1 非参数概率密度估计量

2 多维核密度估计

3 非参数回归模型

- Nadaraya-Watson 估计量
- 局部多项式估计量

4 参考文献

参考文献 I

-  Aït-Sahalia, Yacine (1996). "Testing continuous-time models of the spot interest rate". In: *The Review of Financial Studies* 9.2, pp. 385–426.
-  Chen, Bin and Yongmiao Hong (2012). "Testing for smooth structural changes in time series models via nonparametric regression". In: *Econometrica* 80.3, pp. 1157–1183.
-  Hart, Jeffrey D (1996). "Some automated methods of smoothing time-dependent data". In: *Journal of Nonparametric Statistics* 6.2-3, pp. 115–142.
-  Hong, Yongmiao and Haitao Li (2005). "Nonparametric specification testing for continuous-time models with applications to term structure of interest rates". In: *The Review of Financial Studies* 18.1, pp. 37–84.

参考文献 II

-  Tukey, John W (1961). "Curves as parameters, and touch estimation". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Vol. 4. University of California Press, pp. 681–695.
-  Zhou, Bin (1996). "High-frequency data and volatility in foreign-exchange rates". In: *Journal of Business & Economic Statistics* 14.1, pp. 45–52.

第 8 章金融资产定价模型

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

- 主要资产定价理论：
 - 资本资产定价模型 (CAPM)
 - 套利定价理论 (APT)
 - 基于消费的 CAPM (C-CAPM)
- 将这些理论框架转化为可估计的金融计量模型
- 利用市场数据对模型进行估计与检验
- 为资产定价、投资组合管理等实际问题提供计量工具

从理论模型到计量经济模型：一般步骤

① 模型规范化

- 明确关键经济关系的数学形式（如 CAPM 中的风险-收益关系）
- 为理论模型赋予时间序列结构（将单期 CAPM 拓展为含时间下标的模型）

② 数据收集与处理

- 收集价格、收益率、市场指数、利率等
- 数据清洗、去极端值、对齐频率等

③ 参数估计

- 最小二乘、极大似然、GMM 等

④ 模型检验

- 拟合优度、参数显著性检验、稳健性检验

⑤ 模型应用

- 预期收益、资本成本、事件研究、绩效评估等

目录

1 资本资产定价模型

- CAPM 的检验: Wald 与 LR
- 时变 CAPM 与条件 CAPM

2 多因子模型

- 案例: Fama–French 三因子模型
- SMB/HML/UMD 的构建
- Carhart 四因子与 Fama–French 五因子
- 实践口径与中国市场提示

3 基于消费的资本资产定价模型

- 广义矩估计 (GMM)
- 采用 GMM 估计 C-CAPM 模型
- 时变风险厌恶与股权溢价之谜

4 章节总结

5 参考文献

多资产 CAPM：基本设定

- 令 Z_t 为 $N \times 1$ 的超额收益向量 (N 种风险资产)：

$$Z_t = \alpha + \beta Z_{mt} + \varepsilon_t,$$

- 其中

$$E[\varepsilon_t] = 0, \quad E[\varepsilon_t \varepsilon_t'] = \Sigma,$$

$$E[Z_{mt}] = \mu_m, \quad E[(Z_{mt} - \mu_m)^2] = \sigma_m^2,$$

$$\text{Cov}(Z_{mt}, \varepsilon_t) = 0.$$

- 记各资产期望超额收益向量为 $\mu = E[Z_t]$ ，则

$$\mu = \alpha + \beta \mu_m.$$

多资产 CAPM 的极大似然估计

- 样本均值：

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T Z_t, \quad \hat{\mu}_m = \frac{1}{T} \sum_{t=1}^T Z_{mt}.$$

- 贝塔向量估计量：

$$\hat{\beta} = \frac{\sum_{t=1}^T (Z_t - \hat{\mu})(Z_{mt} - \hat{\mu}_m)}{\sum_{t=1}^T (Z_{mt} - \hat{\mu}_m)^2}.$$

- 截距向量：

$$\hat{\alpha} = \hat{\mu} - \hat{\beta} \hat{\mu}_m.$$

- 协方差矩阵：

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt})(Z_t - \hat{\alpha} - \hat{\beta} Z_{mt})'.$$

MLE 的渐近分布

- 记

$$\hat{\sigma}_m^2 = \frac{1}{T} \sum_{t=1}^T (Z_{mt} - \hat{\mu}_m)^2.$$

- 有

$$\hat{\alpha} \sim N\left(\alpha, \frac{1}{T} \left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right] \Sigma\right),$$

$$\hat{\beta} \sim N\left(\beta, \frac{1}{T} \frac{1}{\hat{\sigma}_m^2} \Sigma\right),$$

$$T\hat{\Sigma} \sim \mathcal{W}_N(T-2, \Sigma).$$

- 协方差:

$$\text{Cov}(\hat{\alpha}, \hat{\beta}') = -\frac{1}{T} \frac{\hat{\mu}_m}{\hat{\sigma}_m^2} \Sigma.$$

- 且 $\hat{\Sigma}$ 、 $\hat{\alpha}$ 与 $\hat{\beta}$ 彼此独立。

威沙特分布简介

设 X 为 $n \times p$ 矩阵，每行

$$X'_{(i)} \sim N_p(\mathbf{0}, \mathbf{V}), \quad i = 1, \dots, n,$$

独立同分布。定义

$$\mathbf{S} = X'X = \sum_{i=1}^n X'_{(i)} X_{(i)},$$

则有

$$\mathbf{S} \sim W_p(\mathbf{V}, n),$$

称为 p 维威沙特分布。



John Wishart

- 当 $p = 1$ 且 $\mathbf{V} = 1$ 时退化为 χ_n^2 .
- 期望: $E[\mathbf{S}] = n\mathbf{V}$ (若 $n \geq p + 1$).

检验思路：超额收益截距是否为零

- 若 CAPM 正确，则每个资产的“异常收益”应为零：

$$H_0 : \alpha = 0, \quad H_1 : \alpha \neq 0.$$

- 检验方法：
 - 沃尔德 (Wald) 检验
 - 似然比 (Likelihood Ratio, LR) 检验
 - 拉格朗日乘数 (Lagrange Multiplier, LM/Score) 检验 (作比较)
- 三种检验在大样本下渐近等价，但实现思路和计算量不同。

Wald 检验

- Wald 统计量:

$$J_0 = \hat{\alpha}' [\text{Var}(\hat{\alpha})]^{-1} \hat{\alpha} = T \left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}.$$

- 在 H_0 下,

$$J_0 \stackrel{a}{\sim} \chi_N^2.$$

- 小样本修正:

$$J_1 = \frac{T - N - 1}{N} \left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F(N, T - N - 1).$$

- 其中 N 为资产个数 (同时也是约束数)。

约束 CAPM 模型与 LR 检验

约束模型令 $\alpha = 0$:

$$Z_t = \beta Z_{mt} + \varepsilon_t.$$

极大似然估计量:

$$\hat{\beta}^* = \frac{\sum_{t=1}^T Z_t Z_{mt}}{\sum_{t=1}^T Z_{mt}^2},$$

$$\hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^T (Z_t - \hat{\beta}^* Z_{mt})(Z_t - \hat{\beta}^* Z_{mt})'.$$

在 H_0 下,

$$\hat{\beta}^* \sim N\left(\beta, \frac{1}{T} \frac{1}{\hat{\mu}_m^2 + \hat{\sigma}_m^2} \Sigma\right),$$

$$T \hat{\Sigma}^* \sim \mathcal{W}_N(T-1, \Sigma).$$

LR 统计量与其渐近分布

- 记

$$\Lambda = \frac{L^*}{L}, \quad \ln \Lambda = \mathcal{L}^* - \mathcal{L} = -\frac{T}{2} \left[\ln |\widehat{\Sigma}| - \ln |\widehat{\Sigma}^*| \right].$$

- LR 统计量:

$$J_2 = -2 \ln \Lambda = T \left[\ln |\widehat{\Sigma}^*| - \ln |\widehat{\Sigma}| \right] \stackrel{a}{\sim} \chi_N^2.$$

- 约束模型与非约束模型估计量间的关系:

$$\widehat{\beta}^* = \widehat{\beta} + \frac{\widehat{\mu}_m}{\widehat{\mu}_m^2 + \widehat{\sigma}_m^2} \widehat{\alpha},$$

$$\widehat{\Sigma}^* = \widehat{\Sigma} + \left(\frac{\widehat{\sigma}_m^2}{\widehat{\mu}_m^2 + \widehat{\sigma}_m^2} \right) \widehat{\alpha} \widehat{\alpha}'.$$

LR 统计量的替代表达与小样本修正

- 行列式关系:

$$|\widehat{\Sigma}^*| = |\widehat{\Sigma}| \left[\left(\frac{\widehat{\sigma}_m^2}{\widehat{\mu}_m^2 + \widehat{\sigma}_m^2} \right) \widehat{\alpha}' \widehat{\Sigma}^{-1} \widehat{\alpha} + 1 \right].$$

- 似然比可写为

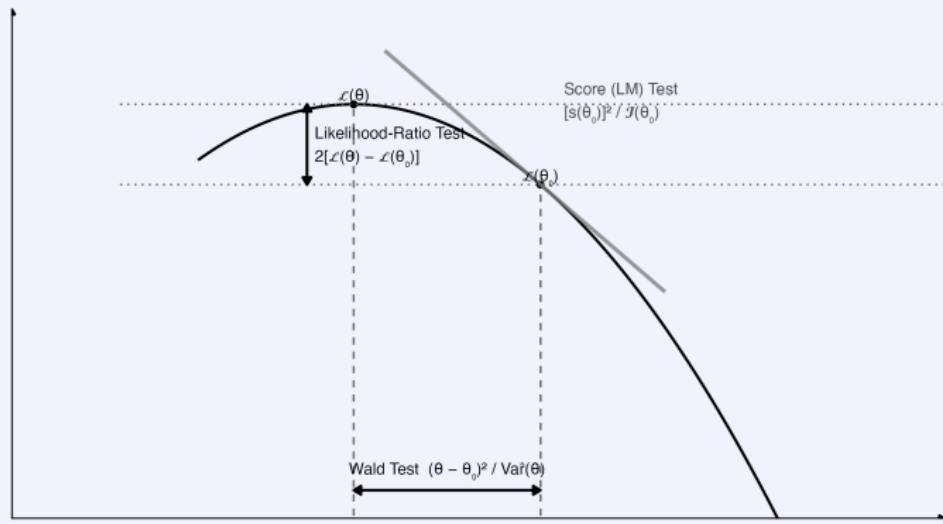
$$\mathcal{LR} = -\frac{T}{2} \ln \left[\left(\frac{\widehat{\sigma}_m^2}{\widehat{\mu}_m^2 + \widehat{\sigma}_m^2} \right) \widehat{\alpha}' \widehat{\Sigma}^{-1} \widehat{\alpha} + 1 \right].$$

- 小样本修正的一种形式:

$$J_3 = \left(T - \frac{N}{2} - 2 \right) \left[\ln |\widehat{\Sigma}^*| - \ln |\widehat{\Sigma}| \right] \implies \chi_N^2.$$

- 当 $T \rightarrow \infty$ 时, J_3 的分布趋近于自由度为 N 的卡方分布。

Wald –LR –LM 检验的比较 (直观图示)



- Wald: 水平距离——参数估计值与约束值之间的差异
- LR: 纵向高度——受限与非受限最大对数似然之差
- LM/Score: 斜率——在受限点处的得分向量是否为零

时变 CAPM：基本形式

对于资产 $i = 1, \dots, n$:

$$Z_{it} = \alpha_i(\mathcal{F}_{t-1}) + \beta_i(\mathcal{F}_{t-1})' Z_{mt} + \varepsilon_{it}, \quad t = 1, \dots, T,$$

$$\mathbb{E}(\varepsilon_{it} | \mathcal{F}_{t-1}) = 0.$$

- Z_{it} : 资产 i 在 t 的超额收益率
- Z_{mt} : 市场组合在 t 的超额收益率
- $\alpha_i(\mathcal{F}_{t-1})$ 、 $\beta_i(\mathcal{F}_{t-1})$: 依赖于信息集 \mathcal{F}_{t-1} 的未知函数系数
- 可写为时变参数:

$$\alpha_{it} = \alpha_i(\mathcal{F}_{t-1}), \quad \beta_{it} = \beta_i(\mathcal{F}_{t-1}).$$

时变 CAPM：状态变量表示与估计

- 设存在状态变量向量 S_t , 使得

$$\alpha_{it} = \alpha_i(S_t), \quad \beta_{it} = \beta_i(S_t),$$

但 $\alpha_i(\cdot)$ 、 $\beta_i(\cdot)$ 的具体函数形式未知。

- 可通过局部加权最小二乘估计：

$$\min_{\{\alpha_i(\cdot), \beta_i(\cdot)\}} \sum_{i=1}^n \sum_{t=1}^T [Z_{it} - \alpha_i(S_t) - \beta_i(S_t)' Z_{mt}]^2 K_h(S_t - z),$$

其中 $K_h(\cdot)$ 为核函数, h 为带宽, z 为评估点。

- 这是函数系数模型思想在 CAPM 中的应用。

条件 CAPM：设定

- 多种风险资产 $i = 1, \dots, n$, 一无风险资产
- 定义时点 t 的超额收益向量：

$$Z_t = (Z_{1t}, \dots, Z_{nt})'$$

- 市场组合超额收益：

$$Z_{mt} = Z_t' W_t,$$

其中 W_t 为 CAPM 均衡下的权重向量，可依赖于 \mathcal{F}_{t-1} 。

- 条件均值向量：

$$\mu_t = (\mu_{1t}, \dots, \mu_{nt})', \quad \mu_{it} = E(Z_{it} | \mathcal{F}_{t-1}).$$

- 条件协方差矩阵：

$$H_t = E[(Z_t - \mu_t)(Z_t - \mu_t)' | \mathcal{F}_{t-1}].$$

条件 CAPM：模型形式

- 条件市场超额收益：

$$\mu_{mt} = \mathbb{E}(Z_{mt} | \mathcal{F}_{t-1}) = \mu_t' W_t,$$

$$\text{Var}(Z_{mt} | \mathcal{F}_{t-1}) = W_t H_t W_t.$$

- 条件 CAPM 的矩形式：

$$\mu_t = \beta_t \mu_{mt} = \frac{\mu_{mt}}{W_t H_t W_t} H_t W_t,$$

其中

$$\beta_t = \frac{H_t W_t}{W_t H_t W_t}$$

为时变市场贝塔。

- 回归形式：

$$Z_t = \beta_t Z_{mt} + \varepsilon_t,$$

$$\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0, \quad \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = H_t.$$

目录

1 资本资产定价模型

- CAPM 的检验: Wald 与 LR
- 时变 CAPM 与条件 CAPM

2 多因子模型

- 案例: Fama–French 三因子模型
- SMB/HML/UMD 的构建
- Carhart 四因子与 Fama–French 五因子
- 实践口径与中国市场提示

3 基于消费的资本资产定价模型

- 广义矩估计 (GMM)
- 采用 GMM 估计 C-CAPM 模型
- 时变风险厌恶与股权溢价之谜

4 章节总结

5 参考文献

多因子定价模型的理论基础

- 多因子定价模型的理论基础主要有两类：
 - 套利定价理论 (Arbitrage Pricing Theory, APT)(Ross 1976);
 - 跨期资本资产定价模型 (Intertemporal CAPM, ICAPM)(Merton 1973);
- **APT**: 基于无套利论证
 - 资产预期收益由若干宏观因子或市场指标解释
 - 若价格偏离因子所隐含的均衡水平，则存在套利机会
- **ICAPM**: 基于跨期均衡
 - 投资者权衡当期风险收益与未来投资机会集
 - 是对传统单期 CAPM 的跨期扩展

APT 与“相对定价”视角

- 学界与教材中，更普遍将无套利视为多因子模型的主要理论基石
 - 对 ICAPM 的系统讨论相对较少
- APT 在金融实务（特别是投资银行定价）中应用广泛
- 与均衡定价模型（如 CAPM、C-CAPM）相比：
 - CAPM/C-CAPM：强调均衡预期收益—风险关系
 - APT：更多是相对定价理论
 - 在已知部分资产价格前提下，对其他相关资产定价

CAPM 与 APT：单因子 vs 多因子

- CAPM：单因子框架
 - 市场组合是唯一系统性风险源
 - 贝塔系数度量资产相对市场的系统性风险
- APT：多因子框架
 - 预期收益由多种宏观经济/金融因子共同驱动
 - 可包括通胀、利率、产出、信用、流动性等因子
- 一般性关系：
 - CAPM 是一种特殊情形（单一市场因子）
 - APT 允许多个风险源，且不要求识别“真正的”市场组合

传统文化视角：分类思想与因子模型

- 《周易》“方以类聚，物以群分”体现了深刻的分类思想
- 现代金融中的因子模型本质上也是在“类聚”：
 - 按市值大小：大盘股、小盘股
 - 按行业：周期类、防御类、成长类等
 - 按财务特征：价值股、成长股、高盈利股、低盈利股
- APT 将这些“共性特征”抽象为因子：
 - 如通胀、利率、规模、价值等
 - 分类有助于更好地进行组合构建与风险管理

APT 的基本含义

- 根据 Ross (1976), APT 意味着

$$\mu \approx \iota\lambda_0 + B\lambda_K,$$

其中

- μ : $N \times 1$ 预期回报向量
 - ι : 所有元素为 1 的 $N \times 1$ 向量
 - λ_0 : 零贝塔参数 (有无风险资产时等于无风险收益率)
 - λ_K : $K \times 1$ 因子风险溢价向量
 - B : $N \times K$ 因子载荷矩阵
- 实务中通常将近似误差视为可忽略:

$$\mu = \iota\lambda_0 + B\lambda_K.$$

K 因子回归模型（实证形式）

- 对单个资产 i :

$$Z_{it} = \alpha_i + \sum_{j=1}^K b_{ij} f_{jt} + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

- 矩阵形式:

$$Z_t = \alpha + B Z_{Kt} + \varepsilon_t,$$

- Z_t : $N \times 1$ 回报或超额回报向量
- Z_{Kt} : $K \times 1$ 因子超额回报向量
- $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$

- 条件矩:

$$E(\varepsilon_t | Z_{K1}, \dots, Z_{KT}) = 0,$$

$$E(\varepsilon_t \varepsilon_t' | Z_{K1}, \dots, Z_{KT}) = \Sigma_\varepsilon.$$

误差结构与 i.i.d. 假设

- 条件期望为零：

$$E(\varepsilon_t | Z_{K1}, \dots, Z_{KT}) = 0$$

- 表明在控制因子后，残差不再有系统性方向偏离

- 条件协方差稳定：

$$E(\varepsilon_t \varepsilon_t' | Z_{K1}, \dots, Z_{KT}) = \Sigma_\varepsilon$$

- 残差相关结构在时间上相对稳定

- 常用简化假设：

$$\varepsilon_t \sim \text{i.i.d. } N(0, \Sigma_\varepsilon),$$

便于采用经典统计推断（MLE、Wald/LR/LM 检验等）。

常见因子类别：概览

① 宏观经济因子

② 市场因子

③ 公司特征因子

④ 统计因子

- 不同类别因子可结合使用，构成多因子定价或风格模型
- 因子选择与构造是实证研究的关键环节

宏观与市场因子

1. 宏观经济因子

- 利率因子：利率变动影响贴现率与资产价格
- 通货膨胀因子：影响资产的**实际收益率**
- 经济增长因子：如 GDP 增长率，影响盈利与风险偏好

2. 市场因子

- 市场组合收益率：如沪深 300、标普 500 等
- 行业组合收益率：反映行业层面的系统性风险
- 市场风险溢价：市场组合收益率减无风险利率

公司特征因子与统计因子

3. 公司特征因子

- 规模因子 (SMB): 小市值组合减大市值组合
- 价值因子 (HML): 高账面市值比组合减低账面市值比组合
- 盈利能力因子: ROE/ROA 等指标构造的组合收益差

4. 统计因子

- 主成分因子: 由 PCA 从收益矩阵中提取
- 波动性/相关性因子: 基于收益波动和共动特征构造

多因子模型与 SURE 结构

- 假设存在无风险资产，因子收益 F 可观测
- 得到多元高斯似不相关回归 (SURE) 模型：

$$Z_t = \alpha + BZ_{Kt} + \varepsilon_t.$$

- 在给定 Z_{K1}, \dots, Z_{KT} 条件下，对数似然为

$$\ell(\alpha, B, \Sigma_\varepsilon) = c - \frac{T}{2} \log \det(\Sigma_\varepsilon) - \frac{1}{2} \sum_{t=1}^T (Z_t - \alpha - BZ_{Kt})' \Sigma_\varepsilon^{-1} (Z_t - \alpha - BZ_{Kt}).$$

无约束 MLE 估计量

- 定义样本矩:

$$\hat{\mu}_K = \frac{1}{T} \sum_{t=1}^T Z_{Kt}, \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^T Z_t,$$

$$\hat{\Sigma}_K = \frac{1}{T} \sum_{t=1}^T (Z_{Kt} - \hat{\mu}_K)(Z_{Kt} - \hat{\mu}_K)',$$

$$\hat{\Sigma}_{ZK} = \frac{1}{T} \sum_{t=1}^T (Z_t - \hat{\mu})(Z_{Kt} - \hat{\mu}_K)'.$$

- 极大似然估计 (在正态假设下与逐方程 OLS 等价):

$$\hat{\alpha} = \hat{\mu} - \hat{B}\hat{\mu}_K, \quad \hat{B} = \hat{\Sigma}_{ZK}\hat{\Sigma}_K^{-1},$$

$$\hat{\Sigma}_\epsilon = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t', \quad \hat{\epsilon}_t = Z_t - \hat{\alpha} - \hat{B}Z_{Kt}.$$

$\alpha = 0$ 的检验与 F 统计量

- 若模型正确，则无套利条件意味着 $\alpha = 0$ 。
- 在正态假设下（给定因子）：

$$\hat{\alpha} \sim N\left(\alpha, \frac{1}{T} \left(1 + \hat{\mu}'_K \hat{\Sigma}_K^{-1} \hat{\mu}_K\right) \Sigma_{\varepsilon}\right).$$

- 原假设：

$$H_0 : \alpha = 0.$$

- Wald/F 检验统计量：

$$F = \frac{T - N - K}{N} \left(1 + \hat{\mu}'_K \hat{\Sigma}_K^{-1} \hat{\mu}_K\right)^{-1} \hat{\alpha}' \hat{\Sigma}_{\varepsilon}^{-1} \hat{\alpha}.$$

- 在 H_0 且正态假设下，

$$F \sim F_{N, T - N - K}.$$

矩阵化表达与其他约束检验

- 逐方程收集参数：

$$\Theta = (\theta_1, \dots, \theta_N)' \in \mathbb{R}^{N \times (K+1)},$$

$$\theta_i = (\alpha_i, b_i')',$$

其中 b_i 为 B 的第 i 行。

- 令 X 为 $T \times (K+1)$ 设计矩阵（截距 + 因子观测）：

$$\hat{\Theta} = Z' X (X' X)^{-1}.$$

- 除了检验 $\alpha = 0$ 外，还可检验 $B_2 = 0$ ：

- 某些因子的载荷是否全部为零
- 通过受限与非受限模型的似然比统计量：

$$J_1 = T \left(\log |\tilde{\Sigma}_\varepsilon| - \log |\hat{\Sigma}_\varepsilon| \right) \approx \chi^2_{K_2}.$$

可交易因子 + 宏观因子的设定

- 假设存在无风险利率 R_{ft} , 因子分为两类:

- 可交易投资组合因子 $f_{1t} \in \mathbb{R}^{K_1}$
- 不可交易宏观因子 $f_{2t} \in \mathbb{R}^{K_2}$

- 无约束模型:

$$R_t - R_{ft} = \alpha + B_1(f_{1t} - R_{ft}) + B_2 f_{2t} + \varepsilon_t,$$

$$E(f_{2t}) = \mu_{K2}.$$

- 与 APT 一致的原假设:

$$H_0 : \alpha = B_2 \gamma_2, \quad \gamma_2 \in \mathbb{R}^{K_2}.$$

约束重写与 LR 检验

- 若 B_2 满秩：

$$\gamma_2 = (B_2' B_2)^{-1} B_2' \alpha.$$

从而约束可写为

$$M_{B_2} \alpha = 0,$$

其中

$$M_{B_2} = I_{K_2} - B_2 (B_2' B_2)^{-1} B_2'$$

为对称幂等矩阵，秩为 $N - K_2$ 。

- 在给定 γ_2 时：

$$R_t - R_{ft} = B_1(f_{1t} - R_{ft}) + B_2(f_{2t} + \gamma_2) + \varepsilon_t.$$

- 通过对 γ_2 的搜索构造受限 MLE，并利用 LR 统计量：

$$LR = T \left(\log |\tilde{\Omega}_\varepsilon| - \log |\hat{\Omega}_\varepsilon| \right) \implies \chi^2_{N-K_2}.$$

Wald 类检验 J_2 (Chamberlain 思路)

- 定义

$$\begin{aligned}\widehat{\theta}_i &= (\alpha_i, b'_{2i})', \quad \widehat{\theta} = (\widehat{\theta}'_1, \dots, \widehat{\theta}'_N)' \in \mathbb{R}^{(K_2+1)N}, \\ \delta &= (\gamma'_2, b'_{21}, \dots, b'_{2N})'.\end{aligned}$$

- 设

$$h(\delta) = (b'_{21}\gamma_2, b'_{21}, \dots, b'_{2N}\gamma_2, b'_{2N})'.$$

- Wald 类统计量:

$$J_2 = T \min_{\delta \in \mathbb{R}^{K_2(1+N)}} (\widehat{\theta} - h(\delta))' \widehat{\Xi}^{-1} (\widehat{\theta} - h(\delta)),$$

其中

$$\widehat{\Xi} = \widehat{\Omega}_\varepsilon \otimes (X'X)^{-1}.$$

- Chamberlain (1984):

$$J_2 \implies \chi^2_{N-K_2} \quad \text{当 } T \rightarrow \infty.$$

Fama–French 三因子模型：基本形式

- Fama and French (1993) 提出的三因子模型：

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_i \text{SMB}_t + h_i \text{HML}_t + \varepsilon_{it}.$$

- 其中

- R_{it} : 个股 i 收益率
- R_{ft} : 无风险收益率
- R_{mt} : 市场组合收益率
- SMB: Small Minus Big (规模因子)
- HML: High Minus Low (价值因子)

- 与 CAPM 相比：

- 引入规模与价值因子，提高解释力
- 广泛用于绩效评估与资产定价实证研究

R 代码流程概览 (Fama–French 三因子)

① 数据获取

- 使用 frenchdata 包下载 Fama/French 三因子日度数据
- 使用 quantmod::getSymbols 下载个股价格

② 数据预处理

- 计算个股日收益率 (dailyReturn)
- 将因子数据与个股收益按日期合并

③ 回归与结果输出

- 逐股票回归：

`daily.returns ~ mkt_excess + smb + hml`

- 保存合并数据与回归结果 (CSV 与 txt)

回归结果示例（某能源股）

Call:

```
lm(formula = daily.returns ~ mkt_excess + smb + hml, data = merged_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.165796	-0.012465	-0.000798	0.011249	0.165252

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.001039	0.000476	2.183	0.0291 *
mkt_excess	0.931388	0.043636	21.344	<2e-16 ***
smb	0.077768	0.079293	0.981	0.3268
hml	-0.035087	0.056460	-0.621	0.5344

Residual standard error: 0.02492 on 2744 degrees of freedom
Multiple R-squared: 0.1513, Adjusted R-squared: 0.1504
F-statistic: 163.1 on 3 and 2744 DF, p-value: < 2.2e-16

- 截距显著 \Rightarrow 存在异常收益或模型设定不完备
- 市场因子显著；SMB/HML 在该样本上不显著

SMB 因子构建思路

SMB (Small Minus Big) 因子

- ① 数据准备：获取股票市值（MV）与收益率，设定形成期与持有期。
- ② 分组排序：在形成时点按市值从小到大排序，将样本划为“小盘”（Small）与“大盘”（Big）。
- ③ 组合收益：计算 Small 与 Big 组合的下期收益（等权或市值加权）。
- ④ 因子值：

$$\text{SMB}_t = R_{\text{Small}, t} - R_{\text{Big}, t}.$$

HTML 因子构建思路

HTML (High Minus Low) 因子

- ① 数据准备：计算账面市值比 ($BM = \text{Book-to-Market}$) 或 PB 的倒数，并获取收益率。
- ② 分组排序：按 BM 排序，取最高 30% 为 High，最低 30% 为 Low。
- ③ 组合收益：计算 High/Low 组合下期收益。
- ④ 因子值：

$$\text{HML}_t = R_{\text{High BM}, t} - R_{\text{Low BM}, t}$$

UMD 因子构建思路

UMD (动量因子)

- ① 数据准备：基于历史价格计算收益率，设定排序期与持有期。
- ② 形成组合：按排序期内累计收益率排序，前 30% 为赢家组合 (Winners)，后 30% 为输家组合 (Losers)。
- ③ 组合收益：计算持有期内赢家与输家组合收益。
- ④ 因子值：

$$\text{UMD}_t = R_{\text{Winners}, t} - R_{\text{Losers}, t}.$$

Carhart 四因子模型

- 在 Fama–French 三因子基础上加入动量因子 UMD:

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_i\text{SMB}_t + h_i\text{HML}_t + u_i\text{UMD}_t + \varepsilon_{it}.$$

- Carhart (1997):
 - UMD (Up Minus Down) 刻画赢家减输家组合收益差
 - 能更好刻画动量效应与组合绩效持续性
- 实务中广泛用于业绩归因与主动管理能力分析。

Fama-French 五因子模型

- Fama and French (2015) 在三因子基础上加入 RMW 与 CMA:

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_i\text{SMB}_t + h_i\text{HML}_t + r_i\text{RMW}_t + c_i\text{CMA}_t$$

- RMW (Robust Minus Weak):
 - 高盈利组合减低盈利组合收益
- CMA (Conservative Minus Aggressive):
 - 保守投资组合减激进投资组合收益
- 通常在解释横截面收益差异方面优于三因子模型。

RMW 与 CMA 的构建要点

RMW (盈利能力因子)

- ① 用 ROE/ROA 等指标衡量盈利能力。
- ② 在形成时点按盈利能力排序：前 30% 为 High (Robust)，后 30% 为 Low (Weak)。
- ③ 计算高/低盈利组合下期收益：

$$\text{RMW}_t = R_{\text{robust},t} - R_{\text{weak},t}.$$

RMW 与 CMA 的构建要点

CMA (投资风格因子)

- ① 用资产增长率等指标衡量投资风格。
- ② 按资产增长率排序：增长率低的前 30% 为 Conservative；高的后 30% 为 Aggressive。
- ③ 定义

$$\text{CMA}_t = R_{\text{conservative},t} - R_{\text{aggressive},t}.$$

实践提示与口径说明（以中国市场为例）

- 形成期与持有期：动量类因子常采用“跳月”(skip-month)。
- 加权方式：等权 vs 市值加权（需明确总市值/流通市值）。
- 分组断点：可按全市场分位点（如 30–40–30）构造因子。
- 收益率口径：前/后复权价格，简单收益率 vs 对数收益率。
- 数据清洗与调整：
 - 处理停牌、涨跌停、新股、ST 等情况
 - 规范除权除息处理
- 极端值与中性化：
 - 1% 分位截断或温莎化
 - 考虑行业中性、市值中性处理
- 无风险利率：明确选用的基准（如中债短期国债收益率、回购利率等）。

目录

1 资本资产定价模型

- CAPM 的检验: Wald 与 LR
- 时变 CAPM 与条件 CAPM

2 多因子模型

- 案例: Fama–French 三因子模型
- SMB/HML/UMD 的构建
- Carhart 四因子与 Fama–French 五因子
- 实践口径与中国市场提示

3 基于消费的资本资产定价模型

- 广义矩估计 (GMM)
- 采用 GMM 估计 C-CAPM 模型
- 时变风险厌恶与股权溢价之谜

4 章节总结

5 参考文献

C-CAPM 的动机与局限性的来源

- 传统 CAPM 与 APT：
 - 搭建了资产定价的基本框架
 - 但忽视了消费决策在资产定价中的核心作用
- 传统模型特点：
 - 主要基于单期财富偏好，投资组合选择在单一时期完成
 - 投资与消费被割裂，难以刻画跨期决策
- 现实中：
 - 投资者在多个时期进行消费与投资组合决策
 - 消费与资产收益紧密相连
 - 需要在跨期框架下联合建模消费与投资

传统模型在动态资产收益解释上的不足

- 真正的系统性风险源自**宏观经济**:
 - 核心是消费的跨期边际替代率
 - 资产价格应当由这些宏观风险内生决定
- 传统 CAPM/APT 的局限:
 - CAPM 中无风险利率、零贝塔收益、市场风险补偿均为外生参数
 - APT 将“单一市场风险价格”扩展为“因子风险价格向量”，但风险价格仍在模型之外
- 现实中的事实:
 - 无风险实际利率与股票超额收益率是**随时间变化的**
 - 传统模型难以合理刻画收益的可预测性与动态行为

C-CAPM 的视角与贡献

- 基于消费的资本资产定价模型 (C-CAPM):
 - 直接将消费纳入资产定价的核心
 - 将系统性风险与消费边际替代率联系起来
- 优点:
 - 在宏观一般均衡框架下刻画资产价格
 - 更清晰地解释风险回报的决定机制
 - 有助于理解资产收益率的动态变化与可预测性
- 本章定位:
 - 第 ?? 节已介绍 C-CAPM 理论基础
 - 本节聚焦: C-CAPM 的广义矩估计 (GMM)方法

GMM 的基本框架：矩条件

- 设经济模型蕴含一组 r 个总体矩约束：

$$E\left\{\underbrace{\mathbf{h}(\theta, \mathbf{w}_t)}_{(r \times 1)}\right\} = 0, \quad (1)$$

其中

- \mathbf{w}_t : t 期已知的 $h \times 1$ 向量
 - θ : $a \times 1$ 未知参数向量
- 核心思想：
 - 选取 θ 使样本矩“尽量接近”总体矩
 - 即样本矩尽量接近 0

样本矩与 GMM 目标函数

- 样本矩定义为

$$\underbrace{\mathbf{g}(\theta; \mathbf{y}_T)}_{(r \times 1)} = \frac{1}{T} \sum_{t=1}^T \mathbf{h}(\theta, \mathbf{w}_t),$$

其中

$$\mathbf{y}_T = (\mathbf{w}'_T, \mathbf{w}'_{T-1}, \dots, \mathbf{w}'_1)' \in \mathbb{R}^{Th}.$$

- GMM 估计量 $\hat{\theta}$ 令下式最小：

$$Q(\theta; \mathbf{y}_T) = \mathbf{g}(\theta; \mathbf{y}_T)' \mathbf{W}_T \mathbf{g}(\theta; \mathbf{y}_T), \quad (2)$$

其中 $\{\mathbf{W}_T\}$ 为 $r \times r$ 正定权重矩阵序列。

识别情形与 Hansen (1982) 的结果

- 当 $r = a$ (刚识别):
 - 可通过令每个样本矩 $\mathbf{g}(\theta; \mathbf{y}_T) = 0$ 来估计
- 当 $r > a$ (过度识别):
 - 不能同时满足所有矩条件
 - 采用式 (2) 的**加权二次型最小化**来估计
- Hansen (1982):
 - 在严格平稳和常规正则条件下, GMM 估计量 $\hat{\theta}$:
 - 一致
 - 收敛速度为 \sqrt{T}
 - 渐近正态

最优权重矩阵与迭代 GMM

- Hansen 证明：最优权重矩阵为

$$\mathbf{W}_T = \mathbf{S}^{-1},$$

其中

$$\mathbf{S} = \sum_{j=-\infty}^{\infty} E\{\mathbf{h}(\theta_0, \mathbf{w}_t) \mathbf{h}(\theta_0, \mathbf{w}_{t-j})'\}.$$

- 实务问题： \mathbf{S} 依赖未知真值 θ_0
- 解决思路：迭代 GMM
 - ① 初始设定权重矩阵（如 $\mathbf{W}_T = I$ ），得到初始估计 $\hat{\theta}^{(1)}$
 - ② 用 $\hat{\theta}^{(1)}$ 估计 $\hat{\mathbf{S}}_T^{(1)}$
 - ③ 令 $\mathbf{W}_T = \hat{\mathbf{S}}_T^{(1)-1}$ ，重新最小化 Q ，得到 $\hat{\theta}^{(2)}$
 - ④ 重复直至收敛或在两步 GMM 处停止（渐近分布相同）

过度识别约束检验 (OID 检验)

- 当 $r > a$ 时，可检验模型设定是否与矩条件一致
- Hansen 的 J 统计量：

$$J_T \equiv T Q(\hat{\theta}; \mathbf{y}_T) \Rightarrow \chi^2(r - a).$$

- 解读：
 - 若模型正确、矩条件成立，则在估计值处， $\mathbf{g}(\hat{\theta}; \mathbf{y}_T)$ 应接近 0
 - J_T 可视作“样本矩偏离零”的加权度量
- 计算便利：
 - 只需在 GMM 估计量 $\hat{\theta}$ 处，乘以样本量 T 即可

标准 C-CAPM：偏好与 Euler 方程

- Hansen (1982) 在提出 GMM 时，即用 C-CAPM 进行估计与检验
- 偏好设为幂效用：

$$\max_{C_t} E_t \left[\sum_{i=0}^{\infty} \beta^i u(C_{t+i}) \right],$$

$$u(C_t) = \begin{cases} \frac{C_t^{1-\gamma}}{1-\gamma}, & \gamma > 0, \\ \ln C_t, & \gamma = 1. \end{cases}$$

- 对可交易资产 $i = 1, \dots, N$ 的一阶条件：

$$C_t^{-\gamma} = \beta E_t \left[(1 + R_{i,t+1}) C_{t+1}^{-\gamma} \right], \quad i = 1, \dots, N.$$

矩条件与消费增长率

- 将 Euler 方程重写为零均值矩条件：

$$0 = E_t \left\{ 1 - \beta(1 + R_{i,t+1}) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right\}, \quad i = 1, \dots, N. \quad (3)$$

- 为满足平稳性要求：

- 采用消费增长率 (C_{t+1}/C_t) 而非水平 C_t , 因后者通常为明显非平稳序列

- 令待估参数：

$$\theta = (\beta, \gamma)'$$

- 对应的边际替代率 (随机贴现因子)：

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}.$$

条件矩与工具变量扩展

- 用 \mathcal{F}_t 表示投资者信息集：

$$0 = E \left\{ \left[1 - \beta(1 + R_{i,t+1}) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right] \middle| \mathcal{F}_t \right\}, \quad i = 1, \dots, N. \quad (4)$$

- 令可观测信息集子集 $\mathbf{x}_t \subseteq \mathcal{F}_t$, 则有无条件矩：

$$0 = E \left\{ \left[1 - \beta(1 + R_{i,t+1}) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right] \mathbf{x}_t \right\}, \quad i = 1, \dots, N.$$

- 若 \mathbf{x}_t 为 $M \times 1$ 向量, 则共有

$$r = N \cdot M$$

个矩条件, 可用于对模型进行估计与检验。

矩向量的构造与 GMM 目标

- 构造矩向量：

$$\mathbf{h}(\theta, \mathbf{w}_{t+1}) = \begin{bmatrix} [1 - \beta(1 + R_{1,t+1}) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}] \mathbf{x}_t \\ [1 - \beta(1 + R_{2,t+1}) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}] \mathbf{x}_t \\ \vdots \\ [1 - \beta(1 + R_{N,t+1}) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}] \mathbf{x}_t \end{bmatrix}. \quad (5)$$

- 对 (5) 取样本均值得到 $\mathbf{g}(\theta; \mathbf{y}_T)$
- Hansen 通过最小化

$$\min_{\theta} Q(\theta; \mathbf{y}_T) = \mathbf{g}(\theta; \mathbf{y}_T)' \hat{\mathbf{S}}_T^{-1} \mathbf{g}(\theta; \mathbf{y}_T)$$

来估计 θ , 其中 $\hat{\mathbf{S}}_T$ 为 \mathbf{S} 的估计。

Hansen (1982) 的经验发现

- 工具变量 x_t :
 - 消费增长率的滞后项
 - 股票收益率的滞后项
- 资产收益 $R_{i,t}$:
 - 股票市场指数
 - 若干行业组合收益率
- 消费定义:
 - 国民收入核算中的非耐用品与服务支出
- 估计结果（多数设定下）:
 - 折现因子 β 的估计值 $\hat{\beta} \approx 0.99$
 - 相对风险厌恶系数 γ 较低，约在 $[0.35, 0.999]$ 范围内

关于股权溢价之谜的注释

- 此处未显现经典的股权溢价之谜 (equity premium puzzle)，原因在于：
 - 模型利用了 x_t 中的条件信息
 - 用的是 $R_{t+1}x_t$ 等“条件收益率”，而非简单“未调整”的股票超额收益
- 股权溢价之谜通常指：
 - 单纯以股票市场超额收益率为对象
 - 在合理风险厌恶下，C-CAPM 预测的股权溢价远低于长期样本中观察到的溢价

股权溢价之谜：问题的提出

- Mehra and Prescott (1985) 首次系统提出“股权溢价之谜”
- 事实：长期样本中
 - 股票年均收益率通常高出无风险资产（如短期国债）数个百分点
- 理论困境：
 - 在合理的风险厌恶系数下，标准 C-CAPM 难以匹配如此之高的股权溢价
 - 即便使用幂效用 + 正态近似，也无法同时匹配消费波动与股权溢价

引入时变参数： β_t 与 γ_t

- 一种可能的解释：令模型参数随时间和状态变化

$$\beta_t = \beta(\mathcal{F}_{t-1}), \quad \gamma_t = \gamma(\mathcal{F}_{t-1}).$$

- 更具体地，以状态向量 $\mathbf{x}_t \in \mathcal{F}_{t-1}$ 为自变量：

$$\beta_t = \beta(\mathbf{x}_t), \quad \gamma_t = \gamma(\mathbf{x}_t),$$

其中 $\beta(\cdot)$ 、 $\gamma(\cdot)$ 为未知光滑函数。

- 经济含义：

- 风险态度与折现因子可随宏观环境、市场情绪或财富水平等状态变化

时变 C-CAPM 的广义回归式

- 由欧拉方程（式(4)）可得：

$$\beta(X_t) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma(X_t)} (1 + R_{i,t+1}) = 1 + \varepsilon_{t+1},$$

其中

- ε_{t+1} 为随机定价误差
- 满足鞅差性质： $E(\varepsilon_{t+1} | \mathcal{F}_t) = 0$

- 思路：

- 视为对 $\beta(\cdot)$ 、 $\gamma(\cdot)$ 的广义非参数回归
- 采用低阶局部多项式（如局部线性）估计这两个函数

局部多项式 GMM 的设想

- 在状态点 x 邻域内，最小化局部广义残差平方和：

$$\min_{\beta(\cdot), \gamma(\cdot)} \sum_{t=1}^T \left[\beta(X_t) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma(X_t)} (1 + R_{i,t+1}) - 1 \right]^2 K_h \left(\frac{x - X_t}{h} \right),$$

其中 K_h 为带宽 h 的核函数。

- 更一般地，将欧拉方程写成时变 GMM 矩条件：

$$E \left\{ \left[\beta(X_t) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma(X_t)} (1 + R_{i,t+1}) - 1 \right] \mathbf{x}_t \mid \mathcal{F}_t \right\} = 0.$$

- 通过局部多项式逼近 $\beta(\cdot)$ 、 $\gamma(\cdot)$ 并最小化局部样本矩二次型，可得到相应的**局部多项式 GMM 估计量**。

目录

1 资本资产定价模型

- CAPM 的检验: Wald 与 LR
- 时变 CAPM 与条件 CAPM

2 多因子模型

- 案例: Fama–French 三因子模型
- SMB/HML/UMD 的构建
- Carhart 四因子与 Fama–French 五因子
- 实践口径与中国市场提示

3 基于消费的资本资产定价模型

- 广义矩估计 (GMM)
- 采用 GMM 估计 C-CAPM 模型
- 时变风险厌恶与股权溢价之谜

4 章节总结

5 参考文献

本章小结：CAPM 与多因子模型 (1/2)

本章系统讨论了多类金融资产定价模型及其计量实现：

- **资本资产定价模型 (CAPM)**

- 给出单因子模型的估计方法 (MLE) 与分布性质
- 使用 Wald 检验与似然比检验 (LR) 检验 $\alpha = 0$
- 讨论了时变 CAPM 与条件 CAPM，以刻画风险与预期收益的时变性

- **多因子模型**

- 从 APT 与 ICAPM 的理论出发，构建多因子回归模型
- 介绍基于投资组合收益与宏观因子的多因子检验框架
- 说明了多因子模型在实际资产定价与风险管理中的重要性

本章小结：实证因子模型与 C-CAPM (2/2)

- Fama–French 三因子模型及扩展

- 以 SMB/HML/UMD/RMW/CMA 等因子为例，展示如何构建基于投资组合收益的因子
- 结合 R 代码演示了因子数据的获取、处理与回归分析流程
- 说明 Carhart 四因子与 Fama–French 五因子在收益横截面解释能力上的提升

- 基于消费的资本资产定价模型（C-CAPM）

- 在宏观一般均衡框架下，将消费与资产定价联系起来
- 通过广义矩估计（GMM）实现模型的估计与检验
- 讨论了股权溢价之谜与时变风险厌恶的可能解释路径

整体而言，本章展示了从传统 CAPM、多因子模型到 C-CAPM 的统一计量框架，帮助读者从理论与实证两个层面更全面地理解资产价格的形成机制。

目录

1 资本资产定价模型

- CAPM 的检验: Wald 与 LR
- 时变 CAPM 与条件 CAPM

2 多因子模型

- 案例: Fama–French 三因子模型
- SMB/HML/UMD 的构建
- Carhart 四因子与 Fama–French 五因子
- 实践口径与中国市场提示

3 基于消费的资本资产定价模型

- 广义矩估计 (GMM)
- 采用 GMM 估计 C-CAPM 模型
- 时变风险厌恶与股权溢价之谜

4 章节总结

5 参考文献

参考文献 I

-  Carhart, Mark M (1997). "On persistence in mutual fund performance". In: *The Journal of Finance* 52.1, pp. 57–82.
-  Chamberlain, Gary (1984). "Panel Data". In: *Handbook of Econometrics*. Ed. by Zvi Griliches and Michael D. Intriligator. Vol. 2. Elsevier.
-  Fama, Eugene F and Kenneth R French (2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116.1, pp. 1–22.
-  — (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics* 33.1, pp. 3–56.
-  Hansen, Lars Peter (1982). "Large sample properties of generalized method of moments estimators". In: *Econometrica*, pp. 1029–1054.
-  Mehra, Rajnish and Edward C Prescott (1985). "The equity premium: A puzzle". In: *Journal of Monetary Economics* 15.2, pp. 145–161.
-  Merton, Robert C (1973). "An intertemporal capital asset pricing model". In: *Econometrica*, pp. 867–887.

参考文献 II

-  Ross, Stephen A (1976). “The arbitrage theory of capital asset pricing”. In: *Journal of Economic Theory* 13.3, pp. 341–360.

第 9 章连续时间模型与高频波动率估计

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

本章导论

- 本章围绕两大主题：
 - 连续时间资产定价模型：以扩散型 SDE 刻画“漂移 + 扩散”的价格动态；
 - 高频波动率估计：把高频成交/报价数据转化为实现波动率、协方差等统计量。
- 理论与实践的衔接：
 - SDE 给出价格在“无穷小时间步”上的近似法则；
 - 高频统计量逼近 SDE 的二次变差，将连续时间理论落地为可计算指标。
- 核心问题：
 - ① 在转移密度难以闭式求解时，如何估计与检验扩散过程？
 - ② 在存在微观结构噪声与异步交易时，如何稳健估计波动率与协方差？

本章结构

- 布朗运动、随机积分与伊藤引理：
 - 从离散时间随机游走到连续时间布朗运动；
 - 伊藤过程与伊藤引理。
- 扩散过程及其统计推断：
 - 漂移与扩散系数的估计；
 - 近似 MLE 与设定检验。
- 高频波动率与协方差估计（后续小节）：
 - RV、CLT 与实现四阶变差 \widehat{IQ} ；
 - 微观结构噪声修正（TSRV、MSRV、RK、PAV）与时间同步；
 - 应用：最小方差组合与风险度量。

目录

1 布朗运动

2 布朗运动的进一步性质

3 伊藤引理 (Itô's Lemma)

4 随机积分

5 扩散过程

6 扩散过程的极大似然估计

- 布莱克–舒尔斯模型的极大似然估计
- Aït-Sahalia 似然展开：近似 MLE
- 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断

7 从高频数据中估计波动率

8 测量误差模型与微观结构噪声

9 基于模拟的 RV–IQ 与 CLT 演示

10 高频协方差矩阵估计方法

- 双频已实现协方差估计量 (TSRV)
- 多元已实现核协方差估计量 (RK)
- 预平均协方差估计量 (PAV)

11 章节总结

12 参考文献

从随机游走到布朗运动

- 设 $\{\varepsilon_t\}$ 独立同分布, $\varepsilon_t \sim N(0, 1)$ 。

- 离散时间随机游走:

$$z_{t+1} - z_t = \varepsilon_t, \quad t = 0, 1, 2, \dots$$

- 累积形式:

$$z_t - z_0 = \sum_{j=0}^{t-1} \varepsilon_{t-1-j}.$$

- 期望与方差:

$$E(z_t - z_0) = 0, \quad \text{Var}(z_t - z_0) = t.$$

- 直观: 粒子在数轴上的位置

- 期望位置不随时间改变 (仍为 z_0);
- 波动随时间增长, 标准差 $\propto \sqrt{t}$ 。

连续时间布朗运动：直观构造

- 在连续时间下，设过程 $\{z_t\}$ 满足：

$$z_{t+\Delta} - z_t \sim N(0, \Delta), \quad \Delta > 0,$$

且增量在两两不相交区间上独立。

- 限制 Δ 为整数 \Rightarrow 离散随机游走；允许任意正实数 \Rightarrow 连续时间极限。
- 满足：
 - 独立增量；
 - 正态增量，方差与间隔长度成正比；
 - 关于时间连续。
- 这样的随机过程称为**布朗运动**（或维纳过程）。

布朗运动的严格定义

定义

若随机过程 $\{X(t), t \geq 0\}$ 满足：

- ① 独立增量：任意不相交区间上的增量相互独立；
- ② 对任意 $s, t > 0$,

$$X(s+t) - X(s) \sim N(0, \sigma^2 t);$$

- ③ 关于 t 连续。

则称 $\{X(t), t \geq 0\}$ 为**布朗运动**（或维纳过程）。当 $\sigma = 1$ 时称为**标准布朗运动**。

标准布朗运动的基本性质

通常用 B_t 表示标准布朗运动，满足：

① **独立增量性**：对任意 $s < t$, $B_t - B_s$ 与 \mathcal{F}_s 独立。

② **正态增量**：

$$B_t - B_s \sim N(0, t - s).$$

③ **非平稳性与平稳增量**：

$$\text{Cov}(B_s, B_t) = \min(s, t),$$

过程本身非平稳，但增量分布仅依赖于 $|t - s|$ 。

④ **无限可分性**：

$$B_t = \sum_{i=1}^n Z_i \sqrt{t_i - t_{i-1}}, \quad 0 = t_0 < \cdots < t_n = t,$$

其中 $Z_i \sim N(0, 1)$ 独立。

R 代码：模拟标准布朗运动路径

