

风险计量分析与工具 R 语言课程作业

---Chapter 9 Fama-French Factor Model

小组成员：周 旭 2019100445

蒋伟林 2019102474

一、Chapter 9 回答的问题

（一）综述

教材的这部分在前一章 CAPM 模型的基础上，构建了 Fama-French 三因子模型，将 CAPM 的一元回归拓展为多元回归。

我们自己选择数据，参考教材的流程，相当于是用新的数据复现了课本的三因子模型。具体来说，我们的工作主要分以下几个部分展开：（1）数据处理；（2）构建 Fama-French 三因子模型；（3）用 ggplot 包来呈现回归结果；（3）用 tidyverse 和 tibble-time 包来回归计算滚动的三因子模型；（4）用 ggplot 包来呈现滚动回归结果。接下来，我们会详细讲述各部分的基本内容和结果。

（二）基本内容

1. 数据来源及说明

我们在“国泰安”数据库中获取了 Fama-French 因子数据。我们选用的基本数据介绍见下表。

表 1 数据基本情况介绍

序号	字段	数据类型	字段标题	字段说明
1	MarketTypeID	Nvarchar	股票市场类型编码	P9706：综合 A 股市场
2	TradingMonth	Datetime	交易月份	以 YYYY-MM 表示
3	RiskPremium2	decimal	市场风险溢价因子（总市值加权）	考虑现金红利再投资的月市场回报率（总市值加权平均法）与月度化无风险利率之差（央行公布三月定存基准利率）。
4	SMB2	decimal	市值因子（总市值加权）	小盘股组合和大盘股组合的月收益率之差，组合划分基于 FAMA 2*3 组合划分方法。组合月收益率的计算采用总市值加权计算。
5	HML2	decimal	账面市值比因子（总市值加权）	高账面市值比组合和低账面市值比组合的月收益率之差，组合划分基于 FAMA 2*3 组合划分方法。组合投资收益率的计算采用总市值加权。

我们选用的数据时间跨度是 2009 年 1 月 1 日到 2018 年 9 月 30 日，选取的是月频率数据。

2. 源代码及注释

```
install.packages("tidyverse")
install.packages("lubridate")
install.packages("readxl")
install.packages("highcharter")
install.packages("tidyquant")
install.packages("timetk")
install.packages("tibbletime")
install.packages("quantmod")
install.packages("PerformanceAnalytics")
install.packages("scales")

library(tidyverse)
library(lubridate)
library(readxl)
library(highcharter)
library(tidyquant)
library(timetk)
library(tibbletime)
library(quantmod)
library(PerformanceAnalytics)
library(scales)

#数据导入
data<-read.csv("C://Users//Administrator//Desktop//Fama-French.csv", header=T)

colnames(data)[1] <- "date"
data$date <- as.character(data$date)
for (i in 1:length(data$date)) {data$date[i]<-
paste(substr(data$date[i], 1, 4), substr(data$date[i], 5, 6), "01", sep=
"-")}
data$date <- as.Date(data$date, format="%Y-%m-%d")
head(data, 3)

#日期处理
data%>%
  select(date) %>%
  mutate(date = lubridate::rollback(date)) %>%
  head(1)
data %>%
  select(date) %>%
  mutate(date = lubridate::rollback(date + months(1))) %>%
```

```

    head(1)
#完整处理
data<-

read.csv("C://Users//Administrator//Desktop//FF.csv",header=T) %>%
  rename(date = X1) %>%
  mutate_at(vars(-date), as.numeric) %>%
  mutate(date =
    ymd(parse_date_time(date, "%Y%m")) %>%
    mutate(date = rollback(date + months(1)))
  )
head(data, 3)
#结果显示

