

## 定量金融证书最终项目简介

### 2023 年 6 月组群

本文件概述了本组可提交的主题。不得提交其他主题。每个主题都有分步说明，为您提供了实施内容和方式的结构（而非限制）。

分数在很大程度上取决于你对数字技术的编码，以及对如何探索和测试定量模型的介绍（PDF 或 HTML 格式的报告）。某些数值方法涉及面太广或对模型有辅助作用，例如，不要重新编码优化或 RNs 生成。如果代码完全由您自己修改，则允许采用代码。

毕业设计项目要求在您的编码环境（如 Python、R、Matlab、C#、C++、Java）中对实现数值方法的软件包进行自学，并有能力使用相关文档。您无需预先批准编码语言和库的使用，包括 Scala、kdb+ 和 q 等非常专业的工具。

仅供 CQF 现有代表使用。不分发。

要完成该项目，您必须从以下选项中选择一个主题，对模型及其数值技术进行编码，并撰写分析报告。如果您是上一届学生中延续下来的，请查看题目说明，因为任务会定期审查。无法提交过去的课题。

1. 一篮子产品的信用利差 (CR)
2. 金融时间序列深度学习 (DL)
3. 货币对交易策略设计与回溯测试 (TS)
4. 利用布莱克-利特曼模型和因子构建投资组合 (PC)

5. 利用高级希腊语进行优化对冲 (DH)
6. 分类 (ML) 混合组合

本届学生的主题列表可在 Canvass Portal 的相关页面查看。

## 项目报告和提交

- 第一条建议：不要“原封不动”地提交 Python Notebook - 要将其转化为分析报告，还有很多工作要做。删除大表格/输出的打印输出。撰写数学部分（使用 LaTeX 标记）。撰写结果和压力测试的分析和比较（或类似内容）。解释图表。像数量学者一样思考计算和统计特性：收敛性/准确性/方差和偏差。将您编码/使用的数字技术制成表格。
- 项目报告必须包含充分的数学模型、数值方法和讨论利弊和进一步发展的充分结论。
- 没有固定的页数。有些代表喜欢在一页纸上呈现多个情节，以便比较，有些代表则选择叙述性更强的风格。
- 将 Python Notebook 报告保存为 HTML 格式是最佳选择，但要包含一个带有页码的 PDF 文件 - 供标记参考。
- 代码必须提交并正常运行。

文件 1. 为使我们的下载和处理脚本正常工作，您必须将项目报告命名为一个文件（pdf 或 html）并上传，

文件中应包含两个字母的项目代码，然后是您在 CQF 门户网站上注册的姓名。

例如：TS John Smith REPORT.pdf 或 PC Xiao Wang REPORT.pdf  
TS John Smith REPORT.pdf 或 PC Xiao Wang REPORT.pdf

文件 2. 所有其他文件、代码和 pdf 声明（如果不是首页）必须作为额外的一个 zip 文件上传，例如 TS John Smith CODE.zip。压缩文件中包括转换后的 PDF、Python 和其他代码文件。请勿提交未压缩的 .py、.cpp 文件，因为我们这边的云杀毒软件可能会闪红。

请勿提交带有通用名称的文件，如 CODE.zip、FinalProject.zip、Final Project Declaration.pdf 等。此类文件将被忽略。

项目提交日期为 2024 年 1 月<sup>22</sup> 日星期一格林尼治标准时间 23:59

---

最终项目没有延期时间。

没有亲笔签名的声明或工作代码的项目是不完整的。

如果没有按照命名说明提交一个报告文件和一个压缩文件，则意味着该项目将错过一次评分分配。

我们会对所有项目的原创性进行检查。我们保留在授予资格前进行口头审查的选择。

## 项目支持

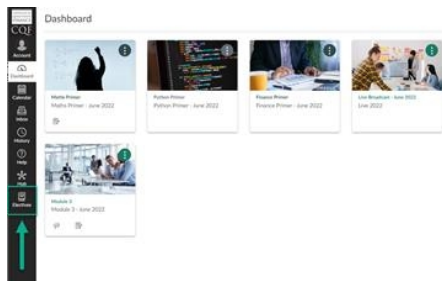
### 高级选修课

为了有针对性地获取背景知识，我们要求您复习两门高级选修课。选修课涉及多个知识领域，可以在撰写分析与讨论（对结果的解释）之前/同时/临近撰写时复习。

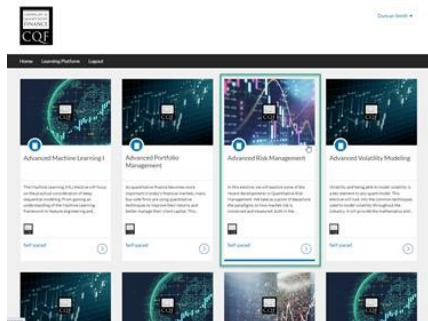
- 项目主题与选修课之间没有直接的匹配关系
- 每个项目主题都可能几种可行的组合
- 一种选修学习策略是选择一门 "专题选修课" 和一门 "编码选修课":

