

# 资本资产套利定价模型

华南理工大学数学学院

刘深泉教授

# 目录

- 1. Markowitz投资组合理论
- 2. Sharpe资本资产定价模型
- 3. Ross资本套利定价模型
- 4. Fama多因子模型
- 5. 量化投资和期货对冲策略
- 6. Matlab 量化投资工具箱
- 7. 统计套利策略和高频交易潜力
- 8. ARCH模型和GARCH模型
- 9. 深度学习和大数据挖掘
- 10. 金融数学建模-经济、数学、计算和深度学习

# 1. Markowitz投资组合理论

- 马克维茨投资组合理论的基本假设为：
- (1) 投资者是风险规避的，追求期望效用最大化；
- (2) 投资者根据收益率的期望值与方差来选择投资组合；
- (3) 所有投资者处于同一单期投资期。
- 马克维茨提出了以期望收益及其方差 ( $E, \delta^2$ ) 确定有效投资组合。
- 以期望收益  $E$  来衡量证券收益，以收益的方差  $\delta^2$  表示投资风险。
- 资产组合的总收益用各个资产预期收益的加权平均值表示，组合资产的风险用收益的方差或标准差表示，则马克维茨优化模型如下：

$$\min \delta^2(r_p) = \sum \sum w_i w_j \text{cov}(r_i, r_j)$$

$$E(r_p) = \sum w_i r_i$$

式中：  $r_p$  —— 组合收益；

$r_i, r_j$  —— 第  $i$  种、第  $j$  种资产的收益；

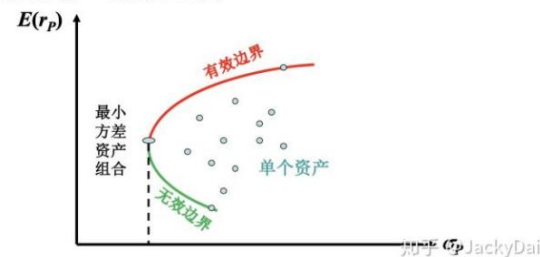
$w_i, w_j$  —— 资产  $i$  和资产  $j$  在组合中的权重；

$\delta^2(r_p)$  —— 组合收益的方差即组合的总体风险；

$\text{cov}(r_i, r_j)$  —— 两种资产之间的协方差。

## 多种风险资产的组合

### ◆ 风险资产有效边界



其中上半段曲线被称为有效前沿 (Efficient Frontier)，曲线内的点被称为可行集。

## 2. Sharpe资本资产定价模型

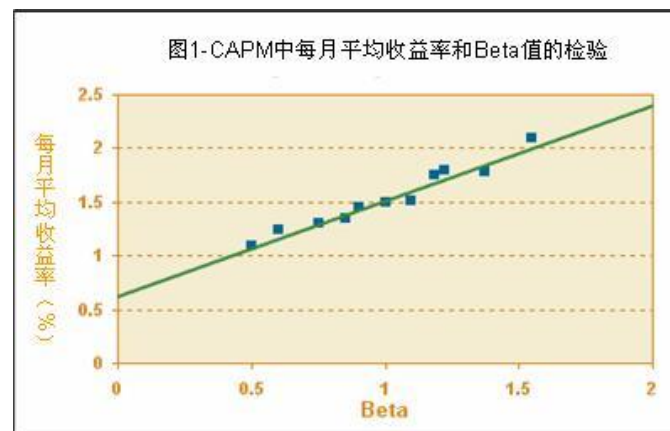
- 马科维茨的理论，要从1500只证券中挑选出有效率的投资组合，1966年，电脑运算耗费150~300美元；需要分析师必须能够持续且精确地估计标的证券的预期报酬、风险及相关系数。
- 夏普探索马科维茨证券投资组合的简化，导致了资本资产定价模型(capital asset pricing model, CAPM)的产生。
- CAPM认为一个资产的预期收益率与衡量该资产风险的一个尺度 $\beta$ 值之间存在正相关关系，单一指数模型。
- Beta系数衡量证券或投资组合相对总体市场的波动。

$$E(r_i) - r_f = \beta_i (E(r_m) - r_f)$$

$$\beta_i = \text{Cov}(r_i, r_m) / \text{Var}(r_m) = \sigma_{im} / \sigma_m^2$$

- $E(r_i)$  : 证券  $i$  的期望收益
- $E(r_m)$  : 市场组合的期望收益
- $r_f$  : 无风险资产的收益
- $\sigma_{im} = \text{Cov}(r_i, r_m)$  : 证券  $i$  收益率和市场组合收益率的协方差
- $\sigma_m^2 = \text{Var}(r_m)$  : 市场组合收益率的方差

**CAPM模型结论：**在均衡条件下，投资者所期望的收益和他所面临的风险的关系可以通过资本市场线、证券市场线和证券特征线等公式来说明。



Source: Black, Jensen and Scholes, 1972

# 3. Ross资本套利定价模型

- 套利定价理论认为证券收益率与一组因子线性相关，这组因子代表证券收益率的一些基本因素。
- 套利定价理论与现代资产组合理论、资本资产定价模型、期权定价模型等一起构成了现代金融学的理论基础。

$$r_i = a_i + \sum_{j=1}^k b_{ij} F_j + \epsilon_i, i = 1, 2, \dots, N \quad (1) \quad \text{套利定价理论的检验}^{[2]}$$

或

$$r = a + B * F + \epsilon \quad (2)$$

其中:

$r = (r_1, \dots, r_N)^T$  代表N种资产收益率组成的列向量.

