



北京大学量化交易协会

量化投资简介和协会安排

2021年9月20日

孙博 余佳豪 叶梦婕 叶文轩

内培说明会

1 协会介绍

2 量化交易简介

3 本学期内培安排

4 内培分组

协会介绍

➤ 协会介绍

北京大学量化交易协会成立于2016年1月，创始人包括13、14、15级的多位量化领域的兴趣爱好者。协会以研究量化交易策略为核心，切合国内金融衍生品市场飞速发展的实际情况，为有志于将来从事量化类工作的会员们提供一个**学术交流、资源共享、人脉传递的专业性学术平台**，助力北京大学汇丰商学院推动量化交易领域的发展。

➤ 协会定位

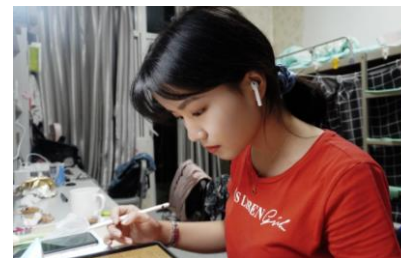
踏实：基础且全面的策略内培和实践项目

探索：了解学术前沿课题，引领业界发展潮流

硬核：极强的数理基础和编程能力



会长：孙博
2020级金融管理



学术副会：薛岚天
2020级数量金融



宣传副会：叶梦婕
2020级金融科技



宣传部长：余佳豪
2020级数量金融



学术部长：国欣然
2020级金融科技



学术部长：聂书涵
2020级数量金融

(AND 还有某位神秘校长)

➤ 量化精品内培（日常）

协会的核心活动，由四个小组轮流向其他成员讲解相关主题。内培全面覆盖量化投资领域最经典、最前沿的策略和技术，包括多因子，大类资产配置，机器学习，衍生品等。

➤ 量化策略实践（配合内培主题）

协会定期举行策略实证研究分享会。内容包括学习WIND数据库的使用，GitHub项目合作，使用python回测平台进行实证研究。成果将形成QTA系列研报永久留存。

➤ 论文研读小组（日常，学术副会组织）

成员轮流阅读量化经典文献或前沿课题，并撰写论文摘要。

➤ 公众号推送（日常，宣传副会组织）

记录协会的活动（内培、团建、分享会）并编辑相关推送。

➤ 量化求职分享会（每年春季开学）

邀请新入职量化岗位的师兄师姐分享实习经验，上届嘉宾分别来自：中信证券，交银施罗德基金，高盛集团，平安资管，WorldQuant，中信建投证券等知名公司。

内培说明会

1 协会介绍

2 量化交易简介

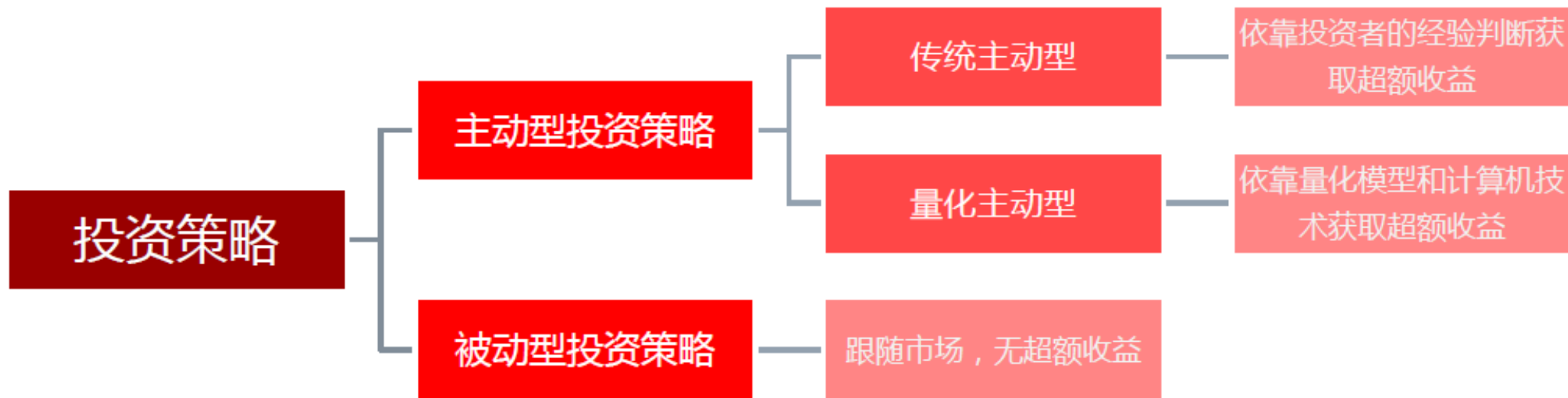
3 本学期内培安排

4 内培分组

量化交易简介

量化投资

- 定义：利用数量化统计分析工具构建相应的**数据模型**，从海量的数据中提炼出**具有投资价值的共性规律**，并借助计算机技术**实现投资思想和投资理念**的一种投资方式。