### 登录 CQF 学习中心

单击 "学习平台" 按钮登录 Canvas



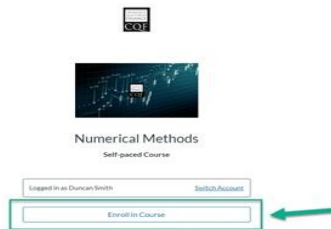
单击全球导航菜单上的 "选修课" 按钮



您将被转到选修课目录，在这里您可以查看和审查所有可供您选择的选修课。在这里可以找到每门选修课的完整说明。



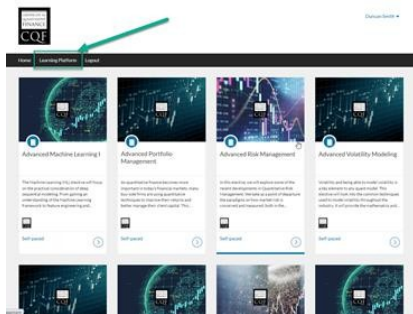
在选修课上单击 "报名" 按钮



您将看到确认页面，点击 *注册课程*

按钮确认选择

您将进入成功注册页面，在此您可以点击开始选修课或返回目录页面



在目录页面，您可以点击 *学习平台* 链接返回 Canvas。您选择的选修课将出现在您的学习仪表板上

## 工作坊和教程

每个项目名称都由一名教师提供支持，同时还提供一系列项目研讨会和辅导。

日期	标题	时间
02/12/2023	最终项目研讨会 I (CR & PC)	13:00 - 15:30 GMT
09/12/2023	最终项目研讨会 II (技术服务、卫生、领导力和管理)	13:00 - 15:30 GMT
18/12/2023	毕业设计辅导 I (TS 和 DH 主题)	18:00 - 19:00 GMT
19/12/2023	最终项目教程 II (DL 和 ML 主题)	18:00 - 19:00 GMT
20/12/2023	最终项目教程 III (PC 主题)	18:00 - 19:00 GMT
21/12/2023	期末项目辅导 IV (CR 专题)	18:00 - 19:00 GMT

### 教师支持

标题： 一篮子产品的信用利差一篮子产品的信

用利差 项目代码CR

领队Riaz Ahmad

标题： 金融时间序列的深度学习金融时间序列

深度学习 项目代码DL

领队Kannan Singaravelu

标题货币对交易策略设计与回溯测试 项目代码

TS

学院领导理查德-戴蒙德

标题： 利用布莱克-利特曼模型和因素构建投资组合利用布莱克-利特曼模型

和因素构建投资组合 项目代码PC

学院领导帕诺斯-帕拉斯

标题利用高级希腊语进行最优对冲 项目代码DH

学院领导理查德-戴蒙德

项目名称混合分类集合（ML）项目代码：ML

学院领导：卡南-辛格拉维鲁



如需就所选主题向故障排除人员提问，请单击门户网站右下角的 "支持 "按钮提交支持票据。

## 定量金融编码

- 选择具有适当优势和设施的编程环境来实现主题（定价模型）。常见的选择有 Python、Java、C++、R、Matlab。作为定量分析人员，锻炼判断力：哪种语言的程序库能让您更快地编码、更轻松地验证。
- 鼓励使用 R/Matlab/Mathematica。通常，Matlab/R 中的特定库可以为稳健协方差矩阵/协整分析任务中的特定模型提供快速解决方案。
- 项目简介中提供了 Matlab 演示的链接，Webex 会议演示了 Python 笔记本{并不意味着您的项目必须基于这些现成的代码
- 拥有 pandas、matplotlib、sklearn 和 tensorow 的 Python 对 Matlab 构成了相当大的挑战，甚至在可视化方面也是如此。Matlab 绘图编辑器很笨拙，而学习 Python 中的各种绘图并不难。
- 脚本化求解 "是指调用工具箱和库中的现成功能，但数值方法的自行编码量极少或根本不存在。这尤其适用于 Matlab/R。
- 仅使用 Excel 电子表格函数完成的项目不可靠、速度慢，而且无法了解基本的数值方法。CQF 提供的 Excel 电子表格是一个起点，有助于验证结果，但希望对数值技术进行编码/使用行业代码库。
- 该项目的目的是使您能够在生产环境中编码数值方法和开发模型原型。纯电子表格或脚本解决方案低于完成项目的预期标准。
- 我应该编码什么？学员应重新编码对模型至关重要的数值方法，并在确定这些方法时做出判断。

作为量化人员，可自行决定是否均衡使用库。

- 在报告中制作一个小表格，列出您实施/调整的方法。如果某项技术使用了现成的函数/借用的代码，请注明这一点，并说明在该代码/标准库中实现的数值方法的局限性。
- 代表们可自行开发测试用例、进行敏感性检查和验证。在实际数据中实施模型时，观察到不规则现象是正常的。如有疑问，请在项目报告中进行反思。
- 必须对代码进行彻底测试并提供完备的文档：必须对每个函数进行说明并使用注释。提供如何运行代码的说明。

## 一揽子产品的信用利差

为 5 个参考名称（一篮子 CDS）的 CDS 投资组合定价公平价差，作为对违约时间联合分布的估算。从分析角度看，该分布是未知的，因此要从协整变量中抽取共同依赖的均匀变量，然后使用危险率的边际期限结构（每个名称分别使用）将其转换为违约时间。通过估计适当的违约相关性来校准 Copula（CDS 差异的历史数据是自然的候选数据，但会带来市场噪声问题）。初步结果是直方图（均匀性检查）和散点图（共同依赖性检查）。实质性结果是通过重新定价进行敏感性分析。

一个成功的项目将同时从高斯和 t 协方函数中进行采样，并对所有 k-th 到默认工具（第 1 到第 5 个）进行定价。采样时，价差收敛可能需要低差异序列（如 Halton、Sobol）。需要对输入进行敏感性分析。

### 数据要求

需要两个**独立**的数据集，以及每个数据集的匹配贴现曲线数据。

1. 特定日期的**信用曲线快照**。债务发行人很可能有一条美元/欧元 CDS 曲线--从该曲线中引导出危害率的期限结构，并利用该结构获得准确的违约时间，即  $u_i \rightarrow \tau_i$ 。在没有数据的情况下，可以假定或从财经媒体的图表中直观地剥离每个期限的利差值。典型的信贷曲线呈凹形（正斜率），1 年期、2 年期、.....、5 年期单调递增。5 年期的单调递增。
2. **历史信用利差时间序列**取自每个参考名称最具流动性的 5 年期。因此，对于 5 个名称，可以计算出  $5 \times 5$  的违约相关矩阵。选择公司名称时，根据股票收益计算相关矩阵要容易得多。

