

金融学类专业前沿系列课程直播

计算机类课程与金融专业课程的融合： 以《R语言与金融数据分析》为例



浙江工商大学 | 金融学院

方霞 | fangxia@zjsu.edu.cn

March 6, 2020

CONTENTS



开课背景



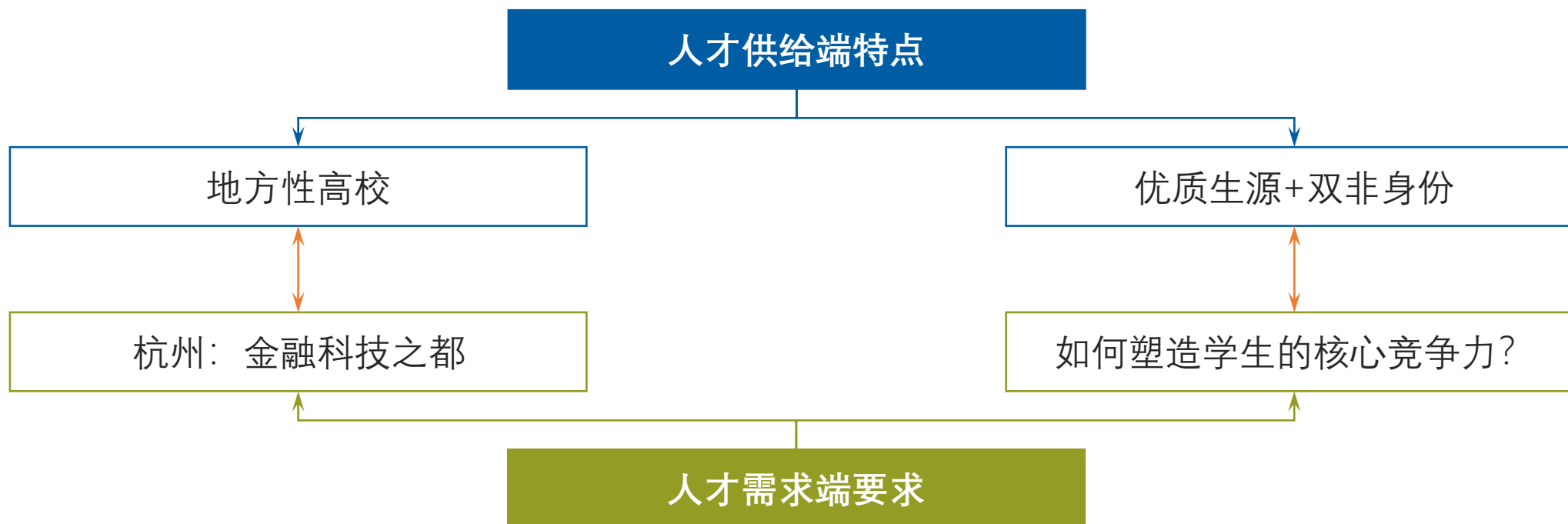
基于混合式教学的课程设计



心得感悟

开课背景

- 供给端：金融专业教育**复合型人才**培养需求



- 需求端：金融机构从业**数据处理能力**极端重要

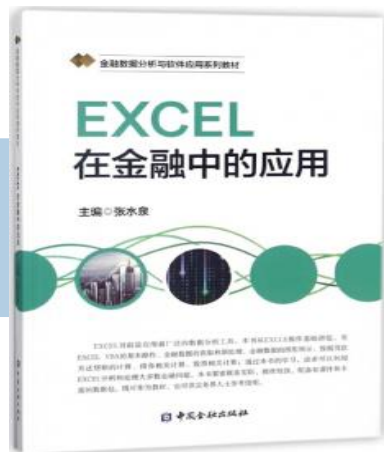
开课背景

- 传统解决方案：增设专门计算机课程（C、Python、算法……）

痛点：纯计算机较少结合金融应用

- 复合型金融人才培养从复合型金融课程体系开始

公司金融
(证券) 投资学



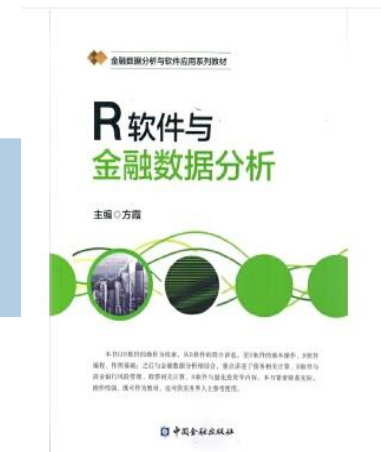
财务估值运用

计量经济学



大数据分析处理

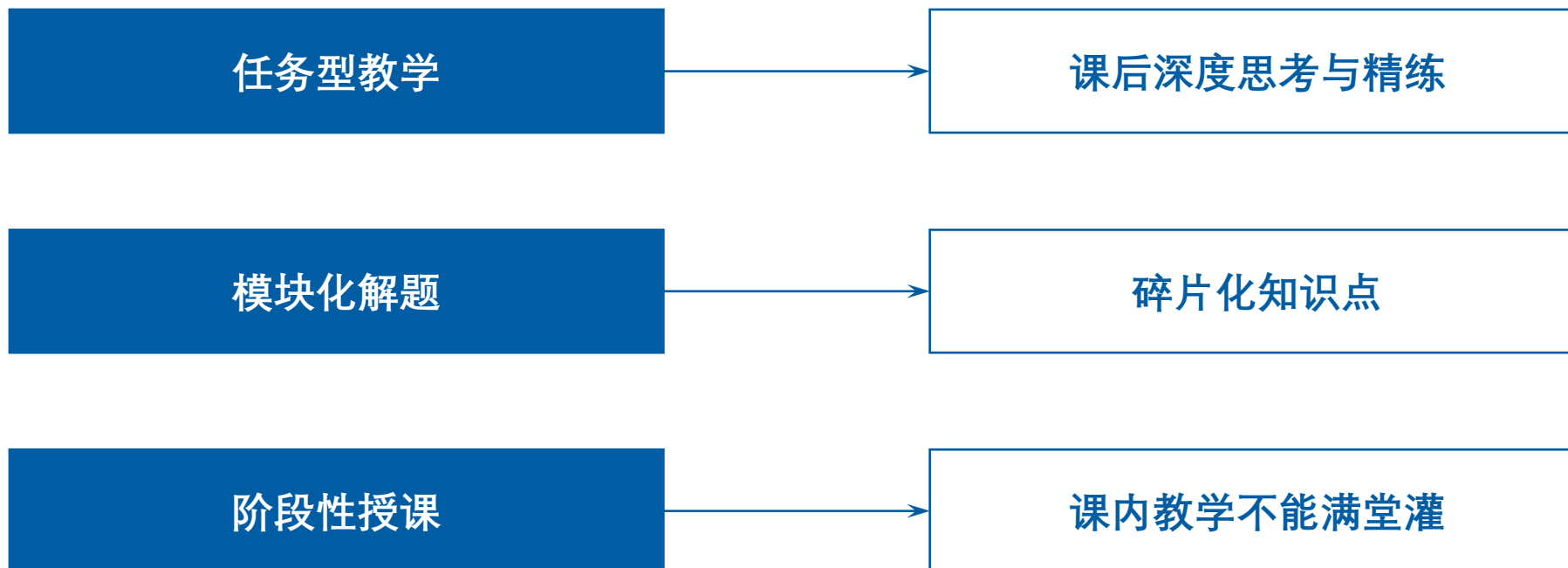
