

《量化投资》课程导论

汪小圈

2025-02-24

关于任课老师

- 汪小圈
- 办公室: 松江校区博萃楼 103 室
- 电子邮箱: xqwang@suike.edu.cn
- 答疑安排:
 - 每周一 8:15-9:45; 13:00-14:30; 18:00-19:30
 - 每周二 8:30-11:30
 - 其他时间需发邮件预约

量化投资：定义与特点

- 定义：

- 量化投资是指以数据为基础，以数学模型为工具，以计算机程序为手段，系统化地执行投资策略，旨在获取稳定、可持续的超额收益的投资方法。
- 实践中，量化投资涵盖数据收集与清洗、策略模型开发、回测与优化、交易执行、风险管理等环节。

- 实践特点：

- 纪律性与客观性：模型信号驱动，克服情绪偏差，程序化交易，客观一致。
- 系统性与流程化：完整流程，规范明确，模块化 (策略、组合、风控、评估)。
- 程序化与可回测性：代码实现 (Python/R/C++)，历史回测，评估策略表现。
- 模型化与迭代优化：数学模型驱动，分析规律，迭代优化，适应市场。

- 量化研究员 (Quant Researcher):

- 职责: 策略研究与模型开发, 因子挖掘, 绩效分析等。
- 技能: 扎实的数理统计基础, 编程能力 (Python/R), 金融市场知识, 数据分析能力, 研究能力。

- 量化开发工程师 (Quant Developer):

- 职责: 策略代码实现, 交易系统开发与维护, 数据平台搭建等。
- 技能: 精通编程 (Python/C++/Java), 软件工程能力, 数据库知识, Linux 系统操作, 高频交易系统经验 (可选)。

- 量化交易员 (Quant Trader):

- 职责: 监控交易系统运行, 处理异常情况, 优化交易执行, 风险控制等。
- 技能: 熟悉交易流程, 风险管理意识, 快速反应能力, 一定的编程基础 (Python), 沟通协调能力。

- 课程名称：《量化投资》
- 教学对象：金融学院本科三年级学生
- 课时安排：16 周，每周 2 学时
- 模块划分：
 - 模块 1：传统因子投资 (9 周)
 - 模块 2：机器学习与另类数据 (7 周)
- 考核方式：策略开发项目 + 期末考试

模块 1：传统因子投资 (第 1-9 周)

- 目标：深入理解因子投资理论，掌握经典多因子模型构建与检验方法
- 核心内容：
 - CAPM 模型与市场有效性
 - Fama-French 三因子模型
 - 多因子模型与异象
 - 因子有效性检验
 - 多因子组合优化
 - 因子择时与动态配置
 - 风险模型与绩效归因

模块 2：机器学习与另类数据 (第 10-16 周)

- 目标：掌握机器学习在因子投资中的应用，融合另类数据提升策略收益
- 核心内容：
 - 监督学习与因子预测
 - 另类数据的获取与处理
 - 另类数据与新因子策略
 - 策略组合增强
 - 风险管理进阶
 - 项目迭代与答辩

考核方式与评分标准

① 策略开发项目 (50%):

• 第一部分 (中期答辩):

- 内容: 传统多因子策略开发 (至少 Fama-French 三因子模型)
- 要求: 代码可复现、因子逻辑清晰、回测报告完整

• 第二部分 (最终答辩):

- 内容: 融合传统因子、另类数据与机器学习模型
- 要求: 至少使用 1 种另类数据源或者 1 种机器学习模型

② 期末考试 (50%):

- 形式: 闭卷笔试 (简答 + 案例分析)
- 范围:
 - 传统因子投资理论
 - 机器学习与另类数据案例分析

策略开发项目要求

小组作业，每组 2-3 人

项目报告要求：

- 项目背景与目标
- 策略开发过程
- 回测结果与评估
- 风险管理措施
- 项目总结与反思
- 项目过程中的收获

项目代码提交要求：

- 包含完整的回测过程、回测评估、回测风险管理

- 学习通平台：
 - 理论课课件
 - 提交作业
 - 课程通知
- 实践课网站：
<https://e.coding.net/g-hcnn4565/lecture/quant.git>
- 课程参考书：《因子投资：方法与实践》



有效市场假说 (Efficient Market Hypothesis, EMH)