---

公司信贷息差不太可能公开；可从彭博或路透终端（通过您的公司或同事）获取。至于主权信用利差，可从 DB Research 获取已准备好的自引导 PD 时间序列<sub>5Y</sub>，但开放访问的情况各不相同。探索数据来源，如 [www.datagrapple.com](http://www.datagrapple.com) 和 [www.quandl.com](http://www.quandl.com)。

即使 CDS<sub>5Y</sub> 和 PD<sub>5Y</sub> 系列以每日为频率，每日变化的共同运动也是市场噪音，而不是违约事件的相关性，因为违约事件很少被观察到。周/月变化为违约相关性提供了更合适的输入，但这需要使用 2-3 年的历史数据，因为我们需要至少 100 个数据点来估计相关性的显著程度。

**如果在获取历史信用利差方面存在问题，可根据历史股票收益率或债务收益率估算违约相关矩阵。**

## 分步说明

1. 对于每个参考名称，从 CDS 报价中引导隐含违约概率，并将其转换为危险利率的期限结构，即  $\tau \sim \text{Exp}(\lambda^{1Y}, \dots, \lambda^{5Y})$ 。
2. 估算默认相关矩阵（近相关矩阵和秩相关矩阵）和 d.f. 参数（即校准 copulae）。您需要通过高斯和 t 协方差分别实现定价。
3. 使用抽样形式 copula 算法，重复以下例程（模拟）：
  - (a) 生成一个相关均匀随机变量向量。
  - (b) 对于每个参考名称，使用其危险利率的期限结构来计算确切的违约时间（或使用半年应计制）。
  - (c) 计算每种工具从第 1 次到第 5 次违约的溢价和违约部分的贴现值。分别进行 MC 或使用一个大的模拟数据集。
4. 分别计算各次模拟的平均溢价和违约脚数。计算公允价差。

## 模型验证

- 第 k 次违约篮子 CDS 的公允价差应小于第 k-1 次违约。为什么？
  - 有关该主题的项目报告应包含**风险和敏感性分析**部分的公平价差。
    1. 参考名称之间的默认相关性：通过恒定的高/低相关性或相关性与实际估计水平的  $\pm$  百分比变化进行压力测试。
    2. 每个单个名称的信用质量（信用价差、信用 Delta 值的变化）以及回收率。
- 确保对所有五种仪器的敏感性进行讨论和比较。
- 确保您解释了默认相关矩阵和 copula 拟合（伪样本的均匀性）的历史采样--即相关性实验和分布拟合实验，这将在项目研讨会上介绍。使用直方图。

## 市场风险的 Copula、CDF 和尾数

关于使用 copula 生成相关样本的最新实用教程，可在以下网址查阅：

<https://www.mathworks.com/help/stats/copulas-generate-correlated-samples.html>

通过半参数 CDF 拟合，我们可以得到中间和尾部的百分位值。在尾部建模时应用了一般化帕累托分布，而 CDF 内部则是高斯核平滑的。这种方法源自极值理论，该理论建议对经验 CDF（核拟合）进行修正，因为尾部会出现超差。

<http://uk.mathworks.com/help/econ/examples/using-extreme-value-theory-and-copulas-to-evaluate-market-risk.html>

<http://uk.mathworks.com/help/stats/examples/nonparametric-estimates-of-cumulative-distribution-functions-and-their-inverses.html>

## 阅读清单

- 您很可能会重温 *CDO 和 Copula 讲座* 材料，特别是说明椭圆 copula 密度和讨论 Cholesky 因式分解的第 48-52 张幻灯片。
- *从 copula 算法中取样*，见 Peter Jaekel（2002 年）所著的《*Workshop and Monte Carlo Methods in Finance*》教科书第 5 章。
- *相关灵敏度讲座* 和 P. Jaekel（2002 年）也介绍了等级相关系数。CR Topic Q&A 文件给出了明确的公式和解释。

# 用于资产预测的深度学习

## 摘要

几十年来，趋势预测已经吸引了大量研究人员使用统计和计算方法（包括机器学习技术）进行研究。趋势预测对投资管理很有价值，因为准确的预测可以确保资产管理者跑赢市场。趋势预测仍然是一项具有挑战性的任务，原因在于市场效率的半强形式、高噪音信号比以及影响资产价格的众多因素，包括但不限于相关工具的随机性质。然而，序列金融时间序列可以使用序列建模方法（如递归神经网络）进行有效建模。

## 目标

您的目标是利用长短期记忆网络建立一个预测积极走势（上升趋势）的模型。您提出的解决方案应包含详细的模型架构，并通过应用于交易策略的回溯测试进行评估。

- 从指数、股票、ETF、加密货币或商品中选择一个您感兴趣的代码。
- 只预测趋势，预测短期回报（例如：每日，6 小时）。将预测局限于二项式分类：因变量的最佳标签为 [0, 1]。避免使用 [-1, 1] 作为类标签。
- 分析应当全面，包括详细的特征工程、数据预处理、模型建立和评估。

**注：**您可以自由选择研究设计，以完成任务。您可以重新定义任务，预测动量符号（相对于回报符号）或波动方向。将探索范围限制在一种资产上。在每一个步骤中，都应该对所遵循的过程进行扩展和详细解释。仅提交 python 代码而不作适当解释的报告将不予接受。报告应详细介绍研究内容，并给出适当的结论。代码的可重现性是必须的，建议使用模块化编程方法。在本课题下，您不得对现有指标、库或优化进行重新编码，以计算神经网络权重和偏差。