$F = (F_1, \dots, F_K)^T$  代表K种因素组成的列向量

$a = (a_1, \dots, a_N)^T$  是常数组成的列向量

$B = (b_{ij})_{N \times K}$  是因素j对风险资产收益率的影响程度,称为灵敏度(sensitivity)/因素负荷(factor loading). 组成灵敏度矩阵.

$\epsilon = (\epsilon_1, \dots, \epsilon_N)^T$  是随机误差列组成的列向量.

并要求:

$$E(\epsilon_i) = 0, 1 \leq i \leq N \quad (3)$$

## 1. 罗尔(Roll)和罗斯的实证研究

与资本资产定价模型一样，套利定价理论也面临着实证检验的问题。Gehr(1975)、Roll and Ross(1980)、Reinganum(1981)

(1)首先对所有股票进行分组。

(2)收集一组股票的日回报率数据，并根据每只股票的回报率计算出该组股票的方差和协方差矩阵。

(3)运用最大似然因子分析法，确定影响收益率变化的因子个数以及因子载荷  $b_{ij}$ 。

(4)用估计出的  $b_{ij}$  来解释不同股票期望收益率在横截面上的差异，进而估计出每个因子相应的风险溢价及其显著性。

(5)对每一组股票重复2~4步。

# 4. Fama多因子模型

- Fama和French 三因子模型来解释股票回报率。
- 投资组合的超额回报率可由三个因子确定：
- 市场资产组合 ( $R_m - R_f$ )、市值因子 (SMB)、账面市值比因子 (HML)。
- $R_{ft}$ 表示时间 $t$ 的无风险收益率；
- $R_{mt}$ 表示时间 $t$ 的市场收益率；
- $R_{it}$ 表示资产 $i$ 在时间 $t$ 的收益率；
- $E(R_{mt}) - R_{ft}$ 是市场风险溢价，
- $SMB_t$ 为时间 $t$ 的市值 (Size) 因子的模拟组合收益率，
- $HML_t$ 为时间 $t$ 的账面市值比因子的模拟组合收益率。
- $\beta_i$ 、 $s_i$ 和 $h_i$ 分别是三个因子的系数，回归模型表示如下：
- $R_{it} - R_{ft} = a_i + \beta_i (R_{mt} - R_{ft}) + s_i SMB_t + h_i HML_t + \varepsilon_{it}$

$$E(R_{it}) - R_{ft} = \beta_i [E(R_{mt}) - R_{ft}] + s_i^E (SMB_t) + h_i^E (HML_t)$$

# Fama-French三因子模型的假设条件

## 1、理论假设

在探讨Fama—French三因子模型的应用时，是以“有限理性”理论假设为基础。并在此基础上得出若干基本假定：

- (1)存在着大量投资者；
- (2)所有投资者都在同一证券持有期计划自己的投资资产组合；
- (3)投资者投资范围仅限于公开金融市场上交易的资产；
- (4)不存在证券交易费用(佣金和服务费用等)及税赋；
- (5)投资者们对于证券回报率的均值、方差及协方差具有相同的期望值；
- (6)所有投资者对证券的评价和经济局势的看法都一致。

## 2、统计假设

从模型的表达式可以看出，FF模型属于多元回归模型。其基本假设为：

- (1)  $(R_m - R_f)$ 、SMB、HML与随机误差项 $u$ 不相关；
- (2)零均值假定： $E(\varepsilon_i) = 0$ ；
- (3)同方差假定，即 $\varepsilon$ 的方差为一常量： $Var(\varepsilon_i) = S^2$ ；
- (4)无自相关假定： $cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$ ；
- (5)解释变量之间不存在线性相关关系。即两个解释变量之间无确切的线性关系；
- (6)假定随机误差项 $\varepsilon$ 服从均值为零，方差为 $S^2$ 正态分布，即 $\varepsilon_i \sim N(0, S^2)$ 。