固定收益证券、金融工程、金融风险管理



综合金融能力培养

为什么选择混合式教学？

- 课程特点



- 本课程混合式教学的实施历程

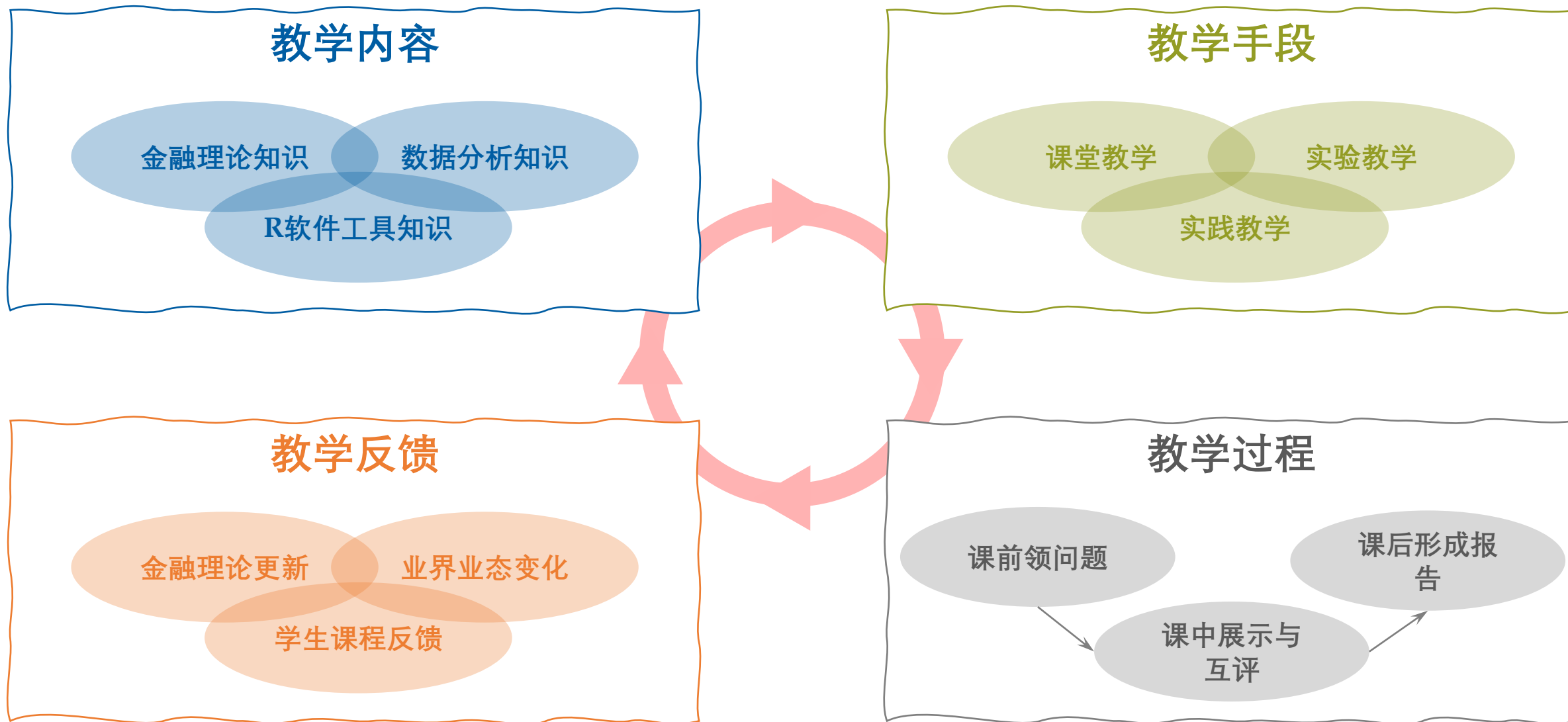
课程设计

- 总学时：48学时
 - 线上学习学时： ≥ 30 学时
 - 线下学习学时：30学时
- 前置课程：
 - 投资学/证券投资学
 - 计量经济学
 - 金融工程
 - 固定收益证券
 - 金融风险管理
 -

首页 > 经济学



教学逻辑：三元融合，动态闭环



课程知识体系

课程内容

R软件的基础内容

金融数据导入，股票收益率和波动率计算，金融技术分析，购买力平价利率，F-H之谜

R软件与固定收益分析收益

应计利息计算，债券内在价值和收益率的计算，债券久期和凸性的计算，综合案例

R软件与金融风险管理

市场风险度量（各种VaR值计算，VaR的回测，改进VaR值），**信用风险管理**（Merton模型实例，KMV模型，统计违约风险模型与PD预测，利差风险与违约密度模型，投资组合信用风险），**操作性风险度量**等