## 分步说明

1. 问题陈述应明确无误，包括基础资产的选择、数据集、时间范围和使用数据的频率。

- 如果要预测短期回报迹象（每日变动），那么最长 5 年的训练和测试就足够了。如果要预测 5D、10D 股票收益率或 1W、1M Fama French 因子，则必须将所需数据增加到至少 10 年。

2. 执行详尽的特征工程 (FE)。

- FE 应该详细，包括列出衍生特征和指定获取/标记。制定如何对极小的近零收益进行分类的方法（从训练样本或正/负收益组中删除）。阈值在很大程度上取决于您的股票代码。例如：0.25% 以下的小正值回报可标记为负值。
- 应通过模型参数或标签定义来解决类别不平衡问题。
- 允许使用协整对和跨资产的特征，但在设计时应讲究策略。没有一套适用于所有资产的推荐特征集，但初始特征集应足够大。财务比率、高级技术指标（包括波动率估算器）和成交量信息可以预测价格方向。
- 可选项 使用新闻热图、信用利差（CDS）、财务比率历史数据、股息历史数据、主要利益相关者（董事交易）或大型基金的购买/处置数据或 Fama-French 因子数据可增强预测效果，这些数据可从您的专业订阅中获取。

3. 进行详细的探索性数据分析 (EDA) 。

- 通过更好地理解特征之间的关系和揭示潜在结构，EDA 有助于降低维度，并有助于检测/解释异常值。特征缩放技术的选择应由 EDA 决定。

4. 必须正确处理数据。使用不同的特征集、回溯长度和数据集需要进行清理和/或估算。

5. 应根据 EDA 进行特征转换。

- 预测因子之间应进行多重共线性分析。
- 展示特征之间关系的多散点图始终是个好主意。
- 大型特征集（包括重复类型和不同回溯）需要降低特征的维度。自组织图 (SOM)、K-Means 聚类或其他方法都可用于降维。避免对非线性数据集/预测因子使用主成分分析 (PCA)。

6. 进行广泛而详尽的模型构建。

- 经过广泛而详尽的研究，设计神经网络架构。
- 只有在进行超参数优化并与基线模型进行比较后，才能提出最佳模型。
- 为获得最佳模型而需要优化的超参数的选择和数量是设计选择。使用 MLFlow 或 TensorBoard 等实验跟踪器来展示您的研究。

7. 应使用多种指标来评估所建议的分类器的性能，包括对应用于交易策略的预测信号进行回溯测试。

- 使用 AUC、混淆矩阵和分类报告调查预测质量，包括平衡精度（如需要）。
- 应将预测信号应用到交易策略中进行评估。

\* \* \*

# 货币对交易策略设计与回溯测试

价格之间的协整关系为套利提供了机会。套利的基础是静态价差的均值回复，这是一种特殊的协整残差。将信号生成和回溯测试作为项目的核心，这不是一次性运行的统计例程。传统上，配对交易是通过相关性进行的，您仍然可以检查一系列资产的高相关性以寻找配对。然而，100% -100% 权重的配对交易是幼稚的，它没有考虑到美元差异，也不利于剩余资产的静态性。资产价格必须与协整（误差修正模型）联系起来，因为它们是非平稳序列， $I(1)$  导致综合布朗运动是一个基本过程的假设。价差是否适合交易取决于 OU 过程拟合和半衰期。

实施的数字技术：矩阵形式自回归、恩格尔-格兰杰方法和统计检验。我们鼓励你大胆尝试多变量协整（VECM、Johansen 程序）和协整权重的稳健性检查，即通过对回归参数进行自适应估计（可选）。多变量协整的优势：交易策略的权重很难从外部猜测。不过，这也带来了  $P \& L$  归属（解释）、 $\beta$  dev 级 Python 库（至 2023 年）的损失。相比之下，Engle-Granger 程序具有很好的仿射性和可控性，因此是开始协整分析的不错选择。它仅适用于两个序列的配对分析。

## 信号生成和回溯测试

- 除了股票配对外，还要有新意：考虑商品期货、VIX 期货、美国/英国债券 ETF 和其他利率工具。
- 通过使用协整系数  $\beta_{Coint}$  作为分配  $w$  来实现 Arb，从而创建一个多空组合，产生均值回复价差。所有项目设计都应包括交易信号生成（来自 OU 过程拟合）和回溯测试（下降图、滚动 SR、滚动 Betas）。
- 累积损益是否符合协整仲裁交易的预期？P&L 是来自几次交易还是多次交易，半衰期是多久？最大回撤和波动率/VaR 的表现？
- 引入流动性和算法流考虑因素（订单流模型）。是否有累积头寸的规则？买卖价差和交易成本会产生什么影响？

## 分步说明

可以利用准备就绪的多元协整（R 软件包 *urca*）先确定您的协整案例，尤其是如果您与系统一起操作，如四种商品期货（到期日不同，但都在交易期间。如果按 EG 分析单独的货币对，则需要 2-3 个货币对。

## 第一部分：配对交易设计

- 即使您使用的是成对变量，也可以重新编码矩阵形式的回归估计--您可以重复使用自己的 OLS 实现。静止变量之间的回归（如 DF 检验回归/差分方程）可选择进行模型规格检验，以便：（a）通过 AIC

