

# 量化投资

## 金融数据分析和Python应用

2025-02

南开大学金融学院

赵博

# 课程相关信息

- 主页: [abzhaobo.github.io/courses/quant\\_2025](https://abzhaobo.github.io/courses/quant_2025)

# 什么是量化投资？

- 狭义：利用计算机技术，借助统计、数学、机器学习等等一切可利用的工具来进行证券投资。
- 广义：把投资理念数字化、模型化、系统化
- 大部分值得钻研的学问，研究越深入，就越来越量化

# 学这门课有什么用？

- 这门课不一定能让你发财
- 能：
  - 理解现代金融理论在实践中的一些应用
  - 基本熟练地掌握 Python 数据分析，不再畏惧编程
  - 掌握一些统计、计量、机器学习的方法，检验投资理论
  - 更快、更准（未必）、更全面地（也未必）看到一些现象
  - 为今后的更进一步研究打下基础

# 一些可能的困惑

- 经济金融的学生一般需要研究生学位，有些工作却技术含量不高
- 需要很长时间的实习（半年甚至更长）
- 入学前：经济、金融赚钱，高大上。电视网络上首席经济学家侃侃而谈，没有什么他们不能分析的
- 入学后：宏微观、专业课学了有什么用？可以用作实际分析吗？首席经济学家们为什么不用模型？为什么不用计量经济学？仅仅是他们怕普通民众听不懂吗？
- 毕业论文：99% 的论文都用计量经济学看一个或者几个  $x$  是不是影响  $y$ 。宏微观理论在哪里？需不需要？
- 机器学习 tmd 又是什么？现在都搞这些东西，我要去学吗？到底怎么就业？拼得过那些搞 AI 的吗？

# 经济学（金融学）方法论简要讨论

- 观点1: 构建理论很容易
- 观点2: 能用科学方法检验的理论不多（科学还是玄学：为什么他们的结果你复现不出来？）
- 观点3: 经济学、金融学理论大部分很难用科学方法检验
  - 即便全面掌握了高深的计量经济学理论
  - 如果了解计量经济学的历史，就多少能理解为什么Prof. Angrist把代表作命名为“Mostly Harmless Econometrics”
- 观点4: 有些理论不需要进行所谓的“科学检验”

# 当你认为他说得对时你在经历什么？

- 1. 你用一整套（较为）科学的工具箱（RCT, IV, DID, RDD, etc.）检验过
- 2. 你凭借着经验、直觉、逻辑在判断

例：<https://fund.eastmoney.com/a/202402282997372976.html>

## 懵了！A股跳水原因是什么？16家基金火速解读

2024年02月28日 20:31 来源：中国基金报 编辑：东方财富网

分享到：

盘后，中国基金报记者采访了易方达、南方、博时、永赢、华宝、诺安、长城、华安、信达澳亚、创金合信、兴业、银河、长盛、[摩根士丹利](#)、恒生前海、金信等合计16家基金公司，对市场调整原因和后市走势进行解读。

受访公募普遍认为，今天A股的回调主要由消息面引起担忧情绪骤升所致，也与市场在持续上涨后部分资金获利了结有关。展望后市，公募对A股表现仍保持乐观。

市场机会方面，基金公司建议在进攻和防守端平衡配置，防守端关注低估值和高[分红](#)的优质标的，进攻端方面，持续发掘成长板块中有长期竞争力和[业绩](#)兑现度的企业。

### 消息面叠加资金面因素致市场调整

对于今天市场调整尤其微盘股大跌的原因，基金公司认为主要由消息面和资金面两方面的因素导致。

[永赢基金](#)认为，今日A股跌幅较大，主要由多方面因素共同驱动：一方面，部分媒体报道DMA业务，小市值及微盘风格风险偏好承压；另一方面，市场自2月6日底部持续反弹，部分抄底资金存在获利了结的可能。

[恒生前海基金](#)进一步分析称，短期抄底资金的获利了结压力是市场调整的主要诱因之一。自上证2635点以来，两市个股尤其是中小盘成长股，几乎均出现30%以上的反弹，部分个股甚至已经翻倍，对于抄底资金来说，短期收益丰厚，获利了结动力充足，同时在连续反弹后市场本身也需要消化短期浮筹。从今天的盘口来看，近期反弹幅度较大的个股纷纷进入跌幅榜也能清楚地看到这个信号。