R软件在股票市场应用

经典投资组合构建：资产 β 系数的估计，投资组合优化

量化投资框架：因子选股策略，配对策略，30分钟突破策略

案例1 R软件的基础操作

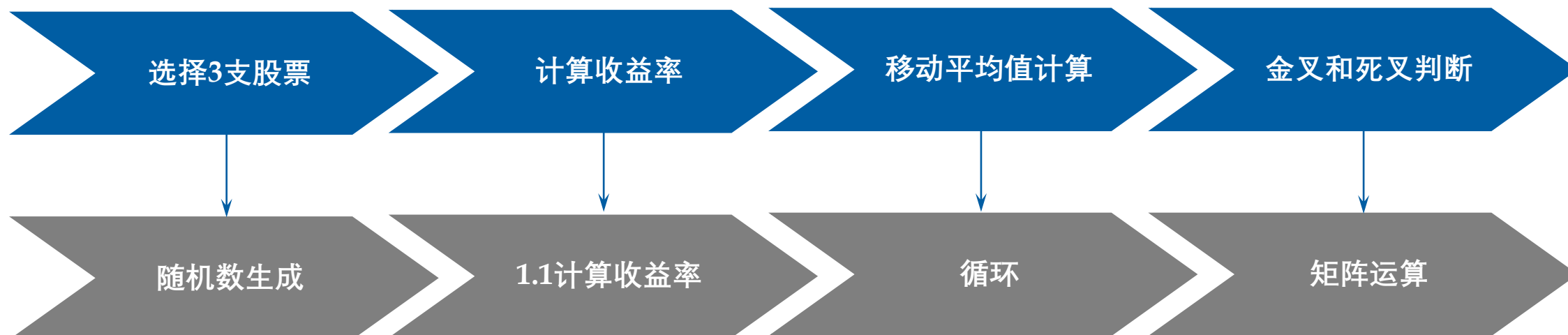
数据处理步骤	软件操作知识点	金融应用
数据导入	连接主流数据库导入数据	股票市场数据，汇率数据，公司财务数据-----quantmod包 宏观数据-----世界银行WDI包 贸易数据----- 联合国商品贸易统计数据库
数据清洗	数据类型、变量重新编码	收益率的计算
数据处理	循环和函数编写	等额本金还款和等额本息还款的计算 技术分析金叉和死叉点判断
基础数据分析	统计分析，回归分析	购买力平价的检验
结果呈现	数据可视化	

1.1 数据清洗

- R软件知识点：数据类型，变量重新编码
- 金融理论知识：金融资产收益率计算
- 实验操作
 - 从quantmod包中下载任意一个股票数据
 - 计算这支股票的日收益率 (ret)
 - 产生一个日收益率分类的变量 (retlevel)
 - 日收益率小于负0.5%标“暴跌”，负0.5%-0%之间标“略跌”，0-0.5%标“略涨”，大于0.5%为暴涨，等于0或者缺失标“干嘛呢？”

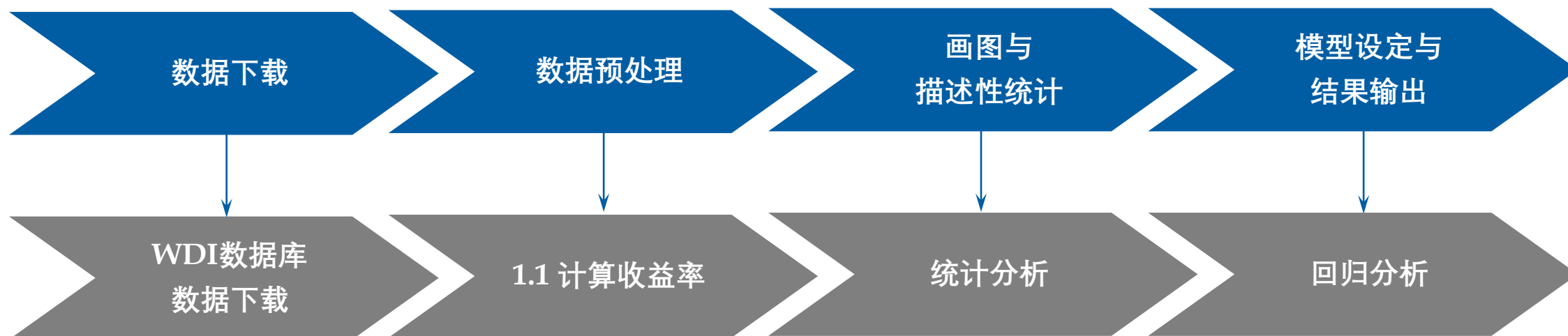
1.2 数据处理

- R软件知识点：循环和函数编写
- 金融理论知识：
 - 等额本金还款和等额本息还款的计算
 - 技术分析金叉和死叉判断



1.3 数据处理

- R软件知识点：统计分析，回归分析
- 金融理论知识：购买力平价理论
- 实验操作
 - 截面回归，时间序列回归，面板回归



1.4 结果呈现

```
plot(c(0,6),c(0,5),type="n", main = "pch=", xlab="",ylab="",  
     axes="FALSE")  
points(1,5,pch=0); text(0.5,5,"0")  
points(1,4,pch=1); text(0.5,4,"1")  
points(1,3,pch=2); text(0.5,3,"2")  
points(1,2,pch=3); text(0.5,2,"3")  
points(1,1,pch=4); text(0.5,1,"4")  
points(2,5,pch=5); text(1.5,5,"5")  
points(2,4,pch=6); text(1.5,4,"6")  
points(2,3,pch=7); text(1.5,3,"7")  
points(2,2,pch=8); text(1.5,2,"8")  
points(2,1,pch=9); text(1.5,1,"9")  
points(3,5,pch=10); text(2.5,5,"10")  
points(3,4,pch=11); text(2.5,4,"11")  
points(3,3,pch=12); text(2.5,3,"12")  
points(3,2,pch=13); text(2.5,2,"13")  
points(3,1,pch=14); text(2.5,1,"14")  
points(4,5,pch=15); text(3.5,5,"15")  
points(4,4,pch=16); text(3.5,4,"16")  
points(4,3,pch=17); text(3.5,3,"17")  
points(4,2,pch=18); text(3.5,2,"18")  
points(4,1,pch=19); text(3.5,1,"19")  
points(5,5,pch=20); text(4.5,5,"20")  
points(5,4,pch=21); text(4.5,4,"21")  
points(5,3,pch=22); text(4.5,3,"22")  
points(5,2,pch=23); text(4.5,2,"23")  
points(5,1,pch=24); text(4.5,1,"24")  
points(6,5,pch=25); text(5.5,5,"25")
```

- R软件知识点：可视化
- 金融理论知识：1.2和1.3
- 教学方法：改代码

pch=

0	□	5	◇	10	⊕	15	■	20	•	25	▽
1	○	6	▽	11	☆	16	●	21	◊		
2	△	7	⊠	12	⊞	17	▲	22	□		
3	+	8	*	13	⊗	18	◆	23	◇		
4	×	9	⊕	14	⊠	19	●	24	△		

案例2 R软件与固定收益分析

- R软件知识点：模块化，函数编写，函数嵌套函数
- 金融理论知识：债券收益率、久期、凸性等计算
 - 上海证券交易所官方网站 (<http://www.sse.com.cn>) 得到在其交易的某债券基本信息