```
# 设置参数
set.seed(123)          # 随机种子
n <- 1000               # 时间步数
T <- 1                  # 总时间
dt <- T / n
t <- seq(0, T, length.out = n + 1)

# 标准正态增量
dW <- rnorm(n, mean = 0, sd = sqrt(dt))

# 布朗运动路径
W <- c(0, cumsum(dW))

# 绘图
plot(t, W, type = 'l',
      main = "标准布朗运动",
      xlab = "时间", ylab = "W(t)",
      col = "blue", lwd = 2)
```

布朗桥 (Brownian Bridge)

定义

设 B_t 为满足 $B_0 = 0$ 的标准布朗运动。给定 $T > 0$, 定义至时间 T 的布朗桥:

$$\mathbb{B}_t = B_t - \frac{t}{T} B_T, \quad 0 \leq t \leq T.$$

- 直观上: 强制过程在 $t = 0$ 与 $t = T$ 取给定端点 (如 0), 中间“悬挂”的随机路径。
- 在金融中常用于:
 - 路径依赖期权 (如亚式期权) 的模拟;
 - 条件在期末达到某价位下的路径构造。

有界 p -变差与布朗运动轨道

定义 (有界 p -变差)

对函数 $f: [0, T] \rightarrow \mathbb{R}$, 若对任意分割 $0 = t_0 < t_1 < \dots < t_n = T$ 有

$$\sup_{\{t_i\}} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|^p < \infty, \quad p > 0,$$

则称 f 具有**有界 p -变差**。

- 连续可微函数 \Rightarrow 有界 1-变差。
- 布朗运动的样本路径在任意紧区间上**无界变差**:
 - “连续但极度锯齿化”, 几乎处处不可导;
 - 这也是传统黎曼积分不再适用、需要随机积分的原因之一。

布朗运动的“微分”与尺度

- 形式上，可写

$$dz_t = \lim_{\Delta \rightarrow 0^+} (z_{t+\Delta} - z_t).$$

对应离散增量：

$$\varepsilon_{t+1} = z_{t+1} - z_t.$$

- 布朗运动增量的方差： $\text{Var}(dz_t) = \Delta$ ，故标准差 $\sim \sqrt{\Delta}$ 。
- 由此得到两个结论：

- 处处不可导**：形式上的“导数” $\frac{\sqrt{\Delta}}{\Delta} \rightarrow \infty$ （直观上斜率趋于无穷大）。
- 无论区间多短，路径都呈现明显随机波动，而非“接近直线”。

处处连续但处处不可导：魏尔斯特拉斯函数



19世纪，魏尔斯特拉斯构造了一个处处连续却处处不可导的函数：

$$W(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x),$$

其中 $0 < a < 1$, b 为正奇数，并满足 $ab > 1 + \frac{3}{2}\pi$ 。
该函数在每一点都没有普通意义的导数。

类比地，若把布朗运动视作时间的函数，它也是一种典型的“处处连续、处处不可导”随机函数。

广义布朗运动与金融含义

- 基于标准布朗运动 B_t , 广义布朗运动定义为

$$dX_t = \mu dt + \sigma dB_t.$$

- 条件期望与方差:

$$\mathbb{E}_t[dX_t] = \mu dt, \quad \text{Var}_t[dX_t] = \sigma^2 dt.$$

- 金融解释:

- 漂移项 μdt : 资金时间价值、长期趋势（与通胀、利率、增长预期等
相关）；
 - 扩散项 σdB_t : 风险波动，反映市场情绪、供需冲击、意外事件等。
- 该 SDE 是许多资产定价模型（如 Black–Scholes）的基础。

目录

1

布朗运动

2

布朗运动的进一步性质

3

伊藤引理 (Itô's Lemma)

4

随机积分

5

扩散过程

6

扩散过程的极大似然估计

- 布莱克–舒尔斯模型的极大似然估计
- Aït-Sahalia 似然展开：近似 MLE
- 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断

7

从高频数据中估计波动率

8

测量误差模型与微观结构噪声

9

基于模拟的 RV–IQ 与 CLT 演示

10

高频协方差矩阵估计方法

- 双频已实现协方差估计量 (TSRV)
- 多元已实现核协方差估计量 (RK)
- 预平均协方差估计量 (PAV)

11

章节总结

12

参考文献

样本路径的 Hölder 连续性

- 对标准布朗运动 B_t , 几乎必然有: 对任意 $\gamma < \frac{1}{2}$,

$$|B_t - B_s| \leq C|t - s|^\gamma$$

在某个有限常数 C 下成立。

- 而对任何 $\gamma > \frac{1}{2}$, 样本路径在局部都不是 Hölder 连续的, 且几乎处处不可微。
- 更细的连续性模量:

$$g(\delta) = (2\delta \ln(1/\delta))^{1/2},$$

满足

$$\mathbb{P} \left(\limsup_{\delta \rightarrow 0^+} \frac{1}{g(\delta)} \max_{\substack{0 \leq s < t \leq 1 \\ t-s \leq \delta}} |B_t - B_s| = 1 \right) = 1.$$

- 说明: 在极小尺度上路径“震荡得恰到好处”, 既不太平滑也不过于剧烈。

停时与穿越时间的定义

- 对布朗运动 B_t , 定义首次离开区间 $[-a, a]$ 的时间:

$$\tau_a = \inf\{t : |B_t| > a\}.$$

- 这是一个停时 (stopping time) 的例子。

定义 (停时)

设过程 $\{X_t\}$ 的自然滤过为 $\mathcal{F}_t^X = \sigma(X_u : 0 \leq u \leq t)$ 。若随机变量 $\tau : \Omega \rightarrow [0, \infty]$ 满足对任意 $s \geq 0$, 事件 $\{\tau \leq s\} \in \mathcal{F}_s^X$, 则称 τ 为 (相对于 \mathcal{F}_t^X 的) 停时。

- 直观: 是否“停止”只能用当前及以前的信息判断, 不能预知未来。
- 金融例子: 资本首次达到(或跌破)某阈值时退出市场的时间。

两侧首次穿越时间的分布（简述）

- 自 Bachelier (1900) 起，首次穿越时间 τ_a 已被系统研究。
- 对标准布朗运动， $B_0 = x$ ，定义

$$\tau_a = \inf\{t : |B_t| > a\}.$$

- 其分布函数为 (Darling and Siegert 1953)：

$$\Pr(\tau_a \leq t \mid x, a) = 1 - \frac{2}{\pi} \sum_{j=0}^{\infty} \frac{(-1)^j}{j + \frac{1}{2}} \cos\left(\left(j + \frac{1}{2}\right) \frac{\pi x}{a}\right) e^{-(j+1/2)^2 \pi^2 t / 2a^2}.$$

- 密度函数 $f_0(t \mid x, a)$ 可由上式对 t 求导得到。
- 特别地，当 $x = 0$ 时， $E(\tau_a) = a^2 / \sigma^2$ 。

单侧穿越概率与熔断机制

定义 (单侧穿越概率)

对随机过程 $X(t)$, 设

$$\tau_a^+ = \min\{t : X(t) > a\}.$$

单侧穿越概率为

$$\Pr(\tau_a^+ \leq t) = \Pr(\min\{s : X(s) > a\} \leq t).$$

单侧穿越概率与熔断机制

- 对标准布朗运动 B_t , $B_0 = 0$, 有 (Feller 1991):

$$\Pr(\tau_a^+ \leq t) = 2\left(1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right),$$

$$f_0(t | 0, a) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right).$$

- 金融应用: 理解价格触及熔断阈值的概率与时机:
 - 如 LSE 等市场设定大盘当日涨跌幅阈值 (如 $\pm 10\%$) 触发熔断;
 - 早盘与尾盘往往更易出现大幅波动, 单侧穿越概率在此时段更高。

目录

1

布朗运动

2

布朗运动的进一步性质

3

伊藤引理 (Itô's Lemma)

4

随机积分

5

扩散过程

6

扩散过程的极大似然估计

- 布莱克–舒尔斯模型的极大似然估计
- Aït-Sahalia 似然展开：近似 MLE
- 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断

7

从高频数据中估计波动率

8

测量误差模型与微观结构噪声

9

基于模拟的 RV–IQ 与 CLT 演示

10

高频协方差矩阵估计方法

- 双频已实现协方差估计量 (TSRV)
- 多元已实现核协方差估计量 (RK)
- 预平均协方差估计量 (PAV)

11

章节总结

12

参考文献

伊藤引理：随机版链式法则

- 问题：若 X_t 遵循布朗运动或更一般的伊藤过程， $f(t, X_t)$ 如何演化？
- 金融中的典型情形：
 - 期权价格是标的资产价格 S_t 的函数；
 - 若 S_t 满足 SDE，如何刻画期权价格 $C(t, S_t)$ 的动态？
- 伊藤引理：是随机微积分中的链式法则，是经典微积分链式法则的推广。
- 与普通泰勒展开的差别：
 - 在随机情形中， $(dB_t)^2$ 与 dt 同阶，不能忽略；
 - 所以需要保留到“二阶”（关于 dB_t 的平方）项。

伊藤微分乘法表

进行随机微分运算的“代数法则”：

- $dt \cdot dt = 0$: 二次及以上的时间增量为高阶无穷小，可忽略。
- $dt \cdot dB_t = dB_t \cdot dt = 0$: 与 dt^2 同阶，可忽略。
- $dB_t \cdot dB_t = dt$: 布朗运动增量的方差与时间间隔 dt 成正比。

	dB_t	dt
dB_t	dt	0
dt	0	0

伊藤引理的具体形式

- 设 X_t 为伊藤过程：

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t.$$

- 对光滑函数 $f(t, x)$, 定义 $Y_t = f(t, X_t)$ 。
- 则伊藤引理给出 Y_t 的微分：

$$\begin{aligned} df(t, X_t) &= \left(\frac{\partial f}{\partial t} + \mu(t, X_t) \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2(t, X_t) \frac{\partial^2 f}{\partial x^2} \right) dt \\ &\quad + \sigma(t, X_t) \frac{\partial f}{\partial x} dB_t. \end{aligned}$$

- 在金融中：将 f 取为期权定价函数 $C(t, S_t)$, 配合无套利条件, 即可推出 Black–Scholes PDE。

目录

1

布朗运动

2

布朗运动的进一步性质

3

伊藤引理 (Itô's Lemma)

4

随机积分

5

扩散过程

6

扩散过程的极大似然估计

- 布莱克–舒尔斯模型的极大似然估计
- Aït-Sahalia 似然展开：近似 MLE
- 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断

7

从高频数据中估计波动率

8

测量误差模型与微观结构噪声

9

基于模拟的 RV–IQ 与 CLT 演示

10

高频协方差矩阵估计方法

- 双频已实现协方差估计量 (TSRV)
- 多元已实现核协方差估计量 (RK)
- 预平均协方差估计量 (PAV)

11

章节总结

12

参考文献

从增量和到随机积分

- 对标准布朗运动 B_t , 可形式地写:

$$\int_0^T dB_t = \lim_{\Delta t \rightarrow 0} \sum_k (B_{t_{k+1}} - B_{t_k}) = B_T - B_0.$$

- 因此,

$$B_T - B_0 \sim N(0, T).$$

- 若 $dX_t = \mu dt + \sigma dB_t$, 则积分得

$$X_T - X_0 = \mu T + \sigma \int_0^T dB_t,$$

故

$$E(X_T - X_0) = \mu T, \quad \text{Var}(X_T - X_0) = \sigma^2 T.$$

- 随机微分方程刻画“瞬时”变化, 随机积分把它还原为“有限时间段”的累积变化。

一般形式的随机积分

- 考虑更一般的形式：

$$x_t = \int_a^t f(B_s, s) dB_s.$$

- 需要给出严谨定义：用增量 Riemann 和在均方意义下的极限。

定义 (随机积分)

设 B 为标准布朗运动。若 $\{f_t, t \in [a, b]\}$ 对 $\{B_s, s \leq t\}$ 适应且

$\int_a^b E[f_t^2] dt < \infty$, 定义

$$\int_a^b f_t dB_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f_{t_i} (B_{t_{i+1}} - B_{t_i}),$$

其中划分 $a = t_0 < \dots < t_n = b$ 的网格宽度 $\max_i |t_{i+1} - t_i| \rightarrow 0$, 极限在均方意义下成立。

随机积分的基本性质

设 $I(f) = \int_a^b f_t dB_t$.

- **线性性**: 对任意可积过程 f, g 与常数 α, β ,

$$I(\alpha f + \beta g) = \alpha I(f) + \beta I(g).$$

- **鞅性质**: 在适当条件下, 过程

$$X_t = \int_a^t f_u dB_u$$

是相对于布朗运动自然滤过的鞅, 即对 $s < t$,

$$E(X_t | \mathcal{F}_s) = X_s.$$

- **Itô 等距 (isometry)**:

$$E\left[\left(\int_a^b f_t dB_t\right)^2\right] = E\left[\int_a^b f_t^2 dt\right].$$

- 该等距是随机积分理论和 L^2 空间之间的重要桥梁。

目录

1

布朗运动

2

布朗运动的进一步性质

3

伊藤引理 (Itô's Lemma)

4

随机积分

5

扩散过程

6

扩散过程的极大似然估计

- 布莱克-舒尔斯模型的极大似然估计
- Aït-Sahalia 似然展开：近似 MLE
- 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断

7

从高频数据中估计波动率

8

测量误差模型与微观结构噪声

9

基于模拟的 RV-IQ 与 CLT 演示

10

高频协方差矩阵估计方法

- 双频已实现协方差估计量 (TSRV)
- 多元已实现核协方差估计量 (RK)
- 预平均协方差估计量 (PAV)

11

章节总结

12

参考文献

马尔可夫过程与扩散过程：基本概念

- 扩散过程通常被理解为：
 - 连续时间；
 - 连续样本路径；
 - 满足强马尔可夫性质的随机过程。
- 首先回顾马尔可夫过程的定义：

定义 (马尔可夫过程)

若随机过程 $\{X_t, t \in [0, T]\}$ 对所有 x, t, s 满足

$$\Pr(X_t \leq x | \mathcal{F}_s) = \Pr(X_t \leq x | X_s),$$

则称其为马尔可夫过程。即在给定 X_s 后， X_t 与 $\{X_r : r < s\}$ 独立。

扩散过程与强马尔可夫性质

定义 (扩散过程)

扩散过程是具有连续样本路径的连续时间**强马尔可夫过程**。

- 强马尔可夫性质：将“当前时间点”从**固定时刻**拓展到**停时** τ 。
- 对任一停时 τ ，在 $\{\tau < \infty\}$ 条件下，对所有 $t \geq 0$ ：
 - $X_{\tau+t}$ 在给定 X_τ 的条件下，与 τ 之前的路径 $\{X_s : s < \tau\}$ 独立；
 - 未来只依赖于“停时状态”而与以往的轨迹无关。
- 很多连续时间金融模型（利率、波动率、价格）都可建模为扩散过程。

扩散过程与随机微分方程 (SDE)

- 扩散过程通常通过随机微分方程来定义：

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t,$$

其中

- $\mu(\cdot, \cdot)$: 漂移项 (drift term);
- $\sigma(\cdot, \cdot)$: 扩散项 (diffusion term);
- B_t : 标准布朗运动。

扩散过程与随机微分方程 (SDE)

定理 (积分形式)

假设 $X_0 = X$, 且

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t.$$

则广义扩散过程可等价写为

$$X_t = X_0 + \int_0^t \mu(X_s, s) ds + \int_0^t \sigma(X_s, s) dB_s, \quad (1)$$

其中第一个积分为黎曼积分, 第二个积分为随机积分。

扩散 SDE 的存在唯一性条件

定理 (存在唯一性与矩条件)

对方程 (1), 下列条件足以保证存在唯一解 $\{X_t, t \in [0, T]\}$, 该解具有连续样本路径、为马尔可夫过程, 且 $\int \mathbb{E}(X_t^2) dt < \infty$:

- ① **矩条件:** $E(X^2) < \infty$;
- ② **Lipschitz 条件:** μ, σ 为 Borel 可测函数, 且存在常数 $K < \infty$, 对所有 $x, y \in \mathbb{R}$,

$$|\mu(x, t) - \mu(y, t)| \leq K|x - y|, \quad |\sigma(x, t) - \sigma(y, t)| \leq K|x - y|;$$

- ③ **增长条件:** 存在 $K > 0$, 对所有 $x \in \mathbb{R}$,

$$|\mu(x, t)| \leq K(1 + x^2)^{1/2}, \quad |\sigma(x, t)| \leq K(1 + x^2)^{1/2}.$$

常见扩散过程举例（一）

- 几何布朗运动 (Black–Scholes 模型):

$$dX_t = \beta X_t dt + \sigma X_t dB_t.$$

- Ornstein–Uhlenbeck 过程 (Vasicek 1977):

$$dX_t = \beta(\alpha - X_t) dt + \sigma dB_t.$$

- Feller 平方根过程 (CIR 模型) (Cox, Ingersoll Jr, and Ross 1985):

$$dX_t = \beta(\alpha - X_t) dt + \sigma \sqrt{X_t} dB_t.$$

- Courtadon (1982):

$$dX_t = \beta(\alpha - X_t) dt + \sigma X_t dB_t.$$

常见扩散过程举例（二）

- Marsh and Rosenfeld (1983):

$$dX_t = (\alpha X_t^{-(1-\delta)} + \beta) dt + \sigma X_t^{\delta/2} dB_t.$$

- Cox (1975):

$$dX_t = \beta(\alpha - X_t) dt + \sigma X_t^\gamma dB_t.$$

- Constantinides (1992):

$$dX_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2) dt + (\sigma_0 + \sigma_1 X_t) dB_t.$$

- 仿射模型** (affine models) (Duffie and Kan 1996; Dai and Singleton 2000):

$$dX_t = \beta(\alpha - X_t) dt + (\sigma_0 + \sigma_1 X_t)^{1/2} dB_t.$$

- 非线性均值回归模型** (Aït-Sahalia 1996):

$$dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \frac{\alpha_{-1}}{X_t} \right) dt + \left(\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3} \right) dB_t.$$

目录

- 1 布朗运动
- 2 布朗运动的进一步性质
- 3 伊藤引理 (Itô's Lemma)
- 4 随机积分
- 5 扩散过程
- 6 扩散过程的极大似然估计
 - 布莱克–舒尔斯模型的极大似然估计
 - Aït-Sahalia 似然展开：近似 MLE
 - 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断
- 7 从高频数据中估计波动率
- 8 测量误差模型与微观结构噪声
- 9 基于模拟的 RV–IQ 与 CLT 演示
- 10 高频协方差矩阵估计方法
 - 双频已实现协方差估计量 (TSRV)
 - 多元已实现核协方差估计量 (RK)
 - 预平均协方差估计量 (PAV)
- 11 章节总结
- 12 参考文献

扩散过程的极大似然估计：思路

- 任何满足马尔可夫性质的连续时间金融模型都可用 MLE 估计。
- 设有 $n + 1$ 个观测：

$$t_0 < t_1 < \cdots < t_n, \quad X_{t_i} = X_{i\Delta}, \quad \Delta = t_i - t_{i-1},$$

为简便假定等间隔采样。

- 目标：在参数向量 θ 上最大化样本联合密度，对应模型

$$dX_t = \mu(X_t; \theta) dt + \sigma(X_t; \theta) dB_t.$$

联合密度分解与转移密度

- 联合密度可按马尔可夫性质分解为：

$$\begin{aligned}
 & \Pr(X_{n\Delta}, X_{(n-1)\Delta}, \dots, X_0; \theta) \\
 &= \Pr(X_{n\Delta} | X_{(n-1)\Delta}, \dots, X_0; \theta) \Pr(X_{(n-1)\Delta}, \dots, X_0; \theta) \\
 &= \Pr(X_{n\Delta} | X_{(n-1)\Delta}; \theta) \cdots \Pr(X_\Delta | X_0; \theta) \Pr(X_0; \theta). \tag{2}
 \end{aligned}$$

- 记 $p_X(\Delta; X_{i\Delta} | X_{(i-1)\Delta}; \theta)$ 为**转移密度** (transition density)。
- 基于观测 $\{X_0, X_\Delta, \dots, X_{n\Delta}\}$, 可构造对数似然函数。

对数似然与信息矩阵

- 对于扩散过程

$$dX_t = \mu(X_t; \theta) dt + \sigma(X_t; \theta) dB_t,$$

其对数似然为

$$l_n(\theta) = \sum_{i=1}^n \ln p_X(\Delta; X_{i\Delta} | X_{(i-1)\Delta}; \theta). \quad (3)$$

- 极大似然估计量：

$$\hat{\theta} = \arg \max_{\theta \in \Theta} l_n(\theta). \quad (4)$$

- 在适当正则条件下：

$$\sqrt{n}(\hat{\theta} - \theta) \implies N(0, \mathcal{I}^{-1}(\theta)),$$

$$\mathcal{I}(\theta) = \lim_{n \rightarrow \infty} -E \left[\frac{1}{n} \frac{\partial^2 l_n(\theta)}{\partial \theta \partial \theta'} \right].$$

MLE 的渐近性质与效率

- 当 n 足够大时,

$$\text{Var}(\hat{\theta}) \approx \frac{1}{n} \mathcal{I}^{-1}(\theta).$$

- 信息矩阵的样本估计:

$$\widehat{\mathcal{I}} = \frac{1}{n} \frac{\partial^2 I_n(\hat{\theta})}{\partial \theta \partial \theta'}.$$

- MLE 的性质:

- 一致 (consistent);
- 统一渐近正态 (Uniformly Asymptotically Normal, UAN);
- 渐近协方差达到 Cramér–Rao 下界;
- 在所有 CUAN 估计量 (Consistent and Uniformly Asymptotically Normal) 中最有效。

- 详细证明可参见 Campbell, Lo, and MacKinlay (1997, A.4 节)。

Black-Scholes 模型：几何布朗运动

- Black-Scholes 模型假定股票价格遵循几何布朗运动：

$$dX_t = \beta X_t dt + \sigma X_t dB_t. \quad (5)$$

- 这是最经典的扩散型资产价格模型，被广泛用于股票及衍生品定价。
- 对数变换 $Y_t = \log X_t$ 后，引入伊藤引理。

对数收益与正态性

- 对 $Y_t = \log X_t$ 应用伊藤引理并代入 (5), 得到

$$d\log X_t = \left(\beta - \frac{1}{2}\sigma^2 \right) dt + \sigma dB_t = \alpha dt + \sigma dB_t, \quad (6)$$

其中 $\alpha = \beta - \frac{1}{2}\sigma^2$ 。

- 连续复利收益率

$$r_t(\Delta) \equiv \log\left(\frac{X_t}{X_{t-\Delta}}\right)$$

为 i.i.d. 正态:

$$r_t(\Delta) \sim N(\alpha\Delta, \sigma^2\Delta).$$

- 因此 $r_t(\Delta)/\sqrt{\Delta}$ 的样本方差是 σ^2 的自然估计量, 并且是 σ 的 MLE。

Black-Scholes 模型的对数似然与 MLE

- 在 (6) 下, 样本 $r_1(\Delta), \dots, r_n(\Delta)$ 的对数似然为

$$l_n(\alpha, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2\Delta) - \frac{1}{2\sigma^2\Delta} \sum_{t=1}^n (r_t(\Delta) - \alpha\Delta)^2. \quad (7)$$

- MLE 为

$$\begin{aligned}\hat{\alpha} &= \frac{1}{n\Delta} \sum_{t=1}^n r_t(\Delta), \\ \hat{\sigma}^2 &= \frac{1}{n\Delta} \sum_{t=1}^n (r_t(\Delta) - \hat{\alpha}\Delta)^2.\end{aligned}$$

- 由于 $r_t(\Delta)$ 独立同分布且正态, 满足常规正则条件, 因此 $\hat{\alpha}, \hat{\sigma}^2$ 在所有一致、渐近无偏正态估计量中渐近最有效。

其他扩散模型的转移密度

- 几何布朗运动 (Black–Scholes) :

$$dX_t = \beta X_t dt + \sigma X_t dB_t$$

的转移密度为对数正态。

- Ornstein–Uhlenbeck 过程:

$$dX_t = \beta(\alpha - X_t) dt + \sigma dB_t$$

的转移密度为正态。

- Feller 平方根过程 (CIR):

$$dX_t = \beta(\alpha - X_t) dt + \sigma \sqrt{X_t} dB_t$$

的转移密度为非中心卡方分布。

- 但对于许多实际扩散模型, ρ_X 无闭式形式 (前述 Courtadon、Marsh、Constantinides、Duffie–Kan、Dai–Singleton、Aït–Sahalia 等模型)。

转移密度未知时的替代方法

- 当 p_X 无闭式形式时，可用多种近似或替代方法：
 - 广义矩估计 (GMM)；
 - 模拟似然法 (simulated likelihood) (Pedersen 1995)；
 - 基于偏微分方程的估计；
 - 各类近似转移密度方法 (如 Aït-Sahalia 封闭形式展开)。
- Jensen and Poulsen (1999) 讨论了多种近似方法的比较。
- Linton (2019) 第 12 章第 4 节给出了 PDE 估计与 GMM 等方法的具体实现。
- 本章重点放在基于 Aït-Sahalia 似然展开的近似 MLE。

Aït-Sahalia 近似方法：总体思路

- 目标：在转移密度 p_X 无闭式形式时，构造其可计算的近似，以实施近似 MLE。
- Aït-Sahalia (1996), Aït-Sahalia (2002), and Aït-Sahalia, Fan, and Peng (2009) 提出的思路：
 - ① 两次变换 $X \mapsto Y \mapsto Z$ ，使得 Z 的分布依分布地逼近正态；
 - ② 围绕“近似正态”的 Z 构造 Hermite 展开的 p_Z ；
 - ③ 通过反变换 $Z \rightarrow Y \rightarrow X$ 得到 $p_X^{(J)}$ 的近似序列。
- 直接对 X 进行 Edgeworth 展开往往发散（如 GBM 右尾过厚），故先做归一化与变换。

第一步： $X \rightarrow Y$ 变换

- 考虑单变量扩散模型：

$$dX_t = \mu(X_t; \theta) dt + \sigma(X_t; \theta) dB_t. \quad (8)$$

- 定义

$$Y_t \equiv \gamma(X_t; \theta) = \int^{X_t} \frac{du}{\sigma(u; \theta)}.$$

- 由伊藤引理可得

$$dY_t = \mu_Y(Y_t; \theta) dt + dB_t, \quad (9)$$

其中

$$\mu_Y(y; \theta) = \frac{\mu(\gamma^{-1}(y; \theta); \theta)}{\sigma(\gamma^{-1}(y; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial x}(\gamma^{-1}(y; \theta); \theta).$$

- 通过该变换，可以消去原 SDE 中的 $\sigma(\cdot)$ 项，使局部方差归一化为 1，但漂移项会变得较复杂。

第二步： $Y \rightarrow Z$ 标准化与 Hermite 展开

- 在固定采样间隔 Δ 下，将 Y 标准化：

$$Z_t \equiv \Delta^{-1/2} (Y_t - y_0).$$

- 只要 Z_t 足够接近正态，就可以对 p_Z 进行 Hermite 展开：

$$H_j(z) \equiv e^{z^2/2} \frac{d^j}{dz^j} [e^{-z^2/2}], \quad \phi(z) \equiv \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

- Z 的密度在阶数 J 下近似为：

$$p_Z^{(J)}(\Delta, z | y_0; \theta) \equiv \phi(z) \sum_{j=0}^J \eta_j(\Delta, y_0; \theta) H_j(z).$$

- 其中系数

$$\eta_j(\Delta, y_0; \theta) = \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) p_Z(\Delta, z | y_0; \theta) dz$$

为在 p_Z 下的期望。

第三步：反变换 $Z \rightarrow Y \rightarrow X$

- 若已知 $p_Z^{(J)}$, 则可得到 $p_Y^{(J)}$:

$$p_Y^{(J)}(\Delta, y | y_0; \theta) = \Delta^{-1/2} p_Z^{(J)}\left(\Delta, \Delta^{-1/2}(y - y_0) | y_0; \theta\right).$$

- 再用链式变换得到 $p_X^{(J)}$:

$$p_X^{(J)}(\Delta, x | x_0; \theta) = \sigma(x; \theta)^{-1} p_Y^{(J)}(\Delta, \gamma(x; \theta) | \gamma(x_0; \theta); \theta).$$

- 注意:

- $Y_t = \gamma(X_t; \theta)$ 是一般的**非线性**变换;
- 因此 $p_X^{(J)}$ 通常不是正态密度, 而是正态密度的非线性变换结果。

收敛性质与极大似然估计

定理 (转移密度近似的收敛)

存在 $\bar{\Delta} > 0$, 使得对每个 $\Delta \in (0, \bar{\Delta})$, 任意 $\theta \in \Theta$ 与 $(x, x_0) \in D_X^2$:

$$p_X^{(J)}(\Delta, x | x_0; \theta) \rightarrow p_X(\Delta, x | x_0; \theta) \quad \text{当 } J \rightarrow \infty$$

收敛在 Θ 上对 θ 一致, 在 D_X 上对 x 一致, 并在 D_X 的任意紧子集上对 x_0 一致。

通过最大化

$$l_n^{(J)}(\theta) \equiv \sum_{i=1}^n \ln p_X^{(J)}(\Delta, X_{i\Delta} | X_{(i-1)\Delta}; \theta)$$

得到的估计量 $\hat{\theta}_n^{(J)}$, 在 $J \rightarrow \infty$ 时收敛到真实 (但不可计算) 的 MLE $\hat{\theta}_n$, 并继承其所有渐近性质。

系数展开与闭式表达

- 为获得 $p_Z^{(\infty)}$:

$$p_Z^{(\infty)}(\Delta, z \mid y_0; \theta) = \phi(z) \sum_{j=0}^{\infty} \eta_j(\Delta, y_0; \theta) H_j(z),$$

需计算 η_j 对 Δ 的泰勒展开（到 Δ^K 阶）。

- 记 $\tilde{p}_Z^{(K)}$ 为该展开，Aït-Sahalia (2002) 给出：

定理

$$\tilde{p}_Z^{(K)}(\Delta, z \mid y_0; \theta) = \Delta^{-1/2} \phi\left(\frac{y - y_0}{\Delta^{1/2}}\right) \exp\left(\int_{y_0}^y \mu_Y(\omega; \theta) d\omega\right) \sum_{k=0}^K c_k(y \mid y_0; \theta) \frac{\Delta^k}{k!}.$$

系数 c_k 与 λ_Y 的递推

- 系数有闭式递推形式：

$$c_0(y | y_0; \theta) = 1,$$

$$c_j(y | y_0; \theta) = j(y - y_0)^{-j} \int_{y_0}^y (\omega - y_0)^{j-1} \left\{ \lambda_Y(\omega; \theta) c_{j-1}(\omega | y_0; \theta) + \frac{1}{2} \frac{\partial^2 c_{j-1}(\omega | y_0; \theta)}{\partial \omega^2} \right\} d\omega,$$

$$\lambda_Y(y; \theta) = -\frac{1}{2} \left(\mu_Y^2(y; \theta) + \frac{\partial \mu_Y(y; \theta)}{\partial y} \right).$$

- 通过截断 K 与 J ，即可获得实际可算的近似转移密度与近似 MLE。
- Aït-Sahalia (2002) 通过仿真表明，该方法相较许多替代方案表现优越。

扩散模型的设定检验

- Aït-Sahalia (1996) 基于
 - 漂移项与扩散项 (μ, σ) 之间的映射；
 - 边缘密度 π_X 与转移密度 p_X ，

构造模型设定检验：

- 使用离散采样下的 π_X 与 p_X ，而非不可观测的“真”连续轨迹；
- 例如：OU 产生高斯转移密度，CIR 产生伽马/非中心卡方形式。
- Hong and Li (2005)：在原假设下，

$$\{P_X(X_{i\Delta} | X_{(i-1)\Delta}, \Delta, \theta)\}$$

为 i.i.d. $U(0, 1)$ 序列，可据此构造检验。

- Aït-Sahalia, Fan, and Peng (2009)：直接比较非参数估计的转移密度与参数模型：

$$\begin{aligned} H_0 : \quad p_X(y | x, \Delta) &= p_X(y | x, \Delta, \theta), \\ H_1 : \quad p_X(y | x, \Delta) &\neq p_X(y | x, \Delta, \theta). \end{aligned}$$

- Chen, Gao, and Tang (2008)：采用核方法非参数估计转移密度。

案例概述：R 包 MLEMVD 的实现

- 连续时间扩散模型的估计与检验可用 R 包 MLEMVD 实现：

<https://rdrr.io/github/mfrdixon/MLEMVD/>

- 安装方式（包不在 CRAN）：

```
install.packages("remotes")
remotes::install_github("mfrdixon/MLEMVD")
```

- GitHub 安装的优势：

- 开发者可快速推送更新与修复；
- 用户可即时获取最新版本；
- 便于基于源码进行二次开发、扩展到新模型与新方法。

估计思路与建模策略

- 数据：示例代码从 Yahoo Finance 抓取 SPY 日频复权收盘价，自 2015-01-01 起。
 - 也可替换为任意自有价格序列 `price`；
 - 采用价格“层级”而非收益率建模。
- 步长：日频 $\Rightarrow \Delta = 1/252$ 。
- 方法管线：
 - ① 对价格做统一缩放，改善数值条件；
 - ② 拟合三次漂移、常数扩散模型 (U6)，使用 Aït-Sahalia 似然展开近似 MLE；
 - ③ 由近似信息矩阵求协方差与标准误，如近奇异则用数值 Hessian 兜底；
 - ④ 作为稳健备选，拟合二次漂移、常数扩散模型 (U4)：
 - 随机搜索可行起点（全样本似然有限）；
 - 使用无导数 Subplex 算法优化；
 - 统一用“信息矩阵优先、Hessian 兜底”的口径给出标准误。
- 目标：在转移密度不可得时，构造近似 MLE + 稳健标准误的“可报告链条”。

R 代码：U6 与 U4 近似 MLE (1/8)

```

# =====
# 扩散模型的 Aït-Sahalia 似然展开：U6 + U4 稳健估计与标准误
# - U6: 三次漂移 + 常数扩散（近似 MLE）
# - U4: 二次漂移 + 常数扩散（稳健备选）
# =====

suppressPackageStartupMessages({
  library(MLEMD)
  library(nloptr)
  library(MASS)      # ginv()
  library(numDeriv)  # hessian()
  library(quantmod)  # 取数用（如已有 price，可跳过）
})

### 0) 数据（如已自备 price，可注释本段）
if (!exists("price")) {
  getSymbols("SPY", src = "yahoo", from = "2015-01-01")
  price <- as.numeric(Ad(SPY))
}
n <- length(price)
del <- 1/252      # 日频步长

### 1) 统一缩放（非常有助于数值稳定）
scale_factor <- 1000
price_s <- price / scale_factor

### 2) 适配器：让 mle() 支持 args
ModelU6_with_args <- function(x, x0, del, param, args = NULL)
  MLEMMD::ModelU6(x, x0, del, param)
ModelU4_with_args <- function(x, x0, del, param, args = NULL)
  MLEMMD::ModelU4(x, x0, del, param)

```

R 代码：U6 与 U4 近似 MLE (2/8)

```
## 3) 工具函数
total_loglik <- function(logdensity_fun, param, x, del) {
  s <- 0
  for (i in 1:(length(x) - 1))
    s <- s + logdensity_fun(x[i+1], x[i], del, param)$llk
  s
}

safe_vcov_from_I <- function(I, n) {
  I <- 0.5 * (I + t(I)) # 对称化
  I <- I + 1e-10 * diag(nrow(I))
  tryCatch(solve(I) / n,
           error = function(e) MASS::ginv(I) / n)
}
```

R 代码：U6 与 U4 近似 MLE (3/8)