[兴业基金](#)认为，市场下跌主要原因在于资金的止盈行为。自底部反弹以来，截至昨日收盘，[上证指数](#)实现10个交易日涨幅11.59%，资金止盈需求增加。今日A股早盘成交巨幅放量，全天两市成交额突破1.36万亿元，多空分歧扩大。

[银河基金](#)总结称，今日行情主要有两方面影响因素。市场反弹接近前高，今日A股早盘成交放量，全天两市成交额突破1.36万亿元，多空分歧加大。盘面上，微盘股与红利风格日内切换，资金再次转向红利风格相关板块。

短期来看，[南方基金](#)判断，本轮伴随着降息等利好政策催化，市场偏好快速好转，对应北向资金和融资买入占比在节后快速提升。当前市场情绪仍在修复途中，但考虑到前期快速的超跌反弹，未来一段时间更可能维持震荡格局，后续能否持续上行需要密切观测两会动态及节后复苏情况。

[创金合信基金](#)指出，中期看，市场上行的进程尽管有所反复，但趋势上仍可以保持积极。短期看，可以尽量避开面临抛盘风险的相关板块，寻找业绩端、政策端、主题等层面的结构性机会。

[金信基金](#)直言，应保持相对乐观的态度。首先，宏观经济已触底企稳，流动性保持宽松。此外，当前政策面较为积极，监管层对于维护市场和投资者利益的决心十分明确。近期重大的经济政策也连续出台，场内情绪已逐步回暖，市场在震荡消化前期涨幅后，行业板块和个股有望持续轮动表现。

[华宝基金](#)指数研发投资部总经理助理蒋俊阳表示，2024年A股市场开局较为波折，震荡幅度较大，但从经济运行的周期角度，以及全年资产配置的角度来看，他认为值得以更积极的态度面对市场并理性做好投资布局。

展望后市，[恒生前海基金](#)认为，市场大幅反弹后可能会面临一定调整压力，后续需进一步跟踪国内经济数据修复程度。但整体来看，中长期维度对后市仍保持乐观。

[博时基金](#)表示，当前美联储开启降息的时间点还具有不确定性，海外流动性也略向紧缩回摆，但美联储[货币](#)政策逐步趋于宽松的趋势不变，A股面临的流动性环境整体将更加友好。后续A股的走势将更多由国内因素决定，本周将要发布的2月[PMI](#)数据是市场关注的焦点。若经济修复能持续得到数据验证，A股将有望继续上行。

[长城基金](#)高级宏观策略研究员汪立提示，当下正处在两会前夕，按过往经验，两会期间A股市场往往不会有较差表现。



# 例：需求定律——价格与需求是反向关系

- 命题：研究生学费越来越贵，考研学生越来越多，是不是违反需求定律？
- 需求定律的推导、检验方式
  - 数学、非数学、计量检验

# 数学推导

- $\max_{x,y} U(x,y)$  subject to  $p_x x + p_y y = I$

$$\text{If } U(x,y) = x^\alpha y^\beta, x^* = \frac{I}{p_x \left(1 + \frac{\beta}{\alpha}\right)}, y^* = \frac{I}{p_y \left(1 + \frac{\alpha}{\beta}\right)}$$

# 非数学推导（反躬自省+逻辑推导）

- 假设 1: 主观价值论。边际效用是人选择商品的依据
- 假设 2: 人们需要多种东西。人们通过机会成本的变化来考虑消费量
  - 推论：边际效用递减。消费商品 A 后，人得到满足，人们希望消费一些其他的東西（这些其他的東西是继续消费商品 A 的机会成本），下一个商品 A 的边际效用减少
- 推论：需求曲线的斜率是负向的：当人的消费量增加时，愿意为这个商品付出的价格减少。当一个商品的价格减少时，愿意买更多这个商品，因为消费其他商品的机会成本变小（也即可以消费更多其他商品）

# 计量检验

$$\bullet \begin{cases} q_t^d = \alpha_0 + \alpha_1 p_t + u_t & (\text{需求}) \\ q_t^s = \beta_0 + \beta_1 p_t + \nu_t & (\text{供给}) \\ q_t^d = q_t^s & (\text{均衡}) \end{cases}$$