# 第七届“泰迪杯”数据挖掘挑战赛

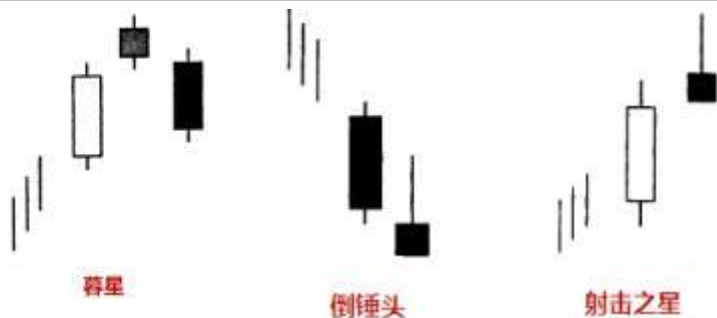
## A题：通过机器学习优化股票多因子模型

- 相比传统的线性多因子模型，机器学习算法能够通过对因子的非线性表达，捕捉到更加精细的市场信号，获取较为稳健的超额收益。根据 2016 年 1 月 1 日至 2018 年 9 月 30 日我国 A 股市场的数据（数据提取方式见附录 2），筛选出各大类股票因子中较优的子因子。在此基础上，分析不同的机器学习算法对提升这些因子的等权重线性模型表现的优劣，并使用“Auto-Trader 策略研究回测引擎”进行策略回测（初始资金为 1000 万元整，手续费为双边千分之 3，每月月初调仓）。可以从以下角度入手进行分析：
- (1) 利用 Auto-Trader 中各大类因子（见附录 3）的日频数据（数据提取方式见附录 4），分别做单因子策略研究和绩效分析，挑选出使得年化夏普比率（Sharpe ratio）最优的各个大类的因子。
- (2) 基于机器学习算法对 (1) 中挑选的因子，进行增强，利用 2016 年 1 月 1 日至 2018 年 9 月 30 日的数据进行选股和回测，比较不同机器学习算法选股策略与等权重线性模型选股策略之间年化夏普比率的优劣。
- (3) 对选股策略进行风险控制，要求将最大回撤控制在 10% 以内，重新完成 (2)。



# 点宽网—数据字典—BP 因子

基础科目衍生类	净营运资本、净债务、留存收益、毛利等
质量类因子	产权比率、超速动比率、非流动资产比率、股东权益比率等
收益风险类因子	120 日方差、股价偏度、历史贝塔、历史波动等
情绪类因子	20 日成交量标准差、20 日平均换手率、20 日换手率与 120 日换手率之比等
成长类因子	营业收入增长率、总资产增长率、5 年收益增长率等
常用技术指标因子	MA10、MA120、MTM、DBCD 等
动量类因子	BIAS20、CMO、PVT 等
价值类因子	PE、PB、PCF、PS 等
每股指标类因子	EPS、EBIPTS、TORPS 等
模式识别类因子	十字暮星、吞噬形态、刺透形态、倒锤头等
行业与分析师类因子	12 月相对强势、分析师盈利预期、投资回报率预测等
特色技术指标	绝对价格振荡器、平均价格、均势指标等



# 第八届“泰迪杯”数据挖掘挑战赛

## A题：基于数据挖掘的上市公司高送转预测

- 近年来，我国证券市场的高速发展催生了一批题材股，根据重大事件的不同分类，可以分为资产重组板块、粤港澳板块、新能源板块等等。在这些题材中间，高送转这一题材无疑是中小投资者强烈追捧的对象。因为实施高送转后股价将做除权处理，投资者可以通过填权行情从二级市场的股票增值中获利。很多股票在公布派送预案的第二天直接涨停，而等除权后再买入可能会面临很大的回撤风险。如果我们能准确预测下一年可能实施高送转的上市公司并提前买入，这对我们投资的安全性具有很大的现实意义。
- 经过研究，影响上市公司实施高送转的因子主要有两类：一是基本因子，包括股价、总股本、上市年限等；二是成长因子，包括每股未分配利润、每股资本公积、每股现金流、每股收益等。除此之外，还有“未来6个月是否存在解禁”、“是否存在定增方案”等因子需要挖掘。

# 4. 量化投资和期货对冲策略

<https://ww2.mathworks.cn/videos/machine-learning-for-algorithmic-trading-1503691224414.html>

哪些技术最有效呢？

表 1 对比基本面分析和量化方法（两者都由分析师进行）与机器学习和大数据方法的有效性。

技术是否有效：	基本面分析	传统量化方法	机器学习和大数据技术
一定时间段内的预测			
长期趋势	是	否	否
短期趋势	否	是	是
当日趋势（高频数据）	否	可能	是
不同数据类型的预测			
结构化数据	是	是	是
非结构化数据	否	否	是
少量数据	是	是	否
大量数据	否	否	是
模型解释难易程度	容易	适中	困难

# 詹姆斯·西蒙斯 (James Simons)



## • 经典名言

- 要像壁虎一样，平时趴在墙上一动不动，蚊子一旦出现就迅速将其吃掉，然后恢复平静，等待下一个机会
- 有些交易模式并非随机，而是有迹可循、具有预测效果的
- 有时候你像个英雄，因为你投资赚了钱，有时候你像个狗熊，因为赔了钱

## • 投资理论及主要策略

- 西蒙斯将他的数学理论背景巧妙运用于股票投资实战中。他通过计算机模型，大量筛选数十亿计单个数据资料，从中挑选出中意的证券买进、卖出。人们将西蒙斯的这种投资方式称为“定量投资”。通过计算机实现交易，也可以有效排除人为因素的干扰。
- 创立“壁虎式投资法”：西蒙斯这种投资方式更多集中于短线套利、频繁交易。

## • 业务模式

- 针对不同市场设计数量化的投资管理模型，并以电脑运算为主导，在全球各种市场上进行短线交易。对于数量分析型对冲基金而言，交易行为更多是基于电脑对价格走势的分析，而非人的主观判断。文艺复兴公司主要由3个部分组成，即电脑和系统专家，研究人员以及交易人员。西蒙斯亲自设计了最初的数学模型，他同时雇用了超过70位拥有数学、物理学或统计学博士头衔的人。西蒙斯每周都要和研究团队见一次面，和他们共同探讨交易细节以及如何使交易策略更加完善。
- 公司对交易品种的选择有三个标准：即公开交易品种、流动性高，同时符合模型设置的某些要求
- 随时都在卖出和买入，依靠活跃赚钱。西蒙斯称，他只寻找那些可以复制的微小获利瞬间，而绝不以“市场终将恢复正常”作为赌注投入资金。