```
make_args <- function(param0, f_upper = 50, maxeval = 1500,
                      method = c("LBFGS", "SBPLX"),
                      deoptim_iter = 0, print_level = 0) {
  method <- match.arg(method)
  p <- length(param0)
  l <- rep(-1, p); u <- rep(1, p)
  l[p] <- 1e-8; u[p] <- f_upper
  algo <- if (method == "LBFGS") "NLOPT_LD_LBFGS" else "NLOPT LN SBPLX"
  list(
    mode = "direct",
    nloptr = list(
      method = algo,
      maxeval = maxeval,
      xtol_rel = 1e-8,
      ftol_rel = 1e-10,
      ftol_abs = 0,
      print_level = print_level,
      l = l, u = u
    ),
    DEoptim = list(maxiter = deoptim_iter,
                  population = 80, strategy = 2),
    eval_g_ineq = NULL, eval_jac_g_ineq = NULL
  )
}
```

R 代码：U6 与 U4 近似 MLE (4/8)

```

series_ll_finite <- function(logdensity_fun, x, del, par) {
  s <- 0
  for (i in 1:(length(x) - 1)) {
    li <- try(logdensity_fun(x[i+1], x[i], del, par)$llk,
               silent = TRUE)
    if (!is.numeric(li) || !is.finite(li)) return(FALSE)
    s <- s + li
  }
  is.finite(s)
}

### 4) U6: 三次漂移 + 常数扩散
start_u6 <- c(a = 0, b = 0, c = 0, d = 0, f = 0.05)
args_u6 <- make_args(start_u6, f_upper = 50,
                      maxeval = 1500, method = "LBFGS")

cat("\n====U6 (近似MLE) ====\n")
fit_u6 <- mle(
  logdensity = ModelU6_with_args,
  x = price_s,
  del = del,
  param0 = start_u6,
  args = args_u6
)
theta_u6 <- setNames(as.numeric(fit_u6$solution), names(start_u6))
print(theta_u6)

```

R 代码：U6 与 U4 近似 MLE (5/8)

```
I_u6 <- as.matrix(logdensity2info(
  logdensity = ModelU6_with_args,
  x = price_s, del = del, param = theta_u6
))
vcov_u6 <- safe_vcov_from_I(I_u6, n)
se_u6 <- sqrt(diag(vcov_u6))

if (any(!is.finite(se_u6))) {
  cat("U6 信息矩阵不稳，改用数值 Hessian。\\n")
  H_u6 <- numDeriv::hessian(
    function(p) -total_loglik(ModelU6_with_args, p, price_s, del),
    theta_u6
  )
  vcov_u6 <- safe_vcov_from_I(H_u6, n)
  se_u6 <- sqrt(diag(vcov_u6))
}
cat("\nU6 估计与标准误：\\n")
print(round(cbind(Estimate = theta_u6, Std.Error = se_u6), 6))
```

R 代码：U6 与 U4 近似 MLE (6/8)

```
## 5) U4: 二次漂移 + 常数扩散 (稳健备选)
set.seed(123)
best_par <- NULL; best_ll <- -Inf
for (k in 1:500) {
  a <- runif(1, -0.8, 0.8)
  b <- runif(1, -0.8, 0.8)
  c <- runif(1, -0.8, 0.8)
  f <- exp(runif(1, log(1e-3), log(50)))
  par <- c(a=a, b=b, c=c, f=f)
  if (series_ll_finite(ModelU4_with_args, price_s, del, par)) {
    ll <- total_loglik(ModelU4_with_args, par, price_s, del)
    if (ll > best_ll) { best_ll <- ll; best_par <- par }
  }
}
if (is.null(best_par))
  stop("未找到 U4 可行起点——可增大 scale_factor 或放宽边界。")
start_u4 <- best_par
message("U4 可行起点：", paste(round(start_u4, 6), collapse = ","))
args_u4 <- make_args(start_u4, f_upper = 100,
                      maxeval = 2500, method = "SBPLX",
                      print_level = 1)
```

R 代码：U6 与 U4 近似 MLE (7/8)

```
cat("\n====U4 ( 稳健近似 MLE) ====\n")
fit_u4 <- mle(
  logdensity = ModelU4_with_args,
  x = price_s,
  del = del,
  param0 = start_u4,
  args = args_u4
)
theta_u4 <- setNames(as.numeric(fit_u4$solution), c("a","b","c","f"))
print(theta_u4)

I_u4_try <- try(as.matrix(logdensity2info(
  logdensity = ModelU4_with_args,
  x = price_s, del = del, param = theta_u4
)), silent = TRUE)
```

R 代码：U6 与 U4 近似 MLE (8/8)

```
if (inherits(l_u4_try, "try-error") || any(!is.finite(l_u4_try))) {  
    cat("U4 信息矩阵不可用，改用数值 Hessian。\\n")  
    H_u4 <- numDeriv::hessian(  
        function(p) -total_loglik(ModelU4_with_args, p, price_s, del),  
        theta_u4  
    )  
    vcov_u4 <- safe_vcov_from_l(H_u4, n)  
} else {  
    vcov_u4 <- safe_vcov_from_l(l_u4_try, n)  
}  
se_u4 <- sqrt(diag(vcov_u4))  
  
cat("\nU4 估计与标准误：\\n")  
print(round(cbind(Estimate = theta_u4, Std.Error = se_u4), 6))  
cat("\\n完成。\\n")
```

目录

- 1 布朗运动
- 2 布朗运动的进一步性质
- 3 伊藤引理 (Itô's Lemma)
- 4 随机积分
- 5 扩散过程
- 6 扩散过程的极大似然估计
 - 布莱克-舒尔斯模型的极大似然估计
 - Aït-Sahalia 似然展开：近似 MLE
 - 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断
- 7 从高频数据中估计波动率
- 8 测量误差模型与微观结构噪声
- 9 基于模拟的 RV-IQ 与 CLT 演示
- 10 高频协方差矩阵估计方法
 - 双频已实现协方差估计量 (TSRV)
 - 多元已实现核协方差估计量 (RK)
 - 预平均协方差估计量 (PAV)
- 11 章节总结
- 12 参考文献

二次变差与累积波动率

定义 (二次变差 (quadratic variation))

对于平方可积过程 X_t , 其在区间 $[0, t]$ 的二次变差定义为

$$\langle X, X \rangle_{0:t} = \text{plim}_{\max_k(t_{k+1} - t_k) \rightarrow 0} \sum_{t_k \leq t} |X_{t_{k+1}} - X_{t_k}|^2,$$

其中 $0 = t_1 < t_2 < \dots < t_n = t$ 为任意分割。

二次变差与累积波动率

- 记二次变差为 QV (quadratic variation 的缩写)。
- 对于**有界变差函数** (定义 ??), 其 QV 存在且为 0。
- Andersen et al. (2003) 指出, 在相当一般的条件下有

$$\mathbb{E}[(X(t+h) - X(t))^2 | \mathcal{F}_t] = \mathbb{E}[\langle X, X \rangle_{t:t+h} | \mathcal{F}_t],$$

即回报的条件方差等于同一时间区间上 QV 的条件期望。

- 换言之, **二次变差可以用来刻画和估计回报的波动率**。

扩散过程的 QV 与半鞅

- 对一般扩散过程

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t,$$

其二次变差 $\langle X \rangle_t$ 一般是随机过程。

- 若 $\sigma^2(X_t) \equiv \sigma^2$ 为常数，则

$$\langle X \rangle_t = t\sigma^2.$$

- 对于连续平方可积的半鞅过程， $\langle X, X \rangle_t$ 总是存在。
- 为此需先引入局部鞅 (local martingale) 与半鞅 (semimartingale) 的概念。

局部鞅 (local martingale)

定义 (局部鞅)

若存在一列停时 $(\tau_k)_{k \geq 1}$ 使得 $\tau_k \uparrow \infty$ a.s., 并且对每个 k , 停止过程

$$M_t^{\tau_k} := M_{t \wedge \tau_k}$$

为鞅, 则称适应过程 M 为局部鞅。

- 每个鞅都是局部鞅;
- 每个有界的局部鞅都是鞅;
- 但并非所有局部鞅都是鞅。
- 无漂移项的某些扩散过程是局部鞅, 但未必是真正的鞅。

半鞅 (semimartingale) 及其意义

定义 (半鞅)

实值过程 X 称为半鞅，若可分解为

$$X(t) = M(t) + A(t),$$

其中 M 为局部鞅， A 为适应过程，具有局部有限（有界）变差的样本路径，且样本路径右连续并具有左极限 (càdlàg)。

半鞅 (semimartingale) 及其意义

- 过程 A 的二次变差为 0, 本质上比 M 变化慢、带来的可预测性有限。
- 半鞅类是**可以定义随机积分**的最大自然过程族:
 - 包含所有鞅；
 - 包含由布朗运动驱动的扩散过程；
 - 也包含带跳跃的过程。
- Girsanov 定理适用于半鞅；
- 资产定价基本定理: 无套利 \Rightarrow 存在等价鞅测度；
- 若资产价格为半鞅, 则可以在某等价测度下成为鞅, 从而排除套利机会。

已实现波动率 (RV): 定义与一致性

- 现在考虑 QV 的一个一致估计量——**已实现波动率** (realized volatility, RV)。
- 假设在区间 $[0, 1]$ 上以等间隔观测到 $n + 1$ 个对数价格 X_t :

$$0, \frac{1}{n}, \dots, \frac{n}{n} = 1.$$

- 定义**

$$RV_X^n = \sum_{l=1}^{n-1} \left(X_{\frac{l+1}{n}} - X_{\frac{l}{n}} \right)^2,$$

即把一天划分为 n 个等长小区间，对应日内高频收益平方和。

- 在适当条件下， RV_X^n 是 $[0, 1]$ 上 X 的二次变差 (QV) 的一致估计。
- Jacod and Protter (1998) 在伊藤类半鞅下建立了 RV_X^n 的中心极限定理；Barndorff-Nielsen and Shephard (2002) 对布朗半鞅建立了一致性与极限分布。

布朗半鞅与 RV 的中心极限定理

定义 (布朗半鞅)

布朗半鞅 X 满足

$$X_t = \int_0^t \mu_v dv + \int_0^t \sigma_v dB_v,$$

其中过程 μ, σ 可预见 (只依赖过去), 且 σ 为右连续、左极限存在的 càdlàg 过程。

布朗半鞅与 RV 的中心极限定理

- 在无杠杆 (no-leverage) 情形下——即 μ, σ 与驱动布朗运动 B 独立——
Barndorff-Nielsen and Shephard (2002) 证明：

$$\sqrt{n} (RV_X^n - QV) \xrightarrow{d} MN\left(0, 2 \int_0^1 \sigma_u^4 du\right) =: MN(0, 2 IQ),$$

其中

$$IQ := \int_0^1 \sigma_u^4 du.$$

- 一致估计量：

$$\hat{IQ} = \frac{n}{3} \sum_{i=1}^n (\Delta_i^n X)^4, \quad \Delta_i^n X := X_{i/n} - X_{(i-1)/n}.$$

RV 的标准化统计量与置信区间

- 由此得到标准化的中心极限定理：

$$\frac{\sqrt{n} (RV_X^n - QV)}{\sqrt{2 \widehat{IQ}}} \xrightarrow{d} N(0, 1). \quad (10)$$

- 应用：
 - 构造 QV (或“积分波动率”) 的渐近置信区间；
 - 对波动率路径进行假设检验 (比如比较不同区间的波动率水平)。
- 后续高频部分：如何在存在微观结构噪声时，修正 RV 估计的上行偏差。

目录

1

布朗运动

2

布朗运动的进一步性质

3

伊藤引理 (Itô's Lemma)

4

随机积分

5

扩散过程

6

扩散过程的极大似然估计

- 布莱克–舒尔斯模型的极大似然估计
- Aït-Sahalia 似然展开：近似 MLE
- 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断

7

从高频数据中估计波动率

8

测量误差模型与微观结构噪声

9

基于模拟的 RV-IQ 与 CLT 演示

10

高频协方差矩阵估计方法

- 双频已实现协方差估计量 (TSRV)
- 多元已实现核协方差估计量 (RK)
- 预平均协方差估计量 (PAV)

11

章节总结

12

参考文献

市场微观结构噪声与测量误差

- 高频金融数据中普遍存在**市场微观结构噪声**:
 - 来自市场摩擦和交易基础设施的局限;
 - 包括价格离散性、最小报价单位、订单执行延迟等;
 - 交易规则与交易员行为也会引入噪声。
- 在高频下, 这些噪声会显著扭曲:
 - 波动率估计;
 - 相关性分析;
 - 市场效率与流动性的评估。
- **波动率签名图** (volatility signature plot):
 - 展示估计波动率随采样频率变化的曲线;
 - 受噪声影响, 超高频下 RV 往往向无穷大发散 (见 Linton (2019, 图 12.4))。
- 若不处理微观结构噪声, 高频推断会系统性偏差, 甚至误读市场动态。

TAQ 数据示例：AA 的逐笔交易与报价

以下示例取自 NYSE TAQ 数据库中美国铝业公司（Alcoa, 代码 AA）2012 年 1 月 3 日的逐笔数据：

2012	1	3	79	8.96	32599	1	8.95	8.96
2012	1	3	80	8.96	100	1	8.95	8.96
2012	1	3	83	8.95	400	-1	8.95	8.96
2012	1	3	85	8.96	5300	1	8.95	8.96
2012	1	3	87	8.96	2200	1	8.95	8.96
2012	1	3	89	8.955	3600	1	8.95	8.96
2012	1	3	94	8.95	200	-1	8.95	8.96
2012	1	3	96	8.95	800	-1	8.95	8.96
2012	1	3	98	8.96	7400	1	8.95	8.96
2012	1	3	109	8.96	6764	1	8.95	8.96
2012	1	3	115	8.96	100	1	8.95	8.96
2012	1	3	118	8.96	100	1	8.95	8.97
2012	1	3	120	8.96	100	1	8.95	8.97
2012	1	3	124	8.96	100	-1	8.955	8.97
2012	1	3	139	8.97	6200	1	8.96	8.97
2012	1	3	140	8.98	200	1	8.96	8.98
2012	1	3	150	8.97	200	-1	8.97	8.98

TAQ 字段含义

- ① Year (年份): 数据记录年份 (如 2012)。
- ② Month (月份): 数据记录月份。
- ③ Day (日期): 当月具体日期。
- ④ Timestamp (时间戳): 交易发生的确切时间。
- ⑤ Price (交易价格): 该笔成交时的交易价格。
- ⑥ Volume (交易量): 该笔成交涉及的股票数量。
- ⑦ Direction (交易方向):
 - +1: 买盘主导;
 - -1: 卖盘主导。
- ⑧ Bid Price (买入价): 买方愿意支付的最高报价。
- ⑨ Ask Price (卖出价): 卖方愿意接受的最低报价。

测量误差模型: $Y_{t_j} = X_{t_j} + \varepsilon_{t_j}$

- 在理论上, 我们假想潜在“真”对数价格过程 X_t ;
- 实际观测到的对数价格为

$$Y_{t_j} = X_{t_j} + \varepsilon_{t_j},$$

其中测量误差 (微观结构噪声) ε_{t_j} 满足:

- 独立同分布, $E[\varepsilon_{t_j}] = 0$;
- $\text{Var}(\varepsilon_{t_j}) = \sigma_\varepsilon^2$;
- ε 与 X 独立。
- 参见 Zhang, Mykland, and Aït-Sahalia (2005) 将 ε 视为“简化的微观结构噪声”:
 - 长期来看 X_{t_j} 主导;
 - 极高频下, ε 可在 RV 中占主导。

有噪声时的已实现波动率: RV_Y^n 的发散

- 基于 Y_{t_j} 定义的已实现波动率:

$$\begin{aligned}
 RV_Y^n &= \sum_{i=1}^{n-1} \left(Y_{\frac{i+1}{n}} - Y_{\frac{i}{n}} \right)^2 \\
 &= \sum_{i=1}^{n-1} \left(X_{\frac{i+1}{n}} - X_{\frac{i}{n}} \right)^2 + \sum_{i=1}^{n-1} \left(\varepsilon_{\frac{i+1}{n}} - \varepsilon_{\frac{i}{n}} \right)^2 \\
 &\quad + 2 \sum_{i=1}^{n-1} \left(\varepsilon_{\frac{i+1}{n}} - \varepsilon_{\frac{i}{n}} \right) \left(X_{\frac{i+1}{n}} - X_{\frac{i}{n}} \right).
 \end{aligned}$$

- 由大数定律:

$$\frac{1}{n} \sum_{i=1}^{n-1} \left(\varepsilon_{\frac{i+1}{n}} - \varepsilon_{\frac{i}{n}} \right)^2 \xrightarrow{P} 2\sigma_\varepsilon^2.$$

有噪声时的已实现波动率: RV_Y^n 的发散

- 由柯西-施瓦茨不等式:

$$\left| \frac{1}{n} \sum_{i=1}^{n-1} (\varepsilon_{\frac{i+1}{n}} - \varepsilon_{\frac{i}{n}})(X_{\frac{i+1}{n}} - X_{\frac{i}{n}}) \right| \xrightarrow{P} 0.$$

- 因而当 $n \rightarrow \infty$ 时, $RV_Y^n \xrightarrow{P} \infty$:

$\Rightarrow RV_Y^n$ 在有噪声时并非波动率的一致估计量。

两种修正思路：低频采样与 TSRV

- 为减弱微观结构噪声影响：
 - ① 降采样：用较低频率数据计算 RV ，减少噪声主导；
 - ② 偏差修正：显式建模误差项，构造一致估计量。
- Zhang, Mykland, and Aït-Sahalia (2005) 提出 **两尺度已实现波动率** (TSRV)：
 - 结合高频与低频两个尺度的 realized variance；
 - 利用线性组合消除噪声引入的加性偏差；
 - 在微观结构噪声存在时仍然是一致估计。

TSRV: 子样本构造

定义 (子样本 realized variance)

设 $K \times (m+1) = n$ 。构造 K 个子样本：

- 第一个子样本： $\{Y_0, Y_{K/n}, \dots, Y_{mK/n}\}$ ；
- 第二个子样本： $\{Y_{1/n}, Y_{(K+1)/n}, \dots, Y_{(mK+1)/n}\}$ ；
- 以此类推。

在第 j 个子样本中 ($j = 1, \dots, K$)，有 $m+1$ 个价格、 m 个收益率。定义

$$RV_{\text{sub}_j} = \sum_{i=1}^m \left(Y_{\frac{j+iK}{n}} - Y_{\frac{j+(i-1)K}{n}} \right)^2.$$

- RV_{sub_j} 对应较低频率（步长为 K/n ）下的 realized variance。
- 无噪声时，当 $m \rightarrow \infty$ ， RV_{sub_j} 对 QV 一致。

TSRV 的线性组合与偏差抵消

- 有测量误差时，类似计算可得到

$$\begin{aligned} \frac{1}{m} RV_{\text{sub},j} &= \frac{1}{m} \sum_{i=1}^m (X_i - \bar{X}_.)^2 + \frac{1}{m} \sum_{i=1}^m (\varepsilon_i - \bar{\varepsilon}_.)^2 \\ &\quad + \frac{2}{m} \sum_{i=1}^m (X_i - \bar{X}_.) (\varepsilon_i - \bar{\varepsilon}_.) \xrightarrow{P} 2\sigma_\varepsilon^2, \end{aligned}$$

表明单个 $RV_{\text{sub},j}$ 仍然存在偏差。

- 考虑线性组合

$$RV_{\text{sub},j} - \frac{m}{n} RV_n,$$

其中 RV_n 为全样本高频 realized variance。

- 噪声项的主导偏差：

$$2m\sigma_\varepsilon^2 - \frac{m}{n} \cdot 2n\sigma_\varepsilon^2 = 0,$$

从而偏差在主阶上被完全抵消。

TSRV 的随机波动与最终估计量

- 主导项删去偏差后，剩余的是噪声驱动的随机项：

$$\sum_{i=1}^m \left\{ (\varepsilon_i - \bar{\varepsilon})^2 - 2\sigma_\varepsilon^2 \right\} = \sqrt{m} \cdot \frac{1}{\sqrt{m}} \sum_{i=1}^m \{ \dots \}.$$

- 对 K 个子样本平均：

$$T = \frac{1}{K} \sum_{j=1}^K \sum_{i=1}^m \left\{ (\varepsilon_i - \bar{\varepsilon}_j)^2 - 2\sigma_\varepsilon^2 \right\}.$$

- T 的期望为 0，每一项都可看作 n 个独立随机变量的和，因此

$$T = O_p \left(\sqrt{\frac{m}{K}} \right).$$

若 $K/m \rightarrow \infty$ ，则 T 更小。

TSRV 的随机波动与最终估计量

由此可构造两尺度已实现波动率 (TSRV):

定义 (TSRV 估计量)

$$\hat{\theta}_{\text{TSRV}} = \frac{1}{K} \sum_{j=1}^K RV_{\text{sub},j} - \frac{m}{n} RV_n.$$

TSRV 的 R 实现示例 (1/2)

```
rm(list = ls())
set.seed(123)  # 设随机种子保证可复现

# 若在 RStudio 中，自动把工作目录设到当前脚本所在路径
if (requireNamespace("rstudioapi", quietly = TRUE) &&
    rstudioapi::isAvailable()) {
  setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
}

library(readxl)

# 读取 Excel 文件（不假定第一行为列名）
aa <- read_excel("AA_daily20120103.xlsx", col_names = FALSE)
```

TSRV 的 R 实现示例 (2/2)

```

TSRV <- function(priceX, K, m, n) {
  # 对数价格
  lprice <- log(priceX)
  # 存每个子样本的 RV
  RV_sub <- numeric(K)
  for (j in 1:K) {
    # 第 j 个子样本：间隔为 K 的抽样
    sub_sample <- lprice[seq(j, by = K, length.out = m + 1)]
    returns <- diff(sub_sample)
    RV_sub[j] <- sum(returns^2)
  }
  # 全样本 RV
  returns_full <- diff(lprice)
  RV_n <- sum(returns_full^2)
  # TSRV 估计量
  TSRV_estimator <- (1 / K) * sum(RV_sub) - (m / n) * RV_n
  TSRV_estimator
}

priceX <- aa$...5  # 假定第 5 列是价格
K      <- 5        # 子样本数
m      <- 100       # 每个子样本中使用的间隔数
n      <- length(priceX)

TSRV_estimate <- TSRV(priceX, K, m, n)
TSRV_estimate

```

多尺度已实现波动率 (MSRV)

- Zhang (2006) 提出多尺度已实现波动率 (MSRV):
 - 结合 $\simeq n^{1/2}$ 个不同时间尺度下的 realized variance;
 - 对二次变差进行加性偏差修正;
 - 进一步平衡噪声与偏差, 提高估计准确性和稳健性。

定义 (多尺度已实现波动率 (MSRV))

$$\widehat{\theta}_{\text{MSRV}} = \sum_{\ell=1}^L \alpha_\ell \frac{1}{K_\ell} \sum_{j=1}^{K_\ell} RV_{\text{sub},j}^{K_\ell}$$

其中权重 $\alpha_1, \dots, \alpha_L$ 满足若干约束 (包括 $\sum_\ell \alpha_\ell \simeq 1$), 并对 L 与 K_ℓ 的增长速度给出条件, 详见 Zhang (2006)。

多尺度已实现波动率 (MSRV)

- Zhang (2006) 证明：在常数波动率情形下，MSRV 的收敛速度达到最优，与高斯 MLE 一致。
- Aït-Sahalia, Mykland, and Zhang (2011) 进一步扩展 TSRV 与 MSRV，使其在带序列相关噪声 下仍保持一致性。

实现核 (Realized Kernels, RK) 与预平均方法

定义 (实现核 (RK) 估计量)

实现核估计量定义为

$$RK_H = \sum_{|h| < n} k\left(\frac{h}{H+1}\right) \gamma_h(Y), \quad \gamma_h(Y) := \sum_{j=h+1}^n Y_{t_j} Y_{t_{j-h}},$$

其中：

- 核函数 k 满足 $k(0) = 1$, 且 $k(s) \rightarrow 0$ 当 $|s| \rightarrow \infty$;
- H 为带宽参数, 用于控制偏差-方差权衡。

实现核 (Realized Kernels, RK) 与预平均方法

- 预平均 (pre-averaging) 方法：见第 10.3 节的预平均方差-协方差估计量。
- 文献：
 - Zhou (1996) 首先讨论独立噪声下的核估计 ($H = 1$)，在 i.i.d. 噪声下无偏但不一致；
 - Hansen and Lunde (2006) 研究其性质，并指出非一致核估计仍能揭示噪声特征；
 - Barndorff-Nielsen et al. (2008) 给出系统理论框架；
 - Jacod et al. (2009) 提出预平均方法；
 - Mykland and Zhang (2012)、Barndorff-Nielsen et al. (2008) 等系统综述高频波动率估计与相关技术。

目录

- 1 布朗运动
- 2 布朗运动的进一步性质
- 3 伊藤引理 (Itô's Lemma)
- 4 随机积分
- 5 扩散过程
- 6 扩散过程的极大似然估计
 - 布莱克–舒尔斯模型的极大似然估计
 - Aït-Sahalia 似然展开：近似 MLE
 - 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断
- 7 从高频数据中估计波动率
- 8 测量误差模型与微观结构噪声
- 9 **基于模拟的 RV-IQ 与 CLT 演示**
- 10 高频协方差矩阵估计方法
 - 双频已实现协方差估计量 (TSRV)
 - 多元已实现核协方差估计量 (RK)
 - 预平均协方差估计量 (PAV)
- 11 章节总结
- 12 参考文献

高频 RV 与 IQ: CLT 的实证演示

- 本节通过模拟展示:

- 如何在等间隔网格上计算 $RV_n = \sum(\Delta X_i)^2$;
- 如何构造积分四次方估计量

$$IQ = \frac{n}{3} \sum (\Delta X_i)^4;$$

- 利用

$$\sqrt{n} \frac{RV_n - QV}{\sqrt{2} IQ} \Rightarrow N(0, 1)$$

构造 QV 的渐近置信区间;

- 在模拟的布朗半鞅下检验 CLT 的近似与区间覆盖率。
- 也可以将模拟代码替换为实际数据（例如从 Yahoo Finance 获取 SPY 的复权收盘价），思路完全类似。

基础函数：给定对数价格计算 RV、IQ 与置信区间

```

# =====
# 高频波动率估计 (RV、IQ) 与 CLT (Barndorff-Nielsen & Shephard)
# - 等间隔观测:  $X_0, X_{\{1/n\}}, \dots, X_1$ 
# -  $RV_n = \sum (\Delta X_i)^2$ 
# -  $IQ_{hat} = (n/3) * \sum (\Delta X_i)^4$ 
# - CLT:  $sqrt(n) (RV_n - QV) / sqrt(2 * IQ) \rightarrow N(0, 1)$ 
# - 置信区间:
#    $QV \quad RV \pm z_{\{ /2} * sqrt(2 * IQ_{hat} / n)$ 
# =====

set.seed(123)

# 1) 基础函数: 给定等间隔对数价格 x, 计算 RV、IQ 与 CI
rv_iq_ci <- function(x, level = 0.95) {
  # x: 等间隔对数价格向量 (长度 n+1)
  n <- length(x) - 1
  dX <- diff(x)                      #  $\Delta X_i$ 
  RV <- sum(dX^2)
  IQ_hat <- (n/3) * sum(dX^4)
  # 基于 CLT 的标准误 (用 RV 估 QV、用 IQ_hat 估 IQ)
  se <- sqrt(2 * IQ_hat / n)
  z <- qnorm((1 + level) / 2)
  ci <- c(RV - z * se, RV + z * se)
  list(RV = RV, IQ_hat = IQ_hat, SE = se, CI = ci)
}

```

模拟布朗半鞅路径：无杠杆情形

```

# 2) 模拟一个“无杠杆”的布朗半鞅：
#    $X_t = \int_0^t s ds + \int_0^t s dB_s$ 
#    $s$  与  $B$  独立， $s$  cadlag
#   这里取： $s = 0; s^2 = 0.04 + 0.02 * \sin(2\pi s)$ 

simulate_semimartingale <- function(n) {
  # 等间隔网格： $t_i = i/n$ 
  t <- seq(0, 1, length.out = n + 1)
  dt <- 1 / n
  # 漂移设为 0 (也可改为缓变函数)
  mu <- rep(0, n)
  # 时间变动的波动率 (与  $B$  独立, 这里取确定性函数)
  sigma2 <- 0.04 + 0.02 * sin(2 * pi * t[-1]) # 使用右端点  $\{t_i\}$ 
  sigma <- sqrt(pmax(sigma2, 1e-10))
  # 标准正态增量  $i$ 
  eps <- rnorm(n)
  #  $\Delta X_i = i * dt + i * \sqrt(dt) * \epsilon_i$ 
  dX <- mu * dt + sigma * sqrt(dt) * eps
  # 对数价格路径 (假设  $X_0 = 0$ )
  X <- c(0, cumsum(dX))
  # 真 QV =  $\int_0^1 s^2 ds$ , 用黎曼和近似
  QV_true <- sum(sigma^2) * dt
  list(path = X, QV_true = QV_true, sigma2 = sigma2, t = t)
}

```

单次模拟：RV、IQ 与标准化统计量

```
# 3) 单次模拟：计算 RV、IQ、标准化统计量与置信区间
single_run <- function(n = 23 * 60) {
  # n: 把 [0,1] 划分为 n 段，可视作 1 天中 n 次高频观测
  sim <- simulate_semimartingale(n)
  out <- rv_iq_ci(sim$path, level = 0.95)
  # 基于真 QV 的标准化统计量（用于检验 CLT）
  Z_true <- sqrt(n) * (out$RV - sim$QV_true) /
    sqrt(2 * (sum(diff(sim$path)^4) * n / 3))
  list(
    n      = n,
    RV     = out$RV,
    IQ_hat = out$IQ_hat,
    SE     = out$SE,
    CI     = out$CI,
    QV_true = sim$QV_true,
    Z_true = Z_true
  )
}

# 示例：把 [0,1] 视作“一天”，n = 23*60 每分钟一笔
res1 <- single_run(n = 23 * 60)
cat("单次模拟结果：\n")
print(res1[c("n", "RV", "QV_true", "SE", "CI", "Z_true")])
```

蒙特卡洛检验：CLT 与置信区间覆盖率

```

# 4) 多次重复：检验 CLT 与置信区间覆盖率
mc_check <- function(n = 23 * 60, R = 500, level = 0.95) {
  z <- qnorm((1 + level) / 2)
  cover <- numeric(R)
  Z <- numeric(R)
  for (r in 1:R) {
    sim <- simulate_semimartingale(n)
    x <- sim$path
    out <- rv_iq_ci(x, level = level)
    # 检查 CI 是否覆盖真 QV
    cover[r] <- (sim$QV_true >= out$CI[1] &&
                  sim$QV_true <= out$CI[2])
    # 基于真 QV 的标准化统计量 (查看是否接近 N(0,1))
    Z[r] <- sqrt(n) * (out$RV - sim$QV_true) /
      sqrt(2 * out$IQ_hat)
  }
  list(
    n       = n,
    R       = R,
    level   = level,
    coverage = mean(cover),
    Z_mean  = mean(Z),
    Z_sd    = sd(Z)
  )
}

res_mc <- mc_check(n = 23 * 60, R = 200, level = 0.95)
cat("\n蒙特卡洛检验：\n")
print(res_mc)

```

目录

- 1 布朗运动
- 2 布朗运动的进一步性质
- 3 伊藤引理 (Itô's Lemma)
- 4 随机积分
- 5 扩散过程
- 6 扩散过程的极大似然估计
 - 布莱克-舒尔斯模型的极大似然估计
 - Aït-Sahalia 似然展开：近似 MLE
 - 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断
- 7 从高频数据中估计波动率
- 8 测量误差模型与微观结构噪声
- 9 基于模拟的 RV-IQ 与 CLT 演示
- 10 高频协方差矩阵估计方法
 - 双频已实现协方差估计量 (TSRV)
 - 多元已实现核协方差估计量 (RK)
 - 预平均协方差估计量 (PAV)
- 11 章节总结
- 12 参考文献

高频协方差矩阵估计：动机与挑战

- 第 ?? 节集中讨论了一维情形（单资产波动率）。
- 高频数据同样可用于估计**方差与协方差矩阵**，服务于多资产风险管理与资产配置。
- 影响高频方差-协方差估计的主要因素：
 - ① **做市商的双边报价**：同一时刻的买入价与卖出价不同，使得观测价格偏离“理想布朗运动”轨迹。
 - ② **价格离散性**：价格变动以最小报价单位为步长，剧烈波动时容易出现跳跃。
 - ③ **异步交易 (non-synchronous trading)**：多个资产的交易时间点通常不一致，导致数据不同步。
- **Epps 效应**：高频下，协方差矩阵的非对角元素趋近于零，导致相关性被系统性低估，进而影响投资组合构建、风险评估与相关性驱动的金融模型。

异步交易与时间同步问题

- 多资产高频数据常见的关键问题：时间不同步。
- 成因：
 - 交易频率差异；
 - 各资产流动性不同；
 - 不同交易时段重叠；
 - 报价和成交时间戳精度差异等。
- 若直接按“原始时间戳”计算跨资产协方差：
 - 非对角元素被严重低估（Epps 效应）；
 - 相关结构被扭曲，影响风险分解与对冲策略。
- 核心任务：在存在价格噪声与异步交易的前提下，构造**稳健且可行**的高维协方差估计量。

时间同步方法：插值 vs 删除

- 为解决多资产时间不同步，常见策略：

① 插值法：

- 用最近一次观测值（上一笔价格）填补；
- 或基于自回归模型预测缺失价格。

② 删除法：

- 剔除无法同步的时间点；
- 只保留在统一时间网格上对齐的数据。

- 实证研究表明：

- 插值法往往引入额外的平滑偏差；
- 删除法更稳健，因而成为主流选择。

- 最常用的删除法：**刷新时间同步法** (refresh time)，见 Harris et al. (1995)，并在第 10.2 节中进一步说明。

本节方法概览

- 高频方差-协方差估计还需显式考虑噪声降噪。
- 重点介绍三类主流稳健已实现协方差估计量：
 - ① 双频已实现协方差 (Two-Scale Realized Variance-Covariance, TSRV)
 - ② 多元已实现核协方差 (Realized Kernel Variance-Covariance, RK)
 - ③ 预平均协方差估计量 (Pre-averaging Variance-Covariance, PAV)
- 本节主要目标：
 - 给出三类方法的表达形式；
 - 展示对应的 R 语言实现；
 - 理论细节与渐近性质参见本章参考文献。

TSRV：基本思想与对角元

- 双频已实现协方差估计量 (TSRV) 由 Zhang (2011) 提出。
- **对角元**: 对应单资产方差, 思路与一维 TSRV 一致。
- 给定时间点 τ_1, \dots, τ_n 对应的 n 个观测价格构成对数价格序列 X , 定义

$$[X, X]_T^{(K)} = \frac{1}{K} \sum_{i=1}^{n-K+1} (X_{t_{i+K}} - X_{t_i})^2,$$

并令

$$\bar{n}_K = \frac{n - K + 1}{K}, \quad \bar{n}_J = \frac{n - J + 1}{J}.$$

- TSRV 方差估计量为

$$\left(1 - \frac{\bar{n}_K}{\bar{n}_J}\right)^{-1} \left([X, X]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} [X, X]_T^{(J)} \right).$$

- 核心思想: 利用**两种不同采样尺度 (K 与 J)**的 realized variance 做线性组合, 抵消噪声造成的偏差。

TSRV：异步交易与刷新时间

- **非对角元**: 协方差矩阵元素，需在多资产之间对齐时间。
- 时间同步采用 Harris et al. (1995) 的**刷新时间** (refresh time):
 - 第一个刷新时间: 所有资产首次均有成交的时间点;
 - 第二个刷新时间: 自上一次刷新时间之后, 所有资产再次完成交易的最早时间;
 - 如此递推直至样本结束。
- 通过刷新时间构造的统一时间网格, 可用 R 包 `highfrequency` 中的 `refreshTime` 函数实现。
- 在刷新时间网格下, 定义配对收益率 $\{(X_{t_i}, Y_{s_i})\}$, 从而构造多资产协方差的 TSRV。

TSRV: 协方差项与半正定问题

- 对两个对数价格序列 X 与 Y , 设各自的交易时间集合为

$$\Gamma = \{\tau_1, \dots, \tau_{N_T}^X\}, \quad \Theta = \{\theta_1, \dots, \theta_{N_T}^Y\}.$$

- 第一个刷新时间:

$$\phi_1 = \max(\tau_1, \theta_1).$$

- 后续刷新时间递推:

$$\phi_{j+1} = \max(\tau_{N_{\phi_j}^X + 1}, \theta_{N_{\phi_j}^Y + 1}).$$

- 对应采样点:

$$t_i = \max\{\tau \in \Gamma : \tau \leq \phi_i\}, \quad s_i = \max\{\theta \in \Theta : \theta \leq \phi_i\}.$$

TSRV：协方差项与半正定问题

- 协方差项：

$$c_N \left([X, Y]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} [X, Y]_T^{(J)} \right),$$

其中

$$[X, Y]_T^{(K)} = \frac{1}{K} \sum_{i=1}^{M_N-K+1} (X_{t_{i+K}} - X_{t_i})(Y_{s_{i+K}} - Y_{s_i}).$$

- 注意：TSRV 协方差矩阵并不总是半正定；使用 `highfrequency::rTSCov` 时，可设置 `makePsd = TRUE` 将负特征值截断为零。

TSRV 的 R 实现示例 (1/2)

```
# TSRV: 双频已实现协方差 (简化版示例)
TSRV <- function(priceX, slows, fasts) {
  # 对数价格
  lprice <- log(priceX)

  # 子网格大小
  bin11 <- floor(nrow(lprice) / slows)
  bin22 <- floor(nrow(lprice) / fasts)

  # 慢频已实现协方差
  rv_slow <- matrix(0, ncol = ncol(lprice), nrow = ncol(lprice))
  for (j in 1:slows) {
    cut <- seq(j + nrow(lprice) %% slows, by = slows, length.out = bin11)
    cut1 <- lprice[cut, ]
    diff_cut1 <- diff(cut1)
    rv_slow <- rv_slow + t(diff_cut1) %*% diff_cut1
  }
  rv_slow <- rv_slow / slows
}
```

TSRV 的 R 实现示例 (2/2)

```
# 快频已实现协方差
rv_fast <- matrix(0, ncol = ncol(lprice), nrow = ncol(lprice))
for (j in 1:fasts) {
  cut <- seq(j + nrow(lprice) %% fasts, by = fasts, length.out = bin22)
  cut1 <- lprice[cut, ]
  diff_cut1 <- diff(cut1)
  rv_fast <- rv_fast + t(diff_cut1) %*% diff_cut1
}
rv_fast <- rv_fast / fasts

# TSRV 估计量
TSRV_mat <- rv_slow - (fasts / slows) * rv_fast

# 年化
TSRV_mat * 252
}
```

RK：从单变量到多元

- 已实现核方法最早用于单变量稳健波动率估计 (Barndorff-Nielsen et al. 2008)。
- 随后, Barndorff-Nielsen et al. (2011) 将其推广至多元协方差矩阵估计。
- 优点：
 - 能处理内生性与序列相关的噪声;
 - 通过核平滑减弱噪声影响;
 - 可构造半正定的协方差估计矩阵。
- 多元 RK 依托刷新时间生成同步价格序列, 再基于高频收益率构造自协方差序列并施加核函数加权。

刷新时间：形式化定义

定义 (刷新时间, Multivariate Refresh Time)

对 $t \in [0, 1]$, 记第 i 个资产的观测时间为 $t_1^{(i)}, \dots, t_{n^{(i)}}^{(i)}$ 。首次刷新时间为

$$\tau_1 = \max(t_1^{(1)}, \dots, t_1^{(d)}),$$

随后递推

$$\tau_{j+1} = \max(t_{N_{\tau_j}^{(1)} + 1}^{(1)}, \dots, t_{N_{\tau_j}^{(d)} + 1}^{(d)}), \quad (11)$$

直到样本结束, 得到刷新时间序列 $\{\tau_j\}$, 总数记为 N 。

刷新时间：形式化定义

- 直观解释：
 - τ_1 是所有资产首次完成交易的“最晚”时刻；
 - τ_{j+1} 是自 τ_j 之后，各资产再次完成一次交易的最晚时刻。
- 保留数据比例可用 $p = dN / \sum_{i=1}^d n^{(i)}$ 衡量；维数越高、不同步越严重， p 越低。

端点平滑与高频收益率

- 已实现核以刷新后的 N 个同步价格生成 n 个高频收益率为基础。
- 为改善开盘与收盘处的端点效应, Barndorff-Nielsen et al. (2011) 建议对端点做平均:

$$n - 1 + 2m = N, \quad m \text{ 适中且远小于 } n.$$

- 令向量观测

$$X_j = X(\tau_{j+m}), \quad j = 1, \dots, n-1,$$

且

$$X_0 = \frac{1}{m} \sum_{j=1}^m X(\tau_j), \quad X_n = \frac{1}{m} \sum_{j=1}^m X(\tau_{N-m+j}).$$

- 定义高频频向量收益率

$$r_j = X_j - X_{j-1}, \quad j = 1, \dots, n.$$

- 通常取 $m \approx 2$, 既平滑端点又不明显损失有效样本。

多元已实现核协方差估计量 (RK)

定义 (多元已实现核协方差估计量)

对同步后的高频收益率 $\{r_j\}_{j=1}^n$, 定义

$$K(X) = \sum_{h=-n}^n k\left(\frac{h}{H+1}\right) \Gamma_h,$$

其中 $k(\cdot)$ 为核函数, H 为带宽, Γ_h 为第 h 阶已实现自协方差:

$$\Gamma_h = \begin{cases} \sum_{j=|h|+1}^n r_j r'_{j-h}, & h \geq 0, \\ \sum_{j=|h|+1}^n r_{j-h} r'_j, & h < 0. \end{cases}$$

多元已实现核协方差估计量 (RK)

- 通过核权重 $k(h/(H+1))$ 对不同滞后阶的自协方差进行加权平滑。
- 合适选择 H 可在偏差-方差之间取得平衡。

核函数的条件

核函数 k 应满足以下条件：

- ① $k(0) = 1, \quad k'(0) = 0;$
- ② k 二阶可导且导数连续，以保证平滑性；
- ③ 矩有界条件：

$$\int_0^1 k(x)^2 dx < \infty, \quad \int_0^1 k'(x)^2 dx < \infty, \quad \int_0^1 k''(x)^2 dx < \infty;$$

- ④ 对任意 $\lambda \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} k(x) e^{ix\lambda} dx \geq 0,$$

从而保证由 k 生成的相关矩阵为半正定。

- 实际中常用 Bartlett 核、Parzen 核等。
- RK 的优势之一：在噪声存在且具有内生性与序列相关时，仍能提供一致且半正定的协方差估计。

RK 的 R 实现示例 (Parzen 核) (1/2)

```
# RK: 多元已实现核协方差 (Parzen 核示例)
RK <- function(priceX, slows, fasts) {
  lprice <- log(priceX)

  bin11 <- floor(nrow(lprice) / slows)
  bin22 <- floor(nrow(lprice) / fasts)

  # 慢频 RV: 估计“信号”方差
  rv_slow <- matrix(0, nrow = slows, ncol = ncol(lprice))
  for (j in 1:slows) {
    cut <- seq(j + nrow(lprice) %% slows, length.out = bin11, by = slows)
    cut1 <- lprice[cut, , drop = FALSE]
    rv_slow[j, ] <- colSums(diff(cut1)^2)
  }
  rv_slow_mean <- colMeans(rv_slow)

  # 快频 RV: 估计噪声方差
  rv_fast <- matrix(0, nrow = fasts, ncol = ncol(lprice))
  for (j in 1:fasts) {
    cut <- seq(j + nrow(lprice) %% fasts, length.out = bin22, by = fasts)
    cut1 <- lprice[cut, , drop = FALSE]
    rv_fast[j, ] <- colSums(diff(cut1)^2)
  }
  rv_fast_mean <- colMeans(rv_fast)