- 如果找了市场上的价格、销量数据，直接回归，得到的是供给还是需求曲线？
- 都不是。需要工具变量
- 如果找到了工具变量，得出需求曲线的回归系数是正的，怎么办？
  - 可能工具变量有问题（比如弱工具变量）
  - 可能数据有问题
  - 可能模型设定有问题
  - 但一般而言，没有人会认为需求定律错了
- 换句话说你通过反躬自省 + 逻辑推导得到的结论，无需经过检验！无论检验是什么结果，你只想要做出你想要的结果

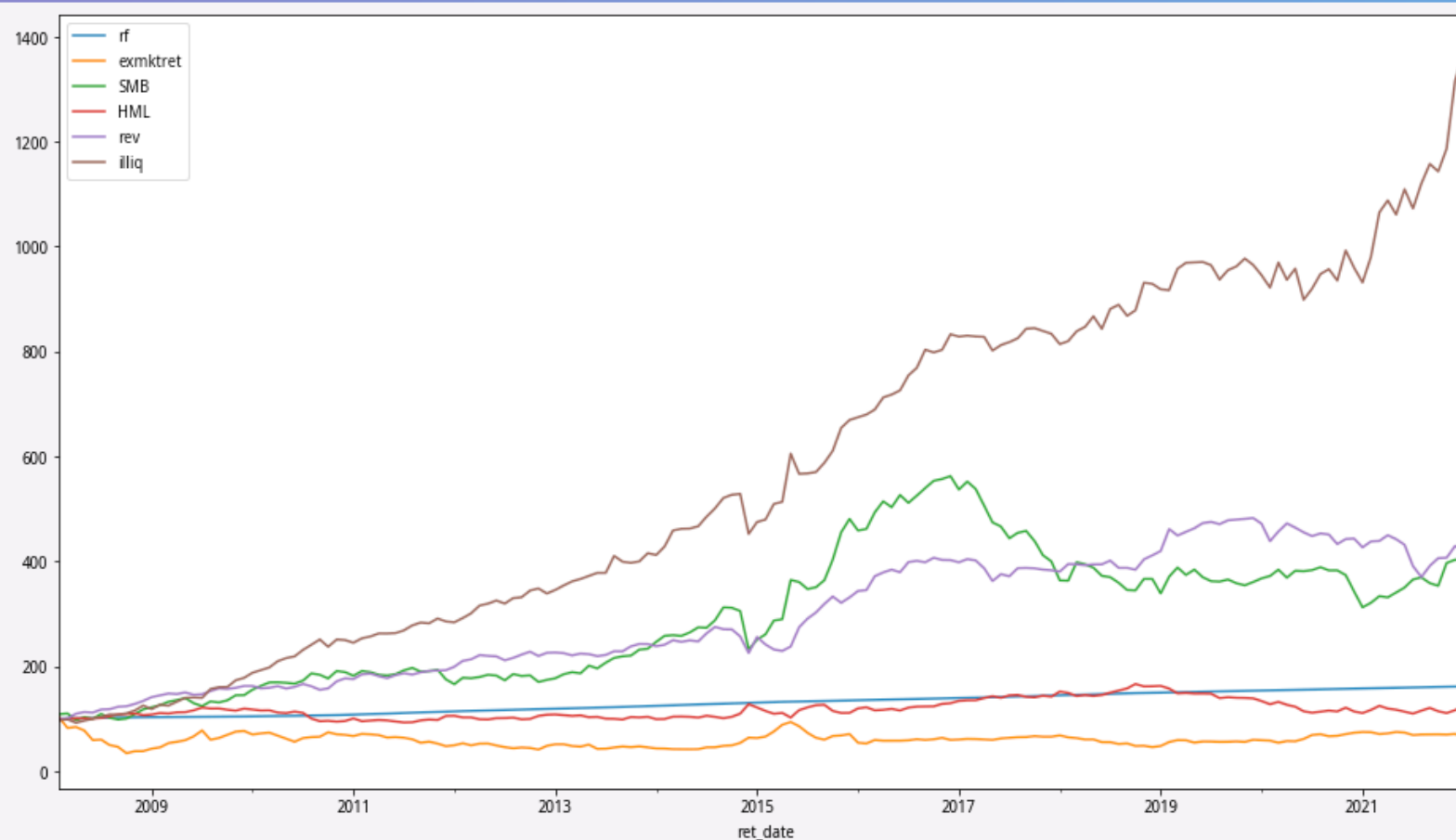
# 因此：

- 在经济学、金融学研究中，不要迷信数学、统计学、计量经济学、机器学习、深度学习.....
- 不要轻视你的直觉、经验、拍脑袋。也不要过度自信。
  - 大胆假设、小心求证
- 量化并不高高在上，但是投资是金融中相对可以进行客观评价的领域（尽管存在很大的运气成份）
- 掌握工具，不要被工具掌握