```

	date	RiskPremium	SMB	HML	Rf
1	<NA>	0.065255	0.031545	-0.036521	0.001856
2	<NA>	0.048575	0.047726	-0.020951	0.001856
3	<NA>	0.140575	0.085989	-0.045720	0.001856

```

#将投资组合设置为 8 只股票的组合
#600000（浦发银行）权重 20%
#600008（首创股份）加权 15%
#600011（华能国际）加权 15%
#600018（上港集团）加权 10%
#600022（山东钢铁）加权 10%
#600030（中信证券）加权 10%
#600048（保利地产）加权 10%
#600054（黄山旅游）加权 10%

#提取数据库中数据
install.packages('DBI')
install.packages('RMySQL')
library(DBI)
library(RMySQL)
symbols<-c("600000","600008","600011","600018","600022","600030",
"600048","600054")
data1<-list() #创建空数据框
for (i in 1:8){
  mydb= dbConnect(MySQL(),user='ktruc002', password='35442fed',
dbname='cn_stock_quote', host='172.19.3.250')
  SQL_statement<-paste("SELECT `day`,`close`
FROM `cn_stock_quote`.`daily_adjusted_quote`
WHERE code=",symbols[i],"and day >='2008-12-31' <='2018-09-30'
ORDER BY `day` DESC ")
  aa <- dbGetQuery(mydb,SQL_statement)
  colnames(aa)[2]<-paste("x",symbols[i],sep="",collaspe="")
}

```

```

    data1[[i]]=aa
  }
stockdata<-data1%%>reduce(merge)
prices<-xts(stockdata[, -1], order.by = as.Date(stockdata[, 1]))
#8 项资产的比例，顺序与 symbols 中的相同
w <- c(0.20, 0.15, 0.15, 0.10, 0.10, 0.10, 0.10, 0.10)
head(prices, 3)

```

#使用 tidyverse 将数据转换为月度回报

```

asset_returns_dplyr_byhand <-
  prices %>%
  to.monthly(indexAt = "lastof", OHLC = FALSE) %>%
  #将索引转换为日期
  data.frame(date = index(.)) %>%
  #删除索引，因为它已转换为行名称
  remove_rownames() %>%
  gather(asset, prices, -date) %>%
  group_by(asset) %>%
  mutate(returns = (log(prices) - log(lag(prices)))) %>%
  select(-prices) %>%
  spread(asset, returns)

```

#除去 NA 项

```

asset_returns_dplyr_byhand <-
  asset_returns_dplyr_byhand %>%
  na.omit()

```

#tidyverse 要求使用 long 格式或整洁格式的数据，其中每个变量都有自己的列，而不是 wide 格式

#为了使资产回报整洁，我们需要一个名为“date”的列、一个名为“asset”的列和一个名为“returns”的列

#asset_returns_long 有 3 列，每个列对应一个变量：日期、资产、回报

```

asset_returns_long <-
  asset_returns_dplyr_byhand %>%
  gather(asset, returns, -date) %>%
  group_by(asset)

```

#使用 tq_portfolio() 将 asset_returns_long 转换为投资组合回报

```

portfolio_returns_tq_rebalanced_monthly <-
  asset_returns_long %>%
  tq_portfolio(assets_col = asset,
               returns_col = returns,
               weights = w,
               col_rename = "returns",
               rebalance_on = "months")

```

#####第九章

left_join 使我们能够将这些数据对象合并在一起，然后将任何不匹配的行消除。我们还将 FF 数据转换为十进制格式，并创建一个新的列称为 R_excess，使 $R_{\text{excess}} = \text{returns} - \text{RF}$

```
ff_portfolio_returns <-  
  portfolio_returns_tq_rebalanced_monthly %>%  
  left_join(data, by = "date") %>%  
  mutate(MKT_RF = MKT_RF,  
         SMB = SMB,  
         HML = HML,  
         RF = RF,  
         R_excess = round(returns - RF, 4)) %>% select(-returns, -RF)  
ff_portfolio_returns <-  
  ff_portfolio_returns %>%  
  na.omit()
```