BIC 检验确定最佳滞后  $p$ ; (b) 稳定性检验。

2. 对每对数据实施 Engle-Granger 程序。第 1 步，使用增量 DF 检验滞后 1 的单位根。第 2 步，列出两个修正方程，并决定哪个更重要。
3. 决定信号：常见的方法是在边界  $\mu_e \pm Z\sigma_{eq}$  上入场，在  $e_t$  reverting 到大约水平  $\mu_e$  上出场。
4. 首先，假设  $z = 1$ 。然后将  $z$  稍微向上和向下改变--计算每种扩大和收紧边界情况下的损益，并给出信号。或者运行一个优化程序，在  $\mu_e \pm Z_{opt}\sigma_{eq}$  的情况下改变  $Z_{opt}$ ，并使累计损益或其他标准最大化。  
注意权衡：更宽的界限可能会带来最高的损益和最低的交易量  
但是，要考虑协整关系破裂的风险。
5. 可选择使用 R 软件包 *urca* 尝试多变量协整--截至 2023 年，Python VECM 模型仅在 *statsapi* 的 Github 开发版本中可用--以便为货币对/篮子交易选择最佳候选货币。

## 第二部分：回溯测试

作为一名量化交易者，您可以自行决定需要介绍哪些关于交易策略的可行性、稳健性和 "非相关收益" 性质的内容。

4. 考虑机器学习启发的回溯测试，如将数据拆分为训练/测试子集、预处理，并在适当可行的情况下进行交叉验证（注意时间序列分析中的交叉验证问题）。
5. 对您的交易策略（货币对交易的收益）进行系统性回溯测试：绘制缩减图、滚动夏普比率、与 S&P500 超额收益相比的至少一个滚动贝塔值。但是，请讨论为什么滚动贝塔系数可能与统计套利和做市商无关。
6. 可选项 学术研究将检验与下列变量的协整关系是否存在突破

LR 检验。协整关系应该持续存在，且  $\beta'$  应该保持不变：

继续提供 3-6 个月的文具，无需更新。这对你们来说现实吗？

讨论通过转移数据 1-2 周定期重新估计协整关系的利弊（切记保留一些未来数据），并报告在滚动  $\beta'$  硬币，而且恩格尔-格兰杰步骤 2 的检验统计量值的历史记录为

EC 项前面的系数。

您是否会实现类似卡尔曼滤波器/粒子滤波器自适应估计 [ap-] 的功能？

用于协整回归]，以查看更新后的  $\beta'$  硬币 和  $\mu_e$ ？参考文献：

[www.thealgoengineer.com/2014/online\\_linear\\_regression\\_kalman\\_filter/](http://www.thealgoengineer.com/2014/online_linear_regression_kalman_filter/)。

**TS 项目研讨会、协整关系讲座和配对交易教程是您的重要资源。**

# 利用布莱克-利特曼模型和因子构建投资组合

## 摘要

构建因子投资组合，计算至少两种优化。在每种优化中，利用布莱克-利特曼模型更新绝对和相对配置。计算三种常见风险规避水平（受托人/市场/凯利投资者）的最优配置。实施系统性回溯测试：包括将投资组合的结果与因子进行回归，以及研究因子本身（与市场超额收益相比）。

优化类型：均值-方差、最大夏普比率、高阶矩（最小斜率、最大峰度）--至少实现两种。最小跟踪误差也是可能的，但你的投资组合选择将根据基准指数来衡量。通过现成公式或二次编程专用公式计算。添加限制条件可提高稳健性：大多数投资者都有保证金限制/借贷能力有限/没有空头头寸。

可选的是，风险贡献也可以事先计算出任何最优分配，而计算 ERC 投资组合则需要求解风险预算方程系统（非线性）。ERC 计算不是优化，但可以 "转换" 为优化--顺序二次编程 (SQP)。

## 投资组合选择与数据

投资组合资产的选择必须体现最佳的多样化。最优性取决于标准。对于资产间最大可能的不相关性，选择相关性最小的资产是最直接的。至于对因子的风险敞口/倾斜度，则需要事先了解因子的贝塔值，并根据目的选择贝塔值高或低的资产。

S&P500 大盘股的天真投资组合完全受市场指数本身这一个因素的影响，这是不够的。针对某一行业、新兴市场、信贷资产的专业化投资组合应有 5 个以上的投资品种，以及 3 个以上不相关的资产，如商品、VIX、债券、信贷、房地产。

因子投资组合更像是一种多空策略，例如，动量因子意味着做多涨幅前 5 位的股票，做空跌幅前 5 位的股票。因子投资组合的设计意味着再平衡（时间多样化）。

- 哈里-马科维茨 (Harry Markowitz) 对简单收益（非对数）的均值-方差优化进行了规定，这些收益超过了  $r_f$ 。对于无风险利率，pandas FRED 数据集上的 3M 美国国债/欧洲央行网站上的欧元利率/一些小的恒定利率/零利率--都可以接受。使用 2-3 年的样本，即大于 500 个日收益率。
- 价格数据的来源是雅虎财经（美国股票和 ETF）。使用代码库访问，如 Google Finance、Quandl、Bloomberg、路透社等。如果没有基准指数，则根据市值（美元价值）计算平衡权重。
- 在 PC 主题的这一变体中，有必要引入 2-3 个因子时间序列，并将其视为可投资资产（5 个法马-法式因子）。如果使用 Smart Beta ETF，根据它们的结构，你可能会发现并没有实际的多空因子，只是收集了一些具有特别高 betas 的纯多头资产。