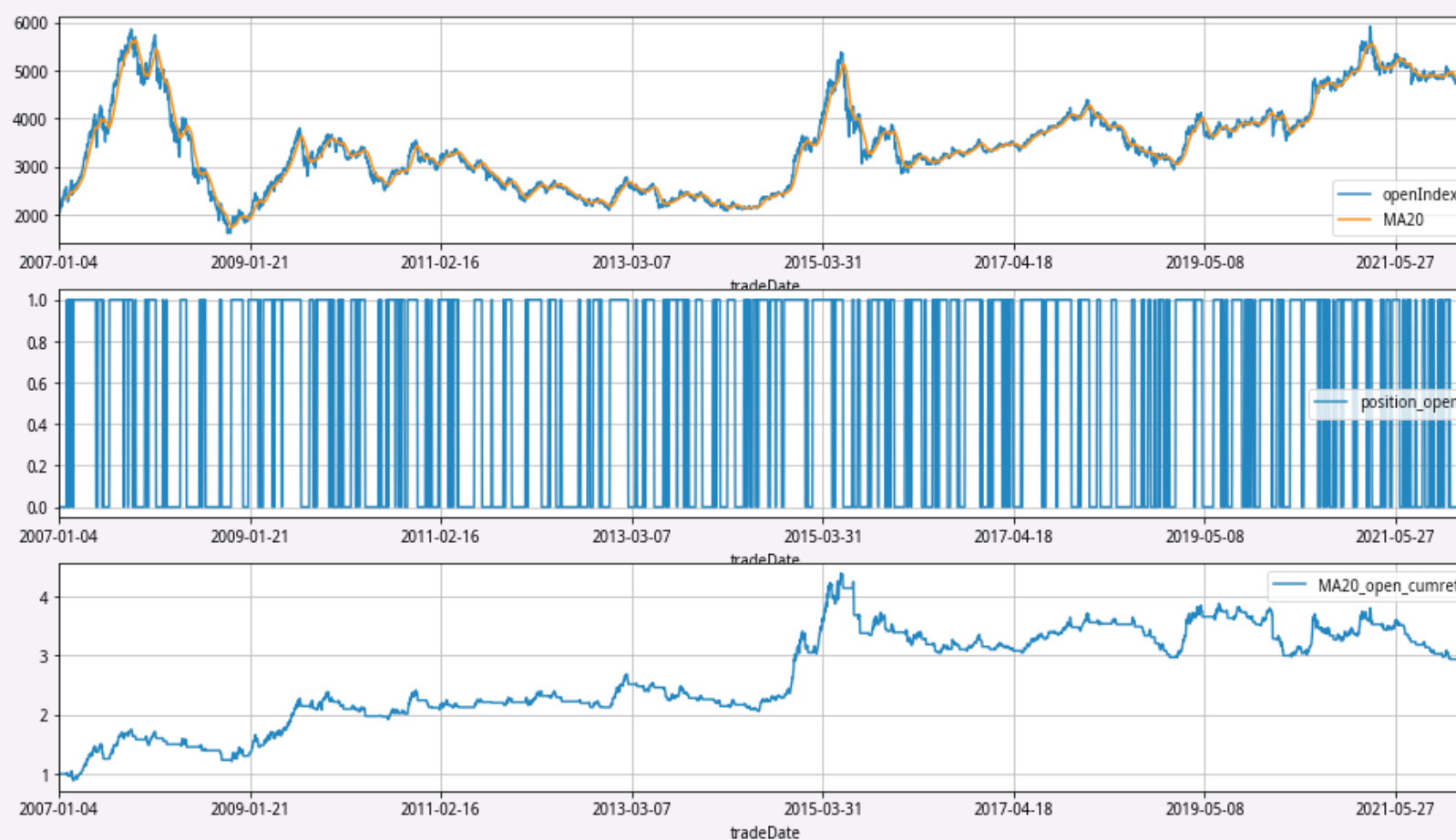
# 本课程主要内容

- 工具：
  - 统计学、计量经济学、机器学习
  - Python: Jupyter Notebook, Uqer量化平台, 量化相关的模块
  - AI类工具
- 思维框架：
  - 常识和经验、因子模型、技术分析、机器学习