## • 核心竞争力

- 与多数投资者关注市场基本面有所不同，西蒙斯的投资策略里充满了数学理论。他通过开发数量分析模型，大量筛选历史数据资料，借助相关性统计来预测期货、货币、股票市场的短期运动，并通过数千次快速的日内短线交易来捕捉稍纵即逝的市场机会。

# 量化对冲策略和高频交易

- 流动性回扣（同时买卖，赚取佣金返还）
- 猎物追踪算法（瞬间拉升杀跌，吸引投资者追涨杀跌的方法）
- 自动做市商（利用数据到达的速度差距赚取）
- 高频交易有违背市场公平，目前在国外也是限制的对象

- 阿尔法套利
- 股指期货套利
- 商品期货套利
- 期权套利
- 统计套利
- 另类套利

# 做空 (Short sale)

- 做空是一个投资术语，是金融资产的一种操作模式。与做多相对，做空是先借入标的资产，然后卖出获得现金，过一段时间之后，再支出现金买入标的资产归还。
- 做空是股票期货市场常见的一种操作方式，操作为预期股票期货市场会有下跌趋势，操作者将手中筹码按市价卖出，等股票期货下跌之后再买入，赚取中间差价。

“金融大鳄” 乔治·索罗斯	搜狐证券		
	<b>管理基金：量子捐赠基金</b>  索罗斯管理的量子捐赠基金据报资产管理规模为200亿美元。根据索罗斯管理的索罗斯基金管理公司向美国证监会呈报的资料，2008年底股票投资规模46.16亿美元。2008年的投资收益率为8%。	<b>经典战役：狙击英镑</b>  在1992年9月16日的黑色星期三，索罗斯以放空100亿英镑而声名大噪。最终英格兰银行只能被迫退出欧洲外汇机制并且让英镑贬值。索罗斯此役获得约11亿美元利润。他也被成为“让英格兰银行破产的男人”。	<b>投资理念：反射理论</b>  索罗斯在《金融炼金术》中详述了他的投资理念。其核心为“反射理论”，简单说是指投资者与市场之间的一个互动影响。在市场转折处进出，利用“羊群效应”逆市主动操控市场进行市场投机。仔细深入研究，坚持自己风格，悄悄撤出资金。
我生来一贫如洗。但决不能死时仍旧贫困潦倒。——索罗斯			
“对赌之王” 约翰·鲍尔森	搜狐证券		
	<b>鲍尔森：“赚钱之神”</b>  对约翰·鲍尔森而言，2008年是一个丰收年：在几乎99%的人都遭遇损失的时候，他却一年赚取了天文数字一般的37亿美元。但在07年之前，他的名字并不为华尔街熟知。鲍尔森1994年创立自己的对冲基金公司。他个人行事低调沉稳，喜欢跑步减压，每天甚至按时回家吃晚饭。	<b>经典战役：做空次贷</b>  2005年，鲍尔森确信投资者远远低估了抵押信贷市场上所存在的风险。他随这个市场会崩溃。在06年和07年，市场泡沫不断增加，鲍尔森不断加注，最终市场走向崩溃，而他也成功登顶2007年度最赚钱基金经理榜首。08年他继续做空英国金融股，继续大赚。	<b>投资理念：打破常规思维</b>  鲍尔森大获全胜，关键是相信独立分析调研。“他一直知道自己的判断是正确的，即便在市场逆行时候，从头到尾他都清楚自己在干什么。”
我不认为自己是一头牛或是一只熊。我只是个现实主义者——鲍尔森			

# 量化投资数学建模的计算 从mathematica编程入手

- Mathematica囊括了大量可立即计算的数据。用户可以通过编程访问这些数据，并且也可以通过Wolfram Research的数据服务器自动更新数据。某些数据如股票价格和天气数据都是实时递送的。
- 数据集包括：
- 天文数据：155,000个天体的99个属性
- 化学数据：34,000个化合物的111个属性，118个化学元素的86个属性以及1000个亚原子粒子的35个属性
- 地缘政治数据：237个国家的225个属性，以及全世界160,000个城市的14个属性
- 金融数据：186,000个股票和金融工具的历史和实时属性
- 数学数据：187种多面体的89个属性，3000种图的258个属性，6种knots的63个属性，21种晶格结构的37个属性，52个测地学方案的32个属性
- 语言数据：149,000个英语单词的37个属性。26个其他语言的词典
- 生物医学数据：所有40,000个人类基因的41个数据，27,000类蛋白质的30个属性
- 天气数据：全球17,000个气象站的43个实时和历史的测量数据
- Wolfram Alpha数据：来自Wolfram Alpha的亿万兆数据