  # 噪声方差估计
  noisev <- rv_fast_mean / ((bin22 - 1) * 2)
}
```

RK 的 R 实现示例 (Parzen 核) (2/2)

```

# 平滑窗口宽度 h (基于 MSE 近似公式)
h <- round(mean(3.51 * (noisev / rv_slow_mean)^0.4 *
                 (nrow(lprice) - 1)^0.6))

# 高频收益率
r <- diff(lprice)

# 已实现核协方差
RK_mat <- t(r) %*% r
for (hh in 1:h) {
  x <- (hh - 1) / h
  if (x <= 0.5) {
    k <- 1 - 6 * x^2 + 6 * x^3
  } else {
    k <- 2 * (1 - x)^3
  }
  RK_mat <- RK_mat +
    k * ( t(r[(hh+1):nrow(r), , drop = FALSE]) %*% r[1:(nrow(r)-hh), , drop = FALSE] +
           t(r[1:(nrow(r)-hh), , drop = FALSE]) %*% r[(hh+1):nrow(r), , drop = FALSE] )
}
# 年化
RK_mat * 252
}

```

PAV：动机与整体思路

- 预平均协方差估计量 (PAV) 由 Christensen, Kinnebrock, and Podolskij (2010) 与 Hautsch and Podolskij (2013) 在 Jacod et al. (2009) 基础上发展。
- 结合了 TSRV 与 RK 的优点：
 - 首先对原始对数价格/收益序列做局部平滑 (预平均)，减弱噪声影响；
 - 再用高频协方差作为噪声估计作差进行偏差修正。
- 考虑多维对数价格模型 (测量误差结构)：

$$Y_\tau = X_\tau + \varepsilon_\tau,$$

其中 X 为潜在布朗半鞅， ε_τ 为 i.i.d. 噪声，且与 X 独立。

- 直观：在零均值 i.i.d. 噪声下，局部加权平均可“平均掉”噪声，从而恢复 X 的连续部分。

单变量预平均：预平均收益与方差估计

- 设某一刷新时段包含 N 个等间隔收益率观测 r_{τ_i} 。
- 定义预平均收益率（单资产）：

$$\bar{r}_{\tau_j}^{(k)} = \sum_{h=1}^{k_N-1} g\left(\frac{h}{k_N}\right) r_{\tau_{j+h}}^{(k)}, \quad (12)$$

其中

$$g: [0, 1] \rightarrow \mathbb{R}, \quad g(x) = \min(x, 1 - x),$$

且 $k_N = \lfloor \theta N^{1/2} \rfloor$ (Hautsch and Podolskij (2013) 建议 $\theta \approx 0.8$)。

单变量预平均：预平均收益与方差估计

- 预平均方差估计量为

$$\hat{C} = \frac{N^{-1/2}}{\theta\psi_2} \sum_{i=0}^{N-k_N+1} r_{\tau_i}^2 - \frac{\psi_1^{k_N} N^{-1}}{2\theta^2\psi_2^{k_N}} \sum_{i=0}^N r_{\tau_i}^2,$$

其中

$$\psi_1^{k_N} = k_N \sum_{j=1}^{k_N} \left(g\left(\frac{j+1}{k_N}\right) - g\left(\frac{j}{k_N}\right) \right)^2,$$

$$\psi_2^{k_N} = \frac{1}{k_N} \sum_{j=1}^{k_N-1} g^2\left(\frac{j}{k_N}\right), \quad \psi_2 = \frac{1}{12}.$$

多变量预平均协方差：PAV 与 PAV^δ

- 多变量预平均协方差估计量：

$$PAV = \frac{N}{N - k_N + 2} \frac{1}{\psi_2 k_N} \sum_{i=0}^{N-k_N+1} \bar{\mathbf{r}}_{\tau_i} \bar{\mathbf{r}}'_{\tau_i} - \frac{\psi_1^{k_N}}{\theta^2 \psi_2^{k_N}} \hat{\Psi}_N,$$

其中

$$\hat{\Psi}_N = \frac{1}{2N} \sum_{i=1}^N \mathbf{r}_{\tau_i} \mathbf{r}'_{\tau_i},$$

为偏差修正项。

- 问题：偏差修正可能导致矩阵并非半正定。
- 实务解决方案：适度增大带宽，取

$$k_N = \lfloor \theta N^{1/2+\delta} \rfloor, \quad \delta > 0,$$

在 $\delta = 0.1$ 时，可得到无需偏差修正也一致且半正定的估计量：

$$PAV^\delta = \frac{N}{N - k_N + 2} \frac{1}{\psi_2 k_N} \sum_{i=0}^{N-k_N+1} \bar{\mathbf{r}}_i \bar{\mathbf{r}}'_i.$$

PAV 的 R 实现与半正定化

- 在 R 中可以使用 `highfrequency` 包的 `rMRCov` 函数计算 PAV (本书中记作 PAV)。
- 由于包含偏差校正项, 估计矩阵不一定半正定; 因此在函数调用中可设 `makePsd = TRUE` 进行半正定化。
- 下面是一个简化的 PAV 实现示例 (单一 θ , 含偏差修正):

PAV 的 R 实现示例

```

# PAV: 预平均协方差估计量 (简化版)
PAV <- function(priceX, theta) {
  lprice <- log(priceX)
  delta_n <- 1 / nrow(lprice)
  K_n    <- round(theta / sqrt(delta_n) / 2) * 2

  # Jacob 等人的常数
  psi1 <- 1
  psi2 <- 1 / 12
  psi_K <- (1 + 2 / K_n^2) / 12

  Z <- matrix(0, ncol = ncol(lprice), nrow = ncol(lprice))

  # 预平均收益并累加外积
  for (j in 1:(nrow(lprice) - K_n + 1)) {
    r <- (1 / K_n) * colSums(
      lprice[(j + K_n/2):(j + K_n - 1), ] -
      lprice[j:(j + K_n/2 - 1), ])
    Z <- Z + r %*% t(r)
  }

  # PAV 估计量 (带偏差修正)
  CX <- sqrt(delta_n) / (theta * psi2) * Z -
    psi1 * delta_n / (2 * theta^2 * psi2) *
    t(diff(lprice)) %*% diff(lprice)

  # 年化
  CX * 252
}

```

TSRV vs RK vs PAV: 比较小结

- **TSRV:**

- 通过不同尺度下的 realized covariance 作差降噪；
- 能同时削弱 Epps 效应与微观结构噪声；
- 协方差矩阵不一定半正定，常需后处理。

- **RK:**

- 引入核平滑，允许噪声内生性与序列相关；
- 理论条件相对宽松；
- 可构造半正定协方差矩阵。

- **PAV:**

- 先预平均平滑，再偏差修正；
- 收敛速度通常较快；
- 带偏差修正时不一定半正定，可通过带宽调节构造 PAV^δ 。

GMVP：基于估计协方差构建组合

- 得到高频方差-协方差估计矩阵 Σ 后，可构造全局最小方差投资组合 (GMVP)。
- 目标：最小化组合方差

$$\min_{\mathbf{w}} \mathbf{w}' \Sigma \mathbf{w},$$

在约束 $\sum_{i=1}^n w_i = 1$ 下。

- 闭式解（允许卖空）：

$$\mathbf{w} = \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}' \Sigma^{-1} \mathbf{1}},$$

其中 $\mathbf{1}$ 为 n 维全 1 列向量。

- 若估计协方差矩阵为半正定，则优化问题良好，可直接求解。

GMVP 权重：无约束情形的 R 代码

```
# 基于协方差矩阵的 GMVP 权重 (允许卖空)
compute_mvp_weights_simple <- function(cov_matrix) {
  n      <- nrow(cov_matrix)          # 资产数
  ones   <- rep(1, n)
  inv_cov <- solve(cov_matrix)       # 协方差矩阵的逆

  num <- inv_cov %*% ones
  den <- as.numeric(t(ones) %*% inv_cov %*% ones)

  weights <- num / den
  as.numeric(weights)
}
```

GMVP：带“禁止卖空”约束的二次规划

- 若市场不允许卖空：

$$\min_{\mathbf{w}} \mathbf{w}' \Sigma \mathbf{w},$$

约束为

$$\sum_{i=1}^n w_i = 1, \quad w_i \geq 0, \quad \forall i.$$

- 可用 quadprog 包中的 solve.QP 求解：

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}' D \mathbf{w} + d' \mathbf{w}$$

在适当的线性约束下（等式 + 不等式）。

- 下一页给出对应的 R 实现示例。

GMVP 权重：禁止卖空的 R 实现示例

```
# GMVP 权重（禁止卖空）：二次规划示例
# 需要 quadprog 包: library(quadprog)
compute_mvp_weights <- function(cov_matrix) {
  n <- nrow(cov_matrix)

  Dmat <- 2 * cov_matrix           # 目标中的二次项
  dvec <- rep(0, n)                # 线性项为 0

  # 约束: sum(w_i) = 1, w_i >= 0
  # solve.QP 使用形式: Amat' * w >= bvec
  # 先构建 sum(w_i) = 1 作为等式约束，接着是 w_i >= 0
  Amat <- cbind(rep(1, n), diag(n)) # 第一列是 sum(w_i)，后面是 w_i
  bvec <- c(1, rep(0, n))          # sum(w_i) >= 1 且 w_i >= 0
  meq <- 1                          # 第一条为等式约束

  sol <- quadprog::solve.QP(Dmat, dvec, Amat, bvec, meq = meq)
  sol$solution
}
```

目录

- 1 布朗运动
- 2 布朗运动的进一步性质
- 3 伊藤引理 (Itô's Lemma)
- 4 随机积分
- 5 扩散过程
- 6 扩散过程的极大似然估计
 - 布莱克-舒尔斯模型的极大似然估计
 - Aït-Sahalia 似然展开：近似 MLE
 - 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断
- 7 从高频数据中估计波动率
- 8 测量误差模型与微观结构噪声
- 9 基于模拟的 RV-IQ 与 CLT 演示
- 10 高频协方差矩阵估计方法
 - 双频已实现协方差估计量 (TSRV)
 - 多元已实现核协方差估计量 (RK)
 - 预平均协方差估计量 (PAV)
- 11 章节总结
- 12 参考文献

本章回顾：从基础到工具

- 本章围绕连续时间序列模型，搭建了由

“基础 → 工具 → 模型 → 估计 → 应用”

构成的完整链条。

- **基础部分：**

- 以布朗运动为起点，系统阐述其独立增量、正态增量与路径处处不可导等性质；
- 引入局部 Hölder 连续性与穿越时间等结果，形成概率与路径两方面直觉。

- **工具部分：**

- 通过随机游走极限定理与等距性质严格定义随机积分；
- 使用伊藤引理将“随机版链式法则”形式化；
- 使得 SDE 中含漂移与扩散项的过程可以进行函数变换与微分运算。

本章回顾：模型估计与设定检验

- 模型层面：

- 以广义布朗运动为核心，将扩散过程写为

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t,$$

并给出解存在与唯一的充分条件。

- 估计层面：

- 在可获得闭式转移密度时，给出 Black–Scholes/几何布朗运动模型的极大似然估计；
- 在转移密度不可得时，引入 Aït–Sahalia 的封闭形式似然展开：通过 $X \rightarrow Y \rightarrow Z$ 的正态化变换与 Hermite 展开构造近似似然，并论证其与“真实” MLE 的一致性与渐近等价。

- 设定检验：

- 基于边际/转移密度的一致性检验；
- 或用非参数估计的转移密度与假设模型对比；
- 为参数估计提供模型诊断与稳健性保障。

本章回顾：高频波动率与协方差

● 一维高频：

- 已实现波动率 (RV) 可一致逼近二次变差；
- 配合 \hat{IQ} 构造 CLT 下的标准化与置信区间；
- 指出微观结构噪声导致高频 RV 上行偏差，需通过降频或两尺度/多尺度修正 (TSRV、MSRV) 校正。

● 多维高频：

- 讨论异步交易引发的 Epps 效应与时间同步问题；
- 引入刷新时间与公共网格对齐；
- 详细介绍 TSRV、RK 与 PAV 三类协方差估计量及 R 实现。

本章回顾：从估计到资产配置

- 实务上，需明确：
 - 取样步长与年化口径；
 - 时间同步规则（刷新时间 vs 公共网格）；
 - TSRV 的 K/J 参数与 PAV 的带宽 θ, δ ；
 - 是否需要对协方差矩阵做半正定化处理。
- 在此基础上：
 - 可以求解 GMVP 等组合优化问题；
 - 对风险进行分解与度量；
 - 为更复杂的结构（跳跃-扩散、随机波动率、仿射模型等）打下基础。
- 后续章节可在本章的“**随机基础 → 计量方法 → 高频估计 → 组合应用**”主线之上，进一步扩展至期权定价与收益率曲线建模等领域。

目录

- 1 布朗运动
- 2 布朗运动的进一步性质
- 3 伊藤引理 (Itô's Lemma)
- 4 随机积分
- 5 扩散过程
- 6 扩散过程的极大似然估计
 - 布莱克-舒尔斯模型的极大似然估计
 - Aït-Sahalia 似然展开：近似 MLE
 - 案例：基于 Aït-Sahalia 展开的近似 MLE 与稳健推断
- 7 从高频数据中估计波动率
- 8 测量误差模型与微观结构噪声
- 9 基于模拟的 RV-IQ 与 CLT 演示
- 10 高频协方差矩阵估计方法
 - 双频已实现协方差估计量 (TSRV)
 - 多元已实现核协方差估计量 (RK)
 - 预平均协方差估计量 (PAV)
- 11 章节总结
- 12 参考文献

参考文献 I

-  Aït-Sahalia, Yacine (2002). "Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach". In: *Econometrica* 70.1, pp. 223–262.
-  — (1996). "Testing continuous-time models of the spot interest rate". In: *The Review of Financial Studies* 9.2, pp. 385–426.
-  Aït-Sahalia, Yacine, Jianqing Fan, and Heng Peng (2009). "Nonparametric transition-based tests for jump diffusions". In: *Journal of the American Statistical Association* 104.487, pp. 1102–1116.
-  Aït-Sahalia, Yacine, Per A Mykland, and Lan Zhang (2011). "Ultra high frequency volatility estimation with dependent microstructure noise". In: *Journal of Econometrics* 160.1, pp. 160–175.
-  Andersen, Torben G et al. (2003). "Modeling and forecasting realized volatility". In: *Econometrica* 71.2, pp. 579–625.
-  Bachelier, Louis (1900). "Théorie de la spéculation". In: *Annales Scientifiques de l'École Normale Supérieure*. Vol. 17, pp. 21–86.

参考文献 II

-  Barndorff-Nielsen, Ole E and Neil Shephard (2002). "Estimating quadratic variation using realized variance". In: *Journal of Applied Econometrics* 17.5, pp. 457–477.
-  Barndorff-Nielsen, Ole E et al. (2008). "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise". In: *Econometrica* 76.6, pp. 1481–1536.
-  — (2011). "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading". In: *Journal of Econometrics* 162.2, pp. 149–169.
-  Campbell, John Y, Andrew W Lo, and A Craig MacKinlay (1997). *The econometrics of financial markets*. Princeton University press.
-  Chen, Song Xi, Jiti Gao, and Cheng Yong Tang (2008). "A test for model specification of diffusion processes". In: *The Annals of Statistics* 36.1.

参考文献 III

-  Christensen, Kim, Silja Kinnebrock, and Mark Podolskij (2010). "Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data". In: *Journal of Econometrics* 159.1, pp. 116–133.
-  Constantinides, George M (1992). "A theory of the nominal term structure of interest rates". In: *The Review of Financial Studies* 5.4, pp. 531–552.
-  Courtadon, G. (1982). "The pricing of options on default-free bonds". In: *Journal of Financial and Quantitative Analysis*, pp. 75–100.
-  Cox, D. R. (1975). "Partial likelihood". In: *Biometrika* 62.2, pp. 269–276.
-  Cox, John C, Jonathan E Ingersoll Jr, and Stephen A Ross (1985). "An intertemporal general equilibrium model of asset prices". In: *Econometrica*, pp. 363–384.

参考文献 IV

-  Dai, Qiang and Kenneth J Singleton (2000). "Specification analysis of affine term structure models". In: *The Journal of Finance* 55.5, pp. 1943–1978.
-  Darling, Donald A and Arnold JF Siegert (1953). "The first passage problem for a continuous Markov process". In: *The Annals of Mathematical Statistics*, pp. 624–639.
-  Duffie, Darrell and Rui Kan (1996). "A yield-factor model of interest rates". In: *Mathematical Finance* 6.4, pp. 379–406.
-  Feller, William (1991). *An introduction to probability theory and its applications, Volume 2*. Vol. 81. John Wiley & Sons.
-  Hansen, Peter R and Asger Lunde (2006). "Realized variance and market microstructure noise". In: *Journal of Business & Economic Statistics* 24.2, pp. 127–161.

参考文献 V

-  Harris, Frederick H. deB et al. (1995). "Cointegration, error correction, and price discovery on informationally linked security markets". In: *Journal of Financial and Quantitative Analysis* 30.4, pp. 563–579.
-  Hautsch, Nikolaus and Mark Podolskij (2013). "Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: theory, implementation, and empirical evidence". In: *Journal of Business & Economic Statistics* 31.2, pp. 165–183.
-  Hong, Yongmiao and Haitao Li (2005). "Nonparametric specification testing for continuous-time models with applications to term structure of interest rates". In: *The Review of Financial Studies* 18.1, pp. 37–84.
-  Jacod, Jean and Philip Protter (1998). "Asymptotic error distributions for the Euler method for stochastic differential equations". In: *The Annals of Probability* 26.1, pp. 267–307.

参考文献 VI

-  Jacod, Jean et al. (2009). "Microstructure noise in the continuous case: the pre-averaging approach". In: *Stochastic Processes and Their Applications* 119.7, pp. 2249–2276.
-  Jensen, Bjarke and Rolf Poulsen (1999). *A comparison of approximation techniques for transition densities of diffusion processes*. Citeseer.
-  Linton, Oliver (2019). *Financial econometrics*. Cambridge University Press.
-  Marsh, T. A. and E. R. Rosenfeld (1983). "Stochastic processes for interest rates and equilibrium bond prices". In: *The Journal of Finance* 38.2, pp. 635–646.
-  Mykland, Per A. and Lan Zhang (2012). "The Econometrics of High-Frequency Data". In: *Statistical Methods for Stochastic Differential Equations*. Ed. by M. Kessler, A. Lindner, and M. Sørensen. Boca Raton: Chapman and Hall/CRC.

参考文献 VII

-  Pedersen, Asger Roer (1995). "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations". In: *Scandinavian Journal of Statistics*, pp. 55–71.
-  Vasicek, Oldrich (1977). "An equilibrium characterization of the term structure". In: *Journal of Financial Economics* 5.2, pp. 177–188.
-  Zhang, Lan (2006). "Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach". In: *Bernoulli* 12.6, pp. 1019–1043.
-  — (2011). "Estimating covariation: Epps effect, microstructure noise". In: *Journal of Econometrics* 160.1, pp. 33–47.
-  Zhang, Lan, Per A Mykland, and Yacine Aït-Sahalia (2005). "A tale of two time scales: Determining integrated volatility with noisy high-frequency data". In: *Journal of the American Statistical Association* 100.472, pp. 1394–1411.

参考文献 VIII

-  Zhou, Bin (1996). "High-frequency data and volatility in foreign-exchange rates". In: *Journal of Business & Economic Statistics* 14.1, pp. 45–52.

第 10 章收益率曲线 (Yield Curve)

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

本章主题：收益率曲线与货币的时间价值

- 本章讨论利率与货币时间价值 (time value of money) 的度量问题。
- 收益率曲线在诸多应用中至关重要：
 - 资产配置；
 - 未来利率预测 (Campbell and Shiller, 1991)；
 - 通胀与国民收入分析 (Estrella and Mishkin, 1997)；
 - 利率衍生品定价与理性预期检验。
- 本章还将讨论收益率曲线的离散时间建模方法。
- 关键对象：
 - 贴现因子 (discount factor)；
 - 零息收益率曲线 (spot curve)；
 - 远期利率曲线 (forward curve)。

收益率曲线的定义

定义 (收益率曲线 (Yield Curve))

收益率曲线描述在特定时点、信用质量相同而期限不同的债券到期收益率随期限变化所形成的曲线。

- 典型形态：
 - 上倾斜：长期利率高于短期利率；
 - 平坦：各期限利率相近；
 - 倒挂：长期利率低于短期利率。
- 倒挂通常被视为经济衰退风险上升的先行信号。
- 实务中关注三类曲线：
 - 零息收益率曲线 (zero curve)；
 - 票面收益率曲线 (par curve)；
 - 远期利率曲线 (forward curve)。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

零息债券与“干净”的利率度量

- 考虑零息债券：到期一次支付面值 M ，期间无息票。
- 零息债券的特点：
 - 将“时间价值”与票息频率、嵌入式期权等因素剥离；
 - 只涉及一个未来确定现金流与其现值的关系。
- 有息票债券定价：
 - 所有现金流可拆为若干面值为 1 的零息债券之和；
 - 因此先在零息债券上定义利率口径，再定价息票债。
- 零息收益率曲线是构建整个收益率曲线的基石。

现货利率 (spot rate) 的定义

定义 (现货利率 (spot rate) 的经济含义与口径)

现货利率 y_n 指在“今天”买入、于 n 期后收回面值 M 的零息债券之到期收益率 (按约定复利口径计算), 满足

$$p = \frac{M}{(1 + y_n)^n}.$$

- p : 当前价格; M : 到期面值; n : 剩余期数。
- 现货利率是期限的函数 $y_n = y(n)$, 而非常数。
- 所有 $\{y_1, \dots, y_n, \dots\}$ 共同构成零息收益率曲线。

现货利率定义的四点补充

① 口径依赖：

- 期数单位与复利频率需与 y_n 一致（年、半年、季度、月、日等）；
- 若采用连续复利 $y^{(c)}(\tau)$ ，则

$$p = M e^{-\tau y^{(c)}(\tau)}.$$

② 期限特异：

- 每个期限对应一个现货利率；
- $\{y_\tau\}$ 形成一条曲线，而非单一利率水平。

③ 无套利解释：

- 若两种无风险策略在 n 期末产生同额确定现金流，其现值必须一致；
- 给定 p 时， y_n 与 p 一一对应。

④ 与 YTM 的区分：

- 有息票债的到期收益率（YTM）是混合多个期限的内部收益率；
- 零息现货利率 y_n 只对应单一到期点。

贴现因子 (discount factor) 的定义

定义 (贴现因子 (discount factor) 的作用与性质)

贴现因子 d_n 是 n 期后 1 单位确定现金流的标准化价格：

$$d_n = \frac{1}{(1 + y_n)^n}.$$

- d_n 与零息利率 y_n 等价，可相互转换。
- 在连续复利下， $d(\tau) = \exp\{-\tau y^{(c)}(\tau)\}$ 。

贴现因子的四个性质

① 可加性与线性定价：

$$p = \sum_{j=1}^m b_{\tau_j} d_{\tau_j},$$

其中 b_{τ_j} 为 τ_j 时刻的确定现金流。

- $\{d_\tau\}$ 构成确定现金流定价的完整基。

② 单调性：

- 在无套利且利率非负时, d_τ 随 τ 单调递减;
- 负利率环境下, 整体仍倾向递减, 但可更平缓或局部变形。

③ 口径统一：

$$d(\tau) = \exp\{-\tau y^{(c)}(\tau)\}, \quad f(\tau) = -\frac{d}{d\tau} \ln d(\tau),$$

便于进行微分、积分及与远期利率的转换。

④ 插值与平滑：

- 实务上常对 $d(\tau)$ 进行平滑或正则化 (如样条、指数样条);
- 再推导 $y(\tau)$ 、 $f(\tau)$, 确保与线性定价一致。

远期利率 (forward rate) 的定义

定义 (远期利率 (forward rate) 的含义、计算与口径)

一期远期利率 f_n 是今天锁定、从第 n 期到第 $n+1$ 期的一期无风险借贷利率，满足

$$1 + f_n = \frac{d_n}{d_{n+1}}.$$

连续复利下的瞬时远期利率 $f(\tau)$ 定义为

$$f(\tau) = -\frac{d}{d\tau} \ln d(\tau), \quad y(\tau) = \frac{1}{\tau} \int_0^\tau f(s) ds.$$

远期利率的四点补充说明

① 离散无套利推导：

- 比较两种无风险策略：
 - (i) 借 n 期再滚动到 $n+1$ 期；
 - (ii) 直接借 $n+1$ 期。
- 令两者 $n+1$ 期末现金流相同，即得 $1 + f_n = d_n / d_{n+1}$ 。

② 度量含义：

- 瞬时远期 $f(\tau)$ 衡量“期限微增一小步”的价格；
- $y(\tau)$ 是区间 $[0, \tau]$ 上远期利率的算术平均。

③ 预期解释与测度：

- 在合适的远期测度下， $f(\tau)$ 等于未来短端利率的条件期望；
- 在物理或风险中性测度下，需要考虑风险溢价。

④ 数值稳定性：

- 对 $d(\tau)$ 数值微分以得 $f(\tau)$ 对噪声敏感；
- 实务中先对 $d(\tau)$ 平滑，再计算 $f(\tau)$ ；
- 或直接假设 $f(\tau)$ 属于某结构族（如 Nelson-Siegel）。

利用贴现因子对一般债券定价

- 若一债券在未来 τ_j 时刻支付 b_j (息票与本金), $j = 1, \dots, m$, 则无套利线性定价为:

$$P = \sum_{j=1}^m b_j d_{\tau_j} = \sum_{j=1}^m \frac{b_j}{(1 + y_{\tau_j})^{\tau_j}}.$$

- 零息收益率曲线 $\{y_\tau\}$ 、贴现曲线 $\{d_\tau\}$ 、远期曲线 $\{f_\tau\}$ 本质等价:
 - 给定其中任意一条, 在边界条件 $d(0) = 1, d(\infty) = 0$ 下可唯一推得其余两条。
- 估计与插值:
 - 在实际市场中, 先利用息票债价格抽取 $\{d_\tau\}$;
 - 再构造 $\{y_\tau\}$ 、 $\{f_\tau\}$ 用于定价与风险管理。

到期收益率 (Yield to Maturity)

定义 (到期收益率 (Yield to Maturity))

面值为 1、到期期限为 n 期、每期付息 c 的息票债券，其价格 p 与到期收益率 y 满足

$$p = \frac{c}{(1+y)} + \frac{c}{(1+y)^2} + \cdots + \frac{1}{(1+y)^n}.$$

若以面值成交 ($p = 1$)，则到期收益率随到期变化构成票面收益率曲线 (par curve)。

- YTM 是“一口价”地刻画整串现金流的内部收益率。
- 与零息现货利率相比，YTM 将不同期限贴现“混合”在一个数中。
- 在曲线建模中，通常先估计零息/贴现曲线，再反推各期限的 par yield。

贴现函数与连续复利收益率曲线

定义 (贴现函数与连续复利收益率曲线)

贴现函数 $d(\tau)$ 为未来 τ 时点 1 美元的现值。连续复利收益率曲线 $y(\tau)$ 与 $d(\tau)$ 的关系为

$$d(\tau) = \exp(-\tau y(\tau)), \quad y(\tau) = \frac{1}{\tau} \int_0^\tau f(s) ds,$$

其中 $f(s)$ 为瞬时远期利率。给定边界条件 $d(0) = 1, d(\infty) = 0$, d, y, f 可相互唯一转换, 且 $d(\tau)$ 单调递减。

- 连续复利口径在理论分析与数值计算中最为方便。
- 曲线平滑与微分操作都可以在 $d(\tau)$ 或 $g(\tau) = \tau y(\tau)$ 上进行。

FRED 数据上的“收益率曲线三件套”

- 数据来源：FRED 公布的美国国债常数到期收益率（DGS 系列）。
- 步骤概览：

- ① 从 FRED 获取同一观测日的各期限收益率（年化简单/债券口径）。
- ② 统一为连续复利 $y(\tau)$ ：

$$y^{(c)}(\tau) = \ln(1 + y_{\text{annual}}(\tau)).$$

- ③ 构造贴现函数：

$$d(\tau) = \exp\{-\tau y^{(c)}(\tau)\}.$$

- ④ 对 $g(\tau) = \tau y^{(c)}(\tau)$ 做样条平滑并求导，得到瞬时远期：

$$f(\tau) = \frac{d}{d\tau}\{\tau y^{(c)}(\tau)\}.$$

- ⑤ 计算相邻期限的一期离散远期：

$$1 + f_{i \rightarrow i+1} = \frac{d(\tau_i)}{d(\tau_{i+1})}.$$

- 输出：

- 按期限排列的表格：原始年化、连续复利、贴现、瞬时远期；
- 简要“三联图”： $y(\tau)$ 、 $d(\tau)$ 、 $f(\tau)$ 。

R 代码：收益率曲线三件套 (1/3)

```

# =====
# 收益率曲线三件套：y( ), d( ), f( ) — 以 FRED 美债常数到期为例
# 依赖: quantmod, xts, stats (base), utils
# 功能：
#   - 从 FRED 拉取常数到期收益率 (DGSx)
#   - 统一为“连续复利”口径：y_cc( )
#   - 计算贴现函数  $d( ) = \exp\{-y_{cc}( )\}$ 
#   - 以样条对  $g( ) = y_{cc}( )$  平滑并求导，得瞬时远期  $f( ) = g'( )$ 
#   - 计算相邻节点离散远期： $1 + f_{disc} = d(-i)/d(-i+1)$ 
#   - 结果打印，并作简单三联图
# =====

suppressPackageStartupMessages(library(quantmod))

# 1) 选择期限（单位：年）与 FRED 代码
taus <- c(0.25, 0.5, 1, 2, 3, 5, 7, 10, 20, 30)
fred <- c("DGS3MO", "DGS6MO", "DGS1", "DGS2", "DGS3",
        "DGS5", "DGS7", "DGS10", "DGS20", "DGS30")

# 2) 拉取 FRED 数据并对齐到“最后一个共同日期”
getSymbols(fred, src = "FRED", auto.assign = TRUE, warnings = FALSE)
Y <- do.call(merge, lapply(fred, get)) # 合并为 xts，一列一个期限
Yc <- Y[stats::complete.cases(Y)] # 各列都有值的日期
stopifnot(nrow(Yc) > 0)
last_row <- tail(Yc, 1)
asof <- index(last_row)
y_annual_pct <- as.numeric(last_row) # 百分比
y_annual <- y_annual_pct / 100 # 年化小数（简单口径）

```

R 代码：收益率曲线三件套

```

# 3) 统一口径：连续复利  $y_{cc}( ) = \ln(1 + y_{annual}( ))$ 
y_cc <- log(1 + y_annual)

# 4) 贴现函数  $d( ) = \exp\{-y_{cc}( )\}$ 
d_tau <- exp(-taus * y_cc)

# 5) 瞬时远期： $f( ) = d/d\{y_{cc}( )\}$  (样条平滑  $g( ) = y_{cc}( )$  再求导)
g_tau <- taus * y_cc
fit <- smooth.spline(x = taus, y = g_tau, spar = NULL)
gprime <- predict(fit, x = taus, deriv = 1)$y
f_tau <- gprime

# 6) 相邻期限的一期“离散远期”： $1 + f_{disc} = d(-i)/d(-\{i+1\})$ 
f_disc <- rep(NA_real_, length(taus) - 1)
for (i in 1:(length(taus) - 1)) {
  f_disc[i] <- d_tau[i] / d_tau[i + 1] - 1
}

```

R 代码：收益率曲线三件套 (2/3)

```

# 7) 结果表：连续复利 y_cc、贴现 d_tau、瞬时远期 f_tau；  

# 并保留原年化 y_annual 对照  

res <- data.frame(  

  asof    = as.character(asof),  

  tau_year= taus,  

  y_annual= y_annual,  

  y_cc    = y_cc,  

  d_tau   = d_tau,  

  f_cc    = f_tau  

)  
  

disc_tab <- data.frame(  

  from_tau = taus[-length(taus)],  

  to_tau   = taus[-1],  

  f_disc   = f_disc  

)  
  

# 8) 打印结果（数值列做四舍五入）  

cat("———FRED Constant Maturity US Treasury —— as of ",  

  as.character(asof), "——\n")  

num_cols_res <- sapply(res, is.numeric)  

res_print <- res  

res_print[num_cols_res] <- lapply(res_print[num_cols_res], round, 6)  

print(res_print, row.names = FALSE)  
  

cat("\n—— Discrete one-period forwards between adjacent nodes ——\n")
num_cols_disc <- sapply(disc_tab, is.numeric)
disc_print <- disc_tab
disc_print[num_cols_disc] <- lapply(disc_print[num_cols_disc], round, 6)
print(disc_print, row.names = FALSE)

```

R 代码：收益率曲线三件套 (3/3)

```
# 9) 简单三联图: y( ) (连续) 、 d( ) 、 f( ) (连续)
op <- par(mfrow = c(1,3), mar = c(4,4,2,1))
plot(taus, y_cc*100, type="b", pch=19,
      xlab="Maturity (years)", ylab="Yield y( )[%p.a., cont.]",
      main="Zero-coupon Yield (cont.)")
plot(taus, d_tau, type="b", pch=19,
      xlab="Maturity (years)", ylab="Discount d( )",
      main="Discount Function")
plot(taus, f_tau*100, type="b", pch=19,
      xlab="Maturity (years)", ylab="Instantaneous f( )[%p.a.]",
      main="Instantaneous Forward (cont.)")
par(op)
```

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

从息票债提取贴现函数：问题设置

- 实务中几乎看不到“全期限”的零息债，多为息票债券。
- 记：
 - 价格： p_i ；
 - 现金流： b_{ij} ；
 - 支付时点： τ_{ij} ；
 - 目标：用样本 $\{p_i, b_{ij}, \tau_{ij}\}$ 提取 $d(\cdot), y(\cdot), f(\cdot)$ 。
- 直接对每个到期点“反演”贴现因子 $d(\tau)$ ：
 - 常出现解不唯一、对噪声极为敏感；
 - 难以保证 $d(\tau)$ 单调（无套利）；
 - 插值与外推表现较差。
- 因此通常采用正则化 + 平滑的统计建模方法。

统计模型：价格 = 现金流 × 贴现因子

定义 (统计模型)

$$p_i = \sum_{j=1}^{m_i} b_{ij} d(\tau_{ij}) + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) < \infty.$$

令 $p = Bd + \varepsilon$, 其中 B 为现金流矩阵, d 为若干节点处的贴现因子向量。形式上的最小二乘为

$$\hat{d} = (B'B)^{-1}B'p.$$

- 当结点数 m 大、矩阵病态或需满足单调/插值约束时，上式并不理想。
- 通常采用：
 - 正则化（惩罚项）；
 - 基函数展开（低维参数化）；
 - 加权与稳健估计；
 - 明确无套利约束。

统计模型的直观解读

- 对任一息票债 i :

$$p_i = \sum_{j=1}^{m_i} b_{ij} d(\tau_{ij}) + \varepsilon_i.$$

- 含义:

- 价格 = 未来各支付时点现金流的贴现值之和;
- ε_i 吸收市场摩擦、个券特征与数据噪声。

- 将所有债券汇总:

- d : 所有支付时点上的贴现因子;
- B : 现金流矩阵;
- $p = Bd + \varepsilon$ 。

- 直接逐点估计 d 的问题:

- 结点多、 $B'B$ 病态, 对噪声极敏感;
- 难以自动满足 $d(\tau)$ 单调、 $d(0) = 1$ 等无套利约束;
- 只在有限支付点有效, 插值与外推缺乏平滑性。

从“逐点 $d(\tau)$ ”到“曲线 $d(\tau)$ ”

- 更合理的做法：把 $d(\tau)$ 看作一条平滑函数。
- 基本思路：
 - 用少量参数刻画整个 $d(\tau)$ 曲线；
 - 在目标函数中加入平滑惩罚；
 - 叠加单调/正性等无套利约束。
- 这样可在三者间取得平衡：
 - ① 拟合误差（价格误差小）；
 - ② 曲线平滑（避免鬼畜振荡）；
 - ③ 无套利约束（ d 单调、正性等）。
- 由 $\hat{d}(\tau)$ 可一并推导：

$$\hat{y}(\tau) = -\frac{\ln \hat{d}(\tau)}{\tau}, \quad \hat{f}(\tau) = -\frac{d}{d\tau} \ln \hat{d}(\tau).$$

基函数展开：用少量系数逼近整条贴现曲线

- 将贴现函数视为平滑曲线，用基函数线性组合逼近：

$$d(t) \approx \sum_{\ell=1}^L \theta_\ell g_\ell(t), \quad L \ll n.$$

- 将定价方程改写为关于 θ 的线性回归：

$$p_i = \sum_{j=1}^{m_i} b_{ij} d(\tau_{ij}) + \varepsilon_i \Rightarrow p_i = \sum_{\ell=1}^L \theta_\ell X_{\ell i} + \varepsilon_i,$$

$$X_{\ell i} = \sum_{j=1}^{m_i} b_{ij} g_\ell(\tau_{ij}).$$

- 估计 $\theta = (\theta_1, \dots, \theta_L)'$ 后，可在任意 t 上计算

$$\hat{d}(t) = \sum_{\ell=1}^L \hat{\theta}_\ell g_\ell(t).$$

由 $\hat{d}(t)$ 推出收益率与远期曲线

- 贴现函数估计完成后，自然得到：

$$\hat{y}(t) = -\frac{\ln \hat{d}(t)}{t}, \quad \hat{f}(t) = -\frac{d}{dt} \ln \hat{d}(t) = \frac{d}{dt} \left\{ t \hat{y}(t) \right\}.$$

- 这一框架在统计与金融上都“站得住脚”，但需要：

- ① 合理选择基函数与复杂度；
- ② 控制曲线平滑与单调性；
- ③ 处理报价误差、流动性等异质性；
- ④ 关注插值与外推行为。

基函数选择与复杂度控制

- 常见基函数：

- 分段多项式 (McCulloch 1971);
- Bernstein 多项式 (Schaefer 1981);
- 指数基/指数样条 (Vasicek and Fong 1982);
- 平滑样条 (Fisher, Nychka, and Zervos 1995);
- B 样条等。

- 经验证明：

- 分段三次样条或指数样条兼顾“平滑—灵活”;
- 基函数数目 L 不宜过大，否则易过拟合、振荡；
- L 与惩罚参数 λ 通常需通过交叉验证或经验调参。

平滑与单调（无套利）约束

- 直接最小二乘：

- 在数据稀疏区易出现局部震荡；
- 甚至违反 $d(t)$ 单调递减。

- 常见做法：

- 在目标函数加入平滑惩罚：

$$\min_{\theta} \sum_i (p_i - \hat{p}_i(\theta))^2 + \lambda \int (d''(t))^2 dt.$$

- 在参数化中直接保证 $d(t)$ 正、递减：

- 例如用“积分+指数”重参数化；
- 或在线性/二次规划中显式加入 $d(t_{k+1}) \leq d(t_k)$ 、 $d(t) \geq 0$ 等约束。

加权与稳健性、插值与外推

- 不同债券的报价误差与流动性差异很大：
 - 可对回归做加权（如按久期或价差的反比）；
 - 或使用稳健损失函数（Huber、Tukey 等）。
- 基函数法天然给出全期限 $\hat{d}(t)$ ：
 - 插值：在估计区间内通常平滑、可解释；
 - 外推：需配合经济约束与尾部形状，避免远端“乱飞”。
- 理论上，在基函数族足够丰富且调参得当时，

$$\sup_{\tau \in [0, \tau^{\max}]} |\hat{d}(\tau) - d(\tau)| \xrightarrow{P} 0,$$

参见 Silverman (1986) 关于平滑估计的一般一致性结果。

R 示例：基函数展开估计贴现曲线（B 样条）(1/3)

```
# =====
# 基函数展开法：由“附息债（基于 FRED 常数到期收益率构造）”估计贴现曲线 d(t)
# =====

suppressPackageStartupMessages({
  library(quantmod)
  library(splines)
})

# 1) 期限与 FRED 代码
taus <- c(0.5, 1, 2, 3, 5, 7, 10, 20, 30)
fred <- c("DGS6MO", "DGS1", "DGS2", "DGS3", "DGS5", "DGS7", "DGS10", "DGS20", "DGS30")

getSymbols(fred, src = "FRED", auto.assign = TRUE, warnings = FALSE)
Y <- do.call(merge, lapply(fred, get))
Yc <- Y[stats::complete.cases(Y)]
stopifnot(nrow(Yc) > 0)
last_row <- tail(Yc, 1)
asof <- index(last_row)
y_ann_pct <- as.numeric(last_row)
y_ann <- y_ann_pct/100
```

R 示例：基函数展开估计贴现曲线（B 样条）

```

# 2) 构造“面值成交”的半年付息债：票息率=当前常到收益率
bond_list <- vector("list", length(taus))
names(bond_list) <- paste0("T", taus, "Y")

for (k in seq_along(taus)) {
  Tm <- taus[k]
  cR <- y_ann[k]
  N <- as.integer(round(2*Tm))
  tps <- (1:N)/2
  cf <- rep(cR/2, N); cf[N] <- cf[N] + 1
  price <- 1
  bond_list [[k]] <- list(t=tps, cf=cf, p=price)
}

# 3) B 样条基函数
t_max <- max(taus)
n_basis <- 12
knots_inner <- seq(0.5, t_max-0.5, length.out = max(0, n_basis-4))
gfun <- function(tt) bs(tt, degree = 3, knots = knots_inner,
                         Boundary.knots = c(0, t_max), intercept = TRUE)

# 构造回归矩阵 X 与价格向量 p
p_vec <- sapply(bond_list, function(b) b$p)
X <- matrix(0, nrow=length(bond_list), ncol=ncol(gfun(c(0, t_max))))
colnames(X) <- paste0("g", seq_len(ncol(X)))

```

R 示例：基函数展开估计贴现曲线（B 样条）(2/3)

```
for (i in seq_along(bond_list)) {  
  bi <- bond_list[[i]]  
  G <- gfun(bi$t)  
  X[i, ] <- colSums( G * bi$cf )  
}  
  
# 4) 在 t=0 处加入 “伪观测” : d(0)=1  
G0 <- gfun(0)  
w0 <- 1e6  
X_aug <- rbind(X, sqrt(w0)*G0)  
p_aug <- c(p_vec, sqrt(w0)*1)  
  
# 5) 二阶差分惩罚 (平滑)  
D2 <- diff(diag(ncol(X)), differences = 2)  
lambda <- 1e-2  
R <- crossprod(D2)  
  
XtX <- crossprod(X_aug)  
Xtp <- crossprod(X_aug, p_aug)  
theta_hat <- solve(XtX + lambda*R, Xtp)
```

R 示例：基函数展开估计贴现曲线（B 样条）(3/3)

```

# 6) 生成 d_hat, y_hat, f_hat
grid <- seq(0, t_max, by = 0.01)
G_grid <- gfun(grid)
d_hat <- as.vector(G_grid %*% theta_hat)
d_hat[d_hat <= 1e-10] <- 1e-10
y_hat <- -log(d_hat) / pmax(grid, 1e-8)
g_tau <- grid * y_hat
f_hat <- c(diff(g_tau)/diff(grid), NA)
f_hat[length(f_hat)] <- f_hat[length(f_hat)-1]

# 7) 回定价检查
p_fitted <- numeric(length(bond_list))
for (i in seq_along(bond_list)) {
  bi <- bond_list[[i]]
  d_i <- approx(grid, d_hat, xout = bi$t, rule = 2)$y
  p_fitted[i] <- sum(bi$cf * d_i)
}
fit_tab <- data.frame(
  tau = taus,
  p_target = p_vec,
  p_fitted = p_fitted,
  abs_err = p_fitted - p_vec
)
cat("As of:", as.character(asof), "\n")
print(round(fit_tab, 6), row.names = FALSE)

```

R 示例：基函数展开的曲线图