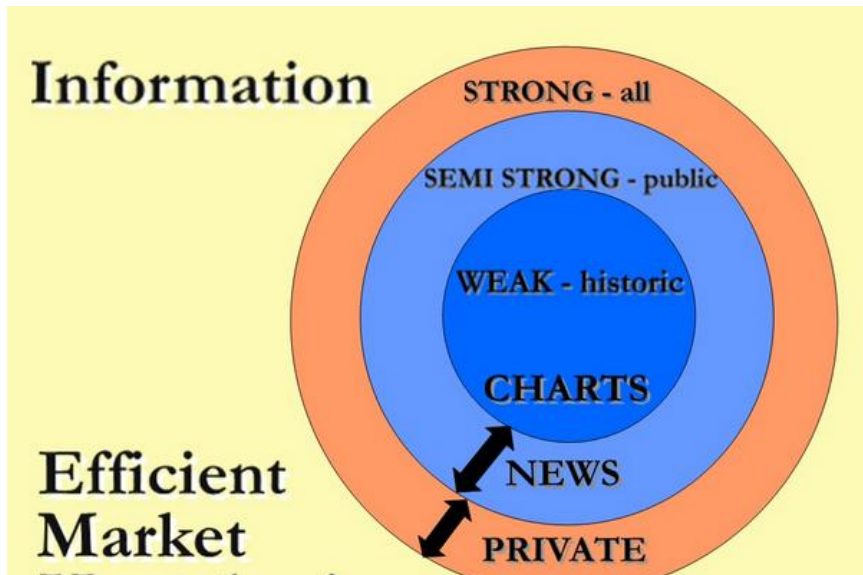
- **核心思想:**

- 在有效市场中, 资产价格充分反映了所有可获得的信息。
- 任何新信息都会迅速、准确地反映到资产价格中, 投资者无法利用公开信息或内幕信息获得超额收益。

- **三种形式:**

- **弱式有效市场:** 价格反映了所有历史市场信息 (价格、成交量)。技术分析无效。
- **半强式有效市场:** 价格反映了所有公开信息 (财务报告、新闻)。基本面分析无效。
- **强式有效市场:** 价格反映了所有信息 (公开 + 内幕信息)。任何分析都无效。

有效市场假说图示



如果市场完全有效，量化投资还有意义吗？

① 弱式有效市场下

- 基本面量化有意义，通过深入分析财务报表、行业数据、宏观经济数据等基本面信息，构建超额收益策略。
- 另类数据量化有意义，通过分析社交媒体、新闻、专利、地理信息等另类数据，构建超额收益策略。

② 半强式有效市场下

- 需要挖掘更深层次的信息或利用行为偏差。
- 行为金融学量化投资策略，通过识别和利用投资者非理性行为导致的市场异象来获取收益。
- 另类数据的重要性进一步提升，因为它们可能提供尚未被市场广泛认知的增量信息。

③ 即使市场是强有效的

- 量化投资仍然可以在风险管理、交易执行等方面发挥优势。

因子 (Factor) 的定义

- 因子定义：
 - 能够解释资产收益率横截面差异的共同特征或风险来源。
 - 可以理解为影响资产收益率的”驱动力”。
- 因子与选股：
 - 通过因子分析，识别影响股票收益的关键因素。
 - 基于因子暴露构建投资组合，获取因子溢价。
- 市场异象 (anomaly)：
 - 市场价格与因子模型预测结果不一致的现象。
 - 因子投资依赖的市场异象通常表现为因子溢价。

因子投资：发展历程

- **早期萌芽 (1950s-1960s):**
 - Markowitz 投资组合理论 (1952)
 - Sharpe CAPM 模型 (1964)
- **经典模型构建 (1970s-1990s):**
 - Fama-French 三因子模型 (1993)
 - Carhart 四因子模型 (1997)
- **因子爆炸与异象研究 (2000s-2010s):**
 - 大量新的因子被发现 (价值、动量、质量等)
 - 行为金融学解释市场异象
- **智能化因子投资 (2010s-至今):**
 - 机器学习在因子挖掘与预测中的应用
 - 另类数据驱动的新型因子策略

因子投资所依赖的“市场异象”是否与 EMH 矛盾?

- 矛盾之处:

- EMH 认为市场价格应充分反映所有信息, 不应存在可被持续利用的、明显的超额收益机会 (即“异象”)。
- 因子投资正是基于对历史数据的研究, 发现某些因子 (如价值、动量等) 在长期内能够带来显著的风险调整后收益, 这似乎与 EMH 的有效性相悖。

- 调和与解释:

- **市场并非完全有效**: 现实市场信息传播和价格调整存在时滞, 为因子投资提供了盈利空间。
- **风险溢价解释**: 因子溢价可能并非真正的“异象”, 而是对承担某种系统性风险的补偿。
- **行为金融学解释**: 市场异象也可能源于投资者的非理性行为, 例如过度自信、羊群效应、处置效应等。行为金融学认为, 投资者并非完全理性, 其行为偏差会导致市场价格出现可预测的模式, 从而产生可被量化策略利用的“异象”。

- 安装 Python 环境（3.8 及以上版本）
- 安装代码编辑器，推荐集成 AI 辅助编程功能的代码编辑器
- 安装量化投资相关库
 - numpy
 - pandas
 - matplotlib
 - scikit-learn
 - pyfolio
 - alphalens
 - cvxpy

- VSCode + GitHub Copilot / 通义灵码插件
 - GitHub Copilot 支持 Claude-3.7-Sonnet, 每月限 50 次免费问答, 申请免费学生账号可以无限使用
 - 通义灵码支持 Qwen2.5 / DeepSeek 大模型, 全免费
- Cursor
 - 支持 Claude-3.7-Sonnet / Gemini 2.0 / DeepSeek 等大模型, 试用期 (两周或前 500 条) 免费, 后续订阅 \$20/月
- Trae
 - 支持 Claude-3.7-Sonnet, 限时全免费
- 推荐注册 GitHub 账号 (校园 IP 可访问)

- Tushare
 - 免费，需要注册账号
 - 获取数据有积分要求 (https://tushare.pro/document/1?doc_id=108)
- CSMAR
 - 学校购买，直接学校 IP 访问
 - 学校账户不可 API 访问
- Wind
 - 学校购买，在 WIND 终端、或个人账号访问
 - 个人账户可 API 访问
- CNRDS
 - 学校购买，账号访问
 - 不可 API 访问