# 量化投资模型的深度计算实现

## 从matlab编程入手

- Matlab Main Toolbox——matlab主工具箱 C
- Control System Toolbox——控制系统工具箱
- Communication Toolbox——通讯工具箱
- Financial Toolbox——财政金融工具箱
- System Identification Toolbox——系统辨识工具箱
- Fuzzy Logic Toolbox——模糊逻辑工具箱
- Higher-Order Spectral Analysis Toolbox——高阶谱分析工具箱
- Image Processing Toolbox——图象处理工具箱
- computer vision system toolbox——计算机视觉工具箱
- LMI Control Toolbox——线性矩阵不等式工具箱
- Model predictive Control Toolbox——模型预测控制工具箱
- $\mu$ -Analysis and Synthesis Toolbox—— $\mu$ 分析工具箱
- Neural Network Toolbox——神经网络工具箱
- Optimization Toolbox——优化工具箱
- Partial Differential Toolbox——偏微分方程工具箱
- Robust Control Toolbox——鲁棒控制工具箱
- Signal Processing Toolbox——信号处理工具箱
- Spline Toolbox——样条工具箱
- Statistics Toolbox——统计工具箱
- Symbolic Math Toolbox——符号数学工具箱
- Simulink Toolbox——动态仿真工具箱
- Wavelet Toolbox——小波工具箱
- DSP system toolbox——DSP处理工具箱



# Financial Toolbox

- [Get Started](#)
- Learn the basics of Financial Toolbox
- [Data Preprocessing](#)
- Financial market data for dates and currencies
- [Timetables in Finance](#)
- Timetables, date transformations and merges, chart technical indicators
- [Financial Data Analytics](#)
- Cash flows and performance metrics, regression analysis, financial data charting
- [Portfolio Optimization and Asset Allocation](#)
- Create portfolios, evaluate composition of assets, perform mean-variance, CVaR, or mean absolute-deviation portfolio optimization, backtest investment strategies
- [Credit Risk](#)
- Credit risk, transition probabilities for credit ratings, credit quality thresholds, credit scorecards
- [Price and Analyze Financial Instruments](#)
- Yield curves, valuation for fixed-income securities, equity derivatives pricing
- [Stochastic Differential Equation \(SDE\) Models](#)
- Parametric models, such as Geometric Brownian Motion (GBM) and Heston Volatility

# Get Started with Financial Instruments Toolbox

- [Use treeviewer to Examine HWTTree and PriceTree When Pricing European Callable Bond](#)

This example demonstrates how to use treeviewer to examine tree information for a Hull-White tree when you price a European callable bond.

- [Interest-Rate Curve Objects and Workflow](#)

Financial Instruments Toolbox class structure supports interest-rate curve objects.

- [Analysis of Bond Futures](#)

This example demonstrates analyzing German Euro-Bund futures traded on Eurex.

- [Portfolio Creation Using Functions](#)

Use the instadd function to create an instrument portfolio or to add new instruments to an existing portfolio using functions.

- [Instrument Construction and Portfolio Management Using Functions](#)

You can create instruments and manage a collection of instruments as a portfolio using functions.

- [Fixed-Rate Mortgage Pool](#)

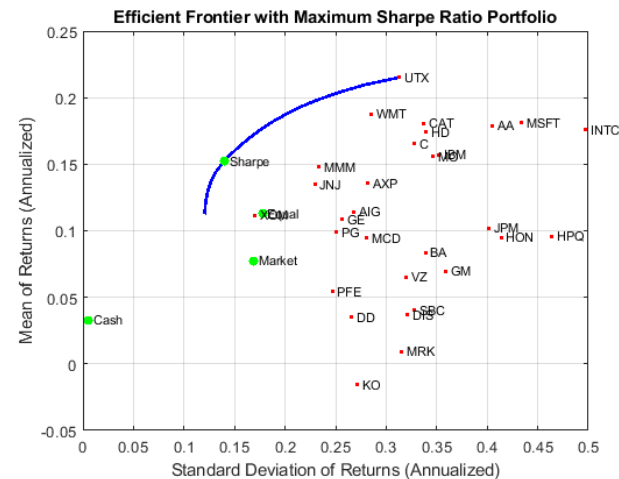
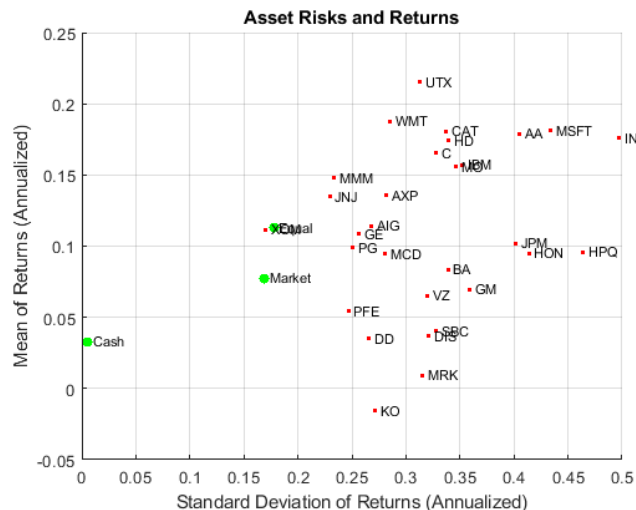
Generic fixed-rate mortgage pools and balloon mortgages have pass-through certificates (PC) that typically have embedded call options in the form of prepayment.

- [Get Started with Workflows Using Object-Based Framework for Pricing Financial Instruments](#)

Use objects to model and price financial instruments.