05国债(4) 010504

基本信息 成交概况 换手率变化

债券概况

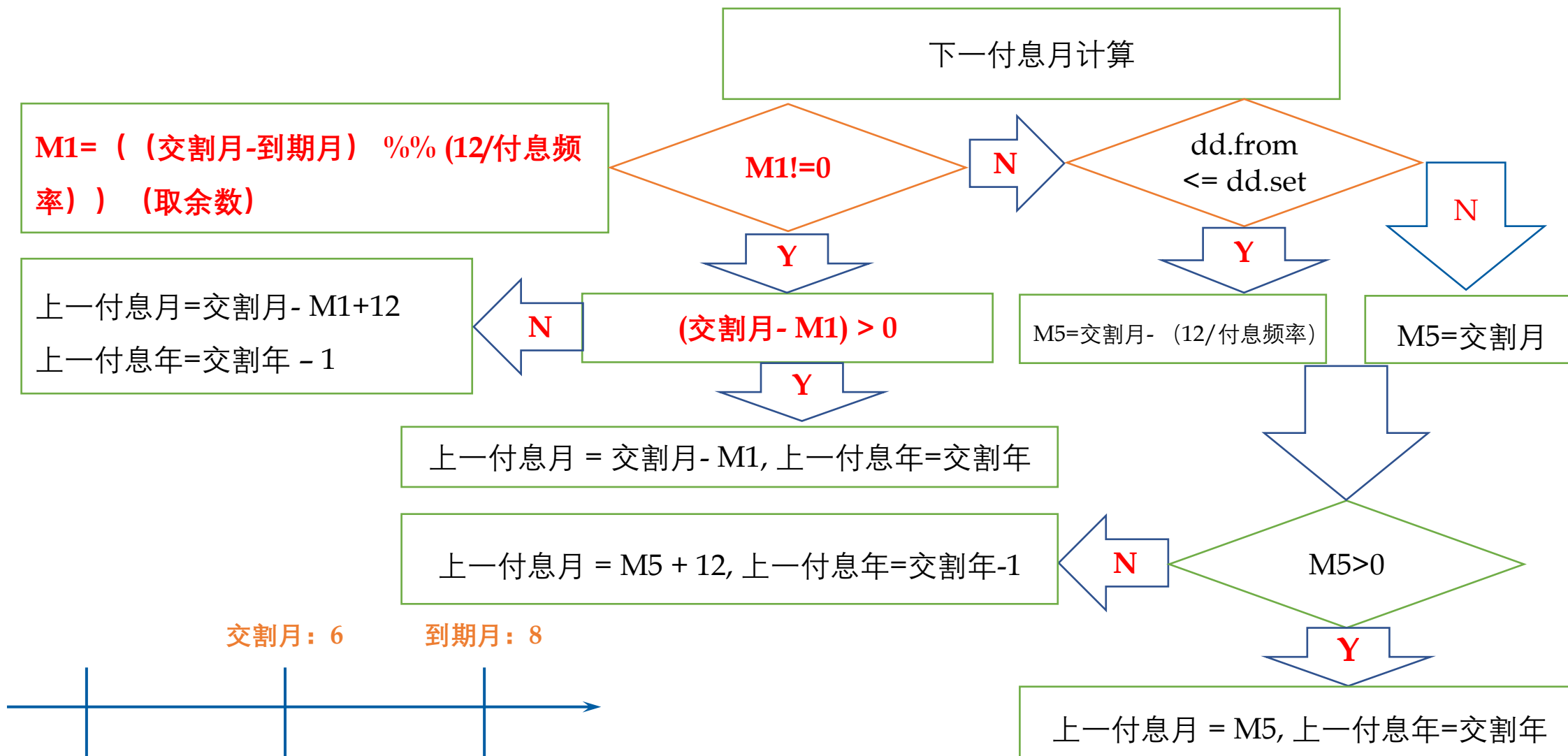
挂牌代码:	010504
挂牌名称:	05国债(4)
国债全称:	2005年记账式(四期)国债
实际发行量:	339.2
发行价格:	100
发行方式:	记账式
期限:	20
到期日:	2025-05-15
票面利率:	4.11
国债付息方式:	按半年付息
计息日期:	2005-05-15
上市日期:	2005-05-25

债券价格计算，收益率计算，久期，凸性计算

应计利息

前一付息日和后一付息日

2.1 应计利息处理



2.2 R软件与固定收益证券的交易与风控

- 债券风控指标的计算

- 久期与凸性
- 有效久期、有效凸性、基点价值
- 主成分久期与关键利率久期

- 债券组合交易与风险分析

- 子弹策略
- 杠铃策略
- 梯式策略

案例建设中

案例3 R软件与金融风险管理

- R软件知识点：模块化，函数编写
- 金融理论知识：各类金融风险测度
- 四大经典风险
 - 市场风险
 - 信用风险
 - 操作风险
 - 流动性风险

3.1 市场风险

- **VaR度量**
 - 历史数据排序法
 - 正态分布法
 - 混合时间序列加权法
 - 正态分布法的延伸——GARCH模型
 - 厚尾分布法——Weibull分布
 - 核密度估计法
 - 蒙特卡洛模拟
- **VaR的回溯测试与压力测试**
- **改进：极值分布理论、条件VaR**
- **基于风险因子的市场风险管理：多因子风险测算与对冲模型**
- 如果你想知道，上证指数在下一个交易日内，有95%的置信水平，收益率不低于多少时，该怎么办？或者换言之，如果你想知道，上证指数在下一个交易日内，有5%的概率损失会超过多少时，该怎么办？
 - 正态分布法：估计均值、标准差
 - 厚尾分布法：Weibull分布，估计 λ 、 k
 - 历史模拟法：取历史数据排序

3.2 信用风险

- 信用风险的成因和度量（理论介绍）
- Merton违约模型及KMV模型
- 统计违约风险模型与PD预测
- 信用利差与违约密度模型
- 投资组合信用风险

案例4 R软件与股票资产组合构建

- 计算组合收益和方差
- 允许卖空下组合优化
- 卖空约束下组合优化

4.1 计算组合的收益与方差

- 历史收益替代
 - 计算个股历史收益均值
 - 估计组合收益协方差矩阵

4.1.1 用个股历史平均收益替代

```
#从之前下载的数据中挑选十只股票，用其月收益率估计协方差
stk_list = list.files("./rets/")
ret_list = NULL
for(stk in stk_list){
  stk = read.csv(paste0("./rets/",stk),sep=',')
  ret_list = cbind(ret_list,stk$Mret)
}
ret_list = na.omit(ret_list)
#期望收益率
aver = apply(ret_list,2,function(x) mean(x,na.rm=T))
#协方差
V = cov(ret_list)
```