P quant 和 Q quant

- 量化世界中主要有两种主流派系，P宗与Q宗，其中P宗目标为量化风险与投资组合管理，主要任务为“预测未来”；Q宗目标为衍生品定价，更加专注于“外推现在”。

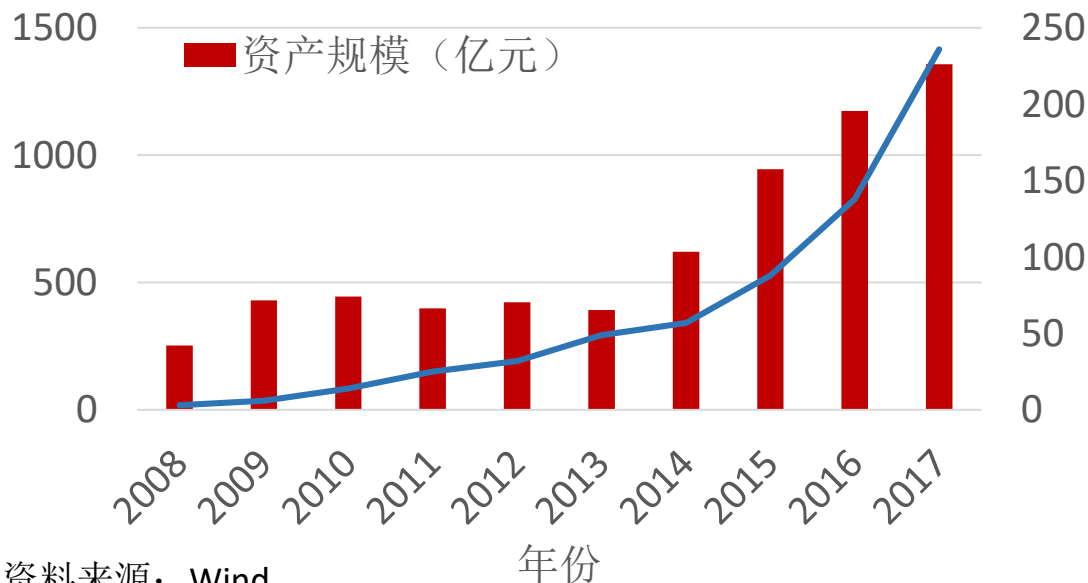
	P 宗	Q 宗
目标	预测未来	外推现在
使用测度	现实测度	风险中性测度
时间描述	离散时间序列	连续时间鞅过程
理论起源	资产资本定价模型CAPM 套利定价理论APT	布朗运动 Black-Scholes期权定价公式
主要应用	投资分析	资产定价
就业方向	公募、私募、券商金工组	券商衍生品部门、做市商

量化交易简介

国内现状

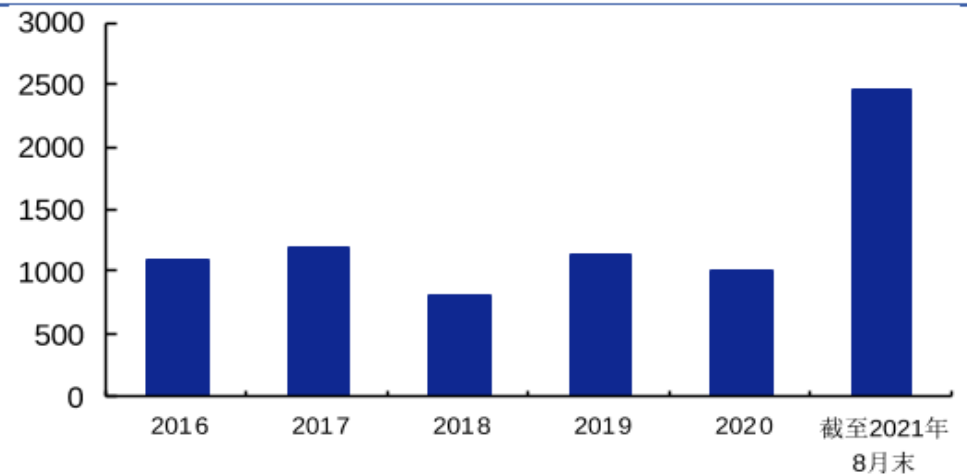
- 量化利用统计学、数学的方法，从海量历史数据中寻找能够带来超额收益的投资策略，主要的应用场景就是各公募或者私募的量化基金。
- 主动型、被动型以及对冲型量化基金占据了大量的市场份额，对量化型金融人才的需求也越来越大。除此之外，大量金融机构开始设立量化研究部门。
- 期权等金融衍生品在我国处于起步期，随着未来规模扩大需要更多的量化人才。