```
# 8) 作图
op <- par(mfrow=c(1,3), mar=c(4,4,2,1))
plot(grid, d_hat, type="l", lwd=2, xlab="t[years]", ylab="d(t)",
      main="贴现函数d(t)")
abline(h=1, v=0, col="grey80", lty=3)
points(0, 1, pch=19, col="steelblue")

plot(grid, y_hat*100, type="l", lwd=2, xlab="t[years]",
      ylab="y(t)[%p.a.,cont.]", main="连续复利收益率y(t)")
# 仅作参考标记：原始点（未严格统一口径）
points(taus, log(1+y_ann)/taus * taus * 100 / taus, pch=19, col="grey40")

plot(grid, f_hat*100, type="l", lwd=2, xlab="t[years]",
      ylab="f(t)[%p.a.]", main="瞬时远期f(t)")
par(op)
```

备注：在实证应用中，可将示例中的 bond_list 直接替换为真实“息票债现金流+价格”样本 (t, cf, p) ，框架不变。

参数法: 用少量参数刻画整条曲线

- 参数法通过给定一个结构化、经济含义明确的函数族来描述曲线。
- 常用模型:

- Nelson–Siegel (NS, Nelson Siegel 1987);
- Svensson 扩展 (NSS, Svensson 1994)。

- NS 模型从瞬时远期利率出发:

$$f(t) = \beta_0 + (\beta_1 + \beta_2 t/\tau_0) \exp(-t/\tau_0),$$

- 其中:

- β_0 : 远端水平 (level);
- β_1 : 斜率 (slope, 短端相对长端);
- β_2, τ_0 : 控制中段曲率 (curvature) 的峰值位置与高度。

Svensson 扩展: 增加次一层曲率

- Svensson 在 NS 基础上加入第二个衰减因子:

$$f(t) = \beta_0 + \beta_1 e^{-t/\tau_1} + \beta_2(t/\tau_1)e^{-t/\tau_1} + \beta_3(t/\tau_2)e^{-t/\tau_2},$$

- 含义:

- β_3, τ_2 增强长端或“第二峰”的刻画能力;
- 能拟合更多样的曲线形状(如双峰、长端隆起)。

- 由 $f(t)$ 可解析积分得到:

- 累积短端积分 $\int_0^t f(s) ds$;
- 贴现函数 $d_\theta(t) = \exp\{-\int_0^t f(s) ds\}$;
- 收益率 $y_\theta(t) = -\ln d_\theta(t)/t$ 。

参数估计: 价格准则 vs 收益率准则

- 估计目标 (价格准则):

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \left(p_i - \sum_{j=1}^{m_i} b_{ij} d_{\theta}(\tau_{ij}) \right)^2.$$

- 也可用“收益率误差”或“远期误差”作为目标, 但:

- 价格准则更符合无套利线性定价;
- 对贴现函数的影响更直接。

- 实务要点:

- ① 初值与尺度: NS/NSS 均为非线性最小二乘, 良好初值很重要;
- ② 约束: 如 $\tau_k > 0$ 、参数幅度不宜过大;
- ③ 加权: 反映不同债券或期限的报价精度与流动性;
- ④ 诊断: 残差与曲线形态检查 (短端、长端是否系统性偏离)。

理论性质与模型错设

- 在模型正确设定且常规正则条件下:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Omega), \quad d_{\hat{\theta}}(\tau) \rightarrow d_{\theta}(\tau) \quad (n \rightarrow \infty).$$

- 但真实市场的曲线形态随制度、税收、基差与流动性时变:

- “模型错设”较常见;
- 不宜把 NS/NSS 看作“真模型”，而应视为**低维近似基**。

- 实务建议:

- 将 NS/NSS 用作稳健的“主干结构”，配合滚动估计与参数平滑；
- 对极端短端或长端可叠加附加修正或局部样条；
- 对异常观测进行适度剔除或稳健处理。

R 示例: NS / NSS 参数估计 (价格准则) (1/2)

```

# =====
# NS / NSS 参数法: 利用“价格准则”估计  $d(t)$ ,  $y(t)$ ,  $f(t)$ 
# =====
suppressPackageStartupMessages(library(quantmod))

# 1) 期限与 FRED 代码
taus <- c(0.5, 1, 2, 3, 5, 7, 10, 20, 30)
fred <- c("DGS6MO", "DGS1", "DGS2", "DGS3", "DGS5",
         "DGS7", "DGS10", "DGS20", "DGS30")

getSymbols(fred, src = "FRED", auto.assign = TRUE, warnings = FALSE)
Y <- do.call(merge, lapply(fred, get))
Yc <- Y[stats::complete.cases(Y)]
stopifnot(nrow(Yc) > 0)
last_row <- tail(Yc, 1)
asof <- index(last_row)
y_ann <- as.numeric(last_row)/100

# 2) 构造“面值成交”的半年付息债
bond_list <- vector("list", length(taus))
names(bond_list) <- paste0("T", taus, "Y")
for (k in seq_along(taus)) {
  Tm <- taus[k]
  cR <- y_ann[k]
  N <- as.integer(round(2*Tm))
  tps <- (1:N)/2
  cf <- rep(cR/2, N); cf[N] <- cf[N] + 1
  price <- 1
  bond_list[[k]] <- list(t=tps, cf=cf, p=price)
}

```

R 示例: NS / NSS 参数估计 (价格准则) (2/2)

```

# 3) NS / NSS 的贴现与远期函数
A_NS <- function(t, theta) {
  b0 <- theta[1]; b1 <- theta[2]; b2 <- theta[3]; tau0 <- theta[4]
  E <- exp(-t/tau0)
  b0*t + b1*tau0*(1 - E) + b2*(tau0 - (t + tau0)*E)
}
d_NS <- function(t, theta) exp( - A_NS(t, theta) )
f_NS <- function(t, theta) {
  b0 <- theta[1]; b1 <- theta[2]; b2 <- theta[3]; tau0 <- theta[4]
  b0 + (b1 + b2*t/tau0) * exp(-t/tau0)
}
A_NSS <- function(t, theta) {
  b0 <- theta[1]; b1 <- theta[2]; b2 <- theta[3]; tau1 <- theta[4]
  b3 <- theta[5]; tau2 <- theta[6]
  E1 <- exp(-t/tau1); E2 <- exp(-t/tau2)
  b0*t + b1*tau1*(1 - E1) +
    b2*(tau1 - (t + tau1)*E1) +
    b3*(tau2 - (t + tau2)*E2)
}
d_NSS <- function(t, theta) exp( - A_NSS(t, theta) )
f_NSS <- function(t, theta) {
  b0 <- theta[1]; b1 <- theta[2]; b2 <- theta[3]; tau1 <- theta[4]
  b3 <- theta[5]; tau2 <- theta[6]
  b0 + b1*exp(-t/tau1) +
    b2*(t/tau1)*exp(-t/tau1) +
    b3*(t/tau2)*exp(-t/tau2)
}

```

R 示例: NS / NSS 拟合与曲线 (1/4)

```
# 4) 价格准则目标函数
price_error_NS <- function(theta) {
  if (theta[4] <= 1e-6) return(1e12)
  err <- 0
  for (i in seq_along(bond_list)) {
    bi <- bond_list[[i]]
    di <- d_NS(bi$t, theta)
    p_hat <- sum(bi$cf * di)
    err <- err + (p_hat - bi$p)^2
  }
  err
}
price_error_NSS <- function(theta) {
  if (theta[4] <= 1e-6 || theta[6] <= 1e-6) return(1e12)
  err <- 0
  for (i in seq_along(bond_list)) {
    bi <- bond_list[[i]]
    di <- d_NSS(bi$t, theta)
    p_hat <- sum(bi$cf * di)
    err <- err + (p_hat - bi$p)^2
  }
  err
}
```

R 示例: NS / NSS 拟合与曲线 (2/4)

```
# 5) 拟合 (L-BFGS-B)
theta0_NS <- c(b0=mean(y_ann), b1=-0.02, b2= 0.02, tau0=2.0)
lower_NS <- c(-0.10, -5, -5, 1e-3)
upper_NS <- c( 0.20, 5, 5, 10.0)
fit_NS <- optim(par=theta0_NS, fn=price_error_NS, method="L-BFGS-B",
                 lower=lower_NS, upper=upper_NS,
                 control=list(maxit=2000))

theta0_NSS <- c(b0=mean(y_ann), b1=-0.03, b2=0.06, tau1=2.0,
                  b3=-0.02, tau2=8.0)
lower_NSS <- c(-0.10, -5, -5, 1e-3, -5, 1e-3)
upper_NSS <- c( 0.20, 5, 5, 10.0, 5, 30.0)
fit_NSS <- optim(par=theta0_NSS, fn=price_error_NSS, method="L-BFGS-B",
                  lower=lower_NSS, upper=upper_NSS,
                  control=list(maxit=3000))

cat("NS\u2225参数:\n"); print(fit_NS$par)
cat("NSS\u2225参数:\n"); print(fit_NSS$par)
```

R 示例: NS / NSS 拟合与曲线 (3/4)

```
# 6) 生成整条曲线并回定价
grid <- seq(0.01, max(taus), by=0.01)
d_ns <- d_NS(grid, fit_NS$par); y_ns <- -log(d_ns)/grid
f_ns <- f_NS(grid, fit_NS$par)
d_nss <- d_NSS(grid, fit_NSS$par); y_nss <- -log(d_nss)/grid
f_nss <- f_NSS(grid, fit_NSS$par)

price_fit_tab <- function(d_fun, theta) {
  DF <- data.frame(
    tau      = taus,
    p_target = sapply(bond_list, function(b) b$p),
    p_fitted = sapply(bond_list, function(b) sum(b$cf * d_fun(b$t, theta)))
  )
  DF$abs_err <- DF$p_fitted - DF$p_target
  DF
}
tab_NS <- price_fit_tab(d_NS, fit_NS$par)
tab_NSS <- price_fit_tab(d_NSS, fit_NSS$par)
cat("NS\u207d回定价误差:\n"); print(round(tab_NS, 6), row.names = FALSE)
cat("NSS\u207d回定价误差:\n"); print(round(tab_NSS, 6), row.names = FALSE)
```

R 示例: NS / NSS 拟合与曲线 (4/4)

```
# 7) 作图
op <- par(mfrow=c(1,3), mar=c(4,4,2,1))
plot(grid, d_ns, type="l", lwd=2, col="steelblue",
      xlab="t (years)", ylab="d(t)", main="贴现函数 d(t)")
lines(grid, d_nss, lwd=2, col="tomato")
legend("topright", c("NS", "NSS"), lty=1,
       col=c("steelblue", "tomato"), bty="n")

plot(grid, 100*y_ns, type="l", lwd=2, col="steelblue",
      xlab="t (years)", ylab="y(t) [% p.a., cont.]",
      main="连续复利收益率 y(t)")
lines(grid, 100*y_nss, lwd=2, col="tomato")

plot(grid, 100*f_ns, type="l", lwd=2, col="steelblue",
      xlab="t (years)", ylab="f(t) [% p.a.]",
      main="瞬时远期 f(t)")
lines(grid, 100*f_nss, lwd=2, col="tomato")
par(op)
```

Fama-Bliss：分段常数远期的序贯引导

- Fama and Bliss (1987) 提出一种“由短到长、逐段锁定远期利率”的思路。
- 基本设定：
 - 将期限区间切分为相邻到期区间 $(\tau^{i-1}, \tau^i]$ ；
 - 在每个区间上假设远期利率为常数 f^i 。
- 优点：
 - 每一步只处理一个小问题，结构简单；
 - 通过严格样本筛选（干净债券）提高稳健性；
 - 逐段构造整条远期与贴现曲线。
- 与其他方法的系统比较：见 Bliss (1997)。

Fama-Bliss：形式化定义

定义 (分段常数远期与序贯定价)

将债券按到期由短到长排序，记第 i 只债券到期为 τ^i , $\tau^0 = 0$ 。在相邻区间 $(\tau^{i-1}, \tau^i]$ 上令

$$f(\tau) = f^i, \quad \tau \in (\tau^{i-1}, \tau^i].$$

据此可将贴现函数写为依赖 $\{f^1, \dots, f^i\}$ 的形式。然后利用第 i 只债券的价格等式

$$p_i = \sum_{j=1}^{m_i} b_i(\tau_{ij}) d(\tau_{ij}),$$

其中 $d(\cdot)$ 由已知的 $\{f^1, \dots, f^{i-1}\}$ 与当前未知的 f^i 共同确定。求得 f^i 后，进入下一期限债券，直至样本最长期限。

样本筛选与“稳健递推”的直观类比

- 实施时，**样本选择**尤为关键：
 - 保留：完全征税、不可赎回、非含权的标准化品种；
 - 剔除：到期不足一年的票据、收益率极端异常的债券；
 - 要求：引入某只债后，曲线形态不应出现反常跳变。
- 可类比为“带筛选的序贯回归”：
 - 先用 OLS 得残差，再只相信残差最小的一小部分样本；
 - 在合理假设下，可与全样本估计在大样本下等价。
- Fama-Bliss 的递推逻辑类似：
 - 用干净子样本逐段锁定远期 f^i ；
 - 在稳健性与可解释性之间取得折中。

Fama-Bliss 的定位

- 统计性质：
 - 并非在所有设定下都有完整极限理论；
 - 分段常数远期在曲线过于光滑或多“弯点”时可能偏粗。
- 较稳妥的用法：
 - 作为“初始曲线”的快速构造工具；
 - 再配合样条平滑或 NS/NSS 作进一步细化；
 - 或在每一步定价等式中叠加温和正则项与局部单调约束。
- 总体而言：
 - Fama-Bliss = 分段常数远期 + 序贯定价 + 严格样本筛选；
 - 不是“一步到位”的全局优化，而是“由近及远”的稳健递推。

R 示例：Fama-Bliss 序贯引导（1/2）

```
# =====
# Fama-Bliss 序贯引导（分段常数远期）
# =====

suppressPackageStartupMessages(library(quantmod))

taus <- c(0.5, 1, 2, 3, 5, 7, 10, 20, 30)
fred <- c("DGS6MO", "DGS1", "DGS2", "DGS3", "DGS5",
         "DGS7", "DGS10", "DGS20", "DGS30")

getSymbols(fred, src = "FRED", auto.assign = TRUE, warnings = FALSE)
Y <- do.call(merge, lapply(fred, get))
Yc <- Y[complete.cases(Y)]
stopifnot(nrow(Yc) > 0)
last_row <- tail(Yc, 1)
asof <- index(last_row)
par_y <- as.numeric(last_row)/100

make_par_bond <- function(Tm, cR) {
  N <- as.integer(round(2*Tm))
  tps <- (1:N)/2
  cf <- rep(cR/2, N); cf[N] <- cf[N] + 1
  list(t = tps, cf = cf, p = 1)
}
bond_list <- mapply(make_par_bond, Tm = taus, cR = par_y, SIMPLIFY = FALSE)
names(bond_list) <- paste0("T", taus, "Y")
```

R 示例：Fama-Bliss 序贯引导 (2/2)

```
# 分段常数远期下的贴现函数
discount_from_piecewise_f <- function(t, tau_knots, f_vec) {
  l <- length(f_vec)
  stopifnot(length(tau_knots) == l + 1)
  integ <- sapply(t, function(tt) {
    if (tt <= 0) return(0)
    k <- max(which(tau_knots < tt)); k <- min(k, l)
    full <- 0
    if (k >= 1) {
      full <- sum( f_vec[1:(k-1)] * diff(tau_knots)[1:(k-1)] )
      rem <- tt - tau_knots[k]
      full <- full + f_vec[k] * rem
    }
    full
  })
  exp(-integ)
}
```

R 示例：Fama-Bliss 引导与曲线 (1/3)

```
# 解出第 i 段的  $f_i^*$ 
solve_f_i <- function(i, f_vec, tau_knots, bond_list, bracket = c(-0.05, 0.20)) {
  bi <- bond_list[[i]]
  obj <- function(fi) {
    f_tmp <- f_vec
    f_tmp[i] <- fi
    d_i <- discount_from_piecewise_f(bi$t, tau_knots[1:(i+1)], f_tmp[1:i])
    sum(bi$cf * d_i) - bi$p
  }
  a <- bracket[1]; b <- bracket[2]
  iter <- 0; max_iter <- 20
  while (iter < max_iter && obj(a)*obj(b) > 0) {
    a <- a - 0.05; b <- b + 0.05; iter <- iter + 1
  }
  if (obj(a)*obj(b) > 0)
    stop("无法在合理区间内找到根：请检查样本或区间。")
  uniroot(obj, interval = c(a, b), tol = 1e-10)$root
}

l <- length(taus)
tau_knots <- c(0, taus)
f_hat <- rep(NA_real_, l)
```

R 示例：Fama-Bliss 引导与曲线（2/3）

```
for (i in 1:l) {
  if (i == 1) {
    f_prev <- numeric(1)
  } else {
    f_prev <- f_hat[1:(i-1)]
  }
  f_hat[i] <- solve_f_i(i,
    f_vec      = c(f_prev, NA),
    tau_knots = tau_knots,
    bond_list = bond_list)
}

grid <- seq(0, max(taus), by = 0.01)
f_grid <- sapply(grid, function(tt) {
  if (tt <= 0) return( f_hat[1] )
  k <- max(which(tau_knots < tt)); k <- min(k, l)
  f_hat[k]
})
d_grid <- discount_from_piecewise_f(grid, tau_knots, f_hat)
d_grid[d_grid <= 1e-12] <- 1e-12
y_grid <- -log(d_grid)/pmax(grid, 1e-8)
```

R 示例：Fama-Bliss 引导与曲线（3/3）

```

p_fitted <- sapply(seq_along(bond_list), function(i) {
  bi <- bond_list[[i]]
  d_i <- discount_from_piecewise_f(bi$t, tau_knots[1:(i+1)], f_hat[1:i])
  sum(bi$cf * d_i)
})
fit_tab <- data.frame(
  tau = taus,
  p_target = 1,
  p_fitted = p_fitted,
  abs_err = p_fitted - 1
)
cat("As of:", as.character(asof), "\n")
print(round(fit_tab, 8), row.names = FALSE)

op <- par(mfrow=c(1,3), mar=c(4,4,2,1))
plot(grid, d_grid, type="l", lwd=2,
      xlab="t (years)", ylab="d(t)",
      main="贴现函数 d(t) (Fama - Bliss) ")
plot(grid, 100*y_grid, type="l", lwd=2,
      xlab="t (years)", ylab="y(t) [% p.a., cont.]",
      main="连续复利收益率 y(t) ")
plot(grid, 100*f_grid, type="s", lwd=2,
      xlab="t (years)", ylab="f(t) [% p.a.]",
      main="分段常数远期 f(t) ")
abline(v = taus, col="grey85", lty=3)
par(op)

```

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama–Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

离散时间下的收益率曲线：问题与目标

- 本节在离散时间框架下讨论收益率曲线建模：
 - 收益率如何随时间演化？
 - 如何在该框架下为固定收益衍生品定价？
- 主要内容：
 - ① 利率的经济学假说：预期、溢价与市场结构；
 - ② 收益率的统计性质：均值、波动与高阶矩；
 - ③ 无套利与随机贴现因子（SDF）框架；
 - ④ Vasicek、CIR 及多因子仿射期限结构模型。
- 目标：在可检验的统计模型中嵌入经济学直觉，为定价与预测提供统一框架。

利率与收益率的经典经济学假说

关于利率与收益率的决定，经典文献中有四类常见假说，从不同角度刻画远期利率、短期利率的预期以及期限溢价之间的关系：

- 预期假说 (Expectations Hypothesis)；
- 流动性（风险）溢价假说 (Liquidity/Risk Premium)；
- 市场分割假说 (Market Segmentation)；
- 偏好栖息地 (Preferred Habitat)。

它们为现代期限结构理论中“预期 vs. 风险溢价”的分解提供了直觉基础。

预期假说与流动性/风险溢价假说

定义 (预期假说 (Expectations Hypothesis))

在适当的测度下，远期利率等于未来对应一期短期即期利率的条件期望。

定义 (流动性 (风险) 溢价假说 (Liquidity/Risk Premium Hypothesis))

远期利率等于未来短期利率的条件期望，加上一项与期限相关的风险溢价（或流动性溢价）。

- 预期假说：远期利率主要反映对未来短端的预期。
- 溢价假说：在此基础上加入与期限相关的风险补偿。

市场分割假说与偏好栖息地

定义 (市场分割假说 (Market Segmentation Hypothesis))

不同期限区间构成相对独立的子市场，各期限的利率主要由该期限特定的供需关系与制度约束决定。

定义 (偏好栖息地 (Preferred Habitat))

投资者对某些期限区间存在偏好；若给予足够溢价，投资者会偏离其偏好期限，从而在不同期限的资产之间形成可交易的风险补偿。

- 市场分割：不同期限“各有一片天”；
- 偏好栖息地：存在偏好，但**价格信号**（溢价）可引导跨期限配置。

现代期限结构理论中的“预期 + 溢价”分解

- 现代模型在上述直觉基础上，对**风险溢价**施加可检验约束：
 - 如仿射模型下的**线性风险价格假设**；
 - 将远期利率分解为“未来短端预期 + 期限溢价”。
- 在数据上可做的事情：
 - 使用 \mathbb{P} 测度下的动态模型估计短端预期路径；
 - 利用 SDF 或 \mathbb{Q} 测度提取期限溢价；
 - 对比不同假说在不同样本期的表现。

收益率的统计性质：经典结论与现代样本

- 早期研究（如 Campbell, Lo, and MacKinlay (1997) 基于 McCulloch and Kwon (1993) 的 1952–1991 样本）发现：
 - 平均收益率曲线整体上行且呈凹形；
 - 单期曲线可上行、下行或驼峰；
 - 收益率标准差对期限不敏感；
 - 短端偏度与超额峰度高且为正，随期限延长下降；
 - 原序列自相关显著且随期限增加而增强。
- 本节用 2000–2024 年的数据做一组可比的更新：
 - FRED 的美债 Constant Maturity Yield, 1–360 个月；
 - 日频样本，考察均值、波动、高阶矩与差分自相关。

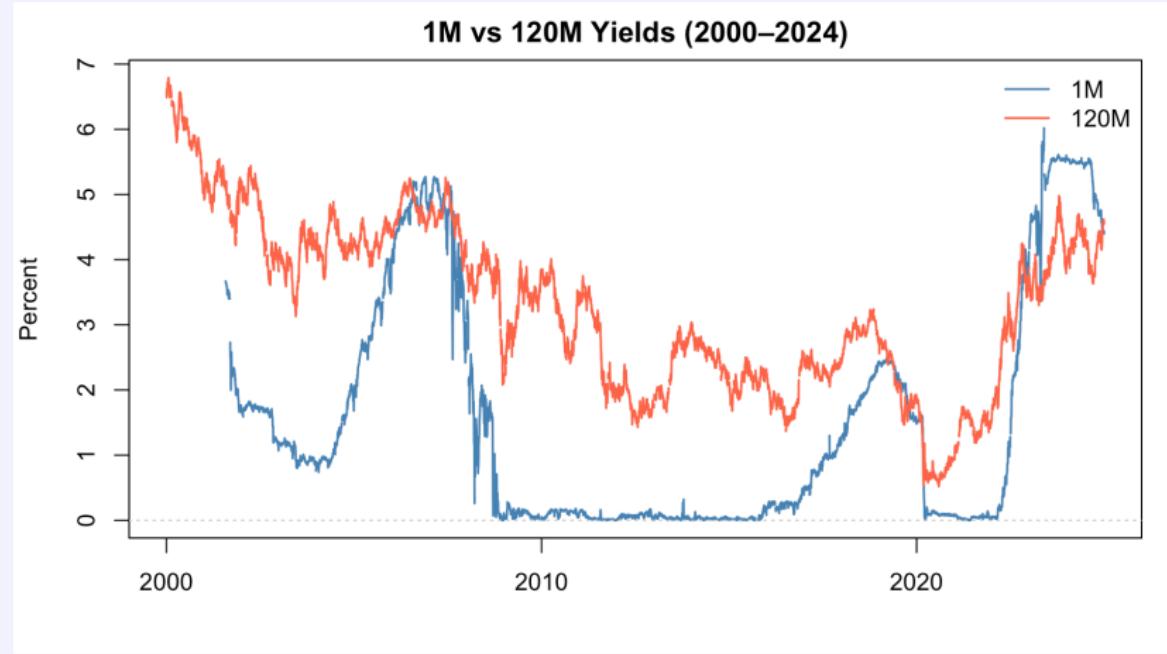
样本与数据描述：2000–2024 年 FRED 美债

- 截面：FRED 公布的“固定期限收益率”(Constant Maturity)：
1, 3, 6, 12, 24, 36, 60, 120, 240, 360 个月.
- 样本区间：2000–2024 年日频数据。
- 数据特点：
 - 1 个月与 30 年期在部分年份数据缺失；
 - 不影响各自可得区间内的描述统计与相关性计算。
- 主要考察对象：
 - 收益率水平与波动的期限结构；
 - 差分收益率 Δy_t 的自相关；
 - 条件累积量随利率水平的变化。

收益率时间序列：1M 与 120M

- 图 1 展示 1M 与 120M 两条典型期限的时间序列。
- 现象：
 - 短端在 2008–2016 年长期贴近零利率下限 (ZLB)；
 - 2022 年后伴随快速加息显著抬升；
 - 10 年期收益率趋势更平缓、波动更小。

收益率时间序列：1M 与 120M

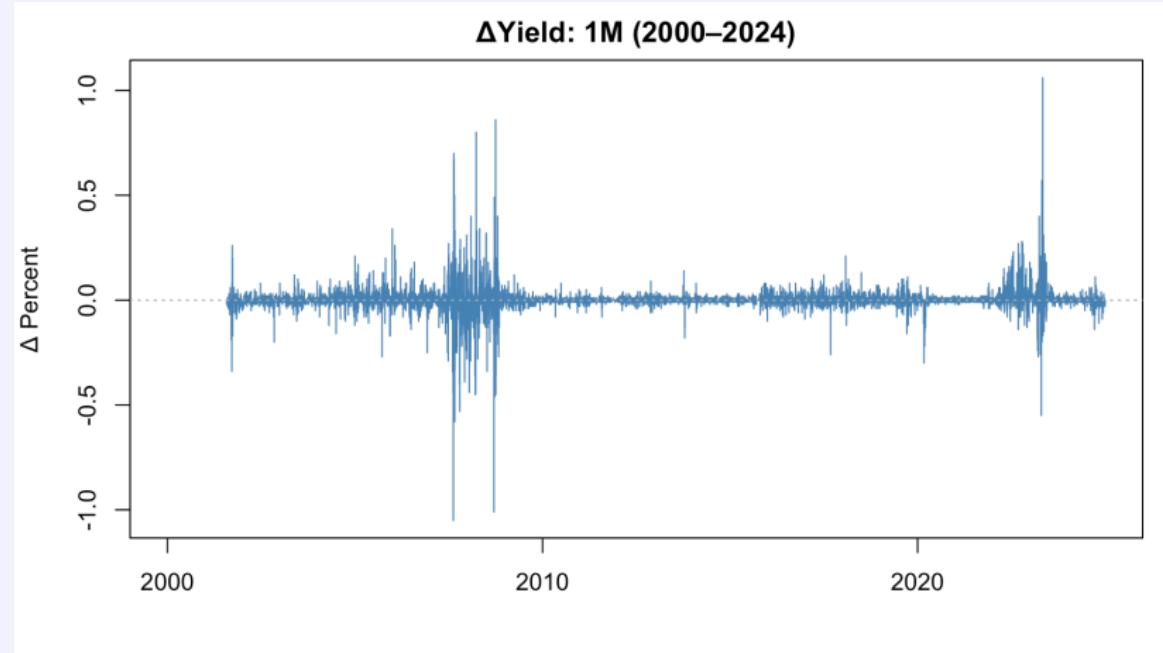


图：1M 与 120M 收益率（2000–2024 年，纵轴为%）。

短端日变化：1M 收益率的波动聚集

- 图 2 给出 1M 收益率的日变化。
- 观测：
 - 2008–2009 年与 2020 年附近波动显著放大；
 - 后续总体回落；
 - 反映危机与流动性冲击聚集在少数时段。

短端日变化：1M 收益率的波动聚集



图：1M 收益率的日变化（2000–2024 年，单位：百分点）。

描述统计：均值、标准差、偏度与超峰度

表：日频收益率的描述统计（2000–2024 年）——均值 m 、标准差 s （单位： $\%$ ），偏度 κ_3 与超峰度 κ_4

统计量	1M	3M	6M	12M	24M	36M	60M	120M	240M	360M
均值 m	1.5525	1.8535	1.9534	2.0237	2.1983	2.3717	2.7284	3.2833	3.7965	3.8742
标准差 s	1.7899	1.9856	1.9846	1.9016	1.7732	1.6630	1.4982	1.3058	1.3021	1.1793
偏度 κ_3	1.0263	0.8257	0.7801	0.7063	0.6392	0.5624	0.4056	0.1212	-0.0287	-0.1052
超峰度 κ_4	-0.2768	-0.7543	-0.8070	-0.8556	-0.7504	-0.7105	-0.6551	-0.7196	-0.9562	-0.8926

差分收益率的自相关：短期相关性集中在短端

表：差分收益率的自相关（1–5 阶，2000–2024 年）

阶数	1M	3M	6M	12M	24M	36M	60M	120M	240M	360M
$\rho_{\Delta}(1)$	0.1412	0.1326	0.0632	0.0433	-0.0185	-0.0086	-0.0051	-0.0037	-0.0165	-0.0099
$\rho_{\Delta}(2)$	-0.0716	-0.0890	-0.0317	-0.0021	-0.0417	-0.0483	-0.0597	-0.0492	-0.0419	-0.0471
$\rho_{\Delta}(3)$	-0.1488	-0.1220	-0.0587	-0.0169	-0.0045	-0.0018	0.0055	0.0064	-0.0024	-0.0075
$\rho_{\Delta}(4)$	0.0125	0.0366	0.0837	0.0324	0.0066	0.0111	0.0080	-0.0049	-0.0108	-0.0094
$\rho_{\Delta}(5)$	0.1110	0.0584	0.0833	-0.0034	-0.0064	-0.0050	-0.0057	-0.0048	0.0022	-0.0034

统计性质总结：现代样本 vs 早期样本

- 2000–2024 年样本中的稳健事实：
 - 平均收益率曲线整体随期限上行；
 - 标准差随期限递减，短端波动更大；
 - 短端收益率偏度显著为正，随期限接近零乃至略为负；
 - 多数期限的超额峰度为负，尖峰厚尾特征较早期样本收敛。
- 差分收益率 Δy_t ：
 - 极短端（1M, 3M）仍有一定短期可预测性；
 - 随期限拉长，自相关迅速衰减并略呈负值；
 - 与“原序列近似单位根、差分后弱相关”的结论一致。
- ZLB 与疫情冲击等现代因素：
 - 使短端水平与波动呈阶段性“极值化”；
 - 长端统计特征相对更稳定。

R 代码：收益率描述统计与差分自相关 (1/3)

```
# =====
# 统计性质复现：2000–2024 日频收益率的描述统计与差分自相关
# 依赖: quantmod, zoo, moments
#
if (requireNamespace("rstudioapi", quietly = TRUE) &&
    rstudioapi::isAvailable()) {
  setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
}
suppressPackageStartupMessages({
  library(quantmod)
  library(zoo)
  library(moments)  # skewness(), kurtosis()
})

mats_month <- c(1,3,6,12,24,36,60,120,240,360)
fred_codes <- c("DGS1MO", "DGS3MO", "DGS6MO", "DGS1", "DGS2", "DGS3",
                 "DGS5", "DGS10", "DGS20", "DGS30")

from <- as.Date("2000-01-01"); to <- as.Date("2024-12-31")
getSymbols(fred_codes, src = "FRED", from = from, to = to, auto.assign = TRUE)

Y <- do.call(merge, mget(fred_codes))
colnames(Y) <- paste0(mats_month, "M")
Ynum <- Y / 100
```

R 代码：收益率描述统计与差分自相关 (2/3)

```
# _____  
stats_one <- function(x) {  
  x <- na.omit(as.numeric(x))  
  if (length(x) < 10) return(c(m=NA, s=NA, k3=NA, k4=NA))  
  c(m = mean(x)*100,  
    s = sd(x)*100,  
    k3 = skewness(x),  
    k4 = kurtosis(x) - 3)  
}  
  
tab1 <- t(apply(Ynum, 2, stats_one))  
tab1 <- as.data.frame(tab1)  
tab1$Maturity <- rownames(tab1)  
tab1 <- tab1[, c("Maturity", "m", "s", "k3", "k4")]  
rownames(tab1) <- NULL  
  
acf_one <- function(x, maxlag=5) {  
  x <- diff(na.omit(as.numeric(x)))  
  if (length(x) < (maxlag+10)) return(rep(NA, maxlag))  
  acf(x, plot=FALSE, na.action=na.pass)$acf[2:(maxlag+1)]  
}  
acf_mat <- t(apply(Ynum, 2, acf_one, maxlag=5))  
colnames(acf_mat) <- paste0("rho_D(", 1:5, ")")  
tab2 <- as.data.frame(acf_mat)  
tab2$Maturity <- rownames(acf_mat)  
tab2 <- tab2[, c("Maturity", colnames(acf_mat))]  
rownames(tab2) <- NULL
```

R 代码：收益率描述统计与差分自相关 (3/3)

```
# =====
cat("== Table 13.1 ==\n")
print(within(tab1, { m=round(m,4); s=round(s,4);
                      k3=round(k3,4); k4=round(k4,4)}),
      row.names=FALSE)