# Portfolio Optimization Examples

The following sequence of examples highlights features of the [Portfolio](#) object in the Financial Toolbox™. Specifically, the examples use the [Portfolio](#) object to show how to set up mean-variance portfolio optimization problems that focus on the two-fund theorem, the impact of transaction costs and turnover constraints, how to obtain portfolios that maximize the Sharpe ratio, and how to set up two popular hedge-fund strategies — dollar-neutral and 130-30 portfolios.



## 6. Matlab量化投资工具箱

Matlab提供了3个重要的量化投资相关工具箱：

- ✓ 金融工具箱（Financial Toolbox）
- ✓ 金融衍生品工具箱（Financial Derivatives）
- ✓ 固定收益箱（Fixed-Income Toolbox）

使用Matlab的知名金融机构主要有：

- IMF、FRB、高盛、摩根斯坦利
- 美国三大评级机构
- 美国Top 15的资产管理公司
- 美国Top 10商业银行有9家
- 美国Top 15对冲基金有12家
- OECD国家85%的中央银行

# Matlab金融工具箱



# Matlab金融衍生品工具箱

## 利率期限结构的计算与应用

利率期限结构的基础知识

利用利率期限结构  
计算价格与敏感性

利率期限结构模型

利率期限结构  
分析与计算

权益类衍生品的分析与定价

## 基于利率的金融衍生品

权益二叉树模型，  
ITT、CRR与EQP模型

奇异期权的介绍与结构

利用ITT、CRR与EQP模型，  
进行期权定价与敏感性计算

用解析解的形式，  
为特定期权定价

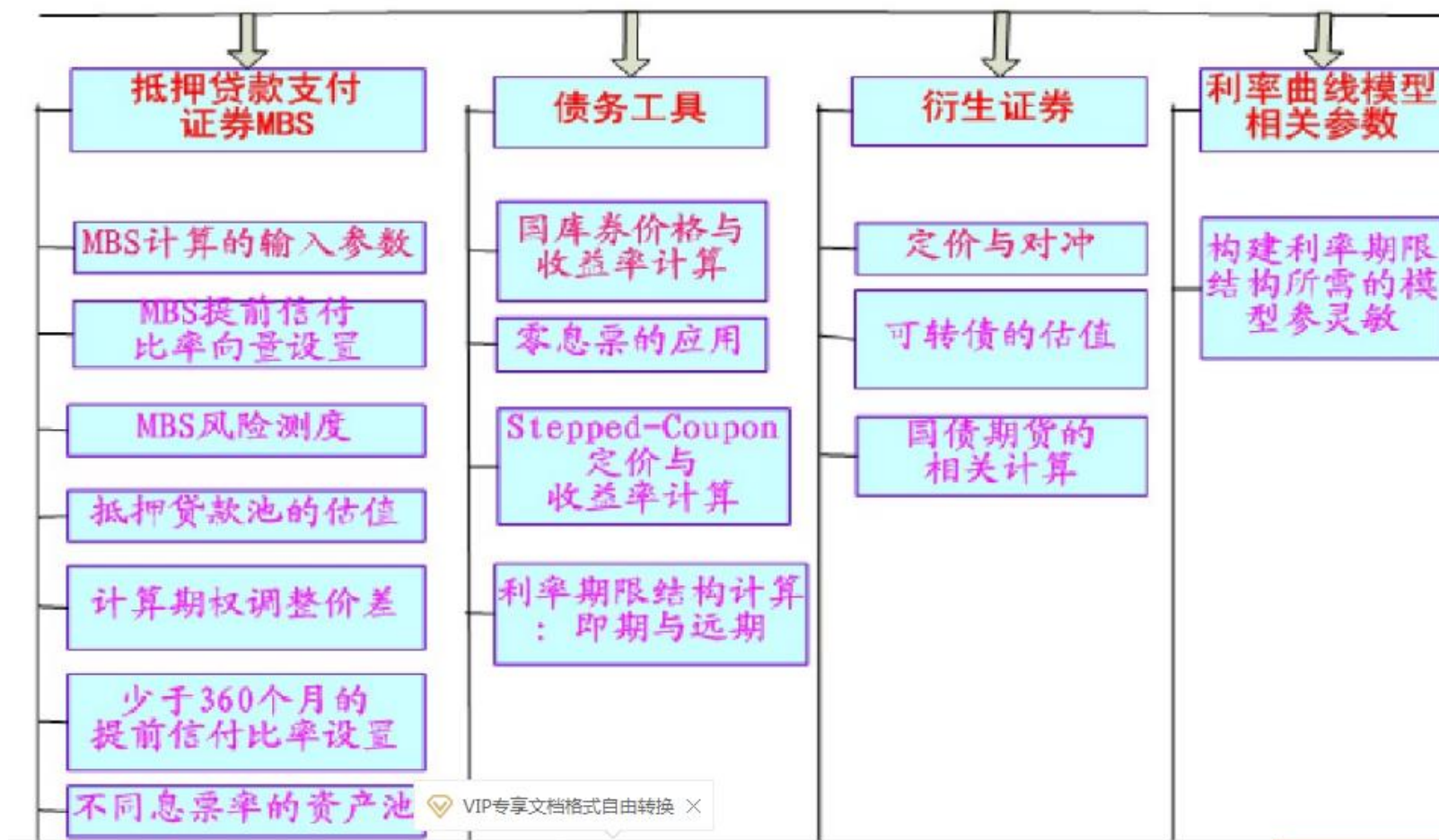
## 资产组合的对冲与优化

资产组合对冲相关函数

约束条件的说明与构建

含有约束的资产组合对冲

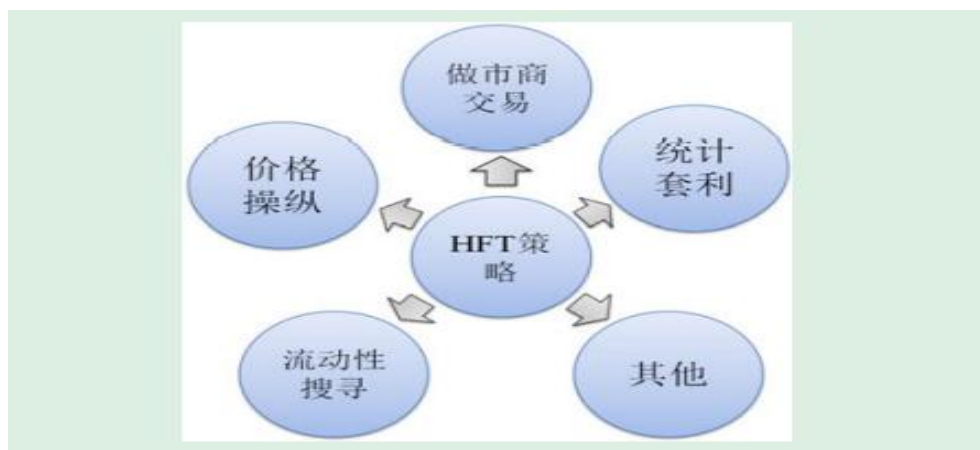
# Matlab固定收益工具箱





## 7. 统计套利策略和高频交易潜力

- 统计套利是基于统计方法挖掘套利机会，通过建立金融资产价格时间序列模型，拟合资产价格变化规律，识别资产组合内部统计错误定价关系，发现两个或者两个以上资产之间存在的套利机会，设定交易阈值，通过计算机自动发出交易信号并执行的投资过程。