量化型公募基金



资料来源: Wind

图 10: 我国量化私募基金每年新成立数量



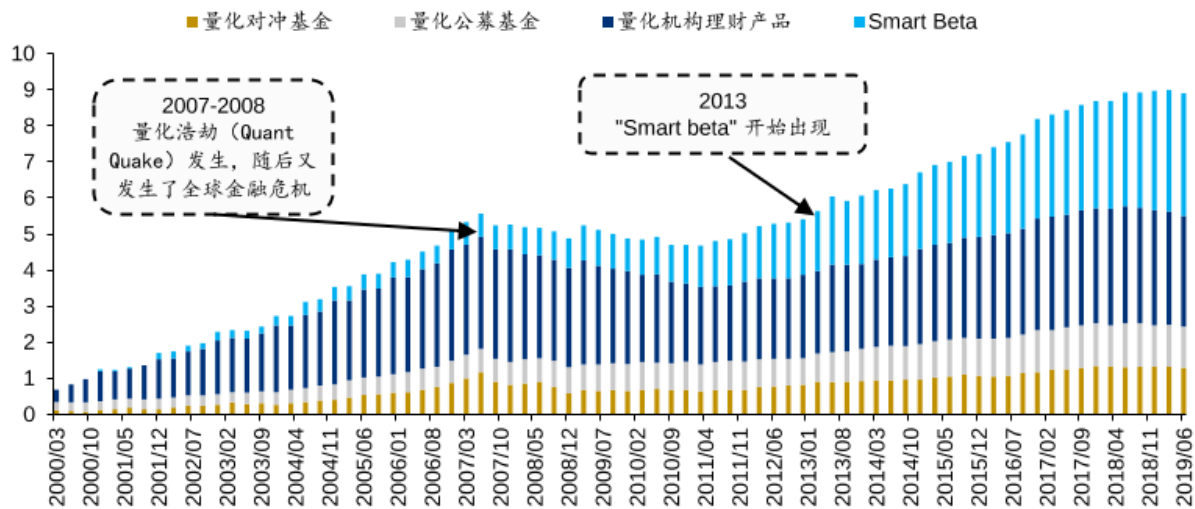
资料来源: Wind, 安信证券研究中心

量化交易简介

国内外对比

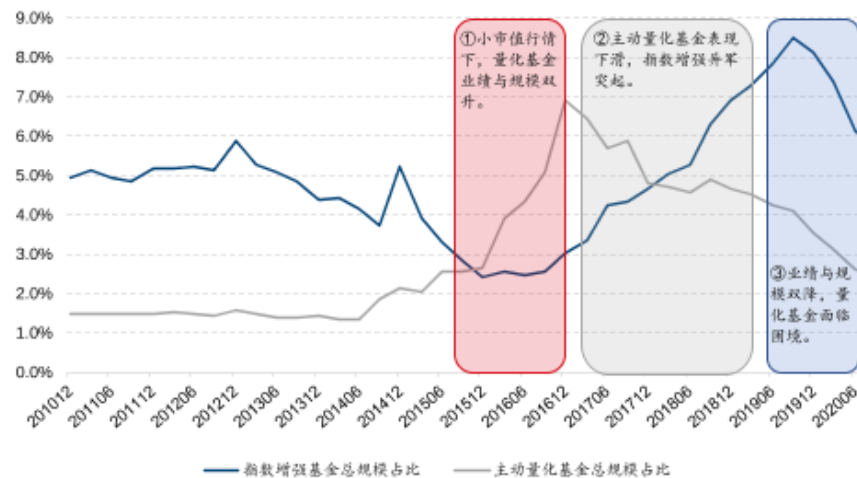
- 国内量化基金的规模相较国外还有很大的发展空间
- 即使在美国这样的发达股票市场，量化基金的规模仍在不断扩大中
- 在衍生品市场，国内和国外的差距更为巨大

图表6：美国各类量化基金管理规模占美股总市值的比重



资料来源: Bloomberg, 华泰研究

图 6：主动量化及指数增强型基金规模占比



资料来源: Wind, 国信证券经济研究所整理

量化交易简介

量化投资 vs 传统投资

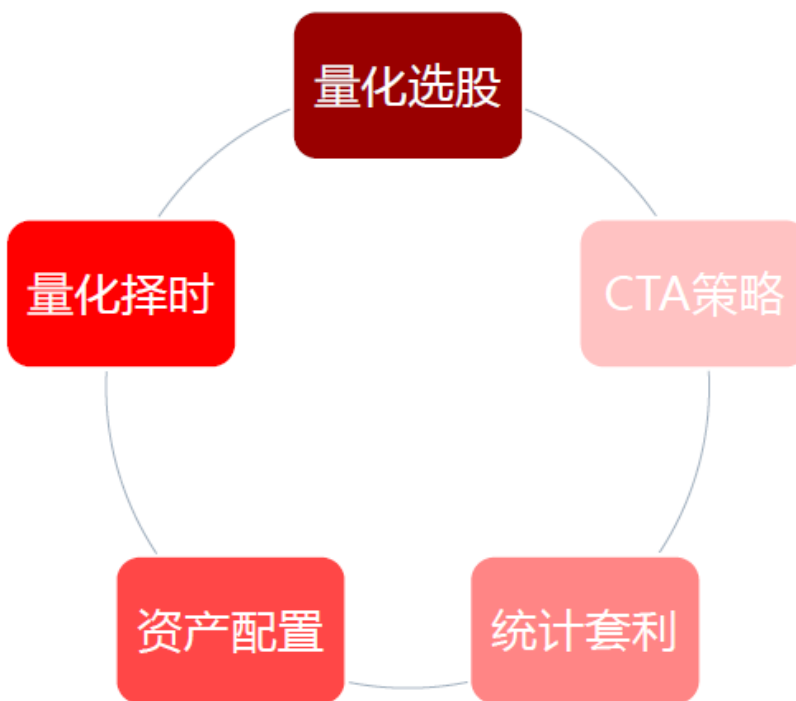
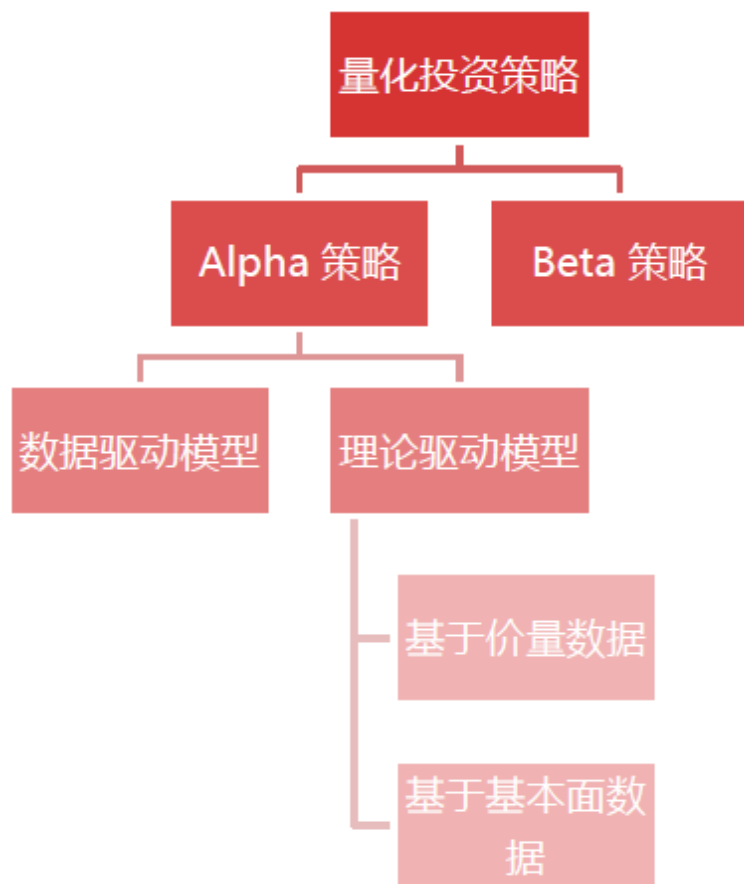
- 量化投资：寻找市场上一类资产的共同特征，依靠大数定律做出投资决策
- 传统投资：深入了解某个公司或者某个行业，根据个人经验做出投资决策