cat("\n== Table 13.2 ==\n")
print(within(tab2, {
  'rho_D(1)'=round('rho_D(1)',4); 'rho_D(2)'=round('rho_D(2)',4);
  'rho_D(3)'=round('rho_D(3)',4); 'rho_D(4)'=round('rho_D(4)',4);
  'rho_D(5)'=round('rho_D(5)',4)
}), row.names=FALSE)
```

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

随机贴现因子与无套利定价

- 令 p_t^n 为 t 时购买、 $t+n$ 兑付 1 美元的价格。
- 一期收益：

$$R_{t+1} = \frac{p_{t+1}^{n-1}}{p_t^n}.$$

- 若存在随机贴现因子 (SDF) M_{t+1} , 则无套利要求：

$$p_t^n = E_t(M_{t+1} p_{t+1}^{n-1}), \quad p_t^n = E_t\left(\prod_{j=1}^n M_{t+j}\right).$$

- SDF 统一描述了时间价值与风险补偿。

风险中性测度与收益率/远期利率

- 在风险中性测度 \mathbb{Q} 下:

$$M_{t+1} = e^{-r_t} \frac{d\mathbb{Q}}{d\mathbb{P}},$$

则

$$p_t^n = E_t^{\mathbb{Q}} \left[e^{-(r_t + \dots + r_{t+n-1})} \right].$$

- 定义收益率与远期利率:

$$y_t^n := -\frac{1}{n} \log p_t^n, \quad 1 + f_t^{n \rightarrow n+1} = \frac{p_t^n}{p_t^{n+1}}.$$

- 关键区分:

- 定价在 \mathbb{Q} 或 SDF 下进行;
- 统计建模多在物理测度 \mathbb{P} 下进行;
- 二者通过“风险价格/期限溢价”相联。

Vasicek 模型：SDF 与状态动态

- 设随机贴现因子与单因子状态变量满足：

$$-\log M_{t+1} = \delta + z_t + \lambda \varepsilon_{t+1},$$

$$z_{t+1} = \varphi z_t + (1 - \varphi)\theta + \sigma \varepsilon_{t+1}, \quad \varepsilon_{t+1} \stackrel{\text{i.i.d.}}{\sim} N(0, 1),$$

则 $E(z_t) = \theta$ 。

- 选取 $\delta = \lambda^2/2$, 可使短端利率与 z_t 对应。

短端利率与长期债券价格的仿射形式

- 由正态矩母函数与无套利定价：

$$p_t^0 = 1, \quad p_t^1 = E_t(M_{t+1}), \quad p_t^2 = E_t(M_{t+1}p_{t+1}^1), \dots$$

- 若 $\delta = \lambda^2/2$, 有：

$$\log p_t^1 = -z_t \quad \Rightarrow \quad r_t := -\log p_t^1 = z_t.$$

即状态 z_t 恰为年化对数短期利率。

- 对长期债券，假设价格为仿射形式：

$$-\log p_t^n = A_n + B_n z_t \iff p_t^n = \exp(-A_n - B_n z_t).$$

Vasicek 模型的 Riccati 递推

将仿射形式代回定价方程并配对常数项与 z_t 系数，可得递推：

$$\begin{aligned} A_{n+1} &= A_n + \delta + B_n(1 - \varphi) \theta - \frac{1}{2} (\lambda + B_n \sigma)^2, & A_1 = 0, \quad B_1 = 1. \\ B_{n+1} &= 1 + \varphi B_n, \end{aligned}$$

- B_n 线性累积均值回复系数 φ ；
- A_n 聚合贴现常数、均值回复与“风险价格-方差”项；
- 属于离散时间的一因子高斯-仿射期限结构模型。

Vasicek 模型的优缺点

- 优点：

- 计算简洁，闭式仿射结构，易于生成 p_t^n, y_t^n, f_t ；
- 短端利率服从均值回复 AR(1)，直观可解释；
- 适合作为基准模型或教学示例。

- 缺点：

- 高斯结构允许利率为负；
- 单因子形状受限，长期端“弯曲程度”不足；
- 难以匹配多期限的协同变动与厚尾特征。

- 实务常用多因子或非高斯扩展提升拟合能力。

CIR 模型：平方根扩散与 SDF

- 离散时间设定：

$$\begin{aligned} z_{t+1} &= \varphi z_t + (1 - \varphi) \theta + \sigma z_t^{1/2} \varepsilon_{t+1}, \\ -\log M_{t+1} &= \left(1 + \frac{\lambda^2}{2}\right) z_t + \lambda z_t^{1/2} \varepsilon_{t+1}, \\ \varepsilon_{t+1} &\stackrel{\text{i.i.d.}}{\sim} N(0, 1). \end{aligned}$$

- 这是连续时间费勒平方根过程在离散时间的对应设定：

- 创新项幅度与利率水平的平方根成正比；
- 短端越高，波动越大；
- 更自然地保持非负性。

CIR 模型的仿射价格形式与递推

- 债券价格保持仿射形式：

$$-\log p_t^n = A_n + B_n z_t \iff p_t^n = \exp(-A_n - B_n z_t).$$

- 系数 (A_n, B_n) 满足 (Riccati 型) 递推：

$$A_{n+1} = A_n + B_n(1 - \varphi)\theta, \quad B_{n+1} = 1 + \frac{\lambda^2}{2} + \varphi B_n - \frac{1}{2}(\lambda + B_n\sigma)^2,$$

$$A_1 = 0, \quad B_1 = 1.$$

- 与连续时间 CIR 在风险中性测度下的 Riccati 微分方程相呼应。

CIR 模型的特征与局限

- 核心特征：“水平–波动联动”：
 - 短端较高时， $\sqrt{z_t}$ 放大创新项；
 - 体现出更强波动性与更厚尾部；
 - 均值回复使利率长期受 θ 锚定。
- 相对 Vasicek 的改进：
 - 更自然地保证利率非负；
 - 对 ZLB 前后更为稳健。
- 局限：
 - 单因子版本曲线形状仍较简单，凸度不足；
 - 不同期限收益率的自相关形状过于类似；
 - 离散时间设定仍用正态创新，难刻画极端尾部与跳跃。

从具体模型到统一的仿射族

- 两条理解路径：

- ① 从若干具体可解的特例出发 (Vasicek、CIR 等)；
- ② 构建一套统一的多因子仿射模型族。

- 具体可解的特例：

- CLM 与 Backus, Foresi, and Telmer (1998) 讨论多种高斯创新情形；
- 例如一因子“平方根短端 + 线性-平方根 SDF”：

$$\begin{aligned}r_{t+1} &= \alpha + \beta r_t + \gamma r_t^{1/2} \varepsilon_{t+1}, \\M_{t+1} &= a + b r_t + c r_t^{1/2} \varepsilon_{t+1}.\end{aligned}$$

具体仿射规格的限制与 Gouriéroux–Monfort–Polimenis 模型族

- 具体规格的限制 (dai2000specification, CLM):
 - ① K 因子时隐含债券协方差矩阵可能退化;
 - ② 波动率函数形式受限, 难反映厚尾/异方差;
 - ③ 长期债券风险溢价往往被约束为固定符号。
- Gouriéroux, Monfort, and Polimenis (2006) 提出多因子仿射模型族:

$$\log E(\exp(u'z_{t+1}) | z_t) = a(u)'z_t + b(u),$$

其中 $z_t \in \mathbb{R}^n$ 为马尔可夫过程, $a(\cdot), b(\cdot)$ 给定。

- SDF 对数设为“当前–未来状态”的仿射函数:

$$m_{t+1} := \log M_{t+1} = \gamma_0 + \gamma_1' z_t + \gamma_2' z_{t+1}.$$

价格-收益率的仿射表示与递推

价格与收益率的仿射形式：

$$P_t^n = \exp(c'_n z_t + d_n), \quad y(t, t+n) = -\frac{c'_n z_t + d_n}{n},$$

其中 $n = 2, 3, \dots$ 的系数 (c_n, d_n) 满足递推：

$$c_n = a(c_{n-1} + \gamma_2) - a(\gamma_2) - e_1,$$

$$d_n = d_{n-1} - b(\gamma_2) + b(c_{n-1} + \gamma_2) - e_1,$$

并取 $e_1 = (1, 0)', c_1 = -e_1, d_1 = 0$ 。

- Vasicek 是其中使递推线性化的特殊情况；
- CIR 对应出现二次项的特殊情况；
- 对复杂现金流，可能需要在此基础上再积分或数值近似。

条件累积量的线性结构

- 关键结论：在给定 $z_t = z$ 时，各阶条件累积量均为 z 的线性函数：

$$E(z_{t+1} \mid z_t = z) = a'(0)z + b'(0),$$

$$\text{Var}(z_{t+1} \mid z_t = z) = a''(0)z + b''(0),$$

$$\kappa_3(z_{t+1} \mid z_t = z) = a'''(0)z + b'''(0),$$

$$\kappa_4(z_{t+1} \mid z_t = z) = a^{(4)}(0)z + b^{(4)}(0).$$

- 类比广义线性模型 (GLM):
 - 条件分布由有限维状态控制；
 - 条件矩结构简单，便于计算与识别。
- 与之对照：GARCH 等模型的条件方差依赖滞后项与二次函数，往往不属于该仿射族。

条件累积量的非参数证据：3M 与 10Y

- 图 3、4：

- 3M（日频）与 10Y（月频）收益率的四个条件累积量；
- 采用局部线性核估计，不同带宽。

- 3M：

- 条件均值近似线性、斜率接近 1，持续性强；
- 条件方差随水平显著非线性；
- 条件偏度与超峰度随水平系统变化。

- 10Y：

- 非线性幅度整体更温和；
- 符合“长端更平稳、短端更敏感”的经验事实。

3M 收益率的条件累积量（非参数估计）

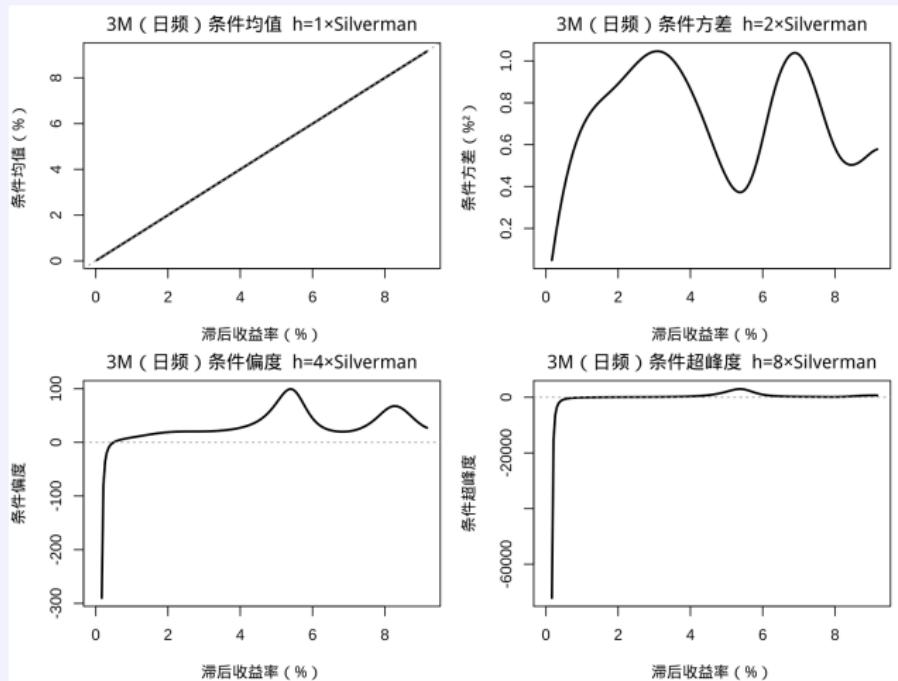


图: 3M (日频) 收益率的四个条件累积量的非参数估计 (局部线性核; 带宽为 1、2、4、8 倍 Silverman 经验法则)。

10Y 收益率的条件累积量 (非参数估计)

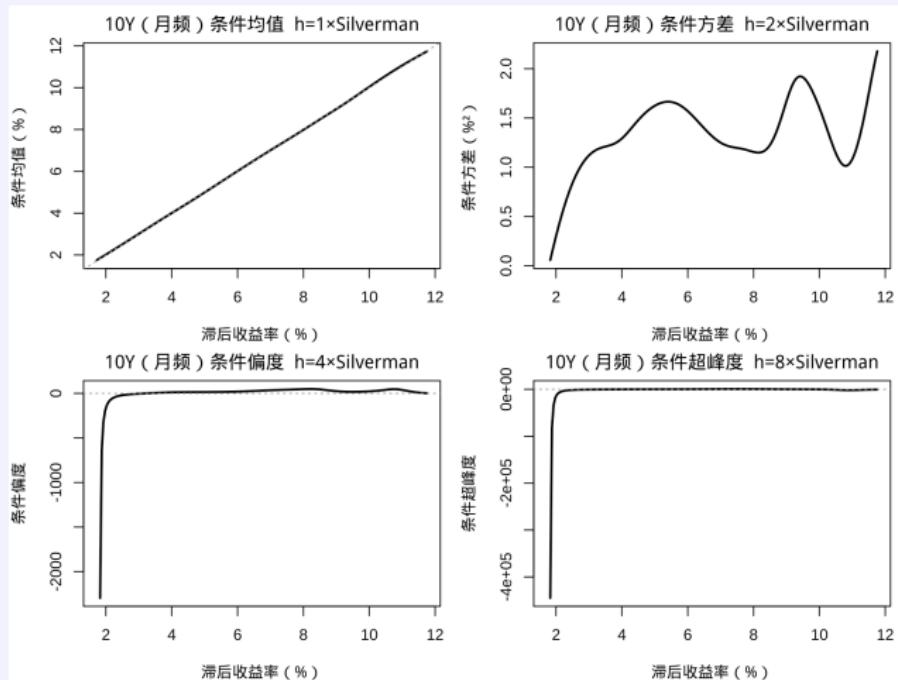


图: 10Y (月频) 收益率的四个条件累积量的非参数估计 (局部线性核; 带宽为 1、2、4、8 倍 Silverman 经验法则)。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

本章小结 (1): 曲线度量与估计

- 核心问题一：如何量化利率期限结构？

- 从零息债出发，梳理贴现函数、收益率曲线和远期利率的关系；
- 用统一口径的真实数据计算“曲线三件套”(y, d, f)。

- 曲线估计的两条实用路径：

- ① 平滑曲线法：

- 将贴现函数视为平滑函数；
- 在“价格 = 现金流 \times 贴现因子”的线性定价框架下，用样条/基函数平滑，并施加（软）无套利约束。

- ② 参数法：

- 使用 NS/NSS 等少量参数控制整体形状；
- 可配合 Fama-Bliss 的分段远期引导。

本章小结（2）：实证事实与仿射模型

- 实证部分（2000–2024 年美债样本）显示：
 - 平均曲线随期限上升，短端波动更大；
 - 差分相关性集中于最短端并快速衰减；
 - ZLB 与疫情冲击带来短端的结构性极值与波动聚集。
- 核心问题二：怎样在无套利下建模？
 - 在“无套利 + 随机贴现因子（SDF）”框架下；
 - 从 Vasicek、CIR 等单因子仿射模型到多因子仿射模型族；
 - 将价格–收益率关系表示为仿射形式，系数由递推确定。
- 操作流程：
 - ① 明确口径（复利、单位、测度）；
 - ② 选方法（样条/参数化/引导），在约束与平滑间平衡；
 - ③ 将曲线用于定价、久期、凸性、预测与风险管理。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

参考文献 |

-  Backus, David K., Silverio Foresi, and Chris I. Telmer (1998). "Discrete-Time Models of Bond Pricing". In: *NBER Macroeconomics Annual* 13, pp. 57–121.
-  Bliss, Robert R. (1997). *Testing Term Structure Estimation Methods*. Working Paper 97-12. Federal Reserve Bank of Atlanta.
-  Campbell, John Y, Andrew W Lo, and A Craig MacKinlay (1997). *The econometrics of financial markets*. Princeton University press.
-  Campbell, John Y. and Robert J. Shiller (1991). "Yield Spreads and Interest Rate Movements: A Bird's Eye View". In: *Review of Economic Studies* 58.3, pp. 495–514.
-  Estrella, Arturo and Frederic S. Mishkin (1997). "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank". In: *European Economic Review* 41.7, pp. 1375–1401.

参考文献 II

-  Fama, Eugene F. and Robert R. Bliss (1987). "The Information in Long-Maturity Forward Rates". In: *American Economic Review* 77.4, pp. 680–692.
-  Fisher, Mark, Douglas W. Nychka, and David Zervos (1995). "Fitting the Term Structure of Interest Rates with Smoothing Splines". In: *Finance and Stochastics* 1.2, pp. 107–134.
-  Gourieroux, Christian, Alain Monfort, and Michael Polimenis (2006). "Affine Models for Discrete-Time Interest Rates". In: *Journal of Econometrics* 131.1–2, pp. 39–73.
-  McCulloch, J. Huston (1971). "Measuring the Term Structure of Interest Rates". In: *Journal of Business* 44.1, pp. 19–31.
-  McCulloch, J. Huston and Heon-Chul Kwon (1993). *U.S. Term Structure Data, 1947–1991*. Working Paper. Ohio State University.

参考文献 III

-  Schaefer, Stephen (1981). "A Note on Nonparametric Estimation of the Term Structure Using Bernstein Polynomials". In: *Journal of Finance* 36.3, pp. 917–930.
-  Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
-  Vasicek, Oldrich and Hsiu-Lang Fong (1982). "Term Structure Modeling Using Exponential Splines". In: *Journal of Finance* 37.2, pp. 339–348.

第 11 章风险管理与尾部风险估计

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

为什么要关注尾部风险？

- 实务与监管长期依赖**标准差/方差 + 正态分布**来度量不确定性：
 - 计算简洁，沟通成本低 (Jorion 1997)；
- 但大量研究（自 Mandelbrot (1963) 以来）表明：
 - 收益分布常呈**尖峰厚尾**，波动具有**时变聚集性**；
 - 极端事件（所谓“ 6σ ”）出现频率远高于正态理论预期；
 - 2008 年危机后监管评估强调：线性-高斯框架可能**系统低估尾部风险**(Turner 2009; Basel Committee on Banking Supervision 2012)。

为什么要“专门研究尾部”？

至少有三方面动因：

- ① 关键决策依赖高置信分位数（99%、99.9%）的损失度量：
 - 正态缩放 (\sqrt{T}) 在这些分位数常严重失真；
- ② 历史样本有限：
 - 真正担心的情形往往位于“样本之外”的更深尾部；
 - 需要在可控假设下对尾部进行**外推**；
- ③ 危机时多资产常“同跌”：
 - 尾部依赖 \neq 常规相关性（线性/均值-方差框架）；
 - 直接关系到分散化是否有效与系统性风险的识别与定价。

中国市场的几类尾部事件

- 2013 年银行间“钱荒”：短端利率在极短时间内飙升，暴露流动性尾部的传导链条。
- 2015 年 A 股异常波动与 2016 年初熔断事件：
 - 提示价格限制与微观结构在极端条件下可能放大波动与相关性。
- 近年来：
 - 房地产及上下游信用事件、城投再融资压力；
 - 衍生品/结构化产品在极端行情下穿仓与追加保证金风险；
 - 尾部体现在**价格、流动性、信用与机制设计**等多个维度。
- 结论：忽视尾部，往往叠加**模型风险与治理风险**。

围绕尾部的工具箱：超出均值-方差

- 分位数/在险价值 (VaR)：作为“通用语言”。
- 稳定分布与非参数分位数估计：稳健刻画中高分位数。
- 极值理论 (EVT) 与正则变化理论：
 - 用尾指数识别尾部厚度；
 - 对“样本之外”的极端分位数进行外推。
- 时间维度：GARCH / CaViaR 等时变风险模型。
- 多维尾部依赖：Copula / CoVaR / 系统性风险分摊。
- VaR + ES：
 - VaR 便于沟通；
 - 期望损失 (ES) 具有更好的理论一致性；
 - 结合形成“可检验、可披露、可操作”的尾部度量闭环。

本章结构：从诊断到沟通

① 诊断：

- 从分位数视角重述 VaR 的优势与局限；
- 对比正态分布与稳定分布下的时间缩放差异。

② 建模与外推：

- 极值理论（EVT）与半参数尾部模型；
- Hill 估计、阈值选择与敏感性分析。

③ 条件 VaR/ES 与回测：

- 条件 VaR/ES 的建模思路与 backtesting 框架；
- 不同置信水平、滚动窗口与样本长度的影响。

④ 多资产与尾部依赖：

- Copula、CoVaR、系统性风险分摊；
- 对分散化与资本配置的意义。

⑤ 度量选型与披露建议：

- 在“可沟通性—一致性—可检验性—可实施性”之间平衡；
- 明确假设与口径，配套回测与敏感性分析。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

分位数的定义与基本性质

定义 (分位数 (Quantile))

设随机变量 X 的分布函数为 $F(x) = \Pr(X \leq x)$, 则 X 的 α 分位数定义为

$$q_\alpha = \inf\{q : \Pr(X \leq q) \geq \alpha\}.$$

若 F 在 q_α 附近严格单调, 则有 $q_\alpha = F^{-1}(\alpha)$ 。

- 仿射变换性质: 对任意 $a > 0$ 与 $b \in \mathbb{R}$,

$$q_\alpha(aX + b) = aq_\alpha(X) + b.$$

- 连续情形下, 左右尾的关系:

$$q_{1-\alpha}(X) = -q_\alpha(-X).$$

- 因此, 只要明确“收益/损失”的号向, 分位数的定义与换算是**一致的**。

正态假设下的 VaR 与时间聚合

- 若 $X \sim N(\mu, \sigma^2)$, 则

$$q_\alpha = \mu + \sigma z_\alpha,$$

其中 z_α 为标准正态分布的 α 分位点 (如 $\alpha = 1\%$ 时 $z_\alpha \approx -2.33$)。

- 样本 X_1, \dots, X_n 下, 正态-i.i.d. 情形的估计:

$$\hat{q}_\alpha = \bar{X} + s z_\alpha,$$

其中 \bar{X}, s^2 为样本均值与方差。

- 时间聚合: 若日收益

$$X_D \sim N\left(\mu_D, \sigma_D^2\right),$$

则 T 日收益之和

$$X_M \sim N\left(T\mu_D, T\sigma_D^2\right),$$

从而

$$q_\alpha(M) = T\mu_D + T^{1/2}\sigma_D z_\alpha.$$

“平方根法则”与其失效

- 当 μ_D 相对 σ_D 很小 (高频收益常见),

$$\text{VaR}_\alpha(T) \approx \sqrt{T} \text{VaR}_\alpha(1).$$

- 正态-i.i.d. 带来的好处:

- 以日度参数快速推导任意持有期 VaR;
- 基于正态近似给出 VaR 的标准误与置信区间。

- 现实中的问题:

- 收益序列具有**尖峰厚尾与时变相关**;
- 正态尾部“过薄”，会把“大亏损”的概率估得过小;
- 序列依赖破坏简单 \sqrt{T} 缩放的有效性;
- \Rightarrow 正态-平方根法则往往**系统低估尾部风险**。

厚尾与稳定分布：时间缩放取决于尾部

- 金融收益常呈厚尾，持有期风险缩放常偏离 \sqrt{T} 。
- 稳定分布用尾指数 θ 刻画极端尾概率与时间缩放：

$$\text{VaR}_\alpha(T) = T^{1/\theta} \text{VaR}_\alpha(1),$$

即使二阶矩不存在也能给出分位数。

- 当 $\theta = 2$: 退化为正态型 \sqrt{T} 缩放；
- 当 $\theta < 2$: 尾更厚, VaR 随 T 增长更快。

稳定分布的定义

定义 (稳定分布)

称实随机变量 X 服从稳定分布，若其特征函数满足

$$\log E\left(e^{iuX}\right) = i\mu u - |\gamma u|^\theta \left(1 + i\beta \operatorname{sign}(u) w(u, \theta)\right), \quad u \in,$$

其中

$$\theta \in (0, 2], \quad \beta \in [-1, 1], \quad \gamma > 0, \quad \mu \in,$$

且

$$w(u, \theta) = \begin{cases} \tan\left(\frac{\pi\theta}{2}\right), & \theta \neq 1, \\ -\frac{2}{\pi} \log |u|, & \theta = 1. \end{cases}$$

稳定分布的定义

- θ : 特征指数/尾指数 (尾越厚 θ 越小);
- β : 偏度 ($\beta = 0$ 为对称);
- γ : 尺度, μ : 位置。

稳定分布的特例与矩存在性

例 (常见特例)

- 正态分布: 当 $\theta = 2$ 时, $w \equiv 0$,

$$\log E(e^{iuX}) = i\mu u - \gamma^2 u^2 \Rightarrow X \sim N(\mu, 2\gamma^2).$$

- 柯西分布: 当 $\theta = 1$ 时, $w(u, 1) = -(2/\pi) \log |u|$, 得到厚尾的 Cauchy 分布。
- 矩存在性:

$$E|X|^p < \infty \Leftrightarrow p < \theta.$$

除 $\theta = 2$ 外, 一般方差不存在; 仅当 $\theta > 1$ 时才有有限均值。

稳定性与 VaR 的持有期缩放

定义 (稳定性/可加性与持有期缩放)

若 X_1, \dots, X_T 独立同分布且各自稳定 (相同 θ, β)，则和 $S_T = \sum_{t=1}^T X_t$ 仍为稳定分布，且

$$S_T \stackrel{d}{=} T^{1/\theta} X_1 + T \mu_0,$$

适当中心化后 (零均值或忽略平移项) 有

$$\text{VaR}_\alpha(T) = T^{1/\theta} \text{VaR}_\alpha(1).$$

- $\theta = 2$: 恢复 \sqrt{T} ;
- $\theta < 2$: 尾更厚, VaR 随 T 的增长更快;
- 解释了厚尾资产“持有期风险膨胀”快于 \sqrt{T} 的现象。

稳定分布在实务中的使用与难点

- 方差并非唯一风险刻画：

- 对 $\theta < 2$ 的稳定分布，方差不存在；
- 分位数/尾部概率仍可清晰定义，适合 VaR/ES 等指标。

- 时间聚合不服从平方根法则：

$$\text{VaR}_\alpha(T) \propto T^{1/\theta}, \quad \theta \neq 2.$$

- 参数估计的难点：

- 稳定分布参数（尤其 θ, β ）的 MLE 数值不稳定；
- 实务中常结合非参数分位数（中等尾部）与EVT（极端尾部）校准。

非参数分位数估计与 EVT 动机

- 经典结果（在温和条件下）：

$$\sqrt{T}(\hat{q}_\alpha - q_\alpha) \xrightarrow{d} N\left(0, \frac{\alpha(1-\alpha)}{[f(q_\alpha)]^2}\right),$$

其中 f 为密度 (Koenker 2005)。

- 可用来构造分位数的近似置信区间。
- 限制：
 - 当 $\alpha < 1/T$ 或 $\alpha > 1 - 1/T$ 时，
 - 非参数方法无法对“样本之外”的极端分位数外推。
- 这正是引入**极值理论 (EVT) 与半参数尾部模型**的动机：
 - 对尾部分布进行结构化建模；
 - 在有限样本下稳健外推更深尾部的 VaR/ES。

VaR 的定义与监管语境

- 若损失随机变量为 X , 则

$$\text{VaR}_\alpha = q_\alpha(X), \quad \Pr(X \leq \text{VaR}_\alpha) = \alpha.$$

- 正态假设下:

$$\text{VaR}_\alpha = \mu + \sigma z_\alpha.$$

- VaR 的优势:

- 以“分位数”为核心, 直观易懂;
- 计算简洁, 便于监管与内部沟通;
- 已被纳入《巴塞尔协议》等监管框架 (Basel Committee on Banking Supervision 2012)。

- 局限:

- 一致性不足 (非次可加);
- 在厚尾下易低估极端损失;
- 需与 EVT、ES 等工具结合。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama–Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

极大值与“尾部行为”的极限

- 从风险管理视角，我们关心样本极端值随样本量增大的极限行为。

- 令样本极大值

$$M_T = \max_{1 \leq t \leq T} \{X_t\}.$$

- 问题：当 T 增大时， M_T 如何变化？

- 总体最大值对应总体分布的 $\alpha = 1$ 分位数；
- M_T 可视作 “ $1 - 1/T$ 分位数”的经验对应物。

- 极小值可通过 $-\max(-X_t)$ 处理，这里集中讨论极大值。

极大值的极限定理与 GEV 分布

定理 (i.i.d. 情形的极值极限定理)

设 X_t 独立同分布。在温和正则条件下，存在常数序列 $a_T > 0, b_T \in \mathbb{R}$ ，使得

$$a_T(M_T - b_T) \xrightarrow{d} \mathcal{E},$$

其中极限随机变量 \mathcal{E} 的分布为

$$G_\gamma(x) = \Pr(\mathcal{E} \leq x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \quad 1 + \gamma x > 0.$$

- G_γ 即广义极值 (GEV) 分布。
- 经典参考：Fisher and Tippett (1928), Gnedenko (1943), and Embrechts, Klüppelberg, and Mikosch (1997)。

GEV 的三种类型与极小值

- 形状参数 γ 决定三种情形：
 - ① $\gamma > 0$: Fréchet (重尾);
 - ② $\gamma = 0$: Gumbel (指数尾);
 - ③ $\gamma < 0$: Weibull (有限端点)。

- 极小值：
 - 对 $-X_t$ 应用同样结果；
 - $\min X_t = -\max(-X_t)$ ；
 - 一般而言，样本极大值与极小值并不独立。

- 向量情形：
 - 可逐分量定义坐标极大值；
 - 进一步研究分量间极端事件的“同步性”或尾部依赖。

例子：均匀分布的极值

例 (均匀分布)

若 $X_t \sim \text{Unif}[0, 1]$ i.i.d., 则 $M_T \rightarrow 1$, 且

$$\Pr(M_T \leq x) = x^T, \quad 0 \leq x \leq 1.$$

取 $x_T = 1 - \frac{x}{T}$ ($x > 0$), 则

$$\Pr(T(M_T - 1) \leq -x) = \Pr(M_T \leq x_T) = \left(1 - \frac{x}{T}\right)^T \rightarrow \exp(-x).$$

因此 $a_T = T$, $b_T = 1$, 极限分布为指数分布 (对应 GEV 中 $\gamma = -1$)。

例子：正态与柯西分布的极值

例 (正态分布)

若 $X_t \sim N(0, 1)$, 则 $M_T \xrightarrow{P} \infty$, 且在 $\gamma = 0$ 下有

$$a_T = \sqrt{2 \ln T}, \quad b_T = \sqrt{2 \ln T} - \frac{\ln(4\pi) + \ln \ln T}{2\sqrt{2 \ln T}},$$

极限 CDF 为 Gumbel:

$$G(x) = \exp(-\exp(-x)).$$

例子：正态与柯西分布的极值

例 (柯西分布)

若 X_t 为标准柯西， $f_{X_t}(x) = 1/\{\pi(1+x^2)\}$ ，则 $M_T \xrightarrow{P} \infty$ ，且在 $\gamma = 1$ 下有

$$a_T = \pi/T, \quad b_T = 0,$$

极限 CDF：

$$G(x) = \exp(-x^{-1}), \quad x > 0,$$

密度 $g(x) = \exp(-x^{-1})/x^2$ 。

弱依赖下的极值: Berman 条件

Proposition (自相关情形: Berman (1964) 条件)

若 $\{X_t\}$ 为边缘分布 $N(0, 1)$ 的平稳高斯过程, 自相关函数 $\rho(k)$ 满足

$$\rho(k) / \ln k \rightarrow 0 \quad (k \rightarrow \infty),$$

则极大值极限定理仍成立, 且可选用与 *i.i.d.* 情形相同的归一化序列 a_T, b_T 。

- 对更一般的非高斯过程也有类似结论:
 - 弱依赖通常不改变收敛速度;
 - 但可能影响极限分布的具体形态 (Embrechts, Klüppelberg, and Mikosch 1997)。

多维高斯与尾部“独立性”的问题

Proposition (多维高斯分布： 极大值的渐近独立性)

若

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad |\rho| < 1,$$

则当 $T \rightarrow \infty$ 时， $\max_{1 \leq t \leq T} \{X_t\}$ 与 $\max_{1 \leq t \leq T} \{Y_t\}$ (以及对应极小值) 渐近独立。

- 在一定弱依赖高斯过程中也成立 (Embrechts, Klüppelberg, and Mikosch 1997)。
- 含义：
 - 高斯世界里，不同资产的极端值不易“同一时点”发生；
 - 与现实中危机期常见的尾部共振、同跌现象不符；
 - 提示高斯假设对尾部依赖的刻画严重不足。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

为什么需要半参数尾模型？

- 样本只能可靠估计

$$\alpha \geq 1/T, \quad \alpha \leq 1 - 1/T$$

对应的分位数；当 T 较小，这覆盖不了极端小/大概率。

- 引入模型：对非常小或非常大的 α 给出合理预期。
- 关注尾部厚度（tail thickness）的若干分布族：
 - 正态分布：薄尾， $\Pr(X > x)$ 衰减很快，所有矩存在；
 - t 分布：允许更厚的尾；
 - Pareto 分布：广泛用于收益/收入分布，对尾部形状更灵活。
- 实务中往往只关心“尾部甚至极端尾部”，无需对整个分布做完全参数化：
 - \Rightarrow 半参数思路：仅在尾部假设 Pareto 型行为。

Pareto 分布与尾指数

定义 (Pareto 分布)

若对 $x \geq L$ 有

$$F(x) = \Pr(X \leq x) = 1 - L^\kappa x^{-\kappa}, \quad \kappa > 0, \quad L > 0,$$

则称 X 在 $[L, \infty)$ 上服从 Pareto 分布。

- 密度: $f(x) = \kappa L^\kappa / x^{\kappa+1}$ 。

- 分位函数:

$$q_{1-\alpha} = L \alpha^{-1/\kappa}.$$

- 尾指数 κ 的含义:

- κ 越小, 尾越厚;
- $E(|X|^\gamma) < \infty \iff \gamma < \kappa$;
- 控制极端事件的频率与矩存在性。

为什么只在尾部用 Pareto ?

- 完全参数化的 Pareto:
 - 可用极大似然直接估 VaR;
 - 但强行用在整个分布上，主体部分可能失真。
- 半参数折中：
 - 仅要求“足够大”的 x 上尾部近似 Pareto;
 - 主体部分可以是任意形状（正态、GED、混合分布等）。
- 例：柯西分布
 - 尾部与 Pareto 同阶 ($\kappa = 1$);
 - 但主体行为与标准 Pareto 完全不同。
- 实务重点：
 - 我们只需要尾部的幂律结构来做极端 VaR 外推。

半参数尾模型：正则变化思想

定义 (半参数尾模型)

设 X 的分布函数为 F , 若当 $x \rightarrow \infty$ 时

$$1 - F(x) \sim L(x) x^{-\kappa},$$

其中 $\kappa > 0$ 为尾指数, $L(x)$ 为常数或慢变函数 (对所有 $a > 0$, $\lim_{x \rightarrow \infty} L(ax)/L(x) = 1$), 则称 X 拥有 Pareto 型尾部。

若 F 有密度 f , 则

$$f(x) \sim \kappa L(x) x^{-(\kappa+1)}.$$

对应分位函数 ($\alpha \rightarrow 0$):

$$q_{1-\alpha} = L^*(1/\alpha) \alpha^{-1/\kappa},$$

其中 $L^*(1/\alpha)$ 为常数或慢变函数。

正则变化尾与尾部厚度

定义 (正则变化尾; 尾部厚度)

若存在 $\kappa \in (0, \infty)$ 使得对任意 $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\kappa},$$

则称 X 的尾部厚度为 κ 。

例 (慢变函数示例)

若 $L(x) = c \log x$, 则

$$\frac{1 - F(tx)}{1 - F(t)} \sim \frac{c \log t + c \log x}{c \log t} x^{-\kappa} \sim x^{-\kappa}, \quad t \rightarrow \infty.$$

正则变化尾与尾部厚度

- $L(x)$ 的影响通常二阶; κ 是核心参数:
 - $\gamma \geq \kappa \Rightarrow E(|X|^\gamma) = \infty$;
 - $\gamma < \kappa \Rightarrow$ 对应矩有限。

模型的“覆盖范围”与排除的情况

- 本模型只约束“ x 很大时”的行为（半参数）：

- 不要求 x 近 0 的形状；
- 正好契合“只关心样本极值”的需求。

- 极值理论中的极值指数 γ 与 κ 关系为

$$\kappa = 1/\gamma.$$

- $\kappa > 0$ 排除两类情形：

- $\gamma \leq 0$: 对应薄尾或有限端点（如正态）；
- 超厚尾（如 log-Cauchy），其衰减比任意 $x^{-\kappa}$ 更慢。

- Pareto 尾模型：

- 在物理、社会科学中广泛出现（Zipf、Gibrat 等）；
- 对高频收益/成交量尾部往往更贴合；
- 比稳定分布族更灵活，更方便与 GARCH 等结合。

从排序到 Hill 估计的思路

- 排序：

$$X_{(1)} > X_{(2)} > \cdots > X_{(T)}.$$

- 取阈值整数 $M < T$, 对大阶次统计量 $j \geq M + 1$:

$$\log \Pr(X > X_{(j)}) \simeq \kappa \log X_{(j)} + \log L(X_{(j)}).$$

- 减去 $j = M + 1$ 的对应式, 并利用慢变性 $L(X_{(j)})/L(X_{(M+1)}) \rightarrow 1$:

$$\log\left(\frac{j}{M+1}\right) \simeq \kappa \log\left(\frac{X_{(j)}}{X_{(M+1)}}\right).$$

- 再用积分近似

$$\log(M+1) - \frac{1}{M} \sum_{j=1}^M \log j \rightarrow 1,$$

$$\frac{1}{M} \sum_{j=1}^M \log\left(\frac{X_{(j)}}{X_{(M+1)}}\right) \simeq \frac{1}{\kappa}.$$

- 由此导出Hill 估计量。

Hill 估计量与极端 VaR 外推

定义 (Hill 估计量)

给定阈值 M , 定义

$$\frac{1}{\hat{\kappa}} = \frac{1}{M} \sum_{j=1}^M \log \frac{X_{(j)}}{X_{(M+1)}},$$

并估计极端分位:

$$\hat{q}_{1-\alpha} = \hat{L}_T \alpha^{-1/\hat{\kappa}}, \quad \hat{L}_T = X_{(M+1)} \left(\frac{M}{T} \right)^{1/\hat{\kappa}}, \quad \alpha < M/T.$$

- 也可用 $\hat{L}_T = X_{(1)} / T^{1/\hat{\kappa}}$:
 - 当 $\alpha = 1/T$ 时, $\hat{q}_{1-\alpha}$ 恰为样本最大值。
- 典型应用: 对高置信水平 VaR 进行**样本外推**。

Hill 估计的渐近性质

定理 (渐近正态性)

若满足 *von Mises* 条件

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{1 - F(x)} = \kappa > 0, \quad f = F',$$

且 $M \rightarrow \infty$, $M/T \rightarrow 0$, 则

$$M^{1/2}(\hat{\kappa} - \kappa) \xrightarrow{d} N(0, \kappa^2).$$

- 渐近标准误 $\approx \kappa/\sqrt{M}$ (i.i.d. 情形)。
- 阈值 M 的选择要在:
 - 太小: 方差大;
 - 太大: 偏差大 (非极端样本混入);
 - 之间取得折中。
- 实务中常用 Hill 曲线 ($\hat{\kappa}$ 随 M 的轨迹) + 诊断图辅助选择 M 。

其他尾指数估计与时间序列扩展

- 对数秩回归 (Gabaix and Ibragimov 2011):

$$\log(i - \delta) = a - b \log X_{(i)}, \quad i = 1, \dots, M, \quad \delta \in [0, 1),$$

斜率 b 估计 κ , 推荐 $\delta = 1/2$ 提升有限样本表现。

- DEM 估计 (Dekkers, Einmahl, and Haan 1989):

$$\hat{\gamma}_{\text{DEM}} = H_T^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(H_T^{(1)})^2}{H_T^{(2)}} \right)^{-1}, \quad H_T^{(j)} = \frac{1}{M} \sum_{i=1}^M (\log X_{(i)} - \log X_{(M+1)})^j,$$

对任意 $\gamma = 1/\kappa \in$ 相合; $\gamma > 0$ 时比 Hill 稍低效。

- 时间序列扩展:

- 在弱依赖条件下, Hill 相合性可推广到平稳序列 (Resnick and Stărică 1998);
- Hill (2010) 给出 $M^{1/2}(\hat{\kappa} - \kappa) \Rightarrow N(0, \Phi)$;
- 一般 $\Phi \geq \kappa^2$, 与依赖结构有关; 部分强 GARCH 中 $\Phi = \kappa^2$ 。

半参数尾模型的实证应用 (S&P 500)

- 以标普 500 日对数收益为例：

- 构造亏损变量 $Y = -r$, 聚焦左尾。
- 使用 Hill 与 DEM 估计尾指数, 绘制 Hill 曲线随阈值 M 的变化。
- 对数秩回归 ($\delta = 1/2$) 作为 rank-size 佐证。
- 诊断图：
 - CCDF 双对数图 (幂律 \Rightarrow 近线性);
 - mean excess 图 (Pareto 尾 \Rightarrow 近线性上升)。

- 在选定 M 下, 用

$$\hat{q}_{1-\alpha} = \hat{L}_T \alpha^{-1/\hat{\kappa}}$$

对极端 VaR 做样本外推, 并与经验分位数对比。

- 注意：外推仅在 $\alpha < M/T$ 的范围内可信。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

从无条件到条件 VaR

- 前述讨论的是无条件分布与尾部。
- 实务中通常拥有丰富的**时间序列信息**：
 - 波动聚集、杠杆效应、结构变化等。

- 给定信息集 \mathcal{F}_{t-1} , 考虑条件分布:

$$X_t \mid \mathcal{F}_{t-1} \Rightarrow q_\alpha(X_t \mid \mathcal{F}_{t-1}).$$

- 动态模型可以:
 - 让 VaR 随市场状态/波动**时变**;
 - 提升危机前夕的风险预警能力。

GARCH(1,1) 模型与条件 VaR

- 设收益服从 GARCH(1,1):

$$X_t = \mu + \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \gamma (X_{t-1} - \mu)^2.$$

- 若 $\varepsilon_t \sim N(0, 1)$, 则

$$q_\alpha(X_t | \mathcal{F}_{t-1}) = \mu + \sigma_t z_\alpha.$$

- 若认为正态性过强, 仅假设 ε_t i.i.d. 且密度 f 未知:

$$q_\alpha(X_t | \mathcal{F}_{t-1}) = \mu_t + \sigma_t w_\alpha,$$

其中 w_α 为 ε_t 的 α 分位数。

- 实务估计:

- 用标准化残差 $\hat{\varepsilon}_t$ 的样本分位估计 w_α ;
- 极端 α 时, 可对 $\hat{\varepsilon}_t$ 应用 Hill 外推。

CaViaR: 直接建模 VaR 动态

定义 (CaViaR 思路 (Engle and Manganelli 2004))

直接对 VaR 序列建模, 使

$$\Pr(X_t < -\text{VaR}_t(\alpha) \mid \mathcal{F}_{t-1}) = \alpha.$$

如若略去 α 显式记号, 典型规格包括:

$$\text{VaR}_t = \text{VaR}_{t-1} + \beta [1\{X_{t-1} \leq -\text{VaR}_{t-1}\} - \alpha],$$

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |X_{t-1} - \mu|,$$

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |X_{t-1}|.$$

- 估计: 最小化分位回归损失

$$Q_T(\beta) = \frac{1}{T} \sum_{t=1}^T \rho_\alpha(X_t - \text{VaR}_t(\beta)).$$

- 在适度平稳/混合条件下, 可得参数的 CLT。

GARCH VaR 与 CaViaR 的关系

- GARCH(1,1) VaR:

$$\text{VaR}_t = \sigma_t z_\alpha = (\omega + \alpha \sigma_{t-1}^2 + \gamma X_{t-1}^2)^{1/2} z_\alpha.$$

- 这可以解读为 CaViaR 的一类间接 GARCH 规格：
 - 先对条件波动建模，再乘以分位点；
 - 与直接对 VaR 做自回归/自回归-外生变量 (CAViaR) 形成对照。
- 对比：
 - GARCH：结构清晰，便于扩展 (t、GED 等)；
 - CAViaR：直接针对 VaR，避免对完整条件密度做假设。

VaR 的回测：超越频率与校准

给定由历史数据 $\{X_{t-T}, \dots, X_{t-1}\}$ 构造的 $\text{VaR}_t(\alpha)$, 在评估期 $\{1, \dots, H\}$ 上:

- 定义“超越”指标:

$$\psi_\alpha(x) = \mathbf{1}\{x > 0\} - (1 - \alpha),$$

$$\hat{\rho}_\alpha = \frac{1}{H} \sum_{t=1}^H \psi_\alpha(X_t - \text{VaR}_t).$$

- 若模型正确，则
 - $E(\psi_\alpha) = 0$;
 - 即超越频率与名义 α 一致。
- 换言之: $\Pr(X_t < -\text{VaR}_t) \approx \alpha$ 。

Kupiec 检验与独立性检验

- Kupiec 无条件覆盖 (UC) 检验：
 - 检查超越次数 n 是否与 $H\alpha$ 一致；
 - 以二项分布为基础构造似然比检验。
- 独立性检验：
 - 构造命中序列 $I_t = \mathbf{1}\{X_t < -\text{VaR}_t\}$ ；
 - 检验 $\{I_t\}$ 是否服从独立 Bernoulli；
 - 可用一阶马尔可夫链的 LR 检验。
- 综合：
 - UC：校准是否正确（频率对不对）；
 - 独立性：是否存在“成串失败”（集中的超越）；
 - 两者合在一起给 VaR 模型打“体检报告”。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法：Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型（离散时间仿射形式）
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

整合实证脚本的目标

- 标的: S&P 500 指数 (Yahoo: ^GSPC) 日对数收益。
- 目标: 从无条件尾部风险到条件 VaR 的全流程实证:
 - ① 模块 A: 滚动 VaR 与样本外回测;
 - ② 模块 B: 尾部厚度的半参数模型;
 - ③ 模块 C: 动态 VaR 模型与回测。
- 技术上:
 - 统一封装 Hill / DEM / VaR 外推 / Kupiec 检验等工具;
 - 自动安装 R 包, 减少环境配置成本;
 - 顶部布尔开关 RUN_A,RUN_B,RUN_C 控制模块。

模块 A: 滚动 VaR 与回测

- 数据: ^GSPC 收盘价, 转为日对数收益。
- 固定窗口 (约 250 日) 滚动估计 1 日 VaR ($\alpha = 1\%, 5\%$):
 - ① 历史模拟: 窗口内经验分位数;
 - ② 高斯 i.i.d.: 滚动均值 + 标准差;
 - ③ GARCH(1,1)- t : 使用 rugarch 的 ugarchroll;
 - ④ EVT-POT: 以 95% 阈值拟合 GPD, 给出无条件 VaR。
- 样本外评估:
 - Kupiec UC 检验 (违约次数与 α 对比);
 - 图形展示: 最近一年收益 vs 四种 99% VaR 轨迹。

模块 B: 尾部厚度的半参数模型

- 构造亏损 $Y = -r$, 假设右尾正则变化。
- 步骤:
 - ① 用 Hill/DEM 估计尾指数 κ , 绘制 Hill 曲线;
 - ② 用滚动标准差寻找“平台段”确定阈值 M ;
 - ③ 对数秩回归 ($\delta = 1/2$) 作 rank-size 佐证;
 - ④ 绘制 CCDF 双对数图与 mean excess 图;
 - ⑤ 在选定 M 下用 Pareto 模型外推极端 VaR。
- 外推结果与经验分位数在双对数坐标下对比:
 - 检查幂律拟合区域;
 - 提示“多远的尾部”可以相信模型。

模块 C: 动态 VaR 模型与回测

- GARCH-norm:
 - 在估计期拟合 GARCH(1,1)-正态;
 - 固定参数滤波全样本, 得到 μ_t, σ_t ;
 - 计算 $q_\alpha = \mu_t + \sigma_t z_\alpha$ 。
- 准非参 VaR:
 - 用估计期残差估计 ε_t 的左尾分位 w_α ;