## 分步说明

### 第一部分：因子数据与研究（回溯测试）

1. 根据您的最佳分散投资方法实施投资组合选择。通常，主要任务是选择少数资产，使其风险调整后的回报与规模更大、自然分散的基准（如标准普尔 500 指数）相同或优于后者。参见研讨会上分发的问答文件。
2. 尝试引入哪些因素，收集其时间序列数据或进行计算。
  - 经典的法马-法式因子是 HML（价值因子）和 SMB（小型企业）。RMW（稳健盈利能力与薄弱盈利能力）和 CMA（保守资本支出与激进资本支出）是新的因子，您可以尝试使用。
  - 行业或风格风险也可被视为一个因素。
  - 建议您引入一个有趣的自定义因素，如 Momentum、BAB（对赌贝塔值）--您可能需要计算其回报的时间序列，但这可以很简单，就像五大科技股的短期投资组合的回报一样。
3. 从源头上更好地解释了对因子进行回溯测试的投资组合范围  
[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)
4. 展示您的因子与市场（您选择的指数）的损益回报和系统回溯测试，其中包括绩效、滚动贝塔图和变化阿尔法图。理想情况下，您可以单独展示每个因子的贝塔值，然后再综合展示。甚至在进行投资组合优化之前，也要展示这项工作

### 第 II 部分：基本法产出的比较分析

1. 规划您的布莱克-利特曼应用。找到一个现成的基准或构建先验：均衡收益可以来自足够广泛的市场指数。实现后验收益率 BL 公式的计算版本。
2. 强加过多的视图会让人很难看出每个视图的影响。
3. 分析描述并计算至少两种优化。使用合理的约束条件（如预算约束条件、“不得持有债券空头头寸”）可以改进优化，但此类不平等约束条件  $w_i > 0$  会引发分配的数值计算。
4. 即使是经典的均值-方差优化（两种优化中的一种），您最终也会得到多组最优分配。请自行选择分析和讨论的重点--最可行、最能说明问题的比较结果。

---

  - 主动风险的最佳分配（你的）与基准。预期收益（ $\bar{r}_p$ ）与隐含均衡收益（类似于 T. Idzorek 所著《BL Guide》中的表 6）。
  - BL 视图不受协方差矩阵的影响--因此，您可以选择通过布莱克-利特曼模型来计算视图移动的分配，并使用天真或稳健协方差。



- 三个级别的风险规避 - 建议您至少探索经典的 Min Var 优化。
5. 该项目没有重新平衡的任务，特别是因为《基本法》规定的后期分配预期是持久的。
  6. 将您的自定义投资组合与因子和市场（滚动贝塔值）的表现进行比较，既可单独比较，也可联合比较。可选择将您的投资组合与  $1/N$  分配/分散比率投资组合/简单风险平价投资组合的表现进行比较，并将该投资组合与因子进行系统性回溯测试。

## 阅读清单

- CQF 关于优化基础知识和投资组合选择应用的讲座
- 托马斯-伊德佐雷克（Thomas Idzorek）2002 年出版的《布莱克-利特曼模式分步指南》讲述了你需要实施的基本内容。
- 布莱克-利特曼方法：Original Model and Extensions Attilio Meucci, 2010. <http://ssrn.com/abstract=1117574>
- 关于 LW 非线性收缩/Marcenko-Pastur 去噪，无论哪种方法都能使协方差矩阵变得稳健，相关研讨会和教程提供了资源和特定代码。



# 利用高级三角洲建模优化套期保值

## 摘要

在本题中，首先要考虑的是在未来实际波动率高于隐含波动率的条件下的简单波动率套利，即  $V_a > V_i$ 。具体做法请参见《理解波动率》讲座解答。用代码实现 delta 复制：做多期权，做空股票。使用欧式看涨期权布莱克-斯科尔斯公式，高/低波动率值自选。提供 Gamma 在  $\Delta t$  期间如何影响损益的可见性。GBM 演进应利用其数值技术和先进的 Monte-Carlo 降低方差和/或低差异序列。无需考虑期权组合或多项资产。

可以通过调整 Black-Scholes Delta 值来改进 Delta 对冲。我们选择最小方差 Delta 法，该方法根据资产  $S_t$  的变化对  $\sigma_{imp}$  的预期变化修正 Delta 值，即  $\frac{\partial E(\sigma_{imp})}{\partial S}$ 。其基本思想很简单：预测隐含波动率的变化幅度，并据此调整 Black-Scholes Delta。

MVD 模型系数  $a$ 、 $b$ 、 $c$  的估计可通过对  $\delta_{BS}$ 、 $\delta$  的回归来完成。<sup>2</sup>  $BS$  作为内直接求解因变量，然后将  $\Delta V - \delta_{BS} \Delta S$  重新排列为因变量。更先进的方法是使用顺序最小平方编程优化器（scipy）来求解

$a$ 、 $b$ 、 $c$  是使 Hull 和 White (2017) 定义的对冲误差  $\sum_M \epsilon^2$  最小化的参数。

在你 MVD 的研究结果中，你可能会提出、说明和讨论以下问题。(1) 你可能会发现 MVD 始终小于 BS Delta，因此，天真的 Delta 对冲通常是过度套期保值。这对价外看涨期权尤其有效。(2) 经过估算，可以发现 MV Delta - BS Delta 的差值 ( $\delta_{MV} - \delta_{BS}$ ) 根据经验给出了一个类似倒抛物线的函数（跨 Delta 桶），因此出现了二次拟合的方法。(3) 高级实施将考虑增益函数，其定义为对冲后残差平方和的减少百分比。

## IV 选项数据

1. S&P500 指数期权数据可从券商和数据提供商 **OptionsDX.com**、Polygon.io 获得。OptionsDX 是当前一个特别好的选择（2023 年），它提供一些免费的日终（EOD）期权报价。在每个交易日和每个到期日/罢工日，你需要 BS 期权价格作为隐含波动率百分比、delta 和 vega:  $(V, \delta_{BS}, v)_{BS,t}$ 。
2. 在没有历史期权数据的情况下，隐含波动率的数值可以从一系列数值中随机生成。对模拟值应用一个乘法因子，该因子从短期到期日到长期到期日各不相同。这个方法模仿了在股票指数期权中观察到的 IV 模式，你可以根据自己对市场中微笑如何发生的了解进行调整。
3. 本项目不需要拟合波动率曲面或 Bloomberg/前台级期权价格数据，但你当然可以使用任何来源的波动率数据。