	定量分析	定性分析
决策依据	数学模型、计算机	个人经验
数据处理量	大	小
管理标的数量	分散	集中
分析深度	低->中	深入
收益来源	共性	特性
反应速度	慢	快

量化交易简介

量化投资策略分类

- 从目标来看，量化策略可以分成选股、择时、配置、套利和对冲五个部分
- 从盈利来源来看，量化策略可以分Alpha策略和Beta策略



内培说明会

1 协会介绍

2 量化交易简介

3 本学期内培安排

4 内培分组

■ 本学期安排

Module 1

- 9月20日（今天） 内培说明会
- 9月27日 学长学姐讲座（TBD）
- 9月29日（周三） 石川老师讲座
- 10月4日 国庆假期，休息
- 10月11日 多因子组内培
- 10月18日 大类资产配置组内培
- 10月25日 机器学习组内培
- 11月1日 衍生品组内培
- 11月8日 所有小组展示研报复现结果
- 11月15日-16日 期末考试，休息

Module 2

- 11月22日 学长学姐经验分享
- 11月29日 学长学姐讲座
- 12月6日 会长团选举与换届
- 12月13日 Smart Beta内培
- 12月20日 统计套利内培
- 12月27日 FOF内培
- 1月3日 元旦假期，休息
- 1月10日 高频交易内培
- 1月17日 所有小组展示研报复现结果
- 1月24日 寒假，休息

内培说明会

1 协会介绍

2 量化交易简介

3 本学期内培安排

4 内培分组

内培分组

➤ 多因子组

指导：叶梦婕
人数：6人



➤ 大类资产配置组

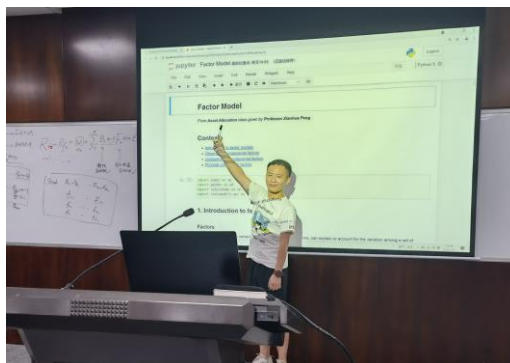
指导：孙博
人数：4人



多因子

➤ 机器学习组

指导：叶文轩
人数：5人



➤ 衍生品组

指导：余佳豪
人数：4人



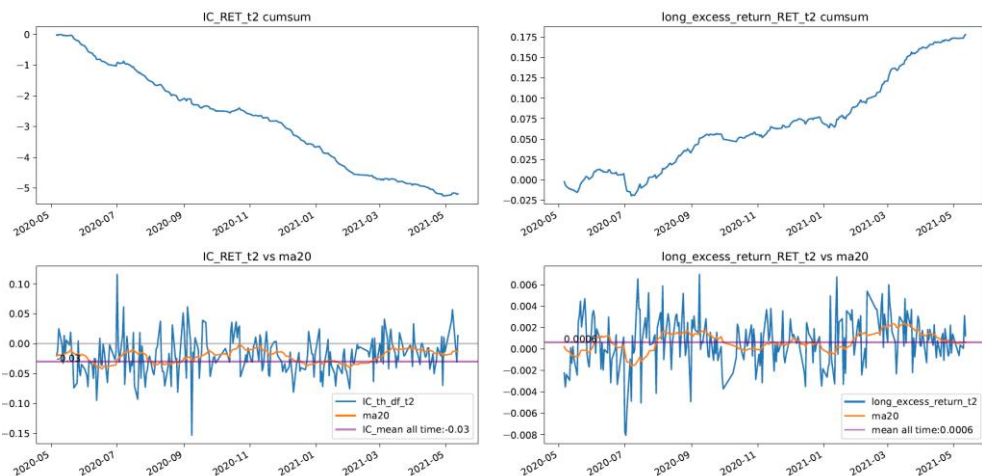
多因子

万物皆可多因子 $y = f(x_1, x_2, \dots, x_n)$

内容安排

- 多因子主题内培
 - 多因子的发展历程：CAPM, FF, Barra等等
 - 常见因子的初步了解：市值、成长、价值、量价等等
 - 常见名词的理解：风险暴露，风格偏好，行业暴露，中性策略等等
 - 因子评价的初步框架及其基本代码实现：脚本的形式即可
 - 发现自己感兴趣的方向：基本面or技术面，中低频or高频

- 因子挖掘项目
 - 基本面：戴维斯双击、PEAD、特定行业的基本面量化等等
 - 量价：日内高频价量数据的深度挖掘、市场微观结构探索
 - 另类数据：新闻舆情情感打分等
 - 构建完整的因子评价体系、因子筛选体系
 - 对市场风格、行业的理解，各种交易心理的探索
 - ML组的输入