开始建模。用之前算出的 portfolio return, 三因子, 通过 lm 函数开始拟合 beta。此外，我们设置 95% 的系数的置信区间，然后重命名。

```
ff_dplyr_byhand <-  
  ff_portfolio_returns %>%  
  do(model =  
      lm(R_excess ~ MKT_RF + SMB + HML,  
         data = .)) %>%  
  tidy(model, conf.int = T, conf.level = .95) %>%  
  rename(beta = estimate)  
## 转换数据格式，调整小数位。表格显示的分别是 beta 的预测值，标准差，  
pvalue 和最高最低置信区间  
ff_dplyr_byhand %>%  
  mutate_if(is.numeric, funs(round(., 3))) %>%  
  select(-statistic, -std.error)
```

通过 ggplot 将 FF 可视化

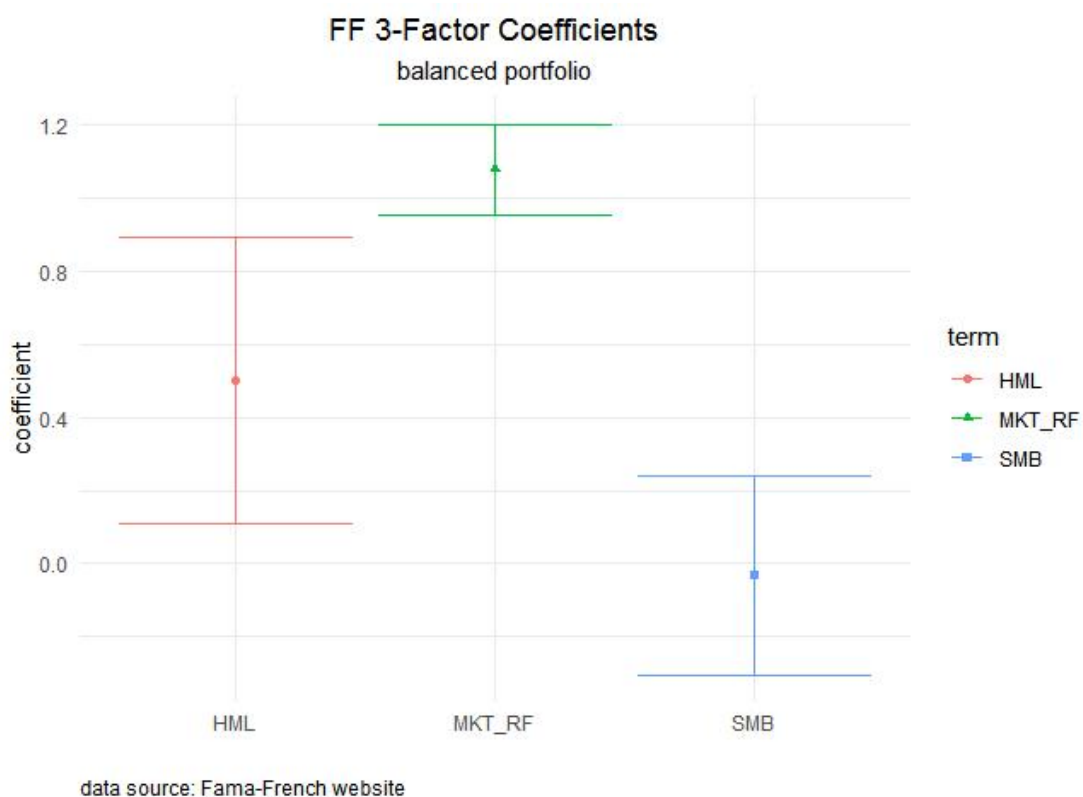
添加置信区间, 用 filter 函数过滤掉 “(Intercept)” 截距，通过 geom_errorbar 展现出带有误差棒的条形图。用 ggplot2 设置图表主题格式布局，包括标题，附注，坐标轴名称。通过 theme 设置主题。

```
ff_dplyr_byhand %>%  
  mutate_if(is.numeric, funs(round(., 3))) %>%  
  filter(term != "(Intercept)") %>%  
  ggplot(aes(x = term,  
             y = beta,  
             shape = term,  
             color = term)) +
```

```

geom_point() +
geom_errorbar(aes(ymin = conf.low,
                  ymax = conf.high)) +
labs(title = "FF 3-Factor Coefficients",
      subtitle = "balanced portfolio",
      x = "",
      y = "coefficient",
      caption = "data source: Fama-French website") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5),
      plot.caption = element_text(hjust = 0))