## 分步说明

### 第一部分：采用改进的 GBM 和 Monte-Carlo 算法的波动率 Arb

1. 考虑改进 GBM 资产演化 (Euler-Maruyana/Milstein 方案)。另外，也可以考虑对具有跳跃性的资产进行建模，如 Merton 跳跃扩散，而不考虑随机波动性，如 Heston-Nandi 方差伽马也与此有关，但适用于具有极端波动的单名资产。
  - 考虑 MC 方差减小技术，例如对立变体；
  - 最佳做法是采用低差异序列，例如采用布朗桥的索博尔序列。
2. 在已知未来实际波动率  $v_a > v_i$  的条件下，通过分析和 Monte-Carlo 验证以下项目。请同时报告折叠  $P \& L_t$  的完整数学计算和  $P \& L_t$  的模拟。
  - 确认实际波动对冲导致已知的总损益；
  - 确认并证明隐含波动率对冲会导致不确定的、与路径相关的损益总额，并说明其取决于哪些参数/希腊语。
3. 想一想其他分析：考虑损益如何按希腊字母分解。与时间相关的 Gamma  $\Gamma_t$  有什么影响？ $r^2 - \sigma_{imp} \delta t$ ？考虑第 II 部分 MVD 模型的结论，用较小的 delta 对冲有什么影响？

### 第二部分：最小方差德尔塔

1. 首先整理你的 IV 数据--或者说每个交易日，你需要 BS 期权价格作为隐含波动率百分比、delta 和 vega： $(v_t, \delta, v_{BSBS})$ 。期权的期限结构为 1M、3M、6M、9M、12M，每周到期的期权没有必要。如果你要研究价外看涨期权的 Delta 值，除了关于 ATM 桶的  $0.45 < \delta_{BS} < 0.55$  之外，这里还有一个关键的选择--每个行权意味着每个到期日都有一个单独的  $a$ 、 $b$ 、 $c$  历史记录。
2. 计算因变量，并对  $\delta_{BS}, \delta$  进行拟合。<sup>2</sup>  $BS$  基于从属方  
你需要  $(\Delta S_t, S_t)$  以及上面提到的希腊文。具体的数据列取决于你如何组织回归或进行 SLSQP。
3. 参数  $a$ 、 $b$ 、 $c$  对于一个研究项目来说可以是常数，但滚动估算本身就是一种校准技术，因为对于每个到期日，你都会有与时间相关的  $a$ 、 $b$ 、 $c$ （而不是 3 个常数）。赫尔-怀特的方法是 3M 滚动窗口，然后将开始日期移动一天（ $3 \times 22$  个观测点）--你可以用更短/更长的周期进行估算，或者移动 5-10 天。另外，请记住，选项 IV 可以通过均匀分布公式进行模拟。
4. 为验证模型，请查看  $a$ 、 $b$ 、 $c$  随时间的变化情况--我们使用回归作为拟合工具，因此它们在统计上可能并不重要。检查  $\delta_{MV} - \delta_{BS}$  是否呈现（倒置的）抛物线形状，绘制 IV 预期变化与 Delta 的对比图。  
。
5.  $E[\Delta \sigma_{imp}]$  预期为负值，但可能不是。您实现的套期保值收益是否接近 15%？

**补充材料--将在项目研讨会（I 或 II）上分发经整理的相关文章小集。模块 3 中与波动性相关的核心讲座也支持本专题 DH。**

# 混合分类组合

## 摘要

趋势预测对投资管理很有价值，因为准确的预测可以确保资产管理者跑赢市场。趋势预测可以利用机器学习算法进行有效建模；然而，选择采用何种学习技术仍然是一项具有挑战性的任务。集合学习是一种功能强大的机器学习算法，它将多个机器学习模型的预测结果结合在一起，通过利用多个模型的集体智慧来减少单个模型中可能存在的误差或偏差，从而实现更精确的预测。

## 目标

您的目标是利用混合组合技术建立一个预测正向移动（上升趋势）的模型。您提出的解决方案应包含详细的模型架构，并通过应用于交易策略的回溯测试进行评估。

- 从指数、股票、ETF、加密货币或商品中选择一个您感兴趣的代码。
- 只预测趋势，预测短期回报（例如：每日，6 小时）。将预测局限于二项式分类：因变量的最佳标签为 [0, 1]。避免使用 [-1, 1] 作为类标签。
- 分析应当全面，包括详细的特征工程、数据预处理、模型建立和评估。
- 本项目只使用机器学习算法。

**注：**您可以自由选择研究设计，以完成任务。您可以重新定义任务，预测动量符号（相对于回报符号）或波动方向。将探索范围限制在一种资产上。在每一个步骤中，都应该对所遵循的过程进行扩展和详细解释。仅提交 python 代码而不作适当解释的报告将不予接受。报告应详细介绍研究内容，并给出适当的结论。代码的可重现性是必须的，建议使用模块化编程方法。在本课题下，不得对现有指标、库或优化算法进行重新编码。

## 分步说明

1. 问题陈述应明确无误，包括基础资产的选择、数据集、时间范围和使用数据的频率。

- 如果要预测短期回报迹象（每日变动），那么最长 5 年的训练和测试就足够了。如果要预测 5D、10D 的股票回报率，或预测 1W、1M 的 Fama French 因子，则必须将所需数据增加到至少 10 年。