多因子

拍了拍我的因子挖掘机

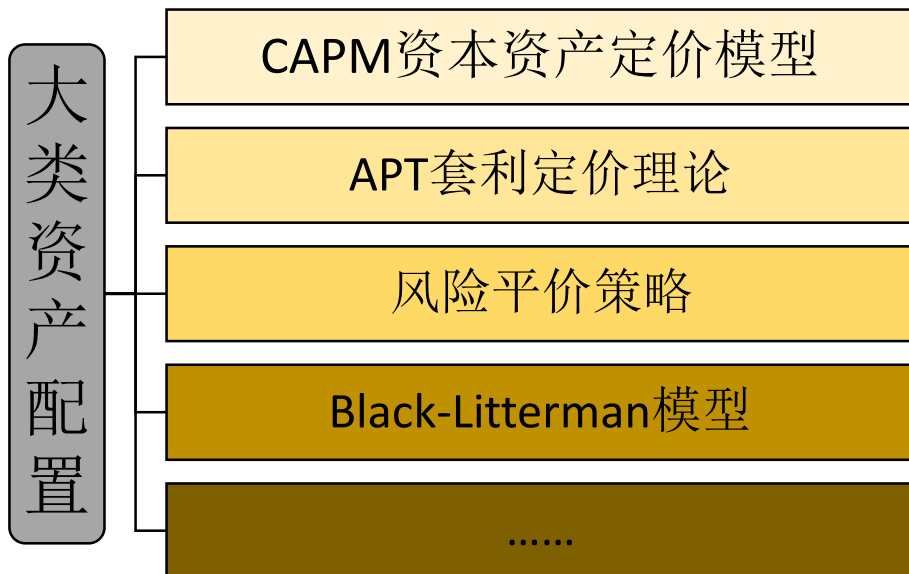
分组建议

- 需要编程，但不需要很厉害的编程
- 更重要的是对市场的理解，宏观、中观、微观结构
- 建议刚接触量化且希望向p quant发展的同学选择多因子组
- 初期读研报、川总讲量化、因子投资等参考资料打好基础
- 后期通过因子挖掘项目逐步上手、加深理解
- 欢迎定期meeting探讨

大类资产配置

大类资产配置

- 从金融的角度来讲，**大类资产**一般可以分为股票、债券、货币、商品、房地产等一些基础资产。
- **大类资产配置**就是指一个投资者怎么把自己的这种流动性的资产配置到各类资产上去，既能够保持一个比较稳定的收益，又能够规避大的风险，特别是在经济和金融周期在不断变化的情况下，也要根据周期的变化来调整自己各种资产的类别，各种资产的比例来达到一个收益和风险的平衡。
- 大类资产配置策略的三步骤：估计资产的预期收益率和协方差矩阵；设定最优化目标；求解最优配置



最优化目标：

$$\begin{aligned} \max \quad & \omega^T \Sigma \omega \\ \text{s. t.} \quad & \sum_{i=1}^n \omega_i = 1 \\ & \omega^T R_p = \alpha \end{aligned}$$

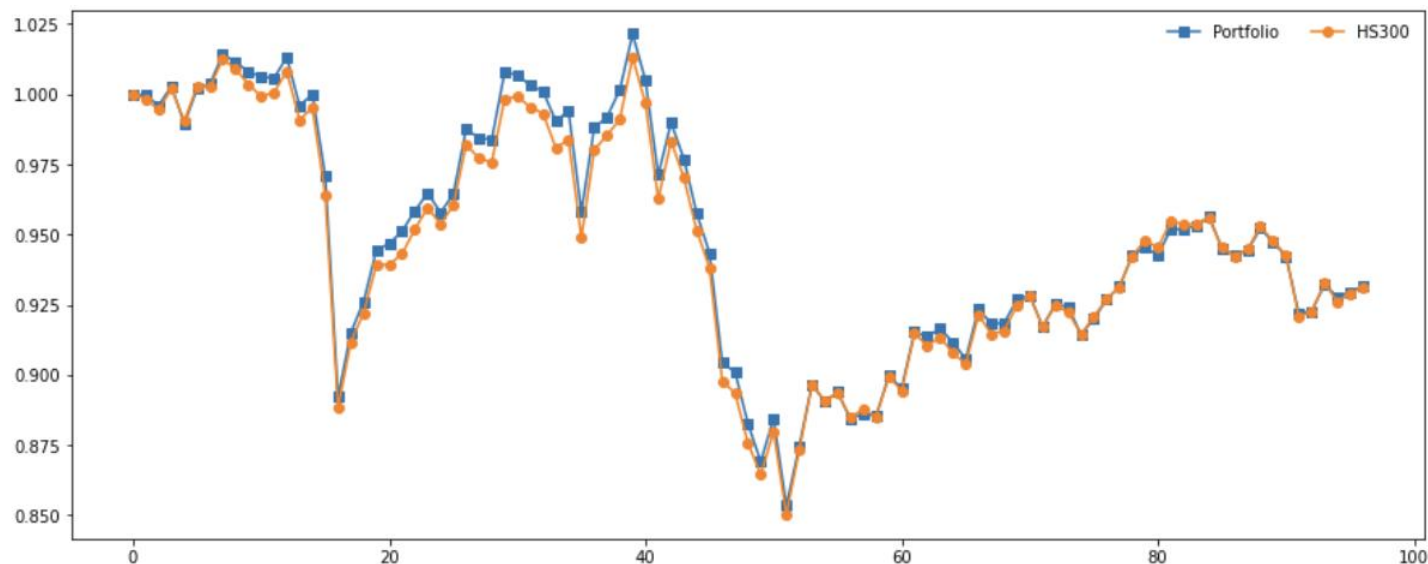
大类资产配置

内容安排

- 学习资料：彭献华老师Asset Allocation课件
- 学习目标：了解主流大类资产配置策略，能够使用Python相关Package求解最优化配置问题。
- 学习形式：自学；定期小组讨论；彭老师在线答疑指导；Coffee & Coding
- 项目展示：使用各类经典大类资产配置策略，创建自己的**A股指数增强组合**，跑赢指数！

去年指数跟踪项目成果：

$$\begin{aligned}
 & \max_{y_j, x_{ij}} \sum_{i=1}^N \sum_{j=1}^N \rho_{ij} x_{ij} \\
 & \text{s.t.} \quad \sum_{i=1}^N y_i = n \\
 & \quad \sum_{j=1}^N x_{ij} = 1 \\
 & \quad x_{ij} \leq y_j \quad \forall i, j \\
 & \quad x_{ij}, y_j \in \{0, 1\} \quad \forall i, j
 \end{aligned}$$



大类资产配置

分组建议

- 大类资产配置组适合：
 - 对数学或运筹学感兴趣的同学；
 - 对金融知识了解不多的同学（大类资产配置不那么需要市场经验，可以以此入门）；
 - 知识体系偏向理论研究，或者未来想在学术界发展的同学；
 - 希望抱彭老师大腿的同学（学完可以给彭老师做TA）；
 - 当然，以上都是充分条件，不是必要条件。

Let's make some interesting stuff!

基于基金“抱团”持股的行业配置策略

Mean-CVaR 大类资产配置框架与实战

大小盘轮动的奥义

.....