```



#回归结果显示，HML 因子和 MKT_RF 因子贡献了主要的收益来源，而 SMB 因子的 beta 值基本在零左右，该因子对组合收益的贡献比较小。

使用 tidyverse 和 tibble-time 滚动 Fama-French

#检验滚动式 Fama-French 结果并探索该模型在不同时间段内情况。

#以下工作特定用于 FF，但通常可以应用于我们希望滚动拟合的任何多元线性回归模型。

#我们首先使用 tibbletime 中的 rollify () 函数定义一个滚动模型

选择 24 个月作为滚动窗口 window <- 24

define a rolling ff model with tibbletime

rolling_lm <-

```

rollify(.f = function(R_excess, MKT_RF, SMB, HML) {

```

```
lm(R_excess ~ MKT_RF + SMB + HML)
}, window = window, unlist = FALSE)
#####我们将资产组合的收益率导入 rolling 函数
rolling_ff_betas <-
```

```
ff_portfolio_returns %>%
mutate(rolling_ff =
      rolling_lm(R_excess,
                  MKT_RF,
                  SMB,
                  HML)) %>%
```

```
slice(-1:-23) %>%
select(date, rolling_ff)
head(rolling_ff_betas, 3)
```

#现在，我们有了一个名为 rolling_ff_betas 的新数据框，我们可以使用 map(rolling_ff, tidy) 函数整理 rolling_ff 列，然后使用 unnest 语法将 map 类型展开成多行，这与 CAPM 的工作非常相似，只是 FF 有多个自变量。

```
###
```

```
rolling_ff_betas <-
ff_portfolio_returns %>%
mutate(rolling_ff =
      rolling_lm(R_excess, MKT_RF,
                  SMB,
                  HML)) %>%
mutate(tidied = map(rolling_ff,
                    tidy,
                    conf.int = T)) %>%
unnest(tidied) %>%
slice(-1:-23) %>%
select(date, term, estimate, conf.low, conf.high) %>%
filter(term != "(Intercept)") %>%
rename(beta = estimate, factor = term) %>%
group_by(factor)
head(rolling_ff_betas, 3)
```

###现在，我们 3 个因子中的每一个都有滚动的 beta 和置信区间。

#我们可以应用相同的代码逻辑来提取模型的滚动 R^2 ，唯一的区别是我们称为 glance() 而不是 tidy()。

```
rolling_ff_rsquared <-
ff_portfolio_returns %>%
mutate(rolling_ff =
      rolling_lm(R_excess,
                  MKT_RF,
                  SMB,
                  HML)) %>%
slice(-1:-23) %>%
```