2. 执行详尽的特征工程 (FE)。

- FE 应该详细，包括列出衍生特征和指定获取/标记。制定如何对极小的近零收益进行分类的方法（从训练样本或正/负收益组中删除）。阈值在很大程度上取决于您的股票代码。例如：小于 0.25 的正收益率
- 应通过模型参数或标签定义来解决类别不平衡问题。
- 允许使用协整对和跨资产的特征，但在设计时应讲究策略。没有一套适用于所有资产的推荐特征集，但初始特征集应足够大。财务比率、高级技术指标（包括波动率估算器）和成交量信息可以预测价格方向。
- 可选项 使用新闻热图、信用利差（CDS）、财务比率历史数据、股息历史数据、主要利益相关者（董事交易）或大型基金的购买/处置数据或 Fama-French 因子数据可增强预测效果，这些数据可从您的专业订阅中获取。

3. 进行详细的探索性数据分析 (EDA) 。

- 通过更好地理解特征之间的关系和揭示潜在结构，EDA 有助于降低维度，并有助于检测/解释异常值。特征缩放技术的选择应由 EDA 决定。

4. 必须正确处理数据。使用不同的特征集、回溯长度和数据集需要进行清理和/或估算。

5. 应根据 EDA 进行特征转换。

- 预测因子之间应进行多重共线性分析。
- 展示特征之间关系的多散点图始终是个好主意。
- 大型特征集（包括重复类型和不同回溯）需要降低特征的维度。自组织图 (SOM)、K-Means 聚类或其他方法都可用于降维。避免对非线性数据集/预测因子使用主成分分析 (PCA)。

6. 进行广泛而详尽的模型构建。

- 堆叠模型的结构应至少包括三个基础学习器。
- 应优化每个基础学习器的超参数。
- 基础学习者的类型和使用的元模型是设计上的选择。

7. 应使用多种指标来评估您建议的分类器的性能，包括对应用于交易策略的预测信号进行回溯测试。

- 使用 AUC、混淆矩阵和分类报告（包括平衡准确率）调查预测质量。
- 应将预测信号应用到交易策略中进行评估。

\* \* \*



## 毕业设计阅读清单（选题）

本清单的目的不是提供教科书，而是提供更具体的章节和范文，您将从中获得专题知识/关键模型知识。下面列出的读物和其他精选书目将随项目研讨会和/或项目教程一起发布，文件名为 "主题 XX - 附加材料.zip"。

### 阅读清单：信贷组合

- 您很可能会重温 CDO 和 Copula 讲座材料，特别是说明椭圆 copula 密度和讨论 Cholesky 因式分解的第 48-52 张幻灯片。
- Peter Jaekel (2002 年) 所著的《Workshop and Monte Carlo Methods in Finance》教科书中载有 copula 算法的取样方法，请参见第 5 章。
- 相关灵敏度讲座》和 P. Jaekel (2002 年) 也介绍了等级相关系数。CR Topic Q&A 文件给出了明确的公式和解释。

### 阅读清单：投资组合构建

- CQF 关于优化基础知识和投资组合选择应用的讲座
- 托马斯-伊德佐雷克 (Thomas Idzorek) 2002 年出版的《布莱克-利特曼模式分步指南》介绍了实施该模式所需的基本知识。
- 布莱克-利特曼方法: Original Model and Extensions Attilio Meucci, 2010.  
<http://ssrn.com/abstract=1117574>
- Marcenko-Pastur 去噪/LW 非线性收缩成为可选方法。这两种方法的目的都是使协方差矩阵的均方差最优化受到噪声程度的干扰较小。可通过相关 FP 研讨会/教程提供资源。

### 阅读清单：协整对

- 金融时间序列建模》，E. Zivot 和 J. Wang，2002 年 - 我们分发了关于协整的第 12 章以及相关的项目研讨会（涉及误差修正方程中的术语）。
- 与其阅读冗长的计量经济学教科书，不如阅读《解释协整分析》：第一部分和第二部分，作者 David Hendry 和 Katarina Juselius，2000 年和 2001 年。能源期刊

- 本作品的附录解释了关键的计量经济学和 OU 过程数学联系，《学习和信任统计套利中的协整》，作者理查德-戴蒙德，*WILMOTT*

[papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2220092](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2220092)。

## 阅读清单：三角洲对冲和局部波动

- *期权的最优三角对冲*》，约翰-赫尔和艾伦-怀特著。《银行与金融期刊》（*Journal of Banking and Finance*），第 82 卷，2017 年 9 月（可在 SSRN 上查阅，随项目研讨会一起分发）。上一版本 2016 年。
- Dong Qu 所著的《*制造和管理客户驱动的衍生产品*》教科书--我们专门针对利率的局部波动模型分发了第 12 章（第 319-331 页）。
- *第 7 讲：局部波动性* 续埃马努埃尔-德曼（Emmanuel Derman）（2008 年）关于波动性微笑的讲座，我们仍然认为这很有用。在任何一本好的教科书中都可以读到有关局部波动性的一般知识。
- *高级波动率建模* 是 CQF 终身图书馆的选修课或校友工作坊（但可能涉及跳跃式波动率和随机波动率）。

## 阅读清单：深度学习与集合学习

- *利用综合深度学习系统预测短期股票市场价格趋势*  
沈静怡、M. Omair Shafiq 著，《大数据期刊》第 7 卷（2020 年）。
- Jimmy Ming-Tai Wu 等人（2021 年）*基于图形的 CNN-LSTM 股价预测算法*。
- Isaac Kofi Nti 著，《大数据期刊》第 7 卷（2020 年），*对用于股票市场预测的集合学习的综合评估*。
- *用于原油价格预测的混合集合学习模型*，Mahmudul Hasan 等著（2022 年）。  
[papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4153206](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4153206)。
- *高级机器学习 I 和 II* 可作为选修课或 CQF 终身图书馆的校友工作坊。