机器学习——项目一：Kaggle比赛总结

目的

- 学会从数据科学的角度如何做金融数据建模，熟悉数据处理、机器学习所用到的编程工具

要点

- 描述这是一个怎么样的比赛，问题的背景是什么，数据有哪些，评价指标是什么，如何提交代码和结果。
- 研究Private Board Top10 队伍的方案(如果有公布)，和【code】、【discussion】板块中高质量的内容。分析他们如何做预处理，他们在EDA(Exploratory Data Analysis)和查阅金融资料的过程中得到了什么启发，他们怎么基于这些启发去做特征工程和模型选择，模型的训练目标怎么设计，用了什么方法防止过拟合，用了什么方法去做模型融合(Ensemble)，前三名相比第八九十名的方案表现好，是运气还是实力，这个比赛中哪些模型表现比较好，为什么。另外，他们数据作图分析做得特别好的、对代码运算耗时做了神奇的优化的，也可以总结。
- *从Top10的方案整合出一套自己的方案，不断改进，尝试打败他们。



Featured Code Competition

Optiver Realized Volatility Prediction

Apply your data science skills to make financial markets better

\$100,000
Prize Money

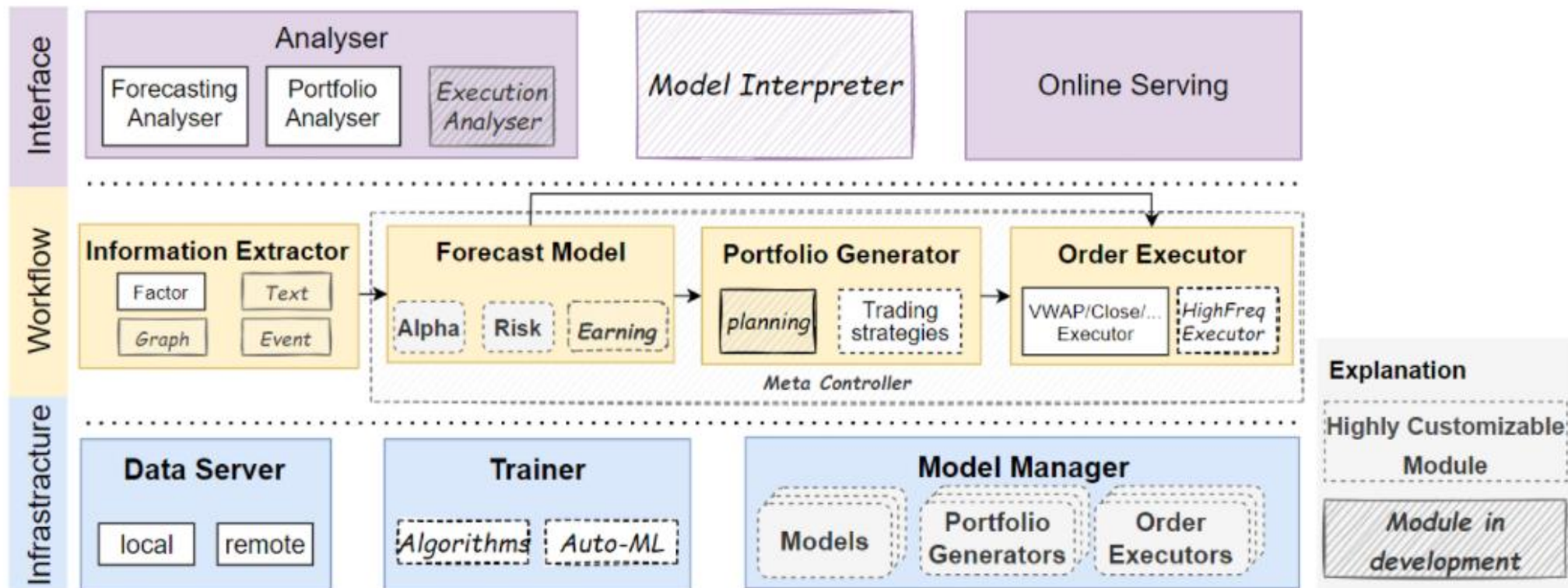
Optiver · 3,689 teams · 8 days to go (21 hours to go until merger deadline)

$\sigma^2 = \frac{1}{T} \sum_{t=1}^T \ln \left(\frac{S_t}{S_{t-1}} \right)^2$

机器学习——项目二：机器学习选股策略

Qlib

- Qlib is an AI-oriented quantitative investment platform, which aims to realize the potential, empower the research, and create the value of AI technologies in quantitative investment.
- It contains the full ML pipeline of data processing, model training, back-testing; and covers the entire chain of quantitative investment: alpha seeking, risk modeling, portfolio optimization, and order execution.