4.1.2 单指数模型法

- 计算个股历史收益率，用单指数模型估计预期收益和协方差
- $E(r_i) = r_f + \beta_i^{historical} \times (r_m - r_f)$

```
betas = read.csv("betaRes.csv", sep=',')  
rf = 0.00314 #一年定期存款利率  
rm = 0.0199 #市场收益  
varm = 0.01008 #市场波动率  
  
n = nrow(betas)  
  
aveRets = rf + betas*(rm-rf)  
covMats = betas%*t(betas)*varm  
portData = list(returns = aveRets, covariance = covMats)  
save(portData, file="portData.RData")
```

4.2 允许卖空时最优组合的构建

- 计算最优权重
- 绘制前沿组合

4.2.1 计算最优权重

- 求解优化问题估计有效边界
- 用前沿组合生成有效边界
- $U - c = \Omega z \Rightarrow z = \Omega^{-1}(U - c)$
- 最优权重 $x = z / z^T z$
- $\text{maximum} \frac{E(R_p) - R_f}{\sqrt{x^T \Omega x}}$

```
##Port optim
c1 = 0
c2 = 0.08
covariance = cov(ret_list)
returns = apply(ret_list,2,mean)

z1 = solve(covariance,returns-c1)
z2 = solve(covariance,returns-c2)
x1 = z1/sum(z1)
x2 = z2/sum(z2)

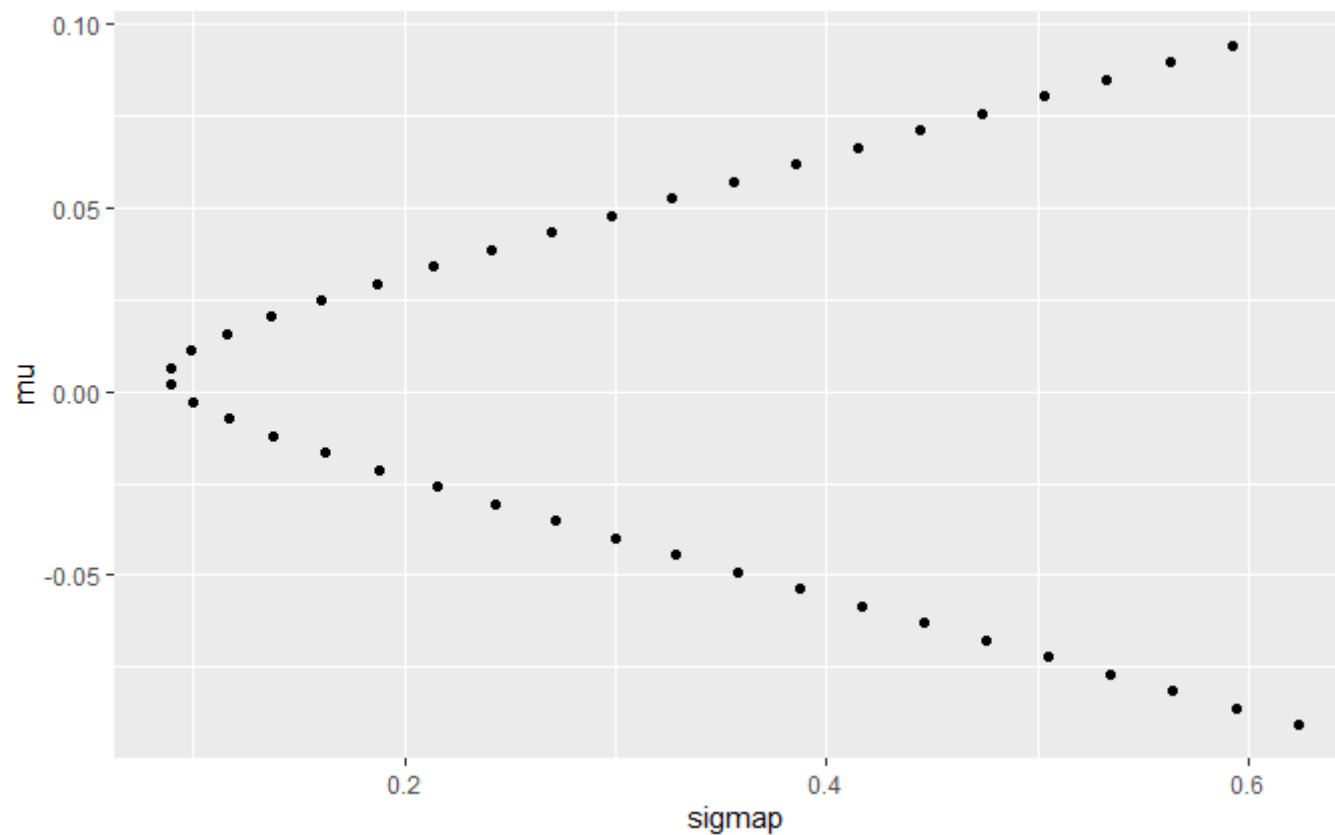
mu1 = t(x1)%*%returns
mu2 = t(x2)%*%returns

var1 = t(x1)%*%covariance%%x1
var2 = t(x2)%*%covariance%%x2
cov12 = t(x1)%*%covariance%%x2

wt = seq(from=-2,to=2,by=0.1)
mup = wt*(mu1[1,1])+(1-wt)*(mu2[1,1])
sigmap = (wt^2*(var1[1,1]) + (1-wt)^2*(var2[1,1])
          + 2*wt*(1-wt)*(cov12[1,1]))^0.5
```

4.2.2 绘制前沿组合

```
port = data.frame(sigmoid, mup)|  
require(ggplot2)  
p = ggplot(mapping=aes(x=sigmoid, y=mup), data=port)  
p + geom_point() + xlab('sigmoid') + ylab('mu')
```



4.3 卖空限制下的组合优化

- 随机抽样寻找前沿组合
- 基于solve.QP()函数
 - solve.QP()函数简介
 - 卖空限制下，整体最小方差
 - 卖空限制下，给定目标收益，整体最小方差
 - 卖空限制下，切点组合