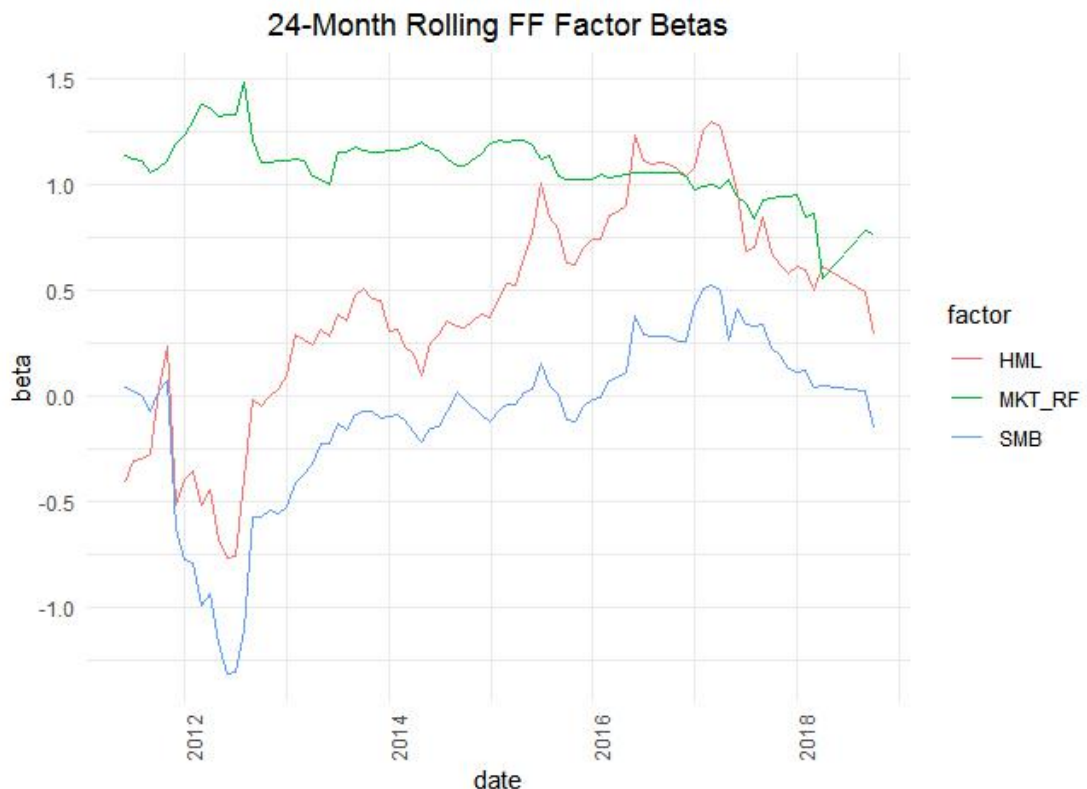
```
mutate(glanced = map(rolling_ff,
                      glance)) %>%
unnest(glanced) %>%
select(date, r.squared, adj.r.squared, p.value)
head(rolling_ff_rsquared, 3)
```

#我们已经提取了滚动 beta 和滚动模型结果，现在我们进行可视化。

使用 ggplot 将 Fama-French 可视化

#我们首先使用 ggplot () 绘制滚动因子 beta 图像。这使我们对每个因子的解释能力如何随时间变化有了直观上的感觉

#：设 x 轴为时间轴，y 轴变量为 beta，用折线图对 24 个月的滚动因子 beta 进行可视化。



#图形描述的是滚动回归计算的 beta 值，揭示了一些有趣的趋势。总体上看，SMB 和 HML 都徘徊在零附近，而 MKT 因子则徘徊在 1 左右。这与我们的带有置信区间的 beta 曲线一致。同时。HML 因子 MB 因子 eta 值波动比较大，而 MKT_RF 因子的 beta 值则波动比较小。

```
rolling_ff_betas %>%
ggplot(aes(x = date,
            y = beta,
            color = factor)) +
geom_line() +
labs(title= "24-Month Rolling FF Factor Betas") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 90))
```

```

#现在，我们对滚动 R2 进行可视化
#我们首先使用 timetk 中的 tk_xts () 函数将 R2 转换为 xts 对象
rolling_ff_rsquared_xts <-
  rolling_ff_rsquared %>%
  tk_xts(date_var = date, silent = TRUE)
#然后将 xts 对象传递给 highchart (类型= “ stock”) 代码流，将滚动 R 平方时间序列添加到 hc_add_series
highchart(type = "stock") %>%
  hc_add_series(rolling_ff_rsquared_xts$r.squared,
    color = "cornflowerblue",
    name = "r-squared") %>%hc_title(text = "Rolling FF
3-Factor R-Squared") %>%
  hc_add_theme(hc_theme_flat()) %>%
  hc_navigator(enabled = FALSE) %>%
  hc_scrollbar(enabled = FALSE) %>%
  hc_exporting(enabled = TRUE)
#虽然结果图比较不稳定，但 R2 从未真正离开介于 0.9 到 0.95 这个区间。
我们可以调整 y 轴的最小和最大值。