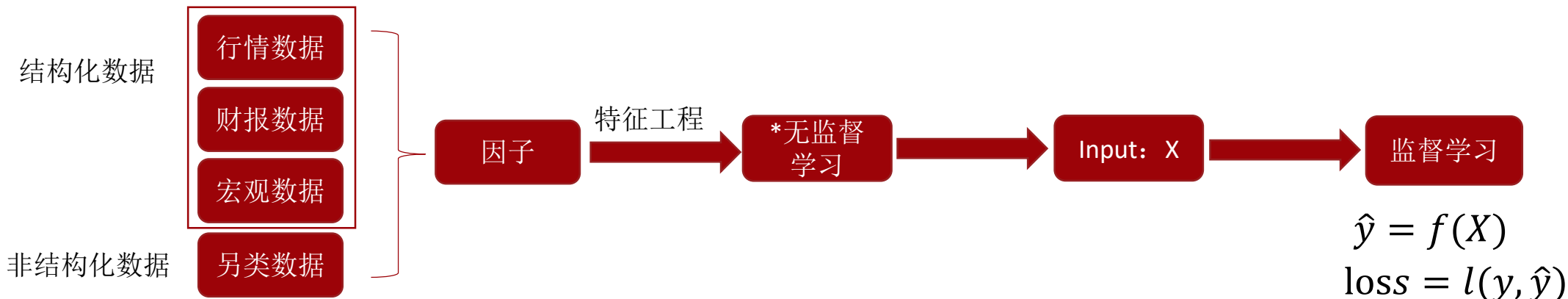
机器学习——项目二：机器学习选股策略

目的

- 学会处理A股数据，熟悉选股策略的整体流程，掌握机器学习的应用，认识机器学习的边界

要点

- 跑通官方Demo，熟悉涉及到的代码
- 介绍Qlib各模块，尤其是数据存储的方案。分析合理与不合理之处。
- 跑通我们自己的数据，模型的输入可以来自于多因子组挖掘出的因子，也可以从EDA与金融理论出发自己设计特征或者设计模型从原始量价数据端到端地学习。
- *不同的loss、不同的打标签方式对结果的影响。
- *模型可解释性分析：模型学到了什么？是否有对应的经济学解释？随时间推移、随市场变化，模型学到的关系是否鲁棒？对之后的模型改进有什么启发？
- *其它任何深入的研究。我们希望研究不止于拿一堆机器学习模型来试，试出表现最好的模型并安个理由。
- ps：我们与多因子组在数据预处理和回测上有交集，可能会一起进行部分模块的开发



机器学习——补充说明

Q&A

Q: 编程与机器学习基础薄弱，可以加入吗？

A: 可以，但是不会手把手教，而是提供有用的资料辅助你们自学，并且解答问题。另外建议机器学习组的成员在写代码时【结对编程】，即两个程序员在一个计算机上共同工作。一个人输入代码，而另一个人审查他输入的每一行代码。两个程序员经常互换角色。这样组员们可以更高效地各取所长，因此某些方面基础差的同学不必过于担心跟不上节奏。