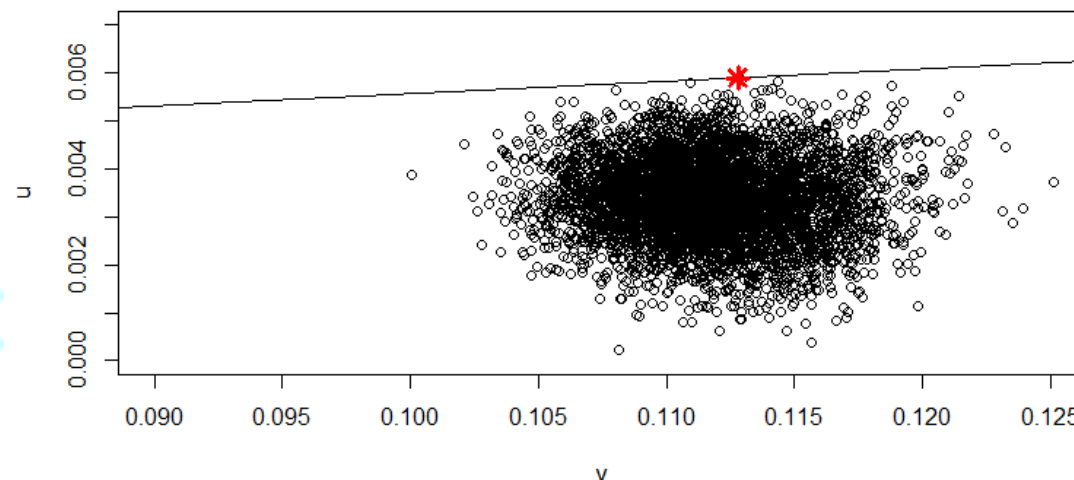
4.3.1 随机抽样计算组合前沿

- 随机生成组合收益与方差

```
#组合优化
N=5000
mat = matrix(runif(10*N,0,1),ncol=10)
weights = mat/apply(mat,1,sum)
u = (weights) %%% (aver)
v = sqrt(diag((weights) %%% v %%% t(weights))
plot(v,u,xlim=c(0.09,0.125),ylim=c(0,0.007))

#找切点
sharpe_ratio = (u-0.003)/v
s = sharpe_ratio[which(sharpe_ratio==max(sharpe_ratio))]
x = seq(0.0,0.13,0.0001)
y = s*x+0.003

lines(x,y)
n=which(sharpe_ratio==max(sharpe_ratio))
points(v[n],u[n],pch=8,col='red',cex=1.5,lwd=3)
```



4.3.2 solve.QP()函数

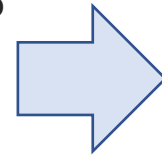
- 求解严格的凸二次规划问题
- `solve.QP(Dmat, dvec, Amat, bvec, meq)`
- 参数说明:
 - `Dmat`: 为Hessian矩阵, 此处指协方差矩阵
 - `dvec`: 为向量, 与`Dmat`一同构成目标函数
 - `Amat`: 为约束的系数矩阵, 默认约束为" \geq "
 - `bvec`: 为向量, 与`Amat`一同构成约束条件
 - `meq`: 表示从哪一行开始`Amat`中的约束为等式

4.3.3 实施算例

- 要求最小化 $Q(x, y) = \frac{1}{2}(x^2 + y^2) - 3x - y$

- 目标函数:

- $Q(x, y) = \frac{1}{2}(x \ y) \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - (3 \ 1) \begin{pmatrix} x \\ y \end{pmatrix} + 6$



- 约束条件:

- $x + y \geq 2$
 - $-x + 2y \geq 1$
 - $2x - y \geq -4$

- $Dmat = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, $dvec = (3 \ 1)$

- $Amat = \begin{pmatrix} 1 & 1 \\ -1 & 2 \\ 2 & -1 \end{pmatrix}$, $bvec = \begin{pmatrix} 2 \\ 1 \\ -4 \end{pmatrix}$

4.3.3.1 求组合最小方差

- 目标函数:

- 组合方差 $\sigma^2 = \omega^T \Omega \omega$
- Dmat= Ω , dvec=0

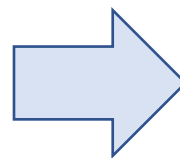
- 约束条件:

- 资产权重之和=1
- 资产权重 ≥ 0

- $$\text{Amat} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & 1 & 0 \\ 0 & 0 & \ddots \\ & \cdots & & 1 \end{bmatrix}, \text{ bvec} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

```
#卖空限制下组合优化|
require('quadprog')
Dmat = covariance
n.asset = nrow(Dmat)
dvec = rep(0,n.asset)
DIAG = diag(n.asset)
ones = rep(1,n.asset)
Amat = t(rbind(ones,DIAG))
bvec = c(1,rep(0,n.asset))
meq = 1

optim = solve.QP(Dmat,dvec,Amat,bvec,meq)
x.gmvp = zapsmall(optim$solution)
mu.gmvp = t(x.gmvp) %% returns
sigma.gmvp = t(x.gmvp) %% covariance %% (x.gmvp)
```



- $$\text{Amat} \times \omega \geq \text{bvec}$$
- $$\begin{cases} \omega_1 + \omega_2 + \cdots + \omega_n = 1 \\ \omega_1 \geq 0 \\ \omega_2 \geq 0 \\ \vdots \\ \omega_n \geq 0 \end{cases}$$

4.3.3.2 给定目标收益率

- 目标函数：组合方差
- 约束条件：
 - 资产权重之和=1
 - 资产加权收益=targetRet
 - 资产权重 ≥ 0

#卖空限制+给定目标收益率

```
targetRet = 0.006
```

```
mu = returns
```

```
Dmat = covariance
```

```
n.asset = nrow(Dmat)
```

```
dvec = rep(0,n.asset)
```

```
DIAG = diag(n.asset)
```

```
Amat = t(rbind(rep(1,n.asset),mu,DIAG))
```

```
bvec = c(1,targetRet,rep(0,n.asset))
```

```
meq = 2
```

```
optim = quadprog::solve.QP(Dmat,dvec,Amat,bvec,meq)
```

```
x2.gmvp = zapsmall(optim$solution)
```

```
mu2.gmvp = t(x.gmvp) %*% returns
```

```
sigma2.gmvp = sqrt(t(x.gmvp) %*% covariance %*% (x.gmvp))
```