 - 极端 α 时对残差尾部做 Hill 外推;
 - 得到 $q_\alpha = \mu_t + \sigma_t w_\alpha$ 。
- CaViaR (SAV):

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |X_{t-1}|.$$
- 在评估期上:
 - 对三种 VaR 做 Kupiec UC + 独立性检验;
 - 作图对比三条 VaR 轨迹的形态与平滑度。

小结：尾厚度 + 动态 VaR 的组合使用

- 半参数尾模型：
 - 提供**极端 VaR 水平**的稳健外推；
 - 关键在于尾指数 κ 与阈值 M 的选择。
- 动态 VaR 模型 (GARCH/CaViaR)：
 - 捕捉波动聚集与状态依赖；
 - 提供**路径上的时变 VaR**。
- 实务建议：
 - 先用半参数模型诊断回报尾部厚度；
 - 在动态模型中使用更合理的分位点（如 w_α 而非 z_α ）；
 - 始终用回测（UC + 独立性 + 图形）检验口径与校准。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

从单变量到多变量尾部风险

- 前文聚焦单变量：尾部厚度、极端 VaR/ES 等。
- 实务中更关心**多资产组合**：
 - 边际（单个资产）风险；
 - 资产之间的**依赖结构**，尤其尾部依赖。
- 常用思路：**边际—依赖分离建模**：
 - 先将各维随机变量用分位映射到 $[0, 1]$ ：

$$Y_1 = F_{X_1}(X_1), \quad Y_2 = F_{X_2}(X_2),$$

则 $Y_1, Y_2 \sim \text{Unif}[0, 1]$ ；

- 再用 Copula C 刻画 (Y_1, Y_2) 的联合分布。
- 这种分解思想由 Sklar 定理形式化。

Sklar 定理与 Copula 的定义

定理 (Sklar 定理 (Sklar 1959))

若 X_1, X_2 连续, 则存在唯一的 $C: [0, 1]^2 \rightarrow [0, 1]$ 使

$$\Pr(X_1 \leq x_1, X_2 \leq x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2)).$$

- 定义 Copula:

$$C(u_1, u_2) = \Pr(Y_1 \leq u_1, Y_2 \leq u_2), \quad u_1, u_2 \in [0, 1].$$

- (X_1, X_2) 的联合分布可写为:

$$F(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2)).$$

- 实现“边际—依赖解耦”:

- 边际 F_{X_1}, F_{X_2} 单独建模;
- Copula C 单独刻画依赖结构 (含尾依赖与不对称)。

高斯 Copula 示例

例 (高斯 Copula)

$$C(u_1, u_2; \rho) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho),$$

其中 Φ 为标准正态 CDF, $\Phi_2(\cdot, \cdot; \rho)$ 为相关系数为 ρ 的二维标准正态 CDF。

对应联合分布与密度:

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2); \rho),$$

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2); \rho) f_1(x_1) f_2(x_2),$$

其中 $c = \partial_{u_1} \partial_{u_2} C$ 为 Copula 密度。

- 高斯 Copula 在信用风险中应用广泛:

- 对边际分布无须正态假设;
- 但尾依赖为零: 极端情形下“同涨同跌”的聚集性可能被低估。

尾依赖与对称 Joe–Clayton (SJC) Copula

定义 (对称 Joe–Clayton (SJC) Copula (Patton 2006))

设

$$C_{\text{SJC}}(u, v) = \frac{1}{2} \left(C_{\text{JC}}(u, v | \tau^U, \tau^L) + C_{\text{JC}}(1 - u, 1 - v | \tau^L, \tau^U) + u + v - 1 \right),$$

$$1 - C_{\text{JC}}(u, v | \tau^U, \tau^L) = \left(1 - [(1 - (1 - u)^\kappa)^{-\gamma} + (1 - (1 - v)^\kappa)^{-\gamma} - 1]^{-1/\gamma} \right)^{1/\kappa},$$

其中 $\kappa = 1 / \log_2(2 - \tau^U)$, $\gamma = -1 / \log_2(\tau^L)$ 。

- 尾依赖系数:

$$\tau^L = \lim_{\varepsilon \rightarrow 0+} \frac{C(\varepsilon, \varepsilon)}{\varepsilon}, \quad \tau^U = \lim_{\varepsilon \rightarrow 1-} \frac{C(\varepsilon, \varepsilon)}{\varepsilon}.$$

- SJC Copula 允许上下尾依赖不对称。
- 实证: Patton (2006) 指出欧元–美元与日元–美元汇率在贬值期相关性更强 (上、下尾不对称)。

系统性风险度量：CoVaR

定义 (系统性风险的 CoVaR (Adrian and Brunnermeier 2016))

机构 j 的条件 VaR 定义为

$$\Pr(X^j \leq \text{CoVaR}_{\alpha}^{j|\mathcal{C}(X^i)} \mid \mathcal{C}(X^i)) = \alpha,$$

其中 $\mathcal{C}(X^i)$ 表示对机构 i 的状态施加的约束。

系统性风险度量：CoVaR

- 常用设定：

$$\mathcal{C}(X^i) : X^i = \text{VaR}_\alpha^i \quad \text{或} \quad X^i = \text{Median}(X^i).$$

- 边际贡献定义：

$$\Delta \text{CoVaR}_{\alpha}^{ji} = \text{CoVaR}_{\alpha}^{j|X^i=\text{VaR}_{\alpha}^i} - \text{CoVaR}_{\alpha}^{j|X^i=\text{Median}(X^i)}.$$

- 直观：衡量“当机构 i 处于自身 VaR 尾部时，机构 j 的 VaR 被额外拖累多少”。

Copula 与 CoVaR 的实证演示（思路）

- 标的： $\hat{\text{GSPC}}$ 与 $\hat{\text{IXIC}}$ 的日对数收益。
- 步骤：
 - ① 用经验 CDF 将收益映射为伪观测 $U \in (0, 1)$ ；
 - ② 分别拟合高斯 Copula 与 Joe–Clayton/BB7 Copula；
 - ③ 报告尾依赖系数 λ_L, λ_U ；
 - ④ 用 Copula 的条件分布逆 (H^{-1}) 计算 CoVaR 与 ΔCoVaR ：
$$\text{CoVaR}_{\alpha}^{j|i} : \Pr(X^j \leq \text{CoVaR}_{\alpha}^{j|X^i=\text{VaR}_{\alpha}^i} \mid X^i = \text{VaR}_{\alpha}^i) = \alpha.$$
- 可视化：
 - U 空间散点图；
 - 原始收益空间中，给定 X^i 极端条件下 X^j 的条件分位线。
- 结论：尾部依赖强度与形状对 CoVaR 与系统性风险溢出有关键影响。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

什么是“好的”风险度量？

- 记 $R(W)$ 为与头寸价值 W 关联的风险度量。
- Artzner et al. (1999) 提出“四公理”，刻画所谓一致性风险度量：
 - 单调性；
 - 现金不变性（平移不变性）；
 - 齐次性；
 - 次可加性。
- $R(W)$ 可理解为“所需资本缓冲”： $R(W)$ 越大，头寸越危险。

一致性风险度量的四公理

定义 (四公理 (Artzner et al. 1999))

对任意头寸 W, W^* 和常数 c :

- **单调性:** 若 W 在一阶随机占优意义上劣于 W^* , 则 $R(W) \geq R(W^*)$;
- **现金不变性:** $R(W+c) = R(W) - c$;
- **齐次性:** $\forall \lambda \geq 0, R(\lambda W) = \lambda R(W)$;
- **次可加性:** $R(W + W^*) \leq R(W) + R(W^*)$ 。

四公理的直观含义

- **单调性：**
 - 若 W 在所有分位上都不优于 W^* ，所需资本不应更低。
- **现金不变性：**
 - 向头寸中加入确定性现金 c ，风险应恰好减少 c ；
 - 体现“现金一元等价抵消一元风险”。
- **齐次性：**
 - 在无规模不经济 / 交易成本假设下，放大 λ 倍，风险线性放大。
- **次可加性（分散化原则）：**
 - 组合风险不应超过分拆后风险之和；
 - 鼓励汇总与分散，而不是人为拆解规避监管。

标准差：部分满足一致性

例 (标准差满足齐次性与次可加性)

由

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y) \leq (\text{sd}(X) + \text{sd}(Y))^2$$

可知 $\text{sd}(X + Y) \leq \text{sd}(X) + \text{sd}(Y)$ 。

- 标准差：
 - 满足齐次性与次可加性；
 - 但对常数平移不满足 $R(W + c) = R(W) - c$, 不现金不变。
- 说明“波动”本身不是完全合格的货币化风险度量。

VaR 的非次可加性：反例

例 (VaR 的非次可加性反例)

设 X, Y 相互独立同分布，且

$$\Pr(X = -100) = 0.04, \Pr(X = 0) = 0.96.$$

则 5% VaR 对单资产均为 0。但

$$X + Y = \begin{cases} -200, & 0.0016 \\ -100, & 0.0768 \\ 0, & 0.9216 \end{cases} \Rightarrow \text{VaR}_{5\%}(X + Y) = 100.$$

$$\Rightarrow \text{VaR}_{5\%}(X + Y) > \text{VaR}_{5\%}(X) + \text{VaR}_{5\%}(Y),$$

VaR 违反次可加性。

- 说明：某些分布下，简单相加组合可能“被罚”得更重。

VaR 何时是“好的”？

- Ibragimov (2009): 在包括 α -稳定分布 ($\alpha > 1$) 的广泛分布类中, VaR 满足次可加性。
- 例如: 若 $X_j \sim N(\mu_j, \sigma_j^2)$ 独立, 则

$$\text{VaR}_\alpha(X_1 + X_2) = \mu_1 + \mu_2 + \sqrt{\sigma_1^2 + \sigma_2^2} z_\alpha \leq \text{VaR}_\alpha(X_1) + \text{VaR}_\alpha(X_2).$$

- 当稳定指数 $\alpha < 1$ 时, VaR 甚至可能超可加 (多样化有害) (Danielsson et al. 2013)。
- 提示: 需结合分布类别与尾部特征来判断 VaR 的适用性。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

ES 的定义与 VaR 的关系

定义 (ES 的定义)

记 q_α 为 X 的 α 分位数, 则

$$ES_\alpha(X) = -E(X | X \leq q_\alpha) = -\frac{E[X \mathbf{1}_{\{X \leq q_\alpha\}}]}{\Pr(X \leq q_\alpha)} = -\frac{1}{\alpha} E[X \mathbf{1}_{\{X \leq q_\alpha\}}].$$

若 X 连续,

$$ES_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\tau(X) d\tau.$$

- ES 是“尾部平均损失”, 而非单点分位;
- 自然反映 VaR 之下更深尾部的严重程度。

ES 的次可加性

定理 (ES 的次可加性)

令 $Z = X + Y$, 则

$$\alpha [ES_\alpha(X) + ES_\alpha(Y) - ES_\alpha(Z)] = E[Z \mathbf{1}_{\{Z \leq q_\alpha(Z)\}} - X \mathbf{1}_{\{X \leq q_\alpha(X)\}} - Y \mathbf{1}_{\{Y \leq q_\alpha(Y)\}}]$$

- 使用指示函数符号比较, 可证明括号内期望非负;
- 因此 ES_α 满足次可加性, 是一致性风险度量。

ES 的估计与渐近性质

- 样本估计：

$$\widehat{\text{ES}}_{\alpha} = -\frac{\sum_{t=1}^T X_t \mathbf{1}_{\{X_t \leq \widehat{q}_{\alpha}\}}}{\sum_{t=1}^T \mathbf{1}_{\{X_t \leq \widehat{q}_{\alpha}\}}},$$

$$\widetilde{\text{ES}}_{\alpha} = -\frac{1}{\alpha} \sum_{t=1}^T X_t \mathbf{1}_{\{X_t \leq \widehat{q}_{\alpha}\}},$$

其中 \widehat{q}_{α} 为分位数/VaR 的估计。

- 两者在 $T \rightarrow \infty$ 下具有相合性与渐近正态性 (Scaillet 2004; Scaillet 2005)。
- 厚尾情形可结合极值理论与半参数尾模型改进 ES 估计。

条件 ES: GARCH 模型中的实现

- 假设条件异方差模型:

$$X_t = \mu_t + \sigma_t \varepsilon_t.$$

- 条件 ES:

$$\text{ES}_\alpha(X_t | \mathcal{F}_{t-1}) = \mu_t + \sigma_t \cdot \text{ES}_\alpha(\varepsilon).$$

- 实务步骤:

- 用 GARCH(1,1) (或更一般模型) 拟合 μ_t, σ_t ;
- 选择创新分布 (正态或 t), 计算

$$m_\alpha := E(\varepsilon | \varepsilon \leq z_\alpha),$$

从而 $\text{ES}_{t,\alpha} = -(\mu_t + \sigma_t m_\alpha)$;

- 若创新呈厚尾, 可对残差尾部再做 EVT/Hill 外推。

ES 的历史模拟与条件 GARCH：实证思路

- 标的： $\sim \text{GSPC}$ 日对数收益。
- 无条件（历史模拟）：
 - 直接用样本分布估计 q_α 与 ES_α ；
 - 构造滚动窗口的一步 VaR/ES 路径；
 - 观察极端损失的时间变化。
- 条件 GARCH：
 - 拟合 GARCH(1,1)-正态或 GARCH(1,1)- t ；
 - 利用闭式条件尾期望 m_α 得到 $ES_{t,\alpha}$ 路径；
 - 与历史模拟路径对比“放大/收缩”效应。
- 可进一步做 ES 回测（如 acerbi2014backtesting 的统计量）。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

黑天鹅：不可预见的极端事件

定义 (黑天鹅 (Taleb 2007))

指先验难以预见、影响巨大且事后被合理化的低概率事件。

要点：

- 超出常用模型与经验外推的“**认知支集**”（模型外部性）；
- 可计算性受限：样本与机制均不足以给出可靠先验；
- 人类事后倾向构造因果叙事（“看后视镜开车”）。

治理思路：

- 承认不可知，做**鲁棒化与反脆弱设计**：
 - 资本缓冲、限仓、熔断机制、尾部对冲等。

龙王：机制不同于常态的极端事件

定义 (龙王 (Sornette 2012))

指在统计上远超幂律尾部、且由不同于常态的内生机制（自激式正反馈、临界转变、网络连锁）所驱动的特异极端事件。

- 与黑天鹅不同：龙王可能可预警。
- 危机前常见信号：
 - 临界减速（恢复变慢、自相关上升）；
 - 方差/偏度上升；
 - 加速增长或对数周期震荡。
- 治理思路：
 - 监测早期预警信号；
 - 削弱正反馈与耦合（降杠杆、限集中度、隔离关键节点）。

灰犀牛：显而易见却被忽视的风险

定义 (灰犀牛 (Wucker 2016))

指大概率、影响大、肉眼可见却被系统性忽视或拖延的风险。

- 本质不是“不可知”，而是“看见但不行动”的治理失败：
 - 债务滚雪球；
 - 期限错配与流动性压力；
 - 气候与金融风险错配等。
- 治理思路：
 - 把显性风险纳入硬约束与考核；
 - 逆周期缓冲、压力测试情景、红线指标与问责机制；
 - 避免“温水煮青蛙”。

气候风险为何是“灰犀牛”？

- 气候相关风险符合灰犀牛三要素：
 - 概率不低；
 - 影响巨大；
 - 长期可见却常被拖延应对。
- 关注的是其金融与宏观传导：
 - **物理风险**：极端天气、海平面上升、慢性干旱、热浪等；
 - **转型风险**：低碳转型的政策、技术、偏好冲击。
- 机制：
 - 资产与产能（保险、地产、供应链、粮价等）的再定价；
 - 搁浅资产（高排放行业融资成本上升、碳价上行、合规成本上升）。
- 若缺乏前瞻管理：
 - 再保险保费跳升；
 - 抵押品折价与信用收缩；
 - 集中爆发时可能被放大为系统性问题。

黑天鹅、龙王与灰犀牛：对比

维度	黑天鹅	龙王	灰犀牛
先验可计算性	极弱 / 不可得	中等：依赖机制与预警信号	较强：风险公开且量化可得
可预警性	低	中高（临界征兆可监测）	高（但常被忽视 / 博弈掉）
驱动机制	模型外部性 / 信息缺失	非线性正反馈、临界转变	累积性、显性失衡
治理策略	鲁棒化、尾部对冲	监测信号、降耦合 / 降杠杆	硬约束、逆周期管理

2008 年危机：灰犀牛 + 龙王 + 黑天鹅

- “灰犀牛”阶段（危机前）：
 - 房地产泡沫、影子银行链条、高杠杆证券化、短融长投等结构性失衡长期累积；
 - 脆弱性可量化、可观测却被忽视。
- “龙王”阶段（临界爆发）：
 - 房价拐头、违约率上升；
 - 抵押品价值下跌、回购融资收缩；
 - “去杠杆—抛售—价跌—进一步去杠杆”的流动性螺旋。
- “黑天鹅”观感（危机中）：
 - 在脆弱系统中，局部违约、评级下调等被极度放大；
 - 若事后简单归为“不可预见”，会遮蔽此前多年治理空间。

A 股极端日：尾部事件并未消失

- 证综指 2007–2020 年间多次出现“极端日”：
 - 2007-02-27：收盘下跌 8.84%；¹
 - 2015-07-27：下跌 8.48%；²
 - 2015-08-24（“黑色星期一”）：下跌 8.49%；³
 - 2020-02-03（春节后复市首日）：下跌约 7.7%。⁴
- 2016–2019 年间单日大跌相对稀少，容易使人误判“风险已消退”。
- 实际上尾部事件只是呈现低频、间歇特征：
 - VaR/ES 估算应显式考虑厚尾与尾部依赖；
 - 结合压力情景，更加关注尾端风险而非平均波动。

¹<https://finance.sina.cn/sa/2007-02-27/detail-ikftssap0992591.d.html?from=wap>

²<https://finance.china.com/stock/pmkx/20150727/3254296.shtml>

³<https://www.cnfin.com/stock-xh08/a/20150824/1543213.shtml>

⁴<https://www.xinhuanet.com/fortune/caiyan/ksh/298.htm>

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

本章小结 (一): 尾部建模与极值理论

- 传统均值-方差与正态近似难以刻画:
 - 尖峰厚尾、波动聚集、时变依赖;
 - 持有期风险也并非总能用 \sqrt{T} 缩放。
- 以 VaR 作为共同语言, 串起三条主线:
 - ① 正态与稳定分布: 理解时间缩放与厚尾现象;
 - ② 非参数分位数: 稳健刻画中等尾部;
 - ③ 极值理论 + 半参数尾模型: 解决“样本之外”的极端分位外推。
- 尾部建模:
 - 块极大 (GEV) 与高阈超额 (POT/GPD) 两条 EVT 路线;
 - 正则变化尾、Hill/DEM 尾指数估计;
 - Hill 曲线、CCDF 双对数与 mean excess 图用于阈值选择与稳健性检查。

本章小结（二）：动态 VaR/ES 与系统性风险

- 时间维度：**条件风险度量**：
 - GARCH 模型刻画时变波动；
 - 结合经验或分布假设得到条件 VaR/ES；
 - CaViaR 模型将 VaR 本身作为状态变量直接建模。
- 多资产维度：
 - Copula 实现“边际—依赖”解耦，线性相关不足以刻画尾依赖；
 - CoVaR 与 Δ CoVaR 度量系统性风险溢出与“同跌”风险。
- 度量选型：
 - VaR 便于沟通但可能不满足次可加性；
 - ES 满足一致性公理，更适合监管与资本计提。

本章小结（三）：从度量到治理

- 度量需要配套：
 - 回测检验（UC、独立性、ES 回测等）；
 - 模型漂移监控；
 - 压力测试与情景分析。
- 实证脚手架：
 - 滚动比较：历史模拟 / 高斯 / GARCH- t / EVT-POT 的 VaR；
 - 半参数尾厚度诊断与极端 VaR 外推；
 - 条件 VaR/ES 与 CaViaR 的轨迹与回测；
 - Copula + CoVaR 评估系统性风险溢出。
- 黑天鹅—龙王—灰犀牛框架：
 - 承认不可知，提升鲁棒性；
 - 监测临界信号；
 - 对显性失衡设硬约束。
- 最终目标：构建“可检验、可实施、可沟通”的尾部风险管理闭环。

目录

1 贴现函数、现货利率与远期利率

- 收益率曲线三件套的实证演示 (R)

2 由息票债估计收益率曲线

- 基函数展开法
- 参数法: Nelson–Siegel / Svensson
- Fama-Bliss 法

3 离散时间的债券定价模型

- 利率的经济学假说
- 收益率的统计性质

4 无套利与随机贴现因子 (SDF)

- Vasicek 模型 (离散时间仿射形式)
- Cox–Ingersoll–Ross (CIR) 模型
- 多因子仿射期限结构

5 本章小结

6 参考文献

参考文献 I

-  Adrian, Tobias and Markus K Brunnermeier (2016). “CoVaR”. In: *American Economic Review* 106.7, pp. 1705–1741.
-  Artzner, Philippe et al. (1999). “Coherent Measures of Risk”. In: *Mathematical Finance* 9.3, pp. 203–228.
-  Basel Committee on Banking Supervision (2012). *Fundamental Review of the Trading Book*. Tech. rep. Bank for International Settlements.
-  Berman, Simeon M. (1964). “Limit Theorems for the Maximum Term in Stationary Sequences”. In: *Annals of Mathematical Statistics* 35.2, pp. 502–516.
-  Danielsson, Jon et al. (2013). “Fat Tails, VaR and Subadditivity”. In: *Journal of Econometrics* 172.2, pp. 283–291.
-  Dekkers, Arnold L. M., John H. J. Einmahl, and Laurens de Haan (1989). “A Moment Estimator for the Index of an Extreme-Value Distribution”. In: *The Annals of Statistics* 17.4, pp. 1833–1855.

参考文献 II

-  Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch (1997). *Modelling Extremal Events: for Insurance and Finance*. Springer.
-  Engle, Robert F. and Simone Manganelli (2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles". In: *Journal of Business & Economic Statistics* 22.4, pp. 367–381.
-  Fisher, R. A. and L. H. C. Tippett (1928). "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample". In: *Proceedings of the Cambridge Philosophical Society* 24, pp. 180–190.
-  Gabaix, Xavier and Rustam Ibragimov (2011). "Rank- 1/2: a simple way to improve the OLS estimation of tail exponents". In: *Journal of Business & Economic Statistics* 29.1, pp. 24–39.
-  Gnedenko, B. (1943). "Sur la distribution limite du terme maximum d'une serie aleatoire". In: *Annals of Mathematics* 44, pp. 423–453.

参考文献 III

-  Hill, J. B. (2010). "Robust Estimation of Tail Thickness Under Dependence". In: *Journal of Econometrics* 158.1, pp. 68–75.
-  Ibragimov, Rustam (2009). "Portfolio Diversification and Value at Risk Under Thick-Tailedness". In: *Operations Research Letters* 37.2, pp. 162–168.
-  Jorion, Philippe (1997). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill.
-  Koenker, Roger (2005). *Quantile Regression*. Cambridge University Press.
-  Mandelbrot, Benoit (1963). "The Variation of Certain Speculative Prices". In: *Journal of Business* 36.4, pp. 394–419.
-  Patton, Andrew J. (2006). "Modelling Asymmetric Exchange Rate Dependence". In: *International Economic Review* 47.2, pp. 527–556.

参考文献 IV

-  Resnick, Sidney I. and Cătălin Stărică (1998). "Tail Index Estimation for Dependent Data". In: *Annals of Applied Probability* 8.4, pp. 1156–1183.
-  Scaillet, Olivier (2004). "Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall". In: *Mathematical Finance* 14.1, pp. 115–129.
-  — (2005). "Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall". In: *Mathematical Finance* 15.3, pp. 283–306.
-  Sklar, Abe (1959). "Fonctions de répartition à n dimensions et leurs marges". In: *Publications de l'Institut de Statistique de l'Université de Paris* 8, pp. 229–231.
-  Sornette, Didier (2012). "Dragon-Kings: Mechanisms, Statistical Methods and Empirical Evidence". In: *The European Physical Journal Special Topics* 205.1, pp. 1–26. DOI: 10.1140/epjst/e2012-01559-5.

参考文献 V

-  Taleb, Nassim Nicholas (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House.
-  Turner, Adair (2009). *The Turner Review: A Regulatory Response to the Global Banking Crisis*. Tech. rep. UK Financial Services Authority.
-  Wucker, Michele (2016). *The Gray Rhino: How to Recognize and Act on the Obvious Dangers We Ignore*. St. Martin's Press.

第 12 章 AI 与金融计量：模型、应用与实践

孙佳婧 † 洪永淼 ‡ 奥利弗 · 林顿 §

中国科学院大学经济与管理学院 †
中国科学院数学与系统科学研究院 ‡
剑桥大学经济学系 §

2025 年 12 月 1 日

本作品采用 CC BY-NC 4.0 国际许可协议发布。
您可以在非商业用途下共享和改编本作品，但需注明作者。

从传统计量到 AI：三重张力

传统计量基石：

- OLS、VAR/ARMA、GARCH、Logit/Probit 等：
 - 结构清晰、检验体系成熟、推断逻辑完备；
 - 围绕参数显著性、区间估计与假设检验形成“可复核、可审计”的证据链。

现代金融数据的三类新张力：

- ① 非线性与高维交互普遍存在，线性 / 准线性关系难以完整刻画；
- ② 非平稳与制度切换（ZLB、疫情冲击、监管与微观结构变化）：
 - 参数不稳定，外推风险上升；
- ③ 数据模态多样化：
 - 文本、图像、逐笔成交、层级账务等，要求跨模态表征与融合能力。

AI 方法（树集成、核方法、深度学习）：提供更强的函数逼近与表征学习能力，可在交叉验证与正则化框架下重塑“预测—解释—合规”流程。

预测型学习 vs 结构 / 因果计量

预测型学习：广义风险最小化

$$R(f) = \mathbb{E}[\ell(Y, f(X))], \quad \hat{f} = \arg \min_{f \in \mathcal{F}} \hat{R}(f),$$

- 强调在给定当前分布下的泛化误差；
- 目标：择时、排序、违约检测等。

结构性 / 因果性计量：

$$Y = g(X; \theta) + \varepsilon, \quad \mathbb{E}[\varepsilon | X] = 0,$$

- 关注数据生成机制与参数 θ 的可解释推断；
- 用于政策评估、规制口径、结构敏感度分析。

AI 的角色：

- 扩展函数类 \mathcal{F} :
 - 预测优先：非线性模型在“弱信号—高噪声—高维协变量”场景下表现更好；
 - 解释 / 因果优先：应嵌入正交化 / 双重稳健框架（如 Double Machine Learning），保持推断有效性与因果可读性。

非平稳环境与时序交叉验证

金融系统很少 i.i.d.:

$$p_t(X) \neq p_s(X) \quad (\text{协变量漂移}), \quad p_t(Y|X) \neq p_s(Y|X) \quad (\text{概念漂移}) .$$

- 对模型的影响:

- 协变量漂移 \Rightarrow 外推偏差;
- 概念漂移 \Rightarrow 策略失效。

本书的底层约束：时间序列交叉验证

- 滚动 / 扩展窗口;

- purged / embargo K-fold:

- 删除与验证标签区间重叠的训练样本;
- 在边界设置“禁运区”(embargo);

- 嵌套交叉验证:

- 外层一次性报告样本外性能;
- 内层调参，将“研究者自由度”与评估严格分离。

治理与合规：可解释—可追溯—可监控

黑箱性并非不可化解：

- 全局 / 局部 SHAP、置换重要度；
- PDP / ALE 与 ICE：回答
 - 哪些特征重要？
 - 在哪些区间敏感？
 - 模型是否稳定？

概率输出与校准：

- Platt 缩放 / 保序回归；
- Brier / ECE 指标；
- 与阈值 / 成本对齐。

漂移与公平性监控：

- 漂移：PSI、JS 散度、基于分类器的两样本检验；
- 公平性：机会均等（TPR 一致）、均衡误差（TPR/FPR 同时一致）、人口比例等；
- 通过阈值分组 / 后处理进行调整。

从算法正确性到业务可用性

制度配套：

- 文档化：
 - 数据血缘、特征字典；
 - 切分与随机种子；
 - 阈值与版本管理。
- 三道防线：
 - 开发（前台 / 模型团队）；
 - 独立验证（模型风险）；
 - 内部审计。
- 人工干预停机开关 / 兜底机制：将“算法正确性”上升为“业务可用性”的前提。

从算法正确性到业务可用性

本章定位：

- 不把 AI 视为传统计量的替代品；
- 而是**补全者与增强器**：
 - 树模型（随机森林 / XGBoost）；
 - 深度神经网络（MLP / LSTM / Transformer）；
 - 超参调优、时序 CV、早停与正则化；
 - 概率校准、阈值—成本对齐与风险—收益评估；
 - SHAP / PDP / 稳定性监控、现实检验 / SPA 与 FDR 控制；
 - VaR / ES 与情景分析下的因子抽取、暴露估计与不确定性量化。

目录

① 树模型（随机森林与 XGBoost）的原理及适用性

② 神经网络与深度学习方法概述及金融场景应用

③ 超参数调优与交叉验证：金融计量视角

④ 案例：AI 模型的典型应用场景

⑤ 模型可解释性与金融稳定性

⑥ 本章小结

决策树模型：基本思想

决策树：

- 通过递归二分划分特征空间，学习一组“如果……则……”的规则；
- 在某个特征的某一阈值上划分，使划分后子集在目标变量上的“纯度”更高。

不纯度度量：

- 分类：
 - 熵 (Entropy)；
 - 基尼指数 (Gini Index)。
- 回归：
 - 残差平方和 (SSE)；
 - 均方误差 (MSE) / 方差。

训练时在候选特征与阈值上贪心地最大化“纯度提升”（不纯度下降）。

分类树：基尼指数与熵

定义（基尼指数）

样本集合 D 有 K 个类别，第 k 类比例为 p_k ：

$$\text{Gini}(D) = 1 - \sum_{k=1}^K p_k^2 = \sum_{k=1}^K p_k(1 - p_k).$$

分类树：基尼指数与熵

定义 (熵与信息增益)

$$\text{Entropy}(D) = - \sum_{k=1}^K p_k \log_2 p_k.$$

按特征 A 划分为若干子集 D_i 时，

$$\text{Gain}(D, A) = \text{Entropy}(D) - \sum_i \frac{|D_i|}{|D|} \text{Entropy}(D_i).$$

训练时选择使 $\text{Gini}_A(D)$ 最小或 $\text{Gain}(D, A)$ 最大的特征 / 阈值。

信息增益率与多值特征

定义 (信息增益率)

$$\text{GainRatio}(D, A) = \frac{\text{Gain}(D, A)}{\text{IV}(A)}, \quad \text{IV}(A) = -\sum_{i=1}^n \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|}.$$

- 纯信息增益对“取值很多”的特征有偏好；
- 信息增益率用 $\text{IV}(A)$ 惩罚这种偏好。

回归树：SSE 与方差减小

回归情形：

- 按特征 A 在阈值 a 分为 D_1, D_2 :

$$\text{SSE}(A, a) = \sum_{x_i \in D_1} (y_i - \bar{y}_1)^2 + \sum_{x_i \in D_2} (y_i - \bar{y}_2)^2.$$

- 最小化 $\text{SSE}(A, a)$ 等价于最大化方差减小。

统一不纯度视角：

$$\Delta I(A, a | D) = I(D) - \frac{|D_1|}{|D|} I(D_1) - \frac{|D_2|}{|D|} I(D_2),$$

- 分类: $I = \text{Gini}$ 或 Entropy ;
- 回归: $I = \text{Var}$ 或归一化的 SSE;
- 划分时选择使 ΔI 最大的特征与阈值。

决策树的预测与规则形式

叶节点预测：

- 分类：输出多数类或类别概率；
- 回归：输出叶内 y 的均值（或中位数，提高鲁棒性）。

规则形式示例：

- 分类：若 $A \leq a$ 且 $B > b$ ，则预测类别 C ；
- 回归：若 $A(x) \leq a, B(x) > b$ ，则

$$\hat{y}(x) = \frac{1}{|D'|} \sum_{x_i \in D'} y_i, \quad D' = \{x_i : A(x_i) \leq a, B(x_i) > b\}.$$

决策树以一组清晰的“可解释规则”刻画非线性与交互效应。

预剪枝：常见超参数的作用

定义（预剪枝超参数（示意））

- max_depth: 最大树深度，控制“根到叶”的路径长度；
- min_samples_split: 继续分裂所需的最小样本数；
- min_samples_leaf: 叶节点最小样本数，保证每个叶子“足够大”；
- max_leaf_nodes: 直接限制叶子总数（决策树的分段数）；
- min_impurity_decrease: 最小不纯度下降阈值：

$$\Delta I(A, a | D) \geq \delta \Rightarrow .$$

调参思路：

- 先用 max_depth/max_leaf_nodes/min_samples_leaf 把握容量；
- 再用 min_impurity_decrease 去除“收益极小”的分裂。

后剪枝：代价复杂度剪枝

定义 (代价复杂度剪枝 (Cost-Complexity Pruning))

已生长的大树 T , 叶数为 $|T|$ 。回归情形经验风险

$$R(T) = \sum_{t \in \text{leaves}(T)} \sum_{x_i \in t} (y_i - \bar{y}_t)^2,$$

正则化目标:

$$R_\alpha(T) = R(T) + \alpha |T|, \quad \alpha \geq 0.$$

- 随 α 增大得到一条嵌套的最优子树序列;
- 可通过交叉验证选择 α , 或采用“一倍标准误差原则”选更简单的树;
- 实现中常以 `ccp_alpha` 暴露。

金融数据中的建模要点（决策树）

- **时间序列评估与防泄露：**

- 滚动 / 扩展窗口，训练集时间早于验证 / 测试；
- 第 m 折：训练集 $\{1, \dots, t_m\}$ ，验证集 $\{t_m + 1, \dots, t_{m+1}\}$ 。

- **指标体系：**

- 不仅是 ACC/MSE，还包括年化收益率、波动率、最大回撤、夏普 / 索提诺；
- 纳入交易成本与滑点后的净表现。

- **类别不平衡 / 稀有事件：**

- 加权损失、阈值优化；
- 指标侧重 PR-AUC、F1、召回率。

金融数据中的建模要点（决策树）

- 概率校准与解释：

- Platt / Isotonic 校准；
- TreeSHAP、置换重要度；
- PDP / ICE 观察局部形状。

- 非平稳与漂移：

- 监控 PSI 等漂移，设置再训练节奏；
- 需要时加入单调约束以符合经济常识。

与集成方法的衔接

单棵树方差较高，实务中多配合集成：

- 随机森林（装袋）：
 - 样本 + 特征双重子采样；
 - 多棵树平均 / 投票，降低方差；
 - 关键参数：树数、深度、特征子采样比例、样本子采样比例。
- 梯度提升树（GBDT / XGBoost / LightGBM）：
 - 以浅树为弱学习器，逐步拟合残差；
 - 偏差低、精度高；
 - 关键参数：学习率、弱学习器数量、深度、叶最小样本数、行 / 列采样率。

决策树 + 集成 = 非线性、高维交互 + 可解释结构，在金融任务中非常实用。

随机森林：构建机制与集成原理

训练阶段的“双重随机化”：

- 样本随机化：

$$D_b = \{(x_i, y_i)\}_{i=1}^n, \quad x_i \sim D \text{ (bootstrap 有放回) ;}$$

- 特征随机化：

- 每个节点从全部 p 个特征中随机抽取 m 个 (`mtry`)；
- 在这 m 个候选中选择最大不纯度下降的划分。

随机森林：构建机制与集成原理

集成预测：

$$\hat{f}_{\text{RF}}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (\text{回归}),$$

$$\hat{y}(x) = \text{mode}\{T_b(x)\}_{b=1}^B \quad (\text{分类}).$$

方差降低机制：

$$\text{Var}\left(\frac{1}{B} \sum_b T_b(x)\right) = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2,$$

通过降低树间相关 ρ + 增加树数 B , 显著降低方差。

袋外误差 (OOB) 与时间序列评估

OOB 思想：

- 每棵树中，某样本“不被抽中”的概率 $(1 - \frac{1}{n})^n \approx e^{-1} \approx 0.368$ ；
- 对每个样本 x_i ，用未包含它的树集合 $\mathcal{B}^{(-i)}$ 做袋外预测：

$$\hat{\gamma}^{(-i)}(x_i) = \frac{1}{|\mathcal{B}^{(-i)}|} \sum_{b \in \mathcal{B}^{(-i)}} T_b(x_i).$$

- 得到近似样本外误差：

$$\hat{\mathcal{E}}_{\text{OOB}} = \frac{1}{n} \sum_i \ell(y_i, \hat{\gamma}^{(-i)}(x_i)).$$

时间序列注意事项：

- 真正严格的评估仍应使用滚动 / 扩展窗口和 out-of-time 留出期；
- 可结合 OOB 作为“训练期内”的快速诊断。

特征重要性与可解释性（随机森林）

基于不纯度下降的重要性：

$$\text{VI}_j^{\text{imp}} = \sum_{t \in T_j} \Delta I_t,$$

- 直观，但对高基数特征有偏好。

置换重要度（更稳健）：

$$\Delta \mathcal{E}(A) = \mathcal{E}(\hat{f}, \pi_A(X); Y) - \mathcal{E}(\hat{f}, X; Y),$$

- π_A 仅对特征 A 进行随机打乱；
- 性能退化越大，特征越重要。

PDP 与 ICE：

$$\text{PD}_j(z) = \frac{1}{n} \sum_i \hat{f}(z, x_{i,-j}),$$

搭配 ICE 曲线可观察个体层面的响应差异。

随机森林在金融中的优势与局限

优势：

- 在高维、噪声大、变量相关性强的情形下稳健；
- 对无效变量有“隐式筛选”效果；
- OOB 评估方便调参与监控。

局限：

- 集成结构使“整体经济含义”的直接归纳不总是直观；
- 近似线性场景中，可能不如线性 / 稀疏模型经济；
- 外推能力有限（本质是分段常数），离训练支持集很远时需谨慎；
- 概率输出用于阈值 / 成本决策前，应先做校准。

随机森林在金融中的优势与局限

策略示例（简化）：

$$p_t = \Pr(\text{Up}_{t+1} \mid x_t), \quad w_t = \mathbf{1}\{p_t > \tau\},$$

$$r_{t+1}^\pi = w_t r_{t+1} - c |w_t - w_{t-1}|,$$

年化夏普：

$$\text{Sharpe} = \frac{\mathbb{E}[r_t^\pi - r_t^f]}{\sqrt{\text{Var}(r_t^\pi - r_t^f)}} \sqrt{12}.$$

梯度提升：可加性模型视角

GBDT 的核心：

- 以浅树为弱学习器，逐轮逼近目标函数；
- 模型为可加形式：

$$F_m(x) = F_{m-1}(x) + \eta \gamma_m h_m(x).$$

负梯度拟合：

$$r_{im} = - \left. \frac{\partial \ell(y_i, F(x_i))}{\partial F(x_i)} \right|_{F=F_{m-1}},$$

拟合 $h_m(x) \approx r_{im}$, 再做线搜索得到 γ_m 。例：二分类对数损失

$$\ell(y, F) = -[y \log p + (1 - y) \log(1 - p)], \quad p = \sigma(F),$$

$$r_{im} = y_i - p_{m-1}(x_i),$$

即每轮拟合“残差概率”。

XGBoost：目标函数与正则化

二阶近似 + 结构正则：

$$\min_{f_m} \sum_i \left[g_i f_m(x_i) + \frac{1}{2} h_i f_m(x_i)^2 \right] + \Omega(f_m),$$

$$\Omega(f) = \gamma T + \frac{\lambda}{2} \sum_{j=1}^T w_j^2,$$

其中 T 为叶数， w_j 为叶权重。聚合量：

$$G_j = \sum_{x_i \in R_j} g_i, \quad H_j = \sum_{x_i \in R_j} h_i,$$

对应最优叶权重与目标值：

$$w_j^* = -\frac{G_j}{H_j + \lambda}, \quad \tilde{\mathcal{L}}(f_m) = -\frac{1}{2} \sum_j \frac{G_j^2}{H_j + \lambda} + \gamma T.$$

XGBoost 的分裂增益 (Gain)

二分划分的增益：

$$\text{Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma.$$

- G_L, H_L : 左子节点的梯度 / 海森和；
- G_R, H_R : 右子节点的梯度 / 海森和；
- 仅当 $\text{Gain} > 0$ 时才值得分裂。

含义：

- 将“二阶信息 + L2 正则 + 结构惩罚”统一入分裂准则；
- 自动平衡“拟合精度”和“模型复杂度”。

可定制损失函数与业务约束

分位数损失：拟合 VaR / ES

$$L_\tau(y, \hat{y}) = (\tau - \mathbf{1}_{\{y < \hat{y}\}}) (y - \hat{y}).$$

排序损失：量化选股

$$\sum_{(i,j):y_i>y_j} \log \left(1 + \exp(-(F(x_i) - F(x_j))) \right).$$

类不平衡：加权对数损失

$$\ell_\alpha(y, p) = -\alpha y \log p - (1 - \alpha)(1 - y) \log(1 - p).$$

约束与校准：

- **单调性约束**：如利率升高 \Rightarrow 违约概率不降；
- **概率校准**：Platt / Isotonic；
- **PSI 等指标监控漂移**，触发再训练或阈值重定标。

随机森林 vs XGBoost：对比与选型

随机森林：

- 并行装袋，深树 + 投票 / 平均，降低方差；
- 调参成本低，OOB 可直接评估样本外；
- 适合作为稳健、快速的基线模型。

XGBoost：

- 串行梯度提升，浅树 + 二阶信息，降低偏差；
- 精度高，可定制损失函数与约束；
- 对学习率、正则与采样更敏感，需要早停与验证集。

共同点：

- 可配合置换重要度、SHAP、PDP 等提升可解释性；
- 在时间序列框架中，需用滚动 / 扩展窗口 + out-of-time 测试。

与金融计量的衔接与注意事项

评估中心：样本外表现

- 遵循时间顺序，滚动/扩展窗口 + 留出期；
- 报告任务相关的收益—风险指标（含交易成本）。

漂移监控：人口稳定性指数（PSI）

$$\text{PSI} = \sum_b (p_b - q_b) \ln \frac{p_b}{q_b},$$

p_b, q_b : 基准期与当前期在第 b 箱中的占比。推断 / 因果问题：

- 在样本分割与正交化框架下，嵌入 Double ML / Causal Forest；
- 报告边际效应及置信区间。

概率质量 / 阈值—成本对齐：

$$\text{BS} = \frac{1}{n} \sum_i (y_i - p_i)^2 \quad (\text{Brier 分数}) ,$$

最优阈值基准：

$$\tau^* = \frac{c_{\text{FP}}}{c_{\text{FP}} + c_{\text{FN}}}.$$

案例 A：随机森林 vs 逻辑回归（滚动样本外）

任务：

- 单资产（如 SPY/^GSPC）月频数据；
- 预测下月方向（Up/Down）。

数据与特征：

- 月度对数收益 $r_t = \Delta \log P_t$ ；
- 特征：6/12 个月动量、3 个月波动率、10 月均线偏离、RSI、MACD 直方等；
- 所有特征统一滞后一月，用 t 的特征预测 $t+1$ 。

评估框架：

- 固定区间（如 2010–2024），训练窗自适应（最多 120 个月）；
- 滚动：用过去 W 个月训练，预测下一个月单点；
- 比较逻辑回归 vs 随机森林的：
 - AUC、准确率；
 - 简单阈值择时 ($p > 0.5$ 满仓，含交易成本)；
 - 年化收益 / 波动 / 夏普，与买入并持有对比。

案例 A：要点总结

- 同一时序框架下，对比“线性 + 稀疏特征”与“非线性 + 高维交互”的收益；
- 显式计入交易成本，防止“纸面优势”；
- 用净值曲线和年化统计衡量择时信号的经济价值；
- 程序中自适应训练窗长度，支持短样本情形；
- 可在此框架上加入概率校准、连续仓位映射及阈值样本外优化。

案例 B: XGBoost (滚动样本外 + 早停)

任务与数据:

- 同样使用月度 SPY/GSPC，预测下月方向；
- 特征构造与滞后方式与随机森林案例相同。

训练与早停:

- 训练窗 W 自适应，末尾取约 20% 作为验证窗；
- 目标: binary:logistic, eval_metric = AUC；