- 高频交易主要策略分为三大类：流动性提供，即做市商交易，统计套利，盘口交易。
- 通过计算机即时地分析数据，选择交易策略，完成交易指令下达和执行具有先天的优势，是套利策略应用的自然选择。





## 8. ARCH模型和GARCH模型

- ARCH模型是一个噪声的发生是服从正态分布。该正态分布的均值为0, 方差是一个随时间变化的量(即为条件异方差)。并且这个随时间变化的方差是过去有限项噪声值平方的线性组合。这样就构成了自回归条件异方差模型。
- GARCH模型是一个专门针对金融数据所量身订做的回归模型, 除去和普通回归模型相同的之处, GARCH对误差的方差进行了进一步的建模。

# ARCH模型的商品期现套利

## 自回归条件异方差模型

ARCH模型的基本思想是指在以前信息集下，某一时刻一个噪声的发生是服从正态分布。该正态分布的均值为零，方差是一个随时间变化的量(即为条件异方差)。并且这个随时间变化的方差是过去有限项噪声值平方的线性组合(即为自回归)。这样就构成了自回归条件异方差模型。

由于需要使用到条件方差，我们这里不采用恩格斯的比较严谨的复杂的数学表达式，而是采取下面的表达方式，以便于我们把握模型的精髓。见如下数学表达：

$$y_t = \beta X_t + \epsilon_t \quad (1)$$
 其中，

★  $y_t$  为因变量，

★  $x_t$  为自变量，

★  $\epsilon_t$  为误差项。

如果误差项的平方服从AR(q)过程，即  $\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \eta_t, t = \{1, 2, 3, \dots\}$

(2) 其中，

$\eta_t$  独立同分布，并满足  $E[\eta_t] = 0, \text{var}(\eta_t^2) = \lambda^2$  则称上述模型是自回归条件异方差模型。简记为ARCH模型。称序列  $\epsilon_t$  服从q阶的ARCH的过程，记作  $\epsilon_t \sim \text{ARCH}(q)$ 。为了保证  $\epsilon_t^2$  为正值，要求  $\alpha_0 \geq 0, \alpha_i > 0, \sum_{i=1}^q \alpha_i < 1, i = \{1, 2, 3, \dots\}$

上面(1)和(2)式构成的模型被称为回归-ARCH模型。ARCH模型通常对主体模型的随机扰动项进行建模分析。以便充分的提取残差中的信息，使得最终的模型残差成为白噪声序列。

从上面的模型中可以看出，由于噪声的方差是过去有限项噪声值平方的回归，也就是说噪声的波动具有一定的记忆性，因此，如果在以前时刻噪声的方差变大，那么在此刻噪声的方差往往也跟着变大；如果在以前时刻噪声的方差变小，那么在此刻噪声的方差往往也跟着变小。体现到期货市场，那就是如果前一阶段期货合约价格波动变大，那么在此刻市场价格波动也往往较大，反之亦然。这就是ARCH模型所具有描述波动的集群性的特性，由此也决定它的无条件分布是一个尖峰胖尾的分布。