4.3.3.3 求切点组合

- 目标函数：组合方差
- 约束条件：
 - 资产超额收益之和=1
 - 资产权重 ≥ 0
 - (有点问题)

```
# 利用solve.QP找切点组合
# 找到令夏普率最高的权重，但solve.QP()函数只能求最小值
# 控制超额收益不变，组合方差最小时夏普比率最大
rf = 0.0034
mu = returns
Dmat = covariance
n.asset = nrow(Dmat)
dvec = rep(0,n.asset)
DIAG = diag(n.asset)
Amat = t(rbind(mu-rf,DIAG))
bvec = c(1,rep(0,n.asset))
meq = 1

optim = quadprog::solve.QP(Dmat,dvec,Amat,bvec,meq)
x3.gmvp = optim$solution
#由于失去了权重之和为1的约束，最终结果需归一化
x.optim = zapsmall(x3.gmvp/sum(x3.gmvp))
mu.gmvp = t(x.optim) %*% returns
sigma.gmvp = sqrt(t(x.optim)%*%covariance%*%(x.optim))
```

案例5 R软件与资本资产定价模型

- R软件知识点：函数编写，if等控制结构，for等循环
- 金融理论知识：
 - 股票 β 系数的含义和估计方法
 - 资本资产定价模型和股票期望收益率的估计
 - 允许卖空和不允许卖空下组合优化策略

案例6 R软件与量化投资

- R软件知识点：函数编写，if等控制结构，for等循环
- 金融理论知识
 - 量化投资盈利的理论基础
 - 量化投资策略
- 简单多因子策略实现
 - 策略的基本思路
 - 数据的获取
 - 多因子选股的实现
 - 回测情况

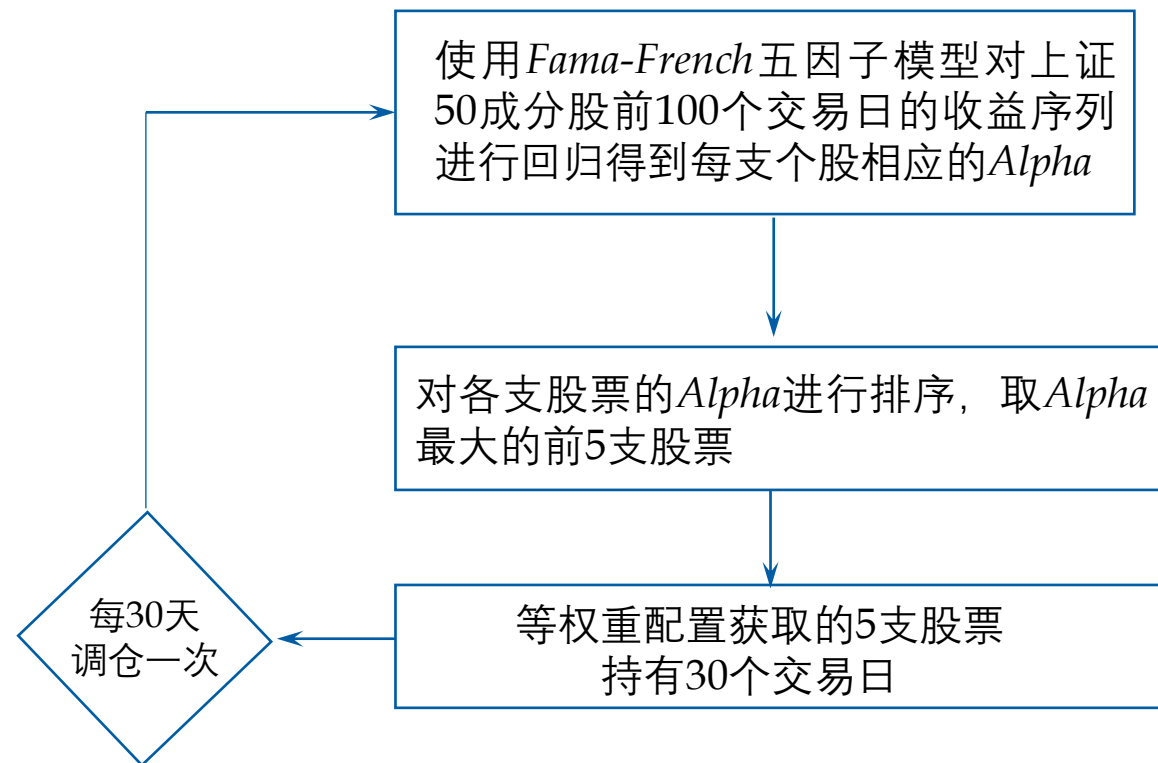
6.1 策略基本思路

- 标的股票池：上证50成分股
- 参考基准：上证50指数（SSE50）
- 样本期间：100交易日
- 调仓频率：30交易日
- Fama-French的五因子模型：

- $$R_t = \alpha + mMAR_t + sSMB_t + hHML_t$$

- $$+ rRMW + cCMA + e_t$$

- 市场因子、规模因子、账面市值比因子、盈利因子、投资因子



6.2 数据获取与基本处理

- Fama五因子数据获取
 - 国泰安数据导入
- 上证50成分股收益率数据获取
 - 长宽数据的转换

```
require(reshape2)
##转置成为更直观的格式、长宽数据转换
stkCls <- dcast(stk,Trddt~Stkcd,value.var = 'Clsprc')
```

- 对数收益率计算

```
names <- colnames(stkCls)
myfun <- function(x){x <- c(NA,diff(log(x)))}
stkRet <- as.data.frame(apply(stkCls[2:ncol(stkCls)],2,myfun))#每列求对数收益率
stkRet <- cbind(stkCls[,1],stkRet)#加上时间列
colnames(stkRet) <- names
```