- 基础超参: 浅树 ($\text{max_depth}=3$)、小学习率、行 / 列子采样；
- 按负/正样本比设置 scale_pos_weight；
- 用验证集 AUC 早停，避免过拟合。

案例 B：评估与特征重要性

评估指标：

- 样本外 AUC 与基于阈值 0.5 的 ACC;
- 简单阈值择时（含交易成本）：

$$r_{t+1}^{\pi} = w_t r_{t+1} - c |w_t - w_{t-1}|;$$

- 与买入并持有对比年化收益 / 波动 / 夏普；
- 绘制两条净值曲线直观比较。

特征重要性：

- 使用 `xgb.importance` 计算 “Gain” 重要度；
- 列出前若干特征及其贡献，用条形图展示；
- 提示 “Gain 是贡献大小，不是因果方向”，需配合 SHAP/置换重要度。

整体框架可以直接迁移到其他标的与特征组合，构成可复现、可解释、贴近实务的样本外实验平台。

小结：树模型在金融计量中的角色

- 决策树：统一的“不纯度下降”视角，规则清晰、可解释；
- 随机森林：稳健、易用的非线性基线，OOB 评估便于快速迭代；
- XGBoost：高精度、可定制目标与约束，配合早停与正则化；
- 在“固定区间 + 滚动样本外 + 严格防泄露”的框架中：
 - 将“强表达力的学习器”嵌入
 - “因果正确的评估—合规完整的治理”闭环；
- 在预测—解释—合规三角中寻找可复用的均衡点，是本章方法论的主旨。

目录

- ① 树模型（随机森林与 XGBoost）的原理及适用性
- ② 神经网络与深度学习方法概述及金融场景应用
- ③ 超参数调优与交叉验证：金融计量视角
- ④ 案例：AI 模型的典型应用场景
- ⑤ 模型可解释性与金融稳定性
- ⑥ 本章小结

背景与动机

人工神经网络 (ANN) :

- 通过堆叠线性变换与非线性激活逼近复杂函数；
- 不预设函数形式，以数据驱动自动学习非线性与高阶交互；
- 相比线性 / 广义线性模型，更适合**结构性非线性与异质性**场景。

典型金融场景：

- 横截面选股、文本情绪、盘口微结构；
- 高维宏观合成指标等。

深度学习 (多层网络) :

- 依靠多级非线性映射构建**表示**；
- 在金融数据中需格外重视：
 - 模型选择与正则化；
 - **严格的样本外评估**（高噪声、弱信号、非平稳）。

多层感知机（MLP）与前向传播

单隐藏层 MLP，输入 $\mathbf{x} \in \mathbb{R}^p$ 、隐藏层宽度 H :

$$\mathbf{a}^{(1)} = \sigma(\mathcal{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) \in \mathbb{R}^H, \quad (1)$$

$$\hat{y} = \sigma_{\text{out}}(\mathcal{W}^{(2)}\mathbf{a}^{(1)} + \mathbf{b}^{(2)}), \quad (2)$$

- $\mathcal{W}^{(1)} \in \mathbb{R}^{H \times p}$, $\mathcal{W}^{(2)} \in \mathbb{R}^{1 \times H}$ (或多维输出);
- 激活 σ : ReLU / tanh 等;
- 输出层映射 σ_{out} :
 - 恒等: 回归;
 - sigmoid: 二分类;
 - softmax: 多分类, $p_k = \exp z_k / \sum_j \exp z_j$ 。

经验风险最小化与典型损失

经验风险最小化 + 正则:

$$\min_{\theta} \mathcal{J}(\theta) = \frac{1}{n} \sum_{i=1}^n \ell(y_i, f_{\theta}(\mathbf{x}_i)) + \lambda \Omega(\theta).$$

常见损失:

- 均方误差:

$$\ell_{\text{MSE}}(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2;$$

- 二分类交叉熵 (最大似然):

$$\ell_{\text{CE}}(y, p) = -[y \log p + (1 - y) \log(1 - p)], \quad p = \sigma(z);$$

- 分位数损失 (VaR 等):

$$\ell_{\tau}(y, \hat{q}) = \left(\tau - \mathbf{1}\{y < \hat{q}\} \right) (y - \hat{q}), \quad \tau \in (0, 1).$$

参数用随机梯度类方法更新:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} \mathcal{J}(\theta),$$

梯度由反向传播计算; 初始化常用 Xavier / He, 激活偏向 ReLU / GELU, 以缓解梯度消失 / 爆炸。

Adam / AdamW 与梯度裁剪

Adam: 记小批次梯度 $g_t = \nabla_{\theta} \mathcal{J}(\theta_t)$,

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2.$$

偏差校正:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}.$$

AdamW 更新:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} - \eta \lambda_{\text{wd}} \theta_t.$$

梯度裁剪:

$$g_t \leftarrow g_t \cdot \min \left(1, \frac{\tau}{\|g_t\|_2} \right),$$

- 控制极端梯度，稳定训练；
- 尤其在深层 RNN / Transformer 中重要。

学习率调度与早停

学习率调度：

- 余弦退火 (cosine annealing)；
- 分段下降 (step decay)；
- warmup + 退火等。

早停 (Early Stopping)：

- 以验证集损失 / 指标为准；
- 若若干轮内无改善，则停止训练；
- 有效防止过拟合，尤其在样本有限的金融任务中。

正则化：权重衰减与 Dropout

过拟合是金融应用的首要风险。 L_2 权重衰减：

$$\Omega(\theta) = \frac{1}{2} \|\theta\|_2^2,$$

- 限制参数规模，提高泛化能力。

Dropout：

- 训练期以概率 ρ 随机屏蔽神经元；
- 测试期按保留率缩放；
- 近似集成大量子网络，降低 co-adaptation。

BatchNorm 与时间序列中的“防泄露”

批归一化 (BatchNorm) :

$$\text{BN}(h) = \gamma \cdot \frac{h - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} + \beta.$$

- 稳定中间分布，加速收敛；
- 可视为带噪正则化。

时间序列中的归一化：

$$\tilde{x}_{t,j} = \frac{x_{t,j} - \mu_j^{(\mathcal{T})}}{\sigma_j^{(\mathcal{T})}}, \quad t > \max(\mathcal{T}).$$

- $\mu^{(\mathcal{T})}, \sigma^{(\mathcal{T})}$ 只在训练窗口拟合；
- 验证 / 测试集只能使用训练期参数进行外推；
- PCA / AE 等降维同理——全部在训练窗拟合以避免信息泄露。

LSTM：门控机制

LSTM 递归单元：

$$\mathbf{i}_t = \sigma(W_i \mathbf{x}_t + U_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad \mathbf{f}_t = \sigma(W_f \mathbf{x}_t + U_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (3)$$

$$\mathbf{o}_t = \sigma(W_o \mathbf{x}_t + U_o \mathbf{h}_{t-1} + \mathbf{b}_o), \quad \tilde{\mathbf{c}}_t = \tanh(W_c \mathbf{x}_t + U_c \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (4)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \quad \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \quad (5)$$

金融序列特点：

- 非平稳、突变多（政策、危机、制度切换）；
- 建议采用窗口化训练与强正则化；
- 样本有限时，浅层 RNN / 一维卷积往往更稳健于深堆叠 LSTM。

自注意力机制与因果约束

单头自注意力：

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK'}{\sqrt{d_k}}\right)V.$$

- 通过相似度加权聚合信息；
- 可并行建模长依赖（相比 RNN）。

计算成本与改进：

- 标准注意力复杂度 $O(T^2)$ ；
- 高频 / 长文本：稀疏注意力、窗口化注意力、低秩近似等。

时间因果约束：

- 利用上三角 mask，禁止访问未来时刻；
- 确保预测仅依赖过去信息，避免前视偏差。

AE / VAE / GAN: 表示与生成

自动编码器 (AE):

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^n \ell(x_i, \psi_{\theta}(\phi_{\theta}(x_i))) + \lambda \Omega(\theta).$$

- 学习压缩表示，用于异常检测、特征压缩。

变分自编码器 (VAE):

$$\log p_{\theta}(x) \geq \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x|z)] - D_{\text{KL}}(q_{\phi}(z|x) \| p(z)).$$

- 显式概率建模，支持生成 / 插值。

生成对抗网络 (GAN):

- 通过极小极大博弈拟合数据分布；
- 可用于极端场景生成与压力测试；
- 但训练稳定性与评估标准需格外谨慎。

概率预测与校准指标

概率而不仅是标签：

- 二分类交叉熵是**合适的评分规则**；
- 实务需要**概率可靠性**，不仅仅是排序。

概率校准：

- Platt 缩放（逻辑回归校准）；
- 保序回归（Isotonic Regression）。

质量指标：

$$\text{Brier} = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2,$$

$$\text{ECE} = \sum_{b=1}^B \frac{|B_b|}{n} |\text{acc}(B_b) - \text{conf}(B_b)|,$$

B_b : 按预测置信度划分的分箱。

风险度量与策略收益

分位数网络：

- 直接最小化 ℓ_τ 拟合 VaR；
- 可扩展到多分位、多期限。

阈值决策：

$$w_t = \mathbf{1}_{\{p_t > \tau\}},$$

含交易成本 c 的策略收益：

$$r_{t+1}^\pi = w_t r_{t+1} - c |w_t - w_{t-1}|.$$

年化夏普比率：

$$\text{Sharpe} = \frac{\text{E}[r_t^\pi - r_t^f]}{\sqrt{\text{Var}(r_t^\pi - r_t^f)}} \sqrt{T},$$

- 作为经济性可用性的核心指标；
- T : 一年内期数（如月度 $T = 12$, 周度 $T = 52$ ）。

典型金融任务与建模要点

单资产择时：

- 输出 $p_t = \Pr(r_{t+1} > 0 | x_t)$ 或直接预测 \hat{r}_{t+1} ；
- 将概率映射为仓位 / 杠杆，并计入成本。

横截面选股：

- 输入“股票—期限”特征至共享网络；
- 输出横截面打分，用排序 / 配对损失稳定名次；
- 典型目标：IC、Rank IC、分层收益。

信用 / 欺诈检测：

- 极端不平衡：需加权损失、采样策略；
- 阈值—成本对齐：误报 / 漏报成本差异巨大。

文本情绪与多模态场景

文本情绪分析：

- 采用预训练语言模型（BERT 等）提取向量 / 情感特征；
- 与结构化因子拼接后输入下游网络；
- 任务：公告情绪、新闻冲击、舆情驱动的 Alpha 等。

多模态融合：

- 文本 + 价格 / 盘口 + 宏观因子；
- 统一在深度模型中进行表示学习与融合；
- 仍需严格时序化处理和样本外评估。

时序交叉验证与漂移监控

时序 CV:

- 折点: $1 < t_1 < \dots < t_M < T$;
- 第 m 折训练集: $\{1, \dots, t_m\}$;
- 验证 / 测试: $\{t_m+1, \dots, t_{m+1}\}$;
- 所有标准化 / PCA / AE 变换: 在训练窗口拟合后外推。

分布漂移 (PSI):

$$\text{PSI} = \sum_b (p_b - q_b) \ln \frac{p_b}{q_b},$$

- 当 PSI 超过阈值时触发再训练或规则回退;
- 非平稳环境下的持续监控尤为重要。

统计显著性:

- 可采用区块自举评估夏普比率差异;
- 兼顾时间依赖和有限样本问题。

积分梯度、SHAP 与经济解释

积分梯度 (IG) :

$$IG_j(x) = (x_j - \tilde{x}_j) \int_0^1 \frac{\partial f(\tilde{\mathbf{x}} + \alpha(\mathbf{x} - \tilde{\mathbf{x}}))}{\partial x_j} d\alpha.$$

- 量化输入维度 j 对预测的边际贡献；
- 基于路径积分，相比单点梯度更稳健。

SHAP / PDP / ICE:

- SHAP：全局 / 局部解释，近似博弈论分摊；
- PDP：平均效应曲线；
- ICE：个体响应曲线，揭示异质性。

目标：

- 不是替代经济学推理；
- 而是在样本外稳定性前提下检验模型是否捕捉合理经济机制。

模型治理与合规要求

可复现性：

- 数据血缘、特征字典；
- 训练 / 验证 / 测试划分；
- 超参数与随机种子；
- 版本控制与文档化。

评估与报告：

- 除统计指标外，报告收益-风险、回撤；
- 换手率与交易成本；
- 违约 / 欺诈等任务中的公平性与可解释性。

生产监控：

- 数据质量、漂移、性能退化监控；
- 应急回退与再训练流程；
- 明确责任边界与人工干预机制。

小结：何时使用深度学习？

神经网络的工具箱：

- MLP：结构化高维数据的非线性整合；
- LSTM / GRU：时间依赖；
- Transformer：长序列与文本；
- AE / VAE / GAN：表示学习与场景生成。

关键：将其纳入金融计量框架

- 建模、正则与样本外评估闭环；
- 以经济目标与解释性作为硬约束；
- 高频、文本、图结构风险传播等“数据丰富、非线性强、结构复杂”领域，深度学习优势明显；
- 在信号稀薄、样本有限的传统任务中：
 - 线性或树模型仍是坚实基线；
 - 深度方法应在严格治理与清晰经济假说约束下使用。

目录

- 1 树模型（随机森林与 XGBoost）的原理及适用性
- 2 神经网络与深度学习方法概述及金融场景应用
- 3 超参数调优与交叉验证：金融计量视角
- 4 案例：AI 模型的典型应用场景
- 5 模型可解释性与金融稳定性
- 6 本章小结

动机与框架

核心问题：

- 复杂模型性能对超参数高度敏感；
- 金融数据具有时间依赖、弱信号、微观结构噪声；
- 还有交易成本、监管约束、风险约束等现实限制。

目标：将调参过程置于统一框架：

- 样本外评估为中心，而非样本内拟合；
- 以经济目标（收益、风险、成本）为指标；
- 严格防止信息泄露和回测过拟合。

风险函数与经验风险

设训练样本 $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$, 学习器由训练算子 \mathcal{A}_λ 与超参数 $\lambda \in \Lambda \subset \mathbb{R}^d$ 决定:

$$f_\lambda = \mathcal{A}_\lambda(\mathcal{D}).$$

损失函数为 $\ell(\cdot, \cdot)$, 则

- 广义风险:

$$R(\lambda) = \mathbb{E}_{(X, Y)} [\ell(Y, f_\lambda(X))];$$

- 经验风险:

$$\hat{R}(\lambda; \mathcal{D}) = \frac{1}{n} \sum_{i=1}^n \ell(y_i, f_\lambda(x_i)).$$

目标是找到在未来数据上风险较小的超参数, 而非仅最小化样本内 \hat{R} 。

K 折交叉验证与模型选择

K 折交叉验证：

- 将索引集划分为互不相交的验证折 V_k 及训练折 T_k , $k = 1, \dots, K$;
- 在每个折上训练 $f_\lambda^{(-k)} = \mathcal{A}_\lambda(\mathcal{D}_{T_k})$;
- CV 风险估计：

$$\hat{R}_{\text{CV}}(\lambda) = \frac{1}{K} \sum_{k=1}^K \frac{1}{|V_k|} \sum_{i \in V_k} \ell(y_i, f_\lambda^{(-k)}(x_i)).$$

模型选择与性能报告：

- 选型： $\hat{\lambda} = \arg \min_{\lambda \in \Lambda} \hat{R}_{\text{CV}}(\lambda)$;
- 性能报告应在独立测试集或外层交叉验证上完成。

嵌套交叉验证 (Nested CV)

问题：在同一验证集上反复调参会导致验证集过拟合。嵌套 CV 思路：

- 外层 K_{out} 折：只用于评估；
- 内层 CV：在外层训练集 $\mathcal{D}_{T_k^{\text{out}}}$ 上调参。

$$\hat{R}_{\text{nested}} = \frac{1}{K_{\text{out}}} \sum_{k=1}^{K_{\text{out}}} \frac{1}{|V_k^{\text{out}}|} \sum_{i \in V_k^{\text{out}}} \ell(y_i, f_{\hat{\lambda}_k}(x_i)),$$

其中

$$\hat{\lambda}_k = \arg \min_{\lambda} \hat{R}_{\text{CV,in}}(\lambda; \mathcal{D}_{T_k^{\text{out}}}).$$

要点：外层不再做任何选择，只打一分。

时间序列交叉验证：滚动 / 扩展窗口

金融时间序列：必须保持因果顺序——训练来自过去，验证来自未来。令折点
 $1 < t_1 < \dots < t_M < T$ ，第 m 折：

$$\mathcal{D}_m^{\text{train}} = \{1, \dots, t_m\}, \quad \mathcal{D}_m^{\text{val}} = \{t_m+1, \dots, t_{m+1}\}.$$

- **扩展窗**：训练集逐步累积历史数据；
- **滚动窗**：每次只用长度为 W 的最近历史。

评估指标通常取各折的**加权平均**（按时间长度或样本数加权）。

Purged / Embargo K-fold：防止标签重叠泄露

若标签由未来窗口构造（如 $r_{t:t+h}$ ），训练样本可能与验证标签窗口重叠。
Purged / Embargo K-fold：

$$\mathcal{D}_m^{\text{train}} \leftarrow \mathcal{D}_m^{\text{train}} \setminus \mathcal{O}(\mathcal{D}_m^{\text{val}}).$$

$$\mathcal{D}_m^{\text{train}} \leftarrow \mathcal{D}_m^{\text{train}} \setminus \text{Embargo}_e(\mathcal{D}_m^{\text{val}}).$$

- $\mathcal{O}(\cdot)$: 与验证标签窗口重叠的训练样本索引；
- Embargo_e : 在验证期边界两侧施加长度为 e 的禁运区；
- 显著降低“标签重叠 \Rightarrow 信息泄露”偏差。

网格搜索与随机搜索

网格搜索：

- 在离散网格上穷举所有超参数组合；
- 高维空间中效率低、易浪费在不重要维度。

随机搜索：

- 从给定分布中**随机采样**超参数；
- 在高维空间更高效（重要维度优先被探索）；
- 实务范式：**随机粗搜** \Rightarrow 在高性能区域**局部细化**。

贝叶斯优化：高斯过程 + 采集函数

目标：最小化噪声目标 $f(\lambda) = \hat{R}_{\text{CV}}(\lambda)$ 。以高斯过程先验 $\mathcal{GP}(m(\lambda), k(\lambda, \lambda'))$ 建模 f 。给定观测 $\mathcal{D}_t = \{(\lambda_j, f_j)\}_{j=1}^t$ ，后验为：

$$\mu_t(\lambda) = \mathbf{k}_t(\lambda)' (K_t + \sigma^2 I)^{-1} \mathbf{y}_t,$$

$$\sigma_t^2(\lambda) = k(\lambda, \lambda) - \mathbf{k}_t(\lambda)' (K_t + \sigma^2 I)^{-1} \mathbf{k}_t(\lambda),$$

其中 $[K_t]_{ij} = k(\lambda_i, \lambda_j)$ 、 $[\mathbf{k}_t(\lambda)]_j = k(\lambda, \lambda_j)$ 。

贝叶斯优化：期望改进 (EI)

期望改进 (EI, 最小化情形):

$$\text{EI}_t(\lambda) = (f^* - \mu_t(\lambda) - \xi) \Phi(z) + \sigma_t(\lambda) \phi(z),$$

$$z = \frac{f^* - \mu_t(\lambda) - \xi}{\sigma_t(\lambda)},$$

其中：

- f^* : 当前已知最优（最小）目标值；
- ϕ, Φ : 标准正态密度 / 分布；
- $\xi \geq 0$: 控制探索—利用权衡。

策略：

$$\lambda_{t+1} = \arg \max_{\lambda} \text{EI}_t(\lambda),$$

再评估 $f(\lambda_{t+1})$, 循环迭代。

Successive Halving / Hyperband

思想：在有限预算 B 下**多臂赌博机式**地分配资源。

- 以较小预算评估大量初始配置；
- 保留表现**较好**的一部分进入下一轮；
- 逐轮加大预算（训练步数 / 迭代数 / 树数等），淘汰劣势配置；
- Hyperband 在多个“起始配置数—预算”轨迹上组合 Successive Halving。

对于深度网络 / 提升树这类“可逐步增加预算”的模型尤其高效。

成本敏感阈值（信用 / 欺诈）

金融任务往往不以准确率或 MSE 为最终目标，而是经济指标。成本敏感阈值：

- 假设误报 (FP) / 漏报 (FN) 成本为 (c_{FP}, c_{FN}) ；
- 在贝叶斯最优意义下，阈值基线：

$$\tau^* = \frac{c_{FP}}{c_{FP} + c_{FN}}.$$

- 调参时可直接以成本敏感损失为目标，而不仅仅是 AUC / logloss。

风险调整收益与策略指标

策略收益：

- 由概率 p_t 或预测 \hat{r}_{t+1} 映射到仓位 w_t ;
- 含交易成本 c 的策略收益：

$$r_{t+1}^{\pi} = w_t r_{t+1} - c |w_t - w_{t-1}|.$$

风险调整收益：

$$\text{Sharpe} = \frac{\text{E}[r_t^{\pi} - r_t^f]}{\sqrt{\text{Var}(r_t^{\pi} - r_t^f)}} \sqrt{A},$$

其中 A 为年化因子（如月度 $A = 12$ ）。

- 调参时可直接以样本外 Sharpe / Sortino / 回撤约束为目标；
- 概率用于阈值或连续仓位前，应先做校准（Platt / 保序回归）。

早停作为容量控制与模型选择

早停 (Early Stopping)：

- 将“训练轮数 / 弱学习器数 / 树数”等视为容量控制超参数；
- 验证集上：训练误差单调下降，样本外误差先降后升；
- 在出现过拟合拐点前停止训练，保留最佳迭代。

本质：

- 一种隐式正则化；
- 必须与内层交叉验证联动：每个内折独立选择早停步数；
- 外层仅报告性能，避免对验证集“反复窥视”。

线性 / 平方损失：谱滤波视角

以最小二乘：

$$\min_{\beta} \frac{1}{2} \|y - X\beta\|_2^2, \quad X = U\Sigma V.$$

满步长梯度下降 ($\eta < 2/\|X\|_2^2$) 第 t 次迭代：

$$\hat{\beta}^{(t)} = \sum_{i=1}^r \left(1 - (1 - \eta\sigma_i^2)^t\right) \frac{u_i'y}{\sigma_i} v_i.$$

解释：

- 第 i 个主方向收缩因子：

$$g_t(\sigma_i^2) = 1 - (1 - \eta\sigma_i^2)^t;$$

- 岭回归收缩因子：

$$g_\lambda(\sigma_i^2) = \frac{\sigma_i^2}{\sigma_i^2 + \lambda}.$$

- 二者都是随谱值单调递增的“低通滤波器”，早停可视为一种谱正则化，在合适 t, η 下形状上逼近岭回归。

GBDT / XGBoost：树数与学习率

梯度提升：

$$F_M(x) = \sum_{m=1}^M \eta h_m(x), \quad h_m \in \mathcal{H},$$

- 在固定弱学习器类 \mathcal{H} 下， M 刻画函数类复杂度；
- 更小的 M 对应更“稀疏”的加性模型。

早停的作用：

- 用验证集选择最优迭代 M^* ；
- 等价于在 ERM 上附加对“树数”的结构惩罚；
- 实务经验：“小学习率 η + 大最大轮数 + 早停”往往优于“少轮数 + 大步长”。

深度网络：早停与显式正则

深度网络中早停的实现：

- 在训练集内划分验证集；
- 设置容忍步数 $\text{patience} = P$ 与最小改进 δ ；
- 若验证指标在连续 P 轮内未改善超过 δ ，则终止训练，并回滚到最佳权重。

与显式正则的配合：

- 结合权重衰减 (L_2)、 L_1 稀疏、Dropout、BatchNorm；
- 先用学习率调度（如 ReduceLROnPlateau）降低学习率，仍无改善再触发早停；
- 早停点视作超参数记录下来，外层不再继续“观望”。

时序金融中的实施细节

切片与因果：

- 外层滚动 / 扩展窗口只用于最终评估；
- 内层时间窗用于所有调参与早停；
- 标签涉及未来窗口（如 $t : t + h$ 收益）时，训练样本需对验证期使用 `purge` / `embargo`。

管道化与可审计：

- 在管道中固化“早停触发逻辑、最佳轮数 t^* 、随机种子、学习率轨迹”；
- 阈值、成本与校准也在内层确定；
- 外层仅做一次性样本外报告。

与显式正则的协同与指标匹配

与显式正则的协同：

- 线性 / GLM: L_2 控制范数, 早停作为谱正则补充; L_1 引入稀疏性;
- GBDT / XGBoost: 用 subsample, colsample_bytree、深度、 γ 、min_child_weight 先稳住方差, 再用早停选 M^* ;
- 深度网络: 权重衰减和 Dropout 稳定表示, 早停控制“拟合程度”。

度量与选择的匹配：

- 早停判据应与最终经济目标一致;
- 分类: AUC / PR-AUC / KS / 成本敏感损失 (校准后);
- 择时 / 配置: 样本外 Sharpe / Sortino / 最大回撤 (含交易成本)。

双降现象与实现准则

双降现象：

- 测试误差可能在插值阈值附近“先升后降”；
- 早停往往停在第一个谷值附近，未必是全局最优；
- 建议将早停与学习率、深度、正则强度等联合搜索，并用区块自助法评估差异显著性。

实践小抄：

- 记录“步数—验证指标”时间序列，用滑动窗口平滑后判定改善；
- 设定最小改进 δ 与容忍步数 P ，避免误触发；
- 对小样本内层折：增大 P 、减小 δ ；
- 提升树 / 深网推荐组合：小学习率 + 大最大轮数 + 早停；
- 概率用于阈值 / 仓位前，内层先做概率校准。

回测过拟合与多重比较

两类隐蔽偏差：

- 回测过拟合：在同一数据上反复试参 / 筛选，挑出样本内最佳，但样本外不可复现；
- 多重比较：同时比较大量候选模型 / 策略，标称显著性水平失效，容易把偶然胜出者当成真信号。

需要系统性缓解：

- 外层嵌套评估；
- CSCV / PBO；
- 现实检验（Reality Check）与 SPA；
- FDR 控制与预注册。

外层评估与 purge / embargo

外层（嵌套）评估：

- 将时间轴切成若干外层滚动 / 扩展窗口；
- 每个外层内：仅在训练窗完成所有调参（含早停、阈值、正则强度等）；
- 外层验证窗上只评估一次并记录结果；
- 合并所有外层验证窗，得到稳健的样本外分布。

时间序列：配合 purge / embargo 打断由标签构造引起的区间重叠与信息泄露。

CSCV / PBO：概率化度量回测过拟合

CSCV (Combinatorially Symmetric CV):

- 将全样本分为 S 个等长子段；
- 枚举对称的训练 / 测试划分（如每次选取 $S/2$ 段训练，其余测试）；
- 每个划分上：用训练段选出样本内最佳策略，比较其在测试段的相对排名。

PBO (Probability of Backtest Overfitting):

- 将这些相对排名映射为对数几率，落在负半轴（样本外劣于中位数）的比例即为 PBO；
- PBO 接近 $1/2$ 或更高：回测过拟合风险大；
- PBO 明显低于 $1/2$ ：支持存在可复现的样本外优势。

现实检验（Reality Check）与 SPA

问题：在多模型同时比较时，如何调整“数据窥探”导致的偏差？基本框架：

$$H_0 : \sup_{k \leq m} E[d_{t,k}] \leq 0,$$

其中 $d_{t,k}$ ：第 k 个候选相对某基准的损失差 / 效用差。

- 统计量取最大平均差异；
- 用平稳区块自助法近似其分布，得到针对“最好者”的家族层面 p 值。

Reality Check vs SPA:

- Reality Check 原理稳健但偏保守；
- SPA 对“明显劣势者”进行截尾与重心调整，改善有限样本下的尺寸与功效。

FDR 控制：Benjamini–Hochberg 等

假发现率 (FDR) 控制在候选众多时比 Bonferroni 更可用。BH 过程：

- 将 m 个 p 值排序： $p_{(1)} \leq \dots \leq p_{(m)}$ ；
- 找最大 k 满足

$$p_{(k)} \leq \frac{k}{m} q;$$

- 拒绝 $p_{(1)}, \dots, p_{(k)}$ 并保证（在独立或正相关下） $\text{FDR} \leq q$ 。

扩展：

- Benjamini–Yekutieli (BY) 修正：适应更复杂相关结构；
- 层级 FDR：对“算法 \times 特征集 \times 阈值”这类分层结构，先控制家族层再控制子层。

预注册与报告透明度

预注册 / 冻结规则：

- 在项目伊始明确：目标指标与业务约束；
- 时间切片与数据保护期；搜索预算（配置数 / 评估次数上限）；
- 候选家族与可调整的超参数边界。

模型清单与报告：

- 记录所有试验与版本（含时间戳、随机种子）；
- 最终报告外层折的完整分布，而非单点均值；
- 用区块自助法给出置信区间和显著性；
- 说明尝试的候选总数与筛选规则。

不同模型族的调参要点

随机森林：

- 关键：`mtry / max_depth / min_samples_leaf`；
- 先用 OOB 误差做**粗筛**；
- 再用**时序 CV**（滚动 / 扩展）微调。

XGBoost / GBDT：

- 以 `eta`（学习率）+ `n_rounds`（早停）为骨架；
- 用 `max_depth`, `min_child_weight`, `subsample`, `colsample_bytree` 控制方差；
- `scale_pos_weight` 处理类不平衡。

深度网络：

- 层数 / 宽度、学习率调度、权重衰减、Dropout、BatchNorm、早停联动控制容量；
- 时序任务优先采用滚动 / 扩展**窗口** + `purge` / `embargo`；
- 小样本优先浅层网络 + 强正则化。

推荐流程：面向时序的调参与评估

- ① 固定外层滚动 / 扩展窗口切片（含 purge / embargo），作为最终报告口径。
- ② 在每个外层训练窗内，运行内层滚动 CV：
 - 随机 / 贝叶斯 / Hyperband 搜索超参数；
 - 模型自带早停在内层确定；
 - 以经济目标（夏普 / 成本敏感）或校准后指标作为选型准则。
- ③ 将内层最优超参数在该外层训练窗上重新训练一次，外层验证窗上仅评估一次，并记录收益、风险、回撤、换手率与成本。
- ④ 合并所有外层折的样本外结果，报告分布与区块自助法下的显著性，而非只给一个数字。

评价指标与报告规范

超越单一统计指标：

- 除 AUC / MSE 外，应报告：
 - 年化收益率、波动率、Sharpe / Sortino；
 - 最大回撤；
 - 平均仓位与年化换手（成本暴露）；
- 说明：
 - 阈值 / 成本设定与校准方法；
 - 时间切片与 embargo 参数；
 - 搜索预算与采集策略；
 - 随机种子与版本信息。

总结：金融中的调参与交叉验证，不只是“找最小验证误差”，而是要在因果正确的时序评估中、围绕经济目标、在受控搜索预算下完成模型选择与不确定性度量。遵循上述流程，才能让高维机器学习在金融计量中既好看又能用。

目录

- ① 树模型（随机森林与 XGBoost）的原理及适用性
- ② 神经网络与深度学习方法概述及金融场景应用
- ③ 超参数调优与交叉验证：金融计量视角
- ④ 案例：AI 模型的典型应用场景
- ⑤ 模型可解释性与金融稳定性
- ⑥ 本章小结

收益率预测：模型谱系与设计思路

目标：预测下一期（月度）对数收益率，并评估方向预测与经济表现。

- 线性端（基线）：

- OLS / 岭回归（Ridge），强调可解释性与稳健性；
- 特征：宏观与期限结构因子（VIX 水平/变动、10Y-3M 利差、信用利差代理）；
- 技术指标：6/12 个月动量、3 个月波动率、均线偏离度等；
- 岭回归用 L_2 约束缓解“小样本-弱信号”。

- 非线性端：树模型：

- XGBoost：分段常数函数 + 剪枝/结构惩罚；
- 自然捕捉非线性与高阶交互，在同一特征集上提升 RMSE 与方向准确率。

- 序列端：深度时序：

- LSTM / Transformer：门控 / 自注意力建模长记忆与状态依赖；
- 对“环境切换、慢变量-快反应”的市场结构更敏感。

收益率预测：时序评估与防泄露

评估口径：

- 使用滚动 / 前进（walk-forward）验证，严格按时间顺序；
- 报告：样本外 RMSE、方向正确率，以及含成本的策略表现。

防泄露与调参：

- 所有特征统一滞后一月，避免前视偏差；
- 训练-验证-测试按时间推进，禁止数据“回看未来”；
- 岭回归：内层验证选择 λ ；
- XGBoost：小学习率 + 早停选择最佳迭代数；
- LSTM：patience 控制容量；
- 阈值、成本与（若有）概率校准均在内层确定，外层只做一次性报告。

收益率预测：R 实证脚手架要点

数据与特征构造：

- 从 Yahoo Finance 抓取：SPY, ^VIX, ^TNX, ^IRX, HYG, LQD；
- 按月末对齐，构造：
 - 6/12 个月动量，3 个月波动率；
 - 均线偏离、VIX 水平与变动；
 - 期限利差水平/变动，信用利差变动等；
 - 目标：下一月对数收益率，所有特征滞后一月。

评估与现实检验：

- 滚动窗口步进（默认训练窗 120 个月）；
- 岭回归：内层选 λ ；XGBoost：早停 + 小 η ；
- LSTM 可选（keras 环境，小样本时缩短序列）；
- 策略：“预测 > 0 持有，否则持有现金”，在 $\{0, 5, 10\}$ bp 成本下，计算净值与年化收益 / 波动 / Sharpe；
- 只保留有效预测的模型，稳健地计算 RMSE / 方向准确率与现实检验指标。

波动率建模：GARCH 家族与其局限

线性-条件异方差端：

- GARCH / EGARCH / GJR：直接建模收益的条件方差；
- 能捕捉“冲击-记忆-杠杆效应”，参数具备经济含义；
- 便于残差诊断与稳健性检验。

局限：

- 面向实现波动率（由高频价格构造）时，单一结构偏刚性；
- 难以同时刻画多尺度记忆与异质行为；
- 需要更灵活的多尺度与非线性结构。

波动率建模：HAR-RV 与 LSTM/注意力

HAR-RV (log)：

- 使用日/周/月实现波动 (RV) 的多尺度信息；
- 线性框架：贴合“长记忆”且保持可解释性；
- 可扩展为 HAR-X (加入 VIX / 信用利差等)。

深度序列与混合方案：

- LSTM / 注意力：直接在序列层面学习非线性与跨尺度依赖；
- 混合：先用 GARCH 抽取条件波动作为可解释主干，再由注意力学习残差结构；
- 在解释与拟合之间实现平衡。

波动率建模：R 脚手架与评估

统一目标：对数实现波动 (log-Vol)

- 用 OHLC 计算 Garman–Klass 日实现波动 (GK)，取 log 稳定方差；
- 比较三类模型：
 - ① GARCH(1,1): ugarchroll 逐日滚动预测，refit.every=5；
 - ② HAR-RV(log): d/w/m 结构，滚动估计与一步预测；
 - ③ LSTM / 注意力 / 混合：默认关闭，提供“快速模式”演示。

评估与调参旋钮：

- 样本外 RMSE + 相关系数，绘制“最近 500 天”真实 vs 预测；
- 速度-精度旋钮：keep_last, n_out, refit.every, L_seq 等；
- 支持 EGARCH/GJR + t 分布、HAR-X 扩展与更复杂的高频特征；
- 在规范的内层调参与早停下，注意力/混合模型的增益更稳定。

信用评分：业务背景与统计框架

任务：将个人/企业特征映射为违约风险度量。**核心量：**

- PD（违约概率）、LGD（损失给定违约）、EAD（违约时风险暴露）；
- 预期损失： $EL = PD \times LGD \times EAD$ 。

制度与应用：

- 巴塞尔 IRB：资本需求随 PD 单调变化；
- IFRS 9：全期限预期信用损失，需要随时间与情景变化的 PD 曲线；
- 微观：授信审批、额度/利率定价、欺诈拦截、催收分流、账户监控；
- 宏观：评分体系稳健性关乎资产质量与宏观审慎。

Logit 基线：可解释评分卡

Logit 模型：

$$y_i \in \{0, 1\}, \quad y_i \mid \mathbf{x}_i \sim \text{Bernoulli}(p_i),$$

$$\text{logit}(p_i) = \log \frac{p_i}{1 - p_i} = \beta_0 + \mathbf{x}'_i \boldsymbol{\beta}, \quad p_i = \frac{1}{1 + \exp(-\eta_i)}.$$

- β_j : 特征 x_{ij} 对对数赔率的边际效应；
- 赔率比: $\exp(\beta_j)$, 解释直观。

评分卡映射：

$$\text{Score} = A - B \cdot \text{logit}(\hat{p}),$$

- 通过 PDO (points to double odds) 约束确定 A, B ;
- 工程实现：哑变量/WOE、单调分箱增强稳健与可解释性；
- 完全分离时可用正则化或 Firth 修正；
- 再抽样/类权重需在校准阶段纠偏。

非线性模型与两段式架构

随机森林 / 浅层神经网络：

- 在非线性强、类别变量基数大的场景提升排序能力（AUC/KS）与召回；
- 相对深网，复杂度可控、训练更稳。

两段式架构（常见折中）：

- 用树模型做变量筛选与衍生（分裂点、交互候选）；
- 再用 Logit 固化为“可解释评分卡”；
- 在边缘群体/灰名单上引入二级 AI 复核，搭建“稳态评分 + 局部提升”。

额外关注：

- 概率质量：Brier、校准曲线；
- 成本敏感阈值：按 (c_{FP}, c_{FN}) 确定 τ ；
- 稳定性/漂移：PSI/CSI；
- 公平性：分组 AUC/TPR 差异、拒绝推断的选择偏差修正。

信用评分：R 流水线概览 (1/2)

数据与切分：

- 优先在线下载 UCI German Credit，失败回退
`caret::GermanCredit;`
- 60/20/20 分层切分训练 / 验证 / 测试；
- 类别做哑变量，数值做标准化。

三类模型：

- Logit (基线，可映射为评分卡雏形)；
- 随机森林 (ranger，概率输出)；
- 浅层神经网络 (nnet，小隐层 + L_2 衰减)。

信用评分：R 流水线概览 (2/2)

校准与评估：

- 验证集做 Platt 校准（预留保序接口），将校准器应用于测试集；
- 报告：AUC、KS、PR-AUC、Brier 与校准曲线；
- 计算成本敏感阈值 τ^* 与混淆矩阵；
- 监测 PSI（训练 → 测试）、按年龄分组的 AUC/TPR 差异；
- 明示例数据为“已通过样本”，生产迁移需嵌入**拒绝推断**与完整治理。

因子提取：从 PCA 到自编码器

目标：在高维资产空间中，将共同波动压缩为少数“因子”。**线性起点：**PCA

- 稳健、简单，给出线性因子结构；
- 适合作为资产定价与风险管理的基准。

非线性扩展：自编码器（AE / VAE）

- 端到端学习**非线性因子**；
- 可与线性风险模型衔接：因子时间序列 + 横截面回归；
- 更好捕捉非线性共动与异质结构。

因子-暴露-风险聚合链条

因子提取：

- 从 Yahoo Finance 抓取多只 ETF 日度价格，构造收益率矩阵；
- 使用 PCA 提取 k 个因子（或设 `run_ae=TRUE` 用 Keras-AE）。

暴露与特质风险：

- 在训练窗口，对每只资产做 OLS 回归，估计 β （因子暴露）与 α ；
- 计算特质波动率（残差方差），得到特质协方差。

组合层风险聚合：

- 用因子协方差 + 组合权重聚合 VaR/ES；
- 默认因子协方差使用 EWMA ($\lambda = 0.97$)；
- 提供“波动率目标化”(`vol_target` 选项)：将 MC 无条件波动缩放到测试期真实水平。

因子 VaR/ES 与不确定性度量

VaR/ES 口径：

- 历史法 VaR/ES：基于测试期真实组合收益；
- 因子 MC VaR/ES：对因子做高斯 MC，叠加特质噪声；
- 支持 EWMA 协方差与波动目标化，缓解 MC VaR 偏大。

自助法（Bootstrap）近似“深度贝叶斯”不确定性：

- 对训练窗口重采样，重估暴露 (β/α)；
- 每个重采样下重新计算组合 VaR/ES；
- 得到 VaR/ES 的分位区间（如 5%/50%/95%），用于传达估计不确定性；
- 这是对深度贝叶斯后验不确定性的轻量近似，易于工程落地。

输出：

- 因子时间序列、资产暴露热力图；
- 历史法 / 因子 MC / 自助中位数 VaR 对比图；
- 关键参数：k_factors, use_ewma, lambda_ewma, vol_target, B_boot 等显式暴露。

目录

- ① 树模型（随机森林与 XGBoost）的原理及适用性
- ② 神经网络与深度学习方法概述及金融场景应用
- ③ 超参数调优与交叉验证：金融计量视角
- ④ 案例：AI 模型的典型应用场景
- ⑤ 模型可解释性与金融稳定性
- ⑥ 本章小结

可解释性挑战：监管视角的四个问题

在金融计量中，模型不仅要“准”，还要可说清楚与可治理。黑箱难点：只能给出 $\hat{y} = f(x)$ ，很难回答：

- ① 为什么是这个结果？——重要变量、方向、相对贡献；
- ② 在什么区间最敏感？——局部斜率、形状与边界；
- ③ 换个时点/子样本是否仍成立？——稳定性与漂移；
- ④ 是否存在系统性偏差？——泄漏、选择偏、群体公平。

数据泄漏与漂移：稳定性的敌人

数据泄漏：

- 标签窗口 $[t, t + h]$ 内才能观测到的特征禁止出现在训练特征中；
- 横截面-时间面板：需净化的时序交叉验证；
- 若验证区间为 $[T_k^{\text{start}}, T_k^{\text{end}}]$ ：
 - 删除与该区间时间重叠的训练样本；
 - 在两端空出禁运带 δ (purge/embargo)。

漂移：

- 协变量漂移： $p_t(x) \neq p_s(x)$ ；
- 概念漂移： $p_t(y | x) \neq p_s(y | x)$ ；
- 人口稳定性指数 (PSI)：

$$\text{PSI} = \sum_b (p_b - q_b) \log \frac{p_b}{q_b};$$

- 当 PSI 超阈值（如 0.1 / 0.25）触发再训练或阈值重校；
- 可辅以 JS 散度或“分类器投票”的两样本检验。

SHAP：基于 Shapley 的特征归因

思想：用 Shapley 值把预测分解到各特征上，同时保持局部-全局一致。**定义**（给定样本 x ）：

$$\phi_j(f, x) = \sum_{S \subseteq \mathcal{F} \setminus \{j\}} \frac{|S|! (M - |S| - 1)!}{M!} \left(f_{S \cup \{j\}}(x) - f_S(x) \right),$$

- \mathcal{F} : 特征集合, $M = |\mathcal{F}|$;
- f_S : 对非 S 特征做条件期望的“子模型”;
- 实现时用背景分布近似，树模型可用 TreeSHAP 精确计算。

性质：

- 局部准确: $f(x) = \phi_0 + \sum_j \phi_j$;
- 一致性与缺省性;
- 全局: 聚合 $E|\phi_j|$ 进行特征排序, **交互 SHAP** 识别二阶交互。

LIME 与 PDP / ICE / ALE (1/2)

LIME：局部线性近似

$$\hat{g} = \arg \min_{g \in \mathcal{G}} \sum_i \pi_x(z_i) (f(z_i) - g(z_i))^2 + \Omega(g).$$

- 在 x 的邻域生成样本 $\{z_i\}$, 用权重 $\pi_x(z_i)$ 约束邻近度;
- \mathcal{G} 为简单模型族 (如线性), Ω 控制稀疏;
- 优点: 直观、模型无关; 缺点: 对邻域与权重敏感, 适合与 SHAP 交叉验证。

LIME 与 PDP / ICE / ALE (2/2)

PDP / ICE / ALE：形状与异质性

- PDP: 平均边际效应, $\hat{f}_S(x_S) = \frac{1}{n} \sum_i f(x_S, x_C^{(i)})$;
- ICE: 固定个体 $x_C^{(j)}$, 观察 x_S 变化轨迹;
- 当特征高度相关时 PDP 可能外推, 建议用 ALE:

$$\text{ALE}_j(x) = \int_{x_0}^x E\left[\frac{\partial f(u, \mathbf{x}_{-j})}{\partial u} \middle| x_j = u \right] du.$$

- 全局重要性可用置换重要度 Δ_j 与 SHAP 排序互证。

稳定性：形状约束、漂移修正与校准

评估设计：

- 滚动/扩展窗口 + purged K-fold (两端禁运 δ);
- 所有特征只用过去信息构造，严禁跨窗信息共享。

形状与符号约束：

- 单调性约束： $\partial f / \partial x_j \geq 0$ 或 ≤ 0 ;
- 对关键变量施加平滑/分段单调先验;
- 必要时用保序回归对输出概率做后校准;
- 监测滚动窗口下 SHAP 排名的 Kendall τ / Spearman ρ 稳定性。

漂移与概率质量：

- 协变量漂移：重要性加权损失 $\sum w(x_i) \ell(y_i, f_\theta(x_i))$, $w = p_t(x) / p_s(x)$;
- 概念漂移：再训练或阈值重定标;
- 概率质量：Brier、校准斜率/截距, Platt/保序校准;
- 区间预测：分位数回归或共形预测，给出有限样本覆盖保证。

合规与治理闭环

三道防线与文档化：

- 开发—独立验证—内审分工明确；
- 文档覆盖：数据来源与处理、特征字典、训练/评估划分、超参数与随机种子、阈值与策略、漂移监测与再训练计划；
- 模型版本与环境封装支持完全复现。

公平性与隐私：

- 明确公平性指标：机会均等 (TPR)、均衡误差 (TPR+FPR)、人口比例 (预测正类率) 等；
- 通过阈值调整、约束学习或后处理实现公平性目标；
- 隐私：差分隐私 (ϵ, δ) -DP、联邦学习等。

运营控制：

- 设置“停机开关”与人工兜底机制；
- 将极端场景压力测试（输入扰动 Δx 、尾部损失 ES）纳入上线前和定期检验。

目录

- ① 树模型（随机森林与 XGBoost）的原理及适用性
- ② 神经网络与深度学习方法概述及金融场景应用
- ③ 超参数调优与交叉验证：金融计量视角
- ④ 案例：AI 模型的典型应用场景
- ⑤ 模型可解释性与金融稳定性
- ⑥ 本章小结

本章小结（一）：方法与深度学习

方法层：树模型与提升算法

- 决策树：以“不纯度下降”统一度量（基尼/熵、信息增益/增益率、方差下降）；
- 随机森林：装袋降方差；
- GBDT / XGBoost：二阶近似 + 结构惩罚，关键超参（学习率、树深、最小叶权重、行/列采样率）。

深度学习：从 MLP 到时序模型

- MLP：经验风险最小化，交叉熵 / 分位数损失，Adam/AdamW + 早停；
- 正则化：权重衰减、Dropout、BatchNorm；
- LSTM / Transformer：门控与自注意力结构；
- 所有标准化 / PCA / AE 等变换在训练窗拟合、验证/测试外推，防止泄漏。

本章小结（二）：调参、评估与回测风险

调参与评估：

- 嵌套交叉验证 + Purged/Embargo K-fold 保证因果与防泄漏；
- 网格/随机/贝叶斯优化与 Hyperband 各司其职；
- 谱滤波视角解释“平方损失 + 线性模型中早停近似 L_2 正则”。

回测过拟合与多重比较：

- CSCV/PBO：量化“样本内最佳策略”在样本外的排名稳定性；
- Reality Check / SPA：针对“最好者”的家族层显著性检验；
- BH/BY FDR 控制：在大规模调参下控制假发现率；
- 建议：外层一次性报告 + 区块自举的经济指标显著性。

本章小结（三）：解释性、治理与应用案例

解释与治理：

- SHAP (TreeSHAP)、LIME、PDP/ICE/ALE 组合回答“谁重要、怎么变、在何区间敏感”；
- PSI/JS 散度做漂移监控，Platt/保序做概率校准；
- 阈值 $\tau^* = c_{FP}/(c_{FP} + c_{FN})$ 对齐成本；
- 三道防线、文档化与可追溯，公平性与隐私约束落地。

应用案例：

- 收益率预测：OLS/岭—XGBoost—LSTM，滚动样本外 + 含成本策略检验；
- 波动率建模：GARCH-HAR-RV-注意力（可选），统一 log-RV 目标；
- 信用评分：Logit vs 随机森林 / NN，校准后评估 AUC / PR-AUC / Brier / 公平性；
- 因子与风险：PCA/AE + OLS 暴露 + EWMA 协方差 + 因子 MC / 历史法 VaR/ES + 自助区间。