# GARCH模型-偏离正态分布

如果方差用ARMA模型来表示，则ARCH模型的变形为GARCH模型（波勒斯勒夫（Bollerslev），1986年）。

GARCH (p, q) 模型为

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2.$$

## IGARCH

IGARCH模型对GARCH的参数做了限制。IGARCH (p, q) 模型可以表示为：

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2,$$

条件是： $\sum_{i=1}^p \alpha_i + \sum_{i=1}^q \beta_i = 1$ 。

## GARCH-M

GARCH-M模型把异方差项引入平均数方程式。一个简单的GARCH-M (1, 1) 模型可以表示为：

$$y_t = \gamma x_t + \phi \sigma_{t-1} + \epsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

残差项  $\epsilon_t$  定义为：

$$\epsilon_t \sim N(0, \sigma_t^2).$$

## ARCH模型的应用

ARCH模型能准确地模拟时间序列变量的波动性的变化，它在金融工程学的实证研究中应用广泛，使人们能更加风险（波动性），尤其是应用在风险价值（Value at Risk）理论中，在华尔街是人尽皆知的工具。 [4]

## ARCH模型的变形和发展

- 波勒斯勒夫（Bollerslev）提出GARCH模型（Generalized ARCH）；
- 利立安（Lilien）提出ARCH-M模型；
- 罗宾斯（Robbins）提出NARCH模型。 [5]

# 9. 深度学习和大数据挖掘

- 机器学习使计算机或机器可直接从数据中学习，而无需通过编程来构建输入数据和输出的关系。从本质上讲，
- 机器学习模型在挖掘数据背后的非线性关系上要强于传统模型。您需要使用现代工具协同处理大数据和机器学习，才能将藏匿于大型数据集中的有价值信息提取出来。
- 在投资中，您能借助机器学习和大数据来做些什么？
  - 资产配置和优化
  - 应用自然语言处理来进行投资情绪分析
  - 异常值和欺诈检测
  - 金融预测和价格预测

# 什么是机器学习？

- 机器学习技术使计算机或机器无需大量编程即可直接从数据中学习。在金融领域，机器学习提供了一套新的方法
- 来开发预测模型。其与目前传统的模型相比，它能够更好地处理非线性关系的数据。
- 机器学习在业内流行着多种分类法。一般来说，根据我们试图解决的问题类型，对机器学习进行分类。例如：
  - 监督式学习：从标记数据中学习。在监督式机器学习中，结果表现为对数据标注正确（或所需）的响应。监督式学习使用分类和回归技术开发预测模型。
  - 无监督学习：从未标记的数据中发现模式。在无监督机器学习中，结果表现为不对数据标注响应。聚类是一种最常用的无监督学习技术。这种技术可通过探索性数据分析发现数据中隐藏的模式或分组。
  - 强化学习：学习行为或动作。强化学习的目的是建立一个模型，该模型能够执行一系列动作以最大化累积奖励。强化学习不是使用已知的输入和输出集，而是优化相对于奖励函数的动作。从根本上说，强化学习就像试错一样，智能体根据其动作从积极和消极的奖励中学习。

# 什么是大数据？

- 现有的技术需要在内存中进行运算。大数据通常是指很难运用现有技术进行处理的大批量数据。金融机构使用或存储的大数据包括：
- 过去 10 年中交易的 1000 多支证券的历史记录数据
- 数十亿笔信用卡交易
- 与过去 10 年中所有证券相关的新闻数据
- 1000 名客服人员的通话记录
- 值得注意的是当数据大于内存容量时，分组计算平均值和标准差的算法，其复杂程度远超数据小于内存容量的情况。
- 必须利用现代工具来处理大数据以及应用机器学习在数据中查找模式，才能提取藏匿于大数据中的有价值信息。

# 在投资中， 您能借助机器学习和大数据来做些什么？

- 遇到的复杂任务或问题中涉及大量数据和变量，但没有现成的处理公式或方程式，这时可以考虑使用机器学习和大数据技术。例如，以下情况适合采用机器学习和大数据技术：
- 数据的性质是非结构化的（例如文本、图像、音频或视频的组合）。
- 需要快速响应大量数据或高速数据，比如交易执行过程中的数据。
- 专家知识、手写规则和方程式对于模型而言太过复杂，比如针对新闻的投资情绪分析。
- 数据本身在不断变化，程序也必须适应这种变化，比如资产配置、自动交易、能量需求预测和价格趋势预测等。

## 10. 金融数学建模-经济、数学、计算和深度学习

- 经济金融问题
- 金融数学建模
- 金融模型计算
- 因子深度学习

W 万联证券 WANLIAN SECURITIES D DigQuant 点宽

### 第一届“大湾区杯” 粤港澳金融数学建模竞赛

2020.9.24-12.18

01

#### \\ 关于竞赛 //

为了开展数学建模方法在金融领域的创新研究，促进数字金融、数字货币和智能金融与传统金融投资领域的融合发展，鼓励数学和金融领域的交叉渗透和原始创新，培养、挖掘金融科技方向的具有国际视野的量化投资创新人才，广东省工业与应用数学学会、万联证券股份有限公司及粤港澳应用数学中心联合举办2020年“大湾区杯”粤港澳金融数学建模竞赛，欢迎高等院校按照参赛规则及有关规定组织学生报名参赛。

主办单位