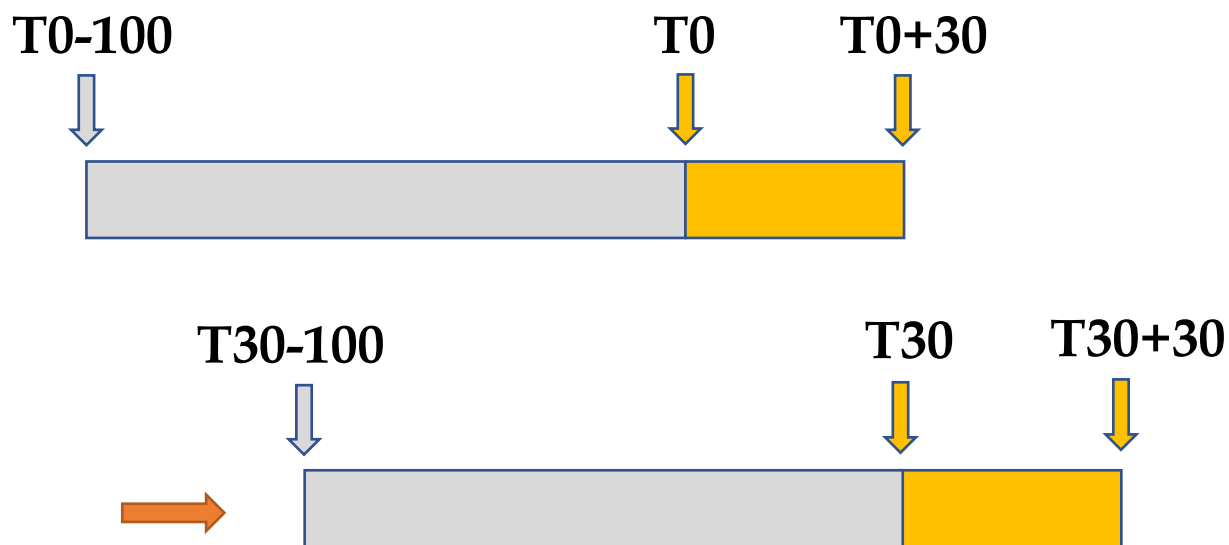
6.3 选股函数

- 使用Fama-French五因子模型对上证50成分股前100个交易日的收益序列进行回归得到相应的Alpha，对每只股票的Alpha排序，取出Alpha最大的五只股票进行配置
- $$R_t = \alpha + mMAR_t + sSMB_t + hHML_t + rRMW + cCMA + e_t$$

```
##使用因子模型选取股票的函数（放入sample直接得出配置的五只股票）
stk_sel <- function(dtSmp){
  sym <- rep(NA,100)
  alpha <- rep(NA,100)
  tgstk <- data.frame(sym,alpha)
  for(i in 7:(ncol(dtSmp)-1)){
    tra <- dtSmp[,c(1:6,i)]
    colnames(tra) <- c(names(Fac),'logRet')
    glm <- glm(formula = 'logRet~MAR+SMB+HML+RMW+CMA',data = tra)
    tgstk$sym[i] <- colnames(dtSmp)[i]
    tgstk$alpha[i] <- glm$coefficients[1]
  }
  tgstk<- na.omit(tgstk[order(-tgstk$alpha),])[1:5,]
  return(tgstk$sym)
}
```

6.4 调仓方式

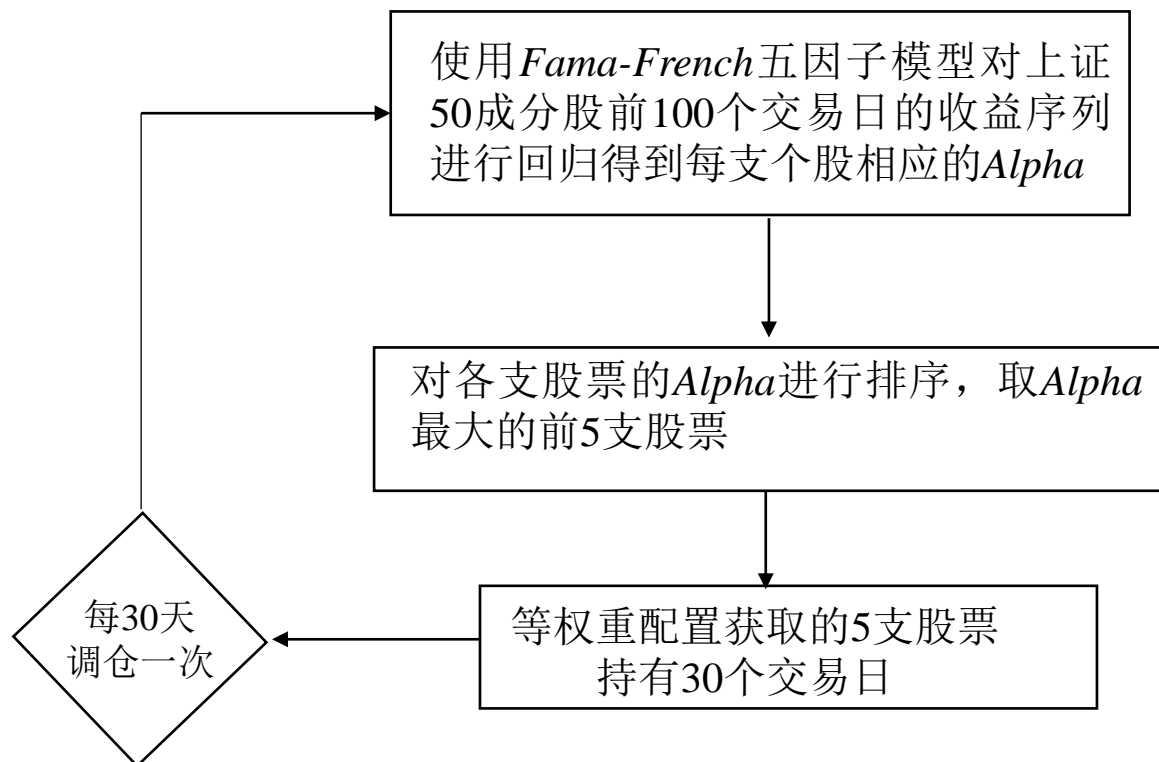
- 样本期间：100个交易日，确定接下来30个交易日要配置的5只股票
- 调仓频率：每隔30个交易日重新计算并调整股票池



6.5 回测结果



因子选股策略



- 标的股票池：上证50成分股
- 参考基准：上证50指数（*SSE50*）
- 样本期间：100交易日
- 调仓频率：30交易日
- **Fama-French的五因子模型：**
 - $R_t = \alpha + mMAR_t + sSMB_t + hHML_t + rRMW + cCMA + e_t$
 - 市场因子、规模因子、账面市值比因子、盈利因子、投资因子

心得与感悟

- 创建视频库

- 打通前置课程和后续课程的障碍
- 打通理论课程和试验课程的通道

- 提高实验课程与实践的融合度

- 真实数据
- 现实案例：产教融合
- 动态调整案例：对接企业需求
- 学生竞赛

- 激发学生自主学习

- 学生寻找教学案例
- 学生干中学

- 痛点

- 任课老师的知识结构问题
- 金融理论知识的难度取舍

金融学类专业前沿系列课程直播

真诚感谢
期待您的批评与指正



浙江工商大学 | 金融学院

方霞 | fangxia@zjsu.edu.cn

March 6, 2020