Q: 学姐学长们的角色是怎么样？

A: 提供大方向，提供数据与资料，分享经验，必要时一起写代码，
非常欢迎各位（跨组亦可）随时提问随时骚扰，但希望是搜索引擎难以搜到的问题。

Q: 我电脑可能跑不动模型

A: 这个不必担心

Q: 在项目正式开始前，我空闲时应该作什么准备？

A: 大体是学好pandas、scikit-learn和pytorch，机器学习基础差的先看李宏毅的《机器学习2021》（B站有）。

anything else



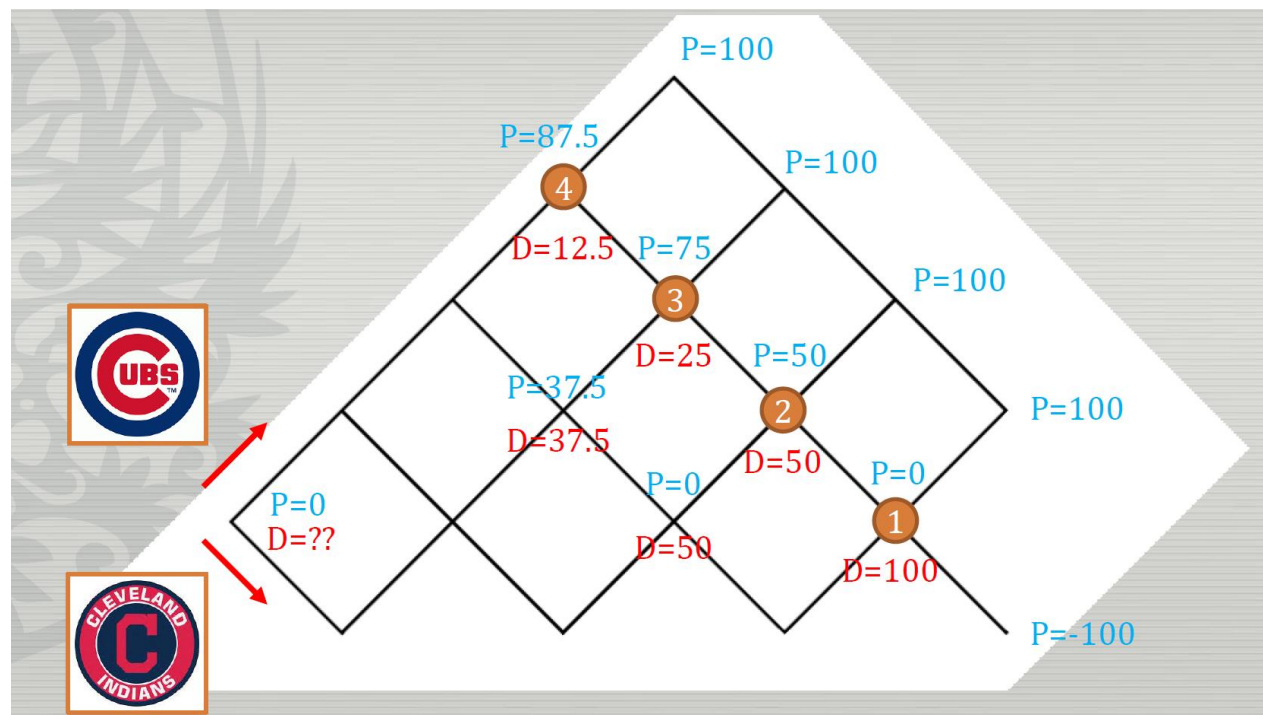
衍生品

内容安排

- 学习国内外的主流衍生品种类和对应市场
- 从Bachelier Model到BSM公式，了解衍生品定价的逻辑
- 尝试定价当前市场最火爆的衍生品：雪球期权！

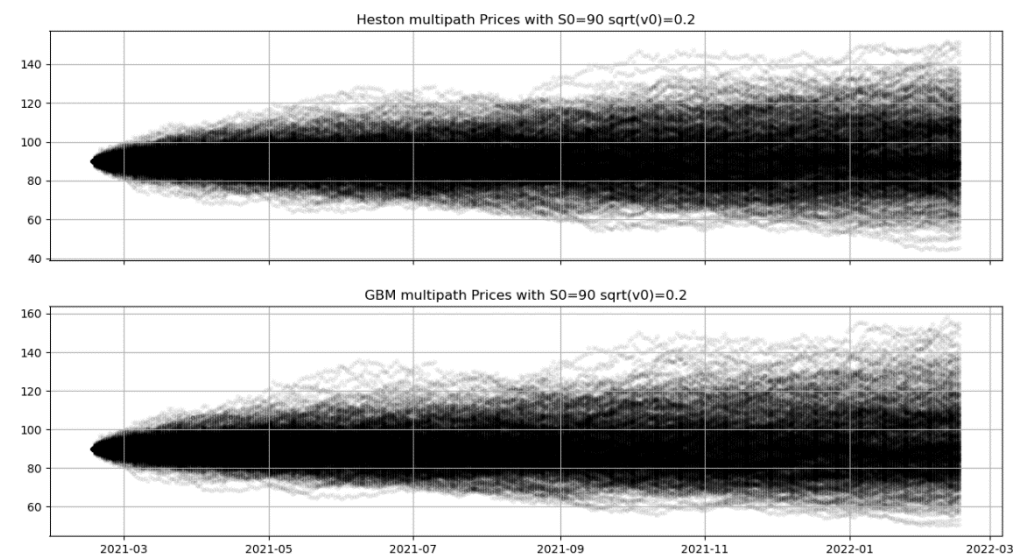
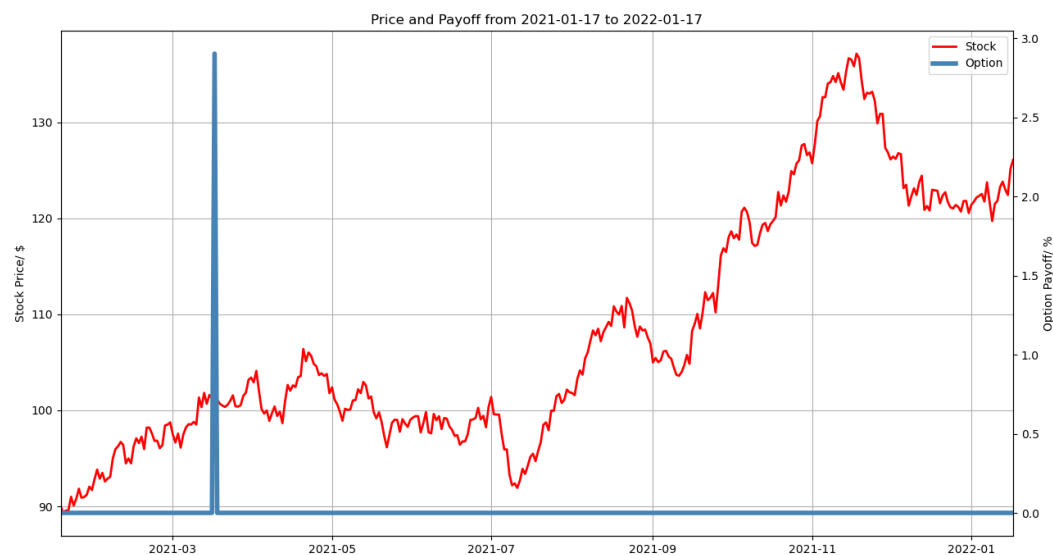
分组建议

- 需要比较扎实的数理统计基础
- 适合数金专业的同学（和必修课Stochastic Finance有很多重叠）
- 合适想往Q quant方向发展的同学
- 不需要很强的编程基础
- 梦想成为赌王的同学考虑一下（体验靠计算赢别人的快感）



随机运动

- Brownian Motion
- Stochastic Volatility Motion



期权期货定义

- 期权（Option）：双方关于未来买卖权利达成的合约，期权的买方（权利方）通过向卖方（义务方）支付一定的费用（权利金），获得一种权利，即有权在约定的时间以约定的价格向期权卖方买入或卖出约定数量的特定资产。当然，买方（权利方）也可以选择放弃行使权利。如果买方决定行使权利，卖方就有义务配合。期权的标的资产是指选择购买或出售的资产。它包括股票、政府债券、货币、股票指数、商品期货等。
- 远期合约（Forward）是指双方同意在未来日期按照固定价格交换金融资产的合约，远期合约是必须履行的协议。期货（Future）是指以约定的时间和价格，买卖某种大众产品如棉花、大豆、石油等及金融资产如股票、债券等的可交易合约。
- 远期合约与期货的区别：前者合约条件是为买卖双方量身定制的，通过场外交易(OTC)达成，而后者则是在交易所买卖的标准化合约。

术语

- 看涨期权（call option）：给期权持有人在将来某个日期以一定价格买入某资产的权利
- 看跌期权（put option）：给期权持有人在将来某个日期以一定价格卖出某资产的权利
- 到期日T（maturity date）：合约中注明的日期
- 执行价格K（strike price）：合约中注明的价格
- 期权头寸：买入期权方为多头（long），卖出期权方为空头（short）。卖出期权方在最初收取期权费用，但该方在今后有潜在的义务，其盈亏曲线与权利方相反。
- 欧式期权（European option）：在到期日方可执行的期权
- 美式期权（American option）：在到期日之前的任何时可均可执行的期权

随机波动率模型

- SABR model [Hagan et al., 2002]:

$$d\sigma_t = \nu \sigma_t dZ_t.$$

- Heston [1993] model (CIR process):

$$dv_t = \kappa(\theta - v_t)dt + \nu\sqrt{v_t}dZ_t.$$

- 3/2 model [Heston, 1997, Lewis, 2000]:

$$dv_t = \kappa v_t(\theta - v_t)dt + \nu v_t^{3/2} dZ_t.$$

- GARCH diffusion model (relatively new):

$$dv_t = \kappa(\theta - v_t)dt + \nu v_t dZ_t.$$

- OU volatility model [Li and Wu, 2019]:

$$d\sigma_t = \kappa(\theta - \sigma_t)dt + \nu dZ_t.$$



北京大学量化交易协会

谢谢大家！