```

#如图



```
highchart(type = "stock") %>%
```



#如上图所示，当 y 轴范围扩大时，我们的 R^2 在投资组合的有效期内波动性总体小了很多。具体来看，2017 年之前， R -Squared 的值基本在 0.7 以上，说明三因子模型对组合收益率的解释力度很好，但是，2018 年开始， R -Squared 的值快速下降，最低降至 0.3 以下，说明这段时期三因子模型对组合的解释力度比较差。

二、运用的金融理论

（一）Fama-French 三因子模型

Fama 和 French 在研究美国资本市场金融资产收益率时发现，不同的因子对收益率有着不同的解释度，其中上市公司市值、账面市值比及市盈率可以较好地分析不同股票的收益率，收益率中的超额部分是 P 对资本资产定价模型中未能解释部分的补偿。1993 年，Fama 和 French 在 CAPM 和 APT 的基础上对美国市场进行了实证分析，得出了上市公司市值因子（SMB）、账面市值比因子（HML）以及市场资产组合可以对不同股票的收益率进行解释的结论。

Fama-French 三因子模型公式：

$$E(R_i) - R_f = b_i [E(R_m - R_f)] + s_i E(SMB) + h_i (HML)$$

其中 R_i 代表金融资产收益率； R_f 代表无风险收益率； R_m 代表市场基准的收益率； $(R_m - R_f)$ 代表市场风险收益超过无风险收益的部分，SMB 代表市值因子的期望收益率，即小市值股票的期望收益率减去大市值股票的期望收益率；SMB 代表账面市值比因子的期望收益率，即高账面市值比组合的期望收益率减去低账面市值比组合的期望收益率。 s_i 反映了这个投资组合对规模因子的敏感度， h_i 则反映了投资组合对价值因子的敏感度。这里的 b_i 是模型中的反映了投资组合对市场风险的敏感情况。

Fama-French 三因子模型将资产收益归结于三个因子，因子在评价模式上由

此出现了按账面市值比划分的价值型、成长型和平衡型股票以及按市值划分的大、中及小盘股。

该方法的优点在于操作简单，方便理解。模型可以看成目前被广泛研究和使用的多因子模型的雏形。虽然，在中国的国内中，模型的形式上已经与当初有了较大的出入。但模型的思想仍然核心仍然贯穿于实务界的研究和开发过程中。

经过国内学术界大量学者的实证检验，Fama-French 模型在国内市场被认为是有效的。也就是说中国的股市是存在规模效应和估值效应的。但 Fama-French 模型中在实际应用也有自己的不足：

(1) Fama-French 三因子模型并没有发展成完整的理论体系来对市场中的所有股票的收益情况进行解释。对于规模效应和估值效应出现的原因。模型只是从统计学的角度通过回测的方式对其进行了验证，而没有从理论层面进行更深入的解读。

(2) Fama-French 三因子模型都只能描述投资标的的较小的一部分特征。在投资实务领域，还有更多的基本面或者技术面的指标被使用者认为对投资标的的超额收益具有解释能力。比如 PE 因子在投资实务中被广泛使用，但是三因子模型却没有对其进行讨论。不排除在不同的时期，影响市场走势的因子是不同的

(3) Fama-French 三因子模型不适用对投资组合短期波段的解读。但对长期的投资具有一定的解释能力。