- 因子选股：



- 技术指标择时：



# 量化的分类

- 技术分析、基本面量化.....
- 股票债券外汇商品、衍生品
- 高频（日内数据）、低频
- 综合



# 量化投资的典型流程

- 有一个想法:
  - 例如: 当股票价格高于近几日均值时, 说明主力开始启动. 此时可以买入. 当股票价格低于近几日均值时, 说明主力开始撤出, 此时应当及时卖出. (择时)
  - 前一个月波动率大的股票在下一个月通常表现比波动率小的股票好, 每个月初调仓, 买入上个月波动率大的股票 (选股)
  - 选股+择时

- 具体化:
  - 当股价高于20日均线5%时全仓买入, 低于20日均线5%时清仓
  - 每月计算股票日收盘价收益率的样本标准差, 排序, 月初第一个交易日全仓买入最高排名前5名的股票, 持有到月底。重复计算。
- 在历史中检验(回测)
- 在现实中检验(实盘检验)
- 交易

# 什么是量化平台?

- 量化平台完成了繁琐的细节工作, 让我们可以专注于投资策略的思考和编写
- 量化平台大致的架构
  - 研究模块: 导入数据, 构建模型
  - 回测模块: 后台提供了一套按时间运行的程序. 在每个时间段(也可以是每笔交易)上运行用户给定的交易算法, 考察收益.
- 常见的量化平台
  - zipline, vnpy, backtrader等
  - Quantopian、聚宽、米筐、优矿、万矿等
- 常见的数据来源
  - CRSP, Compustat, Tushare, 万得, CSMAR

# 如何做量化研究

- ~~一个常见的坑：花费大量时间在编程相关的细节上~~
- 做贴近实际的研究，积累经验、直觉
- 多阅读、多讨论、多动手
- 去交易

# 如何学习？

- “项目思维”取代“过程思维”
- 提问原则：为回答者考虑
  - 只有提问者尽量提供完整的和问题相关的信息，回答者才能帮上忙。
  - 没法回答的提问方式：“老师，我跑你给的代码老出错，是什么原因啊？”“老师，我安装Python总装不上，是什么原因啊？”
  - 正确的提问方式：
    - ▶ “我跑‘噼里啪啦’这段代码时，报告了‘bla bla bla’以下的错误。用搜索引擎搜索过，尝试了‘这个’、‘那个’解决方案，但仍然报‘bling bling bling’的错”
    - ▶ “我的操作系统是Win 10，用Anaconda安装Python 3.9时，报告以下‘bla bla bla’的错误。我已经尝试排除了‘这个’、‘那个’方案，但仍然不能解决问题”
  - 如果可能，应当提供Minimal Working Example (MWE)
- 用好搜索引擎、GPT等工具

# AI时代的学习

- 危：利用（公开）信息差的职业
  - 只会授鱼的教师
- 把握本质，掌握核心思路
- 今年的尝试：进一步减少代码的讲解；增加机器学习的内容

# François Chollet (Keras作者):

## CHAPTER 6 *Deep learning for text and sequences*

### **Markets and machine learning**

Some readers are bound to want to take the techniques we've introduced here and try them on the problem of forecasting the future price of securities on the stock market (or currency exchange rates, and so on). Markets have *very different statistical characteristics* than natural phenomena such as weather patterns. Trying to use machine learning to beat markets, when you only have access to publicly available data, is a difficult endeavor, and you're likely to waste your time and resources with nothing to show for it.

Always remember that when it comes to markets, past performance is *not* a good predictor of future returns—looking in the rear-view mirror is a bad way to drive. Machine learning, on the other hand, is applicable to datasets where the past *is* a good predictor of the future.

# Python以及其他工具简介

- 见notebook